



Guía del usuario de

AWS Clean Rooms



AWS Clean Rooms: Guía del usuario de

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

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas registradas que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, conectados o patrocinados por Amazon.

Table of Contents

¿Qué es AWS Clean Rooms?	1
¿Es la primera vez que lo usa AWS Clean Rooms ?	2
¿Cómo funciona AWS Clean Rooms	2
Servicios relacionados	2
AWS servicios	2
Servicios de terceros	4
Acceder AWS Clean Rooms	4
Precios para AWS Clean Rooms	5
Facturación para AWS Clean Rooms	5
Reglas de análisis	6
Tipos de regla de análisis	7
Regla de análisis de agregación	9
Regla de análisis de lista	29
Regla de análisis personalizada	38
Regla de análisis de tablas de asignación de ID	45
AWS Clean Rooms Privacidad diferencial	56
Privacidad diferencial	57
Cómo funciona Differential Privacy in AWS Clean Rooms	57
Política de privacidad diferencial	58
Capacidades de SQL	60
Consejos y ejemplos de consultas SQL	92
Limitaciones	94
AWS Clean Rooms ML	95
Terminología de aprendizaje automático de AWS Clean Rooms	96
Cómo funciona AWS Clean Rooms ML con AWS los modelos	97
Cómo funciona AWS Clean Rooms ML con modelos personalizados	98
AWS modelos en Clean Rooms ML	100
Modelos personalizados en Clean Rooms ML	110
Computación criptográfica	118
Consideraciones	120
Tipos de archivo y de datos admitidos	122
Nombres de columnas	128
Tipos de columnas	129
Parameters	131

Indicadores opcionales	136
Consultas con C3R	139
Directrices	140
Con AWS Clean Rooms figuración	167
Inscríbase en AWS	167
Configure las funciones de servicio para AWS Clean Rooms	167
Creación de un usuario administrador	168
Creación de un rol de IAM para un miembro de la colaboración	169
Cree un rol de servicio para leer datos de Amazon S3	170
Cree un rol de servicio para leer los datos de Amazon Athena	174
Cree un rol de servicio para leer los datos de Snowflake	178
Cree un rol de servicio para leer el código de un bucket de S3 (rol de plantilla de PySpark análisis)	182
Cree un rol de servicio para escribir los resultados de un PySpark trabajo	185
Creación de un rol de servicio para recibir resultados	189
Configurar funciones de servicio para AWS Clean Rooms ML	193
Configure las funciones de servicio para un modelado similar	193
Configure las funciones de servicio para el modelado personalizado	208
Colaboraciones y suscripciones	224
Crear una colaboración	225
Crear una colaboración para consultas	225
Crear una colaboración para consultas y trabajos	237
Crear una colaboración para el modelado de aprendizaje automático	248
Crear una pertenencia y unirse a una colaboración	257
Editar colaboraciones	264
Edición del nombre y la descripción de la colaboración	264
Actualice el motor de análisis de colaboración	265
Desactiva el almacenamiento de registros	265
Edita la configuración de los registros de colaboración	266
Editar etiquetas de colaboración	267
Edición de etiquetas de pertenencia	268
Agrega un miembro nuevo	269
Edita las habilidades de los miembros existentes	269
Editar la configuración de aprobación automática de la colaboración	269
Editar las etiquetas de tablas asociadas	269
Edición de etiquetas de plantilla de análisis	270

Edición de las etiquetas de la política de privacidad diferencial	270
Solicitudes de cambio	271
Agrega un nuevo miembro a una colaboración	271
Actualice las capacidades de los miembros existentes	272
Editar la configuración de aprobación automática de la colaboración	273
Eliminar colaboraciones	274
Ver colaboraciones	275
Invitar a los miembros	276
Supervisión de miembros	276
Añadir miembros	277
Eliminar miembros	278
Abandonar una colaboración	279
Tablas de datos	281
Formatos de datos	282
Formatos de datos compatibles para los trabajos PySpark	282
Formatos de datos compatibles para consultas SQL	282
Tipos de datos compatibles	283
Tipos de compresión de archivos para AWS Clean Rooms	284
Cifrado del lado del servidor para AWS Clean Rooms	284
Apache Iceberg tablas	285
Tipos de datos admitidos para las tablas de Iceberg	286
Preparar tablas de datos	287
Preparación de tablas de datos en Amazon S3	287
Preparación de tablas de datos en Amazon Athena	290
Preparación de tablas de datos en Snowflake	292
Preparar tablas de datos cifrados	295
Paso 1: completar los requisitos previos	295
Paso 2: descargar el cliente de cifrado de C3R	296
Paso 3: (Opcional) Visualización de los comandos disponibles en el cliente de cifrado de C3R	297
Paso 4: generar un esquema de cifrado para un archivo tabular	297
Paso 5: crear una clave secreta compartida	305
Paso 6: guardar la clave secreta compartida en la variable de entorno	306
Paso 3: cifrar los datos	307
Paso 8: verificar el cifrado de datos	308
(Opcional) Crear un esquema (usuarios avanzados)	309

Descifrar tablas de datos	319
Tablas configuradas	321
Crear una tabla configurada	322
Origen de datos de Amazon S3	323
Fuente de datos de Amazon Athena	326
Fuente de datos de Snowflake	329
Agregación de una regla de análisis a una tabla configurada	333
Agregación de una regla de análisis de agregación a una tabla (flujo guiado)	334
Agregación de una regla de análisis de lista a una tabla (flujo guiado)	338
Agregación de una regla de análisis personalizada a una tabla (flujo guiado)	340
Agregación de una regla de análisis a una tabla (editor JSON)	345
Siguientes pasos	347
Asociar una tabla configurada a una colaboración	347
Presupuesto de acceso a los datos	348
Asocie una tabla configurada	348
Configuración de un presupuesto de acceso de datos	354
Visualización de un presupuesto de acceso a datos	355
Añadir un presupuesto de acceso a datos a una tabla asociada existente	356
Edición de un presupuesto de acceso a datos	357
Eliminar un presupuesto de acceso a datos	359
Agregación de una regla de análisis de colaboración a una tabla configurada	360
Configuración de la política de privacidad diferencial (opcional)	361
Visualización de los registros de uso de la privacidad diferencial	363
Edición de una política de privacidad diferencial	363
Eliminación de una política de privacidad diferencial	364
Visualización de los parámetros de privacidad diferencial calculados	365
Ver tablas y reglas de análisis	366
Edición de una tabla configurada	367
Editar etiquetas de tablas configuradas	367
Edición de la regla de análisis de tablas configurada	368
Eliminar la regla de análisis de tablas configurada	369
Columnas no permitidas de la tabla configurada	369
Editar asociaciones de tablas configuradas	373
Disociar tablas configuradas	373
AWS Entity Resolution in AWS Clean Rooms	375
Espacios de nombres de ID	376

Creación y asociación de un nuevo espacio de nombres de ID	376
Asociación de un espacio de nombres de ID existente	379
Edición de las asociaciones de espacios de nombres de ID	382
Desasociación de las asociaciones de espacios de nombres de ID	383
Tablas de asignación de ID	384
Creación y llenado de una nueva tabla de asignación de ID	385
Llenado de una tabla de asignación de ID existente	393
Edición de una tabla de asignación de ID	394
Eliminación de una tabla de asignación de ID	394
Plantillas de análisis	396
Plantillas de análisis SQL	396
Creación de una plantilla de análisis SQL	397
Revisión de una plantilla de análisis SQL	400
PySpark plantillas de análisis	401
Seguridad	402
Limitaciones	402
Prácticas recomendadas	404
Crear un script de usuario	405
Trabajar con parámetros en plantillas PySpark de análisis	408
Crear un entorno virtual (opcional)	413
Almacenamiento de un script de usuario y un entorno virtual en S3	414
Crear una plantilla PySpark de análisis	416
Revisión de una plantilla PySpark de análisis	420
Plantillas de PySpark análisis para la solución	423
Solución de problemas con el código	424
El trabajo de la plantilla de análisis no se inicia	425
El trabajo de la plantilla de análisis se inicia pero falla durante el procesamiento	426
La configuración del entorno virtual falla	427
Análisis	430
Ejecución de consultas SQL	430
Requisitos previos	431
Configuración de propiedades de Spark para consultas SQL	432
Consulta de tablas configuradas	433
Consulta de tablas de asignación de ID	453
Consulta de tablas configuradas mediante una plantilla de análisis	456
Consulta con el creador de análisis	473

Visualización del impacto de la privacidad diferencial	479
Vista de consultas recientes	479
Visualización de detalles de consultas	480
Ejecutando PySpark trabajos	481
Ejecutar un trabajo mediante una plantilla de análisis	482
Ver trabajos recientes	484
Consultar detalles del trabajo	485
Resultados del análisis	487
Recibir resultados de consultas	488
Recibir los resultados del trabajo	489
Edición de los ajustes predeterminados de los resultados de las consultas	490
Edición de los valores predeterminados de la configuración de los resultados del trabajo	492
Uso del resultado de la consulta en otros Servicios de AWS	493
Modelado personalizado	494
Generación de conjuntos de datos sintéticos con mayor privacidad	495
Consideraciones para la generación de datos sintéticos	496
Crear la colaboración y unirse a ella	499
Crear una colaboración para el aprendizaje automático	500
Unirse a una colaboración	504
Contribuir con datos de formación	505
Configuración de un algoritmo modelo	511
Asociar el algoritmo del modelo configurado	514
Creación de un canal de entrada de ML	519
Crear un modelo entrenado	540
Uso de la formación incremental	544
Uso de formación distribuida	548
Exportación de artefactos de modelos	552
Ejecutar la inferencia en un modelo entrenado	554
Modelado de aprendizaje automático para la formación de proveedores de datos	557
Importación de datos de entrenamiento	558
Crear un modelo similar	559
Configuración de un modelo similar	560
Asociar un modelo similar configurado	562
Actualización de un modelo similar configurado	562
Modelado de aprendizaje automático para proveedores de datos iniciales	564
Crear un segmento similar	564

Exportación de un segmento similar	582
Resolución de problemas	583
Una o más tablas a las que hace referencia la consulta no están accesibles para el rol de servicio asociado. El table/role propietario debe conceder al rol de servicio acceso a la tabla. ...	583
Uno de los conjuntos de datos subyacentes tiene un formato de archivo no compatible.	583
Los resultados de la consulta no son los esperados cuando se utiliza la computación criptográfica para Clean Rooms.	584
AWS Clean Rooms Spark SQL: Faltan datos de partición	584
Seguridad	585
Protección de datos	586
Cifrado en reposo	587
Cifrado en tránsito	588
Cifrado de datos subyacentes	588
Política de claves	588
Cómo utilizar roles vinculados a servicios	594
Permisos de roles vinculados al servicio para AWS Clean Rooms	595
Crear un rol vinculado a un servicio para AWS Clean Rooms	596
Edición de un rol vinculado a un servicio para AWS Clean Rooms	596
Eliminar un rol vinculado a un servicio para AWS Clean Rooms	596
Regiones compatibles para los roles vinculados al servicio AWS Clean Rooms	597
Retención de datos	597
Prácticas recomendadas	598
Mejores prácticas con AWS Clean Rooms	598
Mejores prácticas para utilizar las reglas de análisis en AWS Clean Rooms	599
Gestión de identidad y acceso	600
Público	601
Autenticación con identidades	601
Administración del acceso con políticas	603
¿Cómo AWS Clean Rooms funciona con IAM	606
Ejemplos de políticas basadas en identidades	611
AWS políticas gestionadas	615
Resolución de problemas	622
Prevención de la sustitución confusa entre servicios	624
Comportamientos de IAM para AWS Clean Rooms ML	627
Comportamientos de IAM para los modelos personalizados de Clean Rooms ML	631
Validación de conformidad	633

Resiliencia	633
Seguridad de la infraestructura	633
Seguridad de la red	634
AWS PrivateLink	634
Consideraciones	635
Creación de un punto de conexión de interfaz	635
Supervisión	636
Inicio de sesión de Analysis AWS Clean Rooms	637
Reciba registros de consultas y trabajos	638
Acciones recomendadas para los registros de consultas y trabajos	639
Monitorización detallada CloudWatch con AWS Clean Rooms	640
Tipos de métricas	640
Dimensiones de la métrica	642
¿Quién puede acceder a las métricas	642
Precios	642
CloudTrail registros	643
AWS Clean Rooms información en CloudTrail	643
Descripción de las entradas de los archivos de AWS Clean Rooms registro	644
Ejemplos de AWS Clean Rooms CloudTrail eventos	644
Integrándose en AWS Clean Rooms EDAs	648
AWS Clean Rooms eventos	649
Enrutamiento de AWS Clean Rooms eventos	654
Referencia detallada de los eventos	654
Etiquetado de asignación de costos	664
Recursos etiquetables para la asignación de costos	665
AWS CloudFormation recursos	666
AWS Clean Rooms y CloudFormation plantillas	666
Obtenga más información sobre CloudFormation	666
Cuotas	668
AWS Clean Rooms cuotas	668
AWS Clean Rooms límites de parámetros de recursos	684
Cuotas de aprendizaje automático de AWS Clean Rooms	684
Limpia las cuotas de la API Clean Rooms ML	717
Historial de revisión	723
Glosario	740
Regla de análisis de agregación	740

Reglas de análisis	740
Plantilla de análisis	740
Cliente de cifrado de S3	741
Columna de texto sin cifrar	741
Colaboración	741
Creador de la colaboración	741
Tabla configurada	742
Regla de análisis personalizada	742
Descifrado	742
Privacidad diferencial	743
Cifrado	743
Columna de huella digital	743
Método de flujo de trabajo de asignación de ID	743
tabla de asignación de ID	743
Regla de análisis de tablas de asignación de ID	744
Flujo de trabajo de asignación de ID	744
Espacio de nombres de ID	744
Asociación de espacios de nombres de ID	744
Trabajo	744
Regla de análisis de lista	745
Modelo parecido	745
Un segmento similar	745
Miembro	745
Miembro que puede realizar consultas	745
Miembro que puede ejecutar consultas y trabajos	746
Miembro que puede recibir los resultados	746
Miembro que paga los costos de computación de consultas	746
El miembro paga los costos de consulta y cálculo del trabajo	746
Pertenencia	747
Columna sellada	747
Datos iniciales	747
Motor de análisis Spark	747
Consultar	748
.....	dccxlix

¿Qué es AWS Clean Rooms?

AWS Clean Rooms le ayuda a usted y a sus socios a analizar y colaborar en sus conjuntos de datos colectivos para obtener nuevos conocimientos sin revelarse los datos subyacentes entre sí. AWS Clean Rooms es un espacio de trabajo de colaboración seguro, en el que puede crear sus propias salas limpias en cuestión de minutos y analizar sus conjuntos de datos colectivos en tan solo unos pasos. Tú eliges los socios con los que quieres colaborar, seleccionas sus conjuntos de datos y configuras los controles que mejoran la privacidad de esos socios.

Con él AWS Clean Rooms, puede colaborar con miles de empresas que ya los utilizan. AWS La colaboración no requiere mover los datos de otro proveedor de servicios en la nube AWS ni cargarlos en él. Cuando ejecuta consultas o trabajos, AWS Clean Rooms lee los datos de su ubicación original y aplica reglas de análisis integradas para ayudarle a mantener el control sobre esos datos.

AWS Clean Rooms proporciona controles de acceso a los datos y controles de soporte de auditoría integrados que puede configurar. Estos controles incluyen:

- [Reglas de análisis](#) para restringir las consultas SQL y proporcionar restricciones de salida.
- [Computación criptográfica Clean Rooms para](#) mantener los datos cifrados, incluso mientras se procesan las consultas, a fin de cumplir con las estrictas políticas de manejo de datos.
- [Registros de análisis](#) para revisar las consultas y los trabajos AWS Clean Rooms y ayudar a respaldar las auditorías.
- [Privacidad diferencial](#) para proteger a los usuarios de los intentos de identificación. AWS Clean Rooms La privacidad diferencial es una capacidad totalmente gestionada que protege la privacidad de sus usuarios con técnicas respaldadas matemáticamente y controles intuitivos que puede aplicar en unos pocos pasos.
- [AWS Clean Rooms ML](#) permite a dos partes identificar a usuarios similares en sus datos sin necesidad de compartir sus datos entre sí. La primera parte crea y configura un modelo similar a partir de sus datos de entrenamiento. A continuación, los datos iniciales se traen a la colaboración para crear un segmento similar que se parece a los datos de entrenamiento.

En el siguiente vídeo se explica más sobre AWS Clean Rooms.

[AWS Clean Rooms](#)

¿Es la primera vez que lo usa AWS Clean Rooms ?

Si es la primera vez que lo utiliza AWS Clean Rooms, le recomendamos que comience leyendo las siguientes secciones:

- [¿Cómo funciona AWS Clean Rooms](#)
- [Acceder AWS Clean Rooms](#)
- [Con AWS Clean Rooms figuración](#)
- [AWS Clean Rooms Glosario](#)

¿Cómo funciona AWS Clean Rooms

En AWS Clean Rooms, puedes crear una colaboración y añadir la Cuentas de AWS que quieras invitar, o bien crear una membresía para unirte a una colaboración a la que te hayan invitado. A continuación, vincule los recursos de datos necesarios para su caso de uso: tablas configuradas para datos de eventos, modelos configurados para el modelado de ML o espacios de nombres de ID para la resolución de entidades. Tiene la opción de crear o aprobar plantillas de análisis para acordar de antemano las consultas y los trabajos exactos que desea permitir en una colaboración. Por último, analiza los datos conjuntos mediante la ejecución de consultas o PySpark trabajos de SQL en las tablas configuradas, la resolución de entidades en las tablas de mapeo de ID o el uso de modelos de aprendizaje automático para generar segmentos de audiencia similares.

En el siguiente diagrama se muestra cómo AWS Clean Rooms funciona.

Servicios relacionados

AWS servicios

Los siguientes Servicios de AWS están relacionados con AWS Clean Rooms:

- Amazon Athena

Los miembros de Collaboration pueden almacenar los datos que incorporan AWS Clean Rooms como AWS Glue Data Catalog vistas en Amazon Athena. Para obtener más información, consulte los temas siguientes:

Para obtener más información, consulte los temas siguientes:

[Preparación de tablas de datos para consultas en AWS Clean Rooms](#)

[Creación de una tabla configurada: fuente de datos de Amazon Athena](#)

[¿Qué es Amazon Athena?](#) en la Guía del usuario de Amazon Athena

- CloudFormation

Cree los siguientes recursos en CloudFormation: colaboraciones, tablas configuradas, asociaciones de tablas configuradas y membresías

Para obtener más información, consulte [Creación de AWS Clean Rooms recursos con AWS CloudFormation](#).

- AWS CloudTrail

Úselo AWS Clean Rooms con CloudTrail los registros para mejorar el análisis de la Servicio de AWS actividad.

Para obtener más información, consulte [Registro de llamadas a la AWS Clean Rooms API mediante AWS CloudTrail](#).

- AWS Entity Resolution

Úselo AWS Clean Rooms con AWS Entity Resolution para realizar la resolución de entidades.

Para obtener más información, consulte [AWS Entity Resolution in AWS Clean Rooms](#).

- AWS Glue

Los miembros de Collaboration pueden crear AWS Glue tablas a partir de sus datos en Amazon S3 para usarlas en ellas AWS Clean Rooms.

Para obtener más información, consulte los temas siguientes:

[Preparación de tablas de datos para consultas en AWS Clean Rooms](#)

[¿Qué es AWS Glue?](#) en la Guía para desarrolladores de AWS Glue

- Amazon Simple Storage Service (Amazon S3)

Los miembros de Collaboration pueden almacenar los datos que introduzcan AWS Clean Rooms en Amazon S3.

Para obtener más información, consulte los temas siguientes:

[Preparación de tablas de datos para consultas en AWS Clean Rooms](#)

[Creación de una tabla configurada: fuente de datos de Amazon S3](#)

[¿Qué es Amazon S3](#) en la Guía del usuario de Amazon Simple Storage Service

- AWS Secrets Manager

Los miembros de la colaboración pueden crear secretos para acceder a los datos almacenados en Snowflake y leerlos.

Para obtener más información, consulte los temas siguientes:

[Cree un rol de servicio para leer los datos de Snowflake](#)

[Preparación de tablas de datos para consultas en AWS Clean Rooms](#)

[¿Qué es AWS Secrets Manager?](#) en la Guía del usuario de AWS Secrets Manager

Servicios de terceros

El siguiente servicio de terceros está relacionado con: AWS Clean Rooms

- Snowflake

Los miembros de Collaboration pueden almacenar los datos que introduzcan AWS Clean Rooms en un almacén de Snowflake.

Para obtener más información, consulte los temas siguientes:

[Preparación de tablas de datos para consultas en AWS Clean Rooms](#)

[Creación de una tabla configurada: fuente de datos de Snowflake](#)

Acceder AWS Clean Rooms

Puede acceder a AWS Clean Rooms través de las siguientes opciones:

- Directamente a través de la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms/>.
- Programáticamente a través de la AWS Clean Rooms API. Para obtener más información, consulte la [Referencia de la API de AWS Clean Rooms](#).

Precios para AWS Clean Rooms

Para obtener información acerca de los precios, consulte [Precios de AWS Clean Rooms](#).

Note

En el caso de los miembros de la colaboración que hayan asociado datos almacenados en Snowflake, el proveedor de almacenamiento de datos o el proveedor de la nube respectivos les cobrará tanto la salida como el cálculo de los datos cada vez que se ejecute una consulta que utilice datos almacenados en esas ubicaciones.

Facturación para AWS Clean Rooms

AWS Clean Rooms permite al creador de la colaboración designar qué miembro pagará los costes de cálculo de la consulta o del trabajo en la colaboración.

En la mayoría de los casos, el [miembro que puede realizar consultas](#) es también el [miembro que paga los costos de computación de las consultas](#). No obstante, si el miembro que puede realizar consultas y el miembro que paga los costos de computación de las consultas son diferentes, cuando el miembro que puede realizar consultas ejecuta las consultas en su propio recurso de pertenencia, se factura al recurso de pertenencia del miembro que paga los costos de computación de las consultas.

El miembro que paga los costes de cálculo de las consultas no ve ningún evento relacionado con las consultas que se estén ejecutando en su historial de CloudTrail eventos porque el pagador no es quien ejecuta las consultas ni es el propietario del recurso con el que se ejecutan las consultas. Sin embargo, el pagador ve los cargos generados en su recurso de membresía por todas las consultas realizadas por el miembro que puede ejecutar consultas en la colaboración.

Para obtener más información acerca de cómo crear una colaboración y configurar el miembro que paga los costos de computación de las consultas, consulte [Crear una colaboración](#).

Reglas de análisis en AWS Clean Rooms

Como parte de la habilitación de una tabla para utilizarla en AWS Clean Rooms el análisis de la colaboración, el miembro de la colaboración debe configurar una regla de análisis.

Una regla de análisis es un control que mejora la privacidad y que cada propietario de datos define en una tabla configurada. Una regla de análisis determina cómo se puede analizar la tabla configurada.

La regla de análisis es un control de nivel de cuenta en la tabla configurada (un recurso de nivel de cuenta) y que se aplica a cualquier colaboración a la que esté asociada la tabla configurada. Si no hay ninguna regla de análisis configurada, la tabla configurada se puede asociar a colaboraciones, pero no se puede consultar. Las consultas solo pueden hacer referencia a tablas configuradas con el mismo tipo de regla de análisis.

Para configurar una regla de análisis, primero debe seleccionar un tipo de análisis y, a continuación, especificar la regla de análisis. En ambos pasos, debe tener en cuenta el caso de uso que desea habilitar y cómo desea proteger los datos subyacentes.

AWS Clean Rooms aplica los controles más restrictivos en todas las tablas configuradas a las que se hace referencia en una consulta.

En los siguientes ejemplos se ilustran los controles restrictivos.

Example Control restrictivo: restricción de salida

- El colaborador A tiene una restricción de salida de 100 en la columna de identificador.
- El colaborador B tiene una restricción de salida de 150 en la columna de identificador.

Una consulta de agregación que haga referencia a ambas tablas configuradas requiere al menos 150 valores diferenciados de identificador dentro de una fila de salida para que se muestre en el resultado de la consulta. El resultado de la consulta no indica que los resultados se eliminan debido a la restricción de salida.

Example Control restrictivo: plantilla de análisis no aprobada

- El colaborador A ha permitido una plantilla de análisis con una consulta que hace referencia a tablas configuradas de los colaboradores A y B en su regla de análisis personalizada.
- El colaborador B no ha permitido la plantilla de análisis.

Como el colaborador B no ha permitido la plantilla de análisis, el miembro que puede realizar consultas no puede ejecutar dicha plantilla de análisis.

Tipos de regla de análisis

Existen tres tipos de reglas de análisis: de [agregación](#), de [lista](#) y [personalizadas](#). En las siguientes tablas se comparan los tipos de reglas de análisis. Cada tipo tiene dedicada una sección independiente en la que se describe la especificación de la regla de análisis.

Note

Existe un tipo de regla de análisis llamado regla de análisis de tablas de asignación de ID. Sin embargo, esta regla de análisis la administra AWS Clean Rooms y no se puede modificar. Para obtener más información, consulte [Regla de análisis de tablas de asignación de ID](#).

En las siguientes secciones se describen los casos de uso admitidos y los controles para cada tipo de regla de análisis.

Casos de uso admitidos

En las siguientes tablas se muestra un resumen comparativo de los casos de uso admitidos para cada tipo de regla de análisis.

Caso de uso	Agregación	Lista	Personalizada
Análisis admitidos	Consultas que agregan estadísticas mediante las funciones COUNT, SUM y AVG en dimensiones opcionales	Consultas que generan listas de nivel de fila de la superposición entre varias tablas	Cualquier análisis personalizado, siempre que la plantilla de análisis o el creador del análisis se hayan revisado y permitido

Caso de uso	Agregación	Lista	Personalizada
Casos de uso comunes	Análisis de segmentos, medición y atribución	Enriquecimiento, creación de segmentos	Atribución de primer toque, análisis incrementales, detección de audiencias
Constructos SQL	• Instrucciones JOIN: INNER JOIN • Funciones de agregado: COUNT/COUNT DISTINCT, SUM/SUM DISTINCT y AVG • Funciones escalares: subconjunto limitado	• Instrucciones JOIN: INNER JOIN • Funciones escalares: ninguna	La mayoría de las funciones SQL y constructos SQL están disponibles con el comando SELECT
Subconsultas y expresiones de tabla comunes () CTEs	No	No	Sí
Plantillas de análisis	No	No	Sí

Controles admitidos

En las siguientes tablas se muestra un resumen comparativo del modo en que cada tipo de regla de análisis protege los datos subyacentes.

Controlar	Agregación	Lista	Personalizada
Mecanismo de control	Controle cómo se pueden usar los datos de la tabla en una consulta	Controle cómo se pueden usar los datos de la tabla en una consulta	Controle qué consultas pueden ejecutarse en la tabla

Controlar	Agregación	Lista	Personalizada
	(por ejemplo, permitir COUNT y SUM de la columna hashed_email).	(por ejemplo, permitir el uso de la columna hashed_email solo en combinaciones).	(por ejemplo, permitir solo las consultas definidas en las plantillas de análisis "Consulta personalizada 1").
Técnicas de mejora de la privacidad integrada	• Coincidencia ciega • Agregación obligatoria • Umbral de agregación mín. >= • 2 Estructura de consulta predefinida	• Coincidencia ciega • Solapamiento obligatorio • Estructura de consulta predefinida • Análisis adicionales permitidos	• Privacidad diferencial • Columnas de salida no permitidas
Revisar la consulta antes de que pueda ejecutarse	No	No	Sí, con plantillas de análisis

Para obtener más información sobre las reglas de análisis disponibles en AWS Clean Rooms, consulte los siguientes temas.

- [Regla de análisis de agregación](#)
- [Regla de análisis de lista](#)
- [Regla de análisis personalizada en AWS Clean Rooms](#)

Regla de análisis de agregación

En AWS Clean Rooms, una regla de análisis de agregación genera estadísticas agregadas mediante las funciones COUNT, SUM y and/or AVG en dimensiones opcionales. Cuando la regla de análisis de agregación se agrega a una tabla configurada, permite al miembro que puede realizar la consulta ejecutar consultas en la tabla configurada.

La regla de análisis de agregación admite casos de uso tales como la planificación de campañas, el alcance mediático, la medición de la frecuencia y la atribución.

La estructura y sintaxis de consulta admitidas se definen en [Estructura y sintaxis de consultas de agregación](#).

Los parámetros de la regla de análisis, definidos en [Regla de análisis de agregación: controles de consulta](#), incluyen los controles de consulta y los controles de resultados de las consultas. Sus controles de consulta incluyen la posibilidad de imponer como requisito que una tabla configurada se una a al menos una tabla configurada propiedad del miembro que puede realizar la consulta, ya sea de forma directa o transitiva. Este requisito le permite asegurarse de que la consulta se ejecute en la intersección (INNER JOIN) entre su tabla y la de ellos.

Estructura y sintaxis de consultas de agregación

Las consultas en tablas que tienen una regla de análisis de agregación deben respetar la siguiente sintaxis.

```
--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}]] [,...]]
```

En la siguiente tabla se explica cada una de las expresiones enumeradas en la sintaxis anterior.

Expression	Definición	Ejemplos
<i>select_aggregate_function_expression</i>	Una lista separada por comas que contiene las siguientes expresiones: • <code>select_aggregation_function_expression</code> • <code>select_aggregate_expression</code>  Note Debe haber al menos una <code>select_aggregation_function_expression</code> en la <code>select_aggregate_expression</code>.	<pre>SELECT SUM(PRICE), user_segment</pre>
<i>select_aggregation_function_expression</i>	Una o más funciones de agregación admitidas aplicadas a una o más columnas. Solo se permiten columnas como argumentos de las funciones de agregación.  Note Debe haber al menos una <code>select_aggregation</code>	<pre>AVG(PRICE) COUNT(DISTINCT user_id)</pre>

Expression	Definición	Ejemplos
	<pre> function expression en la select_ag gregate_e xpression . </pre>	
<i>select_grouping_column_expression</i>	Expresión que puede contener cualquier expresión que utilice lo siguiente: • Nombres de columna de la tabla • Funciones escalares admitidas • Literales de cadena • Literales numéricos  Note select_aggregate_expression puede asignar un alias a las columnas con o sin el parámetro AS. Para obtener más información, consulte Referencia de SQL de AWS Clean Rooms.	TRUNC(timestampColumn) UPPER(campaignName)

Expression	Definición	Ejemplos
<i>table_expression</i>	Una tabla o combinación de tablas que conecta expresiones condicionales de unión con <code>join_condition</code> . <code>join_condition</code> devuelve un valor booleano. La <code>table_expression</code> admite: • Un tipo JOIN específico (INNER JOIN) • La condición de comparación de igualdad dentro de una <code>join_condition</code> (=) • Operadores lógicos (AND, OR).	<pre>FROM consumer_table INNER JOIN provider_ table ON consumer_table.ide ntifier1 = provider_ table.identifier1 AND consumer_table .identifier2 = provider_table.ide ntifier2</pre>

Expression	Definición	Ejemplos
<i>where_expression</i>	Una expresión condicional que devuelve un valor booleano. Puede constar de lo siguiente: • Nombres de columna de la tabla • Funciones escalares admitidas • Operadores matemáticos • Literales de cadena • Literales numéricos Las condiciones de comparación admitidas son (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL). Los operadores lógicos admitidos son (AND, OR). La where_expression es opcional.	<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>	Lista separada por comas de expresiones que cumplen con los requisitos de select_grouping_column_expression .	<pre>GROUP BY TRUNC(timestampColumn), UPPER(campaignName), segment</pre>

Expression	Definición	Ejemplos
<i>having_expression</i>	Una expresión condicional que devuelve un valor booleano. Tienen una función de agregación admitida aplicada a una sola columna (por ejemplo, SUM(price)) y se comparan con un literal numérico. Las condiciones admitidas son (=, >, <, <=, >=, <>, !=). Los operadores lógicos admitidos son (AND, OR). La having_expression es opcional.	HAVING SUM(SALES) > 500

Expression	Definición	Ejemplos
<i>order_by_expression</i>	Lista de expresiones separadas por comas que es compatible con los mismos requisitos definidos anteriormente en <code>select_aggregate_expression</code>. La <code>order_by_expression</code> es opcional.  Note <code>order_by_expression</code> admite los parámetros <code>ASC</code> y <code>DESC</code>. Para obtener más información, consulte los parámetros <code>ASC</code> y <code>DESC</code> en Referencia de SQL de AWS Clean Rooms.	<pre>ORDER BY SUM(SALES), UPPER(campaignName)</pre>

En cuanto a la estructura y sintaxis de las consultas de agregación, tenga en cuenta lo siguiente:

- No se admiten comandos SQL distintos de `SELECT`.
- No se admiten subconsultas ni expresiones de tabla comunes (por ejemplo, `WITH`).
- No se admiten operadores que combinen varias consultas (por ejemplo, `UNION`).
- No se admiten los parámetros `TOP`, `LIMIT` ni `OFFSET`.

Regla de análisis de agregación: controles de consulta

Los controles de consulta de agregación permiten controlar cómo se utilizan las columnas de la tabla para consultarla. Por ejemplo, puede controlar qué columna se usa para combinar, qué columna se puede contar o qué columna se puede usar en instrucciones WHERE.

En las secciones siguientes se explica cada uno de los controles.

Temas

- [Controles de agregación](#)
- [Controles de combinación](#)
- [Controles de dimensión](#)
- [Funciones escalares](#)

Controles de agregación

El uso de controles de agregación permite definir qué funciones de agregación se van a permitir y a qué columnas se deben aplicar. Las funciones de agregación se pueden usar en las expresiones SELECT, HAVING y ORDER BY.

Control	Definición	De uso
aggregateColumns	Columnas de tablas configuradas que se permite utilizar en las funciones de agregación.	aggregateColumns se puede usar dentro de una función de agregación en las expresiones SELECT, HAVING y ORDER BY. Algunas aggregateColumns también se pueden categorizar como joinColumn (definición disponible más adelante). La aggregateColumn dada no se puede categorizar también como dimension

Control	Definición	De uso
		Column (definición disponible más adelante).
function	Las funciones COUNT, SUM y AVG que se permite utilizar además de aggregate Columns .	La function se puede aplicar a una aggregate Columns que esté asociada a ella.

Controles de combinación

Se utiliza una cláusula JOIN para combinar filas de dos o más tablas en función de una columna relacionada entre ellas.

Puede usar los controles de unión para controlar cómo se puede unir su tabla a otras tablas `deltatable_expression`. AWS Clean Rooms solo admite INNERJOIN. INNERJOIN las declaraciones solo pueden usar columnas que se hayan clasificado explícitamente como una `joinColumn` en su regla de análisis, sujetas a los controles que usted defina.

INNER JOIN deben operar en una `joinColumn` de la tabla configurada y en una `joinColumn` de otra tabla configurada de la colaboración. Usted decide qué columnas de la tabla se pueden usar como `joinColumn`.

Cada condición de coincidencia de la cláusula ON debe utilizar la condición de comparación de igualdad (=) entre dos columnas.

Las condiciones de coincidencia múltiples dentro de una cláusula ON pueden ser:

- Combinación con el operador lógico AND
- Separación mediante el operador lógico OR

Note

Todas las condiciones de coincidencia JOIN deben coincidir con una fila de cada lado de JOIN. Todos los condicionales conectados por un operador lógico OR o AND también deben cumplir este requisito.

A continuación se muestra un ejemplo de consulta con un operador lógico AND.

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

A continuación se muestra un ejemplo de consulta con un operador lógico OR.

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

Control	Definición	De uso
<code>joinColumns</code>	Las columnas (si las hay) que se desea permitir que el miembro que puede realizar la consulta utilice en la instrucción INNER JOIN.	También se puede categorizar una <code>joinColumn</code> como <code>aggregateColumn</code> (consulte Controles de agregación). La misma columna no se puede usar como <code>joinColumn</code> y <code>dimensionColumns</code> a la vez (consulte más adelante). A menos que también se haya categorizado como <code>aggregateColumn</code>, una <code>joinColumn</code> no se puede usar en ninguna otra parte de la consulta que no sea INNER JOIN.
<code>joinRequired</code>	Controle si necesita una INNER JOIN con una tabla	Si habilita este parámetro, INNER JOIN es obligatorio.

Control	Definición	De uso
	configurada del miembro que puede realizar la consulta.	Si no habilita este parámetro, INNER JOIN es opcional. Presuponiendo que se habilite este parámetro, el miembro que puede realizar la consulta debe incluir una tabla de su propiedad en INNER JOIN. Debe combinar su tabla JOIN con la suya, ya sea de forma directa o transitiva (es decir, combinar su tabla con otra tabla que, a su vez, está combinada con la suya).

A continuación se muestra un ejemplo de transitividad.

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

Note

El miembro que puede realizar la consulta también puede usar el parámetro `joinRequired`. En ese caso, la consulta debe combinar su tabla con al menos otra tabla.

Controles de dimensión

Los controles de dimensión controlan la columna en la cual se pueden filtrar, agrupar o agregar las columnas de agregación.

Control	Definición	De uso
<code>dimensionColumns</code>	Las columnas (si las hay) que se permite que el miembro que puede realizar la consulta utilice en SELECT, WHERE, GROUP BY y ORDER BY.	Se puede usar una <code>dimensionColumn</code> en SELECT (<code>select_grouping_column_expression</code>), WHERE, GROUP BY y ORDER BY. La misma columna no puede ser <code>adimensionColumn</code>, <code>ajoinColumn</code> la vez <code>aggregateColumn</code> . and/or

Funciones escalares

Las funciones escalares controlan qué funciones escalares se pueden usar en las columnas de dimensión.

Control	Definición	De uso
<code>scalarFunctions</code>	Las funciones escalares que se pueden utilizar en <code>dimensionColumns</code> en la consulta.	Especifica las funciones escalares (si las hay) que se permite (por ejemplo, CAST) aplicar a <code>dimensionColumns</code> . Las funciones escalares no se pueden usar además de otras funciones ni dentro de otras funciones. Los argumentos de las funciones escalares pueden ser columnas, literales de cadena o literales numéricos.

Se admiten las siguientes funciones escalares:

- Funciones matemáticas: ABS, CEILING, FLOOR, LOG, LN, ROUND, SQRT
- Funciones de formato de tipo de datos: CAST, CONVERT, TO_CHAR, TO_DATE, TO_NUMBER, TO_TIMESTAMP
- Funciones de cadena: LOWER, UPPER, TRIM, RTRIM, SUBSTRING
 - En el caso de RTRIM, no se permiten conjuntos de caracteres personalizados para recortar.
- Expresiones condicionales: COALESCE
- Funciones de fecha EXTRACT, GETDATE, CURRENT_DATE, DATEADD
- Otras funciones: TRUNC

Para obtener más información, consulte [Referencia de SQL de AWS Clean Rooms](#).

Regla de análisis de agregación: controles de resultados de consulta

Los controles de resultados de consulta de agregación le permiten controlar qué resultados se devuelven especificando una o más condiciones que debe cumplir cada fila de salida. AWS Clean Rooms admite restricciones de agregación en forma de `COUNT (DISTINCT column) >= X`. Este formato requiere que cada fila agregue al menos X valores diferenciados de una selección de la tabla configurada (por ejemplo, un número mínimo de valores `user_id` diferenciados). Este umbral mínimo se aplica automáticamente, incluso si la consulta enviada en sí misma no utiliza la columna especificada. Se aplican de manera conjunta en cada tabla configurada de la consulta desde las tablas configuradas de cada miembro de la colaboración.

Cada tabla configurada debe tener al menos una restricción de agregación en su regla de análisis. Los propietarios de las tablas configuradas pueden añadir varios `columnName` y su `minimum` asociado y estos se aplicarán conjuntamente.

Restricciones de agregación

Las restricciones de agregación controlan qué filas de los resultados de la consulta se devuelven. Para incluirse en los resultados devueltos, una fila debe cumplir con el número mínimo especificado de valores diferenciados en cada columna especificada en la restricción de agregación. Este requisito se aplica incluso si la columna no se menciona explícitamente en la consulta o en otras partes de la regla de análisis.

Control	Definición	De uso
columnName	La aggregateColumn que se usa en la condición que debe cumplir cada fila de salida.	Puede tratarse de cualquier columna de la tabla configurada.
minimum	El número mínimo de valores diferenciados de la aggregateColumn asociada que debe tener la fila de salida (por ejemplo, COUNT DISTINCT) para que se devuelva en los resultados de la consulta.	El minimum debe tener al menos un valor de 2.

Estructura de la regla de análisis de agregación

El siguiente ejemplo muestra una estructura predefinida para una regla de análisis de agregación.

En el siguiente ejemplo, *MyTable* hace referencia a nuestra tabla de datos. Puedes reemplazar cada una *user input placeholder* con tu propia información.

```
{
  "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]
    },
  ],
}
```

Regla de análisis de agregación: ejemplo

El siguiente ejemplo demuestra cómo dos empresas pueden colaborar en el AWS Clean Rooms uso del análisis de agregación.

La empresa A tiene datos de clientes y de ventas. La empresa A está interesada en conocer la actividad de devolución de productos. La empresa B es uno de los minoristas de la empresa A y dispone de datos sobre devoluciones. La empresa B también tiene atributos de segmento sobre los clientes que son útiles para la empresa A (por ejemplo, compra de productos relacionados o uso del servicio de atención al cliente del minorista). La empresa B no quiere proporcionar información sobre atributos ni datos sobre devoluciones de los clientes por fila. La empresa B solo quiere habilitar un conjunto de consultas para que la empresa A obtenga estadísticas agregadas sobre los clientes que se superponen dentro de un umbral de agregación mínimo.

La empresa A y la empresa B deciden colaborar para que la empresa A pueda entender la actividad de devolución de productos y ofrecer mejores productos en la empresa B y en otros canales.

Para crear la colaboración y realizar un análisis de agregación, las empresas hacen lo siguiente:

1. La empresa A crea una colaboración y crea una pertenencia. La colaboración tiene a la empresa B como otro miembro de la colaboración. La empresa A habilita el registro de consultas en la colaboración y habilita el registro de consultas en su cuenta.
2. La empresa B crea una pertenencia en la colaboración. Habilita el registro de consultas en su cuenta.
3. La empresa A crea una tabla configurada de ventas.
4. La empresa A añade la siguiente regla de análisis de agregación a la tabla configurada de ventas.

```
{
  "aggregateColumns": [
    {
      "columnNames": [
        "identifier"
      ],
      "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: la empresa A quiere contar el número de clientes únicos que se superponen entre los datos de ventas y los datos de devoluciones. La empresa A también quiere sumar el número de purchases fabricados para compararlo con el número de returns.

joinColumns: la empresa A desea usar `identifier` para cotejar los clientes de los datos de ventas con los clientes de los datos de devoluciones. Esto ayudará a la empresa A a relacionar las devoluciones con las compras correctas. También ayudará a la empresa A a segmentar los clientes superpuestos.

dimensionColumns: la empresa A usa `dimensionColumns` para filtrar por un producto específico, comparar las compras y las devoluciones correspondientes a un periodo de tiempo

determinado, asegurarse de que la fecha de devolución sea posterior a la fecha del producto y ayudar a segmentar los clientes superpuestos.

`scalarFunctions`: la empresa A selecciona la función escalar CAST para ayudar a actualizar los formatos de tipo de datos si es necesario, basándose en la tabla configurada que la empresa A ha asociado a la colaboración. También añade funciones escalares para ayudar a dar formato a las columnas si es necesario.

`outputConstraints`: la empresa A establece restricciones de salida mínimas. No necesita restringir los resultados, ya que el analista puede ver los datos de la tabla de ventas por fila.

 Note

La empresa A no incluye `joinRequired` en la regla de análisis. Ofrece flexibilidad para que el analista consulte solo la tabla de ventas.

5. La empresa B crea una tabla configurada de devoluciones.
6. La empresa B añade la siguiente regla de análisis de agregación a la tabla configurada de devoluciones.

```
{
  "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: la empresa B permite a la empresa A sumar las `returns` para compararlas con el número de compras. Tienen al menos una columna agregada porque han habilitado una consulta agregada.

joinColumns: la empresa B permite a la empresa A combinar por `identifier` para cotejar los clientes de los datos de devoluciones con los clientes de los datos de ventas. Los datos de `identifier` son especialmente sensibles, y tenerlos como `joinColumn` garantiza que los datos no se incluyan nunca en una consulta.

`joinRequired`: la empresa B impone como requisito que las consultas sobre los datos de devoluciones se superpongan con los datos de ventas. No quieren permitir que la empresa A consulte todas las personas de su conjunto de datos. También acordaron esa restricción en su contrato de colaboración.

`dimensionColumns`: la empresa B permite a la empresa A filtrar y agrupar por `state`, `popularpurchases` y `customerserviceuser`, que son atributos únicos que podrían ayudar a realizar el análisis para la empresa A. La empresa B permite a la empresa A usar `returndate` para filtrar la salida por `returndate` que sea posterior a `purchasedate`. Con este filtrado, la salida es más precisa a la hora de evaluar el impacto del cambio de producto.

`scalarFunctions`: la empresa B habilita lo siguiente:

- `TRUNC` para las fechas
- `LOWER` y `UPPER` en caso de que `producttype` se introduzca en un formato distinto en sus datos
- `CAST` si la empresa A necesita convertir los tipos de datos de las ventas para que sean iguales a los tipos de datos de las devoluciones.

La empresa A no habilita otras funciones escalares porque no cree que sean necesarias para las consultas.

`outputConstraints`: la empresa B establece restricciones de salida mínimas en `hashedemail` para reducir en mayor medida la posibilidad de volver a identificar a los clientes. También añade una restricción de salida mínima en `producttype` para reducir en mayor medida la posibilidad de volver a identificar los productos específicos devueltos. Determinados tipos de productos podrían ser más dominantes en función de las dimensiones de la salida (por ejemplo, `state`). Sus restricciones de salida se aplicarán siempre, independientemente de las restricciones de salida que añada la empresa A a sus datos.

7. La empresa A crea una asociación a la tabla de ventas en la colaboración.
8. La empresa B crea una asociación a la tabla de devoluciones en la colaboración.
9. La empresa A realiza consultas, como las del ejemplo siguiente, para entender mejor la cantidad de devoluciones de la empresa B en comparación con el total de compras por ubicación en 2022.

```
SELECT
  companyB.state,
  SUM(companyB.returns),
  COUNT(DISTINCT companyA.hashedemail)
```

```
FROM
  sales companyA
  INNER JOIN returns companyB ON companyA.identifier = companyB.identifier
WHERE
  companyA.purchasedate BETWEEN '2022-01-01' AND '2022-12-31' AND
  TRUNC(companyB.returndate) > companyA.purchasedate
GROUP BY
  companyB.state;
```

10 La empresa A y la empresa B revisan los registros de consultas. La empresa B comprueba que la consulta se ajusta a lo acordado en el contrato de colaboración.

Solución de problemas relacionados con reglas de análisis de agregación

Utilice la información que se incluye aquí para diagnosticar y solucionar problemas frecuentes cuando utilice reglas de análisis de agregación.

Problemas

- [Mi consulta no ha devuelto ningún resultado](#)

Mi consulta no ha devuelto ningún resultado

Esto puede ocurrir cuando no hay resultados coincidentes o cuando los resultados coincidentes no cumplen uno o más umbrales de agregación mínimos.

Para obtener más información sobre los umbrales de agregación mínimos, consulte [Regla de análisis de agregación: ejemplo](#).

Regla de análisis de lista

En AWS Clean Rooms, una regla de análisis de listas genera listas a nivel de fila de la superposición entre la tabla configurada a la que se agrega y las tablas configuradas del miembro que puede realizar la consulta. El miembro que puede realizar consultas ejecuta consultas que incluyen una regla de análisis de lista.

El tipo de regla de análisis de listas admite casos de uso como el enriquecimiento y la creación de audiencia.

Para obtener más información sobre la estructura y la sintaxis de consulta predefinidas de esta regla de análisis, consulte [Estructura predefinida de la regla de análisis de lista](#).

Los parámetros de la regla de análisis de lista, que se define en [Regla de análisis de lista: controles de consulta](#), tienen controles de consulta. Estos controles de consulta incluyen la posibilidad de seleccionar las columnas que se pueden enumerar en la salida. La consulta debe tener al menos una combinación con una tabla configurada del miembro que puede realizar consultas, ya sea de forma directa o transitiva.

No hay controles de resultados de consulta como los que existen para la [regla de análisis de agregación](#).

Las consultas de lista solo pueden utilizar operadores matemáticos. No pueden usar otras funciones (por ejemplo, de agregación o escalares).

Temas

- [Estructura y sintaxis de las consultas de lista](#)
- [Regla de análisis de lista: controles de consulta](#)
- [Estructura predefinida de la regla de análisis de lista](#)
- [Regla de análisis de lista: ejemplo](#)

Estructura y sintaxis de las consultas de lista

Las consultas de las tablas que tienen una regla de análisis de lista deben respetar la siguiente sintaxis.

```
--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]
```

En la siguiente tabla se explica cada una de las expresiones enumeradas en la sintaxis anterior.

Expression	Definición	Ejemplos
<i>select_list_expression</i>	Una lista separada por comas que contiene al menos un nombre de columna de tabla. Es obligatorio un parámetro DISTINCT.  Note La <code>select_list_expression</code> pueden asignar un alias a las columnas con o sin el parámetro AS. Para obtener más información, consulte Referencia de SQL de AWS Clean Rooms.	SELECT DISTINCT segment
<i>table_expression</i>	Una tabla, o combinación de tablas, con <code>join_condition</code> para conectarla a <code>join_condition</code>. <code>join_condition</code> devuelve un valor booleano. La <code>table_expression</code> admite: • Un tipo JOIN específico (INNER JOIN) • Las condiciones de comparación de igualdad	<pre>FROM consumer_table INNER JOIN provider_table ON consumer_table.identifier1 = provider_table.identifier1 AND consumer_table.identifier2 = provider_table.identifier2</pre>

Expression	Definición	Ejemplos
	dentro de una <code>join_condition</code> (=) Operadores lógicos (AND, OR).	
<i>where_expression</i>	Una expresión condicional que devuelve un valor booleano. Puede constar de los siguientes elementos: - Nombres de columna de la tabla - Operadores matemáticos - Literales de cadena - Literales numéricos Las condiciones de comparación admitidas son (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL). Los operadores lógicos admitidos son (AND, OR). La <code>where_expression</code> es opcional.	<pre>WHERE state + '_' + city = 'NY_NYC'</pre> <pre>WHERE timestampColumn = timestampColumn2 - 14</pre>
<i>limit_expression</i>	Esta expresión debe tener adoptar un entero positivo. La <code>limit_expression</code> es opcional.	<pre>LIMIT 100</pre>

En cuanto a la estructura y sintaxis de las consultas de lista, tenga en cuenta lo siguiente:

- No se admiten comandos SQL distintos de SELECT.
- No se admiten subconsultas ni expresiones de tabla comunes (por ejemplo, WITH).
- No se admiten las cláusulas HAVING, GROUP BY ni ORDER BY.
- No se admite el parámetro OFFSET.

Regla de análisis de lista: controles de consulta

Los controles de consulta de lista le permiten controlar cómo se utilizan las columnas de su tabla a la hora de consultar esta última. Por ejemplo, puede controlar qué columna se usa para combinar, o qué columna se puede usar en una instrucción SELECT y en una cláusula WHERE.

En las secciones siguientes se explica cada uno de los controles.

Temas

- [Controles de combinación](#)
- [Controles de lista](#)

Controles de combinación

Los controles de combinación le permiten controlar cómo se puede combinar su tabla a otras tablas en la `table_expression`. AWS Clean Rooms solo admite INNER JOIN. En la regla de análisis de lista, se requiere al menos una INNER JOIN, y el miembro que puede realizar consultas debe incluir una tabla de su propiedad en la INNER JOIN. Esto significa que deben combinar su tabla con la suya, ya sea de forma directa o transitiva.

A continuación se muestra un ejemplo de transitividad.

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

Las instrucciones INNER JOIN solo pueden usar columnas que se hayan categorizado explícitamente como `joinColumn` en la regla de análisis.

La instrucción INNER JOIN debe operar en una `joinColumn` de su tabla configurada y en una `joinColumn` de otra tabla configurada de la colaboración. Usted decide qué columnas de la tabla se pueden usar como `joinColumn`.

Cada condición de coincidencia de la cláusula ON debe utilizar la condición de comparación de igualdad (=) entre dos columnas.

Las condiciones de coincidencia múltiples dentro de una cláusula ON pueden ser:

- Combinación con el operador lógico AND
- Separación mediante el operador lógico OR

 Note

Todas las condiciones de coincidencia JOIN deben coincidir con una fila de cada lado de JOIN. Todos los condicionales conectados por un operador lógico OR o AND también deben cumplir este requisito.

A continuación se muestra un ejemplo de consulta con un operador lógico AND.

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

A continuación se muestra un ejemplo de consulta con un operador lógico OR.

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

Control	Definición	De uso
<code>joinColumns</code>	Las columnas que se desea permitir que el miembro que	Una misma columna no se puede categorizar como <code>joinColumn</code> y <code>listColumn</code>

Control	Definición	De uso
	puede realizar consultas utilice en la instrucción INNER JOIN.	n a la vez (consulte Controles de lista). joinColumn no se puede usar en ninguna otra parte de la consulta que no sea INNER JOIN.

Controles de lista

Los controles de lista controlan las columnas que se pueden mostrar en el resultado de la consulta (es decir, que se pueden usar en la instrucción SELECT) o que se pueden usar para filtrar los resultados (es decir, que se pueden usar en la instrucción WHERE).

Control	Definición	De uso
listColumns	Las columnas que se permite que el miembro que puede realizar consultas utilice en SELECT y WHERE	Se puede usar una listColumn en SELECT y WHERE. Una misma columna no se puede usar como listColumn y joinColumn a la vez.

Estructura predefinida de la regla de análisis de lista

El siguiente ejemplo incluye una estructura predefinida que muestra cómo completar una regla de análisis de lista.

En el siguiente ejemplo, *MyTable* hace referencia a nuestra tabla de datos. Puede reemplazar cada una de ellas *user input placeholder* con su propia información.

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

Regla de análisis de lista: ejemplo

El siguiente ejemplo demuestra cómo dos empresas pueden colaborar en el AWS Clean Rooms uso del análisis de listas.

La empresa A dispone de datos de administración de las relaciones con los clientes (CRM). La empresa A desea obtener datos de segmento adicionales sobre sus clientes para conocer mejor a estos últimos y, posiblemente, utilizar los atributos como entrada en otros análisis. La empresa B tiene datos de segmento compuestos por atributos de segmento únicos que creó basándose en sus datos de primera parte. La empresa B desea proporcionar los atributos de segmento únicos a la empresa A solo para aquellos clientes que se superponen entre sus datos y los datos de la empresa A.

Las empresas deciden colaborar para que la empresa A pueda enriquecer los datos superpuestos. La empresa A es el miembro que puede realizar consultas y la empresa B es el colaborador.

Para crear una colaboración y ejecutar en ella un análisis de lista, las empresas hacen lo siguiente:

1. La empresa A crea una colaboración y crea una pertenencia. La colaboración tiene a la empresa B como un miembro más de la colaboración. La empresa A habilita el registro de consultas en la colaboración y habilita el registro de consultas en su cuenta.
2. La empresa B crea una pertenencia en la colaboración. Habilita el registro de consultas en su cuenta.
3. La empresa A crea una tabla configurada de CRM.
4. La empresa A agrega la regla de análisis a la tabla configurada por el cliente, como se muestra en el siguiente ejemplo.

```
{
  "joinColumns": [
    "identifier1",
    "identifier2"
  ],
  "listColumns": [
    "internalid",
    "segment1",
    "segment2",
    "customercategory"
  ]
}
```

`joinColumns`— La empresa A quiere utilizarla `hashedemail` and/or `thirdpartyid` (obtenida de un proveedor de identidad) para hacer coincidir los clientes a partir de los datos de CRM con los clientes a partir de los datos segmentados. Esto ayudará a garantizar que la empresa A relacione los datos enriquecidos con los clientes correctos. Tienen dos `JoinColumns` para mejorar potencialmente la tasa de coincidencia del análisis.

`listColumns`: la empresa A usa `listColumns` para obtener columnas enriquecidas junto con un `internalid` que utiliza en sus propios sistemas. Añade `segment1`, `segment2` y `customercategory` para, potencialmente, limitar el enriquecimiento a segmentos específicos mediante su uso en filtros.

5. La empresa B crea una tabla configurada de segmento.
6. La empresa B añade la regla de análisis a la tabla configurada por de segmento.

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

`joinColumns`: la empresa B permite a la empresa A combinarse con `identifier2` para relacionar a los clientes de los datos de segmento con los datos de CRM. La empresa A y la empresa B trabajaron con el proveedor de identidades para obtener un `identifier2` que se ajustara a la colaboración. No añadieron otras `joinColumns` porque creían que `identifier2` proporciona la tasa de coincidencia más alta y precisa y que no se precisaban otros identificadores para las consultas.

`listColumns`: la empresa B permite a la empresa A enriquecer sus datos con los atributos `segment3` y `segment4`, que son atributos únicos que han creado, recopilado y alineado en (con el cliente A) para formar parte del enriquecimiento de datos. Quieren que la empresa A obtenga estos segmentos para solaparlos a nivel de fila, ya que se trata de una colaboración de enriquecimiento de datos.

7. La empresa A crea una asociación a la tabla de CRM en la colaboración.
8. La empresa B crea una asociación a la tabla de segmento en la colaboración.

9. La empresa A ejecuta consultas, como la siguiente, para enriquecer los datos de clientes que se solapan.

```
SELECT companyA.internalid, companyB.segment3, companyB.segment4
INNER JOIN returns companyB
  ON companyA.identifier2 = companyB.identifier2
WHERE companyA.customercategory > 'xxx'
```

- 10 La empresa A y la empresa B revisan los registros de consultas. La empresa B comprueba que la consulta se ajusta a lo acordado en el contrato de colaboración.

Regla de análisis personalizada en AWS Clean Rooms

En AWS Clean Rooms, una regla de análisis personalizada es un nuevo tipo de regla de análisis que permite ejecutar consultas personalizadas en la tabla configurada. Las consultas SQL personalizadas siguen restringidas a tener solo el SELECT comando, pero pueden usar más construcciones SQL que las consultas de [agregación](#) y [lista](#) (por ejemplo, funciones de ventana, OUTER JOIN o subconsultas; consulte la [Referencia de AWS Clean Rooms SQL](#) para obtener una lista completa). Las consultas SQL personalizadas no tienen que seguir una estructura de consulta como las consultas de [agregación](#) y las consultas de [lista](#).

La regla de análisis personalizada admite casos de uso más avanzados que los que admite una regla de análisis de agregación y de lista, como análisis de atribución personalizado, análisis comparativo, análisis de incrementalidad y detección de audiencias. Esto se suma a un superconjunto de casos de uso compatibles con la regla de análisis de agregación y de lista.

La regla de análisis personalizado también admite privacidad diferencial. La privacidad diferencial es un marco matemáticamente riguroso para la protección de la privacidad de los datos. Para obtener más información, consulte [AWS Clean Rooms Privacidad diferencial](#). Al crear una plantilla de análisis, AWS Clean Rooms Differential Privacy comprueba la plantilla para determinar si es compatible con la estructura de consultas de uso general de Differential Privacy. AWS Clean Rooms Esta validación garantiza que no se cree una plantilla de análisis que no esté permitida con una tabla con protección de privacidad diferencial.

Para configurar la regla de análisis personalizada, los propietarios de los datos pueden optar por permitir que determinadas consultas personalizadas, almacenadas en [plantillas de análisis](#), se ejecuten en sus tablas configuradas. Los propietarios de los datos revisan las plantillas de análisis antes de añadirlas al control de análisis permitido en la regla de análisis personalizada. Las plantillas de análisis están disponibles y visibles solo en la colaboración en la que se crean (incluso si la tabla

está asociada a otras colaboraciones) y solo las puede ejecutar el miembro que pueda realizar consultas en esa colaboración.

Como alternativa, los miembros pueden optar por permitir que otros miembros (proveedores de consultas) creen consultas sin revisión. Los miembros añaden las cuentas de proveedores de consultas que los proveedores de consultas permitidos pueden controlar en la regla de análisis personalizada. Si el proveedor de consultas es el miembro que puede realizar la consulta, este puede ejecutar cualquier consulta directamente en la tabla configurada. Los proveedores de consultas también pueden crear consultas mediante la [creación de plantillas de análisis](#). Todas las consultas que hayan creado los proveedores de consultas pueden ejecutarse automáticamente en la tabla en todas las colaboraciones en las que Cuenta de AWS esté presente y la tabla esté asociada.

Los propietarios de los datos solo pueden permitir que las plantillas de análisis o las cuentas creen consultas, pero no ambas. Si el propietario de los datos lo deja en blanco, el miembro que puede realizar la consulta no puede ejecutar consultas en la tabla configurada.

Temas

- [Estructura predefinida de la regla de análisis personalizada](#)
- [Ejemplo de regla de análisis personalizada](#)
- [Regla de análisis personalizada con privacidad diferencial](#)

Estructura predefinida de la regla de análisis personalizada

El siguiente ejemplo incluye una estructura predefinida que muestra cómo completar una regla de análisis personalizada con la privacidad diferencial activada. El valor `userIdentifier` es la columna que identifica de forma exclusiva a los usuarios, como `user_id`. Cuando se activan dos o más tablas con la privacidad diferencial activada en una colaboración, AWS Clean Rooms requiere que configure la misma columna que la columna del identificador de usuario en ambas reglas de análisis para mantener una definición coherente de los usuarios en todas las tablas.

```
{
  "allowedAnalyses": ["ANY_QUERY"] | string[],
  "allowedAnalysisProviders": [],
  "differentialPrivacy": {
    "columns": [
      {
        "name": "userIdentifier"
      }
    ]
  }
}
```

```
]
}
}
```

Puede:

- Agregue una plantilla de análisis ARNs para permitir el control de los análisis. En este caso, el control `allowedAnalysisProviders` no está incluido.

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

- Agregue un miembro Cuenta de AWS IDs al `allowedAnalysisProviders` control. En este caso, se añade `ANY_QUERY` al control `allowedAnalyses`.

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

Ejemplo de regla de análisis personalizada

El siguiente ejemplo demuestra cómo dos empresas pueden colaborar en el AWS Clean Rooms uso de la regla de análisis personalizada.

La empresa A tiene datos de clientes y de ventas. La empresa A está interesada en conocer la incrementalidad de las ventas de una campaña publicitaria en el sitio de la empresa B. La empresa B tiene datos de visualizaciones y atributos de segmento que son útiles para la empresa (por ejemplo, el dispositivo utilizado para ver la publicidad).

La empresa A tiene una consulta de incrementalidad específica que quiere ejecutar en la colaboración.

Para crear una colaboración y ejecutar en ella un análisis personalizado, las empresas hacen lo siguiente:

1. La empresa A crea una colaboración y crea una pertenencia. La colaboración tiene a la empresa B como un miembro más de la colaboración. La empresa A habilita el registro de consultas en la colaboración y habilita el registro de consultas en su cuenta.

2. La empresa B crea una pertenencia en la colaboración. Habilita el registro de consultas en su cuenta.
3. La empresa A crea una tabla configurada de CRM.
4. La empresa A añade una regla de análisis personalizada vacía a la tabla de ventas configurada.
5. La empresa A asocia la tabla configurada de ventas a la colaboración.
6. La empresa B crea una tabla configurada de visualizaciones.
7. La empresa B agrega una regla de análisis personalizada vacía a la tabla configurada de visualizaciones.
8. La empresa B asocia la tabla configurada de visualizaciones a la colaboración.
9. La empresa A visualiza la tabla de ventas y la tabla de visualizaciones asociada a la colaboración y crea una plantilla de análisis, añadiendo la consulta de incrementalidad y el parámetro para el mes de la campaña.

```
{
  "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 La empresa A agrega la cuenta (por ejemplo, 444455556666) al control del proveedor de análisis permitido en la regla de análisis personalizada. Utilizan el control de proveedor de análisis permitido porque quieren permitir que las consultas que creen se ejecuten en su tabla configurada de ventas.

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

11 La empresa B ve la plantilla de análisis creada en la colaboración y revisa su contenido, incluidos la cadena de consulta y el parámetro.

12 La empresa B determina que la plantilla de análisis cumple con el caso de uso de incrementalidad y cumple con sus requisitos de privacidad en cuanto a la forma en que se puede consultar su tabla configurada de visualizaciones.

13 La empresa B agrega la plantilla de análisis ARN al control de análisis permitido en la regla de análisis personalizada de la tabla de visualizaciones. Utilizan el control de análisis permitido porque solo quieren permitir que la consulta de incrementalidad se ejecute en su tabla configurada de visualizaciones.

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

14 La empresa A ejecuta la plantilla de análisis y utiliza el valor del parámetro 05-01-2023.

Regla de análisis personalizada con privacidad diferencial

En AWS Clean Rooms, la regla de análisis personalizado admite la privacidad diferencial. La privacidad diferencial es un marco matemáticamente riguroso para la protección de la privacidad de los datos que le ayuda a proteger sus datos contra los intentos de reidentificación.

La privacidad diferencial permite el análisis agregado, como la planificación de campañas publicitarias, la post-ad-campaign medición, la evaluación comparativa en un consorcio de instituciones financieras y las pruebas A/B para la investigación sanitaria.

La estructura y sintaxis de consulta admitidas se definen en [Estructura y sintaxis de las consultas](#).

Regla de análisis personalizada con privacidad diferencial

Note

AWS Clean Rooms La privacidad diferencial solo está disponible para las colaboraciones en las que los datos se almacenan en Amazon S3.

Analice el [ejemplo de regla de análisis personalizado](#) que se presenta en la sección anterior. En este ejemplo se demuestra cómo puede utilizar la privacidad diferencial para proteger sus datos frente a intentos de reidentificación y, al mismo tiempo, permitir que su socio obtenga información crítica para la empresa a partir de sus datos. Supongamos que la empresa B, que tiene los datos de audiencia, quiere proteger sus datos con una privacidad diferencial. Para completar la configuración de privacidad diferencial, la empresa B realiza los siguientes pasos:

1. La empresa B activa la privacidad diferencial mientras añade una regla de análisis personalizada a la tabla configurada de visualizaciones. La empresa B selecciona `viewershipdata.hashemail` como columna de identificador de usuario.
2. La empresa B [añade una política de privacidad diferencial](#) a la colaboración para que su tabla de datos de audiencia esté disponible para su consulta. La empresa B selecciona la política predeterminada para completar rápidamente la configuración.

La empresa A, que desea comprender la incrementalidad de ventas de una campaña publicitaria en el sitio de la empresa B, utiliza la plantilla de análisis. Como la consulta es compatible con la [estructura](#) de consulta de uso general de AWS Clean Rooms Differential Privacy, la consulta se ejecuta correctamente.

Estructura y sintaxis de las consultas

Las consultas que contengan al menos una tabla y que tengan activada la privacidad diferencial deben seguir la siguiente sintaxis.

```

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 cuanto a la estructura y sintaxis de las consultas de privacidad diferencial, tenga en cuenta lo siguiente:

- No se admiten las subconsultas.
- Las expresiones de tabla comunes (CTEs) deberían emitir la columna de identificador de usuario si una tabla o un CTE incluyen datos protegidos por una privacidad diferencial. Los filtros, agrupaciones y agregaciones deben realizarse en el nivel de usuario.
- Final_select permite las funciones de agregación COUNT DISTINCT, COUNT, SUM, AVG y STDDEV.

Para obtener más información sobre qué palabras clave de SQL se admiten para una privacidad diferencial, consulte [Capacidades SQL de AWS Clean Rooms Differential Privacy](#).

Regla de análisis de tablas de asignación de ID

En AWS Clean Rooms, una regla de análisis de tablas de mapeo de ID no es una regla de análisis independiente. Este tipo de regla de análisis se administra mediante datos de identidad dispares AWS Clean Rooms y se utiliza para unirlos a fin de facilitar las consultas. Se agrega automáticamente a las tablas de asignación de ID y no se puede editar. Hereda los comportamientos de las demás reglas de análisis de la colaboración, siempre que esas reglas de análisis sean homogéneas.

La regla de análisis de la tabla de asignación de ID refuerza la seguridad en una tabla de asignación de ID. Impide que un miembro de la colaboración seleccione o inspeccione directamente la población no superpuesta entre los conjuntos de datos de los dos miembros mediante la tabla de asignación de ID. La regla de análisis de la tabla de asignación de ID se usa para proteger la información confidencial de la tabla de asignación de ID cuando se usa implícitamente en consultas con otras reglas de análisis.

Con la regla de análisis de la tabla de mapeo de ID, se AWS Clean Rooms impone una superposición en ambos lados de la tabla de mapeo de ID en el SQL expandido. Esto le permite realizar las siguientes tareas:

- Utilice la superposición de la tabla de asignación de ID en las declaraciones JOIN.

AWS Clean Rooms permite una INNER o una RIGHT unión en la tabla de mapeo de ID si respeta la superposición. LEFT Para proteger la información de mapeo confidencial, la tabla de mapeo de ID siempre debe estar en el lado inner «» de cualquier JOIN operación. Por ejemplo, las siguientes JOIN operaciones son válidas:

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

Las siguientes JOIN operaciones no son válidas:

- id_mapping_table LEFT JOIN table
- table RIGHT JOIN id_mapping_table

Esto evita la exposición de los registros de mapeo que no tienen las coincidencias correspondientes en su conjunto de datos. Permitir este tipo de operaciones podría revelar información confidencial sobre las asignaciones de datos de otros miembros de la colaboración.

- Utilice las columnas de la tabla de asignación en las declaraciones JOIN.

No puede usar las columnas de la tabla de asignación en las siguientes declaraciones: SELECT, WHERE, HAVING, GROUP BY o ORDER BY (a menos que se modifiquen las protecciones en la asociación del espacio de nombres del ID de origen o la asociación del espacio de nombres del ID de destino).

- En el SQL expandido, AWS Clean Rooms también es compatible con las versiones implícita OUTER JOIN y CROSSJOIN. JOIN Estas uniones no pueden satisfacer los requisitos de superposición. En su lugar, `requireOverlap` se AWS Clean Rooms utiliza para especificar en qué columnas se deben unir.

La estructura y la sintaxis de consulta admitidas se definen en [Estructura y sintaxis de consultas de la tabla de asignación de ID](#).

Los parámetros de la regla de análisis, definidos en [Controles de consulta de reglas de análisis de tablas de asignación de ID](#), incluyen los controles de consulta y los controles de resultados de las consultas. Sus controles de consulta incluyen la posibilidad de requerir la superposición de la tabla de asignación de ID en las instrucciones JOIN (es decir, `requireOverlap`).

Temas

- [Estructura y sintaxis de consultas de la tabla de asignación de ID](#)

- [Controles de consulta de reglas de análisis de tablas de asignación de ID](#)
- [Estructura predefinida de la regla de análisis de tablas de asignación de ID](#)
- [Regla de análisis de tablas de asignación de ID: ejemplo](#)

Estructura y sintaxis de consultas de la tabla de asignación de ID

Las consultas en tablas que tienen una regla de análisis de tabla de asignación de ID deben respetar la siguiente sintaxis.

```
--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
```

Tablas de colaboración

Las siguientes tablas representan las tablas configuradas que existen en una AWS Clean Rooms colaboración. La columna de ID en las tablas `cr_drivers_license` y `cr_insurance` representa una columna que coincide con la tabla de asignación de ID.

`cr_drivers_license`

id	nombre_del_conductor	estado_de_registro
1	Eduard	TX
2	Dana	MA
3	Gweneth	IL

cr_insurance

id	policyholder_email	número_política
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

tabla de asignación de ID

La siguiente tabla representa una tabla de asignación de ID existente que coincide con las tablas cr_drivers_license y cr_insurance. No todas las entradas serán válidas para ambas tablas de colaboración. IDs

cr_drivers_license_id	cr_insurance_id
1	a
2	null
3	b
null	c

La regla de análisis de la tabla de asignación de ID solo permite que las consultas se ejecuten en el conjunto de datos superpuestos, que tendría el siguiente aspecto:

cr_driver s_license_id	cr_insura nce_id	nombre_de l_conductor	estado_de _registro	policyhol der_email	número_po lítica
1	a	Eduard	TX	eduardo@i nternal.c ompany.com	17f9d04e- f5be-4426

					-bdc4-250 ed59c6529
3	b	Gweneth	IL	gwen@inte rnal.comp any.com	3f0092db- 2316-48a8 -8d44-09c f8f6e6c64

Consultas de ejemplo

Los siguientes ejemplos muestran ubicaciones válidas para las uniones de la tabla de asignación de ID:

```
-- 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
;
```

Consideraciones

Para la estructura y sintaxis de las consultas de tabla de asignación de ID, tenga en cuenta lo siguiente:

- No puede editarlo.
- Se aplica a la tabla de asignación de ID de forma predeterminada.
- Utiliza una asociación de espacio de nombres de ID de origen y destino dentro de la colaboración.
- La tabla de asignación de ID está configurada de forma predeterminada para proporcionar protecciones predeterminadas para la columna que proviene del espacio de nombres de ID. Puede modificar esta configuración para que la columna que proviene del espacio de nombres de ID (`sourceID` o `targetID`) se pueda permitir en cualquier parte de la consulta. Para obtener más información, consulte [Espacios de nombres de ID en AWS Clean Rooms](#).
- La regla de análisis de la tabla de asignación de ID hereda las restricciones de SQL de las demás reglas de análisis de la colaboración.

Controles de consulta de reglas de análisis de tablas de asignación de ID

Con los controles de consulta de tablas de mapeo de ID, controla cómo se utilizan las columnas de la tabla para consultarla. AWS Clean Rooms Por ejemplo, controla qué columnas se utilizan para unir y qué columnas deben superponerse. La regla de análisis de tablas de asignación de ID también incluye una funcionalidad que le permite proyectar `sourceID`, `targetID` o ambas sin necesidad de JOIN.

La tabla siguiente explica cada control.

Control	Definición	De uso
<code>joinColumns</code>	Las columnas que el miembro que puede realizar consultas puede usar en la instrucción INNER JOIN.	No puede usar <code>joinColumns</code> en ninguna otra parte de la consulta que no sea INNER JOIN. Para obtener más información, consulte Controles de combinación.
<code>dimensionColumns</code>	Las columnas (si las hay) que el miembro que puede realizar consultas puede utilizar en las instrucciones SELECT y GROUP BY.	Se puede usar una <code>dimensionColumn</code> en SELECT y GROUP BY. Puede aparecer una <code>dimensionColumn</code> como <code>joinKeys</code>. Solo puede utilizar <code>dimensionColumns</code> en la cláusula JOIN si la especifica entre paréntesis.
<code>queryConstraints:RequireOverlap</code>	Las columnas de la tabla de asignación de ID que se deben unir para que se pueda ejecutar la consulta.	Estas columnas se deben usar para UNIR la tabla de asignación de ID y una tabla de colaboración.

Estructura predefinida de la regla de análisis de tablas de asignación de ID

La estructura predefinida para una regla de análisis de tablas de asignación de ID incluye protecciones predeterminadas que se aplican a `sourceID` y `targetID`. Esto significa que la columna con las protecciones aplicadas se debe usar en las consultas.

Puede configurar la regla de análisis de la tabla de asignación de ID de las siguientes maneras:

- `sourceID` y `targetID` protegidos

En esta configuración, no se pueden proyectar `sourceID` y `targetID`. `sourceID` y `targetID` se deben usar en una instrucción `JOIN` cuando se hace referencia a la tabla de asignación de ID.

- Solo `targetID` protegido

En esta configuración, `targetID` no se puede proyectar. `targetID` se debe usar en una instrucción `JOIN` cuando se hace referencia a la tabla de asignación de ID. `sourceID` se puede usar en una consulta.

- Solo `sourceID` protegido

En esta configuración, `sourceID` no se puede proyectar. `sourceID` se debe usar en una instrucción `JOIN` cuando se hace referencia a la tabla de asignación de ID. `targetID` se puede usar en una consulta.

- Ningún `sourceID` o `targetID` protegido

En esta configuración, la tabla de asignación de ID no está sujeta a ninguna aplicación específica que pueda usarse en una consulta.

El siguiente ejemplo muestra una estructura predefinida para una regla de análisis de tablas de asignación de ID con las protecciones predeterminadas que se aplican a `sourceID` y `targetID`. En este ejemplo, la regla de análisis de la tabla de asignación de ID solo permite una instrucción `INNER JOIN` en la columna `sourceID` y en la columna `targetID`.

```
{
  "joinColumns": [
    "source_id",
    "target_id"
  ],
  "queryConstraints": [
    {
      "requireOverlap": {
        "columns": [
          "source_id",
          "target_id"
        ]
      }
    }
  ],
  "dimensionColumns": [] // columns that can be used in SELECT and JOIN
}
```

```
}
```

El siguiente ejemplo muestra una estructura predefinida para una regla de análisis de tablas de asignación de ID con las protecciones que se aplican a `targetID`. En este ejemplo, la regla de análisis de la tabla de asignación de ID solo permite una instrucción `INNER JOIN` en la columna `sourceID`.

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

El siguiente ejemplo muestra una estructura predefinida para una regla de análisis de tablas de asignación de ID con las protecciones que se aplican a `sourceID`. En este ejemplo, la regla de análisis de la tabla de asignación de ID solo permite una instrucción `INNER JOIN` en la columna `targetID`.

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



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

El siguiente ejemplo muestra una estructura predefinida para una regla de análisis de tablas de asignación de ID sin las protecciones que se aplican a sourceID o targetID. En este ejemplo, la regla de análisis de la tabla de asignación de ID permite una instrucción INNER JOIN en la columna sourceID y en la columna targetID.

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

Regla de análisis de tablas de asignación de ID: ejemplo

En lugar de escribir una larga declaración en cascada que haga referencia a la información de identificación personal (PII), por ejemplo, las empresas pueden utilizar la regla de análisis de la tabla de mapeo de identificadores para utilizar la transcodificación LiveRamp multipartita. El siguiente ejemplo demuestra cómo puede colaborar en el AWS Clean Rooms uso de la regla de análisis de tablas de mapeo de ID.

La empresa A es un anunciante que tiene datos de clientes y ventas, que se utilizarán como origen. La empresa A también realiza la transcodificación en nombre de las partes de la colaboración y aporta las LiveRamp credenciales.

La empresa B es un publicador que tiene datos de eventos, que se utilizarán como destino.

 Note

La empresa A o la empresa B pueden proporcionar las credenciales de LiveRamp transcodificación y realizar la transcodificación.

Para crear una colaboración que permita el análisis de tablas de asignación de ID en colaboración, las empresas hacen lo siguiente:

1. La empresa A crea una colaboración y crea una pertenencia. Se agrega la empresa B, que también crea una suscripción en la colaboración.
2. La empresa A asocia una fuente de espacio de nombres de ID existente o crea una nueva mediante la consola. AWS Entity Resolution AWS Clean Rooms

La empresa A crea una tabla configurada con sus datos de ventas y una columna vinculada a `sourceId` en la tabla de asignación de ID.

El origen del espacio de nombres de ID proporciona datos para la transcodificación.

3. La empresa B asocia un destino de espacio de nombres de ID existente o crea uno nuevo mediante la consola. AWS Entity Resolution AWS Clean Rooms

La empresa B crea una tabla configurada con sus datos de eventos y una columna vinculada a `targetId` en la tabla de asignación de ID.

El destino del espacio de nombres de ID no proporciona datos para transcodificar, solo metadatos relacionados con la configuración. LiveRamp

4. La empresa A descubre los dos espacios de nombres de ID asociados a la colaboración y crea y completa una tabla de asignación de ID.
5. La empresa A ejecuta una consulta en los dos conjuntos de datos mediante la unión en la tabla de asignación de ID.

--- 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 Privacidad diferencial

AWS Clean Rooms La privacidad diferencial le ayuda a proteger la privacidad de sus usuarios con una técnica respaldada matemáticamente que se implementa con controles intuitivos en unos pocos clics. Al tratarse de una funcionalidad totalmente gestionada, no es necesario tener experiencia previa en materia de privacidad diferencial para evitar que los usuarios vuelvan a identificarse. AWS Clean Rooms añade automáticamente una cantidad de ruido cuidadosamente calibrada a los resultados de las consultas en tiempo de ejecución para ayudar a proteger sus datos a nivel individual.

AWS Clean Rooms Differential Privacy admite una amplia gama de consultas analíticas y es ideal para una amplia variedad de casos de uso, en los que un pequeño error en los resultados de la consulta no comprometa la utilidad del análisis. Con él, sus socios pueden generar información crítica para la empresa sobre campañas publicitarias, decisiones de inversión, investigaciones clínicas y mucho más, todo ello sin necesidad de ninguna configuración adicional por parte de sus socios.

AWS Clean Rooms Differential Privacy protege contra errores de conversión excesivos o inválidos que utilizan funciones escalares o símbolos de operadores matemáticos de forma malintencionada.

Para obtener más información sobre la privacidad AWS Clean Rooms diferencial, consulte los siguientes temas.

Temas

- [Privacidad diferencial](#)
- [Cómo funciona Differential Privacy in AWS Clean Rooms](#)
- [Política de privacidad diferencial](#)
- [Capacidades SQL de AWS Clean Rooms Differential Privacy](#)
- [Consejos y ejemplos de consultas sobre privacidad diferencial](#)
- [Limitaciones de la AWS Clean Rooms privacidad diferencial](#)

Privacidad diferencial

La privacidad diferencial solo permite obtener información agregada y oculta la contribución de los datos de cualquier persona a esa información. La privacidad diferencial protege los datos de la colaboración para que el miembro pueda obtener resultados al aprender sobre una persona específica. Sin una privacidad diferencial, el miembro que puede recibir los resultados puede intentar deducir los datos de un usuario individual añadiendo o quitando registros sobre una persona y observando la diferencia en los resultados de las consultas.

Cuando se activa la privacidad diferencial, se añade una cantidad específica de ruido a los resultados de la consulta para ocultar la contribución de los usuarios individuales. Si el miembro que puede recibir los resultados intenta observar la diferencia en los resultados de la consulta después de eliminar los registros sobre una persona de su conjunto de datos, la variabilidad del resultado de la consulta ayuda a impedir la identificación de los datos de la persona. AWS Clean Rooms Differential Privacy utiliza el [SampCert](#) muestreador, una implementación comprobada y correcta desarrollada por AWS.

Cómo funciona Differential Privacy in AWS Clean Rooms

El flujo de trabajo para activar la privacidad diferencial AWS Clean Rooms requiere los siguientes pasos adicionales al [completar el flujo de trabajo para AWS Clean Rooms](#):

1. La privacidad diferencial se activa al añadir una [regla de análisis personalizada](#).
2. [Puede configurar la política de privacidad diferencial de la colaboración](#) para que las tablas de datos protegidas con privacidad diferencial estén disponibles para su consulta.

Tras completar estos pasos, el miembro que pueda realizar la consulta podrá empezar a ejecutar consultas sobre datos protegidos por la privacidad diferencial. AWS Clean Rooms devuelve resultados que cumplen con la política de privacidad diferencial. AWS Clean Rooms Differential Privacy registra el número estimado de consultas pendientes que puede ejecutar, de forma similar al indicador de combustible de un automóvil que muestra el nivel de combustible actual del automóvil. El número de consultas que puede ejecutar el miembro que puede realizar la consulta está limitado por los parámetros Presupuesto de privacidad y el Ruido añadido por consulta establecidos en [Política de privacidad diferencial](#).

Consideraciones

Cuando utilices la privacidad diferencial en AWS Clean Rooms, ten en cuenta lo siguiente:

- El miembro que puede recibir los resultados no puede usar la privacidad diferencial. Configuraré una regla de análisis personalizada con la privacidad diferencial desactivada para las tablas configuradas.
- El miembro que puede realizar consultas no puede unir tablas de dos o más proveedores de datos cuando ambos tienen activada la privacidad diferencial.

Política de privacidad diferencial

La política de privacidad diferencial controla cuántas funciones de agregación puede ejecutar el miembro que puede realizar la consulta en una colaboración. El Presupuesto de privacidad define un recurso común y limitado que se aplica a todas las tablas durante una colaboración. El Ruido agregado por consulta determina la velocidad a la que se agota el presupuesto de privacidad.

Es necesaria una política de privacidad diferencial para que las tablas con protección de privacidad diferencial estén disponibles para su consulta. Se trata de un paso único en una colaboración e incluye dos entradas:

- Presupuesto de privacidad: cuantificado en términos de épsilon, el presupuesto de privacidad controla el nivel de protección de la privacidad. Se trata de un recurso común y limitado que se aplica a todas las tablas protegidas con privacidad diferencial durante la colaboración, ya que el objetivo es conservar la privacidad de los usuarios, cuya información puede estar presente en varias tablas.

El presupuesto de privacidad se consume cada vez que se ejecuta una consulta en las tablas. Cuando el presupuesto de privacidad se agote por completo, el miembro de la colaboración que puede realizar consultas no podrá ejecutar consultas adicionales hasta que se aumente o actualice. Al establecer un presupuesto de privacidad mayor, el miembro que puede recibir los resultados puede reducir su incertidumbre acerca de las personas que contienen los datos. Elija un presupuesto de privacidad que equilibre sus requisitos de colaboración con sus necesidades de privacidad y tras consultar con los responsables de la toma de decisiones empresariales.

Si tiene previsto incorporar nuevos datos a la colaboración de forma periódica, puede seleccionar Actualización mensual del presupuesto de privacidad para crear automáticamente un nuevo presupuesto de privacidad cada mes natural. Al elegir esta opción, se mostrarán cantidades arbitrarias de información sobre las filas de datos cuando se consulten repetidamente durante las actualizaciones. Evite elegir esta opción si se van a consultar repetidamente las mismas filas entre las actualizaciones del presupuesto de privacidad.

- El Ruido agregado por consulta se mide en función del número de usuarios cuyas contribuciones desea ocultar. Este valor determina el ritmo al que se agota el presupuesto de privacidad. Un valor de ruido mayor reduce el ritmo al que se agota el presupuesto de privacidad y, por lo tanto, permite ejecutar más consultas sobre los datos. Sin embargo, esto debe equilibrarse con la publicación de información de datos menos precisa. Al establecer este valor, tenga en cuenta la precisión deseada para obtener información sobre la colaboración.

Puede usar la política de privacidad diferencial predeterminada para completar rápidamente la configuración o personalizar su política de privacidad diferencial según su caso de uso. AWS Clean Rooms Differential Privacy proporciona controles intuitivos para configurar la política. AWS Clean Rooms Differential Privacy le permite obtener una vista previa de la utilidad en términos del número de agregaciones posibles en todas las consultas de sus datos y estimar el número de consultas que se pueden ejecutar en una colaboración de datos.

Puede utilizar los ejemplos interactivos para comprender cómo los distintos valores de Presupuesto de privacidad y Ruido añadido por consulta afectarían a los resultados de los distintos tipos de consultas SQL. En general, es necesario equilibrar sus necesidades de privacidad con el número de consultas que desea permitir y la precisión de esas consultas. Un menor Presupuesto de privacidad o un mayor Ruido añadido por consulta puede proteger mejor la privacidad de los usuarios, pero proporciona información menos significativa a sus socios de colaboración.

Si aumenta el Presupuesto de privacidad y mantiene igual el parámetro Ruido añadido por consulta, el miembro que pueda realizar la consulta podrá ejecutar más agregaciones en sus tablas durante la colaboración. Puede aumentar el Presupuesto de privacidad en cualquier momento durante la colaboración. Si reduce el Presupuesto de privacidad y mantiene igual el parámetro Ruido añadido por consulta, el miembro que pueda realizar la consulta podrá ejecutar menos agregaciones. No puede reducir el Presupuesto de privacidad una vez que el miembro que puede realizar la consulta haya empezado a analizar sus datos.

Si aumenta el Ruido añadido por consulta y mantiene igual la entrada de Presupuesto de privacidad, el miembro que pueda realizar la consulta podrá ejecutar más agregaciones en sus tablas durante la colaboración. Si reduce el Ruido añadido por consulta y mantiene igual la entrada de Presupuesto de privacidad, el miembro que pueda realizar la consulta podrá ejecutar menos agregaciones. Puede aumentar o disminuir el Ruido agregado por consulta en cualquier momento durante la colaboración.

La política de privacidad diferencial se gestiona mediante las acciones de la API de la plantilla de presupuesto de privacidad.

Capacidades SQL de AWS Clean Rooms Differential Privacy

AWS Clean Rooms Differential Privacy utiliza una estructura de consultas de uso general para admitir consultas SQL complejas. Las plantillas de análisis personalizadas se validan según esta estructura para garantizar que se puedan ejecutar en tablas protegidas por privacidad diferencial. En la siguiente tabla se indican las funciones compatibles. Para obtener más información, consulte [Estructura y sintaxis de las consultas](#).

Expresión

relación

SELECT

análisis

Spark

admite

estructura

iones

SQL

Función

ANALYZE

función

APPROXIMA

función

COUNT

PERCENTIL

COUNT

STDDEV

AVG

funciones

SUM

COUNT

función

DISTINCT

función

MAX

función

MEDIAN

El **expresión**

relación

de **SELECT**

análisis

Sparknes

admite

de **estructura**

iones

SQL

el **Función**

de **MIN**

generar

datos **PERCENTIL**

con **CONT**

registros

• **Funciones**

a **STDDEV_SA**

nivel

de **MP**

y **STDDEV_PO**

Debe

escribir

• **Funciones**

la **SUM**

expresión

y **SELECT**

en **SUM**

las **DISTINCT**

• **Funciones**

que **VAR_SAMP**

CTES **VAR_POP**

utilicen

el **SELECT**

userIdent

ifierColu

mn...'

formato.

Expresión

relación

SELECT

análisis

Spark

admite

CTEs

iones

SQL

CTEs

WRITE

subconsultas

la

condición

cláusula

WITH

el

CTEs

uso

de

tablas

con

protección

n

de

privacidad

d

diferenci

al

debe

generar

datos

con

registros

a

El **análisis de expresiones** es una capacidad de **Spark** que admite **CTEs** (instrucciones SQL) a nivel de usuario. Debe escribir la expresión **SELECT** en las que **CTEs** utilicen el ``SELECT`` `userIdentifierColumn...` formato.

El **SELECT**

es una

instrucción

de **SQL**

que se utiliza

para

consultar

datos

de

una **base de datos**

relacional

o **data warehouse**

de cualquier

tipo de

datos, como

datos

de

un **anexo**

o **archivo**

de

datos

de

un **anexo**

o

de

un **anexo**

o

de

un

de

un

de

un

de

un

de

un

Expresión

relación

SELECT

análisis

Spark

admite

OTR

iones

SQL

subconsul

ta

que

haga

referenci

a

a

relacione

s

de

privacida

d

diferenci

ales

únicament

e

en

las

cláusulas

FROM

y

JOIN.

El operador

relación

de SELECT

análisis

Sparklines

admite

de instrucciones

iones

SQL

SQL INNER

de las

con LEFT

la JOIN

condición

• SEMIUNIÓN

de IZQUIERDA

que

• ANTIUNIÓN

solo IZQUIERDA

se

• RIGHT

admiten

las JOIN

funciones

JOIN

que

Operador

SQL JOIN

equivale

join

• CROSS

s JOIN

en

las

columnas

del

identific

ador

de

usuario

El **análisis**

relación

de **SELECT**

análisis

Spark

admite

de **instrucciones**

iones

SQL

y

son

obligator

ias

cuando

se

consultan

dos

o

más

tablas

con

la

privacida

d

diferenci

al

activada.

Asegúrese

de

que

las

condicion

es

obligator

ias

de

El propietario de la columna de análisis de Spark no admite estas instrucciones SQL equijoin sean correctas.

Confirme que el propietario de la tabla haya configurado la misma columna de identificación de usuario en todas las

El **análisis**

relación

de **SELECT**

análisis

Spark

admite

instrucciones

iones

SQL

tablas

para

que

la

definición

n

de

usuario

signa

siendo

coherente

en

todas

las

tablas.

Las

funciones

CROSS

JOIN

no

se

admiten

cuando

se

combinan

dos

Expresión

relator

SELECT

análisis

Spark

admite

OTR

iones

SQL

o

más

relacione

s

con

la

privacida

d

diferenci

al

activada.

UNION,pre

UNION tibl

ALL,

INTERSECT

miento

EXCEPTO

||

MINUS

(son

sinónimos

)

Expresión

relación
cláusula
de
SELECT
análisis
Spark
admite
instrucciones
SQL

Funciones

compatibil
integrado
es
con
• Función
la
de
condición
de ventana
de
AVG
que
1a Función
de
columna
de ventana
de
COUNT
identific
acción
de
de
usando
de
CUME_DIST
la Función
cláusula
de
de ventana
partición
DENSE_RANK
de
la
• Función
función
de
de
ventana
ventana

Expresión

relación

SELECT

análisis

Sparklines

admite

OTF Estruct

iones

SQL

FIRST_VAL

obulga

Función

cuando

se ventana

consulte

una

• Función

relación

con ventana

la LAST_VALU

privacida

• Función

diferenci

al ventana

activada.

LEAD

• Funciones

de

ventana

MAX

• Funciones

de

ventana

MEDIAN

• Funciones

de

El **análisis**

relación

de **SELECT**

análisis

Spark

admite

de **instrucciones**

iones

SQL

ventana

MIN

- Función de ventana **DENSE_NTH**

- Las funciones de ventana **STDDEV_SAMP** y **STDDEV_POP** (**STDDEV_SAMP** y **STDDEV** son sinónimos)

- Funciones de ventana **SUM**

Expresión

relación

de

SELECT

función

admite

instrucciones

SQL

- Las funciones de ventana VAR_SAMP y VAR_POP (VAR_SAMP y VARIANCE son sinónimos)

Funciones

de

clasificación

ción

- Función de ventana DENSE_RANK
- Función de

Expresión

relación

SELECT

análisis

Sparklines

admite

OTF Estruct

iones

SQL

ventana

NTILE

- Función de ventana PERCENT_RANK
- Función de ventana RANK
- Función de ventana ROW_NUMBER

~~El~~ Expresión

~~relator~~ cláusula

~~de~~ SELECT

~~análisis~~ análisis

~~Spark~~ Spark

~~admite~~ admite

~~OT~~ Estructura

iones

SQL

~~El~~ Expresión

~~son~~ condición

~~condición~~ condición

~~as~~ CASE

- Expresión
COALESCE
- Funciones
GREATEST
y
LEAST
- Funciones
NVL
y
COALESCE
- NVL2
función
- Función
NULLIF

Expresión

relación

de

SELECT

análisis

Spark

admite

instrucciones

SQL

Condición

entre

sempre

En

pueden

usar

porque

requieren

subconsultas

de

Se

coinciden

admiten

todas

las

patrones

demás.

Condición

es

de

rango

BETWEEN

Condición

nula

Expresión

relación

SELECT

análisis

Sparklines

admite

instrucc

iones

SQL

Flujos de trabajo

de

compatibil

yes

hora

en

transacci

ones

- Operador de concatenación
- Funciones
ADD_MONTHS
- Función
CONVERT_TIMEZONE
- Función
CURRENT_DATE
- Función
DATEADD
- Función
DATEDIFF

Expresión

relación

SELECT

análisis

Sparklines

admite

OTI Estructura

iones

SQL

- Funciones
DATE_PART
- Función
DATE_TRUNC
- Función
EXTRACT
- Función
TO_TIMESTAMP
- Partes
de
fecha
para
funciones
de
fecha
o
marca
temporal

~~El~~ Expresión

~~relación~~ relator

~~de~~ SELECT

~~análisis~~ análisis

~~Spark~~ Sparklines

~~admite~~ admite

~~OT~~ OTRAS

iones

SQL

~~Flujos~~ Flujos

~~com~~ operador

~~compañía~~ compañía

~~es~~ esación)

- Función
BTRIM
- Función
CHAR_LENGTH
- Función
CHARACTER
_LENGTH
- Función
CONCAT
- Funciones
LEFT
y
RIGHT
- Función
LEN
- Función
LENGTH
- Función
LOWER

Expresión

relación

SELECT

análisis

Sparklines

admite

OTI Estructura

iones

SQL

- Funciones
LPAD
y
RPAD
- Función
LTRIM
- Funciones
POSITION
- Función
REGEXP_CO
UNT
- Función
REGEXP_IN
STR
- Función
REGEXP_RE
PLACE
- Función
REGEXP_SU
BSTR
- Función
REPEAT
- Función
REPLACE

Expresión

relación

SELECT

análisis

Spark

admite

instrucciones

iones

SQL

- Función
REVERSE
- Función
RTRIM
- Función
SPLIT_PAR
T
- Función
SUBSTRING
- Función
TRANSLATE
- Funciones
TRIM
- Función
UPPER

Expresión

relación

de

SELECT

análisis

Spark

admite

instrucciones

SQL

Funciones

de

compatible

de

función

tipo

de

datos

• Cadenas

de

formatos

de

fecha

y

hora

• Cadenas

de

formatos

numéricos

Expresión

retosula

DELECT

fantalasis

Spanknes

~~a)~~dmite

Costrucc

iones

SQL

Fluent Energy

ISO 811

~~CONFIDENTIAL~~
ALLS DECRY

esPT

- ENCODE

- ENCODE
- DECODE

- MD5
función

- SHA1
función

- SHA2
función

- XX_
HASH64

Todos los

es:em

***Compatible**

~~SES~~

matemátic



Expresión

relación

SELECT

análisis

Sparklines

admite

instrucciones

iones

SQL

Funciones

relación

función

ACOS

- Función

ASIN

- Función

ATAN

- ATAN2

función

- Función

CBRT

- Función

CEILING

(o

CEIL)

- Función

COS

- Función

COT

- Función

DEGREES

- Función

LTRIM

Expresión

relación

SELECT

función

Sparklines

admite

instrucciones

iones

SQL

- Función EXP
- Función FLOOR
- Función LN
- Función LOG
- Función MOD
- Función PI
- Función POWER
- Función RADIANS
- Función RANDOM
- Función ROUND
- Función SIGN
- Función SIN

Expresión

relación

de SELECT

análisis

Sparklines

admite

de instrucciones

iones

SQL

- Funciones
SQRT
- Función
TRUNC

Flujos

VARBYTE

compatible

es UNBASE64

- MALEFICIO
- HLL_SKETC
H_AGG,
- HLL_SKETC
H_ESTIMAT
E
- HLL_UNION
- HLL_UNION
_AGG

JSON

son GET_JSON_

compatible

es

~~El~~ Expresión

~~relator~~ relator

~~de~~ SELECT

~~análisis~~ análisis

~~Spark~~ Spark

~~admite~~ admite

~~estructura~~ estructura

iones

SQL

~~ARRAYS~~ ARRAYS_CON

~~array~~ array

~~matriz~~ matriz

• MATRIZ_DI

STINCT

• MATRIZ_EX
CEPTO

• ARRAY_INT
ERSECT

• ARRAY_JOI
N

• ARRAY_REM
OVE

• ARRAY_SOR
T

• ARRAY_UNI
ON

~~ARRAY~~ ARRAY

~~CONJUNTOS~~ CONJUNTOS

POR

ENROLLAR,

CUBOS

Expresión

cláusula

SELECT

análisis

Sparklines

admite

instrucc

iones

SQL

ORDER

admite

clasifica

ción

condición

de

que

la

cláusula

ORDER

BY

solo

se

admita

en

la

cláusula

de

partición

de

una

función

de

ventana

cuando

se

El **análisis**

relación

de **SELECT**

análisis

Spark

admite

CTEs

iones

SQL

consultan

tablas

con

la

privacida

d

diferenci

al

activada.

Notas

OFFSET

compatible

es

CTEs

uso

de

tablas

con

protección

n

de

privacida

d

diferenci

al

Expresión

relatoria

de

SELECT

Spark

admite

instrucc

iones

SQL

Almacén

de

tablas

y

columnas

Funciones

matemáticas

as

en

funciones

agregadas

Funciones

escalares

dentro

de

funciones

agregadas

Alternativas comunes para constructos SQL no admitidos

Categoría	Constructo SQL	Alternativa
Funciones de ventana	• LISTAGG • PERCENTILE_CONT • PERCENTILE_DISC	Puede utilizar la función de agregado equivalente con GROUP BY.
Símbolos de operadores matemáticos	• \$column / 2 • \$column / 2 • \$column ^ 2	• CBRT • SQRT • POWER(\$column, 2)
Funciones escalares	• SYSDATE • \$column::integer • convert(type, \$column)	• CURRENT_DATE • CAST \$column AS integer • CAST \$column AS type
Literales	INTERVALO «1 SEGUNDO»	INTERVALO «1» SEGUNDO
Límite de filas	TOP n	LÍMITE n
Join	• USING • NATURAL	La cláusula ON debe contener explícitamente un criterio de unión.

Consejos y ejemplos de consultas sobre privacidad diferencial

AWS Clean Rooms Differential Privacy utiliza una [estructura de consulta de uso general](#) para admitir una amplia variedad de construcciones de SQL, como las expresiones de tabla comunes (CTEs) para la preparación de datos y las funciones agregadas de uso común, como, o. COUNT SUM Con el fin de ocultar la contribución de cualquier posible usuario a los datos y hacer más ruido al agregado de los resultados de las consultas en tiempo de ejecución, AWS Clean Rooms Differential Privacy exige que, en última SELECT statement instancia, las funciones de agregación se ejecuten en datos de nivel de usuario.

En el siguiente ejemplo se utilizan dos tablas denominadas socialco_impressions y socialco_users de un publicador de medios que quiere proteger los datos mediante una privacidad diferencial y, al mismo tiempo, colaborar con una marca deportiva que utiliza datos

`athletic_brand_sales`. El publicador multimedia ha configurado la columna `user_id` como columna de identificador de usuario y, al mismo tiempo, ha habilitado una privacidad diferencial en AWS Clean Rooms. El anunciante no necesita una protección de privacidad diferencial y quiere ejecutar una consulta utilizando datos combinados. CTEs Puesto que su CTE utiliza tablas protegidas por la privacidad diferencial, el anunciante incluye la columna de identificador de usuario de esas tablas protegidas en la lista de columnas de CTE y une las tablas protegidas en la columna de identificador de usuario.

```
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
```

Del mismo modo, si desea ejecutar funciones de ventana en tablas de datos con protección de privacidad diferencial, debe incluir la columna de identificador de usuario en la cláusula `PARTITION BY`.

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


Limitaciones de la AWS Clean Rooms privacidad diferencial

AWS Clean Rooms La privacidad diferencial no aborda las siguientes situaciones:

1. AWS Clean Rooms Differential Privacy solo admite consultas con tablas respaldadas por Amazon S3 AWS Glue . No admite consultas con tablas de Snowflake o Amazon Athena.
2. AWS Clean Rooms Differential Privacy no aborda los ataques cronometrados. Por ejemplo, estos ataques son posibles en situaciones en las que un usuario individual contribuye con un gran número de filas y la adición o eliminación de este usuario cambia considerablemente el tiempo de cálculo de la consulta.
3. AWS Clean Rooms La privacidad diferencial no garantiza la privacidad diferencial cuando una consulta SQL puede provocar desbordamientos o errores de conversión no válidos en tiempo de ejecución debido al uso de determinadas construcciones de SQL.

En la siguiente tabla se muestran algunos constructos de SQL, pero no todos, que pueden producir errores en tiempo de ejecución y que se deben verificar en las plantillas de análisis. Le recomendamos que apruebe plantillas de análisis que minimicen las posibilidades de que se produzcan dichos errores en tiempo de ejecución y que revise periódicamente los registros de consultas para determinar si las consultas se ajustan al acuerdo de colaboración.

Los siguientes constructos de SQL son vulnerables a los errores de desbordamiento:

Categoría	Las construcciones de SQL son vulnerables a errores de desbordamiento en el motor de análisis SQL de Spark
Funciones de agregación	• AVG • SUM/SUM_DISTINCT
Funciones de formato de tipo de datos	• TO_TIMESTAMP • TO_DATE
Funciones de fecha y hora	• ADD_MONTHS • DATEADD • DATEDIFF
Funciones matemáticas	• +, -, *, /

Categoría	Las construcciones de SQL son vulnerables a errores de desbordamiento en el motor de análisis SQL de Spark
	• POWER
Funciones de cadena	• • CONCAT • REPEAT
Funciones de ventana	• AVG • SUM

4. La función de formateo de tipos de datos CAST es vulnerable a errores de conversión no válidos.

Puede configurarlo [CloudWatch para crear un filtro de métricas para un grupo de registros](#) y, a continuación, [crear una CloudWatch alarma](#) en ese filtro de métricas para recibir alertas en caso de que se produzca un posible desbordamiento o error de conversión.

En concreto, debe supervisar los códigos de error `CastError`, `OverflowError` y `ConversionError`. La presencia de estos códigos de error indica un posible ataque de canal lateral, pero podría indicar una consulta SQL errónea.

Para obtener más información, consulte [Inicio de sesión de Analysis AWS Clean Rooms](#).

AWS Clean Rooms ML

El aprendizaje automático de AWS Clean Rooms permite a dos o más partes ejecutar modelos de aprendizaje automático en sus datos sin necesidad de compartirlos entre sí. El servicio proporciona controles que mejoran la privacidad y permiten a los propietarios de los datos proteger sus datos y la IP de su modelo. Puede usar modelos AWS creados o traer su propio modelo personalizado.

Para obtener una explicación más detallada de cómo funciona esto, consulte [Trabajos entre cuentas](#)

Para obtener más información sobre las capacidades de los modelos Clean Rooms ML, consulte los siguientes temas.

Temas

- [Terminología de aprendizaje automático de AWS Clean Rooms](#)

- [Cómo funciona AWS Clean Rooms ML con AWS los modelos](#)
- [Cómo funciona AWS Clean Rooms ML con modelos personalizados](#)
- [AWS modelos en Clean Rooms ML](#)
- [Modelos personalizados en Clean Rooms ML](#)

Terminología de aprendizaje automático de AWS Clean Rooms

Es importante entender la siguiente terminología al utilizar Clean Rooms ML:

- Proveedor de datos de entrenamiento: la parte que aporta los datos de entrenamiento, crea y configura un modelo similar y, a continuación, lo asocia a una colaboración.
- Proveedor de datos iniciales: la parte que aporta los datos iniciales, genera un segmento similar y exporta su segmento similar.
- Datos de entrenamiento: los datos del proveedor de datos de entrenamiento, que se utilizan para generar un modelo similar. Los datos de entrenamiento se utilizan para medir la similitud en los comportamientos de los usuarios.

Los datos de entrenamiento deben contener un ID de usuario, un ID de elemento y una columna de marca de tiempo. De forma opcional, los datos de entrenamiento pueden contener otras interacciones, como características numéricas o categóricas. Algunos ejemplos de interacciones son una lista de los vídeos visualizados, los artículos comprados o los artículos leídos.

- Datos iniciales: los datos del proveedor de datos iniciales, que se utilizan para crear un segmento similar. Los datos iniciales se pueden proporcionar directamente o pueden provenir de los resultados de una AWS Clean Rooms consulta. El resultado del segmento similar es un conjunto de usuarios a partir de los datos de entrenamiento que se parece más a los usuarios iniciales.
- Modelo similar: un modelo de machine learning de los datos de entrenamiento que se utiliza para encontrar usuarios similares en otros conjuntos de datos.

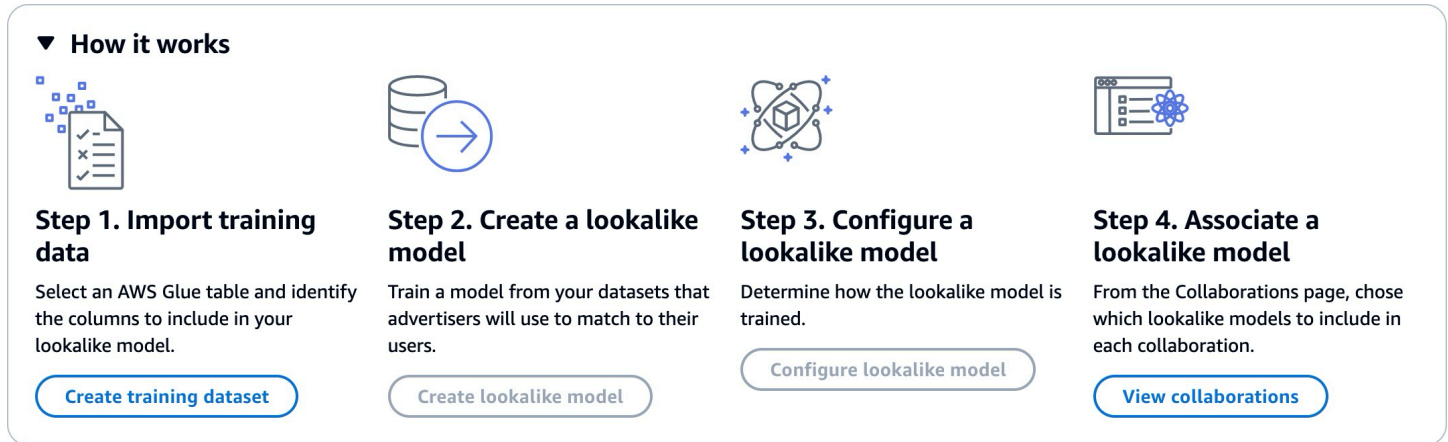
Cuando se utiliza la API, el término modelo de audiencia se utiliza de manera equivalente a modelo similar. Por ejemplo, se utiliza la [CreateAudienceModel](#) API para crear un modelo similar.

- Segmento similar: es un subconjunto de los datos de entrenamiento que se parecen más a los datos iniciales.

Al utilizar la API, se crea un segmento similar con la [StartAudienceGenerationJob](#) API.

Los datos del proveedor de datos de entrenamiento nunca se comparten con el proveedor de datos iniciales, y viceversa. La salida del segmento similar se comparte con el proveedor de datos de entrenamiento, pero nunca con el proveedor de datos iniciales.

Cómo funciona AWS Clean Rooms ML con AWS los modelos



Para trabajar con modelos similares, es necesario que dos partes, un proveedor de datos de formación y un proveedor de datos iniciales, trabajen secuencialmente AWS Clean Rooms para integrar sus datos en una colaboración. Este es el flujo de trabajo que el proveedor de datos de entrenamiento debe completar primero:

1. Los datos del proveedor de datos de formación deben almacenarse en una tabla de catálogo de AWS Glue datos con las interacciones entre usuarios y elementos. Como mínimo, los datos de entrenamiento deben contener una columna de ID de usuario, una columna de ID de interacción y una columna de marca temporal.
2. El proveedor de datos de entrenamiento registra los datos de entrenamiento con AWS Clean Rooms.
3. El proveedor de datos de entrenamiento crea un modelo similar que se puede compartir con varios proveedores de datos iniciales. El modelo similar es una red neuronal profunda que puede tardar hasta 24 horas en entrenarse. No se reentrena automáticamente y le recomendamos que vuelva a entrenar el modelo una vez por semana.
4. El proveedor de datos de entrenamiento configura el modelo similar, que incluye si desea compartir las métricas de relevancia y la ubicación en Amazon S3 de los segmentos de salida. El proveedor de datos de entrenamiento puede crear varios modelos similares configurados a partir de un único modelo similar.
5. El proveedor de datos de entrenamiento asocia el modelo de audiencia configurado a una colaboración que se comparte con un proveedor de datos iniciales.

Este es el flujo de trabajo que el proveedor de datos iniciales debe completar a continuación:

1. Los datos del proveedor de datos iniciales se pueden almacenar en un bucket de Amazon S3 o pueden provenir de los resultados de una consulta.
2. El proveedor de datos iniciales comienza la colaboración que comparte con el proveedor de datos de entrenamiento.
3. El proveedor de datos iniciales crea un segmento similar en la pestaña de Clean Rooms ML de la página de colaboración.
4. El proveedor de datos iniciales puede evaluar las métricas de relevancia, en caso de que se hayan compartido, y exportar el segmento similar para su uso fuera de AWS Clean Rooms.

Cómo funciona AWS Clean Rooms ML con modelos personalizados

Con Clean Rooms ML, los miembros de una colaboración pueden utilizar un algoritmo de modelo personalizado dockerizado que se almacena en Amazon ECR para analizar sus datos de forma conjunta. Para ello, el proveedor del modelo debe crear una imagen y almacenarla en Amazon ECR. Siga los pasos de la [Guía del usuario de Amazon Elastic Container Registry](#) para crear un repositorio privado que contendrá el modelo de aprendizaje automático personalizado.

Cualquier miembro de una colaboración puede ser el proveedor del modelo, siempre que cuente con los permisos correctos. Todos los miembros de una colaboración pueden aportar datos de entrenamiento, datos de inferencia o ambos al modelo. A los efectos de esta guía, los miembros que aportan datos se denominan proveedores de datos. El miembro que crea la colaboración es el creador de la colaboración y puede ser el proveedor del modelo, uno de los proveedores de datos o ambos.

En el nivel más alto, estos son los pasos que se deben completar para realizar un modelado de aprendizaje automático personalizado:

1. El creador de la colaboración crea una colaboración y asigna a cada miembro las capacidades y la configuración de pago adecuadas. En este paso, el creador de la colaboración debe asignar al miembro la capacidad de recibir los resultados del modelo o los resultados de las inferencias al miembro correspondiente, ya que no se puede actualizar una vez creada la colaboración. Para obtener más información, consulte [Crear y unirse a la colaboración en AWS Clean Rooms ML](#).
2. El proveedor del modelo configura y asocia su modelo de aprendizaje automático en contenedores a la colaboración y se asegura de que se establezcan restricciones de privacidad para los datos

- exportados. Para obtener más información, consulte [Configuración de un algoritmo modelo en AWS Clean Rooms ML](#).
3. Los proveedores de datos aportan sus datos a la colaboración y se aseguran de que se especifiquen sus necesidades de privacidad. Los proveedores de datos deben permitir que el modelo acceda a sus datos. Para obtener más información, consulte [Contribución de datos de formación en AWS Clean Rooms ML](#) y [Asociación del algoritmo del modelo configurado en AWS Clean Rooms ML](#).
 4. Un miembro de la colaboración crea la configuración de aprendizaje automático, que define a dónde se exportan los artefactos del modelo o los resultados de las inferencias.
 5. Un miembro de la colaboración crea un canal de entrada de aprendizaje automático que proporciona información al contenedor de formación o al contenedor de inferencias. El canal de entrada de ML es una consulta que define los datos que se utilizarán en el contexto del algoritmo del modelo.
 6. Un miembro de la colaboración invoca el entrenamiento del modelo mediante el canal de entrada ML y el algoritmo del modelo configurado. Para obtener más información, consulte [Creación de un modelo entrenado en AWS Clean Rooms ML](#).
 7. (Opcional) El entrenador de modelos invoca el trabajo de exportación del modelo y los artefactos del modelo se envían al receptor de resultados del modelo. Solo los miembros con una configuración de aprendizaje automático válida y con la capacidad de recibir la salida del modelo pueden recibir artefactos del modelo. Para obtener más información, consulte [Exportación de artefactos modelo desde AWS Clean Rooms ML](#).
 8. (Opcional) Un miembro de la colaboración invoca la inferencia del modelo mediante el canal de entrada ML, el ARN del modelo entrenado y el algoritmo del modelo configurado por inferencia. Los resultados de la inferencia se envían al receptor de salida de la inferencia. Solo los miembros con una configuración de aprendizaje automático válida y con la capacidad de recibir los resultados de la inferencia pueden recibir los resultados de la inferencia.

