aws

用户指南

AWS Clean Rooms

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Clean Rooms: 用户指南

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

什么是 AWS Clean Rooms？

AWS Clean Rooms 帮助您和您的合作伙伴对您的集体数据集进行分析和协作，以获得新的见解，而无需互相透露基础数据。您可以使用 AWS Clean Rooms安全的协作工作空间在几分钟内创建自己的干净室，只需几个步骤即可开始分析您的集体数据集。您可以选择要与之协作的合作伙伴，选择他们的数据集并为参与者配置限制。

借助 AWS Clean Rooms，您可以与成千上万家已经在使用的公司进行协作 AWS。协作不需要将数据移出 AWS 或加载到另一个平台。运行查询时，从数据的原始位置 AWS Clean Rooms 读取数据，并应用内置的分析规则来帮助您保持对查询数据的控制。

AWS Clean Rooms 提供您可以配置的内置数据访问控制和审计支持控件。这些控制包括：

- 用于限制 SQL 查询和提供输出约束的分析规则
- 用于即使在处理查询时也能保持数据加密以遵守严格的数据处理策略的 Clean Rooms 计算加密
- 用于查看查询并帮助支持审计查询日志
- 差异隐私，可防止用户识别尝试。 AWS Clean Rooms 差异隐私是一项完全托管的功能，它通过数学支持的技术和直观的控件来保护用户的隐私，您只需点击几下即可应用这些技术和直观的控件。
- AWS Clean Rooms 机器学习允许双方识别其数据中的相似用户，而无需彼此共享数据。第一方通过其训练数据创建并配置一个相似模型。第二方将其种子数据引入到一个协作中，并创建与训练数据类似的相似细分。

以下视频解释了更多相关信息 AWS Clean Rooms。

AWS Clean Rooms

你是首次 AWS Clean Rooms 使用吗？

如果您是首次使用 AWS Clean Rooms，我们建议您先阅读以下章节：

- 如何 AWS Clean Rooms 运作
- 正在访问 AWS Clean Rooms
- 设置 AWS Clean Rooms
- AWS Clean Rooms 词汇表

如何 AWS Clean Rooms 运作

以下工作流程假设：

- 协作成员已将其数据表上传到 Amazon S3 并创建了一个 AWS Glue 表。
- （可选）仅对于加密数据表，协作成员已经使用 C3R 加密客户端准备了加密数据表。

总而言之，的工作流程 AWS Clean Rooms 如下：

1. 协作创建者执行以下任务：
   - 创建协作。
   - 邀请一个或多个成员参与协作。
   - 为成员分配能力，例如可以查询的成员和可以接收结果的成员。

     如果协作创建者也是可以接收结果的成员，则他们会指定查询结果的目标和格式。他们还提供服务角色 Amazon 资源名称 (ARN)，用于将结果写入查询结果目标。
   - 配置负责支付协作中的查询计算费用的成员。
2. 受邀成员通过创建成员身份资源加入协作。

   如果受邀成员也是可以接收结果的成员，则他们会指定查询结果的目标和格式。他们还提供服务角色 ARN，用于写入查询结果目标。

   如果受邀成员是负责支付查询计算费用的成员，则他们在加入协作之前应接受自己的付款责任。
3. 成员配置现有 AWS Glue 表以供在中使用。 AWS Clean Rooms（除非使用 Clean Rooms 加密计算，否则此步骤可以在加入协作之前或之后完成。）

   > **Note**
   >
   > AWS Clean Rooms 支持 AWS Glue 表格。有关将数据获取到 AWS Glue的更多信息，请参阅步骤 3：将数据表上传到 Amazon S3。

   1. 成员命名配置表并选择要在协作中使用的列。
   2. 成员为配置表配置以下分析规则之一：
      - 聚合分析规则或列表分析规则 - 控制可在表上运行的分析类型。

- 自定义分析规则 - 允许一组特定的预先批准的查询或一组特定的可提供使用您的数据的查询的账户。允许成员开启差别隐私以防范用户识别尝试。

> **Note**
>
> 成员可以在将其配置表与协作关联之前随时配置分析规则。

4. 该成员将其配置的表与协作关联起来，并授予 AWS Clean Rooms 访问其 AWS Glue 表的服务角色。

> **Note**
>
> 此服务角色拥有对表的权限。只有代表可以查询的成员运行 AWS Clean Rooms 允许的查询时，才可以假设服务角色。任何协作成员（数据所有者除外）都无法访问协作中的底层表。数据所有者可以开启差别隐私，以使其表可供其他成员查询。

5. 可以查询的成员对配置表运行 SQL 查询。

   只有当负责支付查询计算费用的成员以活跃成员的身份加入协作时，才能运行查询。

   分析规则和输出约束是自动强制执行的。 AWS Clean Rooms 仅返回符合步骤 3.b 中定义的分析规则的结果。

   对于加密数据的查询，可以接收结果的成员会收到必须解密 AWS Clean Rooms 的加密输出（参见步骤 8）。

6. 可以接收结果的成员在 AWS Clean Rooms 控制台或他们指定的 Amazon S3 存储桶中查看结果。
7. 为查询计算费用付费的成员要对协作中运行的查询付费。
8. （可选）仅对于加密数据表，可以接收结果的成员通过在解密模式下运行 C3R 加密客户端来解密查询结果。

相关服务

以下内容与 AWS 服务 以下内容有关 AWS Clean Rooms：

- Amazon S3

  协作成员可以将他们带入的 Amazon S3 AWS Clean Rooms 中的数据存储。

有关更多信息，请参阅以下主题：

在中为查询准备数据表 AWS Clean Rooms

《Amazon Simple Storage Service 用户指南》中的什么是 Amazon S3？

- AWS Glue

  协作成员可以根据自己在 Amazon S3 中的数据创建 AWS Glue 表以供在中使用 AWS Clean Rooms。

  有关更多信息，请参阅以下主题：

  在中为查询准备数据表 AWS Clean Rooms

  AWS Glue 开发人员指南中的什么是 AWS Glue？

- AWS CloudFormation

  在中创建以下资源 AWS CloudFormation：协作、已配置的表、配置的表关联和成员资格

  有关更多信息，请参阅 使用创建 AWS Clean Rooms 资源 AWS CloudFormation。

- AWS CloudTrail

  AWS Clean Rooms 与 CloudTrail 日志配合使用可增强对 AWS 服务 活动的分析。

  有关更多信息，请参阅 使用 AWS CloudTrail 记录 AWS Clean Rooms API 调用。

正在访问 AWS Clean Rooms

您可以 AWS Clean Rooms 通过以下选项进行访问：

- 直接通过 AWS Clean Rooms 主机访问 https://console.aws.amazon.com/cleanrooms/。
- 通过 AWS Clean Rooms API 以编程方式进行。有关更多信息，请参阅 AWS Clean Rooms API 参考。

的定价 AWS Clean Rooms

有关定价信息，请参阅 AWS Clean Rooms 定价。

的账单 AWS Clean Rooms

AWS Clean Rooms 使协作创建者能够配置哪个成员为协作中的查询计算费用付费。

大多数情况下，可以查询的成员和为查询计算费用付费的成员是相同的人。但是，如果可以查询的成员和为查询计算费用付费的成员不同，则当可以查询的成员针对自己的成员身份资源运行查询时，将按支付查询计算费用的成员的成员身份资源计费。

支付查询计算费用的成员在其 CloudTrail 事件历史记录中看不到任何正在运行的查询的事件，因为付款人既不是运行查询的人，也不是运行查询所针对的资源的所有者。不过，对于可以在协作中运行查询的成员运行的所有查询，付款人将会看到针对其成员资源生成的账单。

有关如何创建协作和配置支付查询计算费用的成员的更多信息，请参阅创建协作。

中的分析规则 AWS Clean Rooms

作为启用表格 AWS Clean Rooms 用于协作分析的一部分，协作成员必须配置分析规则。

分析规则是每个数据所有者在配置表上设置的隐私增强控制。分析规则决定了如何分析配置表。

分析规则是对配置表（账户级资源）的账户级控制，并且在关联了配置表的任何协作中强制执行。如果未配置分析规则，则可以将配置表关联到协作，但无法对其进行查询。查询只能引用具有相同分析规则类型的配置表。

要配置分析规则，首先要选择分析类型，然后指定分析规则。在这两个步骤中，您都应考虑要启用的使用案例以及如何保护底层数据。

AWS Clean Rooms 对查询中引用的所有已配置表强制执行更严格的控制。

以下示例演示了限制性控制。

Example 限制性控制：输出约束

- 协作者 A 对标识符列的输出限制为 100。
- 协作者 B 对标识符列的输出限制为 150。

  引用两个配置表的聚合查询要求输出行中至少有 150 个不同的标识符值，才能在查询输出中显示。查询输出并没有显示由于输出约束而删除了结果。

Example 限制性控制：分析模板未获得批准

- 协作者 A 允许在其自定义分析规则中使用带有引用协作者 A 和协作者 B 配置表的查询的分析模板。
- 合作者 B 不允许使用分析模板。

  由于 Collaborator B 不允许使用分析模板，因此可以查询的成员无法运行该分析模板。

分析规则类型

有三种类型的分析规则：聚合、列表和自定义。下表对这些分析规则类型进行了比较。每种类型具有单独的部分，以描述如何指定分析规则。

下表显示了分析规则类型的比较总结。

支持的使用案例

下表显示了每种分析规则类型支持的使用案例的比较总结。

应用场景	聚合	列表	自定义
支持的分析	使用 COUNT、SUM 和 AVG 函数按可选维度聚合统计数据的查询	输出多个表之间重叠部分的行级列表的查询	经过分析模板或分析创建者审核并允许的任何自定义分析
常见使用案例	细分分析、衡量、归因	扩充、细分构建	首次接触归因、增量分析、受众发现
SQL 构造	• JOIN 语句：内部联接 • 聚合函数：非重复计数/计数、非重复总和和和平均值 • 标量函数：有限子集	• JOIN 语句：内部联接 • 标量函数：无	通过 SELECT 命令可使用的大多数 SQL 函数和 SQL 构造
子查询和通用表表达式 (CTE)	否	否	是
分析模板	否	否	是

