



Guide de l'utilisateur

AWS Clean Rooms



AWS Clean Rooms: Guide de l'utilisateur

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 Clean Rooms ?	1
Utilisez-vous pour la première fois AWS Clean Rooms ?	2
Comment AWS Clean Rooms fonctionne	2
Services connexes	2
AWS services	2
Services tiers	4
Accès AWS Clean Rooms	5
Tarification pour AWS Clean Rooms	5
Facturation pour AWS Clean Rooms	5
Règles d'analyse	6
Types de règles d'analyse	7
Règle d'analyse d'agrégation	9
Règle d'analyse des listes	31
Règle d'analyse personnalisée	40
Règle d'analyse des tables de mappage d'identifiants	47
AWS Clean Rooms Confidentialité différentielle	58
Confidentialité différentielle	59
Comment AWS Clean Rooms fonctionne la confidentialité différentielle	59
Politique de confidentialité différentielle	60
Fonctionnalités SQL	62
Conseils et exemples relatifs aux requêtes SQL	100
Limitations	102
AWS Clean Rooms ML	103
Terminologie AWS Clean Rooms ML	104
Comment AWS Clean Rooms ML fonctionne avec les AWS modèles	105
Comment AWS Clean Rooms ML fonctionne avec les modèles personnalisés	106
AWS modèles dans Clean Rooms ML	109
Modèles personnalisés dans Clean Rooms ML	118
Informatique cryptographique	126
Considérations	128
Types de fichiers et de données pris en charge	130
Noms de colonnes	136
Types de colonnes	137
Parameters	139

Indicateurs facultatifs	144
Requêtes avec C3R	147
Consignes	148
Con AWS Clean Rooms figuration	176
Inscrivez-vous pour AWS	176
Configurer les rôles de service pour AWS Clean Rooms	176
Création d'un utilisateur administrateur	177
Création d'un rôle IAM pour un membre de la collaboration	178
Créez un rôle de service pour lire les données d'Amazon S3	179
Création d'un rôle de service pour lire les données d'Amazon Athena	183
Créez un rôle de service pour lire les données de Snowflake	187
Création d'un rôle de service pour lire le code d'un compartiment S3 (rôle de modèle d'PySpark analyse)	191
Création d'un rôle de service pour écrire les résultats d'une PySpark tâche	194
Créez un rôle de service pour recevoir des résultats	198
Configuration des rôles de service pour AWS Clean Rooms ML	202
Configuration des rôles de service pour la modélisation des ressemblances	202
Configuration des rôles de service pour une modélisation personnalisée	217
Collaborations et adhésions	233
Création d'une collaboration	234
Création d'une collaboration pour les requêtes	234
Création d'une collaboration pour les requêtes et les tâches	246
Création d'une collaboration pour la modélisation ML	258
Création d'un abonnement et adhésion à une collaboration	268
Collaborations d'édition	275
Modifier le nom et la description de la collaboration	275
Mettre à jour le moteur d'analyse de collaboration	276
Désactiver le stockage des journaux	276
Modifier les paramètres des journaux de collaboration	277
Modifier les balises de collaboration	278
Modifier les tags d'adhésion	279
Ajouter un nouveau membre	280
Modifier les capacités des membres existants	280
Modifier les paramètres d'approbation automatique de la collaboration	280
Modifier les balises de table associées	280
Modifier les balises du modèle d'analyse	281

Modifier les balises de politique de confidentialité différentielles	281
Demandes de modification	282
Ajouter un nouveau membre à une collaboration	282
Mettre à jour les capacités des membres existants	283
Modifier les paramètres d'approbation automatique de la collaboration	284
Supprimer des collaborations	286
Afficher les collaborations	286
Inviter des membres	287
Surveillance des membres	287
Ajouter des membres	288
Supprimer des membres	289
Quitter une collaboration	290
Tableaux de données	292
Formats de données	293
Formats de données pris en charge pour les PySpark tâches	293
Formats de données pris en charge pour les requêtes SQL	293
Types de données pris en charge	294
Types de compression de fichiers pour AWS Clean Rooms	295
Chiffrement côté serveur pour AWS Clean Rooms	295
Apache Iceberg tables	296
Types de données pris en charge pour les tables Iceberg	297
Préparation de tableaux de données	298
Préparation des tables de données dans Amazon S3	298
Préparation de tables de données dans Amazon Athena	302
Préparation des tables de données dans Snowflake	304
Préparation de tables de données chiffrées	306
Étape 1 : Exécuter les prérequis	307
Étape 2 : Téléchargez le client de chiffrement C3R	308
Étape 3 : (Facultatif) Afficher les commandes disponibles dans le client de chiffrement C3R	309
Étape 4 : générer un schéma de chiffrement pour un fichier tabulaire	309
Étape 5 : Création d'une clé secrète partagée	317
Étape 6 : Stocker la clé secrète partagée dans une variable d'environnement	318
Étape 7 : Chiffrer les données	319
Étape 8 : vérifier le chiffrement des données	320
(Facultatif) Créez un schéma (utilisateurs avancés)	321

Décryptage des tables de données	331
Tables configurées	333
Création d'une table configurée	334
Source de données Amazon S3	335
Source de données Amazon Athena	338
Source de données Snowflake	341
Ajouter une règle d'analyse à une table configurée	346
Ajouter une règle d'analyse d'agrégation à une table (flux guidé)	347
Ajouter une règle d'analyse de liste à un tableau (flux guidé)	351
Ajouter une règle d'analyse personnalisée à un tableau (flux guidé)	353
Ajouter une règle d'analyse à une table (éditeur JSON)	358
Étapes suivantes	359
Associer une table configurée à une collaboration	360
Budget d'accès aux données	360
Associer une table configurée	361
Configuration d'un budget d'accès aux données	367
Afficher un budget d'accès aux données	368
Ajouter un budget d'accès aux données à une table associée existante	368
Modification d'un budget d'accès aux données	370
Supprimer un budget d'accès aux données	371
Ajouter une règle d'analyse de collaboration à une table configurée	373
Configuration d'une politique de confidentialité différentielle (facultatif)	374
Afficher les journaux d'utilisation différentiels de confidentialité	375
Modifier une politique de confidentialité différentielle	376
Supprimer une politique de confidentialité différentielle	377
Affichage des paramètres de confidentialité différentiels calculés	378
Afficher les tables et les règles d'analyse	379
Modification d'une table configurée	379
Modification des balises de tableau configurées	380
Modification de la règle d'analyse de table configurée	381
Suppression de la règle d'analyse de table configurée	381
Colonnes interdites dans le tableau configuré	382
Modification des associations de tables configurées	386
Dissociation des tables configurées	386
Résolution des entités AWS dans AWS Clean Rooms	388
Espaces de noms d'ID	389

Création et association d'un nouvel espace de noms d'ID	389
Associer un espace de noms d'ID existant	392
Modification des associations d'espaces de noms d'ID	395
Dissociation des associations d'espaces de noms d'ID	396
Tables de mappage d'identifiants	397
Création et remplissage d'une nouvelle table de mappage d'identifiants	398
Remplissage d'une table de mappage d'identifiants existante	406
Modification d'une table de mappage d'identifiants	407
Supprimer une table de mappage d'identifiants	408
Modèles d'analyse	410
Modèles d'analyse SQL	410
Création d'un modèle d'analyse SQL	411
Révision d'un modèle d'analyse SQL	414
PySpark modèles d'analyse	415
Sécurité	416
Limitations	416
Bonnes pratiques	418
Création d'un script utilisateur	419
Utilisation des paramètres dans les modèles PySpark d'analyse	422
Création d'un environnement virtuel (facultatif)	427
Stockage d'un script utilisateur et d'un environnement virtuel dans S3	428
Création d'un modèle PySpark d'analyse	430
Révision d'un modèle PySpark d'analyse	434
Modèles d' PySpark analyse de résolution des problèmes	437
Résolution des problèmes liés à votre code	438
La tâche du modèle d'analyse ne démarre pas	439
La tâche du modèle d'analyse démarre mais échoue pendant le traitement	440
La configuration de l'environnement virtuel échoue	441
Analyse	444
Exécution de requêtes SQL	444
Conditions préalables	446
Configuration des propriétés Spark pour les requêtes SQL	446
Interrogation de tables configurées	447
Interrogation des tables de mappage d'identifiants	466
Interrogation de tables configurées à l'aide d'un modèle d'analyse	469
Interrogation à l'aide du générateur d'analyse	486

Visualisation de l'impact de la confidentialité différentielle	492
Affichage des requêtes récentes	493
Affichage des détails de la requête	493
Exécution de PySpark tâches	494
Exécution d'une tâche à l'aide d'un modèle d'analyse	495
Afficher les offres d'emploi récentes	497
Affichage des détails de la tâche	498
Résultats de l'analyse	500
Réception des résultats de requêtes	501
Recevoir les résultats d'un travail	502
Modification des valeurs par défaut pour les paramètres des résultats de requête	503
Modification des valeurs par défaut pour les paramètres des résultats des tâches	505
Utilisation du résultat de la requête dans d'autres Services AWS	506
Modélisation personnalisée	507
Génération de jeux de données synthétiques garantissant une meilleure confidentialité	508
Considérations relatives à la génération de données synthétiques	509
Création et adhésion à la collaboration	513
Création d'une collaboration pour le machine learning	513
Rejoindre une collaboration	517
Données de formation contributives	519
Configuration d'un algorithme de modèle	525
Associer l'algorithme du modèle configuré	527
Création d'un canal d'entrée ML	533
Création d'un modèle entraîné	553
Utilisation de la formation progressive	557
Utilisation de la formation distribuée	561
Exportation d'artefacts du modèle	565
Exécution de l'inférence sur un modèle entraîné	567
Modélisation ML pour les fournisseurs de données de formation	570
Importation de données d'entraînement	571
Création d'un modèle similaire	572
Configuration d'un modèle similaire	574
Associer un modèle de similarité configuré	575
Mise à jour d'un modèle similaire configuré	575
Modélisation ML pour les fournisseurs de données de départ	577
Création d'un segment similaire	577

Exportation d'un segment similaire	593
Résolution des problèmes	594
Une ou plusieurs tables référencées par la requête ne sont pas accessibles par le rôle de service associé. Le table/role propriétaire doit accorder au rôle de service l'accès à la table.	594
L'un des ensembles de données sous-jacents possède un format de fichier non pris en charge.	594
Les résultats des requêtes ne sont pas ceux attendus lors de l'utilisation de l'informatique cryptographique pour Clean Rooms.	595
AWS Clean Rooms Spark SQL : données de partition manquantes	595
Sécurité	596
Protection des données	597
Chiffrement au repos	598
Chiffrement en transit	599
Chiffrement des données sous-jacentes	599
Stratégie de clé	599
Utilisation des rôles liés à un service	606
Autorisations de rôle liées à un service pour AWS Clean Rooms	606
Création d'un rôle lié à un service pour AWS Clean Rooms	607
Modification d'un rôle lié à un service pour AWS Clean Rooms	607
Supprimer un rôle lié à un service pour AWS Clean Rooms	607
Régions prises en charge pour les rôles AWS Clean Rooms liés à un service	608
Conservation des données	608
Bonnes pratiques	609
Les meilleures pratiques avec AWS Clean Rooms	609
Bonnes pratiques d'utilisation des règles d'analyse dans AWS Clean Rooms	610
Gestion de l'identité et des accès	612
Public ciblé	612
Authentification par des identités	613
Gestion de l'accès à l'aide de politiques	615
Comment AWS Clean Rooms fonctionne avec IAM	617
Exemples de politiques basées sur l'identité	623
AWS politiques gérées	626
Résolution des problèmes	634
Prévention du problème de l'adjoint confus entre services	636
Comportements IAM pour AWS Clean Rooms ML	639
Comportements IAM pour les modèles personnalisés Clean Rooms ML	643

Validation de conformité	644
Résilience	645
Sécurité de l'infrastructure	645
Sécurité du réseau	646
AWS PrivateLink	646
Considérations	647
Création d'un point de terminaison d'interface	647
Contrôle	648
Connexion à une analyse AWS Clean Rooms	649
Recevoir les journaux des requêtes et des tâches	650
Actions recommandées pour les journaux de requêtes et de tâches	651
Surveillance détaillée avec CloudWatch AWS Clean Rooms	652
Types de métriques	652
Dimensions des métriques	654
Qui peut accéder aux métriques	654
Tarification	655
CloudTrail journaux	655
AWS Clean Rooms informations dans CloudTrail	655
Comprendre les entrées du fichier AWS Clean Rooms journal	656
Exemples d' AWS Clean Rooms CloudTrail événements	656
Intégration AWS Clean Rooms dans EDAs	660
AWS Clean Rooms événements	661
AWS Clean Rooms Événements de routage	666
Référence détaillée des événements	667
Balisage de répartition des coûts	676
Ressources étiquetables pour la répartition des coûts	677
AWS CloudFormation ressources	678
AWS Clean Rooms et CloudFormation modèles	678
En savoir plus sur CloudFormation	678
Quotas	680
AWS Clean Rooms quotas	680
AWS Clean Rooms limites des paramètres de ressources	694
Quotas de machine learning d'AWS Clean Rooms	695
Limitation des quotas de l'API Clean Rooms ML	727
Historique de la documentation	734
Glossaire	751

Règle d'analyse d'agrégation	751
Règles d'analyse	751
Modèle d'analyse	751
Client de chiffrement C3R	752
Colonne en texte clair	752
Collaboration	752
Créateur de collaboration	752
Table configurée	753
Règle d'analyse personnalisée	753
Déchiffrement	753
Confidentialité différentielle	754
Chiffrement	754
Colonne d'empreintes digitales	754
Méthode de workflow de mappage des identifiants	754
Table de mappage des identifiants	754
Règle d'analyse des tables de mappage d'identifiants	755
Workflow de mappage des identifiants	755
Espace de noms ID	755
Association d'espaces de noms ID	755
Tâche	755
Règle d'analyse des listes	756
Modèle Lookalike	756
Segment similaire	756
Membre	756
Membre pouvant poser des questions	756
Membre capable d'exécuter des requêtes et des tâches	757
Membre pouvant recevoir les résultats	757
Membre payant les frais de calcul des requêtes	757
Membre payant les frais de recherche et de calcul des tâches	757
Membres	758
Colonne étanche	758
Données sur les semences	758
Moteur d'analyse Spark	758
Query	759
.....	dcclx

Qu'est-ce que c'est AWS Clean Rooms ?

AWS Clean Rooms vous permet, à vous et à vos partenaires, d'analyser et de collaborer sur vos ensembles de données collectifs afin d'obtenir de nouvelles informations sans révéler les données sous-jacentes les uns aux autres. AWS Clean Rooms est un espace de travail collaboratif sécurisé, dans lequel vous pouvez créer vos propres salles blanches en quelques minutes et analyser vos ensembles de données collectifs en quelques étapes seulement. Vous choisissez les partenaires avec lesquels vous souhaitez collaborer, sélectionnez leurs ensembles de données et configurez des contrôles renforçant la confidentialité pour ces partenaires.

Avec AWS Clean Rooms, vous pouvez collaborer avec des milliers d'entreprises qui l'utilisent déjà AWS. La collaboration ne nécessite pas de déplacer des données AWS ou de les charger vers un autre fournisseur de services cloud. Lorsque vous exécutez des requêtes ou des tâches, AWS Clean Rooms lisez les données à partir de leur emplacement d'origine et appliquez des règles d'analyse intégrées pour vous aider à garder le contrôle sur ces données.

AWS Clean Rooms fournit des contrôles d'accès aux données intégrés et des contrôles d'assistance à l'audit que vous pouvez configurer. Ces contrôles incluent :

- [Règles d'analyse](#) pour restreindre les requêtes SQL et fournir des contraintes de sortie.
- [L'informatique cryptographique permet de Clean Rooms crypter les données, même pendant le traitement des requêtes](#), afin de respecter les politiques strictes de traitement des données.
- Des [journaux d'analyse](#) pour examiner les requêtes et les tâches AWS Clean Rooms et aider à soutenir les audits.
- [Confidentialité différentielle](#) pour protéger contre les tentatives d'identification des utilisateurs. AWS Clean Rooms La confidentialité différentielle est une fonctionnalité entièrement gérée qui protège la confidentialité de vos utilisateurs grâce à des techniques mathématiques et à des commandes intuitives que vous pouvez appliquer en quelques étapes.
- [AWS Clean Rooms ML](#) permet à deux parties d'identifier des utilisateurs similaires dans leurs données sans avoir à partager leurs données entre elles. La première partie crée et configure un modèle similaire à partir de ses données d'entraînement. Les données de départ sont ensuite transmises à la collaboration pour créer un segment similaire aux données d'entraînement.

La vidéo suivante explique plus en détail AWS Clean Rooms.

[AWS Clean Rooms](#)

Utilisez-vous pour la première fois AWS Clean Rooms ?

Si vous utilisez ce service pour la première fois AWS Clean Rooms, nous vous recommandons de commencer par lire les sections suivantes :

- [Comment AWS Clean Rooms fonctionne](#)
- [Accès AWS Clean Rooms](#)
- [Con AWS Clean Rooms figuration](#)
- [AWS Clean Rooms Glossaire](#)

Comment AWS Clean Rooms fonctionne

Dans AWS Clean Rooms, vous créez une collaboration et ajoutez celle Comptes AWS que vous souhaitez inviter, ou vous créez un abonnement pour rejoindre une collaboration à laquelle vous avez été invité. Vous liez ensuite les ressources de données nécessaires à votre cas d'utilisation : tables configurées pour les données d'événements, modèles configurés pour la modélisation ML ou espaces de noms d'identification pour la résolution d'entités. Vous avez la possibilité de créer ou d'approuver des modèles d'analyse afin de convenir à l'avance des requêtes et des tâches exactes que vous souhaitez autoriser dans le cadre d'une collaboration. Enfin, vous analysez les données conjointes en exécutant des requêtes ou des PySpark tâches SQL sur les tables configurées, en effectuant la résolution des entités dans les tables de mappage d'identifiants ou en utilisant la modélisation ML pour générer des segments d'audience similaires.

Le schéma suivant montre comment AWS Clean Rooms cela fonctionne.

Services connexes

AWS services

Les éléments suivants Services AWS sont liés à AWS Clean Rooms :

- Amazon Athena

Les membres de la collaboration peuvent stocker les données qu'ils introduisent AWS Clean Rooms sous forme de AWS Glue Data Catalog vues dans Amazon Athena. Pour plus d'informations, consultez les rubriques suivantes :

Pour plus d'informations, consultez les rubriques suivantes :

[Préparation des tables de données pour les requêtes dans AWS Clean Rooms](#)

[Création d'une table configurée — Source de données Amazon Athena](#)

[Qu'est-ce qu'Amazon Athena ?](#) dans le guide de l'utilisateur d'Amazon Athena

- CloudFormation

Créez les ressources suivantes dans CloudFormation : collaborations, tables configurées, associations de tables configurées et adhésions

Pour de plus amples informations, veuillez consulter [Création de AWS Clean Rooms ressources avec AWS CloudFormation](#).

- AWS CloudTrail

AWS Clean Rooms Utilisez-le avec CloudTrail les journaux pour améliorer votre analyse de Service AWS l'activité.

Pour de plus amples informations, veuillez consulter [Journalisation des appels AWS Clean Rooms d'API à l'aide AWS CloudTrail](#).

- Résolution des entités AWS

Utilisez AWS Clean Rooms avec Résolution des entités AWS pour effectuer la résolution d'entités.

Pour de plus amples informations, veuillez consulter [Résolution des entités AWS dans AWS Clean Rooms](#).

- AWS Glue

Les membres de la collaboration peuvent créer AWS Glue des tables à partir de leurs données dans Amazon S3 pour les utiliser dans AWS Clean Rooms.

Pour plus d'informations, consultez les rubriques suivantes :

[Préparation des tables de données pour les requêtes dans AWS Clean Rooms](#)

[Qu'est-ce que AWS Glue ?](#) dans le Guide du développeur AWS Glue

- Amazon Simple Storage Service (Amazon S3)

Les membres de la collaboration peuvent stocker les données qu'ils AWS Clean Rooms introduisent dans Amazon S3.

Pour plus d'informations, consultez les rubriques suivantes :

[Préparation des tables de données pour les requêtes dans AWS Clean Rooms](#)

[Création d'une table configurée — Source de données Amazon S3](#)

[Qu'est-ce qu'Amazon S3](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.

- AWS Secrets Manager

Les membres de la collaboration peuvent créer des secrets pour accéder aux données stockées dans Snowflake et les lire.

Pour plus d'informations, consultez les rubriques suivantes :

[Créez un rôle de service pour lire les données de Snowflake](#)

[Préparation des tables de données pour les requêtes dans AWS Clean Rooms](#)

[Qu'est-ce que AWS Secrets Manager ?](#) dans le Guide de l'utilisateur AWS Secrets Manager

Services tiers

Le service tiers suivant est lié à AWS Clean Rooms :

- Snowflake

Les membres de la collaboration peuvent stocker les données qu'ils introduisent AWS Clean Rooms dans un entrepôt Snowflake.

Pour plus d'informations, consultez les rubriques suivantes :

[Préparation des tables de données pour les requêtes dans AWS Clean Rooms](#)

[Création d'une table configurée — Source de données Snowflake](#)

Accès AWS Clean Rooms

Vous pouvez y accéder AWS Clean Rooms via les options suivantes :

- Directement via la AWS Clean Rooms console à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
- Par programmation via l'API. AWS Clean Rooms Pour plus d'informations, consultez la [Référence des API AWS Clean Rooms](#).

Tarification pour AWS Clean Rooms

Pour de plus amples informations sur la tarification, veuillez consulter [AWS Clean Rooms Pricing](#) (français non garanti).

Note

Pour les membres de la collaboration qui ont associé des données stockées dans Snowflake, leur fournisseur d'entrepôt de données ou leur fournisseur de cloud respectif vous facturera à la fois la sortie des données et le calcul chaque fois qu'une requête utilisant des données stockées dans ces emplacements est exécutée.

Facturation pour AWS Clean Rooms

AWS Clean Rooms permet au créateur de la collaboration de désigner le membre qui prend en charge les coûts de calcul des requêtes ou des tâches dans le cadre de la collaboration.

Dans la plupart des cas, le [membre autorisé à effectuer une requête](#) et le [membre payant les frais de calcul des requêtes](#) sont les mêmes. Toutefois, si le membre autorisé à effectuer des requêtes et le membre payant les frais de calcul des requêtes sont différents, alors, lorsque le membre habilité à effectuer des requêtes exécute des requêtes sur sa propre ressource d'adhésion, la ressource d'adhésion du membre payant les coûts de calcul des requêtes est facturée.

Le membre qui paie les frais de calcul des requêtes ne voit aucun événement lié aux requêtes exécutées dans son historique des CloudTrail événements, car le payeur n'est ni celui qui exécute les requêtes ni le propriétaire de la ressource sur laquelle les requêtes sont exécutées. Cependant, le payeur perçoit des frais sur ses ressources d'adhésion pour toutes les requêtes exécutées par le membre qui peut exécuter des requêtes dans le cadre de la collaboration.

Pour plus d'informations sur la façon de créer une collaboration et de configurer le membre payant les coûts de calcul des requêtes, consultez [Création d'une collaboration](#).

Règles d'analyse dans AWS Clean Rooms

Dans le cadre de l'activation d'une table à des AWS Clean Rooms fins d'analyse de collaboration, le membre de la collaboration doit configurer une règle d'analyse.

Une règle d'analyse est un contrôle renforçant la confidentialité que chaque propriétaire de données met en place sur une table configurée. Une règle d'analyse détermine la manière dont la table configurée peut être analysée.

La règle d'analyse est un contrôle au niveau du compte sur la table configurée (une ressource au niveau du compte) et est appliquée dans toute collaboration où la table configurée est associée. Si aucune règle d'analyse n'est configurée, la table configurée peut être associée à des collaborations, mais elle ne peut pas être interrogée. Les requêtes peuvent uniquement faire référence à des tables configurées avec le même type de règle d'analyse.

Pour configurer une règle d'analyse, vous devez d'abord sélectionner un type d'analyse, puis spécifier la règle d'analyse. Pour les deux étapes, vous devez prendre en compte le cas d'utilisation que vous souhaitez activer et la manière dont vous souhaitez protéger vos données sous-jacentes.

AWS Clean Rooms applique les contrôles les plus restrictifs à toutes les tables configurées référencées dans une requête.

Les exemples suivants illustrent les contrôles restrictifs.

Exemple Contrôle restrictif : contrainte de sortie

- Le collaborateur A a une contrainte de sortie sur la colonne d'identificateur de 100.
- Le collaborateur B a une contrainte de sortie sur la colonne d'identificateur de 150.

Une requête d'agrégation qui fait référence aux deux tables configurées nécessite au moins 150 valeurs distinctes d'identifier dans une ligne de sortie pour qu'elle soit affichée dans la sortie de la requête. Le résultat de la requête n'indique pas que les résultats sont supprimés en raison de la contrainte de sortie.

Exemple Contrôle restrictif : modèle d'analyse non approuvé

- Le collaborateur A a autorisé un modèle d'analyse avec une requête qui fait référence aux tables configurées du collaborateur A et du collaborateur B dans leur règle d'analyse personnalisée.
- Le collaborateur B n'a pas autorisé le modèle d'analyse.

Le collaborateur B n'ayant pas autorisé le modèle d'analyse, le membre autorisé à effectuer une requête ne peut pas exécuter ce modèle d'analyse.

Types de règles d'analyse

Il existe trois types de règles d'analyse : les règles d'[agrégation](#), les règles de [liste](#) et les règles [personnalisées](#). Les tableaux suivants comparent les types de règles d'analyse. Chaque type comporte une section distincte qui décrit la spécification de la règle d'analyse.

Note

Il existe un type de règle d'analyse appelé règle d'analyse de table de mappage d'identifiants. Cependant, cette règle d'analyse est gérée par AWS Clean Rooms et ne peut pas être modifiée. Pour de plus amples informations, veuillez consulter [Règle d'analyse des tables de mappage d'identifiants](#).

Les sections suivantes décrivent les cas d'utilisation et les contrôles pris en charge pour chaque type de règle d'analyse.

Cas d'utilisation pris en charge

Les tableaux suivants présentent un résumé comparatif des cas d'utilisation pris en charge pour chaque type de règle d'analyse.

Cas d'utilisation	Agrégation	List	Personnalisé
Analyses prises en charge	Requêtes qui regroupent des statistiques à l'aide des fonctions COUNT, SUM et AVG	Requêtes qui produisent des listes au niveau des lignes indiquant le	Toute analyse personnalisée, à condition que le modèle d'analyse ou le créateur de

Cas d'utilisation	Agrégation	List	Personnalisé
	selon des dimensions facultatives	chevauchement entre plusieurs tables	l'analyse aient été revus et autorisés
Cas d'utilisation courants	Analyse des segments, mesure, attribution	Enrichissement, création de segments	Attribution directe, analyses incrémentielles, découverte de l'audience
Constructions SQL	• Déclarations JOIN : INNER JOIN • Fonctions d'agrégation : COUNT/COUNT DISTINCT, SUM/SUM DISTINCT et AVG • Fonctions scalaires : sous-ensemble limité	• Déclarations JOIN : INNER JOIN • Fonctions scalaires : Aucune	La plupart des fonctions et constructions SQL sont disponibles avec la commande SELECT
Sous-requêtes et expressions de table communes () CTEs	Non	Non	Oui
Modèles d'analyse	Non	Non	Oui

Contrôles pris en charge

Les tableaux suivants présentent un résumé comparatif de la manière dont chaque type de règle d'analyse protège vos données sous-jacentes.

Contrôle	Agrégation	List	Personnalisé
Mécanisme de commande	Contrôler la manière dont les données de la table peuvent être	Contrôler la manière dont les données de la table peuvent être	Contrôler les requêtes autorisées à s'exécuter sur la table

Contrôle	Agrégation	List	Personnalisé
	utilisées dans une requête (Par exemple, autorisez COUNT et SUM de la colonne hashed_email.)	utilisées dans une requête (Par exemple, autorisez l'utilisation de la colonne hashed_email uniquement pour la connexion.)	(Par exemple, autorisez uniquement les requêtes définies dans les modèles d'analyse « Requête personnalisée 1 ».)
Techniques intégrées d'amélioration de la confidentialité	• Match à l'aveugle • Agrégation requise • Seuil d'agrégation minimum >= • 2 Structure de requête prédéfinie	• Match à l'aveugle • Chevauchement requis • Structure de requête prédéfinie • Analyses supplémentaires autorisées	• Confidentialité différentielle • Colonnes de sortie non autorisées
Vérifiez la requête avant de pouvoir l'exécuter	Non	Non	Oui, en utilisant des modèles d'analyse

Pour plus d'informations sur les règles d'analyse disponibles dans AWS Clean Rooms, consultez les rubriques suivantes.

- [Règle d'analyse d'agrégation](#)
- [Règle d'analyse des listes](#)
- [Règle d'analyse personnalisée dans AWS Clean Rooms](#)

Règle d'analyse d'agrégation

Dans AWS Clean Rooms, une règle d'analyse d'agrégation génère des statistiques agrégées à l'aide des fonctions COUNT, SUM, and/or AVG et de dimensions facultatives. Lorsque la règle d'analyse

d'agrégation est ajoutée à une table configurée, elle permet au membre habilité à effectuer des requêtes sur la table configurée.

La règle d'analyse d'agrégation prend en charge les cas d'utilisation tels que la planification de campagnes, la portée médiatique, la mesure de fréquence et l'attribution.

La structure et la syntaxe de requête prises en charge sont définies dans [Structure et syntaxe des requêtes d'agrégation](#).

Les paramètres de la règle d'analyse, définis dans [Règle d'analyse d'agrégation : contrôles des requêtes](#), incluent les contrôles de requête et les contrôles de résultats de requête. Ses contrôles de requête incluent la possibilité d'exiger qu'une table configurée soit jointe à au moins une table configurée appartenant au membre qui peut effectuer une requête, directement ou de manière transitive. Cette exigence vous permet de vous assurer que la requête est exécutée à l'intersection (INNERJOIN) de votre table et de la leur.

Structure et syntaxe des requêtes d'agrégation

Les requêtes sur les tables dotées d'une règle d'analyse d'agrégation doivent respecter la syntaxe suivante.

```
--select_aggregate_function_expression
SELECT
aggregation_function(column_name) [[AS] column_alias ] [, ...]

--select_grouping_column_expression
[, {column_name|scalar_function(arguments)} [[AS] column_alias ]][, ...]

--table_expression
FROM table_name [[AS] table_alias ]
    [[INNER] JOIN table_name [[AS] table_alias] ON join_condition] [...]

--where_expression
[WHERE where_condition]

--group_by_expression
[GROUP BY {column_name|scalar_function(arguments)}, ...]]

--having_expression
[HAVING having_condition]

--order_by_expression
```

```
[ORDER BY {column_name|scalar_function(arguments)} [{ASC|DESC}]] [,...]]
```

Le tableau suivant explique chaque expression répertoriée dans la syntaxe précédente.

Expression	Définition	Exemples
<i>select_aggregate_function_expression</i>	Une liste séparée par des virgules contenant les expressions suivantes : • <code>select_aggregation_function_expression</code> • <code>select_aggregate_expression</code> Note Il doit y en avoir au moins un <code>select_aggregation_function_expression</code> dans le <code>select_aggregate_expression</code> .	<code>SELECT SUM(PRICE), user_segment</code>
<i>select_aggregation_function_expression</i>	Une ou plusieurs fonctions d'agrégation prises en charge sont appliquées à une ou plusieurs colonnes. Seules les colonnes sont autorisées comme arguments des fonctions d'agrégation.	<code>AVG(PRICE)</code> <code>COUNT(DISTINCT user_id)</code>

Expression	Définition	Exemples
	 Note Il doit y en avoir au moins un select_aggregation_function_expression dans leselect_aggregate_expression .	

Expression	Définition	Exemples
<i>select_grouping_column_expression</i>	Expression qui peut contenir n'importe quelle expression utilisant les éléments suivants : • Nom des colonnes de la table • Fonctions scalaires prises en charge • Littéraux de chaîne • Littéraux numériques	TRUNC(timestampColumn) UPPER(campaignName)

 **Note**

select_aggregate_expression peut créer un alias pour les colonnes avec ou sans le AS paramètre. Pour plus d'informations, consultez la [référence AWS Clean Rooms SQL](#).

Expression	Définition	Exemples
<i>table_expression</i>	Table, ou jointure de tables, reliant des expressions conditionnelles de jointure à <code>join_condition</code> . <code>join_condition</code> renvoie une valeur booléenne. Les <code>table_expression</code> supports : • Un JOIN type spécifique (INNERJOIN) • La condition de comparaison de l'égalité dans un <code>join_condition</code> (=) • Opérateurs logiques (AND,OR).	<pre>FROM consumer_table INNER JOIN provider_ table ON consumer_table.id entifier1 = provider_ table.identified1 AND consumer_table .identified2 = provider_table.id entifier2</pre>

Expression	Définition	Exemples
<i>where_expression</i>	Expression conditionnelle qui renvoie une valeur booléenne . Il peut être composé des éléments suivants : • Nom des colonnes de la table • Fonctions scalaires prises en charge • Opérateurs mathématiques • Littéraux de chaîne • Littéraux numériques Les conditions de comparaison prises en charge sont (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL). Les opérateurs logiques pris en charge sont (AND, OR). L'option where_expression est facultative.	<pre>WHERE where_condition</pre> <pre>WHERE price > 100</pre> <pre>WHERE TRUNC(timestampColumn) = '1/1/2022'</pre> <pre>WHERE timestampColumn = timestampColumn2 - 14</pre>
<i>group_by_expression</i>	Liste d'expressions séparées par des virgules qui répondent aux exigences du. select_grouping_column_expression	<pre>GROUP BY TRUNC(timestampColumn), UPPER(campaignName), segment</pre>

Expression	Définition	Exemples
<i>having_expression</i>	Expression conditionnelle qui renvoie une valeur booléenne . Ils disposent d'une fonction d'agrégation prise en charge appliquée à une seule colonne (par exemple, <code>SUM(price)</code>) et sont comparés à un littéral numérique. Les conditions prises en charge sont (<code>=</code>, <code>></code>, <code><</code>, <code><=</code>, <code>>=</code>, <code><></code>, <code>!=</code>). Les opérateurs logiques pris en charge sont (<code>AND</code>, <code>OR</code>). L'option <code>having_expression</code> est facultative.	<pre>HAVING SUM(SALES) > 500</pre>

Expression	Définition	Exemples
<i>order_by_expression</i>	Liste d'expressions séparées par des virgules qui est compatible avec les mêmes exigences définies dans la section <code>select_aggregate_expression</code> définie précédemment. L'option <code>order_by_expression</code> est facultative.  Note <code>order_by_expression</code> autorisations ASC et DESC paramètres. Pour plus d'informations, consultez la section Paramètres ASC DESC dans le manuel AWS Clean Rooms SQL Reference.	<pre>ORDER BY SUM(SALES), UPPER(campaignName)</pre>

En ce qui concerne la structure et la syntaxe des requêtes d'agrégation, tenez compte des points suivants :

- Les commandes SQL autres que `SELECT` sont pas prises en charge.
- Les sous-requêtes et les expressions de table communes (par exemple, `WITH`) ne sont pas prises en charge.
- Les opérateurs qui combinent plusieurs requêtes (par exemple, `UNION`) ne sont pas pris en charge.

- TOPLIMIT, et les OFFSET paramètres ne sont pas pris en charge.

Règle d'analyse d'agrégation : contrôles des requêtes

Grâce aux commandes de requête d'agrégation, vous pouvez contrôler la manière dont les colonnes de votre table sont utilisées pour interroger la table. Par exemple, vous pouvez contrôler quelle colonne est utilisée pour la jointure, quelle colonne peut être comptée ou quelle colonne peut être utilisée dans WHERE les instructions.

Les sections suivantes expliquent chaque contrôle.

Rubriques

- [Contrôles d'agrégation](#)
- [Commandes de jointure](#)
- [Contrôles dimensionnels](#)
- [Fonctions scalaires](#)

Contrôles d'agrégation

À l'aide des contrôles d'agrégation, vous pouvez définir les fonctions d'agrégation à autoriser et les colonnes auxquelles elles doivent être appliquées. Les fonctions d'agrégation peuvent être utilisées dans les ORDER BY expressions SELECTHAVING, et.

Contrôle	Définition	Usage
aggregateColumns	Colonnes de colonnes de table configurées que vous autorisez à utiliser dans les fonctions d'agrégation.	aggregateColumns peut être utilisé dans une fonction d'agrégation dans les ORDER BY expressions SELECTHAVING,, et. Certains aggregateColumns peuvent également être classés dans la catégorie « A » joinColumn (définis ultérieurement).

Contrôle	Définition	Usage
		Given ne aggregate Column peut pas également être classé dans la catégorie dimensionColumn (défini ultérieurement).
function	Les fonctions COUNT, SUM et AVG que vous autorisez à utiliser en plus de aggregate Columns .	function peut être appliqué à un aggregateColumns objet qui lui est associé.

Commandes de jointure

Une JOIN clause est utilisée pour combiner les lignes de deux tables ou plus, sur la base d'une colonne associée entre elles.

Vous pouvez utiliser les commandes de jointure pour contrôler la manière dont votre table peut être jointe aux autres tables de `table_expression`. AWS Clean Rooms uniquement des supports INNER JOIN. INNER JOIN les instructions ne peuvent utiliser que des colonnes explicitement classées comme telles `joinColumn` dans votre règle d'analyse, sous réserve des contrôles que vous définissez.

Ils INNER JOIN doivent opérer à `joinColumn` partir d'une table configurée et `joinColumn` à partir d'une autre table configurée dans la collaboration. Vous décidez quelles colonnes de votre tableau peuvent être utilisées `joinColumn`.

Chaque condition de correspondance contenue dans la ON clause est requise pour utiliser la condition de comparaison d'égalité (=) entre deux colonnes.

Plusieurs conditions de correspondance au sein d'une même ON clause peuvent être les suivantes :

- Combiné à l'aide de l'opérateur AND logique
- Séparé à l'aide de l'opérateur OR logique

Note

Toutes les JOIN conditions de match doivent correspondre à une ligne de chaque côté du JOIN. Toutes les conditions connectées par un opérateur OR ou un opérateur AND logique doivent également respecter cette exigence.

Voici un exemple de requête avec un opérateur AND logique.

```
SELECT some_col, other_col
FROM table1
JOIN table2
ON table1.id = table2.id AND table1.name = table2.name
```

Voici un exemple de requête avec un opérateur OR logique.

```
SELECT some_col, other_col
FROM table1
JOIN table2
ON table1.id = table2.id OR table1.name = table2.name
```

Contrôle	Définition	Usage
joinColumns	Les colonnes (le cas échéant) que vous souhaitez autoriser le membre autorisé à effectuer une requête à utiliser dans la INNER JOIN déclaration.	Un spécifique joinColumn peut également être classé dans la catégorie aggregate Column (voir Contrôles d'agrégation). La même colonne ne peut pas être utilisée à la fois comme joinColumn et dimension Columns (voir plus loin). À moins qu'il n'ait également été classé dans la catégorie unaggregateColumn , a ne joinColumn peut être

Contrôle	Définition	Usage
		utilisé dans aucune autre partie de la requête que dans le INNERJOIN.
<code>joinRequired</code>	Déterminez si vous avez besoin INNER JOIN d'une table configurée de la part du membre qui peut effectuer la requête.	Si vous activez ce paramètre , un INNER JOIN est requis. Si vous n'activez pas ce paramètre, un INNER JOIN est facultatif. En supposant que vous activiez ce paramètre, le membre autorisé à effectuer une requête doit inclure une table qu'il possède dans le INNERJOIN. Ils doivent joindre JOIN votre table à la leur, soit directement, soit de manière transitive (c'est-à-dire joindre leur table à une autre table, elle-même jointe à la vôtre).

Voici un exemple de transitivité.

```
ON
my_table.identifer = third_party_table.identifier
....
ON
third_party_table.identifier = member_who_can_query_table.id
```

Note

Le membre qui peut effectuer une requête peut également utiliser le `joinRequired` paramètre. Dans ce cas, la requête doit joindre sa table à au moins une autre table.

Contrôles dimensionnels

Les contrôles de dimension contrôlent la colonne le long de laquelle les colonnes d'agrégation peuvent être filtrées, groupées ou agrégées.

Contrôle	Définition	Usage
<code>dimensionColumns</code>	Les colonnes (le cas échéant) que vous autorisez le membre autorisé à effectuer une requête à utiliser dans <code>SELECT</code> WHERE, <code>GROUPBY</code> , et <code>ORDERBY</code> .	A <code>dimensionColumn</code> peut être utilisé dans <code>SELECT (select_grouping_column_expression)WHERE, GROUPBY, et ORDERBY</code>. La même colonne ne peut pas être à la fois a <code>dimensionColumn</code> <code>joinColumn</code> , a et and/or an <code>aggregateColumn</code> .

Fonctions scalaires

Les fonctions scalaires contrôlent les fonctions scalaires qui peuvent être utilisées sur les colonnes de dimension.

Contrôle	Définition	Usage
<code>scalarFunctions</code>	Les fonctions scalaires qui peuvent être utilisées <code>dimensionColumns</code> dans la requête.	Spécifie les fonctions scalaires (le cas échéant) auxquelles vous autorisez (par exemple <code>CAST</code>) l'<code>dimensionColumns</code> application. Les fonctions scalaires ne peuvent pas être utilisées par-dessus d'autres fonctions ou au sein d'autres fonctions.

Contrôle	Définition	Usage
		Les arguments des fonctions scalaires peuvent être des colonnes, des chaînes littérales ou des littéraux numériques.

Les fonctions scalaires suivantes sont prises en charge :

- Fonctions mathématiques : ABS, PLAFOND, PLANCHER, BOIS, LN, ROND, SQRT
- Fonctions de formatage des types de données — CAST, CONVERT, TO_CHAR, TO_DATE, TO_NUMBER, TO_TIMESTAMP
- Fonctions de chaîne : LOWER, UPPER, TRIM, RTRIM, SUBSTRING
 - Pour RTRIM, les jeux de caractères personnalisés à découper ne sont pas autorisés.
- Expressions conditionnelles — COALESCE
- Fonctions de date : EXTRACT, GETDATE, CURRENT_DATE, DATEADD
- Autres fonctions — TRUNC

Pour plus de détails, consultez la [référence AWS Clean Rooms SQL](#).

Règle d'analyse d'agrégation : contrôles des résultats des requêtes

Avec les contrôles des résultats des requêtes d'agrégation, vous pouvez contrôler les résultats renvoyés en spécifiant une ou plusieurs conditions que chaque ligne de sortie doit remplir pour être renvoyée. AWS Clean Rooms prend en charge les contraintes d'agrégation sous la forme de `COUNT (DISTINCT column) >= X`. Ce formulaire exige que chaque ligne agrège au moins X valeurs distinctes d'un choix dans votre table configurée (par exemple, un nombre minimum de `user_id` valeurs distinctes). Ce seuil minimum est automatiquement appliqué, même si la requête soumise elle-même n'utilise pas la colonne spécifiée. Ils sont appliqués collectivement sur chaque table configurée dans la requête à partir des tables configurées de chaque membre de la collaboration.

Chaque table configurée doit comporter au moins une contrainte d'agrégation dans sa règle d'analyse. Les propriétaires de tables configurées peuvent en ajouter plusieurs `columnName` et `minimum` les associer, et elles sont appliquées collectivement.

Contraintes d'agrégation

Les contraintes d'agrégation contrôlent les lignes renvoyées dans les résultats de la requête. Pour être renvoyée, une ligne doit respecter le nombre minimum de valeurs distinctes spécifié dans chaque colonne spécifiée dans la contrainte d'agrégation. Cette exigence s'applique même si la colonne n'est pas explicitement mentionnée dans la requête ou dans d'autres parties de la règle d'analyse.

Contrôle	Définition	Usage
columnName	Le <code>aggregateColumn</code> qui est utilisé dans la condition que chaque ligne de sortie doit remplir.	Il peut s'agir de n'importe quelle colonne de la table configurée.
minimum	Le nombre minimum de valeurs distinctes associées <code>aggregateColumn</code> que la ligne de sortie doit avoir (par exemple, <code>COUNT DISTINCT</code>) pour qu'elle soit renvoyée dans les résultats de la requête.	La valeur minimum doit être au moins égale à 2.

Structure des règles d'analyse d'agrégation

L'exemple suivant montre une structure prédéfinie pour une règle d'analyse d'agrégation.

Dans l'exemple suivant, *MyTable* fait référence à votre table de données. Vous pouvez remplacer chacune d'entre elles *user input placeholder* par vos propres informations.

```
{
  "aggregateColumns": [
    {
      "columnNames": [MyTable column names], "function": [Allowed Agg Functions]
    },
  ],
  "joinRequired": ["QUERY_RUNNER"],
  "joinColumns": [MyTable column names],
```

```
"dimensionColumns": [MyTable column names],
"scalarFunctions": [Allowed Scalar functions],
"outputConstraints": [
  {
    "columnName": [MyTable column names], "minimum": [Numeric value]
  },
]
}
```

Règle d'analyse d'agrégation - exemple

L'exemple suivant montre comment deux entreprises peuvent collaborer à AWS Clean Rooms l'aide de l'analyse d'agrégation.

L'entreprise A possède des données sur les clients et les ventes. L'entreprise A souhaite comprendre l'activité de retour de produits. L'entreprise B est l'un des détaillants de l'entreprise A et possède des données sur les retours. L'entreprise B possède également des attributs de segment relatifs aux clients qui sont utiles à l'entreprise A (par exemple, achat de produits connexes, utilisation du service client du détaillant). L'entreprise B ne souhaite pas fournir de données de retour client au niveau des lignes ni d'informations sur les attributs. L'entreprise B souhaite uniquement activer un ensemble de requêtes pour que l'entreprise A obtienne des statistiques agrégées sur les clients qui se chevauchent à un seuil d'agrégation minimum.

L'entreprise A et l'entreprise B décident de collaborer afin que l'entreprise A puisse comprendre l'activité de retour des produits et fournir de meilleurs produits à l'entreprise B et à d'autres canaux.

Pour créer la collaboration et exécuter une analyse d'agrégation, les entreprises procèdent comme suit :

1. L'entreprise A crée une collaboration et crée une adhésion. La collaboration a la société B comme autre membre de la collaboration. L'entreprise A active la journalisation des requêtes dans la collaboration, et elle permet la journalisation des requêtes dans son compte.
2. L'entreprise B crée une adhésion à la collaboration. Il permet la journalisation des requêtes dans son compte.
3. La société A crée une table configurée pour les ventes.
4. La société A ajoute la règle d'analyse d'agrégation suivante au tableau des ventes configuré.

```
{
  "aggregateColumns": [
    {
```

```

    "columnNames": [
      "identifiant"
    ],
    "function": "COUNT_DISTINCT"
  },
  {
    "columnNames": [
      "purchases"
    ],
    "function": "AVG"
  },
  {
    "columnNames": [
      "purchases"
    ],
    "function": "SUM"
  }
],
"joinColumns": [
  "hashedemail"
],
"dimensionColumns": [
  "demoseg",
  "purchasedate",
  "productline"
],
"scalarFunctions": [
  "CAST",
  "COALESCE",
  "TRUNC"
],
"outputConstraints": [
  {
    "columnName": "hashedemail",
    "minimum": 2,
    "type": "COUNT_DISTINCT"
  },
]
}

```

aggregateColumns— L'entreprise A souhaite compter le nombre de clients uniques entre les données de vente et les données de retours. L'entreprise A souhaite également additionner le nombre de purchases produits fabriqués pour le comparer au nombre de retours.

joinColumns— L'entreprise A souhaite utiliser pour faire correspondre identifier les clients à partir des données de vente aux clients à partir des données de retours. Cela aidera l'entreprise A à faire correspondre les retours aux bons achats. Cela aide également l'entreprise A à segmenter les clients qui se recoupent.

dimensionColumns— L'entreprise A filtre en dimensionColumns fonction du produit spécifique, compare les achats et les retours sur une certaine période, s'assure que la date de retour est postérieure à la date du produit et aide à segmenter les clients qui se recoupent.

scalarFunctions— L'entreprise A sélectionne une fonction CAST scalaire pour aider à mettre à jour les formats des types de données si nécessaire en fonction de la table configurée que l'entreprise A associe à la collaboration. Il ajoute également des fonctions scalaires pour aider à formater les colonnes si nécessaire.

outputConstraints— L'entreprise A définit des contraintes de sortie minimales. Il n'est pas nécessaire de restreindre les résultats car l'analyste est autorisé à voir les données au niveau des lignes depuis son tableau des ventes

 Note

L'entreprise A n'est pas incluse joinRequired dans la règle d'analyse. Cela permet à leur analyste d'interroger seul le tableau des ventes.

5. La société B crée une table de retours configurée.
6. La société B ajoute la règle d'analyse d'agrégation suivante à la table des retours configurés.

```
{
  "aggregateColumns": [
    {
      "columnNames": [
        "identifier"
      ],
      "function": "COUNT_DISTINCT"
    },
    {
      "columnNames": [
        "returns"
      ],
      "function": "AVG"
    }
  ]
}
```

```
    },
    {
      "columnNames": [
        "returns"
      ],
      "function": "SUM"
    }
  ],
  "joinColumns": [
    "hashedemail"
  ],
  "joinRequired": [
    "QUERY_RUNNER"
  ],
  "dimensionColumns": [
    "state",
    "popularpurchases",
    "customerserviceuser",
    "productline",
    "returndate"
  ],
  "scalarFunctions": [
    "CAST",
    "LOWER",
    "UPPER",
    "TRUNC"
  ],
  "outputConstraints": [
    {
      "columnName": "hashedemail",
      "minimum": 100,
      "type": "COUNT_DISTINCT"
    },
    {
      "columnName": "producttype",
      "minimum": 2,
      "type": "COUNT_DISTINCT"
    }
  ]
}
```

`aggregateColumns`— L'entreprise B permet à l'entreprise A de `returns` faire la somme pour comparer le nombre d'achats. Ils ont au moins une colonne d'agrégation car ils activent une requête d'agrégation.

`joinColumns`— L'entreprise B permet à l'entreprise A de se joindre à elle `identifier` pour faire correspondre les clients à partir des données de retour aux clients à partir des données de vente. `identifier` les données sont particulièrement sensibles et leur utilisation `joinColumn` garantit qu'elles ne seront jamais sorties dans une requête.

`joinRequired`— L'entreprise B exige que les requêtes sur les données de retour soient recoupées avec les données de vente. Ils ne veulent pas permettre à l'entreprise A d'interroger tous les individus de leur ensemble de données. Ils ont également convenu de cette restriction dans leur accord de collaboration.

`dimensionColumns`— L'entreprise B permet à l'entreprise A de filtrer et de regrouper par `statepopularpurchases`, et `customerserviceuser` qui sont des attributs uniques qui pourraient aider à effectuer l'analyse pour l'entreprise A. L'entreprise B permet à l'entreprise A d'utiliser `returndate` pour filtrer les résultats en fonction de `returndate` ce qui se produit après `purchasedate`. Grâce à ce filtrage, le résultat est plus précis pour évaluer l'impact du changement de produit.

`scalarFunctions`— La société B permet ce qui suit :

- `TRUNC` pour les dates
- `INFÉRIEUR` et `SUPÉRIEUR` au cas où ils `producttype` sont saisis dans un format différent dans leurs données
- `CAST` si l'entreprise A doit convertir les types de données des ventes pour qu'ils soient identiques aux types de données des retours

La société A n'active pas d'autres fonctions scalaires car elle ne pense pas qu'elles soient nécessaires pour les requêtes.

`outputConstraints`— L'entreprise B impose des contraintes de production minimales `hashedemail` afin de réduire la capacité à réidentifier les clients. Cela ajoute également une contrainte de sortie minimale afin `producttype` de réduire la capacité de réidentifier les produits spécifiques qui ont été renvoyés. Certains types de produits peuvent être plus dominants en fonction des dimensions de la sortie (par exemple, `state`). Leurs contraintes de sortie seront

toujours appliquées, quelles que soient les contraintes de sortie ajoutées par l'entreprise A à ses données.

7. L'entreprise A crée une table de vente associée à la collaboration.
8. L'entreprise B crée une association de tables de retours à la collaboration.
9. L'entreprise A exécute des requêtes, comme dans l'exemple suivant, pour mieux comprendre le nombre de retours dans l'entreprise B par rapport au total des achats par site en 2022.

```
SELECT
  companyB.state,
  SUM(companyB.returns),
  COUNT(DISTINCT companyA.hashemail)
FROM
  sales companyA
  INNER JOIN returns companyB ON companyA.identifiant = companyB.identifiant
WHERE
  companyA.purchasedate BETWEEN '2022-01-01' AND '2022-12-31' AND
  TRUNC(companyB.returndate) > companyA.purchasedate
GROUP BY
  companyB.state;
```

- 10 Les entreprises A et B examinent les journaux de requêtes. L'entreprise B vérifie que la requête est conforme à ce qui a été convenu dans l'accord de collaboration.

Résolution des problèmes liés aux règles d'analyse d'agrégation

Utilisez les informations présentées ici pour vous aider à diagnostiquer et à résoudre les problèmes courants liés à l'utilisation des règles d'analyse d'agrégation.

Problèmes

- [Ma requête n'a renvoyé aucun résultat](#)

Ma requête n'a renvoyé aucun résultat

Cela peut se produire en l'absence de résultats correspondants ou lorsque les résultats correspondants n'atteignent pas un ou plusieurs seuils d'agrégation minimaux.

Pour plus d'informations sur les seuils d'agrégation minimaux, consultez [Règle d'analyse d'agrégation - exemple](#).

Règle d'analyse des listes

Dans AWS Clean Rooms, une règle d'analyse de liste produit des listes au niveau des lignes indiquant le chevauchement entre la table configurée à laquelle elle est ajoutée et les tables configurées du membre qui peut effectuer la requête. Le membre habilité à effectuer des requêtes exécute des requêtes qui incluent une règle d'analyse de liste.

Le type de règle d'analyse de liste prend en charge les cas d'utilisation tels que l'enrichissement et la création d'audience.

Pour plus d'informations sur la structure de requête et la syntaxe prédéfinies pour cette règle d'analyse, consultez [Structure prédéfinie des règles d'analyse des listes](#).

Les paramètres de la règle d'analyse de liste, définis dans [Règle d'analyse des listes : contrôles des requêtes](#), comportent des contrôles de requête. Ses commandes de requête incluent la possibilité de sélectionner les colonnes qui peuvent être répertoriées dans la sortie. La requête doit comporter au moins une jointure avec une table configurée provenant du membre qui peut effectuer la requête, directement ou de manière transitive.

Il n'existe aucun contrôle des résultats de requête comme c'est le cas pour la [règle d'analyse d'agrégation](#).

Les requêtes de liste ne peuvent utiliser que des opérateurs mathématiques. Ils ne peuvent pas utiliser d'autres fonctions (telles que l'agrégation ou le scalaire).

Rubriques

- [Structure et syntaxe des requêtes de liste](#)
- [Règle d'analyse des listes : contrôles des requêtes](#)
- [Structure prédéfinie des règles d'analyse des listes](#)
- [Règle d'analyse des listes - exemple](#)

Structure et syntaxe des requêtes de liste

Les requêtes sur les tables dotées d'une règle d'analyse de liste doivent respecter la syntaxe suivante.

```
--select_list_expression  
SELECT DISTINCT column_name [[AS] column_alias ] [, ...]
```

```
--table_expression
FROM table_name [[AS] table_alias ]
    [[INNER] JOIN table_name [[AS] table_alias] ON join_condition] [...]

--where_expression
[WHERE where_condition]

--limit_expression
[LIMIT number]
```

Le tableau suivant explique chaque expression répertoriée dans la syntaxe précédente.

Expression	Définition	Exemples
<i>select_list_expression</i>	Liste séparée par des virgules contenant au moins un nom de colonne de table. Un DISTINCT paramètre est obligatoire. Note Ils <code>select_list_expression</code> peuvent alier les colonnes avec ou sans le AS paramètre. Pour plus d'informations, consultez la référence AWS Clean Rooms SQL.	SELECT DISTINCT segment
<i>table_expression</i>	Une table, ou une jointure de tables, <code>join_condition</code> à laquelle la connecter <code>join_condition</code> .	<pre>FROM consumer_table INNER JOIN provider_ table ON</pre>

Expression	Définition	Exemples
	<code>join_condition</code> renvoie un booléen. Les <code>table_expression</code> supports : • Un type de JOIN spécifique (INNERJOIN) • Les conditions de comparaison de l'égalité au sein d'un <code>join_condition</code> (=) • Opérateurs logiques (AND,OR).	<pre>consumer_table.identifier1 = provider_table.identifier1 AND consumer_table.identifier2 = provider_table.identifier2</pre>

Expression	Définition	Exemples
<i>where_expression</i>	Expression conditionnelle qui renvoie une valeur booléenne . Il peut être composé des éléments suivants : • Nom des colonnes de la table • Opérateurs mathématiques • Littéraux de chaîne • Littéraux numériques Les conditions de comparaison prises en charge sont (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL). Les opérateurs logiques pris en charge sont (AND, OR). L'option <code>where_expression</code> est facultative.	<pre>WHERE state + '_' + city = 'NY_NYC'</pre> <pre>WHERE timestampColumn = timestampColumn2 - 14</pre>
<i>limit_expression</i>	Cette expression doit prendre un entier positif. L'option <code>limit_expression</code> est facultative.	<code>LIMIT 100</code>

En ce qui concerne la structure et la syntaxe des requêtes de liste, tenez compte des points suivants :

- Les commandes SQL autres que SELECT ne sont pas prises en charge.
- Les sous-requêtes et les expressions de table communes (par exemple, WITH) ne sont pas prises en charge

- Les BY clauses GROUP BY HAVING, et ORDER ne sont pas prises en charge
- Le paramètre OFFSET n'est pas pris en charge

Règle d'analyse des listes : contrôles des requêtes

Avec les commandes de requête de liste, vous pouvez contrôler la manière dont les colonnes de votre table sont utilisées pour interroger la table. Par exemple, vous pouvez contrôler quelle colonne est utilisée pour la jointure ou quelle colonne peut être utilisée dans l'instruction et la WHERE clause SELECT.

Les sections suivantes expliquent chaque contrôle.

Rubriques

- [Commandes de jointure](#)
- [Contrôles de liste](#)

Commandes de jointure

Avec les commandes Join, vous pouvez contrôler la manière dont votre table peut être jointe aux autres tables de la table_expression. AWS Clean Rooms supporte uniquement INNER JOIN. Dans la règle d'analyse de liste, au moins un INNER JOIN est requis et le membre qui peut effectuer une requête doit inclure une table qu'il possède dans le INNER JOIN. Cela signifie qu'ils doivent joindre votre table à la leur, directement ou de manière transitionnelle.

Voici un exemple de transitivité.

```
ON
my_table.identifer = third_party_table.identifier
....
ON
third_party_table.identifier = member_who_can_query_table.id
```

INNERLes instructions JOIN ne peuvent utiliser que des colonnes explicitement classées comme telles joinColumn dans votre règle d'analyse.

Le INNER JOIN doit fonctionner sur une table joinColumn à partir de votre table configurée et joinColumn à partir d'une autre table configurée dans la collaboration. Vous décidez quelles colonnes de votre tableau peuvent être utiliséesjoinColumn.

Chaque condition de correspondance contenue dans la ON clause est requise pour utiliser la condition de comparaison d'égalité (=) entre deux colonnes.

Plusieurs conditions de correspondance au sein d'une ON clause peuvent être les suivantes :

- Combiné à l'aide de l'opérateur AND logique
- Séparé à l'aide de l'opérateur OR logique

 Note

Toutes les JOIN conditions de match doivent correspondre à une ligne de chaque côté du JOIN. Toutes les conditions connectées par un opérateur OR ou un opérateur AND logique doivent également respecter cette exigence.

Voici un exemple de requête avec un opérateur AND logique.

```
SELECT some_col, other_col
FROM table1
    JOIN table2
    ON table1.id = table2.id AND table1.name = table2.name
```

Voici un exemple de requête avec un opérateur OR logique.

```
SELECT some_col, other_col
FROM table1
    JOIN table2
    ON table1.id = table2.id OR table1.name = table2.name
```

Contrôle	Définition	Usage
<code>joinColumns</code>	Les colonnes que vous souhaitez autoriser le membre autorisé à effectuer une requête à utiliser dans l'instruction INNER JOIN.	La même colonne ne peut pas être classée à la fois comme <code>joinColumn</code> et <code>listColumn</code> (voir Contrôles de liste). <code>joinColumn</code> ne peut être utilisé dans aucune autre

Contrôle	Définition	Usage
		partie de la requête que INNER JOIN.

Contrôles de liste

Les contrôles de liste contrôlent les colonnes qui peuvent être répertoriées dans le résultat de la requête (c'est-à-dire utilisées dans l'instruction SELECT) ou utilisées pour filtrer les résultats (c'est-à-dire utilisées dans l'WHERE instruction).

Contrôle	Définition	Usage
<code>listColumns</code>	Les colonnes que vous autorisez le membre qui peut effectuer une requête à utiliser dans le SELECT et WHERE	A <code>listColumn</code> peut être utilisé dans SELECT et WHERE. La même colonne ne peut pas être utilisée à la fois comme <code>listColumn</code> et <code>joinColumn</code>.

Structure prédéfinie des règles d'analyse des listes

L'exemple suivant inclut une structure prédéfinie qui montre comment exécuter une règle d'analyse de liste.

Dans l'exemple suivant, *MyTable* fait référence à votre table de données. Vous pouvez remplacer chacune *user input placeholder* par vos propres informations.

```
{
  "joinColumns": [MyTable column name(s)],
  "listColumns": [MyTable column name(s)],
}
```


Règle d'analyse des listes - exemple

L'exemple suivant montre comment deux entreprises peuvent collaborer à AWS Clean Rooms l'aide de l'analyse de listes.

L'entreprise A dispose de données de gestion de la relation client (CRM). L'entreprise A souhaite obtenir des données sectorielles supplémentaires sur ses clients pour en savoir plus sur leurs clients et éventuellement utiliser des attributs comme données d'entrée dans d'autres analyses. L'entreprise B possède des données de segment composées d'attributs de segment uniques qu'elle a créés sur la base de ses données de première partie. L'entreprise B souhaite fournir les attributs de segment uniques à l'entreprise A uniquement pour les clients dont les données se chevauchent avec celles de l'entreprise A.

Les entreprises décident de collaborer afin que l'entreprise A puisse enrichir les données qui se chevauchent. L'entreprise A est le membre qui peut interroger, et l'entreprise B est le contributeur.

Pour créer une collaboration et exécuter une analyse de liste en collaboration, les entreprises procèdent comme suit :

1. L'entreprise A crée une collaboration et crée une adhésion. La collaboration a la société B comme autre membre de la collaboration. L'entreprise A active la journalisation des requêtes dans la collaboration, et elle active la journalisation des requêtes dans son compte.
2. L'entreprise B crée une adhésion à la collaboration. Il permet la journalisation des requêtes dans son compte.
3. L'entreprise A crée une table configurée pour le CRM
4. L'entreprise A ajoute la règle d'analyse à la table configurée par le client, comme indiqué dans l'exemple suivant.

```
{
  "joinColumns": [
    "identifiant1",
    "identifiant2"
  ],
  "listColumns": [
    "internalid",
    "segment1",
    "segment2",
    "customercategory"
  ]
}
```

```
}
```

`joinColumns`— L'entreprise A souhaite utiliser `hashedemail` and/or `thirdpartyid` (obtenu auprès d'un fournisseur d'identité) pour faire correspondre les clients à partir des données CRM aux clients à partir de données segmentées. Cela permettra de garantir que l'entreprise A associe des données enrichies aux bons clients. Ils ont deux `JoinColumns` pour potentiellement améliorer le taux de correspondance de l'analyse.

`listColumns`— L'entreprise A utilise `listColumns` pour obtenir des colonnes enrichies à côté d'une colonne `internalid` qu'elle utilise dans ses propres systèmes. Ils ajoutent `segment1` et limitent potentiellement l'enrichissement `customercategory` à des segments spécifiques en les utilisant dans des filtres. `segment2`

5. La société B crée une table configurée par segments.
6. L'entreprise B ajoute la règle d'analyse à la table des segments configurés.

```
{
  "joinColumns": [
    "identifier2"
  ],
  "listColumns": [
    "segment3",
    "segment4"
  ]
}
```

`joinColumns`— L'entreprise B permet à l'entreprise A de se joindre à elle pour `identifier2` faire correspondre les clients, qu'il s'agisse de données segmentées ou de données CRM. Les sociétés A et B ont travaillé avec le fournisseur d'identité pour `identifier2` déterminer laquelle correspondrait à cette collaboration. Ils n'en ont pas ajouté d'autres `joinColumns` parce qu'ils pensaient `identifier2` que c'était le taux de correspondance le plus élevé et le plus précis possible et qu'aucun autre identifiant n'était requis pour les requêtes.

`listColumns`— L'entreprise B permet à l'entreprise A d'enrichir ses données `segment3` et ses `segment4` attributs, qui sont des attributs uniques qu'elle a créés, collectés et sur lesquels elle s'est alignée (avec le client A) afin de participer à l'enrichissement des données. Ils souhaitent que l'entreprise A obtienne ces segments pour le chevauchement au niveau des lignes, car il s'agit d'une collaboration d'enrichissement des données.

7. L'entreprise A crée une association de tables CRM pour la collaboration.

8. L'entreprise B crée une association de tables de segments pour la collaboration.
9. L'entreprise A exécute des requêtes, telles que la suivante, pour enrichir les données clients qui se recoupent.

```
SELECT companyA.internalid, companyB.segment3, companyB.segment4
INNER JOIN returns companyB
  ON companyA.identifrier2 = companyB.identifrier2
WHERE companyA.customercategory > 'xxx'
```

- 10 Les entreprises A et B examinent les journaux de requêtes. L'entreprise B vérifie que la requête est conforme à ce qui a été convenu dans l'accord de collaboration.

Règle d'analyse personnalisée dans AWS Clean Rooms

Dans AWS Clean Rooms, une règle d'analyse personnalisée est un nouveau type de règle d'analyse qui permet d'exécuter des requêtes personnalisées sur la table configurée. Les requêtes SQL personnalisées sont toujours limitées à la SELECT commande, mais elles peuvent utiliser davantage de constructions SQL que les requêtes d'[agrégation](#) et de [liste](#) (par exemple, des fonctions de fenêtre, OUTER JOIN ou des sous-requêtes ; voir la [référence AWS Clean Rooms SQL](#) pour une liste complète). Les requêtes SQL personnalisées ne doivent pas nécessairement suivre une structure de requête telle que les requêtes d'[agrégation](#) et de [liste](#).

La règle d'analyse personnalisée prend en charge des cas d'utilisation plus avancés que ceux qui peuvent être pris en charge par la règle d'agrégation et d'analyse de liste, tels que l'analyse d'attribution personnalisée, le benchmarking, l'analyse d'incrémentalité et la découverte d'audience. Cela s'ajoute à un surensemble des cas d'utilisation pris en charge par les règles d'agrégation et d'analyse de listes.

La règle d'analyse personnalisée prend également en charge la confidentialité différentielle. La confidentialité différentielle est un cadre mathématiquement rigoureux pour la protection de la confidentialité des données. Pour de plus amples informations, veuillez consulter [AWS Clean Rooms Confidentialité différentielle](#). Lorsque vous créez un modèle d'analyse, AWS Clean Rooms Differential Privacy vérifie le modèle pour déterminer s'il est compatible avec la structure de requête à usage général pour AWS Clean Rooms Differential Privacy. Cette validation garantit que vous ne créez pas de modèle d'analyse non autorisé avec une table protégée par la confidentialité différentielle.

Pour configurer la règle d'analyse personnalisée, les propriétaires de données peuvent choisir d'autoriser l'exécution de requêtes personnalisées spécifiques, stockées dans des [modèles](#)

[d'analyse](#), sur leurs tables configurées. Les propriétaires de données examinent les modèles d'analyse avant de les ajouter au contrôle d'analyse autorisé dans la règle d'analyse personnalisée. Les modèles d'analyse sont disponibles et visibles uniquement dans la collaboration dans laquelle ils ont été créés (même si la table est associée à d'autres collaborations) et ne peuvent être exécutés que par le membre qui peut effectuer des requêtes dans cette collaboration.

Les membres peuvent également choisir d'autoriser d'autres membres (fournisseurs de requêtes) à créer des requêtes sans révision. Les membres ajoutent les comptes des fournisseurs de requêtes que les fournisseurs de requêtes autorisés contrôlent dans la règle d'analyse personnalisée. Si le fournisseur de requêtes est le membre habilité à effectuer une requête, il peut exécuter n'importe quelle requête directement sur la table configurée. Les fournisseurs de requêtes peuvent également créer des requêtes en [créant des modèles d'analyse](#). Toutes les requêtes créées par les fournisseurs de requêtes sont automatiquement autorisées à s'exécuter sur la table dans toutes les collaborations dans lesquelles elles sont présentes et où la table est associée. Compte AWS

Les propriétaires de données peuvent uniquement autoriser les modèles d'analyse ou les comptes à créer des requêtes, et non les deux. Si le propriétaire des données le laisse vide, le membre autorisé à effectuer des requêtes ne peut pas exécuter de requêtes sur la table configurée.

Rubriques

- [Structure prédéfinie des règles d'analyse personnalisées](#)
- [Exemple de règle d'analyse personnalisée](#)
- [Règle d'analyse personnalisée avec confidentialité différentielle](#)

Structure prédéfinie des règles d'analyse personnalisées

L'exemple suivant inclut une structure prédéfinie qui vous montre comment exécuter une règle d'analyse personnalisée avec la confidentialité différentielle activée. La `userIdentifier` valeur est la colonne qui identifie de manière unique vos utilisateurs, telle que `user_id`. Lorsque la confidentialité différentielle est activée sur deux tables ou plus dans le cadre d'une collaboration, AWS Clean Rooms cela vous oblige à configurer la même colonne que la colonne d'identifiant utilisateur dans les deux règles d'analyse afin de maintenir une définition cohérente des utilisateurs entre les tables.

```
{
  "allowedAnalyses": ["ANY_QUERY"] | string[],
  "allowedAnalysisProviders": [],
  "differentialPrivacy": {
```

```
"columns": [  
  {  
    "name": "userIdentifier"  
  }  
]  
}
```

Vous avez le choix entre les options suivantes :

- Ajoutez un modèle ARNs d'analyse au contrôle des analyses autorisées. Dans ce cas, le `allowedAnalysisProviders` contrôle n'est pas inclus.

```
{  
  allowedAnalyses: string[]  
}
```

- Ajoutez un membre Compte AWS IDs au `allowedAnalysisProviders` contrôle. Dans ce cas, vous ajoutez `ANY_QUERY` au `allowedAnalyses` contrôle.

```
{  
  allowedAnalyses: ["ANY_QUERY"],  
  allowedAnalysisProviders: string[]  
}
```

Exemple de règle d'analyse personnalisée

L'exemple suivant montre comment deux entreprises peuvent collaborer à AWS Clean Rooms l'aide de la règle d'analyse personnalisée.

L'entreprise A possède des données sur les clients et les ventes. L'entreprise A souhaite comprendre l'augmentation des ventes d'une campagne publicitaire sur le site de l'entreprise B. L'entreprise B possède des données d'audience et des attributs de segment utiles à l'entreprise (par exemple, l'appareil utilisé pour visionner la publicité).

L'entreprise A souhaite exécuter une requête d'incrémentalité spécifique dans le cadre de la collaboration.

Pour créer une collaboration et exécuter une analyse personnalisée en collaboration, les entreprises procèdent comme suit :

1. L'entreprise A crée une collaboration et crée une adhésion. La collaboration a la société B comme autre membre de la collaboration. L'entreprise A active la journalisation des requêtes dans la collaboration, et elle active la journalisation des requêtes dans son compte.
2. L'entreprise B crée une adhésion à la collaboration. Il permet la journalisation des requêtes dans son compte.
3. L'entreprise A crée une table configurée pour le CRM
4. La société A ajoute une règle d'analyse personnalisée vide à la table configurée des ventes.
5. L'entreprise A associe la table configurée des ventes à la collaboration.
6. La société B crée une table configurée pour le nombre de vues.
7. La société B ajoute une règle d'analyse personnalisée vide à la table configurée par le nombre de vues.
8. La société B associe la table configurée en termes de nombre de vues à la collaboration.
9. L'entreprise A consulte le tableau des ventes et le tableau d'audience associés à la collaboration et crée un modèle d'analyse, en ajoutant la requête d'incrémentalité et le paramètre pour le mois de la campagne.

```
{
  "analysisParameters": [
    {
      "defaultValue": ""
      "type": "DATE"
      "name": "campaign_month"
    }
  ],
  "description": "Monthly incrementality query using sales and viewership data"
  "format": "SQL"
  "name": "Incrementality analysis"
  "source":
    "WITH labeleddata AS
    (
      SELECT hashedemail, deviceid, purchases, unitprice, purchasedate,
      CASE
        WHEN testvalue IN ('value1', 'value2', 'value3') THEN 0
        ELSE 1
      END AS testgroup
      FROM viewershipdata
    )
    SELECT labeleddata.purchases, provider.impressions
    FROM labeleddata
```

```

    INNER JOIN salesdata
      ON labeleddata.hashedemail = provider.hashedemail
    WHERE MONTH(labeleddata.purchasedate) > :campaignmonth
    AND testgroup = :group
  "
}
```

10 L'entreprise A ajoute son compte (par exemple, 444455556666) au contrôle du fournisseur d'analyse autorisé dans la règle d'analyse personnalisée. Ils utilisent le contrôle du fournisseur d'analyse autorisé car ils souhaitent autoriser l'exécution de toutes les requêtes qu'ils créent sur leur table configurée pour les ventes.

```

{
  "allowedAnalyses": [
    "ANY_QUERY"
  ],
  "allowedAnalysisProviders": [
    "444455556666"
  ]
}
```

11 L'entreprise B voit le modèle d'analyse créé dans la collaboration et en examine le contenu, y compris la chaîne de requête et le paramètre.

12 L'entreprise B détermine que le modèle d'analyse répond au cas d'utilisation de l'incrémentalité et répond à ses exigences de confidentialité quant à la manière dont sa table configurée d'audience peut être interrogée.

13 La société B ajoute l'ARN du modèle d'analyse au contrôle d'analyse autorisé dans la règle d'analyse personnalisée de la table d'audience. Ils utilisent le contrôle d'analyse autorisé car ils souhaitent uniquement autoriser l'exécution de la requête d'incrémentalité sur leur table configurée par affichage.

```

{
  "allowedAnalyses": [
    "arn:aws:cleanrooms:us-east-1:111122223333:membership/41327cc4-bbf0-43f1-b70c-a160dddceb08/analysistemplate/1ff1bf9d-781c-418d-a6ac-2b80c09d6292"
  ]
}
```

14 L'entreprise A exécute le modèle d'analyse et utilise la valeur du paramètre 05-01-2023.

Règle d'analyse personnalisée avec confidentialité différentielle

Dans AWS Clean Rooms, la règle d'analyse personnalisée prend en charge la confidentialité différentielle. La confidentialité différentielle est un cadre mathématiquement rigoureux pour la protection de la confidentialité des données qui vous aide à protéger vos données contre les tentatives de réidentification.

La confidentialité différentielle prend en charge les analyses agrégées telles que la planification de campagnes publicitaires, les post-ad-campaign mesures, l'analyse comparative dans un consortium d'institutions financières et les tests A/B pour la recherche dans le domaine de la santé.

La structure et la syntaxe de requête prises en charge sont définies dans [Structure et syntaxe des requêtes](#).

Exemple de règle d'analyse personnalisée avec confidentialité différentielle

Note

AWS Clean Rooms La confidentialité différentielle n'est disponible que pour les collaborations dans lesquelles les données sont stockées dans Amazon S3.

Examinez l'[exemple de règle d'analyse personnalisée](#) présenté dans la section précédente. Cet exemple montre comment vous pouvez utiliser la confidentialité différentielle pour protéger vos données contre les tentatives de réidentification tout en permettant à votre partenaire de tirer des informations critiques de vos données. Supposons que l'entreprise B, qui possède les données d'audience, souhaite protéger ses données en utilisant une confidentialité différentielle. Pour terminer la configuration de la confidentialité différentielle, l'entreprise B effectue les étapes suivantes :

1. L'entreprise B active la confidentialité différentielle tout en ajoutant une règle d'analyse personnalisée au tableau configuré par le nombre de vues. L'entreprise B sélectionne `viewershipdata.hashemail` comme colonne d'identifiant utilisateur.
2. L'entreprise B [ajoute une politique de confidentialité différentielle](#) à la collaboration afin de rendre sa table de données d'audience disponible pour les requêtes. L'entreprise B sélectionne la politique par défaut pour terminer rapidement la configuration.

L'entreprise A, qui souhaite comprendre l'augmentation des ventes d'une campagne publicitaire sur le site de l'entreprise B, exécute le modèle d'analyse. La requête étant compatible avec la [structure de requête](#) à usage général de AWS Clean Rooms Differential Privacy, elle s'exécute correctement.

Structure et syntaxe des requêtes

Les requêtes contenant au moins une table dont la confidentialité différentielle est activée doivent respecter la syntaxe suivante.

```

query_statement:
    [cte, ...] final_select

cte:
    WITH sub_query AS (
        inner_select
        [ UNION | INTERSECT | UNION_ALL | EXCEPT/MINUS ]
        [ inner_select ]
    )

inner_select:
    SELECT [user_id_column, ] expression [, ...]
    FROM table_reference [, ...]
    [ WHERE condition ]
    [ GROUP BY user_id_column[, expression] [, ...] ]
    [ HAVING condition ]

final_select:
    SELECT [expression, ...] | COUNT | COUNT_DISTINCT | SUM | AVG | STDDEV
    FROM table_reference [, ...]
    [ WHERE condition ]
    [ GROUP BY expression [, ...] ]
    [ HAVING COUNT | COUNT_DISTINCT | SUM | AVG | STDDEV | condition ]
    [ ORDER BY column_list ASC | DESC ]
    [ OFFSET literal ]
    [ LIMIT literal ]

expression:
    column_name [, ...] | expression AS alias | aggregation_functions |
    window_functions_on_user_id | scalar_function | CASE | column_name math_expression [,
    expression]

window_functions_on_user_id:
    function () OVER (PARTITION BY user_id_column, [column_name] [ORDER BY column_list
    ASC|DESC])
  
```

 Note

En ce qui concerne la structure et la syntaxe différentielles des requêtes de confidentialité, tenez compte des points suivants :

- Les sous-requêtes ne sont pas prises en charge.
- Les expressions de table communes (CTEs) doivent émettre la colonne d'identifiant utilisateur si une table ou un CTE implique des données protégées par une confidentialité différentielle. Les filtres, les regroupements et les agrégations doivent être effectués au niveau de l'utilisateur.
- Final_select autorise les fonctions d'agrégation COUNT DISTINCT, COUNT, SUM, AVG et STDDEV.

Pour plus de détails sur les mots clés SQL pris en charge pour une confidentialité différentielle, consultez [Fonctionnalités SQL de AWS Clean Rooms Differential Privacy](#).

Règle d'analyse des tables de mappage d'identifiants

Dans AWS Clean Rooms, une règle d'analyse de table de mappage d'identifiants n'est pas une règle d'analyse autonome. Ce type de règle d'analyse est géré AWS Clean Rooms et utilisé pour joindre des données d'identité disparates afin de faciliter les requêtes. Il est automatiquement ajouté aux tables de mappage des identifiants et ne peut pas être modifié. Elle hérite des comportements des autres règles d'analyse de la collaboration, à condition que ces règles d'analyse soient homogènes.

La règle d'analyse des tables de mappage d'identifiants renforce la sécurité d'une table de mappage d'identifiants. Cela empêche un membre de la collaboration de sélectionner ou d'inspecter directement la population ne se chevauchant pas entre les ensembles de données des deux membres à l'aide de la table de mappage des identifiants. La règle d'analyse de la table de mappage d'identifiants est utilisée pour protéger les données sensibles de la table de mappage d'identifiants lorsqu'elles sont utilisées implicitement dans des requêtes avec d'autres règles d'analyse.

Avec la règle d'analyse de la table de mappage d'identifiants, AWS Clean Rooms impose un chevauchement des deux côtés de la table de mappage d'identifiants dans le SQL étendu. Cela vous permet d'effectuer les tâches suivantes :

- Utilisez le chevauchement de la table de mappage des identifiants dans JOIN les instructions.

AWS Clean Rooms autorise une INNER RIGHT jointure ou une jointure sur la table de mappage des identifiants si elle respecte le chevauchement. LEFT Pour protéger les informations de mappage sensibles, la table de mappage des identifiants doit toujours se trouver du côté inner « » de toute JOIN opération. Par exemple, les JOIN opérations suivantes sont valides :

- table LEFT JOIN id_mapping_table
- id_mapping_table RIGHT JOIN table
- table INNER JOIN id_mapping_table

Les JOIN opérations suivantes ne sont pas valides :

- id_mapping_table LEFT JOIN table
- table RIGHT JOIN id_mapping_table

Cela évite d'exposer les enregistrements cartographiques qui ne présentent pas de correspondance correspondante dans votre ensemble de données. L'autorisation de telles opérations pourrait potentiellement révéler des informations sensibles sur les mappages de données des autres membres de la collaboration.

- Utilisez les colonnes de la table de mappage dans JOIN les instructions.

Vous ne pouvez pas utiliser les colonnes de la table de mappage dans les instructions suivantes :SELECT,, WHERE HAVINGGROUP BY, ou ORDER BY (sauf si les protections sont modifiées sur l'association d'espace de noms d'ID source ou sur l'association d'espace de noms d'ID cible).

- Dans le SQL étendu OUTERJOIN, prend AWS Clean Rooms également en charge l'implicite et JOIN le CROSSJOIN. Ces jointures ne peuvent pas satisfaire aux exigences de chevauchement. AWS Clean Rooms Utilisez-le plutôt requireOverlap pour spécifier les colonnes qui doivent être jointes.

La structure et la syntaxe de requête prises en charge sont définies dans. [Structure et syntaxe des requêtes de la table de mappage d'identifiants](#)

Les paramètres de la règle d'analyse, définis dans[Contrôles des requêtes des règles d'analyse des tables de mappage d'identifiants](#), incluent les contrôles des requêtes et les contrôles des résultats des requêtes. Ses contrôles de requête incluent la possibilité d'exiger le chevauchement de la table de mappage des identifiants dans JOIN les instructions (c'est-à-dire,requireOverlap).

Rubriques

- [Structure et syntaxe des requêtes de la table de mappage d'identifiants](#)
- [Contrôles des requêtes des règles d'analyse des tables de mappage d'identifiants](#)
- [Structure prédéfinie des règles d'analyse des tables de mappage d'identifiants](#)
- [Règle d'analyse des tables de mappage d'identifiants — exemple](#)

Structure et syntaxe des requêtes de la table de mappage d'identifiants

Les requêtes sur les tables dotées d'une règle d'analyse des tables de mappage d'ID doivent respecter la syntaxe suivante.

```
--select_list_expression
SELECT
provider.data_col, consumer.data_col

--table_expression
FROM provider

JOIN idMappingTable idmt ON provider.id = idmt.sourceId

JOIN consumer ON consumer.id = idmt.targetId
```

Tables de collaboration

Les tableaux suivants représentent les tables configurées qui existent dans une AWS Clean Rooms collaboration. La colonne id des tables cr_drivers_license et cr_insurance représente une colonne correspondant à la table de mappage des identifiants.

cr_drivers_license

id	nom_pilote	état_d'enregistrement
1	Eduard	TX
2	Dana	MA
3	Gweneth	IL

cr_insurance

id	courriel du titulaire de la police	numéro_politique
a	eduardo@internal.company.com	17f9d04e-f5be-4426-bdc4-250ed59c6529
b	gwen@internal.company.com	3f0092db-2316-48a8-8d44-09cf8f6e6c64
c	rosa@internal.company.com	d7692e84-3d3c-47b8-b46d-a0d5345f0601

Table de mappage des identifiants

Le tableau suivant représente une table de mappage d'identifiants existante qui correspond aux tables cr_drivers_license et cr_insurance. Toutes les entrées ne seront pas disponibles IDs pour les deux tables de collaboration.

cr_drivers_license_id	cr_insurance_id
1	a
2	null
3	b
null	c

La règle d'analyse de la table de mappage d'identifiants autorise uniquement l'exécution de requêtes sur l'ensemble de données qui se chevauchent, qui se présente comme suit :

cr_driver s_license_id	cr_insura nce_id	nom_pilote	état_d'en registrement	courriel du titulaire de la police	numéro_po litique

1	a	Eduard	TX	eduardo@internal.com	17f9d04e-f5be-4426-bdc4-250ed59c6529
3	b	Gweneth	IL	gwen@internal.com	3f0092db-2316-48a8-8d44-09cf8f6e6c64

Exemples de requêtes

Les exemples suivants montrent des emplacements valides pour les jointures de tables de mappage d'identifiants :

```
-- Single ID mapping table
SELECT
    [ select_items ]FROM
    cr_drivers_license cr_dl
    [ INNER | LEFT ] JOIN cr_identity_mapping_table idmt ON idmt.cr_drivers_license_id
= cr_dl.id
    [ INNER | RIGHT ] JOIN cr_insurance cr_in          ON idmt.cr_insurance_id
= cr_in.id
;
-- Single ID mapping table (Subquery)
SELECT
    [ select_items ]FROM (
    SELECT
        [ select_items ]
    FROM
        cr_drivers_license cr_dl
        [ INNER | LEFT ] JOIN cr_identity_mapping_table idmt ON
idmt.cr_drivers_license_id = cr_dl.id
        [ INNER | RIGHT ] JOIN cr_insurance cr_in          ON idmt.cr_insurance_id
= cr_in.id
    )
;
-- Single ID mapping table (CTE)
WITH
    matched_ids AS (
        SELECT
```

```

        [ select_items ]
    FROM
        cr_drivers_license cr_dl
        [ INNER | LEFT ] JOIN cr_identity_mapping_table idmt ON
idmt.cr_drivers_license_id = cr_dl.id
        [ INNER | RIGHT ] JOIN cr_insurance cr_in          ON
idmt.cr_insurance_id      = cr_in.id
    )SELECT
    [ select_items ]FROM
    matched_ids
;

```

Considérations

En ce qui concerne la structure et la syntaxe des requêtes de la table de mappage d'identifiants, tenez compte des points suivants :

- Vous ne pouvez pas le modifier.
- Il est appliqué à la table de mappage des identifiants par défaut.
- Il utilise une association d'espaces de noms d'ID source et cible au sein de la collaboration.
- La table de mappage des identifiants est configurée par défaut pour fournir des protections par défaut à la colonne provenant de l'espace de noms des identifiants. Vous pouvez modifier cette configuration afin que la colonne provenant de l'espace de noms ID (soit `targetID`) `sourceID` soit autorisée n'importe où dans la requête. Pour de plus amples informations, veuillez consulter [Espaces de noms d'ID dans AWS Clean Rooms](#).
- La règle d'analyse des tables de mappage d'identifiants hérite des restrictions SQL des autres règles d'analyse de la collaboration.

Contrôles des requêtes des règles d'analyse des tables de mappage d'identifiants

Avec les commandes de requête de table de mappage d'ID, vous AWS Clean Rooms contrôlez la manière dont les colonnes de votre table sont utilisées pour interroger la table. Par exemple, il contrôle quelles colonnes sont utilisées pour la jointure et quelles colonnes doivent être superposées. La règle d'analyse des tables de mappage d'identifiants inclut également des fonctionnalités qui vous permettent d'autoriser la `targetID` projection du ou des deux sans nécessiter de JOIN. `sourceID`

Le tableau suivant explique chaque contrôle.

Contrôle	Définition	Usage
<code>joinColumns</code>	Les colonnes que le membre autorisé à interroger peut utiliser dans l'instruction INNER JOIN.	Vous ne pouvez pas l'utiliser <code>joinColumns</code> dans d'autres parties de la requête que dans INNER JOIN. Pour de plus amples informations, veuillez consulter Commandes de jointure.
<code>dimensionColumns</code>	Les colonnes (le cas échéant) que le membre autorisé à interroger peut utiliser dans les instructions SELECT et GROUP BY.	A <code>dimensionColumn</code> peut être utilisé dans SELECT et GROUPBY. A <code>dimensionColumn</code> peut apparaître sous la forme <code>joinKeys</code>. Vous ne pouvez l'utiliser <code>dimensionColumns</code> dans la clause JOIN que si vous la spécifiez entre crochets.
<code>queryConstraints:RequireOverlap</code>	Les colonnes de la table de mappage d'identifiants qui doivent être jointes pour que la requête puisse être exécutée.	Ces colonnes doivent être utilisées pour REJOINDRE la table de mappage des identifiants et une table de collaboration.

Structure prédéfinie des règles d'analyse des tables de mappage d'identifiants

La structure prédéfinie d'une règle d'analyse de table de mappage d'identifiants est assortie de protections par défaut appliquées au `sourceID` et `targetID`. Cela signifie que la colonne avec les protections appliquées doit être utilisée dans les requêtes.

Vous pouvez configurer la règle d'analyse de la table de mappage d'identifiants de la manière suivante :

- À la fois sourceID et targetID protégé

Dans cette configuration, le sourceID et ne targetID peuvent pas être projetés tous les deux. Le sourceID et targetID doit être utilisé dans un JOIN lorsque la table de mappage d'ID est référencée.

- targetIDProtégé uniquement

Dans cette configuration, le ne targetID peut pas être projeté. Le targetID doit être utilisé dans un JOIN lorsque la table de mappage d'ID est référencée. Le sourceID peut être utilisé dans une requête.

- sourceIDProtégé uniquement

Dans cette configuration, le ne sourceID peut pas être projeté. Le sourceID doit être utilisé dans un JOIN lorsque la table de mappage d'ID est référencée. Le targetID peut être utilisé dans une requête.

- Ni l'sourceIDun ni l'targetIDautre

Dans cette configuration, la table de mappage des identifiants n'est soumise à aucune application spécifique pouvant être utilisée dans la requête.

L'exemple suivant montre une structure prédéfinie pour une règle d'analyse de table de mappage d'identifiants avec les protections par défaut appliquées au et. sourceID targetID Dans cet exemple, la règle d'analyse de la table de mappage d'identifiants autorise uniquement un INNER JOIN à la fois sur la sourceID colonne et sur la targetID colonne.

```
{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [
    {
      "requireOverlap": {
        "columns": [
          "source_id",
          "target_id"
        ]
      }
    }
  ]
}
```

```
],  
  "dimensionColumns": [] // columns that can be used in SELECT and JOIN  
}
```

L'exemple suivant montre une structure prédéfinie pour une règle d'analyse de table de mappage d'identifiants avec des protections appliquées au `targetID`. Dans cet exemple, la règle d'analyse de la table de mappage d'ID n'autorise qu'un INNER JOIN sur la `sourceID` colonne.

```
{  
  "joinColumns": [  
    "source_id",  
    "target_id"  
  ],  
  "queryConstraints": [  
    {  
      "requireOverlap": {  
        "columns": [  
          "target_id"  
        ]  
      }  
    }  
  ],  
  "dimensionColumns": [  
    "source_id"  
  ]  
}
```

L'exemple suivant montre une structure prédéfinie pour une règle d'analyse de table de mappage d'identifiants avec des protections appliquées au `sourceID`. Dans cet exemple, la règle d'analyse de la table de mappage d'ID n'autorise qu'un INNER JOIN sur la `targetID` colonne.

```
{  
  "joinColumns": [  
    "source_id",  
    "target_id"  
  ],  
  "queryConstraints": [  
    {  
      "requireOverlap": {  
        "columns": [  
          "source_id"  
        ]  
      }  
    }  
  ]  
}
```

```

    }
  }
],
"dimensionColumns": [
  "target_id"
]
}

```

L'exemple suivant montre une structure prédéfinie pour une règle d'analyse de table de mappage d'identifiants sans protection appliquée au ou. `sourceID` `targetID` Dans cet exemple, la règle d'analyse de la table de mappage d'ID autorise un INNER JOIN à la fois sur la `sourceID` colonne et sur la `targetID` colonne.

```

{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [
    {
      "requireOverlap": {
        "columns": []
      }
    }
  ],
  "dimensionColumns": [
    "source_id",
    "target_id"
  ]
}

```

Règle d'analyse des tables de mappage d'identifiants — exemple

Plutôt que de rédiger une longue déclaration en cascade faisant référence à des informations personnelles identifiables (PII), par exemple, les entreprises peuvent utiliser la règle d'analyse des tables de mappage d'identifiants pour utiliser le transcodage multipartite LiveRamp . L'exemple suivant montre comment vous pouvez collaborer en AWS Clean Rooms utilisant la règle d'analyse des tables de mappage d'identifiants.

L'entreprise A est un annonceur qui dispose de données sur les clients et les ventes, qui seront utilisées comme source. La société A effectue également le transcodage pour le compte des parties à la collaboration et apporte les LiveRamp informations d'identification.

L'entreprise B est un éditeur qui possède des données sur les événements, qui seront utilisées comme cible.

 Note

La société A ou la société B peuvent fournir des informations d'identification pour le LiveRamp transcodage et effectuer le transcodage.

Pour créer une collaboration permettant d'analyser la table de mappage des identifiants en collaboration, les entreprises procèdent comme suit :

1. L'entreprise A crée une collaboration et crée une adhésion. Elle ajoute la société B, qui crée également une adhésion à la collaboration.
2. La société A associe une source d'espace de noms d'ID existante ou en crée une nouvelle en Résolution des entités AWS utilisant la AWS Clean Rooms console.

L'entreprise A crée une table configurée avec ses données de vente et une colonne saisie sur le `sourceId` dans la table de mappage des identifiants.

La source de l'espace de noms ID fournit les données à transcoder.

3. La société B associe une cible d'espace de noms ID existante ou en crée une nouvelle en Résolution des entités AWS utilisant la AWS Clean Rooms console.

La société B crée une table configurée avec ses données d'événements et une colonne saisie sur le `targetId` dans la table de mappage des identifiants.

La cible de l'espace de noms ID ne fournit pas de données à transcoder, mais uniquement des métadonnées relatives à la LiveRamp configuration.

4. L'entreprise A découvre les deux espaces de noms d'identification associés à la collaboration et crée et remplit une table de mappage d'identifiants.
5. L'entreprise A exécute une requête sur les deux ensembles de données en les joignant sur la table de mappage des identifiants.

--- this would be valid for Custom or List

```
SELECT provider.data_col, consumer.data_col
FROM provider
  JOIN idMappingTable-123123123123-myMappingWFName idmt
    ON provider.id = idmt.sourceId
  JOIN consumer
    ON consumer.id = idmt.targetId
```

AWS Clean Rooms Confidentialité différentielle

AWS Clean Rooms La confidentialité différentielle vous aide à protéger la vie privée de vos utilisateurs grâce à une technique basée sur des mathématiques qui est mise en œuvre avec des commandes intuitives en quelques clics. En tant que fonctionnalité entièrement gérée, aucune expérience préalable en matière de confidentialité différentielle n'est nécessaire pour vous aider à empêcher la réidentification de vos utilisateurs. AWS Clean Rooms ajoute automatiquement une quantité de bruit soigneusement calibrée aux résultats de la requête lors de l'exécution afin de protéger vos données au niveau individuel.

AWS Clean Rooms La confidentialité différentielle prend en charge un large éventail de requêtes analytiques et convient parfaitement à une grande variété de cas d'utilisation, dans lesquels une petite quantité d'erreur dans les résultats des requêtes ne compromet pas l'utilité de votre analyse. Grâce à elle, vos partenaires peuvent générer des informations critiques sur les campagnes publicitaires, les décisions d'investissement, la recherche clinique, etc., le tout sans nécessiter de configuration supplémentaire de la part de vos partenaires.

AWS Clean Rooms La confidentialité différentielle protège contre le débordement ou les erreurs de diffusion non valides qui utilisent des fonctions scalaires ou des symboles d'opérateurs mathématiques de manière malveillante.

Pour plus d'informations sur la confidentialité AWS Clean Rooms différentielle, consultez les rubriques suivantes.

Rubriques

- [Confidentialité différentielle](#)
- [Comment AWS Clean Rooms fonctionne la confidentialité différentielle](#)
- [Politique de confidentialité différentielle](#)
- [Fonctionnalités SQL de AWS Clean Rooms Differential Privacy](#)
- [Conseils et exemples de requêtes relatives à la confidentialité différentielle](#)

- [Limites de la confidentialité AWS Clean Rooms différentielle](#)

Confidentialité différentielle

La confidentialité différentielle ne permet que des informations agrégées et masque la contribution des données individuelles à ces informations. La confidentialité différentielle protège les données de collaboration du membre qui peut recevoir des résultats en découvrant une personne en particulier. Sans confidentialité différentielle, le membre qui peut recevoir des résultats peut tenter de déduire des données utilisateur individuelles en ajoutant ou en supprimant des enregistrements concernant un individu et en observant la différence entre les résultats des requêtes.

Lorsque la confidentialité différentielle est activée, une quantité spécifiée de bruit est ajoutée aux résultats de la requête pour masquer la contribution des utilisateurs individuels. Si le membre qui peut recevoir des résultats essaie d'observer la différence entre les résultats de la requête après avoir supprimé des enregistrements concernant un individu de son ensemble de données, la variabilité du résultat de la requête empêche l'identification des données de l'individu. AWS Clean Rooms Differential Privacy utilise le [SampCert](#)sampler, une implémentation d'échantillonneur correcte et éprouvée développée par. AWS

Comment AWS Clean Rooms fonctionne la confidentialité différentielle

Le flux de travail dans lequel vous souhaitez activer la confidentialité différentielle AWS Clean Rooms nécessite les étapes supplémentaires suivantes lors de [l'exécution du flux de travail pour AWS Clean Rooms](#) :

1. Vous activez la confidentialité différentielle lorsque vous ajoutez une [règle d'analyse personnalisée](#).
2. [Vous configurez la politique de confidentialité différentielle pour la collaboration](#) afin que vos tables de données protégées par une confidentialité différentielle soient disponibles pour les requêtes.

Une fois ces étapes terminées, le membre habilité à effectuer des requêtes peut commencer à exécuter des requêtes sur des données protégées par la confidentialité différentielle. AWS Clean Rooms renvoie des résultats conformes à la politique de confidentialité différentielle. AWS Clean Rooms La confidentialité différentielle permet de suivre le nombre estimé de requêtes restantes que vous pouvez exécuter, comme la jauge d'essence d'une voiture qui vous indique le niveau de carburant actuel de la voiture. Le nombre de requêtes que le membre qui peut interroger peut

exécuter est limité par le budget de confidentialité et le bruit ajouté par requête, paramètres définis dans le [Politique de confidentialité différentielle](#).

Considérations

Lorsque vous utilisez la confidentialité différentielle dans AWS Clean Rooms, tenez compte des points suivants :

- Le membre qui peut recevoir les résultats ne peut pas utiliser la confidentialité différentielle. Ils configureront une règle d'analyse personnalisée avec la confidentialité différentielle désactivée pour leurs tables configurées.
- Le membre qui peut effectuer une requête ne peut pas joindre les tables de deux fournisseurs de données ou plus lorsque la confidentialité différentielle est activée dans les deux cas.

Politique de confidentialité différentielle

La politique de confidentialité différentielle contrôle le nombre de fonctions d'agrégation que le membre qui peut interroger est autorisé à exécuter dans le cadre d'une collaboration. Le budget de confidentialité définit une ressource commune limitée qui est appliquée à toutes les tables d'une collaboration. Le bruit ajouté par requête détermine le taux d'épuisement du budget de confidentialité.

Une politique de confidentialité différentielle est requise pour que vos tables protégées par la confidentialité différentielle soient disponibles pour les requêtes. Il s'agit d'une étape unique dans le cadre d'une collaboration qui inclut deux contributions :

- Budget de confidentialité — Quantifié en termes d'epsilon, le budget de confidentialité contrôle le niveau de protection de la vie privée. Il s'agit d'une ressource commune limitée qui est appliquée à toutes vos tables protégées par une confidentialité différentielle dans le cadre de la collaboration, car l'objectif est de préserver la confidentialité de vos utilisateurs dont les informations peuvent être présentes dans plusieurs tables.

Le budget de confidentialité est consommé chaque fois qu'une requête est exécutée sur vos tables. Lorsque le budget de confidentialité est totalement épuisé, le membre de la collaboration qui peut effectuer des requêtes ne peut pas exécuter de requêtes supplémentaires tant qu'il n'est pas augmenté ou actualisé. En établissant un budget de confidentialité plus important, le membre qui peut recevoir les résultats peut réduire son incertitude quant aux individus contenus dans les données. Choisissez un budget de confidentialité qui équilibre vos exigences en matière de

collaboration et vos besoins en matière de confidentialité, après avoir consulté les décideurs commerciaux.

Vous pouvez sélectionner Actualiser le budget de confidentialité tous les mois pour créer automatiquement un nouveau budget de confidentialité chaque mois calendaire, si vous prévoyez d'intégrer régulièrement de nouvelles données à la collaboration. Le choix de cette option permet de révéler des quantités arbitraires d'informations sur les lignes de données lorsqu'elles sont demandées à plusieurs reprises lors des actualisations. Évitez de choisir cette option si les mêmes lignes doivent être consultées à plusieurs reprises entre les actualisations du budget de confidentialité.

- Le bruit ajouté par requête est mesuré en fonction du nombre d'utilisateurs dont vous souhaitez masquer les contributions. Cette valeur détermine le taux d'épuisement du budget de confidentialité. Une valeur de bruit plus élevée réduit le taux d'épuisement du budget de confidentialité et permet donc d'exécuter davantage de requêtes sur vos données. Cependant, cela doit être contrebalancé par la publication d'informations moins précises. Tenez compte de la précision souhaitée pour les informations sur la collaboration lorsque vous définissez cette valeur.

Vous pouvez utiliser la politique de confidentialité différentielle par défaut pour terminer rapidement la configuration ou personnaliser votre politique de confidentialité différentielle en fonction de votre cas d'utilisation. AWS Clean Rooms La confidentialité différentielle fournit des commandes intuitives pour configurer la politique. AWS Clean Rooms La confidentialité différentielle vous permet de prévisualiser l'utilitaire en termes de nombre d'agrégations possibles pour toutes les requêtes portant sur vos données et d'estimer le nombre de requêtes pouvant être exécutées dans le cadre d'une collaboration sur les données.

Vous pouvez utiliser les exemples interactifs pour comprendre l'impact des différentes valeurs du budget de confidentialité et du bruit ajouté par requête sur les résultats des différents types de requêtes SQL. En général, vous devez trouver un équilibre entre vos besoins en matière de confidentialité, le nombre de requêtes que vous souhaitez autoriser et l'exactitude de ces requêtes. Un budget de confidentialité réduit ou une augmentation du bruit ajouté par requête permet de mieux protéger la confidentialité des utilisateurs, mais fournit des informations moins pertinentes à vos partenaires de collaboration.

Si vous augmentez le budget de confidentialité tout en conservant le même paramètre de bruit ajouté par requête, le membre autorisé à effectuer une requête peut exécuter davantage d'agrégations sur vos tables dans le cadre de la collaboration. Vous pouvez augmenter le budget de confidentialité à tout moment au cours de la collaboration. Si vous réduisez le budget de confidentialité tout en

conservant le même paramètre de bruit ajouté par requête, le membre autorisé à effectuer une requête peut exécuter moins d'agrégations. Vous ne pouvez pas réduire le budget consacré à la confidentialité une fois que le membre habilité à interroger a commencé à analyser vos données.

Si vous augmentez le niveau de bruit ajouté par requête tout en conservant le même niveau d'entrée relatif au budget de confidentialité, le membre autorisé à effectuer des requêtes peut exécuter davantage d'agrégations sur vos tables dans le cadre de la collaboration. Si vous réduisez le bruit ajouté par requête tout en conservant le même montant d'entrée relatif au budget de confidentialité, le membre autorisé à effectuer une requête peut exécuter moins d'agrégations. Vous pouvez augmenter ou diminuer le bruit ajouté par requête à tout moment au cours de la collaboration.

La politique de confidentialité différentielle est gérée par les actions de l'API du modèle de budget de confidentialité.

Fonctionnalités SQL de AWS Clean Rooms Differential Privacy

AWS Clean Rooms La confidentialité différentielle utilise une structure de requête polyvalente pour prendre en charge les requêtes SQL complexes. Les modèles d'analyse personnalisés sont validés par rapport à cette structure afin de garantir qu'ils peuvent être exécutés sur des tables protégées par une confidentialité différentielle. Le tableau suivant indique les fonctions prises en charge. Pour plus d'informations, consultez [Structure et syntaxe des requêtes](#).

Opérateurs

SELECT

FROM

WHERE

ORDER BY

GROUP BY

JOIN

UNION

INTERSECT

EXCEPT

WITH

Fonctions

AVG

COUNT

MAX

MIN

Calculer
 SELECT
 table
 tables
 courantes
 (DATE)
 par
 le
 moteur
 d'analyse
 Spark
 offre :
 APERCENTIL
 COUNT
 COUNT
 • Fonction
 DISTINCT,
 AVG
 STDDEV
 • Fonctions
 tables
 COUNT
 SUM
 et
 par
 COUNT
 la
 DISTINCT
 confident
 • Fonction
 ilité
 MAX
 différent
 • Fonction
 telle
 MEDIAN
 doit
 • Fonction
 à MIN
 des
 • Fonction
 données
 PERCENTIL
 contenant
 E_COUNT
 des
 • Fonctions
 enregistr
 STDDEV_SA
 ements
 MP
 au
 et
 niveau

Cliquez sur
 SELECT
 dans
 le
 tableau
 des
 courantes
 (CTEs)
 par
 le
 moteur
 d'analyse
 Spark
 de
 STDDEV_PO
 l'utilisateur
 pour
 Fonctions
 Vous
 SUM
 devez
 écrire
 SUM
 l'expression
 DISTINCT
 ion
 • Fonctions
 SELECT
 VAR_SAMP
 dans
 et
 ceux
 VAR_POP
 qui
 CTEs
 utilisent
 le
 `SELECT
 userIdent
 ifierColu
 mn...'
 format.

Chaque

SELECT

table

tables

courantes

(CTE)

par

le

moteur

d'analyse

Spark

SELECT

WITH,

sous-

requête

et

de

CTE

Willati

on

de

tables

protégées

par

la

confident

ialité

différent

ielle

doit

aboutir

à

des

données

contenant

Chaque requête est envoyée à un moteur d'analyse Spark qui exécute la requête et retourne les résultats. Vous devez écrire l'expression SQL dans ceux qui utilisent le `SELECT` pour identifier les utilisateurs.

Changement
SELECT
ion
table
table
courantes
(DATE)
par
le
moteur
d'analyse
Spark
mn...'
format.

Chaque

SELECT

table

tables

courantes

(DATE)

par

le

moteur

d'analyse

Spark

SELECT

WHERE

à

• JOIN

n'importe

• Condition

quelle

d'adhésio

sous-

n

requ

• FROM

ête

WHERE

ne

fait

pas

référence

à

des

relations

de

confident

ialité

différent

ielles

dans

ces

construct

Chaque table est une table courante (CTE) par le moteur d'analyse Spark
ions.
Vous pouvez avoir n'importe quelle sous-requête qui fait référence à des relations de confidentialité différentielles uniquement dans une clause

Chaque table
isolee
table
en cours
(CITE)
par
le
moteur
d'analyse
Spark
FROM
et
JOIN.

Chapitre 10
10 SELECT
10.1 Table
10.1.1 Tables
10.1.2 courantes
10.1.3 (CTE)
10.1.4 par
10.1.5 le
10.1.6 moteur
10.1.7 d'analyse
10.1.8 Spark
10.2 Clapnet
10.2.1 d'INTERIEUR
10.2.2 JOINTURE
10.2.3 condition
10.2.4 GAUCHE
10.2.5 que
10.2.6 • DEMI-
10.2.7 seules
10.2.8 JOIN
10.2.9 les
10.2.10 fonctions
10.2.11 GAUCHE
10.2.12 JOIN
10.2.13 • ANTI-
10.2.14 qui
10.2.15 JOIN
10.2.16 sont
10.2.17 TURE
10.2.18 des
10.2.19 GAUCHE
10.2.20 jointures
10.2.21 • JOINTURE
10.2.22 égales
10.2.23 DROITE
10.2.24 sur
10.2.25 les
10.2.26 ADHESION
10.2.27 COMPLETE
10.2.28 d'identif
10.2.29 [JOIN]
10.2.30 iant
10.2.31 OU
10.2.32 utilisateur
10.2.33 ur
10.2.34 • CROSS
10.2.35 soient
10.2.36 JOIN
10.2.37 prises
10.2.38 en

Chargement
SELECT
des
table
tables
courantes
(CTE)
par
le
moteur
d'analyse
Spark
charge
et
soient
obligatoi
res
lors
de
l'interro
gation
de
deux
tables
ou
plus
avec
la
confident
ialité
différent
ielle
activée.
Assurez-
v
ous
que

Chargement
SELECT
table
table
courantes
(CTE)
par
le
moteur
d'analyse
Spark
les
condition
s
d'équijoi
nture
obligatoi
res
sont
correctes
.
Vérifiez
que
le
propriéta
ire
de
la
table
a
configuré
la
même
colonne
d'identif
iant

~~Chargement~~

~~SELECT~~

~~Table~~

~~tables~~

~~courantes~~

~~(CTE)~~

par

le

moteur

d'analyse

Spark

utilise

ur

dans

toutes

les

tables

afin

que

la

défini

n

d'un

utilise

ur

reste

cohérente

d'une

table

à

l'autre.

Les

fonctions

CROSS

JOIN

Chargement
ion SELECT
table
table
courantes
(CTE)
par
le
moteur
d'analyse
Spark
ne
sont
pas
prises
en
charge
lors
de
la
combinaison
on
de
deux
relations
ou
plus
lorsque
la
confidentialité
différentielle
est
activée.

Changement

SELECT

Table

tables

courantes

(Date)

par

le

moteur

d'analyse

Spark

UNION,

UNION

Algorithme

INTERSECT

”

EXCEPT

||

MINUS

(ce

sont

des

synonymes

)

Chargement

SELECT

table

tables

courantes

(DATE)

par

le

moteur

d'analyse

Spark

Fonctions

intégrat

fenêtrage

charge

- Fonction

charge

à de

fenêtrage

condition

AVG

que

la Fonction

de

colonne

d'identifi

ifiant

COUNT

utilisant

la fonction

urde

d'fenêtrage

la CUME_DIST

clause

- Fonction

de

partition

fenêtrage

de

DENSE_RAN

la K

fonction

- Fonction

de

de

fenêtre

fenêtrage

Cliquez sur
 SELECT
 dans
 la
 table
 courantes
 (CTE)
 par
 le
 moteur
 d'analyse
 Spark
 FIRST_VAL
 de
 la
 fonction
 de
 interrogation
 LAG
 une
 fonction
 relation
 de
 avec
 fenêtrage
 la
 LAST_VALU
 confident
 ialité
 • Fonction
 différent
 de
 ielle
 fenêtrage
 activée.
 LEAD
 • Fonctions
 de
 fenêtrage
 MAX
 • Fonctions
 de
 la
 fenêtrage
 MEDIAN

Calculs

SELECT

Table

tables

courantes

(DATE)

par

le

moteur

d'analyse

Spark

- Fonctions de la fenêtre MIN
- Fonction de fenêtrage NTH_VALUE
- Fonctions de fenêtre STDDEV_SAMP et STDDEV_POP (STDDEV_SAMP et STDDEV sont des

Calculs

SELECT

FROM

TABLES

courantes

(DATE)

par

le

moteur

d'analyse

Spark

synonymes

)

- Fonctions

de

la

fenêtre

SUM

- Fonctions

de

fenêtre

VAR_SAMP

et

VAR_POP

(VAR_SAMP

et

VARIANCE

sont

des

synonymes

)

Fonctions

de

Chapitre 10

10.1 SELECT

10.1.1 Table

10.1.2 Table

10.1.3 courantes

10.1.4 (CTE)

10.1.5 par

10.1.6 le

10.1.7 moteur

10.1.8 d'analyse

10.1.9 Spark

10.1.10 classemen

10.1.11 t

- Fonction de fenêtrage DENSE_RANK
- Fonction de fenêtrage NTILE
- Fonction de fenêtrage PERCENT_RANK
- Fonction de fenêtrage RANK
- Fonction de fenêtrage

Calculer

SELECT

Table

tables

courantes

(DATE)

par

le

moteur

d'analyse

Spark

ROW_NUMBE

R

Expression

sont

condition

condition

charge

- Expressio

n

COALESCE

- Fonctions

GREATEST

et

LEAST

- Fonctions

NVL

et

COALESCE

- NVL2

fonction

- Fonction

NULLIF

Calculs

SELECT

Table

tables

courantes

(DATE)

par

le

moteur

d'analyse

Spark

Condition

seul

comparais

peuvent

charge

Condition

être

utilisés

logiques

car

Condition

ils

nécessite

de

correspon

des

dance

sous-

de

requ

modèles

êtes.

ENTRE

Tous

les

condition

autres

s

sont

de

pris

gamme

en

Condition

charge.

null

Calculs

SELECT

FROM

TABLES

courantes

(DATE)

par

le

moteur

d'analyse

Spark

Fonctions

concat

pour

enregistrer

charge

les

transacti

ons

- Opérateur de concaténation
- Fonctions
ADD_MONTHS
- Fonction
CONVERT_TIMEZONE
- Fonction
CURRENT_DATE
- Fonction
DATEADD

Chapitre 1

1.1 SELECT

1.1.1 Table

1.1.2 Tables

1.1.3 courantes

1.1.4 (DATE)

1.1.5 par

1.1.6 le

1.1.7 moteur

1.1.8 d'analyse

1.1.9 Spark

- Fonction
DATEDIFF
- fonctions
DATE_PART
- Fonction
DATE_TRUNC
- Fonction
EXTRACT
- Fonction
TO_TIMESTAMP
- Parties
de
date
pour
les
fonctions
de
date
ou
d'horodat
age

Chargement

SELECT

FROM

TABLES

courantes

(DATE)

par

le

moteur

d'analyse

Spark

Fonctions

seul

pré-écrit

en

charge

Fonction

BTRIM

- Fonction

CHAR LENG

TH

- Fonction

CHARACTER

_LENGTH

- Fonction

CONCAT

- Fonctions

LEFT

et

RIGHT

- Fonction

LEN

- Fonction

LENGTH

Chapitre 1

1.1 SELECT

1.1.1 Table

1.1.2 Tables

1.1.3 courantes

1.1.4 (CTE)

1.1.5 par

1.1.6 le

1.1.7 moteur

1.1.8 d'analyse

1.1.9 Spark

- Fonction
LOWER
- Fonctions
LPAD
et
RPAD
- Fonction
LTRIM
- Fonctions
POSITION
- Fonction
REGEXP_CO
UNT
- Fonction
REGEXP_IN
STR
- Fonction
REGEXP_RE
PLACE
- Fonction
REGEXP_SU
BSTR

Calculs

SELECT

Table

Tables

courantes

(CTE)

par

le

moteur

d'analyse

Spark

- Fonction
REPEAT
- Fonction
REPLACE
- Fonction
REVERSE
- Fonction
RTRIM
- Fonction
SPLIT_PAR
T
- Fonction
SUBSTRING
- Fonction
TRANSLATE
- Fonctions
TRIM
- Fonction
UPPER

Chânes

SELECT

Table

tables

courantes

(DATE)

par

le

moteur

d'analyse

Spark

Fonction

CAST

fonctionnement

TO_CHAR

des

fonction

TO_DATE

de

• TO_NUMBER

données

• Chaînes

de

format

datetime

• Chaînes

de

format

numérique

s

Chargement
SELECT
table
tables
courantes
(CTE)
par
le
moteur
d'analyse
Spark

FLAES_ENCRY
sent
message
FLAES_DECRY
en
charge
• ENCODE
• DECODE
• MD5
fonction
• SHA1
fonction
• SHA2
fonction
• XX_
HASH64

Tous
Types
opérate
%
mathémati
@charge

Calculs

SELECT

FROM

TABLES

currentes

(DATE)

par

le

moteur

d'analyse

Spark

Fonction

SQL

Fonction

ACOS

charge

• Fonction

ASIN

• Fonction

ATAN

• ATAN2

fonction

• Fonction

CBRT

• Fonction

CEILING

(ou

CEIL)

• Fonction

COS

• Fonction

COT

• Fonction

DEGREES

Calculs

SELECT

FROM

TABLES

calculs

(CTE)

par

le

moteur

d'analyse

Spark

- Fonction
LTRIM
- Fonction
EXP
- Fonction
FLOOR
- Fonction
LN
- Fonction
LOG
- Fonction
MOD
- Fonction
PI
- Fonction
POWER
- Fonction
RADIANS
- Fonction
RANDOM
- Fonction
ROUND

Chargement

SELECT

FROM

TABLES

courantes

(DATE)

par

le

moteur

d'analyse

Spark

- Fonction
SIGN
- Fonction
SIN
- Fonctions
SQRT
- Fonction
TRUNC

FROM HIVE,

UNBASE64

• SORT

• HLL_SKETC
H_AGG,

• HLL_SKETC
H_ESTIMAT
E

• HL_UNION

• HL_UNION_
AGG

Chargement

SELECT

Table

tables

courantes

(CTE)

par

le

moteur

d'analyse

Spark

JSON

GET_JSON_

OBJECT

en

charge

ARRAY_CON

TAINS

ARRAY_DIS

TABLET

• ARRAY_EXC
EPTÉ

• ARRAY_INT
ERSECT

• ARRAY_JOI
N

• ARRAY_REM
OVE

• ARRAY_SOR
T

• ARRAY_UNI
ON

Chargement
SELECT
table
tables
courantes
(CTE)
par
le
moteur
d'analyse
Spark

ENSEMBLES
PAR
GROUPEMENT
charge
ROLLUP,
CUBE

Calculs

SELECT

Table

tables

courantes

(CTE)

par

le

moteur

d'analyse

Spark

ORDER

BY

taille

condition

qu'une

clause

ORDER

BY

ne

soit

prise

en

charge

dans

la

clause

de

partition

d'une

fonction

de

fenêtre

que

lors

Chaque session SELECT de table pour les courantes (60 fois) par le moteur d'analyse Spark de l'interrogation de tables avec la confidentialité différentielle activée.

Chargement

SELECT

Table

tables

courantes

(CTE)

par

le

moteur

d'analyse

Spark

NOTES,

DÉCALAGE

lignes

charge

CTE

lors

de

l'utilisa

tion

de

tables

protégées

par

la

confident

ialité

différent

ielle

~~Calculs~~

~~ions~~ SELECT

~~de~~ le

~~table~~ tables

~~en~~ courantes

~~(CTE)~~ (CTE)

par

le

moteur

d'analyse

Spark

~~Pris~~ ing

~~en~~

~~table~~ table

et

de

colonnes

~~Pris~~ tions

~~en~~ mathématique

~~charge~~ charge

sur

les

fonctions

d'agrégat

ion

~~Pris~~ tions

~~en~~ laires

~~charge~~ charge

les

fonctions

d'agrégat

ion

Alternatives courantes pour les constructions SQL non prises en charge

Catégorie	construction SQL	Autrement
Fonctions de fenêtrage	• LISTAGG • PERCENTILE_CONT • PERCENTILE_DISC	Vous pouvez utiliser la fonction d'agrégation équivalente avec GROUP BY.
Symboles d'opérateurs mathématiques	• \$column 2 • \$column 2 • \$column ^ 2	• CBRT • SQRT • PUISSANCE (\$column, 2)
Fonctions scalaires	• SYSDATE • \$column : :entier • convertir (type, \$column)	• CURRENT_DATE • CAST \$column EN TANT QU'entier • Type CAST \$column AS
Littéraux	INTERVALLE « 1 SECONDE »	INTERVALLE « 1 » SECONDE
Limitation des lignes	STOP n	LIMITE n
Joindre	• USING • NATURAL	La clause ON doit contenir explicitement un critère de jointure.

Conseils et exemples de requêtes relatives à la confidentialité différentielle

AWS Clean Rooms La confidentialité différentielle utilise une [structure de requête polyvalente](#) pour prendre en charge une grande variété de constructions SQL telles que les expressions de table communes (CTEs) pour la préparation des données et les fonctions d'agrégation couramment utilisées telles que COUNT, ou. SUM Afin de masquer la contribution de tout utilisateur potentiel à vos données en ajoutant du bruit aux résultats des requêtes agrégées au moment de l'exécution, la confidentialité AWS Clean Rooms différentielle exige que les fonctions d'agrégation finales SELECT statement soient exécutées sur des données au niveau de l'utilisateur.

L'exemple suivant utilise deux tables nommées `socialco_impressions` et `socialco_users` provenant d'un éditeur multimédia qui souhaite protéger les données en utilisant une confidentialité différentielle tout en collaborant avec une marque sportive utilisant `athletic_brand_sales` des données. L'éditeur multimédia a configuré la `user_id` colonne comme colonne d'identifiant utilisateur tout en activant la confidentialité différentielle dans AWS Clean Rooms. L'annonceur n'a pas besoin d'une protection différentielle de la confidentialité et souhaite exécuter une requête en utilisant CTEs des données combinées. Comme son CTE utilise des tables protégées de confidentialité différentielles, l'annonceur inclut la colonne d'identifiant d'utilisateur de ces tables protégées dans la liste des colonnes CTE et joint les tables protégées dans la colonne d'identifiant d'utilisateur.

```
WITH matches_table AS(
  SELECT si.user_id, si.campaign_id, s.sale_id, s.sale_price
  FROM socialco_impressions si
  JOIN socialco_users su
    ON su.user_id = si.user_id
  JOIN athletic_brand_sales s
    ON s.emailsha256 = su.emailsha256
  WHERE s.timestamp > si.timestamp

UNION ALL

  SELECT si.user_id, si.campaign_id, s.sale_id, s.sale_price
  FROM socialco_impressions si
  JOIN socialco_users su
    ON su.user_id = si.user_id
  JOIN athletic_brand_sales s
    ON s.phonesha256 = su.phonesha256
  WHERE s.timestamp > si.timestamp
)

SELECT COUNT (DISTINCT user_id) as unique_users
FROM matches_table
GROUP BY campaign_id
ORDER BY COUNT (DISTINCT user_id) DESC
LIMIT 5
```

De même, si vous souhaitez exécuter des fonctions de fenêtre sur des tables de données protégées par la confidentialité différentielle, vous devez inclure la colonne d'identifiant utilisateur dans la `PARTITION BY` clause.


```
ROW_NUMBER() OVER (PARTITION BY conversion_id, user_id ORDER BY match_type, match_age)
AS row
```

Limites de la confidentialité AWS Clean Rooms différentielle

AWS Clean Rooms La confidentialité différentielle ne permet pas de résoudre les situations suivantes :

1. AWS Clean Rooms La confidentialité différentielle prend uniquement en charge les requêtes utilisant des tables basées sur Amazon S3. AWS Glue II ne prend pas en charge les requêtes avec les tables Snowflake ou Amazon Athena.
2. AWS Clean Rooms La confidentialité différentielle ne permet pas de lutter contre les attaques temporelles. Par exemple, ces attaques sont possibles dans les scénarios où un utilisateur individuel fournit un grand nombre de lignes et où l'ajout ou la suppression de cet utilisateur modifie de manière significative le temps de calcul de la requête.
3. AWS Clean Rooms La confidentialité différentielle ne garantit pas une confidentialité différentielle lorsqu'une requête SQL peut entraîner un débordement ou des erreurs de diffusion non valides au moment de l'exécution en raison de l'utilisation de certaines constructions SQL.

Le tableau suivant répertorie certaines constructions SQL, mais pas toutes, susceptibles de générer des erreurs d'exécution et qui doivent être vérifiées dans les modèles d'analyse. Nous vous recommandons d'approuver les modèles d'analyse qui minimisent les risques de telles erreurs d'exécution et de consulter régulièrement les journaux de requêtes pour déterminer si les requêtes sont conformes à l'accord de collaboration.

Les constructions SQL suivantes sont vulnérables aux erreurs de débordement :

Catégorie	Constructions SQL vulnérables aux erreurs de débordement dans le moteur d'analyse Spark SQL
Fonctions d'agrégation	• AVG • SUM/SUM_DISTINCT
Fonctions de formatage des types de données	• TO_TIMESTAMP • TO_DATE

Catégorie	Constructions SQL vulnérables aux erreurs de débordement dans le moteur d'analyse Spark SQL
Fonctions de date et d'heure	• ADD_MONTHS • DATEADD • DATEDIFF
Fonctions mathématiques	• +, -, *, / • POWER
Fonctions de chaîne	• • CONCAT • RÉPÉTITION
Fonctions de fenêtrage	• AVG • SUM

4. La fonction de formatage du type de données CAST est vulnérable aux erreurs de conversion non valides.

Vous pouvez configurer [CloudWatch pour créer un filtre métrique pour un groupe de journaux](#), puis [créer une CloudWatch alarme](#) sur ce filtre métrique afin de recevoir des alertes en cas de dépassement potentiel ou d'erreur de casting.

Plus précisément, vous devez surveiller les codes d'erreur `CastError`, `OverflowError`, `ConversionError`. La présence de ces codes d'erreur indique une attaque potentielle par canal secondaire, mais peut également indiquer une requête SQL erronée.

Pour de plus amples informations, veuillez consulter [Connexion à une analyse AWS Clean Rooms](#).

AWS Clean Rooms ML

AWS Clean Rooms ML permet à deux ou plusieurs parties d'exécuter des modèles d'apprentissage automatique sur leurs données sans avoir à partager leurs données entre elles. Le service fournit des contrôles renforçant la confidentialité qui permettent aux propriétaires de données de protéger

leurs données et leur adresse IP de modèle. Vous pouvez utiliser des AWS modèles créés par des créateurs ou apporter votre propre modèle personnalisé.

Pour une explication plus détaillée de son fonctionnement, voir [Emplois multi-comptes](#).

Pour plus d'informations sur les fonctionnalités des modèles Clean Rooms ML, consultez les rubriques suivantes.