Estos son los pasos que debe completar el proveedor del modelo:

1. Cree una imagen de docker de Amazon ECR compatible con SageMaker IA. Clean Rooms ML solo admite imágenes de docker compatibles con SageMaker IA.
2. Una vez que haya creado una imagen de docker compatible con SageMaker IA, envíe la imagen a Amazon ECR. Siga las instrucciones de la [Guía del usuario de Amazon Elastic Container Registry](#) para crear una imagen de formación sobre contenedores.
3. Configure el algoritmo del modelo para usarlo en Clean Rooms ML.

- a. Proporcione el enlace al repositorio de Amazon ECR y todos los argumentos necesarios para configurar el algoritmo del modelo.
- b. Proporcione un rol de acceso al servicio que permita a Clean Rooms ML acceder al repositorio de Amazon ECR.
- c. Asocie el algoritmo del modelo configurado a la colaboración. Esto incluye proporcionar una política de privacidad que defina los controles para los registros de contenedores, los registros de errores, CloudWatch las métricas y los límites sobre la cantidad de datos que se pueden exportar de los resultados del contenedor.

Estos son los pasos que debe completar el proveedor de datos para colaborar con un modelo de aprendizaje automático personalizado:

1. Configure una AWS Glue tabla existente con una regla de análisis personalizada. Esto permite que un conjunto específico de consultas preaprobadas o cuentas preaprobadas utilice sus datos.
2. Asocie la tabla configurada a una colaboración y proporcione un rol de acceso al servicio que pueda acceder a sus AWS Glue tablas.
3. [Agregue una regla de análisis de colaboración](#) a la tabla que permita a la asociación de algoritmos del modelo configurado acceder a la tabla configurada.
4. Una vez que el modelo y los datos están asociados y configurados en Clean Rooms ML, el miembro con capacidad para ejecutar consultas proporciona una consulta SQL y selecciona el algoritmo del modelo que va a utilizar.

Una vez finalizado el entrenamiento del modelo, ese miembro inicia la exportación de los artefactos del entrenamiento del modelo o de los resultados de las inferencias. Estos artefactos o resultados se envían al miembro para que pueda recibir los resultados del modelo entrenado. El receptor de los resultados debe configurarlos `MachineLearningConfiguration` antes de poder recibir la salida del modelo.

AWS modelos en Clean Rooms ML

AWS Clean Rooms ML proporciona un método que preserva la privacidad para que dos partes identifiquen a usuarios similares en sus datos sin necesidad de compartir sus datos entre sí. La primera parte aporta los datos de entrenamiento para AWS Clean Rooms poder crear y configurar un modelo similar y asociarlo a una colaboración. A continuación, los datos iniciales se traen a la colaboración para crear un segmento similar que se parece a los datos de entrenamiento.

Para obtener una explicación más detallada de cómo funciona esto, consulte [Trabajos entre cuentas](#)

Los siguientes temas proporcionan información sobre cómo crear y configurar un AWS modelo en Clean Rooms ML.

Temas

- [Protecciones de privacidad de AWS Clean Rooms ML](#)
- [Requisitos de datos de entrenamiento para Clean Rooms ML](#)
- [Requisitos de datos iniciales para Clean Rooms ML](#)
- [Métricas de evaluación del modelo de aprendizaje automático de AWS Clean Rooms](#)

Protecciones de privacidad de AWS Clean Rooms ML

Clean Rooms ML está diseñado para reducir el riesgo de ataques de inferencia de suscripción, ya que el proveedor de datos de entrenamiento puede saber quién está en los datos iniciales y el proveedor de datos iniciales puede saber quién está en los datos de entrenamiento. Para prevenir este ataque, se toman varias medidas.

En primer lugar, los proveedores de datos iniciales no observan directamente los resultados de Clean Rooms ML y los proveedores de datos de entrenamiento nunca pueden observar los datos iniciales. Los proveedores de datos iniciales pueden optar por incluir los datos iniciales en el segmento de salida.

A continuación, se crea el modelo similar a partir de una muestra aleatoria de los datos de entrenamiento. Este ejemplo incluye un número significativo de usuarios que no coinciden con la audiencia inicial. Este proceso hace que sea más difícil determinar si un usuario no aparecía en los datos, lo cual es otra forma de deducir la membresía.

Además, se pueden utilizar varios clientes iniciales para cada parámetro del entrenamiento de modelos similares específicos del inicio. Esto limita cuánto puede sobreajustarse el modelo y, por lo tanto, cuánto se puede deducir sobre un usuario. Como resultado, se recomienda que el tamaño mínimo de los datos iniciales sea de 500 usuarios.

Por último, nunca se proporcionan métricas en el nivel de usuario a los proveedores de datos de entrenamiento, lo que elimina otra posibilidad de que se produzca un ataque de inferencia de pertenencia.

Requisitos de datos de entrenamiento para Clean Rooms ML

Para crear correctamente un modelo similar, los datos de entrenamiento deben cumplir los siguientes requisitos:

- Los datos de entrenamiento deben estar en formato Parquet, CSV o JSON.

Note

No se admiten los datos de Parquet comprimidos con Zstandard (ZSTD).

- Tus datos de entrenamiento deben estar catalogados en AWS Glue. Para obtener más información, consulte [Introducción al catálogo de datos de AWS Glue](#) en la Guía para AWS Glue desarrolladores. Le recomendamos que utilice AWS Glue rastreadores para crear las tablas, ya que el esquema se deduce automáticamente.
- El depósito de Amazon S3 que contiene los datos de entrenamiento y los datos iniciales se encuentra en la misma AWS región que sus otros recursos de aprendizaje automático para salas limpias.
- Los datos de entrenamiento deben contener al menos 100 000 usuarios únicos IDs con al menos dos interacciones entre elementos cada uno.
- Los datos de entrenamiento deben contener al menos 1 millón de registros.
- El esquema especificado en la [CreateTrainingDataset](#) acción debe alinearse con el esquema definido cuando se creó la AWS Glue tabla.
- Los campos obligatorios, tal como se definen en la tabla proporcionada, se definen en la acción [CreateTrainingDataset](#).

Tipo de campo	Tipos de datos compatibles	Obligatorio	Description (Descripción)
USER_ID	string, int, bigint	Sí	Un identificador único para cada usuario del conjunto de datos. No

Tipo de campo	Tipos de datos compatibles	Obligatorio	Description (Descripción)
			debe tener un valor de información de identificación personal (PII). Es posible que sea un identificador con hash o un ID de cliente.
ITEM_ID	string, int, bigint	Sí	Un identificador único para cada elemento con el que interactúa un usuario.

Tipo de campo	Tipos de datos compatibles	Obligatorio	Description (Descripción)
TIMESTAMP	bigint, int, timestamp	Sí	La hora en que un usuario interactuó con el elemento. Los valores deben estar en formato de tiempo Unix en formato de segundos.

Tipo de campo	Tipos de datos compatibles	Obligatorio	Description (Descripción)
CATEGORICAL_FEATURE	string, int, float, bigint, double, boolean, array	No	Captura datos categóricos relacionados con el usuario o el elemento. Esto puede incluir cosas como un tipo de evento (como hacer clic o comprar), los datos demográficos de los usuarios (grupo de edad, sexo: anonimizado), la ubicación del usuario (ciudad, país, anonimizado), la categoría del artículo (como ropa

Tipo de campo	Tipos de datos compatibles	Obligatorio	Description (Descripción)
			o artículos electrónicos) o la marca del artículo.

Tipo de campo	Tipos de datos compatibles	Obligatorio	Description (Descripción)
NUMERICAL_FEATURE	double, float, int, bigint	No	Captura datos numéricos relacionados con el usuario o el elemento. Esto puede incluir cosas como el historial de compras del usuario (importe total gastado), el precio del artículo, el número de veces que se visita un artículo o las valoraciones de los usuarios sobre los artículos.

- Opcionalmente, puede proporcionar hasta 10 características categóricas o numéricas en total.

Este es un ejemplo de un conjunto de datos de entrenamiento válido en formato 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
```

Requisitos de datos iniciales para Clean Rooms ML

Los datos iniciales de un modelo similar pueden provenir directamente de un bucket de Amazon S3 o de los resultados de una consulta SQL.

Los datos iniciales que se proporcionan directamente deben cumplir los siguientes requisitos:

- Los datos iniciales deben estar en formato de líneas JSON con una lista de usuarios IDs.
- El tamaño de la semilla debe estar comprendido entre 25 y 500 000 usuarios únicos IDs.
- El número mínimo de usuarios iniciales debe coincidir con el valor mínimo de tamaño inicial coincidente que se especificó al crear el modelo de audiencia configurado.

El siguiente es un ejemplo de un conjunto de datos de entrenamiento válido en formato CSV

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

Métricas de evaluación del modelo de aprendizaje automático de AWS Clean Rooms

Clean Rooms ML calcula la exhaustividad y la puntuación de relevancia para determinar el rendimiento del modelo. La exhaustividad compara la similitud entre los datos similares y los datos de entrenamiento. La puntuación de relevancia se utiliza para decidir cómo de grande debe ser la audiencia, no si el rendimiento del modelo es bueno.

La exhaustividad es una medida imparcial de la similitud del segmento similar con los datos de entrenamiento. La exhaustividad es el porcentaje de usuarios más similares (de forma

predeterminada, el 20 % más similar) de una muestra de los datos de entrenamiento que se incluyen en la audiencia inicial según el trabajo de generación de audiencia. Los valores oscilan entre 0 y 1; los valores más altos indican una mejor audiencia. Un valor de exhaustividad aproximadamente igual al porcentaje máximo de intervalo indica que el modelo de audiencia equivale a una selección aleatoria.

Consideramos que esta métrica de evaluación es mejor que la exactitud, la precisión y las puntuaciones de F1 porque Clean Rooms ML no ha etiquetado con precisión a los usuarios que realmente son negativos al crear su modelo.

La puntuación de relevancia del segmento es una medida de similitud con valores que van desde -1 (menos similar) a 1 (más similar). Clean Rooms ML calcula un conjunto de puntuaciones de relevancia para varios tamaños de segmento a fin de ayudarle a determinar el mejor tamaño de segmento para los datos. Las puntuaciones de relevancia disminuyen de forma monótona a medida que aumenta el tamaño del segmento, por lo que, a medida que aumenta el tamaño del segmento, pueden ser menos similares a los datos iniciales. Cuando la puntuación de relevancia del segmento llega a 0, el modelo predice que todos los usuarios del segmento similar provienen de la misma distribución que los datos iniciales. Al aumentar el tamaño de salida, es probable que se incluyan usuarios del segmento similar que no pertenezcan a la misma distribución que los datos iniciales.

Las puntuaciones de relevancia se normalizan en una sola campaña y no se deben utilizar para comparar campañas. Los puntajes de relevancia no deben utilizarse como una fuente única de evidencia para determinar ningún resultado empresarial, ya que se ven afectados por varios factores complejos además de la relevancia, como la calidad del inventario, el tipo de inventario, el momento en que se publica la publicidad, etc.

Las puntuaciones de relevancia no se deben utilizar para juzgar la calidad del inicio, sino para determinar si se puede aumentar o disminuir. Considere los siguientes ejemplos:

- Todas las puntuaciones son positivas: esto indica que hay más usuarios de salida que se consideran similares que los que se incluyen en el segmento similar. Esto es habitual en el caso de los datos iniciales que forman parte de un mercado grande, como el de todos los que han comprado pasta de dientes en el último mes. Le recomendamos que consulte los datos iniciales más pequeños, como los de todas las personas que han comprado pasta dental más de una vez en el último mes.
- Todas las puntuaciones negativas o negativas para su tamaño de segmento similar deseado: esto indica que Clean Rooms ML predice que no hay suficientes usuarios similares en el tamaño de segmento similar deseado. Esto se puede deber a que los datos iniciales son demasiado

específicos o a que el mercado es demasiado pequeño. Recomendamos aplicar menos filtros a los datos iniciales o ampliar el mercado. Por ejemplo, si los datos iniciales originales eran clientes que habían comprado un cochecito y una silla de coche, podría ampliar el mercado a clientes que hayan comprado varios productos para bebés.

Los proveedores de datos de entrenamiento determinan si se exponen las puntuaciones de relevancia y cuáles son los contenedores de bucket en los que se calculan las puntuaciones de relevancia.

Modelos personalizados en Clean Rooms ML

Con Clean Rooms ML, los miembros de una colaboración pueden utilizar un algoritmo de modelo personalizado dockerizado que se almacena en Amazon ECR para analizar sus datos de forma conjunta. Para ello, el proveedor del modelo debe crear una imagen y almacenarla en Amazon ECR. Siga los pasos de la [Guía del usuario de Amazon Elastic Container Registry](#) para crear un repositorio privado que contendrá el modelo de aprendizaje automático personalizado.

Cualquier miembro de una colaboración puede ser el proveedor del modelo, siempre que cuente con los permisos correctos. Todos los miembros de una colaboración pueden aportar datos al modelo. A los efectos de esta guía, los miembros que aportan datos se denominan proveedores de datos. El miembro que crea la colaboración es el creador de la colaboración y puede ser el proveedor del modelo, uno de los proveedores de datos o ambos.

En los temas siguientes se describe la información necesaria para crear un modelo de aprendizaje automático personalizado

Temas

- [Requisitos previos del modelado ML personalizado](#)
- [Pautas de creación de modelos para el contenedor de formación](#)
- [Directrices de creación de modelos para el contenedor de inferencias](#)
- [Recibir registros y métricas del modelo](#)

Requisitos previos del modelado ML personalizado

Antes de poder realizar un modelado de aprendizaje automático personalizado, debe tener en cuenta lo siguiente:

- Determine si tanto el entrenamiento del modelo como la inferencia sobre el modelo entrenado se realizarán en colaboración.
- Determine la función que desempeñará cada miembro de la colaboración y asígnele las habilidades adecuadas.
 - Asigne la `CAN_QUERY` habilidad al miembro que entrenará el modelo y realizará la inferencia sobre el modelo entrenado.
 - `CAN_RECEIVE_RESULTS` Asígnelo al menos a un miembro de la colaboración.
 - Asigne `CAN_RECEIVE_MODEL_OUTPUT` nuestras `CAN_RECEIVE_INFERENCE_OUTPUT` habilidades al miembro que recibirá las exportaciones de modelos entrenados o el resultado de inferencias, respectivamente. Puedes elegir usar ambas habilidades si así lo requiere tu caso de uso.
- Determine el tamaño máximo de los artefactos del modelo entrenado o de los resultados de inferencia que permitirá exportar.
- Recomendamos que todos los usuarios tengan las `CleanroomsMLFullAccess` políticas `CleanroomsFullAccess` y políticas asociadas a su función. El uso de modelos de aprendizaje automático personalizados requiere utilizar tanto el AWS Clean Rooms ML de AWS Clean Rooms como el de AWS SDKs.
- Tenga en cuenta la siguiente información sobre las funciones de IAM.
 - Todos los proveedores de datos deben tener una función de acceso AWS Clean Rooms al servicio que les permita leer los datos de sus AWS Glue catálogos y tablas, y de las ubicaciones subyacentes de Amazon S3. Estas funciones son similares a las que se requieren para las consultas de SQL. Esto le permite utilizar la `CreateConfiguredTableAssociation` acción. Para obtener más información, consulte [Cree un rol de servicio para crear una asociación de tablas configurada](#).
 - Todos los miembros que deseen recibir métricas deben tener una función de acceso al servicio que les permita escribir CloudWatch métricas y registros. Clean Rooms ML utiliza esta función para escribir todas las métricas y registros del modelo para el miembro Cuenta de AWS durante el entrenamiento y la inferencia del modelo. También ofrecemos controles de privacidad para determinar qué miembros tienen acceso a las métricas y los registros. Esto te permite usar la `CreateMLConfiguration` acción. Para obtener más información, consulte, [Cree un rol de servicio para el modelado de ML personalizado: Configuración de ML](#).

El miembro que reciba los resultados debe proporcionar un rol de acceso al servicio con permisos para escribir en su bucket de Amazon S3. Esta función permite a Clean Rooms ML exportar los resultados (artefactos de modelos entrenados o resultados de inferencias)

a un bucket de Amazon S3. Esto le permite utilizar la `CreateMLConfiguration` acción. Para obtener más información, consulte [Cree un rol de servicio para el modelado de ML personalizado: Configuración de ML](#).

- El proveedor del modelo debe proporcionar un rol de acceso al servicio con permisos para leer su repositorio e imagen de Amazon ECR. Esto le permite utilizar la `CreateConfigureModelAlgorithm` acción. Para obtener más información, consulte [Cree un rol de servicio para proporcionar un modelo de aprendizaje automático personalizado](#).
- El miembro que crea el `MLInputChannel` para generar conjuntos de datos con fines de formación o inferencia debe proporcionar una función de acceso al servicio que permita a Clean Rooms ML ejecutar una consulta SQL en ella. AWS Clean Rooms Esto le permite usar las acciones `CreateTrainedModel` y `StartTrainedModelInferenceJob`. Para obtener más información, consulte [Cree un rol de servicio para consultar un conjunto de datos](#).
- Los autores del modelo deben seguir las [Pautas de creación de modelos para el contenedor de formación](#) y [Directrices de creación de modelos para el contenedor de inferencias](#) asegurarse de que las entradas y salidas del modelo estén configuradas según lo esperado AWS Clean Rooms.

Pautas de creación de modelos para el contenedor de formación

En esta sección se detallan las pautas que los proveedores de modelos deben seguir al crear un algoritmo de modelo de aprendizaje automático personalizado para el aprendizaje automático de salas limpias.

- Utilice la imagen base de contenedores adecuada y compatible con la formación en SageMaker IA, tal y como se describe en la Guía para [desarrolladores de SageMaker IA](#). El siguiente código te permite extraer las imágenes de la base de contenedores compatibles de puntos finales de SageMaker IA públicos.

```
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
```

- Cuando cree el modelo localmente, asegúrese de lo siguiente para poder probarlo localmente, en una instancia de desarrollo, en SageMaker AI Training en su Cuenta de AWS empresa y en Clean Rooms ML.

- Recomendamos escribir un guion de formación que acceda a propiedades útiles del entorno de formación a través de diversas variables de entorno. Clean Rooms ML utiliza los siguientes argumentos para invocar el entrenamiento en el código del modelo: `SM_MODEL_DIR`, `SM_OUTPUT_DIR`, `SM_CHANNEL_TRAIN`, y `FILE_FORMAT`. Clean Rooms ML utiliza estos valores predeterminados para entrenar su modelo de aprendizaje automático en su propio entorno de ejecución con los datos de todas las partes.
- Clean Rooms ML hace que sus canales de entrada de formación estén disponibles a través de los `/opt/ml/input/data/channel-name` directorios del contenedor docker. Cada canal de entrada de ML se mapea en función del correspondiente que `channel_name` se proporciona en la `CreateTrainedModel` solicitud.

```
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"))
```

- Asegúrese de poder generar un conjunto de datos sintético o de prueba basado en el esquema de los colaboradores que se utilizará en el código de su modelo.
- Asegúrese de poder realizar un trabajo de formación en SageMaker IA por su cuenta Cuenta de AWS antes de asociar el algoritmo del modelo a una AWS Clean Rooms colaboración.

El siguiente código contiene un archivo Docker de muestra que es compatible con las pruebas locales, las pruebas del entorno de formación de SageMaker IA y el aprendizaje automático de salas limpias

```
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"]
```

- Para supervisar mejor los fallos de los contenedores, recomendamos exportar los registros y depurarlos en caso de que se produzcan errores. Como `GetTrainedModel` respuesta, Clean Rooms ML devuelve los primeros 1024 caracteres de este archivo. `StatusDetails`
- Una vez que haya realizado cualquier cambio en el modelo y esté listo para probarlo en el entorno de SageMaker IA, ejecute los siguientes comandos en el orden indicado.

```
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
```

Una vez que haya completado el trabajo de SageMaker IA y esté satisfecho con el algoritmo de su modelo, podrá registrar el registro de Amazon ECR en AWS Clean Rooms ML. Utilice la `CreateConfiguredModelAlgorithm` acción para registrar el algoritmo del modelo y asociarlo a una colaboración. `CreateConfiguredModelAlgorithmAssociation`

Directrices de creación de modelos para el contenedor de inferencias

En esta sección se detallan las pautas que los proveedores de modelos deben seguir al crear un algoritmo de inferencia para Clean Rooms ML.

- [Utilice la imagen base de contenedor adecuada compatible con la inferencia de SageMaker IA, tal y como se describe en la SageMaker Guía para desarrolladores de IA.](#) El siguiente código te permite extraer las imágenes de base de contenedores compatibles de puntos finales de IA públicos SageMaker .

```
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
```

- Cuando cree el modelo localmente, asegúrese de lo siguiente para poder probarlo localmente, en una instancia de desarrollo, en SageMaker AI Batch Transform en su Cuenta de AWS entorno y en Clean Rooms ML.
 - Clean Rooms ML hace que los artefactos de su modelo a partir de la inferencia estén disponibles para que los utilice su código de inferencia a través del `/opt/ml/model` directorio del contenedor docker.
 - Clean Rooms ML divide la entrada por línea, utiliza una estrategia por `MultiRecord` lotes y añade un carácter de nueva línea al final de cada registro transformado.
 - Asegúrese de poder generar un conjunto de datos de inferencia sintético o de prueba basado en el esquema de los colaboradores que se utilizará en el código de su modelo.
 - Asegúrese de poder ejecutar un trabajo de transformación por lotes de SageMaker IA por su cuenta Cuenta de AWS antes de asociar el algoritmo del modelo a una AWS Clean Rooms colaboración.

El siguiente código contiene un ejemplo de archivo Docker que es compatible con las pruebas locales, las pruebas de entornos de transformación mediante SageMaker IA y el aprendizaje automático de salas limpias

```
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"]
```

- Una vez que haya realizado cualquier cambio en el modelo y esté listo para probarlo en el entorno de SageMaker IA, ejecute los siguientes comandos en el orden indicado.

```
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
```

Una vez que haya completado el trabajo de SageMaker IA y esté satisfecho con la transformación por lotes, podrá registrar el registro de Amazon ECR en AWS Clean Rooms ML. Utilice la `CreateConfiguredModelAlgorithm` acción para registrar el algoritmo del modelo y asociarlo a una colaboración. `CreateConfiguredModelAlgorithmAssociation`

Recibir registros y métricas del modelo

Para recibir registros y métricas del entrenamiento o la inferencia de modelos personalizados, los miembros deben haber [creado una configuración de aprendizaje automático](#) con un rol válido que proporcione los CloudWatch permisos necesarios (consulte [Crear un rol de servicio para el modelado de aprendizaje automático personalizado: configuración de aprendizaje automático](#)).

Métrica del sistema

Las métricas del sistema, tanto para el entrenamiento como para la inferencia, como el uso de la CPU y la memoria, se publican para todos los miembros en colaboración con las configuraciones de aprendizaje automático válidas. Estas métricas se pueden ver a medida que avanza el trabajo mediante CloudWatch las métricas de los espacios de `/aws/cleanroomsm1/TrainedModelInferenceJobs` nombres `/aws/cleanroomsm1/TrainedModels` o los espacios de nombres, respectivamente.

Registros del modelo

El acceso a los registros del modelo lo proporciona la política de configuración de privacidad de cada algoritmo del modelo configurado. El autor del modelo establece la política de configuración de privacidad al asociar un algoritmo de modelo configurado (mediante la consola o

la `CreateConfiguredModelAlgorithmAssociation` API) a una colaboración. Al establecer la política de configuración de privacidad, se controla qué miembros pueden recibir los registros del modelo.

Además, el autor del modelo puede establecer un patrón de filtro en la política de configuración de privacidad para filtrar los eventos del registro. Todos los registros que envía un contenedor modelo `stdout` o `stderr` que coinciden con el patrón de filtro (si está establecido) se envían a Amazon CloudWatch Logs. Los registros del modelo están disponibles en grupos de CloudWatch registros / `aws/cleanroomsm1/TrainedModels` o `aws/cleanroomsm1/TrainedModelInferenceJobs`, respectivamente.

Métricas definidas personalizadas

Al configurar un algoritmo de modelo (mediante la consola o la `CreateConfiguredModelAlgorithm` API), el autor del modelo puede proporcionar nombres de métricas e instrucciones de expresiones regulares específicos para buscarlos en los registros de salida. Se pueden ver a medida que avanza el trabajo mediante CloudWatch Metrics en el / `aws/cleanroomsm1/TrainedModels` espacio de nombres. Al asociar un algoritmo de modelo configurado, el autor del modelo puede establecer un nivel de ruido opcional en la configuración de privacidad de las métricas para evitar generar datos sin procesar y, al mismo tiempo, ofrecer visibilidad de las tendencias métricas personalizadas. Si se establece un nivel de ruido, las métricas se publican al final del trabajo y no en tiempo real.

Computación criptográfica para Clean Rooms

Computación criptográfica para Clean Rooms [\(C3R\) es una capacidad AWS Clean Rooms que se puede utilizar además de las reglas de análisis](#). Con C3R, las organizaciones pueden recopilar datos confidenciales para obtener nuevos conocimientos a partir del análisis de datos al tiempo que limitan criptográficamente la información que puede conocer cualquiera de las partes que interviene en el proceso. C3R lo pueden utilizar dos o más partes que deseen colaborar con sus datos confidenciales, pero que estén obligadas a usar solo datos cifrados en la nube.

El cliente de cifrado C3R es una herramienta de cifrado del lado del cliente que puede utilizar para [cifrar sus datos](#) y utilizarlos con ellos. AWS Clean Rooms Cuando utiliza el cliente de cifrado C3R, los datos permanecen protegidos criptográficamente mientras se utilizan en una colaboración. AWS Clean Rooms Al igual que en una AWS Clean Rooms colaboración habitual, los datos de entrada son tablas de bases de datos relacionales y el cálculo se expresa como una consulta SQL. Sin embargo, C3R admite solo un subconjunto limitado de consultas SQL en datos cifrados.

En concreto, C3R es compatible con SQL JOIN y SELECT declaraciones sobre datos protegidos criptográficamente. Cada columna de la tabla de entrada se puede utilizar exactamente en uno de los siguientes tipos de instrucciones SQL:

- Columnas que están protegidas criptográficamente para su uso en JOIN las declaraciones se llaman fingerprint columnas.
- Columnas que están protegidas criptográficamente para su uso en SELECT las declaraciones se llaman sealed columnas.
- Columnas que no están protegidas criptográficamente para su uso en JOIN o SELECT las declaraciones se llaman cleartext columnas.

En algunos casos, GROUP BY las declaraciones se apoyan en fingerprint columnas. Para obtener más información, consulte [Columnas Fingerprint](#). Actualmente, C3R no admite el uso de otras construcciones de SQL en datos cifrados, como WHERE cláusulas o funciones agregadas como SUM y AVERAGE, incluso si de otro modo estarían permitidas por las reglas de análisis pertinentes.

C3R está diseñado para proteger los datos de celdas específicas de una tabla. Con la configuración predeterminada de C3R, los datos subyacentes que un cliente pone a disposición de terceros a través de una colaboración permanecen cifrados mientras el contenido se utiliza en AWS Clean Rooms. C3R utiliza el cifrado AES-GCM estándar de la industria para todos sealed columnas y una función pseudoaleatoria estándar del sector, conocida como código de autenticación de mensajes (HMAC) basado en hash, para proteger fingerprint columnas.

A pesar de que C3R cifra los datos de las tablas, es posible que aún se pueda inferir la siguiente información:

- Información sobre las propias tablas, incluido el número de columnas, los nombres de columna y el número de filas de la tabla.
- Como ocurre con la mayoría de las formas de cifrado estándar, C3R no intenta ocultar la longitud de los valores cifrados. C3R ofrece la posibilidad de rellenar los valores cifrados para ocultar la longitud exacta del texto sin cifrar. Sin embargo, aun así se podría revelar un límite superior de la longitud del texto sin cifrar de cada columna a un tercero.
- Información de registro, como cuándo se agregó una fila determinada a una tabla cifrada de C3R.

Para obtener más información acerca de CR3, consulte los siguientes temas.

Temas

- [Consideraciones al utilizar la computación criptográfica para Clean Rooms](#)
- [Tipos de archivo y de datos admitidos en computación criptográfica para Clean Rooms](#)
- [Nombres de columnas en computación criptográfica para Clean Rooms](#)
- [Tipos de columnas en computación criptográfica para Clean Rooms](#)
- [Parámetros de computación criptográfica](#)
- [Indicadores opcionales en computación criptográfica para Clean Rooms](#)
- [Consultas con computación criptográfica para Clean Rooms](#)
- [Directrices para el cliente de cifrado de C3R](#)

Consideraciones al utilizar la computación criptográfica para Clean Rooms

La computación criptográfica para Clean Rooms (C3R) busca maximizar la protección de los datos. Sin embargo, algunos casos de uso pueden beneficiarse de niveles más bajos de protección de datos a cambio de funcionalidades adicionales. Puede hacer estas concesiones específicas modificando C3R a partir de su configuración más segura. Como cliente, debe conocer estas ventajas y desventajas, y determinar si son apropiadas para su caso de uso. Entre las ventajas que debe considerar se incluyen las siguientes:

Temas

- [Permitir cleartext mixto y datos cifrados en sus tablas](#)
- [Permitir valores repetidos en columnas fingerprint](#)
- [Disminuir las restricciones de nomenclatura de las columnas fingerprint](#)
- [Determinar cómo se representan los valores NULL](#)

Para obtener más información sobre cómo establecer parámetros para estos escenarios, consulte [Parámetros de computación criptográfica](#).

Permitir cleartext mixto y datos cifrados en sus tablas

Hacer que todos los datos se cifren en el cliente proporciona la máxima protección de datos. Sin embargo, esto limita ciertos tipos de consultas (por ejemplo, la función de agregación SUM). El riesgo de permitir datos de cleartext es que es factible que cualquier persona con acceso a las tablas cifradas pueda inferir cierta información sobre los valores cifrados. Esto podría hacerse realizando un análisis estadístico de los datos de cleartext y de los datos relacionados.

Por ejemplo, supongamos que tiene las columnas de City y State. La columna City es cleartext y la columna State está cifrada. Cuando ve el valor Chicago en la columna City, esto le ayuda a determinar con una alta probabilidad que el State es Illinois. Por el contrario, si una columna es City y la otra columna es EmailAddress, es poco probable que un cleartext City revele algo sobre una EmailAddress cifrada.

Para obtener más información acerca del parámetro para este escenario, consulte [Parámetro Permitir columnas cleartext](#).

Permitir valores repetidos en columnas fingerprint

Para optar por un método más seguro, presuponemos que cualquier columna fingerprint contiene exactamente una instancia de una variable. En una columna fingerprint no se puede repetir ningún elemento. El cliente de cifrado de C3R asigna estos valores cleartext a valores únicos que son indistinguibles de los valores aleatorios. Por lo tanto, es imposible inferir información acerca del cleartext a partir de estos valores aleatorios.

El riesgo de los valores repetidos en una columna fingerprint es que estos valores repetidos resulten en valores repetidos que parezcan aleatorios. Por lo tanto, cualquier persona que tenga acceso a las tablas cifradas podría, en teoría, realizar un análisis estadístico de las columnas fingerprint que podría revelar información sobre los valores cleartext.

De nuevo, presuponemos que la columna fingerprint es State, y que cada fila de la tabla corresponde a un hogar estadounidense. Un análisis de frecuencia permitiría inferir qué estado es California y cuál es Wyoming con una alta probabilidad. Esta inferencia es posible porque California tiene muchos más residentes que Wyoming. Por el contrario, presuponemos que la columna fingerprint está en un identificador de hogar y que cada hogar aparece en la base de datos entre 1 y 4 veces en una base de datos de millones de entradas. Es poco probable que un análisis de frecuencia revele información útil.

Para obtener más información acerca del parámetro para este escenario, consulte [Parámetro Permitir duplicados](#).

Disminuir las restricciones de nomenclatura de las columnas fingerprint

De forma predeterminada, presuponemos que, cuando se combinan dos tablas utilizando columnas fingerprint cifradas, dichas columnas tienen el mismo nombre en cada tabla. El motivo técnico de este resultado es que, de forma predeterminada, deducimos una clave criptográfica diferente para cifrar cada columna fingerprint. Esa clave se deriva de una combinación de la clave secreta

compartida de la colaboración y el nombre de la columna. Si intentamos combinar dos columnas con nombres de columna diferentes, obtenemos claves diferentes y no podemos procesar una combinación válida.

Para solucionar este problema, puede desactivar la característica que deriva las claves a partir de cada nombre de columna. A continuación, el cliente de cifrado de C3R utiliza una única clave derivada para todas las columnas fingerprint. El riesgo reside en que se pueda realizar otro tipo de análisis de frecuencia que pueda revelar información.

Volvamos al ejemplo de `City` y `State`. Si derivamos los mismos valores aleatorios para cada columna fingerprint (al no incorporar el nombre de columna), `New York` tiene el mismo valor aleatorio en las columnas `City` y `State`. Nueva York es una de las pocas ciudades de EE. UU. donde el nombre de `City` coincide con el nombre del `State` nombre. Por el contrario, si su conjunto de datos tiene valores completamente diferentes en cada columna, no se filtra información.

Para obtener más información acerca del parámetro para este escenario, consulte [Parámetro Permitir JOIN de columnas con nombres diferentes](#).

Determinar cómo se representan los valores NULL

La opción que tiene a su disposición consiste en procesar criptográficamente los valores NULL (cifrado y HMAC) como cualquier otro valor. Si no procesa los valores NULL como cualquier otro valor, es posible que se revele información.

Por ejemplo, supongamos que NULL en la columna `Middle Name` de cleartext señala a las personas sin segundo nombre. Si no cifra esos valores, filtrará qué filas de la tabla cifrada se utilizan para personas sin segundo nombre. Esa información podría ser una señal de identificación para algunas personas de ciertas poblaciones. Sin embargo, si procesa los valores NULL criptográficamente, algunas consultas SQL actúan de forma diferente. Por ejemplo, las cláusulas `GROUP BY` no agruparán los valores fingerprint NULL en columnas fingerprint.

Para obtener más información acerca del parámetro para este escenario, consulte [Parámetro Conservar valores NULL](#).

Tipos de archivo y de datos admitidos en computación criptográfica para Clean Rooms

El cliente de cifrado de C3R reconoce los siguientes tipos de archivo:

- Archivos CSV

- Archivos de Parquet

Puede utilizar el indicador `--fileFormat` del cliente de cifrado de C3R para especificar un formato de archivo de forma explícita. Cuando se especifica de forma explícita, el formato de archivo no viene determinado por la extensión del archivo.

Temas

- [Archivos CSV](#)
- [Archivos de Parquet](#)
- [Cifrar valores que no son de cadena](#)

Archivos CSV

Se presupone que un archivo con la extensión `.csv` tiene formato CSV y contiene texto codificado en UTF-8. El cliente de cifrado de C3R trata todos los valores como cadenas.

Propiedades admitidas en los archivos `.csv`

El cliente de cifrado de C3R requiere que los archivos `.csv` tengan las siguientes propiedades:

- Puede contener o no una fila de encabezado inicial que designe de manera exclusiva cada columna.
- Delimitado por comas. (actualmente no se admiten delimitadores personalizados).
- Texto codificado en UTF-8.

Recorte de espacios en blanco de las entradas `.csv`

Los espacios en blanco iniciales y finales se recortan de las entradas `.csv`.

Codificación NULL personalizada para un archivo `.csv`

Un archivo `.csv` puede utilizar una codificación personalizada NULL.

Con el cliente de cifrado de C3R, puede especificar codificaciones personalizadas para las entradas NULL de los datos de entrada utilizando el indicador `--csvInputNULLValue=<csv-input-null>`. El cliente de cifrado de C3R puede usar codificaciones personalizadas en el archivo de salida generado para las entradas NULL utilizando el indicador `--csvOutputNULLValue=<csv-output-null>`.

 Note

Una entrada NULL se considera carente de contenido, especialmente en el contexto de un formato tabular más enriquecido, como una tabla SQL. Aunque el formato .csv no admite explícitamente esta caracterización por razones históricas, es una convención habitual considerar que una entrada vacía que contiene solo espacios en blanco es NULL. Por lo tanto, ese es el comportamiento predeterminado del cliente de cifrado de C3R, y se puede personalizar según sea necesario.

¿Cómo interpreta C3R las entradas .csv?

La siguiente tabla proporciona ejemplos de cómo se serializan las entradas .csv (de cleartext a cleartext, para mayor claridad) en función de los valores (si los hay) que se proporcionan para los indicadores `--csvInputNULLValue=<csv-input-null>` y `--csvOutputNULLValue=<csv-output-null>`. Los espacios en blanco iniciales y finales fuera de las comillas se recortan antes de que C3R interprete el significado de cualquier valor.

<code><csv-input-null></code>	<code><csv-output-null></code>	Entrada de entrada	Entrada de salida
Ninguno	Ninguna	,AnyProduct,	,AnyProduct,
Ninguna	Ninguna	, AnyProduct ,	,AnyProduct,
Ninguna	Ninguna	,"AnyProduct",	,AnyProduct,
Ninguna	Ninguna	, "AnyProdu ct" ,	,AnyProduct,
Ninguna	Ninguna	,,	,,
Ninguna	Ninguna	, ,	,,
Ninguna	Ninguna	, "",	,,
Ninguna	Ninguna	, " ",	, " ",
Ninguna	Ninguna	, " " ,	, " " ,

<csv-input-null>	<csv-output-null>	Entrada de entrada	Entrada de salida
"AnyProduct"	"NULL"	,AnyProduct,	,NULL,
"AnyProduct"	"NULL"	, AnyProduct ,	,NULL,
"AnyProduct"	"NULL"	,"AnyProduct",	,NULL,
"AnyProduct"	"NULL"	, "AnyProdu ct" ,	,NULL,
Ninguna	"NULL"	,,	,NULL,
Ninguna	"NULL"	, ,	,NULL,
Ninguna	"NULL"	, "",	,NULL,
Ninguna	"NULL"	, " ",	, " ",
Ninguno	"NULL"	, " " ,	, " ",
""	"NULL"	,,	,NULL,
""	"NULL"	, ,	,NULL,
""	"NULL"	, "",	, "",
""	"NULL"	, " ",	, " ",
""	"NULL"	, " " ,	, " ",
"\\\""	"NULL"	,,	,,
"\\\""	"NULL"	, ,	,,
"\\\""	"NULL"	, "",	,NULL,
"\\\""	"NULL"	, " ",	, " ",
"\\\""	"NULL"	, " " ,	, " ",

Archivo CSV sin encabezados

No es necesario que el archivo .csv de origen tenga encabezados en la primera fila que asignen un nombre exclusivo a cada columna. Sin embargo, un archivo .csv sin una fila de encabezados requiere un esquema de cifrado posicional. Se requiere el esquema de cifrado posicional en lugar del esquema mapeado típico que se usa tanto para los archivos .csv con una fila de encabezado como para los archivos de Parquet.

Un esquema de cifrado posicional especifica las columnas de salida por posición en lugar de por nombre. Un esquema de cifrado mapeado asigna los nombres de columnas de origen a nombres de columnas de destino. Para obtener más información, incluida una explicación detallada y ejemplos de ambos formatos de esquema, consulte [Esquemas de tablas mapeados y posicionales](#).

Archivos de Parquet

Se presupone que un archivo con la extensión .parquet tiene formato de Apache Parquet.

Tipos de datos Parquet admitidos

El cliente de cifrado de C3R puede procesar cualquier dato no complejo (es decir, de tipo primitivo) de un archivo Parquet que corresponda a un tipo de datos admitido por AWS Clean Rooms.

Sin embargo, en las columnas sealed solo se pueden usar columnas de cadena.

Se admiten los siguientes tipos de datos de Parquet:

- Tipo primitivo Binary con las siguientes anotaciones lógicas:
 - Ninguno si `--parquetBinaryAsString` está definido (tipo de datos STRING)
 - `Decimal(scale, precision)` (tipo de datos DECIMAL)
 - `String` (tipo de datos STRING)
- Tipo de datos primitivo Boolean sin anotación lógica (tipo de datos BOOLEAN)
- Tipo de datos primitivo Double sin anotación lógica (tipo de datos DOUBLE)
- Tipo primitivo `Fixed_Len_Binary_Array` con anotación lógica `Decimal(scale, precision)` (tipo de datos DECIMAL)
- Tipo de datos primitivo Float sin anotación lógica (tipo de datos FLOAT)
- Tipo primitivo Int32 con las siguientes anotaciones lógicas:
 - Ninguna (tipo de datos INT)

- `Date` (tipo de datos `DATE`)
- `Decimal(scale, precision)` (tipo de datos `DECIMAL`)
- `Int(16, true)` (tipo de datos `SMALLINT`)
- `Int(32, true)` (tipo de datos `INT`)
- Tipo de datos primitivo `Int64` con las siguientes anotaciones lógicas:
 - Ninguna (tipo de datos `BIGINT`)
 - `Decimal(scale, precision)` (tipo de datos `DECIMAL`)
 - `Int(64, true)` (tipo de datos `BIGINT`)
 - `Timestamp(isUTCAdjusted, TimeUnit.MILLIS)` (tipo de datos `TIMESTAMP`)
 - `Timestamp(isUTCAdjusted, TimeUnit.MICROS)` (tipo de datos `TIMESTAMP`)
 - `Timestamp(isUTCAdjusted, TimeUnit.NANOS)` (tipo de datos `TIMESTAMP`)

Cifrar valores que no son de cadena

Actualmente, solo se admiten valores de cadena en las columnas `sealed`.

En el caso de los archivos `.csv`, el cliente de cifrado de C3R trata todos los valores como texto codificado en UTF-8 y no intenta interpretarlos de forma diferente antes del cifrado.

En el caso de las columnas de huella digital, los tipos se agrupan en clases de equivalencia. Una clase de equivalencia es un conjunto de tipos de datos que se pueden comparar de forma inequívoca para determinar su igualdad mediante un tipo de datos representativo.

Las clases de equivalencia permiten asignar huellas dactilares idénticas al mismo valor semántico independientemente de la representación original. Sin embargo, el mismo valor en dos clases de equivalencia no dará como resultado la misma columna de huella digital.

Por ejemplo, al valor `INTEGRAL 42` se le asignará la misma huella digital independientemente de si originalmente era `SMALLINT`, `INT` o `BIGINT`. Además, el valor `INTEGRAL 0` nunca coincidirá con el valor `BOOLEAN FALSE` (que se representa mediante el valor `0`).

Las columnas de huellas digitales admiten las siguientes clases de equivalencia y los tipos de AWS Clean Rooms datos correspondientes:

Clase de equivalencia	Tipo de datos AWS Clean Rooms admitido
BOOLEAN	BOOLEAN
DATE	DATE
INTEGRAL	BIGINT, INT, SMALLINT
STRING	CHAR, STRING, VARCHAR

Nombres de columnas en computación criptográfica para Clean Rooms

De forma predeterminada, los nombres de columnas son importantes en computación criptográfica para Clean Rooms.

Si el valor del parámetro Permitir JOIN de columnas con nombres diferentes es false, se utilizan los nombres de columna durante el cifrado de las columnas fingerprint. Por este motivo, de forma predeterminada, los colaboradores deben coordinarse con antelación y utilizar los mismos nombres de columna de destino para los datos que se utilizarán en las instrucciones JOIN en las consultas. De forma predeterminada, las columnas cifradas para JOIN con nombres diferentes no ejecutan correctamente la instrucción JOIN en ningún valor.

Si el valor del parámetro Permitir JOIN de columnas con nombres diferentes es true, las instrucciones JOIN en columnas cifradas como columnas fingerprint se ejecutan correctamente. Cifrar los datos con este parámetro puede permitir cierta inferencia de los valores cleartext. Por ejemplo, si una fila tiene el mismo valor de código de autenticación de mensajes basado en hash (HMAC) tanto en la columna City como en la columna State, el valor podría ser New York.

Normalización de los nombres de encabezado de columna

El cliente de cifrado de C3R normaliza los nombres de los encabezados de columna. Se eliminan todos los espacios en blanco iniciales y finales, y el nombre de la columna se cambia a minúsculas en el resultado transformado.

La normalización se aplica antes de cualquier otro cómputo, cálculo u otra operación que pueda verse afectada por los nombres de columna. El archivo de salida emitido contiene solo los nombres normalizados.

Tipos de columnas en computación criptográfica para Clean Rooms

En este tema se proporciona información sobre los tipos de columnas en computación criptográfica para Clean Rooms.

Temas

- [Columnas Fingerprint](#)
- [Columnas selladas](#)
- [Columnas Cleartext](#)

Columnas Fingerprint

Las columnas Fingerprint son columnas que están protegidas criptográficamente para su uso en instrucciones JOIN.

Los datos de las columnas fingerprint no se pueden descifrar. Solo se pueden descifrar los datos de las columnas selladas.

Las columnas Fingerprint solo deben usarse en las siguientes cláusulas y funciones SQL:

- JOIN (INNER, OUTER, LEFT, RIGHT, or FULL) frente a otras columnas fingerprint:
 - Si el valor del parámetro `allowJoinsOnColumnsWithDifferentNames` se establece en `false`, ambas columnas fingerprint de la JOIN deben tener también el mismo nombre.
- SELECT COUNT()
- SELECT COUNT(DISTINCT)
- GROUP BY (usar solo si la colaboración ha definido el valor del parámetro `preserveNulls` como `true`).

Las consultas que infrinjan estas restricciones pueden arrojar resultados incorrectos.

Columnas selladas

Las columnas selladas son columnas que están protegidas criptográficamente para su uso en instrucciones SELECT.

Las columnas selladas solo deben usarse en las siguientes cláusulas y funciones SQL:

- SELECT

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

 Note

`SELECT COUNT(DISTINCT )` no se admite.

Las consultas que infrinjan estas restricciones pueden arrojar resultados incorrectos.

Rellenar datos de una columna sealed antes del cifrado

Cuando se especifica que una columna debe ser una columna sealed, C3R pregunta qué tipo de relleno se desea elegir. El relleno de datos antes del cifrado es opcional. Sin relleno (un relleno de tipo `none`), la longitud de los datos cifrados indica el tamaño del cleartext. En determinadas circunstancias, el tamaño del cleartext podría dejar revelar el texto sin formato. Con el relleno (relleno de tipo `fixed` o `max`), todos los valores se rellenan primero hasta alcanzar un tamaño común, y entonces se cifran. Con el relleno, la longitud de los datos cifrados no proporciona información sobre la longitud del cleartext original, salvo por indicar un límite superior de su tamaño.

Si desea rellenar una columna y conoce la longitud máxima en bytes de los datos de esa columna, utilice el relleno `fixed`. Utilice un valor `length` que sea al menos tan grande como la mayor longitud en bytes de esa columna.

 Note

Si un valor supera la `length` proporcionada, se produce un error y el cifrado falla.

Si desea rellenar una columna y no conoce la longitud máxima en bytes de los datos de esa columna, utilice el relleno `max`. Este modo de relleno rellena todos los datos hasta la longitud del valor más largo, más los bytes de `length` adicionales.

 Note

Es posible que desee cifrar los datos por lotes o actualizar sus tablas con nuevos datos de forma periódica. Tenga en cuenta que el relleno `max` rellena las entradas hasta la longitud (más el byte de `length`) de la entrada de texto sin formato más larga de un lote determinado. Esto significa que la longitud del texto cifrado puede variar de un lote a otro.

Por lo tanto, si conoce la longitud máxima en bytes de una columna, entonces debe utilizar el relleno `fixed` en lugar de `max`.

Columnas Cleartext

Las columnas Cleartext son columnas que no están protegidas criptográficamente para su uso en instrucciones JOIN o SELECT.

Las columnas Cleartext se pueden usar en cualquier parte de la consulta SQL.

Parámetros de computación criptográfica

Hay parámetros de computación criptográfica disponibles para las colaboraciones que utilizan la computación criptográfica para Clean Rooms (C3R) al [crear una colaboración](#). Puede crear una colaboración mediante la AWS Clean Rooms consola o la operación de la `CreateCollaboration` API. En la consola, puede establecer valores para los parámetros en Parámetros de computación criptográfica después de activar la opción Admitir computación criptográfica. Para obtener más información, consulte los siguientes temas.

Temas

- [Parámetro Permitir columnas cleartext](#)
- [Parámetro Permitir duplicados](#)
- [Parámetro Permitir JOIN de columnas con nombres diferentes](#)
- [Parámetro Conservar valores NULL](#)

Parámetro Permitir columnas cleartext

En la consola, puede configurar el parámetro Permitir columnas cleartext al [crear una colaboración](#) para especificar si se permiten los datos cleartext en una tabla con datos cifrados.

En la siguiente tabla se describen los valores del parámetro Permitir columnas cleartext.

Valor del parámetro	Description (Descripción)
No	No se permiten columnas Cleartext en la tabla cifrada. Todos los datos están protegidos criptográficamente.

Valor del parámetro	Description (Descripción)
Sí	Se permiten columnas Cleartext en la tabla cifrada. Las columnas Cleartext no están protegidas criptográficamente y se incluyen como cleartext. Debe tomar nota de lo que los datos cleartext de las filas pueden revelar sobre el resto de datos de la tabla. Para ejecutar SUM o AVG en columnas específicas, las columnas deben estar en cleartext.

Mediante la operación `CreateCollaboration` de la API, para el parámetro `dataEncryptionMetadata`, puede establecer el valor de `allowCleartext` en `true` o `false`. Para obtener más información sobre las operaciones de la API, consulte [Referencia de la API de AWS Clean Rooms](#).

Las columnas Cleartext corresponden a columnas que se clasifican como cleartext en el esquema específico de la tabla. Los datos de estas columnas no están cifrados y se pueden utilizar de cualquier forma. Cleartext las columnas pueden resultar útiles si los datos no son confidenciales and/or si se necesita más flexibilidad de la que permite una sealed fingerprint column o columnas cifradas.

Parámetro Permitir duplicados

En la consola, puede configurar el parámetro Permitir duplicados al [crear una colaboración](#) para especificar si las columnas cifradas para consultas JOIN pueden contener valores no NULL duplicados.

Important

Los parámetros Permitir duplicados, [Permitir JOIN de columnas con nombres diferentes](#) y [Conservar los valores NULL](#) tienen efectos distintos pero relacionados.

En la siguiente tabla se describen los valores del parámetro Permitir duplicados.

Valor del parámetro	Description (Descripción)
No	No se permiten valores repetidos en una columna fingerprint. Todos los valores de una misma columna fingerprint deben ser únicos.
Sí	Se permiten valores repetidos en una columna fingerprint. Si necesita combinar columnas con valores repetidos, defina este valor en Sí. Si se establece en Sí, los patrones de frecuencia que aparecen en las columnas fingerprint de la tabla de C3R o en los resultados pueden implicar información adicional sobre la estructura de los datos cleartext.

Mediante la operación `CreateCollaboration` de la API, en el parámetro `dataEncryptionMetadata`, puede establecer el valor de `allowDuplicates` en `true` o `false`. Para obtener más información sobre las operaciones de la API, consulte [Referencia de la API de AWS Clean Rooms](#).

De forma predeterminada, si se deben usar datos cifrados en consultas JOIN, el cliente de cifrado de C3R requiere que esas columnas no tengan valores duplicados. Con este requisito se pretende aumentar la protección de los datos. Este comportamiento puede contribuir a garantizar que los patrones repetidos en los datos no sean observables. Sin embargo, si desea trabajar con datos cifrados en consultas JOIN y no le preocupan los valores duplicados, el parámetro Permitir duplicados puede desactivar esta comprobación conservadora.

Parámetro Permitir JOIN de columnas con nombres diferentes

En la consola, puede configurar el parámetro Permitir JOIN de columnas con nombres diferentes al [crear una colaboración](#) para especificar si se admiten las instrucciones JOIN entre columnas con nombres diferentes.

Para obtener más información, consulte [Normalización de los nombres de encabezado de columna](#)

En la siguiente tabla se describen los valores del parámetro Permitir JOIN de columnas con nombres diferentes.

Valor del parámetro	Description (Descripción)
No	No se admiten las combinaciones de columnas fingerprint con nombres diferentes. Las instrucciones JOIN solo proporcionan resultados precisos en las columnas que tienen el mismo nombre.  Important El valor No proporciona una mayor seguridad de la información, pero requiere que los participantes en la colaboración acuerden previamente los nombres de las columnas. Si dos columnas tienen nombres diferentes cuando se cifran como columnas fingerprint y la opción Permitir JOIN de columnas con nombres diferentes está establecida en No, las instrucciones JOIN en esas columnas no producen ningún resultado. Esto se debe a que, después del cifrado, no se comparten valores entre ellas.
Sí	Se admiten las combinaciones de columnas fingerprint con nombres diferentes. Para mayor flexibilidad, los usuarios pueden establecer este valor en Sí, lo que permite las instrucciones JOIN en las columnas independientemente del nombre de la columna. Si se establece en Sí, el cliente de cifrado de C3R no tiene en cuenta el nombre de la columna al proteger las columnas fingerprint. Como resultado, en la tabla de C3R se pueden observar valores comunes en distintas columnas fingerprint. Por ejemplo, si una fila tiene el mismo valor JOIN cifrado tanto en una columna City como en una columna State, sería razonable inferir que ese valor es New York.

Mediante la operación `CreateCollaboration` de la API, para el parámetro `dataEncryptionMetadata`, puede establecer el valor de `allowJoinsOnColumnsWithDifferentNames` en `true` o `false`. Para obtener más información sobre las operaciones de la API, consulte [Referencia de la API de AWS Clean Rooms](#).

De forma predeterminada, el cifrado de las columnas `fingerprint` se ve afectado por el `targetHeader` de esa columna, establecido en [Paso 4: generar un esquema de cifrado para un archivo tabular](#). Por lo tanto, el mismo valor `cleartext` tiene diferentes representaciones cifradas en cada columna `fingerprint` diferente en la que se cifra.

En algunos casos, este parámetro puede resultar útil para evitar la inferencia de los valores `cleartext`. Por ejemplo, observar el mismo valor cifrado en las columnas `fingerprint City` y `State` podría llevar a inferir de forma razonable que el valor es New York. Sin embargo, el uso de este parámetro requiere una coordinación adicional por adelantado, de modo que todas las columnas que se van a combinar en consultas tengan nombres compartidos.

Puede utilizar el parámetro `Permitir JOIN` de columnas con nombres diferentes para reducir esta restricción. Si el valor del parámetro se establece en `Yes`, permite que todas las columnas cifradas para `JOIN` se utilicen juntas independientemente del nombre.

Parámetro Conservar valores NULL

En la consola, puede configurar el parámetro `Conservar valores NULL` al [crear una colaboración](#) para indicar que no hay ningún valor presente en esa columna.

En la siguiente tabla se describen los valores del parámetro `Conservar valores NULL`.

Valor del parámetro	Description (Descripción)
No	Los valores NULL no se conservan. Los valores NULL no aparecen como NULL en una tabla cifrada. Los valores NULL aparecen como valores aleatorios únicos en una tabla de C3R.
Sí	Los valores NULL se conservan. Los valores NULL aparecen como NULL en una tabla cifrada. Si se requiere la semántica SQL de los NULL, puede establecer este valor en Sí. Como resultado, las entradas NULL aparecen como NULL en la tabla de C3R, independientemente de si la columna está cifrada o de la configuración del parámetro <code>Permitir duplicados</code> .

Mediante la operación `CreateCollaboration` de la API, para el parámetro `dataEncryptionMetadata`, puede establecer el valor de `preserveNulls` en `true` o `false`. Para obtener más información sobre las operaciones de la API, consulte [Referencia de la API de AWS Clean Rooms](#).

Si el parámetro Conservar valores NULL está establecido en No para la colaboración:

1. Las entradas NULL de las columnas `cleartext` permanecen inalteradas.
2. Las entradas NULL de las columnas `fingerprint` cifradas se cifran como valores aleatorios para ocultar su contenido. Combinar una columna cifrada con entradas NULL de la columna `cleartext` no produce ninguna coincidencia para ninguna de las entradas NULL. No se realizan coincidencias porque cada una recibe su propio contenido aleatorio único.
3. Las entradas NULL de las columnas `sealed` cifradas se cifran.

Cuando el valor del parámetro Conservar valores NULL se establece en Sí para la colaboración, las entradas NULL de todas las columnas se mantienen como NULL, independientemente de si la columna está cifrada o no.

El parámetro Conservar valores NULL resulta útil en escenarios tales como el enriquecimiento de datos, donde se desea compartir una falta de información expresada como NULL. El parámetro Conservar valores NULL también es útil en formato `fingerprint` o HMAC cuando se tienen valores NULL en la columna que se desea ejecutar `JOIN` o `GROUP BY`.

Si el valor de los parámetros Permitir duplicados y Conservar valores NULL se establece en No, tener más de una entrada NULL en una columna `fingerprint` produce un error y detiene el cifrado. Si el valor de cualquiera de los parámetros se establece en Sí, no se produce tal error.

Indicadores opcionales en computación criptográfica para Clean Rooms

En las siguientes secciones se describen los indicadores opcionales que puede configurar al [cifrar datos](#) utilizando el cliente de cifrado de C3R para personalizar y probar archivos tabulares.

Temas

- [Indicador `--csvInputNULLValue`](#)
- [Indicador `--csvOutputNULLValue`](#)
- [Indicador `--enableStackTraces`](#)
- [Indicador `--dryRun`](#)

- [Indicador --tempDir](#)

Indicador **--csvInputNULLValue**

Puede usar el indicador `--csvInputNULLValue` para especificar codificaciones personalizadas para las entradas NULL de los datos de entrada al [cifrar datos](#) utilizando el cliente de cifrado de C3R.

En la siguiente tabla se resumen el uso y los parámetros de este indicador.

De uso	Parameters
Opcional. Los usuarios pueden especificar codificaciones personalizadas para las entradas NULL de los datos de entrada.	Codificación especificada por el usuario de valores NULL en el archivo CSV de entrada

Una entrada NULL es una entrada que se considera carente de contenido, específicamente en el contexto de un formato tabular más enriquecido, como una tabla SQL. Aunque el formato .CSV no admite explícitamente esta caracterización por razones históricas, es una convención habitual considerar que una entrada vacía que contiene solo espacios en blanco es NULL. Por lo tanto, ese es el comportamiento predeterminado del cliente de cifrado de C3R, y se puede personalizar según sea necesario.

Indicador **--csvOutputNULLValue**

Puede usar el indicador `--csvOutputNULLValue` para especificar codificaciones personalizadas para las entradas NULL en los datos de salida al [cifrar datos](#) con el cliente de cifrado de C3R.

En la siguiente tabla se resumen el uso y los parámetros de este indicador.

De uso	Parameters
Opcional. Los usuarios pueden especificar codificaciones personalizadas en el archivo de salida generado para las entradas NULL.	Codificación especificada por el usuario de valores NULL en el archivo CSV de salida

Una entrada NULL es una entrada que se considera carente de contenido, específicamente en el contexto de un formato tabular más enriquecido, como una tabla SQL. Aunque el formato .CSV

no admite explícitamente esta caracterización por razones históricas, es una convención habitual considerar que una entrada vacía que contiene solo espacios en blanco es NULL. Por lo tanto, ese es el comportamiento predeterminado del cliente de cifrado de C3R, y se puede personalizar según sea necesario.

Indicador **--enableStackTraces**

Al [cifrar datos](#) utilizando el cliente de cifrado de C3R, utilice el indicador `--enableStackTraces` para proporcionar información contextual adicional para notificar errores cuando C3R detecte un error.

AWS no recopila errores. Si encuentras un error, utiliza el seguimiento de pila para solucionar el error tú mismo o envía el seguimiento de pila a Soporte para obtener ayuda.

En la siguiente tabla se resumen el uso y los parámetros de este indicador.

De uso	Parameters
Opcional. Se utiliza para proporcionar información contextual adicional para notificar errores cuando el cliente de cifrado de C3R detecta un error.	Ninguno

Indicador **--dryRun**

Los comandos [encrypt](#) y [decrypt](#) del cliente de cifrado de C3R incluyen un indicador `--dryRun` opcional. El indicador toma todos los argumentos proporcionados por el usuario y comprueba su validez y coherencia.

Puede usar el indicador `--dryRun` para comprobar si su archivo de esquema es válido y coherente con el archivo de entrada correspondiente.

En la siguiente tabla se resumen el uso y los parámetros de este indicador.

De uso	Parameters
Opcional. Hace que el cliente de cifrado de C3R analice los parámetros y compruebe los	Ninguno

De uso	Parameters
archivos, pero no realiza ningún cifrado ni descifrado.	

Indicador **--tempDir**

Es posible que desee utilizar un directorio temporal, ya que los archivos cifrados a veces pueden superar en tamaño a los no cifrados, dependiendo de la configuración. Los conjuntos de datos también deben cifrarse por colaboración para funcionar correctamente.

Al [cifrar los datos](#) utilizando C3R, utilice el indicador `--tempDir` para especificar la ubicación en la que se pueden crear los archivos temporales mientras se procesa la entrada.

En la siguiente tabla se resumen el uso y los parámetros de este indicador.

De uso	Parameters
Los usuarios pueden especificar la ubicación en la que se pueden crear los archivos temporales mientras se procesa la entrada.	El valor predeterminado es el directorio temporal del sistema.

Consultas con computación criptográfica para Clean Rooms

En este tema se proporciona información sobre cómo escribir consultas que utilicen tablas de datos cifradas mediante computación criptográfica para Clean Rooms.

Temas

- [Consultas que se ramifican en NULL](#)
- [Asignar una columna de origen a varias columnas de destino](#)
- [Usar los mismos datos para las consultas JOIN y SELECT](#)

Consultas que se ramifican en NULL

Tener una ramificación de consulta en una instrucción NULL significa usar una sintaxis similar a `IF x IS NULL THEN 0 ELSE 1`.

Las consultas siempre se pueden ramificar en instrucciones NULL en columnas de cleartext.

Las consultas se pueden ramificar en instrucciones NULL en columnas sealed y en columnas fingerprint solo cuando el valor del parámetro Conservar valores NULL (`preserveNulls`) está establecido en `true`.

Las consultas que infrinjan estas restricciones pueden arrojar resultados incorrectos.

Asignar una columna de origen a varias columnas de destino

Una columna de origen puede asignarse a varias columnas de destino. Por ejemplo, puede que desee utilizar tanto JOIN como SELECT en una columna.

Para obtener más información, consulte [Usar los mismos datos para las consultas JOIN y SELECT](#).

Usar los mismos datos para las consultas JOIN y SELECT

Si los datos de una columna no son confidenciales, pueden aparecer en una columna de destino de cleartext, lo que permite utilizarlos para cualquier fin.

Si los datos de una columna son confidenciales y deben usarse tanto para las consultas JOIN como para las consultas SELECT, asigne dicha columna de origen a dos columnas de destino del archivo de salida. Una columna se cifra con el type como columna fingerprint y la otra columna se cifra con el type como columna sellada. La generación de esquemas interactivos del cliente de cifrado C3R sugiere sufijos de encabezado de `_fingerprint` y `_sealed`. Estos sufijos de encabezado pueden ser una convención útil para diferenciar dichas columnas rápidamente.

Directrices para el cliente de cifrado de C3R

El cliente de cifrado de C3R es una herramienta que permite a las organizaciones recopilar datos confidenciales para obtener nueva información procesable a partir del análisis de datos. La herramienta limita criptográficamente lo que puede aprender cualquier parte y AWS en el proceso. Si bien esto es de vital importancia, el proceso de protección criptográfica de los datos puede suponer una sobrecarga considerable, tanto en términos de recursos de computación como de almacenamiento. Por lo tanto, es importante entender las ventajas y desventajas de usar cada ajuste y saber cómo optimizar la configuración sin dejar de mantener las garantías criptográficas deseadas. Este tema se centra en las implicaciones para el rendimiento de las diferentes configuraciones del cliente de cifrado de C3R y los esquemas.

Todas las configuraciones de cifrado del cliente de cifrado de C3R ofrecen diferentes garantías criptográficas. De forma predeterminada, los ajustes más seguros son aquellos en el nivel de la

colaboración. Al habilitar funciones adicionales al crear una colaboración, se debilitan las garantías de privacidad, ya que se permite realizar actividades como análisis de frecuencia en el texto cifrado. Para obtener más información sobre cómo se utilizan estas configuraciones y cuáles son sus implicaciones, consulte [the section called “Computación criptográfica”](#).

Temas

- [Implicaciones en el rendimiento de los tipos de columnas](#)
- [Solución de problemas relacionados con el aumento imprevisto de tamaño del texto cifrado](#)

Implicaciones en el rendimiento de los tipos de columnas

C3R utiliza tres tipos de columnas: cleartext, fingerprint y sealed. Cada uno de estos tipos de columnas proporciona diferentes garantías criptográficas y tiene distintos usos previstos. En las siguientes secciones se analizan las implicaciones de rendimiento del tipo de columna y el impacto de cada ajuste en el rendimiento.

Temas

- [Columnas Cleartext](#)
- [Columnas Fingerprint](#)
- [Columnas Sealed](#)

Columnas Cleartext

Las columnas Cleartext no se modifican con respecto a su formato original ni se procesan criptográficamente en modo alguno. Este tipo de columna no se puede configurar y no afecta al rendimiento en términos de almacenamiento o de computación.

Columnas Fingerprint

Las columnas Fingerprint están previstas para utilizarse para combinar datos de varias tablas. Para ello, el tamaño del texto cifrado resultante debe ser siempre el mismo. Sin embargo, estas columnas se ven afectadas por los ajustes en el nivel de la colaboración. Las columnas Fingerprint pueden incidir en distinto grado en el tamaño del archivo de salida en función del cleartext contenido en la entrada.

Temas

- [Sobrecarga de base para las columnas fingerprint](#)

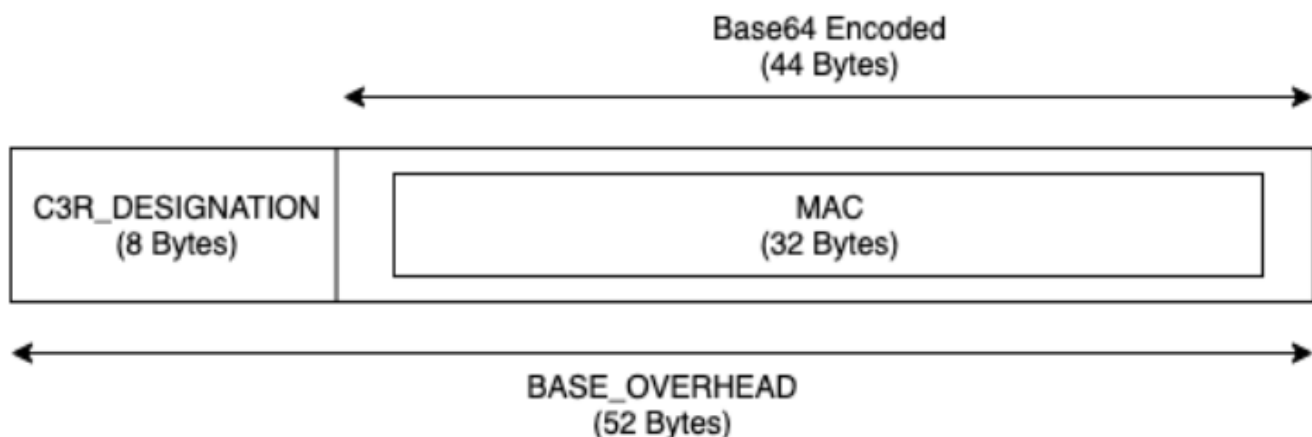
- [Ajustes de colaboración para las columnas fingerprint](#)
- [Datos de ejemplo para una columna fingerprint](#)
- [Solución de problemas con columnas fingerprint](#)

Sobrecarga de base para las columnas fingerprint

Existe una sobrecarga de base para las columnas fingerprint. Esta sobrecarga es constante y sustituye al tamaño de los bytes de cleartext.

Los datos de las columnas fingerprint se procesan criptográficamente mediante una función de código de autenticación de mensajes basado en hash (HMAC), que convierte los datos en un código de autenticación de mensajes (MAC) de 32 bytes. A continuación, estos datos se procesan a través de un codificador base64, lo que aumenta el tamaño del byte aproximadamente un 33 por ciento. Va precedido de una designación C3R de 8 bytes para designar el tipo de columna a la que pertenecen los datos y la versión de cliente que los generó. El resultado final es de 52 bytes. Este resultado se multiplica entonces por el número de filas para obtener la sobrecarga de base total (utilice el número total de valores distintos de null si `preserveNulls` se establece en true).

En la siguiente imagen se muestra cómo $BASE\_OVERHEAD = C3R\_DESIGNATION + (MAC * 1.33)$



El texto cifrado de salida de las columnas fingerprint siempre será de 52 bytes. Esto puede suponer una disminución significativa del almacenamiento si el promedio de los datos de cleartext de entrada es superior a 52 bytes (por ejemplo, las direcciones postales completas). Esto puede suponer un aumento significativo del almacenamiento si el promedio de los datos de cleartext de entrada es inferior a 52 bytes (por ejemplo, las edades de los clientes).

Ajustes de colaboración para las columnas fingerprint

Ajuste **preserveNulls**

Cuando el ajuste de nivel de la colaboración `preserveNulls` es `false` (predeterminado), cada valor `null` se sustituye por 32 bytes únicos y aleatorios y se procesa como si no fuera `null`. El resultado es que cada valor `null` tiene ahora 52 bytes. Esto puede añadir requisitos de almacenamiento importantes para las tablas que contienen datos muy dispersos en comparación con cuando este ajuste es `true` y los valores `null` se transfieren como `null`.

Si no necesita las garantías de privacidad de este ajuste y prefiere retener los valores `null` de sus conjuntos de datos, habilite el ajuste `preserveNulls` en el momento de crear la colaboración. El ajuste `preserveNulls` no se puede cambiar una vez creada la colaboración.

Datos de ejemplo para una columna fingerprint

El siguiente es un ejemplo de conjunto de datos de entrada y salida para una columna fingerprint con los ajustes de configuración que se deben reproducir. Otros ajustes a nivel de la colaboración, como `allowCleartext` y `allowDuplicates`, no afectan a los resultados y se pueden configurar como `true` o `false` si se intenta reproducirlos localmente.

Secreto compartido de ejemplo: `wJa1rXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY`

ID de colaboración de ejemplo: `a1b2c3d4-5678-90ab-cdef-EXAMPLE11111`

`allowJoinsOnColumnsWithDifferentNames`: `True` Este ajuste no afecta al rendimiento ni a los requisitos de almacenamiento. Sin embargo, este ajuste hace que la elección del nombre de la columna sea irrelevante al reproducir los valores que se muestran en las siguientes tablas.

Ejemplo 1

Input	<code>null</code>
<code>preserveNulls</code>	<code>TRUE</code>
Salida	<code>null</code>
Determinista	<code>Yes</code>
Bytes de entrada	<code>0</code>
Bytes de salida	<code>0</code>

Ejemplo 2

Entrada	null
preserveNulls	FALSE
Salida	01:hmac:3lkFjthvV3IUu6mMvFc1a +XAHwgw/E1m0q4p3Yg25kk=
Determinista	No
Bytes de entrada	0
Bytes de salida	52

Ejemplo 3

Input	empty string
preserveNulls	-
Salida	01:hmac:oKTgi3Gba+eUb3JteSz 2EMgXUkF1WgM77UP0Ydw5kPQ=
Determinista	Yes
Bytes de entrada	0
Bytes de salida	52

Ejemplo 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Salida	01:hmac:kU/IqwG7FMmzzshr0B9 scomE0UJUEE7j9keTctplGww=
Determinista	Yes

Bytes de entrada	26
Bytes de salida	52

Ejemplo 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Salida	01:hmac:ks3htnQbw2vdhCRFF6J NzW5LMndJaHG57uvE26mBtSs=
Determinista	Yes
Bytes de entrada	62
Bytes de salida	52

Solución de problemas con columnas fingerprint

¿Por qué el texto cifrado de mis columnas fingerprint es varias veces mayor que el tamaño del cleartext que figuraba en ellas?

El texto cifrado de una columna de fingerprint tiene siempre una longitud de 52 bytes. Si los datos de entrada eran pequeños (por ejemplo, las edades de los clientes), mostrarán un aumento considerable de tamaño. Esto también puede ocurrir si el ajuste `preserveNulls` se define como `false`.

¿Por qué el texto cifrado de mis columnas de fingerprint es varias veces más pequeño que el cleartext que contiene?

El texto cifrado de una columna de fingerprint tiene siempre una longitud de 52 bytes. Si los datos de entrada eran grandes (por ejemplo, las direcciones completas de los clientes), su tamaño se reducirá considerablemente.

¿Cómo sé si necesito las garantías criptográficas que ofrece **preserveNulls**?

Lamentablemente, la respuesta es que depende. Como mínimo, se deben revisar los [the section called “Parameters”](#) en cuanto a la forma en que el ajuste `preserveNulls` protege sus datos. No obstante, le recomendamos que consulte los requisitos de manejo de datos de su organización y cualquier contrato aplicable a la colaboración correspondiente.

¿Por qué tengo que incurrir en la sobrecarga de base64?

Para permitir la compatibilidad con los formatos de archivo tabulares como CSV, es necesaria la codificación base64. Si bien algunos formatos de archivo como Parquet podrían admitir representaciones binarias de datos, es importante que todos los participantes en una colaboración representen los datos de la misma manera para garantizar que los resultados de las consultas sean correctos.

Columns Sealed

Las columnas Sealed están diseñadas para utilizarse para transferir datos entre los miembros de una colaboración. El texto cifrado de estas columnas no es determinista y tiene un impacto significativo tanto en el rendimiento como en el almacenamiento en función de cómo se configuren las columnas. Estas columnas se pueden configurar de forma individual y, a menudo, son las que más influyen en el rendimiento del cliente de cifrado de C3R y en el tamaño del archivo de salida resultante.

Temas

- [Sobrecarga de base para las columnas sealed](#)
- [Ajustes de colaboración para las columnas sealed](#)
- [Columnas sealed de configuración de esquema: tipos de relleno](#)
- [Datos de ejemplo para una columna sealed](#)
- [Solución de problemas con columnas sealed](#)

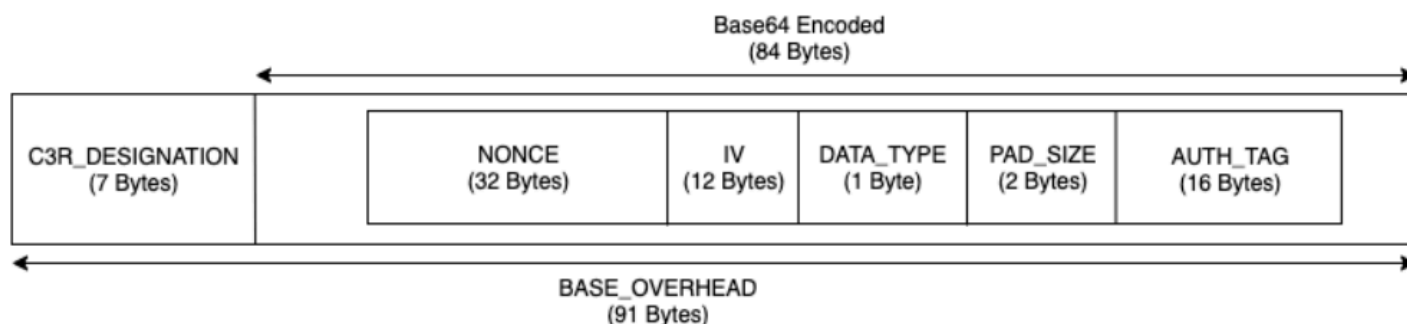
Sobrecarga de base para las columnas sealed

Existe una sobrecarga de base para las columnas sealed. Esta sobrecarga es constante y se suma al tamaño de los bytes de cleartext y de relleno (si los hubiera).

Antes de cualquier cifrado, a los datos de las columnas sealed se les añade como prefijo un carácter de 1 byte que designa el tipo de datos que contienen. Si se selecciona el relleno, los datos se rellenan y se añaden 2 bytes que indican el tamaño del relleno. Tras añadir estos bytes, los datos se procesan criptográficamente mediante AES-GCM y se almacenan con IV (12 bytes), nonce

(32 bytes) y Auth Tag (16 bytes). A continuación, estos datos se procesan a través de un codificador base64, lo que aumenta el tamaño del byte aproximadamente un 33 por ciento. Los datos van precedidos de una designación C3R de 7 bytes para indicar el tipo de columna a la que pertenecen los datos y la versión de cliente utilizada para generarlos. El resultado es una sobrecarga de base final de 91 bytes. A continuación, este resultado se puede multiplicar por el número de filas para obtener la sobrecarga de base total (utilice el número total de valores no nulos si `preserveNulls` está definido como `true`).

En la siguiente imagen se muestra cómo $BASE\_OVERHEAD = C3R\_DESIGNATION + ((NONCE + IV + DATA\_TYPE + PAD\_SIZE + AUTH\_TAG) * 1.33)$



Ajustes de colaboración para las columnas sealed

Ajuste **preserveNulls**

Cuando el ajuste en el nivel de la colaboración `preserveNulls` es `false` (predeterminado), cada valor `null` es único, aleatorio de 32 bytes, y se procesa como si no fuera `null`. El resultado es que cada valor `null` tiene ahora 91 bytes (más si se rellena). Esto puede añadir requisitos de almacenamiento importantes para las tablas que contienen datos muy dispersos en comparación con cuando este ajuste es `true` y los valores `null` se transfieren como `null`.

Si no necesita las garantías de privacidad de este ajuste y prefiere retener los valores `null` de sus conjuntos de datos, habilite el ajuste `preserveNulls` en el momento de crear la colaboración. El ajuste `preserveNulls` no se puede cambiar una vez creada la colaboración.

Columnas sealed de configuración de esquema: tipos de relleno

Temas

- [Tipo de relleno none](#)
- [Tipo de relleno fixed](#)
- [Tipo de relleno max](#)

Tipo de relleno **none**

Seleccionar el tipo de relleno **none** no añade ningún relleno al cleartext ni añade ninguna sobrecarga adicional a la sobrecarga de base descrita anteriormente. La ausencia de relleno se traduce en el tamaño de salida más eficiente en términos de espacio. Sin embargo, no proporciona las mismas garantías de privacidad que los tipos de relleno **fixed** y **max**. Esto se debe a que el tamaño del cleartext subyacente es discernible a partir del tamaño del texto cifrado.

Tipo de relleno **fixed**

Seleccionar un tipo de relleno **fixed** es una medida que salvaguarda la privacidad para ocultar la longitud de los datos contenidos en una columna. Esto se hace rellenando todo el cleartext hasta la `pad_length` antes del cifrado. Cualquier dato que supere ese tamaño provoca un fallo del cliente de cifrado de C3R.

Dado que el relleno se añade al cleartext antes de cifrarlo, AES-GCM utiliza un mapeo 1 a 1 de cleartext a texto cifrado en bytes. La codificación base64 añadirá un 33 por ciento. La sobrecarga de almacenamiento adicional del relleno se puede calcular restando la longitud media del cleartext del valor de `pad_length` y multiplicándola por 1,33. El resultado es la sobrecarga promedio de relleno por registro. Este resultado puede entonces multiplicarse por el número de filas para obtener la sobrecarga de relleno total (utilice el número total de valores no `null` si `preserveNulls` está definido como `true`).

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

Se recomienda seleccionar la `pad_length` mínima que abarque el mayor valor de columna. Por ejemplo, si el mayor valor es de 50 bytes, basta con una `pad_length` de 50. Un valor superior a ese solo añadirá sobrecarga de almacenamiento adicional.

El relleno fijo no añade ninguna sobrecarga de computación significativa.

Tipo de relleno **max**

Seleccionar un tipo de relleno **max** es una medida que salvaguarda la privacidad para ocultar la longitud de los datos contenidos en una columna. Esto se hace rellenando todos los cleartext hasta el valor más alto de la columna, más la `pad_length` adicional, antes de cifrarla. Por lo general, el relleno **max** proporciona las mismas garantías que el relleno **fixed** para un único conjunto de datos, al tiempo que permite no conocer el valor de cleartext más alto de la columna. Sin embargo, es posible que el relleno **max** no ofrezca las mismas garantías de privacidad que el relleno **fixed** en todas las actualizaciones, ya que el valor más alto de cada conjuntos de datos podría variar.

Se recomienda seleccionar una `pad_length` adicional de 0 al usar el relleno `max`. Esta longitud rellena todos los valores para que tengan el mismo tamaño que el mayor valor de la columna. Un valor superior a ese solo añadirá sobrecarga de almacenamiento adicional.

Si conoce el valor de `cleartext` más alto de una columna determinada, le recomendamos que utilice el tipo de relleno `fixed` en su lugar. El uso del relleno `fixed` crea coherencia entre los conjuntos de datos actualizados. El uso del relleno `max` hace que cada subconjunto de datos se rellene hasta el mayor valor presente en el subconjunto.

Datos de ejemplo para una columna `sealed`

El siguiente es un ejemplo de conjunto de datos de entrada y salida para una columna `sealed` con los ajustes de configuración que se deben reproducir. Otros ajustes en el nivel de la colaboración, `allowCleartext`, `allowJoinsOnColumnsWithDifferentNames` y `allowDuplicates`, no afectan a los resultados y se pueden definir como `true` o `false` si se intentan reproducir localmente. Aunque estos son los ajustes básicos que se deben reproducir, la columna `sealed` no es determinista y los valores cambiarán de una vez a otra. El objetivo es mostrar los bytes de entrada en comparación con los bytes de salida. Los valores `pad_length` de ejemplo se han elegido de forma intencionada. Muestran que el relleno `fixed` da como resultado los mismos valores que el relleno `max` con los ajustes de `pad_length` mínima recomendada o cuando se desea un relleno adicional.

Secreto compartido de ejemplo: `wJa1rXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY`

ID de colaboración de ejemplo: `a1b2c3d4-5678-90ab-cdef-EXAMPLE11111`

Temas

- [Tipo de relleno `none`](#)
- [Tipo de relleno `fixed` \(ejemplo 1\)](#)
- [Tipo de relleno `fixed` \(ejemplo 2\)](#)
- [Tipo de relleno `max` \(ejemplo 1\)](#)
- [Tipo de relleno `max` \(ejemplo 2\)](#)

Tipo de relleno **none**

Ejemplo 1

Input	<code>null</code>

<code>preserveNulls</code>	TRUE
Salida	null
Determinista	Yes
Bytes de entrada	0
Bytes de salida	0

Ejemplo 2

Entrada	null
<code>preserveNulls</code>	FALSE
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlmDEyMzQ1Njc40TBqfRY Z98t5KU6aWfssGSPbNIJfG3iXmu 6cbCUrizuV
Determinista	No
Bytes de entrada	0
Bytes de salida	91

Ejemplo 3

Input	empty string
<code>preserveNulls</code>	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlmDEyMzQ1Njc40TBqfRY Z98t5KU6aWfstGSPeM6qR8DWC2P B2GM1X41YK
Determinista	No

Bytes de entrada	0
Bytes de salida	91

Ejemplo 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6pkx9sGL5VLDQeHzh6DmPpyWNuI=
Determinista	No
Bytes de entrada	26
Bytes de salida	127

Ejemplo 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWCv02ckr6plwtH/8tRFnn2rF91bcB9G4+n8GiRfJNmqdP4/Q0Q3cXb/pbvPcnnohrHIGSX54ua+1/JfcVjc=

Determinista	No
Bytes de entrada	62
Bytes de salida	175

Tipo de relleno **fixed** (ejemplo 1)

En este ejemplo, `pad_length` es 62 y la mayor entrada es de 62 bytes.

Ejemplo 1

Input	<code>null</code>
<code>preserveNulls</code>	TRUE
Salida	<code>null</code>
Determinista	Yes
Bytes de entrada	0
Bytes de salida	0

Ejemplo 2

Entrada	<code>null</code>
<code>preserveNulls</code>	FALSE
Salida	<code>01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcgl1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcoNpATs0GzbnLkor4L+/aSuA=</code>
Determinista	No

Bytes de entrada	0
Bytes de salida	175

Ejemplo 3

Input	empty string
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsricoLB53l07VZpA60wkuXu29CA=
Determinista	No
Bytes de entrada	0
Bytes de salida	175

Ejemplo 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfsteEE1GKEPiRzyh0h7t60mWMLTWcV02ckr6pkx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsricutBAc0+Mb9tuU2KIIHH31AWg=

Determinista	No
Bytes de entrada	26
Bytes de salida	175

Ejemplo 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ QQQ3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
Determinista	No
Bytes de entrada	62
Bytes de salida	175

Tipo de relleno **fixed** (ejemplo 2)

En este ejemplo, `pad_length` es 162 y la mayor entrada es de 62 bytes.

Ejemplo 1

Input	null
preserveNulls	TRUE
Salida	null

Determinista	Yes
Bytes de entrada	0
Bytes de salida	0

Ejemplo 2

Entrada	null
preserveNulls	FALSE
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcnkB0xbLWD7zNdAqQGR0rXoSESdW0I0vpNoGcBfv4cJbG0A3h1DvtkSSVc2B8000GppzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6uwrmwv/xAySX+xcntotL703aBTBb
Determinista	No
Bytes de entrada	0
Bytes de salida	307

Ejemplo 3

Input	empty string
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRY

	Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv84lVaT9Yd+6oQx65/+gdVT
Determinista	No
Bytes de entrada	0
Bytes de salida	307

Ejemplo 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbMlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWcV02ckr6pkx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwX5Hn1+Wyf06ks3QMaRDGSf
Determinista	No
Bytes de entrada	26

Bytes de salida	307

Ejemplo 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/Q0Q3cXb/pbvPcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrnwjkJXQZ0gPdeFX9Yr/8a1V5i
Determinista	No
Bytes de entrada	62
Bytes de salida	307

Tipo de relleno **max** (ejemplo 1)

En este ejemplo, `pad_length` es 0 y la mayor entrada es de 62 bytes.

Ejemplo 1

Input	null
preserveNulls	TRUE

Salida	null
Determinista	Yes
Bytes de entrada	0
Bytes de salida	0

Ejemplo 2

Entrada	null
preserveNulls	FALSE
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcoNpATs0GzbnLkor4L+/aSuA=
Determinista	No
Bytes de entrada	0
Bytes de salida	175

Ejemplo 3

Input	empty string
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKL0hK1+7r75Tk+Mx9jy48

	Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcoLB53l07VZp A60wkuXu29CA=
Determinista	No
Bytes de entrada	0
Bytes de salida	175

Ejemplo 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc40TBqfRY Z98t5KU6aWfstEE1GKEPiRzyh0 h7t60mWMLTWCv02ckr6pkx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcutBAc0+Mb9t uU2KIIHH31AWg=
Determinista	No
Bytes de entrada	26
Bytes de salida	175

Ejemplo 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-

Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWcV02ckr6plwtH/8tRFnn2rF91bcB9G4+n8GiRfJNmqdP4/Q0Q3cXb/pbvPcnnohrHIGSX54ua+1/JfcVjc=
Determinista	No
Bytes de entrada	62
Bytes de salida	175

Tipo de relleno **max** (ejemplo 2)

En este ejemplo, `pad_length` es 100 y la mayor entrada es de 62 bytes.

Ejemplo 1

Input	null
<code>preserveNulls</code>	TRUE
Salida	null
Determinista	Yes
Bytes de entrada	0
Bytes de salida	0

Ejemplo 2

Entrada	null
<code>preserveNulls</code>	FALSE

Salida	<pre> 01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv/xAySX+xcntotL703aBTBb </pre>
Determinista	No
Bytes de entrada	0
Bytes de salida	307

Ejemplo 3

Input	empty string
preserveNulls	-
Salida	<pre> 01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKL0hK1+7r75Tk+Mx9jy48 Fcg1y0PvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B8000Gp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv84lVaT9Yd+6oQx65/+gdVT </pre>
Determinista	No

Bytes de entrada	0
Bytes de salida	307

Ejemplo 4

Input	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRYZ98t5KU6aWfstEE1GKEPiRzyh0h7t60mWMLTWcV02ckr6pkx9jy48Fcg1y0PvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcnkB0xbLWD7zNdAqQGR0rXoSESdW0I0vpNoGcBfv4cJbG0A3h1DvtkSSVc2B8000GppzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6uwrmtX5Hn1+Wyf06ks3QMaRDGSf
Determinista	No
Bytes de entrada	26
Bytes de salida	307

Ejemplo 5

Input	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
Salida	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc40TBqfRY

```

Z98t5KU6aWfstEE1GKEPiRzyh0
h7t60mWMLTWcV02ckr6plwtH/8t
RFnn2rF91bcB9G4+n8GiRfJNmqd
P4/Q0Q3cXb/pbvPcnkB0xbLWD7z
NdAqQGR0rXoSESdW0I0vpNoGcBf
v4cJbG0A3h1DvtkSSVc2B8000Gp
pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn
+8o4WtG/ClipNcjDXvXVtK4vfCohcCA6
uwrmwjkJXQZ0gPdeFX9Yr/8a1V5i

```

Determinista	No
Bytes de entrada	62
Bytes de salida	307

Solución de problemas con columnas sealed

¿Por qué el texto cifrado de mis columnas sealed es varias veces mayor que el tamaño del cleartext que figuraba en ellas?

Esto depende de varios factores. Por un lado, el texto cifrado de una columna Cleartext siempre tiene una longitud mínima de 91 bytes. Si los datos de entrada eran pequeños (por ejemplo, las edades de los clientes), mostrarán un aumento considerable de tamaño. En segundo lugar, si `preserveNulls` se hubiera definido como `false` y los datos de entrada contuvieran un gran número de valores `null`, cada uno de esos valores `null` se habrá convertido en 91 bytes de texto cifrado. Por último, si se utiliza el relleno, por definición, se añaden bytes a los datos de cleartext antes de cifrarlos.

La mayoría de los datos de una columna sealed son muy pequeños y necesito utilizar relleno. ¿Puedo simplemente eliminar los valores grandes y procesarlos por separado para ahorrar espacio?

No le recomendamos que elimine los valores grandes y los procese por separado. Al hacerlo, se modifican las garantías de privacidad que ofrece el cliente de cifrado de C3R. Como modelo de amenaza, presuponga que un observador puede ver ambos conjuntos de datos cifrados. Si el observador ve que un subconjunto de datos tiene una columna considerablemente más o menos rellena que otro subconjunto, puede hacer inferencias sobre el tamaño de los datos de cada subconjunto. Por ejemplo, imaginemos que una columna `fullName` se rellena hasta un total de

40 bytes en un archivo y se rellena hasta 800 bytes en otro archivo. Un observador podría deducir que un conjunto de datos contiene el nombre más largo del mundo (747 bytes).

¿Debo proporcionar un relleno adicional al usar el tipo de relleno **max**?

No. Al utilizar el relleno **max**, se recomienda establecer en 0 la `pad_length` (también conocida como relleno adicional complementario al mayor valor de la columna).

¿Puedo elegir una **pad_length** grande al usar el relleno **fixed** para evitar tener que preocuparme de que quepa el valor más grande?

Sí, pero la gran longitud de relleno es ineficiente y utiliza más almacenamiento del necesario. Le recomendamos que compruebe el tamaño del mayor valor y que defina la `pad_length` con ese valor.

¿Cómo sé si necesito las garantías criptográficas que ofrece **preserveNulls**?

Lamentablemente, la respuesta es que depende. Como mínimo, se deben revisar los [Computación criptográfica para Clean Rooms](#) en cuanto a la forma en que el ajuste `preserveNulls` protege sus datos. No obstante, le recomendamos que consulte los requisitos de manejo de datos de su organización y cualquier contrato aplicable a la colaboración correspondiente.

¿Por qué tengo que incurrir en la sobrecarga de base64?

Para permitir la compatibilidad con los formatos de archivo tabulares, como CSV, es necesaria la codificación base64. Si bien algunos formatos de archivo como Parquet podrían admitir representaciones binarias de datos, es importante que todos los participantes en una colaboración representen los datos de la misma manera para garantizar que los resultados de las consultas sean correctos.

Solución de problemas relacionados con el aumento imprevisto de tamaño del texto cifrado

Supongamos que ha cifrado los datos y que el tamaño de los datos resultantes es sorprendentemente grande. Los siguientes pasos pueden ayudarle a identificar dónde se produjo el aumento de tamaño y qué medidas puede adoptar, si las hubiera.

Identificar dónde se produjo el aumento de tamaño

Antes de poder averiguar por qué los datos cifrados son significativamente más grandes que los datos de cleartext, primero debe identificar dónde reside el aumento de tamaño. Las columnas

de Cleartext se pueden ignorar con total seguridad porque permanecen inalteradas. Observe las columnas fingerprint y sealed restantes, y elija una que parezca significativa.

Identificar el motivo por el que se produjo el aumento de tamaño

Una columna de fingerprint o una columna sealed pueden contribuir al aumento de tamaño.

Temas

- [¿Procede el aumento de tamaño de una columna fingerprint?](#)
- [¿Procede el aumento de tamaño de una columna sealed?](#)

¿Procede el aumento de tamaño de una columna fingerprint?

Si la columna que más contribuye al aumento del almacenamiento es una columna fingerprint, es probable que se deba a que los datos de cleartext son pequeños (por ejemplo, la edad de los clientes). Cada texto cifrado de fingerprint resultante tiene una longitud de 52 bytes. Lamentablemente, no se puede hacer nada al respecto sobre una column-by-column base sólida. Para obtener más información, consulte [Sobrecarga de base para las columnas fingerprint](#) para conocer los detalles de esta columna, incluida su incidencia en los requisitos de almacenamiento.

La otra causa posible del aumento de tamaño de una columna fingerprint es el ajuste de la colaboración preserveNulls. Si el ajuste de la colaboración para preserveNulls está deshabilitado (ajuste predeterminado), todos los valores null de las columnas fingerprint se convertirán en 52 bytes de texto cifrado. No hay nada que se pueda hacer al respecto en la colaboración actual. El ajuste preserveNulls se define en el momento en que se crea la colaboración, y todos los colaboradores deben usar el mismo ajuste para garantizar que los resultados de la consulta sean correctos. Para obtener más información sobre el ajuste preserveNulls y sobre cómo su habilitación afecta a las garantías de privacidad de los datos, consulte [the section called “Computación criptográfica”](#).

¿Procede el aumento de tamaño de una columna sealed?

Si la columna que más contribuye al aumento del almacenamiento es una columna sealed, hay algunos detalles que podrían contribuir al aumento de tamaño.

Si los datos de cleartext son pequeños (por ejemplo, las edades de los clientes), cada texto cifrado sealed resultante tiene una longitud mínima de 91 bytes. Lamentablemente, no se puede hacer nada al respecto. Para obtener más información, consulte [Sobrecarga de base para las columnas](#)

[sealed](#) para conocer los detalles de esta columna, incluida su incidencia en los requisitos de almacenamiento.

La segunda causa principal del aumento del almacenamiento en columnas sealed es el relleno. El relleno añade bytes adicionales al cleartext antes del cifrado para ocultar el tamaño de los valores individuales de un conjunto de datos. Se recomienda establecer el relleno en el valor mínimo posible para el conjunto de datos. Como mínimo, se debe definir `pad_length` para el relleno `fixed` de manera que abarque el mayor valor posible de la columna. Todo ajuste superior a ese no aportará garantías de privacidad adicionales. Por ejemplo, si sabe que el mayor valor posible de una columna puede ser de 50 bytes, le recomendamos que defina la `pad_length` en 50 bytes. Sin embargo, si la columna sealed utiliza el relleno `max`, le recomendamos que lo defina la `pad_length` en 0 bytes. Esto se debe a que el relleno `max` se refiere al relleno adicional que se suma al mayor valor de la columna.

La última causa posible del aumento de tamaño de una columna sealed es el ajuste de la colaboración `preserveNulls`. Si el ajuste de la colaboración para `preserveNulls` está deshabilitado (ajuste predeterminado), todos los valores `null` de las columnas sealed se convertirán en 91 bytes de texto cifrado. No hay nada que se pueda hacer al respecto en la colaboración actual. El ajuste `preserveNulls` se define en el momento en que se crea la colaboración, y todos los colaboradores deben usar el mismo ajuste para garantizar que los resultados de la consulta sean correctos. Para obtener más información sobre los efectos de esta configuración y sobre cómo su activación afecta a las garantías de privacidad de sus datos, consulte [the section called “Computación criptográfica”](#).

Con AWS Clean Rooms figuración

En los temas siguientes se explica cómo configurar los AWS Clean Rooms.

Temas

- [Inscríbese en AWS](#)
- [Configure las funciones de servicio para AWS Clean Rooms](#)
- [Configurar funciones de servicio para AWS Clean Rooms ML](#)

Inscríbese en AWS

Antes de poder usar AWS Clean Rooms, o cualquier otro Servicio de AWS, debes registrarte AWS con un Cuenta de AWS.

Si no tienes una Cuenta de AWS, sigue los siguientes pasos para crearlo.

Para suscribirte a una Cuenta de AWS

1. Abrir <https://portal.aws.amazon.com/billing/registro>.
2. Siga las instrucciones que se le indiquen.

Durante el proceso de registro, recibirás una llamada telefónica con un código de verificación que deberás introducir en el teclado del teléfono.

3. Cuando te registras en una Cuenta de AWS, se crea un usuario Cuenta de AWS root. El usuario raíz tendrá acceso a todos los Servicios de AWS y recursos de esa cuenta. Como práctica recomendada de seguridad, [asigne acceso administrativo a un usuario administrativo](#) y utilice únicamente el usuario raíz para realizar [tareas que requieran acceso de usuario raíz](#).

Configure las funciones de servicio para AWS Clean Rooms

En las siguientes secciones se describen las funciones necesarias para realizar cada tarea.

Temas

- [Creación de un usuario administrador](#)
- [Creación de un rol de IAM para un miembro de la colaboración](#)
- [Cree un rol de servicio para leer datos de Amazon S3](#)

- [Cree un rol de servicio para leer los datos de Amazon Athena](#)
- [Cree un rol de servicio para leer los datos de Snowflake](#)
- [Cree un rol de servicio para leer el código de un bucket de S3 \(rol de plantilla de PySpark análisis\)](#)
- [Cree un rol de servicio para escribir los resultados de un PySpark trabajo](#)
- [Creación de un rol de servicio para recibir resultados](#)

Creación de un usuario administrador

Para AWS Clean Rooms utilizarlos, debe crear un usuario administrador para usted y añadir el usuario administrador a un grupo de administradores.

Para crear un usuario administrador, elija una de las siguientes opciones.

Elegir una forma de administrar el administrador	Para	Haga esto	También puede
En IAM Identity Center (recomendado)	Usar credenciales a corto plazo para acceder a AWS. Esto se ajusta a las prácticas recomendadas de seguridad. Para obtener información sobre las prácticas recomendadas, consulta Prácticas recomendadas de seguridad en IAM en	Siga las instrucciones en Introducción en la Guía del usuario de AWS IAM Identity Center .	Configure el acceso mediante programación configurando el AWS CLI que se utilizará AWS IAM Identity Center en la Guía del AWS Command Line Interface usuario.

Elegir una forma de administrar el administrador	Para	Haga esto	También puede
	la Guía del usuario de IAM.		
En IAM (no recomendado)	Usar credenciales a largo plazo para acceder a AWS.	Siguiendo las instrucciones de Crear un usuario de IAM para acceso de emergencia de la Guía del usuario de IAM.	Configure el acceso programático mediante Administrar las claves de acceso de los usuarios de IAM en la Guía del usuario de IAM.

Creación de un rol de IAM para un miembro de la colaboración

Un miembro es un AWS cliente que participa en una colaboración.

Para crear un rol de IAM para un miembro de la colaboración

1. Siga el procedimiento [Creación de un rol para delegar permisos a un usuario de IAM](#) de la Guía del usuario de AWS Identity and Access Management .
2. Para el paso Creación de política, seleccione la pestaña JSON en el Editor de políticas y, a continuación, agregue políticas en función de las capacidades otorgadas al miembro de la colaboración.

AWS Clean Rooms ofrece las siguientes políticas administradas basadas en casos de uso comunes.

Si desea...	Entonces use...
Ver los recursos y metadatos	AWS política gestionada: AWSCleanRoomsReadOnlyAccess

Si desea...	Entonces use...
Consultar	AWS política gestionada: AWSCleanRoomsFullAccess
Consulta y ejecuta trabajos	AWS política gestionada: AWSCleanRoomsFullAccess
Consultar y recibir resultados	AWS política gestionada: AWSCleanRoomsFullAccess
Administración de los recursos de colaboración, pero no consultas	AWS política gestionada: AWSCleanRoomsFullAccessNoQuerying

Para obtener información sobre las diferentes políticas gestionadas que ofrecen AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#),

Cree un rol de servicio para leer datos de Amazon S3

AWS Clean Rooms utiliza un rol de servicio para leer los datos de Amazon S3.

Hay dos formas de crear este rol de servicio.

- Si tiene los permisos de IAM necesarios para crear un rol de servicio, utilice la AWS Clean Rooms consola para crear un rol de servicio.
- Si no tiene `iam:CreateRole` `iam:AttachRolePolicy` permisos o desea crear `iam:CreatePolicy` los roles de IAM manualmente, realice una de las siguientes acciones:
 - Utilice el siguiente procedimiento para crear un rol de servicio mediante políticas de confianza personalizadas.
 - Pídale al administrador que cree el rol de servicio mediante el siguiente procedimiento.

Note

Usted o su administrador de IAM deben seguir este procedimiento únicamente si no tienen los permisos necesarios para crear un rol de servicio mediante la AWS Clean Rooms consola.

Para crear un rol de servicio para leer datos de Amazon S3 mediante políticas de confianza personalizadas

1. Cree un rol mediante políticas de confianza personalizadas. Para obtener más información, consulte el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) en la Guía del AWS Identity and Access Management usuario.
2. Utilice la siguiente política de confianza personalizada de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

 Note

Si quiere asegurarse de que el rol se utilice únicamente en el contexto de una determinada membresía de colaboración, puede ampliar la política de confianza. Para obtener más información, consulte [Prevención de la sustitución confusa entre servicios](#).

JSON

```
{  
  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Sid": "RoleTrustPolicyForCleanRoomsService",  
      "Effect": "Allow",  
      "Principal": {  
        "Service": "cleanrooms.amazonaws.com"  
      },  
      "Action": "sts:AssumeRole"  
    }  
  ]  
}
```

3. Utilice la siguiente política de permisos de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

Note

La siguiente política de ejemplo admite los permisos necesarios para leer los metadatos de AWS Glue y los datos de Amazon S3 correspondientes. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Por ejemplo, si ha configurado una clave de KMS personalizada para sus datos de Amazon S3, es posible que deba modificar esta política con permisos adicionales AWS Key Management Service (AWS KMS).

Sus AWS Glue recursos y los recursos subyacentes de Amazon S3 deben estar en la Región de AWS misma posición que los de la AWS Clean Rooms colaboración.

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

Esta política hace referencia a dos tipos diferentes Cuenta de AWS IDs para respaldar una AWS Clean Rooms colaboración en la que los metadatos del catálogo de datos y el almacenamiento real de los datos estén gestionados por distintas partes:

- 111122223333: es la cuenta propietaria de los recursos del catálogo de AWS Glue datos (bases de datos, tablas y catálogos). La primera instrucción otorga permisos para acceder a los esquemas de tablas, la información de particiones y los metadatos del catálogo de esta cuenta. AWS Glue
- 444455556666: es la cuenta propietaria del bucket de Amazon S3 que contiene los archivos de datos reales. Los permisos de Amazon S3 (estados de cuenta 3 y 4) están restringidos a los buckets propiedad de esta cuenta a través de esta `s3:ResourceAccount` condición.

Esta configuración es compatible con las arquitecturas de datos empresariales habituales, en las que un equipo administra el catálogo de datos y las definiciones del esquema, mientras que otro equipo es propietario de la infraestructura de almacenamiento de datos subyacente. La `s3:ResourceAccount` condición proporciona una capa de seguridad adicional al garantizar que las operaciones de Amazon S3 solo funcionen en los buckets propiedad de la cuenta designada.

4. Reemplace cada *placeholder* por su propia información.
5. Siga el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) para crear el rol.

Cree un rol de servicio para leer los datos de Amazon Athena

AWS Clean Rooms utiliza un rol de servicio para leer los datos de Amazon Athena.

Para crear un rol de servicio para leer datos de Athena mediante políticas de confianza personalizadas

1. Cree un rol mediante políticas de confianza personalizadas. Para obtener más información, consulte el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) en la Guía del AWS Identity and Access Management usuario.
2. Utilice la siguiente política de confianza personalizada de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

 Note

Si quiere asegurarse de que el rol se utilice únicamente en el contexto de una determinada membresía de colaboración, puede ampliar la política de confianza. Para obtener más información, consulte [Prevención de la sustitución confusa entre servicios](#).

JSON

```
{  
  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Sid": "RoleTrustPolicyForCleanRoomsService",  
      "Effect": "Allow",  
      "Principal": {  
        "Service": "cleanrooms.amazonaws.com"  
      },  
      "Action": "sts:AssumeRole"  
    }  
  ]  
}
```

3. Utilice la siguiente política de permisos de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

Note

El siguiente ejemplo de política admite los permisos necesarios para leer AWS Glue los metadatos y sus datos de Athena correspondientes. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Por ejemplo, si ya configuró una clave de KMS personalizada para sus datos de Amazon S3, es posible que deba modificar esta política con AWS KMS permisos adicionales.

```
{
  "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. Reemplace cada *placeholder* por su propia información.
5. Siga el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) para crear el rol.

Configurar permisos de Lake Formation

Si consulta recursos protegidos con permisos de Lake Formation, la función de servicio debe tener permisos de acceso de selección table/view/catalog y descripción en la AWS Glue base de datos y de descripción.

Para obtener más información, consulte lo siguiente:

- [Utilice Athena para consultar los datos registrados AWS Lake Formation](#) en la Guía del usuario de Amazon Athena
- [Cómo incorporar los permisos de Lake Formation](#) en la Guía AWS Lake Formation para desarrolladores

Cree un rol de servicio para leer los datos de Snowflake

AWS Clean Rooms utiliza un rol de servicio para recuperar las credenciales para que Snowflake lea los datos de esta fuente.

Hay dos formas de crear este rol de servicio:

- Si tiene los permisos de IAM necesarios para crear un rol de servicio, utilice la AWS Clean Rooms consola para crear un rol de servicio.
- Si no tiene `iam:CreateRole` `iam:AttachRolePolicy` permisos o desea crear `iam:CreatePolicy` los roles de IAM manualmente, realice una de las siguientes acciones:
 - Utilice el siguiente procedimiento para crear un rol de servicio mediante políticas de confianza personalizadas.
 - Pídale al administrador que cree el rol de servicio mediante el siguiente procedimiento.

Note

Usted o su administrador de IAM deben seguir este procedimiento únicamente si no tienen los permisos necesarios para crear un rol de servicio mediante la AWS Clean Rooms consola.

Para crear un rol de servicio para leer datos de Snowflake mediante políticas de confianza personalizadas

1. Cree un rol mediante políticas de confianza personalizadas. Para obtener más información, consulte el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) en la Guía del AWS Identity and Access Management usuario.

2. Utilice la siguiente política de confianza personalizada de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

 Note

Si quiere asegurarse de que el rol se utilice únicamente en el contexto de una determinada membresía de colaboración, puede ampliar la política de confianza. Para obtener más información, consulte [Prevención de la sustitución confusa entre servicios](#).

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

Esta política de confianza hace referencia a dos tipos diferentes Cuenta de AWS IDs para respaldar una AWS Clean Rooms colaboración en la que las responsabilidades de ejecución de las consultas se distribuyen entre varias partes:

- 111122223333: esta es la cuenta que contiene una membresía que participa en la colaboración. Esta membresía puede ser propietaria de tablas de datos, reglas de análisis u otros recursos de colaboración que requieran acceso desde un rol.
- 444455556666: esta es la cuenta que contiene la membresía responsable de ejecutar las consultas (el «ejecutor de consultas»). Esta membresía ejecuta consultas protegidas y debe asumir esta función para acceder a los recursos informáticos y de datos necesarios.

Esta configuración permite escenarios en los que una parte proporciona plantillas de datos o análisis mientras que otra parte ejecuta las consultas reales. Ambas funciones requieren permisos diferentes pero complementarios a través de la misma función de ejecución. La `aws:SourceArn` condición garantiza que solo AWS Clean Rooms las operaciones que se originen a partir de estas dos membresías específicas puedan asumir esa función, manteniendo la seguridad y, al mismo tiempo, apoyando el flujo de trabajo distribuido de ejecución de tareas y gestión de resultados.

3. Utilice una de las siguientes políticas de permisos según el procedimiento de [creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

Política de permisos para los secretos cifrados con una clave KMS propiedad del cliente

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

Esta política hace referencia a dos opciones diferentes Cuenta de AWS IDs para respaldar un escenario de gestión de secretos entre cuentas:

- 111122223333: esta es la cuenta que posee y almacena el secreto. La primera declaración otorga permiso para recuperar el valor secreto de esta cuenta.
- 444455556666: es la cuenta propietaria de la AWS KMS clave utilizada para cifrar el secreto. La segunda declaración otorga permiso para descifrar el secreto con la AWS KMS clave de esta cuenta.

Esta configuración es común en entornos empresariales en los que:

- Los secretos se administran de forma centralizada en una cuenta (cuenta 1)
- Las claves de cifrado se administran mediante una cuenta de seguridad o de servicios compartidos independiente (cuenta 2)
- La política de AWS KMS claves de la cuenta 2 también debe permitir que el servicio de la cuenta 1 utilice la clave para encryption/decryption las operaciones

Esta `kms:EncryptionContext:SecretARN` condición garantiza que la AWS KMS clave solo se pueda usar para descifrar este secreto específico, lo que proporciona un nivel de seguridad adicional para el acceso entre cuentas.

Política de permisos para los secretos cifrados con un Clave administrada de AWS

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "arn:aws:secretsmanager:us-  
east-1:111122223333:secret:secretIdentifier",
      "Effect": "Allow"
    }
  ]
}
```

4. Reemplace cada *placeholder* por su propia información.
5. Siga el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) para crear el rol.

Cree un rol de servicio para leer el código de un bucket de S3 (rol de plantilla de PySpark análisis)

AWS Clean Rooms usa un rol de servicio para leer el código del bucket de S3 especificado por un miembro de la colaboración cuando usa una plantilla de PySpark análisis.

Para crear un rol de servicio para leer el código de un bucket de S3

1. Cree un rol mediante políticas de confianza personalizadas. Para obtener más información, consulte el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) en la Guía del AWS Identity and Access Management usuario.

2. Utilice la siguiente política de confianza personalizada de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

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

Esta política de confianza hace referencia a dos tipos diferentes Cuenta de AWS IDs para respaldar un escenario de AWS Clean Rooms colaboración entre múltiples partes:

- 111122223333: esta es la cuenta que contiene a los miembros responsables de ejecutar las consultas (los «encargados de ejecutar tareas»). Esta membresía ejecuta los trabajos de análisis y debe asumir esta función para acceder a los recursos necesarios.
- 444455556666: es la cuenta propietaria de la plantilla de análisis y sus miembros asociados (el «propietario de la plantilla de análisis»). Esta membresía define qué

consultas se pueden ejecutar y también debe asumir esta función para administrar y ejecutar el análisis.

Esta configuración es típica de AWS Clean Rooms las colaboraciones en las que varias partes participan en la misma colaboración, cada una con su propia Cuenta de AWS membresía. Tanto el ejecutor de la consulta como el propietario de la plantilla de análisis necesitan acceder a los recursos compartidos. Esta `aws:SourceArn` condición garantiza que solo AWS Clean Rooms las operaciones que se originen a partir de estas dos membresías específicas puedan asumir esa función, lo que proporciona un control de acceso preciso para la colaboración entre múltiples partes.

3. Utilice la siguiente política de permisos de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

 Note

El siguiente ejemplo de política admite los permisos necesarios para leer el código de Amazon S3. No obstante, quizás tenga que modificar esta política en función de cómo haya configurado los datos de S3.

Sus recursos de Amazon S3 deben estar en los Región de AWS mismos que los de la AWS Clean Rooms colaboración.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:GetObjectVersion"
      ],
      "Resource": ["arn:aws:s3:::s3Path"],
      "Condition": {
        "StringEquals": {
          "s3:ResourceAccount": [
```

```
        "s3BucketOwnerAccountId"  
      ]  
    }  
  }  
}
```

4. Sustituya cada uno *placeholder* por su propia información:
 - *s3Path*— La ubicación del código en el depósito S3.
 - *s3BucketOwnerAccountId*— El Cuenta de AWS ID del propietario del bucket de S3.
 - *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
 - *jobRunnerAccountId*— El Cuenta de AWS ID del miembro que puede ejecutar consultas y trabajos.
 - *jobRunnerMembershipId*— El ID de membresía del miembro que puede consultar y ejecutar trabajos. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
 - *analysisTemplateAccountId*— El Cuenta de AWS ID de la plantilla de análisis.
 - *analysisTemplateOwnerMembershipId*— El ID de membresía del miembro propietario de la plantilla de análisis. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración.
5. Siga el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) para crear el rol.

Cree un rol de servicio para escribir los resultados de un PySpark trabajo

AWS Clean Rooms usa un rol de servicio para escribir los resultados de un PySpark trabajo en un bucket de S3 específico.