支持的控制

下表显示了每种分析规则类型如何保护您的基础数据的比较摘要。

控件	聚合	列表	自定义
控制机制	控制如何在查询中使用表中的数据 （例如，允许对列 hashed_email 进行 COUNT 和 SUM。）	控制如何在查询中使用表中的数据 （例如，仅允许使用 hashed_email 列进行联接。）	控制允许在表上运行哪些查询 （例如，仅允许在分析模板“自定义查询 1”中定义的查询。）
内置隐私增强技术	• 给匹配项设盲 • 需要聚合 • 最小聚合阈值 >= • 2 预定义的查询结构	• 给匹配项设盲 • 需要重叠 • 预定义的查询结构	差别隐私
在运行查询之前对其进行审核	否	否	是，正在使用分析模板

有关中提供的分析规则的更多信息 AWS Clean Rooms，请参阅以下主题。

- 聚合分析规则
- 列表分析规则
- 中的自定义分析规则 AWS Clean Rooms

聚合分析规则

在 AWS Clean Rooms 中，聚合分析规则使用 COUNT、SUM 和/或 AVG 函数按可选维度生成聚合统计数据。将聚合分析规则添加到配置表后，可以查询的成员就能在配置表上运行查询。

聚合分析规则支持活动规划、媒体覆盖率、频率测量和归因等使用案例。

支持的查询结构和语法在 聚合查询结构和语法 中定义。

中定义的分析规则的参数包括查询控制和查询结果控制。聚合分析规则 — 查询控制其查询控制包括要求一个配置表至少联接到一个可直接或临时查询的成员所拥有的配置表。此要求可使您确保在您的表及其他人的表的交叉点 (INNER JOIN) 上运行查询。

聚合查询结构和语法

对具有聚合分析规则的表的查询必须遵循以下语法。

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

下表解释前面语法中列出的每个表达式。

Expression	定义	示例
select_aggregate_function_expression	包含以下表达式的逗号分隔列表： • select_aggregation_function_expression • select_aggregate_expression **Note** select_aggregate_expression 中必须至少有一个 select_aggregation_function_expression 。	SELECT SUM(PRICE), user_segment
select_aggregation_function_expression	应用于一个或多个列的一个或多个支持的聚合函数。只允许将列作为聚合函数的参数。 **Note** select_aggregate_expression 中必须至少有一个 select_aggregation	AVG(PRICE) COUNT(DISTINCT user_id)

Expression	定义	示例
	`_function _expression` 。	

select_grouping_column_expression	可以包含任何使用以下元素的表达式的表达式：
	• 表列名称
	• 支持的标量函数
	• 字符串文本
	• 数值文本
	
	Note
	`select_aggregate_expression` 可以带或不带 AS 参数对列设置别名。有关更多信息，请参阅 AWS Clean Rooms SQL 参考。
		`TRUNC(timestampColumn)`
		
		`UPPER(campaignName)`

Expression	定义	示例
table_expression	使用 join_condition 连接联接条件表达式的表或表的联接。 join_condition 返回布尔值。 table_expression 支持： • 特定的 JOIN 类型 (INNER JOIN) • join_condition 中的相等比较条件 (=) • 逻辑运算符（AND、OR）。	FROM consumer_table INNER JOIN provider_table ON consumer_table.identifier1 = provider_table.identifier1 AND consumer_table.identifier2 = provider_table.identifier2
where_expression	返回布尔值的条件表达式。它可能包括以下内容： • 表列名称 • 支持的标量函数 • 数学运算符 • 字符串文本 • 数值文本 支持的比较条件是 (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL)。 支持的逻辑运算符是 (AND, OR)。 where_expression 是可选项。	WHERE where_condition WHERE price > 100 WHERE TRUNC(timestampColumn) = '1/1/2022' WHERE timestampColumn = timestampColumn2 - 14

Expression	定义	示例
group_by_expression	满足 select_grouping_column_expression 要求的表达式的逗号分隔列表。	GROUP BY TRUNC(timestampColumn), UPPER(campaignName), segment
having_expression	返回布尔值的条件表达式。它们具有应用于单列（例如 SUM(price) ）的支持聚合函数，并与数值文字进行比较。 支持的比较条件是 (=, >, <, <=, >=, <>, !=)。 支持的逻辑运算符是 (AND, OR)。 having_expression 是可选项。	HAVING SUM(SALES) > 500
order_by_expression	与前面定义的 select_aggregate_expression 中定义的要求一致的表达式的逗号分隔列表。 order_by_expression 是可选项。 **Note** order_by_expression 允许 ASC 和 DESC 参数。有关更多信息，请参阅 AWS Clean Rooms SQL 参考中的 ASC DESC 参数。	ORDER BY SUM(SALES), UPPER(campaignName)

关于聚合查询的结构和语法，请注意以下几点：

- 不支持除 SELECT 之外的 SQL 命令。
- 不支持子查询和通用表格表达式（例如 WITH）。
- 不支持组合多个查询的运算符（例如 UNION）。
- TOP、LIMIT 和 OFFSET 参数不受支持。

聚合分析规则 — 查询控制

使用聚合查询控制，您可以控制如何使用表中的列来查询表。例如，您可以控制哪一列用于联接，哪一列可以计数，或者 WHERE 语句中可以使用哪一列。

下面几节解释每种控制。

主题

聚合控制

通过使用聚合控制，您可以定义允许哪些聚合函数以及必须将其应用于哪些列。聚合函数可以在 SELECT、HAVING 和 ORDER BY 表达式中使用。

控件	定义	使用量
`aggregateColumns`	允许在聚合函数中使用的已配置表列的列。	`aggregateColumns` 可以在 SELECT、HAVING 和 ORDER BY 表达式中的聚合函数中使用。 有些 `aggregateColumns` 也可以归类为 `joinColumn`（稍后定义）。

控件	定义	使用量
		给定的 aggregateColumn 也不能归类为 dimension Column （稍后定义）。
function	允许在 aggregateColumns 上使用的 COUNT、SUM 和 AVG 函数。	function 可以应用于与之关联的 aggregateColumns 。

联接控制

JOIN 子句用于根据两个或多个表中的相关列合并两个或多个表中的行。

您可以使用联接控制来控制如何将您的表联接到 table_expression 中的其他表。AWS Clean Rooms 仅支持 INNER JOIN。INNER JOIN 语句只能使用在分析规则中明确归类为 joinColumn 的列，但要遵守您定义的控制。

INNER JOIN 必须对您的已配置表中的 joinColumn 和协作中另一个已配置表中的 joinColumn 进行操作。您可以决定表中的哪些列可以用作 joinColumn。

ON 子句中的每个匹配条件都要求在两列之间使用相等比较条件 (=)。

ON 子句中的多个匹配条件可以是：

- 使用 AND 逻辑运算符组合
- 使用 OR 逻辑运算符分隔

> **Note**
>
> 所有 JOIN 匹配条件都必须与 JOIN 两侧各一条记录匹配。所有由 OR 或 AND 逻辑运算符连接的条件也必须遵守此要求。

以下是使用 AND 逻辑运算符的查询示例。

```
SELECT some_col, other_col
FROM table1
```

```
    JOIN table2
    ON table1.id = table2.id AND table1.name = table2.name
```

以下是使用 OR 逻辑运算符的查询示例。

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

控件	定义	使用量
joinColumns	您希望允许可以查询的成员在 INNER JOIN 语句中使用的列（如果有）。	特定的 joinColumn 也可以归类为 aggregate Column （参阅聚合控制）。 同一列不能同时用作 joinColumn 和 dimension Columns （见下文）。 除非它也被归类为 aggregate Column ，否则除了 INNER JOIN 之外，查询的其他部分都不能使用 joinColumn 。
joinRequired	控制您是否要求与可以查询的成员的已配置表进行 INNER JOIN。	如果启用了此参数，则要求 INNER JOIN。如果未启用此参数，则 INNER JOIN 是可选的。 假设您启用了此参数，则可以查询的成员需要在 INNER JOIN 中包含他们拥有的表。他们必须将您的表与他们的表 JOIN，可以是直接，也可以是传递（也就是说，将他们的表联接到另一个表，而另一个表又联接到您的表）。

