



Guide du développeur

AWS Lake Formation



AWS Lake Formation: Guide du développeur

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Qu'est-ce que c'est AWS Lake Formation ?	1
Caractéristiques de Lake Formation	2
Ingestion et gestion des données	2
Gestion de la sécurité	3
Intégration de vos données dans le catalogue de données	5
Comment ça marche	7
Flux de travail de gestion des autorisations de Lake Formation	7
Autorisations relatives aux métadonnées	9
Gestion de l'accès au stockage	12
Partage de données entre comptes dans Lake Formation	14
Composantes de la Lake Formation	15
Console Lake Formation	15
API Lake Formation et interface de ligne de commande	15
Autres AWS services	15
Terminologie des Lake Formation	16
Lac de données	16
Accès aux données	16
Mode d'accès hybride	16
Plan	17
Flux de travail	17
Catalogue de données	17
Données sous-jacentes	18
Principal	18
Administrateur du lac de données	18
AWS intégrations de services avec Lake Formation	18
Ressources supplémentaires sur la Formation du Lac	20
Blogs	21
Discussions techniques et webinaires	21
Architecture moderne	21
Ressources de maillage de données	21
Guides des meilleures pratiques	22
Débuter avec Lake Formation	22
Prise en main	23
Exécution des tâches AWS de configuration initiale	23

Inscrivez-vous pour un Compte AWS	23
Création d'un utilisateur doté d'un accès administratif	24
Octroi d'un accès par programmation	25
Configurez AWS Lake Formation	27
Configurer les ressources de Lake Formation à l'aide d' CloudFormation un modèle	28
Création d'un administrateur de lac de données	29
Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride	34
Attribuer des autorisations aux utilisateurs de Lake Formation	36
Configurer un emplacement Amazon S3 pour votre lac de données	37
(Facultatif) Paramètres de filtrage des données externes	38
(Facultatif) Accordez l'accès à la clé de chiffrement du catalogue de données	40
(Facultatif) Créez un rôle IAM pour les flux de travail	40
Mise à niveau des autorisations de AWS Glue données pour le modèle Lake Formation	42
À propos des autorisations par défaut	42
Répertorier les autorisations existantes	44
Configurer les autorisations de Lake Formation	46
Accorder aux utilisateurs des autorisations IAM	47
Passez au modèle d'autorisations de Lake Formation	48
Étape 5 : Sécuriser les nouvelles ressources du catalogue de données	51
Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM	52
Étape 7 : Nettoyer les politiques IAM existantes	53
Configuration des points de terminaison Amazon VPC (AWS PrivateLink)	53
Considérations relatives aux points de terminaison du VPC de la Lake Formation	54
Création d'un point de terminaison VPC d'interface pour Lake Formation	54
Création d'une politique de point de terminaison VPC pour Lake Formation	55
Tutoriels	57
Création d'un lac de données à partir d'une AWS CloudTrail source	58
Public visé	60
Conditions préalables	60
Étape 1 : créer un utilisateur d'analyste de données	61
Étape 2 : ajouter des autorisations pour lire les AWS CloudTrail journaux au rôle de flux de travail	62
Étape 3 : créer un compartiment Amazon S3 pour le lac de données	62
Étape 4 : enregistrer un chemin Amazon S3	63
Étape 5 : accorder des autorisations de localisation des données	63
Étape 6 : Création d'une base de données dans le catalogue de données	64

Étape 7 : Accorder des autorisations de données	64
Étape 8 : Utiliser un plan pour créer un flux de travail	65
Étape 9 : Exécuter le flux de travail	66
Étape 10 : Autorisez SELECT sur les tables	67
Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena	68
Création d'un lac de données à partir d'une source JDBC	69
Public visé	69
Conditions préalables	70
Étape 1 : créer un utilisateur d'analyste de données	71
Étape 2 : créer une connexion dans AWS Glue	72
Étape 3 : créer un compartiment Amazon S3 pour le lac de données	73
Étape 4 : enregistrer un chemin Amazon S3	73
Étape 5 : accorder des autorisations de localisation des données	73
Étape 6 : Création d'une base de données dans le catalogue de données	74
Étape 7 : Accorder des autorisations de données	74
Étape 8 : Utiliser un plan pour créer un flux de travail	75
Étape 9 : Exécuter le flux de travail	77
Étape 10 : Autorisez SELECT sur les tables	78
Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena	78
Étape 12 : interroger les données du lac de données à l'aide d'Amazon Redshift Spectrum ...	79
Étape 13 : Accorder ou révoquer les autorisations de Lake Formation à l'aide d'Amazon Redshift Spectrum	83
Configuration des autorisations pour les formats de table ouverts dans Lake Formation	84
Public visé	85
Conditions préalables	85
Étape 1 : Approvisionnez vos ressources	87
Étape 2 : configurer les autorisations pour une table Iceberg	88
Étape 3 : configurer les autorisations pour une table Hudi	95
Étape 4 : Configuration des autorisations pour une table Delta Lake	98
Étape 5 : Nettoyer les AWS ressources	100
Gestion d'un lac de données à l'aide d'un contrôle d'accès basé sur des balises	100
Public visé	102
Conditions préalables	103
Étape 1 : Approvisionnez vos ressources	104
Étape 2 : enregistrez l'emplacement de vos données, créez une ontologie LF-Tag et accordez des autorisations	105

Étape 3 : Création de bases de données Lake Formation	109
Étape 4 : Accorder les autorisations relatives aux tables	119
Étape 5 : exécuter une requête dans Amazon Athena pour vérifier les autorisations	121
Étape 6 : Nettoyer les AWS ressources	122
Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes	123
Public visé	123
Prérequis	124
Étape 1 : Approvisionnez vos ressources	125
Étape 2 : Requête sans filtres de données	126
Étape 3 : configurer les filtres de données et accorder des autorisations	128
Étape 4 : Requête à l'aide de filtres de données	130
Étape 5 : Nettoyer les AWS ressources	132
Partagez vos données en toute sécurité à l'aide de Lake Formation	132
Public visé	133
Configurer les paramètres de Lake Formation	135
Étape 1 : provisionnez vos ressources à l'aide AWS CloudFormation de modèles	137
Étape 2 : Conditions préalables au partage entre comptes de Lake Formation	140
Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises	142
Étape 4 : Implémentation de la méthode de ressource nommée	149
Étape 5 : Nettoyer les AWS ressources	153
Partage des ressources du catalogue de données avec des tiers à Comptes AWS l'aide d'un contrôle d'accès précis	153
Public visé	155
Prérequis	156
Étape 1 : Fournir un accès détaillé à un autre compte	157
Étape 2 : fournir un accès détaillé à un utilisateur du même compte	159
Permissions d'intégration à Lake Formation	160
Vue d'ensemble des autorisations relatives à Lake Formation	162
Méthodes de contrôle d'accès précis	164
Contrôle d'accès aux métadonnées	167
Contrôle d'accès aux données sous-jacent	171
Référence des personnalités de Lake Formation et des autorisations IAM	178
AWS Lake Formation personas	178
AWS politiques gérées pour Lake Formation	180
Autorisations suggérées par Persona	188

Modification des paramètres par défaut de votre lac de données	199
Permissions implicites de Lake Formation	202
Référence des autorisations de Lake Formation	204
Permissions de Lake Formation par type de ressource	204
Lake Formation accorde et AWS CLI révoque des commandes	207
Autorisations Lake Formation	212
Intégration d'IAM Identity Center	228
Conditions préalables à l'intégration d'IAM Identity Center à Lake Formation	229
Connecter Lake Formation à l'IAM Identity Center	233
Mise à jour de l'intégration d'IAM Identity Center	238
Suppression d'une connexion Lake Formation avec IAM Identity Center	239
Octroi d'autorisations aux utilisateurs et aux groupes	240
Intégration du contexte utilisateur d'IAM Identity Center dans les journaux CloudTrail	243
Ajouter un emplacement Amazon S3 à votre lac de données	245
Exigences relatives aux rôles utilisés pour enregistrer des sites	246
Enregistrement d'un emplacement Amazon S3	254
Enregistrement d'un emplacement Amazon S3 chiffré	258
Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte	263
Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes	266
Annulation de l'enregistrement d'un site Amazon S3	270
Mode d'accès hybride	271
Cas d'utilisation courants du mode d'accès hybride	273
Comment fonctionne le mode d'accès hybride	275
Configuration du mode d'accès hybride : scénarios courants	276
Supprimer les principes et les ressources du mode d'accès hybride	297
Affichage des principes et des ressources en mode d'accès hybride	298
Ressources supplémentaires	299
Création d'objets dans AWS Glue Data Catalog	299
Création d'un catalogue	301
Création d'une base de données	301
Création de tables	302
Vues du catalogue de données du bâtiment	311
Importation de données à l'aide de workflows	360
Plans et flux de travail	361
Création d'un flux de travail	363
Exécution d'un flux de travail	367

Intégration de vos données dans le catalogue de données	369
Fédération de catalogues	371
Caractéristiques de Catalog Federation	371
Fédérez pour Snowflake	372
Fédérer pour Databricks	380
Intégration des données Amazon Redshift dans le catalogue de données	389
Principaux avantages	392
Rôles et responsabilités	392
Conditions préalables	393
Création de catalogues fédérés Amazon Redshift	399
Affichage des objets du catalogue	409
Mettre à jour un catalogue fédéré	411
Accès à un catalogue fédéré partagé	413
Supprimer un catalogue fédéré	418
Interrogation de catalogues fédérés	419
Ressources supplémentaires	420
Fédération dans des sources de données externes	420
Flux de travail	421
Conditions préalables	422
Création d'un catalogue fédéré	426
Affichage des objets du catalogue	432
Supprimer un catalogue fédéré	434
Interrogation de catalogues fédérés	435
Ressources supplémentaires	435
Intégration des tables Amazon S3 au catalogue de données et à Lake Formation	435
Comment fonctionne l'intégration du catalogue de données et de la Lake Formation	437
Conditions préalables	439
Activation de l'intégration des tables Amazon S3	442
Création de bases de données et de tables	447
Enregistrement d'un compartiment de table Amazon S3 dans un autre AWS compte	450
Octroi d'autorisations	454
Modification des contrôles d'accès pour l'intégration des tables S3	457
Activez Lake Formation grâce à l'intégration des tables S3	457
Passage de Lake Formation à IAM	459
Création d'un catalogue géré par Amazon Redshift	463
Gestion des autorisations pour les données dans un partage de données Amazon Redshift	467

Prérequis	469
Configuration des autorisations pour les partages de données Amazon Redshift	470
Interrogation de bases de données fédérées	474
Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes	475
Flux de travail	478
Conditions préalables	479
Connexion du catalogue de données à un métastore Hive externe	481
Ressources supplémentaires	485
Gestion des autorisations relatives à Lake Formation	486
Octroi d'autorisations de localisation des données	486
Octroi d'autorisations de localisation des données (même compte)	487
Octroi d'autorisations de localisation des données (compte externe)	490
Octroi d'autorisations sur un emplacement de données partagé avec votre compte	493
Octroi d'autorisations de données	494
Autorisations IAM requises pour accorder les autorisations de Lake Formation	496
Utilisation de la méthode des ressources nommées	499
Contrôle d'accès basé sur les étiquettes	526
Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC	590
Contrôle d'accès basé sur les attributs	598
Exemple de scénario d'autorisations	602
Filtrage des données et sécurité au niveau des cellules	604
Filtres de données	606
Support partiQL dans les expressions de filtre de ligne	610
Autorisations requises pour interroger des tables avec filtrage au niveau des cellules	613
Gestion des filtres de données	613
Affichage des autorisations de base de données et de tables	629
Révocation des autorisations à l'aide de la console	634
Partage de données entre comptes	634
Conditions préalables	637
Mise à jour des paramètres de version de partage de données entre comptes	642
Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des entités IAM	649
Octroi d'autorisations sur une base de données ou une table partagée avec votre compte ..	655
Octroi d'autorisations relatives aux liens vers	657
Accès aux données sous-jacentes d'une table partagée	660

Journalisation entre comptes CloudTrail	662
Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation	667
Afficher toutes les subventions entre comptes à l'aide de l'opération GetResourceShares API	670
Accès aux tables et aux bases de données partagées du catalogue de données et affichage ..	671
Accepter une invitation à partager des AWS RAM ressources	673
Affichage des tables et des bases de données partagées du catalogue de données	675
Création de liens vers des ressources	677
Fonctionnement des liens vers des ressources	677
Création d'un lien de ressource vers une table partagée	680
Création d'un lien de ressource vers une base de données partagée	683
Gestion des liens de ressources dans AWS Glue APIs	687
Accès aux tables dans toutes les régions	691
Flux de travail	693
Configuration de l'accès aux tables entre régions	697
Sécurité	701
Protection des données	701
Chiffrement au repos	703
Sécurité de l'infrastructure	703
Prévention du problème de l'adjoint confus entre services	704
Enregistrement des événements de sécurité dans AWS Lake Formation	705
Intégration à Lake Formation	707
Utilisation de l'intégration de l'application Lake Formation	707
Comment fonctionne l'intégration de l'application Lake Formation	708
Rôles et responsabilités dans l'intégration de l'application Lake Formation	710
Lake Formationflux de travail pour les opérations d'API d'intégration d'applications	711
Enregistrement d'un moteur de requête tiers	713
Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications	715
Intégration des applications pour un accès complet aux tables	719
Collaboration avec d'autres AWS services	722
Amazon Athena	726
Support pour les formats de tables transactionnels	728
Ressources supplémentaires	731
Amazon Redshift Spectrum	731

Support pour les types de tables transactionnels	732
Ressources supplémentaires	734
AWS Glue	734
Support pour les types de tables transactionnels	735
Ressources supplémentaires	736
Amazon EMR	736
Support pour les formats de tables transactionnels	737
Ressources supplémentaires	738
Rapide	739
Ressources supplémentaires	739
AWS CloudTrail lac	739
Enregistrement des appels d'API AWS Lake Formation à l'aide de AWS CloudTrail	741
Informations sur la formation du lac en CloudTrail	741
Comprendre les événements liés à la Lake Formation	742
Tracer l'accès aux données CloudTrail grâce à Lake Formation	744
Meilleures pratiques, considérations et limites en matière de formation des lacs	747
Meilleures pratiques et considérations relatives au partage de données entre comptes	747
Limitations des rôles liés aux services	750
Limites d'accès aux données entre régions	751
Considérations et limites relatives aux affichages du catalogue de données	752
Limites du filtrage des données	753
Remarques et restrictions relatives au filtrage au niveau des colonnes	753
Limites du filtrage au niveau des cellules	755
Considérations et limites relatives au mode d'accès hybride	757
Limites liées à l'intégration des données de l'entrepôt de données Amazon Redshift dans le AWS Glue Data Catalog	758
Limites d'intégration du catalogue S3 Tables	760
Considérations et limites relatives au partage des données du magasin de métadonnées Hive	761
Limites du partage de données Amazon Redshift	763
Limites de l'intégration d'IAM Identity Center	765
Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation	766
Considérations relatives au contrôle d'accès basé sur les attributs, limites et régions prises en charge	769
Résolution des problèmes liés à la formation du Lake	770

Résolution de problème généraux	770
Erreur : autorisations insuffisantes pour Lake Formation sur <Amazon S3 location>	770
Erreur : « Autorisations de clé de chiffrement insuffisantes pour l'API Glue »	771
Ma requête Amazon Athena ou celle d'Amazon Redshift qui utilise des manifestes échoue .	771
Erreur : « Autorisations de formation lacustres insuffisantes : création d'une balise requise dans le catalogue »	771
Erreur lors de la suppression d'administrateurs de lacs de données non valides	771
Résolution des problèmes d'accès entre comptes	771
J'ai accordé l'autorisation d'utiliser plusieurs comptes Lake Formation, mais le destinataire ne peut pas voir la ressource	772
Les utilisateurs principaux du compte destinataire peuvent voir la ressource du catalogue de données, mais ne peuvent pas accéder aux données sous-jacentes	773
Erreur : « L'association a échoué car l'appelant n'était pas autorisé » lors de l'acceptation d'une invitation de partage de AWS RAM ressources	773
Erreur : « Non autorisé à accorder des autorisations pour la ressource »	774
Erreur : « Accès refusé pour récupérer les informations de AWS l'organisation »	774
Erreur : « Organisation <organization-ID>introuvable »	774
Erreur : « Permissions de formation lacustres insuffisantes : combinaison illégale »	774
ConcurrentModificationException sur les grant/revoke demandes adressées à des comptes externes	774
Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes	775
Résolution des problèmes liés aux plans et aux flux de travail	776
<role-ARN>Mon plan a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à exécuter : iam : PassRole on ressource : »	776
<role-ARN>Mon flux de travail a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à effectuer : iam : PassRole on ressource : »	777
Un crawler de mon flux de travail a échoué avec le message « La ressource n'existe pas ou le demandeur n'est pas autorisé à accéder aux autorisations demandées »	777
Un crawler de mon flux de travail a échoué avec « Une erreur s'est produite (AccessDeniedException) lors de l'appel de l' CreateTable opération... »	777
Problèmes connus pour AWS Lake Formation	777
Limitation du filtrage des métadonnées des tables	778
Problème lié au changement de nom d'une colonne exclue	779
Problème lié à la suppression de colonnes dans les tableaux CSV	779
Les partitions de table doivent être ajoutées sous un chemin commun	779

Problème lié à la création d'une base de données pendant la création du flux de travail	780
Problème lié à la suppression puis à la recreation d'un utilisateur	780
Les opérations de l'API Data Catalog ne mettent pas à jour la valeur du IsRegisteredWithLakeFormation paramètre	780
Les opérations de Lake Formation ne prennent pas en charge AWS Glue le registre des schémas	781
Message d'erreur mis à jour	781
Lake Formation API	782
Autorisations	783
— operations —	783
— les types de données —	783
Paramètres du lac de données	784
— operations —	784
— les types de données —	784
Intégration à IAM Identity Center	784
— operations —	784
— les types de données —	784
Mode d'accès hybride	785
— operations —	785
— les types de données —	783
Vente d'informations d'identification	785
— operations —	785
— les types de données —	786
Identification	786
— operations —	786
— les types de données —	787
Filtre de données APIs	787
— operations —	787
— les types de données —	787
Types de données courants	788
ErrorDetail	788
Modèles de chaîne	788
Régions prises en charge	790
Disponibilité générale	790
AWS GovCloud (US)	790
Optimisation des transactions et du stockage	790

Historique du document 793

AWS Glossaire 810

..... dcccxi

Qu'est-ce que c'est AWS Lake Formation ?

Bienvenue dans le guide du AWS Lake Formation développeur.

AWS Lake Formation vous permet de gérer, de sécuriser et de partager les données de manière centralisée à l'échelle mondiale à des fins d'analyse et d'apprentissage automatique. Avec Lake Formation, vous pouvez gérer un contrôle d'accès précis pour les données de vos lacs de données sur Amazon Simple Storage Service (Amazon S3) et ses métadonnées dans AWS Glue Data Catalog

Lake Formation fournit son propre modèle d'autorisations qui complète le modèle d'autorisations IAM. Le modèle d'autorisations de Lake Formation permet un accès précis aux données stockées dans des lacs de données ainsi qu'à des sources de données externes telles que les entrepôts de données Amazon Redshift Amazon DynamoDB , les bases de données et les sources de données tierces par le biais d'un simple mécanisme d'autorisation ou de révocation, un peu comme un système de gestion de base de données relationnelle (RDBMS). Les autorisations de Lake Formation sont appliquées à l'aide de contrôles granulaires au niveau des colonnes, des lignes et des cellules dans les services d' AWS analyse et d'apprentissage automatique, notamment Amazon Athena, Amazon Redshift Spectrum, Amazon Quick Amazon EMR et. AWS Glue

Avec le mode d'accès hybride de Lake Formation pour AWS Glue Data Catalog (Data Catalog), vous pouvez sécuriser et accéder aux données cataloguées en utilisant à la fois les autorisations Lake Formation et les politiques d'autorisations IAM pour Amazon S3 et AWS Glue les actions. Grâce au mode d'accès hybride, les administrateurs de données peuvent intégrer les autorisations de Lake Formation de manière sélective et progressive, en se concentrant sur un cas d'utilisation de lac de données à la fois.

Lake Formation vous permet également de partager des données en interne et en externe entre plusieurs AWS organisations ou directement avec les responsables d'IAM sur un autre compte Comptes AWS, offrant ainsi un accès détaillé aux métadonnées du catalogue de données et aux données sous-jacentes.

Rubriques

- [Caractéristiques de Lake Formation](#)
- [AWS Lake Formation : comment ça marche](#)
- [Composantes de la Lake Formation](#)
- [Terminologie des Lake Formation](#)

- [AWS intégrations de services avec Lake Formation](#)
- [Ressources supplémentaires sur la Formation du Lac](#)
- [Débuter avec Lake Formation](#)

Caractéristiques de Lake Formation

Lake Formation vous aide à découpler les données et à combiner différents types de données structurées et non structurées dans un référentiel centralisé. Tout d'abord, identifiez les banques de données existantes dans Amazon S3 ou dans les bases de données relationnelles et NoSQL, puis déplacez les données vers votre lac de données. Ensuite, analysez, cataloguez et préparez les données à des fins d'analyse. Ensuite, offrez à vos utilisateurs un accès sécurisé en libre-service aux données grâce aux services d'analyse de leur choix.

Vous pouvez utiliser la console Lake Formation pour créer des catalogues fédérés à plusieurs niveaux dans le catalogue de données et unifier les données entre les lacs de données Amazon S3 et les entrepôts de données Amazon Redshift. Vous pouvez également intégrer des données provenant de vos bases de données opérationnelles telles que Amazon DynamoDB, et de sources de données tierces telles que Google BigQuery, MySQL, entre autres. Le catalogue de données fournit un référentiel de métadonnées centralisé qui facilite la gestion et la découverte de données sur des systèmes disparates.

Pour de plus amples informations, veuillez consulter [Intégrer vos données dans AWS Glue Data Catalog](#).

Rubriques

- [Ingestion et gestion des données](#)
- [Gestion de la sécurité](#)
- [Intégration de vos données dans le catalogue de données](#)

Ingestion et gestion des données

Importer des données à partir de bases de données déjà présentes AWS

Une fois que vous avez indiqué où se trouvent vos bases de données existantes et que vous avez fourni vos identifiants d'accès, Lake Formation lit les données et leurs métadonnées (schéma) pour comprendre le contenu de la source de données. Il importe ensuite les données dans votre nouveau lac de données et enregistre les métadonnées dans un catalogue central. Avec Lake Formation,

vous pouvez importer des données depuis des bases de données MySQL, PostgreSQL, SQL Server, MariaDB et Oracle exécutées sur Amazon RDS ou hébergées sur Amazon EC2. Le chargement de données en masse et incrémentiel est pris en charge.

Importer des données depuis d'autres sources externes

Vous pouvez utiliser Lake Formation pour déplacer des données depuis des bases de données locales en vous connectant à Java Database Connectivity (JDBC). Identifiez vos sources cibles et fournissez des informations d'accès dans la console. Lake Formation lit et charge vos données dans le lac de données. Pour importer des données à partir de bases de données autres que celles répertoriées ci-dessus, vous pouvez créer des tâches ETL personnalisées avec AWS Glue.

Cataloguez et étiquetez vos données

Vous pouvez utiliser des AWS Glue robots d'exploration pour lire vos données dans Amazon S3, extraire le schéma de base de données et de table et stocker ces données dans un catalogue de données consultable. Utilisez ensuite Lake Formation [Contrôle d'accès basé sur des balises Lake Formation](#) (TBAC) pour gérer les autorisations sur les bases de données, les tables et les colonnes. Pour plus d'informations sur l'ajout de tables au catalogue de données, consultez [Création d'objets dans AWS Glue Data Catalog](#).

Gestion de la sécurité

Définissez et gérez les contrôles d'accès

Lake Formation fournit un endroit unique pour gérer les contrôles d'accès aux données de votre lac de données. Vous pouvez définir des politiques de sécurité qui limitent l'accès aux données au niveau de la base de données, de la table, de la colonne, de la ligne et de la cellule. Ces politiques s'appliquent aux utilisateurs et aux rôles IAM, ainsi qu'aux utilisateurs et aux groupes lors de la fédération via un fournisseur d'identité externe. Vous pouvez utiliser des contrôles précis pour accéder aux données sécurisées par Lake Formation dans Amazon Redshift Spectrum, Athena, ETL et Amazon EMR pour AWS Glue Apache Spark. Chaque fois que vous créez des identités IAM, veillez à suivre les meilleures pratiques IAM. Pour plus d'informations, consultez [la section Bonnes pratiques en matière de sécurité](#) dans le guide de l'utilisateur IAM.

Mode d'accès hybride

Le mode d'accès hybride Lake Formation offre la flexibilité nécessaire pour activer de manière sélective les autorisations Lake Formation pour les bases de données et les tables de votre catalogue de données. Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui

vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants. Pour de plus amples informations, veuillez consulter [Mode d'accès hybride](#).

Mettre en œuvre l'enregistrement des audits

Lake Formation fournit des journaux d'audit complets CloudTrail pour surveiller l'accès et démontrer la conformité aux politiques définies de manière centralisée. Vous pouvez auditer l'historique des accès aux données par le biais de services d'analyse et d'apprentissage automatique qui lisent les données de votre lac de données via Lake Formation. Cela vous permet de voir quels utilisateurs ou quels rôles ont tenté d'accéder à quelles données, avec quels services et à quel moment. Vous pouvez accéder aux journaux d'audit de la même manière que vous accédez à tous CloudTrail les autres journaux à l'aide de la console CloudTrail APIs and. Pour plus d'informations sur les CloudTrail journaux, consultez [Enregistrement des appels d'API AWS Lake Formation à l'aide de AWS CloudTrail](#).

Sécurité au niveau des lignes et des cellules

Lake Formation fournit des filtres de données qui vous permettent de restreindre l'accès à une combinaison de colonnes et de lignes. Utilisez la sécurité au niveau des lignes et des cellules pour protéger les données sensibles telles que les informations personnelles identifiables (PII). Pour plus d'informations sur la sécurité au niveau des lignes, consultez. [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#)

Contrôle d'accès basé sur les étiquettes

Utilisez le [contrôle d'accès basé sur les attributs](#) de Lake Formation pour gérer des centaines, voire des milliers d'autorisations de données en créant des étiquettes personnalisées appelées balises LF. Vous pouvez désormais définir des balises LF et les associer à des bases de données, à des tables ou à des colonnes. Partagez ensuite l'accès contrôlé entre les services d'analyse, d'apprentissage automatique (ML) et d'extraction, de transformation et de chargement (ETL) à des fins de consommation. Les balises LF permettent d'étendre facilement la gouvernance des données en remplaçant les définitions de politique de milliers de ressources par quelques balises logiques. Lake Formation propose une recherche textuelle sur ces métadonnées, afin que vos utilisateurs puissent trouver rapidement les données à analyser.

Contrôle d'accès basé sur les attributs

Utilisez le [contrôle d'accès basé sur les attributs](#) pour accorder l'accès aux objets du catalogue de données. Le contrôle d'accès basé sur les attributs (ABAC) est une stratégie d'autorisation qui définit

les autorisations en fonction des attributs. AWS appelle ces balises d'attributs. Vous pouvez utiliser ABAC pour accorder l'accès aux principaux au sein du même compte ou d'un autre compte sur les ressources du catalogue de données. Tout principal IAM dont les clés et les valeurs de balise IAM ou de balise de session correspondent accède à la ressource. Vous devez disposer d'autorisations pouvant être accordées sur les ressources pour octroyer ces subventions.

Accès entre comptes

Les fonctionnalités de gestion des autorisations de Lake Formation simplifient la sécurisation et la gestion des lacs de données distribués sur plusieurs AWS comptes grâce à une approche centralisée, fournissant un contrôle d'accès précis au catalogue de données et aux sites Amazon S3. Pour de plus amples informations, veuillez consulter [Partage de données entre comptes dans Lake Formation](#).

Intégration de vos données dans le catalogue de données

La fonctionnalité de fédération vous permet de créer des catalogues fédérés et de configurer des autorisations sur les ensembles de données stockés dans différentes sources de données telles qu'Amazon Redshift sans migrer les données ou les métadonnées vers Amazon S3 ou AWS Glue Data Catalog. Vous pouvez utiliser les méthodes suivantes pour importer des données et gérer les autorisations sur des ensembles de données externes dans Lake Formation :

Pour plus d'informations, consultez la [section Transférer vos données dans le AWS Glue Data Catalog](#).

- Intégrer les données des entrepôts de données Amazon Redshift dans le AWS Glue Data Catalog — Enregistrez un espace de noms [Amazon Redshift](#) existant ou un cluster dans le catalogue de données, et créez un catalogue fédéré à plusieurs niveaux dans le catalogue de données.

Vous pouvez accéder à vos données à l'aide de n'importe quel moteur de requête compatible avec les spécifications OpenAPI du catalogue REST Apache Iceberg, tel qu'Amazon EMR Serverless et Amazon Athena.

Pour de plus amples informations, veuillez consulter [Intégrer les données Amazon Redshift dans AWS Glue Data Catalog](#).

- Fédération dans le catalogue de données à partir de sources de données externes : connectez le catalogue de données à des sources de données externes à l'aide de AWS Glue connexions, et créez des catalogues fédérés pour gérer de manière centralisée les autorisations d'accès aux

ensembles de données à l'aide de Lake Formation. Aucune migration de métadonnées dans le catalogue de données n'est nécessaire.

Pour de plus amples informations, veuillez consulter [Fédération en sources de données externes dans le AWS Glue Data Catalog](#).

- Intégration des compartiments de tables Amazon S3 au catalogue de données : vous pouvez publier et cataloguer les tables Amazon S3 sous forme d'objets du catalogue de données et enregistrer le catalogue en tant que localisation des données de Lake Formation depuis la console Lake Formation ou en utilisant AWS Glue APIs.

Pour de plus amples informations, veuillez consulter [Intégration d'Amazon S3 Tables avec AWS Glue Data Catalog et AWS Lake Formation](#).

- Créez des catalogues pour gérer les tables Amazon Redshift dans le catalogue de données : vous ne disposez peut-être pas d'un cluster de producteurs Amazon Redshift ou d'un partage de données Amazon Redshift actuellement, mais vous souhaitez créer et gérer des tables Amazon Redshift à l'aide de Data Catalog. Vous pouvez commencer par créer un catalogue AWS Glue géré à l'aide de l'`glue:CreateCatalogAPI` ou de la AWS Lake Formation console en définissant le type de catalogue Catalog source comme Redshift. Managed

Pour de plus amples informations, veuillez consulter [Création d'un catalogue géré par Amazon Redshift dans AWS Glue Data Catalog](#).

- Intégration de Lake Formation au partage de données Amazon Redshift : utilisez Lake Formation pour gérer de manière centralisée les autorisations d'accès aux bases de données, aux tables, aux colonnes et aux lignes des partages de données Amazon Redshift et pour restreindre l'accès des [utilisateurs](#) aux objets d'un partage de données.
- Connexion du catalogue de données à des métastores externes : connectez-vous AWS Glue Data Catalog à des métastores externes pour gérer les autorisations d'accès aux ensembles de données dans Amazon S3 à l'aide de Lake Formation. Aucune migration de métadonnées dans le catalogue de données n'est nécessaire.

Pour de plus amples informations, veuillez consulter [Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes](#).

- Intégrer Lake Formation à AWS Data Exchange — Lake Formation prend en charge l'octroi de licences d'accès à vos données via AWS Data Exchange. Si vous souhaitez obtenir une licence pour vos données de Lake Formation, [consultez AWS Data Exchange le](#) guide de l'AWS Data Exchange utilisateur.

AWS Lake Formation : comment ça marche

AWS Lake Formation fournit un modèle d'autorisations du système de gestion de base de données relationnelle (RDBMS) pour accorder ou révoquer l'accès aux ressources du catalogue de données telles que les bases de données, les tables et les colonnes contenant des données sous-jacentes dans Amazon S3. Les autorisations Lake Formation, faciles à gérer, remplacent les politiques complexes relatives aux compartiments Amazon S3 et les politiques IAM correspondantes.

Dans Lake Formation, vous pouvez implémenter des autorisations à deux niveaux :

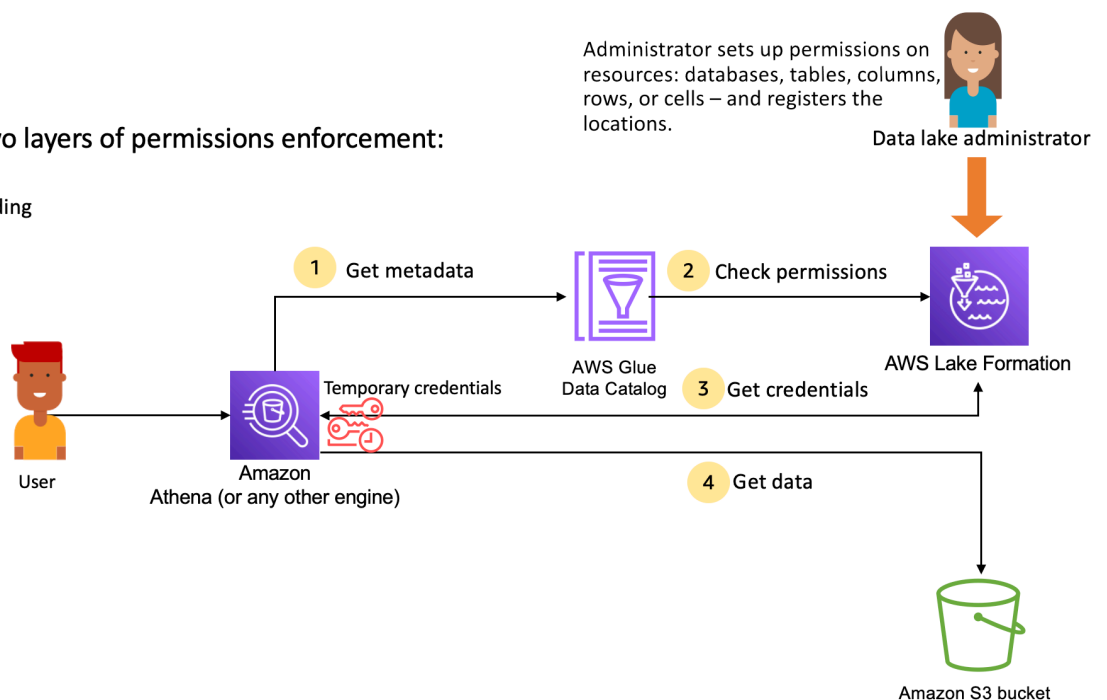
- Application des autorisations au niveau des métadonnées sur les ressources du catalogue de données, telles que les bases de données et les tables
- Gestion des autorisations d'accès au stockage sur les données sous-jacentes stockées dans Amazon S3 pour le compte de moteurs intégrés

Flux de travail de gestion des autorisations de Lake Formation

Lake Formation s'intègre aux moteurs d'analyse pour interroger les magasins de données Amazon S3 et les objets de métadonnées enregistrés auprès de Lake Formation. Le schéma suivant illustre le fonctionnement de la gestion des autorisations dans Lake Formation.

Lake Formation provides two layers of permissions enforcement:

- Metadata layer – Data Catalog
- Storage layer – Credential vending



Étapes de haut niveau de la gestion des autorisations de Lake Formation

Avant que Lake Formation ne puisse fournir des contrôles d'accès aux données de votre lac de données, un [administrateur de lac de données](#) ou un utilisateur disposant d'autorisations administratives définit des politiques utilisateur individuelles pour les tables du catalogue de données afin d'autoriser ou de refuser l'accès aux tables du catalogue de données à l'aide des autorisations de Lake Formation.

Ensuite, l'administrateur du lac de données ou un utilisateur délégué par l'administrateur accorde des autorisations Lake Formation aux utilisateurs sur les bases de données et les tables du catalogue de données, et enregistre l'emplacement de la table sur Amazon S3 auprès de Lake Formation.

1. Obtenir des métadonnées — Un principal (utilisateur) envoie une requête ou un script ETL à un [moteur d'analyse intégré](#) tel qu'Amazon Athena, AWS Glue Amazon EMR ou Amazon Redshift Spectrum. Le moteur d'analyse intégré identifie la table demandée et envoie une demande de métadonnées au catalogue de données.
2. Vérifier les autorisations — Le catalogue de données vérifie les autorisations de l'utilisateur auprès de Lake Formation, et si l'utilisateur est autorisé à accéder à la table, renvoie les métadonnées qu'il est autorisé à voir au moteur.
3. Obtenir des informations d'identification — Le catalogue de données indique au moteur si la table est gérée par Lake Formation ou non. Si les données sous-jacentes sont enregistrées auprès de Lake Formation, le moteur d'analyse demande à Lake Formation de fournir un accès temporaire aux données.
4. Obtenir des données — Si l'utilisateur est autorisé à accéder à la table, Lake Formation fournit un accès temporaire au moteur d'analyse intégré. À l'aide de l'accès temporaire, le moteur d'analyse extrait les données d'Amazon S3 et effectue le filtrage nécessaire, tel que le filtrage par colonne, ligne ou cellule. Lorsque le moteur a terminé d'exécuter la tâche, il renvoie les résultats à l'utilisateur. Ce processus s'appelle la vente [d'informations d'identification](#).

Si la table n'est pas gérée par Lake Formation, le deuxième appel du moteur d'analyse est directement envoyé à Amazon S3. La politique de compartiment Amazon S3 et la politique utilisateur IAM concernées sont évaluées pour l'accès aux données.

Chaque fois que vous utilisez des politiques IAM, veillez à respecter les bonnes pratiques IAM. Pour plus d'informations, consultez la rubrique [Bonnes pratiques IAM](#) du Guide de l'utilisateur IAM.

Rubriques

- [Autorisations relatives aux métadonnées](#)
- [Gestion de l'accès au stockage](#)
- [Partage de données entre comptes dans Lake Formation](#)

Autorisations relatives aux métadonnées

Lake Formation fournit l'autorisation et le contrôle d'accès au catalogue de données. Lorsqu'un rôle IAM appelle l'API du catalogue de données depuis n'importe quel système, le catalogue de données vérifie les autorisations de données de l'utilisateur et renvoie uniquement les métadonnées auxquelles l'utilisateur est autorisé à accéder. Par exemple, si un rôle IAM n'a accès qu'à une seule table dans une base de données et qu'un service ou un utilisateur assumant le rôle effectue l'GetTablesopération, la réponse ne contiendra qu'une seule table, quel que soit le nombre de tables de la base de données.

Paramètres par défaut - autorisations **IAMAllowedPrincipal** de groupe

AWS Lake Formation, par défaut, attribue des autorisations à toutes les bases de données et tables à un groupe virtuel nommé **IAMAllowedPrincipal**. Ce groupe est unique et visible uniquement au sein de Lake Formation. Le **IAMAllowedPrincipal** groupe inclut tous les principaux IAM qui ont accès aux ressources du catalogue de données via les politiques principales IAM et AWS Glue les politiques de ressources. Si cette autorisation existe sur une base de données ou une table, tous les principaux auront accès à la base de données ou à la table.

Si vous souhaitez fournir des autorisations plus détaillées sur une base de données ou une table, supprimez **IAMAllowedPrincipal** l'autorisation et Lake Formation appliquera toutes les autres politiques associées à cette base de données ou table. Par exemple, s'il existe une politique qui permet à l'utilisateur A d'accéder à la base de données A avec **DESCRIBE** des autorisations, et **IAMAllowedPrincipal** qu'elle existe avec toutes les autorisations, l'utilisateur A continuera à effectuer toutes les autres actions jusqu'à ce que l'**IAMAllowedPrincipal** autorisation soit révoquée.

En outre, par défaut, le **IAMAllowedPrincipal** groupe dispose d'autorisations sur toutes les nouvelles bases de données et tables lors de leur création. Deux configurations contrôlent ce comportement. Le premier est au niveau du compte et de la région, ce qui permet cela pour les bases de données nouvellement créées, et le second au niveau de la base de données. Pour modifier le paramètre par défaut, voir [Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride](#).

Octroi d'autorisations

Les administrateurs des lacs de données peuvent accorder des autorisations de catalogue de données aux principaux afin que ceux-ci puissent créer et gérer des bases de données et des tables, et accéder aux données sous-jacentes.

Autorisations au niveau de la base de données et de la table

Lorsque vous accordez des autorisations au sein de Lake Formation, le concédant doit spécifier le principal auquel les autorisations doivent être accordées, les ressources pour lesquelles les autorisations doivent être accordées et les actions que le bénéficiaire doit avoir accès pour effectuer. Pour la plupart des ressources de Lake Formation, la liste principale et les ressources auxquelles accorder des autorisations sont similaires, mais les actions qu'un bénéficiaire peut effectuer varient en fonction du type de ressource. Par exemple, `SELECT` des autorisations sont disponibles pour les tables pour lire les tables, mais `SELECT` pas pour les bases de données. L'`CREATE_TABLE` autorisation est autorisée sur les bases de données, mais pas sur les tables.

Vous pouvez accorder AWS Lake Formation des autorisations de deux manières :

- [Méthode de ressource nommée](#) : vous permet de choisir les noms de base de données et de tables tout en accordant des autorisations aux utilisateurs.
- [Contrôle d'accès basé sur les balises LF \(LF-TBAC\)](#) : les utilisateurs créent des balises LF, les associent aux ressources du catalogue de données, accordent des autorisations sur les balises LF, associent des autorisations à des utilisateurs individuels et rédigent des politiques d'`Describe` autorisation LF à l'aide de balises LF destinées à différents utilisateurs. Ces LF-Tag-based politiques s'appliquent à toutes les ressources du catalogue de données associées à ces valeurs LF-Tag.

Note

Les balises LF sont propres à Lake Formation. Ils ne sont visibles que dans Lake Formation et ne doivent pas être confondus avec les balises de AWS ressources.

Le LF-TBAC est une fonctionnalité qui permet aux utilisateurs de regrouper les ressources dans des catégories définies par l'utilisateur de balises LF et d'appliquer des autorisations à ces groupes de ressources. C'est donc le meilleur moyen d'étendre les autorisations à un grand nombre de ressources du catalogue de données.

Pour de plus amples informations, veuillez consulter [Contrôle d'accès basé sur des balises Lake Formation](#).

Lorsque vous accordez des autorisations à un directeur, Lake Formation évalue les autorisations comme une union de toutes les politiques relatives à cet utilisateur. Par exemple, si vous avez deux politiques sur une table pour un principal où l'une accorde des autorisations aux colonnes col1, col2 et col3 par le biais d'une méthode de ressource nommée, et l'autre politique accorde des autorisations à la même table et au même principal à col5, et col6 via des balises LF, les autorisations effectives seront une union des autorisations qui seraient col1, col2, col3, col5 et col6. Cela inclut également les filtres de données et les lignes.

Autorisations de localisation des données

Les autorisations de localisation des données permettent aux utilisateurs non administrateurs de créer des bases de données et des tables sur des sites Amazon S3 spécifiques. Si un utilisateur tente de créer une base de données ou une table dans un emplacement qu'il n'est pas autorisé à créer, la tâche de création échoue. Cela permet d'empêcher les utilisateurs de créer des tables à des emplacements arbitraires dans le lac de données et de contrôler les endroits où ces utilisateurs peuvent lire et écrire des données. Il existe une autorisation implicite lors de la création de tables dans l'emplacement Amazon S3 au sein de la base de données dans laquelle elles sont créées. Pour de plus amples informations, veuillez consulter [Octroi d'autorisations de localisation des données](#).

Créer des autorisations de table et de base de données

Par défaut, les utilisateurs non administrateurs ne sont pas autorisés à créer des bases de données ou des tables au sein d'une base de données. La création de bases de données est contrôlée au niveau du compte à l'aide des paramètres de Lake Formation afin que seuls les principaux autorisés puissent créer des bases de données. Pour de plus amples informations, veuillez consulter [Création d'une base de données](#). Pour créer une table, le principal a besoin d'une CREATE_TABLE autorisation sur la base de données dans laquelle la table est créée. Pour de plus amples informations, veuillez consulter [Création de tables](#).

Autorisations implicites et explicites

Lake Formation fournit des autorisations implicites en fonction du personnage et des actions qu'il effectue. Par exemple, les administrateurs des lacs de données obtiennent automatiquement DESCRIBE des autorisations pour toutes les ressources du catalogue de données, des autorisations de localisation des données pour tous les emplacements, des autorisations pour créer des bases

de données et des tables dans tous les emplacements, ainsi que Grant Revoke des autorisations sur n'importe quelle ressource. Les créateurs de bases de données obtiennent automatiquement toutes les autorisations de base de données sur les bases de données qu'ils créent, et les créateurs de tables obtiennent toutes les autorisations sur les tables qu'ils créent. Pour de plus amples informations, veuillez consulter [Permissions implicites de Lake Formation](#).

Autorisations pouvant être accordées

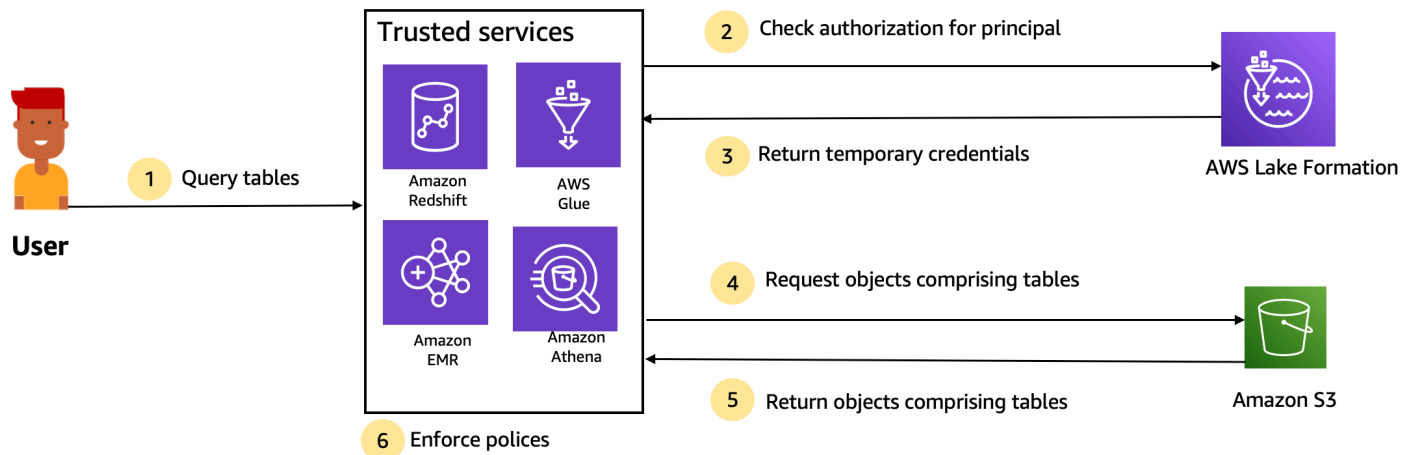
Les administrateurs de data lake ont la possibilité de déléguer la gestion des autorisations à des utilisateurs non administratifs en fournissant des autorisations pouvant être accordées. Lorsqu'un principal reçoit des autorisations pouvant être accordées sur une ressource et un ensemble d'autorisations, ce principal peut accorder des autorisations à d'autres principaux sur cette ressource.

Gestion de l'accès au stockage

Lake Formation utilise la fonctionnalité de distribution [automatique d'informations d'identification](#) pour fournir un accès temporaire aux données Amazon S3. La vente d'informations d'identification, ou de jetons, est un modèle courant qui fournit des informations d'identification temporaires aux utilisateurs, aux services ou à une autre entité dans le but d'accorder un accès à court terme à une ressource.

Lake Formation s'appuie sur ce modèle pour fournir un accès à court terme à des services AWS d'analyse tels qu'Athena afin d'accéder aux données pour le compte du principal appelant. Lorsqu'ils accordent des autorisations, les utilisateurs n'ont pas besoin de mettre à jour leurs politiques de compartiment Amazon S3 ou leurs politiques IAM, et ils n'ont pas besoin d'un accès direct à Amazon S3.

Le schéma suivant montre comment Lake Formation fournit un accès temporaire aux sites enregistrés :



Trusted services enforce AWS Lake Formation policies (distributed enforcement with fail close).

1. Un principal (utilisateur) saisit une requête ou une demande de données pour une table par le biais d'un service intégré fiable tel qu'Athena, Amazon EMR, Redshift Spectrum ou. AWS Glue
2. Le service intégré vérifie l'autorisation de Lake Formation pour le tableau et les colonnes demandées et prend une décision d'autorisation. Si l'utilisateur n'est pas autorisé, Lake Formation refuse l'accès aux données et la requête échoue.
3. Une fois l'autorisation réussie et l'autorisation de stockage activée pour la table et l'utilisateur, le service intégré récupère les informations d'identification temporaires de Lake Formation pour accéder aux données.
4. Le service intégré utilise les informations d'identification temporaires de Lake Formation pour demander des objets à Amazon S3.
5. Amazon S3 fournit les objets Amazon S3 au service intégré. Les objets Amazon S3 contiennent toutes les données de la table.
6. Le service intégré assure l'application nécessaire des politiques relatives à la Formation des Lacs, telles que le filtrage au niveau des colonnes, au niveau des lignes et/ou au niveau des cellules. Le service intégré traite les requêtes et renvoie les résultats à l'utilisateur.

Activer l'application des autorisations au niveau du stockage pour les tables du catalogue de données

Par défaut, l'application au niveau du stockage n'est pas activée pour les tables du catalogue de données. Pour permettre l'application au niveau du stockage, vous devez enregistrer l'emplacement Amazon S3 de vos données sources auprès de Lake Formation et fournir un rôle IAM. Les autorisations au niveau du stockage seront activées pour toutes les tables ayant le même chemin d'emplacement de table ou le même préfixe que l'emplacement Amazon S3.

Lorsqu'un service intégré demande l'accès à l'emplacement des données pour le compte d'un utilisateur, le service Lake Formation assume ce rôle et renvoie les informations d'identification au service demandé avec des autorisations limitées sur la ressource afin que l'accès aux données puisse être effectué. Le rôle IAM enregistré doit disposer de tous les accès requis à l'emplacement Amazon S3, y compris AWS KMS les clés.

Pour de plus amples informations, veuillez consulter [Enregistrement d'un emplacement Amazon S3](#).

AWS Services pris en charge

AWS des services analytiques tels qu'Athena, Redshift Spectrum, Amazon AWS Glue EMR Amazon Quick et Amazon SageMaker AI s'intègrent à AWS Lake Formation à l'aide des opérations de l'API de vente automatique d'informations d'identification de Lake Formation. Pour consulter la liste complète des AWS services intégrés à Lake Formation, ainsi que le niveau de granularité et les formats de table qu'ils prennent en charge, voir [Collaboration avec d'autres AWS services](#).

Partage de données entre comptes dans Lake Formation

Avec Lake Formation, vous pouvez partager les ressources du catalogue de données (bases de données et tables) au sein d'un AWS compte et entre les comptes dans une configuration simple à l'aide de la méthode des ressources nommées ou des balises LF. Vous pouvez partager une base de données complète ou sélectionner des tables d'une base de données avec tous les principaux IAM (rôles et utilisateurs IAM) d'un compte, vers d'autres AWS comptes au niveau du compte ou directement avec les principaux IAM d'un autre compte.

Vous pouvez également partager les tables du catalogue de données avec des filtres de données afin de restreindre l'accès aux détails au niveau des lignes et des cellules. Lake Formation utilise AWS Resource Access Manager (AWS RAM) pour faciliter l'octroi d'autorisations entre comptes. Lorsqu'une ressource est partagée entre deux comptes, AWS RAM envoie des invitations au compte du destinataire. Lorsqu'un utilisateur accepte une invitation de AWS RAM partage, AWS RAM fournit les autorisations nécessaires à Lake Formation pour que les ressources du catalogue de données soient disponibles et pour activer l'application des niveaux de stockage. Pour de plus amples informations, veuillez consulter [Partage de données entre comptes dans Lake Formation](#).

Lorsque l'administrateur du lac de données du compte destinataire accepte le AWS RAM partage, les ressources partagées sont disponibles dans le compte du destinataire. L'administrateur du lac de données accorde d'autres autorisations Lake Formation sur la ressource partagée aux principaux IAM supplémentaires du compte destinataire, s'il dispose d'GRANTABLE autorisations sur la ressource partagée.

Cependant, les principaux ne peuvent pas interroger les ressources partagées à l'aide d'Athena ou de Redshift Spectrum sans lien de ressource. Un lien de ressource est une entité du catalogue de données similaire à un concept Linux-Symlink.

L'administrateur du lac de données du compte destinataire crée un lien de ressource sur la ressource partagée. L'administrateur accorde des `Describe` autorisations sur le lien de ressource avec les autorisations requises sur la ressource partagée d'origine à d'autres utilisateurs. Un utilisateur du compte destinataire peut ensuite utiliser le lien de ressource pour interroger la ressource partagée à l'aide d'Athena et Redshift Spectrum. Pour plus d'informations sur les liens vers des ressources, consultez [Création de liens vers des ressources](#).

Composantes de la Lake Formation

AWS Lake Formation repose sur l'interaction de plusieurs composants pour créer et gérer votre lac de données.

Console Lake Formation

Vous utilisez la console Lake Formation pour définir et gérer votre lac de données et pour accorder et révoquer les autorisations de Lake Formation. Vous pouvez utiliser des plans sur la console pour découvrir, nettoyer, transformer et ingérer des données. Vous pouvez également activer ou désactiver l'accès à la console pour les utilisateurs individuels de Lake Formation.

API Lake Formation et interface de ligne de commande

Lake Formation fournit des opérations d'API via plusieurs langages spécifiques SDKs et le AWS Command Line Interface (AWS CLI). L'API Lake Formation fonctionne conjointement avec le AWS Glue API. L'API Lake Formation se concentre principalement sur la gestion des autorisations de Lake Formation, tandis que le AWS Glue L'API fournit une API de catalogue de données et une infrastructure gérée pour définir, planifier et exécuter des opérations ETL sur vos données.

Pour plus d'informations sur le AWS Glue API, consultez le [guide du AWS Glue développeur](#). Pour plus d'informations sur l'utilisation du AWS CLI, consultez la [référence des AWS CLI commandes](#).

Autres AWS services

Lake Formation utilise les services suivants :

- [AWS Glue](#) pour orchestrer les tâches et les robots d'exploration afin de transformer les données à l'aide du AWS Glue transforme.

- [L'IAM](#) va accorder des politiques d'autorisation aux responsables de Lake Formation. Le modèle d'autorisation Lake Formation complète le modèle d'autorisation IAM pour sécuriser votre lac de données.

Terminologie des Lake Formation

Voici quelques termes importants que vous rencontrerez dans ce guide.

Lac de données

Le lac de données correspond à vos données persistantes stockées dans Amazon S3 et gérées par Lake Formation à l'aide d'un catalogue de données. Un lac de données stocke généralement les éléments suivants :

- Données structurées et non structurées
- Données brutes et données transformées

Pour qu'un chemin Amazon S3 se trouve dans un lac de données, il doit être enregistré auprès de Lake Formation.

Accès aux données

Lake Formation fournit un accès sécurisé et granulaire aux données grâce à un nouveau modèle d'accord/de révocation des autorisations qui renforce les politiques (IAM). Gestion des identités et des accès AWS

Les analystes et les data scientists peuvent utiliser le portefeuille complet de services d' AWS analyse et d'apprentissage automatique, tels qu'Amazon Athena, pour accéder aux données. Les politiques de sécurité configurées de Lake Formation permettent de garantir que les utilisateurs ne peuvent accéder qu'aux données auxquelles ils sont autorisés à accéder.

Mode d'accès hybride

Le mode d'accès hybride vous permet de sécuriser et d'accéder aux données cataloguées à l'aide des autorisations Lake Formation et des autorisations IAM et Amazon S3. Le mode d'accès hybride permet aux administrateurs de données d'intégrer les autorisations de Lake Formation de manière sélective et progressive, en se concentrant sur un cas d'utilisation de lac de données à la fois.

Plan

Un plan est un modèle de gestion des données qui vous permet d'ingérer facilement des données dans un lac de données. Lake Formation fournit plusieurs plans, chacun correspondant à un type de source prédéfini, tel qu'une base de données relationnelle ou AWS CloudTrail des journaux. À partir d'un plan, vous pouvez créer un flux de travail. Les flux de travail se composent de AWS Glue robots d'exploration, de tâches et de déclencheurs générés pour orchestrer le chargement et la mise à jour des données. Les plans utilisent la source de données, la cible de données et le calendrier comme entrées pour configurer le flux de travail.

Flux de travail

Un flux de travail est un conteneur pour un ensemble de AWS Glue jobs, crawlers et déclencheurs. Vous créez le flux de travail dans Lake Formation, qui s'exécute dans AWS Glue service. Lake Formation peut suivre l'état d'un flux de travail en tant qu'entité unique.

Lorsque vous définissez un flux de travail, vous sélectionnez le plan sur lequel il est basé. Vous pouvez ensuite exécuter des flux de travail à la demande ou selon un calendrier.

Les flux de travail que vous créez dans Lake Formation sont visibles dans le AWS Glue console sous forme de graphe acyclique dirigé (DAG). À l'aide du DAG, vous pouvez suivre la progression du flux de travail et résoudre les problèmes.

Catalogue de données

Le catalogue de données est votre magasin de métadonnées permanent. Il s'agit d'un service géré qui vous permet de stocker, d'annoter et de partager des métadonnées dans le AWS cloud de la même manière que vous le feriez dans un métastore Apache Hive. Il fournit un référentiel uniforme dans lequel des systèmes disparates peuvent stocker et trouver des métadonnées pour suivre les données dans des silos de données, puis utiliser ces métadonnées pour interroger et transformer les données. Lake Formation utilise le AWS Glue Catalogue de données pour stocker les métadonnées relatives aux lacs de données, aux sources de données, aux transformations et aux cibles.

Les métadonnées relatives aux sources de données et aux cibles se présentent sous forme de bases de données et de tables. Les tables stockent des informations de schéma, des informations de localisation, etc. Les bases de données sont des ensembles de tables. Lake Formation fournit une hiérarchie d'autorisations pour contrôler l'accès aux bases de données et aux tables du catalogue de données.

Chaque AWS compte possède un catalogue de données par AWS région.

Données sous-jacentes

Les données sous-jacentes font référence aux données sources ou aux données des lacs de données vers lesquels pointent les tables du catalogue de données.

Principal

Un principal est un utilisateur ou un rôle Gestion des identités et des accès AWS (IAM) ou un utilisateur Active Directory.

Administrateur du lac de données

Un administrateur de lac de données est un mandant qui peut accorder à n'importe quel principal (y compris lui-même) n'importe quelle autorisation sur n'importe quelle ressource du catalogue de données ou sur l'emplacement des données. Désignez un administrateur de lac de données comme premier utilisateur du catalogue de données. Cet utilisateur peut ensuite accorder des autorisations plus détaillées sur les ressources à d'autres principaux.

Note

Les utilisateurs administratifs IAM, c'est-à-dire les utilisateurs dotés de la politique `AdministratorAccess` AWS gérée, ne sont pas automatiquement des administrateurs de lacs de données. Par exemple, ils ne peuvent pas accorder d'autorisations Lake Formation sur les objets du catalogue à moins d'en avoir obtenu l'autorisation. Ils peuvent toutefois utiliser la console ou l'API Lake Formation pour se désigner comme administrateurs de lacs de données.

Pour plus d'informations sur les fonctionnalités d'un administrateur de lac de données, consultez [Permissions implicites de Lake Formation](#). Pour plus d'informations sur la désignation d'un utilisateur en tant qu'administrateur de data lake, consultez. [Création d'un administrateur de lac de données](#)

AWS intégrations de services avec Lake Formation

Vous pouvez utiliser Lake Formation pour gérer les autorisations d'accès au niveau des bases de données, des tables et des colonnes sur les données stockées dans Amazon S3. Une fois vos données enregistrées auprès de Lake Formation, vous pouvez utiliser des services AWS d'analyse

tels AWS Glue qu'Amazon Athena, Amazon Redshift Spectrum, Amazon EMR pour interroger les données. Les AWS services suivants s'intègrent aux autorisations de Lake Formation AWS Lake Formation et les respectent.

AWS Service	Détails de l'intégration
AWS Glue	Thème de référence : Utilisation AWS Lake Formation avec AWS Glue AWS Glue et Lake Formation partagent le même catalogue de données. Pour les opérations de console (telles que l'affichage d'une liste de tables) et toutes les opérations d'API, les utilisateurs ne peuvent accéder qu'aux bases de données et aux tables sur lesquelles ils disposent des autorisations Lake Formation.
Amazon Athena	Thème de référence : Utilisation AWS Lake Formation avec Amazon Athena Utilisez Lake Formation pour autoriser ou refuser les autorisations de lecture des données dans Amazon S3. Lorsque Amazon Athena les utilisateurs sélectionnent le AWS Glue catalogue dans l'éditeur de requêtes, ils ne peuvent interroger que les bases de données, les tables et les colonnes sur lesquelles ils sont autorisés à Lake Formation. Les requêtes utilisant des manifestes ne sont pas prises en charge. Actuellement, Lake Formation ne prend pas en charge la gestion des autorisations sur les opérations d'écriture telles que VACUUMMERGE, UPDATE et OPTIMIZE sur les tables dans Open Table Formats. Outre les principaux utilisateurs qui s'authentifient auprès d'Athena via Gestion des identités et des accès AWS (IAM), Lake Formation prend en charge les utilisateurs d'Athena qui se connectent via le pilote JDBC ou ODBC et s'authentifient via SAML. Les fournisseurs SAML pris en charge incluent Okta et Microsoft Active Directory Federation Service (AD FS).
Amazon Redshift Spectrum	Thème de référence : Utilisation AWS Lake Formation avec Amazon Redshift Spectrum

AWS Service	Détails de l'intégration
	Lorsque les utilisateurs d'Amazon Redshift créent un schéma externe sur une base de données dans le AWS Glue Data Catalog, ils peuvent uniquement interroger les tables et les colonnes de ce schéma pour lesquelles ils disposent des autorisations de Lake Formation.
Édition Amazon Quick Enterprise	Référence : Utilisation AWS Lake Formation avec Quick Lorsqu'un utilisateur d'Amazon Quick Enterprise Edition interroge un ensemble de données dans un emplacement Amazon S3, il doit disposer de l' <code>SELECT</code> autorisation Lake Formation sur les données.
Amazon EMR	Référence : Utilisation AWS Lake Formation avec Amazon EMR Vous pouvez intégrer les autorisations de Lake Formation lorsque vous créez un cluster Amazon EMR doté d'un rôle d'exécution. Un rôle d'exécution est un rôle IAM que vous associez à des tâches ou à des requêtes Amazon EMR, puis Amazon EMR utilise ce rôle pour accéder aux ressources. AWS

Lake Formation travaille également avec [AWS Key Management Service](#) (AWS KMS) pour vous permettre de configurer plus facilement ces services intégrés afin de chiffrer et de déchiffrer les données sur les sites Amazon Simple Storage Service (Amazon S3).

Ressources supplémentaires sur la Formation du Lac

Pour plus d'informations AWS Lake Formation, nous vous recommandons de continuer à en apprendre davantage sur les concepts présentés dans ce guide en utilisant les ressources suivantes :

Rubriques

- [Blogs](#)
- [Discussions techniques et webinaires](#)
- [Architecture moderne](#)

- [Ressources de maillage de données](#)
- [Guides des meilleures pratiques](#)

Blogs

- [AWS Lake Formation Bilan de l'année 2022](#)
- [Architecture de données moderne multirégionale hautement résiliente](#)
- [Partage entre comptes à l'aide de balises LF pour diriger les principaux IAM](#)
- [Tableau de bord de l'inventaire des autorisations de Lake Formation](#)
- [Maillage de données piloté par événements](#)

Discussions techniques et webinaires

- re:Invent 2020 — [Lacs de données : créez, sécurisez et partagez facilement](#) avec AWS Lake Formation
- re:Invent 2022 — [Création et exploitation d'un lac de données sur Amazon S3](#)
- AWS Summit SF 2022 — [Comprendre et mettre en place une architecture de données moderne](#)
- AWS Summit ATL 2022 — [Des lacs de données modernes avec AWS Lake Formation Amazon Redshift](#) et AWS Glue
- AWS Summit ANZ 2022 — [Lacs de données, maisons lacustres et maillage de données : quoi, pourquoi et comment ?](#)
- AWS Discussions techniques en ligne — [Simplifier les autorisations et la gouvernance dans votre lac de données](#)

Architecture moderne

- [Modèles d'architecture moderne](#)

Ressources de maillage de données

- [Créez une architecture de données moderne et un modèle de maillage de données à grande échelle à l'aide d'un contrôle d'accès AWS Lake Formation basé sur des balises](#)

- [Comment JPMorgan Chase a créé une architecture de maillage de données pour générer une valeur significative afin d'améliorer sa plateforme de données d'entreprise](#)
- [Créez un maillage de données sur AWS](#)

Guides des meilleures pratiques

- [AWS Lake Formation guides de bonnes pratiques](#)

Débuter avec Lake Formation

La lecture de ces sections est indispensable:

- [AWS Lake Formation : comment ça marche](#)— Découvrez la terminologie essentielle et la façon dont les différents composants interagissent.
- [Débuter avec Lake Formation](#)— Obtenez des informations sur les prérequis et effectuez les tâches de configuration importantes.
- [AWS Lake Formation tutoriels](#)— Suivez les step-by-step tutoriels pour apprendre à utiliser Lake Formation.
- [Sécurité dans le AWS Lake Formation](#)— Découvrez comment vous pouvez contribuer à sécuriser l'accès aux données de Lake Formation.

Débuter avec Lake Formation

Si vous ne vous êtes pas inscrit AWS ou si vous avez besoin d'aide pour démarrer, assurez-vous d'effectuer les tâches suivantes.

Rubriques

- [Exécution des tâches AWS de configuration initiale](#)
- [Configurez AWS Lake Formation](#)
- [Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle](#)
- [AWS Lake Formation et points de terminaison VPC d'interface \(\)AWS PrivateLink](#)

Exécution des tâches AWS de configuration initiale

Pour utiliser AWS Lake Formation, vous devez au préalable effectuer les tâches suivantes :

Rubriques

- [Inscrivez-vous pour un Compte AWS](#)
- [Création d'un utilisateur doté d'un accès administratif](#)
- [Octroi d'un accès par programmation](#)

Inscrivez-vous pour un Compte AWS

Si vous n'en avez pas Compte AWS, procédez comme suit pour en créer un.

Pour vous inscrire à un Compte AWS

1. Ouvrez l'<https://portal.aws.amazon.com/billing/inscription>.
2. Suivez les instructions en ligne.

Dans le cadre de la procédure d'inscription, vous recevrez un appel téléphonique ou un SMS et vous saisissez un code de vérification en utilisant le clavier numérique du téléphone.

Lorsque vous vous inscrivez à un Compte AWS, un Utilisateur racine d'un compte AWS est créé. Par défaut, seul l'utilisateur racine a accès à l'ensemble des Services AWS et des ressources de ce compte. La meilleure pratique de sécurité consiste à attribuer un accès administratif à un

utilisateur, et à utiliser uniquement l'utilisateur racine pour effectuer les [tâches nécessitant un accès utilisateur racine](#).

AWS vous envoie un e-mail de confirmation une fois le processus d'inscription terminé. À tout moment, vous pouvez consulter l'activité actuelle de votre compte et le gérer en accédant à <https://aws.amazon.com/> et en choisissant Mon compte.

Création d'un utilisateur doté d'un accès administratif

Après vous être inscrit à un Compte AWS, sécurisez Utilisateur racine d'un compte AWS AWS IAM Identity Center, activez et créez un utilisateur administratif afin de ne pas utiliser l'utilisateur root pour les tâches quotidiennes.

Sécurisez votre Utilisateur racine d'un compte AWS

1. Connectez-vous en [AWS Management Console](#) tant que propriétaire du compte en choisissant Utilisateur root et en saisissant votre adresse Compte AWS e-mail. Sur la page suivante, saisissez votre mot de passe.

Pour obtenir de l'aide pour vous connecter en utilisant l'utilisateur racine, consultez [Connexion en tant qu'utilisateur racine](#) dans le Guide de l'utilisateur Connexion à AWS .

2. Activez l'authentification multifactorielle (MFA) pour votre utilisateur racine.

Pour obtenir des instructions, voir [Activer un périphérique MFA virtuel pour votre utilisateur Compte AWS root \(console\)](#) dans le guide de l'utilisateur IAM.

Création d'un utilisateur doté d'un accès administratif

1. Activez IAM Identity Center.

Pour obtenir des instructions, consultez [Activation d' AWS IAM Identity Center](#) dans le Guide de l'utilisateur AWS IAM Identity Center .

2. Dans IAM Identity Center, octroyez un accès administratif à un utilisateur.

Pour un didacticiel sur l'utilisation du Répertoire IAM Identity Center comme source d'identité, voir [Configurer l'accès utilisateur par défaut Répertoire IAM Identity Center](#) dans le Guide de AWS IAM Identity Center l'utilisateur.

Connexion en tant qu'utilisateur doté d'un accès administratif

- Pour vous connecter avec votre utilisateur IAM Identity Center, utilisez l'URL de connexion qui a été envoyée à votre adresse e-mail lorsque vous avez créé l'utilisateur IAM Identity Center.

Pour obtenir de l'aide pour vous connecter en utilisant un utilisateur d'IAM Identity Center, consultez la section [Connexion au portail AWS d'accès](#) dans le guide de l'Connexion à AWS utilisateur.

Attribution d'un accès à d'autres utilisateurs

1. Dans IAM Identity Center, créez un ensemble d'autorisations qui respecte la bonne pratique consistant à appliquer les autorisations de moindre privilège.

Pour obtenir des instructions, consultez [Création d'un ensemble d'autorisations](#) dans le Guide de l'utilisateur AWS IAM Identity Center .

2. Attribuez des utilisateurs à un groupe, puis attribuez un accès par authentification unique au groupe.

Pour obtenir des instructions, consultez [Ajout de groupes](#) dans le Guide de l'utilisateur AWS IAM Identity Center .

Octroi d'un accès par programmation

Les utilisateurs ont besoin d'un accès programmatique s'ils souhaitent interagir avec AWS l'extérieur du AWS Management Console. La manière d'accorder un accès programmatique dépend du type d'utilisateur qui y accède AWS.

Pour accorder aux utilisateurs un accès programmatique, choisissez l'une des options suivantes.

Quel utilisateur a besoin d'un accès programmatique ?	À	Méthode
IAM	(Recommandé) Utilisez les informations d'identification de la console comme informations d'identification temporaires pour signer les demandes	Suivez les instructions de l'interface que vous souhaitez utiliser.

Quel utilisateur a besoin d'un accès programmatique ?	À	Méthode
	programmatiques adressées au AWS CLI AWS SDKs, ou AWS APIs.	• Pour le AWS CLI, voir Connexion pour le développement AWS local dans le guide de AWS Command Line Interface l'utilisateur. • Pour AWS SDKs, voir Connexion pour le développement AWS local dans le guide de référence AWS SDKs and Tools.
Identité de la main-d'œuvre (Utilisateurs gérés dans IAM Identity Center)	Utilisez des informations d'identification temporaires pour signer les demandes programmatiques adressées au AWS CLI AWS SDKs, ou AWS APIs.	Suivez les instructions de l'interface que vous souhaitez utiliser. • Pour le AWS CLI, voir Configuration du AWS CLI à utiliser AWS IAM Identity Center dans le guide de AWS Command Line Interface l'utilisateur. • Pour AWS SDKs, outils, et AWS APIs, voir Authentification IAM Identity Center dans le guide de référence AWS SDKs et Tools.
IAM	Utilisez des informations d'identification temporaires pour signer les demandes programmatiques adressées au AWS CLI AWS SDKs, ou AWS APIs.	Suivez les instructions de la section Utilisation d'informations d'identification temporaires avec AWS les ressources du Guide de l'utilisateur IAM.

Quel utilisateur a besoin d'un accès programmatique ?	À	Méthode
IAM	(Non recommandé) Utilisez des informations d'identification à long terme pour signer des demandes programmatiques adressées au AWS CLI AWS SDKs, ou AWS APIs.	Suivez les instructions de l'interface que vous souhaitez utiliser. • Pour le AWS CLI, voir Authentification à l'aide des informations d'identification utilisateur IAM dans le Guide de l'AWS Command Line Interface utilisateur. • Pour les outils AWS SDKs et, voir Authentifier à l'aide d'informations d'identification à long terme dans le guide de référence des outils AWS SDKs et. • Pour AWS APIs, voir Gestion des clés d'accès pour les utilisateurs IAM dans le Guide de l'utilisateur IAM.

Configurez AWS Lake Formation

Les sections suivantes fournissent des informations sur la configuration de Lake Formation pour la première fois. Toutes les rubriques de cette section ne sont pas obligatoires pour commencer à utiliser Lake Formation. Vous pouvez utiliser les instructions pour configurer le modèle d'autorisations de Lake Formation afin de gérer vos AWS Glue Data Catalog objets et emplacements de données existants dans Amazon Simple Storage Service (Amazon S3).

1. [Création d'un administrateur de lac de données](#)
2. [Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride](#)
3. [the section called “Configurer un emplacement Amazon S3 pour votre lac de données”](#)

4. [the section called “Attribuer des autorisations aux utilisateurs de Lake Formation”](#)
5. [the section called “Intégration d'IAM Identity Center”](#)
6. [the section called “\(Facultatif\) Paramètres de filtrage des données externes”](#)
7. [the section called “\(Facultatif\) Accordez l'accès à la clé de chiffrement du catalogue de données”](#)
8. [\(Facultatif\) Créez un rôle IAM pour les flux de travail](#)

Cette section explique comment configurer les ressources de Lake Formation de deux manières différentes :

- Utilisation d'un AWS CloudFormation modèle
- Utilisation de la console Lake Formation

Pour configurer Lake Formation à l'aide de AWS la console, rendez-vous sur [Création d'un administrateur de lac de données](#).

Configurer les ressources de Lake Formation à l'aide d' CloudFormation un modèle

Note

La CloudFormation pile exécute les étapes 1 à 6 ci-dessus, à l'exception des étapes 2 et 5. Effectuez [Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride](#) et [the section called “Intégration d'IAM Identity Center”](#) manuellement depuis la console Lake Formation.

1. Connectez-vous à la AWS CloudFormation console à l'adresse <https://console.aws.amazon.com/cloudformation> en tant qu'administrateur IAM dans la région USA Est (Virginie du Nord).
2. Choisissez [Launch Stack](#).
3. Choisissez Next sur l'écran de création d'une pile.
4. Entrez un nom de pile.
5. Pour DatalakeAdminName et DatalakeAdminPassword, entrez votre nom d'utilisateur et votre mot de passe pour l'utilisateur administrateur de Data Lake.

6. Pour `DatalakeUser1Name` et `DatalakeUser1Password`, entrez votre nom d'utilisateur et votre mot de passe pour l'utilisateur Data Lake Analyst.
7. Pour `DataLakeBucketName`, entrez le nouveau nom du bucket qui sera créé.
8. Choisissez Suivant.
9. Sur la page suivante, choisissez `I acknowledge that CloudFormation might create IAM resources with custom names` Next.
10. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
11. Choisissez Créer.

La création de la pile peut prendre jusqu'à deux minutes.

nettoyer des ressources ;

Si vous souhaitez nettoyer les ressources de la CloudFormation pile :

1. Désenregistrez le compartiment Amazon S3 créé par votre stack et enregistré en tant qu'emplacement de lac de données.
2. Supprimez la CloudFormation pile. Cela supprimera toutes les ressources créées par la pile.

Création d'un administrateur de lac de données

Les administrateurs de lacs de données sont initialement les seuls utilisateurs ou rôles Gestion des identités et des accès AWS (IAM) qui peuvent accorder à Lake Formation des autorisations sur les emplacements des données et les ressources du catalogue de données à n'importe quel principal (y compris à lui-même). Pour plus d'informations sur les fonctionnalités d'administration des lacs de données, consultez [Permissions implicites de Lake Formation](#). Par défaut, Lake Formation vous permet de créer jusqu'à 30 administrateurs de lacs de données.

Vous pouvez créer un administrateur de lac de données à l'aide de la `PutDataLakeSettings` console Lake Formation ou de l'API Lake Formation.

Les autorisations suivantes sont requises pour créer un administrateur de lac de données. L'Administrateur utilisateur dispose implicitement de ces autorisations.

- `lakeformation:PutDataLakeSettings`
- `lakeformation:GetDataLakeSettings`

Si vous accordez cette `AWSLakeFormationDataAdmin` politique à un utilisateur, celui-ci ne pourra pas créer d'autres utilisateurs administrateurs de Lake Formation.

Pour créer un administrateur de lac de données (console)

1. Si l'utilisateur qui doit être administrateur du lac de données n'existe pas encore, utilisez la console IAM pour le créer. Sinon, choisissez un utilisateur existant qui sera l'administrateur du lac de données.

 Note

Nous vous recommandons de ne pas sélectionner d'utilisateur administratif IAM (utilisateur doté de la politique `AdministratorAccess` AWS gérée) comme administrateur du lac de données.

Associez les politiques AWS gérées suivantes à l'utilisateur :

Stratégies	Obligatoire ?	Remarques
<code>AWSLakeFormationDataAdmin</code>	Obligatoire	Autorisations d'administrateur de data lake de base. Cette politique AWS gérée contient un refus explicite du fonctionnement de l'API Lake Formation, <code>PutDataLakeSetting</code> qui empêche les utilisateurs de créer de nouveaux administrateurs de lacs de données.

Stratégies	Obligatoire ?	Remarques
AWSGlueConsoleFullAccess , CloudWatchLogsReadOnlyAccess	Facultatif	Joignez ces politiques si l'administrateur du lac de données doit résoudre les problèmes liés aux flux de travail créés à partir des plans de Lake Formation. Ces politiques permettent à l'administrateur du lac de données de consulter les informations de dépannage dans la AWS Glue console et dans la Amazon CloudWatch Logs console. Pour plus d'informations sur les flux de travail, consultez the section called "Importation de données à l'aide de workflows" .
AWSLakeFormationCrossAccountManager	Facultatif	Joignez cette politique pour permettre à l'administrateur du lac de données d'accorder et de révoquer des autorisations entre comptes sur les ressources du catalogue de données. Pour de plus amples informations, veuillez consulter Partage de données entre comptes dans Lake Formation .
AmazonAthenaFullAccess	Facultatif	Joignez cette politique si l'administrateur du lac de données doit exécuter des requêtes dans Amazon Athena.

- Joignez la politique intégrée suivante, qui accorde à l'administrateur du lac de données l'autorisation de créer le rôle lié au service Lake Formation. Le nom suggéré pour la politique est `LakeFormationSLR`.

Le rôle lié à un service permet à l'administrateur du lac de données d'enregistrer plus facilement les sites Amazon S3 auprès de Lake Formation. Pour plus d'informations sur le rôle lié au service Lake Formation, consultez. [the section called "Utilisation des rôles liés à un service"](#)

 Important

Dans toutes les politiques suivantes, remplacez-le *<account-id>* par un numéro de AWS compte valide.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:CreateServiceLinkedRole",
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "iam:AWSServiceName": "lakeformation.amazonaws.com"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "iam:PutRolePolicy"
      ],
      "Resource": "arn:aws:iam::<account-id>:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess"
    }
  ]
}
```

3. (Facultatif) Attachez la politique PassRole intégrée suivante à l'utilisateur. Cette politique permet à l'administrateur du lac de données de créer et d'exécuter des flux de travail. L'iam:PassRole autorisation permet au flux de travail d'assumer le rôle LakeFormationWorkflowRole de créer des robots d'exploration et des tâches, et d'associer le rôle aux robots et aux tâches créés. Le nom suggéré pour la politique est UserPassRole.

 Important

Remplacez *<account-id>* par un numéro de AWS compte valide.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "PassRolePermissions",
      "Effect": "Allow",
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
        "arn:aws:iam::111122223333:role/LakeFormationWorkflowRole"
      ]
    }
  ]
}
```

4. (Facultatif) Joignez cette politique en ligne supplémentaire si votre compte accordera ou recevra des autorisations entre comptes Lake Formation. Cette politique permet à l'administrateur du lac de données de consulter et d'accepter AWS Resource Access Manager (AWS RAM) les invitations de partage de ressources. En outre, pour les administrateurs de lacs de données du compte de AWS Organizations gestion, la politique inclut une autorisation permettant d'autoriser les subventions entre comptes aux organisations. Pour de plus amples informations, veuillez consulter [Partage de données entre comptes dans Lake Formation](#).

Le nom suggéré pour la politique est RAMAccess.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ram:AcceptResourceShareInvitation",
        "ram:RejectResourceShareInvitation",
        "ec2:DescribeAvailabilityZones",

```

```

        "ram:EnableSharingWithAwsOrganization"
      ],
      "Resource": "*"
    }
  ]
}

```

5. Ouvrez la AWS Lake Formation console <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) ou en tant qu'utilisateur doté d'une politique AWS gérée par AdministratorAccess l'utilisateur.
6. Si une fenêtre Welcome to Lake Formation apparaît, choisissez l'utilisateur IAM que vous avez créé ou sélectionné à l'étape 1, puis choisissez Get started.
7. Si la fenêtre Welcome to Lake Formation ne s'affiche pas, effectuez les étapes suivantes pour configurer un administrateur de Lake Formation.
 - a. Dans le volet de navigation, sous Administration, sélectionnez Administrative roles and tasks. Dans la section Administrateurs du lac de données de la page de console, choisissez Ajouter.
 - b. Dans la boîte de dialogue Ajouter des administrateurs, sous Type d'accès, choisissez Data lake administrator.
 - c. Pour les utilisateurs et les rôles IAM, choisissez l'utilisateur IAM que vous avez créé ou sélectionné à l'étape 1, puis sélectionnez Enregistrer.

Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride

Lake Formation démarre avec les paramètres « Utiliser uniquement le contrôle d'accès IAM » activés pour garantir la compatibilité avec le AWS Glue Data Catalog comportement existant. Ces paramètres vous permettent de gérer l'accès à vos données dans le lac de données et à ses métadonnées par le biais de politiques IAM et de politiques de compartiment Amazon S3.

Pour faciliter la transition des autorisations de lac de données d'un modèle IAM et Amazon S3 vers les autorisations de Lake Formation, nous vous recommandons d'utiliser le mode d'accès hybride pour Data Catalog. Avec le mode d'accès hybride, vous disposez d'un chemin incrémentiel qui vous permet d'activer les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les autres utilisateurs ou charges de travail existants.

Pour de plus amples informations, veuillez consulter [Mode d'accès hybride](#).

Désactivez les paramètres par défaut pour déplacer tous les utilisateurs existants d'une table vers Lake Formation en une seule étape.

 Important

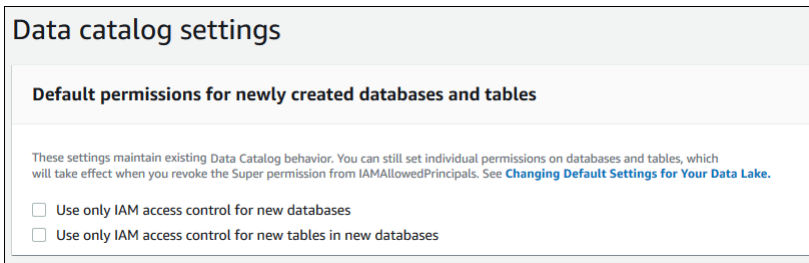
Si vous avez des AWS Glue Data Catalog bases de données et des tables existantes, ne suivez pas les instructions de cette section. Au lieu de cela, suivez les instructions de [the section called “Mise à niveau des autorisations de AWS Glue données pour le modèle Lake Formation”](#).

 Warning

Si vous avez mis en place une automatisation qui crée des bases de données et des tables dans le catalogue de données, les étapes suivantes peuvent entraîner l'échec des tâches d'automatisation et d'extraction, de transformation et de chargement (ETL) en aval. Procédez uniquement après avoir modifié vos processus existants ou accordé des autorisations explicites de Lake Formation aux principaux responsables requis. Pour plus d'informations sur les autorisations de Lake Formation, consultez [the section called “Référence des autorisations de Lake Formation”](#).

Pour modifier les paramètres par défaut du catalogue de données

1. Continuez dans la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>. Assurez-vous d'être connecté en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) ou en tant qu'utilisateur avec la politique AdministratorAccess AWS gérée.
2. Modifiez les paramètres du catalogue de données :
 - a. Dans le volet de navigation, sous Administration, sélectionnez Paramètres du catalogue de données.
 - b. Décochez les deux cases et choisissez Enregistrer.



3. Révoquez IAMAllowedPrincipals l'autorisation accordée aux créateurs de bases de données.
 - a. Dans le volet de navigation, sous Administration, sélectionnez Administrative roles and tasks.
 - b. Sur la page de console des rôles et tâches administratifs, dans la section Créateurs de base de données, sélectionnez le IAMAllowedPrincipals groupe, puis choisissez Révoquer.

La boîte de dialogue Révoquer les autorisations apparaît, indiquant que IAMAllowedPrincipals vous disposez de l'autorisation Créer une base de données.

- c. Choisissez Révoquer.

Attribuer des autorisations aux utilisateurs de Lake Formation

Créez un utilisateur qui aura accès au lac de données dans AWS Lake Formation. Cet utilisateur dispose des autorisations de moindre privilège pour interroger le lac de données.

Pour plus d'informations sur la création d'utilisateurs ou de groupes, consultez la section [Identités IAM](#) dans le guide de l'utilisateur IAM.

Pour attribuer des autorisations à un utilisateur non administrateur afin qu'il puisse accéder aux données de Lake Formation

1. Ouvrez la console IAM à <https://console.aws.amazon.com/iam> et connectez-vous en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) ou en tant qu'utilisateur avec la politique AdministratorAccess AWS gérée.
2. Choisissez Utilisateurs ou Groupes d'utilisateurs.
3. Dans la liste, sélectionnez le nom de l'utilisateur ou du groupe auquel intégrer une politique.

Choisissez Autorisations.

4. Choisissez Ajouter des autorisations, puis choisissez Joindre directement les politiques. Entrez Athena dans le champ de texte Politiques de filtrage. Dans la liste des résultats, cochez la case correspondant à AmazonAthenaFullAccess.
5. Cliquez sur le bouton Créer une politique. Sur la page Créer une politique, choisissez l'onglet JSON. Copiez et collez le code suivant dans l'éditeur de politiques.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "glue:GetTable",
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetPartitions",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:GetLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
      ],
      "Resource": "*"
    }
  ]
}
```

6. Cliquez sur le bouton Suivant en bas jusqu'à ce que la page de révision de la politique s'affiche. Entrez un nom pour la politique, par exemple, DataLakeUserBasic. Choisissez Créer une politique, puis fermez l'onglet Politiques ou la fenêtre du navigateur.

Configurer un emplacement Amazon S3 pour votre lac de données

Pour utiliser Lake Formation afin de gérer et de sécuriser les données de votre lac de données, vous devez d'abord enregistrer un emplacement Amazon S3. Lorsque vous enregistrez un emplacement,

ce chemin Amazon S3 et tous les dossiers situés sous ce chemin sont enregistrés, ce qui permet à Lake Formation de faire appliquer les autorisations de niveau de stockage. Lorsque l'utilisateur demande des données à un moteur intégré tel qu'Amazon Athena, Lake Formation fournit l'accès aux données au lieu d'utiliser les autorisations de l'utilisateur.

Lorsque vous enregistrez un emplacement, vous spécifiez un rôle IAM qui accorde des read/write autorisations sur cet emplacement. Lake Formation assume ce rôle lorsqu'il fournit des informations d'identification temporaires aux AWS services intégrés qui demandent l'accès aux données de l'emplacement Amazon S3 enregistré. Vous pouvez spécifier le rôle lié au service (SLR) de Lake Formation ou créer votre propre rôle.

Utilisez un rôle personnalisé dans les situations suivantes :

- Vous prévoyez de publier des statistiques dans Amazon CloudWatch Logs. Le rôle défini par l'utilisateur doit inclure une politique d'ajout de journaux dans CloudWatch les journaux et de publication de métriques en plus des autorisations SLR. Pour un exemple de politique en ligne qui accorde les CloudWatch autorisations nécessaires, voir [Exigences relatives aux rôles utilisés pour enregistrer des sites](#).
- L'emplacement Amazon S3 existe dans un autre compte. Pour en savoir plus, consultez [the section called “Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte”](#).
- L'emplacement Amazon S3 contient des données chiffrées avec un Clé gérée par AWS. Pour plus d'informations, consultez [Enregistrement d'un emplacement Amazon S3 chiffré](#) et [Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes](#).
- Vous prévoyez d'accéder à l'emplacement Amazon S3 à l'aide d'Amazon EMR. Pour plus d'informations sur les exigences relatives aux rôles, consultez la section [Rôles IAM pour Lake Formation](#) dans le guide de gestion Amazon EMR.

Le rôle que vous choisissez doit disposer des autorisations nécessaires, comme décrit dans [Exigences relatives aux rôles utilisés pour enregistrer des sites](#). Pour obtenir des instructions sur la façon d'enregistrer un emplacement Amazon S3, consultez [Ajouter un emplacement Amazon S3 à votre lac de données](#).

(Facultatif) Paramètres de filtrage des données externes

Si vous avez l'intention d'analyser et de traiter les données de votre lac de données à l'aide de moteurs de requêtes tiers, vous devez autoriser les moteurs externes à accéder aux données gérées

par Lake Formation. Si vous ne vous inscrivez pas, les moteurs externes ne pourront pas accéder aux données des sites Amazon S3 enregistrés auprès de Lake Formation.

Lake Formation prend en charge les autorisations au niveau des colonnes afin de restreindre l'accès à des colonnes spécifiques d'un tableau. Les services d'analyse intégrés tels qu' Amazon Athena Amazon Redshift Spectrum et Amazon EMR extraient les métadonnées de table non filtrées à partir du. AWS Glue Data Catalog Le filtrage proprement dit des colonnes dans les réponses aux requêtes relève de la responsabilité du service intégré. Il est de la responsabilité des administrateurs tiers de gérer correctement les autorisations afin d'éviter tout accès non autorisé aux données.

Pour autoriser les moteurs tiers à accéder aux données et à les filtrer (console)

1. Continuez dans la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>. Assurez-vous que vous êtes connecté en tant que principal disposant de l'autorisation IAM pour le fonctionnement de l'PutDataLakeSettingsAPI Lake Formation. L'utilisateur administrateur IAM que vous avez créé [Inscrivez-vous pour un Compte AWS](#) possède cette autorisation.
2. Dans le volet de navigation, sous Administration, sélectionnez Paramètres d'intégration des applications.
3. Sur la page des paramètres d'intégration des applications, procédez comme suit :
 - a. Cochez la case Autoriser les moteurs externes à filtrer les données dans les sites Amazon S3 enregistrés auprès de Lake Formation.
 - b. Entrez les valeurs des balises de session définies pour les moteurs tiers.
 - c. Pour le AWS compte IDs, saisissez le compte IDs à partir duquel les moteurs tiers sont autorisés à accéder aux sites enregistrés auprès de Lake Formation. Appuyez sur Entrée après chaque identifiant de compte.
 - d. Choisissez Enregistrer.

Pour autoriser les moteurs externes à accéder aux données sans validation des balises de session, voir [Intégration des applications pour un accès complet aux tables](#)

(Facultatif) Accordez l'accès à la clé de chiffrement du catalogue de données

Si elle AWS Glue Data Catalog est cryptée, accordez des autorisations Gestion des identités et des accès AWS (IAM) sur la AWS KMS clé à tous les principaux qui doivent accorder des autorisations à Lake Formation sur les bases de données et les tables du catalogue de données.

Pour plus d'informations, consultez le Guide du développeur AWS Key Management Service .

(Facultatif) Créez un rôle IAM pour les flux de travail

Avec AWS Lake Formation, vous pouvez importer vos données à l'aide de flux de travail exécutés par des AWS Glue robots d'exploration. Un flux de travail définit la source de données et le calendrier d'importation des données dans votre lac de données. Vous pouvez facilement définir des flux de travail à l'aide des plans ou des modèles fournis par Lake Formation.

Lorsque vous créez un flux de travail, vous devez lui attribuer un rôle Gestion des identités et des accès AWS (IAM) qui accorde à Lake Formation les autorisations nécessaires pour ingérer les données.

La procédure suivante suppose une bonne connaissance de l'IAM.

Pour créer un rôle IAM pour les flux de travail

1. Ouvrez la console IAM à <https://console.aws.amazon.com/iam> et connectez-vous en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) ou en tant qu'utilisateur avec la politique AdministratorAccess AWS gérée.
2. Dans le volet de navigation, choisissez Rôles, puis Créer un rôle.
3. Sur la page Create role, choisissez AWS service, puis Glue. Choisissez Suivant.
4. Sur la page Ajouter des autorisations, recherchez la politique AWSGlueServiceRole gérée et cochez la case à côté du nom de la politique dans la liste. Complétez ensuite l'assistant de création de rôle en nommant le rôle LFWorkflowRole. Pour terminer, choisissez Create role.
5. De retour sur la page Rôles LFWorkflowRole, recherchez et choisissez le nom du rôle.
6. Sur la page Résumé des rôles, sous l'onglet Autorisations, choisissez Créer une politique intégrée. Sur l'écran Créer une politique, accédez à l'onglet JSON et ajoutez la politique en ligne suivante. Le nom suggéré pour la politique est LakeFormationWorkflow.

⚠ Important

Dans la politique suivante, remplacez-le `<account-id>` par un Compte AWS numéro valide.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "lakeformation:GrantPermissions"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": ["iam:PassRole"],
      "Resource": [
        "arn:aws:iam::111122223333:role/LakeFormationWorkflowRole"
      ]
    }
  ]
}
```

Voici une brève description des autorisations définies dans cette politique :

- `lakeformation:GetDataAccess` permet aux tâches créées par le flux de travail d'écrire sur l'emplacement cible.
- `lakeformation:GrantPermissions` permet au flux de travail d'accorder l'`SELECT` autorisation sur les tables cibles.
- `iam:PassRole` permet au service d'assumer le rôle de `LakeFormationWorkflowRole` créer des robots d'exploration et des tâches (instances de flux de travail), et d'associer le rôle aux robots et aux tâches créés.

7. Vérifiez que deux politiques sont associées au rôle `LakeFormationWorkflowRole`.
8. Si vous ingérez des données situées en dehors de l'emplacement du lac de données, ajoutez une politique intégrée autorisant la lecture des données sources.

Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle

AWS Lake Formation les autorisations permettent un contrôle d'accès précis pour les données de votre lac de données. Vous pouvez utiliser le modèle d'autorisations de Lake Formation pour gérer vos AWS Glue Data Catalog objets et emplacements de données existants dans Amazon Simple Storage Service (Amazon S3).

Le modèle d'autorisations de Lake Formation utilise des autorisations grossières Gestion des identités et des accès AWS (IAM) pour l'accès aux services d'API. Lake Formation utilise des [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#) fonctionnalités pour restreindre l'accès aux tables au niveau des colonnes, des lignes et des cellules pour les utilisateurs et leurs applications. En comparaison, le AWS Glue modèle accorde l'accès aux données via des [politiques IAM basées sur l'identité et basées sur les ressources](#).

Pour effectuer le changement, suivez les étapes décrites dans ce guide.

Pour de plus amples informations, veuillez consulter [Vue d'ensemble des autorisations relatives à Lake Formation](#).

À propos des autorisations par défaut

Pour maintenir la rétrocompatibilité avec AWS Glue, AWS Lake Formation accorde par défaut l'`Superautorisation` au `IAMAllowedPrincipals` groupe sur toutes les ressources du catalogue de AWS Glue données existantes, et accorde l'`Superautorisation` sur les nouvelles ressources du catalogue de données si les paramètres de contrôle d'accès Utiliser uniquement IAM sont activés. Ainsi, l'accès aux ressources du catalogue de données et aux emplacements Amazon S3 est contrôlé uniquement par des politiques Gestion des identités et des accès AWS (IAM). Le `IAMAllowedPrincipals` groupe inclut tous les utilisateurs et rôles IAM autorisés à accéder aux objets de votre catalogue de données par vos politiques IAM. L'`Superautorisation` permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée.

Vous pouvez commencer à utiliser Lake Formation pour gérer l'accès à vos données en enregistrant les emplacements des ressources du catalogue de données existantes dans Lake Formation ou en utilisant le mode d'accès hybride. Lorsque vous enregistrez l'emplacement Amazon S3 en mode d'accès hybride, vous pouvez activer les autorisations de Lake Formation en optant pour les principes pour les bases de données et les tables situées sous cet emplacement.

Pour faciliter la transition des autorisations de lac de données d'un modèle IAM et Amazon S3 vers les autorisations de Lake Formation, nous vous recommandons d'utiliser le mode d'accès hybride pour Data Catalog. Avec le mode d'accès hybride, vous disposez d'un chemin incrémentiel qui vous permet d'activer les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les autres utilisateurs ou charges de travail existants.

Pour de plus amples informations, veuillez consulter [Mode d'accès hybride](#).

Désactivez les paramètres par défaut du catalogue de données pour déplacer tous les utilisateurs existants d'une table vers Lake Formation en une seule étape.

Pour commencer à utiliser les autorisations de Lake Formation avec vos bases de AWS Glue données et tables de catalogue de données existantes, vous devez effectuer les opérations suivantes :

1. Déterminez les autorisations IAM existantes de vos utilisateurs pour chaque base de données et chaque table.
2. Répliquez ces autorisations dans Lake Formation.
3. Pour chaque site Amazon S3 contenant des données :
 - a. Révoquez l'`Super`autorisation accordée au `IAMAllowedPrincipals` groupe sur chaque ressource du catalogue de données qui fait référence à cet emplacement.
 - b. Enregistrez l'emplacement auprès de Lake Formation.
4. Nettoyez les politiques précises de contrôle d'accès IAM existantes.

Important

Pour ajouter de nouveaux utilisateurs pendant le processus de transition de votre catalogue de données, vous devez configurer des AWS Glue autorisations granulaires dans IAM comme auparavant. Vous devez également reproduire ces autorisations dans Lake Formation, comme décrit dans cette section. Si les nouveaux utilisateurs disposent des politiques IAM grossières décrites dans ce guide, ils peuvent répertorier toutes les bases de

données ou tables auxquelles l'autorisation est accordée. Super IAMAllowedPrincipals Ils peuvent également consulter les métadonnées de ces ressources.

Suivez les étapes décrites dans cette section pour passer au modèle d'autorisations Lake Formation.

Rubriques

- [Étape 1 : Répertorier les autorisations existantes des utilisateurs et des rôles](#)
- [Étape 2 : configurer des autorisations équivalentes pour Lake Formation](#)
- [Étape 3 : Donnez aux utilisateurs l'autorisation IAM d'utiliser Lake Formation](#)
- [Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations Lake Formation](#)
- [Étape 5 : Sécuriser les nouvelles ressources du catalogue de données](#)
- [Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM pour l'accès futur au data lake](#)
- [Étape 7 : Nettoyer les politiques IAM existantes](#)

Étape 1 : Répertorier les autorisations existantes des utilisateurs et des rôles

Pour commencer à utiliser AWS Lake Formation les autorisations avec vos AWS Glue bases de données et tables existantes, vous devez d'abord déterminer les autorisations existantes de vos utilisateurs.

Important

Avant de commencer, assurez-vous d'avoir terminé les tâches dans [Prise en main](#).

Rubriques

- [Utilisation de l'opération API](#)
- [À l'aide du AWS Management Console](#)
- [En utilisant AWS CloudTrail](#)

Utilisation de l'opération API

Utilisez l'opération d'[ListPoliciesGrantingServiceAccess](#) API Gestion des identités et des accès AWS (IAM) pour déterminer les politiques IAM associées à chaque principal (utilisateur ou rôle). À partir des politiques renvoyées dans les résultats, vous pouvez déterminer les autorisations IAM accordées au principal. Vous devez appeler l'API pour chaque principal séparément.

Exemple

L' AWS CLI exemple suivant renvoie les politiques associées à l'utilisateur `glue_user1`.

```
aws iam list-policies-granting-service-access --arn arn:aws:iam::111122223333:user/glue_user1 --service-namespaces glue
```

La commande renvoie des résultats similaires aux suivants.

```
{
  "PoliciesGrantingServiceAccess": [
    {
      "ServiceNamespace": "glue",
      "Policies": [
        {
          "PolicyType": "INLINE",
          "PolicyName": "GlueUserBasic",
          "EntityName": "glue_user1",
          "EntityType": "USER"
        },
        {
          "PolicyType": "MANAGED",
          "PolicyArn": "arn:aws:iam::aws:policy/AmazonAthenaFullAccess",
          "PolicyName": "AmazonAthenaFullAccess"
        }
      ]
    }
  ],
  "IsTruncated": false
}
```

À l'aide du AWS Management Console

Vous pouvez également consulter ces informations sur la console Gestion des identités et des accès AWS (IAM), dans l'onglet Access Advisor de la page Résumé de l'utilisateur ou du rôle :

1. Ouvrez la console IAM à l'adresse <https://console.aws.amazon.com/iam/>.
2. Dans le panneau de navigation, sélectionnez Users (Utilisateurs) ou Roles (Rôles).
3. Choisissez un nom dans la liste pour ouvrir sa page de résumé, puis cliquez sur l'onglet Access Advisor.
4. Examinez chacune des politiques afin de déterminer la combinaison de bases de données, de tables et d'actions pour laquelle chaque utilisateur est autorisé.

N'oubliez pas d'inspecter les rôles en plus des utilisateurs au cours de ce processus, car vos tâches de traitement des données peuvent assumer des rôles pour accéder aux données.

En utilisant AWS CloudTrail

Une autre façon de déterminer vos autorisations existantes consiste à rechercher les appels d'AWS GlueAPI AWS CloudTrail pour lesquels le `additionalEventData` champ des journaux contient une `insufficientLakeFormationPermissions` entrée. Cette entrée répertorie la base de données et la table pour lesquelles l'utilisateur doit disposer des autorisations de Lake Formation pour effectuer la même action.

Il s'agit de journaux d'accès aux données, il n'est donc pas garanti qu'ils produisent une liste complète des utilisateurs et de leurs autorisations. Nous vous recommandons de choisir une plage de temps étendue pour capturer la plupart des modèles d'accès aux données de vos utilisateurs, par exemple plusieurs semaines ou mois.

Pour plus d'informations, consultez la section [Affichage des événements avec l'historique des CloudTrail événements](#) dans le guide de AWS CloudTrail l'utilisateur.

Ensuite, vous pouvez configurer les autorisations de Lake Formation pour qu'elles correspondent aux AWS Glue autorisations. Consultez [Étape 2 : configurer des autorisations équivalentes pour Lake Formation](#).

Étape 2 : configurer des autorisations équivalentes pour Lake Formation

À l'aide des informations collectées dans [Étape 1 : Répertorier les autorisations existantes des utilisateurs et des rôles](#), accordez AWS Lake Formation des autorisations correspondant aux AWS Glue autorisations. Utilisez l'une des méthodes suivantes pour effectuer les subventions :

- Utilisez la console Lake Formation ou le AWS CLI.

Consultez [the section called "Octroi d'autorisations de données"](#).

- Utilisez les opérations de BatchGrantPermissions l'API GrantPermissions or.

Consultez [Autorisations APIs](#).

Pour de plus amples informations, veuillez consulter [Vue d'ensemble des autorisations relatives à Lake Formation](#).

Après avoir configuré les autorisations de Lake Formation, passez à [Étape 3 : Donnez aux utilisateurs l'autorisation IAM d'utiliser Lake Formation](#).

Étape 3 : Donnez aux utilisateurs l'autorisation IAM d'utiliser Lake Formation

Pour utiliser le modèle AWS Lake Formation d'autorisations, les directeurs doivent disposer d'autorisations Gestion des identités et des accès AWS (IAM) sur la Lake Formation. APIs

Créez la politique suivante dans IAM et associez-la à chaque utilisateur ayant besoin d'accéder à votre lac de données. Nommez la stratégie LakeFormationDataAccess.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "LakeFormationDataAccess",
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess"
      ],
      "Resource": "*"
    }
  ]
}
```

Passez ensuite aux autorisations de Lake Formation, un emplacement de données à la fois. Consultez [Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations Lake Formation](#).

Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations Lake Formation

Passez aux autorisations de Lake Formation, un emplacement de données à la fois. Pour ce faire, répétez l'intégralité de cette section jusqu'à ce que vous ayez enregistré tous les chemins Amazon Simple Storage Service (Amazon S3) référencés par votre catalogue de données.

Rubriques

- [Vérifiez les autorisations de Lake Formation](#)
- [Sécuriser les ressources du catalogue de données existantes](#)
- [Activez les autorisations de Lake Formation pour votre site Amazon S3](#)

Vérifiez les autorisations de Lake Formation

Avant d'enregistrer un emplacement, effectuez une étape de vérification pour vous assurer que les bons directeurs disposent des autorisations requises pour la formation de Lake et qu'aucune autorisation de Lake Formation n'est accordée aux directeurs qui ne devraient pas en avoir. À l'aide de l'opération `GetEffectivePermissionsForPathAPI` Lake Formation, identifiez les ressources du catalogue de données qui font référence à l'emplacement Amazon S3, ainsi que les principaux détenteurs d'autorisations sur ces ressources.

L' AWS CLI exemple suivant renvoie les bases de données et les tables du catalogue de données qui font référence au compartiment Amazon S3products.

```
aws lakeformation get-effective-permissions-for-path --resource-arn
arn:aws:s3:::products --profile datalake_admin
```

Notez l'`profile` option. Nous vous recommandons d'exécuter la commande en tant qu'administrateur du lac de données.

Ce qui suit est un extrait des résultats renvoyés.

```
{
  "PermissionsWithGrantOption": [
    "SELECT"
  ],
  "Resource": {
```

```
        "TableWithColumns": {
            "Name": "inventory_product",
            "ColumnWildcard": {},
            "DatabaseName": "inventory"
        },
        "Permissions": [
            "SELECT"
        ],
        "Principal": {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1",
            "DataLakePrincipalType": "IAM_USER"
        }
    }, ...
```

Important

Si votre catalogue de AWS Glue données est chiffré, il ne `GetEffectivePermissionsForPath` renvoie que les bases de données et les tables créées ou modifiées après la mise à disposition générale de Lake Formation.

Sécuriser les ressources du catalogue de données existantes

Révoquez ensuite l'Superautorisation pour chaque table et `IAMAllowedPrincipals` base de données que vous avez identifiées pour l'emplacement.

Warning

Si vous avez mis en place une automatisation qui crée des bases de données et des tables dans le catalogue de données, les étapes suivantes peuvent entraîner l'échec des tâches d'automatisation et d'extraction, de transformation et de chargement (ETL) en aval. Procédez uniquement après avoir modifié vos processus existants ou accordé des autorisations explicites de Lake Formation aux principaux responsables requis. Pour plus d'informations sur les autorisations de Lake Formation, consultez [the section called "Référence des autorisations de Lake Formation"](#).

Pour révoquer **Super** depuis **IAMAllowedPrincipals** une table

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, choisissez Tables.
3. Sur la page Tables, sélectionnez le bouton radio à côté du tableau souhaité.
4. Dans le menu Actions, choisissez Révoquer.
5. Dans la boîte de dialogue Révoquer les autorisations, dans la liste des utilisateurs et des rôles IAM, faites défiler la page jusqu'à l'en-tête Groupe, puis sélectionnez IAMAllowed Principaux.
6. Sous Autorisations relatives aux tables, assurez-vous que Super est sélectionné, puis choisissez Révoquer.

Pour révoquer l'accès **Super** à une **IAMAllowedPrincipals** base de données

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le panneau de navigation, choisissez Databases (Bases de données).
3. Sur la page Bases de données, sélectionnez le bouton radio à côté de la base de données souhaitée.
4. Dans le menu Actions, choisissez Modifier.
5. Sur la page Modifier la base de données, désactivez l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données, puis sélectionnez Enregistrer.
6. De retour sur la page Bases de données, assurez-vous que la base de données est toujours sélectionnée, puis dans le menu Actions, choisissez Révoquer.
7. Dans la boîte de dialogue Révoquer les autorisations, dans la liste des utilisateurs et des rôles IAM, faites défiler la page jusqu'à l'en-tête Groupe, puis sélectionnez IAMAllowed Principaux.
8. Sous Autorisations de base de données, assurez-vous que Super est sélectionné, puis choisissez Révoque.

Activez les autorisations de Lake Formation pour votre site Amazon S3

Enregistrez ensuite l'emplacement Amazon S3 auprès de Lake Formation. Pour ce faire, vous pouvez utiliser le processus décrit dans [Ajouter un emplacement Amazon S3 à votre lac de](#)

[données](#). Vous pouvez également utiliser l'opération d'`RegisterResourceAPI` décrite dans [Vente d'informations d'identification APIs](#).

 Note

Si un établissement parent est enregistré, il n'est pas nécessaire d'enregistrer un établissement pour enfants.

Après avoir terminé ces étapes et vérifié que vos utilisateurs peuvent accéder à leurs données, vous avez passé avec succès aux autorisations Lake Formation. Passez à l'étape suivante, [Étape 5 : Sécuriser les nouvelles ressources du catalogue de données](#).

Étape 5 : Sécuriser les nouvelles ressources du catalogue de données

Sécurisez ensuite toutes les nouvelles ressources du catalogue de données en modifiant les paramètres par défaut du catalogue de données. Désactivez les options permettant d'utiliser uniquement le contrôle d'accès Gestion des identités et des accès AWS (IAM) pour les nouvelles bases de données et tables.

 Warning

Si vous avez mis en place une automatisation qui crée des bases de données et des tables dans le catalogue de données, les étapes suivantes peuvent entraîner l'échec des tâches d'automatisation et d'extraction, de transformation et de chargement (ETL) en aval. Procédez uniquement après avoir modifié vos processus existants ou accordé des autorisations explicites de Lake Formation aux principaux responsables requis. Pour plus d'informations sur les autorisations de Lake Formation, consultez [the section called "Référence des autorisations de Lake Formation"](#).

Pour modifier les paramètres par défaut du catalogue de données

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'utilisateur administratif IAM (l'utilisateur Administrator ou un autre utilisateur doté de la politique AdministratorAccess AWS gérée).
2. Dans le panneau de navigation, sélectionnez Settings (Paramètres).

3. Sur la page des paramètres du catalogue de données, décochez les deux cases, puis choisissez Enregistrer.

L'étape suivante consiste à accorder aux utilisateurs l'accès à des bases de données ou à des tables supplémentaires à l'avenir. Consultez [Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM pour l'accès futur au data lake](#).

Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM pour l'accès futur au data lake

Pour accorder à vos utilisateurs l'accès à des bases de données ou à des tables supplémentaires du catalogue de données à l'avenir, vous devez leur donner la politique générale Gestion des identités et des accès AWS (IAM) intégrée ci-dessous. Nommez la stratégie `GlueFullReadAccess`.

Important

Si vous associez cette politique à un utilisateur avant de révoquer l'accès à chaque `IAMAllowedPrincipals` base Super de données et table de votre catalogue de données, cet utilisateur peut consulter toutes les métadonnées de toutes les ressources pour lesquelles Super il est accordé à `IAMAllowedPrincipals`.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "GlueFullReadAccess",
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "glue:GetTable",
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetPartitions"
      ]
    }
  ],
}
```

```
    "Resource": "*"
  }
}
```

Note

Les politiques intégrées définies dans cette étape et dans les étapes précédentes contiennent des autorisations IAM minimales. Pour les politiques suggérées aux administrateurs de lacs de données, aux analystes de données et à d'autres personnes, consultez [the section called “Référence des personnalités de Lake Formation et des autorisations IAM”](#).

Ensuite, passez à [Étape 7 : Nettoyer les politiques IAM existantes](#).

Étape 7 : Nettoyer les politiques IAM existantes

Après avoir configuré les AWS Lake Formation autorisations et créé et attaché les politiques de contrôle d'accès Gestion des identités et des accès AWS (IAM) grossières, effectuez la dernière étape suivante :

- Supprimez des utilisateurs, des groupes et des rôles les anciennes politiques détaillées [de contrôle d'accès](#) IAM que vous avez répliquées dans Lake Formation.

Ce faisant, vous vous assurez que ces principaux n'ont plus accès directement aux données d'Amazon Simple Storage Service (Amazon S3). Vous pouvez ensuite gérer l'accès au data lake pour ces principaux acteurs entièrement par le biais de Lake Formation.

AWS Lake Formation et points de terminaison VPC d'interface ()AWS PrivateLink

Amazon VPC est un AWS service que vous pouvez utiliser pour lancer AWS des ressources dans un réseau virtuel que vous définissez. Avec un VPC, vous contrôlez des paramètres réseau, tels que la plage d'adresses IP, les sous-réseaux, les tables de routage et les passerelles réseau.

Si vous utilisez Amazon Virtual Private Cloud (Amazon VPC) pour héberger vos AWS ressources, vous pouvez établir une connexion privée entre votre VPC et Lake Formation. Vous utilisez cette connexion pour que Lake Formation puisse communiquer avec les ressources de votre VPC sans passer par l'Internet public.

Vous pouvez établir une connexion privée entre votre VPC et créer un point de terminaison VPC d'interface. Les points de terminaison de l'interface sont alimentés par [AWS PrivateLink](#) une technologie qui vous permet d'accéder à Lake Formation en privé APIs sans passerelle Internet, appareil NAT, connexion VPN ou Direct Connect connexion. Les instances de votre VPC n'ont pas besoin d'adresses IP publiques pour communiquer avec Lake Formation. APIs Le trafic entre votre VPC et Lake Formation ne quitte pas le réseau Amazon.

Chaque point de terminaison d'interface est représenté par une ou plusieurs [interfaces réseau Elastic](#) dans vos sous-réseaux.

Pour de plus amples informations, consultez [Points de terminaison VPC \(AWS PrivateLink\)](#) dans le Guide de l'utilisateur Amazon VPC.

Considérations relatives aux points de terminaison du VPC de la Lake Formation

Avant de configurer un point de terminaison VPC d'interface pour Lake Formation, assurez-vous de consulter les [propriétés et les limites du point de terminaison d'interface](#) dans le guide de l'utilisateur Amazon VPC.

Lake Formation prend en charge les appels à toutes ses actions d'API depuis votre VPC. Vous pouvez utiliser Lake Formation avec des points de terminaison VPC dans tout Régions AWS ce qui prend en charge à la fois les points de terminaison Lake Formation et Amazon VPC.

Création d'un point de terminaison VPC d'interface pour Lake Formation

Vous pouvez créer un point de terminaison VPC pour le service Lake Formation à l'aide de la console Amazon VPC ou du (). AWS Command Line Interface AWS CLI Pour plus d'informations, consultez [Création d'un point de terminaison d'interface](#) dans le Guide de l'utilisateur Amazon VPC.

Créez un point de terminaison VPC pour Lake Formation en utilisant le nom de service suivant :

- `com.amazonaws. region. formation lacustre`

Si vous activez le DNS privé pour le point de terminaison, vous pouvez envoyer des demandes d'API à Lake Formation en utilisant son nom DNS par défaut pour la région, par exemple, `lakeformation.us-east-1.amazonaws.com`.

Pour plus d'informations, consultez [Accès à un service via un point de terminaison d'interface](#) dans le Guide de l'utilisateur Amazon VPC.

Création d'une politique de point de terminaison VPC pour Lake Formation

Lake Formation prend en charge les politiques relatives aux points de terminaison des VPC. Une politique de point de terminaison est une politique basée sur les ressources que vous attachez à un point de terminaison VPC pour contrôler quels AWS principaux peuvent utiliser le point de terminaison pour accéder à un service. AWS

Vous pouvez associer une politique de point de terminaison à votre point de terminaison VPC qui contrôle l'accès à Lake Formation. La politique spécifie les informations suivantes :

- Le principal qui peut exécuter des actions.
- Les actions qui peuvent être effectuées.
- Les ressources sur lesquelles les actions peuvent être exécutées.

Pour plus d'informations, consultez [Contrôle de l'accès aux services avec points de terminaison d'un VPC](#) dans le Guide de l'utilisateur Amazon VPC.

Exemple : politique de point de terminaison VPC pour les actions de Lake Formation

L'exemple de politique de point de terminaison VPC suivant pour Lake Formation permet la vente d'informations d'identification à l'aide des autorisations de Lake Formation. Vous pouvez utiliser cette politique pour exécuter des requêtes à l'aide des autorisations de Lake Formation à partir d'un cluster Amazon Redshift ou d'un Amazon EMR cluster situé dans un sous-réseau privé.

```
{
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "lakeformation:GetDataAccess",
      "Resource": "*",
      "Principal": "*"
    }
  ]
}
```

```
}
```

Note

Si vous n'attachez pas de politique lorsque vous créez un point de terminaison, une politique par défaut autorisant un accès complet au service est jointe.

Pour plus d'informations, consultez les rubriques suivantes dans la documentation Amazon VPC :

- [Qu'est-ce qu'Amazon VPC ?](#)
- [Création d'un point de terminaison d'interface](#)
- [Utiliser les politiques de point de terminaison VPC](#)

AWS Lake Formation tutoriels

Les didacticiels suivants sont organisés en trois parties et fournissent des step-by-step instructions sur la façon de créer un lac de données, d'ingérer des données, de partager et de sécuriser des lacs de données à l'aide AWS Lake Formation de :

1. Créez un lac de données et ingérez des données : apprenez à créer un lac de données et à utiliser des plans pour déplacer, stocker, cataloguer, nettoyer et organiser vos données. Vous apprendrez également à configurer des tables gouvernées. Une table gouvernée est un nouveau type de table Amazon S3 qui prend en charge les transactions atomiques, cohérentes, isolées et durables (ACID).

Avant de commencer, assurez-vous d'avoir terminé les étapes de [Débuter avec Lake Formation](#).

- [Création d'un lac de données à partir d'une AWS CloudTrail source](#)

Créez et chargez votre premier lac de données en utilisant vos propres CloudTrail journaux comme source de données.

- [Création d'un lac de données à partir d'une source JDBC dans Lake Formation](#)

Créez un lac de données en utilisant l'un de vos magasins de données accessibles à JDBC, tel qu'une base de données relationnelle, comme source de données.

2. Sécurisation des lacs de données : apprenez à utiliser des contrôles d'accès basés sur des balises et au niveau des lignes pour sécuriser et gérer efficacement l'accès à vos lacs de données.

- [Configuration des autorisations pour les formats de stockage de tables ouvertes dans Lake Formation](#)

Ce didacticiel explique comment configurer des autorisations pour les formats de tables transactionnels open source (tables Apache Iceberg, Apache Hudi et Linux Foundation Delta Lake) dans Lake Formation.

- [Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation](#)

Apprenez à gérer l'accès aux données d'un lac de données à l'aide du contrôle d'accès basé sur des balises dans Lake Formation.

- [Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes](#)

Apprenez à configurer des autorisations au niveau des lignes qui vous permettent de restreindre l'accès à des lignes spécifiques en fonction des politiques de conformité et de gouvernance des données de Lake Formation.

3. Partage de données : apprenez à partager vos données en toute sécurité à Comptes AWS l'aide du contrôle d'accès basé sur les balises (TBAC) et à gérer les autorisations détaillées sur les ensembles de données partagés entre eux. Comptes AWS

- [Partage d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation et de ressources nommées](#)

Dans ce didacticiel, vous apprendrez à partager vos données en toute sécurité à Comptes AWS l'aide de Lake Formation.

- [Partage d'un lac de données à l'aide du contrôle d'accès précis de Lake Formation](#)

Dans ce didacticiel, vous apprendrez à partager rapidement et facilement des ensembles de données à l'aide de Lake Formation lors de la gestion de plusieurs ensembles de données Comptes AWS avec AWS Organizations.

Rubriques

- [Création d'un lac de données à partir d'une AWS CloudTrail source](#)
- [Création d'un lac de données à partir d'une source JDBC dans Lake Formation](#)
- [Configuration des autorisations pour les formats de stockage de tables ouvertes dans Lake Formation](#)
- [Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation](#)
- [Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes](#)
- [Partage d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation et de ressources nommées](#)
- [Partage d'un lac de données à l'aide du contrôle d'accès précis de Lake Formation](#)

Création d'un lac de données à partir d'une AWS CloudTrail source

Ce didacticiel vous explique les actions à effectuer sur la console Lake Formation pour créer et charger votre premier lac de données à partir d'une AWS CloudTrail source.

Étapes de haut niveau pour créer un lac de données

1. Enregistrez un chemin Amazon Simple Storage Service (Amazon S3) en tant que lac de données.
2. Accordez à Lake Formation l'autorisation d'écrire dans le catalogue de données et dans les emplacements Amazon S3 du lac de données.
3. Créez une base de données pour organiser les tables de métadonnées dans le catalogue de données.
4. Utilisez un plan pour créer un flux de travail. Exécutez le flux de travail pour ingérer les données d'une source de données.
5. Configurez vos autorisations Lake Formation pour permettre à d'autres personnes de gérer les données du catalogue de données et du lac de données.
6. Configurez Amazon Athena pour interroger les données que vous avez importées dans votre lac de données Amazon S3.
7. Pour certains types de banques de données, configurez Amazon Redshift Spectrum pour interroger les données que vous avez importées dans votre lac de données Amazon S3.

Rubriques

- [Public visé](#)
- [Conditions préalables](#)
- [Étape 1 : créer un utilisateur d'analyste de données](#)
- [Étape 2 : ajouter des autorisations pour lire les AWS CloudTrail journaux au rôle de flux de travail](#)
- [Étape 3 : créer un compartiment Amazon S3 pour le lac de données](#)
- [Étape 4 : enregistrer un chemin Amazon S3](#)
- [Étape 5 : accorder des autorisations de localisation des données](#)
- [Étape 6 : Création d'une base de données dans le catalogue de données](#)
- [Étape 7 : Accorder des autorisations de données](#)
- [Étape 8 : Utiliser un plan pour créer un flux de travail](#)
- [Étape 9 : Exécuter le flux de travail](#)
- [Étape 10 : Autorisez SELECT sur les tables](#)
- [Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena](#)

Public visé

Le tableau suivant répertorie les rôles utilisés dans ce didacticiel pour créer un lac de données.

Public visé

Role	Description
Administrateur IAM	Possède la politique AWS gérée :AdministratorAccess . Peut créer des rôles IAM et des compartiments Amazon S3.
Administrateur du lac de données	Utilisateur autorisé à accéder au catalogue de données, à créer des bases de données et à accorder des autorisations Lake Formation à d'autres utilisateurs. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données.
Analyste des données	Utilisateur capable d'exécuter des requêtes sur le lac de données. Dispose uniquement des autorisations suffisantes pour exécuter des requêtes.
Rôle du flux de travail	Rôle doté des politiques IAM requises pour exécuter un flux de travail. Pour de plus amples informations, veuillez consulter (Facultatif) Créez un rôle IAM pour les flux de travail.

Conditions préalables

Avant de commencer :

- Assurez-vous d'avoir terminé les tâches dans [Configurez AWS Lake Formation](#).
- Connaissez l'emplacement de vos CloudTrail journaux.
- Athena a besoin du personnage d'analyste de données pour créer un compartiment Amazon S3 pour stocker les résultats des requêtes avant d'utiliser Athena.

La connaissance de Gestion des identités et des accès AWS (IAM) est supposée. Pour plus d'informations sur IAM, consultez le guide de l'[utilisateur IAM](#).

Étape 1 : créer un utilisateur d'analyste de données

Cet utilisateur dispose du minimum d'autorisations pour interroger le lac de données.

1. Ouvrez la console IAM à l'adresse <https://console.aws.amazon.com/iam>. Connectez-vous en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) ou en tant qu'utilisateur avec la politique AdministratorAccess AWS gérée.
2. Créez un utilisateur nommé `dataLake_user` avec les paramètres suivants :
 - Activez AWS Management Console l'accès.
 - Définissez un mot de passe et n'exigez pas de réinitialisation du mot de passe.
 - Joignez la politique AmazonAthenaFullAccess AWS gérée.
 - Joignez la politique intégrée suivante. Nommez la stratégie `DataLakeUserBasic`.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "glue:GetTable",
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetPartitions",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:GetLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
      ],
      "Resource": "*"
    }
  ]
}
```

Étape 2 : ajouter des autorisations pour lire les AWS CloudTrail journaux au rôle de flux de travail

1. Associez la politique intégrée suivante au rôle `LakeFormationWorkflowRole`. La politique autorise la lecture de vos AWS CloudTrail journaux. Nommez la stratégie `DatalakeGetCloudTrail`.

Pour créer le rôle `LakeFormationWorkflowRole`, consultez [\(Facultatif\) Créez un rôle IAM pour les flux de travail](#).

Important

Remplacez `<your-s3-cloudtrail-bucket>` par l'emplacement Amazon S3 de vos CloudTrail données.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "s3:GetObject",
      "Resource": ["arn:aws:s3:::<your-s3-cloudtrail-bucket>/*"]
    }
  ]
}
```

2. Vérifiez que trois politiques sont associées au rôle.

Étape 3 : créer un compartiment Amazon S3 pour le lac de données

Créez le compartiment Amazon S3 qui doit être l'emplacement racine de votre lac de données.

1. Ouvrez la console Amazon S3 à <https://console.aws.amazon.com/s3/> et connectez-vous en tant qu'utilisateur administrateur dans lequel vous l'avez créée [Création d'un utilisateur doté d'un accès administratif](#).

2. Choisissez Create bucket, puis passez par l'assistant pour créer un bucket nommé `<yourName>-datalake-cloudtrail`, où `<yourName>` sont votre prénom et votre nom de famille. Par exemple : `jdoe-datalake-cloudtrail`.

Pour obtenir des instructions détaillées sur la création d'un compartiment Amazon S3, consultez [Création d'un compartiment](#).

Étape 4 : enregistrer un chemin Amazon S3

Enregistrez un chemin Amazon S3 comme emplacement racine de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
3. Choisissez Enregistrer l'emplacement, puis Parcourir.
4. Sélectionnez le `<yourName>-datalake-cloudtrail` bucket que vous avez créé précédemment, acceptez le rôle IAM par défaut `AWSServiceRoleForLakeFormationDataAccess`, puis choisissez Enregistrer l'emplacement.

Pour plus d'informations sur l'enregistrement des points de vente, consultez [Ajouter un emplacement Amazon S3 à votre lac de données](#).

Étape 5 : accorder des autorisations de localisation des données

Les principaux doivent disposer d'autorisations de localisation des données sur l'emplacement d'un lac de données pour créer des tables de catalogue de données ou des bases de données pointant vers cet emplacement. Vous devez accorder des autorisations de localisation des données au rôle IAM pour les flux de travail afin que le flux de travail puisse écrire sur la destination d'ingestion des données.

1. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données.
2. Choisissez Grant, puis dans la boîte de dialogue Accorder les autorisations, effectuez les sélections suivantes :
 - a. Pour l'utilisateur et les rôles IAM, sélectionnez `LakeFormationWorkflowRole`.

- b. Pour les emplacements de stockage, choisissez votre `<yourName>-datalake-cloudtrail` compartiment.
3. Choisissez Accorder.

Pour plus d'informations sur les autorisations de localisation des données, consultez [Underlying data access control](#).

Étape 6 : Création d'une base de données dans le catalogue de données

Les tables de métadonnées du Lake Formation Data Catalog sont stockées dans une base de données.

1. Dans le volet de navigation, sous Catalogue de données, sélectionnez Bases de données.
2. Choisissez Créer une base de données, puis entrez le nom sous Détails de la base de données `lakeformation_cloudtrail`.
3. Laissez les autres champs vides, puis choisissez Créer une base de données.

Étape 7 : Accorder des autorisations de données

Vous devez accorder des autorisations pour créer des tables de métadonnées dans le catalogue de données. Étant donné que le flux de travail sera exécuté avec le rôle `LakeFormationWorkflowRole`, vous devez accorder ces autorisations au rôle.

1. Dans la console Lake Formation, dans le volet de navigation, sous Catalogue de données, sélectionnez Databases.
2. Choisissez la `lakeformation_cloudtrail` base de données, puis, dans la liste déroulante Actions, choisissez Grant sous le titre Permissions.
3. Dans la boîte de dialogue Accorder les autorisations relatives aux données, effectuez les sélections suivantes :
 - a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez `LakeFormationWorkflowRole`.
 - b. Sous Balises LF ou ressources de catalogue, choisissez Ressources de catalogue de données nommées.
 - c. Pour les bases de données, vous devriez voir que la `lakeformation_cloudtrail` base de données est déjà ajoutée.

- d. Sous Autorisations de base de données, sélectionnez Créer une table, Modifier et Supprimer, puis désélectionnez Super si cette option est sélectionnée.
4. Choisissez Accorder.

Pour plus d'informations sur l'octroi des autorisations de Lake Formation, consultez [Gestion des autorisations relatives à Lake Formation](#).

Étape 8 : Utiliser un plan pour créer un flux de travail

Afin de lire les CloudTrail journaux, de comprendre leur structure et de créer les tables appropriées dans le catalogue de données, nous devons configurer un flux de travail composé d'un robot d'AWS Glueexploration, de tâches, de déclencheurs et de flux de travail. Les plans de Lake Formation simplifient ce processus.

Le flux de travail génère les tâches, les robots d'exploration et les déclencheurs qui découvrent et ingèrent les données dans votre lac de données. Vous créez un flux de travail basé sur l'un des plans prédéfinis de Lake Formation.

1. Dans la console Lake Formation, dans le volet de navigation, choisissez Blueprints sous Ingestion, puis choisissez Use blueprint.
2. Sur la page Utiliser un plan, sous Type de plan, sélectionnez AWS CloudTrail
3. Sous Importer la source, choisissez une CloudTrail source et une date de début.
4. Sous Cible d'importation, spécifiez les paramètres suivants :

Base de données cible	lakeformation_cloudtrail
Emplacement de stockage cible	s3://<yourName> -datalake-cloudtrail
Format de données	Parquet

5. Pour la fréquence d'importation, choisissez Exécuter à la demande.
6. Sous Options d'importation, spécifiez les paramètres suivants :

Nom du flux de travail	lakeformationcloudtrailtest
Rôle IAM	LakeFormationWorkflowRole

Préfixe de table

cloudtrailtest

 Note

Doit être en minuscules.

7. Choisissez Créer et attendez que la console indique que le flux de travail a été créé avec succès.

 Tip

Avez-vous reçu le message d'erreur suivant ?

```
User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...
```

Si tel est le cas, vérifiez que vous avez remplacé `<account-id>` dans la politique en ligne l'utilisateur administrateur du lac de données par un numéro de AWS compte valide.

Étape 9 : Exécuter le flux de travail

Comme vous avez indiqué que le flux de travail est le cas run-on-demand, vous devez le démarrer manuellement.

- Sur la page Blueprints, sélectionnez le flux de travail `lakeformationcloudtrailtest`, puis dans le menu Actions, sélectionnez Démarrer.

Au fur et à mesure que le flux de travail s'exécute, vous pouvez voir sa progression dans la colonne État de la dernière exécution. Cliquez sur le bouton d'actualisation de temps en temps.

Le statut passe de « EN COURS » à « Découverte », puis « Importation », puis « TERMINÉ ».

Lorsque le flux de travail est terminé :

- Le catalogue de données comportera de nouvelles tables de métadonnées.
- Vos CloudTrail journaux seront ingérés dans le lac de données.

Si le flux de travail échoue, procédez comme suit :

- a. Sélectionnez le flux de travail, puis dans le menu Actions, choisissez Afficher le graphique.
Le flux de travail s'ouvre dans la AWS Glue console.
- b. Assurez-vous que le flux de travail est sélectionné, puis sélectionnez l'onglet History (Historique).
- c. Sous Historique, sélectionnez l'exécution la plus récente et choisissez Afficher les détails de la course.
- d. Sélectionnez une tâche ou un robot d'exploration ayant échoué dans le graphique dynamique (d'exécution) et consultez le message d'erreur. Les nœuds défectueux sont rouges ou jaunes.

Étape 10 : Autorisez SELECT sur les tables

Vous devez accorder l'authorisation SELECT sur les nouvelles tables du catalogue de données afin que l'analyste de données puisse interroger les données vers lesquelles pointent les tables.

Note

Un flux de travail accorde automatiquement l'authorisation SELECT sur les tables qu'il crée à l'utilisateur qui l'a exécuté. Étant donné que l'administrateur du lac de données a exécuté ce flux de travail, vous devez accorder une autorisation SELECT à l'analyste de données.

1. Dans la console Lake Formation, dans le volet de navigation, sous Catalogue de données, sélectionnez Databases.
2. Choisissez la `lakeformation_cloudtrail` base de données, puis, dans la liste déroulante Actions, choisissez Grant sous le titre Permissions.
3. Dans la boîte de dialogue Accorder les autorisations relatives aux données, effectuez les sélections suivantes :
 - a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez `datalake_user`
 - b. Sous Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées.

- c. Pour les bases de données, la `lakeformation_cloudtrail` base de données doit déjà être sélectionnée.
 - d. Pour Tables, sélectionnez `cloudtrailtest-cloudtrail`.
 - e. Sous Autorisations relatives aux tables et aux colonnes, choisissez Sélectionner.
4. Choisissez Accorder.

L'étape suivante est exécutée en tant qu'analyste de données.

Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena

Utilisez la Amazon Athena console pour interroger les CloudTrail données de votre lac de données.

1. Ouvrez la console Athena à <https://console.aws.amazon.com/athena/> et connectez-vous en tant qu'analyste de données, utilisateur. `datalake_user`
2. Si nécessaire, choisissez Get Started pour passer à l'éditeur de requêtes Athena.
3. Pour Data source (Source de données), choisissez `AwsDataCatalog`.
4. Pour Database (Base de données), sélectionnez `lakeformation_cloudtrail`.

La liste des tables s'affiche.

5. Dans le menu déroulant (3 points disposés horizontalement) à côté du tableau `cloudtrailtest-cloudtrail`, choisissez Aperçu du tableau, puis Exécuter.

La requête s'exécute et affiche 10 lignes de données.

Si vous n'avez jamais utilisé Athena auparavant, vous devez d'abord configurer un emplacement Amazon S3 dans la console Athena pour stocker les résultats de la requête. Ils `datalake_user` doivent disposer des autorisations nécessaires pour accéder au compartiment Amazon S3 de votre choix.

Note

Maintenant que vous avez terminé le didacticiel, accordez des autorisations relatives aux données et des autorisations de localisation des données aux principaux responsables de votre organisation.

Création d'un lac de données à partir d'une source JDBC dans Lake Formation

Ce didacticiel vous explique les étapes à suivre sur la AWS Lake Formation console pour créer et charger votre premier lac de données à partir d'une source JDBC à l'aide de Lake Formation.

Rubriques

- [Public visé](#)
- [Prérequis pour le didacticiel JDBC](#)
- [Étape 1 : créer un utilisateur d'analyste de données](#)
- [Étape 2 : créer une connexion dans AWS Glue](#)
- [Étape 3 : créer un compartiment Amazon S3 pour le lac de données](#)
- [Étape 4 : enregistrer un chemin Amazon S3](#)
- [Étape 5 : accorder des autorisations de localisation des données](#)
- [Étape 6 : Création d'une base de données dans le catalogue de données](#)
- [Étape 7 : Accorder des autorisations de données](#)
- [Étape 8 : Utiliser un plan pour créer un flux de travail](#)
- [Étape 9 : Exécuter le flux de travail](#)
- [Étape 10 : Autorisez SELECT sur les tables](#)
- [Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena](#)
- [Étape 12 : interroger les données du lac de données à l'aide d'Amazon Redshift Spectrum](#)
- [Étape 13 : Accorder ou révoquer les autorisations de Lake Formation à l'aide d'Amazon Redshift Spectrum](#)

Public visé

Le tableau suivant répertorie les rôles utilisés dans ce didacticiel [AWS Lake Formation JDBC](#).

Role	Description
Administrateur IAM	Utilisateur capable de créer des utilisateurs et des rôles Gestion des identités et des accès

Role	Description
	AWS (IAM) ainsi que des buckets Amazon Simple Storage Service (Amazon S3). Possède la politique AdministratorAccess AWS gérée.
Administrateur du lac de données	Un utilisateur qui peut accéder au catalogue de données, créer des bases de données et accorder des autorisations Lake Formation à d'autres utilisateurs. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données.
Analyste des données	Un utilisateur qui peut exécuter des requêtes sur le lac de données. Dispose uniquement de suffisamment d'autorisations pour exécuter des requêtes.
Rôle du flux de travail	Rôle doté des politiques IAM requises pour exécuter un flux de travail.

Pour plus d'informations sur les conditions requises pour suivre le didacticiel, consultez [Prérequis pour le didacticiel JDBC](#).

Prérequis pour le didacticiel JDBC

Avant de commencer le [didacticiel AWS Lake Formation JDBC](#), assurez-vous d'avoir effectué les opérations suivantes :

- Effectuez les tâches définies dans [Débuter avec Lake Formation](#).
- Choisissez un magasin de données accessible en JDBC que vous souhaitez utiliser pour le didacticiel.
- Rassemblez les informations nécessaires pour créer une AWS Glue connexion de type JDBC. Cet objet de catalogue de données inclut l'URL du magasin de données, les informations de connexion et, si le magasin de données a été créé dans un Amazon Virtual Private Cloud (Amazon VPC), des informations de configuration supplémentaires spécifiques au VPC. Pour plus d'informations,

consultez la section [Définition des connexions dans le catalogue de AWS Glue données](#) du guide du AWS Glue développeur.

Le didacticiel part du principe que vous connaissez Gestion des identités et des accès AWS (IAM). Pour plus d'informations sur IAM, consultez le guide de l'[utilisateur IAM](#).

Pour commencer, passez à [the section called “Étape 1 : créer un utilisateur d'analyste de données”](#).

Étape 1 : créer un utilisateur d'analyste de données

Au cours de cette étape, vous créez un utilisateur Gestion des identités et des accès AWS (IAM) qui sera l'analyste de données de votre lac de données dans AWS Lake Formation.

Cet utilisateur dispose du minimum d'autorisations pour interroger le lac de données.

1. Ouvrez la console IAM à l'adresse <https://console.aws.amazon.com/iam>. Connectez-vous en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) ou en tant qu'utilisateur avec la politique AdministratorAccess AWS gérée.
2. Créez un utilisateur nommé `dataLake_user` avec les paramètres suivants :
 - Activez AWS Management Console l'accès.
 - Définissez un mot de passe et n'exigez pas de réinitialisation du mot de passe.
 - Joignez la politique AmazonAthenaFullAccess AWS gérée.
 - Joignez la politique intégrée suivante. Nommez la stratégie `DataLakeUserBasic`.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "glue:GetTable",
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetPartitions",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",

```

```
        "lakeformation:GetLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
    ],
    "Resource": "*"
  }
]
```

Étape 2 : créer une connexion dans AWS Glue

Note

Ignorez cette étape si vous êtes déjà AWS Glue connecté à votre source de données JDBC.

AWS Lake Formation accède aux sources de données JDBC via une connexion. AWS Glue Une connexion est un objet du catalogue de données qui contient toutes les informations requises pour se connecter à la source de données. Vous pouvez créer une connexion à l'aide de la AWS Glue console.

Pour créer une connexion

1. Ouvrez AWS Glue la console à <https://console.aws.amazon.com/glue/> l'adresse et connectez-vous en tant qu'utilisateur administrateur dans lequel vous l'avez créée [Création d'un utilisateur doté d'un accès administratif](#).
2. Dans le volet de navigation, sous Data catalog (Catalogue de données), choisissez Connexions (Connexions).
3. Sur la page Connecteurs, choisissez Créer une connexion.
4. Sur la page Choisir une source de données, choisissez JDBC comme type de connexion. Ensuite, sélectionnez Suivant.
5. Continuez à utiliser l'assistant de connexion et enregistrez la connexion.

Pour plus d'informations sur la création d'une connexion, consultez les [propriétés de connexion AWS Glue JDBC](#) dans le manuel du AWS Glue développeur.

Étape 3 : créer un compartiment Amazon S3 pour le lac de données

Au cours de cette étape, vous créez le bucket Amazon Simple Storage Service (Amazon S3) qui doit être l'emplacement racine de votre lac de données.

1. Ouvrez la console Amazon S3 à <https://console.aws.amazon.com/s3/> et connectez-vous en tant qu'utilisateur administrateur dans lequel vous l'avez créée [Création d'un utilisateur doté d'un accès administratif](#).
2. Choisissez Create bucket, puis passez par l'assistant pour créer un bucket nommé `<yourName>-datalake-tutorial`, où `<yourName>` sont votre prénom et votre nom de famille. Par exemple : `jdoe-datalake-tutorial`.

Pour obtenir des instructions détaillées sur la création d'un compartiment Amazon S3, consultez [Comment créer un compartiment S3 ?](#) dans le guide de l'utilisateur d'Amazon Simple Storage Service.

Étape 4 : enregistrer un chemin Amazon S3

Au cours de cette étape, vous devez enregistrer un chemin Amazon Simple Storage Service (Amazon S3) comme emplacement racine de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
3. Choisissez Enregistrer l'emplacement, puis choisissez Parcourir.
4. Sélectionnez le `<yourName>-datalake-tutorial` bucket que vous avez créé précédemment, acceptez le rôle IAM par défaut `AWSServiceRoleForLakeFormationDataAccess`, puis choisissez Enregistrer l'emplacement.

Pour plus d'informations sur l'enregistrement des points de vente, consultez [Ajouter un emplacement Amazon S3 à votre lac de données](#).

Étape 5 : accorder des autorisations de localisation des données

Les principaux doivent disposer d'autorisations de localisation des données sur l'emplacement d'un lac de données pour créer des tables de catalogue de données ou des bases de données pointant

vers cet emplacement. Vous devez accorder des autorisations de localisation des données au rôle IAM pour les flux de travail afin que le flux de travail puisse écrire sur la destination d'ingestion des données.

1. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Data locations.
2. Choisissez Grant, puis dans la boîte de dialogue Accorder des autorisations, procédez comme suit :
 - a. Pour l'utilisateur et les rôles IAM, sélectionnez `LakeFormationWorkflowRole`.
 - b. Pour les emplacements de stockage, choisissez votre `<yourName>-datalake-tutorial` compartiment.
3. Choisissez Accorder.

Pour plus d'informations sur les autorisations de localisation des données, consultez [Underlying data access control](#).

Étape 6 : Création d'une base de données dans le catalogue de données

Les tables de métadonnées du Lake Formation Data Catalog sont stockées dans une base de données.

1. Sur la console Lake Formation, dans le volet de navigation, sous Catalogue de données, sélectionnez Databases.
2. Choisissez Créer une base de données, puis entrez le nom sous Détails de la base de données `lakeformation_tutorial`.
3. Laissez les autres champs vides, puis choisissez Créer une base de données.

Étape 7 : Accorder des autorisations de données

Vous devez accorder des autorisations pour créer des tables de métadonnées dans le catalogue de données. Étant donné que le flux de travail s'exécute avec le rôle `LakeFormationWorkflowRole`, vous devez accorder ces autorisations au rôle.

1. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.

2. Choisissez Grant, puis dans la boîte de dialogue Accorder les autorisations relatives aux données, procédez comme suit :
 - a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez `LakeFormationWorkflowRole`
 - b. Sous Balises LF ou ressources de catalogue, choisissez Ressources de catalogue de données nommées.
 - c. Pour Bases de données, choisissez la base de données que vous avez créée précédemment, `lakeformation_tutorial`.
 - d. Sous Autorisations de base de données, sélectionnez Créer une table, Modifier et Supprimer, puis désélectionnez Super si cette option est sélectionnée.
3. Choisissez Accorder.

Pour plus d'informations sur l'octroi des autorisations de Lake Formation, consultez [Vue d'ensemble des autorisations relatives à Lake Formation](#).

Étape 8 : Utiliser un plan pour créer un flux de travail

Le AWS Lake Formation flux de travail génère les AWS Glue tâches, les robots d'exploration et les déclencheurs qui découvrent et ingèrent les données dans votre lac de données. Vous créez un flux de travail basé sur l'un des plans prédéfinis de Lake Formation.

1. Sur la console Lake Formation, dans le volet de navigation, choisissez Blueprints, puis Use blueprint.
2. Sur la page Utiliser un plan, sous Type de plan, choisissez Instantané de base de données.
3. Sous Importer la source, pour Connexion à la base de données, choisissez la connexion que vous venez de créer ou choisissez une connexion existante pour votre source de données. `datalake-tutorial`
4. Dans le champ Chemin des données source, entrez le chemin à partir duquel les données doivent être ingérées dans le formulaire `<database>/<schema>/<table>`.

Vous pouvez remplacer le caractère générique pour le pourcentage (%) par le schéma ou le tableau. Pour les bases de données qui prennent en charge les schémas, entrez `<database>/<schema>/%` pour qu'il corresponde à toutes les tables qu'elles contiennent. `<schema>` `<database>` Oracle Database et MySQL ne prennent pas en charge le schéma

dans le chemin ; entrez plutôt `<database> /%`. Pour Oracle Database, `<database>` c'est l'identifiant du système (SID).

Par exemple, si une base de données Oracle a `orcl` `orcl/%` pour SID, entrez toutes les tables auxquelles l'utilisateur indiqué dans la connexion JDBC a accès.

**Important**

Ce champ est sensible à la casse.

5. Sous Cible d'importation, spécifiez les paramètres suivants :

Base de données cible	lakeformation_tutorial
Emplacement de stockage cible	s3://<yourName> -datalake-tutorial
Format de données	(Choisissez Parquet ou CSV)

6. Pour la fréquence d'importation, choisissez Exécuter à la demande.

7. Sous Options d'importation, spécifiez les paramètres suivants :

Nom du flux de travail	lakeformationjdbctest
Rôle IAM	LakeFormationWorkflowRole
Préfixe de table	jdbctest

**Note**

Doit être en minuscules.

8. Choisissez Créer et attendez que la console indique que le flux de travail a été créé avec succès.

**Tip**

Avez-vous reçu le message d'erreur suivant ?

```
User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...
```

Si tel est le cas, vérifiez que vous avez remplacé **<account-id>** dans la politique en ligne l'utilisateur administrateur du lac de données par un numéro de AWS compte valide.

Étape 9 : Exécuter le flux de travail

Comme vous avez indiqué que le flux de travail est run-on-demand, vous devez le démarrer manuellement dans AWS Lake Formation.

1. Sur la console Lake Formation, sur la page Blueprints, sélectionnez le flux de travail `lakeformationjdbctest`.
2. Choisissez Actions, puis sélectionnez Démarrer.
3. Au fur et à mesure que le flux de travail s'exécute, consultez sa progression dans la colonne État de la dernière exécution. Cliquez sur le bouton d'actualisation de temps en temps.

Le statut passe de « EN COURS » à « Découverte », puis « Importation », puis « TERMINÉ ».

Lorsque le flux de travail est terminé :

- Le catalogue de données contient de nouvelles tables de métadonnées.
- Vos données sont ingérées dans le lac de données.

Si le flux de travail échoue, procédez comme suit :

- a. Sélectionnez un flux de travail. Choisissez Actions, puis Afficher le graphique.

Le flux de travail s'ouvre dans la AWS Glue console.

- b. Sélectionnez le flux de travail et cliquez sur l'onglet Historique.
- c. Sélectionnez l'exécution la plus récente et choisissez Afficher les détails de la course.
- d. Sélectionnez une tâche ou un robot d'exploration ayant échoué dans le graphique dynamique (d'exécution) et consultez le message d'erreur. Les nœuds défectueux sont rouges ou jaunes.

Étape 10 : Autorisez SELECT sur les tables

Vous devez accorder l'authorisation SELECT sur les nouvelles tables du catalogue de données AWS Lake Formation afin que l'analyste de données puisse interroger les données vers lesquelles pointent les tables.

Note

Un flux de travail accorde automatiquement l'authorisation SELECT sur les tables qu'il crée à l'utilisateur qui l'a exécuté. Étant donné que l'administrateur du lac de données a exécuté ce flux de travail, vous devez accorder une autorisation SELECT à l'analyste de données.

1. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.
2. Choisissez Grant, puis dans la boîte de dialogue Accorder les autorisations relatives aux données, procédez comme suit :
 - a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez `datalake_user`
 - b. Sous Balises LF ou ressources de catalogue, choisissez Ressources de catalogue de données nommées.
 - c. Pour Bases de données, sélectionnez `lakeformation_tutorial`.

La liste des tables s'affiche.
 - d. Pour Tables, choisissez une ou plusieurs tables dans votre source de données.
 - e. Sous Autorisations relatives aux tables et aux colonnes, choisissez Sélectionner.
3. Choisissez Accorder.

L'étape suivante est exécutée en tant qu'analyste de données.

Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena

Utilisez la Amazon Athena console pour interroger les données de votre lac de données.

1. Ouvrez la console Athena sur <https://console.aws.amazon.com/athena/>, et connectez-vous en tant qu'analyste de données, utilisateur `datalake_user`
2. Si nécessaire, choisissez Get Started pour passer à l'éditeur de requêtes Athena.

3. Pour Data source (Source de données), choisissez `AwsDataCatalog`.
4. Pour Database (Base de données), sélectionnez `lakeformation_tutorial`.

La liste des tables s'affiche.

5. Dans le menu contextuel situé à côté de l'un des tableaux, choisissez Aperçu du tableau.

La requête s'exécute et affiche 10 lignes de données.

Étape 12 : interroger les données du lac de données à l'aide d'Amazon Redshift Spectrum

Vous pouvez configurer Amazon Redshift Spectrum pour interroger les données que vous avez importées dans votre lac de données Amazon Simple Storage Service (Amazon S3). Créez d'abord un rôle Gestion des identités et des accès AWS (IAM) utilisé pour lancer le cluster Amazon Redshift et pour interroger les données Amazon S3. Accordez ensuite à ce rôle les `Select` autorisations sur les tables que vous souhaitez interroger. Accordez ensuite à l'utilisateur l'autorisation d'utiliser l'éditeur de requêtes Amazon Redshift. Enfin, créez un cluster Amazon Redshift et exécutez des requêtes.

Vous créez le cluster en tant qu'administrateur et vous l'interrogez en tant qu'analyste de données.

Pour plus d'informations sur Amazon Redshift Spectrum, [consultez la section Utilisation d'Amazon Redshift Spectrum pour interroger](#) des données externes dans le manuel Amazon Redshift Database Developer Guide.

Pour configurer les autorisations permettant d'exécuter des requêtes Amazon Redshift

1. Ouvrez la console IAM à l'adresse <https://console.aws.amazon.com/iam/>. Connectez-vous en tant qu'utilisateur administrateur que vous avez créé [Création d'un utilisateur doté d'un accès administratif](#) (nom d'utilisateur `Administrator`) ou en tant qu'utilisateur doté de la politique `AdministratorAccess` AWS gérée.
2. Dans le panneau de navigation, choisissez Politiques.

Si vous sélectionnez Politiques pour la première fois, la page Bienvenue dans les politiques gérées s'affiche. Sélectionnez Get started (Mise en route).

3. Choisissez Create Policy (Créer une politique).
4. Choisissez l'onglet JSON.

5. Collez le document de politique JSON suivant.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "glue:GetTable",
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetPartitions",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:GetLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
      ],
      "Resource": "*"
    }
  ]
}
```

6. Lorsque vous avez terminé, choisissez Review (Vérifier) pour vérifier la politique. Le programme de validation des politiques signale les éventuelles erreurs de syntaxe.
7. Sur la page Réviser la politique, entrez le nom **RedshiftLakeFormationPolicy** de la politique que vous créez. (Facultatif) Entrez une description. Vérifiez le récapitulatif de politique pour voir les autorisations accordées par votre politique. Sélectionnez ensuite Créer une politique pour enregistrer votre travail.
8. Dans le volet de navigation de la console IAM, sélectionnez Roles (Rôles), puis Create role (Créer un rôle).
9. Pour Sélectionner une entité de confiance, choisissez Service AWS .
10. Choisissez le service Amazon Redshift pour assumer ce rôle.
11. Choisissez le cas d'utilisation Redshift Customizable (Redshift personnalisable) pour votre service. Choisissez ensuite Suivant : Autorisations.

12. Recherchez la politique d'autorisation que vous avez créée et cochez la case à côté du nom de la politique dans la liste. `RedshiftLakeFormationPolicy`
13. Choisissez Suivant : Balises.
14. Choisissez Suivant : Vérification.
15. Pour Role name (Nom du rôle), entrez le nom **RedshiftLakeFormationRole**.
16. (Facultatif) Dans le champ Description du rôle, saisissez la description du nouveau rôle.
17. Passez en revue les informations du rôle, puis choisissez Create role (Créer un rôle).

Pour accorder **Select** des autorisations sur la table à interroger dans la base de données Lake Formation

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données, puis choisissez Grant.
3. Saisissez les informations suivantes :
 - Pour les utilisateurs et les rôles IAM, choisissez le rôle IAM que vous avez créé, `RedshiftLakeFormationRole`. Lorsque vous exécutez l'éditeur de requêtes Amazon Redshift, il utilise ce rôle IAM pour l'autorisation des données.
 - Pour Database (Base de données), sélectionnez `lakeformation_tutorial`.

La liste des tables s'affiche.

 - Pour Table, choisissez une table dans la source de données à interroger.
 - Choisissez l'autorisation Sélectionner une table.
4. Choisissez Accorder.

Pour configurer Amazon Redshift Spectrum et exécuter des requêtes

1. Ouvrez la console Amazon Redshift à l'adresse. <https://console.aws.amazon.com/redshift>. Connectez-vous en tant qu'utilisateur Administrator.
2. Choisissez Créer un cluster.
3. Sur la page Créer un cluster, entrez `redshift-lakeformation-demo` l'identifiant du cluster.
4. Pour le type de nœud, sélectionnez `dc2.large`.

5. Faites défiler la page vers le bas et sous Configurations de base de données, entrez ou acceptez les paramètres suivants :
 - Nom d'utilisateur de l'administrateur : `awsuser`
 - Mot de passe de l'utilisateur administrateur : (*Choose a password*)
6. Développez les autorisations du cluster et, dans la zone Rôles IAM disponibles, sélectionnez `RedshiftLakeFormationRole`. Ensuite, choisissez Ajouter un rôle IAM.
7. Si vous devez utiliser un port différent de la valeur par défaut de 5439, à côté de Configurations supplémentaires, désactivez l'option Utiliser les valeurs par défaut. Développez la section consacrée aux configurations de base de données et entrez un nouveau numéro de port de base de données.
8. Choisissez Créer un cluster.

La page Clusters se charge.

9. Patientez jusqu'à ce que l'état du cluster devienne disponible. Choisissez régulièrement l'icône d'actualisation.
10. Accordez à l'analyste de données l'autorisation d'exécuter des requêtes sur le cluster. Pour ce faire, exécutez les étapes suivantes.
 - a. Ouvrez la console IAM à l'<https://console.aws.amazon.com/iam/>adresse et connectez-vous en tant qu'Administratorutilisateur.
 - b. Dans le volet de navigation, choisissez Utilisateurs et associez les politiques gérées suivantes à l'utilisateur `datalake_user`.
 - `AmazonRedshiftQueryEditor`
 - `AmazonRedshiftReadOnlyAccess`
11. Déconnectez-vous de la console Amazon Redshift et reconnectez-vous en tant qu'utilisateur `datalake_user`
12. Dans la barre d'outils verticale de gauche, cliquez sur l'icône EDITOR pour ouvrir l'éditeur de requêtes et vous connecter au cluster. Si la boîte de dialogue Connect to database apparaît, choisissez le nom du cluster `redshift-lakeformation-demo`, puis entrez le nom de la base de données `dev`, le nom `awsuser` d'utilisateur et le mot de passe que vous avez créés. Ensuite, choisissez Se connecter à la base de données.

 Note

Si aucun paramètre de connexion ne vous est demandé et qu'un autre cluster est déjà sélectionné dans l'éditeur de requêtes, choisissez Modifier la connexion pour ouvrir la boîte de dialogue Connexion à la base de données.

13. Dans la zone de texte New Query 1, entrez et exécutez l'instruction suivante pour mapper la base de données lakeformation_tutorial de Lake Formation au nom du schéma Amazon Redshift : redshift_jdbc

 Important

<account-id> Remplacez-le par un numéro de AWS compte valide et *<region>* par un nom de AWS région valide (par exemple, us-east-1).