Para crear un rol de servicio para escribir los resultados de un PySpark trabajo

1. Cree un rol mediante políticas de confianza personalizadas. Para obtener más información, consulte el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) en la Guía del AWS Identity and Access Management usuario.

2. Utilice la siguiente política de confianza personalizada de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

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

Esta política de confianza hace referencia Cuenta de AWS IDs a dos funciones operativas diferentes para respaldar una AWS Clean Rooms colaboración con funciones operativas distintas:

- 111122223333: esta es la cuenta que contiene a los miembros responsables de ejecutar los trabajos de análisis (los «encargados de ejecutar los trabajos»). Esta membresía ejecuta las cargas de trabajo computacionales y debe asumir esta función para acceder a los recursos de procesamiento.

- 444455556666: esta es la cuenta que contiene la membresía con las responsabilidades de receptor de resultados (RR). Esta membresía está autorizada para recibir y acceder a los resultados de los trabajos de análisis, y necesita tener acceso a un rol para escribir los resultados en las ubicaciones designadas.

Esta configuración permite AWS Clean Rooms escenarios en los que una parte ejecuta el análisis computacional mientras que otra parte recibe y administra los resultados. Ambas funciones requieren permisos diferentes pero complementarios a través de la misma función de ejecución. La `aws:SourceArn` condición garantiza que solo AWS Clean Rooms las operaciones que se originen a partir de estas dos membresías específicas puedan asumir esa función, manteniendo la seguridad y, al mismo tiempo, apoyando el flujo de trabajo distribuido de ejecución de tareas y gestión de resultados.

3. Utilice la siguiente política de permisos de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

Note

El siguiente ejemplo de política admite los permisos necesarios para escribir en Amazon S3. Sin embargo, es posible que deba modificar esta política en función de cómo haya configurado S3.

Sus recursos de Amazon S3 deben estar en los Región de AWS mismos que los de la AWS Clean Rooms colaboración.

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. Sustituya cada uno *placeholder* por su propia información:

- *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- *jobRunnerAccountId*— El Cuenta de AWS ID en el que se encuentra el depósito S3.
- *jobRunnerMembershipId*— El ID de membresía del miembro que puede consultar y ejecutar trabajos. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
- *rrAccountId*— El Cuenta de AWS ID en el que se encuentra el depósito de S3.
- *rrMembershipId*— El ID de membresía del miembro que puede recibir los resultados. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
- *bucket*— El nombre y la ubicación del depósito de S3.

- *optionalPrefix*— Un prefijo opcional si desea guardar los resultados con un prefijo S3 específico.
 - *s3BucketOwnerAccountId*— El Cuenta de AWS ID del propietario del bucket de S3.
5. Siga el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) para crear el rol.

Creación de un rol de servicio para recibir resultados

Note

Si es el único miembro que solo puede recibir resultados (en la consola, sus destrezas como miembro solo son Recibir resultados), siga este procedimiento.

Si es un miembro que puede realizar consultas y recibir resultados (en la consola, Sus destrezas como miembro son Consultar y Recibir resultados), puede omitir este procedimiento.

Para los miembros de la colaboración que solo pueden recibir resultados, AWS Clean Rooms utiliza un rol de servicio para escribir los resultados de los datos consultados en la colaboración en el bucket de S3 especificado.

Hay dos formas de crear este rol de servicio:

- Si tiene los permisos de IAM necesarios para crear un rol de servicio, utilice la AWS Clean Rooms consola para crear un rol de servicio.
- Si no tiene `iam:CreateRole` `iam:AttachRolePolicy` permisos o desea crear `iam:CreatePolicy` los roles de IAM manualmente, realice una de las siguientes acciones:
 - Utilice el siguiente procedimiento para crear un rol de servicio mediante políticas de confianza personalizadas.
 - Pídale al administrador que cree el rol de servicio mediante el siguiente procedimiento.

Note

Usted o su administrador de IAM deben seguir este procedimiento únicamente si no tienen los permisos necesarios para crear un rol de servicio mediante la AWS Clean Rooms consola.

Para crear un rol de servicio para recibir resultados mediante políticas de confianza personalizadas

1. Cree un rol mediante políticas de confianza personalizadas. Para obtener más información, consulte el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) en la Guía del AWS Identity and Access Management usuario.
2. Utilice la siguiente política de confianza personalizada de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

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-EXAMPLEaaaaa"
                ]
            }
        }
    }
]
}

```

3. Utilice la siguiente política de permisos de acuerdo con el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#).

Note

La siguiente política de ejemplo admite los permisos necesarios para leer los metadatos de AWS Glue y los datos de Amazon S3 correspondientes. No obstante, quizás tenga que modificar esta política en función de cómo haya configurado los datos de S3. Sus AWS Glue recursos y los recursos subyacentes de Amazon S3 deben estar en la Región de AWS misma posición que los de la AWS Clean Rooms colaboración.

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. Sustituya cada uno *placeholder* por su propia información:

- *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- *a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaaa*— El ID de membresía del miembro que puede realizar la consulta. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
- *arn:aws:cleanrooms:us-east-1:555555555555:membership/a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaaa*— El ARN de membresía único del miembro que puede realizar la consulta. Puede encontrar el ARN de la pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
- *bucket_name*— El nombre de recurso de Amazon (ARN) del bucket de S3. Puede encontrar el Nombre de recurso de Amazon (ARN) en la pestaña Propiedades del bucket en Amazon S3.
- *accountId*— El Cuenta de AWS ID en el que se encuentra el bucket de S3.

bucket_name/optional_key_prefix— El nombre del recurso de Amazon (ARN) del destino de los resultados en Amazon S3. Puede encontrar el Nombre de recurso de Amazon (ARN) en la pestaña Propiedades del bucket en Amazon S3.

5. Siga el procedimiento [Creación de un rol mediante políticas de confianza personalizadas \(consola\)](#) para crear el rol.

Configurar funciones de servicio para AWS Clean Rooms ML

Las funciones necesarias para realizar un modelado similar difieren de las necesarias para utilizar un modelo personalizado. En las siguientes secciones se describen las funciones necesarias para realizar cada tarea.

Temas

- [Configure las funciones de servicio para un modelado similar](#)
- [Configure las funciones de servicio para el modelado personalizado](#)

Configure las funciones de servicio para un modelado similar

Temas

- [Creación de rol de servicio para leer datos de entrenamiento](#)
- [Creación de un rol de servicio para escribir un segmento similar](#)
- [Creación de rol de servicio para leer datos iniciales](#)

Creación de rol de servicio para leer datos de entrenamiento

AWS Clean Rooms usa un rol de servicio para leer los datos de entrenamiento. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Para crear un rol de servicio para entrenar un conjunto de datos

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.
2. En Administración de acceso, seleccione Políticas.

3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue la siguiente política.

 Note

El siguiente ejemplo de política admite los permisos necesarios para leer AWS Glue los metadatos y sus datos correspondientes de Amazon S3. No obstante, quizás tenga que modificar esta política en función de cómo haya configurado los datos de S3. Esta política no incluye una clave de KMS para descifrar los datos.

Sus AWS Glue recursos y los recursos subyacentes de Amazon S3 deben estar en la Región de AWS misma posición que los de la AWS Clean Rooms colaboración.

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 es necesario usar una clave KMS para descifrar los datos, añada esta instrucción de AWS KMS a la plantilla anterior:

```
{
    "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. Sustituya cada uno *placeholder* por su propia información:

- *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- *accountId*— El Cuenta de AWS ID en el que se encuentra el depósito S3.
- *database/databases*, *table/databases/tablescatalog*, y *database/default*: la ubicación de los datos de entrenamiento a los que AWS Clean Rooms hay que acceder.
- *bucket*— El nombre de recurso de Amazon (ARN) del bucket de S3. Puede encontrar el Nombre de recurso de Amazon (ARN) en la pestaña Propiedades del bucket en Amazon S3.
- *bucketFolders*— El nombre de las carpetas específicas del bucket de S3 a AWS Clean Rooms las que se debe acceder.

6. Elija Siguiente.

7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.

8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.

11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.

12. Copie y pegue la siguiente política de confianza personalizada en el editor 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/*"
        }
      }
    }
  ]
}
```

El SourceAccount es siempre tuyo Cuenta de AWS. El SourceArn puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de crear ese conjunto de datos. Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

accountId es el ID Cuenta de AWS que contiene los datos de entrenamiento.

13. Elija Siguiente y, en Agregar permisos, introduzca el nombre de la política que acaba de crear. (es posible que tenga que volver a cargar la página).
14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de `passRole` concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.
- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para AWS Clean Rooms.

Creación de un rol de servicio para escribir un segmento similar

AWS Clean Rooms usa un rol de servicio para escribir segmentos similares en un segmento. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Creación de un rol de servicio para escribir un segmento similar

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.
2. En Administración de acceso, seleccione Políticas.
3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue la siguiente política.

Note

El siguiente ejemplo de política admite los permisos necesarios para leer AWS Glue los metadatos y sus datos correspondientes de Amazon S3. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Esta política no incluye una clave de KMS para descifrar los datos. Sus AWS Glue recursos y los recursos subyacentes de Amazon S3 deben estar en la Región de AWS misma posición que los de la AWS Clean Rooms colaboración.

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 es necesario usar una clave de KMS para cifrar los datos, agregue esta instrucción de AWS KMS a la plantilla:

```

{
    "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. Sustituya cada uno *placeholder* por su propia información:

- *buckets*— El nombre de recurso de Amazon (ARN) del bucket de S3. Puede encontrar el Nombre de recurso de Amazon (ARN) en la pestaña Propiedades del bucket en Amazon S3.
- *accountId*— El Cuenta de AWS ID en el que se encuentra el bucket de S3.
- *bucketFolders*— El nombre de las carpetas específicas del depósito de S3 a AWS Clean Rooms las que hay que acceder.

- **region**: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- **keyId**— La clave KMS necesaria para cifrar los datos.

6. Elija Siguiente.

7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.

8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.

11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.

12. Copie y pegue la siguiente política de confianza personalizada en el editor 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/*"
        }
      }
    }
  ]
}
```

```
}  
]  
}
```

El `SourceAccount` es siempre tuyo Cuenta de AWS. El `SourceArn` puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de crear ese conjunto de datos. Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

13. Elija Siguiente.
14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de `passRole` concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.
- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para. AWS Clean Rooms

Creación de rol de servicio para leer datos iniciales

AWS Clean Rooms utiliza un rol de servicio para leer los datos iniciales. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Para crear un rol de servicio para leer los datos iniciales almacenados en un bucket de S3.

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.

2. En Administración de acceso, seleccione Políticas.
3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue una de las siguientes políticas.

 Note

El siguiente ejemplo de política admite los permisos necesarios para leer AWS Glue los metadatos y sus datos correspondientes de Amazon S3. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Esta política no incluye una clave de KMS para descifrar los datos. Sus AWS Glue recursos y los recursos subyacentes de Amazon S3 deben estar en la Región de AWS misma posición que los de la AWS Clean Rooms colaboración.

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

La siguiente política de ejemplo admite los permisos necesarios para leer los resultados de una consulta SQL y utilizarlos como datos de entrada. No obstante, es posible que tenga que modificar esta política en función de cómo esté estructurada la consulta. Esta política no incluye una clave de KMS para descifrar los datos.

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 es necesario usar una clave KMS para descifrar los datos, añade esta instrucción de AWS KMS a la plantilla:

```

{
    "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. Sustituya cada uno *placeholder* por su propia información:

- *buckets*— El nombre de recurso de Amazon (ARN) del bucket de S3. Puede encontrar el Nombre de recurso de Amazon (ARN) en la pestaña Propiedades del bucket en Amazon S3.
- *accountId*— El Cuenta de AWS ID en el que se encuentra el bucket de S3.

- *bucketFolders*— El nombre de las carpetas específicas del depósito de S3 a AWS Clean Rooms las que hay que acceder.
- *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- *queryRunnerAccountId*— El Cuenta de AWS ID de la cuenta que ejecutará las consultas.
- *queryRunnerMembershipId*— El ID de membresía del miembro que puede realizar la consulta. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
- *keyId*— La clave KMS necesaria para cifrar los datos.

6. Elija Siguiente.

7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.

8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.

11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.

12. Copie y pegue la siguiente política de confianza personalizada en el editor JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAssumeRole",
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-m1.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
```

```

    "StringEqualsIfExists": {
        "aws:SourceAccount": ["111122223333"]
    },
    "ArnLikeIfExists": {
        "aws:SourceArn": "arn:aws:cleanrooms-ml:us-
east-1:111122223333:audience-generation-job/*"
    }
}
]
}

```

El SourceAccount es siempre tuyo Cuenta de AWS. El SourceArn puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de crear ese conjunto de datos. Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

13. Elija Siguiente.
14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de passRole concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.
- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para. AWS Clean Rooms

Configure las funciones de servicio para el modelado personalizado

Temas

- [Cree un rol de servicio para el modelado de ML personalizado: Configuración de ML](#)
- [Cree un rol de servicio para proporcionar un modelo de aprendizaje automático personalizado](#)
- [Cree un rol de servicio para consultar un conjunto de datos](#)
- [Cree un rol de servicio para crear una asociación de tablas configurada](#)

Cree un rol de servicio para el modelado de ML personalizado: Configuración de ML

AWS Clean Rooms usa un rol de servicio para controlar quién puede crear una configuración de ML personalizada. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Esta función le permite utilizar la `MLConfiguration` acción [Put](#).

Para crear un rol de servicio que permita la creación de una configuración de aprendizaje automático personalizada

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.
2. En Administración de acceso, seleccione Políticas.
3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue la siguiente política.

Note

El siguiente ejemplo de política admite los permisos necesarios para acceder a un bucket de S3 y escribirlos y para publicar CloudWatch métricas. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Esta política no incluye una clave de KMS para descifrar los datos. Sus recursos de Amazon S3 deben estar en la Región de AWS mismos que los de la AWS Clean Rooms colaboración.

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. Sustituya cada uno *placeholder* por su propia información:

- *bucket*— El nombre de recurso de Amazon (ARN) del bucket de S3. Puede encontrar el Nombre de recurso de Amazon (ARN) en la pestaña Propiedades del bucket en Amazon S3.
- *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- *accountId*— El Cuenta de AWS ID en el que se encuentra el bucket de S3.
- *keyId*— La clave KMS necesaria para cifrar los datos.

6. Elija Siguiente.

7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.

8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.
11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.
12. Copie y pegue la siguiente política de confianza personalizada en el editor 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"
        }
      }
    }
  ]
}
```

El SourceAccount es siempre tuyo Cuenta de AWS. El SourceArn puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de crear ese conjunto de datos. Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

13. Elija Siguiente.

14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de `passRole` concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.
- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para. AWS Clean Rooms

Cree un rol de servicio para proporcionar un modelo de aprendizaje automático personalizado

AWS Clean Rooms utiliza un rol de servicio para controlar quién puede crear un algoritmo de modelo de aprendizaje automático personalizado. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Este rol le permite usar la [CreateConfiguredModelAlgorithm](#) acción.

Crear un rol de servicio que permita a un miembro proporcionar un modelo de aprendizaje automático personalizado

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.
2. En Administración de acceso, seleccione Políticas.
3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue la siguiente política.

 Note

El siguiente ejemplo de política admite los permisos necesarios para recuperar la imagen del docker que contiene el algoritmo del modelo. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Esta política no incluye una clave de KMS para descifrar los datos.

Sus recursos de Amazon S3 deben estar en los Región de AWS mismos que los de la AWS Clean Rooms colaboración.

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. Sustituya cada uno *placeholder* por su propia información:
 - *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
 - *accountId*— El Cuenta de AWS ID en el que se encuentra el depósito S3.
 - *repoName*— El nombre del repositorio que contiene los datos.
6. Elija Siguiente.
7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.
8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.
11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.
12. Copie y pegue la siguiente política de confianza personalizada en el editor JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Siempre SourceAccount es tuyo. SourceArn Puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de que se haya creado ese conjunto de datos. Cuenta de AWS Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

13. Elija Siguiente.
14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de `passRole` concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.
- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para. AWS Clean Rooms

Cree un rol de servicio para consultar un conjunto de datos

AWS Clean Rooms usa un rol de servicio para controlar quién puede consultar un conjunto de datos que se utilizará para el modelado de aprendizaje automático personalizado. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Este rol te permite usar la acción [Crear MLInput canal](#).

Para crear un rol de servicio que permita a un miembro consultar un conjunto de datos

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.
2. En Administración de acceso, seleccione Políticas.
3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue la siguiente política.

 Note

El siguiente ejemplo de política admite los permisos necesarios para consultar un conjunto de datos que se utilizará para el modelado de aprendizaje automático personalizado. Sin embargo, es posible que tengas que modificar esta política en función

de cómo hayas configurado los datos de Amazon S3. Esta política no incluye una clave de KMS para descifrar los datos.

Sus recursos de Amazon S3 deben estar en los Región de AWS mismos que los de la AWS Clean Rooms colaboración.

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. Sustituya cada uno *placeholder* por su propia información:
 - *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
 - *queryRunnerAccountId*— El Cuenta de AWS ID de la cuenta que ejecutará las consultas.
 - *queryRunnerMembershipId*— El ID de membresía del miembro que puede realizar la consulta. Puede encontrar el ID de pertenencia en la pestaña Detalles de la colaboración. Esto garantiza AWS Clean Rooms que solo asuma el rol cuando este miembro ejecute el análisis en esta colaboración.
6. Elija Siguiente.
7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.
8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.
11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.
12. Copie y pegue la siguiente política de confianza personalizada en el editor JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-m1.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Siempre `SourceAccount` es tuyo. `SourceArn` Puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de que se haya creado ese conjunto de datos. Cuenta de AWS Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

13. Elija Siguiente.
14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de `passRole` concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.
- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para. AWS Clean Rooms

Cree un rol de servicio para crear una asociación de tablas configurada

AWS Clean Rooms usa un rol de servicio para controlar quién puede crear una asociación de tablas configurada. Puede crear este rol mediante la consola si dispone de los permisos de IAM necesarios. Si no tiene permisos `CreateRole`, pida al administrador que cree el rol de servicio.

Este rol le permite usar la `CreateConfiguredTableAssociation` acción.

Para crear un rol de servicio que permita la creación de una asociación de tablas configurada

1. Inicie sesión en la consola de IAM (<https://console.aws.amazon.com/iam/>) con su cuenta de administrador.
2. En Administración de acceso, seleccione Políticas.

3. Elija Create Policy.
4. En el Editor de políticas, seleccione la pestaña JSON y, a continuación, copie y pegue la siguiente política.

 Note

El siguiente ejemplo de política admite la creación de una asociación de tablas configurada. Sin embargo, es posible que tengas que modificar esta política en función de cómo hayas configurado los datos de Amazon S3. Esta política no incluye una clave de KMS para descifrar los datos.

Sus recursos de Amazon S3 deben estar en la Región de AWS mismos que los de la AWS Clean Rooms colaboración.

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"
    }
  ]
}

```

Reemplace el recurso marcador de posición ARNs

Al usar esta política, debes reemplazar los identificadores de recursos marcadores de posición por los reales ARNs de tus recursos:

- AWS KMS Recurso clave: *KMS-key-ID* Sustitúyalo por el identificador de AWS KMS clave real que cifra los datos de Amazon S3. La clave debe estar en la misma cuenta (111122223333) que es propietaria de los recursos del AWS Glue catálogo.
- Recursos del bucket de Amazon S3: *bucket-name* sustitúyalo por el nombre real del bucket de Amazon S3 que contiene los datos de AWS Glue la tabla. Tenga en cuenta

que los buckets de Amazon S3 ARNs no incluyen la cuenta, IDs ya que los nombres de los buckets son únicos a nivel mundial.

- AWS Glue Recursos: sustituya los siguientes marcadores de posición por los nombres reales de sus recursos:
 - *Glue database name*- El nombre de la base de datos AWS Glue
 - *Glue table name*- El nombre de su AWS Glue tabla

Todos los AWS Glue recursos (catálogo, base de datos y tabla) deben estar en el mismo lugar Cuenta de AWS (111122223333) para garantizar la coherencia de los permisos de acceso. Esta cuenta debe ser la misma a la que pertenece la AWS KMS clave utilizada para el cifrado de datos, lo que crea un límite de seguridad unificado para los recursos de AWS Clean Rooms datos.

5. Sustituya cada *placeholder* una por su propia información:

- *KMS key used to encrypt the Amazon S3 data*— La clave de KMS que se utilizó para cifrar los datos de Amazon S3. Para descifrar los datos, debe proporcionar la misma clave KMS que se utilizó para cifrar los datos.
- *Amazon S3 bucket of AWS Glue table*— El nombre del bucket de Amazon S3 que contiene la AWS Glue tabla que contiene sus datos.
- *region*: nombre del Región de AWS. Por ejemplo, **us-east-1**.
- *accountId*— El Cuenta de AWS ID de la cuenta propietaria de los datos.
- *AWS Glue database name*— El nombre de la AWS Glue base de datos que contiene los datos.
- *AWS Glue table name*— El nombre de la AWS Glue tabla que contiene los datos.

6. Elija Siguiente.

7. En Revisar y crear, introduzca un Nombre de política y una Descripción y revise el Resumen.

8. Elija Crear política.

Ha creado una política para AWS Clean Rooms.

9. En Administración de accesos, elija Roles.

Con Roles, puede crear credenciales a corto plazo, que son las recomendadas aumentar la seguridad. También puede elegir Usuarios para crear credenciales a largo plazo.

10. Elija Crear rol.

11. En el asistente Crear rol, en Tipo de entidad de confianza, elija Política de confianza personalizada.
12. Copie y pegue la siguiente política de confianza personalizada en el editor JSON.

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "cleanrooms-ml.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Siempre SourceAccount es tuyo. SourceArn Puede limitarse a un conjunto de datos de entrenamiento específico, pero solo después de que se haya creado ese conjunto de datos. Cuenta de AWS Como aún no conoce el ARN del conjunto de datos de entrenamiento, el comodín se especifica aquí.

13. Elija Siguiente.
14. Seleccione la casilla de verificación situada junto al nombre de la política que creó y, a continuación, elija Siguiente.
15. En Nombre, revisar y crear, introduzca el Nombre del rol y la Descripción.

 Note

El Nombre del rol debe coincidir con el patrón de los permisos de passRole concedidos al miembro que puede realizar consultas y recibir resultados y a roles de miembros.

- a. Revise la sección Seleccionar entidades de confianza y edítela si es necesario.
- b. Revise los permisos en Agregar permisos y edítelos si es necesario.

- c. Revise las Etiquetas y añada etiquetas si es necesario.
- d. Elija Crear rol.

Ha creado el rol de servicio para. AWS Clean Rooms

Colaboraciones y membresías en AWS Clean Rooms

Una colaboración es un límite lógico seguro AWS Clean Rooms en el que los miembros pueden realizar análisis en tablas configuradas.

Cualquier miembro AWS Clean Rooms puede crear una colaboración.

El creador de la colaboración puede designar a un solo miembro para analizar las tablas configuradas y recibir los resultados. Sin embargo, es posible que el creador de la colaboración desee impedir que el miembro que puede ejecutar el análisis tenga acceso a los resultados de la consulta. En ese caso, el creador de la colaboración puede designar a un [miembro para que realice la consulta](#) o a [un miembro para que ejecute las consultas y los trabajos](#) y a otro [miembro para que reciba los resultados](#).

En la mayoría de los casos, el miembro que puede realizar consultas o el miembro que puede consultar y ejecutar trabajos es también el [miembro que paga los costes de procesamiento](#). No obstante, el creador de la colaboración puede configurar a otro miembro para que se encargue de pagar los costos de computación de consultas.

Para obtener información sobre cómo crear una colaboración mediante la AWS SDKs, consulta la [referencia de la AWS Clean Rooms API](#).

Temas

- [Crear una colaboración](#)
- [Crear una pertenencia y unirse a una colaboración](#)
- [Editar colaboraciones](#)
- [Solicitudes de cambio en AWS Clean Rooms](#)
- [Eliminar colaboraciones](#)
- [Ver colaboraciones](#)
- [Invitar a los miembros a una colaboración](#)
- [Supervisión de miembros](#)
- [Añadir miembros a una colaboración](#)
- [Eliminar miembros de una colaboración](#)
- [Abandonar una colaboración](#)

Crear una colaboración

Hay tres formas de crear una colaboración en AWS Clean Rooms.

La forma más básica es la [colaboración para consultas](#). Esta colaboración se centra en el análisis de consultas SQL y mantiene una estructura simple con dos funciones principales: un miembro que puede ejecutar consultas y otro que puede recibir los resultados. Esta configuración de colaboración básica funciona bien para tareas sencillas de análisis de datos.

La segunda forma, la [colaboración para consultas y trabajos](#), amplía la funcionalidad al incorporar tanto consultas SQL como PySpark trabajos. Esta configuración de colaboración mantiene la misma estructura de funciones básica, pero amplía los permisos para incluir la ejecución de tareas. Un requisito importante es que el miembro que crea las plantillas de PySpark análisis también debe ser el que reciba los resultados, lo que garantiza una responsabilidad clara en el proceso de análisis.

La tercera forma, la [colaboración para el modelado de aprendizaje automático](#), está diseñada para los flujos de trabajo de aprendizaje automático. Esta configuración de colaboración añade dos funciones más: una para los usuarios que necesitan los resultados de modelos entrenados y otra para los usuarios que necesitan los resultados del uso de esos modelos para hacer predicciones. Esta configuración de colaboración ayuda a los miembros de la colaboración a trabajar juntos en proyectos de datos complejos y, al mismo tiempo, mantiene claras las funciones y permisos de todos.

En los siguientes temas se explica cómo crear colaboraciones para consultas, trabajos y modelos de aprendizaje automático.

Temas

- [Crear una colaboración para consultas](#)
- [Crear una colaboración para consultas y trabajos](#)
- [Crear una colaboración para el modelado de aprendizaje automático](#)

Crear una colaboración para consultas

En este procedimiento, usted, como [creador de la colaboración](#), realiza las siguientes tareas:

- [Cree una colaboración](#).
- Invita a uno o más [miembros](#) a la [colaboración](#).

- Asigne capacidades a los miembros, como el [miembro que puede realizar consultas](#) y el [miembro que puede recibir los resultados](#).

Si el creador de la colaboración es también el miembro que puede recibir los resultados, especificará el destino y el formato de los resultados. También proporcionan una función de servicio Amazon Resource Name (ARN) para escribir los resultados en el destino de los resultados.

- Configure qué [miembro es responsable de pagar los costes informáticos de la colaboración](#).

Antes de comenzar, asegúrese de que cumple los siguientes requisitos previos:

- Tiene el nombre y el Cuenta de AWS ID de cada miembro al que quiere invitar a la colaboración.
- Tiene permiso para compartir el nombre y la Cuenta de AWS identificación de cada miembro con todos los miembros de la colaboración.

 Note

No puedes añadir más miembros después de crear la colaboración.

Para obtener información sobre cómo crear una colaboración mediante la AWS SDKs, consulta la [referencia de la AWS Clean Rooms API](#).

Para crear una colaboración para consultas

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con la Cuenta de AWS que funcionará como creador de la colaboración.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la esquina superior derecha, elija Crear colaboración.
4. En Paso 1: Definir la colaboración, haga lo siguiente:
 - a. En Detalles, introduzca el Nombre y la Descripción de la colaboración.

Esta información estará visible para los miembros de la colaboración que estén invitados a participar en la colaboración. La información de Nombre y Descripción les ayudará a entender a qué se refiere la colaboración.

- b. En Miembros:

- i. En Miembro 1: usted, introduzca su Nombre de miembro para mostrar tal y como desee que aparezca en la colaboración.

 Note

Tu Cuenta de AWS ID se incluye automáticamente en el Cuenta de AWS ID de miembro.

- ii. En el caso del miembro 2, introduzca el nombre visible y el Cuenta de AWS ID de miembro del miembro al que desee invitar a la colaboración.

El Nombre del miembro para mostrar y el ID de Cuenta de AWS del miembro estarán visibles para todos los invitados a participar en la colaboración. Una vez ingresados y guardados los valores de estos campos, no se podrán editar.

 Note

Debe informar al miembro de la colaboración de que la información de los campos ID de Cuenta de AWS del miembro y Nombre del miembro para mostrar estará visible para todos los colaboradores activos de la colaboración.

- iii. Si desea añadir otro miembro, elija Añadir otro miembro. A continuación, introduzca el nombre de usuario y el Cuenta de AWS ID de miembro de cada miembro que pueda aportar los datos que desee invitar a la colaboración.
- c. Si desea habilitar el registro de análisis, seleccione la casilla Habilitar el registro de análisis.
 - Seleccione la casilla Registrar a partir de consultas en los tipos de registro admitidos.

Recibirás los registros generados a partir de consultas SQL en tu cuenta de Amazon CloudWatch Logs.

- d. Si desea activar la supervisión detallada, seleccione la casilla Habilitar la supervisión detallada.

El responsable del análisis y el pagador configurado pueden optar por habilitar las métricas detalladas al crear su membresía. Cuando esté habilitada, se publicarán métricas de monitoreo detalladas CloudWatch para el monitoreo operativo de las colaboraciones, incluido el rendimiento de las consultas y la utilización de los recursos. Estas métricas

estarán disponibles para el responsable del análisis y el pagador configurado en sus respectivos Cuenta de AWS sistemas.

Para obtener más información sobre CloudWatch los precios, consulte [CloudWatch Precios](#).

- e. En Regiones de resultados de consultas permitidas, selecciona uno o varios Regiones de AWS lugares a los que quieras enviar los resultados de las consultas.

De forma predeterminada, solo se selecciona la región actual (como North Virginia us-east-1).

 Important

Si habilita la entrega de resultados de consultas entre regiones, es posible que sus resultados se procesen y almacenen fuera de la región de origen.

Para obtener más información sobre las regiones, consulte [Regiones y puntos finales](#) en la Referencia general de AWS

- f. (Opcional) Gestione el acceso a sus datos con la aprobación automática de las solicitudes de cambio configurando qué ajustes se pueden cambiar automáticamente sin necesidad de aprobaciones manuales para una solicitud de cambio. De forma predeterminada, algunos ajustes solo se pueden cambiar mediante el envío de una solicitud de cambio que debe ser aprobada por todos los miembros antes de que entre en vigor.
 - Otorgue habilidades a los miembros: elija las habilidades que se pueden otorgar a los miembros de la colaboración sin aprobación manual. Los miembros siempre pueden aportar datos.
 - Elige habilidades:
 - Aportar datos (siempre activado)
 - Recibir resultados
 - Aprueba automáticamente a los nuevos miembros con estas habilidades: si está permitido, cualquier miembro que se añada con las habilidades seleccionadas anteriormente se unirá a la colaboración al instante. Los miembros a los que se añadan otras habilidades seguirán necesitando la aprobación manual para unirse.
 - Habilidades que se pueden revocar automáticamente: elige las habilidades que se pueden revocar sin aprobación manual. Los miembros siempre pueden aportar datos.

- Elige habilidades:
 - Aportar datos (siempre activado)
 - Recibir resultados

Si elige esta opción, puede realizar un seguimiento de todas las modificaciones en la configuración de la colaboración a través del historial de solicitudes de cambio, que se encuentra en la pestaña Detalles de la página de detalles de la colaboración.

- g. (Opcional) Si desea habilitar la capacidad de computación criptográfica, seleccione la casilla **Habilitar la computación criptográfica**.

- i. Elija los siguientes parámetros de cobertura criptográfica:

- Permitir columnas plaintext

Elija No si necesita tablas totalmente cifradas.

Elija Sí si desea permitir las columnas cleartext en la tabla cifrada.

Para ejecutar SUM o AVG en columnas específicas, las columnas deben estar en cleartext.

- Conservar valores NULL

Elija No si no desea conservar los valores NULL. Los valores NULL no aparecerán como NULL en una tabla cifrada.

Elija Sí si desea conservar los valores NULL. Los valores NULL aparecerán como NULL en una tabla cifrada.

- ii. Elija los siguientes parámetros de toma de huellas digitales:

- Permitir duplicados

Elija No si no desea permitir las entradas duplicadas en una columna fingerprint.

Elija Sí si desea permitir las entradas duplicadas en una columna fingerprint.

- Permitir JOIN de columnas con nombres diferentes

Seleccione No si no desea combinar las columnas fingerprint con nombres diferentes.

Elija Sí si desea combinar las columnas fingerprint con nombres diferentes.

Para obtener más información acerca de los parámetros de computación criptográfica, consulte [Parámetros de computación criptográfica](#).

Para obtener más información sobre cómo cifrar los datos para utilizarlos en AWS Clean Rooms, consulte. [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#)

 Note

Compruebe estas configuraciones detenidamente antes de completar el siguiente paso. Tras crear la colaboración, solo podrá editar el nombre y la descripción de la colaboración y si los registros se almacenan en Amazon CloudWatch Logs.

- h. Si desea habilitar la opción de Etiquetas para el recurso de colaboración, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
 - i. Elija Siguiente.
5. En el paso 2: especifique las capacidades de los miembros; en el caso de Análisis mediante consultas y trabajos, en la sección Tipos de análisis compatibles, deje la casilla Consultas seleccionada y tome las medidas recomendadas en función de su objetivo.

Su objetivo	Acción recomendada
Consultar los datos de la colaboración y recibir los resultados	1. Elijase a usted mismo como miembro que puede Ejecutar las consultas. 2. Elija uno de los miembros que pueden recibir los resultados de los análisis de la lista desplegable.
Consultar los datos de la colaboración y designar a un miembro distinto para recibir los resultados	1. Elijase a usted mismo como miembro que puede Ejecutar las consultas. 2. Seleccione el miembro que puede recibir los resultados de los análisis en la lista desplegable.

Su objetivo	Acción recomendada
Recibir los resultados de la consulta en la colaboración y designar a un miembro distinto para consultar los datos	1. Seleccione el miembro que puede Ejecutar las consultas en la lista desplegable. 2. Elija uno de los miembros que pueden recibir los resultados de los análisis de la lista desplegable.
Crear y administrar la colaboración, designar a un miembro distinto para consultar los datos y designar a un miembro distinto para recibir los resultados	1. Seleccione el miembro que puede Ejecutar las consultas en la lista desplegable. 2. Seleccione el miembro que puede recibir los resultados de los análisis en la lista desplegable.

- a. Si utiliza Clean Rooms ML, para modelar el aprendizaje automático mediante flujos de trabajo diseñados específicamente,
 - i. (Opcional) Seleccione en la lista desplegable el miembro que puede recibir resultados de modelos entrenados.
 - ii. (Opcional) Seleccione el miembro que puede recibir los resultados de la inferencia del modelo en la lista desplegable.
 - b. Vea las habilidades de los miembros en la sección Resolución de ID utilizando. AWS Entity Resolution
 - c. Elija Siguiente.
6. Para el paso 3: configurar el pago, en el caso del análisis mediante consultas, realice una de las siguientes acciones en función de su objetivo.

Su objetivo	Acción recomendada
Designar al miembro que puede Ejecutar las consultas como miembro que paga los costos de computación de las consultas	1. En el caso de Análisis mediante consultas , elija el miembro que pagará por las consultas para que sea el mismo que el miembro que puede ejecutar consultas.

Su objetivo	Acción recomendada
	2. Elija Siguiente.
Designar a otro miembro que pague los costos de computación de las consultas	1. En el caso de Analysis mediante consultas , elija usted mismo como el miembro que pagará por las consultas. 2. Elija Siguiente.

En el caso del modelado de aprendizaje automático mediante flujos de trabajo diseñados específicamente, el creador del modelo similar configurado es el miembro que pagará por el modelado similar.

En el caso de la resolución de ID AWS Entity Resolution, el creador de la tabla de mapeo de ID es el miembro que pagará por la tabla de mapeo de ID.

7. Para el paso 4: configurar la membresía, elige una de las siguientes opciones:

Yes, join by creating membership now

1. Para la configuración de resultados predeterminada, para la configuración de resultados de consultas, si usted es el miembro que puede recibir los resultados,
 - a. Para el destino de los resultados en Amazon S3, introduzca el destino de Amazon S3 o elija Browse S3 para seleccionar un bucket de S3.
 - b. Como Formato de los resultados de la consulta, elija CSV o PARQUET.
 - c. (Solo Spark) Para los archivos de resultados, elija Múltiple o Único.
 - d. (Opcional) Para acceder al servicio, si desea enviar consultas que tarden hasta 24 horas en llegar a su destino de S3, active la casilla Añadir una función de servicio para atender las consultas que tarden hasta 24 horas en completarse.

Las consultas de gran tamaño que tarden hasta 24 horas en completarse se enviarán a su destino de S3.

Si no selecciona la casilla de verificación, solo las consultas que se completen en un plazo de 12 horas se enviarán a su ubicación de S3.

- e. Especifique los permisos de Acceso a servicios seleccionando Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.

Si eliges...	Entonces...
Crear y usar un nuevo rol de servicio	• AWS Clean Rooms crea un rol de servicio con la política requerida para esta tabla. • El Nombre del rol de servicio predeterminado es <code>cleanrooms-result-receiver- <timestamp></code> • Debe tener permisos para crear roles y adjuntar políticas.
Usar un rol de servicio existente	- Selecione un Nombre de rol de servicio existente en la lista desplegable. Si tiene permisos de listas de roles, se mostrará la lista de roles. Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar. - Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM. Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible. De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

 Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).
- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes AWS Clean Rooms. Debe agregar la política de rol antes de continuar.
- Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para el rol de servicio.

2. En la configuración de Logs, elige una de las siguientes opciones para el almacenamiento de CloudWatch registros en Amazon Logs:

 Note

La sección de configuración de registros aparece si decide habilitar el registro de consultas.

- a. Selecciona Activar y los registros de consultas relevantes para ti se almacenarán en tu cuenta de Amazon CloudWatch Logs.

Cada miembro solo puede recibir los registros de las consultas que haya iniciado o que contengan sus datos.

El miembro que puede recibir los resultados también recibe los registros de todas las consultas que se ejecutan en una colaboración, aunque no se accede a los datos en una consulta.

En Tipos de registro admitidos, la casilla de verificación Query logs está activada de forma predeterminada.

 Note

Tras activar el registro de consultas, el almacenamiento de registros puede tardar unos minutos en configurarse y empezar a recibir registros en Amazon CloudWatch Logs. Durante este breve período, el miembro que puede realizar consultas puede ejecutar consultas que, en realidad, no enviarán registros.

- b. Selecciona Desactivar y los registros de consultas relevantes para ti no se almacenarán en tu cuenta de Amazon CloudWatch Logs.
3. Si desea habilitar la opción de Etiquetas para el recurso de pertenencia, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
4. Si usted es el miembro que paga por Query Compute, indique su aceptación marcando la casilla Acepto pagar los costos de procesamiento de esta colaboración.

 Note

Debe seleccionar esta casilla de verificación para continuar.
Para obtener más información sobre cómo se calculan los precios, consulte [Precios para AWS Clean Rooms](#).

Si es el [miembro que paga los costes de cálculo](#) de las [consultas, pero no el miembro que puede realizar la consulta](#), se recomienda configurar un presupuesto AWS Clean Rooms y recibir notificaciones cuando se alcance el presupuesto máximo. AWS Budgets Para obtener más información sobre cómo configurar un presupuesto, consulte [Administración de costos con AWS Budgets](#) en la Guía del usuario de AWS Cost Management . Para obtener más información sobre cómo configurar notificaciones, consulte [Creación de un tema de Amazon SNS para las notificaciones de presupuesto](#) en la Guía del usuario de AWS Cost Management . Si se ha alcanzado el presupuesto máximo, puede ponerse en contacto con el miembro que puede realizar consultas o [abandonar la colaboración](#). Si abandona la colaboración, no se podrán ejecutar más consultas y, por lo tanto, ya no se le facturarán costos de cómputo de consultas.

5. Elija Siguiente.

Se crean tanto la colaboración como su pertenencia a ella.

Su estado en la colaboración es activo.

No, I will create a membership later

1. Elija Siguiente.

Se crea solo la colaboración.

Su estado en la colaboración es inactivo.

8. Para el paso 5: revisar y crear, haga lo siguiente:

- a. Revise las selecciones que realizó en los pasos anteriores y edítelas si es necesario.
- b. Elija una de las opciones.

Si ha elegido...	Entonces seleccione...
Crear una pertenencia en la colaboración (Sí, unirme y crear pertenencia ahora)	Crear colaboración y pertenencia
Crear la colaboración y no crear una pertenencia en este momento (No, crearé una pertenencia más tarde)	Crear colaboración

Una vez que la colaboración se haya creado correctamente, podrá ver la página de detalles de la colaboración en Colaboraciones.

Ya puede hacer lo siguiente:

- [Prepare la tabla de datos para analizarla AWS Clean Rooms](#). (Opcional si desea analizar sus propios datos de eventos o si desea consultar datos de identidad).
- [Asociar la tabla configurada a su colaboración](#) (Opcional si quieres analizar los datos de tus propios eventos).
- [Agregue una regla de análisis para la tabla configurada](#). (Opcional si quieres analizar los datos de tus propios eventos).
- [Crear una pertenencia y unirse a una colaboración](#). (Opcional si ya ha creado una suscripción).
- [Invita a los miembros a unirse a la colaboración](#).

Crear una colaboración para consultas y trabajos

En este procedimiento, usted, como [creador de la colaboración](#), realiza las siguientes tareas:

- [Cree una colaboración](#).
- Invita a uno o más [miembros](#) a la [colaboración](#).
- Asigne capacidades a los miembros, como el [miembro que puede ejecutar consultas y trabajos](#) y el [miembro que puede recibir los resultados](#).

Si el creador de la colaboración es también el miembro que puede recibir los resultados, especificará el destino y el formato de los resultados. También proporcionan una función de servicio Amazon Resource Name (ARN) para escribir los resultados en el destino de los resultados.

- Configure qué [miembro es responsable de pagar los costos de consulta y procesamiento del trabajo en la colaboración](#).

Antes de comenzar, asegúrese de que cumple los siguientes requisitos previos:

- Tiene el nombre y el Cuenta de AWS ID de cada miembro al que quiere invitar a la colaboración.
- Tiene permiso para compartir el nombre y la Cuenta de AWS identificación de cada miembro con todos los miembros de la colaboración.

Note

No puedes añadir más miembros después de crear la colaboración.

Para obtener información sobre cómo crear una colaboración mediante la AWS SDKs, consulta la [referencia de la AWS Clean Rooms API](#).

Para crear una colaboración para consultas y trabajos

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con la Cuenta de AWS que funcionará como creador de colaboraciones.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la esquina superior derecha, elija Crear colaboración.
4. En Paso 1: Definir la colaboración, haga lo siguiente:

- a. En Detalles, introduzca el Nombre y la Descripción de la colaboración.

Esta información estará visible para los miembros de la colaboración que estén invitados a participar en la colaboración. La información de Nombre y Descripción les ayudará a entender a qué se refiere la colaboración.

- b. En Miembros:

- i. En Miembro 1: usted, introduzca su Nombre de miembro para mostrar tal y como desee que aparezca en la colaboración.

 Note

Tu Cuenta de AWS ID se incluye automáticamente en el Cuenta de AWS ID de miembro.

- ii. En el caso del miembro 2, introduzca el nombre visible y el Cuenta de AWS ID de miembro del miembro al que desee invitar a la colaboración.

El Nombre del miembro para mostrar y el ID de Cuenta de AWS del miembro estarán visibles para todos los invitados a participar en la colaboración. Una vez ingresados y guardados los valores de estos campos, no se podrán editar.

 Note

Debe informar al miembro de la colaboración de que la información de los campos ID de Cuenta de AWS del miembro y Nombre del miembro para mostrar estará visible para todos los colaboradores activos de la colaboración.

- iii. Si desea añadir otro miembro, elija Añadir otro miembro. A continuación, introduzca el nombre de usuario y el Cuenta de AWS ID de miembro de cada miembro que pueda aportar los datos que desee invitar a la colaboración.
- c. Si desea habilitar el registro de análisis, seleccione la casilla Habilitar el registro de análisis y, a continuación, elija los tipos de registro compatibles.
 - Si desea recibir los registros generados a partir de consultas SQL, seleccione la casilla Registros a partir de consultas.

- Si desea recibir los registros generados a partir de los trabajos que se utilizan PySpark, seleccione la casilla Registros de los trabajos.
- d. Si desea activar la supervisión detallada, seleccione la casilla Habilitar la supervisión detallada.

El responsable del análisis y el pagador configurado pueden optar por habilitar las métricas detalladas al crear su membresía. Cuando esté habilitada, se publicarán métricas de monitoreo detalladas CloudWatch para el monitoreo operativo de las colaboraciones, incluido el rendimiento de las consultas y la utilización de los recursos. Estas métricas estarán disponibles para el responsable del análisis y el pagador configurado en sus respectivos Cuenta de AWS sistemas.

Para obtener más información sobre CloudWatch los precios, consulte [CloudWatch Precios](#).

- e. En Regiones de resultados de consultas permitidas, selecciona uno o varios Regiones de AWS lugares a los que quieras enviar los resultados de las consultas.

De forma predeterminada, solo se selecciona la región actual (como North Virginia us-east-1).

 Important

Si habilita la entrega de resultados de consultas entre regiones, es posible que sus resultados se procesen y almacenen fuera de la región de origen.

Para obtener más información sobre las regiones, consulte [Regiones y puntos finales](#) en la Referencia general de AWS

- f. (Opcional) Gestione el acceso a sus datos con la aprobación automática de las solicitudes de cambio configurando qué ajustes se pueden cambiar automáticamente sin necesidad de aprobaciones manuales para una solicitud de cambio. De forma predeterminada, algunos ajustes solo se pueden cambiar mediante el envío de una solicitud de cambio que debe ser aprobada por todos los miembros antes de que entre en vigor.
- Otorgue habilidades a los miembros: elija las habilidades que se pueden otorgar a los miembros de la colaboración sin aprobación manual. Los miembros siempre pueden aportar datos.
 - Elige habilidades:

- Aportar datos (siempre activado)
- Recibir resultados
- Aprueba automáticamente a los nuevos miembros con estas habilidades: si está permitido, cualquier miembro que se añada con las habilidades seleccionadas anteriormente se unirá a la colaboración al instante. Los miembros a los que se añadan otras habilidades seguirán necesitando la aprobación manual para unirse.
- Habilidades que se pueden revocar automáticamente: elige las habilidades que se pueden revocar sin aprobación manual. Los miembros siempre pueden aportar datos.
 - Elige habilidades:
 - Aportar datos (siempre activado)
 - Recibir resultados

Si elige esta opción, puede realizar un seguimiento de todas las modificaciones en la configuración de la colaboración a través del historial de solicitudes de cambio, que se encuentra en la pestaña Detalles de la página de detalles de la colaboración.

g. (Opcional) Si desea habilitar la capacidad de computación criptográfica, seleccione la casilla **Habilitar la computación criptográfica**.

i. Elija los siguientes parámetros de cobertura criptográfica:

- Permitir columnas plaintext

Elija **No** si necesita tablas totalmente cifradas.

Elija **Sí** si desea permitir las columnas cleartext en la tabla cifrada.

Para ejecutar SUM o AVG en columnas específicas, las columnas deben estar en cleartext.

- Conservar valores NULL

Elija **No** si no desea conservar los valores NULL. Los valores NULL no aparecerán como NULL en una tabla cifrada.

Elija **Sí** si desea conservar los valores NULL. Los valores NULL aparecerán como NULL en una tabla cifrada.

ii. Elija los siguientes parámetros de toma de huellas digitales:

- Permitir duplicados

Elija No si no desea permitir las entradas duplicadas en una columna fingerprint.

Elija Sí si desea permitir las entradas duplicadas en una columna fingerprint.

- Permitir JOIN de columnas con nombres diferentes

Seleccione No si no desea combinar las columnas fingerprint con nombres diferentes.

Elija Sí si desea combinar las columnas fingerprint con nombres diferentes.

Para obtener más información acerca de los parámetros de computación criptográfica, consulte [Parámetros de computación criptográfica](#).

Para obtener más información sobre cómo cifrar los datos para utilizarlos en AWS Clean Rooms, consulte. [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#)

 Note

Compruebe estas configuraciones detenidamente antes de completar el siguiente paso. Tras crear la colaboración, solo podrá editar el nombre y la descripción de la colaboración y si los registros se almacenan en Amazon CloudWatch Logs.

- h. Si desea habilitar la opción de Etiquetas para el recurso de colaboración, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
 - i. Elija Siguiente.
 5. Para el paso 2: especificar las capacidades de los miembros, haga lo siguiente:
 - a. Para el análisis mediante consultas y trabajos, en Tipos de análisis admitidos, seleccione la casilla de verificación Trabajos.

La casilla de verificación Consultas está seleccionada de forma predeterminada.

- i. Seleccione el miembro que puede ejecutar consultas y trabajos en la lista desplegable.
 - ii. Seleccione el miembro que puede recibir los resultados de los análisis en la lista desplegable.

 Note

El miembro que crea la plantilla de PySpark análisis también debe ser el miembro que reciba los resultados.

- b. Si utiliza Clean Rooms ML, para modelar el aprendizaje automático mediante flujos de trabajo diseñados específicamente,
 - i. (Opcional) Seleccione en la lista desplegable el miembro que puede recibir resultados de modelos entrenados.
 - ii. (Opcional) Seleccione el miembro que puede recibir los resultados de la inferencia del modelo en la lista desplegable.
 - c. Vea las habilidades de los miembros en la sección Resolución de ID utilizando. AWS Entity Resolution
 - d. Elija Siguiente.
6. Para el paso 3: configurar el pago,
- a. Para el análisis mediante consultas y trabajos, elija al miembro que pagará por las consultas y los trabajos.

Puede asignar al miembro que puede ejecutar consultas y trabajos para que sea el miembro que pague las consultas y los trabajos calculen los costes.

Puede asignar a un miembro diferente para que pague las consultas y los costes de cálculo de los trabajos.
 - b. En el caso del modelado de aprendizaje automático mediante flujos de trabajo diseñados específicamente, el creador del modelo similar configurado es el miembro que pagará por el modelado similar.
 - c. En el caso de la resolución de ID AWS Entity Resolution, el creador de la tabla de mapeo de ID es el miembro que pagará por la tabla de mapeo de ID.
 - d. Elija Siguiente.
7. Para el paso 4: configurar la membresía, elige una de las siguientes opciones:

Yes, join by creating membership now

1. Para la configuración de resultados predeterminada, para la configuración de resultados de consultas, si usted es el miembro que puede recibir los resultados,
 - a. Seleccione la casilla de verificación Establecer la configuración predeterminada para las consultas. Para el destino de los resultados en Amazon S3, introduzca el destino de Amazon S3 o elija Browse S3 para seleccionar un bucket de S3.
 - b. Como Formato de los resultados de la consulta, elija CSV o PARQUET.
 - c. (Solo Spark) Para los archivos de resultados, elija Múltiple o Único.
 - d. (Opcional) Para acceder al servicio, si desea enviar consultas que tarden hasta 24 horas en llegar a su destino de S3, active la casilla Añadir una función de servicio para atender las consultas que tarden hasta 24 horas en completarse.

Las consultas de gran tamaño que tarden hasta 24 horas en completarse se enviarán a su destino de S3.

Si no selecciona la casilla de verificación, solo las consultas que se completen en un plazo de 12 horas se enviarán a su ubicación de S3.

- e. Especifique los permisos de Acceso a servicios seleccionando Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.

Si eliges...	Entonces...
Crear y usar un nuevo rol de servicio	• AWS Clean Rooms crea un rol de servicio con la política requerida para esta tabla. • El Nombre del rol de servicio predeterminado es <code>cleanrooms-result-receiver-<timestamp></code> • Debe tener permisos para crear roles y adjuntar políticas.

Si eliges...	Entonces...
Usar un rol de servicio existente	i. Seleccione un Nombre de rol de servicio existente en la lista desplegable. Si tiene permisos de listas de roles, se mostrará la lista de roles. Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar. ii. Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM. Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible. De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

 Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).
- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes AWS Clean Rooms. Debe agregar la política de rol antes de continuar.

- Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para el rol de servicio.

2. Para ver los resultados de Job,

Example

Por ejemplo: **s3://bucket/prefix**

- a. Seleccione la casilla Establecer la configuración predeterminada para los trabajos y, a continuación, especifique el destino de los resultados en Amazon S3 introduciendo el destino de S3 o elija Browse S3 para seleccionarlos de una lista de buckets de S3 disponibles.
 - b. Especifique los permisos de acceso al servicio eligiendo un nombre de rol de servicio existente en la lista desplegable.
3. En la configuración de Logs, elige una de las siguientes opciones para el almacenamiento de CloudWatch registros en Amazon Logs:

Note

La sección de configuración de registros aparece si decide habilitar el registro de consultas.

- a. Selecciona Activar y los registros de consultas relevantes para ti se almacenarán en tu cuenta de Amazon CloudWatch Logs.

Cada miembro solo puede recibir los registros de las consultas que haya iniciado o que contengan sus datos.

El miembro que puede recibir los resultados también recibe los registros de todas las consultas que se ejecutan en una colaboración, aunque no se accede a los datos en una consulta.

En Tipos de registro compatibles, elige entre los tipos de registro que el creador de la colaboración ha decidido admitir:

En Tipos de registro compatibles, las casillas de verificación Registros de consultas y Registros de trabajos están activadas de forma predeterminada.

 Note

Tras activar el registro de análisis, el almacenamiento de registros puede tardar unos minutos en configurarse y empezar a recibir registros en Amazon CloudWatch Logs. Durante este breve período, el miembro que puede realizar consultas puede ejecutar consultas que, en realidad, no enviarán registros.

- b. Selecciona Desactivar y los registros de consultas relevantes para ti no se almacenarán en tu cuenta de Amazon CloudWatch Logs.
4. Si quieres activar las etiquetas de membresía para el recurso de membresía, selecciona Añadir nueva etiqueta y, a continuación, introduce el par clave y valor.
5. Si usted es el miembro que paga por Query compute, Job compute o ambos, indique su aceptación seleccionando la casilla Acepto pagar los costos de procesamiento en esta colaboración.

 Note

Debe seleccionar esta casilla de verificación para continuar.
Para obtener más información sobre cómo se calculan los precios, consulte [Precios para AWS Clean Rooms](#).

Si es el [miembro que paga los costes de cálculo](#) de las [consultas, pero no el miembro que puede realizar la consulta](#), se recomienda configurar un presupuesto AWS Clean Rooms y recibir notificaciones cuando se alcance el presupuesto máximo. AWS Budgets Para obtener más información sobre cómo configurar un presupuesto, consulte [Administración de costos con AWS Budgets](#) en la Guía del usuario de AWS Cost Management . Para obtener más información sobre cómo configurar notificaciones, consulte [Creación de un tema de Amazon SNS para las notificaciones de presupuesto](#) en la Guía del usuario de AWS Cost Management . Si se ha alcanzado el presupuesto máximo, puede ponerse en contacto con el miembro que puede realizar consultas o [abandonar la colaboración](#). Si abandona la colaboración, no se podrán ejecutar más consultas y, por lo tanto, ya no se le facturarán costos de cómputo de consultas.

6. Elija Siguiente.

Se crean tanto la colaboración como su pertenencia a ella.

Su estado en la colaboración es activo.

No, I will create a membership later

1. Elija Siguiente.

Se crea solo la colaboración.

Su estado en la colaboración es inactivo.

8. Para el paso 5: revisar y crear, haga lo siguiente:

- a. Revise las selecciones que realizó en los pasos anteriores y edítelas si es necesario.
- b. Elija una de las opciones.

Si ha elegido...	Entonces seleccione...
Crear una pertenencia en la colaboración (Sí, unirme y crear pertenencia ahora)	Crear colaboración y pertenencia
Crear la colaboración y no crear una pertenencia en este momento (No, crearé una pertenencia más tarde)	Crear colaboración

Una vez que la colaboración se haya creado correctamente, podrá ver la página de detalles de la colaboración en Colaboraciones.

Ya puede hacer lo siguiente:

- [Prepare la tabla de datos para analizarla AWS Clean Rooms](#). (Opcional si desea analizar sus propios datos de eventos o si desea consultar datos de identidad).
- [Asociar la tabla configurada a su colaboración](#) (Opcional si quieres analizar los datos de tus propios eventos).
- [Agregue una regla de análisis para la tabla configurada](#). (Opcional si quieres analizar los datos de tus propios eventos).

- [Crear una pertenencia y unirse a una colaboración](#). (Opcional si ya ha creado una suscripción).
- [Invita a los miembros a unirse a la colaboración](#).

Crear una colaboración para el modelado de aprendizaje automático

En este procedimiento, usted, como [creador de la colaboración](#), realiza las siguientes tareas:

- [Cree una colaboración](#).
- Invita a uno o más [miembros](#) a la [colaboración](#).
- Asigne habilidades a los miembros, como
 - [Miembro que puede realizar consultas](#)
 - [Miembro que puede recibir los resultados](#)
 - Miembro que puede recibir resultados de modelos entrenados
 - Miembro que puede generar resultados a partir de la inferencia del modelo

Si el creador de la colaboración es también el miembro que puede recibir los resultados, especificará el destino y el formato de los resultados. También proporcionan una función de servicio Amazon Resource Name (ARN) para escribir los resultados en el destino de los resultados.

- Configure qué [miembro es responsable de pagar los costos de cómputo, la capacitación del modelo y los costos de inferencia del modelo en la colaboración](#).

Antes de comenzar, asegúrese de que cumple los siguientes requisitos previos:

- Tiene el nombre y el Cuenta de AWS ID de cada miembro al que quiere invitar a la colaboración.
- Tiene permiso para compartir el nombre y la Cuenta de AWS identificación de cada miembro con todos los miembros de la colaboración.

Note

No puedes añadir más miembros después de crear la colaboración.

Para obtener información sobre cómo crear una colaboración mediante la AWS SDKs, consulta la [referencia de la AWS Clean Rooms API](#).

Para crear una colaboración para el modelado de aprendizaje automático

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#) con la Cuenta de AWS que funcionará como creador de la colaboración.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la esquina superior derecha, elija Crear colaboración.
4. En Paso 1: Definir la colaboración, haga lo siguiente:
 - a. En Detalles, introduzca el Nombre y la Descripción de la colaboración.

Esta información estará visible para los miembros de la colaboración que estén invitados a participar en la colaboración. La información de Nombre y Descripción les ayudará a entender a qué se refiere la colaboración.

- b. En Miembros:
 - i. En Miembro 1: usted, introduzca su Nombre de miembro para mostrar tal y como desee que aparezca en la colaboración.

 Note

Tu Cuenta de AWS ID se incluye automáticamente en el Cuenta de AWS ID de miembro.

- ii. En el caso del miembro 2, introduzca el nombre visible y el Cuenta de AWS ID de miembro del miembro al que desee invitar a la colaboración.

El Nombre del miembro para mostrar y el ID de Cuenta de AWS del miembro estarán visibles para todos los invitados a participar en la colaboración. Una vez ingresados y guardados los valores de estos campos, no se podrán editar.

 Note

Debe informar al miembro de la colaboración de que la información de los campos ID de Cuenta de AWS del miembro y Nombre del miembro para mostrar estará visible para todos los colaboradores activos de la colaboración.

- iii. Si desea añadir otro miembro, elija Añadir otro miembro. A continuación, introduzca el nombre de usuario y el Cuenta de AWS ID de miembro de cada miembro que pueda aportar los datos que desee invitar a la colaboración.
- c. Si desea habilitar el registro de análisis, seleccione la casilla Habilitar el registro de análisis y, a continuación, en Tipos de registro compatibles, elija Registros a partir de consultas.
- d. Si desea activar la supervisión detallada, seleccione la casilla Habilitar la supervisión detallada.

El responsable del análisis y el pagador configurado pueden optar por habilitar las métricas detalladas al crear su membresía. Cuando esté habilitada, se publicarán métricas de monitoreo detalladas CloudWatch para el monitoreo operativo de las colaboraciones, incluido el rendimiento de las consultas y la utilización de los recursos. Estas métricas estarán disponibles para el responsable del análisis y el pagador configurado en sus respectivos Cuenta de AWS sistemas.

Para obtener más información sobre CloudWatch los precios, consulte [CloudWatch Precios](#).

- e. En Regiones de resultados de consultas permitidas, selecciona uno o varios Regiones de AWS lugares a los que quieras enviar los resultados de las consultas.

De forma predeterminada, solo se selecciona la región actual (como North Virginia us-east-1).

 Important

Si habilita la entrega de resultados de consultas entre regiones, es posible que sus resultados se procesen y almacenen fuera de la región de origen.

Para obtener más información sobre las regiones, consulte [Regiones y puntos finales](#) en la Referencia general de AWS

- f. (Opcional) Gestione el acceso a sus datos con la aprobación automática de las solicitudes de cambio configurando qué ajustes se pueden cambiar automáticamente sin necesidad de aprobaciones manuales para una solicitud de cambio. De forma predeterminada, algunos ajustes solo se pueden cambiar mediante el envío de una solicitud de cambio que debe ser aprobada por todos los miembros antes de que entre en vigor.

- Otorgue habilidades a los miembros: elija las habilidades que se pueden otorgar a los miembros de la colaboración sin aprobación manual. Los miembros siempre pueden aportar datos.
- Elige habilidades:
 - Aportar datos (siempre activado)
 - Recibir resultados
- Aprueba automáticamente a los nuevos miembros con estas habilidades: si está permitido, cualquier miembro que se añada con las habilidades seleccionadas anteriormente se unirá a la colaboración al instante. Los miembros a los que se añadan otras habilidades seguirán necesitando la aprobación manual para unirse.
- Habilidades que se pueden revocar automáticamente: elige las habilidades que se pueden revocar sin aprobación manual. Los miembros siempre pueden aportar datos.
- Elige habilidades:
 - Aportar datos (siempre activado)
 - Recibir resultados

Si elige esta opción, puede realizar un seguimiento de todas las modificaciones en la configuración de la colaboración a través del historial de solicitudes de cambio, que se encuentra en la pestaña Detalles de la página de detalles de la colaboración.

- g. (Opcional) Si desea habilitar la capacidad de computación criptográfica, seleccione la casilla **Habilitar la computación criptográfica**.

- i. Elija los siguientes parámetros de cobertura criptográfica:

- Permitir columnas plaintext

Elija **No** si necesita tablas totalmente cifradas.

Elija **Sí** si desea permitir las columnas cleartext en la tabla cifrada.

Para ejecutar SUM o AVG en columnas específicas, las columnas deben estar en cleartext.

- Conservar valores NULL

Elija No si no desea conservar los valores NULL. Los valores NULL no aparecerán como NULL en una tabla cifrada.

Elija Sí si desea conservar los valores NULL. Los valores NULL aparecerán como NULL en una tabla cifrada.

ii. Elija los siguientes parámetros de toma de huellas digitales:

- Permitir duplicados

Elija No si no desea permitir las entradas duplicadas en una columna fingerprint.

Elija Sí si desea permitir las entradas duplicadas en una columna fingerprint.

- Permitir JOIN de columnas con nombres diferentes

Seleccione No si no desea combinar las columnas fingerprint con nombres diferentes.

Elija Sí si desea combinar las columnas fingerprint con nombres diferentes.

Para obtener más información acerca de los parámetros de computación criptográfica, consulte [Parámetros de computación criptográfica](#).

Para obtener más información sobre cómo cifrar los datos para utilizarlos en AWS Clean Rooms, consulte. [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#)

 Note

Compruebe estas configuraciones detenidamente antes de completar el siguiente paso. Tras crear la colaboración, solo podrá editar el nombre y la descripción de la colaboración y si los registros se almacenan en Amazon CloudWatch Logs.

h. Si desea habilitar la opción de Etiquetas para el recurso de colaboración, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.

i. Elija Siguiente.

5. Para el paso 2: especifique las capacidades de los miembros,

- a. Para el análisis mediante consultas y trabajos, en la sección Tipos de análisis admitidos, deje seleccionada la casilla Consultas.
 - b. En Ejecutar consultas, elija el miembro que iniciará el entrenamiento del modelo
 - c. En Recibir los resultados de los análisis, elija uno o más miembros que recibirán los resultados de la consulta.
 - d. Para el modelado de aprendizaje automático mediante flujos de trabajo diseñados específicamente,
 - i. En Recibir resultados de modelos entrenados, elija el miembro que recibirá los resultados del modelo entrenado, incluidos los artefactos y las métricas del modelo.
 - ii. En Recibir los resultados de la inferencia del modelo, elija el miembro que recibirá los resultados de la inferencia del modelo.
 - e. Vea las habilidades de los miembros bajo la resolución de ID utilizando. AWS Entity Resolution
6. Para el paso 3: configurar el pago,
- a. En Análisis mediante consultas, para Pagar por consultas, lleve a cabo una de las siguientes acciones:
 - Para que el mismo miembro pague y ejecute las consultas, seleccione el mismo miembro que eligió para Ejecutar consultas.
 - Para que otro miembro pague los costes de las consultas, selecciona tu cuenta de miembro.
 - b. Para el modelado de aprendizaje automático mediante flujos de trabajo diseñados específicamente,
 - Elija al miembro que pagará la formación de modelos.
 - c. Elija el miembro que pagará el trabajo de inferencia.
 - d. Para el modelado parecido a Pay for, no es necesario realizar ninguna acción. El creador del modelo similar configurado es el miembro que pagará por el modelado similar.
 - e. (Opcional) Elija al miembro que pagará la generación de datos sintéticos.
 - f. Para la resolución de ID con AWS Entity Resolution, no es necesario realizar ninguna acción. El creador de la tabla de mapeo de ID es el miembro que pagará por la tabla de mapeo de ID.
7. Elija Siguiente.

8. Para el paso 4: configurar la membresía, en Membresía de colaboración, elija una de las siguientes opciones:

Yes, join by creating membership now

1. Para la configuración de resultados predeterminada, para la configuración de resultados de consultas, si usted es el miembro que puede recibir los resultados,
 - a. Seleccione la casilla de verificación Establecer la configuración predeterminada para las consultas.
 - b. Para el destino de los resultados en Amazon S3, introduzca el destino de Amazon S3 o elija Browse S3 para seleccionar un bucket de S3.
 - c. Como Formato de los resultados de la consulta, elija CSV o PARQUET.
 - d. (Solo Spark) Para los archivos de resultados, elija Múltiple o Único.
 - e. (Opcional) Si desea enviar consultas que tarden hasta 24 horas en llegar a su destino de S3, seleccione la casilla Añadir una función de servicio para atender las consultas que tarden hasta 24 horas en completarse.

Las consultas grandes que tarden hasta 24 horas en completarse se enviarán a su destino de S3.

Si no selecciona la casilla de verificación, solo las consultas que se completen en un plazo de 12 horas se enviarán a su ubicación de S3.

- f. Especifique los permisos de Acceso a servicios seleccionando Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.

Si eliges...	Entonces...
Crear y usar un nuevo rol de servicio	• AWS Clean Rooms crea un rol de servicio con la política requerida para esta tabla. • El Nombre del rol de servicio predeterminado es <code>cleanrooms-result-receiver- <timestamp></code> • Debe tener permisos para crear roles y adjuntar políticas.

Si eliges...	Entonces...
Usar un rol de servicio existente	i. Seleccione un Nombre de rol de servicio existente en la lista desplegable. Si tiene permisos de listas de roles, se mostrará la lista de roles. Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar. ii. Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM. Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible. De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

 Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).
- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes AWS Clean Rooms. Debe agregar la política de rol antes de continuar.

- Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para el rol de servicio.

2. Para la configuración de ML,
 - a. Seleccione la casilla de verificación Crear configuración de ML y, a continuación, especifique el destino de salida del modelo en Amazon S3 introduciendo el destino de S3 o elija Browse S3 para seleccionarlo de una lista de buckets de S3 disponibles.
 - b. Especifique los permisos de acceso al servicio eligiendo entre Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.
 - c. Si el depósito de S3 está cifrado, active la casilla de verificación El depósito de destino está cifrado con una clave de KMS y, a continuación, introduzca AWS KMS key o seleccione Crear una AWS KMS key para crear una nueva clave de KMS.
3. Si desea habilitar las etiquetas de membresía para el recurso de membresía, elija Agregar nueva etiqueta y, a continuación, introduzca el par clave y valor.
4. Si usted es el miembro que paga por Query Compute, indique su aceptación marcando la casilla Acepto pagar los costos de computación de esta colaboración.

 Note

Debe seleccionar esta casilla de verificación para continuar.
Para obtener más información sobre cómo se calculan los precios, consulte [Precios para AWS Clean Rooms](#).

Si es el [miembro que paga los costes de cálculo](#) de las [consultas, pero no el miembro que puede realizar la consulta](#), se recomienda configurar un presupuesto AWS Clean Rooms y recibir notificaciones cuando se alcance el presupuesto máximo. AWS Budgets Para obtener más información sobre cómo configurar un presupuesto, consulte [Administración de costos con AWS Budgets](#) en la Guía del usuario de AWS Cost Management . Para obtener más información sobre cómo configurar notificaciones, consulte [Creación de un tema de Amazon SNS para las notificaciones de presupuesto](#) en la Guía del usuario de AWS Cost Management . Si se ha alcanzado el presupuesto máximo, puede ponerse en contacto con el miembro que puede realizar consultas o [abandonar la colaboración](#). Si abandona la colaboración, no se podrán ejecutar más consultas y, por lo tanto, ya no se le facturarán costos de cómputo de consultas.

5. Elija Siguiente.

Se crean tanto la colaboración como su pertenencia a ella.

Su estado en la colaboración es activo.

No, I will create a membership later

1. Elija Siguiente.

Se crea solo la colaboración.

Su estado en la colaboración es inactivo.

9. Para el paso 5: revisar y crear, haga lo siguiente:

- Revise las selecciones que realizó en los pasos anteriores y edítelas si es necesario.
- Elija una de las opciones.

Si ha elegido...	Entonces seleccione...
Crear una pertenencia en la colaboración (Sí, unirme y crear pertenencia ahora)	Crear colaboración y pertenencia
Crear la colaboración y no crear una pertenencia en este momento (No, crearé una pertenencia más tarde)	Crear colaboración

Crear una pertenencia y unirse a una colaboración

Una suscripción es un recurso que se crea cuando un miembro se une a una colaboración en AWS Clean Rooms.

Puedes unirte a una colaboración como

- [miembro que puede realizar consultas](#)
- [miembro que puede realizar consultas y trabajos](#)
- [miembro que puede recibir los resultados](#) de una consulta o un trabajo
- [miembro que paga los costos de cálculo de la consulta](#)

- [miembro que paga por consultas y trabajos](#)

Todos los miembros pueden contribuir con datos.

Para obtener información sobre cómo crear una membresía y unirse a una colaboración mediante la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

En este procedimiento, el miembro invitado [se une a la colaboración creando un recurso de suscripción](#).

Si el miembro invitado es el miembro que puede recibir los resultados, especificará el destino y el formato de los resultados. También proporcionan un ARN de rol de servicio para escribir en el destino de los resultados.

Si el miembro invitado es el responsable de pagar los costes de cálculo, acepta sus responsabilidades de pago antes de unirse a la colaboración.

Para crear una pertenencia y unirse a una colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu miembro Cuenta de AWS.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la pestaña Disponibles para combinar, en Colaboraciones disponibles para combinar, elija el Nombre de la colaboración.
4. En la página de detalles de la colaboración, en la sección Descripción general, consulta los detalles de la colaboración, incluidos los detalles de tu miembro y una lista de los demás miembros.

Compruebe que Cuenta de AWS IDs los miembros de cada uno de los miembros de la colaboración son aquellos con los que piensa iniciar la colaboración.

5. Elija Crear pertenencia.
6. En la página Crear membresía, en la sección Descripción general, consulta el nombre de la colaboración, la descripción de la colaboración, el Cuenta de AWS ID del creador de la colaboración, los detalles de tu miembro y el Cuenta de AWS ID del miembro que pagará las consultas.
7. Si el creador de la colaboración ha decidido habilitar el registro de análisis, elija una de las siguientes opciones para el almacenamiento de CloudWatch registros en Amazon Logs:

Si selecciona...	Entonces...
Activar	Los registros relevantes para usted se almacenan en Amazon CloudWatch Logs. Cada miembro solo puede recibir los registros de las consultas que haya iniciado o que contengan sus datos. El miembro que puede recibir los resultados también recibe los registros de todos los análisis realizados en una colaboración, incluso si no se accede a sus datos en un análisis. En Tipos de registro compatibles, elige entre los tipos de registro que el creador de la colaboración ha decidido admitir: 1. Si desea recibir los registros generados a partir de consultas SQL, seleccione la casilla Registros de consultas. 2. Si desea recibir los registros generados a partir de los trabajos que se utilizan PySpark, seleccione la casilla Registros de los trabajos.
Desactivar	Los registros de consultas relevantes para ti no se almacenan en tu cuenta de Amazon CloudWatch Logs.

 Note

Después de activar el registro de análisis, el almacenamiento de registros puede tardar unos minutos en configurarse y empezar a recibir registros en Amazon CloudWatch

Logs. Durante este breve período, el miembro que puede realizar consultas puede ejecutar consultas que, en realidad, no enviarán registros.

8. Si el creador de la colaboración ha activado la supervisión detallada para esta colaboración, elija si desea recibir métricas de observabilidad detalladas en su CloudWatch cuenta.

Para una supervisión detallada:

Opciones de monitoreo detalladas

Opción	Description (Descripción)
Activar	AWS Clean Rooms publicará métricas de monitoreo detalladas CloudWatch para esta colaboración en su cuenta. Puede utilizar estas métricas para la supervisión operativa, incluido el rendimiento de las consultas y la utilización de los recursos. Se aplican cargos adicionales de CloudWatch. Para obtener más información, consulte CloudWatch Precios.
Desactivar	No se exportará ninguna métrica detallada a tu CloudWatch cuenta. Si han activado esta opción, otros miembros podrán seguir viendo las métricas de monitorización detalladas en sus propias cuentas.

 Note

Las métricas de supervisión detalladas solo están disponibles para los miembros que pueden realizar consultas (ejecutores de análisis) y para los miembros que estén configurados como pagadores de la colaboración.

9. Si las capacidades de sus miembros incluyen Recibir resultados, la configuración de resultados predeterminada es la siguiente:
 - a. Para los resultados de las consultas, seleccione la casilla Establecer la configuración predeterminada para las consultas y, a continuación, especifique el destino de los resultados en Amazon S3 introduciendo el destino de S3 o elija Browse S3 para seleccionarlos de una lista de buckets de S3 disponibles.

Example

Por ejemplo: **s3://bucket/prefix**

- i. Para el formato de resultados, elija CSV o PARQUET.
- ii. (Solo Spark) Para los archivos de resultados, elija Múltiple o Único.
- iii. (Opcional) Para acceder al servicio, si desea enviar consultas que tarden hasta 24 horas en llegar a su destino de S3, active la casilla Añadir una función de servicio para atender las consultas que tarden hasta 24 horas en completarse.

Las consultas de gran tamaño que tarden hasta 24 horas en completarse se enviarán a su destino de S3.

Si no selecciona la casilla de verificación, solo las consultas que se completen en un plazo de 12 horas se enviarán a su ubicación de S3.

Note

Debe seleccionar un rol de servicio existente o tener permisos para crear uno nuevo. Para obtener más información, consulte [Creación de un rol de servicio para recibir resultados](#).

- iv. Especifique los permisos de Acceso a servicios seleccionando Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.

Create and use a new service role

- AWS Clean Rooms crea un rol de servicio con la política requerida para esta tabla.
- El Nombre del rol de servicio predeterminado es `cleanrooms-result-receiver-<timestamp>`
- Debe tener permisos para crear roles y adjuntar políticas.

Use an existing service role

1. Seleccione un Nombre de rol de servicio existente en la lista desplegable.

Si tiene permisos de listas de roles, se mostrará la lista de roles.

Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar.

2. Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM.

Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible.

De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

 Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).
- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes para ello AWS Clean Rooms. Debe agregar la política de rol antes de continuar.
- Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para el rol de servicio.

- b. Para los resultados de los trabajos, seleccione la casilla Establecer la configuración predeterminada para los trabajos y, a continuación, especifique el destino de los resultados en Amazon S3 introduciendo el destino de S3 o elija Browse S3 para seleccionarlos de una lista de depósitos de S3 disponibles.

Example

Por ejemplo: **s3://bucket/prefix**

- Especifique los permisos de acceso al servicio seleccionando el nombre de un rol de servicio existente en la lista desplegable.

10. Si desea habilitar la opción de Etiquetas para el recurso de pertenencia, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.

11. Si el creador de la colaboración lo ha designado como miembro responsable del pago de las consultas o de las consultas y los trabajos, indique su aceptación marcando la casilla Acepto pagar los costes de procesamiento de esta colaboración.

 Note

Debe seleccionar esta casilla de verificación para continuar.

Para obtener más información sobre cómo se calculan los precios, consulte [Precios para AWS Clean Rooms](#).

Si es el [miembro que paga los costes de cálculo](#) de las [consultas o el miembro que paga los costes de las consultas y el trabajo](#), pero no el [miembro que puede realizar](#) las consultas AWS Budgets , se recomienda configurar un presupuesto AWS Clean Rooms y recibir notificaciones cuando se alcance el presupuesto máximo. Para obtener más información sobre cómo configurar un presupuesto, consulte [Administración de costos con AWS Budgets](#) en la Guía del usuario de AWS Cost Management . Para obtener más información sobre cómo configurar notificaciones, consulte [Creación de un tema de Amazon SNS para las notificaciones de presupuesto](#) en la Guía del usuario de AWS Cost Management . Si se ha alcanzado el presupuesto máximo, puedes ponerte en contacto con el miembro que puede gestionar las consultas y los trabajos o [abandonar la colaboración](#). Si abandona la colaboración, no se podrán ejecutar más consultas y, por lo tanto, ya no se le facturarán costos de cómputo de consultas.

12. Si tiene la certeza de que desea crear una pertenencia y unirse a la colaboración, consulte Crear pertenencia.

Se le concede acceso de lectura a los metadatos de la colaboración. Esto incluye información como el nombre para mostrar y la descripción de la colaboración, además de todos los nombres y Cuenta de AWS IDs de los demás miembros.

Ya puede hacer lo siguiente:

- [Preparar su tabla de datos para que se consulte en AWS Clean Rooms](#) (Opcional si desea consultar sus propios datos de eventos o si desea consultar datos de identidad).
- [Asocie la tabla configurada a la colaboración](#): si desea consultar datos de eventos.
- [Agregue una regla de análisis para la tabla configurada](#): si desea consultar datos de eventos.

- [Cree y asocie un nuevo espacio de nombres de ID](#), si desea crear una tabla de mapeo de ID para consultar datos de identidad.

Para obtener información sobre cómo abandonar una colaboración, consulte [Abandonar una colaboración](#).

Editar colaboraciones

Como creador de colaboraciones, puedes editar las distintas partes de una colaboración.

Para obtener información sobre cómo editar una colaboración mediante AWS SDKs, consulte la [referencia de la API de salas limpias de AWS](#).

Temas

- [Edición del nombre y la descripción de la colaboración](#)
- [Actualice el motor de análisis de colaboración](#)
- [Desactiva el almacenamiento de registros](#)
- [Edita la configuración de los registros de colaboración](#)
- [Editar etiquetas de colaboración](#)
- [Edición de etiquetas de pertenencia](#)
- [Agrega un miembro nuevo](#)
- [Edita las habilidades de los miembros existentes](#)
- [Editar la configuración de aprobación automática de la colaboración](#)
- [Editar las etiquetas de tablas asociadas](#)
- [Edición de etiquetas de plantilla de análisis](#)
- [Edición de las etiquetas de la política de privacidad diferencial](#)

Edición del nombre y la descripción de la colaboración

Una vez creada una colaboración, solo su nombre y descripción serán editables.

Para editar el nombre y la descripción de la colaboración

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#).
2. En el panel de navegación izquierdo, elija Colaboraciones.

3. Elija la colaboración que ha creado.
4. En la página de detalles de la colaboración, elija Acciones y, a continuación, elija Editar colaboración.
5. En la página Editar colaboración, en Detalles, edite el nombre y la descripción de la colaboración.
6. Seleccione Save changes (Guardar cambios).

Actualice el motor de análisis de colaboración

Tras crear la colaboración, puedes cambiar el motor de análisis de AWS Clean Rooms SQL a Spark.

Note

Cambiar el motor de análisis de AWS Clean Rooms SQL a Spark podría interrumpir los flujos de trabajo existentes.

Para actualizar el motor de análisis de colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. En la página de detalles de la colaboración, elija Acciones y, a continuación, elija Editar colaboración.
5. Seleccione Save changes (Guardar cambios).

Desactiva el almacenamiento de registros

Si has activado el registro de análisis, puedes editar si los registros de análisis se almacenan en tu cuenta de Amazon CloudWatch Logs.

Para desactivar el almacenamiento de registros

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu Cuenta de AWS (si aún no lo has hecho).

2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elige la colaboración que tenga activado el registro de análisis.
4. En la página de detalles de la colaboración, selecciona Acciones y, a continuación, selecciona Desactivar el almacenamiento de registros.

 Note

Aparece una advertencia que indica lo siguiente:

- Las consultas nuevas ya no se registrarán en su CloudWatch cuenta.
- Los registros existentes se conservarán de acuerdo con su configuración de retención actual.
- Si reactiva el registro en el futuro, solo se aplicará a las consultas realizadas después de la reactivación.
- Este cambio solo afecta a tus registros; la configuración de registro de los demás miembros del equipo permanece inalterada.

5. Seleccione Desactivar.

Edita la configuración de los registros de colaboración

Si has activado el registro de consultas, puedes editar si los registros de consultas se almacenan en tu cuenta de Amazon CloudWatch Logs.

Para editar la configuración de los registros de colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. En la página de detalles de la colaboración, realice una de las siguientes acciones:
 - Elija Acciones y, a continuación, elija Editar la configuración de los registros.
 - En la pestaña Registros, selecciona Editar la configuración de los registros.
5. En el modal Editar configuración de registros, para el almacenamiento de CloudWatch registros en Amazon Logs:

- Si no quieres que los registros relevantes para ti se almacenen en tu cuenta de Amazon CloudWatch Logs, selecciona Activar.
- Si quieres que los registros relevantes para ti se almacenen en tu cuenta de Amazon CloudWatch Logs, selecciona Activar.

Solo puedes recibir los registros de las consultas que hayas iniciado o que contengan tus datos.

El miembro que puede recibir los resultados también recibe los registros de todas las consultas que se ejecutan en una colaboración, aunque no se accede a los datos en una consulta.

1. En Tipos de registro compatibles, elige entre los tipos de registro que el creador de la colaboración ha decidido admitir:
 - Si desea recibir los registros generados a partir de consultas SQL, seleccione la casilla Registros de consultas.
 - Si desea recibir los registros generados a partir de los trabajos que se utilizan PySpark, seleccione la casilla Registros de los trabajos.

6. Seleccione Save changes (Guardar cambios).

Note

Tras activar el registro, el almacenamiento de registros puede tardar unos minutos en configurarse y empezar a recibirlos en Amazon CloudWatch Logs. Durante este breve período, el miembro que puede realizar consultas puede ejecutar consultas que, en realidad, no enviarán registros.

Editar etiquetas de colaboración

Como creador de la colaboración, una vez creada una colaboración, puede administrar las etiquetas del recurso de colaboración.

Para editar las etiquetas de la colaboración

1. Inicia sesión en Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).

2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. Seleccione una de las siguientes opciones:

Si usted es...	Entonces...
El creador de la colaboración y un miembro de la colaboración	Elija la pestaña Detalles.
El creador de la colaboración, pero no un miembro de la colaboración	Desplácese hacia abajo por la página hasta la sección Etiquetas.

5. Para ver los Detalles de la colaboración, elija Administrar etiquetas.
6. En la página Administrar etiquetas, puede hacer lo siguiente:
 - Para eliminar una etiqueta, elija Eliminar.
 - Para añadir una etiqueta, elija Añadir nueva etiqueta.
 - Para guardar los cambios, elija Guardar cambios.

Edición de etiquetas de pertenencia

Como creador de colaboraciones, una vez que haya creado una colaboración, puede administrar las etiquetas del recurso de pertenencia.

Para editar las etiquetas de pertenencia

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. Elija la pestaña Detalles.
5. En Detalles de pertenencia, elija Administrar etiquetas.
6. En la página Administrar etiquetas de pertenencia, puede hacer lo siguiente:
 - Para eliminar una etiqueta, elija Eliminar.
 - Para añadir una etiqueta, elija Añadir nueva etiqueta.

- Para guardar los cambios, elija Guardar cambios.

Agrega un miembro nuevo

Para obtener más información, consulte [Agrega un nuevo miembro a una colaboración](#).

Edita las habilidades de los miembros existentes

Para obtener más información, consulte [Actualice las capacidades de los miembros existentes](#).

Editar la configuración de aprobación automática de la colaboración

Para obtener más información, consulte [Editar la configuración de aprobación automática de la colaboración](#).

Editar las etiquetas de tablas asociadas

Como creador de la colaboración, después de asociar tablas a una colaboración, puede administrar las etiquetas del recurso de tabla asociado.

Para editar las etiquetas de tabla asociadas

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. Seleccione la pestaña Tablas.
5. En Tablas asociadas por usted, elija una tabla.
6. En la página de detalles de la tabla configurada, en Etiquetas, elija Administrar etiquetas.

En la página Administrar etiquetas, puede hacer lo siguiente:

- Para eliminar una etiqueta, elija Eliminar.
- Para añadir una etiqueta, elija Añadir nueva etiqueta.
- Para guardar los cambios, elija Guardar cambios.

Edición de etiquetas de plantilla de análisis

Como creador de la colaboración, una vez creada una colaboración, puede administrar las etiquetas del recurso de plantilla de análisis.

Para editar las etiquetas de pertenencia

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con la tuya Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. Elija la pestaña Plantillas.
5. En la sección Plantillas de análisis creadas por usted, elija la plantilla de análisis.
6. En la página de detalles de la tabla de plantillas de análisis, desplácese hacia abajo hasta la sección Etiquetas.
7. Elija Administrar etiquetas.
8. En la página Administrar etiquetas, puede hacer lo siguiente:
 - Para eliminar una etiqueta, elija Eliminar.
 - Para añadir una etiqueta, elija Añadir nueva etiqueta.
 - Para guardar los cambios, elija Guardar cambios.

Edición de las etiquetas de la política de privacidad diferencial

Como creador de la colaboración, una vez creada una colaboración, puede administrar las etiquetas del recurso de plantilla de análisis.

Para editar las etiquetas de pertenencia

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con la tuya Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que contiene la política de privacidad diferencial que desea editar.
4. Seleccione la pestaña Tablas.
5. En la pestaña Etiquetas, elija Administrar etiquetas.

6. En la página Administrar etiquetas, puede hacer lo siguiente:

- Para eliminar una etiqueta, elija Eliminar.
- Para añadir una etiqueta, elija Añadir nueva etiqueta.
- Para guardar los cambios, elija Guardar cambios.

Solicitudes de cambio en AWS Clean Rooms

Las solicitudes de cambio te permiten proponer cambios en la configuración de colaboración existente para que los aprueben otros miembros de la colaboración. Con las solicitudes de cambio, puede enviar una solicitud para añadir nuevos miembros, actualizar las capacidades de los miembros existentes y modificar la configuración de aprobación automática de la colaboración. Todos los miembros de la colaboración deben aprobar las solicitudes de cambio para que los cambios propuestos surtan efecto.

Las solicitudes de cambio son específicas de la colaboración y el creador de la colaboración puede enviarlas.

Puede enviar una solicitud de cambio de las siguientes maneras:

- Añadir un nuevo miembro a una colaboración
- Actualiza las habilidades de los miembros existentes
- Editar la configuración de aprobación automática de la colaboración

Note

Debe ser el creador de la colaboración para enviar una solicitud de cambio.

Agrega un nuevo miembro a una colaboración

Para añadir un nuevo miembro a una colaboración, debes ser el creador de la colaboración. La adición de un nuevo miembro a una colaboración requiere la aprobación manual de los miembros de la colaboración existentes y, por lo tanto, se presentará una solicitud de cambio.

Para añadir nuevos miembros a una colaboración, sigue estos pasos:

Para añadir un nuevo miembro a una colaboración

1. Inicie sesión en la [AWS Clean Rooms consola de administración de AWS y ábrala](#) con su miembro Cuenta de AWS.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione tu colaboración para acceder a los detalles de tu colaboración.
4. En la página de colaboración, vaya a la pestaña Miembros.
5. En la tabla Miembros, seleccione Editar miembros.
6. Seleccione Añadir otro miembro.
7. Introduzca la información del nuevo miembro:
 - Nombre visible del miembro
 - Cuenta de AWSID de miembro
 - Especifique si el miembro puede recibir los resultados. Marque la casilla para conceder la capacidad a este miembro.
8. Seleccione Save changes (Guardar cambios).
9. Confirme el envío de la solicitud de cambio. En el modal de confirmación, confirma los cambios y selecciona Enviar solicitud de cambio.

Note

Si su colaboración admite los tipos de cambio aprobados automáticamente, es posible que no sea necesaria la aprobación manual de las solicitudes de cambio. Puede revisar qué tipos de cambios no requieren una solicitud de cambio en la sección «Descripción general» de la colaboración. Para obtener más información, consulte [the section called “Editar la configuración de aprobación automática de la colaboración”](#).

Actualice las capacidades de los miembros existentes

Para actualizar las habilidades de los miembros de la colaboración existentes, debes ser el creador de la colaboración. La actualización de las capacidades de los miembros de la colaboración existentes requiere la aprobación manual de los miembros existentes y presentarán una solicitud de cambio.

Las capacidades de los miembros que se pueden actualizar son:

- Puede recibir resultados

Puede actualizar las capacidades de los miembros de la colaboración existentes siguiendo estos pasos:

Para actualizar las capacidades de los miembros existentes

1. Inicia sesión en la [AWS Clean Roomsconsola Consola de administración de AWS](#) y ábrela con tu miembroCuenta de AWS.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Selecciona tu colaboración para acceder a los detalles de tu colaboración.
4. En la página de colaboración, vaya a la pestaña Miembros.
5. En la tabla Miembros, seleccione Editar miembros.
6. Especifique la capacidad de cambio del miembro.
7. Seleccione Save changes (Guardar cambios).
8. Confirme el envío de la solicitud de cambio. En el modal de confirmación, confirma los cambios y selecciona Enviar solicitud de cambio.

Note

Si su colaboración admite los tipos de cambio aprobados automáticamente, es posible que no sea necesaria la aprobación manual de las solicitudes de cambio. Puede revisar qué tipos de cambios no requieren una solicitud de cambio en la sección «Descripción general» de la colaboración. Para obtener más información, consulte [the section called “Editar la configuración de aprobación automática de la colaboración”](#).

Editar la configuración de aprobación automática de la colaboración

Para editar la configuración de la colaboración para las aprobaciones automáticas, debe ser el creador de la colaboración y enviar una solicitud de cambio para que los demás miembros de la colaboración la aprueben.

Puede editar la configuración de aprobación automática de una colaboración siguiendo estos pasos:

Para editar la configuración de aprobación automática de la colaboración

1. Inicia sesión en la [AWS Clean Rooms consola de administración de AWS](#) y ábrela con tu miembro Cuenta de AWS.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Selecciona tu colaboración para acceder a los detalles de tu colaboración.
4. En la página de colaboración, pulse el botón Acciones y seleccione Editar aprobaciones automáticas.
5. Para conceder capacidades de miembros a los miembros existentes sin aprobar manualmente una solicitud de cambio:
 1. Ve a la sección Otorgar habilidades a los miembros.
 2. Especifica qué habilidades de los miembros se pueden conceder automáticamente.

 Note

De forma predeterminada, todos los miembros de la colaboración pueden «aportar datos».

3. (Opcional) Para permitir que los nuevos miembros se unan al instante a una colaboración con las capacidades especificadas sin necesidad de aprobar manualmente las solicitudes de cambio, selecciona Aprobar automáticamente los nuevos miembros con estas habilidades.
6. Para permitir la eliminación de habilidades de los miembros existentes sin necesidad de solicitar un cambio:
 1. Ve a la sección Habilidades que se pueden revocar automáticamente.
 2. Especifica qué habilidades de los miembros se pueden eliminar automáticamente.
7. Seleccione Save changes (Guardar cambios).
8. Confirme el envío de la solicitud de cambio. En el modal de confirmación, confirma los cambios y selecciona Enviar solicitud de cambio.

Eliminar colaboraciones

Como creador de la colaboración, puede eliminar una colaboración que haya creado.

 Note

Al eliminar una colaboración, ni usted ni el resto de miembros pueden realizar consultas, recibir resultados o aportar datos. Cada miembro de la colaboración sigue teniendo acceso a sus propios datos como parte de su pertenencia.

Para eliminar una colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que desea eliminar.
4. En Acciones, selecciona Eliminar colaboración.
5. Confirme la eliminación y luego elija Eliminar.

Ver colaboraciones

Como creador de la colaboración, puede ver todas las colaboraciones que ha creado.

Para ver colaboraciones

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página Colaboraciones, en Usadas por última vez, consulte las últimas 5 colaboraciones utilizadas.
4. En la pestaña Con pertenencia activa, consulte la lista de colaboraciones con pertenencia activa.

Puede ordenar por Nombre, Fecha de creación de la pertenencia y Sus detalles de miembro.

Puede utilizar la barra de búsqueda para buscar una colaboración.

5. En la pestaña Disponibles para unirse, consulte la lista de Colaboraciones disponibles para unirse.
6. En la pestaña Ya no están disponibles, consulte la lista de colaboraciones eliminadas y Pertenencias a colaboraciones que ya no están disponibles (pertenencias eliminadas).

Invitar a los miembros a una colaboración

Como creador de una colaboración, una vez creada una colaboración, puede enviar un enlace de invitación a los miembros que aparecen en la pestaña Miembros.

Para invitar a miembros a una colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. Seleccione la pestaña Miembros.
5. En la tabla Miembros, pulse el botón Copiar el enlace de invitación.

Se copia el enlace de la invitación.

6. Pegue el enlace de la invitación en el método de comunicación seguro que prefiera y envíelo a cada miembro de la colaboración.

Supervisión de miembros

Como creador de una colaboración, una vez creada la colaboración, puede monitorizar el estado de todos los miembros en la pestaña Miembros. Esto ayuda a garantizar un control de acceso y una gestión de costes adecuados en toda su colaboración.

Para supervisar el estado de un miembro

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración que ha creado.
4. Seleccione la pestaña Miembros.
5. En la tabla de miembros, revise el estado de cada miembro.
6. En la tabla de capacidades de los miembros, revise qué miembros pueden realizar consultas, recibir resultados, aportar datos y realizar otras tareas.
7. En la tabla de configuración de pagos, revise qué miembros pagan por las consultas, las tablas de asignación de ID y los modelos de ML.

Añadir miembros a una colaboración

Requisitos previos

- Y Cuenta de AWS con permisos para gestionar las colaboraciones
- El Cuenta de AWS IDs número de miembros que quieres añadir

Para añadir un miembro a una colaboración

1. Inicie sesión en la [AWS Clean Roomsconsola Consola de administración de AWS y ábrala](#).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración a la que quieres añadir miembros.
4. Seleccione la pestaña Miembros.
5. Elija Edit members (Editar miembros).
6. Seleccione Añadir otro miembro e introduce la siguiente información:
 - Nombre visible del miembro
 - Cuenta de AWSID de miembro
 - Especifique si pueden recibir los resultados
7. Seleccione Save changes (Guardar cambios).
8. En el modal de confirmación, compruebe que la información sea correcta y elija Enviar solicitud de cambio.
 - Si la solicitud de cambio requiere la aprobación de otros miembros, todos los miembros existentes deben aprobar la solicitud de cambio antes de añadir al nuevo miembro. Para obtener más información sobre las solicitudes de cambio, consulte [the section called “Solicitudes de cambio”](#).
 - Si se admiten los tipos de cambio de aprobación automática para los nuevos miembros con las habilidades especificadas, este cambio entrará en vigor de inmediato. Para obtener más información, consulte [the section called “Editar la configuración de aprobación automática de la colaboración”](#).
9. En la página de detalles de la colaboración, en la pestaña Miembros, compruebe que el estado de miembro de los miembros añadidos aparezca como Invitado.

Tras completar estos pasos, los miembros invitados pueden unirse a la colaboración. Para obtener más información sobre cómo unirse a una colaboración, consulte [Crear una pertenencia y unirse a una colaboración](#)

Eliminar miembros de una colaboración

Note

Antes de empezar, ten en cuenta que eliminar a un miembro:

- Elimina todos sus conjuntos de datos asociados de la colaboración
- Si el [miembro paga los costes de cálculo de las consultas](#), esta acción detiene la ejecución de todas las consultas en la colaboración.

Al eliminar a un miembro, también se eliminan todos sus conjuntos de datos asociados de la colaboración.

Requisitos previos

- Debe ser un creador de colaboraciones
- No puedes eliminar tu propia cuenta

Para eliminar un miembro de una colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración que desee modificar.
4. Seleccione la pestaña Miembros.
5. Selecciona el botón de opción situado junto al miembro que quieres eliminar.
6. Elija Eliminar.
7. En el cuadro de diálogo de confirmación, escriba **confirm** para comprobar la eliminación.

 Note

Después de eliminar a un miembro, todos los conjuntos de datos asociados a su cuenta también se eliminan de la colaboración.

 Important

Si eliminas al miembro que paga los costes de cálculo de las consultas, no se podrán realizar más consultas en la colaboración hasta que designes un nuevo miembro pagador.

Abandonar una colaboración

Como miembro de una colaboración, puede abandonar una colaboración eliminando tu pertenencia. Si es el creador de la colaboración, solo puede abandonarla si [elimina la colaboración](#).

 Note

Al eliminar su pertenencia, abandonará la colaboración y no podrá volver a unirse a ella. Si es el [miembro que paga los costos de computación de consultas](#) y elimina su pertenencia, no se permite que se ejecuten más consultas.

Para abandonar una colaboración

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En Con pertenencia activa, elija la colaboración de la que es miembro.
4. Elija Acciones.
5. Seleccione Eliminar pertenencia.
6. En el cuadro de diálogo, confirme la decisión de abandonar la colaboración escribiendo **confirm** en el campo de entrada de texto y, a continuación, seleccione Vaciar y eliminar la membresía.

En la consola, verá un mensaje que indica que se ha eliminado la pertenencia.

El creador de la colaboración ve el Estado de miembro como Abandonado.

Prepare tablas de datos en AWS Clean Rooms

Note

Puede preparar las tablas de datos antes o después de unirse a una colaboración. Una vez preparada una tabla, puede reutilizarla en varias colaboraciones siempre que sus necesidades de privacidad para esa tabla sean las mismas.

Como miembro de la colaboración, debe preparar las tablas de datos antes de que el miembro de la colaboración que pueda consultarlas pueda consultarlas. AWS Clean Rooms

Las tablas de datos que se utilizan para las consultas AWS Clean Rooms suelen ser del mismo tipo de tablas de datos que se utilizan para otras aplicaciones. Por ejemplo, los mismos tipos de conjuntos de datos se utilizan con Amazon Athena, Amazon EMR, Amazon Redshift Spectrum y Amazon Quick.

Puede consultar los datos en su formato original directamente desde cualquiera de las siguientes fuentes de datos:

- Amazon Simple Storage Service (Amazon S3)
- Amazon Athena
- Snowflake

AWS Clean Rooms accede al conjunto de datos en tiempo de ejecución de la consulta, lo que garantiza que los miembros que pueden realizar consultas accedan siempre a la mayor cantidad de up-to-date datos. Todos los datos que se lean temporalmente en una AWS Clean Rooms colaboración se eliminan una vez finalizada la consulta. Los resultados de la consulta se escriben en su bucket de Amazon S3.

Si su caso de uso implica la consulta de datos de identidad, consulte [AWS Entity Resolution in AWS Clean Rooms](#).

Temas

- [Formatos de datos para AWS Clean Rooms](#)
- [Apache Iceberg tablas en AWS Clean Rooms](#)

- [Preparación de tablas de datos para consultas en AWS Clean Rooms](#)
- [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#)
- [Descifrar tablas de datos con el cliente de cifrado de C3R](#)

Formatos de datos para AWS Clean Rooms

Para analizar los datos, los conjuntos de datos deben estar en un formato AWS Clean Rooms compatible.

Temas

- [Formatos de datos compatibles para los trabajos PySpark](#)
- [Formatos de datos compatibles para consultas SQL](#)
- [Tipos de datos compatibles](#)
- [Tipos de compresión de archivos para AWS Clean Rooms](#)
- [Cifrado del lado del servidor para AWS Clean Rooms](#)

Formatos de datos compatibles para los trabajos PySpark

AWS Clean Rooms admite los siguientes formatos estructurados para ejecutar PySpark trabajos.

- Parquet
- OpenCSV
- JSON

Formatos de datos compatibles para consultas SQL

AWS Clean Rooms admite los siguientes formatos estructurados para ejecutar consultas SQL.

- [Tablas de Apache Iceberg](#)
- Parquet
- OpenCSV
- JSON

 Note

Un valor `timestamp` de un archivo de texto debe estar en formato `yyyy-MM-dd HH:mm:ss.SSSSSS`. Por ejemplo: `2017-05-01 11:30:59.000000`.

Recomendamos utilizar un formato de archivo de almacenamiento en columnas, como Apache Parquet. Con un formato de archivo de almacenamiento en columnas, puede minimizar el movimiento de datos seleccionando solo las columnas que necesite. Para obtener un rendimiento óptimo, los objetos grandes deben dividirse en objetos de 100 MB a 1 GB.

Tipos de datos compatibles

AWS Clean Rooms admite los siguientes tipos de datos.

- ARRAY
- BIGINT
- BOOLEANO
- BYTE
- CHAR
- DATE
- DECIMAL
- FLOAT
- INTEGER
- INTERVAL
- LONG
- MAP
- REAL
- SHORT
- SMALLINT
- STRUCT
- TIME
- TIMESTAMP_LTZ
- TIMESTAMP_NTZ

- TINYINT
- VARCHAR

Para obtener más información, consulte [Tipos de datos](#) en la referencia de SQL.AWS Clean Rooms

Tipos de compresión de archivos para AWS Clean Rooms

Para reducir el espacio de almacenamiento, mejorar el rendimiento y minimizar costos, es muy recomendable comprimir los archivos de datos.

AWS Clean Rooms reconoce los tipos de compresión de archivos en función de la extensión del archivo y admite los tipos y extensiones de compresión que se muestran en la tabla siguiente.

Algoritmo de compresión	Extensión de archivo
GZIP	.gz
Bzip2	.bz2
Snappy	.snappy

Puede aplicar compresión a distintos niveles. En la mayoría de los casos, se comprime un archivo completo o se comprimen bloques individuales dentro de un archivo. La compresión de formatos de columna por archivo no ofrece beneficios en términos de rendimiento.

Cifrado del lado del servidor para AWS Clean Rooms

Note

El cifrado en el servidor no reemplaza a la computación criptográfica en los casos de uso en los que esta última es obligatoria.

AWS Clean Rooms descifra de forma transparente los conjuntos de datos cifrados mediante las siguientes opciones de cifrado:

- SSE-S3: cifrado en el servidor mediante una clave de cifrado AES-256 administrada por Amazon S3

- SSE-KMS: cifrado del lado del servidor con claves administradas por AWS Key Management Service

Para usar SSE-S3, el rol de AWS Clean Rooms servicio utilizado para asociar la tabla configurada a la colaboración debe tener permisos de descifrado por KMS. Para usar SSE-KMS, la política de claves de KMS también debe permitir que la función de servicio descifre. AWS Clean Rooms

AWS Clean Rooms no admite el cifrado del lado del cliente de Amazon S3. Para obtener más información sobre el cifrado en el servidor, consulte [Protección de datos con el cifrado del lado del servidor](#) en la Guía del usuario de Amazon Simple Storage Service.

Apache Iceberg tablas en AWS Clean Rooms

Apache Iceberg es un formato de tabla de código abierto para lagos de datos. AWS Clean Rooms puede utilizar las estadísticas almacenadas en Apache Iceberg metadatos para optimizar los planes de consultas y reducir el escaneo de archivos durante el procesamiento de consultas en sala limpia. Para obtener más información, consulte la documentación de [Apache Iceberg](#).

Tenga en cuenta lo siguiente cuando las utilice AWS Clean Rooms con tablas Iceberg:

- Tablas Apache Iceberg para S3: Apache Iceberg las tablas deben definirse en AWS Glue Data Catalog función de la [implementación del catálogo Glue de código abierto](#).
- Tablas Apache Iceberg para Athena: [para obtener más información, consulte -iceberg.html https://docs.aws.amazon.com/athena/latest/ug/querying](#)
- [Tablas Apache Iceberg para Snowflake: para obtener más información, consulte user-guide/tables-iceberg https://docs.snowflake.com/en/](#)
- Formato de archivo Parquet: AWS Clean Rooms solo admite tablas Iceberg en el formato de archivo de datos Parquet.
- Compresión GZIP y Snappy: AWS Clean Rooms admite Parquet con GZIP y Snappy compresión.
- Versiones Iceberg: AWS Clean Rooms permite ejecutar consultas en tablas Iceberg de las versiones 1 y 2.
- Particiones: no es necesario añadir particiones manualmente a su Apache Iceberg tablas en AWS Glue. AWS Clean Rooms detecta nuevas particiones en Apache Iceberg tablas automáticamente y no es necesaria ninguna operación manual para actualizar las particiones en la definición de la tabla. Las particiones de Iceberg aparecen como columnas normales en el esquema de tabla de

AWS Clean Rooms , y no por separado como una clave de partición en el esquema de la tabla configurada.

- Limitaciones

- Solo tablas de Iceberg nuevas

Apache Iceberg tablas convertidas desde Apache Parquet no se admiten tablas.

- Consultas de viaje en el tiempo

AWS Clean Rooms no admite consultas sobre viajes en el tiempo con Apache Iceberg tablas.

- Versión 2 del motor Athena

Iceberg no se admiten las tablas creadas con la versión 2 del motor Athena.

- Formatos de archivo

Avro y no se admiten los formatos de archivo optimizados con filas y columnas (ORC).

- Compresión

Zstandard Compresión (Zstd) para Parquet no se admite.

Tipos de datos admitidos para las tablas de Iceberg

AWS Clean Rooms puede consultar Iceberg tablas que contienen los siguientes tipos de datos:

- BOOLEAN
- DATE
- DECIMAL
- DOUBLE
- FLOAT
- INT
- LIST
- LONG
- MAP
- STRING
- STRUCT
- TIMESTAMP WITHOUT TIME ZONE

Para obtener más información sobre los tipos de datos de Iceberg, consulte los [esquemas para Iceberg](#) en la documentación de Apache.

Preparación de tablas de datos para consultas en AWS Clean Rooms

Si no es necesario que traiga sus propios datos, puede omitir este procedimiento.

Si su caso de uso implica la consulta de datos de identidad, consulte [AWS Entity Resolution in AWS Clean Rooms](#).

Para obtener más información sobre los formatos de datos que puede utilizar, consulte [Formatos de datos para AWS Clean Rooms](#).

Temas

- [Preparación de tablas de datos en Amazon S3](#)
- [Preparación de tablas de datos en Amazon Athena](#)
- [Preparación de tablas de datos en Snowflake](#)

Preparación de tablas de datos en Amazon S3

Puede analizar las tablas de datos que se han catalogado AWS Glue y almacenado en Amazon S3. Si sus tablas de datos ya están catalogadas AWS Glue, vaya a [Crear una tabla configurada en AWS Clean Rooms](#)

Important

Al preparar los datos en Amazon S3 para usarlos con ellos AWS Clean Rooms, debe asegurarse de que la ubicación de la AWS Glue tabla coincida exactamente con la ruta del directorio en el que se almacenan los archivos de datos.

Por ejemplo: si sus datos están almacenados en `s3://mybucket/folder/subfolder/data.parquet`, la ubicación de AWS Glue la tabla debe apuntar a `s3://mybucket/folder/subfolder/`. Si se establece la ubicación de la tabla en un directorio principal (`s3://mybucket/folder/``), la tabla aparecerá vacía cuando se consulte.

La preparación de las tablas de datos en Amazon S3 implica los siguientes pasos:

Temas

- [Paso 1: completar los requisitos previos](#)
- [Paso 2: \(opcional\) preparar sus datos para la computación criptográfica](#)
- [Paso 3: cargar la tabla de datos en Amazon S3](#)
- [Paso 4: Crear una AWS Glue tabla](#)
- [Paso 5: Siguiendo pasos](#)

Paso 1: completar los requisitos previos

Para preparar las tablas de datos para usarlas con AWS Clean Rooms ellas, debe cumplir los siguientes requisitos previos:

- Las tablas de datos se guardan como uno de los [formatos de datos compatibles para AWS Clean Rooms](#).
- Las tablas de datos están catalogadas AWS Glue y utilizan los [tipos de datos compatibles](#). AWS Clean Rooms
- Todas las tablas de datos se almacenan en Amazon Simple Storage Service (Amazon S3), en el Región de AWS mismo lugar en el que se creó la colaboración.
- AWS Glue Data Catalog Debe estar en la misma región que la colaboración.
- AWS Glue Data Catalog Está en la misma forma Cuenta de AWS que la membresía.
- El bucket de Amazon S3 no está registrado en AWS Lake Formation.

Paso 2: (opcional) preparar sus datos para la computación criptográfica

(Opcional) Si utiliza la computación criptográfica y su tabla de datos contiene información confidencial que desea cifrar, debe cifrar la tabla de datos mediante el cliente de cifrado de C3R.

Para preparar los datos para la computación criptográfica, siga los procedimientos descritos en [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#).

Paso 3: cargar la tabla de datos en Amazon S3

Note

Si piensa utilizar tablas de datos cifrados en la colaboración, primero debe cifrar los datos para la computación criptográfica antes de cargar la tabla de datos en Amazon S3. Para

obtener más información, consulte [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#).

Para cargar la tabla de datos en Amazon S3

1. Inicie sesión en la consola de Amazon S3 Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/s3/>.
2. Elija Buckets y, a continuación, elija un bucket donde desee almacenar la tabla de datos.
3. Elija Cargar y siga las indicaciones de la pantalla.
4. Seleccione la pestaña Objetos para ver el prefijo donde se almacenan sus datos. Anote el nombre de la carpeta.

Puede seleccionar la carpeta para ver los datos.

Paso 4: Crear una AWS Glue tabla

Si ya tiene una tabla AWS Glue de datos, puede omitir este paso.

En este paso, configuras un rastreador AWS Glue que rastrea todos los archivos del bucket de S3 y crea una AWS Glue tabla. Para obtener más información, consulte [Definición de rastreadores en AWS Glue](#) en la Guía para desarrolladores de AWS Glue .

Para obtener más información sobre AWS Glue Data Catalog los tipos de datos compatibles, consulte. [Tipos de datos compatibles](#)

Note

AWS Clean Rooms actualmente no es compatible con los buckets S3 registrados en. AWS Lake Formation

El siguiente procedimiento describe cómo crear una AWS Glue tabla. Si desea utilizar un AWS Glue Data Catalog objeto cifrado con una clave AWS Key Management Service (AWS KMS), debe configurar la política de permisos de claves de KMS para permitir el acceso a esa tabla cifrada. Para obtener más información, consulte [Configuración del cifrado en AWS Glue](#), en la Guía del desarrollador de AWS Glue .

Para crear una AWS Glue tabla

1. Siga el procedimiento de [trabajo con rastreadores en la AWS Glue consola](#) de la Guía del AWS Glue usuario.
2. Anote el nombre de la AWS Glue base de datos y el nombre de AWS Glue la tabla.

Paso 5: Siguiendo pasos

Ahora que ha preparado las tablas de datos en Amazon S3, está preparado para:

- [Creación de una tabla configurada](#)
- [Creación de un modelo de ML](#)

Las tablas se pueden consultar después de:

- El creador de la colaboración ha configurado una colaboración en AWS Clean Rooms. Para obtener más información, consulte [Crear una colaboración](#).
- El creador de la colaboración le ha enviado el ID de la colaboración como participante en la colaboración.

Preparación de tablas de datos en Amazon Athena

Puede consultar tablas de datos que se hayan creado como vistas AWS Glue Data Catalog (GDC) en Amazon Athena.

Una vista de GDC es una tabla virtual que se crea a partir de una o más tablas subyacentes. AWS Glue Debe crearse con Athena SQL en el catálogo de Athena. `AwsGlueCatalog`

La preparación de las tablas de datos en Amazon Athena implica los siguientes pasos:

Temas

- [Paso 1: completar los requisitos previos](#)
- [Paso 2: \(opcional\) preparar sus datos para la computación criptográfica](#)
- [Paso 3: Próximos pasos](#)

Paso 1: completar los requisitos previos

Para preparar las tablas de datos para usarlas con AWS Clean Rooms ellas, debe cumplir los siguientes requisitos previos:

- Las tablas de datos se guardan como uno de los [formatos de datos compatibles para AWS Clean Rooms](#).
- Las tablas de datos utilizan los [tipos de datos compatibles para AWS Clean Rooms](#).
- Ha creado una vista GDC en su AWS Glue tabla con Athena SQL en el catálogo de AwsDataCatalog Athena.

La vista aparecerá en:

- La consola Athena (debajo deAwsDataCatalog) como vista: <https://console.aws.amazon.com/athena/>
- La AWS Glue consola como AWS Glue mesa: <https://console.aws.amazon.com/glue/>

Para obtener más información, consulte [Uso de vistas del catálogo de datos en Athena](#) en la Guía del usuario de Amazon Athena.

Note

Necesita los permisos adecuados para crear vistas en Athena y. AWS Glue Además, asegúrese de tener acceso a las tablas subyacentes a las que se hace referencia en su View definition.

AWS Clean Rooms solo admite el tipo de AWS Glue catálogo para Athena, no los tipos de catálogo Lambda o Hive.

- Sus tablas de datos o vistas de GDC están catalogadas y registradas en AWS Glue ellas. AWS Lake Formation
- Ha creado un depósito de salida independiente en Amazon S3 para recibir los resultados de Athena.
- Ha configurado un rol de servicio para leer los datos de Amazon Athena. Para obtener más información, consulte [Cree un rol de servicio para leer los datos de Amazon Athena](#).
- El rol de servicio tiene los permisos de acceso Lake Formation Select and Describe en la vista o tabla del GDC.

Paso 2: (opcional) preparar sus datos para la computación criptográfica

(Opcional) Si utiliza la computación criptográfica y su tabla de datos contiene información confidencial que desea cifrar, debe cifrar la tabla de datos mediante el cliente de cifrado de C3R.

Para preparar los datos para la computación criptográfica, siga los procedimientos descritos en [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#).

Paso 3: Próximos pasos

Ahora que ha preparado las tablas de datos en Amazon Athena, está preparado para:

- [Creación de una tabla configurada](#)
- [Creación de un modelo de ML](#)

Las tablas se pueden consultar después de:

- El creador de la colaboración ha configurado una colaboración en AWS Clean Rooms. Para obtener más información, consulte [Crear una colaboración](#).
- El creador de la colaboración le ha enviado el ID de la colaboración como participante en la colaboración.

Preparación de tablas de datos en Snowflake

Puede consultar las tablas de datos que se han almacenado en el almacén de datos de Snowflake.

La preparación de las tablas de datos en Snowflake implica los siguientes pasos:

Temas

- [Paso 1: completar los requisitos previos](#)
- [Paso 2: \(opcional\) preparar sus datos para la computación criptográfica](#)
- [Paso 3: Crea un secreto AWS Secrets Manager](#)
- [Paso 4: Siguiendo los pasos](#)

Paso 1: completar los requisitos previos

Para preparar las tablas de datos para usarlas con ellas AWS Clean Rooms, debe cumplir los siguientes requisitos previos:

- Dispone de Cuenta de AWS los permisos adecuados para leer las tablas de datos. Para obtener más información, consulte [Cree un rol de servicio para leer los datos de Snowflake](#).
- Las tablas de datos se guardan en uno de los [formatos de datos admitidos para AWS Clean Rooms](#).
- Las tablas de datos utilizan los [tipos de datos compatibles para AWS Clean Rooms](#).
- La tabla de datos se almacena en un almacén de Snowflake. Para obtener más información, consulte la documentación de [Snowflake](#).
- Ha configurado un nuevo usuario de Snowflake con privilegios de solo lectura en la tabla de Snowflake que va a asociar a su colaboración.

Paso 2: (opcional) preparar sus datos para la computación criptográfica

(Opcional) Si utiliza la computación criptográfica y su tabla de datos contiene información confidencial que desea cifrar, debe cifrar la tabla de datos mediante el cliente de cifrado de C3R.

Para preparar los datos para la computación criptográfica, siga los procedimientos descritos en [Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms](#).

Paso 3: Crea un secreto AWS Secrets Manager

Para conectarte a Snowflake desde AWS Clean Rooms, tendrás que crear y almacenar tus credenciales de Snowflake en un AWS Secrets Manager secreto y, a continuación, asociarlo a una tabla de Snowflake. AWS Clean Rooms

Note

Le recomendamos que cree un usuario nuevo que sea exclusivo para. AWS Clean Rooms. Ese usuario solo debe tener un rol con permisos de lectura para los datos AWS Clean Rooms a los que desee acceder.

Para crear un AWS Secrets Manager secreto

1. En Snowflake, genere un usuario `snowflakeUser` y configure la autenticación por pares de claves.

 Note

En noviembre de 2025, Snowflake pasará a admitir únicamente la autenticación por pares de claves. Este cambio afectará a la AWS Clean Rooms integración actual con Snowflake, que utiliza la autenticación con nombre de usuario y contraseña. Después de esta fecha, las conexiones de Snowflake AWS Clean Rooms requerirán una autenticación por pares de claves mediante una clave privada de Snowflake Privacy Enhanced Mail (PEM).

2. Determine con qué almacén de Snowflake interactuará este usuario, `snowflakeWarehouse`. Configúrelo como el formulario `DEFAULT_WAREHOUSE` `snowflakeUser` en Snowflake o recuérdelo para el siguiente paso.
3. En [AWS Secrets Manager](#), cree un secreto con sus credenciales de Snowflake. Para crear un secreto en Secrets Manager, sigue el tutorial disponible en [Crear un AWS Secrets Manager secreto](#) en la Guía del AWS Secrets Manager usuario. Después de crear el secreto, conserva el nombre secreto `secretName` para el siguiente paso.
 - Al seleccionar pares clave/valor, crea un par `snowflakeUser` con la clave. `sfUser`
 - Al seleccionar pares clave/valor, cree un par para su clave privada PEM de Snowflake con la clave. `pem_private_key`
 - Al seleccionar pares clave/valor, cree un par con la clave. `snowflakeWarehouse sfWarehouse`

Esto no es necesario si hay un valor predeterminado en Snowflake.

 - Al seleccionar pares clave/valor, cree un par con la clave. `snowflakeRole sfRole`

Paso 4: Sigüientes pasos

Ahora que ha preparado las tablas de datos en Snowflake, está preparado para:

- [Creación de una tabla configurada](#)
- [Creación de un modelo de ML](#)

Las tablas se pueden consultar después de:

- El creador de la colaboración ha configurado una colaboración en AWS Clean Rooms. Para obtener más información, consulte [Crear una colaboración](#).
- El creador de la colaboración le ha enviado el ID de la colaboración como participante en la colaboración.

Preparar tablas de datos cifrados con computación criptográfica para Clean Rooms

La computación criptográfica para Clean Rooms (C3R) es una capacidad de. AWS Clean Rooms. Puede utilizar el C3R para limitar criptográficamente lo que puede aprender cualquiera de las partes y en una colaboración. AWS AWS Clean Rooms

Puede cifrar la tabla de datos mediante el cliente de cifrado C3R, una herramienta de cifrado del lado del cliente, antes de cargar la tabla de datos en su fuente de datos: Amazon Simple Storage Service (Amazon S3), Amazon Athena o Snowflake.

Para obtener más información, consulte [Computación criptográfica para Clean Rooms](#).

Para preparar tablas de datos cifrados con C3R, se deben seguir estos pasos:

Steps

- [Paso 1: completar los requisitos previos](#)
- [Paso 2: descargar el cliente de cifrado de C3R](#)
- [Paso 3: \(Opcional\) Visualización de los comandos disponibles en el cliente de cifrado de C3R](#)
- [Paso 4: generar un esquema de cifrado para un archivo tabular](#)
- [Paso 5: crear una clave secreta compartida](#)
- [Paso 6: guardar la clave secreta compartida en la variable de entorno](#)
- [Paso 3: cifrar los datos](#)
- [Paso 8: verificar el cifrado de datos](#)
- [\(Opcional\) Crear un esquema \(usuarios avanzados\)](#)

Paso 1: completar los requisitos previos

Para preparar las tablas de datos para su uso con C3R, debe cumplir con los siguientes requisitos previos:

- Puede acceder al repositorio Cryptographic Computing for en: Clean Rooms GitHub

<https://github.com/aws/c3r>

- Ha configurado AWS las credenciales para usar el cliente de cifrado C3R. El cliente de cifrado C3R utiliza estas credenciales para las llamadas a la API de solo lectura con el fin de recuperar los metadatos de la colaboración. AWS Clean Rooms Para obtener más información, consulte [Configurar la AWS CLI](#) en la Guía del usuario de la versión 2 de AWS Command Line Interface .
- Tiene Java Runtime Environment (JRE) 11 o una versión posterior instalada en su equipo.
 - [Lo recomendado Java Runtime Environment, Amazon Corretto 11 o superior, se puede descargar de https://aws.amazon.com /corretto.](#)
 - El Java Development Kit (JDK) incluye una versión correspondiente de JRE de la misma versión. Sin embargo, las capacidades adicionales del JDK no son necesarias para ejecutar la computación criptográfica para el cliente de cifrado de Clean Rooms (C3R).
- Sus archivos de datos tabulares (.csv) o archivos Parquet (.parquet) se guardan localmente.
- Usted u otro miembro de la colaboración tienen la capacidad de crear una clave secreta compartida. Para obtener más información, consulte [Paso 5: crear una clave secreta compartida.](#)
- El creador de la colaboración ha creado una colaboración AWS Clean Rooms con la computación criptográfica habilitada para la colaboración. Para obtener más información, consulte [Crear una colaboración.](#)
- El creador de la colaboración le ha enviado el ID de la colaboración como participante en la colaboración. El nombre de recurso de Amazon (ARN) de la colaboración se incluye en la invitación enviada, que contiene el ID de la colaboración.

Paso 2: descargar el cliente de cifrado de C3R

Para descargar el cliente de cifrado C3R desde GitHub

1. [Vaya al repositorio de Computación criptográfica: c3r Clean RoomsAWSGitHub https://github.com/aws/](https://github.com/aws/c3r)
2. Seleccione y descargue los archivos.

El código fuente, las licencias y el material relacionado se pueden clonar o descargar como archivo .zip desde la página de inicio del repositorio de GitHub (vea el botón Código en la parte superior derecha de la lista de contenido del repositorio).

El último cliente de cifrado de C3R con firma Java Executable File (es decir, la aplicación de interfaz de la línea de comandos) se encuentra en la página Versiones del repositorio de GitHub.

El paquete del cliente de cifrado de C3R para Apache Spark (`c3r-cli-spark`) es una versión del `c3r-cli` que debe enviarse como trabajo a un servidor Apache Spark en ejecución. Para obtener más información, consulte [Ejecutar C3R en Apache Spark](#).

Paso 3: (Opcional) Visualización de los comandos disponibles en el cliente de cifrado de C3R

Utilice este procedimiento para familiarizarse con los comandos disponibles en el cliente de cifrado de C3R.

Para ver los comandos disponibles en el cliente de cifrado de C3R.

1. Desde una interfaz de la línea de comandos (CLI), vaya a la carpeta que contiene el archivo `c3r-cli.jar` descargado.
2. Ejecute el comando siguiente: `java -jar c3r-cli.jar`
3. Vea la lista de comandos y opciones disponibles.

Paso 4: generar un esquema de cifrado para un archivo tabular

Para cifrar datos, se requiere un esquema de cifrado que describa cómo se utilizarán los datos. En esta sección se describe cómo el cliente de cifrado de C3R ayuda a generar un esquema de cifrado para un archivo CSV con una fila de encabezado o un archivo Parquet.

Solo tiene que hacerlo una vez por archivo. Una vez creado el esquema, se puede volver a utilizar para cifrar el mismo archivo (o cualquier archivo con nombres de columna idénticos). Si los nombres de las columnas o el esquema de cifrado deseado cambian, debe actualizar el archivo de esquema. Para obtener más información, consulte [\(Opcional\) Crear un esquema \(usuarios avanzados\)](#).

Important

Es imprescindible que todas las partes colaboradoras utilicen la misma clave secreta compartida. Las partes colaboradoras también deben coordinar los nombres de las columnas para que coincidan si se va a ejecutar en ellas una operación JOIN o si se van a comparar

de cualquier otro modo para garantizar la igualdad en las consultas. De lo contrario, las consultas SQL podrían producir resultados inesperados o incorrectos. No obstante, esto no será necesario si el creador de la colaboración habilitó la configuración de cifrado `allowJoinsOnColumnsWithDifferentNames` al crear la colaboración. Para obtener más información acerca de los ajustes relevantes para el cifrado, consulte [Parámetros de computación criptográfica](#).

Cuando se ejecuta en modo de esquema, el cliente de cifrado de C3R revisa el archivo de entrada columna por columna y le pregunta si debe tratarse esa columna y de qué manera. Si el archivo contiene un gran número de columnas que no son necesarias para la salida cifrada, la generación del esquema interactivo podría resultar tediosa, ya que habría que saltarse todas las columnas no deseadas. Para evitarlo, puede escribir un esquema manualmente o crear una versión simplificada del archivo de entrada que incluya solo las columnas deseadas. Entonces, el generador de esquemas interactivo podrá ejecutarse en ese archivo reducido. El cliente de cifrado de C3R genera información sobre el archivo de esquema y le pregunta cómo se deben incluir o cifrar las columnas de origen (si es que se deben incluir) en el resultado de destino.

Para cada columna de origen del archivo de entrada, se le preguntará lo siguiente:

1. Cuántas columnas de destino se deben generar
2. Cómo se debe cifrar cada columna de destino (si procede)
3. El nombre de cada columna de destino
4. Cómo deben rellenarse los datos antes del cifrado si la columna se cifra como columna sealed

 Note

Al cifrar datos de una columna que se ha cifrado como columna sealed, debe determinar qué datos precisan relleno. El cliente de cifrado de C3R sugiere un relleno predeterminado durante la generación del esquema, consistente en rellenar todas las entradas de una columna hasta que alcancen la misma longitud.

Al determinar la longitud `fixed`, tenga en cuenta que el relleno está en bytes, no en bits.

A continuación se proporciona una tabla de decisión para crear el esquema.

Tabla de decisión del esquema

Decisión	¿Número de columnas de destino de la columna de origen <'>? name-of-column	Tipo de columna de destino: [c] cleartext, [f] fingerprint o [s] sealed?	Nombre del encabezado de la columna de destino <default 'name-of-column'>	Añadir un sufijo <sufijo> al encabezado o para indicar cómo se cifró, [y] sí o [n] no <predeterminado 'yes'>	<' name-of-column _sealed'> tipo de relleno: [n] uno, [f] fijo o [m] máximo <default 'max'>
Deje la columna sin cifrar.	1	c	No aplicable	No aplicable	No aplicable
Cifre la columna como columna fingerprint.	1	f	Elija el nombre predeterminado o introduzca un nombre de encabezado nuevo.	Introduzca y para elegir el valor predeterminado (_fingerprint) o introduzcan.	No aplicable
Cifre la columna como columna sealed.	1	s	Elija el nombre predeterminado o introduzca un nombre de encabezado nuevo.	Introduzca y para elegir el valor predeterminado (_sealed) o introduzcan.	Elija el tipo de relleno. Para obtener más información, consulte (Opcional) Crear un esquema (usuarios avanzados) .

Decisión	¿Número de columnas de destino de la columna de origen <'>? name-of-column	Tipo de columna de destino: [c] cleartext, [f] fingerprint o [s] sealed?	Nombre del encabezado de la columna de destino <default 'name-of-column'>	Añadir un sufijo <sufijo> al encabezado o para indicar cómo se cifró, [y] sí o [n] no <predeterminado 'yes'>	<' name-of-column _sealed'> tipo de relleno: [n] uno, [f] fijo o [m] máximo <default 'max'>
Cifre la columna como fingerprint y sealed.	2	Introduzca la primera columna de destino: f. Introduzca la segunda columna de destino: s.	Elija los encabezados de destino para cada columna de destino.	Introduzca y para elegir el valor predeterminado o introduzca n..	Elija el tipo de relleno (solo para columnas sealed). Para obtener más información, consulte (Opcional) Crear un esquema (usuarios avanzados) .

A continuación se muestran dos ejemplos de cómo crear esquemas de cifrado. El contenido exacto de la interacción depende del archivo de entrada y de las respuestas que proporcione.

Ejemplos

- [Ejemplo: Generar un esquema de cifrado para una columna fingerprint y una columna cleartext](#)
- [Ejemplo: Generar un esquema de cifrado con columnas sealed, fingerprint y columnas](#)

Ejemplo: Generar un esquema de cifrado para una columna fingerprint y una columna cleartext

En este ejemplo, para `ads.csv`, solo hay dos columnas: `username` y `ad_variant`. Para estas columnas, queremos lo siguiente:

- Que la columna `username` se cifre como columna `fingerprint`
- Que la columna `ad_variant` sea una columna `cleartext`

Para generar un esquema de cifrado para una columna `fingerprint` y una columna `cleartext`

1. (Opcional) Para garantizar que el archivo `c3r-cli.jar` y el archivo que se va a cifrar estén presentes:
 - a. Desplácese al directorio deseado y ejecute `ls` (si utiliza Mac o Unix/Linux) o `dir` si utiliza Windows).
 - b. Vea la lista de archivos de datos tabulares (por ejemplo, `.csv`) y elija el archivo que desea cifrar.

En este ejemplo, `ads.csv` es el archivo que queremos cifrar.

2. Desde la CLI, ejecute el siguiente comando para crear un esquema de forma interactiva.