Rubriques

- [Terminologie AWS Clean Rooms ML](#)
- [Comment AWS Clean Rooms ML fonctionne avec les AWS modèles](#)
- [Comment AWS Clean Rooms ML fonctionne avec les modèles personnalisés](#)
- [AWS modèles dans Clean Rooms ML](#)
- [Modèles personnalisés dans Clean Rooms ML](#)

Terminologie AWS Clean Rooms ML

Il est important de comprendre la terminologie suivante lors de l'utilisation de Clean Rooms ML :

- Fournisseur de données de formation : partie qui fournit les données de formation, crée et configure un modèle similaire, puis associe ce modèle similaire à une collaboration.
- Fournisseur de données sur les semences : partie qui fournit les données sur les semences, génère un segment similaire et exporte son segment similaire.
- Données d'entraînement : données du fournisseur de données de formation, utilisées pour générer un modèle similaire. Les données d'entraînement sont utilisées pour mesurer la similitude des comportements des utilisateurs.

Les données d'entraînement doivent contenir un ID utilisateur, un ID d'élément et une colonne d'horodatage. Les données d'entraînement peuvent éventuellement contenir d'autres interactions sous forme de caractéristiques numériques ou catégoriques. Des exemples d'interactions sont une liste de vidéos regardées, d'articles achetés ou d'articles lus.

- Données de départ : données du fournisseur de données de départ, utilisées pour créer un segment similaire. Les données de départ peuvent être fournies directement ou provenir des résultats d'une AWS Clean Rooms requête. Le résultat du segment similaire est un ensemble d'utilisateurs issu des données d'entraînement qui ressemble le plus aux utilisateurs initiaux.

- **Modèle similaire** : modèle d'apprentissage automatique des données d'entraînement utilisé pour rechercher des utilisateurs similaires dans d'autres ensembles de données.

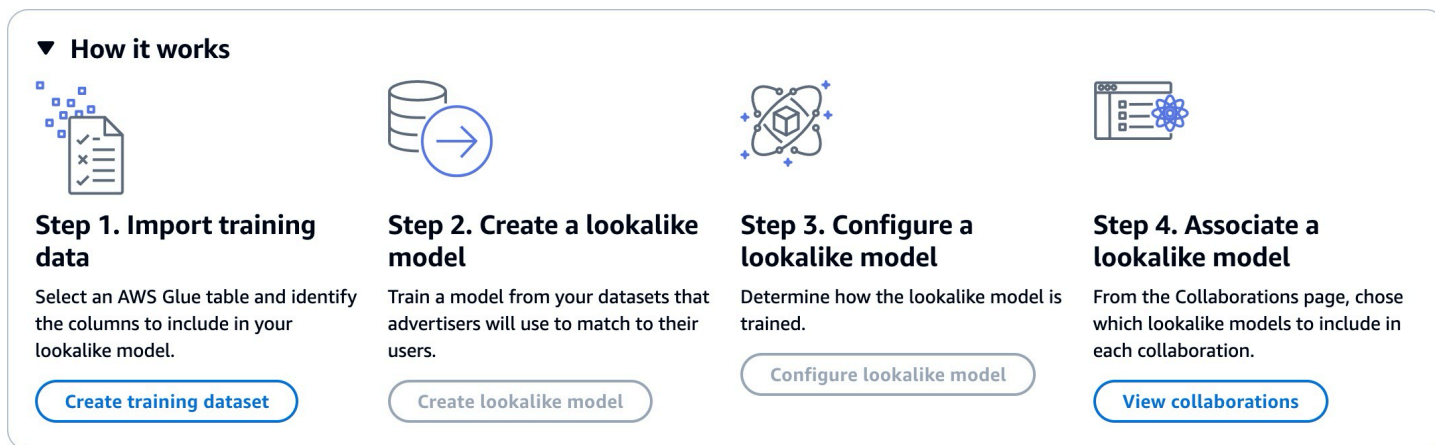
Lors de l'utilisation de l'API, le terme « modèle d'audience » est utilisé de manière équivalente à « modèle similaire ». Par exemple, vous utilisez l'[CreateAudienceModel](#) API pour créer un modèle similaire.

- **Segment similaire** : sous-ensemble des données d'entraînement qui ressemble le plus aux données de départ.

Lorsque vous utilisez l'API, vous créez un segment similaire avec l'[StartAudienceGenerationJob](#) API.

Les données du fournisseur de données de formation ne sont jamais partagées avec le fournisseur de données de départ et les données du fournisseur de données de départ ne sont jamais partagées avec le fournisseur de données de formation. La sortie du segment similaire est partagée avec le fournisseur de données de formation, mais jamais avec le fournisseur de données de départ.

Comment AWS Clean Rooms ML fonctionne avec les AWS modèles



L'utilisation de modèles similaires nécessite que deux parties, un fournisseur de données de formation et un fournisseur de données de départ, travaillent de manière séquentielle AWS Clean Rooms pour intégrer leurs données dans une collaboration. Voici le flux de travail que le fournisseur de données de formation doit effectuer en premier :

1. Les données du fournisseur de données de formation doivent être stockées dans une table de catalogue de AWS Glue données répertoriant les interactions entre les utilisateurs et les éléments. Les données d'entraînement doivent au minimum contenir une colonne d'ID utilisateur, une colonne d'identifiant d'interaction et une colonne d'horodatage.

2. Le fournisseur de données de formation enregistre les données de formation auprès de AWS Clean Rooms.
3. Le fournisseur de données de formation crée un modèle similaire qui peut être partagé avec plusieurs fournisseurs de données initiales. Le modèle similaire est un réseau neuronal profond dont l'entraînement peut prendre jusqu'à 24 heures. Il n'est pas automatiquement réentraîné et nous vous recommandons de le réentraîner chaque semaine.
4. Le fournisseur de données de formation configure le modèle de similarité, notamment en indiquant s'il convient de partager les indicateurs de pertinence et l'emplacement des segments de sortie sur Amazon S3. Le fournisseur de données de formation peut créer plusieurs modèles similaires configurés à partir d'un seul modèle similaire.
5. Le fournisseur de données de formation associe le modèle d'audience configuré à une collaboration partagée avec un fournisseur de données de départ.

Il s'agit du flux de travail que le fournisseur de données de départ doit effectuer ensuite :

1. Les données du fournisseur de données de base peuvent être stockées dans un compartiment Amazon S3 ou peuvent provenir des résultats d'une requête.
2. Le fournisseur de données de départ ouvre la collaboration qu'il partage avec le fournisseur de données de formation.
3. Le fournisseur de données de départ crée un segment similaire à partir de l'onglet Clean Rooms ML de la page de collaboration.
4. Le fournisseur de données de base peut évaluer les indicateurs de pertinence, s'ils ont été partagés, et exporter le segment similaire pour une utilisation externe AWS Clean Rooms.

Comment AWS Clean Rooms ML fonctionne avec les modèles personnalisés

Avec Clean Rooms ML, les membres d'une collaboration peuvent utiliser un algorithme de modèle personnalisé dockerisé stocké dans Amazon ECR pour analyser conjointement leurs données. Pour ce faire, le fournisseur de modèles doit créer une image et la stocker dans Amazon ECR. Suivez les étapes décrites dans le [guide de l'utilisateur d'Amazon Elastic Container Registry](#) pour créer un référentiel privé qui contiendra le modèle de ML personnalisé.

Tout membre d'une collaboration peut être le fournisseur de modèles, à condition de disposer des autorisations appropriées. Tous les membres d'une collaboration peuvent apporter des données

d'entraînement, des données d'inférence ou les deux au modèle. Aux fins du présent guide, les membres fournissant des données sont appelés fournisseurs de données. Le membre qui crée la collaboration est le créateur de la collaboration, et ce membre peut être le fournisseur de modèles, l'un des fournisseurs de données ou les deux.

Au plus haut niveau, voici les étapes à suivre pour effectuer une modélisation ML personnalisée :

1. Le créateur de la collaboration crée une collaboration et attribue à chaque membre les capacités et la configuration de paiement appropriées. Le créateur de la collaboration doit attribuer au membre la capacité de recevoir les sorties du modèle ou de recevoir les résultats d'inférence au membre approprié au cours de cette étape, car il ne peut pas être mis à jour une fois la collaboration créée. Pour de plus amples informations, veuillez consulter [Création et adhésion à la collaboration dans AWS Clean Rooms ML](#).
2. Le fournisseur de modèles configure et associe son modèle de machine learning conteneurisé à la collaboration et s'assure que les contraintes de confidentialité sont définies pour les données exportées. Pour de plus amples informations, veuillez consulter [Configuration d'un algorithme de modèle dans AWS Clean Rooms ML](#).
3. Les fournisseurs de données fournissent leurs données à la collaboration et veillent à ce que leurs besoins en matière de confidentialité soient spécifiés. Les fournisseurs de données doivent autoriser le modèle à accéder à leurs données. Pour plus d'informations, consultez [Contribution de données de formation dans AWS Clean Rooms ML](#) et [Associer l'algorithme du modèle configuré dans AWS Clean Rooms ML](#).
4. Un membre de la collaboration crée la configuration ML, qui définit l'endroit vers lequel les artefacts du modèle ou les résultats d'inférence sont exportés.
5. Un membre de la collaboration crée un canal d'entrée ML qui fournit des informations au conteneur de formation ou au conteneur d'inférence. Le canal d'entrée ML est une requête qui définit les données à utiliser dans le contexte de l'algorithme du modèle.
6. Un membre de la collaboration invoque l'entraînement du modèle à l'aide du canal d'entrée ML et de l'algorithme de modèle configuré. Pour de plus amples informations, veuillez consulter [Création d'un modèle entraîné dans AWS Clean Rooms ML](#).
7. (Facultatif) Le modèle d'entraînement lance la tâche d'exportation du modèle et les artefacts du modèle sont envoyés au récepteur des résultats du modèle. Seuls les membres dotés d'une configuration ML valide et capables de recevoir les résultats du modèle peuvent recevoir des artefacts du modèle. Pour de plus amples informations, veuillez consulter [Exportation d'artefacts de modèles depuis AWS Clean Rooms ML](#).

8. (Facultatif) Un membre de la collaboration invoque l'inférence de modèle à l'aide du canal d'entrée ML, de l'ARN du modèle entraîné et de l'algorithme de modèle configuré par inférence. Les résultats d'inférence sont envoyés au récepteur de sortie d'inférence. Seuls les membres dotés d'une configuration ML valide et capables de recevoir des résultats d'inférence peuvent recevoir des résultats d'inférence.

Voici les étapes qui doivent être effectuées par le fournisseur de modèles :

1. Créez une image de docker Amazon ECR compatible avec l' SageMaker IA. Clean Rooms ML ne prend en charge que les images docker compatibles avec l' SageMaker IA.
2. Après avoir créé une image docker compatible avec l' SageMaker IA, transférez-la vers Amazon ECR. Suivez les instructions du [guide de l'utilisateur d'Amazon Elastic Container Registry](#) pour créer une image de formation sur les conteneurs.
3. Configurez l'algorithme du modèle à utiliser dans Clean Rooms ML.
 - a. Fournissez le lien du référentiel Amazon ECR et tous les arguments nécessaires pour configurer l'algorithme du modèle.
 - b. Fournissez un rôle d'accès au service qui permet à Clean Rooms ML d'accéder au référentiel Amazon ECR.
 - c. Associez l'algorithme du modèle configuré à la collaboration. Cela inclut la fourniture d'une politique de confidentialité qui définit les contrôles pour les journaux des conteneurs, les journaux des défaillances, CloudWatch les métriques et les limites relatives à la quantité de données pouvant être exportées à partir des résultats des conteneurs.

Voici les étapes que le fournisseur de données doit suivre pour collaborer avec un modèle de machine learning personnalisé :

1. Configurez une AWS Glue table existante avec une règle d'analyse personnalisée. Cela permet à un ensemble spécifique de requêtes préapprouvées ou de comptes préapprouvés d'utiliser vos données.
2. Associez votre table configurée à une collaboration et fournissez un rôle d'accès au service qui peut accéder à vos AWS Glue tables.
3. [Ajoutez une règle d'analyse de collaboration](#) à la table qui permet à l'association d'algorithmes du modèle configuré d'accéder à la table configurée.

4. Une fois le modèle et les données associés et configurés dans Clean Rooms ML, le membre capable d'exécuter des requêtes fournit une requête SQL et sélectionne l'algorithme du modèle à utiliser.

Une fois l'entraînement du modèle terminé, ce membre lance l'exportation des artefacts d'entraînement du modèle ou des résultats d'inférence. Ces artefacts ou résultats sont envoyés au membre capable de recevoir les résultats du modèle entraîné. Le récepteur des résultats doit les configurer `MachineLearningConfiguration` avant de pouvoir recevoir la sortie du modèle.

AWS modèles dans Clean Rooms ML

AWS Clean Rooms ML fournit une méthode préservant la confidentialité permettant à deux parties d'identifier des utilisateurs similaires dans leurs données sans avoir à partager leurs données entre elles. La première partie apporte les données d'entraînement AWS Clean Rooms afin de créer et de configurer un modèle similaire et de l'associer à une collaboration. Les données de départ sont ensuite transmises à la collaboration pour créer un segment similaire aux données d'entraînement.

Pour une explication plus détaillée de son fonctionnement, voir [Emplois multi-comptes](#).

Les rubriques suivantes fournissent des informations sur la création et la configuration de AWS modèles dans Clean Rooms ML.

Rubriques

- [Protection de la confidentialité d'AWS Clean Rooms ML](#)
- [Exigences relatives aux données de formation pour Clean Rooms ML](#)
- [Exigences relatives aux données de base pour Clean Rooms ML](#)
- [Mesures d'évaluation du modèle AWS Clean Rooms ML](#)

Protection de la confidentialité d'AWS Clean Rooms ML

Clean Rooms ML est conçu pour réduire le risque d'attaques par inférence d'adhésion, dans le cadre desquelles le fournisseur de données de formation peut savoir qui figure dans les données de départ et le fournisseur de données de départ peut savoir qui figure dans les données d'entraînement. Plusieurs mesures sont prises pour empêcher cette attaque.

Tout d'abord, les fournisseurs de données de départ n'observent pas directement les résultats de Clean Rooms ML et les fournisseurs de données de formation ne peuvent jamais observer les

données de départ. Les fournisseurs de données de départ peuvent choisir d'inclure les données de départ dans le segment de sortie.

Ensuite, le modèle similaire est créé à partir d'un échantillon aléatoire des données d'entraînement. Cet échantillon inclut un nombre important d'utilisateurs qui ne correspondent pas à l'audience initiale. Ce processus rend plus difficile de déterminer si un utilisateur ne figurait pas dans les données, ce qui constitue un autre moyen de déduire son appartenance.

De plus, plusieurs clients de semences peuvent être utilisés pour chaque paramètre de la formation d'un modèle similaire spécifique à une graine. Cela limite le surajustement du modèle, et donc ce qui peut être déduit à propos d'un utilisateur. Par conséquent, nous recommandons que la taille minimale des données de départ soit de 500 utilisateurs.

Enfin, les indicateurs au niveau des utilisateurs ne sont jamais fournis aux fournisseurs de données de formation, ce qui élimine toute autre possibilité d'attaque par inférence d'adhésion.

Exigences relatives aux données de formation pour Clean Rooms ML

Pour réussir à créer un modèle similaire, vos données d'entraînement doivent répondre aux exigences suivantes :

- Les données d'entraînement doivent être au format Parquet, CSV ou JSON.

Note

Les données Parquet compressées au format Zstandard (ZSTD) ne sont pas prises en charge.

- Vos données d'entraînement doivent être cataloguées dans AWS Glue. Pour plus d'informations, consultez [Getting started with the AWS Glue Data Catalog](#) dans le manuel du AWS Glue développeur. Nous vous recommandons d'utiliser AWS Glue des robots d'exploration pour créer vos tables, car le schéma est déduit automatiquement.
- Le compartiment Amazon S3 qui contient les données d'entraînement et les données de départ se trouve dans la même AWS région que vos autres ressources Clean Rooms ML.
- Les données d'entraînement doivent contenir au moins 100 000 utilisateurs uniques IDs ayant chacun au moins deux interactions avec des éléments.
- Les données d'entraînement doivent contenir au moins 1 million d'enregistrements.
- Le schéma spécifié dans l'[CreateTrainingDataset](#) action doit être aligné sur le schéma défini lors de la création de la AWS Glue table.

- Les champs obligatoires, tels que définis dans le tableau fourni, sont définis dans l'[CreateTrainingDataset](#) action.

Type de champ	Types de données pris en charge	Obligatoire	Description
USER_ID	chaîne, int, bigint	Oui	Un identifiant unique pour chaque utilisateur de l'ensemble de données. Il doit s'agir d'une valeur d'information non personnellement identifiable (PII). Il peut s'agir d'un identifiant haché ou d'un identifiant client.

Type de champ	Types de données pris en charge	Obligatoire	Description
ITEM_ID	chaîne, int, bigint	Oui	Un identifiant unique pour chaque élément avec lequel un utilisateur interagit.
TIMESTAMP	bigint, int, horodatage	Oui	Heure à laquelle un utilisateur a interagi avec l'élément. Les valeurs doivent être au format Epoch Time d'Unix en secondes.

Type de champ	Types de données pris en charge	Obligatoire	Description
FONCTIONNALITÉ_CATÉGORIQUE	chaîne, int, float, bigint, double, booléen, tableau	Non	Capture les données catégoriques relatives à l'utilisateur ou à l'article . Cela peut inclure des éléments tels que le type d'événement (tel qu'un clic ou un achat), les données démographiques de l'utilisateur (groupe d'âge, sexe, anonymisé), la localisation de l'utilisateur (ville, pays, anonymisé), la catégorie

Type de champ	Types de données pris en charge	Obligatoire	Description
			d'article (vêtement s ou appareils électroniques, par exemple) ou la marque de l'article.

Type de champ	Types de données pris en charge	Obligatoire	Description
FONCTIONNALITÉ_NUMÉRIQUE	double, float, int, bigint	Non	Capture les données numériques relatives à l'utilisateur ou à l'article . Cela peut inclure des éléments tels que l'historique des achats des utilisateurs (montant total dépensé), le prix de l'article, le nombre de fois où un article a été visité ou les évaluations des utilisateurs pour les articles.

- Vous pouvez éventuellement fournir jusqu'à 10 caractéristiques catégorielles ou numériques au total.

Voici un exemple d'ensemble de données d'entraînement valide au format CSV

```
USER_ID,ITEM_ID,TIMESTAMP,EVENT_TYPE(CATEGORICAL FEATURE),EVENT_VALUE (NUMERICAL FEATURE)
```

```
196,242,881250949,click,15
```

```
186,302,891717742,click,13
```

```
22,377,878887116,click,10
```

```
244,51,880606923,click,20
```

```
166,346,886397596,click,10
```

Exigences relatives aux données de base pour Clean Rooms ML

Les données de départ d'un modèle similaire peuvent provenir directement d'un compartiment Amazon S3 ou des résultats d'une requête SQL.

Les données sur les semences fournies directement doivent répondre aux exigences suivantes :

- Les données de départ doivent être au format de lignes JSON avec une liste d'utilisateurs IDs.
- La taille de la graine doit être comprise entre 25 et 500 000 utilisateurs uniques IDs.
- Le nombre minimum d'utilisateurs de départ doit correspondre à la valeur de taille de départ minimale correspondante spécifiée lors de la création du modèle d'audience configuré.

Voici un exemple d'ensemble de données d'entraînement valide au format CSV

```
{"user_id": "abc"}  
{"user_id": "def"}  
{"user_id": "ghijkl"}  
{"user_id": "123"}  
{"user_id": "456"}  
{"user_id": "7890"}
```

Mesures d'évaluation du modèle AWS Clean Rooms ML

Clean Rooms ML calcule le score de rappel et de pertinence pour déterminer les performances de votre modèle. Recall compare la similitude entre les données similaires et les données d'entraînement. Le score de pertinence est utilisé pour déterminer la taille de l'audience, et non pour déterminer si le modèle est performant.

Le rappel est une mesure impartiale de la similitude entre le segment similaire et les données d'entraînement. Le rappel est le pourcentage d'utilisateurs les plus similaires (par défaut, les 20 % les

plus similaires) à partir d'un échantillon de données de formation inclus dans l'audience initiale par la tâche de génération d'audience. Les valeurs vont de 0 à 1, les valeurs les plus élevées indiquent une meilleure audience. Une valeur de rappel approximativement égale au pourcentage maximal de bacs indique que le modèle d'audience est équivalent à une sélection aléatoire.

Nous considérons qu'il s'agit d'un meilleur indicateur d'évaluation que l'exactitude, la précision et les scores F1, car Clean Rooms ML n'a pas correctement étiqueté les utilisateurs réellement négatifs lors de la création de son modèle.

Le score de pertinence au niveau du segment est une mesure de similarité avec des valeurs allant de -1 (le moins similaire) à 1 (le plus similaire). Clean Rooms ML calcule un ensemble de scores de pertinence pour différentes tailles de segment afin de vous aider à déterminer la meilleure taille de segment pour vos données. Les scores de pertinence diminuent de façon monotone à mesure que la taille du segment augmente. Ainsi, lorsque la taille du segment augmente, il peut être moins similaire aux données de départ. Lorsque le score de pertinence au niveau du segment atteint 0, le modèle prédit que tous les utilisateurs du segment similaire appartiennent à la même distribution que les données initiales. L'augmentation de la taille de sortie est susceptible d'inclure dans le segment similaire des utilisateurs qui ne proviennent pas de la même distribution que les données de départ.

Les scores de pertinence sont normalisés au sein d'une même campagne et ne doivent pas être utilisés pour comparer les campagnes. Les scores de pertinence ne doivent pas être utilisés comme des preuves provenant d'une source unique pour un résultat commercial, car ils sont influencés par de multiples facteurs complexes en plus de la pertinence, tels que la qualité des stocks, le type d'inventaire, le calendrier des publicités, etc.

Les scores de pertinence ne doivent pas être utilisés pour juger de la qualité de la graine, mais plutôt pour déterminer si elle peut être augmentée ou diminuée. Considérez les exemples suivants :

- Tous les scores sont positifs : cela indique que le nombre d'utilisateurs prédits comme similaires est supérieur au nombre d'utilisateurs inclus dans le segment similaire. Cela est courant pour les données sur les semences qui font partie d'un vaste marché, comme pour tous ceux qui ont acheté du dentifrice le mois dernier. Nous vous recommandons de consulter des données sur des semences plus petites, comme celles de tous ceux qui ont acheté du dentifrice plus d'une fois au cours du dernier mois.
- Tous les scores sont négatifs ou négatifs pour la taille de segment de sosie souhaitée : cela indique que Clean Rooms ML prédit qu'il n'y a pas assez d'utilisateurs similaires dans la taille de segment de sosie souhaitée. Cela peut être dû au fait que les données sur les semences sont trop spécifiques ou que le marché est trop petit. Nous recommandons soit d'appliquer moins de

filtres aux données sur les semences, soit d'élargir le marché. Par exemple, si les données initiales concernaient des clients ayant acheté une poussette et un siège auto, vous pourriez étendre le marché aux clients ayant acheté plusieurs produits pour bébés.

Les fournisseurs de données de formation déterminent si les scores de pertinence sont exposés et les compartiments dans lesquels les scores de pertinence sont calculés.

Modèles personnalisés dans Clean Rooms ML

Avec Clean Rooms ML, les membres d'une collaboration peuvent utiliser un algorithme de modèle personnalisé dockerisé stocké dans Amazon ECR pour analyser conjointement leurs données. Pour ce faire, le fournisseur de modèles doit créer une image et la stocker dans Amazon ECR. Suivez les étapes décrites dans le [guide de l'utilisateur d'Amazon Elastic Container Registry](#) pour créer un référentiel privé qui contiendra le modèle de ML personnalisé.

Tout membre d'une collaboration peut être le fournisseur de modèles, à condition de disposer des autorisations appropriées. Tous les membres d'une collaboration peuvent apporter des données au modèle. Aux fins du présent guide, les membres fournissant des données sont appelés fournisseurs de données. Le membre qui crée la collaboration est le créateur de la collaboration, et ce membre peut être le fournisseur de modèles, l'un des fournisseurs de données ou les deux.

Les rubriques suivantes décrivent les informations nécessaires pour créer un modèle de machine learning personnalisé.

Rubriques

- [Prérequis pour la modélisation ML personnalisée](#)
- [Consignes de création de modèles pour le conteneur de formation](#)
- [Directives de création de modèles pour le conteneur d'inférence](#)
- [Réception des journaux et des métriques du modèle](#)

Prérequis pour la modélisation ML personnalisée

Avant de pouvoir effectuer une modélisation ML personnalisée, vous devez prendre en compte les points suivants :

- Déterminez si la formation du modèle et l'inférence sur le modèle entraîné seront effectuées dans le cadre de la collaboration.

- Déterminez le rôle que chaque membre de la collaboration jouera et attribuez-lui les compétences appropriées.
 - Attribuez `CAN_QUERY` cette capacité au membre qui entraînera le modèle et exécutera l'inférence sur le modèle entraîné.
 - Attribuez le `CAN_RECEIVE_RESULTS` à au moins un membre de la collaboration.
 - `CAN_RECEIVE_MODEL_OUTPUT`Attribuez `CAN_RECEIVE_INFERENCE_OUTPUT` des capacités au membre qui recevra des exportations de modèles entraînés ou des sorties d'inférence, respectivement. Vous pouvez choisir d'utiliser les deux capacités si elles sont requises par votre cas d'utilisation.
- Déterminez la taille maximale des artefacts du modèle entraîné ou des résultats d'inférence dont vous autoriserez l'exportation.
- Nous recommandons que tous les utilisateurs aient les `CleanroomsMLFullAccess` politiques `CleanroomsFullAccess` et associées à leur rôle. L'utilisation de modèles de ML personnalisés nécessite d'utiliser à la fois le ML AWS Clean Rooms et AWS Clean Rooms ML SDKs.
- Tenez compte des informations suivantes concernant les rôles IAM.
 - Tous les fournisseurs de données doivent avoir un rôle d'accès aux services qui leur permet AWS Clean Rooms de lire les données de leurs AWS Glue catalogues et de leurs tables, ainsi que des emplacements Amazon S3 sous-jacents. Ces rôles sont similaires à ceux requis pour les requêtes SQL. Cela vous permet d'utiliser l'`CreateConfiguredTableAssociation` action. Pour de plus amples informations, veuillez consulter [Créez un rôle de service pour créer une association de tables configurée](#).
 - Tous les membres qui souhaitent recevoir des métriques doivent disposer d'un rôle d'accès au service qui leur permet de rédiger CloudWatch des métriques et des journaux. Ce rôle est utilisé par Clean Rooms ML pour écrire toutes les métriques et les journaux du modèle dans ceux des membres Compte AWS lors de l'entraînement et de l'inférence du modèle. Nous fournissons également des contrôles de confidentialité pour déterminer quels membres ont accès aux statistiques et aux journaux. Cela vous permet d'utiliser l'`CreateMLConfiguration` action. Pour plus d'informations, voir, [Création d'un rôle de service pour la modélisation ML personnalisée - Configuration ML](#).

Le membre recevant les résultats doit fournir un rôle d'accès au service avec l'autorisation d'écrire dans son compartiment Amazon S3. Ce rôle permet à Clean Rooms ML d'exporter les résultats (artefacts de modèles entraînés ou résultats d'inférence) vers un compartiment Amazon S3. Cela vous permet d'utiliser l'`CreateMLConfiguration` action. Pour de plus

amples informations, veuillez consulter [Création d'un rôle de service pour la modélisation ML personnalisée - Configuration ML](#).

- Le fournisseur de modèles doit fournir un rôle d'accès au service avec des autorisations pour lire son référentiel et son image Amazon ECR. Cela vous permet d'utiliser l'CreateConfigureModelAlgorithm action. Pour de plus amples informations, veuillez consulter [Créez un rôle de service pour fournir un modèle de machine learning personnalisé](#).
- Le membre qui crée le pour générer des ensembles de données MLInputChannel à des fins d'entraînement ou d'inférence doit fournir un rôle d'accès au service qui permet à Clean Rooms ML d'exécuter une requête SQL dans. AWS Clean Rooms Cela vous permet d'utiliser les StartTrainedModelInferenceJob actions CreateTrainedModel et. Pour de plus amples informations, veuillez consulter [Création d'un rôle de service pour interroger un ensemble de données](#).
- Les auteurs du modèle doivent suivre les [Consignes de création de modèles pour le conteneur de formation](#) et s'[Directives de création de modèles pour le conteneur d'inférence](#) assurer que les entrées et sorties du modèle sont configurées comme prévu par AWS Clean Rooms.

Consignes de création de modèles pour le conteneur de formation

Cette section détaille les directives que les fournisseurs de modèles doivent suivre lors de la création d'un algorithme de modèle ML personnalisé pour Clean Rooms ML.

- Utilisez l'image de base de conteneur appropriée prise en charge par la formation à l' SageMaker IA, comme décrit dans le guide du [développeur d'SageMaker IA](#). Le code suivant vous permet d'extraire les images de base de conteneurs prises en charge à partir de points de terminaison d' SageMaker IA publics.

```
ecr_registry_endpoint='763104351884.dkr.ecr.$REGION.amazonaws.com'  
base_image='pytorch-training:2.3.0-cpu-py311-ubuntu20.04-sagemaker'  
aws ecr get-login-password --region $REGION | docker login --username AWS --password-  
stdin $ecr_registry_endpoint  
docker pull $ecr_registry_endpoint/$base_image
```

- Lorsque vous créez le modèle localement, assurez-vous de ce qui suit afin de pouvoir le tester localement, sur une instance de développement, sur SageMaker AI Training in your Compte AWS et sur Clean Rooms ML.
- Nous vous recommandons d'écrire un script d'entraînement qui accède aux propriétés utiles de l'environnement d'entraînement par le biais de diverses variables d'environnement. Clean Rooms

ML utilise les arguments suivants pour appeler l'entraînement sur le code de votre modèle : SM_MODEL_DIR, SM_OUTPUT_DIR, SM_CHANNEL_TRAIN, et FILE_FORMAT. Ces valeurs par défaut sont utilisées par Clean Rooms ML pour entraîner votre modèle ML dans son propre environnement d'exécution avec les données de toutes les parties.

- Clean Rooms ML met à disposition vos canaux d'entrée d'entraînement via les `/opt/ml/input/data/channel-name` répertoires du conteneur docker. Chaque canal d'entrée ML est mappé en fonction du canal correspondant `channel_name` fourni dans la `CreateTrainedModel` demande.

```
parser = argparse.ArgumentParser()# Data, model, and output directories

parser.add_argument('--model_dir', type=str, default=os.environ.get('SM_MODEL_DIR',
"/opt/ml/model"))
parser.add_argument('--output_dir', type=str,
default=os.environ.get('SM_OUTPUT_DIR', "/opt/ml/output/data"))
parser.add_argument('--train_dir', type=str,
default=os.environ.get('SM_CHANNEL_TRAIN', "/opt/ml/input/data/train"))
parser.add_argument('--train_file_format', type=str,
default=os.environ.get('FILE_FORMAT', "csv"))
```

- Assurez-vous de pouvoir générer un ensemble de données synthétique ou de test basé sur le schéma des collaborateurs qui sera utilisé dans le code de votre modèle.
- Assurez-vous de pouvoir exécuter vous-même une tâche de formation à l' SageMaker IA Compte AWS avant d'associer l'algorithme du modèle à une AWS Clean Rooms collaboration.

Le code suivant contient un exemple de fichier Docker compatible avec les tests locaux, les tests de l'environnement SageMaker AI Training et Clean Rooms ML

```
FROM 763104351884.dkr.ecr.us-west-2.amazonaws.com/pytorch-training:2.3.0-cpu-
py311-ubuntu20.04-sagemaker
MAINTAINER $author_name

ENV PYTHONDONTWRITEBYTECODE=1 \
    PYTHONUNBUFFERED=1 \
    LD_LIBRARY_PATH="${LD_LIBRARY_PATH}:/usr/local/lib"

ENV PATH="/opt/ml/code:${PATH}"

# this environment variable is used by the SageMaker PyTorch container to determine
our user code directory
ENV SAGEMAKER_SUBMIT_DIRECTORY /opt/ml/code
```

```
# copy the training script inside the container
COPY train.py /opt/ml/code/train.py
# define train.py as the script entry point
ENV SAGEMAKER_PROGRAM train.py
ENTRYPOINT ["python", "/opt/ml/code/train.py"]
```

- Pour mieux surveiller les défaillances des conteneurs, nous vous recommandons d'exporter les journaux et de procéder au débogage en cas de défaillance. Dans une `GetTrainedModel` réponse, Clean Rooms ML renvoie les 1024 premiers caractères du fichier `ci-dessousStatusDetails`.
- Une fois que vous avez effectué les modifications du modèle et que vous êtes prêt à le tester dans l'environnement d' SageMaker IA, exécutez les commandes suivantes dans l'ordre indiqué.

```
export ACCOUNT_ID=xxx
export REPO_NAME=xxx
export REPO_TAG=xxx
export REGION=xxx

docker build -t $ACCOUNT_ID.dkr.ecr.us-west-2.amazonaws.com/$REPO_NAME:$REPO_TAG

# Sign into AWS $ACCOUNT_ID/ Run aws configure
# Check the account and make sure it is the correct role/credentials
aws sts get-caller-identity
aws ecr create-repository --repository-name $REPO_NAME --region $REGION
aws ecr describe-repositories --repository-name $REPO_NAME --region $REGION

# Authenticate Docker
aws ecr get-login-password --region $REGION | docker login --username AWS --password-stdin $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com

# Push To ECR Images
docker push $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com$REPO_NAME:$REPO_TAG

# Create Sagemaker Training job
# Configure the training_job.json with
# 1. TrainingImage
# 2. Input DataConfig
# 3. Output DataConfig
aws sagemaker create-training-job --cli-input-json file:///training_job.json --region $REGION
```

Une fois que le travail d' SageMaker intelligence artificielle est terminé et que vous êtes satisfait de l'algorithme de votre modèle, vous pouvez enregistrer le registre Amazon ECR auprès d'AWS Clean Rooms ML. Utilisez cette `CreateConfiguredModelAlgorithm` action pour enregistrer l'algorithme du modèle et `CreateConfiguredModelAlgorithmAssociation` pour l'associer à une collaboration.

Directives de création de modèles pour le conteneur d'inférence

Cette section détaille les directives que les fournisseurs de modèles doivent suivre lors de la création d'un algorithme d'inférence pour Clean Rooms ML.

- Utilisez l'image de base de conteneur compatible avec l'inférence SageMaker AI appropriée, comme décrit dans le Guide du [développeur SageMaker AI](#). Le code suivant vous permet d'extraire les images de base de conteneurs prises en charge à partir de points de terminaison d' SageMaker IA publics.

```
ecr_registry_endpoint='763104351884.dkr.ecr.$REGION.amazonaws.com'  
base_image='pytorch-inference:2.3.0-cpu-py311-ubuntu20.04-sagemaker'  
aws ecr get-login-password --region $REGION | docker login --username AWS --password-  
stdin $ecr_registry_endpoint  
docker pull $ecr_registry_endpoint/$base_image
```

- Lorsque vous créez le modèle localement, assurez-vous de ce qui suit afin de pouvoir le tester localement, sur une instance de développement, sur SageMaker AI Batch Transform dans votre entreprise Compte AWS et sur Clean Rooms ML.
 - Clean Rooms ML met les artefacts de votre modèle issus de l'inférence à la disposition de votre code d'inférence via le `/opt/ml/model` répertoire du conteneur docker.
 - Clean Rooms ML divise les entrées par ligne, utilise une stratégie par `MultiRecord` lots et ajoute un caractère de nouvelle ligne à la fin de chaque enregistrement transformé.
 - Assurez-vous de pouvoir générer un jeu de données synthétique ou d'inférence de test basé sur le schéma des collaborateurs qui sera utilisé dans le code de votre modèle.
 - Assurez-vous de pouvoir exécuter vous-même une tâche de transformation par lots basée sur l' SageMaker IA Compte AWS avant d'associer l'algorithme du modèle à une AWS Clean Rooms collaboration.

Le code suivant contient un exemple de fichier Docker compatible avec les tests locaux, les tests d'environnement SageMaker AI Transform et Clean Rooms ML

```
FROM 763104351884.dkr.ecr.us-east-1.amazonaws.com/pytorch-inference:1.12.1-cpu-
py38-ubuntu20.04-sagemaker

ENV PYTHONUNBUFFERED=1

COPY serve.py /opt/ml/code/serve.py
COPY inference_handler.py /opt/ml/code/inference_handler.py
COPY handler_service.py /opt/ml/code/handler_service.py
COPY model.py /opt/ml/code/model.py

RUN chmod +x /opt/ml/code/serve.py

ENTRYPOINT ["/opt/ml/code/serve.py"]
```

- Une fois que vous avez effectué les modifications du modèle et que vous êtes prêt à le tester dans l'environnement d' SageMaker IA, exécutez les commandes suivantes dans l'ordre indiqué.

```
export ACCOUNT_ID=xxx
export REPO_NAME=xxx
export REPO_TAG=xxx
export REGION=xxx

docker build -t $ACCOUNT_ID.dkr.ecr.us-west-2.amazonaws.com/$REPO_NAME:$REPO_TAG

# Sign into AWS $ACCOUNT_ID/ Run aws configure
# Check the account and make sure it is the correct role/credentials
aws sts get-caller-identity
aws ecr create-repository --repository-name $REPO_NAME --region $REGION
aws ecr describe-repositories --repository-name $REPO_NAME --region $REGION

# Authenticate Docker
aws ecr get-login-password --region $REGION | docker login --username AWS --password-
stdin $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com

# Push To ECR Repository
docker push $ACCOUNT_ID.dkr.ecr.$REGION.amazonaws.com$REPO_NAME:$REPO_TAG

# Create Sagemaker Model
# Configure the create_model.json with
# 1. Primary container -
#   a. ModelDataUrl - S3 Uri of the model.tar from your training job
aws sagemaker create-model --cli-input-json file://create_model.json --region $REGION
```

```
# Create Sagemaker Transform Job
# Configure the transform_job.json with
# 1. Model created in the step above
# 2. MultiRecord batch strategy
# 3. Line SplitType for TransformInput
# 4. AssembleWith Line for TransformOutput
aws sagemaker create-transform-job --cli-input-json file://transform_job.json --
region $REGION
```

Une fois que le travail d' SageMaker intelligence artificielle est terminé et que vous êtes satisfait de votre transformation par lots, vous pouvez enregistrer le registre Amazon ECR auprès d'AWS Clean Rooms ML. Utilisez cette `CreateConfiguredModelAlgorithm` action pour enregistrer l'algorithme du modèle et `CreateConfiguredModelAlgorithmAssociation` pour l'associer à une collaboration.

Réception des journaux et des métriques du modèle

Pour recevoir des journaux et des métriques issus de l'entraînement ou de l'inférence d'un modèle personnalisé, les membres doivent avoir [créé une configuration ML](#) avec un rôle valide fournissant les CloudWatch autorisations nécessaires (voir [Création d'un rôle de service pour une modélisation ML personnalisée - Configuration ML](#)).

Métrique du système

Les métriques du système pour l'entraînement et l'inférence, telles que l'utilisation du processeur et de la mémoire, sont publiées à tous les membres en collaboration avec des configurations ML valides. Ces métriques peuvent être consultées au fur et à mesure de l'avancement de la tâche via CloudWatch Metrics in the `/aws/cleanroomsm1/TrainedModels` or `/aws/cleanroomsm1/TrainedModelInferenceJobs` namespaces, respectivement.

Logs de modèles

L'accès aux journaux du modèle est fourni par la politique de configuration de confidentialité de chaque algorithme de modèle configuré. L'auteur du modèle définit la politique de configuration de confidentialité lorsqu'il associe un algorithme de modèle configuré (via la console ou l'`CreateConfiguredModelAlgorithmAssociationAPI`) à une collaboration. La définition de la politique de configuration de confidentialité permet de contrôler quels membres peuvent recevoir les journaux du modèle.

En outre, l'auteur du modèle peut définir un modèle de filtre dans la politique de configuration de confidentialité pour filtrer les événements du journal. Tous les journaux qu'un conteneur modèle envoie vers `stdout` ou `stderr` qui correspondent au modèle de filtre (s'il est défini) sont envoyés à Amazon CloudWatch Logs. Les journaux des modèles sont disponibles dans des groupes de CloudWatch journaux `/aws/cleanroomsm1/TrainedModels` ou `/aws/cleanroomsm1/TrainedModelInferenceJobs`, respectivement.

Métriques définies sur mesure

Lorsque vous configurez un algorithme de modèle (via la console ou l'`CreateConfiguredModelAlgorithmAPI`), l'auteur du modèle peut fournir des noms de métriques spécifiques et des instructions regex à rechercher dans les journaux de sortie. Ils peuvent être consultés au fur et à mesure de l'avancement de la tâche via CloudWatch Metrics dans l'espace de `/aws/cleanroomsm1/TrainedModels` noms. Lorsqu'il associe un algorithme de modèle configuré, l'auteur du modèle peut définir un niveau de bruit facultatif dans la configuration de confidentialité des métriques afin d'éviter de générer des données brutes tout en fournissant une visibilité sur les tendances des métriques personnalisées. Si un niveau de bruit est défini, les mesures sont publiées à la fin de la tâche plutôt qu'en temps réel.

Informatique cryptographique pour Clean Rooms

Informatique cryptographique pour Clean Rooms (C3R) est une fonctionnalité AWS Clean Rooms qui peut être utilisée en plus des [règles d'analyse](#). Avec C3R, les entreprises peuvent rassembler des données sensibles pour tirer de nouvelles informations de l'analyse des données tout en limitant cryptographiquement ce que les parties prenantes peuvent apprendre au cours du processus. Le C3R peut être utilisé par deux ou plusieurs parties qui souhaitent collaborer avec leurs données sensibles, mais doivent uniquement utiliser des données cryptées dans le cloud.

Le client de chiffrement C3R est un outil de chiffrement côté client que vous pouvez utiliser pour [chiffrer](#) vos données à utiliser. AWS Clean Rooms Lorsque vous utilisez le client de chiffrement C3R, les données restent protégées cryptographiquement pendant leur utilisation dans le cadre d'une AWS Clean Rooms collaboration. Comme dans le cas d'une AWS Clean Rooms collaboration normale, les données d'entrée sont des tables de base de données relationnelles, et le calcul est exprimé sous forme de requête SQL. Cependant, C3R ne prend en charge qu'un sous-ensemble limité de requêtes SQL sur des données chiffrées.

Plus précisément, C3R prend en charge le langage SQL JOIN and SELECT déclarations relatives aux données protégées par cryptographie. Chaque colonne de la table d'entrée peut être utilisée dans exactement l'un des types d'instructions SQL suivants :

- Colonne protégée par cryptographie pour une utilisation dans JOIN les déclarations sont appelées fingerprint colonnes.
- Colonne protégée par cryptographie pour une utilisation dans SELECT les déclarations sont appelées sealed colonnes.
- Colonne qui ne sont pas protégées par cryptographie pour être utilisées dans JOIN or SELECT les déclarations sont appelées cleartext colonnes.

Dans certains cas, GROUP BY les déclarations sont prises en charge sur fingerprint colonnes. Pour de plus amples informations, veuillez consulter [Fingerprint colonnes](#). Actuellement, C3R ne prend pas en charge l'utilisation d'autres constructions SQL sur des données chiffrées, telles que WHERE des clauses ou des fonctions agrégées telles que SUM and AVERAGE, même s'ils seraient autrement autorisés par les règles d'analyse pertinentes.

Le C3R est conçu pour protéger les données contenues dans les cellules individuelles d'un tableau. En utilisant la configuration par défaut de C3R, les données sous-jacentes qu'un client met à la disposition de tiers dans le cadre d'une collaboration restent cryptées tant que le contenu est utilisé dans le cadre AWS Clean Rooms de cette collaboration. C3R utilise le cryptage AES-GCM standard pour tous sealed colonnes et une fonction pseudo-aléatoire standard, connue sous le nom de code d'authentification des messages basé sur le hachage (HMAC), pour protéger fingerprint colonnes.

Même si C3R chiffre les données de vos tables, les informations suivantes peuvent toujours être déduites :

- Informations sur les tables elles-mêmes, notamment le nombre de colonnes, les noms des colonnes et le nombre de lignes de votre tableau.
- Comme pour la plupart des formes de chiffrement standard, C3R n'essaie pas de masquer la longueur des valeurs chiffrées. C3R offre la possibilité de compléter des valeurs chiffrées pour masquer la longueur exacte des textes en clair. Cependant, une limite supérieure de la longueur des textes en clair dans chaque colonne pourrait tout de même être révélée à une autre partie.
- Informations au niveau de la journalisation, telles que le moment où une ligne particulière a été ajoutée à une table C3R cryptée.

Pour plus d'informations sur le C3R, consultez les rubriques suivantes.

Rubriques

- [Considérations relatives à l'utilisation de l'informatique cryptographique pour Clean Rooms](#)
- [Types de fichiers et de données pris en charge dans Cryptographic Computing pour Clean Rooms](#)
- [Noms de colonnes dans Cryptographic Computing pour Clean Rooms](#)
- [Types de colonnes dans le calcul cryptographique pour Clean Rooms](#)
- [Paramètres de calcul cryptographique](#)
- [Drapeaux facultatifs dans le calcul cryptographique pour Clean Rooms](#)
- [Requêtes avec informatique cryptographique pour Clean Rooms](#)
- [Directives pour le client de chiffrement C3R](#)

Considérations relatives à l'utilisation de l'informatique cryptographique pour Clean Rooms

Cryptographic Computing for Clean Rooms (C3R) vise à optimiser la protection des données. Toutefois, certains cas d'utilisation peuvent bénéficier de niveaux inférieurs de protection des données en échange de fonctionnalités supplémentaires. Vous pouvez faire ces compromis spécifiques en modifiant C3R à partir de sa configuration la plus sécurisée. En tant que client, vous devez être conscient de ces compromis et déterminer s'ils sont adaptés à votre cas d'utilisation. Les compromis à prendre en compte sont les suivants :

Rubriques

- [Autoriser les données mixtes cleartext et cryptées dans vos tables](#)
- [Autoriser les valeurs répétées dans les fingerprint colonnes](#)
- [Assouplissement des restrictions relatives au fingerprint nom des colonnes](#)
- [Déterminer comment NULL les valeurs sont représentées](#)

Pour plus d'informations sur la façon de définir les paramètres de ces scénarios, consultez [Paramètres de calcul cryptographique](#).

Autoriser les données mixtes cleartext et cryptées dans vos tables

Le chiffrement de toutes les données côté client garantit une protection maximale des données. Cela limite toutefois certains types de requêtes (par exemple, la fonction d'SUMagrégation). Le risque lié à l'autorisation des cleartext données est qu'il est possible que toute personne ayant accès aux tables

cryptées puisse déduire certaines informations sur les valeurs cryptées. Cela pourrait être fait en effectuant une analyse statistique des données cleartext et des données associées.

Par exemple, imaginez que vous disposiez des colonnes de `City` et `State`. La `City` colonne est chiffrée cleartext et `State` elle est cryptée. Lorsque vous voyez la valeur `Chicago` dans la `City` colonne, cela vous permet de déterminer avec une forte probabilité que `State` c'est le `casIllinois`. En revanche, si une colonne l'est `City` et l'autre l'est `EmailAddress`, il cleartext `City` est peu probable que a révèle quoi que ce soit à propos d'un chiffré `EmailAddress`.

Pour plus d'informations sur le paramètre de ce scénario, consultez [Paramètre Autoriser cleartext les colonnes](#).

Autoriser les valeurs répétées dans les fingerprint colonnes

Pour l'approche la plus sûre, nous supposons que chaque fingerprint colonne contient exactement une instance d'une variable. Aucun élément ne peut être répété dans une fingerprint colonne. Le client de chiffrement C3R mappe ces cleartext valeurs en valeurs uniques impossibles à distinguer des valeurs aléatoires. Il est donc impossible de déduire des informations les concernant à cleartext partir de ces valeurs aléatoires.

Le risque de valeurs répétées dans une fingerprint colonne est que les valeurs répétées se traduisent par des valeurs répétées d'apparence aléatoire. Ainsi, toute personne ayant accès aux tables cryptées pourrait, en théorie, effectuer une analyse statistique des fingerprint colonnes susceptible de révéler des informations sur cleartext les valeurs.

Encore une fois, supposons que la fingerprint colonne soit `State`, et que chaque ligne du tableau corresponde à un ménage américain. En effectuant une analyse de fréquence, on pourrait déduire quel état est `California` et lequel est `Wyoming` avec une probabilité élevée. Cette inférence est possible car il `California` compte beaucoup plus de résidents que. `Wyoming` En revanche, supposons que la fingerprint colonne porte sur un identifiant de ménage et que chaque ménage apparaisse dans la base de données entre 1 et 4 fois dans une base de données contenant des millions d'entrées. Il est peu probable qu'une analyse de fréquence révèle des informations utiles.

Pour plus d'informations sur le paramètre de ce scénario, consultez [Paramètre Autoriser les doublons](#).

Assouplissement des restrictions relatives au fingerprint nom des colonnes

Par défaut, nous supposons que lorsque deux tables sont jointes à l'aide de fingerprint colonnes chiffrées, ces colonnes portent le même nom dans chaque table. La raison technique de ce résultat est que, par défaut, nous dérivons une clé cryptographique différente pour chiffrer chaque fingerprint

colonne. Cette clé est dérivée d'une combinaison de la clé secrète partagée pour la collaboration et du nom de colonne. Si nous essayons de joindre deux colonnes portant des noms de colonne différents, nous dérivons des clés différentes et nous ne pouvons pas calculer une jointure valide.

Pour résoudre ce problème, vous pouvez désactiver la fonctionnalité qui déduit les clés du nom de chaque colonne. Le client de chiffrement C3R utilise ensuite une clé dérivée unique pour toutes les fingerprint colonnes. Le risque est qu'un autre type d'analyse de fréquence puisse être effectué qui pourrait révéler des informations.

Utilisons à nouveau l'exemple City et. Si nous dérivons les mêmes valeurs aléatoires pour chaque fingerprint colonne (en n'incorporant pas le nom de la colonne). New York possède la même valeur aléatoire dans les State colonnes City et. New York est l'une des rares villes des États-Unis où le City nom est le même que le State nom. En revanche, si votre ensemble de données contient des valeurs complètement différentes dans chaque colonne, aucune information n'est divulguée.

Pour plus d'informations sur le paramètre de ce scénario, consultez [Paramètre JOIN d'autorisation des colonnes avec des noms différents](#).

Déterminer comment NULL les valeurs sont représentées

L'option qui s'offre à vous est de savoir s'il faut traiter les NULL valeurs de manière cryptographique (chiffrement et HMAC) comme toute autre valeur. Si vous ne traitez pas NULL les valeurs comme les autres valeurs, des informations peuvent être révélées.

Supposons, par exemple, que NULL la Middle Name colonne cleartext indique les personnes sans deuxième prénom. Si vous ne chiffrez pas ces valeurs, vous divulguez les lignes de la table cryptée qui sont utilisées pour les personnes sans deuxième prénom. Ces informations peuvent être un signal d'identification pour certaines personnes dans certaines populations. Mais si vous traitez des NULL valeurs de manière cryptographique, certaines requêtes SQL agissent différemment. Par exemple, les GROUP BY clauses ne regrouperont pas fingerprint NULL les valeurs dans fingerprint des colonnes.

Pour plus d'informations sur le paramètre de ce scénario, consultez [Paramètre de conservation NULL des valeurs](#).

Types de fichiers et de données pris en charge dans Cryptographic Computing pour Clean Rooms

Le client de chiffrement C3R reconnaît les types de fichiers suivants :

- fichiers CSV
- Parquetfichiers

Vous pouvez utiliser l'`--fileFormat`indicateur du client de chiffrement C3R pour spécifier un format de fichier de manière explicite. Lorsqu'il est explicitement spécifié, le format de fichier n'est pas déterminé par l'extension du fichier.

Rubriques

- [fichiers CSV](#)
- [Parquetfichiers](#)
- [Chiffrement de valeurs autres que des chaînes](#)

fichiers CSV

Un fichier portant une extension `.csv` est supposé être au format CSV et contenir du texte codé en UTF-8. Le client de chiffrement C3R traite toutes les valeurs comme des chaînes.

Propriétés prises en charge dans les fichiers `.csv`

Le client de chiffrement C3R nécessite que les fichiers `.csv` possèdent les propriétés suivantes :

- Peut contenir ou non une ligne d'en-tête initiale qui nomme chaque colonne de manière unique.
- Délimité par des virgules. (Actuellement, les délimiteurs personnalisés ne sont pas pris en charge.)
- Texte codé en UTF-8.

Suppression des espaces blancs dans les entrées `.csv`

Les espaces blancs de début et de fin sont supprimés des entrées `.csv`.

NULLEncodage personnalisé pour un fichier `.csv`

Un fichier `.csv` peut utiliser un NULL encodage personnalisé.

Avec le client de chiffrement C3R, vous pouvez spécifier des codages personnalisés pour les NULL entrées dans les données d'entrée à l'aide de l'`--csvInputNULLValue=<csv-input-null>`indicateur. Le client de chiffrement C3R peut utiliser des encodages personnalisés dans le fichier de sortie généré pour les entrées NULL à l'aide de l'`--csvOutputNULLValue=<csv-output-null>`indicateur.

 Note

Une NULL entrée est considérée comme manquant de contenu, en particulier dans le contexte d'un format tabulaire plus riche tel qu'un tableau SQL. Bien que le format .csv ne prenne pas explicitement en charge cette caractérisation pour des raisons historiques, il est courant de considérer qu'une entrée vide contenant uniquement des espaces blancs est considérée NULL comme telle. Il s'agit donc du comportement par défaut du client de chiffrement C3R et il peut être personnalisé selon les besoins.

Comment les entrées .csv sont interprétées par C3R

Le tableau suivant fournit des exemples de la manière dont les entrées .csv sont rassemblées (cleartextcleartextpour plus de clarté) en fonction des valeurs (le cas échéant) fournies pour les indicateurs --csvInputNULLValue=<csv-input-null> et --csvOutputNULLValue=<csv-output-null>. Les espaces blancs de début et de fin situés en dehors des guillemets sont supprimés avant que C3R n'interprète le sens d'une valeur.

<csv-input-null>	<csv-output-null>	Entrée d'entrée	Entrée de sortie
Aucune	Aucun	,AnyProduct,	,AnyProduct,
Aucun	Aucun	, AnyProduct ,	,AnyProduct,
Aucun	Aucun	,"AnyProduct",	,AnyProduct,
Aucun	Aucun	, "AnyProdu ct" ,	,AnyProduct,
Aucun	Aucun	,,	,,
Aucun	Aucun	, ,	,,
Aucun	Aucun	, "",	,,
Aucun	Aucun	, " ",	, " ",
Aucun	Aucun	, " " ,	, " ",

<csv-input-null>	<csv-output-null>	Entrée d'entrée	Entrée de sortie
"AnyProduct"	"NULL"	,AnyProduct,	,NULL,
"AnyProduct"	"NULL"	, AnyProduct ,	,NULL,
"AnyProduct"	"NULL"	,"AnyProduct",	,NULL,
"AnyProduct"	"NULL"	, "AnyProdu ct" ,	,NULL,
Aucun	"NULL"	,,	,NULL,
Aucun	"NULL"	, ,	,NULL,
Aucun	"NULL"	, "",	,NULL,
Aucun	"NULL"	, " ",	, " ",
Aucune	"NULL"	, " " ,	, " ",
""	"NULL"	,,	,NULL,
""	"NULL"	, ,	,NULL,
""	"NULL"	, "",	, "",
""	"NULL"	, " ",	, " ",
""	"NULL"	, " " ,	, " ",
"\\\""	"NULL"	,,	,,
"\\\""	"NULL"	, ,	,,
"\\\""	"NULL"	, "",	,NULL,
"\\\""	"NULL"	, " ",	, " ",
"\\\""	"NULL"	, " " ,	, " ",

Fichier CSV sans en-têtes

Il n'est pas nécessaire que le fichier .csv source comporte des en-têtes dans la première ligne qui nomment chaque colonne de manière unique. Toutefois, un fichier .csv sans ligne d'en-tête nécessite un schéma de chiffrement positionnel. Le schéma de chiffrement positionnel est requis au lieu du schéma mappé classique utilisé à la fois pour les fichiers .csv avec une ligne d'en-tête et pour les fichiers. Parquet

Un schéma de chiffrement positionnel spécifie les colonnes de sortie par position plutôt que par nom. Un schéma de chiffrement mappé associe les noms des colonnes source aux noms des colonnes cibles. Pour plus d'informations, notamment une discussion détaillée et des exemples des deux formats de schéma, consultez [Schémas de tables cartographiées et positionnelles](#).

Parquetfichiers

Un fichier avec une .parquet extension est supposé être au Apache Parquet format.

Types de Parquet données pris en charge

Le client de chiffrement C3R peut traiter toutes les données non complexes (c'est-à-dire de type primitif) dans un Parquet fichier qui représente un type de données pris en charge par. AWS Clean Rooms

Toutefois, seules les colonnes de chaîne peuvent être utilisées pour les sealed colonnes.

Les types de données Parquet suivants sont pris en charge :

- Binarytype primitif avec les annotations logiques suivantes :
 - Aucun si le --parquetBinaryAsString est défini (type de STRING données)
 - Decimal(scale, precision)(type de DECIMAL données)
 - String(type de STRING données)
- Booleantype de données primitif sans annotation logique (type de BOOLEAN données)
- Doubletype de données primitif sans annotation logique (type de DOUBLE données)
- Fixed_Len_Binary_Arraytype primitif avec annotation Decimal(scale, precision) logique (type de DECIMAL données)
- Floattype de données primitif sans annotation logique (type de FLOAT données)
- Int32type primitif avec les annotations logiques suivantes :

- `Aucun` (type de INT données)
- `Date`(type de DATE données)
- `Decimal(scale, precision)`(type de DECIMAL données)
- `Int(16, true)`(type de SMALLINT données)
- `Int(32, true)`(type de INT données)
- `Int64` type de données primitif avec les annotations logiques suivantes :
 - `Aucun` (type de BIGINT données)
 - `Decimal(scale, precision)`(type de DECIMAL données)
 - `Int(64, true)`(type de BIGINT données)
 - `Timestamp(isUTCAdjusted, TimeUnit.MILLIS)`(type de TIMESTAMP données)
 - `Timestamp(isUTCAdjusted, TimeUnit.MICROS)`(type de TIMESTAMP données)
 - `Timestamp(isUTCAdjusted, TimeUnit.NANOS)`(type de TIMESTAMP données)

Chiffrement de valeurs autres que des chaînes

Actuellement, seules les valeurs de chaîne sont prises en charge pour les sealed colonnes.

Pour les fichiers .csv, le client de chiffrement C3R traite toutes les valeurs comme du texte codé en UTF-8 et ne tente pas de les interpréter différemment avant le chiffrement.

Pour les colonnes d'empreintes digitales, les types sont regroupés en classes d'équivalence. Une classe d'équivalence est un ensemble de types de données dont l'égalité peut être comparée sans ambiguïté via un type de données représentatif.

Les classes d'équivalence permettent d'attribuer des empreintes identiques à la même valeur sémantique, quelle que soit la représentation d'origine. Cependant, la même valeur dans deux classes d'équivalence ne produira pas la même colonne d'empreintes digitales.

Par exemple, la même empreinte digitale 42 sera attribuée à la INTEGRAL valeur, qu'il s'agisse à l'origine d'un SMALLINT, ou BIGINT. De plus, la INTEGRAL valeur ne 0 correspondra jamais à la BOOLEAN valeur FALSE (qui est représentée par la valeur 0).

Les classes d'équivalence suivantes et les types de AWS Clean Rooms données correspondants sont pris en charge par les colonnes d'empreintes digitales :

Classe d'équivalence	Type de AWS Clean Rooms données pris en charge
BOOLEAN	BOOLEAN
DATE	DATE
INTEGRAL	BIGINT, INT, SMALLINT
STRING	CHAR, STRING, VARCHAR

Noms de colonnes dans Cryptographic Computing pour Clean Rooms

Par défaut, les noms des colonnes sont importants dans Cryptographic Computing for Clean Rooms.

Si la valeur du paramètre Autoriser les colonnes avec JOIN des noms différents est fausse, les noms des colonnes sont utilisés lors du chiffrement des fingerprint colonnes. C'est pourquoi, par défaut, les collaborateurs doivent se coordonner à l'avance et utiliser les mêmes noms de colonnes cibles pour les données qui utiliseront JOIN des instructions dans les requêtes. Par défaut, les colonnes chiffrées JOIN avec des noms différents ne fonctionnent avec succès JOIN sur aucune valeur.

Si la valeur du paramètre Autoriser les colonnes avec JOIN des noms différents est vraie, les JOIN instructions entre colonnes chiffrées en tant que fingerprint colonnes aboutissent. Le chiffrement des données avec ce paramètre peut permettre une certaine inférence des cleartext valeurs. Par exemple, si une ligne possède la même valeur de code d'authentification de message basé sur le hachage (HMAC) à la fois dans la City colonne et dans la State colonne, la valeur peut être. New York

Normalisation des noms d'en-têtes de colonnes

Les noms des en-têtes de colonnes sont normalisés par le client de chiffrement C3R. Tous les espaces blancs de début et de fin sont supprimés et le nom de la colonne est mis en minuscules pour la sortie transformée.

La normalisation est appliquée avant tous les autres calculs, calculs ou autres opérations susceptibles d'être affectés par les noms de colonnes. Le fichier de sortie émis contient uniquement les noms normalisés.

Types de colonnes dans le calcul cryptographique pour Clean Rooms

Cette rubrique fournit des informations sur les types de colonnes dans Cryptographic Computing for Clean Rooms.

Rubriques

- [Fingerprintcolonnes](#)
- [Colonnes étanches](#)
- [Cleartextcolonnes](#)

Fingerprintcolonnes

Fingerprintles colonnes sont des colonnes protégées cryptographiquement pour être utilisées dans JOIN des instructions.

Les données des fingerprint colonnes ne peuvent pas être déchiffrées. Seules les données provenant de colonnes scellées peuvent être déchiffrées.

Fingerprintles colonnes ne doivent être utilisées que dans les clauses et fonctions SQL suivantes :

- JOIN (INNER, OUTER, LEFT, RIGHT, or FULL) par rapport aux autres fingerprint colonnes :
 - Si la valeur du `allowJoinsOnColumnsWithDifferentNames` paramètre est définie sur `false`, les deux fingerprint colonnes du JOIN doivent également porter le même nom.
- SELECT COUNT()
- SELECT COUNT(DISTINCT)
- GROUP BY (À utiliser uniquement si la collaboration a défini la valeur du `preserveNulls` paramètre sur `true`.)

Les requêtes qui enfreignent ces contraintes peuvent donner des résultats incorrects.

Colonnes étanches

Les colonnes scellées sont des colonnes protégées cryptographiquement pour être utilisées dans SELECT des instructions.

Les colonnes scellées ne doivent être utilisées que dans les clauses et fonctions SQL suivantes :

- SELECT

- `SELECT ... AS`
- `SELECT COUNT()`

 Note

`SELECT COUNT(DISTINCT )` n'est pas pris en charge.

Les requêtes qui enfreignent ces contraintes peuvent donner des résultats incorrects.

Remplissage des données pour une sealed colonne avant le chiffrement

Lorsque vous spécifiez qu'une colonne doit être une sealed colonne, C3R vous demande quel type de rembourrage choisir. Le remplissage des données avant le chiffrement est facultatif. Sans rembourrage (type de `padnone`), la longueur des données cryptées indique la taille du `cleartext`. Dans certaines circonstances, la taille du `cleartext` peut exposer le texte en clair. Avec le rembourrage (un `pad` de type `fixed` ou `max`), toutes les valeurs sont d'abord rembourrées à une taille commune, puis cryptées. Avec le rembourrage, la longueur des données cryptées ne fournit aucune information sur la longueur d'origine, si ce n'est le fait de donner une limite supérieure à sa taille.

Si vous souhaitez un remplissage pour une colonne et que la longueur maximale en octets des données de cette colonne est connue, utilisez le `fixed` rembourrage. Utilisez une `length` valeur au moins égale à la longueur en octets de la valeur la plus longue de cette colonne.

 Note

Une erreur se produit et le chiffrement échoue si une valeur est supérieure à la valeur `length`.

Si vous souhaitez un remplissage pour une colonne et que la longueur maximale en octets des données de cette colonne n'est pas connue, utilisez le `max` rembourrage. Ce mode de remplissage rembourre toutes les données à la longueur de la valeur la plus longue plus des `length` octets supplémentaires.

 Note

Vous souhaitez peut-être chiffrer les données par lots ou mettre régulièrement à jour vos tables avec de nouvelles données. Sachez que le `max` remplissage remplira les entrées à

la longueur (plus l'lengthoctet) de l'entrée en texte brut la plus longue d'un lot donné. Cela signifie que la longueur du texte chiffré peut varier d'un lot à l'autre. Par conséquent, si vous connaissez la longueur maximale en octets d'une colonne, vous devez utiliser à la fixed place de. max

Cleartextcolonnes

Cleartextles colonnes sont des colonnes qui ne sont pas protégées cryptographiquement pour être utilisées dans SELECT des instructions JOIN or.

Cleartextles colonnes peuvent être utilisées dans n'importe quelle partie de la requête SQL.

Paramètres de calcul cryptographique

Les paramètres de calcul cryptographique sont disponibles pour les collaborations utilisant Cryptographic Computing for Clean Rooms (C3R) lors de la [création](#) d'une collaboration. Vous pouvez créer une collaboration à l'aide de la AWS Clean Rooms console ou de l>CreateCollaborationAPI. Dans la console, vous pouvez définir des valeurs pour les paramètres dans Paramètres de calcul cryptographique après avoir activé l'option Support de calcul cryptographique. Pour plus d'informations, consultez les rubriques suivantes.

Rubriques

- [Paramètre Autoriser cleartext les colonnes](#)
- [Paramètre Autoriser les doublons](#)
- [Paramètre JOIN d'autorisation des colonnes avec des noms différents](#)
- [Paramètre de conservation NULL des valeurs](#)

Paramètre Autoriser cleartext les colonnes

Dans la console, vous pouvez définir le paramètre Autoriser les cleartext colonnes lors de la [création d'une collaboration](#) afin de spécifier si cleartext les données sont autorisées dans une table contenant des données chiffrées.

Le tableau suivant décrit les valeurs du paramètre Autoriser cleartext les colonnes.

Valeur de paramètre	Description
Non	Cleartextles colonnes ne sont pas autorisées dans la table cryptée. Toutes les données sont protégées par cryptographie.
Oui	Cleartextles colonnes sont autorisées dans la table cryptée. Cleartextles colonnes ne sont pas protégées par cryptographie et sont incluses en tant quecleartext. Vous devez prendre note de ce que les cleartext données de vos lignes peuvent révéler à propos des autres données du tableau. Pour fonctionner SUM ou AVG sur des colonnes spécifiques, les colonnes doivent être inséréescleartext.

À l'aide de l'opération `CreateCollaboration API`, pour le `dataEncryptionMetadata` paramètre, vous pouvez définir la valeur de `allowCleartext to true` ou `false`. Pour plus d'informations sur les opérations d'API, consultez la [référence des AWS Clean Rooms API](#).

Cleartextles colonnes correspondent aux colonnes classées selon le schéma spécifique cleartext à la table. Les données de ces colonnes ne sont pas cryptées et peuvent être utilisées de quelque manière que ce soit. Cleartextles colonnes peuvent être utiles si les données ne sont pas sensibles and/or si une plus grande flexibilité est requise par rapport à ce que permet une sealed colonne ou une fingerprint colonne cryptée.

Paramètre Autoriser les doublons

Dans la console, vous pouvez définir le paramètre Autoriser les doublons lors de la [création d'une collaboration](#) afin de spécifier si les colonnes chiffrées pour les JOIN requêtes peuvent contenir des NULL non-valeurs dupliquées.

Important

Les paramètres Autoriser les doublons, [Autoriser les colonnes portant JOIN des noms différents](#) et [Préserver NULL les valeurs](#) ont des effets distincts mais connexes.

Le tableau suivant décrit les valeurs du paramètre Autoriser les doublons.

Valeur de paramètre	Description
Non	Les valeurs répétées ne sont pas autorisées dans une fingerprint colonne. Toutes les valeurs d'une seule fingerprint colonne doivent être uniques.
Oui	Les valeurs répétées sont autorisées dans une fingerprint colonne. Si vous devez joindre des colonnes contenant des valeurs répétées, définissez cette valeur sur Oui. Lorsque cette option est définie sur Oui, les modèles de fréquence apparaissant dans les fingerprint colonnes de la table C3R ou des résultats peuvent impliquer des informations supplémentaires sur la structure des cleartext données.

À l'aide de l'opération `CreateCollaboration API`, pour le `dataEncryptionMetadata` paramètre, vous pouvez définir la valeur de `allowDuplicates to true` ou `false`. Pour plus d'informations sur les opérations d'API, consultez la [référence des AWS Clean Rooms API](#).

Par défaut, si des données chiffrées doivent être utilisées dans les JOIN requêtes, le client de chiffrement C3R exige que ces colonnes ne contiennent aucune valeur dupliquée. Cette exigence vise à renforcer la protection des données. Ce comportement permet de garantir que les modèles répétés dans les données ne sont pas observables. Toutefois, si vous souhaitez utiliser des données chiffrées dans des JOIN requêtes et que vous n'êtes pas préoccupé par les valeurs dupliquées, le paramètre Autoriser les doublons peut désactiver cette vérification conservatrice.

Paramètre JOIN d'autorisation des colonnes avec des noms différents

Dans la console, vous pouvez définir le paramètre Autoriser les colonnes JOIN portant des noms différents lors de la [création d'une collaboration](#) afin de spécifier si les JOIN instructions entre des colonnes portant des noms différents sont prises en charge.

Pour de plus amples informations, consultez [Normalisation des noms d'en-têtes de colonnes](#).

Le tableau suivant décrit les valeurs du paramètre Autoriser JOIN les colonnes portant des noms différents.

Valeur de paramètre	Description
Non	Les jointures de fingerprint colonnes portant des noms différents ne sont pas prises en charge. JOIN les instructions ne fournissent des résultats précis que pour les colonnes portant le même nom.  Important La valeur Non renforce la sécurité des informations mais oblige les participants à la collaboration à se mettre d'accord au préalable sur les noms des colonnes. Si deux colonnes portent des noms différents lorsqu'elles sont cryptées sous forme JOIN de fingerprint colonnes et que l'option Autoriser les colonnes portant des noms différents est définie sur Non, les JOIN instructions sur ces colonnes ne produisent aucun résultat. Cela est dû au fait qu'aucune valeur n'est partagée entre eux après le chiffrement.
Oui	Les jointures de fingerprint colonnes portant des noms différents sont prises en charge. Pour plus de flexibilité, les utilisateurs peuvent définir cette valeur sur Oui, ce qui autorise les JOIN instructions sur les colonnes quel que soit le nom de la colonne. S'il est défini sur Oui, le client de chiffrement C3R ne prend pas en compte le nom de colonne lors de la protection des fingerprint colonnes. Par conséquent, les valeurs communes aux différentes fingerprint colonnes sont observables dans le tableau C3R. Par exemple, si une ligne possède la même JOIN valeur chiffrée à la fois dans une City colonne et dans une State colonne, il peut être raisonnable de déduire que cette valeur est New York la même.

À l'aide de l'opération `CreateCollaboration` API, pour le `dataEncryptionMetadata` paramètre, vous pouvez définir la valeur de `allowJoinsOnColumnsWithDifferentNames` to `true` ou `false`. Pour plus d'informations sur les opérations d'API, consultez la [référence des AWS Clean Rooms API](#).

Par défaut, le chiffrement des fingerprint colonnes est affecté par le `targetHeader` paramètre pour cette colonne, défini dans [Étape 4 : générer un schéma de chiffrement pour un fichier tabulaire](#). Par conséquent, la même cleartext valeur possède des représentations cryptées différentes dans chaque fingerprint colonne pour laquelle elle est cryptée.

Ce paramètre peut être utile pour empêcher l'inférence de cleartext valeurs dans certains cas. Par exemple, le fait de voir la même valeur cryptée dans `City` des fingerprint colonnes `State` peut être utilisé pour déduire raisonnablement que la valeur est `New York`. Cependant, l'utilisation de ce paramètre nécessite une coordination supplémentaire à l'avance, de sorte que toutes les colonnes à joindre dans les requêtes portent des noms communs.

Vous pouvez utiliser le paramètre Autoriser JOIN les colonnes portant des noms différents pour assouplir cette restriction. Lorsque la valeur du paramètre est définie sur `Yes`, toutes les colonnes cryptées peuvent JOIN être utilisées ensemble, quel que soit leur nom.

Paramètre de conservation NULL des valeurs

Dans la console, vous pouvez définir le paramètre Conserver NULL les valeurs lors de la [création d'une collaboration](#) pour indiquer qu'aucune valeur n'est présente pour cette colonne.

Le tableau suivant décrit les valeurs du paramètre Conserver NULL les valeurs.

Valeur de paramètre	Description
Non	NULLles valeurs ne sont pas préservées. NULLles valeurs n'apparaissent pas comme NULL dans une table cryptée. NULLles valeurs apparaissent sous forme de valeurs aléatoires uniques dans une table C3R.
Oui	NULLles valeurs sont préservées. NULLles valeurs apparaissent comme NULL dans une table cryptée. Si vous avez besoin d'une sémantique SQL pour les NULL valeurs, vous pouvez définir cette valeur sur Oui. Par conséquent, NULL les entrées apparaissent comme NULL dans la table C3R, que la colonne

Valeur de paramètre	Description
	soit cryptée ou non et quel que soit le paramètre défini pour Autoriser les doublons.

À l'aide de l'opération `CreateCollaboration` API, pour le `dataEncryptionMetadata` paramètre, vous pouvez définir la valeur de `preserveNulls` to `true` ou `false`. Pour plus d'informations sur les opérations d'API, consultez la [référence des AWS Clean Rooms API](#).

Lorsque le paramètre Conserver NULL les valeurs est défini sur Non pour la collaboration :

1. NULLles entrées dans `cleartext` les colonnes restent inchangées.
2. NULLles entrées dans les `fingerprint` colonnes cryptées sont cryptées sous forme de valeurs aléatoires afin de masquer leur contenu. Le fait de rejoindre une colonne cryptée avec des NULL entrées dans la `cleartext` colonne ne produit aucune correspondance pour aucune des NULL entrées. Aucune correspondance n'est établie car ils reçoivent chacun leur propre contenu aléatoire unique.
3. NULLles entrées dans les `sealed` colonnes cryptées sont cryptées.

Lorsque la valeur du paramètre Conserver les NULL valeurs est définie sur Oui pour la collaboration, les NULL entrées de toutes les colonnes restent inchangées, que la colonne NULL soit cryptée ou non.

Le paramètre Conserver NULL les valeurs est utile dans des scénarios tels que l'enrichissement des données, dans lesquels vous souhaitez partager un manque d'informations exprimé sous la forme NULL. Le paramètre Conserver NULL les valeurs est également utile au `fingerprint` format HMAC si vous avez des NULL valeurs dans la colonne que vous souhaitez JOIN ou GROUP BY.

Si la valeur des paramètres Autoriser les doublons et Préserver NULL les valeurs est définie sur Non, la présence de plusieurs NULL entrées dans une `fingerprint` colonne génère une erreur et arrête le chiffrement. Si la valeur de l'un des paramètres est définie sur Oui, aucune erreur de ce type ne se produit.

Drapeaux facultatifs dans le calcul cryptographique pour Clean Rooms

Les sections suivantes décrivent les indicateurs facultatifs que vous pouvez définir lorsque vous [cryptez des données](#) à l'aide du client de chiffrement C3R pour personnaliser et tester des fichiers tabulaires.

Rubriques

- [Indicateur --csvInputNULLValue](#)
- [Indicateur --csvOutputNULLValue](#)
- [Indicateur --enableStackTraces](#)
- [Indicateur --dryRun](#)
- [Indicateur --tempDir](#)

Indicateur --csvInputNULLValue

Vous pouvez utiliser l'--csvInputNULLValueindicateur pour spécifier des codages personnalisés pour les NULL entrées dans les données d'entrée lorsque vous [cryptez des données à l'aide du client](#) de chiffrement C3R.

Le tableau suivant récapitule l'utilisation et les paramètres de cet indicateur.

Usage	Parameters
Facultatif. Les utilisateurs peuvent spécifier des codages personnalisés pour les NULL entrées dans les données d'entrée.	Encodage des NULL valeurs défini par l'utilisateur dans le fichier CSV d'entrée

Une NULL entrée est une entrée considérée comme manquant de contenu, en particulier dans le contexte d'un format tabulaire plus riche tel qu'un tableau SQL. Bien que le format .csv ne prenne pas explicitement en charge cette caractérisation pour des raisons historiques, il est courant de considérer qu'une entrée vide contenant uniquement des espaces blancs l'est. NULL Il s'agit donc du comportement par défaut du client de chiffrement C3R et il peut être personnalisé selon les besoins.

Indicateur --csvOutputNULLValue

Vous pouvez utiliser l'--csvOutputNULLValueindicateur pour spécifier des codages personnalisés pour les NULL entrées dans les données de sortie lorsque vous [cryptez des données à l'aide du client](#) de chiffrement C3R.

Le tableau suivant récapitule l'utilisation et les paramètres de cet indicateur.

Usage	Parameters
Facultatif. Les utilisateurs peuvent spécifier des codages personnalisés dans le fichier de sortie généré pour les NULL entrées.	Encodage des NULL valeurs défini par l'utilisateur dans le fichier CSV de sortie

Une NULL entrée est une entrée considérée comme manquant de contenu, en particulier dans le contexte d'un format tabulaire plus riche tel qu'un tableau SQL. Bien que le format .csv ne prenne pas explicitement en charge cette caractérisation pour des raisons historiques, il est courant de considérer qu'une entrée vide contenant uniquement des espaces blancs l'est. NULL Il s'agit donc du comportement par défaut du client de chiffrement C3R et il peut être personnalisé selon les besoins.

Indicateur **--enableStackTraces**

Lorsque vous [chiffrez des données](#) à l'aide du client de chiffrement C3R, utilisez l'--enableStackTracesindicateur pour fournir des informations contextuelles supplémentaires afin de signaler les erreurs lorsque C3R rencontre une erreur.

AWS ne collecte pas les erreurs. Si vous rencontrez une erreur, utilisez le stack trace pour résoudre vous-même l'erreur ou envoyez le stack trace à Support pour obtenir de l'aide.

Le tableau suivant récapitule l'utilisation et les paramètres de cet indicateur.

Usage	Parameters
Facultatif. Utilisé pour fournir des informations contextuelles supplémentaires afin de signaler les erreurs lorsque le client de chiffrement C3R rencontre une erreur.	Aucune

Indicateur **--dryRun**

[Les commandes du client de chiffrement C3R crypter et déchiffrer incluent un indicateur facultatif.](#)

--dryRun L'indicateur prend tous les arguments fournis par l'utilisateur et vérifie leur validité et leur cohérence.

Vous pouvez utiliser l'option `--dryRun` pour vérifier si votre fichier de schéma est valide et cohérent avec le fichier d'entrée correspondant.

Le tableau suivant récapitule l'utilisation et les paramètres de cet indicateur.

Usage	Parameters
Facultatif. Demande au client de chiffrement C3R d'analyser les paramètres et de vérifier les fichiers, mais n'effectue aucun chiffrement ni déchiffrement.	Aucune

Indicateur `--tempDir`

Vous pouvez utiliser un répertoire temporaire car les fichiers chiffrés peuvent parfois être plus volumineux que les fichiers non chiffrés, en fonction de leurs paramètres. Les ensembles de données doivent également être chiffrés par collaboration pour fonctionner correctement.

Lorsque vous [chiffrez des données](#) à l'aide de C3R, utilisez l'option `--tempDir` pour spécifier l'emplacement où les fichiers temporaires peuvent être créés lors du traitement de l'entrée.

Le tableau suivant récapitule l'utilisation et les paramètres de cet indicateur.

Usage	Parameters
Les utilisateurs peuvent spécifier l'emplacement où les fichiers temporaires peuvent être créés lors du traitement de l'entrée.	Par défaut, c'est le répertoire temporaire du système.

Requêtes avec informatique cryptographique pour Clean Rooms

Cette rubrique fournit des informations sur l'écriture de requêtes utilisant des tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms.

Rubriques

- [Requêtes qui se rattachent à NULL](#)
- [Mappage d'une colonne source vers plusieurs colonnes cibles](#)

- [Utiliser les mêmes données pour JOIN les deux SELECT requêtes](#)

Requêtes qui se rattachent à NULL

Avoir une branche de requête sur une NULL instruction signifie utiliser une syntaxe similaire à `IF x IS NULL THEN 0 ELSE 1`.

Les requêtes peuvent toujours se baser sur NULL des instructions en cleartext colonnes.

Les requêtes peuvent se baser sur NULL des instructions en sealed fingerprint colonnes et en colonnes uniquement lorsque la valeur du paramètre Conserver les valeurs NULL (`preserveNulls`) est définie sur `true`.

Les requêtes qui enfreignent ces contraintes peuvent donner des résultats incorrects.

Mappage d'une colonne source vers plusieurs colonnes cibles

Une colonne source peut être mappée à plusieurs colonnes cibles. Par exemple, vous pouvez vouloir les deux JOIN et SELECT sur une colonne.

Pour de plus amples informations, veuillez consulter [Utiliser les mêmes données pour JOIN les deux SELECT requêtes](#).

Utiliser les mêmes données pour JOIN les deux SELECT requêtes

Si les données d'une colonne ne sont pas sensibles, elles peuvent apparaître dans une colonne cleartext cible, ce qui permet de les utiliser à n'importe quelle fin.

Si les données d'une colonne sont sensibles et doivent être utilisées à la fois pour les SELECT requêtes JOIN et, mappez cette colonne source à deux colonnes cibles dans le fichier de sortie. Une colonne est chiffrée avec la colonne `type` sous forme de fingerprint colonne, et une colonne est chiffrée avec la colonne `type` sous forme de colonne scellée. La génération de schéma interactive du client de chiffrement C3R suggère des suffixes d'en-tête de `et. _fingerprint _sealed` Ces suffixes d'en-tête peuvent constituer une convention utile pour différencier rapidement de telles colonnes.

Directives pour le client de chiffrement C3R

Le client de chiffrement C3R est un outil qui permet aux entreprises de rassembler des données sensibles afin de tirer de nouvelles informations de l'analyse des données. L'outil limite

cryptographiquement ce qui peut être appris par n'importe quelle partie et AWS au cours du processus. Bien que cela soit d'une importance vitale, le processus de sécurisation cryptographique des données peut entraîner une surcharge importante en termes de ressources de calcul et de stockage. Il est donc important de comprendre les inconvénients liés à l'utilisation de chaque paramètre et de savoir comment optimiser les paramètres tout en maintenant les garanties cryptographiques souhaitées. Cette rubrique se concentre sur les implications en termes de performances des différents paramètres du client et des schémas de chiffrement C3R.

Tous les paramètres de chiffrement du client de chiffrement C3R fournissent des garanties cryptographiques différentes. Les paramètres de collaboration sont les plus sécurisés par défaut. L'activation de fonctionnalités supplémentaires lors de la création d'une collaboration affaiblit les garanties de confidentialité, ce qui permet d'effectuer des activités telles que l'analyse des fréquences sur le texte chiffré. Pour plus d'informations sur la manière dont ces paramètres sont utilisés et sur leurs implications, consultez [the section called "Informatique cryptographique"](#).

Rubriques

- [Implications sur les performances pour les types de colonnes](#)
- [Résolution des problèmes liés aux augmentations imprévues de la taille du texte chiffré](#)

Implications sur les performances pour les types de colonnes

C3R utilise trois types de colonnes : cleartextfingerprint, etsealed. Chacun de ces types de colonnes fournit des garanties cryptographiques différentes et a des utilisations prévues différentes. Dans les sections suivantes, les implications du type de colonne sur les performances sont abordées ainsi que l'impact de chaque paramètre sur les performances.

Rubriques

- [Cleartextcolonnes](#)
- [Fingerprintcolonnes](#)
- [Sealedcolonnes](#)

Cleartextcolonnes

Cleartextles colonnes ne sont pas modifiées par rapport à leur format d'origine et ne sont en aucun cas traitées cryptographiquement. Ce type de colonne ne peut pas être configuré et n'a aucune incidence sur les performances de stockage ou de calcul.

Fingerprint colonnes

Fingerprintles colonnes sont destinées à être utilisées pour joindre des données sur plusieurs tables. À cette fin, la taille du texte chiffré obtenu doit toujours être la même. Toutefois, ces colonnes sont affectées par les paramètres de collaboration. Fingerprintles colonnes peuvent avoir différents degrés d'impact sur la taille du fichier de sortie en fonction du cleartext contenu de l'entrée.

Rubriques

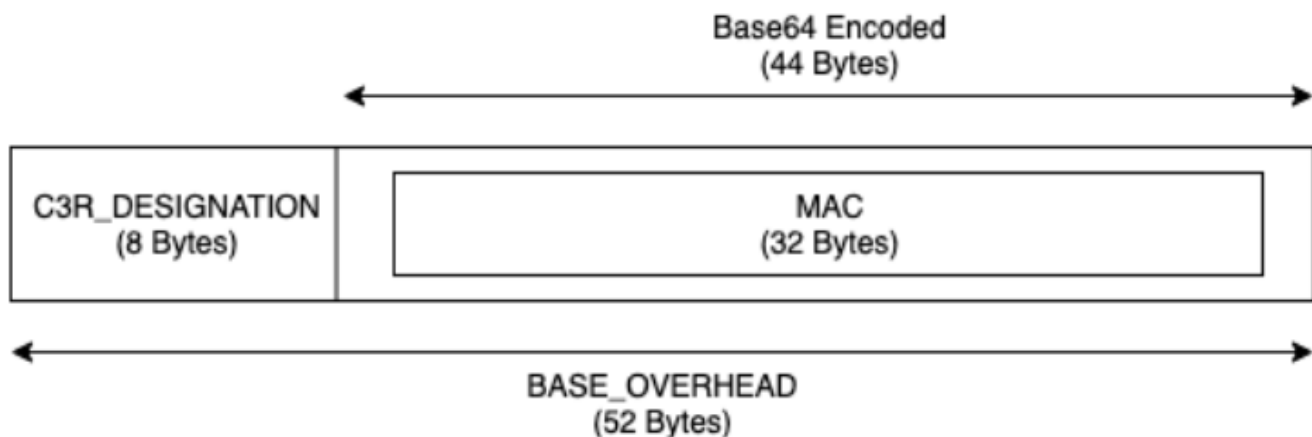
- [Frais généraux de base pour les fingerprint colonnes](#)
- [Paramètres de collaboration pour les fingerprint colonnes](#)
- [Exemple de données pour une fingerprint colonne](#)
- [fingerprintColonnes de dépannage](#)

Frais généraux de base pour les fingerprint colonnes

Il existe un surcoût de base pour les fingerprint colonnes. Cette surcharge est constante et remplace la taille des cleartext octets.

Les données des fingerprint colonnes sont traitées cryptographiquement par le biais d'une fonction HMAC (Hash Based Message Authentication Code), qui transforme les données en un code d'authentification de message (MAC) de 32 octets. Ces données sont ensuite traitées via un encodeur base64, ce qui ajoute environ 33 % à la taille des octets. Il est précédé d'une désignation C3R à 8 octets pour désigner le type de colonne à laquelle appartiennent les données et la version du client qui les a produites. Le résultat final est de 52 octets. Ce résultat est ensuite multiplié par le nombre de lignes pour obtenir le surcoût total de base (utilisez le nombre total de null valeurs non égales s'il `preserveNulls` est défini sur `true`).

L'image suivante montre comment $BASE\_OVERHEAD = C3R\_DESIGNATION + (MAC * 1.33)$



Le texte chiffré de sortie dans les fingerprint colonnes sera toujours de 52 octets. Cela peut entraîner une diminution significative de la capacité de stockage si les cleartext données d'entrée dépassent en moyenne plus de 52 octets (par exemple, les adresses postales complètes). Cela peut représenter une augmentation significative du stockage si les cleartext données d'entrée sont en moyenne inférieures à 52 octets (par exemple, l'âge du client).

Paramètres de collaboration pour les fingerprint colonnes

Paramètre **preserveNulls**

Lorsque le paramètre au niveau de la collaboration `preserveNulls` est `false` (par défaut), chaque `null` valeur est remplacée par 32 octets uniques et aléatoires et traitée comme si ce n'était pas le cas. `null` Le résultat est que chaque `null` valeur est désormais de 52 octets. Cela peut ajouter des exigences de stockage importantes pour les tables contenant très peu de données par rapport à ce paramètre `true` et à ce que `null` les valeurs sont transmises sous forme `null` de.

Si vous n'avez pas besoin des garanties de confidentialité de ce paramètre et que vous préférez conserver `null` les valeurs de vos ensembles de données, activez le `preserveNulls` paramètre au moment de la création de la collaboration. Le `preserveNulls` paramètre ne peut pas être modifié une fois la collaboration créée.

Exemple de données pour une fingerprint colonne

Voici un exemple de jeu de données d'entrée et de sortie pour une fingerprint colonne avec des paramètres à reproduire. D'autres paramètres de collaboration n'ont `allowDuplicates` pas d'impact sur les résultats et peuvent être définis au fur `true` et à mesure que `allowCleartext` `false` vous essayez de reproduire localement.

Exemple de secret partagé : `wJa1rXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY`

Exemple d'identifiant de collaboration : a1b2c3d4-5678-90ab-cdef-EXAMPLE11111

`allowJoinsOnColumnsWithDifferentNames`: `True` ce paramètre n'a aucune incidence sur les performances ou les exigences de stockage. Toutefois, ce paramètre rend le choix du nom de colonne non pertinent lors de la reproduction des valeurs indiquées dans les tableaux suivants.

Exemple 1

Input	null
<code>preserveNulls</code>	TRUE
Output	null
Déterministe	Yes
Octets d'entrée	0
Octets de sortie	0

Exemple 2

Input	null
<code>preserveNulls</code>	FALSE
Output	01: hmac:31kFjthvV3IUu6mMvFc1a +XAHwgw/E1m0q4p3Yg25kk=
Déterministe	No
Octets d'entrée	0
Octets de sortie	52

Exemple 3

Input	empty string
<code>preserveNulls</code>	-

Output	01:hmac:oKTgi3Gba+eUb3JteSz 2EMgXUkF1WgM77UP0Ydw5kPQ=
Déterministe	Yes
Octets d'entrée	0
Octets de sortie	52

Exemple 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:hmac:kU/IqwG7FMmzzshr0B9 scomE0UJUEE7j9keTctp1Gww=
Déterministe	Yes
Octets d'entrée	26
Octets de sortie	52

Exemple 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:hmac:ks3htnQbw2vdhCRFF6J NzW5LMndJaHG57uvE26mBtSs=
Déterministe	Yes
Octets d'entrée	62

Octets de sortie

52

fingerprintColonnes de débannage

Pourquoi le texte chiffré de mes fingerprint colonnes est-il plusieurs fois supérieur à la taille du texte cleartext qui y est inscrit ?

Le texte chiffré d'une fingerprint colonne a toujours une longueur de 52 octets. Si vos données d'entrée étaient petites (par exemple, l'âge des clients), elles indiqueront une augmentation significative de leur taille. Cela peut également se produire si le `preserveNulls` paramètre est réglé sur `false`.

Pourquoi le texte chiffré de mes fingerprint colonnes est-il plusieurs fois plus petit que la taille du texte cleartext qui y figure ?

Le texte chiffré d'une fingerprint colonne a toujours une longueur de 52 octets. Si vos données d'entrée sont volumineuses (par exemple, les adresses complètes des clients), leur taille diminuera considérablement.

Comment savoir si j'ai besoin des garanties cryptographiques fournies par **`preserveNulls`** ?

Malheureusement, la réponse est que cela dépend. Au minimum, [the section called "Parameters"](#) il convient de vérifier la manière dont le `preserveNulls` paramètre protège vos données. Cependant, nous vous recommandons de faire référence aux exigences de traitement des données de votre organisation et à tout contrat applicable à la collaboration correspondante.

Pourquoi dois-je supporter la surcharge de base64 ?

Pour garantir la compatibilité avec les formats de fichiers tabulaires tels que CSV, le codage base64 est nécessaire. Bien que certains formats de fichier Parquet puissent prendre en charge les représentations binaires des données, il est important que tous les participants à une collaboration représentent les données de la même manière afin de garantir des résultats de requête corrects.

Sealedcolonnes

Sealedles colonnes sont destinées à être utilisées pour transférer des données entre les membres d'une collaboration. Le texte chiffré de ces colonnes n'est pas déterministe et a un impact significatif sur les performances et le stockage en fonction de la configuration des colonnes. Ces colonnes

peuvent être configurées individuellement et ont souvent le plus grand impact sur les performances du client de chiffrement C3R et sur la taille du fichier de sortie qui en résulte.

Rubriques

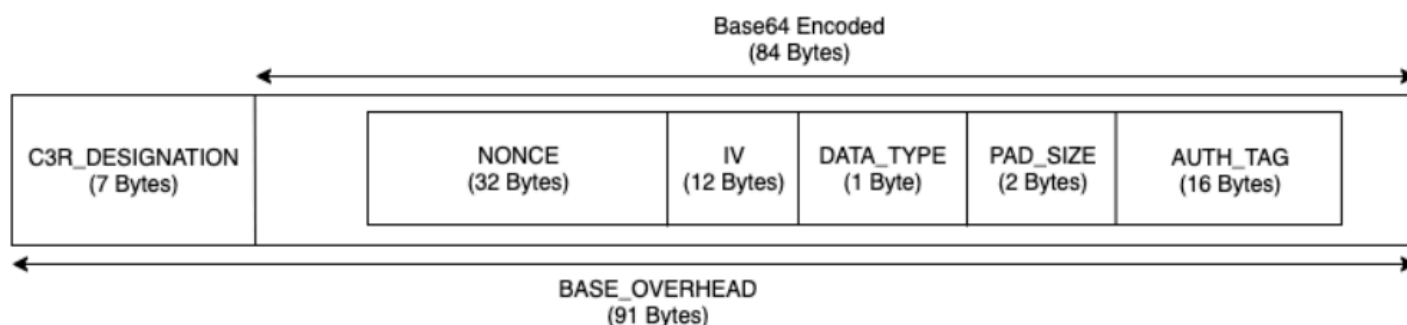
- [Frais généraux de base pour les sealed colonnes](#)
- [Paramètres de collaboration pour les sealed colonnes](#)
- [sealedColonnes des paramètres du schéma : types de rembourrage](#)
- [Exemple de données pour une sealed colonne](#)
- [sealedColonnes de dépannage](#)

Frais généraux de base pour les sealed colonnes

Il existe un surcoût de base pour les sealed colonnes. Cette surcharge est constante et s'ajoute à la taille des octets cleartext et de remplissage (le cas échéant).

Avant tout chiffrement, les données des sealed colonnes sont précédées d'un caractère de 1 octet désignant le type de données contenues. Si le rembourrage est sélectionné, les données sont ensuite complétées et ajoutées avec 2 octets indiquant la taille du pad. Une fois ces octets ajoutés, les données sont traitées cryptographiquement à l'aide d'AES-GCM et stockées avec les IV (12 octets), nonce (32 octets) et Auth Tag (16 octets). Ces données sont ensuite traitées via un encodeur base64, ce qui ajoute environ 33 % à la taille des octets. Les données sont précédées d'une désignation C3R à 7 octets pour indiquer le type de colonne auquel elles appartiennent et la version du client utilisée pour les produire. Le résultat est un surdébit de base final de 91 octets. Ce résultat peut ensuite être multiplié par le nombre de lignes pour obtenir le surcoût total de base (utilisez le nombre total de valeurs non nulles s'il `preserveNulls` est défini sur `true`).

L'image suivante montre comment $BASE\_OVERHEAD = C3R\_DESIGNATION + ((NONCE + IV + DATA\_TYPE + PAD\_SIZE + AUTH\_TAG) * 1.33)$



Paramètres de collaboration pour les sealed colonnes

Paramètre **preserveNulls**

Lorsque le paramètre au niveau de la collaboration `preserveNulls` est `false` (par défaut), chaque `null` valeur est unique, aléatoire de 32 octets et traitée comme si ce n'était pas le cas. `null` Le résultat est que chaque `null` valeur est désormais de 91 octets (plus si elle est complétée). Cela peut ajouter des exigences de stockage importantes pour les tables contenant très peu de données par rapport à ce paramètre `true` et à ce que `null` les valeurs sont transmises sous forme `null` de.

Si vous n'avez pas besoin des garanties de confidentialité de ce paramètre et que vous préférez conserver `null` les valeurs de vos ensembles de données, activez le `preserveNulls` paramètre au moment de la création de la collaboration. Le `preserveNulls` paramètre ne peut pas être modifié une fois la collaboration créée.

sealedColonnes des paramètres du schéma : types de rembourrage

Rubriques

- [Type de pad none](#)
- [Type de pad fixed](#)
- [Type de pad max](#)

Type de pad **none**

La sélection d'un type de pad `none` n'ajoute aucun rembourrage à la surcharge de base décrite précédemment `cleartext` et n'ajoute aucune surcharge supplémentaire à la surcharge de base décrite précédemment. L'absence de rembourrage permet d'obtenir la taille de sortie la plus économe en espace. Cependant, il n'offre pas les mêmes garanties de confidentialité que les types de rembourrage `fixed` et de `max` rembourrage. Cela est dû au fait que la taille du sous-jacent `cleartext` est perceptible à partir de la taille du texte chiffré.

Type de pad **fixed**

La sélection d'un type de pad `fixed` est une mesure de protection de la vie privée qui permet de masquer la longueur des données contenues dans une colonne. Cela se fait en complétant le tout dans le champ `cleartext` fourni `pad_length` avant qu'il ne soit crypté. Toute donnée dépassant cette taille entraîne l'échec du client de chiffrement C3R.

Étant donné que le remplissage est ajouté cleartext avant d'être chiffré, AES-GCM dispose d'un mappage 1 à 1 entre les octets du texte chiffré. Le codage base64 ajoutera 33 %. La surcharge de stockage supplémentaire du rembourrage peut être calculée en soustrayant la longueur moyenne du de la valeur cleartext du `pad_length` et en la multipliant par 1,33. Le résultat est le surcoût moyen de remplissage par enregistrement. Ce résultat peut ensuite être multiplié par le nombre de lignes pour obtenir la surcharge de remplissage totale (utilisez le nombre total de `null` valeurs non égales si la valeur `preserveNulls` est définie sur `true`).

$$PADDING\_OVERHEAD = (PAD\_LENGTH - AVG\_CLEARTEXT\_LENGTH) * 1.33 * ROW\_COUNT$$

Nous vous recommandons de sélectionner le minimum `pad_length` qui englobe la plus grande valeur d'une colonne. Par exemple, si la valeur la plus élevée est de 50 octets, une valeur `pad_length` de 50 est suffisante. Une valeur supérieure à cette valeur ne fera qu'augmenter la charge de stockage.

Le rembourrage fixe n'entraîne aucune surcharge de calcul significative.

Type de pad **max**

La sélection d'un type de pad **max** est une mesure de protection de la vie privée qui permet de masquer la longueur des données contenues dans une colonne. Cela se fait en ajoutant le tout cleartext à la plus grande valeur de la colonne, plus le montant supplémentaire, `pad_length` avant qu'il ne soit chiffré. En général, **max** le remplissage fournit les mêmes garanties que le remplissage **fixed** d'un seul ensemble de données, tout en permettant de ne pas connaître la plus grande cleartext valeur de la colonne. Cependant, le **max** remplissage peut ne pas fournir les mêmes garanties de confidentialité que le **fixed** remplissage entre les mises à jour, car la valeur la plus élevée des ensembles de données individuels peut être différente.

Nous vous recommandons de sélectionner une valeur supplémentaire `pad_length` de 0 lorsque vous utilisez le **max** rembourrage. Cette longueur permet à toutes les valeurs d'avoir la même taille que la plus grande valeur de la colonne. Une valeur supérieure à cette valeur ne fera qu'augmenter la charge de stockage.

Si la cleartext valeur la plus élevée est connue pour une colonne donnée, nous vous recommandons d'utiliser plutôt le type **fixed** pad. L'utilisation **fixed** du rembourrage assure la cohérence entre les ensembles de données mis à jour. L'utilisation du remplissage permet de compléter chaque sous-ensemble de données à la valeur la plus élevée figurant dans le sous-ensemble.

Exemple de données pour une sealed colonne

Voici un exemple de jeu de données d'entrée et de sortie pour une sealed colonne avec des paramètres à reproduire. D'autres paramètres de collaboration tels que `allowClearTextAllowJoinsOnColumnsWithDifferentNames`, et `allowDuplicates` n'ont pas d'impact sur les résultats et peuvent être définis au fur et à mesure que `false` vous essayez de reproduire localement. Bien qu'il s'agisse des paramètres de base à reproduire, la sealed colonne n'est pas déterministe et les valeurs changent à chaque fois. L'objectif est d'afficher les octets entrants par rapport aux octets sortants. Les `pad_length` valeurs d'exemple ont été choisies intentionnellement. Ils montrent que le `fixed` rembourrage donne les mêmes valeurs que le `max` rembourrage avec les `pad_length` paramètres minimaux recommandés ou lorsqu'un rembourrage supplémentaire est souhaité.

Exemple de secret partagé : `wJa1rXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY`

Exemple d'identifiant de collaboration : `a1b2c3d4-5678-90ab-cdef-EXAMPLE11111`

Rubriques

- [Type de pad none](#)
- [Type de tampon fixed \(exemple 1\)](#)
- [Type de tampon fixed \(exemple 2\)](#)
- [Type de tampon max \(exemple 1\)](#)
- [Type de tampon max \(exemple 2\)](#)

Type de pad **none**

Exemple 1

Input	<code>null</code>
<code>preserveNulls</code>	<code>TRUE</code>
Output	<code>null</code>
Déterministe	<code>Yes</code>
Octets d'entrée	<code>0</code>
Octets de sortie	<code>0</code>

Exemple 2

Input	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfssGSPbNIJfG3iXmu 6cbCUrizuV
Déterministe	No
Octets d'entrée	0
Octets de sortie	91

Exemple 3

Input	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfstGSPeM6qR8DWC2P B2GM1X41YK
Déterministe	No
Octets d'entrée	0
Octets de sortie	91

Exemple 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-

Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6pkx9sGL5 VLDQeHzh6DmPpyWNuI=
Déterministe	No
Octets d'entrée	26
Octets de sortie	127

Exemple 5

Input	abcdefghijklmnopqrstuvwxyza BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ Q0Q3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
Déterministe	No
Octets d'entrée	62
Octets de sortie	175

Type de tampon **fixed** (exemple 1)

Dans cet exemple, `pad_length` c'est 62 et la plus grande entrée est 62 octets.

Exemple 1

Input	null
preserveNulls	TRUE
Output	null
Déterministe	Yes
Octets d'entrée	0
Octets de sortie	0

Exemple 2

Input	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmlmDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcoNpATs0GzbnLkor4L+/aSuA=
Déterministe	No
Octets d'entrée	0
Octets de sortie	175

Exemple 3

Input	empty string
preserveNulls	-

Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsicoLB53l07VZp A60wkuXu29CA=
Déterministe	No
Octets d'entrée	0
Octets de sortie	175

Exemple 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6pkx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsicutBAc0+Mb9t uU2KIIHH31AWg=
Déterministe	No
Octets d'entrée	26
Octets de sortie	175

Exemple 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ Q0Q3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
Déterministe	No
Octets d'entrée	62
Octets de sortie	175

Type de tampon **fixed** (exemple 2)

Dans cet exemple, `pad_length` c'est 162 et la plus grande entrée est 62 octets.

Exemple 1

Input	null
preserveNulls	TRUE
Output	null
Déterministe	Yes
Octets d'entrée	0
Octets de sortie	0

Exemple 2

Input	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcgl1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv/xAySX+xcntotL703aBTBb
Déterministe	No
Octets d'entrée	0
Octets de sortie	307

Exemple 3

Input	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcgl1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp

	pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv84lVaT9Yd+6oQx65/+gdVT
Déterministe	No
Octets d'entrée	0
Octets de sortie	307

Exemple 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6pkx9jy48 Fcgl1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwX5Hn1+Wyf06ks3QMaRDGSf
Déterministe	No
Octets d'entrée	26
Octets de sortie	307

Exemple 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/Q0Q3cXb/pbvPcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrnwjkJXQZ0gPdeFX9Yr/8a1V5i
Déterministe	No
Octets d'entrée	62
Octets de sortie	307

Type de tampon **max** (exemple 1)

Dans cet exemple, la valeur `pad_length` est 0 et la plus grande entrée est de 62 octets.

Exemple 1

Input	null
preserveNulls	TRUE
Output	null
Déterministe	Yes

Octets d'entrée	0
Octets de sortie	0

Exemple 2

Input	null
preserveNulls	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTL EZb/hCz7oaIneVsrcoNpATs0GzbnLkor4L+/aSuA=
Déterministe	No
Octets d'entrée	0
Octets de sortie	175

Exemple 3

Input	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTL EZb/hCz7oaIneVsrcoLB53l07VZpA60wkuXu29CA=

Déterministe	No
Octets d'entrée	0
Octets de sortie	175

Exemple 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6pkx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcutBAC0+Mb9tuU2KIIHH31AWg=
Déterministe	No
Octets d'entrée	26
Octets de sortie	175

Exemple 5

Input	abcdefghijklmnopqrstuvwxyzABCDEF GHIJKLMNOPQRSTUVWXYZ0123456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6plwtH/8t

	RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ Q0Q3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
Déterministe	No
Octets d'entrée	62
Octets de sortie	175

Type de tampon **max** (exemple 2)

Dans cet exemple, `pad_length` c'est 100 et la plus grande entrée est 62 octets.

Exemple 1

Input	null
<code>preserveNulls</code>	TRUE
Output	null
Déterministe	Yes
Octets d'entrée	0
Octets de sortie	0

Exemple 2

Input	null
<code>preserveNulls</code>	FALSE
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z

	NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv/xAySX+xcntotL703aBTBb
Déterministe	No
Octets d'entrée	0
Octets de sortie	307

Exemple 3

Input	empty string
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbMlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcgl1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv841VaT9Yd+6oQx65/+gdVT
Déterministe	No
Octets d'entrée	0
Octets de sortie	307

Exemple 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6pkx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwtX5Hn1+Wyf06ks3QMaRDGSf
Déterministe	No
Octets d'entrée	26
Octets de sortie	307

Exemple 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Output	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/Q0Q3cXb/pbvPcnkB0xbLWD7z

```
NdAqQGR0rXoSESdW0I0vpNoGcBf
v4cJbG0A3h1DvtkSSVc2B8000Gp
pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn
+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6
uwrmwjkJXQZ0gPdeFX9Yr/8a1V5i
```

Déterministe	No
Octets d'entrée	62
Octets de sortie	307

sealedColonnes de dépannage

Pourquoi le texte chiffré de mes sealed colonnes est-il plusieurs fois supérieur à la taille du texte cleartext qui y est inscrit ?

Cela dépend de plusieurs facteurs. D'une part, le texte chiffré d'une Cleartext colonne a toujours une longueur d'au moins 91 octets. Si vos données d'entrée étaient petites (par exemple, l'âge des clients), elles indiqueront une augmentation significative de leur taille. Ensuite, si vous avez `preserveNulls` défini ce paramètre sur `false` et que vos données d'entrée contenaient un grand nombre de `null` valeurs, chacune de ces `null` valeurs aura été transformée en 91 octets de texte chiffré. Enfin, si vous utilisez le rembourrage, par définition, des octets sont ajoutés aux cleartext données avant qu'elles ne soient chiffrées.

La plupart de mes données dans une sealed colonne sont très petites et je dois utiliser le rembourrage. Puis-je simplement supprimer les grandes valeurs et les traiter séparément pour économiser de l'espace ?

Nous vous déconseillons de supprimer des valeurs importantes et de les traiter séparément. Cela modifie les garanties de confidentialité fournies par le client de chiffrement C3R. En tant que modèle de menace, supposons qu'un observateur puisse voir les deux ensembles de données chiffrés. Si l'observateur constate qu'une colonne d'un sous-ensemble de données est plus ou moins remplie qu'un autre sous-ensemble, il peut tirer des conclusions sur la taille des données de chaque sous-ensemble. Supposons, par exemple, qu'une `fullName` colonne soit complétée à un total de 40 octets dans un fichier et à 800 octets dans un autre fichier. Un observateur peut supposer qu'un ensemble de données contient le nom le plus long du monde (747 octets).

Dois-je fournir un rembourrage supplémentaire lorsque j'utilise ce type de **max** rembourrage ?

Non. Lorsque vous utilisez le max remboursement, nous recommandons que `lepad_length`, également connu sous le nom de remboursement supplémentaire au-delà de la plus grande valeur de la colonne, soit défini sur 0.

Puis-je simplement en choisir une grande **pad_length** lorsque j'utilise un **fixed** remboursement pour ne pas me demander si la plus grande valeur convient ?

Oui, mais la grande longueur du pad est inefficace et utilise plus d'espace de stockage que nécessaire. Nous vous recommandons de vérifier la taille de la plus grande valeur et de la `pad_length` définir sur cette valeur.

Comment savoir si j'ai besoin des garanties cryptographiques fournies par **preserveNulls** ?

Malheureusement, la réponse est que cela dépend. Au minimum, [Informatique cryptographique pour Clean Rooms](#) il convient de vérifier la manière dont le `preserveNulls` paramètre protège vos données. Cependant, nous vous recommandons de faire référence aux exigences de traitement des données de votre organisation et à tout contrat applicable à la collaboration correspondante.

Pourquoi dois-je supporter la surcharge de base64 ?

Pour garantir la compatibilité avec les formats de fichiers tabulaires tels que CSV, le codage base64 est nécessaire. Bien que certains formats de fichier Parquet puissent prendre en charge les représentations binaires des données, il est important que tous les participants à une collaboration représentent les données de la même manière afin de garantir des résultats de requête corrects.

Résolution des problèmes liés aux augmentations imprévues de la taille du texte chiffré

Supposons que vous ayez chiffré vos données et que la taille des données obtenues soit étonnamment importante. Les étapes suivantes peuvent vous aider à identifier l'endroit où l'augmentation de taille s'est produite et les mesures que vous pouvez prendre, le cas échéant.

Identification de l'endroit où l'augmentation de taille s'est produite

Avant de pouvoir déterminer pourquoi vos données chiffrées sont nettement plus volumineuses que vos cleartext données, vous devez d'abord identifier l'origine de l'augmentation de taille. Cleartextles colonnes peuvent être ignorées en toute sécurité car elles sont inchangées. Examinez le reste fingerprint des sealed colonnes et choisissez-en une qui semble significative.

Identifier la raison pour laquelle l'augmentation de taille s'est produite

Une fingerprint colonne ou une sealed colonne peut contribuer à l'augmentation de la taille.

Rubriques

- [L'augmentation de taille provient-elle d'une fingerprint colonne ?](#)
- [L'augmentation de taille provient-elle d'une sealed colonne ?](#)

L'augmentation de taille provient-elle d'une fingerprint colonne ?

Si la colonne qui contribue le plus à l'augmentation du stockage est une fingerprint colonne, cela est probablement dû au fait que les cleartext données sont petites (par exemple, l'âge du client). Chaque fingerprint texte chiffré obtenu a une longueur de 52 octets. Malheureusement, rien ne peut être fait à ce sujet sur une column-by-column base solide. Pour plus d'informations, voir [Frais généraux de base pour les fingerprint colonnes](#) les détails de cette colonne, notamment son impact sur les exigences de stockage.

L'autre cause possible de l'augmentation de la taille d'une fingerprint colonne est le paramètre de collaboration, `preserveNulls`. Si le paramètre de collaboration pour `preserveNulls` est désactivé (paramètre par défaut), toutes les null valeurs des fingerprint colonnes seront devenues 52 octets de texte chiffré. Rien ne peut être fait pour cela dans le cadre de la collaboration actuelle. Le `preserveNulls` paramètre est défini au moment de la création d'une collaboration et tous les collaborateurs doivent utiliser le même paramètre pour garantir des résultats de requête corrects. Pour plus d'informations sur ce `preserveNulls` paramètre et sur l'impact de son activation sur les garanties de confidentialité de vos données, consultez [the section called "Informatique cryptographique"](#).

L'augmentation de taille provient-elle d'une sealed colonne ?

Si la colonne qui contribue le plus à l'augmentation du stockage est une sealed colonne, certains détails peuvent contribuer à l'augmentation de la taille.

Si les cleartext données sont petites (par exemple, l'âge du client), chaque sealed texte chiffré obtenu a une longueur d'au moins 91 octets. Malheureusement, rien ne peut être fait à ce sujet. Pour plus d'informations, voir [Frais généraux de base pour les sealed colonnes](#) les détails de cette colonne, notamment son impact sur les exigences de stockage.

La deuxième cause principale de l'augmentation du stockage dans les sealed colonnes est le rembourrage. Le remplissage ajoute des octets supplémentaires cleartext avant le chiffrement afin de masquer la taille des valeurs individuelles d'un ensemble de données. Nous vous recommandons de définir le rembourrage à la valeur minimale possible pour votre ensemble de données. Au minimum, le `pad_length fixed` remplissage doit être défini de manière à inclure la plus grande valeur

possible dans la colonne. Tout paramètre supérieur n'ajoute aucune garantie de confidentialité supplémentaire. Par exemple, si vous savez que la plus grande valeur possible d'une colonne peut être de 50 octets, nous vous recommandons de la `pad_length` définir sur 50 octets. Toutefois, si la sealed colonne utilise un max remplissage, nous vous recommandons de définir la `pad_length` valeur sur 0 octet. Cela est dû au fait que le max rembourrage fait référence au rembourrage supplémentaire au-delà de la plus grande valeur de la colonne.

La dernière cause possible de l'augmentation de la taille d'une sealed colonne est le paramètre de collaboration, `preserveNulls`. Si le paramètre de collaboration pour `preserveNulls` est désactivé (paramètre par défaut), toutes les `null` valeurs des sealed colonnes seront devenues 91 octets de texte chiffré. Rien ne peut être fait pour cela dans le cadre de la collaboration actuelle. Le `preserveNulls` paramètre est défini au moment de la création d'une collaboration, et tous les collaborateurs doivent utiliser le même paramètre pour garantir des résultats de requête corrects. Pour plus d'informations sur ce paramètre et sur l'impact de son activation sur les garanties de confidentialité de vos données, consultez [the section called “Informatique cryptographique”](#).

Con AWS Clean Rooms figuration

Les rubriques suivantes expliquent comment procéder à la configuration AWS Clean Rooms.

Rubriques

- [Inscrivez-vous pour AWS](#)
- [Configurer les rôles de service pour AWS Clean Rooms](#)
- [Configuration des rôles de service pour AWS Clean Rooms ML](#)

Inscrivez-vous pour AWS

Avant de pouvoir utiliser AWS Clean Rooms, ou tout autre Service AWS, vous devez vous inscrire AWS avec 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.

Au cours de la procédure d'inscription, vous recevrez un appel téléphonique contenant un code de vérification que vous saisirez sur le clavier du téléphone.

3. Lorsque vous vous inscrivez à un Compte AWS, un utilisateur Compte AWS root 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 administratif](#), et à utiliser uniquement l'utilisateur racine pour effectuer les [tâches nécessitant un accès utilisateur racine](#).

Configurer les rôles de service pour AWS Clean Rooms

Les sections suivantes décrivent les rôles nécessaires à l'exécution de chaque tâche.

Rubriques

- [Création d'un utilisateur administrateur](#)

- [Création d'un rôle IAM pour un membre de la collaboration](#)
- [Créez un rôle de service pour lire les données d'Amazon S3](#)
- [Création d'un rôle de service pour lire les données d'Amazon Athena](#)
- [Créez un rôle de service pour lire les données de Snowflake](#)
- [Création d'un rôle de service pour lire le code d'un compartiment S3 \(rôle de modèle d'PySpark analyse\)](#)
- [Création d'un rôle de service pour écrire les résultats d'une PySpark tâche](#)
- [Créez un rôle de service pour recevoir des résultats](#)

Création d'un utilisateur administrateur

Pour l'utiliser AWS Clean Rooms, vous devez créer un utilisateur administrateur pour vous-même et l'ajouter à un groupe d'administrateurs.

Afin de créer un utilisateur administrateur, choisissez l'une des options suivantes :

Choisissez un moyen de gérer votre administrateur	À	En	Vous pouvez également
Dans IAM Identity Center (Recommandé)	Utiliser des informations d'identification à court terme pour accéder à AWS. C'est conforme aux bonnes pratiques en matière de sécurité. Pour plus d'informations sur les bonnes pratiques, consultez Bonnes pratiques de	Suivant les instructions fournies dans Mise en route dans le Guide de l'utilisateur AWS IAM Identity Center .	Configurez l'accès par programmation en configurant le AWS CLI à utiliser AWS IAM Identity Center dans le guide de l'AWS Command Line Interface utilisateur.

Choisissez un moyen de gérer votre administrateur	À	En	Vous pouvez également
	sécurité dans IAM dans le Guide de l'utilisateur IAM.		
Dans IAM (Non recommandé)	Utiliser les informations d'identification à long terme pour accéder à AWS.	Suivant les instructions fournies dans Création d'un utilisateur IAM pour l'accès d'urgence dans le Guide de l'utilisateur IAM.	Configurer l'accès par programmation en suivant les instructions fournies dans Gestion des clés d'accès pour les utilisateurs IAM dans le Guide de l'utilisateur IAM.

Création d'un rôle IAM pour un membre de la collaboration

Un membre est un AWS client participant à une collaboration.

Pour créer un rôle IAM pour un membre de la collaboration

1. Suivez la procédure [Création d'un rôle pour déléguer des autorisations à un utilisateur IAM](#) dans le Guide de la Gestion des identités et des accès AWS utilisateur.
2. Pour l'étape Créer une politique, sélectionnez l'onglet JSON dans l'éditeur de politiques, puis ajoutez des politiques en fonction des capacités accordées au membre de la collaboration.

AWS Clean Rooms propose les politiques gérées suivantes basées sur des cas d'utilisation courants.

Si vous voulez...	Ensuite, utilisez...
Afficher les ressources et les métadonnées	AWS politique gérée : AWSCleanRoomsReadOnlyAccess

Si vous voulez...	Ensuite, utilisez...
Query	AWS politique gérée : AWSCleanRoomsFullAccess
Interroger et exécuter des tâches	AWS politique gérée : AWSCleanRoomsFullAccess
Interrogez et recevez des résultats	AWS politique gérée : AWSCleanRoomsFullAccess
Gérez les ressources de collaboration sans poser de questions	AWS politique gérée : AWSCleanRoomsFullAccessNoQuerying

Pour plus d'informations sur les différentes politiques gérées proposées par AWS Clean Rooms [AWS politiques gérées pour AWS Clean Rooms](#), voir

Créez un rôle de service pour lire les données d'Amazon S3

AWS Clean Rooms utilise un rôle de service pour lire les données d'Amazon S3.

Il existe deux manières de créer ce rôle de service.

- Si vous disposez des autorisations IAM nécessaires pour créer un rôle de service, utilisez la AWS Clean Rooms console pour créer un rôle de service.
- Si vous ne disposez pas d'iam:CreateRole ou iam:CreatePolicy d'iam:AttachRolePolicy autorisations ou si vous souhaitez créer les rôles IAM manuellement, effectuez l'une des opérations suivantes :
 - Utilisez la procédure suivante pour créer un rôle de service à l'aide de politiques de confiance personnalisées.
 - Demandez à votre administrateur de créer le rôle de service en suivant la procédure suivante.

Note

Vous ou votre administrateur IAM devez suivre cette procédure uniquement si vous ne disposez pas des autorisations nécessaires pour créer un rôle de service à l'aide de la AWS Clean Rooms console.

Pour créer un rôle de service permettant de lire les données d'Amazon S3 à l'aide de politiques de confiance personnalisées

1. Créez un rôle à l'aide de politiques de confiance personnalisées. Pour plus d'informations, consultez la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) dans le guide de Gestion des identités et des accès AWS l'utilisateur.
2. Utilisez la politique de confiance personnalisée suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

Note

Si vous souhaitez garantir que le rôle n'est utilisé que dans le contexte d'une certaine adhésion à une collaboration, vous pouvez affiner davantage la politique de confiance. Pour de plus amples informations, veuillez consulter [Prévention du problème de l'adjoint confus entre services](#).

JSON

```
{  
  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Sid": "RoleTrustPolicyForCleanRoomsService",  
      "Effect": "Allow",  
      "Principal": {  
        "Service": "cleanrooms.amazonaws.com"  
      },  
      "Action": "sts:AssumeRole"  
    }  
  ]  
}
```

}

3. Utilisez la politique d'autorisation suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

 Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire AWS Glue les métadonnées et les données Amazon S3 correspondantes. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données Amazon S3. Par exemple, si vous avez configuré une clé KMS personnalisée pour vos données Amazon S3, vous devrez peut-être modifier cette politique avec des autorisations supplémentaires AWS Key Management Service (AWS KMS).

Vos AWS Glue ressources et les ressources Amazon S3 sous-jacentes doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "NecessaryGluePermissions",
      "Effect": "Allow",
      "Action": [
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartition",
        "glue:GetPartitions",
        "glue:BatchGetPartition"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:database/databaseName",
        "arn:aws:glue:us-east-1:111122223333:table/databaseName/tableName",
        "arn:aws:glue:us-east-1:111122223333:catalog"
      ]
    }
  ]
}
```



```

    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "glue:GetSchema",
      "glue:GetSchemaVersion"
    ],
    "Resource": [
      "*"
    ]
  },
  {
    "Sid": "NecessaryS3BucketPermissions",
    "Effect": "Allow",
    "Action": [
      "s3:GetBucketLocation",
      "s3:ListBucket"
    ],
    "Resource": [
      "arn:aws:s3:::bucket"
    ],
    "Condition": {
      "StringEquals": {
        "s3:ResourceAccount": [
          "444455556666"
        ]
      }
    }
  },
  {
    "Sid": "NecessaryS3ObjectPermissions",
    "Effect": "Allow",
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucket/prefix/*"
    ],
    "Condition": {
      "StringEquals": {
        "s3:ResourceAccount": [
          "444455556666"
        ]
      }
    }
  }
}

```

```

    }
  }
}

```

Note

Cette politique fait référence à deux règles différentes Compte AWS IDs pour soutenir une AWS Clean Rooms collaboration dans laquelle les métadonnées du catalogue de données et le stockage réel des données sont gérés par différentes parties :

- 111122223333 - Il s'agit du compte propriétaire des ressources du catalogue de AWS Glue données (bases de données, tables et catalogue). La première instruction accorde des autorisations pour accéder aux schémas de table, aux informations de partition et aux métadonnées du AWS Glue catalogue de ce compte.
- 444455556666 - Il s'agit du compte propriétaire du compartiment Amazon S3 contenant les fichiers de données réels. Les autorisations Amazon S3 (instructions 3 et 4) sont limitées aux compartiments appartenant à ce compte en vertu de cette `s3:ResourceAccount` condition.

Cette configuration prend en charge les architectures de données d'entreprise courantes dans lesquelles une équipe gère le catalogue de données et les définitions de schéma tandis qu'une autre équipe possède l'infrastructure de stockage de données sous-jacente. Cette `s3:ResourceAccount` condition fournit une couche de sécurité supplémentaire en garantissant que les opérations Amazon S3 ne fonctionnent que sur les buckets appartenant au compte désigné.

4. Remplacez chaque *placeholder* par vos propres informations.
5. Continuez à suivre la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) pour créer le rôle.

Création d'un rôle de service pour lire les données d'Amazon Athena

AWS Clean Rooms utilise un rôle de service pour lire les données d'Amazon Athena.

Pour créer un rôle de service permettant de lire les données d'Athena à l'aide de politiques de confiance personnalisées

1. Créez un rôle à l'aide de politiques de confiance personnalisées. Pour plus d'informations, consultez la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) dans le guide de Gestion des identités et des accès AWS l'utilisateur.
2. Utilisez la politique de confiance personnalisée suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

 Note

Si vous souhaitez garantir que le rôle n'est utilisé que dans le contexte d'une certaine adhésion à une collaboration, vous pouvez affiner davantage la politique de confiance. Pour de plus amples informations, veuillez consulter [Prévention du problème de l'adjoint confus entre services](#).

JSON

```
{  
  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Sid": "RoleTrustPolicyForCleanRoomsService",  
      "Effect": "Allow",  
      "Principal": {  
        "Service": "cleanrooms.amazonaws.com"  
      },  
      "Action": "sts:AssumeRole"  
    }  
  ]  
}
```

3. Utilisez la politique d'autorisation suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire AWS Glue les métadonnées et les données Athena correspondantes. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données Amazon S3. Par exemple, si vous avez déjà configuré une clé KMS personnalisée pour vos données Amazon S3, vous devrez peut-être modifier cette politique avec des AWS KMS autorisations supplémentaires.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "athena:GetWorkGroup",
        "athena:GetTableMetadata",
        "athena:GetDataCatalog",
        "athena:StartQueryExecution",
        "athena:GetQueryExecution",
        "athena:GetQueryResults"
      ],
      "Resource": [
        "arn:aws:athena:region:accountId:workgroup/workgroup",
        "arn:aws:athena:region:accountId:datacatalog/federatedCatalogName"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetDatabase",
        "glue:GetTable",
        "glue:GetCatalog"
      ],
      "Resource": [
        "arn:aws:glue:region:accountId:catalog",
        "arn:aws:glue:region:accountId:catalog/federatedCatalogName",
        "arn:aws:glue:region:accountId:database/federatedCatalogName/databaseName",

```

```

    "arn:aws:glue:region:accountId:table/federatedCatalogName/databaseName/tableName"
  ],
  {
    "Effect": "Allow",
    "Action": [
      "s3:GetObject",
      "s3:GetBucketLocation",
      "s3:AbortMultipartUpload",
      "s3:ListBucket",
      "s3:PutObject",
      "s3:ListMultipartUploadParts"
    ],
    "Resource": [
      "arn:aws:s3::athenaResultsBucket",
      "arn:aws:s3::athenaResultsBucket/*"
    ],
    "Condition": {
      "StringEquals": {
        "aws:ResourceAccount": "accountId"
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": "lakeformation:GetDataAccess",
    "Resource": "*"
  }
]
}

```

4. Remplacez chaque *placeholder* par vos propres informations.
5. Continuez à suivre la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) pour créer le rôle.