以下是传递联接的示例。

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

> **Note**
>
> 可以查询的成员也可以使用 joinRequired 参数。在这种情况下，查询必须将其表与至少一个其他表联接。

维度控制

维度控制控制可以对聚合列进行筛选、分组或聚合的列。

控件	定义	使用量
dimensionColumns	您允许可以查询的成员在 SELECT、WHERE、GROUP 、BY 和 ORDER BY 中使用的列（如果有）。	dimensionColumn 可以在 SELECT (select_grouping_column_expression)、WHERE、GROUP BY 和 ORDER BY 中使用。 同一列不能同时是 dimensionColumn 、joinColumn 和/或 aggregateColumn 。

标量函数

标量函数控制哪些标量函数可以在维度列上使用。

控件	定义	使用量
scalarFunctions	可在查询的 dimension Columns 上使用的标量函数。	指定允许在 dimension Columns 上应用的标量函数（如果有）（例如 CAST）。 标量函数不能在其他函数之上使用，也不能在其他函数中使用。标量函数的参数可以是列、字符串文本或数字文本。

支持以下标量函数：

- 数学函数 — ABS、CEILOR、FLOOR、LN、ROUND、SQRT
- 数据类型格式设置函数 — CAST, CONVERT, TO_CHAR, TO_DATE, TO_NUMBER, TO_TIMESTAMP
- 字符串函数 — 下限、上限、TRIM、RTRIM、SUBSTRING
  - 对于 RTRIM，不允许使用自定义字符集进行修剪。
- 条件表达式 — COALESCE
- 日期函数 — 提取、获取日期、当前日期、日期添加
- 其他函数 — TRUNC

有关详细信息，请参阅 AWS Clean Rooms SQL 参考。

聚合分析规则 — 查询结果控制

使用聚合查询结果控制，可以通过指定每个输出行必须满足的一个或多个条件来控制返回哪些结果。AWS Clean Rooms 支持 `COUNT (DISTINCT column) >= X` 形式的聚合约束。此形式要求每行至少聚合从配置表中选择的 X 个不同值（例如，不同 `user_id` 值的最少个数）。即使提交的查询本身不使用指定的列，也会自动强制执行此最低阈值。它们是在来自协作中每个成员的已配置表的查询中的每个已配置表中共同强制执行的。

每个配置表的分析规则中必须有至少一个聚合约束。配置表所有者可以添加多个 `columnName` 和关联的 `minimum`，这些表将共同强制执行。

聚合约束

聚合约束 控制返回查询结果中的哪些行。要返回，行必须满足聚合约束中指定的每列中指定的最小不同值数。即使在查询或分析规则的其他部分中未明确提及该列，此要求也适用。

控件	定义	使用量
columnName	在每个输出行必须满足的条件中使用的 aggregateColumn 。	可以是已配置表中的任何列。
minimum	要在查询结果中返回输出行（例如 COUNT DISTINCT），关联 aggregateColumn 必须具有的最小不同值个数。	minimum 的值必须至少为 2。

聚合分析规则结构

以下示例显示了聚合分析规则的预定义结构。

在以下示例中，*MyTable* 指您的数据表。您可以将每个*#######*替换为自己的信息。

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

聚合分析规则 — 示例

以下示例演示了两家公司如何使用聚合分析在 AWS Clean Rooms 中进行协作。

A 公司有客户和销售数据。A 公司有兴趣了解产品退货活动。B 公司是 A 公司的零售商之一，有退货数据。B 公司也有对 A 公司有用的客户细分属性（例如，购买过相关产品、使用过零售商的客户服务）。B 公司不想提供行级客户退货数据和属性信息。B 公司只想为 A 公司启用一组查询，以最小聚合阈值获取重叠客户的聚合统计数据。

A 公司和 B 公司决定协作，以便 A 公司能够了解产品退货活动，并在 B 公司和其他渠道提供更好的产品。

为了创建协作并进行聚合分析，两家公司执行以下操作：

1. A 公司创建协作并创建成员身份。协作中的另一个成员是 B 公司。A 公司在协作中启用查询日志记录，并在其账户中启用查询日志记录。
2. B 公司在协作中创建成员身份。它在其账户中启用查询日志记录。
3. A 公司创建销售配置表。
4. A 公司将以下聚合分析规则添加到销售配置表中。

```
{
  "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 — A 公司想要计算销售数据和退货数据之间重叠的唯一客户数量。A 公司还想汇总 purchases 数量，以便与 returns 数量进行比较。

joinColumns — A 公司想要使用 identifier 来匹配销售数据中的客户和退货数据中的客户。这将有助于 A 公司将退货与正确的采购相匹配。它还可以帮助 A 公司细分重叠的客户。

dimensionColumns — A 公司使用 dimensionColumns 来筛选特定产品，比较一段时期内的购买和退货情况，确保退货日期在产品日期之后，并帮助细分重叠客户。

scalarFunctions — A 公司选择 CAST 标量函数，以便在需要时根据 A 公司关联到协作的配置表更新数据类型格式。它还添加了标量函数，以便在需要时帮助格式化列。

outputConstraints — A 公司设定了最低输出约束。它不需要限制结果，因为允许分析师从销售表中查看行级数据

> **Note**
>
> A 公司没有在分析规则中加入 joinRequired。它为他们的分析师提供了单独查询销售表的灵活性。

5. B 公司创建退货配置表。

6. B 公司将以下聚合分析规则添加到退货配置表中。

```
{
  "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 — B 公司让 A 公司汇总 returns 以与购买数量进行比较。它们至少有一个聚合列，因为它们启用了聚合查询。

joinColumns — B 公司让 A 公司在 identifier 上进行联接，以将退货数据中的客户与销售数据中的客户进行匹配。identifier 数据特别敏感，将其作为 joinColumn 可确保数据永远不会在查询中输出。

joinRequired — B 公司要求对退货数据的查询必须与销售数据重叠。他们不想让 A 公司查询其数据集中的所有个人。他们还在协作协议中商定了这一限制。

dimensionColumns — B 公司让 A 公司按 state、popularpurchases 和 customerserviceuser 进行筛选和分组，这些独特的属性有助于 A 公司进行分析。B 公司让 A 公司使用 returndate 筛选在 purchasedate 之后发生的 returndate 的输出。通过这种筛选，输出可以更准确地评估产品变更的影响。

scalarFunctions — B 公司启用以下函数：

- TRUNNC（表示日期）
- LOWER 和 UPPER（如果 producttype 在数据中以不同的格式输入）
- CAST（如果 A 公司需要将销售表中的数据类型转换为与退货中表的数据类型相同）

A 公司不启用其他标量函数，因为他们认为查询不需要这些函数。

outputConstraints — B 公司对 hashedemail 设定了最低输出约束，以帮助降低重新识别客户身份的能力。它还对 producttype 增加了最低输出约束，以降低重新识别退回的特定产品的能力。根据输出的维度（例如 state），某些产品类型可能更占优势。无论 A 公司在其数据中添加了什么输出约束，他们的输出约束都将始终得到执行。

7. A 公司创建了与协作的销售表关联。
8. B 公司创建了与协作的退货表关联。
9. A 公司运行查询（如以下示例），以更好地了解 B 公司的退货数量与 2022 年各地采购总量的对比情况。

```
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;
```

10A 公司和 B 公司查看查询日志。B 公司验证查询是否符合协作协议中上商定的内容。

聚合分析规则问题疑难解答

使用此处的信息可帮助您诊断和修复在使用聚合分析规则时出现的常见问题。

问题

- 我的查询没有返回任何结果

我的查询没有返回任何结果

当没有匹配结果或匹配结果不符合一个或多个最低聚合阈值时，就会发生这种情况。

有关最低聚合阈值的更多信息，请参阅聚合分析规则 — 示例。

列表分析规则

在 AWS Clean Rooms 中，列表分析规则 输出行级列表，显示添加到的配置表与可以查询成员的配置表之间的重叠。可以查询的成员运行包含列表分析规则的查询。

列表分析规则类型支持扩充和受众构建等使用案例。

有关此分析规则的预定义查询结构和语法的更多信息，请参阅列表分析规则预定义结构。

列表分析规则的参数（在列表分析规则 — 查询控制中定义）具有查询控制。它的查询控制包括选择可以在输出中列出的列的功能。查询要求至少有一次与可以查询成员的配置表联接，可以是直接联接，也可以是传递联接。

不存在像聚合分析规则那样的查询结果控制。

列表查询只能使用数学运算符。它们不能使用其他函数（例如聚合或标量）。

主题

- 列表查询结构和语法
- 列表分析规则 — 查询控制
- 列表分析规则预定义结构
- 列表分析规则 — 示例

列表查询结构和语法

对具有列表分析规则的表的查询必须遵循以下语法。

```
--select_list_expression
SELECT
[TOP number ] 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]
```

下表解释前面语法中列出的每个表达式。

Expression	定义	示例
select_list_expression	包含至少一个表列名的逗号分隔列表。 DISTINCT 参数是必需的。 **Note** select_list_expression 可以带或不带 AS 参数对列设置别名。它还支持 TOP 参数。有关更多信息，请参阅 AWS Clean Rooms SQL 参考。	SELECT DISTINCT segment
table_expression	使用 join_condition 连接到 join_condition 的表或表的联接。 join_condition 返回布尔值。 table_expression 支持： • 特定的 JOIN 类型 (INNER JOIN) • join_condition 中的相等比较条件 (=) • 逻辑运算符（AND、OR）。	FROM consumer_table INNER JOIN provider_table ON consumer_table.identifier1 = provider_table.identifier1 AND consumer_table.identifier2 = provider_table.identifier2
where_expression	返回布尔值的条件表达式。它可能包括以下内容： • 表列名称	WHERE state + '_' + city = 'NY_NYC'