```
java -jar c3r-cli.jar schema ads.csv --interactive --output=ads.json
```

Note

- Puede ejecutar `java --jar PATH/T0/c3r-cli.jar`. O bien, si ha añadido `PATH/T0/c3r-cli.jar` a su variable de entorno `CLASSPATH`, también puede ejecutar el nombre de la clase. El cliente de cifrado de C3R buscará en `CLASSPATH` para encontrarlo (por ejemplo, `java com.amazon.psion.cli.Main`).
- El indicador `--interactive` selecciona el modo interactivo para desarrollar el esquema. Esto guía al usuario a través de un asistente para crear el esquema. Los usuarios con conocimientos avanzados pueden crear su propio esquema JSON sin necesidad de utilizar el asistente. Para obtener más información, consulte [\(Opcional\) Crear un esquema \(usuarios avanzados\)](#).
- El indicador `--output` establece un nombre de salida. Si no incluye el indicador `--output`, el cliente de cifrado de C3R intentará elegir un nombre de

salida predeterminado (por ejemplo, `<input>.out.csv` o, para el esquema, `<input>.json`).

3. En **Number of target columns from source column 'username'?**, escriba **1** y luego presione Intro.
4. En **Target column type: [c]leartext, [f]ingerprint, or [s]ealed?**, escriba **f** y luego presione Intro.
5. En **Target column headername** `<default 'username'>`, presione Intro.

Se utiliza el nombre predeterminado 'username'.

6. En **Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o** `<default 'yes'>`, escriba **y** y luego presione Intro.

 Note

El modo interactivo sugiere añadir sufijos a los encabezados de las columnas cifradas (`_fingerprint` para las columnas fingerprint y `_sealed` para las columnas sealed). Los sufijos pueden resultar útiles cuando realizas tareas como cargar datos o crear colaboraciones. Servicios de AWS Clean Rooms Estos sufijos pueden ayudar a indicar qué se puede hacer con los datos cifrados de cada columna. Por ejemplo, la operación no funcionará si cifra una columna como columna sealed (`_sealed`) e intenta una operación JOIN en ella o a la inversa.

7. En **Number of target columns from source column 'ad_variant'?**, escriba **1** y luego presione Intro.
8. En **Target column type: [c]leartext, [f]ingerprint, or [s]ealed?**, escriba **c** y luego presione Intro.
9. En **Target column headername** `<default 'username'>`, presione Intro.

Se utiliza el nombre predeterminado 'ad_variant'.

El esquema se escribe en un nuevo archivo llamado `ads.json`.

 Note

Puede ver el esquema abriéndolo en cualquier editor de texto, por ejemplo, Notepad en Windows o TextEdit en macOS.

10. Ahora ya puede [cifrar datos](#).

Ejemplo: Generar un esquema de cifrado con columnas sealed, fingerprint y columnas

En este ejemplo, para `sales.csv`, hay tres columnas: `username`, `purchased` y `product`. Para estas columnas, queremos lo siguiente:

- Que la `product` column sea una columna `sealed`
- Que la columna `username` se cifre como columna `fingerprint`
- Que la `purchased` column sea una columna `cleartext`

Para generar un esquema de cifrado con columnas sealed, fingerprint y cleartext

1. (Opcional) Para garantizar que el archivo `c3r-cli.jar` y el archivo que se va a cifrar estén presentes:
 - a. Desplácese al directorio deseado y ejecute `ls` (si utiliza Mac o Unix/Linux) o `dir` si utiliza Windows).
 - b. Vea la lista de archivos de datos tabulares (`.csv`) y elija el archivo que desea cifrar.

En este ejemplo, `sales.csv` es el archivo que queremos cifrar.

2. Desde la CLI, ejecute el siguiente comando para crear un esquema de forma interactiva.

```
java -jar c3r-cli.jar schema sales.csv --interactive --  
output=sales.json
```

Note

- El indicador `--interactive` selecciona el modo interactivo para desarrollar el esquema. Esto guía al usuario a través de un flujo de trabajo guiado para crear el esquema.
- Si usted es un usuario avanzado, puede crear su propio esquema JSON sin utilizar el flujo de trabajo guiado. Para obtener más información, consulte [\(Opcional\) Crear un esquema \(usuarios avanzados\)](#).
- Para los archivos `.csv` sin encabezados de columna, consulte el indicador `--noHeaders` del comando de esquema disponible en la CLI.

- El indicador `--output` establece un nombre de salida. Si no incluye el indicador `--output`, el cliente de cifrado de C3R intentará elegir un nombre de salida predeterminado (por ejemplo, `<input>.out` o, para el esquema, `<input>.json`).

3. En `Number of target columns from source column 'username'?`, escriba **1** y luego presione Intro.
4. En `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, escriba **f** y luego presione Intro.
5. En `Target column headername <default 'username'>`, presione Intro.

Se utiliza el nombre predeterminado `'username'`.

6. En `Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'>`, escriba **y** y luego presione Intro.
7. En `Number of target columns from source column 'purchased'?`, escriba **1** y luego presione Intro.
8. En `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, escriba **c** y luego presione Intro.
9. En `Target column headername <default 'purchased'>`, presione Intro.

Se utiliza el nombre predeterminado `'purchased'`.

10. En `Number of target columns from source column 'product'?`, escriba **1** y luego presione Intro.
11. En `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`, escriba **s** y luego presione Intro.
12. En `Target column headername <default 'product'>`, presione Intro.

Se utiliza el nombre predeterminado `'product'`.

13. En `'product_sealed' padding type: [n]one, [f]ixed, or [m]ax <default 'max'>?`, presione Intro para elegir el valor predeterminado.
14. Para `Byte-length beyond max length to pad cleartext to in 'product_sealed' <default '0'>?`, presione Intro para elegir el valor predeterminado.

El esquema se escribe en un nuevo archivo llamado `sales.json`.

15. Ahora ya puede [cifrar datos](#).

Paso 5: crear una clave secreta compartida

Para cifrar las tablas de datos, los participantes de la colaboración deben acordar una clave secreta compartida y compartirla de forma segura.

La clave secreta compartida debe tener al menos 256 bits (32 bytes). Puede especificar una clave de mayor tamaño, pero no le proporcionará ninguna garantía adicional.

Important

Recuerde que la clave y el ID de la colaboración utilizados para el cifrado y el descifrado deben ser idénticos para todos los participantes de la colaboración.

En las siguientes secciones se proporcionan ejemplos de comandos de consola para generar una clave secreta compartida guardada como `secret.key` en el directorio de trabajo actual del terminal correspondiente.

Temas

- [Ejemplo: generación de claves con OpenSSL](#)
- [Ejemplo: generación de claves en Windows con PowerShell](#)

Ejemplo: generación de claves con OpenSSL

Para una biblioteca de criptografía común de uso general, ejecute el comando siguiente para crear una clave secreta compartida.

```
openssl rand 32 > secret.key
```

Si la está utilizando Windows y no ha instalado OpenSSL, puede generar claves utilizando el ejemplo descrito en [Ejemplo: generación de claves en Windows con PowerShell](#).

Ejemplo: generación de claves en Windows con PowerShell

En PowerShell, un terminal de aplicaciones disponible en Windows, ejecute el siguiente comando para crear una clave secreta compartida.

```
$bs = New-Object Byte[](32);  
[Security.Cryptography.RandomNumberGenerator]::Create().GetBytes($bs); Set-  
Content 'secret.key' -Encoding Byte -Value $bs
```

Paso 6: guardar la clave secreta compartida en la variable de entorno

Una variable de entorno es una forma cómoda y ampliable para que los usuarios proporcionen una clave secreta de varios almacenes de claves, por ejemplo, AWS Secrets Manager y la pasen al cliente de cifrado C3R.

El cliente de cifrado C3R puede utilizar claves almacenadas en Servicios de AWS si se utiliza AWS CLI para almacenar esas claves en la variable de entorno correspondiente. Por ejemplo, el cliente de cifrado C3R puede usar una clave de AWS Secrets Manager. Para obtener más información, consulte [Crear y administrar secretos con AWS Secrets Manager](#) en la Guía del usuario de AWS Secrets Manager .

Note

Sin embargo, antes de utilizar un Servicio de AWS como esta AWS Secrets Manager para guardar las claves C3R, compruebe que su caso de uso lo permite. Es posible que algunos casos de uso requieran que se oculte la clave. AWS Esto es para garantizar que los datos cifrados y la clave nunca estén en manos del mismo tercero.

Los únicos requisitos para una clave secreta compartida son que la clave secreta compartida esté codificada en base64 y almacenada en la variable de entorno C3R_SHARED_SECRET.

En las siguientes secciones se describen los comandos de la consola para convertir un archivo `secret.key` en base64 y almacenarlo como una variable de entorno. El archivo `secret.key` podría haberse generado a partir de cualquiera de los comandos enumerados en [Paso 5: crear una clave secreta compartida](#), y este es solo un ejemplo de origen.

Almacenamiento de la clave en una variable de entorno en Windows con PowerShell

Para convertir a base64 y configurar la variable de entorno en Windows con PowerShell, ejecute el siguiente comando.

```
$Bytes=[IO.File]::ReadAllBytes((Get-Location).ToString()+'\secret.key');  
$env:C3R_SHARED_SECRET=[Convert]::ToBase64String($Bytes)
```

Almacenamiento de la clave en una variable de entorno en Linux o macOS

Para convertir a base64 y establecer la variable de entorno en Linux o macOS, ejecute el siguiente comando.

```
export C3R_SHARED_SECRET="$(cat secret.key | base64)"
```

Paso 3: cifrar los datos

Para realizar este paso, debe adquirir el ID de AWS Clean Rooms colaboración y la clave secreta compartida. Para obtener más información, consulte [Requisitos previos](#).

En el siguiente ejemplo, ejecutamos el cifrado en `ads.csv` utilizando el esquema que hemos creado, denominado `ads.json`.

Para cifrar los datos

1. Almacene la clave secreta compartida de la colaboración en [Paso 6: guardar la clave secreta compartida en la variable de entorno](#).
2. Desde la línea de comandos, ingrese el comando siguiente:


```
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. Para *<name of input .csv file>* ello, introduzca el nombre del archivo.csv de entrada.
4. En `schema=`, introduzca el nombre del archivo de esquema de cifrado .json.
5. En `id=`, introduzca el ID de la colaboración.
6. En `output=`, introduzca el nombre del archivo de salida (por ejemplo, `ads-output.csv`).
7. Incluya cualquiera de los indicadores de línea de comandos descritos en [Parámetros de computación criptográfica](#) y [Indicadores opcionales en computación criptográfica para Clean Rooms](#).
8. Ejecute el comando.

En el ejemplo de `ads.csv`, ejecutamos el siguiente comando.

```
java -jar c3r-cli.jar encrypt ads.csv --schema=ads.json --id=123e4567-e89b-42d3-a456-556642440000 --output=ads-output.csv
```

En el ejemplo de `sales.csv`, ejecutamos el siguiente comando.

```
java -jar c3r-cli.jar encrypt sales.csv --schema=sales.json --id=123e4567-e89b-42d3-a456-556642440000
```

 Note

En este ejemplo, no especificamos un nombre de archivo de salida (`--output=sales-output.csv`). Como resultado, se generó el nombre de archivo de salida predeterminado `name-of-file.out.csv`.

Ahora ya puede verificar los datos cifrados.

Paso 8: verificar el cifrado de datos

Para verificar que los datos estaban cifrados

1. Vea el archivo de datos cifrados (por ejemplo, `sales-output.csv`).
2. Compruebe las siguientes columnas:
 - a. Columna 1: cifrada (por ejemplo, `username_fingerprint`).

En las columnas fingerprint (HMAC), después de la versión y el prefijo de tipo (por ejemplo, `01:hmac:`), hay 44 caracteres de datos codificados en base64.

- b. Columna 2: sin cifrar (por ejemplo, `purchased`).
- c. Columna 3: cifrada (por ejemplo, `product_sealed`).

En el caso de las columnas cifradas (SELECT), la longitud del cleartext más cualquier relleno posterior a la versión y el prefijo de tipo (por ejemplo, `01:enc:`) es directamente proporcional a la longitud del cleartext columna cifrada. Es decir, la longitud es el tamaño de la entrada, más aproximadamente un 33 por ciento de sobrecarga debido a la codificación.

Ya puede hacer lo siguiente:

1. [Cargar los datos cifrados a S3.](#)
2. [Cree una AWS Glue tabla.](#)
3. [Crear una tabla configurada en AWS Clean Rooms.](#)

El cliente de cifrado de C3R creará archivos temporales que no contengan datos sin cifrar (a menos que esos datos tampoco estén cifrados en el resultado final). No obstante, es posible que algunos valores cifrados no se rellenen correctamente. Las columnas de huella digital pueden contener

valores duplicados, incluso si el ajuste de colaboración `allowRepeatedFingerprintValue` es `false`. Este problema se debe a que el archivo temporal se escribe antes de comprobar las longitudes de relleno adecuadas y las propiedades de eliminación de duplicados.

Si el cliente de cifrado de C3R falla o se interrumpe durante el cifrado, es posible que se detenga después de escribir el archivo temporal, pero antes de comprobar estas propiedades y eliminar los archivos temporales. Por lo tanto, es posible que estos archivos temporales sigan en el disco. Si este es el caso, el contenido de estos archivos no protege los datos de texto sin formato en los mismos niveles que los de salida. En concreto, estos archivos temporales pueden revelar datos del texto sin formato para los análisis estadísticos, lo que no serviría de nada para el resultado final. El usuario debe eliminar estos archivos (especialmente la base de datos SQLite) para evitar que caigan en manos no autorizadas.

(Opcional) Crear un esquema (usuarios avanzados)

La creación manual de esquemas está reservada para los usuarios avanzados.

A continuación se describe el formato del archivo de esquema JSON para los archivos de entrada con o sin encabezados de columna. Los usuarios avanzados pueden escribir o modificar el esquema directamente si lo desean.

Note

El cliente de cifrado de C3R puede ayudarlo a crear un esquema mediante el proceso interactivo descrito en [Ejemplo: Generar un esquema de cifrado con columnas sealed, fingerprint y columnas](#) o mediante la creación de una plantilla stub.

Esquemas de tablas mapeados y posicionales

En la siguiente sección se describen dos tipos de esquemas de tabla:

- Esquema de tabla mapeado: este esquema se usa para cifrar archivos .csv con una fila de encabezado y archivos Apache Parquet.
- Esquema de tabla posicional: este esquema se usa para cifrar archivos .csv sin fila de encabezado.

El cliente de cifrado de C3R puede cifrar un archivo tabular para una colaboración. Para ello, debe tener un archivo de esquema correspondiente que especifique cómo se debe derivar la salida cifrada a partir de la entrada.

El cliente de cifrado de C3R puede ayudar a generar un esquema para un archivo INPUT ejecutando el comando de esquema del cliente de cifrado de C3R en la línea de comandos. Un ejemplo de comando es `java -jar c3r-cli.jar schema --interactive INPUT`.

El esquema especifica la siguiente información:

1. Qué columnas de origen se asignan a qué columnas transformadas en el archivo de salida mediante sus nombres de encabezado (esquemas mapeados) o su posición (esquemas posicionales)
2. Qué columnas de destino van a permanecer como cleartext
3. Qué columnas de destino se van a cifrar para las consultas SELECT
4. Qué columnas de destino se van a cifrar para las consultas JOIN

Esta información está codificada en un archivo de esquema JSON específico de tabla, que consta de un único objeto cuyo campo `headerRow` es un valor booleano. El valor debe ser `true` para los archivos Parquet y los archivos `.csv` con fila de encabezado, y `false` para el resto.

Esquema de la tabla mapeado

El esquema mapeado tiene la siguiente forma.

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": STRING,
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    },
    ...
  ]
}
```

Si `headerRow` es `true`, es el siguiente campo del objeto es `columns`, que contiene una matriz de esquemas de columnas que asignan los encabezados de origen a los encabezados de destino (es decir, objetos JSON que describen lo que deben contener las columnas de salida).

- `sourceHeader`: el nombre del encabezado `STRING` de la columna de origen a partir de cual se derivan los datos.

 Note

Una misma columna de origen se puede utilizar para varias columnas de destino. Una columna del archivo de entrada que no aparezca enumerada como `sourceHeader` en ninguna parte del esquema no aparecerá en el archivo de salida.

- `targetHeader`: el nombre del encabezado `STRING` de la columna correspondiente del archivo de salida.

 Note

Este campo es opcional para los esquemas mapeados. Si se omite este campo, `sourceHeader` se reutiliza para el nombre del encabezado en el archivo de salida. Se anexa `_fingerprint` o `_sealed` si la columna de salida es una columna `fingerprint` o una columna `sealed`, respectivamente.

- `type`: el `TYPE` de la columna de destino del archivo de salida. Es decir, una de las columnas `cleartext`, `sealed` o `fingerprint` dependiendo de cómo se vaya a utilizar la columna en la colaboración.
- `pad`: un campo de un objeto de esquema de columnas que solo está presente cuando `TYPE` es `sealed`. Su valor correspondiente de `PAD` es un objeto que describe cómo deben rellenarse los datos antes del cifrado.

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

Para especificar el relleno previo al cifrado, se utiliza `type` y `length` de la siguiente manera:

- `PAD_TYPE` como `none`: no se aplicará ningún relleno a los datos de la columna y el campo `length` no es aplicable (es decir, se omite).

- PAD_TYPE como fixed: los datos de la columna se rellenan hasta el número de bytes especificado, length.
- PAD_TYPE como max: los datos de la columna se rellenan hasta el tamaño de la longitud en bytes del valor más largo, más length bytes adicionales.

El siguiente es un ejemplo de esquema mapeado, con una columna de cada tipo.

```
{
  "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
      }
    }
  ]
}
```

Un ejemplo más complejo sería el siguiente ejemplo de un archivo .csv con encabezados.

```

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,C00,9,""
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

```

En el siguiente ejemplo de esquema mapeado, las columnas `FirstName` y `LastName` son columnas `cleartext`. La columna `State` está cifrada como columna `fingerprint` y como columna `sealed` con un relleno de `none`. Las columnas restantes se omiten.

```

{
  "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"
      }
    }
  ]
}

```

```
}

```

El siguiente es el archivo .csv que resulta del esquema mapeado.

```
givenname,surname,state_fingerprint,state
John,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:FQ3n3Ahv9BQQNWQGcugeHzHYzEZE1vapHa2Uu4SRgSatZ3q0bjPA4TcsHt
+B0kMKBcnHWI13BeGG/SBqmj7vKpI=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:KZ5n5GtaXACco65AXk48BQ02durDNR2ULc4YxmMC8NaZZKKJiksU1IwFadAvV4iBQ1
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:mLKpS5HIOSgphdEsrzhd
eN9nB02gAbIygt40Fn4La1Yn9Xyj/XUWXlmn8zFe2T4kyDTD8kG0vpQEUGxAUFk=
Diego,Ramirez,01:hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:rmZhT98Zm
+IIGw1UTjMIJP4IrW/AA1tBLMXcHvnYfRgmWP623VFQ6aUnhsb2MDqEw4G5Uwg5rKKZepUxx5uKbfbk=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhN0d7s2ZiKUe7QiTy08=,01:enc:vVaqWC1VRbhvkf8gnuR7q0z
Terry,Whitlock01:hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:3c9VEwb0D0/
xbQjdGuccLvI7oZTBdPU+SyrJIyr2kudfAxbuMQ2uRdU/q7rbgyJjxZS8M2U35ILJf/1DgTyg7cM=
Jane,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:9RWv46YLveykeNZ/
G0Nd1YFg+AVd0nu05hHyAYTQkPLHnyX+0/jbzD/g9ZT8GCgVE9aB5bV4ooJIXHGBVMXcjrQ=
```

Esquema de tabla posicional

El esquema posicional tiene la siguiente forma.

```
{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": STRING,
        "type": TYPE,
        "pad": PAD
      },
      {
        "targetHeader": STRING,
        "type": TYPE,
        "pad": PAD
      }
    ],
    [],
    ...
  ]
}
```

Si `headerRow` es `false`, es el siguiente campo del objeto es `columns`, que contiene una matriz de entradas. Cada entrada es en sí misma una matriz de cero o más esquemas de columnas posicionales (sin campo `sourceHeader`), que son objetos JSON que describen lo que debe contener la salida.

- `sourceHeader`: el nombre del encabezado `STRING` de la columna de origen a partir de cual se derivan los datos.

 Note

Este campo debe omitirse en los esquemas posicionales. En los esquemas posicionales, la columna de origen se deduce mediante el índice correspondiente de la columna en el archivo de esquema.

- `targetHeader`: el nombre del encabezado `STRING` de la columna correspondiente del archivo de salida.

 Note

Este campo es obligatorio para los esquemas posicionales.

- `type`: el `TYPE` de la columna de destino del archivo de salida. Es decir, una de las columnas `cleartext`, `sealed` o `fingerprint` dependiendo de cómo se vaya a utilizar la columna en la colaboración.
- `pad`: un campo de un objeto de esquema de columnas que solo está presente cuando `TYPE` es `sealed`. Su valor correspondiente de `PAD` es un objeto que describe cómo deben rellenarse los datos antes del cifrado.

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

Para especificar el relleno previo al cifrado, se utiliza `type` y `length` de la siguiente manera:

- `PAD_TYPE` como `none`: no se aplicará ningún relleno a los datos de la columna y el campo `length` no es aplicable (es decir, se omite).
- `PAD_TYPE` como `fixed`: los datos de la columna se rellenan hasta el número de bytes especificado, `length`.

- PAD_TYPE como max: los datos de la columna se rellenan hasta el tamaño de la longitud en bytes del valor más largo, más length bytes adicionales.

 Note

fixed es útil si sabe con antelación cuál es el límite superior del tamaño en bytes de los datos de la columna. Se produce un error si algún dato de esa columna supera la length especificada.

max resulta práctico cuando se desconoce el tamaño exacto de los datos de entrada, ya que funciona independientemente del tamaño de los datos. Sin embargo, max requiere un tiempo de procesamiento adicional porque cifra los datos dos veces. max cifra los datos una vez cuando se leen en el archivo temporal y una vez que se conoce la entrada de datos más larga de la columna.

Además, la longitud del valor más largo no se guarda entre las invocaciones del cliente. Si tiene previsto cifrar los datos por lotes o cifrar datos nuevos periódicamente, tenga en cuenta que la longitud del texto cifrado resultante puede variar de un lote a otro.

A continuación se muestra un ejemplo de esquema posicional.

```
{
  "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
      }
    }
  ]
}

```

Como ejemplo complejo, el siguiente es un ejemplo de un archivo .csv sin la primera fila con los encabezados.

```

Jorge,Souza,12345 Mills Rd,Anytown,SC, 703 -555 -1234,CE0, 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

```

El esquema posicional tiene la siguiente forma.

```

{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": "GivenName",
        "type": "cleartext"
      }
    ],
    [
      {
        "targetHeader": "Surname",
        "type": "cleartext"
      }
    ]
  ]
}

```



```

    }
  ],
  [],
  [],
  [
    {
      "targetHeader": "State_Join",
      "type": "fingerprint"
    },
    {
      "targetHeader": "State",
      "type": "sealed",
      "pad": {
        "type": "none"
      }
    }
  ]
],
[],
[],
[],
[]
]
}

```

El esquema anterior produce el siguiente archivo de salida con una fila de encabezados que contiene los encabezados de destino especificados.

```

givenname,surname,state_fingerprint,state
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:ENS6QD3cMV19vQEGfe9MM
Q8m/Y5SA89dJwKpT5rGPP8e36h6klwDoslpFzGvU0=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhN0d7s2ZiKUe7QiTy08=,01:enc:LKo0zirq2+
+XEIIIMNRjAsGMdyWUDwYaum0B+IFP+rUf1BNeZDJjtFe1Z+zbZfXQWwJy52Rt7HqvAb2WIK1oMmk=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:MyQKyWxJ9kvK1xDQQtX1UNwv3F+yrBRr0xrUY/1BGg5KFg0n9pK+MZ7g
+ZNqZEPcPz4lht1u0t/wbTaQz0CLXFQ=
Jane,Doe,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:Pd8sbITBfb0/
ttUB4svVsgoYkDfnDvgkvxzeCi0Yxq54rLSwccy1o3/B50C3cpkkn56dovCwzgmPNwrmCmYtb4=
Terry,Whitlock01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:Qmtzu3B3GAXKh2KkRYTiEAaMopYedsSdF2e/
ADUiBQ9kv2CxKPzWyYTD3ztmKPMka19dHre5VhUHNp030+j1AQ8=
Diego,Ramirez,01:hmac:UK8s8Cn/WR2J0/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:ysdg
+GHKdeZrS/geBIoo0EPLHG68MsWpx1dh3xjb+fG5rmFmqUcJLNuuYBHhHA1xchM2WVeV1fmHkBX3mvZNvkc=

```

```
John,Doe,01: hmac:UK8s8Cn/WR2J0/To2dTxD73aDEe2ZUXeSHy3Tv+1Mk=,01: enc:9uX0wZu07kAPAx  
+Hf6uvQownkWqFSKtWS7gQIJSe5aXFquKWCK6yZN0X5Ea2N3bn03Uj1kh0agDwoiP9FRZGJA4=
```

Descifrar tablas de datos con el cliente de cifrado de C3R

Siga este procedimiento para las colaboraciones que utilizan la computación criptográfica para Clean Rooms y el cliente de cifrado C3R para cifrar tablas de datos. Utilice este procedimiento después de haber [consultado datos de la colaboración](#).

La clave secreta compartida y el ID de colaboración son necesarios para este procedimiento.

El miembro que puede recibir los resultados descifra los datos con la misma clave secreta compartida y el mismo ID de colaboración que se utilizaron para cifrar los datos de la colaboración.

Note

AWS Clean Rooms las colaboraciones ya limitan quién puede realizar y ver los resultados de las consultas. Para realizar el descifrado, quien tenga acceso a estos resultados necesita la misma clave secreta compartida y el mismo ID de colaboración que se utilizaron para cifrar los datos.

Para descifrar una tabla de datos cifrada

1. (Opcional) [Consulte los comandos disponibles en el cliente de cifrado de C3R](#).
2. (Opcional) Navegue hasta el directorio deseado y ejecute `ls` (macOS) o `dir` (Windows).
 - Compruebe que el `c3r-cli.jar` el archivo y el archivo de datos de resultados de la consulta cifrados se encuentran en el directorio deseado.

Note

Si los resultados de la consulta se descargan desde la interfaz de la AWS Clean Rooms consola, es probable que estén en la carpeta Descargas de su cuenta de usuario. (Por ejemplo, la carpeta Descargas del directorio de usuarios de Windows y macOS.) Se recomienda mover el archivo de resultados de la consulta a la misma carpeta que `c3r-cli.jar`.

3. Guarde la clave secreta compartida en la variable de entorno C3R_SHARED_SECRET. Para obtener más información, consulte [Paso 6: guardar la clave secreta compartida en la variable de entorno](#).

4. Desde AWS Command Line Interface (AWS CLI), ejecute el siguiente comando.

```
java -jar c3r-cli.jar decrypt <name of input .csv file> --id=<collaboration id> --  
output=<output file name>
```

5. Sustituya cada uno *user input placeholder* por su propia información:

- a. En id=, introduzca el ID de la colaboración.
- b. En output=, introduzca el nombre del archivo de salida (por ejemplo, results-decrypt.csv).

Si no especifica un nombre de salida, se mostrará un nombre predeterminado en el terminal.

- c. Vea los datos descifrados en el archivo de salida especificado utilizando el CSV que prefiera o Parquet aplicación de visualización (como Microsoft Excel, un editor de texto u otra aplicación).

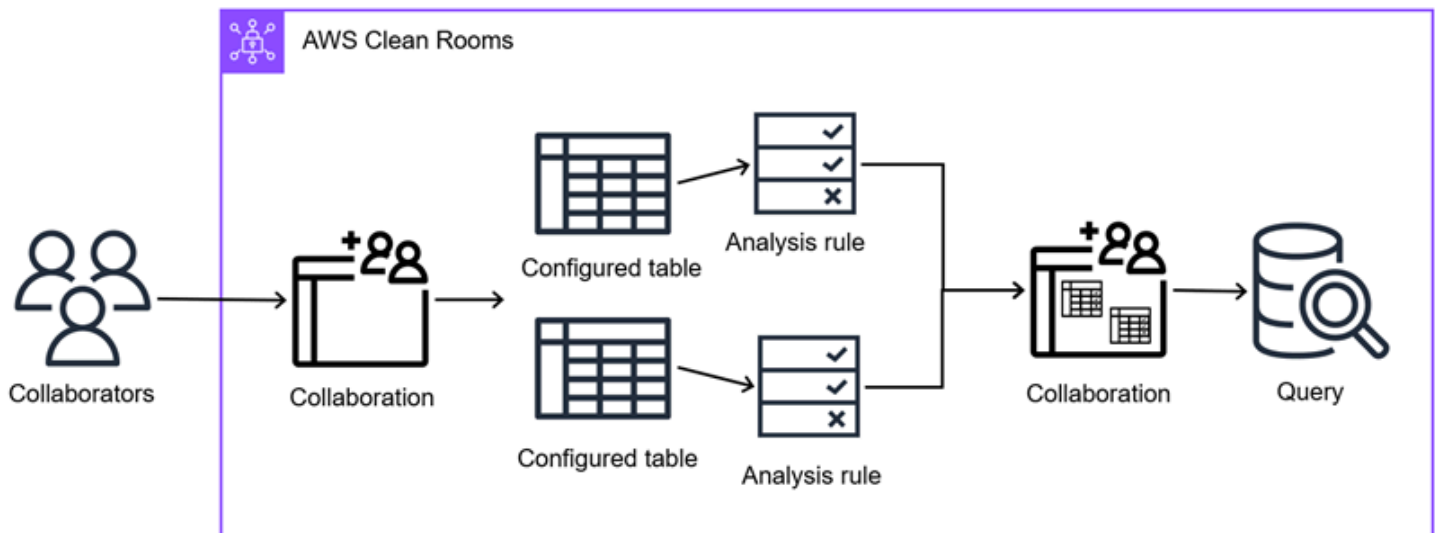
Tablas configuradas en AWS Clean Rooms

Una tabla configurada es una referencia a una tabla existente en una fuente de datos. Contiene una regla de análisis que determina cómo se pueden consultar los datos AWS Clean Rooms y puede incluir un presupuesto de acceso a los datos para controlar el uso de la tabla. Las tablas configuradas se pueden asociar a una o más colaboraciones.

Con él AWS Clean Rooms, puede realizar un análisis de agregación de los datos de los eventos, como el número de compras en comparación con el número de compras. También puede realizar un análisis de listas de datos de eventos, por ejemplo, enriqueciendo los datos de clientes superpuestos pasando de datos de segmentos a datos de CRM. También puede realizar consultas personalizadas y establecer una privacidad diferencial en los datos de los eventos, como los datos de audiencia y los atributos de los segmentos. Para cualquiera de estos tipos de análisis, puede establecer presupuestos de acceso a los datos para supervisar y controlar la cantidad de datos a la que se accede mediante consultas.

En primer lugar, puede crear una colaboración AWS Clean Rooms y añadir la Cuentas de AWS que desee invitar, o bien unirse a una colaboración a la que esté invitado creando una membresía. A continuación, usted y el otro miembro de la colaboración crean tablas configuradas. Puede añadir una regla de análisis a las tablas configuradas (agregación, lista o personalizada) y, si lo desea, establecer presupuestos de acceso a los datos. A continuación, asocie las tablas configuradas a la colaboración. Por último, el miembro que puede realizar la consulta ejecuta una consulta en las dos tablas de datos, lo que consume el presupuesto de acceso a los datos a medida que se ejecutan las consultas.

El siguiente diagrama resume cómo trabajar con los datos de eventos en AWS Clean Rooms.



Temas

- [Crear una tabla configurada en AWS Clean Rooms](#)
- [Agregación de una regla de análisis a una tabla configurada](#)
- [Asociar una tabla configurada a una colaboración](#)
- [Configuración de un presupuesto de acceso de datos](#)
- [Agregación de una regla de análisis de colaboración a una tabla configurada](#)
- [Configuración de la política de privacidad diferencial \(opcional\)](#)
- [Ver tablas y reglas de análisis](#)
- [Edición de una tabla configurada](#)
- [Editar etiquetas de tablas configuradas](#)
- [Edición de la regla de análisis de tablas configurada](#)
- [Eliminar la regla de análisis de tablas configurada](#)
- [Columnas no permitidas de la tabla configurada](#)
- [Editar asociaciones de tablas configuradas](#)
- [Disociar tablas configuradas](#)

Crear una tabla configurada en AWS Clean Rooms

Una tabla configurada es una referencia a una tabla existente en una fuente de datos. Contiene una regla de análisis que determina cómo se pueden consultar los datos en AWS Clean Rooms. Las tablas configuradas se pueden asociar a una o más colaboraciones.

Para obtener información sobre cómo crear una tabla configurada mediante la AWS SDKs, consulta la [referencia de la AWS Clean Rooms API](#).

Temas

- [Creación de una tabla configurada: fuente de datos de Amazon S3](#)
- [Creación de una tabla configurada: fuente de datos de Amazon Athena](#)
- [Creación de una tabla configurada: fuente de datos de Snowflake](#)

Creación de una tabla configurada: fuente de datos de Amazon S3

En este procedimiento, el [miembro](#) realiza las siguientes tareas:

- Configura una AWS Glue tabla existente para usarla en AWS Clean Rooms. (este paso se puede realizar antes o después de unirse a una colaboración, a menos que se utilice la computación criptográfica para Clean Rooms).

Note

AWS Clean Rooms admite AWS Glue tablas. Para obtener más información sobre cómo introducir los datos AWS Glue, consulte [Paso 3: cargar la tabla de datos en Amazon S3](#).

- Asigna un nombre a la [tabla configurada](#) y elige qué columnas usar en la colaboración.

El siguiente procedimiento asume que:

- El miembro de la colaboración ya ha [subido sus tablas de datos a Amazon S3](#) y [ha creado una AWS Glue tabla](#).

Note

El destino de los resultados en Amazon S3 no puede estar dentro del mismo bucket de S3 que cualquier fuente de datos.

- (Opcional) Solo para las tablas de datos [cifrados](#), el miembro de la colaboración ya ha [preparado las tablas de datos cifrados](#) con el cliente de cifrado de C3R.

Puede utilizar la generación de estadísticas proporcionada por AWS Glue para calcular las estadísticas a nivel de columna para las tablas. AWS Glue Data Catalog Tras AWS Glue generar las estadísticas para las tablas del catálogo de datos, Amazon Redshift Spectrum las utiliza automáticamente para optimizar el plan de consultas. Para obtener más información sobre cómo calcular las estadísticas a nivel de columna mediante el uso de estadísticas de columnas AWS Glue, consulte [Optimización del rendimiento de las consultas mediante estadísticas de columnas](#) en la Guía del AWS Glue usuario. Para obtener más información AWS Glue, consulte la [Guía para desarrolladores de AWS Glue](#).

Para crear una tabla configurada: fuente de datos de Amazon S3

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. En la esquina superior derecha, elija Configurar nueva tabla.
4. En Fuente de datos, en Fuentes de AWS datos, elija Amazon S3.
5. En la tabla de Amazon S3:
 - a. Seleccione la región en la que está alojada la tabla S3.

De forma predeterminada, se selecciona la región actual (como North Virginia us-east-1).

 Warning

Si la fuente de datos de Amazon S3 se encuentra en una región diferente a la de su ubicación de procesamiento, el procesamiento de datos puede realizarse temporalmente fuera de la región de origen. Antes de continuar, compruebe que el movimiento de datos entre regiones cumpla con sus requisitos de soberanía de datos, las políticas de conformidad normativa y las normas de gobierno de datos.

Para obtener más información sobre las regiones, consulte [Regiones y puntos finales](#) en la Referencia general de AWS

- b. Elija la base de datos en la lista desplegable.
- c. Seleccione la Tabla que desea configurar en la lista desplegable.

 Note

Para comprobar que se trata de la tabla correcta, realice una de las siguientes acciones:

- Seleccione Ver en AWS Glue.
- Activa Ver esquema desde AWS Glue para ver el esquema.

 Important

En el caso de AWS Glue las tablas en las que los datos están en formato CSV, los nombres y el orden de las columnas en el esquema Glue deben coincidir exactamente con los datos CSV. Si no se alinean, es posible que la lista de columnas permitidas para la tabla configurada no se aplique correctamente.

6. En el caso de las columnas y los métodos de análisis permitidos en las colaboraciones,
 - a. ¿Para qué columnas quieres permitir en las colaboraciones?
 - Seleccione Todas las columnas para permitir que se consulten todas las columnas de la colaboración.
 - Elija Lista personalizada para poder consultar una o más columnas de la lista desplegable Especificar columnas permitidas en la colaboración.
 - b. Para los métodos de análisis permitidos,
 - i. Elija Consulta directa para permitir que las consultas SQL se ejecuten directamente en esta tabla
 - ii. Elija Trabajo directo para permitir que los PySpark trabajos se ejecuten directamente en esta tabla.

Example Ejemplo

Por ejemplo, si desea permitir que los miembros de la colaboración ejecuten consultas SQL directas y PySpark trabajos en todas las columnas, seleccione Todas las columnas, Consulta directa y Trabajo directo.

7. En Detalles de la tabla configurada,

- a. Introduzca un Nombre para la tabla configurada.

Puede usar el nombre predeterminado o cambiar el nombre de esta tabla.

- b. Introduzca una Descripción de la tabla.

La descripción ayuda a diferenciarla de otras tablas configuradas con nombres similares.

8. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
9. Seleccione Configurar nuevo cliente.

Ahora que ha creado una tabla configurada, puede hacer lo siguiente:

- [Agregación de una regla de análisis a la tabla configurada](#)
- [Asociar la tabla configurada a una colaboración](#)

Creación de una tabla configurada: fuente de datos de Amazon Athena

La opción de fuente de datos de Amazon Athena le permite consultar los datos almacenados en Amazon S3, catalogados en el catálogo de AWS Glue datos o en los catálogos federados y acceder a ellos de forma controlada mediante AWS Lake Formation. Se admiten tanto las tablas como las vistas. AWS Glue Data Catalog Los enlaces a los recursos de Lake Formation se pueden usar para compartir tablas y vistas. Regiones de AWS a lo largo de Cuentas de AWS y ancho de la cuenta del AWS Clean Rooms miembro que los une a una AWS Clean Rooms colaboración.

Note

Solo se pueden consultar los conjuntos de datos basados en Amazon S3 mediante la integración de fuentes de datos de Athena.

En este procedimiento, el [miembro](#) realiza las siguientes tareas:

- Configura una tabla o vista existente en el para usarla AWS Glue Data Catalog AWS Clean Rooms
- Asigna un nombre a la [tabla configurada](#) y elige qué columnas usar en la colaboración.

El siguiente procedimiento asume que:

- El miembro de la colaboración ya creó la AWS Glue Data Catalog base de datos y la vista de tabla o GDC.

Para crear una tabla configurada: fuente de datos de Athena

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. En la esquina superior derecha, elija Configurar nueva tabla.
4. En Fuente de datos, en Fuentes AWS de datos, elija Amazon Athena.
5. En la tabla de Amazon Athena:
 - a. Seleccione la región en la que está alojada la mesa de Amazon Athena.

De forma predeterminada, se selecciona la región actual (como North Virginia us-east-1).

 Warning

Si su fuente de datos de Amazon Athena se encuentra en una región diferente a la de su ubicación de procesamiento, el procesamiento de datos puede realizarse temporalmente fuera de la región de origen. Antes de continuar, compruebe que el movimiento de datos entre regiones cumpla con sus requisitos de soberanía de datos, las políticas de conformidad normativa y las normas de gobierno de datos.

Para obtener más información sobre las regiones, consulte [Regiones y puntos finales](#) en la Referencia general de AWS

- b. Elija el catálogo en la lista desplegable.

De forma predeterminada, se selecciona Catálogo de AWS Glue datos.

- AWS Glue Catálogo de datos: el catálogo predeterminado para las tablas de AWS Glue.
- Catálogo federado: disponible si ha configurado AWS Glue Catalog Federation para conectarse a catálogos REST remotos de Apache Iceberg. Para obtener más información, consulte la [federación de catálogos en la AWS Lake Formation Guía](#) para desarrolladores.

- c. Elija la base de datos en la lista desplegable.

- d. Seleccione la Tabla que desea configurar en la lista desplegable.

 Note

Para comprobar que se trata de la tabla correcta, realice una de las siguientes acciones:

- Seleccione Ver en AWS Glue o Ver en AWS Lake Formation (según el tipo de catálogo).
- Activa Ver esquema desde AWS Glue para ver el esquema.

6. Para las configuraciones de Amazon Athena,
- Elija un grupo de trabajo de la lista desplegable.
 - Para la ubicación de salida de S3, elija una acción recomendada en función de uno de los siguientes escenarios.

Escenario	Acción recomendada
Su grupo de trabajo no tiene una ubicación de salida predeterminada.	Introduzca la ubicación de salida de S3 o elija Browse S3.
Su grupo de trabajo impone su ubicación de salida predeterminada.	La ubicación de salida del S3 se elige automáticamente y no puede cambiarla.
Su grupo de trabajo no impone su ubicación de salida predeterminada.	Introduzca la ubicación de salida de S3 o elija Browse S3.

7. En el caso de las columnas permitidas en las colaboraciones, elige una opción en función de tu objetivo.

Su objetivo	Opción recomendada
Permite el uso de todas las columnas AWS Clean Rooms (sujeto a las reglas de análisis)	Todas las columnas
Permita una o más columnas de la lista desplegable Especificar columnas permitidas	Lista personalizada

8. En Detalles de la tabla configurada,

- a. Introduzca un Nombre para la tabla configurada.

Puede usar el nombre predeterminado o cambiar el nombre de esta tabla.

- b. Introduzca una Descripción de la tabla.

La descripción ayuda a diferenciarla de otras tablas configuradas con nombres similares.

- c. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.

9. Seleccione Configurar nuevo cliente.

Ahora que ha creado una tabla configurada, puede hacer lo siguiente:

- [Agregación de una regla de análisis a la tabla configurada](#)
- [Asociar la tabla configurada a una colaboración](#)

Creación de una tabla configurada: fuente de datos de Snowflake

En este procedimiento, el [miembro](#) realiza las siguientes tareas:

- Configura una tabla Snowflake existente para usarla en. AWS Clean Rooms (este paso se puede realizar antes o después de unirse a una colaboración, a menos que se utilice la computación criptográfica para Clean Rooms).
- Asigna un nombre a la [tabla configurada](#) y elige qué columnas usar en la colaboración.

El siguiente procedimiento asume que:

- El miembro de la colaboración ya ha subido sus tablas de datos a Snowflake.
- (Opcional) Solo para las tablas de datos [cifrados](#), el miembro de la colaboración ya ha [preparado las tablas de datos cifrados](#) con el cliente de cifrado de C3R.

Para crear una tabla configurada: fuente de datos de Snowflake

1. [Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en / cleanrooms. https://console.aws.amazon.com](https://console.aws.amazon.com/cleanrooms)

2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. En la esquina superior derecha, elija Configurar nueva tabla.
4. En Fuente de datos, en Nubes y fuentes de datos de terceros, selecciona Snowflake.
5. Especifique las credenciales de Snowflake utilizando un ARN secreto existente o almacenando un secreto nuevo para esta tabla.

Use existing secret ARN

1. Si tienes un ARN secreto, introdúcelo en el campo ARN secreto.

Para buscar tu ARN secreto, selecciona Ir a. AWS Secrets Manager

2. Si ya tiene un secreto de otra tabla, elija Importar el ARN secreto de la tabla existente.



Note

El ARN secreto puede ser multicuenta.

Store a new secret for this table

1. Introduzca las siguientes credenciales de Snowflake:
 - Nombre de usuario de Snowflake
 - Almacén Snowflake
 - Papel de Snowflake
 - Clave privada de Snowflake Privacy Enhanced Mail (PEM)
2. Para el cifrado, realice una de las siguientes acciones:
 - Para usar la opción Clave administrada de AWS (predeterminada), deje desactivada la casilla Personalizar la configuración de cifrado.
 - Para usar una personalizada AWS KMS key:
 - Seleccione la casilla de verificación Personalizar la configuración de cifrado.
 - Para la clave KMS, introduzca la clave ARN o elija una de la lista.
3. Introduzca un nombre secreto que le ayude a encontrar sus credenciales más adelante.
6. Para los detalles de la tabla y el esquema de Snowflake, introdúzcalos manualmente o impórtelos automáticamente.

Enter the details manually

1. Introduzca el identificador de la cuenta de Snowflake.

Para obtener más información, consulte los [identificadores de cuenta](#) en la documentación de Snowflake.

El identificador de su cuenta debe tener el formato utilizado para los conductores de Snowflake. Debe sustituir el punto (.) por un guión (-) para que el identificador tenga el formato siguiente. **<orgname>-<account_name>**

2. Introduzca la base de datos de Snowflake.

Para obtener más información, consulte la [base de datos de Snowflake en la documentación de Snowflake](#).

3. Introduzca el nombre del esquema de Snowflake.

4. Introduzca el nombre de la tabla Snowflake.

Para obtener más información, consulte [Descripción de las estructuras de las tablas de Snowflake](#) en la documentación de Snowflake.

5. Para el esquema, introduzca el nombre de la columna y elija el tipo de datos en la lista desplegable.

6. Seleccione Añadir columna para añadir más columnas.

- Si elige un tipo de datos de objeto, especifique el esquema del objeto.

Example Ejemplo de esquema de objetos

```
name STRING,  
location OBJECT(  
  x INT,  
  y INT,  
  metadata OBJECT(uuid STRING)  
),  
history ARRAY(TEXT)
```

- Si elige un tipo de datos de matriz, especifique el esquema de matriz.

Example Ejemplo de esquema de matriz

```
OBJECT(x INT, y INT)
```

- Si elige un tipo de datos de mapa, especifique el esquema de mapa.

Example Ejemplo de esquema de mapa

```
STRING, OBJECT(x INT, y INT)
```

Automatically import the details

1. Exporte la vista de COLUMNAS de Snowflake como un archivo CSV.

[Para obtener más información sobre la vista COLUMNAS de Snowflake, consulte la vista COLUMNAS en la documentación de Snowflake.](#)

2. Seleccione Importar desde un archivo para importar el archivo CSV y especificar cualquier información adicional.

El nombre de la base de datos, el nombre del esquema, el nombre de la tabla, los nombres de las columnas y los tipos de datos se importan automáticamente.

- Si elige un tipo de datos de objeto, especifique el esquema del objeto.
 - Si elige un tipo de datos de matriz, especifique el esquema de matriz.
 - Si elige un tipo de datos de mapa, especifique el esquema de mapa.
3. Introduzca el identificador de la cuenta de Snowflake.

Para obtener más información, consulte los [identificadores de cuenta](#) en la documentación de Snowflake.

Note

Solo las tablas de S3 catalogadas se AWS Glue pueden usar para recuperar el esquema de la tabla automáticamente.

7. En el caso de las columnas permitidas en las colaboraciones, elige una opción en función de tu objetivo.

Su objetivo	Opción recomendada
Permite el uso de todas las columnas AWS Clean Rooms (sujeto a las reglas de análisis)	Todas las columnas
Permita una o más columnas de la lista desplegable Especificar columnas permitidas	Lista personalizada

8. En Detalles de la tabla configurada,

a. Introduzca un Nombre para la tabla configurada.

Puede usar el nombre predeterminado o cambiar el nombre de esta tabla.

b. Introduzca una Descripción de la tabla.

La descripción ayuda a diferenciarla de otras tablas configuradas con nombres similares.

c. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.

9. Seleccione Configurar nuevo cliente.

Ahora que ha creado una tabla configurada, puede hacer lo siguiente:

- [Agregación de una regla de análisis a la tabla configurada](#)
- [Asociar la tabla configurada a una colaboración](#)

Agregación de una regla de análisis a una tabla configurada

En las secciones siguientes se describe cómo agregar una regla de análisis a la tabla configurada. Al definir las reglas de análisis, puede autorizar al miembro que puede realizar consultas a ejecutar consultas que coincidan con una regla de análisis específica admitida en AWS Clean Rooms.

AWS Clean Rooms admite los siguientes tipos de reglas de análisis:

- [Regla de análisis de agregación](#)
- [Regla de análisis de lista](#)
- [Regla de análisis personalizada en AWS Clean Rooms](#)

Puede haber una sola regla de análisis por tabla configurada. Puede configurar la regla de análisis en cualquier momento antes de asociar las tablas configuradas a la colaboración.

Important

Si utiliza la computación criptográfica para Clean Rooms y tiene tablas de datos cifrados en la colaboración, la regla de análisis que añada a la tabla configurada cifrada debe ser coherente con la forma en que se cifraron los datos. Por ejemplo, si ha cifrado los datos para SELECT (regla de análisis de agregación), no debe agregar la regla de análisis para JOIN (regla de análisis de lista).

Temas

- [Agregación de una regla de análisis de agregación a una tabla \(flujo guiado\)](#)
- [Agregación de una regla de análisis de lista a una tabla \(flujo guiado\)](#)
- [Agregación de una regla de análisis personalizada a una tabla \(flujo guiado\)](#)
- [Agregación de una regla de análisis a una tabla \(editor JSON\)](#)
- [Sigüientes pasos](#)

Agregación de una regla de análisis de agregación a una tabla (flujo guiado)

La regla de análisis de agregación permite realizar consultas que agreguen estadísticas sin revelar información de nivel de fila con las funciones COUNT, SUM, y AVG en dimensiones opcionales.

Este procedimiento describe el proceso de añadir una regla de análisis de agregación a la tabla configurada mediante la opción Flujo guiado de la consola de AWS Clean Rooms .

Note

Las tablas configuradas que utilizan fuentes de datos que no son de S3 solo admiten [reglas de análisis personalizadas](#).

Para añadir la regla de análisis de agregación a una tabla (flujo guiado)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
 2. En el panel de navegación izquierdo, elija Tables (Tablas).
 3. Seleccione la tabla configurada.
 4. En la página de detalles de la tabla configurada, seleccione Configurar regla de análisis.
 5. En el paso 1: elija el tipo de regla de análisis y, en el tipo de regla de análisis, elija la opción Agregación.
 6. En Método de creación, seleccione Flujo guiado y, a continuación, seleccione Siguiente.
 7. En Paso 2: Especificar controles de consulta, en Funciones de agregación:
 - a. Elija una Función de agregación en el menú desplegable:
 - COUNT
 - COUNT DISTINCT
 - SUM
 - SUM DISTINCT
 - AVG
 - b. Elija qué columnas se pueden usar en la Función de agregación en el menú desplegable Columnas.
 - c. (Opcional) Seleccione Añadir otra función para añadir otra función de agregación y asociar una o más columnas a esa función.
-  **Note**

Se requiere al menos una función de agregación.
- d. (Opcional) Seleccione Eliminar para eliminar una función de agregación.
 8. En Controles de combinación,
 - a. Seleccione una opción para Permitir que la tabla se consulte por sí misma:

Si selecciona...	Entonces...
No, solo se puede consultar el solapamiento	La tabla solo se puede consultar si está combinada con una tabla propiedad del miembro que puede realizar consultas.
Sí	La tabla se puede consultar por sí misma o cuando se combina con otras tablas.

- b. En Especificar columnas de combinación, seleccione las columnas que desea permitir que se utilicen en la instrucción INNER JOIN.

Esto es opcional si ha seleccionado Sí en el paso anterior.

- c. En Especificar operadores permitidos para la coincidencia, elija qué operadores (si los hay) se pueden usar para hacer coincidir varias columnas de combinación. Si selecciona dos o más columnas JOIN, será obligatorio el uso de uno de estos operadores.

Si selecciona...	Entonces...
AND	Puede incluir AND en las condiciones de coincidencia INNER JOIN para combinar una columna con otra entre tablas.
OR	Puede incluir OR en las condiciones de coincidencia INNER JOIN para combinar varias coincidencias de columna entre tablas. Este operador lógico es útil para obtener una tasa de coincidencia más alta.

9. (Opcional) En Controles de dimensión, en el menú desplegable Especificar columnas de dimensión, seleccione las columnas que desea permitir que se usen en la instrucción SELECT y las partes WHERE, GROUP BY y ORDER BY de la consulta.

 Note

No se pueden usar columnas de combinación ni de función de agregación como columnas de dimensión.

10. En Funciones escalares, seleccione una opción para ¿Qué funciones escalares desea permitir?

Si selecciona...	Entonces...
Todas ellas son compatibles actualmente con AWS Clean Rooms	Se permiten todas las funciones escalares actualmente admitidas por AWS Clean Rooms. • Puede seleccionar Ver lista para ver la lista completa de Funciones escalares admitidas en AWS Clean Rooms.
Una lista personalizada	Puede personalizar las funciones escalares que desea permitir. • Elija una o más opciones del menú desplegable Especificar las funciones escalares permitidas.
Ninguna	No se permite ninguna función escalar.

Para obtener más información, consulte [Funciones escalares](#).

11. Seleccione Siguiente.

12. En Paso 3: Especificar controles de resultados de consulta, en Restricciones de agregación:

- Seleccione cada Nombre de columna en la lista desplegable.
- Seleccione en la lista desplegable cada Número mínimo de valores diferenciados que se debe cumplir para que se devuelva cada fila de salida después de aplicarle la función COUNT DISTINCT.
- Seleccione Añadir restricción para añadir más restricciones de agregación.
- (Opcional) Seleccione Eliminar para eliminar una restricción de agregación.

13. En análisis adicionales aplicados a los resultados, seleccione una opción en función del objetivo.

Su objetivo	Opción recomendada
Permita solo consultas directas en esta tabla. Impida que se ejecuten análisis adicionales en los resultados de las consultas. La tabla solo se puede utilizar para consultas directas.	No permitido
Permita, pero no exija, tanto las consultas directas como los análisis adicionales en esta tabla.	Permitido
Exija que la tabla solo se pueda utilizar en consultas directas que se procesen con uno de los análisis adicionales necesarios. Las consultas directas de esta tabla se deben procesar más a fondo antes de poder devolverse.	Obligatorio

14. Elija Siguiente.

15. En Paso 4: Revisar y configurar, revise las selecciones que realizó en los pasos anteriores, edítelas si es necesario y, a continuación, seleccione Configurar regla de análisis.

Verá un mensaje de confirmación en el que se indica que ha configurado correctamente una regla de análisis de agregación en la tabla.

Agregación de una regla de análisis de lista a una tabla (flujo guiado)

La regla de análisis de lista permite realizar consultas que generen listas de nivel de fila del solapamiento entre la tabla asociada y una tabla del miembro que puede realizar consultas.

Este procedimiento describe el proceso de agregar la regla de análisis de listas a la tabla configurada mediante la opción de flujo guiado de la AWS Clean Rooms consola.

 Note

Las tablas configuradas que utilizan fuentes de datos que no son de S3 solo admiten [reglas de análisis personalizadas](#).

Para añadir una regla de análisis de lista a una tabla (flujo guiado)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Seleccione la tabla configurada.
4. En la página de detalles de la tabla configurada, seleccione Configurar regla de análisis.
5. En el paso 1: elija el tipo de regla de análisis y, en el tipo de regla de análisis, elija la opción Lista.
6. En Método de creación, seleccione Flujo guiado y, a continuación, seleccione Siguiente.
7. En Paso 2: Especificar controles de consulta, en Controles de combinación:
 - a. En Especificar columnas de combinación, seleccione las columnas que desea permitir que se utilicen en la instrucción INNER JOIN.
 - b. En Especificar operadores permitidos para la coincidencia, elija qué operadores (si los hay) se pueden usar para hacer coincidir varias columnas de combinación. Si selecciona dos o más columnas JOIN, será obligatorio el uso de uno de estos operadores.

Si selecciona...	Entonces...
AND	Puede incluir AND en las condiciones de coincidencia INNER JOIN para combinar una columna con otra entre tablas.
OR	Puede incluir OR en las condiciones de coincidencia INNER JOIN para combinar varias coincidencias de columna entre tablas. Este operador lógico es útil para obtener una tasa de coincidencia más alta.

8. (Opcional) En Controles de lista, en el menú desplegable Especificar columnas de la lista, seleccione las columnas que desea permitir que se utilicen en el resultado de la consulta (es decir, que se usen en la instrucción SELECT) o que se usen para filtrar los resultados (es decir, en la instrucción WHERE).
9. Elija Siguiente.
10. En el Paso 3: Especificar controles de los resultados de las consultas, para aplicar análisis adicionales a los resultados, seleccione una opción en función del objetivo.

Su objetivo	Opción recomendada
Permita solo consultas directas en esta tabla. Impida que se ejecuten análisis adicionales en los resultados de las consultas. La tabla solo se puede utilizar para consultas directas.	No permitido
Permita, pero no exija, tanto las consultas directas como los análisis adicionales en esta tabla.	Permitido
Exija que la tabla solo se pueda utilizar en consultas directas que se procesen con uno de los análisis adicionales necesarios. Las consultas directas de esta tabla se deben procesar más a fondo antes de poder devolverse.	Obligatorio

11. En Paso 4: Revisar y configurar, revise las selecciones que realizó en los pasos anteriores, edítelas si es necesario y, a continuación, seleccione Configurar regla de análisis.

Verá un mensaje de confirmación en el que se indica que ha configurado correctamente una regla de análisis de lista en la tabla.

Agregación de una regla de análisis personalizada a una tabla (flujo guiado)

La regla de análisis personalizada permite realizar consultas o PySpark trabajos de SQL personalizados en una tabla configurada. La regla de análisis personalizada es obligatoria si utiliza:

- [Plantillas de análisis](#) para permitir un conjunto específico de consultas o PySpark trabajos de SQL aprobados previamente o un conjunto específico de cuentas que puedan proporcionar consultas que utilicen sus datos.
- [AWS Clean Rooms Privacidad diferencial](#) para proteger a los usuarios de los intentos de identificación.
- Fuentes de datos que no son de S3, como Amazon Athena o Snowflake.

Este procedimiento describe el proceso de agregar la regla de análisis personalizada a la tabla configurada mediante la opción de flujo guiado de la consola. AWS Clean Rooms

Para añadir una regla de análisis personalizada a una tabla (flujo guiado)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Seleccione la tabla configurada.
4. En la página de detalles de la tabla configurada, seleccione Configurar regla de análisis.
5. En el paso 1: elija el tipo de regla de análisis y, en el tipo de regla de análisis, elija la opción Personalizada.
6. En Método de creación, seleccione Flujo guiado y, a continuación, seleccione Siguiente.
7. En el paso 2: Especificar los controles de análisis, en el caso de los controles de análisis directos, elija una opción en función de su objetivo.

Su objetivo	Acción recomendada
Revise cada análisis nuevo antes de permitir que se ejecute en esta tabla configurada	1. En Plantillas de análisis que se pueden ejecutar, elija Agregar plantilla de análisis. 2. Elija la plantilla de colaboración y de análisis adecuada en las listas desplegadas. 3. Elija Siguiente.
En esta tabla, permita que colaboradores específicos realicen cualquier análisis del tipo elegido sin necesidad de revisarlos	1. En Tipo de análisis,

Su objetivo	Acción recomendada
	- Elija Cualquier consulta para permitir cualquier consulta creada por el usuario Cuenta de AWS que especifique. - Elija Cualquier consulta para permitir que cualquier trabajo sea creado por la Cuenta de AWS persona que especifique. - En Cuentas de AWS Permitido crear cualquier análisis, elija Agregar Cuenta de AWS. - Introduzca un Cuenta de AWS identificador Cuenta de AWS o selecciónelo en la lista desplegable. (Opcional) Selecciona Añadir otro Cuenta de AWS para añadir otro Cuenta de AWS. - Elija Siguiente.

8. En el paso 3: especifique los controles de los resultados del análisis,
- Para los controles de resultados de Job, tenga en cuenta que no se admiten controles de resultados adicionales.
 - En los controles de resultados de la consulta, en el caso de las columnas no permitidas en la salida, elija las columnas que desee que se permitan en la salida de la consulta, en función de su objetivo.

Su objetivo	Acción recomendada
Permitir que se devuelvan todas las columnas en los resultados de las consultas	- Elige Ninguna - Continúe con los análisis adicionales aplicados a la salida.
Impedir que se devuelvan determinadas columnas en los resultados de las consultas	Elija Lista personalizada

Su objetivo	Acción recomendada
	2. En Especificar columnas no permitidas, elija las columnas que desee eliminar de los resultados de la consulta.

- c. Para aplicar análisis adicionales a la salida, elija si se pueden aplicar análisis adicionales a la salida de la consulta, en función de su objetivo.

Su objetivo	Opción recomendada
• Permita solo consultas directas en esta tabla. • Impida que se ejecuten análisis adicionales en los resultados de las consultas. • La tabla solo se puede utilizar para consultas directas.	No permitido
Permita, pero no exija, tanto las consultas directas como los análisis adicionales en esta tabla.	Permitir
• Exija que la tabla solo se pueda utilizar en consultas directas que se procesen con uno de los análisis adicionales necesarios. • Las consultas directas de esta tabla se deben procesar más a fondo antes de poder devolverse.	Obligatorio

- d. Elija Siguiente.

9. (Opcional) En el paso 4: Establecer la privacidad diferencial, determine si desea activar o desactivar la privacidad diferencial.

La privacidad diferencial es una técnica probada matemáticamente para proteger sus datos de los ataques de reidentificación.

 Note

AWS Clean Rooms La privacidad diferencial solo está disponible para las colaboraciones en las que los datos se almacenan en Amazon S3.

En el caso de la privacidad diferencial, elija si desea activar o desactivar la privacidad diferencial en función de su objetivo.

Su objetivo	Acción recomendada
• No necesitas protección contra los intentos de reidentificación • La tabla no tiene datos a nivel de usuario	1. Seleccione Desactivar. 2. Elija Siguiente.

Su objetivo	Acción recomendada
- Necesita protección contra los intentos de reidentificación - Su tabla tiene datos a nivel de usuario	- Elija Turn on. - Seleccione la columna del identificador de usuario que contiene el identificador único de sus usuarios, como la <code>user_id</code> columna cuya privacidad desea proteger. Para activar la privacidad diferencial para dos o más tablas de una colaboración, debe configurar la misma columna que la columna Identificador de usuario en ambas reglas de análisis para mantener una definición coherente de los usuarios en todas las tablas. Si la configuración es incorrecta, el miembro que puede realizar la consulta recibe un mensaje de error que indica que hay dos columnas entre las que elegir para calcular el número de contribuciones de los usuarios (por ejemplo, el número de impresiones de anuncios realizadas por un usuario) al ejecutar la consulta. - Elija Siguiente.

10. En Paso 5: Revisar y configurar, revise las selecciones que realizó en los pasos anteriores, edítelas si es necesario y, a continuación, elija Configurar regla de análisis.

Ve un mensaje de confirmación en el que se indica que ha configurado correctamente una regla de análisis personalizada en la tabla.

Agregación de una regla de análisis a una tabla (editor JSON)

El siguiente procedimiento muestra cómo añadir una regla de análisis a una tabla mediante la opción del editor JSON de la AWS Clean Rooms consola.

 Note

Las tablas configuradas que utilizan fuentes de datos que no son de S3 solo admiten [reglas de análisis personalizadas](#).

Agregación de una regla de análisis de agregación, de lista o personalizada a una tabla (editor JSON)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Seleccione la tabla configurada.
4. En la página de detalles de la tabla configurada, seleccione Configurar regla de análisis.
5. En el paso 1: elija el tipo de regla de análisis y, en el tipo de regla de análisis, elija la opción Agregación, Lista o Personalizada.
6. En Método de creación, seleccione Editor JSON y, a continuación, seleccione Siguiente.
7. En Paso 2: Especificar controles, puede optar por insertar una estructura de consulta (Insertar plantilla) o insertar un archivo (Importar desde archivo).

Sí selecciona...	Entonces...
Insertar plantilla	1. Especifique los parámetros de la regla de análisis seleccionada en Definición de la regla de análisis. 2. Puede pulsar Ctrl + barra espaciadora para habilitar la función de autocompletar. Para obtener más información acerca de los parámetros de la regla de análisis de agregación, consulte Regla de análisis de agregación: controles de consulta. Para obtener más información acerca de los parámetros de la regla de análisis de lista,

Si selecciona...	Entonces...
	consulte Regla de análisis de lista: controles de consulta .
Importar desde archivo	1. Seleccione el archivo JSON en su unidad local. 2. Elija Open. La Definición de la regla de análisis muestra la regla de análisis del archivo cargado.

8. Elija Siguiente.
9. En Paso 3: Revisar y configurar, revise las selecciones que realizó en los pasos anteriores, edítelas si es necesario y, a continuación, seleccione Configurar regla de análisis.

Recibe un mensaje de confirmación en el que se indica que ha configurado correctamente una regla de análisis en la tabla.

Siguientes pasos

Ahora que ha configurado una regla de análisis para la tabla configurada, ya puede:

- [Asociar una tabla configurada a una colaboración](#)
- [Consultar las tablas de datos](#) (como miembro que puede realizar consultas)

Asociar una tabla configurada a una colaboración

Una vez que haya creado una tabla configurada y le haya agregado una regla de análisis, puede asociarla a una colaboración y asignar AWS Clean Rooms una función de servicio para acceder a AWS Glue las tablas.

Note

Este rol de servicio tiene permisos relacionados con las tablas. AWS Clean Rooms solo puede asumir el rol de servicio para ejecutar las consultas permitidas en nombre del miembro que puede realizar consultas. Ningún miembro de la colaboración (salvo el propietario de los

datos) tiene acceso a las tablas subyacentes de la colaboración. El propietario de los datos puede activar la privacidad diferencial para que sus tablas estén disponibles para que otros miembros las consulten.

Presupuesto de acceso a los datos

Al asociar una tabla configurada, puede aplicar un presupuesto de acceso a los datos. Un presupuesto de acceso a los datos controla cuántas veces se puede usar una tabla para consultas, trabajos y canales de entrada de aprendizaje automático en una colaboración. Estos presupuestos ayudan a las organizaciones a administrar la utilización de los recursos y controlar los costos al limitar el uso de las tablas.

Cada vez que se utiliza una tabla en una consulta, un trabajo o un canal de entrada de aprendizaje automático, el presupuesto de esa tabla se reduce en uno. Cuando el presupuesto llega a cero, la tabla no se puede usar en consultas de SQL ni en trabajos de Pyspark ni como parte de los canales de entrada de aprendizaje automático derivados de la tabla.

Puedes establecer un presupuesto por período que se actualice periódicamente, un presupuesto de por vida para el uso general o ambos. De forma predeterminada, el uso de la tabla es ilimitado.

- **Presupuesto por período:** asignación renovable que limita la cantidad de veces que se puede usar esta tabla en un período de tiempo específico. Puede configurar el período como diario, semanal o mensual. Este presupuesto se puede configurar para que se actualice automáticamente de forma diaria, semanal o mensual.
- **Presupuesto total:** una asignación continua que limita la cantidad total de veces que se puede utilizar esta tabla.

Asocie una tabla configurada

En los siguientes temas se describe cómo asociar una tabla configurada y aplicar un presupuesto de acceso a datos a una colaboración mediante la AWS Clean Rooms consola.

Para obtener información sobre cómo asociar las tablas configuradas a la colaboración mediante la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

Paso 1: completar los requisitos previos

Para asociar una tabla configurada, debe cumplir los siguientes requisitos previos:

- Una AWS Glue tabla que apunta a una ubicación de carpeta de Amazon S3 (no a un solo archivo)
- Para AWS Glue tablas cifradas:
 - Un rol de servicio con permisos para usar AWS KMS claves para descifrar tablas AWS Glue
 - Para conjuntos AWS KMS de datos de Amazon S3 cifrados: el rol de servicio también debe tener permisos para usar la AWS KMS clave para descifrar los datos de Amazon S3

Para obtener información sobre la configuración del cifrado, consulte [Configuración del cifrado AWS Glue en](#) la AWS Glue Guía para desarrolladores.

Para comprobar la ubicación AWS Glue de la tabla:

1. Abra la AWS Glue consola en <https://console.aws.amazon.com/glue/>
2. Vea los detalles de la tabla y confirme que la ubicación apunta a una carpeta S3

Paso 2: Asociar una tabla configurada

Para asociar una tabla configurada

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. Elija el método para asociar la tabla:
 - a. Desde la página de detalles de la tabla configurada:
 - i. En el panel de navegación izquierdo, elija Tables (Tablas).
 - ii. Seleccione la tabla configurada.
 - iii. En la página de detalles de la tabla configurada, seleccione Asociar a colaboración.
 - iv. En el cuadro de diálogo Asociar tabla a colaboración, seleccione la Colaboración en la lista desplegable.
 - b. Desde la página de detalles de la colaboración:
 - i. En el panel de navegación izquierdo, elija Colaboraciones.
 - ii. Seleccione la colaboración.
 - iii. En la pestaña Tablas, seleccione Asociar tabla.
3. En la página de la tabla Asociar, realice una de las siguientes acciones:

- Elija una tabla configurada existente: elija el nombre de la tabla configurada que desee asociar a la colaboración en la lista desplegable.
 - Configurar una tabla nueva: elija Configurar nueva tabla y siga las instrucciones de la página Configurar nueva tabla.
 - Ver el esquema y la regla de análisis de la tabla configurada: activa Ver esquema y regla de análisis.
4. En Detalles de asociación de tablas,
- a. Introduzca un Nombre para la tabla asociada.
- Puede usar el nombre predeterminado o cambiar el nombre de esta tabla.
- b. (Opcional) Ingrese una Descripción de la tabla.
- La descripción ayuda a escribir las consultas.
5. Especifique los permisos de Acceso a servicios seleccionando Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.

 Note

Si va a asociar una tabla configurada respaldada por Amazon Athena, elija un nombre de rol de servicio existente en la lista desplegable. Asegúrese de que la función de servicio tenga permisos de IAM y, si es necesario, de Lake Formation para el conjunto de datos.

Si selecciona...	Entonces...
Crear y usar un nuevo rol de servicio	• AWS Clean Rooms crea un rol de servicio con la política requerida para esta tabla. • El Nombre del rol de servicio predeterminado es <code>cleanrooms-<timestamp></code> • Debe tener permisos para crear roles y adjuntar políticas. • Si los datos de entrada están cifrados, puede seleccionar Estos datos están

Si selecciona...	Entonces...
	cifrados con una clave de KMS y, a continuación, ingresar una AWS KMS key que se utilizará para descifrar los datos ingresados.
Usar un rol de servicio existente	1. Seleccione un Nombre de rol de servicio existente en la lista desplegable. Si tiene permisos de listas de roles, se mostrará la lista de roles. Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar. 2. Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM. Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible. De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios. 3. (Opcional) Seleccione la casilla de verificación Agregar una política preconfigurada con los permisos necesarios para este rol para adjuntar los permisos necesarios al rol. Debe tener permisos para modificar roles y crear políticas.

 Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).
- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes para ello AWS Clean Rooms. Debe agregar la política de rol antes de continuar.
- Si no puede modificar la política de rol, recibirá un mensaje de error que indica que AWS Clean Rooms no ha podido encontrar la política del rol de servicio.

6. Si desea habilitar las etiquetas de asociación de tablas configuradas para el recurso de asociación de tablas configurado, elija Agregar nueva etiqueta y, a continuación, introduzca el par clave y valor.
7. Elija Siguiente.
8. En la página Configurar una regla de análisis de colaboración, elija una de las siguientes opciones:
 - Sí, cree una regla de análisis de colaboración ahora: asocie su tabla a esta colaboración y cree una regla de análisis de colaboración
 - No, crearé una regla de análisis de colaboración más adelante. Asocie su tabla únicamente a esta colaboración. Puede crear una regla de análisis de colaboración más adelante.
9. Si selecciona Sí, cree ahora una regla de análisis de colaboración. Para la entrega de resultados, seleccione en la lista desplegable los miembros que pueden recibir los resultados de la consulta.
10. Elija Siguiente.
11. En la página Añadir presupuesto de acceso a datos, en la configuración del presupuesto de acceso a datos, elija una de las siguientes opciones:
 - Sí, añada un presupuesto de acceso a datos ahora: asocie su tabla a esta colaboración y añada un presupuesto de acceso a datos. Puede seleccionar un presupuesto para el período, un presupuesto para toda la vida o ambos.
 - No, añadiré un presupuesto de acceso a los datos más adelante. Asocia tu tabla únicamente a esta colaboración. Puede añadir un presupuesto de acceso a los datos más adelante.

Si seleccionas No, añadiré un presupuesto de acceso a datos más adelante. Continúa con el paso 15.

12. Si elige Sí, agregue un presupuesto de acceso a datos ahora, elija una de las siguientes configuraciones de presupuesto:

Solo presupuesto por período	Solo presupuesto vitalicio	Presupuestos tanto por período como vitalicio
- Deje seleccionada la opción Añadir presupuesto por período. - Introduzca un importe presupuestario por período comprendido entre 1 y 1 000 000 000. - En Período, elija Diario, Semanal o Mensual. (Opcional) Seleccione e Actualizar automáticamente el presupuesto semanalmente para renovar la asignación. - Desactive Añadir presupuesto vitalicio.	- Desactive Agregar presupuesto por período. - Seleccione Añadir presupuesto vitalicio. - Introduce un importe presupuestario de por vida comprendido entre 1 y 1 000 000 000.	- Deje seleccionada la opción Añadir presupuesto por período. - Introduzca un importe presupuestario por período comprendido entre 1 y 1 000 000 000. - En Período, elija Diario, Semanal o Mensual. - Deje seleccionada la opción Actualizar automáticamente el presupuesto semanalmente. - Selecciona Añadir presupuesto vitalicio. - Introduce un importe presupuestario de por vida comprendido entre 1 y 1 000 000 000.

13. Revise sus selecciones en el resumen del presupuesto de acceso a los datos.

Example Ejemplo

Por ejemplo, si ha elegido un importe presupuestario por período de 1 000 000€, ha establecido el período en Semanal, ha dejado seleccionada la casilla Actualizar automáticamente el presupuesto semanalmente y ha establecido el presupuesto vitalicio en 1 000 000 000€, el

resumen del presupuesto de Access mostrará el siguiente mensaje: Cada semana, esta tabla se puede utilizar hasta 1000 veces para ejecutar consultas o tareas. Este presupuesto está configurado para actualizarse automáticamente todos los domingos a las 00:00 UTC y seguirá actualizándose hasta que esta tabla alcance su presupuesto de por vida de 1 000 000 de usos.

14. (Opcional) Si desea habilitar las etiquetas presupuestarias de acceso a los datos para el recurso presupuestario de acceso, seleccione Añadir nueva etiqueta e introduzca un par de clave y valor.
15. Elija Siguiente.
16. Revise la información de la página de revisión y creación.
 - a. Si necesitas editar alguna sección, selecciona Editar.
 - b. Edita las configuraciones y, a continuación, selecciona Siguiente.
17. Seleccione Asociar tabla.

Paso 3: Próximos pasos

Ahora que ha asociado la tabla de datos configurada a la colaboración, ya puede:

- [Agregación de una regla de análisis de colaboración](#) a una tabla configurada
- [Editar la colaboración](#) (si es el creador de la colaboración)
- [Consultar las tablas de datos](#) (como miembro que puede realizar consultas)

Configuración de un presupuesto de acceso de datos

Un colaborador puede ver, añadir, editar y eliminar datos sin presupuesto para establecer un límite en el número de veces que se puede utilizar una tabla en un flujo de trabajo. Utilice estos presupuestos para gestionar los datos y los costes.

Cada vez que se consulta una tabla o se ejecuta un trabajo mediante un canal de entrada de aprendizaje automático derivado de una tabla, el presupuesto de esa tabla se reduce en uno. Cuando el presupuesto llega a cero, no se puede consultar la tabla y los trabajos de aprendizaje automático no se pueden ejecutar con los canales de entrada de aprendizaje automático derivados de la tabla.

Puedes establecer un presupuesto por período que se actualice periódicamente, un presupuesto vitalicio para el uso general o ambos. De forma predeterminada, el uso de la tabla es ilimitado.

- Presupuesto por período: asignación renovable que limita la cantidad de veces que se puede usar esta tabla en un período de tiempo específico. Puede configurar el período como diario, semanal o mensual. Este presupuesto se puede configurar para que se actualice automáticamente de forma diaria, semanal o mensual.
- Presupuesto total: una asignación continua que limita la cantidad total de veces que se puede utilizar esta tabla.

Temas

- [Visualización de un presupuesto de acceso a datos](#)
- [Añadir un presupuesto de acceso a datos a una tabla asociada existente](#)
- [Edición de un presupuesto de acceso a datos](#)
- [Eliminar un presupuesto de acceso a datos](#)

Visualización de un presupuesto de acceso a datos

Puede ver el presupuesto de acceso a los datos desde la pestaña Tablas o desde la página de detalles de la tabla.

Para ver un presupuesto de acceso a datos

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de [AWS](#) y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Realice una de las siguientes acciones:
 - En la columna Presupuesto de acceso a los datos restantes, seleccione el presupuesto para ver los detalles.
 - Elija una tabla y, en la página de detalles de la tabla, desplácese hacia abajo para ver la sección de detalles del presupuesto de acceso a datos.

Añadir un presupuesto de acceso a datos a una tabla asociada existente

Como miembro de la colaboración, puede añadir un presupuesto de acceso a datos a una tabla asociada existente.

Para añadir un presupuesto de acceso a datos a una tabla asociada existente

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Seleccione el botón de opción situado junto a la tabla a la que desee añadir un presupuesto de acceso a los datos.
6. En la lista desplegable Acciones, en Presupuesto de acceso a datos, selecciona Añadir (si aún no hay un presupuesto).
7. Elige una de las siguientes configuraciones de presupuesto:

Solo presupuesto por período	Solo presupuesto vitalicio	Presupuestos tanto por período como vitalicio
1. Deje seleccionada la opción Añadir presupuesto por período. 2. Introduzca un importe presupuestario por período comprendido entre 1 y 1 000 000 000. 3. En Período, elija Diario, Semanal o Mensual. 4. (Opcional) Seleccione Actualizar automáticamente el presupuesto semanalmente para renovar la asignación.	1. Desactive Agregar presupuesto por período. 2. Seleccione Añadir presupuesto vitalicio. 3. Introduce un importe presupuestario de por vida comprendido entre 1 y 1 000 000 000.	1. Deje seleccionada la opción Añadir presupuesto por período. 2. Introduzca un importe presupuestario por período comprendido entre 1 y 1 000 000 000. 3. En Período, elija Diario, Semanal o Mensual. 4. Deje seleccionada la opción Actualizar automáticamente el presupuesto semanalmente.

Solo presupuesto por período	Solo presupuesto vitalicio	Presupuestos tanto por período como vitalicio
5. Desactive Añadir presupuesto vitalicio.		5. Selecciona Añadir presupuesto vitalicio. 6. Introduce un importe presupuestario de por vida comprendido entre 1 y 1 000 000 000.

8. Revise sus selecciones en el resumen del presupuesto de acceso a datos.

9. Example Ejemplo

Por ejemplo, si ha elegido un importe presupuestario por período de 1 000 000€, ha establecido el período en Semanal, ha dejado seleccionada la casilla Actualizar automáticamente el presupuesto semanalmente y ha establecido el presupuesto vitalicio en 1 000 000 000€, el resumen del presupuesto de Access mostrará el siguiente mensaje: Cada semana, esta tabla se puede utilizar hasta 1000 veces para ejecutar consultas o tareas. Este presupuesto está configurado para actualizarse automáticamente todos los domingos a las 00:00 UTC y seguirá actualizándose hasta que esta tabla alcance su presupuesto de por vida de 1 000 000 de usos.

10. (Opcional) Si desea habilitar las etiquetas presupuestarias de acceso a los datos para el recurso presupuestario de acceso, seleccione Añadir nueva etiqueta e introduzca un par de clave y valor.

11. Seleccione Añadir presupuesto de acceso a datos.

Edición de un presupuesto de acceso a datos

Como miembro de la colaboración, puede editar el presupuesto de acceso a los datos. Al editar un presupuesto de acceso a datos, se restablece el saldo presupuestario actual.

Puede editar un presupuesto de acceso a datos desde la pestaña Tablas o desde la página de detalles de la tabla.

Tables tab

Para editar un presupuesto de acceso a datos desde la pestaña Tablas

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Seleccione el botón de opción situado junto a la tabla que desee editar.
6. En la lista desplegable Acciones, en Presupuesto de acceso a datos, selecciona Editar.
7. En la página Editar el presupuesto de acceso a los datos, actualiza la información del presupuesto por período o del presupuesto total.
8. Consulta el resumen del presupuesto de acceso a los datos para comprobar que las modificaciones que has realizado son correctas.
9. Seleccione Save changes (Guardar cambios).

Table details page

Para editar un presupuesto de acceso a datos desde la página de detalles de la tabla

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Elija una tabla.
6. En la página de detalles de la tabla, desplácese hacia abajo hasta la sección de detalles del presupuesto de acceso a datos.
7. En la lista desplegable de Acciones, elija Editar.
8. En la página Editar el presupuesto de acceso a los datos, actualice la información del presupuesto por período o del presupuesto total.
9. Seleccione Save changes (Guardar cambios).

Eliminar un presupuesto de acceso a datos

Puede eliminar un presupuesto de acceso a datos desde la pestaña Tablas o desde la página de detalles de la tabla.

Tables tab

Para eliminar un presupuesto de acceso a datos de la pestaña Tablas

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Seleccione el botón de opción situado junto a la tabla que desee eliminar.
6. En la lista desplegable Acciones, en Presupuesto de acceso a datos, selecciona Eliminar.



Important

No puedes deshacer esta acción y tu presupuesto de acceso a datos pasará a ser ilimitado.

7. Si estás seguro de que quieres eliminar el presupuesto de acceso a los datos, selecciona Eliminar.

Table details page

Para eliminar un presupuesto de acceso a datos de la página de detalles de la tabla

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Elija una tabla.

6. En la página de detalles de la tabla, desplácese hacia abajo hasta la sección de detalles del presupuesto de acceso a datos.
7. En el menú desplegable Acciones, elija Eliminar.

 Important

No puedes deshacer esta acción y tu presupuesto de acceso a datos pasará a ser ilimitado.

8. Si estás seguro de que quieres eliminar el presupuesto de acceso a los datos, selecciona Eliminar.

Agregación de una regla de análisis de colaboración a una tabla configurada

La regla de análisis de colaboración le permite especificar los controles que son específicos de esta colaboración. Estos controles funcionan junto con la regla de análisis de tablas configurada para determinar cómo se puede analizar esta tabla en esta colaboración.

Se agrega una regla de análisis de colaboración a una tabla configurada después de haber [creado una tabla configurada](#), haber [agregado una regla de análisis](#) y haberla [asociado a una colaboración](#). Debe agregar una regla de análisis de colaboración si la tabla está configurada para admitir el análisis directo o para permitir un análisis adicional.

- Análisis directo: la tabla se puede utilizar en consultas que la analicen directamente. Por ejemplo, en una consulta que genera un análisis de medición agregado o una lista de identificadores para su activación.
- Análisis adicional: la tabla también se puede utilizar como entrada en análisis adicionales, además de en consultas que la analizan directamente. Por ejemplo, la tabla se puede usar en una consulta que sea una semilla para un modelo de ML similar o un canal de entrada de ML para un modelo de ML personalizado.

Agregación de la regla de análisis de agregación a una tabla

1. Inicia sesión en Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).

2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Tablas, en Tablas asociadas por usted, consulte la tabla configurada que ha asociado a la colaboración.
 - Si el estado del análisis directo o el estado del análisis adicional tienen el estado Preparado, significa que la tabla está lista para la consulta.
 - Si el estado del análisis directo o el estado del análisis adicional tienen el estado No preparado, seleccione el estado y, a continuación, elija Configurar en el cuadro de diálogo.
5. En la página Configurar regla de análisis de colaboración, expanda Ver regla de análisis de tabla configurada para ver los detalles.
6. Para Análisis adicionales permitidos, elija la opción en función de su objetivo.

Su objetivo	Opción recomendada
Permita los análisis adicionales en la tabla.	Cualquiera
Permita que un miembro específico solo incluya análisis adicionales en la tabla.	Cualquiera realizado por miembros específicos
Permita solo análisis específicos en la tabla.	Lista personalizada

7. Para la entrega de resultados, especifique quién puede recibir los resultados de los miembros autorizados a recibirlos en el menú desplegable de resultados de la consulta.
8. Elija Configurar regla de análisis.

Configuración de la política de privacidad diferencial (opcional)

Note

AWS Clean Rooms La privacidad diferencial solo está disponible para las colaboraciones en las que los datos se almacenan en Amazon S3.

Este procedimiento describe el proceso de configuración de la política de privacidad diferencial en una colaboración mediante la opción de flujo guiado de la AWS Clean Rooms consola. Este paso solo se ejecuta una vez para todas las tablas con protección de privacidad diferencial.

Para configurar los ajustes de privacidad diferencial (flujo guiado)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Tablas de la página de colaboración, elija Configuración de la política de privacidad diferencial.
5. En la página Configuración de la política de privacidad diferencial, elija valores para las siguientes propiedades:
 - Presupuesto de privacidad
 - Actualizar mensualmente el presupuesto de privacidad
 - Ruido agregado por consulta

Puede utilizar los valores predeterminados o introducir valores personalizados que admitan su caso de uso específico. Tras elegir los valores de Presupuesto de privacidad y Ruido añadido por consulta, puede obtener una vista previa de la utilidad resultante en lo que respecta al número de agregaciones posibles en todas las consultas de sus datos.

6. Elija Configurar.

Verá un mensaje de confirmación en el que se indica que ha configurado correctamente la política de privacidad diferencial para la colaboración.

Ahora que ha configurado la privacidad diferencial, puede hacer lo siguiente:

- [Consultar las tablas de datos](#) (como miembro que puede realizar consultas)
- [Colaboraciones](#) (si es el creador de la colaboración)

Visualización de los registros de uso de la privacidad diferencial

Como miembro de una colaboración que protege los datos con privacidad diferencial, después de haber creado una colaboración con privacidad diferencial, puede supervisar el uso del presupuesto de privacidad.

Para ver cuántas agregaciones se ejecutaron y qué parte del presupuesto de privacidad se utilizó

1. [Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en / cleanrooms. https://console.aws.amazon.com](https://console.aws.amazon.com/cleanrooms)
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Seleccione Ver registros de uso (texto azul).
6. Consulte los detalles de uso, incluido el presupuesto de privacidad y cuánta utilidad se ha proporcionado.

Edición de una política de privacidad diferencial

Después de configurar la política de privacidad diferencial, puede actualizarla en cualquier momento para que refleje mejor sus necesidades de privacidad.

Para editar la política de privacidad diferencial

1. [Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en / cleanrooms. https://console.aws.amazon.com](https://console.aws.amazon.com/cleanrooms)
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Tablas de la página de colaboración, en Tablas asociadas por usted, elija Editar.
5. En la página Editar privacidad diferencial, elija nuevos valores para las siguientes propiedades:
 - Presupuesto de privacidad: mueva la barra del control deslizante para aumentar o disminuir el presupuesto en cualquier momento de una colaboración. No puede reducir el presupuesto después de que el miembro que puede realizar la consulta haya empezado a consultar sus datos. Si se aumenta el presupuesto de privacidad, AWS Clean Rooms seguirá utilizando el

presupuesto existente hasta que se agote por completo antes de utilizar el presupuesto de privacidad recién agregado.

- Ruido agregado por consulta: mueva la barra de control deslizante para aumentar o disminuir el ruido agregado por consulta en cualquier momento durante una colaboración.

 Note

Puede elegir Ejemplos interactivos para ver cómo los distintos valores del Presupuesto de privacidad y Ruido agregado por consulta afectan al número de funciones agregadas que puede ejecutar.

No se puede cambiar el valor de la Actualización del presupuesto de privacidad. Para cambiar su selección, debe eliminar la política de privacidad diferencial y crear una nueva.

6. Seleccione Save changes (Guardar cambios).

Aparece un mensaje de confirmación en el que se indica que ha editado correctamente la política de privacidad diferencial.

Eliminación de una política de privacidad diferencial

Puede eliminar la política de privacidad diferencial desde la pestaña Tablas de una colaboración.

Eliminación de la política de privacidad diferencial

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Tablas de la página de colaboración, junto a la Política de privacidad diferencial, seleccione Eliminar.
5. Si tiene la certeza de que desea eliminar la política de privacidad diferencial, seleccione Eliminar.

Tras eliminar una política de privacidad diferencial, no podrá acceder a los registros de uso del presupuesto de privacidad de esa política. Las tablas con la privacidad diferencial activada no se pueden consultar si se elimina la política de privacidad diferencial.

Visualización de los parámetros de privacidad diferencial calculados

Los usuarios con experiencia en privacidad diferencial pueden ver los parámetros de privacidad diferencial calculados en la pestaña Análisis de una colaboración.

Visualización de los parámetros de privacidad diferencial calculados

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Análisis, en la sección Resultados, seleccione Ver los parámetros de privacidad diferencial calculados.

En la tabla Parámetros de privacidad diferencial calculados, puede ver los valores de confidencialidad de las funciones agregadas, que se definen como la cantidad máxima en la que el resultado de una función puede cambiar si se añaden, eliminan o modifican los registros de un solo usuario. La lista incluye los siguientes parámetros de privacidad diferencial:

- El límite de contribución del usuario (UCL) es el número máximo de filas que aporta un usuario en una consulta SQL. Por ejemplo, si quieres contar el número total de impresiones coincidentes en una campaña específica, en la que cada usuario puede tener varias impresiones, AWS Clean Rooms Differential Privacy debe limitar el número de impresiones de un solo usuario para garantizar que el cálculo de la privacidad diferencial sea preciso. En otras palabras, si algún usuario tiene más impresiones que el límite, toma AWS Clean Rooms automáticamente una muestra aleatoria uniforme de las impresiones de ese usuario según el valor de UCL calculado y excluye las impresiones restantes de ese usuario al ejecutar la consulta. El valor de la UCL es igual a 1 si se cuenta el número de usuarios únicos. Esto se debe a que la adición, eliminación o modificación de un solo usuario puede cambiar el recuento de usuarios distintos en 1 como máximo.
- El valor mínimo es el límite inferior de una expresión que se utiliza en una función de agregación, como `sum()`. Por ejemplo, si la expresión es una columna denominada `purchase_value`, el valor mínimo es el límite inferior de la columna.

- El valor máximo es el límite superior de una expresión que se utiliza en una función de agregación, como `sum()`. Por ejemplo, si la expresión es una columna denominada `purchase_value`, el valor máximo es el límite superior de la columna.

En la tabla Parámetros de privacidad diferencial calculados, puede utilizar estos parámetros para comprender mejor la cantidad total de ruido en los resultados de las consultas. Por ejemplo, cuando el ruido agregado por consulta es de 30 usuarios y se ejecuta una `COUNT DISTINCT (user_id)` consulta, AWS Clean Rooms Differential Privacy añade un ruido aleatorio que se sitúa entre -30 y 30, con una alta probabilidad porque la sensibilidad de `COUNT DISTINCT` es 1. En el caso de una consulta `COUNT` con la misma configuración, la privacidad diferencial de AWS Clean Rooms añade ruido estadístico que se escala según el límite de contribución del usuario, ya que un solo usuario podría añadir varias filas al resultado de la consulta. En el caso de `SUM` consultas como `SUM (purchase_value)` las que tienen valores positivos en todas las columnas, el ruido total se escala en función del límite de contribución del usuario multiplicado por el valor máximo. AWS Clean Rooms Differential Privacy calcula automáticamente los parámetros de sensibilidad para añadir ruido en tiempo de ejecución de la consulta y agota el presupuesto de privacidad. Es necesario reducir el presupuesto de privacidad porque los parámetros de sensibilidad dependen de los datos.

Ver tablas y reglas de análisis

Para ver las tablas asociadas a la colaboración y a las reglas de análisis

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. Seleccione una de las siguientes opciones:
 - a. Para ver las tablas asociadas a la colaboración, en Tablas asociadas por usted, elija una tabla (texto azul).
 - b. Para ver otras tablas asociadas a la colaboración, en Tablas asociadas por colaboradores, elija una tabla (texto azul).
6. Vea los detalles de la tabla y las reglas de análisis en la página de detalles de la tabla.

Edición de una tabla configurada

Requisitos previos:

- Y Cuenta de AWS con acceso a AWS Clean Rooms

En las siguientes secciones se explica cómo editar el nombre, la descripción y los detalles de configuración de las tablas de las fuentes de datos de Amazon S3, Amazon Athena y Snowflake.

[Para obtener información sobre cómo editar una tabla configurada mediante la AWS SDKs, consulte la referencia de la AWS Clean Rooms API.](#)

Para editar una tabla configurada

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Elija la tabla configurada que creó.
4. En la página de detalles de la tabla configurada, seleccione Editar.
5. Edite la configuración.
6. Seleccione Save changes (Guardar cambios).

Editar etiquetas de tablas configuradas

Como miembro de la colaboración, después de crear una tabla configurada, puede administrar las etiquetas del recurso de tabla configurada en la pestaña Tablas configuradas.

Para editar las etiquetas de tablas configuradas

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Elija la tabla configurada que creó.
4. En la página de detalles de la tabla configurada, desplácese hacia abajo hasta la sección Etiquetas.
5. Elija Administrar etiquetas.

6. En la página Administrar etiquetas, puede hacer lo siguiente:

- Para eliminar una etiqueta, elija Eliminar.
- Para añadir una etiqueta, elija Añadir nueva etiqueta.
- Para guardar los cambios, elija Guardar cambios.

Edición de la regla de análisis de tablas configurada

Para editar la regla de análisis de tablas configuradas

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Elija la tabla configurada que creó.
4. En la página de detalles de la tabla configurada, desplácese hacia abajo hasta la sección Regla de análisis de agregación, Regla de análisis de lista o Regla de análisis personalizada (su elección dependerá del tipo de regla de análisis que haya elegido para la tabla configurada).
5. Elija Editar.
6. En la página Editar regla de análisis, puede:
 - Modificar la Definición de la regla de análisis de la siguiente manera:
 - Modificando el editor JSON.
 - Seleccione Importar desde archivo para cargar una nueva definición de la regla de análisis.
 - Para obtener una vista previa de lo que verán los miembros de una colaboración, seleccione una de las siguientes opciones:
 - Vista de tabla
 - JSON
 - Consulta de ejemplo
7. Elija Guardar cambios para guardar los cambios.

Eliminar la regla de análisis de tablas configurada

Warning

Esta acción no se puede deshacer y afecta a todos los recursos relacionados.

Para eliminar la regla de análisis de tablas configuradas

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. Elija la tabla configurada que creó.
4. En la página de detalles de la tabla configurada, desplácese hacia abajo hasta la sección Regla de análisis de agregación, Regla de análisis de lista o Regla de análisis personalizada (su elección dependerá del tipo de regla de análisis que haya elegido para la tabla configurada).
5. Elija Eliminar.
6. Si tiene la certeza de que desea eliminar la regla de análisis, elija Eliminar.

Columnas no permitidas de la tabla configurada

La configuración de las columnas de salida no permitidas es un control de la regla de análisis AWS Clean Rooms personalizada que permite definir la lista de columnas (si las hay) que no se pueden proyectar en el resultado de la consulta. Las columnas a las que se hace referencia en esta lista se consideran “columnas de salida no permitidas”. Esto significa que cualquier referencia a dicha columna mediante transformación, alias u otros medios puede no estar presente en la SELECCIÓN (proyección) final de la consulta.

Aunque esta funcionalidad prohíbe que las columnas se proyecten directamente en la salida, no impide por completo que los valores subyacentes se deduzcan indirectamente mediante otros mecanismos. Estas columnas se pueden seguir utilizando en una cláusula de proyección (por ejemplo, en una subconsulta o en una expresión de tabla común [CTE]), siempre y cuando no se haga referencia a ellas en la proyección final.

La configuración de columnas de salida no permitidas le ofrece la flexibilidad de aplicar y codificar el control de la tabla, junto con revisiones de nivel de plantilla de análisis en función de los casos de uso y los requisitos de privacidad correspondientes.

Para obtener más información acerca de cómo definir esta configuración, consulte [Agregación de una regla de análisis personalizada a una tabla \(flujo guiado\)](#).

Ejemplos

Los siguientes ejemplos muestran cómo se aplica el control de columnas de salida no permitidas.

- El miembro A colabora con el miembro B.
- El miembro B es quien puede ejecutar consultas.
- El miembro A define a los usuarios de una tabla con las columnas edad, sexo, correo electrónico y nombre. Las columnas de edad y nombre no son columnas de salida permitidas.
- El miembro B define una tabla mascotas con un conjunto similar de columnas edad, género y owner_name. Sin embargo, no establecen ninguna restricción en las columnas de salida, lo que significa que todas las columnas de la tabla se pueden proyectar libremente en la consulta.

Si el miembro B ejecuta la siguiente consulta, se bloquea porque las columnas de salida no permitidas no se pueden proyectar directamente:

```
SELECT
  age
FROM
  users
```

Si el miembro B ejecuta la siguiente consulta, se bloquea porque las columnas de salida no permitidas no se pueden proyectar implícitamente a través de la estrella del proyecto:

```
SELECT
  *
FROM
  users
```

Si el miembro B ejecuta la siguiente consulta, se bloquea porque las transformaciones de las columnas de salida no permitidas no se pueden proyectar:

```
SELECT
  COUNT(age)
FROM
  users
```

Si el miembro B ejecuta la siguiente consulta, se bloquea porque no se puede hacer referencia a las columnas de salida no permitidas en la proyección final con un alias:

```
SELECT
  count_age
FROM
  (SELECT COUNT(age) AS count_age FROM users)
```

Si el miembro B ejecuta la siguiente consulta, se bloquea porque las columnas restringidas transformadas se proyectan en la salida:

```
SELECT
  CONCAT(name, email)
FROM
  users
```

Si el miembro B ejecuta la siguiente consulta, se bloquea porque no se puede hacer referencia a las columnas de salida no permitidas en CTE en la proyección final:

```
WITH cte AS (
  SELECT
    age AS age_alias
  FROM
    users
)
SELECT age_alias FROM cte
```

Si el miembro B ejecuta la siguiente consulta, se bloquea porque no se pueden usar las columnas de salida no permitidas como claves de clasificación o partición en la proyección final:

```
SELECT
```

```
LISTAGG(gender) WITHIN GROUP (ORDER BY age) OVER (PARTITION BY age)
FROM
  users
```

Si el miembro B ejecuta la siguiente consulta, se realiza correctamente porque las columnas que forman parte de las columnas de salida no permitidas se pueden seguir utilizando en otros constructos de la consulta, como en las cláusulas de unión o filtro.