Configurer les autorisations de Lake Formation

Si vous interrogez des ressources protégées par des autorisations Lake Formation, le rôle de service doit disposer des autorisations d'accès Select and Describe table/view/catalog et Describe sur la AWS Glue base de données.

Pour en savoir plus, consultez :

- [Utilisez Athena pour interroger les données enregistrées AWS Lake Formation dans le guide de l'utilisateur d'Amazon Athena](#)
- [Intégration aux autorisations de Lake Formation](#) dans le guide du AWS Lake Formation développeur

Créez un rôle de service pour lire les données de Snowflake

AWS Clean Rooms utilise un rôle de service pour récupérer vos informations d'identification afin que Snowflake puisse lire vos données à partir de cette source.

Il existe deux manières de créer ce rôle de service :

- Si vous disposez des autorisations IAM nécessaires pour créer un rôle de service, utilisez la AWS Clean Rooms console pour créer un rôle de service.
- Si vous ne disposez pas des autorisations `iam:CreateRole` ou `iam:CreatePolicy` ou si vous souhaitez créer les rôles IAM manuellement, effectuez l'une des opérations suivantes :
 - Utilisez la procédure suivante pour créer un rôle de service à l'aide de politiques de confiance personnalisées.
 - Demandez à votre administrateur de créer le rôle de service en suivant la procédure suivante.

Note

Vous ou votre administrateur IAM devez suivre cette procédure uniquement si vous ne disposez pas des autorisations nécessaires pour créer un rôle de service à l'aide de la AWS Clean Rooms console.

Pour créer un rôle de service permettant de lire les données de Snowflake à l'aide de politiques de confiance personnalisées

1. Créez un rôle à l'aide de politiques de confiance personnalisées. Pour plus d'informations, consultez la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) dans le guide de Gestion des identités et des accès AWS l'utilisateur.

2. Utilisez la politique de confiance personnalisée suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

 Note

Si vous souhaitez garantir que le rôle n'est utilisé que dans le contexte d'une certaine adhésion à une collaboration, vous pouvez affiner davantage la politique de confiance. Pour de plus amples informations, veuillez consulter [Prévention du problème de l'adjoint confus entre services](#).

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowIfSourceArnMatches",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/membershipId",
            "arn:aws:cleanrooms:us-east-1:444455556666:membership/queryRunnerMembershipId"
          ]
        }
      }
    }
  ]
}
```

Note

Cette politique de confiance fait référence Compte AWS IDs à deux règles différentes pour soutenir une AWS Clean Rooms collaboration dans laquelle les responsabilités d'exécution des requêtes sont réparties entre plusieurs parties :

- 111122223333 - Il s'agit du compte qui contient un membre participant à la collaboration. Cette adhésion peut posséder des tables de données, des règles d'analyse ou d'autres ressources de collaboration nécessitant un accès aux rôles.
- 444455556666 - Il s'agit du compte qui contient le membre responsable de l'exécution des requêtes (le « lanceur de requêtes »). Cette adhésion exécute des requêtes protégées et doit assumer ce rôle pour accéder aux ressources de calcul et de données nécessaires.

Cette configuration permet des scénarios dans lesquels une partie fournit des données ou des modèles d'analyse tandis qu'une autre partie exécute les requêtes réelles. Les deux rôles nécessitent des autorisations différentes mais complémentaires via le même rôle d'exécution. Cette `aws:SourceArn` condition garantit que seules les AWS Clean Rooms opérations issues de ces deux adhésions spécifiques peuvent assumer le rôle, préservant ainsi la sécurité tout en soutenant l'exécution distribuée des tâches et le flux de travail de gestion des résultats.

3. Utilisez l'une des politiques d'autorisation suivantes conformément à la procédure [Création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

Politique d'autorisation pour les secrets chiffrés à l'aide d'une clé KMS appartenant au client

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "arn:aws:secretsmanager:us-east-1:111122223333:secret:secretIdentifier",
    }
  ]
}
```



```

        "Effect": "Allow"
      },
      {
        "Sid": "AllowDecryptViaSecretsManagerForKey",
        "Action": "kms:Decrypt",
        "Resource": "arn:aws:kms:us-  
east-1:444455556666:key/keyIdentifier",
        "Effect": "Allow",
        "Condition": {
          "StringEquals": {
            "kms:ViaService": "secretsmanager.us-  
east-1.amazonaws.com",
            "kms:EncryptionContext:SecretARN":  
"arn:aws:secretsmanager:us-east-1:111122223333:secret:secretIdentifier"
          }
        }
      }
    ]
  }
}

```

Note

Cette politique fait référence à deux règles différentes Compte AWS IDs pour prendre en charge un scénario de gestion des secrets entre comptes :

- 111122223333 - Il s'agit du compte qui possède et stocke le secret. La première instruction autorise la récupération de la valeur secrète de ce compte.
- 444455556666 - Il s'agit du compte qui possède la AWS KMS clé utilisée pour chiffrer le secret. La deuxième instruction autorise le déchiffrement du secret à l'aide de la AWS KMS clé de ce compte.

Cette configuration est courante dans les environnements d'entreprise où :

- Les secrets sont gérés de manière centralisée dans un seul compte (compte 1)
- Les clés de chiffrement sont gérées par un compte de sécurité ou de services partagés distinct (compte 2)
- La politique AWS KMS clé du compte 2 doit également autoriser le service du compte 1 à utiliser la clé pour les encryption/decryption opérations

Cette `kms:EncryptionContext:SecretARN` condition garantit que la AWS KMS clé ne peut être utilisée que pour déchiffrer ce secret spécifique, fournissant ainsi une couche de sécurité supplémentaire pour l'accès entre comptes.

Politique d'autorisation pour les secrets chiffrés avec un Clé gérée par AWS

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "arn:aws:secretsmanager:us-  
east-1:111122223333:secret:secretIdentifier",
      "Effect": "Allow"
    }
  ]
}
```

4. Remplacez chaque *placeholder* par vos propres informations.
5. Continuez à suivre la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) pour créer le rôle.

Création d'un rôle de service pour lire le code d'un compartiment S3 (rôle de modèle d'PySpark analyse)

AWS Clean Rooms utilise un rôle de service pour lire le code du compartiment S3 spécifié par un membre de la collaboration lors de l'utilisation d'un modèle d'PySpark analyse.

Pour créer un rôle de service permettant de lire le code d'un compartiment S3

1. Créez un rôle à l'aide de politiques de confiance personnalisées. Pour plus d'informations, consultez la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) dans le guide de Gestion des identités et des accès AWS l'utilisateur.

2. Utilisez la politique de confiance personnalisée suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/jobRunnerMembershipId",
            "arn:aws:cleanrooms:us-east-1:444455556666:membership/analysisTemplateOwnerMembershipId"
          ]
        }
      }
    }
  ]
}
```

Note

Cette politique de confiance fait référence à deux options différentes Compte AWS IDs pour soutenir un scénario de AWS Clean Rooms collaboration multipartite :

- 111122223333 - Il s'agit du compte qui contient le membre responsable de l'exécution des requêtes (le « job runner »). Cette adhésion exécute les tâches d'analyse et doit assumer ce rôle pour accéder aux ressources nécessaires.
- 444455556666 - Il s'agit du compte propriétaire du modèle d'analyse et de ses membres associés (le « propriétaire du modèle d'analyse »). Cette appartenance

définit les requêtes qui peuvent être exécutées et doit également assumer ce rôle pour gérer et exécuter l'analyse.

Cette configuration est typique des AWS Clean Rooms collaborations où plusieurs parties participent à la même collaboration, chacune ayant ses propres Compte AWS membres. L'exécuteur de requêtes et le propriétaire du modèle d'analyse doivent tous deux accéder à des ressources partagées. Cette `aws:SourceArn` condition garantit que seules les AWS Clean Rooms opérations provenant de ces deux adhésions spécifiques peuvent assumer le rôle, fournissant ainsi un contrôle d'accès précis pour la collaboration multipartite.

3. Utilisez la politique d'autorisation suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire votre code depuis Amazon S3. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données S3. Vos ressources Amazon S3 doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:GetObjectVersion"
      ],
      "Resource": ["arn:aws:s3:::s3Path"],
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
```

```

    }
    ]
  }
}
"s3BucketOwnerAccountId"

```

4. Remplacez chacune *placeholder* par vos propres informations :
 - *s3Path*— L'emplacement de votre code dans le compartiment S3.
 - *s3BucketOwnerAccountId*— L' Compte AWS ID du propriétaire du compartiment S3.
 - *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
 - *jobRunnerAccountId*— L' Compte AWS ID du membre autorisé à exécuter des requêtes et à exécuter des tâches.
 - *jobRunnerMembershipId*— L'ID de membre du membre qui peut interroger et exécuter des tâches. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration. Cela garantit qu'il AWS Clean Rooms n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
 - *analysisTemplateAccountId*— L' Compte AWS ID du modèle d'analyse.
 - *analysisTemplateOwnerMembershipId*— L'ID de membre du membre propriétaire du modèle d'analyse. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration.
5. Continuez à suivre la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) pour créer le rôle.

Création d'un rôle de service pour écrire les résultats d'une PySpark tâche

AWS Clean Rooms utilise un rôle de service pour écrire les résultats d'une PySpark tâche dans un compartiment S3 spécifié.

Pour créer un rôle de service afin d'écrire les résultats d'une PySpark tâche

1. Créez un rôle à l'aide de politiques de confiance personnalisées. Pour plus d'informations, consultez la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) dans le guide de Gestion des identités et des accès AWS l'utilisateur.
2. Utilisez la politique de confiance personnalisée suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/jobRunnerMembershipId",
            "arn:aws:cleanrooms:us-east-1:444455556666:membership/rrMembershipId"
          ]
        }
      }
    }
  ]
}
```

 Note

Cette politique de confiance fait référence Compte AWS IDs à deux règles différentes pour soutenir une AWS Clean Rooms collaboration avec des rôles opérationnels distincts :

- 111122223333 - Il s'agit du compte qui contient l'adhésion responsable de l'exécution des tâches d'analyse (le « job runner »). Cette adhésion exécute les charges de travail de calcul et doit assumer ce rôle pour accéder aux ressources de traitement.
- 444455556666 - Il s'agit du compte qui contient l'adhésion avec les responsabilités du destinataire des résultats (RR). Cet abonnement est autorisé à recevoir et à accéder aux résultats des tâches d'analyse, et a besoin d'un accès aux rôles pour écrire les résultats dans des emplacements désignés.

Cette configuration permet AWS Clean Rooms des scénarios dans lesquels une partie exécute l'analyse informatique tandis qu'une autre partie reçoit et gère les résultats. Les deux rôles nécessitent des autorisations différentes mais complémentaires via le même rôle d'exécution. Cette `aws:SourceArn` condition garantit que seules les AWS Clean Rooms opérations issues de ces deux adhésions spécifiques peuvent assumer le rôle, préservant ainsi la sécurité tout en soutenant l'exécution distribuée des tâches et le flux de travail de gestion des résultats.

3. Utilisez la politique d'autorisation suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

 Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour écrire sur Amazon S3. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré S3.

Vos ressources Amazon S3 doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject"
      ],
      "Resource": "arn:aws:s3:::bucket/optionalPrefix/*",
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "s3BucketOwnerAccountId"
          ]
        }
      }
    }
  ]
}
```

```

    },
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket"
      ],
      "Resource": "arn:aws:s3:::bucket",
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "s3BucketOwnerAccountId"
          ]
        }
      }
    }
  ]
}

```

4. Remplacez chacune *placeholder* par vos propres informations :

- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *jobRunnerAccountId*— L' Compte AWS ID dans lequel se trouve le compartiment S3.
- *jobRunnerMembershipId*— L'ID de membre du membre qui peut interroger et exécuter des tâches. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration. Cela garantit qu'il AWS Clean Rooms n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
- *rrAccountId*— L' Compte AWS ID dans lequel se trouve le compartiment S3.
- *rrMembershipId*— Le numéro de membre du membre qui peut recevoir les résultats. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration. Cela garantit qu'il AWS Clean Rooms n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
- *bucket*— Le nom et l'emplacement du compartiment S3.
- *optionalPrefix*— Un préfixe facultatif si vous souhaitez enregistrer vos résultats sous un préfixe S3 spécifique.
- *s3BucketOwnerAccountId*— L' Compte AWS ID du propriétaire du compartiment S3.

5. Continuez à suivre la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) pour créer le rôle.

Créez un rôle de service pour recevoir des résultats

Note

Si vous êtes le membre qui ne peut recevoir que des résultats (dans la console, les capacités de votre membre sont uniquement de recevoir des résultats), suivez cette procédure.

Si vous êtes un membre capable à la fois d'interroger et de recevoir des résultats (dans la console, vos capacités de membre sont à la fois Query et Receive des résultats), vous pouvez ignorer cette procédure.

Pour les membres de la collaboration qui ne peuvent recevoir que des résultats, AWS Clean Rooms utilise un rôle de service pour écrire les résultats des données demandées dans la collaboration dans le compartiment S3 spécifié.

Il existe deux manières de créer ce rôle de service :

- Si vous disposez des autorisations IAM nécessaires pour créer un rôle de service, utilisez la AWS Clean Rooms console pour créer un rôle de service.
- Si vous ne disposez pas `iam:CreateRole` ou `iam:CreatePolicy` d'`iam:AttachRolePolicy` autorisations ou si vous souhaitez créer les rôles IAM manuellement, effectuez l'une des opérations suivantes :
 - Utilisez la procédure suivante pour créer un rôle de service à l'aide de politiques de confiance personnalisées.
 - Demandez à votre administrateur de créer le rôle de service en suivant la procédure suivante.

Note

Vous ou votre administrateur IAM devez suivre cette procédure uniquement si vous ne disposez pas des autorisations nécessaires pour créer un rôle de service à l'aide de la AWS Clean Rooms console.

Pour créer un rôle de service afin de recevoir des résultats à l'aide de politiques de confiance personnalisées

1. Créez un rôle à l'aide de politiques de confiance personnalisées. Pour plus d'informations, consultez la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) dans le guide de Gestion des identités et des accès AWS l'utilisateur.
2. Utilisez la politique de confiance personnalisée suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowIfExternalIdMatches",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ArnLike": {
          "sts:ExternalId":
            "arn:aws:*:region:*:dbuser:*/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaa*"
        }
      }
    },
    {
      "Sid": "AllowIfSourceArnMatches",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "ForAnyValue:ArnEquals": {
          "aws:SourceArn": [
            "arn:aws:cleanrooms:us-east-1:555555555555:membership/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaa"
          ]
        }
      }
    }
  ]
}
```

```

    }
  }
}

```

3. Utilisez la politique d'autorisation suivante conformément à la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#).

Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire AWS Glue les métadonnées et les données Amazon S3 correspondantes. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données S3.

Vos AWS Glue ressources et les ressources Amazon S3 sous-jacentes doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::bucket_name"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceAccount": "accountId"
        }
      }
    }
  ],
}

```

```

    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject"
      ],
      "Resource": [
        "arn:aws:s3:::bucket_name/optional_key_prefix/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceAccount": "accountId"
        }
      }
    }
  ]
}

```

4. Remplacez chacune *placeholder* par vos propres informations :

- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaa*— L'ID de membre du membre qui peut effectuer la demande. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration. Cela garantit qu'il AWS Clean Rooms n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
- *arn:aws:cleanrooms:us-east-1:555555555555:membership/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaa*— L'ARN de membre unique du membre qui peut effectuer une requête. L'ARN d'adhésion se trouve dans l'onglet Détails de la collaboration. Cela garantit qu' AWS Clean Rooms il n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
- *bucket_name*— Le nom de ressource Amazon (ARN) du compartiment S3. Le nom de ressource Amazon (ARN) se trouve dans l'onglet Propriétés du compartiment dans Amazon S3.
- *accountId*— L' Compte AWS ID dans lequel se trouve le compartiment S3.

bucket_name/optional_key_prefix— Le nom de ressource Amazon (ARN) de la destination des résultats dans Amazon S3. Le nom de ressource Amazon (ARN) se trouve dans l'onglet Propriétés du compartiment dans Amazon S3.

5. Continuez à suivre la procédure de [création d'un rôle à l'aide de politiques de confiance personnalisées \(console\)](#) pour créer le rôle.

Configuration des rôles de service pour AWS Clean Rooms ML

Les rôles nécessaires pour effectuer une modélisation similaire sont différents de ceux nécessaires pour utiliser un modèle personnalisé. Les sections suivantes décrivent les rôles nécessaires à l'exécution de chaque tâche.

Rubriques

- [Configuration des rôles de service pour la modélisation des ressemblances](#)
- [Configuration des rôles de service pour une modélisation personnalisée](#)

Configuration des rôles de service pour la modélisation des ressemblances

Rubriques

- [Création d'un rôle de service pour lire les données d'entraînement](#)
- [Créez un rôle de service pour écrire un segment similaire](#)
- [Création d'un rôle de service pour lire les données de départ](#)

Création d'un rôle de service pour lire les données d'entraînement

AWS Clean Rooms utilise un rôle de service pour lire les données d'entraînement. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas `CreateRole` autorisé, demandez à votre administrateur de créer le rôle de service.

Pour créer un rôle de service afin d'entraîner un ensemble de données

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.
2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).
4. Dans l'éditeur de stratégie, sélectionnez l'onglet JSON, puis copiez et collez la politique suivante.

Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire AWS Glue les métadonnées et les données Amazon S3 correspondantes. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez

configuré vos données S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

Vos AWS Glue ressources et les ressources Amazon S3 sous-jacentes doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartitions",
        "glue:GetPartition",
        "glue:BatchGetPartition",
        "glue:GetUserDefinedFunctions"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:database/databases",
        "arn:aws:glue:us-east-1:111122223333:table/databases/tables",
        "arn:aws:glue:us-east-1:111122223333:catalog",
        "arn:aws:glue:us-east-1:111122223333:database/default"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "glue:CreateDatabase"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:database/default"
      ]
    },
    {
      "Effect": "Allow",
```

```

        "Action": [
            "s3:ListBucket",
            "s3:GetBucketLocation"
        ],
        "Resource": [
            "arn:aws:s3:::bucket"
        ],
        "Condition": {
            "StringEquals": {
                "s3:ResourceAccount": [
                    "111122223333"
                ]
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": [
            "s3:GetObject"
        ],
        "Resource": [
            "arn:aws:s3:::bucketFolders/*"
        ],
        "Condition": {
            "StringEquals": {
                "s3:ResourceAccount": [
                    "111122223333"
                ]
            }
        }
    }
]
}

```

Si vous devez utiliser une clé KMS pour déchiffrer des données, ajoutez cette AWS KMS instruction au modèle précédent :

```

{
    "Effect": "Allow",
    "Action": [
        "kms:Decrypt",
    ],

```

```

    "Resource": [
        "arn:aws:kms:region:accountId:key/keyId"
    ],
    "Condition": {
        "ArnLike": {
            "kms:EncryptionContext:aws:s3:arn":
                "arn:aws:s3:::bucketFolders*"
        }
    }
}
]
}

```

5. Remplacez chacune *placeholder* par vos propres informations :

- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *accountId*— L' Compte AWS ID dans lequel se trouve le compartiment S3.
- *database/databases*, *table/databases/tablescatalog*, et *database/default* — L'emplacement des données d'entraînement auxquelles il est AWS Clean Rooms nécessaire d'accéder.
- *bucket*— Le nom de ressource Amazon (ARN) du compartiment S3. Le nom de ressource Amazon (ARN) se trouve dans l'onglet Propriétés du compartiment dans Amazon S3.
- *bucketFolders*— Le nom des dossiers spécifiques du compartiment S3 auxquels il est AWS Clean Rooms nécessaire d'accéder.

6. Choisissez Suivant.

7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.

8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.

9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.

10. Choisissez Créer un rôle.

11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.
12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAssumeRole",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEqualsIfExists": {
          "aws:SourceAccount": ["111122223333"]
        },
        "ArnLikeIfExists": {
          "aws:SourceArn": "arn:aws:cleanrooms-ml:us-east-1:111122223333:training-dataset/*"
        }
      }
    }
  ]
}
```

SourceAccount C'est toujours le vôtre Compte AWS. Ils SourceArn peuvent être limités à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

accountId est l'ID Compte AWS qui contient les données d'entraînement.

13. Choisissez Suivant et sous Ajouter des autorisations, entrez le nom de la politique que vous venez de créer. (Vous devrez peut-être recharger la page.)
14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.

15. Dans Nom, révision et création, entrez le nom et la description du rôle.

 Note

Le nom du rôle doit correspondre au modèle des `passRole` autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Créez un rôle de service pour écrire un segment similaire

AWS Clean Rooms utilise un rôle de service pour écrire des segments similaires dans un compartiment. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas `CreateRole` autorisé, demandez à votre administrateur de créer le rôle de service.

Pour créer un rôle de service, pour écrire un segment similaire

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.
2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).
4. Dans l'éditeur de stratégie, sélectionnez l'onglet JSON, puis copiez et collez la politique suivante.

 Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire AWS Glue les métadonnées et les données Amazon S3 correspondantes. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous

avez configuré vos données Amazon S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

Vos AWS Glue ressources et les ressources Amazon S3 sous-jacentes doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:ListBucket",
        "s3:GetBucketLocation"
      ],
      "Resource": [
        "arn:aws:s3:::buckets"
      ],
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "accountId"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject"
      ],
      "Resource": [
        "arn:aws:s3:::bucketFolders/*"
      ],
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "accountId"
          ]
        }
      }
    }
  ]
}
```

```

    }
  }
]
}

```

Si vous devez utiliser une clé KMS pour chiffrer des données, ajoutez cette AWS KMS instruction au modèle :

```

{
    "Effect": "Allow",
    "Action": [
        "kms:Encrypt",
        "kms:GenerateDataKey*",
        "kms:ReEncrypt*",
    ],
    "Resource": [
        "arn:aws:kms:region:accountId:key/keyId"
    ],
    "Condition": {
        "ArnLike": {
            "kms:EncryptionContext:aws:s3:arn":
                "arn:aws:s3:::bucketFolders*"
        }
    }
}

```

5. Remplacez chacune *placeholder* par vos propres informations :

- *buckets*— Le nom de ressource Amazon (ARN) du compartiment S3. Le nom de ressource Amazon (ARN) se trouve dans l'onglet Propriétés du compartiment dans Amazon S3.
- *accountId*— L'ID Compte AWS dans lequel se trouve le compartiment S3.
- *bucketFolders*— Le nom des dossiers spécifiques du compartiment S3 auxquels il est AWS Clean Rooms nécessaire d'accéder.
- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *keyId*— La clé KMS nécessaire pour chiffrer vos données.

6. Choisissez Suivant.

7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.
8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.

9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.

10. Choisissez Créer un rôle.
11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.
12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAssumeRole",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEqualsIfExists": {
          "aws:SourceAccount": ["111122223333"]
        },
        "ArnLikeIfExists": {
          "aws:SourceArn": "arn:aws:cleanrooms-ml:us-east-1:111122223333:configured-audience-model/*"
        }
      }
    }
  ]
}
```

SourceAccountC'est toujours le vôtre Compte AWS. Ils SourceArn peuvent être limités à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

13. Choisissez Suivant.
14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.
15. Dans Nom, révision et création, entrez le nom et la description du rôle.

 Note

Le nom du rôle doit correspondre au modèle des passRole autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Création d'un rôle de service pour lire les données de départ

AWS Clean Rooms utilise un rôle de service pour lire les données de départ. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas CreateRole autorisé, demandez à votre administrateur de créer le rôle de service.

Créer un rôle de service pour lire les données de départ stockées dans un compartiment S3.

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.
2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).

4. Dans l'éditeur de politiques, sélectionnez l'onglet JSON, puis copiez-collez l'une des politiques suivantes.

 Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire AWS Glue les métadonnées et les données Amazon S3 correspondantes. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données Amazon S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

Vos AWS Glue ressources et les ressources Amazon S3 sous-jacentes doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::buckets"
      ],
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "accountId"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject"
      ],

```

```

    "Resource": [
      "arn:aws:s3:::bucketFolders/*"
    ],
    "Condition": {
      "StringEquals": {
        "s3:ResourceAccount": [
          "accountId"
        ]
      }
    }
  }
]
}

```

Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour lire les résultats d'une requête SQL et les utiliser comme données d'entrée. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la structure de votre requête. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowCleanRoomsStartQuery",
      "Effect": "Allow",
      "Action": [
        "cleanrooms:GetCollaborationAnalysisTemplate",
        "cleanrooms:GetSchema",
        "cleanrooms:StartProtectedQuery"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowCleanRoomsGetAndUpdateQuery",
      "Effect": "Allow",

```



```

        "Action": [
            "cleanrooms:GetProtectedQuery",
            "cleanrooms:UpdateProtectedQuery"
        ],
        "Resource": [
            "arn:aws:cleanrooms:us-east-1:111122223333:membership/queryRunnerMembershipId"
        ]
    }
]
}

```

Si vous devez utiliser une clé KMS pour déchiffrer des données, ajoutez cette AWS KMS instruction au modèle :

```

{
    "Effect": "Allow",
    "Action": [
        "kms:Decrypt",
        "kms:DescribeKey"
    ],
    "Resource": [
        "arn:aws:kms:region:accountId:key/keyId"
    ],
    "Condition": {
        "ArnLike": {
            "kms:EncryptionContext:aws:s3:arn":
                "arn:aws:s3:::bucketFolders*"
        }
    }
}
]
}

```

5. Remplacez chacune *placeholder* par vos propres informations :

- *buckets*— Le nom de ressource Amazon (ARN) du compartiment S3. Le nom de ressource Amazon (ARN) se trouve dans l'onglet Propriétés du compartiment dans Amazon S3.
- *accountId*— L'ID du compte AWS dans lequel se trouve le compartiment S3.
- *bucketFolders*— Le nom des dossiers spécifiques du compartiment S3 auxquels il est AWS Clean Rooms nécessaire d'accéder.

- **region** : Le nom de Région AWS. Par exemple, **us-east-1**.
- **queryRunnerAccountId**— L'ID du compte AWS qui exécutera les requêtes.
- **queryRunnerMembershipId**— L'ID de membre du membre qui peut effectuer la demande. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration. Cela garantit qu'il AWS Clean Rooms n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
- **keyId**— La clé KMS nécessaire pour chiffrer vos données.

6. Choisissez Suivant.

7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.

8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.

9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.

10. Choisissez Créer un rôle.

11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.

12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAssumeRole",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEqualsIfExists": {
```

```

      "aws:SourceAccount": ["111122223333"]
    },
    "ArnLikeIfExists": {
      "aws:SourceArn": "arn:aws:cleanrooms-m1:us-
east-1:111122223333:audience-generation-job/*"
    }
  }
}

```

SourceAccount C'est toujours le vôtre Compte AWS. Ils SourceArn peuvent être limités à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

13. Choisissez Suivant.
14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.
15. Dans Nom, révision et création, entrez le nom et la description du rôle.

Note

Le nom du rôle doit correspondre au modèle des passRole autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Configuration des rôles de service pour une modélisation personnalisée

Rubriques

- [Création d'un rôle de service pour la modélisation ML personnalisée - Configuration ML](#)
- [Créez un rôle de service pour fournir un modèle de machine learning personnalisé](#)
- [Création d'un rôle de service pour interroger un ensemble de données](#)
- [Créez un rôle de service pour créer une association de tables configurée](#)

Création d'un rôle de service pour la modélisation ML personnalisée - Configuration ML

AWS Clean Rooms utilise un rôle de service pour contrôler qui peut créer une configuration ML personnalisée. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas `CreateRole` autorisé, demandez à votre administrateur de créer le rôle de service.

Ce rôle vous permet d'utiliser l'`MLConfigurationaction` [Put](#).

Pour créer un rôle de service afin de permettre la création d'une configuration ML personnalisée

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.
2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).
4. Dans l'éditeur de stratégie, sélectionnez l'onglet JSON, puis copiez et collez la politique suivante.

Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour accéder et écrire des données dans un compartiment S3 et pour publier CloudWatch des métriques. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données Amazon S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

Vos ressources Amazon S3 doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowS3ObjectWriteForExport",
      "Effect": "Allow",
      "Action": [
        "s3:PutObject"
      ],
      "Resource": [
        "arn:aws:s3:::bucket/*"
      ],
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
            "111122223333"
          ]
        }
      }
    },
    {
      "Sid": "AllowS3KMSEncryptForExport",
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:GenerateDataKey*"
      ],
      "Resource": [
        "arn:aws:kms:us-east-1:111122223333:key/keyId"
      ],
      "Condition": {
        "StringLike": {
          "kms:EncryptionContext:aws:s3:arn":
            "arn:aws:s3:::bucket*"
        }
      }
    },
    {
      "Sid": "AllowCloudWatchMetricsPublishingForTrainingJobs",
```

```

        "Action": "cloudwatch:PutMetricData",
        "Resource": "*",
        "Effect": "Allow",
        "Condition": {
            "StringLike": {
                "cloudwatch:namespace": "/aws/cleanroomsm1/*"
            }
        },
    },
    {
        "Sid": "AllowCloudWatchLogsPublishingForTrainingOrInferenceJobs",
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:CreateLogStream",
            "logs:DescribeLogStreams",
            "logs:PutLogEvents"
        ],
        "Resource": [
            "arn:aws:logs:us-east-1:111122223333:log-group:/aws/cleanroomsm1/*"
        ]
    }
]
}

```

5. Remplacez chacune *placeholder* par vos propres informations :

- *bucket*— Le nom de ressource Amazon (ARN) du compartiment S3. Le nom de ressource Amazon (ARN) se trouve dans l'onglet Propriétés du compartiment dans Amazon S3.
- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *accountId*— L'ID Compte AWS dans lequel se trouve le compartiment S3.
- *keyId*— La clé KMS nécessaire pour chiffrer vos données.

6. Choisissez Suivant.

7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.

8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.

9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.

10. Choisissez Créer un rôle.
11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.
12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "111122223333"
        },
        "ArnLike": {
          "aws:SourceArn": "arn:aws:cleanrooms:us-east-1:111122223333:membership/membershipID"
        }
      }
    }
  ]
}
```

SourceAccount C'est toujours le vôtre Compte AWS. Ils SourceArn peuvent être limités à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

13. Choisissez Suivant.

14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.
15. Dans Nom, révision et création, entrez le nom et la description du rôle.

 Note

Le nom du rôle doit correspondre au modèle des `passRole` autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Créez un rôle de service pour fournir un modèle de machine learning personnalisé

AWS Clean Rooms utilise un rôle de service pour contrôler qui peut créer un algorithme de modèle ML personnalisé. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas `CreateRole` autorisé, demandez à votre administrateur de créer le rôle de service.

Ce rôle vous permet d'utiliser l'[CreateConfiguredModelAlgorithm](#) action.

Pour créer un rôle de service permettant à un membre de fournir un modèle de machine learning personnalisé

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.
2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).
4. Dans l'éditeur de stratégie, sélectionnez l'onglet JSON, puis copiez et collez la politique suivante.

 Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour récupérer l'image docker contenant l'algorithme du modèle. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données Amazon S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

Vos ressources Amazon S3 doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowECRIImageDownloadForTrainingAndInferenceJobs",
      "Effect": "Allow",
      "Action": [
        "ecr:BatchGetImage",
        "ecr:BatchCheckLayerAvailability",
        "ecr:GetDownloadUrlForLayer"
      ],
      "Resource": "arn:aws:ecr:us-east-1:111122223333:repository/repoName"
    }
  ]
}
```

5. Remplacez chacune *placeholder* par vos propres informations :

- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *accountId*— L' Compte AWS ID dans lequel se trouve le compartiment S3.
- *repoName*— Le nom du référentiel qui contient vos données.

6. Choisissez Suivant.

7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.

8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.

9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.

10. Choisissez Créer un rôle.

11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.

12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

SourceAccountC'est toujours votre. SourceArn Vous pouvez Compte AWS le limiter à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

13. Choisissez Suivant.

14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.

15. Dans Nom, révision et création, entrez le nom et la description du rôle.

 Note

Le nom du rôle doit correspondre au modèle des `passRole` autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Création d'un rôle de service pour interroger un ensemble de données

AWS Clean Rooms utilise un rôle de service pour contrôler qui peut interroger un ensemble de données qui sera utilisé pour la modélisation ML personnalisée. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas `CreateRole` autorisé, demandez à votre administrateur de créer le rôle de service.

Ce rôle vous permet d'utiliser l'action [Créer un MLInput canal](#).

Pour créer un rôle de service permettant à un membre d'interroger un ensemble de données

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.
2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).
4. Dans l'éditeur de stratégie, sélectionnez l'onglet JSON, puis copiez et collez la politique suivante.

 Note

L'exemple de politique suivant prend en charge les autorisations nécessaires pour interroger un ensemble de données qui sera utilisé pour la modélisation ML personnalisée. Toutefois, il se peut que vous deviez modifier cette politique en fonction

de la façon dont vous avez configuré vos données Amazon S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.
Vos ressources Amazon S3 doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowCleanRoomsStartQueryForMLInputChannel",
      "Effect": "Allow",
      "Action": "cleanrooms:StartProtectedQuery",
      "Resource": "*"
    },
    {
      "Sid": "AllowCleanroomsGetSchemaAndGetAnalysisTemplateForMLInputChannel",
      "Effect": "Allow",
      "Action": [
        "cleanrooms:GetSchema",
        "cleanrooms:GetCollaborationAnalysisTemplate"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowCleanRoomsGetAndUpdateQueryForMLInputChannel",
      "Effect": "Allow",
      "Action": [
        "cleanrooms:GetProtectedQuery",
        "cleanrooms:UpdateProtectedQuery"
      ],
      "Resource": [
        "arn:aws:cleanrooms:us-  
east-1:111122223333:membership/queryRunnerMembershipId"
      ]
    }
  ]
}
```

5. Remplacez chacune *placeholder* par vos propres informations :
 - *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
 - *queryRunnerAccountId*— L'ID du compte AWS du compte qui exécutera les requêtes.
 - *queryRunnerMembershipId*— L'ID de membre du membre qui peut effectuer la demande. L'identifiant de membre se trouve dans l'onglet Détails de la collaboration. Cela garantit qu'il AWS Clean Rooms n'assume le rôle que lorsque ce membre exécute l'analyse dans le cadre de cette collaboration.
6. Choisissez Suivant.
7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.
8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.
9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.
10. Choisissez Créer un rôle.
11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.
12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

```
}
```

SourceAccountC'est toujours votre. SourceArn Vous pouvez Compte AWS le limiter à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

13. Choisissez Suivant.
14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.
15. Dans Nom, révision et création, entrez le nom et la description du rôle.

 Note

Le nom du rôle doit correspondre au modèle des passRole autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Créez un rôle de service pour créer une association de tables configurée

AWS Clean Rooms utilise un rôle de service pour contrôler qui peut créer une association de tables configurée. Vous pouvez créer ce rôle à l'aide de la console si vous disposez des autorisations IAM nécessaires. Si vous n'êtes pas CreateRole autorisé, demandez à votre administrateur de créer le rôle de service.

Ce rôle vous permet d'utiliser l' CreateConfiguredTableAssociation action.

Pour créer un rôle de service afin de permettre la création d'une association de tables configurée

1. Connectez-vous à la console IAM (<https://console.aws.amazon.com/iam/>) avec votre compte d'administrateur.

2. Sous Access Management (Gestion des accès), choisissez Politiques (politiques).
3. Choisissez Create Policy (Créer une politique).
4. Dans l'éditeur de stratégie, sélectionnez l'onglet JSON, puis copiez et collez la politique suivante.

Note

L'exemple de politique suivant prend en charge la création d'une association de tables configurée. Toutefois, il se peut que vous deviez modifier cette politique en fonction de la façon dont vous avez configuré vos données Amazon S3. Cette politique n'inclut pas de clé KMS pour déchiffrer les données.

Vos ressources Amazon S3 doivent être identiques à Région AWS celles de la AWS Clean Rooms collaboration.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "kms:Decrypt",
        "kms:DescribeKey"
      ],
      "Resource": "arn:aws:kms:us-east-1:111122223333:key/KMS-key-ID",
      "Effect": "Allow"
    },
    {
      "Action": [
        "s3:ListBucket",
        "s3:GetBucketLocation"
      ],
      "Resource": "arn:aws:s3:::bucket-name",
      "Effect": "Allow"
    },
    {
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::bucket-name/*",
      "Effect": "Allow"
    }
  ]
}
```

```

    {
      "Action": [
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetTable",
        "glue:GetTables",
        "glue:GetPartitions",
        "glue:GetPartition",
        "glue:BatchGetPartition"
      ],
      "Resource": [
        "arn:aws:glue:us-east-1:111122223333:catalog",
        "arn:aws:glue:us-east-1:111122223333:database/Glue database name",
        "arn:aws:glue:us-east-1:111122223333:table/Glue database name/Glue table name"
      ],
      "Effect": "Allow"
    },
    {
      "Action": [
        "glue:GetSchema",
        "glue:GetSchemaVersion"
      ],
      "Resource": "*",
      "Effect": "Allow"
    }
  ]
}

```

Remplacer la ressource fictive ARNs

Lorsque vous utilisez cette politique, vous devez remplacer les identifiants de ressource réservés par les identifiants réels ARNs de vos ressources :

- AWS KMS Ressource clé : *KMS-key-ID* remplacez-la par l'identifiant de AWS KMS clé réel qui chiffre vos données Amazon S3. La clé doit se trouver dans le même compte (111122223333) qui possède les ressources du catalogue. AWS Glue
- Ressources du compartiment Amazon S3 : remplacez *bucket-name* par le nom réel du compartiment Amazon S3 qui contient les données de votre AWS Glue table. Notez

que le compartiment Amazon S3 ARNs n'inclut pas de compte IDs car les noms de compartiment sont uniques au niveau mondial.

- AWS Glue Ressources : remplacez les espaces réservés suivants par les noms réels de vos ressources :
 - *Glue database name*- Le nom de votre AWS Glue base de données
 - *Glue table name*- Le nom de votre AWS Glue table

Toutes les AWS Glue ressources (catalogue, base de données et table) doivent être identiques Compte AWS (111122223333) pour garantir des autorisations d'accès cohérentes. Ce compte doit être le même que celui qui possède la AWS KMS clé utilisée pour le chiffrement des données, afin de créer une limite de sécurité unifiée pour vos ressources de AWS Clean Rooms données.

5. Remplacez chacune *placeholder* par vos propres informations :

- *KMS key used to encrypt the Amazon S3 data*— La clé KMS utilisée pour chiffrer les données Amazon S3. Pour déchiffrer les données, vous devez fournir la même clé KMS que celle utilisée pour chiffrer les données.
- *Amazon S3 bucket of AWS Glue table*— Le nom du compartiment Amazon S3 qui contient la AWS Glue table contenant vos données.
- *region* : Le nom de Région AWS. Par exemple, **us-east-1**.
- *accountId*— L' Compte AWS identifiant du compte propriétaire des données.
- *AWS Glue database name*— Le nom de la AWS Glue base de données qui contient vos données.
- *AWS Glue table name*— Nom de la AWS Glue table contenant vos données.

6. Choisissez Suivant.

7. Pour Révision et création, entrez le nom et la description de la politique, puis consultez le résumé.

8. Choisissez Create Policy (Créer une politique).

Vous avez créé une politique pour AWS Clean Rooms.

9. Sous Access Management (Gestion des accès), choisissez Roles (Rôles).

Avec les rôles, vous pouvez créer des informations d'identification à court terme, ce qui est recommandé pour renforcer la sécurité. Vous pouvez également sélectionner Utilisateurs pour créer des informations d'identification à long terme.

10. Choisissez Créer un rôle.
11. Dans l'assistant de création de rôle, pour Type d'entité fiable, choisissez Politique de confiance personnalisée.
12. Copiez et collez la politique de confiance personnalisée suivante dans l'éditeur JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

SourceAccount C'est toujours votre. SourceArn Vous pouvez Compte AWS le limiter à un ensemble de données d'entraînement spécifique, mais uniquement après la création de cet ensemble de données. Comme vous ne connaissez pas encore l'ARN du jeu de données d'entraînement, le caractère générique est spécifié ici.

13. Choisissez Suivant.
14. Cochez la case à côté du nom de la politique que vous avez créée, puis choisissez Next.
15. Dans Nom, révision et création, entrez le nom et la description du rôle.

 Note

Le nom du rôle doit correspondre au modèle des passRole autorisations accordées au membre qui peut interroger et recevoir des résultats et des rôles de membre.

- a. Passez en revue Sélectionnez les entités fiables et modifiez-les si nécessaire.
- b. Passez en revue les autorisations dans Ajouter des autorisations et modifiez-les si nécessaire.
- c. Passez en revue les balises et ajoutez-en si nécessaire.
- d. Choisissez Créer un rôle.

Vous avez créé le rôle de service pour AWS Clean Rooms.

Collaborations et adhésions à AWS Clean Rooms

Une collaboration est une limite logique sécurisée AWS Clean Rooms dans laquelle les membres peuvent effectuer des analyses sur des tables configurées.

Tous les membres AWS Clean Rooms peuvent créer une collaboration.

Le créateur de la collaboration peut désigner un seul membre pour analyser les tables configurées et recevoir les résultats. Toutefois, le créateur de la collaboration souhaitera peut-être empêcher le membre autorisé à exécuter l'analyse d'accéder aux résultats de la requête. Dans ce cas, le créateur de la collaboration peut désigner un [membre qui peut effectuer des requêtes](#) ou [un membre qui peut exécuter des requêtes et des tâches](#) et un autre [membre qui peut recevoir des résultats](#).

Dans la plupart des cas, le membre qui peut effectuer des requêtes ou le membre qui peut interroger et exécuter des tâches est également le [membre qui paie les frais de calcul](#). Cependant, le créateur de la collaboration peut configurer un autre membre pour qu'il soit responsable du paiement des coûts de calcul des requêtes.

Pour plus d'informations sur la création d'une collaboration à l'aide du AWS SDKs, consultez la [référence des AWS Clean Rooms API](#).

Rubriques

- [Création d'une collaboration](#)
- [Création d'un abonnement et adhésion à une collaboration](#)
- [Collaborations d'édition](#)
- [Demandes de modification dans AWS Clean Rooms](#)
- [Supprimer des collaborations](#)
- [Afficher les collaborations](#)
- [Inviter des membres à participer à une collaboration](#)
- [Surveillance des membres](#)
- [Ajouter des membres à une collaboration](#)
- [Supprimer des membres d'une collaboration](#)
- [Quitter une collaboration](#)

Création d'une collaboration

Il existe trois manières de créer une collaboration dans AWS Clean Rooms.

La forme la plus élémentaire est la [collaboration pour les requêtes](#). Cette collaboration se concentre sur l'analyse des requêtes SQL et maintient une structure simple avec deux rôles principaux : un membre qui peut exécuter des requêtes et un autre qui peut recevoir des résultats. Cette configuration de collaboration de base fonctionne bien pour les tâches d'analyse de données simples.

La deuxième forme, la [collaboration pour les requêtes et les tâches](#), étend les fonctionnalités en incorporant à la fois des requêtes SQL et PySpark des tâches. Cette configuration de collaboration conserve la même structure de rôles de base mais étend les autorisations pour inclure l'exécution des tâches. Une exigence notable est que le membre qui crée les modèles d' PySpark analyse doit également être celui qui reçoit les résultats, ce qui garantit une responsabilité claire dans le processus d'analyse.

La troisième forme, la [collaboration pour la modélisation du machine learning](#), est conçue pour les flux de travail d'apprentissage automatique. Cette configuration de collaboration ajoute deux rôles supplémentaires : un pour les utilisateurs qui ont besoin des résultats de modèles entraînés, et un autre pour les utilisateurs qui ont besoin des résultats de l'utilisation de ces modèles pour faire des prédictions. Cette configuration de collaboration permet aux membres de la collaboration de travailler ensemble sur des projets de données complexes tout en définissant clairement les rôles et les autorisations de chacun.

Les rubriques suivantes expliquent comment créer des collaborations pour les requêtes, les tâches et la modélisation ML.

Rubriques

- [Création d'une collaboration pour les requêtes](#)
- [Création d'une collaboration pour les requêtes et les tâches](#)
- [Création d'une collaboration pour la modélisation ML](#)

Création d'une collaboration pour les requêtes

Dans cette procédure, en tant que [créateur de la collaboration](#), vous effectuez les tâches suivantes :

- [Créez une collaboration](#).
- Invitez un ou plusieurs [membres](#) à rejoindre la [collaboration](#).

- Attribuez des capacités aux membres, telles que le [membre qui peut interroger](#) et le [membre qui peut recevoir des résultats](#).

Si le créateur de la collaboration est également le membre habilité à recevoir les résultats, il spécifie la destination et le format des résultats. Ils fournissent également un rôle de service Amazon Resource Name (ARN) pour écrire les résultats dans la destination des résultats.

- Configurez quel [membre est responsable du paiement des coûts de calcul dans le cadre de la collaboration](#).

Avant de commencer, assurez-vous de remplir les prérequis suivants :

- Vous avez le nom et l' Compte AWS identifiant de chaque membre que vous souhaitez inviter à rejoindre la collaboration.
- Vous êtes autorisé à partager le nom et l' Compte AWS identifiant de chaque membre avec tous les membres de la collaboration.

 Note

Vous ne pouvez pas ajouter d'autres membres après avoir créé la collaboration.

Pour plus d'informations sur la création d'une collaboration à l'aide du AWS SDKs, consultez la [référence des AWS Clean Rooms API](#).

Pour créer une collaboration pour les requêtes

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Dans le coin supérieur droit, choisissez Créer une collaboration.
4. Pour l'étape 1 : définir la collaboration, procédez comme suit :
 - a. Pour plus de détails, entrez le nom et la description de la collaboration.

Ces informations seront visibles par les membres de la collaboration qui sont invités à participer à la collaboration. Le nom et la description les aident à comprendre à quoi fait référence la collaboration.

b. Pour les membres :

- i. Pour le membre 1 : vous devez saisir le nom d'affichage de votre membre tel que vous souhaitez qu'il apparaisse pour la collaboration.

 Note

Votre Compte AWS identifiant est automatiquement inclus comme Compte AWS identifiant de membre.

- ii. Pour Membre 2, entrez le nom d'affichage du membre et l' Compte AWS ID du membre que vous souhaitez inviter à rejoindre la collaboration.

Le nom d'affichage et l' Compte AWS identifiant du membre seront visibles par toutes les personnes invitées à la collaboration. Une fois que vous avez saisi et enregistré les valeurs de ces champs, vous ne pouvez pas les modifier.

 Note

Vous devez informer le membre de la collaboration que son Compte AWS identifiant de membre et son nom d'affichage seront visibles par tous les collaborateurs invités et actifs de la collaboration.

- iii. Si vous souhaitez ajouter un autre membre, choisissez Ajouter un autre membre. Entrez ensuite le nom d'affichage du membre et l' Compte AWS identifiant de membre pour chaque membre susceptible de fournir les données que vous souhaitez inviter à la collaboration.

- c. Si vous souhaitez activer la journalisation des analyses, cochez la case Activer la journalisation des analyses.

- Cochez la case Journaux issus des requêtes sous les types de journaux pris en charge.

Vous recevrez les journaux générés à partir de requêtes SQL sur votre compte Amazon CloudWatch Logs.

- d. Si vous souhaitez activer la surveillance détaillée, cochez la case Activer la surveillance détaillée.

Le responsable de l'analyse et le payeur configuré peuvent choisir d'activer des métriques détaillées lors de la création de leur adhésion. Lorsque cette option est activée, des mesures de surveillance détaillées seront publiées CloudWatch pour le suivi opérationnel des collaborations, notamment les performances des requêtes et l'utilisation des ressources. Ces statistiques seront mises à la disposition du responsable de l'analyse et du payeur configuré dans leurs versions respectives Compte AWS.

Pour plus d'informations sur la CloudWatch tarification, consultez la section [CloudWatch Tarification](#).

- e. Sous Régions de résultats de requête autorisées, sélectionnez un ou plusieurs Régions AWS endroits où vous souhaitez envoyer les résultats de la requête.

Par défaut, seule la région actuelle (telle que North Virginia us-east-1) est sélectionnée.

 Important

Lorsque vous activez la livraison des résultats de requêtes entre régions, vos résultats peuvent être traités et stockés en dehors de la région source.

Pour plus d'informations sur les régions, voir [Régions et points de terminaison](#) dans le Références générales AWS.

- f. (Facultatif) Gérez l'accès à vos données grâce à l'approbation automatique des demandes de modification en configurant les paramètres qui peuvent être modifiés automatiquement sans approbation manuelle pour une demande de modification. Par défaut, certains paramètres ne peuvent être modifiés qu'en soumettant une demande de modification qui doit être approuvée par tous les membres avant qu'elle ne prenne effet.
 - Accorder des capacités aux membres : choisissez les capacités qui peuvent être accordées aux membres de la collaboration sans approbation manuelle. Les membres peuvent toujours fournir des données.
 - Choisissez des capacités :
 - Contribuer aux données (toujours activé)
 - Recevez les résultats
 - Approuver automatiquement les nouveaux membres dotés de ces capacités : si cela est autorisé, tous les membres ajoutés avec les capacités sélectionnées ci-dessus

rejoindront instantanément la collaboration. Les membres ajoutés avec d'autres capacités devront toujours être approuvés manuellement pour s'inscrire.

- Capacités pouvant être révoquées automatiquement - Choisissez les capacités qui peuvent être révoquées sans approbation manuelle. Les membres peuvent toujours fournir des données.
- Choisissez des capacités :
 - Contribuer aux données (toujours activé)
 - Recevez les résultats

Si vous choisissez cette option, vous pouvez suivre toutes les modifications de configuration de collaboration via l'historique des demandes de modification, situé dans l'onglet Détails de la page des détails de la collaboration.

g. (Facultatif) Si vous souhaitez activer la fonctionnalité de calcul cryptographique, cochez la case Activer le calcul cryptographique.

i. Choisissez les paramètres de couverture cryptographique suivants :

- Autoriser plaintext les colonnes

Choisissez Non si vous avez besoin de tables entièrement chiffrées.

Choisissez Oui si vous souhaitez que cleartext les colonnes soient autorisées dans la table cryptée.

Pour fonctionner SUM ou AVG sur certaines colonnes, les colonnes doivent être insérées cleartext.

- Préservez NULL les valeurs

Choisissez Non si vous ne souhaitez pas conserver NULL les valeurs. NULL les valeurs n'apparaîtront pas comme NULL dans une table cryptée.

Choisissez Oui si vous souhaitez conserver NULL les valeurs. NULL les valeurs apparaîtront comme NULL dans une table cryptée.

ii. Choisissez les paramètres d'empreinte suivants :

- Autoriser les doublons

Choisissez Non si vous ne souhaitez pas que les doublons soient autorisés dans une fingerprint colonne.

Choisissez Oui si vous souhaitez que les entrées dupliquées soient autorisées dans une fingerprint colonne.

- JOINAutorisation de colonnes portant des noms différents

Choisissez Non si vous ne souhaitez pas joindre des fingerprint colonnes portant des noms différents.

Choisissez Oui si vous souhaitez joindre des fingerprint colonnes portant des noms différents.

Pour plus d'informations sur les paramètres de calcul cryptographique, consultez [Paramètres de calcul cryptographique](#).

Pour plus d'informations sur la façon de chiffrer vos données pour les utiliser dans AWS Clean Rooms, consultez [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

 Note

Vérifiez soigneusement ces configurations avant de passer à l'étape suivante. Après avoir créé la collaboration, vous pouvez uniquement modifier le nom et la description de la collaboration et indiquer si les journaux sont stockés dans Amazon CloudWatch Logs.

- h. Si vous souhaitez activer les balises pour la ressource de collaboration, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
 - i. Choisissez Suivant.
5. Pour l'étape 2 : Spécifier les capacités des membres, pour l'analyse à l'aide de requêtes et de tâches, dans la section Types d'analyse pris en charge, laissez la case Requête cochée et prenez les mesures recommandées en fonction de votre objectif.

Votre objectif	Action recommandée
Interrogez les données de la collaboration et recevez les résultats	1. Choisissez-vous comme membre autorisé à exécuter des requêtes. 2. Choisissez-vous comme membre qui peut recevoir les résultats des analyses dans la liste déroulante.
Interrogez les données de la collaboration et désignez un autre membre pour recevoir les résultats	1. Choisissez-vous comme membre autorisé à exécuter des requêtes. 2. Sélectionnez le membre qui peut recevoir les résultats des analyses dans la liste déroulante.
Recevez les résultats de la requête dans la collaboration et désignez un autre membre pour interroger les données	1. Sélectionnez le membre autorisé à exécuter des requêtes dans la liste déroulante. 2. Choisissez-vous comme membre qui peut recevoir les résultats des analyses dans la liste déroulante.
Créez et gérez la collaboration, assignez un autre membre pour interroger les données et affectez un autre membre pour recevoir les résultats	1. Sélectionnez le membre autorisé à exécuter des requêtes dans la liste déroulante. 2. Sélectionnez le membre qui peut recevoir les résultats des analyses dans la liste déroulante.

- a. Si vous utilisez Clean Rooms ML, pour la modélisation du ML à l'aide de flux de travail spécialement conçus,
 - i. (Facultatif) Sélectionnez le membre qui peut recevoir les résultats des modèles entraînés dans la liste déroulante.
 - ii. (Facultatif) Sélectionnez le membre qui peut recevoir le résultat de l'inférence du modèle dans la liste déroulante.

- b. Consultez les capacités des membres sous Résolution d'identification en utilisant Résolution des entités AWS.
 - c. Choisissez Suivant.
6. Pour l'étape 3 : configurer le paiement, pour l'analyse à l'aide de requêtes, effectuez l'une des actions suivantes en fonction de votre objectif.

Votre objectif	Action recommandée
Désignez le membre qui peut exécuter des requêtes comme étant le membre qui paie les coûts de calcul de la requête	1. Pour l'analyse à l'aide de requêtes, choisissez le membre qui paiera pour les requêtes comme étant le même que le membre autorisé à exécuter des requêtes. 2. Choisissez Suivant.
Désignez un membre différent pour payer les coûts de calcul de la requête	1. Pour l'analyse à l'aide de requêtes, choisissez-vous comme membre qui paiera pour les requêtes. 2. Choisissez Suivant.

Pour la modélisation ML à l'aide de flux de travail spécialement conçus, le créateur du modèle de similarité configuré est le membre qui paiera pour la modélisation de similarité.

Pour la résolution des identifiants avec Résolution des entités AWS, le créateur de la table de mappage des identifiants est le membre qui paiera pour la table de mappage des identifiants.

7. Pour l'étape 4 : Configuration de l'adhésion, choisissez l'une des options suivantes :

Yes, join by creating membership now

1. Pour les paramètres de résultats par défaut, pour les paramètres de résultats de requête, si vous êtes le membre autorisé à recevoir les résultats,
 - a. Pour la destination des résultats dans Amazon S3, entrez la destination Amazon S3 ou choisissez Browse S3 pour sélectionner un compartiment S3.
 - b. Pour le format du résultat de la requête, choisissez CSV ou PARQUET.
 - c. (Spark uniquement) Pour les fichiers de résultats, choisissez Multiple ou Single.

- d. (Facultatif) Pour accéder au service, si vous souhaitez envoyer des requêtes qui prennent jusqu'à 24 heures à votre destination S3, cochez la case Ajouter un rôle de service pour prendre en charge les requêtes dont le traitement prend jusqu'à 24 heures.

Les requêtes volumineuses dont le traitement prend jusqu'à 24 heures seront livrées à votre destination S3.

Si vous ne cochez pas cette case, seules les requêtes traitées dans les 12 heures seront livrées à votre agence S3.

- e. Spécifiez les autorisations d'accès au service en sélectionnant Créer et utiliser un nouveau rôle de service ou Utiliser un rôle de service existant.

Si tu choisis de...	Alors...
Création et utilisation d'un nouveau rôle de service	• AWS Clean Rooms crée un rôle de service avec la politique requise pour cette table. • Le nom du rôle de service par défaut est <code>cleanrooms-result-receiver-<timestamp></code> • Vous devez disposer des autorisations nécessaires pour créer des rôles et associer des politiques.

Si tu choisis de...	Alors...
Utiliser un rôle de service existant	i. Choisissez un nom de rôle de service existant dans la liste déroulante. La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles. Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser. ii. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM. S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible. Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

 Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).
- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.

- Si vous ne parvenez pas à modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique pour le rôle de service est introuvable.

2. Pour les paramètres des journaux, choisissez l'une des options suivantes pour le stockage des CloudWatch journaux dans Amazon Logs :

 Note

La section Paramètres des journaux apparaît si vous avez choisi d'activer la journalisation des requêtes.

- a. Choisissez Activer et les journaux de requêtes qui vous concernent seront stockés dans votre compte Amazon CloudWatch Logs.

Chaque membre ne peut recevoir que les journaux des requêtes qu'il a initiées ou qui contiennent ses données.

Le membre qui peut recevoir les résultats reçoit également des journaux pour toutes les requêtes exécutées dans le cadre d'une collaboration, même si ses données ne sont pas accessibles dans le cadre d'une requête.

Sous Types de journaux pris en charge, la case à cocher Journaux des requêtes est activée par défaut.

 Note

Après avoir activé la journalisation des requêtes, la configuration du stockage des journaux et le début de la réception des journaux dans Amazon CloudWatch Logs peuvent prendre quelques minutes. Pendant cette brève période, le membre autorisé à effectuer des requêtes peut exécuter des requêtes qui n'envoient pas réellement de journaux.

- b. Choisissez Désactiver et les journaux de requêtes qui vous concernent ne seront pas stockés dans votre compte Amazon CloudWatch Logs.
3. Si vous souhaitez activer les balises pour la ressource d'adhésion, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.

4. Si vous êtes le membre qui paie pour Query Compute, indiquez votre acceptation en cochant la case J'accepte de payer les coûts de calcul dans le cadre de cette collaboration.

 Note

Vous devez cocher cette case pour continuer.

Pour plus d'informations sur le mode de calcul des prix, consultez [Tarification pour AWS Clean Rooms](#).

Si vous êtes le [membre qui paie les frais de calcul des requêtes, mais que vous n'êtes pas le membre habilité AWS Budgets à effectuer des requêtes](#), il est recommandé de configurer un budget AWS Clean Rooms et de recevoir des notifications une fois le budget maximum atteint. Pour plus d'informations sur la configuration d'un budget, consultez [la section Gérer vos coûts AWS Budgets](#) dans le Guide de AWS Cost Management l'utilisateur. Pour plus d'informations sur la configuration des notifications, consultez la rubrique [Création d'un compte Amazon SNS pour les notifications budgétaires](#) dans le guide de l'AWS Cost Management utilisateur. Si le budget maximum est atteint, vous pouvez contacter le membre qui pourra lancer des requêtes ou [quitter la collaboration](#). Si vous quittez la collaboration, aucune autre requête ne sera autorisée à être exécutée et, par conséquent, les frais de calcul des requêtes ne vous seront plus facturés.

5. Choisissez Suivant.

La collaboration et votre adhésion sont créées.

Votre statut dans la collaboration est actif.

No, I will create a membership later

1. Choisissez Suivant.

Seule la collaboration est créée.

Votre statut dans la collaboration est inactif.

8. Pour l'étape 5 : révision et création, procédez comme suit :

- a. Passez en revue les sélections que vous avez effectuées lors des étapes précédentes et modifiez-les si nécessaire.
- b. Choisissez l'une des options.

Si vous avez choisi de...	Ensuite, choisissez...
Créez un abonnement avec la collaboration (Oui, inscrivez-vous en créant un abonnement maintenant)	Créez une collaboration et une adhésion
Créez la collaboration et ne créez pas d'adhésion pour le moment (Non, je créerai un abonnement plus tard)	Créez une collaboration

Une fois que votre collaboration a été créée avec succès, vous pouvez voir la page des détails de la collaboration sous Collaborations.

Vous êtes maintenant prêt à :

- [Préparez votre tableau de données à analyser AWS Clean Rooms](#). (Facultatif si vous souhaitez analyser vos propres données d'événement ou si vous souhaitez interroger des données d'identité.)
- [Associez le tableau configuré à votre collaboration](#). (Facultatif si vous souhaitez analyser vos propres données d'événement.)
- [Ajoutez une règle d'analyse pour la table configurée](#). (Facultatif si vous souhaitez analyser vos propres données d'événement.)
- [Créez un abonnement et rejoignez une collaboration](#). (Facultatif si vous avez déjà créé un abonnement.)
- [Invitez des membres à rejoindre la collaboration](#).

Création d'une collaboration pour les requêtes et les tâches

Dans cette procédure, en tant que [créateur de la collaboration](#), vous effectuez les tâches suivantes :

- [Créez une collaboration](#).

- Invitez un ou plusieurs [membres](#) à rejoindre la [collaboration](#).
- Attribuez des capacités aux membres, telles que le [membre qui peut exécuter des requêtes et des tâches](#) et le [membre qui peut recevoir des résultats](#).

Si le créateur de la collaboration est également le membre habilité à recevoir les résultats, il spécifie la destination et le format des résultats. Ils fournissent également un rôle de service Amazon Resource Name (ARN) pour écrire les résultats dans la destination des résultats.

- Configurez quel [membre est chargé de payer les coûts de calcul des requêtes et des tâches dans le cadre de la collaboration](#).

Avant de commencer, assurez-vous de remplir les prérequis suivants :

- Vous avez le nom et l' Compte AWS identifiant de chaque membre que vous souhaitez inviter à rejoindre la collaboration.
- Vous êtes autorisé à partager le nom et l' Compte AWS identifiant de chaque membre avec tous les membres de la collaboration.

 Note

Vous ne pouvez pas ajouter d'autres membres après avoir créé la collaboration.

Pour plus d'informations sur la création d'une collaboration à l'aide du AWS SDKs, consultez la [référence des AWS Clean Rooms API](#).

Pour créer une collaboration pour les requêtes et les tâches

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Dans le coin supérieur droit, choisissez Créer une collaboration.
4. Pour l'étape 1 : définir la collaboration, procédez comme suit :
 - a. Pour plus de détails, entrez le nom et la description de la collaboration.

Ces informations seront visibles par les membres de la collaboration qui sont invités à participer à la collaboration. Le nom et la description les aident à comprendre à quoi fait référence la collaboration.

b. Pour les membres :

- i. Pour le membre 1 : vous devez saisir le nom d'affichage de votre membre tel que vous souhaitez qu'il apparaisse pour la collaboration.

 Note

Votre Compte AWS identifiant est automatiquement inclus comme Compte AWS identifiant de membre.

- ii. Pour Membre 2, entrez le nom d'affichage du membre et l' Compte AWS ID du membre que vous souhaitez inviter à rejoindre la collaboration.

Le nom d'affichage et l' Compte AWS identifiant du membre seront visibles par toutes les personnes invitées à la collaboration. Une fois que vous avez saisi et enregistré les valeurs de ces champs, vous ne pouvez pas les modifier.

 Note

Vous devez informer le membre de la collaboration que son Compte AWS identifiant de membre et son nom d'affichage seront visibles par tous les collaborateurs invités et actifs de la collaboration.

- iii. Si vous souhaitez ajouter un autre membre, choisissez Ajouter un autre membre. Entrez ensuite le nom d'affichage du membre et l' Compte AWS identifiant de membre pour chaque membre susceptible de fournir les données que vous souhaitez inviter à la collaboration.
- c. Si vous souhaitez activer la journalisation des analyses, cochez la case Activer la journalisation des analyses, puis choisissez les types de journaux pris en charge.
- Si vous souhaitez recevoir les journaux générés à partir de requêtes SQL, cochez la case Journaux à partir de requêtes.
 - Si vous souhaitez recevoir les journaux générés à partir des tâches à l'aide de PySpark, cochez la case Journaux des tâches.

- d. Si vous souhaitez activer la surveillance détaillée, cochez la case Activer la surveillance détaillée.

Le responsable de l'analyse et le payeur configuré peuvent choisir d'activer des métriques détaillées lors de la création de leur adhésion. Lorsque cette option est activée, des mesures de surveillance détaillées seront publiées CloudWatch pour le suivi opérationnel des collaborations, notamment les performances des requêtes et l'utilisation des ressources. Ces statistiques seront mises à la disposition du responsable de l'analyse et du payeur configuré dans leurs versions respectives Compte AWS.

Pour plus d'informations sur la CloudWatch tarification, consultez la section [CloudWatch Tarification](#).

- e. Sous Régions de résultats de requête autorisées, sélectionnez un ou plusieurs Régions AWS endroits où vous souhaitez envoyer les résultats de la requête.

Par défaut, seule la région actuelle (telle que North Virginia us-east-1) est sélectionnée.

 Important

Lorsque vous activez la livraison des résultats de requêtes entre régions, vos résultats peuvent être traités et stockés en dehors de la région source.

Pour plus d'informations sur les régions, voir [Régions et points de terminaison](#) dans le Références générales AWS.

- f. (Facultatif) Gérez l'accès à vos données grâce à l'approbation automatique des demandes de modification en configurant les paramètres qui peuvent être modifiés automatiquement sans approbation manuelle pour une demande de modification. Par défaut, certains paramètres ne peuvent être modifiés qu'en soumettant une demande de modification qui doit être approuvée par tous les membres avant qu'elle ne prenne effet.
- Accorder des capacités aux membres : choisissez les capacités qui peuvent être accordées aux membres de la collaboration sans approbation manuelle. Les membres peuvent toujours fournir des données.
 - Choisissez des capacités :
 - Contribuer aux données (toujours activé)
 - Recevez les résultats

- Approuver automatiquement les nouveaux membres dotés de ces capacités : si cela est autorisé, tous les membres ajoutés avec les capacités sélectionnées ci-dessus rejoindront instantanément la collaboration. Les membres ajoutés avec d'autres capacités devront toujours être approuvés manuellement pour s'inscrire.
- Capacités pouvant être révoquées automatiquement - Choisissez les capacités qui peuvent être révoquées sans approbation manuelle. Les membres peuvent toujours fournir des données.
- Choisissez des capacités :
 - Contribuer aux données (toujours activé)
 - Recevez les résultats

Si vous choisissez cette option, vous pouvez suivre toutes les modifications de configuration de collaboration via l'historique des demandes de modification, situé dans l'onglet Détails de la page des détails de la collaboration.

g. (Facultatif) Si vous souhaitez activer la fonctionnalité de calcul cryptographique, cochez la case Activer le calcul cryptographique.

i. Choisissez les paramètres de couverture cryptographique suivants :

- Autoriser plaintext les colonnes

Choisissez Non si vous avez besoin de tables entièrement chiffrées.

Choisissez Oui si vous souhaitez que cleartext les colonnes soient autorisées dans la table cryptée.

Pour fonctionner SUM ou AVG sur certaines colonnes, les colonnes doivent être insérées cleartext.

- Préservez NULL les valeurs

Choisissez Non si vous ne souhaitez pas conserver NULL les valeurs. NULL les valeurs n'apparaîtront pas comme NULL dans une table cryptée.

Choisissez Oui si vous souhaitez conserver NULL les valeurs. NULL les valeurs apparaîtront comme NULL dans une table cryptée.

ii. Choisissez les paramètres d'empreinte suivants :

- Autoriser les doublons

Choisissez Non si vous ne souhaitez pas que les doublons soient autorisés dans une fingerprint colonne.

Choisissez Oui si vous souhaitez que les entrées dupliquées soient autorisées dans une fingerprint colonne.

- JOINAutorisation de colonnes portant des noms différents

Choisissez Non si vous ne souhaitez pas joindre des fingerprint colonnes portant des noms différents.

Choisissez Oui si vous souhaitez joindre des fingerprint colonnes portant des noms différents.

Pour plus d'informations sur les paramètres de calcul cryptographique, consultez [Paramètres de calcul cryptographique](#).

Pour plus d'informations sur la façon de chiffrer vos données pour les utiliser dans AWS Clean Rooms, consultez [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

 Note

Vérifiez soigneusement ces configurations avant de passer à l'étape suivante. Après avoir créé la collaboration, vous pouvez uniquement modifier le nom et la description de la collaboration et indiquer si les journaux sont stockés dans Amazon CloudWatch Logs.

- h. Si vous souhaitez activer les balises pour la ressource de collaboration, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
 - i. Choisissez Suivant.
5. Pour l'étape 2 : Spécifier les capacités des membres, procédez comme suit :
 - a. Pour l'analyse à l'aide de requêtes et de tâches, sous Types d'analyse pris en charge, cochez la case Tâches.

La case Requête est cochée par défaut.

- i. Sélectionnez le membre autorisé à exécuter des requêtes et des tâches dans la liste déroulante.
- ii. Sélectionnez le membre qui peut recevoir les résultats des analyses dans la liste déroulante.

 Note

Le membre qui crée le modèle PySpark d'analyse doit également être celui qui reçoit les résultats.

- b. Si vous utilisez Clean Rooms ML, pour la modélisation du ML à l'aide de flux de travail spécialement conçus,
 - i. (Facultatif) Sélectionnez le membre qui peut recevoir les résultats des modèles entraînés dans la liste déroulante.
 - ii. (Facultatif) Sélectionnez le membre qui peut recevoir le résultat de l'inférence du modèle dans la liste déroulante.
 - c. Consultez les capacités des membres sous Résolution d'identification en utilisant Résolution des entités AWS.
 - d. Choisissez Suivant.
6. Pour l'étape 3 : configurer le paiement,
- a. Pour une analyse utilisant des requêtes et des tâches, choisissez le membre qui paiera pour les requêtes et les tâches.

Vous pouvez désigner le membre habilité à exécuter des requêtes et des tâches comme étant le membre qui paie les requêtes et les tâches calculent les coûts.

Vous pouvez désigner un autre membre pour payer les requêtes et les tâches calculent les coûts.

- b. Pour la modélisation ML à l'aide de flux de travail spécialement conçus, le créateur du modèle de similarité configuré est le membre qui paiera pour la modélisation de similarité.
- c. Pour la résolution des identifiants avec Résolution des entités AWS, le créateur de la table de mappage des identifiants est le membre qui paiera pour la table de mappage des identifiants.
- d. Choisissez Suivant.

7. Pour l'étape 4 : Configuration de l'adhésion, choisissez l'une des options suivantes :

Yes, join by creating membership now

1. Pour les paramètres de résultats par défaut, pour les paramètres de résultats de requête, si vous êtes le membre autorisé à recevoir les résultats,
 - a. Cochez la case Définir les paramètres par défaut pour les requêtes. Pour la destination des résultats dans Amazon S3, entrez la destination Amazon S3 ou choisissez Browse S3 pour sélectionner un compartiment S3.
 - b. Pour le format du résultat de la requête, choisissez CSV ou PARQUET.
 - c. (Spark uniquement) Pour les fichiers de résultats, choisissez Multiple ou Single.
 - d. (Facultatif) Pour accéder au service, si vous souhaitez envoyer des requêtes qui prennent jusqu'à 24 heures à votre destination S3, cochez la case Ajouter un rôle de service pour prendre en charge les requêtes dont le traitement prend jusqu'à 24 heures.

Les requêtes volumineuses dont le traitement prend jusqu'à 24 heures seront livrées à votre destination S3.

Si vous ne cochez pas cette case, seules les requêtes traitées dans les 12 heures seront livrées à votre agence S3.

- e. Spécifiez les autorisations d'accès au service en sélectionnant Créer et utiliser un nouveau rôle de service ou Utiliser un rôle de service existant.

Si tu choisis de...	Alors...
Création et utilisation d'un nouveau rôle de service	• AWS Clean Rooms crée un rôle de service avec la politique requise pour cette table. • Le nom du rôle de service par défaut est <code>cleanrooms-result-receiver-<timestamp></code> • Vous devez disposer des autorisations nécessaires pour créer des rôles et associer des politiques.

Si tu choisis de...	Alors...
Utiliser un rôle de service existant	i. Choisissez un nom de rôle de service existant dans la liste déroulante. La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles. Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser. ii. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM. S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible. Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

 Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).
- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.

- Si vous ne parvenez pas à modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique pour le rôle de service est introuvable.

2. Pour les résultats du Job,

Exemple

Par exemple : **s3://bucket/prefix**

- a. Cochez la case Définir les paramètres par défaut pour les tâches, puis spécifiez la destination des résultats dans Amazon S3 en saisissant la destination S3 ou choisissez Parcourir S3 pour effectuer une sélection dans la liste des compartiments S3 disponibles.
 - b. Spécifiez les autorisations d'accès au service en choisissant un nom de rôle de service existant dans la liste déroulante.
3. Pour les paramètres des journaux, choisissez l'une des options suivantes pour le stockage des CloudWatch journaux dans Amazon Logs :

Note

La section Paramètres des journaux apparaît si vous avez choisi d'activer la journalisation des requêtes.

- a. Choisissez Activer et les journaux de requêtes qui vous concernent seront stockés dans votre compte Amazon CloudWatch Logs.

Chaque membre ne peut recevoir que les journaux des requêtes qu'il a initiées ou qui contiennent ses données.

Le membre qui peut recevoir les résultats reçoit également des journaux pour toutes les requêtes exécutées dans le cadre d'une collaboration, même si ses données ne sont pas accessibles dans le cadre d'une requête.

Sous Types de journaux pris en charge, choisissez parmi les types de journaux que le créateur de la collaboration a choisi de prendre en charge :

Sous Types de journaux pris en charge, les cases à cocher Query Logs et Job logs sont activées par défaut.

 Note

Après avoir activé la journalisation d'analyse, la configuration du stockage des journaux et le début de la réception des journaux dans Amazon CloudWatch Logs peuvent prendre quelques minutes. Pendant cette brève période, le membre autorisé à effectuer des requêtes peut exécuter des requêtes qui n'envoient pas réellement de journaux.

- b. Choisissez Désactiver et les journaux de requêtes qui vous concernent ne seront pas stockés dans votre compte Amazon CloudWatch Logs.
4. Si vous souhaitez activer les balises d'adhésion pour la ressource d'adhésion, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
5. Si vous êtes le membre qui paie pour Query Compute, Job Compute, ou les deux, indiquez votre acceptation en cochant la case J'accepte de payer les coûts de calcul dans le cadre de cette collaboration.

 Note

Vous devez cocher cette case pour continuer.

Pour plus d'informations sur le mode de calcul des prix, consultez [Tarification pour AWS Clean Rooms](#).

Si vous êtes le [membre qui paie les frais de calcul des requêtes, mais que vous n'êtes pas le membre habilité AWS Budgets à effectuer des requêtes](#), il est recommandé de configurer un budget AWS Clean Rooms et de recevoir des notifications une fois le budget maximum atteint. Pour plus d'informations sur la configuration d'un budget, consultez [la section Gérer vos coûts AWS Budgets](#) dans le Guide de AWS Cost Management l'utilisateur. Pour plus d'informations sur la configuration des notifications, consultez la rubrique [Création d'un compte Amazon SNS pour les notifications budgétaires](#) dans le guide de l'AWS Cost Management utilisateur. Si le budget maximum est atteint, vous pouvez contacter le membre qui pourra lancer des requêtes ou [quitter la collaboration](#). Si

vous quittez la collaboration, aucune autre requête ne sera autorisée à être exécutée et, par conséquent, les frais de calcul des requêtes ne vous seront plus facturés.

6. Choisissez Suivant.

La collaboration et votre adhésion sont créées.

Votre statut dans la collaboration est actif.

No, I will create a membership later

1. Choisissez Suivant.

Seule la collaboration est créée.

Votre statut dans la collaboration est inactif.

8. Pour l'étape 5 : révision et création, procédez comme suit :

- a. Passez en revue les sélections que vous avez effectuées lors des étapes précédentes et modifiez-les si nécessaire.
- b. Choisissez l'une des options.

Si vous avez choisi de...	Ensuite, choisissez...
Créez un abonnement avec la collaboration (Oui, inscrivez-vous en créant un abonnement maintenant)	Créez une collaboration et une adhésion
Créez la collaboration et ne créez pas d'adhésion pour le moment (Non, je créerai un abonnement plus tard)	Créez une collaboration

Une fois que votre collaboration a été créée avec succès, vous pouvez voir la page des détails de la collaboration sous Collaborations.

Vous êtes maintenant prêt à :

- [Préparez votre tableau de données à analyser AWS Clean Rooms](#). (Facultatif si vous souhaitez analyser vos propres données d'événement ou si vous souhaitez interroger des données d'identité.)
- [Associez le tableau configuré à votre collaboration](#). (Facultatif si vous souhaitez analyser vos propres données d'événement.)
- [Ajoutez une règle d'analyse pour la table configurée](#). (Facultatif si vous souhaitez analyser vos propres données d'événement.)
- [Créez un abonnement et rejoignez une collaboration](#). (Facultatif si vous avez déjà créé un abonnement.)
- [Invitez des membres à rejoindre la collaboration](#).

Création d'une collaboration pour la modélisation ML

Dans cette procédure, en tant que [créateur de la collaboration](#), vous effectuez les tâches suivantes :

- [Créez une collaboration](#).
- Invitez un ou plusieurs [membres](#) à rejoindre la [collaboration](#).
- Attribuez des capacités aux membres, telles que
 - [Membre pouvant poser des questions](#)
 - [Membre pouvant recevoir les résultats](#)
 - Membre pouvant recevoir les résultats de modèles entraînés
 - Membre autorisé à générer des résultats à partir de l'inférence du modèle

Si le créateur de la collaboration est également le membre habilité à recevoir les résultats, il spécifie la destination et le format des résultats. Ils fournissent également un rôle de service Amazon Resource Name (ARN) pour écrire les résultats dans la destination des résultats.

- Configurez quel [membre est responsable du paiement des coûts de calcul, de formation des modèles et des coûts d'inférence des modèles dans le cadre de la collaboration](#).

Avant de commencer, assurez-vous de remplir les prérequis suivants :

- Vous avez le nom et l' ID Compte AWS identifiant de chaque membre que vous souhaitez inviter à rejoindre la collaboration.
- Vous êtes autorisé à partager le nom et l' ID Compte AWS identifiant de chaque membre avec tous les membres de la collaboration.

 Note

Vous ne pouvez pas ajouter d'autres membres après avoir créé la collaboration.

Pour plus d'informations sur la création d'une collaboration à l'aide du AWS SDKs, consultez la [référence des AWS Clean Rooms API](#).

Pour créer une collaboration pour la modélisation ML

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Dans le coin supérieur droit, choisissez Créer une collaboration.
4. Pour l'étape 1 : définir la collaboration, procédez comme suit :
 - a. Pour plus de détails, entrez le nom et la description de la collaboration.

Ces informations seront visibles par les membres de la collaboration qui sont invités à participer à la collaboration. Le nom et la description les aident à comprendre à quoi fait référence la collaboration.

- b. Pour les membres :
 - i. Pour le membre 1 : vous devez saisir le nom d'affichage de votre membre tel que vous souhaitez qu'il apparaisse pour la collaboration.

 Note

Votre Compte AWS identifiant est automatiquement inclus comme Compte AWS identifiant de membre.

- ii. Pour Membre 2, entrez le nom d'affichage du membre et l' Compte AWS ID du membre que vous souhaitez inviter à rejoindre la collaboration.

Le nom d'affichage et l' Compte AWS identifiant du membre seront visibles par toutes les personnes invitées à la collaboration. Une fois que vous avez saisi et enregistré les valeurs de ces champs, vous ne pouvez pas les modifier.

 Note

Vous devez informer le membre de la collaboration que son Compte AWS identifiant de membre et son nom d'affichage seront visibles par tous les collaborateurs invités et actifs de la collaboration.

- iii. Si vous souhaitez ajouter un autre membre, choisissez Ajouter un autre membre. Entrez ensuite le nom d'affichage du membre et l'Identifiant de membre pour chaque membre susceptible de fournir les données que vous souhaitez inviter à la collaboration.
- c. Si vous souhaitez activer la journalisation des analyses, cochez la case Activer la journalisation des analyses, puis sous Types de journaux pris en charge, choisissez Journaux issus des requêtes.
- d. Si vous souhaitez activer la surveillance détaillée, cochez la case Activer la surveillance détaillée.

Le responsable de l'analyse et le payeur configuré peuvent choisir d'activer des métriques détaillées lors de la création de leur adhésion. Lorsque cette option est activée, des mesures de surveillance détaillées seront publiées CloudWatch pour le suivi opérationnel des collaborations, notamment les performances des requêtes et l'utilisation des ressources. Ces statistiques seront mises à la disposition du responsable de l'analyse et du payeur configuré dans leurs versions respectives Compte AWS.

Pour plus d'informations sur la CloudWatch tarification, consultez la section [CloudWatch Tarification](#).

- e. Sous Régions de résultats de requête autorisées, sélectionnez un ou plusieurs Régions AWS endroits où vous souhaitez envoyer les résultats de la requête.

Par défaut, seule la région actuelle (telle que North Virginia us-east-1) est sélectionnée.

 Important

Lorsque vous activez la livraison des résultats de requêtes entre régions, vos résultats peuvent être traités et stockés en dehors de la région source.

Pour plus d'informations sur les régions, voir [Régions et points de terminaison](#) dans le Références générales AWS.

f. (Facultatif) Gérez l'accès à vos données grâce à l'approbation automatique des demandes de modification en configurant les paramètres qui peuvent être modifiés automatiquement sans approbation manuelle pour une demande de modification. Par défaut, certains paramètres ne peuvent être modifiés qu'en soumettant une demande de modification qui doit être approuvée par tous les membres avant qu'elle ne prenne effet.

- Accorder des capacités aux membres : choisissez les capacités qui peuvent être accordées aux membres de la collaboration sans approbation manuelle. Les membres peuvent toujours fournir des données.
- Choisissez des capacités :
 - Contribuer aux données (toujours activé)
 - Recevez les résultats
- Approuver automatiquement les nouveaux membres dotés de ces capacités : si cela est autorisé, tous les membres ajoutés avec les capacités sélectionnées ci-dessus rejoindront instantanément la collaboration. Les membres ajoutés avec d'autres capacités devront toujours être approuvés manuellement pour s'inscrire.
- Capacités pouvant être révoquées automatiquement - Choisissez les capacités qui peuvent être révoquées sans approbation manuelle. Les membres peuvent toujours fournir des données.
- Choisissez des capacités :
 - Contribuer aux données (toujours activé)
 - Recevez les résultats

Si vous choisissez cette option, vous pouvez suivre toutes les modifications de configuration de collaboration via l'historique des demandes de modification, situé dans l'onglet Détails de la page des détails de la collaboration.

g. (Facultatif) Si vous souhaitez activer la fonctionnalité de calcul cryptographique, cochez la case Activer le calcul cryptographique.

- i. Choisissez les paramètres de couverture cryptographique suivants :
- Autoriser plaintext les colonnes

Choisissez Non si vous avez besoin de tables entièrement chiffrées.

Choisissez Oui si vous souhaitez que cleartext les colonnes soient autorisées dans la table cryptée.

Pour fonctionner SUM ou AVG sur certaines colonnes, les colonnes doivent être insérées cleartext.

- Préservez NULL les valeurs

Choisissez Non si vous ne souhaitez pas conserver NULL les valeurs. NULL les valeurs n'apparaîtront pas comme NULL dans une table cryptée.

Choisissez Oui si vous souhaitez conserver NULL les valeurs. NULL les valeurs apparaîtront comme NULL dans une table cryptée.

ii. Choisissez les paramètres d'empreinte suivants :

- Autoriser les doublons

Choisissez Non si vous ne souhaitez pas que les doublons soient autorisés dans une fingerprint colonne.

Choisissez Oui si vous souhaitez que les entrées dupliquées soient autorisées dans une fingerprint colonne.

- JOIN Autorisation de colonnes portant des noms différents

Choisissez Non si vous ne souhaitez pas joindre des fingerprint colonnes portant des noms différents.

Choisissez Oui si vous souhaitez joindre des fingerprint colonnes portant des noms différents.

Pour plus d'informations sur les paramètres de calcul cryptographique, consultez [Paramètres de calcul cryptographique](#).

Pour plus d'informations sur la façon de chiffrer vos données pour les utiliser dans AWS Clean Rooms, consultez [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

 Note

Vérifiez soigneusement ces configurations avant de passer à l'étape suivante. Après avoir créé la collaboration, vous pouvez uniquement modifier le nom et la description de la collaboration et indiquer si les journaux sont stockés dans Amazon CloudWatch Logs.

- Si vous souhaitez activer les balises pour la ressource de collaboration, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
 - Choisissez Suivant.
- Pour l'étape 2 : Spécifier les capacités des membres,
 - Pour l'analyse à l'aide de requêtes et de tâches, dans la section Types d'analyse pris en charge, laissez la case Requête cochée.
 - Pour Exécuter les requêtes, choisissez le membre qui lancera la formation du modèle
 - Pour Recevoir les résultats des analyses, choisissez un ou plusieurs membres qui recevront les résultats de la requête.
 - Pour la modélisation ML à l'aide de flux de travail spécialement conçus,
 - Pour Recevoir les résultats des modèles entraînés, choisissez le membre qui recevra les résultats des modèles entraînés, y compris les artefacts et les métriques du modèle.
 - Pour Recevoir la sortie de l'inférence du modèle, choisissez le membre qui recevra les résultats de l'inférence du modèle.
 - Consultez les capacités des membres sous Résolution d'identification en utilisant Résolution des entités AWS.
- Pour l'étape 3 : configurer le paiement,
 - Sous Analyse à l'aide de requêtes, pour Pay for queries, effectuez l'une des actions suivantes :
 - Pour que le même membre paie et exécute les requêtes, sélectionnez le même membre que vous avez choisi pour Exécuter les requêtes.
 - Pour qu'un autre membre paie les frais de requête, sélectionnez votre compte de membre.
 - Pour la modélisation ML à l'aide de flux de travail spécialement conçus,

- Choisissez le membre qui paiera pour la formation des modèles.
 - c. Choisissez le membre qui paiera pour le travail d'inférence.
 - d. Pour la modélisation Pay for lookalike, aucune action n'est nécessaire. Le créateur du modèle de similarité configuré est le membre qui paiera pour la modélisation de similarité.
 - e. (Facultatif) Choisissez le membre qui paiera pour la génération de données synthétiques.
 - f. Pour la résolution de l'ID avec Résolution des entités AWS, aucune action n'est nécessaire. Le créateur de la table de mappage d'identifiants est le membre qui paiera pour la table de mappage d'identifiants.
7. Choisissez Suivant.
8. Pour l'étape 4 : Configurer l'adhésion, sous Abonnement à la collaboration, choisissez l'une des options suivantes :

Yes, join by creating membership now

1. Pour les paramètres de résultats par défaut, pour les paramètres de résultats de requête, si vous êtes le membre autorisé à recevoir les résultats,
 - a. Cochez la case Définir les paramètres par défaut pour les requêtes.
 - b. Pour la destination des résultats dans Amazon S3, entrez la destination Amazon S3 ou choisissez Browse S3 pour sélectionner un compartiment S3.
 - c. Pour le format du résultat de la requête, choisissez CSV ou PARQUET.
 - d. (Spark uniquement) Pour les fichiers de résultats, choisissez Multiple ou Single.
 - e. (Facultatif) Si vous souhaitez envoyer des requêtes qui prennent jusqu'à 24 heures à votre destination S3, cochez la case Ajouter un rôle de service pour prendre en charge les requêtes dont le traitement prend jusqu'à 24 heures.

Les requêtes volumineuses dont le traitement prend jusqu'à 24 heures seront livrées à votre destination S3.

Si vous ne cochez pas cette case, seules les requêtes traitées dans les 12 heures seront livrées à votre agence S3.

- f. Spécifiez les autorisations d'accès au service en sélectionnant Créer et utiliser un nouveau rôle de service ou Utiliser un rôle de service existant.

Si tu choisis de...	Alors...
Création et utilisation d'un nouveau rôle de service	• AWS Clean Rooms crée un rôle de service avec la politique requise pour cette table. • Le nom du rôle de service par défaut est <code>cleanrooms-result-receiver-<timestamp></code> • Vous devez disposer des autorisations nécessaires pour créer des rôles et associer des politiques.
Utiliser un rôle de service existant	i. Choisissez un nom de rôle de service existant dans la liste déroulante. La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles. Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser. ii. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM. S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible. Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

 Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).
- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.
- Si vous ne parvenez pas à modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique pour le rôle de service est introuvable.

2. Pour la configuration ML,
 - a. Cochez la case Create ML configuration, puis spécifiez la destination de sortie du modèle sur Amazon S3 en saisissant la destination S3 ou choisissez Browse S3 pour effectuer une sélection dans la liste des compartiments S3 disponibles.
 - b. Spécifiez les autorisations d'accès au service en choisissant de créer et d'utiliser un nouveau rôle de service ou d'utiliser un rôle de service existant.
 - c. Si le compartiment S3 est chiffré, cochez la case Le compartiment de destination est chiffré avec une clé KMS, puis entrez AWS KMS key ou sélectionnez Créer un AWS KMS key pour créer une nouvelle clé KMS.
3. Si vous souhaitez activer les balises d'adhésion pour la ressource d'adhésion, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
4. Si vous êtes le membre qui paie pour Query Compute, indiquez votre acceptation en cochant la case J'accepte de payer les coûts de calcul dans le cadre de cette collaboration.

 Note

Vous devez cocher cette case pour continuer.

Pour plus d'informations sur le mode de calcul des prix, consultez [Tarification pour AWS Clean Rooms](#).

Si vous êtes le [membre qui paie les frais de calcul des requêtes, mais que vous n'êtes pas le membre habilité AWS Budgets à effectuer des requêtes](#), il est recommandé de configurer un budget AWS Clean Rooms et de recevoir des notifications une fois le budget maximum atteint. Pour plus d'informations sur la configuration d'un budget, consultez [la section Gérer vos coûts AWS Budgets](#) dans le Guide de AWS Cost Management l'utilisateur. Pour plus d'informations sur la configuration des notifications, consultez la rubrique [Création d'un compte Amazon SNS pour les notifications budgétaires](#) dans le guide de l'AWS Cost Management utilisateur. Si le budget maximum est atteint, vous pouvez contacter le membre qui pourra lancer des requêtes ou [quitter la collaboration](#). Si vous quittez la collaboration, aucune autre requête ne sera autorisée à être exécutée et, par conséquent, les frais de calcul des requêtes ne vous seront plus facturés.

5. Choisissez Suivant.

La collaboration et votre adhésion sont créées.

Votre statut dans la collaboration est actif.

No, I will create a membership later

1. Choisissez Suivant.

Seule la collaboration est créée.

Votre statut dans la collaboration est inactif.

9. Pour l'étape 5 : révision et création, procédez comme suit :

- a. Passez en revue les sélections que vous avez effectuées lors des étapes précédentes et modifiez-les si nécessaire.
- b. Choisissez l'une des options.

Si vous avez choisi de...	Ensuite, choisissez...
Créez un abonnement avec la collaboration (Oui, inscrivez-vous en créant un abonnement maintenant)	Créez une collaboration et une adhésion

Si vous avez choisi de...	Ensuite, choisissez...
Créez la collaboration et ne créez pas d'adhésion pour le moment (Non, je créerai un abonnement plus tard)	Créez une collaboration

Création d'un abonnement et adhésion à une collaboration

Un abonnement est une ressource créée lorsqu'un membre rejoint une collaboration dans AWS Clean Rooms.

Vous pouvez rejoindre une collaboration en tant que

- [membre qui peut interroger](#)
- [membre capable d'exécuter des requêtes et des tâches](#)
- [membre qui peut recevoir les résultats](#) d'une requête ou d'un travail
- [membre payant les frais de calcul des requêtes](#)
- [membre payant pour des requêtes et des emplois](#)

Tous les membres peuvent fournir des données.

Pour plus d'informations sur la façon de créer un abonnement et de rejoindre une collaboration à l'aide du AWS SDKs, consultez la [référence des AWS Clean Rooms API](#).

Dans cette procédure, le membre invité [rejoint la collaboration en créant une ressource d'adhésion](#).

Si le membre invité est celui qui peut recevoir les résultats, il spécifie la destination et le format des résultats. Ils fournissent également un rôle de service ARN pour écrire dans la destination des résultats.

Si le membre invité est le membre chargé de payer les frais de calcul, il accepte ses responsabilités de paiement avant de rejoindre la collaboration.

Pour créer un abonnement et rejoindre une collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre membre Compte AWS.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.

3. Dans l'onglet Disponible pour rejoindre, pour les collaborations disponibles, choisissez le nom de la collaboration.
4. Sur la page des détails de la collaboration, dans la section Vue d'ensemble, consultez les détails de la collaboration, y compris les détails de votre membre et la liste des autres membres.

Vérifiez que Compte AWS IDs les membres de chaque membre de la collaboration sont ceux avec lesquels vous avez l'intention de participer à la collaboration.

5. Choisissez Créer un abonnement.
6. Sur la page Créer un abonnement, dans l'aperçu, consultez le nom de la collaboration, la description de la collaboration, l' Compte AWS identifiant du créateur de la collaboration, les détails de votre membre et l' Compte AWS identifiant du membre qui paiera pour les requêtes.
7. Si le créateur de la collaboration a choisi d'activer la journalisation des analyses, choisissez l'une des options suivantes pour le stockage des CloudWatch journaux dans Amazon Logs :

Si vous choisissez...	Alors...
Allumez	Les journaux qui vous concernent sont stockés dans Amazon CloudWatch Logs. Chaque membre ne peut recevoir que les journaux des requêtes qu'il a initiées ou qui contiennent ses données. Le membre qui peut recevoir les résultats reçoit également des journaux pour toutes les analyses effectuées dans le cadre d'une collaboration, même si ses données ne sont pas accessibles dans le cadre d'une analyse. Sous Types de journaux pris en charge, choisissez parmi les types de journaux que le créateur de la collaboration a choisi de prendre en charge : 1. Si vous souhaitez recevoir les journaux générés à partir de requêtes SQL, cochez la case Journaux à partir de requêtes.

Si vous choisissez...	Alors...
	2. Si vous souhaitez recevoir les journaux générés à partir des tâches à l'aide de PySpark, cochez la case Journaux des tâches.
Éteindre	Les journaux de requêtes qui vous concernent ne sont pas stockés dans votre compte Amazon CloudWatch Logs.

 Note

Après avoir activé la journalisation d'analyse, la configuration du stockage des journaux et le début de la réception des journaux dans Amazon CloudWatch Logs peuvent prendre quelques minutes. Pendant cette brève période, le membre autorisé à effectuer des requêtes peut exécuter des requêtes qui n'envoient pas réellement de journaux.

- Si le créateur de la collaboration a activé le suivi détaillé pour cette collaboration, choisissez si vous souhaitez recevoir des statistiques d'observabilité détaillées dans votre CloudWatch compte.

Pour une surveillance détaillée :

Options de surveillance détaillées

Option	Description
Allumez	AWS Clean Rooms publiera des statistiques de surveillance détaillées dans CloudWatch pour cette collaboration sur votre compte. Vous pouvez utiliser ces mesures pour la surveillance opérationnelle, notamment les performances des requêtes et l'utilisation des ressources. Des CloudWatch frais supplémentaires s'appliquent. Pour plus d'informations, consultez Tarification d'CloudWatch.
Éteindre	Aucune métrique détaillée ne sera exportée vers votre CloudWatch compte. Les autres membres peuvent toujours consulter les statistiq

Option	Description
	ues de surveillance détaillées dans leurs propres comptes s'ils ont activé cette option.

 Note

Les mesures de suivi détaillées ne sont disponibles que pour les membres qui peuvent exécuter des requêtes (analyseurs) et pour les membres configurés comme payeurs pour la collaboration.

9. Si les capacités de votre membre incluent Recevoir des résultats, les paramètres par défaut des résultats sont les suivants :
 - a. Pour les résultats des requêtes, cochez la case Définir les paramètres par défaut pour les requêtes, puis spécifiez la destination des résultats dans Amazon S3 en saisissant la destination S3 ou choisissez Parcourir S3 pour effectuer une sélection dans la liste des compartiments S3 disponibles.

Exemple

Par exemple : **s3://bucket/prefix**

- i. Pour le format du résultat, choisissez CSV ou PARQUET.
- ii. (Spark uniquement) Pour les fichiers de résultats, choisissez Multiple ou Single.
- iii. (Facultatif) Pour accéder au service, si vous souhaitez envoyer des requêtes qui prennent jusqu'à 24 heures à votre destination S3, cochez la case Ajouter un rôle de service pour prendre en charge les requêtes dont le traitement prend jusqu'à 24 heures.

Les requêtes volumineuses dont le traitement prend jusqu'à 24 heures seront livrées à votre destination S3.

Si vous ne cochez pas cette case, seules les requêtes traitées dans les 12 heures seront livrées à votre site S3.

 Note

Vous devez sélectionner un rôle de service existant ou être autorisé à en créer un nouveau. Pour de plus amples informations, veuillez consulter [Créez un rôle de service pour recevoir des résultats](#).

- iv. Spécifiez les autorisations d'accès au service en sélectionnant Créer et utiliser un nouveau rôle de service ou Utiliser un rôle de service existant.

Create and use a new service role

- AWS Clean Rooms crée un rôle de service avec la politique requise pour cette table.
- Le nom du rôle de service par défaut est `cleanrooms-result-receiver-<timestamp>`
- Vous devez disposer des autorisations nécessaires pour créer des rôles et associer des politiques.

Use an existing service role

1. Choisissez le nom d'un rôle de service existant dans la liste déroulante.

La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles.

Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser.

2. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM.

S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible.

Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

 Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).
- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.
- Si vous ne pouvez pas modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique du rôle de service est introuvable.

- b. Pour les résultats des tâches, cochez la case Définir les paramètres par défaut pour les tâches, puis spécifiez la destination des résultats dans Amazon S3 en saisissant la destination S3 ou choisissez Parcourir S3 pour effectuer une sélection dans la liste des compartiments S3 disponibles.

Exemple

Par exemple : **s3://bucket/prefix**

- Spécifiez les autorisations d'accès au service en choisissant un nom de rôle de service existant dans la liste déroulante.
10. Si vous souhaitez activer les balises pour la ressource d'adhésion, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
11. Si le créateur de la collaboration vous a désigné comme membre chargé de payer pour les requêtes ou de payer pour les requêtes et les travaux, indiquez votre acceptation en cochant la case J'accepte de payer les frais de calcul dans cette collaboration.

 Note

Vous devez cocher cette case pour continuer.

Pour plus d'informations sur le mode de calcul des prix, consultez [Tarification pour AWS Clean Rooms](#).

Si vous êtes le [membre qui paie les frais de calcul des requêtes](#) ou le [membre qui paie les requêtes et les coûts de calcul des tâches, mais que vous n'êtes pas le membre habilité AWS Budgets à effectuer des requêtes](#), il est recommandé de configurer un budget pour AWS Clean Rooms et de recevoir des notifications une fois le budget maximum atteint. Pour plus d'informations sur la configuration d'un budget, consultez [la section Gérer vos coûts AWS Budgets](#) dans le Guide de AWS Cost Management l'utilisateur. Pour plus d'informations sur la configuration des notifications, consultez la rubrique [Création d'un compte Amazon SNS pour les notifications budgétaires](#) dans le guide de l'AWS Cost Management utilisateur. Si le budget maximum a été atteint, vous pouvez contacter le membre qui peut exécuter des requêtes et des tâches ou [quitter la collaboration](#). Si vous quittez la collaboration, aucune autre requête ne sera autorisée à être exécutée et, par conséquent, les frais de calcul des requêtes ne vous seront plus facturés.

12. Si vous êtes sûr de vouloir créer un abonnement et rejoindre la collaboration, choisissez Créer un abonnement.

Vous disposez d'un accès en lecture aux métadonnées de collaboration. Cela inclut des informations telles que le nom d'affichage et la description de la collaboration, en plus de tous les noms et Compte AWS IDs des autres membres.

Vous êtes maintenant prêt à :

- [Préparez votre tableau de données à interroger. AWS Clean Rooms](#) (Facultatif si vous souhaitez interroger vos propres données d'événement ou si vous souhaitez interroger des données d'identité.)
- [Associez la table configurée à votre collaboration](#), si vous souhaitez interroger les données d'un événement.
- [Ajoutez une règle d'analyse pour la table configurée](#), si vous souhaitez interroger les données d'événements.
- [Créez et associez un nouvel espace de noms d'identification](#), si vous souhaitez créer une table de mappage d'identifiants pour interroger les données d'identité.

Pour plus d'informations sur la façon de quitter une collaboration, consultez [Quitter une collaboration](#).

Collaborations d'édition

En tant que créateur de collaboration, vous pouvez modifier les différentes parties d'une collaboration.

Pour plus d'informations sur la façon de modifier une collaboration à l'aide d'AWS SDKs, consultez le document de [référence de l'API AWS Clean Rooms](#).

Rubriques

- [Modifier le nom et la description de la collaboration](#)
- [Mettre à jour le moteur d'analyse de collaboration](#)
- [Désactiver le stockage des journaux](#)
- [Modifier les paramètres des journaux de collaboration](#)
- [Modifier les balises de collaboration](#)
- [Modifier les tags d'adhésion](#)
- [Ajouter un nouveau membre](#)
- [Modifier les capacités des membres existants](#)
- [Modifier les paramètres d'approbation automatique de la collaboration](#)
- [Modifier les balises de table associées](#)
- [Modifier les balises du modèle d'analyse](#)
- [Modifier les balises de politique de confidentialité différentielles](#)

Modifier le nom et la description de la collaboration

Après avoir créé la collaboration, vous ne pouvez modifier que le nom et la description de la collaboration.

Pour modifier le nom et la description de la collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Sur la page détaillée de la collaboration, choisissez Actions, puis Modifier la collaboration.

5. Sur la page Modifier la collaboration, dans Détails, modifiez le nom et la description de la collaboration.
6. Sélectionnez Enregistrer les modifications.

Mettre à jour le moteur d'analyse de collaboration

Après avoir créé la collaboration, vous pouvez remplacer le moteur d'analyse AWS Clean Rooms SQL par Spark.

Note

Le passage du moteur d'analyse de AWS Clean Rooms SQL à Spark peut perturber les flux de travail existants.

Pour mettre à jour le moteur d'analyse de collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Sur la page détaillée de la collaboration, choisissez Actions, puis Modifier la collaboration.
5. Sélectionnez Enregistrer les modifications.

Désactiver le stockage des journaux

Si vous avez activé la journalisation des analyses, vous pouvez modifier si les journaux d'analyse sont stockés dans votre compte Amazon CloudWatch Logs.

Pour désactiver le stockage des journaux

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration pour laquelle la journalisation des analyses est activée.

4. Sur la page détaillée de la collaboration, choisissez Actions, puis sélectionnez Désactiver le stockage des journaux.

 Note

Un avertissement apparaît, indiquant ce qui suit :

- Les nouvelles requêtes ne seront plus enregistrées dans votre CloudWatch compte.
- Les journaux existants seront conservés conformément à vos paramètres de conservation actuels.
- Si vous réactivez la connexion à l'avenir, elle ne s'appliquera qu'aux requêtes effectuées après la réactivation.
- Cette modification n'affecte que vos journaux. Les paramètres de journalisation des autres membres de l'équipe restent inchangés.

5. Choisissez Désactiver.

Modifier les paramètres des journaux de collaboration

Si vous avez activé la journalisation des requêtes, vous pouvez modifier si les journaux des requêtes sont stockés dans votre compte Amazon CloudWatch Logs.

Pour modifier les paramètres des journaux de collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Sur la page détaillée de la collaboration, effectuez l'une des opérations suivantes :
 - Choisissez Actions, puis sélectionnez Modifier les paramètres des journaux.
 - Dans l'onglet Journaux, choisissez Modifier les paramètres des journaux.
5. Dans le modal Modifier les paramètres des journaux, pour le stockage des CloudWatch journaux dans Amazon Logs :
 - Si vous ne souhaitez pas que les journaux qui vous concernent soient stockés dans votre compte Amazon CloudWatch Logs, choisissez Turn off.

- Si vous souhaitez que les journaux qui vous concernent soient stockés dans votre compte Amazon CloudWatch Logs, choisissez Turn on.

Vous ne pouvez recevoir des journaux que pour les requêtes que vous avez initiées ou qui contiennent des données vous concernant.

Le membre qui peut recevoir les résultats reçoit également des journaux pour toutes les requêtes exécutées dans le cadre d'une collaboration, même si ses données ne sont pas accessibles dans le cadre d'une requête.

1. Sous Types de journaux pris en charge, choisissez parmi les types de journaux que le créateur de la collaboration a choisi de prendre en charge :
 - Si vous souhaitez recevoir les journaux générés à partir de requêtes SQL, cochez la case Journaux à partir de requêtes.
 - Si vous souhaitez recevoir les journaux générés à partir des tâches à l'aide de PySpark, cochez la case Journaux des tâches.

6. Sélectionnez Enregistrer les modifications.

Note

Après avoir activé la journalisation, la configuration du stockage des journaux et le début de la réception des journaux dans Amazon CloudWatch Logs peuvent prendre quelques minutes. Pendant cette brève période, le membre autorisé à effectuer des requêtes peut exécuter des requêtes qui n'envoient pas réellement de journaux.

Modifier les balises de collaboration

En tant que créateur de collaboration, après avoir créé une collaboration, vous pouvez gérer les balises de la ressource de collaboration.

Pour modifier les balises de collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.

4. Sélectionnez l'une des méthodes suivantes :

Si...	Alors...
Le créateur de la collaboration et un membre de la collaboration	Cliquez sur l'onglet Détails.
Le créateur de la collaboration mais non membre de la collaboration	Faites défiler la page vers le bas jusqu'à la section Tags.

5. Pour plus de détails sur la collaboration, choisissez Gérer les balises.
6. Sur la page de gestion des étiquettes, vous pouvez effectuer les opérations suivantes :
 - Pour supprimer une identification, choisissez Supprimer.
 - Pour ajouter une balise, sélectionnez Ajouter une nouvelle balise.
 - Pour enregistrer vos modifications, choisissez Enregistrer les modifications

Modifier les tags d'adhésion

En tant que créateur de collaboration, après avoir créé une collaboration, vous pouvez gérer les balises de la ressource d'adhésion.

Pour modifier les tags d'adhésion

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Cliquez sur l'onglet Détails.
5. Pour les détails de l'adhésion, choisissez Gérer les tags.
6. Sur la page Gérer les tags d'adhésion, vous pouvez effectuer les opérations suivantes :
 - Pour supprimer une identification, choisissez Supprimer.
 - Pour ajouter une balise, sélectionnez Ajouter une nouvelle balise.
 - Sélectionnez Enregistrer pour enregistrer les modifications.

Ajouter un nouveau membre

Pour de plus amples informations, veuillez consulter [Ajouter un nouveau membre à une collaboration](#).

Modifier les capacités des membres existants

Pour de plus amples informations, veuillez consulter [Mettre à jour les capacités des membres existants](#).

Modifier les paramètres d'approbation automatique de la collaboration

Pour de plus amples informations, veuillez consulter [Modifier les paramètres d'approbation automatique de la collaboration](#).

Modifier les balises de table associées

En tant que créateur de collaboration, après avoir associé des tables à une collaboration, vous pouvez gérer les balises de la ressource de table associée.

Pour modifier les balises de table associées

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Choisissez l'onglet Tables.
5. Pour les tables que vous avez associées, choisissez une table.
6. Sur la page détaillée du tableau configuré, pour Balises, choisissez Gérer les balises.

Sur la page de gestion des étiquettes, vous pouvez effectuer les opérations suivantes :

- Pour supprimer une identification, choisissez Supprimer.
- Pour ajouter une balise, sélectionnez Ajouter une nouvelle balise.
- Sélectionnez Enregistrer pour enregistrer les modifications.

Modifier les balises du modèle d'analyse

En tant que créateur de collaboration, après avoir créé une collaboration, vous pouvez gérer les balises de la ressource du modèle d'analyse.

Pour modifier les tags d'adhésion

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Sélectionnez l'onglet Modèles .
5. Dans la section Modèles d'analyse que vous avez créés, choisissez le modèle d'analyse.
6. Sur la page détaillée du tableau du modèle d'analyse, faites défiler la page vers le bas jusqu'à la section Tags.
7. Choisissez Gérer les balises.
8. Sur la page de gestion des étiquettes, vous pouvez effectuer les opérations suivantes :
 - Pour supprimer une identification, choisissez Supprimer.
 - Pour ajouter une balise, sélectionnez Ajouter une nouvelle balise.
 - Sélectionnez Enregistrer pour enregistrer les modifications.

Modifier les balises de politique de confidentialité différentielles

En tant que créateur de collaboration, après avoir créé une collaboration, vous pouvez gérer les balises de la ressource du modèle d'analyse.

Pour modifier les tags d'adhésion

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration qui contient la politique de confidentialité différentielle que vous souhaitez modifier.
4. Choisissez l'onglet Tables.

5. Dans l'onglet Tables, sélectionnez Gérer les balises.
6. Sur la page de gestion des étiquettes, vous pouvez effectuer les opérations suivantes :
 - Pour supprimer une identification, choisissez Supprimer.
 - Pour ajouter une balise, sélectionnez Ajouter une nouvelle balise.
 - Sélectionnez Enregistrer pour enregistrer les modifications.

Demandes de modification dans AWS Clean Rooms

Les demandes de modification vous permettent de proposer des modifications aux paramètres de collaboration existants pour approbation par les autres membres de la collaboration. Avec les demandes de modification, vous pouvez soumettre une demande pour ajouter de nouveaux membres, mettre à jour les capacités des membres existants et modifier les paramètres d'approbation automatique de la collaboration. Tous les membres de la collaboration doivent approuver les demandes de modification pour que les modifications proposées prennent effet.

Les demandes de modification sont spécifiques à la collaboration et peuvent être soumises par le créateur de la collaboration.

Vous pouvez soumettre une demande de modification de la manière suivante :

- Ajouter un nouveau membre à une collaboration
- Mettre à jour les capacités des membres existants
- Modifier les paramètres d'approbation automatique de la collaboration

Note

Vous devez être le créateur de la collaboration pour soumettre une demande de modification.

Ajouter un nouveau membre à une collaboration

Pour ajouter un nouveau membre à une collaboration, vous devez en être le créateur. L'ajout d'un nouveau membre à une collaboration nécessite l'approbation manuelle des membres existants de la collaboration et soumettra une demande de modification.

Vous pouvez ajouter de nouveaux membres à une collaboration en suivant ces étapes :

Pour ajouter un nouveau membre à une collaboration

1. Connectez-vous à la [AWS Clean Roomsconsole AWS Management Console et ouvrez-la](#) avec votre membreCompte AWS.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sélectionnez votre collaboration pour accéder aux détails de votre collaboration.
4. Sur la page de collaboration, accédez à l'onglet Membres.
5. Dans le tableau Membres, sélectionnez Modifier les membres.
6. Choisissez Ajouter un autre membre.
7. Entrez les informations du nouveau membre :
 - Nom d'affichage du membre
 - Compte AWSID de membre
 - Spécifiez si le membre peut recevoir des résultats. Cochez la case pour accorder cette capacité à ce membre.
8. Sélectionnez Enregistrer les modifications.
9. Confirmez l'envoi de votre demande de modification. Dans le mode de confirmation, confirmez les modifications et sélectionnez Soumettre la demande de modification.

Note

Si les types de modification approuvés automatiquement sont pris en charge dans votre collaboration, l'approbation manuelle des demandes de modification peut ne pas être requise. Vous pouvez vérifier quels types de modifications ne nécessitent pas de demande de modification dans la section « Vue d'ensemble » de la collaboration. Pour de plus amples informations, veuillez consulter [the section called “Modifier les paramètres d'approbation automatique de la collaboration”](#).

Mettre à jour les capacités des membres existants

Pour mettre à jour les capacités existantes des membres de la collaboration, vous devez être le créateur de la collaboration. La mise à jour des capacités pour les membres existants de la collaboration nécessite l'approbation manuelle des membres existants et soumettra une demande de modification.

Les capacités des membres qui peuvent être mises à jour sont les suivantes :

- Peut recevoir des résultats

Vous pouvez mettre à jour les capacités des membres existants de la collaboration en procédant comme suit :

Pour mettre à jour les capacités des membres existants

1. Connectez-vous à la [AWS Clean Roomsconsole AWS Management Console et ouvrez-la](#) avec votre membreCompte AWS.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sélectionnez votre collaboration pour accéder aux détails de votre collaboration.
4. Sur la page de collaboration, accédez à l'onglet Membres.
5. Dans le tableau Membres, sélectionnez Modifier les membres.
6. Spécifiez la capacité du membre à modifier.
7. Sélectionnez Enregistrer les modifications.
8. Confirmez l'envoi de votre demande de modification. Dans le mode de confirmation, confirmez les modifications et sélectionnez Soumettre la demande de modification.

Note

Si les types de modification approuvés automatiquement sont pris en charge dans votre collaboration, l'approbation manuelle des demandes de modification peut ne pas être requise. Vous pouvez vérifier quels types de modifications ne nécessitent pas de demande de modification dans la section « Vue d'ensemble » de la collaboration. Pour de plus amples informations, veuillez consulter [the section called “Modifier les paramètres d'approbation automatique de la collaboration”](#).

Modifier les paramètres d'approbation automatique de la collaboration

Pour modifier les paramètres de collaboration pour les approbations automatiques, vous devez être le créateur de la collaboration et soumettre une demande de modification pour approbation par les autres membres de la collaboration.

Vous pouvez modifier les paramètres d'approbation automatique d'une collaboration en suivant ces étapes :

Pour modifier les paramètres d'approbation automatique de la collaboration

1. Connectez-vous à la [AWS Clean Roomsconsole AWS Management Console et ouvrez-la](#) avec votre membreCompte AWS.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sélectionnez votre collaboration pour accéder aux détails de votre collaboration.
4. Sur la page de collaboration, cliquez sur le bouton Actions et sélectionnez Modifier les approbations automatiques.
5. Pour accorder des capacités aux membres existants sans approuver manuellement une demande de modification :

1. Accédez à la section Capacités des membres de Grant.
2. Spécifiez les capacités des membres qui peuvent être automatiquement accordées.

 Note

Par défaut, tous les membres de la collaboration peuvent « contribuer aux données ».

3. (Facultatif) Pour permettre aux nouveaux membres de rejoindre instantanément une collaboration avec les capacités spécifiées sans approuver manuellement les demandes de modification, sélectionnez Approuver automatiquement les nouveaux membres dotés de ces capacités.
6. Pour autoriser la suppression des capacités des membres existants sans demande de modification, procédez comme suit :
 1. Accédez à la section Capacités pouvant être automatiquement révoquées.
 2. Spécifiez les capacités des membres qui peuvent être supprimées automatiquement.
7. Sélectionnez Enregistrer les modifications.
8. Confirmez l'envoi de votre demande de modification. Dans le mode de confirmation, confirmez les modifications et sélectionnez Soumettre la demande de modification.

Supprimer des collaborations

En tant que créateur de collaboration, vous pouvez supprimer une collaboration que vous avez créée.

Note

Lorsque vous supprimez une collaboration, vous et tous les membres ne pouvez pas exécuter de requêtes, recevoir de résultats ou apporter des données. Chaque membre de la collaboration continue d'avoir accès à ses propres données dans le cadre de son adhésion.

Pour supprimer une collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous souhaitez supprimer.
4. Sous Actions, choisissez Supprimer la collaboration.
5. Confirmez la suppression, puis choisissez Supprimer.

Afficher les collaborations

En tant que créateur de collaboration, vous pouvez consulter toutes les collaborations que vous avez créées.

Pour consulter les collaborations

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, sous Dernière utilisation, consultez les 5 dernières collaborations utilisées.
4. Dans l'onglet Avec adhésion active, consultez la liste des collaborations avec adhésion active.

Vous pouvez trier par nom, date de création de l'adhésion et informations relatives à votre membre.

Vous pouvez utiliser la barre de recherche pour rechercher une collaboration.

5. Dans l'onglet Disponible pour participer, consultez la liste des collaborations disponibles.

6. Dans l'onglet N'est plus disponible, consultez la liste des collaborations supprimées et des adhésions pour les collaborations qui ne sont plus disponibles (adhésions supprimées).

Inviter des membres à participer à une collaboration

En tant que créateur de collaboration, après avoir créé une collaboration, vous pouvez envoyer un lien d'invitation aux membres répertoriés dans l'onglet Membres.

Pour inviter des membres à une collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Choisissez l'onglet Membres.
5. Dans le tableau Membres, cliquez sur le bouton Copier le lien d'invitation.

Le lien d'invitation est copié.

6. Collez le lien d'invitation dans le mode de communication sécurisé de votre choix et envoyez-le à chaque membre de la collaboration.

Surveillance des membres

En tant que créateur de collaboration, après avoir créé une collaboration, vous pouvez suivre le statut de tous les membres dans l'onglet Membres. Cela permet de garantir un contrôle d'accès et une gestion des coûts appropriés dans l'ensemble de votre collaboration.

Pour surveiller le statut d'un membre

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration que vous avez créée.
4. Choisissez l'onglet Membres.
5. Dans le tableau Membres, passez en revue le statut de chaque membre.
6. Dans le tableau des capacités des membres, déterminez quels membres peuvent effectuer des requêtes, recevoir des résultats, fournir des données et effectuer d'autres tâches.

7. Dans le tableau de configuration des paiements, vérifiez quels membres paient pour les requêtes, les tables de mappage d'identifiants et la modélisation du machine learning.

Ajouter des membres à une collaboration

Conditions préalables

- Et Compte AWS avec des autorisations pour gérer les collaborations
- Le Compte AWS IDs nombre de membres que vous souhaitez ajouter

Pour ajouter un membre à une collaboration

1. Connectez-vous à la [AWS Clean Roomsconsole AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sélectionnez la collaboration à laquelle vous souhaitez ajouter des membres.
4. Choisissez l'onglet Membres.
5. Choisissez Edit members (Modifier les membres).
6. Choisissez Ajouter un autre membre et entrez les informations suivantes :
 - Nom d'affichage du membre
 - Compte AWSID de membre
 - Spécifiez s'ils peuvent recevoir des résultats
7. Sélectionnez Enregistrer les modifications.
8. Dans le mode de confirmation, vérifiez que les informations sont correctes, puis choisissez Soumettre la demande de modification.
 - Si la demande de modification nécessite l'approbation des autres membres, tous les membres existants doivent approuver la demande de modification avant que le nouveau membre ne soit ajouté. Pour plus d'informations sur les demandes de modification, consultez [the section called "Demandes de modification"](#).
 - Si les types de modification par approbation automatique sont pris en charge pour les nouveaux membres dotés des capacités spécifiées, cette modification prendra effet immédiatement. Pour de plus amples informations, veuillez consulter [the section called "Modifier les paramètres d'approbation automatique de la collaboration"](#).

9. Sur la page détaillée de la collaboration, sous l'onglet Membres, vérifiez que le statut de membre du ou des membres ajoutés s'affiche comme Invité.

Une fois ces étapes terminées, les membres invités peuvent rejoindre la collaboration. Pour plus d'informations sur l'adhésion à une collaboration, voir [Création d'un abonnement et adhésion à une collaboration](#)

Supprimer des membres d'une collaboration

Note

Avant de commencer, notez que la suppression d'un membre :

- Supprime tous leurs ensembles de données associés de la collaboration
- Si le [membre paie les coûts de calcul des requêtes](#), cette action arrête toutes les exécutions de requêtes dans le cadre de la collaboration.

La suppression d'un membre entraîne également la suppression de tous ses ensembles de données associés de la collaboration.

Prérequis

- Vous devez être un créateur de collaboration
- Vous ne pouvez pas supprimer votre propre compte

Pour supprimer un membre d'une collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sélectionnez la collaboration que vous souhaitez modifier.
4. Choisissez l'onglet Membres.
5. Sélectionnez le bouton d'option situé à côté du membre que vous souhaitez supprimer.
6. Sélectionnez Remove (Supprimer).
7. Dans la boîte de dialogue de confirmation, tapez **confirm** pour vérifier la suppression.

 Note

Après avoir supprimé un membre, tous les ensembles de données associés à son compte sont également supprimés de la collaboration.

 Important

Si vous supprimez le membre qui prend en charge les frais de calcul des requêtes, aucune autre requête ne pourra être exécutée dans le cadre de la collaboration tant que vous n'aurez pas désigné un nouveau membre payant.

Quitter une collaboration

En tant que membre d'une collaboration, vous pouvez quitter une collaboration en supprimant votre adhésion. Si vous êtes le créateur de la collaboration, vous ne pouvez quitter une collaboration qu'en [la supprimant](#).

 Note

Lorsque vous supprimez votre adhésion, vous quittez la collaboration et vous ne pouvez pas la rejoindre à nouveau. Si vous êtes [membre et que vous payez les frais de calcul des requêtes et que vous supprimez votre adhésion, aucune autre requête](#) n'est autorisée à être exécutée.

Pour quitter une collaboration

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Pour Avec adhésion active, choisissez la collaboration dont vous êtes membre.
4. Choisissez Actions.
5. Choisissez Supprimer l'adhésion.
6. Dans la boîte de dialogue, confirmez la décision de quitter la collaboration **confirm** en saisissant du texte dans le champ de saisie, puis en choisissant Vide et supprimer l'adhésion.

Un message s'affiche sur la console indiquant que l'adhésion a été supprimée.

Le créateur de la collaboration considère que le statut de membre est « Gauche ».

Préparez des tableaux de données dans AWS Clean Rooms

Note

La préparation des tableaux de données peut avoir lieu avant ou après avoir rejoint une collaboration. Une fois qu'un tableau est préparé, vous pouvez le réutiliser dans plusieurs collaborations, à condition que vos besoins en matière de confidentialité soient les mêmes pour ce tableau.

En tant que membre de la collaboration, vous devez préparer vos tables de données avant que le membre AWS Clean Rooms de la collaboration puisse les interroger.

Les tables de données que vous utilisez pour les requêtes AWS Clean Rooms sont généralement du même type que celles que vous utilisez pour d'autres applications. Par exemple, les mêmes types de jeux de données sont utilisés avec Amazon Athena, Amazon EMR, Amazon Redshift Spectrum et Amazon Quick.

Vous pouvez interroger les données dans leur format d'origine directement à partir de l'une des sources de données suivantes :

- Amazon Simple Storage Service (Amazon S3)
- Amazon Athena
- Snowflake

AWS Clean Rooms accède à l'ensemble de données au moment de l'exécution de la requête, ce qui garantit que les membres autorisés à effectuer des requêtes accèdent toujours au plus grand nombre de up-to-date données. Toutes les données temporairement lues dans une AWS Clean Rooms collaboration sont supprimées une fois la requête terminée. Les résultats de la requête sont écrits dans votre compartiment Amazon S3.

Si votre cas d'utilisation implique de demander des données d'identité, consultez [Résolution des entités AWS dans AWS Clean Rooms](#).

Rubriques

- [Formats de données pour AWS Clean Rooms](#)
- [Apache Iceberg tables en AWS Clean Rooms](#)

- [Préparation des tables de données pour les requêtes dans AWS Clean Rooms](#)
- [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#)
- [Déchiffrer des tables de données avec le client de chiffrement C3R](#)

Formats de données pour AWS Clean Rooms

Pour analyser les données, les ensembles de données doivent être dans un format AWS Clean Rooms compatible.

Rubriques

- [Formats de données pris en charge pour les PySpark tâches](#)
- [Formats de données pris en charge pour les requêtes SQL](#)
- [Types de données pris en charge](#)
- [Types de compression de fichiers pour AWS Clean Rooms](#)
- [Chiffrement côté serveur pour AWS Clean Rooms](#)

Formats de données pris en charge pour les PySpark tâches

AWS Clean Rooms prend en charge les formats structurés suivants pour exécuter PySpark des tâches.

- Parquet
- OpenCSV
- JSON

Formats de données pris en charge pour les requêtes SQL

AWS Clean Rooms prend en charge les formats structurés suivants pour exécuter des requêtes SQL.

- [Tables Apache Iceberg](#)
- Parquet
- OpenCSV
- JSON

 Note

`timestamp` La valeur d'un fichier texte doit être au format `yyyy-MM-dd HH:mm:ss.SSSSSS`.
Par exemple : `2017-05-01 11:30:59.000000`.

Nous vous recommandons d'utiliser un format de fichier de stockage en colonnes, tel que Apache Parquet. Avec un format de fichier de stockage en colonnes, vous pouvez minimiser les mouvements de données en sélectionnant uniquement les colonnes dont vous avez besoin. Pour des performances optimales, les objets volumineux doivent être divisés en objets de 100 Mo à 1 Go.

Types de données pris en charge

AWS Clean Rooms prend en charge les types de données suivants.

- ARRAY
- BIGINT
- BOOLEAN
- BYTE
- CHAR
- DATE
- DECIMAL
- FLOAT
- INTEGER
- INTERVAL
- LONG
- MAP
- REAL
- SHORT
- SMALLINT
- STRUCT
- TIME
- TIMESTAMP_LTZ

- TIMESTAMP_NTZ
- TINYINT
- VARCHAR

Pour plus d'informations, consultez la section [Types de données](#) dans la référence AWS Clean Rooms SQL.

Types de compression de fichiers pour AWS Clean Rooms

Pour réduire l'espace de stockage, améliorer les performances et minimiser les coûts, nous vous recommandons vivement de compresser vos ensembles de données.

AWS Clean Rooms reconnaît les types de compression de fichiers en fonction de leur extension et prend en charge les types de compression et les extensions indiqués dans le tableau suivant.

Algorithme de compression	Extension de fichier
GZIP	.gz
Bzip2	.bz2
Snappy	.snappy

Vous pouvez appliquer la compression à différents niveaux. Le plus souvent, vous compressez un fichier entier ou des blocs individuels dans un fichier. La compression des formats en colonnes au niveau du fichier n'apporte aucun avantage en termes de performances.

Chiffrement côté serveur pour AWS Clean Rooms

Note

Le chiffrement côté serveur ne remplace pas le calcul cryptographique dans les cas d'utilisation qui l'exigent.

AWS Clean Rooms déchiffre de manière transparente les ensembles de données chiffrés à l'aide des options de chiffrement suivantes :

- SSE-S3 — Chiffrement côté serveur à l'aide d'une clé de chiffrement AES-256 gérée par Amazon S3
- SSE-KMS — Chiffrement côté serveur avec des clés gérées par AWS Key Management Service

Pour utiliser SSE-S3, le rôle de AWS Clean Rooms service utilisé pour associer la table configurée à la collaboration doit disposer des autorisations KMS-Decrypt. Pour utiliser SSE-KMS, la politique de clé KMS doit également autoriser le rôle de AWS Clean Rooms service à déchiffrer.

AWS Clean Rooms ne prend pas en charge le chiffrement côté client Amazon S3. Pour plus d'informations sur le chiffrement côté serveur, consultez la section [Protection des données à l'aide du chiffrement côté serveur dans](#) le guide de l'utilisateur d'Amazon Simple Storage Service.

Apache Iceberg tables en AWS Clean Rooms

Apache Iceberg est un format de table open source pour les lacs de données. AWS Clean Rooms peut utiliser les statistiques stockées dans Apache Iceberg métadonnées pour optimiser les plans de requêtes et réduire le nombre de scans de fichiers pendant le traitement des requêtes en salle blanche. Pour plus d'informations, consultez la documentation d'[Apache Iceberg](#).

Lorsque vous utilisez des tables Iceberg, tenez compte AWS Clean Rooms des points suivants :

- Tables Apache Iceberg pour S3 — Apache Iceberg les tables doivent être définies sur la AWS Glue Data Catalog base de l'[implémentation du catalogue de colle open source](#).
- Tables Apache Iceberg pour Athena — [Pour plus d'informations, consultez le fichier -iceberg.html https://docs.aws.amazon.com/athena/latest/ug/querying](#)
- Tables Apache Iceberg pour Snowflake — [Pour plus d'informations, voir user-guide/tables-iceberg https://docs.snowflake.com/en/](#)
- Format de fichier Parquet : prend AWS Clean Rooms uniquement en charge les tables Iceberg au format de fichier de données Parquet.
- Compression GZIP et Snappy : AWS Clean Rooms supporte Parquet avec GZIP et Snappy compression.
- Versions Iceberg : AWS Clean Rooms permet d'exécuter des requêtes sur les tables Iceberg des versions 1 et 2.
- Partitions — Vous n'avez pas besoin d'ajouter manuellement des partitions pour votre Apache Iceberg tables en AWS Glue. AWS Clean Rooms détecte les nouvelles partitions dans Apache Iceberg les tables sont automatiquement et aucune opération manuelle n'est nécessaire pour

mettre à jour les partitions dans la définition des tables. Les partitions Iceberg apparaissent sous forme de colonnes normales dans le schéma de AWS Clean Rooms table et non séparément sous forme de clé de partition dans le schéma de table configuré.

- Limites

- Nouvelles tables Iceberg uniquement

Apache Iceberg tables converties à partir de Apache Parquet les tables ne sont pas prises en charge.

- Requêtes Time Travel

AWS Clean Rooms ne prend pas en charge les demandes de voyage dans le temps avec Apache Iceberg tables.

- Moteur Athena version 2

Iceberg les tables créées avec la version 2 du moteur Athena ne sont pas prises en charge.

- Formats de fichiers

Avro et les formats de fichier ORC (Optimized Row Columnar) ne sont pas pris en charge.

- Compression

Zstandard (Zstd) compression pour Parquet n'est pas pris en charge.

Types de données pris en charge pour les tables Iceberg

AWS Clean Rooms peut interroger Iceberg tables contenant les types de données suivants :

- BOOLEAN
- DATE
- DECIMAL
- DOUBLE
- FLOAT
- INT
- LIST
- LONG
- MAP

- STRING
- STRUCT
- TIMESTAMP WITHOUT TIME ZONE

Pour en savoir plus sur les types de données Iceberg, consultez [Schemas for Iceberg](#) dans la documentation Apache Iceberg.

Préparation des tables de données pour les requêtes dans AWS Clean Rooms

Si votre cas d'utilisation ne vous oblige pas à apporter vos propres données, vous pouvez ignorer cette procédure.

Si votre cas d'utilisation implique de demander des données d'identité, consultez [Résolution des entités AWS dans AWS Clean Rooms](#).

Pour plus d'informations sur les formats de données que vous pouvez utiliser, consultez [Formats de données pour AWS Clean Rooms](#).

Rubriques

- [Préparation des tables de données dans Amazon S3](#)
- [Préparation de tables de données dans Amazon Athena](#)
- [Préparation des tables de données dans Snowflake](#)

Préparation des tables de données dans Amazon S3

Vous pouvez analyser les tables de données qui ont été cataloguées AWS Glue et stockées dans Amazon S3. Si vos tables de données sont déjà cataloguées AWS Glue, passez à [Création d'une table configurée dans AWS Clean Rooms](#).

Important

Lorsque vous préparez vos données dans Amazon S3 en vue de leur utilisation AWS Clean Rooms, vous devez vous assurer que l'emplacement de votre AWS Glue table correspond exactement au chemin du répertoire dans lequel vos fichiers de données sont stockés.

Par exemple : si vos données sont stockées dans `s3://mybucket/folder/subfolder/data.parquet`, l'emplacement de votre AWS Glue table doit pointer vers `s3://mybucket/folder/subfolder/`. Si vous définissez l'emplacement de la table sur un répertoire parent (`s3://mybucket/folder/``), la table apparaîtra vide lorsqu'elle sera interrogée.

La préparation de vos tables de données dans Amazon S3 implique les étapes suivantes :

Rubriques

- [Étape 1 : Exécuter les prérequis](#)
- [Étape 2 : \(Facultatif\) Préparez vos données pour le calcul cryptographique](#)
- [Étape 3 : Chargez votre tableau de données sur Amazon S3](#)
- [Étape 4 : Création d'une AWS Glue table](#)
- [Étape 5 : étapes suivantes](#)

Étape 1 : Exécuter les prérequis

Pour préparer vos tables de données à utiliser avec AWS Clean Rooms, vous devez remplir les conditions préalables suivantes :

- Vos tables de données sont enregistrées dans l'un des [formats de données pris en charge pour AWS Clean Rooms](#).
- Vos tables de données sont cataloguées AWS Glue et utilisent les [types de données pris en charge pour AWS Clean Rooms](#).
- Toutes vos tables de données sont stockées dans Amazon Simple Storage Service (Amazon S3), là où la Région AWS collaboration a été créée.
- Ils AWS Glue Data Catalog doivent se trouver dans la même région que la collaboration.
- AWS Glue Data Catalog C'est la même chose Compte AWS que l'adhésion.
- Le compartiment Amazon S3 n'est pas enregistré auprès de AWS Lake Formation.

Étape 2 : (Facultatif) Préparez vos données pour le calcul cryptographique

(Facultatif) Si vous utilisez l'informatique cryptographique et que votre table de données contient des informations sensibles que vous souhaitez chiffrer, vous devez chiffrer la table de données à l'aide du client de chiffrement C3R.

Pour préparer vos données pour le calcul cryptographique, suivez les procédures décrites dans [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

Étape 3 : Chargez votre tableau de données sur Amazon S3

Note

Si vous avez l'intention d'utiliser des tables de données chiffrées dans le cadre de la collaboration, vous devez d'abord chiffrer les données pour le calcul cryptographique avant de télécharger votre table de données sur Amazon S3. Pour de plus amples informations, veuillez consulter [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

Pour télécharger votre tableau de données sur Amazon S3

1. Connectez-vous à la console Amazon S3 AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/s3/>.
2. Choisissez Buckets, puis choisissez un bucket dans lequel vous souhaitez stocker votre table de données.
3. Choisissez Upload, puis suivez les instructions.
4. Choisissez l'onglet Objets pour afficher le préfixe dans lequel vos données sont stockées. Notez le nom du dossier.

Vous pouvez sélectionner le dossier pour afficher les données.

Étape 4 : Création d'une AWS Glue table

Si vous disposez déjà d'une table de AWS Glue données, vous pouvez ignorer cette étape.

Au cours de cette étape, vous configurez un robot d'exploration AWS Glue qui analyse tous les fichiers de votre compartiment S3 et crée une AWS Glue table. Pour plus d'informations, consultez la section [Définition des robots d'exploration AWS Glue dans](#) le guide de l'AWS Glue utilisateur.

Pour plus d'informations sur les types de AWS Glue Data Catalog données pris en charge, consultez [Types de données pris en charge](#).

 Note

AWS Clean Rooms ne prend actuellement pas en charge les compartiments S3 enregistrés auprès AWS Lake Formation de.

La procédure suivante décrit comment créer une AWS Glue table. Si vous souhaitez utiliser un AWS Glue Data Catalog objet chiffré avec une clé AWS Key Management Service (AWS KMS), vous devez configurer la politique d'autorisation des clés KMS pour autoriser l'accès à cette table chiffrée. Pour plus d'informations, consultez la section [Configuration du chiffrement dans AWS Glue](#) dans le manuel du AWS Glue développeur.

Pour créer une AWS Glue table

1. Suivez la procédure [relative à l'utilisation des robots d'exploration sur la AWS Glue console](#) du guide de l'AWS Glue utilisateur.
2. Notez le nom de la AWS Glue base de données et le nom de AWS Glue la table.

Étape 5 : étapes suivantes

Maintenant que vous avez préparé vos tables de données dans Amazon S3, vous êtes prêt à :

- [Création d'une table configurée](#)
- [Création d'un modèle ML](#)

Les tables peuvent être consultées après :

- Le créateur de la collaboration a configuré une collaboration dans AWS Clean Rooms. Pour de plus amples informations, veuillez consulter [Création d'une collaboration](#).
- Le créateur de la collaboration vous a envoyé l'identifiant de collaboration en tant que participant à la collaboration.

Préparation de tables de données dans Amazon Athena

Vous pouvez interroger des tables de données créées sous forme de vues AWS Glue Data Catalog (GDC) dans Amazon Athena.

Une vue GDC est une table virtuelle créée à partir d'une ou de plusieurs AWS Glue tables sous-jacentes. Il doit être créé à l'aide d'Athena SQL dans le catalogue Athena. `AwsGlueCatalog`

La préparation de vos tables de données dans Amazon Athena implique les étapes suivantes :

Rubriques

- [Étape 1 : Exécuter les prérequis](#)
- [Étape 2 : \(Facultatif\) Préparez vos données pour le calcul cryptographique](#)
- [Étape 3 : Prochaines étapes](#)

Étape 1 : Exécuter les prérequis

Pour préparer vos tables de données à utiliser avec AWS Clean Rooms, vous devez remplir les conditions préalables suivantes :

- Vos tables de données sont enregistrées dans l'un des [formats de données pris en charge pour AWS Clean Rooms](#).
- Vos tables de données utilisent les [types de données pris en charge pour AWS Clean Rooms](#).
- Vous avez créé une vue GDC sur votre AWS Glue table à l'aide d'Athena SQL dans le catalogue `AwsDataCatalog` Athena.

La vue apparaîtra dans :

- La console Athena (sous le `AwsDataCatalog`) en tant que vue : <https://console.aws.amazon.com/athena/>
- La AWS Glue console sous forme de AWS Glue table : <https://console.aws.amazon.com/glue/>

Pour plus d'informations, consultez la section [Utiliser les vues du catalogue de données dans Athena](#) dans le guide de l'utilisateur d'Amazon Athena.

 Note

Vous avez besoin des autorisations appropriées pour créer des vues dans Athena et. AWS Glue Assurez-vous également que vous avez accès aux tables sous-jacentes référencées dans votre définition de vue.

AWS Clean Rooms ne prend en charge que le type de AWS Glue catalogue pour Athena, et non les types de catalogue Lambda ou Hive.

- Vos tables de données ou vues GDC sont cataloguées AWS Glue et enregistrées auprès de. AWS Lake Formation
- Vous avez créé un compartiment de sortie distinct dans Amazon S3 pour recevoir les résultats d'Athena.
- Vous avez configuré un rôle de service pour lire les données d'Amazon Athena. Pour de plus amples informations, veuillez consulter [Création d'un rôle de service pour lire les données d'Amazon Athena](#).
- Le rôle de service dispose des autorisations d'accès Lake Formation Select et Describe sur la vue ou la table GDC.

Étape 2 : (Facultatif) Préparez vos données pour le calcul cryptographique

(Facultatif) Si vous utilisez l'informatique cryptographique et que votre table de données contient des informations sensibles que vous souhaitez chiffrer, vous devez chiffrer la table de données à l'aide du client de chiffrement C3R.

Pour préparer vos données pour le calcul cryptographique, suivez les procédures décrites dans [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

Étape 3 : Prochaines étapes

Maintenant que vous avez préparé vos tables de données dans Amazon Athena, vous êtes prêt à :

- [Création d'une table configurée](#)
- [Création d'un modèle ML](#)

Les tables peuvent être consultées après :

- Le créateur de la collaboration a configuré une collaboration dans AWS Clean Rooms. Pour de plus amples informations, veuillez consulter [Création d'une collaboration](#).
- Le créateur de la collaboration vous a envoyé l'identifiant de collaboration en tant que participant à la collaboration.

Préparation des tables de données dans Snowflake

Vous pouvez interroger les tables de données qui ont été stockées dans l'entrepôt de données Snowflake.

La préparation de vos tables de données dans Snowflake implique les étapes suivantes :

Rubriques

- [Étape 1 : Exécuter les prérequis](#)
- [Étape 2 : \(Facultatif\) Préparez vos données pour le calcul cryptographique](#)
- [Étape 3 : Créez un AWS Secrets Manager secret](#)
- [Étape 4 : étapes suivantes](#)

Étape 1 : Exécuter les prérequis

Pour préparer vos tables de données à utiliser avec AWS Clean Rooms, vous devez remplir les conditions préalables suivantes :

- Vous disposez d'un compte AWS des autorisations appropriées pour lire vos tableaux de données. Pour de plus amples informations, veuillez consulter [Créez un rôle de service pour lire les données de Snowflake](#).
- Vos tables de données sont enregistrées dans l'un des [formats de données pris en charge pour AWS Clean Rooms](#).
- Vos tables de données utilisent les [types de données pris en charge pour AWS Clean Rooms](#).
- Votre table de données est stockée dans un entrepôt Snowflake. Pour plus d'informations, consultez la documentation de [Snowflake](#).
- Vous avez configuré un nouvel utilisateur Snowflake avec des privilèges de lecture seule sur la table Snowflake que vous allez associer à votre collaboration.

Étape 2 : (Facultatif) Préparez vos données pour le calcul cryptographique

(Facultatif) Si vous utilisez l'informatique cryptographique et que votre table de données contient des informations sensibles que vous souhaitez chiffrer, vous devez chiffrer la table de données à l'aide du client de chiffrement C3R.

Pour préparer vos données pour le calcul cryptographique, suivez les procédures décrites dans [Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms](#).

Étape 3 : Créez un AWS Secrets Manager secret

Pour vous connecter à Snowflake depuis AWS Clean Rooms, vous devez créer et stocker vos informations d'identification Snowflake dans un AWS Secrets Manager secret, puis associer ce secret à une table Snowflake dans AWS Clean Rooms

Note

Nous vous recommandons de créer un nouvel utilisateur exclusivement pour AWS Clean Rooms. Cet utilisateur ne doit avoir un rôle doté d'autorisations de lecture que pour les données auxquelles vous AWS Clean Rooms souhaitez accéder.

Pour créer un AWS Secrets Manager secret

1. Dans Snowflake, générez un utilisateur `snowflakeUser` et configurez l'authentification par paire de clés.

Note

En novembre 2025, Snowflake prendra en charge uniquement l'authentification par paire de clés. Cette modification affectera l' AWS Clean Rooms intégration actuelle avec Snowflake, qui utilise l'authentification par nom d'utilisateur et mot de passe. Après cette date, les connexions Snowflake AWS Clean Rooms nécessiteront une authentification par paire de clés à l'aide d'une clé privée Snowflake Privacy Enhanced Mail (PEM).

2. Déterminez avec quel entrepôt Snowflake cet utilisateur va interagir, `snowflakeWarehouse`. Vous pouvez le `DEFAULT_WAREHOUSE` définir comme forme `snowflakeUser` dans Snowflake ou vous en souvenir pour l'étape suivante.

3. Dans [AWS Secrets Manager](#), créez un secret à l'aide de vos informations d'identification Snowflake. Pour créer un secret dans Secrets Manager, suivez le didacticiel disponible dans la [section Créer un AWS Secrets Manager secret](#) du Guide de l'AWS Secrets Manager utilisateur. Après avoir créé le secret, conservez le nom du secret `secretName` pour l'étape suivante.
- Lorsque vous sélectionnez des paires clé/valeur, créez une paire pour `snowflakeUser` avec la clé. `sfUser`
 - Lorsque vous sélectionnez des paires clé/valeur, créez une paire pour votre clé privée Snowflake PEM avec la clé. `pem_private_key`
 - Lorsque vous sélectionnez des paires clé/valeur, créez une paire pour `snowflakeWarehouse` avec la clé. `sfWarehouse`
- Cela n'est pas nécessaire si une valeur par défaut est définie dans Snowflake.
- Lorsque vous sélectionnez des paires clé/valeur, créez une paire pour `snowflakeRole` avec la clé. `sfRole`

Étape 4 : étapes suivantes

Maintenant que vous avez préparé vos tableaux de données dans Snowflake, vous êtes prêt à :

- [Création d'une table configurée](#)
- [Création d'un modèle ML](#)

Les tables peuvent être consultées après :

- Le créateur de la collaboration a configuré une collaboration dans AWS Clean Rooms. Pour de plus amples informations, veuillez consulter [Création d'une collaboration](#).
- Le créateur de la collaboration vous a envoyé l'identifiant de collaboration en tant que participant à la collaboration.

Préparation de tables de données chiffrées à l'aide de l'informatique cryptographique pour Clean Rooms

L'informatique cryptographique pour Clean Rooms (C3R) est une fonctionnalité de AWS Clean Rooms. Vous pouvez utiliser C3R pour limiter cryptographiquement ce qui peut être appris par n'importe quelle partie et AWS dans le cadre d'une AWS Clean Rooms collaboration.

Vous pouvez chiffrer la table de données à l'aide du client de chiffrement C3R, un outil de chiffrement côté client, avant de télécharger la table de données dans votre source de données : Amazon Simple Storage Service (Amazon S3), Amazon Athena ou Snowflake.

Pour de plus amples informations, veuillez consulter [Informatique cryptographique pour Clean Rooms](#).

La préparation de tables de données chiffrées avec C3R implique les étapes suivantes :

Étapes

- [Étape 1 : Exécuter les prérequis](#)
- [Étape 2 : Téléchargez le client de chiffrement C3R](#)
- [Étape 3 : \(Facultatif\) Afficher les commandes disponibles dans le client de chiffrement C3R](#)
- [Étape 4 : générer un schéma de chiffrement pour un fichier tabulaire](#)
- [Étape 5 : Création d'une clé secrète partagée](#)
- [Étape 6 : Stocker la clé secrète partagée dans une variable d'environnement](#)
- [Étape 7 : Chiffrer les données](#)
- [Étape 8 : vérifier le chiffrement des données](#)
- [\(Facultatif\) Créez un schéma \(utilisateurs avancés\)](#)

Étape 1 : Exécuter les prérequis

Pour préparer vos tables de données en vue de leur utilisation avec C3R, vous devez remplir les conditions préalables suivantes :

- Vous pouvez accéder au Clean Rooms référentiel Cryptographic Computing for sur GitHub :
<https://github.com/aws/c3r>
- Vous avez configuré les AWS informations d'identification pour utiliser le client de chiffrement C3R. Ces informations d'identification sont utilisées par le client de chiffrement C3R pour les appels d'API en lecture seule afin de récupérer les métadonnées AWS Clean Rooms de collaboration. Pour plus d'informations, consultez [la section Configuration du AWS CLI](#) dans le guide de AWS Command Line Interface l'utilisateur de la version 2.
- Vous avez installé Java Runtime Environment (JRE) 11 ou une version ultérieure sur votre machine.

- [La version recommandée Java Runtime Environment, Amazon Corretto 11 ou version ultérieure, peut être téléchargée depuis /corretto. https://aws.amazon.com](https://aws.amazon.com/corretto)
- Le Java Development Kit (JDK) inclut un correspondant JRE de la même version. Toutefois, les fonctionnalités supplémentaires du ne JDK sont pas nécessaires pour exécuter le client de chiffrement Cryptographic Computing for Clean Rooms (C3R).
- Vos fichiers de données tabulaires (.csv) ou vos Parquet fichiers (. parquet) sont enregistrés localement.
- Vous ou un autre membre de la collaboration avez la possibilité de créer une clé secrète partagée. Pour de plus amples informations, veuillez consulter [Étape 5 : Création d'une clé secrète partagée](#).
- Le créateur de la collaboration a créé une collaboration AWS Clean Rooms avec l'informatique cryptographique activée pour la collaboration. Pour de plus amples informations, veuillez consulter [Création d'une collaboration](#).
- Le créateur de la collaboration vous a envoyé l'identifiant de collaboration en tant que participant à la collaboration. Le nom de ressource Amazon (ARN) de la collaboration est inclus dans l'invitation envoyée, qui contient l'ID de collaboration.

Étape 2 : Téléchargez le client de chiffrement C3R

Pour télécharger le client de chiffrement C3R depuis GitHub

1. [Accédez au Clean RoomsAWSGitHub référentiel Cryptographic Computing for : c3r https://github.com/aws/](https://github.com/aws/c3r)
2. Sélectionnez et téléchargez les fichiers.

Le code source, les licences et le matériel connexe peuvent être clonés ou téléchargés sous forme de fichier. zipfichier depuis la page d'accueil du GitHub dépôt. (Voir le bouton Code en haut à droite de la liste du contenu du référentiel).

Le dernier client de chiffrement C3R signé Java Executable File (c'est-à-dire l'application d'interface de ligne de commande) se trouve sur la page Versions du GitHub référentiel.

Le package client de chiffrement C3R pour Apache Spark (`c3r-cli-spark`) est une version du `c3r-cli` qui doit être soumise en tant que tâche à un serveur Apache Spark en cours d'exécution. Pour plus d'informations, consultez [Exécuter C3R sur Apache Spark](#).

Étape 3 : (Facultatif) Afficher les commandes disponibles dans le client de chiffrement C3R

Utilisez cette procédure pour vous familiariser avec les commandes disponibles dans le client de chiffrement C3R.

Pour afficher toutes les commandes disponibles dans le client de chiffrement C3R

1. À partir d'une interface de ligne de commande (CLI), accédez au dossier contenant le `c3r-cli.jar` fichier téléchargé.
2. Exécutez la commande suivante: `java -jar c3r-cli.jar`
3. Consultez la liste des commandes et options disponibles.

Étape 4 : générer un schéma de chiffrement pour un fichier tabulaire

Pour chiffrer des données, un schéma de chiffrement décrivant la manière dont les données seront utilisées est requis. Cette section décrit comment le client de chiffrement C3R aide à générer un schéma de chiffrement pour un fichier CSV avec une ligne d'en-tête ou un Parquet fichier.

Vous ne devez effectuer cette opération qu'une seule fois par fichier. Une fois que le schéma existe, il peut être réutilisé pour chiffrer le même fichier (ou tout autre fichier dont le nom de colonne est identique). Si les noms des colonnes ou le schéma de chiffrement souhaité changent, vous devez mettre à jour le fichier de schéma. Pour de plus amples informations, veuillez consulter [\(Facultatif\) Créez un schéma \(utilisateurs avancés\)](#).

Important

Il est essentiel que toutes les parties collaboratrices utilisent la même clé secrète partagée. Les parties collaboratrices doivent également coordonner les noms des colonnes afin de déterminer s'ils doivent être JOIN édités ou comparés pour garantir l'égalité dans les requêtes. Dans le cas contraire, les requêtes SQL risquent de produire des résultats inattendus ou incorrects. Toutefois, cela n'est pas nécessaire si le créateur de la collaboration a activé le paramètre de `allowJoinsOnColumnsWithDifferentNames` chiffrement lors de la création de la collaboration. Pour plus d'informations sur les paramètres relatifs au chiffrement, consultez. [Paramètres de calcul cryptographique](#)

Lorsqu'il est exécuté en mode schéma, le client de chiffrement C3R parcourt le fichier d'entrée colonne par colonne pour vous demander si et comment cette colonne doit être traitée. Si le fichier contient de nombreuses colonnes qui ne sont pas souhaitées pour la sortie cryptée, la génération de schéma interactif peut devenir fastidieuse car vous devez ignorer chaque colonne indésirable. Pour éviter cela, vous pouvez écrire manuellement un schéma ou créer une version simplifiée du fichier d'entrée contenant uniquement les colonnes souhaitées. Ensuite, le générateur de schéma interactif pourrait être exécuté sur ce fichier réduit. Le client de chiffrement C3R produit des informations sur le fichier de schéma et vous demande comment les colonnes source doivent être incluses ou cryptées (le cas échéant) dans la sortie cible.

Pour chaque colonne source du fichier d'entrée, vous êtes invité à saisir :

1. Combien de colonnes cibles doivent être générées
2. Comment chaque colonne cible doit être cryptée (le cas échéant)
3. Le nom de chaque colonne cible
4. Comment les données doivent être remplies avant le chiffrement si la colonne est chiffrée en tant que sealed colonne

 Note

Lorsque vous chiffrez les données d'une colonne chiffrée en tant que sealed colonne, vous devez déterminer quelles données doivent être rembourrées. Le client de chiffrement C3R suggère un rembourrage par défaut lors de la génération du schéma, afin que toutes les entrées d'une colonne soient garnies de la même longueur.

Lorsque vous déterminez la longueur `fixed`, notez que le remplissage est exprimé en octets et non en bits.

Vous trouverez ci-dessous une table de décision pour créer le schéma.

Tableau de décision relatif au schéma

Décision	Nombre de colonnes cibles depuis la colonne source <'name-of-column'> ?	Type de colonne cible : [c]clear text, [f] fingerprint ou [s] sealed ?	Nom de l'en-tête de la colonne cible <default 'name-of-column'>	Ajouter un suffixe <suffix> à l'en-tête pour indiquer comment il a été chiffré, [y] oui ou [n] non <default 'yes'>	<'name-of-column_sealed'> type de remboursement : [n] un, [f] fixe ou [m] max <default 'max'>
Laissez la colonne non chiffrée.	1	c	Non applicable	Ne s'applique pas	Non applicable
Chiffrez la colonne en tant que fingerprint colonne.	1	f	Choisissez le nom par défaut ou entrez un nouveau nom d'en-tête.	Entrez y pour choisir par défaut (_fingerprint) ou entrezn.	Non applicable
Chiffrez la colonne en tant que sealed colonne.	1	s	Choisissez le nom par défaut ou entrez un nouveau nom d'en-tête.	Entrez y pour choisir par défaut (_sealed) ou entrezn.	Choisissez le type de remboursement. Pour de plus amples informations, veuillez consulter (Facultatif) Créez un schéma (utilisateurs avancés) .

Décision	Nombre de colonnes cibles depuis la colonne source <'name-of-column'> ?	Type de colonne cible : [c]cleartext, [f] fingerprint ou [s] sealed ?	Nom de l'en-tête de la colonne cible <default 'name-of-column'>	Ajouter un suffixe <suffix> à l'en-tête pour indiquer comment il a été chiffré, [y] oui ou [n] non <default 'yes'>	<'name-of-column_sealed'> type de remboursement : [n] un, [f] fixe ou [m] max <default 'max'>
Chiffrez la colonne sous la forme à la fois fingerprint et. sealed	2	Entrez la première colonne cible : f. Entrez la deuxième colonne cible : s.	Choisissez les en-têtes cibles pour chaque colonne cible.	Entrez y pour choisir la valeur par défaut ou entrez n.	Choisissez le type de remboursement (pour les sealed colonnes uniquement). Pour de plus amples informations, veuillez consulter (Facultatif) Créez un schéma (utilisateurs avancés) .

Voici deux exemples de création de schémas de chiffrement. Le contenu exact de votre interaction dépend du fichier d'entrée et des réponses que vous fournissez.

Exemples

- [Exemple : génération d'un schéma de chiffrement pour une fingerprint colonne et une cleartext colonne](#)

- [Exemple : génération d'un schéma de chiffrement avec des cleartext colonnes sealedfingerprint, et](#)

Exemple : génération d'un schéma de chiffrement pour une fingerprint colonne et une cleartext colonne

Dans cet exemple, pour `ads.csv`, il n'y a que deux colonnes : `username` et `ad_variant`. Pour ces colonnes, nous voulons ce qui suit :

- Pour que la `username` colonne soit cryptée en tant que `fingerprint` colonne
- Pour que la `ad_variant` colonne soit une `cleartext` colonne

Pour générer un schéma de chiffrement pour une `fingerprint` colonne et une `cleartext` colonne

1. (Facultatif) Pour vous assurer que le `c3r-cli.jar` fichier et le fichier à chiffrer sont présents :
 - a. Naviguez jusqu'au répertoire souhaité et exécutez `ls` (si vous utilisez a Mac ou Unix/Linux) ou `dir` si vous utilisez Windows).
 - b. Consultez la liste des fichiers de données tabulaires (par exemple, `.csv`) et choisissez un fichier à chiffrer.

Dans cet exemple, `ads.csv` c'est le fichier que nous voulons chiffrer.

2. À partir de la CLI, exécutez la commande suivante pour créer un schéma de manière interactive.