```
create external schema if not exists redshift_jdbc from DATA CATALOG
  database 'lakeformation_tutorial' iam_role 'arn:aws:iam::<account-id>:role/
  RedshiftLakeFormationRole' region '<region>;'
```

14. Dans la liste des schémas sous Sélectionner un schéma, choisissez redshift_jdbc.

La liste des tables s'affiche. L'éditeur de requêtes affiche uniquement les tables pour lesquelles vous avez obtenu des autorisations relatives au lac de données de Lake Formation.

15. Dans le menu contextuel situé à côté du nom d'une table, choisissez Prévisualiser les données.

Amazon Redshift renvoie les 10 premières lignes.

Vous pouvez désormais exécuter des requêtes sur les tables et les colonnes pour lesquelles vous avez des autorisations.

Étape 13 : Accorder ou révoquer les autorisations de Lake Formation à l'aide d'Amazon Redshift Spectrum

Amazon Redshift permet d'accorder et de révoquer les autorisations de Lake Formation sur les bases de données et les tables à l'aide d'instructions SQL modifiées. Ces instructions sont similaires aux

instructions Amazon Redshift existantes. Pour plus d'informations, consultez [GRANT](#) et [REVOKE](#) dans le manuel Amazon Redshift Database Developer Guide.

Configuration des autorisations pour les formats de stockage de tables ouvertes dans Lake Formation

AWS Lake Formation prend en charge la gestion des autorisations d'accès pour les formats Open Table (OTFs) tels qu'[Apache Iceberg](#), [Apache Hudi](#) et [la fondation Linux Delta Lake](#). Dans ce didacticiel, vous allez apprendre à créer Iceberg, Hudi et Delta Lake à l'aide de tables [manifestes](#) de liens symboliques AWS Glue, à configurer des autorisations détaillées à l'AWS Glue Data Catalog aide de Lake Formation et à interroger des données à l'aide d'Amazon Athena.

Note

AWS les services d'analyse ne prennent pas en charge tous les formats de tables transactionnels. Pour de plus amples informations, veuillez consulter [Collaboration avec d'autres AWS services](#). Ce didacticiel couvre la création manuelle d'une nouvelle base de données et d'une table dans le catalogue de données à l'aide de AWS Glue tâches uniquement.

Ce didacticiel inclut un AWS CloudFormation modèle pour une configuration rapide. Vous pouvez le consulter et le personnaliser en fonction de vos besoins.

Rubriques

- [Public visé](#)
- [Conditions préalables](#)
- [Étape 1 : Approvisionnez vos ressources](#)
- [Étape 2 : configurer les autorisations pour une table Iceberg](#)
- [Étape 3 : configurer les autorisations pour une table Hudi](#)
- [Étape 4 : Configuration des autorisations pour une table Delta Lake](#)
- [Étape 5 : Nettoyer les AWS ressources](#)

Public visé

Ce didacticiel est destiné aux administrateurs IAM, aux administrateurs de lacs de données et aux analystes commerciaux. Le tableau suivant répertorie les rôles utilisés dans ce didacticiel pour créer une table gouvernée à l'aide de Lake Formation.

Role	Description
Administrateur IAM	Un utilisateur capable de créer des utilisateurs et des rôles IAM ainsi que des compartiments Amazon S3. Possède la politique AdministratorAccess AWS gérée.
Administrateur du lac de données	Un utilisateur qui peut accéder au catalogue de données, créer des bases de données et accorder des autorisations Lake Formation à d'autres utilisateurs. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données.
Analyste commercial	Un utilisateur qui peut exécuter des requêtes sur le lac de données. Dispose des autorisations nécessaires pour exécuter des requêtes.

Conditions préalables

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant auquel vous pouvez vous connecter en tant qu'utilisateur avec les autorisations appropriées. Pour plus d'informations, consultez [Inscrivez-vous pour un Compte AWS](#) et [Création d'un utilisateur doté d'un accès administratif](#).

Le didacticiel part du principe que vous connaissez les rôles et les politiques IAM. Pour plus d'informations sur IAM, consultez le guide de [l'utilisateur IAM](#).

Vous devez configurer les AWS ressources suivantes pour suivre ce didacticiel :

- Utilisateur administrateur du lac de données

- Données sur la formation des lacs (paramètres des lacs)
- Version 3 du moteur Amazon Athena

Pour créer un administrateur de lac de données

1. Connectez-vous à la console Lake Formation en <https://console.aws.amazon.com/lakeformation/> tant qu'utilisateur administrateur. Vous allez créer des ressources dans la région USA Est (Virginie du Nord) pour ce didacticiel.
2. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Administrative roles and tasks.
3. Sélectionnez Choisir les administrateurs sous Administrateurs du lac de données.
4. Dans la fenêtre contextuelle, Gérer les administrateurs des lacs de données, sous Utilisateurs et rôles IAM, choisissez Utilisateur administrateur IAM.
5. Choisissez Enregistrer.

Pour activer les paramètres du lac de données

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>. Dans le volet de navigation, sous Catalogue de données, sélectionnez Paramètres. Décochez les cases suivantes :
 - Utilisez uniquement le contrôle d'accès IAM pour les nouvelles bases de données.
 - Utilisez uniquement le contrôle d'accès IAM pour les nouvelles tables dans les nouvelles bases de données.
2. Dans les paramètres des versions multi-comptes, choisissez la version 3 comme version multi-comptes.
3. Choisissez Enregistrer.

Pour mettre à niveau le moteur Amazon Athena vers la version 3

1. Ouvrez la console Athena à l'adresse. <https://console.aws.amazon.com/athena/>
2. Sélectionnez le groupe de travail, puis le groupe de travail principal.
3. Assurez-vous que le groupe de travail dispose d'une version minimale de 3. Si ce n'est pas le cas, modifiez le groupe de travail, choisissez Manual for Upgrade query engine, puis sélectionnez la version 3.

4. Sélectionnez Enregistrer les modifications.

Étape 1 : Approvisionnez vos ressources

Cette section explique comment configurer les AWS ressources à l'aide d'un CloudFormation modèle.

Pour créer vos ressources à l'aide d' CloudFormation un modèle

1. Connectez-vous à la AWS CloudFormation console à l'adresse <https://console.aws.amazon.com/cloudformation> en tant qu'administrateur IAM dans la région USA Est (Virginie du Nord).
2. Choisissez [Launch Stack](#).
3. Choisissez Next sur l'écran de création d'une pile.
4. Entrez un nom de pile.
5. Choisissez Suivant.
6. Sur la page suivante, choisissez Next.
7. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
8. Choisissez Créer.

La création de la pile peut prendre jusqu'à deux minutes.

Le lancement de la pile de formation dans le cloud crée les ressources suivantes :

- If-otf-datalake-123456789012 — Compartiment Amazon S3 pour stocker des données

Note

L'identifiant de compte ajouté au nom du compartiment Amazon S3 est remplacé par votre identifiant de compte.

- If-otf-tutorial-123456789012 — Compartiment Amazon S3 pour stocker les résultats des requêtes et les scripts de travail AWS Glue
- Ificebergdb — Base de données Iceberg AWS Glue
- Ifhudidb — Base de données Hudi AWS Glue
- Ifdeltadb — Base de données Delta AWS Glue
- native-iceberg-create — AWS Glue tâche qui crée une table Iceberg dans le catalogue de données

- `native-hudi-create` — AWS Glue tâche qui crée une table Hudi dans le catalogue de données
- `native-delta-create` — AWS Glue tâche qui crée une table Delta dans le catalogue de données
- `LF-OTF-GlueServiceRole` — Rôle IAM auquel vous passez pour AWS Glue exécuter les tâches. Ce rôle est associé aux politiques requises pour accéder aux ressources telles que le catalogue de données, le compartiment Amazon S3, etc.
- `LF-OTF-RegisterRole` — Rôle IAM pour enregistrer le site Amazon S3 auprès de Lake Formation. Ce rôle est `LF-Data-Lake-Storage-Policy` rattaché au rôle.
- `lf-consumer-analystuser` — Utilisateur IAM pour interroger les données à l'aide d'Athena
- `lf-consumer-analystuser-credentials` — Mot de passe de l'utilisateur de l'analyste de données enregistré dans AWS Secrets Manager

Une fois les créations de piles terminées, accédez à l'onglet de sortie et notez les valeurs pour :

- `AthenaQueryResultLocation` — Emplacement Amazon S3 pour la sortie de requête Athena
- `BusinessAnalystUserCredentials` — Mot de passe de l'utilisateur de l'analyste de données

Pour récupérer la valeur du mot de passe :

1. Choisissez la `lf-consumer-analystuser-credentials` valeur en accédant à la console Secrets Manager.
2. Dans la section `Secret value` (Valeur du secret), choisissez `Retrieve secret value` (Récupérer la valeur du secret).
3. Notez la valeur secrète du mot de passe.

Étape 2 : configurer les autorisations pour une table Iceberg

Dans cette section, vous allez apprendre à créer une table Iceberg dans Amazon Athena AWS Glue Data Catalog, à configurer des autorisations de données et à interroger des données à AWS Lake Formation l'aide d'Amazon Athena.

Pour créer une table Iceberg

Au cours de cette étape, vous allez exécuter une AWS Glue tâche qui crée une table transactionnelle Iceberg dans le catalogue de données.

1. Ouvrez la AWS Glue console <https://console.aws.amazon.com/glue/> dans la région USA Est (Virginie du Nord) en tant qu'utilisateur administrateur du lac de données.

2. Choisissez des tâches dans le volet de navigation de gauche.
3. Sélectionnez `native-iceberg-create`.

Create job [Info](#) Create

☒ **Visual with a source and target**
 Start with a source, ApplyMapping transform, and target.

☐ **Visual with a blank canvas**
 Author using an interactive visual interface.

☐ **Spark script editor**
 Write or upload your own Spark code.

☐ **Python Shell script editor**
 Write or upload your own Python shell script.

☐ **Jupyter Notebook**
 Write your own code in a Jupyter Notebook for interactive development.

☐ **Ray script editor** New
 Write your own code to run on Ray.

Source Amazon S3
 JSON, CSV, or Parquet files stored in S3.

Target Amazon S3
 S3 bucket by specifying a bucket path as the data target.

Your jobs (24) [Info](#)

Find jobs

	Job name	Type	Last modified	
☐	native-delta-create	Glue ETL	2/24/2023, 9:22:31 AM	
☒	native-iceberg-create	Glue ETL	2/24/2023, 9:22:31 AM	3.0
☐	native-hudi-create	Glue ETL	2/24/2023, 9:22:30 AM	3.0

Actions ▴ Run job

- Edit job
- Clone job
- Schedule job
- Delete job(s)
- Reset job bookmark

4. Sous Actions, sélectionnez Modifier la tâche.
5. Sous Détails du Job, développez les propriétés avancées et cochez la case à côté de Utiliser AWS Glue Data Catalog comme métastore Hive pour ajouter les métadonnées de la table dans le. AWS Glue Data Catalog Cela permet de AWS Glue Data Catalog définir le métastore pour les ressources du catalogue de données utilisées dans le cadre de la tâche et permet d'appliquer les autorisations de Lake Formation ultérieurement aux ressources du catalogue.
6. Choisissez Enregistrer.
7. Cliquez sur Exécuter. Vous pouvez consulter le statut de la tâche pendant son exécution.

Pour plus d'informations sur les AWS Glue tâches, consultez la section [Utilisation des tâches sur la AWS Glue console](#) dans le Guide du AWS Glue développeur.

Cette tâche crée une table Iceberg nommée `product` dans la `lflakebergdb` base de données. Vérifiez le tableau des produits dans la console Lake Formation.

Pour enregistrer l'emplacement des données auprès de Lake Formation

Enregistrez ensuite le chemin Amazon S3 comme emplacement de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/> en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, choisissez Emplacement des données.
3. Dans le coin supérieur droit de la console, choisissez Enregistrer l'emplacement.
4. Sur la page Enregistrer l'emplacement, entrez les informations suivantes :
 - Chemin Amazon S3 : choisissez Browse et sélectionnez `lf-otf-datalake-123456789012`. Cliquez sur la flèche droite (>) à côté de l'emplacement racine d'Amazon S3 pour accéder à cet `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg` emplacement.
 - Rôle IAM — Choisissez `LF-OTF-RegisterRole` comme rôle IAM.
 - Choisissez Enregistrer l'emplacement.

Register location

Amazon S3 location

Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path

Choose an Amazon S3 path for your data lake.

s3://lf-otf-datalake- /transactionaldata/native-iceberg

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

LF-OTF-GlueServiceRole ▼

☐ Enable Catalog Federation

Lake Formation will only assume a role to access a registered location when accessing a table under a federated database

Cancel

Register location

Pour plus d'informations sur l'enregistrement d'un emplacement de données auprès de Lake Formation, consultez [Ajouter un emplacement Amazon S3 à votre lac de données](#).

Pour accorder des autorisations à Lake Formation sur la table Iceberg

Au cours de cette étape, nous accorderons des autorisations de lac de données à l'utilisateur Business Analyst.

1. Sous Autorisations du lac de données, choisissez Grant.
2. Sur l'écran Accorder les autorisations relatives aux données, choisissez Utilisateurs et rôles IAM.
3. lf-consumer-analystuser Choisissez dans le menu déroulant.

Principals

☒ **IAM users and roles**
Users or roles from this AWS account.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☐ **External accounts**
AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add ▼

lf-consumer-analystuser X
User

4. Choisissez Ressource de catalogue de données nommée.
5. Pour les bases de données, choisissezlficebergdb.
6. Pour Tables, sélectionnezproduct.

LF-Tags or catalog resources

☐ Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ Named data catalog resources
Manager permissions for specific databases or tables, in addition to fine-grained data access.

Databases

Select one or more databases.

Choose databases ▼

Load more

lficebergdb X

Tables - optional

Select one or more tables.

Choose tables ▼

Load more

product X

Data filters - optional

Select one or more data filters.

Choose data filters ▼

Load more

Create new

[Manage data filters](#)

7. Vous pouvez ensuite accorder un accès basé sur les colonnes en spécifiant des colonnes.
 - a. Sous Autorisations relatives aux tables, choisissez Sélectionner.
 - b. Sous Autorisations relatives aux données, choisissez Accès basé sur les colonnes, puis Inclure les colonnes.
 - c. Choisissez product_nameprice, et category colonnes.
 - d. Choisissez Accorder.

Table permissions

Table permissions
Choose specific access permissions to grant.

☒ Select

☐ Insert

☐ Delete

☐ Describe

☐ Alter

☐ Drop

☐ Super
This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions
Choose the permission that may be granted to others.

☐ Select

☐ Insert

☐ Delete

☐ Describe

☐ Alter

☐ Drop

☐ Super
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Data permissions

☐ All data access
Grant access to all data without any restrictions.

☒ Column-based access
Grant data access to specific columns only.

Choose permission filter
Choose whether to include or exclude columns.

☒ Include columns
Grant permissions to access specific columns.

☐ Exclude columns
Grant permissions to access all but specific columns.

Select columns

Choose one or more columns ▼

product_name
string

price
bigint

category
string

CancelGrant

Pour interroger la table Iceberg à l'aide d'Athena

Vous pouvez maintenant commencer à interroger la table Iceberg que vous avez créée à l'aide d'Athena. Si c'est la première fois que vous exécutez des requêtes dans Athena, vous devez configurer l'emplacement des résultats des requêtes. Pour plus d'informations, consultez la section [Spécification de l'emplacement des résultats d'une requête](#).

1. Déconnectez-vous en tant qu'administrateur du lac de données et connectez-vous comme `lf-consumer-analystuser` dans la région USA Est (Virginie du Nord) en utilisant le mot de passe indiqué précédemment dans le CloudFormation résultat.
2. Ouvrez la console à l'adresse <https://console.aws.amazon.com/athena/>.
3. Choisissez Paramètres, puis sélectionnez Gérer.
4. Dans le champ Emplacement du résultat de la requête, entrez le chemin d'accès au compartiment que vous avez créé dans CloudFormation les sorties. Copiez la valeur de **AthenaQueryResultLocation** (`s3://lf-otf-tutorial-123456789012/athena-results/`) et choisissez Enregistrer.
5. Exécutez la requête suivante pour prévisualiser 10 enregistrements stockés dans la table Iceberg :

```
select * from lficebergdb.product limit 10;
```

Pour plus d'informations sur l'interrogation de tables Iceberg à l'aide d'Athena, consultez la section [Interrogation de tables Iceberg dans le guide de l'utilisateur d'Amazon Athena](#).

Étape 3 : configurer les autorisations pour une table Hudi

Dans cette section, vous allez apprendre à créer une table Hudi dans le AWS Glue Data Catalog, à configurer les autorisations de données et à interroger des données à AWS Lake Formation l'aide d'Amazon Athena.

Pour créer une table Hudi

Au cours de cette étape, vous allez exécuter une AWS Glue tâche qui crée une table transactionnelle Hudi dans le catalogue de données.

1. Connectez-vous à la AWS Glue console <https://console.aws.amazon.com/glue/> dans la région USA Est (Virginie du Nord)

en tant qu'utilisateur administrateur du lac de données.
2. Choisissez des tâches dans le volet de navigation de gauche.
3. Sélectionnez `native-hudi-create`.
4. Sous Actions, sélectionnez Modifier la tâche.

5. Sous Détails du Job, développez les propriétés avancées et cochez la case à côté de Utiliser AWS Glue Data Catalog comme métastore Hive pour ajouter les métadonnées de la table dans le. AWS Glue Data Catalog Cela permet de AWS Glue Data Catalog définir le métastore pour les ressources du catalogue de données utilisées dans le cadre de la tâche et permet d'appliquer les autorisations de Lake Formation ultérieurement aux ressources du catalogue.
6. Choisissez Enregistrer.
7. Cliquez sur Exécuter. Vous pouvez consulter le statut de la tâche pendant son exécution.

Pour plus d'informations sur les AWS Glue tâches, consultez la section [Utilisation des tâches sur la AWS Glue console](#) dans le Guide du AWS Glue développeur.

Cette tâche crée une table Hudi (cow) dans la base de données : lfhudidb. Vérifiez le product tableau dans la console Lake Formation.

Pour enregistrer l'emplacement des données auprès de Lake Formation

Enregistrez ensuite un chemin Amazon S3 comme emplacement racine de votre lac de données.

1. Connectez-vous à la console Lake Formation en <https://console.aws.amazon.com/lakeformation/> tant qu'utilisateur administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, choisissez Emplacement des données.
3. Dans le coin supérieur droit de la console, choisissez Enregistrer l'emplacement.
4. Sur la page Enregistrer l'emplacement, entrez les informations suivantes :
 - Chemin Amazon S3 : choisissez Browse et sélectionnez lf-otf-datalake-123456789012. Cliquez sur la flèche droite (>) à côté de l'emplacement racine d'Amazon S3 pour accéder à cet s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi emplacement.
 - Rôle IAM — Choisissez LF-OTF-RegisterRole comme rôle IAM.
 - Choisissez Enregistrer l'emplacement.

Pour accorder des autorisations de data lake sur la table Hudi

Au cours de cette étape, nous accorderons des autorisations de lac de données à l'utilisateur Business Analyst.

1. Sous Autorisations du lac de données, choisissez Grant.

2. Sur l'écran Accorder les autorisations relatives aux données, choisissez Utilisateurs et rôles IAM.
3. `lf-consumer-analystuser` depuis le menu déroulant.
4. Choisissez Ressource de catalogue de données nommée.
5. Pour les bases de données, choisissez `lfhudidb`.
6. Pour Tables, sélectionnez `product`.
7. Vous pouvez ensuite accorder un accès basé sur les colonnes en spécifiant des colonnes.
 - a. Sous Autorisations relatives aux tables, choisissez Sélectionner.
 - b. Sous Autorisations relatives aux données, choisissez Accès basé sur les colonnes, puis Inclure les colonnes.
 - c. Choisissez `product_nameprice`, et `category` colonnes.
 - d. Choisissez Accorder.

Pour interroger la table Hudi à l'aide d'Athena

Commencez maintenant à interroger la table Hudi que vous avez créée avec Athena. Si c'est la première fois que vous exécutez des requêtes dans Athena, vous devez configurer l'emplacement des résultats des requêtes. Pour plus d'informations, consultez la section [Spécification de l'emplacement des résultats d'une requête](#).

1. Déconnectez-vous en tant qu'administrateur du lac de données et connectez-vous comme `lf-consumer-analystuser` dans la région USA Est (Virginie du Nord) en utilisant le mot de passe indiqué précédemment dans le CloudFormation résultat.
2. Ouvrez la console à l'adresse <https://console.aws.amazon.com/athena/>.
3. Choisissez Paramètres, puis sélectionnez Gérer.
4. Dans le champ Emplacement du résultat de la requête, entrez le chemin d'accès au compartiment que vous avez créé dans CloudFormation les sorties. Copiez la valeur de **AthenaQueryResultLocation** (`s3://lf-otf-tutorial-123456789012/athena-results/`) et enregistrez.
5. Exécutez la requête suivante pour prévisualiser 10 enregistrements stockés dans la table Hudi :

```
select * from lfhudidb.product limit 10;
```

Pour plus d'informations sur l'interrogation des tables Hudi, consultez la section [Interrogation des tables Hudi](#) du guide de l'utilisateur Amazon Athena.

Étape 4 : Configuration des autorisations pour une table Delta Lake

Dans cette section, vous allez apprendre à créer une table Delta Lake contenant un fichier manifeste de liens symboliques AWS Glue Data Catalog, à configurer les autorisations de données AWS Lake Formation et à interroger des données à l'aide d'Amazon Athena.

Pour créer une table Delta Lake

Au cours de cette étape, vous allez exécuter une AWS Glue tâche qui crée une table transactionnelle Delta Lake dans le catalogue de données.

1. Connectez-vous à la AWS Glue console <https://console.aws.amazon.com/glue/> dans la région USA Est (Virginie du Nord)

en tant qu'utilisateur administrateur du lac de données.
2. Choisissez des tâches dans le volet de navigation de gauche.
3. Sélectionnez `native-delta-create`.
4. Sous Actions, sélectionnez Modifier la tâche.
5. Sous Détails du Job, développez les propriétés avancées et cochez la case à côté de Utiliser AWS Glue Data Catalog comme métastore Hive pour ajouter les métadonnées de la table dans le. AWS Glue Data Catalog Cela permet de AWS Glue Data Catalog définir le métastore pour les ressources du catalogue de données utilisées dans le cadre de la tâche et permet d'appliquer les autorisations de Lake Formation ultérieurement aux ressources du catalogue.
6. Choisissez Enregistrer.
7. Choisissez Exécuter sous Actions.

Cette tâche crée une table Delta Lake nommée `product` dans la `lfdeltadb` base de données. Vérifiez le `product` tableau dans la console Lake Formation.

Pour enregistrer l'emplacement des données auprès de Lake Formation

Enregistrez ensuite le chemin Amazon S3 comme emplacement racine de votre lac de données.

1. Ouvrez la console Lake Formation chez <https://console.aws.amazon.com/lakeformation/> l'utilisateur administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, choisissez Emplacement des données.
3. Dans le coin supérieur droit de la console, choisissez Enregistrer l'emplacement.

4. Sur la page Enregistrer l'emplacement, entrez les informations suivantes :
 - Chemin Amazon S3 : choisissez Browse et sélectionnez `lf-otf-datalake-123456789012`. Cliquez sur la flèche droite (>) à côté de l'emplacement racine d'Amazon S3 pour accéder à cet `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta` emplacement.
 - Rôle IAM — Choisissez `LF-OTF-RegisterRole` comme rôle IAM.
 - Choisissez Enregistrer l'emplacement.

Pour accorder des autorisations de data lake sur la table Delta Lake

Au cours de cette étape, nous accorderons des autorisations de lac de données à l'utilisateur Business Analyst.

1. Sous Autorisations du lac de données, choisissez Grant.
2. Sur l'écran Accorder les autorisations relatives aux données, choisissez Utilisateurs et rôles IAM.
3. `lf-consumer-analystuser` depuis le menu déroulant.
4. Choisissez Ressource de catalogue de données nommée.
5. Pour les bases de données, choisissez `lfdeltadb`.
6. Pour Tables, sélectionnez `product`.
7. Vous pouvez ensuite accorder un accès basé sur les colonnes en spécifiant des colonnes.
 - a. Sous Autorisations relatives aux tables, choisissez Sélectionner.
 - b. Sous Autorisations relatives aux données, choisissez Accès basé sur les colonnes, puis Inclure les colonnes.
 - c. Choisissez `product_nameprice`, et `category` colonnes.
 - d. Choisissez Accorder.

Pour interroger la table Delta Lake à l'aide d'Athena

Commencez maintenant à interroger la table Delta Lake que vous avez créée à l'aide d'Athena. Si c'est la première fois que vous exécutez des requêtes dans Athena, vous devez configurer l'emplacement des résultats des requêtes. Pour plus d'informations, consultez la section [Spécification de l'emplacement des résultats d'une requête](#).

1. Déconnectez-vous en tant qu'administrateur du lac de données et connectez-vous comme `BusinessAnalystUser` dans la région USA Est (Virginie du Nord) en utilisant le mot de passe indiqué précédemment dans le CloudFormation résultat.
2. Ouvrez la console à l'adresse <https://console.aws.amazon.com/athena/>.
3. Choisissez Paramètres, puis sélectionnez Gérer.
4. Dans le champ Emplacement du résultat de la requête, entrez le chemin d'accès au compartiment que vous avez créé dans CloudFormation les sorties. Copiez la valeur de **AthenaQueryResultLocation** (`s3 ://lf-otf-tutorial-123456789012/athena-results/`) et enregistrez.
5. Exécutez la requête suivante pour prévisualiser 10 enregistrements stockés dans la table Delta Lake :

```
select * from lfdeltadb.product limit 10;
```

Pour plus d'informations sur l'interrogation des tables Delta Lake, consultez la section [Interrogation des tables Delta Lake du](#) guide de l'utilisateur Amazon Athena.

Étape 5 : Nettoyer les AWS ressources

Pour nettoyer des ressources

Pour éviter que des frais supplémentaires ne vous soient facturés Compte AWS, supprimez les AWS ressources que vous avez utilisées pour ce didacticiel.

1. Connectez-vous à la CloudFormation console sur <https://console.aws.amazon.com/cloudformation> en tant qu'administrateur IAM.
2. [Supprimez la pile de formation des nuages](#). Les tables que vous avez créées sont automatiquement supprimées avec la pile.

Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation

Des milliers de clients construisent des lacs de données à l'échelle du pétaoctet. AWS Nombre de ces clients ont l'habitude AWS Lake Formation de créer et de partager facilement leurs lacs de données au sein de l'entreprise. À mesure que le nombre de tables et d'utilisateurs augmente, les

gestionnaires de données et les administrateurs cherchent des moyens de gérer facilement et à grande échelle les autorisations sur les lacs de données. Le contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) résout ce problème en permettant aux gestionnaires de données de créer des balises LF (en fonction de leur classification et de leur ontologie de données) qui peuvent ensuite être attachées aux ressources.

Le LF-TBAC est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans Lake Formation, ces attributs sont appelés balises LF. Vous pouvez associer des balises LF aux ressources du catalogue de données et aux principes de Lake Formation. Les administrateurs des lacs de données peuvent attribuer et révoquer des autorisations sur les ressources de Lake Formation à l'aide de balises LF. Pour plus d'informations sur, voir [Contrôle d'accès basé sur des balises Lake Formation](#).

Ce didacticiel explique comment créer une politique de contrôle d'accès basée sur des balises Lake Formation à l'aide d'un ensemble de données AWS public. En outre, il montre comment interroger des tables, des bases de données et des colonnes associées à des politiques d'accès basées sur des balises Lake Formation.

Vous pouvez utiliser le LF-TBAC dans les cas d'utilisation suivants :

- L'administrateur du lac de données doit accorder l'accès à un grand nombre de tables et de principes.
- Vous souhaitez classer vos données en fonction d'une ontologie et accorder des autorisations en fonction de la classification
- L'administrateur du lac de données souhaite attribuer des autorisations de manière dynamique, d'une manière peu couplée

Voici les étapes de haut niveau pour configurer les autorisations à l'aide de LF-TBAC :

1. Le gestionnaire de données définit l'ontologie des balises à l'aide de deux balises LF : et. `Confidential` Les données sont soumises à des `Confidential=True` contrôles d'accès plus stricts. Les données `Sensitive=True` nécessitent une analyse spécifique de la part de l'analyste.
2. Le data steward attribue différents niveaux d'autorisation à l'ingénieur de données pour créer des tables avec différentes balises LF.
3. L'ingénieur de données crée deux bases de données : `tag_database` et `col_tag_database`. Toutes les tables incluses `tag_database` sont configurées avec `Confidential=True`. Toutes

les tables du `col_tag_database` sont configurées avec `Confidential=False`. Certaines colonnes du tableau `col_tag_database` sont balisées `Sensitive=True` pour répondre à des besoins d'analyse spécifiques.

4. L'ingénieur de données accorde l'autorisation de lecture à l'analyste pour les tables présentant une condition d'expression spécifique `Confidential=True` et `Confidential=False,Sensitive=True`.
5. Grâce à cette configuration, l'analyste de données peut se concentrer sur l'analyse avec les bonnes données.

Rubriques

- [Public visé](#)
- [Conditions préalables](#)
- [Étape 1 : Approvisionnez vos ressources](#)
- [Étape 2 : enregistrez l'emplacement de vos données, créez une ontologie LF-Tag et accordez des autorisations](#)
- [Étape 3 : Création de bases de données Lake Formation](#)
- [Étape 4 : Accorder les autorisations relatives aux tables](#)
- [Étape 5 : exécuter une requête dans Amazon Athena pour vérifier les autorisations](#)
- [Étape 6 : Nettoyer les AWS ressources](#)

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Lorsqu'il s'agit de gérer AWS Glue Data Catalog et d'administrer les autorisations dans Lake Formation, les responsables des données des comptes producteurs ont une propriété fonctionnelle basée sur les fonctions qu'ils prennent en charge et peuvent accorder l'accès à divers consommateurs, organisations externes et comptes.

Le tableau suivant répertorie les rôles utilisés dans ce didacticiel :

Role	Description
Data Steward (administrateur)	L' <code>lf-data-steward</code> utilisateur dispose des droits d'accès suivants :

Role	Description
	• Accès en lecture à toutes les ressources du catalogue de données • Peut créer des balises LF et les associer au rôle d'ingénieur de données pour obtenir des autorisations pouvant être accordées à d'autres principaux
Ingénieur de données	<code>lf-data-engineer</code> l'utilisateur dispose des droits d'accès suivants : • Accès complet en lecture, écriture et mise à jour à toutes les ressources du catalogue de données • Autorisations de localisation des données dans le lac de données • Peut associer des balises LF et les associer au catalogue de données • Peut attacher des balises LF aux ressources, ce qui permet d'accéder aux principaux en fonction des politiques créées par les gestionnaires de données
Analyste des données	<code>lf-data-analyst</code> utilisateur dispose des droits d'accès suivants : • Accès détaillé aux ressources partagées par les politiques d'accès basées sur les balises de Lake Formation

Conditions préalables

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour de plus amples informations, veuillez consulter [Exécution des tâches AWS de configuration initiale](#).

Le didacticiel part du principe que vous êtes familiarisé avec IAM. Pour plus d'informations sur IAM, consultez le guide de l'[utilisateur IAM](#).

Étape 1 : Approvisionnez vos ressources

Ce didacticiel inclut un AWS CloudFormation modèle pour une configuration rapide. Vous pouvez le consulter et le personnaliser en fonction de vos besoins. Le modèle crée trois rôles différents (répertoriés dans [Public visé](#)) pour effectuer cet exercice et copie le nyc-taxi-data jeu de données dans votre compartiment Amazon S3 local.

- Un compartiment Amazon S3
- Les paramètres appropriés de la Lake Formation
- Les ressources Amazon EC2 appropriées
- Trois rôles IAM avec informations d'identification

Créez vos ressources

1. Connectez-vous à la AWS CloudFormation console à l'adresse <https://console.aws.amazon.com/cloudformation> dans la région USA Est (Virginie du Nord).
2. Choisissez [Launch Stack](#).
3. Choisissez Suivant.
4. Dans la section Configuration utilisateur, entrez le mot de passe pour trois rôles :DataStewardUserPassword, DataEngineerUserPassword etDataAnalystUserPassword.
5. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
6. Choisissez Créer.

La création de la pile peut prendre jusqu'à cinq minutes.

Note

Une fois le didacticiel terminé, vous souhaitez peut-être supprimer le stack afin d'éviter CloudFormation de continuer à encourir des frais. Vérifiez que les ressources sont correctement supprimées dans le statut de l'événement pour la pile.

Étape 2 : enregistrez l'emplacement de vos données, créez une ontologie LF-Tag et accordez des autorisations

Au cours de cette étape, l'utilisateur du data steward définit l'ontologie des balises à l'aide de deux balises LF : Confidential et Sensitive donne à des principes IAM spécifiques la possibilité d'associer des balises LF nouvellement créées aux ressources.

Enregistrer un emplacement de données et définir l'ontologie LF-Tag

1. Effectuez la première étape en tant qu'utilisateur responsable de la gestion des données (lf-data-steward) pour vérifier les données dans Amazon S3 et le catalogue de données dans Lake Formation.
 - a. Connectez-vous à la console Lake Formation en utilisant le mot lf-data-steward de passe utilisé lors du déploiement de la CloudFormation pile. <https://console.aws.amazon.com/lakeformation/>
 - b. Dans le volet de navigation, sous Autorisations, choisissez Rôles et tâches administratifs.
 - c. Choisissez Ajouter dans la section Administrateurs du lac de données.
 - d. Sur la page Ajouter un administrateur, pour les utilisateurs et les rôles IAM, choisissez l'utilisateur lf-data-steward.
 - e. Choisissez Enregistrer pour l'ajouter lf-data-steward en tant qu'administrateur de Lake Formation.
2. Ensuite, mettez à jour les paramètres du catalogue de données pour utiliser l'autorisation Lake Formation pour contrôler les ressources du catalogue au lieu du contrôle d'accès basé sur IAM.
 - a. Dans le volet de navigation, sous Administration, sélectionnez Paramètres du catalogue de données.
 - b. Décochez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données.
 - c. Décochez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables dans les nouvelles bases de données.
 - d. Cliquez sur Sauvegarder
3. Enregistrez ensuite l'emplacement des données pour le lac de données.
 - a. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
 - b. Choisissez Enregistrer l'emplacement.

- c. Sur la page Enregistrer l'emplacement, pour le chemin Amazon S3, entrez `s3://lf-tagbased-demo-Account-ID`.
 - d. Pour le rôle IAM, laissez la valeur par défaut `AWSServiceRoleForLakeFormationDataAccess` telle quelle.
 - e. Choisissez Lake Formation comme mode d'autorisation.
 - f. Choisissez Enregistrer l'emplacement.
4. Créez ensuite l'ontologie en définissant une balise LF.
- a. Sous Autorisations dans le volet de navigation, choisissez LF-Tags and permissions. .
 - b. Choisissez Ajouter un tag LF.
 - c. Pour Key (Clé), saisissez `Confidential`.
 - d. Dans le champ Valeurs, ajoutez `True` et `False`.
 - e. Choisissez Ajouter un tag LF.
 - f. Répétez les étapes pour créer le tag LF `Sensitive` avec la valeur. `True`

Vous avez créé toutes les balises LF nécessaires pour cet exercice.

Accorder des autorisations aux utilisateurs IAM

1. Donnez ensuite à des principes IAM spécifiques la possibilité d'associer des balises LF nouvellement créées aux ressources.
 - a. Sous Autorisations dans le volet de navigation, choisissez LF-Tags and permissions.
 - b. Dans la section Autorisations LF-Tag, choisissez Accorder des autorisations.
 - c. Pour le type d'autorisation, choisissez les autorisations de paire clé-valeur LF-Tag.
 - d. Sélectionnez les utilisateurs et les rôles IAM.
 - e. Pour les utilisateurs et les rôles IAM, recherchez et choisissez le `lf-data-engineer` rôle.
 - f. Dans la section LF-Tags, ajoutez la clé `Confidential` avec les valeurs `True` et `False`, et la clé `key Sensitive` avec valeur. `True`
 - g. Sous Autorisations, sélectionnez Décrire et associer pour les autorisations et les autorisations pouvant être accordées.
 - h. Choisissez Accorder.

2. Ensuite, accordez l'autorisation `lf-data-engineer` de créer des bases de données dans notre catalogue de données et dans le compartiment Amazon S3 sous-jacent créé par AWS CloudFormation.
 - a. Sous Administration dans le volet de navigation, sélectionnez Rôles et tâches d'administration.
 - b. Dans la section Créateurs de bases de données, choisissez Grant.
 - c. Pour les utilisateurs et les rôles IAM, choisissez le `lf-data-engineer` rôle.
 - d. Pour les autorisations du catalogue, sélectionnez Créer une base de données.
 - e. Choisissez Accorder.
3. Ensuite, accordez des autorisations sur le compartiment Amazon S3 (`s3://lf-tagbased-demo-Account-ID`) à l'`lf-data-engineer` utilisateur.
 - a. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données.
 - b. Choisissez Accorder.
 - c. Sélectionnez Mon compte.
 - d. Pour les utilisateurs et les rôles IAM, choisissez le `lf-data-engineer` rôle.
 - e. Pour les emplacements de stockage, entrez le compartiment Amazon S3 créé par le CloudFormation modèle(`s3://lf-tagbased-demo-Account-ID`).
 - f. Choisissez Accorder.
4. Ensuite, accordez des autorisations `lf-data-engineer` pouvant être accordées sur les ressources associées à l'expression `LF-Tag. Confidential=True`
 - a. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.
 - b. Choisissez Accorder.
 - c. Sélectionnez les utilisateurs et les rôles IAM.
 - d. Choisissez le rôle `lf-data-engineer`.
 - e. Dans la section Balises LF ou ressources du catalogue, sélectionnez Ressources associées aux balises LF.
 - f. Choisissez Ajouter une paire clé-valeur LF-Tag.
 - g. Ajoutez la clé `Confidential` avec les valeurs `True`.
 - h. Dans la section Autorisations de base de données, sélectionnez Décrire pour les

- i. Dans la section Autorisations relatives aux tables, sélectionnez Décrire, sélectionner et modifier pour les autorisations relatives aux tables et aux autorisations pouvant être accordées.
 - j. Choisissez Accorder.
5. Ensuite, accordez des autorisations `lf-data-engineer` pouvant être accordées sur les ressources associées à l'expression LF-Tag. `Confidential=False`
 - a. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.
 - b. Choisissez Accorder.
 - c. Sélectionnez les utilisateurs et les rôles IAM.
 - d. Choisissez le rôle `lf-data-engineer`.
 - e. Sélectionnez les ressources associées aux balises LF.
 - f. Choisissez Ajouter un tag LF.
 - g. Ajoutez la clé `Confidential` avec la valeur `False`.
 - h. Dans la section Autorisations de base de données, sélectionnez Décrire pour les autorisations de base de données et les autorisations pouvant être accordées.
 - i. Dans la section Autorisations relatives aux tables et aux colonnes, ne sélectionnez rien.
 - j. Choisissez Accorder.
6. Ensuite, nous accordons des autorisations `lf-data-engineer` pouvant être accordées sur les ressources associées aux paires clé-valeur LF-Tag et. `Confidential=False` `Sensitive=True`
 - a. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données.
 - b. Choisissez Accorder.
 - c. Sélectionnez les utilisateurs et les rôles IAM.
 - d. Choisissez le rôle `lf-data-engineer`.
 - e. Dans la section Balises LF ou ressources du catalogue, sélectionnez Ressources associées aux balises LF.
 - f. Choisissez Ajouter un tag LF.
 - g. Ajoutez la clé `Confidential` avec la valeur `False`.
 - h. Choisissez Ajouter une paire clé-valeur LF-Tag

- i. Ajoutez la clé `Sensitive` avec la valeur `True`.
- j. Dans la section Autorisations de base de données, sélectionnez `Décrire` pour les autorisations de base de données et les autorisations pouvant être accordées.
- k. Dans la section Autorisations relatives aux tables, sélectionnez `Décrire`, sélectionner et modifier pour les autorisations relatives aux tables et aux autorisations pouvant être accordées.
- l. Choisissez `Accorder`.

Étape 3 : Création de bases de données Lake Formation

Au cours de cette étape, vous créez deux bases de données et attachez des balises LF aux bases de données et à des colonnes spécifiques à des fins de test.

Créez vos bases de données et votre table pour un accès au niveau de la base de données

1. Créez d'abord la base de données `tag_database`, la table `source_data` et attachez les balises LF appropriées.
 - a. Sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), sous Catalogue de données, sélectionnez `Databases`.
 - b. Choisissez `Créer une base de données`.
 - c. Pour `Nom`, saisissez `tag_database`.
 - d. Dans `Emplacement`, entrez l'emplacement Amazon S3 créé par le CloudFormation modèle (`s3://lf-tagbased-demo-Account-ID/tag_database/`).
 - e. Désélectionnez `Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données`.
 - f. Choisissez `Créer une base de données`.
2. Ensuite, créez une nouvelle table à l'intérieur `tag_database`.
 - a. Sur la page `Bases de données`, sélectionnez la base de données `tag_database`.
 - b. Choisissez `Afficher les tables`, puis cliquez sur `Créer une table`.
 - c. Pour `Nom`, saisissez `source_data`.
 - d. Pour `Database (Base de données)`, choisissez la base de données `tag_database`.
 - e. Pour `Format de tableau`, choisissez `AWS Glue Tableau standard`.
 - f. Si les données se trouvent dans, sélectionnez le chemin spécifié dans mon compte.

- g. Pour Inclure le chemin, entrez le chemin tag_database créé par le CloudFormation modèle(s3://lf-tagbased-demo~~Account-ID~~/tag_database/).
- h. Pour le format des données, sélectionnez CSV.
- i. Sous Charger le schéma, entrez le tableau JSON suivant de structure de colonne pour créer un schéma :

```
[
  {
    "Name": "vendorid",
    "Type": "string"
  },
  {
    "Name": "lpep_pickup_datetime",
    "Type": "string"
  },
  {
    "Name": "lpep_dropoff_datetime",
    "Type": "string"
  },
  {
    "Name": "store_and_fwd_flag",
    "Type": "string"
  },
  {
    "Name": "ratecodeid",
    "Type": "string"
  },
  {
    "Name": "pulocationid",
    "Type": "string"
  },
  {
    "Name": "dolocationid",
    "Type": "string"
  },
  {
    "Name": "passenger_count",
    "Type": "string"
  }
]
```



```
    },  
    {  
      "Name": "trip_distance",  
      "Type": "string"  
    },  
    {  
      "Name": "fare_amount",  
      "Type": "string"  
    },  
    {  
      "Name": "extra",  
      "Type": "string"  
    },  
    {  
      "Name": "mta_tax",  
      "Type": "string"  
    },  
    {  
      "Name": "tip_amount",  
      "Type": "string"  
    },  
    {  
      "Name": "tolls_amount",  
      "Type": "string"  
    },  
    {  
      "Name": "ehail_fee",  
      "Type": "string"  
    },  
    {  
      "Name": "improvement_surcharge",  
      "Type": "string"  
    },  
    {  
      "Name": "total_amount",
```

```
        "Type": "string"
      },
      {
        "Name": "payment_type",
        "Type": "string"
      }
    ]
```

- j. Choisissez Charger. Après avoir chargé le schéma, le schéma de table doit ressembler à la capture d'écran suivante :

#	Column Name	▼	Data type
1	vendorid		string
2	lpep_pickup_datetime		string
3	lpep_dropoff_datetime		string
4	store_and_fwd_flag		string
5	ratecodeid		string
6	pulocationid		string
7	dolocationid		string
8	passenger_count		string
9	trip_distance		string
10	fare_amount		string
11	extra		string
12	mta_tax		string
13	tip_amount		string
14	tolls_amount		string
15	ehail_fee		string
16	improvement_surcharge		string
17	total_amount		string
18	payment_type		string

- k. Sélectionnez Soumettre.
3. Ensuite, attachez des balises LF au niveau de la base de données.
 - a. Sur la page Bases de données, recherchez et sélectionnez `tag_database`.
 - b. Dans le menu Actions, choisissez Modifier les balises LF.
 - c. Choisissez Attribuer un nouveau tag LF.
 - d. Pour les clés assignées, choisissez le `Confidential` tag LF que vous avez créé précédemment.
 - e. Dans le champ Valeurs, sélectionnez `True`.
 - f. Choisissez Enregistrer.

Ceci termine l'attribution du tag LF à la base de données `tag_database`.

Créez votre base de données et votre table pour un accès au niveau des colonnes

Répétez les étapes suivantes pour créer la base de données `col_tag_database` et la table `source_data_col_lvl1`, et attachez des balises LF au niveau de la colonne.

1. Sur la page Bases de données, sélectionnez Créer une base de données.
2. Pour Nom, saisissez `col_tag_database`.
3. Dans Emplacement, entrez l'emplacement Amazon S3 créé par le CloudFormation modèle(`s3://lf-tagbased-demo-Account-ID/col_tag_database/`).
4. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données.
5. Choisissez Créer une base de données.
6. Sur la page Bases de données, sélectionnez votre nouvelle base de données(`col_tag_database`).
7. Choisissez Afficher les tables, puis cliquez sur Créer une table.
8. Pour Nom, saisissez `source_data_col_lvl1`.
9. Dans Base de données, choisissez votre nouvelle base de données(`col_tag_database`).
10. Pour Format de tableau, choisissez AWS Glue Tableau standard.
11. Si les données se trouvent dans, sélectionnez le chemin spécifié dans mon compte.

12. Entrez le chemin Amazon S3 pour col_tag_database(s3://lf-tagbased-demo-*Account-ID*/col_tag_database/).
13. Pour Format des données, sélectionnez CSV.
14. Sous Upload schema, entrez le schéma JSON suivant :

```
[
    {
        "Name": "vendorid",
        "Type": "string"
    },
    {
        "Name": "lpep_pickup_datetime",
        "Type": "string"
    },
    {
        "Name": "lpep_dropoff_datetime",
        "Type": "string"
    },
    {
        "Name": "store_and_fwd_flag",
        "Type": "string"
    },
    {
        "Name": "ratecodeid",
        "Type": "string"
    },
    {
        "Name": "pulocationid",
        "Type": "string"
    },

```

```
{
    "Name": "dolocationid",
    "Type": "string"
},
{
    "Name": "passenger_count",
    "Type": "string"
},
{
    "Name": "trip_distance",
    "Type": "string"
},
{
    "Name": "fare_amount",
    "Type": "string"
},
{
    "Name": "extra",
    "Type": "string"
},
{
    "Name": "mta_tax",
    "Type": "string"
},
{
    "Name": "tip_amount",
    "Type": "string"
},
{
    "Name": "tolls_amount",
```

```
        "Type": "string"

    },
    {
        "Name": "ehail_fee",
        "Type": "string"

    },
    {
        "Name": "improvement_surcharge",
        "Type": "string"

    },
    {
        "Name": "total_amount",
        "Type": "string"

    },
    {
        "Name": "payment_type",
        "Type": "string"

    }

]
```

15. Sélectionnez **Upload**. Après avoir chargé le schéma, le schéma de table doit ressembler à la capture d'écran suivante.

#	Column Name	▼	Data type
1	vendorid		string
2	lpep_pickup_datetime		string
3	lpep_dropoff_datetime		string
4	store_and_fwd_flag		string
5	ratecodeid		string
6	pulocationid		string
7	dolocationid		string
8	passenger_count		string
9	trip_distance		string
10	fare_amount		string
11	extra		string
12	mta_tax		string
13	tip_amount		string
14	tolls_amount		string
15	ehail_fee		string
16	improvement_surcharge		string
17	total_amount		string
18	payment_type		string

16. Choisissez Soumettre pour terminer la création de la table.
17. Associez maintenant le Sensitive=True tag LF aux colonnes vendorid et. fare_amount
 - a. Sur la page Tables, sélectionnez la table que vous avez créée(source_data_col_lvl1).
 - b. Dans le menu Actions, sélectionnez Schéma.
 - c. Sélectionnez la colonne vendorid et choisissez Modifier les balises LF.
 - d. Pour Clés attribuées, choisissez Sensitive.
 - e. Dans le champ Valeurs, sélectionnez Vrai.
 - f. Choisissez Enregistrer.
18. Ensuite, associez le Confidential=False tag LF à. col_tag_database Cela est nécessaire pour lf-data-analyst pouvoir décrire la base de données col_tag_database lorsque vous êtes connecté depuis Amazon Athena.
 - a. Sur la page Bases de données, recherchez et sélectionnez col_tag_database.
 - b. Dans le menu Actions, choisissez Modifier les balises LF.
 - c. Choisissez Attribuer un nouveau tag LF.
 - d. Pour les clés attribuées, choisissez le Confidential tag LF que vous avez créé précédemment.
 - e. Dans le champ Valeurs, sélectionnez False.
 - f. Choisissez Enregistrer.

Étape 4 : Accorder les autorisations relatives aux tables

Accordez des autorisations aux analystes de données pour l'utilisation des bases de données tag_database et de la table à l'col_tag_database aide des balises LF Confidential et Sensitive

1. Procédez comme suit pour accorder à l'lf-data-analyst utilisateur des autorisations sur les objets associés à la balise LF Confidential=True (Database:TAG_DATABASE) afin qu'il dispose de la base de données et des autorisations sur les Describe tables. Select
 - a. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/aslf-data-engineer>.
 - b. Sous Autorisations, sélectionnez Autorisations du lac de données.
 - c. Choisissez Accorder.

- d. Sous Principaux, sélectionnez Utilisateurs et rôles IAM.
 - e. Pour les utilisateurs et les rôles IAM, choisissez `lf-data-analyst`.
 - f. Sous Balises LF ou ressources du catalogue, sélectionnez Ressources associées aux balises LF.
 - g. Choisissez Ajouter un tag LF.
 - h. Pour Key, choisissez `Confidential`.
 - i. Dans le champ Valeurs, sélectionnez `True`.
 - j. Pour les autorisations de base de données, sélectionnez `Describe`.
 - k. Pour les autorisations relatives aux tables, choisissez Sélectionner et décrire.
 - l. Choisissez Accorder.
2. Répétez ensuite les étapes pour accorder des autorisations aux analystes de données pour l'expression LF-Tag for. `Confidential=False` Cette balise LF est utilisée pour décrire le `col_tag_database` et le tableau `source_data_col_lvl1` lorsque vous êtes connecté `lf-data-analyst` depuis Amazon Athena.
- a. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/aslf-data-engineer>.
 - b. Sur la page Bases de données, sélectionnez la base de données `col_tag_database`.
 - c. Choisissez Action et Grant.
 - d. Sous Principaux, sélectionnez Utilisateurs et rôles IAM.
 - e. Pour les utilisateurs et les rôles IAM, choisissez `lf-data-analyst`.
 - f. Sélectionnez les ressources associées aux balises LF.
 - g. Choisissez Ajouter un tag LF.
 - h. Pour Key, choisissez `Confidential`.
 - i. Pour Valeurs, choisissez `False`.
 - j. Pour les autorisations de base de données, sélectionnez `Describe`.
 - k. Pour les autorisations relatives aux tables, ne sélectionnez rien.
 - l. Choisissez Accorder.
3. Répétez ensuite les étapes pour accorder des autorisations aux analystes de données pour l'expression LF-Tag pour `Confidential=False` et `Sensitive=True` Cette balise LF est utilisée pour décrire le `col_tag_database` et le tableau `source_data_col_lvl1` (au niveau des colonnes) lorsque vous êtes connecté depuis Amazon `lf-data-analyst` Athena.

- a. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/aslf-data-engineer>.
- b. Sur la page Bases de données, sélectionnez la base de données `col_tag_database`.
- c. Choisissez Action et Grant.
- d. Sous Principaux, sélectionnez Utilisateurs et rôles IAM.
- e. Pour les utilisateurs et les rôles IAM, choisissez `lf-data-analyst`.
- f. Sélectionnez les ressources associées aux balises LF.
- g. Choisissez Ajouter un tag LF.
- h. Pour Key, choisissez `Confidential`.
- i. Pour Valeurs, choisissez `False`.
- j. Choisissez Ajouter un tag LF.
- k. Pour Key, choisissez `Sensitive`.
- l. Pour Valeurs, choisissez `True`.
- m. Pour les autorisations de base de données, sélectionnez `Describe`.
- n. Pour les autorisations relatives aux tables, sélectionnez `Select` et `Describe`.
- o. Choisissez Accorder.

Étape 5 : exécuter une requête dans Amazon Athena pour vérifier les autorisations

Pour cette étape, utilisez Amazon Athena pour exécuter des SELECT requêtes sur les deux tables. (`source_data` and `source_data_col_lvl1`) Utilisez le chemin Amazon S3 comme emplacement des résultats de la requête (`s3://lf-tagbased-demo-Account-ID/athena-results/`).

1. Connectez-vous à la console Athena à <https://console.aws.amazon.com/athena/> l'adresse `as.lf-data-analyst`
2. Dans l'éditeur de requêtes Athena, choisissez `tag_database` dans le panneau de gauche.
3. Choisissez l'icône d'options de menu supplémentaires (trois points verticaux) à côté `source_data` et choisissez Aperçu du tableau.
4. Choisissez Exécuter la requête.

L'exécution de la requête devrait prendre quelques minutes. La requête affiche toutes les colonnes de la sortie car la balise LF est associée au niveau de la base de données et la source_data table en a automatiquement hérité. LF-tag tag_database

5. Exécutez une autre requête à l'aide col_tag_database de etsource_data_col_lvl1.

La deuxième requête renvoie les deux colonnes étiquetées Non-Confidential et Sensitive.

6. Vous pouvez également vérifier le comportement de la politique d'accès basée sur les balises Lake Formation sur les colonnes pour lesquelles vous n'avez pas de subventions politiques. Lorsqu'une colonne non balisée est sélectionnée dans le tableau source_data_col_lvl1, Athena renvoie une erreur. Par exemple, vous pouvez exécuter la requête suivante pour sélectionner des colonnes geolocationid non balisées :

```
SELECT geolocationid FROM "col_tag_database"."source_data_col_lvl1" limit 10;
```

Étape 6 : Nettoyer les AWS ressources

Pour éviter des frais indésirables Compte AWS, vous pouvez supprimer les AWS ressources que vous avez utilisées pour ce didacticiel.

1. Connectez-vous à la console Lake Formation en tant que lf-data-engineer et supprimez les bases de données tag_database et col_tag_database.
2. Ensuite, connectez-vous en tant que lf-data-steward et nettoyez toutes les autorisations LF-Tag, les autorisations de données et les autorisations de localisation des données accordées ci-dessus qui ont été accordées lf-data-engineer et lf-data-analyst.
3. Connectez-vous à la console Amazon S3 en tant que propriétaire du compte à l'aide des informations d'identification IAM que vous avez utilisées pour déployer la CloudFormation pile.
4. Supprimez les compartiments suivants :
 - lf-tagbased-demo-accesslogs-*acct-id*
 - lf-tagbased-demo-*acct-id*
5. Connectez-vous à CloudFormation la console sur <https://console.aws.amazon.com/cloudformation> et supprimez la pile que vous avez créée. Attendez que le statut de la pile passe à DELETE_COMPLETE.

Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes

AWS Lake Formation les autorisations au niveau des lignes vous permettent d'accéder à des lignes spécifiques d'un tableau en fonction des politiques de conformité et de gouvernance des données. Si vous avez de grandes tables stockant des milliards d'enregistrements, vous devez trouver un moyen de permettre aux différents utilisateurs et équipes d'accéder uniquement aux données qu'ils sont autorisés à consulter. Le contrôle d'accès au niveau des lignes est un moyen simple et performant de protéger les données, tout en permettant aux utilisateurs d'accéder aux données dont ils ont besoin pour effectuer leur travail. Lake Formation fournit des audits et des rapports de conformité centralisés en identifiant quels principaux ont accédé à quelles données, quand et par le biais de quels services.

Dans ce didacticiel, vous apprendrez comment fonctionnent les contrôles d'accès au niveau des lignes dans Lake Formation et comment les configurer.

Ce didacticiel inclut un AWS CloudFormation modèle permettant de configurer rapidement les ressources requises. Vous pouvez le consulter et le personnaliser en fonction de vos besoins.

Rubriques

- [Public visé](#)
- [Prérequis](#)
- [Étape 1 : Approvisionnez vos ressources](#)
- [Étape 2 : Requête sans filtres de données](#)
- [Étape 3 : configurer les filtres de données et accorder des autorisations](#)
- [Étape 4 : Requête à l'aide de filtres de données](#)
- [Étape 5 : Nettoyer les AWS ressources](#)

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Le tableau suivant répertorie les rôles et les responsabilités du propriétaire et du consommateur de données.

Rôle	Description
Administrateur IAM	Utilisateur capable de créer des utilisateurs, des rôles et des compartiments Amazon Simple Storage Service (Amazon S3). Possède la politique AdministratorAccess AWS gérée.
Administrateur du lac de données	Utilisateur chargé de configurer le lac de données, de créer des filtres de données et d'accorder des autorisations aux analystes de données.
Analyste des données	Un utilisateur qui peut exécuter des requêtes sur le lac de données. Les analystes de données résidant dans différents pays (pour notre cas d'utilisation, les États-Unis et le Japon) ne peuvent analyser que les avis des clients situés dans leur propre pays et, pour des raisons de conformité, ne devraient pas être en mesure de consulter les données clients situées dans d'autres pays.

Prérequis

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour de plus amples informations, veuillez consulter [Exécution des tâches AWS de configuration initiale](#).

Le didacticiel part du principe que vous connaissez IAM. Pour plus d'informations sur IAM, consultez le guide de l'[utilisateur IAM](#).

Modifier les paramètres de Lake Formation

Important

Avant de lancer le CloudFormation modèle, désactivez l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données/tables de Lake Formation en suivant les étapes ci-dessous :

1. Connectez-vous à la console Lake Formation <https://console.aws.amazon.com/lakeformation/> dans la région USA Est (Virginie du Nord) ou USA Ouest (Oregon).
2. Sous Catalogue de données, sélectionnez Paramètres.
3. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données et Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables des nouvelles bases de données.
4. Choisissez Save (Enregistrer).

Étape 1 : Approvisionnez vos ressources

Ce didacticiel inclut un CloudFormation modèle pour une configuration rapide. Vous pouvez le consulter et le personnaliser en fonction de vos besoins. Le CloudFormation modèle génère les ressources suivantes :

- Utilisateurs et politiques pour :
 - DataLakeAdmin
 - DataAnalystÉtats-Unis
 - DataAnalystJP
- Lake Formation : paramètres et autorisations du lac
- Fonction Lambda (pour les ressources CloudFormation personnalisées soutenues par Lambda) utilisée pour copier des exemples de fichiers de données du compartiment Amazon S3 public vers votre compartiment Amazon S3
- Un bucket Amazon S3 qui servira de lac de données
- Une AWS Glue Data Catalog base de données, une table et une partition

Créez vos ressources

Suivez ces étapes pour créer vos ressources à l'aide du CloudFormation modèle.

1. Connectez-vous à la CloudFormation console à l'adresse <https://console.aws.amazon.com/cloudformation> dans la région USA Est (Virginie du Nord).
2. Choisissez [Launch Stack](#).
3. Choisissez Next sur l'écran de création d'une pile.
4. Entrez un nom de pile.
5. Pour DatalakeAdminUserNameet DatalakeAdminUserPassword, entrez votre nom d'utilisateur et votre mot de passe IAM pour l'utilisateur administrateur de Data Lake.
6. Pour DataAnalystUsUserNameet DataAnalystUsUserPassword, entrez le nom d'utilisateur et le mot de passe pour le nom d'utilisateur et le mot de passe que vous souhaitez pour l'utilisateur analyste de données responsable du marché américain.
7. Pour DataAnalystJpUserNameet DataAnalystJpUserPassword, entrez le nom d'utilisateur et le mot de passe pour le nom d'utilisateur et le mot de passe que vous souhaitez pour l'utilisateur analyste de données responsable du marché japonais.
8. Pour DataLakeBucketName, entrez le nom de votre compartiment de données.
9. Pour DatabaseName, et TableNamelaissez comme valeur par défaut.
10. Choisissez Next (Suivant)
11. Sur la page suivante, choisissez Next.
12. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela CloudFormation pourrait créer des ressources IAM.
13. Sélectionnez Create (Créer).

La création de la pile peut prendre une minute.

Étape 2 : Requête sans filtres de données

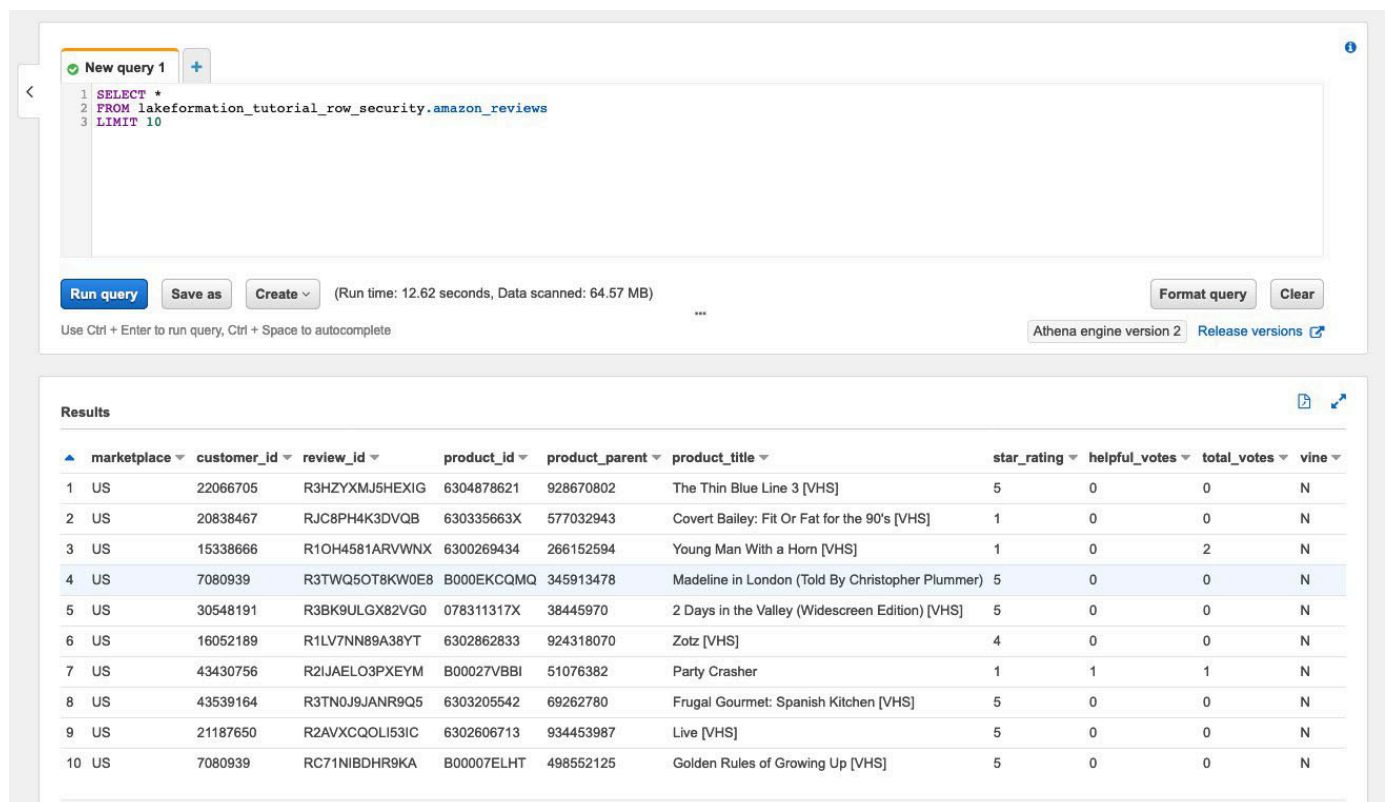
Après avoir configuré l'environnement, vous pouvez consulter le tableau des avis sur les produits. Interrogez d'abord la table sans contrôles d'accès au niveau des lignes pour vous assurer que vous pouvez voir les données. Si vous exécutez des requêtes dans Amazon Athena pour la première fois, vous devez configurer l'emplacement des résultats des requêtes.

Interrogez la table sans contrôle d'accès au niveau des lignes

1. Connectez-vous Athena console en <https://console.aws.amazon.com/athena/> tant qu'`DataLakeAdmin` utilisateur, et exécutez la requête suivante :

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

La capture d'écran suivante montre le résultat de la requête. Ce tableau ne comporte qu'une seule partition. Chaque enregistrement est donc un commentaire d'évaluation pour un produit vidéo. `product_category=Video`



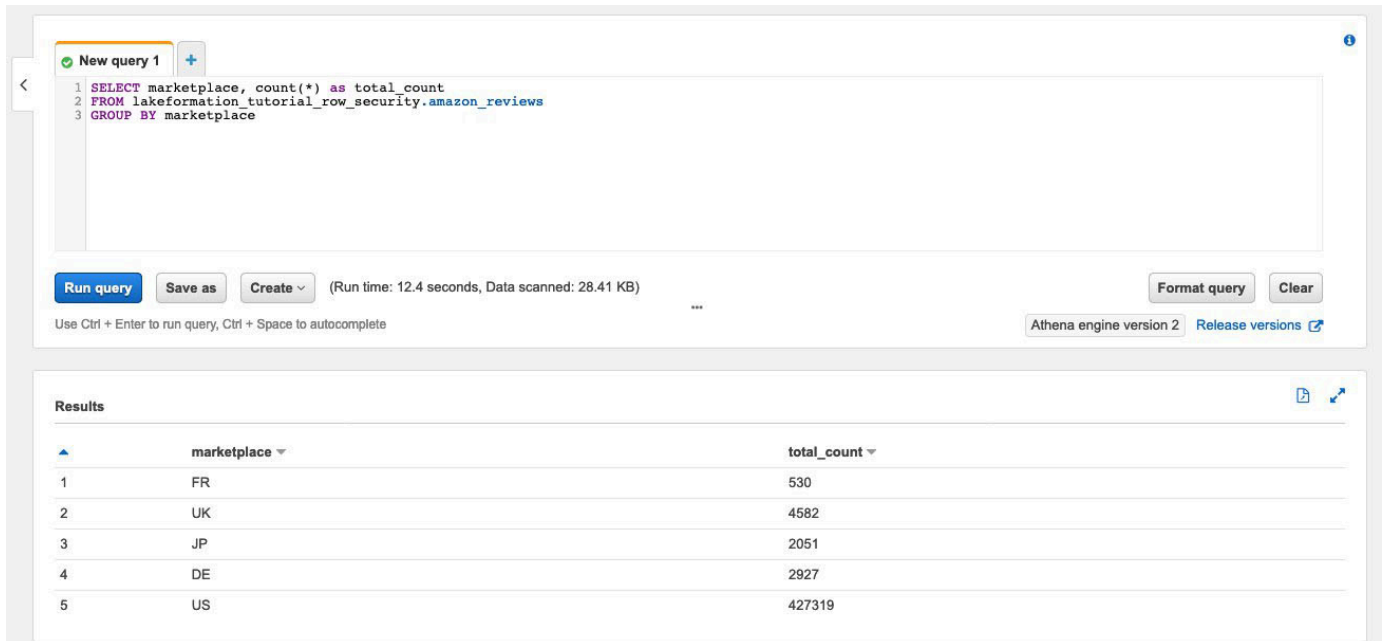
The screenshot shows the AWS Athena console interface. At the top, a query editor displays the SQL query: `SELECT * FROM lakeformation_tutorial_row_security.amazon_reviews LIMIT 10`. Below the editor, the 'Run query' button is highlighted. The console indicates a run time of 12.62 seconds and 64.57 MB of data scanned. Below the query editor, the 'Results' section displays a table with 10 rows of product reviews. The table columns are: marketplace, customer_id, review_id, product_id, product_parent, product_title, star_rating, helpful_votes, total_votes, and vine.

	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	N
3	US	15338666	R1OH4581ARVWNX	6300269434	266152594	Young Man With a Horn [VHS]	1	0	2	N
4	US	7080939	R3TWQ5OT8KW0E8	B000EKCQMQ	345913478	Madeline in London (Told By Christopher Plummer)	5	0	0	N
5	US	30548191	R3BK9ULGX82VG0	078311317X	38445970	2 Days in the Valley (Widescreen Edition) [VHS]	5	0	0	N
6	US	16052189	R1LV7NN89A38YT	6302862833	924318070	Zotz [VHS]	4	0	0	N
7	US	43430756	R2IJAELO3PXEYM	B00027VBBI	51076382	Party Crasher	1	1	1	N
8	US	43539164	R3TNOJ9JANR9Q5	6303205542	69262780	Frugal Gourmet: Spanish Kitchen [VHS]	5	0	0	N
9	US	21187650	R2AVXCQOLI53IC	6302606713	934453987	Live [VHS]	5	0	0	N
10	US	7080939	RC71NIBDHR9KA	B00007ELHT	498552125	Golden Rules of Growing Up [VHS]	5	0	0	N

2. Exécutez ensuite une requête d'agrégation pour récupérer le nombre total d'enregistrements par `marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
GROUP BY marketplace
```

La capture d'écran suivante montre le résultat de la requête. La marketplace colonne comporte cinq valeurs différentes. Au cours des étapes suivantes, vous allez configurer des filtres basés sur des lignes à l'aide de la marketplace colonne.



The screenshot shows the AWS Lake Formation console. At the top, there is a text area for a SQL query: `1 SELECT marketplace, count(*) as total_count`
`2 FROM lakeformation_tutorial_row_security.amazon_reviews`
`3 GROUP BY marketplace`. Below the query area are buttons for 'Run query', 'Save as', and 'Create'. A status bar indicates '(Run time: 12.4 seconds, Data scanned: 28.41 KB)'. To the right are 'Format query' and 'Clear' buttons. Below the query area, the 'Results' section displays a table with two columns: 'marketplace' and 'total_count'. The table contains five rows of data.

	marketplace	total_count
1	FR	530
2	UK	4582
3	JP	2051
4	DE	2927
5	US	427319

Étape 3 : configurer les filtres de données et accorder des autorisations

Ce didacticiel fait appel à deux analystes de données : l'un responsable du marché américain et l'autre du marché japonais. Chaque analyste utilise Athena pour analyser les avis des clients uniquement pour son marché spécifique. Créez deux filtres de données différents, l'un pour l'analyste responsable du marché américain et l'autre pour celui responsable du marché japonais. Accordez ensuite aux analystes leurs autorisations respectives.

Créez des filtres de données et accordez des autorisations

1. Créez un filtre pour restreindre l'accès aux US marketplace données.
 - a. Connectez-vous à la console Lake Formation <https://console.aws.amazon.com/lakeformation/> dans la région USA Est (Virginie du Nord) en tant qu'`DataLakeAdmin` utilisateur.
 - b. Choisissez Filtres de données.
 - c. Choisissez Créer un nouveau filtre.
 - d. Dans Nom du filtre de données, entrez `amazon_reviews_US`.

- e. Pour Base de données cible, choisissez la base de données `lakeformation_tutorial_row_security`.
 - f. Pour Table cible, choisissez la table `amazon_reviews`.
 - g. Pour l'accès au niveau des colonnes, laissez la valeur par défaut.
 - h. Pour Expression du filtre de ligne, entrez `marketplace='US'`.
 - i. Choisissez Create filter (Créer un filtre).
2. Créez un filtre pour restreindre l'accès aux marketplace données japonaises.
 - a. Sur la page Filtres de données, choisissez Créer un nouveau filtre.
 - b. Dans Nom du filtre de données, entrez `amazon_reviews_JP`.
 - c. Pour Base de données cible, choisissez la base de données `lakeformation_tutorial_row_security`.
 - d. Pour le tableau Target, choisissez la table `amazon_reviews`.
 - e. Pour l'accès au niveau des colonnes, laissez la valeur par défaut.
 - f. Pour Expression du filtre de ligne, entrez `marketplace='JP'`.
 - g. Choisissez Create filter (Créer un filtre).
3. Accordez ensuite des autorisations aux analystes de données utilisant ces filtres de données. Procédez comme suit pour accorder des autorisations à l'analyste de données américain (DataAnalystUS) :
 - a. Sous Autorisations, sélectionnez Autorisations du lac de données.
 - b. Sous Autorisation relative aux données, choisissez Grant.
 - c. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis sélectionnez le rôle. DataAnalystUS
 - d. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.
 - e. Pour Database (Base de données), sélectionnez `lakeformation_tutorial_row_security`.
 - f. Pour les tables (facultatif), choisissez. `amazon_reviews`
 - g. Pour les filtres de données (facultatif), sélectionnez `amazon_reviews_US`.
 - h. Pour les autorisations de filtrage de données, sélectionnez Sélectionner.
 - i. Choisissez Accorder.

4. Procédez comme suit pour accorder des autorisations à l'analyste de données japonais (DataAnalystJP) :
 - a. Sous Autorisations, sélectionnez Autorisations du lac de données.
 - b. Sous Autorisation relative aux données, choisissez Grant.
 - c. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis sélectionnez le rôle. DataAnalystJP
 - d. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.
 - e. Pour Database (Base de données), sélectionnez lakeformation_tutorial_row_security.
 - f. Pour les tables (facultatif), choisissez. amazon_reviews
 - g. Pour les filtres de données (facultatif), sélectionnezamazon_reviews_JP.
 - h. Pour les autorisations de filtrage de données, sélectionnez Sélectionner.
 - i. Choisissez Accorder.

Étape 4 : Requête à l'aide de filtres de données

Avec les filtres de données attachés au tableau des avis sur les produits, exécutez quelques requêtes et découvrez comment les autorisations sont appliquées par Lake Formation.

1. Connectez-vous à la console Athena en <https://console.aws.amazon.com/athena/> tant qu'utilisateur. DataAnalystUS
2. Exécutez la requête suivante pour récupérer quelques enregistrements, qui sont filtrés en fonction des autorisations au niveau des lignes que nous avons définies :

```
SELECT *  
FROM lakeformation_tutorial_row_security.amazon_reviews  
LIMIT 10
```

La capture d'écran suivante montre le résultat de la requête.

The screenshot shows the AWS Lake Formation console. At the top, there's a query editor with a SQL query: `SELECT * FROM lakeformation_tutorial_row_security.amazon_reviews LIMIT 10`. Below the editor are buttons for 'Run query', 'Save as', and 'Create'. A status bar indicates '(Run time: 11.9 seconds, Data scanned: 0 KB)'. To the right, there are buttons for 'Format query' and 'Clear'. Below the query editor, the 'Results' section displays a table with 10 rows of data. The columns are: marketplace, customer_id, review_id, product_id, product_parent, product_title, star_rating, helpful_votes, total_votes, vine, verified_purchase, and review_text. All rows show 'US' as the marketplace.

	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine	verified_purchase	review_text
1	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	KI
2	US	20261976	R2QTOLZUQERU5B	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	it
3	US	15947067	R1PHKR75RKZNSU	6303927319	850909689	Biography - Darryl Zanuck [VHS]	5	0	0	N	N	G
4	US	19288153	R1BL2WVE5X34UN	6304032153	479446069	Timon & Pumbaa: Quit Buggin Me [VHS]	5	0	0	N	N	Fi
5	US	19712967	R2DKOCIBS5FSP7	0784017743	35164822	Denise Austin - Hit the Spot: Arms & Bust [VHS]	5	0	0	N	Y	G
6	US	51047097	R2X5HQATT4IVR	0793960142	233936597	I Love Lucy - Lucy's Italian Movie/Ballet [VHS]	5	0	0	N	N	Fi
7	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	KI
8	US	51047097	R1C0H0G6NATZXO	6304872585	233936597	I Love Lucy: Lucy Meets Superman/Freez [VHS]	5	0	1	N	N	Fi
9	US	42808630	R2HXW7UD4IGZLN	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	M
10	US	11682952	R18IURLUPYI4DP	6302993717	42308924	Songs of Christmas [VHS]	1	0	0	N	Y	R

- De même, exécutez une requête pour compter le nombre total d'enregistrements par site de vente.

```
SELECT marketplace , count ( * ) as total_count
FROM lakeformation_tutorial_row_security .amazon_reviews
GROUP BY marketplace
```

Le résultat de la requête n'affiche que le marketplace US dans les résultats. Cela est dû au fait que l'utilisateur n'est autorisé à voir que les lignes dont la valeur de marketplace colonne est égale à US.

- Passez à l'utilisateur `DataAnalystJP` et exécutez la même requête.

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

Le résultat de la requête indique que seuls les enregistrements appartiennent au JP marketplace.

- Exécutez la requête pour compter le nombre total d'enregistrements par marketplace.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
```

```
GROUP BY marketplace
```

Le résultat de la requête montre uniquement la ligne appartenant au JPMarketplace.

Étape 5 : Nettoyer les AWS ressources

Nettoyage des ressources

Pour éviter que des frais supplémentaires ne vous soient facturés Compte AWS, vous pouvez supprimer les AWS ressources que vous avez utilisées pour ce didacticiel.

- [Supprimez la pile de formation des nuages.](#)

Partage d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation et de ressources nommées

Ce didacticiel explique comment vous pouvez configurer AWS Lake Formation pour partager en toute sécurité les données stockées dans un lac de données avec plusieurs entreprises, organisations ou unités commerciales, sans avoir à copier l'intégralité de la base de données. Il existe deux options pour partager vos bases de données et vos tables avec une autre personne Compte AWS en utilisant le contrôle d'accès entre comptes de Lake Formation :

- Contrôle d'accès basé sur des balises Lake Formation (recommandé)

Le contrôle d'accès basé sur les balises de Lake Formation est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans Lake Formation, ces attributs sont appelés balises LF. Pour plus d'informations, consultez [Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation](#).

- Ressources nommées Lake Formation

La méthode des ressources nommées Lake Formation est une stratégie d'autorisation qui définit les autorisations pour les ressources. Les ressources incluent des bases de données, des tables et des colonnes. Les administrateurs des lacs de données peuvent attribuer et révoquer des autorisations sur les ressources du Lake Formation. Pour plus d'informations, consultez [Partage de données entre comptes dans Lake Formation](#).

Nous recommandons d'utiliser des ressources nommées si l'administrateur du lac de données préfère accorder des autorisations explicites à des ressources individuelles. Lorsque vous utilisez la méthode des ressources nommées pour accorder à Lake Formation des autorisations sur une ressource du catalogue de données à un compte externe, Lake Formation utilise AWS Resource Access Manager (AWS RAM) pour partager la ressource.

Rubriques

- [Public visé](#)
- [Configurer les paramètres du catalogue de données de Lake Formation dans le compte du producteur](#)
- [Étape 1 : provisionnez vos ressources à l'aide AWS CloudFormation de modèles](#)
- [Étape 2 : Conditions préalables au partage entre comptes de Lake Formation](#)
- [Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises](#)
- [Étape 4 : Implémentation de la méthode de ressource nommée](#)
- [Étape 5 : Nettoyer les AWS ressources](#)

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Lorsqu'il s'agit de partager des tables du catalogue de données AWS Glue et d'administrer les autorisations dans Lake Formation, les responsables des données des comptes producteurs ont la propriété fonctionnelle des fonctions qu'ils prennent en charge et peuvent accorder l'accès à divers consommateurs, organisations externes et comptes. Le tableau suivant répertorie les rôles utilisés dans ce didacticiel :

Role	Description
DataLakeAdminProducer	L'utilisateur IAM de l'administrateur du lac de données dispose des droits d'accès suivants : • Accès complet en lecture, écriture et mise à jour à toutes les ressources du catalogue de données

Role	Description
	• Possibilité d'accorder des autorisations aux ressources • Peut créer des liens de ressources pour la table partagée • Peut attacher des balises LF aux ressources, ce qui permet d'accéder aux principaux en fonction des politiques créées par les gestionnaires de données
DataLakeAdminConsumer	L'utilisateur IAM de l'administrateur du lac de données dispose des droits d'accès suivants : • Accès complet en lecture, écriture et mise à jour à toutes les ressources du catalogue de données • Possibilité d'accorder des autorisations aux ressources • Peut créer des liens de ressources pour la table partagée • Peut attacher des balises LF aux ressources, ce qui permet d'accéder aux principaux en fonction des politiques créées par les gestionnaires de données
DataAnalyst	L' DataAnalyst utilisateur dispose des droits d'accès suivants : • Accès détaillé aux ressources partagées par les politiques d'accès basées sur les balises de Lake Formation ou à l'aide de la méthode des ressources nommées

Configurer les paramètres du catalogue de données de Lake Formation dans le compte du producteur

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour de plus amples informations, veuillez consulter [Exécution des tâches AWS de configuration initiale](#).

Le didacticiel part du principe que vous connaissez IAM. Pour plus d'informations sur IAM, consultez le guide de l'[utilisateur IAM](#).

Configurer les paramètres du catalogue de données de Lake Formation dans le compte du producteur

Note

Dans ce didacticiel, le compte contenant la table source est appelé compte producteur, et le compte qui a besoin d'accéder à la table source est appelé compte consommateur.

Lake Formation fournit son propre modèle de gestion des autorisations. Pour maintenir la rétrocompatibilité avec le modèle d'autorisation IAM, l'Superautorisation est accordée au groupe IAMAllowedPrincipals sur toutes les AWS Glue Data Catalog ressources existantes par défaut. En outre, les paramètres de contrôle d'accès Use only IAM sont activés pour les nouvelles ressources du catalogue de données. Ce didacticiel utilise un contrôle d'accès détaillé à l'aide des autorisations de Lake Formation et utilise des politiques IAM pour un contrôle d'accès détaillé. Consultez [Méthodes de contrôle d'accès précis](#) pour plus de détails. Par conséquent, avant d'utiliser un AWS CloudFormation modèle pour une configuration rapide, vous devez modifier les paramètres du Lake Formation Data Catalog dans le compte du producteur.

Important

Ce paramètre concerne toutes les bases de données et tables nouvellement créées. Nous vous recommandons donc vivement de suivre ce didacticiel dans un compte autre que celui de production ou dans un nouveau compte. De plus, si vous utilisez un compte partagé (tel que le compte de développement de votre entreprise), assurez-vous que cela n'affecte pas les ressources des autres. Si vous préférez conserver les paramètres de sécurité par défaut, vous devez effectuer une étape supplémentaire lorsque vous partagez des ressources

avec d'autres comptes, au cours de laquelle vous révoquez l'autorisation Super par défaut `IAMAllowedPrincipals` sur la base de données ou la table. Nous aborderons les détails plus loin dans ce didacticiel.

Pour configurer les paramètres du Lake Formation Data Catalog dans le compte du producteur, procédez comme suit :

1. Connectez-vous à l' AWS Management Console aide du compte producteur en tant qu'utilisateur administrateur ou en tant qu'utilisateur autorisé par l'`PutDataLakeSettingsAPI` Lake Formation.
2. Sur la console Lake Formation, dans le volet de navigation, sous Data Catalog, sélectionnez Settings.
3. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données et Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables dans les nouvelles bases de données

Choisissez Enregistrer.

AWS Lake Formation > Data catalog settings

Data catalog settings

Default permissions for newly created databases and tables

These settings maintain existing AWS Glue Data Catalog behavior. You can still set individual permissions on databases and tables, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

- ☐ Use only IAM access control for new databases
- ☐ Use only IAM access control for new tables in new databases

Default permissions for AWS CloudTrail

These settings specify the information being shown in AWS CloudTrail.

Resource owners

Enter resource owners you wish to share your CloudTrail access details with.

Q Enter an AWS account ID

Enter one or more AWS account IDs. Press Enter after each ID.

Cancel

Save

En outre, vous pouvez supprimer CREATE_DATABASE des autorisations pour les créateurs de base de données IAMAllowedPrincipals sous Rôles et tâches administratifs. Ce n'est qu'alors que vous pourrez déterminer qui peut créer une nouvelle base de données grâce aux autorisations de Lake Formation.

Étape 1 : provisionnez vos ressources à l'aide AWS CloudFormation de modèles

Le CloudFormation modèle de compte producteur génère les ressources suivantes :

- Un compartiment Amazon S3 qui servira de lac de données.
- Une fonction Lambda (pour les ressources personnalisées soutenues par CloudFormation Lambda). Nous utilisons cette fonction pour copier des exemples de fichiers de données du compartiment public Amazon S3 vers votre compartiment Amazon S3.

- Utilisateurs et politiques IAM : DataLakeAdminProducer.
- Les paramètres et autorisations appropriés de Lake Formation, notamment :
 - Définition de l'administrateur du lac de données de Lake Formation dans le compte du producteur
 - Enregistrement d'un bucket Amazon S3 comme emplacement du lac de données de Lake Formation (compte du producteur)
- Une AWS Glue Data Catalog base de données, une table et une partition. Comme il existe deux options pour partager des ressources Comptes AWS, ce modèle crée deux ensembles distincts de base de données et de tables.

Le CloudFormation modèle du compte client génère les ressources suivantes :

- Utilisateurs et politiques IAM :
 - DataLakeAdminConsumer
 - DataAnalyst
- Une AWS Glue Data Catalog base de données. Cette base de données permet de créer des liens vers des ressources partagées.

Créez vos ressources dans le compte du producteur

1. Connectez-vous à la AWS CloudFormation console à l'adresse <https://console.aws.amazon.com/cloudformation> dans la région USA Est (Virginie du Nord).
2. Choisissez [Launch Stack](#).
3. Choisissez Suivant.
4. Pour Nom de pile, entrez un nom de pile, tel que stack-producer.
5. Dans la section Configuration utilisateur, entrez le nom d'utilisateur et le mot de passe pour ProducerDataLakeAdminUserName et ProducerDataLakeAdminUserPassword.
6. Pour DataLakeBucketName, entrez le nom de votre bucket de data lake. Ce nom doit être unique au monde.
7. Pour DatabaseName et TableName, laissez les valeurs par défaut.
8. Choisissez Suivant.
9. Sur la page suivante, choisissez Next.

10. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
11. Choisissez Créer.

La création de la pile peut prendre jusqu'à une minute.

Créez vos ressources dans le compte client

1. Connectez-vous à la AWS CloudFormation console à l'adresse <https://console.aws.amazon.com/cloudformation> dans la région USA Est (Virginie du Nord).
2. Choisissez [Launch Stack](#).
3. Choisissez Suivant.
4. Pour Nom de pile, entrez un nom de pile, tel que `stack-consumer`.
5. Dans la section Configuration utilisateur, entrez le nom d'utilisateur et le mot de passe pour `ConsumerDataLakeAdminUserName` et `ConsumerDataLakeAdminUserPassword`.
6. Pour `DataAnalystUserName` et `DataAnalystUserPassword`, entrez le nom d'utilisateur et le mot de passe que vous souhaitez pour l'utilisateur IAM de l'analyste de données.
7. Pour `DataLakeBucketName`, entrez le nom de votre bucket de data lake. Ce nom doit être unique au monde.
8. Pour `DatabaseName`, conservez les valeurs par défaut.
9. Pour `AthenaQueryResultS3BucketName`, entrez le nom du compartiment Amazon S3 qui stocke les résultats des requêtes Amazon Athena. Si vous n'en avez pas, [créez un compartiment Amazon S3](#).
10. Choisissez Suivant.
11. Sur la page suivante, choisissez Next.
12. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
13. Choisissez Créer.

La création de la pile peut prendre jusqu'à une minute.

 Note

Une fois le didacticiel terminé, supprimez le stack in CloudFormation pour éviter d'encourir des frais. Vérifiez que les ressources sont correctement supprimées dans le statut de l'événement pour la pile.

Étape 2 : Conditions préalables au partage entre comptes de Lake Formation

Avant de partager des ressources avec Lake Formation, il existe des conditions préalables à la fois pour la méthode de contrôle d'accès basée sur les balises et pour la méthode des ressources nommées.

Conditions préalables complètes pour le contrôle d'accès basé sur des balises et le partage de données entre comptes

- Pour plus d'informations sur les exigences en matière de partage de données entre comptes, consultez la [Conditions préalables](#) section du chapitre Partage de données entre comptes.

Pour partager les ressources du catalogue de données avec la version 3 ou supérieure des paramètres de version multi-comptes, le concédant doit disposer des autorisations IAM définies dans la politique AWS `AWSLakeFormationCrossAccountManager` gérée de votre compte.

Si vous utilisez la version 1 ou la version 2 des paramètres de version multi-comptes, avant de pouvoir utiliser la méthode de contrôle d'accès basée sur les balises pour accorder l'accès aux ressources entre comptes, vous devez ajouter l'objet d'`JSON` autorisations suivant à la politique de ressources du catalogue de données du compte producteur. Cela donne au compte client l'autorisation d'accéder au catalogue de données lorsque cela `glue:EvaluatedByLakeFormationTags` est vrai. Cette condition s'applique également aux ressources pour lesquelles vous avez accordé une autorisation en utilisant les balises d'autorisation de Lake Formation sur le compte du consommateur. Cette politique est obligatoire pour tous ceux Compte AWS auxquels vous accordez des autorisations.

La politique suivante doit être intégrée à un `Statement` élément. Nous discutons de la politique IAM complète dans la section suivante.

```
{
```

```

    "Effect": "Allow",
    "Action": [
        "glue:*"
    ],
    "Principal": {
        "AWS": [
            "consumer-account-id"
        ]
    },
    "Resource": [
        "arn:aws:glue:region:account-id:table/*",
        "arn:aws:glue:region:account-id:database/*",
        "arn:aws:glue:region:account-id:catalog"
    ],
    "Condition": {
        "Bool": {
            "glue:EvaluatedByLakeFormationTags": true
        }
    }
}

```

Compléter les prérequis relatifs au partage entre comptes de la méthode des ressources nommées

1. S'il n'existe aucune politique de ressources du catalogue de données dans votre compte, le compte croisé Lake Formation vous permet de poursuivre comme d'habitude. Toutefois, s'il existe une politique de ressources pour le catalogue de données, vous devez y ajouter l'instruction suivante pour permettre à vos subventions entre comptes de réussir si elles sont accordées avec la méthode de ressource nommée. Si vous prévoyez d'utiliser uniquement la méthode de ressource nommée ou uniquement la méthode de contrôle d'accès basée sur des balises, vous pouvez ignorer cette étape. Dans ce didacticiel, nous évaluons les deux méthodes et nous devons ajouter la politique suivante.

La politique suivante doit être intégrée à un Statement élément. Nous discutons de la politique IAM complète dans la section suivante.

```

{
    "Effect": "Allow",
    "Action": [
        "glue:ShareResource"
    ],

```

```
"Principal": {
  "Service": "ram.amazonaws.com"
},
"Resource": [
  "arn:aws:glue:region:account-id:table/*/*",
  "arn:aws:glue:region:account-id:database/*",
  "arn:aws:glue:region:account-id:catalog"
]
}
```

2. Ajoutez ensuite la politique de AWS Glue Data Catalog ressources à l'aide du AWS Command Line Interface (AWS CLI).

Si vous accordez des autorisations entre comptes en utilisant à la fois la méthode de contrôle d'accès basée sur les balises et la méthode des ressources nommées, vous devez définir l'EnableHybridargument sur « true » lors de l'ajout des politiques précédentes. Parce que cette option n'est actuellement pas prise en charge sur la console, vous devez utiliser l'glue:PutResourcePolicyAPI et AWS CLI.

Créez d'abord un document de politique (tel que policy.json) et ajoutez les deux politiques précédentes. Remplacez *consumer-account-id* par le Compte AWS bénéficiaire *account ID* de la subvention, *region* par la région du catalogue de données contenant les bases de données et les tables pour lesquelles vous accordez des autorisations, et *account-id* par l'Compte AWS identifiant du producteur.

Entrez la AWS CLI commande suivante. Remplacez *glue-resource-policy* par les valeurs correctes (telles que file : //policy.json).

```
aws glue put-resource-policy --policy-in-json glue-resource-policy --enable-hybrid
TRUE
```

Pour plus d'informations, consultez [put-resource-policy](#).

Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises

Dans cette section, nous vous expliquons les étapes de haut niveau suivantes :

1. Définissez une balise LF.
2. Attribuez le tag LF à la ressource cible.
3. Accordez des autorisations LF-Tag au compte client.
4. Accordez des autorisations de données au compte client.
5. Vous pouvez éventuellement révoquer les autorisations pour la IAMAllowedPrincipals base de données, les tables et les colonnes.
6. Créez un lien de ressource vers la table partagée.
7. Créez une balise LF et attribuez-la à la base de données cible.
8. Accordez des autorisations relatives aux données LF-Tag au compte client.

Définir un LF-Tag

Note

Si vous êtes connecté à votre compte producteur, déconnectez-vous avant de suivre les étapes suivantes.

1. Connectez-vous au compte du producteur en tant qu'administrateur du lac de données à l'adresse <https://console.aws.amazon.com/lakeformation/>. Utilisez le numéro de compte du producteur, le nom d'utilisateur IAM (par défaut DataLakeAdminProducer) et le mot de passe que vous avez spécifiés lors de la création de la CloudFormation pile.
2. Sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), dans le volet de navigation, sous Permissions, choisissez LF-Tags and Permissions.
3. Choisissez Ajouter un tag LF.

Attribuez le tag LF à la ressource cible

Attribuez le tag LF à la ressource cible et accordez des autorisations de données à un autre compte

En tant qu'administrateur de data lake, vous pouvez associer des balises aux ressources. Si vous envisagez d'utiliser un rôle distinct, vous devrez peut-être accorder des autorisations de description et d'attachement au rôle distinct.

1. Dans le volet de navigation, sous Catalogue de données, sélectionnez Bases de données.

2. Sélectionnez la base de données cible (`lakeformation_tutorial_cross_account_database_tbac`) et dans le menu Actions, choisissez Modifier les balises LF.

Dans ce didacticiel, vous allez attribuer une balise LF à une base de données, mais vous pouvez également attribuer des balises LF à des tables et à des colonnes.

3. Choisissez Attribuer un nouveau tag LF.
4. Ajoutez la clé `Confidentiality` et la valeur `public`.
5. Choisissez Enregistrer.

Accorder l'autorisation LF-Tag au compte client

Toujours dans le compte du producteur, autorisez le compte du consommateur à accéder au LF-Tag.

1. Dans le volet de navigation, sous Permissions, sélectionnez LF-Tags and permissions.
2. Choisissez l'onglet LF-Tags, puis choisissez la clé et les valeurs du LF-Tag partagé avec le compte client (clé **Confidentiality** et valeur). `public`
3. Choisissez Grant permissions (Accorder des autorisations).
4. Pour le type d'autorisation, choisissez les autorisations de paire clé-valeur LF-Tag.
5. Pour Principaux, sélectionnez Comptes externes.
6. Entrez l'Compte AWS ID cible.

Comptes AWS au sein de la même organisation apparaissent automatiquement. Dans le cas contraire, vous devez saisir l' Compte AWS identifiant manuellement.

7. Sous Autorisations, sélectionnez Décrire.

Il s'agit des autorisations accordées au compte du consommateur. Les autorisations pouvant être accordées sont des autorisations que le compte client peut accorder à d'autres mandants.

8. Choisissez Accorder.

À ce stade, l'administrateur du lac de données du consommateur devrait être en mesure de trouver le tag de politique partagé via la console Lake Formation du compte consommateur, sous Autorisations, balises LF et autorisations.

Accorder l'autorisation d'accès aux données du compte client

Nous allons maintenant fournir un accès aux données au compte client en spécifiant une expression LF-Tag et en accordant au compte client l'accès à toute table ou base de données correspondant à l'expression.

1. Dans le volet de navigation, sous Autorisations, Autorisations du lac de données, sélectionnez Grant.
2. Pour Principaux, choisissez Comptes externes et entrez l' ID Compte AWS cible.
3. Pour les balises LF ou les ressources du catalogue, choisissez la clé et les valeurs de la balise LF partagée avec le compte du consommateur (clé **Confidentiality** et valeur). `public`
4. Pour les autorisations, sous Ressources associées à des balises LF (recommandé), choisissez Ajouter une balise LF.
5. Sélectionnez la clé et la valeur de la balise partagée avec le compte client (clé `Confidentiality` et valeur `public`).
6. Pour les autorisations de base de données, sélectionnez Décrire sous Autorisations de base de données pour accorder des autorisations d'accès au niveau de la base de données.
7. L'administrateur du lac de données du consommateur doit être en mesure de trouver la balise de politique partagée via le compte du consommateur sur la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>, sous Autorisations, rôles et tâches administratifs, balises LF.
8. Sélectionnez Décrire sous Autorisations pouvant être accordées afin que le compte client puisse accorder des autorisations au niveau de la base de données à ses utilisateurs.
9. Pour les autorisations de table et de colonne, sélectionnez Sélectionner et décrire sous Autorisations de table.
10. Sélectionnez Sélectionner et décrire sous Autorisations pouvant être accordées.
11. Choisissez Accorder.

Révoquez l'autorisation pour la **IAMAllowedPrincipals** base de données, les tables et les colonnes (facultatif).

Au tout début de ce didacticiel, vous avez modifié les paramètres du Lake Formation Data Catalog. Si vous avez ignoré cette partie, cette étape est obligatoire. Si vous avez modifié les paramètres de votre catalogue de données sur les formations des lacs, vous pouvez ignorer cette étape.

Au cours de cette étape, nous devons révoquer l'autorisation Super par défaut IAMAllowedPrincipals sur la base de données ou la table. Consultez [Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations Lake Formation](#) pour plus de détails.

Avant de révoquer l'autorisation pour IAMAllowedPrincipals, assurez-vous d'avoir accordé aux principaux IAM existants les autorisations nécessaires par le biais de Lake Formation. Cela comprend trois étapes :

1. Ajoutez l'autorisation IAM à l'utilisateur ou au rôle IAM cible avec l'GetDataAccessaction Lake Formation (avec la politique IAM).
2. Accordez à l'utilisateur ou au rôle IAM cible doté des autorisations relatives aux données de Lake Formation (modifier, sélectionner, etc.).
3. Révoquez ensuite les autorisations pour IAMAllowedPrincipals. Sinon, après avoir révoqué les autorisations pour IAMAllowedPrincipals, les principaux IAM existants risquent de ne plus être en mesure d'accéder à la base de données ou au catalogue de données cible.

La révocation de l'autorisation Super IAMAllowedPrincipals est requise lorsque vous souhaitez appliquer le modèle d'autorisation Lake Formation (au lieu du modèle de politique IAM) pour gérer l'accès des utilisateurs au sein d'un seul compte ou entre plusieurs comptes à l'aide du modèle d'autorisation Lake Formation. Vous n'êtes pas obligé de révoquer l'autorisation IAMAllowedPrincipals pour les autres tables pour lesquelles vous souhaitez conserver le modèle de politique IAM traditionnel.

À ce stade, l'administrateur du lac de données du compte client doit être en mesure de trouver la base de données et la table partagées via le compte client sur la console Lake Formation <https://console.aws.amazon.com/lakeformation/>, sous Data Catalog, databases. Si ce n'est pas le cas, vérifiez si les éléments suivants sont correctement configurés :

1. La balise de politique et les valeurs correctes sont attribuées aux bases de données et aux tables cibles.
2. L'autorisation de balise et l'autorisation de données correctes sont attribuées au compte client.
3. Révoquez la super autorisation par défaut de la IAMAllowedPrincipals base de données ou de la table.

Création d'un lien de ressource vers la table partagée

Lorsqu'une ressource est partagée entre des comptes et que les ressources partagées ne sont pas placées dans le catalogue de données des comptes consommateurs. Pour les rendre disponibles

et interroger les données sous-jacentes d'une table partagée à l'aide de services tels qu'Athena, nous devons créer un lien de ressource vers la table partagée. Un lien de ressource est un objet du catalogue de données qui est un lien vers une base de données ou une table locale ou partagée. Pour en savoir plus, consultez [Création de liens vers des ressources](#). En créant un lien vers une ressource, vous pouvez :

- Attribuez un nom différent à une base de données ou à une table conforme aux politiques de dénomination des ressources de votre catalogue de données.
- Utilisez des services tels qu'Athena et Redshift Spectrum pour interroger des bases de données ou des tables partagées.

Pour créer un lien vers une ressource, procédez comme suit :

1. Si vous êtes connecté à votre compte client, déconnectez-vous.
2. Connectez-vous en tant qu'administrateur du lac de données du compte client. Utilisez l'identifiant du compte client, le nom d'utilisateur IAM (par défaut DatalakeAdminConsumer) et le mot de passe que vous avez spécifiés lors de la création de la CloudFormation pile.
3. Sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), dans le volet de navigation, sous Data Catalog, Databases, choisissez la base de données partagée `lakeformation_tutorial_cross_account_database_tbac`.

Si la base de données n'apparaît pas, revoyez les étapes précédentes pour vérifier si tout est correctement configuré.

4. Choisissez Afficher les tables.
5. Choisissez la table partagée `amazon_reviews_table_tbac`.
6. Dans le menu Actions, choisissez Créer un lien vers une ressource.
7. Pour le nom du lien de ressource, entrez un nom (pour ce didacticiel, `amazon_reviews_table_tbac_resource_link`).
8. Sous Base de données, sélectionnez la base de données dans laquelle le lien de ressource est créé (pour cet article, la pile CloudFormation n'a créé la base de données `lakeformation_tutorial_cross_account_database_consumer`).
9. Choisissez Créer.

Le lien vers la ressource apparaît sous Catalogue de données, Tables.

Créez une balise LF et attribuez-la à la base de données cible

Les balises Lake Formation se trouvent dans le même catalogue de données que les ressources. Cela signifie que les balises créées dans le compte du producteur ne peuvent pas être utilisées lors de l'octroi de l'accès aux liens vers les ressources du compte client. Vous devez créer un ensemble distinct de balises LF dans le compte client pour utiliser le contrôle d'accès basé sur les balises LF lors du partage des liens de ressources dans le compte client.

1. Définissez le tag LF dans le compte client. Pour ce didacticiel, nous utilisons des clés `Division` et des valeurs `salesmarketing`, et `analyst`.
2. Attribuez la clé `Division` et la valeur du tag LF `analyst` à la base de données `lakeformation_tutorial_cross_account_database_consumer`, dans laquelle le lien de ressource est créé.

Autoriser le consommateur à utiliser les données LF-Tag

Enfin, accordez au consommateur l'autorisation d'utiliser les données LF-Tag.

1. Dans le volet de navigation, sous **Autorisations**, **Autorisations du lac de données**, sélectionnez **Grant**.
2. Pour **Principaux**, choisissez les utilisateurs et les rôles IAM, puis choisissez l'utilisateur `DataAnalyst`.
3. Pour les balises LF ou les ressources du catalogue, choisissez **Ressources associées aux balises LF** (recommandé).
4. Choisissez une division clé et un analyste de valeur.
5. Pour les autorisations de base de données, sélectionnez **Décrire** sous **Autorisations de base de données**.
6. Pour les autorisations de table et de colonne, sélectionnez **Sélectionner et décrire** sous **Autorisations de table**.
7. Choisissez **Accorder**.
8. Répétez ces étapes pour l'utilisateur `DataAnalyst`, où se trouvent la clé LF-Tag `Confidentiality` et la valeur `public`.

À ce stade, l'utilisateur de l'analyste de données du compte client doit être en mesure de trouver la base de données et le lien vers la ressource, et d'interroger la table partagée via la console

Athena à l'adresse. <https://console.aws.amazon.com/athena/> Si ce n'est pas le cas, vérifiez si les éléments suivants sont correctement configurés :

- Le lien de ressource est créé pour la table partagée
- Vous avez accordé à l'utilisateur l'accès au LF-Tag partagé par le compte du producteur
- Vous avez accordé à l'utilisateur l'accès à la balise LF associée au lien de ressource et à la base de données dans lesquels le lien de ressource est créé
- Vérifiez si vous avez attribué la balise LF correcte au lien de ressource et à la base de données dans laquelle le lien de ressource est créé

Étape 4 : Implémentation de la méthode de ressource nommée

Pour utiliser la méthode de ressource nommée, nous vous expliquons les étapes de haut niveau suivantes :

1. Vous pouvez éventuellement révoquer l'autorisation pour la **IAMAllowedPrincipals** base de données, les tables et les colonnes.
2. Accordez l'autorisation d'accès aux données du compte client.
3. Acceptez un partage de ressources depuis AWS Resource Access Manager.
4. Créez un lien de ressource pour la table partagée.
5. Accordez au consommateur l'autorisation d'accéder aux données de la table partagée.
6. Accordez au consommateur l'autorisation d'accès aux données pour le lien vers la ressource.

Révoquer l'autorisation pour la **IAMAllowedPrincipals** base de données, les tables et les colonnes (facultatif)

- Au tout début de ce didacticiel, nous avons modifié les paramètres du Lake Formation Data Catalog. Si vous avez ignoré cette partie, cette étape est obligatoire. Pour obtenir des instructions, reportez-vous à l'étape facultative de la section précédente.

Accorder l'autorisation d'accès aux données du compte client

1.

Note

Si vous êtes connecté au compte producteur en tant qu'autre utilisateur, déconnectez-vous d'abord.

Connectez-vous à la console Lake Formation en <https://console.aws.amazon.com/lakeformation/> utilisant l'administrateur du compte producteur Data Lake à l'aide de l'Compte AWS ID, du nom d'utilisateur IAM (par défaut `DataLakeAdminProducer`) et du mot de passe spécifiés lors de la création de la CloudFormation pile.

2. Sur la page Autorisations, sous Permissions du lac de données, sélectionnez Accorder.
3. Sous Principaux, sélectionnez Comptes externes, puis saisissez une Compte AWS IDs ou plusieurs AWS organisations IDs. Pour plus d'informations, voir : [AWS Organizations](#).

Organisations auxquelles appartient le compte producteur et au Comptes AWS sein de la même organisation apparaissent automatiquement. Dans le cas contraire, entrez manuellement l'identifiant du compte ou l'identifiant de l'organisation.

4. Pour les balises LF ou les ressources du catalogue, choisissez. `Named data catalog resources`
5. Sous Bases de données, sélectionnez la base de données `lakeformation_tutorial_cross_account_database_named_resource`.
6. Choisissez Ajouter un tag LF.
7. Sous Tables, sélectionnez Toutes les tables.
8. Pour les autorisations relatives aux colonnes du tableau, choisissez Sélectionner et Décrire sous Autorisations du tableau.
9. Sélectionnez Sélectionner et décrire, sous Autorisations pouvant être accordées.
10. Facultativement, pour les autorisations relatives aux données, choisissez Accès simple basé sur les colonnes si la gestion des autorisations au niveau des colonnes est requise.
11. Choisissez Accorder.

Si vous n'avez pas révoqué l'autorisation pour `IAMAllowedPrincipals`, le message d'erreur « Échec de l'octroi des autorisations » s'affiche. À ce stade, vous devriez voir la table cible partagée AWS RAM avec le compte client sous Autorisations, Autorisations relatives aux données.

Accepter un partage de ressources depuis AWS RAM

Note

Cette étape est requise uniquement pour le partage Compte AWS basé, et non pour le partage basé sur l'organisation.

1. Connectez-vous à la AWS console en <https://console.aws.amazon.com/connect/> utilisant l'administrateur du lac de données du compte client en utilisant le nom d'utilisateur IAM (par défaut DatalakeAdminConsumer) et le mot de passe spécifiés lors de la création de la CloudFormation pile.
2. Sur la AWS RAM console, dans le volet de navigation, sous Shared with me, Resource shares, choisissez la ressource Lake Formation partagée. Le statut doit être En attente.
3. Choisissez Action et Grant.
4. Confirmez les détails de la ressource, puis choisissez Accepter le partage des ressources.

À ce stade, l'administrateur du lac de données du compte client doit être en mesure de trouver la ressource partagée sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>) sous Catalogue de données, Bases de données.

Création d'un lien de ressource pour la table partagée

- Suivez les instructions de l'étape [Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises](#) 6 pour créer un lien de ressource pour une table partagée. Nommez le lien vers la ressource `amazon_reviews_table_named_resource_resource_link`. Créez le lien vers la ressource dans la base de données `lakeformation_tutorial_cross_account_database_consumer`.

Autoriser le consommateur à accéder aux données de la table partagée

Pour autoriser le consommateur à accéder aux données de la table partagée, procédez comme suit :

1. Sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), sous Autorisations, Autorisations du lac de données, choisissez Grant.

2. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis choisissez l'utilisateur. `DataAnalyst`
3. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.
4. Sous Bases de données, sélectionnez la base de données `lakeformation_tutorial_cross_account_database_named_resource`. Si la base de données ne figure pas dans la liste déroulante, choisissez Load more.
5. Sous Tables, sélectionnez la table `amazon_reviews_table_named_resource`.
6. Pour les autorisations de table et de colonne, sélectionnez Sélectionner et décrire sous Autorisations de table.
7. Choisissez Accorder.

Accorder au consommateur l'autorisation d'accès aux données pour le lien vers la ressource

En plus d'autoriser l'utilisateur du lac de données à accéder à la table partagée, vous devez également accorder à l'utilisateur du lac de données l'autorisation d'accéder au lien de ressource.

1. Sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), sous Autorisations, Autorisations du lac de données, choisissez Grant.
2. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis choisissez l'utilisateur. `DataAnalyst`
3. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.
4. Sous Bases de données, sélectionnez la base de données `lakeformation_tutorial_cross_account_database_consumer`. Si la base de données ne figure pas dans la liste déroulante, choisissez Load more.
5. Sous Tables, sélectionnez la table `amazon_reviews_table_named_resource_resource_link`.
6. Pour les autorisations relatives aux liens vers les ressources, sélectionnez Décrire sous Autorisations relatives aux liens vers les ressources.
7. Choisissez Accorder.

À ce stade, l'utilisateur de l'analyste de données du compte client doit être en mesure de trouver la base de données et le lien vers la ressource, et d'interroger la table partagée via la console Athena.

Si ce n'est pas le cas, vérifiez si les éléments suivants sont correctement configurés :

- Le lien de ressource est créé pour la table partagée
- Vous avez accordé à l'utilisateur l'accès à la table partagée par le compte du producteur
- Vous avez accordé à l'utilisateur l'accès au lien de ressource et à la base de données pour lesquels le lien de ressource est créé

Étape 5 : Nettoyer les AWS ressources

Pour éviter des frais indésirables Compte AWS, vous pouvez supprimer les AWS ressources que vous avez utilisées pour ce didacticiel.

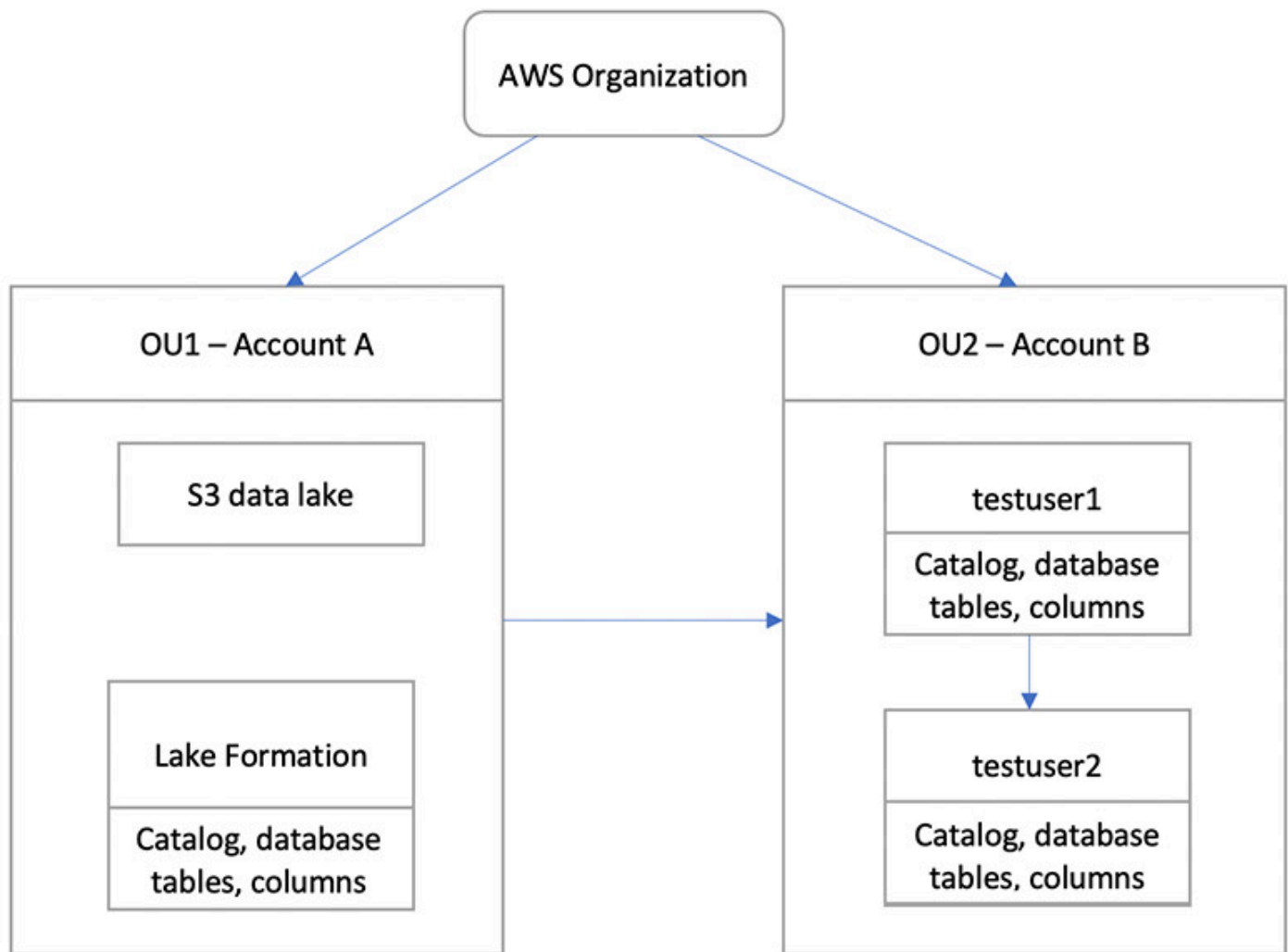
1. Connectez-vous à la console Lake Formation à <https://console.aws.amazon.com/lakeformation/> l'aide du compte producteur et supprimez ou modifiez les éléments suivants :
 - AWS Resource Access Manager partage des ressources
 - Tags Lake Formation
 - CloudFormation pile
 - Paramètres de la Lake Formation
 - AWS Glue Data Catalog
2. Connectez-vous à la console Lake Formation en <https://console.aws.amazon.com/lakeformation/> utilisant le compte client et supprimez ou modifiez les éléments suivants :
 - Tags Lake Formation
 - CloudFormation pile

Partage d'un lac de données à l'aide du contrôle d'accès précis de Lake Formation

Ce didacticiel fournit des step-by-step instructions sur la manière dont vous pouvez partager rapidement et facilement des ensembles de données à l'aide de Lake Formation lorsque vous en gérez plusieurs Comptes AWS avec AWS Organizations. Vous définissez des autorisations détaillées pour contrôler l'accès aux données sensibles.

Les procédures suivantes montrent également comment un administrateur du lac de données du compte A peut fournir un accès détaillé au compte B, et comment un utilisateur du compte B, agissant en tant que gestionnaire de données, peut accorder un accès détaillé à la table partagée aux autres utilisateurs de son compte. Au sein de chaque compte, les responsables des données peuvent déléguer l'accès de manière indépendante à leurs propres utilisateurs, ce qui donne de l'autonomie à chaque équipe ou secteur d'activité (LOB).

Le cas d'utilisation suppose que vous utilisez AWS Organizations pour gérer votre Comptes AWS. L'utilisateur du compte A dans une unité organisationnelle (OU1) accorde l'accès aux utilisateurs du compte B dans OU2. Vous pouvez utiliser la même approche lorsque vous n'utilisez pas Organizations, par exemple lorsque vous n'avez que quelques comptes. Le schéma suivant illustre le contrôle d'accès détaillé des ensembles de données dans un lac de données. Le lac de données est disponible dans le compte A. L'administrateur du lac de données du compte A fournit un accès détaillé au compte B. Le diagramme montre également qu'un utilisateur du compte B fournit un accès au niveau des colonnes de la table du lac de données du compte A à un autre utilisateur du compte B.



Rubriques

- [Public visé](#)
- [Prérequis](#)
- [Étape 1 : Fournir un accès détaillé à un autre compte](#)
- [Étape 2 : fournir un accès détaillé à un utilisateur du même compte](#)

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Le tableau suivant répertorie les rôles utilisés dans ce didacticiel :

Rôle	Description
Administrateur IAM	Utilisateur disposant de la politique AWS gérée : AdministratorAccess .
Administrateur du lac de données	Utilisateur disposant de la politique AWS gérée : AWSLakeFormationDataAdmin attaché au rôle.
Analyste des données	Utilisateur disposant de la politique AWS gérée : AmazonAthenaFullAccess joint.

Prérequis

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour de plus amples informations, veuillez consulter [Exécution des tâches AWS de configuration initiale](#).

Le didacticiel part du principe que vous connaissez IAM. Pour plus d'informations sur IAM, consultez le guide de l'[utilisateur IAM](#).

Vous avez besoin des ressources suivantes pour ce didacticiel :

- Deux unités organisationnelles :
 - OU1 — Contient le compte A
 - OU2 — Contient le compte B
- Emplacement d'un lac de données Amazon S3 (compartiment) dans le compte A.
- Utilisateur administrateur de lac de données dans le compte A. Vous pouvez créer un administrateur de lac de données à l'aide de la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>) ou de l'API Lake Formation. PutDataLakeSettings
- Lake Formation configurée dans le compte A, et l'emplacement du lac de données Amazon S3 enregistré auprès de Lake Formation dans le compte A.
- Deux utilisateurs du compte B avec les politiques gérées par IAM suivantes :
 - testuser1 — contient les politiques AWS AWSLakeFormationDataAdmin gérées.
 - testuser2 — La politique AWS AmazonAthenaFullAccess gérée est attachée.

- Une base de données testdb dans la base de données Lake Formation pour le compte B.

Étape 1 : Fournir un accès détaillé à un autre compte

Découvrez comment l'administrateur du lac de données du compte A fournit un accès détaillé au compte B.

Accorder un accès détaillé à un autre compte

1. Connectez-vous AWS Management Console au compte A <https://console.aws.amazon.com/connect/> en tant qu'administrateur du lac de données.
2. Ouvrez la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), puis choisissez Get started.
3. dans le volet de navigation, sélectionnez Databases.
4. Choisissez Create database (Créer une base de données).
5. Dans la section Détails de la base de données, sélectionnez Base de données.
6. Pour Nom, entrez un nom (pour ce didacticiel, nous utilisons `sampledb01`).
7. Assurez-vous que l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données n'est pas sélectionnée. Si cette option n'est pas sélectionnée, nous pouvons contrôler l'accès depuis Lake Formation.
8. Choisissez Créer une base de données.
9. Sur la page Bases de données, choisissez votre base de données `sampledb01`.
10. Dans le menu Actions, choisissez Grant.
11. Dans la section Accorder des autorisations, sélectionnez Compte externe.
12. Dans le Compte AWS champ ID ou ID d' AWS organisation, entrez le numéro de compte du compte B dans OU2.
13. Pour Table, choisissez la table à laquelle vous souhaitez que le compte B ait accès (pour cet article, nous utilisons la table `acc_a_area`). Vous pouvez éventuellement autoriser l'accès aux colonnes du tableau, ce que nous faisons dans cet article.
14. Pour Inclure les colonnes, choisissez les colonnes auxquelles vous souhaitez que le compte B ait accès (pour cet article, nous accordons des autorisations de type, de nom et d'identifiant).
15. Pour Colonnes, choisissez Inclure les colonnes.
16. Pour les autorisations relatives aux tables, sélectionnez Sélectionner.

17. Pour les autorisations pouvant être accordées, sélectionnez Sélectionner. Des autorisations pouvant être accordées sont requises pour que les utilisateurs administrateurs du compte B puissent accorder des autorisations à d'autres utilisateurs du compte B.
18. Choisissez Accorder.
19. Dans le volet de navigation, choisissez Tables.
20. Vous pouvez voir une connexion active dans la section « Comptes AWS et AWS organisations ayant accès ».

Création d'un lien vers une ressource

Les services intégrés tels qu'Amazon Athena ne peuvent pas accéder directement aux bases de données ou aux tables entre les comptes. Par conséquent, vous devez créer un lien de ressource afin qu'Athéna puisse accéder aux liens de ressources de votre compte vers les bases de données et les tables d'autres comptes. Créez un lien de ressource vers la table (`acc_a_area`) afin que les utilisateurs du compte B puissent interroger ses données auprès d'Athena.

1. Connectez-vous à la AWS console <https://console.aws.amazon.com/connect/> dans le compte B en tant que `testuser1`.
2. Sur la console Lake Formation (<https://console.aws.amazon.com/lakeformation/>), dans le volet de navigation, sélectionnez Tables. Vous devriez voir les tables auxquelles le compte A a fourni l'accès.
3. Choisissez la table `acc_a_area`.
4. Dans le menu Actions, choisissez Créer un lien vers une ressource.
5. Pour le nom du lien de ressource, entrez un nom (pour ce didacticiel, `acc_a_area_r1`).
6. Pour Base de données, choisissez votre base de données (`testdb`).
7. Sélectionnez Create (Créer).
8. Dans le volet de navigation, choisissez Tables.
9. Choisissez la table `acc_b_area_r1`.
10. Dans le menu Actions, choisissez Afficher les données.

Vous êtes redirigé vers la console Athena, où vous devriez voir la base de données et la table.

Vous pouvez désormais exécuter une requête sur la table pour voir la valeur de la colonne pour laquelle l'accès a été accordé à `testuser1` depuis le compte B.

Étape 2 : fournir un accès détaillé à un utilisateur du même compte

Cette section montre comment un utilisateur du compte B (`testuser1`), agissant en tant que gestionnaire de données, fournit à un autre utilisateur du même compte (`testuser2`) un accès détaillé au nom de colonne de la table partagée. `aac_b_area_r1`

Accorder un accès détaillé à un utilisateur du même compte

1. Connectez-vous à la AWS console <https://console.aws.amazon.com/connect/> dans le compte B en tant que `testuser1`.
2. Sur la console Lake Formation, dans le volet de navigation, sélectionnez Tables.

Vous pouvez accorder des autorisations sur une table via son lien de ressource. Pour ce faire, sur la page Tables, sélectionnez le lien vers la ressource `aac_b_area_r1`, puis dans le menu Actions, choisissez Grant on target.

3. Dans la section Accorder des autorisations, sélectionnez Mon compte.
4. Pour les utilisateurs et les rôles IAM, choisissez l'utilisateur `testuser2`.
5. Pour Colonne, choisissez le nom de la colonne.
6. Pour les autorisations relatives aux tables, sélectionnez Sélectionner.
7. Choisissez Accorder.

Lorsque vous créez un lien vers une ressource, vous êtes le seul à pouvoir le consulter et y accéder. Pour permettre aux autres utilisateurs de votre compte d'accéder au lien de ressource, vous devez accorder des autorisations sur le lien de ressource lui-même. Vous devez accorder les autorisations DESCRIBE ou DROP. Sur la page Tables, sélectionnez à nouveau votre tableau et dans le menu Actions, choisissez Grant.

8. Dans la section Accorder des autorisations, sélectionnez Mon compte.
9. Pour les utilisateurs et les rôles IAM, sélectionnez l'utilisateur `testuser2`.
10. Pour les autorisations relatives aux liens vers les ressources, sélectionnez Décrire.
11. Choisissez Accorder.
12. Connectez-vous à la AWS console dans le compte B en tant que `testuser2`.

Sur la console Athena (<https://console.aws.amazon.com/athena/>), vous devriez voir la base de données et la table. `aac_b_area_r1` Vous pouvez désormais exécuter une requête sur la table pour voir la valeur de la colonne à laquelle `testuser2` vous avez accès.

Permissions d'intégration à Lake Formation

AWS Lake Formation utilise le AWS Glue Data Catalog (catalogue de données) pour stocker les métadonnées des lacs de données Amazon S3 et des sources de données externes telles qu'Amazon Redshift sous forme de catalogues, de bases de données et de tables. Les métadonnées du catalogue de données sont organisées selon une hiérarchie de données à trois niveaux comprenant des catalogues, des bases de données et des tables. Il organise les données provenant de diverses sources dans des conteneurs logiques appelés catalogues. Les bases de données sont des ensembles de tables. Le catalogue de données contient également des liens vers des ressources, qui sont des liens vers des bases de données et des tables partagées dans des comptes externes, et sont utilisés pour l'accès entre comptes aux données du lac de données. Chaque AWS compte possède un catalogue de données par AWS région.

Lake Formation fournit un modèle d'autorisations du système de gestion de base de données relationnelle (RDBMS) permettant d'accorder ou de révoquer l'accès aux catalogues, bases de données, tables et colonnes du catalogue de données contenant des données sous-jacentes dans Amazon S3.

Avant de connaître les détails du modèle d'autorisations de Lake Formation, il est utile de consulter les informations générales suivantes :

- Les lacs de données gérés par Lake Formation se trouvent dans des emplacements désignés dans Amazon Simple Storage Service (Amazon S3). Le catalogue de données contient également des objets de catalogue. Chaque catalogue représente des données provenant de sources telles que les entrepôts de données Amazon Redshift, les Amazon DynamoDB bases de données et les sources de données tierces telles que Snowflake, MySQL, et de plus de 30 sources de données externes, qui sont intégrées via des connecteurs fédérés.
- Lake Formation gère un catalogue de données qui contient des métadonnées sur les données sources à importer dans vos lacs de données, telles que les données des journaux et des bases de données relationnelles, et sur les données de vos lacs de données dans Amazon S3. Le catalogue de données contient également des métadonnées relatives aux données provenant de sources de données externes autres qu'Amazon S3. Les métadonnées sont organisées sous forme de catalogues, de bases de données et de tables. Les tables de métadonnées contiennent le schéma, l'emplacement, le partitionnement et d'autres informations sur les données qu'elles représentent. Les bases de métadonnées sont des ensembles de tables.
- Le Lake Formation Data Catalog est le même que celui utilisé par AWS Glue. Vous pouvez utiliser des AWS Glue robots d'exploration pour créer des tables de catalogue de données, et vous

pouvez utiliser des tâches AWS Glue d'extraction, de transformation et de chargement (ETL) pour renseigner les données sous-jacentes de vos lacs de données.

- Les catalogues, les bases de données et les tables du catalogue de données sont appelés ressources du catalogue de données. Les tables du catalogue de données sont appelées tables de métadonnées pour les distinguer des tables des sources de données ou des données tabulaires d'Amazon S3. Les données vers lesquelles pointent les tables de métadonnées dans Amazon S3 ou dans les sources de données sont appelées données sous-jacentes.
- Un principal est un utilisateur ou un rôle, un utilisateur ou un groupe Amazon Quick, un utilisateur ou un groupe qui s'authentifie auprès de Lake Formation via un fournisseur SAML, ou pour le contrôle d'accès entre comptes, un identifiant de AWS compte, un identifiant d'organisation ou un identifiant d'unité organisationnelle.
- AWS Glue crée des tables de métadonnées, mais vous pouvez également créer manuellement des tables de métadonnées à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI). Lorsque vous créez une table de métadonnées, vous devez spécifier un emplacement. Lorsque vous créez une base de données, l'emplacement est facultatif. Les emplacements des tables peuvent être des emplacements Amazon S3 ou des emplacements de sources de données tels qu'une base de données Amazon Relational Database Service (Amazon RDS). Les emplacements de base de données sont toujours des emplacements Amazon S3.
- Les services intégrés à Lake Formation, tels qu'Amazon Athena et Amazon Redshift, peuvent accéder au catalogue de données pour obtenir des métadonnées et vérifier l'autorisation d'exécuter des requêtes. Pour une liste complète des services intégrés, voir [AWS intégrations de services avec Lake Formation](#).

Rubriques

- [Vue d'ensemble des autorisations relatives à Lake Formation](#)
- [Référence des personnalités de Lake Formation et des autorisations IAM](#)
- [Modification des paramètres par défaut de votre lac de données](#)
- [Permissions implicites de Lake Formation](#)
- [Référence des autorisations de Lake Formation](#)
- [Intégration d'IAM Identity Center](#)
- [Ajouter un emplacement Amazon S3 à votre lac de données](#)
- [Mode d'accès hybride](#)
- [Création d'objets dans AWS Glue Data Catalog](#)

- [Importation de données à l'aide de flux de travail dans Lake Formation](#)

Vue d'ensemble des autorisations relatives à Lake Formation

Il existe deux principaux types d'autorisations dans AWS Lake Formation :

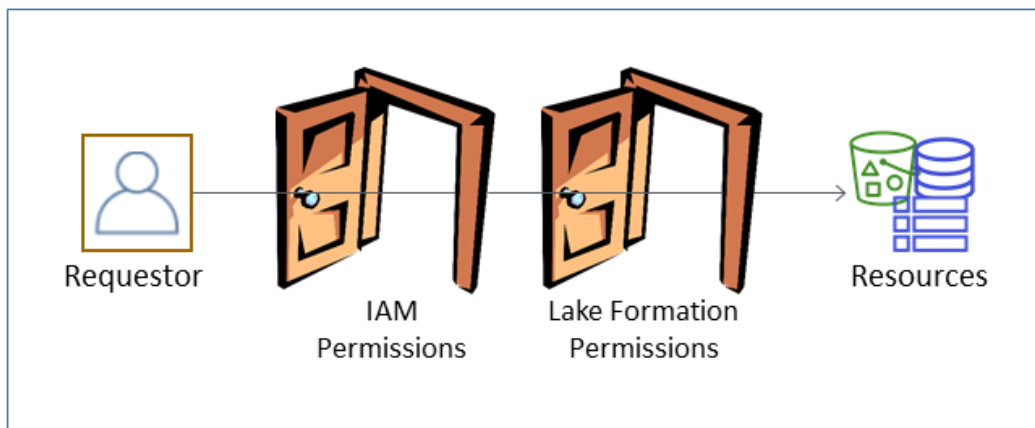
- Accès aux métadonnées : autorisations sur les ressources du catalogue de données (autorisations du catalogue de données).

Ces autorisations permettent aux principaux de créer, de lire, de mettre à jour et de supprimer des bases de données et des tables de métadonnées dans le catalogue de données.

- Accès aux données sous-jacent : autorisations sur les sites Amazon Simple Storage Service (Amazon S3) (autorisations d'accès aux données et autorisations de localisation des données).
 - Les autorisations du lac de données permettent aux principaux de lire et d'écrire des données sur des emplacements Amazon S3 sous-jacents, des données pointées par les ressources du catalogue de données.
 - Les autorisations de localisation des données permettent aux principaux de créer et de modifier des bases de données et des tables de métadonnées qui pointent vers des emplacements Amazon S3 spécifiques.

Pour les deux zones, Lake Formation utilise une combinaison d'autorisations Lake Formation et d'autorisations Gestion des identités et des accès AWS (IAM). Le modèle d'autorisations IAM comprend des politiques IAM. Le modèle d'autorisations de Lake Formation est implémenté sous forme de GRANT/REVOKE commandes de style DBMS, telles que `Grant SELECT on tableName to userName`

Lorsqu'un principal demande d'accès aux ressources du catalogue de données ou aux données sous-jacentes, pour que la demande aboutisse, il doit passer les contrôles d'autorisation par IAM et par Lake Formation.



Les autorisations de Lake Formation contrôlent l'accès aux ressources du catalogue de données, aux sites Amazon S3 et aux données sous-jacentes de ces sites. Les autorisations IAM contrôlent l'accès à la Lake Formation AWS Glue APIs et aux ressources. Ainsi, bien que vous ayez l'autorisation Lake Formation pour créer une table de métadonnées dans le catalogue de données (`CREATE_TABLE`), votre opération échoue si vous ne disposez pas de l'autorisation IAM sur `glue:CreateTableAPI`. (Pourquoi une `glue` : autorisation ? Parce que Lake Formation utilise le catalogue de AWS Glue données.)

Note

Les autorisations de Lake Formation ne s'appliquent que dans la région dans laquelle elles ont été accordées.

AWS Lake Formation exige que chaque principal (utilisateur ou rôle) soit autorisé à effectuer des actions sur les ressources gérées par Lake Formation. Le principal reçoit les autorisations nécessaires de la part de l'administrateur du lac de données ou d'un autre directeur autorisé à accorder les autorisations nécessaires à Lake Formation.

Lorsque vous accordez une autorisation de Lake Formation à un directeur, vous pouvez éventuellement accorder la possibilité de transmettre cette autorisation à un autre principal.

Vous pouvez utiliser l'API Lake Formation, le AWS Command Line Interface (AWS CLI) ou les pages Autorisations de données et Localisation des données de la console Lake Formation pour accorder ou révoquer les autorisations de Lake Formation.

Méthodes de contrôle d'accès précis

Avec un lac de données, l'objectif est de disposer d'un contrôle d'accès précis aux données. Dans Lake Formation, cela signifie un contrôle d'accès précis aux ressources du catalogue de données et aux emplacements Amazon S3. Vous pouvez obtenir un contrôle d'accès précis à l'aide de l'une des méthodes suivantes.

Method	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
Méthode 1 :	Ouvrir	À grain fin	Il s'agit de la méthode par défaut pour la rétrocompatibilité avec AWS Glue. Ouvert signifie que l'autorisation spéciale Super est accordée au groupe <code>IAMAllowedPrincipals</code>, qui <code>IAMAllowedPrincipals</code> est automatiquement créé et inclut tous les utilisateurs et rôles IAM autorisés à accéder aux ressources de votre catalogue de données par vos politiques IAM, et l'autorisation Super permet à un principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée. Ainsi, l'accès aux ressources du catalogue de données et aux emplacements Amazon S3 est contrôlé uniquement par les politiques IAM. Pour plus d'informations, consultez Modification des paramètres par défaut de votre lac de données et Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle.

Method	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
			La précision signifie que les politiques IAM contrôlent tous les accès aux ressources du catalogue de données et aux compartiments Amazon S3 individuels. Sur la console Lake Formation, cette méthode apparaît sous la forme Utiliser uniquement le contrôle d'accès IAM.

Method	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
Méthode 2 :	À grain fin	À gros grains	Il s'agit de la méthode recommandée. • Un accès précis implique l'octroi d'autorisations limitées de Lake Formation à des principaux responsables individuels sur les ressources du catalogue de données, les sites Amazon S3 et les données sous-jacentes de ces sites. • Une granularité grossière signifie des autorisations plus étendues sur les opérations individuelles et sur l'accès aux sites Amazon S3. Par exemple, une politique IAM grossière peut inclure "glue:*" ou "glue:Create*" plutôt que de laisser aux autorisations "glue:CreateTables". Lake Formation le soin de contrôler si un mandant peut ou non créer des objets de catalogue. Cela implique également de donner aux directeurs d'école l'accès à APIs ce dont ils ont besoin pour faire leur travail, tout en bloquant APIs d'autres ressources. Par exemple, vous pouvez créer une politique IAM qui permet à un principal de créer des ressources de catalogue de données et de créer et d'exécuter des flux de travail, mais qui n'autorise pas la création de AWS Glue connexions ou de fonctions

Method	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
			définies par l'utilisateur. Consultez les exemples plus loin dans cette section.

Important

Tenez compte des points suivants :

- Par défaut, les paramètres de contrôle d'accès Use only IAM de Lake Formation sont activés pour garantir la compatibilité avec le comportement existant du catalogue de AWS Glue données. Nous vous recommandons de désactiver ces paramètres une fois que vous serez passé à l'utilisation des autorisations de Lake Formation. Pour de plus amples informations, veuillez consulter [Modification des paramètres par défaut de votre lac de données](#).
- Les administrateurs de lacs de données et les créateurs de bases de données disposent d'autorisations implicites de Lake Formation que vous devez comprendre. Pour de plus amples informations, veuillez consulter [Permissions implicites de Lake Formation](#).

Contrôle d'accès aux métadonnées

Pour le contrôle d'accès aux ressources du catalogue de données, la discussion suivante suppose un contrôle d'accès précis avec les autorisations de Lake Formation et un contrôle d'accès grossier avec des politiques IAM.

Il existe deux méthodes distinctes pour accorder des autorisations à Lake Formation sur les ressources du catalogue de données :

- Contrôle d'accès aux ressources nommées : avec cette méthode, vous accordez des autorisations sur des bases de données ou des tables spécifiques en spécifiant les noms des bases de données ou des tables. Les subventions se présentent sous la forme suivante :

Accordez des autorisations aux directeurs sur les ressources [avec option de subvention].

Avec l'option de subvention, vous pouvez autoriser le bénéficiaire à accorder les autorisations à d'autres principaux.

- **Contrôle d'accès basé sur les balises** : avec cette méthode, vous attribuez une ou plusieurs balises LF aux bases de données, aux tables et aux colonnes du catalogue de données, et vous accordez des autorisations sur une ou plusieurs balises LF aux principaux. Chaque balise LF est une paire clé-valeur, telle que `department=sales`. Un principal dont les balises LF correspondent aux balises LF d'une ressource de catalogue de données peut accéder à cette ressource. Cette méthode est recommandée pour les lacs de données contenant un grand nombre de bases de données et de tables. C'est expliqué en détail dans [Contrôle d'accès basé sur des balises Lake Formation](#).

Les autorisations dont dispose un principal sur une ressource sont l'union des autorisations accordées par les deux méthodes.

Le tableau suivant récapitule les autorisations de Lake Formation disponibles sur les ressources du catalogue de données. Les en-têtes des colonnes indiquent la ressource pour laquelle l'autorisation est accordée.

Catalogue	Base de données	Table
CREATE_DATABASE	CREATE_TABLE	ALTER
	ALTER	DROP
	DROP	DESCRIBE
	DESCRIBE	SELECT*
		INSERT*
		DELETE*

Par exemple, l'`CREATE_TABLE` autorisation est accordée sur une base de données. Cela signifie que le principal est autorisé à créer des tables dans cette base de données.

Les autorisations marquées d'un astérisque (*) sont accordées sur les ressources du catalogue de données, mais elles s'appliquent aux données sous-jacentes. Par exemple, l'`DROP` autorisation sur

une table de métadonnées vous permet de supprimer la table du catalogue de données. Cependant, l'`DELETE` autorisation accordée sur la même table vous permet de supprimer les données sous-jacentes de la table dans Amazon S3, à l'aide, par exemple, d'une `DELETE` instruction SQL. Avec ces autorisations, vous pouvez également consulter le tableau sur la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI. Ainsi, `SELECT`, `INSERT`, et `DELETE` sont à la fois des autorisations de catalogue de données et des autorisations d'accès aux données.

Lorsque vous `SELECT` accordez un accord sur un tableau, vous pouvez ajouter un filtre qui inclut ou exclut une ou plusieurs colonnes. Cela permet un contrôle d'accès précis sur les colonnes des tables de métadonnées, limitant ainsi les colonnes que les utilisateurs des services intégrés peuvent voir lorsqu'ils exécutent des requêtes. Cette fonctionnalité n'est pas disponible uniquement en utilisant les politiques IAM.

Il existe également une autorisation spéciale nommée `Super`. L'`Super` autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée. Cette autorisation peut coexister avec les autres autorisations de Lake Formation. Par exemple, vous pouvez accorder `Super` `SELECT`, et `INSERT` sur une table de métadonnées. Le principal peut effectuer toutes les actions prises en charge sur la table, et lorsque vous les révoquez `Super`, les `INSERT` autorisations `SELECT` et sont conservées.

Pour plus de détails sur chaque autorisation, voir [Référence des autorisations de Lake Formation](#).

Important

Pour pouvoir consulter une table de catalogue de données créée par un autre utilisateur, vous devez disposer d'au moins une autorisation Lake Formation sur la table. Si vous disposez d'au moins une autorisation sur la table, vous pouvez également voir la base de données contenant la table.

Vous pouvez accorder ou révoquer les autorisations du catalogue de données à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI). Voici un exemple de AWS CLI commande qui accorde à l'utilisateur `datalake_user1` autorisation de créer des tables dans la `retail` base de données.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"} }'
```

Voici un exemple de politique IAM de contrôle d'accès grossière qui complète le contrôle d'accès détaillé par des autorisations Lake Formation. Il permet toutes les opérations sur n'importe quelle base de données ou table de métadonnées.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:*Database*",
        "glue:*Table*",
        "glue:*Partition*"
      ],
      "Resource": "*"
    }
  ]
}
```

L'exemple suivant est également grossier mais un peu plus restrictif. Il permet des opérations en lecture seule sur toutes les bases de données et tables de métadonnées du catalogue de données du compte et de la région désignés.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetTable",
        "glue:GetDatabase",
        "glue:GetDatabases"
      ],
      "Resource": "arn:aws:glue:us-east-1:111122223333:*"
    }
  ]
}
```

```

    }
  ]
}

```

Comparez ces politiques à la politique suivante, qui met en œuvre un contrôle d'accès précis basé sur l'IAM. Il accorde des autorisations uniquement sur un sous-ensemble de tables de la base de métadonnées de gestion de la relation client (CRM) du compte et de la région désignés.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetTable",
        "glue:GetDatabase",
        "glue:GetDatabases"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:catalog",
        "arn:aws:glue:us-east-1:111122223333:database/CRM",
        "arn:aws:glue:us-east-1:111122223333:table/CRM/P*"
      ]
    }
  ]
}

```

Pour d'autres exemples de politiques de contrôle d'accès détaillées, voir. [Référence des personnalités de Lake Formation et des autorisations IAM](#)

Contrôle d'accès aux données sous-jacent

Lorsqu'un AWS service intégré demande l'accès aux données d'un site Amazon S3 dont l'accès est contrôlé par AWS Lake Formation, Lake Formation fournit des informations d'identification temporaires pour accéder aux données.

Pour permettre à Lake Formation de contrôler l'accès aux données sous-jacentes sur un site Amazon S3, vous devez enregistrer cet emplacement auprès de Lake Formation.

Après avoir enregistré un site Amazon S3, vous pouvez commencer à accorder les autorisations Lake Formation suivantes :

- Autorisations d'accès aux données (SELECT, INSERT,) et DELETE) sur les tables du catalogue de données qui pointent vers cet emplacement.
- Autorisations de localisation des données à cet emplacement.

Les autorisations de localisation des données de Lake Formation contrôlent la possibilité de créer des ressources de catalogue de données pointant vers des emplacements Amazon S3 particuliers. Les autorisations de localisation des données fournissent un niveau de sécurité supplémentaire aux emplacements du lac de données. Lorsque vous accordez l'ALTER autorisation CREATE_TABLE ou à un principal, vous accordez également des autorisations de localisation des données afin de limiter les emplacements pour lesquels le principal peut créer ou modifier des tables de métadonnées.

Les emplacements Amazon S3 sont des compartiments ou des préfixes situés sous un compartiment, mais pas des objets Amazon S3 individuels.

Vous pouvez accorder des autorisations de localisation des données à un directeur à l'aide de la console Lake Formation, de l'API ou du AWS CLI. La forme générale d'une subvention est la suivante :

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

Si vous incluez `with grant option`, le bénéficiaire peut accorder les autorisations à d'autres directeurs.

N'oubliez pas que les autorisations de Lake Formation fonctionnent toujours en combinaison avec les autorisations Gestion des identités et des accès AWS (IAM) pour un contrôle d'accès précis. Pour read/write les autorisations sur les données Amazon S3 sous-jacentes, les autorisations IAM sont accordées comme suit :

Lorsque vous enregistrez un emplacement, vous spécifiez un rôle IAM qui accorde des autorisations de lecture/écriture sur cet emplacement. Lake Formation assume ce rôle lorsqu'il fournit des informations d'identification temporaires aux AWS services intégrés. Un rôle typique peut être associé à la politique suivante, dans laquelle l'emplacement enregistré est le compartiment `awsexamplebucket`.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:DeleteObject"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-bucket/*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-bucket"
      ]
    }
  ]
}
```

Lake Formation fournit un rôle lié à un service que vous pouvez utiliser lors de votre inscription pour créer automatiquement de telles politiques. Pour de plus amples informations, veuillez consulter [Utilisation de rôles liés à un service pour Lake Formation](#).

Par conséquent, l'enregistrement d'un emplacement Amazon S3 accorde les s3 : autorisations IAM requises pour cet emplacement, les autorisations étant spécifiées par le rôle utilisé pour enregistrer l'emplacement.

 Important

Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle

utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Pour read/write accéder aux données sous-jacentes, en plus des autorisations Lake Formation, les directeurs doivent également disposer de l'autorisation `lakeformation:GetDataAccess` IAM. Avec cette autorisation, Lake Formation accède à la demande d'informations d'identification temporaires pour accéder aux données.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "lakeformation:GetDataAccess",
      "Resource": "*"
    }
  ]
}
```

Dans la politique ci-dessus, vous devez définir le paramètre Resource sur « * » (tous). La spécification d'une autre ressource pour cette autorisation n'est pas prise en charge. Cette configuration permet à Lake Formation de gérer efficacement l'accès aux données dans l'ensemble de votre environnement de lac de données.

Note

Amazon Athena exige que l'utilisateur dispose de cette autorisation.
`lakeformation:GetDataAccess` Les autres services intégrés ont besoin de leur rôle d'exécution sous-jacent pour être `lakeformation:GetDataAccess` autorisés.

Cette autorisation est incluse dans les politiques suggérées dans le [Référence des personnalités de Lake Formation et des autorisations IAM](#).

En résumé, pour permettre aux responsables de Lake Formation de lire et d'écrire des données sous-jacentes avec un accès contrôlé par les autorisations de Lake Formation :

- Enregistrez les sites Amazon S3 contenant les données auprès de Lake Formation.
- Les principaux responsables qui créent des tables de catalogue de données pointant vers des emplacements de données sous-jacents doivent disposer d'autorisations de localisation des données.
- Les directeurs qui lisent et écrivent les données sous-jacentes doivent disposer des autorisations d'accès aux données de Lake Formation sur les tables du catalogue de données qui pointent vers les emplacements des données sous-jacentes.
- Les directeurs qui lisent et écrivent les données sous-jacentes doivent disposer de l'autorisation `lakeformation:GetDataAccess` IAM lorsque l'emplacement des données sous-jacentes est enregistré auprès de Lake Formation.

 Note

Le modèle d'autorisations de Lake Formation n'empêche pas l'accès aux sites Amazon S3 via l'API ou la console Amazon S3 si vous y avez accès via les politiques IAM ou Amazon S3. Vous pouvez associer des politiques IAM aux principaux pour bloquer cet accès.

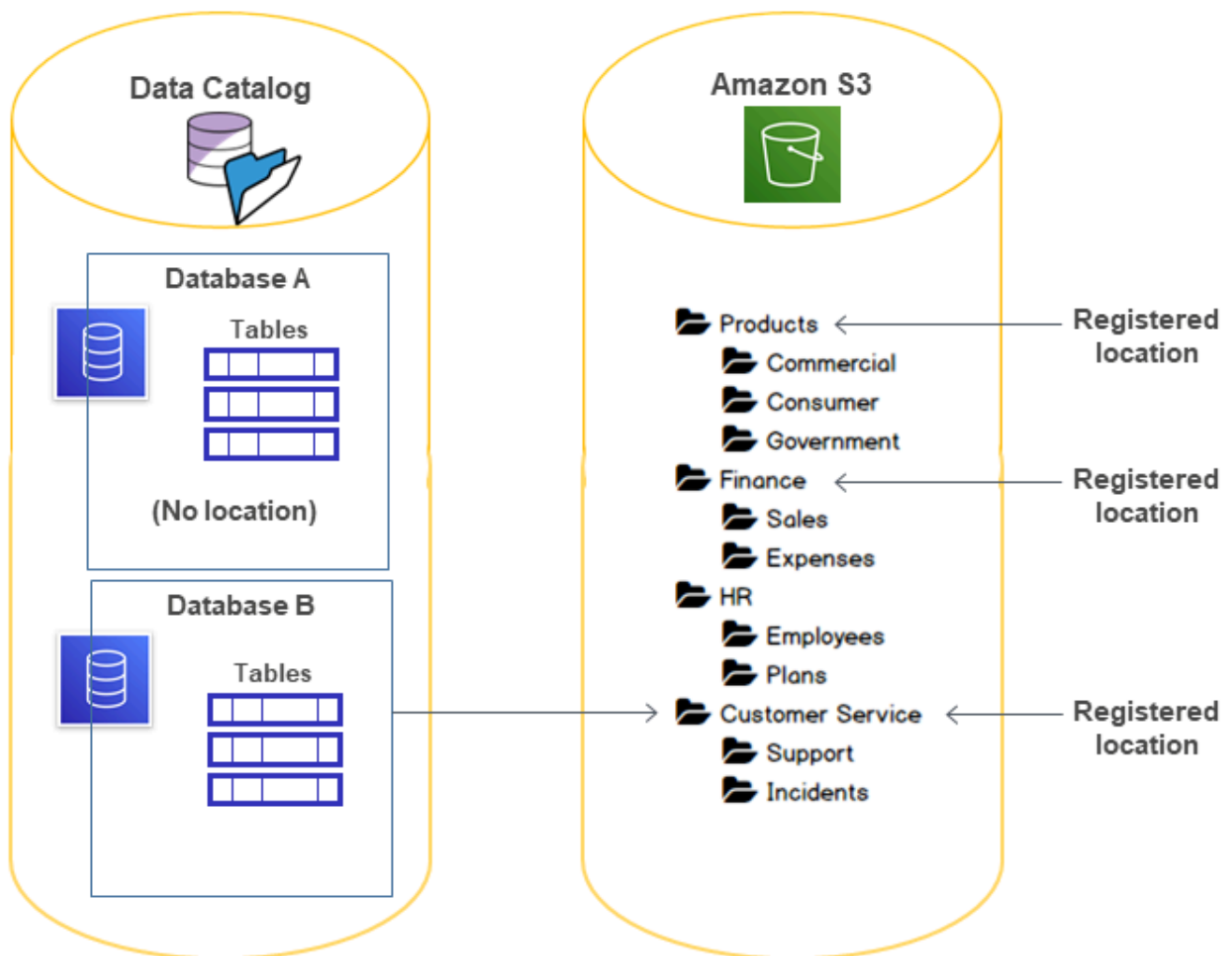
En savoir plus sur les autorisations de localisation des données

Les autorisations de localisation des données régissent le résultat des opérations de création et de mise à jour sur les bases de données et les tables du catalogue de données. Les règles sont les suivantes :

- Un principal doit disposer d'autorisations de localisation de données explicites ou implicites sur un emplacement Amazon S3 pour créer ou mettre à jour une base de données ou une table spécifiant cet emplacement.
- L'autorisation explicite `DATA_LOCATION_ACCESS` est accordée à l'aide de la console, de l'API ou AWS CLI.
- Des autorisations implicites sont accordées lorsqu'une base de données possède une propriété d'emplacement qui pointe vers un emplacement enregistré, que le principal dispose de l'`CREATE_TABLE` autorisation sur la base de données et que le principal essaie de créer une table à cet emplacement ou à un emplacement enfant.

- Si un principal obtient des autorisations de localisation des données sur un emplacement, il dispose des autorisations de localisation des données sur tous les emplacements enfants.
- Le principal n'a pas besoin d'autorisations de localisation des données pour effectuer read/write des opérations sur les données sous-jacentes. Il suffit d'avoir les autorisations SELECT d'accès aux INSERT données. Les autorisations de localisation des données s'appliquent uniquement à la création de ressources de catalogue de données pointant vers l'emplacement.

Examinez le scénario illustré dans le schéma suivant.



Dans ce schéma :

- Les compartiments Amazon S3 ProductsFinance, et Customer Service sont enregistrés auprès de Lake Formation.
- Database A n'a aucune propriété de localisation, Database B mais possède une propriété de localisation qui pointe vers le Customer Service bucket.
- L'utilisateur `datalake_user` possède `CREATE_TABLE` les deux bases de données.
- L'utilisateur `datalake_user` a obtenu des autorisations de localisation des données uniquement sur le Products compartiment.

Voici les résultats obtenus lorsque l'utilisateur `datalake_user` essaie de créer une table de catalogue dans une base de données donnée à un emplacement donné.

Emplacement où **`datalake_user`** essaie de créer une table

Base de données et localisation	Succès ou échec	Raison
Base de données A à Finance/Sales	Échoue	Aucune autorisation de localisation des données
Base de données A à Products	Succès	Dispose d'une autorisation de localisation des données
Base de données A à HR/Plans	Succès	L'emplacement n'est pas enregistré
Base de données B à Customer Service/Incidents	Succès	La base de données possède une propriété de localisation à Customer Service

Pour plus d'informations, consultez les ressources suivantes :

- [Ajouter un emplacement Amazon S3 à votre lac de données](#)
- [Référence des autorisations de Lake Formation](#)
- [Référence des personnalités de Lake Formation et des autorisations IAM](#)

Référence des personnalités de Lake Formation et des autorisations IAM

Cette section répertorie certains personnages de Lake Formation suggérés et leurs autorisations suggérées Gestion des identités et des accès AWS (IAM). Pour plus d'informations sur les autorisations de Lake Formation, consultez [the section called “Référence des autorisations de Lake Formation”](#).

AWS Lake Formation personas

Le tableau suivant répertorie les AWS Lake Formation personnages suggérés.

Lake Formation Personas

Persona	Description
Administrateur IAM (superutilisateur)	(Obligatoire) Utilisateur capable de créer des utilisateurs et des rôles IAM. Possède la politique AdministratorAccess AWS gérée. Possède toutes les autorisations sur toutes les ressources du Lake Formation. Possibilité d'ajouter des administrateurs de data lake. Impossible d'accorder les autorisations de Lake Formation s'il n'est pas également désigné administrateur du lac de données.
Administrateur du lac de données	(Obligatoire) Utilisateur pouvant enregistrer des sites Amazon S3, accéder au catalogue de données, créer des bases de données, créer et exécuter des flux de travail, accorder des autorisations Lake Formation à d'autres utilisateurs et consulter les AWS CloudTrail journaux. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données. Impossible d'ajouter d'autres administrateurs de data lake.
Administrateur en lecture seule	(Facultatif) Utilisateur qui peut consulter les informations principales, les ressources du catalogue de données, les autorisations et les AWS CloudTrail journaux, sans les autorisations nécessaires pour effectuer des mises à jour.

Persona	Description
Ingénieur de données	(Facultatif) Utilisateur capable de créer des bases de données, de créer et d'exécuter des robots d'exploration et des flux de travail, et d'accorder des autorisations à Lake Formation sur les tables du catalogue de données créées par les robots et les flux de travail. Nous vous recommandons de faire de tous les ingénieurs de données des créateurs de bases de données. Pour de plus amples informations, veuillez consulter Création d'une base de données .
Analyste des données	(Facultatif) Utilisateur qui peut exécuter des requêtes sur le lac de données en utilisant, par exemple, Amazon Athena. Dispose uniquement de suffisamment d'autorisations pour exécuter des requêtes.
Rôle du flux de travail	(Obligatoire) Rôle qui exécute un flux de travail pour le compte d'un utilisateur. Vous spécifiez ce rôle lorsque vous créez un flux de travail à partir d'un plan.

 Note

Dans Lake Formation, les administrateurs de lacs de données ajoutés après la création de la base de données peuvent accorder des autorisations mais ne disposent pas automatiquement d'autorisations d'accès aux données telles que SELECT ou DESCRIBE. Les administrateurs qui créent des bases de données reçoivent SUPER des autorisations sur ces bases de données. Ce comportement est intentionnel : bien que tous les administrateurs puissent s'octroyer les autorisations nécessaires, celles-ci ne sont pas automatiquement appliquées aux ressources préexistantes. Par conséquent, les administrateurs doivent s'accorder explicitement l'accès aux bases de données qui existaient avant de se voir attribuer des privilèges d'administrateur.

AWS politiques gérées pour Lake Formation

Vous pouvez accorder les autorisations Gestion des identités et des accès AWS (IAM) requises pour travailler en AWS Lake Formation utilisant des politiques AWS gérées et des politiques intégrées. Les politiques AWS gérées suivantes sont disponibles pour Lake Formation.

AWS politique gérée : AWSLakeFormationDataAdmin

[AWSLakeFormationDataAdmin](#) la politique accorde un accès administratif AWS Lake Formation aux services connexes tels que la gestion AWS Glue des lacs de données.

Vous pouvez associer AWSLakeFormationDataAdmin à vos utilisateurs, groupes et rôles.

Détails de l'autorisation

- **CloudTrail**— Permet aux directeurs de consulter les AWS CloudTrail journaux. Cela est nécessaire pour vérifier toute erreur dans la configuration du lac de données.
- **Glue**— Permet aux principaux d'afficher, de créer et de mettre à jour des tables de métadonnées et des bases de données dans le catalogue de données. Cela inclut les opérations d'API qui commencent par `Get`, `List`, `Create`, `Update`, `Delete`, et `Search`. Cela est nécessaire pour gérer les métadonnées des tables des lacs de données.
- **IAM**— Permet aux principaux de récupérer des informations sur les utilisateurs IAM, les rôles et les politiques associées aux rôles. Cela est nécessaire pour que l'administrateur des données puisse examiner et répertorier les utilisateurs et les rôles IAM afin d'accorder des autorisations à Lake Formation.
- **Lake Formation**— Accorde aux administrateurs de lacs de données les autorisations nécessaires à Lake Formation pour gérer les lacs de données.
- **S3**— Permet aux principaux de récupérer des informations sur les compartiments Amazon S3 et leur emplacement afin de configurer l'emplacement des données pour les lacs de données.