```
SELECT
  u.name,
  p.gender,
  p.age
FROM
  users AS u
JOIN
  pets AS p
ON
  u.name = p.owner_name
```

En el mismo escenario, el miembro B también puede usar la columna de nombres en los usuarios como filtro o clave de clasificación:

```
SELECT
  u.email,
  u.gender
FROM
  users AS u
WHERE
  u.name = 'Mike'
ORDER BY
  u.name
```

Además, las columnas de salida no permitidas por los usuarios se pueden usar en proyecciones intermedias, como subconsultas y CTEs, por ejemplo:

```
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)
```

Editar asociaciones de tablas configuradas

Como miembro de una colaboración, puede editar las asociaciones de tablas configuradas que ha creado.

Para editar asociaciones de tablas configuradas

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. En Tablas asociadas por usted, elija una tabla.
6. En la página de detalles de la tabla, desplácese hacia abajo para ver los Detalles de asociación de tablas.
7. Elija Editar.
8. En la página Editar asociaciones de tablas configuradas, actualice la Descripción o la Información de acceso al servicio.
9. Seleccione Save changes (Guardar cambios).

Disociar tablas configuradas

Como miembro de la colaboración, puede disociar una tabla configurada de la colaboración. Esta acción impide que el miembro que puede realizar consultas consulte la tabla.

Para disociar una tabla configurada

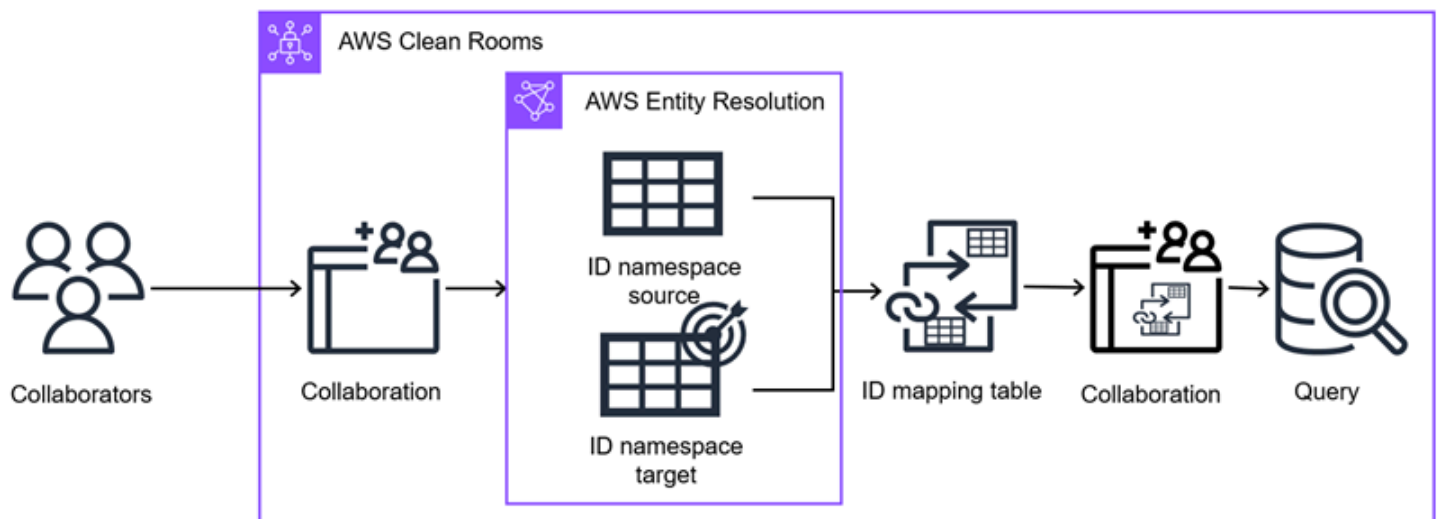
1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Seleccione la pestaña Tablas.
5. En Tablas asociadas por usted, elija el botón de opción situado junto a la tabla que desea disociar.
6. Elija Disociar.
7. En el cuadro de diálogo, confirme la decisión de disociar la tabla configurada, y seleccione Disociar para impedir que el miembro que puede realizar consultas consulte la tabla.

AWS Entity Resolution in AWS Clean Rooms

Con AWS Entity Resolution in AWS Clean Rooms, puede traducir los datos de una fuente a un destino, rellenar una tabla de mapeo de ID con los datos traducidos y consultarlos.

En primer lugar, puedes crear una colaboración en AWS Clean Rooms y añadir la Cuentas de AWS que quieras invitar o unirse a una colaboración a la que estés invitado creando una membresía. A continuación, realice una asignación de ID en dos tablas de datos. Para ello, asocie un origen de espacio de nombres de ID existente o cree uno nuevo en AWS Entity Resolution. El otro miembro de la colaboración asocia un destino de espacio de nombres de ID existente o crea un destino de espacio de nombres de ID nuevo. A continuación, debe crear y rellenar una tabla de asignación de ID a partir de los dos espacios de nombres de ID asociados. Por último, el miembro que puede realizar la consulta ejecuta una consulta en las dos tablas de datos mediante la unión en la tabla de asignación de ID.

En el siguiente diagrama se resume cómo trabajar con ella AWS Entity Resolution . AWS Clean Rooms



Note

El proveedor de servicios de transcodificación compatible actualmente es el siguiente
Regiones de AWS: EE.UU. Este (Norte de Virginia), EE.UU. Este (Ohio) y EE.UU. Oeste (Oregón). LiveRamp

- [Espacios de nombres de ID en AWS Clean Rooms](#)
- [Tablas de mapeo de ID en AWS Clean Rooms](#)

Espacios de nombres de ID en AWS Clean Rooms

Un espacio de nombres de ID es un encapsulador en torno a la tabla de identidades que le permite proporcionar metadatos que explican el conjunto de datos y cómo usarlo en un flujo de trabajo de asignación de ID. Un flujo de trabajo de asignación de ID es un trabajo de procesamiento de datos que asigna datos de un origen de datos de entrada a un destino de datos de entrada en función del método de asignación de ID especificado. Produce una tabla de asignación de ID.

Hay dos tipos de espacios de nombres de ID: origen y destino. El origen contiene configuraciones para los datos de origen que se procesarán en un flujo de trabajo de asignación de ID. El destino contiene una configuración de los datos de destino que resolverán todos los orígenes. Para definir los datos de entrada que desea resolver en dos Cuentas de AWS, cree una fuente de espacio de nombres de ID y un destino de espacio de nombres de ID para traducir los datos de un conjunto (origen) a otro (destino).

Puede crear un nuevo espacio de nombres de ID o asociar uno existente. Para obtener más información sobre cómo crear un espacio de nombres de ID AWS Entity Resolution, consulte [Definir los datos de entrada mediante](#) un espacio de nombres de ID en la Guía del usuario.AWS Entity Resolution

Temas

- [Creación y asociación de un nuevo espacio de nombres de ID](#)
- [Asociación de un espacio de nombres de ID existente](#)
- [Edición de las asociaciones de espacios de nombres de ID](#)
- [Desasociación de las asociaciones de espacios de nombres de ID](#)

Creación y asociación de un nuevo espacio de nombres de ID

Cada miembro de la colaboración debe crear y asociar un Origen de espacio de nombres de ID o un Destino de espacio de nombres de ID antes de crear una tabla de asignación de ID para consultar datos de identidad.

Si ya ha creado un espacio de nombres de ID en, vaya a. AWS Entity Resolution [Asociación de un espacio de nombres de ID existente](#)

Creación y asociación de un nuevo espacio de nombres de ID

1. Inicie sesión en Consola de administración de AWS y abra la AWS Clean Rooms consola en. <https://console.aws.amazon.com/cleanrooms/>
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Resolución de entidades, elija el Espacio de nombres de ID asociado.
5. En la página del Espacio de nombres de ID asociado, para los datos de resolución de entidades, elija Crear espacio de nombres de ID.

La AWS Entity Resolution consola aparece en una pestaña nueva.

6. Siga las peticiones en la página Creación de un espacio de nombres de ID en la consola de AWS Entity Resolution .
 - a. Para obtener más información, ingrese el nombre y la descripción del espacio de nombres de ID y seleccione el tipo de espacio de nombres de ID (origen o destino).
 - b. Para el método del espacio de nombres de ID, elija el método basado en reglas para la coincidencia basada en reglas o los servicios del proveedor para la transcodificación de terceros.
 - c. Especifique el tipo de entrada de datos, en función del método de espacio de nombres de ID que haya elegido.
 - d. Elija Creación de un espacio de nombres de ID.
7. Vuelve a la AWS Clean Rooms consola.
8. En la página del espacio de nombres de ID asociado, para los datos de resolución de entidades, elija el origen o el destino del espacio de nombres de ID de AWS Entity Resolution que desee asociar a la colaboración en la lista desplegable.
9. Para obtener los detalles de la asociación, siga estos pasos.

- a. Ingrese un nombre para el espacio de nombres de ID asociado.

Puede usar el nombre predeterminado o cambiar el nombre de este espacio de nombres de ID.

- b. (Opcional) Ingrese una Descripción del espacio de nombres de ID.

La descripción ayuda a escribir las consultas.

10. Especifique los permisos de acceso de AWS Clean Rooms mediante la selección de una opción y, a continuación, la realización de la acción recomendada.

Opción	Acción recomendada
Permitir AWS Clean Rooms añadir y gestionar la política de permisos	AWS Clean Rooms crea un rol de servicio con la política requerida para esta asociación.
Agregación y administración de los permisos manualmente	Realice una de las siguientes acciones: - Revise la política de recursos y agregue los permisos necesarios a la política. - Use una política existente mediante la elección de Agregar declaración de política. Debe tener permisos para modificar roles y crear políticas.  Note Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para el rol de servicio.

11. (Opcional) Para las configuraciones de tablas de asignación de ID avanzadas, modifique las protecciones predeterminadas para la columna que proviene del espacio de nombres de ID.

La tabla de asignación de ID está configurada de forma predeterminada para permitir solo una INNER JOIN en la columna `sourceID` y en la columna `targetID`. Puede modificar esta configuración para que la columna que proviene de este espacio de nombres de ID (`sourceID` o `targetID`) se pueda permitir en cualquier parte de la consulta.

Su objetivo	Opción recomendada
Clasificación de la columna como una “columna de unión” y aceptación solo en una cláusula INNER JOIN	Sí
Clasifique la columna como una “columna de dimensiones” y colóquela en cualquier parte de la consulta, incluida una cláusula JOIN, instrucciones SELECT, WHERE y GROUP BY de la consulta.	No, permita el uso de cualquier parte de la consulta

12. (Opcional) Si desea habilitar la opción de Etiquetas para el recurso de espacio de nombres de ID, elija Agregar nueva etiqueta y, a continuación, ingrese el par de Clave y Valor.
13. Elija Asociar .
14. En la pestaña Resolución de entidades, en la tabla de espacios de nombres de ID asociados, consulte el espacio de nombres de ID asociado y compruebe que el tipo de espacio de nombres de ID es correcto (origen o destino).

Una vez que todos los miembros de la colaboración hayan asociado sus espacios de nombres de ID, puede [crear una tabla de asignación de ID](#) y consultar los datos.

Asociación de un espacio de nombres de ID existente

En este procedimiento, cada miembro asocia su origen de espacio de nombres de ID existente o su destino de espacio de nombres de ID en la colaboración.

Asociación de un espacio de nombres de ID existente

1. Inicie sesión en Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms/>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Resolución de entidades, elija el Espacio de nombres de ID asociado.

5. En la página del espacio de nombres de ID asociado, para los datos de resolución de entidades, elija el origen o el destino del espacio de nombres de ID de AWS Entity Resolution que desee asociar a la colaboración en la lista desplegable.
6. Para obtener los detalles de la asociación, siga estos pasos.

- a. Ingrese un nombre para el espacio de nombres de ID asociado.

Puede usar el nombre predeterminado o cambiar el nombre de este espacio de nombres de ID.

- b. (Opcional) Ingrese una Descripción del espacio de nombres de ID.

La descripción ayuda a escribir las consultas.

7. Especifique los permisos de acceso de AWS Clean Rooms mediante la selección de una opción y, a continuación, la realización de la acción recomendada.

Opción	Acción recomendada
Permitir AWS Clean Rooms añadir y gestionar la política de permisos	AWS Clean Rooms crea un rol de servicio con la política requerida para esta asociación.
Agregación y administración de los permisos manualmente	Realice una de las siguientes acciones: - Revise la política de recursos y agregue los permisos necesarios a la política. - Use una política existente mediante la elección de Agregar declaración de política. Debe tener permisos para modificar roles y crear políticas.  Note Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean

Opción	Acción recomendada
	Rooms podido encontrar la política para el rol de servicio.

8. (Opcional) Para las configuraciones de tablas de asignación de ID avanzadas, modifique las protecciones predeterminadas para la columna que proviene del espacio de nombres de ID.

La tabla de asignación de ID está configurada de forma predeterminada para permitir solo una INNER JOIN en la columna `sourceID` y en la columna `targetID`. Puede modificar esta configuración para que la columna que proviene de este espacio de nombres de ID (`sourceID` o `targetID`) se pueda permitir en cualquier parte de la consulta.

Su objetivo	Opción recomendada
Clasifique la columna como una “columna de unión” y acéptela solo en una cláusula INNER JOIN.	Sí
Clasifique la columna como una “columna de dimensiones” y colóquela en cualquier parte de la consulta, incluida una cláusula JOIN, instrucciones SELECT, WHERE y GROUP BY de la consulta.	No, permita el uso de cualquier parte de la consulta

9. (Opcional) Si desea habilitar la opción de Etiquetas para el recurso de espacio de nombres de ID, elija Agregar nueva etiqueta y, a continuación, ingrese el par de Clave y Valor.
10. Elija Asociar .
11. En la pestaña Resolución de entidades, en la tabla de espacios de nombres de ID asociados, consulte el espacio de nombres de ID asociado y compruebe que el tipo de espacio de nombres de ID es correcto (origen o destino).

Una vez que todos los miembros de la colaboración hayan asociado sus espacios de nombres de ID, puede [crear una tabla de asignación de ID](#) y consultar los datos.

Edición de las asociaciones de espacios de nombres de ID

Como miembro de una colaboración, puede editar las asociaciones de espacios de nombres de ID que ha creado.

Edición de una asociación de espacios de nombres de ID

1. Inicie sesión en Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms/>
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Elija la pestaña Resolución de entidades.
5. Para los espacios de nombres de ID asociados, elija un espacio de nombres de ID.
6. En la página de detalles del espacio de nombres de ID, desplácese hacia abajo para ver los Detalles de asociación del espacio de nombres de ID.
7. Elija Edit (Edición de).
8. En la página Editar asociaciones de espacios de nombres de ID, edite cualquiera de las siguientes opciones:
 - a. Para ver los detalles de la asociación, actualice el nombre o la descripción.
 - b. (Opcional) Para las configuraciones de tablas de asignación de ID avanzadas, modifique las protecciones predeterminadas para la columna que proviene del espacio de nombres de ID.

La tabla de asignación de ID está configurada de forma predeterminada para permitir solo una INNER JOIN en la columna sourceID y en la columna targetID. Puede modificar esta configuración para que la columna que proviene de este espacio de nombres de ID (sourceID o targetID) se pueda permitir en cualquier parte de la consulta.

Su objetivo	Opción recomendada
Clasificación de la columna como una “columna de unión” y aceptación solo en una cláusula INNER JOIN	Sí
Clasifique la columna como una “columna de dimensiones” y colóquela en cualquier	No, permita el uso de cualquier parte de la consulta

Su objetivo	Opción recomendada
parte de la consulta, incluida una cláusula JOIN, instrucciones SELECT, WHERE y GROUP BY de la consulta.	

9. Seleccione Save changes (Guardar cambios).

Desasociación de las asociaciones de espacios de nombres de ID

Como miembro de la colaboración, puede desasociar un espacio de nombres de ID de la colaboración. Esta acción impide que el miembro que puede realizar consultas consulte la tabla.

Warning

Al desasociar una asociación de un espacio de nombres de ID de una colaboración, se eliminan todos los datos de las tablas de asignación de ID derivadas, por lo que no se pueden consultar.

Por ejemplo, si la asociación del espacio de nombres de ID se utilizó como SOURCE en tres tablas de asignación de ID diferentes, todos los datos de estas tablas de asignación de ID se eliminarán cuando desasocie la asociación de espacio de nombres de ID.

Desasociación de una asociación de espacio de nombres de ID

1. Inicie sesión en Consola de administración de AWS y abra la consola en AWS Clean Rooms .
<https://console.aws.amazon.com/cleanrooms/>
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Elija la pestaña Resolución de entidades.
5. En Nombres de espacios de ID asociados, seleccione el botón de opción situado junto al espacio de nombres de ID que desea desasociar.
6. Elija Desasociar.
7. En el cuadro de diálogo, confirme su decisión de desconectar el espacio de nombres de ID mediante la elección de Desasociar. Esta acción impide que cualquier miembro que pueda realizar consultas acceda a la tabla de asignación de ID.

Si un miembro de la colaboración elimina uno de los espacios de nombres de ID, no podrá volver a rellenar la tabla de asignación de ID si el origen ha abandonado la colaboración.

Aunque la tabla de asignación de ID ya se había rellenado anteriormente, al desasociar el espacio de nombres de ID, ya no se pueden ejecutar consultas en esa tabla.

Tablas de mapeo de ID en AWS Clean Rooms

Una tabla de mapeo de ID es un recurso AWS Clean Rooms que permite el mapeo de identidades multipartitas en una colaboración.

Antes de crear una tabla de asignación de ID, primero debe tener los datos de origen y destino configurados como espacios de nombres de ID.

Tras crear una tabla de asignación de ID, utilice un flujo de trabajo de asignación de ID para traducir el espacio de nombres de ID de origen al espacio de nombres de ID de destino. Puede hacerlo mediante un método basado en reglas o un método de transcodificación de servicios del proveedor.

Un flujo de trabajo de asignación de ID es un trabajo de procesamiento de datos que asigna datos de un origen de datos de entrada a un destino de datos de entrada en función del método de flujo de trabajo de asignación de ID especificado. Este flujo de trabajo rellena una tabla de asignación de ID.

Note

Las tablas de mapeo de ID solo se pueden crear a partir de conjuntos de datos que se almacenan en Amazon S3 y se rastrean en AWS Glue tablas.

Existen dos métodos de flujo de trabajo de asignación de ID: asignación de ID basado en reglas o asignación de ID de servicios de proveedores:

- Asignación de ID basado en reglas: se utilizan reglas de coincidencia para traducir datos propios de un origen a un destino.
- Mapeo de ID de servicios de proveedores: utiliza el servicio de LiveRamp proveedor para traducir datos de terceros de una fuente a un destino.

 Note

El proveedor de servicios de transcodificación compatible actualmente es LiveRamp. Cualquier miembro de la colaboración que tenga una suscripción a LiveRamp Through AWS Data Exchange puede crear la tabla de mapeo de ID. Si ya tienes una suscripción a LiveRamp, pero no a través de ella AWS Data Exchange, ponte en contacto con nosotros LiveRamp para obtener una oferta privada. Para obtener más información, consulte [Suscripción a un servicio de proveedor en AWS Data Exchange](#) en la Guía del usuario de AWS Entity Resolution .

Temas

- [Creación y llenado de una nueva tabla de asignación de ID](#)
- [Llenado de una tabla de asignación de ID existente](#)
- [Edición de una tabla de asignación de ID](#)
- [Eliminación de una tabla de asignación de ID](#)

Creación y llenado de una nueva tabla de asignación de ID

Requisitos previos

Antes de crear una tabla de mapeo de ID, asegúrese de tener:

- Un origen y un destino del espacio de nombres de ID asociados
- Los espacios de nombres de ID configurados para el mapeo de ID basado en reglas o el mapeo de ID de servicios de proveedores

Puede crear dos tipos de tablas de mapeo de ID:

- Basado en reglas: utiliza reglas coincidentes para traducir datos de origen
- Servicios de proveedores: se utiliza para traducir Ramp LiveRamp IDs

Después de crear una tabla de mapeo de ID, puede rellenarla inmediatamente ejecutando un flujo de trabajo de mapeo de ID, o puede esperar a rellenar la tabla más adelante.

Después de que la tabla de asignación de ID se haya rellenado correctamente, puede ejecutar una consulta de unión de varias tablas en la tabla de asignación de ID para unir `sourceId` con `targetId` y analizar los datos.

Temas

- [Creación de una tabla de asignación de ID \(basada en reglas\)](#)
- [Creación de una tabla de asignación de ID \(servicios del proveedor\)](#)

Creación de una tabla de asignación de ID (basada en reglas)

En este tema se describe el proceso de creación de una tabla de asignación de ID que utilice reglas coincidentes para traducir datos propios de un origen a un destino.

Al crear una tabla de mapeo de ID basada en reglas, puede optar por procesar solo los registros nuevos, actualizados o eliminados del flujo de trabajo activando el procesamiento incremental.

Creación y llenado de una nueva tabla de asignación de ID mediante el método basado en reglas

1. Inicie sesión en Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms/>
2. En el panel de navegación, selecciona Colaboraciones.
3. Elija la colaboración y, a continuación, elija la pestaña de resolución de entidades.
4. Elija Crear una tabla de asignación de ID.
5. En la configuración de mapeo de ID, realice una de las siguientes acciones:
 - Para crear un nuevo flujo de trabajo, mantenga seleccionada la opción Crear un nuevo flujo de trabajo de mapeo de ID.
 - Para usar un flujo de trabajo existente, desactive la casilla de verificación, seleccione un flujo de trabajo de la lista desplegable y, a continuación, vaya al paso 9.
6. En Datos de identidad, consulta o configura el origen y el destino.
 - Para un par de espacio de nombres de ID único: revise el origen y el destino preseleccionados.
 - Para varios espacios de nombres de ID: selecciona el origen y el destino en las listas desplegables.
7. (Opcional) Seleccione Activar el procesamiento incremental para procesar solo los registros nuevos, actualizados o eliminados del flujo de trabajo.

AWS Entity Resolution procesa solo los registros nuevos, actualizados o eliminados en el espacio de nombres de ID de origen o de destino, en lugar de volver a crear toda la tabla de mapeo de ID.

Si deja esta opción sin seleccionar, AWS Entity Resolution ejecuta el flujo de trabajo de mapeo de ID de procesamiento por lotes predeterminado en la tabla de mapeo de ID.

8. En Parámetros de regla, configure lo siguiente:

- Controles de reglas: elija si el destino o la fuente proporcionan reglas coincidentes.

Puede ver las reglas activando Mostrar reglas.

Los controles de reglas deben ser compatibles entre el espacio de nombres del ID de origen y el de destino para poder usarse en un flujo de trabajo de asignación de ID. Por ejemplo, si un espacio de nombres de ID de origen limita las reglas al destino, pero el espacio de nombres de ID de destino limita las reglas al origen, se produce un error.

- El tipo de comparación se establece automáticamente en varios campos de entrada.

Esto se debe a que ambos participantes habían seleccionado esta opción anteriormente.

- Coincidencia de registros
 - Una fuente para un destino: almacena un registro coincidente por destino
 - Varias fuentes para un destino: almacena todos los registros coincidentes por destino

 Note

Las limitaciones especificadas para los espacios de nombres de ID de origen y destino deben ser compatibles.

9. Para obtener detalles del mapeo de ID, configure lo siguiente:

- a. Introduzca un nombre para la tabla de mapeo de ID o mantenga el nombre predeterminado.
- b. (Opcional) Ingrese una Descripción de la tabla de asignación de ID.

La descripción ayuda a escribir las consultas.

10. Para AWS Clean Rooms acceder, elija uno:

- Permitir AWS Clean Rooms agregar y administrar la política de permisos: crea un rol de servicio automáticamente.

- Agregue y administre los permisos manualmente: revise y modifique la política de recursos o elija Agregar declaración de política.

 Note

Si no puede modificar la política de funciones, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para la función de servicio.

11. Para AWS Entity Resolution acceder, elige una de las siguientes opciones:

Esta sección solo está visible si está creando una nueva tabla de asignación de ID.

- Crear y usar un nuevo rol de servicio
 - El nombre del rol de servicio predeterminado es `entityresolution-id-mapping-workflow-<timestamp>`
 - (Opcional) Para datos cifrados, seleccione Estos datos están cifrados mediante una clave KMS e introduzca la AWS KMS clave.
- Usar un rol de servicio existente
 - Elija un nombre de función de servicio existente en la lista desplegable o introduzca un ARN de función.

Si tiene permisos de listas de roles, se mostrará la lista de roles.

Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM.

Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible.

De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

- (Opcional) Seleccione Agregar una política preconfigurada con los permisos necesarios a este rol para adjuntar los permisos necesarios al rol.

Debe tener permisos para modificar roles y crear políticas.

12. (Opcional) En Configuración adicional, configure:

- a. Configuración de la tabla de mapeo de ID

- Para habilitar el cifrado personalizado, elija Personalizar la configuración de cifrado e introduzca una AWS KMS clave.

 Note

Esta clave de KMS debe conceder los permisos necesarios para utilizarla dentro de AWS Entity Resolution una política de claves de KMS. `cleanrooms.amazonaws.com` Para obtener más información sobre los permisos necesarios para trabajar con cifrados con un flujo de trabajo de asignación de ID, consulte [Crear un rol de trabajo del flujo de trabajo para AWS Entity Resolution](#) en la Guía del usuario de AWS Entity Resolution .

- Para añadir etiquetas, elija Añadir nueva etiqueta e introduzca los pares clave-valor
- b. Configuración del flujo de trabajo de mapeo de ID (solo flujos de trabajo nuevos):
- Para usar nombres diferentes, desactive Mantener el mismo nombre y descripción de la tabla de mapeo de ID e introduzca valores nuevos.
 - Para añadir etiquetas, seleccione Añadir nueva etiqueta e introduzca pares clave-valor

13. Seleccione una de las siguientes opciones:

- Crear tabla de mapeo de ID: crea una tabla vacía que puede rellenar más adelante () [Llenado de una tabla de asignación de ID existente](#)
- Crear y rellenar la tabla de mapeo de ID: crea y rellena inmediatamente la tabla (puede tardar varias horas)

Comienza el proceso de flujo de trabajo de asignación de ID. Durante este proceso, la tabla de mapeo de ID se rellena con los datos traducidos. IDs Es posible que el flujo de trabajo de asignación de ID tarde unas horas en procesarse.

Después de que la tabla de asignación de ID se haya rellenado correctamente, puede [consultar la tabla de asignación de ID](#) para unir `sourceId` con `targetId` y analizar los datos.

Creación de una tabla de asignación de ID (servicios del proveedor)

En este tema se describe el proceso de creación de una tabla de mapeo de ID que utilice un servicio de proveedor (LiveRamp). Los servicios del LiveRamp proveedor traducen un conjunto de Ramp de origen IDs a otro mediante Ramp mantenido o derivado IDs.

Creación de una nueva tabla de asignación de ID mediante el método de servicios del proveedor

1. Inicie sesión en Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms/>.
2. En el panel de navegación, selecciona Colaboraciones.
3. Elija la colaboración y, a continuación, elija la pestaña de resolución de entidades.
4. Elija Crear una tabla de asignación de ID.
5. En la configuración de mapeo de ID, realice una de las siguientes acciones:
 - Para crear un nuevo flujo de trabajo, mantenga seleccionada la opción Crear un nuevo flujo de trabajo de mapeo de ID.
 - Para usar un flujo de trabajo existente, desactive la casilla de verificación, seleccione un flujo de trabajo de la lista desplegable y, a continuación, vaya al paso 9.
6. En Datos de identidad, consulta o configura el origen y el destino.
 - Para un par de espacio de nombres de ID único: revise el origen y el destino preseleccionados.
 - Para varios espacios de nombres de ID: selecciona el origen y el destino en las listas desplegables.
7. En Método, compruebe que el método de flujo de trabajo de mapeo de ID seleccionado esté transcodificando. LiveRamp
8. Para las LiveRamp configuraciones, realice una de las siguientes acciones:
 - Introduzca el ARN del administrador de LiveRamp ID y el ARN del administrador LiveRamp secreto.
 - Seleccione Importar desde un flujo de trabajo existente.
9. Para ver los detalles del mapeo de ID, configure lo siguiente:
 - a. Introduzca un nombre para la tabla de mapeo de ID o mantenga el nombre predeterminado.
 - b. (Opcional) Ingrese una Descripción de la tabla de asignación de ID.

La descripción ayuda a escribir las consultas.

10. Para AWS Clean Rooms acceder, elija uno:

- Permitir AWS Clean Rooms agregar y administrar la política de permisos: crea un rol de servicio automáticamente.
- Agregue y administre los permisos manualmente: revise y modifique la política de recursos o elija Agregar declaración de política.

 Note

Si no puede modificar la política de funciones, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para la función de servicio.

11. Para AWS Entity Resolution acceder, elige una de las siguientes opciones:

Esta sección solo está visible si está creando una nueva tabla de asignación de ID.

- Crear y usar un nuevo rol de servicio
 - El nombre del rol de servicio predeterminado es `entityresolution-id-mapping-workflow-<timestamp>`
 - (Opcional) Para datos cifrados, seleccione Estos datos están cifrados mediante una clave KMS e introduzca la AWS KMS clave.
- Usar un rol de servicio existente
 - Elija un nombre de función de servicio existente en la lista desplegable o introduzca un ARN de función.

Si tiene permisos de listas de roles, se mostrará la lista de roles.

Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM.

Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible.

De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

- (Opcional) Seleccione Agregar una política preconfigurada con los permisos necesarios a este rol para adjuntar los permisos necesarios al rol.

Debe tener permisos para modificar roles y crear políticas.

12. (Opcional) En Ajustes adicionales, configure:

a. Configuración de la tabla de mapeo de ID

- Para habilitar el cifrado personalizado, elija Personalizar la configuración de cifrado e introduzca una AWS KMS clave.

 Note

Esta clave de KMS debe conceder los permisos necesarios para utilizarla dentro de AWS Entity Resolution una política de claves de KMS. `cleanrooms.amazonaws.com` Para obtener más información sobre los permisos necesarios para trabajar con cifrados con un flujo de trabajo de asignación de ID, consulte [Crear un rol de trabajo del flujo de trabajo para AWS Entity Resolution](#) en la Guía del usuario de AWS Entity Resolution .

- Para añadir etiquetas, elija Añadir nueva etiqueta e introduzca los pares clave-valor.

b. Configuración del flujo de trabajo de mapeo de ID (solo flujos de trabajo nuevos):

- Para usar nombres diferentes, desactive Mantener el mismo nombre y descripción de la tabla de mapeo de ID e introduzca valores nuevos.
- Para añadir etiquetas, elija Añadir nueva etiqueta e introduzca los pares clave-valor.

13. Seleccione una de las siguientes opciones:

- Crear tabla de mapeo de ID: crea una tabla vacía que puede rellenar más adelante () [Llenado de una tabla de asignación de ID existente](#)
- Crear y rellenar la tabla de mapeo de ID: crea y rellena inmediatamente la tabla (puede tardar varias horas)

Comienza el proceso de flujo de trabajo de asignación de ID. Durante este proceso, la tabla de mapeo de ID se rellena con los datos traducidos. IDs Es posible que el flujo de trabajo de asignación de ID tarde unas horas en procesarse.

Después de que la tabla de asignación de ID se haya rellenado correctamente, puede [consultar la tabla de asignación de ID](#) para unir `sourceId` con `targetId` y analizar los datos.

Llenado de una tabla de asignación de ID existente

Cuando se agreguen nuevos datos a un espacio de nombres de ID, utilice este flujo de trabajo. Si ha elegido activar el procesamiento incremental al [crear la tabla de mapeo de ID](#), tiene la posibilidad de procesar solo los registros nuevos, actualizados o eliminados en el espacio de nombres de ID de origen o de destino, en lugar de volver a crear toda la tabla de mapeo de ID.

Llenado de una tabla de asignación de ID existente

1. Inicie sesión en Consola de administración de AWS y abra la consola en AWS Clean Rooms .
<https://console.aws.amazon.com/cleanrooms/>
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Ve a la pestaña de resolución de entidades.
5. En la sección Tablas de mapeo de ID, selecciona una tabla de mapeo de ID.
6. Si no ha activado el procesamiento incremental para esta tabla de mapeo de ID, elija una tabla de mapeo de ID y, a continuación, elija Rellenar.
7. Si ha activado el procesamiento incremental para esta tabla de mapeo de ID, elija Rellenar con y, a continuación, elija una de las siguientes opciones:

- Procesamiento incremental: procesa solo los registros nuevos, actualizados o eliminados en el espacio de nombres del ID de origen o de destino.

Se recomienda para actualizaciones frecuentes, ejecuciones diarias o sincronización de datos en tiempo real.

- Procesamiento por lotes: procesa toda la tabla de mapeo de ID.

Se recomienda para la configuración inicial, las actualizaciones completas periódicas o cuando se producen cambios significativos en los espacios de nombres de los identificadores de origen y destino.

- Eliminar solo el procesamiento: procesa solo los registros eliminados del espacio de nombres del ID de origen y actualiza el espacio de nombres del ID de destino en consecuencia.

Se recomienda para sincronizar rápidamente las eliminaciones.

8. Comienza el proceso de flujo de trabajo de asignación de ID.

Durante este proceso, la tabla de mapeo de ID se rellena con códigos transcodificados. IDs Es posible que el flujo de trabajo de asignación de ID tarde unas horas en procesarse.

Después de que la tabla de asignación de ID se haya relleno correctamente, puede [consultar la tabla de asignación de ID](#) para unir sourceId con targetId.

Edición de una tabla de asignación de ID

Como miembro de una colaboración, puede editar la tabla de asignación de ID que ha creado.

Edición de una tabla de asignación de ID

1. Inicia sesión en la consola Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Elija la pestaña Resolución de entidades.
5. Para las tablas de asignación de ID, elija una tabla.
6. En la página de detalles de la tabla de asignación, desplácese hacia abajo para ver los Detalles de la tabla de asignación de ID.
7. Seleccione Editar.
8. En la página Editar tablas de asignación de ID, actualice la Descripción o la Información de acceso del servicio.
9. Seleccione Save changes (Guardar cambios).

Eliminación de una tabla de asignación de ID

Como miembro de una colaboración, puede eliminar una tabla de asignación de ID que haya creado. Esta acción impide que el miembro que puede realizar consultas consulte la tabla.

Warning

Al eliminar una tabla de asignación, se eliminan permanentemente todos los datos rellenos.

Eliminación de una tabla de asignación de ID

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con la tuya Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. Elija la pestaña Resolución de entidades.
5. Para las tablas de asignación de ID, elija una tabla.
6. En la página de detalles de la tabla de asignación de ID, desplácese hacia abajo para ver las Tablas de asignación de ID.
7. Elija una tabla de asignación de ID y, a continuación, elija Eliminar.
8. Si tiene la certeza de que desea eliminar la tabla de asignación de ID, elija Eliminar.

Plantillas de análisis en AWS Clean Rooms

Las plantillas de análisis funcionan con [Regla de análisis personalizada en AWS Clean Rooms](#). Con una plantilla de análisis, puede definir parámetros que le ayuden a reutilizar la misma consulta. AWS Clean Rooms admite un subconjunto de parametrización con valores literales.

Las plantillas de análisis son específicas de la colaboración. En cada colaboración, los miembros pueden ver solo las consultas de dicha colaboración. Si tiene previsto utilizar la privacidad diferencial en una colaboración, debe asegurarse de que sus plantillas de análisis sean compatibles con la [estructura de consultas de uso general](#) de la privacidad diferencial de AWS Clean Rooms .

Puedes crear una plantilla de análisis de dos maneras: con código SQL o con código Python para Spark.

Temas

- [Plantillas de análisis SQL](#)
- [PySpark plantillas de análisis](#)
- [Plantillas de PySpark análisis para la solución](#)

Plantillas de análisis SQL

Las plantillas de análisis SQL le permiten consultar y analizar datos de diferentes conjuntos de datos en el marco de una colaboración. Puede utilizar estas plantillas para realizar varios tipos de análisis, como identificar las superposiciones de audiencia y calcular las métricas agregadas.

Con las plantillas de análisis de SQL, puede:

- Escribir consultas SQL estándar
- Añada parámetros para hacer que sus consultas sean dinámicas
- Controle el acceso a columnas y tablas específicas
- Establezca los requisitos de agregación para los datos confidenciales
- Defina los datos de entrada para la generación de conjuntos de datos sintéticos con privacidad mejorada para modelos de aprendizaje automático (ML) personalizados

Temas

- [Creación de una plantilla de análisis SQL](#)

- [Revisión de una plantilla de análisis SQL](#)

Creación de una plantilla de análisis SQL

Requisitos previos

Antes de crear una plantilla de análisis SQL, debe tener:

- Una AWS Clean Rooms colaboración activa
- Acceso a al menos una tabla configurada de la colaboración

Para obtener información sobre la configuración de tablas en AWS Clean Rooms, consulte [Crear una tabla configurada en AWS Clean Rooms](#).

- Permisos para crear plantillas de análisis
- Conocimientos básicos de la sintaxis de consultas SQL

El siguiente procedimiento describe el proceso de creación de una plantilla de análisis SQL mediante la [AWS Clean Rooms consola](#).

Para obtener información sobre cómo crear una plantilla de análisis SQL mediante la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

Para crear una plantilla de análisis SQL

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#) con la Cuenta de AWS que funcionará como creador de la colaboración.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Plantillas, vaya a la sección Plantillas de análisis creadas por usted.
5. Seleccione Crear plantilla de análisis.
6. En la página Crear plantilla de análisis, en Detalles,
 - a. Introduzca un nombre para la plantilla de análisis.
 - b. (Opcional) Introduzca una Descripción.
 - c. En Formato, deje seleccionada la opción SQL.
7. En Tablas, consulte las tablas configuradas asociadas a la colaboración.

8. En Definición,

- a. Introduzca la definición de la plantilla de análisis.
- b. Seleccione Importar desde para importar una definición.
- c. (Opcional) Especifique un parámetro en el editor SQL introduciendo dos puntos (:) delante del nombre del parámetro.

Por ejemplo:

```
WHERE table1.date + :date_period > table1.date
```

9. Si ha añadido parámetros anteriormente, en Parámetros: opcional, para cada Nombre de parámetro, elija el Tipo y el Valor predeterminado (opcional).
10. En el caso de los datos sintéticos, si desea generar datos sintéticos para el entrenamiento del modelo, active la casilla de verificación Requerir que el resultado de la plantilla de análisis sea sintético.

Para obtener más información, consulte Generación de conjuntos de datos [sintéticos con privacidad mejorada](#).

- a. Para la clasificación de columnas, elija una columna de la lista desplegable. Se requieren al menos cinco columnas.
 - i. Elija una clasificación de la lista desplegable. Esto identifica el tipo de datos de cada columna.

Los tipos de clasificación incluyen:

- Numérico: valores numéricos continuos, como medidas o recuentos
 - Categórico: valores o categorías discretos, como etiquetas o tipos
- ii. Para eliminar una columna, seleccione Eliminar.
 - iii. Para añadir otra columna, selecciona Añadir otra columna. Elija la columna y la clasificación en las listas desplegables.
 - iv. En Valor predictivo, elija una columna de la lista desplegable. Esta es la columna que el modelo personalizado usa para la predicción después de entrenarse en el conjunto de datos sintético.
- b. La configuración avanzada le permite establecer el nivel de privacidad y el umbral de privacidad. Ajusta la configuración para que se adapte a tus necesidades.

- En el nivel de privacidad, introduzca un valor épsilon para determinar cuánto ruido añade el modelo sintético para proteger la privacidad del conjunto de datos generado. El valor debe estar entre 0,0001 y 10.
- Los valores más bajos añaden más ruido, lo que proporciona una mayor protección de la privacidad, pero puede reducir la utilidad del modelo personalizado posterior basado en estos datos.
- Los valores más altos agregan menos ruido, lo que proporciona una mayor precisión pero reduce potencialmente la protección de la privacidad.

En Umbral de privacidad, introduzca la probabilidad más alta permitida de que un ataque de inferencia de pertenencia pueda identificar a los miembros del conjunto de datos original. El valor debe estar entre 50,0 y 100.

- Las puntuaciones del 50% indican que un ataque de inferencia de miembros no puede distinguir mejor a los miembros de los que no lo son mediante una suposición aleatoria.
- Si no hay límite de privacidad, introduce el 100%.

El valor óptimo depende de su caso de uso específico y los requisitos de privacidad. Si se supera el umbral de privacidad, se produce un error al crear el canal de entrada de ML y no se puede utilizar el conjunto de datos sintético para entrenar un modelo.

Warning

La generación de datos sintéticos evita que se puedan deducir atributos individuales, ya sea que haya personas específicas presentes en el conjunto de datos original o que estén presentes los atributos de aprendizaje de esas personas. Sin embargo, no impide que los valores literales del conjunto de datos original, incluida la información de identificación personal (PII), aparezcan en el conjunto de datos sintético.

Recomendamos evitar los valores en el conjunto de datos de entrada que estén asociados a un solo sujeto de datos, ya que pueden volver a identificar a un sujeto de datos. Por ejemplo, si solo un usuario vive en un código postal, la presencia de ese código postal en el conjunto de datos sintético confirmaría que el usuario estaba en el conjunto de datos original. Para mitigar este riesgo, se pueden utilizar técnicas como

truncar valores de alta precisión o reemplazar catálogos poco comunes por otros. Estas transformaciones pueden formar parte de la consulta utilizada para crear el canal de entrada de ML.

11. Si desea habilitar las etiquetas para el recurso, elija Agregar nueva etiqueta y, a continuación, introduzca el par clave y valor.
12. Seleccione Crear.
13. Ahora puede informar a su colaborador de que puede [revisar una plantilla de análisis](#). (opcional si desea consultar sus propios datos).

Revisión de una plantilla de análisis SQL

Una vez que un miembro de la colaboración haya creado una SQLanalysis plantilla, puede revisarla y aprobarla. Una vez aprobada la plantilla de análisis, se puede utilizar en una consulta en AWS Clean Rooms.

Note

Cuando incorpore su código de análisis a una colaboración, tenga en cuenta lo siguiente:

- AWS Clean Rooms no valida ni garantiza el comportamiento del código de análisis.
 - Si necesita garantizar un comportamiento determinado, revise directamente el código de su socio colaborador o póngase en contacto con un auditor externo de confianza para que lo revise.
- En el modelo de seguridad compartida:
 - Usted (el cliente) es responsable de la seguridad del código que se ejecuta en el entorno.
 - AWS Clean Rooms es responsable de la seguridad del medio ambiente y garantiza que
 - solo se ejecuta el código aprobado
 - solo se puede acceder a las tablas configuradas especificadas
 - el único destino de salida es el depósito S3 del receptor de resultados.

Para revisar una plantilla de análisis SQL mediante la AWS Clean Rooms consola

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#) con la Cuenta de AWS que funcionará como creador de la colaboración.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Plantillas, vaya a la sección Plantillas de análisis creadas por usted.
5. Elija la plantilla de análisis cuyo estado Se puede ejecutar sea No precisa su revisión.
6. Elija Revisar.
7. Revise la Descripción general, la definición y los Parámetros de la regla de análisis (si procede).
8. Revise las tablas configuradas que se enumeran en Tablas a las que se hace referencia en la definición.

El Estado que aparece junto a cada tabla indicará Plantilla no permitida.

9. Elija una tabla.

En el caso de	Entonces elija...
Aprueba la plantilla de análisis	Permitir plantilla en la tabla. Para confirmar su aprobación, seleccione Permitir.
No aprueba la plantilla de análisis	No permitir

Ahora puede consultar la tabla configurada mediante una plantilla de análisis SQL. Para obtener más información, consulte [Ejecución de consultas SQL](#).

PySpark plantillas de análisis

PySpark las plantillas de análisis requieren un script de usuario de Python y un entorno virtual opcional para utilizar bibliotecas personalizadas y de código abierto. Estos archivos se denominan artefactos.

Antes de crear una plantilla de análisis, primero debe crear los artefactos y, a continuación, almacenarlos en un bucket de Amazon S3. AWS Clean Rooms utiliza estos artefactos al ejecutar trabajos de análisis. AWS Clean Rooms solo accede a los artefactos cuando se ejecuta un trabajo.

Antes de ejecutar cualquier código en una plantilla de PySpark análisis, AWS Clean Rooms valida los artefactos de la siguiente manera:

- Comprobar la versión específica del objeto S3 utilizada al crear la plantilla
- Verificación del hash SHA-256 del artefacto
- Falló cualquier trabajo en el que se hayan modificado o eliminado artefactos

 Note

El tamaño máximo de todos los artefactos combinados para una plantilla de PySpark análisis determinada AWS Clean Rooms es de 1 GB.

Seguridad para las plantillas PySpark de análisis

Para preservar un entorno informático seguro, AWS Clean Rooms utiliza una arquitectura informática de dos niveles para aislar el código de usuario de las operaciones del sistema. Esta arquitectura se basa en la tecnología Amazon EMR Serverless Fine Grained Access Control, también conocida como Membrane. Para obtener más información, consulte [Membrane: controles de acceso a datos seguros y eficaces en Apache Spark en presencia de código imperativo](#).

Los componentes del entorno informático se dividen en un espacio de usuario y un espacio de sistema independientes. El espacio de usuario ejecuta el PySpark código de la plantilla de PySpark análisis. AWS Clean Rooms utiliza el espacio del sistema para permitir la ejecución del trabajo, lo que incluye el uso de las funciones de servicio proporcionadas por los clientes para leer los datos a fin de ejecutar el trabajo y la implementación de la lista de columnas permitidas. Como resultado de esta arquitectura, el PySpark código de un cliente que afecta al espacio del sistema, que podría incluir una pequeña cantidad de Spark SQL PySpark DataFrames APIs, queda bloqueado.

PySpark limitaciones en AWS Clean Rooms

Cuando los clientes envían una plantilla de PySpark análisis aprobada, AWS Clean Rooms la ejecutan en su propio entorno informático seguro al que ningún cliente puede acceder. El entorno informático implementa una arquitectura informática con un espacio de usuario y un espacio de sistema para preservar un entorno informático seguro. Para obtener más información, consulte [Seguridad para las plantillas PySpark de análisis](#).

Tenga en cuenta las siguientes limitaciones antes PySpark de usarlo AWS Clean Rooms.

Limitaciones

- Solo se admiten las DataFrame salidas
- Una sola sesión de Spark por ejecución de trabajo

Características no admitidas

- Administración de datos
 - Formatos de tablas Iceberg
 - LakeFormation tablas gestionadas
 - Conjuntos de datos distribuidos resilientes (RDD)
 - Streaming de Spark
 - Control de acceso para columnas anidadas
- Funciones y extensiones personalizadas
 - Funciones de tabla definidas por el usuario () UDTFs
 - Colmena UDFs
 - Clases personalizadas en funciones definidas por el usuario
 - Orígenes de datos personalizados
 - Archivos JAR adicionales para:
 - Extensiones Spark
 - Connectors
 - Configuraciones de Metastore
- Monitorización y análisis
 - Registro de chispas
 - IU de Spark
 - Comandos de ANALYZE TABLE

Important

Estas limitaciones se han establecido para mantener el aislamiento de seguridad entre los espacios del usuario y del sistema.

Todas las restricciones se aplican independientemente de la configuración de colaboración.

Es posible que las actualizaciones futuras añadan compatibilidad con funciones adicionales en función de las evaluaciones de seguridad.

Prácticas recomendadas

Recomendamos las siguientes prácticas recomendadas a la hora de crear plantillas de PySpark análisis.

- Diseñe sus plantillas de análisis teniendo esto [PySpark limitaciones en AWS Clean Rooms](#) en cuenta.
- Pruebe primero su código en un entorno de desarrollo.
- Utilice exclusivamente DataFrame las operaciones compatibles.
- Planifique su estructura de salida para que funcione con DataFrame limitaciones.

Recomendamos las siguientes prácticas recomendadas para gestionar los artefactos

- Guarde todos los artefactos de la plantilla de PySpark análisis en un prefijo o depósito de S3 específico.
- Utilice nombres de versión claros para las diferentes versiones de los artefactos.
- Crea nuevas plantillas de análisis cuando necesites actualizar los artefactos.
- Mantenga un inventario de qué plantillas utilizan qué versiones de artefactos.

Para obtener más información sobre cómo escribir el código de Spark, consulta lo siguiente:

- [Ejemplos de Apache Spark](#)
- [Escribe una aplicación de Spark](#) en la guía de versiones de Amazon EMR
- [Tutorial: Cómo escribir un script AWS Glue para Spark](#) en la guía del AWS Glue usuario

En los temas siguientes se explica cómo crear scripts y bibliotecas de usuario de Python antes de crear y revisar la plantilla de análisis.

Temas

- [Crear un script de usuario](#)
- [Trabajar con parámetros en plantillas PySpark de análisis](#)

- [Crear un entorno virtual \(opcional\)](#)
- [Almacenamiento de un script de usuario y un entorno virtual en S3](#)
- [Crear una plantilla PySpark de análisis](#)
- [Revisión de una plantilla PySpark de análisis](#)

Crear un script de usuario

El script de usuario debe contener una función de punto de entrada (en otras palabras, un controlador). Puede asignar un nombre a su archivo de script de usuario con cualquier nombre de archivo de Python válido.

El siguiente procedimiento describe cómo crear un script de usuario para definir la funcionalidad principal del PySpark análisis.

Requisitos previos

- PySpark 1.0 (corresponde a Python 3.11 y Spark 3.5.3)
- Los conjuntos de datos de Amazon S3 solo se pueden leer como asociaciones de tablas configuradas en la sesión de Spark que defina.
- Su código no puede llamar directamente a Amazon S3 y AWS Glue
- Su código no puede realizar llamadas de red

Para crear un script de usuario

1. Abra el editor de texto o el entorno de desarrollo integrado (IDE) de su elección.

Puede usar cualquier editor de texto o IDE (como Visual Studio Code o Notepad++) que admita archivos de Python. PyCharm

2. Cree un nuevo archivo de Python con el nombre que desee (por ejemplo, **my_analysis.py**).
3. Defina una función de punto de entrada que acepte un parámetro de objeto de contexto.

```
def entrypoint(context)
```

El parámetro de `context` objeto es un diccionario que proporciona acceso a los componentes esenciales de Spark, a las tablas referenciadas y a los parámetros de análisis. Contiene:

Acceda a la sesión de Spark mediante `context['sparkSession']`

Tablas referenciadas mediante `context['referencedTables']`

Parámetros de análisis mediante `context['analysisParameters']` (si los parámetros están definidos en la plantilla)

4. Defina los resultados de la función de punto de entrada:

```
return results
```

`results` Debe devolver un objeto que contenga un diccionario de resultados con nombres de archivos a una salida. `DataFrame`

 Note

AWS Clean Rooms escribe automáticamente los `DataFrame` objetos en el depósito S3 del receptor de resultados.

5. Ya puede hacer lo siguiente:

- a. Guarde este script de usuario en S3. Para obtener más información, consulte [Almacenamiento de un script de usuario y un entorno virtual en S3](#).
- b. Cree el entorno virtual opcional para admitir cualquier biblioteca adicional que necesite su script de usuario. Para obtener más información, consulte [Crear un entorno virtual \(opcional\)](#).

Example Ejemplo 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 estructura de carpetas de este ejemplo es la siguiente:

```

analysis_results/
#
### output1/ # Basic selected columns
# ### part-000000.parquet
# ### _SUCCESS
#
### output2/ # Joined data
# ### part-000000.parquet
# ### _SUCCESS
#
### analysis_summary/ # Aggregated results
### part-000000.parquet
### _SUCCESS

```

Example Ejemplo 2

<caption>The following example demonstrates a more complex user script for a PySpark analysis template.</caption>

```
def entripoint(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
```

Trabajar con parámetros en plantillas PySpark de análisis

Los parámetros aumentan la flexibilidad de sus plantillas de PySpark análisis al permitir que se proporcionen diferentes valores en el momento de enviar los trabajos. Se puede acceder a los parámetros a través del objeto de contexto que se pasa a la función de punto de entrada.

Note

Los parámetros son cadenas proporcionadas por el usuario que pueden contener contenido arbitrario.

- Revise el código para asegurarse de que los parámetros se gestionen de forma segura y evitar comportamientos inesperados en el análisis.

- Diseñe el manejo de los parámetros para que funcione de forma segura, independientemente de los valores de los parámetros que se proporcionen en el momento del envío.

Acceder a los parámetros

Los parámetros están disponibles en el `context['analysisParameters']` diccionario. Todos los valores de los parámetros son cadenas.

Example Acceso seguro a los parámetros

```
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
        }
    }
```

Prácticas recomendadas de seguridad de parámetros

Warning

Los parámetros son cadenas proporcionadas por el usuario que pueden contener contenido arbitrario. Debe gestionar los parámetros de forma segura para evitar vulnerabilidades de seguridad en el código de análisis.

Patrones de manejo de parámetros inseguros para evitar:

- Ejecutar parámetros como código: nunca utilice `eval()` ni `exec()` sobre los valores de los parámetros

```
# UNSAFE - Don't do this
eval(parameters['expression']) # Can execute arbitrary code
```

- Interpolación de cadenas SQL: nunca concatene los parámetros directamente en cadenas SQL

```
# UNSAFE - Don't do this
sql = f"SELECT * FROM table WHERE column = '{parameters['value']}'" # SQL injection risk
```

- Operaciones de rutas de archivos no seguras: nunca utilice los parámetros directamente en las operaciones del sistema de archivos sin validarlos

```
# UNSAFE - Don't do this
file_path = f"/data/{parameters['filename']}" # Path traversal risk
```

Patrones de manejo seguro de parámetros:

- Utilice los parámetros en DataFrame las operaciones: DataFrames maneje los valores de los parámetros de forma segura

```
# SAFE - Use parameters in DataFrame operations
threshold = int(parameters['threshold'])
filtered_df = input_df.filter(input_df.value > threshold)
```

- Valide los valores de los parámetros: compruebe que los parámetros cumplen con los formatos esperados antes de usarlos

```
# 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}")
```

- Utilice listas de valores permitidos para los valores de los parámetros: siempre que sea posible, valide los parámetros comparándolos con valores válidos conocidos

```
# 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}")
```

- Conversión de tipos con gestión de errores: convierta los parámetros de cadena en los tipos esperados de forma segura

```
# 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

Recuerde que los parámetros omiten la revisión del código cuando los ejecutores de tareas proporcionan valores diferentes. Diseñe el manejo de los parámetros para que funcione de forma segura, independientemente de los valores de los parámetros que se proporcionen.

Ejemplo completo de parámetros

Example Utilizar los parámetros de forma segura en 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
```

Crear un entorno virtual (opcional)

Si su script de usuario requiere bibliotecas adicionales, tiene la opción de crear un entorno virtual para almacenar esas bibliotecas. Si no necesita bibliotecas adicionales, puede omitir este paso.

Cuando trabajes con bibliotecas que tienen extensiones nativas, ten en cuenta que PySpark AWS Clean Rooms funciona en Linux con ARM64 arquitectura.

El siguiente procedimiento muestra cómo crear un entorno virtual mediante un comando CLI básico.

Para crear un entorno virtual

1. Abra una terminal o una línea de comandos.
2. Añada el siguiente contenido:

```
# 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. Ahora está listo para almacenar este entorno virtual en S3. Para obtener más información, consulte [Almacenamiento de un script de usuario y un entorno virtual en S3](#).

Para obtener más información sobre cómo trabajar con Docker y Amazon ECR, consulta la Guía de [Amazon ECRUser](#).

Almacenamiento de un script de usuario y un entorno virtual en S3

El siguiente procedimiento explica cómo almacenar un script de usuario y un entorno virtual opcional en Amazon S3. Complete este paso antes de crear una plantilla de PySpark análisis.

Important

No modifique ni elimine artefactos (scripts de usuario o entornos virtuales) después de crear una plantilla de análisis.

Si lo hace, hará lo siguiente:

- Provoca que todos los futuros trabajos de análisis que utilicen esta plantilla fallen.
- Requiere la creación de una nueva plantilla de análisis con nuevos artefactos.
- No afecta a los trabajos de análisis completados anteriormente

Requisitos previos

- Y Cuenta de AWS con los permisos adecuados
- Un archivo de script de usuario (comomy_analysis.py)
- (Opcional, si existe) Un paquete (.tar.gzarchivo) de entorno virtual
- Acceso para crear o modificar funciones de IAM

Console

Para almacenar un script de usuario y un entorno virtual en S3 mediante la consola:

1. Inicie sesión en la consola de Amazon S3 Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/s3/>.
2. Cree un nuevo bucket de S3 o utilice uno existente.
3. Habilite el control de versiones para el bucket.
 - a. Seleccione su depósito.
 - b. Seleccione Propiedades.
 - c. En la sección Control de versiones de buckets, selecciona Editar.
 - d. Seleccione Activar y guardar los cambios.
4. Carga tus artefactos y activa el hash SHA-256.
 - a. Navega hasta tu bucket.
 - b. Seleccione Cargar.
 - c. Selecciona Añadir archivos y añade tu archivo de script de usuario.
 - d. (Opcional, si existe) Añada su archivo.tar.gz.
 - e. Amplíe Propiedades.
 - f. En Sumas de comprobación, para la función de suma de comprobación, seleccione. SHA256
 - g. Seleccione Cargar.
5. Ahora está listo para crear una plantilla de PySpark análisis.

CLI

Para almacenar el script de usuario y el entorno virtual en S3 mediante AWS CLI:

1. Use el siguiente comando:

```
aws s3 cp --checksum-algorithm sha256 pyspark_venv.tar.gz s3://ARTIFACT-BUCKET/  
EXAMPLE-PREFIX/
```

2. Ahora está listo para crear una plantilla PySpark de análisis.

 Note

Si necesita actualizar el script o el entorno virtual:

1. Cargue la nueva versión como un objeto independiente.
2. Cree una nueva plantilla de análisis con los nuevos artefactos.
3. Elimine la plantilla anterior.
4. Guarde los artefactos originales en S3 si la plantilla anterior pudiera seguir siendo necesaria.

Crear una plantilla PySpark de análisis

 Note

Los parámetros son cadenas proporcionadas por el usuario que pueden contener contenido arbitrario.

- Revise el código para asegurarse de que los parámetros se gestionen de forma segura y evitar comportamientos inesperados en el análisis.
- Diseñe el manejo de los parámetros para que funcione de forma segura, independientemente de los valores de los parámetros que se proporcionen en el momento del envío.

Requisitos previos

Antes de crear una plantilla de PySpark análisis, debe tener:

- Ser miembro de una AWS Clean Rooms colaboración activa
- Acceso a al menos una tabla configurada en la colaboración activa
- Permisos para crear plantillas de análisis
- Un script de usuario de Python y un entorno virtual creados y almacenados en S3
 - El bucket de S3 tiene activado el control de versiones. Para obtener más información, consulte [Uso del control de versiones en buckets de S3](#)
 - El bucket de S3 puede calcular las sumas de comprobación del SHA-256 para los artefactos cargados. [Para obtener más información, consulte Uso de sumas de comprobación](#)

- Permisos para leer el código de un bucket de S3

Para obtener información sobre cómo crear el rol de servicio necesario, consulte [Cree un rol de servicio para leer el código de un bucket de S3 \(rol de plantilla de PySpark análisis\)](#).

El siguiente procedimiento describe el proceso de creación de una plantilla de PySpark análisis mediante la [AWS Clean Rooms consola](#). Se supone que ya ha creado un script de usuario y archivos de entorno virtual y que ha almacenado el script de usuario y los archivos del entorno virtual en un bucket de Amazon S3.

 Note

El miembro que crea la plantilla de PySpark análisis también debe ser el miembro que reciba los resultados.

Para obtener información sobre cómo crear una plantilla de PySpark análisis mediante la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

Para crear una plantilla PySpark de análisis

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#) con la Cuenta de AWS que funcionará como creador de la colaboración.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Plantillas, vaya a la sección Plantillas de análisis creadas por usted.
5. Seleccione Crear plantilla de análisis.
6. En la página Crear plantilla de análisis, en Detalles,
 - a. Introduzca un nombre para la plantilla de análisis.
 - b. (Opcional) Introduzca una Descripción.
 - c. En Formato, elija la PySpark opción.
7. En Definición,
 - a. Revise los requisitos previos y asegúrese de que se cumplen todos los requisitos previos antes de continuar.
 - b. Para el archivo de puntos de entrada, introduzca el bucket de S3 o seleccione Browse S3.

- c. (Opcional) Para el archivo de bibliotecas, introduzca el depósito de S3 o seleccione Browse S3.
8. Para los parámetros: opcional, si desea añadir parámetros para que la plantilla de análisis sea reutilizable:
- a. Elija Add parameter (Añadir parámetro).
 - b. Introduzca un nombre de parámetro.

Los nombres de los parámetros deben empezar con una letra o un guión bajo, seguidos de caracteres alfanuméricos o guiones bajos.

- c. En Tipo, STRING se selecciona automáticamente como el único tipo compatible para PySpark las plantillas de análisis.
- d. (Opcional) Introduzca un valor predeterminado para el parámetro.

Si proporciona un valor predeterminado, los ejecutores de tareas pueden usar este valor al ejecutar tareas sin proporcionar explícitamente un valor de parámetro.

- e. Para añadir más parámetros, elija Añadir otro parámetro y repita los pasos anteriores.

 Note

Puede definir hasta 50 parámetros por plantilla PySpark de análisis. El valor de cada parámetro puede tener un máximo de 1000 caracteres.

9. Para las tablas a las que se hace referencia en la definición,
- Si todas las tablas a las que se hace referencia en la definición se han asociado a la colaboración:
 - Deje seleccionada la casilla de verificación Todas las tablas a las que se hace referencia en la definición se han asociado a la colaboración.
 - En Tablas asociadas a la colaboración, seleccione todas las tablas asociadas a las que se hace referencia en la definición.
 - Si todas las tablas a las que se hace referencia en la definición no se han asociado a la colaboración:
 - Desactive la casilla de verificación Todas las tablas a las que se hace referencia en la definición se han asociado a la colaboración.

- En Tablas asociadas a la colaboración, seleccione todas las tablas asociadas a las que se hace referencia en la definición.
- En Tablas que se asociarán más adelante, introduzca un nombre de tabla.
- Seleccione Mostrar otra tabla para mostrar otra tabla.

10. Para configurar los mensajes de error, elija una de las siguientes opciones:

- Mensajes de error básicos: devuelve los mensajes de error básicos sin exponer los datos subyacentes. Recomendado para cargas de trabajo de producción.
- Mensajes de error detallados: devuelve mensajes de error detallados para una solución de problemas más rápida. Se recomienda en entornos de desarrollo y pruebas. Puede exponer datos confidenciales, incluida la información de identificación personal (PII).

 Note

Al utilizar mensajes de error detallados, todos los miembros del proveedor de datos deben aprobar esta configuración para la plantilla.

11. Especifique los permisos de acceso al servicio seleccionando el nombre de un rol de servicio existente en la lista desplegable.

1. Si tiene permisos de listas de roles, se mostrará la lista de roles.

Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar.

2. Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM.

Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible.

De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

 Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).
- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes AWS Clean Rooms. Debe agregar la política de rol antes de continuar.
- Si no puede modificar la política de rol, recibirá un mensaje de error que indica que AWS Clean Rooms no ha podido encontrar la política del rol de servicio.

12. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
13. Seleccione Crear.
14. Ahora puede informar a su miembro de la colaboración de que puede [revisar una plantilla de análisis](#). (opcional si desea consultar sus propios datos).

 Important

No modifique ni elimine artefactos (scripts de usuario o entornos virtuales) después de crear una plantilla de análisis.

Si lo hace, hará lo siguiente:

- Provoca que todos los futuros trabajos de análisis que utilicen esta plantilla fallen.
- Requiere la creación de una nueva plantilla de análisis con nuevos artefactos.
- No afecta a los trabajos de análisis completados anteriormente.

Revisión de una plantilla PySpark de análisis

Cuando otro miembro crea una plantilla de análisis en su colaboración, usted debe revisarla y aprobarla antes de poder utilizarla.

El siguiente procedimiento muestra cómo revisar una plantilla de PySpark análisis, incluidas sus reglas, parámetros y tablas a las que se hace referencia. Como miembro de la colaboración, evaluará si la plantilla se ajusta a sus acuerdos de intercambio de datos y a sus requisitos de seguridad.

Una vez aprobada la plantilla de análisis, se puede utilizar en un trabajo en AWS Clean Rooms.

 Note

Cuando incorpore su código de análisis a una colaboración, tenga en cuenta lo siguiente:

- AWS Clean Rooms no valida ni garantiza el comportamiento del código de análisis.
 - Si necesita garantizar un comportamiento determinado, revise directamente el código de su socio colaborador o póngase en contacto con un auditor externo de confianza para que lo revise.
- AWS Clean Rooms garantiza que los hashes SHA-256 del código que figuran en la plantilla de PySpark análisis coincidan con el código que se ejecuta en el PySpark entorno de análisis.
- AWS Clean Rooms no realiza ninguna auditoría ni análisis de seguridad de las bibliotecas adicionales que incorpore al entorno.
- En el modelo de seguridad compartida:
 - Usted (el cliente) es responsable de la seguridad del código que se ejecuta en el entorno.
 - Usted (el cliente) es responsable de establecer la configuración adecuada de los mensajes de error para el entorno.
 - AWS Clean Rooms es responsable de la seguridad del medio ambiente y garantiza que
 - solo se ejecuta el código aprobado
 - solo se puede acceder a las tablas configuradas especificadas
 - el único destino de salida es el depósito S3 del receptor de resultados.

AWS Clean Rooms genera hashes SHA-256 del script de usuario y del entorno virtual para su revisión. Sin embargo, no se puede acceder directamente al script de usuario ni a las bibliotecas en sí. AWS Clean Rooms

Para validar que el script de usuario y las bibliotecas compartidas son los mismos que los referenciados en la plantilla de análisis, puede crear un hash SHA-256 de los archivos compartidos y compararlo con el hash de la plantilla de análisis creado por AWS Clean Rooms. Los hashes del código ejecutado también estarán en los registros de trabajos.

Requisitos previos

- Sistema operativo Linux/Unix o subsistema Windows para Linux (WSL)
- Archivo de script de usuario que desea cifrar
 - Solicite al creador de la plantilla de análisis que comparta el archivo a través de un canal seguro.
- El hash de la plantilla de análisis creado por AWS Clean Rooms

Para revisar una plantilla de PySpark análisis mediante la AWS Clean Rooms consola

1. Inicie sesión en la [AWS Clean Rooms consola](#) [Consola de administración de AWS y ábrala](#) con la Cuenta de AWS que funcionará como creador de colaboraciones.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Seleccione la colaboración.
4. En la pestaña Plantillas, vaya a la sección Plantillas de análisis creadas por usted.
5. Elija la plantilla de análisis cuyo estado Se puede ejecutar sea No precisa su revisión.
6. Elija Revisar.
7. Revise la Descripción general, la definición y los Parámetros de la regla de análisis (si procede).

Note

Los parámetros permiten a los analizadores enviar valores diferentes en el momento del envío. Si una plantilla de análisis admite parámetros, revise cómo se utilizan los valores de los parámetros en el código de su socio colaborador para asegurarse de que cumple sus requisitos.

8. Compruebe que el script de usuario compartido y las bibliotecas sean los mismos a los que se hace referencia en la plantilla de análisis.
 - a. Cree un hash SHA-256 de los archivos compartidos y compárelo con el hash de la plantilla de análisis creado por AWS Clean Rooms

Para generar un hash, diríjase al directorio que contiene el archivo de script de usuario y, a continuación, ejecute el siguiente comando:

```
sha256sum your_script_filename.py
```

Ejemplo de código de salida:

```
e3b0c44298fc1c149afb4c8996fb92427ae41e4649b934ca495991b7852b855 my_analysis.py
```

- b. Como alternativa, puede utilizar las funciones de suma de comprobación de Amazon S3. Para obtener más información, consulte [Comprobar la integridad de los objetos en Amazon S3](#) en la Guía del usuario de Amazon S3.
 - c. Otra alternativa es ver los hashes del código ejecutado en los registros de trabajos.
9. Revise las tablas configuradas que se enumeran en Tablas a las que se hace referencia en la definición.
- El Estado que aparece junto a cada tabla indicará Plantilla no permitida.
10. Elija una tabla.
- a. Para aprobar la plantilla de análisis, elija Permitir la plantilla en la tabla. Para confirmar su aprobación, seleccione Permitir.
 - b. Para rechazar la aprobación, seleccione No permitir.

Si ha decidido aprobar la plantilla de análisis, el miembro que puede ejecutar los trabajos ahora puede ejecutar un PySpark trabajo en una tabla configurada mediante una plantilla de PySpark análisis. Para obtener más información, consulte [Ejecutando PySpark trabajos](#).

Plantillas de PySpark análisis para la solución

Al ejecutar trabajos con plantillas de PySpark análisis, es posible que se produzcan errores durante la inicialización o la ejecución del trabajo. Estos errores suelen estar relacionados con la configuración del script, los permisos de acceso a los datos o la configuración del entorno.

Para obtener más información sobre PySpark las limitaciones, consulte [PySpark limitaciones en AWS Clean Rooms](#).

Temas

- [Solución de problemas con el código](#)
- [El trabajo de la plantilla de análisis no se inicia](#)
- [El trabajo de la plantilla de análisis se inicia pero falla durante el procesamiento](#)
- [La configuración del entorno virtual falla](#)

Solución de problemas con el código

Para ayudarte a desarrollar tu código y solucionar sus problemas, te sugerimos que AWS Clean Rooms simules tú mismo Cuenta de AWS activando los mensajes de error detallados y ejecutando tareas con tus propios datos de prueba.

También puede simular PySpark AWS Clean Rooms en Amazon EMR Serverless con los siguientes pasos. Tendrá pequeñas diferencias con respecto a las salas limpias de AWS, pero PySpark en su mayoría trata sobre cómo se puede ejecutar el código.

Para simular PySpark AWS Clean Rooms en EMR Serverless

1. Cree un conjunto de datos en Amazon S3, catalóguelo en y configure los permisos de Lake Formation. AWS Glue Data Catalog
2. Registre la ubicación S3 en Lake Formation mediante un rol personalizado.
3. Cree una instancia de Amazon EMR Studio si aún no tiene una (se necesita Amazon EMR Studio para usar Amazon EMR Serverless).
4. Cree una aplicación EMR Serverless
 - Seleccione la versión de lanzamiento emr-7.7.0.
 - Seleccione la arquitectura. ARM64
 - Opte por usar configuraciones personalizadas.
 - Deshabilite la capacidad preinicializada.
 - Si planea realizar un trabajo interactivo, seleccione Punto final interactivo > Habilitar punto final para EMR Studio.
 - Seleccione Configuraciones adicionales > Use Lake Formation para un control de acceso detallado.
 - Cree la aplicación.
5. Utilice EMR-S a través de los portátiles de EMR-Studio o de la API. `StartJobRun`

El trabajo de la plantilla de análisis no se inicia

Causas habituales

Los trabajos de la plantilla de análisis pueden fallar inmediatamente al iniciarse debido a tres problemas de configuración principales:

- El nombre del script es incorrecto y no coincide con el formato requerido
- Falta la función de punto de entrada o tiene un formato incorrecto en el script de usuario
- Versión de Python incompatible en el entorno virtual

Resolución

Para resolverlo:

1. Compruebe su secuencia de comandos de usuario:

- Compruebe que el script de usuario tenga un nombre de archivo de Python válido.

Los nombres de archivo de Python válidos utilizan letras minúsculas, guiones bajos para separar las palabras y la extensión .py.

2. Compruebe la función de punto de entrada. Si el script de usuario no tiene una función de punto de entrada, añada una.

- a. Abre tu script de usuario.
- b. Añada esta función de punto de entrada:

```
def entrypoint(context):  
    # Your analysis code here
```

- c. Asegúrese de que el nombre de la función esté escrito exactamente igual. `entrypoint`
 - d. Compruebe que la función acepta el `context` parámetro.
3. Compruebe la compatibilidad de las versiones de Python:
 - a. Compruebe que su entorno virtual utiliza Python 3.9 o 3.11.
 - b. Para comprobar su versión, ejecute: `python --version`
 - c. Si es necesario, actualice su entorno virtual:

```
conda create -n analysis-env python=3.9
conda activate analysis-env
```

Prevención

- Utilice el código de inicio de la plantilla de análisis proporcionada, que incluye la estructura de archivos correcta.
- Configure un entorno virtual dedicado con Python 3.9 o 3.11 para todas las plantillas de análisis.
- Pruebe su plantilla de análisis localmente con la herramienta de validación de plantillas antes de enviar los trabajos.
- Implemente CI/CD comprobaciones para verificar los requisitos de nomenclatura de los scripts y de las funciones de punto de entrada.

El trabajo de la plantilla de análisis se inicia pero falla durante el procesamiento

Causas habituales

Los trabajos de análisis pueden fallar durante la ejecución por los siguientes motivos de seguridad y formato:

- Intentos de acceso directo no autorizado a AWS servicios como Amazon S3 o AWS Glue
- Devolver el resultado en formatos incorrectos que no coinciden con DataFrame las especificaciones requeridas
- Se bloquearon las llamadas de red debido a restricciones de seguridad en el entorno de ejecución

Resolución

Para resolver

1. Eliminar el acceso directo al AWS servicio:
 - a. Busca tu código para importar y llamar directamente a los AWS servicios.
 - b. Sustituya el acceso directo a S3 por los métodos de sesión de Spark proporcionados.

- c. Utilice únicamente tablas preconfiguradas a través de la interfaz de colaboración.
2. Formatee los resultados correctamente:
 - a. Verifica que todas las salidas sean Spark DataFrames.
 - b. Actualiza tu declaración de devolución para que coincida con este formato:

```
return {  
    "results": {  
        "output1": dataframe1  
    }  
}
```

- c. Elimine cualquier objeto que no se DataFrame devuelva.
3. Eliminar llamadas de red:
 - a. Identifique y elimine cualquier llamada a la API externa.
 - b. Elimine cualquier URL, solicitud o biblioteca de red similar.
 - c. Elimine cualquier conexión de socket o código de cliente HTTP.

Prevención

- Utilice el lector de códigos suministrado para comprobar si hay AWS importaciones y llamadas de red no autorizadas.
- Pruebe los trabajos en un entorno de desarrollo donde las restricciones de seguridad coincidan con las de producción.
- Siga el proceso de validación del esquema de salida antes de implementar los trabajos.
- Revise las directrices de seguridad para conocer los patrones de acceso a los servicios aprobados.

La configuración del entorno virtual falla

Causas habituales

Los errores de configuración del entorno virtual suelen producirse debido a:

- La arquitectura de CPU no coincide entre los entornos de desarrollo y ejecución
- Problemas de formato del código Python que impiden la inicialización adecuada del entorno

- Configuración incorrecta de la imagen base en los ajustes del contenedor

Resolución

Para resolverlo

1. Configure la arquitectura correcta:
 - a. Compruebe su arquitectura actual con `uname -m`.
 - b. Actualice su Dockerfile para especificar: ARM64

```
FROM --platform=linux/arm64 public.ecr.aws/amazonlinux/amazonlinux:2023-minimal
```

- c. Reconstruya su contenedor con `docker build --platform=linux/arm64`.
2. Corrija la hendidura de Python:
 - a. Ejecuta un formateador de código Python como `black` en tus archivos de código.
 - b. Verifica el uso coherente de los espacios o las pestañas (no de ambos).
 - c. Compruebe la hendidura de todos los bloques de códigos:

```
def my_function():  
    if condition:  
        do_something()  
    return result
```

- d. Utilice un IDE con resaltado de hendiduras de Python.
3. Valide la configuración del entorno:
 - a. Ejecute `python -m py_compile your_script.py` para comprobar si hay errores de sintaxis.
 - b. Pruebe el entorno localmente antes de la implementación.
 - c. Compruebe que todas las dependencias estén enumeradas en `requirements.txt`.

Prevención

- Use Visual Studio Code o PyCharm con complementos de formato de Python

- Configure los ganchos de preconfirmación para ejecutar los formateadores de código automáticamente
- Cree y pruebe entornos de forma local con la imagen base proporcionada ARM64
- Implemente la verificación automática del estilo del código en su CI/CD proceso

Analice los datos en una colaboración

En AWS Clean Rooms, puede analizar los datos mediante la ejecución de consultas o trabajos.

Una consulta es un método para acceder a las tablas configuradas en una colaboración y analizarlas, mediante un conjunto compatible de funciones, clases y variables. El lenguaje de consulta que se admite actualmente AWS Clean Rooms es SQL. Hay tres formas de ejecutar una consulta AWS Clean Rooms: escribir código SQL, utilizar una plantilla de análisis SQL aprobada o utilizar la interfaz de usuario de Analysis Builder.

Un trabajo es un método para acceder a las tablas configuradas y analizarlas en una colaboración mediante un conjunto compatible de funciones, clases y variables. El tipo de trabajo que se admite actualmente AWS Clean Rooms es PySpark. Hay una forma de ejecutar un trabajo AWS Clean Rooms: mediante una plantilla de PySpark análisis aprobada.

En los temas siguientes se describe cómo analizar los datos AWS Clean Rooms mediante la ejecución de consultas o PySpark trabajos de SQL.

Temas

- [Ejecución de consultas SQL](#)
- [Ejecutando PySpark trabajos](#)

Ejecución de consultas SQL

Note

Solo puede ejecutar consultas si el miembro responsable de pagar los costos de computación de las consultas se ha unido a la colaboración como miembro activo.

Como [miembro que puede realizar consultas](#), puede ejecutar una consulta SQL de las siguientes maneras:

- Creación manual de una consulta SQL mediante el editor de código SQL.
- Uso de una [plantilla de análisis](#) SQL aprobada.
- Uso de la interfaz de usuario de Analysis Builder para crear una consulta sin tener que escribir código SQL.

Cuando el miembro que puede realizar la consulta ejecuta una consulta SQL en las tablas de la colaboración, AWS Clean Rooms asume las funciones pertinentes para acceder a las tablas en su nombre. AWS Clean Rooms aplica las reglas de análisis necesarias a la consulta de entrada y a su salida.

Las reglas de análisis y las restricciones de salida se aplican automáticamente. AWS Clean Rooms solo devuelve los resultados que cumplen con las reglas de análisis definidas.

AWS Clean Rooms admite consultas SQL que pueden ser diferentes a las de otros motores de consultas. Para ver las especificaciones, consulte la [Referencia de SQL de AWS Clean Rooms](#). Si desea ejecutar consultas en tablas de datos protegidas con privacidad diferencial, debe asegurarse de que sus consultas sean compatibles con la [estructura de consultas de uso general](#) de la privacidad diferencial de AWS Clean Rooms .

Note

Cuando se utiliza la [computación criptográfica para Clean Rooms](#), no todas las operaciones SQL generan resultados válidos. Por ejemplo, puede realizar una operación COUNT en una columna cifrada, pero realizar una operación SUM en números cifrados generará errores. Además, las consultas también pueden arrojar resultados incorrectos. Por ejemplo, las consultas SUM de columnas selladas producen errores. Sin embargo, una consulta GROUP BY en columnas selladas parece ejecutarse correctamente, pero produce grupos diferentes a los que genera una consulta GROUP BY en texto sin formato.

Las consultas ejecutadas en la colaboración se cobran al [miembro que paga los costos de computación de las consultas](#).

El miembro que puede realizar la consulta puede seleccionar varios [miembros que pueden recibir los resultados](#) para recibir los resultados de una sola consulta. Para obtener más información, consulte [Consulta de tablas configuradas con el editor de código SQL](#). Para obtener información general sobre la recepción de los resultados de las consultas, consulte [Recepción y uso de los resultados del análisis](#).

Requisitos previos

Antes de ejecutar una consulta SQL, asegúrese de disponer de lo siguiente:

- Una membresía activa en una AWS Clean Rooms colaboración

- Acceso a al menos una tabla configurada de la colaboración
- Confirmación de que el miembro responsable de los costes informáticos de la consulta es un miembro activo de la colaboración

Para obtener información sobre cómo consultar datos o ver las consultas llamando directamente a la operación de la AWS Clean Rooms [StartProtectedQuery API](#) o utilizando la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

Para obtener más información sobre el registro de consultas, consulte [Inicio de sesión de Analysis AWS Clean Rooms](#).

 Note

Si ejecuta una consulta en tablas de datos [cifrados](#), los resultados de las columnas cifradas se cifran.

Configuración de propiedades de Spark para consultas SQL

AWS Clean Rooms permite personalizar de forma opcional el comportamiento en tiempo de ejecución de Spark mediante la configuración de las propiedades de Spark compatibles para las consultas SQL. Estas propiedades te permiten ajustar el rendimiento, el uso de memoria y los parámetros de ejecución de consultas. Con esta función, tienes un mayor control sobre cómo se procesan las consultas basadas en Spark, lo que te permite optimizarlas en función de tus requisitos específicos de carga de trabajo.

Ahora puedes ajustar ajustes como la reproducción aleatoria de particiones, los umbrales de unión de emisiones y los parámetros de ejecución de consultas adaptables directamente desde la consola. AWS Clean Rooms Esta función es especialmente útil para consultas complejas o conjuntos de datos grandes en los que las configuraciones predeterminadas pueden no ser óptimas. Al ajustar estas propiedades de Spark, puedes mejorar el rendimiento de las consultas, reducir el consumo de recursos y gestionar mejor el uso de la memoria para tus análisis de colaboración basados en Spark.

Para aprovechar esta función, encontrarás una nueva sección de propiedades de Spark en la interfaz de consultas. Puedes seleccionar de una lista de propiedades compatibles y especificar valores personalizados. [También puedes configurar las propiedades de Spark mediante programación mediante la StartProtectedQuery API](#). Esta opción de configuración avanzada permite a los analistas e ingenieros de datos optimizar sus consultas para mejorar la eficiencia y la escalabilidad.

Para obtener más información sobre las propiedades de Spark, incluidos los valores predeterminados, consulte [las propiedades de Spark](#) en la documentación de Apache Spark.

En los siguientes temas se explica cómo consultar datos en una colaboración mediante la consola de AWS Clean Rooms .

Temas

- [Consulta de tablas configuradas con el editor de código SQL](#)
- [Consulta de tablas de asignación de ID con el editor de código SQL](#)
- [Consulta de tablas configuradas mediante una plantilla de análisis SQL](#)
- [Consulta con el creador de análisis](#)
- [Visualización del impacto de la privacidad diferencial](#)
- [Vista de consultas recientes](#)
- [Visualización de detalles de consultas](#)

Consulta de tablas configuradas con el editor de código SQL

Como miembro que puede realizar consultas, puede crear una consulta manualmente escribiendo código SQL en el editor de código SQL. El editor de código SQL se encuentra en la sección Análisis de la pestaña Análisis de la AWS Clean Rooms consola.

El editor de código SQL se muestra de forma predeterminada. Si desea usar el creador de análisis para crear consultas, consulte [Consulta con el creador de análisis](#).

Important

Si empieza a escribir una consulta SQL en el editor de código y luego activa la interfaz de usuario del creador de análisis, la consulta no se guarda.

AWS Clean Rooms admite muchos comandos, funciones y condiciones de SQL. Para obtener más información, consulte la [Referencia de SQL de AWS Clean Rooms](#).

Tip

Si se lleva a cabo un mantenimiento programado mientras se está ejecutando una consulta, esta se termina y se revierte. Deberá reiniciar la consulta.

Consulta de tablas configuradas con el editor de código SQL

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> [Consola de administración de AWS](#) y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Realizar consultas.
4. En la pestaña Análisis, en Tablas, consulte la lista de tablas y el tipo de regla de análisis asociado (regla de análisis de agregación, regla de análisis de listas o regla de análisis personalizada).

Note

Si no ve las tablas que esperaba en la lista, puede deberse a los siguientes motivos:

- Las tablas no se han [asociado](#).
- Las tablas no tienen [una regla de análisis configurada](#).

5. (Opcional) Para ver el esquema de la tabla y los controles de las reglas de análisis, amplíe la tabla seleccionando el icono del signo más (+).
6. En la sección Análisis, en el modo Análisis, seleccione Escribir código SQL.

Note

La sección Análisis solo se muestra si el miembro que puede recibir los resultados y el miembro responsable de pagar los costos de computación de las consultas se han unido a la colaboración como miembros activos.

7. Cree la consulta escribiéndola en el editor de código SQL.

Para obtener más información sobre los comandos y funciones de SQL compatibles, consulte la [Referencia de AWS Clean Rooms SQL](#).

También puede usar las siguientes opciones para crear la consulta.

Use an example query

Para usar una consulta de ejemplo

1. Elija los tres puntos verticales que aparecen junto al nombre de la tabla.

2. En Insertar en el editor, elija Consulta de ejemplo.

 Note

Al insertar una consulta de ejemplo, se anexa a la consulta que ya está en el editor.

Aparece el ejemplo de consulta. Todas las tablas que aparecen en Tablas se incluyen en la consulta.

3. Edite los valores de marcadores de posición de la consulta.

Insert column names or functions

Para insertar un nombre de columna o una función

1. Seleccione los tres puntos verticales situados junto a una columna.
2. En Insertar en el editor, elija Nombre de columna.
3. Para insertar manualmente una función permitida en una columna,
 - a. Seleccione los tres puntos verticales situados junto a una columna.
 - b. Seleccione Insertar en el editor.
 - c. Seleccione el nombre de la función permitida (por ejemplo INNER JOIN, SUM, SUM DISTINCT, o COUNT).
4. Pulse Ctrl + Espacio para ver los esquemas de la tabla en el editor de código.

 Note

Los miembros que pueden realizar consultas pueden ver y usar las columnas de partición de cada asociación de tablas configuradas. Asegúrese de que la columna de particiones esté etiquetada como columna de particiones en la AWS Glue tabla subyacente a la tabla configurada.

5. Edite los valores de marcadores de posición de la consulta.
8. Especifique el tipo de trabajador admitido y el número de trabajadores.

Puede elegir el tipo de instancia y el número de instancias (trabajadores) para ejecutar sus consultas SQL.

Para CR.1X, puede seleccionar hasta 128 trabajadores o un mínimo de 4 trabajadores.

Para CR.4X, puede seleccionar hasta 32 trabajadores o un mínimo de 4 trabajadores.

Utilice la siguiente tabla para determinar el tipo y la cantidad de trabajadores que necesita para su caso de uso.

Tipo de empleado	vCPU	Memoria (GB)	Almacenamiento (GB)	Número de procesos de trabajo	Unidades totales de procesamiento de salas limpias (CRPU)
CR.1X (predeterminado)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

Los diferentes tipos y número de trabajadores tienen costos asociados. Para obtener más información sobre los precios, consulta [AWS Clean Rooms los precios](#).

9. En Enviar los resultados a, especifique quién puede recibir los resultados.

 Note

Para recibir los resultados, el miembro de la colaboración debe estar configurado como receptor de resultados y debe ser un participante activo en la colaboración (estado: Activo)

10. (El miembro solo puede realizar consultas) La casilla Usar la configuración de resultados predeterminada está seleccionada de forma predeterminada. Manténgala seleccionada si desea conservar la configuración de resultados predeterminada.

Si desea especificar una configuración de resultados diferente para esta consulta, desactive la casilla Usar la configuración de resultados predeterminada y, a continuación, seleccione lo siguiente.

- a. Formato de resultado (CSV o PARQUET)
- b. Archivos de resultados (individuales o múltiples)
- c. Destino de los resultados en Amazon S3

Cada miembro que pueda recibir los resultados puede especificar un formato de resultados, archivos de resultados y un destino de resultados diferentes en Amazon S3.

11. Para especificar las propiedades de Spark:

- a. Amplíe Propiedades de Spark.
- b. Selecciona Añadir propiedades de Spark.
- c. En el cuadro de diálogo de propiedades de Spark, elija un nombre de propiedad en la lista desplegable e introduzca un valor.

En la siguiente tabla se proporciona una definición para cada propiedad.

Para obtener más información sobre las propiedades de Spark, consulte [Propiedades de Spark](#) en la documentación de Apache Spark.

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Fallos en Spark.task.max	Controla cuántas veces consecutivas puede fallar una tarea antes de que el trabajo falle. Requiere un valor mayor o igual a 1. El número de reintentos permitidos es igual a este valor menos 1. El recuento de errores se restablece si algún intento tiene éxito. Los errores en las distintas tareas no se acumulan hasta alcanzar este límite.	4
spark.sql.files.maxPartitionBytes	Establece el número máximo de bytes que se deben empaquetar en una sola partición al leer fuentes basadas en archivos, como Parquet, JSON y ORC.	128 MB
Spark.Hadoop.FS.S3.max se reintenta	Establece el número máximo de reintentos para las operaciones con archivos de Amazon S3.	
spark.network.timeout	Establece el tiempo de espera predeterminado para todas las interacciones de la red. Anula los siguientes ajustes de tiempo de espera si no están configurados: • spark.storage.blockManagerHeartbeatTimeoutMs • Tiempo de espera de conexión de Spark.Shuffle.io • Spark.rpc.askTimeout • Se ha agotado el tiempo de espera de Spark.rpc Lookup	120 s

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.rdd.compress	Especifica si se deben comprimir las particiones RDD serializadas mediante spark.io.compression.codec. Se aplica a StorageLevel .MEMORY_ONLY_SER en Java y Scala, o a .MEMORY_ONLY en Python. StorageLevel Reduce el espacio de almacenamiento, pero requiere más tiempo de procesamiento de la CPU.	FALSO
spark.shuffle.spill.compress	Especifica si se deben comprimir los datos de Shuffle Spill mediante spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Establece el tamaño objetivo en bytes para las particiones aleatorias durante la optimización adaptativa cuando spark.sql.adaptive.enabled es true. Controla el tamaño de las particiones al unir particiones pequeñas o dividir particiones asimétricas.	(valor de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)
spark.sql.adaptive.autoBroadcastJoinUmbra	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Se aplica solo en un marco adaptativo. Utiliza el mismo valor predeterminado que spark.sql.adaptive.autoBroadcastJoinUmbra. Configúrelo en -1 para deshabilitar la transmisión.	(ninguno)
spark.sql.adaptive.coalescePartitions.enabled	Especifica si se van a unir particiones aleatorias contiguas en función de spark.sql.adaptive.advisoryPartitionSizeInBytes para optimizar el tamaño de la tarea. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Particiones compuestas por spark.sql.Adaptive.Coalesce.initialPartitionNum	Define el número inicial de particiones de mezcla antes de la fusión. Requiere que spark.sql.adaptive.enabled y spark.sql.Adaptive.coalescePartitions.enabled sean verdaderos. El valor predeterminado es spark.sql.shuffle.partition.	(ninguno)
Spark.sql.Adaptive.CoalescePartitions.minPartitionSize	Establece el tamaño mínimo de las particiones aleatorias combinadas para evitar que las particiones se vuelvan demasiado pequeñas durante la optimización adaptativa.	1 MB
spark.sql.adaptive.coalescePartitions.ParalelismFirst	Especifica si se deben calcular los tamaños de las particiones en función del paralelismo del clúster en lugar de en spark.sql.adaptive.advisoryPartitionSizeInBytes durante la fusión de particiones. Genera tamaños de partición más pequeños que el tamaño objetivo configurado para maximizar el paralelismo. Se recomienda establecer este valor en false en los clústeres ocupados para mejorar la utilización de los recursos y evitar el exceso de tareas pequeñas.	TRUE
spark.sql.adaptive.enabled	Especifica si se debe habilitar la ejecución adaptativa de consultas para volver a optimizar los planes de consultas durante la ejecución de las consultas, en función de estadísticas de tiempo de ejecución precisas.	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoin	Especifica si se debe forzar la activación OptimizeSkewedJoin incluso si se introduce una mezcla adicional.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive.localShuffleReader.habilitado	Especifica si se deben utilizar lectores aleatorios locales cuando no se requiera la creación de particiones aleatorias, por ejemplo, después de la conversión de uniones por ordenación y combinación a uniones por hash de difusión. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE
spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold	Establece el tamaño máximo de partición en bytes para crear mapas hash locales. Prioriza las uniones hash mezcladas sobre las uniones por ordenación y combinación cuando: • Este valor es igual o superior a spark.sql.adaptive.advisoryPartitionSizeInBytes • Todos los tamaños de partición están dentro de este límite Anula spark.sql.join.preferSortMergeJoin a la configuración.	0 bytes
spark.sql.adaptive.optimizeSkewsInRebalancePartitions.habilitado	Especifica si se deben optimizar las particiones aleatorias asimétricas dividiéndolas en particiones más pequeñas según spark.sql.adaptive.advisoryPartitionSizeInBytes. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE
spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor	Define el factor de umbral de tamaño para la fusión de particiones durante la división. Las particiones más pequeñas que este factor se multiplican por spark.sql.adaptive.advisoryPartitionSizeInBytes están fusionadas.	0.2

Nombre de propiedad	Description (Descripción)	Valor predeterminado
<code>spark.sql.adaptive.skewJoin.enabled</code>	Especifica si se debe tratar la asimetría de datos en las uniones mezcladas dividiendo y, opcionalmente, replicando las particiones asimétricas. Se aplica a las uniones hash ordenadas, combinadas y barajadas. Requiere que <code>spark.sql.adaptive.enabled</code> sea verdadero.	TRUE
<code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code>	Determina el factor de tamaño que determina el sesgo de la partición. Una partición está sesgada cuando su tamaño supera las dos condiciones: • Este factor se multiplica por la mediana del tamaño de la partición • El valor de <code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code>	5
<code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code>	Establece el umbral de tamaño en bytes para identificar las particiones asimétricas. Una partición está sesgada cuando su tamaño supera las dos condiciones: • Este umbral • El tamaño medio de la partición multiplicado por <code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code> Recomendamos establecer este valor por encima de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>.	256 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.autoBroadcastJoinThreshold	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Configúrelo en -1 para deshabilitar la transmisión.	10 MB
Spark.sql.WaitTimeBeforeTriggerEmission	Controla el tiempo de espera en segundos para las operaciones de transmisión durante las uniones de transmisión.	300 segundos
spark.sql.cbo.enabled	Especifica si se debe habilitar la optimización basada en costes (CBO) para la estimación de las estadísticas del plan.	FALSO
spark.sql.cbo.joinReorder.dp.star.filter	Especifica si se debe aplicar la heurística del filtro de unión en estrella durante la enumeración de uniones basada en el coste.	FALSO
spark.sql.cbo.joinReorder.dp.Threshold	Establece el número máximo de nodos unidos permitidos en el algoritmo de programación dinámica.	12
spark.sql.cbo.joinReorder.Enabled	Especifica si se debe habilitar el reordenamiento de las uniones en la optimización basada en costes (CBO).	FALSO
Spark.sql.cbo.PlanStats.Enabled	Especifica si se deben obtener los recuentos de filas y las estadísticas de las columnas del catálogo durante la generación del plan lógico.	FALSO
spark.sql.cbo.starSchemaDetection	Especifica si se debe habilitar el reordenamiento de las uniones en función de la detección del esquema en estrella.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.files.maxPartitionNum	Establece el número máximo objetivo de particiones de archivos divididas para fuentes basadas en archivos (Parquet, JSON y ORC). Cambia la escala de las particiones cuando el recuento inicial supera este valor. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.files.maxRecordsPerArchivo	Establece el número máximo de registros que se van a escribir en un único archivo. No se aplica ningún límite cuando se establece en cero o en un valor negativo.	0
spark.sql.files.minPartitionNum	Establece el número mínimo objetivo de particiones de archivos divididas para las fuentes basadas en archivos (Parquet, JSON y ORC). El valor predeterminado es spark.sql.leafNodeDefaultParalelismo. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.inMemoryColumnarAlmacenamiento.Tamaño del lote	Controla el tamaño del lote para el almacenamiento en caché en columnas. El aumento del tamaño mejora la utilización y la compresión de la memoria, pero aumenta el riesgo de out-of-memory errores.	10000
spark.sql.inMemoryColumnarAlmacenamiento.Comprimido	Especifica si se van a seleccionar automáticamente los códecs de compresión para las columnas en función de las estadísticas de datos.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	Especifica si se debe habilitar la lectura vectorizada para el almacenamiento en caché de columnas.	TRUE
spark.sql.legacy.allowHashOnMapType	Especifica si se permiten operaciones de hash en estructuras de datos de tipo mapa. Esta configuración antigua mantiene la compatibilidad con el manejo de tipos de mapas de las versiones anteriores de Spark.	
spark.sql.legacy.allowNegativeScaleOfDecimal	Especifica si se permiten valores de escala negativos en las definiciones de tipo decimal. Esta configuración antigua mantiene la compatibilidad con las versiones anteriores de Spark que admitían escalas decimales negativas.	
spark.sql.legacy.castComplexTypesToString.habilitado	Especifica si se debe habilitar el comportamiento heredado para convertir tipos complejos en cadenas. Mantiene la compatibilidad con las reglas de conversión de tipos de las versiones anteriores de Spark.	
spark.sql.legacy.charVarcharAsCadena	Especifica si se deben tratar los tipos CHAR y VARCHAR como tipos STRING. Esta configuración antigua proporciona compatibilidad con el manejo de tipos de cadenas de las versiones anteriores de Spark.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.createEmptyCollectionUsingStringType	Especifica si se van a crear colecciones vacías mediante elementos de tipo cadena. Esta configuración antigua mantiene la compatibilidad con el comportamiento de inicialización de colecciones de versiones anteriores de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal. Habilitado	Especifica si se deben interpretar los literales exponenciales como tipos decimales. Esta configuración antigua mantiene la compatibilidad con el manejo literal numérico de las versiones anteriores de Spark.	
spark.sql.legacy.json.allowEmptyString.habilitado	Especifica si se permiten cadenas vacías en el procesamiento de JSON. Esta configuración antigua mantiene la compatibilidad con el comportamiento de análisis de JSON de las versiones anteriores de Spark.	
spark.sql.legacy.parquet.int96RebaseModelRead	Especifica si se debe utilizar el modo de rebase de marcas de tiempo heredado al leer los archivos de Parquet. INT96 Esta configuración antigua mantiene la compatibilidad con el manejo de las marcas de tiempo de las versiones anteriores de Spark.	
spark.sql.legacy.timeParserPolicy	Controla el comportamiento del análisis temporal para garantizar la compatibilidad con versiones anteriores. Esta configuración antigua determina cómo se analizan las marcas de tiempo y las fechas a partir de las cadenas.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.TypeCoercion.datetimeToString.habilitado	Especifica si se debe habilitar el comportamiento de coerción de tipo heredado al convertir valores de fecha y hora en cadenas. Mantiene la compatibilidad con las reglas de conversión de fecha y hora de las versiones anteriores de Spark.	
spark.sql.maxSinglePartitionBytes	Establece el tamaño máximo de la partición en bytes. El planificador introduce operaciones de mezcla para particiones más grandes a fin de mejorar el paralelismo.	128 m
Caché de metadatos Spark.sql.TTLSeconds	Controla el time-to-live (TTL) de las cachés de metadatos. Se aplica a los metadatos de los archivos de partición y a las cachés del catálogo de sesiones. Requiere: • Un valor positivo superior a cero • Spark.sql.CatalogImplementation está configurado como hive • spark.sql.hive.filesourcePartitionFileCacheSize mayor que cero • spark.sql.hive.manageFilesourcePartitions establecido en verdadero	-1000 ms
spark.sql.optimizer.collapseProjectAlwaysEn línea	Especifica si se deben contraer las proyecciones adyacentes y las expresiones en línea, incluso cuando esto provoque duplicación.	FALSO
spark.sql.optimizer.dynamicPartitionPruning.habilitado	Especifica si se van a generar predicados para las columnas de partición utilizadas como claves de unión.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimizer.enableCsvExpressionOptimización	Especifica si se deben optimizar las expresiones CSV en el optimizador de SQL eliminando las columnas innecesarias de las operaciones from_csv.	TRUE
spark.sql.optimizer.enableJsonExpressionOptimización	Especifica si se deben optimizar las expresiones JSON en el optimizador de SQL de la siguiente manera: • Eliminar columnas innecesarias de las operaciones de from_json • Simplificar las combinaciones from_json y to_json • Optimización de las operaciones de named_struct	TRUE
Spark.sql.Optimizer.ExcludedRules	Define las reglas del optimizador que se van a deshabilitar, identificadas mediante nombres de reglas separados por comas. Algunas reglas no se pueden deshabilitar porque son necesarias para que sean correctas. El optimizador registra qué reglas se han desactivado correctamente.	(ninguno)
spark.sql.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Establece el tamaño mínimo de escaneo agregado en bytes necesario para inyectar un filtro Bloom en el lado de la aplicación.	10 GB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Filtro Spark.Sql.Optimizer.Runtime.BloomFilter.creationSideThreshold	Define el umbral de tamaño máximo para inyectar un filtro Bloom en el lado de la creación.	10 MB
spark.sql.Optimizer.Runtime.BloomFilter.Enabled	Especifica si se debe insertar un filtro Bloom para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	TRUE
spark.sql.optimizer.Runtime.BloomFilter.expectedNumItems	Define el número predeterminado de elementos esperados en el filtro Bloom en tiempo de ejecución.	1000000
spark.sql.Optimizer.Runtime.BloomFilter.maxNumBits	Establece el número máximo de bits permitido en el filtro Bloom en tiempo de ejecución.	67108864
Spark.Sql.Optimizer.Runtime.BloomFilter.maxNumItems	Establece el número máximo de elementos esperados permitidos en el filtro Bloom en tiempo de ejecución.	4000000
spark.sql.Optimizer.Runtime.BloomFilter.Number.Threshold	Limita el número máximo de filtros de tiempo de ejecución distintos de DPP permitidos por consulta para evitar errores en el controlador. out-of-memory	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Define el número predeterminado de bits que se utilizan en el filtro Bloom en tiempo de ejecución.	8388608

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimize r.runtime. rowlevelO perationGroupFiltro. Habilitado	Especifica si se debe habilitar el filtrado de grupos en tiempo de ejecución para las operaciones a nivel de fila. Permite que las fuentes de datos: • Elimine grupos completos de datos (como archivos o particiones) mediante filtros de fuentes de datos • Ejecute consultas en tiempo de ejecución para identificar los registros coincidentes • Deseche los grupos innecesarios para evitar costosas reescrituras Limitaciones: • No todas las expresiones se pueden convertir en filtros de fuentes de datos • Algunas expresiones requieren la evaluación de Spark (como las subconsultas)	TRUE
spark.sql.Optimize r.RuntimeFilter. semiJoinReduction. habilitado	Especifica si se debe insertar una semiunión para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.parquet.aggregatePushdown	Especifica si se deben enviar los agregados a Parquet para su optimización. Soportes: • MIN y MAX para los tipos booleano, entero, flotante y de fecha • COUNT para todos los tipos de datos Lanza una excepción si faltan estadísticas en el pie de página de un archivo de Parquet.	FALSO
spark.sql.parquet.columnarReaderBatchSize	Controla el número de filas de cada lote de lectores vectorizados Parquet. Elija un valor que equilibre la sobrecarga de rendimiento y el uso de memoria para evitar out-of-memory errores.	4096
Spark.sql.Session.TimeZone	Define la zona horaria de la sesión para gestionar las marcas de tiempo en los literales de cadena y la conversión de objetos Java. Acepta: • area/city Formato basado IDs en regiones (como America/Los_Angeles) • Desplazamientos de zona en formato (+/-) HH, (+/-) HH:mm o (+/-) HH:mm:ss (como -08 o + 01:00) • UTC o Z como alias para + 00:00	(valor de la zona horaria local)
spark.sql.shuffle.partition	Establece el número predeterminado de particiones para la mezcla de datos durante las uniones o agregaciones. No se puede modificar entre reinicios de consultas de streaming estructuradas desde la misma ubicación del punto de control.	200

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.shuffledHashJoinFactor	Define el factor de multiplicación utilizado para determinar la idoneidad de las combinaciones de hash aleatorio. Se selecciona una combinación de hash aleatorio cuando el tamaño de los datos del lado pequeño multiplicado por este factor es menor que el tamaño de los datos del lado grande.	3
spark.sql.sources.parallelPartitionDiscovery.umbra	Establece el número máximo de rutas para la lista de archivos del lado del conductor con fuentes basadas en archivos (Parquet, JSON y ORC). Si se supera durante la detección de particiones, los archivos se enumeran mediante un trabajo distribuido de Spark independiente.	32
spark.sql.statistics.histogram.enabled	Especifica si se deben generar histogramas con la misma altura durante el cálculo de las estadísticas de columnas para mejorar la precisión de la estimación. Requiere un escaneo de tabla adicional al necesario para las estadísticas de columnas básicas.	FALSO

12. Seleccione Ejecutar.

Note

No podrá ejecutar la consulta si el miembro que puede recibir los resultados no ha configurado los ajustes de resultados de consulta.

13. Consulte los resultados.

Para obtener más información, consulte [Recepción y uso de los resultados del análisis](#).

14. Siga ajustando los parámetros y vuelva a ejecutar la consulta, o pulse el botón + para iniciar una nueva consulta en una pestaña nueva.

Note

AWS Clean Rooms tiene como objetivo proporcionar mensajes de error claros. Si un mensaje de error no contiene detalles suficientes para ayudarle a solucionar el problema, póngase en contacto con el equipo de cuentas. Proporciónese una descripción de cómo se produjo el error y el mensaje de error (incluidos los identificadores). Para obtener más información, consulte [Solución de problemas AWS Clean Rooms](#).

Consulta de tablas de asignación de ID con el editor de código SQL

El siguiente procedimiento describe cómo ejecutar una consulta de unión de varias tablas en la tabla de asignación de ID para unir `sourceId` con `targetId`.

Antes de consultar la tabla de asignación de ID, la tabla de asignación de ID debe rellenarse correctamente.

Consulta de tablas de asignación de ID con el editor de código SQL

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Realizar consultas.
4. En la pestaña Análisis, vaya a la sección Análisis.

Note

La sección Análisis solo se muestra si el miembro que puede recibir los resultados y el miembro responsable de pagar los costos de computación de las consultas se han unido a la colaboración como miembros activos.

5. En la pestaña Análisis, en Tablas, consulte la lista de tablas de mapeo de ID (en Administrado por AWS Clean Rooms) y el tipo de regla de análisis asociado (regla de análisis de tablas de mapeo de ID).

 **Note**

Si no ve las tablas de asignación de ID que espera en la lista, es posible que sea porque las tablas de asignación de ID no se han rellenado correctamente. Para obtener más información, consulte [Llenado de una tabla de asignación de ID existente](#).

6. Cree la consulta escribiéndola en el editor de código SQL.

(Opcional) Si desea utilizar una consulta de ejemplo	(Opcional) Si desea insertar un nombre de la tabla
 Note Al insertar un ejemplo de sentencias JOIN, se añade la consulta que ya está en el editor. 1. Elija los tres puntos verticales que aparecen junto al nombre de la tabla. 2. En Insertar en el editor, elija Instrucción JOIN de ejemplo. Aparece la instrucción JOIN de ejemplo. 3. Edite los valores de marcadores de posición de la consulta.	1. Seleccione los tres puntos verticales situados junto a una columna. 2. En Insertar en el editor, elija Nombre de tabla. 3. Edite los valores de marcadores de posición de la consulta.

7. Especifique el tipo de trabajador admitido y el número de trabajadores.

Utilice la siguiente tabla para determinar el tipo y la cantidad de trabajadores que necesita para su caso de uso.

Tipo de empleado	vCPU	Memoria (GB)	Almacenamiento (GB)	Número de procesos de trabajo	Unidades totales de procesamiento de salas limpias (CRPU)
CR.1X (predeterminado)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

Note

Los diferentes tipos y número de trabajadores tienen costos asociados. Para obtener más información sobre los precios, consulta [AWS Clean Rooms los precios](#).

8. Seleccione Ejecutar.

Note

No podrá ejecutar la consulta si el miembro que puede recibir los resultados no ha configurado los ajustes de resultados de consulta.

9. Consulte los resultados.

Para obtener más información, consulte [Recepción y uso de los resultados del análisis](#).

10. Siga ajustando los parámetros y vuelva a ejecutar la consulta, o pulse el botón + para iniciar una nueva consulta en una pestaña nueva.

 Note

AWS Clean Rooms tiene como objetivo proporcionar mensajes de error claros. Si un mensaje de error no contiene detalles suficientes para ayudarle a solucionar el problema, póngase en contacto con el equipo de cuentas. Proporciónese una descripción de cómo se produjo el error y el mensaje de error (incluidos los identificadores). Para obtener más información, consulte [Solución de problemas AWS Clean Rooms](#).

Consulta de tablas configuradas mediante una plantilla de análisis SQL

Este procedimiento muestra cómo utilizar una plantilla de análisis en la AWS Clean Rooms consola para consultar las tablas configuradas con la regla de análisis personalizada.

Para utilizar una plantilla de análisis SQL para consultar tablas configuradas con la regla de análisis personalizada

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Realizar consultas.
4. En la pestaña Análisis, en la sección Tablas, consulte las tablas y el tipo de regla de análisis asociado (regla de análisis personalizada).

 Note

Si no ve las tablas que esperaba en la lista, puede deberse a los siguientes motivos:

- Las tablas no se han [asociado](#).
- Las tablas no tienen [una regla de análisis configurada](#).

5. En la sección Análisis, en el modo Análisis, seleccione Ejecutar plantillas de análisis y, a continuación, elija la plantilla de análisis de la lista desplegable.
6. Los parámetros de la plantilla de análisis SQL se rellenarán automáticamente en la definición.
7. Especifique el tipo de trabajador admitido y el número de trabajadores.

Utilice la siguiente tabla para determinar el tipo y la cantidad de trabajadores que necesita para su caso de uso.

Tipo de empleado	vCPU	Memoria (GB)	Almacenamiento (GB)	Número de procesos de trabajo	Unidades totales de procesamiento de salas limpias (CRPU)
CR.1X (predeterminado)	4	30	100	4	8
				128	256
CR.4X	16	120	400	4	32
				32	256

 Note

Los diferentes tipos y número de trabajadores tienen costos asociados. Para obtener más información sobre los precios, consulta [AWS Clean Rooms los precios](#).

8. Especifica las propiedades de Spark compatibles.
 - a. Selecciona Añadir propiedades de Spark.
 - b. En el cuadro de diálogo de propiedades de Spark, elija un nombre de propiedad en la lista desplegable e introduzca un valor.

En la siguiente tabla se proporciona una definición para cada propiedad.

Para obtener más información sobre las propiedades de Spark, consulte [Propiedades de Spark](#) en la documentación de Apache Spark.