```
java -jar c3r-cli.jar schema ads.csv --interactive --output=ads.json
```

Note

- Tu peux courir `java --jar PATH/T0/c3r-cli.jar`. Ou, si vous l'avez ajoutée `PATH/T0/c3r-cli.jar` à votre variable d'environnement `CLASSPATH`, vous pouvez également exécuter le nom de la classe. Le client de chiffrement C3R regardera le `CLASSPATH` pour le trouver (par exemple,). `java com.amazon.psion.cli.Main`
- L'`--interactive` indicateur sélectionne le mode interactif pour développer le schéma. Cela guide l'utilisateur à travers un assistant de création du schéma. Les utilisateurs ayant des compétences avancées peuvent créer leur propre schéma JSON sans utiliser l'assistant. Pour de plus amples informations, veuillez consulter [\(Facultatif\) Créez un schéma \(utilisateurs avancés\)](#).

- L'option `--outputindicateur` définit un nom de sortie. Si vous n'incluez pas l'option `--outputindicateur`, le client de chiffrement C3R essaie de choisir un nom de sortie par défaut (tel que `<input>.out.csv` ou pour le schéma `<input>.json`).

3. Pour `Number of target columns from source column 'username'?`, entrez, **1** puis appuyez sur Entrée.
4. Pour `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, entrez, **f** puis appuyez sur Entrée.
5. Pour `Target column headername <default 'username'>`, appuyez sur Entrée.

Le nom par défaut « username » est utilisé.

6. Pour `Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'>`, entrez, **y** puis appuyez sur Entrée.

Note

Le mode interactif suggère des suffixes à ajouter aux en-têtes de colonnes chiffrés (`_fingerprint` pour les fingerprint colonnes et `_sealed` pour sealed les colonnes). Les suffixes peuvent être utiles lorsque vous effectuez des tâches telles que le téléchargement de données Services AWS ou la création de collaborations. AWS Clean Rooms Ces suffixes peuvent aider à indiquer ce qui peut être fait avec les données chiffrées de chaque colonne. Par exemple, les choses ne fonctionneront pas si vous chiffrez une colonne sous forme de sealed colonne (`_sealed`) et que vous essayez d'y JOIN accéder ou que vous essayez l'inverse.

7. Pour `Number of target columns from source column 'ad_variant'?`, entrez, **1** puis appuyez sur Entrée.
8. Pour `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, entrez, **c** puis appuyez sur Entrée.
9. Pour `Target column headername <default 'username'>`, appuyez sur Entrée.

Le nom par défaut « ad_variant » est utilisé.

Le schéma est écrit dans un nouveau fichier appelé `ads.json`.

 Note

Vous pouvez afficher le schéma en l'ouvrant dans n'importe quel éditeur de texte, tel que Notepad on Windows ou TextEdit on macOS.

10. Vous êtes maintenant prêt à [chiffrer les données](#).

Exemple : génération d'un schéma de chiffrement avec des cleartext colonnes sealedfingerprint, et

Dans cet exemple, pour sales.csv, il y a trois colonnes : usernamepurchased, et product. Pour ces colonnes, nous voulons ce qui suit :

- Pour que la product colonne soit une sealed colonne
- Pour que la username colonne soit cryptée en tant que fingerprint colonne
- Pour que la purchased colonne soit une cleartext colonne

Pour générer un schéma de chiffrement avec des cleartext colonnes sealedfingerprint, et

1. (Facultatif) Pour vous assurer que le c3r-cli.jar fichier et le fichier à chiffrer sont présents :
 - a. Naviguez jusqu'au répertoire souhaité et exécutez `ls` (si vous utilisez a Mac ou Unix/Linux) ou `dir` si vous utilisez Windows).
 - b. Consultez la liste des fichiers de données tabulaires (.csv) et choisissez un fichier à chiffrer.

Dans cet exemple, sales.csv c'est le fichier que nous voulons chiffrer.

2. À partir de la CLI, exécutez la commande suivante pour créer un schéma de manière interactive.

```
java -jar c3r-cli.jar schema sales.csv --interactive --  
output=sales.json
```

 Note

- L'`--interactive` indicateur sélectionne le mode interactif pour développer le schéma. Cela guide l'utilisateur à travers un flux de travail guidé pour créer le schéma.

- Si vous êtes un utilisateur avancé, vous pouvez créer votre propre schéma JSON sans utiliser le flux de travail guidé. Pour de plus amples informations, veuillez consulter [\(Facultatif\) Créez un schéma \(utilisateurs avancés\)](#).
- Pour les fichiers .csv sans en-têtes de colonne, consultez l'option `--noHeaders` indicateur de la commande de schéma disponible dans la CLI.
- L'option `--output` définit un nom de sortie. Si vous n'incluez pas l'option `--output`, le client de chiffrement C3R essaie de choisir un nom de sortie par défaut (tel que `<input>.out` ou pour le schéma `<input>.json`).

3. Pour `Number of target columns from source column 'username'?`, entrez, **1** puis appuyez sur Entrée.
4. Pour `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, entrez, **f** puis appuyez sur Entrée.
5. Pour `Target column headername <default 'username'>`, appuyez sur Entrée.

Le nom par défaut « username » est utilisé.

6. Pour `Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'>`, entrez, **y** puis appuyez sur Entrée.
7. Pour `Number of target columns from source column 'purchased'?`, entrez, **1** puis appuyez sur Entrée.
8. Pour `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, entrez, **c** puis appuyez sur Entrée.
9. Pour `Target column headername <default 'purchased'>`, appuyez sur Entrée.

Le nom par défaut « purchased » est utilisé.

10. Pour `Number of target columns from source column 'product'?`, entrez, **1** puis appuyez sur Entrée.
11. Pour `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, entrez, **s** puis appuyez sur Entrée.
12. Pour `Target column headername <default 'product'>`, appuyez sur Entrée.

Le nom par défaut « product » est utilisé.

13. Pour `'product_sealed' padding type: [n]one, [f]ixed, or [m]ax <default 'max'>`, appuyez sur Entrée pour choisir la valeur par défaut.

14. Pour sélectionner la valeur par défaut, Byte-length beyond max length to pad cleartext to in 'product_sealed' <default '0'>? appuyez sur Entrée.

Le schéma est écrit dans un nouveau fichier appelé `sales.json`.

15. Vous êtes maintenant prêt à [chiffrer les données](#).

Étape 5 : Création d'une clé secrète partagée

Pour chiffrer les tables de données, les participants à la collaboration doivent s'entendre sur une clé secrète partagée et la partager en toute sécurité.

La clé secrète partagée doit être d'au moins 256 bits (32 octets). Vous pouvez spécifier une clé plus grande, mais cela ne vous apportera aucune sécurité supplémentaire.

Important

N'oubliez pas que la clé et l'identifiant de collaboration utilisés pour le chiffrement et le déchiffrement doivent être identiques pour tous les participants à la collaboration.

Les sections suivantes fournissent des exemples de commandes de console permettant de générer une clé secrète partagée enregistrée `secret.key` dans le répertoire de travail actuel du terminal concerné.

Rubriques

- [Exemple : génération de clés à l'aide de OpenSSL](#)
- [Exemple : génération de clés lors de l'utilisation PowerShell](#)

Exemple : génération de clés à l'aide de OpenSSL

Pour une bibliothèque de cryptographie à usage général, exécutez la commande suivante pour créer une clé secrète partagée.

```
openssl rand 32 > secret.key
```

Si vous utilisez Windows et n'en avez pas OpenSSL installé, vous pouvez générer des clés à l'aide de l'exemple décrit dans [Exemple : génération de clés lors de l'utilisation PowerShell](#).

Exemple : génération de clés lors de l'utilisation PowerShell

Pour PowerShell une application de terminal disponible sur Windows, exécutez la commande suivante pour créer une clé secrète partagée.

```
$bs = New-Object Byte[](32);  
[Security.Cryptography.RandomNumberGenerator]::Create().GetBytes($bs); Set-  
Content 'secret.key' -Encoding Byte -Value $bs
```

Étape 6 : Stocker la clé secrète partagée dans une variable d'environnement

Une variable d'environnement est un moyen pratique et extensible pour les utilisateurs de fournir une clé secrète provenant de différents magasins de clés, par exemple, AWS Secrets Manager et de la transmettre au client de chiffrement C3R.

Le client de chiffrement C3R peut utiliser des clés stockées dans Services AWS si vous utilisez le AWS CLI pour stocker ces clés dans la variable d'environnement correspondante. Par exemple, le client de chiffrement C3R peut utiliser une clé provenant de AWS Secrets Manager. Pour plus d'informations, voir [Création et gestion de secrets AWS Secrets Manager](#) dans le Guide de AWS Secrets Manager l'utilisateur.

Note

Cependant, avant d'utiliser un Service AWS tel AWS Secrets Manager pour maintenir vos clés C3R, vérifiez que votre cas d'utilisation le permet. Certains cas d'utilisation peuvent nécessiter que la clé ne soit pas divulguée. AWS Cela permet de garantir que les données cryptées et la clé ne sont jamais détenues par le même tiers.

Les seules conditions requises pour une clé secrète partagée sont que la clé secrète partagée soit base64 codée et stockée dans la variable C3R_SHARED_SECRET d'environnement.

Les sections suivantes décrivent les commandes de console permettant de convertir un secret.key fichier en variable d'environnement base64 et de le stocker en tant que variable d'environnement. Le secret.key fichier peut avoir été généré à partir de l'une des commandes répertoriées dans [Étape 5 : Création d'une clé secrète partagée](#) et n'est qu'un exemple de source.

Stocker la clé dans une variable d'environnement lors de Windows l'utilisation PowerShell

Pour convertir base64 et définir la variable d'environnement lors de Windows l'utilisation PowerShell, exécutez la commande suivante.

```
$Bytes=[IO.File]::ReadAllBytes((Get-Location).ToString()+'\secret.key');  
$env:C3R_SHARED_SECRET=[Convert]::ToBase64String($Bytes)
```

Stocker la clé dans une variable d'environnement sur Linux ou macOS

Pour convertir base64 et définir la variable d'environnement sur Linux ou macOS, exécutez la commande suivante.

```
export C3R_SHARED_SECRET="$(cat secret.key | base64)"
```

Étape 7 : Chiffrer les données

Pour effectuer cette étape, vous devez acquérir l'ID de AWS Clean Rooms collaboration et la clé secrète partagée. Pour de plus amples informations, veuillez consulter les [Prérequis](#).

Dans l'exemple suivant, nous exécutons le `ads.csv` chiffrement en utilisant le schéma que nous avons créé appelé `ads.json`.

Pour chiffrer des données

1. Stockez la clé secrète partagée pour la collaboration dans [Étape 6 : Stocker la clé secrète partagée dans une variable d'environnement](#).
2. Sur la ligne de commande, entrez la commande suivante.

```
java -jar c3r-cli.jar encrypt <name of input .csv file> --schema=<name of schema .json file> --id=<collaboration id> --output=<name of output.csv file> <optional flags>
```

3. Pour `<name of input .csv file>`, entrez le nom du fichier .csv d'entrée.
4. Pour `schema=`, entrez le nom du fichier de schéma de chiffrement .json.
5. Pour `id=`, entrez l'ID de collaboration.
6. Pour `output=`, entrez le nom du fichier de sortie (par exemple, `ads-output.csv`).

7. Incluez l'un des indicateurs de ligne de commande décrits dans [Paramètres de calcul cryptographique](#) et [Drapeaux facultatifs dans le calcul cryptographique pour Clean Rooms](#).
8. Exécutez la commande .

Dans l'exemple `deads.csv`, nous exécutons la commande suivante.

```
java -jar c3r-cli.jar encrypt ads.csv --schema=ads.json --id=123e4567-e89b-42d3-a456-556642440000 --output=ads-output.csv
```

Dans l'exemple `desales.csv`, nous exécutons la commande suivante.

```
java -jar c3r-cli.jar encrypt sales.csv --schema=sales.json --id=123e4567-e89b-42d3-a456-556642440000
```

Note

Dans cet exemple, nous ne spécifiez pas de nom de fichier de sortie (`--output=sales-output.csv`). Par conséquent, le nom du fichier de sortie par défaut `name-of-file.out.csv` a été généré.

Vous êtes maintenant prêt à vérifier les données cryptées.

Étape 8 : vérifier le chiffrement des données

Pour vérifier que les données ont été cryptées

1. Affichez le fichier de données crypté (par exemple, `sales-output.csv`).
2. Vérifiez les colonnes suivantes :
 - a. Colonne 1 — Chiffré (par exemple, `username_fingerprint`).
 - b. Colonne 2 — Non chiffré (par exemple, `purchased`).
 - c. Colonne 3 — Chiffré (par exemple, `product_sealed`).

Pour les colonnes cryptées (SELECT), la longueur du cleartext plus tout remplissage après le préfixe de version et de type (par exemple, `01:enc:`) est directement proportionnelle à

la longueur de la cleartext colonne cryptée. En d'autres termes, la longueur correspond à la taille de l'entrée plus environ 33 % de surcharge due au codage.

Vous êtes maintenant prêt à :

1. [Téléchargez les données cryptées sur S3.](#)
2. [Créez une AWS Glue table.](#)
3. [Créez une table configurée dans AWS Clean Rooms.](#)

Le client de chiffrement C3R créera des fichiers temporaires qui ne contiennent pas de données non chiffrées (à moins que ces données ne soient également déchiffrées dans la sortie finale). Cependant, certaines valeurs cryptées peuvent ne pas être correctement renseignées. Les colonnes d'empreintes digitales peuvent contenir des valeurs dupliquées, même si le paramètre de collaboration `allowRepeatedFingerprintValue` est `false`. Ce problème se produit parce que le fichier temporaire est écrit avant que les longueurs de remplissage appropriées et les propriétés de suppression des doublons ne soient vérifiées.

Si le client de chiffrement C3R échoue ou est interrompu pendant le chiffrement, il peut s'arrêter après l'écriture du fichier temporaire, mais avant de vérifier ces propriétés et de supprimer les fichiers temporaires. Il se peut donc que ces fichiers temporaires soient toujours sur le disque. Dans ce cas, le contenu de ces fichiers ne protège pas les données en texte brut au même niveau que la sortie. En particulier, ces fichiers temporaires peuvent révéler des données en texte brut pour des analyses statistiques qui ne nuiraient pas au résultat final. L'utilisateur doit supprimer ces fichiers (en particulier une SQLite base de données) pour éviter qu'ils ne tombent entre des mains non autorisées.

(Facultatif) Créez un schéma (utilisateurs avancés)

La création manuelle d'un schéma est réservée aux utilisateurs expérimentés.

Vous trouverez ci-dessous une description du format de fichier de schéma JSON pour les fichiers d'entrée avec ou sans en-têtes de colonne. Les utilisateurs avancés peuvent directement écrire ou modifier le schéma s'ils le souhaitent.

Note

Le client de chiffrement C3R peut vous aider à créer un schéma par le biais du processus interactif décrit dans [Exemple : génération d'un schéma de chiffrement avec des cleartext colonnes sealedfingerprint](#), et ou par la création d'un modèle de stub.

Schémas de tables cartographiées et positionnelles

La section suivante décrit deux types de schémas de table :

- Schéma de table mappé — Ce schéma est utilisé pour chiffrer les fichiers .csv avec une ligne d'en-tête et des fichiers. Apache Parquet
- Schéma de table positionnelle — Ce schéma est utilisé pour chiffrer des fichiers .csv sans ligne d'en-tête.

Le client de chiffrement C3R peut chiffrer un fichier tabulaire pour une collaboration. Pour ce faire, il doit disposer d'un fichier de schéma correspondant qui indique comment la sortie cryptée doit être dérivée de l'entrée.

Le client de chiffrement C3R peut aider à générer un schéma pour un INPUT fichier en exécutant la commande C3R encryption client schema sur la ligne de commande. Voici un exemple de commande `java -jar c3r-cli.jar schema --interactive INPUT`.

Le schéma indique les informations suivantes :

1. Quelles colonnes source correspondent à quelles colonnes transformées dans le fichier de sortie par le biais de leur nom d'en-tête (schémas mappés) ou de leur position (schémas positionnels)
2. Quelles colonnes cibles doivent rester cleartext
3. Quelles colonnes cibles doivent être cryptées pour les SELECT requêtes
4. Quelles colonnes cibles doivent être cryptées pour les JOIN requêtes

Ces informations sont codées dans un fichier de schéma JSON spécifique à une table, composé d'un seul objet dont `headerRow` le champ est une valeur booléenne. La valeur doit être `true` pour Parquet les fichiers et les fichiers .csv avec une ligne d'en-tête, et `false` sinon.

Schéma de table mappé

Le schéma mappé a la forme suivante.

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": STRING,
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    },
    ...
  ]
}
```

Si `headerRow` est le `true`, le champ suivant de l'objet contient un tableau de schémas de colonnes qui mappent les en-têtes source aux en-têtes cibles (c'est-à-dire des objets JSON décrivant ce que les colonnes de sortie doivent contenir). `columns`

- `sourceHeader`— Le nom d'STRING en-tête de la colonne source dont les données sont dérivées.

Note

La même colonne source peut être utilisée pour plusieurs colonnes cibles. Une colonne du fichier d'entrée qui n'est répertoriée `sourceHeader` nulle part dans le schéma n'apparaît pas dans le fichier de sortie.

- `targetHeader`— Le nom d'STRING en-tête de la colonne correspondante dans le fichier de sortie.

Note

Ce champ est facultatif pour les schémas mappés. Si ce champ est omis, il `sourceHeader` est réutilisé comme nom d'en-tête dans la sortie. L'`_fingerprint` ou l'autre `_sealed` est ajouté si la colonne de sortie est une fingerprint colonne ou une sealed colonne respectivement.

- **type**— Celui TYPE de la colonne cible dans le fichier de sortie. C'est-à-dire l'une des options `cleartext` ou `fingerprint` selon la manière dont la colonne sera utilisée dans le cadre de la collaboration. `sealed`
- **pad**— Champ d'un objet de schéma de colonne qui n'est présent que lorsque TYPE c'est le `sealed`. La valeur correspondante de PAD est un objet qui décrit la manière dont les données doivent être remplies avant d'être cryptées.

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

Pour spécifier le rembourrage avant le chiffrement, `type` `length` ils sont utilisés comme suit :

- `PAD_TYPE`as `none` — Aucun remplissage ne sera appliqué aux données de la colonne et le `length` champ n'est pas applicable (c'est-à-dire `omis`).
- `PAD_TYPE`as `fixed` — Les données de la colonne sont complétées au nombre `length` d'octets spécifié.
- `PAD_TYPE`as `max` — Les données de la colonne sont complétées à la taille de l'octet de la valeur la plus longue plus un `length` octet supplémentaire.

Voici un exemple de schéma mappé, avec une colonne de chaque type.

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": "FullName",
      "targetHeader": "name",
      "type": "cleartext"
    },
    {
      "sourceHeader": "City",
      "targetHeader": "city_sealed",
      "type": "sealed",
      "pad": {
        "type": "max",
        "length": 16
      }
    }
  ],
}
```

```
{
  "sourceHeader": "PhoneNumber",
  "targetHeader": "phone_number_fingerprint",
  "type": "fingerprint"
},
{
  "sourceHeader": "PhoneNumber",
  "targetHeader": "phone_number_sealed",
  "type": "sealed",
  "pad": {
    "type": "fixed",
    "length": 20
  }
}
]
```

À titre d'exemple plus complexe, voici un exemple de fichier .csv avec des en-têtes.

```
FirstName,LastName,Address,City,State,PhoneNumber,Title,Level,Notes
Jorge,Souza,12345 Mills Rd,Anytown,SC,703-555-1234,CEO,10,
Paulo,Santos,0 Street,Anytown,MD,404-555-111,CIO,9,This is a really long note that
could really be a paragraph
Mateo,Jackson,1 Two St,Anytown,NY,304-555-1324,COO,9,""
Terry,Whitlock4 N St,Anytown,VA,407-555-8888,EA,7,Secret notes
Diego,Ramirez,9 Hollows Rd,Anytown,VA,407-555-1222,SDE I,4,null
John,Doe,8 Hollows Rd,Anytown,VA,407-555-4321,SDE I,4,Jane's younger brother
Jane,Doe,8 Hollows Rd,Anytown,VA,407-555-4322,SDE II,5,John's older sister
```

Dans l'exemple de schéma mappé suivant, les colonnes `FirstName` et `LastName` sont des cleartext colonnes. La `State` colonne est cryptée en tant que `fingerprint` colonne et en tant que `sealed` colonne avec un rembourrage `denone`. Les autres colonnes sont omises.

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": "FirstName",
      "targetHeader": "GivenName",
      "type": "cleartext"
    },
    {
```



```

    "sourceHeader": "LastName",
    "targetHeader": "Surname",
    "type": "cleartext"
  },
  {
    "sourceHeader": "State",
    "targetHeader": "State_Join",
    "type": "fingerprint"
  },
  {
    "sourceHeader": "State",
    "targetHeader": "State",
    "type": "sealed",
    "pad": {
      "type": "none"
    }
  }
]
}

```

Le fichier .csv qui résulte du schéma mappé est le suivant.

```

givenname,surname,state_fingerprint,state
John,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:FQ3n3Ahv9BQQNWQGcugeHzHYzEZE1vapHa2Uu4SRgSATZ3q0bjPA4TcsHt
+B0kMKBcnHWI13BeGG/SBqmj7vKpI=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:KZ5n5GtaXACco65AXk48BQ02durDNR2ULc4YxmMC8NaZZKKJiksU1IwFadAvV4iBQ1
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:mLKpS5HIOSgphdEsrzhdEd
eN9nB02gAbIygt40Fn4La1Yn9Xyj/XUWXlmn8zFe2T4kyDTD8kG0vpQEUGxAUFk=
Diego,Ramirez,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:rmZhT98Zm
+IIGw1UTjMIJP4IrW/AA1tBLMXcHvnyYfRgmWP623VFQ6aUnhsb2MDqEw4G5Uwg5rKKZepUxx5uKbfbk=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBBYnqQehBVqhN0d7s2ZiKue7QiTy08=,01:enc:vVaqWC1VRbhvkf8gnuR7q0z
Terry,Whitlock01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:3c9VEWb0D0/
xbQjdGuccLvI7oZTBdPU+SyrJIyr2kudfAxbuMQ2uRdU/q7rbgyJjxZS8M2U35ILJf/1DgTyg7cM=
Jane,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:9RWv46YLveykeNZ/
G0Nd1YFg+AVD0nu05hHyAYTQkPLHnyX+0/jbzD/g9ZT8GCgVE9aB5bV4ooJIXHGBVMXcjrQ=

```

Schéma de tableau positionnel

Le schéma positionnel a la forme suivante.

```
{
```

```

"headerRow": false,
"columns": [
  [
    {
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    },
    {
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    }
  ],
  [],
  ...
]
}

```

Si `headerRow` tel est le cas `false`, le champ suivant de l'objet est `columns`, qui contient un tableau d'entrées. Chaque entrée est elle-même un tableau de zéro ou plusieurs schémas de colonnes positionnels (aucun `sourceHeader` champ), qui sont des objets JSON décrivant ce que la sortie doit contenir.

- `sourceHeader`— Le nom d'`STRING` en tête de la colonne source dont les données sont dérivées.

Note

Ce champ doit être omis dans les schémas de position. Dans les schémas positionnels, la colonne source est déduite par l'index correspondant de la colonne dans le fichier de schéma.

- `targetHeader`— Le nom d'`STRING` en tête de la colonne correspondante dans le fichier de sortie.

Note

Ce champ est obligatoire pour les schémas de position.

- `type`— Celui `TYPE` de la colonne cible dans le fichier de sortie. C'est-à-dire l'une des options `cleartext` ou `fingerprint` selon la manière dont la colonne sera utilisée dans le cadre de la collaboration. `sealed`

- `pad`— Champ d'un objet de schéma de colonne qui n'est présent que lorsque `TYPE` c'est le `cassealed`. La valeur correspondante de `PAD` est un objet qui décrit la manière dont les données doivent être remplies avant d'être cryptées.

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

Pour spécifier le rembourrage avant le chiffrement, `type` `length` ils sont utilisés comme suit :

- `PAD_TYPE`as `none` — Aucun remplissage ne sera appliqué aux données de la colonne et le `length` champ n'est pas applicable (c'est-à-dire `omis`).
- `PAD_TYPE`as `fixed` — Les données de la colonne sont complétées au nombre `length` d'octets spécifié.
- `PAD_TYPE`as `max` — Les données de la colonne sont complétées à la taille de l'octet de la valeur la plus longue plus un `length` octet supplémentaire.

Note

`fixed` est utile si vous connaissez à l'avance la limite supérieure de la taille en octets des données de la colonne. Une erreur est générée si les données de cette colonne sont plus longues que celles spécifiées `length`.

`max` est pratique lorsque la taille exacte des données d'entrée est inconnue, car elle fonctionne quelle que soit la taille des données. Cependant, `max` cela nécessite un temps de traitement supplémentaire car il chiffre les données deux fois. `max` chiffre les données une fois lorsqu'elles sont lues dans le fichier temporaire et une fois que l'entrée de données la plus longue dans la colonne est connue.

De plus, la longueur de la valeur la plus longue n'est pas enregistrée entre les appels du client. Si vous prévoyez de chiffrer vos données par lots, ou de chiffrer régulièrement de nouvelles données, sachez que la longueur du texte chiffré peut varier d'un lot à l'autre.

Voici un exemple de schéma positionnel.

```
{
  "headerRow": false,
  "columns": [
```

```
[
  {
    "targetHeader": "name",
    "type": "cleartext"
  }
],
[
  {
    "targetHeader": "city_sealed",
    "type": "sealed",
    "pad": {
      "type": "max",
      "length": 16
    }
  }
],
[
  {
    "targetHeader": "phone_number_fingerprint",
    "type": "fingerprint"
  },
  {
    "targetHeader": "phone_number_sealed",
    "type": "sealed",
    "pad": {
      "type": "fixed",
      "length": 20
    }
  }
]
]
```

À titre d'exemple complexe, voici un exemple de fichier .csv s'il ne contient pas la première ligne avec les en-têtes.

```
Jorge,Souza,12345 Mills Rd,Anytown,SC, 703 -555 -1234,CEO, 10,
Paulo,Santos, 0 Street,Anytown,MD, 404-555-111,CIO, 9,This is a really long note that
could really be a paragraph
Mateo,Jackson, 1 Two St,Anytown,NY, 304-555-1324,C00, 9, ""
Terry,Whitlock, 4 N St,Anytown,VA, 407-555-8888,EA, 7,Secret notes
Diego,Ramirez, 9 Hollows Rd,Anytown,VA, 407-555-1222,SDE I, 4,null
John,Doe, 8 Hollows Rd,Anytown,VA, 407-555-4321,SDE I, 4,Jane's younger brother
```

Jane,Doe, 8 Hollows Rd,Anytown,VA, 407-555-4322,SDE II, 5,John's older sister

Le schéma positionnel se présente sous la forme suivante.

```
{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": "GivenName",
        "type": "cleartext"
      }
    ],
    [
      {
        "targetHeader": "Surname",
        "type": "cleartext"
      }
    ],
    [],
    [],
    [
      {
        "targetHeader": "State_Join",
        "type": "fingerprint"
      },
      {
        "targetHeader": "State",
        "type": "sealed",
        "pad": {
          "type": "none"
        }
      }
    ],
    [],
    [],
    [],
    []
  ]
}
```

Le schéma précédent produit le fichier de sortie suivant avec une ligne d'en-tête contenant les en-têtes cibles spécifiés.

```

givenname,surname,state_fingerprint,state
Mateo,Jackson,01: hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01: enc:ENS6QD3cMV19vQEGfe9MN
Q8m/Y5SA89dJwKpT5rGpP8e36h6klwDoslpFzGvU0=
Jorge,Souza,01: hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhN0d7s2ZiKUe7QiTy08=,01: enc:LKo0zirq2+
+XEIIIMNRjAsGmdyWUDwYaum0B+IFP+rUf1BNeZDJjtFe1Z+zbZfXQWwJy52Rt7HqvAb2WIK1oMmk=
Paulo,Santos,01: hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01: enc:MyQKyWxJ9kvK1xDQQtX1UNwv3F+y1BR10xrUY/1BGg5KFg0n9pK+MZ7g
+ZNqZEPcPz4lht1u0t/wbTaqz0CLXFQ=
Jane,Doe,01: hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01: enc:Pd8sbITBfb0/
ttUB4svVsGoYkDfnDvgkvxzeCi0Yxq54rLSwccy1o3/B50C3cpkkn56dovCwzgmmpNwrmCmYtb4=
Terry,Whitlock01: hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv
+1Mk=,01: enc:Qmtzu3B3GAXKh2KkRYTiEAaMopYedsSdF2e/
ADUiBQ9kv2CxCkPzWyYTD3ztmKPMka19dHre5VhUHNp030+j1AQ8=
Diego,Ramirez,01: hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01: enc:ysdg
+GHKdeZrS/geBIoo0EPLHG68MsWpx1dh3xjb+fG51rmFmqUcJLNuuYBHhHA1xchM2WVeV1fmHkBX3mvZNvkc=
John,Doe,01: hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01: enc:9uX0wZu07kAPAx
+Hf6uvQownkWqFSKtWS7gQIJSe5aXFquKWCK6yZN0X5Ea2N3bn03Uj1kh0agDWoiP9FRZGJA4=

```

Déchiffrer des tables de données avec le client de chiffrement C3R

Suivez cette procédure pour les collaborations qui utilisent l'informatique cryptographique pour Clean Rooms et le client de chiffrement C3R pour chiffrer les tables de données. Utilisez cette procédure après avoir demandé des [données dans le cadre de la collaboration](#).

La clé secrète partagée et l'identifiant de collaboration sont requis pour cette procédure.

Le membre qui peut recevoir les résultats déchiffre les données à l'aide de la même clé secrète partagée et du même identifiant de collaboration que ceux utilisés pour chiffrer les données de la collaboration.

Note

AWS Clean Rooms les collaborations limitent déjà les personnes autorisées à exécuter et à consulter les résultats des requêtes. Pour effectuer le déchiffrement, toute personne ayant accès à ces résultats a besoin de la même clé secrète partagée et du même identifiant de collaboration que ceux utilisés pour chiffrer les données.

Pour déchiffrer une table de données cryptée

1. (Facultatif) [Affichez les commandes disponibles dans le client de chiffrement C3R.](#)
2. (Facultatif) Accédez au répertoire souhaité et exécutez `ls` (macOS) ou `dir` (Windows).
 - Vérifiez que `c3r-cli.jar` le fichier et le fichier de données des résultats de requête chiffrés se trouvent dans le répertoire souhaité.

Note

Si les résultats de la requête sont téléchargés depuis l'interface de la AWS Clean Rooms console, ils se trouvent probablement dans le dossier Téléchargements de votre compte utilisateur. (Par exemple, le dossier Téléchargements de votre répertoire utilisateur sur Windows and macOS.) Nous vous recommandons de déplacer le fichier des résultats de la requête dans le même dossier que le `c3r-cli.jar`.

3. Stockez la clé secrète partagée dans la variable d'`C3R_SHARED_SECRET` environnement. Pour de plus amples informations, veuillez consulter [Étape 6 : Stocker la clé secrète partagée dans une variable d'environnement](#).
4. À partir du AWS Command Line Interface (AWS CLI), exécutez la commande suivante.

```
java -jar c3r-cli.jar decrypt <name of input .csv file> --id=<collaboration id> --output=<output file name>
```
5. Remplacez chacune *user input placeholder* par vos propres informations :
 - a. Pour `id=`, entrez l'ID de collaboration.
 - b. Pour `output=`, entrez le nom du fichier de sortie (par exemple, `results-decrypted.csv`).

Si vous ne spécifiez pas de nom de sortie, un nom par défaut est affiché dans le terminal.
 - c. Affichez les données déchiffrées dans le fichier de sortie spécifié à l'aide de votre fichier CSV préféré ou Parquet application de visualisation (telle que Microsoft Excel, un éditeur de texte ou une autre application).

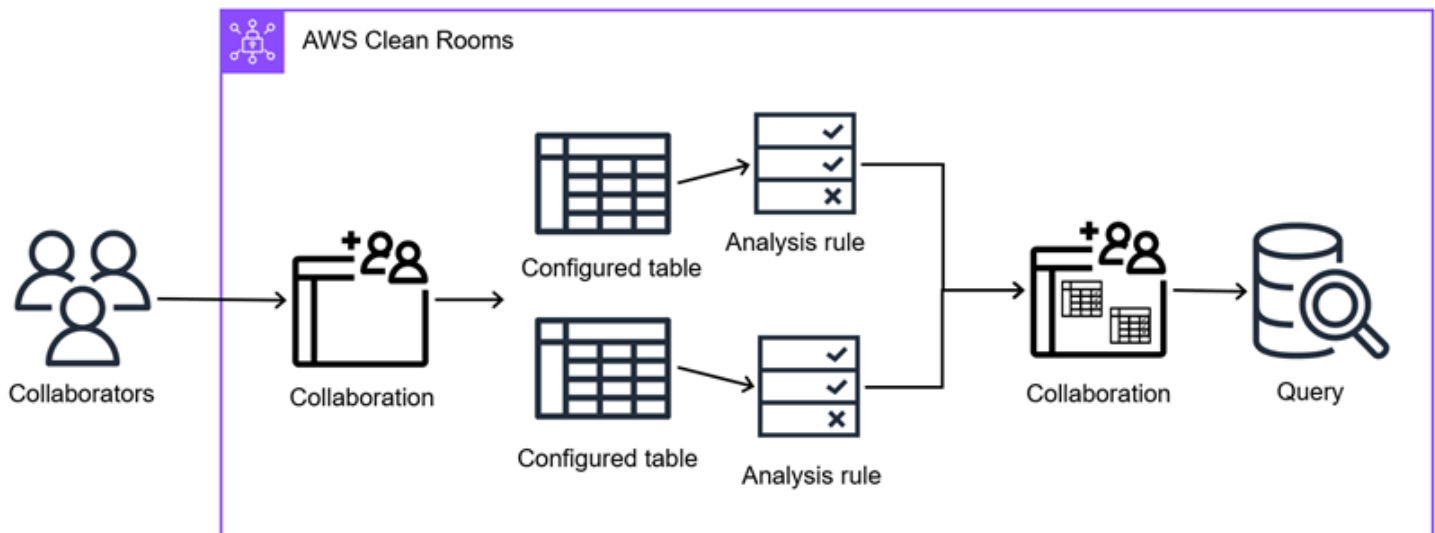
Tables configurées dans AWS Clean Rooms

Une table configurée est une référence à une table existante dans une source de données. Il contient une règle d'analyse qui détermine la manière dont les données peuvent être consultées AWS Clean Rooms et peut inclure un budget d'accès aux données pour contrôler l'utilisation des tables. Les tables configurées peuvent être associées à une ou plusieurs collaborations.

Avec AWS Clean Rooms, vous pouvez effectuer une analyse d'agrégation sur les données d'événements, telles que le nombre d'achats par rapport au nombre d'achats. Vous pouvez également effectuer une analyse de liste sur les données d'événements, par exemple en enrichissant les données clients qui se chevauchent, des données de segment aux données CRM. Vous pouvez également effectuer des requêtes personnalisées et définir une confidentialité différentielle sur les données d'événements, telles que les données d'audience et les attributs de segment. Pour chacun de ces types d'analyse, vous pouvez définir des budgets d'accès aux données afin de surveiller et de contrôler la quantité de données accessible par le biais de requêtes.

Tout d'abord, vous créez une collaboration AWS Clean Rooms et vous y ajoutez celle que Comptes AWS vous souhaitez inviter, ou vous rejoignez une collaboration à laquelle vous êtes invité en créant un abonnement. Ensuite, vous et l'autre membre de la collaboration créez des tables configurées. Vous ajoutez une règle d'analyse aux tables configurées (agrégation, liste ou personnalisée) et définissez éventuellement des budgets d'accès aux données. Vous associez ensuite les tables configurées à la collaboration. Enfin, le membre habilité à effectuer une requête exécute une requête dans les deux tables de données, ce qui consomme le budget d'accès aux données au fur et à mesure que les requêtes sont exécutées.

Le schéma suivant récapitule comment utiliser les données d'événements dans AWS Clean Rooms.



Rubriques

- [Création d'une table configurée dans AWS Clean Rooms](#)
- [Ajouter une règle d'analyse à une table configurée](#)
- [Associer une table configurée à une collaboration](#)
- [Configuration d'un budget d'accès aux données](#)
- [Ajouter une règle d'analyse de collaboration à une table configurée](#)
- [Configuration d'une politique de confidentialité différentielle \(facultatif\)](#)
- [Afficher les tables et les règles d'analyse](#)
- [Modification d'une table configurée](#)
- [Modification des balises de tableau configurées](#)
- [Modification de la règle d'analyse de table configurée](#)
- [Suppression de la règle d'analyse de table configurée](#)
- [Colonnes interdites dans le tableau configuré](#)
- [Modification des associations de tables configurées](#)
- [Dissociation des tables configurées](#)

Création d'une table configurée dans AWS Clean Rooms

Une table configurée est une référence à une table existante dans une source de données. Il contient une règle d'analyse qui détermine la manière dont les données peuvent être consultées. AWS Clean Rooms Les tables configurées peuvent être associées à une ou plusieurs collaborations.

Pour plus d'informations sur la création d'une table configurée à l'aide du AWS SDKs, consultez la [référence de l'AWS Clean Rooms API](#).

Rubriques

- [Création d'une table configurée — Source de données Amazon S3](#)
- [Création d'une table configurée — Source de données Amazon Athena](#)
- [Création d'une table configurée — Source de données Snowflake](#)

Création d'une table configurée — Source de données Amazon S3

Dans cette procédure, le [membre](#) effectue les tâches suivantes :

- Configure une AWS Glue table existante à utiliser dans AWS Clean Rooms (Cette étape peut être effectuée avant ou après avoir rejoint une collaboration, sauf si vous utilisez l'informatique cryptographique pour Clean Rooms.)

Note

AWS Clean Rooms supporte AWS Glue les tables. Pour plus d'informations sur l'introduction de vos données AWS Glue, consultez [Étape 3 : Chargez votre tableau de données sur Amazon S3](#).

- Nomme la [table configurée](#) et choisit les colonnes à utiliser dans la collaboration.

La procédure suivante suppose que :

- Le membre de la collaboration a déjà [chargé ses tables de données sur Amazon S3](#) et en [a créé une AWS Glue](#).

Note

La destination des résultats dans Amazon S3 ne peut pas se trouver dans le même compartiment S3 que n'importe quelle source de données.

- (Facultatif) Pour les tables de données [chiffrées](#) uniquement, le membre de la collaboration a déjà [préparé des tables de données chiffrées](#) à l'aide du client de chiffrement C3R.

Vous pouvez utiliser la génération de statistiques fournie par AWS Glue pour calculer les statistiques au niveau des colonnes pour les tables. AWS Glue Data Catalog Après avoir AWS Glue généré des statistiques pour les tables du catalogue de données, Amazon Redshift Spectrum utilise automatiquement ces statistiques pour optimiser le plan de requête. Pour plus d'informations sur le calcul des statistiques au niveau des colonnes à l'aide de statistiques AWS Glue, consultez la section [Optimisation des performances des requêtes à l'aide des statistiques des colonnes](#) dans le Guide de l'AWS Glue utilisateur. Pour plus d'informations AWS Glue, consultez le manuel [AWS Glue Developer Guide](#).

Pour créer une table configurée — Source de données Amazon S3

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Dans le coin supérieur droit, choisissez Configurer une nouvelle table.
4. Pour Source de données, sous Sources de AWS données, choisissez Amazon S3.
5. Dans le tableau Amazon S3 :
 - a. Sélectionnez la région dans laquelle la table S3 est hébergée.

Par défaut, la région actuelle (telle que North Virginia us-east-1) est sélectionnée.

 Warning

Lorsque votre source de données Amazon S3 se trouve dans une région différente de celle de votre lieu de traitement, le traitement des données peut avoir lieu temporairement en dehors de la région source. Avant de poursuivre, vérifiez que le transfert de données entre régions est conforme à vos exigences en matière de souveraineté des données, à vos politiques de conformité réglementaire et à vos normes de gouvernance des données.

Pour plus d'informations sur les régions, voir [Régions et points de terminaison](#) dans le Références générales AWS.

- b. Choisissez la base de données dans la liste déroulante.
- c. Choisissez la table que vous souhaitez configurer dans la liste déroulante.

 Note

Pour vérifier que ce tableau est correct, effectuez l'une des opérations suivantes :

- Choisissez Afficher dans AWS Glue.
- Activez Afficher le schéma depuis AWS Glue pour afficher le schéma.

 Important

Pour AWS Glue les tableaux dont les données sont au format CSV, les noms et l'ordre des colonnes dans le schéma Glue doivent correspondre exactement aux données CSV. S'ils ne sont pas alignés, la liste des colonnes autorisées pour la table configurée risque de ne pas être appliquée correctement.

6. Pour les colonnes et les méthodes d'analyse autorisées dans les collaborations,
 - a. Pour quelles colonnes souhaitez-vous autoriser les collaborations ?
 - Choisissez Toutes les colonnes pour autoriser toutes les colonnes à être interrogées dans le cadre de la collaboration.
 - Choisissez Liste personnalisée pour autoriser une ou plusieurs colonnes de la liste déroulante Spécifier les colonnes autorisées à être interrogées dans le cadre de la collaboration.
 - b. Pour les méthodes d'analyse autorisées,
 - i. Choisissez Requête directe pour autoriser les requêtes SQL à être exécutées directement sur cette table
 - ii. Choisissez Tâche directe pour autoriser les PySpark tâches à être exécutées directement sur cette table.

Exemple Exemple

Par exemple, si vous souhaitez autoriser les membres de la collaboration à exécuter à la fois des requêtes SQL directes et des PySpark tâches sur toutes les colonnes, choisissez Toutes les colonnes, Requête directe et Tâche directe.

7. Pour les détails de la table configurée,

- a. Entrez un nom pour la table configurée.

Vous pouvez utiliser le nom par défaut ou renommer cette table.

- b. Entrez une description de la table.

La description permet de différencier les autres tables configurées portant des noms similaires.

8. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
9. Choisissez Configurer une nouvelle table.

Maintenant que vous avez créé une table configurée, vous êtes prêt à :

- [Ajouter une règle d'analyse à la table configurée](#)
- [Associer la table configurée à une collaboration](#)

Création d'une table configurée — Source de données Amazon Athena

L'option de source de données Amazon Athena vous permet d'interroger les données stockées dans Amazon S3, cataloguées dans le catalogue de AWS Glue données ou dans des catalogues fédérés, et d'en contrôler l'accès via AWS Lake Formation. Les tables et les AWS Glue Data Catalog vues sont prises en charge. Les liens vers les ressources de Lake Formation peuvent être utilisés pour partager des tableaux Comptes AWS et des vues sur Régions AWS le compte du AWS Clean Rooms membre qui les associe à une AWS Clean Rooms collaboration.

Note

Seuls les ensembles de données basés sur Amazon S3 peuvent être interrogés via l'intégration des sources de données Athena.

Dans cette procédure, le [membre](#) effectue les tâches suivantes :

- Configure une table ou une vue existante dans le AWS Glue Data Catalog AWS Clean Rooms
- Nomme la [table configurée](#) et choisit les colonnes à utiliser dans la collaboration.

La procédure suivante suppose que :

- Le membre de la collaboration a déjà créé la AWS Glue Data Catalog base de données et la table ou la vue GDC.

Pour créer une table configurée : source de données Athena

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Dans le coin supérieur droit, choisissez Configurer une nouvelle table.
4. Dans Source de données, sous Sources de AWS données, sélectionnez Amazon Athena.
5. Sous le tableau Amazon Athena :
 - a. Sélectionnez la région dans laquelle la table Amazon Athena est hébergée.

Par défaut, la région actuelle (telle que North Virginia us-east-1) est sélectionnée.

 Warning

Lorsque votre source de données Amazon Athena se trouve dans une région différente de celle de votre lieu de traitement, le traitement des données peut avoir lieu temporairement en dehors de la région source. Avant de poursuivre, vérifiez que le transfert de données entre régions est conforme à vos exigences en matière de souveraineté des données, à vos politiques de conformité réglementaire et à vos normes de gouvernance des données.

Pour plus d'informations sur les régions, voir [Régions et points de terminaison](#) dans le Références générales AWS.

- b. Choisissez le catalogue dans la liste déroulante.

Par défaut, le catalogue de AWS Glue données est sélectionné.

- AWS Glue Catalogue de données : catalogue par défaut pour les tables dans AWS Glue.
- Catalogue fédéré : disponible si vous avez configuré la fédération de AWS Glue catalogues pour vous connecter à des catalogues REST Apache Iceberg distants. Pour

plus d'informations, consultez la section [Fédération de catalogues](#) dans le guide du AWS Lake Formation développeur.

- c. Choisissez la base de données dans la liste déroulante.
- d. Choisissez la table que vous souhaitez configurer dans la liste déroulante.

 Note

Pour vérifier que ce tableau est correct, effectuez l'une des opérations suivantes :

- Choisissez Afficher dans AWS Glue ou Afficher dans AWS Lake Formation (selon le type de catalogue).
- Activez Afficher le schéma depuis AWS Glue pour afficher le schéma.

6. Pour les configurations Amazon Athena,
 - a. Choisissez un groupe de travail dans la liste déroulante.
 - b. Pour l'emplacement de sortie S3, choisissez une action recommandée, en fonction de l'un des scénarios suivants.

Scénario	Action recommandée
Votre groupe de travail n'a pas d'emplacement de sortie par défaut.	Entrez l'emplacement de sortie S3 ou choisissez Parcourir S3.
Votre groupe de travail applique votre emplacement de sortie par défaut.	L'emplacement de sortie S3 est automatiquement choisi et vous ne pouvez pas le modifier.
Votre groupe de travail n'applique pas votre emplacement de sortie par défaut.	Entrez l'emplacement de sortie S3 ou choisissez Parcourir S3.

7. Pour les colonnes autorisées dans les collaborations, choisissez une option en fonction de votre objectif.

Votre objectif	Option recommandée
Autoriser l'utilisation de toutes les colonnes dans AWS Clean Rooms (sous réserve des règles d'analyse)	Toutes les colonnes
Autoriser une ou plusieurs colonnes dans la liste déroulante Spécifier les colonnes autorisées	Liste personnalisée

8. Pour les détails de la table configurée,

a. Entrez un nom pour la table configurée.

Vous pouvez utiliser le nom par défaut ou renommer cette table.

b. Entrez une description de la table.

La description permet de différencier les autres tables configurées portant des noms similaires.

c. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.

9. Choisissez Configurer une nouvelle table.

Maintenant que vous avez créé une table configurée, vous êtes prêt à :

- [Ajouter une règle d'analyse à la table configurée](#)
- [Associer la table configurée à une collaboration](#)

Création d'une table configurée — Source de données Snowflake

Dans cette procédure, le [membre](#) effectue les tâches suivantes :

- Configure une table Snowflake existante à utiliser dans AWS Clean Rooms (Cette étape peut être effectuée avant ou après avoir rejoint une collaboration, sauf si vous utilisez l'informatique cryptographique pour Clean Rooms.)
- Nomme la [table configurée](#) et choisit les colonnes à utiliser dans la collaboration.

La procédure suivante suppose que :

- Le membre de la collaboration a déjà téléchargé ses tableaux de données sur Snowflake.
- (Facultatif) Pour les tables de données [chiffrées](#) uniquement, le membre de la collaboration a déjà [préparé des tables de données chiffrées](#) à l'aide du client de chiffrement C3R.

Pour créer une table configurée : source de données Snowflake

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Dans le coin supérieur droit, choisissez Configurer une nouvelle table.
4. Dans Source de données, sous Clouds tiers et sources de données, choisissez Snowflake.
5. Spécifiez les informations d'identification Snowflake à l'aide d'un ARN secret existant ou en stockant un nouveau secret pour cette table.

Use existing secret ARN

1. Si vous avez un ARN secret, saisissez-le dans le champ ARN secret.

Vous pouvez rechercher votre ARN secret en choisissant Accéder à AWS Secrets Manager.

2. Si vous possédez un secret provenant d'une autre table, choisissez Importer l'ARN du secret à partir d'une table existante.

 Note

L'ARN secret peut être multicompte.

Store a new secret for this table

1. Entrez les informations d'identification Snowflake suivantes :
 - Nom d'utilisateur Snowflake
 - Entrepôt Snowflake
 - Rôle Snowflake

- Clé privée PEM (Snowflake Privacy Enhanced Mail)
2. Pour le chiffrement, effectuez l'une des opérations suivantes :
 - Pour utiliser le Clé gérée par AWS (par défaut), laissez la case Personnaliser les paramètres de chiffrement décochée.
 - Pour utiliser une personnalisation AWS KMS key :
 - Cochez la case Personnaliser les paramètres de chiffrement.
 - Pour la clé KMS, entrez l'ARN de la clé ou choisissez-en un dans la liste.
 3. Entrez un nom secret pour vous aider à retrouver vos informations d'identification ultérieurement.
6. Pour les détails de la table et du schéma Snowflake, entrez les détails manuellement ou importez-les automatiquement.

Enter the details manually

1. Entrez l'identifiant du compte Snowflake.

Pour plus d'informations, consultez la section [Identifiants de compte](#) dans la documentation de Snowflake.

L'identifiant de votre compte doit être au format utilisé pour les pilotes Snowflake. Vous devez remplacer le point (.) par un trait d'union (-) afin que l'identifiant soit formaté comme suit. **<orgname>-<account_name>**

2. Entrez dans la base de données Snowflake.

Pour plus d'informations, consultez la [base de données Snowflake](#) dans la documentation Snowflake.

3. Entrez le nom du schéma Snowflake.
4. Entrez le nom de la table Snowflake.

Pour plus d'informations, consultez la section [Comprendre les structures des tables Snowflake](#) dans la documentation Snowflake.

5. Pour le schéma, entrez le nom de la colonne et choisissez le type de données dans la liste déroulante.
6. Choisissez Ajouter une colonne pour ajouter d'autres colonnes.
 - Si vous choisissez un type de données d'objet, spécifiez le schéma d'objet.

Exemple Exemple de schéma d'objet

```
name STRING,  
location OBJECT(  
  x INT,  
  y INT,  
  metadata OBJECT(uuid STRING)  
),  
history ARRAY(TEXT)
```

- Si vous choisissez un type de données Array, spécifiez le schéma Array.

Exemple Exemple de schéma de tableau

```
OBJECT(x INT, y INT)
```

- Si vous choisissez un type de données cartographique, spécifiez le schéma cartographique.

Exemple Exemple de schéma cartographique

```
STRING, OBJECT(x INT, y INT)
```

Automatically import the details

1. Exportez votre vue COLUMNS depuis Snowflake sous forme de fichier CSV.

Pour plus d'informations sur la vue Snowflake COLUMNS, voir la [vue COLUMNS](#) dans la documentation Snowflake.

2. Choisissez Importer depuis un fichier pour importer le fichier CSV et spécifier toute information supplémentaire.

Le nom de la base de données, le nom du schéma, le nom de la table, les noms des colonnes et les types de données sont automatiquement importés.

- Si vous choisissez un type de données d'objet, spécifiez le schéma d'objet.
- Si vous choisissez un type de données Array, spécifiez le schéma Array.
- Si vous choisissez un type de données cartographique, spécifiez le schéma cartographique.

3. Entrez l'identifiant du compte Snowflake.

Pour plus d'informations, consultez la section [Identifiants de compte](#) dans la documentation de Snowflake.



Note

Seules les tables S3 cataloguées AWS Glue peuvent être utilisées pour récupérer le schéma de table automatiquement.

7. Pour les colonnes autorisées dans les collaborations, choisissez une option en fonction de votre objectif.

Votre objectif	Option recommandée
Autoriser l'utilisation de toutes les colonnes dans AWS Clean Rooms (sous réserve des règles d'analyse)	Toutes les colonnes
Autoriser une ou plusieurs colonnes dans la liste déroulante Spécifier les colonnes autorisées	Liste personnalisée

8. Pour les détails de la table configurée,

- a. Entrez un nom pour la table configurée.

Vous pouvez utiliser le nom par défaut ou renommer cette table.

- b. Entrez une description de la table.

La description permet de différencier les autres tables configurées portant des noms similaires.

- c. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.

9. Choisissez Configurer une nouvelle table.

Maintenant que vous avez créé une table configurée, vous êtes prêt à :

- [Ajouter une règle d'analyse à la table configurée](#)
- [Associer la table configurée à une collaboration](#)

Ajouter une règle d'analyse à une table configurée

Les sections suivantes décrivent comment ajouter une règle d'analyse à votre table configurée. En définissant les règles d'analyse, vous pouvez autoriser le membre autorisé à exécuter des requêtes correspondant à une règle d'analyse spécifique prise en charge par AWS Clean Rooms.

AWS Clean Rooms prend en charge les types de règles d'analyse suivants :

- [Règle d'analyse d'agrégation](#)
- [Règle d'analyse des listes](#)
- [Règle d'analyse personnalisée dans AWS Clean Rooms](#)

Il ne peut y avoir qu'une seule règle d'analyse par table configurée. Vous pouvez configurer la règle d'analyse à tout moment avant d'associer les tables configurées à la collaboration.

Important

Si vous utilisez l'informatique cryptographique pour la collaboration Clean Rooms et que vous avez chiffré des tables de données, la règle d'analyse que vous ajoutez à la table configurée cryptée doit être cohérente avec la manière dont les données ont été cryptées. Par exemple, si vous avez chiffré les données pour SELECT (règle d'analyse d'agrégation), vous ne devez pas ajouter la règle d'analyse pour JOIN (règle d'analyse de liste).

Rubriques

- [Ajouter une règle d'analyse d'agrégation à une table \(flux guidé\)](#)
- [Ajouter une règle d'analyse de liste à un tableau \(flux guidé\)](#)
- [Ajouter une règle d'analyse personnalisée à un tableau \(flux guidé\)](#)
- [Ajouter une règle d'analyse à une table \(éditeur JSON\)](#)
- [Étapes suivantes](#)

Ajouter une règle d'analyse d'agrégation à une table (flux guidé)

La règle d'analyse d'agrégation autorise les requêtes qui regroupent les statistiques sans révéler d'informations au niveau des lignes en utilisant COUNT et en utilisant des AVG dimensions facultatives. SUM

Cette procédure décrit le processus d'ajout d'une règle d'analyse d'agrégation à votre table configurée à l'aide de l'option Flux guidé de la AWS Clean Rooms console.

Note

Les tables configurées à l'aide de sources de données autres que S3 ne prennent en charge que [les règles d'analyse personnalisées](#).

Pour ajouter la règle d'analyse d'agrégation à une table (flux guidé)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez le tableau configuré.
4. Sur la page détaillée de la table configurée, choisissez Configurer la règle d'analyse.
5. Dans Étape 1 : Choisissez le type de règle d'analyse, sous Type de règle d'analyse, choisissez l'option Agrégation.
6. Sous Méthode de création, sélectionnez Flux guidé, puis Suivant.
7. Dans Étape 2 : Spécifier les contrôles de requête, pour les fonctions d'agrégation :
 - a. Choisissez une fonction d'agrégation dans le menu déroulant :
 - COMPTER
 - DÉCOMPTE DISTINCT
 - SUM
 - SOMME DISTINCTE
 - AVG
 - b. Choisissez les colonnes qui peuvent être utilisées dans la fonction d'agrégation dans le menu déroulant Colonnes.

- c. (Facultatif) Choisissez Ajouter une autre fonction pour ajouter une autre fonction d'agrégation et associer une ou plusieurs colonnes à cette fonction.

 Note

Au moins une fonction d'agrégation est requise.

- d. (Facultatif) Choisissez Supprimer pour supprimer une fonction d'agrégation.

8. Pour les commandes Join,

- a. Choisissez une option pour Autoriser la table elle-même à être interrogée :

Si vous choisissez...	Alors...
Non, seul le chevauchement peut être interrogé	La table ne peut être interrogée que lorsqu'elle est jointe à une table appartenant au membre qui peut effectuer la requête.
Oui	La table peut être interrogée seule ou lorsqu'elle est jointe à d'autres tables.

- b. Sous Spécifier les colonnes de jointure, choisissez les colonnes que vous souhaitez autoriser à utiliser dans l'INNERJOIN instruction.

Cette option est facultative si vous avez sélectionné Oui à l'étape précédente.

- c. Sous Spécifier les opérateurs autorisés pour la mise en correspondance, choisissez quels opérateurs, le cas échéant, peuvent être utilisés pour faire correspondre plusieurs colonnes de jointure. Si vous sélectionnez deux JOIN colonnes ou plus, l'un de ces opérateurs est requis.

Si vous choisissez...	Alors...
ET	Vous pouvez inclure AND dans le INNER JOIN match des conditions permettant de joindre une colonne à une autre entre les tables.

Si vous choisissez...	Alors...
OU	Vous pouvez inclure OR dans les conditions de INNER JOIN correspondance pour combiner plusieurs correspondances de colonnes entre les tables. Cet opérateur logique est utile pour obtenir un taux de correspondance plus élevé.

9. (Facultatif) Pour les contrôles de dimension, dans la liste déroulante Spécifier les colonnes de dimension, choisissez les colonnes que vous souhaitez autoriser à utiliser dans l'instruction SELECT, ainsi que dans les ORDER BY parties WHERE GROUPBY, et de la requête.

 Note

La fonction d'agrégation ou les colonnes de jointure ne peuvent pas être utilisées comme colonnes de dimension.

10. Pour les fonctions scalaires, choisissez une option pour Quelles fonctions scalaires souhaitez-vous autoriser ?

Si vous choisissez...	Alors...
Tous sont actuellement pris en charge par AWS Clean Rooms	Vous autorisez toutes les fonctions scalaires actuellement prises en charge par AWS Clean Rooms. Vous pouvez choisir Afficher la liste pour voir la liste complète des fonctions scalaires prises en charge dans AWS Clean Rooms.
Une liste personnalisée	Vous pouvez personnaliser les fonctions scalaires à autoriser. Choisissez une ou plusieurs options dans la liste déroulante Spécifier les fonctions scalaires autorisées.

Si vous choisissez...	Alors...
Aucun	Vous ne souhaitez autoriser aucune fonction scalaire.

Pour de plus amples informations, veuillez consulter [Fonctions scalaires](#).

11. Choisissez Suivant.
12. Dans Étape 3 : Spécifier les contrôles des résultats de requête, pour les contraintes d'agrégation :
 - a. Sélectionnez la liste déroulante pour chaque nom de colonne.
 - b. Sélectionnez la liste déroulante pour chaque nombre minimum de valeurs distinctes qui doivent être respectées pour que chaque ligne de sortie soit renvoyée, une fois la COUNT DISTINCT fonction appliquée à celle-ci.
 - c. Choisissez Ajouter une contrainte pour ajouter d'autres contraintes d'agrégation.
 - d. (Facultatif) Choisissez Supprimer pour supprimer une contrainte d'agrégation.
13. Pour les analyses supplémentaires appliquées aux résultats, sélectionnez une option en fonction de votre objectif.

Votre objectif	Option recommandée
Autoriser uniquement les requêtes directes sur cette table. Empêchez l'exécution d'analyses supplémentaires sur les résultats des requêtes. La table ne peut être utilisée que pour des requêtes directes.	Non autorisé
Autorisez mais n'exigez pas à la fois des requêtes directes et des analyses supplémentaires sur cette table.	Autorisé
Exigez que la table ne puisse être utilisée que dans des requêtes directes traitées avec l'une des analyses supplémentaires requises. Les requêtes directes sur cette table doivent	Obligatoire

Votre objectif	Option recommandée
être traitées ultérieurement avant de pouvoir être renvoyées.	

14. Choisissez Suivant.

15. Dans Étape 4 : révision et configuration, passez en revue les sélections que vous avez effectuées lors des étapes précédentes, modifiez-les si nécessaire, puis choisissez Configurer la règle d'analyse.

Un message de confirmation s'affiche indiquant que vous avez correctement configuré une règle d'analyse d'agrégation pour la table.

Ajouter une règle d'analyse de liste à un tableau (flux guidé)

La règle d'analyse de liste autorise les requêtes qui produisent des listes au niveau des lignes indiquant le chevauchement entre la table associée et une table du membre autorisé à effectuer la requête.

Cette procédure décrit le processus d'ajout de la règle d'analyse de liste à votre table configurée à l'aide de l'option Flux guidé de la AWS Clean Rooms console.

Note

Les tables configurées à l'aide de sources de données autres que S3 ne prennent en charge que [les règles d'analyse personnalisées](#).

Pour ajouter une règle d'analyse de liste à un tableau (flux guidé)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez le tableau configuré.
4. Sur la page détaillée de la table configurée, choisissez Configurer la règle d'analyse.
5. Sous Étape 1 : Choisissez le type de règle d'analyse, sous Type de règle d'analyse, choisissez l'option Liste.

6. Sous Méthode de création, sélectionnez Flux guidé, puis Suivant.
7. Dans Étape 2 : Spécifier les contrôles de requête, pour les contrôles de jointure :
 - a. Sous Spécifier les colonnes de jointure, choisissez les colonnes que vous souhaitez autoriser à utiliser dans l'INNERJOINinstruction.
 - b. Sous Spécifier les opérateurs autorisés pour la mise en correspondance, choisissez quels opérateurs, le cas échéant, peuvent être utilisés pour faire correspondre plusieurs colonnes de jointure. Si vous sélectionnez deux JOIN colonnes ou plus, l'un de ces opérateurs est requis.

Si vous choisissez...	Alors...
ET	Vous pouvez inclure AND dans le INNER JOIN match des conditions permettant de joindre une colonne à une autre entre les tables.
OU	Vous pouvez inclure OR dans les conditions de INNER JOIN correspondance pour combiner plusieurs correspondances de colonnes entre les tables. Cet opérateur logique est utile pour obtenir un taux de correspondance plus élevé.

8. (Facultatif) Pour les contrôles de liste, dans la liste déroulante Spécifier les colonnes de liste, choisissez les colonnes que vous souhaitez autoriser à utiliser dans le résultat de la requête (c'est-à-dire utilisées dans l'SELECTinstruction) ou utilisées pour filtrer les résultats (c'est-à-dire l'WHEREinstruction).
9. Choisissez Suivant.
10. Dans Étape 3 : Spécifier les contrôles des résultats de requête, pour Analyses supplémentaires appliquées à la sortie, sélectionnez une option en fonction de votre objectif.

Votre objectif	Option recommandée
Autoriser uniquement les requêtes directes sur cette table. Empêchez l'exécution d'analyses supplémentaires sur les résultats	Non autorisé

Votre objectif	Option recommandée
des requêtes. La table ne peut être utilisée que pour des requêtes directes.	
Autorisez mais n'exigez pas à la fois des requêtes directes et des analyses supplémentaires sur cette table.	Autorisé
Exigez que la table ne puisse être utilisée que dans des requêtes directes traitées avec l'une des analyses supplémentaires requises. Les requêtes directes sur cette table doivent être traitées ultérieurement avant de pouvoir être renvoyées.	Obligatoire

11. Dans Étape 4 : révision et configuration, passez en revue les sélections que vous avez effectuées lors des étapes précédentes, modifiez-les si nécessaire, puis choisissez Configurer la règle d'analyse.

Un message de confirmation s'affiche indiquant que vous avez correctement configuré une règle d'analyse de liste pour la table.

Ajouter une règle d'analyse personnalisée à un tableau (flux guidé)

La règle d'analyse personnalisée permet d'effectuer des requêtes ou des PySpark tâches SQL personnalisées sur une table configurée. La règle d'analyse personnalisée est requise si vous utilisez :

- [Modèles d'analyse](#) permettant d'autoriser un ensemble spécifique de requêtes ou de PySpark tâches SQL préapprouvées ou un ensemble spécifique de comptes pouvant fournir des requêtes utilisant vos données.
- [AWS Clean Rooms Confidentialité différentielle](#) pour protéger contre les tentatives d'identification des utilisateurs.
- Sources de données autres que S3, telles qu'Amazon Athena ou Snowflake.

Cette procédure décrit le processus d'ajout de la règle d'analyse personnalisée à votre table configurée à l'aide de l'option Flux guidé de la AWS Clean Rooms console.

Pour ajouter une règle d'analyse personnalisée à un tableau (flux guidé)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez le tableau configuré.
4. Sur la page détaillée de la table configurée, choisissez Configurer la règle d'analyse.
5. Dans Étape 1 : Choisissez le type de règle d'analyse, sous Type de règle d'analyse, choisissez l'option Personnalisée.
6. Sous Méthode de création, sélectionnez Flux guidé, puis Suivant.
7. Dans Étape 2 : Spécifier les contrôles d'analyse, pour les contrôles d'analyse directe, choisissez une option en fonction de votre objectif.

Votre objectif	Action recommandée
Passez en revue chaque nouvelle analyse avant de l'autoriser à être exécutée sur cette table configurée	1. Sous Modèles d'analyse autorisés à être exécutés, choisissez Ajouter un modèle d'analyse. 2. Choisissez le modèle de collaboration et d'analyse approprié dans les listes déroulantes. 3. Choisissez Suivant.
Permettre à des collaborateurs spécifiques d'exécuter n'importe quelle analyse d'un type choisi sans révision sur ce tableau	1. Sous Type d'analyse, a. Choisissez N'importe quelle requête pour autoriser toute requête créée par le Compte AWS que vous spécifiez. b. Choisissez N'importe quelle requête pour autoriser toute tâche créée par le Compte AWS que vous spécifiez. 2. Sous Comptes AWS Autorisé à créer une analyse, choisissez Ajouter Compte AWS. 3. Entrez un Compte AWS identifiant Compte AWS ou choisissez-en un dans la liste déroulante.

Votre objectif	Action recommandée
	4. (Facultatif) Choisissez Ajouter un autre Compte AWS pour en ajouter un autre Compte AWS. 5. Choisissez Suivant.

8. Dans Étape 3 : Spécifier les contrôles des résultats d'analyse,
- Pour les contrôles des résultats des Job, notez qu'aucun contrôle de résultat supplémentaire n'est pris en charge.
 - Sous Contrôles des résultats de la requête, pour Colonnes non autorisées en sortie, choisissez les colonnes que vous souhaitez autoriser dans le résultat de la requête, en fonction de votre objectif.

Votre objectif	Action recommandée
Autoriser le renvoi de toutes les colonnes dans les sorties de requête	1. N'en choisissez aucun 2. Passez à Analyses supplémentaires appliquées à la sortie.
Interdire le renvoi de certaines colonnes dans les résultats des requêtes	1. Choisissez une liste personnalisée 2. Sous Spécifier les colonnes interdites, choisissez les colonnes que vous souhaitez supprimer des résultats de requête.

- Pour les analyses supplémentaires appliquées à la sortie, choisissez si des analyses supplémentaires peuvent être appliquées à la sortie de la requête, en fonction de votre objectif.

Votre objectif	Option recommandée
Autoriser uniquement les requêtes directes sur cette table.	Non autorisé

Votre objectif	Option recommandée
• Empêchez l'exécution d'analyses supplémentaires sur les résultats des requêtes. • La table ne peut être utilisée que pour des requêtes directes.	
Autorisez mais n'exigez pas à la fois des requêtes directes et des analyses supplémentaires sur cette table.	Allow
• Exigez que la table ne puisse être utilisée que dans des requêtes directes traitées avec l'une des analyses supplémentaires requises. • Les requêtes directes sur cette table doivent être traitées ultérieurement avant de pouvoir être renvoyées.	Obligatoire

d. Choisissez Suivant.

9. (Facultatif) À l'étape 4 : définir la confidentialité différentielle, déterminez si vous souhaitez activer ou désactiver la confidentialité différentielle.

La confidentialité différentielle est une technique mathématiquement éprouvée pour protéger vos données contre les attaques de réidentification.

 Note

AWS Clean Rooms La confidentialité différentielle n'est disponible que pour les collaborations dans lesquelles les données sont stockées dans Amazon S3.

Pour la confidentialité différentielle, choisissez d'activer ou de désactiver la confidentialité différentielle, en fonction de votre objectif.

Votre objectif	Action recommandée
• Vous n'avez pas besoin de protection contre les tentatives de réidentification • Votre table ne contient pas de données au niveau de l'utilisateur	1. Choisissez Désactiver. 2. Choisissez Suivant.
• Vous avez besoin d'une protection contre les tentatives de réidentification • Votre tableau contient des données au niveau de l'utilisateur	1. Choisissez Activer. 2. Sélectionnez la colonne Identifiant utilisateur qui contient l'identifiant unique de vos utilisateurs, par exemple la <code>user_id</code> colonne dont vous souhaitez protéger la confidentialité. Pour activer la confidentialité différentielle pour deux tables ou plus dans le cadre d'une collaboration, vous devez configurer la même colonne que la colonne Identifiant utilisateur dans les deux règles d'analyse afin de conserver une définition cohérente des utilisateurs entre les tables. En cas de mauvaise configuration, le membre autorisé à effectuer une requête reçoit un message d'erreur indiquant qu'il a le choix entre deux colonnes afin de calculer le nombre de contributions des utilisateurs (par exemple, le nombre d'impressions publicitaires effectuées par un utilisateur) lors de l'exécution de la requête. 3. Choisissez Suivant.

10. Dans Étape 5 : révision et configuration, passez en revue les sélections que vous avez effectuées lors des étapes précédentes, modifiez-les si nécessaire, puis choisissez Configurer la règle d'analyse.

Un message de confirmation s'affiche indiquant que vous avez correctement configuré une règle d'analyse personnalisée pour la table.

Ajouter une règle d'analyse à une table (éditeur JSON)

La procédure suivante montre comment ajouter une règle d'analyse à une table à l'aide de l'option de l'éditeur JSON dans la AWS Clean Rooms console.

Note

Les tables configurées à l'aide de sources de données autres que S3 ne prennent en charge que [les règles d'analyse personnalisées](#).

Pour ajouter une agrégation, une liste ou une règle d'analyse personnalisée à une table (éditeur JSON)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez le tableau configuré.
4. Sur la page détaillée de la table configurée, choisissez Configurer la règle d'analyse.
5. Dans Étape 1 : Choisissez le type de règle d'analyse, sous Type de règle d'analyse, choisissez l'option Agrégation, Liste ou Personnalisée.
6. Sous Méthode de création, sélectionnez l'éditeur JSON, puis cliquez sur Suivant.
7. À l'étape 2 : Spécifier les contrôles, vous pouvez choisir d'insérer une structure de requête (Insérer un modèle) ou d'insérer un fichier (Importer depuis un fichier).

Si vous choisissez...	Alors...
Insérer un modèle	1. Spécifiez les paramètres de la règle d'analyse sélectionnée dans la définition de la règle d'analyse. 2. Vous pouvez appuyer sur Ctrl + barre d'espace pour activer la saisie automatique.

Si vous choisissez...	Alors...
	Pour plus d'informations sur les paramètres des règles d'analyse d'agrégation, consultez Règle d'analyse d'agrégation : contrôles des requêtes. Pour plus d'informations sur les paramètres des règles d'analyse de liste, consultez Règle d'analyse des listes : contrôles des requêtes.
Importer depuis un fichier	1. Sélectionnez votre fichier JSON sur votre disque local. 2. Choisissez Ouvrir. La définition de la règle d'analyse affiche la règle d'analyse du fichier chargé.

8. Choisissez Suivant.
9. Dans Étape 3 : révision et configuration, passez en revue les sélections que vous avez effectuées lors des étapes précédentes, modifiez-les si nécessaire, puis choisissez Configurer la règle d'analyse.

Vous recevez un message de confirmation indiquant que vous avez correctement configuré une règle d'analyse pour la table.

Étapes suivantes

Maintenant que vous avez configuré une règle d'analyse pour votre table configurée, vous êtes prêt à :

- [Associer une table configurée à une collaboration](#)
- [Interroger les tables de données](#) (en tant que membre pouvant effectuer des requêtes)

Associer une table configurée à une collaboration

Après avoir créé une table configurée et y avoir ajouté une règle d'analyse, vous pouvez l'associer à une collaboration et attribuer AWS Clean Rooms un rôle de service pour accéder à vos AWS Glue tables.

Note

Ce rôle de service dispose d'autorisations d'accès aux tables. Le rôle de service ne peut être assumé que AWS Clean Rooms pour exécuter les requêtes autorisées au nom du membre autorisé à effectuer des requêtes. Aucun membre de la collaboration (autre que le propriétaire des données) n'a accès aux tables sous-jacentes de la collaboration. Le propriétaire des données peut activer la confidentialité différentielle pour que ses tables puissent être consultées par d'autres membres.

Budget d'accès aux données

Lorsque vous associez une table configurée, vous pouvez appliquer un budget d'accès aux données. Un budget d'accès aux données contrôle le nombre de fois qu'une table peut être utilisée pour les requêtes, les tâches et les canaux d'entrée ML dans une collaboration. Ces budgets aident les entreprises à gérer l'utilisation des ressources et à contrôler les coûts en limitant l'utilisation des tables.

Chaque fois qu'une table est utilisée dans une requête, une tâche ou un canal d'entrée ML, le budget de cette table est réduit d'une unité. Lorsque le budget atteint zéro, la table ne peut pas être utilisée dans les requêtes SQL, les jobs Pyspark, ni dans le cadre des canaux d'entrée ML dérivés de la table.

Vous pouvez établir un budget par période actualisé périodiquement, un budget global pour l'utilisation globale, ou les deux. Par défaut, l'utilisation des tables est illimitée.

- **Budget par période** : allocation renouvelable qui limite le nombre de fois que ce tableau peut être utilisé au cours d'une période spécifiée. Vous pouvez définir une période quotidienne, hebdomadaire ou mensuelle. Ce budget peut être configuré pour être actualisé automatiquement sur une base quotidienne, hebdomadaire ou mensuelle.
- **Budget global** : allocation courante qui limite le nombre total de fois que ce tableau peut être utilisé.

Associer une table configurée

Les rubriques suivantes décrivent comment associer une table configurée et appliquer un budget d'accès aux données à une collaboration à l'aide de la AWS Clean Rooms console.

Pour plus d'informations sur la façon d'associer vos tables configurées à la collaboration à l'aide du AWS SDKs, consultez la [référence AWS Clean Rooms d'API](#).

Étape 1 : Exécuter les prérequis

Pour associer une table configurée, vous devez remplir les conditions préalables suivantes :

- Une AWS Glue table qui pointe vers l'emplacement d'un dossier Amazon S3 (pas un seul fichier)
- Pour les AWS Glue tables chiffrées :
 - Rôle de service autorisé à utiliser des AWS KMS clés pour déchiffrer AWS Glue des tables
 - Pour les ensembles de données Amazon S3 AWS KMS chiffrés : le rôle de service doit également être autorisé à utiliser la AWS KMS clé pour déchiffrer les données Amazon S3

Pour plus d'informations sur la configuration du chiffrement, voir [Configuration du chiffrement AWS Glue dans](#) le Guide du AWS Glue développeur.

Pour vérifier l'emplacement de votre AWS Glue table :

1. Ouvrez la AWS Glue console à <https://console.aws.amazon.com/glue/>
2. Consultez les détails de votre table et confirmez que l'emplacement pointe vers un dossier S3

Étape 2 : associer une table configurée

Pour associer une table configurée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Choisissez la méthode pour associer votre table :
 - a. À partir de la page détaillée de la table configurée :
 - i. Dans le volet de navigation de gauche, choisissez Tables.
 - ii. Choisissez le tableau configuré.

- iii. Sur la page détaillée du tableau configuré, choisissez Associer à la collaboration.
 - iv. Dans la boîte de dialogue Associer la table à la collaboration, choisissez la collaboration dans la liste déroulante.
 - b. Sur la page détaillée de la collaboration :
 - i. Dans le volet de navigation de gauche, sélectionnez Collaborations.
 - ii. Choisissez la collaboration.
 - iii. Dans l'onglet Tables, choisissez Associer une table.
3. Sur la page Table associée, effectuez l'une des opérations suivantes :
 - Choisissez une table configurée existante : choisissez le nom de la table configurée que vous souhaitez associer à la collaboration dans la liste déroulante.
 - Configurer une nouvelle table — Choisissez Configurer une nouvelle table et suivez les instructions de la page Configurer une nouvelle table.
 - Afficher le schéma et la règle d'analyse pour la table configurée : activez Afficher le schéma et la règle d'analyse.
4. Pour plus de détails sur l'association des tables,
 - a. Entrez un nom pour la table associée.

Vous pouvez utiliser le nom par défaut ou renommer cette table.
 - b. (Facultatif) Entrez une description de la table.

La description facilite la rédaction de requêtes.
5. Spécifiez les autorisations d'accès au service en sélectionnant Créer et utiliser un nouveau rôle de service ou Utiliser un rôle de service existant.

 Note

Si vous associez une table configurée basée sur Amazon Athena, choisissez un nom de rôle de service existant dans la liste déroulante. Assurez-vous que le rôle de service dispose des autorisations IAM et, si nécessaire, de Lake Formation sur le jeu de données.

Si vous choisissez...	Alors...
Création et utilisation d'un nouveau rôle de service	• AWS Clean Rooms crée un rôle de service avec la politique requise pour cette table. • Le nom du rôle de service par défaut est <code>cleanrooms-<timestamp></code> • Vous devez disposer des autorisations nécessaires pour créer des rôles et associer des politiques. • Si vos données d'entrée sont cryptées, vous pouvez sélectionner Ces données sont cryptées avec une clé KMS, puis saisir une clé AWS KMS key qui sera utilisée pour déchiffrer vos données saisies.

Si vous choisissez...	Alors...
Utiliser un rôle de service existant	1. Choisissez un nom de rôle de service existant dans la liste déroulante. La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles. Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser. 2. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM. S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible. Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires. 3. (Facultatif) Cochez la case Ajouter une politique préconfigurée avec les autorisations nécessaires à ce rôle pour ajouter attacher les autorisations nécessaires au rôle. Vous devez disposer des autorisations nécessaires pour modifier les rôles et créer des politiques.

 Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).

- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.
- Si vous ne pouvez pas modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique du rôle de service est introuvable.

6. Si vous souhaitez activer les balises d'association de tables configurées pour la ressource d'association de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
7. Choisissez Suivant.
8. Sur la page Configurer les règles d'analyse de collaboration, choisissez l'une des options suivantes :
 - Oui, créez une règle d'analyse de collaboration dès maintenant : associez votre table à cette collaboration et créez une règle d'analyse de collaboration
 - Non, je créerai une règle d'analyse de collaboration ultérieurement. Associez votre table à cette collaboration uniquement. Vous pourrez créer une règle d'analyse de collaboration ultérieurement.
9. Si vous choisissez Oui, créez une règle d'analyse de collaboration dès maintenant. Pour la livraison des résultats, choisissez les membres autorisés à recevoir les résultats de la requête dans la liste déroulante.
10. Choisissez Suivant.
11. Sur la page Ajouter un budget d'accès aux données, pour Configuration du budget d'accès aux données, sélectionnez l'une des options suivantes :
 - Oui, ajoutez un budget d'accès aux données dès maintenant : associez votre table à cette collaboration et ajoutez un budget d'accès aux données. Vous pouvez sélectionner un budget périodique, un budget global ou les deux.
 - Non, j'ajouterai un budget d'accès aux données ultérieurement. Associez votre table à cette collaboration uniquement. Vous pouvez ajouter un budget d'accès aux données ultérieurement.

Si vous sélectionnez Non, j'ajouterai un budget d'accès aux données ultérieurement, passez à l'étape 15.

12. Si vous choisissez Oui, ajoutez un budget d'accès aux données dès maintenant, choisissez l'une des configurations budgétaires suivantes :

Budget par période uniquement	Budget à vie uniquement	Budgets par période et par durée de vie
1. Laissez l'option Ajouter un budget par période sélectionnée. 2. Entrez un montant de budget par période compris entre 1 et 1 000 000. 3. Dans Période, choisissez Quotidien, Hebdomadaire ou Mensuel. 4. (Facultatif) Laissez l'option Actualiser automatiquement le budget chaque semaine sélectionnée pour renouveler l'allocation. 5. Désactivez Ajouter un budget à vie.	1. Désactivez Ajouter un budget par période. 2. Sélectionnez Ajouter un budget à vie. 3. Entrez un montant de budget à vie compris entre 1 et 1 000 000.	1. Laissez l'option Ajouter un budget par période sélectionnée. 2. Entrez un montant de budget par période compris entre 1 et 1 000 000. 3. Dans Période, choisissez Quotidien, Hebdomadaire ou Mensuel. 4. Laissez l'option Actualiser automatiquement le budget chaque semaine sélectionnée. 5. Sélectionnez Ajouter un budget à vie. 6. Entrez un montant de budget à vie compris entre 1 et 1 000 000.

13. Passez en revue vos sélections sous Résumé du budget d'accès aux données.

Exemple Exemple

Par exemple, si vous avez choisi un montant de budget par période de 1 000, que vous avez défini la période sur Hebdomadaire, que vous avez laissé la case Actualiser automatiquement le budget hebdomadaire sélectionnée et que vous avez défini le budget global sur 1 000 000, le résumé du budget Access affichera le message suivant : Chaque semaine, ce tableau peut être utilisé jusqu'à 1 000 fois pour exécuter des requêtes ou des tâches. Ce budget est configuré pour

être actualisé automatiquement tous les dimanches à 00h00 UTC et continuera à être actualisé jusqu'à ce que ce tableau atteigne son budget à vie de 1 000 000 d'utilisations.

14. (Facultatif) Si vous souhaitez activer les balises de budget d'accès aux données pour la ressource de budget d'accès, choisissez Ajouter une nouvelle balise et entrez une paire clé/valeur.
15. Choisissez Suivant.
16. Passez en revue les informations de la page Révision et création.
 - a. Si vous devez modifier des sections, choisissez Modifier.
 - b. Modifiez vos configurations, puis choisissez Next.
17. Choisissez Associer une table.

Étape 3 : Prochaines étapes

Maintenant que vous avez associé votre table de données configurée à la collaboration, vous êtes prêt à :

- [Ajouter une règle d'analyse de collaboration](#) au tableau configuré
- [Modifiez la collaboration](#), si vous en êtes le créateur
- [Interroger les tables de données](#) (en tant que membre pouvant effectuer des requêtes)

Configuration d'un budget d'accès aux données

Un collaborateur peut consulter, ajouter, modifier et supprimer un budget d'accès aux données afin de limiter le nombre de fois qu'une table peut être utilisée dans un flux de travail. Utilisez ces budgets pour gérer les données et les coûts.

Chaque fois qu'une table est interrogée ou qu'une tâche est exécutée à l'aide d'un canal d'entrée ML dérivé d'une table, le budget de cette table est réduit d'une unité. Lorsque le budget atteint zéro, la table ne peut pas être interrogée et les tâches ML ne peuvent pas être exécutées à l'aide des canaux d'entrée ML dérivés de la table.

Vous pouvez établir un budget par période actualisé périodiquement, un budget global pour l'utilisation globale, ou les deux. Par défaut, l'utilisation des tables est illimitée.

- Budget par période : allocation renouvelable qui limite le nombre de fois que ce tableau peut être utilisé au cours d'une période spécifiée. Vous pouvez définir une période quotidienne,

hebdomadaire ou mensuelle. Ce budget peut être configuré pour être actualisé automatiquement sur une base quotidienne, hebdomadaire ou mensuelle.

- Budget global : allocation courante qui limite le nombre total de fois que ce tableau peut être utilisé.

Rubriques

- [Afficher un budget d'accès aux données](#)
- [Ajouter un budget d'accès aux données à une table associée existante](#)
- [Modification d'un budget d'accès aux données](#)
- [Supprimer un budget d'accès aux données](#)

Afficher un budget d'accès aux données

Vous pouvez consulter un budget d'accès aux données depuis l'onglet Tables ou depuis la page de détails des tables.

Pour consulter un budget d'accès aux données

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Effectuez l'une des actions suivantes :
 - Dans la colonne Budget d'accès aux données restant, sélectionnez le budget pour afficher les détails.
 - Choisissez un tableau, puis sur la page des détails du tableau, faites défiler la page vers le bas pour afficher la section Détails du budget d'accès aux données.

Ajouter un budget d'accès aux données à une table associée existante

En tant que membre de la collaboration, vous pouvez ajouter un budget d'accès aux données à une table associée existante.

Pour ajouter un budget d'accès aux données à une table associée existante

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Sélectionnez le bouton d'option situé à côté du tableau auquel vous souhaitez ajouter un budget d'accès aux données.
6. Dans la liste déroulante Actions, sous Budget d'accès aux données, sélectionnez Ajouter (s'il n'y a pas déjà de budget).
7. Choisissez l'une des configurations budgétaires suivantes :

Budget par période uniquement	Budget à vie uniquement	Budgets par période et par durée de vie
1. Laissez l'option Ajouter un budget par période sélectionnée. 2. Entrez un montant de budget par période compris entre 1 et 1 000 000. 3. Dans Période, choisissez Quotidien, Hebdomadaire ou Mensuel. 4. (Facultatif) Laissez l'option Actualiser automatiquement le budget chaque semaine sélectionnée pour renouveler l'allocation. 5. Désactivez Ajouter un budget à vie.	1. Désactivez Ajouter un budget par période. 2. Sélectionnez Ajouter un budget à vie. 3. Entrez un montant de budget à vie compris entre 1 et 1 000 000.	1. Laissez l'option Ajouter un budget par période sélectionnée. 2. Entrez un montant de budget par période compris entre 1 et 1 000 000. 3. Dans Période, choisissez Quotidien, Hebdomadaire ou Mensuel. 4. Laissez l'option Actualiser automatiquement le budget chaque semaine sélectionnée. 5. Sélectionnez Ajouter un budget à vie.

Budget par période uniquement	Budget à vie uniquement	Budgets par période et par durée de vie

6. Entrez un montant de budget à vie compris entre 1 et 1 000 000.

8. Passez en revue vos sélections sous Résumé du budget d'accès aux données.

9. Exemple Exemple

Par exemple, si vous avez choisi un montant de budget par période de 1 000, que vous avez défini la période sur Hebdomadaire, que vous avez laissé la case Actualiser automatiquement le budget hebdomadaire sélectionnée et que vous avez défini le budget global sur 1 000 000, le résumé du budget Access affichera le message suivant : Chaque semaine, ce tableau peut être utilisé jusqu'à 1 000 fois pour exécuter des requêtes ou des tâches. Ce budget est configuré pour être actualisé automatiquement tous les dimanches à 00h00 UTC et continuera à être actualisé jusqu'à ce que ce tableau atteigne son budget à vie de 1 000 000 d'utilisations.

10. (Facultatif) Si vous souhaitez activer les balises de budget d'accès aux données pour la ressource de budget d'accès, choisissez Ajouter une nouvelle balise et entrez une paire clé/valeur.

11. Choisissez Ajouter un budget d'accès aux données.

Modification d'un budget d'accès aux données

En tant que membre de la collaboration, vous pouvez modifier le budget d'accès aux données. Lorsque vous modifiez un budget d'accès aux données, il réinitialise le solde budgétaire actuel.

Vous pouvez modifier un budget d'accès aux données à partir de l'onglet Tables ou de la page des détails des tables.

Tables tab

Pour modifier un budget d'accès aux données depuis l'onglet Tables

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.

4. Choisissez l'onglet Tables.
5. Sélectionnez le bouton d'option situé à côté du tableau que vous souhaitez modifier.
6. Dans la liste déroulante Actions, sous Budget d'accès aux données, sélectionnez Modifier.
7. Sur la page Modifier le budget d'accès aux données, mettez à jour les informations relatives au budget par période ou au budget global.
8. Consultez le résumé du budget d'accès aux données pour vérifier que les modifications que vous avez apportées sont correctes.
9. Sélectionnez Enregistrer les modifications.

Table details page

Pour modifier un budget d'accès aux données depuis la page de détails du tableau

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Choisissez une table .
6. Sur la page des détails du tableau, faites défiler la page vers le bas jusqu'à la section Détails du budget d'accès aux données.
7. Dans la liste déroulante Actions, choisissez Modifier.
8. Sur la page Modifier le budget d'accès aux données, mettez à jour les informations relatives au budget par période ou au budget global.
9. Sélectionnez Enregistrer les modifications.

Supprimer un budget d'accès aux données

Vous pouvez supprimer un budget d'accès aux données depuis l'onglet Tables ou depuis la page de détails des tables.

Tables tab

Pour supprimer un budget d'accès aux données depuis l'onglet Tables

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Sélectionnez le bouton d'option situé à côté de la table que vous souhaitez supprimer.
6. Dans la liste déroulante Actions, sous Budget d'accès aux données, sélectionnez Supprimer.

Important

Vous ne pouvez pas annuler cette action et votre budget d'accès aux données sera redéfini comme illimité.

7. Si vous êtes certain de vouloir supprimer le budget d'accès aux données, choisissez Supprimer.

Table details page

Pour supprimer un budget d'accès aux données depuis la page de détails du tableau

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Choisissez une table .
6. Sur la page des détails du tableau, faites défiler la page vers le bas jusqu'à la section Détails du budget d'accès aux données.
7. Dans la liste déroulante Actions, choisissez Supprimer.

 Important

Vous ne pouvez pas annuler cette action et votre budget d'accès aux données sera redéfini comme illimité.

8. Si vous êtes certain de vouloir supprimer le budget d'accès aux données, choisissez Supprimer.

Ajouter une règle d'analyse de collaboration à une table configurée

La règle d'analyse de collaboration vous permet de définir des contrôles spécifiques à cette collaboration. Ces contrôles fonctionnent conjointement avec la règle d'analyse de table configurée pour déterminer comment cette table peut être analysée dans le cadre de cette collaboration.

Vous ajoutez une règle d'analyse de collaboration à une table configurée après avoir [créé une table configurée](#), [ajouté une règle d'analyse](#) et [l'avoir associée à une collaboration](#). Vous devez ajouter une règle d'analyse de collaboration si la table est configurée pour prendre en charge l'analyse directe ou pour permettre une analyse supplémentaire.

- Analyse directe : la table peut être utilisée dans des requêtes qui l'analysent directement. Par exemple, dans une requête qui produit une analyse de mesure agrégée ou une liste d'identifiants à activer.
- Analyse supplémentaire — La table peut également être utilisée comme entrée dans des analyses supplémentaires, en plus des requêtes qui l'analysent directement. Par exemple, la table peut être utilisée dans une requête qui est le point de départ d'un modèle ML similaire ou un canal d'entrée ML pour un modèle ML personnalisé.

Pour ajouter la règle d'analyse de collaboration à un tableau

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Tables, sous Tables associées par vous, consultez la table configurée que vous avez associée à la collaboration.

- Si le statut d'analyse directe ou le statut d'analyse supplémentaire a le statut Prêt, la table est prête à être interrogée.
 - Si le statut d'analyse directe ou le statut d'analyse supplémentaire a le statut Non prêt, sélectionnez le statut, puis choisissez Configurer dans la boîte de dialogue.
5. Sur la page Configurer la règle d'analyse de collaboration, développez Afficher la règle d'analyse de table configurée pour afficher les détails.
 6. Pour les analyses supplémentaires autorisées, choisissez l'option en fonction de votre objectif.

Votre objectif	Option recommandée
Autorisez toute analyse supplémentaire sur la table.	Any
N'autorisez que des analyses supplémentaires sur la table par un membre spécifique.	N'importe lequel par des membres spécifiques
N'autorisez que des analyses spécifiques sur le tableau.	Liste personnalisée

7. Pour la livraison des résultats, spécifiez qui peut recevoir les résultats de la part des membres autorisés à recevoir des résultats pour la liste déroulante des résultats de requête.
8. Choisissez Configurer la règle d'analyse.

Configuration d'une politique de confidentialité différentielle (facultatif)

Note

AWS Clean Rooms La confidentialité différentielle n'est disponible que pour les collaborations dans lesquelles les données sont stockées dans Amazon S3.

Cette procédure décrit le processus de configuration de la politique de confidentialité différentielle dans une collaboration à l'aide de l'option Flux guidé de la AWS Clean Rooms console. Il s'agit d'une étape unique pour toutes les tables dotées d'une protection de confidentialité différentielle.

Pour configurer les paramètres de confidentialité différentiels (flux guidé)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Tables de la page de collaboration, choisissez Configurer la politique de confidentialité différentielle.
5. Sur la page Configurer une politique de confidentialité différentielle, choisissez des valeurs pour les propriétés suivantes :
 - Budget de confidentialité
 - Actualiser le budget de confidentialité tous les mois
 - Bruit ajouté par requête

Vous pouvez utiliser les valeurs par défaut ou saisir des valeurs personnalisées adaptées à votre cas d'utilisation spécifique. Après avoir choisi des valeurs pour le budget de confidentialité et le bruit ajouté par requête, vous pouvez prévisualiser l'utilitaire obtenu en termes de nombre d'agrégations possibles pour toutes les requêtes portant sur vos données.

6. Choisissez Configurer.

Vous verrez un message de confirmation indiquant que vous avez correctement configuré la politique de confidentialité différentielle pour la collaboration.

Maintenant que vous avez configuré la confidentialité différentielle, vous êtes prêt à :

- [Interroger les tables de données](#) (en tant que membre pouvant effectuer des requêtes)
- [Collaborations](#) (si vous êtes le créateur de la collaboration)

Afficher les journaux d'utilisation différentiels de confidentialité

En tant que membre d'une collaboration qui protège les données avec une confidentialité différentielle, une fois que vous avez créé une collaboration avec une confidentialité différentielle, vous pouvez surveiller l'utilisation du budget de confidentialité.

Pour voir combien d'agrégations ont été effectuées et quelle part du budget de confidentialité a été utilisée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Choisissez Afficher les journaux d'utilisation (texte bleu).
6. Consultez les détails d'utilisation, y compris le budget de confidentialité et le niveau d'utilité fourni.

Modifier une politique de confidentialité différentielle

Après avoir configuré la politique de confidentialité différentielle, vous pouvez à tout moment la mettre à jour pour mieux refléter vos besoins en matière de confidentialité.

Pour modifier la politique de confidentialité différentielle

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Tables de la page de collaboration, sous Tables que vous avez associées, choisissez Modifier.
5. Sur la page Modifier la confidentialité différentielle, choisissez de nouvelles valeurs pour les propriétés suivantes :
 - Budget de confidentialité : déplacez le curseur pour augmenter ou diminuer le budget à tout moment au cours d'une collaboration. Vous ne pouvez pas réduire le budget une fois que le membre autorisé à interroger vos données a commencé à interroger vos données. Si le budget de confidentialité est augmenté, AWS Clean Rooms vous continuerez à utiliser le budget existant jusqu'à ce qu'il soit entièrement utilisé avant d'utiliser le budget de confidentialité nouvellement ajouté.
 - Bruit ajouté par requête : déplacez le curseur pour augmenter ou diminuer le bruit ajouté par requête à tout moment au cours d'une collaboration.

 Note

Vous pouvez choisir des exemples interactifs pour découvrir comment les différentes valeurs du budget de confidentialité et du bruit ajouté par requête affectent le nombre de fonctions d'agrégation que vous pouvez exécuter.

Vous ne pouvez pas modifier la valeur de l'actualisation du budget de confidentialité. Pour modifier votre sélection, vous devez supprimer la politique de confidentialité différentielle et en créer une nouvelle.

6. Sélectionnez Enregistrer les modifications.

Un message de confirmation s'affiche indiquant que vous avez correctement modifié la politique de confidentialité différentielle.

Supprimer une politique de confidentialité différentielle

Vous pouvez supprimer la politique de confidentialité différentielle dans l'onglet Tables d'une collaboration.

Pour supprimer la politique de confidentialité différentielle

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Tables de la page de collaboration, à côté de Politique de confidentialité différentielle, sélectionnez Supprimer.
5. Si vous êtes certain de vouloir supprimer la politique de confidentialité différentielle, choisissez Supprimer.

Après avoir supprimé une politique de confidentialité différentielle, vous ne pouvez pas accéder aux journaux d'utilisation du budget de confidentialité contenus dans cette politique. Les tables dans lesquelles la confidentialité différentielle est activée ne peuvent pas être consultées si la politique de confidentialité différentielle est supprimée.

Affichage des paramètres de confidentialité différentiels calculés

Pour les utilisateurs expérimentés en matière de confidentialité différentielle, vous pouvez consulter les paramètres de confidentialité différentielle calculés dans l'onglet Analyse d'une collaboration.

Pour afficher les paramètres de confidentialité différentiels calculés

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Analyse, dans la section Résultats, sélectionnez Afficher les paramètres de confidentialité différentiels calculés.