```
"Statement": [  
  {  
    "Sid": "AWSLakeFormationDataAdminAllow",  
    "Effect": "Allow",  
    "Action": [  
      "lakeformation:*",  
      "cloudtrail:DescribeTrails",  
      "cloudtrail:LookupEvents",  
    ]  
  }  
]
```

```

        "glue:CreateCatalog",
"glue:UpdateCatalog",
        "glue>DeleteCatalog",
"glue:GetCatalog",
        "glue:GetCatalogs",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue>CreateDatabase",
        "glue:UpdateDatabase",
        "glue>DeleteDatabase",
        "glue:GetConnections",
        "glue:SearchTables",
        "glue:GetTable",
        "glue:CreateTable",
        "glue:UpdateTable",
        "glue>DeleteTable",
        "glue:GetTableVersions",
        "glue:GetPartitions",
        "glue:GetTables",
        "glue:ListWorkflows",
        "glue:BatchGetWorkflows",
        "glue>DeleteWorkflow",
        "glue:GetWorkflowRuns",
        "glue:StartWorkflowRun",
        "glue:GetWorkflow",
        "s3:ListBucket",
        "s3:GetBucketLocation",
        "s3:ListAllMyBuckets",
        "s3:GetBucketAcl",
        "iam:ListUsers",
        "iam:ListRoles",
        "iam:GetRole",
        "iam:GetRolePolicy"
    ],
    "Resource": "*"
},
{
    "Sid": "AWSLakeFormationDataAdminDeny",
    "Effect": "Deny",
    "Action": [
        "lakeformation:PutDataLakeSettings"
    ],
    "Resource": "*"
}

```

```
]
}
```

Note

La `AWSLakeFormationDataAdmin` politique n'accorde pas toutes les autorisations requises aux administrateurs de lacs de données. Des autorisations supplémentaires sont nécessaires pour créer et exécuter des flux de travail et enregistrer des sites avec le rôle lié au service `AWSServiceRoleForLakeFormationDataAccess`. Pour plus d'informations, consultez [Création d'un administrateur de lac de données](#) et [Utilisation de rôles liés à un service pour Lake Formation](#).

AWS politique gérée : `AWSLakeFormationCrossAccountManager`

[AWSLakeFormationCrossAccountManager](#) la politique fournit un accès multicompte aux AWS Glue ressources via Lake Formation et accorde un accès en lecture à d'autres services requis tels que AWS Organizations et AWS RAM.

Vous pouvez associer `AWSLakeFormationCrossAccountManager` à vos utilisateurs, groupes et rôles.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes.

- `Glue`— Permet aux principaux de définir ou de supprimer la politique de ressources du catalogue de données pour le contrôle d'accès.
- `Organizations`— Permet aux responsables de récupérer les informations relatives aux comptes et aux unités organisationnelles (UO) d'une organisation.
- `ram:CreateResourceShare`— Permet aux principaux de créer un partage de ressources.
- `ram:UpdateResourceShare`: permet aux principaux de modifier certaines propriétés du partage de ressources spécifié.
- `ram>DeleteResourceShare`— Permet aux principaux de supprimer le partage de ressources spécifié.
- `ram:AssociateResourceShare`— Permet aux principaux d'ajouter la liste de principes et la liste de ressources spécifiées à un partage de ressources.

- **ram:DisassociateResourceShare**— Permet aux principaux d'empêcher les principaux ou les ressources spécifiés de participer au partage de ressources spécifié.
- **ram:GetResourceShares**— Permet aux principaux de récupérer des informations sur les partages de ressources que vous possédez ou qui sont partagés avec vous.
- **ram:RequestedResourceType**— Permet aux principaux de récupérer le type de ressource (base de données, table ou catalogue).
- **AssociateResourceSharePermission**— Permet aux principaux d'ajouter ou de remplacer l' AWS RAM autorisation pour un type de ressource inclus dans un partage de ressources. Vous pouvez avoir exactement une autorisation associée à chaque type de ressource dans le partage de ressources.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Sid": "AllowCreateResourceShare",
    "Effect": "Allow",
    "Action": [
      "ram:CreateResourceShare"
    ],
    "Resource": "*",
    "Condition": {
      "StringLikeIfExists": {
        "ram:RequestedResourceType": [
          "glue:Table",
          "glue:Database",
          "glue:Catalog"
        ]
      }
    }
  },
  {
    "Sid": "AllowManageResourceShare",
    "Effect": "Allow",
    "Action": [
      "ram:UpdateResourceShare",
      "ram>DeleteResourceShare",
      "ram:AssociateResourceShare",
```

```

        "ram:DisassociateResourceShare",
        "ram:GetResourceShares"
    ],
    "Resource": "*",
    "Condition": {
        "StringLike": {
            "ram:ResourceShareName": [
                "LakeFormation*"
            ]
        }
    }
},
{
    "Sid": "AllowManageResourceSharePermissions",
    "Effect": "Allow",
    "Action": [
        "ram:AssociateResourceSharePermission"
    ],
    "Resource": "*",
    "Condition": {
        "ArnLike": {
            "ram:PermissionArn": [
                "arn:aws:ram::aws:permission/AWSRAMLFEnabled*"
            ]
        }
    }
},
{
    "Sid": "AllowXAcctManagerPermissions",
    "Effect": "Allow",
    "Action": [
        "glue:PutResourcePolicy",
        "glue>DeleteResourcePolicy",
        "organizations:DescribeOrganization",
        "organizations:DescribeAccount",
        "ram:Get*",
        "ram:List*"
    ],
    "Resource": "*"
},
{
    "Sid": "AllowOrganizationsPermissions",
    "Effect": "Allow",
    "Action": [

```

```

        "organizations:ListRoots",
        "organizations:ListAccountsForParent",
        "organizations:ListOrganizationalUnitsForParent"
    ],
    "Resource": "*"
}
]
}

```

AWS politique gérée : AWSGlue ConsoleFullAccess

[AWSGlueConsoleFullAccess](#) la politique accorde un accès complet aux AWS Glue ressources lorsqu'une identité à laquelle la politique est attachée utilise le AWS Management Console. Si vous suivez la convention de dénomination pour les ressources spécifiées dans la politique, les utilisateurs bénéficient des capacités totales de la console. Cette politique s'applique généralement aux utilisateurs de la AWS Glue console.

Lake Formation assume également le rôle `AWSGlueServiceRole` de service permettant d'accéder aux services connexes, notamment Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3) et Amazon. AWS Glue CloudWatch

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

Cette politique est attachée à un rôle lié au service nommé `ServiceRoleForLakeFormationDataAccess` qui permet au service d'effectuer des actions sur les ressources à votre demande. Vous ne pouvez pas associer cette politique à vos identités IAM.

Cette politique permet aux AWS services intégrés de Lake Formation tels qu'Amazon Redshift Amazon Athena d'utiliser le rôle lié au service pour découvrir les ressources Amazon S3.

Pour plus d'informations, voir, [Utilisation de rôles liés à un service pour Lake Formation](#).

Détails de l'autorisation

Cette politique inclut l'autorisation suivante.

- `s3:ListAllMyBuckets`— Renvoie la liste de tous les buckets appartenant à l'expéditeur authentifié de la demande.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "LakeFormationDataAccessServiceRolePolicy",
      "Effect": "Allow",
      "Action": [
        "s3:ListAllMyBuckets"
      ],
      "Resource": [
        "arn:aws:s3:::*"
      ]
    }
  ]
}
```

Lake Formation met à jour les politiques AWS gérées

Consultez les détails des mises à jour des politiques AWS gérées pour Lake Formation depuis que ce service a commencé à suivre ces modifications.

Modifier	Description	Date
AWSLakeFormationCrossAccountManager Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManager politique en remplaçant l'opérateur de StringLike condition par l'ArnLikeopérateur qui permet à IAM d'effectuer la vérification du format ARN.	janvier 2025
AWSLakeFormationDataAdmin Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationDataAdmin politique en ajoutant le AWS Glue Data Catalog CRUD suivant dans le APIs cadre de la fonctionnalité multi-catalogue.	décembre 2024

Modifier	Description	Date
	• colle : CreateCatalog • colle : UpdateCatalog • colle : DeleteCatalog • colle : GetCatalog • colle : GetCatalogs Ce changement de politique géré vise à garantir que l'administrateur de Lake Formation dispose par défaut de l'autorisation IAM pour ces nouvelles opérations.	
AWSLakeFormationCrossAccountManager Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManager politique en ajoutant des éléments Sid à la déclaration de politique.	mars 2024
AWSLakeFormationDataAdmin Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationDataAdmin politique en ajoutant un élément Sid à la déclaration de politique et en supprimant une action redondante.	mars 2024
LakeFormationDataAccessServiceRolePolicy Politique mise à jour de Lake Formation.	Lake Formation a amélioré la LakeFormationDataAccessServiceRolePolicy politique en ajoutant un élément Sid à la déclaration de politique .	février 2024

Modifier	Description	Date
AWSLakeFormationCrossAccountManager Politique mise à jour de Lake Formation.	Lake Formation a amélioré cette AWSLakeFormationCrossAccountManager politique en ajoutant une nouvelle autorisation permettant le partage de données entre comptes en mode d'accès hybride.	octobre 2023
AWSLakeFormationCrossAccountManager Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManager politique de création d'un seul partage de ressources par compte bénéficiaire lors du premier partage d'une ressource. Toutes les ressources partagées par la suite avec le même compte sont associées au même partage de ressources.	6 mai 2022
Lake Formation a commencé à suivre les changements.	Lake Formation a commencé à suivre les modifications apportées AWS à ses politiques gérées.	6 mai 2022

Autorisations suggérées par Persona

Les autorisations suggérées pour chaque personnage sont les suivantes. L'administrateur IAM n'est pas inclus car il dispose de toutes les autorisations sur toutes les ressources.

Rubriques

- [Autorisations d'administrateur du lac de données](#)
- [Autorisations d'administrateur en lecture seule](#)
- [Autorisations d'ingénieur de données](#)
- [Autorisations d'analyse de données](#)
- [Autorisations relatives aux rôles du flux](#)

Autorisations d'administrateur du lac de données

⚠ Important

Dans les politiques suivantes, remplacez *<account-id>* par un numéro de AWS compte valide et remplacez *<workflow_role>* par le nom d'un rôle autorisé à exécuter un flux de travail, comme défini dans [Autorisations relatives aux rôles du flux](#).

Type de politique	Politique
AWS politiques gérées	• <code>AWSLakeFormationDataAdmin</code> • <code>LakeFormationDataAccessServiceRolePolicy</code> (politique des rôles liés au service) • <code>AWSGlueConsoleFullAccess</code> (facultatif) • <code>CloudWatchLogsReadOnlyAccess</code> (facultatif) • <code>AWSLakeFormationCrossAccountManager</code> (facultatif) • <code>AmazonAthenaFullAccess</code> (facultatif) Pour plus d'informations sur les politiques AWS gérées facultatives, consultez the section called “Création d'un administrateur de lac de données”.
Politique en ligne (pour créer le rôle lié au service Lake Formation)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": "iam:CreateServiceLinkedRole", "Resource": "*", "Condition": { "StringEquals": { "iam:AWSServiceName": "lakeformation.amazonaws.com" } } }] }</pre>

Type de politique	Politique
	<pre> } }, { "Effect": "Allow", "Action": ["iam:PutRolePolicy"], "Resource": "arn:aws:iam:: <account-id> :role/aws-service-role/lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess" }] } </pre>
(Facultatif) Politique intégrée (politique de passe-rôle pour le rôle de flux de travail). Cela n'est nécessaire que si l'administrateur du lac de données crée et exécute des flux de travail.	<pre> { "Version": "2012-10-17", "Statement": [{ "Sid": "PassRolePermissions", "Effect": "Allow", "Action": ["iam:PassRole"], "Resource": ["arn:aws:iam:: 111122223333 :role/<workflow_role> "] }] } </pre>

Type de politique	Politique
(Facultatif) Politique en ligne (si votre compte accorde ou reçoit des autorisations entre comptes Lake Formation). Cette politique permet d'accepter ou de rejeter les invitations à partager des AWS RAM ressources et d'autoriser l'octroi d'autorisations entre comptes aux organisations. <code>ram:EnableSharingWithAwsOrganization</code> est obligatoire uniquement pour les administrateurs de data lake dans le compte AWS Organisations de gestion.	<pre> { "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["ram:AcceptResourceShareInvitation", "ram:RejectResourceShareInvitation", "ec2:DescribeAvailabilityZones", "ram:EnableSharingWithAwsOrganization"], "Resource": "*" }] }</pre>

Autorisations d'administrateur en lecture seule

Type de politique	Politique
Politique en ligne (de base)	<pre> { "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetEffectivePermissionsForPath", "lakeformation:ListPermissions", "lakeformation:ListDataCellsFilter", "lakeformation:GetDataCellsFilter", "lakeformation:SearchDatabasesByLFTags", </pre>

Type de politique	Politique
	<pre> "lakeformation:SearchTablesByLFTags" , "lakeformation:GetLFTag", "lakeformation:ListLFTags", "lakeformation:GetResourceLFTags", "lakeformation:ListLakeFormationOpti ns", "cloudtrail:DescribeTrails", "cloudtrail:LookupEvents", "glue:GetDatabase", "glue:GetDatabases", "glue:GetConnections", "glue:SearchTables", "glue:GetTable", "glue:GetTableVersions", "glue:GetPartitions", "glue:GetTables", "glue:GetWorkflow", "glue:ListWorkflows", "glue:BatchGetWorkflows", "glue:GetWorkflowRuns", "glue:GetWorkflow", "s3:ListBucket", "s3:GetBucketLocation", "s3:ListAllMyBuckets", "s3:GetBucketAcl", "iam:ListUsers", "iam:ListRoles", "iam:GetRole", "iam:GetRolePolicy"], "Resource": "*" }, { "Effect": "Deny", "Action": ["lakeformation:PutDataLakeSettings"], "Resource": "*" }] </pre>

Type de politique	Politique
	<pre>}</pre>

Autorisations d'ingénieur de données

 Important

Dans les politiques suivantes, remplacez *<account-id>* par un numéro de AWS compte valide et remplacez *<workflow_role>* par le nom du rôle du flux de travail.

Type de politique	Politique
AWS politique gérée	AWSGlueConsoleFullAccess
Politique en ligne (de base)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions", "lakeformation:RevokePermissions", "lakeformation:BatchGrantPermissions", "lakeformation:BatchRevokePermissions", "lakeformation:ListPermissions", "lakeformation:AddLFTagsToResource", "lakeformation:RemoveLFTagsFromResource", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags"] }] }</pre>

Type de politique	Politique
	<pre>"lakeformation:SearchDatabasesByLFTa gs", "lakeformation:GetWorkUnits", "lakeformation:GetWorkUnitResults", "lakeformation:StartQueryPlanning", "lakeformation:GetQueryState", "lakeformation:GetQueryStatistics"], "Resource": "*" }]</pre>

Politique intégrée (pour les opérations sur les tables gouvernées, y compris les opérations au sein des transactions)

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:StartTransaction",
        "lakeformation:CommitTransaction",
        "lakeformation:CancelTransaction",
        "lakeformation:ExtendTransaction",
        "lakeformation:DescribeTransaction",
        "lakeformation:ListTransactions",
        "lakeformation:GetTableObjects",
        "lakeformation:UpdateTableObjects",
        "lakeformation>DeleteObjectsOnCancel
"
      ],
      "Resource": "*"
    }
  ]
}
```

Type de politique	Politique
Politique en ligne (pour le contrôle d'accès aux métadonnées à l'aide de la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC))	<pre> { "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:AddLFTagsToResource", "lakeformation:RemoveLFTagsFromResource", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags"], "Resource": "*" }] }</pre>
Politique intégrée (politique de passe-rôle pour le rôle de flux de travail)	<pre> { "Version": "2012-10-17", "Statement": [{ "Sid": "PassRolePermissions", "Effect": "Allow", "Action": ["iam:PassRole"], "Resource": ["arn:aws:iam:: 111122223333 :role/<workflow_role> "] }] }</pre>

Autorisations d'analyse de données

Type de politique	Politique
AWS politique gérée	AmazonAthenaFullAccess
Politique en ligne (de base)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "glue:GetTable", "glue:GetTables", "glue:SearchTables", "glue:GetDatabase", "glue:GetDatabases", "glue:GetPartitions", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags"], "Resource": "*" }] }</pre>

(Facultatif) Politique intégrée (pour les opérations sur les tables gouvernées, y compris les opérations au sein des transactions)

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:StartTransaction",
        "lakeformation:CommitTransaction",
        "lakeformation:CancelTransaction",
        "lakeformation:ExtendTransaction",
```

Type de politique	Politique
	<pre> "lakeformation:DescribeTransaction", "lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation>DeleteObjectsOnCancel "], "Resource": "*" }] } </pre>

Autorisations relatives aux rôles du flux

Ce rôle dispose des autorisations requises pour exécuter un flux de travail. Vous spécifiez un rôle doté de ces autorisations lorsque vous créez un flux de travail.

Important

Dans les politiques suivantes, remplacez-le *<region>* par un identifiant de AWS région valide (par exemple *us-east-1*), *<account-id>* par un numéro de AWS compte valide, *<workflow_role>* par le nom du rôle du flux de travail et *<your-s3-cloudtrail-bucket>* par le chemin Amazon S3 vers vos AWS CloudTrail journaux.

Type de politique	Politique
AWS politique gérée	AWSGlueServiceRole
Politique en ligne (accès aux données)	<pre> { "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", </pre>

Type de politique	Politique
	<pre> "lakeformation:GrantPermissions"], "Resource": "*" }] }</pre>

Politique intégrée (politique de passe-rôle pour le rôle de flux de travail)

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "PassRolePermissions",
      "Effect": "Allow",
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
        "arn:aws:iam:: 111122223333 :role/<workflow_role> "
      ]
    }
  ]
}
```

Politique en ligne (pour l'ingestion de données en dehors du lac de données, AWS CloudTrail par exemple des journaux)

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": ["s3:GetObject", "s3:ListBucket"],
      "Resource": ["arn:aws:s3::: <your-s3-cloudtrail-bucket> /*"]
    }
  ]
}
```


Modification des paramètres par défaut de votre lac de données

Pour maintenir la rétrocompatibilité avec AWS Glue, AWS Lake Formation possède les paramètres de sécurité initiaux suivants :

- L'`Super`autorisation est accordée au groupe `IAMAllowedPrincipals` sur toutes les ressources du catalogue de AWS Glue données existantes.
- Les paramètres « Utiliser uniquement le contrôle d'accès IAM » sont activés pour les nouvelles ressources du catalogue de données.

Ces paramètres font en sorte que l'accès aux ressources du catalogue de données et aux emplacements Amazon S3 soit contrôlé uniquement par des politiques Gestion des identités et des accès AWS (IAM). Les autorisations individuelles de Lake Formation ne sont pas en vigueur.

Le `IAMAllowedPrincipals` groupe inclut tous les utilisateurs et rôles IAM autorisés à accéder aux ressources de votre catalogue de données par vos politiques IAM. L'`Super`autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée.

Pour modifier les paramètres de sécurité afin que l'accès aux ressources du catalogue de données (bases de données et tables) soit géré par les autorisations de Lake Formation, procédez comme suit :

1. Modifiez les paramètres de sécurité par défaut pour les nouvelles ressources. Pour obtenir des instructions, veuillez consulter [Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride](#).
2. Modifiez les paramètres des ressources du catalogue de données existantes. Pour obtenir des instructions, veuillez consulter [Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle](#).

Modification des paramètres de sécurité par défaut à l'aide de l'opération d'`PutDataLakeSettings` API Lake Formation

Vous pouvez également modifier les paramètres de sécurité par défaut à l'aide de l'opération d'`PutDataLakeSettings` API Lake Formation. Cette action prend comme arguments un ID de catalogue facultatif et une `DataLakeSettings` structure.

Pour renforcer le contrôle d'accès aux métadonnées et aux données sous-jacentes par Lake Formation sur les nouvelles bases de données et tables, codez la `DataLakeSettings` structure comme suit.

 Note

<AccountID> Remplacez-le par un identifiant de AWS compte valide et *<Username>* par un nom d'utilisateur IAM valide. Vous pouvez définir plusieurs utilisateurs en tant qu'administrateurs de data lake.

```
{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier":
"arn:aws:iam::<AccountID>:user/<Username>"
      }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": []
  }
}
```

Vous pouvez également coder la structure comme suit. L'omission du `CreateTableDefaultPermissions` paramètre `CreateDatabaseDefaultPermissions` or équivaut à transmettre une liste vide.

```
{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier":
"arn:aws:iam::<AccountID>:user/<Username>"
      }
    ]
  }
}
```

Cette action révoque effectivement toutes les autorisations de Lake Formation accordées au IAMAllowedPrincipals groupe sur les nouvelles bases de données et tables. Lorsque vous créez une base de données, vous pouvez annuler ce paramètre.

Pour appliquer le contrôle d'accès aux métadonnées et aux données sous-jacentes uniquement par IAM sur les nouvelles bases de données et tables, codez la DataLakeSettings structure comme suit.

```
{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier":
"arn:aws:iam::<AccountId>:user/<Username>"
      }
    ],
    "CreateDatabaseDefaultPermissions": [
      {
        "Principal": {
          "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
        },
        "Permissions": [
          "ALL"
        ]
      }
    ],
    "CreateTableDefaultPermissions": [
      {
        "Principal": {
          "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
        },
        "Permissions": [
          "ALL"
        ]
      }
    ]
  }
}
```

Cela permet à la Super Lake Formation d'autoriser le IAMAllowedPrincipals groupe à utiliser de nouvelles bases de données et de nouvelles tables. Lorsque vous créez une base de données, vous pouvez annuler ce paramètre.

 Note

Dans la `DataLakeSettings` structure précédente, la seule valeur autorisée pour `DataLakePrincipalIdentifier` est `IAM_ALLOWED_PRINCIPALS`, et la seule valeur autorisée pour `Permissions` est `ALL`.

Permissions implicites de Lake Formation

AWS Lake Formation accorde les autorisations implicites suivantes aux administrateurs de lacs de données, aux créateurs de bases de données et aux créateurs de tables.

Administrateurs de lacs de données

- `Describe` Accédez à toutes les ressources du catalogue de données, à l'exception des ressources partagées directement depuis un autre compte avec un autre principal. Cet accès ne peut pas être révoqué auprès d'un administrateur.
- Disposez d'autorisations de localisation des données partout dans le lac de données.
- Peut accorder ou révoquer l'accès à toutes les ressources du catalogue de données à n'importe quel principal (y compris lui-même). Cet accès ne peut pas être révoqué auprès d'un administrateur.
- Peut créer des bases de données dans le catalogue de données.
- Peut accorder l'autorisation de créer une base de données à un autre utilisateur.

 Note

Les administrateurs de data lake peuvent enregistrer des sites Amazon S3 uniquement s'ils disposent des autorisations IAM pour le faire. Les politiques proposées aux administrateurs des lacs de données dans ce guide accordent ces autorisations. En outre, les administrateurs des lacs de données ne disposent pas des autorisations implicites leur permettant de supprimer des bases de données ou alter/drop des tables créées par d'autres. Ils peuvent toutefois s'octroyer l'autorisation de le faire.

Pour plus d'informations sur les administrateurs de data lake, consultez [Création d'un administrateur de lac de données](#).

Créateurs de catalogues

- Disposez de toutes les autorisations de catalogue sur les catalogues qu'ils créent, d'autorisations sur les bases de données et les tables qu'ils créent dans le catalogue, et vous pouvez accorder aux autres principaux du même AWS compte l'autorisation de créer des bases de données et des tables dans le catalogue. Un créateur de catalogue qui dispose également de la politique `AWSLakeFormationCrossAccountManager` AWS gérée peut accorder des autorisations sur le catalogue à d'autres AWS comptes ou organisations.

Les administrateurs de data lake peuvent utiliser la console ou l'API Lake Formation pour désigner les créateurs de catalogues.

Note

Les créateurs de catalogues ne disposent pas implicitement d'autorisations sur les bases de données et les tables créées par d'autres utilisateurs dans le catalogue.

Pour plus d'informations sur la création de catalogues, consultez [Intégrer vos données dans AWS Glue Data Catalog](#).

créateurs de bases de données

- Disposer de toutes les autorisations de base de données sur les bases de données qu'ils créent, d'autorisations sur les tables qu'ils créent dans la base de données et peuvent accorder aux autres principaux du même AWS compte l'autorisation de créer des tables dans la base de données. Un créateur de base de données qui dispose également de la politique `AWSLakeFormationCrossAccountManager` AWS gérée peut accorder des autorisations sur la base de données à d'autres AWS comptes ou organisations.

Les administrateurs de data lake peuvent utiliser la console ou l'API Lake Formation pour désigner les créateurs de bases de données.

Note

Les créateurs de base de données ne disposent pas implicitement d'autorisations sur les tables créées par d'autres utilisateurs dans la base de données.

Pour de plus amples informations, veuillez consulter [Création d'une base de données](#).

Créateurs de tables

- Disposent de toutes les autorisations sur les tables qu'ils créent.
- Ils peuvent accorder des autorisations sur toutes les tables qu'ils créent aux principaux d'un même AWS compte.
- Ils peuvent accorder des autorisations sur toutes les tables qu'ils créent à d'autres AWS comptes ou organisations s'ils disposent de la politique `AWSLakeFormationCrossAccountManager` AWS gérée.
- Peut afficher les bases de données qui contiennent les tables qu'ils créent.

Référence des autorisations de Lake Formation

Pour effectuer des AWS Lake Formation opérations, les directeurs ont besoin à la fois des autorisations Lake Formation et des autorisations Gestion des identités et des accès AWS (IAM). Vous accordez généralement des autorisations IAM à l'aide de politiques de contrôle d'accès grossières, comme décrit dans [the section called “Vue d'ensemble des autorisations relatives à Lake Formation”](#) Vous pouvez accorder des autorisations à Lake Formation à l'aide de la console, de l'API ou du AWS Command Line Interface (AWS CLI).

Pour savoir comment accorder ou révoquer les autorisations de Lake Formation, consultez [the section called “Octroi d'autorisations de données”](#) et [the section called “Octroi d'autorisations de localisation des données”](#).

Note

Les exemples de cette section montrent comment accorder des autorisations aux principaux d'un même AWS compte. Pour des exemples de subventions entre comptes, voir [the section called “Partage de données entre comptes”](#).

Permissions de Lake Formation par type de ressource

Voici les autorisations valides de Lake Formation disponibles pour chaque type de ressource :

Ressource	Autorisations
Catalog	ALL(Super), Super utilisateur

Ressource	Autorisations	
	ALTER	
	CREATE_DATABASE	
	DESCRIBE	
	DROP	
Database	ALL (Super)	
	ALTER	
	CREATE_TABLE	
	DESCRIBE	
	DROP	
Table	ALL (Super)	
	ALTER	
	DELETE	
	DESCRIBE	
	DROP	
	INSERT	
	SELECT	
View	ALL (Super)	
	SELECT	
	DESCRIBE	
	DROP	

Ressource	Autorisations	
Data Catalog	CREATE_DATABASE	
Amazon S3 location	DATA_LOCATION_ACCESS	
LF-Tags	DROP	
	ALTER	
LF-Tag values	ASSOCIATE	
	DESCRIBE	
	GrantWithLFTagExpression	
LF-Tag policy - Database	ALL (Super)	
	ALTER	
	CREATE_TABLE	
	DESCRIBE	
	DROP	
LF-Tag policy - Table	ALL (Super)	
	ALTER	
	DESCRIBE	
	DELETE	
	DROP	
	INSERT	
	SELECT	

Ressource	Autorisations	
Resource link - Database or Table	DESCRIBE	
	DROP	
Table with data filters	DESCRIBE	
	DROP	
	SELECT	
Table with column filter	SELECT	

Rubriques

- [Lake Formation accorde et AWS CLI révoque des commandes](#)
- [Autorisations Lake Formation](#)

Lake Formation accorde et AWS CLI révoque des commandes

Chaque description d'autorisation présentée dans cette section inclut des exemples d'octroi de l'autorisation à l'aide d'une AWS CLI commande. Voici les synopses de la Lake Formation grant-permissions et revoke-permissions AWS CLI des commandes.

```
grant-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

```
revoke-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
```

```
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

Pour une description détaillée de ces commandes, voir [grant-permissions et revoke-permissions](#) dans la référence des commandes.AWS CLI Cette section fournit des informations supplémentaires sur `--principal` cette option.

La valeur de l'`--principal` option est l'une des suivantes :

- Amazon Resource Name (ARN) pour un utilisateur ou un rôle Gestion des identités et des accès AWS (IAM)
- ARN pour un utilisateur ou un groupe qui s'authentifie via un fournisseur SAML, tel que Microsoft Active Directory Federation Service (AD FS)
- ARN pour un utilisateur ou un groupe Amazon Quick
- Pour les autorisations entre comptes, un identifiant de AWS compte, un identifiant d'organisation ou un identifiant d'unité organisationnelle
- Pour l'utilisateur ou le groupe IAM Identity Center, l'ARN de l'utilisateur ou du groupe IAM Identity Center.

Vous trouverez ci-dessous la syntaxe et des exemples pour tous les `--principal` types.

Le principal est un utilisateur IAM

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

Le directeur est un rôle IAM

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

Le principal est un utilisateur s'authentifiant via un fournisseur SAML

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-provider/<SAMLproviderName>:user/<user-name>
```

Exemples :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/idp1:user/datalake_user1
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/AthenaLakeFormation0kta:user/athena-user@example.com
```

Le principal est un groupe s'authentifiant via un fournisseur SAML

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-provider/<SAMLproviderName>:group/<group-name>
```

Exemples :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/idp1:group/data-scientists
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/AthenaLakeFormation0kta:group/my-group
```

Le principal est un utilisateur d'Amazon Quick Enterprise Edition

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-id>:user/<namespace>/<user-name>
```

 Note

Pour *<namespace>*, vous devez spécifier default.

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-east-1:111122223333:user/default/bi_user1
```

Principal est un groupe Amazon Quick Enterprise Edition

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-id>:group/<namespace>/<group-name>
```

 Note

Pour *<namespace>*, vous devez spécifier default.

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-east-1:111122223333:group/default/data_scientists
```

Le principal est un AWS compte

Syntaxe :

```
--principal DataLakePrincipalIdentifier=<account-id>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=111122223333
```

Le principal est une organisation

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-id>:organization/<organization-id>
```

Exemple :

```
--principal
DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl
```

Le directeur est une unité organisationnelle

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-id>:ou/<organization-id>/<organizational-unit-id>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-
abcdefghijkl/ou-ab00-cdefghij
```

Le principal est un utilisateur ou un groupe d'identités IAM Identity Center

Exemple : utilisateur

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

Exemple : Groupe :

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

Le principal est un groupe IAM - **IAMAllowedPrincipals**

Lake Formation attribue des Super autorisations sur toutes les bases de données et tables du catalogue de données à un groupe appelé IAMAllowedPrincipals par défaut. Si cette autorisation de groupe existe sur une base de données ou une table, tous les principaux de votre compte auront accès à la ressource par le biais des politiques principales IAM pour. AWS Glue II assure une rétrocompatibilité lorsque vous commencez à utiliser les autorisations de

Lake Formation pour sécuriser les ressources du catalogue de données qui étaient auparavant protégées par les politiques IAM pour AWS Glue.

Lorsque vous utilisez Lake Formation pour gérer les autorisations pour les ressources de votre catalogue de données, vous devez d'abord révoquer l'`IAMAllowedPrincipals` autorisation sur les ressources, ou activer les principes et les ressources en mode d'accès hybride pour que les autorisations de Lake Formation fonctionnent.

Exemple :

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

Le principal est un groupe IAM - **ALLIAMPrincipals**

Lorsque vous autorisez un `ALLIAMPrincipals` groupe sur une ressource de catalogue de données, chaque principal du compte a accès à la ressource de catalogue de données à l'aide des autorisations Lake Formation et des autorisations IAM.

Exemple :

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Autorisations Lake Formation

Cette section contient les autorisations de Lake Formation disponibles que vous pouvez accorder aux directeurs.

ALTER

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
ALTER	DATABASE	glue:UpdateDatabase
ALTER	TABLE	glue:UpdateTable
ALTER	LF-Tag	lakeformation:UpdateLFTag

Un directeur disposant de cette autorisation peut modifier les métadonnées d'une base de données ou d'une table dans le catalogue de données. Pour les tables, vous pouvez modifier le schéma des colonnes et ajouter des paramètres de colonne. Vous ne pouvez pas modifier les colonnes des données sous-jacentes vers lesquelles pointe une table de métadonnées.

Si la propriété modifiée est un emplacement enregistré auprès d'Amazon Simple Storage Service (Amazon S3), le principal doit disposer des autorisations de localisation des données sur le nouvel emplacement.

Exemple

L'exemple suivant accorde l'ALTER autorisation à l'utilisateur `datalake_user1` sur la base de données `retail` dans le AWS compte 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"} }'
```

Exemple

L'exemple suivant accorde ALTER à l'utilisateur le droit d'`datalake_user1` accéder à la table `inventory` de la base de données `retail`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"} }'
```

CREATE_DATABASE

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
CREATE_DATABASE	Catalogue de données	glue:CreateDatabase

Un directeur disposant de cette autorisation peut créer une base de données de métadonnées ou un lien de ressource dans le catalogue de données. Le principal peut également créer des tables dans la base de données.

Exemple

L'exemple suivant octroie `CREATE_DATABASE` à l'utilisateur `datalake_user1` du AWS compte `1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {} }'
```

Lorsqu'un principal crée une base de données dans le catalogue de données, aucune autorisation sur les données sous-jacentes n'est accordée. Les autorisations de métadonnées supplémentaires suivantes sont accordées (ainsi que la possibilité d'accorder ces autorisations à d'autres personnes) :

- `CREATE_TABLE` dans la base de données
- Base de données `ALTER`
- Base de données `DROP`

Lors de la création d'une base de données, le principal peut éventuellement spécifier un emplacement Amazon S3. Selon que le principal dispose ou non d'autorisations de localisation des données, celles-ci `CREATE_DATABASE` peuvent ne pas être suffisantes pour créer des bases de données dans tous les cas. Il est important de garder à l'esprit les trois cas suivants.

Créer un cas d'utilisation de base de données	Autorisations nécessaires
La propriété de localisation n'est pas spécifiée.	<code>CREATE_DATABASE</code> est suffisant.
La propriété de localisation est spécifiée, et l'emplacement n'est pas géré par Lake Formation (il n'est pas enregistré).	<code>CREATE_DATABASE</code> est suffisant.
La propriété de localisation est spécifiée, et l'emplacement est géré par Lake Formation (est enregistré).	<code>CREATE_DATABASE</code> est obligatoire, ainsi que les autorisations de localisation des données à l'emplacement spécifié.

CREATE_TABLE

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
CREATE_TABLE	DATABASE	glue:CreateTable

Un directeur disposant de cette autorisation peut créer une table de métadonnées ou un lien de ressource dans le catalogue de données de la base de données spécifiée.

Exemple

L'exemple suivant accorde à l'utilisateur l'`datalake_user1` autorisation de créer des tables dans la `retail` base de données sous le AWS compte 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"} }'
```

Lorsqu'un directeur crée une table dans le catalogue de données, toutes les autorisations de Lake Formation sur la table sont accordées au principal, avec la possibilité d'accorder ces autorisations à d'autres personnes.

Subventions entre comptes

Si un compte propriétaire de base de données accorde une autorisation `CREATE_TABLE` à un compte destinataire et qu'un utilisateur du compte destinataire crée avec succès une table dans la base de données du compte propriétaire, les règles suivantes s'appliquent :

- L'utilisateur et les administrateurs du lac de données du compte destinataire disposent de toutes les autorisations relatives à Lake Formation. Ils peuvent accorder des autorisations sur la table aux autres principaux de leur compte. Ils ne peuvent pas accorder d'autorisations aux principaux sur le compte du propriétaire ou sur tout autre compte.
- Les administrateurs du lac de données du compte propriétaire peuvent accorder des autorisations sur la table aux autres principaux de leur compte.

Autorisations de localisation des données

Lorsque vous tentez de créer une table pointant vers un emplacement Amazon S3, selon que vous disposez ou non des autorisations de localisation des données, ces `CREATE_TABLE` autorisations peuvent ne pas être suffisantes pour créer une table. Il est important de garder à l'esprit les trois cas suivants.

Création d'un cas d'utilisation d'une table	Autorisations nécessaires
L'emplacement spécifié n'est pas géré par Lake Formation (il n'est pas enregistré).	<code>CREATE_TABLE</code> est suffisant.
L'emplacement spécifié est géré par Lake Formation (il est enregistré), et la base de données qui le contient ne possède aucune propriété de localisation ou possède une propriété de localisation qui n'est pas un préfixe Amazon S3 de l'emplacement de la table.	<code>CREATE_TABLE</code> est obligatoire, ainsi que les autorisations de localisation des données à l'emplacement spécifié.
L'emplacement spécifié est géré par Lake Formation (il est enregistré), et la base de données contenant possède une propriété de localisation qui pointe vers un emplacement enregistré et qui est un préfixe Amazon S3 de l'emplacement de la table.	<code>CREATE_TABLE</code> est suffisant.

DATA_LOCATION_ACCESS

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
<code>DATA_LOCATION_ACCESS</code>	Emplacement Amazon S3	(Autorisations Amazon S3 sur l'emplacement, qui doivent être spécifiées par le rôle utilisé pour enregistrer l'emplacement.)

Il s'agit de la seule autorisation de localisation des données. Un principal disposant de cette autorisation peut créer une base de données ou une table de métadonnées pointant vers l'emplacement Amazon S3 spécifié. L'emplacement doit être enregistré. Un directeur qui dispose d'autorisations de localisation des données sur un emplacement possède également des autorisations de localisation sur les emplacements des enfants.

Exemple

L'exemple suivant accorde des autorisations de localisation des données `s3://products/retail` à l'utilisateur du AWS compte `datalake_user1 1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
{"ResourceArn":"arn:aws:s3::products/retail"} }'
```

`DATA_LOCATION_ACCESS` n'est pas nécessaire pour interroger ou mettre à jour les données sous-jacentes. Cette autorisation s'applique uniquement à la création de ressources de catalogue de données.

Pour plus d'informations sur les autorisations de localisation des données, consultez [Underlying data access control](#).

DELETE

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
DELETE	TABLE	(Aucune autorisation IAM supplémentaire n'est requise si l'emplacement est enregistré.)

Un principal disposant de cette autorisation peut insérer, mettre à jour et lire les données sous-jacentes à l'emplacement Amazon S3 indiqué dans le tableau. Le directeur peut également consulter le tableau sur la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS Glue API.

Exemple

L'exemple suivant accorde l'`DELETE` autorisation à l'utilisateur `datalake_user1` sur la table de la base de données `inventory` dans le AWS compte `retail 1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"}}'
```

Cette autorisation s'applique uniquement aux données d'Amazon S3, et non aux données d'autres magasins de données tels qu'Amazon Relational Database Service (Amazon RDS).

DESCRIBE

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
DESCRIBE	Lien vers les ressources du tableau	glue:GetTable
	Lien vers les ressources de la base	glue:GetDatabase
DESCRIBE	DATABASE	glue:GetDatabase
DESCRIBE	TABLE	glue:GetTable
DESCRIBE	LF-Tag	glue:GetTable
		glue:GetDatabase
		lakeformation:GetResourceLFTags
		lakeformation:ListLFTags
		lakeformation:GetLFTag

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
		lakeformation:SearchTablesByLFTags lakeformation:SearchDatabasesByLFTags

Un principal disposant de cette autorisation peut consulter la base de données, la table ou le lien de ressource spécifié. Aucune autre autorisation de catalogue de données n'est accordée implicitement, et aucune autorisation d'accès aux données n'est accordée implicitement. Les bases de données et les tables apparaissent dans les éditeurs de requêtes des services intégrés, mais aucune requête ne peut être effectuée à leur encontre à moins que d'autres autorisations de Lake Formation (par exemple, SELECT) ne soient accordées.

Par exemple, un utilisateur qui possède une DESCRIBE base de données peut voir la base de données et toutes les métadonnées de la base de données (description, emplacement, etc.). Toutefois, l'utilisateur ne peut pas savoir quelles tables contient la base de données et ne peut pas supprimer, modifier ou créer des tables dans la base de données. De même, un utilisateur qui possède DESCRIBE une table peut voir la table et ses métadonnées (description, schéma, emplacement, etc.), mais ne peut pas supprimer, modifier ou exécuter des requêtes sur la table.

Voici quelques règles supplémentaires pour DESCRIBE :

- Si un utilisateur dispose d'autres autorisations Lake Formation sur une base de données, une table ou un lien de ressource, elles DESCRIBE sont implicitement accordées.
- Si un utilisateur ne dispose SELECT que d'un sous-ensemble de colonnes pour un tableau (partielSELECT), il est limité à l'affichage de ces colonnes.
- Vous ne pouvez pas accorder d'autorisation DESCRIBE à un utilisateur qui a effectué une sélection partielle sur une table. À l'inverse, vous ne pouvez pas spécifier de listes d'inclusion ou d'exclusion de colonnes pour les DESCRIBE tables accordées sur.

Exemple

L'exemple suivant accorde l'DESCRIBE autorisation à l'utilisateur `datalake_user1` sur le lien de ressource de la table dans la base de données `inventory-link` dans le AWS compte `retail-1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory-link"}}'
```

DROP

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
DROP	DATABASE	glue:DeleteDatabase
DROP	TABLE	glue:DeleteTable
DROP	LF-Tag	lakeformation:DeleteLFTag
DROP	Lien vers les ressources de la base Lien vers les ressources du tableau	glue:DeleteDatabase glue:DeleteTable

Un principal disposant de cette autorisation peut supprimer une base de données, une table ou un lien de ressource dans le catalogue de données. Vous ne pouvez pas accorder l'autorisation DROP sur une base de données à un compte ou à une organisation externe.

 **Warning**

La suppression d'une base de données entraîne la suppression de toutes les tables de la base de données.

Exemple

L'exemple suivant accorde l'DROP autorisation à l'utilisateur `datalake_user1` sur la base de données `retail` dans le AWS compte `1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
  DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Exemple

L'exemple suivant accorde des DROP autorisations à l'utilisateur `datalake_user1` sur la table de `inventory` la base de données `retail`.

```
aws lakeformation grant-permissions --principal
  DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"}}'
```

Exemple

L'exemple suivant accorde DROP à l'utilisateur `datalake_user1` sur la table `un lien de ressource inventory-link` dans la base de données `retail`.

```
aws lakeformation grant-permissions --principal
  DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
INSERT	TABLE	(Aucune autorisation IAM supplémentaire n'est requise si l'emplacement est enregistré.)

Un principal disposant de cette autorisation peut insérer, mettre à jour et lire les données sous-jacentes à l'emplacement Amazon S3 indiqué dans le tableau. Le directeur peut également consulter le tableau dans la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI.

Exemple

L'exemple suivant accorde l'INSERT autorisation à l'utilisateur `datalake_user1` sur la table de la base de données `inventory` dans le AWS compte `retail 1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"}}'
```

Cette autorisation s'applique uniquement aux données d'Amazon S3, et non aux données d'autres magasins de données tels qu'Amazon RDS.

SELECT

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
SELECT	• TABLE	(Aucune autorisation IAM supplémentaire n'est requise si l'emplacement est enregistré.)

Un directeur disposant de cette autorisation peut consulter une table dans le catalogue de données et interroger les données sous-jacentes dans Amazon S3 à l'emplacement spécifié par la table. Le directeur peut consulter le tableau dans la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI. Si le filtrage des colonnes a été appliqué lorsque cette autorisation a été accordée, le principal peut consulter les métadonnées uniquement pour les colonnes incluses et ne peut interroger les données qu'à partir des colonnes incluses.

 Note

Il est de la responsabilité du service d'analyse intégré d'appliquer le filtrage des colonnes lors du traitement d'une requête.

Exemple

L'exemple suivant accorde l'SELECT autorisation à l'utilisateur `datalake_user1` sur la table de la base de données `inventory` dans le AWS compte `retail 1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"}}'
```

Cette autorisation s'applique uniquement aux données d'Amazon S3, et non aux données d'autres magasins de données tels qu'Amazon RDS.

Vous pouvez filtrer (restreindre l'accès à) des colonnes spécifiques à l'aide d'une liste d'inclusion ou d'exclusion facultative. Une liste d'inclusion indique les colonnes accessibles. Une liste d'exclusion indique les colonnes inaccessibles. En l'absence de liste d'inclusion ou d'exclusion, toutes les colonnes du tableau sont accessibles.

Les résultats de `glue:GetTable` renvoient uniquement les colonnes que l'appelant est autorisé à consulter. Les services intégrés tels qu'Amazon Athena et Amazon Redshift respectent les listes d'inclusion et d'exclusion des colonnes.

Exemple

L'exemple suivant octroie une SELECT autorisation à l'utilisateur figurant `datalake_user1` sur la table `inventory` à l'aide d'une liste d'inclusion.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"retail",
"Name":"inventory", "ColumnNames": ["prodcode","location","period","withdrawals"]}}'
```

Exemple

L'exemple suivant donne des autorisations SELECT sur la inventory table à l'aide d'une liste d'exclusion.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"retail",
"Name":"inventory", "ColumnWildcard": {"ExcludedColumnNames": ["intkey",
"prodcode"]}}}'
```

Les restrictions suivantes s'appliquent à l'SELECT autorisation :

- Lors de l'octroi SELECT, vous ne pouvez pas inclure l'option d'octroi si le filtrage des colonnes est appliqué.
- Vous ne pouvez pas restreindre le contrôle d'accès aux colonnes qui sont des clés de partition.
- Un principal disposant de l'SELECT autorisation sur un sous-ensemble de colonnes d'une table ne peut pas obtenir l'INSERT autorisation ALTER DROP, DELETE, ou sur cette table. De même, un principal disposant de l'INSERT autorisation ALTER DROP DELETE,, ou sur une table ne peut pas SELECT obtenir l'autorisation de filtrage des colonnes.

L'SELECT autorisation apparaît toujours sur la page Autorisations relatives aux données de la console Lake Formation sous forme de ligne séparée. L'image suivante montre que cela SELECT est accordé aux utilisateurs datalake_user2 et datalake_user3 sur toutes les colonnes du inventory tableau.

Data permissions (8)						
Choose a database or table for which to review, grant or revoke user permissions.						
Find by properties < 1 > ⚙						
Database: retail X Table: inventory X Clear filter						
	Principal	Principal type	Resource type	Resource	Owner account ID	Permissions
☐	datalake_user3	IAM user	Table	inventory	111122223333	Insert
☐	datalake_user3	IAM user	Column	retail.inventory.*	111122223333	Select
☐	datalake_user2	AD user	Table	inventory	111122223333	Delete, Insert
☐	datalake_user2	AD user	Column	retail.inventory.*	111122223333	Select

Super

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
Super	DATABASE	glue:*Database*
Super	TABLE	glue:*Table*, glue:*Partition*

Cette autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table. Vous ne pouvez pas attribuer une Super base de données à un compte externe.

Cette autorisation peut coexister avec les autres autorisations de Lake Formation. Par exemple, vous pouvez accorder les INSERT autorisations SuperSELECT, et sur une table de métadonnées. Le principal peut ensuite effectuer toutes les opérations prises en charge sur la table. Lorsque vous révoquez Super, les INSERT autorisations SELECT et sont conservées et le principal ne peut effectuer que des opérations de sélection et d'insertion.

Au lieu de l'octroyer Super à un directeur individuel, vous pouvez l'accorder au groupe IAMAllowedPrincipals. Le IAMAllowedPrincipals groupe est créé automatiquement et inclut tous les utilisateurs et rôles IAM autorisés à accéder aux ressources de votre catalogue de données par vos politiques IAM. Lorsqu'il Super est accordé IAMAllowedPrincipals pour une ressource de catalogue de données, l'accès à la ressource est effectivement contrôlé uniquement par les politiques IAM.

Vous pouvez Super obtenir l'autorisation d'accéder automatiquement IAMAllowedPrincipals aux nouvelles ressources du catalogue en tirant parti des options de la page Paramètres de la console Lake Formation.

Data catalog settings

Default permissions for newly created databases and tables

These settings maintain existing Data Catalog behavior. You can still set individual permissions on databases and tables, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

☐ Use only IAM access control for new databases

☐ Use only IAM access control for new tables in new databases

- Pour l'accorder Super à IAMAllowedPrincipals toutes les nouvelles bases de données, sélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données.
- IAMAllowedPrincipals Pour l'accorder Super à toutes les nouvelles tables des nouvelles bases de données, sélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables des nouvelles bases de données.

Note

Cette option entraîne la sélection par défaut de la case Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données dans la boîte de dialogue Créer une base de données. Il ne fait rien de plus que cela. C'est la case à cocher de la boîte de dialogue Créer une base de données qui permet d'accorder Super à IAMAllowedPrincipals.

Ces options de la page Paramètres sont activées par défaut. Pour plus d'informations, consultez les ressources suivantes :

- [the section called “Modification des paramètres par défaut de votre lac de données”](#)
- [the section called “Mise à niveau des autorisations de AWS Glue données pour le modèle Lake Formation”](#)

SUPER_USER

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
Super user	Catalog	glue:GetCatalog

Vous ne pouvez accorder l'Super user autorisation qu'à des entités spécifiques sur les catalogues du catalogue de données par défaut. Vous ne pouvez pas accorder d'Super user autorisation sur le catalogue par défaut ou sur d'autres types de ressources tels que les bases de données et les tables, ou aux principaux de comptes externes. L'Super user autorisation d'autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur les bases de données et les tables du catalogue accordé.

Avec cette `Super user` autorisation, le principal (bénéficiaire) est en mesure d'effectuer les actions suivantes sur les ressources (catalogues, bases de données et tables) du catalogue :

- `CREATE_DATABASE`, `DESCRIBE` autorisations sur le catalogue.
- `DROP`, `ALTER`, `CREATE_TABLE`, autorisations `DESCRIBE` (effectivement `SUPER`) sur toutes les bases de données du catalogue.
- `DROP`, `ALTER`, `DESCRIBE`, `SELECT`, `INSERT`, `DELETE` (effectivement `SUPER`) autorisations sur toutes les tables de toutes les bases de données du catalogue.
- All autorisations (en fait `SUPER`) sur les catalogues du catalogue.
- Autorisations accordables (possibilité d'accorder ces autorisations à d'autres principaux) sur tous les catalogues, bases de données et tables du catalogue.

Avec l'`Super user` autorisation sur une ressource du catalogue, le bénéficiaire n'est pas autorisé à effectuer ou à déléguer `ALTER` des `DROP` actions sur le catalogue.

ASSOCIATE

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
ASSOCIATE	LF-Tag	<code>glue:GetDatabase</code> <code>glue:GetTable</code> <code>lakeformation:AddLFTagsToResource"</code> <code>lakeformation:RemoveLFTagsFromResource"</code> <code>lakeformation:GetResourceLFTags</code> <code>lakeformation:ListLFTags</code> <code>lakeformation:GetLFTag</code>

Autorisations	Accordé sur cette ressource	Le bénéficiaire a également besoin
		lakeformation:SearchTablesByLFTags
		lakeformation:SearchDatabasesByLFTags

Un directeur disposant de cette autorisation sur une balise LF peut attribuer la balise LF à une ressource de catalogue de données. Accorder ASSOCIATE implicitement des subventions DESCRIBE.

Exemple

Cet exemple accorde à l'utilisateur `datalake_user1` l'ASSOCIATE autorisation d'utiliser le tag LF avec la clé. `module` Il autorise l'affichage et l'attribution de toutes les valeurs pour cette clé, comme indiqué par l'astérisque (*).

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}'
```

Intégration d'IAM Identity Center

Vous pouvez ainsi vous connecter à des fournisseurs d'identité (IdPs) et gérer de manière centralisée l'accès des utilisateurs et des groupes à travers les services AWS d'analyse. AWS IAM Identity Center Vous pouvez intégrer des fournisseurs d'identité tels qu'Okta, Ping et Microsoft Entra ID (anciennement Azure Active Directory) à IAM Identity Center pour que les utilisateurs de votre organisation puissent accéder aux données via une expérience d'authentification unique. IAM Identity Center prend également en charge la connexion à d'autres fournisseurs d'identité tiers.

Pour plus d'informations, reportez-vous à la section [Fournisseurs d'identité pris en charge](#) dans le Guide de AWS IAM Identity Center l'utilisateur.

Vous pouvez AWS Lake Formation le configurer en tant qu'application activée dans IAM Identity Center, et les administrateurs de data lake peuvent accorder des autorisations détaillées aux utilisateurs et aux groupes autorisés sur les ressources. AWS Glue Data Catalog

Les utilisateurs de votre organisation peuvent se connecter à n'importe quelle application compatible avec Identity Center en utilisant le fournisseur d'identité de votre organisation et interroger des ensembles de données en appliquant les autorisations de Lake Formation. Grâce à cette intégration, vous pouvez gérer l'accès aux AWS services sans créer plusieurs rôles IAM.

[La propagation fiable des identités](#) est une AWS IAM Identity Center fonctionnalité que les administrateurs de Connected Services AWS peuvent utiliser pour accorder et auditer l'accès aux données de service. L'accès à ces données est basé sur les attributs utilisateur tels que les associations de groupe. La mise en place d'une propagation d'identité sécurisée nécessite une collaboration entre les administrateurs de Connected Services AWS et les administrateurs d'IAM Identity Center. Pour plus d'informations, consultez [Prérequis et considérations](#).

Pour connaître les limitations, veuillez consulter [Limites de l'intégration d'IAM Identity Center](#).

Rubriques

- [Conditions préalables à l'intégration d'IAM Identity Center à Lake Formation](#)
- [Connecter Lake Formation à l'IAM Identity Center](#)
- [Mise à jour de l'intégration d'IAM Identity Center](#)
- [Suppression d'une connexion Lake Formation avec IAM Identity Center](#)
- [Octroi d'autorisations aux utilisateurs et aux groupes](#)
- [Intégration du contexte utilisateur d'IAM Identity Center dans les journaux CloudTrail](#)

Conditions préalables à l'intégration d'IAM Identity Center à Lake Formation

Les conditions préalables à l'intégration d'IAM Identity Center à Lake Formation sont les suivantes.

1. Activer le centre d'identité IAM : l'activation d'IAM Identity Center est une condition préalable à la prise en charge de l'authentification et de la propagation des identités.
2. Choisissez votre source d'identité : après avoir activé IAM Identity Center, vous devez disposer d'un fournisseur d'identité pour gérer les utilisateurs et les groupes. Vous pouvez utiliser le répertoire intégré Identity Center comme source d'identité ou utiliser un IdP externe, tel que Microsoft Entra ID ou Okta.

Pour plus d'informations, voir [Gérer votre source d'identité](#) et [Se connecter à un fournisseur d'identité externe](#) dans le guide de AWS IAM Identity Center l'utilisateur.

3. Créer un rôle IAM — Le rôle qui crée la connexion IAM Identity Center nécessite des autorisations pour créer et modifier la configuration de l'application dans Lake Formation et IAM Identity Center, conformément à la politique en ligne suivante.

Vous devez ajouter des autorisations conformément aux meilleures pratiques IAM. Les autorisations spécifiques sont détaillées dans les procédures qui suivent. Pour plus d'informations, consultez [Getting Started with IAM Identity Center](#).

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
        "sso:CreateApplication",
        "sso:PutApplicationAssignmentConfiguration",
        "sso:PutApplicationAuthenticationMethod",
        "sso:PutApplicationGrant",
        "sso:PutApplicationAccessScope"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

Si vous partagez des ressources du catalogue de données avec des organisations externes Comptes AWS ou avec des organisations, vous devez disposer des autorisations AWS Resource Access Manager (AWS RAM) pour créer des partages de ressources. Pour plus d'informations sur les autorisations requises pour partager des ressources, consultez la section [Conditions requises pour le partage de données entre comptes](#).

Les politiques intégrées suivantes contiennent les autorisations spécifiques requises pour afficher, mettre à jour et supprimer les propriétés de l'intégration de Lake Formation à IAM Identity Center.

- Utilisez la politique intégrée suivante pour autoriser un rôle IAM à visualiser une intégration de Lake Formation avec IAM Identity Center.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
        "sso:DescribeApplication"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

- Utilisez la politique intégrée suivante pour autoriser un rôle IAM à mettre à jour une intégration de Lake Formation avec IAM Identity Center. La politique inclut également les autorisations facultatives requises pour partager des ressources avec des comptes externes.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",
        "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
        "sso:DescribeApplication",
        "sso:UpdateApplication"
      ]
    }
  ]
}
```

```

        "Resource": [
            "*"
        ]
    }
]
}

```

- Utilisez la politique intégrée suivante pour autoriser un rôle IAM à supprimer une intégration de Lake Formation avec IAM Identity Center.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
        "sso:DeleteApplication"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}

```

- Pour les autorisations IAM requises pour accorder ou révoquer des autorisations de lac de données pour les utilisateurs et les groupes IAM Identity Center, voir. [Autorisations IAM requises pour accorder ou révoquer les autorisations de Lake Formation](#)

Description des autorisations

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration`— Crée la configuration iDC de Lake Formation.
- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration`— Décrit une configuration iDC existante.
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration`— Permet de supprimer une configuration iDC de Lake Formation existante.

- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration`— Utilisé pour modifier la configuration d'une Lake Formation existante.
- `sso:CreateApplication` : sert à créer une application IAM Identity Center.
- `sso:DeleteApplication` : sert à supprimer une application IAM Identity Center.
- `sso:UpdateApplication` : sert à mettre à jour une application IAM Identity Center.
- `sso:PutApplicationGrant` : sert à modifier les informations relatives à l'émetteur de jetons approuvé.
- `sso:PutApplicationAuthenticationMethod`— Accorde un accès d'authentification à Lake Formation.
- `sso:GetApplicationGrant` : sert à répertorier les informations relatives à l'émetteur de jetons approuvé.
- `sso:DeleteApplicationGrant` : supprime les informations relatives à l'émetteur de jetons approuvé.
- `sso:PutApplicationAccessScope`— Ajoute ou met à jour la liste des cibles autorisées pour une étendue d'accès à l'IAM Identity Center pour une application.
- `sso:PutApplicationAssignmentConfiguration`— Utilisé pour configurer la manière dont les utilisateurs accèdent à une application.

Connecter Lake Formation à l'IAM Identity Center

Avant de pouvoir utiliser IAM Identity Center pour gérer les identités afin d'accorder l'accès aux ressources du catalogue de données à l'aide de Lake Formation, vous devez suivre les étapes suivantes. Vous pouvez créer l'intégration IAM Identity Center à l'aide de la console Lake Formation ou AWS CLI.

AWS Management Console

Pour connecter Lake Formation à l'IAM Identity Center

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, sélectionnez Intégration à IAM Identity Center.

Create IAM Identity Center Integration

Enable IAM Identity Center and then create Lake Formation - IAM Identity Center integration to manage identities from IAM Identity Center (external IdPs like Azure AD or Okta Universal Directory). [Learn more](#)

▼ How it works

Enable IAM Identity Center

Enable IAM Identity Center for your account or organization and select an identity provider.

Create Lake Formation integration

Integrate Lake Formation with IAM Identity Center to permit Lake Formation to access users from your selected identity provider.

Grant permissions

Grant permissions to users on Data Catalog databases and tables using fine-grained Lake Formation permissions.

Connect Lake Formation to IAM Identity Center



Connect to organization instance of IAM Identity Center

Manage access to Lake Formation by assigning users and groups from the Identity Center directory for your organization. [Learn more](#)

Recommended



Connect to account instance of IAM Identity Center

Manage access to Lake Formation by assigning existing or creating dedicated users and groups from your Identity Center directory. [Learn more](#)

instance of IAM Identity Center

Manage access to Lake Formation by assigning users and groups from your Identity Center directory.

 `arn:aws:sso::instance/ssoins-6987513bf5410c2f`

Add AWS account and organization IDs

Add AWS accounts and organizations whose users need access to Lake Formation managed resources.

AWS Accounts and AWS organizations

Enter one or more AWS account IDs and AWS organization IDs. Press Enter after each ID.



► Lake Formation application integration - optional

Add application IDs that can access S3 data locations registered with Lake Formation on behalf of the user.

 After this step, you can't edit the connection. You can edit AWS accounts, organizations, and applications. If you want to modify the connection, delete it and create a new connection.

Create

3. (Facultatif) Entrez une ou plusieurs organisations IDs ou unités and/or organisationnelles valides Compte AWS IDs IDs pour autoriser les comptes externes à accéder aux ressources du catalogue de données. Lorsque des utilisateurs ou des groupes d'IAM Identity Center

tentent d'accéder aux ressources du catalogue de données géré par Lake Formation, Lake Formation assume un rôle IAM pour autoriser l'accès aux métadonnées. Si le rôle IAM appartient à un compte externe qui n'a pas de politique de AWS Glue ressources ni de partage de AWS RAM ressources, les utilisateurs et les groupes de l'IAM Identity Center ne pourront pas accéder à la ressource même s'ils disposent des autorisations de Lake Formation.

Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour partager la ressource avec des comptes et des organisations externes. AWS RAM envoie une invitation au compte du bénéficiaire pour qu'il accepte ou rejette le partage des ressources.

Pour de plus amples informations, veuillez consulter [Acceptation d'une invitation de partage de ressources de AWS RAM](#).

 Note

Lake Formation permet aux rôles IAM issus de comptes externes d'agir en tant que rôles de support au nom des utilisateurs et des groupes d'IAM Identity Center pour accéder aux ressources du catalogue de données, mais les autorisations ne peuvent être accordées que sur les ressources du catalogue de données du compte propriétaire. Si vous essayez d'accorder des autorisations aux utilisateurs et aux groupes d'IAM Identity Center sur les ressources du catalogue de données d'un compte externe, Lake Formation génère le message d'erreur suivant : « Les autorisations entre comptes ne sont pas prises en charge pour le principal ».

4. (Facultatif) Sur l'écran d'intégration ARNs de Create Lake Formation, spécifiez les applications tierces qui peuvent accéder aux données des sites Amazon S3 enregistrés auprès de Lake Formation. Lake Formation fournit des informations d'identification temporaires limitées sous forme de AWS STS jetons aux sites Amazon S3 enregistrés en fonction des autorisations effectives, afin que les applications autorisées puissent accéder aux données pour le compte des utilisateurs.
5. (Facultatif) Sur l'écran d'intégration de Create Lake Formation, cochez la case Amazon Redshift Connect dans Trusted Identity Propagation pour activer la découverte des autorisations fédérées par Amazon Redshift via IDC. Lake Formation propage l'identité vers l'aval en fonction des autorisations effectives, afin que les applications autorisées puissent accéder aux données au nom des utilisateurs.
6. Sélectionnez Soumettre.

Une fois que l'administrateur de Lake Formation a terminé les étapes et créé l'intégration, les propriétés de l'IAM Identity Center apparaissent dans la console Lake Formation. L'exécution de ces tâches fait de Lake Formation une application compatible avec IAM Identity Center. Les propriétés de la console incluent l'état de l'intégration. L'état de l'intégration indique Success quand elle est terminée. Ce statut indique si la configuration du centre d'identité IAM est terminée.

AWS CLI

- L'exemple suivant montre comment créer une intégration de Lake Formation avec IAM Identity Center. Vous pouvez également spécifier le Status (ENABLED,DISABLED) des applications.

```
aws lakeformation create-lake-formation-identity-center-configuration \
  --catalog-id <123456789012> \
  --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
  --share-recipients '[{"DataLakePrincipalIdentifier": "<123456789012>"},
                      {"DataLakePrincipalIdentifier": "<555555555555>"}]' \
  --external-filtering '{"AuthorizedTargets": [<app arn1>, "<app arn2>"],
                      "Status": "ENABLED"}'
```

- L'exemple suivant montre comment visualiser une intégration de Lake Formation avec IAM Identity Center.

```
aws lakeformation describe-lake-formation-identity-center-configuration
  --catalog-id <123456789012>
```

- L'exemple suivant montre comment activer Redshift :Connect l'autorisation. L'autorisation peut être ACTIVÉE ou DÉSACTIVÉE.

```
aws lakeformation create-lake-formation-identity-center-configuration \
  --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
  --service-integrations '[{
    "Redshift": [{
      "RedshiftConnect": {
        "Authorization": "ENABLED"
      }
    }
  ]]
```

```
}]'
```

- Utilisez la `describe-lake-formation-identity-center-configuration` commande pour décrire l'application Lake Formation Identity Center. Redshift:Connect l'intégration des services est essentielle pour la propagation de l'identité IDC entre services et entre clusters :

```
aws lakeformation describe-lake-formation-identity-center-configuration --catalog-id <123456789012>
```

Réponse :

```
{
  "CatalogId": "CATALOG ID",
  "InstanceArn": "INSTANCE ARN",
  "ApplicationArn": "APPLICATION ARN",
  "ShareRecipients": [],
  "ServiceIntegrations": [
    {
      "Redshift": [
        {
          "RedshiftConnect": {
            "Authorization": "ENABLED"
          }
        }
      ]
    }
  ]
}
```

Utilisation d'IAM Identity Center sur plusieurs Régions AWS

Lake Formation prend en charge l'IAM Identity Center à plusieurs Régions AWS reprises. Vous pouvez étendre IAM Identity Center de votre région principale Région AWS à d'autres régions pour améliorer les performances grâce à la proximité des utilisateurs et à la fiabilité. Lorsqu'une nouvelle région est ajoutée dans IAM Identity Center, vous pouvez créer des applications Lake Formation Identity Center dans la nouvelle région sans répliquer les identités de la région principale. Pour plus de détails sur les premiers pas avec IAM Identity Center dans plusieurs régions, consultez la section [Multi-Region IAM Identity Center](#) dans le guide de l'utilisateur d'IAM Identity Center.

Mise à jour de l'intégration d'IAM Identity Center

Après avoir créé la connexion, vous pouvez ajouter des applications tierces pour l'intégration d'IAM Identity Center afin de les intégrer à Lake Formation et d'accéder aux données Amazon S3 pour le compte des utilisateurs. Vous pouvez également supprimer des applications existantes de l'intégration d'IAM Identity Center. Vous pouvez ajouter ou supprimer des applications à l'aide de la console Lake Formation et en utilisant [UpdateLakeFormationIdentityCenterConfiguration](#) Operation. AWS CLI

Note

Après avoir créé l'intégration IAM Identity Center, vous ne pouvez pas mettre à jour l'instanceARN.

AWS Management Console

Pour mettre à jour une connexion IAM Identity Center existante avec Lake Formation

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, sélectionnez Intégration à IAM Identity Center.
3. Sélectionnez Ajouter sur la page d'intégration d'IAM Identity Center.
4. Entrez une ou plusieurs organisations IDs ou unités and/or organisationnelles valides Compte AWS IDs IDs pour autoriser les comptes externes à accéder aux ressources du catalogue de données.
5. Sur l'écran Ajouter des applications, entrez l'application IDs des applications tierces que vous souhaitez intégrer à Lake Formation.
6. Sélectionnez Ajouter.
7. (Facultatif) Sur la page d'intégration d'IAM Identity Center, vous pouvez soit activer la propagation d'identité sécurisée pour Amazon Redshift Connect, soit la désactiver. Lake Formation propage l'identité vers l'aval en fonction des autorisations effectives, afin que les applications autorisées puissent accéder aux données au nom des utilisateurs.

AWS CLI

Vous pouvez ajouter ou supprimer des applications tierces pour l'intégration d'IAM Identity Center en exécutant la AWS CLI commande suivante. Lorsque vous définissez l'état du filtrage externe sur `ENABLED`, cela permet à l'IAM Identity Center de fournir une gestion des identités permettant aux applications tierces d'accéder aux données gérées par Lake Formation. Vous pouvez également activer ou désactiver l'intégration d'IAM Identity Center en définissant le statut de l'application.

```
aws lakeformation update-lake-formation-identity-center-configuration \
  --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"], "Status": "ENABLED"}' \
  --share-recipients '[{"DataLakePrincipalIdentifier": "<444455556666>"} {"DataLakePrincipalIdentifier": "<777788889999>"}]' \
  --application-status ENABLED
```

Si vous avez une application LF IDC existante, mais que vous souhaitez ajouter l'`Redshift:Connect` autorisation, vous pouvez utiliser ce qui suit pour mettre à jour votre application Lake Formation IDC. L'autorisation peut être `ACTIVÉE` ou `DÉSACTIVÉE`.

```
aws lakeformation update-lake-formation-identity-center-configuration \
  --service-integrations '[{
    "Redshift": [{
      "RedshiftConnect": {
        "Authorization": "ENABLED"
      }
    }]
}]'
```

Suppression d'une connexion Lake Formation avec IAM Identity Center

Si vous souhaitez supprimer une intégration IAM Identity Center existante, vous pouvez le faire à l'aide de la console Lake Formation ou d'une [DeleteLakeFormationIdentityCenterConfiguration](#) opération. AWS CLI

AWS Management Console

Pour supprimer une connexion IAM Identity Center existante avec Lake Formation

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, sélectionnez Intégration à IAM Identity Center.
3. Sélectionnez Supprimer sur la page d'intégration d'IAM Identity Center.
4. Sur l'écran Confirmer l'intégration, confirmez l'action, puis sélectionnez Supprimer.

AWS CLI

Vous pouvez supprimer l'intégration d'IAM Identity Center en exécutant la AWS CLI commande suivante.

```
aws lakeformation delete-lake-formation-identity-center-configuration \
  --catalog-id <123456789012>
```

Octroi d'autorisations aux utilisateurs et aux groupes

L'administrateur de votre lac de données peut accorder des autorisations aux utilisateurs et aux groupes d'IAM Identity Center sur les ressources du catalogue de données (bases de données, tables et vues) afin de faciliter l'accès aux données. Pour accorder ou révoquer des autorisations de lac de données, le concédant a besoin d'autorisations pour les actions suivantes de l'IAM Identity Center.

- [DescribeUser](#)
- [DescribeGroup](#)
- [DescribeInstance](#)

Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS CLI.

Pour plus d'informations sur l'octroi d'autorisations, consultez [the section called “Octroi d'autorisations de données”](#).

Note

Vous ne pouvez accorder des autorisations que sur les ressources de votre compte. Pour attribuer des autorisations en cascade aux utilisateurs et aux groupes sur les ressources partagées avec vous, vous devez utiliser AWS RAM des partages de ressources.

AWS Management Console

Pour accorder des autorisations aux utilisateurs et aux groupes

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Sélectionnez Autorisations du lac de données sous Autorisations dans la console Lake Formation.
3. Sélectionnez Grant.
4. Sur la page Accorder les autorisations du lac de données, sélectionnez Utilisateurs et groupes IAM Identity Center.
5. Sélectionnez Ajouter pour choisir les utilisateurs et les groupes auxquels accorder des autorisations.

Grant permissions

Principals

Choose the principals to grant permissions.

☒ **IAM users and roles**
Users or roles from this AWS account.

☐ **IAM Identity Center - new**
Users and groups configured in IAM Identity Center.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☐ **External accounts**
AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add ▼

AWSServiceRoleForRedshift X
Role

6. Sur l'écran Attribuer des utilisateurs et des groupes, choisissez les and/or groupes d'utilisateurs auxquels vous accordez des autorisations.

Sélectionnez Attribuer.

Assign users and groups [X]

Search by user display name or group name

Users

user1 [Remove]

user2 [Remove]

Groups

DataStewards [Remove]

[Manage groups]

[Learn more about managing groups from IAM Identity Center](#)

[Cancel] [Assign]

7. Ensuite, choisissez la méthode d'octroi des autorisations.

Pour obtenir des instructions sur l'octroi d'autorisations à l'aide de la méthode des ressources nommées, consultez [Octroi d'autorisations de données à l'aide de la méthode de ressource nommée](#).

Pour obtenir des instructions sur l'octroi d'autorisations à l'aide de balises LF, consultez [Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC](#)

8. Choisissez les ressources du catalogue de données pour lesquelles vous souhaitez accorder des autorisations.

9. Choisissez les autorisations de catalogue de données à accorder.
10. Sélectionnez Grant.

AWS CLI

L'exemple suivant montre comment accorder l'`SELECT` autorisation utilisateur d'IAM Identity Center sur une table.

```
aws lakeformation grant-permissions \  
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \  
--permissions "SELECT" \  
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Pour effectuer une récupération `UserId` depuis IAM Identity Center, voir le [GetUserId](#) fonctionnement dans le manuel de référence de l'API IAM Identity Center.

Intégration du contexte utilisateur d'IAM Identity Center dans les journaux CloudTrail

Lake Formation utilise la fonctionnalité de distribution [automatique d'informations d'identification](#) pour fournir un accès temporaire aux données Amazon S3. Par défaut, lorsqu'un utilisateur de l'IAM Identity Center soumet une requête à un service d'analyse intégré, les CloudTrail journaux incluent uniquement le rôle IAM assumé par le service pour fournir un accès à court terme. Si vous utilisez un rôle défini par l'utilisateur pour enregistrer l'emplacement des données Amazon S3 auprès de Lake Formation, vous pouvez choisir d'inclure le contexte de l'utilisateur IAM Identity Center dans les CloudTrail événements, puis de suivre les utilisateurs qui accèdent à vos ressources.

Important

Pour inclure les demandes d'API Amazon S3 au niveau de l'objet dans le CloudTrail, vous devez activer la journalisation des CloudTrail événements pour le compartiment et les objets Amazon S3. Pour plus d'informations, consultez la section [Activation de la journalisation des CloudTrail événements pour les buckets et les objets Amazon S3](#) dans le guide de l'utilisateur Amazon S3.

Pour activer l'audit des distributeurs automatiques d'informations d'identification sur les sites des lacs de données enregistrés avec des rôles définis par l'utilisateur

1. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, développez Administration, puis sélectionnez Paramètres du catalogue de données.
3. Sous Audit amélioré, choisissez Propager le contexte fourni.
4. Choisissez Enregistrer.

Vous pouvez également activer l'option d'audit améliorée en définissant l'`Parametersattribut` dans l'[PutDataLakeSettings](#) opération. Par défaut, la valeur du `SET_CONTEXT` paramètre est définie sur « true ».

```
{
  "DataLakeSettings": {
    "Parameters": {"SET_CONTEXT": "true"},
  }
}
```

Ce qui suit est un extrait d'un CloudTrail événement avec l'option d'audit améliorée. Ce journal inclut à la fois le contexte de session de l'utilisateur IAM Identity Center et le rôle IAM défini par l'utilisateur assumé par Lake Formation pour accéder à l'emplacement des données Amazon S3. Consultez le `onBehalfOf` paramètre dans l'extrait suivant.

```
{
  "eventVersion": "1.09",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "AR0AW7F7M0X40YE6FLIFN:access-grants-
e653760c-4e8b-44fd-94d9-309e035b75ab",
    "arn": "arn:aws:sts::123456789012:assumed-role/accessGrantsTestRole/access-
grants-e653760c-4e8b-44fd-94d9-309e035b75ab",
    "accountId": "123456789012",
    "accessKeyId": "ASIAW7F7M0X4CQLD4JIZN",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "AR0AW7F7M0X40YE6FLIFN",
```

```
        "arn": "arn:aws:iam::123456789012:role/accessGrantsTestRole",
        "accountId": "123456789012",
        "userName": "accessGrantsTestRole"
    },
    "attributes": {
        "creationDate": "2023-08-09T17:24:02Z",
        "mfaAuthenticated": "false"
    }
},
"onBehalfOf": {
    "userId": "<identityStoreUserId>",
    "identityStoreArn": "arn:aws:identitystore::<restOfIdentityStoreArn>"
}
},
"eventTime": "2023-08-09T17:25:43Z",
"eventSource": "s3.amazonaws.com",
"eventName": "GetObject",
....
```

Ajouter un emplacement Amazon S3 à votre lac de données

Pour ajouter un emplacement de données en tant que stockage dans votre lac de données, vous devez enregistrer l'emplacement (emplacement du lac de données) auprès de AWS Lake Formation. Vous pouvez ensuite utiliser les autorisations de Lake Formation pour un contrôle d'accès précis aux AWS Glue Data Catalog objets pointant vers cet emplacement et aux données sous-jacentes de cet emplacement.

Lake Formation permet également d'enregistrer un emplacement de données en mode d'accès hybride et vous offre la flexibilité d'activer de manière sélective les autorisations Lake Formation pour les bases de données et les tables de votre catalogue de données. Avec le mode d'accès hybride, vous disposez d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Pour plus d'informations sur la configuration du mode d'accès hybride, voir [Mode d'accès hybride](#)

Lorsque vous enregistrez un emplacement, ce chemin Amazon S3 et tous les dossiers situés sous ce chemin sont enregistrés.

Supposons, par exemple, que vous disposiez d'une organisation des chemins Amazon S3 telle que la suivante :

```
/mybucket/accounting/sales/
```

Si vous vous inscrivez S3 : `//mybucket/accounting`, le `sales` dossier est également enregistré et placé sous Gestion de Lake Formation.

Pour plus d'informations sur l'enregistrement des points de vente, consultez [Underlying data access control](#).

Note

Les autorisations Lake Formation sont recommandées pour les données structurées (organisées dans des tableaux avec des lignes et des colonnes). Si vos données contiennent des données non structurées basées sur des objets, pensez à utiliser les autorisations d'accès Amazon S3 pour gérer l'accès aux données.

Rubriques

- [Exigences relatives aux rôles utilisés pour enregistrer des sites](#)
- [Enregistrement d'un emplacement Amazon S3](#)
- [Enregistrement d'un emplacement Amazon S3 chiffré](#)
- [Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte](#)
- [Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes](#)
- [Annulation de l'enregistrement d'un site Amazon S3](#)

Exigences relatives aux rôles utilisés pour enregistrer des sites

Vous devez spécifier un rôle Gestion des identités et des accès AWS (IAM) lorsque vous enregistrez un emplacement Amazon Simple Storage Service (Amazon S3). AWS Lake Formation assume ce rôle lors de l'accès aux données à cet emplacement.

Vous pouvez utiliser l'un des types de rôles suivants pour enregistrer un emplacement :

- Le rôle lié au service Lake Formation. Ce rôle accorde les autorisations requises sur l'emplacement. L'utilisation de ce rôle est le moyen le plus simple d'enregistrer l'emplacement.

Pour plus d'informations, consultez [Utilisation de rôles liés à un service pour Lake Formation](#) et [Limitations des rôles liés aux services](#).

- Rôle défini par l'utilisateur. Utilisez un rôle défini par l'utilisateur lorsque vous devez accorder plus d'autorisations que le rôle lié à un service n'en fournit.

Vous devez utiliser un rôle défini par l'utilisateur dans les cas suivants :

- Lors de l'enregistrement d'une position dans un autre compte.

Pour plus d'informations, consultez [the section called “Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte”](#) et [the section called “Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes”](#).

- Si vous avez utilisé une AWS clé CMK gérée (aws/s3) pour chiffrer l'emplacement Amazon S3.

Pour de plus amples informations, veuillez consulter [Enregistrement d'un emplacement Amazon S3 chiffré](#).

- Si vous prévoyez d'accéder à l'emplacement via Amazon EMR.

Si vous avez déjà enregistré un emplacement avec le rôle lié au service et que vous souhaitez commencer à y accéder avec Amazon EMR, vous devez annuler l'enregistrement du point de vente et le réenregistrer avec un rôle défini par l'utilisateur. Pour de plus amples informations, veuillez consulter [the section called “Annulation de l'enregistrement d'un site Amazon S3”](#).

Utilisation de rôles liés à un service pour Lake Formation

AWS Lake Formation utilise un rôle Gestion des identités et des accès AWS lié à un service (IAM). Un rôle lié à un service est un type unique de rôle IAM directement lié à Lake Formation. Le rôle lié au service est prédéfini par Lake Formation et inclut toutes les autorisations dont le service a besoin pour appeler d'autres AWS services en votre nom.

Un rôle lié à un service facilite la configuration de Lake Formation, car il n'est pas nécessaire de créer un rôle et d'ajouter manuellement les autorisations nécessaires. Lake Formation définit les autorisations associées à son rôle lié aux services et, sauf indication contraire, seule Lake Formation peut assumer ses rôles. Les autorisations définies comprennent la politique d'approbation et la politique d'autorisation. De plus, cette politique d'autorisation ne peut pas être attachée à une autre entité IAM.

Ce rôle lié à un service fait confiance aux services suivants pour assumer le rôle :

- `lakeformation.amazonaws.com`

Lorsque vous utilisez un rôle lié à un service dans le compte A pour enregistrer un emplacement Amazon S3 appartenant au compte B, la politique de compartiment Amazon S3 (une politique basée sur les ressources) du compte B doit accorder des autorisations d'accès au rôle lié au service dans le compte A.

Pour plus d'informations sur l'utilisation d'un rôle lié à un service pour enregistrer un emplacement de données, consultez [Limitations des rôles liés aux services](#)

 Note

Les politiques de contrôle des services (SCPs) n'affectent pas les rôles liés aux services. Pour plus d'informations, voir [Politiques de contrôle des services \(SCPs\)](#) dans le guide de AWS Organizations l'utilisateur.

Autorisations de rôle liées à un service pour Lake Formation

Lake Formation utilise le rôle lié au service nommé.

`AWSServiceRoleForLakeFormationDataAccess` Ce rôle fournit un ensemble d'autorisations Amazon Simple Storage Service (Amazon S3) qui permettent au service intégré Lake Formation (Amazon Athena tel que) d'accéder aux emplacements enregistrés. Lorsque vous enregistrez un emplacement de lac de données, vous devez fournir un rôle disposant des read/write autorisations Amazon S3 requises pour cet emplacement. Au lieu de créer un rôle avec les autorisations Amazon S3 requises, vous pouvez utiliser ce rôle lié à un service.

La première fois que vous nommez le rôle lié au service comme le rôle avec lequel enregistrer un chemin, le rôle lié au service et une nouvelle politique IAM sont créés en votre nom. Lake Formation ajoute le chemin à la politique en ligne et l'associe au rôle lié au service. Lorsque vous enregistrez les chemins suivants avec le rôle lié au service, Lake Formation ajoute le chemin à la politique existante.

Lorsque vous êtes connecté en tant qu'administrateur du lac de données, enregistrez l'emplacement d'un lac de données. Ensuite, dans la console IAM, recherchez le rôle `AWSServiceRoleForLakeFormationDataAccess` et consultez les politiques qui lui sont associées.

Par exemple, une fois que vous avez enregistré l'emplacement `s3://my-kinesis-test/logs`, Lake Formation crée la politique en ligne suivante et l'attache à `AWSServiceRoleForLakeFormationDataAccess`.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "LakeFormationDataAccessPermissionsForS3",
      "Effect": "Allow",
      "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:DeleteObject",
        "s3:AbortMultipartUpload",
        "s3:ListMultipartUploadParts"
      ],
      "Resource": [
        "arn:aws:s3:::my-kinesis-test/logs/*"
      ]
    },
    {
      "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
      "Effect": "Allow",
      "Action": [
        "s3:ListBucket",
        "s3:ListBucketMultipartUploads"
      ],
      "Resource": [
        "arn:aws:s3:::my-kinesis-test"
      ]
    }
  ]
}
```

Création d'un rôle lié à un service pour Lake Formation

Vous n'avez pas besoin de créer manuellement un rôle lié à un service. Lorsque vous enregistrez un site Amazon S3 avec Lake Formation dans l' AWS Management Console AWS API AWS CLI, Lake Formation crée le rôle lié au service pour vous.

Important

Ce rôle lié à un service peut apparaître dans votre compte si vous avez effectué une action dans un autre service qui utilise les fonctions prises en charge par ce rôle. Pour en savoir plus, consultez [Un nouveau rôle est apparu dans mon compte IAM](#).

Si vous supprimez ce rôle lié à un service et que vous avez ensuite besoin de le recréer, vous pouvez utiliser la même procédure pour recréer le rôle dans votre compte. Lorsque vous enregistrez un site Amazon S3 auprès de Lake Formation, Lake Formation crée à nouveau le rôle lié au service pour vous.

Vous pouvez également utiliser la console IAM pour créer un rôle lié à un service avec le cas d'utilisation de Lake Formation. Dans l'API AWS CLI ou dans l' AWS API, créez un rôle lié à un service avec le nom du `lakeformation.amazonaws.com` service. Pour plus d'informations, consultez [Création d'un rôle lié à un service](#) dans le Guide de l'utilisateur IAM. Si vous supprimez ce rôle lié à un service, vous pouvez utiliser ce même processus pour créer le rôle à nouveau.

Modification d'un rôle lié à un service pour Lake Formation

Lake Formation ne vous permet pas de modifier le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service. Après avoir créé un rôle lié à un service, vous ne pouvez pas changer le nom du rôle, car plusieurs entités peuvent faire référence à ce rôle. Néanmoins, vous pouvez modifier la description du rôle à l'aide d'IAM. Pour en savoir plus, consultez [Modification d'un rôle lié à un service](#) dans le Guide de l'utilisateur IAM.

Suppression d'un rôle lié à un service pour Lake Formation

Si vous n'avez plus besoin d'utiliser une fonctionnalité ou un service qui nécessite un rôle lié à un service, nous vous recommandons de supprimer ce rôle. De cette façon, vous n'avez aucune entité inutilisée qui n'est pas surveillée ou gérée activement. Cependant, vous devez nettoyer les ressources de votre rôle lié à un service avant de pouvoir les supprimer manuellement.

 Note

Si le service Lake Formation utilise le rôle lorsque vous essayez de supprimer les ressources, la suppression risque d'échouer. Si cela se produit, patientez quelques minutes et réessayez.

Pour supprimer les ressources de la Formation du Lac utilisées par la Formation du Lac

- Si vous avez utilisé le rôle lié à un service pour enregistrer des sites Amazon S3 auprès de Lake Formation, avant de supprimer le rôle lié au service, vous devez désenregistrer l'emplacement et le réenregistrer à l'aide d'un rôle personnalisé.

Pour supprimer manuellement le rôle lié au service à l'aide d'IAM

Utilisez la console IAM, le AWS CLI, ou l' AWS API pour supprimer le rôle lié au `AWSServiceRoleForLakeFormationDataAccess` service. Pour plus d'informations, veuillez consulter [Suppression d'un rôle lié à un service](#) dans le Guide de l'utilisateur IAM.

Les conditions requises pour un rôle défini par l'utilisateur sont les suivantes :

- Lors de la création du nouveau rôle, sur la page Créer un rôle de la console IAM, sélectionnez AWS service, puis sous Choisir un cas d'utilisation, choisissez Lake Formation.

Si vous créez le rôle en utilisant un chemin différent, assurez-vous que le rôle entretient une relation de confiance avec `lakeformation.amazonaws.com`. Pour plus d'informations, consultez la section [Modification d'une politique d'approbation des rôles \(console\)](#).

- Le rôle doit avoir une politique intégrée qui accorde des read/write autorisations à Amazon S3 sur le site. Voici une politique typique.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:DeleteObject"
      ]
    }
  ]
}
```

```

    ],
    "Resource": [
        "arn:aws:s3:::awsexamplebucket/*"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
        "s3:ListBucket"
    ],
    "Resource": [
        "arn:aws:s3:::awsexamplebucket"
    ]
  }
]
}

```

- Ajoutez la politique de confiance suivante au rôle IAM pour permettre au service Lake Formation d'assumer le rôle et de vendre des informations d'identification temporaires aux moteurs d'analyse intégrés.

Pour inclure le contexte utilisateur d'IAM Identity Center dans les CloudTrail journaux, la politique de confiance doit autoriser `sts:SetContextaction`.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "DataCatalogViewDefinerAssumeRole1",
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "lakeformation.amazonaws.com"
        ]
      },
      "Action": [
        "sts:AssumeRole",
        "sts:SetContext"
      ]
    }
  ]
}

```

```
}
```

- L'administrateur du lac de données qui enregistre l'emplacement doit disposer de l'`iam:PassRole` autorisation associée au rôle.

Voici une politique intégrée qui accorde cette autorisation. Remplacez `<account-id>` par un numéro de AWS compte valide et remplacez `<role-name>` par le nom du rôle.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "PassRolePermissions",
      "Effect": "Allow",
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
        "arn:aws:iam::111122223333:role/<role-name>"
      ]
    }
  ]
}
```

- Pour permettre à Lake Formation d'ajouter des journaux dans CloudWatch les journaux et de publier des métriques, ajoutez la politique en ligne suivante.

Note

L'écriture dans CloudWatch Logs entraîne des frais.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Sid1",
```

```
    "Effect": "Allow",
    "Action": [
        "logs:CreateLogStream",
        "logs:CreateLogGroup",
        "logs:PutLogEvents"
    ],
    "Resource": [
        "arn:aws:logs:us-east-1:111122223333:log-group:/aws-lakeformation-acceleration/*",
        "arn:aws:logs:us-east-1:111122223333:log-group:/aws-lakeformation-acceleration/*:log-stream:*"
    ]
}
}
```

Enregistrement d'un emplacement Amazon S3

Vous devez spécifier un rôle Gestion des identités et des accès AWS (IAM) lorsque vous enregistrez un emplacement Amazon Simple Storage Service (Amazon S3). Lake Formation assume ce rôle lorsqu'elle accorde des informations d'identification temporaires aux AWS services intégrés qui accèdent aux données à cet endroit.

Important

Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Vous pouvez utiliser la AWS Lake Formation console, l'API Lake Formation ou AWS Command Line Interface (AWS CLI) pour enregistrer un emplacement Amazon S3.

Avant de commencer

Passez en revue [les exigences relatives au rôle utilisé pour enregistrer l'emplacement](#).

Pour enregistrer un emplacement (console)

Important

Les procédures suivantes supposent que le site Amazon S3 se trouve dans le même AWS compte que le catalogue de données et que les données du site ne sont pas chiffrées. Les autres sections de ce chapitre traitent de l'enregistrement entre comptes et de l'enregistrement des emplacements chiffrés.

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
2. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
3. Choisissez Register location, puis Browse pour sélectionner un chemin Amazon Simple Storage Service (Amazon S3).
4. (Facultatif, mais fortement recommandé) Sélectionnez Vérifier les autorisations de localisation pour afficher la liste de toutes les ressources existantes dans l'emplacement Amazon S3 sélectionné et leurs autorisations.

L'enregistrement de l'emplacement sélectionné peut permettre aux utilisateurs de votre Lake Formation d'accéder aux données déjà présentes à cet emplacement. La consultation de cette liste vous permet de garantir la sécurité des données existantes.

5. Pour le rôle IAM, choisissez le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service (par défaut) ou un rôle IAM personnalisé répondant aux exigences de. [the section called "Exigences relatives aux rôles utilisés pour enregistrer des sites"](#)

Vous pouvez mettre à jour un emplacement enregistré ou d'autres informations uniquement lorsque vous l'enregistrez à l'aide d'un rôle IAM personnalisé. Pour modifier un point de vente enregistré à l'aide d'un rôle lié à un service, vous devez le désenregistrer, puis l'enregistrer à nouveau.

6. Choisissez l'option Enable Data Catalog Federation pour autoriser Lake Formation à assumer un rôle et à vendre des informations d'identification temporaires aux AWS services intégrés afin d'accéder aux tables des bases de données fédérées. Si un emplacement est enregistré auprès de Lake Formation et que vous souhaitez utiliser le même emplacement pour une table dans une base de données fédérée, vous devez enregistrer le même emplacement à l'aide de l'option Enable Data Catalog Federation.

7. Choisissez le mode d'accès hybride pour ne pas activer les autorisations de Lake Formation par défaut. Lorsque vous enregistrez l'emplacement Amazon S3 en mode d'accès hybride, vous pouvez activer les autorisations de Lake Formation en optant pour les principes pour les bases de données et les tables situées sous cet emplacement.

Pour plus d'informations sur la configuration du mode d'accès hybride, consultez [Mode d'accès hybride](#).

8. Sélectionnez Enregistrer l'emplacement.

Pour enregistrer un point de vente (AWS CLI)

1. Enregistrez un nouvel emplacement avec Lake Formation

Cet exemple utilise un rôle lié à un service pour enregistrer l'emplacement. Vous pouvez plutôt utiliser l'option `--role-arn` pour indiquer votre propre rôle.

Remplacez-le `<s3-path>` par un chemin Amazon S3 valide, un numéro de AWS compte associé à un compte valide et `<s3-access-role>` par un rôle IAM autorisé à enregistrer un emplacement de données.

 Note

Vous ne pouvez pas modifier les propriétés d'un point de vente enregistré s'il est enregistré à l'aide d'un rôle lié à un service.

```
aws lakeformation register-resource \  
  --resource-arn arn:aws:s3:::<s3-path> \  
  --use-service-linked-role
```

L'exemple suivant utilise un rôle personnalisé pour enregistrer l'emplacement.

```
aws lakeformation register-resource \  
  --resource-arn arn:aws:s3:::<s3-path> \  
  --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
```

2. Pour mettre à jour une position enregistrée auprès de Lake Formation

Vous ne pouvez modifier un emplacement enregistré que s'il est enregistré à l'aide d'un rôle IAM personnalisé. Pour un emplacement enregistré avec un rôle lié à un service, vous devez le désenregistrer et l'enregistrer à nouveau. Pour de plus amples informations, veuillez consulter [the section called “Annulation de l'enregistrement d'un site Amazon S3”](#).

```
aws lakeformation update-resource \  
  --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \  
  --resource-arn arn:aws:s3:::<s3-path>
```

```
aws lakeformation update-resource \  
  --resource-arn arn:aws:s3:::<s3-path> \  
  --use-service-linked-role
```

3. Enregistrer un emplacement de données en mode d'accès hybride avec fédération

```
aws lakeformation register-resource \  
  --resource-arn arn:aws:s3:::<s3-path> \  
  --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \  
  --hybrid-access-enabled
```

```
aws lakeformation register-resource \  
  --resource-arn arn:aws:s3:::<s3-path> \  
  --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \  
  --with-federation
```

```
aws lakeformation update-resource \  
  --resource-arn arn:aws:s3:::<s3-path> \  
  --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \  
  --hybrid-access-enabled
```

Pour plus d'informations, consultez la section Fonctionnement de l'[RegisterResource](#) API.

 Note

Une fois que vous avez enregistré un emplacement Amazon S3, toute AWS Glue table pointant vers cet emplacement (ou l'un de ses emplacements enfants) renvoie la valeur du `IsRegisteredWithLakeFormation` paramètre comme `true` dans l'`GetTable` appel. Il existe une limite connue selon laquelle les opérations de l'API du catalogue de données, telles que `GetTables` la mise à jour ou non de la valeur du `IsRegisteredWithLakeFormation` paramètre, renvoient la valeur par défaut, qui est fausse. `SearchTables` Il est recommandé d'utiliser l'`GetTableAPI` pour afficher la valeur correcte du `IsRegisteredWithLakeFormation` paramètre.

Enregistrement d'un emplacement Amazon S3 chiffré

Lake Formation s'intègre à [AWS Key Management Service](#) (AWS KMS) pour vous permettre de configurer plus facilement d'autres services intégrés pour chiffrer et déchiffrer les données sur les sites Amazon Simple Storage Service (Amazon S3).

Ils Clés gérées par AWS sont tous deux gérés par AWS KMS keys le client et pris en charge. Actuellement, le côté client n' encryption/decryption est pris en charge qu'avec Athena.

Vous devez spécifier un rôle Gestion des identités et des accès AWS (IAM) lorsque vous enregistrez un emplacement Amazon S3. Pour les sites Amazon S3 chiffrés, le rôle doit être autorisé à chiffrer et à déchiffrer les données avec le AWS KMS key, ou la politique de clé KMS doit accorder des autorisations sur la clé du rôle.

 Important

Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Lake Formation utilise un rôle lié à un service pour enregistrer l'emplacement de vos données. Ce rôle présente toutefois plusieurs [limites](#). En raison de ces contraintes, nous recommandons plutôt de créer et d'utiliser un rôle IAM personnalisé pour plus de flexibilité et de contrôle. Le rôle personnalisé

que vous créez pour enregistrer l'emplacement doit répondre aux exigences spécifiées dans [the section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”](#).

 Important

Si vous avez utilisé un Clé gérée par AWS pour chiffrer l'emplacement Amazon S3, vous ne pouvez pas utiliser le rôle lié au service Lake Formation. Vous devez utiliser un rôle personnalisé et ajouter des autorisations IAM sur la clé du rôle. Les détails sont fournis plus loin dans cette section.

Les procédures suivantes expliquent comment enregistrer un emplacement Amazon S3 chiffré à l'aide d'une clé gérée par le client ou d'un Clé gérée par AWS.

- [Enregistrement d'un emplacement chiffré à l'aide d'une clé gérée par le client](#)
- [Enregistrer une position cryptée avec un Clé gérée par AWS](#)

Avant de commencer

Passez en revue [les exigences relatives au rôle utilisé pour enregistrer l'emplacement](#).

Pour enregistrer un emplacement Amazon S3 chiffré à l'aide d'une clé gérée par le client

 Note

Si la clé KMS ou l'emplacement Amazon S3 ne se trouvent pas dans le même AWS compte que le catalogue de données, suivez [the section called “Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes”](#) plutôt les instructions indiquées.

1. Ouvrez la AWS KMS console à l'<https://console.aws.amazon.com/adresse/kms> et connectez-vous en tant qu'utilisateur administratif Gestion des identités et des accès AWS (IAM) ou en tant qu'utilisateur pouvant modifier la politique de clé KMS utilisée pour chiffrer l'emplacement.
2. Dans le volet de navigation, sélectionnez Clés gérées par le client, puis choisissez le nom de la clé KMS souhaitée.
3. Sur la page de détails des clés KMS, choisissez l'onglet Politique clé, puis effectuez l'une des opérations suivantes pour ajouter votre rôle personnalisé ou le rôle lié au service Lake Formation en tant qu'utilisateur clé KMS :

- Si la vue par défaut s'affiche (avec les sections Administrateurs clés, Suppression des clés, Utilisateurs clés et Autres AWS comptes), dans la section Utilisateurs clés, ajoutez votre rôle personnalisé ou le rôle lié au service Lake Formation. `AWSServiceRoleForLakeFormationDataAccess`
- Si la politique clé (JSON) s'affiche, modifiez la politique pour ajouter votre rôle personnalisé ou le rôle lié au service Lake Formation `AWSServiceRoleForLakeFormationDataAccess` à l'objet « Autoriser l'utilisation de la clé », comme indiqué dans l'exemple suivant.

 Note

Si cet objet est manquant, ajoutez-le avec les autorisations indiquées dans l'exemple. L'exemple utilise le rôle lié à un service.

```
...
{
  "Sid": "Allow use of the key",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess",
      "arn:aws:iam::111122223333:user/keyuser"
    ]
  },
  "Action": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
  ],
  "Resource": "*"
},
...
```

4. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
5. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
6. Choisissez Register location, puis Browse pour sélectionner un chemin Amazon Simple Storage Service (Amazon S3).
7. (Facultatif, mais fortement recommandé) Choisissez Vérifier les autorisations de localisation pour afficher la liste de toutes les ressources existantes dans l'emplacement Amazon S3 sélectionné ainsi que leurs autorisations.

L'enregistrement de l'emplacement sélectionné peut permettre aux utilisateurs de votre Lake Formation d'accéder aux données déjà présentes à cet emplacement. La consultation de cette liste vous permet de garantir la sécurité des données existantes.

8. Pour le rôle IAM, choisissez soit le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service (par défaut), soit votre rôle personnalisé qui répond aux [the section called "Exigences relatives aux rôles utilisés pour enregistrer des sites"](#)
9. Choisissez Enregistrer l'emplacement.

Pour de plus amples informations sur le rôle lié à un service, veuillez consulter [Autorisations de rôle liées à un service pour Lake Formation](#).

Pour enregistrer une position Amazon S3 chiffrée à l'aide d'un Clé gérée par AWS

 Important

Si l'emplacement Amazon S3 n'est pas enregistré dans le même AWS compte que le catalogue de données, suivez [the section called "Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes"](#) plutôt les instructions indiquées dans la section.

1. Créez un rôle IAM à utiliser pour enregistrer l'emplacement. Assurez-vous qu'il répond aux exigences répertoriées dans [the section called "Exigences relatives aux rôles utilisés pour enregistrer des sites"](#).
2. Ajoutez la politique intégrée suivante au rôle. Il accorde des autorisations sur la clé du rôle. La Resource spécification doit indiquer le nom de ressource Amazon (ARN) du Clé gérée par AWS. Vous pouvez obtenir l'ARN depuis la AWS KMS console. Pour obtenir le bon ARN,

assurez-vous de vous connecter à la AWS KMS console avec le même AWS compte et la même région Clé gérée par AWS que ceux utilisés pour chiffrer l'emplacement.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "arn:aws:kms:us-east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
    }
  ]
}
```

Vous pouvez utiliser des alias de clé KMS au lieu de l'ID de clé -

arn:aws:kms:region:account-id:key/alias/your-key-alias

Pour plus d'informations, consultez la AWS KMS section [Alias](#) du Guide du AWS Key Management Service développeur.

3. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation lakeformation:RegisterResource IAM.
4. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
5. Choisissez Register location, puis Browse pour sélectionner un chemin Amazon S3.
6. (Facultatif, mais fortement recommandé) Choisissez Vérifier les autorisations de localisation pour afficher la liste de toutes les ressources existantes dans l'emplacement Amazon S3 sélectionné ainsi que leurs autorisations.

L'enregistrement de l'emplacement sélectionné peut permettre aux utilisateurs de votre Lake Formation d'accéder aux données déjà présentes à cet emplacement. La consultation de cette liste vous permet de garantir la sécurité des données existantes.

7. Pour le rôle IAM, choisissez le rôle que vous avez créé à l'étape 1.
8. Choisissez Enregistrer l'emplacement.

Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte

AWS Lake Formation vous permet d'enregistrer des sites AWS Amazon Simple Storage Service (Amazon S3) sur plusieurs comptes. Par exemple, s'il AWS Glue Data Catalog se trouve dans le compte A, un utilisateur du compte A peut enregistrer un compartiment Amazon S3 dans le compte B.

L'enregistrement d'un compartiment Amazon S3 dans le AWS compte B à l'aide d'un rôle Gestion des identités et des accès AWS (IAM) dans le AWS compte A nécessite les autorisations suivantes :

- Le rôle du compte A doit accorder des autorisations sur le compartiment du compte B.
- La politique de compartiment du compte B doit accorder des autorisations d'accès au rôle dans le compte A.

Important

Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Vous ne pouvez pas utiliser le rôle lié au service Lake Formation pour enregistrer une position sur un autre compte. Vous devez plutôt utiliser un rôle défini par l'utilisateur. Le rôle doit répondre aux exigences de [the section called "Exigences relatives aux rôles utilisés pour enregistrer des sites"](#). Pour de plus amples informations sur le rôle lié à un service, veuillez consulter [Autorisations de rôle liées à un service pour Lake Formation](#).

Avant de commencer

Passez en revue [les exigences relatives au rôle utilisé pour enregistrer l'emplacement](#).

Pour enregistrer une position dans un autre AWS compte

 Note

Si l'emplacement est crypté, suivez [the section called “Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes”](#) plutôt les instructions indiquées.

La procédure suivante suppose qu'un mandant du compte 1111-2222-3333, qui contient le catalogue de données, souhaite enregistrer le compartiment Amazon S3, qui se trouve dans le compte awsexamplebucket1 1234-5678-9012.

1. Dans le compte 1111-2222-3333, connectez-vous à la console IAM AWS Management Console et ouvrez-la à l'adresse. <https://console.aws.amazon.com/iam/>
2. Créez un nouveau rôle ou consultez un rôle existant qui répond aux exigences de [the section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”](#). Assurez-vous que le rôle accorde des autorisations à Amazon S3 sur awsexamplebucket1.
3. Ouvrez la console Amazon S3 à l'adresse <https://console.aws.amazon.com/s3/>. Connectez-vous avec le compte 1234-5678-9012.
4. Dans la liste Nom du compartiment, choisissez le nom du compartiment, awsexamplebucket1.
5. Choisissez Autorisations.
6. Sur la page Autorisations, choisissez Bucket Policy.
7. Dans l'éditeur de politique Bucket, collez la politique suivante. Remplacez *<role-name>* par le nom de votre rôle.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/<role-name>"
      },

```

```
        "Action": "s3:ListBucket",
        "Resource": "arn:aws:s3:::awsexamplebucket1"
    },
    {
        "Effect": "Allow",
        "Principal": {
            "AWS": "arn:aws:iam::111122223333:role/<role-name>"
        },
        "Action": [
            "s3:DeleteObject",
            "s3:GetObject",
            "s3:PutObject"
        ],
        "Resource": "arn:aws:s3:::awsexamplebucket1/*"
    }
]
```

8. Choisissez Enregistrer.
9. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous au compte 1111-2222-3333 en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant des autorisations suffisantes pour enregistrer des emplacements.
10. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
11. Sur la page des emplacements des Data Lake, choisissez Enregistrer l'emplacement.
12. Sur la page Enregistrer l'emplacement, pour le chemin Amazon S3, entrez le nom du compartiments3://awsexamplebucket1.

 Note

Vous devez saisir le nom du bucket car les buckets multi-comptes n'apparaissent pas dans la liste lorsque vous choisissez Browse.

13. Pour le rôle IAM, choisissez votre rôle.
14. Choisissez Enregistrer l'emplacement.

Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes

AWS Lake Formation s'intègre à [AWS Key Management Service](#) (AWS KMS) pour vous permettre de configurer plus facilement d'autres services intégrés pour chiffrer et déchiffrer les données sur les sites Amazon Simple Storage Service (Amazon S3).

Les deux clés sont gérées par le client et Clés gérées par AWS sont prises en charge. Le côté client n'encryption/decryption est pas pris en charge.

Important

Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Cette section explique comment enregistrer un emplacement Amazon S3 dans les circonstances suivantes :

- Les données de l'emplacement Amazon S3 sont chiffrées à l'aide d'une clé KMS créée dans AWS KMS.
- L'emplacement Amazon S3 n'est pas enregistré dans le même AWS compte que le AWS Glue Data Catalog.
- La clé KMS se trouve ou non dans le même AWS compte que le catalogue de données.

L'enregistrement d' AWS KMS un compartiment Amazon S3 chiffré dans le AWS compte B à l'aide d'un rôle Gestion des identités et des accès AWS (IAM) dans le AWS compte A nécessite les autorisations suivantes :

- Le rôle du compte A doit accorder des autorisations sur le compartiment du compte B.
- La politique de compartiment du compte B doit accorder des autorisations d'accès au rôle dans le compte A.
- Si la clé KMS se trouve dans le compte B, la politique clé doit accorder l'accès au rôle dans le compte A, et le rôle dans le compte A doit accorder des autorisations sur la clé KMS.

Dans la procédure suivante, vous allez créer un rôle dans le AWS compte qui contient le catalogue de données (compte A dans la discussion précédente). Vous utilisez ensuite ce rôle pour enregistrer l'emplacement. Lake Formation assume ce rôle lors de l'accès aux données sous-jacentes dans Amazon S3. Le rôle assumé dispose des autorisations requises sur la clé KMS. Par conséquent, vous n'avez pas à accorder d'autorisations sur la clé KMS aux principaux accédant aux données sous-jacentes via des tâches ETL ou des services intégrés tels que Amazon Athena.

⚠ Important

Vous ne pouvez pas utiliser le rôle lié au service Lake Formation pour enregistrer une position sur un autre compte. Vous devez plutôt utiliser un rôle défini par l'utilisateur. Le rôle doit répondre aux exigences de [the section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”](#). Pour de plus amples informations sur le rôle lié à un service, veuillez consulter [Autorisations de rôle liées à un service pour Lake Formation](#).

Avant de commencer

Passez en revue [les exigences relatives au rôle utilisé pour enregistrer l'emplacement](#).

Pour enregistrer une position Amazon S3 chiffrée sur plusieurs AWS comptes

1. Dans le même AWS compte que le catalogue de données, connectez-vous à la console IAM AWS Management Console et ouvrez-la à <https://console.aws.amazon.com/iam/> l'adresse.
2. Créez un nouveau rôle ou consultez un rôle existant qui répond aux exigences de [the section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”](#). Assurez-vous que le rôle inclut une politique qui accorde des autorisations à Amazon S3 sur le site.
3. Si la clé KMS ne se trouve pas dans le même compte que le catalogue de données, ajoutez au rôle une politique intégrée qui accorde les autorisations requises sur la clé KMS. Voici un exemple de politique . Remplacez la région et l'ID de compte par la région et le numéro de compte de la clé KMS. Remplacez *<key-id>* par l'identifiant de la clé.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
```

```

    "Effect": "Allow",
    "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
    ],
    "Resource": "arn:aws:kms:us-east-1:111122223333:key/<key-id>"
  }
]
}

```

4. Sur la console Amazon S3, ajoutez une politique de compartiment accordant les autorisations Amazon S3 requises au rôle. Voici un exemple de stratégie de compartiment. Remplacez l'ID de AWS compte par le numéro de compte du catalogue de données, *<role-name>* par le nom de votre rôle et *<bucket-name>* par le nom du bucket.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/<role-name>"
      },
      "Action": "s3:ListBucket",
      "Resource": "arn:aws:s3:::<bucket-name>"
    },
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/<role-name>"
      },
      "Action": [
        "s3:DeleteObject",
        "s3:GetObject",
        "s3:PutObject"
      ],
      "Resource": "arn:aws:s3:::<bucket-name>/*"
    }
  ]
}

```

```

    }
  ]
}

```

5. Dans AWS KMS, ajoutez le rôle en tant qu'utilisateur de la clé KMS.
 - a. Ouvrez la AWS KMS console à l'adresse <https://console.aws.amazon.com/kms>. Connectez-vous ensuite en tant qu'administrateur ou en tant qu'utilisateur pouvant modifier la politique de clé de la clé KMS utilisée pour chiffrer l'emplacement.
 - b. Dans le volet de navigation, choisissez Customer managed keys, puis choisissez le nom de la clé KMS.
 - c. Sur la page des détails de la clé KMS, sous l'onglet Stratégie clé, si la vue JSON de la politique clé ne s'affiche pas, choisissez Basculer vers la vue politique.
 - d. Dans la section Politique clé, choisissez Modifier et ajoutez le nom de ressource Amazon (ARN) du rôle à l'Allow use of the key objet, comme indiqué dans l'exemple suivant.

 Note

Si cet objet est manquant, ajoutez-le avec les autorisations indiquées dans l'exemple.

```

...
{
  "Sid": "Allow use of the key",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::<catalog-account-id>:role/<role-name>"
    ]
  },
  "Action": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
  ],
  "Resource": "*"
},

```

...

Pour plus d'informations, voir [Autoriser les utilisateurs d'autres comptes à utiliser une clé KMS](#) dans le Guide du AWS Key Management Service développeur.

6. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous au AWS compte Data Catalog en tant qu'administrateur du lac de données.
7. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
8. Choisissez Enregistrer l'emplacement.
9. Sur la page Enregistrer l'emplacement, pour le chemin Amazon S3, entrez le chemin de localisation sous la forme **s3://<bucket>/<prefix>**. **<bucket>** Remplacez-le par le nom du compartiment et **<prefix>** par le reste du chemin correspondant à l'emplacement.

 Note

Vous devez saisir le chemin car les compartiments multicomptes n'apparaissent pas dans la liste lorsque vous choisissez Parcourir.

10. Pour le rôle IAM, choisissez le rôle à l'étape 2.
11. Choisissez Enregistrer l'emplacement.

Annulation de l'enregistrement d'un site Amazon S3

Vous pouvez annuler l'enregistrement d'un site Amazon Simple Storage Service (Amazon S3) si vous ne souhaitez plus qu'il soit géré par Lake Formation. L'annulation de l'enregistrement d'un emplacement n'affecte pas les autorisations de localisation des données de Lake Formation accordées pour cet emplacement. Vous pouvez réenregistrer un emplacement que vous avez désenregistré, et les autorisations de localisation des données restent en vigueur. Vous pouvez utiliser un autre rôle pour réenregistrer l'emplacement.

Pour désenregistrer une position (console)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
2. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.

3. Sélectionnez un emplacement, puis dans le menu Actions, choisissez Supprimer.
4. Lorsque vous êtes invité à confirmer, choisissez Supprimer.

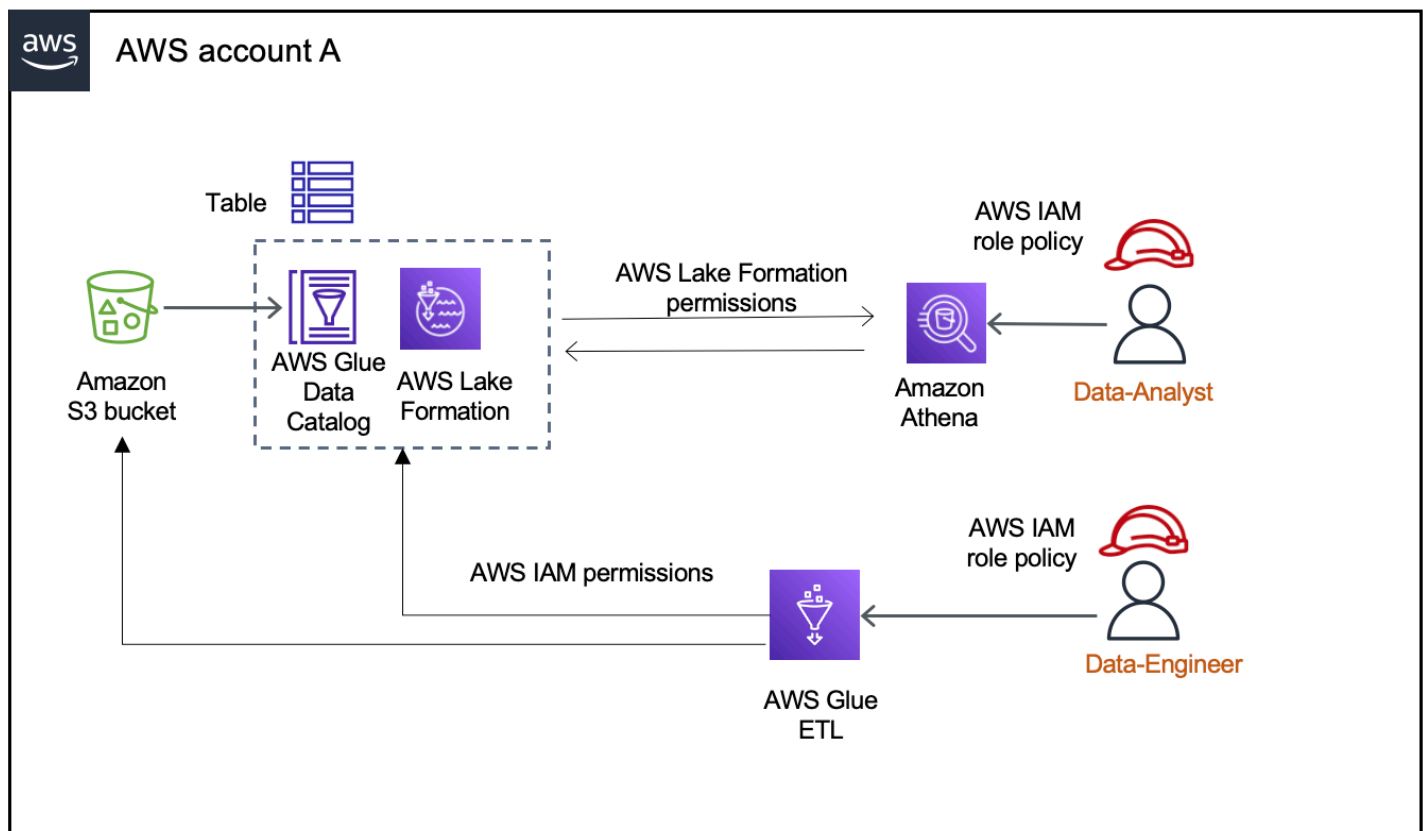
Mode d'accès hybride

AWS Lake Formation le mode d'accès hybride prend en charge deux voies d'autorisation vers les mêmes AWS Glue Data Catalog objets.

Dans le premier parcours, Lake Formation vous permet de sélectionner des principes spécifiques et de leur accorder les autorisations Lake Formation pour accéder aux catalogues, aux bases de données, aux tables et aux vues en vous inscrivant. La deuxième voie permet à tous les autres principaux d'accéder à ces ressources via les politiques principales IAM par défaut pour Amazon S3 et AWS Glue les actions.

Lorsque vous enregistrez un site Amazon S3 auprès de Lake Formation, vous avez la possibilité d'appliquer les autorisations de Lake Formation à toutes les ressources de cet emplacement ou d'utiliser le mode d'accès hybride. Le mode d'accès hybride applique uniquement `CREATE_TABLE` `CREATE_PARTITION` les `UPDATE_TABLE` autorisations par défaut. Lorsqu'un site Amazon S3 est en mode hybride, vous pouvez activer les autorisations de Lake Formation en choisissant des principes pour les objets du catalogue de données situés sous cet emplacement. Cela signifie que les autorisations Lake Formation et les autorisations IAM peuvent contrôler l'accès à ces données. Cela signifie que les principaux participants auront besoin à la fois des autorisations Lake Formation et des autorisations IAM pour accéder aux données, tandis que non-opted-in les principaux continueront d'accéder aux données en utilisant uniquement les autorisations IAM.

Ainsi, le mode d'accès hybride offre la flexibilité nécessaire pour activer de manière sélective Lake Formation pour les catalogues, les bases de données et les tables de votre catalogue de données pour un ensemble spécifique d'utilisateurs sans interrompre l'accès pour les autres utilisateurs ou charges de travail existants.



Pour les considérations et les restrictions, consultez [Considérations et limites relatives au mode d'accès hybride](#).

Termes et définitions

Voici les définitions des ressources du catalogue de données en fonction de la façon dont vous configurez les autorisations d'accès :

Ressource Lake Formation

Une ressource enregistrée auprès de Lake Formation. Les utilisateurs ont besoin des autorisations de Lake Formation pour accéder à la ressource.

AWS Glue ressource

Une ressource qui n'est pas enregistrée auprès de Lake Formation. Les utilisateurs n'ont besoin que d'autorisations IAM pour accéder à la ressource, car celle-ci dispose d'autorisations de `IAMAllowedPrincipals` groupe. Les autorisations de Lake Formation ne sont pas appliquées.

Pour plus d'informations sur les autorisations de `IAMAllowedPrincipals` groupe, consultez [Autorisations relatives aux métadonnées](#).

Ressource hybride

Une ressource enregistrée en mode d'accès hybride. En fonction des utilisateurs accédant à la ressource, celle-ci passe de manière dynamique à une ressource de Lake Formation ou à une AWS Glue ressource.

Cas d'utilisation courants du mode d'accès hybride

Vous pouvez utiliser le mode d'accès hybride pour fournir un accès dans le cadre de scénarios de partage de données à compte unique ou entre comptes :

Scénarios de compte unique

- Convertir une AWS Glue ressource en ressource hybride : dans ce scénario, vous n'utilisez pas actuellement Lake Formation mais souhaitez adopter les autorisations Lake Formation pour les objets du catalogue de données. Lorsque vous enregistrez l'emplacement Amazon S3 en mode d'accès hybride, vous pouvez accorder des autorisations Lake Formation aux utilisateurs qui optent pour des bases de données et des tables spécifiques pointant vers cet emplacement.
- Conversion d'une ressource Lake Formation en ressource hybride — Actuellement, vous utilisez les autorisations Lake Formation pour contrôler l'accès à une base de données de catalogue de données, mais vous souhaitez fournir l'accès à de nouveaux principaux en utilisant les autorisations IAM pour Amazon S3 et AWS Glue sans interrompre les autorisations Lake Formation existantes.

Lorsque vous mettez à jour l'enregistrement d'un emplacement de données en mode d'accès hybride, les nouveaux principaux peuvent accéder à la base de données du catalogue de données pointant vers l'emplacement Amazon S3 en utilisant les politiques d'autorisation IAM sans interrompre les autorisations Lake Formation des utilisateurs existants.

Avant de mettre à jour l'enregistrement de la localisation des données pour activer le mode d'accès hybride, vous devez d'abord activer les principaux qui accèdent actuellement à la ressource avec les autorisations de Lake Formation.

Cela permet d'éviter toute interruption potentielle du flux de travail en cours.

Vous devez également Super autoriser le `IAMAllowedPrincipal` groupe à accéder aux tables de la base de données.

Scénarios de partage de données entre comptes

- Partagez AWS Glue des ressources à l'aide du mode d'accès hybride : dans ce scénario, le compte producteur possède des tables dans une base de données qui sont actuellement partagées avec un compte client conformément aux politiques d'autorisation IAM pour Amazon S3 et aux AWS Glue actions. L'emplacement des données de la base de données n'est pas enregistré auprès de Lake Formation.

Avant d'enregistrer l'emplacement des données en mode d'accès hybride, vous devez mettre à jour les paramètres de version du compte Cross vers la version 4. La version 4 fournit les nouvelles politiques AWS RAM d'autorisation requises pour le partage entre comptes lorsque le `IAMAllowedPrincipal` groupe dispose d'une `Super` autorisation sur la ressource. Pour les ressources disposant d'autorisations de `IAMAllowedPrincipal` groupe, vous pouvez accorder des autorisations de Lake Formation à des comptes externes et les autoriser à utiliser les autorisations de Lake Formation. L'administrateur du lac de données du compte destinataire peut accorder des autorisations Lake Formation aux principaux du compte et les autoriser à appliquer les autorisations Lake Formation.

- Partagez les ressources de Lake Formation en mode d'accès hybride — Actuellement, le compte producteur contient des tables dans une base de données qui sont partagées avec un compte consommateur appliquant les autorisations de Lake Formation. L'emplacement des données de la base de données est enregistré auprès de Lake Formation.

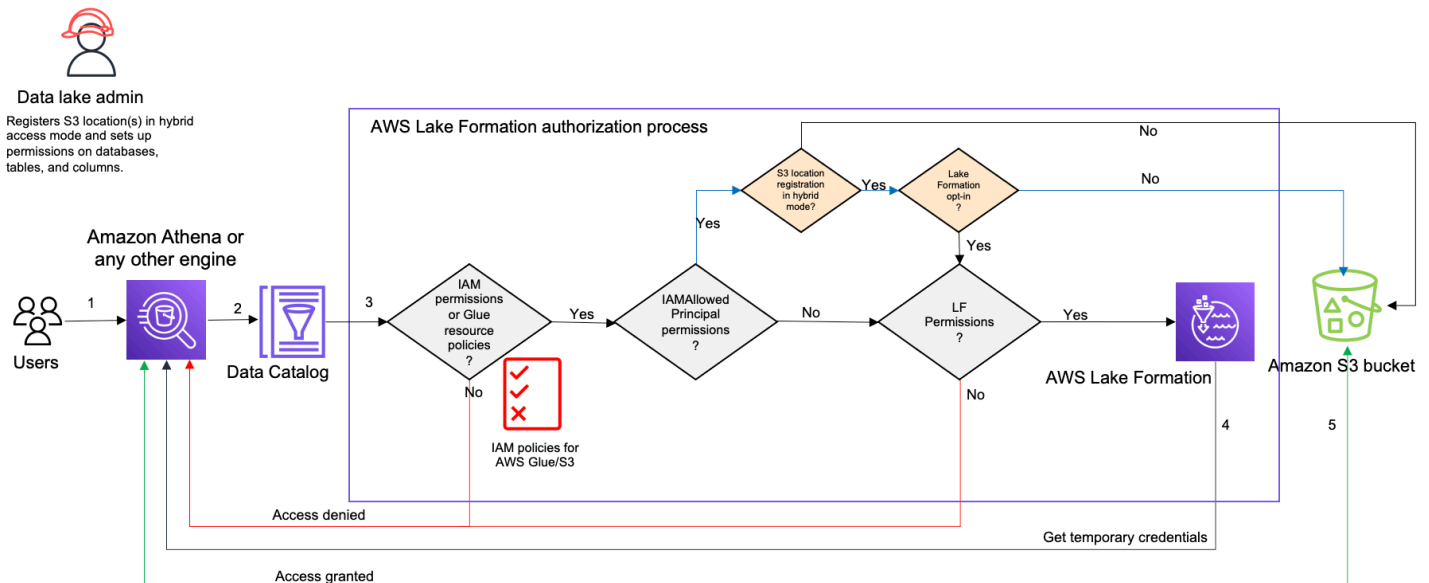
Dans ce cas, vous pouvez mettre à jour l'enregistrement de l'emplacement Amazon S3 en mode d'accès hybride et partager les données d'Amazon S3 et les métadonnées de Data Catalog en utilisant les politiques de compartiment Amazon S3 et les politiques de ressources du catalogue de données avec les principaux du compte client. Vous devez réoctroyer les autorisations Lake Formation existantes et accepter les principales avant de mettre à jour l'enregistrement de la position Amazon S3. Vous devez également `Super` autoriser le `IAMAllowedPrincipals` groupe à accéder aux tables de la base de données.

Rubriques

- [Comment fonctionne le mode d'accès hybride](#)
- [Configuration du mode d'accès hybride : scénarios courants](#)
- [Supprimer les principes et les ressources du mode d'accès hybride](#)
- [Affichage des principes et des ressources en mode d'accès hybride](#)
- [Ressources supplémentaires](#)

Comment fonctionne le mode d'accès hybride

Le schéma suivant montre comment fonctionne l'autorisation de Lake Formation en mode d'accès hybride lorsque vous interrogez les ressources du catalogue de données.



Avant d'accéder aux données de votre lac de données, un administrateur de lac de données ou un utilisateur disposant d'autorisations administratives définit des politiques utilisateur individuelles pour les tables du catalogue de données afin d'autoriser ou de refuser l'accès aux tables de votre catalogue de données. Ensuite, un directeur autorisé à effectuer une `RegisterResource` opération enregistre l'emplacement de la table sur Amazon S3 auprès de Lake Formation en mode d'accès hybride. Si aucun emplacement de données n'est enregistré auprès de Lake Formation, l'administrateur accorde des autorisations Lake Formation à des utilisateurs spécifiques sur les bases de données et les tables du catalogue de données et les autorise à utiliser les autorisations de Lake Formation pour ces bases de données et tables en mode d'accès hybride.

1. Soumet une requête : un directeur soumet une requête ou un script ETL à l'aide d'un service intégré tel qu'Amazon Athena, AWS Glue Amazon EMR ou Amazon Redshift Spectrum.
2. Demande des données : le moteur d'analyse intégré identifie la table demandée et envoie la demande de métadonnées au catalogue de données (`GetTable`, `GetDatabase`).
3. Vérifie les autorisations - Le catalogue de données vérifie les autorisations d'accès du principal demandeur auprès de Lake Formation.
 - a. Si aucune autorisation de `IAMAllowedPrincipals` groupe n'est attachée à la table, les autorisations de Lake Formation sont appliquées.

- b. Si le principal a choisi d'utiliser les autorisations de Lake Formation en mode d'accès hybride et que des autorisations de `IAMAllowedPrincipals` groupe sont associées à la table, les autorisations de Lake Formation sont appliquées. Le moteur de requête applique les filtres qu'il a reçus de Lake Formation et renvoie les données à l'utilisateur.
- c. Si l'emplacement de la table n'est pas enregistré auprès de Lake Formation et que le directeur n'a pas choisi d'utiliser les autorisations de Lake Formation en mode d'accès hybride, le catalogue de données vérifie si des autorisations de `IAMAllowedPrincipals` groupe sont associées à la table. Si cette autorisation existe sur la table, tous les principaux du compte obtiennent `Super` ou obtiennent `All` des autorisations sur la table.

La vente d'identifiants Lake Formation n'est pas disponible, même en cas d'inscription, sauf si l'emplacement des données est enregistré auprès de Lake Formation.

4. Obtenir des informations d'identification — Le catalogue de données vérifie et indique au moteur si l'emplacement de la table est enregistré auprès de Lake Formation ou non. Si les données sous-jacentes sont enregistrées auprès de Lake Formation, le moteur d'analyse demande à Lake Formation des informations d'identification temporaires pour accéder aux données du compartiment Amazon S3.
5. Obtenir des données — Si le principal est autorisé à accéder aux données de la table, Lake Formation fournit un accès temporaire au moteur d'analyse intégré. À l'aide de l'accès temporaire, le moteur d'analyse extrait les données d'Amazon S3 et effectue le filtrage nécessaire, tel que le filtrage par colonne, ligne ou cellule. Lorsque le moteur a terminé d'exécuter la tâche, il renvoie les résultats à l'utilisateur. Ce processus est appelé distributeur d'informations d'identification. Pour plus d'informations, voir [Intégration à Lake Formation](#).
6. Si l'emplacement des données de la table n'est pas enregistré auprès de Lake Formation, le deuxième appel du moteur d'analyse est directement envoyé à Amazon S3. La politique de compartiment Amazon S3 et la politique utilisateur IAM concernées sont évaluées pour l'accès aux données. Chaque fois que vous utilisez des politiques IAM, veillez à respecter les bonnes pratiques IAM. Pour plus d'informations, consultez [la section Bonnes pratiques en matière de sécurité dans IAM dans le guide de l'utilisateur d'IAM](#).

Configuration du mode d'accès hybride : scénarios courants

Comme pour les autorisations de Lake Formation, vous disposez généralement de deux types de scénarios dans lesquels vous pouvez utiliser le mode d'accès hybride pour gérer l'accès aux

données : fournir un accès aux principaux au sein d'un seul Compte AWS et fournir un accès à un externe Compte AWS ou à un principal.

Cette section fournit des instructions pour configurer le mode d'accès hybride dans les scénarios suivants :

Gérez les autorisations en mode d'accès hybride en une seule fois Compte AWS

- [Conversion d'une AWS Glue ressource en ressource hybride](#)— Vous fournissez actuellement l'accès aux tables d'une base de données à tous les principaux de votre compte à l'aide des autorisations IAM pour Amazon S3, AWS Glue mais vous souhaitez adopter Lake Formation pour gérer les autorisations de manière progressive.
- [Conversion d'une ressource de la Lake Formation en une ressource hybride](#)— Vous utilisez actuellement Lake Formation pour gérer l'accès aux tables d'une base de données pour tous les principaux de votre compte, mais vous souhaitez utiliser Lake Formation uniquement pour des principes spécifiques. Vous souhaitez donner accès à de nouveaux principaux en utilisant les autorisations IAM pour AWS Glue Amazon S3 sur la même base de données et les mêmes tables.

Gérez les autorisations en mode d'accès hybride sur Compte AWS s

- [Partage d'une AWS Glue ressource à l'aide du mode d'accès hybride](#)— Vous n'utilisez actuellement pas Lake Formation pour gérer les autorisations associées à une table, mais vous souhaitez appliquer les autorisations Lake Formation pour permettre aux principaux d'accéder à un autre compte.
- [Partage d'une ressource de Lake Formation à l'aide du mode d'accès hybride](#)— Vous utilisez Lake Formation pour gérer l'accès à une table, mais vous souhaitez fournir l'accès aux principaux d'un autre compte en utilisant les autorisations IAM pour AWS Glue Amazon S3 sur la même base de données et les mêmes tables.

Configuration du mode d'accès hybride — Étapes de haut niveau

1. Enregistrez l'emplacement des données Amazon S3 auprès de Lake Formation en sélectionnant le mode d'accès hybride.
2. Les principaux doivent être DATA_LOCATION autorisés à accéder à l'emplacement d'un lac de données pour créer des tables ou des bases de données du catalogue de données pointant vers cet emplacement.

3. Définissez le paramètre de version multi-comptes sur Version 4.
4. Accordez des autorisations détaillées à des utilisateurs ou à des rôles IAM spécifiques sur les bases de données et les tables. Dans le même temps, assurez-vous de définir Super des All autorisations pour le IAMAllowedPrincipals groupe sur la base de données et sur toutes les tables ou sur certaines d'entre elles.
5. Choisissez les principes et les ressources. Les autres principaux du compte peuvent continuer à accéder aux bases de données et aux tables en utilisant les politiques d'autorisation IAM pour AWS Glue les actions Amazon S3.
6. Nettoyez éventuellement les politiques d'autorisation IAM pour Amazon S3 pour les principaux utilisateurs qui ont choisi d'utiliser les autorisations de Lake Formation.

Conditions préalables à la configuration du mode d'accès hybride

Les conditions requises pour configurer le mode d'accès hybride sont les suivantes :

Note

Nous recommandons à un administrateur de Lake Formation d'enregistrer l'emplacement Amazon S3 en mode d'accès hybride et d'opter pour les principes et les ressources.

1. Accordez l'autorisation de localisation des données (DATA_LOCATION_ACCESS) pour créer des ressources de catalogue de données pointant vers les sites Amazon S3. Les autorisations de localisation des données contrôlent la possibilité de créer des catalogues de données, des bases de données et des tables pointant vers des emplacements Amazon S3 particuliers.
2. Pour partager les ressources du catalogue de données avec un autre compte en mode d'accès hybride (sans supprimer les autorisations de IAMAllowedPrincipals groupe associées à la ressource), vous devez mettre à jour les paramètres de version multi-comptes vers la version 4 ou supérieure. Pour mettre à jour la version à l'aide de la console Lake Formation, choisissez la version 4 ou la version 5 dans les paramètres de version entre comptes sur la page des paramètres du catalogue de données.

Vous pouvez également utiliser la `put-data-lake-settings` AWS CLI commande pour définir le `CROSS_ACCOUNT_VERSION` paramètre sur la version 4 ou 5 :

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
file://settings
```



```
{
  "DataLakeAdmins": [
    {
      "DataLakePrincipalIdentifier": "arn:aws:iam::<111122223333>:user/<user-name>"
    }
  ],
  "CreateDatabaseDefaultPermissions": [],
  "CreateTableDefaultPermissions": [],
  "Parameters": {
    "CROSS_ACCOUNT_VERSION": "5"
  }
}
```

3.

Pour accorder des autorisations entre comptes en mode d'accès hybride, le concédant doit disposer des autorisations IAM requises pour AWS Glue les services et les services. AWS RAM La politique AWS gérée AWSLakeFormationCrossAccountManager accorde les autorisations requises.

Pour permettre le partage de données entre comptes en mode d'accès hybride, nous avons mis à jour la politique AWSLakeFormationCrossAccountManager gérée en ajoutant deux nouvelles autorisations IAM :

- RAM : ListResourceSharePermissions
- RAM : AssociateResourceSharePermission

Note

Si vous n'utilisez pas la politique AWS gérée pour le rôle de donateur, ajoutez les politiques ci-dessus à vos politiques personnalisées.

Emplacement du compartiment Amazon S3 et accès utilisateur

Lorsque vous créez un catalogue, une base de données ou une table dans le AWS Glue Data Catalog, vous pouvez spécifier l'emplacement du compartiment Amazon S3 des données sous-jacentes et les enregistrer auprès de Lake Formation. Les tableaux ci-dessous décrivent le fonctionnement des autorisations pour AWS Glue les utilisateurs (principaux) de Lake Formation en fonction de l'emplacement des données Amazon S3 de la table ou de la base de données.

Emplacement Amazon S3 enregistré auprès de Lake Formation

Emplacement d'une base de données sur Amazon S3	AWS Glue utilisateurs	Utilisateurs de Lake Formation
Enregistré auprès de Lake Formation (en mode d'accès hybride ou en mode Lake Formation)	read/write Accédez à l'emplacement des données Amazon S3 en héritant des autorisations du groupe IAMAllowed Principals (super accès).	Héritez des autorisations permettant de créer des tables à partir de l'autorisation CREATE TABLE qui leur a été accordée.
Aucun emplacement Amazon S3 associé	L'autorisation DATA LOCATION explicite est requise pour exécuter les instructions CREATE TABLE et INSERT TABLE.	L'autorisation DATA LOCATION explicite est requise pour exécuter les instructions CREATE TABLE et INSERT TABLE.

IsRegisteredWithLakeFormation propriété de la table

La `IsRegisteredWithLakeFormation` propriété d'une table indique si l'emplacement des données de la table est enregistré auprès de Lake Formation pour le demandeur. Si le mode d'autorisation de l'emplacement est enregistré sous le nom de Lake Formation, la `IsRegisteredWithLakeFormation` propriété est `true` réservée à tous les utilisateurs accédant à l'emplacement des données, car tous les utilisateurs sont considérés comme ayant opté pour cette table. Si l'emplacement est enregistré en mode d'accès hybride, la valeur est définie sur `true` uniquement pour les utilisateurs qui ont opté pour cette table.

Fonctionnement d'`IsRegisteredWithLakeFormation`

Mode d'autorisation	Utilisateurs/Rôles	IsRegisteredWithLakeFormation	Description
Lake Formation	Tous	True	Lorsqu'un emplacement est enregistré auprès de Lake Formation

Mode d'autorisation	Utilisateurs/Rôles	IsRegisteredWithLakeFormation	Description
			, la IsRegisteredWithLakeFormation propriété est définie sur <code>true</code> pour tous les utilisateurs. Cela signifie que les autorisations définies dans Lake Formation s'appliquent à l'emplacement enregistré. La vente des accréditations sera effectuée par Lake Formation.
Mode d'accès hybride	Inscrit	True	Pour les utilisateurs qui ont choisi d'utiliser Lake Formation pour l'accès aux données et la gouvernance d'une table, la IsRegisteredWithLakeFormation propriété sera définie sur <code>true</code> pour cette table. Ils sont soumis aux politiques d'autorisation définies dans Lake Formation pour le site enregistré.

Mode d'autorisation	Utilisateurs/Rôles	IsRegisteredWithLakeFormation	Description
Mode d'accès hybride	Non inscrit	False	Pour les utilisateurs qui n'ont pas choisi d'utiliser les autorisations de Lake Formation, la <code>IsRegisteredWithLakeFormation</code> propriété est définie sur <code>false</code> . Ils ne sont pas soumis aux politiques d'autorisation définies dans Lake Formation pour le site enregistré. Les utilisateurs suivront plutôt les politiques d'autorisation d'Amazon S3.

Conversion d'une AWS Glue ressource en ressource hybride

Suivez ces étapes pour enregistrer un site Amazon S3 en mode d'accès hybride et intégrer de nouveaux utilisateurs de Lake Formation sans interrompre l'accès aux données des utilisateurs existants du catalogue de données.

Description du scénario - L'emplacement des données n'est pas enregistré auprès de Lake Formation, et l'accès des utilisateurs à la base de données et aux tables du catalogue de données est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions. Le `IAMAllowedPrincipals` groupe dispose par défaut d'`Super` autorisations sur toutes les tables de la base de données.

Pour activer le mode d'accès hybride pour un emplacement de données non enregistré auprès de Lake Formation

1. Enregistrez un emplacement Amazon S3 permettant le mode d'accès hybride.

Console

1. Connectez-vous à la [console Lake Formation](#) en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sélectionnez Emplacements des lacs de données sous Administration.
3. Choisissez Enregistrer l'emplacement.

Register location

Amazon S3 location

Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path

Choose an Amazon S3 path for your data lake.

e.g.: s3://bucket/prefix/

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess ▼

 Do not select the service linked role if you plan to use EMR.

☐ **Enable Data Catalog Federation**

Checking this box will allow Lake Formation to assume a role to access tables in a federated database.

Permission mode

Select the permission mode you want to use to manage access.

☒ **Hybrid access mode - new**

Lake Formation permissions can co-exist with IAM permission policies for AWS Glue and S3 actions to manage access. [Learn more](#) 

☐ **Lake Formation**

Only Lake Formation permissions are enforced.

Cancel

Register location

4. Dans la fenêtre Enregistrer l'emplacement, choisissez le chemin Amazon S3 que vous souhaitez enregistrer auprès de Lake Formation.
5. Pour le rôle IAM, choisissez le rôle **AWSServiceRoleForLakeFormationDataAccess** lié au service (par défaut) ou un rôle IAM personnalisé rôle qui répond aux exigences de [Exigences relatives aux rôles utilisés pour enregistrer des sites](#).

6. Choisissez le mode d'accès hybride pour appliquer des politiques précises de contrôle d'accès à Lake Formation aux principes d'adhésion ainsi qu'aux bases de données et aux tables du catalogue de données pointant vers l'emplacement enregistré.

Choisissez Lake Formation pour permettre à Lake Formation d'autoriser les demandes d'accès à l'emplacement enregistré.

7. Choisissez Enregistrer l'emplacement.

AWS CLI

Voici un exemple d'enregistrement d'un emplacement de données auprès de Lake Formation HybridAccessEnabled avec:true/false. La valeur par défaut du HybridAccessEnabled paramètre est false. Remplacez le chemin, le nom du rôle et l'identifiant du AWS compte Amazon S3 par des valeurs valides.

```
aws lakeformation register-resource --cli-input-json file:file path
json:
{
  "ResourceArn": "arn:aws:s3:::s3-path",
  "UseServiceLinkedRole": false,
  "RoleArn": "arn:aws:iam::<123456789012>:role/<role-name>",
  "HybridAccessEnabled": true
}
```

2. Accordez des autorisations et autorisez les principaux à utiliser les autorisations de Lake Formation pour les ressources en mode d'accès hybride

Avant d'activer les principes et les ressources en mode d'accès hybride, vérifiez que les bases de données et les tables dont l'emplacement est enregistré auprès de Lake Formation en mode d'accès hybride existent Super ou que les All autorisations de IAMAllowedPrincipals regroupement existent.

Note

Vous ne pouvez pas accorder d'autorisation au IAMAllowedPrincipals groupe All tables dans une base de données. Vous devez sélectionner chaque table séparément dans le menu déroulant et accorder des autorisations. En outre, lorsque vous créez de nouvelles tables dans la base de données, vous pouvez choisir de les utiliser Use only IAM access control for new tables in new databases dans les paramètres

du catalogue de données. Cette option accorde automatiquement Super l'autorisation au IAMAllowedPrincipals groupe lorsque vous créez de nouvelles tables dans la base de données.

Console

1. Sur la console Lake Formation, sous Data Catalog, sélectionnez Catalogues, Databases ou Tables.
2. Sélectionnez un catalogue, une base de données ou une table dans la liste, puis choisissez Grant dans le menu Actions.
3. Choisissez des principes pour accorder des autorisations sur la base de données, les tables et les colonnes à l'aide d'une méthode de ressource nommée ou de balises LF.

Vous pouvez également choisir Autorisations relatives aux données, sélectionner les principaux auxquels accorder les autorisations dans la liste, puis sélectionner Accorder.

Pour plus de détails sur l'octroi d'autorisations de données, consultez [Octroi d'autorisations sur les ressources du catalogue de données](#).

Note

Si vous accordez à un principal l'autorisation de créer une table, vous devez également accorder des autorisations de localisation des données (DATA_LOCATION_ACCESS) au principal. Cette autorisation n'est pas nécessaire pour mettre à jour les tables.

Pour de plus amples informations, veuillez consulter [Octroi d'autorisations de localisation des données](#).

4. Lorsque vous utilisez la méthode des ressources nommées pour accorder des autorisations, l'option permettant d'activer les principes et les ressources est disponible dans la section inférieure de la page d'autorisation des données d'octroi.

Choisissez Rendre les autorisations Lake Formation effectives immédiatement pour activer les autorisations Lake Formation pour les principaux et les ressources.

Hybrid access mode

In hybrid access mode, Lake Formation and IAM policies for AWS Glue and S3 work together.

☒ **Make Lake Formation permissions effective immediately**

Lake Formation permissions are enforced for databases, tables, and principals.

ⓘ You might get access denied.

If the checkbox is selected, your Lake Formation permissions are enforced. Make sure that you've completed the required setup for Lake Formation for the permissions to work. If the checkbox is clear, you can go to [hybrid access mode](#) to add resources and principals. [Learn more](#)

Cancel

Grant

5. Choisissez Accorder.

Lorsque vous activez le principal A sur la table A qui pointe vers un emplacement de données, cela permet au principal A d'accéder à l'emplacement de cette table en utilisant les autorisations de Lake Formation si l'emplacement des données est enregistré en mode hybride.

AWS CLI

Voici un exemple d'activation d'un principal et d'une table en mode d'accès hybride.

Remplacez le nom du rôle, l'identifiant du AWS compte, le nom de la base de données et le nom de la table par des valeurs valides.

```
aws lakeformation create-lake-formation-opt-in --cli-input-json file://file path
json:
{
  "Principal": {
    "DataLakePrincipalIdentifier":
    "arn:aws:iam::<123456789012>:role/<hybrid-access-role>"
  },
  "Resource": {
    "Table": {
      "CatalogId": "<123456789012>",
      "DatabaseName": "<hybrid_test>",
      "Name": "<hybrid_test_table>"
    }
  }
}
```

```
}
```

- a. (Optional) Si vous choisissez les balises LF pour octroyer des autorisations, vous pouvez activer les principes pour utiliser les autorisations de Lake Formation lors d'une étape séparée. Vous pouvez le faire en choisissant le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche.
- b. Dans la section inférieure de la page du mode d'accès hybride, choisissez Ajouter pour ajouter des ressources et des principes au mode d'accès hybride.
- c. Sur la page Ajouter des ressources et des principes, choisissez les catalogues, les bases de données et les tables enregistrés en mode d'accès hybride.

Vous pouvez choisir d'autoriser l'accès All tables dans une base de données.

Add resources and principals

Choose databases, tables, and principals to add in hybrid access mode. Lake Formation permissions will be enforced.

[Learn more](#) 

Resources

Catalogs

Databases

Select one or more databases.

Choose databases

testdb

X

Tables - optional

Select one or more tables.

Choose tables

testtable

X

- d. Choisissez les principaux et acceptez d'utiliser les autorisations de Lake Formation en mode d'accès hybride.
 - Principaux — Vous pouvez choisir les utilisateurs et les rôles IAM dans le même compte ou dans un autre compte. Vous pouvez également choisir des utilisateurs et des groupes SAML.

- Attributs : sélectionnez les attributs pour accorder des autorisations en fonction des attributs.

Principals

☐ **IAM users and roles**
Users or roles from this AWS account.

☐ **External accounts**
AWS account, AWS organization or IAM principal outside of this account.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☒ **Attributes - new**
Define permissions using attributes (IAM session tags) with a Cedar policy expression.

Attribute expression

Add tag key-value pairs from IAM roles or [IAM session tags](#). Separate values with commas. **AND** logic applies between pairs, **OR** logic within keys. Verify tag entries for accuracy.

Key	Value
Department	Sales

Add key-value pair

Expression review

The attribute expression above will be written in the following Cedar policy. Use CLI/SDK to edit or write custom Cedar policy.

```
permit (principal, action, resource)
when {
  context.iam.principalTags.hasTag("Department")
  && context.iam.principalTags.getTag("Department") == "Sales"
};
```

Permission scope

☒ This account

☐ External account

- Entrez la paire clé-valeur pour créer une subvention basée sur les attributs. Passez en revue l'expression de politique Cedar sur la console. Pour plus d'informations sur le cèdre, voir [Qu'est-ce que le cèdre ? | Référence linguistique des politiques de Cedar GuideLink](#).
- Choisissez Ajouter.

L'accès est accordé à tous rôles/utilisateurs IAM dont les attributs correspondent.

- e. Choisissez Ajouter.

Conversion d'une ressource de la Lake Formation en une ressource hybride

Dans les cas où vous utilisez actuellement les autorisations Lake Formation pour les bases de données et les tables de votre catalogue de données, vous pouvez modifier les propriétés d'enregistrement des emplacements pour activer le mode d'accès hybride. Cela vous permet de fournir aux nouveaux principaux l'accès aux mêmes ressources en utilisant les politiques d'autorisation IAM pour Amazon S3 et les AWS Glue actions sans interrompre les autorisations Lake Formation existantes.

Description du scénario - Les étapes suivantes supposent que vous avez un emplacement de données enregistré auprès de Lake Formation et que vous avez défini des autorisations pour les principaux sur les bases de données, les tables ou les colonnes pointant vers cet emplacement. Si l'emplacement a été enregistré avec un rôle lié à un service, vous ne pouvez pas mettre à jour les paramètres de localisation et activer le mode d'accès hybride. Le `IAMAllowedPrincipals` groupe dispose par défaut de super autorisations sur la base de données et toutes ses tables.

Important

Ne mettez pas à jour l'enregistrement d'un emplacement en mode d'accès hybride sans avoir sélectionné les principaux utilisateurs qui accèdent aux données dans cet emplacement.

Activation du mode d'accès hybride pour un emplacement de données enregistré auprès de Lake Formation

1.

Warning

Nous ne recommandons pas de convertir un emplacement de données géré par Lake Formation en mode d'accès hybride pour éviter d'interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Ajoutez les principaux existants qui ont les autorisations de Lake Formation.

1. Répertoriez et passez en revue les autorisations que vous avez accordées aux principaux sur les catalogues, les bases de données et les tables. Pour de plus amples informations, veuillez consulter [Affichage des autorisations de base de données et de tables dans Lake Formation](#).

2. Choisissez le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche, puis choisissez Ajouter.
 3. Sur la page Ajouter des principes et des ressources, choisissez les catalogues, les bases de données et les tables de l'emplacement de données Amazon S3 que vous souhaitez utiliser en mode d'accès hybride. Choisissez les principaux qui disposent déjà des autorisations de Lake Formation.
 4. Choisissez Ajouter pour activer les principes permettant d'utiliser les autorisations de Lake Formation en mode d'accès hybride.
2. Mettez à jour l' bucket/prefix enregistrement Amazon S3 en choisissant l'option Mode d'accès hybride.

Console

1. Connectez-vous à la console Lake Formation en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
3. Sélectionnez un emplacement, puis dans le menu Actions, choisissez Modifier.
4. Choisissez le mode d'accès hybride.
5. Choisissez Enregistrer.
6. Sous Catalogue de données, sélectionnez la base de données ou la table et accordez Super All des autorisations au groupe virtuel appelé `IAMAllowedPrincipals`.
7. Vérifiez que l'accès de vos utilisateurs actuels de Lake Formation n'est pas interrompu lorsque vous avez mis à jour les propriétés d'enregistrement de la localisation. Connectez-vous à la console Athena en tant que responsable de Lake Formation et exécutez un exemple de requête sur une table pointant vers l'emplacement mis à jour.

De même, vérifiez l'accès des AWS Glue utilisateurs qui utilisent les politiques d'autorisation IAM pour accéder à la base de données et aux tables.

AWS CLI

Voici un exemple d'enregistrement d'un emplacement de données auprès de Lake Formation `HybridAccessEnabled` avec `true/false`. La valeur par défaut du `HybridAccessEnabled` paramètre est `false`. Remplacez le chemin, le nom du rôle et l'identifiant du AWS compte Amazon S3 par des valeurs valides.

```
aws lakeformation update-resource --cli-input-json file://file path
```

```
json:
{
  "ResourceArn": "arn:aws:s3:::<s3-path>",
  "RoleArn": "arn:aws:iam::<123456789012>:role/<test>",
  "HybridAccessEnabled": true
}
```

Partage d'une AWS Glue ressource à l'aide du mode d'accès hybride

Partagez des données avec une autre personne Compte AWS ou un responsable d'un autre en Compte AWS appliquant les autorisations de Lake Formation sans interrompre l'accès basé sur l'IAM des utilisateurs existants du catalogue de données.

Description du scénario - Le compte producteur dispose d'une base de données de catalogue de données dont l'accès est contrôlé à l'aide des principales politiques et AWS Glue actions IAM pour Amazon S3. L'emplacement des données de la base de données n'est pas enregistré auprès de Lake Formation. Le `IAMAllowedPrincipals` groupe dispose par défaut d'Superautorisations sur la base de données et sur toutes ses tables.

Octroi d'autorisations entre comptes Lake Formation en mode d'accès hybride

1. Configuration du compte producteur

1. Connectez-vous à la console Lake Formation à l'aide d'un rôle autorisé par `lakeformation:PutDataLakeSettings` IAM.
2. Accédez aux paramètres du catalogue de données et choisissez `Version 4` les paramètres de version entre comptes.

Si vous utilisez actuellement la version 1 ou 2, consultez [Mise à jour des paramètres de version de partage de données entre comptes](#) les instructions relatives à la mise à jour vers la version 3.

Aucune modification de la politique d'autorisation n'est requise lors de la mise à niveau de la version 3 vers la version 4.

3. Enregistrez l'emplacement Amazon S3 de la base de données ou de la table que vous prévoyez de partager en mode d'accès hybride.

4. Vérifiez que l'`Superautorisation d'accès au IAMAllowedPrincipals` groupe existe sur les bases de données et les tables dont vous avez enregistré l'emplacement des données en mode d'accès hybride à l'étape ci-dessus.
5. Accordez des autorisations Lake Formation à AWS des organisations, à des unités organisationnelles (OUs) ou directement auprès d'un directeur IAM sur un autre compte.
6. Si vous accordez des autorisations directement à un directeur IAM, activez le principal depuis le compte client pour appliquer les autorisations Lake Formation en mode d'accès hybride en activant l'option Rendre les autorisations Lake Formation effectives immédiatement.

Si vous accordez des autorisations entre comptes à un autre AWS compte, lorsque vous activez le compte, les autorisations de Lake Formation ne sont appliquées qu'aux administrateurs de ce compte. L'administrateur du lac de données du compte destinataire doit répartir les autorisations en cascade et sélectionner les principaux du compte pour appliquer les autorisations de Lake Formation aux ressources partagées en mode d'accès hybride.

Si vous choisissez l'option Ressources correspondant à des balises LF pour accorder des autorisations entre comptes, vous devez d'abord terminer l'étape d'octroi des autorisations. Vous pouvez activer les principes et les ressources pour le mode d'accès hybride lors d'une étape séparée en choisissant le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche de la console Lake Formation. Choisissez ensuite Ajouter pour ajouter les ressources et les principes auxquels vous souhaitez appliquer les autorisations de Lake Formation.

2. Configuration du compte client

1. Connectez-vous à la console Lake Formation en <https://console.aws.amazon.com/lakeformation/> tant qu'administrateur de lac de données.
2. Rentrez <https://console.aws.amazon.com/ram/chez> vous et acceptez l'invitation de partage de ressources. L'onglet Partagé avec moi de la AWS RAM console affiche la base de données et les tables partagées avec votre compte.
3. Créez un lien de ressource vers la and/or table de base de données partagée dans Lake Formation.
4. Accordez `Describe` l'autorisation sur le lien vers la ressource et l'`Grant on targetautorisation` (sur la ressource partagée d'origine) aux responsables IAM de votre compte (consommateur).
5. Accordez aux responsables de votre compte des autorisations sur la base de données ou la table partagée avec vous à Lake Formation. Choisissez les principes et les ressources

nécessaires pour appliquer les autorisations de Lake Formation en mode d'accès hybride en activant l'option Rendre les autorisations Lake Formation effectives immédiatement.

6. Testez les autorisations du principal sur la Lake Formation en exécutant des exemples de requêtes Athena. Testez l'accès existant de vos AWS Glue utilisateurs à l'aide des principales politiques et AWS Glue actions IAM pour Amazon S3.

(Facultatif) Supprimez la politique de compartiment Amazon S3 pour l'accès aux données et les politiques principales IAM AWS Glue et d'accès aux données Amazon S3 pour les principaux que vous avez configurés pour utiliser les autorisations de Lake Formation.

Partage d'une ressource de Lake Formation à l'aide du mode d'accès hybride

Autorisez les nouveaux utilisateurs du catalogue de données d'un compte externe à accéder aux bases de données et aux tables du catalogue de données à l'aide de politiques basées sur l'IAM sans interrompre les autorisations de partage entre comptes existantes de Lake Formation.

Description du scénario - Le compte producteur possède une base de données gérée par Lake Formation et des tables partagées avec un compte externe (consommateur) au niveau du compte ou au niveau principal de l'IAM. L'emplacement des données de la base de données est enregistré auprès de Lake Formation. Le IAMAllowedPrincipals groupe ne dispose pas d'Superautorisation sur la base de données et ses tables.

Accorder un accès multicompte aux nouveaux utilisateurs du catalogue de données via des politiques basées sur l'IAM sans interrompre les autorisations existantes de Lake Formation

1. Configuration du compte producteur

1. Connectez-vous à la console Lake Formation en utilisant un rôle `quilakeformation:PutDataLakeSettings`.
2. Sous Paramètres du catalogue de données, choisissez Version 4 les paramètres de version multi-comptes.

Si vous utilisez actuellement la version 1 ou 2, consultez [Mise à jour des paramètres de version de partage de données entre comptes](#) les instructions relatives à la mise à jour vers la version 3.

Aucune modification de la politique d'autorisation n'est requise pour passer de la version 3 à la version 4.

3. Répertoriez les autorisations que vous avez accordées aux principaux sur les bases de données et les tables. Pour de plus amples informations, veuillez consulter [Affichage des autorisations de base de données et de tables dans Lake Formation](#).
4. Réaccordez les autorisations de comptes croisés de Lake Formation existantes en optant pour les principes et les ressources.

 Note

Avant de mettre à jour un enregistrement de localisation de données en mode d'accès hybride afin d'accorder des autorisations entre comptes, vous devez réaccorder au moins un partage de données entre comptes par compte. Cette étape est nécessaire pour mettre à jour les autorisations AWS RAM gérées associées au partage de AWS RAM ressources.

En juillet 2023, Lake Formation a mis à jour les autorisations AWS RAM gérées utilisées pour le partage de bases de données et de tables :

- `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase`(politique de partage au niveau de la base de données)
- `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite`(politique de partage au niveau de la table)

Les autorisations entre comptes accordées avant juillet 2023 ne sont pas assorties de ces AWS RAM autorisations mises à jour.

Si vous avez accordé des autorisations multi-comptes directement aux principaux, vous devez les accorder de nouveau individuellement aux principaux. Si vous ignorez cette étape, les principaux accédant à la ressource partagée risquent de recevoir une erreur de combinaison illégale.

5. Rentre <https://console.aws.amazon.com/ram/chez-toi>.
6. L'onglet Shared by me de la AWS RAM console affiche les noms de base de données et de tables que vous avez partagés avec un compte ou un principal externe.

Assurez-vous que les autorisations associées à la ressource partagée possèdent le bon ARN.

7. Vérifiez que le Associated statut des ressources du AWS RAM partage est valide. Si le statut est défini comme telAssociating, attendez qu'ils passent à Associated l'état. Si

le statut devient le même `Failed`, arrêtez-vous et contactez l'équipe de service de Lake Formation.

8. Choisissez le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche, puis choisissez Ajouter.
 9. La page Ajouter des principes et des ressources indique les bases de données, and/or les tables et les principes auxquels ils ont accès. Vous pouvez effectuer les mises à jour requises en ajoutant ou en supprimant des principes et des ressources.
 10. Choisissez les principes dotés des autorisations Lake Formation pour la base de données et les tables que vous souhaitez passer en mode d'accès hybride. Choisissez les bases de données et les tables.
 11. Choisissez Ajouter pour activer les principes permettant d'appliquer les autorisations de Lake Formation en mode d'accès hybride.
 12. Accordez l'Superautorisation au groupe virtuel `IAMAllowedPrincipals` sur votre base de données et sur les tables sélectionnées.
 13. Modifiez l'enregistrement du site Amazon S3 Lake Formation en mode d'accès hybride.
 14. Accordez des autorisations aux AWS Glue utilisateurs du compte externe (consommateur) en utilisant les politiques d'autorisation IAM pour les AWS Glue actions Amazon S3.
2. Configuration du compte client
 1. Connectez-vous à la console Lake Formation en <https://console.aws.amazon.com/lakeformation/> tant qu'administrateur de lac de données.
 2. Accédez à la <https://console.aws.amazon.com/ram/page d'accueil> et acceptez l'invitation de partage de ressources. L'onglet Ressources partagées avec moi de la AWS RAM page affiche les noms de base de données et de tables partagés avec votre compte.

Pour le AWS RAM partage, assurez-vous que l'autorisation attachée possède le bon ARN de l' AWS RAM invitation partagée. Vérifiez si les ressources du AWS RAM partage sont en `Associated` état. Si le statut est défini comme `telAssociating`, attendez qu'ils passent à `Associated` l'état. Si le statut devient le même `Failed`, arrêtez-vous et contactez l'équipe de service de Lake Formation.
 3. Créez un lien de ressource vers la and/or table de base de données partagée dans Lake Formation.
 4. Accordez `Describe` l'autorisation sur le lien vers la ressource et l'`Grant on targetautorisation` (sur la ressource partagée d'origine) aux responsables IAM de votre compte (consommateur).

5. Configurez ensuite les autorisations de Lake Formation pour les principaux utilisateurs de votre compte sur la base de données ou la table partagée.

Dans la barre de navigation de gauche, sous Autorisations, choisissez le mode d'accès hybride.

6. Choisissez Ajouter dans la section inférieure de la page du mode d'accès hybride pour activer les principes et la base de données ou la table partagée avec vous depuis le compte du producteur.
7. Accordez des autorisations aux AWS Glue utilisateurs de votre compte en utilisant les politiques d'autorisation IAM pour les AWS Glue actions Amazon S3.
8. Testez les autorisations et AWS Glue autorisations des utilisateurs de Lake Formation en exécutant des exemples de requêtes distincts sur la table à l'aide d'Athena

(Facultatif) Nettoyez les politiques d'autorisation IAM pour Amazon S3 pour les principaux qui sont en mode d'accès hybride.

Supprimer les principes et les ressources du mode d'accès hybride

Procédez comme suit pour supprimer les bases de données, les tables et les principes du mode d'accès hybride.

Console

1. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Sous Autorisations, choisissez le mode d'accès hybride.
3. Sur la page Mode d'accès hybride, cochez la case à côté du nom de la base de données ou de la table et choisissez Remove.
4. Un message d'avertissement vous invite à confirmer l'action. Cliquez sur Supprimer.

Lake Formation n'applique plus les autorisations pour ces ressources, et l'accès à ces ressources sera contrôlé à l'aide de l'IAM et AWS Glue des autorisations. Cela peut empêcher l'utilisateur d'avoir accès à cette ressource s'il ne dispose pas des autorisations IAM appropriées.

AWS CLI

L'exemple suivant montre comment supprimer des ressources du mode d'accès hybride.