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Fallos en Spark.task.max	Controla cuántas veces consecutivas puede fallar una tarea antes de que el trabajo falle. Requiere un valor mayor o igual a 1. El número de reintentos permitidos es igual a este valor menos 1. El recuento de errores se restablece si algún intento tiene éxito. Los errores en las distintas tareas no se acumulan hasta alcanzar este límite.	4
spark.sql.files.maxPartitionBytes	Establece el número máximo de bytes que se deben empaquetar en una sola partición al leer fuentes basadas en archivos, como Parquet, JSON y ORC.	128 MB
Spark.Hadoop.FS.S3.max se reintenta	Establece el número máximo de reintentos para las operaciones con archivos de Amazon S3.	
spark.network.timeout	Establece el tiempo de espera predeterminado para todas las interacciones de la red. Anula los siguientes ajustes de tiempo de espera si no están configurados: • spark.storage.blockManagerHeartbeatTimeoutMs • Tiempo de espera de conexión de Spark.Shuffle.io • Spark.rpc.askTimeout • Se ha agotado el tiempo de espera de Spark.rpc Lookup	120 s

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.rdd.compress	Especifica si se deben comprimir las particiones RDD serializadas mediante spark.io.compression.codec. Se aplica a StorageLevel .MEMORY_ONLY_SER en Java y Scala, o a .MEMORY_ONLY en Python. StorageLevel Reduce el espacio de almacenamiento, pero requiere más tiempo de procesamiento de la CPU.	FALSO
spark.shuffle.spill.compress	Especifica si se deben comprimir los datos de Shuffle Spill mediante spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Establece el tamaño objetivo en bytes para las particiones aleatorias durante la optimización adaptativa cuando spark.sql.adaptive.enabled es true. Controla el tamaño de las particiones al unir particiones pequeñas o dividir particiones asimétricas.	(valor de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)
spark.sql.adaptive.autoBroadcastJoinUmbra	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Se aplica solo en un marco adaptativo. Utiliza el mismo valor predeterminado que spark.sql.adaptive.autoBroadcastJoinUmbra. Configúrelo en -1 para deshabilitar la transmisión.	(ninguno)
spark.sql.adaptive.coalescePartitions.enabled	Especifica si se van a unir particiones aleatorias contiguas en función de spark.sql.adaptive.advisoryPartitionSizeInBytes para optimizar el tamaño de la tarea. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Particiones compuestas por spark.sql.Adaptive.Coalesce.initialPartitionNum	Define el número inicial de particiones de mezcla antes de la fusión. Requiere que spark.sql.adaptive.enabled y spark.sql.Adaptive.coalescePartitions.enabled sean verdaderos. El valor predeterminado es spark.sql.shuffle.partition.	(ninguno)
Spark.sql.Adaptive.CoalescePartitions.minPartitionSize	Establece el tamaño mínimo de las particiones aleatorias combinadas para evitar que las particiones se vuelvan demasiado pequeñas durante la optimización adaptativa.	1 MB
spark.sql.adaptive.coalescePartitions.ParalelismFirst	Especifica si se deben calcular los tamaños de las particiones en función del paralelismo del clúster en lugar de en spark.sql.adaptive.advisoryPartitionSizeInBytes durante la fusión de particiones. Genera tamaños de partición más pequeños que el tamaño objetivo configurado para maximizar el paralelismo. Se recomienda establecer este valor en false en los clústeres ocupados para mejorar la utilización de los recursos y evitar el exceso de tareas pequeñas.	TRUE
spark.sql.adaptive.enabled	Especifica si se debe habilitar la ejecución adaptativa de consultas para volver a optimizar los planes de consultas durante la ejecución de las consultas, en función de estadísticas de tiempo de ejecución precisas.	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoin	Especifica si se debe forzar la activación OptimizeSkewedJoin incluso si se introduce una mezcla adicional.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive.localShuffleReader.habilitado	Especifica si se deben utilizar lectores aleatorios locales cuando no se requiera la creación de particiones aleatorias, por ejemplo, después de la conversión de uniones por ordenación y combinación a uniones por hash de difusión. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE
spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold	Establece el tamaño máximo de partición en bytes para crear mapas hash locales. Prioriza las uniones hash mezcladas sobre las uniones por ordenación y combinación cuando: • Este valor es igual o superior a spark.sql.adaptive.advisoryPartitionSizeInBytes • Todos los tamaños de partición están dentro de este límite Anula spark.sql.join.preferSortMergeJoin a la configuración.	0 bytes
spark.sql.adaptive.optimizeSkewsInRebalancePartitions.habilitado	Especifica si se deben optimizar las particiones aleatorias asimétricas dividiéndolas en particiones más pequeñas según spark.sql.adaptive.advisoryPartitionSizeInBytes. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE
spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor	Define el factor de umbral de tamaño para la fusión de particiones durante la división. Las particiones más pequeñas que este factor se multiplican por spark.sql.adaptive.advisoryPartitionSizeInBytes están fusionadas.	0.2

Nombre de propiedad	Description (Descripción)	Valor predeterminado
<code>spark.sql.adaptive.skewJoin.enabled</code>	Especifica si se debe tratar la asimetría de datos en las uniones mezcladas dividiendo y, opcionalmente, replicando las particiones asimétricas. Se aplica a las uniones hash ordenadas, combinadas y barajadas. Requiere que <code>spark.sql.adaptive.enabled</code> sea verdadero.	TRUE
<code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code>	Determina el factor de tamaño que determina el sesgo de la partición. Una partición está sesgada cuando su tamaño supera las dos condiciones: • Este factor se multiplica por la mediana del tamaño de la partición • El valor de <code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code>	5
<code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code>	Establece el umbral de tamaño en bytes para identificar las particiones asimétricas. Una partición está sesgada cuando su tamaño supera las dos condiciones: • Este umbral • El tamaño medio de la partición multiplicado por <code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code> Recomendamos establecer este valor por encima de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> .	256 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.autoBroadcastJoinThreshold	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Configúrelo en -1 para deshabilitar la transmisión.	10 MB
Spark.sql.Tiempo de espera de emisión	Controla el tiempo de espera en segundos para las operaciones de transmisión durante las uniones de transmisión.	300 segundos
spark.sql.cbo.enabled	Especifica si se debe habilitar la optimización basada en costes (CBO) para la estimación de las estadísticas del plan.	FALSO
spark.sql.cbo.joinReorder.dp.star.filter	Especifica si se debe aplicar la heurística del filtro de unión en estrella durante la enumeración de uniones basada en el coste.	FALSO
spark.sql.cbo.joinReorder.dp.Threshold	Establece el número máximo de nodos unidos permitidos en el algoritmo de programación dinámica.	12
spark.sql.cbo.joinReorder.Enabled	Especifica si se debe habilitar el reordenamiento de las uniones en la optimización basada en costes (CBO).	FALSO
Spark.sql.cbo.PlanStats.Enabled	Especifica si se deben obtener los recuentos de filas y las estadísticas de las columnas del catálogo durante la generación del plan lógico.	FALSO
spark.sql.cbo.starSchemaDetection	Especifica si se debe habilitar el reordenamiento de las uniones en función de la detección del esquema en estrella.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.files.maxPartitionNum	Establece el número máximo objetivo de particiones de archivos divididas para fuentes basadas en archivos (Parquet, JSON y ORC). Cambia la escala de las particiones cuando el recuento inicial supera este valor. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.files.maxRecordsPerArchivo	Establece el número máximo de registros que se van a escribir en un único archivo. No se aplica ningún límite cuando se establece en cero o en un valor negativo.	0
spark.sql.files.minPartitionNum	Establece el número mínimo objetivo de particiones de archivos divididas para las fuentes basadas en archivos (Parquet, JSON y ORC). El valor predeterminado es spark.sql.leafNodeDefaultParalelismo. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.inMemoryColumnarAlmacenamiento.Tamaño del lote	Controla el tamaño del lote para el almacenamiento en caché en columnas. El aumento del tamaño mejora la utilización y la compresión de la memoria, pero aumenta el riesgo de out-of-memory errores.	10000
spark.sql.inMemoryColumnarAlmacenamiento.Comprimido	Especifica si se van a seleccionar automáticamente los códecs de compresión para las columnas en función de las estadísticas de datos.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	Especifica si se debe habilitar la lectura vectorizada para el almacenamiento en caché de columnas.	TRUE
spark.sql.legacy.allowHashOnMapType	Especifica si se permiten operaciones de hash en estructuras de datos de tipo mapa. Esta configuración antigua mantiene la compatibilidad con el manejo de tipos de mapas de las versiones anteriores de Spark.	
spark.sql.legacy.allowNegativeScaleOfDecimal	Especifica si se permiten valores de escala negativos en las definiciones de tipo decimal. Esta configuración antigua mantiene la compatibilidad con las versiones anteriores de Spark que admitían escalas decimales negativas.	
spark.sql.legacy.castComplexTypesToString.habilitado	Especifica si se debe habilitar el comportamiento heredado para convertir tipos complejos en cadenas. Mantiene la compatibilidad con las reglas de conversión de tipos de las versiones anteriores de Spark.	
spark.sql.legacy.charVarcharAsCadena	Especifica si se deben tratar los tipos CHAR y VARCHAR como tipos STRING. Esta configuración antigua proporciona compatibilidad con el manejo de tipos de cadenas de las versiones anteriores de Spark.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.createEmptyCollectionUsingStringType	Especifica si se van a crear colecciones vacías mediante elementos de tipo cadena. Esta configuración antigua mantiene la compatibilidad con el comportamiento de inicialización de colecciones de versiones anteriores de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal. Habilitado	Especifica si se deben interpretar los literales exponenciales como tipos decimales. Esta configuración antigua mantiene la compatibilidad con el manejo literal numérico de las versiones anteriores de Spark.	
spark.sql.legacy.json.allowEmptyString.habilitado	Especifica si se permiten cadenas vacías en el procesamiento de JSON. Esta configuración antigua mantiene la compatibilidad con el comportamiento de análisis de JSON de las versiones anteriores de Spark.	
spark.sql.legacy.parquet.int96RebaseModelRead	Especifica si se debe utilizar el modo de rebase de marcas de tiempo heredado al leer los archivos de Parquet. INT96 Esta configuración antigua mantiene la compatibilidad con el manejo de las marcas de tiempo de las versiones anteriores de Spark.	
spark.sql.legacy.timeParserPolicy	Controla el comportamiento del análisis temporal para garantizar la compatibilidad con versiones anteriores. Esta configuración antigua determina cómo se analizan las marcas de tiempo y las fechas a partir de las cadenas.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.TypeCoercion.datetimeToString.habilitado	Especifica si se debe habilitar el comportamiento de coerción de tipo heredado al convertir valores de fecha y hora en cadenas. Mantiene la compatibilidad con las reglas de conversión de fecha y hora de las versiones anteriores de Spark.	
spark.sql.maxSinglePartitionBytes	Establece el tamaño máximo de la partición en bytes. El planificador introduce operaciones de mezcla para particiones más grandes a fin de mejorar el paralelismo.	128 m
Caché de metadatos Spark.sql.TTLSeconds	Controla el time-to-live (TTL) de las cachés de metadatos. Se aplica a los metadatos de los archivos de partición y a las cachés del catálogo de sesiones. Requiere: • Un valor positivo superior a cero • Spark.sql.CatalogImplementation está configurado como hive • spark.sql.hive.filesourcePartitionFileCacheSize mayor que cero • spark.sql.hive.manageFilesourcePartitions establecido en verdadero	-1000 ms
spark.sql.optimizer.collapseProjectAlwaysEn línea	Especifica si se deben contraer las proyecciones adyacentes y las expresiones en línea, incluso cuando esto provoque duplicación.	FALSO
spark.sql.optimizer.dynamicPartitionPruning.habilitado	Especifica si se van a generar predicados para las columnas de partición utilizadas como claves de unión.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimizer.enableCsvExpressionOptimización	Especifica si se deben optimizar las expresiones CSV en el optimizador de SQL eliminando las columnas innecesarias de las operaciones from_csv.	TRUE
spark.sql.optimizer.enableJsonExpressionOptimización	Especifica si se deben optimizar las expresiones JSON en el optimizador de SQL de la siguiente manera: • Eliminar columnas innecesarias de las operaciones de from_json • Simplificar las combinaciones from_json y to_json • Optimización de las operaciones de named_struct	TRUE
Spark.sql.Optimizer.ExcludedRules	Define las reglas del optimizador que se van a deshabilitar, identificadas mediante nombres de reglas separados por comas. Algunas reglas no se pueden deshabilitar porque son necesarias para que sean correctas. El optimizador registra qué reglas se han desactivado correctamente.	(ninguno)
spark.sql.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Establece el tamaño mínimo de escaneo agregado en bytes necesario para inyectar un filtro Bloom en el lado de la aplicación.	10 GB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Filtro Spark.Sql.Optimizer.Runtime.BloomFilter.creationSideThreshold	Define el umbral de tamaño máximo para inyectar un filtro Bloom en el lado de la creación.	10 MB
spark.sql.Optimizer.Runtime.BloomFilter.Enabled	Especifica si se debe insertar un filtro Bloom para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	TRUE
spark.sql.optimizer.Runtime.BloomFilter.expectedNumItems	Define el número predeterminado de elementos esperados en el filtro Bloom en tiempo de ejecución.	1000000
spark.sql.Optimizer.Runtime.BloomFilter.maxNumBits	Establece el número máximo de bits permitido en el filtro Bloom en tiempo de ejecución.	67108864
Spark.Sql.Optimizer.Runtime.BloomFilter.maxNumItems	Establece el número máximo de elementos esperados permitidos en el filtro Bloom en tiempo de ejecución.	4000000
spark.sql.Optimizer.Runtime.BloomFilter.Number.Threshold	Limita el número máximo de filtros de tiempo de ejecución distintos de DPP permitidos por consulta para evitar errores en el controlador. out-of-memory	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Define el número predeterminado de bits que se utilizan en el filtro Bloom en tiempo de ejecución.	8388608

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimize r.runtime. rowlevelO perationGroupFiltro. Habilitado	Especifica si se debe habilitar el filtrado de grupos en tiempo de ejecución para las operaciones a nivel de fila. Permite que las fuentes de datos: • Elimine grupos completos de datos (como archivos o particiones) mediante filtros de fuentes de datos • Ejecute consultas en tiempo de ejecución para identificar los registros coincidentes • Deseche los grupos innecesarios para evitar costosas reescrituras Limitaciones: • No todas las expresiones se pueden convertir en filtros de fuentes de datos • Algunas expresiones requieren la evaluación de Spark (como las subconsultas)	TRUE
spark.sql.Optimize r.RuntimeFilter. semiJoinReduction. habilitado	Especifica si se debe insertar una semiunión para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.parquet.aggregatePushdown	Especifica si se deben enviar los agregados a Parquet para su optimización. Soportes: • MIN y MAX para los tipos booleano, entero, flotante y de fecha • COUNT para todos los tipos de datos Lanza una excepción si faltan estadísticas en el pie de página de un archivo de Parquet.	FALSO
spark.sql.parquet.columnarReaderBatchSize	Controla el número de filas de cada lote de lectores vectorizados Parquet. Elija un valor que equilibre la sobrecarga de rendimiento y el uso de memoria para evitar out-of-memory errores.	4096
Spark.sql.Session.TimeZone	Define la zona horaria de la sesión para gestionar las marcas de tiempo en los literales de cadena y la conversión de objetos Java. Acepta: • area/city Formato basado IDs en regiones (como America/Los_Angeles) • Desplazamientos de zona en formato (+/-) HH, (+/-) HH:mm o (+/-) HH:mm:ss (como -08 o + 01:00) • UTC o Z como alias para + 00:00	(valor de la zona horaria local)
spark.sql.shuffle.partition	Establece el número predeterminado de particiones para la mezcla de datos durante las uniones o agregaciones. No se puede modificar entre reinicios de consultas de streaming estructuradas desde la misma ubicación del punto de control.	200

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.shuffleHashJoinFactor	Define el factor de multiplicación utilizado para determinar la idoneidad de las combinaciones de hash aleatorio. Se selecciona una combinación de hash aleatorio cuando el tamaño de los datos del lado pequeño multiplicado por este factor es menor que el tamaño de los datos del lado grande.	3
spark.sql.sources.parallelPartitionDiscovery.umbra	Establece el número máximo de rutas para la lista de archivos del lado del conductor con fuentes basadas en archivos (Parquet, JSON y ORC). Si se supera durante la detección de particiones, los archivos se enumeran mediante un trabajo distribuido de Spark independiente.	32
spark.sql.statistics.histogram.enabled	Especifica si se deben generar histogramas con la misma altura durante el cálculo de las estadísticas de columnas para mejorar la precisión de la estimación. Requiere un escaneo de tabla adicional al necesario para las estadísticas de columnas básicas.	FALSO

9. Seleccione Ejecutar.

Note

No podrá ejecutar la consulta si el miembro que puede recibir los resultados no ha configurado los ajustes de resultados de consulta.

10. Siga ajustando los parámetros y vuelva a ejecutar la consulta, o pulse el botón + para iniciar una nueva consulta en una pestaña nueva.

Consulta con el creador de análisis

Puede usar el creador de análisis para crear consultas sin tener que escribir código SQL. Con el creador de análisis, puede crear una consulta para una colaboración que tenga:

- Una sola tabla que utilice la [regla de análisis de agregación](#) sin necesidad de usar JOIN
- Dos tablas (una de cada miembro) que utilicen la [regla de análisis de agregación](#)
- Dos tablas (una de cada miembro) que utilicen la [regla de análisis de lista](#)
- Dos tablas (una de cada miembro) que utilicen la regla de análisis de agregación y dos tablas (una de cada miembro) que utilicen la regla de análisis de lista

Si desea escribir consultas SQL manualmente, consulte [Consulta de tablas configuradas con el editor de código SQL](#).

El generador de análisis aparece como la opción de interfaz de usuario del generador de análisis en la sección Análisis de la pestaña Análisis de la AWS Clean Rooms consola.

Important

Si activa Interfaz de usuario del creador de análisis, comienza a crear una consulta en el creador de análisis y, a continuación, desactiva la Interfaz de usuario del creador de análisis, la consulta no se guarda.

Tip

Si se lleva a cabo un mantenimiento programado mientras se está ejecutando una consulta, esta se termina y se revierte. Deberá reiniciar la consulta.

En los siguientes temas se explica cómo utilizar el creador de análisis.

Temas

- [Uso del creador de análisis para consultar una sola tabla \(agregación\)](#)
- [Uso del creador de análisis para consultar dos tablas \(agregación o lista\)](#)

Uso del creador de análisis para consultar una sola tabla (agregación)

Este procedimiento muestra cómo utilizar la interfaz de usuario del generador de análisis de la AWS Clean Rooms consola para crear una consulta. La consulta se refiere a una colaboración que tiene una sola tabla que utiliza la [regla de análisis de agregación](#) sin necesidad de JOIN.

Para usar el creador de análisis para consultar una sola tabla

Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.

1. En el panel de navegación izquierdo, elija Colaboraciones.
2. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Realizar consultas.
3. En la pestaña Análisis, en Tablas, consulte la tabla y su tipo de regla de análisis asociada. (el tipo de regla de análisis debe ser Regla de análisis de agregación).

Note

Si no ve la tabla que esperaba, puede deberse a los siguientes motivos:

- La tabla no se ha [asociado](#).
- La tabla no tiene una [regla de análisis configurada](#).

4. En la sección Análisis, active la Interfaz de usuario del creador de análisis.
5. Cree una consulta.

Si desea ver todas las métricas de agregación, vaya al paso 9.

- a. En Elegir métricas, revise las métricas agregadas que se han preseleccionado de forma predeterminada y elimine cualquier métrica si es necesario.
- b. (Opcional) En Añadir segmentos: opcional, elija uno o varios parámetros.

Note

Añadir segmentos: opcional solo se muestra si se especifican las dimensiones para la tabla.

- c. (Opcional) En Añadir filtros: opcional, elija Añadir filtro y, a continuación, elija un Parámetro, un operador y un Valor.

Para añadir más filtros, elija Añadir otro filtro.

Para eliminar un filtro, elija Eliminar.

 Note

ORDER BY no se admite en las consultas de agregación.
En los filtros, solo se admite el operador AND.

- d. (Opcional) En Añadir descripción: opcional, introduzca una descripción que ayude a identificar la consulta en la lista de consultas.
6. Amplíe Vista previa del código SQL.
 - a. Consulte el código SQL que se genera a partir del creador de análisis.
 - b. Para copiar el código SQL, elija Copiar.
 - c. Para editar el código SQL, elija Editar en el editor de código SQL.
 7. Elija Ejecutar.
-  Note
- No podrá ejecutar la consulta si el miembro que puede recibir los resultados no ha configurado los ajustes de resultados de consulta.
8. Siga ajustando los parámetros y vuelva a ejecutar la consulta, o pulse el botón + para iniciar una nueva consulta en una pestaña nueva.

 Note

AWS Clean Rooms tiene como objetivo proporcionar mensajes de error claros. Si un mensaje de error no contiene detalles suficientes para ayudarle a solucionar el problema, póngase en contacto con el equipo de cuentas. Proporcióneles una descripción de cómo se produjo el error y el mensaje de error (incluidos los identificadores). Para obtener más información, consulte [Solución de problemas AWS Clean Rooms](#).

Uso del creador de análisis para consultar dos tablas (agregación o lista)

Este procedimiento describe cómo utilizar el generador de análisis de la AWS Clean Rooms consola para crear una consulta para una colaboración que tenga:

- Dos tablas (una de cada miembro) que utilicen la [regla de análisis de agregación](#)
- Dos tablas (una de cada miembro) que utilicen la [regla de análisis de lista](#)
- Dos tablas (una de cada miembro) que utilicen la regla de análisis de agregación y dos tablas (una de cada miembro) que utilicen la regla de análisis de lista

Para usar el creador de análisis para consultar dos tablas

Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.

1. En el panel de navegación izquierdo, elija Colaboraciones.
2. Elija la colaboración que tenga el estado Sus habilidades de miembro de Ejecutar consultas..
3. En la pestaña Análisis, en Tablas, consulte las dos tablas y el tipo de regla de análisis asociado (regla de análisis de agregación o regla de análisis de listas).

Note

Si no ve las tablas que espera en la lista, puede deberse a los siguientes motivos:

- Las tablas no se han [asociado](#).
- Las tablas no tienen [una regla de análisis configurada](#).

4. En la sección Análisis, active la Interfaz de usuario del creador de análisis.
5. Cree una consulta.

Si la colaboración contiene dos tablas que utilizan la Regla de análisis de agregación y dos tablas que utilizan la Regla de análisis de lista, elija primero Agregación o Lista y, a continuación, siga las instrucciones según la regla de análisis seleccionada.

Si las dos tablas utilizan la regla de análisis de agregación

1. En Elegir métricas, revise las métricas agregadas que se han preseleccionado de forma predeterminada y elimine cualquier métrica si es necesario.
2. En Registros de coincidencia, elija uno o más registros.

Note

Al utilizar el creador de análisis, solo puede hacer coincidir un único par de columnas.

3. (Opcional) En Añadir segmentos: opcional, elija uno o varios parámetros.

Note

Añadir segmentos: opcional solo se muestra si se especifican las dimensiones para la tabla.

4. (Opcional) En Añadir filtros: opcional, elija Añadir filtro y, a continuación, elija un parámetro, un operador y un valor.

Para añadir más filtros, elija Añadir otro filtro.

Para eliminar un filtro, elija Eliminar.

Si las dos tablas utilizan la regla de análisis de lista

1. En Elegir atributos, revise la lista de atributos que se han preseleccionado de forma predeterminada y elimine cualquier métrica si es necesario.
2. En Registros de coincidencia, elija uno o más registros.

Note

Al utilizar el creador de análisis, solo puede hacer coincidir un único par de columnas.

3. (Opcional) En Añadir filtros: opcional, elija Añadir filtro y, a continuación, elija un parámetro, un operador y un valor.

Para añadir más filtros, elija Añadir otro filtro.

Para eliminar un filtro, elija Eliminar.

Note

Las consultas de lista no admiten LIMIT.
En los filtros, solo se admite el operador AND.

4. (Opcional) En Añadir descripción: opcional, introduzca una descripción que ayude a identificar la consulta en la lista de consultas recientes.

Si las dos tablas utilizan la regla de análisis de agregación

Si las dos tablas utilizan la regla de análisis de lista

 Note

ORDER BY no se admite en las consultas de agregación.
En los filtros, solo se admite el operador AND.

5. (Opcional) En Añadir descripción: opcional, introduzca una descripción que ayude a identificar la consulta en la lista de consultas recientes.

6. Amplíe Vista previa del código SQL.

- a. Consulte el código SQL que se genera a partir del creador de análisis.
- b. Para copiar el código SQL, elija Copiar.
- c. Para editar el código SQL, elija Editar en el editor de código SQL.

7. Elija Ejecutar.

 Note

No podrá ejecutar la consulta si el miembro que puede recibir los resultados no ha configurado los ajustes de resultados de consulta.

8. Siga ajustando los parámetros y vuelva a ejecutar la consulta, o pulse el botón + para iniciar una nueva consulta en una pestaña nueva.

 Note

AWS Clean Rooms tiene como objetivo proporcionar mensajes de error claros. Si un mensaje de error no contiene detalles suficientes para ayudarle a solucionar el problema, póngase en contacto con el equipo de cuentas. Proporcióneles una descripción de cómo

se produjo el error y el mensaje de error (incluidos los identificadores). Para obtener más información, consulte [Solución de problemas AWS Clean Rooms](#).

Visualización del impacto de la privacidad diferencial

En general, la escritura y la ejecución de consultas no cambian cuando se activa la privacidad diferencial. Sin embargo, no puede ejecutar una consulta si no queda suficiente presupuesto de privacidad. A medida que ejecuta consultas y consume el presupuesto de privacidad, puede ver aproximadamente cuántas agregaciones puede ejecutar y cómo podrían afectar a las consultas futuras.

Para ver el impacto de la privacidad diferencial en una colaboración

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus detalles de miembro sea Ejecutar consultas.
4. En la pestaña Análisis, en Tablas, consulta el presupuesto de privacidad restante. Se muestra como el número estimado de funciones de agregación restantes y la utilidad utilizada (representada como porcentaje).

Note

El número estimado de funciones de agregación restantes y el porcentaje de la utilidad utilizada solo muestran para el miembro que puede realizar la consulta.

5. Seleccione Ver impacto para ver cuánto ruido se inyecta en los resultados y aproximadamente cuántas funciones de agregación puede ejecutar.

Vista de consultas recientes

Puede ver las consultas que se han realizado en los últimos 90 días en la pestaña Análisis.

 Note

Si su única capacidad de miembro es Contribuir datos y no es el [miembro que paga los costes de cálculo de las consultas](#), la pestaña Análisis no aparecerá en la consola.

Para ver consultas recientes

Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.

1. En el panel de navegación izquierdo, elija Colaboraciones.
2. Elija una colaboración.
3. En la pestaña Análisis, en Análisis, seleccione Todas las consultas en el menú desplegable y consulte las consultas que se han realizado en los últimos 90 días.
4. Para ordenar las consultas recientes por Estado, seleccione un estado en la lista desplegable Todos los estados.

Los estados son: Enviada, Iniciada, Cancelada, Éxito, Error y Agotado tiempo de espera.

Visualización de detalles de consultas

Puede ver los detalles de la consulta como miembro que puede realizar consultas o como miembro que puede recibir los resultados.

Para ver los detalles de la consulta

1. Inicie sesión en Consola de administración de AWS <https://console.aws.amazon.com/cleanrooms> y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija una colaboración.
4. En la pestaña Análisis, realice una de las siguientes acciones:
 - Elija el botón de opción correspondiente a la consulta específica que desea ver y, a continuación, elija Ver detalles.
 - Elija el ID de consulta protegida.
5. En la página de Detalles de la consulta,

- Si es el miembro que puede realizar consultas, vea los Detalles de la consulta, el Texto SQL y los Resultados.
- Ve un mensaje que confirma que los resultados de la consulta se entregaron al miembro que puede recibir los resultados.
- Si usted es el miembro que puede recibir los resultados, vea los Detalles de la consulta y los Resultados.

Ejecutando PySpark trabajos

Como [miembro que puede realizar consultas](#), puede ejecutar un PySpark trabajo en una tabla configurada mediante una [plantilla de PySpark análisis](#) aprobada.

Requisitos previos

Antes de ejecutar un PySpark trabajo, debe tener:

- Ser miembro activo de AWS Clean Rooms Collaboration
- Acceso a al menos una plantilla de análisis de la colaboración
- Acceso a al menos una tabla configurada de la colaboración
- Permisos para escribir los resultados de un PySpark trabajo en un bucket de S3 específico

Para obtener información sobre cómo crear el rol de servicio requerido, consulte [Cree un rol de servicio para escribir los resultados de un PySpark trabajo](#).

- El miembro responsable de pagar los costes informáticos se ha unido a la colaboración como miembro activo

Para obtener información sobre cómo consultar datos o ver las consultas llamando directamente a la operación de la AWS Clean Rooms StartProtectedJob API o utilizando la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

Para obtener información sobre el registro de trabajos, consulte [Inicio de sesión de Analysis AWS Clean Rooms](#).

Para obtener información sobre cómo recibir los resultados de los trabajos, consulte [Recepción y uso de los resultados del análisis](#).

En los temas siguientes se explica cómo ejecutar un PySpark trabajo en una tabla configurada en una colaboración mediante la AWS Clean Rooms consola.

Temas

- [Ejecutar un PySpark trabajo en una tabla configurada mediante una plantilla PySpark de análisis](#)
- [Ver trabajos recientes](#)
- [Consultar detalles del trabajo](#)

Ejecutar un PySpark trabajo en una tabla configurada mediante una plantilla PySpark de análisis

Este procedimiento muestra cómo utilizar una plantilla de PySpark análisis en la AWS Clean Rooms consola para analizar las tablas configuradas con la regla de análisis personalizada.

Para ejecutar un PySpark trabajo en una tabla configurada mediante una plantilla PySpark de análisis

Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.

1. En el panel de navegación izquierdo, elija Colaboraciones.
2. Elija la colaboración que tenga el estado Sus habilidades de miembro de Ejecutar trabajos.
3. En la pestaña Análisis, en la sección Tablas, consulte las tablas y el tipo de regla de análisis asociado (regla de análisis personalizada).

Note

Si no ve las tablas que esperaba en la lista, puede deberse a los siguientes motivos:

- Las tablas no se han [asociado](#).
- Las tablas no tienen [una regla de análisis configurada](#).

4. En la sección Análisis, en el modo Análisis, seleccione Ejecutar plantillas de análisis.
5. Elija la plantilla de PySpark análisis en la lista desplegable de plantillas de análisis.

Los parámetros de la plantilla de PySpark análisis se rellenarán automáticamente en la definición.

6. Si la plantilla de análisis tiene parámetros definidos, en **Parámetros**, proporcione valores para los parámetros:
 - a. Para cada parámetro, consulte el nombre del parámetro y el valor por defecto (si está configurado).
 - b. Introduzca un valor para cada parámetro que desee anular.

 Note

Si no proporciona un valor pero existe un valor predeterminado, se utilizará el valor predeterminado.

 Important

Los valores de los parámetros pueden tener hasta 1000 caracteres y admiten la codificación UTF-8. Todos los valores de los parámetros se tratan como cadenas y se pasan al script de usuario a través del objeto de contexto. Asegúrese de que el script de usuario valide y gestione los valores de los parámetros de forma segura. Para obtener más información sobre la gestión segura de parámetros, consulte [Trabajar con parámetros en plantillas PySpark de análisis](#).

7. Especifique el tipo de trabajador admitido y el número de trabajadores.

Utilice la siguiente tabla para determinar el tipo y la cantidad de trabajadores que necesita para su caso de uso.

Tipo de empleado	vCPU	Memoria (GB)	Almacenamiento (GB)	Número de procesos de trabajo	Unidades totales de procesamiento de salas limpias (CRPU)
CR.1X (predeterminado)	4	30	100	4	8
				128	256

Tipo de empleado	vCPU	Memoria (GB)	Almacenamiento (GB)	Número de procesos de trabajo	Unidades totales de procesamiento de salas limpias (CRPU)
CR.4X	16	120	400	4	32
				32	256

Note

Los diferentes tipos y número de trabajadores tienen costos asociados. Para obtener más información sobre los precios, consulta [AWS Clean Rooms los precios](#).

8. Seleccione Ejecutar.

Note

No puedes ejecutar el trabajo si el miembro que puede recibir los resultados no ha configurado los ajustes de los resultados del trabajo.

9. Siga ajustando los parámetros y vuelva a ejecutar el trabajo, o pulse el botón + para iniciar un nuevo trabajo en una pestaña nueva.

Ver trabajos recientes

Puede ver los trabajos ejecutados en los últimos 90 días en la pestaña Análisis.

Note

Si su única capacidad de miembro es Contribuir datos y no es el [miembro que paga los costes de cálculo del trabajo](#), la pestaña Análisis no aparecerá en la consola.

Para ver los trabajos recientes

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija una colaboración.
4. En la pestaña Análisis, en Análisis, seleccione Todos los trabajos en el menú desplegable y vea los trabajos que se han ejecutado en los últimos 90 días.
5. Para ordenar los trabajos recientes por estado, seleccione un estado de la lista desplegable Todos los estados.

Los estados son: Enviada, Iniciada, Cancelada, Éxito, Error y Agotado tiempo de espera.

Consultar detalles del trabajo

Puede ver los detalles del trabajo como el miembro que puede ejecutar los trabajos o como el miembro que puede recibir los resultados.

Para ver los detalles del trabajo

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija una colaboración.
4. En la pestaña Análisis, en Análisis, seleccione Todos los trabajos en el menú desplegable y, a continuación, realice una de las siguientes acciones:
 - Elija el botón de opción para el trabajo específico que desee ver y, a continuación, elija Ver detalles.
 - Elige el ID de trabajo protegido.
5. En la página de detalles del trabajo,
 - Si usted es el miembro que puede ejecutar trabajos, consulte los detalles del trabajo, el trabajo y los resultados.

Aparece un mensaje que confirma que los resultados del trabajo se entregaron al miembro que puede recibirlos.

- Si usted es el miembro que puede recibir los resultados, consulte los detalles del trabajo y los resultados.

Recepción y uso de los resultados del análisis

[Los miembros que puedan recibir los resultados](#) pueden revisar los resultados de la consulta en la AWS Clean Rooms consola o en los buckets de Amazon S3 que especificaron al unirse a la colaboración.

Note

Solo en el caso de las tablas de datos cifrados, el miembro que puede recibir los resultados descifra los resultados de la consulta ejecutando el cliente de cifrado de C3R en modo de [descifrado](#).

El destino de los resultados en Amazon S3 no puede estar dentro del mismo bucket de S3 que cualquier fuente de datos.

En los temas siguientes se explica cómo recibir los resultados de los análisis mediante la AWS Clean Rooms consola.

Temas

- [Recibir resultados de consultas](#)
- [Recibir los resultados del trabajo](#)
- [Edición de los ajustes predeterminados de los resultados de las consultas](#)
- [Edición de los valores predeterminados de la configuración de los resultados del trabajo](#)
- [Uso del resultado de la consulta en otros Servicios de AWS](#)

Para obtener información sobre cómo consultar datos o ver las consultas llamando a la AWS Clean Rooms API directamente o utilizando la AWS SDKs, consulte la [referencia de la AWS Clean Rooms API](#).

Para obtener más información sobre el registro de consultas, consulte [Inicio de sesión de Analysis AWS Clean Rooms](#).

Note

Si ejecuta una consulta en tablas de datos cifrados, los resultados de las columnas cifradas se cifran.

Recibir resultados de consultas

 Note

El destino de los resultados en Amazon S3 no puede estar dentro del mismo bucket de S3 que cualquier fuente de datos.

Los resultados de la consulta se encuentran en la sección de valores predeterminados de la configuración de resultados de la pestaña Análisis de la AWS Clean Rooms consola.

Para recibir resultados de consultas

1. Inicie sesión en la [AWS Clean Rooms consola Consola de administración de AWS y ábrala](#) con su Cuenta de AWS (si aún no lo ha hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Recibir resultados.
4. Para recibir los resultados de la consulta directamente AWS Clean Rooms, en la pestaña Análisis, en Análisis, selecciona Todas las consultas en el menú desplegable y, a continuación, en la columna ID de consulta protegida, selecciona la consulta.
5. En la página Detalles de la consulta, en Resultados, realice una de las siguientes acciones:

Si desea...	Entonces elija...
Copiar los resultados.	Copiar
Descargar los resultados.	Download  Note De forma predeterminada, el nombre del archivo descargado es el Query id correspondiente que se mostró cuando al ejecutar la consulta en AWS Clean Rooms.
Consulte los resultados en Amazon S3.	Ver en Amazon S3

Si desea...	Entonces elija...
	Se abre la consola de Amazon S3 en una pestaña aparte.

6. Si utiliza datos cifrados, ahora puede [descifrar](#) las tablas de datos.

Para obtener más información, consulte [Descifrar tablas de datos con el cliente de cifrado de C3R](#).

Recibir los resultados del trabajo

Note

El destino de los resultados en Amazon S3 no puede estar dentro del mismo bucket de S3 que cualquier fuente de datos.

Los resultados del trabajo se encuentran en la sección de valores predeterminados de la configuración de resultados de la pestaña Análisis de la AWS Clean Rooms consola.

Para recibir los resultados del trabajo

1. Inicia sesión Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Recibir resultados.
4. Para recibir los resultados del trabajo directamente AWS Clean Rooms, en la pestaña Análisis, en Análisis, selecciona Todos los trabajos en el menú desplegable y, a continuación, en la columna de ID de trabajo protegido, selecciona el trabajo.
5. En la página de detalles del trabajo, en Resultados, copia el ID del trabajo.

Vuelva a la pestaña Análisis y amplíe los valores predeterminados de la configuración de resultados.

En Destino de los resultados, seleccione el enlace para ver los resultados en Amazon S3.

Se abre la consola de Amazon S3 en una pestaña aparte.

En Amazon S3, pegue el ID del trabajo en la barra de búsqueda y pulse Entrar.

Aparece la carpeta que contiene los resultados. Seleccione la carpeta para ver los resultados del trabajo.

Edición de los ajustes predeterminados de los resultados de las consultas

Note

El destino de los resultados en Amazon S3 no puede estar dentro del mismo bucket de S3 que cualquier fuente de datos.

Como miembro que puede recibir resultados, puede editar los valores predeterminados de la configuración de los resultados de las consultas en la AWS Clean Rooms consola.

Para editar los ajustes predeterminados de los resultados de las consultas

1. Inicia sesión en la [AWS Clean Rooms consola Consola de administración de AWS](#) y ábrela con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Recibir resultados.
4. En la pestaña Análisis, en los valores predeterminados de la configuración de resultados, selecciona Editar.
5. En la página Editar los valores predeterminados de la configuración de los resultados, modifique cualquiera de las siguientes opciones según sea necesario:
 - a. En Resultados de la consulta, modifique el destino de los resultados en Amazon S3, el formato de los resultados o los archivos de resultados.
 - b. (Opcional) Para acceder al servicio, si desea enviar consultas que tardan hasta 24 horas en llegar a su destino de S3, seleccione la casilla Añadir una función de servicio para atender las consultas que tarden hasta 24 horas en completarse.

Las consultas grandes que tarden hasta 24 horas en completarse se enviarán a su destino de S3.

Si no selecciona la casilla de verificación, solo las consultas que se completen en un plazo de 12 horas se enviarán a su ubicación de S3.

- Especifique los permisos de Acceso a servicios seleccionando Crear y usar un nuevo rol de servicio o Usar un rol de servicio existente.

Create and use a new service role

- AWS Clean Rooms crea un rol de servicio con la política requerida para esta tabla.
- El Nombre del rol de servicio predeterminado es `cleanrooms-query-receiver-<timestamp>`
- Debe tener permisos para crear roles y adjuntar políticas.

Use an existing service role

1. Seleccione un Nombre de rol de servicio existente en la lista desplegable.

Si tiene permisos de listas de roles, se mostrará la lista de roles.

Si no tiene permisos de listas de roles, puede ingresar el nombre de recurso de Amazon (ARN) del rol que desea usar.

2. Consulte el rol de servicio mediante la elección del enlace externo Ver en IAM.

Si no hay ningún rol de servicio existente, la opción Usar un rol de servicio existente no estará disponible.

De forma predeterminada, AWS Clean Rooms no intenta actualizar la política de roles existente para añadir los permisos necesarios.

Note

- AWS Clean Rooms requiere permisos para realizar consultas de acuerdo con las reglas de análisis. Para obtener más información sobre los permisos para AWS Clean Rooms, consulte [AWS políticas gestionadas para AWS Clean Rooms](#).

- Si el rol no tiene permisos suficientes AWS Clean Rooms, recibirá un mensaje de error que indica que el rol no tiene permisos suficientes AWS Clean Rooms. Debe agregar la política de rol antes de continuar.
- Si no puede modificar la política de roles, recibirá un mensaje de error que indica que no ha AWS Clean Rooms podido encontrar la política para el rol de servicio.

6. Seleccione Save changes (Guardar cambios).
7. Los Ajustes de los resultados de las consultas aparecen en la página de detalles de la colaboración.

Edición de los valores predeterminados de la configuración de los resultados del trabajo

Note

El destino de los resultados en Amazon S3 no puede estar dentro del mismo bucket de S3 que cualquier fuente de datos.

Como miembro que puede recibir resultados, puede editar los valores predeterminados de la configuración de los resultados de los trabajos en la AWS Clean Rooms consola.

Para editar los valores predeterminados de la configuración de los resultados del trabajo

1. Inicie sesión en Consola de administración de AWS y abra la [AWS Clean Rooms consola](#) con su Cuenta de AWS (si aún no lo ha hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. Elija la colaboración cuyo estado de Sus capacidades como miembro sea Recibir resultados.
4. En la pestaña Análisis, en los valores predeterminados de la configuración de resultados, selecciona Editar.
5. En la página Editar los valores predeterminados de la configuración de los resultados, modifique cualquiera de las siguientes opciones según sea necesario:
 - a. En Job results, modifique el destino de los resultados en Amazon S3.

- b. En Acceso al servicio, modifique el nombre del rol de servicio existente.
6. Seleccione Save changes (Guardar cambios).
7. La configuración actualizada de los resultados del trabajo aparece en la página de detalles de la colaboración.

Uso del resultado de la consulta en otros Servicios de AWS

El resultado de una consulta SQL se puede utilizar para los datos iniciales para un modelo de Clean Rooms ML. Para obtener más información, consulte [AWS Clean Rooms ML](#).

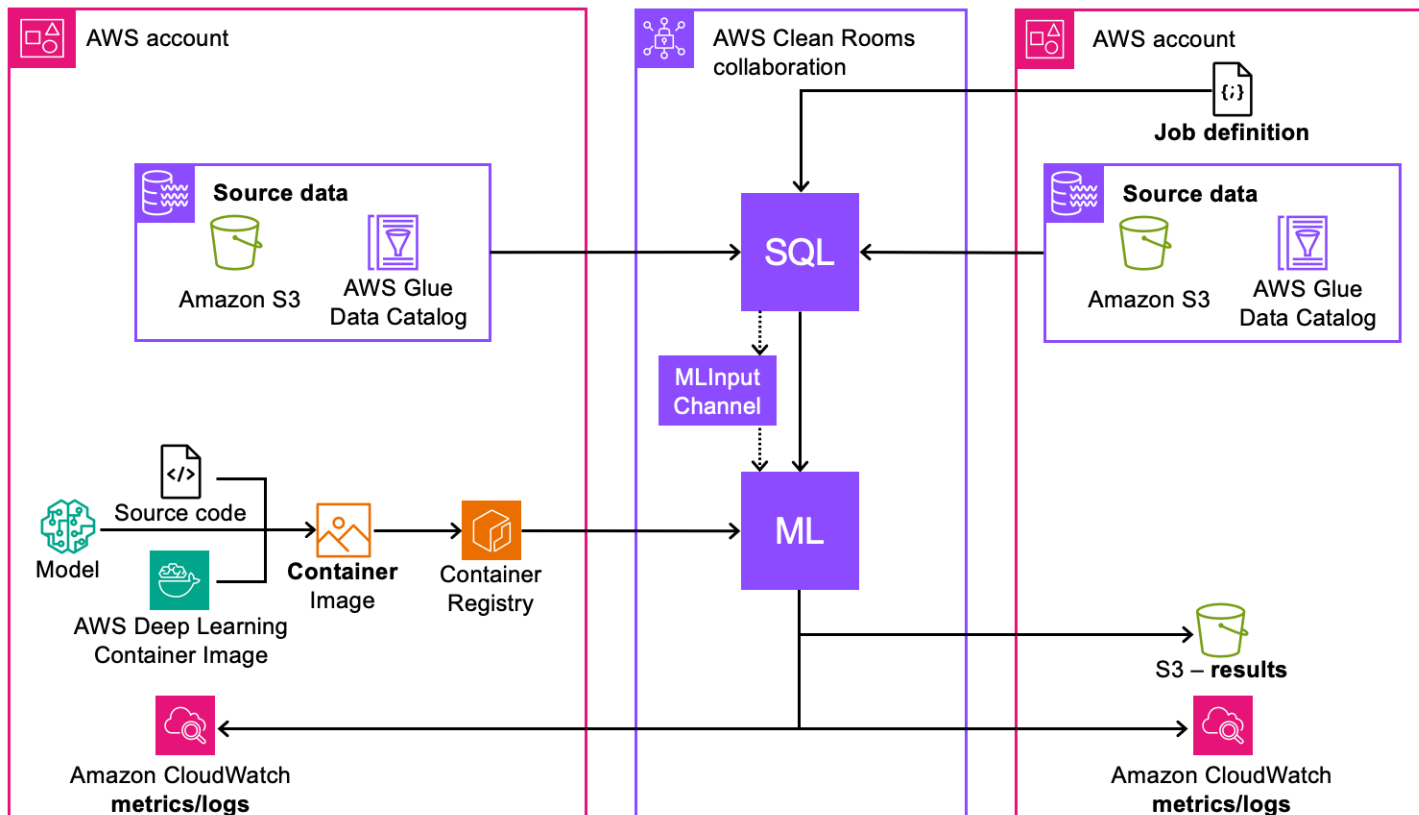
El resultado de la consulta AWS Clean Rooms está disponible en la consola (si la consola se utiliza para ejecutar consultas) y se descarga en un bucket de Amazon S3 específico. A partir de ahí, puede utilizar el resultado de la consulta en otros Servicios de AWS, como Amazon Quick y Amazon SageMaker AI, en función del uso que esos servicios hagan de los datos de Amazon S3.

Para obtener más información sobre Amazon Quick, consulta la [documentación de Amazon Quick](#).

Para obtener más información sobre Amazon SageMaker AI, consulte la [documentación de Amazon SageMaker AI](#).

Modelado personalizado en AWS Clean Rooms ML

Desde un punto de vista técnico, en el siguiente diagrama se describe cómo funciona el modelado de aprendizaje automático personalizado en AWS Clean Rooms ML.



Así es como funciona el modelado de aprendizaje automático personalizado en Clean Rooms ML:

1. Configuración de la fuente de datos

- Los datos de origen se pueden almacenar en el catálogo de Amazon S3, en o en Snowflake AWS Glue Data Catalog
- AWS Glue Data Catalog se utiliza para organizar y catalogar
- Los datos de varios se Cuentas de AWS pueden utilizar en la misma colaboración

2. Consultas SQL y procesamiento de datos

- Las consultas SQL se utilizan para acceder a los datos de origen y procesarlos
- Las consultas se ejecutan dentro de los límites de AWS Clean Rooms colaboración
- Los datos procesados se introducen en los canales de entrada de ML para el entrenamiento de modelos

3. Desarrollo de modelos ML

- El código fuente del modelo se puede desarrollar utilizando AWS Deep Learning Container Images
- Las imágenes de contenedores personalizadas deben crearse y almacenarse en Amazon Elastic Container Registry

4. Componentes de infraestructura

- Amazon Elastic Container Registry almacena y administra los contenedores del modelo ML
- El procesamiento del aprendizaje automático se realiza dentro de un entorno de AWS Clean Rooms colaboración seguro

5. Monitoreo y registro

- Amazon CloudWatch proporciona métricas y registros para ambas partes colaboradoras
- El monitoreo está disponible para todas las Cuentas de AWS partes involucradas en la colaboración
- Las partes pertinentes pueden acceder a las métricas de rendimiento y los registros operativos

6. Gestión de resultados

- El acceso a los resultados se controla de acuerdo con los permisos de colaboración

Antes de empezar, consulte la [Requisitos previos del modelado ML personalizado](#) y [Pautas de creación de modelos para el contenedor de formación](#) para obtener más información.

Generación de conjuntos de datos sintéticos con mayor privacidad

Un conjunto de datos sintético tiene propiedades estadísticas similares al conjunto de datos original en el que se basa, pero no contiene las observaciones del mundo real presentes en el conjunto de datos original. Al utilizar conjuntos de datos sintéticos que mejoran la privacidad, puedes descubrir nuevos casos de uso del entrenamiento con modelos de aprendizaje automático (ML) que antes se impedían por cuestiones de privacidad de datos. Al crear un canal de entrada de aprendizaje automático, puede generar datos sintéticos para proteger la información confidencial y, al mismo tiempo, entrenar modelos de aprendizaje automático.

Al crear una plantilla con datos sintéticos, debe:

- Exija que el resultado de la plantilla sea sintético
- Clasifique las columnas del esquema de salida como numéricas o categóricas
- Personalice los datos sintéticos en función de las necesidades de la organización

- Ajusta la configuración de privacidad:
 - Establece el nivel de privacidad (épsilon)
 - Configure el umbral de privacidad

Warning

La generación de datos sintéticos evita la inferencia de atributos individuales, ya sea que haya personas específicas presentes en el conjunto de datos original o que estén presentes los atributos de aprendizaje de esas personas. Sin embargo, no impide que los valores literales del conjunto de datos original, incluida la información de identificación personal (PII), aparezcan en el conjunto de datos sintético.

Recomendamos evitar los valores en el conjunto de datos de entrada que estén asociados a un solo sujeto de datos, ya que pueden volver a identificar a un sujeto de datos. Por ejemplo, si solo un usuario vive en un código postal, la presencia de ese código postal en el conjunto de datos sintético confirmaría que el usuario estaba en el conjunto de datos original. Para mitigar este riesgo, se pueden utilizar técnicas como truncar valores de alta precisión o reemplazar catálogos poco comunes por otros. Estas transformaciones pueden formar parte de la consulta utilizada para crear el canal de entrada de ML.

Para obtener más información sobre cómo generar datos sintéticos para el entrenamiento de modelos personalizados, consulte [Creación de una plantilla de análisis SQL](#).

Las plantillas de análisis con salidas sintéticas solo se pueden utilizar para crear canales de entrada de aprendizaje automático. Para obtener más información, consulte [Creación de un canal de entrada de ML en AWS Clean Rooms ML](#).

Consideraciones para la generación de datos sintéticos

Con AWS Clean Rooms ML, los miembros de la colaboración pueden crear un conjunto de datos sintético que desidentifique de forma irreversible los sujetos del conjunto de datos original de sus conjuntos de datos colectivos para entrenar un modelo de aprendizaje automático personalizado. Al crear la colaboración, debe configurar la información de pago para especificar quién paga por la generación de datos sintéticos. Estos son los pasos básicos para generar un conjunto de datos sintético y entrenar un modelo de aprendizaje automático personalizado:

1. Un miembro de la colaboración crea una plantilla de análisis que incluye:

- El SQL necesario para definir el conjunto de datos que se va a sintetizar.
 - Las configuraciones relacionadas con la privacidad se utilizan para garantizar que los datos sintéticos cumplan con los requisitos de conformidad de los proveedores de datos.
2. Una vez que todos los proveedores de datos aprueban la plantilla de análisis, el ejecutor de consultas de colaboración crea un canal de entrada de aprendizaje automático (ML) utilizando la plantilla.
 3. Clean Rooms ML genera el conjunto de datos sintético y verifica que cumpla con los umbrales de privacidad especificados en la plantilla de análisis.
 4. Si se cumplen todos los umbrales, el canal de entrada de ML se rellena con el conjunto de datos sintéticos.
 5. Luego, los clientes pueden usar este canal de entrada de aprendizaje automático para entrenar el modelo de aprendizaje automático personalizado asociado a la colaboración.

Consideraciones importantes:

- Los datos sintéticos generados en Clean Rooms ML no eliminan, censuran, ocultan ni sanear ningún valor individual, incluida la información de identificación personal (PII) que se encuentra en el conjunto de datos original. El conjunto de datos sintético se genera mediante el muestreo de valores, pero no de registros completos, del conjunto de datos original.
- Si el conjunto de datos original contiene filas similares, es posible que los datos sintéticos contengan filas que parezcan idénticas a las filas del conjunto de datos original.

Preparación del conjunto de datos:

- Evite las columnas con una distribución de clases significativamente desequilibrada. Esto es especialmente importante para el valor previsto o la columna «Y». Los desequilibrios extremos reducen la privacidad general del conjunto de datos sintéticos.
- Clean Rooms ML no admite la generación de datos sintéticos a partir de datos de series temporales, por lo que es importante mantener las correlaciones entre los registros secuenciales.
- Clean Rooms ML no admite la generación de datos sintéticos a partir de texto o datos no estructurados.
- Los siguientes tipos de datos son compatibles:

Nombre
del
tipo
de
datos

BIGINT

BOOLEANO

CHAR

DATE

DECIMAL

FLOAT

INTEGER

LONG

REAL

SHORT

SMALLINT

TIME

TIMESTAMP
_LTZ

TIMESTAMP
_NTZ

TINYINT

VARCHAR

Limitaciones:

- Para la generación de datos sintéticos, el número máximo de columnas predictivas es una.
- Si la columna de destino es categórica, el número máximo de categorías en el conjunto de datos original es 100.
- En el conjunto de datos original, el número de filas debe estar entre 1500 y 2,5 millones y el número máximo de columnas es 1000. Para valores no nulos en la columna de destino, el número mínimo de filas es 1000.

Métricas de privacidad:

- Clean Rooms ML proporciona una puntuación de privacidad que mide el grado de protección de los datos sintéticos generados contra los ataques de inferencia de miembros (MIAs). El servicio retiene el 5% de los datos originales del proceso de síntesis para calcular esta puntuación.
- Los puntajes cercanos al 50% se consideran buenos; los puntajes más altos indican una menor protección en contra. MIAs Los puntajes significativamente inferiores al 50% son poco frecuentes y pueden deberse a que los datos sintetizados no están representados en los patrones de los datos originales.

Modelo personalizado descendente:

- Los datos sintéticos generados en Clean Rooms ML son los más adecuados para entrenar modelos de clasificación binaria y modelos de clasificación multiclase con hasta cinco clases.
- El entrenamiento de modelos de regresión con datos sintéticos generados en el aprendizaje automático de salas limpias puede provocar una baja precisión del modelo, medida mediante el error cuadrático medio (RMSE).

Crear y unirse a la colaboración en AWS Clean Rooms ML

El creador de la colaboración es responsable de crear la colaboración, invitar a los miembros y asignar sus funciones. Los miembros invitados se unen a la colaboración y especifican la configuración de los resultados, modelan los artefactos entrenados, la configuración del destino y aceptan las responsabilidades de pago, en función de cómo esté configurada la colaboración.

Crear una colaboración para el aprendizaje automático

El siguiente procedimiento muestra cómo crear una colaboración para el aprendizaje automático, invitar a uno o más miembros y asignar miembros que puedan iniciar el entrenamiento con modelos, recibir resultados, recibir resultados de modelos entrenados, incluidos los artefactos y métricas del modelo, y recibir resultados de inferencias de modelos. El creador de la colaboración también asigna a un miembro que pagará los costes de cálculo de consultas, formación e inferencia de modelos.

Console

Crear una colaboración para el aprendizaje automático (consola)

1. [Cree una colaboración e invite a uno o más miembros a unirse a ella](#)
2. Asigne las siguientes habilidades a los miembros para el análisis mediante consultas y trabajos:
 - Asigne las consultas de ejecución al miembro que comenzará el entrenamiento con el modelo.
 - Asigne la opción Recibir los resultados del análisis a los miembros que recibirán los resultados de la consulta.
3. Asigne las siguientes capacidades a los miembros para el modelado de aprendizaje automático mediante flujos de trabajo diseñados específicamente para tal fin:
 - Asigne los resultados de los modelos entrenados al miembro que recibirá los resultados del modelo entrenado, incluidos los artefactos y las métricas del modelo.
 - Asigne la salida de Receive de la inferencia del modelo al miembro que recibirá los resultados de la inferencia del modelo.
4. En Configure el pago, especifique los miembros que pagarán los costos de procesamiento de consultas, entrenamiento de modelos, inferencia de modelos y generación de conjuntos de datos sintéticos.

Un conjunto de datos sintético tiene propiedades estadísticas similares al conjunto de datos original en el que se basa, pero no contiene las observaciones del mundo real presentes en el conjunto de datos original. Al utilizar conjuntos de datos sintéticos que mejoran la privacidad, puede descubrir nuevos casos de uso de la formación en modelos de aprendizaje automático que anteriormente se impedían por cuestiones de privacidad de datos. Para obtener más información, consulte Generación de datos sintéticos con [privacidad mejorada](#).

Cada uno de estos costos se puede asignar a los mismos miembros o a diferentes miembros. Si un miembro invitado es el miembro responsable de pagar los costes de pago, debe aceptar sus responsabilidades de pago antes de unirse a la colaboración.

5. En el caso de la membresía de Configure, el creador de la colaboración puede decidir unirse a la membresía ahora o crear una membresía más adelante. A continuación, el creador de la colaboración debe configurar la configuración de ML.
 - a. Si el creador de la colaboración también es el receptor de los resultados, también debe especificar el destino y el formato de los resultados de la consulta en los valores predeterminados de la configuración de resultados.
 - b. La configuración de ML proporciona una función para que Clean Rooms ML publique métricas en un Cuenta de AWS. Si el creador de la colaboración también recibe artefactos de modelos entrenados, puede especificar el bucket de Amazon S3 utilizado para recibir los resultados.
 - c. En la sección Configuraciones de ML, seleccione Crear configuración de ML y, a continuación, especifique el destino de salida del modelo en Amazon S3 y la función de acceso al servicio necesaria para acceder a esta ubicación.
 - d. Si el creador de la colaboración es el miembro responsable de pagar los costes de pago, debe aceptar sus responsabilidades de pago antes de crear la colaboración.

API

Para crear una colaboración para el aprendizaje automático (API)

1. [Cree una colaboración e invite a uno o más miembros a unirse a la colaboración](#)
2. Asigne las siguientes funciones a los miembros de la colaboración:
 - CAN_QUERY- asignado al miembro que iniciará el entrenamiento y la inferencia del modelo.
 - CAN_RECEIVE_MODEL_OUTPUT- se asigna a los miembros que recibirán los resultados de los modelos entrenados.
 - CAN_RECEIVE_INFERENCE_OUTPUT- se asigna a los miembros que recibirán los resultados de la inferencia del modelo.

Si el creador de la colaboración es también el receptor de los resultados, también debe especificar el destino y el formato de los resultados de la consulta durante la creación de la

colaboración. También asignan una función de servicio Amazon Resource Name (ARN) para escribir los resultados en el destino de los resultados de la consulta.

3. Especifique los miembros que pagarán los costes de cálculo de consultas, formación de modelos e inferencia de modelos. Cada uno de estos costos se puede asignar a los mismos miembros o a miembros diferentes. Si un miembro invitado es el miembro responsable de pagar los costes de pago, debe aceptar sus responsabilidades de pago antes de unirse a la colaboración.
4. El siguiente código crea una colaboración, invita a un miembro que puede realizar consultas y recibir resultados, y especifica al creador de la colaboración como receptor de los artefactos modelo.

```
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. A continuación, el creador de la colaboración debe configurar la configuración del aprendizaje automático. La configuración de ML proporciona una función para que Clean Rooms ML publique métricas y registros en un Cuenta de AWS. Si el creador de la colaboración también recibe resultados (artefactos del modelo o resultados de inferencias), puede especificar el bucket de Amazon S3 utilizado para recibir los resultados.

```

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"
            }
        }
    }
)

```

```
}  
}  
)
```

Unirse a una colaboración

Una vez que el creador de la colaboración haya terminado sus tareas, los miembros invitados deberán completar las suyas.

Console

Para crear una membresía y unirse a una colaboración (consola)

1. El miembro invitado [crea una membresía y se une a la colaboración](#).
2. Si el miembro invitado es el miembro responsable de pagar, incluidos los costos de consulta, capacitación e inferencia de modelos, debe aceptar sus responsabilidades de pago antes de unirse a la colaboración.
3. El miembro invitado configura la configuración de ML, que proporciona a Clean Rooms ML la función de publicar las métricas del modelo en un Cuenta de AWS. Si también es el miembro que recibe artefactos de modelos entrenados, deben entregar un bucket de Amazon S3 donde se almacenan los artefactos de modelos entrenados.

API

Para crear una membresía y unirse a una colaboración (API)

1. Si el miembro invitado es el miembro que puede recibir los resultados, especificará el destino y el formato de los resultados de las consultas. También proporcionan un rol de servicio (ARN) que permite al servicio escribir en el destino de los resultados de la consulta.

Si el miembro invitado es el miembro responsable de pagar, incluidos los costos de procesamiento de consultas, capacitación e inferencia de modelos, debe aceptar sus responsabilidades de pago antes de unirse a la colaboración.

Si el miembro invitado es el responsable de pagar la formación de modelos y la inferencia de modelos para el modelado personalizado, debe aceptar sus responsabilidades de pago antes de unirse a la colaboración.

El siguiente código crea una membresía con el registro de consultas activado.

```
import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_membership(
    membershipIdentifier='membership_id',
    queryLogStatus='ENABLED'
)
```

2. El miembro invitado configura la configuración de ML, que proporciona a Clean Rooms ML la función de publicar las métricas del modelo en un Cuenta de AWS. Si también es el miembro que recibe los artefactos de modelos entrenados, deben proporcionar un depósito de Amazon S3 donde se almacenan los artefactos de modelos entrenados.

```
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"
            }
        }
    }
)
```

Contribución de datos de formación en AWS Clean Rooms ML

Una vez que el creador de la colaboración haya creado la colaboración y los miembros invitados se hayan unido a ella, estará listo para aportar datos de formación a la colaboración. Cualquier miembro puede aportar datos de formación.

Console

Para aportar datos de entrenamiento (consola)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Tables (Tablas).
3. En la página Tablas, elija Configurar nueva tabla.
4. Para Configurar una nueva tabla, para Fuente de datos, elija Amazon S3, Amazon Athena o Snowflake y complete los siguientes pasos, en función de su fuente de datos:

Si está utilizando	Entonces
Amazon S3	1. Elija una base de datos de la lista desplegable y, a continuación, seleccione la tabla de la base de datos. 2. En Columnas permitidas en colaboraciones, seleccione Todas las columnas o Lista personalizada. 3. En los detalles de la tabla configurada, proporcione el nombre y una descripción opcional para esta tabla. 4. Si desea informar sobre las métricas del modelo, introduzca el nombre de las métricas y la instrucción Regexp que buscará la métrica en los registros de salida. 5. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
Amazon Athena	1. Elija una base de datos de la lista desplegable y, a continuación, seleccione la tabla de la base de datos.

Si está utilizando	Entonces
	2. En Columnas permitidas en colaboraciones, seleccione Todas las columnas o Lista personalizada. 3. En los detalles de la tabla configurada, proporcione el nombre y una descripción opcional para esta tabla. 4. Si desea informar sobre las métricas del modelo, introduzca el nombre de las métricas y la instrucción Regex que buscará la métrica en los registros de salida. 5. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
Snowflake	1. Especifique las credenciales de Snowflake utilizando un ARN secreto existente o almacenando un secreto nuevo para esta tabla. 2. Para los detalles de la tabla y el esquema de Snowflake, introduzca los detalles manualmente o impórtelos automáticamente. 3. Para el esquema, introduzca el nombre de la columna y elija el tipo de datos en la lista desplegable. 4. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.

5. Seleccione Configurar nuevo cliente.
6. En la página de detalles de la tabla, elija Configurar regla de análisis para configurar una regla de análisis personalizada para esta tabla. Una regla de análisis personalizada limita el acceso a los datos. Puede permitir un conjunto específico de consultas preautorizadas en sus datos o permitir que un conjunto específico de cuentas consulte sus datos.
 - a. Para el tipo de regla de análisis, elija Personalizado y para Método de creación, elija Flujo guiado.
 - b. Elija Siguiente.
 - c. En Especificar los controles de análisis, elija entre Revisar cada análisis nuevo y Permitir cualquier análisis realizado por colaboradores específicos.
 - d. Elija Siguiente.
 - e. (Opcional) En los controles Especificar los resultados del análisis, en el caso de Columnas no permitidas en la salida, especifique si desea excluir alguna columna de la salida. Si selecciona Ninguna, no se excluirá ninguna columna de la salida. Si elige Lista personalizada, puede especificar determinadas columnas que se eliminarán de la salida.
 - f. Para los análisis adicionales aplicados a la salida, especifique si desea permitir, denegar o requerir un análisis adicional antes de que se generen los resultados.
 - g. Elija Siguiente.
 - h. (Opcional) En Establecer privacidad diferencial, seleccione Desactivar.
 - i. Elija Siguiente.
 - j. Revise la información de la página Revisar y configurar y, a continuación, seleccione Configurar regla de análisis.
7. En la página de detalles de la tabla, elija Asociar a la colaboración.
8. En el cuadro de diálogo Asociar tabla, seleccione la colaboración a la que desee asociar esta tabla y elija Elegir colaboración.
9. En la página Asociar la tabla, revise y verifique la información de los detalles de la asociación de tablas, el acceso al servicio y las etiquetas. Seleccione Asociar tabla.
10. En la tabla Tablas asociadas a su tabla, seleccione el botón de opción situado junto a la tabla que acaba de asociar. En el menú Acciones, seleccione Configurar en el grupo de reglas de análisis de colaboración.
11. En la página Configurar la regla de análisis de colaboración, en la sección Análisis adicionales permitidos, elija si algún miembro de la colaboración o miembro específico de la colaboración puede realizar análisis adicionales.

En la opción Entrega de resultados, elija qué miembros pueden recibir los resultados de las consultas.

12. Elija Configurar regla de análisis.

API

Para aportar datos de formación (API)

1. Configure una AWS Glue tabla existente para usarla en ella AWS Clean Rooms proporcionando la tabla y las columnas que se pueden usar.

Ejecute el siguiente código con sus parámetros específicos.

```
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. Configure una regla de análisis personalizada que limite el acceso a sus datos. Puede permitir un conjunto específico de consultas preautorizadas en sus datos o permitir que un conjunto específico de cuentas consulte sus datos.

Ejecuta el siguiente código con tus parámetros específicos.

```
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"
            }
        }
    }
)

```

En este ejemplo, una cuenta específica puede ejecutar cualquier consulta sobre los datos y se requiere un análisis adicional.

3. Asocie una tabla configurada a la colaboración y proporcione una función de acceso al servicio a las AWS Glue tablas.

Ejecute el siguiente código con sus parámetros específicos.

```

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

Este rol de servicio tiene permisos relacionados con las tablas. La función de servicio solo se puede asumir si ejecuta AWS Clean Rooms las consultas permitidas en nombre del miembro que puede realizar la consulta. Ningún miembro de la colaboración (salvo el propietario de los datos) tiene acceso a las tablas subyacentes de la colaboración. El propietario de los datos puede desactivar la privacidad diferencial para que sus tablas estén disponibles para que otros miembros las consulten.

4. Por último, añada una regla de análisis a la asociación de tablas configurada.

Ejecute el siguiente código con sus parámetros específicos.

```

import boto3
acr_client= boto3.client('cleanrooms')

acr_client.create_configured_table_association_analysis_rule(

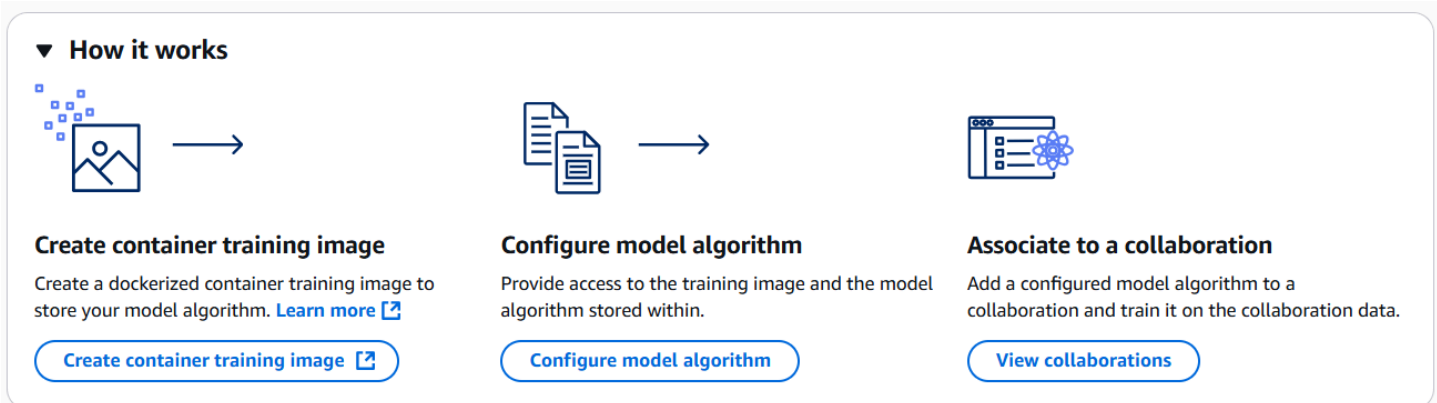
    configuredTableAssociationIdentifier='configured_table_association_identifier',
    membershipIdentifier='membership_id',
    configuredTableIdentifier='configured_table_id',
    analysisRuleType = 'CUSTOM',
    analysisRulePolicy= {
        'v1': {
            'custom': {
                'allowedAdditionalAnalyses':
                ['configured_model_algorithm_association_arns'],
                'allowedResultReceivers': ['query_runner_account']
            }
        }
    }
)

```

Configuración de un algoritmo modelo en AWS Clean Rooms ML

Una vez que haya [creado una imagen de entrenamiento en contenedores](#), debe configurar el algoritmo del modelo. La configuración de un algoritmo modelo permite que esté disponible para asociarlo a una colaboración.

La siguiente imagen muestra la configuración de un algoritmo modelo como un paso que se realiza después de crear la imagen de entrenamiento del contenedor y antes de asociarla a la colaboración.



Console

Para configurar un algoritmo de modelo de aprendizaje automático personalizado (consola)

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija modelos ML personalizados.
3. En la página de modelos ML personalizados, elija Configurar algoritmo de modelo.
4. En la página Configurar el algoritmo del modelo, en los detalles del algoritmo del modelo, introduzca un nombre y, si lo desea, una descripción.
5. Si desea realizar un entrenamiento de modelos, para ver los detalles del contenedor ECR de la imagen de entrenamiento,
 - a. Seleccione la casilla de verificación Especificar el URI de la imagen de entrenamiento.
 - b. Seleccione el repositorio que contiene el modelo de entrenamiento, el contenedor de inferencias o ambos en la lista desplegable.
 - c. Seleccione la imagen.
 - d. (Opcional) Introduzca el valor de los puntos de entrada para acceder a la imagen de entrenamiento.
 - e. (Opcional) Introduzca el valor de los argumentos.
6. (Opcional) Si desea generar informes sobre las métricas del modelo, en el campo de las métricas de entrenamiento, introduzca el nombre de las métricas y la instrucción Regex que buscará la métrica en los registros de salida.
7. Si desea realizar una inferencia de modelos, para ver los detalles del contenedor ECR de la imagen de inferencia,
 - a. Seleccione la casilla de verificación Especificar el URI de la imagen de inferencia.
 - b. Seleccione el repositorio en la lista desplegable.
 - c. Seleccione la imagen.
8. En Acceso a los servicios, elija el Nombre del rol de servicio existente que se utilizará para acceder a esta tabla.
9. Para el cifrado, elija Personalizar la configuración de cifrado para especificar su propia clave de KMS y la información relacionada. De lo contrario, Clean Rooms ML gestionará el cifrado.
10. Si desea activar las etiquetas, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par clave y valor.

11. Elija Configurar el algoritmo del modelo.

API

Para configurar un algoritmo de modelo de aprendizaje automático (API) personalizado

1. Crea una imagen de docker compatible con SageMaker IA. Clean Rooms ML solo admite imágenes de docker compatibles con SageMaker IA.
2. Después de crear una imagen de docker compatible con SageMaker IA, utilice Amazon ECR para crear una imagen de entrenamiento. Siga las instrucciones de la [Guía del usuario de Amazon Elastic Container Registry](#) para crear una imagen de formación sobre contenedores.
3. Configure el algoritmo del modelo para usarlo en Clean Rooms ML. Debe proporcionar la siguiente información:
 - El enlace al repositorio de Amazon ECR y argumentos adicionales para entrenar el modelo y ejecutar inferencias. Clean Rooms ML permite ejecutar trabajos de transformación por lotes en un contenedor de inferencias.
 - Un rol de acceso al servicio que permite a Clean Rooms ML acceder al repositorio.
 - (Opcional) Un contenedor de inferencias. Aunque puede proporcionarlo en un algoritmo modelo configurado por separado, le recomendamos que lo haga en este paso para que tanto el contenedor de entrenamiento como el de inferencia se administren como parte del mismo recurso.

Ejecute el siguiente código con sus parámetros específicos.

```
import boto3
acr_ml_client= boto3.client('cleanroomsm1')

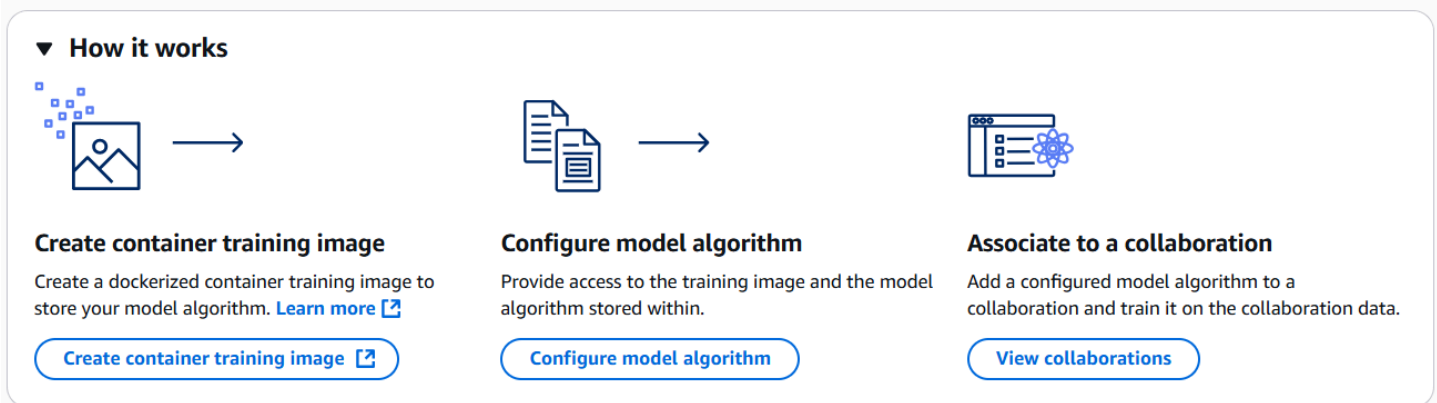
acr_ml_client.create_configured_model_algorithm(
    name='configured_model_algorithm_name',
    trainingContainerConfig={
        'imageUri': 'account.dkr.ecr.region.amazonaws.com/image_name:tag',
        'metricDefinitions': [
            {
                'name': 'custom_metric_name_1',
                'regex': 'custom_metric_regex_1'
            }
        ]
    },
    inferenceContainerConfig={
```

```
'imageUri': 'account.dkr.ecr.region.amazonaws.com/image_name:tag',  
}  
roleArn='arn:aws:iam::account:role/role_name'  
)
```

Asociación del algoritmo del modelo configurado en AWS Clean Rooms ML

Una vez que haya configurado el algoritmo modelo, estará listo para asociarlo a una colaboración. Al asociar un algoritmo modelo, el algoritmo modelo estará disponible para todos los miembros de la colaboración.

La siguiente imagen muestra la asociación del algoritmo del modelo configurado como último paso, después de crear la imagen de entrenamiento del contenedor y configurar un algoritmo del modelo.



Console

Note

Una vez asociado el algoritmo del modelo, no se puede editar. Para realizar cambios, puede eliminar el algoritmo del modelo asociado y asociar uno nuevo.

Para asociar un algoritmo de modelo de aprendizaje automático personalizado (consola)

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija modelos ML personalizados.

3. En la página de modelos de aprendizaje automático personalizados, elija el algoritmo de modelo configurado que desee asociar a una colaboración y, a continuación, elija Asociar a la colaboración.
4. En la ventana Asociar el algoritmo del modelo configurado, elija la colaboración a la que desee asociarse.
5. Seleccione Elegir colaboración.
6. En la página Asociar el algoritmo del modelo, en los detalles de la asociación del algoritmo del modelo, introduzca un nombre y, si lo desea, una descripción.
7. En Algoritmo modelo, elija un algoritmo modelo configurado.
8. Para las configuraciones de privacidad de exportación del modelo entrenado,
 - a. Para exportar archivos de modelo, seleccione la casilla de verificación Archivos de modelo.
 - b. Para exportar los archivos de salida, seleccione la casilla de verificación Archivos de salida.
 - c. Introduzca un valor de tamaño máximo para los datos exportados. El valor debe estar comprendido entre 0,01 y 10.
9. (Opcional) Si desea enviar registros de errores completos o resúmenes de errores más cortos a los miembros, según la configuración de privacidad de tareas de inferencia de modelos entrenados,
 - a. En Registros completos, selecciona una o más cuentas IDs de la lista desplegable.
 - b. (Opcional) Si desea enviar registros que coincidan con un patrón de filtro, introduzca un patrón de filtro.
 - c. (Opcional) Si desea añadir otra cuenta y un patrón de filtro opcional, elija Añadir política de registro.
 - d. En Resúmenes de errores, selecciona una o más cuentas IDs de la lista desplegable.
 - e. (Opcional) Seleccione una o más entidades para redactarlas y especifique qué entidades se eliminarán del registro de errores o de los resúmenes de errores.
 - PII: redacta la información de identificación personal
 - Números: redacta números
 - Personalizado: redacción basada en el patrón de redacción personalizado

- i. Si has elegido Personalizado en el paso anterior, introduce un patrón de redacción personalizado. Esto registra la información que coincide con este patrón.
 - ii. (Opcional) Si quieres añadir otro patrón de redacción personalizado, selecciona Añadir otro patrón personalizado.
10. (Opcional) Si desea configurar métricas de modelos entrenados, en Configuración de métricas de modelos entrenados, seleccione un nivel de ruido de la lista desplegable.

Puede elegir Ninguno, Bajo, Medio y Alto.
11. (Opcional) Si desea establecer el tamaño máximo de los artefactos, en Configuración de artefactos, introduzca el valor del tamaño máximo de los artefactos. El valor debe estar comprendido entre 0,01 y 10.
12. (Opcional) Si desea activar las etiquetas, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par clave y valor.
13. Elija Asociar .

API

Para asociar un algoritmo de modelo de aprendizaje automático (API) personalizado

Ejecute el siguiente código con sus parámetros específicos.

También proporciona una política de privacidad que define quién tiene acceso a los diferentes registros, permite a los clientes definir las expresiones regulares y cuántos datos se pueden exportar de los resultados del modelo de entrenamiento o de las inferencias.

Note

Las asociaciones de algoritmos de modelos configurados son inmutables.

```
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'
}
)

```

Una vez que el algoritmo modelo configurado esté asociado a la colaboración, los proveedores de datos de formación deben añadir una regla de análisis de colaboración a su tabla. Esta regla permite que la asociación de algoritmos del modelo configurado acceda a su tabla configurada. Todos los proveedores de datos de formación que contribuyan deben ejecutar el siguiente código:

```

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

Como las asociaciones de algoritmos de modelos configurados son inmutables, recomendamos que los proveedores de datos de formación que deseen permitir su uso en listas de modelos utilicen caracteres comodín `allowedAdditionalAnalyses` durante las primeras iteraciones de la configuración del modelo personalizado. Esto permite a los proveedores de modelos iterar su código sin necesidad de que otros proveedores de formación vuelvan a asociarlo antes de entrenar su código de modelo actualizado con datos.

Creación de un canal de entrada de ML en AWS Clean Rooms ML

Requisitos previos:

- Y Cuenta de AWS con acceso a AWS Clean Rooms
- Una colaboración configurada en la AWS Clean Rooms que desea crear el canal de entrada ML
- Permisos para consultar datos y crear canales de entrada de aprendizaje automático en la colaboración.
- (Opcional) Un algoritmo modelo existente para asociarlo al canal de entrada de ML o permisos para crear uno nuevo
- (Opcional) Tablas con reglas de análisis que se pueden ejecutar para el modelo especificado.
- (Opcional) Una plantilla de consulta o análisis SQL existente que se utilizará para generar el conjunto de datos
- (Opcional) Un rol de servicio existente con los permisos adecuados o permisos para crear un nuevo rol de servicio
- (Opcional) Una AWS KMS clave personalizada si quieres usar tu propia clave de cifrado
- Permisos adecuados para crear y administrar modelos de aprendizaje automático en la colaboración

Un canal de entrada de ML es un conjunto de datos que se crea a partir de una consulta de datos específica. Los miembros con la capacidad de consultar datos pueden preparar sus datos para

el entrenamiento y la inferencia mediante la creación de un canal de entrada de aprendizaje automático. La creación de un canal de entrada de aprendizaje automático permite que los datos se utilicen en diferentes modelos de entrenamiento dentro de la misma colaboración. Debe crear canales de entrada de aprendizaje automático independientes para el entrenamiento y la inferencia.

Para crear un canal de entrada de ML, debe especificar la consulta SQL que se utiliza para consultar los datos de entrada y crear el canal de entrada de ML. Los resultados de esta consulta nunca se comparten con ningún miembro y permanecen dentro de los límites de Clean Rooms ML. El nombre de recurso de Amazon (ARN) de referencia se utiliza en los siguientes pasos para entrenar un modelo o ejecutar una inferencia.

Console

Para crear un canal de entrada de aprendizaje automático (consola)

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página de colaboraciones, elija la colaboración en la que desee crear un canal de entrada de aprendizaje automático.
4. Cuando se abra la colaboración, selecciona la pestaña de modelos de aprendizaje automático.
5. En Modelos ML personalizados, en la sección Canales de entrada ML, selecciona Crear canal de entrada ML.
6. En la página Crear canal de entrada ML, para ver los detalles del canal de entrada ML, haga lo siguiente:
 - a. En Nombre, introduce un nombre único para tu canal.
 - b. (Opcional) En Descripción, introduce una descripción de tu canal.
 - c. En Algoritmo del modelo asociado, seleccione el algoritmo que desee utilizar.

Elija Asociar algoritmo de modelo para agregar uno nuevo.

7. En Dataset, elige un método para generar el conjunto de datos de entrenamiento:
 - Elija una consulta SQL para usar los resultados de una consulta SQL como conjunto de datos de entrenamiento.

Si eligió una consulta SQL, introduzca la consulta en el campo de consulta SQL.

(Opcional) Para importar una consulta que haya utilizado recientemente, seleccione Importar de consultas recientes.

- Elija una plantilla de análisis para usar los resultados de una plantilla de análisis como conjunto de datos de entrenamiento.

 Warning

La generación de datos sintéticos protege contra la inferencia de atributos individuales, ya sea que haya personas específicas presentes en el conjunto de datos original o que estén presentes los atributos de aprendizaje de esas personas. Sin embargo, no impide que los valores literales del conjunto de datos original, incluida la información de identificación personal (PII), aparezcan en el conjunto de datos sintético.

Recomendamos evitar los valores en el conjunto de datos de entrada que estén asociados a un solo sujeto de datos, ya que pueden volver a identificar a un sujeto de datos. Por ejemplo, si solo un usuario vive en un código postal, la presencia de ese código postal en el conjunto de datos sintético confirmaría que el usuario estaba en el conjunto de datos original. Para mitigar este riesgo, se pueden utilizar técnicas como truncar valores de alta precisión o reemplazar catálogos poco comunes por otros. Estas transformaciones pueden formar parte de la consulta utilizada para crear el canal de entrada de ML.

- a. Si no hay tablas asociadas, elija Asociar tabla para agregar tablas con una regla de análisis que se pueda ejecutar para el modelo especificado.
- b. Elija el tipo de trabajador que se utilizará al crear este canal de datos. El tipo de trabajador predeterminado es CR.1X. Especifique el número de trabajadores que se van a utilizar. El número de trabajador predeterminado es 16. Para especificar las propiedades de Spark:
 - i. Amplíe Propiedades de Spark.
 - ii. Selecciona Añadir propiedades de Spark.
 - iii. En el cuadro de diálogo de propiedades de Spark, elija un nombre de propiedad en la lista desplegable e introduzca un valor.

En la siguiente tabla se proporciona una definición para cada propiedad.

Para obtener más información sobre las propiedades de Spark, consulte [Propiedades de Spark](#) en la documentación de Apache Spark.

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Fallos en Spark.task.max	Controla cuántas veces consecutivas puede fallar una tarea antes de que el trabajo falle. Requiere un valor mayor o igual a 1. El número de reintentos permitidos es igual a este valor menos 1. El recuento de errores se restablece si algún intento tiene éxito. Los errores en las distintas tareas no se acumulan hasta alcanzar este límite.	4
spark.sql.files.maxPartitionBytes	Establece el número máximo de bytes que se deben empaquetar en una sola partición al leer fuentes basadas en archivos, como Parquet, JSON y ORC.	128 MB
Spark.HadoopFS.S3.max retries	Establece el número máximo de reintentos para las operaciones con archivos de Amazon S3.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.network.timeout	Establece el tiempo de espera predeterminado para todas las interacciones de la red. Anula los siguientes ajustes de tiempo de espera si no están configurados: • spark.storage.blockManagerHeartbeatTimeoutMs • Tiempo de espera de conexión de Spark.Shuffle.io • Spark.rpc.askTimeout • Se ha agotado el tiempo de espera de Spark.rpc Lookup	120 s
spark.rdd.compress	Especifica si se deben comprimir las particiones RDD serializadas mediante spark.io.compression.codec. Se aplica a StorageLevel .MEMORY_ONLY_SER en Java y Scala, o a .MEMORY_ONLY en Python. StorageLevel Reduce el espacio de almacenamiento, pero requiere más tiempo de procesamiento de la CPU.	FALSO
spark.shuffle.spill.compress	Especifica si se deben comprimir los datos de Shuffle Spill mediante spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Establece el tamaño objetivo en bytes para las particiones aleatorias durante la optimización adaptativa cuando spark.sql.adaptive.enabled es true. Controla el tamaño de las particiones al unir particiones pequeñas o dividir particiones asimétricas.	(valor de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive. .autoBroadcastJoinUmbral	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Se aplica solo en un marco adaptativo. Utiliza el mismo valor predeterminado que spark.sql.adaptive.autoBroadcastJoinUmbral. Configúrelo en -1 para deshabilitar la transmisión.	(ninguno)
spark.sql.adaptive. .coalescePartitions.enabled	Especifica si se van a unir particiones aleatorias contiguas en función de spark.sql.adaptive.advisoryPartitionSizeInBytes para optimizar el tamaño de la tarea. Requiere que spark.sql.adaptive.coalescePartitions.enabled sea verdadero.	TRUE
Particiones spark.sql.Adaptive. .Coalesce.initialPartitionNum	Define el número inicial de particiones de mezcla antes de la fusión. Requiere que spark.sql.adaptive.enabled y spark.sql.Adaptive.coalescePartitions.enabled sean verdaderos. El valor predeterminado es spark.sql.shuffle.partition.	(ninguno)
Spark.sql.Adaptive. .CoalescePartitions. minPartitionSize	Establece el tamaño mínimo de las particiones aleatorias combinadas para evitar que las particiones se vuelvan demasiado pequeñas durante la optimización adaptativa.	1 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive.coalescePartitions.ParalelismFirst	Especifica si se deben calcular los tamaños de las particiones en función del paralelismo del clúster en lugar de en spark.sql.adaptive.advisoryPartitionSizeInBytes durante la fusión de particiones. Genera tamaños de partición más pequeños que el tamaño objetivo configurado para maximizar el paralelismo. Se recomienda establecer este valor en false en los clústeres ocupados para mejorar la utilización de los recursos y evitar el exceso de tareas pequeñas.	TRUE
spark.sql.adaptive.enabled	Especifica si se debe habilitar la ejecución adaptativa de consultas para volver a optimizar los planes de consultas durante la ejecución de las consultas, en función de estadísticas de tiempo de ejecución precisas.	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoin;Únete!	Especifica si se debe forzar la activación OptimizeSkewedJoin incluso si se introduce una mezcla adicional.	FALSO
spark.sql.adaptive.localShuffleReader.habilitado	Especifica si se deben utilizar lectores aleatorios locales cuando no se requiera la creación de particiones aleatorias, por ejemplo, después de la conversión de uniones por ordenación y combinación a uniones por hash de difusión. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	Establece el tamaño máximo de partición en bytes para crear mapas hash locales. Prioriza las uniones hash mezcladas sobre las uniones por ordenación y combinación cuando: - Este valor es igual o superior a <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> - Todos los tamaños de partición están dentro de este límite Anula <code>spark.sql.join.preferSortMergeUnion</code> a la configuración.	0 bytes
<code>spark.sql.adaptive.optimizeShuffleRebalancePartitions.enabled</code>	Especifica si se deben optimizar las particiones aleatorias asimétricas dividiéndolas en particiones más pequeñas según <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> . Requiere que <code>spark.sql.adaptive.enabled</code> sea verdadero.	TRUE
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	Define el factor de umbral de tamaño para la fusión de particiones durante la división. Las particiones más pequeñas que este factor se multiplican por <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> están fusionadas.	0.2

Nombre de propiedad	Description (Descripción)	Valor predeterminado
<code>spark.sql.adaptive.skewJoin.enabled</code>	Especifica si se debe tratar la asimetría de datos en las uniones mezcladas dividiendo y, opcionalmente, replicando las particiones asimétricas. Se aplica a las uniones hash ordenadas, combinadas y barajadas. Requiere que <code>spark.sql.adaptive.enabled</code> sea verdadero.	TRUE
<code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code>	Determina el factor de tamaño que determina el sesgo de la partición. Una partición está sesgada cuando su tamaño supera las dos condiciones: • Este factor se multiplica por la mediana del tamaño de la partición • El valor de <code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code>	5
<code>spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes</code>	Establece el umbral de tamaño en bytes para identificar las particiones asimétricas. Una partición está sesgada cuando su tamaño supera las dos condiciones: • Este umbral • El tamaño medio de la partición multiplicado por <code>spark.sql.adaptive.skewJoin.skewedPartitionFactor</code> Recomendamos establecer este valor por encima de <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>.	256 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.autoBroadcastJoinThreshold	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Configúrelo en -1 para deshabilitar la transmisión.	10 MB
Spark.sql.Tiempo de espera de emisión	Controla el tiempo de espera en segundos para las operaciones de transmisión durante las uniones de transmisión.	300 segundos
spark.sql.cbo.enabled	Especifica si se debe habilitar la optimización basada en costes (CBO) para la estimación de las estadísticas del plan.	FALSO
spark.sql.cbo.joinReorder.dp.starFilter	Especifica si se debe aplicar la heurística del filtro de unión en estrella durante la enumeración de uniones basada en el coste.	FALSO
spark.sql.cbo.joinReorder.dp.Threshold	Establece el número máximo de nodos unidos permitidos en el algoritmo de programación dinámica.	12
spark.sql.cbo.joinReorder.Enabled	Especifica si se debe habilitar el reordenamiento de las uniones en la optimización basada en costes (CBO).	FALSO
Spark.sql.cbo.PlanStats.Enabled	Especifica si se deben obtener los recuentos de filas y las estadísticas de las columnas del catálogo durante la generación del plan lógico.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.cbo.starSchemaDetection	Especifica si se debe habilitar el reordenamiento de las uniones en función de la detección del esquema en estrella.	FALSO
spark.sql.files.maxPartitionNum	Establece el número máximo objetivo de particiones de archivos divididas para fuentes basadas en archivos (Parquet, JSON y ORC). Cambia la escala de las particiones cuando el recuento inicial supera este valor. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.files.maxRecordsPerFile	Establece el número máximo de registros que se van a escribir en un único archivo. No se aplica ningún límite cuando se establece en cero o en un valor negativo.	0
spark.sql.files.minPartitionNum	Establece el número mínimo objetivo de particiones de archivos divididas para las fuentes basadas en archivos (Parquet, JSON y ORC). El valor predeterminado es spark.sql.leafNodeDefaultParallelism. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.inMemoryColumnarStorageSizeInMemoryColumnarStorage	Controla el tamaño del lote para el almacenamiento en caché en columnas. El aumento del tamaño mejora la utilización y la compresión de la memoria, pero aumenta el riesgo de out-of-memory errores.	10000

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.inMemoryColumnarAlchemyComprimido	Especifica si se seleccionan automáticamente los códecs de compresión para las columnas en función de las estadísticas de datos.	TRUE
spark.sql.inMemoryColumnarAlchemy.enableVectorizedReader	Especifica si se debe habilitar la lectura vectorizada para el almacenamiento en caché de columnas.	TRUE
spark.sql.legacy.allowHashOnMapType	Especifica si se permiten operaciones de hash en estructuras de datos de tipo mapa. Esta configuración antigua mantiene la compatibilidad con el manejo de tipos de mapas de las versiones anteriores de Spark.	
spark.sql.legacy.allowNegativeScaleOfDecimal	Especifica si se permiten valores de escala negativos en las definiciones de tipo decimal. Esta configuración antigua mantiene la compatibilidad con las versiones anteriores de Spark que admitían escalas decimales negativas.	
spark.sql.legacy.castComplexTypesToString.habilitado	Especifica si se debe habilitar el comportamiento heredado para convertir tipos complejos en cadenas. Mantiene la compatibilidad con las reglas de conversión de tipos de las versiones anteriores de Spark.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.charVarcharAsCadena	Especifica si se deben tratar los tipos CHAR y VARCHAR como tipos STRING. Esta configuración antigua proporciona compatibilidad con el manejo de tipos de cadenas de las versiones anteriores de Spark.	
spark.sql.legacy.createEmptyCollectionUsingStringType	Especifica si se van a crear colecciones vacías mediante elementos de tipo cadena. Esta configuración antigua mantiene la compatibilidad con el comportamiento de inicialización de colecciones de versiones anteriores de Spark.	
spark.sql.legacy.exponentLiteralAsDecimal. Habilitado	Especifica si se deben interpretar los literales exponenciales como tipos decimales. Esta configuración antigua mantiene la compatibilidad con el manejo literal numérico de las versiones anteriores de Spark.	
spark.sql.legacy.json.allowEmptyString.habilitado	Especifica si se permiten cadenas vacías en el procesamiento de JSON. Esta configuración antigua mantiene la compatibilidad con el comportamiento de análisis de JSON de las versiones anteriores de Spark.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.parquet.int96RebaseModelRead	Especifica si se debe utilizar el modo de rebase de marcas de tiempo heredado al leer los archivos de Parquet. INT96 Esta configuración antigua mantiene la compatibilidad con el manejo de las marcas de tiempo de las versiones anteriores de Spark.	
spark.sql.legacy.timeParserPolicy	Controla el comportamiento del análisis temporal para garantizar la compatibilidad con versiones anteriores. Esta configuración antigua determina cómo se analizan las marcas de tiempo y las fechas a partir de las cadenas.	
spark.sql.legacy.TypeCoercion.datetimeToString.habilitado	Especifica si se debe habilitar el comportamiento de coerción de tipo heredado al convertir valores de fecha y hora en cadenas. Mantiene la compatibilidad con las reglas de conversión de fecha y hora de las versiones anteriores de Spark.	
spark.sql.maxSinglePartitionBytes	Establece el tamaño máximo de la partición en bytes. El planificador introduce operaciones de mezcla para particiones más grandes a fin de mejorar el paralelismo.	128 m

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Caché de metadatos Spark.sql.TTLSeconds	Controla el time-to-live (TTL) de las cachés de metadatos. Se aplica a los metadatos de los archivos de partición y a las cachés del catálogo de sesiones. Requiere: • Un valor positivo superior a cero • Spark.sql.CatalogImplementation está configurado como hive • spark.sql.hive.filesourcePartitionFileCacheSize mayor que cero • spark.sql.hive.manageFilesourcePartitions establecido en verdadero	-1000 ms
spark.sql.optimizeProjectAlwaysEn línea	Especifica si se deben contraer las proyecciones adyacentes y las expresiones en línea, incluso cuando esto provoque duplicación.	FALSO
spark.sql.optimizeDynamicPartitionPruning.habilitado	Especifica si se van a generar predicados para las columnas de partición utilizadas como claves de unión.	TRUE
spark.sql.optimizeEnableCsvExpressionOptimización	Especifica si se deben optimizar las expresiones CSV en el optimizador de SQL eliminando las columnas innecesarias de las operaciones from_csv.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimizer.enableJsonExpressionOptimization	Especifica si se deben optimizar las expresiones JSON en el optimizador de SQL de la siguiente manera: • Eliminar columnas innecesarias de las operaciones de from_json • Simplificar las combinaciones from_json y to_json • Optimización de las operaciones de named_struct	TRUE
Spark.sql.Optimizer.ExcludedRules	Define las reglas del optimizador que se van a deshabilitar, identificadas mediante nombres de reglas separados por comas. Algunas reglas no se pueden deshabilitar porque son necesarias para que sean correctas. El optimizador registra qué reglas se han desactivado correctamente.	(ninguno)
spark.sql.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Establece el tamaño mínimo de escaneo agregado en bytes necesario para inyectar un filtro Bloom en el lado de la aplicación.	10 GB
Filtro Spark.Sql.Optimizer.Runtime.BloomFilter.creationSideThreshold	Define el umbral de tamaño máximo para inyectar un filtro Bloom en el lado de la creación.	10 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Spark.Sql.Optimizer.Runtime.BloomFilter.Enabled	Especifica si se debe insertar un filtro Bloom para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	TRUE
spark.sql.optimizer.Runtime.BloomFilter.expectedNumItems	Define el número predeterminado de elementos esperados en el filtro Bloom en tiempo de ejecución.	1000000
spark.sql.Optimizer.Runtime.BloomFilter.maxNumBits	Establece el número máximo de bits permitido en el filtro Bloom en tiempo de ejecución.	67108864
Spark.Sql.Optimizer.Runtime.BloomFilter.maxNumItems	Establece el número máximo de elementos esperados permitidos en el filtro Bloom en tiempo de ejecución.	4000000
spark.sql.Optimizer.Runtime.BloomFilter.Number.Threshold	Limita el número máximo de filtros de tiempo de ejecución distintos de DPP permitidos por consulta para evitar errores en el controlador. out-of-memory	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Define el número predeterminado de bits que se utilizan en el filtro Bloom en tiempo de ejecución.	8388608

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimizer.runtime.rowlevelOperationGroupFiltero. Habilitado	Especifica si se debe habilitar el filtrado de grupos en tiempo de ejecución para las operaciones a nivel de fila. Permite que las fuentes de datos: • Elimine grupos completos de datos (como archivos o particiones) mediante filtros de fuentes de datos • Ejecute consultas en tiempo de ejecución para identificar los registros coincidentes • Deseche los grupos innecesarios para evitar costosas reescrituras Limitaciones: • No todas las expresiones se pueden convertir en filtros de fuentes de datos • Algunas expresiones requieren la evaluación de Spark (como las subconsultas)	TRUE
spark.sql.Optimizer.RuntimeFilter.semiJoinReduction.habilitado	Especifica si se debe insertar una semiunión para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.parquet.aggregatePushdown	Especifica si se deben enviar los agregados a Parquet para su optimización. Soporta: • MIN y MAX para los tipos booleano, entero, flotante y de fecha • COUNT para todos los tipos de datos Lanza una excepción si faltan estadísticas en el pie de página de un archivo de Parquet.	FALSO
spark.sql.parquet.columnarReaderBatchSize	Controla el número de filas de cada lote de lectores vectorizados Parquet. Elija un valor que equilibre la sobrecarga de rendimiento y el uso de memoria para evitar out-of-memory errores.	4096
Spark.sql.Session.TimeZone	Define la zona horaria de la sesión para gestionar las marcas de tiempo en los literales de cadena y la conversión de objetos Java. Acepta: • area/city Formato basado IDs en regiones (como America/Los_Angeles) • Desplazamientos de zona en formato (+/-) HH, (+/-) HH:mm o (+/-) HH:mm:ss (como -08 o + 01:00) • UTC o Z como alias para + 00:00	(valor de la zona horaria local)

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.shuffle.partition	Establece el número predeterminado de particiones para la mezcla de datos durante las uniones o agregaciones. No se puede modificar entre reinicios de consultas de streaming estructuradas desde la misma ubicación del punto de control.	200
spark.sql.shuffleJoinFactor	Define el factor de multiplicación utilizado para determinar la idoneidad de las combinaciones de hash aleatorio. Se selecciona una combinación de hash aleatorio cuando el tamaño de los datos del lado pequeño multiplicado por este factor es menor que el tamaño de los datos del lado grande.	3
spark.sql.sources.parallelPartitionDiscovery.umbra	Establece el número máximo de rutas para la lista de archivos del lado del conductor con fuentes basadas en archivos (Parquet, JSON y ORC). Si se supera durante la detección de particiones, los archivos se enumeran mediante un trabajo distribuido de Spark independiente.	32
spark.sql.statistics.histogram.enabled	Especifica si se deben generar histogramas con la misma altura durante el cálculo de las estadísticas de columnas para mejorar la precisión de la estimación. Requiere un escaneo de tabla adicional al necesario para las estadísticas de columnas básicas.	FALSO

- c. En Retención de datos en días, introduzca el número de días que desea conservar los datos.
 - d. En el formato de resultado, elija CSV o Parquet como formato de datos que debe usar el canal de entrada ML.
8. Para acceder al servicio, elija el nombre del rol de servicio existente que se usará para acceder a esta tabla o elija Crear y usar un nuevo rol de servicio.
 9. Para el cifrado, elija el secreto de cifrado con una clave de KMS personalizada para especificar su propia clave de KMS y la información relacionada. De lo contrario, Clean Rooms ML gestionará el cifrado.
 10. Elija Crear canal de entrada ML.

La creación del canal de entrada ML tardará unos minutos. Puede ver una lista de canales de entrada ML en la pestaña de modelos ML.

 Note

Una vez creado el canal de entrada ML, no podrá editarlo.

API

Para crear un canal de entrada (API) de ML

Ejecute el siguiente código con sus parámetros específicos:

```
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']

```

Creación de un modelo entrenado en AWS Clean Rooms ML

Requisitos previos:

- Y Cuenta de AWS con acceso a AWS Clean Rooms
- Una colaboración creada en AWS Clean Rooms
- Un algoritmo modelo configurado asociado a la colaboración
- Al menos un canal de entrada ML configurado
- Permisos adecuados para crear y administrar modelos de aprendizaje automático en la colaboración

Una vez que haya asociado el algoritmo del modelo configurado a una colaboración y, a continuación, haya creado y configurado un canal de entrada de aprendizaje automático, estará listo para crear un modelo entrenado. Los miembros de una colaboración utilizan un modelo entrenado para analizar sus datos de forma conjunta.

Puede crear un modelo entrenado mediante el siguiente procedimiento.

Como alternativa, puede utilizar el entrenamiento incremental para mejorar un modelo existente con nuevos datos o el entrenamiento distribuido para entrenar modelos en varias instancias de procesamiento.

Temas

- [Uso de la formación incremental en AWS Clean Rooms ML](#)
- [Uso de la formación distribuida en AWS Clean Rooms ML](#)

Console

Para crear un modelo entrenado (consola)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página de colaboraciones, elija la colaboración en la que desee crear un modelo entrenado.
4. Cuando se abra la colaboración, selecciona la pestaña de modelos de aprendizaje automático.
5. En Modelos de aprendizaje automático personalizados, en la sección Modelos entrenados, elija Crear modelo entrenado.
6. En la página Crear un modelo entrenado, en Algoritmo de modelo asociado, especifique el algoritmo.
7. En los detalles del modelo entrenado, introduzca lo siguiente:
 - a. En Nombre, introduzca un nombre exclusivo para el modelo de la colaboración.
 - b. (Opcional) En Descripción, introduzca una descripción del modelo entrenado.
 - c. En el modo de entrada de datos de entrenamiento, elija una de las siguientes opciones:
 - Seleccione Archivo si tiene un conjunto de datos más pequeño que pueda caber en el volumen de almacenamiento de aprendizaje automático y si prefiere el acceso al sistema de archivos tradicional para su guion de entrenamiento.
 - Seleccione Pipe para obtener conjuntos de datos de gran tamaño para transmitir datos directamente desde S3 y así evitar tener que descargar todo al disco, lo que puede mejorar la velocidad de entrenamiento y reducir los requisitos de almacenamiento.

- Seleccione FastFile esta opción si desea combinar las ventajas del streaming desde S3 con el acceso al sistema de archivos, especialmente para leer datos de forma secuencial o si utiliza menos archivos para acelerar los tiempos de inicio.

8. Para obtener los detalles del canal de entrada ML, haga lo siguiente:

- a. Para el canal de entrada ML, especifique el canal de entrada ML que proporciona datos al algoritmo del modelo.

Para añadir otro canal, elija Añadir otro canal de entrada ML. Puede añadir hasta 19 canales de entrada ML adicionales.

- b. En Nombre del canal, introduzca el nombre del canal de entrada ML.
- c. Para el tipo de distribución de datos de Amazon S3, elija una de las siguientes opciones:
- Seleccione Totalmente replicado para proporcionar a cada instancia de entrenamiento una copia completa de su conjunto de datos. Esto funciona mejor cuando el conjunto de datos es lo suficientemente pequeño como para caber en la memoria o cuando cada instancia necesita acceder a todos los datos.
 - Selecciona Fragmentado por la clave S3 para dividir el conjunto de datos entre las instancias de entrenamiento en función de las claves S3. Cada instancia recibe aproximadamente $1/n$ del total de objetos de S3, donde «n» es el número de instancias. Esto funciona mejor para conjuntos de datos grandes que desee procesar en paralelo.

 Note

Tenga en cuenta el tamaño del conjunto de datos y los requisitos de formación al seleccionar un tipo de distribución. La replicación completa proporciona un acceso completo a los datos, pero requiere más almacenamiento, mientras que la clave Sharded by S3 permite el procesamiento distribuido de grandes conjuntos de datos.

9. En Duración máxima del entrenamiento, elige la cantidad máxima de tiempo que desees entrenar tu modelo.
10. En el caso de los hiperparámetros, especifique los parámetros específicos del algoritmo y sus valores previstos. Los hiperparámetros son específicos del modelo que se está entrenando y se utilizan para ajustar el entrenamiento del modelo.

11. En el caso de las variables de entorno, especifique cualquier variable específica del algoritmo y sus valores previstos. Las variables de entorno se configuran en el contenedor de Docker.
12. En el caso del cifrado, para usar una personalizada AWS KMS key, selecciona la casilla de verificación Cifrar el secreto con una clave KMS personalizada.
13. Para la configuración de recursos de EC2, especifique la información sobre los recursos informáticos que se utilizan para el entrenamiento del modelo.
 - a. En Tipo de instancia, elija el tipo de instancia que desee ejecutar.
 - b. En Recuento de instancias, introduzca el número de instancias.
 - c. Para el tamaño del volumen en GB, introduzca el tamaño del volumen de almacenamiento de ML.
14. Elija Crear modelo entrenado.

API

Para crear un modelo entrenado (API)

El miembro con la capacidad de entrenar un modelo comienza a entrenar seleccionando el canal de entrada de ML y el algoritmo del modelo.

Ejecute el siguiente código con sus parámetros específicos:

```
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

Una vez creado el modelo entrenado, no podrá editarlo. Para realizar cambios, elimine el modelo entrenado y cree uno nuevo.

Uso de la formación incremental en AWS Clean Rooms ML

Requisitos previos:

- Y Cuenta de AWS con acceso a AWS Clean Rooms
- Un modelo formado existente en una colaboración
- Un conjunto de datos nuevo o actualizado para el entrenamiento incremental
- Permisos adecuados para crear y gestionar modelos de aprendizaje automático en la colaboración
- Familiaridad con los hiperparámetros y la configuración del modelo existente

Con el entrenamiento incremental, puede usar los artefactos de un modelo existente y un conjunto de datos actualizado para entrenar un modelo nuevo. La capacitación incremental le ahorra tiempo y recursos.

Utilice la capacitación incremental para:

- Entrene un modelo nuevo con un conjunto de datos ampliado que tenga un patrón subyacente que no se tuvo en cuenta en el entrenamiento anterior.
- Entrene varias variantes de un modelo, ya sea con diferentes hiperparámetros o utilizando diferentes conjuntos de datos.

Console

Para ejecutar un trabajo de entrenamiento incremental (consola)

1. Inicie sesión en <https://console.aws.amazon.com/cleanrooms> Consola de administración de AWS y abra la AWS Clean Rooms consola.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página de colaboraciones, elija la colaboración en la que estén presentes los artefactos modelo que desee utilizar para el entrenamiento incremental.
4. Cuando se abra la colaboración, seleccione la pestaña de modelos de aprendizaje automático.
5. En Modelos de aprendizaje automático personalizados, en la sección Modelos entrenados, selecciona el botón de radio situado junto al modelo entrenado que desees entrenar de forma incremental.
6. En la página de descripción general, en Versiones,
 - a. Selecciona el botón de radio situado junto al modelo entrenado que desees entrenar de forma incremental.
 - b. Elige Tren desde la versión.
7. En la página Crear un modelo entrenado a partir de una versión, en Versión de modelo entrenado, elija la versión.

La versión del modelo base se selecciona automáticamente. Puede cambiar esta versión si existen otras versiones.

8. Para ver los detalles del modelo entrenado, introduzca lo siguiente:
 - a. En Nombre, introduzca un nombre exclusivo para el modelo de la colaboración.
 - b. (Opcional) En Descripción, introduzca una descripción del modelo entrenado.
 - c. En el modo de entrada de datos de entrenamiento, elija una de las siguientes opciones:
 - Seleccione Archivo si tiene un conjunto de datos más pequeño que pueda caber en el volumen de almacenamiento de aprendizaje automático y si prefiere el acceso al sistema de archivos tradicional para su guion de entrenamiento.
 - Seleccione Pipe para obtener conjuntos de datos de gran tamaño para transmitir datos directamente desde S3 y así evitar tener que descargar todo al disco, lo que puede mejorar la velocidad de entrenamiento y reducir los requisitos de almacenamiento.

- Seleccione FastFile esta opción si desea combinar las ventajas del streaming desde S3 con el acceso al sistema de archivos, especialmente para leer datos de forma secuencial o si utiliza menos archivos para acelerar los tiempos de inicio.
- d. En el nombre del canal de formación incremental, introduzca un nombre para el canal de formación incremental

 Note

Si especificas el nombre del canal de entrenamiento incremental sin un identificador de versión, el sistema utiliza el modelo base para el entrenamiento incremental.

9. Para obtener los detalles del canal de entrada de ML, haga lo siguiente:
 - a. Para el canal de entrada ML, especifique el canal de entrada ML que proporciona datos al algoritmo del modelo.

Para añadir otro canal, elija Añadir otro canal de entrada ML. Puede añadir hasta 19 canales de entrada ML adicionales.
 - b. En Nombre del canal, introduzca el nombre del canal de entrada ML.
 - c. Para el tipo de distribución de datos de Amazon S3, elija una de las siguientes opciones:
 - Seleccione Totalmente replicado para proporcionar a cada instancia de entrenamiento una copia completa de su conjunto de datos. Esto funciona mejor cuando el conjunto de datos es lo suficientemente pequeño como para caber en la memoria o cuando cada instancia necesita acceder a todos los datos.
 - Selecciona Fragmentado por la clave S3 para dividir el conjunto de datos entre las instancias de entrenamiento en función de las claves S3. Cada instancia recibe aproximadamente $1/n$ del total de objetos de S3, donde «n» es el número de instancias. Esto funciona mejor para conjuntos de datos grandes que desee procesar en paralelo.

 Note

Tenga en cuenta el tamaño del conjunto de datos y los requisitos de formación al seleccionar un tipo de distribución. La replicación completa proporciona un acceso completo a los datos, pero requiere más almacenamiento, mientras

que la clave Sharded by S3 permite el procesamiento distribuido de grandes conjuntos de datos.

10. En Duración máxima del entrenamiento, elige la cantidad máxima de tiempo que deseas entrenar tu modelo.
11. En el caso de los hiperparámetros, especifique los parámetros específicos del algoritmo y sus valores previstos. Los hiperparámetros son específicos del modelo que se está entrenando y se utilizan para ajustar el entrenamiento del modelo.
12. En el caso de las variables de entorno, especifique cualquier variable específica del algoritmo y sus valores previstos. Las variables de entorno se configuran en el contenedor de Docker.
13. En el caso del cifrado, para usar una personalizada AWS KMS key, selecciona la casilla de verificación Cifrar el secreto con una clave KMS personalizada.
14. Para la configuración de recursos de EC2, especifique la información sobre los recursos informáticos que se utilizan para el entrenamiento del modelo.
 - a. En Tipo de instancia, elija el tipo de instancia que desee ejecutar.
 - b. En Recuento de instancias, introduzca el número de instancias.
 - c. Para el tamaño del volumen en GB, introduzca el tamaño del volumen de almacenamiento de ML.
15. Elija Crear un modelo entrenado a partir de la versión.

API

Para ejecutar un trabajo de formación incremental (API)

Ejecute el siguiente código con sus parámetros específicos:

```
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

Límite: máximo 20 canales en total (incluidos ambos `dataChannels` y `incrementalTrainingDataChannels`).

 Note

Una vez creado el modelo entrenado, no podrá editarlo. Para realizar cambios, elimine el modelo entrenado y cree uno nuevo.

Uso de la formación distribuida en AWS Clean Rooms ML

Requisitos previos:

- Y Cuenta de AWS con acceso a AWS Clean Rooms
- Una colaboración creada en AWS Clean Rooms
- Un algoritmo modelo configurado que admite el entrenamiento distribuido

- Un conjunto de datos grande adecuado para el procesamiento distribuido
- Permisos adecuados para crear y administrar modelos de aprendizaje automático en la colaboración
- Cuota de Amazon EC2 suficiente para ejecutar varias instancias para una formación distribuida

El entrenamiento distribuido aprovecha la potencia de muchos nodos de computación que trabajan en paralelo para procesar grandes cantidades de datos y actualizar los parámetros del modelo de manera eficiente.

Para obtener más información sobre la formación distribuida, consulte [Conceptos de formación distribuida](#) en la Guía para desarrolladores de Amazon SageMaker AI.

Console

Para ejecutar un trabajo de formación distribuido (consola)

1. Inicie sesión en la AWS Clean Rooms consola Consola de administración de AWS y ábrala en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página de colaboraciones, elija la colaboración en la que desee crear un modelo entrenado.
4. Cuando se abra la colaboración, selecciona la pestaña de modelos de aprendizaje automático.
5. En Modelos de aprendizaje automático personalizados, en la sección Modelos entrenados, elija Crear modelo entrenado.
6. En la página Crear un modelo entrenado, en Algoritmo de modelo asociado, especifique el algoritmo
7. En los detalles del modelo entrenado, introduzca lo siguiente:
 - a. En Nombre, introduzca un nombre exclusivo para el modelo de la colaboración.
 - b. (Opcional) En Descripción, introduzca una descripción del modelo entrenado.
 - c. En el modo de entrada de datos de entrenamiento, elija una de las siguientes opciones:
 - Seleccione Archivo si tiene un conjunto de datos más pequeño que pueda caber en el volumen de almacenamiento de aprendizaje automático y si prefiere el acceso al sistema de archivos tradicional para su guion de entrenamiento.

- Selecciona Pipe para obtener conjuntos de datos de gran tamaño para transmitir datos directamente desde S3 y así evitar tener que descargar todo al disco, lo que puede mejorar la velocidad de entrenamiento y reducir los requisitos de almacenamiento.
- Seleccione FastFile esta opción si desea combinar las ventajas del streaming desde S3 con el acceso al sistema de archivos, especialmente para leer datos de forma secuencial o si utiliza menos archivos para acelerar los tiempos de inicio.

8. Para obtener los detalles del canal de entrada ML, haga lo siguiente:

- a. Para el canal de entrada ML, especifique el canal de entrada ML que proporciona datos al algoritmo del modelo.

Para añadir otro canal, elija Añadir otro canal de entrada ML. Puede añadir hasta 19 canales de entrada ML adicionales.

- b. En Nombre del canal, introduzca el nombre del canal de entrada ML.
- c. Para el tipo de distribución de datos de Amazon S3, elija una de las siguientes opciones:
- Seleccione Totalmente replicado para proporcionar a cada instancia de entrenamiento una copia completa de su conjunto de datos. Esto funciona mejor cuando el conjunto de datos es lo suficientemente pequeño como para caber en la memoria o cuando cada instancia necesita acceder a todos los datos.
 - Selecciona Fragmentado por la clave S3 para dividir el conjunto de datos entre las instancias de entrenamiento en función de las claves S3. Cada instancia recibe aproximadamente $1/n$ del total de objetos de S3, donde «n» es el número de instancias. Esto funciona mejor para conjuntos de datos grandes que desee procesar en paralelo.

 Note

Tenga en cuenta el tamaño del conjunto de datos y los requisitos de formación al seleccionar un tipo de distribución. La replicación completa proporciona un acceso completo a los datos, pero requiere más almacenamiento, mientras que la clave Sharded by S3 permite el procesamiento distribuido de grandes conjuntos de datos.

9. En Duración máxima del entrenamiento, elige la cantidad máxima de tiempo que deseas entrenar tu modelo.

10. En el caso de los hiperparámetros, especifique los parámetros específicos del algoritmo y sus valores previstos. Los hiperparámetros son específicos del modelo que se está entrenando y se utilizan para ajustar el entrenamiento del modelo.
11. En el caso de las variables de entorno, especifique cualquier variable específica del algoritmo y sus valores previstos. Las variables de entorno se configuran en el contenedor de Docker.
12. En el caso del cifrado, para usar una personalizada AWS KMS key, selecciona la casilla de verificación Cifrar el secreto con una clave KMS personalizada.
13. Para la configuración de recursos de EC2, especifique la información sobre los recursos informáticos que se utilizan para el entrenamiento del modelo.
 - a. En Tipo de instancia, elija el tipo de instancia que desee ejecutar.