Expression	定义	示例
	• 数学运算符 • 字符串文本 • 数值文本 支持的比较条件是 (=, >, <, <=, >=, <>, !=, NOT, IN, NOT IN, LIKE, IS NULL, IS NOT NULL)。 支持的逻辑运算符是 (AND, OR)。 where_expression 是可选项。	WHERE timestampColumn = timestampColumn2 - 14
limit_expression	此表达式必须采用正整数。它也可以与 TOP 参数互换。 limit_expression 是可选项。	LIMIT 100

关于列表查询的结构和语法，请注意以下几点：

- 不支持除 SELECT 之外的 SQL 命令。
- 不支持子查询和通用表格表达式（例如 WITH）。
- 不支持 HAVING、GROUP BY 和 ORDER BY 子句
- 不支持 OFFSET 参数

列表分析规则 — 查询控制

使用列表查询控制，您可以控制如何使用表中的列来查询表。例如，您可以控制哪一列用于联接，或者 SELECT 语句和 WHERE 子句中可以使用哪一列。

下面几节解释每种控制。

主题

联接控制

使用联接控制，您可以控制如何将您的表连接到 table_expression 中的其他表。AWS Clean Rooms 仅支持 INNER JOIN。在列表分析规则中，至少需要一个 INNER JOIN，并且可以查询的成员必须在 INNER JOIN 中包含自己拥有的表。这意味着他们必须直接或通过传递方式将您的表与他们的表联接起来。

以下是传递联接的示例。

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

INNER JOIN 语句只能使用在分析规则中明确归类为 `joinColumn` 的列。

INNER JOIN 必须对您的已配置表中的 `joinColumn` 和协作中另一个已配置表中的 `joinColumn` 进行操作。您可以决定表中的哪些列可以用作 `joinColumn`。

ON 子句中的每个匹配条件都要求在两列之间使用相等比较条件 (=)。

ON 子句中的多个匹配条件可以是：

- 使用 AND 逻辑运算符组合
- 使用 OR 逻辑运算符分隔

> **Note**
>
> 所有 JOIN 匹配条件都必须与 JOIN 两侧各一条记录匹配。所有由 OR 或 AND 逻辑运算符连接的条件也必须遵守此要求。

以下是使用 AND 逻辑运算符的查询示例。

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

以下是使用 OR 逻辑运算符的查询示例。

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

控件	定义	使用量
joinColumns	您希望允许可以查询的成员在 INNER JOIN 语句中使用的列。	同一列不能同时归类为 joinColumn 和 listColumn （参阅列表控制）。 除了 INNER JOIN 之外，不能在查询的任何其他部分中使用 joinColumn 。

列表控制

列表控件控制可在查询输出中列出（即在 SELECT 语句中使用）或用于筛选结果（即在WHERE语句中使用）的列。

控件	定义	使用量
listColumns	您允许可以查询的成员在 SELECT 和 WHERE 中使用的列。	listColumn 可以在 SELECT 和 WHERE 中使用。 同一列不能同时用作 listColumn 和 joinColumn 。

列表分析规则预定义结构

以下示例包括一个预定义的结构，该结构向您展示了如何完成列表分析规则。

在以下示例中，*MyTable* 指您的数据表。您可以将每个*#######*替换为自己的信息。

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

列表分析规则 — 示例

以下示例演示了两家公司如何使用列表分析在 AWS Clean Rooms 中进行协作。

A 公司有客户关系管理 (CRM) 数据。A 公司希望获得有关其客户的更多细分数据，以进一步了解他们的客户，并有可能使用属性作为其他分析的输入。B 公司的细分数据由他们根据第一方数据创建的独特细分属性组成。B 公司只想向 A 公司提供其数据与 A 公司数据重叠的客户的唯一细分属性。

两家公司决定进行协作，以便 A 公司能够扩充重叠的数据。A 公司是可以查询的成员，B 公司是贡献者。

为创建协作并在协作中运行列表分析，两家公司执行以下操作：

1. A 公司创建协作并创建成员身份。协作中的另一个成员是 B 公司。A 公司在协作中启用查询日志记录，并在其账户中启用查询日志记录。
2. B 公司在协作中创建成员身份。它在其账户中启用查询日志记录。
3. A 公司创建 CRM 配置表。
4. A 公司将分析规则添加到客户配置表中，如以下示例所示。

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

```
}
```

joinColumns — A 公司希望使用 hashedemail 和/或 thirdpartyid（从身份供应商处获得）将 CRM 数据中的客户与细分数据中的客户进行匹配。这将有助于确保 A 公司为合适的客户匹配扩充的数据。他们有两个 JoinColumns，可以提高分析的匹配率。

listColumns — A 公司使用 listColumns 来获取他们在自己系统中使用的 internalid 旁边的扩充列。他们添加了 segment1、segment2 和 customercategory，以便通过在筛选器中使用它们，将扩充限制到特定的细分。

5. B 公司创建细分配置表。
6. B 公司将分析规则添加到细分配置表中。

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

joinColumns— B 公司让 A 公司在 identifier2 上进行联接，以便将细分数据中的客户与 CRM 数据相匹配。A 公司和 B 公司与身份供应商合作，以获得与此协作相匹配的 identifier2。他们之所以没有添加其他 joinColumns，是因为他们认为 identifier2 可以提供最高和最准确的匹配率，而且查询不需要其他标识符。

listColumns — B 公司让 A 公司使用 segment3 和 segment4 属性来扩充其数据，这些属性是他们（与客户 A）一起创建、收集和调整的独特属性，是数据扩充的一部分。他们希望 A 公司在行级获取这些重叠的细分，因为这是一项数据扩充协作。

7. A 公司创建了与协作的 CRM 表关联。
8. B 公司创建了与协作的细分表关联。
9. A 公司运行查询（例如以下查询）以扩充重叠的客户数据。

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

10A 公司和 B 公司查看查询日志。B 公司验证查询是否符合协作协议中上商定的内容。

中的自定义分析规则 AWS Clean Rooms

在中 AWS Clean Rooms，自定义分析规则是一种新型的分析规则，它允许对配置的表运行自定义查询。自定义 SQL 查询仍然仅限于只有 SELECT 命令，但与聚合和列表查询相比，可以使用更多的 SQL 构造（例如，窗口函数、OUTER JOIN、CTE 或子查询；有关完整列表，请参阅 AWS Clean Rooms SQL 参考）。自定义 SQL 查询不必遵循聚合和列表查询之类的查询结构。

与聚合和列表分析规则支持的使用案例相比，自定义分析规则支持更高级的使用案例，例如自定义归因分析、基准测试、增量分析和受众发现。这是对聚合和列表分析规则支持的使用案例的超集的补充。

自定义分析规则还支持差别隐私。差别隐私是一种在数学上非常严格的数据隐私保护框架。有关更多信息，请参阅 AWS Clean Rooms 差异隐私。创建分析模板时， AWS Clean Rooms 差异隐私会检查该模板以确定其是否与 AWS Clean Rooms 差异隐私的通用查询结构兼容。此验证可确保您不会创建不允许使用差异隐私保护表的分析模板。

要配置自定义分析规则，数据所有者可以选择允许存储在分析模板中的特定自定义查询在其配置表上运行。数据所有者在将分析模板添加到自定义分析规则中允许的分析控制之前应先审核这些模板。分析模板仅在创建这些模板的协作中可用和可见（即使该表与其他协作关联），并且只能由可以在该协作中进行查询的成员运行。

或者，成员可以选择允许其他成员（查询提供者）无需审核即可创建查询。成员在自定义分析规则添加允许的查询提供者控制的查询提供者账户。如果查询提供者是可以查询的成员，则他们可以直接在配置表上运行任何查询。查询提供者还可以通过创建分析模板来创建查询。在存在查询提供程序和关联表的所有协作中，自动允许查询提供者创建的任何查询在表上运行。 AWS 账户

数据所有者只能允许分析模板或账户创建查询，不能同时允许两者创建。如果数据所有者将其留空，则可以查询的成员将无法对配置表运行查询。

主题

- 自定义分析规则预定义结构
- 自定义分析规则示例
- 具有差别隐私的自定义分析规则

自定义分析规则预定义结构

以下示例包含一个预定义的结构，说明了如何完成开启差别隐私的自定义分析规则。`userIdentifier` 值是唯一地标识您的用户的列，例如 user_id。当您在协作中启用了差异隐私的两个或更多表时， AWS Clean Rooms 需要您在两个分析规则中配置与用户标识符列相同的列，以保持各表中用户定义的一致性。

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

您可以：

- 将分析模板 ARN 添加到允许的分析控制。在这种情况下，不包括 `allowedAnalysisProviders` 控制。

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

- 向`allowedAnalysisProviders`控件添加成员 AWS 账户 ID。在这种情况下，您可以将 ANY_QUERY 添加到 `allowedAnalyses` 控制。