```
aws lakeformation delete-lake-formation-opt-in --cli-input-json file://file path

json:
{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::<123456789012>:role/role name"
  },
  "Resource": {
    "Table": {
      "CatalogId": "<123456789012>",
      "DatabaseName": "<database name>",
      "Name": "<table name>"
    }
  }
}
```

Affichage des principes et des ressources en mode d'accès hybride

Procédez comme suit pour afficher les bases de données, les tables et les principaux en mode d'accès hybride.

Console

1. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Sous Autorisations, choisissez le mode d'accès hybride.
3. La page Mode d'accès hybride affiche les ressources et les principaux qui sont actuellement en mode d'accès hybride.

AWS CLI

L'exemple suivant montre comment répertorier tous les principes et ressources optionnels en mode d'accès hybride.

```
aws lakeformation list-lake-formation-opt-ins
```

L'exemple suivant montre comment répertorier l'opt-in pour une paire principale-ressource spécifique.

```
aws lakeformation list-lake-formation-opt-ins --cli-input-json file://file path

json:
{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::<account-id>:role/<role name>"
  },
  "Resource": {
    "Table": {
      "CatalogId": "<account-id>",
      "DatabaseName": "<database name>",
      "Name": "<table name>"
    }
  }
}
```

Ressources supplémentaires

Dans le billet de blog suivant, nous vous expliquons comment intégrer les autorisations Lake Formation en mode d'accès hybride pour certains utilisateurs alors que la base de données est déjà accessible aux autres utilisateurs via les autorisations IAM et Amazon S3. Nous passerons en revue les instructions pour configurer le mode d'accès hybride au sein d'un AWS compte et entre deux comptes.

- [Présentation du mode d'accès hybride pour sécuriser l'accès AWS Glue Data Catalog à l'aide de Lake Formation et des politiques IAM et Amazon S3.](#)

Création d'objets dans AWS Glue Data Catalog

AWS Lake Formation utilise le AWS Glue Data Catalog (catalogue de données) pour stocker les métadonnées relatives aux lacs de données, aux sources de données, aux transformations et aux

cibles. Les métadonnées sont des données relatives aux données sous-jacentes de votre jeu de données. Chaque AWS compte possède un catalogue de données par AWS région.

Les métadonnées du catalogue de données sont organisées selon une hiérarchie de données à trois niveaux comprenant des catalogues, des bases de données et des tables. Il organise les données provenant de diverses sources dans des conteneurs logiques appelés catalogues. Chaque catalogue représente des données provenant de sources telles que les entrepôts de données Amazon Redshift, les Amazon DynamoDB bases de données et les sources de données tierces telles que Snowflake, MySQL, et de plus de 30 sources de données externes, qui sont intégrées via des connecteurs fédérés. Vous pouvez également créer de nouveaux catalogues dans le catalogue de données pour stocker les données dans des compartiments de table S3 ou dans le stockage géré Redshift (RMS).

Les tables stockent des informations sur les données sous-jacentes, notamment les informations de schéma, les informations de partition et l'emplacement des données. Les bases de données sont des ensembles de tables. Le catalogue de données contient également des liens vers des ressources, qui sont des liens vers des catalogues, des bases de données et des tables partagés dans des comptes externes, et sont utilisés pour l'accès entre comptes aux données du lac de données.

Le catalogue de données est un objet de catalogue imbriqué qui contient des catalogues, des bases de données et des tables. Il est référencé par l'ID Compte AWS et constitue le catalogue par défaut d'un compte et d'une Région AWS. Le catalogue de données utilise une hiérarchie à trois niveaux (catalog.database.table) pour organiser les tables.

- Catalogue : niveau supérieur de la hiérarchie des métadonnées à trois niveaux du catalogue de données. Vous pouvez ajouter plusieurs catalogues dans un catalogue de données par le biais de la fédération.
- Base de données : deuxième niveau de la hiérarchie des métadonnées comprenant des tables et des vues. Une base de données est également appelée schéma dans de nombreux systèmes de données tels qu'Amazon Redshift et Trino.
- Tableau et affichage : troisième niveau de la hiérarchie des données à 3 niveaux du catalogue de données.

Toutes les tables Iceberg d'Amazon S3 sont stockées dans le catalogue de données par défaut avec Catalog ID = Compte AWS ID. Vous pouvez créer des catalogues fédérés dans lesquels sont stockées AWS Glue Data Catalog les définitions des tables dans Amazon Redshift, Amazon S3 Table Storage ou dans d'autres sources de données tierces par le biais de la fédération.

Rubriques

- [Création d'un catalogue](#)
- [Création d'une base de données](#)
- [Création de tables](#)
- [AWS Glue Data Catalog Vues du bâtiment](#)

Création d'un catalogue

Les catalogues représentent le niveau le plus élevé ou le plus élevé de la hiérarchie de métadonnées à trois niveaux du. AWS Glue Data Catalog Vous pouvez utiliser plusieurs méthodes pour intégrer des données dans le catalogue de données et créer des catalogues à plusieurs niveaux.

Pour plus d'informations sur la création de catalogues à partir de sources de données externes, consultez [Intégrer vos données dans AWS Glue Data Catalog](#).

Pour créer un catalogue à l'aide de la console Lake Formation, vous devez être connecté en tant qu'administrateur de lac de données ou créateur de catalogue. Un créateur de catalogue est un directeur qui a obtenu l'CREATE_CATALOG autorisation de Lake Formation. Vous pouvez consulter la liste des créateurs de catalogues sur la page Rôles et tâches administratifs de la console Lake Formation. Pour consulter cette liste, vous devez disposer de l'autorisation lakeformation:ListPermissions IAM et être connecté en tant qu'administrateur de lac de données ou en tant que créateur de catalogue avec l'option d'octroi sur l'CREATE_CATALOG autorisation.

Création d'une base de données

Les tables de métadonnées du catalogue de données sont stockées dans des bases de données. Vous pouvez créer autant de bases de données que vous le souhaitez, et vous pouvez accorder différentes autorisations Lake Formation pour chaque base de données.

Les bases de données peuvent avoir une propriété d'emplacement facultative. Cet emplacement se trouve généralement dans un site Amazon Simple Storage Service (Amazon S3) enregistré auprès de Lake Formation. Lorsque vous spécifiez un emplacement, les principaux n'ont pas besoin d'autorisations de localisation des données pour créer des tables de catalogue de données qui pointent vers des emplacements au sein de l'emplacement de la base de données. Pour de plus amples informations, veuillez consulter [Underlying data access control](#).

Pour créer une base de données à l'aide de la console Lake Formation, vous devez être connecté en tant qu'administrateur de lac de données ou créateur de base de données. Un créateur

de base de données est un directeur qui a obtenu l'`CREATE_DATABASE` autorisation de Lake Formation. Vous pouvez consulter la liste des créateurs de bases de données sur la page Rôles et tâches administratifs de la console Lake Formation. Pour consulter cette liste, vous devez disposer de l'autorisation `lakeformation:ListPermissions` IAM et être connecté en tant qu'administrateur de lac de données ou en tant que créateur de base de données avec l'option d'octroi sur l'`CREATE_DATABASE` autorisation.

Pour créer une base de données

1. Ouvrez la AWS Lake Formation console sur <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'administrateur de lac de données ou créateur de base de données.
2. Dans le volet de navigation, sous Catalogue de données, sélectionnez Bases de données.
3. Choisissez Créer une base de données.
4. Dans la boîte de dialogue Créer une base de données, entrez un nom de base de données, un emplacement facultatif et une description facultative.
5. Sélectionnez éventuellement Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données.

Pour de plus amples informations sur cette option, veuillez consulter [the section called “Modification des paramètres par défaut de votre lac de données”](#).

6. Choisissez Créer une base de données.

Création de tables

AWS Lake Formation les tables de métadonnées contiennent des informations sur les données du lac de données, notamment des informations sur le schéma, les informations de partition et l'emplacement des données. Ces tables sont stockées dans le catalogue AWS Glue de données. Vous les utilisez pour accéder aux données sous-jacentes du lac de données et pour gérer ces données avec les autorisations de Lake Formation. Les tables sont stockées dans les bases de données du catalogue de données.

Il existe plusieurs méthodes pour créer des tables de catalogue de données :

- Lancez un crawler dedans. AWS Glue Consultez [la section Définition des robots](#) d'exploration dans le guide du AWS Glue développeur.
- Créez et exécutez un flux de travail. Consultez [the section called “Importation de données à l'aide de workflows”](#).

- Créez une table manuellement à l'aide de la console Lake Formation, de AWS Glue l'API ou AWS Command Line Interface (AWS CLI).
- Créez un tableau à l'aide de Amazon Athena.
- Créez un lien de ressource vers une table dans un compte externe. Consultez [the section called “Création de liens vers des ressources”](#).

Création de tables Apache Iceberg

AWS Lake Formation prend en charge la création de tables Apache Iceberg qui utilisent le format de données Apache Parquet dans AWS Glue Data Catalog les données résidant dans Amazon S3. Une table dans le catalogue de données correspond à la définition des métadonnées qui représente les données dans un entrepôt de données. Par défaut, Lake Formation crée des tables Iceberg v2. Pour connaître la différence entre les tables v1 et v2, consultez la section [Modifications de version de format](#) dans la documentation Apache Iceberg.

[Apache Iceberg](#) est un format de table ouvert pour les jeux de données analytiques très volumineux. Iceberg permet de modifier facilement votre schéma, également ce que l'on appelle l'évolution du schéma, ce qui signifie que les utilisateurs peuvent ajouter, renommer ou supprimer des colonnes d'une table de données sans perturber les données sous-jacentes. Iceberg fournit également une prise en charge pour la gestion des versions des données, ce qui permet aux utilisateurs de suivre les modifications apportées aux données au fil du temps. Cela active la fonctionnalité Time Travel, qui permet aux utilisateurs d'accéder aux versions historiques des données, de les interroger, mais aussi d'analyser les modifications apportées aux données entre les mises à jour et les suppressions.

Vous pouvez utiliser la console Lake Formation ou l'`CreateTable` opération de l' AWS Glue API pour créer une table Iceberg dans le catalogue de données. Pour plus d'informations, consultez [CreateTable action \(Python : create_table\)](#).

Lorsque vous créez une table Iceberg dans le catalogue de données, vous devez spécifier le format de la table et le chemin d'accès au fichier de métadonnées dans Amazon S3 pour pouvoir effectuer des lectures et des écritures.

Vous pouvez utiliser Lake Formation pour sécuriser votre table Iceberg à l'aide d'autorisations de contrôle d'accès précises lorsque vous enregistrez l'emplacement des données Amazon S3 auprès de celui-ci. AWS Lake Formation Pour les données source dans Amazon S3 et les métadonnées qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions. Pour de plus amples informations, veuillez consulter [Gestion des autorisations relatives à Lake Formation](#).

 Note

Le catalogue de données ne prend pas en charge la création de partitions ni l'ajout de propriétés de table Iceberg.

Rubriques

- [Conditions préalables](#)
- [Création d'une table Iceberg](#)

Conditions préalables

Pour créer des tables Iceberg dans le catalogue de données et configurer les autorisations d'accès aux données de Lake Formation, vous devez remplir les conditions suivantes :

1. Autorisations requises pour créer des tables Iceberg sans les données enregistrées auprès de Lake Formation :

Outre les autorisations requises pour créer une table dans le catalogue de données, le créateur de table exige les autorisations suivantes :

- `s3:PutObject` sur la ressource `arn:aws:s3:::{bucketName}`
 - `s3:GetObject` sur la ressource `arn:aws:s3:::{bucketName}`
 - `s3:DeleteObject` sur la ressource `arn:aws:s3:::{bucketName}`
2. Autorisations requises pour créer des tables Iceberg avec des données enregistrées auprès de Lake Formation :

Pour utiliser Lake Formation afin de gérer et de sécuriser les données de votre lac de données, enregistrez votre site Amazon S3 contenant les données pour les tables auprès de Lake Formation. Cela permet à Lake Formation de vendre des informations d'identification à AWS des services d'analyse tels qu'Athena, Redshift Spectrum et Amazon EMR pour accéder aux données. Pour plus d'informations sur l'enregistrement d'un site Amazon S3, consultez [Ajouter un emplacement Amazon S3 à votre lac de données](#).

Un principal qui lit et écrit les données sous-jacentes enregistrées auprès de Lake Formation doit disposer des autorisations suivantes :

- `lakeformation:GetDataAccess`

- DATA_LOCATION_ACCESS

Un principal qui possède des autorisations de localisation des données sur un emplacement possède également des autorisations d'emplacement sur tous les sites enfant.

Pour plus d'informations sur les autorisations de localisation des données, consultez [Contrôle d'accès aux données sous-jacent](#).

Pour activer le compactage, le service doit endosser un rôle IAM autorisé à mettre à jour les tables dans le catalogue de données. Pour plus de détails, consultez la section [Conditions préalables à l'optimisation des tables](#).

Création d'une table Iceberg

Vous pouvez créer des tables Iceberg v1 et v2 à l'aide de la console Lake Formation ou AWS Command Line Interface comme indiqué sur cette page. Vous pouvez également créer des tables Iceberg à l'aide de AWS Glue la console ou AWS Glue crawler. Pour plus d'informations, consultez [Data Catalog and Crawlers](#) dans le Guide du développeur AWS Glue .

Pour créer une table Iceberg

Console

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Sous catalogue de données, choisissez Tables, puis utilisez le bouton Créer une table pour spécifier les attributs suivants :
 - Nom de la table : entrez le nom de la table. Si vous utilisez Athena pour accéder à des tables, utilisez ces [conseils de dénomination](#) figurant dans le Guide de l'utilisateur Amazon Athena.
 - Base de données : Choisissez une base de données existante ou créez-en une nouvelle.
 - Description : description de la table. Vous pouvez écrire une description vous aidant à comprendre le contenu de la table.
 - Format de tableau : pour Format de tableau, choisissez Apache Iceberg.

Table format [Info](#)

Data Catalog managed tables support data compaction for Apache Iceberg table type. [Learn more](#)

☐ **Standard AWS Glue table (default)**
Create a standard AWS Glue table.

☒ **Apache Iceberg table**
Create a table in Apache Iceberg table format.

Table optimization
Enable optimization capabilities for Apache Iceberg tables enhance query performance.

☒ **Compaction**
Combine small data files into larger, more efficient files to optimize table performance.

☒ **Snapshot retention - new**
Optimize table storage by removing old snapshots from metadata overhead and expiring files that are no longer needed.

☒ **Orphan file deletion - new**
Automatically clean up orphan files periodically.

- **Optimisation des tables**
 - **Compaction** : les fichiers de données sont fusionnés et réécrits pour supprimer les données obsolètes et regrouper les données fragmentées dans des fichiers plus volumineux et plus efficaces.
 - **Conservation des instantanés** : les instantanés sont des versions horodatées d'une table Iceberg. Les configurations de conservation des instantanés permettent aux clients de définir la durée de conservation des instantanés et le nombre d'instantanés à conserver. La configuration d'un optimiseur de conservation des instantanés peut aider à gérer la charge de stockage en supprimant les instantanés plus anciens et inutiles, ainsi que leurs fichiers sous-jacents associés.
 - **Suppression de fichiers orphelins** : les fichiers orphelins sont des fichiers qui ne sont plus référencés par les métadonnées de la table Iceberg. Ces fichiers peuvent s'accumuler au fil du temps, en particulier après des opérations telles que la suppression de tables ou l'échec de tâches ETL. L'activation de la suppression des fichiers orphelins permet AWS Glue d'identifier et de supprimer périodiquement ces fichiers inutiles, libérant ainsi de l'espace de stockage.

Pour plus d'informations, consultez [Optimisation des tables Iceberg](#).

- **Rôle IAM** : pour exécuter le compactage, le service assume un rôle IAM en votre nom. Vous pouvez choisir un rôle IAM à l'aide de la liste déroulante. Assurez-vous que le rôle dispose des autorisations requises pour activer le compactage.

Pour en savoir plus sur les autorisations requises, consultez la section [Conditions préalables à l'optimisation des tables](#).

- **Emplacement** : Spécifiez le chemin d'accès au dossier dans Amazon S3 qui stocke la table de métadonnées. Iceberg a besoin d'un fichier de métadonnées et d'un emplacement dans le catalogue de données pour pouvoir effectuer des lectures et des écritures.
- **Schéma** : choisissez Ajouter des colonnes pour ajouter des colonnes et les types de données des colonnes. Vous avez la possibilité de créer une table vide et de mettre à jour le schéma ultérieurement. Le catalogue de données prend en charge les types de données Hive. Pour plus d'informations, consultez [Hive data types](#).

Iceberg vous permet de faire évoluer le schéma et la partition après avoir créé la table. Vous pouvez utiliser des [requêtes Athena](#) pour mettre à jour le schéma de table et des [requêtes Spark](#) pour mettre à jour les partitions.

AWS CLI

```
aws glue create-table \  
  --database-name iceberg-db \  
  --region us-west-2 \  
  --open-table-format-input '{  
    "IcebergInput": {  
      "MetadataOperation": "CREATE",  
      "Version": "2"  
    }  
  }' \  
  --table-input '{"Name":"test-iceberg-input-demo",  
    "TableType": "EXTERNAL_TABLE",  
    "StorageDescriptor":{  
      "Columns":[  
        {"Name":"col1", "Type":"int"},  
        {"Name":"col2", "Type":"int"},  
        {"Name":"col3", "Type":"string"}  
      ],  
      "Location":"s3://DOC_EXAMPLE_BUCKET_ICEBERG/"  
    }  
  }'
```

Optimisation des tables Iceberg

Lake Formation prend en charge plusieurs options d'optimisation des tables afin d'améliorer la gestion et les performances des tables Apache Iceberg utilisées par les moteurs AWS d'analyse et les tâches ETL. Ces optimiseurs permettent d'utiliser efficacement le stockage, d'améliorer les performances des requêtes et de gérer efficacement les données. Trois types d'optimiseurs de tables sont disponibles dans Lake Formation :

- **Compactage** : le compactage des données compacte les petits fichiers de données afin de réduire l'utilisation du stockage et d'améliorer les performances de lecture. Les fichiers de données sont fusionnés et réécrits pour supprimer les données obsolètes et regrouper les données fragmentées dans des fichiers plus volumineux et plus efficaces. Le compactage peut être configuré pour s'exécuter automatiquement ou être déclenché manuellement selon les besoins.
- **Conservation des instantanés** : les instantanés sont des versions horodatées d'une table Iceberg. Les configurations de conservation des instantanés permettent aux clients de définir la durée de conservation des instantanés et le nombre d'instantanés à conserver. La configuration d'un optimiseur de conservation des instantanés peut aider à gérer la charge de stockage en supprimant les instantanés plus anciens et inutiles, ainsi que leurs fichiers sous-jacents associés.
- **Suppression de fichiers orphelins** : les fichiers orphelins sont des fichiers qui ne sont plus référencés par les métadonnées de la table Iceberg. Ces fichiers peuvent s'accumuler au fil du temps, en particulier après des opérations telles que la suppression de tables ou l'échec de tâches ETL. L'activation de la suppression des fichiers orphelins permet AWS Glue d'identifier et de supprimer périodiquement ces fichiers inutiles, libérant ainsi de l'espace de stockage.

Vous pouvez activer ou désactiver le compactage, la conservation des instantanés et les optimiseurs de suppression de fichiers orphelins pour les tables Iceberg individuelles du catalogue de données à l'aide de la AWS Glue console ou des opérations de l'API AWS CLI. AWS Glue

Pour plus d'informations, consultez la section [Optimisation des tables Iceberg](#) dans le guide du AWS Glue développeur.

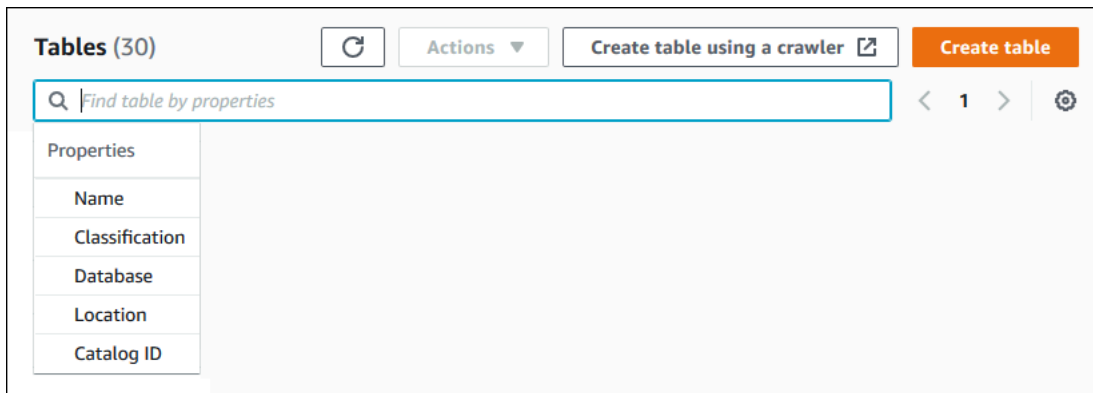
Recherche de tables

Vous pouvez utiliser la AWS Lake Formation console pour rechercher des tables du catalogue de données par nom, emplacement, base de données contenant, etc. Les résultats de recherche n'affichent que les tables pour lesquelles vous êtes autorisé à accéder à Lake Formation.

Pour rechercher des tables (console)

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, choisissez Tables.
3. Positionnez le curseur dans le champ de recherche en haut de la page. Le champ contient le texte d'espace réservé Rechercher le tableau par propriétés.

Le menu Propriétés apparaît et indique les différentes propriétés de table à partir desquelles effectuer une recherche.



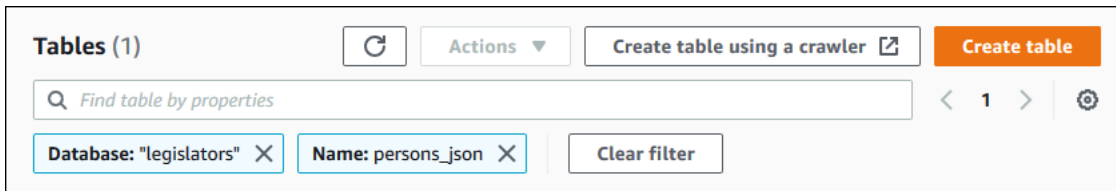
4. Effectuez l'une des actions suivantes :

- Recherche par base de données contenant.
 1. Choisissez Base de données dans le menu Propriétés, puis choisissez une base de données dans le menu Bases de données qui apparaît ou tapez un nom de base de données et appuyez sur Entrée.

Les tables sur lesquelles vous avez des autorisations dans la base de données sont répertoriées.

2. (Facultatif) Pour réduire la liste à une seule table de la base de données, positionnez à nouveau le curseur dans le champ de recherche, choisissez Nom dans le menu Propriétés, puis choisissez un nom de table dans le menu Tables qui apparaît ou tapez un nom de table et appuyez sur Entrée.

La table unique est répertoriée, et le nom de la base de données et le nom de la table apparaissent sous forme de vignettes sous le champ de recherche.



Pour ajuster le filtre, fermez l'une des vignettes ou choisissez Effacer le filtre.

- Effectuez une recherche par autres propriétés.

1. Choisissez une propriété de recherche dans le menu Propriétés.

Pour effectuer une recherche par numéro de AWS compte, choisissez Numéro de catalogue dans le menu Propriétés, entrez un identifiant de AWS compte valide (par exemple, 111122223333), puis appuyez sur Entrée.

Pour effectuer une recherche par lieu, choisissez Emplacement dans le menu Propriétés, puis sélectionnez un lieu dans le menu Emplacements qui apparaît. Toutes les tables situées à l'emplacement racine de l'emplacement sélectionné (par exemple, Amazon S3) sont renvoyées.

Recherche de tables à l'aide de AWS CLI

- L'exemple suivant montre comment exécuter une recherche partielle. Le `--search-text` paramètre vous permet de rechercher des tables contenant le texte spécifié dans leurs métadonnées. Dans ce cas, il renvoie toutes les tables dont le nom, la description ou d'autres champs de métadonnées contiennent « client ».

```
aws glue search-tables
  --search-text "customer"
  --region Région AWS
  --max-results 10
  --sort-criteria "FieldName=Name,Sort=ASC"
```

Partage des tables et des bases de données du catalogue de données entre les AWS comptes

Vous pouvez partager les ressources du catalogue de données (bases de données et tables) avec AWS des comptes externes en accordant à Lake Formation des autorisations sur les ressources aux comptes externes. Les utilisateurs peuvent ensuite exécuter des requêtes et des tâches qui joignent

et interrogent des tables sur plusieurs comptes. Sous réserve de certaines restrictions, lorsque vous partagez une ressource de catalogue de données avec un autre compte, les principaux associés de ce compte peuvent utiliser cette ressource comme si elle figurait dans leur catalogue de données.

Vous ne partagez pas de ressources avec des responsables spécifiques dans des AWS comptes externes ; vous partagez les ressources avec un compte ou une AWS organisation. Lorsque vous partagez une ressource avec une AWS organisation, vous la partagez avec tous les comptes à tous les niveaux de cette organisation. L'administrateur du lac de données de chaque compte externe doit ensuite accorder des autorisations sur les ressources partagées aux principaux de son compte.

Pour plus d'informations, consultez [Partage de données entre comptes dans Lake Formation](#) et [Octroi d'autorisations sur les ressources du catalogue de données](#).

 Voir aussi :

- [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#)
- [Conditions préalables](#)

AWS Glue Data Catalog Vues du bâtiment

Dans le AWS Glue Data Catalog, une vue est une table virtuelle dont le contenu est défini par une requête SQL qui fait référence à une ou plusieurs tables. Vous pouvez créer une vue du catalogue de données qui référence jusqu'à 10 tables à l'aide d'éditeurs SQL pour Amazon Athena, Amazon Redshift ou Apache Spark à l'aide d'EMR Serverless ou de la version 5.0. AWS Glue Les tables de référence sous-jacentes d'une vue peuvent appartenir à la même base de données ou à différentes bases Compte AWS de données du même catalogue de données.

Vous pouvez référencer AWS Glue des tables standard et des tables dans des formats de table ouverts (OTF) tels que [Apache Hudi](#), Linux Foundation [Delta Lake](#) et [Apache Iceberg](#), avec des données sous-jacentes stockées dans des emplacements Amazon S3 enregistrés auprès de. AWS Lake Formation En outre, vous pouvez créer des vues à partir de tables fédérées à partir de partages de données Amazon Redshift partagés avec Lake Formation.

Différencier les vues du catalogue de données des autres types de vues

Les vues du catalogue de données sont différentes des vues Apache Hive, Apache Spark et Amazon Athena. La vue du catalogue de données est une fonctionnalité native du AWS Glue Data Catalog. Il

s'agit d'une vue créée par un définisseur multidialecte. Vous pouvez créer une vue de catalogue de données à l'aide de l'un des services d'analyse pris en charge, tels qu'Athena ou Amazon Redshift Spectrum, et accéder à la même vue à l'aide d'autres services d'analyse pris en charge. D'autre part, les vues Apache Hive, Apache Spark et Athena sont créées indépendamment dans chaque service d'analyse, tel qu'Athena et Amazon Redshift, et sont visibles et accessibles uniquement au sein de ce service.

Qu'est-ce qu'une vue définissante ?

Une vue de définition est une vue SQL qui fonctionne en fonction des autorisations du principal qui l'a créée. Le rôle de définition dispose des autorisations nécessaires pour accéder aux tables référencées, et il exécute l'instruction SQL qui définit la vue. Le définisseur crée la vue et la partage avec les autres utilisateurs grâce à un AWS Lake Formation contrôle d'accès précis.

Lorsqu'un utilisateur interroge la vue du définisseur, le moteur de requête utilise les autorisations du rôle du définisseur pour accéder aux tables de référence sous-jacentes. Cette approche permet aux utilisateurs d'interagir avec la vue sans avoir besoin d'accéder directement aux tables sources, ce qui améliore la sécurité et simplifie la gestion de l'accès aux données.

Pour configurer une vue de définition, le rôle IAM de définition peut se trouver dans le même AWS compte que les tables de base, ou dans un autre compte en utilisant des rôles de définition entre comptes. Pour plus d'informations sur les autorisations requises pour le rôle de définisseur, consultez [Conditions préalables à la création de vues](#).

Un cadre pour des vues multidialectes

Le catalogue de données permet de créer des vues à l'aide de plusieurs dialectes SQL (Structured Query Language). Le SQL est un langage utilisé pour stocker et traiter des informations dans une base de données relationnelle et chaque moteur AWS d'analyse utilise sa propre variante de SQL, ou dialecte SQL.

Vous créez une vue du catalogue de données dans un dialecte SQL à l'aide de l'un des moteurs de requêtes d'analyse pris en charge. Vous pouvez ensuite mettre à jour la vue à l'aide de l'ALTER VIEW instruction dans un autre dialecte SQL au sein de tout autre moteur d'analyse pris en charge. Cependant, chaque dialecte doit faire référence au même ensemble de tables, de colonnes et de types de données.

Vous pouvez accéder aux multiples dialectes disponibles pour la vue à l'aide de l'GetTableAPI AWS CLI et de la AWS console. Ainsi, la vue du catalogue de données est visible et disponible pour effectuer des requêtes sur les différents moteurs d'analyse pris en charge.

En définissant un schéma de vue commun et un objet de métadonnées que vous pouvez interroger à partir de plusieurs moteurs, les vues du catalogue de données vous permettent d'utiliser des vues uniformes sur l'ensemble de votre lac de données.

Pour plus de détails sur la façon dont le schéma est résolu pour chaque dialecte, consultez le [lien vers la référence de l'API](#). Pour plus de détails sur les règles de correspondance pour les différents types, voir le [lien vers la section correspondante dans le document de l'API](#).

Intégration aux autorisations de Lake Formation

Vous pouvez l'utiliser AWS Lake Formation pour centraliser la gestion des autorisations sur les AWS Glue Data Catalog vues pour les utilisateurs. Vous pouvez accorder des autorisations détaillées sur les vues du catalogue de données à l'aide de la méthode des ressources nommées ou des balises LF, et les partager entre les AWS organisations et les unités organisationnelles. Comptes AWS Vous pouvez également partager et accéder aux vues du catalogue de données Régions AWS via des liens vers des ressources. Cela permet aux utilisateurs d'accéder aux données sans dupliquer la source de données et sans partager les tables sous-jacentes.

L'instruction `CREATE VIEW` DDL d'une vue du catalogue de données peut faire référence aux AWS Glue tables standard et aux tables dans des formats de table ouverts (OTF) tels que Hudi, Delta Lake et Iceberg avec des données sous-jacentes stockées dans des emplacements Amazon S3 enregistrés auprès de Lake Formation, ainsi qu'aux tables fédérées du partage de données Amazon Redshift partagées avec Lake Formation. Les tables peuvent être de n'importe quel format de fichier, à condition que le moteur utilisé pour interroger la vue prenne en charge ce format. Vous pouvez également faire référence aux fonctions intégrées du moteur sur lequel il est exécuté, mais d'autres ressources spécifiques au moteur peuvent ne pas être autorisées. Pour plus d'informations, consultez [Considérations et limites relatives aux affichages du catalogue de données](#).

Cas d'utilisation

Voici les principaux cas d'utilisation des vues du catalogue de données :

- Créez et gérez les autorisations sur un schéma de vue unique. Cela vous permet d'éviter le risque d'autorisations incohérentes sur les vues dupliquées créées dans plusieurs moteurs.
- Accordez des autorisations aux utilisateurs sur une vue qui référence plusieurs tables sans accorder d'autorisations directement sur les tables de référence sous-jacentes.
- Réalisez un filtrage au niveau des lignes sur les tables à l'aide de balises LF (les balises LF ne sont mises en cascade que jusqu'au niveau des colonnes) en appliquant des balises LF aux vues et en accordant aux utilisateurs des autorisations basées sur les balises LF.

Services AWS d'analyse pris en charge pour les vues

Les services AWS d'analyse suivants permettent de créer des vues de catalogue de données :

- Amazon Redshift
- Amazon Athena version 3
- Apache Spark sur EMR sans serveur
- Apache Spark sur AWS Glue la version 5.0

Ressources supplémentaires

Vous pouvez en savoir plus sur le catalogue de données dans ce guide, ainsi qu'en utilisant les ressources suivantes :

La vidéo suivante montre comment créer des vues et les interroger depuis Athena et Amazon Redshift.

Rubriques

- [Conditions préalables à la création de vues](#)
- [Création de vues de catalogue de données à l'aide d'instructions DDL](#)
- [Création de vues de catalogue de données à l'aide de AWS Glue APIs](#)
- [Octroi d'autorisations sur les vues du catalogue de données](#)
- [Vues matérialisées](#)

Conditions préalables à la création de vues

- Pour créer des vues dans Data Catalog, vous devez enregistrer les emplacements de données Amazon S3 sous-jacents des tables de référence auprès de Lake Formation. Pour plus de détails sur l'enregistrement des données auprès de Lake Formation, voir [Ajouter un emplacement Amazon S3 à votre lac de données](#).
- Seuls les rôles IAM peuvent créer des vues de catalogue de données. Les autres identités IAM ne peuvent pas créer de vues de catalogue de données.
- Le rôle IAM qui définit la vue doit disposer des autorisations suivantes :
 - `SELECT` Autorisation Lake Formation avec `Grantable` option sur toutes les tables de référence, toutes les colonnes incluses.

- CREATE_TABLEAutorisation Lake Formation sur la base de données cible dans laquelle les vues sont créées.
- Une politique de confiance pour que la Lake Formation et AWS Glue les services assument ce rôle.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "DataCatalogViewDefinerAssumeRole1",
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "glue.amazonaws.com",
          "lakeformation.amazonaws.com"
        ]
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

- Le but : PassRole l'autorisation pour AWS Glue Lake Formation.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "DataCatalogViewDefinerPassRole1",
      "Action": [
        "iam:PassRole"
      ],
      "Effect": "Allow",
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": [
            "glue.amazonaws.com",

```

```

        "lakeformation.amazonaws.com"
    ]
}
}
}
}
}

```

- AWS Glue et autorisations de Lake Formation.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "Glue:GetDatabase",
        "Glue:GetDatabases",
        "Glue:CreateTable",
        "Glue:GetTable",
        "Glue:GetTables",
        "Glue:BatchGetPartition",
        "Glue:GetPartitions",
        "Glue:GetPartition",
        "Glue:GetTableVersion",
        "Glue:GetTableVersions",
        "Glue:PassConnection",
        "lakeFormation:GetDataAccess"
      ],
      "Resource": "*"
    }
  ]
}

```

- Vous ne pouvez pas créer de vues dans une base de données Super dotée d'une ALL autorisation accordée au IAMAllowedPrincipals groupe. Vous pouvez soit révoquer l'Superautorisation accordée au IAMAllowedPrincipals groupe sur une base de données, voir [Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations Lake Formation](#), soit créer une nouvelle base de données en décochant la case Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données sous Autorisations par défaut pour les tables nouvellement créées.

Création de vues de catalogue de données à l'aide d'instructions DDL

Vous pouvez créer des AWS Glue Data Catalog vues à l'aide des éditeurs SQL pour Athena, Amazon Redshift et à l'aide du/. AWS Glue APIs AWS CLI

Pour créer une vue de catalogue de données à l'aide des éditeurs SQL, choisissez Athena ou Redshift Spectrum, puis créez la vue à l'aide d'une instruction DDL (`CREATE VIEW` Data Definition Language). Après avoir créé une vue dans le dialecte du premier moteur, vous pouvez utiliser une instruction `ALTER VIEW` DDL du second moteur pour ajouter les dialectes supplémentaires.

Lors de la définition des vues, il est important de prendre en compte les points suivants :

- **Définition de vues multidialectes** : lorsque vous définissez une vue comportant plusieurs dialectes, les schémas des différents dialectes doivent correspondre. Chaque dialecte SQL aura une spécification de syntaxe légèrement différente. La syntaxe de requête définissant la vue du catalogue de données doit aboutir à la même liste de colonnes, y compris les types et les noms, dans tous les dialectes. Ces informations sont stockées dans `StorageDescriptor` la vue. Les dialectes doivent également faire référence aux mêmes objets de table sous-jacents du catalogue de données.

Pour ajouter un autre dialecte à une vue à l'aide du DDL, vous pouvez utiliser l'`ALTER VIEW` instruction. Si une `ALTER VIEW` instruction tente de mettre à jour la définition de la vue, par exemple en modifiant le descripteur de stockage ou les tables sous-jacentes de la vue, l'instruction indique par erreur « Incompatibilité entre le descripteur de stockage en entrée et le descripteur de stockage existant ». Vous pouvez utiliser les opérations SQL Cast pour vous assurer que les types de colonnes de vue correspondent.

- **Mettre à jour une vue** — Pour mettre à jour la vue, vous pouvez utiliser l'`UpdateTableAPI`. Si vous mettez à jour la vue sans faire correspondre les descripteurs de stockage ou les tables de référence, vous pouvez fournir l'`FORCE` indicateur (voir la documentation SQL du moteur pour la syntaxe). Après une mise à jour forcée, la vue prendra en compte les tables forcées `StorageDescriptor` et de référence. Tout autre `ALTER VIEW` DDL doit correspondre aux valeurs modifiées. Une vue qui a été mise à jour pour contenir des dialectes incompatibles sera considérée comme « obsolète ». L'état de la vue est visible dans la console Lake Formation et à l'aide de l'`GetTable` opération.
- **Référencer un type de colonne varchar sous forme de chaîne** — Il n'est pas possible de convertir un type de colonne varchar de Redshift Spectrum en chaîne. Si une vue est créée dans Redshift Spectrum avec un type de colonne varchar et qu'un dialecte ultérieur essaie de référencer ce

champ sous forme de chaîne, le catalogue de données le traitera comme une chaîne sans avoir besoin de l'indicateur. FORCE

- Traitement des champs de type complexe : Amazon Redshift traite tous les types complexes comme des [SUPER types](#), tandis qu'Athena spécifie le type complexe. Si une vue possède un champ de SUPER type et qu'un autre moteur référence cette colonne comme un type complexe particulier, tel que struct (<street_address:struct<street_number:int, street_name:string, street_type:string>>), le catalogue de données suppose que le champ est du type complexe spécifique et l'utilise dans le descripteur de stockage, sans avoir besoin de l'Forceindicateur.

Pour plus d'informations sur la syntaxe de création et de gestion des vues du catalogue de données, voir :

- [Utilisation des AWS Glue Data Catalog vues](#) dans le guide de l'utilisateur d'Amazon Athena.
- [Syntaxe des requêtes d'affichage du catalogue Glue Data](#) dans le guide de l'utilisateur d'Amazon Athena.
- [Création de vues AWS Glue Data Catalog dans le manuel](#) du développeur de base de données Amazon Redshift.

Pour plus d'informations sur les commandes SQL associées aux vues du catalogue de données, consultez [CREATE EXTERNAL VIEW](#), [ALTER EXTERNAL VIEW](#) et [DROP EXTERNAL VIEW](#).

Une fois que vous avez créé une vue du catalogue de données, les détails de la vue sont disponibles dans la console Lake Formation.

1. Choisissez Views sous Data Catalog dans la console Lake Formation.
2. La liste des vues disponibles apparaît sur la page des vues.
3. Choisissez une vue dans la liste et la page de détails affiche les attributs de la vue.

[AWS Lake Formation](#) > [Views](#) > europe_players

europe_players

Version 1 (Current version) ▾ Actions ▾

Details

Name	Database	Definer role
europe_players	views_demo_database	admin
Last updated	Status	Description
November 22, 2023 at 10:41 PM UTC	Ready	-

Schema

[SQL definitions](#)

LF-Tags

Cross-account access

Underlying tables

SQL definitions (2)

Add SQL definition ▾

List of available SQL definitions in different engines. Choose an engine from the list to add or edit the definition.

Find engine

< 1 >

Engine name	Version	Status	SQL statement	Edit definition
Athena	3	Ready	View	Amazon Athena
Redshift	1.0	Ready	View	Amazon Redshift

Schema

Choisissez une CoLumn ligne, puis sélectionnez Modifier les balises LF pour mettre à jour les valeurs des balises ou attribuer de nouvelles balises LF.

Définitions SQL

Vous pouvez consulter la liste des définitions SQL disponibles. Sélectionnez Ajouter une définition SQL, puis choisissez un moteur de requête pour ajouter une définition SQL. Choisissez un moteur de requête (Athena ou Amazon Redshift) sous Edit définition la colonne pour mettre à jour une définition SQL.

Étiquettes LF

Choisissez Modifier les balises LF pour modifier les valeurs d'une balise ou attribuer de nouvelles balises. Vous pouvez utiliser les balises LF pour accorder des autorisations sur les vues.

Accès intercomptes

Vous pouvez consulter la Comptes AWS liste des organisations et unités organisationnelles (OUs) qui ont partagé la vue du catalogue de données avec vous.

Tableaux sous-jacents

Les tables sous-jacentes référencées dans la définition SQL utilisée pour créer la vue sont affichées sous cet onglet.

Création de vues de catalogue de données à l'aide de AWS Glue APIs

Vous pouvez utiliser AWS Glue [CreateTable](#) et [UpdateTable](#) APIs pour créer et mettre à jour des vues dans le catalogue de données. Les UpdateTable opérations CreateTable et ont une nouvelle TableInput structure pour ViewDefinitionSearchTables, tandis que les GetTableVersions opérations GetTableGetTables,GetTableVersion,, fournissent la syntaxe des vues ViewDefinition dans leur sortie. En outre, il existe un nouveau Status champ dans la sortie de l'GetTableAPI.

Deux nouvelles AWS Glue connexions sont disponibles pour valider le dialecte SQL pour chaque moteur de requête pris en charge, Amazon Athena ainsi que pour Amazon Redshift.

Les CreateTable et UpdateTable APIs sont asynchrones lorsqu'ils sont utilisés avec des vues. Lorsque ceux-ci APIs sont appelés avec plusieurs dialectes SQL, l'appel est validé par chaque moteur afin de déterminer si le dialecte peut être exécuté sur ce moteur et si le schéma obtenu de la vue de chaque dialecte correspond. Le AWS Glue service utilise ces connexions pour effectuer des appels internes aux moteurs d'analyse. Ces appels simulent ce que fait le moteur pour valider si un DDL CREATE VIEW ou ALTER VIEW SQL a été exécuté sur le moteur.

Si le code SQL fourni est valide et que les schémas correspondent aux différents dialectes de vue, l'AWS Glue API valide le résultat de manière atomique. Atomicity permet de créer ou de modifier des vues avec plusieurs dialectes sans aucune interruption.

Rubriques

- [Création de AWS Glue connexions pour valider le statut](#)
- [Validation de l'état de génération de vues](#)
- [États et opérations asynchrones](#)
- [Afficher les scénarios d'échec de création lors d'opérations asynchrones](#)

Création de AWS Glue connexions pour valider le statut

Pour créer ou mettre à jour une AWS Glue Data Catalog vue à l'aide `UpdateTable` des opérations `CreateTable` or, vous devez créer un nouveau type de AWS Glue connexion à des fins de validation et le fournir au moteur d'analyse pris en charge. Ces connexions sont nécessaires pour utiliser les vues du catalogue de données avec Athena ou Amazon Redshift. Vous pouvez créer ces connexions uniquement à l'aide du AWS CLI AWS SDKs, ou AWS Glue APIs. Vous ne pouvez pas utiliser le AWS Management Console pour créer la AWS Glue connexion.

Note

Si le rôle de définition de la vue et le rôle appelant `CreateTable` ou `UpdateTable` sont différents, ils ont tous deux besoin d'une `glue:PassConnection` autorisation dans leur déclaration de politique IAM.

Pour plus d'informations, consultez la documentation relative à la [création d'une connexion](#) AWS CLI .

AWS CLI commande pour créer une connexion

Voici une AWS CLI commande permettant de créer une connexion :

```
aws glue create-connection --region us-east-1
--endpoint-url https://glue.us-east-1.amazonaws.com
--cli-input-json file:///root/path/to/create-connection.json
```

AWS CLI entrée JSON

Pour Amazon Redshift :

```
{
  "CatalogId": "123456789012",
  "ConnectionInput": {
    "ConnectionType": "VIEW_VALIDATION_REDSHIFT",
    "Name": "views-preview-cluster-connection-2",
    "Description": "My first Amazon Redshift validation connection",
    "ConnectionProperties": {
      "DATABASE": "dev",
      "CLUSTER_IDENTIFIER": "glue-data-catalog-views-preview-cluster"
    }
  }
}
```

```
}  
}
```

Pour Amazon Athena :

```
{  
  "CatalogId": "123456789012",  
  "ConnectionInput": {  
    "ConnectionType": "VIEW_VALIDATION_ATHENA",  
    "Name": "views-preview-cluster-connection-3",  
    "Description": "My first Amazon Athena validation connection",  
    "ConnectionProperties": {  
      "WORKGROUP_NAME": "workgroup-name"  
    }  
  }  
}
```

Validation de l'état de génération de vues

Lorsque vous exécutez les UpdateTable opérations CreateTable or, le Status champ de sortie de l'GetTableAPI affiche les détails de l'état de création de la vue. Pour les create demandes pour lesquelles la table n'existe pas déjà, AWS Glue crée une table vide pendant la durée du processus asynchrone. Lorsque vous appelezGetTable, vous pouvez transmettre un indicateur booléen facultatifIncludeStatusDetails, qui affiche les informations de diagnostic relatives à la demande. En cas d'échec, cet indicateur affiche un message d'erreur avec les statuts individuels de chaque dialecte.

Des erreurs lors des opérations de création, de lecture, de mise à jour et de suppression de vues (CRUD) peuvent survenir soit pendant le traitement dans le service AWS Glue/Lake Formation, soit lors de la validation SQL de la vue dans Amazon Redshift ou Athena. Lorsqu'une erreur se produit lors de la validation dans un moteur, le AWS Glue service fournit le message d'erreur renvoyé par le moteur.

Champs de statut

Les champs de statut sont les suivants :

- Statut : statut générique, indépendant des différents types d'emplois :
 - QUEUED
 - EN_COURS

- SUCCESS
- ÉCHEC
- Action : indique quelle action a été appelée sur la table, actuellement uniquement CREATE ou quelles UPDATE opérations sont disponibles.

Il est important de faire la distinction entre CREATE les opérations UPDATE et lorsque vous travaillez avec des vues. Le type d'opération détermine la manière dont vous devez interroger les tables.

Une UPDATE opération signifie que la table existe déjà dans le catalogue de données. Dans ce cas, vous pouvez continuer à interroger la table créée précédemment sans aucun problème. D'autre part, une CREATE opération indique que la table n'a jamais été créée avec succès auparavant. Si une table est marquée pour CREATE, toute tentative d'interrogation échouera car la table n'existe pas encore dans le système. Il est donc essentiel d'identifier le type d'opération (UPDATE ou CREATE) avant de tenter d'interroger une table.

- RequestedBy — L'ARN de l'utilisateur qui a demandé la modification asynchrone.
- UpdatedBy — L'ARN de l'utilisateur qui a modifié manuellement pour la dernière fois le processus de modification asynchrone, par exemple en demandant une annulation ou une modification.
- Erreur — Ce champ n'apparaît que lorsque l'état est FAILED. Il s'agit d'un message d'exception au niveau du parent. Il peut y avoir des erreurs différentes pour chaque dialecte.
 - ErrorCode — Le type d'exception.
 - ErrorMessage — une brève description de l'exception.
- RequestTime — une chaîne de date au format ISO 8601 indiquant l'heure à laquelle la modification a été initiée.
- UpdateTime — une chaîne de date au format ISO 8601 indiquant l'heure à laquelle l'état a été mis à jour pour la dernière fois.

États et opérations asynchrones

Lorsque vous exécutez une `glue:CreateTable` demande, la création asynchrone de la vue du catalogue de données commence. Dans les sections suivantes, ce document décrit une AWS Glue vue disponible dans une `glue:GetTable` réponse. Status Par souci de concision, cette section omet la réponse complète.

```
{  
  "Table": {
```

```
    ...
    "Status": {
        ...
        "Action": "CREATE",
        "State": "QUEUED",
    }
}
```

Les deux attributs ci-dessus représentent des informations de diagnostic importantes qui indiquent l'état de l'opération asynchrone, ainsi que les actions qui peuvent être effectuées sur cette vue. Vous trouverez ci-dessous les valeurs possibles que ces attributs peuvent prendre.

1. Status.Action

- a. CREATE
- b. UPDATE

2. Status.State

- a. QUEUED
- b. EN_COURS
- c. SUCCESS
- d. ÉCHEC

Il est également important de noter que certaines mises à jour d'une vue du catalogue de données ne nécessitent pas d'opération asynchrone. Par exemple, on peut souhaiter mettre à jour l'`Description` attribut de la table. Comme cela ne nécessite aucune opération asynchrone, les métadonnées de table résultantes n'en comporteront aucune `Status`, et l'attribut le sera. `NULL`

```
{
  "Table": {
    ...,
    "Description": "I changed this attribute!"
  }
}
```

Ensuite, cette rubrique explique comment les informations d'état ci-dessus peuvent avoir un impact sur les opérations pouvant être effectuées sur une AWS Glue vue.

colle : CreateTable

Il n'y a aucun changement pour cette API par rapport au `glue:CreateTable` fonctionnement de n'importe quelle table Glue. `CreateTable` peut être appelé pour n'importe quel nom de table qui n'existe pas déjà.

colle : `UpdateTable`

Cette opération ne peut pas être effectuée sur une AWS Glue vue dont les informations d'état sont les suivantes :

1. Action = CRÉER et état = EN FILE D'ATTENTE
2. Action == CREATE et État == IN_PROGRESS
3. Action == CRÉATION et état == ÉCHEC
4. Action == MISE À JOUR et état == EN ATTENTE
5. Action == MISE À JOUR et état == IN_PROGRESS

En résumé, vous ne pouvez mettre à jour une vue du catalogue de données que si elle répond aux exigences suivantes.

1. Il a été créé avec succès pour la première fois.
 - a. Action == CRÉATION et état == SUCCÈS
2. Il a atteint l'état terminal après une opération de mise à jour asynchrone.
 - a. Action == MISE À JOUR et état == SUCCÈS
 - b. Action == MISE À JOUR et état == ÉCHEC
3. Il possède un attribut d'NULL état à la suite d'une mise à jour synchrone.

colle : `DeleteTable`

Aucune modification n'est apportée à cette opération par rapport au `glue>DeleteTable` fonctionnement de n'importe quelle AWS Glue table. Vous pouvez supprimer une vue du catalogue de données quel que soit son état.

colle : `GetTable`

Aucune modification n'est apportée à cette opération par rapport au `glue:GetTable` fonctionnement de n'importe quelle AWS Glue table. Toutefois, vous ne pouvez pas interroger une vue du catalogue de données à partir des moteurs d'analyse tant qu'elle n'a pas été créée avec succès pour la première fois. Action == CREATE and State == SUCCESS. Après avoir créé

une vue du catalogue de données avec succès pour la première fois, vous pouvez interroger la vue quel que soit son statut.

Note

Toutes les informations de cette section s'appliquent à tous les tableaux lusGetTable, APIs tels queGetTables, etSearchTables.

Afficher les scénarios d'échec de création lors d'opérations asynchrones

Les exemples suivants sont représentatifs des types d'erreurs susceptibles de résulter d'appels d'API CreateTable ou de leur UpdateTable affichage. Ils ne sont pas exhaustifs car la surface d'erreur associée aux échecs de requêtes SQL est assez importante.

Scénario 1 : échec de la requête Amazon Redshift

La requête fournie pour Amazon Redshift inclut un nom de table mal orthographié introuvable dans le catalogue de données lors de la validation. L'erreur qui en résulte est affichée dans le Status champ de la GetTable réponse à la vue.

GetTabledemande :

```
{
  "CatalogId": "123456789012",
  "DatabaseName": "async-view-test-db",
  "TableInput": {
    "Name": "view-athena-redshift-72",
    "Description": "This is an atomic operation",
    "StorageDescriptor": {
      "Columns": [
        { "Name": "col1", "Type": "int" },
        { "Name": "col2", "Type": "string" },
        { "Name": "col3", "Type": "double" }
      ]
    },
    "ViewDefinition": {
      "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
      "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-playground-db/table_1" ],
      "Representations": [
        {
```



```

        "Dialect": "ATHENA",
        "DialectVersion": "3",
        "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"\",
        "ValidationConnection": "athena-connection"
    },
    {
        "Dialect": "REDSHIFT",
        "DialectVersion": "1.0",
        "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table__1\";",
        "ValidationConnection": "redshift-connection"
    }
]
}
}
}

```

GetTableréponse :

```

IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",

```

```

        "UpdateTime": "2024-07-11T11:40:06-07:00",
        "Action": "CREATE",
        "State": "FAILED"
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation
using a connection: [Connection Name: redshift-connection | Query Execution
Id: ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu Jul 11 18:39:37 UTC
2024]: Redshift returned error for the statement: ERROR: AwsClientException:
EntityNotFoundException from glue - Entity Not Found"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-72",

```

```

    "DatabaseName": "async-view-test-db",
    "Description": "This is an atomic operation",
    "Retention": 0,
    "StorageDescriptor": {
      "Columns": [
        {
          "Name": "col1",
          "Type": "int"
        },
        {
          "Name": "col2",
          "Type": "string"
        },
        {
          "Name": "col3",
          "Type": "double"
        }
      ],
      "Compressed": false,
      "NumberOfBuckets": 0,
      "SortColumns": [],
      "StoredAsSubDirectories": false
    },
    "TableType": "VIRTUAL_VIEW",
    "IsRegisteredWithLakeFormation": false,
    "CatalogId": "123456789012",
    "IsRowFilteringEnabled": false,
    "VersionId": "-1",
    "DatabaseId": "<databaseID>",
    "ViewDefinition": {
      "IsProtected": true,
      "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
      "SubObjects": [
        "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
      ],
      "Representations": [
        {
          "Dialect": "ATHENA",
          "DialectVersion": "3",
          "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
          "IsStale": false
        }
      ],

```

```

    {
      "Dialect": "REDSHIFT",
      "DialectVersion": "1.0",
      "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-schema\".\"table__1\";",
      "IsStale": false
    }
  ],
  "IsMultiDialectView": true
},
"ViewValidations": [
  {
    "Dialect": "ATHENA",
    "DialectVersion": "3",
    "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db\".\"table_1\"",
    "UpdateTime": "2024-07-11T11:40:06-07:00",
    "State": "SUCCESS"
  },
  {
    "Dialect": "REDSHIFT",
    "DialectVersion": "1.0",
    "ViewValidationText": "SELECT * FROM \"gdc-view-playground-external-schema\".\"table__1\";",
    "UpdateTime": "2024-07-11T11:39:37-07:00",
    "State": "FAILED",
    "Error": {
      "ErrorCode": "QueryExecutionException",
      "ErrorMessage": "Error received during view SQL validation using a connection: [Connection Name: redshift-connection | Query Execution Id: ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu Jul 11 18:39:37 UTC 2024]: Redshift returned error for the statement: ERROR: AwsClientException: EntityNotFoundException from glue - Entity Not Found"
    }
  }
]
}
}
}
}

```

Scénario 2 : connexion Amazon Redshift non valide

Dans l'exemple suivant, la connexion Amazon Redshift est mal formée car elle fait référence à une base de données Amazon Redshift qui n'existe pas dans le point de terminaison fourni. cluster/serverless Amazon Redshift n'est pas en mesure de valider la vue et le Status champ de la GetTable réponse indique l'erreur ("State": "FAILED"provenant d'Amazon Redshift).

GetTabledemande :

```
{
  "CatalogId": "123456789012",
  "DatabaseName": "async-view-test-db",
  "TableInput": {
    "Name": "view-athena-redshift-73",
    "Description": "This is an atomic operation",
    "StorageDescriptor": {
      "Columns": [
        { "Name": "col1", "Type": "int" },
        { "Name": "col2", "Type": "string" },
        { "Name": "col3", "Type": "double" }
      ]
    },
    "ViewDefinition": {
      "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
      "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-playground-db/table_1" ],
      "Representations": [
        {
          "Dialect": "ATHENA",
          "DialectVersion": "3",
          "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".\"table_1\"",
          "ValidationConnection": "athena-connection"
        },
        {
          "Dialect": "REDSHIFT",
          "DialectVersion": "1.0",
          "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-schema\".\"table_1\";",
          "ValidationConnection": "redshift-connection-malformed"
        }
      ]
    }
  }
}
```

```
}
}
```

GetTableréponse :

```
IncludeStatusDetails = FALSE
{
  "Table": {
    "Name": "view-athena-redshift-73",
    "DatabaseName": "async-view-test-db",
    "Description": "",
    "CreateTime": "2024-07-11T11:43:27-07:00",
    "UpdateTime": "2024-07-11T11:43:27-07:00",
    "Retention": 0,
    "ViewOriginalText": "",
    "ViewExpandedText": "",
    "TableType": "",
    "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
    "IsRegisteredWithLakeFormation": false,
    "CatalogId": "123456789012",
    "IsRowFilteringEnabled": false,
    "VersionId": "-1",
    "DatabaseId": "<databaseID>",
    "IsMultiDialectView": false,
    "Status": {
      "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "RequestTime": "2024-07-11T11:43:27-07:00",
      "UpdateTime": "2024-07-11T11:43:40-07:00",
      "Action": "CREATE",
      "State": "FAILED"
    }
  }
}

IncludeStatusDetails = TRUE
{
  "Table": {
    "Name": "view-athena-redshift-73",
    "DatabaseName": "async-view-test-db",
    "Description": "",
    "CreateTime": "2024-07-11T11:43:27-07:00",
    "UpdateTime": "2024-07-11T11:43:27-07:00",
```

```

"Retention": 0,
"ViewOriginalText": "",
"ViewExpandedText": "",
"TableType": "",
"CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
"IsRegisteredWithLakeFormation": false,
"CatalogId": "123456789012",
"IsRowFilteringEnabled": false,
"VersionId": "-1",
"DatabaseId": "<databaseID>",
"IsMultiDialectView": false,
"Status": {
  "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
  "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
  "RequestTime": "2024-07-11T11:43:27-07:00",
  "UpdateTime": "2024-07-11T11:43:40-07:00",
  "Action": "CREATE",
  "State": "FAILED",
  "Error": {
    "ErrorCode": "QueryExecutionException",
    "ErrorMessage": "Error received during view SQL validation using a
connection: [Connection Name: redshift-connection-malformed | Query Execution Id:
69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Timestamp: Thu Jul 11 18:43:38 UTC 2024]:
Redshift returned error for the statement: FATAL: database \"devo00\" does not exist"
  },
  "Details": {
    "RequestedChange": {
      "Name": "view-athena-redshift-73",
      "DatabaseName": "async-view-test-db",
      "Description": "This is an atomic operation",
      "Retention": 0,
      "StorageDescriptor": {
        "Columns": [
          {
            "Name": "col1",
            "Type": "int"
          },
          {
            "Name": "col2",
            "Type": "string"
          },
          {
            "Name": "col3",
            "Type": "double"
          }
        ]
      }
    }
  }
}

```

```

        }
    ],
    "Compressed": false,
    "NumberOfBuckets": 0,
    "SortColumns": [],
    "StoredAsSubDirectories": false
},
"TableType": "VIRTUAL_VIEW",
"IsRegisteredWithLakeFormation": false,
"CatalogId": "123456789012",
"IsRowFilteringEnabled": false,
"VersionId": "-1",
"DatabaseId": "<databaseID>",
"ViewDefinition": {
    "IsProtected": true,
    "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
    "SubObjects": [
        "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
    ],
    "Representations": [
        {
            "Dialect": "ATHENA",
            "DialectVersion": "3",
            "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
            "IsStale": false
        },
        {
            "Dialect": "REDSHIFT",
            "DialectVersion": "1.0",
            "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
            "IsStale": false
        }
    ]
},
    ],
    "IsMultiDialectView": true
},
"ViewValidations": [
    {
        "Dialect": "ATHENA",
        "DialectVersion": "3",

```



```

        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db\n\".\n\"table_1\n\"",
        "UpdateTime": "2024-07-11T11:43:40-07:00",
        "State": "SUCCESS"
    },
    {
        "Dialect": "REDSHIFT",
        "DialectVersion": "1.0",
        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-\nexternal-schema\n\".\n\"table_1\n\";",
        "UpdateTime": "2024-07-11T11:43:38-07:00",
        "State": "FAILED",
        "Error": {
            "ErrorCode": "QueryExecutionException",
            "ErrorMessage": "Error received during view SQL validation\nusing a connection: [Connection Name: redshift-connection-malformed | Query Execution\nId: 69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Time\nstamp: Thu Jul 11 18:43:38 UTC 2024]: Redshift returned error for the statement: FATAL:\ndatabase \"devoovv\" does not exist"
        }
    }
]
}
}
}

```

Scénario 3 : échec de la requête Athena

Le code SQL pour Athena ici n'est pas valide car la requête a mal orthographié le nom de la base de données. La validation de la requête Athena détecte cela et l'erreur qui en résulte est détectée dans l'`StatusObjet` lors d'un appel. `GetTable`

GetTabledemande :

```
{
  "CatalogId": "123456789012",
  "DatabaseName": "async-view-test-db",
  "TableInput": {
    "Name": "view-athena-redshift-70",
    "Description": "This is an atomic operation",
  }
}
```

```

    "StorageDescriptor": {
      "Columns": [
        { "Name": "col1", "Type": "int" },
        { "Name": "col2", "Type": "string" },
        { "Name": "col3", "Type": "double" }
      ]
    },
    "ViewDefinition": {
      "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
      "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-playground-db/table_1" ],
      "Representations": [
        {
          "Dialect": "ATHENA",
          "DialectVersion": "3",
          "ViewOriginalText": "SELECT * FROM \"gdc--view-playground-db\".\"table_1\"",
          "ValidationConnection": "athena-connection"
        },
        {
          "Dialect": "REDSHIFT",
          "DialectVersion": "1.0",
          "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-schema\".\"table_1\";",
          "ValidationConnection": "redshift-connection"
        }
      ]
    }
  }
}

```

GetTableRéponse :

```

IncludeStatusDetails = FALSE
{
  "Table": {
    "Name": "view-athena-redshift-70",
    "DatabaseName": "async-view-test-db",
    "Description": "",
    "CreateTime": "2024-07-11T11:09:53-07:00",
    "UpdateTime": "2024-07-11T11:09:53-07:00",
    "Retention": 0,
    "ViewOriginalText": ""
  }
}

```

```

    "ViewExpandedText": "",
    "TableType": "",
    "CreatedBy": "arn:aws:iam::123456789012:user/",
    "IsRegisteredWithLakeFormation": false,
    "CatalogId": "123456789012",
    "IsRowFilteringEnabled": false,
    "VersionId": "-1",
    "DatabaseId": "<databaseID>",
    "IsMultiDialectView": false,
    "Status": {
      "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "RequestTime": "2024-07-11T11:09:54-07:00",
      "UpdateTime": "2024-07-11T11:10:41-07:00",
      "Action": "CREATE",
      "State": "FAILED",
    }
  }
}

```

IncludeStatusDetails = TRUE

```

{
  "Table": {
    "Name": "view-athena-redshift-70",
    "DatabaseName": "async-view-test-db",
    "Description": "",
    "CreateTime": "2024-07-11T11:09:53-07:00",
    "UpdateTime": "2024-07-11T11:09:53-07:00",
    "Retention": 0,
    "ViewOriginalText": "",
    "ViewExpandedText": "",
    "TableType": "",
    "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
    "IsRegisteredWithLakeFormation": false,
    "CatalogId": "123456789012",
    "IsRowFilteringEnabled": false,
    "VersionId": "-1",
    "DatabaseId": "<databaseID>",
    "IsMultiDialectView": false,
    "Status": {
      "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "RequestTime": "2024-07-11T11:09:54-07:00",
      "UpdateTime": "2024-07-11T11:10:41-07:00",
    }
  }
}

```

```

    "Action": "CREATE",
    "State": "FAILED",
    "Error": {
      "ErrorCode": "QueryExecutionException",
      "ErrorMessage": "Error received during view SQL validation using
a connection: [Connection Name: athena-connection | Query Execution Id: d9bb1e6d-
ce26-4b35-8276-8a199af966aa | Timestamp: Thu Jul 11 18:10:
41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType: 1301,Retryable:
false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does not exist}"
    },
    "Details": {
      "RequestedChange": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "This is an atomic operation",
        "Retention": 0,
        "StorageDescriptor": {
          "Columns": [
            {
              "Name": "col1",
              "Type": "int"
            },
            {
              "Name": "col2",
              "Type": "string"
            },
            {
              "Name": "col3",
              "Type": "double"
            }
          ],
          "Compressed": false,
          "NumberOfBuckets": 0,
          "SortColumns": [],
          "StoredAsSubDirectories": false
        },
        "TableType": "VIRTUAL_VIEW",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "ViewDefinition": {
          "IsProtected": true,

```

```

        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
        "SubObjects": [
            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
        ],
        "Representations": [
            {
                "Dialect": "ATHENA",
                "DialectVersion": "3",
                "ViewOriginalText": "SELECT * FROM \"gdc--view-
playground-db\".\"table_1\"",
                "IsStale": false
            },
            {
                "Dialect": "REDSHIFT",
                "DialectVersion": "1.0",
                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                "IsStale": false
            }
        ]
    },
    "IsMultiDialectView": true
},
"ViewValidations": [
    {
        "Dialect": "ATHENA",
        "DialectVersion": "3",
        "ViewValidationText": "SELECT * FROM \"gdc--view-playground-db
\".\"table_1\"",
        "UpdateTime": "2024-07-11T11:10:41-07:00",
        "State": "FAILED",
        "Error": {
            "ErrorCode": "QueryExecutionException",
            "ErrorMessage": "Error received during view SQL validation
using a connection: [Connection Name: athena-connection | Query Execution Id:
d9bb1e6d-ce26-4b35-8276-8a199af966aa | Timestamp: Thu J
ul 11 18:10:41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType:
1301,Retryable: false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does
not exist}"
        }
    },
    {
        "Dialect": "REDSHIFT",

```

```

        "DialectVersion": "1.0",
        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-external-schema\".\"table_1\";",
        "UpdateTime": "2024-07-11T11:10:41-07:00",
        "State": "SUCCESS"
    }
}
}
}
}
}

```

Scénario 4 : les descripteurs de stockage ne correspondent pas

Le code SQL fourni pour le dialecte Athena sélectionne col1 et col2 tandis que le SQL pour Redshift sélectionne uniquement. col1 Cela entraîne une erreur de non-concordance du descripteur de stockage.

GetTabledemande :

```

{
  "CatalogId": "123456789012",
  "DatabaseName": "async-view-test-db",
  "TableInput": {
    "Name": "view-athena-redshift-71",
    "Description": "This is an atomic operation",
    "StorageDescriptor": {
      "Columns": [
        { "Name": "col1", "Type": "int" },
        { "Name": "col2", "Type": "string" },
        { "Name": "col3", "Type": "double" }
      ]
    },
    "ViewDefinition": {
      "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
      "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-playground-db/table_1" ],
      "Representations": [
        {
          "Dialect": "ATHENA",
          "DialectVersion": "3",
          "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-playground-db\".\"table_1\"",

```

```

        "ValidationConnection": "athena-connection"
    },
    {
        "Dialect": "REDSHIFT",
        "DialectVersion": "1.0",
        "ViewOriginalText": "SELECT col1 FROM \"gdc-view-playground-external-schema\".\"table_1\";",
        "ValidationConnection": "redshift-connection"
    }
]
}
}
}
}

```

GetTableréponse :

```
IncludeStatusDetails = FALSE
```

```

{
  "Table": {
    "Name": "view-athena-redshift-71",
    "DatabaseName": "async-view-test-db",
    "Description": "",
    "CreateTime": "2024-07-11T11:22:02-07:00",
    "UpdateTime": "2024-07-11T11:22:02-07:00",
    "Retention": 0,
    "ViewOriginalText": "",
    "ViewExpandedText": "",
    "TableType": "",
    "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
    "IsRegisteredWithLakeFormation": false,
    "CatalogId": "123456789012",
    "IsRowFilteringEnabled": false,
    "VersionId": "-1",
    "DatabaseId": "<databaseID>",
    "IsMultiDialectView": false,
    "Status": {
      "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "RequestTime": "2024-07-11T11:22:02-07:00",
      "UpdateTime": "2024-07-11T11:23:19-07:00",
      "Action": "CREATE",
      "State": "FAILED"
    }
  }
}

```

```

    }
  }
}

IncludeStatusDetails = TRUE

{
  "Table": {
    "Name": "view-athena-redshift-71",
    "DatabaseName": "async-view-test-db",
    "Description": "",
    "CreateTime": "2024-07-11T11:22:02-07:00",
    "UpdateTime": "2024-07-11T11:22:02-07:00",
    "Retention": 0,
    "ViewOriginalText": "",
    "ViewExpandedText": "",
    "TableType": "",
    "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
    "IsRegisteredWithLakeFormation": false,
    "CatalogId": "123456789012",
    "IsRowFilteringEnabled": false,
    "VersionId": "-1",
    "DatabaseId": "<databaseID>",
    "IsMultiDialectView": false,
    "Status": {
      "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
      "RequestTime": "2024-07-11T11:22:02-07:00",
      "UpdateTime": "2024-07-11T11:23:19-07:00",
      "Action": "CREATE",
      "State": "FAILED",
      "Error": {
        "ErrorCode": "InvalidInputException",
        "ErrorMessage": "Engine and existing storage descriptor mismatch"
      },
    },
    "Details": {
      "RequestedChange": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
        "Description": "This is an atomic operation",
        "Retention": 0,
        "StorageDescriptor": {
          "Columns": [
            {

```



```

        "Name": "col1",
        "Type": "int"
    },
    {
        "Name": "col2",
        "Type": "string"
    },
    {
        "Name": "col3",
        "Type": "double"
    }
],
"Compressed": false,
"NumberOfBuckets": 0,
"SortColumns": [],
"StoredAsSubDirectories": false
},
"TableType": "VIRTUAL_VIEW",
"IsRegisteredWithLakeFormation": false,
"CatalogId": "123456789012",
"IsRowFilteringEnabled": false,
"VersionId": "-1",
"DatabaseId": "<databaseID>",
"ViewDefinition": {
    "IsProtected": true,
    "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
    "SubObjects": [
        "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
    ],
    "Representations": [
        {
            "Dialect": "ATHENA",
            "DialectVersion": "3",
            "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
            "IsStale": false
        },
        {
            "Dialect": "REDSHIFT",
            "DialectVersion": "1.0",
            "ViewOriginalText": "SELECT col1 FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
            "IsStale": false
        }
    ]
}

```

```
    },
    "IsMultiDialectView": true
},
"ViewValidations": [
  {
    "Dialect": "ATHENA",
    "DialectVersion": "3",
    "ViewValidationText": "SELECT col1, col2 FROM \"gdc-view-playground-db\".\"table_1\"",
    "UpdateTime": "2024-07-11T11:23:19-07:00",
    "State": "FAILED",
    "Error": {
      "ErrorCode": "InvalidInputException",
      "ErrorMessage": "Engine and existing storage descriptor mismatch"
    }
  },
  {
    "Dialect": "REDSHIFT",
    "DialectVersion": "1.0",
    "ViewValidationText": "SELECT col1 FROM \"gdc-view-playground-external-schema\".\"table_1\";",
    "UpdateTime": "2024-07-11T11:22:49-07:00",
    "State": "FAILED",
    "Error": {
      "ErrorCode": "InvalidInputException",
      "ErrorMessage": "Engine and existing storage descriptor mismatch"
    }
  }
]
}
}
```

Octroi d'autorisations sur les vues du catalogue de données

Après avoir créé des vues dans le AWS Glue Data Catalog, vous pouvez accorder des autorisations de lac de données sur Comptes AWS les vues aux responsables des organisations et des unités organisationnelles. Vous pouvez accorder des autorisations à l'aide de balises LF ou de

la méthode de ressource nommée. Pour plus d'informations sur le balisage des ressources, consultez [Contrôle d'accès basé sur des balises Lake Formation](#). Pour plus d'informations sur l'octroi direct d'autorisations sur les vues, consultez [Octroi d'autorisations sur les vues à l'aide de la méthode de ressource nommée](#).

Vues matérialisées

Rubriques

- [Différenciation des vues matérialisées des autres types de vues](#)
- [Cas d'utilisation](#)
- [Concepts clés](#)
- [Autorisations pour les vues matérialisées](#)
- [Création et gestion de vues matérialisées](#)
- [Stockage et accès aux données](#)
- [Intégration avec AWS Lake Formation les autorisations](#)
- [Surveillance et débogage](#)
- [Gestion des tâches d'actualisation](#)
- [Surveillance et résolution des problèmes](#)
- [Considérations et restrictions](#)

Dans le catalogue de AWS Glue données, une vue matérialisée est une table gérée qui stocke le résultat précalculé d'une requête SQL au format Apache Iceberg. Contrairement aux vues standard du catalogue de données qui exécutent la requête à chaque accès, les vues matérialisées stockent physiquement les résultats de la requête et les mettent à jour à mesure que les tables sources sous-jacentes changent. Vous pouvez créer des vues matérialisées à l'aide d'Apache Spark version 3.5.6+ dans Amazon Athena, Amazon EMR ou. AWS Glue

Les vues matérialisées font référence aux tables Apache Iceberg enregistrées dans le catalogue de AWS Glue données, les données précalculées étant stockées sous forme de tables Apache Iceberg dans des compartiments Amazon S3 Tables ou des compartiments à usage général Amazon S3, ce qui les rend accessibles depuis plusieurs moteurs de requêtes, notamment Amazon Athena, Amazon Redshift et des moteurs tiers compatibles avec Iceberg.

Différenciation des vues matérialisées des autres types de vues

Les vues matérialisées diffèrent fondamentalement AWS Glue des vues du catalogue de données, des vues Apache Spark et des vues Amazon Athena. Alors que les vues du catalogue de données sont des tables virtuelles qui exécutent la définition de la requête SQL à chaque accès, les vues matérialisées stockent physiquement les résultats des requêtes précalculés. Cela élimine les calculs redondants et améliore considérablement les performances des requêtes pour les transformations complexes fréquemment utilisées.

Les vues matérialisées diffèrent également des pipelines de transformation de données traditionnels créés avec des tâches AWS Glue ETL ou Spark personnalisées. Au lieu d'écrire du code personnalisé pour gérer la détection des modifications, les mises à jour incrémentielles et l'orchestration du flux de travail, vous définissez des vues matérialisées à l'aide de la syntaxe SQL standard. Le catalogue de AWS Glue données surveille automatiquement les tables sources, détecte les modifications et actualise les vues matérialisées à l'aide d'une infrastructure informatique entièrement gérée.

Cas d'utilisation

Les principaux cas d'utilisation des vues matérialisées sont les suivants :

- Accélérez les requêtes analytiques complexes : créez des vues matérialisées qui précalculent les jointures, les agrégations et les fonctions de fenêtre coûteuses. Les moteurs Spark réécrivent automatiquement les requêtes suivantes pour utiliser les résultats précalculés, réduisant ainsi la latence des requêtes et les coûts de calcul.
- Simplifiez les pipelines de transformation des données : remplacez les tâches ETL complexes qui gèrent la détection des modifications, les mises à jour incrémentielles et l'orchestration du flux de travail par de simples définitions de vues matérialisées basées sur SQL. Le catalogue AWS Glue de données gère automatiquement toute la complexité opérationnelle.
- Activez l'analyse en libre-service avec un accès régi aux données : créez des vues matérialisées sélectionnées qui transforment les données brutes en ensembles de données prêts à être utilisés par l'entreprise. Accordez aux utilisateurs l'accès à des vues matérialisées sans exposer les tables sources sous-jacentes, ce qui simplifie la gestion de la sécurité tout en renforçant les analyses en libre-service.
- Optimisez l'ingénierie des fonctionnalités pour l'apprentissage automatique : définissez des vues matérialisées qui implémentent des transformations de fonctionnalités pour les modèles ML. La fonctionnalité d'actualisation automatique garantit que les magasins de fonctionnalités restent à

jour au fur et à mesure de l'évolution des données sources, tandis que l'actualisation incrémentielle minimise les coûts de calcul.

- Mettez en œuvre un partage de données efficace : créez des vues matérialisées qui filtrent et transforment les données pour des consommateurs spécifiques. Partagez des vues matérialisées entre les comptes et les régions en AWS Lake Formation éliminant le besoin de duplication des données tout en maintenant une gouvernance centralisée.

Concepts clés

Actualisation automatique

L'actualisation automatique est une fonctionnalité qui surveille en permanence vos tables sources et met à jour les vues matérialisées selon un calendrier que vous définissez. Lorsque vous créez une vue matérialisée, vous pouvez spécifier une fréquence d'actualisation à l'aide d'une planification basée sur le temps avec des intervalles allant jusqu'à une heure. Le catalogue de AWS Glue données utilise l'infrastructure informatique gérée de Spark pour exécuter les opérations d'actualisation en arrière-plan, en gérant de manière transparente tous les aspects de la détection des modifications et des mises à jour incrémentielles.

Lorsque les données source changent entre les intervalles d'actualisation, la vue matérialisée devient temporairement obsolète. Les requêtes accédant directement à la vue matérialisée peuvent renvoyer des résultats obsolètes jusqu'à la fin de la prochaine actualisation planifiée. Pour les scénarios nécessitant un accès immédiat aux données les plus récentes, vous pouvez exécuter une actualisation manuelle à l'aide de la commande `REFRESH MATERIALIZED VIEW SQL`.

Actualisation incrémentielle

L'actualisation incrémentielle est une technique d'optimisation qui traite uniquement les données modifiées dans les tables sources depuis la dernière actualisation, plutôt que de recalculer l'intégralité de la vue matérialisée. Le catalogue de AWS Glue données utilise la couche de métadonnées d'Apache Iceberg pour suivre efficacement les modifications apportées aux tables sources et déterminer quelles parties de la vue matérialisée nécessitent des mises à jour.

Cette approche réduit considérablement les coûts de calcul et la durée d'actualisation par rapport aux opérations d'actualisation complètes, en particulier pour les grands ensembles de données où seul un faible pourcentage des données change entre les cycles d'actualisation. Le mécanisme d'actualisation incrémentielle fonctionne automatiquement ; il n'est pas nécessaire d'écrire une logique personnalisée pour détecter ou traiter les données modifiées.

Réécriture automatique des requêtes

La réécriture automatique des requêtes est une fonctionnalité d'optimisation des requêtes disponible dans les moteurs Spark d'Amazon Athena, Amazon EMR et AWS Glue. Lorsque vous exécutez une requête sur des tables de base, l'optimiseur Spark analyse votre plan de requête et détermine automatiquement si les vues matérialisées disponibles peuvent répondre plus efficacement à la requête. S'il existe une vue matérialisée appropriée, l'optimiseur réécrit la requête de manière transparente afin d'utiliser les résultats précalculés au lieu de traiter les tables de base.

Cette optimisation s'effectue sans qu'il soit nécessaire de modifier le code de votre application ou les instructions de requête. L'optimiseur Spark garantit que la réécriture automatique des requêtes ne s'applique que lorsque la vue matérialisée est actuelle et peut produire des résultats précis. Si une vue matérialisée est obsolète ou ne répond pas entièrement aux exigences de la requête, l'optimiseur exécute le plan de requête d'origine par rapport aux tables de base, en privilégiant l'exactitude par rapport aux performances.

Afficher le rôle du définisseur

Une vue matérialisée fonctionne en fonction des autorisations du rôle IAM qui l'a créée, connu sous le nom de rôle de définition de vue. Le rôle de définisseur doit avoir un accès en lecture à toutes les tables de base référencées dans la définition de la vue matérialisée et créer des autorisations de table sur la base de données cible. Lorsque le catalogue de AWS Glue données actualise une vue matérialisée, il assume le rôle de définisseur pour accéder aux tables sources et écrire les résultats mis à jour.

Ce modèle de sécurité vous permet d'accorder aux utilisateurs l'accès aux vues matérialisées sans leur accorder d'autorisations directes sur les tables sources sous-jacentes. Si le rôle de définition de vues perd l'accès à une table de base, les opérations d'actualisation suivantes échoueront jusqu'à ce que les autorisations soient restaurées.

Autorisations pour les vues matérialisées

Pour créer et gérer des vues matérialisées, vous devez configurer AWS Lake Formation les autorisations. Le rôle IAM qui crée la vue matérialisée (le rôle de définition) nécessite des autorisations spécifiques sur les tables sources et les bases de données cibles.

Autorisations requises pour le rôle de définisseur

Le rôle de définition doit disposer des autorisations Lake Formation suivantes :

- Sur les tables sources : autorisations SELECT ou ALL sans filtres de ligne, de colonne ou de cellule
- Sur la base de données cible : autorisation CREATE_TABLE
- Sur le catalogue AWS Glue de données GetTable et les autorisations CreateTable d'API

Lorsque vous créez une vue matérialisée, l'ARN du rôle de définition est stocké dans la définition de la vue. Le catalogue AWS Glue de données assume ce rôle lors de l'exécution des opérations d'actualisation automatique. Si le rôle de définition perd l'accès aux tables sources, les opérations d'actualisation échoueront tant que les autorisations ne seront pas restaurées.

Autorisations IAM pour les tâches AWS Glue

Le rôle IAM de votre AWS Glue travail nécessite les autorisations suivantes :

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetCatalog",
        "glue:GetCatalogs",
        "glue:GetTable",
        "glue:GetTables",
        "glue:CreateTable",
        "glue:UpdateTable",
        "glue>DeleteTable",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "cloudwatch:PutMetricData"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:PutObject",
        "s3:DeleteObject"
      ],
      "Resource": [
```

```

        "arn:aws:s3:::amzn-s3-demo-bucket/*"
    ],
    },
    {
        "Effect": "Allow",
        "Action": [
            "s3:ListBucket"
        ],
        "Resource": [
            "arn:aws:s3:::amzn-s3-demo-bucket"
        ]
    },
    {
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:CreateLogStream",
            "logs:PutLogEvents"
        ],
        "Resource": [
            "arn:aws:logs:*:*:*:/aws-glue/*"
        ]
    },
    {
        "Effect": "Allow",
        "Action": [
            "lakeformation:GetDataAccess"
        ],
        "Resource": "*"
    }
]
}

```

Le rôle que vous utilisez pour l'actualisation automatique de Materialized View doit disposer de l'PassRole autorisation iam : sur le rôle.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [

```



```
    "iam:PassRole"
  ],
  "Resource": [
    "arn:aws:iam::111122223333:role/materialized-view-role-name"
  ]
}
]
```

Pour que Glue actualise automatiquement la vue matérialisée pour vous, le rôle doit également respecter la politique de confiance suivante qui permet au service d'assumer le rôle.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "glue.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Si la vue matérialisée est stockée dans des compartiments de tables S3, vous devez également ajouter l'autorisation suivante au rôle.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3tables:PutTableMaintenanceConfiguration"
      ],
      "Resource": "arn:aws:s3tables:*:123456789012:*"
    }
  ]
}
```

```
]
}
```

Octroi de l'accès aux vues matérialisées

Pour autoriser d'autres utilisateurs à interroger une vue matérialisée, accordez l' AWS Lake Formation autorisation SELECT sur la table des vues matérialisées. Les utilisateurs peuvent interroger la vue matérialisée sans avoir besoin d'accéder directement aux tables sources sous-jacentes.

Pour des informations détaillées sur la configuration des autorisations de Lake Formation, consultez la section Octroi et révocation d'autorisations sur les ressources du catalogue de données dans le manuel du AWS Lake Formation développeur.

Création et gestion de vues matérialisées

Vous créez des vues matérialisées à l'aide de l'CREATE MATERIALIZED VIEW instruction SQL dans les moteurs Spark. La définition de la vue spécifie la requête SQL qui définit la logique de transformation, la base de données cible et le nom de la table, ainsi que la configuration d'actualisation facultative. Vous pouvez définir des transformations complexes, notamment des agrégations, des jointures entre plusieurs tables, des filtres et des fonctions de fenêtre.

```
CREATE MATERIALIZED VIEW sales_summary
AS
SELECT
    region,
    product_category,
    SUM(sales_amount) as total_sales,
    COUNT(DISTINCT customer_id) as unique_customers
FROM sales_transactions
WHERE transaction_date >= current_date - interval '90' day
GROUP BY region, product_category;
```

Pour configurer l'actualisation automatique, incluez le calendrier d'actualisation dans la définition de votre vue :

```
CREATE MATERIALIZED VIEW sales_summary
SCHEDULE REFRESH EVERY 1 HOUR
AS
```

```
SELECT region, product_category, SUM(sales_amount) as total_sales
FROM sales_transactions
GROUP BY region, product_category;
```

Vous pouvez actualiser manuellement une vue matérialisée à tout moment à l'aide de la `REFRESH MATERIALIZED VIEW` commande :

```
REFRESH MATERIALIZED VIEW sales_summary;
```

Pour modifier le calendrier d'actualisation d'une vue matérialisée existante, utilisez l'`ALTER MATERIALIZED VIEW` instruction suivante :

```
ALTER MATERIALIZED VIEW sales_summary
ADD SCHEDULE REFRESH EVERY 2 HOURS;
```

Vues matérialisées imbriquées

Vous pouvez créer des vues matérialisées qui font référence à d'autres vues matérialisées sous forme de tables de base, ce qui permet des transformations de données en plusieurs étapes. Lorsque vous créez des vues matérialisées imbriquées, le catalogue de AWS Glue données suit les dépendances et propage automatiquement les mises à jour dans la hiérarchie des vues matérialisées. Lorsqu'une vue matérialisée de base est actualisée, toutes les vues matérialisées en aval qui en dépendent sont mises à jour en conséquence.

Cette fonctionnalité vous permet de décomposer les transformations complexes en étapes logiques, d'améliorer la maintenabilité et de permettre l'actualisation sélective des couches de transformation en fonction de vos exigences en matière de fraîcheur des données.

Stockage et accès aux données

Les vues matérialisées stockent les résultats précalculés sous forme de tables Apache Iceberg dans des compartiments S3 Tables ou de compartiments S3 à usage général au sein de votre compte. AWS Le catalogue de AWS Glue données gère tous les aspects de la maintenance des tables Iceberg, y compris le compactage et la conservation des instantanés, grâce aux fonctionnalités d'optimisation automatisées de S3 Tables.

Les vues matérialisées étant stockées sous forme de tables Iceberg, vous pouvez les lire directement à partir de n'importe quel moteur compatible avec Iceberg, notamment Amazon Athena, Amazon

Redshift et des plateformes d'analyse tierces. Cette accessibilité multimoteur garantit que vos données précalculées restent accessibles dans l'ensemble de votre écosystème d'analyse sans duplication des données ni conversion de format.

Intégration avec AWS Lake Formation les autorisations

Vous pouvez l'utiliser AWS Lake Formation pour gérer des autorisations précises sur les vues matérialisées. Le créateur de la vue devient automatiquement propriétaire de la vue matérialisée et peut accorder des autorisations à d'autres utilisateurs ou rôles à l'aide AWS Lake Formation de sa méthode de ressource nommée ou de balises LF.

Lorsque vous accordez à un utilisateur l'`SELECT` autorisation d'accéder à une vue matérialisée, il peut interroger les résultats précalculés sans avoir besoin d'accéder aux tables sources sous-jacentes. Ce modèle de sécurité simplifie la gestion de l'accès aux données et vous permet de mettre en œuvre le principe du moindre privilège, en permettant aux utilisateurs d'accéder uniquement aux transformations de données spécifiques dont ils ont besoin.

Vous pouvez partager des vues matérialisées entre AWS comptes, AWS organisations et unités organisationnelles à l'aide des fonctionnalités AWS Lake Formation de partage entre comptes. Vous pouvez également accéder à des vues matérialisées dans toutes AWS les régions à l'aide de liens vers des ressources, ce qui permet une gouvernance centralisée des données avec un accès distribué aux données.

Surveillance et débogage

Le catalogue de AWS Glue données publie toutes les opérations d'actualisation des vues matérialisées et les métriques associées sur Amazon CloudWatch. Vous pouvez surveiller l'heure de début, l'heure de fin, la durée, le volume de données traité et l'état de l'actualisation grâce à CloudWatch des métriques. Lorsque les opérations d'actualisation échouent, les messages d'erreur et les informations de diagnostic sont enregistrés dans CloudWatch les journaux.

Vous pouvez configurer des CloudWatch alarmes pour recevoir des notifications lorsque les tâches d'actualisation dépassent la durée prévue ou échouent à plusieurs reprises. Le catalogue de AWS Glue données publie également les événements de modification relatifs aux exécutions d'actualisation réussies ou non, ce qui vous permet d'intégrer les opérations de visualisation matérialisée dans une automatisation plus large des flux de travail.

Pour vérifier l'état actuel d'une vue matérialisée, utilisez la commande `DESCRIBE MATERIALIZED VIEW SQL`, qui renvoie des métadonnées telles que l'état d'obsolescence, l'horodatage de la dernière actualisation et la configuration du calendrier d'actualisation.

Gestion des tâches d'actualisation

Lancer une actualisation manuelle

Déclenchez une actualisation immédiate en dehors de l'intervalle planifié.

Autorisation requise : les AWS informations d'identification utilisées pour effectuer l'appel d'API doivent être `glue:GetTable` autorisées à accéder à la vue matérialisée.

Pour le catalogue de tables S3 :

```
aws glue start-materialized-view-refresh-task-run \  
  --catalog-id <ACCOUNT_ID>:s3tablescatalog/<CATALOG_NAME> \  
  --database-name <DATABASE_NAME> \  
  --table-name <MV_TABLE_NAME>
```

Pour Root Catalog :

```
aws glue start-materialized-view-refresh-task-run \  
  --catalog-id <ACCOUNT_ID> \  
  --database-name <DATABASE_NAME> \  
  --table-name <MV_TABLE_NAME>
```

Vérification de l'état d'actualisation

Obtenez le statut d'une tâche d'actualisation spécifique :

```
aws glue get-materialized-view-refresh-task-run \  
  --catalog-id <CATALOG_ID> \  
  --materialized-view-refresh-task-run-id <TASK_RUN_ID>
```

Historique d'actualisation des annonces

Affichez toutes les tâches d'actualisation pour obtenir une vue matérialisée :

```
aws glue list-materialized-view-refresh-task-runs \  
  --catalog-id <CATALOG_ID> \  
  --database-name <DATABASE_NAME> \  
  --table-name <MV_TABLE_NAME>
```

 Note

<ACCOUNT_ID>:s3tables/catalog/<CATALOG_NAME> À utiliser pour les tables S3 ou <ACCOUNT_ID> pour le catalogue racine.

Arrêt d'une actualisation en cours

Annuler une tâche d'actualisation en cours :

```
aws glue stop-materialized-view-refresh-task-run \  
  --catalog-id <CATALOG_ID> \  
  --database-name <DATABASE_NAME> \  
  --table-name <MV_TABLE_NAME>
```

Surveillance et résolution des problèmes

Il existe trois méthodes pour surveiller les tâches d'actualisation des vues matérialisées :

CloudWatch Métriques

Consultez les statistiques agrégées pour toutes vos tâches d'actualisation de vues matérialisées dans CloudWatch :

Métriques disponibles :

- AWS/Glue l'espace de noms avec les dimensions suivantes :
 - CatalogId: identifiant de votre catalogue
 - DatabaseName: base de données contenant la vue matérialisée
 - TableName: nom de la vue matérialisée
 - TaskType: défini sur « MaterializedViewRefresh »

Affichage dans la console :

1. Accédez à la CloudWatch console → Métriques
2. Sélectionnez l'espace de AWS noms /Glue
3. Filtrer par dimensions : CatalogId, DatabaseName, TableName, TaskType

4. Afficher les indicateurs de réussite, d'échec et de durée des tâches

Exemple de requête de CloudWatch métriques :

```
{AWS/Glue,CatalogId,DatabaseName,TableName,TaskType} MaterializedViewRefresh
```

En utilisant AWS CLI :

```
aws cloudwatch get-metric-statistics \
  --namespace AWS/Glue \
  --metric-name <MetricName> \
  --dimensions Name=CatalogId,Value=<CATALOG_ID> \
    Name=DatabaseName,Value=<DATABASE_NAME> \
    Name=TableName,Value=<TABLE_NAME> \
    Name=TaskType,Value=MaterializedViewRefresh \
  --start-time <START_TIME> \
  --end-time <END_TIME> \
  --period 3600 \
  --statistics Sum \
  --region <REGION>
```

CloudWatch Journaux

Consultez les journaux d'exécution détaillés pour chaque exécution de tâches d'actualisation :

Groupe de journaux : /aws-glue/materialized-views/<task_run_id>

Où se <task_run_id> trouve un UUID (par exemple, abc12345-def6-7890-ghij-klmnopqrstuv).

Affichage des journaux :

```
# List log streams for a task run
aws logs describe-log-streams \
  --log-group-name /aws-glue/materialized-views/<TASK_RUN_ID> \
  --region <REGION>

# Get log events
aws logs get-log-events \
  --log-group-name /aws-glue/materialized-views/<TASK_RUN_ID> \
  --log-stream-name <LOG_STREAM_NAME> \
  --region <REGION>
```

Dans CloudWatch la console :

1. Accédez à CloudWatch → Groupes de journaux
2. Recherchez /aws-glue/materialized-views/
3. Sélectionnez le groupe de journaux avec votre ID d'exécution de tâche
4. Afficher les journaux d'exécution détaillés, les erreurs et les résultats des tâches Spark

Notifications

Abonnez-vous aux événements pour recevoir des notifications en temps réel concernant les modifications de l'état des tâches d'actualisation :

Types d'événements disponibles :

- La tâche d'actualisation de Glue Materialized View a démarré
- La tâche d'actualisation de Glue Materialized View a réussi
- Echec de la tâche d'actualisation de Glue Materialized View
- Échec de l'invocation de l'actualisation automatique de Glue Materialized View

Création d'une règle :

```
aws events put-rule \  
  --name materialized-view-refresh-notifications \  
  --event-pattern '{  
    "source": ["aws.glue"],  
    "detail-type": [  
      "Glue Materialized View Refresh Task Started",  
      "Glue Materialized View Refresh Task Succeeded",  
      "Glue Materialized View Refresh Task Failed",  
      "Glue Materialized View Auto-Refresh Invocation Failure"  
    ]  
  }' \  
  --region <REGION>
```

Ajouter une cible (par exemple, une rubrique SNS) :

```
aws events put-targets \  
  --rule materialized-view-refresh-notifications \  
  --targets '[{"id": "sns-topic", "arn": "arn:aws:sns:us-east-1:123456789012:my-topic"}']
```



```
--targets "Id"="1", "Arn"="arn:aws:sns:<REGION>:<ACCOUNT_ID>:<TOPIC_NAME>" \  
--region <REGION>
```

Affichage de l'état d'actualisation

Vérifiez l'état de vos tâches d'actualisation de vues matérialisées à l'aide de l' AWS Glue API :

```
aws glue get-materialized-view-refresh-task-run \  
  --catalog-id <CATALOG_ID> \  
  --materialized-view-refresh-task-run-id <TASK_RUN_ID> \  
  --region <REGION>
```

Ou listez toutes les opérations d'actualisation récentes :

```
aws glue list-materialized-view-refresh-task-runs \  
  --catalog-id <CATALOG_ID> \  
  --database-name <DATABASE_NAME> \  
  --table-name <MV_TABLE_NAME> \  
  --region <REGION>
```

Cela montre :

- Heure de la dernière actualisation
- État d'actualisation (RÉUSSI, ÉCHEC, EXÉCUTION, ARRÊT)
- ID d'exécution de la tâche
- Messages d'erreur (en cas d'échec)

États d'actualisation courants :

- EN COURS D'EXÉCUTION : le travail d'actualisation est en cours d'exécution
- RÉUSSI : L'actualisation s'est terminée avec succès
- ÉCHEC : une erreur s'est produite lors de l'actualisation
- ARRÊTÉ : L'actualisation a été annulée manuellement

Résolution des problèmes d'échec des actualisations :

Si l'actualisation échoue, vérifiez :

1. Autorisations IAM : assurez-vous que le rôle de définisseur a accès à toutes les tables de base et à l'emplacement de la vue matérialisée
2. Disponibilité des tables de base : Vérifiez que toutes les tables référencées existent et sont accessibles
3. Validité de la requête : vérifiez que la requête SQL est valide pour le dialecte Spark SQL
4. Limites de ressources : vérifiez si vous avez atteint les limites d'actualisation simultanées de votre compte

Utilisez l' `GetMaterializedViewRefreshTaskRun` API pour récupérer des messages d'erreur détaillés.

Considérations et restrictions

- Les vues matérialisées peuvent uniquement faire référence aux tables Apache Iceberg enregistrées dans le catalogue de AWS Glue données en tant que tables de base.
- La création de vues et la réécriture automatique des requêtes ne sont disponibles qu'à partir des moteurs Spark dans les versions 3.5.6 et supérieures d'Apache Spark sur Amazon Athena AWS Glue , Amazon EMR et (version 5.1).
- Les vues matérialisées sont finalement cohérentes avec les tables de base. Au cours de la fenêtre d'actualisation, les requêtes accédant directement à la vue matérialisée peuvent renvoyer des données obsolètes. Pour accéder immédiatement aux données actuelles, effectuez une actualisation manuelle.
- L'intervalle d'actualisation automatique minimum est d'une heure. Pour les cas d'utilisation nécessitant des mises à jour plus fréquentes, exécutez des actualisations manuelles par programmation à l'aide de la commande. `REFRESH MATERIALIZED VIEW`
- La réécriture des requêtes privilégie l'exactitude par rapport aux performances. Si une vue matérialisée est obsolète ou ne peut pas répondre avec précision aux exigences de la requête, les moteurs Spark exécutent la requête d'origine sur les tables de base.

Importation de données à l'aide de flux de travail dans Lake Formation

Avec AWS Lake Formation, vous pouvez importer vos données à l'aide de flux de travail. Un flux de travail définit la source de données et le calendrier d'importation des données dans votre lac de données. Il s'agit d'un conteneur pour AWS Glue les robots d'exploration, les tâches et les

déclencheurs utilisés pour orchestrer les processus de chargement et de mise à jour du lac de données.

Rubriques

- [Plans et flux de travail dans Lake Formation](#)
- [Création d'un flux de travail](#)
- [Exécution d'un flux de travail](#)

Plans et flux de travail dans Lake Formation

Un flux de travail encapsule une activité complexe d'extraction, de transformation et de chargement (ETL) multi-tâches. Les flux de travail génèrent des AWS Glue robots, des tâches et des déclencheurs pour orchestrer le chargement et la mise à jour des données. Lake Formation exécute et suit un flux de travail en tant qu'entité unique. Vous pouvez configurer un flux de travail pour qu'il s'exécute à la demande ou selon un calendrier.

Note

Spark Parquet Writer ne prend pas en charge les caractères spéciaux dans les noms de colonnes. Il s'agit d'une limitation technique du rédacteur lui-même, et non d'un problème de configuration.

Les flux de travail que vous créez dans Lake Formation sont visibles dans la AWS Glue console sous la forme d'un graphe acyclique dirigé (DAG). Chaque nœud DAG est une tâche, un robot d'exploration ou un déclencheur. Pour suivre les progrès et résoudre les problèmes, vous pouvez suivre l'état de chaque nœud du flux de travail.

Lorsqu'un flux de travail Lake Formation est terminé, l'utilisateur qui l'a exécuté reçoit l'autorisation Lake Formation sur les tables du catalogue de données créées par le flux de travail.

Vous pouvez également créer des flux de travail dans AWS Glue. Cependant, dans la mesure où Lake Formation vous permet de créer un flux de travail à partir d'un plan, la création de flux de travail est beaucoup plus simple et automatisée dans Lake Formation. Lake Formation fournit les types de plans suivants :

- **Instantané de base de données** : charge ou recharge les données de toutes les tables dans le lac de données à partir d'une source JDBC. Vous pouvez exclure certaines données de la source selon un modèle d'exclusion.
- **Base de données incrémentielle** : charge uniquement les nouvelles données dans le lac de données à partir d'une source JDBC, en fonction des signets définis précédemment. Vous spécifiez les tables individuelles à inclure dans la base de données source JDBC. Pour chaque tableau, vous choisissez les colonnes des signets et l'ordre de tri des favoris afin de suivre les données précédemment chargées. La première fois que vous exécutez un plan de base de données incrémentiel sur un ensemble de tables, le flux de travail charge toutes les données des tables et définit des signets pour la prochaine exécution du plan de base de données incrémentiel. Vous pouvez donc utiliser un plan de base de données incrémentiel au lieu du plan de capture de base de données pour charger toutes les données, à condition de spécifier chaque table de la source de données en tant que paramètre.
- **Fichier journal** : charge en bloc des données à partir de sources de fichiers journaux AWS CloudTrail, notamment les journaux Elastic Load Balancing et les journaux Application Load Balancer.

Utilisez le tableau suivant pour déterminer s'il convient d'utiliser un instantané de base de données ou un plan de base de données incrémentiel.

Utilisez un instantané de base de données lorsque...	Utiliser une base de données incrémentielle lorsque...
• L'évolution du schéma est flexible. (Les colonnes sont renommées, les colonnes précédentes sont supprimées et de nouvelles colonnes sont ajoutées à leur place.) • Une cohérence complète est nécessaire entre la source et la destination.	• L'évolution du schéma est progressive. (Il n'y a que des ajouts successifs de colonnes.) • Seules les nouvelles lignes sont ajoutées ; les lignes précédentes ne sont pas mises à jour.

Note

Les utilisateurs ne peuvent pas modifier les plans et les flux de travail créés par Lake Formation.

Création d'un flux de travail

Avant de commencer, assurez-vous d'avoir accordé au rôle les autorisations de données et les autorisations de localisation des données requises `LakeFormationWorkflowRole`. Le flux de travail peut ainsi créer des tables de métadonnées dans le catalogue de données et écrire des données vers des emplacements cibles dans Amazon S3. Pour plus d'informations, consultez [\(Facultatif\) Créez un rôle IAM pour les flux de travail](#) et [Vue d'ensemble des autorisations relatives à Lake Formation](#).

Note

Lake Formation utilise `GetTemplateInstance`/`GetTemplateInstances`, et effectue des `InstantiateTemplate` opérations pour créer des flux de travail à partir de plans. Ces opérations ne sont pas accessibles au public et ne sont utilisées qu'en interne pour créer des ressources en votre nom. Vous recevez des CloudTrail événements pour créer des flux de travail.

Pour créer un flux de travail à partir d'un plan

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant d'autorisations d'ingénieur de données. Pour de plus amples informations, veuillez consulter [Référence des personnalités de Lake Formation et des autorisations IAM](#).
2. Dans le volet de navigation, choisissez Blueprints, puis choisissez Use Blueprint.
3. Sur la page Utiliser un plan, choisissez une vignette pour sélectionner le type de plan.
4. Sous Source d'importation, spécifiez la source de données.

Si vous effectuez une importation à partir d'une source JDBC, spécifiez les éléments suivants :

- Connexion à la base de données —Choisissez une connexion dans la liste. Créez des connexions supplémentaires à l'aide de la AWS Glue console. Le nom d'utilisateur et le mot de passe JDBC de la connexion déterminent les objets de base de données auxquels le flux de travail a accès.
- Chemin des données source :
entrez `<database>/<schema>/<table>` ou `<database>/<table>`, selon le produit de

base de données. Oracle Database et MySQL ne prennent pas en charge le schéma dans le chemin. Vous pouvez remplacer le pourcentage (%) par `<schema>` ou `<table>`. Par exemple, pour une base de données Oracle dont l'identifiant système (SID) est égal `orcl/%` à `orcl`, entrez pour importer toutes les tables auxquelles l'utilisateur nommé dans la connexion a accès.

 Important

Ce champ distingue les majuscules et minuscules. Le flux de travail échouera s'il existe une incompatibilité majuscules/minuscules pour l'un des composants.

Si vous spécifiez une base de données MySQL, AWS Glue ETL utilise le pilote JDBC Mysql5 par défaut, donc My SQL8 n'est pas pris en charge nativement. Vous pouvez modifier le script de tâche ETL pour utiliser un `customJdbcDriverS3Path` paramètre tel que décrit dans la section [JDBC ConnectionType Values](#) du manuel du AWS Glue développeur afin d'utiliser un autre pilote JDBC compatible avec My. SQL8

Si vous effectuez une importation à partir d'un fichier journal, assurez-vous que le rôle que vous spécifiez pour le flux de travail (le « rôle de flux de travail ») dispose des autorisations IAM requises pour accéder à la source de données. Par exemple, pour importer AWS CloudTrail des journaux, l'utilisateur doit disposer des `cloudtrail:LookupEvents` autorisations `cloudtrail:DescribeTrails` et pour consulter la liste des CloudTrail journaux lors de la création du flux de travail, et le rôle du flux de travail doit disposer d'autorisations sur l' CloudTrail emplacement dans Amazon S3.

5. Effectuez l'une des actions suivantes :

- Pour le type de plan instantané de base de données, identifiez éventuellement un sous-ensemble de données à importer en spécifiant un ou plusieurs modèles d'exclusion. Ces modèles d'exclusion sont des modèles de style Unix. `glob` Ils sont stockés en tant que propriété des tables créées par le flux de travail.

Pour plus de détails sur les modèles d'exclusion disponibles, consultez la section [Include et d'exclusion des modèles](#) dans le manuel du AWS Glue développeur.

- Pour le type de plan de base de données incrémentiel, spécifiez les champs suivants. Ajoutez une ligne pour chaque table à importer.

Nom de la table

Tableau à importer. Tout doit être en minuscules.

Clés de signet

Liste de noms de colonnes séparés par des virgules qui définissent les clés des signets. Si ce champ est vide, la clé primaire est utilisée pour déterminer les nouvelles données. Le cas de chaque colonne doit correspondre au cas défini dans la source de données.

Note

La clé primaire est considérée comme la clé de signet par défaut uniquement si elle augmente ou diminue de manière séquentielle (sans interruption). Si vous souhaitez utiliser la clé primaire comme clé de signet et qu'elle comporte des lacunes, vous devez nommer la colonne de clé primaire comme clé de signet.

Marquer la commande

Lorsque vous choisissez Ascending, les lignes dont les valeurs sont supérieures aux valeurs enregistrées dans les favoris sont identifiées comme de nouvelles lignes. Lorsque vous choisissez Décroissant, les lignes dont les valeurs sont inférieures aux valeurs enregistrées dans les favoris sont identifiées comme de nouvelles lignes.

Schéma de partitionnement

(Facultatif) Liste des colonnes clés de partitionnement, délimitées par des barres obliques (/). Exemple : year/month/day.

Incremental data
Enter tables in the data source to import along with bookmark columns to determine previously imported data.

Table name	Bookmark keys	Bookmark order	Partitioning scheme - optional	
Enter a table name	Enter a bookmark key Comma-delimited list of bookmark columns.	Choose a sort. ▼	Type partitioning	Remove
Add				

Pour plus d'informations, consultez la section [Suivi des données traitées à l'aide des signets de tâches](#) dans le Guide du AWS Glue développeur.

6. Sous Cible d'importation, spécifiez la base de données cible, l'emplacement Amazon S3 cible et le format des données.

Assurez-vous que le rôle de flux de travail dispose des autorisations Lake Formation requises sur la base de données et sur l'emplacement cible Amazon S3.

 Note

À l'heure actuelle, les plans ne prennent pas en charge le chiffrement des données sur la cible.

7. Choisissez une fréquence d'importation.

Vous pouvez spécifier une cron expression à l'aide de l'option Personnalisée.

8. Sous Options d'importation :
 - a. Entrez un nom de flux de travail.
 - b. Pour rôle, choisissez le rôle LakeFormationWorkflowRole que vous avez créé dans [\(Facultatif\) Créez un rôle IAM pour les flux de travail](#).
 - c. Spécifiez éventuellement un préfixe de table. Le préfixe est ajouté aux noms des tables du catalogue de données créées par le flux de travail.
9. Choisissez Créer et attendez que la console indique que le flux de travail a été créé avec succès.

 Tip

Avez-vous reçu le message d'erreur suivant ?

```
User: arn:aws:iam::<account-id>:user/<username> is not authorized
to perform: iam:PassRole on resource:arn:aws:iam::<account-
id>:role/<rolename>...
```

Si tel est le cas, vérifiez que vous l'avez *<account-id>* remplacé par un numéro de AWS compte valide dans toutes les polices.

 Voir aussi :

- [Plans et flux de travail dans Lake Formation](#)

Exécution d'un flux de travail

Vous pouvez exécuter un flux de travail à l'aide de la console Lake Formation, de la AWS Glue console, de l'interface de ligne de commande de AWS Glue (AWS CLI) ou de l'API.

Pour exécuter un flux de travail (console Lake Formation)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant d'autorisations d'ingénieur de données. Pour de plus amples informations, veuillez consulter [Référence des personnalités de Lake Formation et des autorisations IAM](#).
2. Dans le panneau de navigation, sélectionnez Blueprints (Modèles).
3. Sur la page Blueprints, sélectionnez le flux de travail. Ensuite, dans le menu Actions, choisissez Démarrer.
4. Au fur et à mesure que le flux de travail s'exécute, consultez sa progression dans la colonne État de la dernière exécution. Cliquez sur le bouton d'actualisation de temps en temps.

Le statut passe de « EN COURS » à « Découverte », puis « Importation », puis « TERMINÉ ».

Lorsque le flux de travail est terminé :

- Le catalogue de données contient de nouvelles tables de métadonnées.
- Vos données sont ingérées dans le lac de données.

Si le flux de travail échoue, procédez comme suit :

- a. Sélectionnez un flux de travail. Choisissez Actions, puis Afficher le graphique.

Le flux de travail s'ouvre dans la AWS Glue console.

- b. Assurez-vous que le flux de travail est sélectionné, puis sélectionnez l'onglet History (Historique).
- c. Sous Historique, sélectionnez l'exécution la plus récente et choisissez Afficher les détails de la course.
- d. Sélectionnez une tâche ou un robot d'exploration ayant échoué dans le graphique dynamique (d'exécution) et consultez le message d'erreur. Les nœuds défectueux sont rouges ou jaunes.

 Voir aussi :

- [Plans et flux de travail dans Lake Formation](#)

Intégrer vos données dans AWS Glue Data Catalog

Vous pouvez créer des catalogues fédérés dans le AWS Glue Data Catalog (catalogue de données) et unifier les données entre les lacs de données Amazon S3 et les entrepôts de données Amazon Redshift. Vous pouvez également intégrer des données provenant de vos bases de données opérationnelles telles que Amazon DynamoDB, et de sources de données tierces telles que PostgreSQL BigQuery, Google, MySQL, entre autres. Le catalogue de données fournit un référentiel de métadonnées centralisé qui facilite la gestion et la découverte de données sur des systèmes disparates.

Le catalogue de données s'intègre à plus de 30 sources de données externes via des connecteurs fédérés. Grâce à cette intégration, vous pouvez interroger des données provenant de ces sources externes sans avoir à créer des pipelines de données pour les intégrer au AWS préalable.

Après avoir catalogué les données externes, vous pouvez les utiliser AWS Lake Formation pour gérer de manière centralisée les autorisations d'accès aux données dans le catalogue de données. Les administrateurs de data lake peuvent accorder des autorisations d'accès détaillées à d'autres principaux IAM (utilisateurs ou rôles) au sein d'un même compte ou entre plusieurs comptes à l'aide d'un contrôle d'accès basé sur des balises (balises LF) et de méthodes de ressources nommées.

En utilisant des balises LF, les administrateurs de données peuvent organiser les ressources de manière logique en fonction d'attributs tels que le domaine et le niveau de sensibilité, simplifiant ainsi la gestion des autorisations tout en garantissant des contrôles d'accès cohérents entre les services d'analyse et d'apprentissage automatique tels qu'Athena, Amazon EMR ou Redshift Spectrum. AWS Glue

Le catalogue de données fournit les méthodes suivantes pour gérer les données et les autorisations sur les ensembles de données externes et les métastores externes :

- Intégrer les compartiments de tables Amazon S3 au catalogue de données : vous pouvez publier et cataloguer les tables Amazon S3 sous forme d'objets du catalogue de données et enregistrer le catalogue en tant que localisation des données de Lake Formation depuis la console Lake Formation ou à l'aide d'opérations d' AWS Glue API.
- Importez les données des entrepôts de données Amazon Redshift dans le AWS Glue Data Catalog — Enregistrez un espace de noms [Amazon Redshift](#) existant ou un cluster dans le catalogue de données, et créez un catalogue fédéré à plusieurs niveaux dans le catalogue de données.

Vous pouvez accéder à vos données à l'aide de n'importe quel moteur de requête compatible avec les spécifications OpenAPI du catalogue REST Apache Iceberg, tel qu'Amazon EMR Serverless et Amazon Athena.

- Fédérez les Iceberg REST catalogues distants au catalogue de données : fédérez les Iceberg REST catalogues distants au catalogue de données et accédez en toute sécurité aux Iceberg tables distantes stockées dans Amazon S3 à l'aide de moteurs d'analyse. AWS
- Créez une fédération dans le catalogue de données à partir de sources de données externes : connectez le catalogue de données à des sources de données externes à l'aide de AWS Glue connexions, et créez des catalogues fédérés pour gérer de manière centralisée les autorisations d'accès aux ensembles de données à l'aide de Lake Formation. Aucune migration de métadonnées dans le catalogue de données n'est nécessaire.
- Créez des catalogues pour gérer les tables Amazon Redshift dans le catalogue de données : vous ne disposez peut-être pas d'un cluster de producteurs Amazon Redshift ou d'un partage de données Amazon Redshift actuellement, mais vous souhaitez créer et gérer des tables Amazon Redshift à l'aide de Data Catalog. Vous pouvez commencer par créer un catalogue AWS Glue géré à l'aide de l'opération `glue:CreateCatalog` API ou de la AWS Lake Formation console en définissant le type de catalogue `Catalog source` comme Redshift. Managed
- Publiez des partages de données Amazon Redshift avec Data Catalog : publiez des partages de données [Amazon Redshift sur](#) Data Catalog et utilisez Lake Formation pour gérer de manière centralisée l'accès aux données des partages de données et restreindre l'accès des utilisateurs.

Vous pouvez interroger vos données à l'aide d'Amazon Redshift Spectrum.

- Connect Data Catalog à des métastores Hive externes : connectez le catalogue de données à des métastores externes pour gérer les autorisations d'accès aux ensembles de données dans Amazon S3 à l'aide de Lake Formation. Aucune migration de métadonnées dans le catalogue de données n'est nécessaire.
- Intégrez Lake Formation à AWS Data Exchange — Lake Formation prend en charge l'accès sous licence à vos données via AWS Data Exchange. Si vous souhaitez obtenir une licence pour vos données de Lake Formation, consultez la section [Contenu AWS Data Exchange](#) du guide de AWS Data Exchange l'utilisateur.

Rubriques

- [Fédération de catalogues vers des catalogues Iceberg distants](#)
- [Intégrer les données Amazon Redshift dans AWS Glue Data Catalog](#)

- [Fédération en sources de données externes dans le AWS Glue Data Catalog](#)
- [Intégration d'Amazon S3 Tables avec AWS Glue Data Catalog et AWS Lake Formation](#)
- [Modification des contrôles d'accès pour l'intégration des tables S3](#)
- [Création d'un catalogue géré par Amazon Redshift dans AWS Glue Data Catalog](#)
- [Gestion des autorisations pour les données dans un partage de données Amazon Redshift](#)
- [Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes](#)

Fédération de catalogues vers des catalogues Iceberg distants

La fédération de catalogues AWS Glue fournit un accès direct et sécurisé aux tables Iceberg, stockées dans Amazon S3 et cataloguées dans des catalogues distants, à l'aide AWS de moteurs d'analyse. La fédération de catalogues synchronise les métadonnées entre Data Catalog et les catalogues distants lorsque vous accédez à des tables distantes. Il est pris en charge par une grande variété de moteurs d'analyse, notamment Amazon Redshift, Amazon EMR, Amazon Athena, des moteurs tiers tels qu' AWS Glue Apache Spark, etc.

La fédération de catalogues AWS Glue Data Catalog permet de communiquer avec les systèmes de catalogues distants pour découvrir les tables et avec Lake Formation pour autoriser l'accès aux données des tables dans Amazon S3. Lorsque vous interrogez une table fédérée, Data Catalog découvre les dernières informations de table dans le catalogue distant au moment de la requête, obtenant ainsi l'emplacement Amazon S3 de la table, le schéma actuel et les informations de partition. Votre moteur d'analyse (Amazon Athena, Amazon Redshift, Amazon EMR) utilise ensuite ces informations pour accéder aux fichiers de données Iceberg directement depuis Amazon S3. Lake Formation gère l'accès aux tables en distribuant des informations d'identification limitées aux données des tables stockées dans Amazon S3, ce qui permet aux moteurs d'appliquer des autorisations détaillées aux tables fédérées.

Caractéristiques de Catalog Federation

Gouverné à l'aide de Lake Formation

Les catalogues d'icebergs fédérés du catalogue de données sont des ressources enregistrées dans Lake Formation, ce qui vous permet d'accorder des autorisations détaillées au niveau des lignes, des colonnes et des cellules aux tables Iceberg des catalogues d'icebergs fédérés à l'aide de subventions Lake Formation. Les catalogues Federated Iceberg et les objets associés peuvent être partagés en toute sécurité entre les comptes. AWS Les catalogues Federated Iceberg fonctionnent également

avec le contrôle d'accès basé sur les balises Lake Formation, ce qui vous permet d'étendre la gouvernance à l'aide de balises.

Configurations réseau

La fédération de catalogues prend en charge les connexions directes aux sources de catalogue distantes à l'aide de la connectivité HTTPS standard. Il prend également en charge la connectivité via Amazon VPC lorsque vous souhaitez maintenir l'isolation du réseau et la connectivité à l'aide de proxy lorsque vous souhaitez une communication sécurisée via les pare-feux de l'entreprise.

Rubriques

- [Fédérer vers le catalogue Snowflake Iceberg](#)
- [Fédérer pour Databricks Unity Catalog](#)

Fédérer vers le catalogue Snowflake Iceberg

AWS Glue Data Catalog se connecte à Snowflake l'utilisation des OAuth2 informations d'identification d'un directeur de Snowflake service. Vous pouvez utiliser cette fédération pour vous connecter à Snowflake Horizon des Snowflake Polaris catalogues. Ce mécanisme d'authentification permet à Data Catalog d'accéder aux métadonnées de divers objets (tels que les catalogues, les bases de données et les tables) de votre catalogue, en fonction des privilèges associés au principal de service. Pour garantir l'accès aux bons objets, il est essentiel d'accorder au principal du service Snowflake les autorisations nécessaires pour lire les métadonnées de ces objets.

Conditions préalables

Avant de créer un catalogue fédéré dans Data Catalog régi par Lake Formation, assurez-vous de disposer des autorisations suivantes :

Votre principal IAM (utilisateur ou rôle) doit disposer des autorisations suivantes :

- Permissions de Lake Formation —`lakeformation:RegisterResource`,
`lakeformation:DescribeResource`
- AWS Glue autorisations —
`glue:CreateConnection`,`glue:CreateCatalog`,`glue:GetConnection`,
`glue:PassConnection`
- Autorisations du Secrets Manager —`secretsmanager:CreateSecret`,
`secretsmanager:GetSecretValue`

- Autorisations IAM —`iam:CreateRole`, `iam:AttachRolePolicy` `iam:PassRole`

Vous devez être un administrateur du lac de Lake Formation ou avoir l'`CREATE_CATALOG` autorisation d'accéder au catalogue de données

Création d'un catalogue fédéré

Utilisation de la console

1. Connectez-vous à la console et ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Choisissez la AWS région préférée dans la section supérieure droite de la page.
3. Dans le volet de navigation de gauche, choisissez Catalogues.
4. Choisissez Créer un catalogue pour ouvrir le flux de travail de création de catalogue.
5. Dans l'étape Choisir une source de données, Snowflake sélectionnez l'une des options disponibles.
6. À l'étape Définir les détails du catalogue, vous fournissez trois informations : les détails du catalogue, les détails de connexion et les détails d'enregistrement.
7. Dans le conteneur des détails du catalogue, attribuez un nom unique à votre catalogue AWS Glue fédéré et entrez le nom du Snowflake catalogue existant.
8. Dans le conteneur des détails des connexions, vous pouvez soit choisir une connexion existante à laquelle vous avez accès, soit fournir une configuration pour créer un nouveau connecteur.
9. Les nouvelles configurations de connexion incluent :
 - Nom de connexion : nom unique de l'objet de AWS Glue connexion.
 - URL de l'instance : URL du point de terminaison de votre Snowflake compte existant.
 - Authentification — Spécifiez la configuration d'authentification AWS Glue utilisée pour se connecter au serveur de catalogue distant. AWS Glue prend en charge les deux OAuth2 et l'authentification personnalisée.
 - URL du jeton : spécifiez l'URL du fournisseur d'identité du catalogue distant.
 - OAuth2 ID client — Spécifiez l'ID client des OAuth2 informations d'identification associées à votre catalogue distant.
 - Secret — Stockez et utilisez OAuth2 le secret du client en utilisant AWS Secrets Manager ou entrez la valeur du secret dans la zone de texte. Lorsque vous entrez le secret manuellement dans la console, AWS Glue crée le secret en votre nom.

- Étendue de l'URL du jeton : spécifiez l' OAuth étendue de l'authentification.
 - Filtre de boîtier de catalogue : choisissez d'importer les objets en minuscules ou en majuscules de votre catalogue distant vers Data Catalog.
10. Créez un rôle IAM que les responsables du service Lake Formation AWS Glue et les responsables du service Lake Formation pourront utiliser respectivement pour accéder aux emplacements secrets AWS Secrets Manager et aux emplacements Amazon S3 des tables Iceberg distantes. Sélectionnez le rôle IAM dans le menu déroulant d'inscription. Reportez-vous aux étapes 2 et 3 de la section CLI suivante pour plus de détails sur la politique IAM.
 11. Sélectionnez Tester la connexion pour vérifier si les propriétés de votre connexion et l'accès au rôle IAM sont correctement configurés.
 12. Sélectionnez Suivant pour vérifier vos paramètres.
 13. Sélectionnez Créer un catalogue sur la page de révision.

Utilisation de l'interface de ligne de commande (CLI)

1. Créez un AWS Secrets Manager secret

Le AWS Glue connecteur prend en charge deux types d'authentification : OAuth2 l'authentification personnalisée. Lorsque vous utilisez l' OAuth2 option, AWS Secrets Manager utilisez-la pour stocker le secret client du principal de Snowflake service. Vous utiliserez ce secret ultérieurement lors de la création de la AWS Glue connexion. Pour l'authentification personnalisée, utilisez-le AWS Secrets Manager pour stocker et récupérer le jeton d'accès.

Dans l'exemple suivant, remplacez *your-snowflake-secretclient_secret*, *region* par vos propres informations.

```
aws secretsmanager create-secret \  
--name your-snowflake-secret \  
--description "Snowflake secret" \  
--secret-string '{  
  "USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET": "client_secret"  
}' \  
--region region
```


 Note

`USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET` est un mot clé réservé AWS Glue utilisé pour faire référence à une valeur secrète du client dans le secret. Utilisez également le même mot clé lorsque vous créez le secret dans la console Lake Formation.

2. Créez un rôle IAM qui permet à l'objet de AWS Glue connexion d'accéder au secret créé à l'étape précédente

L'objet de AWS Glue connexion a besoin d'accéder au AWS Secrets Manager secret lorsque vous l'utilisez AWS Secrets Manager pour stocker, récupérer et actualiser votre jeton OAuth secret. L'objet de AWS Glue connexion nécessite également un accès pour créer, décrire et utiliser les interfaces réseau Amazon VPC lorsque vous utilisez un point de terminaison Amazon VPC pour restreindre la connectivité à votre compte. Snowflake

Créez une politique IAM et associez-la à un rôle IAM. Ajoutez le principal de AWS Glue service à la politique de confiance.

Dans l'exemple suivant, remplacez *your-secrets-manager-ARN* et *your-subnet-id1* par vos propres informations.

Exemple Stratégie IAM

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue",
      "secretsmanager:DescribeSecret",
      "secretsmanager:PutSecretValue"
    ],
    "Resource": [
      "your-secrets-manager-ARN"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
```

```

        "ec2:CreateNetworkInterface",
        "ec2:DeleteNetworkInterface",
        "ec2:DescribeNetworkInterfaces"
    ],
    "Resource": "*",
    "Condition": {
        "ArnEquals": {
            "ec2:Vpc": "arn:aws:ec2:region:account-id:vpc/your-vpc-id",
            "ec2:Subnet": ["arn:aws:ec2:region:account-id:subnet/your-
subnet-id1"]
        }
    }
}

```

Exemple Stratégie d'approbation

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Principal": {
      "Service": "glue.amazonaws.com"
    },
    "Action": "sts:AssumeRole"
  }]
}

```

3. Créez une politique IAM qui donne à Lake Formation un accès en lecture à l'emplacement Amazon S3 du catalogue

En tant que propriétaire d'un catalogue fédéré dans Data Catalog, vous utilisez Lake Formation pour accorder un accès grossier aux tables, un accès fin (au niveau des colonnes, des lignes et au niveau des cellules) et un accès basé sur des balises à vos équipes chargées des données. Lake Formation utilise un rôle IAM qui lui permet d'accéder aux emplacements Amazon S3 sous-jacents de vos tables Iceberg distantes. Cet accès permet à Lake Formation de vendre des identifiants d'accès délimités aux moteurs d'analyse interrogeant des tables distantes.

Créez une politique IAM et attachez-la à un rôle IAM. Ajoutez le principal de service de Lake Formation à la politique de confiance des rôles.

Dans l'exemple suivant, remplacez *amzn-s3-demo-bucketN* et *your-kms-key* par vos propres informations.

Exemple Stratégie IAM

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::amzn-s3-demo-bucket1/*",
      "arn:aws:s3:::amzn-s3-demo-bucket2/*"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "s3:ListBucket"
    ],
    "Resource": [
      "arn:aws:s3:::amzn-s3-demo-bucket1",
      "arn:aws:s3:::amzn-s3-demo-bucket2"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt",
      "kms:Encrypt"
    ],
    "Resource": [
      "your-kms-key"
    ]
  }
]
```

Exemple Stratégie d'approbation

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Sid": "",
    "Effect": "Allow",
    "Principal": {
      "Service": "lakeformation.amazonaws.com"
    },
    "Action": "sts:AssumeRole"
  }]
}
```

Note

Lorsque vous utilisez la console Lake Formation pour créer un catalogue fédéré, la console utilise un seul rôle IAM auquel les deux politiques sont associées pour terminer la configuration.

4. Création d'un objet AWS Glue de connexion

AWS Glue supports de connecteurs OAuth2 et méthodes d'authentification personnalisées. L'objet de connexion au catalogue de données prend en charge la CATALOG_CASING_FILTER configuration qui vous permet d'importer dans Data Catalog des objets en minuscules ou en majuscules de votre catalogue distant.

L'exemple suivant utilise la configuration OAuth2 d'authentification pour créer une AWS Glue connexion. Remplacez *highlighted sections* par vos informations.

```
aws glue create-connection \
  --connection-input '{
    "Name": "your-glue-connection-to-snowflake-account",
    "ConnectionType": "SNOWFLAKEICEBERGRESTCATALOG",
    "ConnectionProperties": {
      "INSTANCE_URL": "your-snowflake-account-URL",
      "ROLE_ARN": "your-IAM-role-for-secrets-and-VPC-access",
      "CATALOG_CASING_FILTER": "LOWERCASE_ONLY"
    },
    "AuthenticationConfiguration": {
```

```

    "AuthenticationType": "OAUTH2",
    "OAuth2Properties": {
      "OAuth2GrantType": "CLIENT_CREDENTIALS",
      "TokenUrl": "your-internal-or-external-token-server-url",
      "OAuth2ClientApplication": {
        "UserManagedClientApplicationClientId": "our-client-id"
      },
      "TokenUrlParametersMap": {
        "scope": "all-apis"
      }
    },
    "SecretArn": "arn:aws:secretsmanager:your-aws-region:your-aws-account-id:secret:snowflake-secret"
  }
}'

```

5. Enregistrer AWS Glue la connexion en tant que ressource Lake Formation

À l'aide de l'objet de AWS Glue connexion (créé à l'étape 4) et du rôle IAM (créé à l'étape 3), vous pouvez désormais enregistrer l'objet de AWS Glue connexion en tant que ressource gérée par Lake Formation.

Remplacez *your-glue-connector-arn* et *your-IAM-role-ARN-having-LF-access* par vos informations.

```

aws lakeformation register-resource \
  --resource-arn your-glue-connector-arn \
  --role-arn your-IAM-role-ARN-having-LF-access \
  --with-federation \
  --with-privileged-access

```

6. Création d'un catalogue fédéré dans Data Catalog

Après avoir créé un objet de AWS Glue connexion et l'avoir enregistré auprès de Lake Formation, vous pouvez créer un catalogue fédéré dans le catalogue de données :

Fournissez au catalogue fédéré un nom unique à *your-federated-catalog-name*, référencez le catalogue dans Snowflake at *catalog-name-in-Snowflake* et un nom de connexion d'entrée créé précédemment à *your-glue-connection-name*.

```

aws glue create-catalog \
  --name your-federated-catalog-name \

```

```
--catalog-input '{
  "FederatedCatalog": {
    "Identifier": catalog-name-in-Snowflake,
    "ConnectionName": your-glue-connection-name
  },
  "CreateTableDefaultPermissions": [],
  "CreateDatabaseDefaultPermissions": []
}'
```

Considérations relatives à l'intégration avec Snowflake

- Lorsque vous déposez des ressources (telles que des bases de données et des tables) Snowflake, Lake Formation ne révoque pas automatiquement les autorisations accordées sur cette ressource fédérée. Pour supprimer les autorisations d'accès, vous devez révoquer explicitement les autorisations précédemment accordées sur la ressource fédérée à l'aide de Lake Formation.
- Lorsque vous montez un catalogue distant avec `CATALOG_CASING_FILTER= 'UPPERCASE_ONLY'` configuration, les bases de données et les tables avec des identifiants en majuscules sont fédérées, mais pas les objets avec des identifiants en minuscules.
- Vous pouvez réutiliser la même AWS Glue connexion pour créer plusieurs catalogues fédérés. La suppression d'un catalogue n'entraîne pas la suppression de la connexion associée. Pour supprimer une connexion, utilisez la `aws glue delete-connection` commande AWS CLI et assurez-vous que tous les catalogues associés sont d'abord supprimés.
- Les espaces de noms imbriqués dans le Polaris catalogue ne sont pas pris en charge. En d'autres termes, la fédération de catalogues peut accéder aux Iceberg tables distantes qui suivent une notation `catalog.database.table` en 3 parties.

Fédérer pour Databricks Unity Catalog

AWS Glue Data Catalog se connecte à Databricks l'utilisation des OAuth2 informations d'identification d'un directeur de Databricks service. Ce mécanisme d'authentification permet d'accéder AWS Glue Data Catalog aux métadonnées de divers objets (tels que des catalogues, des bases de données et des tables) en Databricks Unity Catalog fonction des privilèges associés au principal de service. Pour garantir l'accès aux bons objets, il est essentiel d'accorder au principal du service les autorisations nécessaires Databricks pour lire les métadonnées de ces objets.

Ensuite, la fédération de catalogues permet de découvrir et d'interroger Iceberg les tables de votre catalogue Databricks Unity. Pour lire les tables delta, assurez-vous que Iceberg les métadonnées sont disponibles pour ces tables à l'aide de Uniform. Suivez le Databricks didacticiel et la documentation pour créer le principal de service et les privilèges associés dans votre Databricks espace de travail.

Conditions préalables

Avant de créer un catalogue fédéré dans Data Catalog régi par Lake Formation, assurez-vous de disposer des autorisations suivantes :

Votre principal IAM (utilisateur ou rôle) doit disposer des autorisations suivantes :

- Permissions de Lake Formation —`lakeformation:RegisterResource`,
`lakeformation:DescribeResource`
- AWS Glue autorisations —
`glue:CreateConnection`,`glue:CreateCatalog`,`glue:GetConnection`,
`glue:PassConnection`
- Autorisations du Gestionnaire de Secrets —`secretsmanager:CreateSecret`,
`secretsmanager:GetSecretValue`
- Autorisations IAM —`iam:CreateRole`, `iam:AttachRolePolicy` `iam:PassRole`

Vous devez être un administrateur du lac de Lake Formation ou avoir l'`CREATE_CATALOG` autorisation d'accéder au catalogue de données

Création d'un catalogue fédéré

Utilisation de la console

1. Connectez-vous à la console et ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Choisissez la AWS région préférée dans la section supérieure droite de la page.
3. Dans le volet de navigation de gauche, choisissez Catalogues.
4. Choisissez Créer un catalogue pour ouvrir le flux de travail de création de catalogue.
5. Dans l'étape Choisir une source de données, Databricks sélectionnez l'une des options disponibles.

6. À l'étape Définir les détails du catalogue, vous fournissez trois informations : les détails du catalogue, les détails de connexion et les détails d'enregistrement.
7. Dans le conteneur des détails du catalogue, attribuez un nom unique à votre catalogue AWS Glue fédéré et entrez le nom du Databricks catalogue existant.
8. Dans le conteneur des détails des connexions, vous pouvez soit choisir une connexion existante à laquelle vous avez accès, soit fournir une configuration pour créer un nouveau connecteur.
9. Les nouvelles configurations de connexion incluent :
 - Nom de connexion : nom unique de l'objet de AWS Glue connexion.
 - URL de l'espace de travail : URL du point de terminaison de votre Databricks espace de travail existant.
 - Authentification — Spécifiez la configuration d'authentification AWS Glue utilisée pour se connecter au serveur de catalogue distant. AWS Glue prend en charge les deux OAuth2 et l'authentification personnalisée.
 - URL du jeton : spécifiez l'URL du fournisseur d'identité du catalogue distant.
 - OAuth2 ID client — Spécifiez l'ID client des OAuth2 informations d'identification associées à votre catalogue distant.
 - Secret — Stockez et utilisez OAuth2 le secret du client en utilisant AWS Secrets Manager ou entrez la valeur du secret dans la zone de texte. Lorsque vous entrez le secret manuellement dans la console, AWS Glue crée le secret en votre nom.
 - Étendue de l'URL du jeton : spécifiez l' OAuth étendue de l'authentification.
10. Créez un rôle IAM que les responsables du service Lake Formation AWS Glue et les responsables du service Lake Formation peuvent utiliser pour accéder respectivement aux emplacements secrets et Amazon S3 des Iceberg tables distantes. Sélectionnez le rôle IAM dans le menu déroulant d'inscription. Reportez-vous aux étapes 2 et 3 de la section CLI suivante pour plus de détails sur la politique IAM.
11. Sélectionnez Tester la connexion pour vérifier si les propriétés de votre connexion et l'accès au rôle IAM sont correctement configurés. La fonctionnalité de test de connexion n'est pas disponible lors de la connexion à Databricks l'aide d'Amazon VPC.
12. Sélectionnez Suivant pour vérifier vos paramètres.
13. Sélectionnez Créer un catalogue sur la page de révision.

Utilisation de l'interface de ligne de commande (CLI)

1. Créez un AWS Secrets Manager secret

Le AWS Glue connecteur prend en charge deux types d'authentification : OAuth2 et l'authentification personnalisée. Lorsque vous utilisez l' OAuth2 option, AWS Secrets Manager utilise-la pour stocker le secret client du principal de Databricks service. Vous utiliserez ce secret ultérieurement lors de la création de la AWS Glue connexion. Pour l'authentification personnalisée, utilisez-le AWS Secrets Manager pour stocker et récupérer le jeton d'accès.

Dans l'exemple suivant, remplacez `<databricks-secret>`, `<client_secret>`, `<region>` par vos propres informations.

```
aws secretsmanager create-secret \
--name <databricks-secret> \
--description "Databricks secret" \
--secret-string '{
"USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET": "<client_secret>"
}' \
--region <region>
```

Note

USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET est un mot clé réservé AWS Glue utilisé pour faire référence à une valeur secrète du client dans le secret. Utilisez également le même mot clé lorsque vous créez le secret dans la console Lake Formation.

2. Créez un rôle IAM qui permet à l'objet de AWS Glue connexion d'accéder au secret créé à l'étape précédente

L'objet de AWS Glue connexion a besoin d'accéder au AWS Secrets Manager secret lorsque vous l'utilisez AWS Secrets Manager pour stocker, récupérer et actualiser votre jeton OAuth secret. L'objet de AWS Glue connexion nécessite également un accès pour créer, décrire et utiliser les interfaces réseau Amazon VPC lorsque vous utilisez un point de terminaison Amazon VPC pour restreindre la connectivité à votre espace de travail. Databricks

Créez une politique IAM et associez-la à un rôle IAM. Ajoutez le principal de AWS Glue service à la politique de confiance.

Dans l'exemple suivant, remplacez *<your-secrets-manager-ARN>**<your-vpc-id>*, et *<your-subnet-id1>* par vos propres informations.

Exemple Stratégie IAM

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue",
      "secretsmanager:DescribeSecret",
      "secretsmanager:PutSecretValue"
    ],
    "Resource": [
      "<your-secrets-manager-ARN>"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "ec2:CreateNetworkInterface",
      "ec2:DeleteNetworkInterface",
      "ec2:DescribeNetworkInterfaces"
    ],
    "Resource": "*",
    "Condition": {
      "ArnEquals": {
        "ec2:Vpc": "arn:aws:ec2:region:account-id:vpc/<your-vpc-id>",
        "ec2:Subnet": ["arn:aws:ec2:region:account-id:subnet/<your-
subnet-id1>"]
      }
    }
  }
]
```

Exemple Stratégie d'approbation

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Principal": {
      "Service": "glue.amazonaws.com"
    },
    "Action": "sts:AssumeRole"
  }]
}
```

3. Créez une politique IAM qui donne à Lake Formation un accès en lecture à l'emplacement Amazon S3 du catalogue

En tant que propriétaire d'un catalogue fédéré dans Data Catalog, vous utilisez Lake Formation pour accorder un accès grossier aux tables, un accès fin (au niveau des colonnes, des lignes et au niveau des cellules) et un accès basé sur des balises à vos équipes chargées des données. Lake Formation utilise un rôle IAM qui lui permet d'accéder aux emplacements Amazon S3 sous-jacents de vos tables Iceberg distantes. Cet accès permet à Lake Formation de vendre des identifiants d'accès délimités aux moteurs d'analyse interrogeant des tables distantes.

Créez une politique IAM et attachez-la à un rôle IAM. Ajoutez le principal de service Lake Formation à la politique de confiance des rôles IAM.

Dans l'exemple suivant, remplacez *<your-s3-bucket-N>* et *<your-kms-key>* par vos propres informations.

Exemple Stratégie IAM

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::<your-s3-bucket-1>/*",
      "arn:aws:s3:::<your-s3-bucket-2>/*"
    ]
  }]
}
```

```

    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "s3:ListBucket"
    ],
    "Resource": [
      "arn:aws:s3:::<your-s3-bucket-1>",
      "arn:aws:s3:::<your-s3-bucket-2>"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt",
      "kms:Encrypt"
    ],
    "Resource": [
      "<your-kms-key>"
    ]
  }
]
}

```

Exemple Stratégie d'approbation

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Sid": "",
    "Effect": "Allow",
    "Principal": {
      "Service": "lakeformation.amazonaws.com"
    },
    "Action": "sts:AssumeRole"
  }]
}

```

Note

Lorsque vous utilisez la console Lake Formation pour créer un catalogue fédéré, la console utilise un seul rôle IAM auquel les deux politiques sont associées pour terminer la configuration.

4. Création d'un objet AWS Glue de connexion

Data Catalog prend en charge ConnectionType DATABRICKSICEBERGRESTCATALOG : pour connecter Data Catalog à Databricks. Ce AWS Glue connecteur prend en charge OAuth2 les méthodes d'authentification personnalisées.

L'exemple suivant utilise la configuration OAuth2 d'authentification pour créer une AWS Glue connexion. Remplacez *highlighted sections* par vos informations.

```
aws glue create-connection \
  --connection-input '{
    "Name": "<your-glue-connection-to-databricks-unity-account>",
    "ConnectionType": "DATABRICKSICEBERGRESTCATALOG",
    "ConnectionProperties": {
      "INSTANCE_URL": "<your-databricks-workspace-catalog-URL>",
      "ROLE_ARN": "<your-IAM-role-for-secrets-and-VPC-access>"
    },
    "AuthenticationConfiguration": {
      "AuthenticationType": "OAUTH2",
      "OAuth2Properties": {
        "OAuth2GrantType": "CLIENT_CREDENTIALS",
        "TokenUrl": "<your-internal-or-external-token-server-url>",
        "OAuth2ClientApplication": {
          "UserManagedClientApplicationClientId": "<your-client-id>"
        },
        "TokenUrlParametersMap": {
          "scope": "all-apis"
        }
      },
      "SecretArn": "arn:aws:secretsmanager:<aws-region>:<your-aws-account-id>:secret:<databricks-secret>"
    }
  }'
```

5. Enregistrer AWS Glue la connexion en tant que ressource Lake Formation

À l'aide de l'objet de AWS Glue connexion (créé à l'étape 4) et du rôle IAM (créé à l'étape 3), vous pouvez désormais enregistrer l'objet de AWS Glue connexion en tant que ressource gérée par Lake Formation.

Remplacez *<your-glue-connector-arn>* et *<your-IAM-role-ARN-having-LF-access>* par vos informations.

```
aws lakeformation register-resource \  
  --resource-arn <your-glue-connector-arn> \  
  --role-arn <your-IAM-role-ARN-having-LF-access> \  
  --with-federation \  
  --with-privileged-access
```

6. Création d'un catalogue fédéré dans Data Catalog

Après avoir créé un objet de AWS Glue connexion et l'avoir enregistré auprès de Lake Formation, vous pouvez créer un catalogue fédéré dans le catalogue de données :

Fournissez au catalogue fédéré un nom unique à *<your-federated-catalog-name>*, référencez le catalogue dans Databricks à *"<catalog-name-in-Databricks>"* et un nom de connexion d'entrée créé précédemment à *"<your-glue-connection-name>"*.

```
aws glue create-catalog \  
  --name <your-federated-catalog-name> \  
  --catalog-input '{  
    "FederatedCatalog": {  
      "Identifier": "<catalog-name-in-Databricks>",  
      "ConnectionName": "<your-glue-connection-name>"  
    },  
    "CreateTableDefaultPermissions": [],  
    "CreateDatabaseDefaultPermissions": []  
  }'
```

Considérations relatives à l'intégration avec Databricks

- Lorsque vous déposez des ressources (telles que des bases de données et des tables) dans le Databricks, Lake Formation ne révoque pas automatiquement les autorisations accordées sur cette ressource fédérée. Pour supprimer les autorisations d'accès, vous devez révoquer

explicitement les autorisations précédemment accordées sur la ressource fédérée à l'aide de Lake Formation.

- Vous pouvez interroger les tables Iceberg stockées dans Amazon S3 à l'aide de cette intégration. Lorsque vous utilisez un autre format de table ou un autre stockage d'objets, vous pouvez fédérer des métadonnées dans des catalogues distants AWS Glue et répertorier ses bases de données et ses tables, mais les opérations de requête de ce type `SELECT ColumnFoo from TableBar` échoueront lors de la requête avec l'erreur « Impossible de lire la table ». Apache Iceberg L'emplacement de stockage des objets n'est pas pris en charge. '
- Vous pouvez réutiliser la même AWS Glue connexion pour créer plusieurs catalogues fédérés. La suppression d'un catalogue n'entraîne pas la suppression de l'objet de connexion associé. Pour supprimer un objet de connexion, utilisez la `aws glue delete-connection` commande AWS CLI et assurez-vous que tous les catalogues associés sont d'abord supprimés.

Intégrer les données Amazon Redshift dans AWS Glue Data Catalog

Vous pouvez gérer les données analytiques dans les entrepôts de données Amazon Redshift dans AWS Glue Data Catalog le (catalogue de données) et unifier les lacs de données Amazon S3 et les entrepôts de données Amazon Redshift. Amazon Redshift est un service d'entrepôt de données entièrement géré de plusieurs pétaoctets dans le cloud. AWS Un entrepôt des données Amazon Redshift est un ensemble de ressources informatiques appelées nœuds, qui sont organisées en un groupe appelé cluster. Chaque cluster exécute un moteur Amazon Redshift et contient une ou plusieurs bases de données.

Dans Amazon Redshift, vous pouvez créer des clusters provisionnés par Amazon Redshift et des espaces de noms sans serveur, et les enregistrer dans le catalogue de données. Vous pouvez ainsi unifier les données dans le stockage géré Amazon Redshift (RMS) et les compartiments Amazon S3, et accéder aux données à partir de moteurs d'analyse compatibles avec Apache Iceberg.

En enregistrant des espaces de noms et des clusters, vous pouvez donner accès aux données sans avoir à les copier ou à les déplacer. Pour plus d'informations sur l'enregistrement de clusters et d'espaces de noms dans Amazon Redshift, [consultez la section Enregistrement de clusters et d'espaces de noms Amazon Redshift](#) auprès du. AWS Glue Data Catalog

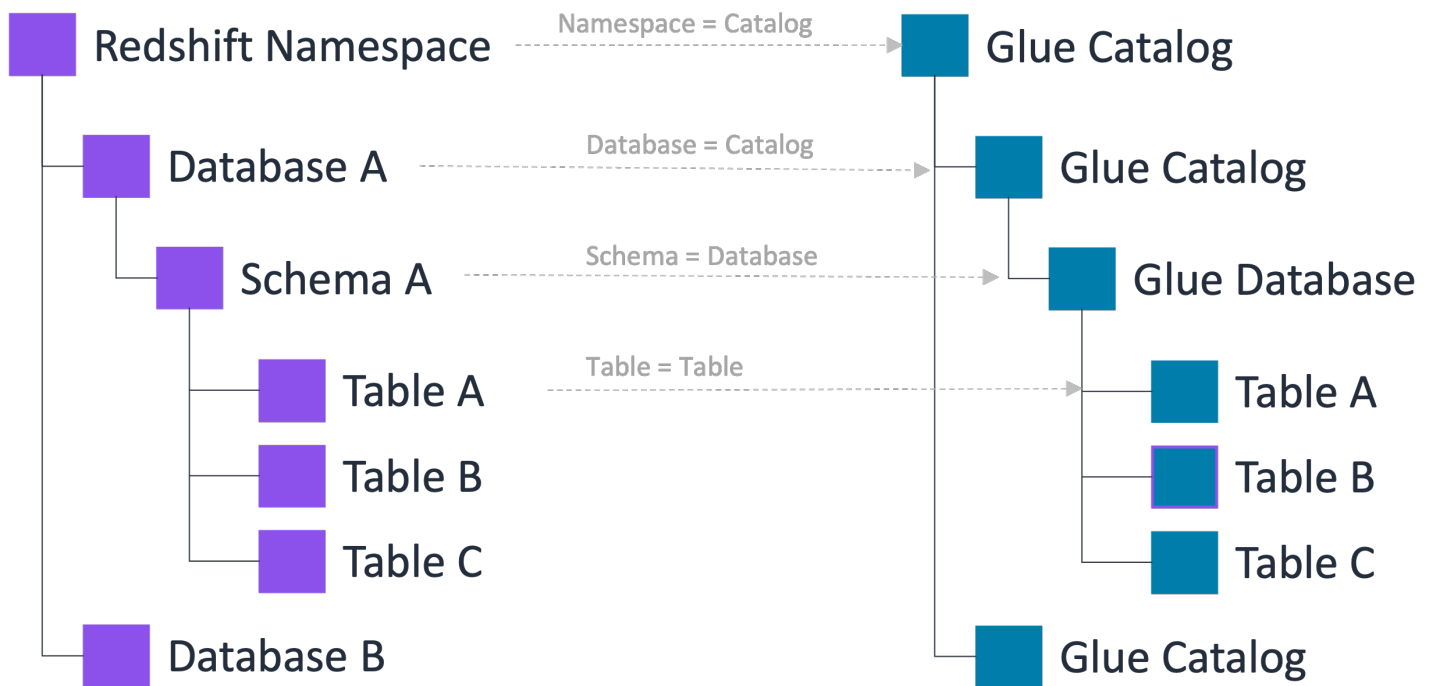
Dans Amazon Redshift, vous pouvez partager des données par le biais de partages de données ou en enregistrant des espaces de noms et des clusters auprès de Data Catalog. Dans le cas

des partages de données, qui fonctionnent au niveau de chaque objet de base de données, vous devez activer le partage pour chaque table ou vue. En revanche, la publication d'espaces de noms fonctionne au niveau du cluster ou de l'espace de noms. Lorsque vous enregistrez un cluster ou un espace de noms dans le catalogue de données, toutes les bases de données et les tables qu'il contient sont automatiquement partagées, sans que vous ayez à configurer le partage pour des objets individuels.

Dans le catalogue de données, vous pouvez créer un catalogue fédéré pour chaque espace de noms ou cluster. Un catalogue est appelé catalogue fédéré lorsqu'il pointe vers une entité extérieure au catalogue de données. Les tables et les vues de l'espace de noms Amazon Redshift sont répertoriées sous forme de tables individuelles dans le catalogue de données. Vous pouvez partager des bases de données et des tables du catalogue fédéré avec des principaux IAM et des utilisateurs SAML sélectionnés au sein du même compte ou d'un autre compte avec Lake Formation. Vous pouvez également inclure des expressions de filtre de ligne et de colonne pour restreindre l'accès à certaines données. Pour de plus amples informations, veuillez consulter [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#).

Le catalogue de données prend en charge une hiérarchie de métadonnées à trois niveaux comprenant des catalogues, des bases de données et des tables (et des vues). Lorsque vous enregistrez un espace de noms dans le catalogue de données, la hiérarchie de données Amazon Redshift est mappée à la hiérarchie à 3 niveaux du catalogue de données comme suit :

- L'espace de noms Amazon Redshift devient un catalogue à plusieurs niveaux dans le catalogue de données.
- La base de données Amazon Redshift associée est enregistrée en tant que catalogue dans le catalogue de données.
- Le schéma Amazon Redshift devient une base de données dans le catalogue de données.
- La table Amazon Redshift devient une table dans le catalogue de données.



Grâce à cette hiérarchie de métadonnées à trois niveaux, vous pouvez accéder aux tables Amazon Redshift en utilisant la notation en trois parties « catalog1/catalog2.database.table » dans le catalogue de données. Les équipes chargées des données peuvent également conserver la même organisation qu'Amazon Redshift utilise pour organiser les tables au sein du compte Data Catalog.

Dans Lake Formation, vous pouvez gérer en toute sécurité les données d'Amazon Redshift à l'aide d'un contrôle d'accès précis pour les ressources du catalogue de données. Grâce à cette intégration, vous pouvez gérer, sécuriser et interroger des données analytiques à partir d'un catalogue unique doté d'un mécanisme de contrôle d'accès commun.

Pour connaître les limitations, veuillez consulter [Limites liées à l'intégration des données de l'entrepôt de données Amazon Redshift dans le AWS Glue Data Catalog](#).