Los tipos de instancias compatibles para la formación distribuida son:

- ml.m5.4xlarge
- ml.m5.12xlarge
- ml.m5.2xlarge
- ml.g5.12xlarge
- ml.g5.24xlarge

- b. En Recuento de instancias, introduce el número de instancias.
- c. Para el tamaño del volumen en GB, introduzca el tamaño del volumen de almacenamiento de ML.

14. Elija Crear modelo entrenado.

API

Para ejecutar un trabajo de formación distribuido (API)

Ejecute el siguiente código con sus parámetros específicos:

```
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

Una vez creado el modelo entrenado, no podrá editarlo. Para realizar cambios, elimine el modelo entrenado y cree uno nuevo.

Exportación de artefactos modelo desde AWS Clean Rooms ML

Esta tarea es opcional y debe completarse cuando haya asignado la capacidad de CAN_RECEIVE_MODEL_OUTPUT miembro a un miembro de la colaboración.

Una vez finalizada la formación del modelo, el miembro que lo entrenó puede iniciar la exportación de los artefactos del modelo. El miembro que entrenó el modelo elige quién recibirá los artefactos del modelo, siempre que ese miembro pueda recibir los resultados y una configuración de aprendizaje automático válida.

Console

Para configurar un algoritmo de modelo de aprendizaje automático personalizado (consola)

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página Colaboraciones, elija la colaboración que contenga el modelo personalizado que desee exportar.
4. Cuando se abra la colaboración, seleccione la pestaña Modelos de aprendizaje automático y, a continuación, elija su modelo de la tabla de modelos entrenados personalizados
5. En la página de detalles del modelo entrenado personalizado, haga clic en Exportar la salida del modelo.
6. En Exportar la salida del modelo, en Exportar detalles de salida del modelo, introduzca un nombre y, si lo desea, una descripción.

Elija qué miembro recibirá los artefactos del modelo en la lista desplegable del modelo exportado a los miembros de la colaboración.

7. Seleccione Exportar.

Los resultados se exportan a la siguiente ruta en la ubicación de Amazon S3 que se especificó en la configuración de `ML:yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`. Solo se exportan los archivos que se van a exportar, hasta el tamaño máximo de archivo especificado, que haya seleccionado al asociar el algoritmo del modelo configurado.

API

Para configurar un algoritmo de modelo ML (API) personalizado

Inicie la exportación del modelo ejecutando el siguiente código:

```
import boto3
acr_ml_client= boto3.client('cleanroomsm')

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'
)
```

Los resultados se exportan a la siguiente ruta en la ubicación de Amazon S3 que se especificó en la configuración de ML:`yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`. Solo se exporta el algoritmo del modelo configurado, hasta el `maxSize` especificado, que haya seleccionado al asociar. `filesToExport`

Ejecución de inferencias en un modelo entrenado en AWS Clean Rooms ML

Los miembros con la capacidad de ejecutar consultas también pueden iniciar el trabajo de inferencia una vez finalizado el trabajo de formación. Escogen el conjunto de datos de inferencias con el que quieren ejecutar la inferencia y hacen referencia a los resultados del modelo entrenado con los que desean ejecutar el contenedor de inferencias.

Al miembro que recibirá el resultado de la inferencia se le debe conceder la habilidad de miembro. `CAN_RECEIVE_INFERENCE_OUTPUT`

Console

Para crear un trabajo de inferencia de modelos (consola)

1. Inicie sesión Consola de administración de AWS y abra la AWS Clean Rooms consola en <https://console.aws.amazon.com/cleanrooms>.
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la página Colaboraciones, elija la colaboración que contenga el modelo personalizado sobre el que desee crear un trabajo de inferencia.
4. Cuando se abra la colaboración, seleccione la pestaña Modelos de aprendizaje automático y, a continuación, elija su modelo de la tabla de modelos entrenados personalizados.

5. En la página de detalles del modelo entrenado personalizado, haga clic en Iniciar trabajo de inferencia.
6. En Iniciar trabajo de inferencia, en Detalles del trabajo de inferencia, introduzca un nombre y, si lo desea, una descripción.

Introduzca la información siguiente:

- Algoritmo de modelo asociado: el algoritmo de modelo asociado que se utiliza durante el trabajo de inferencia.
 - Detalles del canal de entrada ML: el canal de entrada ML que proporcionará los datos para este trabajo de inferencia.
 - Recursos de transformación: la instancia de procesamiento que se utiliza para realizar la función de transformación del trabajo de inferencia.
 - Configuración de salida: quién recibirá el resultado del trabajo de inferencia y el tipo MIME del resultado.
 - Cifrado: elija Personalizar la configuración de cifrado para especificar su propia clave de KMS y la información relacionada. De lo contrario, Clean Rooms ML gestionará el cifrado.
 - Transformar los detalles del trabajo: la carga útil máxima del trabajo de inferencia, en MB.
 - Variables de entorno: cualquier variable de entorno necesaria para acceder a la imagen del contenedor de trabajos de inferencia.
7. Elija Iniciar trabajo de inferencia.

Los resultados se exportan a la siguiente ruta en la ubicación de Amazon S3 que se especificó en la configuración de ML: `yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName`.

API

Para crear un trabajo de inferencia de modelos (API)

Inicie el trabajo de inferencia ejecutando el siguiente código:

```
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'
    }
  ]
}
)
```

Los resultados se exportan a la siguiente ruta en la ubicación de Amazon S3 que se especificó en la configuración de ML:yourSpecifiedS3Path/collaborationIdentifier/trainedModelName/callerAccountId/jobName.

Cree modelos de aprendizaje automático de AWS Clean Rooms como proveedor de datos de formación

Un modelo similar es un modelo de los datos de un proveedor de datos de entrenamiento que permite a un proveedor de datos iniciales crear un segmento similar de los datos del proveedor de datos de entrenamiento que se parezca más a sus datos iniciales. Para crear un modelo similar que se pueda utilizar en una colaboración, debe importar los datos de entrenamiento, crear un modelo similar, configurar ese modelo similar y, después, asociarlo a una colaboración.

Trabajar con modelos similares requiere que dos partes, un proveedor de datos de formación y un proveedor de datos iniciales, trabajen secuencialmente AWS Clean Rooms para integrar sus datos en una colaboración. Este es el flujo de trabajo que el proveedor de datos de entrenamiento debe completar primero:

1. Los datos del proveedor de datos de formación deben almacenarse en una tabla de catálogo de AWS Glue datos con las interacciones entre usuarios y elementos. Como mínimo, los datos de entrenamiento deben contener una columna de ID de usuario, una columna de ID de interacción y una columna de marca temporal.
2. El proveedor de datos de entrenamiento registra los datos de entrenamiento con AWS Clean Rooms.
3. El proveedor de datos de entrenamiento crea un modelo similar que se puede compartir con varios proveedores de datos iniciales. El modelo similar es una red neuronal profunda que puede tardar hasta 24 horas en entrenarse. No se reentrena automáticamente y le recomendamos que vuelva a entrenar el modelo una vez por semana.
4. El proveedor de datos de entrenamiento configura el modelo similar, que incluye si desea compartir las métricas de relevancia y la ubicación en Amazon S3 de los segmentos de salida. El proveedor de datos de entrenamiento puede crear varios modelos similares configurados a partir de un único modelo similar.
5. El proveedor de datos de entrenamiento asocia el modelo de audiencia configurado a una colaboración que se comparte con un proveedor de datos iniciales.

Una vez que el proveedor de datos de entrenamiento haya terminado de crear el modelo de aprendizaje automático, el [proveedor de datos iniciales puede crear y exportar el segmento similar](#).

Temas

- [Importación de datos de entrenamiento](#)
- [Crear un modelo similar](#)
- [Configuración de un modelo similar](#)
- [Asociar un modelo similar configurado](#)
- [Actualización de un modelo similar configurado](#)

Importación de datos de entrenamiento

Note

Solo puede proporcionar un conjunto de datos de entrenamiento para usarlo en un modelo similar a Clean Rooms ML que tenga datos almacenados en Amazon S3. Sin embargo, puede proporcionar los datos iniciales para un modelo similar mediante SQL que agrupe los datos almacenados en cualquier fuente de datos compatible.

Antes de crear un modelo similar, debe especificar la AWS Glue tabla que contiene los datos de entrenamiento. Clean Rooms ML no almacena una copia de estos datos, solo los metadatos que le permiten acceder a los datos.

Para importar los datos de entrenamiento en AWS Clean Rooms

1. Inicia sesión Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, selecciona los modelos AWS ML.
3. En la pestaña Conjuntos de datos de entrenamiento, seleccione Crear conjunto de datos de entrenamiento.
4. En la página Crear conjunto de datos de entrenamiento, en los detalles del conjunto de datos de entrenamiento, ingrese un nombre y, de forma opcional, una descripción.
5. Elija el origen de datos de entrenamiento seleccionando la base de datos y la tabla que desee configurar en las listas desplegables.

Note

Para comprobar que se trata de la tabla correcta, realice una de las siguientes acciones:

- Seleccione Ver en AWS Glue.
- Active Ver esquema para ver el esquema.

6. Para Detalles de entrenamiento, elija la Columna de identificador de usuario, la Columna de identificador de elemento y la Columna de marca temporal de las listas desplegables. Los datos de entrenamiento deben contener estas tres columnas. También puede seleccionar cualquier otra columna que quiera incluir en los datos de entrenamiento.

Los datos de la Columna de marca temporal deben estar en formato de tiempo Unix en formato de segundos.

7. (Opcional) Si tiene columnas adicionales que entrenar, elija el nombre y el tipo de la columna de las listas desplegables.
8. En Acceso al servicio, debe especificar un rol de servicio que pueda acceder a los datos y proporcionar una clave de KMS si los datos están cifrados. Elija Crear y usar un nuevo rol de servicio y Clean Rooms ML creará automáticamente un rol de servicio y agregará la política de permisos necesaria. Elija Usar un rol de servicio existente e ingréselo en el campo Nombre del rol de servicio si tiene un rol de servicio específico que quiera usar.

Si los datos están cifrados, ingrese la clave de KMS en el campo AWS KMS key o haga clic en Crear una AWS KMS key para generar una nueva clave de KMS.

9. Si desea habilitar la opción de Etiquetas para el conjunto de datos de entrenamiento, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
10. Elija Crear el conjunto de datos de entrenamiento.

Para ver la acción de API correspondiente, consulte [CreateTrainingDataset](#).

Crear un modelo similar

Una vez que haya creado un conjunto de datos de entrenamiento, estará listo para crear un modelo similar. Puede crear muchos modelos similares a partir de un único conjunto de datos de entrenamiento.

Debe crear una base de datos predeterminada en su función AWS Glue Data Catalog o incluir el `glue:createDatabase` permiso en la función proporcionada.

Para crear un modelo similar en AWS Clean Rooms

1. Inicia sesión en la consola Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, selecciona los modelos AWS ML.
3. En la pestaña Modelos similares, seleccione Crear un modelo similar.
4. En la página Crear modelo similar, para Detalles de modelo similar, ingrese un Nombre y una Descripción opcional.
 - a. Seleccione el Conjunto de datos de entrenamiento que desea modelar en la lista desplegable.

Note

Para comprobar que se trata del conjunto de datos de entrenamiento correcto, active Mostrar detalles del conjunto de datos de entrenamiento para ver los detalles. Para crear un nuevo conjunto de datos de entrenamiento, elija Crear conjunto de datos de entrenamiento.

- b. (Opcional) Ingrese una ventana de entrenamiento.
5. Si desea activar la configuración de cifrado personalizada para el modelo similar, elija Personalizar la configuración de cifrado y, a continuación, introduzca la clave de KMS.
 6. Si desea habilitar la opción de Etiquetas para el modelo similar, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
 7. Seleccione Crear un modelo similar.

Note

El entrenamiento del modelo puede tardar entre varias horas y 2 días.

Para ver la acción de API correspondiente, consulte [CreateAudienceModel](#).

Configuración de un modelo similar

Una vez que haya creado un modelo similar, estará listo para configurarlo para su uso en una colaboración. Puede crear varios modelos similares configurados a partir de un único modelo similar.

Para configurar un modelo similar en AWS Clean Rooms

1. Inicia sesión en la consola Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, selecciona los modelos AWS ML.
3. En la pestaña Modelos similares configurados, elija Configurar un modelo similar.
4. En la página Configurar modelo similar, para Detalles de modelo similar configurado, ingrese un Nombre y una Descripción opcional.

- a. Seleccione el Modelo similar que desea configurar en la lista desplegable.

Note

Para comprobar que se trata del modelo similar correcto, active Mostrar detalles del modelo similar para ver los detalles.

Para crear un nuevo modelo parecido, elija Crear modelo parecido.

- b. Elija el Tamaño mínimo de coincidencia de la lista inicial que desee. Es el número mínimo de usuarios de los datos del proveedor de datos iniciales que se superponen con los usuarios de los datos de entrenamiento. Este valor debe ser mayor que 0.
5. En Métricas para compartir con otros miembros, elija si desea que el proveedor de datos iniciales de su colaboración reciba métricas del modelo, incluidas las puntuaciones de relevancia.
 6. Para la ubicación de destino del segmento similar, introduzca el depósito de Amazon S3 al que se exporta el segmento similar. Este bucket debe estar en la misma región que los demás recursos.
 7. En Acceso a los servicios, elija el Nombre del rol de servicio existente que se utilizará para acceder a esta tabla.
 8. Para la configuración avanzada del tamaño de los contenedores, especifique el tipo de tamaño de la audiencia como un número absoluto o un porcentaje.
 9. Si desea habilitar la opción de Etiquetas para el recurso de tabla configurada, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
 10. Elija Configurar un modelo similar.

Para ver la acción de API correspondiente, consulte [CreateConfiguredAudienceModel](#).

Asociar un modelo similar configurado

Después de haber configurado un modelo similar, puede asociarlo a una colaboración.

Para asociar un modelo similar configurado en AWS Clean Rooms

1. Inicia sesión en Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la pestaña Con pertenencia activa, elija una colaboración.
4. En la pestaña Modelos ML, en Modelos eady-to-use parecidos a R, selecciona Asociar modelo parecido.
5. En la página Asociar modelo parecido configurado, para ver los detalles de la asociación de modelos similares configurados:
 - a. Introduzca un Nombre para el modelo de audiencia configurado asociado.
 - b. Introduzca una Descripción de la tabla.

La descripción ayuda a diferenciarlo de otros modelos de audiencia configurados asociados con nombres similares.

6. En Modelo similar configurado, elija un modelo similar configurado de la lista desplegable.
7. Elija Asociar .

Para ver la acción de API correspondiente, consulte. [CreateConfiguredAudienceModelAssociation](#)

Actualización de un modelo similar configurado

Una vez que haya asociado un modelo similar configurado, puede actualizarlo para cambiar información como el nombre, las métricas que desea compartir o la ubicación de salida de Amazon S3.

Para actualizar un modelo similar configurado asociado en AWS Clean Rooms

1. Inicie sesión en Consola de administración de AWS y abra la [AWS Clean Rooms consola](#) con su Cuenta de AWS (si aún no lo ha hecho).
2. En el panel de navegación izquierdo, selecciona los modelos AWS ML.

3. En la pestaña Modelos similares configurados, en Ready-to-use Modelos similares, elija un modelo similar configurado y seleccione Editar.
4. En la página de edición, para ver los detalles de la asociación de modelos similares configurados:
 - a. Actualice un Nombre y una Descripción opcional.
 - b. Elija el Modelo similar que desea configurado de la lista desplegable.
 - c. Elija el Tamaño mínimo de coincidencia de la lista inicial que desee. Es el número mínimo de usuarios de los datos del proveedor de datos iniciales que se superponen con los usuarios de los datos de entrenamiento. Este valor debe ser mayor que 0.
5. En Métricas para compartir con otros miembros, elija si desea que el proveedor de datos iniciales de su colaboración reciba métricas del modelo, incluidas las puntuaciones de relevancia.
6. En Ubicación de destino del segmento similar, introduzca el bucket de Amazon S3 al que se exporta el segmento similar. Este bucket debe estar en la misma región que los demás recursos.
7. En Acceso a los servicios, elija el Nombre del rol de servicio existente que se utilizará para acceder a esta tabla.
8. Para la configuración avanzada del tamaño de los contenedores, elija cómo desea configurar los tamaños de los contenedores de audiencia.
9. Seleccione Save changes (Guardar cambios).

Para ver la acción de API correspondiente, consulte. [UpdateConfiguredAudienceModel](#)

Creación de modelos de aprendizaje automático de AWS Clean Rooms como proveedor de datos iniciales

Una vez que el proveedor de datos de entrenamiento haya terminado de crear el modelo de aprendizaje automático, el proveedor de datos iniciales podrá crear y exportar el segmento similar. El segmento similar es un subconjunto de los datos de entrenamiento que más se parece a los datos iniciales.

Este es el flujo de trabajo que debe completar el proveedor de datos iniciales:

1. Los datos del proveedor de datos iniciales se pueden almacenar en un bucket de Amazon S3 o pueden provenir de los resultados de una consulta.
2. El proveedor de datos iniciales comienza la colaboración que comparte con el proveedor de datos de entrenamiento.
3. El proveedor de datos iniciales crea un segmento similar en la pestaña de Clean Rooms ML de la página de colaboración.
4. El proveedor de datos iniciales puede evaluar las métricas de relevancia, en caso de que se hayan compartido, y exportar el segmento similar para su uso fuera de AWS Clean Rooms.

Temas

- [Crear un segmento similar](#)
- [Exportación de un segmento similar](#)

Crear un segmento similar

Note

Solo puede proporcionar un conjunto de datos de entrenamiento para usarlo en un modelo similar a Clean Rooms ML que tenga datos almacenados en Amazon S3. Sin embargo, puede proporcionar los datos iniciales para un modelo similar mediante SQL que agrupe los datos almacenados en cualquier fuente de datos compatible.

Un segmento similar es un subconjunto de los datos de entrenamiento que se parece más a los datos iniciales.

Para crear un segmento similar en AWS Clean Rooms

1. Inicia sesión en Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la pestaña Con pertenencia activa, elija una colaboración.
4. En la pestaña Modelos ML, selecciona Crear segmento similar.
5. En la página Crear segmento similar, en Modelo similar configurado asociado, elija el modelo similar configurado asociado para usarlo en este segmento similar.
6. Para Detalles de segmento similar ingrese un Nombre y una Descripción opcional.
7. Para Perfiles iniciales, elija el Método inicial mediante la selección de una opción y, a continuación, realizando la acción recomendada.

Opción	Acción recomendada
Ruta de Amazon S3	1. Seleccione una ubicación de Amazon S3. 2. (Opcional) Elija Incluir perfiles iniciales en la salida.
Consulta SQL	Escriba una consulta SQL y utilice sus resultados como datos iniciales.
Plantilla de análisis	Elija una plantilla de análisis de la lista desplegable y utilice los resultados creados por una plantilla de análisis.

8. Elija el tipo de trabajador que se utilizará al crear esta fuente de datos. El tipo de trabajador predeterminado es CR.1X. Especifique el número de trabajadores que se van a utilizar. El número de trabajador predeterminado es 16. Para especificar las propiedades de Spark:
 - a. Amplíe Propiedades de Spark.
 - b. Selecciona Añadir propiedades de Spark.

- c. En el cuadro de diálogo de propiedades de Spark, elija un nombre de propiedad en la lista desplegable e introduzca un valor.

En la siguiente tabla se proporciona una definición para cada propiedad.

Para obtener más información sobre las propiedades de Spark, consulte [Propiedades de Spark](#) en la documentación de Apache Spark.

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Fallos en Spark.task.max	Controla cuántas veces consecutivas puede fallar una tarea antes de que el trabajo falle. Requiere un valor mayor o igual a 1. El número de reintentos permitidos es igual a este valor menos 1. El recuento de errores se restablece si algún intento tiene éxito. Los errores en las distintas tareas no se acumulan hasta alcanzar este límite.	4
spark.sql.files.maxPartitionBytes	Establece el número máximo de bytes que se deben empaquetar en una sola partición al leer fuentes basadas en archivos, como Parquet, JSON y ORC.	128 MB
Spark.Hadoop.FS.S3.max se reintenta	Establece el número máximo de reintentos para las operaciones con archivos de Amazon S3.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.network.timeout	Establece el tiempo de espera predeterminado para todas las interacciones de la red. Anula los siguientes ajustes de tiempo de espera si no están configurados: • spark.storage.blockManagerHeartbeatTimeoutMs • Tiempo de espera de conexión de Spark.Shuffle.io • Spark.rpc.askTimeout • Se ha agotado el tiempo de espera de Spark.rpc Lookup.	120 s
spark.rdd.compress	Especifica si se deben comprimir las particiones RDD serializadas mediante spark.io.compression.codec. Se aplica a StorageLevel .MEMORY_ONLY_SER en Java y Scala, o a .MEMORY_ONLY en Python. StorageLevel Reduce el espacio de almacenamiento, pero requiere más tiempo de procesamiento de la CPU.	FALSO
spark.shuffle.spill.compress	Especifica si se deben comprimir los datos de Shuffle Spill mediante spark.io.compression.codec.	TRUE
spark.sql.adaptive.advisoryPartitionSizeInBytes	Establece el tamaño objetivo en bytes para las particiones aleatorias durante la optimización adaptativa cuando spark.sql.adaptive.enabled es true. Controla el tamaño de las particiones al unir particiones pequeñas o dividir particiones asimétricas.	(valor de spark.sql.adaptive.shuffle.targetPostShuffleInputSize)

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive.autoBroadcastJoinUmbra	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Se aplica solo en un marco adaptativo. Utiliza el mismo valor predeterminado que spark.sql.adaptive.autoBroadcastJoinUmbra. Configúrelo en -1 para deshabilitar la transmisión.	(ninguno)
spark.sql.adaptive.coalescePartitions.enabled	Especifica si se van a unir particiones aleatorias contiguas en función de spark.sql.adaptive.advisoryPartitionSizeInBytes para optimizar el tamaño de la tarea. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE
Particiones spark.sql.Adaptive.Coalesce.initialPartitionNum	Define el número inicial de particiones de mezcla antes de la fusión. Requiere que spark.sql.adaptive.enabled y spark.sql.Adaptive.coalescePartitions.enabled sean verdaderos. El valor predeterminado es spark.sql.shuffle.partition.	(ninguno)
Spark.sql.Adaptive.CoalescePartitions.minPartitionSize	Establece el tamaño mínimo de las particiones aleatorias combinadas para evitar que las particiones se vuelvan demasiado pequeñas durante la optimización adaptativa.	1 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive.coalescePartitions.ParalelismFirst	Especifica si se deben calcular los tamaños de las particiones en función del paralelismo del clúster en lugar de en spark.sql.adaptive.advisoryPartitionSizeInBytes durante la fusión de particiones. Genera tamaños de partición más pequeños que el tamaño objetivo configurado para maximizar el paralelismo. Se recomienda establecer este valor en false en los clústeres ocupados para mejorar la utilización de los recursos y evitar el exceso de tareas pequeñas.	TRUE
spark.sql.adaptive.enabled	Especifica si se debe habilitar la ejecución adaptativa de consultas para volver a optimizar los planes de consultas durante la ejecución de las consultas, en función de estadísticas de tiempo de ejecución precisas.	TRUE
spark.sql.adaptive.forceOptimizeSkewedJoinÚnete!	Especifica si se debe forzar la activación OptimizeSkewedJoin incluso si se introduce una mezcla adicional.	FALSO
spark.sql.adaptive.localShuffleReader.habilitado	Especifica si se deben utilizar lectores aleatorios locales cuando no se requiera la creación de particiones aleatorias, por ejemplo, después de la conversión de uniones por ordenación y combinación a uniones por hash de difusión. Requiere que spark.sql.adaptive.enabled sea verdadero.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
<code>spark.sql.adaptive.maxShuffledHashJoinLocalMapThreshold</code>	Establece el tamaño máximo de partición en bytes para crear mapas hash locales. Prioriza las uniones hash mezcladas sobre las uniones por ordenación y combinación cuando: • Este valor es igual o superior a <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> • Todos los tamaños de partición están dentro de este límite Anula <code>spark.sql.join.preferSortMergeUnirse</code> a la configuración.	0 bytes
<code>spark.sql.adaptive.optimizeSkewsInRebalancePartitions.habilitado</code>	Especifica si se deben optimizar las particiones aleatorias asimétricas dividiéndolas en particiones más pequeñas según <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>. Requiere que <code>spark.sql.adaptive.enabled</code> sea verdadero.	TRUE
<code>spark.sql.adaptive.rebalancePartitionsSmallPartitionFactor</code>	Define el factor de umbral de tamaño para la fusión de particiones durante la división. Las particiones más pequeñas que este factor se multiplican por <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> están fusionadas.	0.2
<code>spark.sql.adaptive.skewJoin.enabled</code>	Especifica si se debe tratar la asimetría de datos en las uniones mezcladas dividiendo y, opcionalmente, replicando las particiones asimétricas. Se aplica a las uniones hash ordenadas, combinadas y barajadas. Requiere que <code>spark.sql.adaptive.enabled</code> sea verdadero.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.adaptive.skewJoin.skewedPartitionFactor	Determina el factor de tamaño que determina el sesgo de la partición. Una partición está sesgada cuando su tamaño supera ambos: • Este factor se multiplica por la mediana del tamaño de la partición • El valor de spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	5
spark.sql.adaptive.skewJoin.skewedPartitionThresholdInBytes	Establece el umbral de tamaño en bytes para identificar las particiones asimétricas. Una partición está sesgada cuando su tamaño supera ambos: • Este umbral • El tamaño medio de la partición multiplicado por spark.sql.adaptive.skewJoin.skewedPartitionFactor Recomendamos establecer este valor por encima de spark.sql.adaptive.advisoryPartitionSizeInBytes.	256 MB
spark.sql.autoBroadcastJoinUmbral	Establece el tamaño máximo de la tabla en bytes para la transmisión a los nodos trabajadores durante las uniones. Configúrelo en -1 para deshabilitar la transmisión.	10 MB
Spark.sql.Tiempo de espera de emisión	Controla el tiempo de espera en segundos para las operaciones de transmisión durante las uniones de transmisión.	300 segundos

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.cbo.enabled	Especifica si se debe habilitar la optimización basada en costes (CBO) para la estimación de las estadísticas del plan.	FALSO
spark.sql.cbo.joinReorder.dp.star.filter	Especifica si se debe aplicar la heurística del filtro de unión en estrella durante la enumeración de uniones basada en el coste.	FALSO
spark.sql.cbo.joinReorder.dp.Threshold	Establece el número máximo de nodos unidos permitidos en el algoritmo de programación dinámica.	12
spark.sql.cbo.joinReorder.Enabled	Especifica si se debe habilitar el reordenamiento de las uniones en la optimización basada en costes (CBO).	FALSO
Spark.sql.cbo.PlanStats.Enabled	Especifica si se deben obtener los recuentos de filas y las estadísticas de las columnas del catálogo durante la generación del plan lógico.	FALSO
spark.sql.cbo.starSchemaDetection	Especifica si se debe habilitar el reordenamiento de las uniones en función de la detección del esquema en estrella.	FALSO
spark.sql.files.maxPartitionNum	Establece el número máximo objetivo de particiones de archivos divididas para fuentes basadas en archivos (Parquet, JSON y ORC). Cambia la escala de las particiones cuando el recuento inicial supera este valor. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.files.maxRecordsPerFile	Establece el número máximo de registros que se van a escribir en un único archivo. No se aplica ningún límite cuando se establece en cero o en un valor negativo.	0
spark.sql.files.minPartitionNum	Establece el número mínimo objetivo de particiones de archivos divididas para las fuentes basadas en archivos (Parquet, JSON y ORC). El valor predeterminado es spark.sql.files.defaultParalelism. Se trata de un objetivo sugerido, no de un límite garantizado.	(ninguno)
spark.sql.inMemoryColumnarStorageSizeInBytes	Controla el tamaño del lote para el almacenamiento en caché en columnas. Al aumentar el tamaño, se mejora la utilización y la compresión de la memoria, pero aumenta el riesgo de out-of-memory errores.	10000
spark.sql.inMemoryColumnarStorageCompressed	Especifica si se seleccionan automáticamente los códecs de compresión para las columnas en función de las estadísticas de datos.	TRUE
spark.sql.inMemoryColumnarStorage.enableVectorizedReader	Especifica si se debe habilitar la lectura vectorizada para el almacenamiento en caché de columnas.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.allowHashOnMapType	Especifica si se permiten operaciones de hash en estructuras de datos de tipo mapa. Esta configuración antigua mantiene la compatibilidad con el manejo de tipos de mapas de las versiones anteriores de Spark.	
spark.sql.legacy.allowNegativeScaleOfDecimal	Especifica si se permiten valores de escala negativos en las definiciones de tipo decimal. Esta configuración antigua mantiene la compatibilidad con las versiones anteriores de Spark que admitían escalas decimales negativas.	
spark.sql.legacy.castComplexTypesToString.habilitado	Especifica si se debe habilitar el comportamiento heredado para convertir tipos complejos en cadenas. Mantiene la compatibilidad con las reglas de conversión de tipos de las versiones anteriores de Spark.	
spark.sql.legacy.charVarcharAsCadena	Especifica si se deben tratar los tipos CHAR y VARCHAR como tipos STRING. Esta configuración antigua proporciona compatibilidad con el manejo de tipos de cadenas de las versiones anteriores de Spark.	
spark.sql.legacy.createEmptyCollectionUsingStringType	Especifica si se van a crear colecciones vacías mediante elementos de tipo cadena. Esta configuración antigua mantiene la compatibilidad con el comportamiento de inicialización de colecciones de versiones anteriores de Spark.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.legacy.exponentLiteralAsDecimal. Habilitado	Especifica si se deben interpretar los literales exponenciales como tipos decimales. Esta configuración antigua mantiene la compatibilidad con el manejo literal numérico de las versiones anteriores de Spark.	
spark.sql.legacy.json.allowEmptyString.habilitado	Especifica si se permiten cadenas vacías en el procesamiento de JSON. Esta configuración antigua mantiene la compatibilidad con el comportamiento de análisis de JSON de las versiones anteriores de Spark.	
spark.sql.legacy.parquet.int96RebaseModelRead	Especifica si se debe utilizar el modo de rebase de marcas de tiempo heredado al leer los archivos de Parquet. INT96 Esta configuración antigua mantiene la compatibilidad con el manejo de las marcas de tiempo de las versiones anteriores de Spark.	
spark.sql.legacy.timeParserPolicy	Controla el comportamiento del análisis temporal para garantizar la compatibilidad con versiones anteriores. Esta configuración antigua determina cómo se analizan las marcas de tiempo y las fechas a partir de las cadenas.	
spark.sql.legacy.TypeCoercion.datetimeToString.habilitado	Especifica si se debe habilitar el comportamiento de coerción de tipo heredado al convertir valores de fecha y hora en cadenas. Mantiene la compatibilidad con las reglas de conversión de fecha y hora de las versiones anteriores de Spark.	

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql. maxSingle PartitionBytes	Establece el tamaño máximo de la partición en bytes. El planificador introduce operaciones de mezcla para particiones más grandes a fin de mejorar el paralelismo.	128 m
Caché de metadatos Spark.sql. TTLSeconds	Controla el time-to-live (TTL) de las cachés de metadatos. Se aplica a los metadatos de los archivos de partición y a las cachés del catálogo de sesiones. Requiere: • Un valor positivo superior a cero • Spark.sql.CatalogImplementation está configurado como hive • spark.sql.hive. filesourcePartitionFileCacheSize mayor que cero • spark.sql.hive. manageFilesourcePartitions establecido en verdadero	-1000 ms
spark.sql.optimizer. collapseProjectAlwaysEn línea	Especifica si se deben contraer las proyecciones adyacentes y las expresiones en línea, incluso cuando esto provoque duplicación.	FALSO
spark.sql.optimizer. dynamicPartitionPruning.habilitado	Especifica si se van a generar predicados para las columnas de partición utilizadas como claves de unión.	TRUE
spark.sql.optimizer. enableCsvExpressionOptimización	Especifica si se deben optimizar las expresiones CSV en el optimizador de SQL eliminando las columnas innecesarias de las operaciones from_csv.	TRUE

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimizer.enableJsonExpressionOptimización	Especifica si se deben optimizar las expresiones JSON en el optimizador de SQL de la siguiente manera: • Eliminar columnas innecesarias de las operaciones de from_json • Simplificar las combinaciones from_json y to_json • Optimización de las operaciones de named_struct	TRUE
Spark.sql.Optimizer.ExcludedRules	Define las reglas del optimizador que se van a deshabilitar, identificadas mediante nombres de reglas separados por comas. Algunas reglas no se pueden deshabilitar porque son necesarias para que sean correctas. El optimizador registra qué reglas se han desactivado correctamente.	(ninguno)
spark.sql.Optimizer.Runtime.BloomFilter.applicationSideScanSizeThreshold	Establece el tamaño mínimo de escaneo agregado en bytes necesario para inyectar un filtro Bloom en el lado de la aplicación.	10 GB
Filtro Spark.Sql.Optimizer.Runtime.BloomFilter.creationSideThreshold	Define el umbral de tamaño máximo para inyectar un filtro Bloom en el lado de la creación.	10 MB

Nombre de propiedad	Description (Descripción)	Valor predeterminado
Spark.Sql.Optimizer.Runtime.BloomFilter.Enabled	Especifica si se debe insertar un filtro Bloom para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	TRUE
spark.sql.optimizer.Runtime.BloomFilter.expectedNumItems	Define el número predeterminado de elementos esperados en el filtro Bloom en tiempo de ejecución.	1000000
spark.sql.Optimizer.Runtime.BloomFilter.maxNumBits	Establece el número máximo de bits permitido en el filtro Bloom en tiempo de ejecución.	67108864
Spark.Sql.Optimizer.Runtime.BloomFilter.maxNumItems	Establece el número máximo de elementos esperados permitidos en el filtro Bloom en tiempo de ejecución.	4000000
spark.sql.Optimizer.Runtime.BloomFilter.Number.Threshold	Limita el número máximo de filtros de tiempo de ejecución distintos de DPP permitidos por consulta para evitar errores en el controlador. out-of-memory	10
spark.sql.Optimizer.Runtime.BloomFilter.NumBits	Define el número predeterminado de bits que se utilizan en el filtro Bloom en tiempo de ejecución.	8388608

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.optimize r.runtime. rowlevelO perationGroupFiltro. Habilitado	Especifica si se debe habilitar el filtrado de grupos en tiempo de ejecución para las operaciones a nivel de fila. Permite que las fuentes de datos: • Elimine grupos completos de datos (como archivos o particiones) mediante filtros de fuentes de datos • Ejecute consultas en tiempo de ejecución para identificar los registros coincidentes • Deseche los grupos innecesarios para evitar costosas reescrituras Limitaciones: • No todas las expresiones se pueden convertir en filtros de fuentes de datos • Algunas expresiones requieren la evaluación de Spark (como las subconsultas)	TRUE
spark.sql.Optimize r.RuntimeFilter. semiJoinReduction. habilitado	Especifica si se debe insertar una semiunión para reducir los datos de mezcla aleatoria cuando un lado de una unión aleatoria tiene un predicado selectivo.	FALSO

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.parquet.aggregatePushdown	Especifica si se deben enviar los agregados a Parquet para su optimización. Soportes: • MIN y MAX para los tipos booleano, entero, flotante y de fecha • COUNT para todos los tipos de datos Lanza una excepción si faltan estadísticas en el pie de página de un archivo de Parquet.	FALSO
spark.sql.parquet.columnarReaderBatchTamaño	Controla el número de filas de cada lote de lectores vectorizados Parquet. Elija un valor que equilibre la sobrecarga de rendimiento y el uso de memoria para evitar out-of-memory errores.	4096
Spark.sql.Session.TimeZone	Define la zona horaria de la sesión para gestionar las marcas de tiempo en los literales de cadena y la conversión de objetos Java. Acepta: • area/city Formato basado IDs en regiones (como America/Los_Angeles) • Desplazamientos de zona en formato (+/-) HH, (+/-) HH:mm o (+/-) HH:mm:ss (como -08 o + 01:00) • UTC o Z como alias para + 00:00	(valor de la zona horaria local)
spark.sql.shuffle.partition	Establece el número predeterminado de particiones para la mezcla de datos durante las uniones o agregaciones. No se puede modificar entre reinicios de consultas de streaming estructuradas desde la misma ubicación del punto de control.	200

Nombre de propiedad	Description (Descripción)	Valor predeterminado
spark.sql.shuffledHashJoinFactor	Define el factor de multiplicación utilizado para determinar la idoneidad de las combinaciones de hash aleatorio. Se selecciona una combinación de hash aleatorio cuando el tamaño de los datos del lado pequeño multiplicado por este factor es menor que el tamaño de los datos del lado grande.	3
spark.sql.sources.parallelPartitionDiscovery.umbra	Establece el número máximo de rutas para la lista de archivos del lado del conductor con fuentes basadas en archivos (Parquet, JSON y ORC). Si se supera durante la detección de particiones, los archivos se enumeran mediante un trabajo distribuido de Spark independiente.	32
spark.sql.statistics.histogram.enabled	Especifica si se deben generar histogramas con la misma altura durante el cálculo de las estadísticas de columnas para mejorar la precisión de la estimación. Requiere un escaneo de tabla adicional al necesario para las estadísticas de columnas básicas.	FALSO

9. En Acceso a los servicios, elija el Nombre del rol de servicio existente que se utilizará para acceder a esta tabla.
10. Si desea habilitar la opción de Etiquetas para el conjunto de datos de entrenamiento, seleccione Añadir nueva etiqueta y, a continuación, introduzca el par de Clave y Valor.
11. Elija Crear segmento similar.

Para ver la acción de API correspondiente, consulte [StartAudienceGenerationJob](#).

Exportación de un segmento similar

Después de crear un segmento similar, puede exportar esos datos a un bucket de Amazon S3.

Para exportar un segmento similar a AWS Clean Rooms

1. Inicia sesión en Consola de administración de AWS y abre la [AWS Clean Rooms consola](#) con tu Cuenta de AWS (si aún no lo has hecho).
2. En el panel de navegación izquierdo, elija Colaboraciones.
3. En la pestaña Con pertenencia activa, elija una colaboración.
4. En la pestaña Modelos ML, selecciona un segmento similar y selecciona Exportar.
5. En Exportar modelo similar, para Exportar detalles de modelo similar, introduzca un Nombre y una Descripción opcional.
6. En Tamaño del segmento, elija el tamaño que desee para el segmento exportado.
7. Seleccione Exportar.

Para ver la acción de API correspondiente, consulte [StartAudienceExportJob](#).

Solución de problemas AWS Clean Rooms

En esta sección se describen algunos problemas comunes que pueden surgir al utilizarlos AWS Clean Rooms y cómo solucionarlos.

Problemas

- [Una o más tablas a las que hace referencia la consulta no están accesibles para el rol de servicio asociado. El table/role propietario debe conceder al rol de servicio acceso a la tabla.](#)
- [Uno de los conjuntos de datos subyacentes tiene un formato de archivo no compatible.](#)
- [Los resultados de la consulta no son los esperados cuando se utiliza la computación criptográfica para Clean Rooms.](#)
- [AWS Clean Rooms Spark SQL: Faltan datos de partición](#)

Una o más tablas a las que hace referencia la consulta no están accesibles para el rol de servicio asociado. El table/role propietario debe conceder al rol de servicio acceso a la tabla.

- Asegúrese de que los permisos del rol de servicio estén configurados como corresponda. Para obtener más información, consulte [Con AWS Clean Rooms figuración](#).

Uno de los conjuntos de datos subyacentes tiene un formato de archivo no compatible.

- Asegúrese de que su conjunto de datos esté en uno de los formatos de archivo compatibles:
 - Parquet
 - RCFile
 - TextFile
 - SequenceFile
 - RegexSerde
 - OpenCSV
 - AVRO

- JSON

Para obtener más información, consulte [Formatos de datos para AWS Clean Rooms](#).

Los resultados de la consulta no son los esperados cuando se utiliza la computación criptográfica para Clean Rooms.

Si utiliza la computación criptográfica para Clean Rooms (C3R), compruebe que la consulta utiliza correctamente las columnas cifradas:

- Las columnas sealed solo se utilizan en cláusulas SELECT.
- Las columnas de fingerprint solo se utilizan en cláusulas JOIN (y en cláusulas GROUP BY en ciertas condiciones).
- Solo está JOINing columnas fingerprint con el mismo nombre si los ajustes de la colaboración lo requieren.

Para obtener más información, consulte [the section called “Computación criptográfica”](#) y [the section called “Tipos de columnas”](#).

AWS Clean Rooms Spark SQL: Faltan datos de partición

Todas las particiones del también AWS Glue Data Catalog deben tener datos en S3. El motor usa la configuración Spark `spark.sql.files.ignoreMissingFiles=False`

Para obtener más información, consulte <https://spark.apache.org/docs/latest/sql-data-sources-generic-options.html#ignore-missing-files>.

Si encuentra este error, recibirá el siguiente mensaje de error: "Missing partition data: One of the configured tables is partitioned and one or more of the partitions does not have data".

Compare los datos presentes en Amazon S3 con las particiones que aparecen en la AWS Glue Data Catalog tabla. Elimine las particiones que no tengan los datos correspondientes en S3.

Seguridad en AWS Clean Rooms

La seguridad en la nube AWS es la máxima prioridad. Como AWS cliente, usted se beneficia de una arquitectura de centro de datos y red diseñada para cumplir con los requisitos de las organizaciones más sensibles a la seguridad.

La seguridad es una responsabilidad compartida entre usted AWS y usted. El [modelo de responsabilidad compartida](#) la describe como seguridad de la nube y seguridad en la nube:

- Seguridad de la nube: AWS es responsable de proteger la infraestructura que ejecuta AWS los servicios en la AWS nube. AWS también le proporciona servicios que puede utilizar de forma segura. Los auditores externos prueban y verifican periódicamente la eficacia de nuestra seguridad como parte de los [AWS programas](#) de de . Para obtener más información sobre los programas de conformidad aplicables AWS Clean Rooms, consulte [Servicios de AWS dentro del alcance por programa de conformidad](#) .
- Seguridad en la nube: su responsabilidad viene determinada por el AWS servicio que utilice. También es responsable de otros factores, incluida la confidencialidad de los datos, los requisitos de la empresa y la legislación y los reglamentos aplicables

Esta documentación le ayuda a comprender cómo aplicar el modelo de responsabilidad compartida cuando se utiliza AWS Clean Rooms. Le muestra cómo configurarlo para AWS Clean Rooms cumplir sus objetivos de seguridad y conformidad. También aprenderá a utilizar otros AWS servicios que le ayudan a supervisar y proteger sus AWS Clean Rooms recursos.

Contenido

- [Protección de datos en AWS Clean Rooms](#)
- [Uso de funciones vinculadas al servicio para AWS Clean Rooms](#)
- [Retención de datos en AWS Clean Rooms](#)
- [Mejores prácticas para la colaboración de datos en AWS Clean Rooms](#)
- [Identity and Access Management para AWS Clean Rooms](#)
- [Validación de conformidad para AWS Clean Rooms](#)
- [Resiliencia en AWS Clean Rooms](#)
- [Seguridad de la infraestructura en AWS Clean Rooms](#)
- [Acceda al aprendizaje AWS Clean Rooms automático mediante AWS Clean Rooms un punto final de interfaz \(\)AWS PrivateLink](#)

Protección de datos en AWS Clean Rooms

El modelo de [responsabilidad AWS compartida modelo](#) se aplica a la protección de datos en AWS Clean Rooms. Como se describe en este modelo, AWS es responsable de proteger la infraestructura global que ejecuta todos los Nube de AWS. Eres responsable de mantener el control sobre el contenido alojado en esta infraestructura. También eres responsable de las tareas de administración y configuración de seguridad para los Servicios de AWS que utiliza. Para obtener más información sobre la privacidad de los datos, consulte las [Preguntas frecuentes sobre la privacidad de datos](#). Para obtener información sobre la protección de datos en Europa, consulte la publicación de blog sobre el [Modelo de responsabilidad compartida de AWS y GDPR](#) en el Blog de seguridad de AWS .

Con fines de protección de datos, le recomendamos que proteja Cuenta de AWS las credenciales y configure los usuarios individuales con AWS IAM Identity Center o AWS Identity and Access Management (IAM). De esta manera, solo se otorgan a cada usuario los permisos necesarios para cumplir sus obligaciones laborales. También recomendamos proteger sus datos de la siguiente manera:

- Utiliza la autenticación multifactor (MFA) en cada cuenta.
- Se utiliza SSL/TLS para comunicarse con AWS los recursos. Exigimos TLS 1.2 y recomendamos TLS 1.3.
- Configure la API y el registro de actividad de los usuarios con AWS CloudTrail. Para obtener información sobre el uso de CloudTrail senderos para capturar AWS actividades, consulte [Cómo trabajar con CloudTrail senderos](#) en la Guía del AWS CloudTrail usuario.
- Utilice soluciones de AWS cifrado, junto con todos los controles de seguridad predeterminados Servicios de AWS.
- Utiliza servicios de seguridad administrados avanzados, como Amazon Macie, que lo ayuden a detectar y proteger la información confidencial almacenada en Amazon S3.
- Si necesita módulos criptográficos validados por FIPS 140-3 para acceder a AWS través de una interfaz de línea de comandos o una API, utilice un punto final FIPS. Para obtener más información sobre los puntos de conexión de FIPS disponibles, consulte [Estándar de procesamiento de la información federal \(FIPS\) 140-3](#).

Se recomienda encarecidamente no introducir nunca información confidencial o sensible, como por ejemplo, direcciones de correo electrónico de clientes, en etiquetas o campos de formato libre, tales como el campo Nombre. Esto incluye cuando trabaja con AWS Clean Rooms o Servicios de AWS utiliza la consola, la API o AWS CLI AWS SDKs Cualquier dato que introduzca en etiquetas

o campos de formato libre utilizados para los nombres se pueden emplear para los registros de facturación o diagnóstico. Si proporciona una URL a un servidor externo, recomendamos encarecidamente que no incluya la información de las credenciales en la URL para validar la solicitud para ese servidor.

Cifrado en reposo

AWS Clean Rooms siempre cifra todos los metadatos del servicio en reposo sin necesidad de ninguna configuración adicional. Este cifrado es automático cuando se utiliza AWS Clean Rooms.

Clean Rooms ML cifra todos los datos almacenados en el servicio en reposo. AWS KMS Si decide proporcionar su propia clave KMS, el contenido de sus modelos similares y de los trabajos de generación de segmentos similares se cifra en reposo con su clave KMS.

Cuando se utilizan modelos de aprendizaje automático AWS Clean Rooms personalizados, el servicio cifra todos los datos almacenados en reposo. AWS KMS AWS Clean Rooms admite el uso de claves simétricas administradas por el cliente que usted crea, posee y administra para cifrar los datos en reposo. Si no se especifican las claves administradas por el cliente, Claves propiedad de AWS se utilizan de forma predeterminada.

AWS Clean Rooms utiliza concesiones y políticas clave para acceder a las claves gestionadas por el cliente. Puede revocar el acceso a la concesión o eliminar el acceso del servicio a la clave administrada por el cliente en cualquier momento. Si lo hace, AWS Clean Rooms no podrá acceder a ninguno de los datos cifrados por la clave gestionada por el cliente, lo que afectará a las operaciones que dependen de esos datos. Por ejemplo, si intenta crear un modelo entrenado a partir de un canal de entrada de aprendizaje automático cifrado al que no AWS Clean Rooms se puede acceder, la operación devolverá un `ValidationException` error.

Note

Puede utilizar las opciones de cifrado de Amazon S3 para proteger sus datos en reposo. Para obtener más información, consulte [Especificación del cifrado de Amazon S3](#) en la Guía del usuario de Amazon S3.

Cuando se utiliza una tabla de mapeo de ID interna AWS Clean Rooms, el servicio cifra todos los datos almacenados en reposo. AWS KMS Si decides proporcionar tu propia clave de KMS, el contenido de la tabla de mapeo de ID se cifra en reposo con tu clave de KMS mediante AWS Entity Resolution. Para obtener más información sobre los permisos necesarios para trabajar con cifrados

con un flujo de trabajo de asignación de ID, consulte [Crear un rol de trabajo del flujo de trabajo para AWS Entity Resolution](#) en la Guía del usuario de AWS Entity Resolution .

Cifrado en tránsito

AWS Clean Rooms utiliza Transport Layer Security (TLS) para el cifrado en tránsito. La comunicación siempre AWS Clean Rooms se realiza a través de HTTPS, por lo que sus datos siempre se cifran en tránsito, independientemente de si están almacenados en Amazon S3, Amazon Athena o Snowflake. Esto incluye todos los datos en tránsito cuando se utiliza Clean Rooms ML.

Cifrado de datos subyacentes

Para obtener más información sobre cómo cifrar sus datos subyacentes, consulte [Computación criptográfica para Clean Rooms](#).

Política de claves

Las políticas de clave controlan el acceso a la clave administrada por el cliente. Cada clave administrada por el cliente debe tener exactamente una política de clave, que contiene instrucciones que determinan quién puede usar la clave y cómo puede utilizarla. Cuando crea la clave administrada por el cliente, puede especificar una política de clave. Para obtener más información, consulte Administrar el acceso a las claves administradas por el cliente en la AWS Key Management Service Guía para desarrolladores.

Para utilizar la clave gestionada por el cliente con los modelos de aprendizaje automático AWS Clean Rooms personalizados, la política de claves debe permitir las siguientes operaciones de API:

- `kms:DescribeKey`— Proporciona los detalles de la clave gestionada por el cliente AWS Clean Rooms para permitir su validación.
- `kms:Decrypt`— Proporciona acceso AWS Clean Rooms para descifrar los datos cifrados y utilizarlos en tareas relacionadas.
- `kms:CreateGrant`- Clean Rooms ML cifra las imágenes de formación e inferencia inactivas en Amazon ECR mediante la creación de subvenciones para Amazon ECR. Para obtener más información, consulte [Encryption at Rest in Amazon ECR](#). Clean Rooms ML también utiliza Amazon SageMaker AI para realizar trabajos de formación e inferencia, y crea subvenciones para que SageMaker AI cifre los volúmenes de Amazon EBS adjuntos a las instancias, así como los datos de salida en Amazon S3. Para obtener más información, consulte [Proteger los datos en reposo mediante el cifrado en Amazon SageMaker AI](#).

- kms:GenerateDataKey- Clean Rooms ML cifra los datos en reposo almacenados en Amazon S3 mediante el cifrado del lado del servidor con. AWS KMS keys Para obtener más información, consulte [Uso del cifrado del lado del servidor con AWS KMS keys \(SSE-KMS\) en Amazon S3](#).

Los siguientes son ejemplos de declaraciones de políticas que puede añadir a los AWS Clean Rooms siguientes recursos:

Canal de entrada ML con datos sintéticos

```
{
  "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 de entrada ML sin datos sintéticos

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": "*"
    }
  ]
}

```

Trabajo de modelo entrenado o trabajo de inferencia de modelo entrenado

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 no permite especificar el contexto de cifrado del servicio o el contexto de origen en las políticas clave administradas por el cliente. Los clientes pueden ver el contexto de cifrado utilizado internamente por el servicio CloudTrail.

Uso de funciones vinculadas al servicio para AWS Clean Rooms

AWS Clean Rooms [usa roles vinculados al AWS Identity and Access Management servicio \(IAM\)](#). Un rol vinculado a un servicio es un tipo único de rol de IAM al que se vincula directamente. AWS Clean

Rooms Los roles vinculados al servicio están predefinidos AWS Clean Rooms e incluyen todos los permisos que el servicio requiere para llamar a otros AWS servicios en su nombre.

Un rol vinculado a un servicio facilita la configuración AWS Clean Rooms , ya que no es necesario añadir manualmente los permisos necesarios. AWS Clean Rooms define los permisos de sus funciones vinculadas al servicio y, a menos que se defina lo contrario, solo AWS Clean Rooms puede asumir sus funciones. Los permisos definidos incluyen las políticas de confianza y de permisos, y que la política de permisos no se pueda adjuntar a ninguna otra entidad de IAM.

Solo es posible eliminar un rol vinculado a un servicio después de eliminar sus recursos relacionados. Esto protege sus AWS Clean Rooms recursos porque no puede eliminar inadvertidamente el permiso de acceso a los recursos.

Para obtener información sobre otros servicios que admiten funciones vinculadas a servicios, consulte los [AWS servicios que funcionan con IAM y busque los servicios con](#) la palabra Sí en la columna Funciones vinculadas a servicios. Elija una opción Sí con un enlace para ver la documentación acerca del rol vinculado al servicio en cuestión.

Permisos de roles vinculados al servicio para AWS Clean Rooms

AWS Clean Rooms utiliza el rol vinculado al servicio denominado `AWSServiceRoleForAWSCleanRooms` para publicar métricas relacionadas con las salas CloudWatch limpias en su cuenta. AWS

El rol vinculado al servicio de `AWSService RoleFor AWSClean Rooms` confía en los siguientes servicios para asumir el rol:

- `cleanrooms.amazonaws.com`

La política de permisos de roles denominada `AWSClean RoomsServiceRolePolicy` permite AWS Clean Rooms realizar las siguientes acciones en los recursos especificados:

- Acción: `cloudwatch:PutMetricData` en all AWS resources, restricted to the AWS Clean Rooms namespace

Debe configurar los permisos para permitir a sus usuarios, grupos o funciones, crear, editar o eliminar la descripción de un rol vinculado al servicio. Para obtener más información, consulte [Permisos de roles vinculados a servicios](#) en la Guía del usuario de IAM.

Crear un rol vinculado a un servicio para AWS Clean Rooms

Puedes usar la consola de IAM para crear un rol vinculado a un servicio con el caso de uso de Rooms. `AWSServiceRoleForAWSCleanRooms` En la API AWS CLI o en la AWS API, crea un rol vinculado a un servicio con el nombre del servicio. `cleanrooms.amazonaws.com` Para obtener más información, consulte [Crear un rol vinculado a un servicio](#) en la Guía del usuario de IAM. Si elimina este rol vinculado al servicio, puede utilizar este mismo proceso para volver a crear el rol.

Edición de un rol vinculado a un servicio para AWS Clean Rooms

AWS Clean Rooms no permite editar el rol vinculado al servicio de `AWSServiceRoleForAWSCleanRooms`. Después de crear un rol vinculado al servicio, no podrá cambiar el nombre del rol, ya que varias entidades podrían hacer referencia al rol. Sin embargo, sí puede editar la descripción del rol con IAM. Para obtener más información, consulte [Editar un rol vinculado a servicios](#) en la Guía del usuario de IAM.

Eliminar un rol vinculado a un servicio para AWS Clean Rooms

Si ya no necesita usar una característica o servicio que requieran un rol vinculado a un servicio, le recomendamos que elimine dicho rol. Así no tendrá una entidad no utilizada que no se supervise ni mantenga de forma activa. Sin embargo, debe limpiar los recursos de su rol vinculado al servicio antes de eliminarlo manualmente.

Para eliminar las `AWSServiceRoleForAWSCleanRooms`, primero debes eliminar todas las [colaboraciones](#) y [membresías](#) de tu. Cuenta de AWS

Note

Si el AWS Clean Rooms servicio utiliza el rol cuando intentas eliminar los recursos, es posible que la eliminación no se realice correctamente. En tal caso, espere unos minutos e intente de nuevo la operación.

Para eliminar manualmente el rol vinculado a servicios mediante IAM

Usa la consola de IAM AWS CLI, la o la AWS API para eliminar la función vinculada al servicio de `AWSServiceRoleForAWSCleanRooms`. Para obtener más información, consulte [Eliminación de un rol vinculado a servicios](#) en la Guía del usuario de IAM.

Regiones compatibles para los roles vinculados al servicio AWS Clean Rooms

AWS Clean Rooms admite el uso de funciones vinculadas al servicio en todas las regiones en las que el servicio está disponible. Para obtener más información, consulte [Puntos de conexión y Regiones de AWS](#).

Retención de datos en AWS Clean Rooms

Todos los datos que se lean temporalmente en una AWS Clean Rooms colaboración se eliminan una vez finalizada la consulta.

Al crear un modelo similar, Clean Rooms ML lee los datos de entrenamiento, los transforma en un formato adecuado para nuestro modelo de ML y almacena los parámetros del modelo entrenado dentro de Clean Rooms ML. Clean Rooms ML no conserva una copia de sus datos de entrenamiento. AWS Clean Rooms Las consultas SQL no retienen ninguno de sus datos una vez ejecutada la consulta. A continuación, Clean Rooms ML utiliza el modelo entrenado para resumir el comportamiento de todos los usuarios. Clean Rooms ML almacena un conjunto de datos en el nivel de usuario para cada usuario de los datos durante el tiempo que el modelo similar esté activo.

Al iniciar un trabajo de generación de segmentos similares, Clean Rooms ML lee los datos iniciales, lee los resúmenes de comportamiento del modelo similar asociado y crea un segmento similar que se almacena en el servicio. AWS Clean Rooms Clean Rooms ML no retiene ninguna copia de los datos iniciales. Clean Rooms ML almacena la salida del trabajo en el nivel de usuario mientras el trabajo esté activo.

Si los datos iniciales provienen de una consulta SQL, el resultado de esa consulta solo se almacena en el servicio mientras dure el trabajo. Los resultados de la consulta están cifrados en reposo y en tránsito.

Utilice la API para eliminar los datos del trabajo de generación de segmentos o modelos similares. Clean Rooms ML elimina de forma asíncrona todos los datos asociados al modelo o trabajo. Una vez finalizado este proceso, Clean Rooms ML elimina los metadatos del modelo o trabajo y ya no están visibles en la API. Clean Rooms ML retiene los datos eliminados durante 3 días para evitar la recuperación ante desastres. Una vez que el trabajo o el modelo ya no estén visibles en la API y hayan transcurrido 3 días, todos los datos asociados al modelo o al trabajo se eliminarán permanentemente.

Mejores prácticas para la colaboración de datos en AWS Clean Rooms

En este tema se describen las prácticas recomendadas para realizar colaboraciones de datos en AWS Clean Rooms.

AWS Clean Rooms sigue el [modelo de responsabilidad AWS compartida](#). AWS Clean Rooms ofrece [reglas de análisis](#) que puede configurar para reforzar su capacidad de proteger los datos confidenciales en una colaboración. Las reglas de análisis que configure AWS Clean Rooms impondrán las restricciones (controles de consulta y controles de salida de consultas) que haya configurado. Usted es responsable de determinar las restricciones y configurar las reglas de análisis como corresponda.

Las colaboraciones de datos pueden implicar algo más que su uso AWS Clean Rooms. Para ayudarlo a maximizar los beneficios de las colaboraciones de datos, le recomendamos que lleve a cabo las siguientes prácticas recomendadas con el uso de las reglas de análisis AWS Clean Rooms y, específicamente, con ellas.

Temas

- [Mejores prácticas con AWS Clean Rooms](#)
- [Mejores prácticas para utilizar las reglas de análisis en AWS Clean Rooms](#)

Mejores prácticas con AWS Clean Rooms

Usted es responsable de evaluar el riesgo de cada colaboración de datos y compararlo con sus requisitos de privacidad (por ejemplo, con sus programas y políticas de cumplimiento externos e internos). Le recomendamos que tome medidas adicionales al usar AWS Clean Rooms. Estas acciones pueden ayudarle a administrar mejor los riesgos y a protegerse de los intentos de reidentificación de los datos por parte de terceros (por ejemplo, ataques diferenciales o ataques de canal lateral).

Por ejemplo, considere la posibilidad de realizar comprobaciones de diligencia debida en sus otros colaboradores y de formalizar acuerdos legales con ellos antes de iniciar una colaboración. Para monitorizar el uso de sus datos, considere también la posibilidad de adoptar otros mecanismos de auditoría cuando utilice AWS Clean Rooms.

Mejores prácticas para utilizar las reglas de análisis en AWS Clean Rooms

Las reglas de análisis AWS Clean Rooms permiten restringir las consultas que se pueden ejecutar configurando los controles de consulta en una tabla configurada. Por ejemplo, puede definir un control de consulta sobre cómo se puede combinar una tabla configurada y qué columnas se pueden seleccionar. También puede restringir el resultado de la consulta configurando controles de resultados de consulta, como umbrales de agregación en las filas de salida. El servicio rechaza cualquier consulta y elimina las filas que no cumplen las reglas de análisis establecidas por los miembros en sus tablas configuradas de la consulta.

Es aconsejable que siga estas 10 prácticas recomendadas a la hora de usar reglas de análisis en su tabla configurada:

- Cree tablas configuradas distintas para distintos casos de uso de consultas (por ejemplo, planificación de audiencias o atribución). Puede crear varias tablas configuradas con la misma tabla de AWS Glue subyacente.
- Especifique las columnas de la regla de análisis (por ejemplo, columnas de dimensión, columnas de lista, columnas de combinación) que sean necesarias para las consultas de una colaboración. Esto puede contribuir a mitigar el riesgo de ataques diferenciales o de que otros miembros apliquen técnicas de ingeniería inversa a sus datos. Use la característica de lista de columnas permitidas para anotar otras columnas que quizás desee habilitar para consulta en el futuro. Para personalizar las columnas que se pueden usar para una colaboración determinada, cree tablas configuradas adicionales con la misma AWS Glue tabla subyacente.
- Especifique las funciones de la regla de análisis que son necesarias para el análisis en la colaboración. Esto puede contribuir a mitigar el riesgo derivado de errores de función poco frecuentes que puedan presentar información sobre un punto de datos concreto. Para personalizar las funciones que se pueden utilizar en una determinada colaboración, cree tablas configuradas adicionales con la misma tabla de AWS Glue subyacente.
- Añada restricciones de agregación a aquellas columnas cuyos valores a nivel de fila sean confidenciales. Esto incluye las columnas de la tabla configurada que también existen en las tablas y reglas de análisis de otros miembros de la colaboración como una restricción de agregación. También incluye las columnas de su tabla configurada que no se pueden consultar; es decir, las columnas que figuran en su tabla configurada, pero no en la regla de análisis. Las restricciones de agregación pueden contribuir a mitigar el riesgo de correlacionar los resultados de las consultas con datos externos a la colaboración.
- Cree colaboraciones y reglas de análisis de prueba para probar las restricciones creadas con las reglas de análisis especificadas.

- Revise las tablas configuradas de los colaboradores y las reglas de análisis de los miembros en las tablas configuradas para comprobar que coinciden con las condiciones acordadas para la colaboración. Esto puede contribuir a mitigar el riesgo de que otros miembros manipulen sus propios datos para ejecutar consultas no acordadas.
- Revise la consulta de ejemplo proporcionada (solo en la consola) que se habilita en la tabla configurada después de configurar la regla de análisis.

 Note

Además de la consulta de ejemplo proporcionada, es posible realizar otras consultas en función de la regla de análisis y de otras tablas y reglas de análisis de los miembros de la colaboración.

- Puede agregar o actualizar una regla de análisis para una tabla configurada en una colaboración. Cuando lo haga, revise todas las colaboraciones a las que está asociada la tabla configurada y el efecto resultante. Esto le ayuda a asegurarse de que ninguna colaboración utilice reglas de análisis obsoletas.
- Revise las consultas ejecutadas en la colaboración para comprobar que coincidan con los casos de uso o con las consultas acordados para la colaboración (las consultas están disponibles en los registros de consultas cuando la característica Registro de consultas está activada). Esto puede contribuir a mitigar el riesgo de que los miembros ejecuten análisis no acordados y de que se produzcan posibles ataques (por ejemplo, ataques de canal lateral).
- Revise las columnas de la tabla configurada que se utilizan en las reglas de análisis de los miembros de la colaboración y en las consultas para asegurarse de que coinciden con lo acordado en la colaboración (las consultas están disponibles en los registros de consultas cuando la característica está activada). Esto puede contribuir a mitigar el riesgo de que otros miembros manipulen sus propios datos para ejecutar consultas no acordadas.

Identity and Access Management para AWS Clean Rooms

AWS Identity and Access Management (IAM) es una herramienta Servicio de AWS que ayuda al administrador a controlar de forma segura el acceso a AWS los recursos. Los administradores de IAM controlan quién puede autenticarse (iniciar sesión) y quién puede autorizarse (tener permisos) para usar los recursos. AWS Clean Rooms La IAM es una Servicio de AWS opción que puede utilizar sin coste adicional.

Temas

- [Público](#)
- [Autenticación con identidades](#)
- [Administración del acceso con políticas](#)
- [¿Cómo AWS Clean Rooms funciona con IAM](#)
- [Ejemplos de políticas basadas en la identidad para AWS Clean Rooms](#)
- [AWS políticas gestionadas para AWS Clean Rooms](#)
- [Solución de problemas AWS Clean Rooms de identidad y acceso](#)
- [Prevención de la sustitución confusa entre servicios](#)
- [Comportamientos de IAM para AWS Clean Rooms ML](#)
- [Comportamientos de IAM para los modelos personalizados de Clean Rooms ML](#)

Público

La forma de usar AWS Identity and Access Management (IAM) varía según la función que desempeñes:

- Usuario del servicio: solicite permisos al administrador si no puede acceder a las características (consulte [Solución de problemas AWS Clean Rooms de identidad y acceso](#)).
- Administrador del servicio: determine el acceso de los usuarios y envíe las solicitudes de permiso (consulte [¿Cómo AWS Clean Rooms funciona con IAM](#)).
- Administrador de IAM: escribe las políticas para administrar el acceso (consulte [Ejemplos de políticas basadas en la identidad para AWS Clean Rooms](#)).

Autenticación con identidades

La autenticación es la forma en que inicias sesión AWS con tus credenciales de identidad. Debe estar autenticado (con quien haya iniciado sesión AWS) como usuario de IAM o asumiendo una función de IAM. Usuario raíz de la cuenta de AWS

Puede iniciar sesión AWS como una identidad federada mediante las credenciales proporcionadas a través de una fuente de identidad. AWS IAM Identity Center Los usuarios (Centro de identidades de IAM) o la autenticación de inicio de sesión único de su empresa son ejemplos de identidades

federadas. Al iniciar sesión como una identidad federada, su administrador habrá configurado previamente la federación de identidades mediante roles de IAM. Cuando accede AWS mediante la federación, asume indirectamente un rol.

Según el tipo de usuario que sea, puede iniciar sesión en el portal Consola de administración de AWS o en el de AWS acceso. Para obtener más información sobre cómo iniciar sesión AWS, consulte [Cómo iniciar sesión Cuenta de AWS en su](#) Guía del AWS Sign-In usuario.

Si accede AWS mediante programación, AWS proporciona un kit de desarrollo de software (SDK) y una interfaz de línea de comandos (CLI) para firmar criptográficamente sus solicitudes con sus credenciales. Si no utilizas AWS herramientas, debes firmar las solicitudes tú mismo. Para obtener más información sobre el uso del método recomendado para firmar las solicitudes usted mismo, consulte [Proceso de firma de Signature Version 4](#) en la Referencia general de AWS.

Independientemente del método de autenticación que use, es posible que deba proporcionar información de seguridad adicional. Por ejemplo, le AWS recomienda que utilice la autenticación multifactor (MFA) para aumentar la seguridad de su cuenta. Para obtener más información, consulte [Autenticación multifactor](#) en la Guía del usuario de AWS IAM Identity Center y [Uso de la autenticación multifactor \(MFA\) en AWS](#) en la Guía del usuario de IAM.

Cuenta de AWS usuario root

Al crear una Cuenta de AWS, comienza con una identidad de inicio de sesión que tiene acceso completo a todos Servicios de AWS los recursos de la cuenta. Esta identidad recibe el nombre de usuario raíz de la Cuenta de AWS y se accede a ella iniciando sesión con el email y la contraseña que utilizó para crear la cuenta. Recomendamos que no utilice el usuario raíz para las tareas cotidianas. Proteja las credenciales del usuario raíz y utilícelas sólo para las tareas que el usuario raíz pueda realizar. Para obtener la lista completa de tareas que requieren que inicie sesión como usuario raíz, consulte [Credenciales de Usuario raíz de la cuenta de AWS e identidades de IAM](#) en la Referencia general de AWS.

Identidad federada

Como práctica recomendada, exija a los usuarios humanos que utilicen la federación con un proveedor de identidades para acceder Servicios de AWS mediante credenciales temporales.

Una identidad federada es un usuario del directorio empresarial, del proveedor de identidades web o al Directory Service que se accede Servicios de AWS mediante credenciales de una fuente de identidad. Las identidades federadas asumen roles que proporcionan credenciales temporales.

Para una administración de acceso centralizada, se recomienda AWS IAM Identity Center. Para obtener más información, consulte [¿Qué es el Centro de identidades de IAM?](#) en la Guía del usuario de AWS IAM Identity Center .

Usuarios y grupos de IAM

Un [usuario de IAM](#) es una identidad con permisos específicos para una sola persona o aplicación. Recomendamos el uso de credenciales temporales en lugar de usuarios de IAM con credenciales de larga duración. Para obtener más información, consulte [Exigir a los usuarios humanos que utilicen la federación con un proveedor de identidad para acceder AWS mediante credenciales temporales](#) en la Guía del usuario de IAM.

Un [grupo de IAM](#) especifica un conjunto de usuarios de IAM y facilita la administración de los permisos para grupos grandes de usuarios. Para obtener más información, consulte [Casos de uso para usuarios de IAM](#) en la Guía del usuario de IAM.

Roles de IAM

Un [Rol de IAM](#) es una identidad con permisos específicos que proporciona credenciales temporales. Puede asumir un rol [cambiando de un rol de usuario a uno de IAM \(consola\)](#) o llamando a una AWS CLI operación de AWS API. Para obtener más información, consulte [Métodos para asumir un rol](#) en la Guía del usuario de IAM.

Los roles de IAM son útiles para el acceso de usuario federado, los permisos de usuario de IAM temporales, el acceso entre cuentas, el acceso entre servicios y las aplicaciones que se ejecutan en Amazon EC2. Para obtener más información, consulte [Acceso a recursos entre cuentas en IAM](#) en la Guía del usuario de IAM.

Administración del acceso con políticas

AWS Para controlar el acceso, puede crear políticas y adjuntarlas a AWS identidades o recursos. Una política es un objeto AWS que, cuando se asocia a una identidad o un recurso, define sus permisos. AWS evalúa estas políticas cuando un director (usuario, usuario raíz o sesión de rol) realiza una solicitud. Los permisos de las políticas determinan si la solicitud está permitida o denegada. La mayoría de las políticas se almacenan AWS como documentos JSON. Para obtener más información sobre la estructura y el contenido de los documentos de política JSON, consulte [Información general de políticas JSON](#) en la Guía del usuario de IAM.

Los administradores pueden usar las políticas de AWS JSON para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y bajo qué condiciones.

Cada entidad de IAM (usuario o rol) comienza sin permisos. De forma predeterminada, los usuarios no pueden hacer nada, ni siquiera cambiar sus propias contraseñas. Para conceder permiso a un usuario para hacer algo, el administrador debe asociar una política de permisos a un usuario. O bien el administrador puede agregar al usuario a un grupo que tenga los permisos necesarios. Cuando el administrador concede permisos a un grupo, todos los usuarios de ese grupo obtienen los permisos.

Las políticas de IAM definen permisos para una acción independientemente del método que se utilice para realizar la operación. Por ejemplo, suponga que dispone de una política que permite la acción `iam:GetRole`. Un usuario con esa política puede obtener información sobre el rol de la API Consola de administración de AWS, la AWS CLI, la o la AWS API.

Políticas basadas en identidades

Las políticas basadas en identidad son documentos de políticas de permisos JSON que puede asociar a una identidad, como un usuario de IAM, un grupo de usuarios o un rol. Estas políticas controlan qué acciones pueden realizar los usuarios y los roles, en qué recursos y en qué condiciones. Para obtener más información sobre cómo crear una política basada en identidad, consulte [Creación de políticas de IAM](#) en la Guía del usuario de IAM.

Las políticas basadas en identidades pueden clasificarse además como políticas insertadas o políticas administradas. Las políticas insertadas se integran directamente en un único usuario, grupo o rol. Las políticas administradas son políticas independientes que puedes adjuntar a varios usuarios, grupos y roles de su Cuenta de AWS. Las políticas administradas incluyen políticas AWS administradas y políticas administradas por el cliente. Para más información sobre cómo elegir una política administrada o una política insertada, consulte [Elegir entre políticas administradas y políticas insertadas](#) en la Guía del usuario de IAM.

Políticas basadas en recursos

Las políticas basadas en recursos son documentos de política JSON que se asocian a un recurso. Los ejemplos de políticas basadas en recursos son las políticas de confianza de roles de IAM y las políticas de bucket de Amazon S3. En los servicios que admiten políticas basadas en recursos, los administradores de servicios pueden utilizarlos para controlar el acceso a un recurso específico. Para el recurso al que se asocia la política, la política define qué acciones puede realizar una entidad principal especificada en ese recurso y en qué condiciones. Debe [especificar una entidad principal](#)

en una política basada en recursos. Los principales pueden incluir cuentas, usuarios, roles, usuarios federados o. Servicios de AWS

Las políticas basadas en recursos son políticas insertadas que se encuentran en ese servicio. No puedes usar políticas AWS gestionadas de IAM en una política basada en recursos.

Otros tipos de políticas

AWS admite tipos de políticas adicionales y menos comunes. Estos tipos de políticas pueden establecer el máximo de permisos que los tipos de políticas más frecuentes le conceden.

- **Límites de permisos:** un límite de permisos es una característica avanzada que le permite establecer los permisos máximos que una política basada en identidad puede conceder a una entidad de IAM (usuario o rol de IAM). Puede establecer un límite de permisos para una identidad. Los permisos resultantes son la intersección de las políticas basadas en identidad de la entidad y los límites de sus permisos. Las políticas basadas en recursos que especifiquen el usuario o rol en el campo `Principal` no estarán restringidas por el límite de permisos. Una denegación explícita en cualquiera de estas políticas anulará el permiso. Para obtener más información sobre los límites de los permisos, consulte [Límites de permisos para las entidades de IAM](#) en la Guía del usuario de IAM.
- **Políticas de control de servicios (SCPs):** SCPs son políticas de JSON que especifican los permisos máximos para una organización o unidad organizativa (OU). AWS Organizations es un servicio para agrupar y administrar de forma centralizada varios de los Cuentas de AWS que son propiedad de su empresa. Si habilitas todas las funciones de una organización, puedes aplicar políticas de control de servicios (SCPs) a una o a todas tus cuentas. El SCP limita los permisos de las entidades en las cuentas de los miembros, incluidas las de cada una Usuario raíz de la cuenta de AWS. Para obtener más información sobre Organizations SCPs, consulte [Cómo SCPs trabajar](#) en la Guía del AWS Organizations usuario.
- **Políticas de sesión:** las políticas de sesión son políticas avanzadas que se pasan como parámetro cuando se crea una sesión temporal mediante programación para un rol o un usuario federado. Los permisos de la sesión resultantes son la intersección de las políticas basadas en identidades del rol y las políticas de la sesión. Los permisos también pueden proceder de una política en función de recursos. Una denegación explícita en cualquiera de estas políticas anulará el permiso. Para más información, consulte [Políticas de sesión](#) en la Guía del usuario de IAM.

Varios tipos de políticas

Cuando se aplican varios tipos de políticas a una solicitud, los permisos resultantes son más complicados de entender. Para saber cómo se AWS determina si se debe permitir una solicitud cuando se trata de varios tipos de políticas, consulte la [lógica de evaluación de políticas](#) en la Guía del usuario de IAM.

¿Cómo AWS Clean Rooms funciona con IAM

Antes de utilizar IAM para gestionar el acceso AWS Clean Rooms, infórmese sobre las funciones de IAM disponibles para su uso. AWS Clean Rooms

Funciones de IAM que puede utilizar con AWS Clean Rooms

Característica de IAM	AWS Clean Rooms soporte
Políticas basadas en identidades	Sí
Políticas basadas en recursos	Parcial
Acciones de políticas	Sí
Recursos de políticas	Sí
Claves de condición de política (específicas del servicio)	Parcial
ACLs	No
ABAC (etiquetas en políticas)	Sí
Credenciales temporales	Sí
Sesiones de acceso directo (FAS)	Sí
Roles de servicio	Sí
Roles vinculados al servicio	No

Para obtener una visión general de cómo AWS Clean Rooms y otras funciones Servicios de AWS funcionan con la mayoría de las funciones de IAM, consulte Servicios de AWS la Guía del usuario de [IAM sobre cómo funcionan con IAM](#).

Políticas basadas en la identidad para AWS Clean Rooms

Compatibilidad con las políticas basadas en identidad: sí

Las políticas basadas en identidad son documentos de políticas de permisos JSON que puede asociar a una identidad, como un usuario de IAM, un grupo de usuarios o un rol. Estas políticas controlan qué acciones pueden realizar los usuarios y los roles, en qué recursos y en qué condiciones. Para obtener más información sobre cómo crear una política basada en la identidad, consulte [Definición de permisos de IAM personalizados con políticas administradas por el cliente](#) en la Guía del usuario de IAM.

Con las políticas basadas en identidades de IAM, puede especificar las acciones y los recursos permitidos o denegados, así como las condiciones en las que se permiten o deniegan las acciones. Para obtener más información sobre los elementos que puede utilizar en una política de JSON, consulte [Referencia de los elementos de la política de JSON de IAM](#) en la Guía del usuario de IAM.

Ejemplos de políticas basadas en la identidad para AWS Clean Rooms

Para ver ejemplos de políticas AWS Clean Rooms basadas en la identidad, consulte. [Ejemplos de políticas basadas en la identidad para AWS Clean Rooms](#)

Políticas basadas en recursos incluidas AWS Clean Rooms

Compatibilidad con las políticas basadas en recursos: parcial

Las políticas basadas en recursos son documentos de política JSON que se asocian a un recurso. Los ejemplos de políticas basadas en recursos son las políticas de confianza de roles de IAM y las políticas de bucket de Amazon S3. En los servicios que admiten políticas basadas en recursos, los administradores de servicios pueden utilizarlos para controlar el acceso a un recurso específico. Para el recurso al que se asocia la política, la política define qué acciones puede realizar una entidad principal especificada en ese recurso y en qué condiciones. Debe [especificar una entidad principal](#) en una política basada en recursos. Los principales pueden incluir cuentas, usuarios, roles, usuarios federados o. Servicios de AWS

Para habilitar el acceso entre cuentas, puede especificar toda una cuenta o entidades de IAM de otra cuenta como la entidad principal de una política en función de recursos. Para obtener más información, consulte [Acceso a recursos entre cuentas en IAM](#) en la Guía del usuario de IAM.

El AWS Clean Rooms servicio solo admite un tipo de política basada en recursos denominada política de recursos administrados de modelo similar configurada, que se adjunta a un modelo similar configurado. Esta política define qué entidades principales pueden realizar acciones en el modelo similar configurado.

Para obtener información sobre cómo asociar una política basada en recursos a un modelo similar configurado, consulte [Comportamientos de IAM para AWS Clean Rooms ML](#).

Acciones políticas para AWS Clean Rooms

Compatibilidad con las acciones de políticas: sí

Los administradores pueden usar las políticas de AWS JSON para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

El elemento `Action` de una política JSON describe las acciones que puede utilizar para conceder o denegar el acceso en una política. Incluya acciones en una política para conceder permisos y así llevar a cabo la operación asociada.

Para ver una lista de AWS Clean Rooms acciones, consulte [las acciones definidas AWS Clean Rooms](#) en la Referencia de autorización del servicio.

Las acciones políticas AWS Clean Rooms utilizan el siguiente prefijo antes de la acción.

```
cleanrooms
```

Para especificar varias acciones en una única instrucción, sepárelas con comas.

```
"Action": [  
    "cleanrooms:action1",  
    "cleanrooms:action2"  
]
```

Para ver ejemplos de políticas AWS Clean Rooms basadas en la identidad, consulte. [Ejemplos de políticas basadas en la identidad para AWS Clean Rooms](#)

Recursos de políticas para AWS Clean Rooms

Compatibilidad con los recursos de políticas: sí

Los administradores pueden usar las políticas de AWS JSON para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

El elemento `Resource` de la política JSON especifica el objeto u objetos a los que se aplica la acción. Como práctica recomendada, especifique un recurso utilizando el [Nombre de recurso de Amazon \(ARN\)](#). En el caso de las acciones que no admiten permisos por recurso, utilice un carácter comodín (*) para indicar que la instrucción se aplica a todos los recursos.

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

Para ver una lista de los tipos de AWS Clean Rooms recursos y sus tipos ARNs, consulte [los recursos definidos AWS Clean Rooms](#) en la Referencia de autorización de servicios. Para obtener información sobre las acciones con las que puede especificar el ARN de cada recurso, consulte [Acciones definidas por AWS Clean Rooms](#).

Para ver ejemplos de políticas AWS Clean Rooms basadas en la identidad, consulte. [Ejemplos de políticas basadas en la identidad para AWS Clean Rooms](#)

Claves de condición de la política para AWS Clean Rooms

Compatibilidad con claves de condición de políticas específicas del servicio: parcial

Los administradores pueden usar las políticas de AWS JSON para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

El elemento `Condition` especifica cuándo se ejecutan las instrucciones en función de criterios definidos. Puede crear expresiones condicionales que utilizan [operadores de condición](#), tales como igual o menor que, para que la condición de la política coincida con los valores de la solicitud. Para ver todas las claves de condición AWS globales, consulte las claves de [contexto de condición AWS globales](#) en la Guía del usuario de IAM.

Para obtener información sobre cómo AWS Clean Rooms ML utiliza las claves de condición de las políticas, consulte [Comportamientos de IAM para AWS Clean Rooms ML](#).

ACLs in AWS Clean Rooms

Soportes ACLs: No

Las listas de control de acceso (ACLs) controlan qué directores (miembros de la cuenta, usuarios o roles) tienen permisos para acceder a un recurso. ACLs son similares a las políticas basadas en recursos, aunque no utilizan el formato de documento de políticas JSON.

ABAC con AWS Clean Rooms

Admite ABAC (etiquetas en las políticas): sí

El control de acceso basado en atributos (ABAC) es una estrategia de autorización que define permisos en función de atributos denominados etiquetas. Puede adjuntar etiquetas a las entidades y AWS los recursos de IAM y, a continuación, diseñar políticas de ABAC para permitir las operaciones cuando la etiqueta del principal coincida con la etiqueta del recurso.

Para controlar el acceso en función de etiquetas, debe proporcionar información de las etiquetas en el [elemento de condición](#) de una política utilizando las claves de condición `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` o `aws:TagKeys`.

Si un servicio admite las tres claves de condición para cada tipo de recurso, el valor es Sí para el servicio. Si un servicio admite las tres claves de condición solo para algunos tipos de recursos, el valor es Parcial.

Para obtener más información sobre ABAC, consulte [Definición de permisos con la autorización de ABAC](#) en la Guía del usuario de IAM. Para ver un tutorial con los pasos para configurar ABAC, consulte [Uso del control de acceso basado en atributos \(ABAC\)](#) en la Guía del usuario de IAM.

Utilizar credenciales temporales con AWS Clean Rooms

Compatibilidad con credenciales temporales: sí

Las credenciales temporales proporcionan acceso a AWS los recursos a corto plazo y se crean automáticamente cuando se utiliza la federación o se cambia de rol. AWS recomienda generar credenciales temporales de forma dinámica en lugar de utilizar claves de acceso a largo plazo. Para obtener más información, consulte [Credenciales de seguridad temporales en IAM](#) y [Servicios de AWS que funcionan con IAM](#) en la Guía del usuario de IAM.

Sesiones de acceso directo para AWS Clean Rooms

Admite sesiones de acceso directo (FAS): sí

Las sesiones de acceso directo (FAS) utilizan los permisos de la persona principal que llama Servicio de AWS, junto con la solicitud, Servicio de AWS para realizar solicitudes a los servicios descendentes. Para obtener información sobre las políticas a la hora de realizar solicitudes de FAS, consulte [Sesiones de acceso directo](#).

Funciones de servicio para AWS Clean Rooms

Compatible con roles de servicio: sí

Un rol de servicio es un [rol de IAM](#) que asume un servicio para realizar acciones en su nombre. Un administrador de IAM puede crear, modificar y eliminar un rol de servicio desde IAM. Para obtener más información, consulte [Crear un rol para delegar permisos a un Servicio de AWS](#) en la Guía del usuario de IAM.

Warning

Cambiar los permisos de un rol de servicio podría interrumpir AWS Clean Rooms la funcionalidad. Edite las funciones de servicio solo cuando se AWS Clean Rooms proporcionen instrucciones para hacerlo.

Funciones vinculadas al servicio para AWS Clean Rooms

Compatibilidad con roles vinculados al servicio: no

Un rol vinculado a un servicio es un tipo de rol de servicio que está vinculado a un. Servicio de AWS El servicio puede asumir el rol para realizar una acción en su nombre. Los roles vinculados al servicio aparecen en usted Cuenta de AWS y son propiedad del servicio. Un administrador de IAM puede ver, pero no editar, los permisos de los roles vinculados a servicios.

Para más información sobre cómo crear o administrar roles vinculados a servicios, consulta [Servicios de AWS que funcionan con IAM](#). Busque un servicio en la tabla que incluya Yes en la columna Rol vinculado a un servicio. Seleccione el vínculo Sí para ver la documentación acerca del rol vinculado a servicios para ese servicio.

Ejemplos de políticas basadas en la identidad para AWS Clean Rooms

De forma predeterminada, los usuarios y roles no tienen permiso para crear, ver ni modificar recursos de AWS Clean Rooms . Un administrador de IAM puede crear políticas de IAM para conceder permisos a los usuarios para realizar acciones en los recursos que necesitan.

Para obtener información acerca de cómo crear una política basada en identidades de IAM mediante el uso de estos documentos de políticas JSON de ejemplo, consulte [Creación de políticas de IAM \(consola\)](#) en la Guía del usuario de IAM.

Para obtener más información sobre las acciones y los tipos de recursos definidos AWS Clean Rooms, incluido el formato de cada uno de los tipos de recursos, consulte [las claves de condición, recursos y acciones de la Referencia AWS Clean Rooms](#) de autorización de servicios. ARNs

Temas

- [Prácticas recomendadas sobre las políticas](#)
- [Uso de la consola AWS Clean Rooms](#)
- [Cómo permitir a los usuarios consultar sus propios permisos](#)

Prácticas recomendadas sobre las políticas

Las políticas basadas en la identidad determinan si alguien puede crear AWS Clean Rooms recursos de tu cuenta, acceder a ellos o eliminarlos. Estas acciones pueden generar costos adicionales para su Cuenta de AWS. Siga estas directrices y recomendaciones al crear o editar políticas basadas en identidades:

- Comience con las políticas AWS administradas y avance hacia los permisos con privilegios mínimos: para empezar a conceder permisos a sus usuarios y cargas de trabajo, utilice las políticas AWS administradas que otorgan permisos para muchos casos de uso comunes. Están disponibles en su Cuenta de AWS. Le recomendamos que reduzca aún más los permisos definiendo políticas administradas por el AWS cliente que sean específicas para sus casos de uso. Con el fin de obtener más información, consulte las [políticas administradas por AWS](#) o las [políticas administradas por AWS para funciones de tarea](#) en la Guía de usuario de IAM.
- Aplique permisos de privilegio mínimo: cuando establezca permisos con políticas de IAM, conceda solo los permisos necesarios para realizar una tarea. Para ello, debe definir las acciones que se pueden llevar a cabo en determinados recursos en condiciones específicas, también conocidos como permisos de privilegios mínimos. Con el fin de obtener más información sobre el uso de IAM para aplicar permisos, consulte [Políticas y permisos en IAM](#) en la Guía del usuario de IAM.
- Utilice condiciones en las políticas de IAM para restringir aún más el acceso: puede agregar una condición a sus políticas para limitar el acceso a las acciones y los recursos. Por ejemplo, puede escribir una condición de políticas para especificar que todas las solicitudes deben enviarse utilizando SSL. También puedes usar condiciones para conceder el acceso a las acciones del servicio si se utilizan a través de una acción específica Servicio de AWS, por ejemplo

CloudFormation. Para obtener más información, consulte [Elementos de la política de JSON de IAM: Condición](#) en la Guía del usuario de IAM.

- Utiliza el analizador de acceso de IAM para validar las políticas de IAM con el fin de garantizar la seguridad y funcionalidad de los permisos: el analizador de acceso de IAM valida políticas nuevas y existentes para que respeten el lenguaje (JSON) de las políticas de IAM y las prácticas recomendadas de IAM. El analizador de acceso de IAM proporciona más de 100 verificaciones de políticas y recomendaciones procesables para ayudar a crear políticas seguras y funcionales. Para más información, consulte [Validación de políticas con el Analizador de acceso de IAM](#) en la Guía del usuario de IAM.
- Requerir autenticación multifactor (MFA): si tiene un escenario que requiere usuarios de IAM o un usuario raíz en Cuenta de AWS su cuenta, active la MFA para mayor seguridad. Para exigir la MFA cuando se invoquen las operaciones de la API, añada condiciones de MFA a sus políticas. Para más información, consulte [Acceso seguro a la API con MFA](#) en la Guía del usuario de IAM.

Para obtener más información sobre las prácticas recomendadas de IAM, consulte [Prácticas recomendadas de seguridad en IAM](#) en la Guía del usuario de IAM.

Uso de la consola AWS Clean Rooms

Para acceder a la AWS Clean Rooms consola, debe tener un conjunto mínimo de permisos. Estos permisos deben permitirle enumerar y ver detalles sobre los AWS Clean Rooms recursos de su cuenta Cuenta de AWS. Si crea una política basada en identidades que sea más restrictiva que el mínimo de permisos necesarios, la consola no funcionará del modo esperado para las entidades (usuarios o roles) que tengan esa política.

No es necesario que concedas permisos mínimos de consola a los usuarios que solo realicen llamadas a la API AWS CLI o a la AWS API. En su lugar, permita el acceso únicamente a las acciones que coincidan con la operación de API que intentan realizar.

Para garantizar que los usuarios y los roles puedan seguir utilizando la AWS Clean Rooms consola, adjunte también la política *ReadOnly* AWS gestionada AWS Clean Rooms *FullAccess* o la política gestionada a las entidades. Para obtener más información, consulte [Adición de permisos a un usuario](#) en la Guía del usuario de IAM:

Cómo permitir a los usuarios consultar sus propios permisos

En este ejemplo, se muestra cómo podría crear una política que permita a los usuarios de IAM ver las políticas administradas e insertadas que se asocian a la identidad de sus usuarios. Esta política

incluye permisos para completar esta acción en la consola o mediante programación mediante la API AWS CLI o 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 políticas gestionadas para AWS Clean Rooms

Una política AWS administrada es una política independiente creada y administrada por AWS. Las políticas administradas están diseñadas para proporcionar permisos para muchos casos de uso comunes, de modo que pueda empezar a asignar permisos a usuarios, grupos y funciones.

Ten en cuenta que es posible que las políticas AWS administradas no otorguen permisos con privilegios mínimos para tus casos de uso específicos, ya que están disponibles para que los usen todos los AWS clientes. Se recomienda definir [políticas administradas por el cliente](#) específicas para sus casos de uso a fin de reducir aún más los permisos.

No puedes cambiar los permisos definidos en AWS las políticas administradas. Si AWS actualiza los permisos definidos en una política AWS administrada, la actualización afecta a todas las identidades principales (usuarios, grupos y roles) a las que está asociada la política. AWS es más probable que actualice una política AWS administrada cuando Servicio de AWS se lance una nueva o cuando estén disponibles nuevas operaciones de API para los servicios existentes.

Para obtener más información, consulte [Políticas administradas por AWS](#) en la Guía del usuario de IAM.

AWS política gestionada: **AWSCleanRoomsReadOnlyAccess**

Puede adjuntar **AWSCleanRoomsReadOnlyAccess** a sus entidades principales de IAM.

Esta política otorga permisos de solo lectura a recursos y metadatos de una colaboración de **AWSCleanRoomsReadOnlyAccess**.

Detalles de los permisos

Esta política incluye los permisos siguientes:

- **CleanRoomsRead**: concede a las entidades principales acceso de solo lectura al servicio.
- **ConsoleDisplayTables**— Permite a los directores acceder de solo lectura a los AWS Glue metadatos necesarios para mostrar los datos sobre AWS Glue las tablas subyacentes en la consola.
- **ConsoleLogSummaryQueryLogs**: concede a las entidades principales permiso para ver los registros de consultas.
- **ConsoleLogSummaryObtainLogs**: concede a las entidades principales permiso para recuperar los resultados de registro.

Para obtener una lista en JSON de los detalles de la política, consulte la Guía [AWSCleanRoomsReadOnlyAccess](#) de referencia de políticas AWS administradas.

AWS política gestionada: **AWSCleanRoomsFullAccess**

Puede adjuntar **AWSCleanRoomsFullAccess** a sus entidades principales de IAM.

Esta política otorga permisos administrativos que permiten el acceso total (lectura, escritura y actualización) a los recursos y metadatos de una AWS Clean Rooms colaboración. Esta política incluye el acceso para realizar consultas.

Detalles de los permisos

Esta política incluye los permisos siguientes:

- **CleanRoomsAccess**— Otorga acceso total a todas las acciones de todos los recursos para AWS Clean Rooms.
- **PassServiceRole**: concede acceso para transferir un rol de servicio únicamente al servicio (condición **PassedToService**) cuyo nombre contenga "cleanrooms".
- **ListRolesToPickServiceRole**— Permite a los directores enumerar todas sus funciones para poder elegir una función de servicio cuando la utilicen AWS Clean Rooms.
- **GetRoleAndListRolePoliciesToInspectServiceRole**: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- **ListPoliciesToInspectServiceRolePolicy**: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- **GetPolicyToInspectServiceRolePolicy**: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- **ConsoleDisplayTables**— Permite a los directores acceder de solo lectura a los AWS Glue metadatos necesarios para mostrar los datos sobre AWS Glue las tablas subyacentes de la consola.
- **ConsolePickQueryResultsBucketListAll**: permite a las entidades principales elegir, de una lista de todos los buckets de S3 disponibles, un bucket de Amazon S3 en el que se escriban los resultados de sus consultas.
- **SetQueryResultsBucket**: permite a las entidades principales elegir un bucket de S3 en el que se escriban los resultados de sus consultas.
- **ConsoleDisplayQueryResults**: permite a las entidades principales mostrar al cliente los resultados de las consultas, leídos desde el bucket de S3.

- **WriteQueryResults**: permite a las entidades principales escribir los resultados de las consultas en un bucket de S3 propiedad del cliente.
- **EstablishLogDeliveries**— Permite a los directores enviar registros de consultas al grupo de registros de Amazon CloudWatch Logs de un cliente.
- **SetupLogGroupsDescribe**— Permite a los directores utilizar el proceso de creación de grupos de CloudWatch registros de Amazon Logs.
- **SetupLogGroupsCreate**— Permite a los directores crear un grupo de CloudWatch registros de Amazon Logs.
- **SetupLogGroupsResourcePolicy**— Permite a los directores configurar una política de recursos en el grupo de CloudWatch registros de Amazon Logs.
- **ConsoleLogSummaryQueryLogs**: concede a las entidades principales permiso para ver los registros de consultas.
- **ConsoleLogSummaryObtainLogs**: concede a las entidades principales permiso para recuperar los resultados de registro.

Para ver una lista en JSON de los detalles de la política, consulte [AWSCleanRoomsFullAccess](#) la Guía de referencia de políticas AWS administradas.

AWS política gestionada: **AWSCleanRoomsFullAccessNoQuerying**

Puede adjuntar **AWSCleanRoomsFullAccessNoQuerying** a sus IAM principals.

Esta política otorga permisos administrativos que permiten el acceso total (lectura, escritura y actualización) a los recursos y metadatos de una AWS Clean Rooms colaboración. Esta política no incluye el acceso para realizar consultas.

Detalles de los permisos

Esta política incluye los permisos siguientes:

- **CleanRoomsAccess**— Otorga acceso total a todas las acciones de todos los recursos AWS Clean Rooms, excepto para las consultas en las colaboraciones.
- **CleanRoomsNoQuerying**: deniega explícitamente **StartProtectedQuery** y **UpdateProtectedQuery** para impedir las consultas.
- **PassServiceRole**: concede acceso para transferir un rol de servicio únicamente al servicio (condición **PassedToService**) cuyo nombre contenga "cleanrooms".

- `ListRolesToPickServiceRole`— Permite a los directores enumerar todas sus funciones para poder elegir una función de servicio cuando la utilicen. AWS Clean Rooms
- `GetRoleAndListRolePoliciesToInspectServiceRole`: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- `ListPoliciesToInspectServiceRolePolicy`: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- `GetPolicyToInspectServiceRolePolicy`: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- `ConsoleDisplayTables`— Permite a los directores acceder de solo lectura a los AWS Glue metadatos necesarios para mostrar los datos sobre AWS Glue las tablas subyacentes de la consola.
- `EstablishLogDeliveries`— Permite a los directores enviar registros de consultas al grupo de registros de Amazon CloudWatch Logs de un cliente.
- `SetupLogGroupsDescribe`— Permite a los directores utilizar el proceso de creación de grupos de CloudWatch registros de Amazon Logs.
- `SetupLogGroupsCreate`— Permite a los directores crear un grupo de CloudWatch registros de Amazon Logs.
- `SetupLogGroupsResourcePolicy`— Permite a los directores configurar una política de recursos en el grupo de CloudWatch registros de Amazon Logs.
- `ConsoleLogSummaryQueryLogs`: concede a las entidades principales permiso para ver los registros de consultas.
- `ConsoleLogSummaryObtainLogs`: concede a las entidades principales permiso para recuperar los resultados de registro.
- `cleanrooms`: administre las colaboraciones, las plantillas de análisis, las tablas configuradas, las suscripciones y los recursos asociados dentro del servicio de AWS Clean Rooms . Realice diversas operaciones, como crear, actualizar, eliminar, mostrar y recuperar información sobre estos recursos.
- `iam`— Transfiere al servicio las funciones de AWS Clean Rooms servicio cuyos nombres contengan `cleanrooms` «». Enumere las funciones y políticas e inspeccione las funciones y políticas de servicio relacionadas con el AWS Clean Rooms servicio.
- `glue`— Recupere información sobre bases de datos, tablas, particiones y esquemas de AWS Glue. Esto es necesario para que el AWS Clean Rooms servicio muestre las fuentes de datos subyacentes e interactúe con ellas.

- **logs**— Gestione las entregas de registros, los grupos de registros y las políticas de recursos para CloudWatch los registros. Consulte y recupere los registros relacionados con el AWS Clean Rooms servicio. Estos permisos son necesarios para la supervisión, la auditoría y la solución de problemas dentro del servicio.

La política también deniega explícitamente las acciones `cleanrooms:StartProtectedQuery` y `cleanrooms:UpdateProtectedQuery` para evitar que los usuarios ejecuten o actualicen directamente las consultas protegidas, lo que debe hacerse a través de los mecanismos controlados de AWS Clean Rooms .

Para obtener una lista en JSON de los detalles de la política, consulte [AWSCleanRoomsFullAccessNoQuerying](#) la Guía de referencia de políticas AWS gestionadas.

AWS política gestionada: **AWSCleanRoomsMLReadOnlyAccess**

Puede adjuntar `AWSCleanRoomsMLReadOnlyAccess` a sus entidades principales de IAM.

Esta política otorga permisos de solo lectura a recursos y metadatos de una colaboración de `AWSCleanRoomsMLReadOnlyAccess`.

Esta política incluye los permisos siguientes:

- **CleanRoomsConsoleNavigation**— Otorga acceso para ver las pantallas de la AWS Clean Rooms consola.
- **CleanRoomsMLRead**: permite a las entidades principales el acceso de solo lectura al servicio de Clean Rooms ML.
- **PassCleanRoomsResources**— Otorga acceso para pasar AWS Clean Rooms recursos específicos.

Para ver una lista en JSON de los detalles de la política, consulta [AWSCleanHabitaciones MLRead OnlyAccess](#) en la Guía de referencia sobre políticas AWS gestionadas.

AWS política gestionada: **AWSCleanRoomsMLFullAccess**

Puede adjuntar `AWSCleanRoomsMLFullAccess` a sus entidades principales de IAM. Esta política concede permisos administrativos que permiten el acceso total (lectura, escritura y actualización) a los recursos y metadatos necesarios de Clean Rooms ML.

Detalles de los permisos

Esta política incluye los permisos siguientes:

- `CleanRoomsMLFullAccess`: concede acceso a todas las acciones de Clean Rooms ML.
- `PassServiceRole`: concede acceso para transferir un rol de servicio únicamente al servicio (condición `PassedToService`) cuyo nombre contenga "cleanrooms-ml".
- `CleanRoomsConsoleNavigation`— Otorga acceso para ver las pantallas de la AWS Clean Rooms consola.
- `CollaborationMembershipCheck`— Al iniciar un trabajo de generación de audiencia (segmento similar) dentro de una colaboración, el servicio Clean Rooms ML llama `ListMembers` para comprobar que la colaboración es válida, que la persona que llama es un miembro activo y que el propietario del modelo de audiencia configurado es un miembro activo. Este permiso siempre es obligatorio; el SID de navegación de la consola solo es necesario para los usuarios de la consola.
- `PassCleanRoomsResources`— Otorga acceso para transferir recursos específicos AWS Clean Rooms .
- `AssociateModels`: permite a las entidades principales asociar un modelo de Clean Rooms ML a la colaboración.
- `TagAssociations`: permite a las entidades principales añadir etiquetas a la asociación entre un modelo similar y una colaboración.
- `ListRolesToPickServiceRole`— Permite a los directores enumerar todas sus funciones para poder elegir una función de servicio cuando la utilicen AWS Clean Rooms.
- `GetRoleAndListRolePoliciesToInspectServiceRole`: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- `ListPoliciesToInspectServiceRolePolicy`: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- `GetPolicyToInspectServiceRolePolicy`: permite a las entidades principales ver el rol de servicio y la política correspondiente en IAM.
- `ConsoleDisplayTables`— Permite a los directores acceder de solo lectura a los AWS Glue metadatos necesarios para mostrar los datos sobre AWS Glue las tablas subyacentes de la consola.
- `ConsolePickOutputBucket`: permite a las entidades principales seleccionar buckets de Amazon S3 para las salidas configuradas del modelo de audiencia.
- `ConsolePickS3Location`: permite a las entidades principales seleccionar la ubicación dentro de un bucket para los resultados configurados del modelo de audiencia.

- `ConsoleDescribeECRRepositories`— Permite a los directores describir los repositorios e imágenes de Amazon ECR.

Para ver una lista en JSON de los detalles de la política, consulte [AWSCleanRooms MLFull Access](#) en la Guía de referencia de políticas AWS gestionadas.

AWS Clean Rooms actualizaciones de las políticas AWS gestionadas

Consulte los detalles sobre las actualizaciones de las políticas AWS administradas AWS Clean Rooms desde que este servicio comenzó a realizar el seguimiento de estos cambios. Para recibir alertas automáticas sobre los cambios en esta página, suscríbase a la fuente RSS de la página del historial del AWS Clean Rooms documento.

Cambio	Descripción	Fecha
AWSCleanRoomsFullAccessNoQuerying — Actualización de la política existente	Se agregaron salas limpias: UpdateConfiguredTableAllowe dColumns y salas limpias: UpdateCon figuredTableReference también. CleanRoomsAccess	29 de julio de 2025
AWSCleanRoomsMLReadOnlyAcce ss : actualización de una política actual	Se agregó PassCleanRoomsReso urces a AWSCleanRoomsMLRea dOnlyAccess. Se agregaron PassClean RoomsResources y ConsoleDe scribeECRRepositories a AWSCleanR oomsMLFullAccess.	10 de enero de 2025
AWSCleanRoomsFullAccessNoQuerying : actualización de una política actual	Se agregó cleanrooms:BatchGe tSchemaAnalysisRule a CleanRoom sAccess.	13 de mayo de 2024
AWSCleanRoomsFullAccess : actualización de una política actual	Se ha actualizado el ID de la declaraci ón en AWSCleanRoomsFullAccess de ConsolePickQueryResultsBucket a SetQueryResultsBucket en esta política para representar mejor los permisos, ya que los permisos son necesarios	21 de marzo de 2024

Cambio	Descripción	Fecha
	para configurar el bucket de resultados de las consultas con y sin la consola.	
AWSCleanRoomsMLReadOnlyAccess : política nueva AWSCleanRoomsMLFullAccess : política nueva	Se agregó AWSCleanRoomsMLReadOnlyAccess y AWSCleanRoomsMLFullAccess es compatible con AWS Clean Rooms ML.	29 de noviembre de 2023
AWSCleanRoomsFullAccessNoQuering : actualización de una política actual	Se ha agregado cleanrooms:CreateAnalysisTemplate, cleanrooms:GetAnalysisTemplate, cleanrooms:UpdateAnalysisTemplate, cleanrooms>DeleteAnalysisTemplate, cleanrooms:ListAnalysisTemplates, cleanrooms:GetCollaborationAnalysisTemplate, cleanrooms:BatchGetCollaborationAnalysisTemplate y cleanrooms:ListCollaborationAnalysisTemplates a CleanRoomsAccess para habilitar la nueva característica de plantillas de análisis.	31 de julio de 2023
AWSCleanRoomsFullAccessNoQuering : actualización de una política actual	Se ha añadido cleanrooms:ListTagsForResource, cleanrooms:UntagResource y cleanrooms:TagResource a CleanRoomsAccess para habilitar el etiquetado de recursos.	21 de marzo de 2023
AWS Clean Rooms comenzó a rastrear los cambios	AWS Clean Rooms comenzó a realizar un seguimiento de los cambios de sus políticas AWS gestionadas.	12 de enero de 2023

Solución de problemas AWS Clean Rooms de identidad y acceso

Utilice la siguiente información como ayuda para diagnosticar y solucionar los problemas más comunes que pueden surgir al trabajar con una AWS Clean Rooms IAM.

Temas

- [No estoy autorizado a realizar ninguna acción en AWS Clean Rooms](#)
- [No estoy autorizado a realizar tareas como: PassRole](#)
- [Quiero permitir que personas ajenas a mí accedan Cuenta de AWS a mis AWS Clean Rooms recursos](#)

No estoy autorizado a realizar ninguna acción en AWS Clean Rooms

Si recibe un error que indica que no tiene autorización para realizar una acción, las políticas se deben actualizar para permitirle realizar la acción.

En el siguiente ejemplo, el error se produce cuando el usuario de IAM mateojackson intenta utilizar la consola para consultar los detalles acerca de un recurso ficticio *my-example-widget*, pero no tiene los permisos ficticios `cleanrooms:GetWidget`.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
cleanrooms:GetWidget on resource: my-example-widget
```

En este caso, la política de Mateo se debe actualizar para permitirle acceder al recurso *my-example-widget* mediante la acción `cleanrooms:GetWidget`.

Si necesita ayuda, póngase en contacto con su AWS administrador. El administrador es la persona que le proporcionó las credenciales de inicio de sesión.

No estoy autorizado a realizar tareas como: PassRole

Si recibe un error que indica que no tiene autorización para realizar la acción `iam:PassRole`, las políticas deben actualizarse a fin de permitirle pasar un rol a AWS Clean Rooms.

Algunos Servicios de AWS permiten transferir una función existente a ese servicio en lugar de crear una nueva función de servicio o una función vinculada a un servicio. Para ello, debe tener permisos para transferir la función al servicio.

En el siguiente ejemplo, el error se produce cuando un usuario de IAM denominado marymajor intenta utilizar la consola para realizar una acción en AWS Clean Rooms. Sin embargo, la acción requiere que el servicio cuente con permisos que otorguen un rol de servicio. Mary no tiene permisos para transferir el rol al servicio.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
iam:PassRole
```

En este caso, las políticas de Mary se deben actualizar para permitirle realizar la acción `iam:PassRole`.

Si necesita ayuda, póngase en contacto con su AWS administrador. El administrador es la persona que le proporcionó las credenciales de inicio de sesión.

Quiero permitir que personas ajenas a mí accedan Cuenta de AWS a mis AWS Clean Rooms recursos

Se puede crear un rol que los usuarios de otras cuentas o las personas externas a la organización puedan utilizar para acceder a sus recursos. Se puede especificar una persona de confianza para que asuma el rol.

Para obtener más información, consulte lo siguiente:

- Para saber si AWS Clean Rooms es compatible con estas funciones, consulte [¿Cómo AWS Clean Rooms funciona con IAM](#).
- Para obtener información sobre cómo proporcionar acceso a los recursos de su Cuentas de AWS propiedad, consulte [Proporcionar acceso a un usuario de IAM en otro usuario de su propiedad Cuenta de AWS en](#) la Guía del usuario de IAM.
- Para obtener información sobre cómo proporcionar acceso a tus recursos a terceros Cuentas de AWS, consulta [Cómo proporcionar acceso a recursos que Cuentas de AWS son propiedad de terceros](#) en la Guía del usuario de IAM.
- Para obtener información sobre cómo proporcionar acceso mediante una federación de identidades, consulte [Proporcionar acceso a usuarios autenticados externamente \(identidad federada\)](#) en la Guía del usuario de IAM.
- Para obtener información sobre la diferencia entre los roles y las políticas basadas en recursos para el acceso entre cuentas, consulte [Cómo los roles de IAM difieren de las políticas basadas en recursos](#) en la Guía del usuario de IAM.

Prevención de la sustitución confusa entre servicios

El problema de la sustitución confusa es un problema de seguridad en el que una entidad que no tiene permiso para realizar una acción puede obligar a una entidad con más privilegios a realizar la

acción. En AWS, la suplantación de identidad entre servicios puede provocar el confuso problema de un diputado. La suplantación entre servicios puede producirse cuando un servicio (el servicio que lleva a cabo las llamadas) llama a otro servicio (el servicio al que se llama). El servicio que lleva a cabo las llamadas se pueden manipular para utilizar sus permisos a fin de actuar en función de los recursos de otro cliente de una manera en la que no debe tener permiso para acceder. Para evitarlo, AWS proporciona herramientas que lo ayudan a proteger sus datos para todos los servicios con entidades principales de servicio a las que se les ha dado acceso a los recursos de su cuenta.

Se recomienda utilizar las claves de contexto de condición global [aws:SourceArn](#) en las políticas de recursos para limitar los permisos que AWS Clean Rooms concede a otro servicio para el recurso. Utiliza `aws:SourceArn` si desea que solo se asocie un recurso al acceso entre servicios.

La forma más eficaz de protegerse contra el problema de la sustitución confusa es utilizar la clave de contexto de condición global de `aws:SourceArn` con el ARN completo del recurso. En AWS Clean Rooms, también tienes que compararla con la clave de condición. `sts:ExternalId`

El valor de `aws:SourceArn` debe definirse con el ARN de pertenencia del rol asumido.

El siguiente ejemplo muestra cómo usar la clave de contexto de condición global `aws:SourceArn` en AWS Clean Rooms para evitar el problema del suplente confuso.

Note

La política de ejemplo se aplica a la política de confianza del rol de servicio que se AWS Clean Rooms utiliza para acceder a los datos y metadatos de una tabla configurada. El valor de `<query-runner-membership-id>` debe establecerse en el ID de membresía del ejecutor de consultas. Todos los miembros de la colaboración pueden ver los metadatos de la tabla configurada, por lo que cada ARN de membresía debe incluirse en la lista de miembros. ARNs

Note

Cuando se crea un rol de servicio a través de la AWS Clean Rooms consola, todos los miembros actuales de la colaboración se incluyen de forma predeterminada en la confusa condición de suplentes. Si va a añadir nuevos miembros a una colaboración que ya tiene tablas configuradas asociadas, asegúrese de actualizar la confusa condición de suplente de su rol de servicio con el ARN de membresía del nuevo miembro.

Si no actualiza la confusa condición de suplente de su rol de servicio después de agregar un nuevo miembro, ese nuevo miembro no podrá acceder a la información AWS Clean Rooms que se recupere utilizando ese rol.

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>"
          ]
        }
      }
    }
  ]
}
```

```
]
}
```

Comportamientos de IAM para AWS Clean Rooms ML

Trabajos entre cuentas

Clean Rooms ML permite que otra Cuenta de AWS persona acceda de forma segura Cuenta de AWS a determinados recursos creados por una persona en su cuenta. Cuando un cliente de Cuenta de AWS A recurre `StartAudienceGenerationJob` a un `ConfiguredAudienceModel` recurso propiedad de Cuenta de AWS B, Clean Rooms ML crea dos ARNs para esa tarea. Un ARN en Cuenta de AWS A y otro en B. Cuenta de AWS ARNs Son idénticos excepto por sus. Cuenta de AWS

Clean Rooms ML crea dos ARNs para el trabajo a fin de garantizar que ambas cuentas puedan aplicar sus propias políticas de IAM a los trabajos. Por ejemplo, ambas cuentas pueden usar el control de acceso basado en etiquetas y aplicar las políticas de su AWS organización. El trabajo procesa los datos de ambas cuentas, de modo que ambas cuentas pueden eliminar el trabajo y sus datos asociados. Ninguna de las dos cuentas puede impedir que la otra elimine el trabajo.

Solo hay una ejecución de trabajo y ambas cuentas pueden ver la tarea cuando llaman a `ListAudienceGenerationJobs`. Ambas cuentas pueden llamar al `GetDelete`, y `Export APIs` estar trabajando usando el ARN con su propia Cuenta de AWS identificación.

Ninguno de los dos Cuenta de AWS puede acceder al trabajo cuando se utiliza un ARN con el otro Cuenta de AWS ID.

El nombre del trabajo debe ser único dentro de una Cuenta de AWS. El nombre en Cuenta de AWS B es `$accountA-$name`. El nombre elegido por Cuenta de AWS A lleva el prefijo Cuenta de AWS A cuando el trabajo se visualiza en B. Cuenta de AWS

Para que una cuenta cruzada `StartAudienceGenerationJob` tenga éxito, Cuenta de AWS B debe permitir esa acción tanto en el nuevo trabajo de B como `ConfiguredAudienceModel` en el de Cuenta de AWS Cuenta de AWS B mediante una política de recursos similar a la del siguiente ejemplo:

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

Esta política de recursos de aprendizaje AWS Clean Rooms automático hace referencia a dos tipos diferentes Cuenta de AWS IDs para facilitar la generación de audiencias entre cuentas:

- 111122223333: esta es la cuenta que contiene al principal (usuario, rol o servicio) autorizado a iniciar trabajos de generación de audiencia. Esta cuenta inicia el flujo de trabajo de procesamiento de aprendizaje automático.
- 444455556666: es la cuenta propietaria de los recursos de aprendizaje AWS Clean Rooms automático (el modelo de audiencia configurado y los trabajos de generación de audiencia). Esta cuenta aloja los modelos de aprendizaje automático y administra la ejecución de los trabajos.