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

自定义分析规则示例

以下示例演示了两家公司如何合作 AWS Clean Rooms 使用自定义分析规则。

A 公司有客户和销售数据。A 公司有兴趣了解 B 公司网站上广告活动的销售增量。B 公司拥有对 A 公司有用的观众数据和细分属性（例如，他们观看广告时使用的设备）。

A 公司想在协作中运行一个特定的增量查询。

为创建协作并在协作中运行自定义分析，两家公司执行以下操作：

1. A 公司创建协作并创建成员身份。协作中的另一个成员是 B 公司。A 公司在协作中启用查询日志记录，并在其账户中启用查询日志记录。
2. B 公司在协作中创建成员身份。它在其账户中启用查询日志记录。
3. A 公司创建 CRM 配置表。
4. A 公司向销售配置表添加空的自定义分析规则。
5. A 公司将销售配置表与协作关联起来。
6. B 公司创建观众配置表。
7. B 公司在观众配置表中添加一个空的自定义分析规则。
8. B 公司将观众配置表与协作关联起来。
9. A 公司查看与协作关联的销售表和观众表，并创建分析模板，为活动月份添加增量查询和参数。

```
{
    "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.公司 A 将其帐户（例如 444455556666）添加到自定义分析规则中允许的分析提供者控件中。他们之所以使用允许的分析提供者控制，是因为他们希望允许在销售配置表上运行他们创建的任何查询。

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

11B 公司在协作中看到创建的分析模板并查看其内容，包括查询字符串和参数。

12B 公司确定分析模板实现了增量使用案例，并满足如何查询其观众配置表的隐私要求。

13B 公司将分析模板 ARN 添加到观众表的自定义分析规则允许的分析控制中。他们之所以使用允许的分析控制，是因为他们只想允许在观众配置表上运行增量查询。

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

14A 公司运行分析模板并使用参数值 `05-01-2023`。

具有差别隐私的自定义分析规则

在中 AWS Clean Rooms，自定义分析规则支持差异隐私。差别隐私是一种在数学上非常严格的数据隐私保护框架，可以帮助保护您的数据以防范重新识别尝试。

差异隐私支持综合分析，例如广告活动规划、 post-ad-campaign 衡量、金融机构联盟中的基准测试以及医疗保健研究的 A/B 测试。

支持的查询结构和语法在 查询结构和语法 中定义。

具有差别隐私的自定义分析规则示例

考虑上一节中介绍的自定义分析规则示例。该示例说明了如何使用差别隐私保护您的数据以防范重新识别尝试，同时允许您的合作伙伴从您的数据中了解业务关键型见解。假设 B 公司具有观众数据，并希望使用差别隐私保护其数据。为了完成差别隐私设置，B 公司完成以下步骤：

1. B 公司开启差别隐私，同时在观众配置表中添加自定义分析规则。B 公司选择 `viewershipdata.hashedemail` 以作为用户标识符列。
2. B 公司在协作中添加差别隐私策略，以使其观众数据表可供查询。B 公司选择默认策略以快速完成设置。

A 公司希望了解 B 公司网站上的广告活动的销售增量，并运行分析模板。由于该查询与 AWS Clean Rooms Differential Privacy 的通用查询结构兼容，因此，查询成功运行。

查询结构和语法

包含至少一个开启了差别隐私的表的查询必须遵循以下语法。

```
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

对于差别隐私查询结构和语法，请注意以下事项：

- 不支持子查询。
- 如果表或通用表表达式 (CTE) 涉及受差别隐私保护的数据，CTE 应生成用户标识符列。应在用户级别完成筛选、分组和聚合。
- final_select 允许使用 COUNT DISTINCT、COUNT、SUM、AVG 和 STDDEV 聚合函数。

有关差别隐私支持哪些 SQL 关键字的更多详细信息，请参阅 AWS Clean Rooms 差异隐私的 SQL 功能。

AWS Clean Rooms 差异隐私

AWS Clean Rooms 差异隐私通过一种以数学为依据的技术帮助您保护用户的隐私，该技术只需单击几下即可通过直观的控件实现。作为一项完全托管的功能，无需事先体验差异化隐私即可帮助您防止重新识别用户。 AWS Clean Rooms 在运行时自动向查询结果添加经过精心校准的噪音量，以帮助保护您的个人级别数据。

AWS Clean Rooms Difersition Privacy 支持广泛的分析查询，非常适合各种用例，在这些用例中，查询结果中的少量错误不会影响分析的实用性。通过使用该功能，您的合作伙伴可以生成有关广告活动、投资决策、临床研究等的业务关键型见解，合作伙伴无需进行任何额外的设置。

AWS Clean Rooms 差异隐私可防止恶意使用标量函数或数学运算符符号的溢出或无效强制转换错误。

有关 AWS Clean Rooms 差分隐私的更多信息，请参阅以下主题。

主题

差别隐私

差别隐私仅允许聚合的见解，并掩盖任何个人数据在这些见解中的贡献。差别隐私保护协作数据，以防止可以接收结果的成员了解特定个人的数据。如果没有差别隐私，可以接收结果的成员可能会尝试添加或删除有关个人的记录，并观察查询结果差异以推断个人用户数据。

在开启差别隐私后，将在查询结果中添加指定数量的噪声以掩盖各个用户的贡献。如果能够接收结果的成员在从其数据集中删除有关个人的记录后试图观察查询结果的差异，则查询结果的可变性有助于阻止识别该个人的数据。 AWS Clean Rooms Difersial Privacy 使用SampCert采样器，这是由开发的经过验证的正确采样器实现。 AWS

差分隐私 AWS Clean Rooms 的工作原理

在完成以下工作流程时，开启差异隐私的工作流程 AWS Clean Rooms需要执行以下额外步骤 AWS Clean Rooms：

1. 在添加自定义分析规则时，您可以开启差别隐私。
2. 您为协作配置差别隐私策略，以使受差别隐私保护的数据表可供查询。

完成这些步骤后，可以查询的成员可以开始对受差异隐私保护的数据进行查询。 AWS Clean Rooms 返回符合差异隐私政策的结果。 AWS Clean Rooms Differation Privacy 会跟踪您可以运行的剩余查询的估计数量，类似于显示汽车当前燃油水平的汽车中的汽油表。可以查询的成员可以运行的查询数量受差别隐私策略中设置的隐私预算和每个查询添加的噪声参数的限制。

注意事项

在中使用差分隐私时 AWS Clean Rooms，请考虑以下几点：

- 能够收到结果的成员不能使用差异隐私。他们将为配置的表配置自定义分析规则，并关闭差别隐私。
- 如果两个或更多数据提供者都开启了差别隐私，可以查询的成员无法联接来自这些数据提供者的表。

差别隐私策略

差别隐私策略控制允许可以查询的成员在协作中运行多少个聚合函数。隐私预算定义一种通用的有限资源，该资源应用于协作中的所有表。每个查询添加的噪声控制隐私预算的耗尽速率。

需要具有差别隐私策略，才能使受差别隐私保护的表可供查询。这是协作中的一次性步骤，其中包括两个输入：

- 隐私预算 - 以 epsilon 量化，隐私预算控制隐私保护级别。这是一种通用的有限资源，应用于协作中受差别隐私保护的所有表，因为目标是保护可能在多个表中包含信息的用户的隐私。

  每次对表运行查询时，都会使用隐私预算。在隐私预算用完时，可以查询的协作成员无法运行额外的查询，直到增加或刷新隐私预算。通过设置较大的隐私预算，可以接收结果的成员可以减少他们对数据中的个人的不确定性。在咨询业务决策者后，选择一个兼顾您的协作要求和隐私需求的隐私预算。

  如果您计划定期将新数据引入到一个协作中，您可以选择每月刷新隐私预算，以在每个日历月自动创建新的隐私预算。如果选择该选项，在两次刷新之间重复查询时，可能会泄露任意数量的数据行相关信息。如果在隐私预算刷新之间重复查询相同的行，请避免选择该选项。

- 每个查询添加的噪声是根据您希望掩盖其贡献的用户数量测量的。该值控制隐私预算的耗尽速率。较大的噪声值降低隐私预算的耗尽速率，因此，允许对数据运行更多查询。不过，这会导致发布的数据见解不太准确。在设置该值时，请考虑协作见解所需的准确性。

您可以使用默认的差异隐私策略来快速完成设置或根据您的用例自定义差异隐私政策。 AWS Clean Rooms 差异隐私提供了用于配置策略的直观控件。 AWS Clean Rooms Difersition Privacy 允许您根据数据的所有查询中可能的聚合数量来预览该实用程序，并估算在数据协作中可以运行多少查询。

您可以使用交互式示例，以了解隐私预算和每个查询添加的噪声的不同值如何影响不同类型的 SQL 查询的结果。一般来说，您需要兼顾隐私需求以及要允许的查询数量和这些查询的准确性。较小的隐私预算或较大的每个查询添加的噪声可以更好地保护用户隐私，但为协作合作伙伴提供不太有意义的见解。

如果您增加隐私预算，同时将每个查询添加的噪声参数保持不变，则可以查询的成员可以在协作中对您的表运行更多的聚合。您可以在协作期间随时增加隐私预算。如果您减少隐私预算，同时将每个查询添加的噪声参数保持不变，则可以查询的成员可以运行更少的聚合。在可以查询的成员开始分析您的数据后，您无法减少隐私预算。