Dans le tableau des paramètres de confidentialité différentiels calculés, vous pouvez voir les valeurs de sensibilité des fonctions d'agrégation, définies comme la valeur maximale selon laquelle le résultat d'une fonction peut changer si les enregistrements d'un seul utilisateur sont ajoutés, supprimés ou modifiés. La liste inclut les paramètres de confidentialité différentiels suivants :

- La limite de contribution utilisateur (UCL) est le nombre maximum de lignes ajoutées par un utilisateur dans une requête SQL. Par exemple, si vous souhaitez compter le nombre total d'impressions correspondantes dans une campagne spécifique où chaque utilisateur peut avoir plusieurs impressions, la confidentialité AWS Clean Rooms différentielle doit limiter le nombre d'impressions d'un seul utilisateur afin de garantir l'exactitude du calcul de la confidentialité différentielle. En d'autres termes, si un utilisateur a plus d'impressions que la limite, il prend AWS Clean Rooms automatiquement un échantillon aléatoire uniforme des impressions de cet utilisateur conformément à la valeur UCL calculée et exclut les impressions restantes de cet utilisateur lors de l'exécution de la requête. La valeur UCL est égale à 1 si vous comptez le nombre d'utilisateurs uniques. Cela est dû au fait que l'ajout, la suppression ou la modification d'un seul utilisateur peut modifier le nombre d'utilisateurs distincts d'au plus 1.
- La valeur minimale est la limite inférieure d'une expression utilisée dans une fonction d'agrégation telle que `sum()`. Par exemple, si l'expression est une colonne connue sous le nom `depurchase_value`, la valeur minimale est la limite inférieure de la colonne.
- La valeur maximale est la limite supérieure d'une expression utilisée dans une fonction d'agrégation telle que `sum()`. Par exemple, si l'expression est une colonne connue sous le nom `depurchase_value`, la valeur maximale est la limite supérieure de la colonne.

Dans le tableau des paramètres de confidentialité différentiels calculés, vous pouvez utiliser ces paramètres pour mieux comprendre la quantité totale de bruit dans les résultats des requêtes. Par exemple, lorsque le bruit configuré ajouté par requête est de 30 utilisateurs et qu'une `COUNT DISTINCT (user_id)` requête est exécutée, la confidentialité AWS Clean Rooms différentielle ajoute un bruit aléatoire compris entre -30 et 30 avec une probabilité élevée car la sensibilité de `COUNT DISTINCT` est de 1. Dans le cas d'une `COUNT` requête avec la même configuration, la confidentialité AWS Clean Rooms différentielle ajoute du bruit statistique qui est ajusté en fonction de la limite de contribution de l'utilisateur, car un seul utilisateur peut ajouter plusieurs lignes au résultat de la requête. Dans le cas d'une `SUM` requête `SUM (purchase_value)` où toutes les valeurs des colonnes sont positives, le bruit total est redimensionné en fonction de la limite de contribution de l'utilisateur multipliée par la valeur maximale. AWS Clean Rooms La confidentialité différentielle calcule automatiquement les paramètres de sensibilité pour ajouter du bruit au moment de l'exécution de la requête et épuise le budget de confidentialité. L'épuisement du budget de confidentialité est nécessaire car les paramètres de sensibilité dépendent des données.

Afficher les tables et les règles d'analyse

Pour afficher les tables associées à la collaboration et aux règles d'analyse

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Sélectionnez l'une des méthodes suivantes :
 - a. Pour afficher les tables associées à la collaboration, pour Tables associées par vous, choisissez une table (texte bleu).
 - b. Pour afficher les autres tables associées à la collaboration, pour Tables associées par des collaborateurs, choisissez une table (texte bleu).
6. Consultez les détails de la table et les règles d'analyse sur la page des détails de la table.

Modification d'une table configurée

Prérequis :

- Et Compte AWS avec accès à AWS Clean Rooms

Les sections suivantes expliquent comment modifier le nom, la description et les détails de configuration des tables pour les sources de données Amazon S3, Amazon Athena et Snowflake.

Pour plus d'informations sur la façon de modifier une table configurée à l'aide du AWS SDKs, consultez la [référence de l'AWS Clean Rooms API](#).

Pour modifier une table configurée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez la table configurée que vous avez créée.
4. Sur la page détaillée du tableau configuré, choisissez Modifier.
5. Modifiez votre configuration.
6. Sélectionnez Enregistrer les modifications.

Modification des balises de tableau configurées

En tant que membre de la collaboration, après avoir créé une table configurée, vous pouvez gérer les balises de la ressource de table configurée dans l'onglet Tables configurées.

Pour modifier les balises de table configurées

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez la table configurée que vous avez créée.
4. Sur la page détaillée du tableau configuré, faites défiler la page vers le bas jusqu'à la section Tags.
5. Choisissez Gérer les balises.
6. Sur la page de gestion des étiquettes, vous pouvez effectuer les opérations suivantes :
 - Pour supprimer une identification, choisissez Supprimer.

- Pour ajouter une balise, sélectionnez Ajouter une nouvelle balise.
- Sélectionnez Enregistrer pour enregistrer les modifications.

Modification de la règle d'analyse de table configurée

Pour modifier la règle d'analyse de table configurée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez la table configurée que vous avez créée.
4. Sur la page détaillée de la table configurée, faites défiler la page vers le bas jusqu'à la section Règle d'analyse d'agrégation, Règle d'analyse de liste ou Règle d'analyse personnalisée. (Votre choix dépend du type de règle d'analyse que vous avez choisi pour la table configurée.)
5. Choisissez Modifier.
6. Sur la page Modifier la règle d'analyse, vous pouvez :
 - Modifiez la définition de la règle d'analyse en :
 - Modification de l'éditeur JSON.
 - Choisissez Importer depuis un fichier pour télécharger une nouvelle définition de règle d'analyse.
 - Prévisualisez ce que les membres verront dans une collaboration en sélectionnant l'une des options suivantes :
 - Vue du tableau
 - JSON
 - Exemple de requête
7. Choisissez Enregistrer les Modifications pour enregistrer vos Modifications.

Suppression de la règle d'analyse de table configurée

Warning

Cette action est irréversible et a un impact sur toutes les ressources associées.

Pour supprimer la règle d'analyse de table configurée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Choisissez la table configurée que vous avez créée.
4. Sur la page détaillée de la table configurée, faites défiler la page vers le bas jusqu'à la section Règle d'analyse d'agrégation, Règle d'analyse de liste ou Règle d'analyse personnalisée. (Votre choix dépend du type de règle d'analyse que vous avez choisi pour la table configurée.)
5. Sélectionnez Delete (Supprimer).
6. Si vous êtes certain de vouloir supprimer la règle d'analyse, choisissez Supprimer.

Colonnes interdites dans le tableau configuré

La configuration des colonnes de sortie interdites est un contrôle de la règle d'analyse AWS Clean Rooms personnalisée qui vous permet de définir la liste des colonnes (le cas échéant) dont vous n'autorisez pas la projection dans le résultat de la requête. Les colonnes référencées dans cette liste sont considérées comme des « colonnes de sortie interdites ». Cela signifie que toute référence à une telle colonne par transformation, aliasing ou autre moyen peut ne pas être présente dans le SELECT final (projection) de la requête.

Bien que cette fonctionnalité empêche les colonnes d'être directement projetées dans la sortie, elle n'empêche pas totalement les valeurs sous-jacentes d'être déduites indirectement par d'autres mécanismes. Ces colonnes peuvent toujours être utilisées dans une clause de projection (telle que dans une sous-requête ou une expression de table commune (CTE)), tant qu'elles ne sont pas référencées dans la toute dernière projection.

La configuration des colonnes de sortie interdites vous donne la flexibilité d'appliquer et de codifier le contrôle sur votre table en combinaison avec des révisions au niveau des modèles d'analyse basées sur des cas d'utilisation et des exigences de confidentialité correspondantes.

Pour plus d'informations sur la façon de définir cette configuration, consultez [Ajouter une règle d'analyse personnalisée à un tableau \(flux guidé\)](#).

Exemples

Les exemples suivants montrent comment le contrôle des colonnes de sortie interdites est appliqué.

- Le membre A est en collaboration avec le membre B.
- Le membre B est le membre qui peut exécuter des requêtes.
- Le membre A définit les utilisateurs d'une table avec les colonnes âge, sexe, e-mail et nom. L'âge et le nom des colonnes sont des colonnes de sortie interdites.
- Le membre B définit un animal de table avec un ensemble similaire de colonnes age, sexe et nom_propriétaire. Cependant, ils ne définissent aucune contrainte sur les colonnes de sortie, ce qui signifie que toutes les colonnes de la table peuvent être projetées librement dans la requête.

Si le membre B exécute la requête suivante, elle est bloquée car les colonnes de sortie interdites ne peuvent pas être projetées directement :

```
SELECT
  age
FROM
  users
```

Si le membre B exécute la requête suivante, elle est bloquée car les colonnes de sortie interdites ne peuvent pas être implicitement projetées via project star :

```
SELECT
  *
FROM
  users
```

Si le membre B exécute la requête suivante, elle est bloquée car les transformations des colonnes de sortie interdites ne peuvent pas être projetées :

```
SELECT
  COUNT(age)
FROM
  users
```

Si le membre B exécute la requête suivante, elle est bloquée car les colonnes de sortie interdites ne peuvent pas être référencées dans la projection finale à l'aide d'un alias :

```
SELECT
  count_age
FROM
  (SELECT COUNT(age) AS count_age FROM users)
```

Si le membre B exécute la requête suivante, elle est bloquée car les colonnes restreintes transformées sont projetées dans la sortie :

```
SELECT
  CONCAT(name, email)
FROM
  users
```

Si le membre B exécute la requête suivante, elle est bloquée car les colonnes de sortie interdites définies dans CTE ne peuvent pas être référencées dans la projection finale :

```
WITH cte AS (
  SELECT
    age AS age_alias
  FROM
    users
)
SELECT age_alias FROM cte
```

Si le membre B exécute la requête suivante, elle est bloquée car les colonnes de sortie interdites ne peuvent pas être utilisées comme clés de tri ou de partition dans la projection finale :

```
SELECT
  LISTAGG(gender) WITHIN GROUP (ORDER BY age) OVER (PARTITION BY age)
FROM
  users
```

Si le membre B exécute la requête suivante, elle aboutit car les colonnes faisant partie des colonnes de sortie interdites peuvent toujours être utilisées dans d'autres constructions de la requête, par exemple dans les clauses de jointure ou de filtrage.

```
SELECT
  u.name,
  p.gender,
  p.age
FROM
  users AS u
JOIN
  pets AS p
ON
  u.name = p.owner_name
```

Dans le même scénario, le membre B peut également utiliser la colonne de nom dans les utilisateurs comme filtre ou clé de tri :

```
SELECT
  u.email,
  u.gender
FROM
  users AS u
WHERE
  u.name = 'Mike'
ORDER BY
  u.name
```

En outre, les colonnes de sortie interdites aux utilisateurs peuvent être utilisées dans des projections intermédiaires telles que des sous-requêtes et CTEs, par exemple :

```
WITH cte AS (
  SELECT
    u.gender,
    u.id,
    u.first_name
  FROM
    users AS u
)
SELECT
  first_name
FROM
  (SELECT cte.gender, cte.id, cte.first_name FROM cte)
```

Modification des associations de tables configurées

En tant que membre de la collaboration, vous pouvez modifier les associations de tables configurées que vous avez créées.

Pour modifier les associations de tables configurées

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.
5. Pour les tables que vous avez associées, choisissez une table.
6. Sur la page des détails de la table, faites défiler la page vers le bas pour afficher les détails de l'association de tables.
7. Choisissez Modifier.
8. Sur la page Modifier les associations de tables configurées, mettez à jour la description ou les informations d'accès au service.
9. Sélectionnez Enregistrer les modifications.

Dissociation des tables configurées

En tant que membre de la collaboration, vous pouvez dissocier une table configurée de la collaboration. Cette action empêche le membre autorisé à interroger la table.

Pour dissocier une table configurée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Tables.

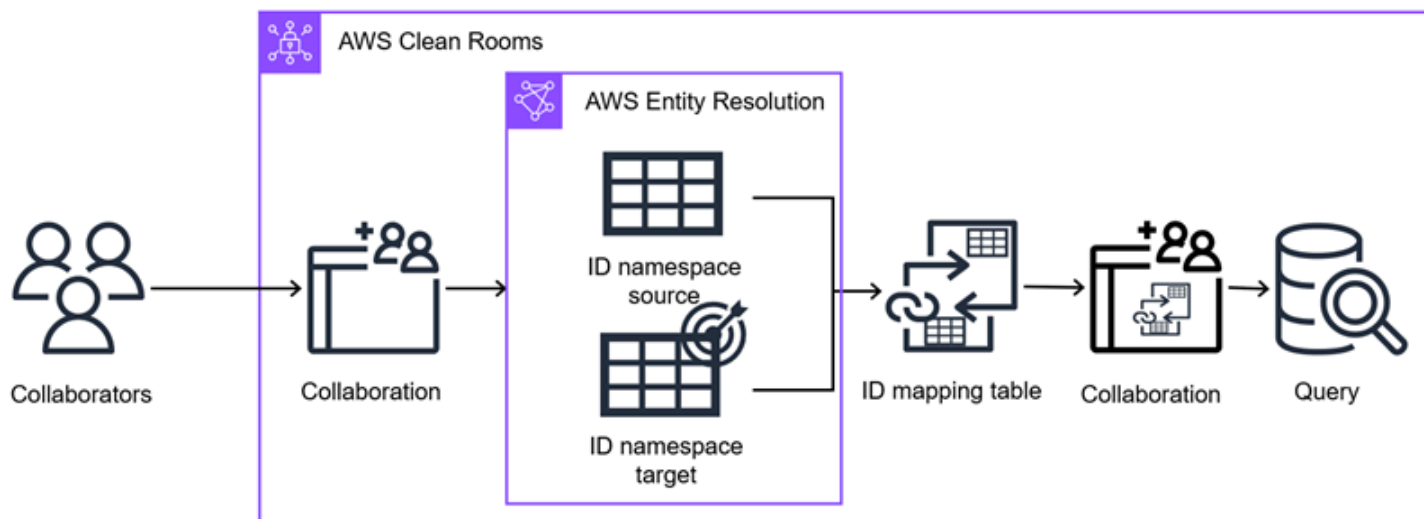
5. Pour les tables que vous avez associées, sélectionnez le bouton d'option situé à côté de la table que vous souhaitez dissocier.
6. Choisissez Dissocier.
7. Dans la boîte de dialogue, confirmez la décision de dissocier la table configurée et empêchez le membre autorisé à interroger la table en choisissant Dissocier.

Résolution des entités AWS dans AWS Clean Rooms

Grâce à la Résolution des entités AWS dans AWS Clean Rooms, vous pouvez traduire les données d'une source vers une cible, remplir une table de mappage d'identifiants avec les données traduites et interroger les données.

Tout d'abord, vous créez une collaboration AWS Clean Rooms et vous y ajoutez celle que vous souhaitez inviter, ou vous rejoignez une collaboration à laquelle vous êtes invité en créant un abonnement. Ensuite, vous effectuez le mappage des identifiants sur deux tables de données. Pour ce faire, associez une source d'espace de noms d'ID existante ou créez-en une nouvelle dans la Résolution des entités AWS. L'autre membre de la collaboration associe une cible d'espace de noms d'ID existante ou crée une nouvelle cible d'espace de noms d'ID. Ensuite, vous créez et remplissez une table de mappage d'ID à partir des deux espaces de noms d'ID associés. Enfin, le membre qui peut effectuer une requête exécute une requête dans les deux tables de données en se joignant à la table de mappage d'identifiants.

Le schéma suivant résume la manière d'utiliser la Résolution des entités AWS dans AWS Clean Rooms.



Note

Le fournisseur de services de transcodage actuellement pris en charge est LiveRamp, qui est disponible dans les Régions AWS pays suivants : USA Est (Virginie du Nord), USA Est (Ohio) et USA Ouest (Oregon).

- [Espaces de noms d'ID dans AWS Clean Rooms](#)
- [Tables de mappage d'identifiants dans AWS Clean Rooms](#)

Espaces de noms d'ID dans AWS Clean Rooms

Un espace de noms d'identification est une enveloppe entourant votre table d'identité qui vous permet de fournir des métadonnées expliquant votre ensemble de données et expliquant comment l'utiliser dans un flux de travail de mappage d'identifiants. Un flux de travail de mappage d'ID est une tâche de traitement de données qui mappe les données d'une source de données d'entrée vers une cible de données d'entrée en fonction de la méthode de mappage d'ID spécifiée. Il produit une table de mappage des identifiants.

Il existe deux types d'espaces de noms d'ID : source et cible. La source contient des configurations pour les données source qui seront traitées dans un flux de travail de mappage d'identifiants. La cible contient une configuration des données cibles vers laquelle toutes les sources seront résolues. Pour définir les données d'entrée que vous souhaitez résoudre entre deux Comptes AWS, créez une source d'espace de noms ID et une cible d'espace de noms ID pour traduire vos données d'un ensemble (Source) à un autre (Target).

Vous pouvez soit créer un nouvel espace de noms d'ID, soit associer un espace existant. Pour plus d'informations sur la création d'un espace de noms d'ID dans Résolution des entités AWS, voir [Définir les données d'entrée à l'aide d'un espace de noms d'ID](#) dans le Guide de l'Utilisateur des entités AWS.

Rubriques

- [Création et association d'un nouvel espace de noms d'ID](#)
- [Associer un espace de noms d'ID existant](#)
- [Modification des associations d'espaces de noms d'ID](#)
- [Dissociation des associations d'espaces de noms d'ID](#)

Création et association d'un nouvel espace de noms d'ID

Chaque membre de la collaboration doit créer et associer un espace de noms ID Source ou un espace de noms ID Target avant de créer une table de mappage d'identifiants pour interroger les données d'identité.

Si vous avez déjà créé un espace de noms d'ID dans Résolution des entités AWS, passez directement à [Associer un espace de noms d'ID existant](#).

Pour créer et associer un nouvel espace de noms d'ID

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Résolution de l'entité, choisissez l'espace de noms Associate ID.
5. Sur la page espace de noms Associate ID, pour les données de résolution d'entité, choisissez Create ID namespace.

La Résolution des entités AWS console apparaît dans un nouvel onglet.

6. Suivez les instructions de la page Create ID namespace de la Résolution des entités AWS console.
 - a. Pour plus de détails, entrez le nom de l'espace de noms d'ID, la description, puis sélectionnez le type d'espace de noms d'ID (source ou cible).
 - b. Pour la méthode de l'espace de noms ID, choisissez soit la méthode basée sur des règles pour le rapprochement basé sur des règles, soit les services du fournisseur pour le transcodage tiers.
 - c. Spécifiez le type de saisie des données, en fonction de la méthode d'espace de noms ID que vous avez choisie.
 - d. Choisissez Create ID namespace.
7. Retournez à la AWS Clean Rooms console.
8. Sur la page Espace de noms d'ID associé, pour les données de résolution d'entité, choisissez la source ou la cible de l'espace de noms d'ID de Résolution des entités AWS ID que vous souhaitez associer à la collaboration dans la liste déroulante.
9. Pour obtenir des informations sur l'association, procédez comme suit.
 - a. Entrez un nom pour l'espace de noms ID associé.

Vous pouvez utiliser le nom par défaut ou renommer cet espace de noms d'ID.

- b. (Facultatif) Entrez une description de l'espace de noms ID.

La description facilite la rédaction de requêtes.

10. Spécifiez les autorisations AWS Clean Rooms d'accès en sélectionnant une option, puis en prenant les mesures recommandées.

Option	Action recommandée
AWS Clean Rooms Autoriser l'ajout et la gestion de la politique d'autorisation	AWS Clean Rooms crée un rôle de service avec la politique requise pour cette association.
Ajouter et gérer les autorisations manuellement	Effectuez l'une des actions suivantes : • Passez en revue la politique des ressources et ajoutez-y les autorisations nécessaires. • Utilisez une politique existante en choisissant Ajouter une déclaration de politique. Vous devez disposer des autorisations nécessaires pour modifier les rôles et créer des politiques.  Note Si vous ne pouvez pas modifier la politique de rôle, vous recevez un message d'erreur indiquant que vous n'avez pas AWS Clean Rooms trouvé la politique pour le rôle de service.

11. (Facultatif) Pour les configurations avancées des tables de mappage d'identifiants, modifiez les protections par défaut pour la colonne provenant de l'espace de noms d'identifiants.

La table de mappage des identifiants est configurée par défaut pour autoriser uniquement un à la fois INNER JOIN sur la sourceID colonne et sur la targetID colonne. Vous pouvez modifier

cette configuration afin que la colonne provenant de cet espace de noms d'ID (soit `targetID`) `sourceID` soit autorisée n'importe où dans la requête.

Votre objectif	Option recommandée
Catégoriser la colonne en tant que « colonne de jointure » et ne l'autoriser que dans une clause <code>INNER JOIN</code>	Oui
Catégorisez la colonne en tant que « colonne de dimension » et autorisez-la n'importe où dans la requête, y compris dans une <code>JOIN</code> clause <code>WHERE</code> et dans les <code>GROUP BY</code> instructions de la requête. <code>SELECT</code>	Non, autoriser n'importe où dans la requête

- (Facultatif) Si vous souhaitez activer les balises pour la ressource ID namespace, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
- Choisissez Associer.
- Dans l'onglet Résolution des entités, sous le tableau des espaces de noms ID associés, consultez l'espace de noms ID associé et vérifiez que le type d'espace de noms ID est correct (source ou cible).

Une fois que tous les membres de la collaboration ont associé leurs espaces de noms d'identification, vous pouvez [créer une table de mappage d'identifiants](#) et interroger les données.

Associer un espace de noms d'ID existant

Dans cette procédure, chaque membre associe sa source d'espace de noms d'ID existante ou sa cible d'espace de noms d'ID dans la collaboration.

Pour associer un espace de noms d'ID existant

- Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
- Dans le volet de navigation de gauche, sélectionnez Collaborations.
- Choisissez la collaboration.
- Dans l'onglet Résolution de l'entité, choisissez l'espace de noms Associate ID.

5. Sur la page Espace de noms d'ID associé, pour les données de résolution d'entité, choisissez la source ou la cible de l'espace de noms d'Résolution des entités AWS ID que vous souhaitez associer à la collaboration dans la liste déroulante.
6. Pour obtenir des informations sur l'association, procédez comme suit.

- a. Entrez un nom pour l'espace de noms ID associé.

Vous pouvez utiliser le nom par défaut ou renommer cet espace de noms d'ID.

- b. (Facultatif) Entrez une description de l'espace de noms ID.

La description facilite la rédaction de requêtes.

7. Spécifiez les autorisations AWS Clean Rooms d'accès en sélectionnant une option, puis en prenant les mesures recommandées.

Option	Action recommandée
AWS Clean Rooms Autoriser l'ajout et la gestion de la politique d'autorisation	AWS Clean Rooms crée un rôle de service avec la politique requise pour cette association.
Ajouter et gérer les autorisations manuellement	Effectuez l'une des actions suivantes : • Passez en revue la politique des ressources et ajoutez-y les autorisations nécessaires. • Utilisez une politique existante en choisissant Ajouter une déclaration de politique. Vous devez disposer des autorisations nécessaires pour modifier les rôles et créer des politiques.  Note Si vous ne pouvez pas modifier la politique de rôle, vous recevez un

Option	Action recommandée
	message d'erreur indiquant que vous n'avez pas AWS Clean Rooms trouvé la politique pour le rôle de service.

8. (Facultatif) Pour les configurations avancées des tables de mappage d'identifiants, modifiez les protections par défaut pour la colonne provenant de l'espace de noms d'identifiants.

La table de mappage des identifiants est configurée par défaut pour autoriser uniquement un à la fois INNER JOIN sur la sourceID colonne et sur la targetID colonne. Vous pouvez modifier cette configuration afin que la colonne provenant de cet espace de noms d'ID (soit targetID) sourceID soit autorisée n'importe où dans la requête.

Votre objectif	Option recommandée
Catégorisez la colonne en tant que « colonne de jointure » et autorisez-la uniquement dans une INNER JOIN clause.	Oui
Catégorisez la colonne en tant que « colonne de dimension » et autorisez-la n'importe où dans la requête, y compris dans une JOIN clause SELECT, WHERE, et des GROUP BY instructions de la requête.	Non, autoriser n'importe où dans la requête

9. (Facultatif) Si vous souhaitez activer les balises pour la ressource ID namespace, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
10. Choisissez Associer.
11. Dans l'onglet Résolution des entités, sous le tableau des espaces de noms ID associés, consultez l'espace de noms ID associé et vérifiez que le type d'espace de noms ID est correct (source ou cible).

Une fois que tous les membres de la collaboration ont associé leurs espaces de noms d'identification, vous pouvez [créer une table de mappage d'identifiants](#) et interroger les données.

Modification des associations d'espaces de noms d'ID

En tant que membre de la collaboration, vous pouvez modifier les associations d'espaces de noms d'ID que vous avez créées.

Pour modifier une association d'espaces de noms d'ID

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Résolution de l'entité.
5. Pour les espaces de noms d'ID associés, choisissez un espace de noms d'ID.
6. Sur la page des détails de l'espace de nommage ID, faites défiler la page vers le bas pour afficher les détails de l'association de l'espace de nommage ID.
7. Choisissez Modifier.
8. Sur la page Modifier les associations d'espaces de noms ID, modifiez l'un des éléments suivants :
 - a. Pour les détails de l'association, mettez à jour le nom ou la description.
 - b. (Facultatif) Pour les configurations avancées des tables de mappage d'identifiants, modifiez les protections par défaut pour la colonne provenant de l'espace de noms d'identifiants.

La table de mappage des identifiants est configurée par défaut pour autoriser uniquement un à la fois INNER JOIN sur la sourceID colonne et sur la targetID colonne. Vous pouvez modifier cette configuration afin que la colonne provenant de cet espace de noms d'ID (soit targetID) sourceID soit autorisée n'importe où dans la requête.

Votre objectif	Option recommandée
Catégoriser la colonne en tant que « colonne de jointure » et ne l'autoriser que dans une clause INNER JOIN	Oui
Catégorisez la colonne en tant que « colonne de dimension » et autorisez-la	Non, autoriser n'importe où dans la requête

Votre objectif	Option recommandée
n'importe où dans la requête, y compris dans une JOIN clause WHERE et dans les GROUP BY instructions de la requête. SELECT	

9. Sélectionnez Enregistrer les modifications.

Dissociation des associations d'espaces de noms d'ID

En tant que membre de la collaboration, vous pouvez dissocier un espace de noms ID de la collaboration. Cette action empêche le membre autorisé à interroger la table.

Warning

Dissocier une association d'espaces de noms d'ID d'une collaboration supprime toutes les données des tables de mappage d'identifiants dérivées, les rendant ainsi ininterrogeables. Par exemple, si votre association d'espace de noms d'ID a été utilisée comme SOURCE dans trois tables de mappage d'ID différentes, toutes les données de ces tables de mappage d'ID seront supprimées lorsque vous dissocierez votre association d'espace de noms d'ID.

Pour dissocier une association d'espaces de noms ID

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Résolution de l'entité.
5. Pour les espaces de noms d'ID associés, sélectionnez le bouton d'option à côté de l'espace de noms d'ID que vous souhaitez dissocier.
6. Choisissez Dissocier.
7. Dans la boîte de dialogue, confirmez votre décision de déconnecter l'espace de noms ID en choisissant Dissocier. Cette action empêche tout membre autorisé à effectuer une requête d'accéder à la table de mappage des identifiants.

Si un membre de la collaboration supprime l'un des espaces de noms d'identification, vous ne pouvez pas remplir à nouveau la table de mappage des identifiants si la source a quitté la collaboration.

Même si la table de mappage des identifiants a déjà été remplie, la dissociation de l'espace de noms des identifiants signifie que vous ne pouvez plus exécuter de requêtes sur cette table.

Tables de mappage d'identifiants dans AWS Clean Rooms

Une table de mappage d'identifiants est une ressource AWS Clean Rooms qui permet le mappage d'identité multipartite dans le cadre d'une collaboration.

Avant de créer une table de mappage d'ID, vous devez d'abord configurer les données source et cible en tant qu'espaces de noms d'ID.

Après avoir créé une table de mappage d'ID, vous utilisez un flux de travail de mappage d'ID pour traduire l'espace de noms d'ID source en espace de noms d'ID cible. Pour ce faire, vous pouvez utiliser une méthode basée sur des règles ou une méthode de transcodage des services du fournisseur.

Un flux de travail de mappage d'identification est une tâche de traitement de données qui mappe les données d'une source de données d'entrée vers une cible de données d'entrée en fonction de la méthode de flux de travail de mappage d'identification spécifiée. Ce flux de travail remplit une table de mappage d'identifiants.

Note

Les tables de mappage d'identifiants ne peuvent être créées qu'à partir d'ensembles de données stockés dans Amazon S3 et analysés dans AWS Glue des tables.

Il existe deux méthodes de mappage des identifiants : le mappage des identifiants basé sur des règles ou le mappage des identifiants des fournisseurs de services :

- Mappage d'ID basé sur des règles : vous utilisez des règles correspondantes pour convertir les données internes sources en cibles.
- Mappage des identifiants des services du LiveRamp fournisseur : vous utilisez le service du fournisseur pour traduire des données tierces d'une source vers une cible.

 Note

Le fournisseur de services de transcodage actuellement pris en charge est LiveRamp. Tout membre de la collaboration abonné à LiveRamp Through AWS Data Exchange peut créer la table de mappage des identifiants. Si vous avez déjà souscrit un abonnement LiveRamp, mais pas par le biais de AWS Data Exchange celui-ci, contactez-nous LiveRamp pour obtenir une offre privée. Pour plus d'informations, consultez la section [Abonnement à un service fournisseur AWS Data Exchange dans le](#) guide de Résolution des entités AWS l'utilisateur.

Rubriques

- [Création et remplissage d'une nouvelle table de mappage d'identifiants](#)
- [Remplissage d'une table de mappage d'identifiants existante](#)
- [Modification d'une table de mappage d'identifiants](#)
- [Supprimer une table de mappage d'identifiants](#)

Création et remplissage d'une nouvelle table de mappage d'identifiants

Conditions préalables

Avant de créer une table de mappage d'identifiants, assurez-vous d'avoir :

- Un ID, un espace de noms, une source et une cible associées
- Les espaces de noms d'ID configurés soit pour le mappage d'ID basé sur des règles, soit pour le mappage d'ID des services du fournisseur

Vous pouvez créer deux types de tables de mappage d'identifiants :

- Basé sur des règles — Utilise des règles de correspondance pour traduire les données de première partie
- Prestataire de services — Utilisations LiveRamp pour traduire « Ramp » IDs

Après avoir créé une table de mappage d'ID, vous pouvez soit la remplir immédiatement en exécutant un flux de travail de mappage d'ID, soit attendre de remplir la table ultérieurement.

Une fois que la table de mappage des identifiants est correctement remplie, vous pouvez exécuter une requête de jointure multitable sur la table de mappage des identifiants pour les `sourceId` joindre aux données `targetId` et les analyser.

Rubriques

- [Création d'une table de mappage d'identifiants \(basée sur des règles\)](#)
- [Création d'une table de mappage des identifiants \(services du fournisseur\)](#)

Création d'une table de mappage d'identifiants (basée sur des règles)

Cette rubrique décrit le processus de création d'une table de mappage d'identifiants qui utilise des règles de correspondance pour traduire les données de première partie d'une source vers une cible.

Lorsque vous créez une table de mappage d'identifiants basée sur des règles, vous pouvez choisir de traiter uniquement les enregistrements nouveaux, mis à jour ou supprimés dans le flux de travail en activant le traitement incrémentiel.

Pour créer et remplir une nouvelle table de mappage d'identifiants à l'aide de la méthode basée sur des règles

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
2. Dans le volet de navigation, sélectionnez Collaborations.
3. Choisissez la collaboration, puis cliquez sur l'onglet Résolution des entités.
4. Choisissez Créer une table de mappage d'identifiants.
5. Dans Paramètres de mappage des identifiants, effectuez l'une des opérations suivantes :
 - Pour créer un nouveau flux de travail, maintenez l'option Créer un nouveau flux de travail de mappage d'identifiants sélectionnée.
 - Pour utiliser un flux de travail existant, décochez la case et sélectionnez un flux de travail dans la liste déroulante, puis passez à l'étape 9.
6. Sous Données d'identité, consultez ou configurez la source et la cible.

- Pour une paire d'espaces de noms ID unique : passez en revue la source et la cible présélectionnées.
 - Pour plusieurs espaces de noms d'identification : sélectionnez la source et la cible dans les listes déroulantes.
7. (Facultatif) Sélectionnez Activer le traitement incrémentiel pour traiter uniquement les enregistrements nouveaux, mis à jour ou supprimés dans le flux de travail.

Résolution des entités AWS traite uniquement les enregistrements nouveaux, mis à jour ou supprimés dans l'espace de noms source ou cible, plutôt que de recréer l'intégralité de la table de mappage des identifiants.

Si vous ne sélectionnez pas cette option, Résolution des entités AWS exécute le flux de travail de mappage d'identifiants de traitement par lots par défaut sur la table de mappage d'identifiants.

8. Sous Paramètres de la règle, configurez les éléments suivants :
- Contrôles des règles : choisissez si la cible ou la source fournit des règles correspondantes.

Vous pouvez consulter les règles en activant Afficher les règles.

Les contrôles de règles doivent être compatibles entre l'espace de noms d'ID source et cible à utiliser dans un flux de travail de mappage d'ID. Par exemple, si un espace de noms d'ID source limite les règles à la cible mais que l'espace de noms d'ID cible limite les règles à la source, cela entraîne une erreur.

- Le type de comparaison est automatiquement défini sur Plusieurs champs de saisie.

Cela est dû au fait que les deux participants avaient précédemment sélectionné cette option.

- Correspondance des enregistrements
 - Une source pour une cible — Stocke un enregistrement correspondant par cible
 - Plusieurs sources vers une seule cible — Stocke tous les enregistrements correspondants par cible

 Note

Les limites spécifiées pour les espaces de noms d'ID source et cible doivent être compatibles.

9. Pour obtenir des informations détaillées sur le mappage des identifiants, configurez les éléments suivants :

- a. Entrez le nom d'une table de mappage d'identifiants ou conservez le nom par défaut.
- b. (Facultatif) Entrez une description de la table de mappage des identifiants.

La description facilite la rédaction de requêtes.

10. Pour AWS Clean Rooms y accéder, choisissez-en un :

- AWS Clean Rooms Autoriser l'ajout et la gestion des politiques d'autorisation : crée automatiquement un rôle de service.
- Ajouter et gérer les autorisations manuellement : passez en revue et modifiez la politique de ressources ou choisissez Ajouter une déclaration de politique.

 Note

Si vous ne parvenez pas à modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique pour le rôle de service est introuvable.

11. Pour Résolution des entités AWS y accéder, choisissez-en un :

Cette section n'est visible que si vous créez une nouvelle table de mappage d'identifiants.

- Création et utilisation d'un nouveau rôle de service
 - Le nom du rôle de service par défaut est `entityresolution-id-mapping-workflow-
<timestamp>`
 - (Facultatif) Pour les données chiffrées, sélectionnez Ces données sont chiffrées par une clé KMS et entrez la AWS KMS clé.
- Utiliser un rôle de service existant
 - Choisissez un nom de rôle de service existant dans la liste déroulante ou entrez un ARN de rôle.

La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles.

Affichez le rôle de service en choisissant le lien externe Afficher dans IAM.

S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible.

Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

- (Facultatif) Sélectionnez Ajouter une politique préconfigurée avec les autorisations nécessaires pour ce rôle pour associer les autorisations nécessaires au rôle.

Vous devez disposer des autorisations nécessaires pour modifier les rôles et créer des politiques.

12. (Facultatif) Sous Paramètres supplémentaires, configurez :

a. Paramètres de la table de mappage des identifiants

- Pour activer le chiffrement personnalisé, choisissez Personnaliser les paramètres de chiffrement et entrez une AWS KMS clé.

Note

Cette clé KMS doit accorder les autorisations requises pour être utilisée dans le cadre de `cleanrooms.amazonaws.com` l'utilisation Résolution des entités AWS d'une politique de clé KMS. Pour plus de détails sur les autorisations requises pour utiliser des chiffrements avec un flux de travail de mappage d'identifiants, voir [Créer un rôle de travail dans le flux de travail Résolution des entités AWS](#) dans le guide de l'Utilisateur de la Résolution des entités AWS.

- Pour ajouter des balises, choisissez Ajouter une nouvelle balise et entrez des paires clé-valeur

b. Paramètres du flux de travail de mappage des identifiants (nouveaux flux de travail uniquement) :

- Pour utiliser des noms différents, décochez la case Conserver le même nom et la même description de la table de mappage d'identifiants et entrez de nouvelles valeurs.
- Pour ajouter des balises, choisissez Ajouter une nouvelle balise et entrez des paires clé-valeur

13. Sélectionnez l'une des méthodes suivantes :

- Créer une table de mappage d'identifiants — Crée une table vide que vous pourrez remplir ultérieurement () [Remplissage d'une table de mappage d'identifiants existante](#)
- Création et remplissage de la table de mappage des identifiants : crée et remplit immédiatement la table (cela peut prendre plusieurs heures)

Le processus de mappage des identifiants commence. Au cours de ce processus, la table de mappage des identifiants est remplie avec une traduction IDs. Le traitement du flux de travail de mappage des identifiants peut prendre quelques heures.

Une fois que la table de mappage des identifiants est correctement remplie, vous pouvez [interroger la table de mappage des identifiants](#) pour les `sourceId` joindre aux données `targetId` et les analyser.

Création d'une table de mappage des identifiants (services du fournisseur)

Cette rubrique décrit le processus de création d'une table de mappage d'identifiants utilisant un service fournisseur (LiveRamp). Les services du LiveRamp fournisseur traduisent un ensemble de Ramp source IDs en un autre en utilisant Ramp maintenu ou dérivé IDs.

Pour créer une nouvelle table de mappage d'identifiants à l'aide de la méthode des services du fournisseur

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.
2. Dans le volet de navigation, sélectionnez Collaborations.
3. Choisissez la collaboration, puis cliquez sur l'onglet Résolution des entités.
4. Choisissez Créer une table de mappage d'identifiants.
5. Dans Paramètres de mappage des identifiants, effectuez l'une des opérations suivantes :
 - Pour créer un nouveau flux de travail, maintenez l'option Créer un nouveau flux de travail de mappage d'identifiants sélectionnée.
 - Pour utiliser un flux de travail existant, décochez la case et sélectionnez un flux de travail dans la liste déroulante, puis passez à l'étape 9.
6. Sous Données d'identité, consultez ou configurez la source et la cible.

- Pour une paire d'espaces de noms ID unique : passez en revue la source et la cible présélectionnées.
 - Pour plusieurs espaces de noms d'identification : sélectionnez la source et la cible dans les listes déroulantes.
7. Sous Méthode, vérifiez que la méthode de flux de travail de mappage d'identifiants sélectionnée est un LiveRamp transcodage.
 8. Pour les LiveRamp configurations, effectuez l'une des opérations suivantes :
 - Entrez l'ARN du gestionnaire LiveRamp d'identifiants et l'ARN du gestionnaire de LiveRamp secrets.
 - Choisissez Importer à partir d'un flux de travail existant.
 9. Pour obtenir des informations détaillées sur le mappage des identifiants, configurez les éléments suivants :
 - a. Entrez le nom d'une table de mappage d'identifiants ou conservez le nom par défaut.
 - b. (Facultatif) Entrez une description de la table de mappage des identifiants.

La description facilite la rédaction de requêtes.
 10. Pour AWS Clean Rooms y accéder, choisissez-en un :
 - AWS Clean Rooms Autoriser l'ajout et la gestion des politiques d'autorisation : crée automatiquement un rôle de service.
 - Ajouter et gérer les autorisations manuellement : passez en revue et modifiez la politique de ressources ou choisissez Ajouter une déclaration de politique.

 Note

Si vous ne parvenez pas à modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique pour le rôle de service est introuvable.

11. Pour Résolution des entités AWS y accéder, choisissez-en un :

Cette section n'est visible que si vous créez une nouvelle table de mappage d'identifiants.

- Création et utilisation d'un nouveau rôle de service

- Le nom du rôle de service par défaut est `entityresolution-id-mapping-workflow-
<timestamp>`
- (Facultatif) Pour les données chiffrées, sélectionnez Ces données sont chiffrées par une clé KMS et entrez la AWS KMS clé.
- Utiliser un rôle de service existant
- Choisissez un nom de rôle de service existant dans la liste déroulante ou entrez un ARN de rôle.

La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles.

Affichez le rôle de service en choisissant le lien externe Afficher dans IAM.

S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible.

Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

- (Facultatif) Sélectionnez Ajouter une politique préconfigurée avec les autorisations nécessaires pour ce rôle pour associer les autorisations nécessaires au rôle.

Vous devez disposer des autorisations nécessaires pour modifier les rôles et créer des politiques.

12. (Facultatif) Sous Paramètres supplémentaires, configurez :

a. Paramètres de la table de mappage des identifiants

- Pour activer le chiffrement personnalisé, choisissez Personnaliser les paramètres de chiffrement et entrez une AWS KMS clé.

Note

Cette clé KMS doit accorder les autorisations requises pour être utilisée dans le cadre de `cleanrooms.amazonaws.com` l'utilisation Résolution des entités AWS d'une politique de clé KMS. Pour plus de détails sur les autorisations requises pour utiliser des chiffrements avec un flux de travail de mappage d'identifiants, voir [Créer un rôle de travail dans le flux de travail Résolution des entités AWS](#) dans le guide de l'utilisateur de la Résolution des entités AWS.

- Pour ajouter des balises, choisissez Ajouter une nouvelle balise et entrez des paires clé-valeur.
- b. Paramètres du flux de travail de mappage des identifiants (nouveaux flux de travail uniquement) :
 - Pour utiliser des noms différents, décochez la case Conserver le même nom et la même description de la table de mappage d'identifiants et entrez de nouvelles valeurs.
 - Pour ajouter des balises, choisissez Ajouter une nouvelle balise et entrez des paires clé-valeur.

13. Sélectionnez l'une des méthodes suivantes :

- Créer une table de mappage d'identifiants — Crée une table vide que vous pourrez remplir ultérieurement () [Remplissage d'une table de mappage d'identifiants existante](#)
- Création et remplissage de la table de mappage des identifiants : crée et remplit immédiatement la table (cela peut prendre plusieurs heures)

Le processus de mappage des identifiants commence. Au cours de ce processus, la table de mappage des identifiants est remplie avec une traduction IDs. Le traitement du flux de travail de mappage des identifiants peut prendre quelques heures.

Une fois que la table de mappage des identifiants est correctement remplie, vous pouvez [interroger la table de mappage des identifiants](#) pour les `sourceId` joindre aux données `targetId` et les analyser.

Remplissage d'une table de mappage d'identifiants existante

Lorsque de nouvelles données sont ajoutées à un espace de noms d'ID, utilisez ce flux de travail. Si vous avez choisi d'activer le traitement incrémentiel lors de [la création de la table de mappage d'identifiants](#), vous avez la possibilité de traiter uniquement les enregistrements nouveaux, mis à jour ou supprimés dans l'espace de noms source ou cible, plutôt que de recréer la table de mappage d'identifiants dans son intégralité.

Pour remplir une table de mappage d'identifiants existante

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms/>.

2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Accédez à l'onglet Résolution de l'entité.
5. Dans la section Tables de mappage d'ID, sélectionnez une table de mappage d'ID.
6. Si vous n'avez pas activé le traitement incrémentiel pour cette table de mappage d'identifiants, choisissez-en une, puis choisissez Remplir.
7. Si vous avez activé le traitement incrémentiel pour cette table de mappage d'identifiants, choisissez Remplir avec, puis choisissez l'une des options suivantes :

- Traitement incrémentiel : traite uniquement les enregistrements nouveaux, mis à jour ou supprimés dans l'espace de noms Source ou Target ID.

Recommandé pour les mises à jour fréquentes, les exécutions quotidiennes ou la synchronisation des données en temps réel.

- Traitement par lots : traite l'intégralité de la table de mappage des identifiants.

Recommandé pour la configuration initiale, les actualisations complètes périodiques ou lorsque des modifications importantes se produisent à la fois dans les espaces de noms Source et Target ID.

- Traitement uniquement par suppression : traite uniquement les enregistrements supprimés de l'espace de noms Source ID et met à jour l'espace de noms Target ID en conséquence.

Recommandé pour synchroniser rapidement les suppressions.

8. Le processus de mappage des identifiants commence.

Au cours de ce processus, la table de mappage des identifiants est remplie de données transcodées IDs. Le processus de mappage des identifiants peut prendre quelques heures.

Une fois que la table de mappage d'identifiants est correctement remplie, vous pouvez [interroger la table de mappage d'identifiants](#) pour la sourceId joindre à targetId.

Modification d'une table de mappage d'identifiants

En tant que membre de la collaboration, vous pouvez modifier la table de mappage des identifiants que vous avez créée.

Pour modifier une table de mappage d'identifiants

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Résolution de l'entité.
5. Pour les tables de mappage d'identifiants, choisissez une table.
6. Sur la page des détails de la table de mappage des identifiants, faites défiler la page vers le bas pour afficher les détails de la table de mappage des identifiants.
7. Choisissez Modifier.
8. Sur la page de la table de mappage Modifier l'ID, mettez à jour la description ou les informations d'accès au service.
9. Sélectionnez Enregistrer les modifications.

Supprimer une table de mappage d'identifiants

En tant que membre de la collaboration, vous pouvez supprimer une table de mappage d'identifiants que vous avez créée. Cette action empêche le membre autorisé à interroger la table.

Warning

La suppression d'une table de mappage supprime définitivement toutes les données renseignées.

Pour supprimer une table de mappage d'identifiants

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Choisissez l'onglet Résolution de l'entité.
5. Pour les tables de mappage d'identifiants, choisissez une table.

6. Sur la page des détails de la table de mappage des identifiants, faites défiler la page vers le bas pour afficher les tables de mappage des identifiants.
7. Choisissez une table de mappage d'identifiants, puis choisissez Supprimer.
8. Si vous êtes certain de vouloir supprimer la table de mappage des identifiants, choisissez Supprimer.

Modèles d'analyse dans AWS Clean Rooms

Les modèles d'analyse fonctionnent avec [Règle d'analyse personnalisée dans AWS Clean Rooms](#). Avec un modèle d'analyse, vous pouvez définir des paramètres pour vous aider à réutiliser la même requête. AWS Clean Rooms prend en charge un sous-ensemble de paramétrage avec des valeurs littérales.

Les modèles d'analyse sont spécifiques à la collaboration. Pour chaque collaboration, les membres peuvent uniquement voir les requêtes de cette collaboration. Si vous envisagez d'utiliser la confidentialité différentielle dans le cadre d'une collaboration, vous devez vous assurer que vos modèles d'analyse sont compatibles avec la [structure de requête à usage général](#) de AWS Clean Rooms Differential Privacy.

Vous pouvez créer un modèle d'analyse de deux manières : en utilisant du code SQL ou en utilisant du code Python pour Spark.

Rubriques

- [Modèles d'analyse SQL](#)
- [PySpark modèles d'analyse](#)
- [Modèles d' PySpark analyse de résolution des problèmes](#)

Modèles d'analyse SQL

Les modèles d'analyse SQL vous permettent d'interroger et d'analyser les données de différents ensembles de données dans le cadre d'une collaboration. Vous pouvez utiliser ces modèles pour effectuer différents types d'analyse, tels que l'identification des chevauchements d'audience et le calcul de mesures agrégées.

Les modèles d'analyse SQL vous permettent de :

- Écrire des requêtes SQL standard
- Ajoutez des paramètres pour dynamiser vos requêtes
- Contrôlez l'accès à des colonnes et à des tables spécifiques
- Définissez les exigences d'agrégation pour les données sensibles
- Définissez les données d'entrée pour la génération d'ensembles de données synthétiques respectueux de la vie privée pour les modèles d'apprentissage automatique (ML) personnalisés

Rubriques

- [Création d'un modèle d'analyse SQL](#)
- [Révision d'un modèle d'analyse SQL](#)

Création d'un modèle d'analyse SQL

Conditions préalables

Avant de créer un modèle d'analyse SQL, vous devez disposer des éléments suivants :

- Une AWS Clean Rooms collaboration active
- Accès à au moins une table configurée dans la collaboration

Pour plus d'informations sur la configuration des tables dans AWS Clean Rooms, consultez [Création d'une table configurée dans AWS Clean Rooms](#).

- Autorisations pour créer des modèles d'analyse
- Connaissances de base de la syntaxe des requêtes SQL

La procédure suivante décrit le processus de création d'un modèle d'analyse SQL à l'aide de la [AWS Clean Rooms console](#).

Pour plus d'informations sur la création d'un modèle d'analyse SQL à l'aide du AWS SDKs, consultez le manuel de [référence des AWS Clean Rooms API](#).

Pour créer un modèle d'analyse SQL

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Modèles, accédez à la section Modèles d'analyse que vous avez créés.
5. Choisissez Créer un modèle d'analyse.
6. Sur la page Créer un modèle d'analyse, pour plus de détails,
 - a. Entrez un nom pour le modèle d'analyse.
 - b. (Facultatif) Entrez une description.

- c. Pour Format, laissez l'option SQL sélectionnée.
7. Pour les tables, consultez les tables configurées associées à la collaboration.
8. Pour la définition,
 - a. Entrez la définition du modèle d'analyse.
 - b. Choisissez Importer depuis pour importer une définition.
 - c. (Facultatif) Spécifiez un paramètre dans l'éditeur SQL en saisissant deux points (:) devant le nom du paramètre.

Par exemple :

```
WHERE table1.date + :date_period > table1.date
```

9. Si vous avez déjà ajouté des paramètres, sous Paramètres — facultatif, pour chaque nom de paramètre, choisissez le type et la valeur par défaut (facultatif).
10. Pour les données synthétiques, si vous souhaitez générer des données synthétiques pour l'entraînement du modèle, cochez la case Exiger que la sortie du modèle d'analyse soit synthétique.

Pour plus d'informations, voir Génération de jeux de données [synthétiques améliorée en termes de confidentialité](#).

- a. Pour la classification par colonne, choisissez une colonne dans la liste déroulante. Au moins cinq colonnes sont requises.
 - i. Choisissez une classification dans la liste déroulante. Cela permet d'identifier le type de données pour chaque colonne.

Les types de classification incluent :

- Numérique — Valeurs numériques continues telles que des mesures ou des dénombrements
 - Catégoriel — Valeurs ou catégories discrètes telles que des étiquettes ou des types
- ii. Pour supprimer une colonne, sélectionnez Supprimer.
 - iii. Pour ajouter une autre colonne, sélectionnez Ajouter une autre colonne. Choisissez la colonne et la classification dans les listes déroulantes.

- iv. Pour Valeur prédictive, choisissez une colonne dans la liste déroulante. Il s'agit de la colonne que le modèle personnalisé utilise pour la prédiction après avoir été entraîné sur le jeu de données synthétique.
- b. Les paramètres avancés vous permettent de définir le niveau de confidentialité et le seuil de confidentialité. Ajustez les paramètres en fonction de vos besoins.
 - Pour le niveau de confidentialité, entrez une valeur epsilon pour déterminer le niveau de bruit ajouté par le modèle synthétique afin de protéger la confidentialité de votre ensemble de données généré. La valeur doit être comprise entre 0,0001 et 10.
 - Des valeurs plus faibles ajoutent du bruit, ce qui renforce la protection de la confidentialité, mais réduit potentiellement l'utilité du modèle personnalisé en aval formé à partir de ces données.
 - Des valeurs supérieures ajoutent moins de bruit, ce qui renforce la précision tout en réduisant potentiellement la protection de la vie privée.

Pour le seuil de confidentialité, entrez la probabilité maximale autorisée qu'une attaque par inférence d'appartenance puisse identifier les membres du jeu de données d'origine. La valeur doit être comprise entre 50,0 et 100.

- Des scores de 50 % indiquent qu'une attaque par inférence d'appartenance ne permet pas de distinguer les membres des non-membres mieux qu'une estimation aléatoire.
- Pour aucune limite de confidentialité, entrez 100 %.

La configuration que vous choisissez dépend de votre cas d'utilisation spécifique et de vos exigences en matière de confidentialité. Si le seuil de confidentialité est dépassé, la création du canal d'entrée ML échoue et vous ne pouvez pas utiliser le jeu de données synthétique pour entraîner un modèle.

 Warning

La génération de données synthétiques empêche de déduire des attributs individuels, que des individus spécifiques soient présents dans l'ensemble de données d'origine ou que des attributs d'apprentissage de ces individus soient présents. Cependant, cela n'empêche pas les valeurs littérales de l'ensemble de données d'origine, y compris les

informations personnelles identifiables (PII), d'apparaître dans l'ensemble de données synthétique.

Nous recommandons d'éviter dans le jeu de données d'entrée les valeurs associées à une seule personne concernée, car elles peuvent permettre de réidentifier une personne concernée. Par exemple, si un seul utilisateur vit dans un code postal, la présence de ce code postal dans le jeu de données synthétique confirmera que cet utilisateur figurait dans le jeu de données d'origine. Des techniques telles que la troncation de valeurs de haute précision ou le remplacement de catalogues peu courants par d'autres peuvent être utilisées pour atténuer ce risque. Ces transformations peuvent faire partie de la requête utilisée pour créer le canal d'entrée ML.

11. Si vous souhaitez activer les balises pour la ressource, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
12. Choisissez Créer.
13. Vous êtes maintenant prêt à informer le membre de votre collaboration qu'il peut [réviser un modèle d'analyse](#). (Facultatif si vous souhaitez interroger vos propres données.)

Révision d'un modèle d'analyse SQL

Une fois qu'un membre de la collaboration a créé un SQLAnalysis modèle, vous pouvez le consulter et l'approuver. Une fois le modèle d'analyse approuvé, il peut être utilisé dans une requête dans AWS Clean Rooms.

Note

Lorsque vous intégrez votre code d'analyse à une collaboration, tenez compte des points suivants :

- AWS Clean Rooms ne valide ni ne garantit le comportement du code d'analyse.
 - Si vous devez vous assurer de certains comportements, consultez directement le code de votre partenaire de collaboration ou confiez-le à un auditeur tiers de confiance.
- Dans le modèle de sécurité partagé :
 - Vous (le client) êtes responsable de la sécurité du code exécuté dans l'environnement.
 - AWS Clean Rooms est responsable de la sécurité de l'environnement, en veillant à ce que
 - seul le code approuvé s'exécute

- seules les tables configurées spécifiées sont accessibles
- la seule destination de sortie est le compartiment S3 du récepteur des résultats.

Pour consulter un modèle d'analyse SQL à l'aide de la AWS Clean Rooms console

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Modèles, accédez à la section Modèles d'analyse créés par d'autres membres.
5. Choisissez le modèle d'analyse dont le statut Peut être exécuté est Non nécessite votre révision.
6. Choisissez Examiner.
7. Consultez la présentation, la définition et les paramètres des règles d'analyse (le cas échéant).
8. Passez en revue les tables configurées répertoriées sous Tables référencées dans la définition.

Le statut à côté de chaque table indiquera Modèle non autorisé.

9. Choisissez une table .

Si vous	Choisissez ensuite
Approuver le modèle d'analyse	Autoriser le modèle sur la table. Confirmez votre approbation en choisissant Autoriser.
Ne pas approuver le modèle d'analyse	Interdire

Vous êtes maintenant prêt à interroger la table configurée à l'aide d'un modèle d'analyse SQL. Pour de plus amples informations, veuillez consulter [Exécution de requêtes SQL](#).

PySpark modèles d'analyse

PySpark les modèles d'analyse nécessitent un script utilisateur Python et un environnement virtuel facultatif pour utiliser des bibliothèques personnalisées et open source. Ces fichiers sont appelés artefacts.

Avant de créer un modèle d'analyse, vous devez d'abord créer les artefacts, puis les stocker dans un compartiment Amazon S3. AWS Clean Rooms utilise ces artefacts lors de l'exécution de tâches d'analyse. AWS Clean Rooms accède aux artefacts uniquement lors de l'exécution d'une tâche.

Avant d'exécuter du code sur un modèle d' PySpark analyse, AWS Clean Rooms valide les artefacts en :

- Vérification de la version spécifique de l'objet S3 utilisée lors de la création du modèle
- Vérification du hachage SHA-256 de l'artefact
- Échec d'une tâche où des artefacts ont été modifiés ou supprimés

Note

La taille maximale de tous les artefacts combinés pour un modèle d' PySpark analyse donné AWS Clean Rooms est de 1 Go.

Sécurité pour les modèles PySpark d'analyse

Pour préserver un environnement informatique sécurisé, AWS Clean Rooms utilise une architecture informatique à deux niveaux pour isoler le code utilisateur des opérations du système. Cette architecture est basée sur la technologie de contrôle d'accès fine sans serveur Amazon EMR, également connue sous le nom de Membrane. Pour plus d'informations, voir [Membrane — Contrôles d'accès aux données sûrs et performants dans Apache Spark en présence de code impératif](#).

Les composants de l'environnement informatique sont divisés en un espace utilisateur et un espace système distincts. L'espace utilisateur exécute le PySpark code dans le modèle PySpark d'analyse. AWS Clean Rooms utilise l'espace système pour permettre l'exécution de la tâche, notamment en utilisant les rôles de service fournis par les clients pour lire les données nécessaires à l'exécution de la tâche et en implémentant la liste des colonnes autorisées. En raison de cette architecture, le PySpark code d'un client qui affecte l'espace système, qui peut inclure un petit nombre de Spark SQL PySpark DataFrames APIs, est bloqué.

PySpark limites dans AWS Clean Rooms

Lorsque les clients soumettent un modèle d' PySpark analyse approuvé, il l' AWS Clean Rooms exécute sur son propre environnement informatique sécurisé auquel aucun client ne peut accéder.

L'environnement informatique met en œuvre une architecture informatique avec un espace utilisateur et un espace système afin de préserver un environnement informatique sécurisé. Pour de plus amples informations, veuillez consulter [Sécurité pour les modèles PySpark d'analyse](#).

Tenez compte des limites suivantes avant de PySpark l'utiliser AWS Clean Rooms.

Limites

- Seules les DataFrame sorties sont prises en charge
- Une seule session Spark par exécution de tâche

Fonctions non prises en charge

- Gestion des données
 - Formats de tableaux Iceberg
 - LakeFormation tables gérées
 - Jeux de données distribués résilients (RDD)
 - Spark Streaming
 - Contrôle d'accès pour les colonnes imbriquées
- Fonctions et extensions personnalisées
 - Fonctions de table définies par l'utilisateur () UDTFs
 - Ruche UDFs
 - Classes personnalisées dans les fonctions définies par l'utilisateur
 - Sources de données personnalisées
 - Fichiers JAR supplémentaires pour :
 - Extensions Spark
 - Connecteurs
 - Configurations de métastore
- Surveillance et analyse
 - Journalisation Spark
 - Interface utilisateur Spark
 - Commandes de l'ANALYZE TABLE

Important

Ces limites sont en place pour maintenir l'isolation de sécurité entre les espaces utilisateur et système.

Toutes les restrictions s'appliquent quelle que soit la configuration de collaboration.

Les mises à jour futures peuvent ajouter la prise en charge de fonctionnalités supplémentaires en fonction des évaluations de sécurité.

Bonnes pratiques

Nous recommandons les meilleures pratiques suivantes lors de la création de modèles PySpark d'analyse.

- Concevez vos modèles d'analyse [PySpark limites dans AWS Clean Rooms](#) en gardant cela à l'esprit.
- Testez d'abord votre code dans un environnement de développement.
- Utilisez exclusivement les DataFrame opérations prises en charge.
- Planifiez votre structure de sortie pour qu'elle fonctionne avec DataFrame les limites.

Nous recommandons les meilleures pratiques suivantes pour gérer les artefacts

- Conservez tous les artefacts du modèle d' PySpark analyse dans un compartiment ou un préfixe S3 dédié.
- Utilisez un nom de version clair pour les différentes versions d'artefacts.
- Créez de nouveaux modèles d'analyse lorsque des mises à jour d'artefacts sont nécessaires.
- Tenez à jour un inventaire des modèles qui utilisent les différentes versions des artefacts.

Pour plus d'informations sur la façon d'écrire du code Spark, consultez ce qui suit :

- [Exemples d'Apache Spark](#)
- [Écrire une application Spark](#) dans le guide de mise à jour d'Amazon EMR
- [Tutoriel : Écrire un script AWS Glue pour Spark](#) dans le guide de AWS Glue l'utilisateur

Les rubriques suivantes expliquent comment créer des bibliothèques et des scripts utilisateur Python avant de créer et de réviser le modèle d'analyse.

Rubriques

- [Création d'un script utilisateur](#)
- [Utilisation des paramètres dans les modèles PySpark d'analyse](#)
- [Création d'un environnement virtuel \(facultatif\)](#)
- [Stockage d'un script utilisateur et d'un environnement virtuel dans S3](#)
- [Création d'un modèle PySpark d'analyse](#)
- [Révision d'un modèle PySpark d'analyse](#)

Création d'un script utilisateur

Le script utilisateur doit contenir une fonction de point d'entrée (en d'autres termes, un gestionnaire). Vous pouvez nommer votre fichier de script utilisateur avec n'importe quel nom de fichier Python valide.

La procédure suivante explique comment créer un script utilisateur pour définir les fonctionnalités principales de votre PySpark analyse.

Conditions préalables

- PySpark 1.0 (correspond à Python 3.11 et Spark 3.5.3)
- Les ensembles de données d'Amazon S3 ne peuvent être lus que sous forme d'associations de tables configurées dans la session Spark que vous définissez.
- Votre code ne peut pas appeler directement Amazon S3 et AWS Glue
- Votre code ne peut pas passer d'appels réseau

Pour créer un script utilisateur

1. Ouvrez un éditeur de texte ou un environnement de développement intégré (IDE) de votre choix.

Vous pouvez utiliser n'importe quel éditeur de texte ou IDE (tel que Visual Studio Code ou Notepad++) qui prend en charge les fichiers Python. PyCharm

2. Créez un nouveau fichier Python avec le nom de votre choix (par exemple, `my_analysis.py`).

3. Définissez une fonction de point d'entrée qui accepte un paramètre d'objet contextuel.

```
def entrypoint(context)
```

Le paramètre `context` object est un dictionnaire qui donne accès aux composants essentiels de Spark, aux tables référencées et aux paramètres d'analyse. Il contient :

Accès à la session Spark via `context['sparkSession']`

Tables référencées via `context['referencedTables']`

Paramètres d'analyse via `context['analysisParameters']` (si les paramètres sont définis dans le modèle)

4. Définissez les résultats de la fonction `entrypoint` :

```
return results
```

Le `results` doit renvoyer un objet contenant un dictionnaire de résultats de noms de fichiers vers une sortie. `DataFrame`

Note

AWS Clean Rooms écrit automatiquement les `DataFrame` objets dans le compartiment S3 du récepteur des résultats.

5. Vous êtes maintenant prêt à :
 - a. Stockez ce script utilisateur dans S3. Pour de plus amples informations, veuillez consulter [Stockage d'un script utilisateur et d'un environnement virtuel dans S3](#).
 - b. Créez l'environnement virtuel facultatif pour prendre en charge les bibliothèques supplémentaires requises par votre script utilisateur. Pour de plus amples informations, veuillez consulter [Création d'un environnement virtuel \(facultatif\)](#).

Exemple Exemple 1

<caption>The following example demonstrates a generic user script for a PySpark analysis template.</caption>

```
# File name: my_analysis.py
```

```
def entrypoint(context):
    try:
        # Access Spark session
        spark = context['sparkSession']

        # Access input tables
        input_table1 = context['referencedTables']['table1_name']
        input_table2 = context['referencedTables']['table2_name']

        # Example data processing operations
        output_df1 = input_table1.select("column1", "column2")
        output_df2 = input_table2.join(input_table1, "join_key")
        output_df3 = input_table1.groupBy("category").count()

        # Return results - each key creates a separate output folder
        return {
            "results": {
                "output1": output_df1,          # Creates output1/ folder
                "output2": output_df2,          # Creates output2/ folder
                "analysis_summary": output_df3 # Creates analysis_summary/ folder
            }
        }

    except Exception as e:
        print(f"Error in main function: {str(e)}")
        raise e
```

La structure de dossiers de cet exemple est la suivante :

```
analysis_results/
#
### output1/ # Basic selected columns
# ### part-00000.parquet
# ### _SUCCESS
#
### output2/ # Joined data
# ### part-00000.parquet
# ### _SUCCESS
#
### analysis_summary/ # Aggregated results
### part-00000.parquet
### _SUCCESS
```


Example Exemple 2

<caption>The following example demonstrates a more complex user script for a PySpark analysis template.</caption>

```
def entrypoint(context):
    try:
        # Get DataFrames from context
        emp_df = context['referencedTables']['employees']
        dept_df = context['referencedTables']['departments']

        # Apply Transformations
        emp_dept_df = emp_df.join(
            dept_df,
            on="dept_id",
            how="left"
        ).select(
            "emp_id",
            "name",
            "salary",
            "dept_name"
        )

        # Return Dataframes
        return {
            "results": {
                "outputTable": emp_dept_df
            }
        }

    except Exception as e:
        print(f"Error in entrypoint function: {str(e)}")
        raise e
```

Utilisation des paramètres dans les modèles PySpark d'analyse

Les paramètres augmentent la flexibilité de vos modèles d' PySpark analyse en permettant de fournir différentes valeurs au moment de la soumission des tâches. Les paramètres sont accessibles via l'objet de contexte transmis à votre fonction de point d'entrée.

 Note

Les paramètres sont des chaînes fournies par l'utilisateur qui peuvent contenir du contenu arbitraire.

- Passez en revue le code pour vous assurer que les paramètres sont gérés en toute sécurité afin d'éviter tout comportement inattendu dans votre analyse.
- Concevez la gestion des paramètres de manière à ce qu'elle fonctionne en toute sécurité, quelles que soient les valeurs des paramètres fournies au moment de la soumission.

Accès aux paramètres

Les paramètres sont disponibles dans le `context['analysisParameters']` dictionnaire. Toutes les valeurs des paramètres sont des chaînes de caractères.

Exemple Accès aux paramètres en toute sécurité

```
def entrypoint(context):
    # Access parameters from context
    parameters = context['analysisParameters']
    threshold = parameters['threshold']
    table_name = parameters['table_name']

    # Continue with analysis using parameters
    spark = context['sparkSession']
    input_df = context['referencedTables'][table_name]

    # Convert threshold value
    threshold_val = int(threshold)

    # Use parameter in DataFrame operation
    filtered_df = input_df.filter(input_df.amount > threshold_val)

    return {
        "results": {
            "output": filtered_df
        }
    }
```

Bonnes pratiques de sécurité pour les paramètres

Warning

Les paramètres sont des chaînes fournies par l'utilisateur qui peuvent contenir du contenu arbitraire. Vous devez gérer les paramètres en toute sécurité pour éviter les failles de sécurité dans votre code d'analyse.

Modèles de gestion des paramètres non sécurisés à éviter :

- Exécution de paramètres sous forme de code — Ne jamais utiliser `eval()` ou `exec()` sur des valeurs de paramètres

```
# UNSAFE - Don't do this
eval(parameters['expression']) # Can execute arbitrary code
```

- Interpolation de chaînes SQL — Ne concaténez jamais de paramètres directement dans des chaînes SQL

```
# UNSAFE - Don't do this
sql = f"SELECT * FROM table WHERE column = '{parameters['value']}'" # SQL injection risk
```

- Opérations de chemin de fichier non sécurisées : n'utilisez jamais de paramètres directement dans les opérations du système de fichiers sans validation

```
# UNSAFE - Don't do this
file_path = f"/data/{parameters['filename']}" # Path traversal risk
```

Modèles de gestion des paramètres sécurisés :

- Utiliser les paramètres dans les DataFrame opérations — Spark DataFrames gère les valeurs des paramètres en toute sécurité

```
# SAFE - Use parameters in DataFrame operations
threshold = int(parameters['threshold'])
filtered_df = input_df.filter(input_df.value > threshold)
```

- Valider les valeurs des paramètres — Vérifiez que les paramètres répondent aux formats attendus avant utilisation

```
# SAFE - Validate parameters before use
def validate_date(date_str):
    try:
        from datetime import datetime
        datetime.strptime(date_str, '%Y-%m-%d')
        return True
    except ValueError:
        return False

date_param = parameters['date_filter'] or '2024-01-01'
if not validate_date(date_param):
    raise ValueError(f"Invalid date format: {date_param}")
```

- Utiliser des listes d'autorisation pour les valeurs des paramètres : dans la mesure du possible, validez les paramètres par rapport aux bonnes valeurs connues

```
# SAFE - Use allowlists
allowed_columns = ['column1', 'column2', 'column3']
column_param = parameters['column_name']
if column_param not in allowed_columns:
    raise ValueError(f"Invalid column: {column_param}")
```

- Conversion de type avec gestion des erreurs — Convertissez les paramètres de chaîne en types attendus en toute sécurité

```
# SAFE - Convert with error handling
try:
    batch_size = int(parameters['batch_size'] or '1000')
    if batch_size <= 0 or batch_size > 10000:
        raise ValueError(f"Batch size must be between 1 and 10000")
except ValueError as e:
    print(f"Invalid parameter: {e}")
    raise
```

⚠ Important

N'oubliez pas que les paramètres ignorent la révision du code lorsque les exécuteurs de tâches fournissent des valeurs différentes. Concevez votre gestion des paramètres pour qu'elle fonctionne en toute sécurité, quelles que soient les valeurs de paramètres fournies.

Exemple de paramètre complet

Exemple Utilisation sécurisée des paramètres dans un PySpark script

```
def entrypoint(context):
    try:
        # Access Spark session and tables
        spark = context['sparkSession']
        input_table = context['referencedTables']['sales_data']

        # Access parameters - fail fast if analysisParameters missing
        parameters = context['analysisParameters']

        # Validate and convert numeric parameter (handles empty strings with default)
        try:
            threshold = int(parameters['threshold'] or '100')
            if threshold <= 0:
                raise ValueError("Threshold must be positive")
        except (ValueError, TypeError) as e:
            print(f"Invalid threshold parameter: {e}")
            raise

        # Validate date parameter (handles empty strings with default)
        date_filter = parameters['start_date'] or '2024-01-01'
        from datetime import datetime
        try:
            datetime.strptime(date_filter, '%Y-%m-%d')
        except ValueError:
            raise ValueError(f"Invalid date format: {date_filter}")

        # Use parameters safely in DataFrame operations
        filtered_df = input_table.filter(
            (input_table.amount > threshold) &
            (input_table.date >= date_filter)
        )
```

```
result_df = filtered_df.groupBy("category").agg(
    {"amount": "sum"}
)

return {
    "results": {
        "filtered_results": result_df
    }
}

except Exception as e:
    print(f"Error in analysis: {str(e)}")
    raise
```

Création d'un environnement virtuel (facultatif)

Si des bibliothèques supplémentaires sont requises par votre script utilisateur, vous avez la possibilité de créer un environnement virtuel pour stocker ces bibliothèques. Si vous n'avez pas besoin de bibliothèques supplémentaires, vous pouvez ignorer cette étape.

Lorsque vous travaillez avec des bibliothèques dotées d'extensions natives, gardez à l'esprit qu'elles AWS Clean Rooms fonctionnent PySpark sous Linux avec une ARM64 architecture.

La procédure suivante explique comment créer un environnement virtuel à l'aide d'une commande CLI de base.

Pour créer un environnement virtuel

1. Ouvrez un terminal ou une invite de commande.
2. Ajoutez le contenu suivant :

```
# create and activate a python virtual environment
python3 -m venv pyspark_venvsource
source pyspark_venvsource/bin/activate

# install the python packages
pip3 install pandas # add packages here

# package the virtual environment into an archive
pip3 install venv-pack
venv-pack -f -o pyspark_venv.tar.gz
```

```
# optionally, remove the virtual environment directory
deactivate
rm -fr pyspark_venvsource
```

3. Vous êtes maintenant prêt à stocker cet environnement virtuel dans S3. Pour de plus amples informations, veuillez consulter [Stockage d'un script utilisateur et d'un environnement virtuel dans S3](#).

Pour plus d'informations sur l'utilisation de Docker et d'Amazon ECR, consultez le guide [Amazon ECR User](#).

Stockage d'un script utilisateur et d'un environnement virtuel dans S3

La procédure suivante explique comment stocker un script utilisateur et un environnement virtuel facultatif dans Amazon S3. Effectuez cette étape avant de créer un modèle d'PySpark analyse.

Important

Ne modifiez ni ne supprimez d'artefacts (scripts utilisateur ou environnements virtuels) après avoir créé un modèle d'analyse.

Cela permettra de :

- Faire échouer toutes les futures tâches d'analyse utilisant ce modèle.
- Exiger la création d'un nouveau modèle d'analyse avec de nouveaux artefacts.
- Pas d'incidence sur les tâches d'analyse effectuées précédemment

Conditions préalables

- Et Compte AWS avec les autorisations appropriées
- Un fichier de script utilisateur (tel `quemy_analysis.py`)
- (Facultatif, s'il existe) Un package d'environnement virtuel (`.tar.gz` fichier)
- Accès pour créer ou modifier des rôles IAM

Console

Pour stocker un script utilisateur et un environnement virtuel dans S3 à l'aide de la console :

1. Connectez-vous à la console Amazon S3 AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/s3/>.
2. Créez un nouveau compartiment S3 ou utilisez-en un existant.
3. Activez la gestion des versions pour le compartiment.
 - a. Sélectionnez votre compartiment.
 - b. Choisissez Propriétés.
 - c. Dans la section Gestion des versions des compartiments, choisissez Modifier.
 - d. Sélectionnez Activer et enregistrez les modifications.
4. Téléchargez vos artefacts et activez le hachage SHA-256.
 - a. Accédez à votre compartiment.
 - b. Choisissez Charger.
 - c. Choisissez Ajouter des fichiers et ajoutez votre fichier de script utilisateur.
 - d. (Facultatif, s'il en existe un) Ajoutez votre fichier .tar.gz.
 - e. Développez les propriétés.
 - f. Sous Checksums, pour la fonction Checksum, sélectionnez. SHA256
 - g. Choisissez Charger.
5. Vous êtes maintenant prêt à créer un modèle d' PySpark analyse.

CLI

Pour stocker le script utilisateur et l'environnement virtuel dans S3 à l'aide de AWS CLI :

1. Exécutez la commande suivante :

```
aws s3 cp --checksum-algorithm sha256 pyspark_venv.tar.gz s3://ARTIFACT-BUCKET/  
EXAMPLE-PREFIX/
```

2. Vous êtes maintenant prêt à créer un modèle d' PySpark analyse.

 Note

Si vous devez mettre à jour le script ou l'environnement virtuel :

1. Téléchargez la nouvelle version en tant qu'objet distinct.
2. Créez un nouveau modèle d'analyse à l'aide des nouveaux artefacts.
3. Dépréciez l'ancien modèle.
4. Conservez les artefacts d'origine dans S3 si l'ancien modèle est toujours nécessaire.

Création d'un modèle PySpark d'analyse

 Note

Les paramètres sont des chaînes fournies par l'utilisateur qui peuvent contenir du contenu arbitraire.

- Passez en revue le code pour vous assurer que les paramètres sont gérés en toute sécurité afin d'éviter tout comportement inattendu dans votre analyse.
- Concevez la gestion des paramètres de manière à ce qu'elle fonctionne en toute sécurité, quelles que soient les valeurs des paramètres fournies au moment de la soumission.

Conditions préalables

Avant de créer un modèle d' PySpark analyse, vous devez disposer des éléments suivants :

- L'adhésion à une AWS Clean Rooms collaboration active
- Accès à au moins une table configurée dans la collaboration active
- Autorisations pour créer des modèles d'analyse
- Un script utilisateur Python et un environnement virtuel créés et stockés dans S3
 - La gestion des versions est activée dans le compartiment S3. Pour plus d'informations, voir [Utilisation de la gestion des versions dans les compartiments S3](#)
 - Le compartiment S3 peut calculer les sommes de contrôle SHA-256 pour les artefacts téléchargés. Pour plus d'informations, voir [Utilisation des checksums](#)
- Autorisations pour lire le code d'un compartiment S3

Pour plus d'informations sur la création du rôle de service requis, consultez [Création d'un rôle de service pour lire le code d'un compartiment S3 \(rôle de modèle d'PySpark analyse\)](#).

La procédure suivante décrit le processus de création d'un modèle d' PySpark analyse à l'aide de la [AWS Clean Rooms console](#). Cela suppose que vous avez déjà créé un script utilisateur et des fichiers d'environnement virtuel et que vous avez stocké votre script utilisateur et vos fichiers d'environnement virtuel dans un compartiment Amazon S3.

 Note

Le membre qui crée le modèle PySpark d'analyse doit également être celui qui reçoit les résultats.

Pour plus d'informations sur la création d'un modèle d' PySpark analyse à l'aide du AWS SDKs, consultez la [référence de l'AWS Clean Rooms API](#).

Pour créer un modèle PySpark d'analyse

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Modèles, accédez à la section Modèles d'analyse que vous avez créés.
5. Choisissez Créer un modèle d'analyse.
6. Sur la page Créer un modèle d'analyse, pour plus de détails,
 - a. Entrez un nom pour le modèle d'analyse.
 - b. (Facultatif) Entrez une description.
 - c. Pour Format, choisissez l'PySparkoption.
7. Pour la définition,
 - a. Passez en revue les prérequis et assurez-vous que chaque condition préalable est remplie avant de continuer.
 - b. Pour le fichier du point d'entrée, entrez dans le compartiment S3 ou choisissez Browse S3.

- c. (Facultatif) Pour le fichier Libraries, entrez dans le compartiment S3 ou choisissez Browse S3.
8. Pour les paramètres : facultatif, si vous souhaitez ajouter des paramètres afin de rendre votre modèle d'analyse réutilisable :
- a. Choisissez Ajouter un paramètre.
 - b. Entrez un nom de paramètre.

Les noms des paramètres doivent commencer par une lettre ou un trait de soulignement, suivi de caractères alphanumériques ou de traits de soulignement.

- c. Pour Type, STRING est automatiquement sélectionné comme seul type pris en charge pour les modèles PySpark d'analyse.
- d. (Facultatif) Entrez une valeur par défaut pour le paramètre.

Si vous fournissez une valeur par défaut, les exécuteurs de tâches peuvent utiliser cette valeur lorsqu'ils exécutent des tâches sans fournir explicitement de valeur de paramètre.

- e. Pour ajouter d'autres paramètres, choisissez Ajouter un autre paramètre et répétez les étapes précédentes.

 Note

Vous pouvez définir jusqu'à 50 paramètres par modèle PySpark d'analyse. Chaque valeur de paramètre peut comporter jusqu'à 1 000 caractères.

9. Pour les tables référencées dans la définition,
- Si toutes les tables référencées dans la définition ont été associées à la collaboration :
 - Laissez la case Toutes les tables référencées dans la définition ont été associées à la collaboration cochée.
 - Sous Tables associées à la collaboration, choisissez toutes les tables associées référencées dans la définition.
 - Si toutes les tables référencées dans la définition n'ont pas été associées à la collaboration :
 - Décochez la case Toutes les tables référencées dans la définition ont été associées à la collaboration.

- Sous Tables associées à la collaboration, choisissez toutes les tables associées référencées dans la définition.
- Sous Tables qui seront associées ultérieurement, entrez un nom de table.
- Choisissez Répertoire une autre table pour répertorier une autre table.

10. Pour la configuration des messages d'erreur, choisissez l'une des options suivantes :

- Messages d'erreur de base : renvoie les messages d'erreur de base sans exposer les données sous-jacentes. Recommandé pour les charges de travail de production.
- Messages d'erreur détaillés : renvoie des messages d'erreur détaillés pour accélérer le dépannage. Recommandé dans les environnements de développement et de test. Peut exposer des données sensibles, y compris des informations personnelles identifiables (PII).

 Note

Lorsque vous utilisez des messages d'erreur détaillés, tous les membres du fournisseur de données doivent approuver ce paramètre pour le modèle.

11. Spécifiez les autorisations d'accès au service en sélectionnant un nom de rôle de service existant dans la liste déroulante.

1. La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles.

Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser.

2. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM.

S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible.

Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

 Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).
- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.
- Si vous ne pouvez pas modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique du rôle de service est introuvable.

12. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
13. Choisissez Créer.
14. Vous êtes maintenant prêt à informer le membre de votre collaboration qu'il peut [réviser un modèle d'analyse](#). (Facultatif si vous souhaitez interroger vos propres données.)

 Important

Ne modifiez ni ne supprimez d'artefacts (scripts utilisateur ou environnements virtuels) après avoir créé un modèle d'analyse.

Cela permettra de :

- Faire échouer toutes les futures tâches d'analyse utilisant ce modèle.
- Exiger la création d'un nouveau modèle d'analyse avec de nouveaux artefacts.
- Cela n'affecte pas les tâches d'analyse effectuées précédemment.

Révision d'un modèle PySpark d'analyse

Lorsqu'un autre membre crée un modèle d'analyse dans votre collaboration, vous devez le consulter et l'approuver avant de pouvoir l'utiliser.

La procédure suivante explique comment examiner un modèle d' PySpark analyse, notamment ses règles, ses paramètres et ses tables référencées. En tant que membre de la collaboration, vous évalueriez si le modèle est conforme à vos accords de partage de données et à vos exigences de sécurité.

Une fois le modèle d'analyse approuvé, il peut être utilisé dans une tâche dans AWS Clean Rooms.

 Note

Lorsque vous intégrez votre code d'analyse à une collaboration, tenez compte des points suivants :

- AWS Clean Rooms ne valide ni ne garantit le comportement du code d'analyse.
 - Si vous devez vous assurer de certains comportements, consultez directement le code de votre partenaire de collaboration ou confiez-le à un auditeur tiers de confiance.
- AWS Clean Rooms garantit que les hachages SHA-256 du code répertorié dans le modèle d' PySparkanalyse correspondent au code exécuté dans l' PySpark environnement d'analyse.
- AWS Clean Rooms n'effectue aucun audit ni aucune analyse de sécurité des bibliothèques supplémentaires que vous importez dans l'environnement.
- Dans le modèle de sécurité partagé :
 - Vous (le client) êtes responsable de la sécurité du code exécuté dans l'environnement.
 - Vous (le client) êtes responsable de définir la configuration des messages d'erreur appropriée pour l'environnement.
 - AWS Clean Rooms est responsable de la sécurité de l'environnement, en veillant à ce que
 - seul le code approuvé s'exécute
 - seules les tables configurées spécifiées sont accessibles
 - la seule destination de sortie est le compartiment S3 du récepteur des résultats.

AWS Clean Rooms génère des hachages SHA-256 du script utilisateur et de l'environnement virtuel pour votre examen. Cependant, le script utilisateur et les bibliothèques eux-mêmes ne sont pas directement accessibles depuis AWS Clean Rooms.

Pour vérifier que le script utilisateur et les bibliothèques partagés sont identiques à ceux référencés dans le modèle d'analyse, vous pouvez créer un hachage SHA-256 des fichiers partagés et le comparer au hachage du modèle d'analyse créé par AWS Clean Rooms. Les hachages du code exécuté figureront également dans les journaux des tâches.

Conditions préalables

- Système d'exploitation Linux/Unix ou sous-système Windows pour Linux (WSL)
- Fichier de script utilisateur que vous souhaitez hacher
 - Demandez au créateur du modèle d'analyse de partager le fichier via un canal sécurisé.
- Le hachage du modèle d'analyse créé par AWS Clean Rooms

Pour consulter un modèle d'analyse PySpark à l'aide de la AWS Clean Rooms console

1. Connectez-vous à la console AWS Management Console et ouvrez-la avec la [AWS Clean Rooms console](#) Compte AWS qui fonctionnera en tant que créateur de collaboration.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration.
4. Dans l'onglet Modèles, accédez à la section Modèles d'analyse créés par d'autres membres.
5. Choisissez le modèle d'analyse dont le statut Peut être exécuté est Non nécessite votre révision.
6. Choisissez Examiner.
7. Consultez la présentation, la définition et les paramètres des règles d'analyse (le cas échéant).

Note

Les paramètres permettent aux analyseurs de soumettre différentes valeurs au moment de la soumission. Si un modèle d'analyse prend en charge les paramètres, vérifiez comment les valeurs des paramètres sont utilisées dans le code de votre partenaire de collaboration pour vous assurer qu'il répond à vos exigences.

8. Vérifiez que le script utilisateur partagé et les bibliothèques sont identiques à ceux référencés dans le modèle d'analyse.
 - a. Créez un hachage SHA-256 des fichiers partagés et comparez-le au hachage du modèle d'analyse créé par AWS Clean Rooms

Vous pouvez générer un hachage en accédant au répertoire contenant votre fichier de script utilisateur, puis en exécutant la commande suivante :

```
sha256sum your_script_filename.py
```

Exemple de sortie :

```
e3b0c44298fc1c149afb4c8996fb92427ae41e4649b934ca495991b7852b855 my_analysis.py
```

- b. Vous pouvez également utiliser les fonctionnalités de somme de contrôle d'Amazon S3. Pour plus d'informations, consultez la section [Vérification de l'intégrité des objets dans Amazon S3](#) dans le guide de l'utilisateur Amazon S3.
 - c. Une autre solution consiste à afficher les hachages du code exécuté dans les journaux des tâches.
9. Passez en revue les tables configurées répertoriées sous Tables référencées dans la définition.
- Le statut à côté de chaque table indiquera Modèle non autorisé.
10. Choisissez une table .
- a. Pour approuver le modèle d'analyse, choisissez Autoriser le modèle sur la table. Confirmez votre approbation en choisissant Autoriser.
 - b. Pour refuser l'approbation, choisissez Disallow.

Si vous avez choisi d'approuver le modèle d'analyse, le membre autorisé à exécuter des tâches peut désormais exécuter une PySpark tâche sur une table configurée à l'aide d'un modèle d' PySpark analyse. Pour de plus amples informations, veuillez consulter [Exécution de PySpark tâches](#).

Modèles d' PySpark analyse de résolution des problèmes

Lorsque vous exécutez des tâches à l'aide de modèles d' PySpark analyse, vous pouvez rencontrer des échecs lors de l'initialisation ou de l'exécution des tâches. Ces échecs sont généralement liés à la configuration des scripts, aux autorisations d'accès aux données ou à la configuration de l'environnement.

Pour plus d'informations sur PySpark les limitations, consultez [PySpark limites dans AWS Clean Rooms](#).

Rubriques

- [Résolution des problèmes liés à votre code](#)
- [La tâche du modèle d'analyse ne démarre pas](#)
- [La tâche du modèle d'analyse démarre mais échoue pendant le traitement](#)
- [La configuration de l'environnement virtuel échoue](#)

Résolution des problèmes liés à votre code

Pour vous aider à développer votre code et à le résoudre, nous vous suggérons de le simuler vous-même AWS Clean Rooms en Compte AWS activant les messages d'erreur détaillés et en exécutant des tâches à l'aide de vos propres données de test.

Vous pouvez également effectuer une simulation PySpark AWS Clean Rooms dans Amazon EMR Serverless en suivant les étapes suivantes. Il présente de petites différences par rapport PySpark à AWS Clean Rooms, mais couvre principalement la manière dont votre code peut être exécuté.

Pour simuler PySpark AWS Clean Rooms dans EMR Serverless

1. Créez un ensemble de données dans Amazon S3, cataloguez-le dans le AWS Glue Data Catalog et configurez les autorisations de Lake Formation.
2. Enregistrez l'emplacement S3 auprès de Lake Formation à l'aide d'un rôle personnalisé.
3. Créez une instance Amazon EMR Studio si vous n'en avez pas déjà une (Amazon EMR Studio est nécessaire pour utiliser Amazon EMR Serverless).
4. Création d'une application EMR sans serveur
 - Sélectionnez la version de publication emr-7.7.0.
 - Sélectionnez ARM64 l'architecture.
 - Choisissez Utiliser les paramètres personnalisés.
 - Désactivez la capacité préinitialisée.
 - Si vous prévoyez d'effectuer un travail interactif, sélectionnez Point de terminaison interactif > Activer le point de terminaison pour le studio EMR.
 - Sélectionnez Configurations supplémentaires > Utiliser Lake Formation pour un contrôle d'accès précis.
 - Créez l'application.

5. Utilisez EMR-S via les blocs-notes EMR-Studio ou l'API. StartJobRun

La tâche du modèle d'analyse ne démarre pas

Causes courantes

Les tâches du modèle d'analyse peuvent échouer immédiatement au démarrage en raison de trois problèmes de configuration principaux :

- Nom de script incorrect ne correspondant pas au format requis
- Fonction de point d'entrée manquante ou mal formatée dans le script utilisateur
- Version de Python incompatible dans l'environnement virtuel

Résolution

Pour résoudre le problème :

1. Vérifiez votre script utilisateur :

- Vérifiez que le nom de fichier Python de votre script utilisateur est valide.

Les noms de fichiers Python valides utilisent des lettres minuscules, des traits de soulignement pour séparer les mots et l'extension .py.

2. Vérifiez la fonction du point d'entrée. Si votre script utilisateur ne possède pas de fonction de point d'entrée, ajoutez-en une.

- Ouvrez votre script utilisateur.
- Ajoutez cette fonction de point d'entrée :

```
def entrypoint(context):  
    # Your analysis code here
```

- Assurez-vous que le nom de la fonction est orthographié exactement comme `entrypoint`.
- Vérifiez que la fonction accepte le `context` paramètre.

3. Vérifiez la compatibilité des versions de Python :

- Vérifiez que votre environnement virtuel utilise Python 3.9 ou 3.11.
- Pour vérifier votre version, exécutez : `python --version`

- c. Si nécessaire, mettez à jour votre environnement virtuel :

```
conda create -n analysis-env python=3.9
conda activate analysis-env
```

Prévention

- Utilisez le code de démarrage du modèle d'analyse fourni qui inclut la structure de fichier appropriée.
- Configurez un environnement virtuel dédié avec Python 3.9 ou 3.11 pour tous les modèles d'analyse.
- Testez votre modèle d'analyse localement à l'aide de l'outil de validation du modèle avant de soumettre des tâches.
- Mettez en œuvre CI/CD des contrôles pour vérifier la dénomination des scripts et les exigences relatives aux fonctions de point d'entrée.

La tâche du modèle d'analyse démarre mais échoue pendant le traitement

Causes courantes

Les tâches d'analyse peuvent échouer en cours d'exécution pour les raisons de sécurité et de formatage suivantes :

- Tentatives d'accès direct non autorisées à AWS des services tels qu'Amazon S3 ou AWS Glue
- Renvoyer une sortie dans des formats incorrects qui ne correspondent pas aux DataFrame spécifications requises
- Appels réseau bloqués en raison de restrictions de sécurité dans l'environnement d'exécution

Résolution

Pour résoudre

1. Supprimer l'accès direct au AWS service :
 - a. Recherchez votre code pour les importations AWS de services et les appels directs.
 - b. Remplacez l'accès direct au S3 par les méthodes de session Spark fournies.

- c. Utilisez uniquement des tables préconfigurées via l'interface de collaboration.
2. Formatez correctement les sorties :
 - a. Vérifiez que toutes les sorties sont Spark DataFrames.
 - b. Mettez à jour votre relevé de retour pour qu'il corresponde au format suivant :

```
return {  
    "results": {  
        "output1": dataframe1  
    }  
}
```

- c. Supprimez tous les objets non DataFrame retournables.
3. Supprimer les appels réseau :
 - a. Identifiez et supprimez tous les appels d'API externes.
 - b. Supprimez toutes les URLLib, les requêtes ou les bibliothèques réseau similaires.
 - c. Supprimez toutes les connexions socket ou le code client HTTP.

Prévention

- Utilisez le linter de code fourni pour vérifier la présence d' AWS importations et d'appels réseau non autorisés.
- Tester des tâches dans un environnement de développement où les restrictions de sécurité correspondent à celles de la production.
- Suivez le processus de validation du schéma de sortie avant de déployer des tâches.
- Consultez les consignes de sécurité pour connaître les modèles d'accès aux services approuvés.

La configuration de l'environnement virtuel échoue

Causes courantes

Les défaillances de configuration de l'environnement virtuel se produisent généralement pour les raisons suivantes :

- Architecture de processeur inadaptée entre les environnements de développement et d'exécution
- Problèmes de formatage du code Python empêchant l'initialisation correcte de l'environnement

- Configuration d'image de base incorrecte dans les paramètres du conteneur

Résolution

Pour résoudre

1. Configurez l'architecture appropriée :

- a. Vérifiez votre architecture actuelle avec `uname -m`.
- b. Mettez à jour votre Dockerfile pour spécifier : ARM64

```
FROM --platform=linux/arm64 public.ecr.aws/amazonlinux/amazonlinux:2023-minimal
```

- c. Reconstituez votre conteneur avec `docker build --platform=linux/arm64`.

2. Corriger l'indentation en Python :

- a. Exécutez un formateur de code Python comme `black` sur vos fichiers de code.
- b. Vérifiez l'utilisation cohérente des espaces ou des onglets (pas les deux).
- c. Vérifiez l'indentation de tous les blocs de code :

```
def my_function():  
    if condition:  
        do_something()  
    return result
```

- d. Utilisez un IDE avec mise en évidence des indentations en Python.

3. Validez la configuration de l'environnement :

- a. Exécutez `python -m py_compile your_script.py` pour vérifier les erreurs de syntaxe.
- b. Testez l'environnement localement avant le déploiement.
- c. Vérifiez que toutes les dépendances sont répertoriées dans `requirements.txt`.

Prévention

- Utilisez Visual Studio Code ou des plugins PyCharm de formatage Python
- Configurer des hooks de pré-validation pour exécuter automatiquement les formateurs de code

- Créez et testez des environnements localement à l'aide de l'image ARM64 de base fournie
- Implémentez la vérification automatique du style de code dans votre CI/CD pipeline

Analyser les données dans le cadre d'une collaboration

Dans AWS Clean Rooms, vous pouvez analyser les données en exécutant des requêtes ou des tâches.

Une requête est une méthode permettant d'accéder aux tables configurées et de les analyser dans le cadre d'une collaboration, à l'aide d'un ensemble de fonctions, de classes et de variables pris en charge. Le langage de requête actuellement pris en charge dans SQL AWS Clean Rooms est le langage SQL. Il existe trois méthodes pour exécuter une requête AWS Clean Rooms : écrire du code SQL, utiliser un modèle d'analyse SQL approuvé ou utiliser l'interface utilisateur du générateur d'analyse.

Une tâche est une méthode permettant d'accéder aux tables configurées et de les analyser dans le cadre d'une collaboration à l'aide d'un ensemble pris en charge de fonctions, de classes et de variables. Le type d'emploi actuellement pris en charge dans AWS Clean Rooms est PySpark. Il existe une méthode pour exécuter une tâche AWS Clean Rooms : en utilisant un modèle d' PySpark analyse approuvé.

Les rubriques suivantes décrivent comment analyser les données en AWS Clean Rooms exécutant des requêtes ou des PySpark tâches SQL.

Rubriques

- [Exécution de requêtes SQL](#)
- [Exécution de PySpark tâches](#)

Exécution de requêtes SQL

Note

Vous ne pouvez exécuter des requêtes que si le membre chargé de payer les coûts de calcul des requêtes a rejoint la collaboration en tant que membre actif.

En tant que [membre habilité à effectuer une requête](#), vous pouvez exécuter une requête SQL en :

- Création manuelle d'une requête SQL à l'aide de l'éditeur de code SQL.

- À l'aide d'un [modèle d'analyse](#) SQL approuvé.
- Utilisation de l'interface utilisateur du générateur d'analyse pour créer une requête sans avoir à écrire de code SQL.

Lorsque le membre habilité à effectuer une requête exécute une requête SQL sur les tables de la collaboration, il AWS Clean Rooms assume les rôles appropriés pour accéder aux tables en son nom. AWS Clean Rooms applique les règles d'analyse nécessaires à la requête d'entrée et à sa sortie.

Les règles d'analyse et les contraintes de sortie sont appliquées automatiquement. AWS Clean Rooms renvoie uniquement les résultats conformes aux règles d'analyse définies.

AWS Clean Rooms prend en charge les requêtes SQL qui peuvent être différentes des autres moteurs de requêtes. Pour les spécifications, consultez la [référence AWS Clean Rooms SQL](#). Si vous souhaitez exécuter des requêtes sur des tables de données protégées par une confidentialité différentielle, vous devez vous assurer que vos requêtes sont compatibles avec la [structure de requête à usage général](#) de AWS Clean Rooms Differential Privacy.

Note

Lorsque vous utilisez le [calcul cryptographique pour Clean Rooms](#), toutes les opérations SQL ne génèrent pas de résultats valides. Par exemple, vous pouvez effectuer un COUNT sur une colonne cryptée, mais effectuer un SUM sur des numéros chiffrés entraîne des erreurs. En outre, les requêtes peuvent également donner des résultats incorrects. Par exemple, les requêtes dont les colonnes sont SUM scellées produisent des erreurs. Cependant, une GROUP BY requête sur des colonnes scellées semble réussir mais produit des groupes différents de ceux produits par une GROUP BY requête sur du texte clair.

Le [membre qui paie les coûts de calcul des requêtes](#) est facturé pour les requêtes exécutées dans le cadre de la collaboration.

Le membre autorisé à effectuer une requête peut sélectionner plusieurs [membres susceptibles de recevoir des résultats](#) pour recevoir les résultats d'une seule requête. Pour de plus amples informations, veuillez consulter [Interrogation de tables configurées à l'aide de l'éditeur de code SQL](#). Pour obtenir des informations générales sur la réception des résultats de requêtes, consultez [Réception et utilisation des résultats d'analyse](#).

Conditions préalables

Avant d'exécuter une requête SQL, assurez-vous que vous disposez des éléments suivants :

- Une adhésion active à la AWS Clean Rooms collaboration
- Accès à au moins une table configurée dans la collaboration
- Confirmation que le membre responsable des coûts de calcul des requêtes est un membre actif de la collaboration

Pour plus d'informations sur la façon d'interroger des données ou d'afficher des requêtes en appelant directement l'opération d' AWS Clean Rooms [StartProtectedQuery API](#) ou en utilisant le AWS SDKs, consultez la [référence de l'AWS Clean Rooms API](#).

Pour plus d'informations sur la journalisation des requêtes, consultez [Connexion à une analyse AWS Clean Rooms](#).

Note

Si vous exécutez une requête sur des tables de données [chiffrées](#), les résultats des colonnes chiffrées sont chiffrés.

Configuration des propriétés Spark pour les requêtes SQL

AWS Clean Rooms vous permet de personnaliser éventuellement le comportement d'exécution de Spark en configurant les propriétés Spark prises en charge pour les requêtes SQL. Ces propriétés vous permettent d'affiner les performances, l'utilisation de la mémoire et les paramètres d'exécution des requêtes. Grâce à cette fonctionnalité, vous avez un meilleur contrôle sur le traitement de vos requêtes basées sur Spark, ce qui permet une optimisation en fonction de vos exigences spécifiques en matière de charge de travail.

Vous pouvez désormais ajuster des paramètres tels que le shuffle des partitions, les seuils de jointure de diffusion et les paramètres d'exécution adaptative des requêtes directement depuis la AWS Clean Rooms console. Cette fonctionnalité est particulièrement utile pour les requêtes complexes ou les grands ensembles de données pour lesquels les configurations par défaut peuvent ne pas être optimales. En affinant ces propriétés de Spark, vous pouvez potentiellement améliorer les performances des requêtes, réduire la consommation de ressources et mieux gérer l'utilisation de la mémoire pour vos analyses de collaboration basées sur Spark.

Pour tirer parti de cette fonctionnalité, vous trouverez une nouvelle section sur les propriétés de Spark dans l'interface de requête. Vous pouvez effectuer une sélection dans la liste des propriétés prises en charge et spécifier des valeurs personnalisées. [Vous pouvez également configurer les propriétés de Spark par programmation à l'aide de l'StartProtectedQuery API](#). Cette option de configuration avancée permet aux analystes de données et aux ingénieurs d'optimiser leurs requêtes pour une efficacité et une évolutivité accrues.

Pour plus d'informations sur les propriétés de Spark, y compris les valeurs par défaut, consultez la section [Propriétés de Spark](#) dans la documentation d'Apache Spark.

Les rubriques suivantes expliquent comment interroger des données dans le cadre d'une collaboration à l'aide de la AWS Clean Rooms console.

Rubriques

- [Interrogation de tables configurées à l'aide de l'éditeur de code SQL](#)
- [Interrogation des tables de mappage d'identifiants à l'aide de l'éditeur de code SQL](#)
- [Interrogation de tables configurées à l'aide d'un modèle d'analyse SQL](#)
- [Interrogation à l'aide du générateur d'analyse](#)
- [Visualisation de l'impact de la confidentialité différentielle](#)
- [Affichage des requêtes récentes](#)
- [Affichage des détails de la requête](#)

Interrogation de tables configurées à l'aide de l'éditeur de code SQL

En tant que membre habilité à effectuer des requêtes, vous pouvez créer une requête manuellement en écrivant du code SQL dans l'éditeur de code SQL. L'éditeur de code SQL se trouve dans la section Analyse de l'onglet Analyse de la AWS Clean Rooms console.

L'éditeur de code SQL s'affiche par défaut. Si vous souhaitez utiliser le générateur d'analyse pour créer des requêtes, consultez [Interrogation à l'aide du générateur d'analyse](#).

Important

Si vous commencez à écrire une requête SQL dans l'éditeur de code, puis que vous activez l'interface utilisateur du générateur d'analyse, votre requête n'est pas enregistrée.

AWS Clean Rooms prend en charge de nombreuses commandes, fonctions et conditions SQL. Pour plus d'informations, consultez la [référence AWS Clean Rooms SQL](#).

 Tip

Si une maintenance planifiée a lieu pendant l'exécution d'une requête, celle-ci est interrompue et annulée. Vous devez relancer la requête.

Pour interroger les tables configurées à l'aide de l'éditeur de code SQL

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est défini sur Exécuter des requêtes.
4. Dans l'onglet Analyse, sous Tables, consultez la liste des tables et le type de règle d'analyse associé (règle d'analyse d'agrégation, règle d'analyse de liste ou règle d'analyse personnalisée).

 Note

Si les tables attendues ne figurent pas dans la liste, c'est peut-être pour les raisons suivantes :

- Les tables n'ont pas été [associées](#).
- Aucune [règle d'analyse n'est configurée pour](#) les tables.

5. (Facultatif) Pour afficher le schéma et les contrôles des règles d'analyse du tableau, développez le tableau en sélectionnant l'icône du signe plus (+).
6. Dans la section Analyse, pour le mode Analyse, sélectionnez Écrire du code SQL.

 Note

La section Analyse s'affiche uniquement si le membre qui peut recevoir les résultats et le membre chargé de payer les coûts de calcul des requêtes ont rejoint la collaboration en tant que membre actif.

7. Créez la requête en la saisissant dans l'éditeur de code SQL.

Pour plus d'informations sur les commandes et fonctions SQL prises en charge, consultez la [référence AWS Clean Rooms SQL](#).

Vous pouvez également utiliser les options suivantes pour créer votre requête.

Use an example query

Pour utiliser un exemple de requête

1. Sélectionnez les trois points verticaux à côté du tableau.
2. Sous Insérer dans l'éditeur, sélectionnez Exemple de requête.

Note

L'insertion d'une requête d'exemple l'ajoute à la requête déjà présente dans l'éditeur.

L'exemple de requête apparaît. Toutes les tables répertoriées sous Tables sont incluses dans la requête.

3. Modifiez les valeurs de l'espace réservé dans la requête.

Insert column names or functions

Pour insérer un nom ou une fonction de colonne

1. Sélectionnez les trois points verticaux à côté d'une colonne.
2. Sous Insérer dans l'éditeur, sélectionnez Nom de colonne.
3. Pour insérer manuellement une fonction autorisée sur une colonne,
 - a. Sélectionnez les trois points verticaux à côté d'une colonne.
 - b. Sélectionnez Insérer dans l'éditeur.
 - c. Sélectionnez le nom de la fonction autorisée (par exemple INNER JOIN SUM, SUM DISTINCT, ou COUNT).
4. Appuyez sur Ctrl + Espace pour afficher les schémas de table dans l'éditeur de code.

Note

Les membres autorisés à effectuer des requêtes peuvent consulter et utiliser les colonnes de partition dans chaque association de tables configurée. Assurez-vous que la colonne de partition est étiquetée en tant que colonne de partition dans la AWS Glue table sous-jacente à la table configurée.

5. Modifiez les valeurs de l'espace réservé dans la requête.

8. Spécifiez le type de travailleur pris en charge et le nombre de travailleurs.

Vous pouvez choisir le type d'instance et le nombre d'instances (travailleurs) pour exécuter vos requêtes SQL.

Pour le CR.1X, vous pouvez sélectionner jusqu'à 128 travailleurs ou un minimum de 4 travailleurs.

Pour le CR.4X, vous pouvez sélectionner jusqu'à 32 travailleurs ou un minimum de 4 travailleurs.

Utilisez le tableau suivant pour déterminer le type et le nombre de travailleurs dont vous avez besoin pour votre cas d'utilisation.

Type d'employé	vCPU	Mémoire (Go)	Stockage (Go)	Nombre d'employés	Nombre total d'unités de traitement pour salles blanches (CRPU)
CR.1X (par défaut)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

Les différents types de travailleurs et le nombre de travailleurs entraînent des coûts associés. Pour en savoir plus sur les tarifs, consultez la section [AWS Clean Rooms Tarifs](#).

9. Pour Envoyer les résultats à, spécifiez qui peut recevoir les résultats.

 Note

Pour recevoir des résultats, le membre de la collaboration doit être configuré en tant que destinataire des résultats et doit participer activement à la collaboration (statut : actif)

10. (Membre autorisé à interroger uniquement) La case Utiliser vos paramètres de résultats par défaut est cochée par défaut. Gardez cette option sélectionnée si vous souhaitez conserver vos paramètres de résultats par défaut.

Si vous souhaitez définir des paramètres de résultats différents pour cette requête, décochez la case Utiliser vos paramètres de résultats par défaut, puis choisissez ce qui suit.

- a. Format du résultat (CSV ou PARQUET)
- b. Fichiers de résultats (simples ou multiples)
- c. Destination des résultats dans Amazon S3

Chaque membre autorisé à recevoir des résultats peut spécifier un format de résultat, des fichiers de résultats et une destination de résultats différents dans Amazon S3.

11. Pour définir les propriétés de Spark :

- a. Développez les propriétés Spark.
- b. Choisissez Ajouter des propriétés Spark.
- c. Dans la boîte de dialogue des propriétés de Spark, choisissez un nom de propriété dans la liste déroulante et entrez une valeur.

Le tableau suivant fournit une définition pour chaque propriété.

Pour plus d'informations sur les propriétés de Spark, consultez la section [Propriétés de Spark](#) dans la documentation d'Apache Spark.

Nom de propriété	Description	Valeur par défaut
Défaillances de Spark.Task.Max	Contrôle le nombre de fois consécutives qu'une tâche peut échouer avant que la tâche échoue. Nécessite une valeur supérieure ou égale à 1. Le nombre de tentatives autorisées est égal à cette valeur moins 1. Le nombre d'échecs est réinitialisé en cas de réussite d'une tentative. Les échecs liés aux différentes tâches ne s'accumulent pas jusqu'à atteindre cette limite.	4
fichiers spark.sql.maxPartitionBytes	Définit le nombre maximum d'octets à regrouper dans une seule partition lors de la lecture à partir de sources basées sur des fichiers telles que Parquet, JSON et ORC.	128 MO
Spark.Hadoop.FS.S3.max Réessaie	Définit le nombre maximal de tentatives pour les opérations sur les fichiers Amazon S3.	
spark.network.timeout	Définit le délai d'expiration par défaut pour toutes les interactions réseau. Remplace les paramètres de délai d'expiration suivants s'ils ne sont pas configurés : - spark.storage.blockManagerHeartbeatTimeoutMs - Spark.shuffle.io. Délai de connexion - Spark.rpc.AskTimeout - Spark.rpc. Délai de recherche expiré	années 120

Nom de propriété	Description	Valeur par défaut
spark.rdd.compress	Spécifie s'il faut compresser les partitions RDD sérialisées à l'aide de spark.io.compression.codec. S'applique à StorageLevel .MEMORY_ONLY_SER en Java et Scala, ou à .MEMORY_ONLY en Python. StorageLevel Réduit l'espace de stockage mais nécessite un temps de traitement supplémentaire du processeur.	FALSE
spark.shuffle.spill.compress	Spécifie s'il faut compresser les données du shuffle spill à l'aide de spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Définit la taille cible en octets pour les partitions shuffle lors de l'optimisation adaptative lorsque spark.sql.adaptive.enabled est vrai. Contrôle la taille de la partition lors de la fusion de petites partitions ou du fractionnement de partitions asymétriques.	(valeur de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)
spark.sql.adaptive.autoBroadcastJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. S'applique uniquement dans le cadre adaptatif. Utilise la même valeur par défaut que spark.sql.autoBroadcastJoinSeuil. Réglez sur -1 pour désactiver la diffusion.	(aucun)
spark.sql.adaptive.CoalescePartitions.Activé	Spécifie s'il faut fusionner les partitions shuffle contiguës sur la base de spark.sql.adaptive.advisoryPartitionSizeInBytes pour optimiser la taille des tâches. Nécessite que spark.sql.adaptive.enabled soit vrai.	TRUE

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive. .CoalescePartitions. initialPartitionNum	Définit le nombre initial de partitions shuffle avant la fusion. Nécessite que spark.sql.adaptive.enabled et spark.sql.adaptive.CoalescePartitions.enabled soient vrais. La valeur par défaut est spark.sql.shuffle.partitions.	(aucun)
spark.sql.adaptive. .CoalescePartitions. minPartitionSize	Définit la taille minimale des partitions shuffle coalescées afin d'éviter que les partitions ne deviennent trop petites lors de l'optimisation adaptative.	1 Mo
spark.sql.adaptive. .coalescepartition s.ParallelismFirst	Spécifie s'il faut calculer la taille des partitions en fonction du parallélisme des clusters plutôt que de spark.sql.adaptive.advisoryPartitionSizeInBytes lors de la fusion des partitions. Génère des tailles de partition inférieures à la taille cible configurée afin de maximiser le parallélisme. Nous vous recommandons de définir ce paramètre sur false sur les clusters occupés afin d'améliorer l'utilisation des ressources en évitant de trop petites tâches.	TRUE
spark.sql.adaptive. .enabled	Spécifie s'il faut activer l'exécution adaptative des requêtes afin de réoptimiser les plans de requêtes pendant l'exécution des requêtes, sur la base de statistiques d'exécution précises.	TRUE
spark.sql.adaptive. forceOptimizeSkewed Joignez-vous	Spécifie s'il faut forcer l'activation OptimizeSkewedJoin même si cela introduit un shuffle supplémentaire.	FALSE

Nom de propriété	Description	Valeur par défaut
<code>spark.sql.adaptive.localShuffleReader</code> .activé	Spécifie s'il faut utiliser des lecteurs de shuffle locaux lorsque le partitionnement aléatoire n'est pas requis, par exemple après la conversion des jointures par tri-fusion en jointures par hachage par diffusion. Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	Définit la taille de partition maximale en octets pour la création de cartes de hachage locales. Privilégie les jointures par hachage mélangées par rapport aux jointures par tri-fusion lorsque : - Cette valeur est égale ou supérieure à <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> - Toutes les tailles de partition se situent dans cette limite Remplace <code>spark.sql.join.preferSortMergeParameter</code> de jointure.	0 octet
<code>spark.sql.adaptive.optimizeSkewsInRebalancePartitions</code> .activé	Spécifie s'il faut optimiser les partitions shuffle asymétriques en les divisant en partitions plus petites sur la base de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> . Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	Définit le facteur de seuil de taille pour la fusion des partitions lors du fractionnement. Partitions inférieures à ce facteur multipliées par <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> sont fusionnées.	0.2

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive .SkewJoin.Enabled	Spécifie s'il faut gérer l'inclinaison des données dans les jointures mélangées en divisant et éventuellement en répliquant les partitions asymétriques. S'applique aux jointures par tri-fusion et aux jointures de hachage mélangées. Nécessite que spark.sql.adaptive.enabled soit vrai.	TRUE
spark.sql.adaptive .skewJoin. skewedPartitionFactor	Détermine le facteur de taille qui détermine l'inclinaison de la partition. Une partition est inclinée lorsque sa taille dépasse les deux valeurs suivantes : - Ce facteur multiplié par la taille médiane de la partition - La valeur de spark.sql.adaptive.SkewJoin.skewedPartitionThresholdInBytes	5
spark.sql.adaptive .skewJoin. skewedPartitionThresholdInBytes	Définit le seuil de taille en octets pour identifier les partitions asymétriques. Une partition est inclinée lorsque sa taille dépasse les deux valeurs suivantes : - Ce seuil - Taille de partition médiane multipliée par spark.sql.adaptive.skewJoin.skewedPartitionFactor Nous recommandons de définir cette valeur à une valeur supérieure à spark.sql.adaptive.advisoryPartitionSizeInBytes.	256 MO

Nom de propriété	Description	Valeur par défaut
spark.sql. autoBroad castJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. Réglez sur -1 pour désactiver la diffusion.	10 MO
Spark.SQL. Délai de diffusion	Contrôle le délai d'expiration en secondes pour les opérations de diffusion lors des jointures de diffusion.	300 secondes
spark.sql.cbo.enab led	Spécifie s'il faut activer l'optimisation basée sur les coûts (CBO) pour l'estimation des statistiques du plan.	FALSE
spark.sql.cbo.join reorder.dp.star.filter	Spécifie s'il faut appliquer l'heuristique du filtre de jointure en étoile lors de l'énumération des jointures basée sur les coûts.	FALSE
spark.sql.cbo.join reorder.dp.Threshold	Définit le nombre maximum de nœuds joints autorisés dans l'algorithme de programmation dynamique.	12
spark.sql.cbo.Join Reorder.Enabled	Spécifie s'il faut activer la réorganisation des jointures dans le cadre de l'optimisation basée sur les coûts (CBO).	FALSE
spark.sql.cbo.plan stats.activé	Spécifie s'il faut extraire le nombre de lignes et les statistiques des colonnes du catalogue lors de la génération du plan logique.	FALSE
spark.sql.cbo. starSchemaDetectio n	Spécifie s'il faut activer la réorganisation des jointures en fonction de la détection du schéma en étoile.	FALSE

Nom de propriété	Description	Valeur par défaut
fichiers spark.sql. maxPartitionNum	Définit le nombre maximal cible de partitions de fichiers fractionnés pour les sources basées sur des fichiers (Parquet, JSON et ORC). Redimensionne les partitions lorsque le nombre initial dépasse cette valeur. Il s'agit d'un objectif suggéré et non d'une limite garantie.	(aucun)
fichiers spark.sql. . maxRecord sPerDossier	Définit le nombre maximum d'enregistrements à écrire dans un seul fichier. Aucune limite ne s'applique lorsqu'il est réglé sur zéro ou sur une valeur négative.	0
fichiers spark.sql. minPartitionNum	Définit le nombre minimum cible de partitions de fichiers fractionnés pour les sources basées sur des fichiers (Parquet, JSON et ORC). La valeur par défaut est spark.sql.leafNodeDefaultParallélisme. Il s'agit d'un objectif suggéré et non d'une limite garantie.	(aucun)
spark.sql. inMemoryC olumnarStockage. Taille du lot	Contrôle la taille du lot pour la mise en cache en colonnes. L'augmentation de la taille améliore l'utilisation de la mémoire et la compression, mais augmente le risque d'out-of-memoryerreurs.	10 000
spark.sql. inMemoryC olumnarStockage compressé	Spécifie s'il faut sélectionner automatiquement les codecs de compression pour les colonnes en fonction des statistiques de données.	TRUE
spark.sql. inMemoryC olumnarRangement enableVectorizedRe ader	Spécifie s'il faut activer la lecture vectorisée pour la mise en cache en colonnes.	TRUE

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.allowHashOnMapType	Spécifie s'il faut autoriser les opérations de hachage sur les structures de données de type carte. Cet ancien paramètre assure la compatibilité avec la gestion des types de carte dans les anciennes versions de Spark.	
spark.sql.legacy.allowNegativeScaleOfDecimal	Spécifie s'il faut autoriser les valeurs d'échelle négatives dans les définitions de type décimal. Cet ancien paramètre maintient la compatibilité avec les anciennes versions de Spark qui prenaient en charge les échelles décimales négatives.	
spark.sql.legacy.castComplexTypesToString.activated	Spécifie s'il faut activer le comportement existant pour convertir des types complexes en chaînes. Maintient la compatibilité avec les règles de conversion de type des anciennes versions de Spark.	
spark.sql.legacy.charVarcharAsString	Spécifie s'il faut traiter les types CHAR et VARCHAR comme des types STRING. Cet ancien paramètre assure la compatibilité avec la gestion des types de chaînes des anciennes versions de Spark.	
spark.sql.legacy.createEmptyCollectionUsingStringType	Spécifie s'il faut créer des collections vides à l'aide d'éléments de type chaîne. Cet ancien paramètre maintient la compatibilité avec le comportement d'initialisation des collections des anciennes versions de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal.activated	Spécifie s'il faut interpréter les littéraux exponentiels comme des types décimaux. Cet ancien paramètre assure la compatibilité avec la gestion littérale numérique des anciennes versions de Spark.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.json.allowEmptyString.activé	Spécifie s'il faut autoriser les chaînes vides dans le traitement JSON. Cet ancien paramètre maintient la compatibilité avec le comportement d'analyse JSON des anciennes versions de Spark.	
spark.sql.legacy.parquet.int96RebaseModelRead	Spécifie s'il faut utiliser l'ancien mode de rebase d' INT96 horodatage lors de la lecture des fichiers Parquet. Cet ancien paramètre assure la compatibilité avec la gestion des horodatages des anciennes versions de Spark.	
spark.sql.legacy.timeParserPolicy	Contrôle le comportement d'analyse temporelle à des fins de rétrocompatibilité. Cet ancien paramètre détermine la manière dont les horodatages et les dates sont analysés à partir de chaînes.	
spark.sql.Legacy.TypeCoercion.datetimeToString.activé	Spécifie s'il faut activer le comportement de coercion de type existant lors de la conversion de valeurs date/heure en chaînes. Maintient la compatibilité avec les règles de conversion date/heure des anciennes versions de Spark.	
spark.sql. maxSinglePartitionOctets	Définit la taille maximale de la partition en octets. Le planificateur introduit des opérations de shuffle pour les partitions plus grandes afin d'améliorer le parallélisme.	128 m

Nom de propriété	Description	Valeur par défaut
Cache de métadonnées Spark.SQL TTLSeconds	Contrôle le time-to-live (TTL) pour les caches de métadonnées. S'applique aux métadonnées des fichiers de partition et aux caches du catalogue de sessions. Nécessite : • Une valeur positive supérieure à zéro • spark.sql.CatalogImplementation défini sur hive • spark.sql.hive. fileSourcePartitionFileCacheSize supérieur à zéro • spark.sql.hive. manageFileSourcePartitions défini sur true	-1000 ms
spark.sql.optimizer.collapseProjectAlwaysEn ligne	Spécifie s'il faut réduire les projections adjacentes et les expressions en ligne, même si cela entraîne une duplication.	FALSE
spark.sql.optimizer.dynamicPartitionPruning.activé	Spécifie s'il faut générer des prédicats pour les colonnes de partition utilisées comme clés de jointure.	TRUE
spark.sql.optimizer.enableCsvExpressionOptimisation	Spécifie s'il faut optimiser les expressions CSV dans l'optimiseur SQL en supprimant les colonnes inutiles des opérations from_csv.	TRUE
spark.sql.optimizer.enableJsonExpressionOptimisation	Spécifie s'il faut optimiser les expressions JSON dans l'optimiseur SQL en : • Supprimer les colonnes inutiles des opérations from_json • Simplification des combinaisons from_json et to_json • Optimisation des opérations named_struct	TRUE

Nom de propriété	Description	Valeur par défaut
Spark.SQL.Optimizer.Règles exclues	Définit les règles d'optimisation à désactiver, identifiées par des noms de règles séparés par des virgules. Certaines règles ne peuvent pas être désactivées car elles sont nécessaires pour être correctes. L'optimiseur enregistre les règles qui ont été désactivées avec succès.	(aucun)
Spark.SQL.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Définit la taille de scan agrégée minimale en octets requise pour injecter un filtre Bloom côté application.	10 GO
Spark.SQL.Optimizer.Runtime.BloomFilter.creationSideThreshold	Définit le seuil de taille maximale pour injecter un filtre Bloom côté création.	10 MO
Spark.SQL.Optimizer.Runtime.BloomFilter.Activé	Spécifie s'il faut insérer un filtre Bloom pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	TRUE
Spark.SQL.Optimizer.Runtime.BloomFilter.expectedNumItems	Définit le nombre par défaut d'éléments attendus dans le filtre Bloom d'exécution.	1000000
Spark.SQL.Optimizer.Runtime.BloomFilter.maxNumBits	Définit le nombre maximum de bits autorisés dans le filtre Bloom d'exécution.	67108864
Spark.SQL.Optimizer.Runtime.BloomFilter.maxNumItems	Définit le nombre maximum d'éléments attendus autorisés dans le filtre Bloom d'exécution.	4000000

Nom de propriété	Description	Valeur par défaut
spark.sql.Optimizer.Runtime.BloomFilter.Number.Threshold	Limite le nombre maximum de filtres d'exécution non DPP autorisés par requête afin d'éviter les out-of-memory erreurs dans le pilote.	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Définit le nombre de bits par défaut utilisés dans le filtre Bloom d'exécution.	8388608
spark.sql.optimize r.runtime. rowlevelO perationGroupFiltr e.activé	Spécifie s'il faut activer le filtrage des groupes d'exécution pour les opérations au niveau des lignes. Permet aux sources de données de : • Réduisez des groupes entiers de données (tels que des fichiers ou des partitions) à l'aide de filtres de source de données • Exécuter des requêtes d'exécution pour identifier les enregistrements correspondants • Supprimez les groupes inutiles pour éviter des réécritures coûteuses Limites: • Toutes les expressions ne peuvent pas être converties en filtres de source de données • Certaines expressions nécessitent une évaluation Spark (telles que les sous-requêtes)	TRUE

Nom de propriété	Description	Valeur par défaut
Spark.SQL.Optimize r.RuntimeFilter. semiJoinReduction. activé	Spécifie s'il faut insérer une semi-jointure pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	FALSE
spark.sql.parquet. AggregatePushdown	Spécifie s'il faut transférer les agrégats vers Parquet à des fins d'optimisation. Supporte : • MIN et MAX pour les types booléen, entier, flottant et date • COUNT pour tous les types de données Lance une exception si les statistiques sont absentes du pied de page d'un fichier Parquet.	FALSE
spark.sql.parquet. columnarReaderBatch Taille	Contrôle le nombre de lignes dans chaque lot de lecteurs vectorisés Parquet. Choisissez une valeur qui équilibre le surcoût des performances et l'utilisation de la mémoire afin d'éviter out-of-memory les erreurs.	4096
spark.sql.session. TimeZone	Définit le fuseau horaire de session pour la gestion des horodatages sous forme de chaînes littérales et pour la conversion d'objets Java. Accepte : • area/city Format basé sur la région IDs (tel que America/Los_Angeles) • Décalages de zone au format (+/-) HH, (+/-) HH:mm ou (+/-) HH:mm:ss (tel que -08 ou + 01:00) • UTC ou Z comme alias pour + 00:00	(valeur du fuseau horaire local)

Nom de propriété	Description	Valeur par défaut
spark.sql.shuffle.partitions	Définit le nombre de partitions par défaut pour le brassage des données lors des jointures ou des agrégations. Ne peut pas être modifié entre les redémarrages de requêtes de streaming structurées à partir du même emplacement de point de contrôle.	200
spark.sql.shuffleHashJoinFactor	Définit le facteur de multiplication utilisé pour déterminer l'éligibilité à la jointure par hachage aléatoire. Une jointure par hachage aléatoire est sélectionnée lorsque la taille des données du petit côté multipliée par ce facteur est inférieure à la taille des données du grand côté.	3
spark.sql.sources.parallelPartitionDiscovery.seuil	Définit le nombre maximum de chemins pour la liste des fichiers côté pilote avec des sources basées sur des fichiers (Parquet, JSON et ORC). En cas de dépassement lors de la découverte de partitions, les fichiers sont répertoriés à l'aide d'une tâche distribuée Spark distincte.	32
spark.sql.statistics.histogram.enabled	Spécifie s'il faut générer des histogrammes à hauteur égale lors du calcul des statistiques des colonnes afin d'améliorer la précision de l'estimation. Nécessite une analyse de table supplémentaire en plus de celle requise pour les statistiques de base sur les colonnes.	FALSE

12. Cliquez sur Exécuter.

 Note

Vous ne pouvez pas exécuter la requête si le membre qui peut recevoir les résultats n'a pas configuré les paramètres des résultats de la requête.

13. Consultez les résultats.

Pour de plus amples informations, veuillez consulter [Réception et utilisation des résultats d'analyse](#).

14. Continuez à ajuster les paramètres et réexécutez votre requête, ou cliquez sur le bouton + pour démarrer une nouvelle requête dans un nouvel onglet.

 Note

AWS Clean Rooms vise à fournir un message d'erreur clair. Si un message d'erreur ne contient pas suffisamment de détails pour vous aider à résoudre le problème, contactez l'équipe chargée du compte. Fournissez-leur une description de la façon dont l'erreur s'est produite et du message d'erreur (y compris les éventuels identifiants). Pour de plus amples informations, veuillez consulter [Résolution des problèmes AWS Clean Rooms](#).

Interrogation des tables de mappage d'identifiants à l'aide de l'éditeur de code SQL

La procédure suivante décrit comment exécuter une requête de jointure multitable sur la table de mappage d'identifiants pour joindre `sourceId` et `targetId`.

Avant de demander la table de mappage d'identifiants, celle-ci doit être correctement renseignée.

Pour interroger les tables de mappage d'ID à l'aide de l'éditeur de code SQL

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est défini sur Exécuter des requêtes.

4. Dans l'onglet Analyse, accédez à la section Analyse.

 Note

La section Analyse s'affiche uniquement si le membre qui peut recevoir les résultats et le membre chargé de payer les coûts de calcul des requêtes ont rejoint la collaboration en tant que membre actif.

5. Dans l'onglet Analyse, sous Tables, consultez la liste des tables de mappage d'ID (sous Géré par AWS Clean Rooms) et le type de règle d'analyse associé (règle d'analyse de table de mappage d'ID).