Notas de configuración adicionales:

- ID de la declaración (Sid): CAMA-ID sustitúyalo por el identificador real de la solicitud de modelo de AWS Clean Rooms audiencia (CAMA) para que la declaración de política sea fácilmente identificable.
- Recurso IDs: *id* Sustitúyalo por el ID real del modelo de audiencia configurado y *UUID* por el ID de colaboración específico.
- Condición: la `cleanrooms-ml:CollaborationId` condición garantiza que los trabajos de generación de audiencia solo se puedan iniciar en el contexto de la AWS Clean Rooms colaboración especificada, lo que proporciona un límite de seguridad adicional.

Esta configuración multicuenta permite situaciones en las que una organización administra los modelos y la infraestructura del aprendizaje automático y, al mismo tiempo, permite a los socios autorizados iniciar los procesos de generación de audiencia dentro de los límites de su acuerdo de colaboración.

Si utiliza la [API de aprendizaje automático de AWS Clean Rooms](#) para crear un modelo similar `manageResourcePolicies` configurado con el valor `true`, AWS Clean Rooms crea esta política automáticamente.

Además, la política de identidad de la persona que llama en Cuenta de AWS A necesita `StartAudienceGenerationJob` autorización. `arn:aws:cleanrooms-ml:us-west-1:AccountA:audience-generation-job/*` Por lo tanto, hay tres recursos de acción de `IAMStartAudienceGenerationJob`: el trabajo Cuenta de AWS A, el trabajo Cuenta de AWS B y el Cuenta de AWS trabajo B. `ConfiguredAudienceModel`

Warning

El Cuenta de AWS que inició el trabajo recibe un evento de registro de AWS CloudTrail auditoría sobre el trabajo. La Cuenta de AWS propietaria de `ConfiguredAudienceModel` no recibe ningún evento de registro de auditoría de AWS CloudTrail .

Etiquetado de trabajos

Al establecer el parámetro `childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE` de `CreateConfiguredAudienceModel`, todos los trabajos de generación de segmentos similares de

su cuenta que se creen a partir de ese modelo similar configurado tendrán de forma predeterminada las mismas etiquetas que el modelo similar configurado. El modelo similar configurado es el principal y el trabajo de generación de segmentos similares es el secundario.

Si está creando un trabajo en su propia cuenta, las etiquetas de solicitud del trabajo anulan las etiquetas principales. Los trabajos creados por otras cuentas nunca crean etiquetas en su cuenta. Si establece `childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE` y otra cuenta crea un trabajo, hay dos copias del trabajo. La copia de su cuenta tiene las etiquetas del recurso principal y la copia de la cuenta del remitente del trabajo tiene las etiquetas de la solicitud.

Validación de colaboradores

Al conceder permisos a otros miembros de una AWS Clean Rooms colaboración, la política de recursos debe incluir la clave de condición `cleanrooms-ml:CollaborationId`. Esto exige que el `collaborationId` parámetro esté incluido en la [StartAudienceGenerationJob](#) solicitud. Cuando el parámetro `collaborationId` se incluye en la solicitud, Clean Rooms ML valida que la colaboración existe, que el remitente del trabajo es un miembro activo de la colaboración y que el propietario del modelo similar configurado es un miembro activo de la colaboración.

Cuando AWS Clean Rooms administre la política de recursos del modelo similar configurada (el `manageResourcePolicies` parámetro está `TRUE` en la [CreateConfiguredAudienceModelAssociation](#) solicitud), esta clave de condición se establecerá en la política de recursos. Por lo tanto, debe especificar la entrada `collaborationId`.
[StartAudienceGenerationJob](#)

Acceso entre cuentas

Solo se puede llamar a `StartAudienceGenerationJob` entre cuentas. El resto de Clean Rooms ML solo se APIs pueden utilizar con los recursos de su propia cuenta. Esto garantiza que sus datos de entrenamiento, la configuración del modelo similar y otra información permanezcan privados.

Clean Rooms ML nunca revela Amazon S3 ni AWS Glue las ubicaciones de las cuentas. La ubicación de los datos de entrenamiento, la ubicación de salida del modelo similar configurado y la ubicación inicial del trabajo de generación de segmentos similar nunca están visibles entre cuentas. A menos que el registro de consultas esté habilitado en la colaboración, las cuentas no pueden ver si los datos iniciales provienen de una consulta SQL o la consulta en sí. Si `Get` un trabajo de generación de audiencia enviado por otra cuenta, el servicio no mostrará la ubicación inicial.

Comportamientos de IAM para los modelos personalizados de Clean Rooms ML

Trabajos entre cuentas

Clean Rooms ML permite que otra Cuenta de AWS persona acceda de forma segura a determinados recursos asociados Cuenta de AWS a una colaboración creada por una persona desde su cuenta. Un cliente de Cuenta de AWS A con la capacidad de realizar consultas como miembro puede llamar `CreateTrainedModel` o utilizar `StartTrainedModelInferenceJob` un `ConfiguredModelAlgorithmAssociation` recurso propiedad de otro miembro de la colaboración, siempre que lo permita la `ConfiguredModelAlgorithmAssociation` regla de análisis personalizada creada con ella `CreateConfiguredTableAnalysisRule`. `CreateMLInputChannel`

Además, cualquier miembro activo de una colaboración puede eliminar los datos asociados a un modelo entrenado o a un canal de entrada de aprendizaje automático mediante el comando `DeleteTrainedModelOutput` and `DeleteMLInputChannelData` APIs.

Acceso entre cuentas

Clean Rooms ML permite a los usuarios recuperar metadatos sobre los recursos creados por otras cuentas a través de `GetCollaboration` y `ListCollaboration` APIs. Clean Rooms ML no revela la clave del KMS ARNs, las etiquetas, las variables de entorno ni los hiperparámetros (para la `TrainedModel` acción) a otras cuentas.

Acceso a membresías y colaboraciones

Al acceder a los recursos de membresía y colaboración en el contexto de los modelos personalizados de Clean Rooms ML, la política de identidad del usuario necesita permisos para las acciones `cleanrooms:PassMembership` `cleanrooms:PassCollaboration`, o ambos. Todos los APIs que aceptan `membershipId` necesitan el `cleanrooms:PassMembership` permiso y todos los APIs que aceptan `collaborationId` necesitan el `cleanrooms:PassCollaboration` permiso. Se proporciona un ejemplo de política de identidad para un rol al que se puede llamar `createTrainedModel` en el contexto de un ID de membresía y al que se puede llamar `GetCollaborationTrainedModel` en el contexto de un ID de colaboración.

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"
      ]
    }
  ]
}
```

Validación de conformidad para AWS Clean Rooms

Para saber si uno Servicio de AWS está dentro del ámbito de aplicación de programas de cumplimiento específicos, consulte [Servicios de AWS Alcance por programa de cumplimiento](#) [Servicios de AWS](#) de cumplimiento y elija el programa de cumplimiento que le interese. Para obtener información general, consulte Programas de [AWS cumplimiento > Programas AWS](#).

Puede descargar informes de auditoría de terceros utilizando AWS Artifact. Para obtener más información, consulte [Descarga de informes en AWS Artifact](#).

Su responsabilidad de cumplimiento al Servicios de AWS utilizarlos viene determinada por la confidencialidad de sus datos, los objetivos de cumplimiento de su empresa y las leyes y reglamentos aplicables. Para obtener más información sobre su responsabilidad de conformidad al utilizarlos Servicios de AWS, consulte [AWS la documentación de seguridad](#).

Resiliencia en AWS Clean Rooms

La infraestructura AWS global se basa en AWS regiones y zonas de disponibilidad. Las regiones proporcionan varias zonas de disponibilidad físicamente independientes y aisladas que se encuentran conectadas mediante redes con un alto nivel de rendimiento y redundancia, además de baja demora. Con las zonas de disponibilidad, puede diseñar y utilizar aplicaciones y bases de datos que realizan una conmutación por error automática entre las zonas sin interrupciones. Las zonas de disponibilidad tienen una mayor disponibilidad, tolerancia a errores y escalabilidad que las infraestructuras tradicionales de uno o varios centros de datos.

Para obtener más información sobre AWS las regiones y las zonas de disponibilidad, consulte [Infraestructura AWS global](#).

Seguridad de la infraestructura en AWS Clean Rooms

Como servicio gestionado, AWS Clean Rooms está protegido por la seguridad de la red AWS global. Para obtener información sobre los servicios AWS de seguridad y cómo se AWS protege la infraestructura, consulte [Seguridad AWS en la nube](#). Para diseñar su AWS entorno utilizando las mejores prácticas de seguridad de la infraestructura, consulte [Protección de infraestructuras en un marco](#) de buena AWS arquitectura basado en el pilar de la seguridad.

Utiliza las llamadas a la API AWS publicadas para acceder a AWS Clean Rooms través de la red. Los clientes deben admitir lo siguiente:

- Seguridad de la capa de transporte (TLS). Exigimos TLS 1.2 y recomendamos TLS 1.3.
- Conjuntos de cifrado con confidencialidad directa total (PFS) como DHE (Ephemeral Diffie-Hellman) o ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). La mayoría de los sistemas modernos como Java 7 y posteriores son compatibles con estos modos.

Seguridad de la red

Cuando se AWS Clean Rooms lee desde el bucket de S3 durante la ejecución de la consulta, el tráfico entre Amazon S3 AWS Clean Rooms y Amazon S3 se enruta de forma segura a través de la red AWS privada. El tráfico en vuelo se firma mediante el protocolo Amazon Signature versión 4 (SIGv4) y se cifra mediante HTTPS. Este tráfico se autoriza en función del rol de servicio de IAM que se haya configurado para la tabla configurada.

Puede conectarse mediante programación a AWS Clean Rooms través de un punto final. Para obtener una lista de puntos de conexión de servicio, consulte [Puntos de conexión y cuotas de AWS Clean Rooms](#) en la Referencia general de AWS.

Todos los puntos de conexión de servicio son únicamente HTTPS. Puede utilizar los puntos de conexión de Amazon Virtual Private Cloud (VPC) en caso de que desee conectarse desde AWS Clean Rooms su VPC y no desee tener conectividad a Internet. Para obtener más información, consulte [Acceder a AWS los servicios AWS PrivateLink](#) en la Guía.AWS PrivateLink

Puede asignar políticas de IAM a sus directores de IAM que utilicen las [claves aws: SourceVpce context](#) para restringir a su director de IAM a fin de que solo pueda realizar llamadas a través de AWS Clean Rooms un punto final de VPC y no a través de Internet.

Acceda al aprendizaje AWS Clean Rooms automático mediante AWS Clean Rooms un punto final de interfaz ()AWS PrivateLink

Puede usarlo AWS PrivateLink para crear una conexión privada entre su nube privada virtual (VPC) y ML AWS Clean Rooms . AWS Clean Rooms Puede acceder a nuestro AWS Clean Rooms AWS Clean Rooms ML como si estuviera en su VPC, sin el uso de una puerta de enlace a Internet, un dispositivo NAT, una conexión VPN o Direct Connect una conexión. Las instancias de la VPC no necesitan direcciones IP públicas para acceder a AWS Clean Rooms.

Esta conexión privada se establece mediante la creación de un punto de conexión de interfaz alimentado por AWS PrivateLink. Creamos una interfaz de red de punto de conexión en cada subred

habilitada para el punto de conexión de interfaz. Se trata de interfaces de red administradas por el solicitante que sirven como punto de entrada para el tráfico destinado a AWS Clean Rooms.

Para obtener más información, consulte [Acceso a los Servicios de AWS a través de AWS PrivateLink](#) en la Guía de AWS PrivateLink .

Consideraciones sobre AWS Clean Rooms

Antes de configurar un punto final de interfaz para AWS Clean Rooms, consulte [las consideraciones](#) de la AWS PrivateLink guía.

AWS Clean Rooms y AWS Clean Rooms ML permiten realizar llamadas a todas sus acciones de API a través del punto final de la interfaz.

Las políticas de puntos de conexión de VPC no son compatibles con el ML AWS Clean Rooms . AWS Clean Rooms De forma predeterminada, se permite el acceso total AWS Clean Rooms y el AWS Clean Rooms aprendizaje automático a través del punto final de la interfaz. Como alternativa, puede asociar un grupo de seguridad a las interfaces de red de los puntos finales para controlar el tráfico hacia AWS Clean Rooms o el aprendizaje AWS Clean Rooms automático a través del punto final de la interfaz.

Cree un punto final de interfaz para AWS Clean Rooms

Puede crear un punto de enlace de interfaz para AWS Clean Rooms el AWS Clean Rooms aprendizaje automático mediante la consola de Amazon VPC o el AWS Command Line Interface (AWS CLI). Para obtener más información, consulte [Creación de un punto de conexión de interfaz](#) en la Guía de AWS PrivateLink .

Cree un punto final de interfaz para AWS Clean Rooms usar el siguiente nombre de servicio.

```
com.amazonaws.region.cleanrooms
```

Cree un punto final de interfaz para AWS Clean Rooms ML con el siguiente nombre de servicio.

```
com.amazonaws.region.cleanrooms-ml
```

Si habilita DNS privado para el punto de conexión de interfaz, puede realizar solicitudes a la API para AWS Clean Rooms usando su nombre de DNS predeterminado para la región. Por ejemplo, `cleanrooms-ml.us-east-1.amazonaws.com`.

Monitorización AWS Clean Rooms

La supervisión es una parte importante del mantenimiento de la confiabilidad, la disponibilidad y el rendimiento de AWS Clean Rooms AWS las demás soluciones. AWS proporciona las siguientes herramientas de monitoreo para observar AWS Clean Rooms, informar cuando algo anda mal y tomar medidas automáticas cuando sea apropiado:

- Amazon CloudWatch Logs le permite supervisar, almacenar y acceder a sus archivos de registro desde instancias de Amazon EC2 y otras fuentes. AWS CloudTrail Amazon CloudWatch Logs puede supervisar la información de los archivos de registro y notificarle cuando se alcancen determinados umbrales. También se pueden archivar los datos del registro en un almacenamiento de larga duración. Para obtener más información, consulta la [Guía del usuario CloudWatch de Amazon Logs](#).

Clean Rooms ML permite realizar trabajos entre cuentas para determinadas acciones de la API. La Cuenta de AWS persona que inició el trabajo recibe el evento del registro de AWS CloudTrail auditoría del trabajo. Para obtener más información, consulte [Comportamientos de IAM para AWS Clean Rooms ML](#)

- AWS CloudTrail captura las llamadas a la API y los eventos relacionados realizados por usted o en su nombre Cuenta de AWS y entrega los archivos de registro a un bucket de Amazon S3 que especifique. Puede identificar qué usuarios y cuentas llamaron AWS, la dirección IP de origen desde la que se realizaron las llamadas y cuándo se produjeron. Para obtener más información, consulte la [Guía del usuario de AWS CloudTrail](#).
- Amazon EventBridge es un servicio de bus de eventos sin servidor que facilita la conexión de sus aplicaciones con datos de diversas fuentes. EventBridge ofrece un flujo de datos en tiempo real desde sus propias aplicaciones, aplicaciones Software-as-a-Service (SaaS) y AWS servicios, y dirige esos datos a destinos como Lambda. Esto le permite monitorear los eventos que ocurren en los servicios y crear arquitecturas basadas en eventos. Para obtener más información, consulta la [Guía del EventBridge usuario de Amazon](#) y la [referencia de Amazon EventBridge Events](#).
- Administración de facturación y costos de AWS es un servicio que ofrece funciones que le ayudan a pagar sus facturas y a optimizar sus costes. Amazon Web Services factura a tu cuenta en función del uso, lo que garantiza que solo pagues por lo que utilices. AWS también le permite aplicar [etiquetas de asignación de costos](#) a sus recursos para realizar un seguimiento y categorizar los costos de AWS. Por ejemplo, en AWS Clean Rooms, puede aplicar etiquetas a las membresías de una colaboración para realizar un seguimiento de los costos por colaboración. Para obtener más información, consulte la [Guía del usuario de AWS Billing](#).

Inicio de sesión de Analysis AWS Clean Rooms

El registro de análisis es una función de AWS Clean Rooms. Al [crear una colaboración](#) y activar el registro de análisis, los miembros pueden almacenar los registros relevantes de las consultas o los registros de los trabajos en Amazon CloudWatch Logs.

Con los registros de consultas y los registros de trabajos, los miembros pueden determinar si las consultas cumplen con las reglas de análisis y se ajustan al acuerdo de colaboración. Además, los registros de consultas respaldan las auditorías.

Cuando la opción de registro de análisis está activada en la AWS Clean Rooms consola, los registros de consultas incluyen lo siguiente:

- `analysisRule`: la regla de análisis de la tabla configurada.
- `analysisTemplateArn`: la plantilla de análisis que se ejecutó (aparece en función de la regla de análisis).
- `collaborationId`: el identificador único de la colaboración en la que se ejecutó la consulta.
- `configuredTableID`: el identificador único de la tabla configurada a la que hace referencia la consulta.
- `directQueryAnalysisRulePolicy.custom.allowedAnalysis`: la plantilla de análisis que se puede ejecutar en una tabla configurada (aparece según la regla de análisis).
- `directQueryAnalysisRulePolicy.v1.custom.allowedAnalysisProviders`: los proveedores de consultas que tienen permitido crear consultas (aparece según la regla de análisis).
- `errorCode`: el código de error que se produce cuando una consulta no se ejecuta correctamente.
- `errorMessage`: el mensaje de error que se produce cuando una consulta no se ejecuta correctamente.
- `eventID`: el identificador único de ejecución de la consulta. Desde el 31 de agosto de 2023, el identificador único coincidirá con el `protectedQueryID`.
- `eventTimestamp`: el momento de ejecución de la consulta.
- `parameters.parametervalue`: los valores de los parámetros (aparecen en función del texto de la consulta).
- `queryText`: la definición SQL de la ejecución de la consulta. Si hay parámetros, se etiquetan como `:parametervalue`.

- `queryValidationErrors`: los errores de consulta durante la validación de la consulta.
- `schemaName`: el nombre de la asociación de tablas configuradas a la que se hace referencia en la consulta.
- `status`: el estado de la ejecución de la consulta.

Reciba registros de consultas y trabajos

No es necesario realizar ninguna acción aparte de AWS Clean Rooms configurar los registros de consultas y trabajos. AWS Clean Rooms crea grupos de registro para las colaboraciones después de que cada miembro de la colaboración [cree una membresía](#).

Los miembros que puedan realizar consultas, los miembros que puedan ejecutar consultas y trabajos, los miembros que puedan recibir resultados y los miembros cuyas tablas de configuración estén referenciadas en la consulta recibirán un registro de consultas o un registro de trabajos.

El miembro que puede realizar consultas y el miembro que puede recibir los resultados recibirán registros de consultas para cada tabla configurada a la que se haga referencia en la consulta. Si no son el propietario de la tabla configurada, no podrán ver el ID de la tabla configurada (`configuredTableID`).

El miembro que pueda ejecutar consultas y trabajos y el miembro que pueda recibir los resultados recibirán los registros de trabajos de cada tabla configurada a la que se haga referencia en el trabajo. Si no son el propietario de la tabla configurada, no podrán ver el ID de la tabla configurada (`configuredTableID`).

Si un miembro tiene varias asociaciones de tablas configuradas a las que se hace referencia en la consulta, recibirá un registro de consultas por cada tabla configurada.

Si un miembro tiene varias asociaciones de tablas configuradas a las que se hace referencia en el trabajo, recibirá un registro de trabajos para cada tabla configurada.

Se crean registros para las consultas que contienen SQL admitido y no admitido en AWS Clean Rooms. Para obtener más información, consulte [Referencia de SQL de AWS Clean Rooms](#).

Los registros también se crean cuando las consultas o los trabajos hacen referencia a tablas configuradas que no están asociadas a la colaboración.

Los registros pueden contener información sobre un SQL incorrecto.

Los registros de consultas y trabajos indican el estado de una consulta, pero no indican si el resultado de la consulta se entregó. Confirman que el miembro que puede realizar la consulta ha enviado una consulta o un trabajo. Los registros de consultas también confirman que la consulta contiene SQL compatible AWS Clean Rooms y hace referencia a las tablas configuradas asociadas a la colaboración.

Example

Por ejemplo, no se genera un registro si la consulta se canceló después de AWS Clean Rooms validar su conformidad con las reglas de análisis y durante el procesamiento de la consulta.

Si elimina el grupo de registros, debe volver a crearlo manualmente con el mismo nombre de grupo de registros (ID de colaboración de la colaboración). O bien, puede activar y desactivar el registro en su pertenencia.

Para obtener más información sobre cómo activar el registro de análisis, consulte [Crear una colaboración](#).

Para obtener más información sobre Amazon CloudWatch Logs, consulta la [Guía del usuario de Amazon CloudWatch Logs](#).

Acciones recomendadas para los registros de consultas y trabajos

Es aconsejable que los miembros tomen las siguientes medidas de forma periódica:

- Para comprobar que las consultas y los trabajos coinciden con los casos de uso o las consultas que se acordaron para la colaboración, revise las consultas y los trabajos que se ejecutan en la colaboración.

Para obtener más información sobre cómo ver consultas recientes, consulte [Visualización de consultas recientes](#).

Para obtener más información sobre cómo ver los trabajos recientes, consulte [Ver trabajos recientes](#).

- Revisar las columnas de la tabla configurada que se utilizan en las reglas de análisis de los miembros de la colaboración y en las consultas para asegurarse de que las columnas de la tabla configurada se ajustan a lo acordado para la colaboración.

Para obtener más información sobre cómo ver las columnas configuradas, consulte [Ver tablas y reglas de análisis](#).

Monitorización detallada CloudWatch con AWS Clean Rooms

La supervisión detallada es una función opcional de AWS Clean Rooms. Al [crear una colaboración](#) y activar la supervisión detallada, los miembros que pueden realizar consultas y configurar los pagadores pueden optar por recibir métricas de observabilidad detalladas en sus cuentas. CloudWatch

Con una supervisión detallada, los miembros pueden supervisar el rendimiento de las consultas, realizar un seguimiento del uso de los recursos y crear paneles de control para obtener información operativa. Estas métricas ayudan a planificar la capacidad, analizar los costes y solucionar problemas de rendimiento.

Note

La disponibilidad de las métricas varía en función de las fuentes de datos de las tablas configuradas:

- Para los análisis que hacen referencia a tablas configuradas con Athena como fuente de datos, las QueryFileScanned métricas BytesRead y siempre devuelven 0.
- En el caso de los análisis que hacen referencia a tablas configuradas con Snowflake como fuente de datos BytesReadQueryFileScanned, las RecordsRead métricas y las métricas siempre devuelven 0.
- Para los análisis que hacen referencia a varias tablas configuradas con diferentes fuentes de datos (por ejemplo, tablas respaldadas por Snowflake y S3) BytesRead y QueryFileScanned devuelven métricas solo para AWS Glue tablas respaldadas por S3. AWS Glue RecordsRead devuelve las métricas de las AWS Glue tablas respaldadas por S3 y las tablas de Athena.

Tipos de métricas

AWS Clean Rooms publica dos categorías de métricas para: CloudWatch

Métricas de monitoreo detalladas agregadas

Estas métricas proporcionan una visibilidad global de todos los análisis:

AnalysisRuntime

Muestra la duración total de un análisis en milisegundos.

CRPUConsumed

Las CRPU (unidades de procesamiento de sala limpia) consumidas para un análisis en particular.

[Para obtener más información sobre el cálculo de la CRPU, consulte AWS Clean Rooms Precios.](#)

ConcurrentVCPUs

vCPU simultánea utilizada en un momento dado (en relación con las cuotas de servicio).

AnalysesCount

Número de análisis ejecutados durante un período de tiempo para una dimensión específica.

ConcurrentQueries

Recuento de consultas simultáneas en un momento dado (relacionado con las cuotas de servicio).

Métricas de supervisión detalladas a nivel de consulta

Estas métricas proporcionan información detallada sobre la ejecución de consultas individuales:

BytesRead

Datos totales (en bytes) leídos de las tablas de origen para un análisis.

BytesWritten

Datos totales (en bytes) escritos para un análisis.

RecordsRead

Número total de registros leídos de las tablas de origen para un análisis.

RecordsWritten

Número total de registros devueltos en el resultado del análisis.

QueryFileScanned

Número total de archivos de datos escaneados para un análisis.

Memory utilization

Recurso de memoria utilizado para un análisis.

vCPU utilization

Recurso de vCPU utilizado para un análisis.

Disk utilization

Recurso de almacenamiento utilizado para un análisis.

Dimensiones de la métrica

Las métricas se organizan según las siguientes dimensiones:

- Estado del análisis: completado
- Tipo de análisis: SparkSQL
- Membresía: el identificador de membresía
- Consulta: ID de consulta individual (para métricas a nivel de consulta)
- Cuenta o región: tuya y región Cuenta de AWS

Todas las métricas se publican en el espacio de AWS/CleanRooms nombres de. CloudWatch

¿Quién puede acceder a las métricas

Las métricas de supervisión detalladas están disponibles para:

- Ejecutores de consultas: miembros que pueden ejecutar consultas en la colaboración
- Pagadores configurados: miembros designados para pagar los costos de ejecución de las consultas

Los miembros que solo aportan datos no pueden acceder a las métricas de monitoreo detalladas.

Precios

AWS Clean Rooms no cobra por publicar métricas en CloudWatch. Sin embargo, se aplican CloudWatch cargos estándar por el almacenamiento de métricas, el uso del panel de control y las alarmas. Para obtener más información, consulta los [CloudWatch precios de Amazon](#).

Registro de llamadas a la AWS Clean Rooms API mediante AWS CloudTrail

AWS Clean Rooms está integrado con AWS CloudTrail un servicio que proporciona un registro de las acciones realizadas por un usuario, un rol o un Servicio de AWS usuario AWS Clean Rooms. CloudTrail captura todas las llamadas a la API AWS Clean Rooms como eventos. Las llamadas capturadas incluyen llamadas desde la AWS Clean Rooms consola y llamadas en código a las operaciones de la AWS Clean Rooms API. Si crea una ruta, puede habilitar la entrega continua de CloudTrail eventos a un bucket de Amazon S3, incluidos los eventos para AWS Clean Rooms. Si no configura una ruta, podrá ver los eventos más recientes en la CloudTrail consola, en el historial de eventos. Con la información recopilada por usted CloudTrail, puede determinar a AWS Clean Rooms qué dirección IP se realizó la solicitud, quién la realizó, cuándo se realizó y detalles adicionales.

Para obtener más información CloudTrail, consulte la [Guía AWS CloudTrail del usuario](#).

AWS Clean Rooms información en CloudTrail

CloudTrail está habilitada en tu cuenta Cuenta de AWS al crear la cuenta. Cuando se produce una actividad en AWS Clean Rooms, esa actividad se registra en un CloudTrail evento junto con otros Servicio de AWS eventos del historial de eventos. Puede ver, buscar y descargar eventos recientes en su Cuenta de AWS. Para obtener más información, consulte [Visualización de eventos con el historial de CloudTrail eventos](#).

Para tener un registro continuo de tus eventos Cuenta de AWS, incluidos los eventos para AWS Clean Rooms ti, crea una ruta. Un rastro permite CloudTrail entregar archivos de registro a un bucket de Amazon S3. De forma predeterminada, cuando se crea un registro de seguimiento en la consola, el registro de seguimiento se aplica a todas las Regiones de AWS. La ruta registra los eventos de todas las regiones de la AWS partición y envía los archivos de registro al bucket de Amazon S3 que especifique. Además, puede configurar otros Servicios de AWS para que analicen más a fondo los datos de eventos recopilados en los CloudTrail registros y actúen en función de ellos. Para más información, consulte los siguientes temas:

- [Introducción a la creación de registros de seguimiento](#)
- [CloudTrail servicios e integraciones compatibles](#)
- [Configuración de las notificaciones de Amazon SNS para CloudTrail](#)
- [Recibir archivos de CloudTrail registro de varias regiones](#)
- [Recibir archivos de CloudTrail registro de varias cuentas](#)

Todas AWS Clean Rooms las acciones se registran CloudTrail y se documentan en la [Referencia de la AWS Clean Rooms API](#).

Cada entrada de registro o evento contiene información sobre quién generó la solicitud. La información de identidad del usuario le ayuda a determinar lo siguiente:

- Si la solicitud se realizó con las credenciales del usuario raíz o del usuario de IAM.
- Si la solicitud se realizó con credenciales de seguridad temporales de un rol o fue un usuario federado.
- Si la solicitud la realizó otro Servicio de AWS.

Para obtener más información, consulte el [elemento `userIdentity` de CloudTrail](#).

Descripción de las entradas de los archivos de AWS Clean Rooms registro

Un rastro es una configuración que permite la entrega de eventos como archivos de registro a un bucket de Amazon S3 que usted especifique. CloudTrail Los archivos de registro contienen una o más entradas de registro. Un evento representa una solicitud única de cualquier fuente e incluye información sobre la acción solicitada, la fecha y la hora de la acción, los parámetros de la solicitud, etc. CloudTrail Los archivos de registro no son un registro ordenado de las llamadas a la API pública, por lo que no aparecen en ningún orden específico.

Ejemplos de AWS Clean Rooms CloudTrail eventos

En los siguientes ejemplos se muestran CloudTrail los eventos de:

Temas

- [StartProtectedQuery \(correcto\)](#)
- [StartProtectedQuery \(error\)](#)

StartProtectedQuery (correcto)

```
{
  "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 (error)

```

{
    "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) [identifier] 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"
}

```

Integración AWS Clean Rooms en aplicaciones basadas en eventos mediante Amazon EventBridge

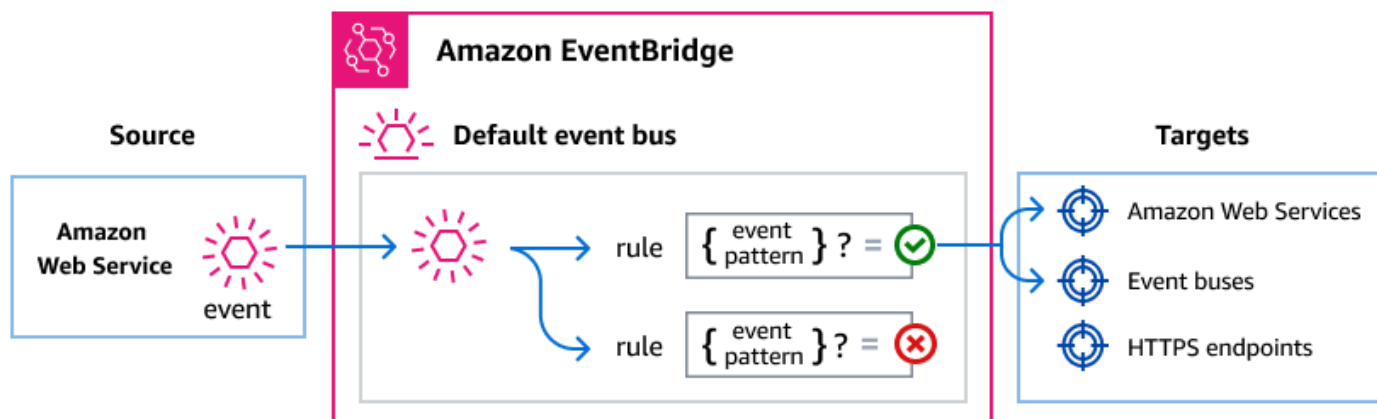
Puede AWS Clean Rooms incorporarlas a aplicaciones basadas en eventos (EDAs) que utilizan los eventos que se producen AWS Clean Rooms para comunicarse entre los componentes de la aplicación e iniciar procesos posteriores. Para ello, utiliza Amazon EventBridge para enrutar eventos desde AWS Clean Rooms otros componentes de software. Amazon EventBridge es un servicio sin servidor que utiliza eventos para conectar los componentes de la aplicación entre sí, lo que facilita la integración de AWS servicios, por ejemplo, en arquitecturas basadas AWS Clean Rooms en eventos sin código ni operaciones adicionales.

La arquitectura basada en eventos es un estilo de creación de sistemas de software de acoplamiento flexible que funcionan juntos emitiendo eventos y respondiendo a ellos. En este modelo, un evento representa un cambio en un recurso o entorno.

Así es como EventBridge funciona con AWS Clean Rooms:

Como ocurre con muchos AWS servicios, AWS Clean Rooms genera y envía eventos al bus de eventos EventBridge predeterminado. Un bus de eventos es un enrutador que recibe eventos y los envía a destinos u objetivos, que se especifican. Los objetivos pueden incluir otros AWS servicios, aplicaciones personalizadas y aplicaciones de socios de SaaS.

EventBridge enruta los eventos de acuerdo con las reglas que usted crea en el bus de eventos. Para cada regla, se debe especificar un filtro o patrón de eventos para seleccionar solo los eventos que desee. Cada vez que se envía un evento al bus de eventos, lo EventBridge compara con cada regla. Si el evento coincide con la regla, lo EventBridge enruta a los objetivos especificados.



Por ejemplo, supongamos que quiere saber cada vez que se crea una nueva AWS Clean Rooms colaboración en su cuenta. Puede crear una regla en el bus de eventos predeterminado. En la regla, crearía un patrón de eventos que especificara los eventos de AWS Clean Rooms denominados **Collaboration Created**. Cada vez que EventBridge recibía un evento que coincidiera con esas propiedades, lo redireccionaba al flujo de trabajo especificado.

AWS Clean Rooms eventos

AWS los servicios pueden enviar eventos directamente al bus de eventos EventBridge predeterminado. Además, también AWS CloudTrail envía eventos que se originan en EventBridge numerosos AWS servicios. Estos eventos pueden incluir llamadas a la API, inicios de sesión y acciones en la consola, eventos de servicio e CloudTrail información valiosa. Para obtener más información, consulte [los eventos AWS de servicio que se AWS CloudTrail](#) ofrecen en la Guía del EventBridge usuario.

Para obtener una lista completa de AWS Clean Rooms los eventos enviados EventBridge, consulte el AWS Clean Rooms tema de la [Referencia de EventBridge eventos](#).

Tipo de detalle del evento	Description (Descripción)
Plantilla de análisis creada	El propietario de la plantilla de análisis y todos los miembros activos de la colaboración reciben una notificación cuando se crea una plantilla de análisis.
Plantilla de análisis actualizada	El propietario de la plantilla de análisis y todos los miembros activos de la colaboración que tengan acceso a la actualización reciben una notificación cuando se actualiza una plantilla de análisis.
Plantilla de análisis eliminada	El propietario de la plantilla de análisis y todos los miembros activos de la colaboración reciben una notificación cuando se elimina una plantilla de análisis.
Colaboración creada	El propietario de la colaboración recibe una notificación cuando se crea una colaboración.
Colaboración actualizada	El propietario de la colaboración y todos los miembros activos de la colaboración que tienen acceso a la actualización reciben una notificación cuando se actualiza una colaboración.

Tipo de detalle del evento	Description (Descripción)
<u>Se ha creado una solicitud de cambio de colaboración</u>	El propietario de la colaboración y todos los miembros activos de la colaboración reciben una notificación cuando se crea una solicitud de cambio de colaboración.
<u>Solicitud de cambio de colaboración aprobada</u>	El propietario de la colaboración y todos los miembros activos de la colaboración reciben una notificación cuando se aprueba una solicitud de cambio de colaboración.
<u>Solicitud de cambio de colaboración cancelada</u>	Cuando se cancela una solicitud de cambio de colaboración, se notifica al propietario de la colaboración y a todos los miembros activos de la colaboración.
<u>Solicitud de cambio de colaboración confirmada</u>	El propietario de la colaboración y todos los miembros activos de la colaboración reciben una notificación cuando se confirma una solicitud de cambio de colaboración.
<u>Se ha creado una asociación de tablas configuradas</u>	El propietario de la asociación de tablas configuradas y todos los miembros activos de la colaboración reciben una notificación cuando se crea una asociación de tablas configuradas.
<u>Se ha actualizado la asociación de tablas configuradas</u>	El propietario de la asociación de tablas configuradas y todos los miembros activos de la colaboración que tienen acceso a la actualización reciben una notificación cuando se actualiza una asociación de tablas configuradas.
<u>Se ha eliminado la asociación de tablas configuradas</u>	Cuando se elimina una asociación de tablas configuradas, se notifica al propietario de la asociación de tablas configuradas y a todos los miembros activos de la colaboración.
<u>Se ha creado una regla de análisis de asociación de tablas configuradas</u>	El propietario de la regla de análisis de asociación de tablas configurada y todos los miembros activos de la colaboración reciben una notificación cuando se crea una regla de análisis de asociación de tablas configurada.

Tipo de detalle del evento	Description (Descripción)
<u>Se actualiza la regla de análisis de asociación de tablas configuradas</u>	Cuando se actualiza una regla de análisis de asociación de tablas configurada, se notifica al propietario de la regla de análisis de asociación de tablas configurada y a todos los miembros activos de la colaboración que tienen acceso a la actualización.
<u>Se ha eliminado la regla de análisis de asociación de tablas configurada</u>	Cuando se elimina una regla de análisis de asociación de tablas configurada, se notifica al propietario de la regla de análisis de asociación de tablas configurada y a todos los miembros activos de la colaboración.
<u>Se creó la tabla de mapeo de ID</u>	El propietario de la tabla de mapeo de identificadores y todos los miembros activos de la colaboración reciben una notificación cuando se crea una tabla de mapeo de identificadores.
<u>Se actualiza la tabla de mapeo de ID</u>	El propietario de la tabla de mapeo de identificadores y todos los miembros activos de la colaboración que tengan acceso a la actualización reciben una notificación cuando se actualiza una tabla de mapeo de identificadores.
<u>Se eliminó la tabla de mapeo de ID</u>	El propietario de la tabla de mapeo de identificadores y todos los miembros activos de la colaboración reciben una notificación cuando se elimina una tabla de mapeo de identificadores.
<u>Se ha creado la asociación de espacios de nombres de Id</u>	El propietario de la asociación de espacios de nombres de ID y todos los miembros activos de la colaboración reciben una notificación cuando se crea una asociación de espacios de nombres de ID.
<u>Se ha actualizado la asociación de espacios de nombres de ID</u>	El propietario de la Asociación de Espacios de Nombres de ID y todos los miembros activos de la colaboración que tengan acceso a la actualización reciben una notificación cuando se actualiza una asociación de espacios de nombres de ID.

Tipo de detalle del evento	Description (Descripción)
<u>Se ha eliminado la asociación de espacios de nombres de Id.</u>	El propietario de la asociación de espacios de nombres de ID y todos los miembros activos de la colaboración reciben una notificación cuando se elimina una asociación de espacios de nombres de ID.
<u>Invitado a la colaboración</u>	El miembro invitado recibe una notificación cuando es invitado a una colaboración.
<u>Membresía creada</u>	El propietario de la membresía y todos los miembros activos de la colaboración reciben una notificación cuando se crea una membresía.
<u>Membresía actualizada</u>	El propietario de la membresía recibe una notificación cuando se actualiza una membresía, a menos que la membresía se haya eliminado de la colaboración, en cuyo caso se notificará a todos los miembros activos de la colaboración.
<u>Membresía eliminada</u>	El propietario de la membresía y todos los miembros activos de la colaboración reciben una notificación cuando se elimina una membresía.
<u>Job protegido enviado</u>	El candidato, el pagador del trabajo y el receptor de resultados de un trabajo protegido reciben una notificación cuando se envía el trabajo protegido.
<u>Trabajo protegido iniciado</u>	El Job Runner, el pagador del trabajo y el receptor de resultados de un trabajo protegido reciben una notificación cuando se inicia el trabajo protegido.
<u>Cancelación de trabajo protegida</u>	El Job Runner, el pagador del trabajo y el receptor de resultados de un trabajo protegido reciben una notificación cuando el trabajo protegido se cancela.
<u>Trabajo protegido cancelado</u>	El Job Runner, el pagador del trabajo y el receptor de resultados de un trabajo protegido reciben una notificación cuando se cancela el trabajo protegido.

Tipo de detalle del evento	Description (Descripción)
El trabajo protegido se realizó correctamente	El Job Runner, el pagador del trabajo y el receptor de resultados de un trabajo protegido reciben una notificación cuando el trabajo protegido ha tenido éxito.
Error en el trabajo protegido	El Job Runner, el pagador del trabajo y el receptor de resultados de un trabajo protegido reciben una notificación cuando el trabajo protegido ha fallado.
Consulta protegida enviada	El ejecutor de la consulta, el pagador de la consulta y el receptor o receptores de los resultados de una consulta protegida reciben una notificación cuando se envía la consulta protegida.
Se inició la consulta protegida	El ejecutor de consultas, el pagador de consultas y los receptores de resultados de una consulta protegida reciben una notificación cuando se inicia la consulta protegida.
Cancelación de consultas protegida	El ejecutor de consultas, el pagador de consultas y el receptor de resultados de una consulta protegida reciben una notificación cuando la consulta protegida se cancela.
Consulta protegida cancelada	El ejecutor de consultas, el pagador de consultas y los receptores de resultados de una consulta protegida reciben una notificación cuando se cancela la consulta protegida.
La consulta protegida se realizó correctamente	El ejecutor de consultas, el pagador de consultas y el receptor de resultados de una consulta protegida reciben una notificación cuando la consulta protegida se ha realizado correctamente.
Error en la consulta protegida	El ejecutor de consultas, el pagador de consultas y los receptores de resultados de una consulta protegida reciben una notificación cuando la consulta protegida ha fallado.
Se agotó el tiempo de espera de la consulta protegida	El ejecutor de consultas, el pagador de consultas y los receptores de resultados de una consulta protegida reciben una notificación cuando se agota el tiempo de espera de la consulta protegida.

Enrutar eventos mediante AWS Clean Rooms EventBridge

Para que AWS Clean Rooms los eventos de EventBridge ruta se dirijan a un destino, debe crear una regla. Cada regla contiene un patrón de eventos que EventBridge coincide con cada evento recibido en el bus de eventos. Si los datos del evento coinciden con el patrón de eventos especificado, EventBridge enruta ese evento a los objetivos de la regla.

Para obtener instrucciones detalladas sobre cómo crear reglas de bus de eventos, consulte [Creación de reglas que reaccionan a eventos](#) en la Guía del usuario de EventBridge .

Crear patrones de eventos que coincidan con AWS Clean Rooms los eventos

Cada patrón de eventos es un objeto JSON que contiene:

- (Opcional): `source` atributo que identifica el servicio que envía el evento. En el AWS Clean Rooms caso de los eventos, la fuente es `aws.cleanrooms`.
- (Opcional): un atributo `detail-type` que contiene una matriz de los nombres de eventos que deben coincidir.
- (Opcional): un atributo `detail` que contiene cualquier otro dato de evento con el que coincidir.

Por ejemplo, el siguiente patrón de eventos coincide con todos los eventos de `Id Namespace Association Updated` de los que se eliminó la colaboración: AWS Clean Rooms

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"],
  "detail": {
    "status": ["COLLABORATION_DELETED"]
  }
}
```

Para obtener más información sobre la escritura de los patrones de eventos, consulte [Patrones de eventos](#) en la Guía del usuario de EventBridge .

AWS Clean Rooms referencia detallada de los eventos

Todos los eventos de los AWS servicios tienen un conjunto común de campos que contienen metadatos sobre el evento, como el AWS servicio que es el origen del evento, la hora en que se

generó el evento, la cuenta y la región en las que tuvo lugar el evento, etc. Para ver las definiciones de estos campos generales, consulte [Estructura de eventos en Amazon EventBridge Events Reference](#).

Además, cada evento tiene un campo `detail` que contiene datos específicos de ese evento en particular. La siguiente referencia define los campos de detalle para los distintos eventos de AWS Clean Rooms .

Cuando se utilice EventBridge para seleccionar y gestionar AWS Clean Rooms eventos, es útil tener en cuenta lo siguiente:

- El `source` campo para todos los eventos de AWS Clean Rooms está establecido en `aws.cleanrooms`.
- El campo `detail-type` especifica el tipo de evento.

Por ejemplo, `Collaboration Created`.

- El campo `detail` contiene los datos específicos de ese evento en particular.

Para obtener información sobre cómo crear patrones de eventos que permitan que las reglas coincidan con AWS Clean Rooms los eventos, consulte [Patrones de eventos](#) en la Guía del EventBridge usuario de Amazon.

Plantilla de análisis: evento creado

A continuación, se muestran los campos de detalle del evento `Analysis Template Created`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Created"]
}
```

Evento actualizado de la plantilla de análisis

A continuación, se muestran los campos de detalle del evento `Analysis Template Updated`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Updated"]
}
```

Evento eliminado de la plantilla de análisis

A continuación, se muestran los campos de detalle del evento `Analysis Template Deleted`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Analysis Template Deleted"]
}
```

Colaboración: evento creado

A continuación, se muestran los campos de detalle del evento `Collaboration Created`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Created"]
}
```

Evento de colaboración actualizado

A continuación, se muestran los campos de detalle del evento `Collaboration Updated`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Updated"]
}
```

Solicitud de cambio de colaboración: evento creado

A continuación, se muestran los campos de detalle del evento `Collaboration Change Request Created`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Created"]
}
```

Evento aprobado por la solicitud de cambio de colaboración

A continuación, se muestran los campos de detalle del evento `Collaboration Change Request Approved`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Approved"]
}
```

Evento cancelado por solicitud de cambio de colaboración

A continuación, se muestran los campos de detalle del evento Collaboration Change Request Cancelled.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Cancelled"]
}
```

Evento comprometido con la solicitud de cambio de colaboración

A continuación, se muestran los campos de detalle del evento Collaboration Change Request Committed.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Collaboration Change Request Committed"]
}
```

Asociación de tablas configuradas: evento creado

A continuación, se muestran los campos de detalle del evento Configured Table Association Created.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Created"]
}
```

Asociación de tablas configuradas: evento actualizado

A continuación, se muestran los campos de detalle del evento Configured Table Association Updated.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Updated"]
}
```

Evento eliminado de la asociación de tablas configuradas

A continuación, se muestran los campos de detalle del evento Configured Table Association Deleted.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Deleted"]
}
```

Evento creado por la regla de análisis de asociación de tablas configurada

A continuación, se muestran los campos de detalle del evento Configured Table Association Analysis Rule Created.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Created"]
}
```

Evento actualizado de la regla de análisis de asociación de tablas configuradas

A continuación, se muestran los campos de detalle del evento Configured Table Association Analysis Rule Updated.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Updated"]
}
```

Evento eliminado de la regla de análisis de asociación de tablas configurada

A continuación, se muestran los campos de detalle del evento Configured Table Association Analysis Rule Deleted.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Configured Table Association Analysis Rule Deleted"]
}
```

Evento creado por Id Mapping Table

A continuación, se muestran los campos de detalle del evento Id Mapping Table Created.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Created"]
}
```

Evento actualizado de Id Mapping Table

A continuación, se muestran los campos de detalle del evento Id Mapping Table Updated.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Updated"]
}
```

Evento eliminado de Id Mapping Table

A continuación, se muestran los campos de detalle del evento Id Mapping Table Deleted.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Mapping Table Deleted"]
}
```

Evento creado por Id Namespace Association

A continuación, se muestran los campos de detalle del evento Id Namespace Association Created.

```
{
  "source": ["aws.cleanrooms"],
```

```
"detail-type": ["Id Namespace Association Created"]
}
```

Evento actualizado de Id Namespace Association

A continuación, se muestran los campos de detalle del evento Id Namespace Association Updated.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"]
}
```

Note

Cuando se elimina una colaboración, el evento se captura en el evento Id Namespace Association Updated del campo. `detail.status`

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Updated"],
  "detail": {
    "status": ["COLLABORATION_DELETED"]
  }
}
```

Id: Namespace Association: evento eliminado

A continuación, se muestran los campos de detalle del evento Id Namespace Association Deleted.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Id Namespace Association Deleted"]
}
```

Invitado al evento de colaboración

A continuación, se muestran los campos de detalle del evento Invited To Collaboration.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Invited To Collaboration"]
}
```

Evento creado por miembros

A continuación, se muestran los campos de detalle del evento `Membership Created`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Created"]
}
```

Evento actualizado de membresía

A continuación, se muestran los campos de detalle del evento `Membership Updated`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Updated"]
}
```

Evento de membresía eliminada

A continuación, se muestran los campos de detalle del evento `Membership Deleted`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Membership Deleted"]
}
```

Evento enviado por Protected Job

A continuación, se muestran los campos de detalle del evento `Protected Job Submitted`.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Submitted"]
}
```


Evento Protected Job Started

A continuación, se muestran los campos de detalle del evento Protected Job Started.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Started"]
}
```

Evento de cancelación de Protected Job

A continuación, se muestran los campos de detalle del evento Protected Job Cancelling.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Cancelling"]
}
```

Evento Protected Job Cancelado

A continuación, se muestran los campos de detalle del evento Protected Job Cancelled.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Cancelled"]
}
```

Evento Protected Job Succeed

A continuación, se muestran los campos de detalle del evento Protected Job Succeeded.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Succeeded"]
}
```

Evento Protected Job Failed

A continuación, se muestran los campos de detalle del evento Protected Job Failed.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Job Failed"]
}
```

Evento enviado por consulta protegida

A continuación, se muestran los campos de detalle del evento Protected Query Submitted.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Submitted"]
}
```

Evento iniciado en Protected Query

A continuación, se muestran los campos de detalle del evento Protected Query Started.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Started"]
}
```

Evento de cancelación de consultas protegidas

A continuación, se muestran los campos de detalle del evento Protected Query Cancelling.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Cancelling"]
}
```

Evento cancelado de Protected Query

A continuación, se muestran los campos de detalle del evento Protected Query Cancelled.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Cancelled"]
}
```

```
}
```

Evento de consulta protegida realizado correctamente

A continuación, se muestran los campos de detalle del evento Protected Query Succeeded.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Succeeded"]
}
```

Evento de consulta protegida fallida

A continuación, se muestran los campos de detalle del evento Protected Query Failed.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Failed"]
}
```

Se agotó el tiempo de espera de Protected Query

A continuación, se muestran los campos de detalle del evento Protected Query Timed Out.

```
{
  "source": ["aws.cleanrooms"],
  "detail-type": ["Protected Query Timed Out"]
}
```

Etiquetado de asignación de costos

AWS Clean Rooms admite el uso de etiquetas de asignación de costos para realizar un seguimiento de sus AWS costos. Estas etiquetas se activan en el Administración de facturación y costos de AWS panel de control. AWS utiliza las etiquetas para clasificar los costes y entregarle un informe mensual de asignación de costes. Las etiquetas de asignación de costos definidas por el usuario se pueden aplicar a AWS Clean Rooms los recursos para ayudar a rastrear y asignar los costos en toda la colaboración.

Recursos etiquetables para la asignación de costos

En la siguiente tabla se enumeran los AWS Clean Rooms recursos que generan cargos y se pueden etiquetar con etiquetas de asignación de costos definidas por el usuario para realizar un seguimiento y organizar los costos.

Recurso facturado	Recurso etiquetado
Consulta de SQL	Pertenencia
PySpark Job	Pertenencia
Job de formación en ML	Pertenencia
Job de inferencia de ML	Pertenencia
Generación de datos sintéticos	Pertenencia
Entrenamiento con modelos parecidos	Modelo parecido
Exportación de segmentos parecidos	Segmento parecido

Creación de AWS Clean Rooms recursos con AWS CloudFormation

AWS Clean Rooms está integrado con AWS CloudFormation un servicio que le ayuda a modelar y configurar sus AWS recursos. Como resultado de esta integración, puede dedicar menos tiempo a crear y administrar sus recursos e infraestructura. Usted crea una plantilla que describe todos los AWS recursos que desea y los CloudFormation aprovisiona y configura automáticamente. Algunos ejemplos de recursos son las colaboraciones, las tablas configuradas, las asociaciones de tablas configuradas y las pertenencias.

Cuando la utilice CloudFormation, podrá reutilizar la plantilla para configurar los AWS Clean Rooms recursos de forma coherente y repetida. Describa sus recursos una vez y, a continuación, aprovisiona los mismos recursos una y otra vez en varios Cuentas de AWS y Regiones de AWS.

AWS Clean Rooms y CloudFormation plantillas

Para aprovisionar y configurar los recursos AWS Clean Rooms y los servicios relacionados, debe conocer [CloudFormation las plantillas](#). Las plantillas son archivos de texto con formato JSON o YAML. Estas plantillas describen los recursos que desea aprovisionar en sus CloudFormation pilas. Si no estás familiarizado con JSON o YAML, puedes usar CloudFormation Designer para ayudarte a empezar con CloudFormation las plantillas. Para obtener más información, consulte [¿Qué es Designer de CloudFormation ?](#) en la Guía del usuario de AWS CloudFormation .

AWS Clean Rooms permite crear colaboraciones, tablas configuradas, asociaciones de tablas configuradas y membresías en ellas. CloudFormationPara obtener más información, incluidos ejemplos de plantillas JSON y YAML para colaboraciones, tablas configuradas, asociaciones de tablas configuradas y membresías, consulte las referencias a los tipos de recursos de aprendizaje [automático de AWS Clean RoomsAWS Clean Rooms](#) en la Guía del AWS CloudFormation usuario.

Obtenga más información sobre CloudFormation

Para obtener más información CloudFormation, consulte los siguientes recursos:

- [AWS CloudFormation](#)
- [AWS CloudFormation Guía del usuario](#)
- [CloudFormation Referencia de la API](#)

- [Guía del usuario de la interfaz de la línea de comandos de AWS CloudFormation](#)

Cuotas para AWS Clean Rooms

Tienes Cuenta de AWS cuotas predeterminadas, antes denominadas límites, para cada una de ellas Servicio de AWS. A menos que se indique lo contrario, cada cuota es específica de un Región de AWS. Puede solicitar aumentos para algunas cuotas, pero otras no se pueden aumentar.

Para ver las cuotas AWS Clean Rooms, abra la [consola Service Quotas](#). En el panel de navegación, elija Servicios de AWS y seleccione AWS Clean Rooms.

Para solicitar un aumento de cuota, consulte [Solicitud de aumento de cuota](#) en la Guía del usuario de Service Quotas. Si la cuota no está todavía disponible en Service Quotas, utilice el [Formulario de aumento del límite de servicio](#).

Temas

- [AWS Clean Rooms cuotas](#)
- [Cuotas de aprendizaje automático de AWS Clean Rooms](#)

AWS Clean Rooms cuotas

Cuenta de AWS Tiene las siguientes cuotas relacionadas con. AWS Clean Rooms

Name	Predeterminado	Ajuste	Description (Descripción)
Tamaño de regla de análisis	Cada región admitida: 100 kilobytes	No	Tamaño máximo de JSON para una regla de análisis
Plantillas de análisis por membresía	Cada región admitida: 25	Sí	Número máximo de plantillas de análisis por membresía
Colaboraciones creadas por cuenta	Cada región admitida: 10	Sí	Número máximo de colaboraciones creadas por cuenta
Lista de columnas permitidas por tabla configurada	Cada región admitida: 100	Sí	Número máximo de columnas que se pueden

Name	Predeterminado	Ajuste	Description (Descripción)
			incluir en la lista permitida por tabla configurada
Uso de vCPU en PySpark tareas simultáneas por cuenta	us-east-1:1.024 Cada una de las demás regiones compatibles: 512	Sí	Uso máximo total de vCPU de todos los trabajos que se ejecutan PySpark simultáneamente por cuenta
Trabajos simultáneos PySpark por cuenta	us-east-1: 5 Cada una de las demás regiones compatibles: 2	Sí	Número máximo de PySpark trabajos en ejecución simultánea por cuenta
Consultas SQL simultáneas por cuenta	us-east-1: 5 Cada una de las demás regiones compatibles: 2	Sí	Número máximo de modelos que se ejecutan simultáneamente por cuenta.
Uso simultáneo de vCPU de consultas SQL por cuenta	Cada región admitida: 512	Sí	Uso máximo total de vCPU de todas las consultas SQL que se ejecutan simultáneamente por cuenta
Trabajo simultáneo en curso por membresía	Cada región admitida: 1	No	Número máximo de consultas simultáneas en curso por pertenencia
Consultas simultáneas en curso por membresía	Cada región admitida: 5	No	Número máximo de consultas simultáneas en curso por membresía

Name	Predeterminado	Ajuste	Description (Descripción)
Asociaciones de modelos similares (modelo de audiencia) configuradas por pertenencia	Cada región admitida: 5	Sí	Número máximo de asociaciones de modelos similares configuradas por pertenencia.
Tablas configuradas por cuenta	Cada región admitida: 250	Sí	Número máximo de tablas configuradas creadas por cuenta
Tablas configuradas por consulta protegida	Cada región admitida: 15	Sí	Número máximo de tablas configuradas en una consulta protegida
Tablas de asignación de ID por suscripción	Cada región admitida: 5	Sí	Número máximo de tablas de asignación de ID por suscripción.
Asociaciones de espacios de nombres de ID por suscripción	Cada región admitida: 10	Sí	Número máximo de asociaciones de espacio de nombres de ID por suscripción.
Miembros invitados por colaboración	Cada región admitida: 5	Sí	Número máximo de miembros invitados por colaboración
Membresías por cuenta	Cada región admitida: 100	Sí	Número máximo de membresías por cuenta
Plantillas de presupuestos de privacidad para acceder a los presupuestos por colaboración	Cada región admitida: 25	Sí	Número máximo de plantillas de presupuesto de privacidad para acceder a los presupuestos por colaboración

Name	Predeterminado	Ajuste	Description (Descripción)
Longitud del texto de la consulta (utilizando la privacidad diferencial)	Cada región admitida: 8 kilobytes	No	Longitud máxima del texto para una instrucción de consulta SQL con privacidad diferencial
Longitud del texto de consulta en el motor de análisis SQL de Clean Rooms	Cada región compatible: 90 kilobytes	No	Longitud máxima del texto para una instrucción de consulta SQL en el motor de análisis SQL de Clean Rooms
Longitud del texto de consulta en el motor de análisis Spark	Cada región admitida: 500 kilobytes	No	Longitud máxima del texto de una instrucción de consulta SQL
Tasa de solicitudes BatchGetCollaborationAnalysisTemplate	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la BatchGetCollaborationAnalysisTemplate API por segundo
Tasa de BatchGetSchema solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la BatchGetSchema API por segundo
Tasa de BatchGetSchemaAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la BatchGetSchemaAnalysisRule API por segundo
Tasa de CreateAnalysisTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateAnalysisTemplate API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de CreateCollaboration solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateCollaboration API por segundo
Tasa de CreateCollaborationChangeRequest solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateCollaborationChangeRequest API por segundo
Tasa de CreateConfiguredAudienceModelAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateConfiguredAudienceModelAssociation API por segundo
Tasa de CreateConfiguredTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateConfiguredTable API por segundo
Tasa de CreateConfiguredTableAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateConfiguredTableAnalysisRule API por segundo
Tasa de CreateConfiguredTableAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateConfiguredTableAssociation API por segundo
Tasa de CreateConfiguredTableAssociationAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateConfiguredTableAssociationAnalysisRule API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de CreateIdMappingTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateIdMappingTable API por segundo
Tasa de CreateIdNamespaceAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateIdNamespaceAssociation API por segundo
Tasa de CreateMembership solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreateMembership API por segundo
Tasa de CreatePrivacyBudgetTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la CreatePrivacyBudgetTemplate API por segundo
Tasa de DeleteAnalysisTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteAnalysisTemplate API por segundo
Tasa de DeleteCollaboration solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteCollaboration API por segundo
Tasa de DeleteConfiguredAudienceModelAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteConfiguredAudienceModelAssociation API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de DeleteConfiguredTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteConfiguredTable API por segundo
Tasa de DeleteConfiguredTableAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteConfiguredTableAnalysisRule API por segundo
Tasa de DeleteConfiguredTableAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteConfiguredTableAssociation API por segundo
Tasa de DeleteConfiguredTableAssociationAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteConfiguredTableAssociationAnalysisRule API por segundo
Tasa de DeleteIdMappingTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteIdMappingTable API por segundo
Tasa de DeleteIdNamespaceAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteIdNamespaceAssociation API por segundo
Tasa de DeleteMember solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteMember API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de DeleteMembership solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeleteMembership API por segundo
Tasa de DeletePrivacyBudgetTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la DeletePrivacyBudgetTemplate API por segundo
Tasa de GetAnalysisTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetAnalysisTemplate API por segundo
Tasa de GetCollaboration solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetCollaboration API por segundo
Tasa de GetCollaborationAnalysisTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetCollaborationAnalysisTemplate API por segundo
Tasa de GetCollaborationChangeRequest solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetCollaborationChangeRequest API por segundo
Tasa de GetCollaborationConfiguredAudienceModelAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetCollaborationConfiguredAudienceModelAssociation API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de GetCollaborationIdNamespace Association solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetCollaborationIdNamespace Association API por segundo
Tasa de GetCollaborationPrivacyBudgetTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetCollaborationPrivacyBudgetTemplate API por segundo
Tasa de GetConfiguredAudienceModelAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetConfiguredAudienceModelAssociation API por segundo
Tasa de GetConfiguredTable solicitudes	Cada región admitida: 20 por segundo	Sí	Número máximo de llamadas a la GetConfiguredTable API por segundo
Tasa de GetConfiguredTableAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetConfiguredTableAnalysisRule API por segundo
Tasa de GetConfiguredTableAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetConfiguredTableAssociation API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de GetConfiguredTableAssociationAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetConfiguredTableAssociationAnalysisRule API por segundo
Tasa de GetIdMappingTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetIdMappingTable API por segundo
Tasa de GetIdNamespaceAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetIdNamespaceAssociation API por segundo
Tasa de GetMembership solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetMembership API por segundo
Tasa de GetPrivacyBudgetTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetPrivacyBudgetTemplate API por segundo
Tasa de GetProtectedJob solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetProtectedJob API por segundo
Tasa de GetProtectedQuery solicitudes	Cada región admitida: 20 por segundo	Sí	Número máximo de llamadas a la GetProtectedQuery API por segundo
Tasa de GetSchema solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetSchema API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de GetSchemaAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la GetSchemaAnalysisRule API por segundo
Tasa de ListAnalysisTemplates solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListAnalysisTemplates API por segundo
Tasa de ListCollaborationAnalysisTemplates solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborationAnalysisTemplates API por segundo
Tasa de ListCollaborationChangeRequests solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborationChangeRequests API por segundo
Tasa de ListCollaborationConfiguredAudienceModelAssociations solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborationConfiguredAudienceModelAssociations API por segundo
Tasa de ListCollaborationIdNamespaceAssociations solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborationIdNamespaceAssociations API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de ListCollaborationPrivacyBudgetTemplates solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborationPrivacyBudgetTemplates API por segundo
Tasa de ListCollaborationPrivacyBudgets solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborationPrivacyBudgets API por segundo
Tasa de ListCollaborations solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListCollaborations API por segundo
Tasa de ListConfiguredAudienceModelAssociations solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListConfiguredAudienceModelAssociations API por segundo
Tasa de ListConfiguredTableAssociations solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListConfiguredTableAssociations API por segundo
Tasa de ListConfiguredTables solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListConfiguredTables API por segundo
Tasa de ListIdMappingTables solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListIdMappingTables API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de ListIdNamespaceAssociations solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListIdNamespaceAssociations API por segundo
Tasa de ListMembers solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListMembers API por segundo
Tasa de ListMemberships solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListMemberships API por segundo
Tasa de ListPrivacyBudgetTemplates solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListPrivacyBudgetTemplates API por segundo
Tasa de ListPrivacyBudgets solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListPrivacyBudgets API por segundo
Tasa de ListProtectedJobs solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListProtectedJobs API por segundo
Tasa de ListProtectedQueries solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListProtectedQueries API por segundo
Tasa de ListSchemas solicitudes	Cada región admitida: 5	Sí	Número máximo de llamadas a la ListSchemas API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de ListTagsForResource solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la ListTagsForResource API por segundo
Tasa de PopulateIdMappingTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la PopulateIdMappingTable API por segundo
Tasa de PreviewPrivacyImpact solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la PreviewPrivacyImpact API por segundo
Tasa de StartProtectedJob solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la StartProtectedJob API por segundo
Tasa de StartProtectedQuery solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la StartProtectedQuery API por segundo
Tasa de TagResource solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la TagResource API por segundo
Tasa de UntagResource solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UntagResource API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de UpdateAnalysisTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateAnalysisTemplate API por segundo
Tasa de UpdateCollaboration solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateCollaboration API por segundo
Tasa de UpdateCollaborationChangeRequest solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateCollaborationChangeRequest API por segundo
Tasa de UpdateConfiguredAudienceModelAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateConfiguredAudienceModelAssociation API por segundo
Tasa de UpdateConfiguredTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateConfiguredTable API por segundo
Tasa de UpdateConfiguredTableAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateConfiguredTableAnalysisRule API por segundo
Tasa de UpdateConfiguredTableAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateConfiguredTableAssociation API por segundo

Name	Predeterminado	Ajuste	Description (Descripción)
Tasa de UpdateConfiguredTableAssociationAnalysisRule solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateConfiguredTableAssociationAnalysisRule API por segundo
Tasa de UpdateIdMappingTable solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateIdMappingTable API por segundo
Tasa de UpdateIdNamespaceAssociation solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateIdNamespaceAssociation API por segundo
Tasa de UpdatePrivacyBudgetTemplate solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdatePrivacyBudgetTemplate API por segundo
Tasa de UpdateProtectedJob solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateProtectedJob API por segundo
Tasa de UpdateProtectedQuery solicitudes	Cada región admitida: 5 por segundo	Sí	Número máximo de llamadas a la UpdateProtectedQuery API por segundo
Asociaciones de tabla por membresía	Cada región admitida: 25	Sí	Número máximo de asociaciones de tabla por membresía

AWS Clean Rooms límites de parámetros de recursos

Recurso	Predeterminado	Description (Descripción)
Longitud del texto de consulta	90 KB	Longitud máxima del texto de una instrucción de consulta SQL
Longitud del texto de la consulta (utilizando la privacidad diferencial)	90 KB	Longitud máxima del texto para una instrucción de consulta SQL con privacidad diferencial
Tiempo de ejecución de consultas	24 horas	Tiempo máximo que se ejecuta una consulta antes de que se agote el tiempo de espera

Cuotas de aprendizaje automático de AWS Clean Rooms

Cuenta de AWS Tiene las cuotas predeterminadas, anteriormente denominadas límites, para AWS Clean Rooms ML.

Para ver las cuotas de servicio de AWS Clean Rooms ML, realice una de las siguientes acciones:

- Siga los pasos que se indican en [Visualización de cuotas](#) de servicio en la Guía del usuario de Service Quotas y seleccione AWS Clean Rooms ML como servicio.
- Consulte las [cuotas de servicio de aprendizaje automático de AWS Clean Rooms](#) en Referencia general de Amazon Web Services.

Para mantener el rendimiento del servicio y garantizar un uso adecuado del aprendizaje automático de AWS Clean Rooms, es posible que las cuotas predeterminadas asignadas a una cuenta se actualicen en función de los factores regionales, el historial de pagos, el uso fraudulento y la and/or aprobación de una solicitud de aumento de cuota.

Name	Predeterminado	Ajuste	Description (Descripción)
Trabajos de exportación de audiencia activa por trabajo de generación de audiencia	Cada región admitida: 25	No	El número máximo de trabajos de exportación de audiencia activa para un trabajo de generación de audiencia
Asociaciones activas de algoritmos de modelos configurados por membresía	Cada región admitida: 1000	Sí	El número máximo de asociaciones de algoritmos del modelo configurados activas por membresía
Algoritmos de modelo configurados activos por membresía	Cada región admitida: 1000	Sí	El número máximo de algoritmos del modelo configurados activos por membresía
Canales de entrada de modelos personalizados activos por membresía	Cada región admitida: 100	Sí	El número máximo de canales de entrada del modelo personalizado activos por membresía
Número máximo de instancias por trabajo de entrenamiento	Cada región admitida: 10	Sí	El número máximo de instancias de entrenamiento activas que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de versiones de modelos entrenadas para el active/pending/in progreso por modelo entrenado	Cada región admitida: 100	Sí	El número máximo de versiones de modelos entrenados según el active/pending/in progreso que se

Name	Predeterminado	Ajuste	Description (Descripción)
			pueden crear por modelo entrenado.
Número máximo de columnas de entrada para generar datos sintéticos	Cada región admitida: 1000	No	El número máximo de columnas de entrada para generar datos sintéticos
Número máximo de filas de entrada para generar datos sintéticos	Cada región compatible: 2 500 000	No	El número máximo de filas de entrada para generar datos sintéticos
Número máximo de instancias de entrenamiento ml.c4.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c4.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c4.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c4.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c4.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c4.8xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.c4.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c4.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5.18xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5.18xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5.9xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.c5.9xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.c5.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5n.18xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5n.18xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5n.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5n.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5n.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5n.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c5n.9xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5n.9xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.c5n.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c5n.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c6i.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c6i.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c6i.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c6i.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c6i.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c6i.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c6i.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c6i.2xlarge que se pueden utilizar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.c6i.32xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.c6i.32xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c6i.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c6i.4xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.c6i.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.c6i.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c6i.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c6i.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.12xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.c7i.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.48xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.4xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.c7i.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.large que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.c7i.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.c7i.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g4dn.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.g4dn.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g4dn.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.g4dn.16xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.g4dn.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.g4dn.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g4dn.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.g4dn.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g4dn.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.g4dn.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g4dn.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g4dn.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g5.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.12xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.g5.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g5.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g5.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g5.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.48xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g5.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.4xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.g5.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.g5.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g5.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.24xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.g6.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.48xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6.xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.g6e.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6e.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6e.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6e.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6e.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.48xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.g6e.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6e.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.g6e.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.g6e.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m4.10xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m4.10xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m4.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m4.16xlarge que se pueden utilizar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.m4.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m4.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m4.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m4.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m4.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m4.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m5.12xlarge por cuenta	Cada región admitida: 3	Sí	El número máximo de instancias de entrenamiento de ml.m5.12xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m5.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m5.24xlarge que se pueden utilizar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.m5.2xlarge por cuenta	Cada región admitida: 3	Sí	El número máximo de instancias de entrenamiento de ml.m5.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m5.4xlarge por cuenta	Cada región admitida: 3	Sí	El número máximo de instancias de entrenamiento de ml.m5.4xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m5.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m5.large que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m5.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m5.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.m6i.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.12xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.m6i.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.16xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m6i.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m6i.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.2xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m6i.32xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.32xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m6i.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.4xlarge que se pueden utilizar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.m6i.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.8xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m6i.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.large que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m6i.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m6i.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.16xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.m7i.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.2xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.48xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.4xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.8xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.m7i.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.large que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.m7i.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.m7i.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.p2.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p2.16xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.p2.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p2.8xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.p2.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p2.xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.p4d.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p4d.24xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.p4de.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p4de.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.p5.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p5.48xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.p5en.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.p5en.48xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.r5.12xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.r5.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.r5.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r5.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5.2x large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r5.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5.4xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.r5.8xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.r5.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5.large que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5d.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5d.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5d.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.r5d.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5d.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5d.24xlarge que se pueden utilizar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.r5d.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5d.2xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5d.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5d.4xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.r5d.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento ml.r5d.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r5d.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5d.large que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r5d.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r5d.xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.r7i.12xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.12xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.16xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.16xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.24xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.24xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.48xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.48xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento de ml.r7i.4xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.4xlarge que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.8xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.8xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.large que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento de ml.r7i.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.r7i.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.t3.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.t3.2xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.t3.large por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.t3.large que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.t3.medium por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.t3.medium que se pueden utilizar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.t3.xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.t3.xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.trn1.2xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.trn1.2xlarge que se pueden usar por cuenta para crear modelos entrenados.
Número máximo de instancias de entrenamiento ml.trn1.32xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.trn1.32xlarge que se pueden usar por cuenta para crear modelos entrenados.

Name	Predeterminado	Ajuste	Description (Descripción)
Número máximo de instancias de entrenamiento ml.trn1n.32xlarge por cuenta	Cada región admitida: 0	Sí	El número máximo de instancias de entrenamiento de ml.trn1n.32xlarge que se pueden usar por cuenta para crear modelos entrenados.
Valores únicos máximos en la columna de predicción categórica	Cada región admitida: 100	No	Los valores únicos máximos en la columna de predicción categórica
Número mínimo de columnas de entrada necesarias para generar datos sintéticos	Cada región admitida: 5	No	El número mínimo de columnas de entrada necesario para generar datos sintéticos
Número mínimo de filas de entrada necesario para generar datos sintéticos	Cada región admitida: 1500	No	El número mínimo de filas de entrada necesario para generar datos sintéticos
Los trabajos de exportación de audiencia pendientes o en curso por cliente	Cada región admitida: 20	No	El número máximo de trabajos de exportación de audiencia pendientes o en curso por cliente
Los trabajos de generación de audiencia pendientes o en curso por cliente	Cada región admitida: 10	Sí	El número máximo de trabajos de generación de audiencia pendientes o en curso por cliente
Los modelos de audiencia pendientes o en curso por cliente	Cada región admitida: 2	Sí	El número máximo de trabajos de entrenamiento del modelo de audiencia pendientes o en curso por cliente

Name	Predeterminado	Ajuste	Description (Descripción)
Los trabajos de inferencia de modelos personalizados pendientes o en curso por cuenta	Cada región admitida: 10	Sí	El número máximo de trabajos de entrenamiento del modelo de audiencia pendientes o en curso por cliente
Los trabajos de inferencia de modelos personalizados pendientes o en curso por membresía	Cada región admitida: 5	Sí	El número máximo de trabajos de inferencia de modelos personalizados pendientes o en curso por membresía
Los trabajos de entrenamiento de modelos de personalización pendientes o en curso por cuenta	Cada región admitida: 10	Sí	El número máximo de trabajos de entrenamiento del modelo de audiencia pendientes o en curso por cliente
Los trabajos de entrenamiento de modelos de personalización pendientes o en curso por membresía	Cada región admitida: 5	Sí	El número máximo de trabajos de entrenamiento del modelo de audiencia pendientes o en curso por cliente
Empleos de generación de MLIC sintéticos pendientes o en curso por cuenta	Cada región admitida: 2	Sí	El número máximo de empleos de generación de MLIC sintéticos pendientes o en curso por cuenta

AWS Clean Rooms ML tiene cuotas adicionales que se indican a continuación

Cuotas de Clean Rooms ML

Recurso	Predeterminado	Description (Descripción)
Duración máxima de un trabajo de inferencia de modelos personalizados	25 horas	
Número máximo de interacciones	20 mil millones	Número máximo de interacciones permitidas en los datos de entrenamiento. Las entradas más grandes se muestrean.
Número mínimo de interacciones	1 millón	
Número máximo de usuarios distintos para el entrenamiento de modelos similares	100 millones	Si se incluyen más, solo se utilizan los primeros 100 millones, ordenados por número de interacciones.
Número mínimo de usuarios distintos para el entrenamiento de modelos similares	100 000	
Número mínimo de usuarios para el trabajo de segmentos similares (audiencia) de exportación	10 000	
Número máximo de elementos distintos que se utilizan para el entrenamiento de modelos.	1 millón	Puede incluir hasta 50 millones de artículos, pero solo se utiliza el 1 millón más popular.
Número máximo de columnas de características en el	10	

Recurso	Predeterminado	Description (Descripción)
conjunto de datos de entrenamiento.		
Número mínimo de elementos distintos por usuario	2	AWS Clean Rooms ML requiere que cada fila o usuario tenga dos o más elementos, incluidos los elementos repetidos.
Tamaño máximo de la audiencia inicial	500.000	
Tamaño mínimo de la audiencia inicial	500	El proveedor de datos de entrenamiento puede establecer este valor tan bajo como 25.
APIs	por cliente	
Número total de conjuntos de datos de entrenamiento activos.	500	
Número total de modelos similares activos (modelos de audiencia)	500	
Número total de modelos similares configurados y activos (modelos de audiencia)	10 000	
Número total de trabajos de generación de segmentos similares (audiencia) completados	Sin límite	

Recurso	Predeterminado	Description (Descripción)
Número total de trabajos de segmentos similares de exportación (audiencia) completados	Sin límite	
Duración máxima de un trabajo de generación de modelos similares (modelo de audiencia)	1 día (24 horas)	
Duración máxima de un trabajo de generación de segmentos similares (audiencia)	10 horas	Tras proporcionar un origen, Clean Rooms ML tarda un máximo de 10 horas en generar un segmento similar. Si utiliza una consulta de SQL como dato inicial, la consulta puede tardar hasta 12 horas en ejecutarse, además de las 10 horas que se tardan en generar el segmento similar.
Porcentaje mínimo para el tamaño de la papelera de un segmento (audiencia)	1%	
Porcentaje máximo para el tamaño de la papelera de un segmento (audiencia)	20%	
Tamaño absoluto mínimo para el tamaño de la papelera de un segmento (audiencia)	El 1 % del número de usuarios distintos	
Tamaño absoluto máximo para el tamaño de la papelera de un segmento (audiencia)	El 20 % del número de usuarios distintos	

Limpia las cuotas de la API Clean Rooms ML

Cuenta de AWS Tiene las siguientes cuotas de transacciones de API por segundo (TPS) por cuenta y punto final.

Recurso	Límite de frecuencia	Description (Descripción)
Tasa de solicitudes de <code>CreateAudienceModel</code>	Tasa de 1 TPS, ráfaga de 3 TPS	Número máximo de llamadas a la API <code>CreateAudienceModel</code> por segundo
Tasa de solicitudes <code>CreateConfiguredAudienceModel</code>	10 TPS	Número máximo de llamadas a la API <code>CreateConfiguredAudienceModel</code> por segundo
Tasa de solicitudes <code>CreateTrainingDataset</code>	10 TPS	Número máximo de llamadas a la API <code>CreateTrainingDataset</code> por segundo
Tasa de solicitudes <code>DeleteAudienceGenerationJob</code>	Tasa de 2 TPS, ráfaga de 10 TPS	Número máximo de llamadas a la API <code>DeleteAudienceGenerationJob</code> por segundo
Tasa de solicitudes <code>DeleteAudienceModel</code>	Tasa de 2 TPS, ráfaga de 10 TPS	Número máximo de llamadas a la API <code>DeleteAudienceModel</code> por segundo
Tasa de solicitudes <code>DeleteConfiguredAudienceModel</code>	10 TPS	Número máximo de llamadas a la API <code>DeleteConfiguredAudienceModel</code> por segundo
Tasa de solicitudes <code>DeleteConfiguredAudienceModelPolicy</code>	25 TPS	Número máximo de llamadas a la API <code>DeleteConfiguredAudienceModelPolicy</code>

Recurso	Límite de frecuencia	Description (Descripción)
		figuredAudienceModelPolicy por segundo
Tasa de solicitudes DeleteTrainingDataset	10 TPS	Número máximo de llamadas a la API DeleteTrainingDataset por segundo
Tasa de solicitudes GetAudienceGenerationJob	50 TPS	Número máximo de llamadas a la API GetAudienceGenerationJob por segundo
Tasa de solicitudes GetAudienceModel	50 TPS	Número máximo de llamadas a la API GetAudienceModel por segundo
Tasa de solicitudes GetConfiguredAudienceModel	50 TPS	Número máximo de llamadas a la API GetConfiguredAudienceModel por segundo
Tasa de solicitudes GetConfiguredAudienceModelPolicy	50 TPS	Número máximo de llamadas a la API GetConfiguredAudienceModelPolicy por segundo
Tasa de solicitudes GetTrainingDataset	50 TPS	Número máximo de llamadas a la API GetTrainingDataset por segundo
Tasa de solicitudes ListAudienceExportJobs	50 TPS	Número máximo de llamadas a la API ListAudienceExportJobs por segundo

Recurso	Límite de frecuencia	Description (Descripción)
Tasa de solicitudes <code>ListAudienceGenerationJobs</code>	50 TPS	Número máximo de llamadas a la API <code>ListAudienceGenerationJobs</code> por segundo
Tasa de solicitudes <code>ListAudienceModels</code>	50 TPS	Número máximo de llamadas a la API <code>ListAudienceModels</code> por segundo
Tasa de solicitudes <code>ListConfiguredAudienceModels</code>	50 TPS	Número máximo de llamadas a la API <code>ListConfiguredAudienceModels</code> por segundo
Tasa de solicitudes <code>ListTagsForResource</code>	50 TPS	Número máximo de llamadas a la API <code>ListTagsForResource</code> por segundo
Tasa de solicitudes <code>ListTrainingDatasets</code>	50 TPS	Número máximo de llamadas a la API <code>ListTrainingDatasets</code> por segundo
Tasa de solicitudes <code>PutConfiguredAudienceModelPolicy</code>	25 TPS	Número máximo de llamadas a la API <code>PutConfiguredAudienceModelPolicy</code> por segundo
Tasa de solicitudes <code>StartAudienceExportJob</code>	Tasa de 1 TPS, ráfaga de 3 TPS	Número máximo de llamadas a la API <code>StartAudienceExportJob</code> por segundo
Tasa de solicitudes <code>StartAudienceGenerationJob</code>	Tasa de 1 TPS, ráfaga de 5 TPS	Número máximo de llamadas a la API <code>StartAudienceGenerationJob</code> por segundo

Recurso	Límite de frecuencia	Description (Descripción)
Tasa de solicitudes TagResource	10 TPS	Número máximo de llamadas a la API TagResource por segundo
Tasa de solicitudes UntagResource	50 TPS	Número máximo de llamadas a la API UntagResource por segundo
Tasa de solicitudes UpdateConfiguredAudienceModel	10 TPS	Número máximo de llamadas a la API UpdateConfiguredAudienceModel por segundo
Tasa de solicitudes CreateConfiguredModelAlgorithm	10 TPS	Número máximo de llamadas a la CreateConfiguredModelAlgorithm API por segundo.
Tasa de solicitudes CreateConfiguredModelAlgorithmAssociation	10 TPS	Número máximo de llamadas a la CreateConfiguredModelAlgorithmAssociation API por segundo.
Tasa de solicitudes PutMLConfiguration	10 TPS	Número máximo de llamadas a la PutMLConfiguration API por segundo.
Tasa de solicitudes CreateTrainedModel	Tasa de 1 TPS, ráfaga de 3 TPS	Número máximo de llamadas a la CreateTrainedModel API por segundo.
Tasa de solicitudes StartTrainedModelExportJob	10 TPS	Número máximo de llamadas a la StartTrainedModelExportJob API por segundo.

Recurso	Límite de frecuencia	Description (Descripción)
Tasa de solicitudes StartTrainedModelInferenceJob	Tasa de 1 TPS, tasa de 3 TPS	Número máximo de llamadas a la StartTrainedModelInferenceJob API por segundo.
Tasa de GetConfiguredModelAlgorithm solicitud	50 TPS	Número máximo de llamadas a la GetConfiguredModelAlgorithm API por segundo.
Tasa de GetConfiguredModelAlgorithm Association solicitud	50 TPS	Número máximo de llamadas a la GetConfiguredModelAlgorithmAssociation API por segundo.
Tasa de solicitudes GetTrainedModel	50 TPS	Número máximo de llamadas a la GetTrainedModel API por segundo.
Tasa de solicitudes GetMLConfiguration	50 TPS	Número máximo de llamadas a la GetMLConfiguration API por segundo.
Tasa de solicitudes GetTrainedModelInferenceJob	50 TPS	Número máximo de llamadas a la GetTrainedModelInferenceJob API por segundo.
Tasa de solicitudes ListConfiguredModelAlgorithm	50 TPS	Número máximo de llamadas a la ListConfiguredModelAlgorithm API por segundo.

Recurso	Límite de frecuencia	Description (Descripción)
Tasa de solicitudes ListConfiguredModelAlgorithmAssociations	50 TPS	Número máximo de llamadas a la ListConfiguredModelAlgorithmAssociations API por segundo.
Tasa de solicitudes ListTrainedModels	50 TPS	Número máximo de llamadas a la ListTrainedModels API por segundo.
Tasa de solicitudes ListCollaborationTrainedModelExportJobs	50 TPS	Número máximo de llamadas a la ListCollaborationTrainedModelExportJobs API por segundo.
Tasa de solicitudes ListCollaborationTrainedModelInferenceJobs	50 TPS	Número máximo de llamadas a la ListCollaborationTrainedModelInferenceJobs API por segundo.
Tasa de solicitudes DeleteConfiguredModelAlgorithm	Tasa de 2 TPS, ráfaga de 10 TPS	Número máximo de llamadas a la DeleteConfiguredModelAlgorithm API por segundo.
Tasa de solicitudes DeleteConfiguredModelAlgorithmAssociation	Tasa de 2 TPS, ráfaga de 10 TPS	Número máximo de solicitudes de DeleteConfiguredModelAlgorithmAssociation API por segundo.
Tasa de solicitudes DeleteMLConfiguration	Tasa de 2 TPS, ráfaga de 10 TPS	Número máximo de solicitudes de DeleteMLConfiguration API por segundo.
Tasa de solicitudes DeleteTrainedModelOutput	Tasa de 2 TPS, ráfaga de 10 TPS	Número máximo de solicitudes de DeleteTrainedModelOutput API por segundo.

Historial de documentos de la Guía AWS Clean Rooms del usuario

En la siguiente tabla se describen las versiones de la documentación de AWS Clean Rooms.

Para obtener notificaciones sobre las actualizaciones de esta documentación, puede suscribirse a la fuente RSS. Para suscribirse a las actualizaciones RSS, debe tener un complemento de RSS habilitado para el navegador que esté utilizando.

Cambio	Descripción	Fecha
Documentación actualizada para la sección de etiquetado de la asignación de costos	Se agregó documentación Administración de facturación y costos de AWS y se creó una sección para los recursos etiquetables relacionados con la asignación de costos.	20 de febrero de 2026
Support para la configuración cleanrooms-ml spark compute	Los clientes ahora pueden personalizar el comportamiento en tiempo de ejecución de Spark y optimizar el rendimiento de SQL configurando las propiedades de Spark compatibles para las consultas de SQL. Esta capacidad ya está disponible para la generación de audiencias, la creación de puestos de trabajo y la creación de MLInput canales .	5 de enero de 2026
Support para una supervisión detallada con CloudWatch	Los clientes ahora pueden optar por una supervisión detallada con CloudWatch métricas para la supervisión	2 de enero de 2026

operativa de las colaboraciones, incluidos el rendimiento de las consultas y la utilización de los recursos.

[Support for Collaboration Change Request: eventos EventBridge de notificación](#)

Se agregaron nuevos EventBridge eventos para los eventos del ciclo de vida de las solicitudes de cambio de colaboración.

19 de diciembre de 2025

[Support for change requests to manage collaboration, membership and settings](#)

Los creadores de colaboraciones ahora pueden enviar solicitudes de cambio para añadir nuevos miembros, actualizar las capacidades de los miembros existentes y modificar la configuración de aprobación automática. Todos los miembros de la colaboración deben aprobar las solicitudes de cambio antes de que los cambios surtan efecto.

18 de diciembre de 2025

[Plazo de finalización del soporte para el motor de análisis de AWS Clean Rooms SQL](#)

Support para el motor de análisis AWS Clean Rooms SQL finaliza el 17 de diciembre de 2025. Los clientes ahora solo pueden usar el motor de análisis SQL de Spark.

17 de diciembre de 2025

[Support para eventos EventBridge de notificación adicionales](#)

Se agregaron nuevos EventBridge eventos para la plantilla de análisis, la asociación de tablas configuradas, la regla de análisis de asociación de tablas configuradas, la tabla de mapeo de ID, la asociación de espacios de nombres de ID, la asociación de espacios de nombres de ID, la tarea protegida y la consulta protegida.

15 de diciembre de 2025

[Support para la generación de conjuntos de datos sintéticos con privacidad mejorada para el entrenamiento de modelos de aprendizaje automático personalizados](#)

Los clientes ahora pueden generar conjuntos de datos sintéticos con mayor privacidad para entrenar modelos de aprendizaje automático personalizados y, al mismo tiempo, proteger los datos confidenciales.

30 de noviembre de 2025

[Support para propiedades configurables de Spark en consultas SQL](#)

Los clientes ahora pueden personalizar el comportamiento en tiempo de ejecución de Spark configurando las propiedades de Spark compatibles para las consultas SQL.

30 de octubre de 2025

[Support for cross-region collaboration](#)

AWS Clean Rooms ahora admite la colaboración entre regiones para AWS Glue las tablas de datos de Athena, lo que permite a los clientes procesar los datos almacenados en Región de AWS una sola y entregar los resultados a las regiones aprobadas por todos los miembros de la colaboración.

3 de octubre de 2025

[Support for data access budget](#)

Los clientes ahora pueden establecer y administrar los presupuestos de acceso a los datos para controlar el uso de las mesas.

1 de octubre de 2025

[Complete la tabla de mapeo de ID: actualice](#)

Los clientes ahora pueden utilizar el procesamiento incremental en un flujo de trabajo de mapeo de ID basado en reglas para procesar grandes conjuntos de datos de forma más eficiente. Los clientes también pueden eliminar registros de un flujo de trabajo de mapeo de identidad para cumplir con las normas de administración de datos.

23 de septiembre de 2025

[Support para tamaños de carga de trabajo variables para PySpark trabajos](#)

Los clientes ahora pueden elegir y ajustar su tamaño de cómputo para ejecutar los PySpark trabajos de forma más eficaz.

4 de septiembre de 2025

[Support para añadir nuevos miembros a una colaboración existente](#)

Los clientes ahora pueden añadir nuevos miembros a las colaboraciones existentes. Esta función incluye opciones para configurar los ajustes de aprobación automática y bloquear las solicitudes de membresía.

3 de septiembre de 2025

[Se agregó soporte para configurar los resúmenes de errores y la redacción de registros de modelos personalizados](#)

Los clientes ahora pueden configurar los resúmenes de errores y la redacción de los registros de modelos personalizados en una asociación de algoritmos de modelos configurada.

3 de septiembre de 2025

[Support for Snowflake, opción de autenticación](#)

Los clientes ahora pueden usar la autenticación por pares de claves con una clave privada PEM de Snowflake al almacenar sus credenciales de Snowflake para las conexiones. AWS Secrets Manager AWS Clean Rooms

29 de agosto de 2025

[Se PySpark agregó una configuración de errores para las plantillas de análisis](#)

Los clientes ahora pueden configurar los niveles de mensajes de error básicos o detallados en las plantillas de PySpark análisis para gestionar la visibilidad de los errores y las capacidades de solución de problemas.

19 de agosto de 2025

Support para EventBridge	Los clientes ahora pueden usar Amazon EventBridge para enrutar eventos desde AWS Clean Rooms otros componentes de software.	31 de julio de 2025
Actualización de una política existente	Se han agregado los siguientes permisos nuevos a la política <code>AWSCleanRoomsFullAccessNoQuerying</code> administrada: <code>cleanrooms:UpdateConfiguredTableAllowedColumns</code> y <code>cleanrooms:UpdateConfiguredTableReference</code> .	29 de julio de 2025
Mayores capacidades de edición para las tablas configuradas	Los clientes ahora pueden modificar las columnas y los campos de referencia de tabla permitidos para las tablas configuradas.	29 de julio de 2025
El canal de entrada ML admite el tipo de archivo Parquet	Los clientes ahora pueden elegir el tipo de archivo Parquet como formato de datos que puede utilizar el canal de entrada ML.	17 de julio de 2025
Support para una privacidad diferencial en el motor de análisis SQL de Spark	Los clientes ahora pueden utilizar la privacidad diferencial con el motor de análisis SQL de Spark. Además, la longitud máxima del texto de una declaración de consulta SQL que utilice la privacidad diferencial es ahora de 90 KB.	16 de julio de 2025

[Plazo de finalización del soporte para el motor de análisis de AWS Clean Rooms SQL](#)

El soporte para nuevas colaboraciones mediante el motor de análisis SQL de Clean Rooms finaliza el 16 de julio de 2025. A partir del 17 de julio de 2025, no podrás crear nuevas colaboraciones con el antiguo motor de análisis SQL de Clean Rooms. AWS Clean Rooms dejará de ser compatible con el motor de análisis SQL de Clean Rooms el 17 de diciembre de 2025.

16 de julio de 2025

[Support para formación incremental y distribuida](#)

Los clientes ahora pueden utilizar la formación incremental y la formación distribuida para sus modelos personalizados.

1 de julio de 2025

[Plazo de finalización del soporte para el motor de análisis SQL de Clean Rooms](#)

A partir de ahora y hasta el 16 de julio de 2025, las nuevas colaboraciones que utilicen el antiguo motor de análisis SQL de Clean Rooms requieren un aumento del límite. A partir del 17 de julio de 2025, no podrás crear nuevas colaboraciones con este motor. AWS Clean Rooms finaliza la compatibilidad con el motor de análisis SQL de Clean Rooms el 17 de diciembre de 2025.

22 de mayo de 2025

Actualización de las plantillas PySpark de análisis	Los clientes ahora pueden asignar un nombre a un archivo de script de usuario con cualquier nombre de archivo Python válido.	15 de mayo de 2025
Support para múltiples receptores de resultados	Los clientes ahora pueden seleccionar varios miembros para recibir los resultados de una sola consulta.	30 de abril de 2025
Support para migrar colaboraciones a Spark SQL	AWS Clean Rooms SQL ahora admite reglas de agregación y análisis de listas, además de reglas de análisis personalizadas. Además, los clientes pueden actualizar una colaboración existente para usar el motor de análisis de Spark que impulsa Spark SQL.	2 de abril de 2025
Support for PySpark jobs	Los clientes ahora pueden analizar los datos mediante la ejecución de tareas mediante plantillas de PySpark análisis aprobadas.	18 de marzo de 2025

[Actualización de políticas existentes](#)

Se ha agregado el permiso nuevo siguiente a la política administrada `AWSCleanRoomsMLReadOnlyAccess` : `PassCleanRoomsResources` . Se han agregado los siguientes permisos nuevos a la política `AWSCleanRoomsMLFullAccess` administrada: `PassCleanRoomsResources` y `ConsoleDescribeECRRepositories` .

10 de enero de 2025

[Support para varios trabajadores de cómputo](#)

Los clientes ahora pueden especificar qué tipo de trabajadores informáticos y cuántos aprovisionar al crear un segmento similar.

17 de diciembre de 2024

[Support para múltiples fuentes de datos y nubes](#)

Los clientes ahora pueden usar múltiples fuentes de datos y nubes, como Amazon Athena y Snowflake, para colaborar con los conjuntos de datos de sus socios.

1 de diciembre de 2024

[El modelado personalizado de Clean Rooms ML ya está disponible](#)

Los clientes ahora pueden usar sus propios modelos de aprendizaje automático personalizados en colaboración.

7 de noviembre de 2024

Nuevo motor de análisis	Los clientes que tienen conjuntos de datos de gran tamaño ahora pueden ejecutar consultas complejas mediante las funciones de SQL compatibles con el motor de análisis SQL de Spark.	29 de octubre de 2024
Protección de privacidad mejorada, creación de audiencias parecidas, elección de varios receptores de resultados	Puede proteger los datos y, al mismo tiempo, permitir consultas de activación complejas mediante los análisis adicionales y la regla de análisis de colaboración. Puede crear modelos de audiencia similares a partir de consultas SQL o plantillas de análisis. Puede elegir varios miembros para recibir los resultados.	24 de julio de 2024
Resolución de entidades en AWS Clean Rooms	Con AWS Entity Resolution in AWS Clean Rooms, puede crear una tabla de mapeo de ID entre dos espacios de nombres de ID para consultar datos de eventos en espacios de identidad dispares.	23 de julio de 2024
Actualización de una política existente	Se ha agregado el permiso nuevo siguiente a la política administrada <code>AWSCleanRoomsFullAccessNoQuerying :cleanrooms:BatchGetSchemaAnalysisRule</code> .	13 de mayo de 2024

[AWS Clean Rooms ML ya está totalmente disponible](#)

AWS Clean Rooms ML proporciona un método que mejora la privacidad para que dos partes identifiquen usuarios similares en sus datos sin necesidad de compartir sus datos entre sí.

3 de abril de 2024

[Actualización de una política existente](#)

El ID de declaración de la política administrada por `AWSCleanRoomsFullAccess` se ha actualizado de `ConsolePickQueryResultsBucket` a `SetQueryResultsBucket` para representar mejor los permisos desde los permisos.

21 de marzo de 2024

[Nuevas políticas administradas para AWS Clean Rooms ML](#)

Se han añadido dos nuevas políticas administradas: `AWSCleanRoomsMLReadOnlyAccess` y `AWSCleanRoomsMLFullAccess`.

29 de noviembre de 2023

[AWS Clean Rooms ML \(versión preliminar\)](#)

AWS Clean Rooms ML proporciona un método que mejora la privacidad para que dos partes identifiquen usuarios similares en sus datos sin necesidad de compartir sus datos entre sí.

29 de noviembre de 2023

[AWS Clean Rooms Privacidad diferencial \(versión preliminar\)](#)

Los clientes ahora pueden usar AWS Clean Rooms Differential Privacy para ayudar a proteger la privacidad de sus usuarios.

29 de noviembre de 2023

[Configuración de pagos](#)

El creador de la colaboración ahora puede designar al miembro que puede ejecutar consultas o a otro miembro de la colaboración para que se le facturen los costos de computación de las consultas.

14 de noviembre de 2023

[Tiempo de ejecución de consulta: actualización](#)

El tiempo máximo que se ejecuta una consulta antes de que se agote el tiempo de espera se ha actualizado de 4 a 12 horas.

6 de octubre de 2023

[CloudFormation recursos: actualización](#)

AWS Clean Rooms ha agregado los siguientes recursos nuevos:

```
AWS::CleanRooms::Membership Protected
QueryOutputConfiguration
AWS::CleanRooms::Membership ProtectedQueryResultConfiguration
,
yAWS::CleanRooms::Membership ProtectedQueryS3OutputConfiguration .
```

7 de septiembre de 2023

CloudFormation recursos: actualizar	AWS Clean Rooms ha añadido los siguientes recursos nuevos: AWS::CleanRooms::AnalysisTemplate y AWS::CleanRooms::ConfiguredTableAnalysisRuleCustom .	31 de agosto de 2023
Capacidades independientes de los miembros	El creador de la colaboración ahora puede designar a un miembro que realice las consultas y a otro miembro que reciba los resultados. Esto permite al creador de la colaboración asegurarse de que el miembro que puede realizar las consultas no tenga acceso a los resultados de la consulta.	30 de agosto de 2023
AWS Clean Rooms Glosario	Actualización exclusiva de la documentación para añadir un glosario de términos. AWS Clean Rooms	30 de agosto de 2023
Compatibilidad para tablas de Apache Iceberg (vista previa)	AWS Clean Rooms ahora admite Apache Iceberg tablas (vista previa).	25 de agosto de 2023
Actualización de cuotas	La sección Cuotas se ha actualizado para reflejar la nueva cuota predeterminada de pertenencias por cuenta.	9 de agosto de 2023

Actualización de la política existente

Se han añadido los siguientes nuevos permisos a la política administrada de `AWSCleanRoomsFullAccessNoQuerying` :

- `cleanrooms:CreateAnalysisTemplate` ,
- `cleanrooms:GetAnalysisTemplate` ,
- `cleanrooms:UpdateAnalysisTemplate` ,
- `cleanrooms>DeleteAnalysisTemplate` ,
- `cleanrooms:ListAnalysisTemplates` ,
- `cleanrooms:GetCollaborationAnalysisTemplate` ,
- `cleanrooms:BatchGetCollaborationAnalysisTemplate` y
- `cleanrooms>ListCollaborationAnalysisTemplates` .

31 de julio de 2023

Plantillas de análisis y regla de análisis personalizada	AWS Clean Rooms ahora admite plantillas de análisis y la regla de análisis personalizada. Las plantillas de análisis permiten a los colaboradores crear o importar su propia consulta SQL personalizada para utilizarla en la colaboración. Con la regla de análisis personalizada, el propietario de la tabla puede aprobar consultas SQL personalizadas en sus tablas configuradas.	31 de julio de 2023
Las reglas de análisis admiten la condición lógica OR	AWS Clean Rooms las reglas de análisis ahora admiten la condición OR lógica de la JOIN cláusula.	29 de junio de 2023
CloudFormation integration	AWS Clean Rooms ahora se integra con CloudFormation.	15 de junio de 2023
Creador de análisis	Los miembros que pueden realizar consultas y recibir resultados ahora tienen la posibilidad de ejecutar consultas en algunas tablas sin escribir código SQL utilizando la interfaz de usuario del creador de análisis.	15 de junio de 2023
Funciones SQL	Actualización solo de la documentación para aclarar las funciones SQL admitidas.	5 de mayo de 2023

Solución de problemas	Actualización solo de la documentación para añadir una sección de solución de problemas para los problemas más frecuentes.	27 de abril de 2023
Tipos de datos compatibles para AWS Clean Rooms	Actualización exclusiva de la documentación para añadir una nueva sección que muestre los tipos de AWS Glue Data Catalog datos compatibles.	26 de abril de 2023
Ejemplos de eventos AWS CloudTrail	Actualización solo de la documentación para agregar ejemplos de CloudTrail eventos StartProtectedQuery (exitosos) y StartProtectedQuery (fallidos).	20 de abril de 2023
Actualización de la política existente	Se han añadido los siguientes nuevos permisos a la política administrada AWSCleanRoomsFullAccessNoQuerying : cleanrooms:ListTagsForResource , cleanrooms:UntagResource y cleanrooms:TagResource . Para obtener más información, consulte Políticas administradas de AWS .	21 de marzo de 2023
Disponibilidad general	AWS Clean Rooms ahora está disponible de forma general.	21 de marzo de 2023

[Versión de prueba](#)

Versión preliminar de la
Guía AWS Clean Rooms del
usuario

12 de enero de 2023

AWS Clean Rooms Glosario

Consulte este glosario para familiarizarse con la terminología utilizada en AWS Clean Rooms.

Regla de análisis de agregación

La restricción de consultas que permite realizar consultas que agreguen análisis utilizando funciones COUNT, SUM o AVG en dimensiones opcionales. Estas consultas no revelarán información de nivel de fila.

Admite casos de uso como la planificación de campañas, el alcance mediático, la frecuencia y la medición de conversiones.

Otros tipos de reglas de análisis son las [personalizadas](#) y las de [lista](#).

Reglas de análisis

Las restricciones de consulta que autorizan un tipo específico de consulta.

El tipo de regla de análisis determina qué tipo de análisis se puede ejecutar en la tabla configurada. Cada tipo tiene una estructura de consulta predefinida. Usted controla cómo se pueden utilizar las columnas de la tabla en la estructura mediante los controles de consulta.

Los tipos de reglas de análisis son de [agregación](#), de [lista](#) y [personalizadas](#).

Plantilla de análisis

Una consulta preaprobada y específica de la colaboración que se puede reutilizar.

Formatos compatibles: código SQL o código Python para Spark.

Si utiliza SQL, la plantilla de análisis puede contener parámetros siempre que aparezca un valor literal en una consulta SQL. Para obtener más información sobre los tipos de parámetros admitidos, consulte [Tipos de datos](#) en la Referencia de SQL de AWS Clean Rooms .

Las plantillas de análisis solo funcionan con la [regla de análisis personalizada](#).

Cliente de cifrado de S3

El cliente de cifrado de computación criptográfica para Clean Rooms (C3R).

C3R, que se utiliza para cifrar y descifrar datos, es un SDK de cifrado del cliente con una interfaz de línea de comandos.

Columna de texto sin cifrar

Una columna que no está protegida criptográficamente para un constructo de SQL JOIN ni SELECT.

Las columnas de texto sin cifrar se pueden usar en cualquier parte de la consulta SQL.

Colaboración

Un límite lógico seguro AWS Clean Rooms en el que los miembros pueden realizar consultas SQL en tablas configuradas.

Las colaboraciones las crea el [creador de la colaboración](#).

Solo los miembros que hayan sido invitados a la colaboración pueden unirse a ella.

Una colaboración solo puede tener un [miembro que pueda consultar](#) datos o un [miembro que pueda ejecutar consultas y trabajos](#).

Una colaboración solo puede tener un [miembro que pueda recibir los resultados](#).

Una colaboración solo puede tener un [miembro que pague los costos de procesamiento de consultas](#) o un [miembro que pague los costos de procesamiento de consultas y trabajos](#).

Todos los miembros pueden ver la lista de participantes invitados a la colaboración antes de unirse a ella.

Creador de la colaboración

El miembro que crea una colaboración.

Solo hay un creador por colaboración.

El creador de la colaboración es el único que puede eliminar miembros de la colaboración o eliminar la colaboración.

Tabla configurada

Cada tabla configurada representa una referencia a una tabla existente en la AWS Glue Data Catalog que se ha configurado para su uso en AWS Clean Rooms. Una tabla configurada contiene una regla de análisis que determina cómo se pueden utilizar los datos.

Actualmente, AWS Clean Rooms admite la asociación de datos almacenados en Amazon Simple Storage Service (Amazon S3) que están catalogados mediante AWS Glue

Para obtener más información al respecto AWS Glue, consulte la Guía para [AWS Glue desarrolladores](#).

Las tablas configuradas se pueden asociar a una o más colaboraciones.

Note

AWS Clean Rooms actualmente no es compatible con las ubicaciones de bucket de Amazon S3 en las que esté registrado AWS Lake Formation.

Regla de análisis personalizada

La restricción de consultas que permite un conjunto específico de consultas preaprobadas ([plantillas de análisis](#)) o permite que un conjunto específico de cuentas puedan proporcionar consultas o trabajos que utilicen sus datos.

Admite casos de uso como la atribución de primer toque, los análisis incrementales y los análisis de detección de audiencia.

Admite la privacidad diferencial.

Otros tipos de reglas de análisis son de [agregación](#) y [lista](#).

Descifrado

El proceso de transformar los datos cifrados para devolverles su forma original. El descifrado solo se puede realizar si se tiene el acceso a la clave secreta.

Privacidad diferencial

Una técnica matemáticamente rigurosa que protege los datos de la colaboración para que el miembro pueda obtener resultados al aprender sobre una persona específica.

Cifrado

Proceso de codificación de datos en un formato aparentemente aleatorio utilizando un valor secreto denominado clave. Es imposible determinar el texto sin formato original sin tener acceso a la clave.

Columna de huella digital

Columna protegida criptográficamente para un constructo SQL JOIN.

Método de flujo de trabajo de asignación de ID

Cómo desea que se realice la asignación de ID.

Existen dos métodos de flujo de trabajo de asignación de ID:

- Asignación de ID basado en reglas: el método por el que se utilizan reglas de coincidencia para traducir datos propios de un origen a un destino en un flujo de trabajo de asignación de ID.
- Asignación de ID de servicios de proveedores: el método por el que se utiliza un servicio de proveedores para traducir datos codificados de terceros de un origen a un destino en un flujo de trabajo de asignación de ID.

AWS Clean Rooms actualmente admite LiveRamp como método de flujo de trabajo de mapeo de ID basado en servicios de proveedores. Debe tener una suscripción para LiveRamp poder AWS Data Exchange utilizar este método. Para obtener más información, consulte [Suscripción a un servicio de proveedor en AWS Data Exchange](#) en la Guía del usuario de AWS Entity Resolution .

tabla de asignación de ID

Recurso AWS Clean Rooms que permite establecer reglas de coincidencia propias o transcodificar identidades multipartitas en una colaboración.

Una tabla de asignación de ID es una referencia a una tabla existente en AWS Glue Data Catalog. Contiene una [regla de análisis de tabla de asignación de ID](#) que determina cómo se pueden

consultar los datos en AWS Clean Rooms. Las tablas de asignación de ID se pueden asociar a una o más colaboraciones.

Regla de análisis de tablas de asignación de ID

Un tipo de regla de análisis se administra mediante AWS Clean Rooms y se usa para unir datos de identidad dispares a fin de facilitar las consultas. Se agrega automáticamente a las [tablas de asignación de ID](#) y no se puede editar. Hereda los comportamientos de las demás reglas de análisis de la colaboración, siempre que esas reglas de análisis sean homogéneas.

Flujo de trabajo de asignación de ID

Un trabajo de procesamiento de datos que asigna datos de un origen a un destino en función del [método de flujo de trabajo de asignación de ID](#) especificado. Produce una [tabla de asignación de ID](#).

Espacio de nombres de ID

[Un recurso AWS Clean Rooms que contiene metadatos que explican los conjuntos de datos de varios conjuntos de datos Cuentas de AWS y cómo utilizarlos en un flujo de trabajo de mapeo de identidades.](#)

Asociación de espacios de nombres de ID

Una asociación de un recurso de espacio de nombres de ID que le ayuda a detectar las entradas del [flujo de trabajo de asignación de ID](#).

Trabajo

Método para acceder a las tablas configuradas y analizarlas en una colaboración mediante un conjunto compatible de funciones, clases y variables.

AWS Clean Rooms actualmente admite el tipo de PySpark trabajo.

AWS Clean Rooms actualmente admite la ejecución de trabajos mediante una plantilla de PySpark análisis.

Regla de análisis de lista

La restricción de consultas que permite realizar consultas que generen un análisis de atributos de nivel de fila del solapamiento entre esta tabla y las tablas del miembro que puede realizar consultas.

Admite casos de uso como el enriquecimiento y la creación o supresión de audiencias.

Otros tipos de reglas de análisis son de [agregación](#) y [personalizadas](#).

Modelo parecido

Un modelo de los datos de un proveedor de datos de entrenamiento que permite a un proveedor de datos iniciales crear un [segmento similar de los](#) datos del proveedor de datos de entrenamiento que se parezca más a sus datos [iniciales](#).

Un segmento similar

Un subconjunto de los datos de entrenamiento que más se parece a los datos [iniciales](#).

Miembro

Un AWS cliente que participa en una [colaboración](#).

Un miembro se identifica mediante su Cuenta de AWS.

Todos los miembros pueden contribuir con datos.

Miembro que puede realizar consultas

El miembro que puede consultar los datos de la [colaboración](#).

Solo hay un miembro que puede realizar consultas por colaboración y ese miembro es inmutable.

Un usuario administrativo puede usar los permisos AWS Identity and Access Management (de IAM) para controlar cuáles de sus directores de IAM (como los usuarios o las funciones) pueden consultar los datos de la colaboración. Para obtener más información, consulte [Cree un rol de servicio para leer datos de Amazon S3](#).

Miembro que puede ejecutar consultas y trabajos

El miembro que puede ejecutar consultas y tareas con los datos de la [colaboración](#).

Solo hay un miembro que puede ejecutar consultas y trabajos por colaboración, y ese miembro es inmutable.

Un usuario administrativo puede usar los permisos AWS Identity and Access Management (de IAM) para controlar cuáles de sus directores de IAM (como usuarios o roles) pueden ejecutar consultas y trabajos en la colaboración. Para obtener más información, consulte [Cree un rol de servicio para leer datos de Amazon S3](#).

Miembro que puede recibir los resultados

Un miembro que puede recibir los resultados de las consultas. Un miembro que puede recibir los resultados especifica la configuración de los resultados de la consulta para el destino de Amazon S3 y el formato de los resultados de la consulta (CSV o Parquet).

En el caso de los análisis con el motor de análisis Spark, el miembro que puede recibir los resultados también especifica si los archivos deben generarse como un solo archivo o como varios archivos.

Puede haber más de un miembro que pueda recibir los resultados en una colaboración.

Miembro que paga los costos de computación de consultas

El miembro responsable de pagar los costos de computación de consultas.

Solo hay un miembro responsable de pagar los costos de computación de consultas por colaboración, y ese miembro es inmutable.

Si el creador de la colaboración no ha designado a nadie como miembro que paga los costos de computación de consultas, el [miembro que puede realizar consultas](#) es el pagador predeterminado.

El miembro que paga los costos de computación de las consultas recibe una factura por las consultas que se han ejecutado en la colaboración.

El miembro paga los costos de consulta y cálculo del trabajo

El miembro responsable de pagar los costos de consulta y cálculo del trabajo.

Solo hay un miembro que es responsable de pagar los costes de consulta y cálculo del trabajo por colaboración, y ese miembro es inmutable.

Si el creador de la colaboración no ha especificado a nadie como miembro que paga los costes de consulta y cálculo del trabajo, el [miembro que puede realizar la consulta](#) es el pagador predeterminado.

El miembro que pague los costes de consulta y cálculo del trabajo recibirá una factura por las consultas que se hayan realizado en la colaboración.

Pertenencia

Un recurso que se crea cuando un [miembro](#) se une a una [colaboración](#).

Todos los recursos que el miembro asocia a una colaboración forman parte de la pertenencia o están asociados a ella.

Solo el miembro propietario de la pertenencia puede añadir, eliminar o editar los recursos de esa pertenencia.

Columna sellada

Columna protegida criptográficamente para un constructo SQL SELECT.

Datos iniciales

Los datos del proveedor de datos iniciales, que se utilizan para crear un [segmento similar](#). Los datos iniciales se pueden proporcionar directamente o pueden provenir de los resultados de una consulta de AWS Clean Rooms . El resultado del segmento similar es un conjunto de usuarios a partir de los datos de entrenamiento que se parece más a los usuarios iniciales.

Motor de análisis Spark

Una opción de análisis AWS Clean Rooms que permite a los clientes ejecutar consultas complejas en grandes conjuntos de datos almacenados en Amazon S3, Amazon Athena o Snowflake mediante las funciones SQL de Apache Spark. También admite el análisis en. PySpark AWS Clean Rooms

Cuando creas una colaboración mediante la [CreateCollaborationAPI](#), el valor del motor de análisis de Spark lo es SPARK.

Consultar

Método para acceder a las tablas configuradas de una colaboración y analizarlas, mediante un conjunto compatible de funciones, clases y variables.

AWS Clean Rooms actualmente es compatible con el lenguaje de consultas SQL.

AWS Clean Rooms actualmente admite la ejecución de consultas SQL directas o la ejecución de consultas mediante una plantilla de análisis SQL.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la version original de inglés, prevalecerá la version en inglés.