如果您增加每个查询添加的噪声，同时将隐私预算输入保持不变，则可以查询的成员可以在协作中对您的表运行更多的聚合。如果您减少每个查询添加的噪声，同时将隐私预算输入保持不变，则可以查询的成员可以运行更少的聚合。您可以在协作期间随时增加或减少每个查询添加的噪声。

差别隐私策略是通过隐私预算模板 API 操作管理的。

AWS Clean Rooms 差异隐私的 SQL 功能

AWS Clean Rooms 差异隐私使用通用查询结构来支持复杂的 SQL 查询。根据此结构对自定义分析模板进行验证，以确保它们可以在受差异隐私保护的表上运行。下表指示支持哪些函数。请参阅查询结构和语法了解更多信息。

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
聚合函数	• ANY_VALUE 函数 • APPROXIMATE PERCENTILE_DISC 函数 • AVG 函数	支持使用差异隐私保护表的 CTE 必须生成包含用户级记录的数据。你应该使用`SELECT userIdentifierColumn...' 格式在那些	支持的聚合：AVG、COUNT、COUNT DISTINCT、STDDEV 和 SUM。

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
	• COUNT 和 COUNT DISTINCT 函数 • LISTAGG 函数 • MAX 函数 • MEDIAN 函数 • MIN 函数 • PERCENTILE_CONT 函数 • STDDEV_SAMP 和 STDDEV_POP 函数 • SUM 和 SUM DISTINCT 函数 • VAR_SAMP 和 VAR_POP 函数	CTE 中编写 SELECT 表达式。	
CTE	WITH 子句、WITH 子句子查询	支持使用差异隐私保护表的 CTE 必须生成包含用户级记录的数据。你应该使用`SELECT userIdentifierColumn...' 格式在那些 CTE 中编写 SELECT 表达式。	不适用
子查询	选择列表子查询、FROM 子句子查询、WHERE 子句子查询	不支持。不支持查询中引用开启差分隐私功能的表的子查询。将您的子查询重写为公用表表达式 (CTE)。	

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
联接条款	• INNER JOIN • LEFT JOIN • RIGHT JOIN • FULL JOIN • [JOIN] OR 运算符 • CROSS JOIN	支持的条件是，仅支持对用户标识符列进行等值联接的 JOIN 函数；在查询两个或更多开启了差别隐私的表时，必须使用这些函数。确保必需的等值联接条件是正确的。确认表所有者在所有表中配置了相同的用户标识符列，以便用户的定义在表之间保持一致。 在合并两个或更多开启了差别隐私的关系时，不支持 CROSS JOIN 函数。	
集合运算符	UNION、UNION ALL、INTERSECT、除外 | 减号（这些是同义词）	全部支持	不支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
窗口函数	聚合函数 • AVG 窗口函数 • COUNT 窗口函数 • CUME_DIST 开窗函数 • DENSE_RANK 窗口函数 • FIRST_VALUE 窗口函数 • LAG 窗口函数 • LAST_VALUE 窗口函数 • LEAD 窗口函数 • MAX 窗口函数 • MEDIAN 窗口函数 • MIN 窗口函数 • NTH_VALUE 窗口函数 • RATIO_TO_REPORT 开窗函数 • STDDEV_SAMP 和 STDDEV_POP 窗口函数（STDDEV_SAMP 和 STDDEV 是同义词） • SUM 窗口函数 • VAR_SAMP 和 VAR_POP 窗口函数（VAR_SAMP 和	在查询开启差异隐私的关系时，窗口函数的分区子句中的用户标识符列是必填的，条件是所有这些都支持。	不支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
	VARIANCE 是同义词)		
	排名函数 • DENSE_RANK 窗口函数 • NTILE 窗口函数 • PERCENT_RANK 开窗函数 • RANK 窗口函数 • ROW_NUMBER 窗口函数		
条件表达式	• CASE 条件表达式 • COALESCE 表达式 • GREATEST 和 LEAST 函数 • NVL 和 COALESCE 函数 • NVL2 函数 • NULLIF 函数	全部支持	全部支持
Conditions	• 比较条件 • 逻辑条件 • 模式匹配条件 • BETWEEN 范围条件 • Null 条件	EXISTS并且IN不能使用，因为它们需要子查询。支持所有其他内容。	全部支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
日期时间函数	• 事务中的日期和时间函数 • 串联运算符 • ADD_MONTHS 函数 • CONVERT_TIMEZONE 函数 • CURRENT_DATE 函数 • DATEADD 函数 • DATEDIFF 函数 • DATE_PART 函数 • DATE_TRUNC 函数 • EXTRACT 函数 • GETDATE 函数 • TIMEOFDAY 函数 • TO_TIMESTAMP 函数 • 日期或时间戳函数的日期部分	全部支持	全部支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
字符串函数	• ||（串联）运算符 • BTRIM 函数 • CHAR_LENGTH 函数 • CHARACTER _LENGTH 函数 • CHARINDEX 函数 • CONCAT 函数 • LEFT 和 RIGHT 函数 • LEN 函数 • LENGTH 函数 • LOWER 函数 • LPAD 和 RPAD 函数 • LTRIM 函数 • POSITION 函数 • REGEXP_COUNT 函数 • REGEXP_INSTR 函数 • REGEXP_RE PLACE 函数 • REGEXP_SUBSTR 函数 • REPEAT 函数 • REPLACE 函数 • REPLICATE 函数 • REVERSE 函数 • RTRIM 函数	全部支持	全部支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
	• SOUNDEX 函数 • SPLIT_PART 函数 • STRPOS 函数 • SUBSTRING 函数 • TEXTLEN 函数 • TRANSLATE 函数 • TRIM 函数 • UPPER 函数		
数据类型格式设置函数	• CAST 函数 • TO_CHAR • TO_DATE 函数 • TO_NUMBER • 日期时间格式字符串 • 数字格式字符串	全部支持	全部支持
哈希函数	• MD5 函数 • SHA 函数 • SHA1 函数 • SHA2 函数 • MURMUR3_32_HASH	全部支持	全部支持
数学运算符符号	+、-、*、/、% 和 @	全部支持	全部支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
数学函数	• ABS 函数 • ACOS 函数 • ASIN 函数 • ATAN 函数 • ATAN2 函数 • CBRT 函数 • CEILING（或 CEIL）函数 • COS 函数 • COT 函数 • DEGREES 函数 • DEXP 函数 • LTRIM 函数 • DLOG1 函数 • DLOG10 函数 • EXP 函数 • FLOOR 函数 • LN 函数 • LOG 函数 • MOD 函数 • PI 函数 • POWER 函数 • RADIANS 函数 • RANDOM 函数 • ROUND 函数 • SIGN 函数 • SIN 函数 • SQRT 函数 • TRUNC 函数	全部支持	全部支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
SUPER 类型信息函数	• DECIMAL_PRECISION 函数 • DECIMAL_SCALE 函数 • IS_ARRAY 函数 • IS_BIGINT 函数 • IS_CHAR 函数 • IS_DECIMAL 函数 • IS_FLOAT 函数 • IS_INTEGER 函数 • IS_OBJECT 函数 • IS_SCALAR 函数 • IS_SMALLINT 函数 • IS_VARCHAR 函数 • JSON_TYPEOF 函数	全部支持	全部支持
VARBYTE 函数	• FROM_HEX 函数 • FROM_VARBYTE 函数 • TO_HEX 函数 • TO_VARBYTE 函数	全部支持	全部支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
JSON	• CAN_JSON_PARSE 函数 • JSON_EXTRACT_ARRAY_ELEMENT_TEXT 函数 • JSON_EXTRACT_PATH_TEXT 函数 • JSON_PARSE 函数 • JSON_SERIALIZE 函数 • JSON_SERALIZE_TO_VARBYTE 函数	全部支持	全部支持
数组函数	• 数组函数 • array_concat 函数 • array_flatten 函数 • get_array_length 函数 • split_to_array 函数 • 子数组函数	不支持	不支持
扩展分组依据	分组集、汇总、立方体	不支持	不支持
排序操作	ORDER BY	支持 ORDER BY 子句，条件是只有在开启差分隐私的情况下查询表时，窗口函数的分区子句才支持 ORDER BY 子句。	支持

短名称	SQL 构造	通用表表达式 (CTE)	最终 SELECT 子句
行数限制	LIMIT、OFFSET	使用差异隐私保护表的 CTE 不支持	全部支持
表和列别名		支持	支持
聚合函数上的数学函数		支持	支持
聚合函数中的标量函数		支持	支持

不支持的 SQL 构造的常见替代方案

类别	SQL 构造	或者
窗口函数	• LISTAGG • PERCENTILE_CONT • PERCENTILE_DISC	您可以将等效的聚合函数与 GROUP BY 一起使用。
数学运算符符号	• $column ||/ 2 • $column |/ 2 • $column ^ 2	• CBRT • SQRT • POWER($column, 2)
标量函数	• SYSDATE • $column::integer • convert(type, $column)	• CURRENT_DATE • CAST $column AS integer • CAST $column AS type
文本	间隔 '1 秒'	间隔 '1' 秒
行限制	TOP n	限制 n
联接	• USING • NATURAL	ON 子句应明确包含连接标准。