Rubriques

- [Principaux avantages](#)
- [Rôles et responsabilités](#)
- [Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog](#)
- [Création de catalogues fédérés Amazon Redshift](#)
- [Affichage des objets du catalogue](#)

- [Mettre à jour un catalogue fédéré](#)
- [Accès à un catalogue fédéré partagé](#)
- [Supprimer un catalogue fédéré](#)
- [Interrogation de catalogues fédérés](#)
- [Ressources supplémentaires](#)

Principaux avantages

L'enregistrement de clusters et d'espaces de noms Amazon Redshift dans les AWS Glue Data Catalog lacs de données Amazon S3 et les entrepôts de données Amazon Redshift et leur unification offrent les avantages suivants :

- Expérience d'interrogation uniforme : interrogez vos données gérées par Amazon Redshift et les données contenues dans les compartiments Amazon S3 à l'aide de n'importe quel moteur de requête compatible avec Apache Iceberg, tel qu'Amazon EMR Serverless et Amazon Athena, sans avoir à déplacer ou à copier des données.
- Accès aux données cohérent entre les services : vous n'avez pas besoin de mettre à jour les noms des bases de données et des tables dans vos pipelines de données lorsque vous accédez aux mêmes sources de données fédérées à partir de différents services AWS d'analyse, car les sources de données sont enregistrées dans le catalogue de données.
- Contrôle d'accès détaillé — Vous pouvez appliquer des autorisations de Lake Formation pour gérer l'accès aux sources de données fédérées à l'aide d'autorisations de contrôle d'accès détaillées.

Rôles et responsabilités

Rôle	Responsabilité
Administrateur du cluster de producteurs Amazon Redshift	Enregistre le cluster ou l'espace de noms dans le catalogue de données.
Administrateur du lac de données de Lake Formation	Accepte l'invitation du cluster ou de l'espace de noms, crée des catalogues fédérés et accorde l'accès aux catalogues fédérés à d'autres principaux.

Administrateur en lecture seule de Lake Formation	Découvre le catalogue fédéré, interroge les tables Amazon Redshift dans le catalogue fédéré.
Rôle de transfert de données	Amazon Redshift se charge en votre nom de transférer les données vers et depuis le compartiment Amazon S3.

Voici les étapes de haut niveau permettant aux utilisateurs d'accéder à un espace de noms Amazon Redshift :

1. Dans Amazon Redshift, l'administrateur du cluster de producteurs enregistre un cluster ou un espace de noms dans le catalogue de données.
2. L'administrateur du lac de données accepte l'invitation à l'espace de noms de l'administrateur du cluster de producteurs Amazon Redshift et crée un catalogue fédéré dans le catalogue de données.

Une fois cette étape terminée, vous pouvez gérer le catalogue d'espaces de noms Amazon Redshift dans le catalogue de données.

3. Accordez des autorisations aux utilisateurs sur les catalogues, les bases de données et les tables. Vous pouvez partager l'intégralité du catalogue d'espaces de noms ou un sous-ensemble de tables avec des utilisateurs du même compte ou d'un autre compte.

Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog

1. Création d'un administrateur de lac de données : créez un rôle IAM autorisé à accepter l'invitation d'espace de noms, à créer les AWS Glue Data Catalog objets (catalogues, bases de données, tables/vues) et à accorder les autorisations de Lake Formation aux autres utilisateurs.

Pour step-by-step obtenir des instructions sur la création d'un administrateur de lac de données, consultez [Création d'un administrateur de lac de données](#).

2. Mettez à jour les autorisations d'administrateur du lac de données.

Outre les autorisations d'administrateur du lac de données, l'administrateur du lac de données a besoin des autorisations suivantes pour accepter une invitation à un espace de noms Amazon

Redshift dans Lake Formation, créer ou mettre à jour les ressources du catalogue de données et autoriser l'accès au lac de données :

JSON

```
{
  "Version": "2012-10-17",
  "Id": "glue-enable-datalake-access",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "redshift:AssociateDataShareConsumer",
      "redshift:DescribeDataSharesForConsumer",
      "redshift:DescribeDataShares",
      "redshift-serverless:CreateNamespace",
      "redshift-serverless:CreateWorkgroup",
      "redshift-serverless>DeleteNamespace",
      "redshift-serverless>DeleteWorkgroup",
      "ec2:DescribeAccountAttributes",
      "ec2:DescribeSubnets",
      "ec2:DescribeSecurityGroups",
      "ec2:DescribeAvailabilityZones",
      "s3:createBucket",
      "s3:deleteBucket",
      "s3:putBucketPolicy",
      "s3:putEncryptionConfiguration",
      "s3:putLifecycleConfiguration",
      "s3:putBucketVersioning",
      "iam:CreateRole"
    ],
    "Resource": "*"
  },
  {
    "Action": [
      "iam:PassRole"
    ],
    "Effect": "Allow",
    "Resource": "arn:aws:iam::*:role/data transfer role name",
    "Condition": {
      "StringLike": {
        "iam:PassedToService": [
          "glue.amazonaws.com"
        ]
      }
    }
  }
]
```

```

    }
  }
}

```

3. Si le rôle IAM utilisé pour créer des catalogues fédérés n'est pas un administrateur de data lake, vous devez lui accorder l'autorisation. `Create catalog`

Pour créer des créateurs de catalogues

- a. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
- b. Choisissez Rôles et tâches administratifs sous Administration.
- c. Choisissez Accorder.
- d. Sur l'écran Accorder des autorisations, choisissez un utilisateur ou un rôle IAM.
- e. Sélectionnez Créer une autorisation de catalogue.
- f. Facultativement, vous pouvez également accorder une autorisation de création de catalogue pouvant être accordée. L'autorisation pouvant être accordée permet au créateur du catalogue d'accorder l'`Create catalog` autorisation à d'autres personnes principales.
- g. Choisissez Accorder.

AWS CLI exemple d'octroi d'autorisations pour créer un catalogue fédéré.

```

aws lakeformation grant-permissions \
--cli-input-json \
'{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/Admin"
  },
  "Resource": {
    "Catalog": {
    }
  },
  "Permissions": [
    "CREATE_CATALOG",
    "DESCRIBE"
  ]
}'

```

4. Créez un rôle d'administrateur en lecture seule pour découvrir les catalogues fédérés Amazon Redshift dans le catalogue de données d'Amazon Redshift Query Editor v2.

Pour interroger les tables Amazon Redshift dans le catalogue fédéré à partir d'Amazon Redshift Query Editor v2, assurez-vous que la politique de rôle d'administrateur en lecture seule contient l'ARN du rôle lié au service Amazon Redshift. `AWSServiceRoleForRedshift`

```
aws lakeformation put-data-lake-settings
    --region us-east-1 \
    --data-lake-settings \
    '{
      "DataLakeAdmins": [{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/Admin"}],
      "ReadOnlyAdmins":[{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/aws-service-role/redshift.amazonaws.com/AWSServiceRoleForRedshift"}],
      "CreateDatabaseDefaultPermissions":[],
      "CreateTableDefaultPermissions":[],
      "Parameters":{"CROSS_ACCOUNT_VERSION":"4","SET_CONTEXT":"TRUE"}
    }'
```

5. Créez un rôle de transfert de données qu'Amazon Redshift peut assumer en votre nom pour transférer des données vers et depuis le compartiment Amazon S3.

Lorsque vous activez l'accès aux lacs de données pour les moteurs de requêtes compatibles Apache Iceberg tels qu'Athena, Amazon EMR ou Amazon EC2 pour accéder aux ressources Amazon Redshift du catalogue de données, vous devez créer un rôle IAM doté des autorisations requises pour effectuer le transfert de données vers et depuis le compartiment Amazon S3.

JSON

```
{
  "Version":"2012-10-17",
  "Id": "glue-enable-datalake-access",
  "Statement": [{
    "Sid": "DataTransferRolePolicy",
    "Effect": "Allow",
    "Action": [ "glue:GetCatalog",
                "glue:GetDatabase",
                "kms:GenerateDataKey",
                "kms:Decrypt"],
    "Resource": "*"
  }]
```

```

    }
  ]
}

```

6. Ajoutez la politique de confiance suivante au rôle de transfert de données pour que AWS Glue les services Amazon Redshift assument le rôle de transfert de données vers et depuis le compartiment Amazon S3.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "redshift.amazonaws.com",
          "glue.amazonaws.com"
        ]
      },
      "Action": "sts:AssumeRole"
    }
  ]
}

```

7. Ajoutez la politique de clé suivante à la AWS KMS clé si vous utilisez une clé gérée par le client pour chiffrer les données dans le cluster/espace de noms Amazon Redshift. Remplacez le numéro de compte par un numéro de AWS compte valide et spécifiez le nom du rôle de transfert de données. Par défaut, les données du cluster Amazon Redshift sont chiffrées à l'aide d'une clé KMS. Lake Formation propose une option permettant de créer votre clé KMS personnalisée pour le chiffrement. Si vous utilisez une clé gérée par le client, vous devez ajouter des politiques clés spécifiques à la clé.

Pour plus d'informations sur la gestion des autorisations d'une clé gérée par le client, consultez [Clés gérées par le client](#).

JSON

```

{
  "Version": "2012-10-17",

```

```

    "Id": "auto-redshift-3",
    "Statement": [{
      "Sid": "RedshiftAllowAccessPolicy",
      "Effect": "Allow",
      "Principal": {
        "AWS": "*"
      },
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:CreateGrant",
        "kms:DescribeKey"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "kms:CallerAccount": "111122223333",
          "kms:ViaService": "redshift.us-east-1.amazonaws.com"
        }
      }
    },
    {
      "Sid": "RedshiftServerlessAllowAccessPolicy",
      "Effect": "Allow",
      "Principal": {
        "AWS": "*"
      },
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:CreateGrant",
        "kms:DescribeKey"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "kms:CallerAccount": "111122223333",
          "kms:ViaService": "redshift-serverless.us-
east-1.amazonaws.com"
        }
      }
    }
  ]
}

```



```

    }
  },
  {
    "Sid": "DirectMetadataAccess",
    "Effect": "Allow",
    "Principal": {
      "AWS": "arn:aws:iam::111122223333:root"
    },
    "Action": [
      "kms:Describe*",
      "kms:Get*",
      "kms:List*",
      "kms:RevokeGrant"
    ],
    "Resource": "*"
  },
  {
    "Sid": "GenerateDataKeyDecryptDataTransferRole",
    "Effect": "Allow",
    "Principal": {
      "AWS": "arn:aws:iam::111122223333:role/data-transfer-role-name"
    },
    "Action": [
      "kms:GenerateDataKey",
      "kms:Decrypt"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "kms:ViaService": "s3.us-east-1.amazonaws.com"
      }
    }
  }
]
}

```

Création de catalogues fédérés Amazon Redshift

Cette rubrique décrit les étapes à suivre pour accepter une invitation de cluster ou d'espace de noms, créer un catalogue fédéré à plusieurs niveaux et accorder des autorisations à d'autres principaux.

Vous pouvez effectuer ces tâches à l'aide de la console Lake Formation, du AWS Command Line Interface (AWS CLI) ou du APIs/SDKs. Les exemples présentés dans cette rubrique présentent le cluster/espace de noms du producteur, le catalogue de données et le consommateur de données dans le même compte.

Pour en savoir plus sur les fonctionnalités multicomptes de Lake Formation, voir [Partage de données entre comptes dans Lake Formation](#).

Pour gérer un espace de noms Amazon Redshift dans le catalogue de données

1. Passez en revue une invitation à un espace de noms et acceptez-la.

Console

1. Connectez-vous à la console Lake Formation en tant qu'administrateur de lac de données à l'adresse <https://console.aws.amazon.com/lakeformation/>. Accédez à la page Catalogues sous Catalogue de données.
2. Passez en revue l'invitation à l'espace de noms à laquelle vous êtes autorisé à accéder. La colonne Status indique votre statut de participation actuel pour l'espace de noms. Le statut Non accepté indique que vous avez été ajouté à l'espace de noms, mais que vous ne l'avez pas encore accepté ou que vous avez rejeté l'invitation.

Catalogs

▼ How it works

Create a catalog
Register Redshift databases as catalogs in the Data Catalog. [Learn more](#)

Manage catalog permissions

Manage permissions for specific catalogs, databases, tables and fine-grained data access. [Learn more](#)

Access from query editors

Access catalog objects from [Redshift Query Editor v2](#) and [Athena Console](#).

Create a federated catalog for your S3 Table Buckets.

Enable S3 Table integration

Pending catalog invitations (4)

View and manage Redshift namespace/cluster invitations in the AWS Glue Data Catalog.

Find invitations

< 1 >

Name	Source account ID	Received	Status
arn:aws:redshift-serverless:us-east-2:451785580005:namespace/c0381d75-3f21-49f2-b2c3-44d000803a71		November 20, 2024 at 10:16 PM UTC	Accepted, catalog not created
arn:aws:redshift-serverless:us-east-2:451785580005:namespace/4a798b4c-71d8-4df4-b77f-52cff8ac80a1		November 20, 2024 at 5:38 PM UTC	Accepted, catalog not created
arn:aws:redshift:us-east-2:451785580005:namespace:48a491a6-d5d8-415b-b5b2-3832a4affb08		November 26, 2024 at 3:45 PM UTC	Accepted, catalog not created
arn:aws:redshift:us-east-2:451785580005:namespace:a77f139c-5a19-4b53-a662-c2f50db2fc28		December 3, 2024 at 2:21 PM UTC	Accepted, catalog not created

Catalogs (1)

A catalog is the top level in the Data Catalog's three-level data hierarchy and contains Data Catalog objects.

Find catalogs by name

< 1 >

Name	Type	Source	Owner account...	Shared resource	Shared resour...	Shared resource owner region
	Default	Default catalog	-	-	-	

3. Pour répondre à une invitation à un espace de noms ou à un cluster, sélectionnez le nom de l'invitation, puis sélectionnez Vérifier l'invitation. Dans Accepter ou rejeter l'invitation,

prenez en compte les détails de l'invitation. Choisissez Accepter pour accepter l'invitation ou Refuser pour refuser l'invitation. Vous n'avez pas accès à l'espace de noms si vous rejetez l'invitation.

AWS CLI

Les exemples suivants montrent comment afficher, accepter et enregistrer l'invitation. Remplacez l'identifiant de compte AWS par un identifiant de compte AWS valide. Remplacez le `data-share-arn` par le véritable Amazon Resource Name (ARN) qui fait référence à l'espace de noms.

1. Afficher une invitation en attente.

```
aws redshift describe-data-shares \
  --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
```

2. Acceptez une invitation.

```
aws redshift associate-data-share-consumer \
  --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
  --consumer-arn 'arn:aws:glue:us-east-1:123456789012:catalog'
```

3. Enregistrez le cluster ou l'espace de noms dans le compte Lake Formation. Utilisez l'opération [RegisterResourceAPI](#) pour enregistrer le partage de données dans Lake Formation. `DataShareArn` est le paramètre d'entrée pour `ResourceArn`.

Note

Il s'agit d'une étape obligatoire.

```
aws lakeformation register-resource \
  --resource-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace'
```

2. Créez un catalogue fédéré.

Après avoir accepté une invitation, vous devez créer un catalogue fédéré dans le catalogue de données qui met en correspondance les objets de l'espace de noms Amazon Redshift avec le catalogue de données. Vous devez être un administrateur de lac de données, un utilisateur ou un rôle disposant des autorisations requises pour créer un catalogue.

Console

1. Après avoir accepté l'invitation à l'espace de noms, la page Définir les détails du catalogue s'affiche.
2. Sur la page Définir les détails du catalogue, entrez un nom unique pour le catalogue. Utilisez des minuscules pour les noms de catalogue. Les noms de catalogue doivent comporter une longueur inférieure ou égale à 255 caractères. Vous utilisez cet identifiant pour mapper l'espace de noms en interne dans la hiérarchie des métadonnées (Catalogid.DBName.Schema.Table).
3. Entrez une description pour le catalogue. La description doit comporter une longueur inférieure ou égale à 2 048 caractères.
4. Cochez ensuite la case Accéder à ce catalogue depuis des moteurs compatibles Iceberg pour permettre l'accès aux ressources Amazon Redshift à l'aide de moteurs d'analyse compatibles Apache Iceberg tels qu'Athena et Apache Spark sur Amazon EMR.

Il n'est pas nécessaire d'activer l'accès au lac de données pour accéder aux catalogues fédérés à l'aide d'Amazon Redshift.

Catalog details

A catalog is the top level in the Data Catalog's three-level data hierarchy and contains Data Catalog objects.

Name

Catalog name is required, in lowercase characters, and no longer than 255 characters.

Type

Federated

Source

Redshift

Description - optional

Descriptions can be up to 2048 characters long.

Access from engines

You can access this catalog from open source engines as well as Amazon Redshift.

☒ Access this catalog from Iceberg compatible engines.

Choose this option to access the data catalog using with Apache Spark running on an EMR cluster.

IAM role

Role used by Redshift for loading data to and from S3 bucket that is created for the managed workgroup.

[View](#)[Create an IAM role](#)

Encryption options

Your data is encrypted by default with a key that AWS owns and manages for you. To choose a different key, customize your encryption settings.

☐ Customize encryption settings

To use the default key, clear this option.

[Cancel](#)[Skip to Review and create](#)[Next](#)

5. Pour permettre à ces moteurs de requêtes de lire et d'écrire dans les espaces de noms Amazon Redshift, AWS Glue crée un cluster Amazon Redshift géré avec les ressources de calcul et de stockage nécessaires pour effectuer des opérations de lecture et d'écriture sans impact sur les charges de travail de l'entrepôt de données Amazon Redshift.

Vous devez également fournir un rôle IAM avec les autorisations requises pour transférer des données vers et depuis le compartiment Amazon S3.

6. Par défaut, les données du cluster Amazon Redshift sont chiffrées à l'aide d'une clé AWS gérée. Lake Formation propose une option permettant de créer votre clé KMS personnalisée pour le chiffrement. Si vous utilisez une clé gérée par le client, vous devez ajouter des politiques clés spécifiques à la clé.

Choisissez l'option Personnaliser les paramètres de chiffrement si vous utilisez une clé gérée par le client pour chiffrer les données du cluster/espace de noms Amazon Redshift. Pour utiliser une clé personnalisée, vous devez ajouter une stratégie de clé gérée personnalisée supplémentaire à votre clé KMS. Pour de plus amples informations, veuillez consulter [Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog](#).

AWS CLI

Utilisez l'exemple de code suivant pour créer un catalogue avec les données Amazon Redshift publiées dans le catalogue de données à l'aide du AWS CLI

```
aws glue create-catalog
--cli-input-json \
'{
  "Name": "nscatalog",
  "CatalogInput": {
    "Description": "Redshift federated catalog",
    "CreateDatabaseDefaultPermissions" : [],
    "CreateTableDefaultPermissions": [],
    "FederatedCatalog": {
      "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
      "ConnectionName": "aws:redshift"
    },
    "CatalogProperties": {
      "DataLakeAccessProperties" : {
        "DataLakeAccess" : true,
        "DataTransferRole" :
"arn:aws:iam::123456789012:role/DataTransferRole"
      }
    }
  }
}'
```

3. Accordez des autorisations aux utilisateurs de votre compte ou de comptes externes.

AWS Management Console

1. Choisissez Next pour accorder des autorisations à d'autres utilisateurs sur les catalogues, bases de données et tables partagés.
2. Sur l'écran Ajouter des autorisations, choisissez les principes et les types d'autorisations à accorder.

Add permissions

Principals

Choose the principals to grant permissions.

☒ **IAM users and roles**
Users or roles from this AWS account.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☐ **External accounts**
AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

UserRole X
Role

Catalog permissions

Choose the permissions to grant on the catalog. Choosing Super user overwrites individual permissions, granting unrestricted administrative access.

☐ **Super user**
A super user has unrestricted administrative privileges to perform any operation on all resources within the catalog (databases, tables, and views).

Catalog permissions

Choose specific access permissions to grant.

☒ **Create database**
☐ **Describe**
☐ **Alter**

☐ **Drop**

☐ **Super**
This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions

Choose the permission that can be granted to others.

☒ **Create database**
☐ **Describe**
☐ **Alter**

☐ **Drop**

☐ **Super**
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel

Add

- a. Dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels octroyer des autorisations.
 - Utilisateurs et rôles IAM : choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

- Utilisateurs et groupes SAML : pour le SAML et Amazon Quick les utilisateurs et groupes, entrez un ou plusieurs Amazon Resource Names (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou pour les utilisateurs ou groupes ARNs Amazon Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de créer les ARNs, voir les AWS CLI commandes AWS CLI d'octroi et de révocation.

- Comptes externes : pour AWS, AWS organisation ou directeur IAM, entrez un ou plusieurs AWS comptes, organisations IDs IDs, unités organisationnelles ou ARN valides pour l'utilisateur ou le rôle IAM. IDs Appuyez sur Entrée après chaque identifiant. Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres. L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.
- b. Dans la section Autorisations, sélectionnez les autorisations et les autorisations octroyables.

Sous Autorisations du catalogue, sélectionnez une ou plusieurs autorisations à octroyer. Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS Cette option n'est pas prise en charge lorsque vous octroyez des autorisations à un principal IAM à partir d'un compte externe.

Choisissez Super user pour accorder à l'utilisateur des autorisations illimitées sur les ressources (bases de données, tables, vues) du catalogue.

3. Choisissez Ajouter.

AWS CLI

Utilisez les exemples suivants pour accorder des autorisations de catalogue, de base de données et de table à l'aide de AWS CLI :

- L'exemple suivant montre comment accorder des autorisations sur le catalogue fédéré.

```
aws lakeformation grant-permissions
--cli-input-cli-json \
```

```
{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/
non-admin"
  },
  "Resource": {
    "Catalog": {
      "Id": "123456789012:nscatalog"
    }
  },
  "Permissions": [
    "DESCRIBE", "CREATE_CATALOG"
  ],
  "PermissionsWithGrantOption": [
  ]
}
```

- Utilisez l'exemple suivant pour accorder des autorisations sur une base de données.

```
aws lakeformation grant-permissions \
  --cli-input-json \
  '{
    "Principal": {

      "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
    },
    "Resource": {
      "Database": {
        "CatalogId": "123456789012:nscatalog/dev",
        "Name": "public"
      }
    },
    "Permissions": [
      "ALL"
    ]
  }'
```

- L'exemple suivant montre comment accorder des autorisations sur une table de la base de données Amazon Redshift.

```
aws lakeformation grant-permissions \
  --cli-input-json \
  '{
```

```
"Principal": {  
  "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"  
},  
  "Resource": {  
    "Table": {  
      "CatalogId": "123456789012:nscatalog2/dev",  
      "DatabaseName": "public",  
      "TableWildcard" : {}  
    }  
  },  
  "Permissions": [  
    "ALL"  
  ]  
}'
```

4. Choisissez Next pour consulter les détails du catalogue et créer un catalogue fédéré. Le catalogue fédéré nouvellement créé et les objets du catalogue apparaissent sur la page Catalogues.

Un catalogue fédéré Amazon Redshift est référencé avec. catalogID =
123456789012:Redshift-federated catalog id

Affichage des objets du catalogue

Après avoir créé le catalogue fédéré, vous pouvez afficher les objets du catalogue à l'aide de la console Lake Formation ou AWS CLI.

AWS Management Console

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Choisissez Catalogues sous Catalogue de données.
3. Choisissez un catalogue fédéré dans la liste de la page Catalogues.
4. La page de résumé du catalogue indique les objets du catalogue (bases de données et tables) pour lesquels vous avez des autorisations. L'onglet Autorisations indique les principaux IAM auxquels des autorisations ont été accordées sur ces objets.

AWS CLI

- L' AWS CLI exemple suivant montre comment demander le catalogue de niveau supérieur.

```
aws glue get-catalog \  
--catalog-id 123456789012:nscatalog
```

Réponse

```
{  
  "Catalog": {  
    "CatalogId": "123456789012:nscatalog",  
    "Name": "nscatalog",  
    "ResourceArn": "arn:aws:glue:us-east-1:123456789012:catalog/nscatalog",  
    "Description": "Redshift published Catalog",  
    "CreateTime": "2024-09-05T14:49:16-07:00",  
    "FederatedCatalog": {  
      "Identifier": "arn:aws:redshift:us-east-1:123456789012:datashare:b1234589-e823-4a14-ad8e-077085540a50/ds_internal_namespace",  
      "ConnectionName": "aws:redshift"  
    },  
    "CatalogProperties": {  
      "DataLakeAccessProperties": {  
        "DataLakeAccess": true,  
        "DataTransferRole": "arn:aws:iam::123456789012:role/DataTransferRole",  
        "KmsKey": "AWS_OWNED_KMS_KEY",  
        "ManagedWorkgroupName": "123456789012:nscatalog",  
        "ManagedWorkgroupStatus": "AVAILABLE",  
        "RedshiftDatabaseName": "dev"  
      },  
      "CatalogIdentifier": "e2309c2c2fb048f1a3069dfdc1c7883e",  
      "CreateTableDefaultPermissions": [],  
      "CreateDatabaseDefaultPermissions": []  
    }  
  }  
}
```

- L'exemple suivant montre comment demander tous les catalogues du compte.

```
aws glue get-catalogs \  

```

```
--recursive
```

- L'exemple de demande suivant montre comment obtenir un catalogue Amazon Redshift au niveau de la base de données.

```
aws glue get-catalog \  
--catalog-id 123456789012:namespace catalog name/redshift database name
```

- L'exemple de demande suivant montre comment obtenir les bases de données dans le catalogue au niveau de la base de données Amazon Redshift.

```
aws glue get-databases \  
--catalog-id 123456789012:namespace catalog name/redshift database name
```

- L'exemple de demande suivant montre comment obtenir une table Amazon Redshift dans le catalogue.

```
aws glue get-table \  
--catalog-id 123456789012:parent catalog name/redshift database \  
--database-name redshift schema name \  
--name table name
```

- L'exemple suivant montre comment obtenir toutes les tables de la base de données Amazon Redshift.

```
aws glue get-tables \  
--catalog-id 123456789012:namespace catalog name/redshift database name \  
--database-name RS schema name
```

Mettre à jour un catalogue fédéré

Vous pouvez mettre à jour un catalogue fédéré Amazon Redshift dans le catalogue de données à l'aide de la console Lake Formation AWS CLI ou de l'opération API. [UpdateCatalog](#)

AWS Management Console

Suivez ces étapes pour mettre à jour votre catalogue fédéré à l'aide de la console Lake Formation.

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, sélectionnez Catalogues sous Catalogue de données.
3. Sur la page Catalogues, choisissez le catalogue fédéré Amazon Redshift que vous souhaitez mettre à jour.
4. Sous Actions, sélectionnez Modifier.
5. Sur l'écran Définir les détails du catalogue, dans la section Accès depuis les moteurs, sélectionnez Accéder à ce catalogue depuis les moteurs compatibles Iceberg. Si vous cochez cette option, l'accès au lac de données sera activé pour les moteurs de requêtes compatibles avec Apache Iceberg.
6. Créez ensuite un nouveau rôle IAM ou choisissez un rôle IAM existant dont la politique accorde des autorisations pour effectuer le transfert de données vers et depuis le compartiment Amazon S3.

Pour plus d'informations sur les autorisations, consultez [Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog](#).

7. Par défaut, vos données dans le cluster Amazon Redshift sont chiffrées à l'aide d'une clé gérée par AWS. Si vous choisissez de chiffrer les données à l'aide d'une clé gérée par le client, créez une clé KMS ou choisissez une clé existante dotée des autorisations définies dans la [Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog](#) section.
8. Choisissez Enregistrer.

Une fois l'opération terminée, la page des détails du catalogue affiche le nom du groupe de travail géré avec le statut « Succès ».

AWS CLI

Voici un exemple d'entrée `update-catalog` CLI avec l'accès au lac de données désactivé en définissant la valeur du `DataLakeAccess` paramètre comme `false`.

```
aws glue update-catalog --cli-input-json \  
{  
  "Name": "nscatalog",  
  "CatalogInput": {  
    "Description": "Redshift published catalog",
```

```

    "CreateDatabaseDefaultPermissions" : [],
    "CreateTableDefaultPermissions": [],
    "FederatedCatalog": {
        "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
        "ConnectionName": "aws:redshift"
    },
    "CatalogProperties": {
        "DataLakeAccessProperties" : {
            "DataLakeAccess" : false
        }
    }
}
}'

```

Accès à un catalogue fédéré partagé

AWS Lake Formation les fonctionnalités inter-comptes permettent aux utilisateurs de partager en toute sécurité des lacs de données distribués entre plusieurs Compte AWS entreprises ou AWS organisations, ou directement avec les responsables IAM d'un autre compte, offrant ainsi un accès détaillé aux métadonnées et aux données sous-jacentes.

Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour faciliter le partage des ressources. Lorsque vous partagez une ressource de catalogue avec un autre compte, AWS RAM envoie une invitation au compte bénéficiaire pour qu'il accepte ou rejette la subvention de ressource.

Les services analytiques intégrés tels qu'Amazon Athena et Redshift Spectrum nécessitent des liens de ressources pour pouvoir inclure des ressources partagées dans les requêtes. Les directeurs doivent créer un lien vers une ressource partagée depuis une autre Compte AWS ressource. AWS Glue Data Catalog Pour plus d'informations sur les liens vers des ressources, consultez [Comment fonctionnent les liens vers des ressources dans Lake Formation](#).

Un conteneur de liens de catalogue est un objet de catalogue de données qui fait référence à un catalogue de base de données fédéré local ou entre comptes provenant d'autres comptes. AWS Vous pouvez également créer des liens de base de données et des liens de table dans un conteneur de liens de catalogue. Lorsque vous créez un lien de base de données ou un lien de table, vous devez spécifier une ressource cible résidant dans le même catalogue cible au niveau de la base de données Amazon Redshift (base de données Amazon Redshift).

Pour créer un conteneur de liens de catalogue, vous avez besoin de la Lake Formation `CREATE_CATALOG` ou de l'`glue:CreateCatalog` autorisation.

Vous devez disposer des paramètres de version inter-comptes version 4 ou supérieure pour partager des bases de données ou des tables dans le catalogue fédéré entre Compte AWS s.

Création d'un conteneur de liens de catalogue vers un catalogue fédéré entre comptes

Vous pouvez créer un conteneur de liens de catalogue qui pointe vers un catalogue fédéré au niveau de la base de données Redshift dans n'importe quelle AWS région à l'aide de la AWS Lake Formation console, AWS Glue `CreateCatalog` de l'API ou (). AWS Command Line Interface AWS CLI

Pour créer un conteneur de liens de catalogue vers un catalogue partagé (console)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant que directeur disposant de l'`CREATE_CATALOG` autorisation de Lake Formation.
2. Dans le volet de navigation, choisissez Catalogues, puis sélectionnez Créer un catalogue.
3. Sur la page Définir les détails du catalogue, fournissez les informations suivantes :

Nom

Entrez un nom qui respecte les mêmes règles qu'un nom de catalogue. Le nom peut être identique à celui du catalogue partagé cible.

Type

Choisissez le conteneur de liens de catalogue comme type de catalogue.

Source

Sélectionnez Redshift.

Catalogue Target Redshift

Sélectionnez un catalogue fédéré au niveau de la base de données Redshift ou choisissez un catalogue local (détenu) dans la liste.

La liste contient tous les catalogues partagés avec votre compte. Notez que l'ID de compte du propriétaire du catalogue est indiqué avec chaque catalogue. Si aucun catalogue dont vous savez qu'il a été partagé avec votre compte ne s'affiche, vérifiez les points suivants :

- Si vous n'êtes pas un administrateur de lac de données, vérifiez que l'administrateur du lac de données vous a accordé les autorisations de Lake Formation sur le catalogue.
- Si vous êtes administrateur d'un lac de données et que votre compte n'appartient pas à la même AWS organisation que le compte octroyant, assurez-vous d'avoir accepté l'invitation AWS Resource Access Manager (AWS RAM) au partage de ressources pour le catalogue. Pour de plus amples informations, veuillez consulter [Acceptation d'une invitation de partage de ressources de AWS RAM](#).

 Note

Lorsque vous créez un conteneur de liens de catalogue via la console, la liste déroulante Target Redshift Catalog peut s'afficher No matches lorsque vous essayez de sélectionner un catalogue Redshift entre comptes. Malgré cet affichage, vous pouvez saisir manuellement l'ARN cible du catalogue fédéré au niveau de la base de données Amazon Redshift (base de données Amazon Redshift) dans le champ de saisie, et le formulaire fonctionnera toujours correctement. Par exemple : `arn:aws:glue:us-east-1:123456789012:catalog/federated-catalog-redshift/dev`.

Ce comportement se produit car la console ne peut rechercher des candidats potentiels que dans le compte actuellement connecté. La liste déroulante est conçue comme une fonctionnalité de saisie automatique, mais vous pouvez toujours la saisir manuellement ARNs pour un accès entre comptes.

4. Pour permettre aux moteurs de requêtes Apache Iceberg de lire et d'écrire dans les espaces de noms Amazon Redshift AWS Glue , créez un cluster Amazon Redshift géré avec les ressources de calcul et de stockage nécessaires pour effectuer des opérations de lecture et d'écriture sans impact sur les charges de travail de l'entrepôt de données Amazon Redshift. Vous devez fournir un rôle IAM avec les autorisations requises pour transférer des données depuis et vers le compartiment Amazon S3.
5. Choisissez Suivant.
6. (Facultatif) Choisissez Ajouter des autorisations pour accorder des autorisations à d'autres principaux.

Toutefois, l'octroi d'autorisations sur un conteneur de liens de catalogue n'accorde pas d'autorisations sur le catalogue cible (lié). Vous devez accorder des autorisations sur le catalogue cible séparément pour que le lien du catalogue soit visible dans Athena.

7. Passez ensuite en revue les détails du conteneur de liens du catalogue et sélectionnez Créer un catalogue.

Vous pouvez ensuite afficher le nom du conteneur de liens sur la page Catalogues.

Vous pouvez désormais créer des liens de base de données et des liens de table dans le conteneur de liens de catalogue pour permettre l'accès depuis les moteurs de requête.

Exemple de CLI pour créer un conteneur de liens de catalogue

- Dans l'exemple suivant, l'`TargetRedshiftCatalog` spécifie l'ARN du catalogue fédéré au niveau de la base de données Amazon Redshift (base de données Amazon Redshift). Le `DataLakeAccess` doit être activé lorsque vous créez le conteneur de liens de catalogue.

```
aws glue create-catalog \  
  --cli-input-json \  
  '{  
    "Name": "linkcontainer",  
    "CatalogInput": {  
      "TargetRedshiftCatalog": {  
        "CatalogArn": "arn:aws:us-east-1:123456789012:catalog/nscatalog/dev"  
      },  
      "CatalogProperties": {  
        "DataLakeAccessProperties" : {  
          "DataLakeAccess" : true,  
          "DataTransferRole" : "arn:aws:iam::111122223333:role/  
DataTransferRole"  
        }  
      }  
    }  
  }'
```

Création de liens vers des ressources sous le conteneur de liens du catalogue

Vous pouvez créer des liens vers des ressources vers des bases de données et des tables dans un conteneur de liens de catalogue. Lorsque vous créez des liens vers des ressources de base de données ou des liens vers des ressources de table, vous devez spécifier une ressource cible résidant dans le même catalogue cible au niveau de la base de données Amazon Redshift (base de données Amazon Redshift) que celui vers lequel pointe le conteneur de liens.

Vous pouvez créer un lien de ressource vers une base de données Amazon Redshift partagée ou une table à l'aide de la AWS Lake Formation console, de l'API ou AWS Command Line Interface (AWS CLI).

- Pour obtenir des instructions complètes, consultez [Création d'un lien de ressource vers une base de données de catalogue de données partagée](#).

Voici un AWS CLI exemple de création d'un lien vers une ressource de base de données sous un conteneur de liens de catalogue.

```
aws glue create-database \  
  --cli-input-json \  
  '{  
    "CatalogId": "111122223333:linkcontainer",  
    "DatabaseInput": {  
      "Name": "dblink",  
      "TargetDatabase": {  
        "CatalogId": "123456789012:nscatalog/dev",  
        "DatabaseName": "schema1"  
      }  
    }  
  }'
```

- Pour créer un lien vers une ressource de table sous un conteneur de liens de catalogue, vous devez d'abord créer une AWS Glue base de données dans le répertoire local AWS Glue Data Catalog pour contenir le lien vers une ressource de table.

Pour plus d'informations sur la création de liens de ressources vers des tables partagées, consultez [Création d'un lien de ressource vers une table de catalogue de données partagée](#).

- Création d'une base de données contenant la table (exemple de lien de ressource)

```
aws glue create-database \  
  --cli-input-json \  
  '{  
    "CatalogId": "111122223333:linkcontainer",  
    "DatabaseInput": {  
      "Name": "db1",  
      "Description": "creating parent database for table link"  
    }  
  }'
```

- Exemple de lien vers les ressources d'une table

```
aws glue create-table \  
  --cli-input-json \  
  '{  
    "CatalogId": "111122223333:linkcontainer",  
    "DatabaseName": "db1",  
    "TableInput": {  
      "Name": "tablelink",  
      "TargetTable": {  
        "CatalogId": "123456789012:nscatalog/dev",  
        "DatabaseName": "schema1",  
        "Name": "table1"  
      }  
    }  
  }'
```

Supprimer un catalogue fédéré

Vous pouvez supprimer les catalogues fédérés que vous avez créés à l' AWS Glue Data Catalog aide de l' `glue:DeleteCatalog` opération ou de la console. AWS Lake Formation

Pour supprimer un catalogue fédéré (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données.
3. Choisissez le catalogue que vous souhaitez supprimer dans la liste des catalogues.
4. Choisissez Supprimer des actions.
5. Choisissez Supprimer pour confirmer et le catalogue fédéré sera supprimé du catalogue de données.

Delete catalog gluebqcatalog



Permanently delete catalog **gluebqcatalog**? This action can't be undone.

 Proceeding with this action will delete the catalog.

To confirm this deletion, type **gluebqcatalog**.

Cancel

Drop

Pour supprimer un catalogue fédéré (CLI)

- ```
aws glue delete-catalog
--catalog-id 123456789012:catalog name
```

## Interrogation de catalogues fédérés

Une fois que vous avez accordé des autorisations aux autres principaux, ils peuvent se connecter et commencer à interroger les tables des catalogues fédérés en se connectant aux outils SQL via Amazon Redshift, Amazon EMR et ETL. Amazon Athena AWS Glue

Pour plus d'informations sur la connexion au point de terminaison de AWS Glue Data Catalog l'extension Apache Iceberg Rest ou à l'application Spark autonome, consultez la section [Accès à la AWS Glue Data Catalog section du](#) Guide du AWS Glue développeur.

Vous pouvez utiliser les requêtes DDL (Data Definition Language) pour créer et gérer des tables dans la base de données à l'aide d'Apache Spark sur Amazon EMR. Pour créer et supprimer des tables dans la base de données Amazon Redshift, le principal doit disposer des autorisations Lake Formation `Create table`. `Drop`

Pour plus d'informations sur l'octroi d'autorisations au catalogue de données, consultez [Octroi d'autorisations sur les ressources du catalogue de données](#).

Pour plus d'informations sur l'interrogation des ressources du catalogue Amazon Athena, consultez la section [Interrogation depuis le guide AWS Glue Data Catalog de l' Amazon Athena utilisateur d'Amazon Athena](#).

## Ressources supplémentaires

Vous pouvez utiliser [Amazon SageMaker Lakehouse](#) pour obtenir un accès unifié aux données dans les entrepôts de données et les lacs de données. Grâce à SageMaker Lakehouse, vous pouvez utiliser des moteurs d'analyse, d'apprentissage automatique et de business intelligence préférés via une API REST Apache Iceberg ouverte pour garantir un accès sécurisé aux données grâce à des contrôles d'accès cohérents et précis.

- [SageMaker Atelier Amazon](#)
- [Simplifiez l'accès aux données pour votre entreprise à l'aide d'Amazon SageMaker Lakehouse](#)

## Fédération en sources de données externes dans le AWS Glue Data Catalog

Vous pouvez connecter le AWS Glue Data Catalog (catalogue de données) à des entrepôts de données tels qu'Amazon Redshift, Snowflake, à des bases de données cloud telles qu'Amazon RDS, Oracle, Amazon DynamoDB à des services de streaming tels qu'Amazon MSK, et à des systèmes sur site tels que Teradata à l'aide de connexions. AWS Glue Ces connexions sont enregistrées dans le AWS Glue Data Catalog et enregistrées auprès de celui-ci AWS Lake Formation, ce qui vous permet de créer un catalogue fédéré pour chaque source de données disponible.

Un catalogue fédéré est un conteneur de niveau supérieur qui pointe vers une base de données dans un système de données externe. Il vous permet d'interroger les données directement depuis le système de données externe sans processus d'extraction, de transformation et de chargement (ETL).

Pour plus d'informations sur AWS Glue les connexions, consultez la section [Connexion aux données](#) dans le Guide du AWS Glue développeur.

Les administrateurs de data lake peuvent créer des catalogues fédérés à l'aide d'[Amazon SageMaker Lakehouse](#) ou. [Amazon Athena](#)

Les administrateurs de data lake peuvent ensuite accorder des autorisations détaillées sur les objets du catalogue à l'aide de Lake Formation, en contrôlant l'accès à différents niveaux tels que le catalogue, la base de données, la table, la colonne, la ligne ou la cellule. Les analystes de données peuvent découvrir et interroger les sources de données cataloguées à l'aide d'Athena, Lake Formation appliquant les politiques d'accès définies. Les analystes peuvent joindre des données provenant de plusieurs sources en une seule requête sans avoir à se connecter à chaque source individuellement.

## Rubriques

- [Flux de travail](#)
- [Conditions requises pour connecter le catalogue de données à des sources de données externes](#)
- [Création d'un catalogue fédéré à l'aide d'une connexion AWS Glue](#)
- [Affichage des objets du catalogue](#)
- [Supprimer un catalogue fédéré](#)
- [Interrogation de catalogues fédérés](#)
- [Ressources supplémentaires](#)

## Flux de travail

Un administrateur de lac de données ou un utilisateur disposant des autorisations requises effectue les étapes suivantes pour le connecter AWS Glue Data Catalog à une source de données externe.

1. Crée une AWS Glue connexion à la source de données. Lorsque vous enregistrez la connexion, le rôle IAM utilisé pour enregistrer la connexion doit avoir accès à la fonction Lambda et à l'emplacement du compartiment de déversement Amazon S3.
2. Enregistre le lien avec Lake Formation.
3. Crée un catalogue fédéré dans le catalogue de données à l'aide d'une AWS Glue connexion pour se connecter aux sources de données disponibles. Les bases de données, les tables et les vues sont automatiquement cataloguées dans le catalogue de données et enregistrées auprès de Lake Formation.
4. Accorde l'accès à des catalogues, bases de données et tables spécifiques aux analystes de données à l'aide des autorisations de Lake Formation. Des politiques de contrôle d'accès précises peuvent être définies pour les lacs de données, les entrepôts et les sources OLTP à l'aide de Lake Formation, ce qui permet d'activer des filtres de sécurité au niveau des lignes et des colonnes.

Les analystes de données peuvent ensuite accéder à toutes les données via le catalogue de données à l'aide de requêtes SQL dans Athena, sans avoir besoin de connexions distinctes ni d'informations d'identification de source de données. Les analystes peuvent exécuter des requêtes SQL fédérées qui analysent les données provenant de sources multiples, en joignant les données sur place sans pipelines de données complexes.

## Conditions requises pour connecter le catalogue de données à des sources de données externes

Pour vous connecter AWS Glue Data Catalog à des sources de données externes, enregistrer la connexion auprès de Lake Formation et configurer des catalogues fédérés, vous devez remplir les conditions suivantes :

### Note

Nous recommandons à un administrateur du lac de données de Lake Formation de créer les AWS Glue connexions pour se connecter aux sources de données externes et de créer les catalogues fédérés.

### 1. Création des rôles IAM.

- Créez un rôle doté des autorisations nécessaires pour déployer les ressources (fonction Lambda, compartiment de déversement Amazon S3, rôle IAM et AWS Glue connexion) requises pour créer une connexion à la source de données externe.
- Créez un rôle doté des autorisations minimales nécessaires pour accéder aux propriétés de AWS Glue connexion (fonction Lambda et compartiment de déversement Amazon S3). C'est le rôle que vous allez inclure lorsque vous enregistrez la connexion avec Lake Formation.

Pour utiliser Lake Formation pour gérer et sécuriser les données de votre lac de données, vous devez enregistrer la AWS Glue connexion auprès de Lake Formation. Lake Formation peut ainsi vendre des informations d'identification à Amazon Athena pour interroger les sources de données fédérées.

Le rôle doit disposer `Select` d'`Describe` autorisations sur le compartiment Amazon S3 et sur la fonction Lambda.



- s3 : ListBucket
- s3 : GetObject
- lambda : InvokeFunction

JSON

```
{
 "Version": "2012-10-17",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "s3:*"
],
 "Resource": [
 "arn:aws:s3:::amzn-s3-demo-bucket1/object/*",
 "arn:aws:s3:::amzn-s3-demo-bucket1/object"
]
 },
 {
 "Sid": "lambdainvoke",
 "Effect": "Allow",
 "Action": "lambda:InvokeFunction",
 "Resource": "arn:aws:lambda:us-east-1:123456789012:function:example-lambda-function"
 },
 {
 "Sid": "gluepolicy",
 "Effect": "Allow",
 "Action": "glue:*",
 "Resource": "*"
 }
]
}
```

- Ajoutez la politique de confiance suivante au rôle IAM utilisé lors de l'enregistrement de la connexion :

JSON

```
{
```

```

 "Version": "2012-10-17",
 "Statement": [
 {
 "Effect": "Allow",
 "Principal": {
 "Service": [
 "lakeformation.amazonaws.com"
]
 },
 "Action": "sts:AssumeRole"
 }
]
 }
}

```

- L'administrateur du lac de données qui enregistre la connexion doit disposer de l'`iam:PassRole` autorisation associée au rôle.

Voici une politique intégrée qui accorde cette autorisation. Remplacez `<account-id>` par un numéro de AWS compte valide et remplacez `<role-name>` par le nom du rôle.

JSON

```

{
 "Version": "2012-10-17",
 "Statement": [
 {
 "Sid": "PassRolePermissions",
 "Effect": "Allow",
 "Action": [
 "iam:PassRole"
],
 "Resource": [
 "arn:aws:iam::111122223333:role/example-role-name"
]
 }
]
}

```

- Pour créer des catalogues fédérés dans Data Catalog, assurez-vous que le rôle IAM que vous utilisez est un administrateur de lac de données de Lake Formation en vérifiant les paramètres du lac de données (). `aws lakeformation get-data-lake-settings`

Si vous n'êtes pas administrateur de data lake, vous devez disposer de l'`CREATE_CATALOG` autorisation Lake Formation pour créer un catalogue. L'exemple suivant montre comment accorder les autorisations requises pour créer des catalogues.

```
aws lakeformation grant-permissions \
--cli-input-json \
 '{
 "Principal": {
 "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
 },
 "Resource": {
 "Catalog": {
 }
 },
 "Permissions": [
 "CREATE_CATALOG",
 "DESCRIBE"
]
 }'
```

2. Ajoutez la politique de clé suivante à la AWS KMS clé si vous utilisez une clé gérée par le client pour chiffrer les données de la source de données. Remplacez le numéro de compte par un numéro de AWS compte valide et spécifiez le nom du rôle. Par défaut, les données sont chiffrées à l'aide d'une clé KMS. Lake Formation propose une option permettant de créer votre clé KMS personnalisée pour le chiffrement. Si vous utilisez une clé gérée par le client, vous devez ajouter des politiques clés spécifiques à la clé.

Pour plus d'informations sur la gestion des autorisations d'une clé gérée par le client, consultez [Clés gérées par le client](#).

## JSON

```
{
 "Version": "2012-10-17",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
```

```
 "kms:Encrypt",
 "kms:Decrypt",
 "kms:ReEncrypt*",
 "kms:GenerateDataKey*",
 "kms:DescribeKey"
],
 "Resource": "arn:aws:kms:us-east-1:123456789012:key/key-1"
}
]
```

## Création d'un catalogue fédéré à l'aide d'une connexion AWS Glue

Pour connecter les sources AWS Glue Data Catalog de données externes, vous devez utiliser des AWS Glue connexions qui permettent la communication avec les sources de données externes. Vous pouvez créer des AWS Glue connexions à l'aide de la AWS Glue console, de l'API [Create connection](#) et de la console Amazon SageMaker Lakehouse.

Pour obtenir des instructions détaillées sur la création d'une AWS Glue connexion, consultez [Connexion aux données](#) dans le guide du AWS Glue développeur ou [Création de connexions dans Amazon SageMaker Lakehouse](#).

Lorsqu'un utilisateur exécute une requête sur des tables fédérées, Lake Formation envoie des informations d'identification qui invoquent une AWS Lambda fonction spécifiée dans la AWS Glue connexion pour récupérer des objets de métadonnées depuis la source de données.

### AWS Management Console

Pour créer un catalogue fédéré à partir d'une source de données externe et configurer des autorisations (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données.
3. Sélectionnez l'option Créer un catalogue.
4. Sur la page de détails du catalogue, entrez les informations suivantes :

- Step 1  
 ● **Set catalog details**  
 Step 2 - optional  
 ● Grant permissions  
 Step 3  
 ● Review and create

## Set catalog details

Create a catalog in the Data Catalog.

### Catalog details

A catalog is the top level in the Data Catalog's three-level data hierarchy and contains Data Catalog objects.

#### Name

Catalog name is required, in lowercase characters, and no longer than 255 characters.

#### Type

#### Source

#### Connection

#### Description - optional

Descriptions can be up to 2048 characters long.

### Register Glue connection with Lake Formation

You can access this catalog from AWS Glue data connections.

#### IAM role

Choose a role that has permissions to invoke an AWS Glue connector.


[View](#)
[Create an IAM role](#)

☒ **Activate the connector and connect to the data source.**

A connector is a piece of code that runs on AWS Lambda that translates between the target data source and query engine (Athena).

#### Encryption options

Your data is encrypted by default with a key that AWS owns and manages for you. To choose a different key, customize your encryption settings.

☐ **Customize encryption settings**

To use the default key, clear this option.

[Cancel](#)
[Skip to Review and create](#)
[Next](#)

- Nom : nom unique pour votre catalogue fédéré. Le nom ne peut pas être modifié et doit être en minuscules. Le nom peut comporter un maximum de 255 caractères. compte.
  - Type — Choisissez le catalogue fédéré comme type de catalogue.
  - Source : choisissez une source de données dans la liste déroulante. Les sources de données pour lesquelles vous avez créé des connexions s'affichent. Pour plus d'informations sur la création d'une AWS Glue connexion à une source de données externe, consultez [Création de connexions pour les connecteurs](#) dans le Guide du AWS Glue développeur ou [Création de connexions dans Amazon SageMaker Lakehouse](#).
  - Connexion : choisissez une AWS Glue connexion existante à la source de données.
  - Description — Entrez une description pour le catalogue créé à partir de la source de données.
5. Choisissez un rôle IAM que Lake Formation assumera pour vendre les informations d'identification permettant au moteur de requête d'accéder aux données depuis la source de données. Ce rôle doit disposer des autorisations requises pour accéder à la AWS Glue

connexion et appeler la fonction Lambda pour accéder aux données de la source de données externe.

Vous pouvez également créer un nouveau rôle dans la console IAM.

Consultez la [Conditions requises pour connecter le catalogue de données à des sources de données externes](#) section pour connaître les autorisations requises.

6. Sélectionnez l'option Activer le connecteur pour vous connecter à la source de données afin de permettre à Athena d'exécuter des requêtes fédérées.

Pour obtenir la liste des connecteurs pris en charge, consultez la section [Enregistrer votre connexion](#) dans le guide de l'utilisateur Amazon Athena.

7. Options de chiffrement — Choisissez l'option Personnaliser les paramètres de chiffrement si vous souhaitez utiliser une clé personnalisée pour chiffrer le catalogue. Pour utiliser une clé personnalisée, vous devez ajouter une stratégie de clé gérée personnalisée supplémentaire à votre clé KMS.
8. Choisissez Next pour accorder des autorisations aux autres principaux.
9. Sur la page Accorder des autorisations, choisissez Ajouter des autorisations.
10. Sur l'écran Ajouter des autorisations, choisissez les principes et les types d'autorisations à accorder.

## Add permissions

### Principals

Choose the principals to grant permissions.

☒ **IAM users and roles**  
Users or roles from this AWS account.

☐ **SAML users and groups**  
SAML users and group or QuickSight ARNs.

☐ **External accounts**  
AWS account, AWS organization or IAM principal outside of this account

#### IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

UserRole X  
Role

### Catalog permissions

Choose the permissions to grant on the catalog. Choosing Super user overwrites individual permissions, granting unrestricted administrative access.

☐ **Super user**  
A super user has unrestricted administrative privileges to perform any operation on all resources within the catalog (databases, tables, and views).

#### Catalog permissions

Choose specific access permissions to grant.

☒ **Create database**

☐ Describe

☐ Alter

☐ Drop

☐ **Super**  
This permission is the union of all the individual permissions to the left, and supersedes them.

#### Grantable permissions

Choose the permission that can be granted to others.

☒ **Create database**

☐ Describe

☐ Alter

☐ Drop

☐ **Super**  
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel

Add

- Dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels octroyer des autorisations.

- Utilisateurs et rôles IAM : choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.
- Utilisateurs et groupes SAML : pour le SAML et Amazon Quick les utilisateurs et groupes, entrez un ou plusieurs Amazon Resource Names (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou pour les utilisateurs ou groupes ARNs Amazon Quick. Appuyez sur Entrée après chaque ARN.
- Dans la section Autorisations, sélectionnez les autorisations et les autorisations octroyables.

Sous Autorisations du catalogue, sélectionnez une ou plusieurs autorisations à octroyer.

Choisissez Super utilisateur pour octroyer des autorisations administratives illimitées sur toutes les ressources du catalogue.

Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS Cette option n'est pas prise en charge lorsque vous octroyez des autorisations à un principal IAM à partir d'un compte externe.

11. Choisissez Next pour consulter les informations et créer le catalogue. La liste des catalogues présente le nouveau catalogue fédéré.

La liste des emplacements des données indique la connexion fédérée récemment enregistrée.

Data lake locations (7)

Find data lake storage

Actions

Register location

< 1 >

|                       | Data lake location                               | IAM role                                          | Location Type        | Permission mode    | Last modified                     |
|-----------------------|--------------------------------------------------|---------------------------------------------------|----------------------|--------------------|-----------------------------------|
| <input type="radio"/> | <a href="#">ddb_ds_3</a>                         | <a href="#">SageMakerStudioQueryExecutionR...</a> | Federated connection | Lake Formation     | November 26, 2024 at 10:34 PM UTC |
| <input type="radio"/> | <a href="#">postgre_db2</a>                      | <a href="#">SageMakerStudioQueryExecutionR...</a> | Federated connection | Lake Formation     | November 24, 2024 at 11:12 AM UTC |
| <input type="radio"/> | <a href="#">sf_ds2</a>                           | <a href="#">SageMakerStudioQueryExecutionR...</a> | Federated connection | Lake Formation     | November 24, 2024 at 3:27 AM UTC  |
| <input type="radio"/> | <a href="#">s3://amazon-sagemaker-5390106...</a> | <a href="#">datazone_usr_role_50wwm8ts855...</a>  | Amazon S3            | Lake Formation     | November 24, 2024 at 3:10 AM UTC  |
| <input type="radio"/> | <a href="#">ddb_ds_2</a>                         | <a href="#">SageMakerStudioQueryExecutionR...</a> | Federated connection | Lake Formation     | November 24, 2024 at 3:05 AM UTC  |
| <input type="radio"/> | <a href="#">s3://amazon-sagemaker-5390106...</a> | <a href="#">datazone_usr_role_adtmv7d4lm98...</a> | Amazon S3            | Lake Formation     | November 23, 2024 at 9:15 PM UTC  |
| <input type="radio"/> | <a href="#">s3://data-lake-pk-us-east-2</a>      | <a href="#">AWSServiceRoleForLakeFormation...</a> | Amazon S3            | Hybrid access mode | November 21, 2024 at 7:40 PM UTC  |



## AWS CLI

Pour créer un catalogue fédéré à partir d'une source de données externe et configurer des autorisations

1. L'exemple suivant montre comment créer une AWS Glue connexion.

```
aws glue create-connection
--connection-input \
 '{
 "Name": "DynamoDB connection",
 "ConnectionType": "DYNAMODB",
 "Description": "A connection created for DynamoDB",
 "ConnectionProperties": {},
 "AthenaProperties": {"spill_prefix": "your_spill_prefix",
 "lambda_function_arn": "Lambda_function_arn",
 "spill_bucket": "Your_Bucket_name",
 "AuthenticationConfiguration": {}
 }'
```

2. L'exemple suivant montre comment enregistrer une AWS Glue connexion avec Lake Formation.

```
aws lakeformation register-resource
--cli-input-json \
 '{"ResourceArn": "arn:aws:glue:us-east-1:123456789012:connection/
dynamo", "RoleArn": "arn:aws:iam::123456789012:role/
AdminTelemetry", "WithFederation": true}'
```

3. L'exemple suivant montre comment créer un catalogue fédéré.

```
aws glue create-catalog
--cli-input-json \
 '{
 "Name": "ddbcatalog",
 "CatalogInput": {"CatalogProperties": {"DataLakeAccessProperties":
{"DataTransferRole": "arn:aws:iam::123456789012:role/role name"}},
 "CreateDatabaseDefaultPermissions": [],
 "CreateTableDefaultPermissions": [],
 "FederatedCatalog": {"ConnectionName": "dynamo", "Identifier": "dynamo"}
 }'
```

```
}'
```

## Affichage des objets du catalogue

Pour chaque source de données disponible, AWS Glue crée un catalogue correspondant dans le AWS Glue Data Catalog. Après avoir créé un catalogue, vous pouvez afficher les bases de données et les tables du catalogue à l'aide de la console Lake Formation ou AWS CLI. Dans

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Choisissez Catalogues sous Catalogue de données. La page des catalogues affiche les catalogues pour lesquels vous avez des autorisations.

Catalogs (11)

A catalog is the top level in the Data Catalog's three-level data hierarchy and contains Data Catalog objects.

Find catalogs by name

< 1 > ⚙

|                       | Name                | Type      | Source                 | Owner account... | Shared resource | Shared resourc... | Shared resource owner region |
|-----------------------|---------------------|-----------|------------------------|------------------|-----------------|-------------------|------------------------------|
| <input type="radio"/> | 30005               | Default   | Default catalog        | 0005             | -               | -                 | -                            |
| <input type="radio"/> | bkaiyuan_nscatal... | Federated | Redshift               | 0005             | -               | -                 | -                            |
| <input type="radio"/> | bkaiyuan_test_ca... | Federated | -                      | 0005             | -               | -                 | -                            |
| <input type="radio"/> | linkcontainer-leon  | Managed   | Catalog Link container | 0005             | -               | -                 | -                            |
| <input type="radio"/> | mymulticatalog      | Federated | TPCDS                  | 0005             | -               | -                 | -                            |
| <input type="radio"/> | test-bug-share      | Federated | Redshift               | 0005             | -               | -                 | -                            |
| <input type="radio"/> | test-zetl           | Managed   | Redshift               | 0005             | -               | -                 | -                            |
| <input type="radio"/> | test-zetl-mwg       | Managed   | Redshift               | 0005             | -               | -                 | -                            |
| <input type="radio"/> | tpcdscatalog        | Federated | TPCDS                  | 0005             | -               | -                 | -                            |
| <input type="radio"/> | yansoncatalog2      | Federated | Redshift               | 0005             | -               | -                 | -                            |
| <input type="radio"/> | zetltest123         | Managed   | Redshift               | 0005             | -               | -                 | -                            |

3. Choisissez un catalogue dans la liste pour afficher les bases de données et les tables qu'il contient. La liste contient les bases de données de votre compte et les liens vers les ressources, qui sont des liens vers des bases de données partagées et des tables dans des comptes externes, et sont utilisés pour l'accès entre comptes aux données du lac de données.

**Catalog summary**

|                                                                   |                             |                                      |
|-------------------------------------------------------------------|-----------------------------|--------------------------------------|
| <b>Name</b><br>451785580005                                       | <b>Data encryption</b><br>- | <b>IAM role</b><br>-                 |
| <b>Catalog ARN</b><br>arn:aws:glue:us-west-2:451785580005:catalog |                             | <b>KMS key for optimization</b><br>- |

**Objects** | Permissions | Table optimizations

**Databases (1/14)**

Find databases

| Name                                         | Owner account ID | Lake Fo... | Default ...  | Shared ...    | Shared ...  | Shared ... | Amazo...  | Descript...  |
|----------------------------------------------|------------------|------------|--------------|---------------|-------------|------------|-----------|--------------|
| <b>arfarajpostgresqldb</b>                   |                  | -          | Lake Form... | -             | -           | -          | -         | -            |
| <b>aws:cloudtrail</b>                        |                  | -          | Lake Form... | -             | -           | -          | -         | -            |
| <b>default</b>                               |                  | -          | Lake Form... | -             | -           | -          | -         | -            |
| <b>gluedynamodb</b>                          |                  | -          | Lake Form... | -             | -           | -          | -         | -            |
| <b>mynowflakedb</b>                          |                  | -          | Lake Form... | -             | -           | -          | -         | -            |
| <b>snowflakedb</b>                           |                  | -          | Lake Form... | -             | -           | -          | -         | -            |
| <b>test-db-0737fa687d584b2d9ab72fbd...</b>   |                  | -          | Lake Form... | test-db-0...  | 45178558... | -          | -         | -            |
| <b>test-db-1927b03560764a4b81a216e9...</b>   |                  | -          | Lake Form... | test-db-1...  | 45178558... | -          | -         | -            |
| <b>test-db-32ee54b6949b4fdd85b8ff066...</b>  |                  | -          | Lake Form... | test-db-3...  | 45178558... | -          | -         | -            |
| <b>test-db-5978c2e076aa4583a28df124f...</b>  |                  | -          | Lake Form... | test-db-5...  | 45178558... | -          | -         | -            |
| <b>test-db-5cebe417bf734eafbbeaf7d3d6...</b> |                  | -          | Lake Form... | test-db-5c... | 45178558... | -          | -         | -            |
| <b>test-db-7fa7ca8de2b84232ae3e8dcf...</b>   |                  | -          | Lake Form... | -             | -           | -          | http://db | database ... |
| <b>test-db-bb19fb486dc14ad68813612...</b>    |                  | -          | Lake Form... | -             | -           | -          | http://db | database ... |
| <b>test-db-cdec6ecaa0f143b7b4d6f24a8...</b>  |                  | -          | Lake Form... | test-db-cd... | 45178558... | -          | -         | -            |

4. Choisissez l'option Tables sous Afficher pour afficher et gérer les tables de la base de données.

## AWS CLI exemples de visualisation de catalogues et de bases de données

L'exemple suivant montre comment afficher un catalogue à l'aide de AWS CLI

```
aws glue get-catalog \
--catalog-id 123456789012:dynamodbcatalog
```

L'exemple suivant montre comment demander tous les catalogues du compte.

```
aws glue get-catalogs \
--recursive
```

L'exemple de demande suivant montre comment obtenir les bases de données dans le catalogue.

```
aws glue get-database \
--catalog-id 123456789012:dynamodbcatalog
```

```
--database-name database name
```

## Supprimer un catalogue fédéré

Vous pouvez supprimer les catalogues fédérés que vous avez créés à l' AWS Glue Data Catalog aide de l'`glue:DeleteCatalog`opération ou de la console. AWS Lake Formation

Pour supprimer un catalogue fédéré (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données.
3. Choisissez le catalogue que vous souhaitez supprimer dans la liste des catalogues.
4. Choisissez Supprimer des actions.
5. Choisissez Supprimer pour confirmer et le catalogue fédéré sera supprimé du catalogue de données.

### Delete catalog gluebqcatalog



Permanently delete catalog **gluebqcatalog**? This action can't be undone.

 Proceeding with this action will delete the catalog.

To confirm this deletion, type **gluebqcatalog**.

Cancel

Drop

Pour supprimer un catalogue fédéré (CLI)

- ```
aws glue delete-catalog  
--catalog-id 123456789012:catalog name
```

Interrogation de catalogues fédérés

Une fois que vous avez accordé des autorisations aux autres principaux, ils peuvent se connecter et commencer à interroger les tables des catalogues fédérés à l'aide d'Athena.

Pour créer et supprimer des tables dans la base de données fédérée, le principal doit disposer des `Drop` autorisations `Lake FormationCreate table`.

Pour plus d'informations sur l'octroi d'autorisations au catalogue de données, consultez [Octroi d'autorisations sur les ressources du catalogue de données](#).

Pour plus d'informations sur l'interrogation du catalogue de données Amazon Athena, consultez la section [Interrogation depuis le guide AWS Glue Data Catalog de l' Amazon Athena utilisateur d'Amazon Athena](#).

Ressources supplémentaires

Dans ce billet de blog, nous expliquons comment les analystes de données peuvent désormais accéder et interroger en toute sécurité les données stockées en dehors des lacs de données S3, y compris les entrepôts de données et les Amazon DynamoDB bases de données Amazon Redshift, le tout via une expérience unique et unifiée. Les administrateurs peuvent désormais appliquer des contrôles d'accès à différents niveaux de granularité pour garantir la protection des données sensibles tout en élargissant l'accès aux données. Cela permet aux entreprises d'accélérer les initiatives relatives aux données tout en préservant la sécurité et la conformité, ce qui permet de prendre des décisions plus rapides et axées sur les données.

- [Cataloguez et gérez les requêtes fédérées Amazon Athena avec Amazon Lakehouse SageMaker](#)

Intégration d'Amazon S3 Tables avec AWS Glue Data Catalog et AWS Lake Formation

[Les tables Amazon S3](#) fournissent un stockage S3 spécifiquement optimisé pour les charges de travail d'analyse, améliorant ainsi les performances des requêtes tout en réduisant les coûts. Les données figurant dans S3 Tables sont stockées dans un nouveau type de compartiment : un compartiment de tables, qui stocke les tables en tant que sous-ressources. Les tables S3 prennent en charge la norme Apache Iceberg, qui vous permet d'interroger facilement des données tabulaires dans des compartiments de tables Amazon S3 à l'aide de moteurs de requête populaires tels qu'Apache Spark.

Vous pouvez intégrer les tables Amazon S3 à AWS Glue Data Catalog l'aide de contrôles d'accès IAM ou de subventions IAM et Lake Formation :

- Contrôle d'accès IAM : utilise des politiques IAM pour contrôler l'accès aux tables S3 et au catalogue de données. Dans cette approche de contrôle d'accès, vous avez besoin d'autorisations IAM sur les ressources des tables S3 et sur les objets du catalogue de données pour accéder aux ressources.
- Contrôle d'accès à Lake Formation : utilise AWS Lake Formation des autorisations en plus des autorisations AWS Glue IAM pour contrôler l'accès aux tables S3 via le catalogue de données. Dans ce mode, les principaux ont besoin d'autorisations IAM pour interagir avec le catalogue de données, et les subventions de Lake Formation déterminent les ressources du catalogue (bases de données, tables, colonnes, lignes) auxquelles le principal peut accéder. Ce mode prend en charge à la fois un contrôle d'accès grossier (autorisations au niveau de la base de données et au niveau de la table) et un contrôle d'accès détaillé (sécurité au niveau des colonnes et au niveau des lignes). Lorsqu'un rôle enregistré est configuré et que la vente d'informations d'identification est activée, les autorisations IAM de S3 Tables ne sont pas requises pour le principal, car Lake Formation vend les informations d'identification au nom du principal en utilisant le rôle enregistré. Le contrôle d'accès à Lake Formation prend également en charge la vente d'informations d'identification pour les moteurs d'analyse tiers.

Cette section fournit des conseils pour configurer l'intégration avec AWS Lake Formation pour les scénarios suivants :

- Scénario A : vous avez intégré les tables S3 et le catalogue de données à l'aide des contrôles d'accès IAM et vous prévoyez maintenant de les utiliser AWS Lake Formation. Pour en savoir plus, consultez [Modification des contrôles d'accès pour l'intégration des tables S3](#).
- Scénario B : Vous prévoyez d'intégrer les tables S3 et le catalogue de données à l'aide de votre compte AWS Lake Formation et de votre région, mais vous ne les avez pas intégrés à votre compte et à votre région aujourd'hui. Commencez par la [Conditions préalables à l'intégration du catalogue de tables Amazon S3 au catalogue de données et à Lake Formation](#) section et suivez [Activation de l'intégration des tables Amazon S3](#).
- Scénario C : Vous avez intégré les tables S3 et le catalogue de données à l'aide d'IAM AWS Lake Formation et vous prévoyez maintenant d'utiliser IAM. Pour en savoir plus, consultez [Modification des contrôles d'accès pour l'intégration des tables S3](#).

Assurez-vous de suivre les étapes décrites dans la [section Intégration des tables S3 aux services d' AWS analyse](#) afin de disposer des autorisations appropriées pour accéder aux AWS Glue Data Catalog ressources de vos tables et pour utiliser les services d' AWS analyse.

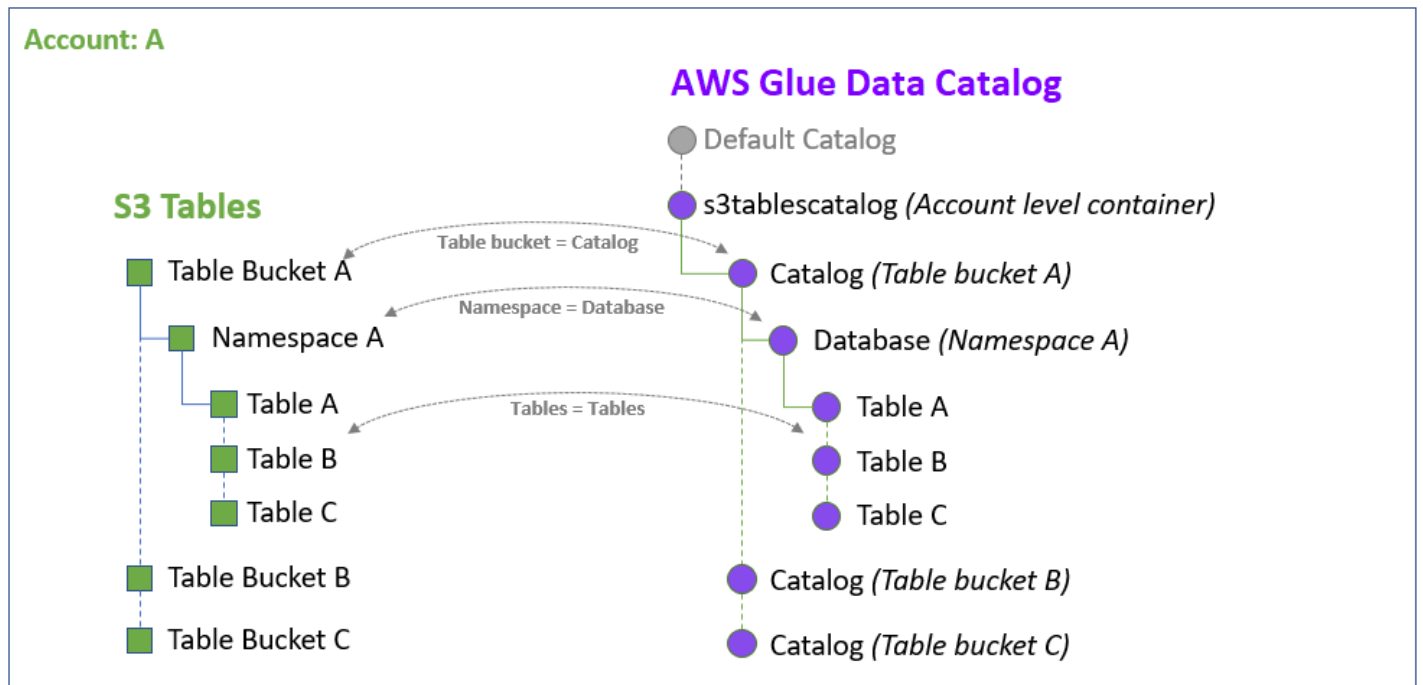
Rubriques

- [Comment fonctionne l'intégration du catalogue de données et de la Lake Formation](#)
- [Conditions préalables à l'intégration du catalogue de tables Amazon S3 au catalogue de données et à Lake Formation](#)
- [Activation de l'intégration des tables Amazon S3](#)
- [Création de bases de données et de tables dans le catalogue de tables S3](#)
- [Enregistrement d'un compartiment de table Amazon S3 dans un autre AWS compte](#)
- [Octroi d'autorisations](#)

Comment fonctionne l'intégration du catalogue de données et de la Lake Formation

Lorsque vous intégrez le catalogue de tables S3 au catalogue de données et à Lake Formation, le AWS Glue service crée un catalogue fédéré unique appelé `s3tablescatalog` dans le catalogue de données par défaut de votre compte, spécifique à votre Région AWS compte. L'intégration mappe toutes les ressources du bucket de tables Amazon S3 de votre compte Région AWS et du catalogue fédéré de la manière suivante :

- Les compartiments de table Amazon S3 deviennent un catalogue à plusieurs niveaux dans le catalogue de données.
- L'espace de noms Amazon S3 associé est enregistré en tant que base de données dans le catalogue de données.
- Les tables Amazon S3 du compartiment de tables deviennent des tables du catalogue de données.



Après l'intégration à Lake Formation, vous pouvez créer des tables Apache Iceberg dans le catalogue de compartiments de tables et y accéder via des moteurs AWS d'analyse intégrés tels qu' Amazon Athena Amazon EMR ainsi que des moteurs d'analyse tiers.

Lorsque vous activez également l'intégration de Lake Formation, cela permet un contrôle d'accès précis. AWS Lake Formation Cette approche de sécurité signifie qu'en plus des autorisations Gestion des identités et des accès AWS (IAM), vous devez accorder à votre principal IAM des autorisations Lake Formation sur vos tables avant de pouvoir travailler avec elles.

Il existe deux principaux types d'autorisations dans AWS Lake Formation :

- Les autorisations d'accès aux métadonnées contrôlent la capacité de créer, de lire, de mettre à jour et de supprimer des bases de données et des tables de métadonnées dans le catalogue de données.
- Les autorisations d'accès aux données sous-jacentes contrôlent la capacité à lire et à écrire des données dans les emplacements Amazon S3 sous-jacents vers lesquels pointent les ressources du catalogue de données.

Lake Formation utilise une combinaison de son propre modèle d'autorisations et du modèle d'autorisations IAM pour contrôler l'accès aux ressources des catalogues de données et aux données sous-jacentes :

- Pour qu'une demande d'accès aux ressources du catalogue de données ou aux données sous-jacentes aboutisse, elle doit passer les contrôles d'autorisation par IAM et Lake Formation.
- Les autorisations IAM contrôlent l'accès à la Formation AWS Glue APIs et aux ressources du lac, tandis que les autorisations Lake Formation contrôlent l'accès aux ressources du catalogue de données, aux sites Amazon S3 et aux données sous-jacentes.

Les autorisations Lake Formation ne s'appliquent que dans la région dans laquelle elles ont été accordées, et un principal doit être autorisé par un administrateur de lac de données ou un autre principal disposant des autorisations nécessaires pour obtenir les autorisations Lake Formation.

Conditions préalables à l'intégration du catalogue de tables Amazon S3 au catalogue de données et à Lake Formation

Vous trouverez ci-dessous les conditions requises pour activer l'intégration des tables Amazon S3 avec AWS Glue Data Catalog et AWS Lake Formation.

1. Lorsque vous activez l'intégration des tables Amazon S3, Lake Formation enregistre automatiquement l'emplacement des tables S3. Pour enregistrer l'emplacement du bucket de table auprès de Lake Formation, vous avez besoin d'un IAM role/user avec `lakeformation:RegisterResource` et `lakeformation:CreateCatalog` autorisations. Lorsqu'un utilisateur non administrateur disposant de ces autorisations enregistre l'emplacement d'un catalogue, Lake Formation lui accorde automatiquement l'`DATA_LOCATION_ACCESS` autorisation pour cet emplacement, ce qui permet au principal appelant d'effectuer toutes les opérations de Lake Formation prises en charge sur l'emplacement de données enregistré.
2. Lorsque vous activez l'intégration des tables S3, vous devez choisir un rôle IAM pour que Lake Formation fournisse des informations d'identification permettant l'accès aux données. Créez un rôle IAM pour accéder aux données de Lake Formation à vos compartiments de tables S3. Le rôle IAM utilisé lors de l'enregistrement du bucket de table auprès de Lake Formation nécessite les autorisations suivantes :

JSON

```
{  
  "Version": "2012-10-17",  
  "Statement": [  

```

```

    {
      "Sid": "LakeFormationPermissionsForS3ListTableBucket",
      "Effect": "Allow",
      "Action": [
        "s3tables:ListTableBuckets"
      ],
      "Resource": [
        "*"
      ]
    },
    {
      "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
      "Effect": "Allow",
      "Action": [
        "s3tables:CreateTableBucket",
        "s3tables:GetTableBucket",
        "s3tables:CreateNamespace",
        "s3tables:GetNamespace",
        "s3tables:ListNamespaces",
        "s3tables>DeleteNamespace",
        "s3tables>DeleteTableBucket",
        "s3tables:CreateTable",
        "s3tables>DeleteTable",
        "s3tables:GetTable",
        "s3tables:ListTables",
        "s3tables:RenameTable",
        "s3tables:UpdateTableMetadataLocation",
        "s3tables:GetTableMetadataLocation",
        "s3tables:GetTableData",
        "s3tables:PutTableData"
      ],
      "Resource": [
        "arn:aws:s3tables:us-east-1:123456789012:bucket/*"
      ]
    }
  ]
}

```

Pour de plus amples informations, veuillez consulter [Exigences relatives aux rôles utilisés pour enregistrer des sites](#).

3. Ajoutez la politique de confiance suivante au rôle IAM pour permettre au service Lake Formation d'assumer le rôle et de vendre des informations d'identification temporaires aux moteurs d'analyse intégrés.

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "lakeformation.amazonaws.com"
  },
  "Action": [
    "sts:AssumeRole",
    "sts:SetSourceIdentity",
    "sts:SetContext" # add action to trust relationship when using IAM Identity
                    # center principals with Lake Formation
  ]
}
```

Note

Le processus d'intégration des services d' AWS analyse a été mis à jour. Si vous avez configuré l'intégration avec la version préliminaire, vous pouvez continuer à utiliser votre intégration actuelle. Cependant, le processus d'intégration mis à jour permet d'améliorer les performances. Pour mettre à jour l'intégration :

1. Supprimez d'abord votre catalogue de tables S3 existant dans Lake Formation. Pour supprimer le catalogue, sélectionnez-le dans `s3tablescatalog` la liste des catalogues, puis choisissez Supprimer des actions.
2. Ensuite, désenregistrez l'emplacement des données pour. `s3tablescatalog`
 - a. Sur la console Lake Formation, dans la section Administrations, choisissez Data Locations.
 - b. Sélectionnez un emplacement, puis dans le menu Actions, choisissez Supprimer.
 - c. Lorsque vous êtes invité à confirmer, choisissez Supprimer.

Pour obtenir des instructions détaillées sur le désenregistrement d'un emplacement de données, consultez la section. [Annulation de l'enregistrement d'un site Amazon S3](#)

3. Suivez ensuite les étapes d'intégration mises à jour dans la [Activation de l'intégration des tables Amazon S3](#) section.

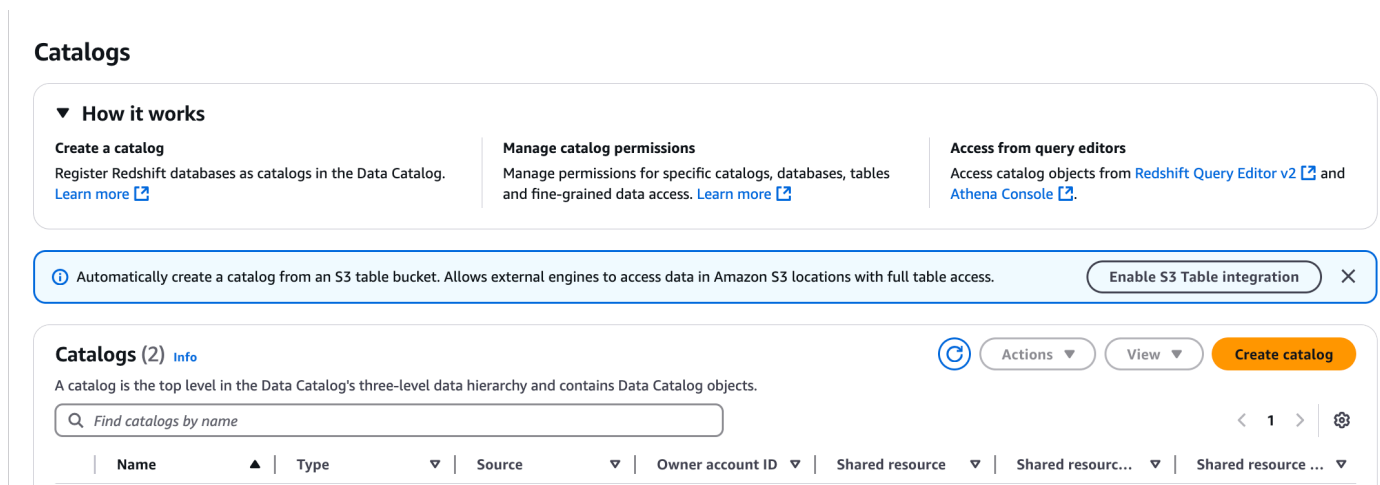
Activation de l'intégration des tables Amazon S3

Vous pouvez créer des compartiments de table Amazon S3 à l'aide de la console Amazon S3 et les intégrer aux services AWS d'analyse. Pour plus d'informations, consultez [Utilisation des tables Amazon S3 avec des services AWS d'analyse](#).

Dans AWS Lake Formation, vous pouvez activer l'intégration d'Amazon S3 Tables avec AWS Glue Data Catalog et AWS Lake Formation à l'aide de la console Lake Formation ou utiliser AWS CLI.

Pour intégrer les tables Amazon S3 au catalogue de données et à Lake Formation (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données.
3. Choisissez Activer l'intégration des tables S3 sur la page Catalogues.



4. Choisissez un rôle IAM doté des autorisations requises pour que Lake Formation puisse assumer la vente d'informations d'identification aux moteurs de requêtes analytiques. Pour connaître les autorisations requises pour que le rôle puisse accéder aux données, consultez [step3-permissions](#) la section sur les conditions préalables.

Enable S3 Table integration

Once integration is enabled, every table bucket in this account and region will automatically be available under the s3tablescatalog catalog in AWS Data Catalog.

Select a principal to register

AWS LakeFormation needs to be able to call S3 Tables APIs on your behalf to retrieve S3 Table buckets, namespace, and tables. S3 tables must be registered with AWS LakeFormation with an IAM role that can be assumed.

Select a role to register

☐ Allow external engines to access data in Amazon S3 locations with full table access

Cancel Enable

i When you register an S3 tables catalog in Lake Formation, the registering principal automatically receives permissions to perform all supported Lake Formation operations on the registered data location.

5. Sélectionnez Autoriser les moteurs externes à accéder aux données dans les emplacements Amazon S3 avec l'option d'accès complet aux tables. Lorsque vous activez l'accès complet aux tables pour des moteurs tiers, Lake Formation renvoie directement les informations d'identification au moteur tiers sans effectuer de validation des balises de session IAM. Cela signifie que vous ne pouvez pas appliquer de contrôles d'accès précis à Lake Formation aux tables auxquelles vous accédez.
6. Sélectionnez Activer. Le nouveau catalogue des tables S3 est ajouté à la liste des catalogues. Lorsque vous activez l'intégration du catalogue de tables S3, le service enregistre l'emplacement des données du compartiment de tables S3 auprès de Lake Formation.
7. Choisissez le catalogue pour afficher les objets du catalogue et accorder des autorisations aux autres principaux.

Success
Successfully created catalog s3tablescatalog.

s3tablescatalog

Actions

Catalog summary

Name	Permissions for newly created tables	Description
s3tablescatalog	-	-

Catalog connection details

Connect to data stored in lakes, warehouses, and other external data sources and publish to unified Iceberg data catalog.

Access for Open Source Engine	IAM role
-	-
Namespace register status	KMS key
Registered to AWS Data Catalog	-

Objects Permissions

Data permissions for catalog s3tablescatalog (1)

View all permissions Revoke Grant

Filter permissions by property or value

Principal	Principal	Principal	Resource	Database	Table	Resource	Catalog	LF-Tag ex...	Permissions	Grantable	RAM
Admin	IAM role	arn:aws:ia...	Catalog	-	-	-	45178558...	-	All, Alter, ...	All, Alter, ...	-

Pour créer des catalogues à plusieurs niveaux, consultez la section [Création d'un compartiment de tables](#) dans le guide de l'utilisateur d'Amazon Simple Storage Service.

Pour intégrer les tables Amazon S3 au catalogue de données et au Lake Formation (CLI)

En suivant la section sur les conditions préalables, créez un rôle de service IAM qui permet à Lake Formation d'accéder aux ressources de votre table.

1. Créez un fichier Role-Trust-Policy.json contenant la stratégie d'approbation suivante :

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "LakeFormationDataAccessPolicy",
      "Effect": "Allow",
      "Principal": {
        "Service": "lakeformation.amazonaws.com"
      },
      "Action": [
        "sts:AssumeRole",
        "sts:SetContext",
        "sts:SetSourceIdentity"
      ]
    }
  ]
}
```

```

    "Condition": {
      "StringEquals": {
        "aws:SourceAccount": "111122223333"
      }
    }
  }
]
}

```

2. Créez le rôle de service IAM à l'aide de la commande suivante :

```

aws iam create-role \
  --role-name S3TablesRoleForLakeFormation \
  --assume-role-policy-document file://Role-Trust-Policy.json

```

3. Créez un fichier `LF-GluePolicy.json` contenant la stratégie suivante :

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "LakeFormationPermissionsForS3ListTableBucket",
      "Effect": "Allow",
      "Action": [
        "s3tables:ListTableBuckets"
      ],
      "Resource": [
        "*"
      ]
    },
    {
      "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
      "Effect": "Allow",
      "Action": [
        "s3tables:CreateTableBucket",
        "s3tables:GetTableBucket",
        "s3tables:CreateNamespace",
        "s3tables:GetNamespace",
        "s3tables:ListNamespaces",
        "s3tables>DeleteNamespace",
        "s3tables>DeleteTableBucket",
        "s3tables:CreateTable",
        "s3tables>DeleteTable",

```

```

        "s3tables:GetTable",
        "s3tables:ListTables",
        "s3tables:RenameTable",
        "s3tables:UpdateTableMetadataLocation",
        "s3tables:GetTableMetadataLocation",
        "s3tables:GetTableData",
        "s3tables:PutTableData"
    ],
    "Resource": [
        "arn:aws:s3tables:us-east-1:111122223333:bucket/*"
    ]
}
]
}

```

4. Attachez la stratégie au rôle en utilisant la commande suivante :

```

aws iam put-role-policy \
  --role-name S3TablesRoleForLakeFormation \
  --policy-name LakeFormationDataAccessPermissionsForS3TableBucket \
  --policy-document file://LF-GluePolicy.json

```

5. Créez un fichier `input.json` contenant ce qui suit :

```

{
  "ResourceArn": "arn:aws:s3tables:us-east-1:111122223333:bucket/*",
  "WithFederation": true,
  "RoleArn": "arn:aws:iam::111122223333:role/S3TablesRoleForLakeFormation"
}

```

6. Enregistrez les compartiments de table auprès de Lake Formation en utilisant la commande suivante :

```

aws lakeformation register-resource \
  --region us-east-1 \
  --with-privileged-access \
  --cli-input-json file://input.json

```

7. Créez un fichier `catalog.json` contenant le catalogue qui suit :

```

{
  "Name": "s3tablescatalog",
  "CatalogInput": {

```



```
"FederatedCatalog": {
  "Identifier": "arn:aws:s3tables:us-east-1:111122223333:bucket/*",
  "ConnectionName": "aws:s3tables"
},
"CreateDatabaseDefaultPermissions": [],
"CreateTableDefaultPermissions": [],
"AllowFullTableExternalDataAccess": "True"
}
```

8. Créez le catalogue `s3tablescatalog` à l'aide de la ligne de commande suivante. La création de ce catalogue remplit le AWS Glue Data Catalog avec des objets correspondant à des compartiments de tables, à des espaces de noms et à des tables.

```
aws glue create-catalog \
  --region us-east-1 \
  --cli-input-json file://catalog.json
```

9. Vérifiez que le `s3tablescatalog` catalogue a été ajouté à AWS Glue l'aide de la commande suivante :

```
aws glue get-catalog --catalog-id s3tablescatalog
```

Création de bases de données et de tables dans le catalogue de tables S3

Vous pouvez créer des bases de données pour organiser vos tables Apache Iceberg, ainsi que des tables pour définir le schéma et l'emplacement de vos données dans le catalogue de tables S3.

Création d'une base de données (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'administrateur de lac de données ou créateur de base de données.
2. Dans le volet de navigation, sélectionnez Bases de données sous Catalogue de données.
3. Choisissez Créer une base de données.
4. Sur la page Créer une base de données, choisissez l'option Base de données et entrez les informations suivantes :
 - Nom : nom unique pour la base de données

- Catalogue de données : choisissez le catalogue de tables S3. La base de données résidera dans ce catalogue.
- Description — (Facultatif) Ajoutez une description et un emplacement.
- Contrôle d'accès IAM pour les nouvelles tables : sélectionnez éventuellement Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données. Pour plus d'informations sur cette option, consultez la section [Modification des paramètres par défaut de votre lac de données](#).
- Choisissez Créer une base de données. Vous pouvez voir la base de données créée dans le catalogue des tables S3.

Créez une base de données à l'aide de AWS CLI

La commande CLI suivante montre comment créer une base de données dans le catalogue de tables S3.

```
aws glue create-database
--region us-east-1 \
--catalog-id "123456789012:s3tablescatalog/test" \
--database-input \
'{ "Name": "testglueclidbcreation" }'
```

Création d'une table (AWS Management Console)

Vous pouvez créer des tables de métadonnées Apache Iceberg dans le catalogue de tables S3 à l'aide de la console Lake Formation ou de l' AWS Glue CreateTableAPI.

1. Ouvrez la console Lake Formation à l'<https://console.aws.amazon.com/lakeformation/>adresse et connectez-vous en tant qu'administrateur de lac de données ou en tant qu'utilisateur CreateTable autorisé.
2. Dans le volet de navigation, sélectionnez Tables sous Catalogue de données.
3. Choisissez Créer un tableau.
4. Sur la page Créer une table, entrez les détails de la table :

Create table [Info](#)

Table details

Create a table in the Data Catalog.

Name

If you plan to access the table from Amazon Athena, then the name should be under 256 characters and contain only lowercase letters (a-z), numbers (0-9), and underscore (_). For more information, see [Athena names](#).

Catalog

Table is contained within this catalog.



[Create a catalog](#)

Database

Table is contained within this database.



[Create database](#)

Table format

☒ Apache Iceberg Table

Create a table in the Apache Iceberg table format

Schema [Info](#)

View and manage table schema.

[Upload schema](#)[Delete](#)[Edit](#)[Add column](#)

< 1 >

No available schema

[Cancel](#)[Submit](#)

- Nom — Entrez un nom unique pour la table.
- Catalogue — Choisissez le catalogue des tables S3 comme catalogue.
- Base de données — Choisissez la base de données dans le catalogue des tables S3.

- Description — Entrez une description pour le tableau.
- Schéma — Choisissez Ajouter des colonnes pour ajouter des colonnes et les types de données des colonnes. Vous avez la possibilité de créer une table vide et de mettre à jour le schéma ultérieurement. Iceberg vous permet de faire évoluer le schéma et la partition après avoir créé la table. Vous pouvez utiliser des requêtes Athena pour mettre à jour le schéma de table et des requêtes Spark pour mettre à jour les partitions.

5. Sélectionnez Soumettre.

Création d'une table (AWS CLI)

```
aws glue create-table \  
--database-name "testglueclidbcreation" \  
--catalog-id "123456789012:s3tablescatalog/test" \  
--region us-east-1 \  
--table-input \  
'{ "Name": "testtablegluecli", "Parameters": { "format": "ICEBERG" },  
  "StorageDescriptor": { "Columns": [ {"Name": "x", "Type": "int", "Parameters":  
    {"required": "true"}} ] } }'
```

Enregistrement d'un compartiment de table Amazon S3 dans un autre AWS compte

Vous pouvez enregistrer des compartiments de table Amazon S3 individuels (format ARN :arn:aws:s3tables:us-east-1:*account-id*:bucket/*bucket-name*) à partir d'un AWS compte auprès de Lake Formation sur un autre compte. Par exemple, vous pouvez enregistrer un seau de tables depuis le compte A dans Lake Formation du compte B.

Conditions préalables

Avant de commencer l'enregistrement du bucket entre comptes :

- Créez un compartiment de table dans le compte A.
- Créez un rôle IAM dans le compte B avec les autorisations appropriées pour l'enregistrement du bucket.

Pour plus d'informations sur les autorisations requises pour enregistrer un bucket de table auprès de Lake Formation, consultez [Conditions préalables à l'intégration du catalogue de tables Amazon S3 au catalogue de données et à Lake Formation](#).

- Enregistrez le compartiment de tables dans le compte sur lequel vous allez créer le catalogue de tables S3.
- Pour l'accès entre comptes, le rôle spécifié lors de l'enregistrement du compartiment de table doit être un rôle intégré au compte avec les autorisations appropriées pour accéder au compartiment entre comptes. Le rôle nécessite les actions IAM S3 Tables nécessaires pour accéder aux ressources du bucket.

Enregistrement d'un bucket de table entre comptes

Dans les procédures suivantes, le compte A est le compte propriétaire de la ressource, et le compte B est l'endroit où le bucket de table sera enregistré pour gérer les autorisations d'accès.

1. Connectez-vous AWS Management Console au compte A.

Ouvrez la console Amazon S3 à l'adresse <https://console.aws.amazon.com/s3/>.

2. Créez un compartiment de tables. Pour plus d'informations, consultez la section [Création d'un compartiment de table](#) dans le guide de l'utilisateur Amazon S3.
3. Enregistrez le bucket de table dans le compte B.

Utilisez le AWS CLI pour enregistrer le seau de table du compte A avec Lake Formation dans le compte B.

```
aws lakeformation register-resource \  
--resource-arn 'arn:aws:s3tables:us-east-1:account-A-id:bucket/single-bucket-name' \  
\  
--role-arn arn:aws:iam::account-B-id:role/role-name \  
--region us-east-1
```

Remplacez account-A-id, single-bucket-name account-B-id, et role-name par vos valeurs spécifiques.

4. Créez ensuite un catalogue pour le compartiment de table dans le compte B.

Créez un catalogue à l'aide du AWS CLI.

```
aws glue create-catalog --region us-east-1 \  
--cli-input-json \  
'{  
  "Name": "catalog-name",  
  "CatalogInput" : {  
    "FederatedCatalog": {  
      "Identifier": "arn:aws:s3tables:us-east-1:account A:bucket/single-bucket-  
name",  
      "ConnectionName": "aws:s3tables"  
    },  
    "CreateDatabaseDefaultPermissions": [],  
    "CreateTableDefaultPermissions": []  
  }  
'
```

Remplacez *catalog-name* et *bucket-name* par vos valeurs spécifiques. *account-A-id*

L'exemple de CLI suivant montre comment afficher les détails du catalogue.

```
aws glue get-catalog \  
--catalog-id account-id:catalog-name \  
--region us-east-1
```

5. Ensuite, créez des bases de données et des tables dans le catalogue nouvellement créé dans le compte B.

Créez une base de données.

```
aws glue create-database \  
--region us-east-1 \  
--catalog-id "account-B-id:catalog-name" \  
--database-input \  
'{  
  "Name": "database-name"  
'
```

Créer une table.

```
aws glue create-table \  
--database-name database-name \  

```

```
--catalog-id account-B-id:catalog-name \  
--region us-east-1 \  
--table-input \  
{  
  "Name": "table-name",  
  "Parameters": {  
    "format": "ICEBERG"  
  },  
  "StorageDescriptor": {  
    "Columns": [  
      {"Name": "x", "Type": "int", "Parameters": {"required": "true"}}  
    ]  
  }  
}
```

Remplacez database-name account-B-id, catalog-name et table-name par vos valeurs spécifiques.

6. Les exemples suivants montrent comment afficher les objets du catalogue.

Afficher les détails de la base de données.

```
aws glue get-database \  
--name database-name \  
--catalog-id account-B-id:catalog-name \  
--region us-east-1
```

Afficher les détails du tableau.

```
aws glue get-table \  
--name table-name \  
--database-name database-name \  
--catalog-id account-B-id:catalog-name \  
--region us-east-1
```

Remplacez database-name account-B-id, catalog-name et table-name par vos valeurs spécifiques.

Octroi d'autorisations

Après avoir intégré vos tables S3 AWS Lake Formation, vous pouvez accorder des autorisations sur le catalogue des tables S3 et les objets du catalogue (compartiments de tables, bases de données, tables) à d'autres rôles IAM et utilisateurs de votre compte. Les autorisations de Lake Formation vous permettent de définir des contrôles d'accès au niveau de la granularité au niveau des tables, des colonnes et des lignes pour les utilisateurs de moteurs d'analyse intégrés tels qu'Amazon Redshift Spectrum et Athena.

Vous pouvez accorder des autorisations en utilisant la méthode des ressources nommées ou la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC). Avant d'accorder des autorisations à l'aide de balises LF et d'expressions de balises LF, vous devez les définir et les affecter aux objets du catalogue de données.

Pour de plus amples informations, veuillez consulter [Gestion des balises LF pour le contrôle d'accès aux métadonnées](#).

Vous pouvez partager des bases de données et des tables avec AWS des comptes externes en accordant des autorisations à Lake Formation aux comptes externes. Les utilisateurs peuvent ensuite exécuter des requêtes et des tâches qui joignent et interrogent des tables sur plusieurs comptes. Lorsque vous partagez une ressource de catalogue avec un autre compte, les principaux titulaires de ce compte peuvent utiliser cette ressource comme si elle figurait dans leur catalogue de données.

Lorsque vous partagez des bases de données et des tables avec des comptes externes, l'autorisation de super utilisateur n'est pas disponible.

Pour obtenir des instructions détaillées sur l'octroi d'autorisations, consultez la [Gestion des autorisations relatives à Lake Formation](#) section.

AWS CLI exemple d'octroi d'autorisations sur une table Amazon S3

```
aws lakeformation grant-permissions \
--cli-input-json \
'{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:role/DataAnalystRole"
  },
  "Resource": {
    "Table": {
      "CatalogId": "111122223333:s3tablescatalog/amzn-s3-demo-bucket1",
      "DatabaseName": "S3 table bucket namespace <example_namespace>",
```



```

        "Name": "S3 table bucket table name <example_table>"
    },
    "Permissions": [
        "SELECT"
    ]
}

```

Les paramètres à inclure dans la commande sont les suivants :

- **DataLakePrincipalIdentifier** — ARN d'utilisateur, de rôle ou de groupe IAM pour accorder des autorisations
- **CatalogId** — Identifiant de AWS compte à 12 chiffres auquel appartient le catalogue de données
- **DatabaseName** — Nom de l'espace de nom du compartiment de table Amazon S3
- **Nom** : nom de la table de compartiment de table Amazon S3
- **Autorisations** : autorisations à accorder. Les options incluent : SELECT, INSERT, DELETE, DESCRIBE, ALTER, DROP, ALL et SUPER

Accès aux tables Amazon S3 partagées

Après avoir accordé des autorisations entre comptes sur une base de données ou une table du catalogue de tables S3, pour accéder aux ressources, vous devez créer des liens de ressources vers les bases de données et les tables partagées.

1. Dans le compte de destination (le compte qui reçoit les ressources partagées), créez un lien vers les ressources de base de données. Pour obtenir des instructions complètes, consultez [Création d'un lien de ressource vers une base de données de catalogue de données partagée](#).

Exemple de CLI pour créer un lien vers une ressource de base de données

```

aws glue create-database
--region us-east-1
--catalog-id "111122223333"
--database-input \
'{
  "Name": "s3table_resourcelink",
  "TargetDatabase": {
    "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-bucket-011426214932",

```

```

    "DatabaseName": "s3_table_ns"
  },
  "CreateTableDefaultPermissions": []
}'

```

2. Accordez une autorisation entre comptes sur la table.

Exemple de CLI pour l'octroi d'autorisations entre comptes

```

aws lakeformation grant-permissions \
--region us-east-1 \
--cli-input-json \
'{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:role/
S3TablesTestExecRole"
  },
  "Resource": {
    "Table": {
      "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-
bucket-011426214932",
      "DatabaseName": "s3_table_ns",
      "Name": "test_s3_iceberg_table"
    }
  },
  "Permissions": [
    "ALL"
  ]
}'

```

3. DESCRIBE Autorisation de Grant Lake Formation sur le lien vers la ressource.

Exemple de CLI pour accorder l'autorisation de description sur le lien de ressource.

```

aws lakeformation grant-permissions \
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/
S3TablesTestExecRole
--resource Database='{CatalogId=111122223333;, Name=s3table_resourcelink}' \
--permissions DESCRIBE

```

Modification des contrôles d'accès pour l'intégration des tables S3

Après avoir intégré les tables Amazon S3 au AWS Glue Data Catalog, vous pouvez modifier la façon dont l'accès aux ressources de votre catalogue est contrôlé. Cette section explique comment modifier le contrôle d'accès en fonction du modèle de contrôle d'accès actuel et souhaité. L'activation de Lake Formation vous permet d'utiliser des autorisations détaillées telles que la sécurité au niveau des colonnes et au niveau des lignes grâce aux subventions de Lake Formation, et permet à Lake Formation de vendre des informations d'identification temporaires au nom des principaux via un rôle enregistré. Le passage du contrôle d'accès AWS Lake Formation à IAM ramène le contrôle d'accès aux politiques IAM standard, ce qui peut être approprié si vos charges de travail ne nécessitent pas un accès précis et si vous préférez gérer les autorisations entièrement via IAM. Les deux voies de migration impliquent la mise à jour des valeurs par défaut du catalogue de données, l'ajustement des enregistrements des ressources auprès de Lake Formation et la coordination des autorisations accordées pour éviter les interruptions d'accès pendant la transition.

Rubriques

- [Activez Lake Formation grâce à l'intégration des tables S3 au catalogue de données](#)
- [Passer du contrôle d'accès AWS Lake Formation à IAM](#)

Activez Lake Formation grâce à l'intégration des tables S3 au catalogue de données

Cette section décrit le flux de travail permettant de migrer le contrôle d'accès des privilèges IAM vers IAM avec AWS Lake Formation des autorisations pour les tables Amazon S3 intégrées au. AWS Glue Data Catalog

Important

AWS Lake Formation L'activation du contrôle d'accès révoquera tous les accès existants basés sur IAM à vos ressources S3 Tables. Une fois l'étape 1 terminée, les utilisateurs et les rôles qui avaient précédemment accédé aux données via des autorisations IAM en perdront immédiatement l'accès. Vous devez accorder les autorisations de Lake Formation à l'étape 2 avant que les utilisateurs puissent à nouveau interroger des données. Planifiez cette migration pendant une période de maintenance et coordonnez-la avec votre équipe chargée des données.

Conditions préalables

Pour read/write accéder aux tables S3, en plus des autorisations Lake Formation, les directeurs doivent également disposer de l'autorisation `lakeformation:GetDataAccess` IAM. Avec cette autorisation, Lake Formation accède à la demande d'informations d'identification temporaires pour accéder aux données.

En utilisant AWS CLI

1. Étape 1 : enregistrer un bucket auprès de Lake Formation à l'aide du rôle IAM

Enregistrez la ressource S3 Tables auprès de Lake Formation.

Note

Si vous avez déjà un rôle, assurez-vous que l'accès hybride est faux.

```
aws lakeformation register-resource \  
  --resource-arn "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*" \  
  --role-arn "arn:aws:iam::AWSAccountID:role/service-role/LFAccessRole" \  
  --with-federation
```

2. Étape 2 : Mettre à jour le AWS Glue catalogue pour activer le contrôle d'accès à Lake Formation

Mettez à jour le catalogue avec un `CreateDatabaseDefaultPermissions` champ vide et `CreateTableDefaultPermissions` (défini sur[]) et défini `OverwriteChildResourcePermissionsWithDefault` sur `Accept`. Cela supprime l'accès basé sur l'IAM à toutes les ressources enfants existantes et permet de gérer le catalogue et ses objets à l'aide des subventions de Lake Formation.

```
aws glue update-catalog \  
  --catalog-id "s3tablescatalog" \  
  --catalog-input '{  
    "FederatedCatalog": {  
      "Identifier": "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*",  
      "ConnectionName": "aws:s3tables"  
    },  
    "CreateDatabaseDefaultPermissions": [],  
    "CreateTableDefaultPermissions": [],  
    "OverwriteChildResourcePermissionsWithDefault": "Accept",  
  }
```

```
"AllowFullTableExternalDataAccess": "True"
}'
```

3. Étape 3 : Accordez les autorisations de Lake Formation à votre équipe chargée des données

Accordez des autorisations de Lake Formation aux principaux (rôles, utilisateurs ou groupes) qui ont besoin d'un accès. Par exemple, pour accorder à un rôle un accès en lecture complète au tableau :

```
aws lakeformation grant-permissions \
  --principal
  DataLakePrincipalIdentifier=arn:aws:iam::AWSAccountID:role/DataTeamRole \
  --resource '{
    "Table": {
      "CatalogId": "AWSAccountID",
      "DatabaseName": "s3tablescatalog/table-bucket-name/namespace",
      "TableWildcard": {}
    }
  }' \
  --permissions "SELECT" "DESCRIBE"
```

Répétez l'opération pour chaque combinaison principale et ressource selon les besoins.

Passer du contrôle d'accès AWS Lake Formation à IAM

Cette section décrit le flux de travail permettant de passer du contrôle d'accès AWS Lake Formation des autorisations aux privilèges IAM pour les tables Amazon S3 intégrées au AWS Glue Data Catalog.

Important

La modification du contrôle d'accès AWS Lake Formation des autorisations à IAM révoquera tous les accès existants basés sur Lake Formation à vos ressources S3 Tables. Une fois l'étape 2 terminée, les utilisateurs et les rôles qui accédaient auparavant aux données par le biais des subventions de Lake Formation perdront immédiatement leur accès. Vous devez accorder l'accès IAM à l'étape 1 avant de mettre à jour le catalogue. Planifiez cette migration pendant une période de maintenance et coordonnez-la avec votre équipe chargée des données.

Important

Les contrôles d'accès précis, tels que l'accès au niveau des colonnes et les filtres de cellules de données, avec les objets du catalogue de données sont disponibles uniquement en cas d'utilisation. AWS Lake Formation Avant de procéder à la migration des contrôles d'accès AWS Lake Formation vers IAM, vérifiez vos autorisations Lake Formation existantes à l'aide de politiques IAM équivalentes `aws lakeformation list-permissions` et déterminez si des politiques IAM équivalentes peuvent fournir l'accès dont vos utilisateurs ont besoin. Tout directeur qui s'est appuyé sur des subventions détaillées de Lake Formation aura besoin d'un accès IAM complet au niveau de la table après la migration du contrôle d'accès.

Conditions préalables

Avant de commencer, assurez-vous de ce qui suit :

- Vous avez identifié toutes les subventions de Lake Formation actuellement en vigueur pour les ressources en cours de migration. Courez `aws lakeformation list-permissions --resource-type TABLE` pour les examiner.
- Vous avez préparé des politiques IAM qui fournissent un accès équivalent à tous les principaux concernés.
- Le rôle IAM enregistré auprès de Lake Formation existe toujours `lakeformation:GetDataAccess` (nécessaire pendant la période de transition hybride).

En utilisant AWS CLI

1. Étape 1 : octroyer des autorisations IAM aux principaux

Associez des politiques IAM aux utilisateurs ou aux rôles qui ont besoin d'un accès. La politique doit inclure à la fois les autorisations AWS Glue relatives aux métadonnées et les autorisations relatives aux données des tables S3.

Note

L'exemple de politique suivant fournit uniquement un accès en lecture.

```

aws iam put-user-policy \
  --user-name GlueIAMAccessUser \
  --policy-name S3TablesIAMAccessPolicy \
  --policy-document '{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Sid": "GlueMetadataAccess",
        "Effect": "Allow",
        "Action": [
          "glue:GetCatalog",
          "glue:GetDatabase",
          "glue:GetTable"
        ],
        "Resource": [
          "arn:aws:glue:us-east-1:AWSAccountID:catalog/s3tablescatalog",
          "arn:aws:glue:us-east-1:AWSAccountID:database/
s3tablescatalog/table-bucket-name/namespace",
          "arn:aws:glue:us-east-1:AWSAccountID:table/s3tablescatalog/table-
bucket-name/namespace/*"
        ]
      },
      {
        "Sid": "S3TablesDataAccess",
        "Effect": "Allow",
        "Action": [
          "s3tables:GetTableBucket",
          "s3tables:GetTable",
          "s3tables:GetTableMetadataLocation",
          "s3tables:GetTableData"
        ],
        "Resource": [
          "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/table-bucket-name",
          "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/table-bucket-name/
table/*"
        ]
      }
    ]
  }'

```

Vérifiez que tous les utilisateurs et rôles concernés peuvent accéder aux tables attendues à l'aide de leurs informations d'identification IAM avant de continuer.

2. Étape 2 : mettre à jour le catalogue pour restaurer les autorisations par défaut d'IAM

Mettez à jour le catalogue en conséquence `CreateDatabaseDefaultPermissions` et `CreateTableDefaultPermissions` ALL accordez-le à `IAM_ALLOWED_PRINCIPALS`. Définissez cette `OverwriteChildResourcePermissionsWithDefault` option pour `Accept` que la modification se propage à toutes les ressources enfants existantes, et pas seulement à celles qui viennent d'être créées.

```
aws glue update-catalog \  
  --catalog-id "s3tablescatalog" \  
  --catalog-input '{  
    "FederatedCatalog": {  
      "Identifier": "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*",  
      "ConnectionName": "aws:s3tables"  
    },  
    "CreateDatabaseDefaultPermissions": [{  
      "Principal": {"DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"},  
      "Permissions": ["ALL"]  
    }],  
    "CreateTableDefaultPermissions": [{  
      "Principal": {"DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"},  
      "Permissions": ["ALL"]  
    }],  
    "OverwriteChildResourcePermissionsWithDefault": "Accept"  
  }'
```

3. Étape 3 : Désenregistrer la ressource de Lake Formation

Une fois que vous avez confirmé que tous les accès fonctionnent conformément aux politiques IAM et qu'aucun principe ne dépend des subventions de Lake Formation, vous pouvez annuler l'enregistrement de la ressource auprès de Lake Formation pour terminer la migration.

```
aws lakeformation deregister-resource \  
  --resource-arn "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*"
```


 Note

Après avoir annulé l'enregistrement de la ressource, supprimez `lakeformation:GetDataAccess` de l'IAM les principaux qui n'en ont plus besoin.

Aucune `revoke-permissions` étape n'est requise.

Création d'un catalogue géré par Amazon Redshift dans AWS Glue Data Catalog

Vous n'avez peut-être pas de cluster de producteurs Amazon Redshift ou de partage de données Amazon Redshift disponibles aujourd'hui, mais vous souhaitez créer et gérer des tables Amazon Redshift à l'aide du. AWS Glue Data Catalog Vous pouvez commencer par créer un catalogue AWS Glue géré à l'aide de `glue:CreateCatalogAPI` ou de la AWS Lake Formation console en définissant le type de catalogue `Catalog source` comme `Redshift. Managed` Cette étape permet d'effectuer les opérations suivantes :

- Crée un catalogue dans le catalogue de données
- Enregistre le catalogue en tant que localisation des données de Lake Formation
- crée un groupe de travail sans serveur géré par Amazon Redshift
- Lie le groupe de travail sans serveur Amazon Redshift et le catalogue de données à l'aide d'un objet de partage de données

Pour créer un catalogue géré et configurer des autorisations (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données.
3. Sélectionnez l'option Créer un catalogue.
4. Sur la page de détails du catalogue, entrez les informations suivantes :
 - Nom : nom unique pour votre catalogue géré. Le nom ne peut pas être modifié et doit être en minuscules. Le nom peut comporter un maximum de 255 caractères. compte.
 - Type — Choisissez `Managed catalog` comme type de catalogue.

- Stockage — Choisissez Redshift le stockage.
 - Description — Entrez une description pour le catalogue créé à partir de la source de données.
5. Vous pouvez utiliser les applications Apache Spark exécutées sur Amazon EMR sur Amazon EC2 pour accéder aux bases de données Amazon Redshift dans le AWS Glue Data Catalog

Pour permettre à Apache Spark de lire et d'écrire sur le stockage géré Amazon Redshift, AWS Glue crée un cluster Amazon Redshift géré avec les ressources de calcul et de stockage nécessaires pour effectuer des opérations de lecture et d'écriture sans impact sur les charges de travail de l'entrepôt de données Amazon Redshift. Vous devez également fournir un rôle IAM avec les autorisations requises pour transférer des données vers et depuis le compartiment Amazon S3. Pour connaître les autorisations requises pour le rôle de transfert de données, consultez l'étape 5 de la [Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog](#) section.

6. Par défaut, les données du cluster Amazon Redshift sont chiffrées à l'aide d'une clé AWS gérée. Lake Formation propose une option permettant de créer votre clé KMS personnalisée pour le chiffrement. Si vous utilisez une clé gérée par le client, vous devez ajouter des politiques clés spécifiques à la clé.
7. Choisissez l'option Personnaliser les paramètres de chiffrement si vous utilisez une clé gérée par le client pour chiffrer les données du stockage géré Amazon Redshift. Pour utiliser une clé personnalisée, vous devez ajouter une stratégie de clé gérée personnalisée supplémentaire à votre clé KMS. Pour de plus amples informations, veuillez consulter [Conditions préalables à la gestion des espaces de noms Amazon Redshift dans AWS Glue Data Catalog](#).
8. Options de chiffrement — Choisissez l'option Personnaliser les paramètres de chiffrement si vous souhaitez utiliser une clé personnalisée pour chiffrer le catalogue. Pour utiliser une clé personnalisée, vous devez ajouter une stratégie de clé gérée personnalisée supplémentaire à votre clé KMS.
9. Choisissez Next pour accorder des autorisations aux autres principaux.
10. Sur la page Accorder des autorisations, choisissez Ajouter des autorisations.
11. Sur l'écran Ajouter des autorisations, choisissez les principes et les types d'autorisations à accorder.

Add permissions



Principals

Choose the principals to grant permissions.



IAM users and roles

Users or roles from this AWS account.



SAML users and groups

SAML users and group or QuickSight ARNs.



External accounts

AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add



UserRole X
Role

Catalog permissions

Choose the permissions to grant on the catalog. Choosing Super user overwrites individual permissions, granting unrestricted administrative access.

☐ Super user

A super user has unrestricted administrative privileges to perform any operation on all resources within the catalog (databases, tables, and views).

Catalog permissions

Choose specific access permissions to grant.



Create
database



Describe



Alter



Drop



Super

This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions

Choose the permission that can be granted to others.



Create
database



Describe



Alter



Drop



Super

This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel

Add

- Dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels octroyer des autorisations.

- Utilisateurs et rôles IAM : choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.
- Utilisateurs et groupes SAML : pour le SAML et Amazon Quick les utilisateurs et groupes, entrez un ou plusieurs Amazon Resource Names (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou pour les utilisateurs ou groupes ARNs Amazon Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de créer les ARNs, voir les AWS CLI commandes AWS CLI d'octroi et de révocation.

- Dans la section Autorisations, sélectionnez les autorisations et les autorisations octroyables.

Sous Autorisations du catalogue, sélectionnez une ou plusieurs autorisations à octroyer.

Choisissez Super utilisateur pour octroyer des autorisations administratives illimitées sur toutes les ressources du catalogue.

Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS Cette option n'est pas prise en charge lorsque vous octroyez des autorisations à un principal IAM à partir d'un compte externe.

12. Choisissez Next pour consulter les informations et créer le catalogue. La liste des catalogues affiche le nouveau catalogue géré.

Pour créer un catalogue fédéré (CLI)

- L'exemple suivant montre comment créer un catalogue fédéré.

```
aws glue create-catalog --cli-input-json file://input.json

{
  "Name": "CatalogName",
  "CatalogInput": {
    "Description": "Redshift published Catalog",
    "CreateDatabaseDefaultPermissions" : [],
    "CreateTableDefaultPermissions": [],
    "CatalogProperties": {
      "DataLakeAccessProperties" : {
        "DataLakeAccess" : "true",
        "DataTransferRole" : "DTR arn",
```

```

        "KMSKey": "kms key arn", // Optional
        "CatalogType": "aws:redshift"
    }
}
}
}

```

Réponse de Glue Get-Catalog

```

aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1

Response:
{
  "Catalog": {
    "Name": "CatalogName",
    "Description": "Glue Catalog for Redshift z-etl use case",
    "CreateDatabaseDefaultPermissions" : [],
    "CreateTableDefaultPermissions": [],
    "CatalogProperties": {
      "DataLakeAccessProperties" : {
        "DataLakeAccess": "true",
        "DataTransferRole": "DTR arn",
        "KMSKey": "kms key arn",
        "ManagedWorkgroupName": "MWG name",
        "ManagedWorkgroupStatus": "MWG status",
        "RedshiftDatabaseName": "RS db name",
        "NamespaceArn": "namespace key arn",
        "CatalogType": "aws:redshift"
      }
    }
  }
}

```

Gestion des autorisations pour les données dans un partage de données Amazon Redshift

Avec AWS Lake Formation, vous pouvez gérer les données en toute sécurité dans un partage de données d'Amazon Redshift. Amazon Redshift est un service d'entrepôt de données entièrement

géré de plusieurs pétaoctets dans le cloud. AWS Grâce à la fonctionnalité de partage de données, Amazon Redshift vous permet de partager des données entre différentes entités. Comptes AWS Pour plus d'informations sur le partage de données Amazon Redshift, consultez [Présentation du partage de données dans Amazon Redshift](#).

Dans Amazon Redshift, l'administrateur du cluster de producteurs crée un partage de données et le partage avec l'administrateur du lac de données. Pour step-by-step obtenir des instructions sur la création d'un administrateur de lac de données, consultez [Création d'un administrateur de lac de données](#).

Une fois que vous (administrateur du lac de données) avez accepté le partage de données, vous devez créer une AWS Glue Data Catalog base de données pour le partage de données spécifique. Cela vous permet de contrôler l'accès à celui-ci à l'aide des autorisations de Lake Formation. Lake Formation fait correspondre chaque partage de données à une base de données du catalogue de données correspondante. Elles apparaissent sous forme de bases de données fédérées dans le catalogue de données.

Une base de données est qualifiée de base de données fédérée lorsqu'elle pointe vers une entité extérieure au catalogue de données. Les tables et les vues du partage de données Amazon Redshift sont répertoriées sous forme de tables individuelles dans le catalogue de données. Vous pouvez partager la base de données fédérée avec des principaux IAM et des utilisateurs SAML sélectionnés au sein du même compte ou d'un autre compte avec Lake Formation. Vous pouvez également inclure des expressions de filtre de ligne et de colonne pour restreindre l'accès à certaines données. Pour de plus amples informations, veuillez consulter [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#).

Pour permettre aux utilisateurs d'accéder à un partage de données Amazon Redshift, vous devez effectuer les opérations suivantes :

1. Mettez à jour les paramètres du catalogue de données pour activer les autorisations de Lake Formation.
2. Acceptez l'invitation de partage de données envoyée par l'administrateur du cluster de producteurs Amazon Redshift et enregistrez le partage de données dans Lake Formation.

Une fois cette étape terminée, vous pouvez gérer le partage de données dans le Lake Formation Data Catalog.

3. Créez une base de données fédérée et définissez les autorisations sur cette base de données.

4. Accordez des autorisations aux utilisateurs sur les bases de données et les tables. Vous pouvez partager l'intégralité de la base de données ou un sous-ensemble de tables avec des utilisateurs du même compte ou d'un autre compte.

Pour connaître les limitations, veuillez consulter [Limites du partage de données Amazon Redshift](#).

Rubriques

- [Conditions préalables à la configuration des autorisations sur les partages de données Amazon Redshift](#)
- [Configuration des autorisations pour les partages de données Amazon Redshift](#)
- [Interrogation de bases de données fédérées](#)

Conditions préalables à la configuration des autorisations sur les partages de données Amazon Redshift

Mettre à jour les paramètres par défaut du catalogue de données

Pour activer les autorisations Lake Formation pour les ressources du catalogue de données, nous vous recommandons de désactiver les paramètres par défaut du catalogue de données dans Lake Formation. Pour de plus amples informations, veuillez consulter [Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride](#).

Mettre à jour les autorisations

Outre les autorisations d'administrateur du lac de données (AWSLakeFormationDataAdmin), les autorisations suivantes sont également requises pour accepter un partage de données Amazon Redshift dans Lake Formation :

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`
- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

L'utilisateur IAM administrateur du lac de données dispose implicitement des autorisations suivantes.

- `accès_location_aux données`

- créer_une base de données
- Formation lacustre : Enregistrer une ressource

Configuration des autorisations pour les partages de données Amazon Redshift

Cette rubrique décrit les étapes à suivre pour accepter une invitation de partage de données, créer une base de données fédérée et accorder des autorisations. Vous pouvez utiliser la console Lake Formation ou le AWS Command Line Interface (AWS CLI). Les exemples présentés dans cette rubrique montrent le cluster de producteurs, le catalogue de données et le consommateur de données dans le même compte.

Pour en savoir plus sur les fonctionnalités multicomptes de Lake Formation, voir [Partage de données entre comptes dans Lake Formation](#).

Pour configurer des autorisations pour un partage de données

1. Passez en revue une invitation de partage de données et acceptez-la.

Console

1. Connectez-vous à la console Lake Formation en tant qu'administrateur de lac de données à l'adresse <https://console.aws.amazon.com/lakeformation/>. Accédez à la page Partage de données.
2. Vérifiez les partages de données auxquels vous êtes autorisé à accéder. La colonne État indique votre statut de participation actuel au partage de données. Le statut En attente indique que vous avez été ajouté à un partage de données, mais que vous ne l'avez pas encore accepté ou que vous avez rejeté l'invitation.
3. Pour répondre à une invitation de partage de données, sélectionnez le nom du partage de données et choisissez Vérifier l'invitation. Dans Accepter ou rejeter le partage de données, passez en revue les détails de l'invitation. Choisissez Accepter pour accepter l'invitation ou Refuser pour refuser l'invitation. Vous n'avez pas accès au partage de données si vous rejetez l'invitation.

AWS CLI

Les exemples suivants montrent comment afficher, accepter et enregistrer l'invitation. Remplacez l' Compte AWS identifiant par un Compte AWS identifiant valide. Remplacez le data-share-arn par le véritable Amazon Resource Name (ARN) qui fait référence au partage de données.

1. Afficher une invitation en attente.

```
aws redshift describe-data-shares \  
  --data-share-arn 'arn:aws:redshift:us-  
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/  
federatedds' \  

```

2. Acceptez un partage de données.

```
aws redshift associate-data-share-consumer \  
  --data-share-arn 'arn:aws:redshift:us-  
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/  
federatedds' \  
  --consumer-arn 'arn:aws:glue:us-east-1:111122223333:catalog  

```

3. Enregistrez le partage de données dans le compte Lake Formation. Utilisez l'opération [RegisterResourceAPI](#) pour enregistrer le partage de données dans Lake Formation. DataShareArnest le paramètre d'entrée pourResourceArn.

Note

Il s'agit d'une étape obligatoire.

```
aws lakeformation register-resource \  
  --resource-arn 'arn:aws:redshift:us-  
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/  
federatedds' \  

```

2. Créez une base de données.

Après avoir accepté une invitation de partage de données, vous devez créer une base de données qui pointe vers la base de données Amazon Redshift associée au partage de données. Vous devez être administrateur de data lake pour créer une base de données.

Console

1. Sélectionnez le partage de données dans le volet Invitations et choisissez Définir les détails de la base de données.
2. Dans Définir les détails de la base de données, entrez un nom et un identifiant uniques pour le partage de données. Vous utilisez cet identifiant pour mapper le partage de données en interne dans la hiérarchie des métadonnées (DBName.Schema.Table).
3. Choisissez Next pour accorder des autorisations à d'autres utilisateurs sur la base de données et les tables partagées.

AWS CLI

Utilisez l'exemple de code suivant pour créer une base de données qui pointe vers la base de données Amazon Redshift partagée avec Lake Formation à l'aide du AWS CLI

```
aws glue create-database --cli-input-json \  
  
'{  
  "CatalogId": "111122223333",  
  "DatabaseInput": {  
    "Name": "tahoedb",  
    "FederatedDatabase": {  
      "Identifier": "arn:aws:redshift:us-  
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/federatedds",  
      "ConnectionName": "aws:redshift"  
    }  
  }  
}'
```

3. Octroyez les autorisations.

Après avoir créé la base de données, vous pouvez accorder des autorisations aux utilisateurs de votre compte ou à Comptes AWS des organisations externes. Vous ne pourrez pas accorder d'autorisations d'écriture (insertion, suppression) et de métadonnées (modification, suppression, création) sur la base de données fédérée mappée à un partage de données Amazon Redshift.

Pour plus d'informations sur l'octroi d'autorisations, consultez [Gestion des autorisations relatives à Lake Formation](#).

Note

En tant qu'administrateur de data lake, vous pouvez uniquement consulter les tables des bases de données fédérées. Pour effectuer toute autre action, vous devez vous accorder davantage d'autorisations sur ces tables.

Console

1. Sur l'écran Accorder des autorisations, sélectionnez les utilisateurs auxquels vous souhaitez accorder des autorisations.
2. Choisissez Accorder.

AWS CLI

Utilisez les exemples suivants pour accorder des autorisations de base de données et de table à l'aide de AWS CLI :

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/non-admin"
  },
  "Resource": {
    "Database": {
      "CatalogId": "111122223333",
      "Name": "tahoedb"
    }
  },
  "Permissions": [
    "DESCRIBE"
  ],
  "PermissionsWithGrantOption": [
  ]
}
```

```
}
```

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
    "Principal": {
        "DataLakePrincipalIdentifier":
"arn:aws:iam::111122223333:user/non-admin"
    },
    "Resource": {
        "Table": {
            "CatalogId": "111122223333",
            "DatabaseName": "tahoedb",
            "Name": "public.customer"
        }
    },
    "Permissions": [
        "SELECT"
    ],
    "PermissionsWithGrantOption": [
        "SELECT"
    ]
}
```

Interrogation de bases de données fédérées

Après avoir accordé les autorisations, les utilisateurs peuvent se connecter et commencer à interroger la base de données fédérée à l'aide d'Amazon Redshift. Les utilisateurs peuvent désormais utiliser le nom de la base de données locale pour faire référence au partage de données Amazon Redshift dans les requêtes SQL. Dans Amazon Redshift, la table client du schéma public partagé via le partage de données sera associée à une table correspondante créée comme `public.customer` dans le catalogue de données.

1. Avant d'interroger la base de données fédérée à l'aide d'Amazon Redshift, l'administrateur du cluster crée une base de données à partir de la base de données du catalogue de données à l'aide de la commande suivante :

```
CREATE DATABASE sharedcustomerdb FROM ARN  
'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA  
tahoedb
```

2. L'administrateur du cluster accorde des autorisations d'utilisation sur la base de données.

```
GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
```

3. Vous (l'utilisateur fédéré) pouvez désormais vous connecter aux outils SQL pour interroger la table.

```
Select * from sharedcustomerdb.public.customer limit 10;
```

Pour plus d'informations, consultez la section [Requête AWS Glue Data Catalog dans](#) le guide de gestion Amazon Redshift.

Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes

Grâce à la fédération des AWS Glue Data Catalog métadonnées (fédération du catalogue de données), vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation Il n'est pas nécessaire de migrer les métadonnées du métastore externe vers le catalogue de données.

Le catalogue de données fournit un référentiel de métadonnées centralisé qui facilite la gestion et la découverte de données sur des systèmes disparates. Lorsque votre organisation gère les données du catalogue de données, vous pouvez les utiliser AWS Lake Formation pour contrôler l'accès à vos ensembles de données dans Amazon S3.

Note

Actuellement, nous prenons uniquement en charge la fédération de métastores Apache Hive (version 3 et supérieure).

Pour configurer la fédération de catalogues de données, nous fournissons une application AWS Serverless Application Model (AWS SAM) appelée [GlueDataCatalogFederation- HiveMetastore](#) dans le AWS Serverless Application Repository.

L'implémentation de référence est fournie GitHub sous forme de projet open source sur [AWS Glue Data Catalog Federation - Hive Metastore](#).

L' AWS SAM application crée et déploie les ressources suivantes qui sont nécessaires pour connecter le catalogue de données au métastore Hive :

- Une AWS Lambda fonction : héberge l'implémentation du service de fédération qui communique entre le catalogue de données et le métastore Hive. AWS Glue invoque cette fonction Lambda pour récupérer des objets de métadonnées depuis le métastore Hive.
- Amazon API Gateway— Le point de connexion de votre métastore Hive qui agit comme un proxy pour acheminer toutes les invocations vers la fonction Lambda.
- Rôle IAM : rôle doté des autorisations nécessaires pour créer la connexion entre le catalogue de données et le métastore Hive.
- AWS Glue connexion — Amazon API Gateway Type de AWS Glue connexion qui stocke le Amazon API Gateway point de terminaison et un rôle IAM pour l'appeler.

Lorsque vous interrogez des tables, le AWS Glue service effectue un appel d'exécution vers le métastore Hive et récupère les métadonnées. La fonction Lambda agit comme un traducteur entre le métastore Hive et le catalogue de données.

Après avoir établi la connexion, afin de synchroniser les métadonnées du métastore Hive avec le catalogue de données, vous devez créer une base de données fédérée dans le catalogue de données à l'aide des détails de connexion du métastore Hive, et mapper cette base de données à la base de données Hive. Une base de données est qualifiée de base de données fédérée lorsqu'elle pointe vers une entité extérieure au catalogue de données.

Vous pouvez appliquer les autorisations de Lake Formation à l'aide du contrôle d'accès basé sur des balises et de la méthode des ressources nommées sur la base de données fédérée Comptes AWS AWS Organizations, et les partager entre plusieurs unités organisationnelles (OU). Vous pouvez également partager la base de données fédérée directement avec les principaux IAM depuis un autre compte.

Vous pouvez définir des autorisations détaillées au niveau des colonnes, des lignes et des cellules à l'aide des filtres de données Lake Formation sur les tables Hive externes. Vous pouvez utiliser

Amazon Athena, Amazon Redshift ou Amazon EMR pour interroger les tables Hive externes gérées par Lake Formation.

Pour plus d'informations sur le partage de données entre comptes et le filtrage des données, voir :

- [Partage de données entre comptes dans Lake Formation](#)
- [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#)

Étapes générales de fédération des métadonnées du catalogue de données

1. Vous créez des utilisateurs et des rôles IAM dotés des autorisations appropriées pour déployer l' AWS SAM application et créer des bases de données fédérées.
2. Vous enregistrez l'emplacement des données Amazon S3 auprès de Lake Formation en sélectionnant l'Enable Data Catalog federation option pour les ensembles de données qui utilisent un métastore Hive externe.
3. Vous configurez les paramètres de l' AWS SAM application (nom de AWS Glue connexion, URL du métastore Hive et paramètres de la fonction Lambda) et déployez l'application. AWS SAM
4. L' AWS SAM application déploie les ressources nécessaires pour connecter le métastore Hive externe au catalogue de données.
5. Pour appliquer les autorisations Lake Formation à la base de données et aux tables Hive, vous créez une base de données dans le catalogue de données à l'aide des détails de connexion au métastore Hive, et vous mappez cette base de données à la base de données Hive.
6. Accordez des autorisations sur les bases de données fédérées aux principaux de votre compte ou d'un autre compte.

Note

Vous pouvez connecter le catalogue de données à un métastore Hive externe, créer des bases de données fédérées et exécuter des requêtes et des scripts ETL sur des bases de données et des tables Hive sans appliquer les autorisations de Lake Formation. Pour les données source dans Amazon S3 qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions.

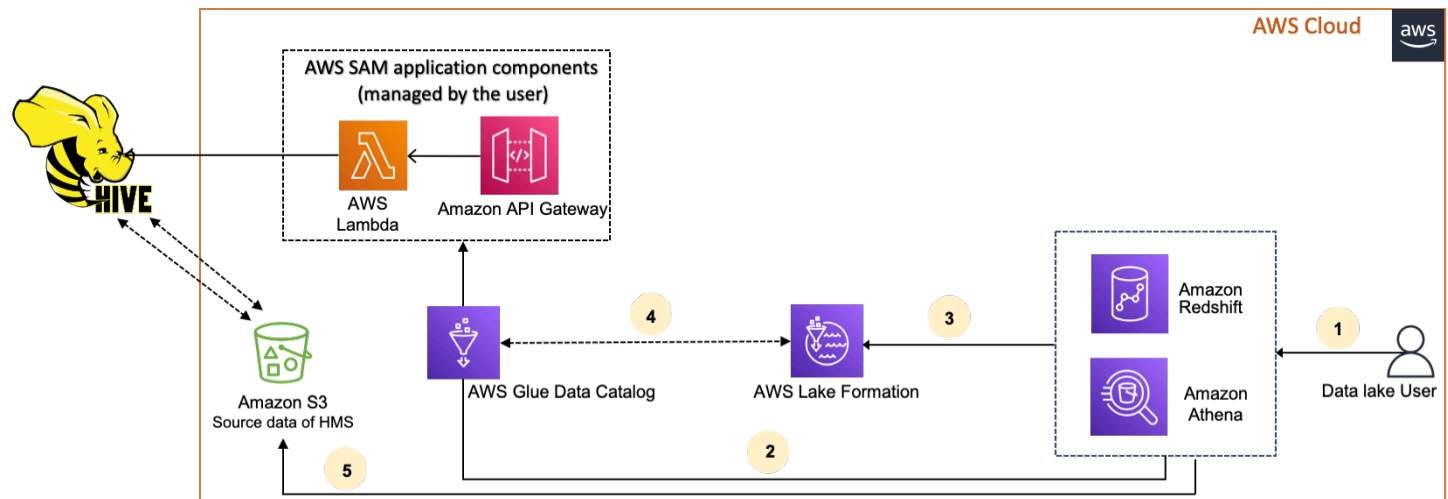
Pour connaître les limitations, veuillez consulter [Considérations et limites relatives au partage des données du magasin de métadonnées Hive](#).

Rubriques

- [Flux de travail](#)
- [Conditions préalables à la connexion du catalogue de données au métastore Hive](#)
- [Connexion du catalogue de données à un métastore Hive externe](#)
- [Ressources supplémentaires](#)

Flux de travail

Le schéma suivant montre le flux de travail pour connecter le AWS Glue Data Catalog à un métastore Hive externe.



1. Un directeur soumet une requête à l'aide d'un service intégré tel qu'Athena ou Redshift Spectrum.
2. Le service intégré appelle le catalogue de données pour obtenir les métadonnées, qui à son tour appelle le point de terminaison du métastore Hive disponible derrière Amazon API Gateway, et reçoit les réponses aux demandes de métadonnées.
3. Le service intégré envoie la demande à Lake Formation pour vérifier les informations de la table et les informations d'identification pour accéder à la table.
4. Lake Formation autorise la demande et fournit des informations d'identification temporaires à l'application intégrée, qui permet l'accès aux données.
5. À l'aide des informations d'identification temporaires reçues de Lake Formation, le service intégré lit les données d'Amazon S3 et partage les résultats avec le principal.

Conditions préalables à la connexion du catalogue de données au métastore Hive

Pour connecter le AWS Glue Data Catalog à un métastore Apache Hive externe et configurer les autorisations d'accès aux données, vous devez remplir les conditions suivantes :

Note

Nous recommandons qu'un administrateur de Lake Formation déploie l' AWS SAM application, et seul un utilisateur privilégié utilise la connexion au métastore Hive pour créer les bases de données fédérées correspondantes.

1. Création des rôles IAM.

Pour déployer l' AWS SAM application

- Créez un rôle doté des autorisations nécessaires pour déployer les ressources (fonction Lambda Amazon API Gateway, rôle IAM et AWS Glue connexion) requises pour créer une connexion au métastore Hive.

Pour créer des bases de données fédérées

Les autorisations suivantes sont requises pour les ressources :

- `glue:CreateDatabase` on resource `arn:aws:glue:region:account-id:database/gluedatabasename`
- `glue:PassConnection` on resource `arn:aws:glue:region:account-id:connection/hms_connection`

2. Enregistrez le site Amazon S3 auprès de Lake Formation.

Pour utiliser Lake Formation afin de gérer et de sécuriser les données de votre lac de données, vous devez enregistrer l'emplacement Amazon S3 contenant les données des tables du métastore Hive auprès de Lake Formation. Lake Formation peut ainsi vendre des informations d'identification à des services AWS d'analyse tels qu'Athena, Redshift Spectrum et Amazon EMR.

Pour plus d'informations sur l'enregistrement d'un site Amazon S3, consultez [Ajouter un emplacement Amazon S3 à votre lac de données](#).

Lorsque vous enregistrez l'emplacement Amazon S3, cochez la case Enable Data Catalog Federation pour autoriser Lake Formation à assumer un rôle d'accès aux tables d'une base de données fédérée.

[AWS Lake Formation](#) > [Data lake locations](#) > Register location

Register location

Amazon S3 location

Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path

Choose an Amazon S3 path for your data lake.

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

▼

 Do not select the service linked role if you plan to use EMR.

☐ **Enable Data Catalog Federation**
Checking this box will allow Lake Formation to assume a role to access tables in a federated database.

Cancel

Register location

Pour plus d'informations sur l'enregistrement d'un emplacement de données auprès de Lake Formation, consultez [Configurer un emplacement Amazon S3 pour votre lac de données](#).

3. Utilisez la bonne version d'Amazon EMR.

Pour utiliser Amazon EMR avec les bases de données de métastore Hive fédérées, vous devez disposer de la version 3.x ou supérieure de Hive et de la version 6.x ou supérieure d'Amazon EMR.

Connexion du catalogue de données à un métastore Hive externe

[Pour le connecter AWS Glue Data Catalog à un métastore Hive, vous devez déployer une AWS SAM application appelée -. GlueDataCatalogFederation HiveMetastore](#) Il crée les ressources nécessaires pour connecter le métastore Hive externe au catalogue de données. Vous pouvez accéder à l' AWS SAM application dans le AWS Serverless Application Repository.

L' AWS SAM application crée la connexion pour le métastore Hive derrière Amazon API Gateway à l'aide d'une fonction Lambda. L' AWS SAM application utilise un identifiant de ressource uniforme (URI) comme entrée par l'utilisateur et connecte le métastore Hive externe au catalogue de données. Lorsqu'un utilisateur exécute une requête sur des tables Hive, le catalogue de données appelle le point de terminaison API Gateway. Le point de terminaison invoque la fonction Lambda pour récupérer les métadonnées des tables Hive.

Pour connecter le catalogue de données au métastore Hive et configurer les autorisations

1. Déployez AWS SAM l'application.
 1. Connectez-vous au AWS Management Console et ouvrez le AWS Serverless Application Repository.
 2. Dans le volet de navigation, choisissez Applications.
 3. Choisissez Applications publiques.
 4. Sélectionnez l'option Show apps that create custom IAM roles or resource policies (Afficher les applications qui créent des rôles IAM ou des politiques de ressources personnalisés).
 5. Dans le champ de recherche, entrez le nom GlueDataCatalogFederation- HiveMetastore.
 6. Choisissez l'HiveMetastoreapplication GlueDataCatalogFederation-.
 7. Sous Paramètres de l'application, entrez les paramètres minimaux requis suivants pour votre fonction Lambda :
 - Nom de l'application : nom de votre AWS SAM application.
 - GlueConnectionName- Un nom pour la connexion.

- HiveMetastoreURIs- L'URI de votre hôte de métastore Hive.
 - LambdaMemory- La quantité de mémoire Lambda en Mo comprise entre 128 et 10240. La valeur par défaut est 1024.
 - LambdaTimeout- Durée maximale d'invocation Lambda en secondes. La valeur par défaut est 30.
 - VPCSecurityGroupIds VPCSubnetIdentifiers : informations relatives au VPC sur lequel se trouve le métastore Hive.
8. Sélectionnez I acknowledge that this app creates custom IAM roles and resource policies (Je reconnais que cette application crée des politiques de ressources et rôles IAM personnalisés). Pour de plus amples informations, veuillez cliquer sur le lien Info.
 9. En bas à droite de la section Application settings (Paramètres de l'application), choisissez Deploy (Déployer). Lorsque le déploiement est terminé, la fonction Lambda apparaît dans la section Resources (Ressources) dans la console Lambda.

L'application est déployée sur Lambda. Son nom est précédé de serverlessrepo- pour indiquer que l'application a été déployée à partir du AWS Serverless Application Repository La sélection de l'application vous amène à la page Ressources où chacune des ressources de l'application qui ont été déployées est répertoriée. Les ressources incluent la fonction Lambda qui permet la communication entre le catalogue de données et le métastore Hive, la AWS Glue connexion et d'autres ressources nécessaires à la fédération de bases de données.

2. Créez une base de données fédérée dans le catalogue de données.

Après avoir créé une connexion au métastore Hive, vous pouvez créer des bases de données fédérées dans le catalogue de données qui pointent vers les bases de données du métastore Hive externes. Vous devez créer une base de données correspondante dans le catalogue de données pour chaque base de données de métastore Hive que vous connectez au catalogue de données.

Lake Formation console

1. Sur la page Partage de données, choisissez l'onglet Bases de données partagées, puis sélectionnez Créer une base de données.
2. Pour Nom de la connexion, choisissez le nom de votre connexion au métastore Hive dans le menu déroulant.

3. Entrez un nom de base de données unique et l'identifiant de source de fédération pour la base de données. Il s'agit du nom que vous utilisez dans vos instructions SQL lorsque vous interrogez des tables. Le nom peut comporter un maximum de 255 caractères et doit être unique au sein de votre compte.
4. Choisissez Créer une base de données.

AWS CLI

```
aws glue create-database \  
'{  
  "CatalogId": "<111122223333>",  
  "database-input": {  
    "Name": "<fed_glue_db>",  
    "FederatedDatabase": {  
      "Identifier": "<hive_db_on_emr>",  
      "ConnectionName": "<hms_connection>"  
    }  
  }  
'
```

3. Affichez les tables de la base de données fédérée.

Après avoir créé la base de données fédérée, vous pouvez consulter la liste des tables de votre métastore Hive à l'aide de la console Lake Formation ou du AWS CLI

Lake Formation console

1. Sélectionnez le nom de la base de données dans l'onglet Bases de données partagées.
2. Sur la page Bases de données, choisissez Afficher les tables.

AWS CLI

Les exemples suivants montrent comment récupérer la définition de connexion, le nom de la base de données et certaines ou toutes les tables de la base de données. Remplacez l'ID du catalogue de données par l'ID Compte AWS valide que vous avez utilisé pour créer la base de données. Remplacez `hms_connection` par le nom de la connexion.

```
aws glue get-connection \  

```

```
--name <hms_connection> \  
--catalog-id 111122223333
```

```
aws glue get-database \  
--name <fed_glu_db> \  
--catalog-id 111122223333
```

```
aws glue get-tables \  
--database-name <fed_glue_db> \  
--catalog-id 111122223333
```

```
aws glue get-table \  
--database-name <fed_glue_db> \  
--name <hive_table_name> \  
--catalog-id 111122223333
```

4. Octroyez les autorisations.

Après avoir créé la base de données, vous pouvez accorder des autorisations à d'autres utilisateurs et rôles IAM dans votre compte ou à Comptes AWS des organisations externes. Vous ne pourrez pas accorder d'autorisations d'écriture (insertion, suppression) et de métadonnées (modification, suppression, création) sur les bases de données fédérées. Pour plus d'informations sur l'octroi d'autorisations, consultez [Gestion des autorisations relatives à Lake Formation](#).

5. Interrogez les bases de données fédérées.

Après avoir accordé les autorisations, les utilisateurs peuvent se connecter et commencer à interroger la base de données fédérée à l'aide d'Athena et d'Amazon Redshift. Les utilisateurs peuvent désormais utiliser le nom de la base de données locale pour faire référence à la base de données Hive dans les requêtes SQL.

Exemple de syntaxe Amazon Athena de requête

fed_glue_db Remplacez-le par le nom de base de données locale que vous avez créé précédemment.

```
Select * from fed_glue_db.customers limit 10;
```

Ressources supplémentaires

Le billet de blog suivant contient des instructions détaillées pour configurer les autorisations de Lake Formation sur une base de données et des tables de métastore Hive, et pour les interroger à l'aide d'Athena. Nous illustrons également un cas d'utilisation du partage entre comptes, dans lequel un responsable de Lake Formation inscrit sur le compte producteur A partage une base de données Hive fédérée et des tables en utilisant le tag LF sur le compte client B.

- [Interrogez votre métastore Apache Hive avec des autorisations AWS Lake Formation](#)

Gestion des autorisations relatives à Lake Formation

Lake Formation fournit des contrôles d'accès centralisés aux données de votre lac de données. Vous pouvez définir des règles basées sur des politiques de sécurité pour vos utilisateurs et applications par rôle dans Lake Formation, et l'intégration permet d' Gestion des identités et des accès AWS authentifier ces utilisateurs et ces rôles. Une fois les règles définies, Lake Formation applique vos contrôles d'accès au niveau de la table, de la colonne et de la granularité au niveau des rôles pour les utilisateurs d'Amazon Redshift Spectrum et Amazon Athena.

Rubriques

- [Octroi d'autorisations de localisation des données](#)
- [Octroi d'autorisations sur les ressources du catalogue de données](#)
- [Exemple de scénario d'autorisations](#)
- [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#)
- [Affichage des autorisations de base de données et de tables dans Lake Formation](#)
- [Révocation de l'autorisation à l'aide de la console Lake Formation](#)
- [Partage de données entre comptes dans Lake Formation](#)
- [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#)
- [Création de liens vers des ressources](#)
- [Accès aux tables dans toutes les régions](#)

Octroi d'autorisations de localisation des données

Les autorisations de localisation des données AWS Lake Formation permettent aux principaux de créer et de modifier les ressources du catalogue de données qui pointent vers des sites Amazon S3 enregistrés désignés. Les autorisations de localisation des données s'ajoutent aux autorisations relatives aux données de Lake Formation pour sécuriser les informations contenues dans votre lac de données.

Lake Formation n'utilise pas le service AWS Resource Access Manager (AWS RAM) pour octroyer des autorisations de localisation des données. Vous n'avez donc pas besoin d'accepter des invitations à partager des ressources pour obtenir des autorisations de localisation des données.

Vous pouvez accorder des autorisations de localisation des données à l'aide de la console Lake Formation, de l'API ou AWS Command Line Interface (AWS CLI).

 Note

Pour qu'une subvention soit acceptée, vous devez d'abord enregistrer l'emplacement des données auprès de Lake Formation.

 Voir aussi :

- [Underlying data access control](#)

Rubriques

- [Octroi d'autorisations de localisation des données \(même compte\)](#)
- [Octroi d'autorisations de localisation des données \(compte externe\)](#)
- [Octroi d'autorisations sur un emplacement de données partagé avec votre compte](#)

Octroi d'autorisations de localisation des données (même compte)

Suivez ces étapes pour accorder des autorisations de localisation des données aux principaux de votre AWS compte. Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

AWS Management Console

Pour accorder des autorisations de localisation des données (même compte)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant que principal ayant accordé des autorisations sur l'emplacement de données souhaité.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données.
3. Choisissez Accorder.
4. Dans la boîte de dialogue Octroyer des autorisations, assurez-vous que la vignette Mon compte est sélectionnée. Fournissez ensuite les informations suivantes :
 - Pour les utilisateurs et les rôles IAM, choisissez un ou plusieurs principaux.

- Pour les utilisateurs et les groupes SAML et Amazon Quick, entrez un ou plusieurs noms de ressources Amazon (ARNs) pour les utilisateurs ou les groupes fédérés via SAML ou pour les utilisateurs ou groupes ARNs Amazon Quick.

Entrez un ARN à la fois, puis appuyez sur Entrée après chaque ARN. Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

- Pour les emplacements de stockage, choisissez Browse, puis choisissez un emplacement de stockage Amazon Simple Storage Service (Amazon S3). L'emplacement doit être enregistré auprès de Lake Formation. Choisissez à nouveau Parcourir pour ajouter un autre emplacement. Vous pouvez également saisir le lieu, mais assurez-vous de le faire précéder de. s3://
- Dans Emplacement du compte enregistré, entrez le numéro du AWS compte sur lequel le point de vente est enregistré. Il s'agit par défaut de votre identifiant de compte. Dans un scénario multicompte, les administrateurs du lac de données d'un compte destinataire peuvent spécifier ici le compte propriétaire lorsqu'ils accordent l'autorisation de localisation des données aux autres principaux du compte destinataire.
- (Facultatif) Pour permettre aux principaux sélectionnés d'accorder des autorisations de localisation des données sur l'emplacement sélectionné, sélectionnez Accordable.

Grant permissions ✕

Add access permissions for specific storage locations.

☒ **My account**
User or role from this AWS account.

☐ **External account**
AWS account or AWS organization outside of my account.

IAM users and roles
Add one or more IAM users or roles.

Choose IAM principals to add

datalake_user ✕
User

SAML and Amazon QuickSight users and groups
Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: `arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>`

Storage locations
Choose one or more data lake locations.

Registered account location
The account where this storage location is registered in AWS Lake Formation.

☐ Grantable

5. Choisissez Accorder.

AWS CLI

Pour accorder des autorisations de localisation des données (même compte)

- Exécutez une `grant-permissions` commande et `DATA_LOCATION_ACCESS` accordez-la au principal, en spécifiant le chemin Amazon S3 comme ressource.

Exemple

L'exemple suivant accorde des autorisations de localisation des données `s3://retail` à l'utilisateur `datalake_user1`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1
```

```
--permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":  
{"ResourceArn":"arn:aws:s3:::retail"}}'
```

Exemple

L'exemple suivant accorde des autorisations de localisation des données `s3://retail` à `ALLIAMPrincipals` un groupe.

```
aws lakeformation grant-permissions --principal  
DataLakePrincipalIdentifier=111122223333:IAMPrincipals --  
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":  
{"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

 Voir aussi :

- [Référence des autorisations de Lake Formation](#)

Octroi d'autorisations de localisation des données (compte externe)

Suivez ces étapes pour accorder des autorisations de localisation des données à un AWS compte ou à une organisation externe.

Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

Avant de commencer

Assurez-vous que toutes les conditions d'accès entre comptes sont satisfaites. Pour de plus amples informations, veuillez consulter [Conditions préalables](#).

AWS Management Console

Pour accorder des autorisations de localisation des données (compte externe, console)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données, puis choisissez Grant.

3. Dans la boîte de dialogue Octroyer des autorisations, choisissez la vignette Compte externe.
4. Saisissez les informations suivantes :
 - Pour l'ID de AWS compte ou AWS l'ID d'organisation, entrez des numéros de AWS compte, d'organisation IDs ou d'unité organisationnelle valides IDs.

Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

Un identifiant d'unité organisationnelle se compose de « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (l'ID de la racine qui contient l'unité d'organisation). Cette chaîne est suivie d'un deuxième « - » (trait d'union) et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

- Sous Emplacements de stockage, choisissez Browse, puis choisissez un emplacement de stockage Amazon Simple Storage Service (Amazon S3). L'emplacement doit être enregistré auprès de Lake Formation.

Grant permissions ×

Add access permissions for specific storage locations.

☐ My account
User or role from this AWS account.

☒ External account
AWS account or AWS organization outside of my account.

AWS account ID or AWS organization ID

Q Enter AWS account ID or AWS organization ID

111122223333 ×
Account

Enter one or more AWS account IDs or AWS organization IDs. Press Enter after each ID.

Storage locations
Choose one or more data lake locations.

s3://retail/transactions/2020q1 Browse

☒ Grantable

Cancel Grant

5. Sélectionnez Grantable.
6. Choisissez Accorder.

AWS CLI

Pour accorder des autorisations de localisation des données (compte externe, AWS CLI)

- Pour accorder des autorisations à un AWS compte externe, entrez une commande similaire à la suivante.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333 --permissions
"DATA_LOCATION_ACCESS" --permissions-with-grant-option
"DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
{"CatalogId":"123456789012", "ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

Cette commande accorde DATA_LOCATION_ACCESS avec l'option grant le compte 1111-2222-3333 sur le site Amazon S3s3://retail/transactions/2020q1, qui appartient au compte 1234-5678-9012.

Pour accorder des autorisations à une organisation, entrez une commande similaire à la suivante.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "DATA_LOCATION_ACCESS" --permissions-
with-grant-option "DATA_LOCATION_ACCESS" --resource '{"DataLocation":
{"CatalogId":"123456789012", "ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

Cette commande accorde DATA_LOCATION_ACCESS une option d'autorisation à l'organisation o-abcdefghijkl sur le site Amazon S3s3://retail/transactions/2020q1, qui appartient au compte 1234-5678-9012.

Pour accorder des autorisations à un mandant sur un AWS compte externe, entrez une commande similaire à la suivante.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
{"ResourceArn":"arn:aws:s3::retail/transactions/2020q1", "CatalogId":
"123456789012"}}'
```

Cette commande est accordée DATA_LOCATION_ACCESS à un mandant du compte 1111-2222-3333 sur le site Amazon S3s3://retail/transactions/2020q1, qui appartient au compte 1234-5678-9012.

Exemple

L'exemple suivant accorde des autorisations de localisation des données s3://retail pour les ALLIAMPrincipals regrouper dans un compte externe.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
{"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}'}
```

 Voir aussi :

- [Référence des autorisations de Lake Formation](#)

Octroi d'autorisations sur un emplacement de données partagé avec votre compte

Une fois qu'une ressource de catalogue de données est partagée avec votre AWS compte, en tant qu'administrateur du lac de données, vous pouvez accorder des autorisations sur la ressource aux autres principaux de votre compte. Si l'ALTERautorisation est accordée sur une table partagée et que la table pointe vers un emplacement Amazon S3 enregistré, vous devez également accorder des autorisations de localisation des données sur cet emplacement. De même, si l'ALTERautorisation CREATE_TABLE ou est accordée sur une base de données partagée et que la base de données possède une propriété d'emplacement qui pointe vers un emplacement enregistré, vous devez également accorder des autorisations de localisation des données sur cet emplacement.

Pour accorder des autorisations de localisation des données sur un emplacement partagé à un responsable de votre compte, votre compte doit avoir obtenu l'DATA_LOCATION_ACCESSautorisation sur le lieu avec l'option d'octroi. Lorsque vous accordez ensuite DATA_LOCATION_ACCESS l'autorisation à un autre principal sur votre compte, vous devez inclure l'ID du catalogue de données

(ID de AWS compte) du compte propriétaire. Le compte propriétaire est le compte qui a enregistré l'emplacement.

Vous pouvez utiliser la AWS Lake Formation console, l'API ou le AWS Command Line Interface (AWS CLI pour accorder des autorisations de localisation des données.

Pour accorder des autorisations sur un emplacement de données partagé avec votre compte (console)

- Suivez les étapes de [Octroi d'autorisations de localisation des données \(même compte\)](#).

Pour les emplacements de stockage, vous devez saisir les emplacements. Dans Emplacement du compte enregistré, entrez le AWS numéro de compte du propriétaire.