 Note

Si vous ne voyez pas les tables de mappage d'identifiants que vous attendez dans la liste, cela peut être dû au fait que les tables de mappage d'identifiants n'ont pas été correctement remplies. Pour de plus amples informations, veuillez consulter [Remplissage d'une table de mappage d'identifiants existante](#).

6. Créez la requête en la saisissant dans l'éditeur de code SQL.

(Facultatif) Si vous souhaitez utiliser un exemple de requête

1. Sélectionnez les trois points verticaux à côté du tableau.
2. Sous Insérer dans l'éditeur, choisissez Exemple JOIN statement.

 Note

L'insertion d'un exemple d'instructions JOIN ajoute la requête déjà dans l'éditeur.

L'exemple d'instruction JOIN apparaît.

(Facultatif) Si vous souhaitez insérer un nom de table

1. Sélectionnez les trois points verticaux à côté d'une colonne.
2. Sous Insérer dans l'éditeur, choisissez Nom de la table.
3. Modifiez les valeurs de l'espace réservé dans la requête.

(Facultatif) Si vous souhaitez utiliser un exemple de requête

(Facultatif) Si vous souhaitez insérer un nom de table

3. Modifiez les valeurs de l'espace réservé dans la requête.

7. Spécifiez le type de travailleur pris en charge et le nombre de travailleurs.

Utilisez le tableau suivant pour déterminer le type et le nombre de travailleurs dont vous avez besoin pour votre cas d'utilisation.

Type d'employé	vCPU	Mémoire (Go)	Stockage (Go)	Nombre d'employés	Nombre total d'unités de traitement pour salles blanches (CRPU)
CR.1X (par défaut)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

Les différents types de travailleurs et le nombre de travailleurs entraînent des coûts associés. Pour en savoir plus sur les tarifs, consultez la section [AWS Clean Rooms Tarifs](#).

8. Cliquez sur Exécuter.

 Note

Vous ne pouvez pas exécuter la requête si le membre qui peut recevoir les résultats n'a pas configuré les paramètres des résultats de la requête.

9. Consultez les résultats.

Pour de plus amples informations, veuillez consulter [Réception et utilisation des résultats d'analyse](#).

10. Continuez à ajuster les paramètres et réexécutez votre requête, ou cliquez sur le bouton + pour démarrer une nouvelle requête dans un nouvel onglet.

 Note

AWS Clean Rooms vise à fournir un message d'erreur clair. Si un message d'erreur ne contient pas suffisamment de détails pour vous aider à résoudre le problème, contactez l'équipe chargée du compte. Fournissez-leur une description de la façon dont l'erreur s'est produite et du message d'erreur (y compris les éventuels identifiants). Pour de plus amples informations, veuillez consulter [Résolution des problèmes AWS Clean Rooms](#).

Interrogation de tables configurées à l'aide d'un modèle d'analyse SQL

Cette procédure explique comment utiliser un modèle d'analyse dans la AWS Clean Rooms console pour interroger des tables configurées à l'aide de la règle d'analyse personnalisée.

Pour utiliser un modèle d'analyse SQL pour interroger les tables configurées à l'aide de la règle d'analyse personnalisée

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est défini sur Exécuter des requêtes.
4. Dans l'onglet Analyse, sous la section Tables, visualisez les tables et le type de règle d'analyse associé (règle d'analyse personnalisée).

Note

Si les tables attendues ne figurent pas dans la liste, c'est peut-être pour les raisons suivantes :

- Les tables n'ont pas été [associées](#).
- Aucune [règle d'analyse n'est configurée pour](#) les tables.

5. Dans la section Analyse, pour le mode Analyse, sélectionnez Exécuter les modèles d'analyse, puis choisissez le modèle d'analyse dans la liste déroulante.
6. Les paramètres du modèle d'analyse SQL seront automatiquement renseignés dans la définition.
7. Spécifiez le type de travailleur pris en charge et le nombre de travailleurs.

Utilisez le tableau suivant pour déterminer le type et le nombre de travailleurs dont vous avez besoin pour votre cas d'utilisation.

Type d'employé	vCPU	Mémoire (Go)	Stockage (Go)	Nombre d'employés	Nombre total d'unités de traitement pour salles blanches (CRPU)
CR.1X (par défaut)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

Les différents types de travailleurs et le nombre de travailleurs entraînent des coûts associés. Pour en savoir plus sur les tarifs, consultez la section [AWS Clean Rooms Tarifs](#).

8. Spécifiez les propriétés Spark prises en charge.
 - a. Sélectionnez Ajouter des propriétés Spark.
 - b. Dans la boîte de dialogue des propriétés de Spark, choisissez un nom de propriété dans la liste déroulante et entrez une valeur.

Le tableau suivant fournit une définition pour chaque propriété.

Pour plus d'informations sur les propriétés de Spark, consultez la section [Propriétés de Spark](#) dans la documentation d'Apache Spark.

Nom de propriété	Description	Valeur par défaut
Défaillances de Spark.Task.Max	Contrôle le nombre de fois consécutives qu'une tâche peut échouer avant que la tâche échoue. Nécessite une valeur supérieure ou égale à 1. Le nombre de tentatives autorisées est égal à cette valeur moins 1. Le nombre d'échecs est réinitialisé en cas de réussite d'une tentative. Les échecs liés aux différentes tâches ne s'accumulent pas jusqu'à atteindre cette limite.	4
fichiers spark.sql.maxPartitionBytes	Définit le nombre maximum d'octets à regrouper dans une seule partition lors de la lecture à partir de sources basées sur des fichiers telles que Parquet, JSON et ORC.	128 MO

Nom de propriété	Description	Valeur par défaut
Spark.Hadoop.FS.S3.max Réessaie	Définit le nombre maximal de tentatives pour les opérations sur les fichiers Amazon S3.	
spark.network.timeout	Définit le délai d'expiration par défaut pour toutes les interactions réseau. Remplace les paramètres de délai d'expiration suivants s'ils ne sont pas configurés : • spark.storage.blockManagerHeartbeatTimeoutMs • Spark.shuffle.io. Délai de connexion • Spark.rpc.AskTimeout • Spark.rpc. Délai de recherche expiré	années 120
spark.rdd.compress	Spécifie s'il faut compresser les partitions RDD sérialisées à l'aide de spark.io.compression.codec. S'applique à StorageLevel .MEMORY_ONLY_SER en Java et Scala, ou à .MEMORY_ONLY en Python. StorageLevel Réduit l'espace de stockage mais nécessite un temps de traitement supplémentaire du processeur.	FALSE
spark.shuffle.spill.compress	Spécifie s'il faut compresser les données du shuffle spill à l'aide de spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Définit la taille cible en octets pour les partitions shuffle lors de l'optimisation adaptative lorsque spark.sql.adaptive.enabled est vrai. Contrôle la taille de la partition lors de la fusion de petites partitions ou du fractionnement de partitions asymétriques.	(valeur de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive.autoBroadcastJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. S'applique uniquement dans le cadre adaptatif. Utilise la même valeur par défaut que spark.sql.autoBroadcastJoinSeuil. Réglez sur -1 pour désactiver la diffusion.	(aucun)
spark.sql.adaptive.CoalescePartitions.Activé	Spécifie s'il faut fusionner les partitions shuffle contiguës sur la base de spark.sql.adaptive.advisoryPartitionSizeInBytes pour optimiser la taille des tâches. Nécessite que spark.sql.adaptive.enabled soit vrai.	TRUE
spark.sql.adaptive.CoalescePartitions.initialPartitionNum	Définit le nombre initial de partitions shuffle avant la fusion. Nécessite que spark.sql.adaptive.enabled et spark.sql.adaptive.CoalescePartitions.enabled soient vrais. La valeur par défaut est spark.sql.shuffle.partitions.	(aucun)
spark.sql.adaptive.CoalescePartitions.minPartitionSize	Définit la taille minimale des partitions shuffle coalescées afin d'éviter que les partitions ne deviennent trop petites lors de l'optimisation adaptative.	1 Mo

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive.coalescepartition.s.ParallelismFirst	Spécifie s'il faut calculer la taille des partitions en fonction du parallélisme des clusters plutôt que de spark.sql.adaptive.advisoryPartitionSizeInBytes lors de la fusion des partitions. Génère des tailles de partition inférieures à la taille cible configurée afin de maximiser le parallélisme. Nous vous recommandons de définir ce paramètre sur false sur les clusters occupés afin d'améliorer l'utilisation des ressources en évitant de trop petites tâches.	TRUE
spark.sql.adaptive.enabled	Spécifie s'il faut activer l'exécution adaptative des requêtes afin de réoptimiser les plans de requêtes pendant l'exécution des requêtes, sur la base de statistiques d'exécution précises.	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoignez-vous	Spécifie s'il faut forcer l'activation OptimizeSkewedJoin même si cela introduit un shuffle supplémentaire.	FALSE
spark.sql.adaptive.localShuffleReader.activé	Spécifie s'il faut utiliser des lecteurs de shuffle locaux lorsque le partitionnement aléatoire n'est pas requis, par exemple après la conversion des jointures par tri-fusion en jointures par hachage par diffusion. Nécessite que spark.sql.adaptive.enabled soit vrai.	TRUE

Nom de propriété	Description	Valeur par défaut
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	Définit la taille de partition maximale en octets pour la création de cartes de hachage locales. Privilégie les jointures par hachage mélangées par rapport aux jointures par tri-fusion lorsque : - Cette valeur est égale ou supérieure à <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> - Toutes les tailles de partition se situent dans cette limite Remplace <code>spark.sql.join.preferSortMergePar</code> amètre de jointure.	0 octet
<code>spark.sql.adaptive.optimizeSkewsInRebalancePartitions.activé</code>	Spécifie s'il faut optimiser les partitions shuffle asymétriques en les divisant en partitions plus petites sur la base de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> . Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	Définit le facteur de seuil de taille pour la fusion des partitions lors du fractionnement. Partitions inférieures à ce facteur multipliées par <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> sont fusionnées.	0.2
<code>spark.sql.adaptive.SkewJoin.Enabled</code>	Spécifie s'il faut gérer l'inclinaison des données dans les jointures mélangées en divisant et éventuellement en répliquant les partitions asymétriques. S'applique aux jointures par tri-fusion et aux jointures de hachage mélangées. Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive.skewJoin.skewedPartitionFactor	Détermine le facteur de taille qui détermine l'inclinaison de la partition. Une partition est inclinée lorsque sa taille dépasse les deux : - Ce facteur multiplié par la taille médiane de la partition - La valeur de spark.sql.adaptive.SkewJoin.skewedPartitionThresholdInBytes	5
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	Définit le seuil de taille en octets pour identifier les partitions asymétriques. Une partition est inclinée lorsque sa taille dépasse les deux : - Ce seuil - Taille de partition médiane multipliée par spark.sql.adaptive.skewJoin.skewedPartitionFactor Nous recommandons de définir cette valeur à une valeur supérieure à spark.sql.adaptive.advisoryPartitionSizeInBytes.	256 MO
spark.sql. autoBroadcastJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. Réglez sur -1 pour désactiver la diffusion.	10 MO
Spark.SQL. Délai de diffusion	Contrôle le délai d'expiration en secondes pour les opérations de diffusion lors des jointures de diffusion.	300 secondes
spark.sql.cbo.enabled	Spécifie s'il faut activer l'optimisation basée sur les coûts (CBO) pour l'estimation des statistiques du plan.	FALSE

Nom de propriété	Description	Valeur par défaut
spark.sql.cbo.join reorder.dp.star.filter	Spécifie s'il faut appliquer l'heuristique du filtre de jointure en étoile lors de l'énumération des jointures basée sur les coûts.	FALSE
spark.sql.cbo.join reorder.dp.Threshold	Définit le nombre maximum de nœuds joints autorisés dans l'algorithme de programmation dynamique.	12
spark.sql.cbo.Join Reorder.Enabled	Spécifie s'il faut activer la réorganisation des jointures dans le cadre de l'optimisation basée sur les coûts (CBO).	FALSE
spark.sql.cbo.plan stats.activé	Spécifie s'il faut extraire le nombre de lignes et les statistiques des colonnes du catalogue lors de la génération du plan logique.	FALSE
spark.sql.cbo. starSchemaDetection	Spécifie s'il faut activer la réorganisation des jointures en fonction de la détection du schéma en étoile.	FALSE
fichiers spark.sql. maxPartitionNum	Définit le nombre maximal cible de partitions de fichiers fractionnées pour les sources basées sur des fichiers (Parquet, JSON et ORC). Redimensionne les partitions lorsque le nombre initial dépasse cette valeur. Il s'agit d'un objectif suggéré et non d'une limite garantie.	(aucun)
fichiers spark.sql . maxRecordsPerDossier	Définit le nombre maximum d'enregistrements à écrire dans un seul fichier. Aucune limite ne s'applique lorsqu'il est réglé sur zéro ou sur une valeur négative.	0

Nom de propriété	Description	Valeur par défaut
spark.sql.minPartitionNum	Définit le nombre minimum cible de partitions de fichiers fractionnées pour les sources basées sur des fichiers (Parquet, JSON et ORC). La valeur par défaut est spark.sql.leafNodeDefaultParallélisme. Il s'agit d'un objectif suggéré et non d'une limite garantie.	(aucun)
spark.sql.inMemoryColumnarStockage.Taille du lot	Contrôle la taille du lot pour la mise en cache en colonnes. L'augmentation de la taille améliore l'utilisation de la mémoire et la compression, mais augmente le risque d'out-of-memory erreurs.	10 000
spark.sql.inMemoryColumnarStockage.compressé	Spécifie s'il faut sélectionner automatiquement les codecs de compression pour les colonnes en fonction des statistiques de données.	TRUE
spark.sql.inMemoryColumnarRangement.enableVectorizedReader	Spécifie s'il faut activer la lecture vectorisée pour la mise en cache en colonnes.	TRUE
spark.sql.legacy.allowHashOnMapType	Spécifie s'il faut autoriser les opérations de hachage sur les structures de données de type carte. Cet ancien paramètre assure la compatibilité avec la gestion des types de carte dans les anciennes versions de Spark.	
spark.sql.legacy.allowNegativeScaleOfDecimal	Spécifie s'il faut autoriser les valeurs d'échelle négatives dans les définitions de type décimal. Cet ancien paramètre maintient la compatibilité avec les anciennes versions de Spark qui prenaient en charge les échelles décimales négatives.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.castComplexTypesToString.activé	Spécifie s'il faut activer le comportement existant pour convertir des types complexes en chaînes. Maintient la compatibilité avec les règles de conversion de type des anciennes versions de Spark.	
spark.sql.legacy.charVarcharAsString.activé	Spécifie s'il faut traiter les types CHAR et VARCHAR comme des types STRING. Cet ancien paramètre assure la compatibilité avec la gestion des types de chaînes des anciennes versions de Spark.	
spark.sql.legacy.createEmptyCollectionUsingStringType	Spécifie s'il faut créer des collections vides à l'aide d'éléments de type chaîne. Cet ancien paramètre maintient la compatibilité avec le comportement d'initialisation des collections des anciennes versions de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal.activé	Spécifie s'il faut interpréter les littéraux exponentiels comme des types décimaux. Cet ancien paramètre assure la compatibilité avec la gestion littérale numérique des anciennes versions de Spark.	
spark.sql.legacy.json.allowEmptyString.activé	Spécifie s'il faut autoriser les chaînes vides dans le traitement JSON. Cet ancien paramètre maintient la compatibilité avec le comportement d'analyse JSON des anciennes versions de Spark.	
spark.sql.legacy.parquet.int96RebaseModelRead	Spécifie s'il faut utiliser l'ancien mode de rebase d' INT96 horodatage lors de la lecture des fichiers Parquet. Cet ancien paramètre assure la compatibilité avec la gestion des horodatages des anciennes versions de Spark.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.timeParserPolicy	Contrôle le comportement d'analyse temporelle à des fins de rétrocompatibilité. Cet ancien paramètre détermine la manière dont les horodatages et les dates sont analysés à partir de chaînes.	
spark.sql.Legacy.TypeCoercion.datetimeToString.activated	Spécifie s'il faut activer le comportement de coercition de type existant lors de la conversion de valeurs date/heure en chaînes. Maintient la compatibilité avec les règles de conversion date/heure des anciennes versions de Spark.	
spark.sql.maxSinglePartitionOctets	Définit la taille maximale de la partition en octets. Le planificateur introduit des opérations de shuffle pour les partitions plus grandes afin d'améliorer le parallélisme.	128 m
Cache de métadonnées Spark.SQL TTLSeconds	Contrôle le time-to-live (TTL) pour les caches de métadonnées. S'applique aux métadonnées des fichiers de partition et aux caches du catalogue de sessions. Nécessite : • Une valeur positive supérieure à zéro • spark.sql.CatalogImplementation défini sur hive • spark.sql.hive.filesourcePartitionFileCacheSize supérieur à zéro • spark.sql.hive.manageFilesourcePartitions défini sur true	-1000 ms
spark.sql.optimizer.collapseProjectAlwaysEn ligne	Spécifie s'il faut réduire les projections adjacentes et les expressions en ligne, même si cela entraîne une duplication.	FALSE

Nom de propriété	Description	Valeur par défaut
spark.sql.optimizer.dynamicPartitionPruning.activé	Spécifie s'il faut générer des prédicats pour les colonnes de partition utilisées comme clés de jointure.	TRUE
spark.sql.optimizer.enableCsvExpressionOptimisation	Spécifie s'il faut optimiser les expressions CSV dans l'optimiseur SQL en supprimant les colonnes inutiles des opérations from_csv.	TRUE
spark.sql.optimizer.enableJsonExpressionOptimisation	Spécifie s'il faut optimiser les expressions JSON dans l'optimiseur SQL en : • Supprimer les colonnes inutiles des opérations from_json • Simplification des combinaisons from_json et to_json • Optimisation des opérations named_struct	TRUE
Spark.SQL.Optimizer.Règles exclues	Définit les règles d'optimisation à désactiver, identifiées par des noms de règles séparés par des virgules. Certaines règles ne peuvent pas être désactivées car elles sont nécessaires pour être correctes. L'optimiseur enregistre les règles qui ont été désactivées avec succès.	(aucun)
Spark.SQL.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Définit la taille de scan agrégée minimale en octets requise pour injecter un filtre Bloom côté application.	10 GO
Spark.SQL.Optimizer.Runtime.BloomFilter.creationSideThreshold	Définit le seuil de taille maximale pour injecter un filtre Bloom côté création.	10 MO

Nom de propriété	Description	Valeur par défaut
Spark.SQL.Optimize r.Runtime.BloomFilter.Activé	Spécifie s'il faut insérer un filtre Bloom pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	TRUE
Spark.SQL.Optimize r.Runtime.BloomFilter.expectedNumItems	Définit le nombre par défaut d'éléments attendus dans le filtre Bloom d'exécution.	1000000
Spark.SQL.Optimize r.Runtime.BloomFilter.maxNumBits	Définit le nombre maximum de bits autorisés dans le filtre Bloom d'exécution.	67108864
Spark.SQL.Optimize r.Runtime.BloomFilter.maxNumItems	Définit le nombre maximum d'éléments attendus autorisés dans le filtre Bloom d'exécution.	4000000
spark.sql.Optimize r.Runtime.BloomFilter.Number.Threshold	Limite le nombre maximum de filtres d'exécution non DPP autorisés par requête afin d'éviter les out-of-memory erreurs dans le pilote.	10
spark.sql.Optimize r.Runtime.BloomFilter.NumBits	Définit le nombre de bits par défaut utilisés dans le filtre Bloom d'exécution.	8388608

Nom de propriété	Description	Valeur par défaut
spark.sql.optimize r.runtime. rowlevelO perationGroupFiltr e.activé	Spécifie s'il faut activer le filtrage des groupes d'exécution pour les opérations au niveau des lignes. Permet aux sources de données de : • Réduisez des groupes entiers de données (tels que des fichiers ou des partitions) à l'aide de filtres de source de données • Exécuter des requêtes d'exécution pour identifier les enregistrements correspondants • Supprimez les groupes inutiles pour éviter des réécritures coûteuses Limites: • Toutes les expressions ne peuvent pas être converties en filtres de source de données • Certaines expressions nécessitent une évaluation Spark (telles que les sous-requêtes)	TRUE
Spark.SQL.Optimize r.RuntimeFilter. semiJoinReduction. activé	Spécifie s'il faut insérer une semi-jointure pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	FALSE

Nom de propriété	Description	Valeur par défaut
spark.sql.parquet. AggregatePushdown	Spécifie s'il faut transférer les agrégats vers Parquet à des fins d'optimisation. Supporte : • MIN et MAX pour les types booléen, entier, flottant et date • COUNT pour tous les types de données Lance une exception si les statistiques sont absentes du pied de page d'un fichier Parquet.	FALSE
spark.sql.parquet. columnarReaderBatchTaille	Contrôle le nombre de lignes dans chaque lot de lecteurs vectorisés Parquet. Choisissez une valeur qui équilibre le surcoût des performances et l'utilisation de la mémoire afin d'éviter out-of-memory les erreurs.	4096
spark.sql.session. TimeZone	Définit le fuseau horaire de session pour la gestion des horodatages sous forme de chaînes littérales et pour la conversion d'objets Java. Accepte : • area/city Format basé sur la région IDs (tel que America/Los_Angeles) • Décalages de zone au format (+/-) HH, (+/-) HH:mm ou (+/-) HH:mm:ss (tel que -08 ou + 01:00) • UTC ou Z comme alias pour + 00:00	(valeur du fuseau horaire local)
spark.sql.shuffle. partitions	Définit le nombre de partitions par défaut pour le brassage des données lors des jointures ou des agrégations. Ne peut pas être modifié entre les redémarrages de requêtes de streaming structurées à partir du même emplacement de point de contrôle.	200

Nom de propriété	Description	Valeur par défaut
spark.sql.shuffledHashJoinFacteur	Définit le facteur de multiplication utilisé pour déterminer l'éligibilité à la jointure par hachage aléatoire. Une jointure par hachage aléatoire est sélectionnée lorsque la taille des données du petit côté multipliée par ce facteur est inférieure à la taille des données du grand côté.	3
spark.sql.sources.parallelPartitionDiscovery.seuil	Définit le nombre maximum de chemins pour la liste des fichiers côté pilote avec des sources basées sur des fichiers (Parquet, JSON et ORC). En cas de dépassement lors de la découverte de partitions, les fichiers sont répertoriés à l'aide d'une tâche distribuée Spark distincte.	32
spark.sql.statistics.histogram.enabled	Spécifie s'il faut générer des histogrammes à hauteur égale lors du calcul des statistiques des colonnes afin d'améliorer la précision de l'estimation. Nécessite une analyse de table supplémentaire en plus de celle requise pour les statistiques de base sur les colonnes.	FALSE

9. Cliquez sur Exécuter.

 Note

Vous ne pouvez pas exécuter la requête si le membre qui peut recevoir les résultats n'a pas configuré les paramètres des résultats de la requête.

10. Continuez à ajuster les paramètres et réexécutez votre requête, ou cliquez sur le bouton + pour démarrer une nouvelle requête dans un nouvel onglet.

Interrogation à l'aide du générateur d'analyse

Vous pouvez utiliser le générateur d'analyse pour créer des requêtes sans avoir à écrire de code SQL. Avec le générateur d'analyse, vous pouvez créer une requête pour une collaboration qui possède :

- Une table unique qui utilise la [règle d'analyse d'agrégation](#) sans qu'aucun JOIN ne soit requis
- Deux tables (une pour chaque membre) qui utilisent toutes deux la [règle d'analyse d'agrégation](#)
- Deux tables (une pour chaque membre) qui utilisent toutes deux la [règle d'analyse de liste](#)
- Deux tables (une pour chaque membre) qui utilisent toutes deux la règle d'analyse d'agrégation et deux tables (une pour chaque membre) qui utilisent toutes deux la règle d'analyse de liste

Si vous souhaitez écrire des requêtes SQL manuellement, consultez [Interrogation de tables configurées à l'aide de l'éditeur de code SQL](#).

Le générateur d'analyse apparaît sous forme d'option d'interface utilisateur du générateur d'analyse dans la section Analyse de l'onglet Analyse de la AWS Clean Rooms console.

Important

Si vous activez l'interface utilisateur du générateur d'analyse, que vous commencez à créer une requête dans le générateur d'analyse, puis que vous désactivez l'interface utilisateur du générateur d'analyse, votre requête n'est pas enregistrée.

Tip

Si une maintenance planifiée a lieu pendant l'exécution d'une requête, celle-ci est interrompue et annulée. Vous devez relancer la requête.

Les rubriques suivantes expliquent comment utiliser le générateur d'analyse.

Rubriques

- [Utiliser le générateur d'analyse pour interroger une seule table \(agrégation\)](#)
- [Utilisez le générateur d'analyse pour interroger deux tables \(agrégation ou liste\)](#)

Utiliser le générateur d'analyse pour interroger une seule table (agrégation)

Cette procédure explique comment utiliser l'interface utilisateur du générateur d'analyse dans la AWS Clean Rooms console pour créer une requête. La requête concerne une collaboration comportant une seule table qui utilise la [règle d'analyse d'agrégation](#) sans qu'elle soit JOIN requise.

Pour utiliser le générateur d'analyse pour interroger une seule table

Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.

1. Dans le volet de navigation de gauche, sélectionnez Collaborations.
2. Choisissez la collaboration dont le statut de capacité de vos membres est défini sur Exécuter des requêtes.
3. Dans l'onglet Analyse, sous Tables, affichez le tableau et le type de règle d'analyse associé. (Le type de règle d'analyse doit être la règle d'analyse d'agrégation.)

Note

Si le tableau que vous attendez ne s'affiche pas, c'est peut-être pour les raisons suivantes :

- La table n'a pas été [associée](#).
- Aucune [règle d'analyse n'est configurée](#) dans le tableau.

4. Dans la section Analyse, activez l'interface utilisateur du générateur d'analyse.
5. Créez une requête.

Si vous souhaitez voir toutes les mesures d'agrégation, passez à l'étape 9.

- a. Pour Choose metrics, passez en revue les métriques agrégées qui ont été présélectionnées par défaut et supprimez toute métrique si nécessaire.
- b. (Facultatif) Pour Ajouter des segments — facultatif, choisissez un ou plusieurs paramètres.

Note

Ajouter des segments : cette option n'est affichée que si des dimensions sont spécifiées pour le tableau.

- c. (Facultatif) Pour Ajouter des filtres : facultatif, choisissez Ajouter un filtre, puis choisissez un paramètre, un opérateur et une valeur.

Pour ajouter d'autres filtres, choisissez Ajouter un autre filtre.

Pour supprimer un filtre, choisissez Supprimer.

 Note

ORDER BY n'est pas pris en charge pour les requêtes d'agrégation.
Seul l'AND opérateur est pris en charge dans les filtres.

- d. (Facultatif) Pour Ajouter une description — facultatif, entrez une description pour aider à identifier la requête dans la liste des requêtes.
6. Développez le code SQL de prévisualisation.
 - a. Affichez le code SQL généré à partir du générateur d'analyse.
 - b. Pour copier le code SQL, choisissez Copier.
 - c. Pour modifier le code SQL, choisissez Modifier dans l'éditeur de code SQL.
 7. Cliquez sur Exécuter.

 Note

Vous ne pouvez pas exécuter la requête si le membre qui peut recevoir les résultats n'a pas configuré les paramètres des résultats de la requête.

8. Continuez à ajuster les paramètres et réexécutez votre requête, ou cliquez sur le bouton + pour démarrer une nouvelle requête dans un nouvel onglet.

 Note

AWS Clean Rooms vise à fournir un message d'erreur clair. Si un message d'erreur ne contient pas suffisamment de détails pour vous aider à résoudre le problème, contactez l'équipe chargée du compte. Fournissez-leur une description de la façon dont l'erreur s'est produite et du message d'erreur (y compris les éventuels identifiants). Pour de plus amples informations, veuillez consulter [Résolution des problèmes AWS Clean Rooms](#).

Utilisez le générateur d'analyse pour interroger deux tables (agrégation ou liste)

Cette procédure décrit comment utiliser le générateur d'analyse de la AWS Clean Rooms console pour créer une requête pour une collaboration qui possède :

- Deux tables (une pour chaque membre) qui utilisent toutes deux la [règle d'analyse d'agrégation](#)
- Deux tables (une pour chaque membre) qui utilisent toutes deux la [règle d'analyse de liste](#)
- Deux tables (une pour chaque membre) qui utilisent toutes deux la règle d'analyse d'agrégation et deux tables (une pour chaque membre) qui utilisent toutes deux la règle d'analyse de liste

Pour utiliser le générateur d'analyse pour interroger deux tables

Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.

1. Dans le volet de navigation de gauche, sélectionnez Collaborations.
2. Choisissez la collaboration dont les capacités de vos membres ont le statut Exécuter des requêtes.
3. Dans l'onglet Analyse, sous Tables, visualisez les deux tables et le type de règle d'analyse associé (règle d'analyse d'agrégation ou règle d'analyse de liste).

Note

Si les tables attendues ne figurent pas dans la liste, c'est peut-être pour les raisons suivantes :

- Les tables n'ont pas été [associées](#).
- Aucune [règle d'analyse n'est configurée pour](#) les tables.

4. Dans la section Analyse, activez l'interface utilisateur du générateur d'analyse.
5. Créez une requête.

Si la collaboration contient deux tables qui utilisent la règle d'analyse d'agrégation et deux tables qui utilisent la règle d'analyse de liste, choisissez d'abord Agrégation ou Liste, puis suivez les instructions en fonction de la règle d'analyse sélectionnée.

Si les deux tables utilisent la règle d'analyse d'agrégation

1. Pour Choose metrics, passez en revue les métriques agrégées qui ont été présélectionnées par défaut et supprimez toute métrique si nécessaire.
2. Pour Match records, sélectionnez un ou plusieurs enregistrements.

Note

Lorsque vous utilisez le générateur d'analyse, vous ne pouvez effectuer de correspondance que sur une seule paire de colonnes.

3. (Facultatif) Pour Ajouter des segments — facultatif, choisissez un ou plusieurs paramètres.

Note

Ajouter des segments : cette option n'est affichée que si des dimensions sont spécifiées pour le tableau.

4. (Facultatif) Pour Ajouter des filtres : facultatif, choisissez Ajouter un filtre, puis choisissez un paramètre, un opérateur et une valeur.

Pour ajouter d'autres filtres, choisissez Ajouter un autre filtre.

Si les deux tables utilisent la règle d'analyse de liste

1. Pour Choisir les attributs, passez en revue les attributs de liste qui ont été présélectionnés par défaut et supprimez toute métrique si nécessaire.
2. Pour Match records, sélectionnez un ou plusieurs enregistrements.

Note

Lorsque vous utilisez le générateur d'analyse, vous ne pouvez effectuer de correspondance que sur une seule paire de colonnes.

3. (Facultatif) Pour Ajouter des filtres : facultatif, choisissez Ajouter un filtre, puis choisissez un paramètre, un opérateur et une valeur.

Pour ajouter d'autres filtres, choisissez Ajouter un autre filtre.

Pour supprimer un filtre, choisissez Supprimer.

Note

LIMIT n'est pas pris en charge pour les requêtes de liste.
Seul l'AND opérateur est pris en charge dans les filtres.

4. (Facultatif) Pour Ajouter une description — facultatif, entrez une description pour

Si les deux tables utilisent la règle d'analyse d'agrégation

Pour supprimer un filtre, choisissez Supprimer.

Note

ORDER BY n'est pas pris en charge pour les requêtes d'agrégation.

Seul l'AND opérateur est pris en charge dans les filtres.

Si les deux tables utilisent la règle d'analyse de liste

aider à identifier la requête dans la liste des requêtes récentes.

5. (Facultatif) Pour Ajouter une description — facultatif, entrez une description pour aider à identifier la requête dans la liste des requêtes récentes.

6. Développez le code SQL de prévisualisation.
 - a. Affichez le code SQL généré à partir du générateur d'analyse.
 - b. Pour copier le code SQL, choisissez Copier.
 - c. Pour modifier le code SQL, choisissez Modifier dans l'éditeur de code SQL.
7. Cliquez sur Exécuter.

Note

Vous ne pouvez pas exécuter la requête si le membre qui peut recevoir les résultats n'a pas configuré les paramètres des résultats de la requête

8. Continuez à ajuster les paramètres et réexécutez votre requête, ou cliquez sur le bouton + pour démarrer une nouvelle requête dans un nouvel onglet.

 Note

AWS Clean Rooms vise à fournir un message d'erreur clair. Si un message d'erreur ne contient pas suffisamment de détails pour vous aider à résoudre le problème, contactez l'équipe chargée du compte. Fournissez-leur une description de la façon dont l'erreur s'est produite et du message d'erreur (y compris les éventuels identifiants). Pour de plus amples informations, veuillez consulter [Résolution des problèmes AWS Clean Rooms](#).

Visualisation de l'impact de la confidentialité différentielle

En général, l'écriture et l'exécution de requêtes ne changent pas lorsque la confidentialité différentielle est activée. Toutefois, vous ne pouvez pas exécuter de requête s'il ne reste pas suffisamment de budget de confidentialité. Au fur et à mesure que vous exécutez des requêtes et que vous consommez le budget de confidentialité, vous pouvez voir approximativement le nombre d'agrégations que vous pouvez exécuter et l'impact que cela pourrait avoir sur les requêtes futures.

Pour voir l'impact de la confidentialité différentielle dans une collaboration

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut « Vos informations de membre » est « Exécuter des requêtes ».
4. Dans l'onglet Analyse, sous Tableaux, consultez le budget de confidentialité restant. Ceci est affiché sous la forme du nombre estimé de fonctions d'agrégation restantes et de l'utilitaire utilisé (rendu sous forme de pourcentage).

 Note

Le nombre estimé de fonctions d'agrégation restantes et le pourcentage de l'utilitaire utilisé ne s'affichent que pour le membre autorisé à effectuer des requêtes.

5. Choisissez Afficher l'impact pour voir le niveau de bruit injecté dans les résultats et le nombre approximatif de fonctions d'agrégation que vous pouvez exécuter.

Affichage des requêtes récentes

Vous pouvez consulter les requêtes exécutées au cours des 90 derniers jours dans l'onglet Analyse.

Note

Si votre seule capacité de membre concerne les données Contribute et que vous n'êtes pas le [membre qui paie les coûts de calcul des requêtes](#), l'onglet Analyse n'apparaît pas sur la console.

Pour consulter les requêtes récentes

Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.

1. Dans le volet de navigation de gauche, sélectionnez Collaborations.
2. Choisissez une collaboration.
3. Dans l'onglet Analyse, sous Analyses, sélectionnez Toutes les requêtes dans le menu déroulant et consultez les requêtes exécutées au cours des 90 derniers jours.
4. Pour trier les requêtes récentes par statut, sélectionnez un statut dans la liste déroulante Tous les statuts.

Les statuts sont les suivants : Soumis, Commencé, Annulé, Réussite, Échec et Expéré.

Affichage des détails de la requête

Vous pouvez consulter les détails de la requête en tant que membre habilité à exécuter des requêtes ou en tant que membre habilité à recevoir les résultats.

Pour afficher les détails de la requête

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez une collaboration.
4. Dans l'onglet Analyse, effectuez l'une des opérations suivantes :

- Cliquez sur le bouton d'option correspondant à la requête spécifique que vous souhaitez consulter, puis sélectionnez Afficher les détails.
- Choisissez l'ID de requête protégé.

5. Sur la page Détails de la requête,

- Si vous êtes le membre autorisé à exécuter des requêtes, consultez les détails de la requête, le texte SQL et les résultats.

Un message s'affiche pour confirmer que les résultats de la requête ont été transmis au membre autorisé à recevoir les résultats.

- Si vous êtes le membre autorisé à recevoir les résultats, consultez les détails de la requête et les résultats.

Exécution de PySpark tâches

En tant que [membre habilité à effectuer une requête](#), vous pouvez exécuter une PySpark tâche sur une table configurée à l'aide d'un [modèle d' PySpark analyse](#) approuvé.

Conditions préalables

Avant d'exécuter une PySpark tâche, vous devez disposer des éléments suivants :

- Une adhésion active à la AWS Clean Rooms collaboration
- Accès à au moins un modèle d'analyse dans la collaboration
- Accès à au moins une table configurée dans la collaboration
- Autorisations permettant d'écrire les résultats d'une PySpark tâche dans un compartiment S3 spécifié

Pour plus d'informations sur la création du rôle de service requis, consultez [Création d'un rôle de service pour écrire les résultats d'une PySpark tâche](#).

- Le membre chargé de payer les frais de calcul a rejoint la collaboration en tant que membre actif

Pour plus d'informations sur la façon d'interroger des données ou d'afficher des requêtes en appelant directement l'opération d' AWS Clean Rooms StartProtectedJobAPI ou en utilisant le AWS SDKs, consultez la [référence de l'AWS Clean Rooms API](#).

Pour plus d'informations sur la journalisation des tâches, consultez [Connexion à une analyse AWS Clean Rooms](#).

Pour plus d'informations sur la réception des résultats d'une tâche, consultez [Réception et utilisation des résultats d'analyse](#).

Les rubriques suivantes expliquent comment exécuter une PySpark tâche sur une table configurée dans le cadre d'une collaboration à l'aide de la AWS Clean Rooms console.

Rubriques

- [Exécution d'une PySpark tâche sur une table configurée à l'aide d'un modèle d' PySpark analyse](#)
- [Afficher les offres d'emploi récentes](#)
- [Affichage des détails de la tâche](#)

Exécution d'une PySpark tâche sur une table configurée à l'aide d'un modèle d' PySpark analyse

Cette procédure explique comment utiliser un modèle d' PySpark analyse dans la AWS Clean Rooms console pour analyser des tables configurées à l'aide de la règle d'analyse personnalisée.

Pour exécuter une PySpark tâche sur une table configurée à l'aide d'un modèle d' PySpark analyse

Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.

1. Dans le volet de navigation de gauche, sélectionnez Collaborations.
2. Choisissez la collaboration dont le statut de compétences de vos membres est Run jobs.
3. Dans l'onglet Analyse, sous la section Tables, visualisez les tables et le type de règle d'analyse associé (règle d'analyse personnalisée).

Note

Si les tables attendues ne figurent pas dans la liste, c'est peut-être pour les raisons suivantes :

- Les tables n'ont pas été [associées](#).
- Aucune [règle d'analyse n'est configurée pour](#) les tables.

4. Dans la section Analyse, pour le mode Analyse, sélectionnez Exécuter les modèles d'analyse.
5. Choisissez le modèle PySpark d'analyse dans la liste déroulante des modèles d'analyse.

Les paramètres du modèle PySpark d'analyse seront automatiquement renseignés dans la définition.

6. Si des paramètres sont définis dans le modèle d'analyse, sous Paramètres, fournissez les valeurs des paramètres :
 - a. Pour chaque paramètre, consultez le nom du paramètre et la valeur par défaut (si configuré).
 - b. Entrez une valeur pour chaque paramètre que vous souhaitez remplacer.

 Note

Si vous ne fournissez aucune valeur mais qu'une valeur par défaut existe, la valeur par défaut sera utilisée.

 Important

Les valeurs des paramètres peuvent comporter jusqu'à 1 000 caractères et sont compatibles avec le codage UTF-8. Toutes les valeurs de paramètres sont traitées comme des chaînes et transmises à votre script utilisateur via l'objet de contexte. Assurez-vous que votre script utilisateur valide et gère les valeurs des paramètres en toute sécurité. Pour plus d'informations sur la gestion sécurisée des paramètres, consultez [Utilisation des paramètres dans les modèles PySpark d'analyse](#).

7. Spécifiez le type de travailleur pris en charge et le nombre de travailleurs.

Utilisez le tableau suivant pour déterminer le type et le nombre de travailleurs dont vous avez besoin pour votre cas d'utilisation.

Type d'employé	vCPU	Mémoire (Go)	Stockage (Go)	Nombre d'employés	Nombre total d'unités de traitement pour salles blanches (CRPU)
CR.1X (par défaut)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

Les différents types de travailleurs et le nombre de travailleurs entraînent des coûts associés. Pour en savoir plus sur les tarifs, consultez la section [AWS Clean Rooms Tarifs](#).

8. Cliquez sur Exécuter.

 Note

Vous ne pouvez pas exécuter le travail si le membre qui peut recevoir les résultats n'a pas configuré les paramètres des résultats du travail.

9. Continuez à ajuster les paramètres et réexécutez votre tâche, ou cliquez sur le bouton + pour démarrer une nouvelle tâche dans un nouvel onglet.

Afficher les offres d'emploi récentes

Vous pouvez consulter les tâches exécutées au cours des 90 derniers jours dans l'onglet Analyse.

 Note

Si votre seule capacité de membre concerne les données Contribute et que vous n'êtes pas le [membre qui paie les coûts de calcul des tâches](#), l'onglet Analyse n'apparaît pas sur la console.

Pour consulter les offres d'emploi récentes

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez une collaboration.
4. Dans l'onglet Analyse, sous Analyses, sélectionnez Toutes les tâches dans la liste déroulante et consultez les tâches exécutées au cours des 90 derniers jours.
5. Pour trier les offres d'emploi récentes par statut, sélectionnez un statut dans la liste déroulante Tous les statuts.

Les statuts sont les suivants : Soumis, Commencé, Annulé, Réussite, Échec et Expéré.

Affichage des détails de la tâche

Vous pouvez consulter les détails des tâches en tant que membre habilité à exécuter des tâches ou en tant que membre habilité à recevoir les résultats.

Pour consulter les détails de la tâche

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez une collaboration.
4. Dans l'onglet Analyse, sous Analyses, sélectionnez Toutes les tâches dans le menu déroulant, puis effectuez l'une des opérations suivantes :
 - Cliquez sur le bouton d'option correspondant à la tâche spécifique que vous souhaitez consulter, puis sélectionnez Afficher les détails.

- Choisissez l'ID de tâche protégé.

5. Sur la page des détails du Job,

- Si vous êtes le membre autorisé à exécuter des tâches, consultez les détails de la tâche, la tâche et les résultats.

Un message s'affiche pour confirmer que les résultats du travail ont été transmis au membre autorisé à recevoir les résultats.

- Si vous êtes le membre autorisé à recevoir les résultats, consultez les détails du Job et les résultats.

Réception et utilisation des résultats d'analyse

[Les membres autorisés à recevoir des résultats](#) peuvent consulter les résultats de la requête dans la AWS Clean Rooms console ou dans les compartiments Amazon S3 qu'ils ont spécifiés lorsqu'ils ont rejoint la collaboration.

Note

[Pour les tables de données chiffrées uniquement, le membre qui peut recevoir les résultats déchiffre les résultats de la requête en exécutant le client de chiffrement C3R en mode déchiffrement.](#)

La destination des résultats dans Amazon S3 ne peut pas se trouver dans le même compartiment S3 que n'importe quelle source de données.

Les rubriques suivantes expliquent comment recevoir des résultats d'analyse à l'aide de la AWS Clean Rooms console.

Rubriques

- [Réception des résultats de requêtes](#)
- [Recevoir les résultats d'un travail](#)
- [Modification des valeurs par défaut pour les paramètres des résultats de requête](#)
- [Modification des valeurs par défaut pour les paramètres des résultats des tâches](#)
- [Utilisation du résultat de la requête dans d'autres Services AWS](#)

Pour plus d'informations sur la façon d'interroger des données ou d'afficher des requêtes en appelant directement l' AWS Clean Rooms API ou en utilisant le AWS SDKs, consultez la [référence de l'AWS Clean Rooms API](#).

Pour plus d'informations sur la journalisation des requêtes, consultez [Connexion à une analyse AWS Clean Rooms](#).

Note

Si vous exécutez une requête sur des tables de données chiffrées, les résultats des colonnes chiffrées sont chiffrés.

Réception des résultats de requêtes

Note

La destination des résultats dans Amazon S3 ne peut pas se trouver dans le même compartiment S3 que n'importe quelle source de données.

Les résultats de la requête se trouvent dans la section Paramètres par défaut des résultats de l'onglet Analyse de la AWS Clean Rooms console.

Pour recevoir les résultats d'une requête

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est de recevoir des résultats.
4. Pour recevoir les résultats de la requête directement depuis AWS Clean Rooms, dans l'onglet Analyse, sous Analyses, sélectionnez Toutes les requêtes dans le menu déroulant, puis dans la colonne ID de requête protégée, sélectionnez la requête.
5. Sur la page Détails de la requête, sous Résultats, effectuez l'une des opérations suivantes :

Si vous souhaitez	Choisissez ensuite...
Copiez les résultats.	Copy
Téléchargez les résultats.	Download
	Note Par défaut, le nom du fichier téléchargé est celui Query id qui était affiché lors de l'exécution de la requête. AWS Clean Rooms
Consultez les résultats dans Amazon S3.	Afficher dans Amazon S3

Si vous souhaitez

Choisissez ensuite...

La console Amazon S3 s'ouvre dans un onglet séparé.

6. Si vous utilisez des données chiffrées, vous pouvez désormais [déchiffrer](#) les tables de données.

Pour de plus amples informations, veuillez consulter [Déchiffrer des tables de données avec le client de chiffrement C3R](#).

Recevoir les résultats d'un travail

Note

La destination des résultats dans Amazon S3 ne peut pas se trouver dans le même compartiment S3 que n'importe quelle source de données.

Les résultats de la tâche se trouvent dans la section Paramètres des résultats par défaut de l'onglet Analyse de la AWS Clean Rooms console.

Pour recevoir les résultats d'une offre d'emploi

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est de recevoir des résultats.
4. Pour recevoir les résultats du travail directement depuis AWS Clean Rooms, dans l'onglet Analyse, sous Analyses, sélectionnez Tous les travaux dans le menu déroulant, puis sous la colonne ID du travail protégé, sélectionnez le travail.
5. Sur la page Détails du Job, sous Résultats, copiez l'ID du Job.

Retournez à l'onglet Analyse et développez les paramètres de résultat par défaut.

Sous Destination des résultats, sélectionnez le lien pour afficher les résultats dans Amazon S3.

La console Amazon S3 s'ouvre dans un onglet séparé.

Dans Amazon S3, collez le Job ID dans la barre de recherche et appuyez sur Entrée.

Le dossier contenant les résultats apparaît. Sélectionnez le dossier pour afficher les résultats du travail.

Modification des valeurs par défaut pour les paramètres des résultats de requête

Note

La destination des résultats dans Amazon S3 ne peut pas se trouver dans le même compartiment S3 que n'importe quelle source de données.

En tant que membre habilité à recevoir des résultats, vous pouvez modifier les valeurs par défaut des paramètres des résultats de requête dans la AWS Clean Rooms console.

Pour modifier les valeurs par défaut des paramètres des résultats de requête

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est de recevoir des résultats.
4. Dans l'onglet Analyse, sous Paramètres par défaut des résultats, sélectionnez Modifier.
5. Sur la page Modifier les paramètres par défaut des résultats, modifiez les éléments suivants, selon vos besoins :
 - a. Sous Résultats de la requête, modifiez la destination des résultats dans Amazon S3, le format des résultats ou les fichiers de résultats.
 - b. (Facultatif) Pour accéder au service, si vous souhaitez envoyer des requêtes qui prennent jusqu'à 24 heures à votre destination S3, cochez la case Ajouter un rôle de service pour prendre en charge les requêtes dont le traitement prend jusqu'à 24 heures.

Les requêtes volumineuses dont le traitement prend jusqu'à 24 heures seront livrées à votre destination S3.

Si vous ne cochez pas cette case, seules les requêtes traitées dans les 12 heures seront livrées à votre agence S3.

- Spécifiez les autorisations d'accès au service en sélectionnant Créer et utiliser un nouveau rôle de service ou Utiliser un rôle de service existant.

Create and use a new service role

- AWS Clean Rooms crée un rôle de service avec la politique requise pour cette table.
- Le nom du rôle de service par défaut est `cleanrooms-query-receiver-
<timestamp>`
- Vous devez disposer des autorisations nécessaires pour créer des rôles et associer des politiques.

Use an existing service role

1. Choisissez un nom de rôle de service existant dans la liste déroulante.

La liste des rôles s'affiche si vous êtes autorisé à répertorier les rôles.

Si vous n'êtes pas autorisé à répertorier les rôles, vous pouvez saisir le nom de ressource Amazon (ARN) du rôle que vous souhaitez utiliser.

2. Affichez le rôle de service en choisissant le lien externe Afficher dans IAM.

S'il n'existe aucun rôle de service existant, l'option Utiliser un rôle de service existant n'est pas disponible.

Par défaut, AWS Clean Rooms ne tente pas de mettre à jour la politique de rôle existante pour ajouter les autorisations nécessaires.

Note

- AWS Clean Rooms nécessite des autorisations pour effectuer des requêtes conformément aux règles d'analyse. Pour plus d'informations sur les autorisations pour AWS Clean Rooms, voir [AWS politiques gérées pour AWS Clean Rooms](#).

- Si le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms, vous recevez un message d'erreur indiquant que le rôle ne dispose pas d'autorisations suffisantes pour AWS Clean Rooms. La politique de rôle doit être ajoutée avant de continuer.
- Si vous ne parvenez pas à modifier la politique de rôle, vous recevez un message d'erreur indiquant que AWS Clean Rooms la politique pour le rôle de service est introuvable.

6. Sélectionnez Enregistrer les modifications.
7. Les paramètres de résultats de requête mis à jour apparaissent sur la page détaillée de la collaboration.

Modification des valeurs par défaut pour les paramètres des résultats des tâches

Note

La destination des résultats dans Amazon S3 ne peut pas se trouver dans le même compartiment S3 que n'importe quelle source de données.

En tant que membre habilité à recevoir les résultats, vous pouvez modifier les valeurs par défaut des paramètres des résultats des tâches dans la AWS Clean Rooms console.

Pour modifier les valeurs par défaut des paramètres des résultats des tâches

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Choisissez la collaboration dont le statut de capacité de vos membres est de recevoir des résultats.
4. Dans l'onglet Analyse, sous Paramètres par défaut des résultats, sélectionnez Modifier.
5. Sur la page Modifier les paramètres par défaut des résultats, modifiez les éléments suivants, selon vos besoins :

- a. Sous Résultats du Job, modifiez la destination des résultats dans Amazon S3.
 - b. Sous Accès au service, modifiez le nom du rôle de service existant.
6. Sélectionnez Enregistrer les modifications.
7. Les paramètres de résultats de Job mis à jour apparaissent sur la page détaillée de la collaboration.

Utilisation du résultat de la requête dans d'autres Services AWS

La sortie de requête SQL peut être utilisée pour les données de départ d'un modèle Clean Rooms ML. Pour plus d'informations, consultez [AWS Clean Rooms ML](#).

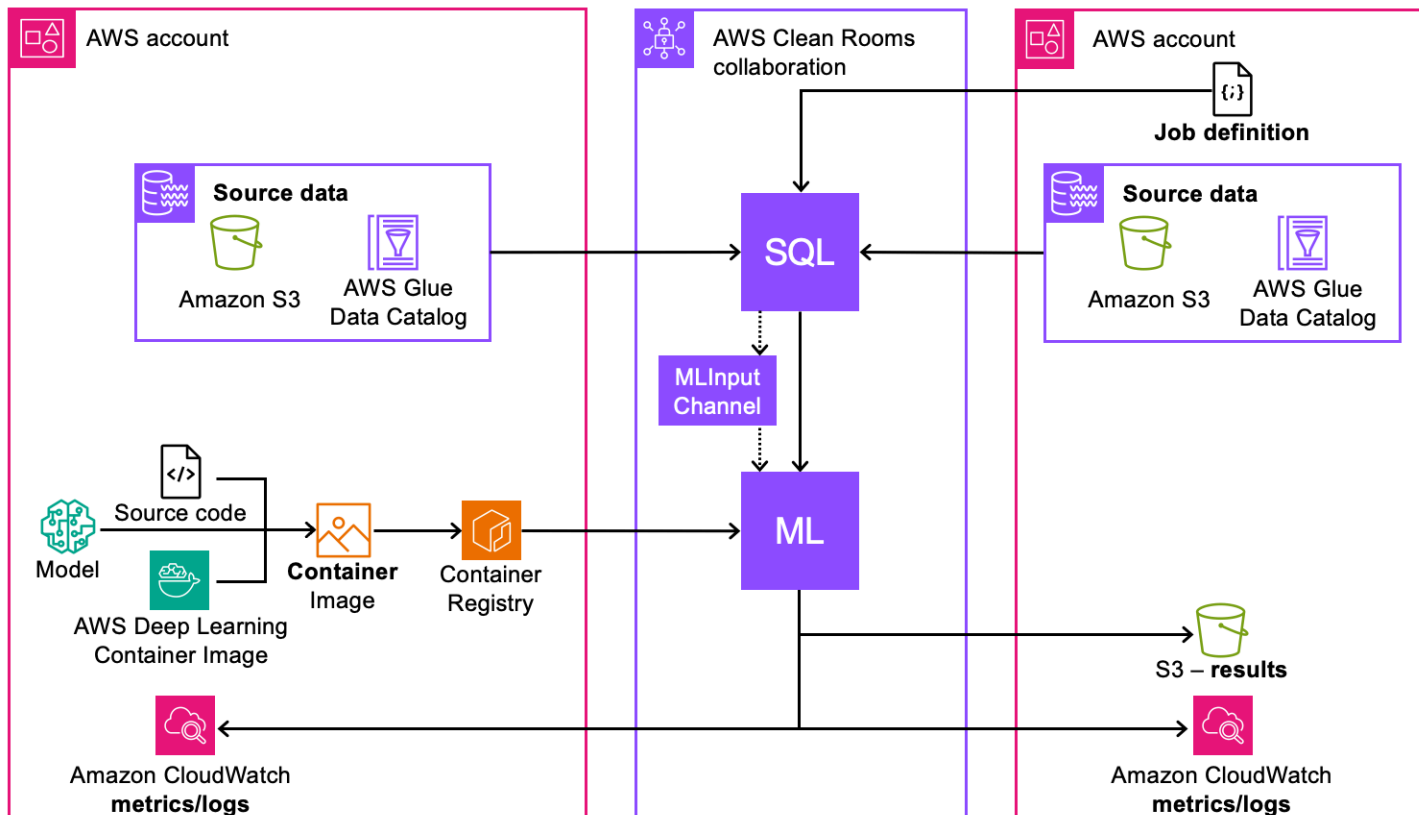
Le résultat de la requête AWS Clean Rooms est disponible sur la console (si la console est utilisée pour exécuter des requêtes) et est téléchargé dans un compartiment Amazon S3 spécifié. À partir de là, vous pouvez utiliser le résultat de la requête dans d'autres services Services AWS, tels qu'Amazon Quick et Amazon SageMaker AI, en fonction de la manière dont ces services utilisent les données d'Amazon S3.

Pour plus d'informations sur Amazon Quick, consultez la [documentation Amazon Quick](#).

Pour plus d'informations sur Amazon SageMaker AI, consultez la [documentation Amazon SageMaker AI](#).

Modélisation personnalisée dans AWS Clean Rooms ML

D'un point de vue technique, le schéma suivant décrit le fonctionnement de la modélisation ML personnalisée dans AWS Clean Rooms ML.



Voici comment fonctionne la modélisation ML personnalisée dans Clean Rooms ML :

1. Configuration de la source de données

- Les données sources peuvent être stockées dans le catalogue Amazon S3, dans le AWS Glue Data Catalog ou dans Snowflake
- AWS Glue Data Catalog est utilisé pour organiser et cataloguer
- Les données provenant de plusieurs Comptes AWS sources peuvent être utilisées dans le cadre d'une même collaboration

2. Requête SQL et traitement des données

- Les requêtes SQL sont utilisées pour accéder aux données sources et les traiter
- Les requêtes s'exécutent dans les limites de la AWS Clean Rooms collaboration
- Les données traitées sont transmises aux canaux d'entrée ML pour l'entraînement des modèles

3. Développement de modèles ML

- Le code source du modèle peut être développé à l'aide de AWS Deep Learning Container Images
- Les images de conteneur personnalisées doivent être créées et stockées dans Amazon Elastic Container Registry

4. Composantes de l'infrastructure

- Amazon Elastic Container Registry stocke et gère les conteneurs du modèle ML
- Le traitement du ML s'effectue dans un environnement de AWS Clean Rooms collaboration sécurisé

5. Surveillance et journalisation

- Amazon CloudWatch fournit des statistiques et des journaux aux deux parties collaboratrices
- Le suivi est disponible pour toutes les Comptes AWS personnes impliquées dans la collaboration
- Les indicateurs de performance et les journaux opérationnels sont accessibles aux parties concernées

6. Gestion des résultats

- L'accès aux résultats est contrôlé en fonction des autorisations de collaboration

Avant de commencer, consultez le [Prérequis pour la modélisation ML personnalisée](#) et [Consignes de création de modèles pour le conteneur de formation](#) pour plus d'informations.

Génération de jeux de données synthétiques garantissant une meilleure confidentialité

Un jeu de données synthétique possède des propriétés statistiques similaires à celles du jeu de données d'origine sur lequel il est basé, mais ne contient pas les observations du monde réel présentes dans le jeu de données d'origine. En utilisant des ensembles de données synthétiques optimisés pour la confidentialité, vous pouvez découvrir de nouveaux cas d'utilisation de modèles d'apprentissage automatique (ML) que les problèmes de confidentialité des données évitaient auparavant. Lorsque vous créez un canal d'entrée ML, vous pouvez générer des données synthétiques pour protéger les informations sensibles lors de l'entraînement des modèles de ML.

Lorsque vous créez un modèle avec des données synthétiques, vous devez :

- Exiger que la sortie du modèle soit synthétique
- Classer les colonnes du schéma de sortie en tant que colonnes numériques ou catégoriques
- Personnalisez les données synthétiques en fonction des besoins de l'organisation
- Ajustez les paramètres de confidentialité :
 - Définir le niveau de confidentialité (epsilon)
 - Configurer le seuil de confidentialité

Warning

La génération de données synthétiques empêche de déduire des attributs individuels, que des individus spécifiques soient présents dans l'ensemble de données d'origine ou que des attributs d'apprentissage de ces individus soient présents. Cependant, cela n'empêche pas les valeurs littérales de l'ensemble de données d'origine, y compris les informations personnelles identifiables (PII), d'apparaître dans l'ensemble de données synthétique. Nous recommandons d'éviter dans le jeu de données d'entrée les valeurs associées à une seule personne concernée, car elles peuvent permettre de réidentifier une personne concernée. Par exemple, si un seul utilisateur vit dans un code postal, la présence de ce code postal dans le jeu de données synthétique confirmera que cet utilisateur figurait dans le jeu de données d'origine. Des techniques telles que la troncation de valeurs de haute précision ou le remplacement de catalogues peu courants par d'autres peuvent être utilisées pour atténuer ce risque. Ces transformations peuvent faire partie de la requête utilisée pour créer le canal d'entrée ML.

Pour plus d'informations sur la façon de générer des données synthétiques pour l'entraînement de modèles personnalisés, consultez [Création d'un modèle d'analyse SQL](#).

Les modèles d'analyse avec des sorties synthétiques ne peuvent être utilisés que pour créer des canaux d'entrée ML. Pour de plus amples informations, veuillez consulter [Création d'un canal d'entrée ML dans AWS Clean Rooms ML](#).

Considérations relatives à la génération de données synthétiques

Avec AWS Clean Rooms ML, les membres de la collaboration peuvent créer un ensemble de données synthétique qui désidentifie de manière irréversible les sujets du jeu de données d'origine de leurs ensembles de données collectifs afin de créer un modèle d'apprentissage automatique

personnalisé. Lors de la création de la collaboration, vous devez configurer les informations de paiement pour spécifier qui paie pour la génération de données synthétiques. Voici les étapes de haut niveau pour générer un ensemble de données synthétique et entraîner un modèle d'apprentissage automatique personnalisé :

1. Un membre de la collaboration crée un modèle d'analyse qui inclut :
 - Le code SQL nécessaire pour définir l'ensemble de données à synthétiser.
 - Configurations liées à la confidentialité utilisées pour garantir que les données synthétiques répondent aux exigences de conformité des fournisseurs de données.
2. Une fois que tous les fournisseurs de données ont approuvé le modèle d'analyse, le lanceur de requêtes collaboratif crée un canal d'entrée d'apprentissage automatique (ML) à l'aide du modèle.
3. Clean Rooms ML génère le jeu de données synthétique et vérifie qu'il respecte les seuils de confidentialité spécifiés dans le modèle d'analyse.
4. Si tous les seuils sont satisfaits, le canal d'entrée ML est rempli avec le jeu de données synthétique.
5. Les clients peuvent ensuite utiliser ce canal de saisie ML pour entraîner le modèle de ML personnalisé associé à la collaboration.

Considérations importantes :

- Les données synthétiques générées dans Clean Rooms ML ne suppriment, ne suppriment, ne masquent ni ne nettoient aucune valeur individuelle, y compris les informations personnelles identifiables (PII) trouvées dans le jeu de données d'origine. Le jeu de données synthétique est généré en échantillonnant des valeurs, mais pas des enregistrements complets, à partir du jeu de données d'origine.
- Si le jeu de données d'origine contient des lignes similaires, il est possible que les données synthétiques contiennent des lignes identiques à celles du jeu de données d'origine.

Préparation du jeu de données :

- Évitez les colonnes dont la distribution des classes est significativement déséquilibrée. Ceci est particulièrement important pour la valeur prédite ou la colonne « Y ». Les déséquilibres extrêmes réduisent la confidentialité globale de l'ensemble de données synthétique.

- Clean Rooms ML ne prend pas en charge la génération de données synthétiques à partir de séries chronologiques pour lesquelles il est important de maintenir des corrélations entre les enregistrements séquentiels.
- Clean Rooms ML ne prend pas en charge la génération de données synthétiques à partir de texte ou de données non structurées.
- Les types de données suivants sont pris en charge :

Nom

du

type

de

données

BIGINT

BOOLEAN

CHAR

DATE

DECIMAL

FLOAT

INTEGER

LONG

REAL

SHORT

SMALLINT

TIME

TIMESTAMP

_LTZ

Nom
du
type
de
données

TIMESTAMP
_NTZ

TINYINT

VARCHAR

Limites:

- Pour la génération de données synthétiques, le nombre maximum de colonnes prédictives est d'une.
- Si la colonne cible est catégorique, le nombre maximum de catégories dans le jeu de données d'origine est de 100.
- Dans le jeu de données d'origine, le nombre de lignes doit être compris entre 1 500 et 2,5 millions et le nombre maximum de colonnes est de 1 000. Pour les valeurs non nulles de la colonne cible, le nombre minimum de lignes est de 1 000.

Indicateurs de confidentialité :

- Clean Rooms ML fournit un score de confidentialité qui mesure le niveau de protection des données synthétiques générées contre les attaques par inférence d'adhésion (MIAs). Le service utilise 5 % des données originales issues du processus de synthèse pour calculer ce score.
- Des scores proches de 50 % sont considérés comme bons ; des scores élevés indiquent une protection moindre contre MIAs. Les scores significativement inférieurs à 50 % sont rares et peuvent être dus à la non-représentation des modèles issus des données d'origine dans les données synthétisées.

Modèle personnalisé en aval :

- Les données synthétiques générées dans Clean Rooms ML conviennent parfaitement à l'entraînement de modèles de classification binaires et de modèles de classification à classes multiples comportant jusqu'à cinq classes.
- L'entraînement de modèles de régression utilisant des données synthétiques générées dans Clean Rooms ML peut entraîner une faible précision du modèle, telle que mesurée par l'erreur quadratique moyenne (RMSE).

Création et adhésion à la collaboration dans AWS Clean Rooms ML

Le créateur de la collaboration est chargé de créer la collaboration, d'inviter les membres et d'attribuer leurs rôles. Les membres invités rejoignent la collaboration et spécifient les paramètres des résultats, les paramètres de destination des artefacts du modèle formé et acceptent les responsabilités de paiement, en fonction de la manière dont la collaboration est configurée.

Création d'une collaboration pour le machine learning

La procédure suivante explique comment créer une collaboration pour le machine learning, inviter un ou plusieurs membres et affecter des membres capables de démarrer la formation des modèles, de recevoir des résultats, de recevoir des résultats de modèles entraînés, y compris des artefacts et des métriques de modèle, et de recevoir des résultats d'inférence de modèles. Le créateur de la collaboration désigne également un membre qui prendra en charge les coûts liés au calcul des requêtes, à la formation des modèles et à l'inférence des modèles.

Console

Pour créer une collaboration pour le machine learning (console)

1. [Créez une collaboration et invitez un ou plusieurs membres à la rejoindre](#)
2. Attribuez aux membres les capacités d'analyse suivantes à l'aide de requêtes et de tâches :
 - Affectez les requêtes Run au membre qui commencera la formation des modèles.
 - Attribuez les résultats de l'analyse aux membres qui recevront les résultats de la requête.
3. Attribuez les capacités suivantes aux membres pour la modélisation du machine learning à l'aide de flux de travail spécialement conçus :

- Affectez la réception des résultats des modèles entraînés au membre qui recevra les résultats des modèles entraînés, y compris les artefacts et les métriques du modèle.
 - Affectez le résultat de l'inférence du modèle au membre qui recevra les résultats de l'inférence du modèle.
4. Pour Configurer le paiement, spécifiez les membres qui paieront les coûts de calcul des requêtes, de formation des modèles, d'inférence de modèles et de génération de jeux de données synthétiques.

Un jeu de données synthétique possède des propriétés statistiques similaires à celles du jeu de données d'origine sur lequel il est basé, mais ne contient pas les observations du monde réel présentes dans le jeu de données d'origine. En utilisant des ensembles de données synthétiques optimisés pour la confidentialité, vous pouvez découvrir de nouveaux cas d'utilisation du modèle de machine learning que les problèmes de confidentialité des données évitaient auparavant. Pour plus d'informations, consultez la section [Génération de données synthétiques améliorée en termes de confidentialité](#).

Chacun de ces coûts peut être attribué au même membre ou à des membres différents. Si un membre invité est le membre responsable du paiement des frais de paiement, il doit accepter ses responsabilités de paiement avant de rejoindre la collaboration.

5. Pour configurer l'adhésion, le créateur de la collaboration peut décider de rejoindre l'adhésion maintenant ou de créer une adhésion ultérieurement. Le créateur de la collaboration doit ensuite configurer la configuration ML.
 - a. Si le créateur de la collaboration est également le destinataire des résultats, il doit également spécifier la destination et le format des résultats de la requête dans les paramètres par défaut des résultats.
 - b. La configuration ML permet à Clean Rooms ML de publier des métriques sur un Compte AWS. Si le créateur de la collaboration reçoit également des artefacts de modèles entraînés, il peut spécifier le compartiment Amazon S3 utilisé pour recevoir les résultats.
 - c. Dans la section Configurations ML, sélectionnez Créer une configuration ML, puis spécifiez la destination de sortie du modèle sur Amazon S3 et le rôle d'accès au service requis pour accéder à cet emplacement.
 - d. Si le créateur de la collaboration est le membre chargé de payer les frais de paiement, il doit accepter ses responsabilités de paiement avant de créer la collaboration.

API

Pour créer une collaboration pour l'apprentissage automatique (API)

1. [Créez une collaboration et invitez un ou plusieurs membres à la rejoindre](#)
2. Attribuez les rôles suivants aux membres de la collaboration :
 - CAN_QUERY- attribué au membre qui commencera la formation et l'inférence du modèle.
 - CAN_RECEIVE_MODEL_OUTPUT- attribué aux membres qui recevront les résultats du modèle formé.
 - CAN_RECEIVE_INFERENCE_OUTPUT- attribué aux membres qui recevront les résultats d'inférence du modèle.

Si le créateur de la collaboration est également le destinataire des résultats, il doit également spécifier la destination et le format des résultats de la requête lors de la création de la collaboration. Ils attribuent également un rôle de service Amazon Resource Name (ARN) pour écrire les résultats dans la destination des résultats de la requête.

3. Spécifiez les membres qui paieront les coûts de calcul des requêtes, de formation des modèles et d'inférence des modèles. Chacun de ces coûts peut être attribué au même membre ou à des membres différents. Si un membre invité est le membre responsable du paiement des frais de paiement, il doit accepter ses responsabilités de paiement avant de rejoindre la collaboration.
4. Le code suivant crée une collaboration, invite un membre capable d'exécuter des requêtes et de recevoir des résultats, et indique que le créateur de la collaboration est le récepteur des artefacts du modèle.

```
import boto3
acr_client= boto3.client('cleanrooms')

collaboration = a_acr_client.create_collaboration(
    members=[
        {
            'accountId': 'invited_member_accountId',
            'memberAbilities':["CAN_QUERY","CAN_RECEIVE_RESULTS"],
            'displayName': 'member_display_name'
        }
    ],
    name='collaboration_name',
    description=collaboration_description,
    creatorMLMemberAbilities= {
```

```

        'customMLMemberAbilities':["CAN_RECEIVE_MODEL_OUTPUT",
"CAN_RECEIVE_INFERENCE_OUTPUT"],
    },
    creatorDisplayName='creator_display_name',
    queryLogStatus="ENABLED",
    analyticsEngine="SPARK",
    creatorPaymentConfiguration={
        "queryCompute": {
            "isResponsible": True
        },
        "machineLearning": {
            "modelTraining": {
                "isResponsible": True
            },
            "modelInference": {
                "isResponsible": True
            }
        }
    }
}
)

collaboration_id = collaboration['collaboration']['id']
print("collaborationId: {collaboration_id}")

member_membership = a_acr_client.create_membership(
    collaborationIdentifier = collaboration_id,
    queryLogStatus = 'ENABLED',
    paymentConfiguration={
        "queryCompute": {
            "isResponsible": True
        },
        "machineLearning": {
            "modelTraining": {
                "isResponsible": True
            },
            "modelInference": {
                "isResponsible": True
            }
        }
    }
)

```

5. Le créateur de la collaboration doit ensuite configurer la configuration ML. La configuration ML permet à Clean Rooms ML de publier des métriques et des journaux sur un Compte AWS. Si

le créateur de la collaboration reçoit également des résultats (artefacts du modèle ou résultats d'inférence), il peut spécifier le compartiment Amazon S3 utilisé pour recevoir les résultats.

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.put_ml_configuration(
    membershipIdentifier=membership_id,
    defaultOutputLocation={
        'roleArn': 'arn:aws:iam::account:role/roleName',
        'destination':{
            's3Destination':{
                's3Uri': "s3://bucketName/prefix"
            }
        }
    }
)
```

Rejoindre une collaboration

Une fois que le créateur de la collaboration a terminé ses tâches, les membres invités doivent terminer les leurs.

Console

Pour créer un abonnement et rejoindre une collaboration (console)

1. Le membre invité [crée une adhésion et rejoint la collaboration](#).
2. Si le membre invité est le membre responsable du paiement, notamment des coûts liés au calcul des requêtes, à la formation des modèles et à l'inférence des modèles, il doit accepter ses responsabilités de paiement avant de rejoindre la collaboration.
3. Le membre invité configure la configuration ML, qui permet à Clean Rooms ML de publier les métriques du modèle sur un Compte AWS. S'il est également le membre qui reçoit les artefacts du modèle entraîné, il doit fournir un compartiment Amazon S3 dans lequel les artefacts du modèle entraîné sont stockés.

API

Pour créer un abonnement et rejoindre une collaboration (API)

1. Si le membre invité est celui qui peut recevoir les résultats, il spécifie la destination et le format des résultats de la requête. Ils fournissent également un rôle de service (ARN) qui permet au service d'écrire dans la destination des résultats de la requête.

Si le membre invité est le membre responsable du paiement, notamment des coûts liés au calcul des requêtes, à la formation des modèles et à l'inférence des modèles, il doit accepter ses responsabilités de paiement avant de rejoindre la collaboration.

Si le membre invité est le membre responsable du paiement de la formation des modèles et de l'inférence des modèles pour la modélisation personnalisée, il doit accepter ses responsabilités de paiement avant de rejoindre la collaboration.

Le code suivant crée une adhésion avec l'enregistrement des requêtes activé.

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_membership(
    membershipIdentifier='membership_id',
    queryLogStatus='ENABLED'
)
```

2. Le membre invité configure la configuration ML, qui permet à Clean Rooms ML de publier les métriques du modèle sur un Compte AWS. S'il est également le membre qui reçoit les artefacts du modèle entraîné, il doit fournir un compartiment Amazon S3 dans lequel les artefacts du modèle entraîné sont stockés.

```
import boto3
acr_ml_client= boto3.client('cleanroomsmml')

acr_ml_client.put_ml_configuration(
    membershipIdentifier='membership_id',
    defaultOutputLocation={
        'roleArn':"arn:aws:iam::account:role/role_name",
        'destination':{
            's3Destination':{
                's3Uri':"s3://bucket_name/prefix"
            }
        }
    }
)
```

Contribution de données de formation dans AWS Clean Rooms ML

Une fois que le créateur de la collaboration a créé la collaboration et que les membres invités l'ont rejoint, vous êtes prêt à apporter des données de formation à la collaboration. Tout membre peut fournir des données de formation.

Console

Pour fournir des données d'entraînement (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, choisissez Tables.
3. Sur la page Tables, choisissez Configurer une nouvelle table.
4. Pour Configurer une nouvelle table, pour Source de données, choisissez Amazon S3, Amazon Athena ou Snowflake et effectuez les étapes suivantes, en fonction de votre source de données :

Si vous utilisez	Then
Amazon S3	1. Choisissez une base de données dans la liste déroulante, puis sélectionnez la table dans la base de données. 2. Pour les colonnes autorisées dans les collaborations, sélectionnez Toutes les colonnes ou Liste personnalisée. 3. Pour les détails de la table configurée, fournissez le nom et une description facultative pour cette table. 4. Si vous souhaitez signaler les métriques du modèle, entrez le nom des métriques et l'instruction Regex qui recherchera la métrique dans les journaux de sortie. 5. Si vous souhaitez activer les balises pour la ressource de table configurée,

Si vous utilisez	Then
	choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
Amazon Athena	1. Choisissez une base de données dans la liste déroulante, puis sélectionnez la table dans la base de données. 2. Pour les colonnes autorisées dans les collaborations, sélectionnez Toutes les colonnes ou Liste personnalisée. 3. Pour les détails de la table configurée, fournissez le nom et une description facultative pour cette table. 4. Si vous souhaitez signaler les métriques du modèle, entrez le nom des métriques et l'instruction Regex qui recherchera la métrique dans les journaux de sortie. 5. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.

Si vous utilisez	Then
Snowflake	1. Spécifiez les informations d'identification Snowflake à l'aide d'un ARN secret existant ou en stockant un nouveau secret pour cette table. 2. Pour les détails de la table et du schéma Snowflake, entrez les détails manuellement ou importez-les automatiquement. 3. Pour le schéma, entrez le nom de la colonne et choisissez le type de données dans la liste déroulante. 4. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.

5. Choisissez Configurer une nouvelle table.
6. Sur la page des détails de la table, choisissez Configurer la règle d'analyse pour configurer une règle d'analyse personnalisée pour cette table. Une règle d'analyse personnalisée limite l'accès à vos données. Vous pouvez autoriser un ensemble spécifique de requêtes préautorisées sur vos données ou autoriser un ensemble spécifique de comptes à interroger vos données.
 - a. Pour le type de règle d'analyse, choisissez Personnalisé et pour Méthode de création, choisissez Flux guidé.
 - b. Choisissez Suivant.
 - c. Pour Spécifier les contrôles d'analyse, choisissez entre Vérifier chaque nouvelle analyse et Autoriser toute analyse réalisée par des collaborateurs spécifiques.
 - d. Choisissez Suivant.
 - e. (Facultatif) Pour les commandes Spécifier les résultats d'analyse, pour les colonnes non autorisées en sortie, indiquez si vous souhaitez exclure des colonnes de la sortie. Si vous choisissez Aucune, aucune colonne n'est exclue de la sortie. Si vous choisissez Liste personnalisée, vous pouvez spécifier certaines colonnes qui seront supprimées de la sortie.

- f. Pour les analyses supplémentaires appliquées à la sortie, spécifiez si vous souhaitez autoriser, refuser ou exiger une analyse supplémentaire avant que les résultats ne soient générés.
 - g. Choisissez Suivant.
 - h. (Facultatif) Pour Définir la confidentialité différentielle, choisissez Désactiver.
 - i. Choisissez Suivant.
 - j. Consultez les informations de la page Révision et configuration, puis choisissez Configurer la règle d'analyse.
7. Sur la page des détails du tableau, choisissez Associer à la collaboration.
8. Dans la boîte de dialogue Associer une table, sélectionnez la collaboration à laquelle vous souhaitez associer cette table et choisissez Choisir une collaboration.
9. Sur la page Associer la table, passez en revue et vérifiez les informations figurant dans Détails de l'association de tables, Accès au service et Tags. Choisissez Associer une table.
10. Dans le tableau Tables associées à votre compte, sélectionnez le bouton radio situé à côté du tableau que vous venez d'associer. Dans le menu Actions, choisissez Configurer dans le groupe de règles d'analyse de collaboration.
11. Sur la page Configurer les règles d'analyse de collaboration, dans Analyses supplémentaires autorisées, indiquez si des membres de la collaboration ou des membres spécifiques de la collaboration peuvent effectuer des analyses supplémentaires.

Pour la livraison des résultats, choisissez les membres autorisés à recevoir des résultats à partir des résultats des requêtes.

12. Choisissez Configurer la règle d'analyse.

API

Pour fournir des données d'entraînement (API)

1. Configurez une AWS Glue table existante à utiliser en AWS Clean Rooms fournissant la table et les colonnes qui peuvent être utilisées.

Exécutez le code suivant avec vos paramètres spécifiques.

```
import boto3
acr_client= boto3.client('cleanrooms')
```

```
acr_client.create_configured_table(
    name='configured_table_name',
    tableReference= {
        'glue': {
            'tableName': 'glue_table_name',
            'databaseName': 'glue_database_name'
        }
    },
    analysisMethod="DIRECT_QUERY",
    allowedColumns=["column1", "column2", "column3",...]
)
```

2. Configurez une règle d'analyse personnalisée qui limite l'accès à vos données. Vous pouvez autoriser un ensemble spécifique de requêtes préautorisées sur vos données ou autoriser un ensemble spécifique de comptes à interroger vos données.

Exécutez le code suivant avec vos paramètres spécifiques.

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_analysis_rule(
    configuredTableIdentifier='configured_table_id',
    analysisRuleType='CUSTOM',
    analysisRulePolicy= {
        'v1': {
            'custom': {
                'allowedAnalyses': ['ANY_QUERY'],
                'allowedAnalysisProviders': ['query_runner_account'],
                'additionalAnalyses': "REQUIRED"
            }
        }
    }
)
```

Dans cet exemple, un compte spécifique est autorisé à exécuter n'importe quelle requête sur les données et une analyse supplémentaire est requise.

3. Associez une table configurée à la collaboration et attribuez un rôle d'accès aux services aux AWS Glue tables.

Exécutez le code suivant avec vos paramètres spécifiques.

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association(
    name='configured_table_association_name',
    membershipIdentifier='membership_id',
    configuredTableIdentifier='configured_table_id',
    roleArn='arn:aws:iam::account:role/role_name'
)
```

Note

Ce rôle de service dispose d'autorisations d'accès aux tables. Le rôle de service ne peut être assumé que AWS Clean Rooms pour exécuter les requêtes autorisées au nom du membre autorisé à effectuer des requêtes. Aucun membre de la collaboration (autre que le propriétaire des données) n'a accès aux tables sous-jacentes de la collaboration. Le propriétaire des données peut désactiver la confidentialité différentielle pour que ses tables puissent être consultées par d'autres membres.

4. Enfin, ajoutez une règle d'analyse à l'association de tables configurée.

Exécutez le code suivant avec vos paramètres spécifiques.

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association_analysis_rule(

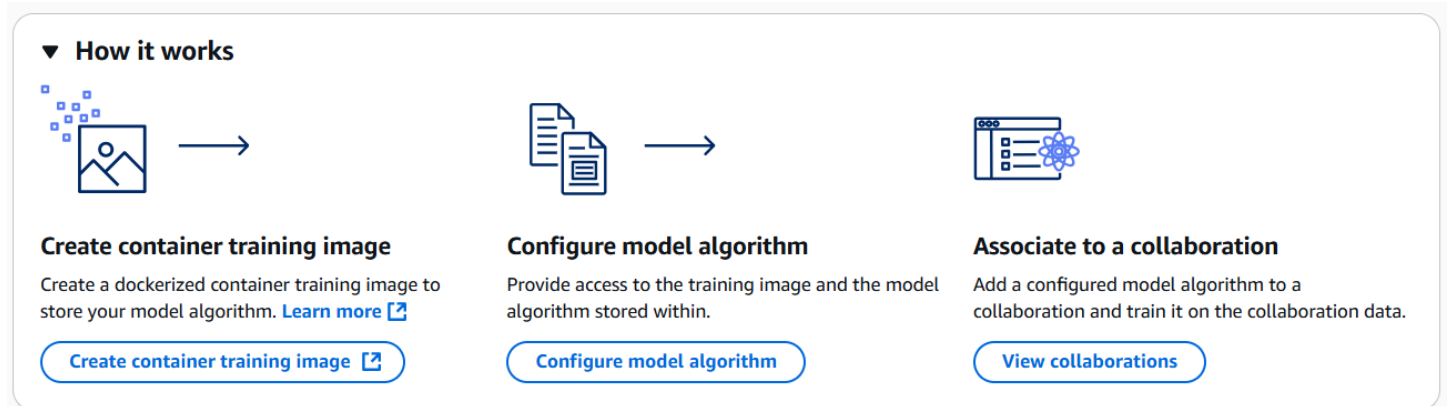
    configuredTableAssociationIdentifier='configured_table_association_identifier',
    membershipIdentifier='membership_id',
    configuredTableIdentifier='configured_table_id',
    analysisRuleType = 'CUSTOM',
    analysisRulePolicy= {
        'v1': {
            'custom': {
                'allowedAdditionalAnalyses':
                ['configured_model_algorithm_association_arns'],
                'allowedResultReceivers': ['query_runner_account']
            }
        }
    }
)
```

```
}  
)
```

Configuration d'un algorithme de modèle dans AWS Clean Rooms ML

Après avoir [créé une image d'entraînement du conteneur](#), vous devez configurer l'algorithme de votre modèle. La configuration d'un algorithme de modèle permet de l'associer à une collaboration.

L'image suivante montre la configuration d'un algorithme de modèle en tant qu'étape qui se produit après la création de l'image d'entraînement du conteneur et avant de l'associer à la collaboration.



Console

Pour configurer un algorithme de modèle ML personnalisé (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Modèles ML personnalisés.
3. Sur la page Modèles ML personnalisés, choisissez Configurer l'algorithme du modèle.
4. Sur la page Configurer l'algorithme du modèle, pour les détails de l'algorithme du modèle, entrez un nom et une description facultative.
5. Si vous souhaitez effectuer un entraînement sur un modèle, pour les détails du conteneur ECR de l'image d'entraînement,
 - a. Cochez la case Spécifier l'URI de l'image d'entraînement.

- b. Sélectionnez le référentiel contenant le modèle d'entraînement, le conteneur d'inférence, ou les deux, dans la liste déroulante.
 - c. Sélectionnez l'image.
 - d. (Facultatif) Entrez la valeur des points d'entrée pour accéder à l'image d'entraînement.
 - e. (Facultatif) Entrez la valeur des arguments.
6. (Facultatif) Si vous souhaitez signaler les métriques du modèle, pour les métriques d'entraînement, entrez le nom des métriques et l'instruction Regex qui effectuera une recherche dans les journaux de sortie pour trouver la métrique.
7. Si vous souhaitez effectuer une inférence de modèle, pour les détails du conteneur ECR de l'image d'inférence,
 - a. Cochez la case Spécifier l'URI de l'image d'inférence.
 - b. Sélectionnez le référentiel dans la liste déroulante.
 - c. Sélectionnez l'image.
8. Pour l'accès au service, choisissez le nom du rôle de service existant qui sera utilisé pour accéder à cette table.
9. Pour le chiffrement, choisissez Personnaliser les paramètres de chiffrement pour spécifier votre propre clé KMS et les informations associées. Dans le cas contraire, Clean Rooms ML gèrera le chiffrement.
10. Si vous souhaitez activer les balises, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
11. Choisissez Configurer l'algorithme du modèle.

API

Pour configurer un algorithme de modèle ML (API) personnalisé

1. Créez une image de docker compatible avec l' SageMaker IA. Clean Rooms ML ne prend en charge que les SageMaker images docker compatibles avec l'IA.
2. Après avoir créé une image docker compatible avec l' SageMaker IA, utilisez Amazon ECR pour créer une image d'entraînement. Suivez les instructions du [guide de l'utilisateur d'Amazon Elastic Container Registry](#) pour créer une image de formation sur les conteneurs.
3. Configurez l'algorithme du modèle à utiliser dans Clean Rooms ML. Vous devez fournir les informations suivantes :

- Le lien vers le référentiel Amazon ECR et des arguments supplémentaires pour entraîner le modèle et exécuter l'inférence. Clean Rooms ML prend en charge l'exécution de tâches de transformation par lots sur un conteneur d'inférence.
- Rôle d'accès au service qui permet à Clean Rooms ML d'accéder au référentiel.
- (Facultatif) Un conteneur d'inférence. Bien que vous puissiez le fournir dans un algorithme de modèle configuré distinct, nous vous recommandons de le fournir à cette étape afin que le conteneur d'entraînement et le conteneur d'inférence soient gérés dans le cadre de la même ressource.

Exécutez le code suivant avec vos paramètres spécifiques.

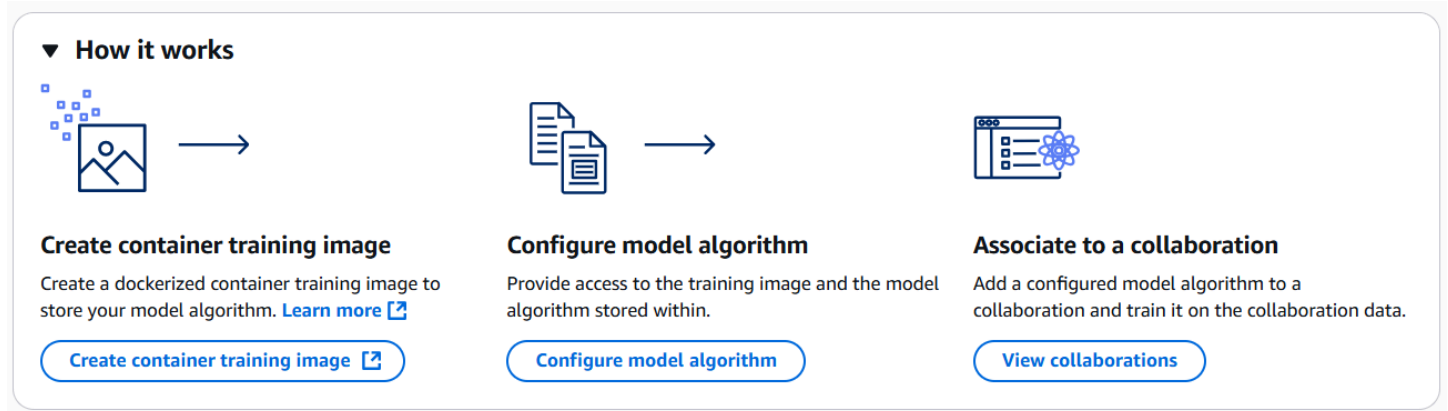
```
import boto3
acr_ml_client= boto3.client('cleanroomsmml')

acr_ml_client.create_configured_model_algorithm(
    name='configured_model_algorithm_name',
    trainingContainerConfig={
        'imageUri': 'account.dkr.ecr.region.amazonaws.com/image_name:tag',
        'metricDefinitions': [
            {
                'name': 'custom_metric_name_1',
                'regex': 'custom_metric_regex_1'
            }
        ]
    },
    inferenceContainerConfig={
        'imageUri': 'account.dkr.ecr.region.amazonaws.com/image_name:tag',
    },
    roleArn='arn:aws:iam::account:role/role_name'
)
```

Associer l'algorithme du modèle configuré dans AWS Clean Rooms ML

Après avoir configuré l'algorithme du modèle, vous êtes prêt à l'associer à une collaboration. L'association d'un algorithme de modèle met celui-ci à la disposition de tous les membres de la collaboration.

L'image suivante montre l'association de l'algorithme de modèle configuré comme dernière étape, après avoir créé l'image d'apprentissage du conteneur et configuré un algorithme de modèle.



Console

Note

Une fois que l'algorithme du modèle est associé, il ne peut pas être modifié. Pour apporter des modifications, vous pouvez supprimer l'algorithme du modèle associé et en associer un nouveau.

Pour associer un algorithme de modèle ML personnalisé (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Modèles ML personnalisés.
3. Sur la page Modèles ML personnalisés, choisissez l'algorithme de modèle configuré que vous souhaitez associer à une collaboration, puis choisissez Associer à la collaboration.
4. Dans la fenêtre Associer un algorithme de modèle configuré, choisissez la collaboration à laquelle vous souhaitez vous associer.
5. Choisissez Choisir une collaboration.
6. Sur la page Associer un algorithme de modèle, pour les détails de l'association d'algorithmes de modèle, entrez un nom et une description facultative.
7. Pour Algorithme de modèle, choisissez un algorithme de modèle configuré.
8. Pour les configurations de confidentialité des exportations du modèle Trained,

- a. Pour exporter des fichiers modèles, cochez la case Fichiers modèles.
 - b. Pour exporter les fichiers de sortie, cochez la case Fichiers de sortie.
 - c. Entrez une valeur de taille maximale pour les données exportées. La valeur doit être comprise entre 0,01 et 10.
9. (Facultatif) Si vous souhaitez envoyer des journaux d'erreurs complets ou des résumés d'erreurs plus courts aux membres, sous Configuration de la confidentialité des tâches d'inférence de modèles entraînés,
- a. Sous Journaux complets, sélectionnez un ou plusieurs comptes IDs dans la liste déroulante.
 - b. (Facultatif) Si vous souhaitez envoyer des journaux correspondant à un modèle de filtre, entrez un modèle de filtre.
 - c. (Facultatif) Si vous souhaitez ajouter un autre compte et un modèle de filtre facultatif, choisissez Ajouter une politique de journalisation.
 - d. Sous Résumés des erreurs, sélectionnez un ou plusieurs comptes IDs dans la liste déroulante.
 - e. (Facultatif) Sélectionnez une ou plusieurs entités à supprimer pour spécifier les entités qui seront supprimées du journal des erreurs ou des résumés d'erreurs.
 - PII — supprimer les informations personnellement identifiables
 - Chiffres — expurgez les numéros
 - Personnalisé : rédaction basée sur le modèle de rédaction personnalisé
 - i. Si vous avez choisi Personnalisé à l'étape précédente, entrez un modèle de rédaction personnalisé. Cela enregistre les informations correspondant à ce modèle.
 - ii. (Facultatif) Si vous souhaitez ajouter un autre modèle de rédaction personnalisé, choisissez Ajouter un autre modèle personnalisé.
10. (Facultatif) Si vous souhaitez configurer les métriques du modèle entraîné, sous Configuration des métriques du modèle entraîné, sélectionnez un niveau de bruit dans la liste déroulante.

Vous pouvez choisir Aucun, Faible, Moyen et Élevé.

11. (Facultatif) Si vous souhaitez définir la taille maximale des artefacts, sous Configuration des artefacts, entrez la valeur de la taille maximale des artefacts. La valeur doit être comprise entre 0,01 et 10.
12. (Facultatif) Si vous souhaitez activer les balises, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
13. Choisissez Associer.

API

Pour associer un algorithme de modèle ML (API) personnalisé

Exécutez le code suivant avec vos paramètres spécifiques.

Vous fournissez également une politique de confidentialité qui définit qui a accès aux différents journaux, permet aux clients de définir l'expression régulière et quelle quantité de données peut être exportée à partir des sorties du modèle d'entraînement ou des résultats d'inférence.

Note

Les associations d'algorithmes de modèles configurées sont immuables.

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_configured_model_algorithm_association(
    name='configured_model_algorithm_association_name',
    description='purpose of the association',
    configuredModelAlgorithmArn='arn:aws:cleanrooms-ml:region:account:membership/
membershipIdentifier/configured-model-algorithm/identifier',
    privacyConfiguration={
        "policies": {
            "trainedModelExports": {
                "filesToExport": ['files to export'],
                "containerLogs": [
                    {
                        "allowedAccountIds": ['member_account_id'],
                        "filterPattern": ['filter pattern'],
                        "logRedactionConfiguration": {
                            "entitiesToRedact": [
```

```

        'ALL_PERSONALLY_IDENTIFIABLE_INFORMATION',
        'NUMBERS',
        'CUSTOM'
    ],
    "customEntityConfig": {
        "customDataIdentifiers": [
            'custom_regex_1',
            'custom_regex_2'
        ]
    }
},
"containerMetrics": {
    "noiseLevel": 'noise value'
},
"maxArtifactSize": {
    "unit": 'unit',
    "value": 'number'
}
},
"trainedModelInferenceJobs": {
    "containerLogs": [
        {
            "allowedAccountIds": ['member_account_id'],
            "filterPattern": ['filter pattern'],
            "logRedactionConfiguration": {
                "entitiesToRedact": [
                    'ALL_PERSONALLY_IDENTIFIABLE_INFORMATION',
                    'NUMBERS',
                    'CUSTOM'
                ],
                "customEntityConfig": {
                    "customDataIdentifiers": [
                        'custom_regex_1',
                        'custom_regex_2'
                    ]
                }
            }
        ]
    },
    "maxOutputSize": {
        "unit": 'unit',
        "value": 'number'
    }
}

```

```

    }
  }
},
tags={
  'tag': 'tag'
}
)

```

Une fois que l'algorithme du modèle configuré est associé à la collaboration, les fournisseurs de données de formation doivent ajouter une règle d'analyse de collaboration à leur table. Cette règle permet à l'association d'algorithmes du modèle configuré d'accéder à sa table configurée. Tous les fournisseurs de données de formation contributeurs doivent exécuter le code suivant :

```

import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association_analysis_rule(
    membershipIdentifier= 'membership_id',
    configuredTableAssociationIdentifier= 'configured_table_association_id',
    analysisRuleType= 'CUSTOM',
    analysisRulePolicy = {
        'v1': {
            'custom': {
                'allowedAdditionalAnalyses': ['arn:aws:cleanrooms-
ml:region:*:membership/*/configured-model-algorithm-association/*'],
                'allowedResultReceivers': []
            }
        }
    }
)

```

Note

Les associations d'algorithmes de modèles configurés étant immuables, nous recommandons aux fournisseurs de données de formation qui souhaitent autoriser l'utilisation de modèles de liste d'utiliser des caractères génériques `allowedAdditionalAnalyses` lors des premières itérations de configuration de modèles personnalisés. Cela permet aux fournisseurs de modèles d'itérer leur code sans

obliger les autres prestataires de formation à s'associer à nouveau avant d'entraîner leur code de modèle mis à jour avec des données.

Création d'un canal d'entrée ML dans AWS Clean Rooms ML

Prérequis :

- Et Compte AWS avec accès à AWS Clean Rooms
- Une collaboration configurée dans l' AWS Clean Rooms endroit où vous souhaitez créer le canal d'entrée ML
- Autorisations pour interroger des données et créer des canaux d'entrée ML dans le cadre de la collaboration.
- (Facultatif) Un algorithme de modèle existant à associer au canal d'entrée ML, ou des autorisations pour en créer un nouveau
- (Facultatif) Tables avec des règles d'analyse qui peuvent être exécutées pour le modèle que vous avez spécifié.
- (Facultatif) Modèle de requête ou d'analyse SQL existant à utiliser pour générer le jeu de données
- (Facultatif) Un rôle de service existant avec les autorisations appropriées, ou des autorisations pour créer un nouveau rôle de service
- (Facultatif) Une AWS KMS clé personnalisée si vous souhaitez utiliser votre propre clé de chiffrement
- Autorisations appropriées pour créer et gérer des modèles de machine learning dans le cadre de la collaboration

Un canal d'entrée ML est un ensemble de données créé à partir d'une requête de données spécifique. Les membres capables d'interroger des données peuvent préparer leurs données pour l'entraînement et l'inférence en créant un canal d'entrée ML. La création d'un canal d'entrée ML permet d'utiliser ces données dans différents modèles d'entraînement au sein d'une même collaboration. Vous devez créer des canaux d'entrée ML distincts pour l'entraînement et l'inférence.

Pour créer un canal d'entrée ML, vous devez spécifier la requête SQL utilisée pour interroger les données d'entrée et créer le canal d'entrée ML. Les résultats de cette requête ne sont jamais partagés avec aucun membre et restent dans les limites de Clean Rooms ML. Le nom de ressource

Amazon (ARN) de référence est utilisé dans les étapes suivantes pour entraîner un modèle ou exécuter une inférence.

Console

Pour créer un canal d'entrée ML (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, choisissez la collaboration dans laquelle vous souhaitez créer un canal d'entrée ML.
4. Une fois la collaboration ouverte, choisissez l'onglet Modèles ML.
5. Sous Modèles ML personnalisés, dans la section Canaux d'entrée ML, choisissez Créer un canal d'entrée ML.
6. Sur la page Créer un canal d'entrée ML, pour les détails du canal d'entrée ML, procédez comme suit :
 - a. Dans Nom, entrez un nom unique pour votre chaîne.
 - b. (Facultatif) Dans Description, entrez une description de votre chaîne.
 - c. Pour Algorithme de modèle associé, sélectionnez l'algorithme à utiliser.

Choisissez Associer un algorithme de modèle pour en ajouter un nouveau.

7. Pour Dataset, choisissez une méthode pour générer le jeu de données d'entraînement :
 - Choisissez SQL query pour utiliser les résultats d'une requête SQL comme ensemble de données d'apprentissage.

Si vous avez choisi SQL query, saisissez votre requête dans le champ de requête SQL.

(Facultatif) Pour importer une requête que vous avez utilisée récemment, choisissez Importer à partir de requêtes récentes.

- Choisissez le modèle d'analyse pour utiliser les résultats d'un modèle d'analyse comme jeu de données d'apprentissage.

 Warning

La génération de données synthétiques empêche de déduire des attributs individuels, que des individus spécifiques soient présents dans l'ensemble de données d'origine ou que des attributs d'apprentissage de ces individus soient présents. Cependant, cela n'empêche pas les valeurs littérales de l'ensemble de données d'origine, y compris les informations personnelles identifiables (PII), d'apparaître dans l'ensemble de données synthétique.

Nous recommandons d'éviter dans le jeu de données d'entrée les valeurs associées à une seule personne concernée, car elles peuvent permettre de réidentifier une personne concernée. Par exemple, si un seul utilisateur vit dans un code postal, la présence de ce code postal dans le jeu de données synthétique confirmera que cet utilisateur figurait dans le jeu de données d'origine. Des techniques telles que la troncation de valeurs de haute précision ou le remplacement de catalogues peu courants par d'autres peuvent être utilisées pour atténuer ce risque. Ces transformations peuvent faire partie de la requête utilisée pour créer le canal d'entrée ML.

- a. Si aucune table n'est associée, choisissez Associer une table pour ajouter des tables avec une règle d'analyse pouvant être exécutée pour le modèle spécifié.
- b. Choisissez le type de travailleur à utiliser lors de la création de ce canal de données. Le type de travailleur par défaut est CR.1X. Spécifiez le nombre de travailleurs à utiliser. Le numéro de travailleur par défaut est 16. Pour définir les propriétés de Spark :
 - i. Développez les propriétés Spark.
 - ii. Choisissez Ajouter des propriétés Spark.
 - iii. Dans la boîte de dialogue des propriétés de Spark, choisissez un nom de propriété dans la liste déroulante et entrez une valeur.

Le tableau suivant fournit une définition pour chaque propriété.

Pour plus d'informations sur les propriétés de Spark, consultez la section [Propriétés de Spark](#) dans la documentation d'Apache Spark.

Nom de propriété	Description	Valeur par défaut
Défaillances de Spark.Task.Max	Contrôle le nombre de fois consécutives qu'une tâche peut échouer avant que la tâche échoue. Nécessite une valeur supérieure ou égale à 1. Le nombre de tentatives autorisées est égal à cette valeur moins 1. Le nombre d'échecs est réinitialisé en cas de réussite d'une tentative. Les échecs liés aux différentes tâches ne s'accumulent pas jusqu'à atteindre cette limite.	4
fichiers spark.sql.maxPartitionBytes	Définit le nombre maximum d'octets à regrouper dans une seule partition lors de la lecture à partir de sources basées sur des fichiers telles que Parquet, JSON et ORC.	128 MO
Spark.HadoopFS.S3.maxRéessaie	Définit le nombre maximal de tentatives pour les opérations sur les fichiers Amazon S3.	
spark.network.timeout	Définit le délai d'expiration par défaut pour toutes les interactions réseau. Remplace les paramètres de délai d'expiration suivants s'ils ne sont pas configurés : • stockage d'étincelles. blockManagerHeartbeatTimeoutMs • Spark.shuffle.io. Délai de connexion • Spark.rpc.AskTimeout • Spark.rpc. Délai de recherche expiré	années 120

Nom de propriété	Description	Valeur par défaut
<code>spark.rdd.compress</code>	Spécifie s'il faut compresser les partitions RDD sérialisées à l'aide de <code>spark.io.compression.codec</code> . S'applique à <code>StorageLevel.MEMORY_ONLY_SER</code> en Java et Scala, ou à <code>.MEMORY_ONLY</code> en Python. <code>StorageLevel</code> Réduit l'espace de stockage mais nécessite un temps de traitement supplémentaire du processeur.	FALSE
<code>spark.shuffle.spill.compress</code>	Spécifie s'il faut compresser les données du shuffle spill à l'aide de <code>spark.io.compression.codec</code> .	TRUE
<code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>	Définit la taille cible en octets pour les partitions shuffle lors de l'optimisation adaptative lorsque <code>spark.sql.adaptive.enabled</code> est vrai. Contrôle la taille de la partition lors de la fusion de petites partitions ou du fractionnement de partitions asymétriques.	(valeur de <code>spark.sql.adaptive.shuffle.targetPostShuffleInputSize</code>)
<code>spark.sql.adaptive.autoBroadcastJoinSeuil</code>	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. S'applique uniquement dans le cadre adaptatif. Utilise la même valeur par défaut que <code>spark.sql.autoBroadcastJoinSeuil</code> . Réglez sur -1 pour désactiver la diffusion.	(aucun)
<code>spark.sql.adaptive.CoalescePartitions.Activé</code>	Spécifie s'il faut fusionner les partitions shuffle contiguës sur la base de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> pour optimiser la taille des tâches. Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE

Nom de propriété	Description	Valeur par défaut
<code>spark.sql.adaptive.CoalescePartitions.initialPartitionNum</code>	Définit le nombre initial de partitions shuffle avant la fusion. Nécessite que <code>spark.sql.adaptive.enabled</code> et <code>spark.sql.adaptive.CoalescePartitions.enabled</code> soient vrais. La valeur par défaut est <code>spark.sql.shuffle.partitions</code> .	(aucun)
<code>spark.sql.adaptive.CoalescePartitions.minPartitionSize</code>	Définit la taille minimale des partitions shuffle coalescées afin d'éviter que les partitions ne deviennent trop petites lors de l'optimisation adaptative.	1 Mo
<code>spark.sql.adaptive.coalescepartitions.ParallelismFirst</code>	Spécifie s'il faut calculer la taille des partitions en fonction du parallélisme des clusters plutôt que de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> lors de la coalescence de la partition. Génère des tailles de partition inférieures à la taille cible configurée afin de maximiser le parallélisme. Nous vous recommandons de définir ce paramètre sur <code>false</code> sur les clusters occupés afin d'améliorer l'utilisation des ressources en évitant de trop petites tâches.	TRUE
<code>spark.sql.adaptive.enabled</code>	Spécifie s'il faut activer l'exécution adaptative des requêtes afin de réoptimiser les plans de requêtes pendant l'exécution des requêtes, sur la base de statistiques d'exécution précises.	TRUE
<code>spark.sql.adaptive.forceOptimizeSkewedJoints</code>	Spécifie s'il faut forcer l'activation <code>OptimizeSkewedJoin</code> même si cela introduit un shuffle supplémentaire.	FALSE

Nom de propriété	Description	Valeur par défaut
<code>spark.sql.adaptive.localShuffleReader.activated</code>	Spécifie s'il faut utiliser des lecteurs de shuffle locaux lorsque le partitionnement aléatoire n'est pas requis, par exemple après la conversion des jointures par tri-fusion en jointures par hachage par diffusion. Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	Définit la taille de partition maximale en octets pour la création de cartes de hachage locales. Privilégie les jointures par hachage mélangées par rapport aux jointures par tri-fusion lorsque : - Cette valeur est égale ou supérieure à <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> - Toutes les tailles de partition se situent dans cette limite Remplace <code>spark.sql.join.preferSortMerge</code> Paramètre de jointure.	0 octet
<code>spark.sql.adaptive.optimizeSkewnessInRebalancePartitions.activated</code>	Spécifie s'il faut optimiser les partitions shuffle asymétriques en les divisant en partitions plus petites sur la base de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> . Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE

Nom de propriété	Description	Valeur par défaut
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	Définit le facteur de seuil de taille pour la fusion des partitions lors du fractionnement. Partitions inférieures à ce facteur multipliées par <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> sont fusionnées.	0.2
<code>spark.sql.adaptive.SkewJoin.Enabled</code>	Spécifie s'il faut gérer l'inclinaison des données dans les jointures mélangées en divisant et éventuellement en répliquant les partitions asymétriques. S'applique aux jointures par tri-fusion et aux jointures de hachage mélangées. Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code>	Détermine le facteur de taille qui détermine l'inclinaison de la partition. Une partition est inclinée lorsque sa taille dépasse les deux valeurs suivantes : - Ce facteur multiplié par la taille médiane de la partition - La valeur de <code>spark.sql.adaptive.SkewJoin.skewedPartitionThresholdInBytes</code>	5

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	Définit le seuil de taille en octets pour identifier les partitions asymétriques. Une partition est inclinée lorsque sa taille dépasse les deux valeurs suivantes : - Ce seuil - Taille de partition médiane multipliée par spark.sql.adaptive.skewJoin.skewedPartitionFactor Nous recommandons de définir cette valeur à une valeur supérieure à spark.sql.adaptive.advisoryPartitionSizeInBytes.	256 MO
spark.sql.autoBroadcastJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. Réglez sur -1 pour désactiver la diffusion.	10 MO
Spark.SQL. Délai de diffusion	Contrôle le délai d'expiration en secondes pour les opérations de diffusion lors des jointures de diffusion.	300 secondes
spark.sql.cbo.enabled	Spécifie s'il faut activer l'optimisation basée sur les coûts (CBO) pour l'estimation des statistiques du plan.	FALSE
spark.sql.cbo.joinreorder.dp.starfilter	Spécifie s'il faut appliquer l'heuristique du filtre de jointure en étoile lors de l'énumération des jointures basée sur les coûts.	FALSE

Nom de propriété	Description	Valeur par défaut
spark.sql.cbo.joinReorder.dp.Threshold	Définit le nombre maximum de nœuds joints autorisés dans l'algorithme de programmation dynamique.	12
spark.sql.cbo.JoinReorder.Enabled	Spécifie s'il faut activer la réorganisation des jointures dans le cadre de l'optimisation basée sur les coûts (CBO).	FALSE
spark.sql.cbo.planstats.activé	Spécifie s'il faut extraire le nombre de lignes et les statistiques des colonnes du catalogue lors de la génération du plan logique.	FALSE
spark.sql.cbo.starSchemaDetection	Spécifie s'il faut activer la réorganisation des jointures en fonction de la détection du schéma en étoile.	FALSE
fichiers spark.sql.maxPartitionNum	Définit le nombre maximal cible de partitions de fichiers fractionnées pour les sources basées sur des fichiers (Parquet, JSON et ORC). Redimensionne les partitions lorsque le nombre initial dépasse cette valeur. Il s'agit d'une cible suggérée et non d'une limite garantie.	(aucun)
fichiers spark.sql.maxRecordSizePerDossier	Définit le nombre maximum d'enregistrements à écrire dans un seul fichier. Aucune limite ne s'applique lorsqu'il est réglé sur zéro ou sur une valeur négative.	0

Nom de propriété	Description	Valeur par défaut
spark.sql.minPartitionNum	Définit le nombre minimum cible de partitions de fichiers fractionnées pour les sources basées sur des fichiers (Parquet, JSON et ORC). La valeur par défaut est spark.sql. leafNodeDefaultParallélisme. Il s'agit d'une cible suggérée et non d'une limite garantie.	(aucun)
spark.sql.inMemoryColumnarStockage. Taille du lot	Contrôle la taille du lot pour la mise en cache en colonnes. L'augmentation de la taille améliore l'utilisation de la mémoire et la compression, mais augmente le risque d' out-of-memoryerreurs.	10 000
spark.sql.inMemoryColumnarStockage compressé	Spécifie s'il faut sélectionner automatiquement les codecs de compression pour les colonnes en fonction des statistiques de données.	TRUE
spark.sql.inMemoryColumnarRange enableVectorizedReader	Spécifie s'il faut activer la lecture vectorisée pour la mise en cache en colonnes.	TRUE
spark.sql.legacy. allowHashOnMapType	Spécifie s'il faut autoriser les opérations de hachage sur les structures de données de type carte. Cet ancien paramètre assure la compatibilité avec la gestion des types de carte dans les anciennes versions de Spark.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.allowNegativeScaleOfDecimal	Spécifie s'il faut autoriser les valeurs d'échelle négatives dans les définitions de type décimal. Cet ancien paramètre est compatible avec les anciennes versions de Spark qui prenaient en charge les échelles décimales négatives.	
spark.sql.legacy.castComplexTypesToString.activé	Spécifie s'il faut activer le comportement existant pour convertir des types complexes en chaînes. Maintient la compatibilité avec les règles de conversion de type des anciennes versions de Spark.	
spark.sql.legacy.charVarcharAsString	Spécifie s'il faut traiter les types CHAR et VARCHAR comme des types STRING. Cet ancien paramètre assure la compatibilité avec la gestion des types de chaînes des anciennes versions de Spark.	
spark.sql.legacy.createEmptyCollectionUsingStringType	Spécifie s'il faut créer des collections vides à l'aide d'éléments de type chaîne. Cet ancien paramètre assure la compatibilité avec le comportement d'initialisation des collections des anciennes versions de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal.activé	Spécifie s'il faut interpréter les littéraux exponentiels comme des types décimaux. Cet ancien paramètre assure la compatibilité avec la gestion littérale numérique des anciennes versions de Spark.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.json.allowEmptyString.activé	Spécifie s'il faut autoriser les chaînes vides dans le traitement JSON. Cet ancien paramètre maintient la compatibilité avec le comportement d'analyse JSON des anciennes versions de Spark.	
spark.sql.legacy.parquet.int96RebaseModelRead	Spécifie s'il faut utiliser le mode de rebase d' INT96 horodatage existant lors de la lecture des fichiers Parquet. Cet ancien paramètre assure la compatibilité avec la gestion des horodatages des anciennes versions de Spark.	
spark.sql.legacy.timeParserPolicy	Contrôle le comportement d'analyse temporelle à des fins de rétrocompatibilité. Cet ancien paramètre détermine la manière dont les horodatages et les dates sont analysés à partir de chaînes.	
spark.sql.Legacy.TypeCoercion.datetimeToString.activé	Spécifie s'il faut activer le comportement de coercion de type existant lors de la conversion de valeurs date/heure en chaînes. Maintient la compatibilité avec les règles de conversion date/heure des anciennes versions de Spark.	
spark.sql.maxSinglePartitionOctets	Définit la taille maximale de la partition en octets. Le planificateur introduit des opérations de shuffle pour les partitions plus grandes afin d'améliorer le parallélisme.	128 m

Nom de propriété	Description	Valeur par défaut
Cache de métadonnées Spark.SQL TTLSeconds	Contrôle le time-to-live (TTL) pour les caches de métadonnées. S'applique aux métadonnées des fichiers de partition et aux caches du catalogue de sessions. Nécessite : • Une valeur positive supérieure à zéro • spark.sql.CatalogImplementation défini sur hive • spark.sql.hive.filesourcePartitionFileCacheSize supérieur à zéro • spark.sql.hive.manageFilesourcePartitions défini sur true	-1000 ms
spark.sql.optimize r.collapsePro jectAlwaysEn ligne	Spécifie s'il faut réduire les projections adjacentes et les expressions en ligne, même si cela entraîne une duplication.	FALSE
spark.sql.optimize r.dynamicPa rtitionPruning.act ivé	Spécifie s'il faut générer des prédicats pour les colonnes de partition utilisées comme clés de jointure.	TRUE
spark.sql.optimize r.enableCsv Expressio nOptimisation	Spécifie s'il faut optimiser les expressions CSV dans l'optimiseur SQL en supprimant les colonnes inutiles des opérations from_csv.	TRUE

Nom de propriété	Description	Valeur par défaut
spark.sql.optimize r.enableJo inExpressi onOptimisation	Spécifie s'il faut optimiser les expressions JSON dans l'optimiseur SQL en : • Supprimer les colonnes inutiles des opérations from_json • Simplification des combinaisons from_json et to_json • Optimisation des opérations named_struct	TRUE
Spark.SQL .Optimizer.Règles exclues	Définit les règles d'optimisation à désactiver, identifiées par des noms de règles séparés par des virgules. Certaines règles ne peuvent pas être désactivées car elles sont nécessaires pour être correctes. L'optimiseur enregistre les règles qui ont été désactivées avec succès.	(aucun)
Spark.SQL .Optimizer.Runtime .BloomFilter. applicationSideScan SizeThreshold	Définit la taille de scan agrégée minimale en octets requise pour injecter un filtre Bloom côté application.	10 GO
Spark.SQL .Optimize r.Runtime.BloomFil ter.creationS ideThreshold	Définit le seuil de taille maximale pour injecter un filtre Bloom côté création.	10 MO
Spark.SQL .Optimizer.Runtime .BloomFilter.Activé	Spécifie s'il faut insérer un filtre Bloom pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	TRUE

Nom de propriété	Description	Valeur par défaut
Spark.SQL .Optimizer.Runtime .BloomFilter. expectedNumItems	Définit le nombre par défaut d'éléments attendus dans le filtre Bloom d'exécution.	1000000
Spark.SQL .Optimizer.Runtime .BloomFilter. maxNumBits	Définit le nombre maximum de bits autorisés dans le filtre Bloom d'exécution.	67108864
Spark.SQL .Optimizer.Runtime .BloomFilter. maxNumItems	Définit le nombre maximum d'éléments attendus autorisés dans le filtre Bloom d'exécution.	4000000
spark.sql.Optimizer.Runtime.BloomFilter.NumberThreshold	Limite le nombre maximum de filtres d'exécution non DPP autorisés par requête afin d'éviter les out-of-memory erreurs dans le pilote.	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Définit le nombre de bits par défaut utilisés dans le filtre Bloom d'exécution.	8388608

Nom de propriété	Description	Valeur par défaut
spark.sql .optimizer.runtime .rowlevelO perationGroupFiltr e.activé	Spécifie s'il faut activer le filtrage des groupes d'exécution pour les opérations au niveau des lignes. Permet aux sources de données de : - Réduisez des groupes entiers de données (tels que des fichiers ou des partitions) à l'aide de filtres de source de données - Exécuter des requêtes d'exécution pour identifier les enregistrements correspondants - Supprimez les groupes inutiles pour éviter des réécritures coûteuses Limites: - Toutes les expressions ne peuvent pas être converties en filtres de source de données - Certaines expressions nécessitent une évaluation Spark (telles que les sous-requêtes)	TRUE
Spark.SQL .Optimizer.Runtime Filter. semiJoinR eduction.activé	Spécifie s'il faut insérer une semi-jointure pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	FALSE

Nom de propriété	Description	Valeur par défaut
spark.sql.parquet. Aggregate Pushdown	Spécifie s'il faut transférer les agrégats vers Parquet à des fins d'optimisation. Supporte : • MIN et MAX pour les types booléen, entier, flottant et date • COUNT pour tous les types de données Lance une exception si les statistiques sont absentes du pied de page d'un fichier Parquet.	FALSE
spark.sql.parquet. columnarR eaderBatchTaille	Contrôle le nombre de lignes dans chaque lot de lecteurs vectorisés Parquet. Choisissez une valeur qui équilibre le surcoût des performances et l'utilisation de la mémoire afin d'éviter out-of-memory les erreurs.	4096
spark.sql.session. TimeZone	Définit le fuseau horaire de session pour la gestion des horodatages sous forme de chaînes littérales et pour la conversion d'objets Java. Accepte : • area/city Format basé sur la région IDs (tel que America/Los_Angeles) • Décalages de zone au format (+/-) HH, (+/-) HH:mm ou (+/-) HH:mm:ss (tel que -08 ou + 01:00) • UTC ou Z comme alias pour + 00:00	(valeur du fuseau horaire local)

Nom de propriété	Description	Valeur par défaut
spark.sql.shuffle.partitions	Définit le nombre de partitions par défaut pour le brassage des données lors des jointures ou des agrégations. Ne peut pas être modifié entre les redémarrages de requêtes de streaming structurées à partir du même emplacement de point de contrôle.	200
spark.sql.shuffleJoinFactor	Définit le facteur de multiplication utilisé pour déterminer l'éligibilité à la jointure par hachage aléatoire. Une jointure par hachage aléatoire est sélectionnée lorsque la taille des données du petit côté multipliée par ce facteur est inférieure à la taille des données du grand côté.	3
spark.sql.sources.parallelPartitionDiscovery.seuil	Définit le nombre maximum de chemins pour la liste des fichiers côté pilote avec des sources basées sur des fichiers (Parquet, JSON et ORC). En cas de dépassement lors de la découverte de partitions, les fichiers sont répertoriés à l'aide d'une tâche distribuée Spark distincte.	32
spark.sql.statistics.histogram.enabled	Spécifie s'il faut générer des histogrammes à hauteur égale lors du calcul des statistiques des colonnes afin d'améliorer la précision de l'estimation. Nécessite une analyse de table supplémentaire en plus de celle requise pour les statistiques de base sur les colonnes.	FALSE

- c. Pour Conservation des données en jours, entrez le nombre de jours pendant lesquels les données sont conservées.

- d. Pour le format des résultats, choisissez CSV ou Parquet comme format de données que le canal d'entrée ML doit utiliser.
8. Pour l'accès au service, choisissez le nom du rôle de service existant qui sera utilisé pour accéder à cette table ou choisissez Créer et utiliser un nouveau rôle de service.
9. Pour le chiffrement, choisissez le secret de chiffrement avec une clé KMS personnalisée pour spécifier votre propre clé KMS et les informations associées. Dans le cas contraire, Clean Rooms ML gérera le chiffrement.
10. Choisissez Create ML input channel.

La création du canal d'entrée ML prendra quelques minutes. Vous pouvez voir une liste des canaux d'entrée ML dans l'onglet Modèles ML.

Note

Une fois le canal d'entrée ML créé, vous ne pouvez pas le modifier.

API

Pour créer un canal d'entrée ML (API)

Exécutez le code suivant avec vos paramètres spécifiques :

```
import boto3
acr_client = boto3.client('cleanroomsm1')

acr_client.create_ml_input_channel(
    name="ml_input_channel_name",
    membershipIdentifier='membership_id',

    configuredModelAlgorithmAssociations=[configured_model_algorithm_association_arn],
    retentionInDays=1,
    inputChannel={
        "dataSource": {
            "protectedQueryInputParameters": {
                "sqlParameters": {
                    "queryString": "select * from table",
                    "computeConfiguration": {
                        "worker": {
```

```

        "type": "CR.1X",
        "number": 16,
        "properties": {
            "spark": {
                "spark configuration key": "spark configuration
value",
            }
        }
    },
    "resultFormat": "PARQUET"
}
},
"roleArn": "arn:aws:iam::111122223333:role/role_name"
}
)
channel_arn = resp['ML Input Channel ARN']

```

Création d'un modèle entraîné dans AWS Clean Rooms ML

Prérequis :

- Et Compte AWS avec accès à AWS Clean Rooms
- Une collaboration mise en place dans AWS Clean Rooms
- Un algorithme de modèle configuré associé à la collaboration
- Au moins un canal d'entrée ML configuré
- Autorisations appropriées pour créer et gérer des modèles de machine learning dans le cadre de la collaboration

Après avoir associé l'algorithme de modèle configuré à une collaboration, puis créé et configuré un canal d'entrée ML, vous êtes prêt à créer un modèle entraîné. Un modèle entraîné est utilisé par les membres d'une collaboration pour analyser conjointement leurs données.

Vous pouvez créer un modèle entraîné à l'aide de la procédure suivante.

Vous pouvez également utiliser l'entraînement incrémentiel pour améliorer un modèle existant avec de nouvelles données, ou l'entraînement distribué pour entraîner des modèles sur plusieurs instances de calcul.

Rubriques

- [Utilisation de la formation progressive dans AWS Clean Rooms ML](#)
- [Utilisation de la formation distribuée dans AWS Clean Rooms ML](#)

Console

Pour créer un modèle entraîné (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, choisissez la collaboration pour laquelle vous souhaitez créer un modèle entraîné.
4. Une fois la collaboration ouverte, choisissez l'onglet Modèles ML.
5. Sous Modèles ML personnalisés, dans la section Modèles entraînés, sélectionnez Créer un modèle entraîné.
6. Sur la page Créer un modèle entraîné, dans Algorithme de modèle associé, spécifiez l'algorithme.
7. Pour obtenir des informations sur le modèle entraîné, entrez les informations suivantes :
 - a. Dans Nom, entrez un nom unique pour le modèle de la collaboration.
 - b. (Facultatif) Dans Description, entrez une description du modèle entraîné.
 - c. Pour le mode de saisie des données d'entraînement, choisissez l'une des options suivantes :
 - Sélectionnez Fichier si vous disposez d'un ensemble de données plus petit pouvant tenir sur le volume de stockage ML et si vous préférez un accès au système de fichiers traditionnel pour votre script de formation.
 - Sélectionnez Pipe pour les grands ensembles de données afin de diffuser des données directement depuis S3, évitant ainsi d'avoir à tout télécharger sur disque, ce qui peut améliorer la vitesse d'entraînement et réduire les besoins en stockage.
 - Choisissez FastFiles si vous souhaitez combiner les avantages du streaming depuis S3 avec l'accès au système de fichiers, en particulier pour la lecture séquentielle des données ou lorsque vous traitez moins de fichiers pour des temps de démarrage plus rapides.

8. Pour les détails du canal d'entrée ML, procédez comme suit :

- a. Pour le canal d'entrée ML, spécifiez le canal d'entrée ML qui fournit des données à l'algorithme du modèle.

Pour ajouter un autre canal, choisissez Ajouter un autre canal d'entrée ML. Vous pouvez ajouter jusqu'à 19 canaux d'entrée ML supplémentaires.

- b. Dans Nom du canal, entrez le nom du canal d'entrée ML.

- c. Pour le type de distribution de données Amazon S3, choisissez l'une des options suivantes :

- Sélectionnez Entièrement répliqué pour fournir à chaque instance de formation une copie complète de votre ensemble de données. Cela fonctionne mieux lorsque votre ensemble de données est suffisamment petit pour tenir en mémoire ou lorsque chaque instance doit accéder à toutes les données.
- Sélectionnez Sharded by S3 key pour diviser votre ensemble de données entre les instances d'entraînement en fonction des clés S3. Chaque instance reçoit environ $1/n$ du total des objets S3, où « n » est le nombre d'instances. Cela fonctionne mieux pour les grands ensembles de données que vous souhaitez traiter en parallèle.

 Note

Tenez compte de la taille de votre jeu de données et des exigences de formation lorsque vous sélectionnez un type de distribution. La réplication complète fournit un accès complet aux données mais nécessite davantage de stockage, tandis que la clé Sharded by S3 permet le traitement distribué de grands ensembles de données.

9. Pour Durée d'entraînement maximale, choisissez la durée maximale pendant laquelle vous souhaitez entraîner votre modèle.
10. Pour les hyperparamètres, spécifiez les paramètres spécifiques à l'algorithme et leurs valeurs prévues. Les hyperparamètres sont spécifiques au modèle en cours d'entraînement et sont utilisés pour affiner l'entraînement du modèle.
11. Pour les variables d'environnement, spécifiez les variables spécifiques à l'algorithme et leurs valeurs prévues. Les variables d'environnement sont définies dans le conteneur Docker.

12. Pour le chiffrement, pour utiliser une option personnalisée AWS KMS key, cochez la case Chiffrer le secret avec une clé KMS personnalisée.
13. Pour la configuration des ressources EC2, spécifiez les informations relatives aux ressources de calcul utilisées pour l'entraînement des modèles.
 - a. Dans Type d'instance, choisissez le type d'instance que vous souhaitez exécuter.
 - b. Dans Nombre d'instances, entrez le nombre d'instances.
 - c. Pour Taille du volume en Go, entrez la taille du volume de stockage ML.
14. Choisissez Créer un modèle entraîné.

API

Pour créer un modèle entraîné (API)

Le membre capable d'entraîner un modèle commence l'entraînement en sélectionnant le canal d'entrée ML et l'algorithme du modèle.