Differential Privacy 查询技巧和示例

AWS Clean Rooms 差异隐私使用通用查询结构来支持各种 SQL 结构，例如用于数据准备的公用表表达式 (CTE) 和常用的聚合函数，例如COUNT、或。SUM为了通过在运行时向聚合查询结果添加噪音来混淆任何可能的用户在数据中的贡献，Difersient Privacy 要求最终SELECT statement版本中的聚合函数在用户级数据上运行。 AWS Clean Rooms

以下示例使用来自一个媒体发布者的两个名为 socialco_impressions 和 socialco_users 的表，该发布者希望使用差别隐私保护数据，同时与一个具有 athletic_brand_sales 数据的运动品牌协作。该媒体发布者已将 user_id 列配置为用户标识符列，同时在 AWS Clean Rooms中启用差别隐私。广告商不需要差别隐私保护，并希望使用 CTE 对合并的数据运行查询。由于他们的 CTE 使用受差别隐私保护的表，因此，广告商将这些受保护的表中的用户标识符列包含在 CTE 列的列表中，并根据用户标识符列联接这些受保护的表。

```
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
```

同样，如果要对受差别隐私保护的数据表运行窗口函数，您必须在 PARTITION BY 子句中包含用户标识符列。

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

AWS Clean Rooms 差异隐私的局限性

AWS Clean Rooms 差异隐私不能解决以下情况：

1. AWS Clean Rooms 差分隐私不能解决定时攻击。例如，如果单个用户贡献大量的行，并且添加或删除该用户显著改变查询计算时间，则可能会受到这些攻击。
2. 当 SQL 查询由于使用某些 SQL 结构而在运行时可能导致溢出或无效的强制转换错误时，AWS Clean Rooms 差异隐私不能保证差异隐私。下表列出了一些（但不是全部）SQL 结构，这些结构可能会产生运行时错误，应在分析模板中进行验证。我们建议您批准能够最大限度地减少出现此类运行时错误机会的分析模板，并定期查看查询日志，以确定查询是否符合协作协议。

   以下 SQL 结构容易出现溢出错误：

   - 聚合函数-AVG、LISTAVG、PERCENTILE_COUNT、PERCETILE_DISC、SUM/SUM_DISTINCT
   - 数据类型格式化函数-TO_TIMESTAMP、TO_DATE
   - 日期和时间函数-ADD_MONTHS、DATEADD、DATEDIFF
   - 数学函数-+、-、*、/、POWER
   - 字符串函数-||、CONCAT、重复、复制
   - 窗口函数——AVG、LISTAGG、PERCENTILE_COUNT、PERCENTILE_DISC、RATIO_TO_REPORT、SUM

   CAST 数据类型格式化函数容易出现无效的强制转换错误。

AWS Clean Rooms ML

AWS Clean Rooms ML

AWS Clean Rooms 机器学习为双方提供了一种隐私保护方法，可以识别其数据中的相似用户，而无需彼此共享数据。第一方将训练数据带到， AWS Clean Rooms 这样他们就可以创建和配置外观相似的模型并将其与协作关联起来。然后，第二方将其种子数据带到 AWS Clean Rooms 并生成类似于训练数据的相似区段。

有关其工作方式的更详细说明，请参阅跨账户作业。

- 训练数据提供者 - 贡献训练数据、创建和配置相似模型并将该相似模型与一个协作关联的一方。
- 种子数据提供者 - 贡献种子数据、生成相似细分并导出其相似细分的一方。
- 训练数据 - 训练数据提供者的数据，用于生成相似模型。训练数据用于测量用户行为的相似性。

  训练数据必须包含用户 ID、项目 ID 和时间戳列。（可选）训练数据可以包含其他交互作为数值或分类特征。举例而言，交互可以是观看的视频、购买的物品或阅读的文章列表。
- 种子数据 - 种子数据提供者的数据，用于创建相似细分。相似细分输出是训练数据中与种子用户最相似的一组用户。
- 相似模型 - 训练数据的机器学习模型，用于在其他数据集中查找相似用户。

  在使用 API 时，受众模型 术语等同于相似模型。例如，您可以使用CreateAudience模型 API 来创建外观相似的模型。
- Lookalik e segment — 与种子数据最为相似的训练数据子集。

  使用 API 时，您可以使用 StartAudienceGenerationJobAPI 创建相似的区段。

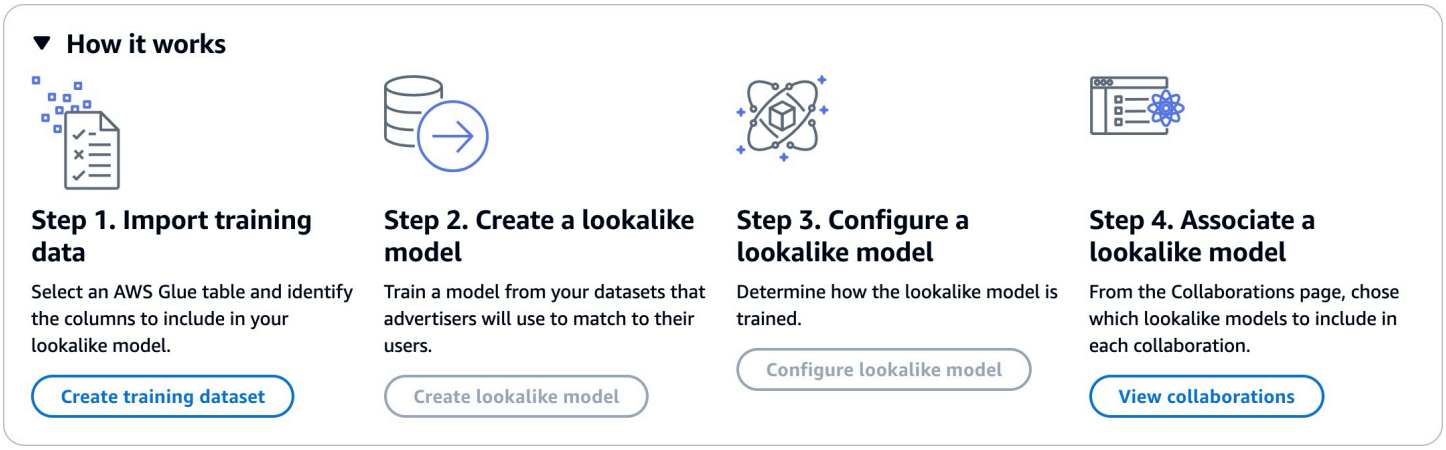
训练数据提供者的数据绝不会与种子数据提供者共享，并且种子数据提供者的数据绝不会与训练数据提供者共享。相似细分输出与训练数据提供者共享，但绝不会与种子数据提供者共享。

有关相似模型的更多信息，请参阅以下主题。

主题

- AWS Clean Rooms 机器学习的工作原理

AWS Clean Rooms 机器学习的工作原理



Clean Rooms ML 要求两方，即训练数据提供者和种子数据提供者，按顺序协作，将其数据整合到协作中。 AWS Clean Rooms 以下是训练数据提供者必须先完成的工作流程：

1. 训练数据提供者的数据必须存储在用户-项目交互 AWS Glue 的数据目录表中。训练数据必须至少包含用户 ID 列、交互 ID 列和时间戳列。
2. 训练数据提供者向注册训练数据 AWS Clean Rooms。
3. 训练数据提供者创建一个相似模型，可以将其与多个种子数据提供者共享。相似模型是一种深度神经网络，训练时间可能长达 24 小时。该模型不会自动重新训练，我们建议您每周重新训练一次。
4. 训练数据提供者配置相似模型，包括是否共享相关性指标以及输出细分的 Amazon S3 位置。训练数据提供者可以通过单个相似模型创建多个配置的相似模型。
5. 训练数据提供者将配置的受众模型关联到与某个种子数据提供者共享的协作。

以下是种子数据提供者接下来必须完成的工作流程：

1. 种子数据提供者的数据必须存储在 Amazon S3 存储桶中。
2. 种子数据提供者开启与训练数据提供者共享的协作。
3. 种子数据提供者从协作页面的 “Clean Rooms ML” 选项卡中创建一个相似的区段。
4. 种子数据提供者可以评估相关性指标（如果已共享），并导出相似细分以在 AWS Clean Rooms外部使用。

AWS Clean Rooms 机器学习的隐私保护

Clean Rooms ML 旨在降低成员资格推断攻击的风险，在这种攻击中，训练数据提供者可以了解谁在种子数据中，种子数据提供者可以了解谁在训练数据中。我们采取了一些措施以防范这种攻击。

首先，种子数据提供者不直接观察 Clean Rooms ML 输出，训练数据提供者也永远无法观察种子数据。种子数据提供者可以选择将种子数据包含在输出细分中。

接下来，通过训练数据的随机样本创建相似模型。该样本包含大量与种子受众不匹配的用户。此过程使得确定用户是否不在数据中变得更加困难，这是推断成员资格的另一种途径。

此外，可以在种子特定的相似模型训练的每个参数中使用多个种子客户。这限制了模型可以过度拟合的程度，从而限制了可以推断的用户相关数据量。因此，我们建议种子数据的最小大小为 500 个用户。