Pour accorder des autorisations sur un emplacement de données partagé avec votre compte (AWS CLI)

- Entrez l'une des commandes suivantes pour accorder des autorisations à un utilisateur ou à un rôle.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
--permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
{"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
--permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
{"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
```

Octroi d'autorisations sur les ressources du catalogue de données

Vous pouvez accorder des autorisations de données aux principaux AWS Lake Formation afin que ceux-ci puissent créer et gérer les ressources du catalogue de données et accéder aux données sous-jacentes. Vous pouvez accorder des autorisations Data Lake sur les catalogues, les bases de données, les tables et les vues. Lorsque vous accordez des autorisations sur des tables, vous pouvez limiter l'accès à des colonnes ou à des lignes de table spécifiques pour un contrôle d'accès encore plus précis.

Vous pouvez accorder des autorisations sur des catalogues, des bases de données, des tables et des vues individuels, ou avec une seule opération d'autorisation, vous pouvez accorder des autorisations sur toutes les bases de données, tables et vues d'un catalogue ou d'une base de données. Si vous accordez des autorisations sur toutes les tables d'une base de données aux principaux IAM, vous accordez implicitement l'`DESCRIBE` autorisation sur la base de données. La base de données apparaît ensuite sur la page Bases de données de la console et est renvoyée par l'opération `GetDatabases` d'API. Le même principe s'applique au niveau du catalogue : lorsque vous recevez des autorisations pour les bases de données d'un catalogue, vous obtenez également `DESCRIBE` des autorisations pour ce catalogue.

 Important

L'`DESCRIBE` autorisation implicite s'applique uniquement lors de l'octroi d'autorisations aux principaux IAM au sein du même AWS compte. Pour les ressources entre comptes, vous devez explicitement accorder `DESCRIBE` des autorisations. L'octroi automatique des `DESCRIBE` autorisations ne s'applique pas lors de l'utilisation du contrôle d'accès basé sur les attributs (ABAC). Lorsque vous accordez des autorisations sur toutes les tables d'une base de données à l'aide d'attributs, Lake Formation n'accorde pas implicitement d'`DESCRIBE` autorisation à la base de données.

Vous pouvez accorder des autorisations en utilisant la méthode des ressources nommées ou la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC).

Vous pouvez accorder des autorisations aux directeurs d'un même compte ou d'une organisation Compte AWS ou à un compte externe. Lorsque vous accordez des autorisations à des comptes ou à des organisations externes, vous partagez des objets du catalogue de données dont vous êtes propriétaire avec ces comptes ou organisations. Les responsables de ces comptes ou organisations peuvent ensuite accéder aux objets du catalogue de données dont vous êtes propriétaire et aux données sous-jacentes.

 Note

Actuellement, la méthode LF-TBAC prend en charge l'octroi d'autorisations entre comptes aux principaux Comptes AWS, aux organisations et aux unités organisationnelles IAM (). OUs

Lorsque vous accordez des autorisations à des comptes ou à des organisations externes, vous devez inclure l'option d'octroi. Seul l'administrateur du lac de données du compte externe peut accéder aux objets partagés jusqu'à ce qu'il accorde des autorisations sur les objets partagés aux autres principaux du compte externe.

Vous pouvez accorder des autorisations au catalogue de données à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Note

Lorsque vous supprimez un objet du catalogue de données, toutes les autorisations associées à cet objet ne sont plus valides. Le fait de recréer la même ressource avec le même nom ne récupérera pas les autorisations de Lake Formation. Les utilisateurs devront à nouveau configurer de nouvelles autorisations.

Voir aussi :

- [Partage des tables et des bases de données du catalogue de données entre les AWS comptes](#)
- [Contrôle d'accès aux métadonnées](#)
- [Référence des autorisations de Lake Formation](#)

Autorisations IAM requises pour accorder ou révoquer les autorisations de Lake Formation

Tous les principaux, y compris l'administrateur du lac de données, ont besoin des autorisations Gestion des identités et des accès AWS (IAM) suivantes pour accorder ou révoquer les autorisations de catalogue de AWS Lake Formation données ou les autorisations de localisation des données avec l'API Lake Formation ou le : AWS CLI

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`

- `glue:GetTable`, `glue:GetDatabase`, ou `glue:GetCatalog` pour une table, une base de données ou un catalogue auxquels vous accordez des autorisations à l'aide de la méthode de ressource nommée.

Note

Les administrateurs des lacs de données disposent d'autorisations implicites pour accorder et révoquer les autorisations relatives à Lake Formation. Mais ils ont toujours besoin des autorisations IAM sur l'octroi de l'autorisation Lake Formation et de la révocation des opérations d'API.

Les rôles IAM dotés d'une politique `AWSLakeFormationDataAdmin` AWS gérée ne peuvent pas ajouter de nouveaux administrateurs de lacs de données, car cette politique contient un refus explicite du fonctionnement de l'API Lake Formation. `PutDataLakeSetting`

La politique IAM suivante est recommandée aux directeurs qui ne sont pas des administrateurs de lacs de données et qui souhaitent accorder ou révoquer des autorisations à l'aide de la console Lake Formation.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:ListPermissions",
        "lakeformation:GrantPermissions",
        "lakeformation:BatchGrantPermissions",
        "lakeformation:RevokePermissions",
        "lakeformation:BatchRevokePermissions",
        "glue:GetCatalogs",
        "glue:GetDatabases",
        "glue:SearchTables",
        "glue:GetTables",
        "glue:GetCatalog",
        "glue:GetDatabase",
        "glue:GetTable",

```

```

        "iam:ListUsers",
        "iam:ListRoles",
        "sso-directory:DescribeUser",
        "sso-directory:DescribeGroup",
        "sso:DescribeInstance"
    ],
    "Resource": "*"
}
]
}

```

Toutes les `iam:` autorisations glue: et autorisations de cette politique sont disponibles dans la politique AWS gérée `AWSGlueConsoleFullAccess`.

Pour accorder des autorisations à l'aide du contrôle d'accès basé sur les balises Lake Formation (LF-TBAC), les principaux ont besoin d'autorisations IAM supplémentaires. Pour plus d'informations, consultez [Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation](#) et [Référence des personnalités de Lake Formation et des autorisations IAM](#).

Autorisations entre comptes

Les utilisateurs qui souhaitent accorder des autorisations entre comptes Lake Formation à l'aide de la méthode des ressources nommées doivent également disposer des autorisations définies dans la politique `AWSLakeFormationCrossAccountManager` AWS gérée.

Les administrateurs des lacs de données ont besoin des mêmes autorisations pour accorder des autorisations entre comptes, ainsi que de l'autorisation AWS Resource Access Manager (AWS RAM) pour autoriser l'octroi d'autorisations aux organisations. Pour de plus amples informations, veuillez consulter [Autorisations d'administrateur du lac de données](#).

L'utilisateur administratif

Un directeur disposant d'autorisations administratives (par exemple, dans le cadre de la politique `AdministratorAccess` AWS gérée) est autorisé à accorder des autorisations à Lake Formation et à créer des administrateurs de lacs de données. Pour refuser à un utilisateur ou à un rôle l'accès aux opérations de l'administrateur de Lake Formation, joignez ou ajoutez à sa politique une `Deny` déclaration concernant les opérations d'API de l'administrateur.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "lakeformation:GetDataLakeSettings",
        "lakeformation:PutDataLakeSettings"
      ],
      "Effect": "Deny",
      "Resource": [
        "*"
      ]
    }
  ]
}
```

Important

Pour empêcher les utilisateurs de s'ajouter en tant qu'administrateurs à l'aide d'un script d'extraction, de transformation et de chargement (ETL), assurez-vous que l'accès à ces opérations d'API est refusé à tous les utilisateurs et rôles non administrateurs. La politique `AWSLakeFormationDataAdmin` AWS gérée contient un refus explicite du fonctionnement de l'API Lake Formation, `PutDataLakeSetting` qui empêche les utilisateurs d'ajouter de nouveaux administrateurs de lacs de données.

Octroi d'autorisations de données à l'aide de la méthode de ressource nommée

La méthode de ressource nommée Data Catalog permet d'accorder des autorisations à des AWS Glue Data Catalog objets, tels que des catalogues, des bases de données, des tables, des colonnes et des vues, en utilisant une approche centralisée. Il vous permet de définir des politiques basées sur les ressources qui contrôlent l'accès à des ressources spécifiques au sein de votre lac de données.

Lorsque vous utilisez la méthode de ressource nommée pour accorder des autorisations, vous pouvez spécifier le type de ressource et les autorisations que vous souhaitez accorder ou révoquer

pour cette ressource. Vous pouvez également révoquer l'autorisation ultérieurement si nécessaire, supprimant ainsi les autorisations des ressources associées.

Vous pouvez accorder des autorisations à l'aide de la AWS Lake Formation console ou du AWS Command Line Interface (AWS CLI). APIs

Rubriques

- [Octroi d'autorisations de catalogue à l'aide de la méthode de ressource nommée](#)
- [Octroi d'autorisations de base de données via la méthode de ressource nommée](#)
- [Octroi d'autorisations de table à l'aide de la méthode de ressource nommée](#)
- [Octroi d'autorisations sur les vues à l'aide de la méthode de ressource nommée](#)

Octroi d'autorisations de catalogue à l'aide de la méthode de ressource nommée

Les étapes suivantes expliquent comment accorder des autorisations de catalogue à l'aide de la méthode des ressources nommées.

Console

Utilisez la page Accorder des autorisations sur la console Lake Formation. La page est divisée selon les sections suivantes :

- Type de principal : vous pouvez accorder des autorisations à des principaux spécifiques ou utiliser des balises d'attribut.
- Principaux : utilisateurs, rôles, utilisateurs et groupes IAM Identity Center, utilisateurs et groupes SAML, AWS comptes, organisations ou unités organisationnelles auxquels accorder les autorisations.

Principal par attributs — Ajoutez des paires clé-valeur de balise à partir de balises de IAMroles session IAM. Les principaux dont les attributs correspondent ont accès à la ressource spécifiée.

- Balises LF ou ressources de catalogue : catalogues, bases de données, tables, vues ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

 Note

Pour accorder des autorisations sur un lien de ressource de base de données, voir [Octroi d'autorisations relatives aux liens vers](#).

1. Ouvrez la page Accorder des autorisations.

Ouvrez la AWS Lake Formation console sur <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'administrateur du lac de données, créateur du catalogue ou utilisateur IAM disposant d'autorisations Grantable sur le catalogue.

Effectuez l'une des actions suivantes :

- Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données. Ensuite, choisissez Accorder.
- Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données. Ensuite, sur la page Catalogues, choisissez un catalogue, puis dans le menu Actions, sous Autorisations, choisissez Grant.

 Note

Vous pouvez accorder des autorisations sur un catalogue via son lien de ressource. Pour ce faire, sur la page Catalogues, choisissez un conteneur de liens de catalogue, puis dans le menu Actions, choisissez Grant on target. Pour de plus amples informations, veuillez consulter [Mode de fonctionnement des liens des ressources dans Lake Formation](#).

2. Ensuite, dans la section Type de principal, choisissez les principes ou spécifiez les attributs attachés aux principaux.

Grant permissions

Principal type

Grant access by IAM principal or attribute tags.



Principals

Grant permissions directly to specific IAM users, roles, or Identity Center principals.



Principals by attributes

Grant permissions using attribute tags with Cedar policy expressions.

Principals

Choose the principals to grant permissions.



IAM users and roles

Users or roles from this AWS account.



IAM Identity Center - *new*

Users and groups configured in IAM Identity Center.



SAML users and groups

SAML users and group or QuickSight ARNs.



External accounts

AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add



lf-data-analyst



User

Spécifier les principes

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes. Sélectionnez Ajouter pour ajouter d'autres utilisateurs ou groupes.

Utilisateurs et groupes SAML

Pour les utilisateurs et les groupes SAML et Quick, entrez un ou plusieurs noms de ressources Amazon (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou

pour les utilisateurs ou groupes ARNs Amazon Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

 Note

L'intégration de Lake Formation à Quick n'est prise en charge que pour Quick Enterprise Edition.

Comptes externes

Pour Compte AWS, AWS organisation ou directeur IAM, entrez un ou plusieurs AWS comptes IDs, organisations IDs, unités organisationnelles ou ARN valides pour l'utilisateur ou le rôle IAM. IDs Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

Principaux par attributs

Attributs

Ajoutez les paires clé-valeur de la balise IAM à partir du rôle IAM.

Étendue de l'autorisation

Spécifiez si vous accordez des autorisations aux principaux ayant les mêmes attributs dans le même compte ou dans un autre compte.

3. Dans la section Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées.

LF-Tags or catalog resources

Choose a method to grant permissions.

☐ Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ Named Data Catalog resources
Manage permissions for specific databases or tables, in addition to fine-grained data access.

Catalogs

Choose catalogs

mycatalogdemo X

Databases

Select one or more databases.

Choose databases

tpcds1 X
mycatalogdemo

Tables - optional

Select one or more tables.

Choose tables

All tables X
mycatalogdemo

Views - optional

Select one or more views.

Choose views

Data filters - optional

Select one or more data filters.

Choose data filters

Load
more

Create new

[Manage data filters](#)

4. Choisissez un ou plusieurs catalogues dans la liste des catalogues. Vous pouvez également choisir une ou plusieurs bases de données, tables, filtres de and/or données.
5. Dans la section Autorisations du catalogue, sélectionnez les autorisations et les autorisations pouvant être accordées. Sous Autorisations du catalogue, sélectionnez une ou plusieurs autorisations à octroyer.

Catalog permissions

Choose the permissions to grant on the catalog. Choosing Super user overwrites individual permissions, granting unrestricted administrative access.

☐ Super user

A super user has unrestricted administrative privileges to perform any operation on all resources within the catalog (databases, tables, and views).

Catalog permissions

Choose specific access permissions to grant.

☐ Create database

☐ Describe

☐ Alter

☐ Drop

☐ Super

This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions

Choose the permission that can be granted to others.

☐ Create database

☐ Describe

☐ Alter

☐ Drop

☐ Super

This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel

Grant

Choisissez Super utilisateur pour accorder des privilèges administratifs illimités afin d'effectuer n'importe quelle opération sur toutes les ressources du catalogue (bases de données, tables et vues).

Note

Après avoir accordé `Create database` ou `Alter` sur un catalogue doté d'une propriété de localisation pointant vers un emplacement enregistré, veuillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour de plus amples informations, veuillez consulter [Octroi d'autorisations de localisation des données](#).

- (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS Cette option n'est pas prise en charge lorsque vous octroyez des autorisations à un principal IAM à partir d'un compte externe.
- Choisissez Accorder.

La page Autorisations relatives aux données affiche les détails des autorisations. Si vous avez utilisé l'option Principaux par attribut pour accorder des autorisations, vous pouvez afficher l'autorisation accordée ALLPrincipals dans la liste.

AWS CLI

Pour accorder des autorisations de catalogue à l'aide de AWS CLI, voir [Création de catalogues fédérés Amazon Redshift](#).

Octroi d'autorisations de base de données via la méthode de ressource nommée

Les étapes suivantes expliquent comment accorder des autorisations de base de données à l'aide de la méthode des ressources nommées.

Console

Utilisez la page Accorder des autorisations sur la console Lake Formation. La page est divisée selon les sections suivantes :

- Type principal : la section Principaux inclut les utilisateurs, les rôles, les utilisateurs et groupes IAM Identity Center, les utilisateurs et groupes SAML, les AWS comptes, les organisations ou les unités organisationnelles habilités à accorder des autorisations. Dans la section Principaux par attributs, vous pouvez spécifier la clé et les valeurs des attributs attachés aux rôles IAM.
- Balises LF ou ressources du catalogue : bases de données, tables, vues ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

Note

Pour accorder des autorisations sur un lien de ressource de base de données, voir [Octroi d'autorisations relatives aux liens vers](#).

1. Ouvrez la page Accorder des autorisations.

Ouvrez la AWS Lake Formation console sur <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'administrateur du lac de données, créateur

de base de données ou utilisateur IAM disposant d'autorisations Grantable sur la base de données.

Effectuez l'une des actions suivantes :

- Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données. Ensuite, choisissez Accorder.
- Dans le volet de navigation, sélectionnez Bases de données sous Catalogue de données. Ensuite, sur la page Bases de données, choisissez une base de données, puis dans le menu Actions, sous Autorisations, choisissez Grant.

Note

Vous pouvez accorder des autorisations sur une base de données via son lien de ressource. Pour ce faire, sur la page Bases de données, choisissez un lien vers une ressource, puis dans le menu Actions, choisissez Grant on target. Pour de plus amples informations, veuillez consulter [Mode de fonctionnement des liens des ressources dans Lake Formation](#).

2. Dans la section Type de principal, spécifiez les principaux ou accordez des autorisations aux principaux à l'aide d'attributs.

Grant permissions

Principals

Choose the principals to grant permissions.

☒ IAM users and roles

Users or roles from this AWS account.

☐ IAM Identity Center

- *new*
Users and groups configured in IAM Identity Center.

☐ SAML users and groups

SAML users and group or QuickSight ARNs.

☐ External accounts

AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

AWSServiceRoleForRedshift ✕
Role

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes. Sélectionnez Ajouter pour ajouter d'autres utilisateurs ou groupes.

Utilisateurs et groupes SAML

Pour les utilisateurs et les groupes SAML et Quick, entrez un ou plusieurs noms de ressources Amazon (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou pour les utilisateurs ou groupes ARNs Amazon Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

Note

L'intégration de Lake Formation à Quick n'est prise en charge que pour Quick Enterprise Edition.

Comptes externes

Pour Compte AWS, AWS organisation ou directeur IAM, entrez un ou plusieurs AWS comptes IDs, organisations IDs, unités organisationnelles ou ARN valides pour l'utilisateur ou le rôle IAM. IDs Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

Principaux par attributs

Spécifiez la clé et la ou les valeurs de l'attribut. Si vous choisissez plusieurs valeurs, vous créez une expression attributaire avec un opérateur OR. Cela signifie que si l'une des valeurs de balise d'attribut attribuées à un rôle ou à un utilisateur IAM correspond, il role/user obtient des autorisations d'accès à la ressource.

Choisissez l'étendue des autorisations en spécifiant si vous accordez des autorisations aux principaux ayant les mêmes attributs dans le même compte ou dans un autre compte.

3. Dans la section Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées.

LF-Tags or catalog resources

Choose a method to grant permissions.

☐ Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ Named Data Catalog resources
Manage permissions for specific databases or tables, in addition to fine-grained data access.

Catalogs

Choose catalogs ▼

Default catalog

Databases

Select one or more databases.

Choose databases ▼

sales ✕

Tables - optional

Select one or more tables.

Choose tables ▼

Views - optional

Select one or more views.

Choose views ▼

Data filters - optional

Select one or more data filters.

Choose data filters ▼

Load more

Create new

[Manage data filters](#) 

4. Choisissez une ou plusieurs bases de données dans la liste des bases de données. Vous pouvez également choisir un ou plusieurs filtres de and/or données de tables.
5. Dans la section Autorisations, sélectionnez les autorisations et les autorisations octroyables. Sous Autorisations de base de données, sélectionnez une ou plusieurs autorisations à accorder.

Database permissions

Database permissions

Choose specific access permissions to grant.

☐ Create table
 ☐ Alter
 ☐ Drop

☐ Describe

Grantable permissions

Choose the permission that may be granted to others.

☐ Create table
 ☐ Alter
 ☐ Drop

☐ Describe

☐ Super

This permission is the union of all the individual permissions to the left, and supersedes them.

☐ Super

This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Note

Après avoir accordé Create Table ou Alter sur une base de données dotée d'une propriété d'emplacement pointant vers un emplacement enregistré, veuillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour de plus amples informations, veuillez consulter [Octroi d'autorisations de localisation des données](#).

- (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS Cette option n'est pas prise en charge lorsque vous octroyez des autorisations à un principal IAM à partir d'un compte externe.
- Choisissez Accorder.

AWS CLI

Vous pouvez accorder des autorisations de base de données en utilisant la méthode de ressource nommée et le AWS Command Line Interface (AWS CLI).

Pour accorder des autorisations de base de données à l'aide du AWS CLI

- Exécutez une `grant-permissions` commande et spécifiez une base de données ou le catalogue de données comme ressource, en fonction de l'autorisation accordée.

Dans les exemples suivants, remplacez-le `<account-id>` par un identifiant de AWS compte valide.

Exemple— Subvention pour créer une base de données

Cet exemple accorde des autorisations `CREATE_DATABASE` à l'utilisateur `datalake_user1`. Étant donné que la ressource pour laquelle cette autorisation est accordée est le catalogue de données, la commande spécifie une `CatalogResource` structure vide comme `resource` paramètre.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {} }'
```

Exemple— Autorisation de créer des tables dans une base de données désignée

L'exemple suivant octroie la `CREATE_TABLE` base de données `retail` à l'utilisateur `datalake_user1`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": { "Name": "retail" } }'
```

Exemple— Subvention à un AWS compte externe avec l'option Grant

L'exemple suivant octroie `CREATE_TABLE` l'option grant sur la base de données `retail` au compte externe 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
--permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
{ "Name": "retail" } }'
```

Exemple— Subvention à une organisation

L'exemple suivant octroie `ALTER` à l'organisation l'option grant sur la base issues de données `so-abcdefghijkl`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "ALTER" --permissions-with-grant-option "ALTER" --
resource '{ "Database": {"Name":"issues"}}'
```

Exemple- Accordez ALLIAMPPrincipals à sur le même compte

L'exemple suivant accorde CREATE_TABLE l'autorisation d'accéder à la base retail de données à tous les principaux d'un même compte. Cette option permet à chaque principal du compte de créer une table dans la base de données et de créer un lien de ressource de table permettant aux moteurs de requêtes intégrés d'accéder aux bases de données et aux tables partagées. Cette option est particulièrement utile lorsqu'un directeur reçoit une subvention entre comptes et n'est pas autorisé à créer des liens vers des ressources. Dans ce scénario, l'administrateur du lac de données peut créer une base de données d'espaces réservés et accorder des CREATE_TABLE autorisations au ALLIAMPPrincipal groupe, permettant ainsi à chaque responsable IAM du compte de créer des liens de ressources dans la base de données d'espaces réservés.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333:IAMPPrincipals
--permissions "CREATE_TABLE" --resource '{ "Database":
{"Name":"temp","CatalogId":"111122223333"}}'
```

Exemple- Subvention ALLIAMPPrincipals à un compte externe

L'exemple suivant accorde des autorisations CREATE_TABLE d'accès à la base retail de données à tous les principaux d'un compte externe. Cette option permet à tous les principaux du compte de créer une table dans la base de données.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333:IAMPPrincipals
--permissions "CREATE_TABLE" --resource '{ "Database":
{"Name":"retail","CatalogId":"123456789012"}}'
```

Note

Après avoir accordé `CREATE_TABLE` ou `ALTER` sur une base de données dotée d'une propriété d'emplacement pointant vers un emplacement enregistré, veuillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour de plus amples informations, veuillez consulter [Octroi d'autorisations de localisation des données](#).

Consultez aussi

- [Référence des autorisations de Lake Formation](#)
- [Octroi d'autorisations sur une base de données ou une table partagée avec votre compte](#)
- [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#)

Octroi d'autorisations de table à l'aide de la méthode de ressource nommée

Vous pouvez utiliser la console Lake Formation ou AWS CLI accorder des autorisations Lake Formation sur les tables du catalogue de données. Vous pouvez accorder des autorisations sur des tables individuelles, ou avec une seule opération d'autorisation, vous pouvez accorder des autorisations sur toutes les tables d'une base de données.

Si vous accordez des autorisations sur toutes les tables d'une base de données, vous les accordez implicitement sur la `DESCRIBE` base de données. La base de données apparaît ensuite sur la page Bases de données de la console et est renvoyée par l'opération `GetDatabases` d'API. Cette `DESCRIBE` autorisation automatique ne s'applique pas lors de l'utilisation du contrôle d'accès basé sur les attributs (ABAC). Lorsque vous accordez des autorisations sur toutes les tables d'une base de données à l'aide d'attributs, Lake Formation n'accorde pas implicitement d'`DESCRIBE` autorisation à la base de données.

Lorsque vous choisissez `SELECT` l'autorisation à accorder, vous avez la possibilité d'appliquer un filtre de colonne, un filtre de ligne ou un filtre de cellule.

Console

Les étapes suivantes expliquent comment accorder des autorisations de table à l'aide de la méthode de ressource nommée et de la page Accorder des autorisations de lac de données sur la console Lake Formation. La page est divisée en sections suivantes :

- Types de principes : utilisateurs, rôles, AWS comptes, organisations ou unités organisationnelles auxquels accorder des autorisations. Vous pouvez également accorder des autorisations aux directeurs dont les attributs correspondent.
- Balises LF ou ressources du catalogue : bases de données, tables ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

Note

Pour accorder des autorisations sur le lien d'une ressource de table, consultez [Octroi d'autorisations relatives aux liens vers](#).

1. Ouvrez la page Accorder des autorisations.

Ouvrez la AWS Lake Formation console à <https://console.aws.amazon.com/lakeformation/> l'adresse et connectez-vous en tant qu'administrateur du lac de données, créateur de la table ou utilisateur ayant obtenu des autorisations sur la table avec l'option d'autorisation.

Effectuez l'une des actions suivantes :

- Dans le volet de navigation, sélectionnez Autorisations relatives aux données sous Autorisations. Ensuite, choisissez Accorder.
- Dans le volet de navigation, choisissez Tables. Ensuite, sur la page Tables, choisissez un tableau, puis dans le menu Actions, sous Autorisations, choisissez Accorder.

Note

Vous pouvez accorder des autorisations sur une table via son lien de ressource. Pour ce faire, sur la page Tables, choisissez un lien vers une ressource, puis dans le

menu Actions, choisissez Grant on target. Pour de plus amples informations, veuillez consulter [Mode de fonctionnement des liens des ressources dans Lake Formation](#).

2. Ensuite, dans la section Principaux types, spécifiez les principaux ou les principaux avec les attributs correspondants pour accorder des autorisations.

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes.

Utilisateurs et groupes SAML

Pour les utilisateurs et les groupes SAML et Quick, entrez un ou plusieurs Amazon Resource Names (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou ARNs pour les utilisateurs ou groupes Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

Note

L'intégration de Lake Formation à Quick n'est prise en charge que pour Quick Enterprise Edition.

Comptes externes

Pour Compte AWS , AWS organisation ou principal IAM, entrez une ou plusieurs organisations Compte AWS IDs IDs, unités organisationnelles ou ARN valides pour l'utilisateur ou le rôle IAM. IDs Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette

chaîne est suivie d'un deuxième caractère « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

Principaux par attributs

Spécifiez la clé et la ou les valeurs de l'attribut. Si vous choisissez plusieurs valeurs, vous créez une expression attributaire avec un opérateur OR. Cela signifie que si l'une des valeurs de balise d'attribut attribuées à un rôle ou à un utilisateur IAM correspond, les autorisations d'accès role/user à la ressource sont obtenues.

Choisissez l'étendue des autorisations en spécifiant si vous accordez des autorisations aux principaux ayant les mêmes attributs dans le même compte ou dans un autre compte.

3. Dans la section Balises LF ou ressources du catalogue, choisissez une base de données. Choisissez ensuite une ou plusieurs tables, ou toutes les tables.

LF-Tags or catalog resources

☐ Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ Named data catalog resources
Manager permissions for specific databases or tables, in addition to fine-grained data access.

Databases

Select one or more databases.

Choose databases ▼

Load more

retail X

Tables - optional

Select one or more tables.

Choose tables ▼

Load more

inventory X
No description available

4. Spécifiez les autorisations sans filtrage des données.

Dans la section Autorisations, sélectionnez les autorisations de table à accorder, et sélectionnez éventuellement les autorisations pouvant être accordées.

Table and column permissions

Table permissions
Choose specific access permissions to grant.

☒ Alter ☒ Insert ☐ Drop
☐ Delete ☒ Select ☒ Describe

☐ Super
This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions
Choose the permission that may be granted to others.

☐ Alter ☐ Insert ☐ Drop
☐ Delete ☐ Select ☐ Describe

☐ Super
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Si vous autorisez Select, la section Autorisations relatives aux données apparaît sous la section Autorisations relatives aux tables et aux colonnes, l'option Accès à toutes les données étant sélectionnée par défaut. Acceptez la valeur par défaut.

Data permissions

☒ All data access
Grant access to all data without any restrictions.

☐ Simple column-based access
Grant data access to specific columns only.

☐ Advanced cell-level filters
Grant access to specific columns and/or rows with data filters.

5. Choisissez Accorder.
6. Spécifiez l'autorisation de sélection avec filtrage des données

Sélectionnez l'autorisation Select. Ne sélectionnez aucune autre autorisation.

La section Autorisations relatives aux données apparaît sous la section Autorisations relatives aux tables et aux colonnes.

7. Effectuez l'une des actions suivantes :
 - Appliquez uniquement un filtrage par colonne simple.
 1. Choisissez Accès simple basé sur des colonnes.

Table and column permissions

Table permissions

Choose specific access permissions to grant.

☐ Alter
 ☐ Insert
 ☐ Drop

☐ Delete
 ☒ Select
 ☐ Describe

☐ Super

This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions

Choose the permission that may be granted to others.

☐ Alter
 ☐ Insert
 ☐ Drop

☐ Delete
 ☒ Select
 ☐ Describe

☐ Super

This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Data permissions

☐ All data access

Grant access to all data without any restrictions.

☒ Simple column-based access

Grant data access to specific columns only.

☐ Advanced cell-level filters

Grant access to specific columns and/or rows with data filters.

Choose permission filter

Choose whether to include or exclude columns.

☒ Include columns

Grant permissions to access specific columns.

☐ Exclude columns

Grant permissions to access all but specific columns.

Select columns

Grantable permissions

Choose the permission that may be granted to others.

☐ Select

2. Choisissez d'inclure ou d'exclure des colonnes, puis choisissez les colonnes à inclure ou à exclure.

Seules les listes d'inclusion sont prises en charge lors de l'octroi d'autorisations à un AWS compte ou à une organisation externe.

3. (Facultatif) Sous Autorisations pouvant être accordées, activez l'option d'octroi pour l'autorisation Select.

Si vous incluez l'option de subvention, le bénéficiaire de la subvention ne peut accorder des autorisations que sur les colonnes que vous lui accordez.

Note

Vous pouvez également appliquer le filtrage des colonnes uniquement en créant un filtre de données qui spécifie un filtre de colonne et spécifie toutes les lignes comme filtre de ligne. Cependant, cela nécessite d'autres étapes.

- Appliquez un filtrage par colonne, ligne ou cellule.
1. Choisissez Filtres avancés au niveau des cellules.

The screenshot shows the 'Data permissions' configuration page. At the top, there are three radio button options: 'All data access' (unselected), 'Simple column-based access' (unselected), and 'Advanced cell-level filters' (selected). Below these is a link 'View existing permissions'. The main section is 'Data filters to grant', which includes a search bar with the placeholder 'Find filter', a 'Manage filters' button, and a 'Create new filter' button. Below the search bar is a table with the following data:

☐	Filter name	Table	Database	Table catalog ID
☐	restrict-pharma	orders	sales	111122223333
☐	no-pharma	orders	sales	111122223333

2. (Facultatif) Agrandir Afficher les autorisations existantes.
3. (Facultatif) Choisissez Créer un nouveau filtre.
4. (Facultatif) Pour afficher les détails des filtres répertoriés, créer de nouveaux filtres ou supprimer des filtres existants, choisissez Gérer les filtres.

La page Filtres de données s'ouvre dans une nouvelle fenêtre de navigateur.

Lorsque vous avez terminé sur la page Filtres de données, retournez à la page Accorder des autorisations et, si nécessaire, actualisez la page pour afficher les nouveaux filtres de données que vous avez créés.

5. Sélectionnez un ou plusieurs filtres de données à appliquer à la subvention.

 Note

Si la liste ne contient aucun filtre de données, cela signifie qu'aucun filtre de données n'a été créé pour la table sélectionnée.

8. Choisissez Accorder.

AWS CLI

Vous pouvez accorder des autorisations de table en utilisant la méthode de ressource nommée et le AWS Command Line Interface (AWS CLI).

Pour accorder des autorisations de table à l'aide du AWS CLI

- Exécutez une `grant-permissions` commande et spécifiez une table comme ressource.

Exemple— Subvention sur une seule table, sans filtrage

L'exemple suivant accorde SELECT et ALTER à l'utilisateur `datalake_user1` dans le AWS compte 1111-2222-3333 sur la table de la base de données. `inventory retail`

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
"Name":"inventory"}}'
```

 Note

Si vous accordez l'ALTER autorisation sur une table dont les données sous-jacentes se trouvent dans un emplacement enregistré, veillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour de plus amples informations, veuillez consulter [Octroi d'autorisations de localisation des données](#).

Exemple— Subvention sur toutes les tables avec l'option Grant : aucun filtrage

L'exemple suivant accorde des autorisations SELECT avec l'option grant sur toutes les tables de la base de données `retail`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --permissions-with-grant-option "SELECT" --resource '{ "Table":
{ "DatabaseName": "retail", "TableWildcard": {} } }'
```

Exemple— Subvention avec filtrage simple par colonne

L'exemple suivant accorde des SELECT autorisations sur un sous-ensemble de colonnes de la table. `persons` Il utilise un simple filtrage par colonne.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"hr",
"Name":"persons", "ColumnNames":["family_name", "given_name", "gender"]}}'
```

Exemple— Subvention avec filtre de données

Cet exemple accorde des SELECT autorisations sur la `orders` table et applique le filtre de `restrict-pharma` données.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Le contenu du fichier est le suivant `grant-params.json`.

```
{
  "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
  "Resource": {
    "DataCellsFilter": {
      "TableCatalogId": "111122223333",
      "DatabaseName": "sales",
      "TableName": "orders",
      "Name": "restrict-pharma"
    }
  },
  "Permissions": ["SELECT"],
  "PermissionsWithGrantOption": ["SELECT"]
}
```

 Consultez aussi

- [Vue d'ensemble des autorisations relatives à Lake Formation](#)
- [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#)
- [Référence des personnalités de Lake Formation et des autorisations IAM](#)
- [Octroi d'autorisations relatives aux liens vers](#)
- [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#)

Octroi d'autorisations sur les vues à l'aide de la méthode de ressource nommée

Les étapes suivantes expliquent comment accorder des autorisations sur les vues à l'aide de la méthode de ressource nommée et de la page Accorder des autorisations. La page est divisée selon les sections suivantes :

- Principaux types : utilisateurs IAM, rôles, utilisateurs et groupes de l'IAM Identity Center Comptes AWS, organisations ou unités organisationnelles auxquels octroyer les autorisations. Vous pouvez également accorder des autorisations aux directeurs dont les attributs correspondent.
- Balises LF ou ressources du catalogue : bases de données, tables, vues ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations : autorisations à accorder au lac de données.

Ouvrez la page Accorder les autorisations

1. Ouvrez la AWS Lake Formation console sur <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou utilisateur IAM disposant d'autorisations Grantable sur la base de données.
2. Effectuez l'une des actions suivantes :
 - Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données. Ensuite, choisissez Accorder.
 - Dans le volet de navigation, sélectionnez Views sous Catalogue de données. Ensuite, sur la page Vues, choisissez une vue, puis dans le menu Actions, sous Autorisations, choisissez Accorder.

 Note

Vous pouvez accorder des autorisations sur une vue via son lien de ressource. Pour ce faire, sur la page Vues, choisissez un lien vers une ressource, puis dans le menu Actions, choisissez Grant on target. Pour de plus amples informations, veuillez consulter [Mode de fonctionnement des liens des ressources dans Lake Formation](#).

Spécifiez les principaux types

Dans la section Principaux types, choisissez Principaux ou Principaux par attributs. Si vous choisissez Principaux, les options suivantes sont disponibles :

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes.

Utilisateurs et groupes SAML

Pour les utilisateurs et les groupes SAML et Quick, entrez un ou plusieurs noms de ressources Amazon (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou pour les utilisateurs ou groupes ARNs Amazon Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

 Note

L'intégration de Lake Formation à Quick n'est prise en charge que pour Quick Enterprise Edition.

Comptes externes

Pour Compte AWS, AWS organisation ou directeur IAM, entrez un ou plusieurs AWS comptes IDs, organisations IDs, unités organisationnelles ou ARN valides pour l'utilisateur ou le rôle IAM. IDs Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

Voir aussi

- [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#)

Principaux par attributs

Spécifiez la clé et la ou les valeurs de l'attribut. Si vous choisissez plusieurs valeurs, vous créez une expression attributaire avec un opérateur OR. Cela signifie que si l'une des valeurs de balise d'attribut attribuées à un rôle ou à un utilisateur IAM correspond, les autorisations d'accès role/user à la ressource sont obtenues.

Choisissez l'étendue des autorisations en spécifiant si vous accordez des autorisations aux principaux ayant les mêmes attributs dans le même compte ou dans un autre compte.

Spécifiez les vues

Dans la section LF-Tags ou ressources du catalogue, choisissez une ou plusieurs vues pour lesquelles vous souhaitez accorder des autorisations.

1. Choisissez Ressources de catalogue de données nommées.
2. Choisissez une ou plusieurs vues dans la liste des vues. Vous pouvez également choisir un ou plusieurs catalogues, bases de données, tables, filtres de and/or données.

L'octroi d'autorisations de lac de données au All tables sein d'une base de données permettra au bénéficiaire de disposer d'autorisations sur toutes les tables et vues de la base de données.

Spécifiez les autorisations

Dans la section Autorisations, sélectionnez les autorisations et les autorisations octroyables.

View permissions

View permissions

Choose specific access permissions to grant.

☐ Select

☐ Describe

☐ Drop

☐ Super

This permission is the union of all the individual permissions to the left, and supersedes them.

Grantable permissions

Choose the permission that may be granted to others.

☐ Select

☐ Describe

☐ Drop

☐ Super

This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel

Grant

1. Sous Afficher les autorisations, sélectionnez une ou plusieurs autorisations à accorder.
2. (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux dans son domaine.
Compte AWS Cette option n'est pas prise en charge lorsque vous octroyez des autorisations à un principal IAM à partir d'un compte externe.
3. Choisissez Accorder.

Voir aussi

- [Référence des autorisations de Lake Formation](#)
- [Octroi d'autorisations sur une base de données ou une table partagée avec votre compte](#)

Contrôle d'accès basé sur des balises Lake Formation

Le contrôle d'accès basé sur les balises de Lake Formation (LF-TBAC) est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans Lake Formation, ces attributs

sont appelés balises LF. Vous pouvez associer des balises LF aux ressources du catalogue de données et accorder des autorisations aux responsables de Lake Formation sur ces ressources à l'aide de ces balises LF. Lake Formation autorise les opérations sur ces ressources lorsque le principal a accordé l'accès à une valeur de balise qui correspond à la valeur de l'étiquette de ressource.

Le LF-TBAC est utile dans les environnements qui se développent rapidement et dans les situations où la gestion des politiques devient fastidieuse.

La méthode LF-TBAC est recommandée pour accorder des autorisations à Lake Formation lorsqu'il existe un grand nombre d'objets du catalogue de données, notamment des catalogues fédérés, des bases de données, des tables et des vues. Lake Formation prend en charge le contrôle d'accès basé sur des balises pour les catalogues fédérés de tables Amazon S3, les entrepôts de Amazon Redshift données et les sources de données fédérées telles que SQL Server et Amazon DynamoDB Snowflake.

Note

Les balises IAM sont différentes des balises LF. Ces étiquettes ne sont pas interchangeables. Les balises LF sont utilisées pour accorder des autorisations à Lake Formation et les balises IAM sont utilisées pour définir les politiques IAM.

Comment fonctionne le contrôle d'accès basé sur des balises Lake Formation

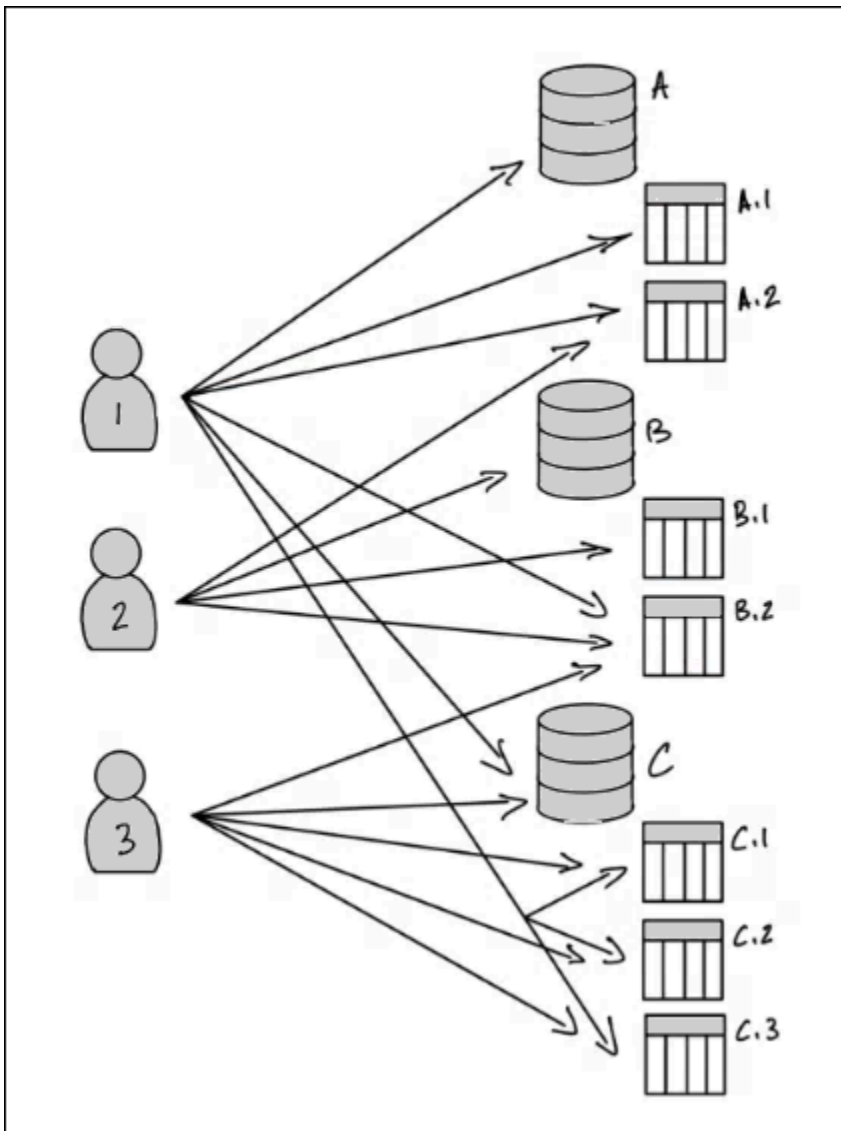
Chaque balise LF est une paire clé-valeur, telle que `ou. department=sales classification=restricted`. Une clé peut avoir plusieurs valeurs définies, telles que `department=sales,marketing,engineering,finance`.

Pour utiliser la méthode LF-TBAC, les administrateurs de lacs de données et les ingénieurs de données effectuent les tâches suivantes.

Sous-tâche	Détails de la tâche
1. Définissez les propriétés et les relations des balises LF.	-
2. Créez les créateurs de balises LF dans Lake Formation.	Ajouter des créateurs de LF-Tag

Sous-tâche	Détails de la tâche
3. Créez le tag LF dans Lake Formation.	Création de balises LF
4. Attribuez des balises LF aux ressources du catalogue de données.	Affectation de balises LF aux ressources du catalogue de données
5. Accordez des autorisations à d'autres principaux pour attribuer des balises LF aux ressources, éventuellement avec l'option d'octroi.	Gestion des autorisations relatives aux valeurs des balises LF
6. Accordez des expressions LF-Tag aux principaux, éventuellement avec l'option grant.	Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC
7. (Recommandé) Après avoir vérifié que les principaux ont accès aux bonnes ressources par le biais de la méthode LF-TBAC, révoquez les autorisations accordées à l'aide de la méthode des ressources nommées.	-

Imaginons le cas où vous devez accorder des autorisations à trois principaux sur trois bases de données et sept tables.



Pour obtenir les autorisations indiquées dans le schéma précédent en utilisant la méthode des ressources nommées, vous devez accorder 17 autorisations, comme suit (en pseudo-code).

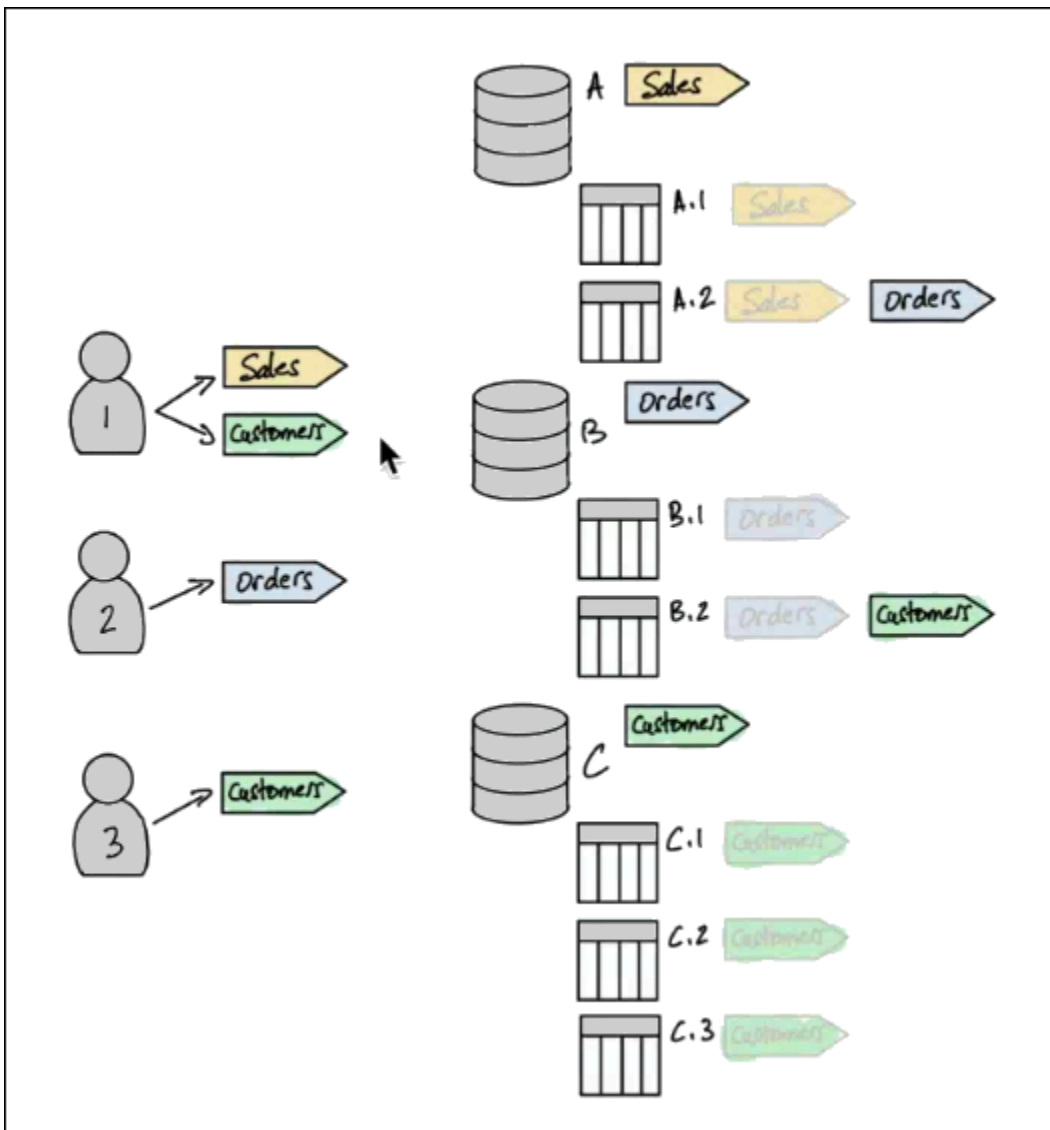
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

Réfléchissez maintenant à la manière dont vous accorderiez des autorisations en utilisant le LF-TBAC. Le schéma suivant indique que vous avez attribué des balises LF aux bases de données et aux tables, et que vous avez accordé des autorisations sur les balises LF aux principaux.

Dans cet exemple, les balises LF représentent des zones du lac de données qui contiennent des analyses pour différents modules d'une suite d'applications de planification des ressources d'entreprise (ERP). Vous pouvez les utiliser pour contrôler l'accès aux données d'analyse des différents modules. Toutes les balises LF ont la clé `module` et les valeurs possibles `SalesOrders`, et `Customers`. Voici un exemple de balise LF :

```
module=Sales
```

Le diagramme montre uniquement les valeurs des balises LF.



Attributions de balises aux ressources et à l'héritage du catalogue de données

Les tables héritent des balises LF des bases de données et les colonnes héritent des balises LF des tables. Les valeurs héritées peuvent être remplacées. Dans le schéma précédent, les balises LF grisées sont héritées.

En raison de l'héritage, l'administrateur du lac de données doit uniquement attribuer les cinq balises LF suivantes aux ressources (en pseudo-code).

```
ASSIGN TAGS module=Sales TO database A
ASSIGN TAGS module=Orders TO table A.2
ASSIGN TAGS module=Orders TO database B
ASSIGN TAGS module=Customers TO table B.2
ASSIGN TAGS module=Customers TO database C
```

Tag : subventions aux directeurs d'école

Après avoir attribué des balises LF aux bases de données et aux tables, l'administrateur du lac de données ne doit accorder que quatre balises LF aux principaux, comme suit (en pseudo-code).

```
GRANT TAGS module=Sales TO Principal 1
GRANT TAGS module=Customers TO Principal 1
GRANT TAGS module=Orders TO Principal 2
GRANT TAGS module=Customers TO Principal 3
```

Désormais, un principal doté de la module=Sales balise LF peut accéder aux ressources du catalogue de données avec la module=Sales balise LF (par exemple, la base de données A), un principal doté de la balise module=Customers LF peut accéder aux ressources avec la module=Customers balise LF, etc.

Les commandes d'autorisation précédentes sont incomplètes. En effet, bien qu'ils indiquent par le biais de balises LF les ressources du catalogue de données sur lesquelles les principaux sont autorisés, ils n'indiquent pas exactement quelles autorisations de Lake Formation (telles que SELECT, ALTER) les principaux ont sur ces ressources. Par conséquent, les commandes de pseudo-code suivantes sont une représentation plus précise de la manière dont les autorisations Lake Formation sont accordées sur les ressources du catalogue de données via des balises LF.

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Sales TO Principal 1
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=Sales TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 1
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=Customers TO Principal 1
```

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Orders TO Principal 2
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=Orders TO Principal 2
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 3
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=Customers TO Principal 3
```

Assemblage : autorisations résultantes sur les ressources

Compte tenu des balises LF attribuées aux bases de données et aux tables dans le schéma précédent, et des balises LF accordées aux principaux dans le diagramme, le tableau suivant répertorie les autorisations de Lake Formation dont disposent les principaux sur les bases de données et les tables.

Principal	Autorisations accordées via des balises LF
Principal 1	• CREATE_TABLE sur la base de données A • SELECT, INSERT sur le tableau A.1 • SELECT, INSERT sur le tableau B.2 • CREATE_TABLE sur la base de données C • SELECT, INSERT sur le tableau C.1 • SELECT, INSERT sur le tableau C.2 • SELECT, INSERT sur le tableau C.3
Principal 2	• SELECT, INSERT sur le tableau A.2 • CREATE_TABLE sur la base de données B • SELECT, INSERT sur le tableau B.1
Principal 3	• SELECT, INSERT sur le tableau B.2 • CREATE_TABLE sur la base de données C • SELECT, INSERT sur le tableau C.1 • SELECT, INSERT sur le tableau C.2 • SELECT, INSERT sur le tableau C.3

Conclusion

Dans cet exemple simple, à l'aide de cinq opérations d'attribution et de huit opérations de subvention, l'administrateur du lac de données a pu spécifier 17 autorisations. Lorsqu'il existe des dizaines de

bases de données et des centaines de tables, l'avantage de la méthode LF-TBAC par rapport à la méthode des ressources nommées devient évident. Dans le cas hypothétique de la nécessité d'accorder à chaque principal l'accès à chaque ressource, et où $n(P)$ sont le nombre de principaux et $n(R)$ le nombre de ressources :

- Avec la méthode des ressources nommées, le nombre de subventions requises est de $n(P) \times n(R)$.
- Avec la méthode LF-TBAC, en utilisant un seul tag LF, le total du nombre de subventions aux directeurs et d'affectations aux ressources est de $+ n(P) n(R)$

 Consultez aussi

- [Gestion des balises LF pour le contrôle d'accès aux métadonnées](#)
- [Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC](#)

Rubriques

- [Gestion des balises LF pour le contrôle d'accès aux métadonnées](#)
- [Gestion des expressions LF-Tag pour le contrôle d'accès aux métadonnées](#)
- [Gestion des autorisations relatives aux valeurs des balises LF](#)

Gestion des balises LF pour le contrôle d'accès aux métadonnées

Pour utiliser la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) afin de sécuriser les objets du catalogue de données tels que les catalogues, les bases de données, les tables, les vues et les colonnes, vous créez des balises LF, vous les attribuez aux ressources et vous accordez des autorisations LF-Tag aux principaux.

Avant de pouvoir attribuer des balises LF aux objets du catalogue de données ou accorder des autorisations aux principaux, vous devez définir des balises LF. Seul un administrateur de lac de données ou un directeur disposant des autorisations de création de balises LF peut créer des balises LF.

Créateurs de LF-Tag

Le créateur de balises LF est un administrateur principal non administrateur autorisé à créer et à gérer des balises LF. Les administrateurs de data lake peuvent ajouter des créateurs de balises LF à l'aide de la console Lake Formation ou de la CLI. Les créateurs de balises LF ont des autorisations implicites de Lake Formation pour mettre à jour et supprimer les balises LF, pour attribuer des balises LF aux ressources et pour accorder des autorisations de balise LF et des autorisations de valeur de balise LF à d'autres directeurs.

Grâce aux rôles de créateur de balises LF, les administrateurs des lacs de données peuvent déléguer des tâches de gestion des balises, telles que la création et la mise à jour des clés et des valeurs des balises, à des personnes qui ne sont pas des administrateurs principaux. Les administrateurs de lacs de données peuvent également accorder aux créateurs de balises LF des autorisations susceptibles d'être accordées `Create LF-Tag`. Ensuite, le créateur de balises LF peut accorder l'autorisation de créer des balises LF à d'autres personnes principales.

Vous pouvez accorder deux types d'autorisations sur les LF-Tags :

- Autorisations LF-Tag - `Create LF-Tag``Alter`, et `Drop` Ces autorisations sont requises pour créer, mettre à jour et supprimer des balises LF.

Les administrateurs du lac de données et les créateurs de balises LF disposent implicitement de ces autorisations sur les balises LF qu'ils créent et peuvent accorder ces autorisations explicitement aux principaux pour gérer les balises dans le lac de données.

- Autorisations de paire clé-valeur LF-Tag - `Assign`, et `Describe Grant with LF-Tag` expressions Ces autorisations sont nécessaires pour attribuer des balises LF aux objets du catalogue de données et pour accorder des autorisations sur les ressources aux principaux à l'aide du contrôle d'accès basé sur les balises Lake Formation. Les créateurs de balises LF reçoivent implicitement ces autorisations lors de la création de balises LF.

Après avoir reçu l'`Create LF-Tag` autorisation et créé avec succès les balises LF, le créateur de balises LF peut attribuer des balises LF aux ressources et accorder des autorisations de balises LF (`Create LF-Tag`, `AlterDrop`, et) à d'autres entités non administratives pour gérer les balises dans le lac de données. Vous pouvez gérer les balises LF à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

 Note

Les administrateurs des lacs de données disposent des autorisations implicites de Lake Formation pour créer, mettre à jour et supprimer des balises LF, attribuer des balises LF aux ressources et accorder des autorisations LF-Tag aux principaux.

Pour connaître les meilleures pratiques et les considérations, voir [Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation](#)

Rubriques

- [Ajouter des créateurs de LF-Tag](#)
- [Création de balises LF](#)
- [Mise à jour des balises LF](#)
- [Suppression des balises LF](#)
- [Répertorier les balises LF](#)
- [Affectation de balises LF aux ressources du catalogue de données](#)
- [Afficher les balises LF attribuées à une ressource](#)
- [Afficher les ressources auxquelles un tag LF est attribué](#)
- [Cycle de vie d'un LF-Tag](#)
- [Comparaison entre le contrôle d'accès basé sur les balises Lake Formation et le contrôle d'accès basé sur les attributs IAM](#)

 Consultez aussi

- [Gestion des autorisations relatives aux valeurs des balises LF](#)
- [Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC](#)
- [Contrôle d'accès basé sur des balises Lake Formation](#)

Ajouter des créateurs de LF-Tag

Par défaut, les administrateurs des lacs de données peuvent créer, mettre à jour et supprimer des balises LF, attribuer des balises aux objets du catalogue de données et accorder des autorisations

de balises aux principaux. Si vous souhaitez déléguer les opérations de création et de gestion des balises à des administrateurs non administrateurs, l'administrateur du lac de données peut créer des rôles de créateur de balises LF et `Create LF-Tag` autoriser Lake Formation à accéder à ces rôles. Avec une `Create LF-Tag` autorisation pouvant être accordée, les créateurs de balises LF peuvent déléguer les tâches de création et de maintenance des balises à d'autres personnes non administratives.

Pour que les administrateurs de lacs de données puissent attribuer des balises LF aux ressources du catalogue de données, ils doivent s'octroyer des autorisations d'association sur les balises LF qu'ils n'ont pas créées.

Note

Les autorisations accordées entre comptes ne peuvent inclure que `Describe` des `Associate` autorisations. Vous ne pouvez pas accorder `Create LF-TagDrop`, `Alter`, et `Grant with LFTag` expressions autorisations aux principaux d'un autre compte.

Rubriques

- [Autorisations IAM requises pour créer des balises LF](#)
- [Ajouter des créateurs de LF-Tag](#)

Consultez aussi

- [Gestion des autorisations relatives aux valeurs des balises LF](#)
- [Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC](#)
- [Contrôle d'accès basé sur des balises Lake Formation](#)

Autorisations IAM requises pour créer des balises LF

Vous devez configurer les autorisations pour permettre à un directeur de Lake Formation de créer des balises LF. Ajoutez la déclaration suivante à la politique d'autorisation pour le principal qui doit être un créateur de balises LF.

 Note

Bien que les administrateurs des lacs de données disposent des autorisations implicites de Lake Formation pour créer, mettre à jour et supprimer des balises LF, pour attribuer des balises LF aux ressources et pour accorder des balises LF aux principaux, les administrateurs des lacs de données ont également besoin des autorisations IAM suivantes.

Pour de plus amples informations, veuillez consulter [Référence des personnalités de Lake Formation et des autorisations IAM](#).

```
{
  "Sid": "Transformational",
  "Effect": "Allow",
  "Action": [
    "lakeformation:AddLFTagsToResource",
    "lakeformation:RemoveLFTagsFromResource",
    "lakeformation:GetResourceLFTags",
    "lakeformation:ListLFTags",
    "lakeformation:CreateLFTag",
    "lakeformation:GetLFTag",
    "lakeformation:UpdateLFTag",
    "lakeformation>DeleteLFTag",
    "lakeformation:SearchTablesByLFTags",
    "lakeformation:SearchDatabasesByLFTags"
  ]
}
```

Les principaux qui attribuent des balises LF aux ressources et accordent des balises LF aux principaux doivent avoir les mêmes autorisations, à l'exception des autorisations, et. CreateLFTag UpdateLFTag DeleteLFTag

Ajouter des créateurs de LF-Tag

Un créateur de balise LF peut créer une balise LF, mettre à jour la clé et les valeurs de balise, supprimer des balises, associer des balises aux ressources du catalogue de données et accorder des autorisations sur les ressources du catalogue de données aux principaux à l'aide de la méthode LF-TBAC. Le créateur du LF-Tag peut également accorder ces autorisations aux principaux.

Vous pouvez créer des rôles de créateur de balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

console

Pour ajouter un créateur de balises LF

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du datalake.

2. Dans le volet de navigation, sous Permissions, sélectionnez LF-Tags and permissions.

Sur la page des balises LF et des autorisations, choisissez la section des créateurs de balises LF, puis choisissez Ajouter des créateurs de balises LF.

Add LF-Tag creators

LF-Tag creators can create and manage LF-Tags. [Learn more](#)

LF-Tag creator details

IAM users and roles
Add IAM users or roles.

Choose IAM principals to add

lf-developer X
User

Permission
Choose the permission to grant.
☒ Create LF-Tag

Grantable permission
Choose the permission that may be granted to others.
☒ Create LF-Tag

Cancel Add

3. Sur la page Ajouter des créateurs de balises LF, choisissez un rôle ou un utilisateur IAM disposant des autorisations requises pour créer des balises LF.
4. Activez Create LF-Tag la case à cocher d'autorisation.

5. (Facultatif) Pour permettre aux principaux sélectionnés d'accorder des Create LF-Tag autorisations aux principaux, choisissez Autorisation accordable. Create LF-Tag
6. Choisissez Ajouter.

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
  },
  "Resource": {
    "Catalog": {}
  },
  "Permissions": [
    "CreateLFTag"
  ],
  "PermissionsWithGrantOption": [
    "CreateLFTag"
  ]
}
```

Les autorisations disponibles pour un rôle de créateur de balises LF sont les suivantes :

Autorisations	Description
Drop	Un directeur disposant de cette autorisation sur un tag LF peut supprimer un tag LF du lac de données. Le principal obtient une Describe autorisation implicite sur toutes les valeurs de balise d'une ressource LF-Tag.
Alter	Un directeur disposant de cette autorisation sur une balise LF peut ajouter ou supprimer une valeur de balise à une balise LF. Le principal obtient une Alter autorisation implicite sur toutes les valeurs de balise d'une balise LF.
Describe	Un directeur disposant de cette autorisation sur un LF-Tag peut voir le LF-Tag et ses valeurs lorsqu'il attribue des LF-Tags à des ressources

Autorisations	Description
	ou accorde des autorisations sur des LF-Tags. Vous pouvez accorder une autorisation <code>Describe</code> sur toutes les valeurs clés ou sur des valeurs spécifiques.
<code>Associate</code>	Un directeur disposant de cette autorisation sur une balise LF peut attribuer la balise LF à une ressource de catalogue de données. Accorder <code>Associate</code> implicitement des subventions <code>Describe</code> .
<code>Grant with LF-Tag expression</code>	Un directeur disposant de cette autorisation sur un tag LF peut accorder des autorisations sur les ressources d'un catalogue de données à l'aide de la clé et des valeurs du tag LF. Accorder <code>Grant with LF-Tag expression</code> implicitement des subventions <code>Describe</code> .

Ces autorisations peuvent être accordées. Un directeur qui a obtenu ces autorisations avec l'option d'octroi peut les accorder à d'autres principaux.

Création de balises LF

Toutes les balises LF doivent être définies dans Lake Formation avant de pouvoir être utilisées. Une balise LF se compose d'une clé et d'une ou de plusieurs valeurs possibles pour la clé.

Une fois que l'administrateur du lac de données a configuré les autorisations IAM et Lake Formation requises pour le rôle de créateur de balises LF, le principal peut créer une balise LF. Le créateur de la balise LF obtient l'autorisation implicite de mettre à jour ou de supprimer toute valeur de balise de la balise LF et de supprimer la balise LF.

Vous pouvez créer des balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

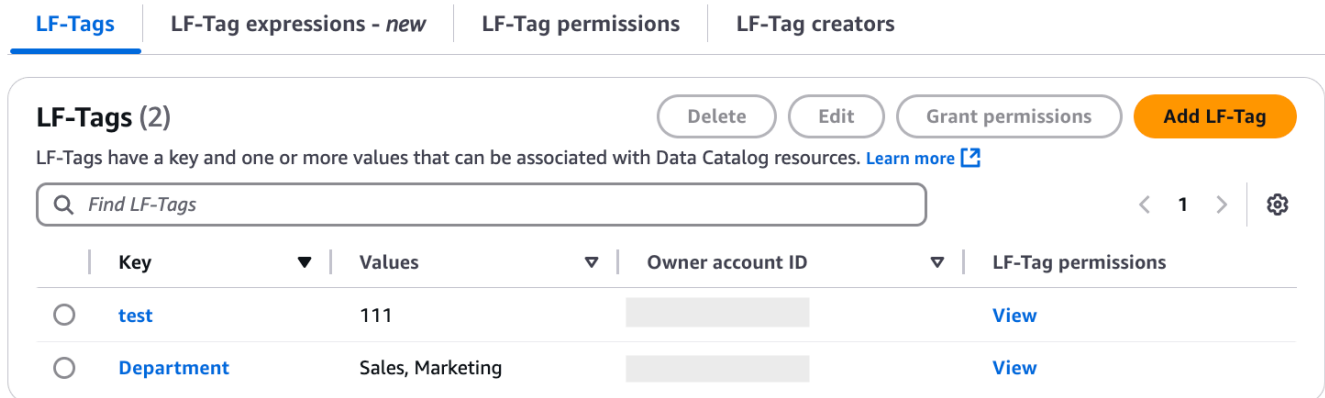
Pour créer un LF-Tag

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant que principal avec les autorisations de création de balises LF ou en tant qu'administrateur du lac de données.

2. Dans le volet de navigation, sous Permissions, LF-Tags et permissions, choisissez LF-Tags.

La page LF-Tags apparaît.



3. Choisissez Ajouter un tag LF.
4. Dans la boîte de dialogue Ajouter une balise LF, entrez une clé et une ou plusieurs valeurs.

Chaque clé doit avoir au moins une valeur. Pour saisir plusieurs valeurs, entrez une liste séparée par des virgules puis appuyez sur Entrée, ou entrez une valeur à la fois et choisissez Ajouter après chacune d'elles. Le nombre maximum de valeurs autorisées est de 1 000.

5. Choisissez Ajouter une balise.

AWS CLI

Pour créer un LF-Tag

- Entrez une `create-lf-tag` commande.

L'exemple suivant crée une balise LF avec une clé `module` et des valeurs `Customers` et `Orders`

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

En tant que créateur de balise, le principal obtient l'Altération autorisation d'utiliser cette balise LF et peut mettre à jour ou supprimer toute valeur de balise de cette balise LF. Le créateur principal du tag LF peut également Alter autoriser un autre principal à mettre à jour et à supprimer les valeurs des balises sur ce tag LF.

Mise à jour des balises LF

Vous mettez à jour une balise LF sur laquelle vous avez l'Alte^rautorisation en ajoutant ou en supprimant des valeurs clés autorisées. Vous ne pouvez pas modifier la touche LF-Tag. Pour modifier la clé, supprimez le tag LF et ajoutez-en un avec la clé requise. Outre l'Alte^rautorisation, vous avez également besoin de l'autorisation `lakeformation:UpdateLFTag` IAM pour mettre à jour les valeurs.

Lorsque vous supprimez une valeur de balise LF, aucune vérification n'est effectuée pour vérifier la présence de cette valeur de balise LF sur aucune ressource du catalogue de données. Si la valeur de balise LF supprimée est associée à une ressource, elle n'est plus visible pour cette ressource, et les principaux auxquels des autorisations ont été accordées sur cette paire clé-valeur ne disposent plus de ces autorisations.

Avant de supprimer une valeur de balise LF, vous pouvez éventuellement utiliser la [remove-lf-tags-from-resourcecommande](#) pour supprimer la balise LF des ressources du catalogue de données contenant la valeur que vous souhaitez supprimer, puis réétiqueter la ressource avec les valeurs que vous souhaitez conserver.

Seuls les administrateurs du lac de données, le créateur du tag LF et les principaux détenteurs d'Alte^rautorisations sur le tag LF peuvent mettre à jour un tag LF.

Vous pouvez mettre à jour une balise LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

Pour mettre à jour un LF-Tag (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, créateur du tag LF ou principal Alte^r autorisé à utiliser le tag LF.

2. Dans le volet de navigation, sous Permissions, LF-Tags et permissions, choisissez LF-Tags.
3. Sur la page Balises LF, sélectionnez une balise LF, puis choisissez Modifier.
4. Dans la boîte de dialogue Modifier le tag LF, ajoutez ou supprimez des valeurs du tag LF.

Pour ajouter plusieurs valeurs, dans le champ Valeurs, entrez une liste séparée par des virgules et appuyez sur Entrée, ou entrez une valeur à la fois ou choisissez Ajouter après chacune d'elles.

5. Choisissez Enregistrer.

AWS CLI

Pour mettre à jour un LF-Tag ()AWS CLI

- Entrez une `update-lf-tag` commande. Fournissez l'un des arguments suivants ou les deux :
 - `--tag-values-to-add`
 - `--tag-values-to-delete`

Exemple

L'exemple suivant remplace la valeur `vp` par la valeur de `vice-president` la clé LF-Tag. `level`

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

Suppression des balises LF

Vous pouvez supprimer les balises LF qui ne sont plus utilisées. Aucune vérification n'est effectuée pour vérifier la présence de la balise LF sur une ressource de catalogue de données. Si le tag LF supprimé est associé à une ressource, il n'est plus visible pour la ressource, et les principaux auxquels des autorisations ont été accordées sur ce tag LF ne le sont plus.

Avant de supprimer une balise LF, vous pouvez éventuellement utiliser la [remove-lf-tags-from-resource](#) commande pour supprimer la balise LF de toutes les ressources.

Seuls les administrateurs du lac de données, le créateur du tag LF ou un principal Drop autorisé à utiliser le tag LF peuvent supprimer un tag LF. Outre l'ID d'autorisation, le principal doit également disposer de l'autorisation `lakeformation:DeleteLFTag` IAM pour supprimer un tag LF.

Vous pouvez supprimer une balise LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

Console

Pour supprimer un tag LF (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données.

2. Dans le volet de navigation, sous Permissions, LF-Tags et permissions, choisissez LF-Tags.
3. Sur la page Balises LF, sélectionnez une balise LF, puis choisissez Supprimer.
4. Dans l'environnement Delete tag ? boîte de dialogue, pour confirmer la suppression, entrez la valeur de la clé LF-Tag dans le champ désigné, puis choisissez Supprimer.

AWS CLI

Pour supprimer un tag LF (AWS CLI)

- Entrez une `delete-lf-tag` commande. Indiquez la clé du tag LF à supprimer.

Exemple

L'exemple suivant supprime le tag LF avec la clé. `region`

```
aws lakeformation delete-lf-tag --tag-key region
```

Répertorier les balises LF

Vous pouvez répertorier les balises LF pour lesquelles vous avez les `Associate` autorisations `Describe` ou les autorisations. Les valeurs répertoriées avec chaque clé LF-Tag sont les valeurs pour lesquelles vous avez des autorisations.

Le créateur de balises LF dispose d'autorisations implicites pour voir les balises LF qu'il a créées.

Les administrateurs du data lake peuvent voir toutes les balises LF définies dans le AWS compte local et toutes les balises LF pour lesquelles les `Associate` autorisations `Describe` et les autorisations ont été accordées au compte local à partir de comptes externes. L'administrateur du lac de données peut voir toutes les valeurs de toutes les balises LF.

Vous pouvez répertorier les balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

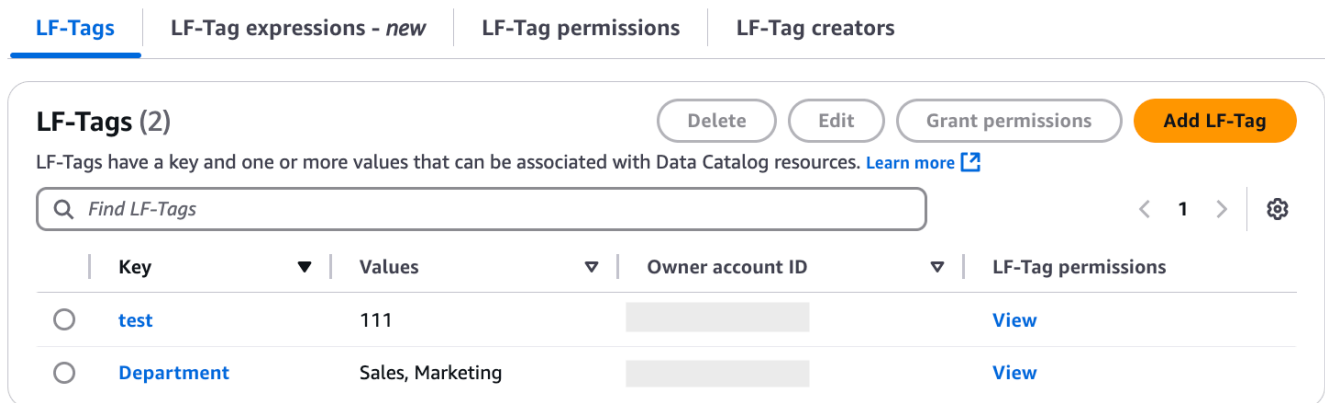
Pour répertorier les balises LF (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant que créateur de balises LF, en tant qu'administrateur du lac de données ou en tant que principal ayant obtenu des autorisations sur les balises LF et disposant de l'autorisation IAM. `lakeformation:ListLFTags`

2. Dans le volet de navigation, sous Permissions, LF-Tags et permissions, choisissez LF-Tags.

La page LF-Tags apparaît.



LF-Tags (2) Delete Edit Grant permissions Add LF-Tag

LF-Tags have a key and one or more values that can be associated with Data Catalog resources. [Learn more](#)

	Key	Values	Owner account ID	LF-Tag permissions
☐	test	111		View
☐	Department	Sales, Marketing		View

Consultez la colonne ID du compte propriétaire pour déterminer les balises LF qui ont été partagées avec votre compte depuis un compte externe.

AWS CLI

Pour répertorier les balises LF (AWS CLI)

- Exécutez la commande suivante en tant qu'administrateur du lac de données ou en tant que principal ayant obtenu des autorisations sur les balises LF et disposant de l'autorisation IAM. `lakeformation:ListLFTags`

```
aws lakeformation list-lf-tags
```

La sortie est similaire à ce qui suit.

```
{
  "LFTags": [
    {
      "CatalogId": "111122223333",
      "TagKey": "level",
      "TagValues": [
        "director",
        "vp",
        "c-level"
      ]
    },
    {
      "CatalogId": "111122223333",
      "TagKey": "module",
      "TagValues": [
        "Orders",
        "Sales",
        "Customers"
      ]
    }
  ]
}
```

Pour voir également les balises LF attribuées à partir de comptes externes, incluez l'option de commande. `--resource-share-type ALL`

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

La sortie est similaire à ce qui suit. Notez la `NextToken` clé, qui indique qu'il y en a d'autres à répertorier.

```
{
  "LFTags": [
    {
      "CatalogId": "111122223333",
      "TagKey": "level",
      "TagValues": [
        "director",
        "vp",
```

```

        "c-level"
      ]
    },
    {
      "CatalogId": "111122223333",
      "TagKey": "module",
      "TagValues": [
        "Orders",
        "Sales",
        "Customers"
      ]
    }
  ],
  "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}

```

Répétez la commande et ajoutez l'option `--next-token` pour afficher les balises LF locales restantes et les balises LF accordées par des comptes externes. Les balises LF des comptes externes se trouvent toujours sur une page séparée.

```

aws lakeformation list-lf-tags --resource-share-type ALL
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==

```

```

{
  "LFTags": [
    {
      "CatalogId": "123456789012",
      "TagKey": "region",
      "TagValues": [
        "central",
        "south"
      ]
    }
  ]
}

```

API

Vous pouvez utiliser le SDKs paramètre available for Lake Formation pour répertorier les balises que le demandeur est autorisé à consulter.

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50
)
```

Cette commande renvoie un dict objet dont la structure est la suivante :

```
{
  'LFTags': [
    {
      'CatalogId': 'string',
      'TagKey': 'string',
      'TagValues': [
        'string',
      ]
    },
  ],
  'NextToken': 'string'
}
```

Pour plus d'informations sur les autorisations requises, consultez [Référence des personnalités de Lake Formation et des autorisations IAM](#).

Affectation de balises LF aux ressources du catalogue de données

Vous pouvez attribuer des balises LF aux ressources du catalogue de données (bases de données, tables et colonnes) pour contrôler l'accès à ces ressources. Seuls les principaux auxquels des balises LF correspondantes sont attribuées (et les principaux auxquels l'accès est accordé avec la méthode de ressource nommée) peuvent accéder aux ressources.

Si une table hérite d'une balise LF d'une base de données ou si une colonne hérite d'une balise LF d'une table, vous pouvez remplacer la valeur héritée en affectant une nouvelle valeur à la clé de balise LF.

Le nombre maximum de balises LF que vous pouvez attribuer à une ressource est de 50.

Rubriques

- [Exigences relatives à la gestion des balises attribuées aux ressources](#)
- [Attribuer des balises LF à une colonne de tableau](#)
- [Attribuer des balises LF à une ressource de catalogue de données](#)
- [Mettre à jour les balises LF d'une ressource](#)
- [Supprimer le tag LF d'une ressource](#)

Exigences relatives à la gestion des balises attribuées aux ressources

Pour attribuer une balise LF à une ressource du catalogue de données, vous devez :

- Ayez l'ASSOCIATE autorisation de Lake Formation sur le LF-Tag.
- Ayez l'`lakeformation:AddLFTagsToResource` autorisation IAM.
- Avoir de la colle : GetDatabase autorisation sur une base de données Glue.
- Soyez le propriétaire (créateur) de la ressource, Super détenez l'autorisation Lake Formation sur la ressource avec l'GRANT option, ou disposez des autorisations suivantes avec l'GRANT option :
 - Pour les bases de données du même AWS compte : DESCRIBE, CREATE_TABLE, ALTER, et DROP
 - Pour les bases de données d'un compte externe : DESCRIBE, CREATE_TABLE et ALTER
 - Pour les tables (et les colonnes) : DESCRIBE, ALTER, DROP, INSERT, SELECT, et DELETE

De plus, le tag LF et la ressource à laquelle il est attribué doivent se trouver dans le même AWS compte.

Pour supprimer une balise LF d'une ressource de catalogue de données, vous devez répondre à ces exigences et disposer de l'autorisation `lakeformation:RemoveLFTagsFromResource` IAM.

Attribuer des balises LF à une colonne de tableau

Pour attribuer des balises LF à une colonne de tableau (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'utilisateur répondant aux exigences répertoriées ci-dessus.

2. Dans le volet de navigation, choisissez Tables.
3. Choisissez un nom de table (et non le bouton d'option situé à côté du nom de la table).

4. Sur la page des détails de la table, dans la section Schéma, choisissez Modifier le schéma.
5. Sur la page Modifier le schéma, sélectionnez une ou plusieurs colonnes, puis choisissez Modifier les balises LF.

 Note

Si vous avez l'intention d'ajouter ou de supprimer des colonnes et d'enregistrer une nouvelle version, faites-le d'abord. Modifiez ensuite les balises LF.

La boîte de dialogue Modifier les balises LF apparaît et affiche toutes les balises LF héritées du tableau.

 **Edit LF-Tags: product_id** [Learn More](#) 

LF-Tags

After they are associated with catalog resources, LF-Tags allow you to create scalable permissions.

Inherited keys	Values
level	director (inherited) ▼
module	Orders (inherited) ▼

Assign new LF-Tag

You can add 50 more tags.

Cancel

Save

6. (Facultatif) Dans la liste des valeurs située à côté d'un champ Clés héritées, choisissez une valeur qui remplacera la valeur héritée.
7. (Facultatif) Choisissez Attribuer un nouveau tag LF. Ensuite, pour Clés attribuées, choisissez une clé, et pour Valeurs, choisissez une valeur pour la clé.

Edit LF-Tags: product_id [Learn More](#) 

LF-Tags

After they are associated with catalog resources, LF-Tags allow you to create scalable permissions.

Inherited keys

Values

director (inherited) ▼

Orders (inherited) ▼

Assigned keys



Assign new LF-Tag

You can add 49 more tags.

Values

Production ▲

Production

Development

Remove

Cancel

Save

8. (Facultatif) Choisissez à nouveau Attribuer un nouveau tag LF pour ajouter un autre tag LF.
9. Choisissez Enregistrer.

Attribuer des balises LF à une ressource de catalogue de données

Console

Pour attribuer des balises LF à une base de données ou à une table de catalogue de données

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'utilisateur répondant aux exigences répertoriées précédemment.

2. Dans le volet de navigation, sous Catalogue de données, effectuez l'une des opérations suivantes :
 - Pour attribuer des balises LF aux bases de données, choisissez Databases.

- Pour attribuer des balises LF aux tables, choisissez Tables.
3. Choisissez une base de données ou une table, puis dans le menu Actions, choisissez Modifier les balises LF.

La boîte de *resource-name* dialogue Modifier les balises LF : apparaît.

Si une table hérite des balises LF de la base de données qui la contient, la fenêtre affiche les balises LF héritées. Sinon, le texte « Aucune balise LF héritée n'est associée à la ressource » s'affiche.

Edit LF-Tags: inventory [Learn More](#) 

LF-Tags
After they are associated with catalog resources, LF-Tags allow you to create scalable permissions.

Inherited keys

Values

director (inherited) ▼

Assigned keys



Assign new LF-Tag
You can add 49 more tags.

Values

Enter LF-Tag value ▲

Orders

Sales

Customers

Remove

Cancel

Save

4. (Facultatif) Si une table possède des balises LF héritées, dans la liste des valeurs située à côté d'un champ Clés héritées, vous pouvez choisir une valeur qui remplacera la valeur héritée.
5. Pour attribuer de nouvelles balises LF, procédez comme suit :
 - a. Choisissez Attribuer un nouveau tag LF.
 - b. Dans le champ Clés assignées, choisissez une clé LF-Tag, et dans le champ Valeurs, choisissez une valeur.

- c. (Facultatif) Choisissez à nouveau Assigner un nouveau tag LF pour attribuer un tag LF supplémentaire.
6. Choisissez Enregistrer.

AWS CLI

Pour attribuer des balises LF à une ressource de catalogue de données

- Exécutez la commande `add-lf-tags-to-resource`.

L'exemple suivant attribue la balise LF à la table `module=orders` de la base de données `orders.erp`. Il utilise la syntaxe du raccourci pour l'argument `--lf-tags`. La propriété `CatalogId` pour `--lf-tags` est facultative. S'il n'est pas fourni, l'ID de catalogue de la ressource (dans ce cas, la table) est supposé.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":  
  {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags  
CatalogId=111122223333,TagKey=module,TagValues=orders
```

Le résultat suivant est le résultat si la commande aboutit.

```
{  
  "Failures": []  
}
```

L'exemple suivant affecte deux balises LF à la table `sales` et utilise la syntaxe JSON pour l'argument `--lf-tags`.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":  
  {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":  
  "module", "TagValues": ["sales"]}, {"TagKey": "environment", "TagValues":  
  ["development"]}']
```

L'exemple suivant affecte la balise LF `level=director` à la colonne `total` de la table `sales`.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":  
{"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags  
TagKey=level,TagValues=director
```

Mettre à jour les balises LF d'une ressource

Pour mettre à jour une balise LF pour une ressource de catalogue de données ()AWS CLI

- Utilisez la `add-lf-tags-to-resource` commande, comme décrit dans la procédure précédente.

L'ajout d'un tag LF avec la même clé qu'un tag LF existant, mais avec une valeur différente met à jour la valeur existante.

Supprimer le tag LF d'une ressource

Pour supprimer une balise LF pour une ressource de catalogue de données ()AWS CLI

- Exécutez la commande `remove-lf-tags-from-resource`.

Si une table possède une valeur de balise LF qui remplace la valeur héritée de la base de données parent, la suppression de cette balise LF de la table restaure la valeur héritée. Ce comportement s'applique également à une colonne qui remplace les valeurs clés héritées de la table.

L'exemple suivant supprime le tag LF `level=director` de la `total` colonne du `sales` tableau. La `CatalogID` propriété pour `--lf-tags` est facultative. S'il n'est pas fourni, l'ID de catalogue de la ressource (dans ce cas, la table) est supposé.

```
aws lakeformation remove-lf-tags-from-resource  
--resource '{ "TableWithColumns":  
{ "DatabaseName": "erp", "Name": "sales", "ColumnNames": [ "total" ] } }'  
--lf-tags CatalogId=111122223333,TagKey=level,TagValues=director
```

Afficher les balises LF attribuées à une ressource

Vous pouvez afficher les balises LF attribuées à une ressource de catalogue de données. Vous devez disposer de l'ASSOCIATE autorisation DESCRIBE or sur un LF-Tag pour le visualiser.

Console

Pour afficher les balises LF attribuées à une ressource (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, propriétaire de la ressource ou utilisateur ayant obtenu les autorisations de Lake Formation sur la ressource.

2. Dans le volet de navigation, sous le titre Catalogue de données, effectuez l'une des opérations suivantes :
 - Pour afficher les balises LF attribuées à une base de données, sélectionnez Databases.
 - Pour afficher les balises LF attribuées à une table, choisissez Tables.
3. Sur la page Tables ou bases de données, choisissez le nom de la base de données ou de la table. Ensuite, sur la page de détails, faites défiler la page vers le bas jusqu'à la section LF-Tags.

La capture d'écran suivante montre les balises LF attribuées à une `customers` table contenue dans la `retail` base de données. La module balise LF est héritée de la base de données. Le `level=vp` tag LF est attribué à la `credit_limit` colonne.

LF-Tags (3)

[Edit tags](#)

LF-Tags are key-value pairs that you can assign to data catalog resources, such as databases, tables, and columns. You can then grant permissions to principals based on these tags to control access to the resources. Table columns inherit all LF-Tags that are assigned to the table. [Learn More](#) 

< 1 > 

Resource ▲	Key ▼	Value ▼	Inherited from
customers (table)	module	Customers	retail
customers (table)	environment	Production	-
credit_limit (column)	level	vp	-

AWS CLI

Pour afficher les balises LF attribuées à une ressource ()AWS CLI

- Utilisez une commande similaire à la suivante.

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
"Name":"sales"}}'
```

La commande renvoie le résultat suivant.

```
{
  "TableTags": [
    {
      "CatalogId": "111122223333",
      "TagKey": "module",
      "TagValues": [
        "sales"
      ]
    },
    {
```

```

        "CatalogId": "111122223333",
        "TagKey": "environment",
        "TagValues": [
            "development"
        ]
    },
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}

```

Cette sortie affiche uniquement les balises LF attribuées de manière explicite et non héritées. Si vous souhaitez voir toutes les balises LF sur toutes les colonnes, y compris les balises LF héritées, omettez cette option. `--show-assigned-lf-tags`

Afficher les ressources auxquelles un tag LF est attribué

Vous pouvez afficher toutes les ressources du catalogue de données auxquelles une clé LF-Tag particulière est attribuée. Pour ce faire, vous devez disposer des autorisations Lake Formation suivantes :

- `Describeou Associate` sur le LF-Tag.
- `Describeou` toute autre autorisation de Lake Formation sur la ressource.

En outre, vous devez disposer des autorisations Gestion des identités et des accès AWS (IAM) suivantes :

- `lakeformation:SearchDatabasesByLFTags`

- `lakeformation:SearchTablesByLFTags`

Console

Pour afficher les ressources auxquelles un tag LF est attribué (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur répondant aux exigences répertoriées précédemment.

2. Dans le volet de navigation, sous Permissions et balises LF et autorisations, choisissez LF-Tags.
3. Choisissez une touche LF-Tag (et non le bouton d'option à côté du nom de la clé).

La page de détails du LF-Tag affiche une liste des ressources auxquelles le LF-Tag a été attribué.

module

LF-Tag

DeleteEdit

Key

module

Values

Orders, Sales, Customers

Associated data catalog resources (12)

Q Find resource

Key	Values ▾	Resource type ▾	Resource ▾
module	Customers	DATABASE	retail
module	Customers	TABLE	customers
module	Orders	TABLE	inventory
module	Customers	COLUMN	customers.cust_first_name
module	Customers	COLUMN	customers.work_phone_number
module	Customers	COLUMN	customers.company_name
module	Customers	COLUMN	customers.credit_limit

AWS CLI

Pour afficher les ressources auxquelles un tag LF est attribué

- Exécutez une `search-databases-by-lf-tags` commande `search-tables-by-lf-tags` or.

Exemple

L'exemple suivant répertorie les tables et les colonnes auxquelles le `level=vp` tag LF est attribué. Pour chaque table ou colonne répertoriée, toutes les balises LF attribuées à la table ou à la colonne sont affichées, et pas seulement l'expression de recherche.

```
aws lakeformation search-tables-by-lf-tags --expression
TagKey=level,TagValues=vp
```

Pour plus d'informations sur les autorisations requises, consultez [Référence des personnalités de Lake Formation et des autorisations IAM](#).

Cycle de vie d'un LF-Tag

1. Michael, le créateur du LF-Tag, crée un LF-Tag. `module=Customers`
2. Michael octroie Associate le LF-Tag à l'ingénieur de données Eduardo. Accorder Associate implicitement des subventionsDescribe.
3. Michael accorde une subvention Super sur la table Custs à Eduardo avec l'option de subvention, afin qu'Eduardo puisse attribuer des balises LF à la table. Pour de plus amples informations, veuillez consulter [Affectation de balises LF aux ressources du catalogue de données](#).
4. Eduardo attribue le LF-tag à `module=customers` la table. Custs
5. Michael accorde la subvention suivante à l'ingénieure de données Sandra (en pseudo-code).

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra accorde la subvention suivante à l'analyste de données Maria.

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Maria peut désormais exécuter des requêtes sur la Custs table.

 Consultez aussi

- [Contrôle d'accès aux métadonnées](#)

Comparaison entre le contrôle d'accès basé sur les balises Lake Formation et le contrôle d'accès basé sur les attributs IAM

Le contrôle d'accès par attributs (ABAC) est une stratégie d'autorisation qui définit des autorisations en fonction des attributs. Dans AWS, ces attributs sont appelés balises. Vous pouvez associer des balises aux ressources IAM, notamment aux entités IAM (utilisateurs ou rôles) et aux AWS ressources. Vous pouvez créer une seule politique ABAC ou un petit nombre de politiques pour vos principaux IAM. Ces politiques ABAC sont conçues pour autoriser des opérations lorsque la balise du principal correspond à celle de la ressource. L'ABAC est utile dans les environnements qui connaissent une croissance rapide et pour les cas où la gestion des politiques devient fastidieuse.

Les équipes de sécurité et de gouvernance du cloud utilisent IAM pour définir des politiques d'accès et des autorisations de sécurité pour toutes les ressources, y compris les compartiments Amazon S3, les instances Amazon EC2 et toutes les ressources auxquelles vous pouvez faire référence avec un ARN. Les politiques IAM définissent des autorisations générales (grossières) sur les ressources de votre lac de données, par exemple, pour autoriser ou refuser l'accès au niveau du compartiment ou du préfixe Amazon S3 ou au niveau de la base de données. Pour plus d'informations sur IAM ABAC, voir [À quoi sert ABAC ? AWS](#) dans le guide de l'utilisateur IAM.

Par exemple, vous pouvez créer trois rôles avec la clé de balise `project-access`. Définissez la valeur de la balise du premier rôle sur `Dev`, celle du deuxième sur `Marketing`, et celle du troisième sur `Support`. Attribuez des balises avec la valeur appropriée aux ressources. Vous pouvez alors utiliser une seule politique qui autorise l'accès lorsque le rôle et la ressource sont balisés avec la même valeur pour `project-access`.

Les équipes de gouvernance des données utilisent Lake Formation pour définir des autorisations précises pour des ressources de lacs de données spécifiques. Les balises LF sont attribuées aux ressources du catalogue de données (bases de données, tables et colonnes) et sont accordées aux principaux. Un principal dont les balises LF correspondent aux balises LF d'une ressource peut accéder à cette ressource. Les autorisations de Lake Formation sont secondaires aux autorisations IAM. Par exemple, si les autorisations IAM n'autorisent pas un utilisateur à accéder à un lac de données, Lake Formation n'accorde l'accès à aucune ressource de ce lac de données à cet utilisateur, même si le principal et la ressource ont des balises LF identiques.

Le contrôle d'accès basé sur des balises Lake Formation (LF-TBAC) fonctionne avec IAM ABAC pour fournir des niveaux d'autorisations supplémentaires pour vos données et ressources de Lake Formation.

- Les autorisations TBAC de Lake Formation évoluent avec l'innovation. L'administrateur n'a plus besoin de mettre à jour les politiques existantes pour autoriser l'accès aux nouvelles ressources. Supposons, par exemple, que vous utilisiez une stratégie IAM ABAC avec la `project-access` balise pour fournir un accès à des bases de données spécifiques au sein de Lake Formation. À l'aide du LF-TBAC, le tag `LF Project=SuperApp` est attribué à des tables ou à des colonnes spécifiques, et le même tag LF est accordé à un développeur pour ce projet. Grâce à IAM, le développeur peut accéder à la base de données, et les autorisations LF-TBAC lui permettent d'accéder davantage à des tables ou à des colonnes spécifiques au sein de tables. Si une nouvelle table est ajoutée au projet, l'administrateur de Lake Formation n'a qu'à attribuer la balise à la nouvelle table pour que le développeur puisse y accéder.
- Lake Formation TBAC nécessite moins de politiques IAM. Comme vous utilisez les politiques IAM pour accorder un accès de haut niveau aux ressources de Lake Formation et le TBAC de Lake Formation pour gérer un accès aux données plus précis, vous créez moins de politiques IAM.
- Grâce au Lake Formation TBAC, les équipes peuvent changer et se développer rapidement. En effet, les autorisations d'accès aux nouvelles ressources sont automatiquement accordées en fonction des attributs. Par exemple, si un nouveau développeur rejoint le projet, il est facile de lui accorder l'accès en associant le rôle IAM à l'utilisateur, puis en lui attribuant les balises LF requises. Il n'est pas nécessaire de modifier la politique IAM pour prendre en charge un nouveau projet ou pour créer de nouveaux LF-Tags.
- Des autorisations plus précises sont possibles grâce au Lake Formation TBAC. Les politiques IAM accordent l'accès aux ressources de haut niveau, telles que les bases de données ou les tables du catalogue de données. À l'aide de Lake Formation TBAC, vous pouvez autoriser l'accès à des tables ou à des colonnes spécifiques contenant des valeurs de données spécifiques.

Note

Les balises IAM sont différentes des balises LF. Ces étiquettes ne sont pas interchangeables. Les balises LF sont utilisées pour accorder des autorisations à Lake Formation et les balises IAM sont utilisées pour définir les politiques IAM.

Gestion des expressions LF-Tag pour le contrôle d'accès aux métadonnées

Les expressions LF-Tag sont des expressions logiques composées d'une ou plusieurs balises LF (paires clé-valeur) utilisées pour octroyer des autorisations sur les ressources. AWS Glue Data Catalog Les expressions LF-Tag vous permettent de définir des règles qui régissent l'accès à vos

ressources de données en fonction de leurs balises de métadonnées. Vous pouvez enregistrer ces expressions et les réutiliser dans le cadre de plusieurs autorisations accordées, afin de garantir la cohérence et de faciliter la gestion des modifications apportées à l'ontologie des balises au fil du temps.

Dans une expression LF-Tag donnée, les clés des balises sont combinées à l'aide de l'opération AND, tandis que les valeurs sont combinées à l'aide de l'opération OR. Par exemple, l'expression de balise `content_type:Sales AND location:US` représente les ressources liées aux données de vente aux États-Unis.

Vous pouvez créer jusqu'à 1 000 expressions LF-Tag dans un. Compte AWS Ces expressions constituent un moyen flexible et évolutif de gérer les autorisations en fonction des balises de métadonnées, en garantissant que seuls les utilisateurs ou applications autorisés peuvent accéder à des ressources de données spécifiques en fonction des règles de balises définies.

Les expressions LF-Tag offrent les avantages suivants :

- **Réutilisabilité** — En définissant et en enregistrant des expressions LF-Tag, vous n'avez plus besoin de répliquer manuellement les mêmes expressions lorsque vous attribuez des autorisations à d'autres ressources ou à d'autres principaux.
- **Cohérence** — La réutilisation des expressions LF-Tag pour plusieurs autorisations garantit la cohérence dans la manière dont les autorisations sont accordées et gérées.
- **Gestion des ontologies de balises** : les expressions de balises LF aident à gérer les modifications apportées à l'ontologie des balises au fil du temps, car vous pouvez mettre à jour les expressions enregistrées au lieu de modifier les autorisations accordées individuellement.

Pour plus d'informations sur le contrôle d'accès basé sur des balises, veuillez consulter le [Contrôle d'accès basé sur des balises Lake Formation](#).

Créateurs d'expressions LF-Tag

Le créateur d'expressions LF-Tag est un directeur autorisé à créer et à gérer des expressions LF-Tag. Les administrateurs de data lake peuvent ajouter des créateurs d'expressions LF-Tag à l'aide de la console, de la CLI, de l'API ou du SDK Lake Formation. Les créateurs d'expressions LF-Tag disposent des autorisations implicites de Lake Formation pour créer, mettre à jour et supprimer des expressions LF-Tag, et pour accorder des autorisations d'expression LF-Tag à d'autres principaux.

Les créateurs d'expressions LF-Tag qui ne sont pas des administrateurs de lacs de données reçoivent des `Grant with LF-Tag expression` autorisations implicites `AlterDrop,Describe`, et uniquement pour les expressions qu'ils ont créées.

Les administrateurs de data lake peuvent également accorder aux créateurs d'expressions LF-Tag des autorisations susceptibles d'être accordées `Create LF-Tag expression`. Le créateur de l'expression LF-Tag peut ensuite accorder l'autorisation de créer des expressions LF-Tag à d'autres principaux.

Rubriques

- [Autorisations IAM requises pour créer des expressions LF-Tag](#)
- [Ajouter des créateurs d'expressions LF-Tag](#)
- [Création d'expressions LF-Tag](#)
- [Mise à jour des expressions LF-Tag](#)
- [Suppression d'expressions de balises LF](#)
- [Répertorier les expressions de balises LF](#)

Consultez aussi

- [Gestion des autorisations relatives aux valeurs des balises LF](#)
- [Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC](#)
- [Contrôle d'accès basé sur des balises Lake Formation](#)

Autorisations IAM requises pour créer des expressions LF-Tag

Vous devez configurer les autorisations pour permettre à un directeur de Lake Formation de créer des expressions LF-Tag. Ajoutez l'instruction suivante à la politique d'autorisation pour le principal qui doit être un créateur d'expressions LF-Tag.

Note

Bien que les administrateurs de lacs de données disposent d'autorisations implicites de Lake Formation pour créer, mettre à jour et supprimer des balises LF et des expressions de balises LF, pour attribuer des balises LF aux ressources et pour accorder des autorisations

de balises LF et d'expression de balises LF aux principaux, les administrateurs de lacs de données ont également besoin des autorisations IAM suivantes.

Pour de plus amples informations, veuillez consulter [Référence des personnalités de Lake Formation et des autorisations IAM](#).

```
{
  "Sid": "Transformational",
  "Effect": "Allow",
  "Action": [
    "lakeformation:AddLFTagsToResource",
    "lakeformation:RemoveLFTagsFromResource",
    "lakeformation:GetResourceLFTags",
    "lakeformation:ListLFTags",
    "lakeformation:CreateLFTag",
    "lakeformation:GetLFTag",
    "lakeformation:UpdateLFTag",
    "lakeformation>DeleteLFTag",
    "lakeformation:SearchTablesByLFTags",
    "lakeformation:SearchDatabasesByLFTags",
    "lakeformation:CreateLFTagExpression",
    "lakeformation>DeleteLFTagExpression",
    "lakeformation:UpdateLFTagExpression",
    "lakeformation:GetLFTagExpression",
    "lakeformation:ListLFTagExpressions",
    "lakeformation:GrantPermissions",
    "lakeformation:RevokePermissions",
    "lakeformation:BatchGrantPermissions",
    "lakeformation:BatchRevokePermissions"
  ]
}
```

Ajouter des créateurs d'expressions LF-Tag

Les créateurs d'expressions LF-Tag peuvent créer et enregistrer des expressions LF-Tag réutilisables, mettre à jour la clé et les valeurs des balises, supprimer des expressions et accorder des autorisations sur les ressources du catalogue de données aux principaux à l'aide de la méthode LF-TBAC. Le créateur de l'expression LF-Tag peut également accorder ces autorisations aux principaux.

Vous pouvez créer des rôles de créateur d'expressions LF-Tag à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

console

Pour ajouter un créateur d'expressions LF-Tag

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données.

2. Dans le volet de navigation, sous Permissions, sélectionnez LF-Tags and permissions.
3. Choisissez l'onglet Expressions LF-Tag.
4. Dans la section Créateurs d'expressions LF-Tag, choisissez Ajouter des créateurs d'expressions LF-Tag.

Add LF-Tag expression creators

LF-Tag expression creators can create and manage LF-Tags expressions.

LF-Tag expression creator details

IAM users and roles
Add IAM users or roles.

Choose IAM principals to add ▼

datalake_user X

User

Permission
Choose the permission to grant.

☒ Create LF-Tag expression

Grantable permission
Choose the permission that may be granted to others.

☒ Create LF-Tag expression

[Cancel](#)[Add](#)

5. Sur la page Ajouter des créateurs d'expressions LF-Tag, choisissez un rôle ou un utilisateur IAM disposant des autorisations requises pour créer des expressions LF-Tag.
6. Cochez Create LF-Tag expression la case d'autorisation.

7. (Facultatif) Pour permettre aux principaux sélectionnés d'accorder des Create LF-Tag expression autorisations aux principaux, choisissez Autorisation accordable. Create LF-Tag expression
8. Choisissez Ajouter.

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
  },
  "Resource": {
    "Catalog": {}
  },
  "Permissions": [
    "CreateLFTagExpression"
  ],
  "PermissionsWithGrantOption": [
    "CreateLFTagExpression"
  ]
}
```

Le rôle de créateur d'expressions LF-Tag permet de créer, mettre à jour ou supprimer des expressions LF-Tag.

Autorisations	Description
Create	Un directeur disposant de cette autorisation peut ajouter des expressions LF-Tag dans le lac de données.
Drop	Un principal disposant de cette autorisation sur une expression LF-Tag peut supprimer une expression LF-Tag du lac de données.
Alter	Un principal disposant de cette autorisation sur une expression LF-Tag peut mettre à jour le corps de l'expression d'une expression LF-Tag.

Autorisations	Description
Describe	Un principal disposant de cette autorisation sur une expression LF-Tag peut voir le contenu d'une expression LF-Tag.
Grant with LF-Tag expression	Cette autorisation permet au destinataire d'utiliser l'expression du tag comme ressource lorsqu'il accorde des autorisations d'accès aux données ou aux métadonnées. Accorder Grant with LF-Tag expression implicitement des subventionsDescribe.
Super	Pour les expressions de balise LF, l'Superautorisation donne la possibilité deDescribe, AlterDrop, et d'accorder des autorisations sur l'expression de balise à d'autres principaux.

Ces autorisations peuvent être accordées. Un directeur qui a obtenu ces autorisations avec l'option d'octroi peut les accorder à d'autres principaux.

Création d'expressions LF-Tag

Vous devez définir toutes les balises LF de Lake Formation et les affecter aux ressources du catalogue de données avant de pouvoir les utiliser pour créer des expressions. Une expression LF-Tag se compose d'une clé supplémentaire et d'une ou de plusieurs valeurs possibles pour chaque clé.

Une fois que l'administrateur du lac de données a configuré les autorisations IAM et Lake Formation requises pour le rôle de créateur d'expressions LF-Tag, le principal peut créer des expressions LF-Tag réutilisables. Le créateur de l'expression LF-Tag obtient des autorisations implicites pour mettre à jour le corps de l'expression et supprimer l'expression LF-Tag.

Vous pouvez créer des expressions LF-Tag à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

Pour créer une expression LF-Tag

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant que principal avec les autorisations de création d'expressions LF-Tag ou en tant qu'administrateur du lac de données.

2. Dans le volet de navigation, sous Permissions, sélectionnez LF-Tags and permissions.
3. Choisissez les expressions LF-Tag. La page Ajouter des expressions de balise LF apparaît.

Add LF-Tag Expression

LF-Tag expression creators can create and manage LF-Tag expressions

Expression details

Name
Enter a name that describes the expression. Expression name cannot be edited after creation.

Name must be less than 1000 characters.

Description - optional

Description can be up to 2048 characters.

Expression
Choose the keys and values for this expression. When multiple keys are specified, the keys are joined by an AND operator and when multiple values are specified, the values are joined by an OR operator.

Key	Values
Department	Choose LF-Tag values
	sales

[Add LF-Tag key-value pair](#)

You can add 49 more LF-Tags.

Expression review
The LF-Tag expression above will be interpreted in the following way.

Department = sales

► **Grant permissions - optional**
Grant permissions to allow others to manage or grant permissions with the expression.

[Cancel](#)[Add](#)

4. Entrez les informations suivantes :

- Nom — Entrez un nom unique pour l'expression. Vous ne pouvez pas mettre à jour le nom de l'expression.

- **Description** — Fournissez une description facultative de l'expression avec les détails de l'expression.
- **Expression** — Créez l'expression en spécifiant les clés de balise et leurs valeurs associées. Vous pouvez ajouter jusqu'à 50 clés par expression. Vous devez disposer de l'autorisation de `Grant with LF-Tags` Lake Formation sur toutes les balises du corps de l'expression.

Chaque clé doit avoir au moins une valeur. Pour saisir plusieurs valeurs, entrez une liste séparée par des virgules puis appuyez sur Entrée, ou entrez une valeur à la fois et choisissez Ajouter après chacune d'elles. Le nombre maximum de valeurs autorisées par clé est de 1 000.

Lake Formation utilise la AND/OR logique pour combiner plusieurs clés et valeurs dans une expression. Au sein d'une seule paire (clé : liste de valeurs), les valeurs sont combinées à l'aide de l'opérateur logique OR. Par exemple, si la paire est (Department : [Sales, Marketing]), cela signifie que le tag correspond si la ressource possède le tag Department avec la valeur Sales OR Marketing.

Lorsque vous spécifiez plusieurs clés, celles-ci sont jointes par un opérateur logique AND. Ainsi, si l'expression complète est (Department : [Sales, Marketing]) AND (Location : [États-Unis, Canada]), elle fait correspondre les ressources dont le tag Department a la valeur Sales OR Marketing, ET dont le tag Location a la valeur US OR Canada. Voici un autre exemple avec plusieurs clés et valeurs :

Expression LF-Tag : (ContentType : [Vidéo, audio]) ET (Région : [Europe, Asie]) ET (Département : [Ingénierie, ProductManagement]).

Cette expression correspondrait aux ressources qui ont : - La ContentType balise avec la valeur Video OR Audio AND - La balise Region avec la valeur Europe OU Asie ET - La balise Department avec la valeur Engineering OR ProductManagement.

Vous pouvez également enregistrer une expression de balise lorsque vous accordez des autorisations de lac de données à l'aide de balises LF. Choisissez les paires clé/valeur, puis sélectionnez l'option Enregistrer en tant que nouvelle expression. Entrez un nom qui décrit l'expression.

LF-Tags or catalog resources

Choose a method to grant permissions.

☒ **Resources matched by LF-Tags (recommended)**

Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☐ **Named Data Catalog resources**

Manage permissions for specific databases or tables, in addition to fine-grained data access.

☒ **LF-Tag key-value pairs**

☐ **Saved LF-Tag expressions - *new***

Key

Department ▼

Values

Choose LF-Tag values ▼

marketing ✕

sales ✕

Add LF-Tag key-value pair

You can add 49 more LF-Tags.

Expression review

The LF-Tag expression above will be interpreted in the following way.

Department = (marketing OR sales)

☒ **Save as new expression**

Use saved expressions to grant permissions. Create LF-Tag expression permissions are needed.

New LF-Tag expression name

Enter a name that describes the expression. Expression name cannot be edited after creation.

General access

Name must be less than 1000 characters.

5. (Facultatif) Choisissez ensuite les utilisateurs/rôles et les autorisations sur l'expression que vous souhaitez leur accorder dans le compte. Vous pouvez également choisir des autorisations pouvant être accordées qui permettent aux utilisateurs d'accorder ces autorisations à d'autres utilisateurs du compte. Vous ne pouvez pas accorder d'autorisations entre comptes sur les expressions de balise.

▼ Grant permissions - optional

Grant permissions to allow others to manage or grant permissions with the expression.

IAM users and roles

Users or roles from this AWS account.

Choose IAM principals to add

Permissions

Choose the specific LF-Tag permissions to grant.

- ☐ **Describe**
See keys and values.
- ☐ **Alter**
Update or delete LF-Tag expressions.
- ☐ **Drop**
Delete LF-Tag expressions.
- ☐ **Grant with LF-Tag expression**
Allow principals to grant access permissions using LF-Tag expressions.
- ☐ **Super**
This permission supersedes the individual permissions set above.

Grantable permissions

Choose the permissions that the recipient can grant to other principals.

- ☐ **Describe**
See keys and values.
- ☐ **Alter**
Update or delete LF-Tag expressions.
- ☐ **Drop**
Delete LF-Tag expressions.
- ☐ **Grant with LF-Tag expression**
Allow principals to grant access permissions using LF-Tag expressions.
- ☐ **Super**
This permission supersedes the individual permissions set above.

Cancel

Add

6. Choisissez Ajouter.

AWS CLI

Pour créer une expression LF-Tag

- Entrez une `create-lf-tag-expression` commande.

L'exemple suivant crée une expression LF-Tag avec la balise contenant Department des valeurs Sales et Marketing la balise Location contenant la valeur. US

```
aws lakeformation create-lf-tag-expression \
```

```
-- name "my-tag-expression" \  
-- catalog-id "123456789012" \  
-- expression '{"Expression":[{"TagKey":"Department","TagValues":  
["Sales","Marketing"]},{"TagKey":"Location","TagValues":["US"]}]}'
```

Cette commande CLI crée une nouvelle expression LF-Tag dans le AWS Glue Data Catalog. L'expression peut être utilisée pour accorder des autorisations aux ressources du catalogue de données telles que les bases de données, les tables, les vues ou les colonnes en fonction de leurs balises associées. Dans cet exemple, l'expression fera correspondre les ressources qui ont la Department clé avec des valeurs Sales ou Marketing, et la Location clé avec la valeur US.

En tant que créateur d'expressions de balise, le principal obtient l'Alter autorisation d'utiliser cette expression de balise LF et peut mettre à jour ou supprimer l'expression. Le créateur principal de l'expression LF-Tag peut également Alter autoriser un autre principal à mettre à jour et à supprimer cette expression.

Mise à jour des expressions LF-Tag

Seuls les administrateurs du lac de données, le créateur de l'expression LF-Tag et les principaux qui ont Alter l'Superautorisation d'utiliser l'expression LF-Tag peuvent mettre à jour une expression LF-Tag. Outre Alter l'autorisation, vous avez également besoin de l'autorisation lakeformation:UpdateLFTagExpression IAM et de l'Grant with LF-Tag autorisation sur toutes les valeurs-clés sous-jacentes du nouveau corps d'expression pour mettre à jour les expressions.

Vous mettez à jour une expression LF-Tag en mettant à jour la description, le corps de l'expression et les autorisations accordées à l'expression. Vous ne pouvez pas modifier le nom de l'expression LF-Tag. Pour modifier le nom, supprimez l'expression LF-Tag et ajoutez-en une avec les paramètres requis.

Vous pouvez mettre à jour une expression de balise LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

Pour mettre à jour une expression de balise LF

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, créateur du tag LF ou principal Alter autorisé à utiliser le tag LF.

2. Dans le volet de navigation, sous Autorisations, choisissez LF-Tags and permissions.
3. Choisissez l'onglet Expressions LF-Tag.
4. Dans la section Expressions de balise LF, sélectionnez une expression de balise LF, puis choisissez Modifier.
5. Dans la boîte de dialogue Modifier l'expression LF-Tag, mettez à jour la description et mettez à jour le corps de l'expression en ajoutant ou en supprimant des clés et des valeurs.

Pour ajouter plusieurs valeurs, dans le champ Valeurs, sélectionnez les valeurs dans le menu déroulant.

6. Choisissez Enregistrer.

AWS CLI

La `update-lf-tag-expression` commande de Lake Formation vous permet de mettre à jour une expression LF-Tag existante.

```
aws lakeformation update-lf-tag-expression \
-- name expression_name \
-- description new_description \
-- catalog-id catalog_id \
-- expression '{"Expression": [{"TagKey": "tag_key", "TagValues": ["tag_value1", "tag_value2", ...]}]}'
```

Voici ce que signifient les paramètres de la commande fournie :

- `name` — Le nom de l'expression de balise nommée existante que vous souhaitez mettre à jour.
- `description` — Nouvelle description de l'expression.

`catalog-id` — ID du catalogue de données dans lequel réside l'expression de balise nommée.

- **expression** — La nouvelle chaîne d'expression de balise avec laquelle vous souhaitez mettre à jour l'expression.

Suppression d'expressions de balises LF

Vous pouvez supprimer les expressions LF-Tag qui ne sont plus utilisées. Si vous avez accordé des autorisations aux principaux sur les ressources du catalogue de données à l'aide de l'expression LF-Tag, ils n'en auront plus.

Seuls les administrateurs du lac de données, le créateur de l'expression LF-Tag ou un mandant Drop autorisé à utiliser l'expression LF-Tag peuvent supprimer une expression LF-Tag. Outre l'Drop autorisation, le principal a également besoin de l'autorisation `lakeformation:DeleteLFTagExpression` IAM pour supprimer une expression LF-Tag.

Vous pouvez supprimer une expression de balise LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

Pour supprimer une expression de balise LF (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, créateur de l'expression LF-Tag ou principal autorisé à supprimer l'expression.

2. Dans le volet de navigation, sous Permissions, sélectionnez LF-Tags and permissions.
3. Choisissez l'onglet expression LF-Tag.
4. Dans la section Expressions de balise LF, sélectionnez une expression de balise LF, puis choisissez Supprimer.
5. Dans l'expression Supprimer le tag LF ? boîte de dialogue, pour confirmer la suppression, entrez le nom de l'expression LF-Tag dans le champ prévu à cet effet, puis choisissez Supprimer.

AWS CLI

Pour supprimer un tag LF ()AWS CLI

- Entrez une `delete-lf-tag-expression` commande. Indiquez le nom de l'expression et l'ID du catalogue à supprimer.

Exemple

L'exemple suivant supprime l'expression LF-Tag portant le nom du catalogue `my-tag-expression` de données avec ID. `123456789012` Le `catalog-id` paramètre est facultatif si vous utilisez le même compte que celui utilisé pour votre AWS CLI configuration. Après avoir supprimé une expression LF-Tag, Lake Formation nettoie les enregistrements d'autorisation associés à cette expression. Cela inclut à la fois les enregistrements d'autorisation individuels et les enregistrements d'autorisation agrégés contenant l'expression supprimée.

```
aws lakeformation delete-lf-tag-expression \  
--name "my-tag-expression" \  
--catalog-id "123456789012"
```

Répertorier les expressions de balises LF

Vous pouvez répertorier les expressions LF-Tag pour lesquelles vous disposez des autorisations `Describe`. Les administrateurs de data lake, les créateurs d'expressions de balises LF et les administrateurs en lecture seule peuvent implicitement voir toutes les expressions de balises de leur compte.

Vous pouvez répertorier les expressions LF-Tag à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

Console

Pour répertorier les expressions de balises LF (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant que créateur d'expressions LF-Tag, en tant qu'administrateur de data lake ou en tant que principal ayant obtenu des autorisations sur les expressions LF-Tag et disposant de l'autorisation IAM. `lakeformation:ListLFTagExpressions`

2. Dans le volet de navigation, sous Autorisations, balises LF et autorisations.
3. Choisissez l'onglet LF-Tag Expressions pour voir les expressions. Cette section présente des informations sur les expressions LF-Tag existantes, notamment le nom de l'expression, l'expression elle-même avec des liens vers les balises incluses et les options permettant de créer, de modifier ou de supprimer des expressions.

AWS CLI

Pour répertorier les balises LF ()AWS CLI

- Pour répertorier les expressions LF-Tag à l'aide de AWS CLI, vous pouvez utiliser la `list-lf-tag-expressions` commande. La syntaxe de la demande est la suivante :

```
aws lakeformation list-lf-tag-expressions \  
-- catalog-id "123456789012" \  
-- max-items "100" \  
-- next-token "next-token"
```

Où :

- `catalog-id` est l'ID de AWS compte du catalogue de données pour lequel vous souhaitez répertorier les expressions de balises.
- `max-items` indique le nombre maximal d'expressions de balise à renvoyer. Si ce paramètre n'est pas utilisé, la valeur par défaut est 100.
- `next-token` est un jeton de continuation si les résultats ont été tronqués lors d'une demande précédente.

La réponse inclura une liste d'expressions LF-Tag et un jeton suivant, le cas échéant.

Gestion des autorisations relatives aux valeurs des balises LF

Vous pouvez accorder les `Alter` autorisations sur `Drop` les balises LF aux principaux afin de gérer les expressions de valeur des balises LF. Vous pouvez également accorder `DescribeAssociate`,

et des `Grant with LF-Tag` expressions autorisations sur les balises LF aux principaux pour qu'ils puissent visualiser les balises LF et les attribuer aux ressources du catalogue de données (bases de données, tables et colonnes). Lorsque des balises LF sont attribuées aux ressources du catalogue de données, vous pouvez utiliser la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC) pour sécuriser ces ressources. Pour de plus amples informations, veuillez consulter [Contrôle d'accès basé sur des balises Lake Formation](#).

Vous pouvez accorder ces autorisations avec l'option d'octroi afin que les autres principaux puissent les accorder. Les `Associate` autorisations `Grant with LF-Tag` expressions `Describe`, et sont expliquées dans [Ajouter des créateurs de LF-Tag](#).

Vous pouvez accorder les `Associate` autorisations `Describe` et sur un LF-Tag à un compte externe AWS. L'administrateur du lac de données de ce compte peut ensuite accorder ces autorisations aux autres principaux du compte. Les principaux auxquels l'administrateur du lac de données du compte externe accorde l'`Associate` autorisation peuvent ensuite attribuer des balises LF aux ressources du catalogue de données que vous avez partagées avec leur compte.

Lorsque vous accordez à un compte externe, vous devez inclure l'option de subvention.

Vous pouvez accorder des autorisations sur les balises LF à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

Rubriques

- [Répertorier les autorisations LF-Tag à l'aide de la console](#)
- [Octroi d'autorisations LF-Tag à l'aide de la console](#)
- [Gestion des autorisations LF-Tag à l'aide du AWS CLI](#)

Pour plus d'informations, consultez [Gestion des balises LF pour le contrôle d'accès aux métadonnées](#) et [Contrôle d'accès basé sur des balises Lake Formation](#).

Répertorier les autorisations LF-Tag à l'aide de la console

Vous pouvez utiliser la console Lake Formation pour consulter les autorisations accordées sur les LF-Tags. Vous devez être un créateur de balises LF, un administrateur de lac de données ou avoir l'`Associate` autorisation `Describe` ou l'autorisation sur une balise LF pour le voir.

Pour répertorier les autorisations LF-Tag (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant que créateur de balises LF, administrateur du lac de données ou en tant qu'utilisateur à qui les Drop Describe autorisations relatives aux balises LF ont été accordées.

Alter Associate

2. Dans le volet de navigation, sous Autorisations, choisissez LF-Tags et autorisations, puis choisissez la section Autorisations LF-Tag.

La section des autorisations LF-Tag présente un tableau contenant le principal, les clés de balise, les valeurs et les autorisations.

	Principal	Principal type	Keys	Values	LF-Tag permissions	LF-Tag value permissions	Grantable
☐	arn:aws:iam::[redacted]:role/Admin	IAM role	module	All values	Alter, Drop	-	Alter, Drop
☐	arn:aws:iam::[redacted]:role/Admin	IAM role	module	All values	-	Describe	Describe
☐	arn:aws:iam::[redacted]:role/Admin	IAM role	module	All values	-	Associate	Associate
☐	arn:aws:iam::[redacted]:role/Admin	IAM role	module	All values	-	Grant with LF-Tag expression	Grant with LF-Tag expression
☐	arn:aws:iam::[redacted]:role/Admin	IAM role	LF-Test	All values	-	Describe	Describe
☐	arn:aws:iam::[redacted]:role/Admin	IAM role	LF-Test	All values	-	Associate	Associate

Octroi d'autorisations LF-Tag à l'aide de la console

Les étapes suivantes expliquent comment accorder des autorisations sur les balises LF à l'aide de la page Accorder des autorisations sur les balises LF de la console Lake Formation. La page est divisée en sections suivantes :

- Types d'autorisation : type d'autorisation à accorder.
- Principaux : utilisateurs ou rôles IAM, ou utilisateurs ou rôles SAML auxquels accorder des autorisations.
- Autorisations d'autorisation des paires clé-valeur des balises LF : paires clé-valeur des balises LF auxquelles accorder des autorisations.
- Autorisations LF-Tag — Les LF-Tags auxquels accorder des autorisations.
- Autorisations d'autorisation des expressions LF-Tag : les balises LF auxquelles accorder des autorisations.
- Autorisations : autorisations à accorder.

Ouvrez la page Accorder les autorisations LF-Tag

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
Connectez-vous en tant que créateur du tag LF, administrateur du lac de données ou en tant qu'utilisateur. Les autorisations LF-Tag ou paire clé-valeur LF-Tag sur les balises LF ont été accordées avec cette option. Grant
2. Dans le volet de navigation, choisissez LF-Tags et permissions, choisissez la section LF-Tag permissions.
3. Choisissez Grant permissions (Accorder des autorisations).

Spécifiez le type d'autorisation

Dans la section Type d'autorisations, choisissez un type d'autorisation.

Autorisations LF-Tag

Choisissez les autorisations LF-Tag pour permettre aux principaux de mettre à jour les valeurs des LF-Tag ou de supprimer des LF-Tag.

Autorisations relatives à la paire clé-valeur LF-Tag

Choisissez les autorisations relatives à la paire clé-valeur LF-Tag pour permettre aux principaux d'attribuer des balises LF aux ressources du catalogue de données, d'afficher les balises LF et d'accorder aux principaux des autorisations basées sur les balises LF sur les ressources du catalogue de données.

Les options disponibles dans les sections suivantes dépendent du type d'autorisations.

Autorisations d'expression LF-Tag

Choisissez les autorisations d'expression LF-Tag pour permettre aux principaux de mettre à jour les expressions ou de supprimer des expressions.

Spécifiez les principes

Note

Vous ne pouvez pas accorder d'autorisations LF-Tag (AlteretDrop) à des comptes externes ou à des directeurs d'un autre compte.

Dans la section Principaux, choisissez un type de principal et spécifiez les principaux auxquels accorder des autorisations.

Principals

☒ **IAM users and roles**
Users or roles from this AWS account.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☐ **External accounts**
AWS account, AWS organization or IAM principal outside of this account

IAM users and roles
Add one or more IAM users or roles.

Choose IAM principals to add ▼

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

Utilisateurs et groupes SAML

Pour les utilisateurs et les groupes SAML et Quick, entrez un ou plusieurs Amazon Resource Names (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou ARNs pour les utilisateurs ou groupes Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

Note

L'intégration de Lake Formation à Quick n'est prise en charge que pour Quick Enterprise Edition.

Comptes externes

Dans le AWS champ Compte, saisissez un ou plusieurs AWS comptes valides IDs. Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

Pour le principal IAM, entrez l'ARN de l'utilisateur ou du rôle IAM.

Spécifiez les balises LF

Pour accorder des autorisations sur les balises LF, dans la section Autorisations des balises LF, spécifiez les balises LF auxquelles accorder des autorisations.

LF-Tag permissions

LF-Tags
Choose the LF-Tags you want to grant permissions to.

Choose one or more LF-Tags ▼

Department X

Permissions
Choose the specific LF-Tag permissions to grant.

☒ **Alter**
Update or delete key values.

☒ **Drop**
Delete tag(s).

Grantable permissions
Choose the permissions that the grant recipient(s) can grant to other principals.

☒ **Alter**
Update or delete key values.

☒ **Drop**
Delete tag(s).

Cancel

Grant

- Choisissez un ou plusieurs LF-Tag à l'aide de la liste déroulante.

Spécifiez les paires clé-valeur LF-Tag

1. Pour accorder des autorisations sur les paires clé-valeur LF-Tag, (vous devez d'abord choisir les autorisations de paire clé-valeur LF-Tag comme type d'autorisation), choisissez Ajouter une paire clé-valeur LF-Tag pour afficher la première ligne de champs permettant de spécifier la clé et les valeurs LF-Tag.

LF-Tag key-value pair permissions

Key

Values

Remove

Add LF-Tag key-value pair

You can add 50 more LF-Tags.

Permissions

Choose the specific key-value pair permissions to grant.

☐ Describe
See keys and values.

☐ Associate
Assign LF-Tags to databases, tables, and columns.

☐ Grant with LF-Tag expression
Allow the principal(s) to grant access permissions using the LF-Tag(s).

Grantable permissions

Choose the permissions that the grant recipient(s) can grant to other principals.

☐ Describe
See keys and values.

☐ Associate
Assign LF-Tags to databases, tables, and columns.

☐ Grant with LF-Tag expression
Allow the principal(s) to grant access permissions using the LF-Tag(s).

Cancel

Grant

2. Positionnez le curseur dans le champ Clé, commencez éventuellement à taper pour affiner la liste de sélection, puis sélectionnez une touche LF-Tag.
3. Dans la liste des valeurs, sélectionnez une ou plusieurs valeurs, puis appuyez sur Tab ou cliquez ou appuyez en dehors du champ pour enregistrer les valeurs sélectionnées.

Note

Si l'une des lignes de la liste des valeurs est sélectionnée, appuyez sur Entrée pour sélectionner ou désactiver la case à cocher.

Les valeurs sélectionnées apparaissent sous forme de vignettes sous la liste des valeurs.

Cliquez sur le ✕ pour supprimer une valeur. Choisissez Supprimer pour supprimer le LF-tag dans son intégralité.

4. Pour ajouter un autre tag LF, choisissez à nouveau Ajouter un tag LF et répétez les deux étapes précédentes.

Spécifiez les expressions LF-Tag

1. Pour accorder des autorisations sur les expressions LF-Tag, (vous devez d'abord choisir les autorisations des expressions LF-Tag comme type d'autorisation).

Permission type

Choose the type of permission to grant. [Learn more](#)

☐ **LF-Tag permissions**
Grant permissions on LF-Tags to create, update, and delete LF-Tags.

☐ **LF-Tag key-value pair permissions**
Grant permissions on LF-Tag key-value pairs to assign LF-Tags to Data Catalog resources and grant permissions on the resources to principals.

☒ **LF-Tag expression permissions - new**
Grant permissions on LF-Tag expressions.

Principals

Choose the principals to grant permissions.

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

datalake_user ✕
User

2. Choisissez une expression LF-Tag.

3. Les expressions sélectionnées apparaissent sous forme de vignettes sous la liste des expressions LF-Tag. Cliquez sur le ✕ pour supprimer une expression.
4. Pour ajouter une autre expression LF-Tag, choisissez-en une autre.

Spécifiez les autorisations

Cette section affiche les autorisations LF-Tag ou les autorisations à valeur LF-Tag en fonction du type d'autorisation que vous avez choisi à l'étape précédente.

Selon le type d'autorisation que vous avez choisi d'accorder, sélectionnez les autorisations LF-Tag ou les autorisations par paire clé-valeur LF-Tag, et les autorisations pouvant être accordées.

1. Sous Autorisations LF-Tag, sélectionnez les autorisations à accorder.

Si vous accordez Drop and Alter, vous accordez implicitement Describe.

Vous devez accorder les autorisations Alter and Drop sur toutes les valeurs des balises.

2. Sous Autorisations à valeur clé-valeur LT-Tag, sélectionnez les autorisations à accorder.

L'associé octroie implicitement l'autorisation à Describe. Choisissez Grant avec expression LF-Tag pour permettre au bénéficiaire de la subvention d'accorder ou de révoquer les autorisations d'accès aux ressources du catalogue de données à l'aide de la méthode LF-TBAC.

3. Sous Autorisations d'expression LF-Tag, sélectionnez les autorisations à accorder.

Si vous accordez Drop and Alter, vous accordez implicitement Describe.

Si vous accordez une super autorisation, toutes les autorisations disponibles sont accordées.

4. (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS
5. Choisissez Accorder.

Gestion des autorisations LF-Tag à l'aide du AWS CLI

Vous pouvez accorder, révoquer et répertorier les autorisations sur les balises LF en utilisant le AWS Command Line Interface ().AWS CLI

Pour répertorier les autorisations LF-Tag (AWS CLI)

- Entrez une `list-permissions` commande. Vous devez être le créateur du tag LF, un administrateur du lac de données ou avoir l'`Grant with LF-Tag permissions` autorisation `Drop, Alter, Describe Associate`, sur un tag LF pour le voir.

La commande suivante demande toutes les balises LF pour lesquelles vous avez des autorisations.

```
aws lakeformation list-permissions --resource-type LF_TAG
```

Voici un exemple de sortie destiné à un administrateur de lac de données, qui voit toutes les balises LF accordées à tous les principaux. Les utilisateurs non administrateurs ne voient que les balises LF qui leur sont accordées. Les autorisations LF-Tag accordées à partir d'un compte externe apparaissent sur une page de résultats distincte. Pour les voir, répétez la commande et fournissez à l'`--next-token` argument le jeton renvoyé lors de l'exécution de la commande précédente.

```
{
  "PrincipalResourcePermissions": [
    {
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/datalake_admin"
      },
      "Resource": {
        "LFTag": {
          "CatalogId": "111122223333",
          "TagKey": "environment",
          "TagValues": [
            "*"
          ]
        }
      },
      "Permissions": [
        "ASSOCIATE"
      ],
      "PermissionsWithGrantOption": [
        "ASSOCIATE"
      ]
    },
  ],
}
```

```

    {
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
      },
      "Resource": {
        "LFTag": {
          "CatalogId": "111122223333",
          "TagKey": "module",
          "TagValues": [
            "Orders",
            "Sales"
          ]
        }
      },
      "Permissions": [
        "DESCRIBE"
      ],
      "PermissionsWithGrantOption": []
    },
    ...
  ],
  "NextToken": "eyJzaG91bGRRdWVy...Wlzc2lvdnMiOnRydWV9"
}

```

Vous pouvez répertorier toutes les subventions pour une clé LF-Tag spécifique. La commande suivante renvoie toutes les autorisations accordées sur le LF-Tagmodule.

```

aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}'

```

Vous pouvez également répertorier les valeurs de balise LF accordées à un principal spécifique pour une balise LF spécifique. Lorsque vous fournissez l'--principalargument, vous devez le --resource fournir. Par conséquent, la commande ne peut effectivement demander que les valeurs accordées à un principal spécifique pour une clé LF-Tag spécifique. La commande suivante montre comment procéder pour la touche principale datalake_user1 et la touche LF-Tag. module

```

aws lakeformation list-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/

```

```
datalake_user1 --resource-type LF_TAG --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}]'
```

Voici un exemple de sortie.

```
{
  "PrincipalResourcePermissions": [
    {
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
      },
      "Resource": {
        "LFTag": {
          "CatalogId": "111122223333",
          "TagKey": "module",
          "TagValues": [
            "Orders",
            "Sales"
          ]
        }
      },
      "Permissions": [
        "ASSOCIATE"
      ],
      "PermissionsWithGrantOption": []
    }
  ]
}
```

Pour accorder des autorisations sur les balises LF (AWS CLI)

1. Utilisez une commande similaire à la suivante. Cet exemple accorde à l'utilisateur `datalake_user1` l'Associate autorisation d'utiliser le tag LF avec la clé. `module` Il autorise l'affichage et l'attribution de toutes les valeurs pour cette clé, comme indiqué par l'astérisque (*).

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}]'
```

L'octroi de l'Associate autorisation octroie implicitement l'Describe autorisation.

L'exemple suivant accorde Associate au AWS compte externe 1234-5678-9012 sur le LF-Tag avec la clé, avec l'option d'octroi. module Il accorde des autorisations pour afficher et attribuer uniquement les valeurs sales et orders.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=123456789012 --permissions "ASSOCIATE"
--permissions-with-grant-option "ASSOCIATE" --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}]'
```

2. L'octroi de l'GrantWithLFTagExpression autorisation octroie implicitement l'Describe autorisation.

L'exemple suivant accorde GrantWithLFTagExpression à un utilisateur sur le tag LF avec la clé module, avec l'option d'octroi. Il accorde des autorisations pour consulter et octroyer des autorisations sur les ressources du catalogue de données en utilisant uniquement les valeurs sales et orders.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333 --permissions "GrantWithLFTagExpression"
--permissions-with-grant-option "GrantWithLFTagExpression" --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}]'
```

3. L'exemple suivant accorde des Drop autorisations à un utilisateur sur le tag LF avec la clé module, avec l'option d'octroi. Il accorde l'autorisation de supprimer le tag LF. Pour supprimer une balise LF, vous devez disposer d'autorisations sur toutes les valeurs de cette clé.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
--permissions-with-grant-option "DROP" --resource '{ "LFTag":
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}]'
```

4. L'exemple suivant accorde des Alter autorisations à l'utilisateur sur le tag LF avec la clé module, avec l'option d'octroi. Il accorde l'autorisation de supprimer le tag LF. Pour mettre à jour une balise LF, vous devez disposer d'autorisations sur toutes les valeurs de cette clé.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
```

```
--permissions-with-grant-option "ALTER" --resource '{ "LFTag":  
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Pour révoquer les autorisations sur les balises LF ()AWS CLI

- Utilisez une commande similaire à la suivante. Cet exemple révoque l'Associate autorisation sur le tag LF avec la clé module de l'utilisateur. `datalake_user1`

```
aws lakeformation revoke-permissions --principal  
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/  
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":  
{"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC

Vous pouvez accorder les autorisations DESCRIBE et ASSOCIATE Lake Formation sur les balises LF aux principaux afin qu'ils puissent visualiser les balises LF et les attribuer aux ressources du catalogue de données (bases de données, tables, vues et colonnes). Lorsque des balises LF sont attribuées aux ressources du catalogue de données, vous pouvez utiliser la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC) pour sécuriser ces ressources. Pour de plus amples informations, veuillez consulter [Contrôle d'accès basé sur des balises Lake Formation](#).

Dans un premier temps, seul l'administrateur du lac de données peut accorder ces autorisations. Si l'administrateur du lac de données accorde ces autorisations avec l'option grant, d'autres principaux peuvent les accorder. Les ASSOCIATE autorisations DESCRIBE et sont expliquées dans [Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation](#).

Vous pouvez accorder les ASSOCIATE autorisations DESCRIBE et sur un LF-Tag à un compte externe AWS . L'administrateur du lac de données de ce compte peut ensuite accorder ces autorisations aux autres principaux du compte. Les principaux auxquels l'administrateur du lac de données du compte externe accorde l'ASSOCIATE autorisation peuvent ensuite attribuer des balises LF aux ressources du catalogue de données que vous avez partagées avec leur compte.

Lorsque vous accordez à un compte externe, vous devez inclure l'option de subvention.

Vous pouvez accorder des autorisations sur les balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

 Note

Les étapes suivantes ne sont pas nécessaires pour les catalogues S3 Tables. Vous pouvez utiliser les balises LF pour accorder des autorisations sur les catalogues de tables S3 existants sans les supprimer ni les recréer.

Activation de la prise en charge des balises LF pour les catalogues fédérés existants qui utilisent les autorisations de Lake Formation

Procédez comme suit si vous possédez des catalogues fédérés existants qui utilisent les autorisations de Lake Formation, tels qu'Amazon Redshift Amazon DynamoDB ou des catalogues créés avant que le support des balises LF ne soit disponible pour les catalogues fédérés.

1. Supprimer le catalogue existant : appelez l'opération `deleteCatalog` API pour supprimer le catalogue fédéré existant qui utilise les autorisations de Lake Formation.
2. Créer un nouveau catalogue fédéré : créez un nouveau catalogue et orientez le nouveau catalogue vers votre espace de noms/partage de données existant.

Utiliser un nouveau nom pour le catalogue : ce processus met à jour vos catalogues fédérés préexistants afin de prendre en charge la fonctionnalité LF-Tag. Si vous souhaitez utiliser le même nom de catalogue, contactez AWS l'équipe d'assistance pour obtenir de l'aide.

Rubriques

- [Octroi d'autorisations au catalogue de données](#)

 Consultez aussi

- [Gestion des autorisations relatives aux valeurs des balises LF](#)
- [Gestion des balises LF pour le contrôle d'accès aux métadonnées](#)
- [Contrôle d'accès basé sur des balises Lake Formation](#)

Octroi d'autorisations au catalogue de données

Utilisez la console Lake Formation ou accordez AWS CLI à Lake Formation des autorisations sur les bases de données, les tables, les vues et les colonnes du catalogue de données à l'aide de la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC).

Console

Les étapes suivantes expliquent comment accorder des autorisations à l'aide de la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) et de la page Accorder des autorisations aux lacs de données sur la console Lake Formation. La page est divisée selon les sections suivantes :

- Principaux — Les utilisateurs, les rôles et les autorisations Comptes AWS auxquelles accorder des autorisations.
- Balises LF ou ressources du catalogue : bases de données, tables ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

1. Ouvrez la page des autorisations du lac de données Grant.

Ouvrez la AWS Lake Formation console sur <https://console.aws.amazon.com/lakeformation/> et connectez-vous en tant qu'administrateur de lac de données ou en tant qu'utilisateur ayant obtenu les autorisations de Lake Formation sur les ressources du catalogue de données via LF-TBAC avec l'option d'autorisation.

Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données. Ensuite, choisissez Accorder.

2. Spécifiez les principes.

Dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels accorder des autorisations.

Grant permissions

Principals

Choose the principals to grant permissions.

☒ **IAM users and roles**
Users or roles from this AWS account.

☐ **IAM Identity Center**
- *new*
Users and groups configured in IAM Identity Center.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☐ **External accounts**
AWS account, AWS organization or IAM principal outside of this account

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add ▼

AWSServiceRoleForRedshift X

Role

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou dans la liste Utilisateurs et groupes.

Utilisateurs et groupes SAML

Pour les utilisateurs et les groupes SAML et Quick, entrez un ou plusieurs Amazon Resource Names (ARNs) pour les utilisateurs ou les groupes fédérés via SAML, ou ARNs pour les utilisateurs ou groupes Quick. Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire le ARNs, voir [Lake Formation accorde et AWS CLI révoque des commandes](#).

Note

L'intégration de Lake Formation à Quick n'est prise en charge que pour Quick Enterprise Edition.

Comptes externes

Pour Comptes AWS AWS l'organisation ou le principal IAM, entrez une ou plusieurs organisations Compte AWS IDs IDs IDs, unités organisationnelles ou ARN valides pour l'utilisateur ou le rôle IAM. Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

3. Spécifiez les balises LF.

Assurez-vous que l'option Resources matching by LF-Tags est sélectionnée. Choisissez des paires clé-valeur LF-Tag ou des expressions LF-Tag enregistrées.

1. Si vous choisissez l'option paires clé-valeur LF-Tag, choisissez les clés et les valeurs.

Si vous choisissez plusieurs valeurs, vous créez une expression LF-Tag à l'aide d'un OR opérateur. Cela signifie que si l'une des valeurs de balise LF correspond à une balise LF attribuée à une ressource de catalogue de données, des autorisations sur cette ressource vous sont accordées.

LF-Tags or catalog resources

Choose a method to grant permissions.

☒ **Resources matched by LF-Tags (recommended)**

Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☐ **Named Data Catalog resources**

Manage permissions for specific databases or tables, in addition to fine-grained data access.

☒ **LF-Tag key-value pairs**

☐ **Saved LF-Tag expressions - new**

Key

Location

Values

Choose LF-Tag values

Remove

US

Department

Choose LF-Tag values

Remove

marketing

sales

Add LF-Tag key-value pair

You can add 48 more LF-Tags.

Expression review

The LF-Tag expression above will be interpreted in the following way.

Location = US
AND Department = (marketing OR sales)

☒ **Save as new expression**

Use saved expressions to grant permissions. Create LF-Tag expression permissions are needed.

New LF-Tag expression name

Enter a name that describes the expression. Expression name cannot be edited after creation.

new-expression

Name must be less than 1000 characters.

2. (Facultatif) Choisissez à nouveau Ajouter une paire clé-valeur de balise LF pour spécifier une autre balise LF.

Si vous spécifiez plusieurs balises LF, vous créez une expression de balise LF à l'aide d'un opérateur. AND Le principal n'obtient des autorisations sur une ressource de catalogue de données que si une balise LF correspondante a été attribuée à la ressource pour chaque balise LF de l'expression LF-Tag.

3. Choisissez l'option Enregistrer en tant que nouvelle expression pour réutiliser l'expression.

Vous devez Create LF-Tag expression enregistrer les expressions.

Pour plus d'informations sur les expressions LF-Tag, consultez. [Gestion des expressions LF-Tag pour le contrôle d'accès aux métadonnées](#)

4. Spécifiez les autorisations.

Spécifiez les autorisations que vous souhaitez accorder au principal pour faire correspondre les ressources du catalogue de données. Les ressources correspondantes sont les ressources auxquelles des balises LF ont été attribuées qui correspondent à l'une des expressions de balise LF accordées au principal.

Vous pouvez spécifier les autorisations à accorder sur les bases de données correspondantes, les tables correspondantes et les vues correspondantes.

▼ Database permissions

Database permissions

Choose specific access permissions to grant.

☐ Create table
 ☐ Alter
 ☐ Drop
 ☐ Super

☐ Describe

Grantable permissions

Choose the permission that may be granted to others.

☐ Create table
 ☐ Alter
 ☐ Drop
 ☐ Super

☐ Describe

This permission is the union of all the individual permissions to the left, and supersedes them.

▼ Table permissions

Table permissions

Choose specific access permissions to grant.

☐ Alter
 ☐ Insert
 ☐ Drop
 ☐ Super

☐ Delete
 ☐ Select
 ☐ Describe

Grantable permissions

Choose the permission that may be granted to others.

☐ Alter
 ☐ Insert
 ☐ Drop
 ☐ Super

☐ Delete
 ☐ Select
 ☐ Describe

This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Sous Autorisations de base de données, sélectionnez les autorisations de base de données à accorder au principal sur les bases de données correspondantes.

Sous Autorisations relatives aux tables, sélectionnez les autorisations de table ou de vue à accorder au principal sur les tables et les vues correspondantes.

Vous pouvez également choisir `SelectDescribe`, et `Drop` les autorisations dans le tableau, les autorisations à appliquer aux vues.

5. Choisissez `Accorder`.

AWS CLI

Vous pouvez utiliser la méthode AWS Command Line Interface (AWS CLI) et la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) pour accorder à Lake Formation des autorisations sur les bases de données, les tables et les colonnes du catalogue de données.

Octroi d'autorisations de data lake à l'aide de la méthode AWS CLI et de la méthode LF-TBAC

- Utilisez la commande `grant-permissions`.

Exemple

L'exemple suivant accorde l'expression LF-tag « `module=*` » (toutes les valeurs de la clé `module` LF-Tag) à l'utilisateur. `datalake_user1` Cet utilisateur aura l'`CREATE_TABLE` autorisation d'accéder à toutes les bases de données correspondantes, c'est-à-dire aux bases de données auxquelles a été attribuée la balise LF avec la clé `module`, quelle que soit la valeur.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "CREATE_TABLE" --resource '{ "LFTagPolicy":
{"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["*"]}]]}'
```

Exemple

L'exemple suivant accorde l'expression LF-tag « `(level=director) AND (region=west OR region=south)` » à l'utilisateur. `datalake_user1` Cet utilisateur aura les `DROP` autorisations `SELECTALTER`, et avec l'option d'autorisation sur les tables correspondantes, c'est-à-dire les tables auxquelles les deux `level=director` et `(region=westouregion=south)` ont été assignés.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
```

```
datalake_user1 --permissions "SELECT" "ALTER" "DROP" --permissions-  
with-grant-option "SELECT" "ALTER" "DROP" --resource '{ "LFTagPolicy":  
  {"CatalogId":"111122223333","ResourceType":"TABLE","Expression": [{"TagKey":  
    "level","TagValues": ["director"]},{ "TagKey": "region","TagValues": ["west",  
    "south"]}]]}'
```

Exemple

L'exemple suivant accorde l'expression LF-tag « module=orders » au AWS compte 1234-5678-9012. L'administrateur du lac de données de ce compte peut ensuite accorder l'expression module=orders « » aux principaux de son compte. Ces principaux auront alors l'`CREATE_TABLE` autorisation de faire correspondre les bases de données détenues par le compte 1111-2222-3333 et partagées avec le compte 1234-5678-9012 en utilisant soit la méthode des ressources nommées, soit la méthode LF-TBAC.

```
aws lakeformation grant-permissions --principal  
DataLakePrincipalIdentifier=123456789012 --permissions "CREATE_TABLE" --  
permissions-with-grant-option "CREATE_TABLE" --resource '{ "LFTagPolicy":  
  {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":  
    [{"TagKey":"module","TagValues":["orders"]}]]}'
```

Contrôle d'accès basé sur les attributs

Dans AWS Lake Formation, vous pouvez accorder l'accès à AWS Glue Data Catalog des objets tels que des catalogues, des bases de données, des tables et des filtres de données à l'aide d'attributs qui sont des balises IAM et des balises de session associées à des entités IAM telles que les rôles et les utilisateurs.

Pour plus d'informations sur l'utilisation des balises de session, voir [assume-role](#) dans le guide de l'AWS CLI utilisateur.

Le contrôle d'accès basé sur les attributs (ABAC) est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. AWS appelle ces balises d'attributs. Vous pouvez utiliser ABAC pour accorder l'accès aux principaux au sein du même compte ou d'un autre compte sur les ressources du catalogue de données. Tout principal IAM dont les clés et les valeurs de balise IAM ou de balise de session correspondent accède à la ressource. Vous devez disposer d'autorisations pouvant être accordées sur les ressources pour octroyer ces subventions.

ABAC vous permet d'accorder l'accès à plusieurs utilisateurs en même temps. Lorsque de nouveaux utilisateurs rejoignent l'organisation, leur accès aux données peut être déterminé automatiquement en fonction de leurs attributs, tels que leur fonction ou leur service, sans que les administrateurs n'aient à attribuer manuellement des rôles ou des autorisations spécifiques. En utilisant des attributs plutôt que des rôles, ABAC fournit une méthode plus rationalisée et plus facile à gérer pour gérer l'accès aux données dans divers systèmes et environnements, améliorant ainsi la gouvernance et la conformité des données.

Pour plus d'informations sur la définition des attributs, voir [Définir des autorisations en fonction des attributs dotés d'une autorisation ABAC](#).

Pour plus d'informations sur les limites, les considérations et les AWS régions prises en charge, consultez [Considérations relatives au contrôle d'accès basé sur les attributs, limites et régions prises en charge](#).

Rubriques

- [Conditions préalables à l'octroi d'autorisations à l'aide d'attributs](#)
- [Octroi d'autorisations à l'aide du contrôle d'accès basé sur les attributs](#)

Conditions préalables à l'octroi d'autorisations à l'aide d'attributs

Pour accorder des autorisations à l'aide du contrôle d'accès basé sur les attributs (ABAC), vous devez remplir les conditions préalables suivantes :

- Mettez à jour les paramètres du catalogue de données pour activer les autorisations de Lake Formation pour les objets du catalogue de données. Pour plus d'informations, consultez la section [Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride](#).
- Définissez les paramètres de version multi-comptes sur deux ou plus.
- [Attachez des attributs](#) aux entités IAM qui nécessitent un accès.
- Seul un administrateur de lac de données ou un utilisateur IAM disposant des autorisations requises peut accorder l'accès aux objets du catalogue de données. Pour plus d'informations sur les autorisations requises, consultez la section [Autorisations IAM](#).

Octroi d'autorisations à l'aide du contrôle d'accès basé sur les attributs

Cette rubrique décrit les étapes à suivre pour accorder des autorisations d'accès basées sur les attributs aux ressources du catalogue de données. Vous pouvez utiliser la console Lake Formation ou l'interface de ligne de AWS commande (AWS CLI).

Octroi d'autorisations à l'aide d'ABAC ()AWS Management Console

1. Ouvrez la console Lake Formation à l'<https://console.aws.amazon.com/lakeformation/>adresse et connectez-vous en tant qu'administrateur du lac de données, créateur de ressources ou utilisateur IAM disposant d'autorisations accordables sur la ressource.
2. Effectuez l'une des actions suivantes :
 - Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données. Ensuite, choisissez Accorder.
 - Dans le volet de navigation, sélectionnez Catalogues sous Catalogue de données. Choisissez ensuite un objet de catalogue (catalogues, bases de données, tables et filtres de données), puis dans le menu Actions, sous Autorisations, sélectionnez Accorder.
3. Sur la page Accorder des autorisations, sélectionnez Principaux par attribut.
4. Spécifiez la clé et les valeurs de l'attribut. Si vous choisissez plusieurs valeurs, vous créez une expression attributaire à l'aide d'un OR opérateur. Cela signifie que si l'une des valeurs de balise d'attribut attribuées à un rôle ou à un utilisateur IAM correspond, il role/user obtient des autorisations d'accès à la ressource.

Si vous spécifiez plusieurs balises d'attribut, vous créez une expression d'attribut à l'aide d'un AND opérateur. Le principal n'obtient des autorisations sur une ressource de catalogue de données que si une balise correspondante role/user a été attribuée à l'IAM pour chaque balise d'attribut de l'expression d'attribut.

Passez en revue l'expression de politique Cedar résultante affichée dans la console.

Grant permissions

Principal type

Grant access by IAM principal or attribute tags.

☐

Principals

Grant permissions directly to specific IAM users, roles, or Identity Center principals.

☒

Principals by attributes

Grant permissions using attribute tags with Cedar policy expressions.

Attributes

Grant permissions using an attribute expression based on user, resource, and contextual attributes.

Attribute expression

Add tag key-value pairs from IAM roles or [IAM session tags](#). Separate values with commas. **AND** logic applies between pairs, **OR** logic within keys. Verify tag entries for accuracy.

Key

Department

Value

sales

Add key-value pair

Expression review

The attribute expression above will be written in the following Cedar policy. Use CLI/SDK to edit or write custom Cedar policy.

```
permit (principal, action, resource)
when {
  context.iam.principalTags.hasTag("Department")
  && context.iam.principalTags.getTag("Department") == "sales"
};
```

Permission scope

☒

This account

☐

External account

5. Choisissez l'étendue des autorisations. Si les bénéficiaires appartiennent à un compte externe, choisissez Compte externe et entrez l'identifiant du AWS compte.
6. Choisissez ensuite le compte Data Catalog ou des comptes externes. Vous devez disposer des autorisations pouvant être accordées correspondantes sur les ressources pour terminer avec succès les autorisations accordées.

7. Spécifiez les actions que vous souhaitez autoriser les principaux (utilisateurs ou rôles) dotés d'attributs correspondants à effectuer. L'accès est accordé aux entités IAM auxquelles ont été attribuées des balises et des valeurs correspondant à au moins une des expressions d'attributs que vous avez spécifiées. Vérifiez l'expression de politique Cedar dans la console. Pour plus d'informations sur le cèdre, voir [Qu'est-ce que le cèdre ? | Référence linguistique des politiques de Cedar GuideLink](#).
8. Choisissez ensuite les ressources du catalogue de données auxquelles vous souhaitez accorder l'accès. Vous pouvez définir ces autorisations pour différentes ressources du catalogue de données, notamment les catalogues, les bases de données, les tables et les filtres de données.
9. Choisissez Accorder.

Cette approche vous permet de contrôler l'accès en fonction des attributs, en veillant à ce que seuls les utilisateurs ou les rôles dotés des balises appropriées puissent effectuer des actions spécifiques sur les ressources désignées.

Octroi d'autorisations à l'aide d'ABAC ()AWS CLI

L'exemple suivant montre une expression d'attribut qui doit être respectée pour recevoir toutes les autorisations disponibles sur la ressource. Vous pouvez également spécifier des autorisations individuelles telles que `SelectDescribe`, ou `Drop`. L'expression utilise l'expression de politique Cedar. Pour plus d'informations sur le cèdre, voir [Qu'est-ce que le cèdre ? | Référence linguistique des politiques de Cedar GuideLink](#).

Cette condition vérifie si le principal IAM possède une department balise et si la valeur de la department balise est égale à sales.

```
aws lakeformation grant-permissions
--principal '{"DataLakePrincipalIdentifier": "111122223333:IAMPrincipals"}' \
--resource '{"Database": {"CatalogId": 111122223333, "Name": "abac-db"}}' \
--permissions ALL \
--condition '{"Expression": "context.iam.principalTags.hasTag(\"department\") \
&& context.iam.principalTags.getTag(\"department\") == \"sales\""}'
```

Exemple de scénario d'autorisations

Le scénario suivant montre comment configurer des autorisations pour sécuriser l'accès aux données dans AWS Lake Formation.

Shirley est administratrice de données. Elle souhaite créer un lac de données pour son entreprise AnyCompany. Actuellement, toutes les données sont stockées dans Amazon S3. John est directeur marketing et a besoin d'un accès écrit aux informations d'achat des clients (contenues dans `s3://customerPurchases`). Diego, analyste marketing, rejoint John cet été. John doit être en mesure d'autoriser Diego à effectuer des requêtes sur les données sans impliquer Shirley.

Mateo, du service financier, a besoin d'accéder aux données comptables (par exemple, `s3://transactions`). Il souhaite interroger les données des transactions dans les tables d'une base de données (`Finance_DB`) utilisée par l'équipe financière. Son manager, Arnav, peut lui donner accès au `Finance_DB`. Bien qu'il ne devrait pas être en mesure de modifier les données comptables, il doit être capable de convertir les données dans un format (schéma) adapté aux prévisions. Ces données seront stockées dans un compartiment séparé (`s3://financeForecasts`) qu'il pourra modifier.

Pour résumer :

- Shirley est l'administratrice du lac de données.
- John a besoin `CREATE_DATABASE` d'une `CREATE_TABLE` autorisation pour créer de nouvelles bases de données et de nouvelles tables dans le catalogue de données.
- John a également besoin `SELECT` d'`INSERTDELETE` autorisations et d'autorisations sur les tables qu'il crée.
- Diego a besoin `SELECT` d'une autorisation sur la table pour exécuter des requêtes.

Les employés de AnyCompany exécutent les actions suivantes pour configurer les autorisations. Les opérations d'API présentées dans ce scénario présentent une syntaxe simplifiée pour plus de clarté.