Exécutez le code suivant avec vos paramètres spécifiques :

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_trained_model(
    membershipIdentifier= 'membership_id',
    configuredModelAlgorithmAssociationArn = 'arn:aws:cleanrooms-
ml:region:account:membership/membershipIdentifier/configured-model-algorithm-
association/identifier',
    name='trained_model_name',
    resourceConfig={
        'instanceType': "ml.m5.xlarge",
        'volumeSizeInGB': 1
    },
    dataChannels=[
        {
            "mlInputChannelArn": channel_arn_1,
            "channelName": "channel_name"
        },
        {
            "mlInputChannelArn": channel_arn_2,
            "channelName": "channel_name"
        }
    ]
)
```

```
}  
]  
)
```

Note

Une fois le modèle entraîné créé, vous ne pouvez pas le modifier. Pour apporter des modifications, supprimez le modèle entraîné et créez-en un nouveau.

Utilisation de la formation progressive dans AWS Clean Rooms ML

Prérequis :

- Et Compte AWS avec accès à AWS Clean Rooms
- Un modèle formé existant dans le cadre d'une collaboration
- Un ensemble de données nouveau ou mis à jour pour un entraînement progressif
- Autorisations appropriées pour créer et gérer des modèles de machine learning dans le cadre de la collaboration
- Connaissance des hyperparamètres et de la configuration du modèle existant

Grâce à l'entraînement incrémentiel, vous pouvez utiliser les artefacts d'un modèle existant et d'un jeu de données mis à jour pour entraîner un nouveau modèle. L'entraînement incrémentiel permet de gagner du temps et des ressources.

Employez l'entraînement incrémentiel pour :

- Entraînez un nouveau modèle à l'aide d'un ensemble de données étendu dont le schéma sous-jacent n'avait pas été pris en compte lors de l'entraînement précédent.
- Entraînez plusieurs variantes d'un modèle, soit avec différents hyperparamètres, soit en utilisant différents ensembles de données.

Console

Pour exécuter une tâche d'entraînement incrémentielle (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, choisissez la collaboration dans laquelle se trouvent les artefacts du modèle que vous souhaitez utiliser pour une formation incrémentielle.
4. Une fois la collaboration ouverte, choisissez l'onglet Modèles ML.
5. Sous Modèles ML personnalisés, dans la section Modèles entraînés, cliquez sur le bouton radio à côté du modèle entraîné que vous souhaitez entraîner progressivement.
6. Sur la page Vue d'ensemble, sous Versions,
 - a. Cliquez sur le bouton radio situé à côté du modèle entraîné que vous souhaitez entraîner progressivement.
 - b. Choisissez Train depuis la version.
7. Sur la page Créer un modèle entraîné à partir de la version, pour Version du modèle entraîné, sélectionnez la version.

La version du modèle de base est automatiquement sélectionnée. Vous pouvez modifier cette version si d'autres versions existent.

8. Pour obtenir des informations sur le modèle entraîné, entrez les informations suivantes :
 - a. Dans Nom, entrez un nom unique pour le modèle de la collaboration.
 - b. (Facultatif) Dans Description, entrez une description du modèle entraîné.
 - c. Pour le mode de saisie des données d'entraînement, choisissez l'une des options suivantes :
 - Sélectionnez Fichier si vous disposez d'un ensemble de données plus petit pouvant tenir sur le volume de stockage ML et si vous préférez un accès au système de fichiers traditionnel pour votre script de formation.
 - Sélectionnez Pipe pour les grands ensembles de données afin de diffuser des données directement depuis S3, évitant ainsi d'avoir à tout télécharger sur disque, ce qui peut améliorer la vitesse d'entraînement et réduire les besoins en stockage.

- Choisissez FastFiles si vous souhaitez combiner les avantages du streaming depuis S3 avec l'accès au système de fichiers, en particulier pour la lecture séquentielle des données ou lorsque vous traitez moins de fichiers pour des temps de démarrage plus rapides.
- d. Pour Nom du canal d'entraînement incrémental, entrez un nom pour le canal d'entraînement incrémental

 Note

Si vous spécifiez le nom du canal d'entraînement incrémental sans numéro de version, le système utilise le modèle de base pour l'entraînement incrémentiel.

9. Pour les détails du canal d'entrée ML, procédez comme suit :

- a. Pour le canal d'entrée ML, spécifiez le canal d'entrée ML qui fournit des données à l'algorithme du modèle.

Pour ajouter un autre canal, choisissez Ajouter un autre canal d'entrée ML. Vous pouvez ajouter jusqu'à 19 canaux d'entrée ML supplémentaires.

- b. Dans Nom du canal, entrez le nom du canal d'entrée ML.
- c. Pour le type de distribution de données Amazon S3, choisissez l'une des options suivantes :
- Sélectionnez Entièrement répliqué pour fournir à chaque instance de formation une copie complète de votre ensemble de données. Cela fonctionne mieux lorsque votre ensemble de données est suffisamment petit pour tenir en mémoire ou lorsque chaque instance doit accéder à toutes les données.
 - Sélectionnez Sharded by S3 key pour diviser votre ensemble de données entre les instances d'entraînement en fonction des clés S3. Chaque instance reçoit environ $1/n$ du total des objets S3, où « n » est le nombre d'instances. Cela fonctionne mieux pour les grands ensembles de données que vous souhaitez traiter en parallèle.

 Note

Tenez compte de la taille de votre jeu de données et des exigences de formation lorsque vous sélectionnez un type de distribution. La réplication complète fournit un accès complet aux données mais nécessite davantage de stockage, tandis

que la clé Sharded by S3 permet le traitement distribué de grands ensembles de données.

10. Pour Durée d'entraînement maximale, choisissez la durée maximale pendant laquelle vous souhaitez entraîner votre modèle.
11. Pour les hyperparamètres, spécifiez les paramètres spécifiques à l'algorithme et leurs valeurs prévues. Les hyperparamètres sont spécifiques au modèle en cours d'entraînement et sont utilisés pour affiner l'entraînement du modèle.
12. Pour les variables d'environnement, spécifiez les variables spécifiques à l'algorithme et leurs valeurs prévues. Les variables d'environnement sont définies dans le conteneur Docker.
13. Pour le chiffrement, pour utiliser une option personnalisée AWS KMS key, cochez la case Chiffrer le secret avec une clé KMS personnalisée.
14. Pour la configuration des ressources EC2, spécifiez les informations relatives aux ressources de calcul utilisées pour l'entraînement des modèles.
 - a. Dans Type d'instance, choisissez le type d'instance que vous souhaitez exécuter.
 - b. Dans Nombre d'instances, entrez le nombre d'instances.
 - c. Pour Taille du volume en Go, entrez la taille du volume de stockage ML.
15. Choisissez Créer un modèle entraîné à partir de la version.

API

Pour exécuter une tâche de formation incrémentielle (API)

Exécutez le code suivant avec vos paramètres spécifiques :

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_trained_model(
    membershipIdentifier= 'membership_id',
    configuredModelAlgorithmAssociationArn = 'arn:aws:cleanrooms-
ml:region:account:membership/membershipIdentifier/configured-model-algorithm-
association/identifier',
    name='trained_model_name',
    resourceConfig={
        'instanceType': 'ml.m5.xlarge',
        'volumeSizeInGB': 1
```

```
},
incrementalTrainingDataChannels=[
  {
    'trainedModelArn': trained_model_arn,
    'channelName': 'channel_name'
  },
]
dataChannels=[
  {
    'mlInputChannelArn': channel_arn_1,
    'channelName': 'channel_name'
  },
  {
    'mlInputChannelArn': channel_arn_2,
    'channelName': 'channel_name'
  }
]
)
```

 Note

Limite : maximum de 20 canaux au total (y compris dataChannels les deux incrementalTrainingDataChannels).

 Note

Une fois le modèle entraîné créé, vous ne pouvez pas le modifier. Pour apporter des modifications, supprimez le modèle entraîné et créez-en un nouveau.

Utilisation de la formation distribuée dans AWS Clean Rooms ML

Prérequis :

- Et Compte AWS avec accès à AWS Clean Rooms
- Une collaboration mise en place dans AWS Clean Rooms
- Un algorithme de modèle configuré qui prend en charge la formation distribuée
- Un ensemble de données volumineux adapté au traitement distribué

- Autorisations appropriées pour créer et gérer des modèles de machine learning dans le cadre de la collaboration
- Quota Amazon EC2 suffisant pour exécuter plusieurs instances dans le cadre d'une formation distribuée

La formation distribuée tire parti de la puissance de nombreux nœuds informatiques travaillant en parallèle pour traiter de grandes quantités de données et mettre à jour les paramètres du modèle de manière efficace.

Pour plus d'informations sur la formation distribuée, consultez la section [Concepts de formation distribuée](#) dans le manuel Amazon SageMaker AI Developer Guide.

Console

Pour exécuter une tâche de formation distribuée (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, choisissez la collaboration pour laquelle vous souhaitez créer un modèle entraîné.
4. Une fois la collaboration ouverte, choisissez l'onglet Modèles ML.
5. Sous Modèles ML personnalisés, dans la section Modèles entraînés, sélectionnez Créer un modèle entraîné.
6. Sur la page Créer un modèle entraîné, dans Algorithme de modèle associé, spécifiez l'algorithme
7. Pour obtenir des informations sur le modèle entraîné, entrez les informations suivantes :
 - a. Dans Nom, entrez un nom unique pour le modèle de la collaboration.
 - b. (Facultatif) Dans Description, entrez une description du modèle entraîné.
 - c. Pour le mode de saisie des données d'entraînement, choisissez l'une des options suivantes :
 - Sélectionnez Fichier si vous disposez d'un ensemble de données plus petit pouvant tenir sur le volume de stockage ML et si vous préférez un accès au système de fichiers traditionnel pour votre script de formation.

- Sélectionnez Pipe pour les grands ensembles de données afin de diffuser des données directement depuis S3, évitant ainsi d'avoir à tout télécharger sur disque, ce qui peut améliorer la vitesse d'entraînement et réduire les besoins en stockage.
- Choisissez FastFiles si vous souhaitez combiner les avantages du streaming depuis S3 avec l'accès au système de fichiers, en particulier pour la lecture séquentielle des données ou lorsque vous traitez moins de fichiers pour des temps de démarrage plus rapides.

8. Pour les détails du canal d'entrée ML, procédez comme suit :

- a. Pour le canal d'entrée ML, spécifiez le canal d'entrée ML qui fournit des données à l'algorithme du modèle.

Pour ajouter un autre canal, choisissez Ajouter un autre canal d'entrée ML. Vous pouvez ajouter jusqu'à 19 canaux d'entrée ML supplémentaires.

- b. Dans Nom du canal, entrez le nom du canal d'entrée ML.

- c. Pour le type de distribution de données Amazon S3, choisissez l'une des options suivantes :

- Sélectionnez Entièrement répliqué pour fournir à chaque instance de formation une copie complète de votre ensemble de données. Cela fonctionne mieux lorsque votre ensemble de données est suffisamment petit pour tenir en mémoire ou lorsque chaque instance doit accéder à toutes les données.
- Sélectionnez Sharded by S3 key pour diviser votre ensemble de données entre les instances d'entraînement en fonction des clés S3. Chaque instance reçoit environ $1/n$ du total des objets S3, où « n » est le nombre d'instances. Cela fonctionne mieux pour les grands ensembles de données que vous souhaitez traiter en parallèle.

 Note

Tenez compte de la taille de votre jeu de données et des exigences de formation lorsque vous sélectionnez un type de distribution. La réplication complète fournit un accès complet aux données mais nécessite davantage de stockage, tandis que la clé Sharded by S3 permet le traitement distribué de grands ensembles de données.

9. Pour Durée d'entraînement maximale, choisissez la durée maximale pendant laquelle vous souhaitez entraîner votre modèle.
10. Pour les hyperparamètres, spécifiez les paramètres spécifiques à l'algorithme et leurs valeurs prévues. Les hyperparamètres sont spécifiques au modèle en cours d'entraînement et sont utilisés pour affiner l'entraînement du modèle.
11. Pour les variables d'environnement, spécifiez les variables spécifiques à l'algorithme et leurs valeurs prévues. Les variables d'environnement sont définies dans le conteneur Docker.
12. Pour le chiffrement, pour utiliser une option personnalisée AWS KMS key, cochez la case Chiffrer le secret avec une clé KMS personnalisée.
13. Pour la configuration des ressources EC2, spécifiez les informations relatives aux ressources de calcul utilisées pour l'entraînement des modèles.
 - a. Dans Type d'instance, choisissez le type d'instance que vous souhaitez exécuter.

Les types d'instances pris en charge pour la formation distribuée sont les suivants :

- ml.m5.4xlarge
- ml.m5.12xlarge
- ml.m5.2xlarge
- ml.g5.12xlarge
- ml.g5.24xlarge

- b. Dans Nombre d'instances, entrez le nombre d'instances.
- c. Pour Taille du volume en Go, entrez la taille du volume de stockage ML.

14. Choisissez Créer un modèle entraîné.

API

Pour exécuter une tâche de formation distribuée (API)

Exécutez le code suivant avec vos paramètres spécifiques :

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

acr_ml_client.create_trained_model(
    membershipIdentifier= 'membership_id',
```

```

    configuredModelAlgorithmAssociationArn = 'arn:aws:cleanrooms-
ml:region:account:membership/membershipIdentifier/configured-model-algorithm-
association/identifier',
    name='trained_model_name',
    trainingInputMode: "File",
    resourceConfig={
        'instanceCount': "3"
        'instanceType': "ml.m5.xlarge",
        'volumeSizeInGB': 3
    },
    dataChannels=[
        {
            "mlInputChannelArn": channel_arn_1,
            "channelName": "channel_name",
            "S3DataDistributionType": "FullyReplicated"
        }
    ]
)

```

Note

Une fois le modèle entraîné créé, vous ne pouvez pas le modifier. Pour apporter des modifications, supprimez le modèle entraîné et créez-en un nouveau.

Exportation d'artefacts de modèles depuis AWS Clean Rooms ML

Cette tâche est facultative et doit être terminée lorsque vous avez attribué la capacité de CAN_RECEIVE_MODEL_OUTPUT membre à un membre de la collaboration.

Une fois la formation du modèle terminée, le membre qui a entraîné le modèle peut lancer l'exportation des artefacts du modèle. Le membre qui a entraîné le modèle choisit qui recevra les artefacts du modèle, à condition que ce membre puisse recevoir des résultats et une configuration ML valide.

Console

Pour configurer un algorithme de modèle ML personnalisé (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, choisissez la collaboration qui contient le modèle personnalisé que vous souhaitez exporter.
4. Une fois la collaboration ouverte, choisissez l'onglet Modèles ML, puis choisissez votre modèle dans le tableau des modèles entraînés personnalisés
5. Sur la page de détails du modèle entraîné personnalisé, cliquez sur Exporter la sortie du modèle.
6. Pour Exporter la sortie du modèle, pour Exporter les détails de sortie du modèle, entrez un nom et une description facultative.

Choisissez le membre qui recevra les artefacts du modèle dans la liste déroulante Sortie du modèle exportée vers les membres de la collaboration.

7. Cliquez sur Exporter.

Les résultats sont exportés vers le chemin suivant dans l'emplacement Amazon S3 spécifié dans la configuration ML :`yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`. Seuls les fichiers à exporter, jusqu'à la taille de fichier maximale spécifiée, que vous avez sélectionnés lors de l'association de l'algorithme de modèle configuré sont exportés.

API

Pour configurer un algorithme de modèle ML (API) personnalisé

Lancez l'exportation du modèle en exécutant le code suivant :

```
import boto3
acr_ml_client= boto3.client('cleanroomsmml')

acr_ml_client.start_trained_model_export_job(
    membershipIdentifier='membership_id',
```

```
trainedModelArn='arn:aws:cleanrooms-ml:region:account:membership/
membershipIdentifier/trained-model/identifier',
outputConfiguration={
  'member': {
    'accountId': 'model_output_receiver_account'
  }
},
name='export_job_name'
)
```

Les résultats sont exportés vers le chemin suivant dans l'emplacement Amazon S3 spécifié dans la configuration ML : `yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`. Seul l'`filesToExport` algorithme de modèle configuré que vous avez sélectionné lors de l'association de l'algorithme de modèle configuré est exporté, dans la limite de ce qui est `maxSize` spécifié.

Exécution d'une inférence sur un modèle entraîné dans AWS Clean Rooms ML

Les membres capables d'exécuter des requêtes peuvent également lancer une tâche d'inférence une fois la tâche de formation terminée. Ils choisissent le jeu de données d'inférence par rapport auquel ils souhaitent exécuter l'inférence et font référence aux sorties du modèle entraîné avec lesquelles ils souhaitent exécuter le conteneur d'inférence.

Le membre qui recevra les résultats d'inférence doit avoir la capacité `CAN_RECEIVE_INFERENCE_OUTPUT` de membre.

Console

Pour créer une tâche d'inférence de modèle (console)

1. Connectez-vous à la AWS Clean Rooms console AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/cleanrooms>.
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Sur la page Collaborations, choisissez la collaboration contenant le modèle personnalisé sur lequel vous souhaitez créer une tâche d'inférence.
4. Une fois la collaboration ouverte, choisissez l'onglet Modèles ML, puis choisissez votre modèle dans le tableau des modèles entraînés personnalisés.

5. Sur la page de détails du modèle entraîné personnalisé, cliquez sur Démarrer la tâche d'inférence.
6. Pour Démarrer la tâche d'inférence, pour les détails de la tâche d'inférence, entrez un nom et une description facultative.

Entrez les informations suivantes :

- Algorithme de modèle associé : algorithme de modèle associé utilisé pendant le travail d'inférence.
 - Détails du canal d'entrée ML — Le canal d'entrée ML qui fournira les données pour cette tâche d'inférence.
 - Ressources de transformation : instance de calcul utilisée pour exécuter la fonction de transformation de la tâche d'inférence.
 - Configuration de sortie — Qui recevra la sortie de la tâche d'inférence et le type MIME de la sortie.
 - Chiffrement — Choisissez l'option Personnaliser les paramètres de chiffrement pour spécifier votre propre clé KMS et les informations associées. Dans le cas contraire, Clean Rooms ML gérera le chiffrement.
 - Détails de la tâche de transformation : charge utile maximale de la tâche d'inférence, en Mo.
 - Variables d'environnement : toutes les variables d'environnement nécessaires pour accéder à l'image du conteneur de la tâche d'inférence.
7. Choisissez Démarrer la tâche d'inférence.

Les résultats sont exportés vers le chemin suivant dans l'emplacement Amazon S3 spécifié dans la configuration ML : `yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`.

API

Pour créer une tâche d'inférence de modèle (API)

Lancez la tâche d'inférence en exécutant le code suivant :

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')
```

```
acr_ml_client.start_trained_model_inference_job(  
    name="inference_job",  
    membershipIdentifier='membership_id',  
    trainedModelArn='arn:aws:cleanrooms-ml:region:account:membership/  
membershipIdentifier/trained-model/identifier',  
  
    dataSource={  
        "mlInputChannelArn": 'channel_arn_3'  
    },  
    resourceConfig={'instanceType': 'ml.m5.xlarge'},  
    outputConfiguration={  
        'accept': 'text/csv',  
        'members': [  
            {  
                "accountId": 'member_account_id'  
            }  
        ]  
    }  
)
```

Les résultats sont exportés vers le chemin suivant dans l'emplacement Amazon S3 spécifié dans la configuration ML :yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName.

Créez des modèles AWS Clean Rooms ML en tant que fournisseur de données de formation

Un modèle de similarité est un modèle de données d'un fournisseur de données de formation qui permet à un fournisseur de données de départ de créer un segment similaire des données du fournisseur de données de formation qui ressemble le plus à ses données de départ. Pour créer un modèle de similarité utilisable dans une collaboration, vous devez importer vos données de formation, créer un modèle de similarité, configurer ce modèle de similarité, puis l'associer à une collaboration.

L'utilisation de modèles similaires nécessite que deux parties, un fournisseur de données de formation et un fournisseur de données de départ, travaillent de manière séquentielle AWS Clean Rooms pour intégrer leurs données dans une collaboration. Voici le flux de travail que le fournisseur de données de formation doit effectuer en premier :

1. Les données du fournisseur de données de formation doivent être stockées dans une table de catalogue de AWS Glue données répertoriant les interactions entre les utilisateurs et les éléments. Les données d'entraînement doivent au minimum contenir une colonne d'ID utilisateur, une colonne d'identifiant d'interaction et une colonne d'horodatage.
2. Le fournisseur de données de formation enregistre les données de formation auprès de AWS Clean Rooms.
3. Le fournisseur de données de formation crée un modèle similaire qui peut être partagé avec plusieurs fournisseurs de données initiales. Le modèle similaire est un réseau neuronal profond dont l'entraînement peut prendre jusqu'à 24 heures. Il n'est pas automatiquement réentraîné et nous vous recommandons de le réentraîner chaque semaine.
4. Le fournisseur de données de formation configure le modèle de similarité, notamment en indiquant s'il convient de partager les indicateurs de pertinence et l'emplacement des segments de sortie sur Amazon S3. Le fournisseur de données de formation peut créer plusieurs modèles similaires configurés à partir d'un seul modèle similaire.
5. Le fournisseur de données de formation associe le modèle d'audience configuré à une collaboration partagée avec un fournisseur de données de départ.

Une fois que le fournisseur de données de formation a fini de créer le modèle ML, [le fournisseur de données de départ peut créer et exporter le segment similaire](#).

Rubriques

- [Importation de données d'entraînement](#)
- [Création d'un modèle similaire](#)
- [Configuration d'un modèle similaire](#)
- [Associer un modèle de similarité configuré](#)
- [Mise à jour d'un modèle similaire configuré](#)

Importation de données d'entraînement

Note

Vous pouvez uniquement fournir un ensemble de données d'entraînement à utiliser dans un modèle similaire à Clean Rooms ML dont les données sont stockées dans Amazon S3. Toutefois, vous pouvez fournir les données de départ d'un modèle similaire à l'aide de SQL qui analyse les données stockées dans n'importe quelle source de données prise en charge.

Avant de créer un modèle similaire, vous devez spécifier la AWS Glue table contenant les données d'entraînement. Clean Rooms ML ne stocke pas de copie de ces données, mais uniquement des métadonnées qui lui permettent d'accéder aux données.

Pour importer des données d'entraînement dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si ce n'est pas encore fait).
2. Dans le volet de navigation de gauche, choisissez AWS ML models.
3. Dans l'onglet Ensembles de données d'entraînement, choisissez Créer un jeu de données d'entraînement.
4. Sur la page Créer un jeu de données d'entraînement, pour les détails du jeu de données d'entraînement, entrez un nom et une description facultative.
5. Choisissez la source de données d'entraînement en sélectionnant la base de données et la table que vous souhaitez configurer dans les listes déroulantes.

Note

Pour vérifier que ce tableau est correct, effectuez l'une des opérations suivantes :

- Choisissez Afficher dans AWS Glue.
- Activez Afficher le schéma pour afficher le schéma.

6. Pour les détails de la formation, choisissez la colonne Identifiant utilisateur, la colonne Identifiant de l'article et la colonne Horodatage dans les listes déroulantes. Les données d'entraînement doivent contenir ces trois colonnes. Vous pouvez également sélectionner les autres colonnes que vous souhaitez inclure dans les données d'entraînement.

Les données de la colonne Horodatage doivent être au format Unix Epoch en secondes.

7. (Facultatif) Si vous avez des colonnes supplémentaires à entraîner, choisissez le nom et le type de colonne dans les listes déroulantes.
8. Dans Accès aux services, vous devez spécifier un rôle de service qui peut accéder à vos données et fournir une clé KMS si vos données sont chiffrées. Choisissez Créer et utiliser un nouveau rôle de service et Clean Rooms ML créera automatiquement un rôle de service et ajoutera la politique d'autorisation nécessaire. Choisissez Utiliser un rôle de service existant et saisissez-le dans le champ Nom du rôle de service si vous souhaitez utiliser un rôle de service spécifique.

Si vos données sont chiffrées, entrez votre clé KMS dans le AWS KMS keychamp ou cliquez sur Créer une AWS KMS key pour générer une nouvelle clé KMS.

9. Si vous souhaitez activer les balises pour le jeu de données d'entraînement, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
10. Choisissez Créer un jeu de données d'entraînement.

Pour l'action d'API correspondante, consultez [CreateTrainingDataset](#).

Création d'un modèle similaire

Après avoir créé un jeu de données d'entraînement, vous êtes prêt à créer un modèle similaire. Vous pouvez créer de nombreux modèles similaires à partir d'un seul jeu de données d'entraînement.

Vous devez créer une base de données par défaut dans votre rôle AWS Glue Data Catalog ou inclure `l'glue:createDatabaseautorisation` dans le rôle fourni.

Pour créer un modèle similaire dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si ce n'est pas encore fait).
2. Dans le volet de navigation de gauche, choisissez AWS ML models.
3. Dans l'onglet Modèles similaires, choisissez Créer un modèle similaire.
4. Sur la page Créer un modèle similaire, pour les détails du modèle similaire, entrez un nom et une description facultative.
 - a. Choisissez le jeu de données d'entraînement que vous souhaitez modéliser dans la liste déroulante.

Note

Pour vérifier qu'il s'agit du bon jeu de données d'entraînement, activez Afficher les détails du jeu de données d'entraînement pour afficher les détails.
Pour créer un nouveau jeu de données d'entraînement, choisissez Créer un jeu de données d'entraînement.

- b. (Facultatif) Entrez dans une fenêtre d'entraînement.
5. Si vous souhaitez activer les paramètres de chiffrement personnalisés pour le modèle similaire, choisissez Personnaliser les paramètres de chiffrement, puis entrez la clé KMS.
6. Si vous souhaitez activer les balises pour le modèle similaire, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
7. Choisissez Créer un modèle similaire.

Note

La formation des modèles peut prendre de plusieurs heures à deux jours.

Pour l'action d'API correspondante, consultez [CreateAudienceModel](#).

Configuration d'un modèle similaire

Après avoir créé un modèle similaire, vous êtes prêt à le configurer pour une utilisation dans le cadre d'une collaboration. Vous pouvez créer plusieurs modèles similaires configurés à partir d'un seul modèle similaire.

Pour configurer un modèle similaire dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si ce n'est pas encore fait).
2. Dans le volet de navigation de gauche, choisissez AWS ML models.
3. Dans l'onglet Modèles similaires configurés, choisissez Configurer le modèle similaire.
4. Sur la page Configurer le modèle similaire, pour les détails du modèle similaire configuré, entrez un nom et une description facultative.

- a. Choisissez le modèle Lookalike que vous souhaitez configurer dans la liste déroulante.

Note

Pour vérifier qu'il s'agit du bon modèle similaire, activez Afficher les détails du modèle similaire pour afficher les détails.

Pour créer un nouveau modèle similaire, choisissez Créer un modèle similaire.

- b. Choisissez la taille de graine minimale correspondante que vous souhaitez. Il s'agit du nombre minimum d'utilisateurs dans les données du fournisseur de données de départ qui se chevauchent avec les utilisateurs dans les données de formation. Cette valeur doit être supérieure à 0.
5. Pour que les métriques soient partagées avec d'autres membres, choisissez si vous souhaitez que le fournisseur de données de base de votre collaboration reçoive les métriques du modèle, y compris les scores de pertinence.
 6. Pour l'emplacement de destination du segment similaire, entrez le compartiment Amazon S3 dans lequel le segment similaire est exporté. Ce compartiment doit être situé dans la même région que vos autres ressources.
 7. Pour l'accès au service, choisissez le nom du rôle de service existant qui sera utilisé pour accéder à cette table.
 8. Pour la configuration avancée de la taille des bacs, spécifiez le type de taille de l'audience sous la forme d'un nombre absolu ou d'un pourcentage.

9. Si vous souhaitez activer les balises pour la ressource de table configurée, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
10. Choisissez Configurer le modèle similaire.

Pour l'action d'API correspondante, consultez [CreateConfiguredAudienceModel](#).

Associer un modèle de similarité configuré

Après avoir configuré un modèle similaire, vous pouvez l'associer à une collaboration.

Pour associer un modèle similaire configuré dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si ce n'est pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Dans l'onglet Avec adhésion active, choisissez une collaboration.
4. Dans l'onglet Modèles ML, sous Modèles easy-to-use similaires R, choisissez Associer un modèle similaire.
5. Sur la page Associer un modèle similaire configuré, pour les détails de l'association du modèle similaire configuré :
 - a. Entrez un nom pour le modèle d'audience configuré associé.
 - b. Entrez une description de la table.

La description permet de différencier les autres modèles d'audience configurés associés portant des noms similaires.

6. Pour Modèle similaire configuré, choisissez un modèle similaire configuré dans la liste déroulante.
7. Choisissez Associer.

Pour l'action d'API correspondante, consultez [CreateConfiguredAudienceModelAssociation](#).

Mise à jour d'un modèle similaire configuré

Après avoir associé un modèle similaire configuré, vous pouvez le mettre à jour pour modifier des informations telles que le nom, les métriques à partager ou la position Amazon S3 en sortie.

Pour mettre à jour un modèle similaire configuré associé dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si ce n'est pas encore fait).
2. Dans le volet de navigation de gauche, choisissez AWS ML models.
3. Dans l'onglet Modèles similaires configurés, sous Modèles Ready-to-use similaires, choisissez un modèle similaire configuré et sélectionnez Modifier.
4. Sur la page Modifier, pour les détails de l'association de modèles similaires configurés :
 - a. Mettez à jour le nom et la description facultative.
 - b. Choisissez le modèle Lookalike que vous souhaitez configurer dans la liste déroulante.
 - c. Choisissez la taille de graine minimale correspondante que vous souhaitez. Il s'agit du nombre minimum d'utilisateurs dans les données du fournisseur de données de départ qui se chevauchent avec les utilisateurs dans les données de formation. Cette valeur doit être supérieure à 0.
5. Pour que les métriques soient partagées avec d'autres membres, choisissez si vous souhaitez que le fournisseur de données de base de votre collaboration reçoive les métriques du modèle, y compris les scores de pertinence.
6. Pour l'emplacement de destination du segment Lookalike, entrez le compartiment Amazon S3 dans lequel le segment Lookalike est exporté. Ce compartiment doit être situé dans la même région que vos autres ressources.
7. Pour l'accès au service, choisissez le nom du rôle de service existant qui sera utilisé pour accéder à cette table.
8. Pour la configuration avancée de la taille des bacs, choisissez la manière dont vous souhaitez configurer les tailles des bacs d'audience.
9. Sélectionnez Enregistrer les modifications.

Pour l'action d'API correspondante, consultez [UpdateConfiguredAudienceModel](#).

Création de modèles AWS Clean Rooms ML en tant que fournisseur de données de base

Une fois que le fournisseur de données de formation a fini de créer le modèle ML, le fournisseur de données de départ peut créer et exporter le segment similaire. Le segment similaire est un sous-ensemble des données d'apprentissage qui ressemble le plus aux données de départ.

Voici le flux de travail que le fournisseur de données de départ doit effectuer :

1. Les données du fournisseur de données de base peuvent être stockées dans un compartiment Amazon S3 ou peuvent provenir des résultats d'une requête.
2. Le fournisseur de données de départ ouvre la collaboration qu'il partage avec le fournisseur de données de formation.
3. Le fournisseur de données de départ crée un segment similaire à partir de l'onglet Clean Rooms ML de la page de collaboration.
4. Le fournisseur de données de base peut évaluer les indicateurs de pertinence, s'ils ont été partagés, et exporter le segment similaire pour une utilisation en dehors AWS Clean Rooms.

Rubriques

- [Création d'un segment similaire](#)
- [Exportation d'un segment similaire](#)

Création d'un segment similaire

Note

Vous pouvez uniquement fournir un ensemble de données d'entraînement à utiliser dans un modèle similaire à Clean Rooms ML dont les données sont stockées dans Amazon S3. Cependant, vous pouvez fournir les données de départ d'un modèle similaire à l'aide de SQL qui analyse les données stockées dans n'importe quelle source de données prise en charge.

Un segment similaire est un sous-ensemble des données d'apprentissage qui ressemble le plus aux données de départ.

Pour créer un segment similaire dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Dans l'onglet Avec adhésion active, choisissez une collaboration.
4. Dans l'onglet ML Models, choisissez Create lookalike segment.
5. Sur la page Créer un segment de similitude, pour Modèle de similitude configuré associé, choisissez le modèle de similitude configuré associé à utiliser pour ce segment de similitude.
6. Pour les détails du segment Lookalike, entrez un nom et une description facultative.
7. Pour les profils Seed, choisissez votre méthode Seed en sélectionnant une option, puis en prenant l'action recommandée.

Option	Action recommandée
Chemin Amazon S3	1. Sélectionnez un emplacement Amazon S3. 2. (Facultatif) Choisissez Inclure les profils de départ dans la sortie.
Requête SQL	Rédigez une requête SQL et utilisez ses résultats comme données de départ.
Modèle d'analyse	Choisissez un modèle d'analyse dans la liste déroulante et utilisez les résultats créés par un modèle d'analyse.

8. Choisissez le type de travailleur à utiliser lors de la création de cette source de données. Le type de travailleur par défaut est CR.1X. Spécifiez le nombre de travailleurs à utiliser. La valeur par défaut est le travailleur numéro 16. Pour définir les propriétés de Spark :
 - a. Développez les propriétés Spark.
 - b. Choisissez Ajouter des propriétés Spark.
 - c. Dans la boîte de dialogue des propriétés de Spark, choisissez un nom de propriété dans la liste déroulante et entrez une valeur.

Le tableau suivant fournit une définition pour chaque propriété.

Pour plus d'informations sur les propriétés de Spark, consultez la section [Propriétés de Spark](#) dans la documentation d'Apache Spark.

Nom de propriété	Description	Valeur par défaut
Défaillances de Spark.Task.Max	Contrôle le nombre de fois consécutives qu'une tâche peut échouer avant que la tâche échoue. Nécessite une valeur supérieure ou égale à 1. Le nombre de tentatives autorisées est égal à cette valeur moins 1. Le nombre d'échecs est réinitialisé en cas de réussite d'une tentative. Les échecs liés aux différentes tâches ne s'accumulent pas jusqu'à atteindre cette limite.	4
fichiers spark.sql.maxPartitionBytes	Définit le nombre maximum d'octets à regrouper dans une seule partition lors de la lecture à partir de sources basées sur des fichiers telles que Parquet, JSON et ORC.	128 MO
Spark.Hadoop.FS.S3.max Réessaie	Définit le nombre maximal de tentatives pour les opérations sur les fichiers Amazon S3.	
spark.network.timeout	Définit le délai d'expiration par défaut pour toutes les interactions réseau. Remplace les paramètres de délai d'expiration suivants s'ils ne sont pas configurés : • stockage d'étincelles. blockManagerHeartbeatTimeoutMs • Spark.shuffle.io. Délai de connexion • Spark.rpc.AskTimeout • Spark.rpc. Délai de recherche expiré	années 120

Nom de propriété	Description	Valeur par défaut
spark.rdd.compress	Spécifie s'il faut compresser les partitions RDD sérialisées à l'aide de spark.io.compression.codec. S'applique à StorageLevel .MEMORY_ONLY_SER en Java et Scala, ou à .MEMORY_ONLY en Python. StorageLevel Réduit l'espace de stockage mais nécessite un temps de traitement supplémentaire du processeur.	FALSE
spark.shuffle.spill.compress	Spécifie s'il faut compresser les données du shuffle spill à l'aide de spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Définit la taille cible en octets pour les partitions shuffle lors de l'optimisation adaptative lorsque spark.sql.adaptive.enabled est vrai. Contrôle la taille de la partition lors de la fusion de petites partitions ou du fractionnement de partitions asymétriques.	(valeur de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)
spark.sql.adaptive.autoBroadcastJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. S'applique uniquement dans le cadre adaptatif. Utilise la même valeur par défaut que spark.sql.autoBroadcastJoinSeuil. Réglez sur -1 pour désactiver la diffusion.	(aucun)
spark.sql.adaptive.CoalescePartitions.Activé	Spécifie s'il faut fusionner les partitions shuffle contiguës sur la base de spark.sql.adaptive.advisoryPartitionSizeInBytes pour optimiser la taille des tâches. Nécessite que spark.sql.adaptive.enabled soit vrai.	TRUE

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive. .CoalescePartitions. initialPartitionNum	Définit le nombre initial de partitions shuffle avant la fusion. Nécessite que spark.sql.adaptive.enabled et spark.sql.adaptive.CoalescePartitions.enabled soient vrais. La valeur par défaut est spark.sql.shuffle.partitions.	(aucun)
spark.sql.adaptive. .CoalescePartitions. minPartitionSize	Définit la taille minimale des partitions shuffle coalescées afin d'éviter que les partitions ne deviennent trop petites lors de l'optimisation adaptative.	1 Mo
spark.sql.adaptive. .coalescepartition s.ParallelismFirst	Spécifie s'il faut calculer la taille des partitions en fonction du parallélisme des clusters plutôt que de spark.sql.adaptive.advisoryPartitionSizeInBytes pendant la coalescence de la partition. Génère des tailles de partition inférieures à la taille cible configurée afin de maximiser le parallélisme. Nous vous recommandons de définir ce paramètre sur false sur les clusters occupés afin d'améliorer l'utilisation des ressources en évitant de trop petites tâches.	TRUE
spark.sql.adaptive. .enabled	Spécifie s'il faut activer l'exécution adaptative des requêtes afin de réoptimiser les plans de requêtes pendant l'exécution des requêtes, sur la base de statistiques d'exécution précises.	TRUE
spark.sql.adaptive. forceOptimizeSkewed Joignez-vous	Spécifie s'il faut forcer l'activation OptimizeSkewedJoin même si cela introduit un shuffle supplémentaire.	FALSE

Nom de propriété	Description	Valeur par défaut
<code>spark.sql.adaptive.localShuffleReader</code> .activé	Spécifie s'il faut utiliser des lecteurs de shuffle locaux lorsque le partitionnement aléatoire n'est pas requis, par exemple après la conversion des jointures par tri-fusion en jointures par hachage par diffusion. Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	Définit la taille de partition maximale en octets pour la création de cartes de hachage locales. Privilégie les jointures par hachage mélangées par rapport aux jointures par tri-fusion lorsque : - Cette valeur est égale ou supérieure à <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> - Toutes les tailles de partition se situent dans cette limite Remplace <code>spark.sql.join.preferSortMergeParameter</code> de jointure.	0 octet
<code>spark.sql.adaptive.optimizeSkewsInRebalancePartitions</code> .activé	Spécifie s'il faut optimiser les partitions shuffle asymétriques en les divisant en partitions plus petites sur la base de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> . Nécessite que <code>spark.sql.adaptive.enabled</code> soit vrai.	TRUE
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	Définit le facteur de seuil de taille pour la fusion des partitions lors du fractionnement. Partitions inférieures à ce facteur multipliées par <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> sont fusionnées.	0.2

Nom de propriété	Description	Valeur par défaut
spark.sql.adaptive. .SkewJoin.Enabled	Spécifie s'il faut gérer l'inclinaison des données dans les jointures mélangées en divisant et éventuellement en répliquant les partitions asymétriques. S'applique aux jointures par tri-fusion et aux jointures de hachage mélangées. Nécessite que spark.sql.adaptive.enabled soit vrai.	TRUE
spark.sql.adaptive. .skewJoin. skewedPartitionFactor	Détermine le facteur de taille qui détermine l'inclinaison de la partition. Une partition est inclinée lorsque sa taille dépasse les deux : - Ce facteur multiplié par la taille médiane de la partition - La valeur de spark.sql.adaptive.SkewJoin.skewedPartitionThresholdInBytes	5
spark.sql.adaptive. .skewJoin. skewedPartitionThresholdInBytes	Définit le seuil de taille en octets pour identifier les partitions asymétriques. Une partition est inclinée lorsque sa taille dépasse les deux : - Ce seuil - Taille de partition médiane multipliée par spark.sql.adaptive.skewJoin.skewedPartitionFactor Nous recommandons de définir cette valeur à une valeur supérieure à spark.sql.adaptive.advisoryPartitionSizeInBytes.	256 MO
spark.sql. autoBroad castJoinSeuil	Définit la taille maximale de la table en octets pour la diffusion vers les nœuds de travail lors des jointures. Réglez sur -1 pour désactiver la diffusion.	10 MO

Nom de propriété	Description	Valeur par défaut
Spark.SQL. Délai de diffusion	Contrôle le délai d'expiration en secondes pour les opérations de diffusion lors des jointures de diffusion.	300 secondes
spark.sql.cbo.enabled	Spécifie s'il faut activer l'optimisation basée sur les coûts (CBO) pour l'estimation des statistiques du plan.	FALSE
spark.sql.cbo.joinreorder.dp.star.filter	Spécifie s'il faut appliquer l'heuristique du filtre de jointure en étoile lors de l'énumération des jointures basée sur les coûts.	FALSE
spark.sql.cbo.joinreorder.dp.Threshold	Définit le nombre maximum de nœuds joints autorisés dans l'algorithme de programmation dynamique.	12
spark.sql.cbo.JoinReorder.Enabled	Spécifie s'il faut activer la réorganisation des jointures dans le cadre de l'optimisation basée sur les coûts (CBO).	FALSE
spark.sql.cbo.planstats.Enabled	Spécifie s'il faut extraire le nombre de lignes et les statistiques des colonnes du catalogue lors de la génération du plan logique.	FALSE
spark.sql.cbo.starSchemaDetection	Spécifie s'il faut activer la réorganisation des jointures en fonction de la détection du schéma en étoile.	FALSE
fichiers spark.sql.maxPartitionNum	Définit le nombre maximal cible de partitions de fichiers fractionnées pour les sources basées sur des fichiers (Parquet, JSON et ORC). Redimensionne les partitions lorsque le nombre initial dépasse cette valeur. Il s'agit d'une cible suggérée et non d'une limite garantie.	(aucun)

Nom de propriété	Description	Valeur par défaut
fichiers spark.sql. . maxRecord sPerDossier	Définit le nombre maximum d'enregistrements à écrire dans un seul fichier. Aucune limite ne s'applique lorsqu'il est réglé sur zéro ou sur une valeur négative.	0
fichiers spark.sql. minPartitionNum	Définit le nombre minimum cible de partitions de fichiers fractionnées pour les sources basées sur des fichiers (Parquet, JSON et ORC). La valeur par défaut est spark.sql.leafNodeDefaultParallélisme. Il s'agit d'une cible suggérée et non d'une limite garantie.	(aucun)
spark.sql. inMemoryC olumnarStockage. Taille du lot	Contrôle la taille du lot pour la mise en cache en colonnes. L'augmentation de la taille améliore l'utilisation de la mémoire et la compression, mais augmente le risque d' out-of-memoryerreurs.	10 000
spark.sql. inMemoryC olumnarStockage compressé	Spécifie s'il faut sélectionner automatiquement les codecs de compression pour les colonnes en fonction des statistiques de données.	TRUE
spark.sql. inMemoryC olumnarRangement. enableVectorizedRe ader	Spécifie s'il faut activer la lecture vectorisée pour la mise en cache en colonnes.	TRUE
spark.sql.legacy. allowHash OnMapType	Spécifie s'il faut autoriser les opérations de hachage sur les structures de données de type carte. Cet ancien paramètre assure la compatibilité avec la gestion des types de carte dans les anciennes versions de Spark.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.allowNegativeScaleOfDecimal	Spécifie s'il faut autoriser les valeurs d'échelle négatives dans les définitions de type décimal. Cet ancien paramètre maintient la compatibilité avec les anciennes versions de Spark qui prenaient en charge les échelles décimales négatives.	
spark.sql.legacy.castComplexTypesToString.activé	Spécifie s'il faut activer le comportement existant pour convertir des types complexes en chaînes. Maintient la compatibilité avec les règles de conversion de type des anciennes versions de Spark.	
spark.sql.legacy.charVarcharAsString	Spécifie s'il faut traiter les types CHAR et VARCHAR comme des types STRING. Cet ancien paramètre assure la compatibilité avec la gestion des types de chaînes des anciennes versions de Spark.	
spark.sql.legacy.createEmptyCollectionUsingStringType	Spécifie s'il faut créer des collections vides à l'aide d'éléments de type chaîne. Cet ancien paramètre assure la compatibilité avec le comportement d'initialisation des collections des anciennes versions de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal.activé	Spécifie s'il faut interpréter les littéraux exponentiels comme des types décimaux. Cet ancien paramètre assure la compatibilité avec la gestion littérale numérique des anciennes versions de Spark.	
spark.sql.legacy.json.allowEmptyString.activé	Spécifie s'il faut autoriser les chaînes vides dans le traitement JSON. Cet ancien paramètre maintient la compatibilité avec le comportement d'analyse JSON des anciennes versions de Spark.	

Nom de propriété	Description	Valeur par défaut
spark.sql.legacy.parquet.int96RebaseModelRead	Spécifie s'il faut utiliser l'ancien mode de rebase d' INT96 horodatage lors de la lecture des fichiers Parquet. Cet ancien paramètre assure la compatibilité avec la gestion des horodatages des anciennes versions de Spark.	
spark.sql.legacy.timeParserPolicy	Contrôle le comportement d'analyse temporelle à des fins de rétrocompatibilité. Cet ancien paramètre détermine la manière dont les horodatages et les dates sont analysés à partir de chaînes.	
spark.sql.Legacy.TypeCoercion.datetimeToString.activated	Spécifie s'il faut activer le comportement de coercition de type existant lors de la conversion de valeurs date/heure en chaînes. Maintient la compatibilité avec les règles de conversion date/heure des anciennes versions de Spark.	
spark.sql. maxSinglePartitionOctets	Définit la taille maximale de la partition en octets. Le planificateur introduit des opérations de shuffle pour les partitions plus grandes afin d'améliorer le parallélisme.	128 m

Nom de propriété	Description	Valeur par défaut
Cache de métadonnées Spark.SQL TTLSeconds	Contrôle le time-to-live (TTL) pour les caches de métadonnées. S'applique aux métadonnées des fichiers de partition et aux caches du catalogue de sessions. Nécessite : • Une valeur positive supérieure à zéro • spark.sql.CatalogImplementation défini sur hive • spark.sql.hive.filesourcePartitionFileCacheSize supérieur à zéro • spark.sql.hive.manageFilesourcePartitions défini sur true	-1000 ms
optimiseur spark.sql . collapseProjectAlwaysEn ligne	Spécifie s'il faut réduire les projections adjacentes et les expressions en ligne, même si cela entraîne une duplication.	FALSE
optimiseur spark.sql . dynamicPartitionPruning.activé	Spécifie s'il faut générer des prédicats pour les colonnes de partition utilisées comme clés de jointure.	TRUE
optimiseur spark.sql. enableCsvExpressionOptimisation	Spécifie s'il faut optimiser les expressions CSV dans l'optimiseur SQL en supprimant les colonnes inutiles des opérations from_csv.	TRUE
optimiseur spark.sql. enableJsonExpressionOptimisation	Spécifie s'il faut optimiser les expressions JSON dans l'optimiseur SQL en : • Supprimer les colonnes inutiles des opérations from_json • Simplification des combinaisons from_json et to_json • Optimisation des opérations named_struct	TRUE

Nom de propriété	Description	Valeur par défaut
Spark.SQL.Optimizer.Règles exclues	Définit les règles d'optimisation à désactiver, identifiées par des noms de règles séparés par des virgules. Certaines règles ne peuvent pas être désactivées car elles sont nécessaires pour être correctes. L'optimiseur enregistre les règles qui ont été désactivées avec succès.	(aucun)
Spark.SQL.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Définit la taille de scan agrégée minimale en octets requise pour injecter un filtre Bloom côté application.	10 GO
Spark.SQL.Optimizer.Runtime.BloomFilter.creationSideThreshold	Définit le seuil de taille maximale pour injecter un filtre Bloom côté création.	10 MO
Spark.SQL.Optimizer.Runtime.BloomFilter.Activé	Spécifie s'il faut insérer un filtre Bloom pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	TRUE
Spark.SQL.Optimizer.Runtime.BloomFilter.expectedNumItems	Définit le nombre par défaut d'éléments attendus dans le filtre Bloom d'exécution.	1000000
Spark.SQL.Optimizer.Runtime.BloomFilter.maxNumBits	Définit le nombre maximum de bits autorisés dans le filtre Bloom d'exécution.	67108864
Spark.SQL.Optimizer.Runtime.BloomFilter.maxNumItems	Définit le nombre maximum d'éléments attendus autorisés dans le filtre Bloom d'exécution.	4000000

Nom de propriété	Description	Valeur par défaut
spark.sql.Optimizer.Runtime.BloomFilter.Number.Threshold	Limite le nombre maximum de filtres d'exécution non DPP autorisés par requête afin d'éviter les out-of-memory erreurs dans le pilote.	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Définit le nombre de bits par défaut utilisés dans le filtre Bloom d'exécution.	8388608
spark.sql.optimize r.runtime. rowlevelO perationGroupFiltr e.activé	Spécifie s'il faut activer le filtrage des groupes d'exécution pour les opérations au niveau des lignes. Permet aux sources de données de : • Réduisez des groupes entiers de données (tels que des fichiers ou des partitions) à l'aide de filtres de source de données • Exécuter des requêtes d'exécution pour identifier les enregistrements correspondants • Supprimez les groupes inutiles pour éviter des réécritures coûteuses Limites: • Toutes les expressions ne peuvent pas être converties en filtres de source de données • Certaines expressions nécessitent une évaluation Spark (telles que les sous-requêtes)	TRUE

Nom de propriété	Description	Valeur par défaut
Spark.SQL.Optimize r.RuntimeFilter. semiJoinReduction. activé	Spécifie s'il faut insérer une semi-jointure pour réduire les données de mélange lorsqu'un côté d'une jointure aléatoire comporte un prédicat sélectif.	FALSE
spark.sql.parquet. AggregatePushdown	Spécifie s'il faut transférer les agrégats vers Parquet à des fins d'optimisation. Supporte : • MIN et MAX pour les types booléen, entier, flottant et date • COUNT pour tous les types de données Lance une exception si les statistiques sont absentes du pied de page d'un fichier Parquet.	FALSE
spark.sql.parquet. columnarReaderBatch Taille	Contrôle le nombre de lignes dans chaque lot de lecteurs vectorisés Parquet. Choisissez une valeur qui équilibre le surcoût des performances et l'utilisation de la mémoire pour éviter out-of-memory les erreurs.	4096
spark.sql.session. TimeZone	Définit le fuseau horaire de session pour la gestion des horodatages sous forme de chaînes littérales et pour la conversion d'objets Java. Accepte : • area/city Format basé sur la région IDs (tel que America/Los_Angeles) • Décalages de zone au format (+/-) HH, (+/-) HH:mm ou (+/-) HH:mm:ss (tel que -08 ou + 01:00) • UTC ou Z comme alias pour + 00:00	(valeur du fuseau horaire local)

Nom de propriété	Description	Valeur par défaut
spark.sql.shuffle.partitions	Définit le nombre de partitions par défaut pour le brassage des données lors des jointures ou des agrégations. Ne peut pas être modifié entre les redémarrages de requêtes de streaming structurées à partir du même emplacement de point de contrôle.	200
spark.sql.shuffleHashJoinFactor	Définit le facteur de multiplication utilisé pour déterminer l'éligibilité à la jointure par hachage aléatoire. Une jointure par hachage aléatoire est sélectionnée lorsque la taille des données du petit côté multipliée par ce facteur est inférieure à la taille des données du grand côté.	3
spark.sql.sources.parallelPartitionDiscovery.seuil	Définit le nombre maximum de chemins pour la liste des fichiers côté pilote avec des sources basées sur des fichiers (Parquet, JSON et ORC). En cas de dépassement lors de la découverte de partitions, les fichiers sont répertoriés à l'aide d'une tâche distribuée Spark distincte.	32
spark.sql.statistics.histogram.enabled	Spécifie s'il faut générer des histogrammes à hauteur égale lors du calcul des statistiques des colonnes afin d'améliorer la précision de l'estimation. Nécessite une analyse de table supplémentaire en plus de celle requise pour les statistiques de base sur les colonnes.	FALSE

9. Pour l'accès au service, choisissez le nom du rôle de service existant qui sera utilisé pour accéder à cette table.
10. Si vous souhaitez activer les balises pour le jeu de données d'entraînement, choisissez Ajouter une nouvelle balise, puis entrez la paire clé/valeur.
11. Choisissez Créer un segment similaire.

Pour l'action d'API correspondante, consultez [StartAudienceGenerationJob](#).

Exportation d'un segment similaire

Après avoir créé un segment similaire, vous pouvez exporter ces données vers un compartiment Amazon S3.

Pour exporter un segment similaire dans AWS Clean Rooms

1. Connectez-vous à la [AWS Clean Rooms console AWS Management Console et ouvrez-la](#) avec votre Compte AWS (si vous ne l'avez pas encore fait).
2. Dans le volet de navigation de gauche, sélectionnez Collaborations.
3. Dans l'onglet Avec adhésion active, choisissez une collaboration.
4. Dans l'onglet Modèles ML, sélectionnez un segment similaire et choisissez Exporter.
5. Pour Exporter un modèle similaire, pour Exporter les détails du modèle similaire, entrez un nom et une description facultative.
6. Pour Taille du segment, choisissez la taille que vous souhaitez pour le segment exporté.
7. Cliquez sur Exporter.

Pour l'action d'API correspondante, consultez [StartAudienceExportJob](#).

Résolution des problèmes AWS Clean Rooms

Cette section décrit certains problèmes courants susceptibles de survenir lors de l'utilisation AWS Clean Rooms et explique comment les résoudre.

Problèmes

- [Une ou plusieurs tables référencées par la requête ne sont pas accessibles par le rôle de service associé. Le table/role propriétaire doit accorder au rôle de service l'accès à la table.](#)
- [L'un des ensembles de données sous-jacents possède un format de fichier non pris en charge.](#)
- [Les résultats des requêtes ne sont pas ceux attendus lors de l'utilisation de l'informatique cryptographique pour Clean Rooms.](#)
- [AWS Clean Rooms Spark SQL : données de partition manquantes](#)

Une ou plusieurs tables référencées par la requête ne sont pas accessibles par le rôle de service associé. Le table/role propriétaire doit accorder au rôle de service l'accès à la table.

- Vérifiez que les autorisations pour le rôle de service sont configurées conformément aux exigences. Pour plus d'informations, consultez [Con AWS Clean Rooms figuration](#).

L'un des ensembles de données sous-jacents possède un format de fichier non pris en charge.

- Assurez-vous que votre ensemble de données est dans l'un des formats de fichier pris en charge :
 - Parquet
 - RCFile
 - TextFile
 - SequenceFile
 - RegexSerde
 - OpenCSV
 - AVRO

- JSON

Pour de plus amples informations, veuillez consulter [Formats de données pour AWS Clean Rooms](#).

Les résultats des requêtes ne sont pas ceux attendus lors de l'utilisation de l'informatique cryptographique pour Clean Rooms.

Si vous utilisez l'informatique cryptographique pour Clean Rooms (C3R), vérifiez que votre requête utilise correctement les colonnes chiffrées :

- Les sealed colonnes ne sont utilisées que dans les SELECT clauses.
- Les fingerprint colonnes ne sont utilisées que dans JOIN les clauses (et GROUP BY les clauses sous certaines conditions).
- Que vous n'êtes que JOINing fingerprint des colonnes portant le même nom si les paramètres de collaboration l'exigent.

Pour plus d'informations, consultez [the section called "Informatique cryptographique"](#) et [the section called "Types de colonnes"](#).

AWS Clean Rooms Spark SQL : données de partition manquantes

Toutes les partitions du AWS Glue Data Catalog doivent également contenir des données dans S3. Le moteur utilise le paramètre Spark `spark.sql.files.ignoreMissingFiles=False`

Pour plus d'informations, voir <https://spark.apache.org/docs/latest/sql-data-sources-generic-options.html#ignore-missing-files>.

Si vous rencontrez cette erreur, le message d'erreur suivant s'affichera : "Missing partition data: One of the configured tables is partitioned and one or more of the partitions does not have data".

Comparez vos données présentes dans Amazon S3 avec les partitions répertoriées dans AWS Glue Data Catalog le tableau correspondant. Supprimez les partitions qui n'ont pas de données correspondantes dans S3.

Sécurité dans AWS Clean Rooms

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 ceci comme la sécurité du cloud et la 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 [AWS conformité Programmes](#) de de conformité. Pour en savoir plus sur les programmes de conformité applicables à AWS Clean Rooms, consultez la section [Services AWS 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 la sensibilité de vos données, les exigences de votre entreprise et la législation et la réglementation applicables.

Cette documentation vous aide à comprendre comment appliquer le modèle de responsabilité partagée lors de son utilisation AWS Clean Rooms. Il vous explique comment procéder à la configuration AWS Clean Rooms pour atteindre vos objectifs de sécurité et de conformité. Vous apprendrez également à utiliser d'autres AWS services qui vous aident à surveiller et à sécuriser vos AWS Clean Rooms ressources.

Table des matières

- [Protection des données dans AWS Clean Rooms](#)
- [Utilisation de rôles liés à un service pour AWS Clean Rooms](#)
- [Conservation des données dans AWS Clean Rooms](#)
- [Bonnes pratiques en matière de collaboration de données dans AWS Clean Rooms](#)
- [Identity and Access Management pour AWS Clean Rooms](#)
- [Validation de conformité pour AWS Clean Rooms](#)
- [Résilience dans AWS Clean Rooms](#)
- [Sécurité de l'infrastructure dans AWS Clean Rooms](#)

- [Accès AWS Clean Rooms ou AWS Clean Rooms ML à l'aide d'un point de terminaison d'interface \(AWS PrivateLink\)](#)

Protection des données dans AWS Clean Rooms

Le [modèle de responsabilité AWS partagée](#) de s'applique à la protection des données dans AWS Clean Rooms. 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 AWS Clean Rooms 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 Clean Rooms chiffre toujours toutes les métadonnées du service au repos sans nécessiter de configuration supplémentaire. Ce chiffrement est automatique lorsque vous l'utilisez AWS Clean Rooms.

Clean Rooms ML chiffre toutes les données stockées dans le service au repos avec AWS KMS. Si vous choisissez de fournir votre propre clé KMS, le contenu de vos modèles similaires et de vos tâches de génération de segments similaires est chiffré au repos avec votre clé KMS.

Lorsque vous utilisez AWS Clean Rooms des modèles de machine learning personnalisés, le service chiffre toutes les données stockées au repos avec AWS KMS. AWS Clean Rooms prend en charge l'utilisation de clés symétriques gérées par le client que vous créez, détenez et gérez pour chiffrer les données au repos. Si les clés gérées par le client ne sont pas spécifiées, Clés détenues par AWS elles sont utilisées par défaut.

AWS Clean Rooms utilise des autorisations et des politiques clés pour accéder aux clés gérées par le client. Vous pouvez révoquer l'accès à l'octroi ou supprimer l'accès du service à la clé gérée par le client à tout moment. Si vous le faites, vous AWS Clean Rooms ne pourrez accéder à aucune des données chiffrées par la clé gérée par le client, ce qui affectera les opérations qui dépendent de ces données. Par exemple, si vous tentez de créer un modèle entraîné à partir d'un canal d'entrée ML crypté AWS Clean Rooms auquel il est impossible d'accéder, l'opération renverra une `ValidationException` erreur.

Note

Vous pouvez utiliser les options de chiffrement d'Amazon S3 pour protéger vos données au repos.

Pour plus d'informations, consultez la section [Spécification du chiffrement Amazon S3](#) dans le guide de l'utilisateur Amazon S3.

Lorsque vous utilisez une table de mappage d'identifiants à l'intérieur AWS Clean Rooms, le service chiffre toutes les données stockées au repos avec AWS KMS. Si vous choisissez de fournir votre propre clé KMS, le contenu de votre table de mappage d'identifiants est chiffré au repos avec votre clé KMS via Résolution des entités AWS. Pour plus de détails sur les autorisations requises pour utiliser des chiffrements avec un flux de travail de mappage d'identifiants, voir [Créer un rôle de travail dans le flux de travail Résolution des entités AWS](#) dans le guide de l'utilisateur Résolution des entités AWS.

Chiffrement en transit

AWS Clean Rooms utilise le protocole TLS (Transport Layer Security) pour le chiffrement en transit. La communication se fait toujours via HTTPS, de sorte que vos données sont toujours chiffrées en transit, qu'elles soient stockées sur Amazon S3, Amazon Athena ou Snowflake. AWS Clean Rooms Cela inclut toutes les données en transit lors de l'utilisation de Clean Rooms ML.

Chiffrement des données sous-jacentes

Pour plus d'informations sur le chiffrement de vos données sous-jacentes, consultez [Informatique cryptographique pour Clean Rooms](#).

Stratégie de clé

Les stratégies de clé contrôlent l'accès à votre clé gérée par le client. Chaque clé gérée par le client doit avoir exactement une stratégie de clé, qui contient des instructions qui déterminent les personnes pouvant utiliser la clé et comment elles peuvent l'utiliser. Lorsque vous créez votre clé gérée par le client, vous pouvez spécifier une stratégie de clé. Pour plus d'informations, consultez la section Gestion de l'accès aux clés gérées par le client dans le Guide du AWS Key Management Service développeur.

Pour utiliser votre clé gérée par le client avec vos modèles de ML AWS Clean Rooms personnalisés, les opérations d'API suivantes doivent être autorisées dans la politique des clés :

- `kms:DescribeKey`— Fournit les informations clés gérées par le client AWS Clean Rooms pour permettre de valider la clé.

- `kms:Decrypt`— Permet d'accéder AWS Clean Rooms aux données cryptées pour les déchiffrer et les utiliser dans des tâches connexes.
- `kms:CreateGrant`- Clean Rooms ML chiffre les images d'entraînement et d'inférence au repos dans Amazon ECR en créant des subventions pour Amazon ECR. Pour en savoir plus, consultez [Encryption at Rest dans Amazon ECR](#). Clean Rooms ML utilise également Amazon SageMaker AI pour exécuter des tâches de formation et d'inférence, et crée des subventions permettant à l' SageMaker IA de chiffrer les volumes Amazon EBS attachés aux instances ainsi que les données de sortie dans Amazon S3. Pour en savoir plus, consultez [Protéger les données au repos à l'aide du chiffrement dans Amazon SageMaker AI](#).
- `kms:GenerateDataKey`- Clean Rooms ML chiffre les données au repos stockées dans Amazon S3 à l'aide du chiffrement côté serveur avec AWS KMS keys Pour en savoir plus, consultez [Utilisation du chiffrement côté serveur avec AWS KMS keys \(SSE-KMS\) dans Amazon S3](#).

Vous trouverez ci-dessous des exemples de déclarations de politique que vous pouvez ajouter AWS Clean Rooms pour les ressources suivantes :

Canal d'entrée ML avec données synthétiques

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Allow access to principals authorized to use AWS Clean Rooms ML",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::666666666666:role/ExampleRole"
      },
      "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        },
        "ForAllValues:StringEquals": {
          "kms:GrantOperations": [
            "Decrypt",

```

```

        "Encrypt",
        "GenerateDataKeyWithoutPlaintext",
        "ReEncryptFrom",
        "ReEncryptTo",
        "CreateGrant",
        "DescribeKey",
        "RetireGrant",
        "GenerateDataKey"
    ]
},
    "BoolIfExists": {
        "kms:GrantIsForAWSResource": true
    }
},
{
    "Sid": "Allow describe key for principals authorized to use AWS Clean Rooms
ML",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::444455556666:role/ExampleRole"
    },
    "Action": [
        "kms:DescribeKey"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": {
            "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        }
    }
},
{
    "Sid": "Allow grant operations for AWS Clean Rooms ML service principal",
    "Effect": "Allow",
    "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
    },
    "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
    ],
    "Resource": "*",

```

```

    "Condition": {
      "ForAllValues:StringEquals": {
        "kms:GrantOperations": [
          "Decrypt",
          "Encrypt",
          "GenerateDataKeyWithoutPlaintext",
          "ReEncryptFrom",
          "ReEncryptTo",
          "CreateGrant",
          "DescribeKey",
          "RetireGrant",
          "GenerateDataKey"
        ]
      }
    },
    {
      "Sid": "Allow describe key for AWS Clean Rooms ML service principal",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": [
        "kms:DescribeKey"
      ],
      "Resource": "*"
    }
  ]
}

```

Canal d'entrée ML sans données synthétiques

JSON

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Allow access to principals authorized to use AWS Clean Rooms ML",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::666666666666:role/ExampleRole"
      },
    },
  ],
}

```

```

    "Action": [
        "kms:DescribeKey",
        "kms:GenerateDataKey",
        "kms:Decrypt"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": {
            "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        }
    }
},
{
    "Sid": "Allow access to AWS Clean Rooms ML service principal",
    "Effect": "Allow",
    "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
    },
    "Action": [
        "kms:DescribeKey",
        "kms:GenerateDataKey",
        "kms:Decrypt"
    ],
    "Resource": "*"
}
]
}

```

Poste de modèle entraîné ou travail d'inférence de modèle entraîné

JSON

```

{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Allow grant operations for principals authorized to use AWS  
Clean Rooms ML",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::666666666666:role/ExampleRole"
            },

```

```

    "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": {
            "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        },
        "ForAllValues:StringEquals": {
            "kms:GrantOperations": [
                "Decrypt",
                "Encrypt",
                "GenerateDataKeyWithoutPlaintext",
                "ReEncryptFrom",
                "ReEncryptTo",
                "CreateGrant",
                "DescribeKey",
                "RetireGrant",
                "GenerateDataKey"
            ]
        },
        "BoolIfExists": {
            "kms:GrantIsForAWSResource": true
        }
    }
},
{
    "Sid": "Allow describe key for principals authorized to use AWS Clean
Rooms ML",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::444455556666:role/ExampleRole"
    },
    "Action": [
        "kms:DescribeKey"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": {
            "kms:ViaService": "cleanrooms-ml.us-east-1.amazonaws.com"
        }
    }
}

```

```

    },
    {
      "Sid": "Allow grant operations for AWS Clean Rooms ML service
principal",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": [
        "kms:GenerateDataKey",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": "*",
      "Condition": {
        "ForAllValues:StringEquals": {
          "kms:GrantOperations": [
            "Decrypt",
            "Encrypt",
            "GenerateDataKeyWithoutPlaintext",
            "ReEncryptFrom",
            "ReEncryptTo",
            "CreateGrant",
            "DescribeKey",
            "RetireGrant",
            "GenerateDataKey"
          ]
        }
      }
    },
    {
      "Sid": "Allow describe key for AWS Clean Rooms ML service principal",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": [
        "kms:DescribeKey"
      ],
      "Resource": "*"
    }
  ]
}

```


Clean Rooms ML ne prend pas en charge la spécification du contexte de chiffrement des services ou du contexte source dans les politiques relatives aux clés gérées par le client. Le contexte de chiffrement utilisé par le service en interne est visible par les clients dans CloudTrail.

Utilisation de rôles liés à un service pour AWS Clean Rooms

AWS Clean Rooms utilise des Gestion des identités et des accès AWS rôles liés à un [service](#) (IAM). Un rôle lié à un service est un type unique de rôle IAM directement lié à. AWS Clean Rooms Les rôles liés au service sont prédéfinis par AWS Clean Rooms et incluent 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 AWS Clean Rooms car vous n'avez pas à ajouter manuellement les autorisations nécessaires. AWS Clean Rooms définit les autorisations associées à ses rôles liés aux services et, sauf indication contraire, seul AWS Clean Rooms peut assumer ses rôles. Les autorisations définies comprennent la politique de confiance et la politique d'autorisation. De plus, cette politique d'autorisation ne peut pas être attachée à une autre entité IAM.

Vous pouvez supprimer un rôle lié à un service uniquement après la suppression préalable de ses ressources connexes. Cela protège vos AWS Clean Rooms ressources car vous ne pouvez pas supprimer par inadvertance l'autorisation d'accès aux ressources.

Pour plus d'informations sur les autres services qui prennent en charge les rôles liés à un service, consultez la section [AWS Services qui fonctionnent avec IAM](#) et recherchez les services dont la valeur est Oui dans la colonne Rôles liés à un service. Sélectionnez un Oui ayant un lien pour consulter la documentation du rôle lié à un service, pour ce service.

Autorisations de rôle liées à un service pour AWS Clean Rooms

AWS Clean Rooms utilise le rôle lié au service nommé `AWSServiceRoleForAWSCleanRooms` pour publier les CloudWatch statistiques relatives à Clean Rooms sur votre compte. AWS

Le rôle lié au service `AWSService RoleFor AWSClean Rooms` fait confiance aux services suivants pour assumer ce rôle :

- `cleanrooms.amazonaws.com`

La politique d'autorisations de rôle nommée `AWSClean RoomsServiceRolePolicy` AWS Clean Rooms permet d'effectuer les actions suivantes sur les ressources spécifiées :

- Action : `cloudwatch:PutMetricData` sur all AWS resources, restricted to the AWS Clean Rooms namespace

Vous devez configurer les autorisations de manière à permettre à vos utilisateurs, groupes ou rôles de créer, modifier ou supprimer un rôle lié à un service. Pour plus d'informations, consultez [Autorisations de rôles liés à un service](#) dans le Guide de l'utilisateur IAM.

Création d'un rôle lié à un service pour AWS Clean Rooms

Vous pouvez utiliser la console IAM pour créer un rôle lié à un service avec le cas d'utilisation de `AWSServiceRoleForAWSCleanRooms`. Dans l'API AWS CLI ou dans l'AWS API, créez un rôle lié à un service avec le nom du `cleanrooms.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 AWS Clean Rooms

AWS Clean Rooms ne vous permet pas de modifier le rôle lié au service `AWSServiceRoleForAWSCleanRooms`. 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 plus d'informations, consultez [Modification d'un rôle lié à un service](#) dans le Guide de l'utilisateur IAM.

Supprimer un rôle lié à un service pour AWS Clean Rooms

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.

Pour supprimer les `AWSServiceRoleForAWSCleanRooms`, vous devez d'abord supprimer toutes les [collaborations](#) et [adhésions](#) de votre Compte AWS.

Note

Si le AWS Clean Rooms service 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 manuellement le rôle lié au service à l'aide d'IAM

Utilisez la console IAM AWS CLI, le ou l' AWS API pour supprimer le rôle lié au service AWSService RoleFor AWSClean Rooms. Pour plus d'informations, consultez la section [Suppression d'un rôle lié à un service](#) dans le Guide de l'utilisateur IAM.

Régions prises en charge pour les rôles AWS Clean Rooms liés à un service

AWS Clean Rooms prend en charge l'utilisation de rôles liés au service dans toutes les régions où le service est disponible. Pour plus d'informations, consultez [AWS Régions et points de terminaison](#).

Conservation des données dans AWS Clean Rooms

Toutes les données temporairement lues dans une AWS Clean Rooms collaboration sont supprimées une fois la requête terminée.

Lorsque vous créez un modèle similaire, Clean Rooms ML lit vos données d'entraînement, les transforme dans un format adapté à notre modèle ML et stocke les paramètres du modèle entraîné dans Clean Rooms ML. Clean Rooms ML ne conserve aucune copie de vos données d'entraînement. AWS Clean Rooms Les requêtes SQL ne conservent aucune de vos données une fois la requête exécutée. Clean Rooms ML utilise ensuite le modèle entraîné pour résumer le comportement de tous vos utilisateurs. Clean Rooms ML stocke un ensemble de données au niveau utilisateur pour chaque utilisateur de vos données tant que votre modèle de sosie est actif.

Lorsque vous lancez une tâche de génération de segments similaires, Clean Rooms ML lit les données initiales, lit les résumés des comportements à partir du modèle de similarité associé et crée un segment similaire qui est stocké dans le service. AWS Clean Rooms Clean Rooms ML ne conserve aucune copie de vos données de départ. Clean Rooms ML stocke les résultats de la tâche au niveau utilisateur tant que celle-ci est active.

Si vos données de départ proviennent d'une requête SQL, le résultat de cette requête n'est stocké dans le service que pendant la durée de la tâche. Les résultats de la requête sont chiffrés au repos et en transit.

Si vous souhaitez supprimer votre modèle de similarité ou les données de travail de génération de segments similaires, utilisez l'API pour les supprimer. Clean Rooms ML supprime de manière asynchrone toutes les données associées au modèle ou à la tâche. Une fois ce processus terminé,

Clean Rooms ML supprime les métadonnées du modèle ou de la tâche et celles-ci ne sont plus visibles dans l'API. Clean Rooms ML conserve les données supprimées pendant 3 jours à des fins de prévention de la reprise après sinistre. Une fois que la tâche ou le modèle n'est plus visible dans l'API et que 3 jours se sont écoulés, toutes les données associées au modèle ou à la tâche sont définitivement supprimées.

Bonnes pratiques en matière de collaboration de données dans AWS Clean Rooms

Cette rubrique décrit les meilleures pratiques pour mener des collaborations de données dans AWS Clean Rooms.

AWS Clean Rooms suit le [modèle de responsabilitéAWS partagée](#). AWS Clean Rooms propose des [règles d'analyse](#) que vous pouvez configurer pour renforcer votre capacité à protéger les données sensibles dans le cadre d'une collaboration. Les règles d'analyse que vous configurez dans AWS Clean Rooms appliqueront les restrictions (contrôles de requête et contrôles de sortie de requête) que vous avez configurées. Il vous incombe de déterminer les restrictions et de configurer les règles d'analyse en conséquence.

Les collaborations en matière de données peuvent impliquer bien plus que votre simple utilisation de AWS Clean Rooms. Pour vous aider à tirer le meilleur parti des collaborations de données, nous vous recommandons de suivre les meilleures pratiques suivantes en utilisant AWS Clean Rooms et en particulier en ce qui concerne les règles d'analyse.

Rubriques

- [Les meilleures pratiques avec AWS Clean Rooms](#)
- [Bonnes pratiques d'utilisation des règles d'analyse dans AWS Clean Rooms](#)

Les meilleures pratiques avec AWS Clean Rooms

Vous êtes chargé d'évaluer le risque lié à chaque collaboration sur les données et de le comparer à vos exigences en matière de confidentialité, telles que les programmes et politiques de conformité externes et internes. Nous vous recommandons de prendre des mesures supplémentaires lors de l'utilisation de AWS Clean Rooms. Ces actions peuvent contribuer à mieux gérer les risques et à vous prémunir contre les tentatives de tiers visant à réidentifier vos données (par exemple, des attaques différenciées ou des attaques par canal secondaire).

Par exemple, envisagez de faire preuve de diligence raisonnable à l'égard de vos autres collaborateurs et de conclure des accords juridiques avec eux avant de vous engager dans une collaboration. Pour surveiller l'utilisation de vos données, envisagez également d'adopter d'autres mécanismes d'audit lorsque vous utilisez AWS Clean Rooms.

Bonnes pratiques d'utilisation des règles d'analyse dans AWS Clean Rooms

Les règles d'analyse vous AWS Clean Rooms permettent de limiter les requêtes pouvant être exécutées en définissant des contrôles de requête sur une table configurée. Par exemple, vous pouvez définir un contrôle de requête indiquant comment une table configurée peut être jointe et quelles colonnes peuvent être sélectionnées. Vous pouvez également restreindre le résultat de la requête en définissant des contrôles des résultats de requête tels que des seuils d'agrégation sur les lignes de sortie. Le service rejette toute requête et supprime les lignes non conformes aux règles d'analyse définies par les membres sur leurs tables configurées dans la requête.

Nous recommandons les 10 meilleures pratiques suivantes pour utiliser les règles d'analyse sur votre table configurée :

- Créez des tables configurées distinctes pour des cas d'utilisation de requêtes distincts (par exemple, planification d'audience ou attribution). Vous pouvez créer plusieurs tables configurées avec la même AWS Glue table sous-jacente.
- Spécifiez les colonnes de la règle d'analyse (par exemple, les colonnes de dimension, les colonnes de liste, les colonnes de jointure) qui sont nécessaires pour les requêtes dans le cadre d'une collaboration. Cela peut contribuer à atténuer le risque de différenciation des attaques ou de permettre à d'autres membres de rétroconcevoir vos données. Utilisez la fonctionnalité des colonnes autorisées pour noter les autres colonnes que vous souhaitez peut-être rendre interrogeables à l'avenir. Pour personnaliser les colonnes qui peuvent être utilisées pour une collaboration donnée, créez des tables configurées supplémentaires avec la même AWS Glue table sous-jacente.
- Spécifiez dans la règle d'analyse les fonctions nécessaires à l'analyse dans le cadre de la collaboration. Cela peut contribuer à atténuer les risques liés à de rares erreurs de fonctionnement susceptibles de présenter des informations sur un point de données individuel. Pour personnaliser les fonctions qui peuvent être utilisées pour une collaboration donnée, créez des tables configurées supplémentaires avec la même AWS Glue table sous-jacente.
- Ajoutez des contraintes d'agrégation à toutes les colonnes dont les valeurs au niveau des lignes sont sensibles. Cela inclut les colonnes de votre table configurée qui existent également dans les tables des autres membres de la collaboration et les règles d'analyse en tant que contrainte

d'agrégation. Cela inclut également les colonnes de votre table configurée qui ne sont pas interrogeables, c'est-à-dire les colonnes qui se trouvent dans votre table configurée mais qui ne figurent pas dans la règle d'analyse. Les contraintes d'agrégation peuvent contribuer à atténuer les risques liés à la corrélation des résultats des requêtes avec des données extérieures à la collaboration.

- Créez des collaborations de test et des règles d'analyse pour tester les restrictions créées avec des règles d'analyse spécifiées.
- Passez en revue les tables configurées par le collaborateur et les règles d'analyse des membres sur les tables configurées pour vérifier qu'elles correspondent à ce qui a été convenu pour la collaboration. Cela peut aider à atténuer les risques liés à l'ingénierie par les autres membres de leurs propres données pour exécuter des requêtes qui n'ont pas été convenues.
- Consultez l'exemple de requête fourni (console uniquement) qui est activé sur votre table configurée après avoir configuré la règle d'analyse.

 Note

Outre l'exemple de requête fourni, d'autres requêtes sont possibles en fonction de la règle d'analyse, d'autres tables de membres de collaboration et de règles d'analyse.

- Vous pouvez ajouter ou mettre à jour une règle d'analyse pour une table configurée dans une collaboration. Lorsque vous le faites, passez en revue toutes les collaborations auxquelles la table configurée est associée et l'impact qui en résulte. Cela permet de s'assurer qu'aucune collaboration n'utilise de règles d'analyse obsolètes.
- Passez en revue les requêtes exécutées dans le cadre de la collaboration pour vérifier qu'elles correspondent aux cas d'utilisation ou aux requêtes convenus pour la collaboration. (Les requêtes sont disponibles dans les journaux des requêtes lorsque la fonctionnalité de journalisation des requêtes est activée.) Cela peut aider à atténuer les risques liés à l'exécution par les membres d'analyses non approuvées et aux attaques potentielles telles que les attaques par canal secondaire.
- Passez en revue les colonnes de table configurées utilisées dans les règles d'analyse des membres de la collaboration et dans les requêtes pour vérifier qu'elles correspondent à ce qui a été convenu dans le cadre de la collaboration. (Les requêtes sont disponibles dans les journaux de requêtes lorsque cette fonctionnalité est activée.) Cela peut aider à atténuer les risques liés à l'ingénierie par les autres membres de leurs propres données pour effectuer des requêtes qui n'ont pas été convenues.

Identity and Access Management pour AWS Clean Rooms

Gestion des identités et des accès AWS (IAM) est un outil Service AWS qui permet à un administrateur de contrôler en toute sécurité l'accès aux AWS ressources. Les administrateurs IAM contrôlent qui peut être authentifié (connecté) et autorisé (autorisé) à utiliser AWS Clean Rooms les ressources. IAM est un Service AWS outil que vous pouvez utiliser sans frais supplémentaires.

Rubriques

- [Public ciblé](#)
- [Authentification par des identités](#)
- [Gestion de l'accès à l'aide de politiques](#)
- [Comment AWS Clean Rooms fonctionne avec IAM](#)
- [Exemples de politiques basées sur l'identité pour AWS Clean Rooms](#)
- [AWS politiques gérées pour AWS Clean Rooms](#)
- [Résolution des problèmes AWS Clean Rooms d'identité et d'accès](#)
- [Prévention du problème de l'adjoint confus entre services](#)
- [Comportements IAM pour AWS Clean Rooms ML](#)
- [Comportements IAM pour les modèles personnalisés Clean Rooms ML](#)

Public ciblé

La façon dont vous utilisez Gestion des identités et des accès AWS (IAM) varie en fonction de votre rôle :

- Utilisateur du service : demandez des autorisations à votre administrateur si vous ne pouvez pas accéder aux fonctionnalités (voir [Résolution des problèmes AWS Clean Rooms d'identité et d'accès](#))
- Administrateur du service : déterminez l'accès des utilisateurs et soumettez les demandes d'autorisation (voir [Comment AWS Clean Rooms fonctionne avec IAM](#))
- Administrateur IAM : rédigez des politiques pour gérer l'accès (voir [Exemples de politiques basées sur l'identité pour AWS Clean Rooms](#))

Authentification par des identités

L'authentification est la façon dont vous vous connectez à AWS l'aide de vos informations d'identification. Vous devez être authentifié (connecté à AWS) en tant qu'utilisateur IAM ou en assumant un rôle IAM. Utilisateur racine d'un compte AWS

Vous pouvez vous connecter en AWS tant qu'identité fédérée en utilisant les informations d'identification fournies par le biais d'une source d'identité. AWS IAM Identity Center Les utilisateurs (IAM Identity Center) ou l'authentification unique de votre entreprise sont des exemples d'identités fédérées. Lorsque vous vous connectez avec une identité fédérée, votre administrateur aura précédemment configuré une fédération d'identités avec des rôles IAM. Lorsque vous accédez à AWS l'aide de la fédération, vous assumez indirectement un rôle.

Selon le type d'utilisateur que vous êtes, vous pouvez vous connecter au portail AWS Management Console ou au portail AWS d'accès. Pour plus d'informations sur la connexion à AWS, consultez la section [Comment vous connecter à votre compte Compte AWS dans](#) le guide de Connexion à AWS l'utilisateur.

Si vous y accédez AWS par programmation, AWS fournit un kit de développement logiciel (SDK) et une interface de ligne de commande (CLI) pour signer cryptographiquement vos demandes à l'aide de vos informations d'identification. Si vous n'utilisez pas d' AWS outils, vous devez signer vous-même les demandes. Pour plus d'informations sur l'utilisation de la méthode recommandée pour signer vos demandes vous-même, veuillez consulter la rubrique [Processus de signature Signature Version 4](#) dans la Références générales AWS.

Quelle que soit la méthode d'authentification que vous utilisez, vous devrez peut-être fournir des informations de sécurité supplémentaires. Par exemple, il vous AWS recommande d'utiliser l'authentification multifactorielle (MFA) pour renforcer la sécurité de votre compte. Pour en savoir plus, consultez [Authentification multifactorielle](#) dans le Guide de l'utilisateur AWS IAM Identity Center et [Utilisation de l'authentification multifactorielle \(MFA\) dans l'interface AWS](#) dans le Guide de l'utilisateur IAM.

Compte AWS utilisateur root

Lorsque vous créez un Compte AWS, vous commencez par une identité de connexion unique qui donne un accès complet à toutes Services AWS les ressources du compte. Cette identité est appelée Compte AWS utilisateur root. Vous pouvez y accéder en vous connectant à l'aide de l'adresse électronique et du mot de passe que vous avez utilisés pour créer le compte. Il est vivement recommandé de ne pas utiliser l'utilisateur root pour vos tâches quotidiennes. Protégez

vos informations d'identification d'utilisateur root et utilisez-les pour effectuer les tâches que seul l'utilisateur root peut effectuer. Pour obtenir la liste complète des tâches qui nécessitent que vous vous connectiez en tant qu'utilisateur root, consultez les [Utilisateur racine d'un compte AWS informations d'identification et les identités IAM](#) dans le Références générales AWS.

Identité fédérée

Il est recommandé d'obliger les utilisateurs humains à utiliser la fédération avec un fournisseur d'identité pour accéder à Services AWS l'aide d'informations d'identification temporaires.

Une identité fédérée est un utilisateur provenant de l'annuaire de votre entreprise, de votre fournisseur d'identité Web ou Directory Service qui y accède à Services AWS l'aide d'informations d'identification provenant d'une source d'identité. Les identités fédérées assument des rôles qui fournissent des informations d'identification temporaires.

Pour une gestion des accès centralisée, nous vous recommandons d'utiliser AWS IAM Identity Center. Pour plus d'informations, consultez [Qu'est-ce que IAM Identity Center ?](#) dans le Guide de l'utilisateur AWS IAM Identity Center .

Utilisateurs et groupes IAM

Un [utilisateur IAM](#) est une identité qui dispose d'autorisations spécifiques pour une seule personne ou application. Nous vous recommandons d'utiliser ces informations d'identification temporaires au lieu des utilisateurs IAM avec des informations d'identification à long terme. Pour plus d'informations, voir [Exiger des utilisateurs humains qu'ils utilisent la fédération avec un fournisseur d'identité pour accéder à AWS l'aide d'informations d'identification temporaires](#) dans le guide de l'utilisateur IAM.

[Les groupes IAM](#) spécifient une collection d'utilisateurs IAM et permettent de gérer plus facilement les autorisations pour de grands ensembles d'utilisateurs. Pour plus d'informations, consultez [Cas d'utilisation pour les utilisateurs IAM](#) dans le Guide de l'utilisateur IAM.

Rôles IAM

Un [rôle IAM](#) est une identité dotée d'autorisations spécifiques qui fournit des informations d'identification temporaires. Vous pouvez assumer un rôle en [passant d'un rôle utilisateur à un rôle IAM \(console\)](#) ou en appelant une opération AWS CLI ou AWS API. Pour plus d'informations, consultez [Méthodes pour endosser un rôle](#) dans le Guide de l'utilisateur IAM.

Les rôles IAM sont utiles pour l'accès des utilisateurs fédérés, les autorisations temporaires des utilisateurs IAM, les accès intercompte, les accès entre services et les applications exécutées sur

Amazon EC2. Pour plus d'informations, consultez [Accès intercompte aux ressources dans IAM](#) dans le Guide de l'utilisateur IAM.

Gestion de l'accès à l'aide de politiques

Vous contrôlez l'accès en AWS créant des politiques et en les associant à AWS des identités ou à des ressources. Une politique est un objet AWS qui, lorsqu'il est associé à une identité ou à une ressource, définit leurs autorisations. AWS évalue ces politiques lorsqu'un principal (utilisateur, utilisateur root ou session de rôle) fait une demande. Les autorisations dans les politiques déterminent si la demande est autorisée ou refusée. La plupart des politiques sont stockées AWS sous forme de documents JSON. Pour plus d'informations sur la structure et le contenu des documents de politique JSON, consultez [Vue d'ensemble des politiques JSON](#) dans le Guide de l'utilisateur IAM.

Les administrateurs peuvent utiliser les politiques AWS JSON pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions, ainsi que les ressources sur lesquelles il peut le faire et dans quelles conditions.

Chaque entité IAM (utilisateur ou rôle) démarre sans autorisation. Par défaut, les utilisateurs ne peuvent rien faire, pas même changer leurs propres mots de passe. Pour autoriser un utilisateur à effectuer une opération, un administrateur doit lui associer une politique d'autorisations. Il peut également ajouter l'utilisateur à un groupe disposant des autorisations prévues. Lorsqu'un administrateur accorde des autorisations à un groupe, tous les utilisateurs de ce groupe se voient octroyer ces autorisations.

Les politiques IAM définissent les autorisations d'une action, quelle que soit la méthode que vous utilisez pour exécuter l'opération. Par exemple, supposons que vous disposiez d'une politique qui autorise l'action `iam:GetRole`. Un utilisateur appliquant cette politique peut obtenir des informations sur le rôle à partir de AWS Management Console AWS CLI, de ou de l' AWS API.

Politiques basées sur l'identité

Les politiques basées sur l'identité sont des documents de politique d'autorisations JSON que vous pouvez attacher à une identité telle qu'un utilisateur, un groupe d'utilisateurs ou un rôle IAM. Ces politiques contrôlent quel type d'actions des utilisateurs et des rôles peuvent exécuter, sur quelles ressources et dans quelles conditions. Pour découvrir comment créer une politique basée sur l'identité, consultez [Définition d'autorisations IAM personnalisées avec des politiques gérées par le client](#) dans le Guide de l'utilisateur IAM.

Les politiques basées sur l'identité peuvent être classées comme des politiques en ligne ou des politiques gérées. Les politiques en ligne sont intégrées directement à un utilisateur, groupe ou rôle. Les politiques gérées sont des politiques autonomes que vous pouvez attacher à plusieurs utilisateurs, groupes et rôles dans votre Compte AWS. Les politiques gérées incluent les politiques AWS gérées et les politiques gérées par le client. Pour découvrir comment choisir entre une politique gérée et une politique en ligne, consultez [Choix entre les politiques gérées et les politiques en ligne](#) dans le Guide de l'utilisateur IAM.

Politiques basées sur les ressources

Les politiques basées sur les ressources sont des documents de politique JSON que vous attachez à une ressource. Par exemple, les politiques de confiance de rôle IAM et les politiques de compartiment Amazon S3 sont des politiques basées sur les ressources. Dans les services qui sont compatibles avec les politiques basées sur les ressources, les administrateurs de service peuvent les utiliser pour contrôler l'accès à une ressource spécifique. Pour la ressource dans laquelle se trouve la politique, cette dernière définit quel type d'actions un principal spécifié peut effectuer sur cette ressource et dans quelles conditions. Vous devez [spécifier un principal](#) dans une politique basée sur les ressources. Les principaux peuvent inclure des comptes, des utilisateurs, des rôles, des utilisateurs fédérés ou. Services AWS

Les politiques basées sur les ressources sont des politiques en ligne situées dans ce service. Vous ne pouvez pas utiliser les politiques AWS gérées par IAM dans une stratégie basée sur les ressources.

Autres types de politique

AWS prend en charge d'autres types de politiques moins courants. Ces types de politiques peuvent définir le nombre maximum d'autorisations qui vous sont accordées par des types de politiques plus courants.

- **Limite d'autorisations** : une limite d'autorisations est une fonctionnalité avancée dans laquelle vous définissez le nombre maximal d'autorisations qu'une politique basée sur l'identité peut accorder à une entité IAM (utilisateur ou rôle IAM). Vous pouvez définir une limite d'autorisations pour une entité. Les autorisations obtenues représentent la combinaison des politiques basées sur l'identité de l'entité et de ses limites d'autorisations. Les politiques basées sur les ressources qui spécifient l'utilisateur ou le rôle dans le champ `Principal` ne sont pas limitées par les limites d'autorisations. Un refus explicite dans l'une de ces politiques annule l'autorisation. Pour plus d'informations sur les limites d'autorisations, consultez [Limites d'autorisations pour des entités IAM](#) dans le Guide de l'utilisateur IAM.

- **Politiques de contrôle des services (SCPs) :** SCPs politiques JSON qui spécifient les autorisations maximales pour une organisation ou une unité organisationnelle (UO) dans AWS Organizations. AWS Organizations est un service permettant de regrouper et de gérer de manière centralisée Comptes AWS les multiples propriétés de votre entreprise. Si vous activez toutes les fonctionnalités d'une organisation, vous pouvez appliquer des politiques de contrôle des services (SCPs) à l'un ou à l'ensemble de vos comptes. Le SCP limite les autorisations pour les entités figurant dans les comptes des membres, y compris chacune Utilisateur racine d'un compte AWS d'entre elles. Pour plus d'informations sur les Organizations SCPs, voir [Comment SCPs travailler](#) dans le Guide de AWS Organizations l'utilisateur.
- **Politiques de séance :** les politiques de séance sont des politiques avancées que vous utilisez en tant que paramètre lorsque vous créez par programmation une séance temporaire pour un rôle ou un utilisateur fédéré. Les autorisations de séance en résultant sont une combinaison des politiques basées sur l'identité de l'utilisateur ou du rôle et des politiques de séance. Les autorisations peuvent également provenir d'une politique basée sur les ressources. Un refus explicite dans l'une de ces politiques annule l'autorisation. Pour plus d'informations, consultez [Politiques de session](#) dans le Guide de l'utilisateur IAM.

Plusieurs types de politique

Lorsque plusieurs types de politiques s'appliquent à la requête, les autorisations en résultant sont plus compliquées à comprendre. Pour savoir comment AWS déterminer s'il faut autoriser une demande lorsque plusieurs types de politiques sont impliqués, consultez la section [Logique d'évaluation des politiques](#) dans le guide de l'utilisateur IAM.

Comment AWS Clean Rooms fonctionne avec IAM

Avant d'utiliser IAM pour gérer l'accès à AWS Clean Rooms, découvrez les fonctionnalités IAM disponibles. AWS Clean Rooms

Fonctionnalités IAM que vous pouvez utiliser avec AWS Clean Rooms

Fonctionnalité IAM	AWS Clean Rooms soutien
Politiques basées sur l'identité	Oui
Politiques basées sur les ressources	Partielle

Fonctionnalité IAM	AWS Clean Rooms soutien
Actions de politique	Oui
Ressources de politique	Oui
Clés de condition de politique (spécifiques au service)	Partielle
ACLs	Non
ABAC (étiquettes dans les politiques)	Oui
Informations d'identification temporaires	Oui
Transmission des sessions d'accès (FAS)	Oui
Rôles de service	Oui
Rôles liés à un service	Non

Pour obtenir une vue d'ensemble de la façon dont la plupart des fonctionnalités IAM Services AWS fonctionnent AWS Clean Rooms et des autres fonctionnalités, consultez Services AWS le [guide de l'utilisateur IAM](#) consacré à leur utilisation.

Politiques basées sur l'identité pour AWS Clean Rooms

Prend en charge les politiques basées sur l'identité : oui

Les politiques basées sur l'identité sont des documents de politique d'autorisations JSON que vous pouvez attacher à une identité telle qu'un utilisateur, un groupe d'utilisateurs ou un rôle IAM. Ces politiques contrôlent quel type d'actions des utilisateurs et des rôles peuvent exécuter, sur quelles ressources et dans quelles conditions. Pour découvrir comment créer une politique basée sur l'identité, consultez [Définition d'autorisations IAM personnalisées avec des politiques gérées par le client](#) dans le Guide de l'utilisateur IAM.

Avec les politiques IAM basées sur l'identité, vous pouvez spécifier des actions et ressources autorisées ou refusées, ainsi que les conditions dans lesquelles les actions sont autorisées ou refusées. Pour découvrir tous les éléments que vous utilisez dans une politique JSON, consultez [Références des éléments de politique JSON IAM](#) dans le Guide de l'utilisateur IAM.

Exemples de politiques basées sur l'identité pour AWS Clean Rooms

Pour consulter des exemples de politiques AWS Clean Rooms basées sur l'identité, consultez.

[Exemples de politiques basées sur l'identité pour AWS Clean Rooms](#)

Politiques basées sur les ressources au sein de AWS Clean Rooms

Prise en charge des politiques basées sur les ressources : partielle

Les politiques basées sur les ressources sont des documents de politique JSON que vous attachez à une ressource. Par exemple, les politiques de confiance de rôle IAM et les politiques de compartiment Amazon S3 sont des politiques basées sur les ressources. Dans les services qui sont compatibles avec les politiques basées sur les ressources, les administrateurs de service peuvent les utiliser pour contrôler l'accès à une ressource spécifique. Pour la ressource dans laquelle se trouve la politique, cette dernière définit quel type d'actions un principal spécifié peut effectuer sur cette ressource et dans quelles conditions. Vous devez [spécifier un principal](#) dans une politique basée sur les ressources. Les principaux peuvent inclure des comptes, des utilisateurs, des rôles, des utilisateurs fédérés ou. Services AWS

Pour permettre un accès intercompte, vous pouvez spécifier un compte entier ou des entités IAM dans un autre compte en tant que principal dans une politique basée sur les ressources. Pour plus d'informations, consultez [Accès intercompte aux ressources dans IAM](#) dans le Guide de l'utilisateur IAM.

Le AWS Clean Rooms service ne prend en charge qu'un seul type de politique basée sur les ressources, appelée politique de ressources gérées par modèle similaire configuré, qui est attachée à un modèle similaire configuré. Cette politique définit les principaux autorisés à effectuer des actions sur le modèle similaire configuré.

Pour savoir comment associer une politique basée sur les ressources à un modèle similaire configuré, consultez. [Comportements IAM pour AWS Clean Rooms ML](#)

Actions politiques pour AWS Clean Rooms

Prend en charge les actions de politique : oui

Les administrateurs peuvent utiliser les politiques AWS JSON pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'élément `Action` d'une politique JSON décrit les actions que vous pouvez utiliser pour autoriser ou refuser l'accès à une politique. Intégration d'actions dans une politique afin d'accorder l'autorisation d'exécuter les opérations associées.

Pour consulter la liste des AWS Clean Rooms actions, voir [Actions définies par AWS Clean Rooms](#) dans la référence d'autorisation de service.

Les actions de politique en AWS Clean Rooms cours utilisent le préfixe suivant avant l'action.

```
cleanrooms
```

Pour indiquer plusieurs actions dans une seule déclaration, séparez-les par des virgules.

```
"Action": [  
    "cleanrooms:action1",  
    "cleanrooms:action2"  
]
```

Pour consulter des exemples de politiques AWS Clean Rooms basées sur l'identité, consultez [Exemples de politiques basées sur l'identité pour AWS Clean Rooms](#)

Ressources politiques pour AWS Clean Rooms

Prend en charge les ressources de politique : oui

Les administrateurs peuvent utiliser les politiques AWS JSON pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'élément de politique JSON `Resource` indique le ou les objets auxquels l'action s'applique. Il est recommandé de définir une ressource à l'aide de son [Amazon Resource Name \(ARN\)](#). Pour les actions qui ne sont pas compatibles avec les autorisations de niveau ressource, utilisez un caractère générique (*) afin d'indiquer que l'instruction s'applique à toutes les ressources.

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

Pour consulter la liste des types de AWS Clean Rooms ressources et leurs caractéristiques ARNs, consultez la section [Ressources définies par AWS Clean Rooms](#) dans la référence d'autorisation

de service. Pour savoir grâce à quelles actions vous pouvez spécifier l'ARN de chaque ressource, consultez [Actions définies par AWS Clean Rooms](#).

Pour consulter des exemples de politiques AWS Clean Rooms basées sur l'identité, consultez [Exemples de politiques basées sur l'identité pour AWS Clean Rooms](#)

Clés de conditions de politique pour AWS Clean Rooms

Prend en charge les clés de condition de politique spécifiques au service : partiel

Les administrateurs peuvent utiliser les politiques AWS JSON pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'élément `Condition` indique à quel moment les instructions s'exécutent en fonction de critères définis. Vous pouvez créer des expressions conditionnelles qui utilisent des [opérateurs de condition](#), tels que les signes égal ou inférieur à, pour faire correspondre la condition de la politique aux valeurs de la demande. Pour voir toutes les clés de condition AWS globales, voir les clés de [contexte de condition AWS globales](#) dans le guide de l'utilisateur IAM.

Pour savoir comment AWS Clean Rooms ML utilise les clés de conditions de politique, consultez [Comportements IAM pour AWS Clean Rooms ML](#).

ACLs in AWS Clean Rooms

Supports ACLs : Non

Les listes de contrôle d'accès (ACLs) contrôlent les principaux (membres du compte, utilisateurs ou rôles) autorisés à accéder à une ressource. ACLs sont similaires aux politiques basées sur les ressources, bien qu'elles n'utilisent pas le format de document de politique JSON.

ABAC avec AWS Clean Rooms

Prise en charge d'ABAC (balises dans les politiques) : Oui

Le contrôle d'accès par attributs (ABAC) est une stratégie d'autorisation qui définit les autorisations en fonction des attributs appelés balises. Vous pouvez associer des balises aux entités et aux AWS ressources IAM, puis concevoir des politiques ABAC pour autoriser les opérations lorsque la balise du principal correspond à la balise de la ressource.

Pour contrôler l'accès basé sur des étiquettes, vous devez fournir les informations d'étiquette dans l'[élément de condition](#) d'une politique utilisant les clés de condition `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` ou `aws:TagKeys`.

Si un service prend en charge les trois clés de condition pour tous les types de ressources, alors la valeur pour ce service est Oui. Si un service prend en charge les trois clés de condition pour certains types de ressources uniquement, la valeur est Partielle.

Pour plus d'informations sur ABAC, consultez [Définition d'autorisations avec l'autorisation ABAC](#) dans le Guide de l'utilisateur IAM. Pour accéder à un didacticiel décrivant les étapes de configuration de l'ABAC, consultez [Utilisation du contrôle d'accès par attributs \(ABAC\)](#) dans le Guide de l'utilisateur IAM.

Utilisation d'informations d'identification temporaires avec AWS Clean Rooms

Prend en charge les informations d'identification temporaires : oui

Les informations d'identification temporaires fournissent un accès à court terme aux AWS ressources et sont automatiquement créées lorsque vous utilisez la fédération ou que vous changez de rôle. AWS recommande de générer dynamiquement des informations d'identification temporaires au lieu d'utiliser des clés d'accès à long terme. Pour plus d'informations, consultez [Informations d'identification de sécurité temporaires dans IAM](#) et [Services AWS compatibles avec IAM](#) dans le Guide de l'utilisateur IAM.

Transférer les sessions d'accès pour AWS Clean Rooms

Prend en charge les sessions d'accès direct (FAS) : oui

Les sessions d'accès direct (FAS) utilisent les autorisations du principal appelant et Service AWS, combinées Service AWS à la demande d'envoi de demandes aux services en aval. Pour plus de détails sur la politique relative à la transmission de demandes FAS, consultez la section [Sessions de transmission d'accès](#).

Rôles de service pour AWS Clean Rooms

Prend en charge les rôles de service : oui

Un rôle de service est un [rôle IAM](#) qu'un service endosse pour accomplir des actions en votre nom. Un administrateur IAM peut créer, modifier et supprimer un rôle de service à partir d'IAM. Pour plus d'informations, consultez [Création d'un rôle pour la délégation d'autorisations à un Service AWS](#) dans le Guide de l'utilisateur IAM.

 Warning

La modification des autorisations associées à un rôle de service peut perturber AWS Clean Rooms les fonctionnalités. Modifiez les rôles de service uniquement lorsque AWS Clean Rooms vous êtes invité à le faire.

Rôles liés à un service pour AWS Clean Rooms

Prend en charge les rôles liés à un service : non

Un rôle lié à un service est un type de rôle de service lié à un. Service AWS Le service peut endosser le rôle afin d'effectuer une action en votre nom. Les rôles liés à un service apparaissent dans votre Compte AWS répertoire et appartiennent au service. Un administrateur IAM peut consulter, mais ne peut pas modifier, les autorisations concernant les rôles liés à un service.

Pour plus d'informations sur la création ou la gestion des rôles liés à un service, consultez [Services AWS qui fonctionnent avec IAM](#). Recherchez un service dans le tableau qui inclut un Yes dans la colonne Rôle lié à un service. Choisissez le lien Oui pour consulter la documentation du rôle lié à ce service.

Exemples de politiques basées sur l'identité pour AWS Clean Rooms

Par défaut, les utilisateurs et les rôles ne sont pas autorisés à créer ou modifier les ressources AWS Clean Rooms . Pour octroyer aux utilisateurs des autorisations d'effectuer des actions sur les ressources dont ils ont besoin, un administrateur IAM peut créer des politiques IAM.

Pour apprendre à créer une politique basée sur l'identité IAM à l'aide de ces exemples de documents de politique JSON, consultez [Création de politiques IAM \(console\)](#) dans le Guide de l'utilisateur IAM.

Pour plus de détails sur les actions et les types de ressources définis par AWS Clean Rooms, y compris le format de ARNs pour chacun des types de ressources, voir [Actions, ressources et clés de condition AWS Clean Rooms](#) dans la référence d'autorisation de service.

Rubriques

- [Bonnes pratiques en matière de politiques](#)
- [Utilisation de la AWS Clean Rooms console](#)
- [Autorisation accordée aux utilisateurs pour afficher leurs propres autorisations](#)

Bonnes pratiques en matière de politiques

Les politiques basées sur l'identité déterminent si quelqu'un peut créer, accéder ou supprimer AWS Clean Rooms des ressources dans votre compte. Ces actions peuvent entraîner des frais pour votre Compte AWS. Lorsque vous créez ou modifiez des politiques basées sur l'identité, suivez ces instructions et recommandations :

- Commencez AWS par les politiques gérées et passez aux autorisations du moindre privilège : pour commencer à accorder des autorisations à vos utilisateurs et à vos charges de travail, utilisez les politiques AWS gérées qui accordent des autorisations pour de nombreux cas d'utilisation courants. Ils sont disponibles dans votre Compte AWS. Nous vous recommandons de réduire davantage les autorisations en définissant des politiques gérées par les AWS clients spécifiques à vos cas d'utilisation. Pour plus d'informations, consultez [politiques gérées par AWS](#) ou [politiques gérées par AWS pour les activités professionnelles](#) dans le Guide de l'utilisateur IAM.
- Accordez les autorisations de moindre privilège : lorsque vous définissez des autorisations avec des politiques IAM, accordez uniquement les autorisations nécessaires à l'exécution d'une seule tâche. Pour ce faire, vous définissez les actions qui peuvent être entreprises sur des ressources spécifiques dans des conditions spécifiques, également appelées autorisations de moindre privilège. Pour plus d'informations sur l'utilisation d'IAM pour appliquer des autorisations, consultez [politiques et autorisations dans IAM](#) dans le Guide de l'utilisateur IAM.
- Utilisez des conditions dans les politiques IAM pour restreindre davantage l'accès : vous pouvez ajouter une condition à vos politiques afin de limiter l'accès aux actions et aux ressources. Par exemple, vous pouvez écrire une condition de politique pour spécifier que toutes les demandes doivent être envoyées via SSL. Vous pouvez également utiliser des conditions pour accorder l'accès aux actions de service si elles sont utilisées par le biais d'un service spécifique Service AWS, tel que CloudFormation. Pour plus d'informations, consultez [Conditions pour éléments de politique JSON IAM](#) dans le Guide de l'utilisateur IAM.
- Utilisez l'Analyseur d'accès IAM pour valider vos politiques IAM afin de garantir des autorisations sécurisées et fonctionnelles : l'Analyseur d'accès IAM valide les politiques nouvelles et existantes de manière à ce que les politiques IAM respectent le langage de politique IAM (JSON) et les bonnes pratiques IAM. IAM Access Analyzer fournit plus de 100 vérifications de politiques et des recommandations exploitables pour vous aider à créer des politiques sécurisées et fonctionnelles. Pour plus d'informations, consultez [Validation de politiques avec IAM Access Analyzer](#) dans le Guide de l'utilisateur IAM.
- Exiger l'authentification multifactorielle (MFA) : si vous avez un scénario qui nécessite des utilisateurs IAM ou un utilisateur root, activez l'authentification MFA pour une sécurité accrue.

Compte AWS Pour exiger la MFA lorsque des opérations d'API sont appelées, ajoutez des conditions MFA à vos politiques. Pour plus d'informations, consultez [Sécurisation de l'accès aux API avec MFA](#) dans le Guide de l'utilisateur IAM.

Pour plus d'informations sur les bonnes pratiques dans IAM, consultez [Bonnes pratiques de sécurité dans IAM](#) dans le Guide de l'utilisateur IAM.

Utilisation de la AWS Clean Rooms console

Pour accéder à la AWS Clean Rooms console, vous devez disposer d'un ensemble minimal d'autorisations. Ces autorisations doivent vous permettre de répertorier et d'afficher les détails AWS Clean Rooms des ressources de votre Compte AWS. Si vous créez une politique basée sur l'identité qui est plus restrictive que l'ensemble minimum d'autorisations requis, la console ne fonctionnera pas comme prévu pour les entités (utilisateurs ou rôles) tributaires de cette politique.

Il n'est pas nécessaire d'accorder des autorisations de console minimales aux utilisateurs qui appellent uniquement l'API AWS CLI ou l' AWS API. Autorisez plutôt l'accès à uniquement aux actions qui correspondent à l'opération d'API qu'ils tentent d'effectuer.

Pour garantir que les utilisateurs et les rôles peuvent toujours utiliser la AWS Clean Rooms console, associez également la politique AWS Clean Rooms *FullAccess* ou la politique *ReadOnly* AWS gérée aux entités. Pour plus d'informations, consultez [Ajout d'autorisations à un utilisateur](#) dans le Guide de l'utilisateur IAM.

Autorisation accordée aux utilisateurs pour afficher leurs propres autorisations

Cet exemple montre comment créer une politique qui permet aux utilisateurs IAM d'afficher les politiques en ligne et gérées attachées à leur identité d'utilisateur. Cette politique inclut les autorisations permettant d'effectuer cette action sur la console ou par programmation à l'aide de l'API AWS CLI or AWS .

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ViewOwnUserInfo",
      "Effect": "Allow",
      "Action": [
        "iam:GetUserPolicy",
        "iam:ListGroupsForUser",
```

```

        "iam:ListAttachedUserPolicies",
        "iam:ListUserPolicies",
        "iam:GetUser"
    ],
    "Resource": ["arn:aws:iam::*:user/${aws:username}"]
},
{
    "Sid": "NavigateInConsole",
    "Effect": "Allow",
    "Action": [
        "iam:GetGroupPolicy",
        "iam:GetPolicyVersion",
        "iam:GetPolicy",
        "iam:ListAttachedGroupPolicies",
        "iam:ListGroupPolicies",
        "iam:ListPolicyVersions",
        "iam:ListPolicies",
        "iam:ListUsers"
    ],
    "Resource": "*"
}
]
}

```

AWS politiques gérées pour AWS Clean Rooms

Une politique AWS gérée est une politique autonome créée et administrée par AWS. AWS les politiques gérées sont conçues pour fournir des autorisations pour de nombreux cas d'utilisation courants afin que vous puissiez commencer à attribuer des autorisations aux utilisateurs, aux groupes et aux rôles.

N'oubliez pas que les politiques AWS gérées peuvent ne pas accorder d'autorisations de moindre privilège pour vos cas d'utilisation spécifiques, car elles sont accessibles à tous les AWS clients. Nous vous recommandons de réduire encore les autorisations en définissant des [politiques gérées par le client](#) qui sont propres à vos cas d'utilisation.

Vous ne pouvez pas modifier les autorisations définies dans les politiques AWS gérées. Si les autorisations définies dans une politique AWS gérée sont AWS mises à jour, la mise à jour affecte toutes les identités principales (utilisateurs, groupes et rôles) auxquelles la politique est attachée. AWS est le plus susceptible de mettre à jour une politique AWS gérée lorsqu'une nouvelle politique

Service AWS est lancée ou lorsque de nouvelles opérations d'API sont disponibles pour les services existants.

Pour plus d'informations, consultez [Politiques gérées par AWS](#) dans le Guide de l'utilisateur IAM.

AWS politique gérée : **AWSCleanRoomsReadOnlyAccess**

Vous pouvez vous rattacher **AWSCleanRoomsReadOnlyAccess** à vos principaux IAM.

Cette politique accorde des autorisations en lecture seule aux ressources et aux métadonnées dans le cadre d'une **AWSCleanRoomsReadOnlyAccess** collaboration.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes :

- **CleanRoomsRead**— Permet aux principaux d'accéder au service en lecture seule.
- **ConsoleDisplayTables**— Permet aux principaux d'accéder en lecture seule aux AWS Glue métadonnées nécessaires pour afficher les données relatives aux AWS Glue tables sous-jacentes sur la console.
- **ConsoleLogSummaryQueryLogs**— Permet aux principaux de consulter les journaux de requêtes.
- **ConsoleLogSummaryObtainLogs**— Permet aux principaux de récupérer les résultats du journal.

Pour obtenir une liste JSON des détails de la politique, consultez

[AWSCleanRoomsReadOnlyAccess](#) le Guide de référence des politiques AWS gérées.

AWS politique gérée : **AWSCleanRoomsFullAccess**

Vous pouvez vous rattacher **AWSCleanRoomsFullAccess** à vos principaux IAM.

Cette politique accorde des autorisations administratives qui permettent un accès complet (lecture, écriture et mise à jour) aux ressources et aux métadonnées dans le cadre d'une AWS Clean Rooms collaboration. Cette politique inclut l'accès pour effectuer des requêtes.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes :

- **CleanRoomsAccess**— Accorde un accès complet à toutes les actions sur toutes les ressources pour AWS Clean Rooms.

- **PassServiceRole**— Accorde l'accès pour transmettre un rôle de service uniquement au service (**PassedToServicecondition**) dont le nom contient `cleanrooms` « ».
- **ListRolesToPickServiceRole**— Permet aux directeurs de répertorier tous leurs rôles afin de choisir un rôle de service lors de l'utilisation AWS Clean Rooms.
- **GetRoleAndListRolePoliciesToInspectServiceRole**— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- **ListPoliciesToInspectServiceRolePolicy**— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- **GetPolicyToInspectServiceRolePolicy**— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- **ConsoleDisplayTables**— Permet aux principaux d'accéder en lecture seule aux AWS Glue métadonnées nécessaires pour afficher les données relatives aux AWS Glue tables sous-jacentes sur la console.
- **ConsolePickQueryResultsBucketListAll**— Permet aux principaux de choisir un compartiment Amazon S3 dans une liste de tous les compartiments S3 disponibles dans lesquels les résultats de leurs requêtes sont écrits.
- **SetQueryResultsBucket**— Permet aux principaux de choisir un compartiment S3 dans lequel les résultats de leurs requêtes sont écrits.
- **ConsoleDisplayQueryResults**— Permet aux principaux d'afficher les résultats de la requête au client, lus depuis le compartiment S3.
- **WriteQueryResults**— Permet aux principaux d'écrire les résultats de la requête dans un compartiment S3 appartenant au client.
- **EstablishLogDeliveries**— Permet aux principaux de fournir des journaux de requêtes au groupe de CloudWatch journaux Amazon Logs d'un client.
- **SetupLogGroupsDescribe**— Permet aux principaux d'utiliser le processus de création de groupes de CloudWatch journaux Amazon Logs.
- **SetupLogGroupsCreate**— Permet aux principaux de créer un groupe de CloudWatch journaux Amazon Logs.
- **SetupLogGroupsResourcePolicy**— Permet aux principaux de définir une politique de ressources sur le groupe de CloudWatch journaux Amazon Logs.
- **ConsoleLogSummaryQueryLogs**— Permet aux principaux de consulter les journaux de requêtes.
- **ConsoleLogSummaryObtainLogs**— Permet aux principaux de récupérer les résultats du journal.

Pour obtenir une liste JSON des détails de la politique, consultez [AWSCleanRoomsFullAccessle](#) Guide de référence des politiques AWS gérées.

AWS politique gérée : **AWSCleanRoomsFullAccessNoQuerying**

Vous pouvez joindre `AWSCleanRoomsFullAccessNoQuerying` à votre IAM principaux.

Cette politique accorde des autorisations administratives qui permettent un accès complet (lecture, écriture et mise à jour) aux ressources et aux métadonnées dans le cadre d'une AWS Clean Rooms collaboration. Cette politique exclut l'accès pour effectuer des requêtes.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes :

- `CleanRoomsAccess`— Accorde un accès complet à toutes les actions sur toutes les ressources AWS Clean Rooms, à l'exception des requêtes dans le cadre de collaborations.
- `CleanRoomsNoQuerying`— Refuse explicitement `StartProtectedQuery` et `UpdateProtectedQuery` empêche les requêtes.
- `PassServiceRole`— Accorde l'accès pour transmettre un rôle de service uniquement au service (`PassedToServicecondition`) dont le nom contient `cleanrooms` « ».
- `ListRolesToPickServiceRole`— Permet aux directeurs de répertorier tous leurs rôles afin de choisir un rôle de service lors de l'utilisation AWS Clean Rooms.
- `GetRoleAndListRolePoliciesToInspectServiceRole`— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- `ListPoliciesToInspectServiceRolePolicy`— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- `GetPolicyToInspectServiceRolePolicy`— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- `ConsoleDisplayTables`— Permet aux principaux d'accéder en lecture seule aux AWS Glue métadonnées nécessaires pour afficher les données relatives aux AWS Glue tables sous-jacentes sur la console.
- `EstablishLogDeliveries`— Permet aux principaux de fournir des journaux de requêtes au groupe de CloudWatch journaux Amazon Logs d'un client.
- `SetupLogGroupsDescribe`— Permet aux principaux d'utiliser le processus de création de groupes de CloudWatch journaux Amazon Logs.

- `SetupLogGroupsCreate`— Permet aux principaux de créer un groupe de CloudWatch journaux Amazon Logs.
- `SetupLogGroupsResourcePolicy`— Permet aux principaux de définir une politique de ressources sur le groupe de CloudWatch journaux Amazon Logs.
- `ConsoleLogSummaryQueryLogs`— Permet aux principaux de consulter les journaux de requêtes.
- `ConsoleLogSummaryObtainLogs`— Permet aux principaux de récupérer les résultats du journal.
- `cleanrooms`— Gérez les collaborations, les modèles d'analyse, les tables configurées, les adhésions et les ressources associées au sein du AWS Clean Rooms service. Effectuez diverses opérations telles que la création, la mise à jour, la suppression, la liste et la récupération d'informations sur ces ressources.
- `iam`— Transmettez les rôles de service dont les noms contiennent `cleanrooms` « » au AWS Clean Rooms service. Répertoriez les rôles, les politiques et inspectez les rôles de service et les politiques liés au AWS Clean Rooms service.
- `glue`— Récupérez des informations sur les bases de données, les tables, les partitions et les schémas à partir de AWS Glue. Cela est nécessaire pour que le AWS Clean Rooms service affiche et interagisse avec les sources de données sous-jacentes.
- `logs`— Gérez les livraisons de journaux, les groupes de journaux et les politiques de ressources pour les CloudWatch journaux. Interrogez et récupérez les journaux relatifs au AWS Clean Rooms service. Ces autorisations sont nécessaires à des fins de surveillance, d'audit et de dépannage au sein du service.

La politique refuse également explicitement les actions `cleanrooms:StartProtectedQuery` et empêche les utilisateurs `cleanrooms:UpdateProtectedQuery` d'exécuter ou de mettre à jour directement les requêtes protégées, ce qui doit être fait par le biais des mécanismes AWS Clean Rooms contrôlés.

Pour obtenir une liste JSON des détails de la politique, consultez [AWSCleanRoomsFullAccessNoQuerying](#) le Guide de référence des politiques AWS gérées.

AWS politique gérée : **AWSCleanRoomsMLReadOnlyAccess**

Vous pouvez vous rattacher `AWSCleanRoomsMLReadOnlyAccess` à vos principaux IAM.

Cette politique accorde des autorisations en lecture seule aux ressources et aux métadonnées dans le cadre d'une `AWSCleanRoomsMLReadOnlyAccess` collaboration.

Cette politique inclut les autorisations suivantes :

- **CleanRoomsConsoleNavigation**— Permet d'accéder aux écrans de la AWS Clean Rooms console.
- **CleanRoomsMLRead**— Permet aux principaux d'accéder en lecture seule au service Clean Rooms ML.
- **PassCleanRoomsResources**— Permet de transmettre des AWS Clean Rooms ressources spécifiées.

Pour une liste JSON des détails de la politique, voir [AWSCleanRooms MLRead OnlyAccess](#) dans le AWS Managed Policy Reference Guide.

AWS politique gérée : **AWSCleanRoomsMLFullAccess**

Vous pouvez vous rattacher **AWSCleanRoomsMLFullAccess** à vos principaux IAM. Cette politique accorde des autorisations administratives qui permettent un accès complet (lecture, écriture et mise à jour) aux ressources et aux métadonnées nécessaires à Clean Rooms ML.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes :

- **CleanRoomsMLFullAccess**— Accorde l'accès à toutes les actions de Clean Rooms ML.
- **PassServiceRole**— Accorde l'accès pour transmettre un rôle de service uniquement au service (PassedToServicecondition) dont le nom contient cleanrooms-ml « ».
- **CleanRoomsConsoleNavigation**— Permet d'accéder aux écrans de la AWS Clean Rooms console.
- **CollaborationMembershipCheck**— Lorsque vous lancez une tâche de génération d'audience (segment similaire) dans le cadre d'une collaboration, le service Clean Rooms ML appelle **ListMembers** pour vérifier que la collaboration est valide, que l'appelant est un membre actif et que le propriétaire du modèle d'audience configuré est un membre actif. Cette autorisation est toujours requise ; le SID de navigation dans la console n'est requis que pour les utilisateurs de la console.
- **PassCleanRoomsResources**— Permet de transmettre des AWS Clean Rooms ressources spécifiées.
- **AssociateModels**— Permet aux directeurs d'associer un modèle Clean Rooms ML à votre collaboration.

- **TagAssociations**— Permet aux principaux d'ajouter des balises à l'association entre un modèle similaire et une collaboration.
- **ListRolesToPickServiceRole**— Permet aux directeurs de répertorier tous leurs rôles afin de choisir un rôle de service lors de l'utilisation AWS Clean Rooms.
- **GetRoleAndListRolePoliciesToInspectServiceRole**— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- **ListPoliciesToInspectServiceRolePolicy**— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- **GetPolicyToInspectServiceRolePolicy**— Permet aux principaux de voir le rôle du service et la politique correspondante dans IAM.
- **ConsoleDisplayTables**— Permet aux principaux d'accéder en lecture seule aux AWS Glue métadonnées nécessaires pour afficher les données relatives aux AWS Glue tables sous-jacentes sur la console.
- **ConsolePickOutputBucket**— Permet aux principaux de sélectionner des compartiments Amazon S3 pour les sorties du modèle d'audience configurées.
- **ConsolePickS3Location**— Permet aux principaux de sélectionner l'emplacement dans un compartiment pour les sorties du modèle d'audience configurées.
- **ConsoleDescribeECRRepositories**— Permet aux principaux de décrire les référentiels et les images Amazon ECR.

Pour une liste JSON des détails de la politique, voir [AWSCleanRooms MLFull Access](#) dans le AWS Managed Policy Reference Guide.

AWS Clean Rooms mises à jour des politiques AWS gérées

Consultez les détails des mises à jour des politiques AWS gérées AWS Clean Rooms depuis que ce service a commencé à suivre ces modifications. Pour recevoir des alertes automatiques concernant les modifications apportées à cette page, abonnez-vous au flux RSS sur la page Historique du AWS Clean Rooms document.

Modifier	Description	Date
AWSCleanRoomsFullAccessNoQuering — Mise à jour de la politique existante	Ajout de salles blanches : UpdateConfiguredTableAllowedColumns et de	29 juillet 2025

Modifier	Description	Date
	salles blanches : UpdateConfiguredTableReference à. CleanRoomsAccess	
AWSCleanRoomsMLReadOnlyAccess : mise à jour de la politique existante	Ajout de PassCleanRoomsResources à AWSCleanRoomsMLReadOnlyAccess. Ajout de PassCleanRoomsResources et ConsoleDescribeECRRepositories à AWSCleanRoomsMLFullAccess.	10 janvier 2025
AWSCleanRoomsFullAccessNoQuering : mise à jour de la politique existante	Ajout de cleanrooms:BatchGetSchemaAnalysisRule à CleanRoomsAccess.	13 mai 2024
AWSCleanRoomsFullAccess : mise à jour de la politique existante	L'ID AWSCleanRoomsFullAccess de déclaration a été mis ConsolePickQueryResultsBucket à jour SetQueryResultsBucket dans cette politique afin de mieux représenter les autorisations, car celles-ci sont nécessaires pour définir le compartiment des résultats des requêtes avec et sans la console.	21 mars 2024
AWSCleanRoomsMLReadOnlyAccess : nouvelle politique AWSCleanRoomsMLFullAccess : nouvelle politique	Ajouté AWSCleanRoomsMLReadOnlyAccess et AWSCleanRoomsMLFullAccess pour prendre en charge AWS Clean Rooms ML.	29 novembre 2023

Modifier	Description	Date
AWSCleanRoomsFullAccessNoQuering : mise à jour de la politique existante	Ajout de cleanrooms:CreateAnalysisTemplate,cleanrooms:GetAnalysisTemplate,cleanrooms:UpdateAnalysisTemplate, cleanrooms:DeleteAnalysisTemplate,cleanrooms:ListAnalysisTemplates,cleanrooms:GetCollaborationAnalysisTemplate,cleanrooms:BatchGetCollaborationAnalysisTemplate, et cleanrooms:ListCollaborationAnalysisTemplates CleanRoomsAccess pour activer la nouvelle fonctionnalité de modèles d'analyse.	31 juillet 2023
AWSCleanRoomsFullAccessNoQuering : mise à jour de la politique existante	Ajoutécleanrooms:ListTagsForResource,cleanrooms:UntagResource, et cleanrooms:TagResource pour CleanRoomsAccess activer le balisage des ressources.	21 mars 2023
AWS Clean Rooms a commencé à suivre les modifications	AWS Clean Rooms a commencé à suivre les modifications apportées AWS à ses politiques gérées.	12 janvier 2023

Résolution des problèmes AWS Clean Rooms d'identité et d'accès

Utilisez les informations suivantes pour vous aider à diagnostiquer et à résoudre les problèmes courants que vous pouvez rencontrer lorsque vous travaillez avec AWS Clean Rooms IAM.

Rubriques

- [Je ne suis pas autorisé à effectuer une action dans AWS Clean Rooms](#)
- [Je ne suis pas autorisé à effectuer iam : PassRole](#)
- [Je souhaite permettre à des personnes extérieures Compte AWS à moi d'accéder à mes AWS Clean Rooms ressources](#)

Je ne suis pas autorisé à effectuer une action dans AWS Clean Rooms

Si vous recevez une erreur qui indique que vous n'êtes pas autorisé à effectuer une action, vos politiques doivent être mises à jour afin de vous permettre d'effectuer l'action.

L'exemple d'erreur suivant se produit quand l'utilisateur IAM `mateojackson` tente d'utiliser la console pour afficher des informations détaillées sur une ressource `my-example-widget` fictive, mais ne dispose pas des autorisations `cleanrooms:GetWidget` fictives.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
cleanrooms:GetWidget on resource: my-example-widget
```

Dans ce cas, la stratégie de Mateo doit être mise à jour pour l'autoriser à accéder à la ressource `my-example-widget` à l'aide de l'action `cleanrooms:GetWidget`.

Si vous avez besoin d'aide, contactez votre AWS administrateur. Votre administrateur vous a fourni vos informations d'identification de connexion.

Je ne suis pas autorisé à effectuer iam : PassRole

Si vous recevez une erreur selon laquelle vous n'êtes pas autorisé à exécuter `iam:PassRole` l'action, vos stratégies doivent être mises à jour afin de vous permettre de transmettre un rôle à AWS Clean Rooms.

Certains services AWS permettent de transmettre un rôle existant à ce service au lieu de créer un nouveau rôle de service ou un rôle lié à un service. Pour ce faire, vous devez disposer des autorisations nécessaires pour transmettre le rôle au service.