1. Shirley enregistre le chemin Amazon S3 contenant les informations d'achat des clients auprès de Lake Formation.

```
RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
```

2. Shirley accorde à John l'accès au chemin Amazon S3 contenant les informations d'achat des clients.

```
GrantPermissions(John, S3Location("s3://customerPurchases"),  
[DATA_LOCATION_ACCESS]) )
```

3. Shirley autorise John à créer des bases de données.

```
GrantPermissions(John, catalog, [CREATE_DATABASE])
```

- John crée la base de données `John_DB`. John a automatiquement `CREATE_TABLE` l'autorisation d'accéder à cette base de données parce qu'il l'a créée.

```
CreateDatabase(John_DB)
```

- John crée le tableau `John_Table` pointant vers `s3://customerPurchases`. Comme il a créé la table, il dispose de toutes les autorisations et peut accorder des autorisations sur celle-ci.

```
CreateTable(John_DB, John_Table)
```

- John autorise son analyste, Diego, à accéder à la table `John_Table`.

```
GrantPermissions(Diego, John_Table, [SELECT])
```

- John autorise son analyste, Diego, à accéder à `s3://customerPurchases/London/`. Shirley étant déjà inscrite à `s3://customerPurchases`, ses sous-dossiers sont enregistrés auprès de Lake Formation.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],  
S3Location("s3://customerPurchases/London/") )
```

- John permet à son analyste, Diego, de créer des tables dans la base de données `John_DB`.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],  
[] )
```

- Diego crée une table dans `John_DB` à `s3://customerPurchases/London/` et obtient automatiquement `ALTER`, `DROP`, `SELECT`, `INSERT`, et `DELETE` les autorisations.

```
CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Filtrage des données et sécurité au niveau des cellules dans Lake Formation

Lorsque vous accordez des autorisations à Lake Formation sur une table du catalogue de données, vous pouvez inclure des spécifications de filtrage des données afin de restreindre l'accès à certaines

données dans les résultats des requêtes et les moteurs intégrés à Lake Formation. Lake Formation utilise le filtrage des données pour garantir la sécurité au niveau des colonnes, au niveau des lignes et au niveau des cellules. Vous pouvez définir et appliquer des filtres de données sur des colonnes imbriquées si vos données sources contiennent des structures imbriquées.

Grâce aux fonctionnalités de filtrage des données de Lake Formation, vous pouvez mettre en œuvre les niveaux de sécurité des données suivants.

Sécurité au niveau des colonnes

L'octroi d'autorisations sur une table du catalogue de données avec sécurité au niveau des colonnes (filtrage des colonnes) permet aux utilisateurs de n'afficher que les colonnes spécifiques et les colonnes imbriquées auxquelles ils ont accès dans la table. Prenons l'exemple d'une table utilisée dans plusieurs applications pour une grande entreprise de communication multirégionale. L'octroi d'autorisations sur les tables du catalogue de données avec filtrage par colonne peut empêcher les utilisateurs qui ne travaillent pas dans le service des ressources humaines de voir des informations personnelles identifiables (PII) telles qu'un numéro de sécurité sociale ou une date de naissance. Vous pouvez également définir des politiques de sécurité et n'autoriser l'accès qu'à des sous-structures partielles de colonnes imbriquées.

Sécurité au niveau des lignes

L'octroi d'autorisations sur une table du catalogue de données avec sécurité au niveau des lignes (filtrage des lignes) permet aux utilisateurs de n'afficher que les lignes de données spécifiques auxquelles ils ont accès dans la table. Le filtrage est basé sur les valeurs d'une ou de plusieurs colonnes. Vous pouvez inclure des structures de colonnes imbriquées lorsque vous définissez des expressions de filtre de ligne. Par exemple, si les différents bureaux régionaux de l'entreprise de communication disposent de leur propre service des ressources humaines, vous pouvez limiter les dossiers personnels que les employés des ressources humaines peuvent consulter uniquement aux dossiers des employés de leur région.

Sécurité au niveau des cellules

La sécurité au niveau des cellules combine le filtrage des lignes et le filtrage des colonnes pour créer un modèle d'autorisations extrêmement flexible. Si vous visualisez les lignes et les colonnes d'un tableau sous forme de grille, en utilisant la sécurité au niveau des cellules, vous pouvez restreindre l'accès à des éléments individuels (cellules) de la grille n'importe où dans les deux dimensions. En d'autres termes, vous pouvez restreindre l'accès aux différentes colonnes en fonction de la ligne. Ceci est illustré par le schéma suivant, dans lequel les colonnes restreintes sont ombrées.

	Col1	Col2	Col3	Col4	Col5	Col6
Row1						
Row2						
Row3						
Row4						
Row5						

En reprenant l'exemple de la table des personnes, vous pouvez créer un filtre de données au niveau de la cellule qui restreint l'accès à la colonne des adresses postales si la colonne du pays est définie sur « Royaume-Uni », mais autorise l'accès à la colonne des adresses postales si la colonne du pays est définie sur « États-Unis ».

Les filtres s'appliquent uniquement aux opérations de lecture. Par conséquent, vous ne pouvez accorder que l'autorisation à la SELECT Lake Formation avec des filtres.

Sécurité au niveau des cellules sur les colonnes imbriquées

Lake Formation vous permet de définir et d'appliquer des filtres de données avec une sécurité au niveau des cellules sur des colonnes imbriquées. Cependant, les moteurs d'analyse intégrés tels qu'Amazon Athena, Amazon EMR et Amazon Redshift Spectrum permettent d'exécuter des requêtes sur des tables imbriquées gérées par Lake Formation avec une sécurité au niveau des lignes et des colonnes.

Pour connaître les limitations, veuillez consulter [Limites du filtrage des données](#).

Rubriques

- [Filtres de données dans Lake Formation](#)
- [Support partiQL dans les expressions de filtre de ligne](#)
- [Autorisations requises pour interroger des tables avec filtrage au niveau des cellules](#)
- [Gestion des filtres de données](#)

Filtres de données dans Lake Formation

Vous pouvez implémenter la sécurité au niveau des colonnes, des lignes et des cellules en créant des filtres de données. Vous sélectionnez un filtre de données lorsque vous accordez l'autorisation SELECT Lake Formation sur les tables. Si votre table contient des structures de colonnes imbriquées, vous pouvez définir un filtre de données en incluant ou en excluant les colonnes enfants et définir des expressions de filtre au niveau des lignes sur les attributs imbriqués.

Chaque filtre de données appartient à une table spécifique de votre catalogue de données. Un filtre de données inclut les informations suivantes :

- Nom du filtre
- Le catalogue IDs de la table associée au filtre
- Nom de la table
- Nom de la base de données contenant la table
- Spécification des colonnes : liste de colonnes et de colonnes imbriquées (avec `struct` types de données) à inclure ou à exclure dans les résultats de la requête.
- Expression de filtre de ligne : expression qui spécifie les lignes à inclure dans les résultats de la requête. Sous réserve de certaines restrictions, l'expression possède la syntaxe d'une `WHERE` clause du langage partiQL. Pour spécifier toutes les lignes, choisissez Accès à toutes les lignes sous Accès au niveau des lignes dans la console ou Utiliser `AllRowsWildcard` dans les appels d'API.

Pour plus d'informations sur les éléments pris en charge dans les expressions de filtre de ligne, consultez [Support partiQL dans les expressions de filtre de ligne](#).

Le niveau de filtrage obtenu dépend de la façon dont vous renseignez le filtre de données.

- Lorsque vous spécifiez le caractère générique « `all columns` » (« toutes les colonnes ») et que vous fournissez une expression de filtre de ligne, vous établissez uniquement une sécurité au niveau de la ligne (filtrage des lignes).
- Lorsque vous incluez ou excluez des colonnes spécifiques et des colonnes imbriquées, et que vous spécifiez « toutes les lignes » en utilisant le caractère générique pour toutes les lignes, vous établissez une sécurité au niveau des colonnes (filtrage des colonnes) uniquement.
- Lorsque vous incluez ou excluez des colonnes spécifiques et que vous fournissez également une expression de filtre de ligne, vous établissez une sécurité au niveau de la cellule (filtrage des cellules).

La capture d'écran suivante de la console Lake Formation montre un filtre de données qui effectue un filtrage au niveau des cellules. Pour les requêtes portant sur la `orders` table, cela restreint l'accès à la `customer_name` colonne et les résultats de la requête ne renvoient que les lignes où la `product_type` colonne contient « `pharma` ».

Create data filter



Data filter name

Enter a name that describes this data access filter.

restrict-pharma

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.

Target database

Select the database that contains the target table.

Choose databases ▼

Load more

sales



054881201579

Target table

Select the table for which the data filter will be created.

Choose tables ▼

Load more

orders



054881201579

Column-level access

Choose whether this filter should have column-level restrictions.

- ☐ Access to all columns
Filter won't have any column restrictions.
- ☐ Include columns
Filter will only allow access to specific columns.
- ☒ Exclude columns
Filter will allow access to all but specific columns.

Select columns

Choose one or more columns ▼

customer_name



string

Row filter expression

Enter the rest of the following query statement "SELECT * FROM orders WHERE..."
Please see the documentation for examples of filter expressions.

product_type='pharma'

Notez l'utilisation de guillemets simples pour entourer la chaîne littérale, 'pharma'.

Vous pouvez utiliser la console Lake Formation pour créer ce filtre de données, ou vous pouvez fournir l'objet de requête suivant à l'opération `CreateDataCellsFilter` d'API.

```
{
  "Name": "restrict-pharma",
  "DatabaseName": "sales",
  "TableName": "orders",
  "TableCatalogId": "111122223333",
  "RowFilter": {"FilterExpression": "product_type='pharma'"},
  "ColumnWildcard": {
    "ExcludedColumnNames": ["customer_name"]
  }
}
```

Vous pouvez créer autant de filtres de données que nécessaire pour une table. Pour ce faire, vous avez besoin d'une `SELECT` autorisation avec l'option d'octroi sur une table. Par défaut, les administrateurs de Data Lake sont autorisés à créer des filtres de données sur toutes les tables de ce compte. Vous n'utilisez généralement qu'un sous-ensemble des filtres de données possibles lorsque vous accordez des autorisations sur la table à un principal. Par exemple, vous pouvez créer un deuxième filtre de données pour la `orders` table qui est un filtre de `row-security-only` données. En vous référant à la capture d'écran précédente, vous pouvez choisir l'option `Accès à toutes les colonnes` et inclure une expression de filtre de ligne `product_type<>'pharma'`. Le nom de ce filtre de données peut être `no-pharma`. Il restreint l'accès à toutes les lignes dont la `product_type` colonne est définie sur « `pharma` ».

L'objet de demande pour l'opération `CreateDataCellsFilter` d'API pour ce filtre de données est le suivant.

```
{
  "Name": "no-pharma",
  "DatabaseName": "sales",
  "TableName": "orders",
  "TableCatalogId": "111122223333",
  "RowFilter": {"FilterExpression": "product_type<>'pharma'"},
  "ColumnNames": ["customer_id", "customer_name", "order_num",
    "product_id", "purchase_date", "product_type",
    "product_manufacturer", "quantity", "price"]
}
```

Vous pouvez ensuite accorder une autorisation SELECT sur le `orders` tableau contenant le filtre de `restrict-pharma` données à un utilisateur administratif, et SELECT sur le `orders` tableau contenant le filtre de `no-pharma` données à des utilisateurs non administratifs. Pour les utilisateurs du secteur de la santé, vous accorderiez SELECT sur le `orders` tableau un accès complet à toutes les lignes et colonnes (aucun filtre de données), ou peut-être un autre filtre de données restreignant l'accès aux informations tarifaires.

Vous pouvez inclure ou exclure des colonnes imbriquées lorsque vous spécifiez la sécurité au niveau des colonnes et au niveau des lignes dans un filtre de données. Dans l'exemple suivant, l'accès au `product.offer` champ est spécifié à l'aide de noms de colonnes qualifiés (entre guillemets). Ceci est important pour les champs imbriqués afin d'éviter que des erreurs ne se produisent lorsque les noms de colonnes contiennent des caractères spéciaux et pour maintenir la rétrocompatibilité avec les définitions de sécurité de niveau supérieur des colonnes.

```
{
  "Name": "example_dcf",
  "DatabaseName": "example_db",
  "TableName": "example_table",
  "TableCatalogId": "111122223333",
  "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
  "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

 Consultez aussi

- [Gestion des filtres de données](#)

Support partiQL dans les expressions de filtre de ligne

Vous pouvez créer des expressions de filtre de ligne à l'aide d'un sous-ensemble de types de données, d'opérateurs et d'agrégations PartiQL. Lake Formation n'autorise aucune fonction PartiQL standard ou définie par l'utilisateur dans l'expression du filtre. Vous pouvez utiliser des opérateurs de comparaison pour comparer des colonnes à des constantes (par exemple, `views >= 10000`), mais vous ne pouvez pas comparer des colonnes à d'autres colonnes.

Une expression de filtre Row peut être une expression simple ou une expression composite. La longueur totale de l'expression doit être inférieure à 2 048 caractères.

Expression simple

Une expression simple aura le format suivant : `<column name > <comparison operator ><value >`

- Nom de colonne

Il peut s'agir d'une colonne de données de niveau supérieur, d'une colonne de partition ou d'une colonne imbriquée présente dans le schéma de table et doit appartenir à la [Types de données pris en charge](#) liste ci-dessous.

- Opérateur de comparaison

Les opérateurs pris en charge sont les suivants : `=`, `>`, `<`, `>=`, `<=`, `<>`, `!=`, `BETWEEN`, `IN`, `LIKE`, `NOT`, `IS [NOT] NULL`

- Toutes les comparaisons de chaînes et les correspondances de LIKE modèles distinguent les majuscules et minuscules. Vous ne pouvez pas utiliser l'opérateur `IS [NOT] NULL` sur les colonnes de partition.
- Valeur de colonne

La valeur de colonne doit correspondre au type de données du nom de colonne.

Expression composite

Une expression composite sera au format : `( <simple expression > ) <AND/OR > ( <simple expression > )`. Les expressions composites peuvent être combinées ultérieurement à l'aide d'opérateurs logiques `AND/OR`.

Types de données pris en charge

Les filtres de ligne qui font référence à une AWS Glue Data Catalog table contenant un type de données non pris en charge provoqueront une erreur. Les types de données pris en charge pour les colonnes de table et les constantes sont les suivants, qui sont mappés aux types de Amazon Redshift données :

- `STRING`, `CHAR`, `VARCHAR`
- `INT`, `LONG`, `BIGINT`, `FLOAT`, `DECIMAL`, `DOUBLE`
- `BOOLEAN`
- `STRUCT`

Pour plus d'informations sur les types de données dans Amazon Redshift, consultez la section [Types de données](#) dans le guide du développeur de base de données Amazon Redshift.

Expressions de filtre de ligne

Exemple

Voici des exemples d'expressions de filtre de ligne valides pour une table comportant des colonnes : `country` (String), `id` (Long), `year` (partition column of type Integer), `month` (partition column of type Integer)

- `year > 2010 and country != 'US'`
- `(year > 2010 and country = 'US') or (month < 8 and id > 23)`
- `(country between 'Z' and 'U') and (year = 2018)`
- `(country like '%ited%') and (year > 2000)`

Exemple

Voici des exemples valides d'expressions de filtre de ligne pour une table comportant des colonnes imbriquées : `year > 2010 and customer.customerId <> 1`

Les champs imbriqués sous les colonnes de partition ne doivent pas être référencés lors de la définition d'expressions imbriquées au niveau des lignes.

Les constantes de chaîne doivent être placées entre guillemets simples.

Mots-clés réservés

Si votre expression de filtre de ligne contient des mots clés partiQL, vous recevrez une erreur d'analyse car les noms des colonnes peuvent entrer en conflit avec les mots clés. Dans ce cas, évitez les noms de colonnes en utilisant des guillemets doubles. Voici quelques exemples de mots clés réservés : « premier », « dernier », « asc », « manquant ». Consultez la spécification partiQL pour obtenir la liste des mots clés réservés.

Référence PartiQL

Pour plus d'informations sur PartiQL, consultez. <https://partiql.org/>

Autorisations requises pour interroger des tables avec filtrage au niveau des cellules

Les autorisations Gestion des identités et des accès AWS (IAM) suivantes sont requises pour exécuter des requêtes sur des tables avec un filtrage au niveau des cellules.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:StartQueryPlanning",
        "lakeformation:GetQueryState",
        "lakeformation:GetWorkUnits",
        "lakeformation:GetWorkUnitResults"
      ],
      "Resource": "*"
    }
  ]
}
```

Pour plus d'informations sur les autorisations de Lake Formation, consultez [Référence des personnalités de Lake Formation et des autorisations IAM](#).

Gestion des filtres de données

Pour implémenter la sécurité au niveau des colonnes, des lignes et des cellules, vous pouvez créer et gérer des filtres de données. Chaque filtre de données appartient à une table du catalogue de données. Vous pouvez créer plusieurs filtres de données pour une table, puis utiliser un ou plusieurs d'entre eux lorsque vous accordez des autorisations sur la table. Vous pouvez également définir et appliquer des filtres de données sur des colonnes imbriquées dont les struct types de données permettent aux utilisateurs d'accéder uniquement aux sous-structures des colonnes imbriquées.

Vous avez besoin SELECT d'une autorisation avec l'option d'octroi pour créer ou afficher un filtre de données. Pour permettre aux principaux utilisateurs de votre compte de consulter et d'utiliser un filtre de données, vous pouvez leur accorder l'DESCRIBE autorisation.

 Note

Lake Formation n'est pas favorable à l'octroi d'une Describe autorisation sur un filtre de données partagé depuis un autre compte.

Vous pouvez gérer les filtres de données à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Pour plus d'informations sur les filtres de données, voir [Filtres de données dans Lake Formation](#)

Création d'un filtre de données

Vous pouvez créer un ou plusieurs filtres de données pour chaque table du catalogue de données.

Pour créer un filtre de données pour une table du catalogue de données (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, propriétaire de la table cible ou principal disposant d'une autorisation Lake Formation sur la table cible.

2. Dans le volet de navigation, sous Catalogue de données, sélectionnez Filtres de données.
3. Sur la page Filtres de données, choisissez Créer un nouveau filtre.
4. Dans la boîte de dialogue Créer un filtre de données, entrez les informations suivantes :
 - Nom du filtre de données
 - Base de données cible : spécifiez la base de données qui contient la table.
 - Tableau cible
 - Accès au niveau des colonnes : laissez ce paramètre défini sur Accès à toutes les colonnes pour spécifier le filtrage des lignes uniquement. Choisissez Inclure les colonnes ou Exclure les colonnes pour définir le filtrage des colonnes ou des cellules, puis spécifiez les colonnes à inclure ou à exclure.

Colonnes imbriquées : si vous appliquez le filtre à une table contenant des colonnes imbriquées, vous pouvez spécifier explicitement les sous-structures des colonnes de structure imbriquées dans un filtre de données.

Lorsque vous accordez l'autorisation SELECT à un principal sur ce filer, le principal exécutant la requête suivante ne verra que les données pour `customer.customerName` et `noncustomer.customerId`.

```
SELECT "customer" FROM "example_db"."example_table";
```

Column-level access

Choose whether this filter should have column-level restrictions.

Column-level access

Choose whether this filter should have column-level restrictions.

- ☐ Access to all columns
- Filter won't have any column restrictions.
- ☒ Include columns
- Filter will only allow access to specific columns.
- ☐ Exclude columns
- Filter will allow access to all but specific columns.

Included columns (4/11)

Choose the columns for column-level access

Find column

< 1 >

	Name	Type
☐	customer	struct
☐	customerId	string
☒	customerName	string
☒	customerapplication	struct
☐	appId	string
☒	product	struct
☐	offer	struct
☐	listingId	string
☐	prodId	string
☐	type	string
☒	purchaseId	string

Row-level access

Choose whether this filter should have row-level restrictions.

- ☐ Access to all rows
- ☒ Filter rows

Row filter expression

Enter the rest of the following query statement `SELECT * FROM nested-table WHERE...`
Please see the documentation for examples of filter expressions.

customer.customerName <> 'John'

Lorsque vous accordez des autorisations à la `customer` colonne, le principal reçoit l'accès à la colonne et aux champs imbriqués situés sous la colonne (`customerName` et `customerID`).

- Expression de filtre de ligne : entrez une expression de filtre pour spécifier le filtrage des lignes ou des cellules. Pour les types de données et les opérateurs pris en charge, consultez [Support partiQL dans les expressions de filtre de ligne](#). Choisissez Accès à toutes les lignes pour accorder l'accès à toutes.

Vous pouvez inclure des structures de colonnes partielles issues de colonnes imbriquées dans une expression de filtre de ligne afin de filtrer les lignes contenant une valeur spécifique.

Lorsqu'un principal est autorisé à accéder à une table avec une expression `Select * from example_nestedtable where customer.customerName <> 'John'` de filtre de ligne et que l'accès au niveau des colonnes est défini sur Accès à toutes les colonnes, les résultats de la requête n'affichent que les lignes dont la valeur est `customerName <> 'John'` vraie.

La capture d'écran suivante montre un filtre de données qui implémente le filtrage des cellules. Dans les requêtes portant sur la `orders` table, il refuse l'accès à la `customer_name` colonne et affiche uniquement les lignes contenant « `pharma` » dans la `product_type` colonne.

Create data filter



Data filter name

Enter a name that describes this data access filter.

restrict-pharma

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.

Target database

Select the database that contains the target table.

Choose databases ▼

Load more

sales



054881201579

Target table

Select the table for which the data filter will be created.

Choose tables ▼

Load more

orders



054881201579

Column-level access

Choose whether this filter should have column-level restrictions.

- ☐ Access to all columns
Filter won't have any column restrictions.
- ☐ Include columns
Filter will only allow access to specific columns.
- ☒ Exclude columns
Filter will allow access to all but specific columns.

Select columns

Choose one or more columns ▼

customer_name
string



5. Choisissez Create filter (Créer un filtre).

Pour créer un filtre de données avec des politiques de filtre cellulaire sur un champ imbriqué

Cette section utilise l'exemple de schéma suivant pour montrer comment créer un filtre de cellules de données :

```
[
  { name: "customer", type: "struct<customerId:string,customerName:string>" },
  { name: "customerApplication", type: "struct<appId:string>" },
  { name: "product", type:
"struct<offer:struct<prodId:string,listingId:string>,type:string>" },
  { name: "purchaseId", type: "string" },
]
```

1. Sur la page Créer un filtre de données, entrez le nom du filtre de données.
2. Ensuite, utilisez le menu déroulant pour choisir le nom de la base de données et le nom de la table.
3. Dans la section Accès au niveau des colonnes, choisissez Colonnes incluses, puis sélectionnez une colonne imbriquée (). `customer.customerName`
4. Dans la section Accès au niveau des lignes, choisissez l'option Accès à toutes les lignes.
5. Choisissez Create filter (Créer un filtre).

Lorsque vous SELECT autorisez ce filtre, le principal a accès à toutes les lignes de la `customerName` colonne.

6. Définissez ensuite un autre filtre de données pour la même base de données/table.
7. Dans la section Accès au niveau des colonnes, choisissez Colonnes incluses, puis sélectionnez une autre colonne imbriquée (). `customer.customerId`
8. Dans la section Accès au niveau des lignes, choisissez Filtrer les lignes, puis entrez une expression de filtre de ligne () `customer.customerId <> 5`.
9. Choisissez Create filter (Créer un filtre).

Lorsque vous SELECT autorisez ce filtre, le principal a accès à toutes les lignes des `customerId` champs `customerName`, à l'exception de la cellule dont la valeur est 5 dans la `customerId` colonne.

Octroi d'autorisations de filtrage de données

Vous pouvez accorder les SELECT autorisations DESCRIBE et DROP Lake Formation sur les filtres de données aux principaux.

Dans un premier temps, vous êtes le seul à pouvoir visualiser les filtres de données que vous créez pour une table. Pour permettre à un autre principal de consulter un filtre de données et d'accorder des autorisations au catalogue de données avec le filtre de données, vous devez soit :

- Accordez SELECT sur un tableau au directeur avec l'option de subvention, et appliquez le filtre de données à la subvention.
- Accordez l'DROP autorisation DESCRIBE or sur le filtre de données au principal.

Vous pouvez accorder l'SELECT autorisation à un AWS compte externe. L'administrateur du lac de données de ce compte peut ensuite accorder cette autorisation aux autres principaux du compte. Lorsque vous accordez des autorisations à un compte externe, vous devez inclure l'option d'octroi afin que l'administrateur du compte externe puisse étendre l'autorisation aux autres utilisateurs du his/her compte. Lorsque vous accordez à un mandant de votre compte, l'option de subvention est facultative.

Vous pouvez accorder et révoquer des autorisations sur les filtres de données à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.
3. Sur la page Autorisations, dans la section Autorisations relatives aux données, choisissez Grant.
4. Sur la page Accorder les autorisations relatives aux données, choisissez les principaux auxquels accorder les autorisations.
5. Dans la section Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées. Choisissez ensuite la base de données, la table et le filtre de données pour lesquels vous souhaitez accorder des autorisations.

LF-Tags or catalog resources

☐ Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ Named data catalog resources
Manager permissions for specific databases or tables, in addition to fine-grained data access.

Databases

Select one or more databases.

Choose databases ▼

cloudtrail 106567286946 ✕

Load more

Tables - optional

Select one or more tables.

Choose tables ▼

cloudtrail_logs_awslogs 106567286946 ✕

Load more

Data filters - optional

Select one or more data filters.

Choose data filters ▼

cloudtrail_lakeformation_filter 106567286946 ✕

Load more

Create new

[Manage data filters](#)

6. Dans la section Autorisations du filtre de données, choisissez les autorisations que vous souhaitez accorder aux principaux sélectionnés.

Data filter permissions

Data filter permissions

Choose specific access permissions to grant.

☐ Select ☒ Describe ☒ Drop

Grantable permissions

Choose the permission that may be granted to others.

☐ Select ☐ Describe ☐ Drop

AWS CLI

- Entrez une `grant-permissions` commande. Spécifiez `DataCellsFilter` pour l'`resource` argument, et spécifiez `DESCRIBE` ou `DROP` pour l'`Permissions` argument et, éventuellement, pour l'`PermissionsWithGrantOption` argument.

L'exemple suivant octroie l'option `grant DESCRIBE` à l'utilisateur `datalake_user1` sur le filtre de données `restrict-pharma`, qui appartient à la `orders` table de la `sales` base de données dans le AWS compte `1111-2222-3333`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Voici le contenu du fichier `grant-params.json`.

```
{
  "Principal": {"DataLakePrincipalIdentifier":
    "arn:aws:iam::111122223333:user/datalake_user1"},
  "Resource": {
    "DataCellsFilter": {
      "TableCatalogId": "111122223333",
      "DatabaseName": "sales",
      "TableName": "orders",
      "Name": "restrict-pharma"
    }
  },
  "Permissions": ["DESCRIBE"],
  "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

Octroi des autorisations de données fournies par les filtres de données

Les filtres de données représentent un sous-ensemble de données d'une table. Pour fournir un accès aux données aux principaux, `SELECT` des autorisations doivent être accordées à ces principaux.

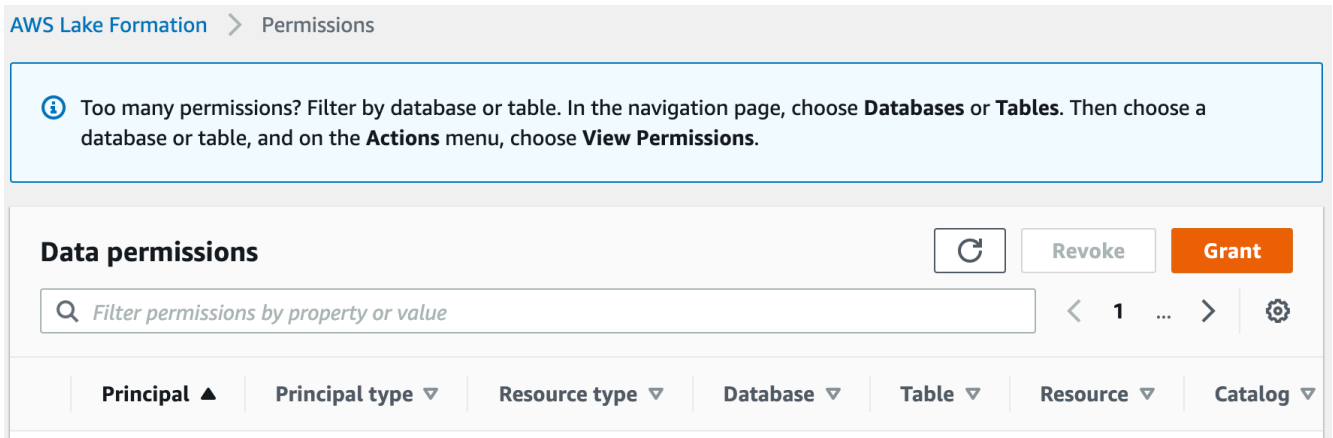
Avec cette autorisation, les directeurs peuvent :

- Afficher le nom réel de la table dans la liste des tables partagées avec leur compte.
- Créez des filtres de données sur la table partagée et accordez des autorisations à leurs utilisateurs sur ces filtres de données.

Console

Pour accorder des autorisations SELECT

1. Accédez à la page Autorisations de la console Lake Formation, puis choisissez Grant.



2. Sélectionnez les principaux auxquels vous souhaitez donner accès, puis sélectionnez Ressources de catalogue de données nommées.

LF-Tags or catalog resources

☐ Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ Named data catalog resources
Manager permissions for specific databases or tables, in addition to fine-grained data access.

Databases

Select one or more databases.

Choose databases ▼

cloudtrail 106567286946 ✕

Load more

Tables - optional

Select one or more tables.

Choose tables ▼

cloudtrail_logs_awslogs 106567286946 ✕

Load more

Data filters - optional

Select one or more data filters.

Choose data filters ▼

cloudtrail_lakeformation_filter 106567286946 ✕

Load more

Create new

[Manage data filters](#) 

3. Pour donner accès aux données représentées par le filtre, choisissez Sélectionner sous Autorisations du filtre de données.

Data filter permissions

Data filter permissions

Choose specific access permissions to grant.

☒ Select ☐ Describe ☐ Drop

Grantable permissions

Choose the permission that may be granted to others.

☐ Select ☐ Describe ☐ Drop

 Select permissions on data filters will grant access to the table 'cloudtrail_logs_awslogs'.

CLI

Entrez une `grant-permissions` commande. Spécifiez `DataCellsFilter` l'argument ressource et spécifiez `SELECT` l'argument `Permissions`.

L'exemple suivant octroie `SELECT` l'option `grant` à l'utilisateur `datalake_user1` sur le filtre de données `restrict-pharma`, qui appartient à la `orders` table de la `sales` base de données dans Compte AWS 1111-2222-3333.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Voici le contenu du fichier `grant-params.json`.

```
{
  "Principal": {
    "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
  },
  "Resource": {
    "DataCellsFilter": {
      "TableCatalogId": "111122223333",
      "DatabaseName": "sales",
      "TableName": "orders",
      "Name": "restrict-pharma"
    }
  },
}
```

```
"Permissions": ["SELECT"]
}
```

Affichage des filtres de données

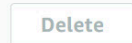
Vous pouvez utiliser la console Lake Formation ou AWS CLI l'API Lake Formation pour afficher les filtres de données.

Pour afficher les filtres de données, vous devez être un administrateur de Data Lake ou disposer des autorisations requises sur les filtres de données.

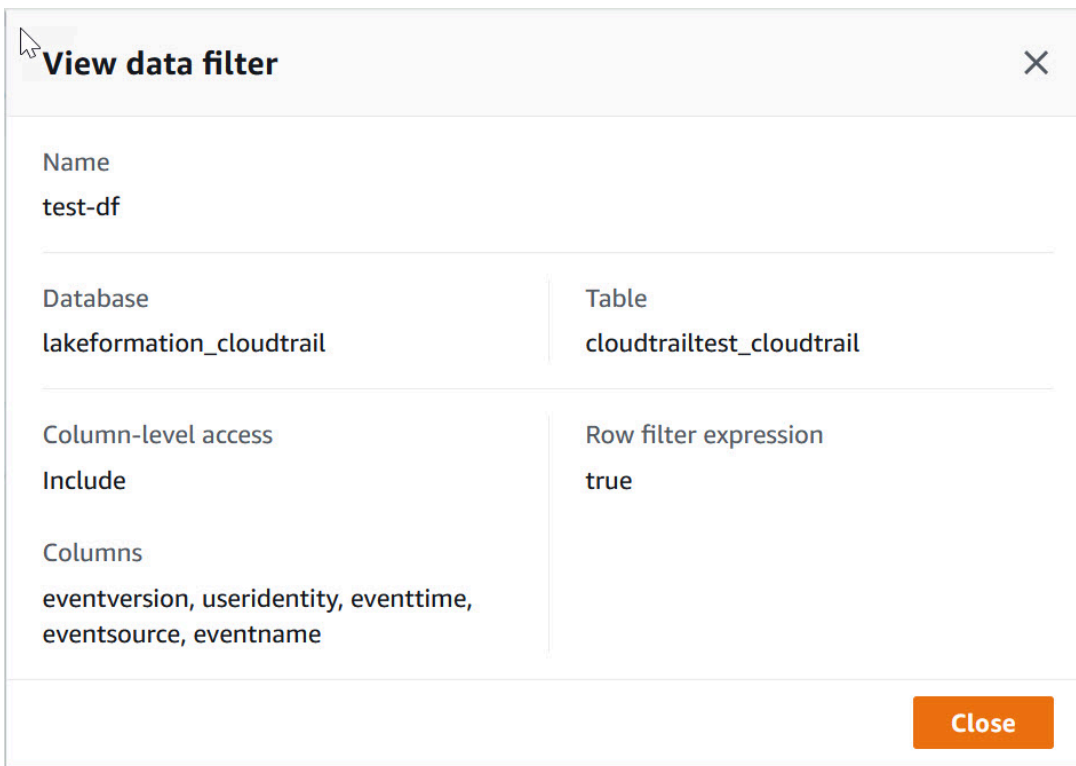
Console

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sous Catalogue de données, sélectionnez Filtres de données.

La page affiche les filtres de données auxquels vous avez accès.

Data filters (1)							
 Find filter				< 1 > 			
	Filter name ▼	Table ▼	Database ▼	Table catalog ID ▼			
☐	test-df	cloudtrailtest_cloudtrail	lakeformation_cloudtrail				

3. Pour afficher les détails du filtre de données, choisissez le filtre de données, puis sélectionnez Afficher. Une nouvelle fenêtre apparaît avec les informations détaillées du filtre de données.



The image shows a 'View data filter' dialog box with a close button (X) in the top right corner. It contains the following fields:

- Name:** test-df
- Database:** lakeformation_cloudtrail
- Table:** cloudtrailtest_cloudtrail
- Column-level access:** Include
- Row filter expression:** true
- Columns:** eventversion, useridentity, eventtime, eventsource, eventname

A 'Close' button is located at the bottom right of the dialog.

AWS CLI

Entrez une `list-data-cells-filter` commande et spécifiez une ressource de table.

L'exemple suivant répertorie les filtres de données de la `cloudtrailtest_cloudtrail` table.

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}
```

API/SDK

Utilisez l'`ListDataCellsFilterAPI` et spécifiez une ressource de table.

L'exemple suivant utilise Python pour répertorier les 20 premiers filtres de données de la `myTable` table.

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
```

```
},
MaxResults=20
)
```

Autorisations de filtrage des données de liste

Vous pouvez utiliser la console Lake Formation pour consulter les autorisations accordées sur les filtres de données.

Pour consulter les autorisations sur un filtre de données, vous devez être un administrateur de Data Lake ou disposer des autorisations requises sur le filtre de données.

Console

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données.
3. Sur la page Autorisations relatives aux données, cliquez ou appuyez dans le champ de recherche, puis dans le menu Propriétés, choisissez Type de ressource.
4. Dans le menu Type de ressource, choisissez Type de ressource : filtre de cellules de données.

Les filtres de données pour lesquels vous avez des autorisations sont répertoriés. Vous devrez peut-être faire défiler la page horizontalement pour voir les colonnes Permissions et Accordable.

Data Permissions (58)							Refresh Revoke Grant
7 matches							Clear filter
Resource type: Data cell filter X							
Principal	Resource type	Database	Table	Resource	Catalog	Permissions	
☐ datalake_admin	Data cell filter	sales	orders	no-pharma	111122223333	Describe, Drop, Select	
☐ datalake_admin	Data cell filter	sales	orders	restrict-pharma	111122223333	Describe, Drop, Select	
☐ datalake_user1	Data cell filter	sales	orders	restrict-pharma	111122223333	Describe	
☐ datalake_user2	Data cell filter	sales	orders	restrict-pharma	111122223333	Select	

AWS CLI

- Entrez une `list-permissions` commande. Spécifiez `DataCellsFilter` pour l'`resource` argument, et spécifiez `DESCRIBE` ou `DROP` pour l'`Permissions` argument et, éventuellement, pour l'`PermissionsWithGrantOption` argument.

L'exemple suivant répertorie les `DESCRIBE` autorisations avec l'option d'octroi sur le filtre de données `restrict-pharma`. Les résultats sont limités aux autorisations accordées pour le principal `datalake_user1` et la `orders` table de la `sales` base de données dans le AWS compte `1111-2222-3333`.

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

Voici le contenu du fichier `grant-params.json`.

```
{
  "Principal": {"DataLakePrincipalIdentifier":
    "arn:aws:iam::111122223333:user/datalake_user1"},
  "Resource": {
    "DataCellsFilter": {
      "TableCatalogId": "111122223333",
      "DatabaseName": "sales",
      "TableName": "orders",
      "Name": "restrict-pharma"
    }
  },
  "Permissions": ["DESCRIBE"],
  "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

Affichage des autorisations de base de données et de tables dans Lake Formation

Vous pouvez consulter les autorisations de Lake Formation accordées sur une base de données ou une table de catalogue de données. Vous pouvez le faire en utilisant la console Lake Formation, l'API ou le AWS Command Line Interface (AWS CLI).

À l'aide de la console, vous pouvez consulter les autorisations à partir des pages Bases de données ou Tables, ou à partir de la page Autorisations relatives aux données.

 Note

Si vous n'êtes ni administrateur de base de données ni propriétaire de ressource, vous pouvez consulter les autorisations accordées aux autres principaux sur la ressource uniquement si vous disposez d'une autorisation Lake Formation sur la ressource avec l'option d'attribution.

Outre les autorisations requises pour Lake Formation, vous avez besoin des autorisations Gestion des identités et des accès AWS (IAM) `glue:GetDatabases`, `glue:GetDatabase`, `glue:GetTables`, `glue:GetTable`, et `lakeformation:ListPermissions`.

Pour afficher les autorisations sur une base de données (console, à partir de la page Bases de données)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

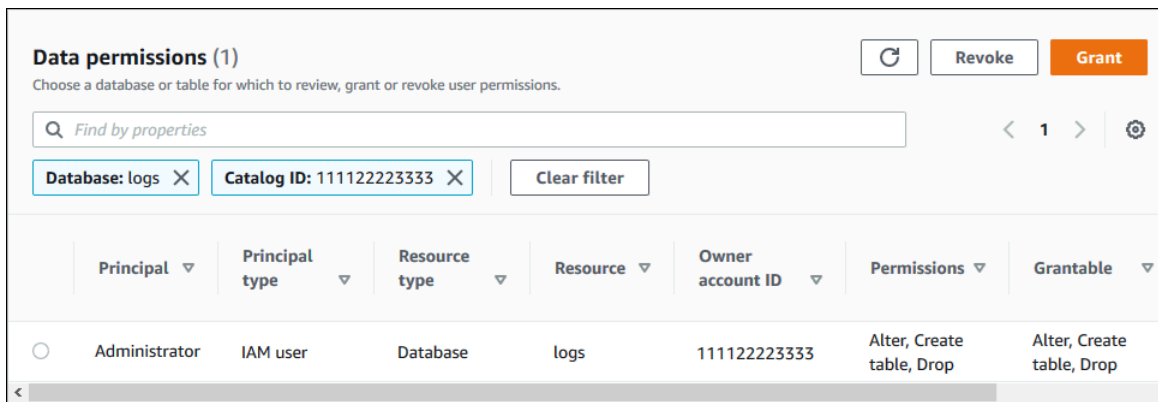
Connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la base de données avec l'option `grant`.

2. Dans le panneau de navigation, choisissez Databases (Bases de données).
3. Choisissez une base de données, puis dans le menu Actions, choisissez Afficher les autorisations.

 Note

Si vous choisissez un lien de ressource de base de données, Lake Formation affiche les autorisations sur le lien de ressource, et non sur la base de données cible du lien de ressource.

La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la base de données. Le nom de la base de données et l'ID de catalogue (ID de AWS compte) du propriétaire de la base de données apparaissent sous forme d'étiquettes sous la zone de recherche. Les vignettes indiquent qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette base de données. Vous pouvez ajuster le filtre en fermant une vignette ou en choisissant Effacer le filtre.



Pour afficher les autorisations sur une base de données (console, à partir de la page Autorisations relatives aux données)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la base de données avec l'option grant.

2. Dans le volet de navigation, sélectionnez Autorisations relatives aux données.
3. Positionnez le curseur dans le champ de recherche en haut de la page, puis dans le menu Propriétés qui apparaît, sélectionnez Base de données.
4. Dans le menu Bases de données qui apparaît, choisissez une base de données.

Note

Si vous choisissez un lien de ressource de base de données, Lake Formation affiche les autorisations sur le lien de ressource, et non sur la base de données cible du lien de ressource.

La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la base de données. Le nom de la base de données apparaît sous forme de vignette sous la zone de recherche. La vignette indique qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette base de données. Vous pouvez supprimer le filtre en fermant la vignette ou en choisissant Effacer le filtre.

Pour afficher les autorisations sur une table (console, à partir de la page Tables)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

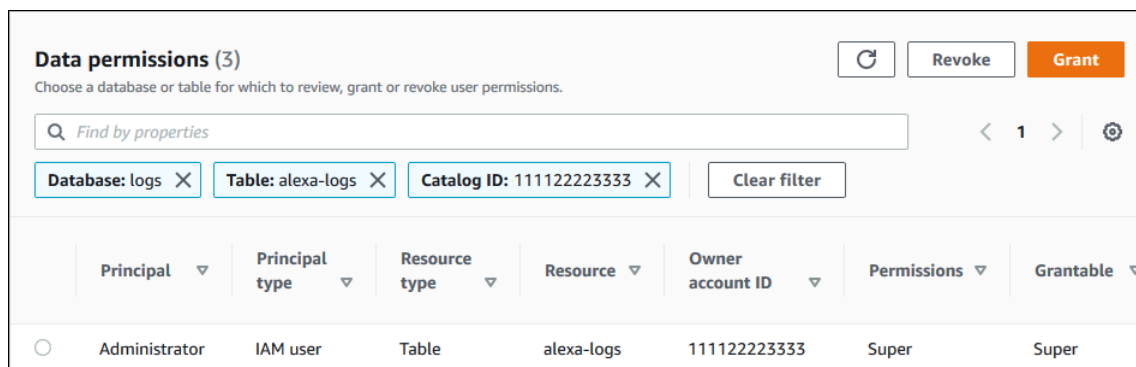
Connectez-vous en tant qu'administrateur du lac de données, créateur de la table ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la table avec l'option d'attribution.

2. Dans le volet de navigation, choisissez Tables.
3. Choisissez un tableau, puis dans le menu Actions, choisissez Afficher les autorisations.

Note

Si vous choisissez un lien de ressource de table, Lake Formation affiche les autorisations sur le lien de ressource, et non sur la table cible du lien de ressource.

La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la table. Le nom de la table, le nom de la base de données contenant la table et l'ID de catalogue (ID de AWS compte) du propriétaire de la table apparaissent sous forme d'étiquettes sous la zone de recherche. Les étiquettes indiquent qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette table. Vous pouvez ajuster le filtre en fermant une étiquette ou en choisissant Effacer le filtre.



Principal	Principal type	Resource type	Resource	Owner account ID	Permissions	Grantable
Administrator	IAM user	Table	alexa-logs	111122223333	Super	Super

Pour afficher les autorisations sur une table (console, à partir de la page Autorisations relatives aux données)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données, créateur de la table ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la table avec l'option d'attribution.

2. Dans le volet de navigation, sélectionnez Autorisations relatives aux données.
3. Positionnez le curseur dans le champ de recherche en haut de la page, puis dans le menu Propriétés qui apparaît, sélectionnez Base de données.
4. Dans le menu Bases de données qui apparaît, choisissez une base de données.

 Important

Si vous souhaitez consulter les autorisations sur une table partagée avec votre AWS compte à partir d'un compte externe, vous devez choisir la base de données du compte externe qui contient la table, et non un lien de ressource vers la base de données.

La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la base de données.

5. Positionnez à nouveau le curseur dans la zone de recherche, puis dans le menu Propriétés qui apparaît, sélectionnez Tableau.
6. Dans le menu Tables qui apparaît, choisissez un tableau.

La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la table. Le nom de la table et le nom de la base de données contenant la table apparaissent sous forme de vignettes sous la zone de recherche. Les vignettes indiquent qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette table. Vous pouvez ajuster le filtre en fermant une vignette ou en choisissant Effacer le filtre.

Pour afficher les autorisations sur une table (AWS CLI)

- Entrez une `list-permissions` commande.

L'exemple suivant répertorie les autorisations sur une table partagée depuis un compte externe. La `CatalogId` propriété est l'ID de AWS compte du compte externe, et le nom de la base de données fait référence à la base de données du compte externe qui contient la table.

```
aws lakeformation list-permissions --resource-type TABLE --resource '{ "Table":  
  {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"} }'
```

Révocation de l'autorisation à l'aide de la console Lake Formation

Vous pouvez utiliser la console pour révoquer tous les types d'autorisations relatives à Lake Formation : autorisations relatives au catalogue de données, autorisations relatives aux balises de politique, autorisations de filtrage des données et autorisations de localisation.

Pour révoquer les autorisations de Lake Formation sur une ressource (console)

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur ayant obtenu des autorisations avec l'option d'octroi sur la ressource.

2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données, balises LF et autorisations, ou Emplacements des données.
3. Sélectionnez l'autorisation ou l'emplacement, puis choisissez Révoquer.
4. Dans la boîte de dialogue qui s'ouvre, choisissez Révoquer.

Partage de données entre comptes dans Lake Formation

Les fonctionnalités multi-comptes de Lake Formation permettent aux utilisateurs de partager en toute sécurité des lacs de données distribués entre plusieurs AWS organisations ou directement avec les responsables IAM d'un autre compte Comptes AWS, offrant ainsi un accès détaillé aux métadonnées du catalogue de données et aux données sous-jacentes. Les grandes entreprises en utilisent généralement plusieurs Comptes AWS, et bon nombre de ces comptes peuvent avoir besoin d'accéder à un lac de données géré par un seul Compte AWS. Les utilisateurs et les tâches AWS Glue d'extraction, de transformation et de chargement (ETL) peuvent interroger et joindre des tables sur plusieurs comptes tout en bénéficiant de la protection des données au niveau des tables et des colonnes de Lake Formation.

Lorsque vous accordez des autorisations sur une ressource du catalogue de données à un compte externe ou directement à un responsable IAM d'un autre compte, Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour partager la ressource. Si le compte du bénéficiaire appartient à la même organisation que le compte du concédant, la ressource partagée est immédiatement accessible au bénéficiaire. Si le compte du bénéficiaire n'appartient pas à la même organisation, AWS RAM envoie une invitation au compte du bénéficiaire pour qu'il accepte ou rejette la subvention de ressources. Ensuite, pour rendre la ressource partagée disponible,

l'administrateur du lac de données du compte bénéficiaire doit utiliser la AWS RAM console ou AWS CLI accepter l'invitation.

Lake Formation prend en charge le partage des ressources du catalogue de données avec des comptes externes en mode d'accès hybride. Le mode d'accès hybride offre la flexibilité d'activer de manière sélective les autorisations de Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog.

Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Pour de plus amples informations, veuillez consulter [Mode d'accès hybride](#).

Partage direct entre comptes

Les principaux autorisés peuvent partager des ressources de manière explicite avec un directeur IAM sur un compte externe. Cette fonctionnalité est utile lorsqu'un propriétaire de compte souhaite contrôler les utilisateurs du compte externe qui peuvent accéder aux ressources. Les autorisations que recevra le directeur de l'IAM seront une combinaison de subventions directes et de subventions au niveau du compte, qui seront répercutées en cascade sur les principaux. Seul le bénéficiaire de l'autorisation peut consulter les autorisations directes entre comptes. Le principal qui reçoit la part de ressources ne peut pas partager la ressource avec d'autres principaux.

Méthodes de partage des ressources du catalogue de données

Avec une seule opération de subvention Lake Formation, vous pouvez accorder des autorisations entre comptes sur les ressources du catalogue de données suivantes.

- Une base de données
- Un tableau individuel (avec filtrage des colonnes optionnel)
- Quelques tables sélectionnées
- Toutes les tables d'une base de données (en utilisant le caractère générique Toutes les tables)

Il existe deux options pour partager vos bases de données et vos tables avec un autre compte Compte AWS ou avec les principaux IAM d'un autre compte.

- Contrôle d'accès basé sur des balises Lake Formation (LF-TBAC) (recommandé)

Le contrôle d'accès basé sur les balises de Lake Formation est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Vous pouvez utiliser le contrôle d'accès basé sur des balises pour partager les ressources du catalogue de données (bases de données, tables et colonnes) avec des responsables, des Comptes AWS organisations et des unités organisationnelles IAM externes (). OUs Dans Lake Formation, ces attributs sont appelés balises LF. Pour plus d'informations, voir [Gestion d'un lac de données à l'aide du contrôle d'accès basé sur les balises Lake Formation](#).

 Note

La méthode LF-TBAC d'octroi des autorisations de catalogue de données est utilisée AWS Resource Access Manager pour les autorisations entre comptes.

Lake Formation prend désormais en charge l'octroi d'autorisations entre comptes aux Organisations et aux unités organisationnelles à l'aide de la méthode LF-TBAC.

Pour activer cette fonctionnalité, vous devez mettre à jour les paramètres de version des comptes Cross vers la version 3 ou supérieure.

Pour de plus amples informations, veuillez consulter [Mise à jour des paramètres de version de partage de données entre comptes](#).

- Ressources nommées Lake Formation

La méthode de partage de données entre comptes Lake Formation à l'aide de ressources nommées vous permet d'accorder des autorisations Lake Formation avec une option d'octroi sur les tables et les bases de données du catalogue de données à des responsables externes Comptes AWS, à des organisations ou à des unités organisationnelles IAM. L'opération de subvention partage automatiquement ces ressources.

 Note

Vous pouvez également autoriser le AWS Glue robot d'exploration à accéder à un magasin de données dans un autre compte à l'aide des informations d'identification de Lake Formation. Pour plus d'informations, consultez la [section Exploration entre comptes dans le Guide](#) du AWS Glue développeur.

Les services intégrés tels qu'Athena et Amazon Redshift Spectrum nécessitent des liens de ressources pour pouvoir inclure des ressources partagées dans les requêtes. Pour plus d'informations sur les liens vers des ressources, consultez [Mode de fonctionnement des liens des ressources dans Lake Formation](#).

Pour les considérations et les limites, voir [Meilleures pratiques et considérations relatives au partage de données entre comptes](#).

Rubriques

- [Conditions préalables](#)
- [Mise à jour des paramètres de version de partage de données entre comptes](#)
- [Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des entités IAM](#)
- [Octroi d'autorisations sur une base de données ou une table partagée avec votre compte](#)
- [Octroi d'autorisations relatives aux liens vers](#)
- [Accès aux données sous-jacentes d'une table partagée](#)
- [Journalisation entre comptes CloudTrail](#)
- [Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation](#)
- [Afficher toutes les subventions entre comptes à l'aide de l'opération GetResourceShares API](#)

Rubriques en relation

- [Vue d'ensemble des autorisations relatives à Lake Formation](#)
- [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#)
- [Création de liens vers des ressources](#)
- [Résolution des problèmes d'accès entre comptes](#)

Conditions préalables

Avant que votre AWS compte puisse partager les ressources du catalogue de données (catalogues, bases de données et tables) avec un autre compte ou des principaux d'un autre compte, et avant de

pouvoir accéder aux ressources partagées avec votre compte, les conditions préalables suivantes doivent être remplies.

Exigences générales en matière de partage de données entre comptes

- Pour partager des bases de données et des tables du catalogue de données en mode d'accès hybride et partager des objets dans les catalogues fédérés, vous devez mettre à jour les paramètres de version entre comptes vers la version 4.
- Avant d'accorder des autorisations entre comptes sur une ressource de catalogue de données, vous devez révoquer toutes les autorisations de Lake Formation accordées au `IAMAllowedPrincipals` groupe pour cette ressource. Si le principal appelant dispose d'autorisations intercomptes pour accéder à une ressource et que `IAMAllowedPrincipals` cette autorisation existe sur la ressource, Lake Formation lance la requête `AccessDeniedException`.

Cette exigence s'applique uniquement lorsque vous enregistrez l'emplacement des données sous-jacentes en mode Lake Formation. Si vous enregistrez l'emplacement des données en mode hybride, les autorisations de `IAMAllowedPrincipals` groupe peuvent exister sur la base de données ou la table partagée.

- Pour les bases de données contenant des tables que vous avez l'intention de partager, vous devez empêcher que les nouvelles tables soient associées par défaut `Super` à `IAMAllowedPrincipals`. Sur la console Lake Formation, modifiez la base de données et désactivez `Utiliser uniquement le contrôle d'accès IAM` pour les nouvelles tables de cette base de données ou entrez la AWS CLI commande suivante en la database remplaçant par le nom de la base de données. Si l'emplacement des données sous-jacent est enregistré en mode d'accès hybride, il n'est pas nécessaire de modifier ce paramètre par défaut. En mode d'accès hybride, Lake Formation vous permet d'appliquer de manière sélective les autorisations Lake Formation et les politiques d'autorisations IAM pour Amazon S3 et AWS Glue sur la même ressource.

```
aws glue update-database --name database --database-input  
'{"Name": "database", "CreateTableDefaultPermissions": []}'
```

- Pour accorder des autorisations entre comptes, le concédant doit disposer des autorisations Gestion des identités et des accès AWS (IAM) requises sur et sur le AWS Glue service. AWS RAM La politique AWS gérée `AWSLakeFormationCrossAccountManager` accorde les autorisations requises.

Les administrateurs de lacs de données des comptes recevant des partages de ressources en utilisant AWS RAM doivent appliquer la politique supplémentaire suivante. Il permet à l'administrateur d'accepter AWS RAM des invitations de partage de ressources. Cela permet également à l'administrateur d'activer le partage des ressources avec les organisations.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ram:AcceptResourceShareInvitation",
        "ram:RejectResourceShareInvitation",
        "ec2:DescribeAvailabilityZones",
        "ram:EnableSharingWithAwsOrganization"
      ],
      "Resource": "*"
    }
  ]
}
```

- Si vous souhaitez partager des ressources du catalogue de données avec AWS Organizations ou des unités organisationnelles, le partage avec les organisations doit être activé dans AWS RAM.

Pour plus d'informations sur la manière d'activer le partage avec les organisations, voir [Activer le partage avec AWS les organisations](#) dans le Guide de AWS RAM l'utilisateur.

Vous devez être `ram:EnableSharingWithAwsOrganization` autorisé à activer le partage avec les organisations.

- Pour partager des ressources directement avec le principal IAM d'un autre compte, vous devez mettre à jour les paramètres de version entre comptes vers la version 3. Ce paramètre est disponible sur la page des paramètres du catalogue de données. Si vous utilisez la version 1, consultez les instructions pour mettre à jour le paramètre [Mise à jour des paramètres de version de partage de données entre comptes](#).
- Vous ne pouvez pas partager les ressources du catalogue de données chiffrées avec une clé gérée par le AWS Glue service avec un autre compte. Vous ne pouvez partager que les ressources du catalogue de données chiffrées avec la clé de chiffrement du client, et le compte recevant le

partage des ressources doit disposer des autorisations sur la clé de chiffrement du catalogue de données pour déchiffrer les objets.

Partage de données entre comptes selon les exigences du LF-TBAC

- Pour partager les ressources du catalogue de données avec AWS Organizations des unités organisationnelles (OUs), vous devez mettre à jour les paramètres de version entre comptes vers la version 3 ou supérieure.
- Pour partager les ressources du catalogue de données avec la version 3 des paramètres de version entre comptes, le concédant doit disposer des autorisations IAM définies dans la politique AWS `AWSLakeFormationCrossAccountManager` gérée de votre compte.
- Si vous utilisez la version 1 ou la version 2 des paramètres de version entre comptes, vous devez disposer d'une politique de ressources du catalogue de données (`glue:PutResourcePolicy`) qui active le LF-TBAC. Pour de plus amples informations, veuillez consulter [Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation](#).
- Si vous utilisez actuellement une politique de ressources du catalogue de AWS Glue données pour partager des ressources et que vous souhaitez accorder des autorisations entre comptes à l'aide de la version 3 des paramètres de version entre comptes, vous devez ajouter `glue:ShareResourceautorisation` dans les paramètres du catalogue de données à l'aide de l'opération `glue:PutResourcePolicyAPI`, comme indiqué dans la [Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation](#) section. Cette politique n'est pas requise si votre compte n'a accordé aucune autorisation entre comptes en utilisant la politique de ressources du catalogue de AWS Glue données (`glue:PutResourcePolicyautorisation` d'utilisation des versions 1 et 2) pour accorder un accès entre comptes.

```
{
  "Effect": "Allow",
  "Action": [
    "glue:ShareResource"
  ],
  "Principal": {"Service": [
    "ram.amazonaws.com"
  ]},
  "Resource": [
    "arn:aws:glue:<region>:<account-id>:table/*/*",
    "arn:aws:glue:<region>:<account-id>:database/*",
    "arn:aws:glue:<region>:<account-id>:catalog"
  ]
}
```

```
}
```

- Si votre compte a effectué des partages entre comptes en utilisant la politique de ressources du catalogue de AWS Glue données et que vous utilisez actuellement la méthode des ressources nommées ou le LF-TBAC avec la version 3 des paramètres entre comptes pour partager des ressources, qui utilise AWS RAM pour partager des ressources, vous devez définir l'EnableHybridargument sur 'true' lorsque vous appelez l'opération d'API. `glue:PutResourcePolicy` Pour de plus amples informations, veuillez consulter [Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation](#).

Configuration requise pour chaque compte accédant à la ressource partagée

- Si vous partagez des ressources avec Comptes AWS, au moins un utilisateur du compte client doit être un administrateur de data lake pour consulter les ressources partagées. Pour plus d'informations sur la création d'un administrateur de lac de données, consultez [Création d'un administrateur de lac de données](#).

L'administrateur du lac de données peut accorder des autorisations de Lake Formation sur les ressources partagées aux autres principaux du compte. Les autres principaux ne peuvent pas accéder aux ressources partagées tant que l'administrateur du lac de données ne leur a pas accordé d'autorisations sur les ressources.

- Les services intégrés tels qu'Athena et Redshift Spectrum nécessitent des liens de ressources pour pouvoir inclure des ressources partagées dans les requêtes. Les directeurs doivent créer un lien de ressource dans leur catalogue de données vers une ressource partagée par une autre Compte AWS personne. Pour plus d'informations sur les liens vers des ressources, consultez [Mode de fonctionnement des liens des ressources dans Lake Formation](#).
- Lorsqu'une ressource est partagée directement avec un principal IAM, pour interroger la table à l'aide d'Athena, le principal doit créer un lien vers une ressource. Pour créer un lien vers une ressource, le directeur a besoin de la Lake Formation `CREATE_TABLE` `glue:CreateTable` ou `CREATE_DATABASE` de l'autorisation `glue:CreateDatabase` IAM.

Si le compte producteur partage une table différente dans la même base de données avec le même principal ou un autre principal, ce principal peut immédiatement interroger la table.

 Note

Pour l'administrateur du lac de données et pour les principaux auxquels l'administrateur du lac de données a accordé des autorisations, les ressources partagées apparaissent dans le catalogue de données comme s'il s'agissait de ressources locales (détenues). Les tâches d'extraction, de transformation et de chargement (ETL) peuvent accéder aux données sous-jacentes des ressources partagées.

Pour les ressources partagées, les pages Tables et Bases de données de la console Lake Formation affichent l'ID de compte du propriétaire.

Lorsque les données sous-jacentes d'une ressource partagée sont accessibles, les événements du CloudTrail journal sont générés à la fois dans le compte du destinataire de la ressource partagée et dans le compte du propriétaire de la ressource. Les CloudTrail événements peuvent contenir l'ARN du principal qui a accédé aux données, mais uniquement si le compte du destinataire choisit d'inclure l'ARN principal dans les journaux. Pour de plus amples informations, veuillez consulter [Journalisation entre comptes CloudTrail](#).

Mise à jour des paramètres de version de partage de données entre comptes

Met à AWS Lake Formation jour de temps à autre les paramètres de partage de données entre comptes afin de distinguer les modifications apportées à l'AWS RAM utilisation et de prendre en charge les mises à jour apportées à la fonctionnalité de partage de données entre comptes. Lorsque Lake Formation effectue cette opération, il crée une nouvelle version des paramètres de version du compte Cross.

Principales différences entre les paramètres des versions multi-comptes

Pour plus d'informations sur le fonctionnement du partage de données entre comptes selon les différents paramètres de version entre comptes, consultez les sections suivantes.

 Note

Pour partager des données avec un autre compte, le donateur doit avoir `AWSLakeFormationCrossAccountManager` géré les autorisations liées à la politique IAM. Il s'agit d'une condition préalable pour toutes les versions.

La mise à jour des paramètres de version de plusieurs comptes n'a aucune incidence sur les autorisations dont dispose le destinataire sur les ressources partagées. Cela s'applique lors

de la mise à jour de la version 1 vers la version 2, de la version 2 vers la version 3 et de la version 1 vers la version 3. Consultez les considérations répertoriées ci-dessous lors de la mise à jour des versions.

Version 1

Méthode de ressource nommée : mappe chaque autorisation de Lake Formation accordée entre comptes à un partage de AWS RAM ressources. L'utilisateur (rôle du concédant ou principal) n'a pas besoin d'autorisations supplémentaires.

Méthode LF-TBAC : les autorisations accordées entre comptes sur la Lake Formation ne sont pas AWS RAM utilisées pour partager des données. L'utilisateur doit disposer d'une `glue:PutResourcePolicy` autorisation.

Avantages de la mise à jour des versions : Version initiale, non applicable.

Considérations relatives à la mise à jour des versions : Version initiale - non applicable

Version 2

Méthode des ressources nommées : optimise le nombre de partages de AWS RAM ressources en mappant plusieurs autorisations accordées entre comptes avec un seul partage de AWS RAM ressources. L'utilisateur n'a pas besoin d'autorisations supplémentaires.

Méthode LF-TBAC : les autorisations accordées entre comptes sur la Lake Formation ne sont pas AWS RAM utilisées pour partager des données. L'utilisateur doit disposer d'une `glue:PutResourcePolicy` autorisation.

Avantages de la mise à jour des versions : configuration multi-comptes évolutive grâce à une utilisation optimale de la AWS RAM capacité.

Considérations à prendre en compte lors de la mise à jour des versions : les utilisateurs qui souhaitent accorder des autorisations multi-comptes à Lake Formation doivent disposer des autorisations définies dans la politique `AWSLakeFormationCrossAccountManager` AWS gérée. Dans le cas contraire, vous devez disposer `ram:AssociateResourceShare` des `ram:DisassociateResourceShare` autorisations nécessaires pour partager correctement des ressources avec un autre compte.

Version 3

Méthode des ressources nommées : optimise le nombre de partages de AWS RAM ressources en mappant plusieurs autorisations accordées entre comptes avec un seul partage de AWS RAM ressources. L'utilisateur n'a pas besoin d'autorisations supplémentaires.

Méthode LF-TBAC : Lake Formation utilise AWS RAM pour les subventions entre comptes. L'utilisateur doit ajouter l'`ShareResource` instruction glue : à `l'glue:PutResourcePolicy` autorisation. Le destinataire doit accepter les invitations à partager des ressources provenant de AWS RAM.

Avantages de la mise à jour des versions : prend en charge les fonctionnalités suivantes :

- Permet de partager des ressources de manière explicite avec un directeur IAM dans un compte externe.

Pour de plus amples informations, veuillez consulter [Octroi d'autorisations sur les ressources du catalogue de données](#).

- Permet les partages entre comptes à l'aide de la méthode LF-TBAC pour les Organisations ou les unités organisationnelles (). OUs
- Supprime les frais liés à la gestion de AWS Glue politiques supplémentaires pour les subventions entre comptes.

Considérations relatives à la mise à jour des versions : Lorsque vous utilisez la méthode LF-TBAC pour partager des ressources, si le concédant utilise une version inférieure à la version 3 et que le destinataire utilise la version 3 ou supérieure, le concédant reçoit le message d'erreur suivant : « Demande de subvention entre comptes non valide. Le compte client a opté pour la version multi-comptes : v3. Veuillez passer `CrossAccountVersion DataLakeSetting` à la version minimale v3 (Service : `AmazonDataCatalog` ; Code d'état : 400 ; Code d'erreur : `InvalidInputException`) ». Toutefois, si le concédant utilise la version 3 et que le destinataire utilise la version 1 ou la version 2, les subventions entre comptes utilisant des balises LF sont traitées avec succès.

Les subventions entre comptes accordées à l'aide de la méthode des ressources nommées sont compatibles entre les différentes versions. Même si le compte du concédant utilise une ancienne version (version 1 ou 2) et que le compte du bénéficiaire utilise une version plus récente (version 3 ou supérieure), la fonctionnalité d'accès entre comptes fonctionne parfaitement sans aucun problème de compatibilité ni erreur.

Pour partager des ressources directement avec les responsables IAM sur un autre compte, seul le donateur doit utiliser la version 3.

Les subventions entre comptes accordées à l'aide de la méthode LF-TBAC nécessitent que les utilisateurs disposent d'une politique de AWS Glue Data Catalog ressources dans le compte. Lorsque vous passez à la version 3, le LF-TBAC autorise des utilisations. AWS RAM Pour que les subventions AWS RAM basées sur plusieurs comptes soient couronnées de succès, vous devez ajouter la `glue:ShareResource` déclaration à vos politiques de ressources de catalogue de données existantes, comme indiqué dans la [Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation](#) section.

Version 4

Le concédant a besoin de la version 4 ou supérieure pour partager les ressources du catalogue de données en mode d'accès hybride ou pour partager des objets dans un catalogue fédéré.

Version 5

La version 5 entre comptes améliore le partage des ressources entre comptes en vous permettant de partager un nombre illimité de tables avec un autre compte, éliminant ainsi les limites d'association de ressources par type de ressource. Pour commencer, passez à la version 5 multi-comptes via la console ou l'API Lake Formation. Toute nouvelle autorisation accordée entre comptes utilisera automatiquement des modèles génériques dans le partage des ressources au lieu d'associations de ressources individuelles. Toutes les actions entre comptes existantes continuent de fonctionner, et toutes les Lake Formation existantes APIs restent compatibles.

Avantages de la mise à jour des versions : Cross-account v5 améliore le partage entre comptes, vous permettant de partager des centaines de milliers de tables entre comptes.

Considérations relatives à la mise à jour des versions : les nouvelles autorisations après la mise à niveau de la version 5 ajouteront des modèles de ressources génériques aux partages de ressources AWS Resource Manager existants ou créeront de nouveaux partages avec des modèles génériques. Une fois la mise à niveau vers la version 5, la rétrogradation n'est pas prise en charge.

Optimisez le partage AWS RAM des ressources

Les nouvelles versions (version 2 et supérieures) des subventions entre comptes utilisent de manière optimale la AWS RAM capacité afin de maximiser l'utilisation entre comptes. Lorsque vous partagez

une ressource avec un responsable externe Compte AWS ou un directeur IAM, Lake Formation peut créer un nouveau partage de ressources ou associer la ressource à un partage existant. En s'associant à des actions existantes, Lake Formation réduit le nombre d'invitations à partager des ressources qu'un consommateur doit accepter. La version 5 optimise encore l'utilisation de la RAM en utilisant des modèles de ressources basés sur des caractères génériques au lieu d'associations de ressources individuelles, réduisant ainsi considérablement les associations de ressources par partage de ressources.

Activez AWS RAM les partages via TBAC ou partagez les ressources directement avec les principaux

Pour partager des ressources directement avec les responsables IAM d'un autre compte ou pour activer les partages entre comptes TBAC vers des Organisations ou des unités organisationnelles, vous devez mettre à jour les paramètres de version entre comptes vers la version 3. Pour plus d'informations sur les limites de AWS RAM ressources, consultez [Meilleures pratiques et considérations relatives au partage de données entre comptes](#).

Autorisations requises pour mettre à jour les paramètres des versions entre comptes

Si un fournisseur d'autorisations entre comptes a `AWSLakeFormationCrossAccountManager` géré les autorisations de politique IAM, aucune configuration d'autorisation supplémentaire n'est requise pour le rôle ou le principal du concédant d'autorisations entre comptes. Toutefois, si le concédant entre comptes n'utilise pas la politique gérée, le rôle ou le principal du concédant doit disposer des autorisations IAM suivantes accordées pour que la nouvelle version de la subvention entre comptes soit couronnée de succès.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
        "ram:AssociateResourceShare",
        "ram:DisassociateResourceShare",
        "ram:GetResourceShares"
      ]
    }
  ]
}
```



```
    ],  
    "Resource": "*",  
    "Condition": {  
      "StringLike": {  
        "ram:ResourceShareName": "LakeFormation*"  
      }  
    }  
  }  
]  
}
```

Pour activer la nouvelle version

Procédez comme suit pour mettre à jour les paramètres de version de plusieurs comptes via la AWS Lake Formation console ou le AWS CLI.

Console

1. Choisissez la version 2, la version 3, la version 4 ou la version 5 dans les paramètres de version multi-comptes sur la page des paramètres du catalogue de données. Si vous sélectionnez la version 1, Lake Formation utilisera le mode de partage des ressources par défaut.

AWS Lake Formation > Data catalog settings

Data catalog settings

Default permissions for newly created databases and tables

These settings maintain existing AWS Glue Data Catalog behavior. You can still set individual permissions on databases and tables, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

- ☐ Use only IAM access control for new databases
- ☐ Use only IAM access control for new tables in new databases

Default permissions for AWS CloudTrail

These settings specify the information being shown in AWS CloudTrail.

Resource owners

Enter resource owners you wish to share your CloudTrail access details with.

Enter one or more AWS account IDs. Press Enter after each ID.

Cross account version settings

Version 1

Version 2

Version 3

Version 3

cross account permissions. See

Cancel

Save

2. Choisissez Enregistrer.

AWS Command Line Interface (AWS CLI)

Utilisez la `put-data-lake-settings` AWS CLI commande pour définir le `CROSS_ACCOUNT_VERSION` paramètre. Les valeurs acceptées sont 1, 2, 3, 4 et 5.

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
file://settings
{
```

```
"DataLakeAdmins": [  
  {  
    "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/test"  
  }  
],  
"CreateDatabaseDefaultPermissions": [],  
"CreateTableDefaultPermissions": [],  
"Parameters": {  
  "CROSS_ACCOUNT_VERSION": "3"  
}  
}
```

Important

Une fois que vous aurez choisi la version 2 ou la version 3, toutes les nouvelles subventions de ressources nommées passeront par le nouveau mode d'attribution entre comptes. Pour utiliser de manière optimale la AWS RAM capacité de vos partages entre comptes existants, nous vous recommandons de révoquer les autorisations accordées avec l'ancienne version et de les réattribuer dans le nouveau mode.

Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des entités IAM

Cette section contient des instructions sur la manière d'accorder des autorisations entre comptes sur les ressources du catalogue de données à un AWS compte externe, à un responsable IAM, à une AWS organisation ou à une unité organisationnelle. L'opération de subvention partage automatiquement ces ressources.

Rubriques

- [Partage de données à l'aide du contrôle d'accès basé sur des balises](#)
- [Partage de données entre comptes à l'aide de la méthode des ressources nommées](#)

Partage de données à l'aide du contrôle d'accès basé sur des balises

AWS Lake Formation le contrôle d'accès basé sur des balises (LF-TBAC) est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Les étapes suivantes expliquent comment accorder des autorisations entre comptes à l'aide de balises LF.

Configuration requise sur le producer/grantor compte

1. Ajoutez des balises LF.
 1. Connectez-vous à la console Lake Formation en tant qu'administrateur de data lake ou créateur de balises LF.
 2. Dans la barre de navigation de gauche, choisissez Permissions et LF-Tags et autorisations.
 3. Choisissez Ajouter un tag LF.

Pour obtenir des instructions détaillées sur la création de balises LF, consultez. [Création de balises LF](#)

2. Accorder and/or Describe Associate des autorisations LF-tag pour les paires clé-valeur aux principaux IAM de votre compte ou de comptes externes.

L'octroi d'autorisations sur les paires clé-valeur des balises LF permet aux principaux de visualiser les balises LF et de les attribuer aux ressources du catalogue de données (bases de données, tables et colonnes).

3. Ensuite, l'administrateur du lac de données ou un responsable IAM disposant de l'autorisation Associate peut attribuer le tag LF aux bases de données, aux tables ou aux colonnes. Pour de plus amples informations, veuillez consulter [Affectation de balises LF aux ressources du catalogue de données](#).
4. Accordez ensuite l'autorisation d'accès aux données à des comptes externes à l'aide d'expressions LF-Tag. Cela permet au bénéficiaire ou au destinataire des autorisations d'accéder aux ressources du catalogue de données qui sont étiquetées avec les mêmes clés et valeurs.
 1. Dans le volet de navigation, sélectionnez Autorisations et autorisations de données.
 2. Choisissez Accorder.
 3. Sur la page Autorisations d'octroi, pour les principaux, choisissez External accounts et entrez l' Compte AWS ID du bénéficiaire ou le rôle IAM du principal ou le nom de ressource Amazon (ARN) du principal (ARN principal) si vous accordez une subvention directe entre comptes à un principal externe. Vous devez appuyer sur Entrée après avoir saisi l'identifiant du compte.

Principals

Choose the principals to grant permissions.

☐ **IAM users and roles**
Users or roles from this AWS account.

☐ **IAM Identity Center** - *new*
Users and groups configured in IAM Identity Center.

☐ **SAML users and groups**
SAML users and group or QuickSight ARNs.

☒ **External accounts**
AWS account, AWS organization or IAM principal outside of this account

AWS account, AWS organization, or IAM principal outside of this account

Enter one or more AWS account IDs, AWS organization IDs, or IAM principal ARNs. Press Enter after each ID or ARN.

Granting data permissions to organizations is not supported when granting permissions by using LF-Tags.

LF-Tags or catalog resources

Choose a method to grant permissions.

☒ **Resources matched by LF-Tags (recommended)**
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☐ **Named Data Catalog resources**
Manage permissions for specific databases or tables, in addition to fine-grained data access.

☒ LF-Tag key-value pairs
☐ Saved LF-Tag expressions - *new*

Key

Department
▼

Values

Choose LF-Tag values
▼

Sales X

Marketing X

Add LF-Tag key-value pair

4. Pour les balises LF ou les ressources du catalogue, choisissez Resources matching by LF-Tags (recommandé).
 - a. Choisissez l'option Paires clé-valeur LF-Tag ou Expressions LF-Tag enregistrées.
 - b. Si vous choisissez des paires clé-valeur LF-Tag, entrez la clé et la ou les valeurs du LF-Tag associées à la ressource du catalogue de données partagée avec le compte du bénéficiaire.

Le bénéficiaire reçoit des autorisations sur les ressources du catalogue de données auxquelles une balise LF correspondante a été attribuée dans l'expression LF-Tag. Si

l'expression LF-Tag spécifie plusieurs valeurs par clé de balise, n'importe laquelle des valeurs de balise peut correspondre.

5. Choisissez les autorisations au niveau de la base de données ou au niveau de la table à accorder aux ressources qui correspondent à l'expression LF-Tag.

 Important

Étant donné que l'administrateur du lac de données doit accorder des autorisations sur les ressources partagées aux principaux du compte bénéficiaire, vous devez toujours accorder des autorisations entre comptes avec l'option d'octroi.

Pour de plus amples informations, veuillez consulter [Octroi d'autorisations LF-Tag à l'aide de la console](#).

 Note

Les directeurs qui reçoivent des subventions directes entre comptes n'auront pas l'option Autorisations accordables.

Configuration requise sur le receiving/grantee compte

1. Connectez-vous à la console Lake Formation en tant qu'administrateur du lac de données du compte client.
2. Ensuite, recevez le partage des ressources sur le compte du consommateur.
 1. Ouvrez la AWS RAM console.
 2. Dans le volet de navigation, sous Partagé avec moi, choisissez Resource shares.
 3. Sélectionnez les partages de ressources, choisissez Accepter le partage de ressources.
3. Lorsque vous partagez une ressource avec un autre compte, elle appartient toujours au compte du producteur et n'est pas visible dans la console Athena. Pour que la ressource soit visible dans la console Athena, vous devez créer un lien de ressource pointant vers la ressource partagée. Pour obtenir des instructions sur la création d'un lien vers une ressource, consultez [Création d'un lien de ressource vers une table de catalogue de données partagée](#) et [Création d'un lien de ressource vers une base de données de catalogue de données partagée](#)

1. Choisissez Bases de données ou tables dans le catalogue de données.
2. Sur la Databases/Tables page, choisissez Create, Resource link.
3. Entrez les informations suivantes pour un lien vers une ressource de base de données :
 - Nom du lien vers la ressource : nom unique pour le lien vers la ressource.
 - Catalogue de destination : catalogue dans lequel vous créez le lien vers la ressource.
 - Région de base de données partagée : région de la base de données partagée avec vous si vous créez le lien de ressource dans une autre région.
 - Base de données partagée — Choisissez la base de données partagée.
 - ID de catalogue de la base de données partagée : entrez l'ID de catalogue de la base de données partagée.
4. Choisissez Créer. Vous pouvez voir le lien de ressource nouvellement créé dans la liste des bases de données.

De même, vous pouvez créer un lien de ressource vers une table partagée.

4. Accordez maintenant l'autorisation Describe sur le lien de la ressource aux principaux IAM avec lesquels vous partagez la ressource.
 1. Sur la page Bases de données/tables, sélectionnez le lien vers la ressource, puis dans le menu Actions, choisissez Grant.
 2. Dans la section Accorder des autorisations, sélectionnez les utilisateurs et les rôles IAM.
 3. Choisissez le rôle IAM auquel vous souhaitez accorder l'accès au lien de ressource.
 4. Dans la section Autorisations relatives aux liens vers les ressources, sélectionnez Décrire.
 5. Choisissez Accorder.
5. Ensuite, accordez des autorisations relatives à la valeur clé-tag LF-Tag aux principaux du compte client.

Vous devriez pouvoir trouver les balises LF partagées avec vous dans le compte client de la console Lake Formation, sous Autorisations, balises LF et autorisations. Vous pouvez associer des balises partagées par le donateur aux ressources partagées depuis le compte du donateur, notamment les bases de données, les tables et les colonnes. Vous pouvez également accorder des autorisations sur les ressources à d'autres principaux.

LF-Tags | LF-Tag expressions - new | **LF-Tag permissions** | LF-Tag creators

LF-Tag permissions (8)  [View](#) [Revoke](#) [Grant permissions](#)

View and manage the permissions granted on LF-Tags. [Learn more](#)

< 1 > 

	Principal	Principal type	Resource name	Resource type	LF-Tag permissions	LF-Tag value permissions
☐		AWS account	Department=Marketing, Sales	LF-Tag key value pair	-	Describe
☐		AWS account	Department=Marketing, Sales	LF-Tag key value pair	-	Associate
☐	arn:aws:iam::():/Admin	IAM role	Department	LF-Tag	Alter, Drop	-
☐	arn:aws:iam::():/Admin	IAM role	Department	LF-Tag	-	Describe
☐	arn:aws:iam::():/Admin	IAM role	Department	LF-Tag	-	Associate
☐	arn:aws:iam::():/Admin	IAM role	Department	LF-Tag	-	Grant with LF-Tag expression
☐	arn:aws:iam::():/Admin	IAM role	LF-Test	LF-Tag	-	Describe
☐	arn:aws:iam::():/Admin	IAM role	LF-Test	LF-Tag	-	Associate

1. Dans le volet de navigation, sous Autorisations, Autorisations relatives aux données, choisissez Grant.
2. Sur la page Accorder des autorisations, sélectionnez Utilisateurs et rôles IAM.
3. Choisissez ensuite les utilisateurs et les rôles IAM de votre compte pour accorder l'accès aux bases de données/tables partagées.
4. Ensuite, pour les balises LF ou les ressources du catalogue, choisissez Ressources matching by LF-Tags.
5. Ensuite, choisissez la clé et les valeurs du tag LF qui est partagé avec vous.
6. Choisissez ensuite les autorisations de base de données et de table que vous souhaitez accorder aux utilisateurs et aux rôles IAM. Vous pouvez également choisir des autorisations accordables qui permettent aux utilisateurs et aux rôles IAM d'accorder des autorisations à d'autres utilisateurs/rôles.
7. Choisissez Accorder.
8. Vous pouvez consulter les autorisations accordées sous Autorisations relatives aux données sur la console Lake Formation.

Partage de données entre comptes à l'aide de la méthode des ressources nommées

Vous pouvez accorder des autorisations directement aux principaux d'un autre AWS compte, ou à un compte externe Comptes AWS ou AWS Organizations. Accorder des autorisations de Lake Formation à des organisations ou à des unités organisationnelles revient à accorder l'autorisation Compte AWS à tous les membres de cette organisation ou unité organisationnelle.

Lorsque vous accordez des autorisations à des comptes ou à des organisations externes, vous devez inclure l'option Autorisations pouvant être accordées. Seul l'administrateur du lac de données du compte externe peut accéder aux ressources partagées jusqu'à ce qu'il accorde des autorisations sur les ressources partagées aux autres principaux du compte externe.

 Note

L'option d'autorisations pouvant être accordées n'est pas prise en charge lors de l'octroi d'autorisations directement aux principaux IAM à partir de comptes externes.

Suivez les instructions [Octroi d'autorisations de base de données via la méthode de ressource nommée](#) pour accorder des autorisations entre comptes à l'aide de la méthode de ressource nommée.

Octroi d'autorisations sur une base de données ou une table partagée avec votre compte

Une fois qu'une ressource de catalogue de données appartenant à un autre AWS compte est partagée avec votre AWS compte, en tant qu'administrateur du lac de données, vous pouvez accorder des autorisations sur la ressource partagée aux autres principaux de votre compte. Vous ne pouvez toutefois pas accorder d'autorisations sur la ressource à d'autres AWS comptes ou organisations.

Vous pouvez utiliser la AWS Lake Formation console, l'API ou le AWS Command Line Interface (AWS CLI) pour accorder les autorisations.

Pour accorder des autorisations sur une base de données partagée (méthode de ressource nommée, console)

- Suivez les instructions de la section [Octroi d'autorisations de base de données via la méthode de ressource nommée](#). Dans la liste des bases de données, sous balises LF ou ressources du catalogue, assurez-vous de sélectionner la base de données dans le compte externe, et non un lien de ressource pour la base de données.

Si la base de données ne figure pas dans la liste des bases de données, assurez-vous d'avoir accepté l'invitation de partage de ressources AWS Resource Access Manager (AWS RAM) pour la base de données. Pour de plus amples informations, veuillez consulter [Acceptation d'une invitation de partage de ressources de AWS RAM](#).

De plus, pour les ALTER autorisations CREATE_TABLE et, suivez les [Octroi d'autorisations de localisation des données \(même compte\)](#) instructions fournies et assurez-vous de saisir l'identifiant du compte propriétaire dans le champ Emplacement du compte enregistré.

Pour accorder des autorisations sur une table partagée (méthode de ressource nommée, console)

- Suivez les instructions de la section [Octroi d'autorisations de table à l'aide de la méthode de ressource nommée](#). Dans la liste des bases de données, sous balises LF ou ressources du catalogue, assurez-vous de sélectionner la base de données dans le compte externe, et non un lien de ressource pour la base de données.

Si le tableau ne figure pas dans la liste des tableaux, assurez-vous d'avoir accepté l'invitation de partage de AWS RAM ressources pour le tableau. Pour de plus amples informations, veuillez consulter [Acceptation d'une invitation de partage de ressources de AWS RAM](#).

De plus, pour ALTER obtenir l'autorisation, suivez les [Octroi d'autorisations de localisation des données \(même compte\)](#) instructions fournies et assurez-vous de saisir l'identifiant du compte propriétaire dans le champ Emplacement du compte enregistré.

Pour accorder des autorisations sur des ressources partagées (méthode LF-TBAC, console)

- Suivez les instructions de la section [Octroi d'autorisations au catalogue de données](#). Dans la section Balises LF ou ressources du catalogue, accordez l'expression de balise LF exacte que le compte externe a accordée à votre compte, ou un sous-ensemble de cette expression.

Par exemple, si un compte externe a accordé l'expression LF-tag `module=customers AND environment=production` à votre compte avec l'option d'attribution, en tant qu'administrateur du lac de données, vous pouvez accorder cette même expression, `module=customers` ou `environment=production` à un mandant de votre compte. Vous ne pouvez accorder que les mêmes autorisations ou un sous-ensemble des autorisations de Lake Formation (par exemple, SELECT, ALTER, etc.) qui ont été accordées aux ressources via l'expression LF-Tag.

Pour accorder des autorisations sur une table partagée (méthode de ressource nommée, AWS CLI)

- Utilisez une commande similaire à la suivante. Dans cet exemple :
 - L'identifiant de votre AWS compte est 1111-2222-3333.

- Le compte propriétaire de la table et qui l'a attribuée à votre compte est le 1234-5678-9012.
- L'`SELECT` autorisation est accordée à l'utilisateur sur la table `pageviews` partagée `datalake_user1`. Cet utilisateur est le principal de votre compte.
- La `pageviews` table se trouve dans la `analytics` base de données, qui appartient au compte 1234-5678-9012.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
--permissions "SELECT" --resource '{ "Table": {"CatalogId":"123456789012",
"DatabaseName":"analytics", "Name":"pageviews"}}'
```

Notez que le compte propriétaire doit être spécifié dans la `CatalogId` propriété de l'`resource` argument.

Octroi d'autorisations relatives aux liens vers

Suivez ces étapes pour accorder AWS Lake Formation des autorisations sur un ou plusieurs liens de ressources à un responsable de votre AWS compte.

Après avoir créé un lien vers une ressource, vous êtes le seul à pouvoir le consulter et y accéder. (Cela suppose que l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données n'est pas activée pour la base de données.) Pour permettre aux autres utilisateurs de votre compte d'accéder au lien de la ressource, accordez au moins l'`DESCRIBE` autorisation.

Important

L'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations à la cible séparément.

Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

console

Pour accorder des autorisations relatives aux liens vers des ressources à l'aide de la console Lake Formation

1. Effectuez l'une des actions suivantes :

- Pour les liens vers des ressources de base de données, suivez les étapes [Octroi d'autorisations de base de données via la méthode de ressource nommée](#) décrites dans. pour effectuer les opérations suivantes :
 1. Sélectionnez le lien vers la ressource dans la liste des bases de données sous Catalogue de données, Bases de données.
 2. Choisissez Grant pour ouvrir la page Accorder les autorisations.
 3. Spécifiez les principaux auxquels accorder les autorisations.
 4. Les champs Catalogues et Bases de données sont renseignés.
 - Pour les liens vers les ressources des tables, suivez [Octroi d'autorisations de table à l'aide de la méthode de ressource nommée](#) les étapes décrites ci-dessous :
 1. Sélectionnez le lien vers la ressource dans la liste des tables sous Catalogue de données, Tables.
 2. Ouvrez la page Accorder des autorisations.
 3. Spécifiez les principes.
 4. Les champs Catalogues, Bases de données, Tables sont renseignés.
 5. Spécifiez les principes.
2. Sous Autorisations, sélectionnez les autorisations à accorder. Sélectionnez éventuellement les autorisations pouvant être accordées.

LF-Tags or catalog resources

Choose a method to grant permissions.

☐ **Resources matched by LF-Tags (recommended)**
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.

☒ **Named Data Catalog resources**
Manage permissions for specific databases or tables, in addition to fine-grained data access.

Catalogs

Choose catalogs

Default catalog

Databases

Select one or more databases.

Choose databases

resourcelink

Tables - optional

Select one or more tables.

Choose tables

Views - optional

Select one or more views.

Choose views

Data filters - optional

Select one or more data filters.

Choose data filters

Load more

Create new

[Manage data filters](#)

Resource link permissions

Resource link permissions

Choose specific access permissions to grant.

☐ Describe ☐ Drop

Grantable permissions

Choose the permission that can be granted to others.

☐ Describe ☐ Drop

Cancel

Grant

3. Choisissez Accorder.

AWS CLI

Pour accorder des autorisations de lien vers des ressources à l'aide de AWS CLI

- Exécutez la `grant-permissions` commande en spécifiant un lien de ressource comme ressource.

Exemple

Cet exemple accorde `DESCRIBE` à l'utilisateur `dataLake_user1` sur la table un lien de ressource `incidents-link` dans la base de données du AWS compte issues `1111-2222-3333`.

```
aws lakeformation grant-permissions --principal
DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/dataLake_user1
--permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"issues",
"Name":"incidents-link"}}'
```

 Voir aussi :

- [Création de liens vers des ressources](#)
- [Référence des autorisations de Lake Formation](#)

Accès aux données sous-jacentes d'une table partagée

Supposons que le AWS compte A partage une table du catalogue de données `SELECT` avec le compte B, par exemple, en octroyant au compte B. Pour que le principal du compte B puisse lire les données sous-jacentes de la table partagée, les conditions suivantes doivent être remplies :

- L'administrateur du lac de données du compte B doit accepter le partage. (Cela n'est pas nécessaire si les comptes A et B appartiennent à la même organisation ou si la subvention a été accordée selon la méthode de contrôle d'accès basée sur les balises Lake Formation.)
- L'administrateur du lac de données doit réaccorder au principal l'`SELECT` autorisation Lake Formation accordée par le compte A sur la table partagée.
- Le principal doit disposer des autorisations IAM suivantes sur la table, la base de données qui la contient et le compte A Data Catalog.

 Note

Dans la politique IAM suivante :

- Remplacez *<account-id-A>* par le AWS numéro de compte du compte A.
- Remplacez *<region>* par une région valide.
- Remplacez *<database>* par le nom de la base de données du compte A qui contient la table partagée.
- Remplacez *<table>* par le nom de la table partagée.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartition",
        "glue:GetPartitions",
        "glue:BatchGetPartition",
        "glue:GetDatabase",
        "glue:GetDatabases"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:table/<database>/<table>",
        "arn:aws:glue:us-east-1:111122223333:database/<database>",
        "arn:aws:glue:us-east-1:111122223333:catalog"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

```
    ]  
  }  
]  
}
```

 Voir aussi :

- [Acceptation d'une invitation de partage de ressources de AWS RAM](#)

Journalisation entre comptes CloudTrail

Lake Formation fournit une piste d'audit centralisée de tous les accès entre comptes aux données de votre lac de données. Lorsqu'un AWS compte destinataire accède aux données d'une table partagée, Lake Formation copie l' CloudTrail événement dans les CloudTrail journaux du compte propriétaire. Les événements copiés incluent des requêtes portant sur les données par des services intégrés tels qu' Amazon Athena Amazon Redshift Spectrum, et des accès aux AWS Glue données par des tâches.

CloudTrail les événements relatifs aux opérations entre comptes sur les ressources du catalogue de données sont copiés de la même manière.

En tant que propriétaire de ressources, si vous activez la journalisation au niveau des objets dans Amazon S3, vous pouvez exécuter des requêtes associant des événements S3 à CloudTrail des événements Lake Formation CloudTrail afin de déterminer les comptes qui ont accédé à vos compartiments S3.

Rubriques

- [Inclure les identités principales dans les journaux intercomptes CloudTrail](#)
- [Consultation des CloudTrail journaux pour l'accès entre comptes Amazon S3](#)

Inclure les identités principales dans les journaux intercomptes CloudTrail

Par défaut, les CloudTrail événements intercomptes ajoutés aux journaux du destinataire de la ressource partagée et copiés dans les journaux du propriétaire de la ressource contiennent uniquement l'identifiant AWS principal du compte externe, et non le nom de ressource Amazon (ARN) lisible par l'homme du principal (ARN principal). Lorsque vous partagez des ressources dans des

limites fiables, par exemple au sein d'une même organisation ou d'une même équipe, vous pouvez choisir d'inclure l'ARN principal dans les CloudTrail événements. Les comptes propriétaires des ressources peuvent ensuite suivre les principaux des comptes destinataires qui accèdent à leurs propres ressources.

Important

En tant que destinataire de ressources partagées, pour voir l'ARN principal dans les événements de vos propres CloudTrail journaux, vous devez choisir de partager l'ARN principal avec le compte du propriétaire.

Si l'accès aux données se fait via un lien de ressource, deux événements sont enregistrés dans le compte du destinataire de la ressource partagée : un pour l'accès au lien de ressource et un pour l'accès à la ressource cible. L'événement pour l'accès au lien de ressource inclut l'ARN principal. L'événement pour l'accès à la ressource cible n'inclut pas l'ARN principal sans l'opt-in. L'événement d'accès au lien de ressource n'est pas copié sur le compte du propriétaire.

Ce qui suit est un extrait d'un CloudTrail événement multicompte par défaut (sans opt-in). Le compte effectuant l'accès aux données est le 1111-2222-3333. Il s'agit du journal affiché à la fois dans le compte d'appel et dans le compte du propriétaire de la ressource. Lake Formation remplit les journaux dans les deux comptes dans le cas de comptes croisés.

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "AWSAccount",
    "principalId": "AROAQGFTBBBG0BWV2EMZA:GlueJobRunnerSession",
    "accountId": "111122223333"
  },
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GetDataAccess",
  ...
  ...
  "additionalEventData": {
    "requesterService": "GLUE_JOB",
    "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
  },
  ...
}
```

En tant que consommateur de ressources partagées, lorsque vous choisissez d'inclure l'ARN principal, l'extrait devient le suivant. Le `lakeFormationPrincipal` champ représente le rôle final ou l'utilisateur exécutant la requête via Amazon Athena, Amazon Redshift Spectrum ou des jobs.

AWS Glue

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "AWSAccount",
    "principalId": "AROAQGFTBBBG0BWV2EMZA:GlueJobRunnerSession",
    "accountId": "111122223333"
  },
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GetDataAccess",
  ...
  ...
  "additionalEventData": {
    "requesterService": "GLUE_JOB",
    "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
    "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
  },
  ...
}
```

Pour choisir d'inclure le principal ARNs dans les journaux intercomptes CloudTrail

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'Administrateur utilisateur ou en tant qu'utilisateur avec la politique Administrator Access IAM.

2. Dans le panneau de navigation, sélectionnez Settings (Paramètres).
3. Sur la page des paramètres du catalogue de données, dans la AWS CloudTrail section Autorisations par défaut pour, pour les propriétaires de ressources, entrez un ou plusieurs comptes de propriétaire de AWS ressources IDs.

Appuyez sur Entrée après chaque identifiant de compte.

4. Choisissez Enregistrer.

Désormais, les CloudTrail événements entre comptes stockés dans les journaux du destinataire de la ressource partagée et du propriétaire de la ressource contiennent l'ARN principal.

Consultation des CloudTrail journaux pour l'accès entre comptes Amazon S3

En tant que propriétaire de ressources partagées, vous pouvez interroger les CloudTrail journaux S3 pour déterminer les comptes qui ont accédé à vos compartiments Amazon S3 (à condition que vous ayez activé la journalisation au niveau des objets dans Amazon S3). Cela s'applique uniquement aux sites S3 que vous avez enregistrés auprès de Lake Formation. Si les consommateurs de ressources partagées choisissent d'inclure le principal ARNs dans les CloudTrail journaux de Lake Formation, vous pouvez déterminer les rôles ou les utilisateurs qui ont accédé aux compartiments.

Lorsque vous exécutez des requêtes avec Amazon Athena, vous pouvez joindre les CloudTrail événements Lake Formation et les CloudTrail événements S3 sur la propriété du nom de session. Les requêtes peuvent également filtrer les événements Lake Formation sur `eventName="GetDataAccess"` et les événements S3 sur `eventName="GetObject"` ou `eventName="PutObject"`.

Ce qui suit est un extrait d'un CloudTrail événement inter-comptes de Lake Formation au cours duquel les données d'un emplacement S3 enregistré ont été consultées.

```
{
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GetDataAccess",
  .....
  .....
  "additionalEventData": {
    "requesterService": "GLUE_JOB",
    "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
    "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-B8JSAjo5QA"
  }
}
```

La valeur de la `lakeFormationRoleSessionName` `AWSLF-00-GL-111122223333-B8JSAjo5QA` clé peut être jointe au nom de session dans la `principalId` clé de l' CloudTrail événement S3. Ce qui suit est un extrait de l' CloudTrail événement S3. Il indique l'emplacement du nom de session.

```
{
  "eventSource": "s3.amazonaws.com",
  "eventName": "GetObject"
  .....
  .....
```

```

"principalId": "AROAQSOX5XXUR7D6RMYLR:AWSLF-00-GL-111122223333-B8JSAjo5QA",
"arn": "arn:aws:sets::111122223333:assumed-role/Deformationally/AWSLF-00-GL-111122223333-B8JSAjo5QA",
"session Context": {
  "session Issuer": {
    "type": "Role",
    "principalId": "AROAQSOX5XXUR7D6RMYLR",
    "arn": "arn:aws:iam::111122223333:role/aws-service-role/lakeformation.amazonaws.com/Deformationally",
    "accountId": "111122223333",
    "user Name": "Deformationally"
  },
  .....
  .....
}

```

Le nom de session est formaté comme suit :

```
AWSLF-<version-number>-<query-engine-code>-<account-id>-<suffix>
```

version-number

La version de ce format, actuellement 00. Si le format du nom de session change, la prochaine version sera la suivante 01.

query-engine-code

Indique l'entité qui a accédé aux données. Les valeurs actuelles sont les suivantes :

GL AWS Glue Tâche ETL

AT Athena

RE Amazon Redshift Spectrum

account-id

L'identifiant du AWS compte qui a demandé les informations d'identification à Lake Formation.

suffix

Chaîne générée de manière aléatoire.

Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation

Il est possible d'accorder un accès entre comptes aux ressources du catalogue de données et aux données sous-jacentes en utilisant l'un AWS Glue ou AWS Lake Formation l'autre des deux.

Dans AWS Glue, vous accordez des autorisations entre comptes en créant ou en mettant à jour une politique de ressources du catalogue de données. Dans Lake Formation, vous accordez des autorisations entre comptes en utilisant le modèle d'GRANT/REVOKE autorisations Lake Formation et le fonctionnement de l'Grant Permissions API.

Tip

Nous vous recommandons de vous fier uniquement aux autorisations de Lake Formation pour sécuriser votre lac de données.

Vous pouvez consulter les subventions multicomptes de Lake Formation à l'aide de la console Lake Formation ou de la console AWS Resource Access Manager (AWS RAM). Toutefois, ces pages de console n'affichent pas les autorisations entre comptes accordées par la politique de ressources du catalogue de données de AWS Glue. De même, vous pouvez consulter les autorisations entre comptes dans la politique de ressources du catalogue de données à l'aide de la page Paramètres de la AWS Glue console, mais cette page n'affiche pas les autorisations entre comptes accordées via Lake Formation.

Pour vous assurer de ne manquer aucune subvention lorsque vous consultez et gérez les autorisations entre comptes, Lake Formation vous AWS Glue demande d'effectuer les actions suivantes pour indiquer que vous êtes au courant et que vous autorisez les subventions croisées par Lake Formation et AWS Glue

Lorsque vous accordez des autorisations entre comptes à l'aide de la politique de ressources du catalogue de données de AWS Glue

Si votre compte (compte du donateur ou compte du producteur) n'a accordé aucune subvention entre comptes destinée AWS RAM à partager les ressources, vous pouvez enregistrer une politique de ressources du catalogue de données comme d'habitude dans AWS Glue. Toutefois, si des subventions impliquant AWS RAM des partages de ressources ont déjà été accordées, vous devez effectuer l'une des opérations suivantes pour garantir le succès de l'enregistrement de la politique de ressources :

- Lorsque vous enregistrez la politique de ressources sur la page Paramètres de la AWS Glue console, celle-ci émet une alerte indiquant que les autorisations définies dans la politique s'ajouteront à celles accordées à l'aide de la console Lake Formation. Vous devez choisir Proceed pour enregistrer la politique.
- Lorsque vous enregistrez la politique de ressources à l'aide de l'opération `glue:PutResourcePolicy` API, vous devez définir le `EnableHybrid` champ sur `TRUE` « » (type = chaîne).

Pour mettre à jour une politique de ressources existante, utilisez l'opération `glue:GetResourcePolicy` API pour récupérer d'abord votre politique actuelle, puis modifiez-la selon vos besoins avant de l'appeler `glue:PutResourcePolicy`.

 Note

Lorsque vous créez AWS Glue des politiques de ressources pour l'accès entre comptes, accordez uniquement les autorisations minimales requises pour votre cas d'utilisation spécifique.

Pour plus d'informations, consultez [PutResourcePolicy Action \(Python : put_resource_policy\)](#) dans le manuel du développeur.AWS Glue

Lorsque vous accordez des autorisations entre comptes à l'aide de la méthode des ressources nommées de Lake Formation

S'il n'y a aucune politique de ressources du catalogue de données sur votre compte (compte producteur), les subventions croisées de Lake Formation que vous accordez se poursuivent comme d'habitude. Toutefois, s'il existe une politique de ressources pour le catalogue de données, vous devez y ajouter l'instruction suivante pour permettre à vos subventions entre comptes de réussir si elles sont accordées avec la méthode de ressource nommée. `<region>` Remplacez-le par un nom de région valide et `<account-id>` par votre numéro de AWS compte (numéro de compte producteur).

```
{
  "Effect": "Allow",
  "Action": [
    "glue:ShareResource"
  ],
```

```
"Principal": {"Service": [
  "ram.amazonaws.com"
]},
"Resource": [
  "arn:aws:glue:<region>:<account-id>:table/*/*",
  "arn:aws:glue:<region>:<account-id>:database/*",
  "arn:aws:glue:<region>:<account-id>:catalog"
]
}
```

Sans cette déclaration supplémentaire, la subvention Lake Formation est acceptée, mais elle est bloquée et le compte du bénéficiaire ne peut pas accéder à la ressource accordée. AWS RAM

Important

Lorsque vous utilisez la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) pour accorder des autorisations entre comptes, vous devez disposer d'une politique de ressources du catalogue de données avec au moins les autorisations spécifiées dans. [Conditions préalables](#)

Voir aussi :

- [Contrôle d'accès aux métadonnées](#) (pour une discussion sur la méthode des ressources nommées par rapport à la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC)).
- [Affichage des tables et des bases de données partagées du catalogue de données](#)
- [Utilisation des paramètres du catalogue de données sur la AWS Glue console](#) dans le manuel du AWS Glue développeur
- [Octroi d'un accès entre comptes](#) dans le guide du AWS Glue développeur (pour des exemples de politiques relatives aux ressources du catalogue de données)

Afficher toutes les subventions entre comptes à l'aide de l'opération GetResourceShares API

Si votre entreprise accorde des autorisations entre comptes en utilisant à la fois une politique de AWS Glue Data Catalog ressources et des subventions de Lake Formation, le seul moyen de consulter toutes les autorisations entre comptes en un seul endroit est d'utiliser l'opération `glue:GetResourceShares` API.

Lorsque vous accordez des autorisations à Lake Formation sur plusieurs comptes à l'aide de la méthode des ressources nommées, AWS Resource Access Manager (AWS RAM) crée une politique de ressources Gestion des identités et des accès AWS (IAM) et l'enregistre dans votre AWS compte. La politique accorde les autorisations requises pour accéder à la ressource. AWS RAM crée une politique de ressources distincte pour chaque subvention intercomptes. Vous pouvez consulter toutes ces politiques à l'aide de l'opération `glue:GetResourceShares` API.

 Note

Cette opération renvoie également la politique de ressources du catalogue de données. Toutefois, si vous avez activé le chiffrement des métadonnées dans les paramètres du catalogue de données et que vous n'êtes pas autorisé à utiliser la AWS KMS clé, l'opération ne renverra pas la politique de ressources du catalogue de données.

Pour voir toutes les subventions entre comptes

- Entrez la AWS CLI commande suivante.

```
aws glue get-resource-policies
```

Voici un exemple de politique de ressources qui AWS RAM crée et stocke lorsque vous accordez des autorisations sur une table t dans la base de données db1 au AWS compte 1111-2222-3333.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
```



```
{
  "Effect": "Allow",
  "Action": [
    "glue:GetTable",
    "glue:GetTables",
    "glue:GetTableVersion",
    "glue:GetTableVersions",
    "glue:GetPartition",
    "glue:GetPartitions",
    "glue:BatchGetPartition",
    "glue:SearchTables"
  ],
  "Principal": {"AWS": [
    "111122223333"
  ]},
  "Resource": [
    "arn:aws:glue:us-east-1:111122223333:table/db1/t"
  ]
}
```

 Voir aussi :

- [GetResourceShares Action \(Python : `get\_resource\_policies`\)](#) dans le manuel du développeur AWS Glue

Accès aux tables et aux bases de données partagées du catalogue de données et affichage

Pour l'administrateur du lac de données et pour les principaux auxquels des autorisations ont été accordées, les ressources partagées avec votre AWS compte apparaissent dans le catalogue de données comme s'il s'agissait de ressources de votre compte. La console affiche le compte propriétaire de la ressource.

Vous pouvez consulter les ressources partagées avec votre compte à l'aide de la console Lake Formation. Vous pouvez également utiliser la console AWS Resource Access Manager (AWS RAM)

pour afficher à la fois les ressources partagées avec votre compte et les ressources que vous avez partagées avec d'autres AWS comptes en utilisant la méthode des ressources nommées.

Important

Lorsqu'une personne utilise la méthode de ressource nommée pour accorder des autorisations entre comptes sur une ressource du catalogue de données à votre compte ou à votre AWS organisation, Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour partager la ressource. Si votre compte appartient à la même AWS organisation que le compte donateur, la ressource partagée est immédiatement disponible. Toutefois, si votre compte n'appartient pas à la même organisation, AWS RAM envoie une invitation à votre compte pour accepter ou refuser le partage des ressources. Ensuite, pour rendre la ressource partagée disponible, l'administrateur du lac de données de votre compte doit utiliser la AWS RAM console ou la CLI pour accepter l'invitation.

La console Lake Formation affiche une alerte si une invitation de partage de AWS RAM ressources attend d'être acceptée. Seuls les utilisateurs autorisés à consulter les AWS RAM invitations reçoivent l'alerte.

Voir aussi :

- [Partage des tables et des bases de données du catalogue de données entre les AWS comptes](#)
- [Partage de données entre comptes dans Lake Formation](#)
- [Accès aux données sous-jacentes d'une table partagée](#)
- [Contrôle d'accès aux métadonnées](#) (pour plus d'informations sur la méthode des ressources nommées par rapport à la méthode LF-TBAC pour le partage des ressources.)

Rubriques

- [Acceptation d'une invitation de partage de ressources de AWS RAM](#)
- [Affichage des tables et des bases de données partagées du catalogue de données](#)

Acceptation d'une invitation de partage de ressources de AWS RAM

Si une ressource du catalogue de données est partagée avec votre AWS compte et que celui-ci n'appartient pas à la même AWS organisation que le compte de partage, vous n'avez pas accès à la ressource partagée tant que vous n'avez pas accepté une invitation de partage de ressources de AWS Resource Access Manager (AWS RAM). En tant qu'administrateur du lac de données, vous devez d'abord AWS RAM demander les invitations en attente, puis accepter l'invitation.

Vous pouvez utiliser la AWS RAM console, l'API ou AWS Command Line Interface (AWS CLI) pour consulter et accepter les invitations.

Pour consulter et accepter une invitation de partage de ressources depuis AWS RAM (console)

1. Assurez-vous de disposer des autorisations Gestion des identités et des accès AWS (IAM) requises pour consulter et accepter les invitations de partage de ressources.

Pour plus d'informations sur les politiques IAM suggérées pour les administrateurs de data lake, consultez [the section called “Autorisations d'administrateur du lac de données”](#).

2. Suivez les instructions de la section [Acceptation et rejet des invitations](#) dans le guide de l'AWS RAM utilisateur.

Pour consulter et accepter une invitation de partage de ressources depuis AWS RAM (AWS CLI)

1. Assurez-vous de disposer des autorisations Gestion des identités et des accès AWS (IAM) requises pour consulter et accepter les invitations de partage de ressources.

Pour plus d'informations sur les politiques IAM suggérées pour les administrateurs de data lake, consultez [the section called “Autorisations d'administrateur du lac de données”](#).

2. Entrez la commande suivante pour afficher les invitations de partage de ressources en attente.

```
aws ram get-resource-share-invitations
```

La sortie doit ressembler à ce qui suit.

```
{
  "resourceShareInvitations": [
    {
```

```

        "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
        "resourceShareName": "111122223333-123456789012-uswuU",
        "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
        "senderAccountId": "111122223333",
        "receiverAccountId": "123456789012",
        "invitationTimestamp": 1589576601.79,
        "status": "PENDING"
    }
]
}

```

Notez le statut de PENDING.

3. Copiez la valeur de la `resourceShareInvitationArn` clé dans le presse-papiers.
4. Collez la valeur dans la commande suivante, en la remplaçant *<invitation-arn>*, puis entrez la commande.

```

aws ram accept-resource-share-invitation --resource-share-invitation-
arn <invitation-arn>

```

La sortie doit ressembler à ce qui suit.

```

{
  "resourceShareInvitations": [
    {
      "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
      "resourceShareName": "111122223333-123456789012-uswuU",
      "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
      "senderAccountId": "111122223333",
      "receiverAccountId": "123456789012",
      "invitationTimestamp": 1589576601.79,
      "status": "ACCEPTED"
    }
  ]
}

```

Notez le statut de `ACCEPTED`.

Affichage des tables et des bases de données partagées du catalogue de données

Vous pouvez consulter les ressources partagées avec votre compte à l'aide de la console Lake Formation ou de la AWS CLI. Vous pouvez également utiliser la console AWS Resource Access Manager (AWS RAM) ou la CLI pour afficher à la fois les ressources partagées avec votre compte et les ressources que vous avez partagées avec d'autres AWS comptes.

Pour afficher les ressources partagées à l'aide de la console Lake Formation

1. Ouvrez la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.

Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur ayant obtenu des autorisations sur une table partagée.

2. Pour consulter les ressources partagées avec votre AWS compte, effectuez l'une des opérations suivantes :
 - Pour afficher les tables partagées avec votre compte, dans le volet de navigation, sélectionnez **Tables**.
 - Pour afficher les bases de données partagées avec votre compte, dans le volet de navigation, sélectionnez **Bases de données**.

La console affiche une liste de bases de données ou de tables à la fois présentes dans votre compte et partagées avec votre compte. Pour les ressources partagées avec votre compte, la console affiche l'identifiant du AWS compte du propriétaire dans la colonne ID du compte du propriétaire (la troisième colonne de la capture d'écran suivante).

Tables (11)

Actions

Create table using a crawler

Create table

Find table by properties

< 1 >

	Name	Database	Owner account ...	Shared resource	Shared resource owner
☐	adviews	analytics	111122223333	-	-
☐	pageviews	analytics	111122223333	-	-
☐	blackholes	hubble	123456789012	-	-
☐	celestial-events	hubble	123456789012	-	-
☐	suns	hubble	123456789012	-	-

3. Pour afficher les ressources que vous avez partagées avec d'autres AWS comptes ou organisations, dans le volet de navigation, sélectionnez Autorisations relatives aux données.

Les ressources que vous avez partagées sont répertoriées sur la page Autorisations relatives aux données avec le numéro de compte externe indiqué dans la colonne Principal, comme indiqué dans l'image suivante.

Data permissions (4)

Revoke

Grant

Find by properties

< 1 >

Database: analytics

Table: clickthroughs

Clear filter

	Principal	Principal type	Resource type	Resource	Owner account ID	Permissions
☐	datalake_admin	IAM user	Table	clickthroughs	123456789012	Super, Alter, Delete, Drop, Insert
☐	datalake_admin	IAM user	Column	analytics.clickthroughs.*	123456789012	Select
☐	111122223333	AWS account	Table	clickthroughs	123456789012	Insert
☐	111122223333	AWS account	Column	analytics.clickthroughs.*	123456789012	Select

Pour afficher les ressources partagées à l'aide de la AWS RAM console

1. Assurez-vous de disposer des autorisations Gestion des identités et des accès AWS (IAM) requises pour consulter les ressources partagées à l'aide AWS RAM de.

Au minimum, vous devez avoir l'autorisation`ram:ListResources`. Cette autorisation est incluse dans la politique gérée par AWS. `AWSLakeFormationCrossAccountManager`

2. Connectez-vous à la AWS RAM console AWS Management Console et ouvrez-la à la <https://console.aws.amazon.com/ram/maison>.
3. Effectuez l'une des actions suivantes :
 - Pour voir les ressources que vous avez partagées, dans le volet de navigation, sous Shared by me, sélectionnez Shared resources.
 - Pour voir les ressources partagées avec vous, dans le volet de navigation, sous Partagé avec moi, choisissez Ressources partagées.

Création de liens vers des ressources

Les liens de ressources sont des objets du catalogue de données qui sont des liens vers des bases de données et des tables de métadonnées, généralement vers des bases de données partagées et des tables d'autres AWS comptes. Ils aident à permettre l'accès entre comptes aux données du lac de données dans toutes les AWS régions.

Note

Lake Formation permet d'interroger les tables du catalogue de données dans toutes les AWS régions. Vous pouvez accéder aux bases de données et aux tables du catalogue de données depuis n'importe quelle AWS région en créant des liens de ressources dans ces régions qui pointent vers des bases de données et des tables partagées dans différentes régions.

Rubriques

- [Mode de fonctionnement des liens des ressources dans Lake Formation](#)
- [Création d'un lien de ressource vers une table de catalogue de données partagée](#)
- [Création d'un lien de ressource vers une base de données de catalogue de données partagée](#)
- [Gestion des liens de ressources dans AWS Glue APIs](#)

Mode de fonctionnement des liens des ressources dans Lake Formation

Un lien de ressource est un objet du catalogue de données qui est un lien vers une base de données ou une table locale ou partagée. Après avoir créé un lien de ressource vers une base de données ou une table, vous pouvez utiliser le nom du lien de ressource partout où vous utiliseriez le nom de la base de données ou de la table. Outre les tables que vous possédez ou les tables partagées avec

vous, les liens vers les ressources des tables sont renvoyés par la page Tables de la console Lake Formation `glue:GetTables()` et apparaissent sous forme d'entrées sur celle-ci. Les liens vers les ressources vers les bases de données agissent de la même manière.

La création d'un lien de ressource vers une base de données ou une table vous permet d'effectuer les opérations suivantes :

- Attribuez un nom différent à une base de données ou à une table dans votre catalogue de données. Cela est particulièrement utile si différents AWS comptes partagent des bases de données ou des tables portant le même nom, ou si plusieurs bases de données de votre compte possèdent des tables portant le même nom.
- Accédez aux bases de données et aux tables du catalogue de données depuis n'importe quelle AWS région en créant des liens de ressources dans ces régions pointant vers la base de données et les tables d'une autre région. Vous pouvez exécuter des requêtes dans n'importe quelle région avec ces liens de ressources à l'aide d'Athena, Amazon EMR et exécuter des tâches AWS Glue ETL Spark, sans copier les données source ni les métadonnées dans Glue Data Catalog.
- Utilisez AWS des services intégrés tels qu' Amazon Athena Amazon Redshift Spectrum pour exécuter des requêtes qui accèdent à des bases de données ou à des tables partagées. Certains services intégrés ne peuvent pas accéder directement aux bases de données ou aux tables d'un compte à l'autre. Ils peuvent toutefois accéder aux liens de ressources de votre compte vers les bases de données et les tables d'autres comptes.

Note

Il n'est pas nécessaire de créer un lien de ressource pour référencer une base de données ou une table partagée dans les scripts AWS Glue d'extraction, de transformation et de chargement (ETL). Toutefois, pour éviter toute ambiguïté lorsque plusieurs AWS comptes partagent une base de données ou une table portant le même nom, vous pouvez soit créer et utiliser un lien de ressource, soit spécifier l'ID du catalogue lorsque vous appelez des opérations ETL.

L'exemple suivant montre la page Tables de la console Lake Formation, qui répertorie deux liens vers des ressources. Les noms des liens vers les ressources sont toujours affichés en italique. Chaque lien de ressource est affiché avec le nom et le propriétaire de la ressource partagée associée. Dans cet exemple, un administrateur de lac de données du AWS compte 1111-2222-3333 a partagé les

tables `inventory` et `incidents` avec le compte 1234-5678-9012. Un utilisateur de ce compte a ensuite créé des liens de ressources vers ces tables partagées.

Tables (30)					
Find table by properties		Actions	Create table using a crawler	Create table	
	Name	Database	Owner account ...	Shared resource	Shared resource owner
☐	inventory-link	retail	123456789012	inventory	111122223333
☐	incidents-link	issues-local	123456789012	incidents	111122223333
☐	site-logs	logs	123456789012	-	-
☐	alexa-logs	logs	123456789012	-	-

Les remarques et restrictions relatives aux liens vers les ressources sont les suivantes :

- Des liens de ressources sont nécessaires pour permettre aux services intégrés tels qu'Athena et Redshift Spectrum d'interroger les données sous-jacentes des tables partagées. Les requêtes de ces services intégrés sont construites en fonction des noms des liens vers les ressources.
- En supposant que le paramètre Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données soit désactivé pour la base de données qui la contient, seul le principal qui a créé un lien de ressource peut le consulter et y accéder. Pour permettre aux autres utilisateurs de votre compte d'accéder à un lien vers une ressource, accordez l'`DESCRIBE` autorisation correspondante. Pour permettre à d'autres personnes de supprimer un lien vers une ressource, `DROP` accordez-lui l'autorisation. Les administrateurs du data lake peuvent accéder à tous les liens vers les ressources du compte. Pour supprimer un lien de ressource créé par un autre principal, l'administrateur du lac de données doit d'abord s'accorder l'`DROP` autorisation sur le lien de ressource. Pour de plus amples informations, veuillez consulter [Référence des autorisations de Lake Formation](#).

Important

L'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations à la cible séparément.

- Pour créer un lien vers une ressource, vous avez besoin de la Lake Formation `CREATE_TABLE` ou de l'`CREATE_DATABASE` autorisation `glue:CreateTable` ou `glue:CreateDatabase` Gestion des identités et des accès AWS (IAM).

- Vous pouvez créer des liens vers des ressources locales (détenues) du catalogue de données, ainsi que vers des ressources partagées avec votre AWS compte.
- Lorsque vous créez un lien vers une ressource, aucune vérification n'est effectuée pour vérifier si la ressource partagée cible existe ou si vous disposez d'autorisations entre comptes sur la ressource. Cela vous permet de créer le lien vers la ressource et la ressource partagée dans n'importe quel ordre.
- Si vous supprimez un lien vers une ressource, la ressource partagée associée n'est pas supprimée. Si vous supprimez une ressource partagée, les liens vers cette ressource ne sont pas supprimés.
- Il est possible de créer des chaînes de liens entre les ressources. Cependant, cela ne sert à rien, car ils ne APIs suivent que le premier lien de ressource.

 Voir aussi :

- [Octroi d'autorisations sur les ressources du catalogue de données](#)

Création d'un lien de ressource vers une table de catalogue de données partagée

Vous pouvez créer un lien de ressource vers une table partagée dans n'importe quelle AWS région à l'aide de la AWS Lake Formation console, de l'API ou AWS Command Line Interface (AWS CLI).

Pour créer un lien de ressource vers une table partagée (console)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant que directeur CREATE_TABLE autorisé à accéder à la base de données pour contenir le lien vers la ressource dans la base de données.
2. Dans le volet de navigation, choisissez Tables sous Catalogue de données, puis choisissez Create, Resource link.
3. Sur la page Créer un lien vers une ressource, fournissez les informations suivantes :

Nom du lien vers la ressource

Entrez un nom qui respecte les mêmes règles qu'un nom de table. Le nom peut être identique à celui de la table partagée cible.

Base de données

La base de données du catalogue de données local qui doit contenir le lien vers la ressource.

Région du propriétaire de la table partagée

Si vous créez le lien de ressource dans une autre région, sélectionnez la région de la table partagée cible.

Table partagée

Sélectionnez une table partagée dans la liste ou entrez un nom de table locale (détenue) ou partagée.

La liste contient toutes les tables partagées avec votre compte. Notez la base de données et l'ID du compte propriétaire répertoriés dans chaque table. Si vous ne voyez aucune table dont vous savez qu'elle a été partagée avec votre compte, vérifiez les points suivants :

- Si vous n'êtes pas un administrateur de lac de données, vérifiez que l'administrateur du lac de données vous a accordé les autorisations de Lake Formation sur la table.
- Si vous êtes administrateur d'un lac de données et que votre compte n'appartient pas à la même AWS organisation que le compte octroyant, assurez-vous d'avoir accepté l'invitation de partage de ressources AWS Resource Access Manager (AWS RAM) pour la table. Pour de plus amples informations, veuillez consulter [Acceptation d'une invitation de partage de ressources de AWS RAM](#).

Base de données de la table partagée

Si vous avez sélectionné une table partagée dans la liste, ce champ est rempli avec la base de données de la table partagée dans le compte externe. Sinon, entrez une base de données locale (pour un lien de ressource vers une table locale) ou la base de données de la table partagée dans le compte externe.

Propriétaire de la table partagée

Si vous avez sélectionné une table partagée dans la liste, ce champ est renseigné avec l'ID de compte du propriétaire de la table partagée. Sinon, entrez votre identifiant de AWS compte (pour un lien de ressource vers une table locale) ou l'identifiant du AWS compte qui a partagé la table.

4. Choisissez Créer pour créer le lien vers la ressource.

Vous pouvez ensuite afficher le nom du lien vers la ressource dans la colonne Nom de la page Tables.

5. (Facultatif) Accordez l'DESCRIBE autorisation de Lake Formation sur le lien de la ressource aux principaux qui doivent être en mesure de voir le lien et d'accéder à la table cible.

Toutefois, l'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations sur la base de données cible séparément pour que le table/ressource lien soit visible dans Athena.

Pour créer un lien de ressource vers une table partagée dans la même région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-table --database-name myissues --table-input
'{"Name": "my_customers", "TargetTable":
{"CatalogId": "111122223333", "DatabaseName": "issues", "Name": "customers"}}'
```

Cette commande crée un lien de ressource nommé `my_customers` vers la table partagée `customers`, qui se trouve dans la base de données `issues` sous le AWS compte 1111-2222-3333. Le lien vers la ressource est stocké dans la base de données locale `myissues`.

2. (Facultatif) Accordez l'DESCRIBE autorisation de Lake Formation sur le lien de la ressource aux principaux qui doivent être en mesure de voir le lien et d'accéder à la table cible.

Toutefois, l'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la table cible (liée). Vous devez accorder des autorisations sur la base de données cible séparément pour que le table/ressource lien soit visible dans Athena.

Pour créer un lien de ressource vers une table partagée dans une autre région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-table --region eu-west-1 --cli-input-json '{
  "CatalogId": "111122223333",
  "DatabaseName": "ireland_db",
  "TableInput": {
    "Name": "rl_useast1salestb_ireland",
    "TargetTable": {
      "CatalogId": "444455556666",
```

```
        "DatabaseName": "useast1_salesdb",
        "Region": "us-east-1",
        "Name": "useast1_salestb"
    }
}
```

Cette commande crée un lien de ressource nommé `rl_useast1salestb_ireland` dans la région Europe (Irlande) vers la `useast1_salestb` table partagée, qui se trouve dans la `useast1_salesdb` base de données du compte 444455556666 dans AWS la région USA Est (Virginie du Nord). Le lien vers la ressource est stocké dans la base de données `localeireland_db`.

2. Accordez l'`DESCRIBE` autorisation de la Lake Formation aux principaux qui doivent être en mesure de voir le lien et d'accéder à la cible du lien via le lien.

Toutefois, l'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la table cible (liée). Vous devez accorder des autorisations sur la table cible séparément pour que le table/ressource lien soit visible dans Athena.

 Voir aussi :

- [Mode de fonctionnement des liens des ressources dans Lake Formation](#)
- [DESCRIBE](#)

Création d'un lien de ressource vers une base de données de catalogue de données partagée

Vous pouvez créer un lien de ressource vers une base de données partagée à l'aide de la AWS Lake Formation console, de l'API ou AWS Command Line Interface (AWS CLI).

Pour créer un lien de ressource vers une base de données partagée (console)

1. Ouvrez la AWS Lake Formation console à l'adresse <https://console.aws.amazon.com/lakeformation/>. Connectez-vous en tant qu'administrateur du lac de données ou en tant que créateur de base de données.

Un créateur de base de données est un directeur qui a obtenu l'`CREATE_DATABASE` autorisation de Lake Formation.

2. Dans le volet de navigation, choisissez Databases, puis Create, Resource link.
3. Sur la page Créer un lien vers une ressource, fournissez les informations suivantes :

Nom du lien vers la ressource

Entrez un nom qui respecte les mêmes règles qu'un nom de base de données. Le nom peut être identique à celui de la base de données partagée cible.

Catalogue de destinations

Sélectionnez le catalogue de destination pour le lien de ressource de base de données.

Région du propriétaire de la base de données partagée

Si vous créez le lien de ressource dans une autre région, sélectionnez la région de la base de données partagée cible.

Base de données partagée

Choisissez une base de données dans la liste ou entrez un nom de base de données locale (détenue) ou partagée.

La liste contient toutes les bases de données partagées avec votre compte. Notez l'ID du compte propriétaire répertorié dans chaque base de données. Si vous ne trouvez aucune base de données dont vous savez qu'elle a été partagée avec votre compte, vérifiez les points suivants :

- Si vous n'êtes pas un administrateur de lac de données, vérifiez que l'administrateur du lac de données vous a accordé les autorisations de Lake Formation sur la base de données.
- Si vous êtes administrateur d'un lac de données et que votre compte n'appartient pas à la même AWS organisation que le compte octroyant, assurez-vous d'avoir accepté l'invitation de partage de ressources AWS Resource Access Manager (AWS RAM) pour la base de données. Pour de plus amples informations, veuillez consulter [Acceptation d'une invitation de partage de ressources de AWS RAM](#).

Propriétaire de la base de données partagée

Si vous avez sélectionné une base de données partagée dans la liste, ce champ est renseigné avec l'ID de compte du propriétaire de la base de données partagée. Sinon, entrez

votre identifiant de AWS compte (pour un lien de ressource vers une base de données locale) ou l'identifiant du AWS compte qui a partagé la base de données.

ID de catalogue de la base de données partagée

Entrez l'ID du catalogue pour la base de données partagée. Lorsque vous créez un lien de ressource vers une base de données partagée depuis un autre AWS compte, vous devez spécifier cet ID de catalogue pour identifier le catalogue de données du compte qui contient la base de données source.

Lorsque vous sélectionnez une base de données partagée dans le menu déroulant, le système saisit automatiquement l'ID de catalogue du compte qui possède et a partagé cette base de données avec vous.

Create resource link [Info](#)

Database resource link details

Create a database resource link in the Data Catalog.

Resource link name

If you plan to access the database from Amazon Athena, then the name should be under 256 characters and contain only lowercase letters (a-z), numbers (0-9), and underscore (_). For more information, see [Athena names](#).

Destination catalog

Select the destination catalog for the database resource link.

Shared database's region

Select the region of the shared database.

Shared database

Enter or choose a shared database.

Shared database's catalog ID

Enter the catalog ID of the shared database.

[Cancel](#)[Create](#)

4. Choisissez Créer pour créer le lien vers la ressource.

Vous pouvez ensuite afficher le nom du lien vers la ressource dans la colonne Nom de la page Bases de données.

5. (Facultatif) Accordez l'DESCRIBE autorisation de la Lake Formation sur le lien de la ressource aux responsables de la région Europe (Irlande) qui doivent être en mesure de consulter le lien et d'accéder à la base de données cible.

Toutefois, l'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations sur la base de données cible séparément pour que le table/ressource lien soit visible dans Athena.

Pour créer un lien de ressource vers une base de données partagée dans la même région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

Cette commande crée un lien de ressource nommé myissues vers la base de données partagée issues, qui se trouve dans le AWS compte 1111-2222-3333.

2. (Facultatif) Accordez l'DESCRIBE autorisation de la Lake Formation aux principaux sur le lien de ressource qui doivent être en mesure de visualiser le lien et d'accéder à la base de données ou à la table cible.

Toutefois, l'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations sur la base de données cible séparément pour que le table/ressource lien soit visible dans Athena.

Pour créer un lien de ressource vers une base de données partagée dans une autre région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-database --region eu-west-1 --cli-input-json '{
  "CatalogId": "111122223333",
  "DatabaseInput": {
    "Name": "r1_useast1shared_irelanddb",
```



```

    "TargetDatabase": {
      "CatalogId": "444455556666",
      "DatabaseName": "useast1shared_db",
      "Region": "us-east-1"
    }
  }
}'

```

Cette commande crée un lien de ressource nommé `rl_useeast1shared_irelanddb` dans le AWS compte 11112223333 dans la région Europe (Irlande) vers la base de données partagée `useast1shared_db`, qui se trouve dans le AWS compte 444455556666 dans la région USA Est (Virginie du Nord).

2. Accordez l'`DESCRIBE` autorisation de la Lake Formation aux responsables de la région Europe (Irlande) qui doivent être en mesure de voir le lien et d'accéder à la cible du lien via le lien.

 Voir aussi :

- [Mode de fonctionnement des liens des ressources dans Lake Formation](#)
- [DESCRIBE](#)

Gestion des liens de ressources dans AWS Glue APIs

Les tableaux suivants expliquent comment le catalogue de AWS Glue données APIs gère les liens entre les ressources de base de données et de tables. Pour toutes les opérations `Get*` d'API, seules les bases de données et les tables sur lesquelles l'appelant est autorisé sont renvoyées. En outre, lorsque vous accédez à une base de données ou à une table cible via un lien de ressource, vous devez disposer à la fois des autorisations Gestion des identités et des accès AWS (IAM) et de Lake Formation sur la cible et sur le lien de ressource. L'autorisation Lake Formation requise pour les liens de ressources est `DESCRIBE`. Pour de plus amples informations, veuillez consulter [DESCRIBE](#).

Opérations de l'API de base

Opération API	Gestion des liens vers les ressources
<code>CreateDatabase</code>	Si la base de données est un lien de ressource, crée le lien de ressource vers la base de données cible désignée.

Opération API	Gestion des liens vers les ressources
UpdateDatabase	Si la base de données désignée est un lien de ressource, suivez le lien et mettez à jour la base de données cible. Si le lien de ressource doit être modifié pour être lié à une autre base de données, vous devez le supprimer et en créer une nouvelle.
DeleteDatabase	Supprime le lien vers la ressource. Il ne supprime pas la base de données liée (cible).
GetDatabase	Si l'appelant dispose d'autorisations sur la cible, suivez le lien pour renvoyer les propriétés de la cible. Dans le cas contraire, elle renvoie les propriétés du lien.
GetDatabases	Renvoie une liste de bases de données, y compris des liens vers des ressources. Pour chaque lien de ressource du jeu de résultats, l'opération suit le lien pour obtenir les propriétés de la cible du lien. Vous devez spécifier <code>ResourceShareType = ALL</code> pour voir les bases de données partagées avec votre compte.

Opérations de l'API de table

Opération API	Gestion des liens vers les ressources
CreateTable	Si la base de données est un lien de ressource, suit le lien de base de données et crée une table dans la base de données cible. Si la table est un lien de ressource, l'opération crée le lien de ressource dans la base de données désignée. La création d'un lien de ressource de table via un lien de ressource de base de données n'est pas prise en charge.
UpdateTable	Si la table ou la base de données désignée est un lien de ressource, met à jour la table cible. Si la table et la base de données sont des liens de ressources, l'opération échoue.
DeleteTable	Si la base de données désignée est un lien de ressource, suit le lien et supprime le lien de ressource de table ou de table dans la base de données cible. Si la table est un lien de ressource, l'opération

Opération API	Gestion des liens vers les ressources
	supprime le lien de ressource de table dans la base de données désignée. La suppression d'un lien vers une ressource de table ne supprime pas la table cible.
BatchDeleteTable	Identique à DeleteTable .
GetTable	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie le lien de ressource de table ou de table depuis la base de données cible. Sinon, si la table est un lien vers une ressource, l'opération suit le lien et renvoie les propriétés de la table cible.
GetTables	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie les tables et les liens de ressources de table depuis la base de données cible. Si la base de données cible est une base de données partagée provenant d'un autre AWS compte, l'opération renvoie uniquement les tables partagées de cette base de données. Il ne suit pas les liens vers les ressources de la table dans la base de données cible. Sinon, si la base de données désignée est une base de données locale (détenue), l'opération renvoie toutes les tables de la base de données locale et suit chaque lien de ressource de table pour renvoyer les propriétés de la table cible.
SearchTables	Renvoie les tables et les liens vers les ressources des tables. Il ne suit pas les liens pour renvoyer les propriétés de la table cible. Vous devez spécifier ResourceShareType = ALL pour voir les tables partagées avec votre compte.
GetTableVersion	Identique à GetTable.
GetTableVersions	Identique à GetTable.
DeleteTableVersion	Identique à DeleteTable .
BatchDeleteTableVersion	Identique à DeleteTable .

Opérations de l'API de partition

Opération API	Gestion des liens vers les ressources
CreatePartition	Si la base de données désignée est un lien de ressource, suit le lien de base de données et crée une partition dans la table désignée de la base de données cible. Si la table est un lien de ressource, l'opération suit le lien de ressource et crée la partition dans la table cible. La création d'une partition via un lien de ressource de table et un lien de ressource de base de données n'est pas prise en charge.
BatchCreatePartitions	Identique à CreatePartition .
UpdatePartition	Si la base de données désignée est un lien de ressource, suit le lien de base de données et met à jour la partition dans la table désignée dans la base de données cible. Si la table est un lien de ressource, l'opération suit le lien de ressource et met à jour la partition dans la table cible. La mise à jour d'une partition via un lien de ressource de table et un lien de ressource de base de données n'est pas prise en charge.
DeletePartition	Si la base de données désignée est un lien de ressource, suit le lien de base de données et supprime la partition dans la table désignée dans la base de données cible. Si la table est un lien de ressource , l'opération suit le lien de ressource et supprime la partition dans la table cible. La suppression d'une partition via un lien de ressource de table et un lien de ressource de base de données n'est pas prise en charge.
BatchDeletePartitions	Identique à DeletePartition .
GetPartition	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie les informations de partition à partir de la table désignée. Sinon, si la table est un lien de ressource , l'opération suit le lien et renvoie les informations de partition.

Opération API	Gestion des liens vers les ressources
	Si la table et la base de données sont toutes deux des liens de ressources, elles renvoient un jeu de résultats vide.
GetPartitions	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie les informations de partition pour toutes les partitions de la table désignée. Sinon, si la table est un lien de ressource, l'opération suit le lien et renvoie les informations de partition. Si la table et la base de données sont toutes deux des liens de ressources, elles renvoient un jeu de résultats vide.
BatchGetPartition	Identique à GetPartition .

Fonctions définies par l'utilisateur et opérations d'API

Opération API	Gestion des liens vers les ressources
(Toutes les opérations d'API)	Si la base de données est un lien de ressource, suit le lien de ressource et exécute l'opération sur la base de données cible.

 Voir aussi :

- [Mode de fonctionnement des liens des ressources dans Lake Formation](#)

Accès aux tables dans toutes les régions

Lake Formation permet d'interroger les tables du catalogue de données dans toutes les AWS régions. Vous pouvez accéder aux données d'une région depuis d'autres régions à l'aide d'Amazon Athena, Amazon EMR et AWS Glue ETL en [créant des liens de ressources](#) dans d'autres régions pointant vers les bases de données et les tables sources. Grâce à l'accès aux tables entre régions, vous pouvez accéder aux données entre les régions sans copier les données sous-jacentes ou les métadonnées dans le catalogue de données.

Par exemple, vous pouvez partager une base de données ou une table d'un compte producteur avec un compte client de la région A. Après avoir accepté l'invitation de partage de ressources dans la

région A, l'administrateur du lac de données du compte client peut créer des liens vers la ressource partagée dans la région B. L'administrateur du compte client peut accorder des autorisations sur la ressource partagée aux principaux IAM de ce compte dans la région A et peut accorder des autorisations de lien de ressource dans la région B. À l'aide du lien de ressource, les principaux du compte client peuvent interroger les données partagées de la région B.

Vous pouvez également héberger la source de données Amazon S3 dans la région A dans un compte producteur et enregistrer l'emplacement des données dans un compte central dans la région B. Vous pouvez créer des ressources de catalogue de données dans le compte central, configurer les autorisations de Lake Formation et partager des données avec les consommateurs de votre compte ou avec des comptes externes dans la région B. La fonctionnalité inter-régions permet aux utilisateurs d'accéder à ces tables du catalogue de données depuis la région C à l'aide de liens de ressources.

Grâce à cette fonctionnalité, vous pouvez interroger des bases de données fédérées dans les métastores Apache Hive d'une région à l'autre, et également joindre des tables de la région locale à des tables d'une autre région lors de l'exécution de requêtes.

Lake Formation prend en charge les fonctionnalités suivantes avec un accès aux tables interrégional :

- Contrôle d'accès basé sur des balises LF
- Autorisations de contrôle d'accès détaillées
- Opérations d'écriture sur la base de données ou la table partagée avec les autorisations appropriées
- Partage de données entre comptes au niveau du compte et directement avec les responsables IAM

Les utilisateurs non administrateurs dotés `Create_Database` d'`Create_Table` autorisations peuvent créer des liens de ressources interrégionaux.

Note

Vous pouvez créer des liens de ressources entre régions dans n'importe quelle région et accéder aux données sans appliquer les autorisations de Lake Formation. Pour les données source dans Amazon S3 qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions.

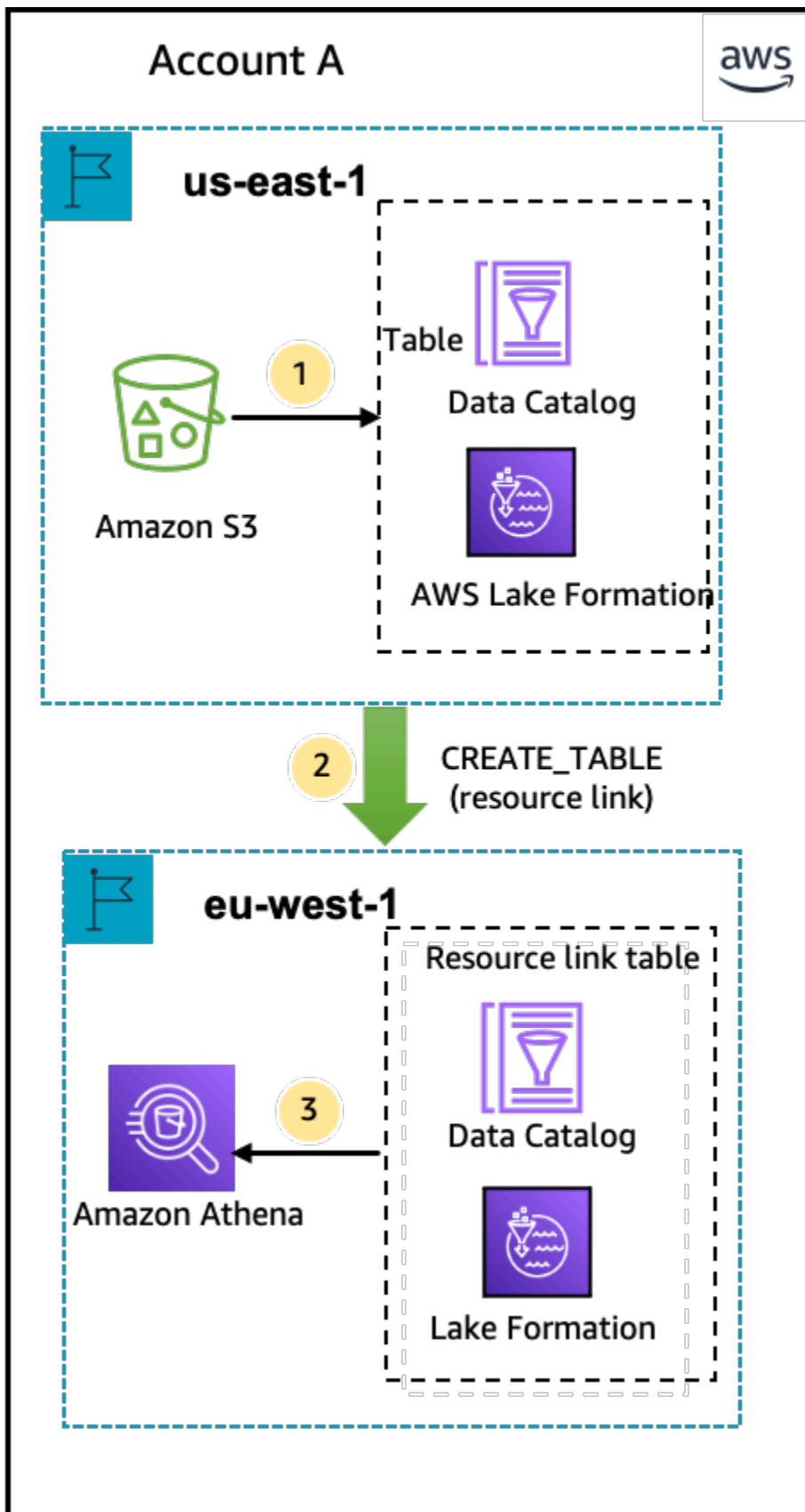
Pour connaître les limitations, veuillez consulter [Limites d'accès aux données entre régions](#).

Flux de travail

Les diagrammes suivants montrent les flux de travail permettant d'accéder aux données entre AWS les régions à partir du même AWS compte et d'un compte externe.

Flux de travail pour accéder aux tables partagées au sein d'un même AWS compte

Dans le schéma ci-dessous, les données sont partagées avec un utilisateur du même AWS compte dans la région USA Est (Virginie du Nord), et l'utilisateur interroge les données partagées depuis la région Europe (Irlande).



L'administrateur du lac de données effectue les activités suivantes (étapes 1 et 2) :

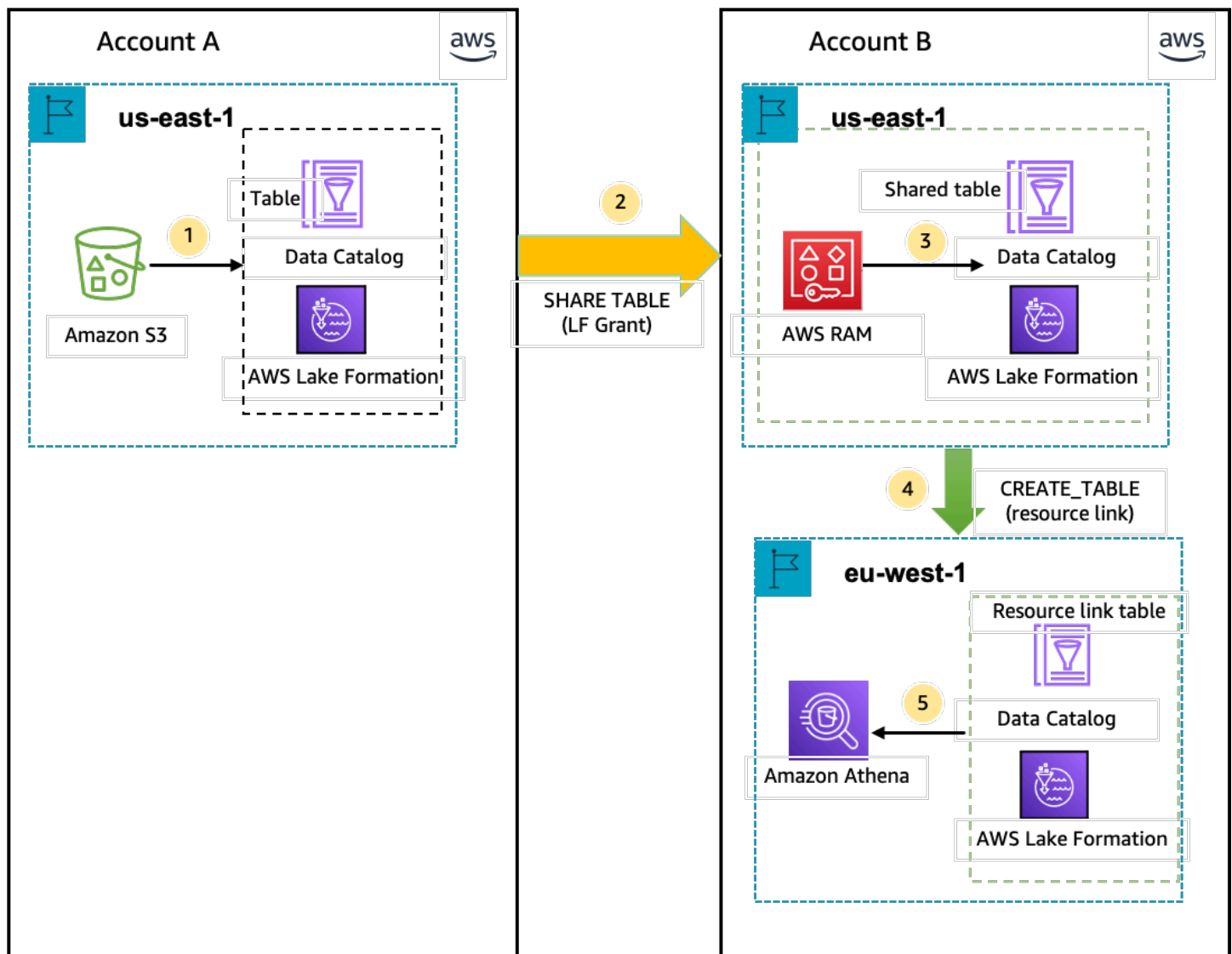
1. Un administrateur de lac de données crée un AWS compte auprès des bases de données et des tables du catalogue de données et enregistre un emplacement de données Amazon S3 auprès de Lake Formation dans la région de l'est des États-Unis (Virginie du Nord).

Accorde une `Select` autorisation sur une ressource du catalogue de données (tableau des produits dans le diagramme) à un principal (utilisateur) du même compte.

2. Crée un lien de ressource dans la région Europe (Irlande) pointant vers la table source dans la région USA Est (Virginie du Nord). Accorde l'`DESCRIBE` autorisation sur le lien de ressource de la région Europe (Irlande) vers le principal.
3. L'utilisateur interroge le tableau depuis la région Europe (Irlande) à l'aide d'Athena.

Flux de travail pour accéder aux tables partagées avec un AWS compte externe

Dans le schéma ci-dessous, le compte producteur (compte A) héberge le compartiment Amazon S3, enregistre l'emplacement des données et partage une table du catalogue de données avec un compte consommateur (compte B) dans la région USA Est (Virginie du Nord) et un utilisateur du compte consommateur (compte B) interroge le tableau depuis la région Europe (Irlande).



1. Un administrateur de lac de données crée un AWS compte (compte producteur) avec les ressources du catalogue de données et un emplacement de données Amazon S3 enregistré auprès de Lake Formation dans la région de l'est des États-Unis (Virginie du Nord).
2. L'administrateur du lac de données du compte producteur partage une table du catalogue de données avec un compte client.
3. L'administrateur du lac de données du compte client accepte l'invitation à partager des données dans la région USA Est (Virginie du Nord) et accorde l'`Select` autorisation d'utiliser la table partagée à un mandant de la même région.
4. L'administrateur du lac de données du compte client crée un lien de ressource dans la région Europe (Irlande) pointant vers la table partagée cible dans la région USA Est (Virginie du Nord) et

accorde à l'DESCRIBEutilisateur l'autorisation d'utiliser le lien de ressource depuis la région Europe (Irlande).

5. L'utilisateur interroge les données de la région Europe (Irlande) à l'aide d'Athena.

Configuration de l'accès aux tables entre régions

Pour accéder aux données d'une autre région, vous devez d'abord configurer les bases de données et les tables du catalogue de données dans la région dans laquelle vous enregistrez votre emplacement de données Amazon S3. Vous pouvez partager les bases de données et les tables du catalogue de données avec les principaux de votre compte ou d'un autre compte. Vous devez ensuite créer des administrateurs de lacs de données qui peuvent créer des liens de ressources pointant vers l'emplacement des données partagées cible dans les régions où les utilisateurs interrogent les données.

Pour interroger des données partagées au sein d'un même compte depuis une autre région

Dans cette section, la table partagée cible Region est appelée Région A et les utilisateurs exécutent des requêtes depuis la Région B.

1. Configuration du compte dans la région A (où vous créez et partagez les données)

Un administrateur de lac de données doit effectuer les actions suivantes :

- a. Enregistrez un emplacement de données Amazon S3.

Pour de plus amples informations, veuillez consulter [Ajouter un emplacement Amazon S3 à votre lac de données](#).

- b. Créez des bases de données et des tables dans le compte. Cela peut également être fait par un utilisateur non administratif autorisé à créer des bases de données et des tables.
- c. Accordez des autorisations de données sur une table aux principaux avec `GrantTablePermissions`.

Pour plus d'informations, voir, [Octroi d'autorisations sur les ressources du catalogue de données](#).

2. Configuration du compte dans la région B (où vous accédez aux données)

Un administrateur de lac de données doit effectuer les actions suivantes :

- a. Créez un lien de ressource dans la région B pointant vers la table partagée cible dans la région A. Spécifiez la région propriétaire de la table partagée sur l'écran Créer une table.

Create table

Table details

Create a table in the AWS Glue Data Catalog.

☐ **Table**
Create a table in my account.

☒ **Resource link**
Create a resource link to a shared table.

Resource link name

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (_), and must be less than 256 characters long.

Database
Resource link will be contained in this database.

Shared table owner region
Select the region where the table is shared

Shared table
Enter or choose a shared table.

Shared table's database
Enter the database containing the shared table.

Shared table's owner ID
Enter the AWS account ID of the shared table owner.

Cancel **Create**

Pour obtenir des instructions sur la création de liens de ressources vers des bases de données et des tables, consultez [Création de liens vers des ressources](#).

- b. Accordez Describe l'autorisation aux responsables de l'IAM sur le lien de ressource dans la région B.

Pour plus d'informations sur l'octroi d'autorisations sur les liens vers des ressources, consultez [Octroi d'autorisations relatives aux liens vers](#).

Les responsables IAM de la région B peuvent interroger la table cible via le lien en utilisant Athena.

Pour accéder aux données entre comptes d'une autre région

1. Configuration du compte producteur/concédant

Un administrateur de lac de données doit effectuer les actions suivantes :

- a. Configurez le producer/grantor compte dans la région A.
- b. Enregistrez un emplacement de données Amazon S3 dans la région A.
- c. Créez des bases de données et des tables. Cela peut être fait par un utilisateur non administrateur autorisé à créer des tables.
- d. Accordez des autorisations de données au consumer/grantee compte sur une table de la région A avec `GrantTable permissions`.

Pour de plus amples informations, veuillez consulter [Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des entités IAM](#).

2. Configuration du compte consommateur/bénéficiaire

Un administrateur de lac de données doit effectuer les actions suivantes :

- a. Acceptez l'invitation de partage de ressources depuis AWS RAM la région A.
- b. Créez un lien de ressource dans la région B pointant vers la table partagée. La région B est l'endroit où les utilisateurs voudront interroger la table.
- c. Accordez des autorisations de données sur la table partagée aux principaux IAM de la région A.

Note

Vous devez accorder des autorisations à la table partagée dans la même région où la table a été partagée.

- d. Accordez des autorisations aux directeurs sur le lien de ressource dans la région B.

Les responsables du compte client de la région B interrogent ensuite la table partagée depuis la région B à l'aide d'Athena.

Sécurité dans le AWS Lake Formation

La sécurité du cloud AWS est la priorité absolue. En tant que AWS client, vous bénéficiez d'un centre de données et d'une architecture réseau conçus pour répondre aux exigences des entreprises les plus sensibles en matière de sécurité.

La sécurité est une responsabilité partagée entre vous AWS et vous. Le [modèle de responsabilité partagée](#) décrit cette notion par les termes sécurité du cloud et sécurité dans le cloud :

- **Sécurité du cloud** : AWS est chargée de protéger l'infrastructure qui exécute les AWS services dans le AWS cloud. AWS vous fournit également des services que vous pouvez utiliser en toute sécurité. Des auditeurs tiers testent et vérifient régulièrement l'efficacité de notre sécurité dans le cadre des [programmes de conformité AWS](#). Pour en savoir plus sur les programmes de conformité qui s'appliquent à AWS Lake Formation, consultez la section [AWS Services concernés par programme de conformité](#).
- **Sécurité dans le cloud** — Votre responsabilité est déterminée par le AWS service que vous utilisez. Vous êtes également responsable d'autres facteurs, y compris de la sensibilité de vos données, des exigences de votre entreprise, ainsi que de la législation et de la réglementation applicables.

Cette documentation vous aide à comprendre comment appliquer le modèle de responsabilité partagée lors de l'utilisation de Lake Formation. Les rubriques suivantes expliquent comment configurer Lake Formation pour répondre à vos objectifs de sécurité et de conformité. Vous apprendrez également à utiliser d'autres AWS services qui vous aident à surveiller et à sécuriser les ressources de vos Lake Formation.

Rubriques

- [Protection des données dans le Lake Formation](#)
- [Sécurité des infrastructures dans AWS le Lake Formation](#)
- [Prévention du problème de l'adjoint confus entre services](#)
- [Enregistrement des événements de sécurité dans AWS Lake Formation](#)

Protection des données dans le Lake Formation

Le [modèle de responsabilité AWS partagée](#) de s'applique à la protection des données dans. Comme décrit dans ce modèle, AWS est chargé de protéger l'infrastructure mondiale qui gère tous les

AWS Cloud. La gestion du contrôle de votre contenu hébergé sur cette infrastructure relève de votre responsabilité. Vous êtes également responsable des tâches de configuration et de gestion de la sécurité des Services AWS que vous utilisez. Pour plus d'informations sur la confidentialité des données, consultez [Questions fréquentes \(FAQ\) sur la confidentialité des données](#). Pour en savoir plus sur la protection des données en Europe, consultez le billet de blog [Modèle de responsabilité partagée d'AWS et RGPD \(Règlement général sur la protection des données\)](#) sur le Blog de sécuritéAWS .

À des fins de protection des données, nous vous recommandons de protéger les Compte AWS informations d'identification et de configurer les utilisateurs individuels avec AWS IAM Identity Center ou Gestion des identités et des accès AWS (IAM). Ainsi, chaque utilisateur se voit attribuer uniquement les autorisations nécessaires pour exécuter ses tâches. Nous vous recommandons également de sécuriser vos données comme indiqué ci-dessous :

- Utilisez l'authentification multifactorielle (MFA) avec chaque compte.
- SSL/TLS À utiliser pour communiquer avec AWS les ressources. Nous exigeons TLS 1.2 et recommandons TLS 1.3.
- Configurez l'API et la journalisation de l'activité des utilisateurs avec AWS CloudTrail. Pour plus d'informations sur l'utilisation des CloudTrail sentiers pour capturer AWS des activités, consultez la section [Utilisation des CloudTrail sentiers](#) dans le guide de AWS CloudTrail l'utilisateur.
- Utilisez des solutions de AWS chiffrement, ainsi que tous les contrôles de sécurité par défaut qu'ils contiennent Services AWS.
- Utilisez des services de sécurité gérés avancés tels qu'Amazon Macie, qui contribuent à la découverte et à la sécurisation des données sensibles stockées dans Amazon S3.
- Si vous avez besoin de modules cryptographiques validés par la norme FIPS 140-3 pour accéder AWS via une interface de ligne de commande ou une API, utilisez un point de terminaison FIPS. Pour plus d'informations sur les points de terminaison FIPS disponibles, consultez [Norme FIPS \(Federal Information Processing Standard\) 140-3](#).

Nous vous recommandons fortement de ne jamais placer d'informations confidentielles ou sensibles, telles que les adresses e-mail de vos clients, dans des balises ou des champs de texte libre tels que le champ Nom. Cela inclut lorsque vous travaillez avec ou d'autres Services AWS utilisateurs de la console, de l'API ou AWS SDKs. AWS CLI Toutes les données que vous entrez dans des balises ou des champs de texte de forme libre utilisés pour les noms peuvent être utilisées à des fins de facturation ou dans les journaux de diagnostic. Si vous fournissez une adresse URL à un serveur

externe, nous vous recommandons fortement de ne pas inclure d'informations d'identification dans l'adresse URL permettant de valider votre demande adressée à ce serveur.

Chiffrement au repos

AWS Lake Formation prend en charge le chiffrement des données dans les domaines suivants :

- Données de votre lac de données Amazon Simple Storage Service (Amazon S3).

Lake Formation prend en charge le chiffrement des données avec [AWS Key Management Service](#) (AWS KMS). Les données sont généralement écrites dans le lac de données au moyen de tâches d'AWS Glue extraction, de transformation et de chargement (ETL). Pour plus d'informations sur le chiffrement des données écrites par les AWS Glue tâches, consultez la section [Chiffrement des données écrites par les robots d'exploration, les tâches et les points de terminaison de développement](#) dans le guide du développeur AWS Glue.

- Le AWS Glue Data Catalog, où Lake Formation stocke les tables de métadonnées décrivant les données du lac de données.

Pour plus d'informations, consultez la section [Chiffrer votre catalogue de données](#) dans le guide du AWS Glue développeur.

Pour ajouter un emplacement Amazon S3 en tant que stockage dans votre lac de données, vous devez enregistrer l'emplacement auprès de AWS Lake Formation. Vous pouvez ensuite utiliser les autorisations de Lake Formation pour un contrôle d'accès précis aux AWS Glue Data Catalog objets pointant vers cet emplacement, ainsi qu'aux données sous-jacentes de cet emplacement.

Lake Formation prend en charge l'enregistrement d'un emplacement Amazon S3 contenant des données chiffrées. Pour de plus amples informations, veuillez consulter [Enregistrement d'un emplacement Amazon S3 chiffré](#).

Sécurité des infrastructures dans AWS le Lake Formation

En tant que service géré, AWS Lake Formation est protégé par les procédures de sécurité du réseau AWS mondial décrites dans le livre blanc [Amazon Web Services : présentation des processus de sécurité](#).

Vous utilisez des appels d'API AWS publiés pour accéder à Lake Formation via le réseau. Les clients doivent supporter le protocole TLS (Sécurité de la couche transport) 1.0 ou une version ultérieure. Nous recommandons TLS 1.2 ou version ultérieure. Les clients doivent aussi prendre en charge

les suites de chiffrement PFS (Perfect Forward Secrecy) comme Ephemeral Diffie-Hellman (DHE) ou Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La plupart des systèmes modernes tels que Java 7 et les versions ultérieures prennent en charge ces modes.

En outre, les demandes doivent être signées à l'aide d'un ID de clé d'accès et d'une clé d'accès secrète associée à un principal IAM. Vous pouvez également utiliser [AWS Security Token Service](#) (AWS STS) pour générer des informations d'identification de sécurité temporaires et signer les demandes.

Prévention du problème de l'adjoint confus entre services

Le problème de délégué confus est un problème de sécurité dans lequel une entité qui n'est pas autorisée à effectuer une action peut contraindre une entité plus privilégiée à le faire. En AWS, l'usurpation d'identité interservices peut entraîner la confusion des adjoints. L'usurpation d'identité entre services peut se produire lorsqu'un service (le service appelant) appelle un autre service (le service appelé). Le service appelant peut être manipulé et ses autorisations utilisées pour agir sur les ressources d'un autre client auxquelles on ne serait pas autorisé à accéder autrement. Pour éviter cela, AWS fournit des outils qui vous aident à protéger vos données pour tous les services auprès des principaux fournisseurs de services qui ont obtenu l'accès aux ressources de votre compte.

Nous vous recommandons d'utiliser les clés de contexte de condition globale [aws:SourceArn](#) et [aws:SourceAccount](#) dans les politiques de ressources afin de limiter les autorisations à la ressource octroyées par AWS Lake Formation à un autre service. Si vous utilisez les deux clés de contexte de condition globale, la valeur `aws:SourceAccount` et le compte de la valeur `aws:SourceArn` doit utiliser le même ID de compte lorsqu'il est utilisé dans la même déclaration de stratégie.

Actuellement, Lake Formation ne prend `aws:SourceArn` en charge que les formats suivants :

```
arn:aws:lakeformation:aws-region:account-id:*
```

L'exemple suivant montre comment utiliser les clés contextuelles `aws:SourceArn` et les clés de contexte de condition `aws:SourceAccount` globale dans Lake Formation pour éviter le problème de confusion des adjoints.

JSON

```
{
```

```
"Version": "2012-10-17",
"Statement": [
  {
    "Sid": "ConfusedDeputyPreventionExamplePolicy",
    "Effect": "Allow",
    "Principal": {
      "Service": "lakeformation.amazonaws.com"
    },
    "Action": [
      "sts:AssumeRole"
    ],
    "Condition": {
      "StringEquals": {
        "aws:SourceAccount": "123456789012"
      },
      "ArnEquals": {
        "aws:SourceArn": "arn:aws:lakeformation:aws-region:123456789012:*"
      }
    }
  }
]
```

Enregistrement des événements de sécurité dans AWS Lake Formation

AWS Lake Formation est intégré à AWS CloudTrail un service qui fournit un enregistrement des actions entreprises par un utilisateur, un rôle ou un AWS service dans Lake Formation. CloudTrail capture tous les appels d'API pour Lake Formation sous forme d'événements. Les appels capturés incluent des appels provenant de la console Lake Formation AWS Command Line Interface, des appels et des appels de code vers les opérations de l'API Lake Formation.

Pour plus d'informations sur la journalisation des événements dans Lake Formation, consultez [Enregistrement des appels d'API AWS Lake Formation à l'aide de AWS CloudTrail](#).

Note

GetTableObjectsUpdateTableObjects, et GetWorkUnitResults sont des opérations de plan de données à volume élevé. Les appels qui leur APIs sont adressés ne sont actuellement pas enregistrés CloudTrail. Pour plus d'informations sur les opérations du plan

de données dans CloudTrail, consultez la section [Enregistrement des événements liés aux données pour les sentiers](#) dans le guide de AWS CloudTrail l'utilisateur.

Les modifications de la Lake Formation destinées à soutenir d'autres CloudTrail événements seront documentées à l'adresse [Historique du document pour AWS Lake Formation](#).

Intégration de services tiers avec Lake Formation

L'intégration AWS Lake Formation permet aux services tiers d'accéder en toute sécurité aux données de leurs lacs de données basés sur Amazon S3. Vous pouvez utiliser Lake Formation comme moteur d'autorisation pour gérer ou appliquer les autorisations d'accès à votre lac de données grâce à AWS des services intégrés tels que AWS Glue ETL, Amazon Athena, Amazon EMR et Redshift Spectrum. Lake Formation propose deux options pour intégrer les services :

1. Paramètres d'intégration de l'application Lake Formation : Lake Formation peut vendre des informations d'identification temporaires limitées sous forme de jetons AWS STS aux sites Amazon S3 enregistrés en fonction des autorisations effectives, afin que les applications autorisées puissent accéder aux données pour le compte des utilisateurs.
2. Application centralisée : les opérations d'[API d'interrogation](#) de Lake Formation récupèrent les données d'Amazon S3 et filtrent les résultats en fonction des autorisations effectives. Le moteur ou l'application qui s'intègre au fonctionnement de l'API de requête peut dépendre de Lake Formation pour évaluer les autorisations de l'identité appelante et filtrer les données en toute sécurité en fonction de ces autorisations. Les moteurs de requêtes tiers ne voient et ne fonctionnent que sur les données filtrées.

La vente d'informations d'identification de Lake Formation ne s'intègre pas aux requêtes SQL de Spark. La vente d'informations d'identification ne fonctionne qu'avec les requêtes qui passent par la bibliothèque AWS Glue ETL.

Rubriques

- [Utilisation de l'intégration de l'application Lake Formation](#)

Utilisation de l'intégration de l'application Lake Formation

Lake Formation permet à des services tiers de s'intégrer à Lake Formation et d'obtenir un accès temporaire aux données Amazon S3 pour le compte de leurs utilisateurs en utilisant [GetTemporaryGlueTableCredentials](#) et en effectuant [GetTemporaryGluePartitionCredentials](#) des opérations. Cela permet aux services tiers d'utiliser la même fonctionnalité d'autorisation et de vente d'informations d'identification que les autres services d'AWS analyse. Cette section décrit comment utiliser ces opérations d'API pour intégrer un moteur de requête tiers à Lake Formation.

Ces opérations d'API sont désactivées par défaut. Il existe deux options pour autoriser Lake Formation à intégrer des applications :

- Configurer les balises de session IAM qui sont validées chaque fois que les opérations de l'API d'intégration de l'application sont appelées

Pour de plus amples informations, veuillez consulter [Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications](#).

- Activez l'option qui permet aux moteurs externes d'accéder aux données des emplacements Amazon S3 avec un accès complet aux tables

Cette option permet aux moteurs de requête et aux applications d'obtenir des informations d'identification sans balises de session IAM si l'utilisateur dispose d'un accès complet à la table. Il fournit des avantages en termes de performances aux moteurs de requêtes et aux applications, tout en simplifiant l'accès aux données. Amazon EMR sur Amazon EC2 est en mesure de tirer parti de ce paramètre.

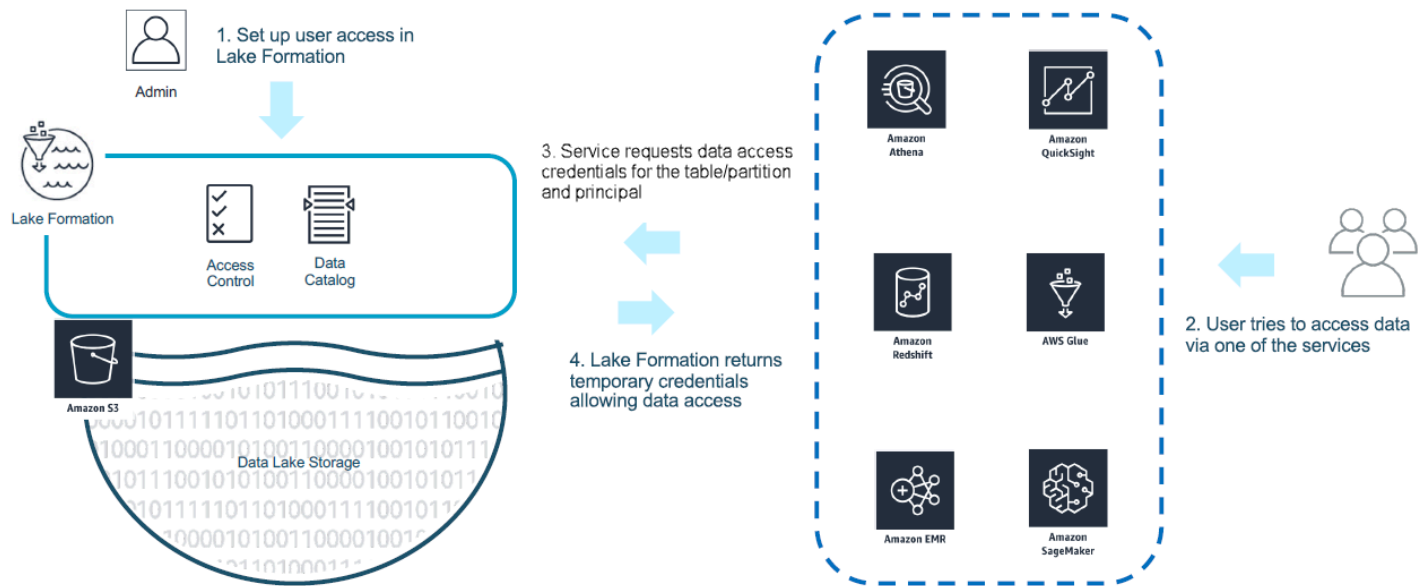
Pour de plus amples informations, veuillez consulter [Intégration des applications pour un accès complet aux tables](#).

Rubriques

- [Comment fonctionne l'intégration de l'application Lake Formation](#)
- [Rôles et responsabilités dans l'intégration de l'application Lake Formation](#)
- [Lake Formation flux de travail pour les opérations d'API d'intégration d'applications](#)
- [Enregistrement d'un moteur de requête tiers](#)
- [Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications](#)
- [Intégration des applications pour un accès complet aux tables](#)

Comment fonctionne l'intégration de l'application Lake Formation

Cette section décrit comment utiliser les opérations de l'API d'intégration d'applications pour intégrer une application tierce (moteur de requête) à Lake Formation.



1. L'Lake Formation administrateur exécute les activités suivantes :

- Enregistre un site Amazon S3 auprès de Lake Formation en fournissant un rôle IAM (utilisé pour les informations d'identification automatiques) doté des autorisations appropriées pour accéder aux données du site Amazon S3
- Enregistre une application tierce pour pouvoir appeler les opérations de l'API de vente d'informations d'identification de Lake Formation. Consultez [the section called “Enregistrement d'un moteur de requête tiers”](#)
- Accorde aux utilisateurs des autorisations d'accès aux bases de données et aux tables

Par exemple, si vous souhaitez publier un ensemble de données de sessions utilisateur comprenant des colonnes contenant des informations personnelles identifiables (PII), pour restreindre l'accès, vous devez attribuer à ces colonnes une balise [LF-TBAC](#) nommée « classification » avec la valeur « sensible ». Vous définissez ensuite une autorisation qui permet à un analyste commercial d'accéder aux données des sessions utilisateur, mais vous excluez les colonnes étiquetées avec classification = sensitive.

2. Un principal (utilisateur) soumet une requête à un service intégré.
3. L'application intégrée envoie la demande à Lake Formation pour demander les informations de la table et les informations d'identification pour accéder à la table.
4. Si le principal demandeur est autorisé à accéder à la table, Lake Formation renvoie les informations d'identification à l'application intégrée, qui autorise l'accès aux données.

 Note

Lake Formation n'accède pas aux données sous-jacentes lors de la vente d'informations d'identification.

5. Le service intégré lit les données d'Amazon S3, filtre les colonnes en fonction des politiques reçues et renvoie les résultats au principal.

 Important

Lake Formation les opérations de l'API de vente d'informations d'identification permettent une application distribuée avec un modèle de refus explicite en cas d'échec (fermeture en cas d'échec). Cela introduit un modèle de sécurité tripartite entre les clients, les services tiers et Lake Formation. Les services intégrés sont fiables pour appliquer correctement Lake Formation les autorisations (application distribuée).

Le service intégré est chargé de filtrer les données lues depuis Amazon S3 en fonction des politiques renvoyées Lake Formation avant que les données filtrées ne soient renvoyées à l'utilisateur. Les services intégrés suivent un modèle de fermeture automatique, ce qui signifie qu'ils doivent échouer à la requête s'ils ne sont pas en mesure d'appliquer les autorisations requises Lake Formation.

Rôles et responsabilités dans l'intégration de l'application Lake Formation

Vous trouverez ci-dessous les rôles et les responsabilités associées pour permettre l'intégration d'applications tierces avec AWS Lake Formation.

Role	Responsabilité
Le client	• Activez le paramètre d'intégration de l'application Lake Formation (voir the section called “Enregistrement d'un moteur de requête tiers”). • Enregistre explicitement les tiers agréés auprès de Lake Formation (voir the section called “Enregistrement d'un moteur de requête tiers”). • Teste et valide les solutions tierces avec les autorisations de Lake Formation.

Role	Responsabilité
	• Surveille et audite l'utilisation par des tiers des opérations de l'API de distribution automatique d'informations d'identification de Lake Formation.
Le tiers	• Documente publiquement les fonctionnalités prises en charge pour chaque révision logicielle et fournit des instructions pour les activer correctement. • Annonce avec précision les fonctionnalités prises en charge lors de l'appel aux opérations de l'API de vente d'informations d'identification de Lake Formation (conformément à la documentation). • Stocke et gère en toute sécurité les informations d'identification vendues afin d'éviter les fuites d'informations d'identification et l'augmentation des privilèges. • Applique les autorisations en fonction des fonctionnalités prises en charge et renvoie uniquement les données filtrées aux utilisateurs • Échoue la requête lorsqu'il est impossible d'appliquer correctement les autorisations requises
AWS Lake Formation	• Dérive et renvoie correctement les autorisations effectives pour un principal donné. • Valide les fonctionnalités prises en charge par des tiers sur la call-by-call base du fonctionnement de l'API. • Renvoie des informations d'identification IAM limitées uniquement lorsque les fonctionnalités annoncées du moteur correspondent à celles définies dans les ressources du catalogue, sinon renvoie une erreur.

Lake Formationflux de travail pour les opérations d'API d'intégration d'applications

Le flux de travail pour les opérations d'API d'intégration d'applications est le suivant :

1. Un utilisateur soumet une requête ou une demande de données à l'aide d'un moteur de requête tiers intégré. Le moteur de requête assume un rôle IAM qui représente l'utilisateur ou un groupe

- d'utilisateurs et récupère des informations d'identification fiables à utiliser lors de l'appel des opérations de l'API d'intégration de l'application.
2. Le moteur de requête appelle `GetUnfilteredTableMetadata`, et s'il s'agit d'une table partitionnée, le moteur de requête appelle `GetUnfilteredPartitionsMetadata` pour récupérer les métadonnées et les informations de politique à partir du catalogue de données.
 3. Lake Formation autorise la demande. Si l'utilisateur ne dispose pas des autorisations appropriées sur la table, il `AccessDeniedException` est renvoyé.
 4. Dans le cadre de la demande, le moteur de requête envoie le filtrage qu'il prend en charge. Deux indicateurs peuvent être envoyés dans un tableau : `COLUMN_PERMISSIONS` et `CELL_FILTER_PERMISSION`. Si le moteur de requête ne prend en charge aucune de ces fonctionnalités et qu'une politique existe sur la table pour la fonctionnalité, un `PermissionTypeMismatchException` est émis et la requête échoue. Cela permet d'éviter les fuites de données.
 5. La réponse renvoyée contient les éléments suivants :
 - Le schéma complet de la table afin que les moteurs de requête puissent l'utiliser pour analyser les données stockées.
 - Liste des colonnes autorisées auxquelles l'utilisateur a accès. Si la liste des colonnes autorisées est vide, cela indique que l'utilisateur dispose d'`DESCRIBE` autorisations, mais pas d'`SELECT` autorisations, et la requête échoue.
 - Un drapeau `IsRegisteredWithLakeFormation`, qui indique si Lake Formation peut fournir des informations d'identification à ces données de ressources. Si le résultat est faux, les informations d'identification des clients doivent être utilisées pour accéder à Amazon S3.
 - Une liste indiquant, le `CellFilters` cas échéant, à appliquer aux lignes de données. Cette liste contient des colonnes et une expression permettant d'évaluer chaque ligne. Ce champ ne doit être renseigné que si `CELL_FILTER_PERMISSION` est envoyé dans le cadre de la demande et s'il existe un filtre de données dans la table pour l'utilisateur appelant.
 6. Une fois les métadonnées récupérées, le moteur de requête appelle `GetTemporaryGlueTableCredentials` ou `GetTemporaryGluePartitionCredentials` pour obtenir des AWS informations d'identification afin de récupérer les données depuis l'emplacement Amazon S3.
 7. Le moteur de requête lit les objets pertinents depuis Amazon S3, filtre les données en fonction des politiques reçues à l'étape 2 et renvoie les résultats à l'utilisateur.

Les opérations de l'API d'intégration d'applications Lake Formation contiennent du contenu supplémentaire pour configurer l'intégration avec des moteurs de requêtes tiers. Vous pouvez consulter les détails des opérations dans la section [Opérations de l'API Credential vending](#).

QuerySessionContextIl s'agit d'une structure à laquelle les moteurs de requêtes peuvent également envoyer des informations Lake Formation pour ces opérations d'API d'intégration d'applications. Il permet Lake Formation de stocker et d'utiliser un contexte supplémentaire pour une requête donnée. Voici un exemple de la façon dont [QuerySessionContext](#) convient de l'utiliser :

1. Le moteur de requête effectue un GetInternalUnfilteredMetadata appel en transmettant une structure QSC contenant un identifiant de requête unique dans la demande :

```
{
  "QuerySessionContext": {
    "QueryId": "your-unique-identifiant-here"
  }
}
```

2. L'GetInternalUnfilteredMetadataappel aura renvoyé une QueryAuthorizationId chaîne dans la réponse. Lors du prochain appel de requête (et de tout appel ultérieur) qui accepte une structure QSC en entrée, le moteur de requête transmet la même structure QSC qui contient désormais également le résultat QueryAuthorizationId renvoyé par. Lake Formation Supposons que le prochain appel soit GetTemporaryGlueTableCredentials ; la demande contiendra :

```
{
  "QuerySessionContext": {
    "QueryAuthorizationId": "lf-returned-query-authz-id-here",
    "QueryId": "your-unique-identifiant-here"
  },
}
```

Enregistrement d'un moteur de requête tiers

Avant qu'un moteur de requête tiers puisse utiliser les opérations de l'API d'intégration d'applications, vous devez explicitement autoriser le moteur de requêtes à appeler les opérations d'API en votre nom. Cela se fait en quelques étapes :

1. Vous devez spécifier les AWS comptes et les balises de session IAM qui nécessitent une autorisation pour appeler les opérations de l'API d'intégration des applications via la AWS Lake Formation console, l'API/SDK AWS CLI ou l'API/le.
2. Lorsque le moteur de requête tiers assume le rôle d'exécution dans votre compte, il doit associer une balise de session enregistrée auprès de Lake Formation représentant le moteur tiers. Lake Formation utilise cette balise pour le valider si la demande provient d'un moteur approuvé. Pour plus d'informations sur les balises de session, consultez la section [Balises de session](#) dans le guide de l'utilisateur IAM.
3. Lorsque vous configurez un rôle d'exécution de moteur de requête tiers, vous devez disposer des autorisations minimales suivantes dans la politique IAM :

JSON

```
{
  "Version": "2012-10-17",
  "Statement": {
    "Effect": "Allow",
    "Action": [
      "lakeformation:GetDataAccess",
      "glue:GetTable",
      "glue:GetTables",
      "glue:GetDatabase",
      "glue:GetDatabases",
      "glue>CreateDatabase",
      "glue:GetUserDefinedFunction",
      "glue:GetUserDefinedFunctions",
      "glue:GetPartition",
      "glue:GetPartitions"
    ],
    "Resource": "*"
  }
}
```

4. Configurez une politique d'approbation des rôles sur le rôle d'exécution du moteur de requête afin de contrôler avec précision la paire clé-valeur de balise de session qui peut être attachée à ce rôle. Dans l'exemple suivant, ce rôle est uniquement autorisé à associer une clé de balise de session "LakeFormationAuthorizedCaller" et une valeur de balise "engine1" de session, et aucune autre paire clé-valeur de balise de session n'est autorisée.

```
{
  "Sid": "AllowPassSessionTags",
```

```
"Effect": "Allow",
"Principal": {
  "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
},
"Action": "sts:TagSession",
"Condition": {
  "StringLike": {
    "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"
  }
}
```

Lorsque LakeFormationAuthorizedCaller l'opération STS : AssumeRole API est appelée pour récupérer les informations d'identification à utiliser par le moteur de requête, le tag de session doit être inclus dans la [AssumeRole demande](#). Les informations d'identification temporaires renvoyées peuvent être utilisées pour effectuer des demandes d'API d'intégration d'Lake Formation applications.

Lake Formation les opérations de l'API d'intégration d'applications nécessitent que le principal appelant soit un rôle IAM. Le rôle IAM doit inclure une balise de session avec une valeur prédéterminée enregistrée avec Lake Formation. Cette balise permet Lake Formation de vérifier que le rôle utilisé pour appeler les opérations de l'API d'intégration de l'application est autorisé à le faire.

Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications

Procédez comme suit pour autoriser un moteur de requête tiers à appeler les opérations de l'API d'intégration d'applications via la AWS Lake Formation console, l'API/SDK AWS CLI ou l'API/le.

Console

Pour enregistrer votre compte pour le filtrage externe des données :

1. Connectez-vous au et ouvrez AWS Management Console la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, développez Administration, puis choisissez Paramètre d'intégration des applications.
3. Sur la page des paramètres d'intégration des applications, choisissez l'option Autoriser les moteurs externes à filtrer les données dans les sites Amazon S3 enregistrés auprès de Lake Formation.

4. Entrez les balises de session que vous avez créées pour le moteur tiers. Pour plus d'informations sur les balises de session, voir [Transmission de balises de session dans AWS STS](#) dans le Guide de Gestion des identités et des accès AWS l'utilisateur.
5. Entrez le compte IDs pour les utilisateurs qui peuvent utiliser le moteur tiers pour accéder aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données des ressources du compte courant.

Vous pouvez également utiliser le champ ID du AWS compte pour configurer l'accès entre comptes.

Application integration settings [Learn more](#)

Application integration settings
Use the options below to control which third-party engines are allowed to read and filter data in Amazon S3 locations registered with Lake Formation.

☒ **Allow external engines to filter data in Amazon S3 locations registered with Lake Formation**
Check this box to allow third-party engines to access data in Amazon S3 locations that are registered with Lake Formation.

Session tag values
Enter one or more strings that match the LakeFormationAuthorizedCaller session tag defined for third-party engines.

Enter one or several string values separated by comma.

AWS account IDs
Enter the external AWS account IDs from where third-party engines are allowed to access locations registered with Lake Formation.

Account Account

Enter one or more AWS account IDs. Press enter after each ID.

☐ **Allow external engines to access data in Amazon S3 locations with full table access.**
When you enable this option, Lake Formation will return credentials to the integrated application directly without IAM session tag validation.

CLI

Utilisez la commande `put-data-lake-settings` CLI pour définir les paramètres suivants.

Trois champs sont à configurer lors de l'utilisation de cette AWS CLI commande :

- `allow-external-data-filtering` — (booléen) Indique qu'un moteur tiers peut accéder aux informations de métadonnées non filtrées et aux informations d'accès aux données des ressources du compte courant.

- `external-data-filtering-allow-list`— (tableau) Liste de comptes IDs pouvant accéder aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données des ressources du compte courant lors de l'utilisation d'un moteur tiers. Lorsqu'elle `AllowExternalDataFiltering` est définie sur `true`, la `ExternalDataFilteringAllowList` propriété doit inclure au moins un identifiant de compte. Une liste vide n'est pas autorisée.
- `authorized-sessions-tag-value-list`— (tableau) Liste des valeurs de balises de session autorisées (chaînes). Si un identifiant de rôle IAM a été associé à une paire clé-valeur autorisée, si le tag de session est inclus dans la liste, la session a accès aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données sur les ressources du compte configuré. La clé de balise de session autorisée est définie comme `*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess`- (booléen) S'il faut autoriser un moteur de requête tiers à obtenir des informations d'accès aux données sans balises de session lorsqu'un appelant dispose d'autorisations d'accès complètes aux données.

Par exemple :

```
aws lakeformation put-data-lake-settings --cli-input-json file://
datalakesettings.json

{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/lakeAdmin"
      }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "TrustedResourceOwners": [],
    "AllowExternalDataFiltering": true,
    "ExternalDataFilteringAllowList": [
      {"DataLakePrincipalIdentifier": "111111111111"}
    ],
    "AuthorizedSessionTagValueList": ["engine1"],
    "AllowFullTableExternalDataAccess": false
  }
}
```

```
}
```

API/SDK

Utilisez l'opération `PutDataLakeSetting` API pour définir les paramètres suivants.

Trois champs doivent être configurés lors de l'utilisation de cette opération d'API :

- `AllowExternalDataFiltering`— (Booléen) Indique si un moteur tiers peut accéder aux informations de métadonnées non filtrées et aux informations d'accès aux données des ressources du compte courant.
- `ExternalDataFilteringAllowList`— (tableau) Liste de comptes IDs pouvant accéder à des informations de métadonnées non filtrées et aux informations d'identification d'accès aux données des ressources du compte courant à l'aide d'un moteur tiers.
- `AuthorizedSectionsTagValueList`— (tableau) Liste des valeurs de balises autorisées (chaînes). Si un identifiant de rôle IAM a été associé à une balise autorisée, la session a accès aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données sur les ressources du compte configuré. La clé de balise de session autorisée est définie comme `*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess`- (booléen) S'il faut autoriser un moteur de requête tiers à obtenir des informations d'accès aux données sans balises de session lorsqu'un appelant dispose d'autorisations d'accès complètes aux données.

Par exemple :

```
//Enable session tag on existing data lake settings
public void sessionTagSetUpForExternalFiltering(AWSLakeFormationClient
lakeformation) {
    GetDataLakeSettingsResult getDataLakeSettingsResult =
    lfClient.getDataLakeSettings(new GetDataLakeSettingsRequest());
    DataLakeSettings dataLakeSettings =
    getDataLakeSettingsResult.getDataLakeSettings();

    //set account level flag to allow external filtering
    dataLakeSettings.setAllowExternalDataFiltering(true);

    //set account that are allowed to call credential vending or Glue
    GetFilteredMetadata API
    List<DataLakePrincipal> allowlist = new ArrayList<>();
```



```
allowlist.add(new
DataLakePrincipal().withDataLakePrincipalIdentifier("111111111111"));
dataLakeSettings.setWhitelistedForExternalDataFiltering(allowlist);

//set registered session tag values
List<String> registeredTagValues = new ArrayList<>();
registeredTagValues.add("engine1");
dataLakeSettings.setAuthorizedSessionTagValueList(registeredTagValues);

lakeformation.putDataLakeSettings(new
PutDataLakeSettingsRequest().withDataLakeSettings(dataLakeSettings));
}
```

Intégration des applications pour un accès complet aux tables

Procédez comme suit pour permettre aux moteurs de requêtes tiers d'accéder aux données sans validation du tag de session IAM :

Console

1. Connectez-vous à la console Lake Formation à l'adresse <https://console.aws.amazon.com/lakeformation/>.
2. Dans le volet de navigation de gauche, développez Administration, puis sélectionnez Paramètres d'intégration des applications.
3. Sur la page des paramètres d'intégration des applications, choisissez l'option Autoriser les moteurs externes à accéder aux données dans les emplacements Amazon S3 avec accès complet aux tables.

Lorsque vous activez cette option, Lake Formation renvoie les informations d'identification à l'application demandeuse directement sans validation des balises de session IAM.

Application integration settings [Learn more](#)

Application integration settings

Use the options below to control which third-party engines are allowed to read and filter data in Amazon S3 locations registered with Lake Formation.

☐ **Allow external engines to filter data in Amazon S3 locations registered with Lake Formation**

Check this box to allow third-party engines to access data in Amazon S3 locations that are registered with Lake Formation.

Session tag values

Enter one or more strings that match the LakeFormationAuthorizedCaller session tag defined for third-party engines.

engine 1 ✕

engine 2 ✕

session 1 ✕

Enter one or several string values separated by comma.

AWS account IDs

Enter the external AWS account IDs from where third-party engines are allowed to access locations registered with Lake Formation.

111111111111 ✕
Account

222222222222 ✕
Account

Enter one or more AWS account IDs. Press enter after each ID.

☒ **Allow external engines to access data in Amazon S3 locations with full table access.**

When you enable this option, Lake Formation will return credentials to the integrated application directly without IAM session tag validation.

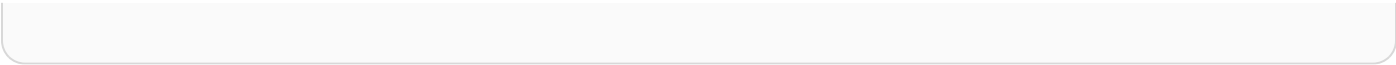
Cancel

Save

AWS CLI

Utilisez la commande `put-data-lake-settings` CLI pour définir le `AllowFullTableExternalDataAccess` paramètre.

```
aws lakeformation put-data-lake-settings --cli-input-json file://put-data-lake-
settings.json --region ap-northeast-1
{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/
lakeAdmin"
      }
    ],
    "AllowFullTableExternalDataAccess": true
  }
}
```



Collaboration avec d'autres AWS services

AWS des services tels qu'Amazon Athena AWS Glue, Amazon Redshift Spectrum et Amazon EMR peuvent être AWS Lake Formation utilisés pour accéder en toute sécurité aux données des sites Amazon S3 enregistrés auprès de Lake Formation. Avec Lake Formation, vous pouvez définir et gérer des autorisations de contrôle d'accès détaillées (FGAC) pour vos tables dans le. AWS Glue Data Catalog Chacun de ces AWS services est un interlocuteur fiable de Lake Formation, et Lake Formation fournit un accès aux données stockées dans Amazon S3 via des informations d'identification temporaires. Pour de plus amples informations, veuillez consulter [Comment fonctionne l'intégration de l'application Lake Formation](#).

Pour bénéficier de ces fonctionnalités, Lake Formation vous demande d'abord d'enregistrer l'emplacement Amazon S3 et d'attribuer les autorisations appropriées au principal IAM pour accéder à la table, à la base de données et à l'emplacement Amazon S3. Pour plus d'informations, voir, [Gestion des autorisations relatives à Lake Formation](#).

Les tableaux suivants répertorient les types d'autorisations Lake Formation pris en charge par Amazon Athena, Amazon EMR et AWS Glue Amazon Redshift Spectrum pour accéder aux données à AWS Glue partir de tables standard et de tables transactionnelles ([Apache Iceberg](#), [Apache Hudi](#) et [Linux foundation Delta Lake](#)) avec des données stockées dans Amazon S3 et des métadonnées de table dans le catalogue de données.

AWS services et types d'autorisations pris en charge pour les tables et les vues AWS Glue standard

AWS service	Autorisations au niveau du tableau	Autorisations au niveau des colonnes	Autorisations au niveau des lignes et des cellules
Athena SQL	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture
Athena Spark	Non pris en charge	Non pris en charge	Non pris en charge
Redshift Spectrum sur un cluster provisionné ou Amazon Redshift sans serveur	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture

AWS service	Autorisations au niveau du tableau	Autorisations au niveau des colonnes	Autorisations au niveau des lignes et des cellules
Apache Spark sur Amazon EMR (EC2)	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture
Apache Hive sur Amazon EMR (EC2)	Accès en lecture/écriture	Accès à la lecture	Non pris en charge
Apache Spark sur EMR sans serveur	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture
Apache Hive sur EMR sans serveur	Non pris en charge	Non pris en charge	Non pris en charge
Amazon EMR on EKS	Non pris en charge	Non pris en charge	Non pris en charge
AWS Glue ETL	Accès en lecture/écriture	AWS Glue La version 5.0 ou supérieur prend en charge l'accès en lecture.	AWS Glue La version 5.0 ou supérieur prend en charge l'accès en lecture.

Considérations et restrictions

- Athena Spark ne permet pas d'interroger les tables du catalogue de données avec les autorisations Lake Formation.
- Les utilisateurs basés sur Athena SAML peuvent lire les sources de données sécurisées à l'aide des autorisations Lake Formation en activant la fédération basée sur SAML 2.0. Les utilisateurs de SAML peuvent insérer des données dans les tables Parquet.
- Apache Spark sur EMR Serverless ne prend pas en charge l'interrogation des vues du catalogue de données.
- Apache Hive sur EMR Serverless ne prend pas en charge l'interrogation de tables avec les autorisations Lake Formation.
- AWS Glue La version 5.0 ou supérieure prend en charge des contrôles d'accès précis sur les tables Iceberg et Hive du catalogue de données soutenues par S3. Cette fonctionnalité vous

permet de configurer des contrôles d'accès au niveau des tables, des lignes, des colonnes et des cellules pour les requêtes de lecture dans vos AWS Glue tâches Apache Spark.

Pour plus d'informations, consultez [Version d'AWS Glue](#).

AWS services et types d'autorisation pris en charge pour les formats de tables transactionnels

AWS service	Iceberg	Hudi	Delta Lake (natif)	Lac Delta (tableaux à liens symboliques)
Athena SQL	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture nécessitent un accès complet à la table.	Prend en charge les opérations de lecture et de création sur les tables avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	Athena (version 3 du moteur) prend en charge la lecture des tables natives de Delta Lake avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	Athena (version 3 du moteur) prend en charge la lecture des tables Delta Lake par liens symboliques avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.
Redshift Spectrum sur un cluster provisionné	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des	Non pris en charge	Permet de lire les tables de Delta Lake via un manifeste de liens symboliques avec des autorisations

AWS service	Iceberg	Hudi	Delta Lake (natif)	Lac Delta (tableaux à liens symboliques)
	lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.		au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.
Apache Spark sur Amazon EMR (EC2)	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture nécessitent un accès complet à la table.	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture nécessitent un accès complet à la table.	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture nécessitent un accès complet à la table.

AWS service	Iceberg	Hudi	Delta Lake (natif)	Lac Delta (tableaux à liens symboliques)
AWS Glue ETL	AWS Glue La version 5.0 ou supérieure prend en charge la lecture de tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules.	Supports read/write sur les tables avec des autorisations au niveau des tables.	Supports read/write sur les tables avec des autorisations au niveau des tables.	Supports read/write sur les tables avec des autorisations au niveau des tables.

Rubriques

- [Utilisation AWS Lake Formation avec Amazon Athena](#)
- [Utilisation AWS Lake Formation avec Amazon Redshift Spectrum](#)
- [Utilisation AWS Lake Formation avec AWS Glue](#)
- [Utilisation AWS Lake Formation avec Amazon EMR](#)
- [Utilisation AWS Lake Formation avec Quick](#)
- [Utilisation AWS Lake Formation avec AWS CloudTrail Lake](#)

Utilisation AWS Lake Formation avec Amazon Athena

[Amazon Athena](#) est un service de requête sans serveur qui vous aide à analyser les données structurées, semi-structurées et non structurées stockées dans Amazon S3. Vous pouvez utiliser Athena SQL pour interroger des données aux formats de données CSV, JSON, Parquet et Avro. [Athena SQL prend également en charge les formats de table tels qu'Apache Hive, Apache Hudi et Apache Iceberg.](#) Athena s'intègre au système AWS Glue Data Catalog pour stocker les métadonnées de vos ensembles de données dans Amazon S3. Athena peut utiliser Lake Formation pour définir et maintenir des politiques de contrôle d'accès sur ces ensembles de données.

Voici quelques cas d'utilisation courants dans lesquels vous pouvez utiliser Lake Formation avec Athena.

- Utilisez les autorisations de Lake Formation pour accéder aux ressources du catalogue de données (base de données et tables) depuis Athena. Vous pouvez utiliser la méthode de ressource nommée ou des balises LF pour définir des autorisations sur la base de données et les tables. Pour en savoir plus, consultez :
 - [Octroi d'autorisations de base de données via la méthode de ressource nommée](#)
 - [Contrôle d'accès basé sur des balises Lake Formation](#)

 Note

Les autorisations de Lake Formation s'appliquent uniquement lorsque vous utilisez Athena SQL pour interroger les données sources d'Amazon S3 et les métadonnées du catalogue de données.

Athena Spark ne permet pas d'interroger les tables du catalogue de données avec les autorisations Lake Formation. Les autorisations de Lake Formation prennent en charge les opérations de lecture et d'écriture sur les bases de données et les tables.

 Note

Vous ne pouvez pas appliquer de filtres de données lorsque vous utilisez des balises LF pour gérer les autorisations sur les ressources du catalogue de données.

- Contrôlez les résultats des requêtes en [Filtres de données dans Lake Formation](#) sécurisant les tables de vos lacs de données Amazon S3 en accordant des autorisations au niveau des colonnes, des lignes et des cellules. Consultez les [limites relatives à la projection de partitions](#) dans le guide de l'utilisateur d'Amazon Athena.
- Appliquez un contrôle d'accès précis aux données accessibles à l'utilisateur Athena basé sur SAML lors de l'exécution de requêtes fédérées.

Les pilotes JDBC et ODBC Athena prennent en charge la configuration de l'accès fédéré à votre source de données à l'aide d'un fournisseur d'identité (IdP) basé sur SAML. Utilisez Quick intégré à Lake Formation avec votre rôle IAM existant ou avec des utilisateurs ou des groupes SAML pour visualiser les résultats des requêtes Athena.

Note

Les autorisations de Lake Formation pour les utilisateurs et les groupes SAML ne s'appliquent que lorsque vous soumettez des requêtes à Athena à l'aide du pilote JDBC ou ODBC.

Pour plus d'informations, consultez la section [Utilisation de Lake Formation et des pilotes JDBC et ODBC Athena pour un accès fédéré](#) à Athena.

Note

Actuellement, l'autorisation d'accès aux identités SAML dans Lake Formation n'est pas prise en charge dans les régions suivantes :

- Moyen-Orient (Bahreïn) – me-south-1
- Asie-Pacifique (Hong Kong) – ap-east-1
- Afrique (Le Cap) – af-south-1
- Chine (Ningxia) – cn-northwest-1
- Asie-Pacifique (Osaka) – ap-northeast-3

- [Partage de données entre comptes dans Lake Formation](#) À utiliser pour interroger les tables d'un autre compte.

Note

Pour plus d'informations sur les limites liées à l'utilisation des autorisations de Lake Formation pour Views, voir [Considérations et limites](#).

Support pour les formats de tables transactionnels

L'application des autorisations Lake Formation vous permet de sécuriser vos données transactionnelles dans vos lacs de données basés sur Amazon S3. Le tableau ci-dessous répertorie les formats de tables transactionnels pris en charge dans les autorisations Athena et Lake Formation. Lake Formation applique ces autorisations lorsque les utilisateurs d'Athena exécutent leurs requêtes.

Format de tableau	Description et opérations autorisées	Autorisations relatives à la formation du lac prises en charge dans Athena
Apache Hudi	Format utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Athena prend en charge les opérations de création et de lecture à l'aide des formats de table Apache Hudi sur les ensembles de données Amazon S3 pour les types de tables Copy on Write (CoW) et Merge On Read (MoR) Hudi. Athena ne prend pas en charge les opérations d'écriture sur les tables Hudi. Utilisez Athena pour interroger les ensembles de données Hudi.	Filtrage des données et sécurité au niveau des cellules dans Lake Formation À utiliser pour sécuriser la table Hudi à l'aide d'autorisations au niveau des tables, des colonnes, des lignes et des cellules.
Apache Iceberg	Format de table ouvert qui gère de grandes collections de fichiers sous forme de tables et prend en charge les opérations de lac de données analytiques modernes, telles que les requêtes d'insertion, de mise à jour, de suppression et de voyage dans le temps au niveau des enregistrements.	Les autorisations au niveau des tables, des colonnes, des lignes et des cellules sont prises en charge. Actuellement, Lake Formation ne prend pas en charge la gestion des autorisations sur les opérations d'écriture telles que VACUUMMERGE, UPDATE et OPTIMIZE sur les tables dans Open Table Formats.

Format de tableau	Description et opérations autorisées	Autorisations relatives à la formation du lac prises en charge dans Athena
	Pour plus d'informations sur la prise en charge des tables Iceberg par Athena, voir Utilisation des tables Iceberg .	
Linux Foundation Delta Lake	Delta Lake est un projet open source qui aide à implémenter des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Athena prend en charge les tables Delta Lake créées à l'aide d'une définition de table manifeste basée sur des liens symboliques à AWS Glue Data Catalog partir d'une table Delta Lake. Pour plus d'informations, consultez les tables Crawl Delta Lake à l'aide de AWS Glue crawlers. Athena (version 3 du moteur) prend en charge la lecture des tables natives de Delta Lake. Pour plus d'informations, voir Présentation du support de table natif de Delta Lake avec des AWS Glue crawlers.	Les autorisations au niveau des tables, des colonnes, des lignes et des cellules sont prises en charge pour les tables de liens symboliques et les tables natives de Delta Lake.

Ressources supplémentaires

Articles de blog, vidéos et ateliers

- [Interrogez un ensemble de données Apache Hudi dans un lac de données Amazon S3 avec Amazon Athena](#)
- [Créez un lac de données Apache Iceberg à l'aide d'Amazon Athena, Amazon EMR et AWS Glue](#)
- [Insérer, mettre à jour, supprimer sur Amazon S3 avec Athena et Apache Iceberg](#)
- Atelier Lake Formation sur le [contrôle d'accès basé sur des balises LF-Tag](#) sur l'interrogation d'un lac de données.

Utilisation AWS Lake Formation avec Amazon Redshift Spectrum

[Amazon Redshift](#) Spectrum vous permet d'interroger et de récupérer des données dans les lacs de données Amazon S3 sans charger de données dans les nœuds du cluster Amazon Redshift.

Redshift Spectrum propose deux méthodes pour enregistrer un catalogue de AWS Glue données externe activé avec Lake Formation.

- Utilisation d'un rôle IAM attaché à un cluster autorisé à accéder au catalogue de données

Pour créer un rôle IAM, suivez les étapes décrites dans la procédure ci-dessous.

[Contrôle de l'accès au AWS Glue Data Catalog](#)

- Utilisation d'une identité IAM fédérée configurée pour gérer l'accès aux ressources externes AWS Glue Data Catalog

Redshift Spectrum permet d'interroger les tables de Lake Formation à l'aide d'identités IAM fédérées. Les identités IAM peuvent être un utilisateur IAM ou un rôle IAM. Pour plus d'informations sur la fédération d'identité IAM dans Redshift Spectrum, [consultez Utilisation d'une identité fédérée pour gérer l'accès d'Amazon Redshift aux ressources locales et aux tables externes Redshift Spectrum](#).

Grâce à l'intégration de Lake Formation à Redshift Spectrum, vous pouvez définir des autorisations de contrôle d'accès au niveau des lignes, des colonnes et des cellules sur les tables une fois vos données enregistrées auprès de Lake Formation.

Pour plus d'informations, consultez la section [Utilisation de Redshift Spectrum](#) avec. AWS Lake Formation

Redshift Spectrum prend en charge les lectures ou les SELECT requêtes sur les tables de schéma externes gérées par Lake Formation.

Pour plus d'informations, consultez [Création de schémas externes pour Redshift Spectrum](#).

Support pour les types de tables transactionnels

Ce tableau répertorie les formats de tables transactionnels pris en charge dans Redshift Spectrum et les autorisations Lake Formation applicables.

Formats de tableau pris en charge

Format de tableau	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Redshift Spectrum
Apache Hudi	Format utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Redshift Spectrum prend en charge les opérations d'écriture et d'insertion, de suppression et d'insertion à l'aide du format de table Apache Hudi Copy on Write (CoW) sur Amazon S3. Pour plus d'informations, voir Création de tables externes pour les données gérées dans Apache Hudi.	Filtrage des données et sécurité au niveau des cellules dans Lake Formation À utiliser pour sécuriser les tables Hudi à l'aide d'autorisations au niveau des tables, des colonnes, des lignes et des cellules.
Apache Iceberg	Format de table ouvert qui gère de grandes collections de fichiers sous forme de tables et prend en charge les	Redshift Spectrum prend en charge les tables Apache Iceberg pour les requêtes.

Format de tableau	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Redshift Spectrum
	opérations de lac de données analytiques modernes, telles que les requêtes d'insertion, de mise à jour, de suppression et de voyage dans le temps au niveau des enregistrements. Pour plus d'informations, consultez la section Utilisation des tables Apache Iceberg avec Amazon Redshift.	
Linux Foundation Delta Lake	Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Redshift Spectrum prend en charge l'interrogation des tables Delta Lake. Pour plus d'informations, voir Création de tables externes pour les données gérées dans Delta Lake.	Les autorisations au niveau des tables, des colonnes, des lignes et des cellules sont prises en charge.

Ressources supplémentaires

Articles de blog et ateliers

- [Centralisez la gouvernance de votre lac de données AWS Lake Formation tout en mettant en place une architecture de données moderne avec Amazon Redshift Spectrum](#)
- [Utilisez Redshift Spectrum pour interroger les tables Apache HUDI Copy On Write \(CoW\) dans le lac de données Amazon S3](#)

Utilisation AWS Lake Formation avec AWS Glue

Les ingénieurs des données et DevOps les professionnels utilisent AWS Glue Extract, Transform and Load (ETL) avec Apache Spark pour transformer leurs ensembles de données dans Amazon S3 et charger les données transformées dans des lacs de données et des entrepôts de données à des fins d'analyse, d'apprentissage automatique et de développement d'applications. Différentes équipes accédant au même ensemble de données dans Amazon S3, il est impératif d'accorder et de restreindre les autorisations en fonction de leurs rôles.

AWS Lake Formation est basé sur AWS Glue, et les services interagissent de la manière suivante :

- Lake Formation et AWS Glue partagent le même catalogue de données.
- Les fonctionnalités suivantes de la console Lake Formation invoquent la AWS Glue console :
 - Tâches — Pour plus d'informations, consultez la section [Ajouter des tâches](#) dans le Guide du AWS Glue développeur.
 - Crawlers — Pour plus d'informations, voir [Catalogage de tables à l'aide d'un robot d'exploration](#) dans le Guide du AWS Glue développeur.
- Les flux de travail générés lorsque vous utilisez un plan de Lake Formation sont des AWS Glue flux de travail. Vous pouvez consulter et gérer ces flux de travail à la fois dans la console Lake Formation et dans la AWS Glue console.
- Les transformations d'apprentissage automatique sont fournies avec Lake Formation et reposent sur des opérations AWS Glue d'API. Vous créez et gérez des transformations de machine learning sur la AWS Glue console. Pour plus d'informations, consultez [Machine Learning Transforms](#) dans le Guide du AWS Glue développeur.

Vous pouvez utiliser le contrôle précis des accès de Lake Formation pour gérer les ressources de votre catalogue de données existantes et les emplacements de données Amazon S3.

 Note

AWS Glue La version 5.0 ou supérieure prend en charge des contrôles d'accès précis sur les tables Iceberg et Hive soutenues par S3. Cette fonctionnalité vous permet de configurer des contrôles d'accès au niveau des tables, des lignes, des colonnes et des cellules pour les requêtes de lecture dans vos AWS Glue tâches Apache Spark.

Support pour les types de tables transactionnels

L'application des autorisations Lake Formation vous permet de sécuriser vos données transactionnelles dans vos lacs de données basés sur Amazon S3. Le tableau ci-dessous répertorie les formats de tables transactionnels pris en charge dans les autorisations Lake Formation AWS Glue et les autorisations. Lake Formation applique ces autorisations pour les AWS Glue opérations.

Formats de tableau pris en charge

Format de tableau	Description et opérations autorisées	Permissions de Lake Formation prises en charge dans AWS Glue
Apache Hudi	Format de table ouvert utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Pour des exemples, voir Utilisation du framework Hudi dans AWS Glue.	Des autorisations au niveau des tables sont disponibles pour les tables Hudi. Pour plus d'informations, consultez Limites.
Apache Iceberg	Format de tableau ouvert qui gère de grandes collections de fichiers sous forme de tableaux. Pour des exemples, voir Utilisation du framework Iceberg dans AWS Glue.	AWS Glue la version 5.0 et supérieure vous permet de configurer des contrôles d'accès au niveau des tables, des lignes, des colonnes et des cellules pour les requêtes de lecture dans vos AWS Glue

Format de tableau	Description et opérations autorisées	Permissions de Lake Formation prises en charge dans AWS Glue
		tâches Apache Spark pour les tables Iceberg. Pour plus d'informations, consultez Limites.
Linux Foundation Delta Lake	Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Pour des exemples, voir Utilisation du framework Delta Lake dans AWS Glue.	Des autorisations au niveau des tables sont disponibles pour les tables Delta Lake. Pour plus d'informations, consultez Limites.

Ressources supplémentaires

Articles de blog et référentiels

- [Utilisez le AWS Glue connecteur pour lire et écrire des tables Apache Iceberg avec des transactions ACID et voyager dans le temps](#)
- [Écrire dans des tables Apache Hudi à l'aide d'un connecteur AWS Glue personnalisé](#)
- AWS référentiel du [modèle Cloudformation et d'un exemple de code pyspark](#) pour analyser les données de streaming à l'aide d' AWS Glue Apache Hudi et d'Amazon S3.

Utilisation AWS Lake Formation avec Amazon EMR

Amazon EMR est une plateforme de cluster AWS gérée flexible sur laquelle vous pouvez exécuter n'importe quel code personnalisé sur des frameworks de données compatibles tels que Hadoop Map-

Reduce, Spark, Hive, Presto, etc. Organisations utilisent également Amazon EMR pour exécuter des applications de traitement de données par lots et en flux sur un cluster hautement distribué. À l'aide d'Apache Spark sur Amazon EMR, vous pouvez exécuter vos transformations de données et votre code personnalisé sur une base de données et des tables dont les autorisations sont gérées par Lake Formation.

Il existe trois options pour déployer Amazon EMR :

- EMR sur EC2
- EMR sans serveur
- Amazon EMR on EKS

Pour plus d'informations, consultez [Intégrer Amazon EMR à Lake Formation](#) ou Utiliser [EMR sans serveur](#) pour un contrôle d'accès précis AWS Lake Formation

Support pour les formats de tables transactionnels

Les versions 6.15.0 et ultérieures d'Amazon EMR incluent la prise en charge des autorisations de contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules de Lake Formation sur les formats de table [Apache Hudi](#), [Apache Iceberg](#) et [Delta Lake](#) lorsque vous lisez et écrivez des données avec Spark SQL.

Pour connaître les limites, consultez [Considérations relatives à Amazon EMR with Lake Formation](#).

Formats de tableau pris en charge

Format de tableau	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Amazon EMR
Apache Hudi	Format de table ouvert utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Pour une liste des opérations prises en charge, consultez	Amazon EMR prend en charge le contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules avec Apache Hudi.

Format de tableau	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Amazon EMR
	Apache Hudi et Lake Formation.	
Apache Iceberg	Format de tableau ouvert qui gère de grandes collections de fichiers sous forme de tableaux. Pour une liste des opérations prises en charge, consultez Apache Iceberg et Lake Formation.	Amazon EMR prend en charge le contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules avec Apache Iceberg.
Linux Foundation Delta Lake	Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Pour une liste des opérations prises en charge, voir Delta Lake and Lake Formation.	Amazon EMR prend en charge le contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules avec les tables Delta Lake.

Ressources supplémentaires

Guide de l'utilisateur, articles de blog et ateliers

- [Intégration à Amazon EMR à l'aide de rôles d'exécution](#)
- [Démarrez rapidement avec Apache Hudi, Apache Iceberg et Delta Lake avec Amazon EMR sur EKS](#)

- [Utilisation de Delta Lake OSS avec EMR Serverless](#)

Utilisation AWS Lake Formation avec Quick

Quick permet d'explorer les ensembles de données gérés par les autorisations Lake Formation dans Amazon S3 à l'aide d'Athena.

Les utilisateurs des éditions Standard et Enterprise de Quick s'intègrent à Lake Formation, mais de manière légèrement différente.

- Édition Enterprise — Accordez des autorisations de contrôle d'accès détaillé (FGAC) à des utilisateurs individuels et à des groupes de Quick pour accéder aux bases de données et aux tables.
- Édition standard : accordez des autorisations aux rôles IAM pour accéder aux bases de données et aux tables.

Note

Par défaut, Quick utilise un rôle nommé `aws-quicksight-service-role-v0`. Vous pouvez également définir des rôles personnalisés avec les autorisations requises pour permettre à Quick d'accéder à Athena.

Pour plus d'informations, voir [Autoriser les connexions via AWS Lake Formation](#)

Ressources supplémentaires

Billets de blogs

- [Activez des autorisations détaillées pour les auteurs rapides dans AWS Lake Formation](#)
- [Analysez vos données en toute sécurité avec AWS Lake Formation et Quick](#)

Utilisation AWS Lake Formation avec AWS CloudTrail Lake

AWS CloudTrail Lake permet d'explorer les magasins de données d'événements à l'Amazon Athena aide d'autorisations détaillées dans. AWS Lake Formation

 Note

CloudTrail Le lac ne peut être interrogé que par son intermédiaire. Amazon Athena

Pour enregistrer votre banque de données d'événements CloudTrail Lake auprès de Lake Formation, voir [Fédérer une banque de données d'événements](#).

Enregistrement des appels d'API AWS Lake Formation à l'aide de AWS CloudTrail

AWS Lake Formation est intégré à AWS CloudTrail un service qui fournit un enregistrement des actions entreprises par un utilisateur, un rôle ou un AWS service dans Lake Formation. CloudTrail capture tous les appels de l'API Lake Formation sous forme d'événements. Les appels capturés incluent les appels provenant de la console Lake Formation AWS Command Line Interface, les appels et les appels de code vers les actions de l'API Lake Formation. Si vous créez un suivi, vous pouvez activer la diffusion continue des CloudTrail événements vers un compartiment Amazon S3, y compris les événements relatifs à Lake Formation. Si vous ne configurez pas de suivi, vous pouvez toujours consulter les événements les plus récents dans la CloudTrail console dans Historique des événements. À l'aide des informations collectées par CloudTrail, vous pouvez déterminer la demande qui a été faite à Lake Formation, l'adresse IP à partir de laquelle la demande a été faite, qui a fait la demande, quand elle a été faite et des détails supplémentaires.

Pour en savoir plus CloudTrail, consultez le [guide de AWS CloudTrail l'utilisateur](#).

Informations sur la formation du lac en CloudTrail

CloudTrail est activé par défaut lorsque vous créez un nouveau AWS compte. Lorsqu'une activité se produit dans Lake Formation, cette activité est enregistrée en tant qu' CloudTrail événement avec les autres événements de AWS service dans l'historique des événements. Un événement représente une demande émise par une source et comprend des informations sur l'action demandée, la date et l'heure de l'action, et les paramètres de la demande. En outre, chaque événement ou entrée de journal contient des informations sur l'auteur de la demande. Les informations relatives à l'identité permettent de déterminer les éléments suivants :

- Si la demande a été faite avec les informations d'identification de l'utilisateur root ou Gestion des identités et des accès AWS (IAM).
- Si la demande a été effectuée avec les informations d'identification de sécurité temporaires d'un rôle ou d'un utilisateur fédéré.
- Si la demande a été faite par un autre AWS service.

Pour de plus amples informations, veuillez consulter l'[élément userIdentity CloudTrail](#) .

Vous pouvez consulter, rechercher et télécharger les événements récents pour votre AWS compte. Pour plus d'informations, consultez la section [Affichage des événements avec l'historique des CloudTrail événements](#).

Pour un enregistrement continu des événements enregistrés sur votre AWS compte, y compris les événements liés à Lake Formation, créez un parcours. Un suivi permet CloudTrail de fournir des fichiers journaux à un compartiment Amazon S3. Par défaut, lorsque vous créez un journal de suivi dans la console, il s'applique à toutes les régions AWS. Le journal enregistre les événements de toutes les régions de la AWS partition et transmet les fichiers journaux au compartiment Amazon S3 que vous spécifiez. En outre, vous pouvez configurer d'autres AWS services, par exemple Amazon Athena pour analyser plus en profondeur et agir sur les données d'événements collectées dans CloudTrail les journaux. CloudTrail peut également envoyer des fichiers journaux à Amazon CloudWatch Logs and CloudWatch Events.

Pour plus d'informations, consultez les ressources suivantes :

- [Présentation de la création d'un journal de suivi](#)
- [CloudTrail services et intégrations pris en charge](#)
- [Configuration des notifications Amazon SNS pour CloudTrail](#)
- [Réception de fichiers CloudTrail journaux de plusieurs régions](#) et [réception de fichiers CloudTrail journaux de plusieurs comptes](#)

Comprendre les événements liés à la Lake Formation

Toutes les actions de l'API Lake Formation sont enregistrées CloudTrail et documentées dans le Guide du AWS Lake Formation développeur. Par exemple, les appels aux `PutDataLakeSettingsGrantPermissions`, et `RevokePermissions` les actions génèrent des entrées dans les fichiers CloudTrail journaux.

L'exemple suivant montre un CloudTrail événement associé à l'`GrantPermissions` action. L'entrée inclut l'utilisateur qui a accordé l'autorisation (`datalake_admin`), le principal auquel l'autorisation a été accordée (`datalake_user1`) et l'autorisation accordée (`CREATE_TABLE`). L'entrée indique également que l'autorisation a échoué car la base de données cible n'a pas été spécifiée dans l'`resource` argument.

```
{
  "eventVersion": "1.08",
  "userIdentity": {
```



```

    "type": "IAMUser",
    "principalId": "AIDAZKE67KM3P775X74U2",
    "arn": "arn:aws:iam::111122223333:user/datalake_admin",
    "accountId": "111122223333",
    "accessKeyId": "...",
    "userName": "datalake_admin"
  },
  "eventTime": "2021-02-06T00:43:21Z",
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GrantPermissions",
  "awsRegion": "us-east-1",
  "sourceIPAddress": "72.21.198.65",
  "userAgent": "aws-cli/1.19.0 Python/3.6.12
Linux/4.9.230-0.1.ac.223.84.332.metal1.x86_64 botocore/1.20.0",
  "errorCode": "InvalidInputException",
  "errorMessage": "Resource must have one of the have either the catalog, table or
database field populated.",
  "requestParameters": {
    "principal": {
      "dataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
    },
    "resource": {},
    "permissions": [
      "CREATE_TABLE"
    ]
  },
  "responseElements": null,
  "requestID": "b85e863f-e75d-4fc0-9ff0-97f943f706e7",
  "eventID": "8d2ccef0-55f3-42d3-9ede-3a6faedaa5c1",
  "readOnly": false,
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "eventCategory": "Management",
  "recipientAccountId": "111122223333"
}

```

L'exemple suivant montre une entrée de CloudTrail journal pour l'GetDataAccesssaction. Les directeurs n'appellent pas directement cette API. GetDataAccesssII est plutôt enregistré chaque fois qu'un AWS service principal ou intégré demande des informations d'identification temporaires pour accéder aux données d'un lac de données enregistré auprès de Lake Formation.

```
{
```

```
"eventVersion": "1.05",
"userIdentity": {
  "type": "AWSAccount",
  "principalId": "AROAQGFTBBBG0BWV2EMZA:GlueJobRunnerSession",
  "accountId": "111122223333"
},
"eventSource": "lakeformation.amazonaws.com",
"eventName": "GetDataAccess",
...
...
"additionalEventData": {
  "requesterService": "GLUE_JOB",
  "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
  "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
},
...
}
```

Tracer l'accès aux données CloudTrail grâce à Lake Formation

Lake Formation fournit des informations d'identification temporaires et limitées aux moteurs de requête, tels qu'Amazon EMR AWS Glue et, pour accéder à l'emplacement des données des tables dans Amazon S3. Lake Formation utilise le rôle IAM que vous spécifiez lorsque vous [enregistrez un emplacement de données](#) auprès de Lake Formation pour vendre ces informations d'identification.

Modification du comportement de CloudTrail journalisation (février 2026)

Avant février 2026, lorsqu'un moteur de requête `s3:GetObject` appelait pour récupérer les données de requête, les CloudTrail journaux indiquaient le rôle IAM d'enregistrement de Lake Formation en tant que principal.

À compter de février 2026, CloudTrail les journaux des `s3:GetObject` appels incluent également l'[identifiant unique IAM](#) (commençant par `AROA`) du rôle IAM à l'origine de la requête. Cet identifiant correspond au rôle IAM qui a lancé la requête, la tâche Amazon EMR AWS Glue ou la tâche ETL.

Suivi de l'accès aux données depuis le lancement de la requête vers Amazon S3

Chaque `s3:GetObject` appel entrant CloudTrail correspond à un `lakeformation:GetDataAccess` appel. Vous pouvez utiliser l'identifiant de rôle IAM présent dans les deux entrées du journal pour suivre le rôle IAM de l'appelant depuis le lancement de la requête jusqu'à l'accès aux données dans Amazon S3.

Pour activer les informations relatives à l'identité de la source dans les CloudTrail journaux des événements de données Amazon S3, suivez les étapes ci-dessous. Pour activer les événements de données Amazon S3 dans CloudTrail, consultez [Activation de la journalisation des CloudTrail événements pour les buckets et les objets Amazon S3](#).

 Important

L'activation des événements de données Amazon S3 CloudTrail entraîne des frais supplémentaires.

Pour activer le suivi de l'identité de la source dans CloudTrail les journaux

1. En utilisant, exécutez la `get-data-lake-settings` commande. Dans la réponse, modifiez le `Parameters` champ à ajouter `"SET_SOURCE_IDENTITY": "TRUE"`.

Exécutez la commande `put-data-lake-settings`.

```
"Parameters": {
  "CROSS_ACCOUNT_VERSION": "4",
  "SET_SOURCE_IDENTITY": "TRUE"
},
```

2. Dans la politique de confiance du rôle IAM utilisé pour enregistrer l'emplacement de vos données Amazon S3 auprès de Lake Formation, ajoutez `sts:SetSourceIdentity` :

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "lakeformation.amazonaws.com"
        ]
      },
      "Action": [
        "sts:AssumeRole",
        "sts:SetSourceIdentity"
      ]
    }
  ]
}
```

```
]
}
```

Le journal suivant présente les informations d'identité de la source dans un exemple de `s3:GetObject` CloudTrail journal.

```
"sessionContext": {
  "sessionIssuer": {
    "type": "Role",
    "principalId": "AROAZEXAMPLEXYZ3AB6CDKLM",
    "arn": "arn:aws:iam::111122223333:role/LFRegistrationRole",
    "accountId": "111122223333",
    "userName": "LFRegistrationRole"
  },
  "sourceIdentity": "V2.111122223333.AROAZEXAMPLEPQR7ST9UVWXY"
},
"invokedBy": "glue.amazonaws.com"
```

Voir aussi

- Pour plus d'informations sur le suivi du contexte utilisateur d'IAM Identity Center dans CloudTrail les journaux, consultez [Intégration d'IAM Identity Center](#).

Surveillance de l'accès aux catalogues fédérés Amazon Redshift

Les catalogues fédérés Amazon Redshift figurent dans les journaux « AWS Glue Data Catalog Ne génèrent pas d'événements dans les journaux de GetDataAccess Lake Formation ». CloudTrail Pour suivre l'accès aux données pour les tables Amazon Redshift dans le catalogue de données, le moniteur GetTable et les BatchGetTable événements, qui capturent les modèles d'accès aux métadonnées pour les tables fédérées Redshift, notamment le nom de la table, le nom de la base de données et les informations d'identification du catalogue.

Voir aussi

- [Journalisation entre comptes CloudTrail](#)

Meilleures pratiques, considérations et limites en matière de formation des lacs

Utilisez cette section pour trouver rapidement les meilleures pratiques, les considérations et les limites qu'elle contient AWS Lake Formation.

Consultez la section [Quotas de service](#) pour connaître le nombre maximal de ressources de service ou d'opérations pour votre Compte AWS.

Rubriques

- [Meilleures pratiques et considérations relatives au partage de données entre comptes](#)
- [Limitations des rôles liés aux services](#)
- [Limites d'accès aux données entre régions](#)
- [Considérations et limites relatives aux affichages du catalogue de données](#)
- [Limites du filtrage des données](#)
- [Considérations et limites relatives au mode d'accès hybride](#)
- [Limites liées à l'intégration des données de l'entrepôt de données Amazon Redshift dans le AWS Glue Data Catalog](#)
- [Limites d'intégration du catalogue S3 Tables](#)
- [Considérations et limites relatives au partage des données du magasin de métadonnées Hive](#)
- [Limites du partage de données Amazon Redshift](#)
- [Limites de l'intégration d'IAM Identity Center](#)
- [Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation](#)
- [Considérations relatives au contrôle d'accès basé sur les attributs, limites et régions prises en charge](#)

Meilleures pratiques et considérations relatives au partage de données entre comptes

Les fonctionnalités multi-comptes de Lake Formation permettent aux utilisateurs de partager en toute sécurité des lacs de données distribués entre plusieurs AWS organisations ou directement avec les

responsables IAM d'un autre compte Comptes AWS, offrant ainsi un accès détaillé aux métadonnées du catalogue de données et aux données sous-jacentes.

Tenez compte des meilleures pratiques suivantes lorsque vous utilisez le partage de données entre comptes de Lake Formation :

- Il n'y a pas de limite au nombre d'autorisations de Lake Formation que vous pouvez accorder aux directeurs d'école pour votre propre AWS compte. Cependant, Lake Formation utilise AWS Resource Access Manager (AWS RAM) la capacité pour les subventions entre comptes que votre compte peut accorder avec la méthode de ressource nommée. Pour optimiser la AWS RAM capacité, suivez les meilleures pratiques suivantes pour la méthode de ressource nommée :
 - Utilisez le nouveau mode de subvention entre comptes (version 3 et supérieure, sous Paramètres des versions entre comptes) pour partager une ressource avec un externe Compte AWS. Pour de plus amples informations, veuillez consulter [Mise à jour des paramètres de version de partage de données entre comptes](#).
 - Organisez AWS les comptes en organisations et accordez des autorisations à des organisations ou à des unités organisationnelles. Une subvention accordée à une organisation ou à une unité organisationnelle est considérée comme une subvention.

L'octroi à des organisations ou à des unités organisationnelles élimine également le besoin d'accepter une AWS Resource Access Manager (AWS RAM) invitation à partager des ressources pour la subvention. Pour de plus amples informations, veuillez consulter [Accès aux tables et aux bases de données partagées du catalogue de données et affichage](#).

- Au lieu d'accorder des autorisations sur de nombreuses tables individuelles d'une base de données, utilisez le caractère générique spécial Toutes les tables pour accorder des autorisations sur toutes les tables de la base de données. L'octroi d'une subvention sur toutes les tables est considéré comme une subvention unique. Pour de plus amples informations, veuillez consulter [Octroi d'autorisations sur les ressources du catalogue de données](#).

Note

Pour plus d'informations sur la demande d'une limite plus élevée pour le nombre de partages de ressources dans AWS RAM, voir les [quotas de AWS service](#) dans le Références générales AWS.

- Vous devez créer un lien de ressource vers une base de données partagée pour que cette base de données apparaisse dans les éditeurs de requêtes Amazon Redshift Spectrum Amazon Athena

et Amazon Redshift. De même, pour pouvoir interroger des tables partagées à l'aide d'Athena et Redshift Spectrum, vous devez créer des liens de ressources vers les tables. Les liens vers les ressources apparaissent ensuite dans la liste des tables des éditeurs de requêtes.

Au lieu de créer des liens de ressources pour de nombreuses tables individuelles à des fins d'interrogation, vous pouvez utiliser le caractère générique Toutes les tables pour accorder des autorisations sur toutes les tables d'une base de données. Ensuite, lorsque vous créez un lien de ressource pour cette base de données et que vous sélectionnez ce lien de ressource de base de données dans l'éditeur de requêtes, vous aurez accès à toutes les tables de cette base de données pour votre requête. Pour de plus amples informations, veuillez consulter [Création de liens vers des ressources](#).

- Lorsque vous partagez des ressources directement avec les principaux d'un autre compte, le principal IAM du compte destinataire n'est peut-être pas autorisé à créer des liens vers des ressources pour interroger les tables partagées à l'aide d'Athena et d'Amazon Redshift Spectrum. Au lieu de créer un lien de ressource pour chaque table partagée, l'administrateur du lac de données peut créer une base de données fictive et accorder des CREATE_TABLE autorisations au ALLIAMPrincipal groupe. Tous les principaux IAM du compte destinataire peuvent ensuite créer des liens de ressources dans la base de données d'espaces réservés et commencer à interroger les tables partagées.

Consultez l'exemple de commande CLI pour accorder des autorisations ALLIAMPrincipals d'entrée [Octroi d'autorisations de base de données via la méthode de ressource nommée](#).

- Lorsque des autorisations entre comptes sont accordées directement à un mandant, seul le bénéficiaire de la subvention peut consulter ces autorisations. L'administrateur du lac de données du AWS compte du bénéficiaire ne peut pas consulter ces subventions directes.
- Athena et Redshift Spectrum prennent en charge le contrôle d'accès au niveau des colonnes, mais uniquement pour l'inclusion, et non pour l'exclusion. Le contrôle d'accès au niveau des colonnes n'est pas pris en charge dans les tâches AWS Glue ETL.
- Lorsqu'une ressource est partagée avec votre AWS compte, vous pouvez accorder des autorisations sur cette ressource uniquement aux utilisateurs de votre compte. Vous ne pouvez pas accorder d'autorisations sur la ressource à d'autres AWS comptes, à des organisations (pas même à votre propre organisation) ou au IAMAllowedPrincipals groupe.
- Vous ne pouvez pas accorder DROP ou attribuer une Super base de données à un compte externe.

- Révoquez les autorisations entre comptes avant de supprimer une base de données ou une table. Sinon, vous devez supprimer les partages de ressources orphelins dans AWS Resource Access Manager.

 Consultez aussi

- [Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation](#)
- [CREATE_TABLE](#) dans le [Référence des autorisations de Lake Formation](#) pour plus de règles et de limitations d'accès entre comptes.

Limitations des rôles liés aux services

Un rôle lié à un service est un type spécial de rôle IAM directement lié à AWS Lake Formation. Ce rôle possède des autorisations prédéfinies qui permettent à Lake Formation d'effectuer des actions en votre nom sur l'ensemble AWS des services.

Les limites suivantes s'appliquent lors de l'utilisation d'un rôle lié à un service (SLR) pour enregistrer des emplacements de données auprès de Lake Formation.

- Vous ne pouvez pas modifier les politiques de rôle liées aux services une fois créées.
- Un rôle lié à un service ne prend pas en charge le partage crypté des ressources de catalogue entre les comptes. Les ressources chiffrées nécessitent des autorisations AWS KMS clés spécifiques. Les rôles liés à un service sont dotés d'autorisations prédéfinies qui n'incluent pas la possibilité d'utiliser des ressources de catalogue chiffrées sur plusieurs comptes.
- Lorsque vous enregistrez plusieurs sites Amazon S3, l'utilisation d'un rôle lié à un service peut vous amener à dépasser rapidement les limites de votre politique IAM. Cela se produit parce qu'avec les rôles liés à un service, elle AWS écrit la politique pour vous, et elle s'incrémente sous la forme d'un gros bloc qui inclut toutes vos inscriptions. Vous pouvez rédiger des politiques gérées par le client de manière plus efficace, répartir les autorisations entre plusieurs politiques ou utiliser différents rôles pour différentes régions.
- Amazon EMR sur EC2 ne peut pas accéder aux données. Vous enregistrez des emplacements de données avec des rôles liés à un service.
- Les opérations de rôle liées au service contournent vos politiques de contrôle des AWS services.

- Lorsque vous enregistrez des emplacements de données avec un rôle lié à un service, les politiques IAM sont mises à jour de manière cohérente. Pour plus d'informations, consultez la documentation relative à la [résolution des problèmes liés à l'IAM](#) dans le guide de l'utilisateur d'IAM.
- Vous ne pouvez pas définir `SET_CONTEXT = TRUE` les paramètres du lac de données de Lake Formation lorsque vous utilisez des rôles liés à un service, et vous utilisez IAM Identity Center. La raison en est que les rôles liés à un service sont soumis à des politiques de confiance immuables incompatibles avec la propagation d'identité fiable nécessaire à l'`SetContextaudit` auprès des principaux IAM Identity Center.

Limites d'accès aux données entre régions

Lake Formation permet d'interroger les tables du catalogue de données. Régions AWS Vous pouvez accéder aux données d'une région depuis d'autres régions à l'aide Amazon Athena d'Amazon EMR et d' AWS Glue ETL en créant des liens de ressources dans d'autres régions pointant vers les bases de données et les tables sources. Grâce à l'accès aux tables entre régions, vous pouvez accéder aux données entre les régions sans copier les données sous-jacentes ou les métadonnées dans le catalogue de données.

Les restrictions suivantes s'appliquent à l'accès aux tables entre régions.

- Lake Formation ne permet pas d'interroger les tables du catalogue de données d'une autre région à l'aide d'Amazon Redshift Spectrum.
- Dans la console Lake Formation, les vues de base de données et de table n'affichent pas les noms des bases de données/tables de la région source.
- Pour afficher la liste des tables d'une base de données partagée d'une autre région, vous devez d'abord créer un lien de ressource vers la base de données partagée, puis sélectionner le lien de ressource et choisir Afficher les tables.
- Lake Formation ne prend pas en charge les appels de liens de ressources interrégionaux effectués par les utilisateurs de SAML.
- La fonctionnalité interrégionale de Lake Formation n'implique pas de frais supplémentaires pour les transferts de données.

Considérations et limites relatives aux affichages du catalogue de données

Les considérations et limites suivantes s'appliquent aux vues du catalogue de données.

- Vous ne pouvez pas créer une vue du catalogue de données à partir de la console Lake Formation. Vous pouvez créer des vues à l'aide du SDK AWS CLI ou.
- Vous pouvez créer des vues de catalogue de données à partir de 10 tables. C'est une limite stricte. Les tables de référence sous-jacentes d'une vue peuvent appartenir à la même base de données ou à différentes bases de données au sein du même AWS compte.
- Pour d'autres considérations et limitations spécifiques à la création de vues de catalogue de données à l'aide de Redshift, consultez la section [Considérations et limites relatives aux vues de catalogue de données](#) dans le guide du développeur de base de données Amazon Redshift. Pour Athena, consultez la section [Considérations et limites relatives aux vues du catalogue de données](#) dans le guide de l'utilisateur d'Amazon Athena.
- Vous pouvez créer des vues de catalogue de données sur des tables enregistrées auprès de Lake Formation à la fois en mode d'accès hybride et en mode Lake Formation.

Lorsque vous utilisez des vues du catalogue de données avec le mode d'accès hybride Lake Formation, il est recommandé de s'assurer que les principaux consommateurs de vues sont autorisés à accéder aux autorisations Lake Formation pour les tables de base référencées dans la vue sans accorder d'accès. Cela garantit que les tables de base ne seront pas révélées aux consommateurs via les autorisations AWS Glue IAM.

- Il n'existe aucune restriction quant à la version de partage entre comptes pour partager des vues.
- Les vues sont versionnées comme les tables du catalogue de données, lorsque vous utilisez l'`ALTER VIEW` instruction pour un dialecte de vue déjà créé. Vous ne pouvez pas revenir à une vue précédente car la version de la vue change en fonction des modifications des données sous-jacentes. Vous pouvez supprimer une version d'affichage et elle utilisera par défaut la version la plus récente disponible. Lorsque vous modifiez la version de vue, assurez-vous que vos données sont synchronisées avec le schéma de version de vue sélectionné.
- Aucun nouveau catalogue de données APIs n'est introduit. Les existants `CreateTableUpdateTable`, `DeleteTable` et `GetTable` APIs sont mis à jour.
- Amazon Redshift crée toujours des vues avec des colonnes varchar à partir de tables contenant des chaînes. Vous devez convertir les colonnes de chaîne en varchar avec une longueur explicite lorsque vous ajoutez des dialectes provenant d'autres moteurs.

- L'octroi d'autorisations de lac de données au All tables sein d'une base de données permettra au bénéficiaire de disposer d'autorisations sur toutes les tables et vues de la base de données.
- Vous ne pouvez pas créer de vues :
 - Cela fait référence à d'autres points de vue.
 - Lorsque la table de référence est un lien vers une ressource.
 - Lorsque la table de référence se trouve dans un autre compte.
 - À partir de métastores Hive externes.
- Les rôles de définition entre comptes ne sont pas pris en charge pour les vues Redshift Spectrum Dialect.
- Les liens de ressources pour le dialecte Athena dans l'éditeur de requêtes Athena ne sont pas pris en charge. Pour utiliser des rôles de définition entre comptes pour le dialecte Athena, ajoutez le compte hébergeant les tables de base en tant que source de données dans Athena.

Limites du filtrage des données

Lorsque vous accordez des autorisations à Lake Formation sur une table du catalogue de données, vous pouvez inclure des spécifications de filtrage des données afin de restreindre l'accès à certaines données dans les résultats des requêtes et les moteurs intégrés à Lake Formation. Lake Formation utilise le filtrage des données pour garantir la sécurité au niveau des colonnes, au niveau des lignes et au niveau des cellules. Vous pouvez définir et appliquer des filtres de données sur des colonnes imbriquées si vos données sources contiennent des structures imbriquées.

Remarques et restrictions relatives au filtrage au niveau des colonnes

Il existe trois méthodes pour définir le filtrage des colonnes :

- En utilisant des filtres de données
- En utilisant un filtrage par colonne simple ou un filtrage par colonne imbriqué.
- En utilisant TAGs.

Le filtrage simple des colonnes indique simplement une liste de colonnes à inclure ou à exclure. La console Lake Formation, l'API et l'API prennent en AWS CLI charge un filtrage simple par colonne. Pour obtenir un exemple, consultez [Grant with Simple Column Filtering](#).

Les remarques et restrictions suivantes s'appliquent au filtrage des colonnes :

- AWS Glue La version 5.0 ou supérieure prend en charge le contrôle d'accès précis via Lake Formation uniquement pour les tables Apache Hive et Apache Iceberg.
- Pour octroyer SELECT avec l'option d'autorisation et le filtrage des colonnes, vous devez utiliser une liste d'inclusion et non une liste d'exclusion. Sans l'option d'autorisation, vous pouvez utiliser des listes d'inclusion ou d'exclusion.
- Pour accorder une autorisation SELECT sur une table avec filtrage par colonne, vous devez avoir été autorisé SELECT sur la table avec l'option d'attribution et sans aucune restriction de ligne. Vous devez avoir accès à toutes les lignes.
- Si vous accordez une subvention SELECT avec l'option d'attribution et le filtrage des colonnes à un principal de votre compte, ce principal doit spécifier le filtrage des colonnes pour les mêmes colonnes ou pour un sous-ensemble des colonnes accordées lorsqu'il octroie à un autre principal. Si vous accordez SELECT à un compte externe à l'aide de l'option d'attribution et du filtrage des colonnes, l'administrateur du lac de données du compte externe peut accorder une autorisation SELECT sur toutes les colonnes à un autre principal de son compte. Cependant, même dans toutes SELECT les colonnes, ce principal n'aura de visibilité que sur les colonnes accordées au compte externe.
- Vous ne pouvez pas appliquer le filtrage des colonnes aux clés de partition.
- Un principal SELECT autorisé autorisé à accéder à un sous-ensemble de colonnes d'une table ne peut pas obtenir l'INSERTautorisation ALTERDROP,DELETE, ou sur cette table. Pour un directeur disposant de l'INSERTautorisation ALTERDROP,DELETE, ou sur une table, si vous accordez l'SELECTautorisation avec filtrage par colonne, cela n'a aucun effet.

Les remarques et restrictions suivantes s'appliquent au filtrage des colonnes imbriquées :

- Vous pouvez inclure ou exclure cinq niveaux de champs imbriqués dans un filtre de données.

Exemple

Col1.Col1_1.Col1_1_1.Col1_1_1_1.Col1_1_1_1_1

- Vous ne pouvez pas appliquer de filtrage de colonnes à des champs imbriqués dans des colonnes de partition.
- Si le schéma de votre table contient un nom de colonne de premier niveau (« client »). » adresse ») qui a le même modèle de représentation de champs imbriqués dans un filtre de données (une colonne imbriquée avec un nom de colonne de premier niveau customer et un nom de champ imbriqué address est spécifiée comme "customer". "address" dans un filtre de données), vous ne pouvez pas spécifier explicitement l'accès à une colonne de niveau supérieur ou à un

champ imbriqué car les deux sont représentés selon le même modèle dans les listes. inclusion/exclusion Cela est ambigu, et Lake Formation ne peut pas être résolu si vous spécifiez la colonne de niveau supérieur ou le champ imbriqué.

- Si le nom d'une colonne ou d'un champ imbriqué de niveau supérieur contient un guillemet double, vous devez inclure un deuxième guillemet double lorsque vous spécifiez l'accès à un champ imbriqué dans la liste d'inclusion et d'exclusion d'un filtre de cellules de données.

Exemple

Exemple de nom de colonne imbriqué entre guillemets : `a.b.double"quote`

Exemple

Exemple de représentation de colonnes imbriquées dans un filtre de données :
`"a"."b"."double""quote"`

Limites du filtrage au niveau des cellules

Tenez compte des remarques et restrictions suivantes concernant le filtrage au niveau des lignes et au niveau des cellules.

- La sécurité au niveau des cellules n'est pas prise en charge sur les colonnes, les vues et les liens de ressources imbriqués.
- Toutes les expressions prises en charge sur les colonnes de niveau supérieur sont également prises en charge sur les colonnes imbriquées. Cependant, les champs imbriqués sous les colonnes de partition ne doivent PAS être référencés lors de la définition d'expressions imbriquées au niveau des lignes.
- La sécurité au niveau des cellules est disponible dans toutes les régions lorsque vous utilisez le moteur Athena version 3 ou Amazon Redshift Spectrum. Pour les autres services, la sécurité au niveau des cellules n'est disponible que dans les régions mentionnées sur le [Régions prises en charge](#)
- Les instructions `SELECT INTO` ne sont pas prises en charge.
- Les types `array` de map données et ne sont pas pris en charge dans les expressions de filtre de ligne. Le type `struct` de données est pris en charge.
- Il n'y a pas de limite au nombre de filtres de données pouvant être définis sur une table, mais il existe une limite de 100 filtres de données pour un seul principal sur une table.

- Pour appliquer un filtre de données avec une expression de filtre de ligne, vous devez avoir SELECT l'option grant sur toutes les colonnes du tableau. Cette restriction ne s'applique pas aux administrateurs des comptes externes lorsque la subvention a été accordée au compte externe.
- Si un directeur est membre d'un groupe et que le principal et le groupe reçoivent des autorisations sur un sous-ensemble de lignes, les autorisations de ligne effectives du principal sont l'union des autorisations du principal et des autorisations du groupe.
- Les noms de colonnes suivants sont restreints dans un tableau pour le filtrage au niveau des lignes et au niveau des cellules :
 - ctid
 - oid
 - xmin
 - cmin
 - xmax
 - cmax
 - tabloïd
 - insérez un identifiant
 - supprimer exid
 - importoïde
 - identifiant unique du chat rouge
- Si vous appliquez l'expression de filtre toutes les lignes sur un tableau en même temps que d'autres expressions de filtre contenant des prédicats, l'expression de toutes les lignes prévaudra sur toutes les autres expressions de filtre.
- Lorsque des autorisations sur un sous-ensemble de lignes sont accordées à un AWS compte externe et que l'administrateur du lac de données du compte externe accorde ces autorisations au principal de ce compte, le prédicat de filtre effectif du principal est l'intersection du prédicat du compte et de tout prédicat directement accordé au principal.

Par exemple, si le compte dispose d'autorisations de ligne avec le prédicat dept= 'hr' et que le principal a été autorisé séparément pour country= 'us', le principal n'a accès qu'aux lignes avec dept= 'hr' et country= 'us'.

Pour plus d'informations sur le filtrage au niveau des cellules, consultez. [Filtrage des données et sécurité au niveau des cellules dans Lake Formation](#)

Pour connaître les points à prendre en compte et les limites lors de l'interrogation de tables à l'aide d'Amazon Redshift Spectrum avec des politiques de sécurité au niveau des lignes, consultez la section [Considérations et limites relatives à l'utilisation des politiques RLS dans le](#) manuel Amazon Redshift Database Developer Guide.

Considérations et limites relatives au mode d'accès hybride

Le mode d'accès hybride offre la flexibilité d'activer de manière sélective les autorisations de Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog.

Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Les considérations et limitations suivantes s'appliquent au mode d'accès hybride.

Limitations

- Mettre à jour l'enregistrement d'une position Amazon S3 : vous ne pouvez pas modifier les paramètres d'une position enregistrée auprès de Lake Formation à l'aide d'un rôle lié à un service.
- Option d'activation lors de l'utilisation des balises LF — Lorsque vous pouvez accorder des autorisations à Lake Formation à l'aide de balises LF, vous pouvez choisir des principes pour appliquer les autorisations de Lake Formation par étapes successives en choisissant des bases de données et des tables auxquelles des balises LF sont attachées.
- Accès en mode d'accès hybride — L'accès au mode d'accès hybride dans Lake Formation est limité aux utilisateurs disposant d'autorisations d'administrateur de lac de données ou d'administrateur en lecture seule.
- Principaux d'adhésion — Actuellement, seul un rôle d'administrateur de lac de données peut attribuer des principes aux ressources.
- Activer toutes les tables d'une base de données : dans le cas des autorisations entre comptes, lorsque vous accordez des autorisations et que vous activez toutes les tables d'une base de données, vous devez également activer la base de données pour que les autorisations fonctionnent.

Considérations

- Mise à jour de l'emplacement Amazon S3 enregistré auprès de Lake Formation en mode d'accès hybride — Nous déconseillons de convertir un emplacement de données Amazon S3 déjà enregistré auprès de Lake Formation en mode d'accès hybride, bien que cela soit possible.
- Comportements de l'API lorsqu'un emplacement de données est enregistré en mode d'accès hybride
 - CreateTable — L'emplacement est considéré comme enregistré auprès de Lake Formation, quels que soient le drapeau du mode d'accès hybride et le statut d'inscription. L'utilisateur a donc besoin de l'autorisation de localisation des données pour créer une table.
 - CreatePartition/BatchCreatePartitions/UpdatePartitions(lorsque l'emplacement de la partition est mis à jour pour pointer vers l'emplacement enregistré auprès de l'hybride) — L'emplacement Amazon S3 est considéré comme enregistré auprès de Lake Formation, quels que soient l'indicateur du mode d'accès hybride et le statut d'option d'inscription. L'utilisateur a donc besoin de l'autorisation de localisation des données pour créer ou mettre à jour une base de données.
 - CreateDatabase/UpdateDatabase (lorsque l'emplacement de la base de données est mis à jour pour pointer vers l'emplacement enregistré en mode d'accès hybride) — L'emplacement est considéré comme enregistré auprès de Lake Formation, quels que soient le drapeau du mode d'accès hybride et le statut d'option d'inscription. L'utilisateur a donc besoin de l'autorisation de localisation des données pour créer ou mettre à jour une base de données.
 - UpdateTable (lorsqu'un emplacement de table est mis à jour pour pointer vers l'emplacement enregistré en mode d'accès hybride) — L'emplacement est considéré comme enregistré auprès de Lake Formation, quels que soient le drapeau du mode d'accès hybride et le statut d'option d'inscription. L'utilisateur a donc besoin d'une autorisation de localisation des données pour mettre à jour le tableau. Si l'emplacement de la table n'est pas mis à jour ou s'il pointe vers un emplacement qui n'est pas enregistré auprès de Lake Formation, l'utilisateur n'a pas besoin d'autorisation de localisation des données pour mettre à jour la table.

Limites liées à l'intégration des données de l'entrepôt de données Amazon Redshift dans le AWS Glue Data Catalog

Vous pouvez cataloguer et gérer l'accès aux données analytiques dans les entrepôts de données Amazon Redshift à l'aide du. AWS Glue Data Catalog Les limites suivantes s'appliquent :

- Le partage entre comptes n'est pas pris en charge au niveau du catalogue fédéré. Cependant, vous pouvez partager des bases de données et des tables individuelles à partir d'un catalogue fédéré entre Compte AWS nous.
- Vous devez disposer des paramètres de version inter-comptes version 4 ou supérieure pour partager des bases de données ou des tables dans le catalogue fédéré entre Compte AWS s.
- Le catalogue de données prend uniquement en charge la création de catalogues de haut niveau.
- Vous pouvez uniquement mettre à jour la description des catalogues dans le Redshift Managed Storage (RMS).
- La configuration d'autorisations sur les catalogues fédérés ainsi que sur les bases de données et les tables du catalogue fédéré à IAMAllowedPrincipals regrouper n'est pas prise en charge.
- Les opérations DDL (Data Definition Language) sur le catalogue à partir de moteurs tels qu'Athena, Amazon EMR Spark ou d'autres, y compris la définition des configurations du catalogue, ne sont pas prises en charge.
- L'exécution d'opérations DDL sur des tables RMS à l'aide d'Athena n'est pas prise en charge.
- La création de vues matérialisées n'est pas prise en charge, que ce soit via Athena, Apache Spark, AWS Glue Data Catalog le client Amazon Redshift ou Amazon Redshift.
- Athena ne prend pas en charge une expérience multi-catalogues. Il ne peut se connecter qu'à un seul catalogue spécifique à la fois. Athena ne peut pas accéder à plusieurs catalogues ou les interroger simultanément.
- Les opérations de balisage et de branchement sur les tables Iceberg via Athena et Amazon Redshift ne sont pas prises en charge.
- Le voyage dans le temps sur les tables RMS n'est pas pris en charge.
- Les catalogues multiniveaux avec des tables de lacs de données ne sont pas pris en charge. Toutes les données stockées dans Amazon S3 destinées à être utilisées avec les tables de lacs de données doivent résider dans la version par défaut AWS Glue Data Catalog et ne peuvent pas être organisées dans des catalogues à plusieurs niveaux.
- Dans Amazon Redshift, les partages de données ne sont pas ajoutés à l'espace de noms enregistré. Les clusters et les espaces de noms sont synonymes. Une fois que vous avez publié un cluster sur le AWS Glue Data Catalog, vous ne pouvez pas ajouter de nouvelles données.
- Amazon EMR sur EC2 ne prend pas en charge la jonction entre les tables RMS et les tables Amazon S3. Seul EMR Serverless prend en charge cette fonctionnalité.
- Les schémas et tables externes ne sont pas pris en charge.

- Les tables RMS ne sont accessibles que depuis le point de terminaison de l'extension dans le catalogue REST d' AWS Glue Iceberg.
- Les tables Hive ne sont pas accessibles depuis des moteurs tiers connectés au catalogue REST d' AWS Glue Iceberg.
- Le niveau d'isolation `read_committed` sur les tables RMS via Spark sera pris en charge.
- Les noms des bases de données Redshift sont traités sans distinction majuscules/minuscules AWS Glue Data Catalog, limités à 128 caractères, et peuvent être alphanumériques avec des tirets (-) et des traits de soulignement (_).
- Les noms de catalogue ne distinguent pas les majuscules des minuscules, sont limités à 50 caractères et peuvent être alphanumériques avec des tirets (-) et des traits de soulignement (_).
- Amazon Redshift ne prend pas en charge l'utilisation des commandes GRANT et REVOKE de style Lake Formation SQL pour gérer les autorisations d'accès sur les tables publiées sur le AWS Glue Data Catalog
- Les politiques de sécurité au niveau des lignes et de masquage dynamique des données associées au cluster Amazon Redshift producteur (source) ne seront pas appliquées. Au lieu de cela, les autorisations d'accès définies dans Lake Formation seront appliquées aux données partagées.
- L'exécution d'opérations en langage de définition de données (DDL) et en langage de manipulation de données (DML) sur les liens de table n'est pas prise en charge.
- Si les mots clés réservés ne sont pas correctement échappés, cela entraînera des échecs ou des erreurs.
- Le chiffrement des données dans les scénarios à catalogues multiples n'est pas pris en charge.

Limites d'intégration du catalogue S3 Tables

Amazon S3 Tables s'intègre à AWS Glue Data Catalog (Data Catalog) et enregistre le catalogue en tant que localisation de données de Lake Formation. Vous pouvez configurer cet enregistrement depuis la console Lake Formation ou en utilisant le service APIs.

Note

Si vous avez une politique basée sur les ressources IAM ou S3 Tables qui restreint les utilisateurs IAM et les rôles IAM en fonction des balises principales, vous devez associer les mêmes balises principales au rôle IAM que Lake Formation utilise pour accéder à vos données S3 (par exemple `LakeFormationDataAccessRole`,) et accorder à ce rôle les

autorisations nécessaires. Cela est nécessaire pour que votre politique de contrôle d'accès basée sur des balises fonctionne correctement avec votre intégration d'analyse S3 Tables.

Les limites suivantes s'appliquent à l'intégration du catalogue de tables S3 au catalogue de données et à Lake Formation :

- AWS Glue et Lake Formation ne prennent pas en charge les noms de colonnes mixtes et convertissent tous les noms de colonnes en minuscules. Vous devez vérifier que les noms des colonnes des tables sont uniques lorsqu'ils sont convertis en minuscules. Utiliser `customer_id` au lieu de `customerId`. L'utilisation de noms de colonnes en majuscules et minuscules n'était prise en charge que lors de la version préliminaire.
- L' `CreateCatalog` API ne peut pas créer de compartiments de tables dans Amazon S3.
- L' `SearchTables` API ne peut pas effectuer de recherche dans les tables S3.

Considérations et limites relatives au partage des données du magasin de métadonnées Hive

Grâce à la fédération des AWS Glue Data Catalog métadonnées (fédération du catalogue de données), vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation

Les considérations et limitations suivantes s'appliquent aux bases de données fédérées créées à partir de bases de données Hive :

Considérations

- AWS SAM support des applications : vous êtes responsable de la disponibilité des ressources de l'application AWS SAM déployée (Amazon API Gateway et de la fonction Lambda). Assurez-vous que la connexion entre le métastore AWS Glue Data Catalog et le métastore Hive fonctionne lorsque les utilisateurs exécutent des requêtes.
- Exigence de version du métastore Hive — Vous ne pouvez créer des bases de données fédérées qu'à l'aide d'Apache Hive version 3 ou supérieure.
- Exigence de base de données mappée — Chaque base de données Hive doit être mappée à une nouvelle base de données dans Lake Formation.

- Support de fédération au niveau de la base de données : vous pouvez vous connecter au métastore Hive uniquement au niveau de la base de données.
- Autorisations sur les bases de données fédérées : les autorisations appliquées à une base de données fédérée ou aux tables d'une base de données fédérée sont conservées même lorsqu'une table source ou une base de données est supprimée. Lorsque la base de données ou la table source est recréée, il n'est pas nécessaire de réoctroyer les autorisations. Lorsqu'une table fédérée dotée d'autorisations Lake Formation est supprimée à la source, les autorisations Lake Formation sont toujours visibles et vous pouvez les révoquer si nécessaire.

Si un utilisateur supprime une base de données fédérée, toutes les autorisations correspondantes sont perdues. La recréation de la même base de données portant le même nom ne permet pas de récupérer les autorisations de Lake Formation. Les utilisateurs devront à nouveau configurer de nouvelles autorisations.

- IAMAllowedAutorisations de groupe principales sur les bases de données fédérées — Sur la base de `celaDataLakeSettings`, Lake Formation peut attribuer des autorisations à toutes les bases de données et tables à un groupe virtuel nommé `IAMAllowedPrincipal`. Il `IAMAllowedPrincipal` fait référence à tous les principaux IAM qui ont accès aux ressources du catalogue de données par le biais des politiques principales IAM et AWS Glue des politiques de ressources. Si ces autorisations existent sur une base de données ou une table, tous les principaux ont accès à la base de données ou à la table.

Cependant, Lake Formation n'autorise pas `IAMAllowedPrincipal` les autorisations sur les tables des bases de données fédérées. Lorsque vous créez des bases de données fédérées, assurez-vous de transmettre le `CreateTableDefaultPermissions` paramètre sous forme de liste vide.

Pour de plus amples informations, veuillez consulter [Modification des paramètres par défaut de votre lac de données](#).

- Joindre des tables dans des requêtes : vous pouvez joindre des tables de métastore Hive à des tables natives de Data Catalog pour exécuter des requêtes.

Limitations

- Limitation de synchronisation des métadonnées entre le métastore AWS Glue Data Catalog et le métastore Hive — Après avoir établi la connexion au métastore Hive, vous devez créer une base de données fédérée pour synchroniser les métadonnées du métastore Hive avec le. AWS

Glue Data Catalog Les tables de la base de données fédérée sont synchronisées au moment de l'exécution lorsque les utilisateurs exécutent des requêtes.

- Limitation relative à la création de nouvelles tables dans une base de données fédérée : vous ne pourrez pas créer de nouvelles tables dans des bases de données fédérées.
- Limitation des autorisations relatives aux données : le support pour les autorisations sur les vues tabulaires du métastore Hive n'est pas disponible.

Limites du partage de données Amazon Redshift

AWS Lake Formation vous permet de gérer en toute sécurité les données d'un partage de données d'Amazon Redshift. Amazon Redshift est un service d'entrepôt de données entièrement géré de plusieurs pétaoctets dans le cloud. AWS Grâce à la fonctionnalité de partage de données, Amazon Redshift vous permet de partager des données entre différentes entités. Comptes AWS Pour plus d'informations sur le partage de données Amazon Redshift, consultez [Présentation du partage de données dans Amazon Redshift](#).

Les remarques et restrictions suivantes s'appliquent aux bases de données fédérées créées à partir de partages de données Amazon Redshift :

- Exigence de base de données mappée — Chaque partage de données Amazon Redshift doit être mappé vers une nouvelle base de données dans Lake Formation. Cela est nécessaire pour conserver des noms de table uniques lorsque la représentation des objets de partage de données est aplatie dans la base de données du catalogue de données.
- Limitation relative à la création de nouvelles tables dans une base de données fédérée : vous ne pourrez pas créer de nouvelles tables dans des bases de données fédérées.
- Autorisations sur les bases de données fédérées : les autorisations appliquées à une base de données fédérée ou à des tables d'une base de données fédérée sont conservées même lorsqu'une table source ou une base de données est supprimée. Lorsque la base de données ou la table source est recréée, il n'est pas nécessaire de réoctroyer les autorisations. Lorsqu'une table fédérée avec des autorisations Lake Formation est supprimée à la source, les autorisations Lake Formation seront toujours visibles et vous pouvez les révoquer si nécessaire.

Si un utilisateur supprime une base de données fédérée, toutes les autorisations correspondantes sont perdues. La recréation de la même base de données portant le même nom ne permet pas de récupérer les autorisations de Lake Formation. Les utilisateurs devront à nouveau configurer de nouvelles autorisations.

- `IAMAllowedAutorisations` de groupe principales sur les bases de données fédérées — Sur la base de `celaDataLakeSettings`, Lake Formation peut attribuer des autorisations à toutes les bases de données et tables à un groupe virtuel nommé `IAMAllowedPrincipal`. Il `IAMAllowedPrincipal` fait référence à tous les principaux IAM qui ont accès aux ressources du catalogue de données par le biais des politiques principales IAM et AWS Glue des politiques de ressources. Si ces autorisations existent sur une base de données ou une table, tous les principaux ont accès à la base de données ou à la table.

Cependant, Lake Formation n'autorise pas `IAMAllowedPrincipal` les autorisations sur les tables des bases de données fédérées. Lorsque vous créez des bases de données fédérées, assurez-vous de transmettre le `CreateTableDefaultPermissions` paramètre sous forme de liste vide.

Pour de plus amples informations, veuillez consulter [Modification des paramètres par défaut de votre lac de données](#).

- Filtrage des données : dans Lake Formation, vous pouvez accorder des autorisations sur une table d'une base de données fédérée avec un filtrage au niveau des colonnes et au niveau des lignes. Toutefois, vous ne pouvez pas combiner le filtrage au niveau des colonnes et au niveau des lignes pour restreindre l'accès au niveau de la granularité au niveau des cellules aux tables des bases de données fédérées.
- Identifiant distinguant majuscules et minuscules : les objets de partage de données Amazon Redshift gérés par Lake Formation prennent en charge les noms de tables et de colonnes uniquement en minuscules. N'activez pas l'identifiant majuscules/minuscules pour les bases de données, les tables et les colonnes dans les partages de données Amazon Redshift, s'ils doivent être partagés et gérés à l'aide de Lake Formation.
- Assistance aux requêtes — Vous pouvez interroger les partages de données Amazon Redshift gérés par Lake Formation avec Amazon Redshift. Athena ne prend pas en charge l'interrogation des partages de données Amazon Redshift gérés par Lake Formation.

Pour plus d'informations sur les limites liées à l'utilisation de partages de données dans Amazon Redshift, [consultez la section Limitations relatives au partage de données dans le manuel](#) Amazon Redshift Database Developer Guide.

Limites de l'intégration d'IAM Identity Center

Vous pouvez ainsi vous connecter à des fournisseurs d'identité (IdPs) et gérer de manière centralisée l'accès des utilisateurs et des groupes à travers les services AWS d'analyse. AWS IAM Identity Center Vous pouvez AWS Lake Formation le configurer en tant qu'application activée dans IAM Identity Center, et les administrateurs de data lake peuvent accorder des autorisations détaillées aux utilisateurs et aux groupes autorisés sur les ressources. AWS Glue Data Catalog

Les limites suivantes s'appliquent à l'intégration de Lake Formation à IAM Identity Center :

- Vous ne pouvez pas affecter des utilisateurs et des groupes IAM Identity Center en tant qu'administrateurs de data lake ou administrateurs en lecture seule dans Lake Formation.

Les utilisateurs et les groupes IAM Identity Center peuvent interroger les ressources chiffrées du catalogue de données si vous utilisez un rôle IAM qui AWS Glue peut assumer en votre nom le chiffrement et le déchiffrement du catalogue de données. AWS les clés gérées ne prennent pas en charge la propagation d'identités fiables.

- Les utilisateurs et les groupes IAM Identity Center peuvent uniquement invoquer les opérations d'API répertoriées dans la `AWSIAMIdentityCenterAllowListForIdentityContext` politique fournie par IAM Identity Center.
- Lake Formation permet aux rôles IAM issus de comptes externes d'agir en tant que rôles de support au nom des utilisateurs et des groupes d'IAM Identity Center pour accéder aux ressources du catalogue de données, mais les autorisations ne peuvent être accordées que sur les ressources du catalogue de données du compte propriétaire. Si vous essayez d'accorder des autorisations aux utilisateurs et aux groupes d'IAM Identity Center sur les ressources du catalogue de données d'un compte externe, Lake Formation génère le message d'erreur suivant : « Les autorisations entre comptes ne sont pas prises en charge pour le principal ».
- Lorsque vous utilisez Lake Formation avec IAM Identity Center, la configuration d'attribution des applications est définie `false` par défaut. Si vous modifiez cette configuration directement via l'[API IAM Identity Center](#), vous devez ensuite gérer manuellement toutes les attributions d'applications à l'aide de l'API. Lake Formation ne synchronise ni ne gère automatiquement les modifications d'affectation effectuées en dehors de ses flux de travail standard, ce qui peut avoir un impact sur les modèles d'accès et les flux d'autorisation au sein de votre environnement de lac de données.

Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation

Vous pouvez créer, gérer et attribuer des balises LF pour contrôler l'accès aux bases de données, aux tables et aux colonnes du catalogue de données.

Tenez compte des meilleures pratiques suivantes lorsque vous utilisez le contrôle d'accès basé sur des balises Lake Formation :

- Toutes les balises LF doivent être prédéfinies avant de pouvoir être attribuées aux ressources du catalogue de données ou accordées aux principaux.

L'administrateur du lac de données peut déléguer les tâches de gestion des balises en créant des balises LF avec les autorisations IAM requises. Les ingénieurs de données et les analystes décident des caractéristiques et des relations des balises LF. Les créateurs de balises LF créent et maintiennent ensuite les balises LF dans Lake Formation.

- Vous pouvez attribuer plusieurs balises LF aux ressources du catalogue de données. Une seule valeur pour une clé donnée peut être affectée à une ressource donnée.

Par exemple, vous pouvez attribuer `module=Orders,region=West,division=Consumer`, etc. à une base de données, une table ou une colonne. Vous ne pouvez pas attribuer `module=Orders,Customers`.

- Vous ne pouvez pas attribuer de balises LF aux ressources lorsque vous les créez. Vous ne pouvez ajouter des balises LF qu'aux ressources existantes.
- Vous pouvez accorder des expressions de balises LF, et pas simplement des balises LF uniques, à un principal.

Une expression LF-Tag ressemble à ce qui suit (en pseudo-code).

```
module=sales AND division=(consumer OR commercial)
```

Un principal auquel cette expression LF-Tag est attribuée ne peut accéder qu'aux ressources du catalogue de données (bases de données, tables et colonnes) attribuées `module=sales` et soit `division=consumer`, `division=commercial`. Si vous souhaitez que le directeur puisse accéder à des ressources qui incluent `module=sales` ou `division=commercial` non les deux dans la même subvention. Faites deux subventions, une pour `module=sales` et une pour `division=commercial`.

L'expression LF-Tag la plus simple consiste en une seule balise LF, telle que `module=sales`

- Un principal autorisé à accéder à une balise LF comportant plusieurs valeurs peut accéder aux ressources du catalogue de données avec l'une ou l'autre de ces valeurs. Par exemple, si un utilisateur se voit attribuer une balise LF avec `key= module` et `values=orders, customers`, il a accès aux ressources attribuées soit `module=orders` `module=customers`
- Vous devez être `Grant with LF-Tag expressions` autorisé à accorder des autorisations de données sur les ressources du catalogue de données à l'aide de la méthode LF-TBAC. L'administrateur du lac de données et le créateur du tag LF reçoivent implicitement cette autorisation. Un principal `Grant with LFTag expressions` autorisé peut accorder des autorisations de données sur les ressources en utilisant :
 - la méthode de ressource nommée
 - la méthode LF-TBAC, mais en utilisant uniquement la même expression LF-Tag

Supposons, par exemple, que l'administrateur du lac de données accorde l'autorisation suivante (en pseudo-code).

```
GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH  
GRANT OPTION
```

Dans ce cas, `user1` vous pouvez attribuer des tables `SELECT` à d'autres principaux en utilisant la méthode LF-TBAC, mais uniquement avec l'expression LF-Tag complète.
`module=customers, region=west,south`

- Si un principal obtient des autorisations sur une ressource à la fois avec la méthode LF-TBAC et la méthode de ressource nommée, les autorisations que le principal possède sur la ressource sont l'union des autorisations accordées par les deux méthodes.
- Lake Formation prend en charge l'attribution `DESCRIBE` et `ASSOCIATE` l'attribution de balises LF entre les comptes, ainsi que l'octroi d'autorisations sur les ressources du catalogue de données entre les comptes à l'aide de la méthode LF-TBAC. Dans les deux cas, le principal est un identifiant de AWS compte.

Note

Lake Formation soutient les subventions intercomptes aux organisations et aux unités organisationnelles en utilisant la méthode LF-TBAC. Pour utiliser cette fonctionnalité, vous devez mettre à jour les paramètres de version entre comptes vers la version 3.

Pour de plus amples informations, veuillez consulter [Partage de données entre comptes dans Lake Formation](#).

- Les ressources du catalogue de données créées dans un compte ne peuvent être étiquetées qu'à l'aide de balises LF créées dans le même compte. Les balises LF créées dans un compte ne peuvent pas être associées à des ressources partagées depuis un autre compte.
- L'utilisation du contrôle d'accès basé sur les balises de Lake Formation (LF-TBAC) pour accorder un accès entre comptes aux ressources du catalogue de données nécessite des ajouts à la politique de ressources du catalogue de données de votre compte. AWS Pour de plus amples informations, veuillez consulter [Conditions préalables](#).
- Les touches LF-Tag et les valeurs LF-Tag ne peuvent pas dépasser 50 caractères.
- Le nombre maximum de balises LF pouvant être attribuées à une ressource de catalogue de données est de 50.
- Les limites suivantes sont des limites souples :
 - Le nombre maximum de balises LF pouvant être créées est de 1 000.
 - Le nombre maximum de valeurs pouvant être définies pour une balise LF est de 1 000.
- Les balises, les clés et les valeurs sont converties en minuscules lors de leur enregistrement.
- Une seule valeur pour une balise LF peut être affectée à une ressource particulière.
- Si plusieurs balises LF sont accordées à un principal avec une seule autorisation, le principal ne peut accéder qu'aux ressources du catalogue de données contenant toutes les balises LF.
- Si l'évaluation d'une expression LF-Tag ne donne accès qu'à un sous-ensemble de colonnes de table, mais que l'autorisation Lake Formation accordée en cas de correspondance est l'une des autorisations nécessitant un accès complet aux colonnes, à savoir `Alter`, ou `Drop InsertDelete`, alors aucune de ces autorisations n'est accordée. Au lieu de cela, seul `Describe` est accordé. Si l'autorisation accordée est `All (Super)`, alors uniquement `Select` et `Describe` sont accordées.
- Les caractères génériques ne sont pas utilisés avec les balises LF. Pour attribuer une balise LF à toutes les colonnes d'une table, vous attribuez la balise LF à la table, et toutes les colonnes de la table héritent de la balise LF. Pour attribuer une balise LF à toutes les tables d'une base de données, vous attribuez la balise LF à la base de données, et toutes les tables de la base de données héritent de cette balise LF.
- Vous pouvez créer jusqu'à 1 000 expressions LF-Tag dans un compte.

- Vous pouvez utiliser jusqu'à 50 expressions LF-Tag pour accorder des autorisations à un principal sur les ressources du catalogue de données.
- Lorsque vous accordez ou révoquez des autorisations sur une expression de balise LF en ligne, la taille de l'expression de balise LF ne peut pas dépasser 900 octets. Pour accorder des autorisations sur des expressions de balise LF de plus grande taille, utilisez les expressions de balise LF enregistrées. Pour de plus amples informations, veuillez consulter [Création d'expressions LF-Tag](#).
- Pour ajouter le LF-Tag aux catalogues fédérés Redshift existants créés avant la sortie de disponibilité générale du support LF-Tag pour les catalogues fédérés, vous devez contacter l'équipe d'assistance pour obtenir de l'aide. AWS

Considérations relatives au contrôle d'accès basé sur les attributs, limites et régions prises en charge

Les considérations et limitations suivantes s'appliquent au contrôle d'accès basé sur les attributs (ABAC).

- ABAC ne prend pas en charge l'octroi d'accès à l'aide de politiques LF-Tag.
- Les autorisations pouvant être accordées ne sont pas disponibles avec ABAC.
- ABAC ne prend pas en charge l'octroi d'autorisations aux utilisateurs d'IAM Identity Center.
- Lorsque vous utilisez des autorisations ABAC sur une table dans Lake Formation, Lake Formation n'accorde aucune DESCRIBE autorisation à la base de données ou au catalogue parent. Cela diffère des scénarios non ABAC, dans lesquels Lake Formation fournit des DESCRIBE autorisations implicites aux ressources parentes.
- Tous les principaux dotés de la clé AmazonDataZoneProject tag sont toujours considérés comme ayant souscrit à Lake Formation pour toutes les ressources du catalogue de données.
- ABAC ne prend en charge que les attributs de chaîne.

Résolution des problèmes liés à la formation du Lake

Si vous rencontrez des problèmes lorsque vous travaillez avec AWS Lake Formation, consultez les rubriques de cette section.

Rubriques

- [Résolution de problème généraux](#)
- [Résolution des problèmes d'accès entre comptes](#)
- [Résolution des problèmes liés aux plans et aux flux de travail](#)
- [Problèmes connus pour AWS Lake Formation](#)
- [Message d'erreur mis à jour](#)

Résolution de problème généraux

Utilisez les informations fournies ici pour vous aider à diagnostiquer et à résoudre divers problèmes liés à la Lake Formation.

Erreur : autorisations insuffisantes pour Lake Formation sur <Amazon S3 location>

Une tentative a été faite pour créer ou modifier une ressource de catalogue de données sans autorisation de localisation des données sur l'emplacement Amazon S3 indiqué par la ressource.

Si une base de données ou une table de catalogue de données pointe vers un emplacement Amazon S3, lorsque vous accordez les autorisations CREATE_TABLE à Lake Formation ALTER, vous devez également accorder l'AUTHORIZATION_ACCESS autorisation sur cet emplacement. Si vous accordez ces autorisations à des comptes externes ou à des organisations, vous devez inclure l'option d'octroi.

Une fois ces autorisations accordées à un compte externe, l'administrateur du lac de données de ce compte doit les accorder aux principaux (utilisateurs ou rôles) du compte. Lorsque vous DATA_LOCATION_ACCESS accordez l'autorisation reçue d'un autre compte, vous devez spécifier l'ID de catalogue (ID de AWS compte) du compte propriétaire. Le compte propriétaire est le compte qui a enregistré l'emplacement.

Pour plus d'informations, consultez [Contrôle d'accès aux données sous-jacent](#) et [Octroi d'autorisations de localisation des données](#).

Erreur : « Autorisations de clé de chiffrement insuffisantes pour l'API Glue »

Une tentative a été faite pour accorder des autorisations à Lake Formation sans autorisations Gestion des identités et des accès AWS (IAM) sur la clé de AWS KMS chiffrement d'un catalogue de données chiffré.

Ma requête Amazon Athena ou celle d'Amazon Redshift qui utilise des manifestes échoue

Lake Formation ne prend pas en charge les requêtes utilisant des manifestes.

Erreur : « Autorisations de formation lacustres insuffisantes : création d'une balise requise dans le catalogue »

user/role Il doit être un administrateur de data lake.

Erreur lors de la suppression d'administrateurs de lacs de données non valides

Vous devez supprimer simultanément tous les administrateurs de lacs de données non valides (rôles IAM supprimés définis comme administrateurs de lacs de données). Si vous essayez de supprimer séparément les administrateurs de data lake non valides, Lake Formation génère une erreur principale non valide.

Résolution des problèmes d'accès entre comptes

Utilisez les informations fournies ici pour vous aider à diagnostiquer et à résoudre les problèmes d'accès entre comptes.

Rubriques

- [J'ai accordé l'autorisation d'utiliser plusieurs comptes Lake Formation, mais le destinataire ne peut pas voir la ressource](#)
- [Les utilisateurs principaux du compte destinataire peuvent voir la ressource du catalogue de données, mais ne peuvent pas accéder aux données sous-jacentes](#)
- [Erreur : « L'association a échoué car l'appelant n'était pas autorisé » lors de l'acceptation d'une invitation de partage de AWS RAM ressources](#)

- [Erreur : « Non autorisé à accorder des autorisations pour la ressource »](#)
- [Erreur : « Accès refusé pour récupérer les informations de AWS l'organisation »](#)
- [Erreur : « Organisation <organization-ID>introuvable »](#)
- [Erreur : « Permissions de formation lacustres insuffisantes : combinaison illégale »](#)
- [ConcurrentModificationException sur les grant/revoke demandes adressées à des comptes externes](#)
- [Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes](#)

J'ai accordé l'autorisation d'utiliser plusieurs comptes Lake Formation, mais le destinataire ne peut pas voir la ressource

- L'utilisateur du compte destinataire est-il un administrateur de data lake ? Seuls les administrateurs de data lake peuvent voir la ressource au moment du partage.
- Partagez-vous avec un compte externe à votre organisation en utilisant la méthode des ressources nommées ? Dans ce cas, l'administrateur du lac de données du compte destinataire doit accepter une invitation à partager des ressources dans AWS Resource Access Manager (AWS RAM).

Pour de plus amples informations, veuillez consulter [the section called “Accepter une invitation à partager des AWS RAM ressources”](#).

- Utilisez-vous des politiques de ressources au niveau du compte (catalogue de données) dans ? AWS Glue Dans l'affirmative, si vous utilisez la méthode des ressources nommées, vous devez inclure une déclaration spéciale dans la politique qui autorise le partage AWS RAM des politiques en votre nom.

Pour de plus amples informations, veuillez consulter [the section called “Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation”](#).

- Disposez-vous des autorisations Gestion des identités et des accès AWS (IAM) requises pour accorder un accès entre comptes ?

Pour de plus amples informations, veuillez consulter [the section called “Conditions préalables”](#).

- La ressource pour laquelle vous avez accordé des autorisations ne doit pas avoir d'autorisation Lake Formation accordée au IAMAllowedPrincipals groupe.
- La politique au niveau du compte comporte-t-elle une deny déclaration sur la ressource ?

Les utilisateurs principaux du compte destinataire peuvent voir la ressource du catalogue de données, mais ne peuvent pas accéder aux données sous-jacentes

Les principaux associés au compte du destinataire doivent disposer des autorisations Gestion des identités et des accès AWS (IAM) requises. Pour en savoir plus, consultez [Accès aux données sous-jacentes d'une table partagée](#).

Erreur : « L'association a échoué car l'appelant n'était pas autorisé » lors de l'acceptation d'une invitation de partage de AWS RAM ressources

Après avoir accordé l'accès à une ressource à un autre compte, lorsque le compte destinataire tente d'accepter l'invitation de partage de ressources, l'action échoue.

```
$ aws ram get-resource-share-associations --association-type PRINCIPAL --resource-share-arns arn:aws:ram:aws-region:444444444444:resource-share/e1d1f4ba-xxxx-xxxx-xxxx-xxxxxxxx5d8d
{
  "resourceShareAssociations": [
    {
      "resourceShareArn": "arn:aws:ram:aws-region:444444444444:resource-share/e1d1f4ba-xxxx-xxxx-xxxx-xxxxxxxx5d8d",
      "resourceShareName": "LakeFormation-MMCC0XQBH3Y",
      "associatedEntity": "5815803XXXXX",
      "associationType": "PRINCIPAL",
      "status": "FAILED",
      "statusMessage": "Association failed because the caller was not authorized.",
      "creationTime": "2021-07-12T02:20:10.267000+00:00",
      "lastUpdatedTime": "2021-07-12T02:20:51.830000+00:00",
      "external": true
    }
  ]
}
```

L'erreur se produit parce que le `glue:PutResourcePolicy` est invoqué AWS Glue lorsque le compte destinataire accepte l'invitation de partage de ressources. Pour résoudre le problème, autorisez `glue:PutResourcePolicy` par le rôle assumé utilisé par le producteur/grantor compte.

Erreur : « Non autorisé à accorder des autorisations pour la ressource »

Une tentative a été faite pour accorder des autorisations entre comptes sur une base de données ou une table appartenant à un autre compte. Lorsqu'une base de données ou une table est partagée avec votre compte, en tant qu'administrateur de data lake, vous ne pouvez accorder des autorisations à ce sujet qu'aux utilisateurs de votre compte.

Erreur : « Accès refusé pour récupérer les informations de AWS l'organisation »

Votre compte est un compte de gestion des AWS Organisations et vous ne disposez pas des autorisations requises pour récupérer les informations relatives à l'organisation, telles que les unités organisationnelles du compte.

Pour de plus amples informations, veuillez consulter [Required permissions for cross-account grants](#).

Erreur : « Organisation <organization-ID>introuvable »

Une tentative a été faite pour partager une ressource avec une organisation, mais le partage avec les organisations n'est pas activé. Activez le partage des ressources avec les organisations.

Pour plus d'informations, consultez la section [Activer le partage avec AWS les organisations](#) dans le guide de AWS RAM l'utilisateur.

Erreur : « Permissions de formation lacustres insuffisantes : combinaison illégale »

Un utilisateur a partagé une ressource de catalogue de données alors que les autorisations de Lake Formation étaient accordées au IAMAllowedPrincipals groupe pour cette ressource. L'utilisateur doit révoquer toutes les autorisations de Lake Formation IAMAllowedPrincipals avant de partager la ressource.

ConcurrentModificationException sur les grant/revoke demandes adressées à des comptes externes

Lorsque les utilisateurs font plusieurs demandes d'autorisation et de and/or révocation d'autorisation simultanées pour un principal sur des politiques LF-Tag, Lake Formation les lance. ConcurrentModificationException Les utilisateurs doivent intercepter l'exception

et réessayer si la grant/revoke demande a échoué. L'utilisation de versions par lots des opérations `GrantPermissions/RevokePermissionsAPI` [BatchRevokePermissions](#) permet [BatchGrantPermissions](#) de résoudre ce problème dans une certaine mesure en réduisant le nombre de grant/revoke demandes simultanées.

Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes

Lorsque vous utilisez Amazon EMR pour accéder aux données partagées avec vous depuis un autre compte, certaines bibliothèques Spark tentent d'appeler l'opération d'API `Glue:GetUserDefinedFunctions`. Les versions 1 et 2 des autorisations AWS RAM gérées ne prenant pas en charge cette action, le message d'erreur suivant s'affiche :

```
"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-spark-role/i-06ab8c2b59299508a is not authorized to perform: glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource because no resource-based policy allows the glue:GetUserDefinedFunctions action"
```

Pour résoudre cette erreur, l'administrateur du lac de données qui a créé le partage de ressources doit mettre à jour les autorisations AWS RAM gérées associées au partage de ressources. La version 3 des autorisations gérées AWS RAM permet aux nœuds principaux d'effectuer l'action `glue:GetUserDefinedFunctions`.

Si vous créez un nouveau partage de ressources, Lake Formation applique la dernière version de l'autorisation AWS RAM gérée par défaut, et aucune action n'est requise de votre part. Pour activer l'accès aux données entre comptes pour les partages de ressources existants, vous devez mettre à jour les autorisations AWS RAM gérées vers la version 3.

Vous pouvez consulter les AWS RAM autorisations attribuées aux ressources partagées avec vous dans AWS RAM. Les autorisations suivantes sont incluses dans la version 3 :

Databases

- `AWSRAMPermissionGlueDatabaseReadWriteForCatalog`
- `AWSRAMPermissionGlueDatabaseReadWrite`

Tables

- `AWSRAMPermissionGlueTableReadWriteForCatalog`
- `AWSRAMPermissionGlueTableReadWriteForDatabase`

AllTables

```
AWSRAMPermissionGlueAllTablesReadWriteForCatalog  
AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Pour mettre à jour la version des autorisations AWS RAM gérées des partages de ressources existants

Vous (administrateur du lac de données) pouvez soit [mettre à jour les autorisations AWS RAM gérées vers une version plus récente](#) en suivant les instructions du guide de AWS RAM l'utilisateur, soit révoquer toutes les autorisations existantes pour le type de ressource et les réaccorder. Si vous révoquez les autorisations, le partage AWS RAM de AWS RAM ressources associé au type de ressource est supprimé. Lorsque vous réaccordez des autorisations, AWS RAM de nouveaux partages de ressources sont créés en y joignant la dernière version des autorisations AWS RAM gérées.

Résolution des problèmes liés aux plans et aux flux de travail

Utilisez les informations fournies ici pour vous aider à diagnostiquer et à résoudre les problèmes liés au plan et au flux de travail.

Rubriques

- [<role-ARN>Mon plan a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à exécuter : iam : PassRole on resource : »](#)
- [<role-ARN>Mon flux de travail a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à effectuer : iam : PassRole on resource : »](#)
- [Un crawler de mon flux de travail a échoué avec le message « La ressource n'existe pas ou le demandeur n'est pas autorisé à accéder aux autorisations demandées »](#)
- [Un crawler de mon flux de travail a échoué avec « Une erreur s'est produite \(AccessDeniedException\) lors de l'appel de l' CreateTable opération... »](#)

<role-ARN>Mon plan a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à exécuter : iam : PassRole on resource : »

Un utilisateur qui ne dispose pas des autorisations suffisantes pour transmettre le rôle choisi a tenté de créer un plan.

Mettez à jour la politique IAM de l'utilisateur pour pouvoir transmettre le rôle, ou demandez-lui de choisir un autre rôle avec les autorisations de mot de passe requises.

Pour de plus amples informations, veuillez consulter [the section called “Référence des personnalités de Lake Formation et des autorisations IAM”](#).

<role-ARN>Mon flux de travail a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à effectuer : iam : PassRole on resource : »

Le rôle que vous avez spécifié pour le flux de travail ne comportait pas de politique intégrée permettant au rôle de se transmettre de lui-même.

Pour de plus amples informations, veuillez consulter [the section called “\(Facultatif\) Créez un rôle IAM pour les flux de travail”](#).

Un crawler de mon flux de travail a échoué avec le message « La ressource n'existe pas ou le demandeur n'est pas autorisé à accéder aux autorisations demandées »

L'une des causes possibles est que le rôle transmis ne disposait pas des autorisations suffisantes pour créer une table dans la base de données cible. Accordez au rôle l'CREATE_TABLE autorisation d'accéder à la base de données.

Un crawler de mon flux de travail a échoué avec « Une erreur s'est produite (AccessDeniedException) lors de l'appel de l' CreateTable opération... »

L'une des causes possibles est que le rôle de flux de travail ne disposait pas d'autorisations de localisation des données sur l'emplacement de stockage cible. Accordez des autorisations de localisation des données au rôle.

Pour de plus amples informations, veuillez consulter [the section called “DATA_LOCATION_ACCESS”](#).

Problèmes connus pour AWS Lake Formation

Passez en revue ces problèmes connus pour AWS Lake Formation.

Rubriques

- [Limitation du filtrage des métadonnées des tables](#)

- [Problème lié au changement de nom d'une colonne exclue](#)
- [Problème lié à la suppression de colonnes dans les tableaux CSV](#)
- [Les partitions de table doivent être ajoutées sous un chemin commun](#)
- [Problème lié à la création d'une base de données pendant la création du flux de travail](#)
- [Problème lié à la suppression puis à la recréation d'un utilisateur](#)
- [Les opérations de l'API Data Catalog ne mettent pas à jour la valeur du `IsRegisteredWithLakeFormation` paramètre](#)
- [Les opérations de Lake Formation ne prennent pas en charge AWS Glue le registre des schémas](#)

Limitation du filtrage des métadonnées des tables

AWS Lake Formation les autorisations au niveau des colonnes peuvent être utilisées pour restreindre l'accès à des colonnes spécifiques d'un tableau. Lorsqu'un utilisateur extrait des métadonnées relatives à la table à l'aide de la console ou d'une API similaire `glue:GetTable`, la liste des colonnes de l'objet de table contient uniquement les champs auxquels il a accès. Il est important de comprendre les limites de ce filtrage des métadonnées.

Bien que Lake Formation mette à disposition des métadonnées relatives aux autorisations des colonnes pour les services intégrés, le filtrage des colonnes dans les réponses aux requêtes relève de la responsabilité du service intégré. Les clients de Lake Formation qui prennent en charge le filtrage au niveau des colonnes, notamment Amazon Athena, Amazon Redshift Spectrum et Amazon EMR, filtrent les données en fonction des autorisations de colonne enregistrées auprès de Lake Formation. Les utilisateurs ne pourront pas lire les données auxquelles ils ne devraient pas avoir accès. Actuellement, l'AWS Glue ETL ne prend pas en charge le filtrage des colonnes.

Note

Les clusters EMR ne sont pas entièrement gérés par AWS. Il est donc de la responsabilité des administrateurs EMR de sécuriser correctement les clusters afin d'éviter tout accès non autorisé aux données.

Certaines applications ou certains formats peuvent stocker des métadonnées supplémentaires, notamment des noms et des types de colonnes, dans la `Parameters` carte sous forme de propriétés de table. Ces propriétés sont renvoyées telles quelles et sont accessibles à tous les utilisateurs `SELECT` autorisés sur n'importe quelle colonne.

Par exemple, l'[Avro SerDe](#) stocke une représentation JSON du schéma de table dans une propriété de table nommée `avro.schema.literal`, qui est accessible à tous les utilisateurs ayant accès à la table. Nous vous recommandons d'éviter de stocker des informations sensibles dans les propriétés des tables et de savoir que les utilisateurs peuvent apprendre le schéma complet des tables au format Avro. Cette limitation est spécifique aux métadonnées relatives à une table.

AWS Lake Formation supprime toute propriété de table en commençant par « `spark.sql.sources.schema` lors de la réponse à une demande `glue:GetTable` ou à une demande similaire » si l'appelant n'a pas d'autorisations `SELECT` sur toutes les colonnes de la table. Cela empêche les utilisateurs d'accéder à des métadonnées supplémentaires concernant les tables créées avec Apache Spark. Lorsqu'elles sont exécutées sur Amazon EMR, les applications Apache Spark peuvent toujours lire ces tables, mais certaines optimisations risquent de ne pas être appliquées et les noms de colonnes distinguant majuscules et minuscules ne sont pas pris en charge. Si l'utilisateur a accès à toutes les colonnes de la table, Lake Formation renvoie la table non modifiée avec toutes les propriétés de la table.

Problème lié au changement de nom d'une colonne exclue

Si vous utilisez des autorisations au niveau des colonnes pour exclure une colonne puis renommer la colonne, celle-ci n'est plus exclue des requêtes, telles que `SELECT *`

Problème lié à la suppression de colonnes dans les tableaux CSV

Si vous créez une table de catalogue de données au format CSV, puis que vous supprimez une colonne du schéma, les requêtes peuvent renvoyer des données erronées et les autorisations au niveau des colonnes risquent de ne pas être respectées.

Solution : créez plutôt une nouvelle table.

Les partitions de table doivent être ajoutées sous un chemin commun

Lake Formation s'attend à ce que toutes les partitions d'une table suivent un chemin commun défini dans le champ d'emplacement de la table. Lorsque vous utilisez le robot pour ajouter des partitions à un catalogue, cela fonctionne parfaitement. Mais si vous ajoutez des partitions manuellement et que ces partitions ne se trouvent pas à l'emplacement défini dans la table parent, l'accès aux données ne fonctionne pas.

Problème lié à la création d'une base de données pendant la création du flux de travail

Lorsque vous créez un flux de travail à partir d'un plan à l'aide de la console Lake Formation, vous pouvez créer la base de données cible si elle n'existe pas. Dans ce cas, l'utilisateur connecté obtient l'`CREATE_TABLE` autorisation d'accéder à la base de données créée. Cependant, le robot généré par le flux de travail assume le rôle du flux de travail lorsqu'il tente de créer une table. Cela échoue car le rôle n'a pas l'`CREATE_TABLE` autorisation d'accéder à la base de données.

Solution : Si vous créez la base de données via la console lors de la configuration du flux de travail, avant d'exécuter le flux de travail, vous devez accorder au rôle associé au flux de travail l'`CREATE_TABLE` autorisation sur la base de données que vous venez de créer.

Problème lié à la suppression puis à la recreation d'un utilisateur

Le scénario suivant se traduit par des autorisations erronées de Lake Formation renvoyées par `lakeformation:ListPermissions` :

1. Créez un utilisateur et accordez les autorisations de Lake Formation.
2. Supprimez l'utilisateur.
3. Recréez l'utilisateur portant le même nom.

`ListPermissions` renvoie deux entrées, une pour l'ancien utilisateur et une pour le nouvel utilisateur. Si vous essayez de révoquer les autorisations accordées à l'ancien utilisateur, les autorisations sont révoquées pour le nouvel utilisateur.

Les opérations de l'API Data Catalog ne mettent pas à jour la valeur du **IsRegisteredWithLakeFormation** paramètre

Il existe une limite connue selon laquelle les opérations de l'API du catalogue de données, telles que `GetTables` la mise à jour ou non de la valeur du `IsRegisteredWithLakeFormation` paramètre, renvoient la valeur par défaut, qui est fausse. `SearchTables` Il est recommandé d'utiliser l'`getTableAPI` pour afficher la valeur correcte du `IsRegisteredWithLakeFormation` paramètre.

Les opérations de Lake Formation ne prennent pas en charge AWS Glue le registre des schémas

Les opérations de Lake Formation ne prennent pas en charge AWS Glue les tables contenant un `SchemaReference` dans le `StorageDescriptor` à utiliser dans le [registre des schémas](#).

Message d'erreur mis à jour

AWS Lake Formation a mis à jour les exceptions spécifiques aux ressources par rapport au message `EntityNotFound` d'erreur général pour les opérations d'API suivantes afin de répondre aux objectifs de sécurité et de conformité.

- `RevokePermissions`
- `GrantPermissions`
- `GetResourceLFTags`
- `GetTable`
- `GetDatabase`

AWS Lake Formation API

Note

[La référence d'API](#) mise à jour pour le AWS Lake Formation service est désormais disponible.

Table des matières

- [Autorisations APIs](#)
 - [Opérations](#)
 - [Les types de données](#)
- [Paramètres du lac de données APIs](#)
 - [Opérations](#)
 - [Les types de données](#)
- [Intégration à IAM Identity Center APIs](#)
 - [Opérations](#)
 - [Les types de données](#)
- [Mode d'accès hybride APIs](#)
 - [Opérations](#)
 - [Les types de données](#)
- [Vente d'informations d'identification APIs](#)
 - [Opérations](#)
 - [Les types de données](#)
- [Marquage APIs](#)
 - [Opérations](#)
 - [Les types de données](#)
- [Filtre de données APIs](#)
 - [Opérations](#)
 - [Types de données](#)
- [Types de données courants](#)

- [ErrorDetail structure](#)
- [Modèles de chaîne](#)

Autorisations APIs

La section Permissions API décrit les opérations et les types de données requis pour accorder et révoquer des autorisations dans AWS Lake Formation. Consultez le [guide de référence de l'API Lake Formation](#) pour toutes les opérations AWS Lake Formation d'API et tous les types de données.

Opérations

- [GrantPermissions](#)
- [RevokePermissions](#)
- [BatchGrantPermissions](#)
- [BatchRevokePermissions](#)
- [GetEffectivePermissionsForPath](#)
- [ListPermissions](#)
- [GetDataLakePrincipal](#)

Les types de données

- [Ressource](#)
- [DatabaseResource](#)
- [TableResource](#)
- [TableWithColumnsResource](#)
- [DataCellsFilterResource](#)
- [DataLocationResource](#)
- [DataLakePrincipal](#)
- [PrincipalPermissions](#)
- [PrincipalResourcePermissions](#)
- [DetailsMap](#)
- [ColumnWildcard](#)

- [BatchPermissionsRequestEntry](#)
- [BatchPermissionsFailureEntry](#)

Paramètres du lac de données APIs

Cette section contient les paramètres du lac de données, les opérations de l'API et les types de données permettant de gérer les administrateurs du lac de données.

Opérations

- [GetDataLakeSettings](#)
- [PutDataLakeSettings](#)

Les types de données

- [DataLakeSettings](#)

Intégration à IAM Identity Center APIs

Cette section contient les opérations de création et de gestion de l'intégration de Lake Formation avec IAM Identity Center.

Opérations

- [CreateLakeFormationIdentityCenterConfiguration](#)
- [DeleteLakeFormationIdentityCenterConfiguration](#)
- [DescribeLakeFormationIdentityCenterConfiguration](#)
- [UpdateLakeFormationIdentityCenterConfiguration](#)

Les types de données

- [ExternalFilteringConfiguration](#)

Mode d'accès hybride APIs

La section API du mode d'accès hybride décrit les opérations et les types de données requis pour configurer le mode d'accès hybride dans AWS Lake Formation. Consultez le [guide de référence de l'API Lake Formation](#) pour toutes les opérations AWS Lake Formation d'API et tous les types de données.

Opérations

- [CreateLakeFormationOptIn](#)
- [DeleteLakeFormationOptIn](#)
- [ListLakeFormationOptIns](#)

Les types de données

- [Ressource](#)
- [DatabaseResource](#)
- [TableResource](#)
- [Informations sur les ressources](#)
- [LakeFormationOptInsInfo](#)
- [DataLocationResource](#)

Vente d'informations d'identification APIs

La section API Credential Vending décrit les opérations et les types de données liés à l'utilisation du AWS Lake Formation service pour vendre des informations d'identification et pour enregistrer et gérer une ressource de lac de données.

Opérations

- [RegisterResource](#)
- [DeregisterResource](#)
- [ListResources](#)
- [GetUnfilteredTableMetadata](#)

- [GetUnfilteredPartitionsMetadata](#)
- [GetTemporaryGluePartitionCredentials](#)
- [GetTemporaryGlueTableCredentials](#)
- [UpdateResource](#)

Les types de données

- [FilterCondition](#)
- [RowFilter](#)
- [ResourceInfo](#)

Marquage APIs

La section API de balisage décrit les opérations et les types de données liés à une stratégie d'autorisation qui définit un modèle d'autorisation sur les attributs ou les balises de paires clé-valeur.

Opérations

- [Obtenir une LFTag expression](#)
- [LFTagExpressions de liste](#)
- [Supprimer LFTag l'expression](#)
- [Mettre à jour LFTag l'expression](#)
- [Créer une LFTag expression](#)
- [AdditionLFTagsToResource](#)
- [SupprimezLFTagsFromResource](#)
- [GetResourceLFTags](#)
- [ListLFTags](#)
- [CréerLFTag](#)
- [Obtenez LFTag](#)
- [Mettre à jourLFTag](#)
- [SuppressionLFTag](#)
- [SearchTablesByLFTags](#)

- [SearchDatabasesByLFTags](#)

Les types de données

- [LFTagKeyResource](#)
- [LFTagPolicyResource](#)
- [TaggedTable](#)
- [TaggedDatabase](#)
- [LFTag](#)
- [LFTagPaire](#)
- [LFTagError \(Erreur\)](#)
- [ColonneLFTag](#)

Filtre de données APIs

Le filtre de données APIs décrit comment gérer les filtres de cellules de données dans AWS Lake Formation.

Opérations

- [CreateDataCellsFilter](#)
- [DeleteDataCellsFilter](#)
- [ListDataCellsFilter](#)
- [GetDataCellsFilter](#)
- [UpdateDataCellsFilter](#)

Types de données

- [DataCellsFilter](#)
- [RowFilter](#)

Types de données courants

Les types de données courants décrivent des types de données courants variés dans AWS Lake Formation.

ErrorDetail structure

Contient des informations détaillées sur une erreur.

Champs

- **ErrorCode** – Chaîne UTF-8, d'une longueur comprise entre 1 et 255 octets, correspondant au [Single-line string pattern](#).

Code associé à cette erreur.

- `ErrorMessage` – Chaîne de description, d'une longueur maximale de 2 048 octets, correspondant au [URI address multi-line string pattern](#).

Message décrivant l'erreur.

Modèles de chaîne

L'API utilise les expressions régulières suivantes pour définir le contenu valide pour différents paramètres et membres de chaîne :

- Modèle de chaîne à ligne unique – "[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*"
- Modèle de chaîne à plusieurs lignes d'adresse URI – "[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*"
- Schéma de chaîne personnalisé #3 — « ^\w+\.\w+\.\w+\$ »
- Schéma de chaîne personnalisé #4 — « ^\w+\.\w+\$ »
- Schéma de chaîne personnalisé #5 — « arn:aws:iam::[0-9]*:role/. * »
- Schéma de chaîne personnalisé #6 — « arn:aws:iam::[0-9]*:user/. * »
- Schéma de chaîne personnalisé #7 — « arn:aws:iam::[0-9]*:group/. * »
- Modèle de chaîne personnalisée #8 – "arn:aws:iam::[0-9]*:saml-provider/. *"
- Modèle de chaîne personnalisée #9 – "^([\p{L}\p{Z}\p{N} \.:/=+~-@%]*)\$"

- Modèle de chaîne personnalisée #10 – `"^([\p{L}\p{Z}\p{N}_.\:*\/=+\-@%]*)$"`
- Modèle de chaîne personnalisée #11 – `"[\p{L}\p{N}\p{P}]*"`

Régions prises en charge

Cette section contient des informations sur la prise en charge Régions AWS et les fonctionnalités de Lake Formation.

Disponibilité générale

Pour les services Régions AWS pris en charge par AWS Lake Formation, consultez [la liste des AWS services disponibles par région](#).

Pour une liste des points de terminaison du service Lake Formation pour chaque région et des quotas du service Lake Formation, voir [AWS Lake Formation Points de terminaison et quotas](#).

AWS GovCloud (US)

Pour un aperçu des différences entre AWS GovCloud (US) la région et la norme Régions AWS, voir [En quoi AWS Lake Formation diffère pour AWS GovCloud \(US\)](#).

Optimisation des transactions et du stockage

Les tables gouvernées, la prise en charge des transactions et les fonctionnalités d'optimisation du stockage pour Lake Formation sont disponibles dans les versions suivantes : Régions AWS

Nom de la région	Paramètre de région	Endpoint
USA Est (Virginie du Nord)	us-east-1	lakeformation.us-east-1.amazonaws.com
		lakeformation-fips.us-east-1.amazonaws.com
USA Est (Ohio)	us-east-2	lakeformation.us-east-2.amazonaws.com
		lakeformation-fips.us-east-2.amazonaws.com

Nom de la région	Paramètre de région	Endpoint
USA Ouest (Oregon)	us-west-2	lakeformation.us-west-2.amazonaws.com lakeformation-fips.us-west-2.amazonaws.com
Asie-Pacifique (Mumbai)	ap-south-1	lakeformation.ap-south-1.amazonaws.com
Asie-Pacifique (Séoul)	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
Asie-Pacifique (Singapour)	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
Asie-Pacifique (Sydney)	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
Asie-Pacifique (Tokyo)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
Europe (Francfort)	eu-central-1	lakeformation.eu-central-1.amazonaws.com
Europe (Irlande)	eu-west-1	lakeformation.eu-west-1.amazonaws.com
Europe (Londres)	eu-west-2	lakeformation.eu-west-2.amazonaws.com
Europe (Stockholm)	eu-north-1	lakeformation.eu-north-1.amazonaws.com

Nom de la région	Paramètre de région	Endpoint
Canada (Centre)	ca-central-1	lakeformation.ca-central-1.amazonaws.com
Amérique du Sud (São Paulo)	sa-east-1	lakeformation.sa-east-1.amazonaws.com

Historique du document pour AWS Lake Formation

Le tableau suivant décrit les modifications importantes apportées à la documentation de AWS Lake Formation.

Modification	Description	Date
Documentation mise à jour pour le contrôle d'accès basé sur les attributs	Lake Formation vous permet d'accorder des autorisations aux principaux dotés d'attributs correspondants sur les ressources du catalogue de données. Pour plus d'informations, consultez la section Contrôle d' accès basé sur les attributs .	17 avril 2025
Intégration d'Amazon S3 Tables avec AWS Lake Formation et AWS Glue Data Catalog	Vous pouvez désormais intégrer et cataloguer les tables S3 sous forme AWS Glue Data Catalog d'objets et enregistrer le catalogue en tant qu'emplacement de données de Lake Formation . Pour plus d'informations, consultez Création d'un catalogue de tables Amazon S3 dans le AWS Glue Data Catalog .	13 mars 2025
AWSLakeFormationCrossAccountManager Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManager politique en remplaçant l'opérateur de StringLike condition par l'ArnLikeopérateur qui	25 janvier 2025

permet à IAM d'effectuer la vérification du format ARN.

[Modification de politique mise à jour](#)

A documenté le changement apporté à la [AWSLakeFormationDataAdmin](#) politique.

3 décembre 2024

[Mises à jour de plusieurs catalogues](#)

AWS Glue Data Catalog vous permet de créer des catalogues fédérés et d'unifier les données entre les lacs de données Amazon S3 et les entrepôts de données Amazon Redshift, ainsi que d'intégrer les données de vos bases de données opérationnelles telles qu'Amazon DynamoDB et de sources de données tierces telles que Snowflake, MySQL, etc. Pour plus d'informations, consultez la [section Transférer vos données dans le AWS Glue Data Catalog](#).

3 décembre 2024

[Documentation mise à jour pour les expressions LF-Tag](#)

Vous pouvez enregistrer des expressions LF-Tag et les réutiliser pour accorder des autorisations sur les ressources du catalogue de données. Pour plus d'informations, consultez la section [Gestion des expressions LF-Tag](#).

7 novembre 2024

Documentation mise à jour pour les vues du catalogue de données	Vous pouvez créer des vues en AWS Glue Data Catalog utilisant AWS Glue APIs en plus d' DDLs Amazon Athena et Amazon Redshift. Pour plus d'informations, consultez la section Création de vues du catalogue de données .	7 août 2024
Ajout de documentation pour la vente d'informations d'identification vérifiables	Lake Formation vous permet d'inclure le contexte de l'utilisateur d'IAM Identity Center dans les CloudTrail événements, puis de suivre les utilisateurs qui accèdent à vos ressources. Pour plus d'informations, voir Inclure le contexte utilisateur d'IAM Identity Center dans les CloudTrail journaux .	14 juillet 2024
Modification de politique mise à jour	A documenté la modification (ajout d'une instruction IDs et suppression des autorisations redondantes) apportée aux AWSLakeFormationDataAdmin politiques AWSLakeFormationCrossAccountManager et.	14 mars 2024
Mise à jour de la configuration de Lake Formation	Les étapes de la AWS Lake Formation section Configuration ont été mises à jour.	7 février 2024

Modification de politique mise à jour	De nouvelles autorisations ont été ajoutées à la politique en ligne du rôle lié au service. Pour plus d'informations, consultez la section Utilisation de rôles liés à un service pour Lake Formation .	7 février 2024
Modification de politique mise à jour	A documenté le changement apporté à la LakeFormationDataAccessServiceRolePolicy politique.	2 février 2024
Limites de la Formation lacustre consolidée	Création d'une section unifiée pour les limites et les considérations relatives à la Formation des Lacs. Pour plus d'informations, consultez la section Limitations relatives aux Lake Formation .	15 décembre 2023
Ajout de documentation pour le compactage des icebergs	Pour améliorer les performances de lecture des services AWS d'analyse tels qu'Athena et Amazon EMR, ainsi que pour les tâches AWS Glue ETL, AWS Glue Data Catalog fournit un compactage géré (un processus qui compacte de petits objets Amazon S3 en objets plus grands) pour les tables Iceberg du catalogue de données. Pour plus d'informations, consultez Optimisation des tables Iceberg .	25 novembre 2023

[Documentation ajoutée pour l'intégration d'IAM Identity Center](#)

Les intégrations d'IAM Identity Center permettent aux utilisateurs et aux groupes d'accéder aux ressources du catalogue de données en appliquant les autorisations de Lake Formation. Pour plus d'informations, consultez la section [Intégration d'IAM Identity Center](#).

25 novembre 2023

[Documentation ajoutée pour les vues du catalogue de données](#)

Vous pouvez créer des vues AWS Glue Data Catalog faisant référence à un maximum de 10 tables à l'aide des éditeurs SQL pour Amazon Athena ou Amazon Redshift. Pour plus d'informations, consultez la section [Création de vues](#).

25 novembre 2023

[Mise à jour du changement de politique](#)

A documenté le changement apporté à la [AWSLakeFormationCrossAccountManager](#) politique.

25 octobre 2023

[Documentation ajoutée pour le mode d'accès hybride](#)

Le mode d'accès hybride offre la flexibilité d'activer de manière sélective les autorisations de Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog. Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants. Pour plus d'informations, consultez la section [Mode d'accès hybride](#).

26 septembre 2023

[Ajout de documentation pour la création de tables Apache Iceberg](#)

Vous pouvez désormais créer des tables Apache Iceberg qui utilisent le format de données Apache Parquet AWS Glue Data Catalog avec des données résidant dans Amazon S3. Pour plus d'informations, consultez la section [Création de tables Iceberg](#).

16 août 2023

[Ajout de documentation pour l'accès aux données entre régions](#)

Lake Formation permet d'interroger les tables du catalogue de données dans toutes les AWS régions. Vous pouvez accéder aux données d'une région depuis d'autres régions à l'aide d'Athena, Amazon EMR, et exécuter l' AWS Glue ETL en créant des liens de ressources dans d'autres régions pointant vers les bases de données et les tables sources. Vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation Pour plus d'informations, consultez la section [Accès aux tables entre les régions](#).

30 juin 2023

[Contenu réorganisé](#)

Les chapitres du guide ont été réorganisés pour correspondre au parcours des utilisateurs de Lake Formation.

15 mai 2023

[Ajout de documentation pour la fédération HMS](#)

Vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation Pour plus d'informations, consultez la section [Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes.](#)

15 avril 2023

[Ajout de documentation pour le partage de données Amazon Redshift](#)

Vous pouvez désormais gérer en toute sécurité les données d'un partage de données depuis Amazon Redshift à l'aide des autorisations de Lake Formation. Lake Formation prend en charge l'accès sous licence à vos données via AWS Data Exchange. Pour plus d'informations, consultez la section [Partage de données dans AWS Lake Formation.](#)

30 novembre 2022

[Support pour le partage de données entre comptes directement avec les donneurs d'ordre](#)

Ajout d'informations sur le partage de données directement avec les responsables IAM dans un autre compte. Pour plus d'informations, voir [Partage de données entre comptes dans AWS Lake Formation](#).

10 novembre 2022

[Support pour le partage de données AWS RAM activé à l'aide de TBAC](#)

[Ajout d'informations sur la méthode LF-TBAC d'octroi d'autorisations de catalogue de données utilisée AWS Resource Access Manager pour les autorisations entre comptes](#).

10 novembre 2022

[Ajout d'une section sur le travail avec d'autres services](#)

Informations supplémentaires sur la manière dont AWS des services tels qu'Athena, AWS Glue Redshift Spectrum et Amazon EMR peuvent utiliser Lake Formation pour accéder en toute sécurité aux données des sites Amazon S3 enregistrés auprès de Lake Formation. Pour de plus amples informations, veuillez consulter [Collaboration avec d'autres AWS services](#).

10 novembre 2022

[???](#)

Ajout d'informations sur la résolution d'une erreur lors de l'utilisation d'Amazon EMR pour accéder aux données entre comptes. Pour de plus amples informations, veuillez consulter [Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes](#).

7 novembre 2022

[Mises à jour du partage des ressources entre comptes](#)

Ajout d'une description du fonctionnement des [partages de ressources entre comptes](#) dans Lake Formation. A documenté le changement apporté à la [AWSLakeFormationCrossAccountManager](#) politique.

6 mai 2022

[Nouveaux tutoriels](#)

Ajout de nouveaux didacticiels pour la création de tables gouvernées, la sécurisation des lacs de données et le partage de lacs de données. Pour plus de détails, consultez la section [Démarrage](#).

20 avril 2022

[Page d'accueil de New Lake Formation](#)

Mise à jour de la page d'accueil de [Lake Formation](#) pour inclure des liens vers des didacticiels fournissant des step-by-step instructions sur la façon de créer un lac de données, d'ingérer des données, de partager et de sécuriser des lacs de données à l'aide de Lake Formation.

20 avril 2022

[Support pour la vente d'informations d'identification](#)

Ajout d'informations sur la vente d'informations d'identification, qui permettent à Lake Formation de permettre à des services tiers de s'intégrer à Lake Formation en utilisant les opérations de l'API de vente d'informations d'identification. Pour plus d'informations, consultez [Comment fonctionne le distributeur automatique d'informations d'identification à Lake Formation](#).

28 février 2022

[Support pour les tables gouvernées et le filtrage avancé des données](#)

Ajout d'informations sur les tables gouvernées, qui prennent en charge les transactions ACID, le compactage automatique des données et les requêtes de voyage dans le temps. Ajout d'informations sur la création de filtres de données pour prendre en charge la sécurité au niveau des colonnes, la sécurité au niveau des lignes et la sécurité au niveau des cellules. Pour plus d'informations, consultez les sections [Tables gouvernées dans la Lake Formation](#) et [Filtrage des données et Sécurité au niveau des cellules dans la Lake Formation](#).

30 novembre 2021

[Support pour les points de terminaison d'interface VPC](#)

Ajout d'informations sur la création d'un point de terminaison d'interface de cloud privé virtuel (VPC) pour Lake Formation, afin que les communications entre votre VPC et Lake Formation soient effectuées de manière complète et sécurisée au sein du réseau. AWS Pour plus d'informations, consultez la section [Utilisation de Lake Formation avec des points de terminaison VPC](#).

11 octobre 2021

[Support des politiques de point de terminaison d'un VPC](#)

Ajout d'informations sur la prise en charge des politiques relatives aux points de terminaison Virtual Private Cloud (VPC) dans Lake Formation. Pour plus d'informations, consultez la section [Utilisation de Lake Formation avec des points de terminaison VPC.](#)

11 octobre 2021

[Support pour le contrôle d'accès basé sur des balises](#)

Le contrôle d'accès basé sur les balises Lake Formation fournit une nouvelle méthode plus évolutive pour gérer l'accès aux ressources du catalogue de données et aux données sous-jacentes à l'aide de balises LF. Pour plus d'informations, consultez la section [Contrôle d'accès basé sur des balises Lake Formation.](#)

7 mai 2021

[Nouvelle exigence d'inscription pour le filtrage des données sur Amazon EMR.](#)

Ajout d'informations concernant l'obligation d'opter pour autoriser Amazon EMR à filtrer les données gérées par Lake Formation. Pour plus d'informations, consultez [Autoriser le filtrage des données sur Amazon EMR.](#)

9 octobre 2020

[Support pour l'octroi d'autorisations complètes entre comptes sur les bases de données du catalogue de données](#)

Ajout d'informations sur l'octroi d'autorisations complètes à Lake Formation sur les bases de données du catalogue de données pour tous AWS les comptes, notamment CREATE_TABLE . Pour plus d'informations, voir [Partage de bases de données de catalogues de données](#).

1er octobre 2020

[Support pour l'authentification Amazon Athena des utilisateurs via SAML.](#)

Ajout d'informations sur la prise en charge des utilisateurs d'Athena qui se connectent via le pilote JDBC ou ODBC et s'authentifient via des fournisseurs d'identité SAML tels qu'Okta et Microsoft Active Directory Federation Service (AD FS). Pour plus d'informations, voir [Intégrations de AWS services avec Lake Formation](#).

30 septembre 2020

[Support pour l'accès entre comptes avec un catalogue de données crypté](#)

Ajout d'informations sur l'octroi d'autorisations entre comptes lorsque le catalogue de données est crypté. Pour plus d'informations, consultez la [section Conditions préalables relatives à l'accès entre comptes](#).

30 juillet 2020

[Support pour l'accès entre comptes au lac de données](#)

Ajout d'informations sur l'octroi d' AWS Lake Formation autorisations sur les bases de données et les tables du catalogue de données à AWS des comptes et organisations externes, et sur l'accès aux objets du catalogue de données partagés à partir de comptes externes. Pour plus d'informations, consultez la section [Accès entre comptes](#).

7 juillet 2020

[Intégration à Amazon Quick](#)

Ajout d'informations sur la façon d'accorder des autorisations Lake Formation aux utilisateurs de Quick Enterprise Edition afin qu'ils puissent accéder aux ensembles de données résidant dans des sites Amazon S3 enregistrés. Pour plus d'informations, consultez la section [Octroi d'autorisations au catalogue de données](#).

29 juin 2020

[Mises à jour des chapitres relatifs à la configuration et à la mise en route](#)

Réorganisation et amélioration des chapitres Configuration et Démarrage. Les autorisations recommandées Gestion des identités et des accès AWS (IAM) pour l'administrateur du lac de données ont été mises à jour.

27 février 2020

[Support pour AWS Key Management Service](#)

Ajout d'informations sur la façon dont le support de Lake Formation pour AWS Key Management Service (AWS KMS) simplifie la configuration de services intégrés pour lire et écrire des données chiffrées sur des sites Amazon Simple Storage Service (Amazon S3) enregistrés. Ajout d'informations sur la procédure d'enregistrement des sites Amazon S3 chiffrés avec AWS KMS keys. Pour de plus amples informations, veuillez consulter [the section called “Ajouter un emplacement Amazon S3 à votre lac de données”](#).

27 février 2020

[Mises à jour des plans et des politiques IAM des administrateurs de lacs de données](#)

Paramètres d'entrée clarifiés pour les plans de base de données incrémentiels. Mise à jour des politiques IAM requises pour un administrateur de lac de données.

20 décembre 2019

[Réécriture du chapitre sur la sécurité et révisions du chapitre sur la mise à niveau](#)

Amélioration des chapitres relatifs à la sécurité et à la mise à niveau.

29 octobre 2019

[La super autorisation remplace toutes les autorisations](#)

Mise à jour des chapitres sur la sécurité et la mise à niveau pour refléter le remplacement de l'autorisation All par Super.

10 octobre 2019

[Ajouts, corrections et clarifications](#)

A apporté des ajouts, des corrections et des clarifications en fonction des commentaires. Le chapitre sur la sécurité a été révisé. Mise à jour des chapitres sur la sécurité et la mise à niveau pour refléter le remplacement du groupe Everyone par IAMAllowedPrincipals .

11 septembre 2019

[Nouveau guide](#)

Il s'agit de la première version du Guide du développeur AWS Lake Formation .

8 août 2019

AWS Glossaire

Pour la AWS terminologie la plus récente, consultez le [AWS glossaire](#) dans la Glossaire AWS référence.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.