L'exemple d'erreur suivant se produit lorsqu'un utilisateur IAM nommé `marymajor` essaie d'utiliser la console pour exécuter une action dans AWS Clean Rooms. Toutefois, l'action nécessite que le service ait des autorisations accordées par une fonction de service. Mary n'est pas autorisée à transmettre le rôle au service.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
iam:PassRole
```

Dans ce cas, les politiques de Mary doivent être mises à jour pour lui permettre d'exécuter l'action `iam:PassRole`.

Si vous avez besoin d'aide, contactez votre AWS administrateur. Votre administrateur vous a fourni vos informations d'identification de connexion.

Je souhaite permettre à des personnes extérieures Compte AWS à moi d'accéder à mes AWS Clean Rooms ressources

Vous pouvez créer un rôle que les utilisateurs provenant d'autres comptes ou les personnes extérieures à votre organisation pourront utiliser pour accéder à vos ressources. Vous pouvez spécifier qui est autorisé à assumer le rôle.

Pour plus d'informations, consultez les éléments suivants :

- Pour savoir si ces fonctionnalités sont prises AWS Clean Rooms en charge, consultez [Comment AWS Clean Rooms fonctionne avec IAM](#).
- Pour savoir comment fournir l'accès à vos ressources sur celles Comptes AWS que vous possédez, consultez la section [Fournir l'accès à un utilisateur IAM dans un autre utilisateur Compte AWS que vous possédez](#) dans le Guide de l'utilisateur IAM.
- Pour savoir comment fournir l'accès à vos ressources à des tiers Comptes AWS, consultez la section [Fournir un accès à des ressources Comptes AWS détenues par des tiers](#) dans le guide de l'utilisateur IAM.
- Pour savoir comment fournir un accès par le biais de la fédération d'identité, consultez [Fournir un accès à des utilisateurs authentifiés en externe \(fédération d'identité\)](#) dans le Guide de l'utilisateur IAM.
- Pour découvrir quelle est la différence entre l'utilisation des rôles et l'utilisation des stratégies basées sur les ressources pour l'accès comptes multiples, consultez [Différence entre les rôles IAM et les stratégies basées sur les ressources](#) dans le Guide de l'utilisateur IAM.

Prévention du problème de l'adjoint confus entre services

Le problème de député 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 un problème de confusion chez les 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 avec des principaux de service qui ont eu accès aux ressources de votre compte.

Nous recommandons d'utiliser les clés de contexte de condition [aws:SourceArn](#) globale dans les politiques de ressources afin de limiter les autorisations que AWS Clean Rooms accordent un autre

service à la ressource. Utilisez `aws:SourceArn` si vous souhaitez qu'une seule ressource soit associée à l'accès entre services.

Le moyen le plus efficace de se protéger contre le problème de député confus consiste à utiliser la clé de contexte de condition globale `aws:SourceArn` avec l'ARN complet de la ressource. Dans AWS Clean Rooms, vous devez également comparer avec la clé de `sts:ExternalId` condition.

La valeur de `aws:SourceArn` doit être définie sur l'ARN de l'appartenance au rôle assumé.

L'exemple suivant montre comment vous pouvez utiliser la clé de contexte de condition `aws:SourceArn` globale AWS Clean Rooms pour éviter le problème de confusion des adjoints.

Note

L'exemple de politique s'applique à la politique de confiance du rôle de service AWS Clean Rooms utilisé pour accéder aux données et aux métadonnées d'une table configurée.

La valeur de `<query-runner-membership-id>` doit être définie sur l'ID de membre du lanceur de requêtes.

Tous les membres de la collaboration peuvent consulter les métadonnées de table configurées, de sorte que chaque ARN d'adhésion doit être inclus dans la liste des membres ARNs.

Note

Lorsqu'un rôle de service est créé via la AWS Clean Rooms console, tous les membres actuels de la collaboration sont inclus par défaut dans la condition d'adjoint confus.

Si vous ajoutez de nouveaux membres à une collaboration à laquelle des tables configurées sont déjà associées, veillez à mettre à jour la condition d'adjoint confuse associée à votre rôle de service avec l'ARN de membre du nouveau membre.

Si vous ne mettez pas à jour la condition d'adjoint confuse associée à votre rôle de service après avoir ajouté un nouveau membre, ce nouveau membre ne pourra pas accéder aux informations récupérées à AWS Clean Rooms l'aide de ce rôle.

JSON

```
{
```



```

"Version": "2012-10-17",
"Statement": [
  {
    "Sid": "AllowIfExternalIdMatches",
    "Effect": "Allow",
    "Principal": {
      "Service": "cleanrooms.amazonaws.com"
    },
    "Action": "sts:AssumeRole",
    "Condition": {
      "StringLike": {
        "sts:ExternalId": "arn:aws:*:us-east-1*:dbuser:*/<query-
runner-membership-id>*"
      }
    }
  },
  {
    "Sid": "AllowIfSourceArnMatches",
    "Effect": "Allow",
    "Principal": {
      "Service": "cleanrooms.amazonaws.com"
    },
    "Action": "sts:AssumeRole",
    "Condition": {
      "ForAnyValue:ArnEquals": {
        "aws:SourceArn": [
          "arn:aws:cleanrooms:us-
east-1:111122223333:membership/<member-1-membership-id>",
          "arn:aws:cleanrooms:us-
east-1:444455556666:membership/<member-2-membership-id>"
        ]
      }
    }
  }
]
}

```

Comportements IAM pour AWS Clean Rooms ML

Emplois multi-comptes

Clean Rooms ML permet à une autre personne d'accéder en toute sécurité Compte AWS à certaines ressources créées par l'un sur son compte Compte AWS. Lorsqu'un client situé dans Compte AWS A fait appel `StartAudienceGenerationJob` à une `ConfiguredAudienceModel` ressource appartenant à Compte AWS B, Clean Rooms ML en crée deux ARNs pour la tâche. Un ARN dans Compte AWS A et un autre dans Compte AWS B. Ils ARNs sont identiques sauf pour le leur Compte AWS.

Clean Rooms ML en crée deux ARNs pour les tâches afin que les deux comptes puissent appliquer leurs propres politiques IAM aux tâches. Par exemple, les deux comptes peuvent utiliser le contrôle d'accès basé sur des balises et appliquer les politiques de leur AWS organisation. La tâche traite les données des deux comptes, de sorte que les deux comptes peuvent supprimer la tâche et les données associées. Aucun des deux comptes ne peut empêcher l'autre compte de supprimer la tâche.

Il n'y a qu'une seule exécution de tâche et les deux comptes peuvent voir la tâche lorsqu'ils appellent `ListAudienceGenerationJobs`. Les deux comptes peuvent appeler le `GetDelete`, et `Export APIs` au travail en utilisant l'ARN avec leur propre Compte AWS identifiant.

Aucun des deux ne Compte AWS peut accéder à la tâche en utilisant un ARN avec l'autre Compte AWS ID.

Le nom de la tâche doit être unique dans un Compte AWS. Le nom en Compte AWS B est `$accountA-$name`. Le nom choisi par Compte AWS A est préfixé par Compte AWS A lorsque le travail est affiché dans B. Compte AWS

Pour qu'un compte croisé réussisse, Compte AWS B doit autoriser cette action `StartAudienceGenerationJob` à la fois sur la nouvelle tâche en Compte AWS B et sur la nouvelle tâche `ConfiguredAudienceModel` en Compte AWS B en utilisant une politique de ressources similaire à l'exemple suivant :

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
```

```

    "Sid": "Clean-Rooms-CAMA-ID",
    "Effect": "Allow",
    "Principal": {
        "AWS": [
            "111122223333"
        ]
    },
    "Action": [
        "cleanrooms-ml:StartAudienceGenerationJob"
    ],
    "Resource": [
        "arn:aws:cleanrooms-ml:us-east-1:444455556666:configured-audience-model/id",
        "arn:aws:cleanrooms-ml:us-east-1:444455556666:audience-generation-job/*"
    ],
    "Condition": {"StringEquals": {"cleanrooms-ml:CollaborationId": "UUID"}}
}
]
}

```

Note

Cette politique de ressources AWS Clean Rooms ML fait référence à deux solutions différentes Compte AWS IDs pour favoriser la génération d'audience entre comptes :

- 111122223333 - Il s'agit du compte qui contient le principal (utilisateur, rôle ou service) autorisé à démarrer des tâches de génération d'audience. Ce compte lance le flux de travail de traitement ML.
- 444455556666 - Il s'agit du compte propriétaire des ressources AWS Clean Rooms ML (le modèle d'audience configuré et les tâches de génération d'audience). Ce compte héberge les modèles ML et gère l'exécution des tâches.

Remarques de configuration supplémentaires :

- ID de déclaration (Sid) : remplacez-le CAMA-ID par votre identifiant actuel d' AWS Clean Rooms Audience Model Application (CAMA) pour que la déclaration de politique soit facilement identifiable.

- Ressource IDs : *id* remplacez-le par l'identifiant réel de votre modèle d'audience configuré et *UUID* par votre identifiant de collaboration spécifique.
- Condition : cette `cleanrooms-ml:CollaborationId` condition garantit que les tâches de génération d'audience ne peuvent être démarrées que dans le contexte de la AWS Clean Rooms collaboration spécifiée, fournissant ainsi une limite de sécurité supplémentaire.

Cette configuration multi-comptes permet des scénarios dans lesquels une organisation gère les modèles et l'infrastructure de machine learning tout en permettant aux partenaires autorisés de lancer des processus de génération d'audience dans les limites de leur accord de collaboration.

Si vous utilisez l'[API AWS Clean Rooms ML](#) pour créer un modèle similaire configuré avec `manageResourcePolicies` défini sur `true`, AWS Clean Rooms créez cette politique pour vous.

De plus, la politique d'identité de l'appelant dans Compte AWS A doit être `StartAudienceGenerationJob` autorisée. `arn:aws:cleanrooms-ml:us-west-1:AccountA:audience-generation-job/*` Il existe donc trois ressources IAM pour agir `StartAudienceGenerationJob` : la tâche Compte AWS A, la tâche Compte AWS B et la tâche Compte AWS B. `ConfiguredAudienceModel`

Warning

La Compte AWS personne qui a démarré la tâche reçoit un événement du journal AWS CloudTrail d'audit concernant la tâche. Le Compte AWS propriétaire du `ConfiguredAudienceModel` ne reçoit aucun événement du journal d' AWS CloudTrail audit.

Tâches de balisage

Lorsque vous définissez le `childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE` paramètre de `CreateConfiguredAudienceModel`, toutes les tâches de génération de segments similaires de votre compte créées à partir de ce modèle de similarité configuré comportent par défaut les mêmes balises que le modèle de similarité configuré. Le modèle de similarité configuré est le parent et la tâche de génération de segments de similarité est l'enfant.

Si vous créez une tâche dans votre propre compte, les balises de requête de la tâche remplacent les balises parentes. Les offres d'emploi créées par d'autres comptes ne créent jamais de tags dans votre compte. Si vous définissez une tâche `childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE` et qu'un autre compte la crée, il existe deux copies de la tâche. La copie de votre compte contient les balises de ressource parent et la copie du compte de l'auteur de la tâche contient les balises de la demande.

Validation des collaborateurs

Lorsque vous accordez des autorisations à d'autres membres d'une AWS Clean Rooms collaboration, la politique de ressources doit inclure la clé de condition `cleanrooms-ml:CollaborationId`. Cela garantit que le `collaborationId` paramètre est inclus dans la [StartAudienceGenerationJob](#) demande. Lorsque le `collaborationId` paramètre est inclus dans la demande, Clean Rooms ML confirme que la collaboration existe, que l'auteur de la tâche est un membre actif de la collaboration et que le propriétaire du modèle similaire configuré est un membre actif de la collaboration.

Lorsque AWS Clean Rooms vous gère la politique de ressources de votre modèle similaire configurée (le `manageResourcePolicies` paramètre est `TRUE` dans la [CreateConfiguredAudienceModelAssociation](#) demande), cette clé de condition sera définie dans la politique de ressources. Par conséquent, vous devez spécifier le `collaborationId` in [StartAudienceGenerationJob](#).

Accès intercomptes

Ne `StartAudienceGenerationJob` peut être appelé que sur plusieurs comptes. Tous les autres Clean Rooms ML ne APIs peuvent être utilisés qu'avec les ressources de votre propre compte. Cela garantit la confidentialité de vos données d'entraînement, de la configuration de votre modèle similaire et d'autres informations.

Clean Rooms ML ne révèle jamais Amazon S3 ni les AWS Glue emplacements d'un compte à l'autre. L'emplacement des données de formation, l'emplacement de sortie du modèle similaire configuré et l'emplacement de départ des tâches pour la génération de segments similaires ne sont jamais visibles sur tous les comptes. À moins que la journalisation des requêtes ne soit activée dans la collaboration, les données initiales ne sont pas visibles d'un compte à l'autre si les données de départ proviennent d'une requête SQL et si la requête elle-même n'est pas visible. Si vous avez `Get` une tâche de génération d'audience soumise par un autre compte, le service n'indique pas l'emplacement initial.

Comportements IAM pour les modèles personnalisés Clean Rooms ML

Emplois multi-comptes

Clean Rooms ML permet à une autre personne d'accéder en toute sécurité à certaines ressources associées Compte AWS à une collaboration créée par une personne depuis son compte Compte AWS. Un client de Compte AWS A ayant la capacité d'exécuter des requêtes peut appeler `CreateTrainedModelCreateMLInputChannel`, ou `StartTrainedModelInferenceJob` sur une `ConfiguredModelAlgorithmAssociation` ressource appartenant à un autre membre de la collaboration, à condition que cela `ConfiguredModelAlgorithmAssociation` soit autorisé par la règle d'analyse personnalisée créée avec `CreateConfiguredTableAnalysisRule`.

En outre, tout membre actif d'une collaboration peut supprimer les données associées à un modèle entraîné ou à un canal d'entrée ML via le `DeleteTrainedModelOutput` et `DeleteMLInputChannelData` APIs.

Accès intercomptes

Clean Rooms ML permet aux utilisateurs de récupérer des métadonnées sur les ressources créées par d'autres comptes via le `GetCollaboration` et `ListCollaboration` APIs. Clean Rooms ML ne révèle pas la clé KMS ARNs, les balises, les variables d'environnement ou les hyperparamètres (pour l'`TrainedModelAction`) aux autres comptes.

Accès à l'adhésion et à la collaboration

Lors de l'accès aux ressources d'adhésion et de collaboration dans le contexte des modèles personnalisés de Clean Rooms ML, la politique d'identité d'un utilisateur nécessite des autorisations pour les actions `cleanrooms:PassMembershipcleanrooms:PassCollaboration`, ou les deux. Tous ceux APIs qui acceptent `membershipId` ont besoin de l'`cleanrooms:PassMembershipautorisation`, et tous ceux APIs qui acceptent `collaborationId` ont besoin de l'`cleanrooms:PassCollaborationautorisation`. Un exemple de politique d'identité pour un rôle pouvant être appelé `createTrainedModel` dans le contexte d'un identifiant de membre pouvant appeler `GetCollaborationTrainedModel` dans le contexte d'un identifiant de collaboration est fourni.

JSON

```
{
```

```

"Version": "2012-10-17",
"Statement": [
  {
    "Sid": "AllowCleanroomsMLActions",
    "Effect": "Allow",
    "Action": [
      "cleanrooms:PassCollaboration",
      "cleanrooms:PassMembership"
    ],
    "Resource": [
      "*"
    ]
  },
  {
    "Sid": "AllowMembershipAccess",
    "Effect": "Allow",
    "Action": [
      "cleanrooms:GetMembership"
    ],
    "Resource": [
      "arn:aws:cleanrooms:us-east-1:111122223333:membership/memberId"
    ]
  },
  {
    "Sid": "AllowCollaborationAccess",
    "Effect": "Allow",
    "Action": [
      "cleanrooms:GetCollaboration"
    ],
    "Resource": [
      "arn:aws:cleanrooms:us-east-1:111122223333:collaboration/collaborationId"
    ]
  }
]
}

```

Validation de conformité pour AWS Clean Rooms

Pour savoir si un [programme Services AWS de conformité Service AWS s'inscrit dans le champ d'application de programmes de conformité](#) spécifiques, consultez Services AWS la section de

conformité et sélectionnez le programme de conformité qui vous intéresse. Pour des informations générales, voir Programmes de [AWS conformité Programmes AWS](#) de .

Vous pouvez télécharger des rapports d'audit tiers à l'aide de AWS Artifact. Pour plus d'informations, voir [Téléchargement de rapports dans AWS Artifact](#) .

Votre responsabilité en matière de conformité lors de l'utilisation Services AWS est déterminée par la sensibilité de vos données, les objectifs de conformité de votre entreprise et les lois et réglementations applicables. Pour plus d'informations sur votre responsabilité en matière de conformité lors de l'utilisation Services AWS, consultez [AWS la documentation de sécurité](#).

Résilience dans AWS Clean Rooms

L'infrastructure AWS mondiale est construite autour des AWS régions et des zones de disponibilité. Les régions fournissent plusieurs zones de disponibilité physiquement séparées et isolées, reliées par un réseau à latence faible, à débit élevé et à forte redondance. Avec les zones de disponibilité, vous pouvez concevoir et exploiter des applications et des bases de données qui basculent automatiquement d'une zone à l'autre sans interruption. Les zones de disponibilité sont davantage disponibles, tolérantes aux pannes et ont une plus grande capacité de mise à l'échelle que les infrastructures traditionnelles à un ou plusieurs centres de données.

Pour plus d'informations sur AWS les régions et les zones de disponibilité, consultez la section [Infrastructure AWS mondiale](#).

Sécurité de l'infrastructure dans AWS Clean Rooms

En tant que service géré, AWS Clean Rooms il est protégé par la sécurité du réseau AWS mondial. Pour plus d'informations sur les services AWS de sécurité et sur la manière dont AWS l'infrastructure est protégée, consultez la section [Sécurité du AWS cloud](#). Pour concevoir votre AWS environnement en utilisant les meilleures pratiques en matière de sécurité de l'infrastructure, consultez la section [Protection de l'infrastructure](#) dans le cadre AWS bien architecturé du pilier de sécurité.

Vous utilisez des appels d'API AWS publiés pour accéder AWS Clean Rooms via le réseau. Les clients doivent prendre en charge les éléments suivants :

- Protocole TLS (Transport Layer Security). Nous exigeons TLS 1.2 et recommandons TLS 1.3.
- Ses suites de chiffrement PFS (Perfect Forward Secrecy) comme DHE (Ephemeral Diffie-Hellman) ou ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). La plupart des systèmes modernes tels que Java 7 et les versions ultérieures prennent en charge ces modes.

Sécurité du réseau

Lors des AWS Clean Rooms lectures depuis votre compartiment S3 pendant l'exécution d'une requête, le trafic entre Amazon S3 AWS Clean Rooms et Amazon S3 est acheminé de manière sécurisée via le réseau AWS privé. Le trafic en vol est signé à l'aide du protocole Amazon Signature Version 4 (SIGv4) et crypté à l'aide du protocole HTTPS. Ce trafic est autorisé en fonction du rôle de service IAM que vous avez défini pour votre table configurée.

Vous pouvez vous connecter par programmation AWS Clean Rooms via un point de terminaison. Pour obtenir la liste des points de terminaison de service, consultez la section [AWS Clean Rooms Points de terminaison et quotas](#) dans le. Références générales AWS

Tous les points de terminaison de service fonctionnent uniquement en HTTPS. Vous pouvez utiliser les points de terminaison Amazon Virtual Private Cloud (VPC) au cas où vous souhaiteriez vous connecter à partir de AWS Clean Rooms votre VPC sans avoir de connexion Internet. Pour plus d'informations, consultez la section [Accès aux AWS services AWS PrivateLink](#) dans le AWS PrivateLink Guide.

Vous pouvez attribuer des politiques IAM à vos principaux IAM en utilisant les [clés de SourceVpce contexte aws](#) : pour empêcher votre principal IAM de ne pouvoir passer des appels que via AWS Clean Rooms un point de terminaison VPC et non via Internet.

Accès AWS Clean Rooms ou AWS Clean Rooms ML à l'aide d'un point de terminaison d'interface (AWS PrivateLink)

Vous pouvez l'utiliser AWS PrivateLink pour créer une connexion privée entre votre cloud privé virtuel (VPC) AWS Clean Rooms et/ou AWS Clean Rooms ML. Vous pouvez accéder au AWS Clean Rooms AWS Clean Rooms ML comme s'il se trouvait dans votre VPC, sans utiliser de passerelle Internet, de périphérique NAT, de connexion VPN ou Direct Connect de connexion. Les instances de votre VPC n'ont pas besoin d'adresses IP publiques pour y accéder. AWS Clean Rooms

Vous établissez cette connexion privée en créant un point de terminaison d'interface optimisé par AWS PrivateLink. Nous créons une interface réseau de point de terminaison dans chaque sous-réseau que vous activez pour le point de terminaison d'interface. Il s'agit d'interfaces réseau gérées par le demandeur qui servent de point d'entrée pour le trafic destiné à AWS Clean Rooms.

Pour plus d'informations, consultez [Accès aux Services AWS via AWS PrivateLink](#) dans le Guide AWS PrivateLink .

Considérations relatives à AWS Clean Rooms

Avant de configurer un point de terminaison d'interface pour AWS Clean Rooms, consultez les [considérations](#) du AWS PrivateLink guide.

AWS Clean Rooms et le AWS Clean Rooms ML prennent en charge l'envoi d'appels à toutes leurs actions d'API via le point de terminaison de l'interface.

Les politiques de point de terminaison VPC ne sont pas prises en charge pour le ML AWS Clean Rooms . AWS Clean Rooms Par défaut, l'accès complet à AWS Clean Rooms et AWS Clean Rooms ML est autorisé via le point de terminaison de l'interface. Vous pouvez également associer un groupe de sécurité aux interfaces réseau du point de terminaison afin de contrôler le trafic vers AWS Clean Rooms ou le AWS Clean Rooms ML via le point de terminaison de l'interface.

Créez un point de terminaison d'interface pour AWS Clean Rooms

Vous pouvez créer un point de terminaison d'interface pour AWS Clean Rooms ou AWS Clean Rooms ML 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 AWS PrivateLink .

Créez un point de terminaison d'interface pour AWS Clean Rooms utiliser le nom de service suivant.

```
com.amazonaws.region.cleanrooms
```

Créez un point de terminaison d'interface pour AWS Clean Rooms ML en utilisant le nom de service suivant.

```
com.amazonaws.region.cleanrooms-ml
```

Si vous activez le DNS privé pour le point de terminaison de l'interface, vous pouvez envoyer des demandes d'API AWS Clean Rooms en utilisant son nom DNS régional par défaut. Par exemple, `cleanrooms-ml.us-east-1.amazonaws.com`.

Surveillance AWS Clean Rooms

La surveillance joue un rôle important dans le maintien de la fiabilité, de la disponibilité AWS Clean Rooms et des performances de vos autres AWS solutions. AWS fournit les outils de surveillance suivants pour surveiller AWS Clean Rooms, signaler tout problème et prendre des mesures automatiques le cas échéant :

- Amazon CloudWatch Logs vous permet de surveiller, de stocker et d'accéder à vos fichiers journaux à partir d'instances Amazon EC2 et d'autres sources. AWS CloudTrail Amazon CloudWatch Logs peut surveiller les informations contenues dans les fichiers journaux et vous avertir lorsque certains seuils sont atteints. Vous pouvez également archiver vos données de journaux dans une solution de stockage hautement durable. Pour plus d'informations, consultez le [guide de l'utilisateur Amazon CloudWatch Logs](#).

Clean Rooms ML autorise les tâches entre comptes pour certaines actions d'API. La personne Compte AWS qui a démarré la tâche reçoit l'événement du journal AWS CloudTrail d'audit correspondant à la tâche. Pour de plus amples informations, consultez [Comportements IAM pour AWS Clean Rooms ML](#).

- AWS CloudTrail capture les appels d'API et les événements associés effectués par vous ou en votre nom Compte AWS et envoie les fichiers journaux dans un compartiment Amazon S3 que vous spécifiez. Vous pouvez identifier les utilisateurs et les comptes appelés AWS, l'adresse IP source à partir de laquelle les appels ont été effectués et la date des appels. Pour plus d'informations, consultez le [Guide de l'utilisateur AWS CloudTrail](#).
- Amazon EventBridge est un service de bus d'événements sans serveur qui permet de connecter facilement vos applications à des données provenant de diverses sources. EventBridge fournit un flux de données en temps réel à partir de vos propres applications, applications Software-as-a-Service (SaaS) et AWS services et achemine ces données vers des cibles telles que Lambda. Cela vous permet de surveiller les événements qui se produisent dans les services et de créer des architectures basées sur les événements. Pour plus d'informations, consultez le [guide de EventBridge l'utilisateur Amazon](#) et le manuel [Amazon EventBridge Events Reference](#).
- AWS Billing and Cost Management est un service qui fournit des fonctionnalités pour vous aider à payer vos factures et à optimiser vos coûts. Amazon Web Services facture votre compte en fonction de l'utilisation, ce qui garantit que vous ne payez que pour ce que vous utilisez. AWS vous permet également d'appliquer [des balises de répartition des coûts](#) à vos ressources afin de suivre et de classer vos coûts AWS. Par exemple, dans AWS Clean Rooms, vous pouvez appliquer des

balises aux adhésions aux collaborations pour suivre les coûts par collaboration. Pour de plus amples informations, veuillez consulter le [Guide de l'utilisateur AWS Billing](#).

Connexion à une analyse AWS Clean Rooms

La journalisation des analyses est une fonctionnalité de AWS Clean Rooms. Lorsque vous [créez une collaboration](#) et que vous activez la journalisation des analyses, les membres peuvent stocker les journaux pertinents issus de requêtes ou les journaux de tâches dans Amazon CloudWatch Logs.

Grâce aux journaux des requêtes et aux journaux des tâches, les membres peuvent déterminer si les requêtes sont conformes aux règles d'analyse et à l'accord de collaboration. En outre, les journaux de requêtes facilitent les audits.

Lorsque l'option de journalisation de l'analyse est activée dans la AWS Clean Rooms console, les journaux de requêtes incluent les éléments suivants :

- `analysisRule`— Règle d'analyse pour la table configurée.
- `analysisTemplateArn`— Le modèle d'analyse qui a été exécuté (apparaît en fonction de la règle d'analyse).
- `collaborationId`— Identifiant unique pour la collaboration dans laquelle la requête a été exécutée.
- `configuredTableID`— Identifiant unique de la table configurée référencée dans la requête.
- `directQueryAnalysisRulePolicy.custom.allowedAnalysis`— Le modèle d'analyse autorisé à s'exécuter sur une table configurée (apparaît en fonction de la règle d'analyse).
- `directQueryAnalysisRulePolicy.v1.custom.allowedAnalysisProviders`— Les fournisseurs de requêtes autorisés à créer des requêtes (apparaissent en fonction de la règle d'analyse).
- `errorCode`— Code d'erreur lorsqu'une requête ne s'exécute pas correctement.
- `errorMessage`— Le message d'erreur lorsqu'une requête ne s'est pas exécutée correctement.
- `eventID`— Identifiant unique de la requête exécutée. Après le 31 août 2023, l'identifiant unique est le même que `protectedQueryID`.
- `eventTimestamp`— Durée d'exécution de la requête.
- `parameters.parametervalue`— Les valeurs des paramètres (apparaissent en fonction du texte de la requête).

- `queryText`— Définition SQL de la requête exécutée. S'il existe des paramètres, ils sont étiquetés comme `:parametervalue`.
- `queryValidationErrors`— Les erreurs de requête lors de la validation de la requête.
- `schemaName`— Nom de l'association de tables configurée référencée dans la requête.
- `status`— État d'exécution de la requête.

Recevoir les journaux des requêtes et des tâches

Vous n'avez pas besoin d'effectuer d'actions en dehors de la AWS Clean Rooms configuration des journaux de requêtes et des journaux de tâches. AWS Clean Rooms crée des groupes de journaux pour les collaborations une fois que chaque membre de la collaboration [a créé une adhésion](#).

Les membres autorisés à effectuer des requêtes, à exécuter des requêtes et à exécuter des tâches, à recevoir des résultats et aux membres dont les tables de configuration sont référencées dans la requête recevront un journal des requêtes ou un journal des tâches.

Le membre autorisé à effectuer une requête et le membre autorisé à recevoir des résultats recevront des journaux de requêtes pour chaque table configurée référencée dans la requête. S'ils ne possèdent pas la table configurée, ils ne pourront pas voir l'ID de table configuré (`configuredTableID`).

Le membre qui peut exécuter des requêtes et des tâches et le membre qui peut recevoir des résultats recevront des journaux de tâches pour chaque table configurée référencée dans la tâche. S'ils ne possèdent pas la table configurée, ils ne pourront pas voir l'ID de table configuré (`configuredTableID`).

Si un membre possède plusieurs associations de tables configurées référencées dans la requête, il recevra un journal des requêtes pour chaque table configurée.

Si un membre possède plusieurs associations de tables configurées référencées dans la tâche, il recevra un journal des tâches pour chaque table configurée.

Des journaux sont créés pour les requêtes contenant du code SQL non pris en charge et pris en charge dans AWS Clean Rooms. Pour plus de détails, consultez la [référence AWS Clean Rooms SQL](#).

Des journaux sont également créés lorsque des requêtes ou des tâches font référence à des tables configurées qui ne sont pas associées à la collaboration.

Les journaux peuvent contenir des informations relatives à un code SQL incorrect.

Les journaux des requêtes et des tâches indiquent le statut d'une requête mais n'indiquent pas si le résultat de la requête a été fourni. Ils confirment qu'une requête ou un travail a été soumis par le membre habilité à effectuer la requête. Les journaux de requêtes confirment également que la requête contient du code SQL pris en charge AWS Clean Rooms et fait référence aux tables configurées associées à la collaboration.

Exemple

Par exemple, aucun journal n'est produit si la requête a été annulée après avoir AWS Clean Rooms validé sa conformité aux règles d'analyse et pendant le traitement de la requête.

Si vous supprimez le groupe de journaux, vous devez le recréer manuellement avec le même nom de groupe de journaux (ID de collaboration de la collaboration). Vous pouvez également désactiver et activer la déconnexion dans votre abonnement.

Pour plus d'informations sur la façon d'activer la journalisation des analyses, consultez [Création d'une collaboration](#).

Pour plus d'informations sur Amazon CloudWatch Logs, consultez le [guide de l'utilisateur Amazon CloudWatch Logs](#).

Actions recommandées pour les journaux de requêtes et de tâches

Nous recommandons aux membres de prendre régulièrement les mesures suivantes :

- Pour vérifier que les requêtes et les tâches correspondent aux cas d'utilisation ou aux requêtes convenus pour la collaboration, passez en revue les requêtes et les tâches exécutées dans le cadre de la collaboration.

Pour plus d'informations sur l'affichage des requêtes récentes, consultez la section [Affichage des requêtes récentes](#).

Pour plus d'informations sur la façon de consulter les offres d'emploi récentes, consultez [Afficher les offres d'emploi récentes](#).

- Pour vérifier que les colonnes de table configurées correspondent à ce qui a été convenu pour la collaboration, passez en revue les colonnes de table configurées qui sont utilisées dans les règles d'analyse des membres de la collaboration et dans les requêtes.

Pour plus d'informations sur la façon d'afficher les colonnes configurées, consultez la section [Affichage des tables et des règles d'analyse](#).

Surveillance détaillée avec CloudWatch AWS Clean Rooms

La surveillance détaillée est une fonctionnalité facultative. AWS Clean Rooms Lorsque vous [créez une collaboration](#) et que vous activez le suivi détaillé, les membres qui peuvent exécuter des requêtes et configurer les payeurs peuvent choisir de recevoir des mesures d'observabilité détaillées sur leur CloudWatch compte.

Grâce à une surveillance détaillée, les membres peuvent surveiller les performances des requêtes, suivre l'utilisation des ressources et créer des tableaux de bord pour obtenir des informations opérationnelles. Ces indicateurs aident à planifier les capacités, à analyser les coûts et à résoudre les problèmes de performance.

Note

La disponibilité des mesures varie en fonction des sources de données de table configurées :

- Pour les analyses qui font référence à des tables configurées avec Athena comme source de données, les QueryFileScanned métriques BytesRead and renvoient toujours 0.
- Pour les analyses qui font référence à des tables configurées avec Snowflake comme source de données, les RecordsRead métriques BytesReadQueryFileScanned, et renvoient toujours 0.
- Pour les analyses qui font référence à plusieurs tables configurées avec différentes sources de données (par exemple, Snowflake et les tables basées sur S3) et qui QueryFileScanned renvoient des métriques uniquement pour les AWS Glue tables basées sur S3. BytesRead AWS Glue RecordsReadrenvoie des métriques pour les tables basées sur S3 et AWS Glue les tables Athena.

Types de métriques

AWS Clean Roomspublie deux catégories de métriques pour CloudWatch :

Mesures de surveillance détaillées agrégées

Ces indicateurs fournissent une visibilité globale sur l'ensemble des analyses :

AnalysisRuntime

Indique la durée totale d'une analyse en millisecondes.

CRPUConsumed

Les unités CRPU (Clean Room Processing Units) utilisées pour une analyse particulière. Pour plus d'informations sur le calcul du CRPU, consultez la section [AWS Clean Rooms Tarification](#).

ConcurrentVCPUs

vCPU utilisés simultanément à un moment donné (en fonction des quotas de service).

AnalysesCount

Nombre d'analyses effectuées sur une période donnée pour une dimension spécifiée.

ConcurrentQueries

Nombre de requêtes simultanées à un moment donné (lié aux quotas de service).

Mesures de surveillance détaillées au niveau des requêtes

Ces indicateurs fournissent des informations détaillées sur l'exécution de requêtes individuelles :

BytesRead

Nombre total de données (en octets) lues à partir des tables sources pour une analyse.

BytesWritten

Total des données (en octets) écrites pour une analyse.

RecordsRead

Nombre total d'enregistrements lus à partir des tables sources pour une analyse.

RecordsWritten

Nombre total d'enregistrements renvoyés dans le résultat de l'analyse.

QueryFileScanned

Nombre total de fichiers de données scannés pour une analyse.

Memory utilization

Ressource mémoire utilisée pour une analyse.

vCPU utilization

Ressource vCPU utilisée pour une analyse.

Disk utilization

Ressource de stockage utilisée pour une analyse.

Dimensions des métriques

Les métriques sont organisées selon les dimensions suivantes :

- État de l'analyse : terminée
- Type d'analyse : SparkSQL
- Adhésion : l'identifiant de membre
- Requête : ID de requête individuel (pour les métriques au niveau de la requête)
- Compte/région : votre région et votre Compte AWS région

Toutes les métriques sont publiées dans l'espace de AWS/CleanRooms noms dans CloudWatch.

Qui peut accéder aux métriques

Des mesures de surveillance détaillées sont disponibles pour :

- Exécuteurs de requêtes : membres autorisés à exécuter des requêtes dans le cadre de la collaboration
- Payeurs configurés : membres désignés pour payer les coûts d'exécution des requêtes

Les membres qui fournissent uniquement des données ne peuvent pas accéder aux mesures de surveillance détaillées.

Tarification

AWS Clean Rooms ne facture pas la publication de statistiques sur CloudWatch. Toutefois, des CloudWatch frais standard s'appliquent pour le stockage des métriques, l'utilisation du tableau de bord et les alarmes. Pour plus d'informations, consultez [Amazon CloudWatch Pricing](#).

Journalisation des appels AWS Clean Rooms d'API à l'aide AWS CloudTrail

AWS Clean Rooms est intégré à AWS CloudTrail un service qui fournit un enregistrement des actions entreprises par un utilisateur, un rôle ou un Service AWS utilisateur AWS Clean Rooms. CloudTrail capture tous les appels d'API AWS Clean Rooms sous forme d'événements. Les appels capturés incluent des appels provenant de la AWS Clean Rooms console et des appels de code vers les opérations de l' AWS Clean Rooms API. Si vous créez un suivi, vous pouvez activer la diffusion continue d' CloudTrail événements vers un compartiment Amazon S3, y compris les événements pour AWS Clean Rooms. 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 AWS Clean Rooms, 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](#).

AWS Clean Rooms informations dans CloudTrail

CloudTrail est activé sur votre compte Compte AWS lorsque vous créez le compte. Lorsqu'une activité se produit dans AWS Clean Rooms, cette activité est enregistrée dans un CloudTrail événement avec d'autres Service AWS événements dans l'historique des événements. Vous pouvez afficher, rechercher et télécharger les événements récents dans votre Compte AWS. Pour plus d'informations, consultez la section [Affichage des événements à l'aide de l'historique des CloudTrail événements](#).

Pour un enregistrement continu des événements de votre région Compte AWS, y compris des événements pour AWS Clean Rooms, créez un parcours. Un suivi permet CloudTrail de fournir des fichiers journaux à un compartiment Amazon S3. Par défaut, lorsque vous créez un journal d'activité dans la console, il s'applique à toutes les régions 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 en configurer d'autres Services AWS pour analyser plus en détail les données d'événements collectées dans les CloudTrail journaux et agir en conséquence. 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 provenant de plusieurs régions](#)
- [Réception de fichiers CloudTrail journaux provenant de plusieurs comptes](#)

Toutes les AWS Clean Rooms actions sont enregistrées CloudTrail et documentées dans la [référence de l'AWS Clean Rooms API](#).

Chaque événement ou entrée de journal contient des informations sur la personne ayant initié la demande. Les informations relatives à l'identité permettent de déterminer :

- Si la demande a été effectuée avec des informations d'identification d'utilisateur root ou d'utilisateur 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 requête a été effectuée par un autre Service AWS.

Pour de plus amples informations, veuillez consulter l'[élément userIdentity CloudTrail](#).

Comprendre les entrées du fichier AWS Clean Rooms journal

Un suivi est une configuration qui permet de transmettre des événements sous forme de fichiers journaux à un compartiment Amazon S3 que vous spécifiez. CloudTrail les fichiers journaux contiennent une ou plusieurs entrées de journal. Un événement représente une demande unique provenant de n'importe quelle source et inclut des informations sur l'action demandée, la date et l'heure de l'action, les paramètres de la demande, etc. CloudTrail les fichiers journaux ne constituent pas une trace ordonnée des appels d'API publics, ils n'apparaissent donc pas dans un ordre spécifique.

Exemples d' AWS Clean Rooms CloudTrail événements

Les exemples suivants illustrent CloudTrail des événements pour :

Rubriques

- [StartProtectedQuery \(réussite\)](#)
- [StartProtectedQuery\(échec\)](#)

StartProtectedQuery (réussite)

```
{
  "eventVersion": "1.08",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "EXAMPLE_PRINCIPAL_ID",
    "arn": "arn:aws:sts::123456789012:assumed-role/query-runner/jdoe",
    "accountId": "123456789012",
    "accessKeyId": "EXAMPLE_KEY_ID",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "EXAMPLE_PRINCIPAL_ID",
        "arn": "arn:aws:iam::123456789012:role/query-runner",
        "accountId": "123456789012",
        "userName": "query-runner"
      },
      "webIdFederationData": {},
      "attributes": {
        "creationDate": "2023-04-07T19:34:32Z",
        "mfaAuthenticated": "false"
      }
    }
  },
  "eventTime": "2023-04-07T19:53:32Z",
  "eventSource": "cleanrooms.amazonaws.com",
  "eventName": "StartProtectedQuery",
  "awsRegion": "us-east-2",
  "sourceIPAddress": "203.0.113.1",
  "userAgent": "aws-internal/3",
  "requestParameters": {
    "resultConfiguration": {
      "outputConfiguration": {
        "s3": {
          "resultFormat": "CSV",
          "bucket": "cleanrooms-queryresults-jdoe-test",
          "keyPrefix": "test"
        }
      }
    }
  }
}
```

```

        }
      },
      "sqlParameters": "****",
      "membershipIdentifier": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
      "type": "SQL"
    },
    "responseElements": {
      "Access-Control-Expose-Headers": "x-amzn-RequestId,x-amzn-ErrorType,x-amzn-ErrorMessage,Date",
      "protectedQuery": {
        "createTime": 1680897212.279,
        "id": "f5988bf1-771a-4141-82a8-26fcc4e41c9f",
        "membershipArn": "arn:aws:cleanrooms:us-east-2:123456789012:membership/a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
        "membershipId": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
        "resultConfiguration": {
          "outputConfiguration": {
            "s3": {
              "bucket": "cleanrooms-queryresults-jdoe-test",
              "keyPrefix": "test",
              "resultFormat": "CSV"
            }
          }
        }
      },
      "sqlParameters": "****",
      "status": "SUBMITTED"
    }
  },
  "requestID": "7464211b-2277-4b55-9723-fb4f259aefd2",
  "eventID": "f7610f5e-74b9-420f-ae43-206571ebcbf7",
  "readOnly": false,
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "recipientAccountId": "123456789012",
  "eventCategory": "Management"
}

```

StartProtectedQuery(échec)

```

{
  "eventVersion": "1.08",
  "userIdentity": {

```

```

    "type": "AssumedRole",
    "principalId": "EXAMPLE_PRINCIPAL_ID",
    "arn": "arn:aws:sts::123456789012:assumed-role/query-runner/jdoe",
    "accountId": "123456789012",
    "accessKeyId": "EXAMPLE_KEY_ID",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "EXAMPLE_PRINCIPAL_ID",
        "arn": "arn:aws:iam::123456789012:role/query-runner",
        "accountId": "123456789012",
        "userName": "query-runner"
      },
      "webIdFederationData": {},
      "attributes": {
        "creationDate": "2023-04-07T19:34:32Z",
        "mfaAuthenticated": "false"
      }
    },
    "eventTime": "2023-04-07T19:47:27Z",
    "eventSource": "cleanrooms.amazonaws.com",
    "eventName": "StartProtectedQuery",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "203.0.113.1",
    "userAgent": "aws-internal/3",
    "errorCode": "ValidationException",
    "requestParameters": {
      "resultConfiguration": {
        "outputConfiguration": {
          "s3": {
            "resultFormat": "CSV",
            "bucket": "cleanrooms-queryresults-jdoe-test",
            "keyPrefix": "test"
          }
        }
      },
      "sqlParameters": "****",
      "membershipIdentifier": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
      "type": "SQL"
    },
    "responseElements": {
      "Access-Control-Expose-Headers": "x-amzn-RequestId,x-amzn-ErrorType,x-amzn-ErrorMessage,Date",

```

```
    "message": "Column(s) [identifiant] is not allowed in select"
  },
  "requestID": "e29f9f74-8299-4a83-9d18-5ddce7302f07",
  "eventID": "c8ee3498-8e4e-44b5-87e4-ab9477e56eb5",
  "readOnly": false,
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "recipientAccountId": "123456789012",
  "eventCategory": "Management"
}
```

Intégration AWS Clean Rooms dans des applications axées sur les événements à l'aide d'Amazon EventBridge

Vous pouvez les AWS Clean Rooms intégrer à des applications pilotées par des événements (EDAs) qui utilisent des événements qui se produisent AWS Clean Rooms pour communiquer entre les composants de l'application et lancer des processus en aval. Pour ce faire, utilisez Amazon EventBridge pour acheminer les événements depuis AWS Clean Rooms d'autres composants logiciels. Amazon EventBridge est un service sans serveur qui utilise des événements pour connecter les composants de l'application entre eux, ce qui vous permet d'intégrer plus facilement des AWS services, par exemple AWS Clean Rooms dans des architectures pilotées par des événements, sans code ni opérations supplémentaires.

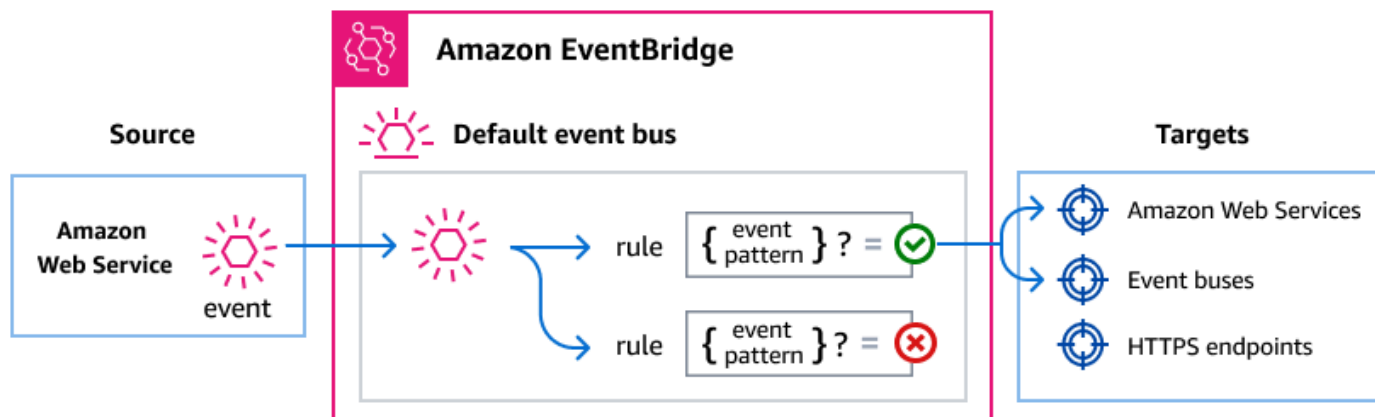
Une architecture pilotée par les événements est un style de création de systèmes logiciels faiblement couplés qui fonctionnent ensemble en émettant des événements et en y répondant. Dans ce modèle, un événement représente une modification d'une ressource ou d'un environnement.

Voici comment EventBridge fonctionne avec AWS Clean Rooms :

Comme c'est le cas pour de nombreux AWS services, AWS Clean Rooms génère et envoie des événements au bus d'événements EventBridge par défaut. Un bus d'événements est un routeur qui reçoit des événements et les achemine vers les destinations, ou cibles, que vous spécifiez. Les cibles peuvent inclure d'autres AWS services, des applications personnalisées et des applications partenaires SaaS.

EventBridge achemine les événements conformément aux règles que vous créez sur le bus d'événements. Pour chaque règle, vous spécifiez un filtre, ou un modèle d'événement, afin de sélectionner uniquement les événements souhaités. Chaque fois qu'un événement est envoyé au

bus d'événements EventBridge, comparez-le à chaque règle. Si l'événement correspond à la règle, EventBridge achemine l'événement vers la ou les cibles spécifiées.



Supposons, par exemple, que vous souhaitiez être informé à chaque fois qu'une nouvelle AWS Clean Rooms collaboration est créée dans votre compte. Vous pouvez créer une règle sur le bus d'événements par défaut. Dans la règle, vous créeriez un modèle d'événement nommant les événements provenant d'AWS Clean Rooms **Collaboration Created**. Chaque fois que EventBridge reçoit un événement correspondant à ces propriétés, il achemine l'événement vers le flux de travail spécifié.

AWS Clean Rooms événements

AWS les services peuvent envoyer des événements directement au bus d'événements EventBridge par défaut. En outre, AWS CloudTrail envoie également des événements provenant de nombreux AWS services à EventBridge. Ces événements peuvent inclure des appels d'API, des connexions et des actions à la console, des événements de service et CloudTrail Insights. Pour plus d'informations, consultez les [événements AWS de service fournis via AWS CloudTrail](#) dans le guide de EventBridge l'utilisateur.

Pour obtenir la liste complète des AWS Clean Rooms événements envoyés à EventBridge, reportez-vous à la AWS Clean Rooms rubrique correspondante dans la [référence EventBridge des événements](#).

Type de détail d'événement	Description
<u>Modèle d'analyse créé</u>	Le propriétaire du modèle d'analyse et tous les membres actifs de la collaboration sont avertis lorsqu'un modèle d'analyse est créé.
<u>Modèle d'analyse mis à jour</u>	Le propriétaire du modèle d'analyse et tous les membres actifs de la collaboration ayant une visibilité sur la mise à jour sont avertis lorsqu'un modèle d'analyse est mis à jour.
<u>Modèle d'analyse supprimé</u>	Le propriétaire du modèle d'analyse et tous les membres actifs de la collaboration sont avertis lorsqu'un modèle d'analyse est supprimé.
<u>Collaboration créée</u>	Le propriétaire de la collaboration est averti lorsqu'une collaboration est créée.
<u>Collaboration mise à jour</u>	Le propriétaire de la collaboration et tous les membres actifs de la collaboration ayant accès à la mise à jour sont avertis lorsqu'une collaboration est mise à jour.
<u>Demande de modification de collaboration créée</u>	Le propriétaire de la collaboration et tous les membres actifs de la collaboration sont avertis lorsqu'une demande de modification de collaboration est créée.
<u>Demande de modification de collaboration approuvée</u>	Le propriétaire de la collaboration et tous les membres actifs de la collaboration sont avertis lorsqu'une demande de modification de collaboration est approuvée.
<u>Demande de modification de collaboration annulée</u>	Le propriétaire de la collaboration et tous les membres actifs de la collaboration sont avertis lorsqu'une demande de modification de collaboration est annulée.
<u>Demande de modification de la collaboration validée</u>	Le propriétaire de la collaboration et tous les membres actifs de la collaboration sont avertis lorsqu'une demande de modification de collaboration est validée.

Type de détail d'événement	Description
<u>Association de tables configurée créée</u>	Le propriétaire de l'association de tables configurée et tous les membres actifs de la collaboration sont avertis lorsqu'une association de tables configurée est créée.
<u>Association de tables configurée mise à jour</u>	Le propriétaire de l'association de tables configurée et tous les membres actifs de la collaboration ayant une visibilité sur la mise à jour sont avertis lorsqu'une association de tables configurée est mise à jour.
<u>Association de tables configurée supprimée</u>	Le propriétaire de l'association de tables configurée et tous les membres actifs de la collaboration sont avertis lorsqu'une association de tables configurée est supprimée.
<u>Règle d'analyse des associations de tables configurée créée</u>	Le propriétaire de la règle d'analyse des associations de tables configurée et tous les membres actifs de la collaboration sont avertis lorsqu'une règle d'analyse des associations de tables configurée est créée.
<u>Règle d'analyse des associations de tables configurée mise à jour</u>	Le propriétaire de la règle d'analyse des associations de tables configurée et tous les membres actifs de la collaboration ayant une visibilité sur la mise à jour sont avertis lorsqu'une règle d'analyse des associations de tables configurée est mise à jour.
<u>Règle d'analyse des associations de tables configurée supprimée</u>	Le propriétaire de la règle d'analyse des associations de tables configurée et tous les membres actifs de la collaboration sont avertis lorsqu'une règle d'analyse des associations de tables configurée est supprimée.
<u>Table de mappage d'identifiants créée</u>	Le propriétaire de la table de mappage d'identifiants et tous les membres actifs de la collaboration sont avertis lorsqu'une table de mappage d'identifiants est créée.
<u>Table de mappage des identifiants mise à jour</u>	Le propriétaire de la table de mappage d'identifiants et tous les membres actifs de la collaboration ayant une visibilité sur la mise à jour sont avertis lorsqu'une table de mappage d'identifiants est mise à jour.

Type de détail d'événement	Description
Table de mappage d'identifiants supprimée	Le propriétaire de la table de mappage d'identifiants et tous les membres actifs de la collaboration sont avertis lorsqu'une table de mappage d'identifiants est supprimée.
ID : association d'espace de noms créée	Le propriétaire de l'association Id Namespace et tous les membres actifs de la collaboration sont avertis lorsqu'une association d'espace de noms Id est créée.
Association d'espace de noms Id mise à jour	Le propriétaire de l'association Id Namespace et tous les membres actifs de la collaboration ayant accès à la mise à jour sont avertis lorsqu'une association d'espace de noms Id est mise à jour.
ID : association d'espace de noms supprimée	Le propriétaire de l'association Id Namespace et tous les membres actifs de la collaboration sont avertis lorsqu'une association d'espace de noms Id est supprimée.
Invité à collaborer	Le membre invité est averti lorsqu'il est invité à participer à une collaboration.
Adhésion créée	Le propriétaire de l'adhésion et tous les membres actifs de la collaboration sont avertis lorsqu'une adhésion est créée.
Adhésion mise à jour	Le propriétaire de l'adhésion est averti lorsqu'une adhésion est mise à jour, sauf si l'adhésion a été supprimée de la collaboration, auquel cas tous les membres actifs de la collaboration sont avertis.
Abonnement supprimé	Le propriétaire de l'adhésion et tous les membres actifs de la collaboration sont avertis lorsqu'une adhésion est supprimée.
Job protégé soumis	Le Job Runner, le Job Payer et le récepteur des résultats d'un Job Protégé sont avertis lorsque le Job Protégé est soumis.
Job protégé commencé	Le Job Runner, le Job Payer et le Récepteur des résultats d'un Job Protégé sont avertis lorsque le Job Protégé démarre.

Type de détail d'événement	Description
<u>Annulation d'un emploi protégée</u>	Le Job Runner, le Job Payer et le Récepteur des résultats d'un Job Protégé sont avertis lorsque le Job Protégé est annulé.
<u>Job protégé annulé</u>	Le Job Runner, le Job Payer et le Récepteur des résultats d'un Job Protégé sont avertis lorsque le Job Protégé est annulé.
<u>Job protégé réussi</u>	Le Job Runner, le Job Payer et le Récepteur des résultats d'un Job Protégé sont avertis lorsque le Job Protégé est réussi.
<u>Echec du Job protégé</u>	Le Job Runner, le Job Payer et le récepteur des résultats d'un Job Protégé sont avertis lorsque le Job Protégé a échoué.
<u>Requête protégée soumise</u>	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque la requête protégée est soumise.
<u>Requête protégée démarrée</u>	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque la requête protégée est lancée.
<u>Annulation d'une requête protégée</u>	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque la requête protégée est annulée.
<u>Requête protégée annulée</u>	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque la requête protégée est annulée.
<u>Requête protégée réussie</u>	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque la requête protégée est réussie.
<u>Échec de la requête protégée</u>	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque la requête protégée a échoué.

Type de détail d'événement	Description
Expération du délai imparti pour la requête protégée	Le lanceur de requêtes, le payeur de requêtes et le ou les destinataires des résultats d'une requête protégée sont avertis lorsque le délai imparti pour la requête protégée est expiré.

Routage AWS Clean Rooms des événements à l'aide EventBridge

Pour EventBridge acheminer AWS Clean Rooms les événements vers une cible, vous devez créer une règle. Chaque règle contient un modèle d'événement, qui EventBridge correspond à chaque événement reçu sur le bus d'événements. Si les données d'événement correspondent au modèle d'événement spécifié, EventBridge achemine cet événement vers les cibles de la règle.

Pour obtenir des instructions complètes sur la création de règles de bus d'événements, voir [Création de règles réagissant aux événements](#) dans le Guide de EventBridge l'utilisateur.

Création de modèles d'événements qui correspondent aux AWS Clean Rooms événements

Chaque modèle d'événement est un objet JSON qui contient :

- (Facultatif) : `source` attribut identifiant le service qui envoie l'événement. Pour les AWS Clean Rooms événements, la source est `aws.cleanrooms`.
- (Facultatif) : `detail-type` attribut qui contient un tableau des noms d'événements à associer.
- (Facultatif) : un attribut `detail` qui contient toute autre donnée d'événement à rechercher.

Par exemple, le modèle d'événement suivant correspond à tous les événements `Id Namespace Association Updated` dont la collaboration a été supprimée AWS Clean Rooms :

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"],
  "detail": {
    "status": ["COLLABORATION_DELETED"]
  }
}
```

Pour plus d'informations sur la rédaction de modèles d'événements, consultez la section [Modèles d'événements](#) dans le guide de EventBridge l'utilisateur.

AWS Clean Rooms référence détaillée des événements

Tous les événements issus AWS des services ont un ensemble commun de champs contenant des métadonnées relatives à l'événement, telles que le AWS service à l'origine de l'événement, l'heure à laquelle l'événement a été généré, le compte et la région dans lesquels l'événement a eu lieu, etc. Pour les définitions de ces champs généraux, consultez la section [Structure des événements](#) dans le Amazon EventBridge Events Reference.

En outre, chaque événement possède un champ `detail` qui contient des données spécifiques à cet événement en particulier. La référence ci-dessous définit les champs de détail des différents AWS Clean Rooms événements.

Lorsque vous l'utilisez EventBridge pour sélectionner et gérer AWS Clean Rooms des événements, il est utile de garder à l'esprit les points suivants :

- Le source champ pour tous les événements de AWS Clean Rooms est défini `suraws.cleanrooms`.
- Le champ `detail-type` indique le type d'événement.

Par exemple, `Collaboration Created`.

- Le champ `detail` contient les données spécifiques à cet événement en particulier.

Pour plus d'informations sur la création de modèles d'événements permettant aux règles de correspondre aux AWS Clean Rooms événements, consultez la section [Modèles d'événements](#) dans le guide de EventBridge l'utilisateur Amazon.

Modèle d'analyse | Événement créé

Vous trouverez ci-dessous les champs détaillés de l'`Analysis Template Created` événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Created"]
}
```

Événement mis à jour du modèle d'analyse

Vous trouverez ci-dessous les champs détaillés de l'Analysis Template Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Updated"]
}
```

Modèle d'analyse : événement supprimé

Vous trouverez ci-dessous les champs détaillés de l'Analysis Template Deleted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Deleted"]
}
```

Événement créé en collaboration

Vous trouverez ci-dessous les champs détaillés de l'Collaboration Created événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Created"]
}
```

Collaboration : événement mis à jour

Vous trouverez ci-dessous les champs détaillés de l'Collaboration Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Updated"]
}
```

Demande de modification de collaboration Événement créé

Vous trouverez ci-dessous les champs détaillés de l'Collaboration Change Request Created événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Created"]
}
```

Demande de modification de collaboration : événement approuvé

Vous trouverez ci-dessous les champs détaillés de l'Collaboration Change Request Approved événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Approved"]
}
```

Demande de modification de collaboration Événement annulé

Vous trouverez ci-dessous les champs détaillés de l'Collaboration Change Request Cancelled événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Cancelled"]
}
```

Demande de modification de collaboration Événement validé

Vous trouverez ci-dessous les champs détaillés de l'Collaboration Change Request Committed événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Committed"]
}
```

Association de tables configurée Événement créé

Vous trouverez ci-dessous les champs détaillés de l'Configured Table Association Created événement.


```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Created"]
}
```

Événement mis à jour de l'association de tables configurée

Vous trouverez ci-dessous les champs détaillés de l'Configured Table Association Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Updated"]
}
```

Événement supprimé de l'association de tables configurée

Vous trouverez ci-dessous les champs détaillés de l'Configured Table Association Deleted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Deleted"]
}
```

Règle d'analyse des associations de tables configurée Événement créé

Vous trouverez ci-dessous les champs détaillés de l'Configured Table Association Analysis Rule Created événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Created"]
}
```

Événement mis à jour de la règle d'analyse des associations de tables configurée

Vous trouverez ci-dessous les champs détaillés de l'Configured Table Association Analysis Rule Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Updated"]
}
```

Règle d'analyse des associations de tables configurée Événement supprimé

Vous trouverez ci-dessous les champs détaillés de l'Configured Table Association Analysis Rule Deleted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Deleted"]
}
```

Événement créé par la table de mappage d'identifiants

Vous trouverez ci-dessous les champs détaillés de l'Id Mapping Table Created événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Created"]
}
```

Événement mis à jour de la table de mappage

Vous trouverez ci-dessous les champs détaillés de l'Id Mapping Table Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Updated"]
}
```

Événement supprimé de la table de mappage d'identifiants

Vous trouverez ci-dessous les champs détaillés de l'Id Mapping Table Deleted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Deleted"]
}
```

```
}
```

Événement créé par l'association Id Namespace

Vous trouverez ci-dessous les champs détaillés de l'Id Namespace Association Created événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Created"]
}
```

Événement mis à jour de l'association Id Namespace

Vous trouverez ci-dessous les champs détaillés de l'Id Namespace Association Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"]
}
```

Note

Lorsqu'une collaboration est supprimée, l'événement est capturé dans l'événement Id Namespace Association Updated `detail.status` sur le terrain.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"],
  "detail": {
    "status": ["COLLABORATION_DELETED"]
  }
}
```

Id Namespace Association (événement supprimé)

Vous trouverez ci-dessous les champs détaillés de l'Id Namespace Association Deleted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Deleted"]
}
```

Invité à un événement de collaboration

Vous trouverez ci-dessous les champs détaillés de l'Invited To Collaboration événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Invited To Collaboration"]
}
```

Événement créé par l'adhésion

Vous trouverez ci-dessous les champs détaillés de l'Membership Created événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Created"]
}
```

Événement actualisé pour les membres

Vous trouverez ci-dessous les champs détaillés de l'Membership Updated événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Updated"]
}
```

Événement supprimé de l'adhésion

Vous trouverez ci-dessous les champs détaillés de l'Membership Deleted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Deleted"]
}
```

Événement « Protected Job Submitted »

Vous trouverez ci-dessous les champs détaillés de l'Protected Job Submitted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Submitted"]
}
```

Événement « Protected Job Started »

Vous trouverez ci-dessous les champs détaillés de l'Protected Job Started événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Started"]
}
```

Événement d'annulation d'un emploi protégé

Vous trouverez ci-dessous les champs détaillés de l'Protected Job Cancelling événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Cancelling"]
}
```

Protected Job (événement annulé)

Vous trouverez ci-dessous les champs détaillés de l'Protected Job Cancelled événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Cancelled"]
}
```

Événement Protected Job Succeeded

Vous trouverez ci-dessous les champs détaillés de l'Protected Job Succeeded événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Succeeded"]
}
```

Événement d'échec du projet Protected Job

Vous trouverez ci-dessous les champs détaillés de l'Protected Job Failed événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Failed"]
}
```

Événement soumis à une requête protégée

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Submitted événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Submitted"]
}
```

Événement protégé par requête démarrée

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Started événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Started"]
}
```

Événement d'annulation d'une requête protégée

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Cancelling événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Cancelling"]
}
```

Requête protégée Événement annulé

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Cancelled événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Cancelled"]
}
```

Événement de réussite de la requête protégée

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Succeeded événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Succeeded"]
}
```

Événement d'échec de la requête protégée

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Failed événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Failed"]
}
```

Événement protégé Query Timed Out

Vous trouverez ci-dessous les champs détaillés de l'Protected Query Timed Out événement.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Timed Out"]
}
```

Balisage de répartition des coûts

AWS Clean Rooms prend en charge l'utilisation d'étiquettes de répartition des coûts pour suivre vos AWS coûts. Vous activez ces balises sur le AWS Billing and Cost Management tableau de bord.

AWS utilise les balises pour classer vos coûts et vous fournir un rapport mensuel de répartition des coûts. Des balises de répartition des coûts définies par l'utilisateur peuvent être appliquées aux AWS Clean Rooms ressources pour faciliter le suivi et la répartition des coûts dans l'ensemble de votre collaboration.

Ressources étiquetables pour la répartition des coûts

Le tableau suivant répertorie les AWS Clean Rooms ressources facturées et pouvant être étiquetées avec des balises de répartition des coûts définies par l'utilisateur pour le suivi et l'organisation des coûts.

Ressource facturée	Ressource taguée
Requête SQL	Membres
PySpark Job	Membres
Job de formation ML	Membres
Job d'inférence ML	Membres
Génération de données synthétiques	Membres
Formation sur les modèles Lookalike	Modèle Lookalike
Exportation de segments similaires	Segment similaire

Création de AWS Clean Rooms ressources avec AWS CloudFormation

AWS Clean Rooms est intégré à AWS CloudFormation un service qui vous aide à modéliser et à configurer vos AWS ressources. Grâce à cette intégration, vous pouvez passer moins de temps à créer et à gérer vos ressources et votre infrastructure. Vous créez un modèle qui décrit toutes les AWS ressources que vous souhaitez, et qui CloudFormation fournit et configure ces ressources pour vous. Les exemples de ressources incluent les collaborations, les tables configurées, les associations de tables configurées et les adhésions.

Lorsque vous l'utilisez CloudFormation, vous pouvez réutiliser votre modèle pour configurer vos AWS Clean Rooms ressources de manière cohérente et répétée. Décrivez vos ressources une seule fois, puis fournissez les mêmes ressources à plusieurs reprises Comptes AWS et Régions AWS.

AWS Clean Rooms et CloudFormation modèles

Pour fournir et configurer des ressources AWS Clean Rooms et des services associés, vous devez comprendre les [CloudFormation modèles](#). Les modèles sont des fichiers texte formatés en JSON ou YAML. Ces modèles décrivent les ressources que vous souhaitez mettre à disposition dans vos CloudFormation piles. Si vous n'êtes pas familiarisé avec JSON ou YAML, vous pouvez utiliser CloudFormation Designer pour vous aider à démarrer avec les CloudFormation modèles. Pour plus d'informations, consultez [Qu'est-ce que CloudFormation Designer](#) dans le Guide de l'utilisateur AWS CloudFormation .

AWS Clean Rooms prend en charge la création de collaborations, de tables configurées, d'associations de tables configurées et d'adhésions à CloudFormation. Pour plus d'informations, notamment des exemples de modèles JSON et YAML pour les collaborations, des tables configurées, des associations de tables configurées et des adhésions, consultez les références aux types de ressources et [AWS Clean RoomsAWS Clean Rooms ML](#) dans le guide de l'AWS CloudFormation utilisateur.

En savoir plus sur CloudFormation

Pour en savoir plus CloudFormation, consultez les ressources suivantes :

- [AWS CloudFormation](#)

- [AWS CloudFormation Guide de l'utilisateur](#)
- [CloudFormation API Reference](#)
- [Guide de l'utilisateur de l'interface de ligne de commande AWS CloudFormation](#)

Quotas pour AWS Clean Rooms

Vous Compte AWS disposez de quotas par défaut, anciennement appelés limites, pour chacun d'entre eux Service AWS. Sauf indication contraire, chaque quota est spécifique à un Région AWS. Vous pouvez demander des augmentations pour certains quotas, et d'autres quotas ne peuvent pas être augmentés.

Pour consulter les quotas pour AWS Clean Rooms, ouvrez la [console Service Quotas](#). Dans le panneau de navigation, choisissez AWS services (Services AWS) et sélectionnez AWS Clean Rooms.

Pour demander une augmentation de quota, consultez [Demander une augmentation de quota](#) dans le Guide de l'utilisateur de Service Quotas. Si le quota n'est pas encore disponible dans Quotas de service, utilisez le [formulaire d'augmentation des limites de service](#).

Rubriques

- [AWS Clean Rooms quotas](#)
- [Quotas de machine learning d'AWS Clean Rooms](#)

AWS Clean Rooms quotas

Vous Compte AWS disposez des quotas suivants relatifs à AWS Clean Rooms.

Nom	Par défaut	Ajusté	Description
Taille de la règle d'analyse	Chaque région prise en charge : 100 kilo-octets	Non	Taille maximale du JSON pour une règle d'analyse
Modèles d'analyse par membre	Chaque région prise en charge : 25	Oui	Nombre maximum de modèles d'analyse par membre
Collaborations créées par compte	Par région prise en charge : 10	Oui	Nombre maximum de collaborations créées par compte

Nom	Par défaut	Ajusté	Description
Colonnes par liste d'autorisation de table configurée	Chaque région prise en charge : 100	Oui	Nombre maximum de colonnes pouvant être autorisées par table configurée
Utilisation PySpark simultanée du vCPU par compte	us-east-1 : 1 024 Chacune des autres régions prises en charge : 512	Oui	Utilisation totale maximale du vCPU pour toutes les tâches exécutées PySpark simultanément par compte
PySpark Tâches simultanées par compte	us-east-1 : 5 Chacune des autres régions prises en charge : 2	Oui	Nombre maximum de PySpark tâches exécutées simultanément par compte
Requêtes SQL simultanées par compte	us-east-1 : 5 Chacune des autres régions prises en charge : 2	Oui	Nombre maximum de requêtes SQL exécutées simultanément par compte
Utilisation simultanée du vCPU par compte pour les requêtes SQL	Chaque région prise en charge : 512	Oui	Utilisation totale maximale du vCPU pour toutes les requêtes SQL exécutées simultanément par compte
Travail continu simultané par adhésion	Par région prise en charge : 1	Non	Nombre maximum d'emplois permanents simultanés par adhésion

Nom	Par défaut	Ajusté	Description
Demandes continues simultanées par membre	Chaque région prise en charge : 5	Non	Nombre maximum de requêtes en cours simultanées par membre
Associations de modèles d'audience configurées par membre	Chaque Région prise en charge : 5	Oui	Nombre maximum d'associations de modèles d'audience configurées par membre
Tables configurées par compte	Chaque région prise en charge : 250	Oui	Nombre maximum de tables configurées créées par compte
Tables configurées par requête protégée	Chaque région prise en charge : 15	Oui	Nombre maximal de tables configurées dans une requête protégée
Tables de mappage des identifiants par membre	Chaque Région prise en charge : 5	Oui	Nombre maximum de tables de mappage d'identifiants par membre
Associations d'espaces de noms d'ID par membre	Par région prise en charge : 10	Oui	Nombre maximum d'associations d'espaces de noms d'identification par membre
Membres invités par collaboration	Chaque Région prise en charge : 5	Oui	Nombre maximum de membres invités par collaboration
Abonnements par compte	Chaque région prise en charge : 100	Oui	Nombre maximum d'abonnements par compte

Nom	Par défaut	Ajusté	Description
Modèles de budget de confidentialité pour les budgets d'accès par collaboration	Chaque région prise en charge : 25	Oui	Nombre maximum de modèles de budget de confidentialité pour les budgets d'accès par collaboration
Longueur du texte de la requête (en utilisant la confidentialité différentielle)	Chaque région prise en charge : 8 kilo-octets	Non	Longueur de texte maximale pour une instruction de requête SQL utilisant la confidentialité différentielle
Longueur du texte de requête sur le moteur d'analyse SQL de Clean Rooms	Chaque région prise en charge : 90 kilo-octets	Non	Longueur de texte maximale pour une instruction de requête SQL sur le moteur d'analyse SQL de Clean Rooms
Longueur du texte de requête sur le moteur d'analyse Spark	Chaque région prise en charge : 500 kilo-octets	Non	Longueur de texte maximale pour une instruction de requête SQL sur le moteur d'analyse Spark
Taux de BatchGetCollaborationAnalysisTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' BatchGetCollaborationAnalysisTemplate API par seconde
Taux de BatchGetSchema demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' BatchGetSchema API par seconde

Nom	Par défaut	Ajusté	Description
Taux de BatchGetSchemaAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' BatchGetSchemaAnalysisRule API par seconde
Taux de CreateAnalysisTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateAnalysisTemplate API par seconde
Taux de CreateCollaboration demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateCollaboration API par seconde
Taux de CreateCollaborationChangeRequest demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateCollaborationChangeRequest API par seconde
Taux de CreateConfiguredAudienceModelAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateConfiguredAudienceModelAssociation API par seconde
Taux de CreateConfiguredTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateConfiguredTable API par seconde
Taux de CreateConfiguredTableAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateConfiguredTableAnalysisRule API par seconde
Taux de CreateConfiguredTableAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateConfiguredTableAssociation API par seconde

Nom	Par défaut	Ajusté	Description
Taux de CreateConfiguredTableAssociationAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateConfiguredTableAssociationAnalysisRule API par seconde
Taux de CreateIdMappingTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateIdMappingTable API par seconde
Taux de CreateIdNamespaceAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateIdNamespaceAssociation API par seconde
Taux de CreateMembership demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreateMembership API par seconde
Taux de CreatePrivacyBudgetTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' CreatePrivacyBudgetTemplate API par seconde
Taux de DeleteAnalysisTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteAnalysisTemplate API par seconde
Taux de DeleteCollaboration demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteCollaboration API par seconde
Taux de DeleteConfiguredAudienceModelAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteConfiguredAudienceModelAssociation API par seconde

Nom	Par défaut	Ajusté	Description
Taux de DeleteConfiguredTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteConfiguredTable API par seconde
Taux de DeleteConfiguredTableAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteConfiguredTableAnalysisRule API par seconde
Taux de DeleteConfiguredTableAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteConfiguredTableAssociation API par seconde
Taux de DeleteConfiguredTableAssociationAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteConfiguredTableAssociationAnalysisRule API par seconde
Taux de DeleteIdMappingTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteIdMappingTable API par seconde
Taux de DeleteIdNamespaceAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteIdNamespaceAssociation API par seconde
Taux de DeleteMember demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteMember API par seconde
Taux de DeleteMembership demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeleteMembership API par seconde

Nom	Par défaut	Ajusté	Description
Taux de DeletePrivacyBudgetTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' DeletePrivacyBudgetTemplate API par seconde
Taux de GetAnalysisTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetAnalysisTemplate API par seconde
Taux de GetCollaboration demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetCollaboration API par seconde
Taux de GetCollaborationAnalysisTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetCollaborationAnalysisTemplate API par seconde
Taux de GetCollaborationChangeRequest demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetCollaborationChangeRequest API par seconde
Taux de GetCollaborationConfiguredAudienceModelAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetCollaborationConfiguredAudienceModelAssociation API par seconde
Taux de GetCollaborationIdNamespaceAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetCollaborationIdNamespaceAssociation API par seconde

Nom	Par défaut	Ajusté	Description
Taux de GetCollaborationPrivacyBudgetTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetCollaborationPrivacyBudgetTemplate API par seconde
Taux de GetConfiguredAudienceModelAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetConfiguredAudienceModelAssociation API par seconde
Taux de GetConfiguredTable demandes	Chaque région prise en charge : 20 par seconde	Oui	Nombre maximal d'appels d' GetConfiguredTable API par seconde
Taux de GetConfiguredTableAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetConfiguredTableAnalysisRule API par seconde
Taux de GetConfiguredTableAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetConfiguredTableAssociation API par seconde
Taux de GetConfiguredTableAssociationAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetConfiguredTableAssociationAnalysisRule API par seconde
Taux de GetIdMappingTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetIdMappingTable API par seconde
Taux de GetIdNamespaceAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetIdNamespaceAssociation API par seconde

Nom	Par défaut	Ajusté	Description
Taux de GetMembership demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetMembership API par seconde
Taux de GetPrivacyBudgetTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetPrivacyBudgetTemplate API par seconde
Taux de GetProtectedJob demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetProtectedJob API par seconde
Taux de GetProtectedQuery demandes	Chaque région prise en charge : 20 par seconde	Oui	Nombre maximal d'appels d' GetProtectedQuery API par seconde
Taux de GetSchema demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetSchema API par seconde
Taux de GetSchemaAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' GetSchemaAnalysisRule API par seconde
Taux de ListAnalysisTemplates demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListAnalysisTemplates API par seconde
Taux de ListCollaborationAnalysisTemplates demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborationAnalysisTemplates API par seconde
Taux de ListCollaborationChangeRequests demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborationChangeRequests API par seconde

Nom	Par défaut	Ajusté	Description
Taux de ListCollaborationConfiguredAudienceModelAssociations demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborationConfiguredAudienceModelAssociations API par seconde
Taux de ListCollaborationIdNamespaceAssociations demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborationIdNamespaceAssociations API par seconde
Taux de ListCollaborationPrivacyBudgetTemplates demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborationPrivacyBudgetTemplates API par seconde
Taux de ListCollaborationPrivacyBudgets demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborationPrivacyBudgets API par seconde
Taux de ListCollaborations demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListCollaborations API par seconde
Taux de ListConfiguredAudienceModelAssociations demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListConfiguredAudienceModelAssociations API par seconde
Taux de ListConfiguredTableAssociations demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListConfiguredTableAssociations API par seconde

Nom	Par défaut	Ajusté	Description
Taux de ListConfiguredTables demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListConfiguredTables API par seconde
Taux de ListIdMappingTables demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListIdMappingTables API par seconde
Taux de ListIdNamespaceAssociations demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListIdNamespaceAssociations API par seconde
Taux de ListMembers demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListMembers API par seconde
Taux de ListMemberships demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListMemberships API par seconde
Taux de ListPrivacyBudgetTemplates demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListPrivacyBudgetTemplates API par seconde
Taux de ListPrivacyBudgets demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListPrivacyBudgets API par seconde
Taux de ListProtectedJobs demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListProtectedJobs API par seconde
Taux de ListProtectedQueries demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListProtectedQueries API par seconde

Nom	Par défaut	Ajusté	Description
Taux de ListSchemas demandes	Chaque Région prise en charge : 5	Oui	Nombre maximal d'appels d' ListSchemas API par seconde
Taux de ListTagsForResource demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' ListTagsForResource API par seconde
Taux de PopulateldMappingTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' PopulateldMappingT able API par seconde
Taux de PreviewPrivacyImpact demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' PreviewPrivacyImpact API par seconde
Taux de StartProtectedJob demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' StartProtectedJob API par seconde
Taux de StartProtectedQuery demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' StartProtectedQuery API par seconde
Taux de TagResource demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' TagResource API par seconde
Taux de UntagResource demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UntagResource API par seconde
Taux de UpdateAnalysisTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateAnalysisTemp late API par seconde

Nom	Par défaut	Ajusté	Description
Taux de UpdateCollaboration demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateCollaboration API par seconde
Taux de UpdateCollaborationChangeRequest demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateCollaborationChangeRequest API par seconde
Taux de UpdateConfiguredAudienceModelAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateConfiguredAudienceModelAssociation API par seconde
Taux de UpdateConfiguredTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateConfiguredTable API par seconde
Taux de UpdateConfiguredTableAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateConfiguredTableAnalysisRule API par seconde
Taux de UpdateConfiguredTableAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateConfiguredTableAssociation API par seconde
Taux de UpdateConfiguredTableAssociationAnalysisRule demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateConfiguredTableAssociationAnalysisRule API par seconde
Taux de UpdateIdMappingTable demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateIdMappingTable API par seconde

Nom	Par défaut	Ajusté	Description
Taux de UpdateIdNamespaceAssociation demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateIdNamespaceAssociation API par seconde
Taux de UpdatePrivacyBudgetTemplate demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdatePrivacyBudgetTemplate API par seconde
Taux de UpdateProtectedJob demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateProtectedJob API par seconde
Taux de UpdateProtectedQuery demandes	Chaque région prise en charge : 5 par seconde	Oui	Nombre maximal d'appels d' UpdateProtectedQuery API par seconde
Table des associations par membre	Chaque région prise en charge : 25	Oui	Nombre maximum d'associations de tables par membre

AWS Clean Rooms limites des paramètres de ressources

Ressource	Par défaut	Description
Longueur du texte de la requête	90 KO	Longueur de texte maximale pour une instruction de requête SQL
Longueur du texte de la requête (en utilisant la confidentialité différentielle)	90 KO	Longueur de texte maximale pour une instruction de requête SQL utilisant la confidentialité différentielle

Ressource	Par défaut	Description
Durée d'exécution de la requête	24 heures	Durée maximale d'exécution d'une requête avant le délai d'expiration

Quotas de machine learning d'AWS Clean Rooms

Vous Compte AWS disposez des quotas par défaut, anciennement appelés limites, pour AWS Clean Rooms ML.

Pour consulter les quotas de service pour AWS Clean Rooms ML, effectuez l'une des opérations suivantes :

- Suivez les étapes décrites à la [section Affichage des quotas](#) de service dans le Guide de l'utilisateur des Quotas de service et sélectionnez AWS Clean Rooms ML comme service.
- Reportez-vous aux [quotas du service AWS Clean Rooms ML](#) dans le Référence générale d'Amazon Web Services.

Pour maintenir les performances du service et garantir une utilisation appropriée d'AWS Clean Rooms ML, les quotas par défaut attribués à un compte peuvent être mis à jour en fonction de facteurs régionaux, de l'historique des paiements, de l'utilisation frauduleuse ou de and/or l'approbation d'une demande d'augmentation de quota.

Nom	Par défaut	Ajuste	Description
Nombre d'emplois d'exportation d'audience actif par travail de génération d'audience	Chaque région prise en charge : 25	Non	Le nombre maximum de tâches d'exportation d'audience actives pour une tâche de génération d'audience
Associations d'algorithmes de modèles configurées actives par membre	Chaque Région prise en charge : 1 000	Oui	Le nombre maximum d'associations d'algorithmes de modèles

Nom	Par défaut	Ajusté	Description
			configurées actives par membre
Algorithmes de modèles configurés actifs par membre	Chaque Région prise en charge : 1 000	Oui	Le nombre maximum d'algorithmes de modèle configurés actifs par membre
Canaux de saisie de modèles personnalisés actifs par membre	Chaque région prise en charge : 100	Oui	Le nombre maximum de canaux de saisie de modèles personnalisés actifs par membre
Nombre maximum d'instances de formation actives par compte	Par région prise en charge : 10	Oui	Nombre maximal d'instances de formation actives pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum de versions de modèles entraînés par active/pending/in progression par modèle entraîné	Chaque région prise en charge : 100	Oui	Nombre maximal de versions de modèles entraînés par active/pending/in progression qui peuvent être créées par modèle entraîné.
Nombre maximal de colonnes d'entrée pour générer des données synthétiques	Chaque Région prise en charge : 1 000	Non	Le nombre maximum de colonnes d'entrée pour générer des données synthétiques
Nombre maximal de lignes d'entrée pour générer des données synthétiques	Chaque région prise en charge : 2 500 000	Non	Le nombre maximum de lignes d'entrée pour générer des données synthétiques

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.c4.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c4.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c4.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c4.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c4.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximum d'instances de formation ml.c4.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c4.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.c4.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5.18xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.c5.18xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.c5.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5.9xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5.9xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.c5.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5n.18xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5n.18xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.c5n.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5n.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5n.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5n.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5n.9xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5n.9xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c5n.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c5n.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c6i.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.c6i.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c6i.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c6i.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c6i.32xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.32xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c6i.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.c6i.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c6i.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c6i.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.c7i.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.c7i.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.c7i.large pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.c7i.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.c7i.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g4dn.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g4dn.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g4dn.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g4dn.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g4dn.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g4dn.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g4dn.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g4dn.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.g4dn.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g4dn.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g4dn.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g4dn.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g5.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g5.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g5.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.g5.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g5.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g5.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximum d'instances de formation ml.g5.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximum d'instances de formation ml.g5.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g5.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g5.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.g6.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g6.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g6.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g6.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g6.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.g6.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximum d'instances de formation ml.g6.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximum d'instances de formation ml.g6.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g6.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6e.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6e.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.g6e.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6e.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6e.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6e.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.g6e.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.g6e.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.g6e.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.g6e.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m4.10xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m4.10xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m4.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m4.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m4.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m4.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m4.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m4.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.m4.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m4.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m5.12xlarge par compte	Chaque région prise en charge : 3	Oui	Nombre maximal d'instances de formation ml.m5.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m5.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m5.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m5.2xlarge par compte	Chaque région prise en charge : 3	Oui	Nombre maximal d'instances d'entraînement ml.m5.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m5.4xlarge par compte	Chaque région prise en charge : 3	Oui	Nombre maximal d'instances d'entraînement ml.m5.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.m5.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m5.large pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m5.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m5.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.m6i.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.32xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.32xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m6i.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m6i.large pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.m6i.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m6i.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.m7i.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.m7i.large pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.m7i.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.m7i.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.p2.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.p2.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.p2.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.p2.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.p2.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.p2.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.p4d.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.p4d.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.p4de.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.p4de.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.p5.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.p5.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.p5en.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.p5en.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r5.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r5.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.r5.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r5.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r5.large pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r5.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.r5d.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5d.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5d.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5d.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5d.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.r5d.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5d.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r5d.large pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r5d.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r5d.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.12xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.12xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.16xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.16xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.r7i.24xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.24xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.48xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.48xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.4xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.4xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.8xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.8xlarge pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajuste	Description
Nombre maximum d'instances de formation ml.r7i.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.r7i.large pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.r7i.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.r7i.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.t3.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.t3.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.t3.large par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.t3.large pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.t3.medium par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.t3.medium pouvant être utilisées par compte pour créer des modèles entraînés.

Nom	Par défaut	Ajusté	Description
Nombre maximum d'instances de formation ml.t3.xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances de formation ml.t3.xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.trn1.2xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.trn1.2xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.trn1.32xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.trn1.32xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Nombre maximum d'instances de formation ml.trn1n.32xlarge par compte	Chaque région prise en charge : 0	Oui	Nombre maximal d'instances d'entraînement ml.trn1n.32xlarge pouvant être utilisées par compte pour créer des modèles entraînés.
Valeurs uniques maximales dans la colonne de prédiction catégorique	Chaque région prise en charge : 100	Non	Les valeurs uniques maximales dans la colonne de prédiction catégorique

Nom	Par défaut	Ajusté	Description
Nombre minimal de colonnes d'entrée requises pour générer des données synthétiques	Chaque région prise en charge : 5	Non	Le nombre minimum de colonnes d'entrée requises pour générer des données synthétiques
Nombre minimal de lignes d'entrée requises pour générer des données synthétiques	Chaque région prise en charge : 1 500	Non	Le nombre minimum de lignes d'entrée requises pour générer des données synthétiques
Tâches d'exportation d'audience en attente ou en cours par client	Chaque Région prise en charge : 20	Non	Le nombre maximum de tâches d'exportation d'audience en attente ou en cours par client
Tâches de génération d'audience en attente ou en cours par client	Par région prise en charge : 10	Oui	Le nombre maximum de tâches de génération d'audience en attente ou en cours par client
Modèles d'audience en attente ou en cours par client	Chaque région prise en charge : 2	Oui	Le nombre maximum de tâches de formation sur les modèles d'audience en attente ou en cours par client
Tâches d'inférence de modèles personnalisés en attente ou en cours par compte	Par région prise en charge : 10	Oui	Le nombre maximum de tâches d'inférence de modèles personnalisés en attente ou en cours par compte

Nom	Par défaut	Ajuste	Description
Tâches d'inférence de modèles personnalisés en attente ou en cours par abonnement	Chaque Région prise en charge : 5	Oui	Le nombre maximum de tâches d'inférence de modèles personnalisés en attente ou en cours par adhésion
Modèles de formation personnalisés en attente ou en cours par compte	Par région prise en charge : 10	Oui	Le nombre maximum de tâches de formation sur modèle personnalisé en attente ou en cours par compte
Modèles de formation personnalisés en attente ou en cours par abonnement	Chaque Région prise en charge : 5	Oui	Le nombre maximum de tâches de formation sur modèle personnalisé en attente ou en cours par adhésion
Emplois de génération de MLIC synthétiques en attente ou en cours par compte	Chaque région prise en charge : 2	Oui	Le nombre maximum de tâches de génération de MLIC synthétiques en attente ou en cours par compte

AWS Clean Rooms ML possède des quotas supplémentaires répertoriés ci-dessous

Quotas ML pour Clean Rooms

Ressource	Par défaut	Description
Durée maximale d'une tâche d'inférence de modélisation personnalisée	25 heures	
Nombre maximum d'interactions	20 milliards	Nombre maximum d'interactions autorisées dans les

Ressource	Par défaut	Description
		données d'entraînement. Les entrées plus importantes sont échantillonnées vers le bas.
Nombre minimal d'interactions	1 million	
Nombre maximum d'utilisateurs distincts pour la formation sur les modèles similaires	100 millions	Si d'autres sont inclus, seuls les 100 millions les plus populaires sont utilisés, classés en fonction du nombre d'interactions.
Nombre minimal d'utilisateurs distincts pour la formation sur les modèles similaires	100 000	
Nombre minimum d'utilisateurs pour la tâche d'exportation portant sur un segment similaire (public)	10 000	
Nombre maximum d'éléments distincts utilisés pour l'entraînement des modèles.	1 million	Vous pouvez inclure jusqu'à 50 millions d'articles, mais seul le million le plus populaire est utilisé.
Nombre maximal de colonnes d'entités dans le jeu de données d'entraînement.	10	
Nombre minimum d'éléments distincts par utilisateur	2	AWS Clean Rooms ML exige que chaque ligne ou utilisateur possède au moins deux éléments, y compris des éléments répétés.

Ressource	Par défaut	Description
Taille maximale de l'audience initiale	500 000	
Taille minimale de l'audience initiale	500	Le fournisseur de données de formation peut définir cette valeur à une valeur aussi basse que 25.
APIs	par client	
Nombre total de jeux de données d'entraînement actifs	500	
Nombre total de modèles similaires actifs (modèles d'audience)	500	
Nombre total de modèles similaires configurés actifs (modèles d'audience)	10 000	
Nombre total de jobs de génération de segments similaires (audience) terminés	Aucune limite	
Nombre total de jobs terminés dans un segment similaire à l'exportation (public)	Aucune limite	
Durée maximale d'une tâche de génération de modèle similaire (modèle d'audience)	1 jour (24 heures)	

Ressource	Par défaut	Description
Durée maximale d'une tâche de génération de segments similaires (audience)	10 heures	Une fois que vous avez fourni une graine, Clean Rooms ML met au maximum 10 heures pour générer un segment similaire. Si vous utilisez une requête SQL comme données de départ, l'exécution de la requête peut prendre jusqu'à 12 heures, en plus des 10 heures nécessaires pour générer le segment similaire.
Pourcentage minimum pour une classe de taille de segment (audience)	1 %	
Pourcentage maximal pour un groupe de taille de segment (audience)	20 %	
Taille absolue minimale pour un compartiment de taille de segment (audience)	1 % du nombre d'utilisateurs distincts	
Taille absolue maximale pour un compartiment de taille de segment (audience)	20 % du nombre d'utilisateurs distincts	

Limitation des quotas de l'API Clean Rooms ML

Vous Compte AWS disposez des quotas de transaction API par seconde (TPS) suivants par compte et par point de terminaison.

Ressource	Limite de débit	Description
Taux de CreateAudienceModel demandes	Taux de 1 TPS, rafale de 3 TPS	Nombre maximal d'appels d>CreateAudienceModel API par seconde
Taux de CreateConfiguredAudienceModel demandes	10 TPS	Nombre maximal d'appels d>CreateConfiguredAudienceModel API par seconde
Taux de CreateTrainingDataset demandes	10 TPS	Nombre maximal d'appels d>CreateTrainingDataset API par seconde
Taux de DeleteAudienceGenerationJob demandes	Débit de 2 TPS, rafale de 10 TPS	Nombre maximal d'appels d>DeleteAudienceGenerationJob API par seconde
Taux de DeleteAudienceModel demandes	Débit de 2 TPS, rafale de 10 TPS	Nombre maximal d'appels d>DeleteAudienceModel API par seconde
Taux de DeleteConfiguredAudienceModel demandes	10 TPS	Nombre maximal d'appels d>DeleteConfiguredAudienceModel API par seconde
Taux de DeleteConfiguredAudienceModelPolicy demandes	25 TPS	Nombre maximal d'appels d>DeleteConfiguredAudienceModelPolicy API par seconde
Taux de DeleteTrainingDataset demandes	10 TPS	Nombre maximal d'appels d>DeleteTrainingDataset API par seconde

Ressource	Limite de débit	Description
Taux de GetAudienceGenerationJob demandes	50 TPS	Nombre maximal d'appels d'GetAudienceGenerationJob API par seconde
Taux de GetAudienceModel demandes	50 TPS	Nombre maximal d'appels d'GetAudienceModel API par seconde
Taux de GetConfiguredAudienceModel demandes	50 TPS	Nombre maximal d'appels d'GetConfiguredAudienceModel API par seconde
Taux de GetConfiguredAudienceModelPolicy demandes	50 TPS	Nombre maximal d'appels d'GetConfiguredAudienceModelPolicy API par seconde
Taux de GetTrainingDataset demandes	50 TPS	Nombre maximal d'appels d'GetTrainingDataset API par seconde
Taux de ListAudienceExportJobs demandes	50 TPS	Nombre maximal d'appels d>ListAudienceExportJobs API par seconde
Taux de ListAudienceGenerationJobs demandes	50 TPS	Nombre maximal d'appels d>ListAudienceGenerationJobs API par seconde
Taux de ListAudienceModels demandes	50 TPS	Nombre maximal d'appels d>ListAudienceModels API par seconde

Ressource	Limite de débit	Description
Taux de <code>ListConfiguredAudienceModels</code> demandes	50 TPS	Nombre maximal d'appels d' <code>ListConfiguredAudienceModels</code> API par seconde
Taux de <code>ListTagsForResource</code> demandes	50 TPS	Nombre maximal d'appels d' <code>ListTagsForResource</code> API par seconde
Taux de <code>ListTrainingDatasets</code> demandes	50 TPS	Nombre maximal d'appels d' <code>ListTrainingDatasets</code> API par seconde
Taux de <code>PutConfiguredAudienceModelPolicy</code> demandes	25 TPS	Nombre maximal d'appels d' <code>PutConfiguredAudienceModelPolicy</code> API par seconde
Taux de <code>StartAudienceExportJob</code> demandes	Taux de 1 TPS, rafale de 3 TPS	Nombre maximal d'appels d' <code>StartAudienceExportJob</code> API par seconde
Taux de <code>StartAudienceGenerationJob</code> demandes	Taux de 1 TPS, rafale de 5 TPS	Nombre maximal d'appels d' <code>StartAudienceGenerationJob</code> API par seconde
Taux de <code>TagResource</code> demandes	10 TPS	Nombre maximal d'appels d' <code>TagResource</code> API par seconde
Taux de <code>UntagResource</code> demandes	50 TPS	Nombre maximal d'appels d' <code>UntagResource</code> API par seconde

Ressource	Limite de débit	Description
Taux de UpdateConfiguredAudienceModel demandes	10 TPS	Nombre maximal d'appels d'UpdateConfiguredAudienceModel API par seconde
Taux de CreateConfiguredModelAlgorithm demandes	10 TPS	Nombre maximal d'appels d>CreateConfiguredModelAlgorithm API par seconde.
Taux de CreateConfiguredModelAlgorithmAssociation demandes	10 TPS	Nombre maximal d'appels d>CreateConfiguredModelAlgorithmAssociation API par seconde.
Taux de PutMLConfiguration demandes	10 TPS	Nombre maximal d'appels d'PutMLConfiguration API par seconde.
Taux de CreateTrainedModel demandes	Taux de 1 TPS, rafale de 3 TPS	Nombre maximal d'appels d>CreateTrainedModel API par seconde.
Taux de StartTrainedModelExportJob demandes	10 TPS	Nombre maximal d'appels d'StartTrainedModelExportJob API par seconde.
Taux de StartTrainedModelInferenceJob demandes	Débit 1 TPS, débit 3 TPS	Nombre maximal d'appels d'StartTrainedModelInferenceJob API par seconde.
Taux de GetConfiguredModelAlgorithm demande	50 TPS	Nombre maximal d'appels d'GetConfiguredModelAlgorithm API par seconde.

Ressource	Limite de débit	Description
Taux de GetConfiguredModelAlgorithmAssociation demande	50 TPS	Nombre maximal d'appels d'GetConfiguredModelAlgorithmAssociation API par seconde.
Taux de GetTrainedModel demandes	50 TPS	Nombre maximal d'appels d'GetTrainedModel API par seconde.
Taux de GetMLConfiguration demandes	50 TPS	Nombre maximal d'appels d'GetMLConfiguration API par seconde.
Taux de GetTrainedModelInferenceJob demandes	50 TPS	Nombre maximal d'appels d'GetTrainedModelInferenceJob API par seconde.
Taux de ListConfiguredModelAlgorithm demandes	50 TPS	Nombre maximal d'appels d'ListConfiguredModelAlgorithm API par seconde.
Taux de ListConfiguredModelAlgorithmAssociations demandes	50 TPS	Nombre maximal d'appels d'ListConfiguredModelAlgorithmAssociations API par seconde.
Taux de ListTrainedModels demandes	50 TPS	Nombre maximal d'appels d'ListTrainedModels API par seconde.
Taux de ListCollaborationTrainedModelExportJobs demandes	50 TPS	Nombre maximal d'appels d'ListCollaborationTrainedModelExportJobs API par seconde.

Ressource	Limite de débit	Description
Taux de ListCollaborationTrainedModelInferenceJobs demandes	50 TPS	Nombre maximal d'appels d'ListCollaborationTrainedModelInferenceJobs API par seconde.
Taux de DeleteConfiguredModelAlgorithm demandes	Débit de 2 TPS, rafale de 10 TPS	Nombre maximal d'appels d>DeleteConfiguredModelAlgorithm API par seconde.
Taux de DeleteConfiguredModelAlgorithmAssociation demandes	Débit de 2 TPS, rafale de 10 TPS	Nombre maximal de demandes d>DeleteConfiguredModelAlgorithmAssociation API par seconde.
Taux de DeleteMLCconfiguration demandes	Débit de 2 TPS, rafale de 10 TPS	Nombre maximal de demandes d>DeleteMLCconfiguration API par seconde.
Taux de DeleteTrainedModelOutput demandes	Débit de 2 TPS, rafale de 10 TPS	Nombre maximal de demandes d>DeleteTrainedModelOutput API par seconde.

Historique du document pour le guide de AWS Clean Rooms l'utilisateur

Le tableau suivant décrit les versions de documentation pour AWS Clean Rooms.

Pour recevoir les notifications sur les mises à jour de cette documentation, vous pouvez vous abonner au Flux RSS. Pour vous abonner aux mises à jour RSS, votre navigateur doit disposer d'un plug-in RSS activé.

Modification	Description	Date
Documentation mise à jour pour la section de balisage de la répartition des coûts	Ajout de documentation AWS Billing and Cost Management et création d'une section pour les ressources étiquetables liées à la répartition des coûts.	20 février 2026
Support pour la configuration de cleanrooms-ml Spark Compute	Les clients peuvent désormais personnaliser le comportement d'exécution de Spark et optimiser les performances SQL en configurant les propriétés Spark prises en charge pour les requêtes SQL. Cette fonctionnalité est désormais disponible pour la création d'emplois de génération d'audience et la création de MLInput chaînes .	5 janvier 2026
Support pour une surveillance détaillée avec CloudWatch	Les clients peuvent désormais opter pour une surveillance détaillée avec des CloudWatch indicateurs pour le suivi opérationnel des collaborations, notamment les	2 janvier 2026

performances des requêtes et l'utilisation des ressources.

[Support pour les événements de EventBridge notification des demandes de modification liées à la collaboration](#)

Ajout de nouveaux EventBridge événements pour les événements du cycle de vie des demandes de modification de collaboration.

19 décembre 2025

[Support pour les demandes de modification visant à gérer les adhésions et les paramètres de collaboration](#)

Les créateurs de collaborations peuvent désormais soumettre des demandes de modification pour ajouter de nouveaux membres, mettre à jour les capacités des membres existants et modifier les paramètres d'approbation automatique. Tous les membres de la collaboration doivent approuver les demandes de modification avant que les modifications ne prennent effet.

18 décembre 2025

[Calendrier de fin de support pour le moteur d'analyse AWS Clean Rooms SQL](#)

Support pour le moteur d'analyse AWS Clean Rooms SQL prend fin le 17 décembre 2025. Les clients peuvent désormais utiliser uniquement le moteur d'analyse Spark SQL.

17 décembre 2025

[Support pour les événements EventBridge de notification supplémentaires](#)

Ajout de nouveaux EventBridge événements pour le modèle d'analyse, l'association de tables configurée, la règle d'analyse des associations de tables configurée, la table de mappage d'identifiants, l'association d'espaces de noms d'identifiants, le job protégé et la requête protégée.

15 décembre 2025

[Support à la génération de jeux de données synthétiques garantissant la confidentialité pour la formation de modèles d'apprentissage automatique personnalisés](#)

Les clients peuvent désormais générer des ensembles de données synthétiques respectueux de la vie privée pour former des modèles d'apprentissage automatique personnalisés tout en protégeant les données sensibles.

30 novembre 2025

[Support des propriétés Spark configurables dans les requêtes SQL](#)

Les clients peuvent désormais personnaliser le comportement d'exécution de Spark en configurant les propriétés Spark prises en charge pour les requêtes SQL.

30 octobre 2025

[Support à la collaboration interrégionale](#)

AWS Clean Rooms prend désormais en charge la collaboration interrégionale pour AWS Glue les tables de données Athena, ce qui permet aux clients de traiter les données stockées dans Région AWS une seule et de fournir des résultats aux régions approuvées par tous les membres de la collaboration.

3 octobre 2025

[Support pour les budgets d'accès aux données](#)

Les clients peuvent désormais définir et gérer les budgets d'accès aux données afin de contrôler l'utilisation des tables.

1er octobre 2025

[Remplir le tableau de mappage des identifiants — mise à jour](#)

Les clients peuvent désormais utiliser le traitement incrémentiel dans un flux de travail de mappage d'identité basé sur des règles afin de traiter plus efficacement de grands ensembles de données. Les clients peuvent également supprimer des enregistrements d'un flux de travail de mappage des identifiants pour se conformer aux réglementations en matière de gestion des données.

23 septembre 2025

<u>Support pour des charges de travail variables pour les PySpark tâches</u>	Les clients peuvent désormais choisir et ajuster leur taille de calcul pour exécuter les PySpark tâches plus efficacement.	4 septembre 2025
<u>Support pour l'ajout de nouveaux membres à une collaboration existante</u>	Les clients peuvent désormais ajouter de nouveaux membres aux collaborations existantes. Cette fonctionnalité inclut des options permettant de configurer les paramètres d'approbation automatique et de bloquer les demandes d'adhésion.	3 septembre 2025
<u>Ajout de la prise en charge de la configuration des résumés des erreurs et de la rédaction des journaux de modèles personnalisés</u>	Les clients peuvent désormais configurer des résumés d'erreurs et une rédaction pour les journaux de modèles personnalisés dans le cadre d'une association d'algorithmes de modèles configurée.	3 septembre 2025
<u>Support pour l'option d'authentification Snowflake</u>	Les clients peuvent désormais utiliser l'authentification par paire de clés avec une clé privée Snowflake PEM lorsqu'ils stockent leurs informations d'identification Snowflake pour les connexions. AWS Secrets Manager AWS Clean Rooms	29 août 2025

<u>Configuration d'erreur ajoutée pour les modèles PySpark d'analyse</u>	Les clients peuvent désormais configurer des niveaux de message d'erreur de base ou détaillés dans les modèles d'PySpark analyse afin de gérer la visibilité des erreurs et les capacités de résolution des problèmes.	19 août 2025
<u>Support pour EventBridge</u>	Les clients peuvent désormais utiliser Amazon EventBridge pour acheminer des événements depuis AWS Clean Rooms d'autres composants logiciels.	31 juillet 2025
<u>Mise à jour de la politique existante</u>	Les nouvelles autorisations suivantes ont été ajoutées à la politique AWSCleanRoomsFullAccessNoQuering gérée : cleanrooms:UpdateConfiguredTableAllowedColumns et cleanrooms:UpdateConfiguredTableReference .	29 juillet 2025
<u>Capacités d'édition accrues pour les tables configurées</u>	Les clients peuvent désormais modifier les colonnes autorisées et les champs de référence des tables pour les tables configurées.	29 juillet 2025

[Le canal d'entrée ML prend en charge le type de fichier Parquet](#)

Les clients peuvent désormais choisir le type de fichier Parquet comme format de données que le canal d'entrée ML peut utiliser.

17 juillet 2025

[Support de la confidentialité différentielle sur le moteur d'analyse Spark SQL](#)

Les clients peuvent désormais utiliser la confidentialité différentielle avec le moteur d'analyse Spark SQL. En outre, la longueur maximale du texte d'une instruction de requête SQL utilisant la confidentialité différentielle est désormais de 90 Ko.

16 juillet 2025

[Calendrier de fin de support pour le moteur d'analyse AWS Clean Rooms SQL](#)

Support pour les nouvelles collaborations utilisant le moteur d'analyse SQL de Clean Rooms prend fin le 16 juillet 2025. À compter du 17 juillet 2025, vous ne pourrez plus créer de nouvelles collaborations avec l'ancien moteur d'analyse SQL de Clean Rooms. AWS Clean Rooms met fin au support du moteur d'analyse SQL de Clean Rooms le 17 décembre 2025.

16 juillet 2025

[Support pour la formation progressive et distribuée](#)

Les clients peuvent désormais utiliser la formation incrémentielle et la formation distribuée pour leurs modèles personnalisés.

1er juillet 2025

[Calendrier de fin de support pour le moteur d'analyse SQL de Clean Rooms](#)

À compter du 16 juillet 2025, les nouvelles collaborations utilisant l'ancien moteur d'analyse SQL Clean Rooms nécessitent une augmentation de la limite. À compter du 17 juillet 2025, vous ne pourrez plus créer de nouvelles collaborations avec ce moteur. AWS Clean Rooms met fin au support du moteur d'analyse SQL de Clean Rooms le 17 décembre 2025.

22 mai 2025

[Mise à jour des modèles PySpark d'analyse](#)

Les clients peuvent désormais nommer un fichier de script utilisateur avec n'importe quel nom de fichier Python valide.

15 mai 2025

[Support pour plusieurs récepteurs de résultats](#)

Les clients peuvent désormais sélectionner plusieurs membres pour recevoir les résultats d'une seule requête.

30 avril 2025

[Support pour la migration des collaborations vers Spark SQL](#)

AWS Clean Rooms SQL prend désormais en charge les règles d'agrégation et d'analyse de listes, en plus des règles d'analyse personnalisées. En outre, les clients peuvent mettre à jour une collaboration existante pour utiliser le moteur d'analyse Spark qui alimente Spark SQL.

2 avril 2025

Support à l' PySpark emploi	Les clients peuvent désormais analyser les données en exécutant des tâches à l'aide de modèles PySpark d'analyse approuvés.	18 mars 2025
Mise à jour des politiques existantes	La nouvelle autorisation suivante a été ajoutée à la politique AWSCleanRoomsMLReadOnlyAccess gérée : PassCleanRoomsResources . Les nouvelles autorisations suivantes ont été ajoutées à la politique AWSCleanRoomsMLFullAccess gérée : PassCleanRoomsResources etConsoleDescribeECRRepositories .	10 janvier 2025
Support pour plusieurs informaticiens	Les clients peuvent désormais spécifier le type de personnel informatique et le nombre à affecter lors de la création d'un segment similaire.	17 décembre 2024
Support pour plusieurs sources de données et clouds	Les clients peuvent désormais utiliser plusieurs sources de données et clouds, tels qu'Amazon Athena et Snowflake, pour collaborer avec les ensembles de données de leurs partenaires.	1er décembre 2024

[La modélisation personnalisée de Clean Rooms ML est désormais disponible](#)

Les clients peuvent désormais utiliser leurs propres modèles de machine learning personnalisés dans le cadre d'une collaboration.

7 novembre 2024

[Nouveau moteur d'analyse](#)

Les clients disposant de grands ensembles de données peuvent désormais exécuter des requêtes complexes à l'aide des fonctions SQL prises en charge par le moteur d'analyse Spark SQL.

29 octobre 2024

[Protection de la confidentialité améliorée, création d'audiences similaires, choix de plusieurs destinataires de résultats](#)

Vous pouvez protéger vos données tout en autorisant les requêtes d'activation complexes à l'aide d'analyses supplémentaires et de la règle d'analyse de collaboration. Vous pouvez créer des modèles d'audience similaires à partir de requêtes SQL ou de modèles d'analyse. Vous pouvez sélectionner plusieurs membres pour recevoir les résultats.

24 juillet 2024

[Résolution de l'entité en AWS Clean Rooms](#)

Avec Résolution des entités AWS in AWS Clean Rooms, vous pouvez créer une table de mappage d'identifiants entre deux espaces de noms d'identification pour interroger les données d'événements dans des espaces d'identité disparates.

23 juillet 2024

<u>Mise à jour de la politique existante</u>	La nouvelle autorisation suivante a été ajoutée à la politique <code>AWSCleanRoomsFullAccessNoQuerying</code> gérée : <code>cleanrooms:BatchGetSchemaAnalysisRule</code> .	13 mai 2024
<u>AWS Clean Rooms ML est désormais entièrement disponible</u>	AWS Clean Rooms ML fournit une méthode d'amélioration de la confidentialité permettant à deux parties d'identifier des utilisateurs similaires dans leurs données sans avoir à partager leurs données entre elles.	3 avril 2024
<u>Mise à jour de la politique existante</u>	L'ID de déclaration dans la politique <code>AWSCleanRoomsFullAccess</code> gérée a été mis à jour de <code>ConsolePickQueryResultsBucket</code> à <code>SetQueryResultsBucket</code> pour mieux représenter les autorisations depuis les autorisations.	21 mars 2024
<u>Nouvelles politiques gérées pour AWS Clean Rooms ML</u>	Deux nouvelles politiques gérées ont été ajoutées : <code>AWSCleanRoomsMLReadOnlyAccess</code> et <code>AWSCleanRoomsMLFullAccess</code> .	29 novembre 2023

AWS Clean Rooms ML (version préliminaire)	AWS Clean Rooms ML fournit une méthode d'amélioration de la confidentialité permettant à deux parties d'identifier des utilisateurs similaires dans leurs données sans avoir à partager leurs données entre elles.	29 novembre 2023
AWS Clean Rooms Confidentialité différentielle (aperçu)	Les clients peuvent désormais utiliser la confidentialité AWS Clean Rooms différentielle pour protéger la vie privée de leurs utilisateurs.	29 novembre 2023
Configuration du paiement	Le créateur de la collaboration peut désormais configurer le membre autorisé à exécuter des requêtes ou un autre membre de la collaboration pour qu'il soit facturé pour les coûts de calcul des requêtes.	14 novembre 2023
Durée d'exécution de la requête - mise à jour	La durée maximale d'exécution d'une requête avant la mise à jour du délai d'expiration passe de 4 heures à 12 heures.	6 octobre 2023

[CloudFormation ressources - mise à jour](#)

AWS Clean Rooms a ajouté les nouvelles ressources suivantes :
 AWS::CleanRooms::Membership Protected QueryOutputConfiguration
 AWS::CleanRooms::Membership ProtectedQueryResultConfiguration ,
 etAWS::CleanRooms::Membership Protected QueryS3OutputConfiguration .

7 septembre 2023

[CloudFormation ressources - mise à jour](#)

AWS Clean Rooms a ajouté les nouvelles ressources suivantes : AWS::CleanRooms::AnalysisTemplate etAWS::CleanRooms::ConfiguredTable AnalysisRuleCustom .

31 août 2023

[Capacités distinctes des membres](#)

Le créateur de la collaboration peut désormais désigner un membre en tant que membre habilité à effectuer des requêtes et un autre membre en tant que membre habilité à recevoir les résultats. Cela permet au créateur de la collaboration de s'assurer que le membre autorisé à effectuer une requête n'a pas accès aux résultats de la requête.

30 août 2023

AWS Clean Rooms Glossaire	Mise à jour portant uniquement sur la documentation pour ajouter un glossaire des termes. AWS Clean Rooms	30 août 2023
Support pour Apache Iceberg les tableaux (aperçu)	AWS Clean Rooms supporte désormais Apache Iceberg les tables (aperçu).	25 août 2023
Mise à jour des quotas	La section Quotas a été mise à jour pour refléter le nouveau quota par défaut d'abonnements par compte.	9 août 2023
Mise à jour de la politique existante	Les nouvelles autorisations suivantes ont été ajoutées à la politique AWSCleanRoomsFullAccessNoQuering gérée : cleanrooms:CreateAnalysisTemplates cleanrooms:GetAnalysisTemplate , cleanrooms:UpdateAnalysisTemplate cleanrooms>DeleteAnalysisTemplate , cleanrooms>ListAnalysisTemplates , cleanrooms:GetCollaborationAnalysisTemplate , cleanrooms:BatchGetCollaborationAnalysisTemplate , et cleanrooms>ListCollaborationAnalysisTemplates .	31 juillet 2023

Modèles d'analyse et règle d'analyse personnalisée	AWS Clean Rooms prend désormais en charge les modèles d'analyse et la règle d'analyse personnalisée. Les modèles d'analyse permettent aux collaborateurs de créer ou d'importer leur propre requête SQL personnalisée à utiliser dans le cadre de la collaboration. Avec la règle d'analyse personnalisée, le propriétaire de la table peut approuver des requêtes SQL personnalisées sur ses tables configurées.	31 juillet 2023
Les règles d'analyse soutiennent la condition OR logique	AWS Clean Rooms les règles d'analyse prennent désormais en charge la condition OR logique de la JOIN clause.	29 juin 2023
CloudFormation intégration	AWS Clean Rooms s'intègre désormais à CloudFormation.	15 juin 2023
Générateur d'analyses	Les membres autorisés à effectuer des requêtes et à recevoir des résultats peuvent désormais exécuter des requêtes sur certaines tables sans écrire de code SQL à l'aide de l'interface utilisateur du générateur d'analyse.	15 juin 2023
fonctions SQL	Mise à jour uniquement destinée à la documentation pour clarifier les fonctions SQL prises en charge.	5 mai 2023

Résolution des problèmes	Mise à jour basée uniquement sur la documentation pour ajouter une section de résolution des problèmes courants.	27 avril 2023
Types de données pris en charge pour AWS Clean Rooms	Mise à jour relative à la documentation uniquement pour ajouter une nouvelle section répertoriant les types de AWS Glue Data Catalog données pris en charge.	26 avril 2023
Exemples d' AWS CloudTrail événements	Mise à jour de documentation uniquement pour ajouter des exemples d' CloudTrail événements pour StartProtectedQuery (réussite) et StartProtectedQuery (échec).	20 avril 2023
Mise à jour de la politique existante	Les nouvelles autorisations suivantes ont été ajoutées à la politique AWSCleanRoomsFullAccessNoQuerying gérée :cleanrooms:ListTagsForResource ,cleanrooms:UntagResource , etcleanrooms:TagResource . Pour plus d'informations, consultez la section Politiques AWS gérées .	21 mars 2023
Disponibilité générale	AWS Clean Rooms est désormais disponible pour tous.	21 mars 2023

[Version préliminaire](#)

Version préliminaire du guide
de AWS Clean Rooms l'utilisa
teur

12 janvier 2023

AWS Clean Rooms Glossaire

Consultez ce glossaire pour vous familiariser avec la terminologie utilisée pour AWS Clean Rooms.

Règle d'analyse d'agrégation

La restriction de requête qui autorise les requêtes qui regroupent l'analyse COUNT en SUM utilisant ou AVG fonctionnant selon des dimensions facultatives. Ces requêtes ne révéleront pas d'informations au niveau des lignes.

Prend en charge des cas d'utilisation tels que la planification des campagnes, la portée médiatique, la fréquence et la mesure des conversions.

Les autres types de règles d'analyse sont [personnalisées](#) et [listées](#).

Règles d'analyse

Les restrictions de requête qui autorisent un type de requête spécifique.

Le type de règle d'analyse détermine le type d'analyse qui peut être exécuté sur la table configurée. Chaque type possède une structure de requête prédéfinie. Vous contrôlez la manière dont les colonnes de votre table peuvent être utilisées dans la structure par le biais des commandes de requête.

Les types de règles d'analyse sont [l'agrégation](#), [la liste](#) et les règles [personnalisées](#).

Modèle d'analyse

Une requête pré-approuvée spécifique à la collaboration qui peut être réutilisée.

Formats pris en charge : code SQL ou code Python pour Spark.

Si vous utilisez le langage SQL, le modèle d'analyse peut contenir des paramètres là où une valeur littérale peut généralement apparaître dans une requête SQL. Pour plus d'informations sur les types de paramètres pris en charge, consultez la section [Types de données](#) de la référence AWS Clean Rooms SQL.

Les modèles d'analyse fonctionnent uniquement avec la [règle d'analyse personnalisée](#).

Client de chiffrement C3R

Le client de chiffrement Cryptographic Computing for Clean Rooms (C3R).

Utilisé pour chiffrer et déchiffrer des données, C3R est un SDK de chiffrement côté client doté d'une interface en ligne de commande.

Colonne en texte clair

Colonne qui n'est protégée par cryptographie ni pour une construction A JOIN ni pour une construction SELECT SQL.

Les colonnes en texte clair peuvent être utilisées dans n'importe quelle partie de la requête SQL.

Collaboration

Limite logique sécurisée AWS Clean Rooms dans laquelle les membres peuvent exécuter des requêtes SQL sur des tables configurées.

Les collaborations sont créées par le [créateur de la collaboration](#).

Seuls les membres qui ont été invités à participer à la collaboration peuvent rejoindre la collaboration.

Une collaboration ne peut avoir qu'un seul [membre qui peut interroger](#) des données ou un seul [membre qui peut exécuter des requêtes et des tâches](#).

Une collaboration ne peut avoir qu'un seul [membre qui peut recevoir les résultats](#).

Dans une collaboration, un seul [membre peut payer les coûts de calcul des requêtes](#) ou un seul [membre payant les coûts de calcul des requêtes et des tâches](#).

Tous les membres peuvent consulter la liste des participants invités à la collaboration avant de rejoindre la collaboration.

Créateur de collaboration

Le membre qui crée une collaboration.

Il n'y a qu'un seul créateur de collaboration par collaboration.

Seul le créateur de la collaboration peut retirer des membres de la collaboration ou supprimer la collaboration.

Table configurée

Chaque table configurée représente une référence à une table existante dans le AWS Glue Data Catalog qui a été configurée pour être utilisée dans AWS Clean Rooms. Une table configurée contient une règle d'analyse qui détermine la manière dont les données peuvent être utilisées.

Actuellement, AWS Clean Rooms prend en charge l'association de données stockées dans Amazon Simple Storage Service (Amazon S3) qui sont cataloguées via. AWS Glue

Pour plus d'informations à ce sujet AWS Glue, consultez le [guide du AWS Glue développeur](#).

Les tables configurées peuvent être associées à une ou plusieurs collaborations.

Note

AWS Clean Rooms ne prend actuellement pas en charge les emplacements de compartiment Amazon S3 enregistrés auprès de AWS Lake Formation.

Règle d'analyse personnalisée

La restriction de requête qui autorise un ensemble spécifique de requêtes préapprouvées ([modèles d'analyse](#)) ou autorise un ensemble spécifique de comptes capables de fournir des requêtes ou des tâches utilisant vos données.

Prend en charge des cas d'utilisation tels que l'attribution au premier contact, les analyses incrémentielles et les analyses de découverte d'audience.

Soutient la confidentialité différentielle.

Les autres types de règles d'analyse sont [l'agrégation](#) et [la liste](#).

Déchiffrement

Le processus qui consiste à remettre les données chiffrées dans leur forme d'origine. Le déchiffrement ne peut être effectué que si vous avez accès à la clé secrète.

Confidentialité différentielle

Une technique mathématiquement rigoureuse qui protège les données de collaboration contre le membre qui peut recevoir des résultats en apprenant sur une personne en particulier.

Chiffrement

Processus consistant à coder des données sous une forme qui semble aléatoire à l'aide d'une valeur secrète appelée clé. Il est impossible de déterminer le texte brut d'origine sans accéder à la clé.

Colonne d'empreintes digitales

Colonne protégée cryptographiquement pour une construction JOIN SQL.

Méthode de workflow de mappage des identifiants

Comment souhaitez-vous que le mappage des identifiants soit effectué.

Il existe deux méthodes de flux de travail de mappage d'identifiants :

- Mappage d'ID basé sur des règles : méthode par laquelle vous utilisez des règles de correspondance pour traduire des données de première partie d'une source vers une cible dans un flux de travail de mappage d'ID.
- Mappage des identifiants des services fournisseurs : méthode par laquelle vous utilisez un service fournisseur pour traduire des données codées par des tiers d'une source vers une cible dans un flux de travail de mappage d'identifiants.

AWS Clean Rooms est actuellement prise en charge en LiveRamp tant que méthode de flux de travail de mappage des identifiants basée sur les services des fournisseurs de services. Vous devez être abonné à LiveRamp through AWS Data Exchange pour utiliser cette méthode. Pour plus d'informations, consultez la section [Abonnement à un service fournisseur AWS Data Exchange dans le](#) guide de Résolution des entités AWS l'utilisateur.

Table de mappage des identifiants

Une ressource AWS Clean Rooms qui permet soit des règles de correspondance de première partie, soit un transcodage d'identité multipartite dans le cadre d'une collaboration.

Une table de mappage d'identifiants est une référence à une table existante dans le AWS Glue Data Catalog. Il contient une [règle d'analyse des tables de mappage d'identifiants](#) qui détermine la manière dont les données peuvent être consultées. Les tables de mappage d'identifiants peuvent être associées à une ou plusieurs collaborations.

Règle d'analyse des tables de mappage d'identifiants

Type de règle d'analyse géré par AWS Clean Rooms et utilisé pour joindre des données d'identité disparates afin de faciliter les requêtes. Il est automatiquement ajouté aux [tables de mappage des identifiants](#) et ne peut pas être modifié. Elle hérite des comportements des autres règles d'analyse de la collaboration, à condition que ces règles d'analyse soient homogènes.

Workflow de mappage des identifiants

Une tâche de traitement de données qui mappe les données d'une source vers une cible en fonction de la [méthode de flux de travail de mappage d'identifiants](#) spécifiée. Il produit une [table de mappage des identifiants](#).

Espace de noms ID

Une ressource AWS Clean Rooms qui contient des métadonnées expliquant les ensembles de données sur plusieurs Comptes AWS et expliquant comment utiliser ces ensembles de données dans un flux de travail de [mappage d'identifiants](#).

Association d'espaces de noms ID

Association d'une ressource d'espace de noms d'identification qui vous aide à découvrir les entrées dans leur [flux de travail de mappage d'identifiants](#).

Tâche

Méthode permettant d'accéder aux tables configurées et de les analyser dans le cadre d'une collaboration à l'aide d'un ensemble pris en charge de fonctions, de classes et de variables.

AWS Clean Rooms prend actuellement en charge le type de PySpark tâche.

AWS Clean Rooms prend actuellement en charge l'exécution de tâches à l'aide d'un modèle d'PySpark analyse.

Règle d'analyse des listes

La restriction de requête qui autorise les requêtes qui génèrent une analyse attributaire au niveau des lignes du chevauchement entre cette table et les tables du membre qui peut effectuer la requête.

Prend en charge des cas d'utilisation tels que l'enrichissement et la création ou la suppression d'audience.

Les autres types de règles d'analyse sont [l'agrégation](#) et les règles [personnalisées](#).

Modèle Lookalike

Modèle de données d'un fournisseur de données de formation qui permet à un fournisseur de données de départ de créer un [segment similaire](#) des données du fournisseur de données de formation qui ressemble le plus à ses [données de départ](#).

Segment similaire

Sous-ensemble des données d'entraînement qui ressemble le plus aux [données de départ](#).

Membre

Un AWS client participant à une [collaboration](#).

Un membre est identifié à l'aide de son Compte AWS.

Tous les membres peuvent fournir des données.

Membre pouvant poser des questions

Le membre qui peut interroger des données dans le cadre de la [collaboration](#).

Un seul membre peut effectuer une requête par collaboration, et ce membre est immuable.

Un utilisateur administratif peut utiliser les autorisations Gestion des identités et des accès AWS (IAM) pour contrôler lequel de ses principaux IAM (tels que les utilisateurs ou les rôles) peut interroger des données dans le cadre de la collaboration. Pour de plus amples informations, veuillez consulter [Créez un rôle de service pour lire les données d'Amazon S3](#).

Membre capable d'exécuter des requêtes et des tâches

Le membre qui peut exécuter des requêtes et des tâches sur les données de la [collaboration](#).

Un seul membre peut exécuter des requêtes et des tâches par collaboration, et ce membre est immuable.

Un utilisateur administratif peut utiliser les autorisations Gestion des identités et des accès AWS (IAM) pour contrôler lesquels de ses principaux IAM (tels que les utilisateurs ou les rôles) peuvent exécuter des requêtes et des tâches dans le cadre de la collaboration. Pour de plus amples informations, veuillez consulter [Créez un rôle de service pour lire les données d'Amazon S3](#).

Membre pouvant recevoir les résultats

Un membre qui peut recevoir les résultats des requêtes. Un membre autorisé à recevoir des résultats définit les paramètres des résultats de requête pour la destination Amazon S3 et le format des résultats de la requête (CSV ou Parquet).

Pour les analyses effectuées avec le moteur d'analyse Spark, un membre qui peut recevoir les résultats indique également si les fichiers doivent être générés sous la forme d'un seul fichier ou de plusieurs fichiers.

Plusieurs membres peuvent recevoir des résultats dans le cadre d'une collaboration.

Membre payant les frais de calcul des requêtes

Membre chargé de payer les frais de calcul des requêtes.

Un seul membre est responsable du paiement des coûts de calcul des requêtes par collaboration, et ce membre est immuable.

Si le créateur de la collaboration n'a indiqué aucun membre payant les frais de calcul des requêtes, le [membre habilité à effectuer les requêtes](#) est le payeur par défaut.

Le membre qui paie les frais de calcul des requêtes reçoit une facture pour les requêtes exécutées dans le cadre de la collaboration.

Membre payant les frais de recherche et de calcul des tâches

Le membre chargé de payer les frais de recherche et de calcul des tâches.

Un seul membre est responsable du paiement des coûts de requête et de calcul des tâches par collaboration, et ce membre est immuable.

Si le créateur de la collaboration n'a indiqué aucun membre payant les frais de requête et de calcul des tâches, le [membre habilité à effectuer les requêtes](#) est le payeur par défaut.

Le membre qui paie les frais de requête et de calcul des tâches reçoit une facture pour les requêtes exécutées dans le cadre de la collaboration.

Membres

Ressource créée lorsqu'un [membre](#) rejoint une [collaboration](#).

Toutes les ressources que le membre associe à une collaboration font partie de l'adhésion ou sont associées à l'adhésion.

Seul le membre propriétaire de l'adhésion peut ajouter, supprimer ou modifier les ressources de cette adhésion.

Colonne étanche

Colonne protégée cryptographiquement pour une construction SELECT SQL.

Données sur les semences

Les données du fournisseur de données de départ, qui sont utilisées pour créer un [segment similaire](#). Les données de départ peuvent être fournies directement ou provenir des résultats d'une AWS Clean Rooms requête. Le résultat du segment similaire est un ensemble d'utilisateurs issu des données d'entraînement qui ressemble le plus aux utilisateurs initiaux.

Moteur d'analyse Spark

Une option d'analyse AWS Clean Rooms qui permet aux clients d'exécuter des requêtes complexes sur de grands ensembles de données stockés dans Amazon S3, Amazon Athena ou Snowflake à l'aide des fonctions SQL d'Apache Spark. Il prend également en charge PySpark l'analyse dans AWS Clean Rooms.

Lorsque vous créez une collaboration à l'aide de l'[CreateCollaborationAPI](#), la valeur du moteur d'analyse Spark est SPARK la suivante :

Query

Méthode permettant d'accéder aux tables configurées et de les analyser dans le cadre d'une collaboration, à l'aide d'un ensemble de fonctions, de classes et de variables pris en charge.

AWS Clean Rooms supporte actuellement le langage de requête SQL.

AWS Clean Rooms prend actuellement en charge l'exécution de requêtes SQL directes ou l'exécution de requêtes à l'aide d'un modèle d'analyse SQL.

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.