最后，一定不要向训练数据提供者提供用户级指标，这可以阻断成员身份推断攻击的另一种途径。

AWS Clean Rooms 机器学习模型评估指标

Clean Rooms ML 会计算召回率和相关性分数，以确定模型的表现如何。Recall 比较了相似数据和训练数据之间的相似性。相关性分数用于决定受众应该有多大，而不是模型是否表现良好。

召回率是衡量相似区段与训练数据的相似程度的公正衡量标准。召回率是受众生成作业包含在种子受众中的训练数据样本中最相似的用户（默认为最相似的20%）的百分比。值范围为 0-1，值越大表示受众越好。召回值大致等于最大分区百分比表示受众模型等同于随机选择。

我们认为这是比准确性、精度和 F1 分数更好的评估指标，因为 Clean Rooms ML 在构建模型时没有准确标记真正的负面用户。

细分级相关性分数 是一个相似性指标，值范围从 -1（最不相似）到 1（最相似）。Clean Rooms ML 会针对各种区段大小计算一组相关性分数，以帮助您确定数据的最佳区段大小。随着区段大小的增加，相关性分数会单调降低，因此，随着区段大小的增加，它可能与种子数据不太相似。在细分级相关性分数达到 0 时，模型预测相似细分中的所有用户来自与种子数据相同的分布。增加输出大小可能会包括相似细分中来自与种子数据不同的分布的用户。

相关性分数是在单个活动中标准化的，不应用于比较不同的活动。相关性分数不应用作任何业务成果的单一来源证据，因为除了相关性以外，它们还受到多种复杂因素的影响，例如库存质量、库存类型、广告时间等。

相关性分数不应用于判断种子质量，而应用于判断它是否可以增加或减少。考虑以下示例：

- 全部为正分 - 这表明预测为相似的输出用户比相似细分中包含的用户多。这对于属于大型市场的种子数据来说很常见，例如，过去一个月内购买过牙膏的每个人。我们建议查看较小的种子数据，例如，过去一个月内多次购买牙膏的每个人。
- 对于你想要的相似区段大小，所有负分或负数 — 这表明 Clean Rooms ML 预测在所需的相似区段大小中没有足够的相似用户。这可能是因为，种子数据太具体或市场太小。我们建议为种子数据应用更少的筛选条件，或者扩大市场。例如，如果原始种子数据是购买婴儿车和汽车座椅的客户，您可以将市场扩大到购买多种婴儿产品的客户。

训练数据提供者确定是否公开相关性分数以及计算相关性分数的桶区间。

使用 AWS Clean Rooms 机器学习

相似模型 是训练数据提供者的数据的模型，它允许种子数据提供者创建训练数据提供者数据的相似细分，该细分与其种子数据最相似。要创建可以在协作中使用的相似模型，您必须导入训练数据，创建相似模型，配置该相似模型，然后将其与一个协作相关联。

在训练数据提供者创建完机器学习模型后，种子数据提供者可以创建并导出种子细分。

主题

使用相似模型（训练数据提供者）

导入训练数据

在创建相似模型之前，必须指定包含训练数据的 AWS Glue 表。Clean Rooms ML 不存储这些数据的副本，只存储允许其访问数据的元数据。

要在中导入训练数据 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择机器学习建模。
3. 在训练数据集选项卡上，选择创建训练数据集。

4. 输入名称和可选描述。
5. 对于数据源，请选择您的 AWS Glue 表：

   a. 从下拉列表中选择要配置的数据库。

   b. 通过从下拉列表中选择要配置的数据库和表来选择训练数据源。

> **Note**
>
> 要验证是否是正确的表，请执行以下任一操作：
>
> - 选择 “在” 中查看 AWS Glue。
> - 打开查看架构以查看架构。

6. 要了解培训详情，请从您的数据中选择用户标识符列、项目标识符列和时间戳列。训练数据必须包含这三列。您也可以选择要在训练数据中包含的任何其他列。

   “时间戳” 列中的数据必须采用 Unix 纪元时间（以秒为单位）格式。
7. 在服务访问中，您必须指定可以访问您的数据的服务角色，如果您的数据已加密，则必须提供 KMS 密钥。选择创建并使用新的服务角色，Clean Rooms ML 将自动创建服务角色并添加必要的权限策略。如果您要使用特定的服务角色，请选择 “使用现有服务角色”，然后在 “服务角色名称” 字段中输入该服务角色。

   如果您的数据已加密，请在AWS KMS key字段中输入您的 KMS 密钥，或者单击 “创建” AWS KMS key以生成新的 KMS 密钥。
8. 如果要为训练数据集启用标签，请选择添加新标签，然后输入键和值对。
9. 选择创建训练数据集。

有关相应的 API 操作，请参阅CreateTraining数据集。

创建相似模型

在创建训练数据集后，您就可以创建相似模型了。您可以通过单个训练数据集创建很多相似模型。

您必须在中创建默认数据库， AWS Glue Data Catalog 或者在提供的角色中包含该`glue:createDatabase`权限。

要在中创建外观相似的模型 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择机器学习建模。
3. 在相似模型选项卡上，选择创建相似模型。
4. 对于创建相似模型，为相似模型详细信息：
    a. 输入名称和可选描述。
    b. 从下拉列表中选择要建模的训练数据集。
    c. 输入可选的训练窗口。
5. 如果要为相似模型启用自定义加密设置，请选择自定义加密设置，然后输入 KMS 密钥。
6. 如果要为相似模型启用标签，请选择添加新标签，然后输入键和值对。
7. 选择创建相似模型。

有关相应的 API 操作，请参阅CreateAudience模型。

配置相似模型

在创建相似模型后，您就可以对其进行配置以在协作中使用。您可以通过单个相似模型创建多个配置的相似模型。

要在中配置外观相似的模型 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择机器学习建模。
3. 在配置的相似模型选项卡上，选择配置相似模型。
4. 对于配置相似模型，为配置的相似模型详细信息：
    a. 输入名称和可选描述。
    b. 从下拉列表中选择您要配置的相似模型。
    c. 选择您希望的最小匹配种子大小。这是种子数据提供者数据中与训练数据中的用户重叠的最小用户数。此值必须大于 0。
5. 对于与其他成员共享的指标，选择您是否希望协作中的种子数据提供者接收模型指标，包括相关性分数。

6. 对于相似细分目标位置，输入将相似细分导出到的 Amazon S3 存储桶。此存储桶必须与您的其他资源位于同一区域。
7. 对于服务访问，选择将用于访问该表的现有服务角色名称。
8. 选择 “配置相似模型”。
9. 如果要为已配置的表资源启用标签，请选择添加新标签，然后输入键和值对。

有关相应的 API 操作，请参阅CreateConfiguredAudienceModel。

关联配置的相似模型

在配置相似模型后，您可以将其与一个协作相关联。

将配置的相似模型关联到 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 在具有活跃成员身份选项卡上，选择一个协作。
4. 在机器学习建模选项卡上，选择关联相似模型。
5. 对于关联配置的相似模型，为关联相似模型详细信息：
   a. 输入关联的配置受众模型的名称。
   b. 输入表的描述。

      该描述有助于区分具有相似名称的其他关联的配置受众模型。
6. 对于配置的相似模型，从下拉列表中选择一个配置的相似模型。
7. 选择关联。

有关相应的 API 操作，请参阅CreateConfiguredAudienceModel关联。

更新已配置的相似模型

关联已配置的相似模型后，您可以对其进行更新以更改诸如名称、要共享的指标或输出 Amazon S3 位置之类的信息。

要在中更新相关配置的相似模型 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择 ML 建模。
3. 在配置的相似模型选项卡上，选择已配置的相似模型，然后选择编辑。
4. 对于配置相似模型，为配置的相似模型详细信息：
   a. 从下拉列表中选择要配置的相似模型。
   b. 选择您希望的最小匹配种子大小。这是种子数据提供者数据中与训练数据中的用户重叠的最小用户数。此值必须大于 0。
5. 对于与其他成员共享的指标，选择您是否希望协作中的种子数据提供者接收模型指标，包括相关性分数。
6. 对于相似细分目标位置，输入将相似细分导出到的 Amazon S3 存储桶。此存储桶必须与您的其他资源位于同一区域。
7. 对于服务访问，选择将用于访问该表的现有服务角色名称。
8. 对于高级素材箱大小配置，请选择要如何配置受众素材箱大小。
9. 选择保存更改。

有关相应的 API 操作，请参阅UpdateConfiguredAudienceModel。

使用相似的区段（种子数据提供者）

创建相似细分

相似细分是与种子数据最相似的训练数据子集。

要在中创建外观相似的区段 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 在具有活跃成员身份选项卡上，选择一个协作。
4. 在机器学习建模选项卡上，选择创建相似细分。
5. 对于创建相似细分，为相似细分详细信息输入名称和可选描述。

6. 对于种子配置文件，选择存储种子数据的 Amazon S3 输入源。
7. 对于服务访问，选择将用于访问该表的现有服务角色名称。
8. 如果要为训练数据集启用标签，请选择添加新标签，然后输入键和值对。
9. 选择创建相似细分。

有关相应的 API 操作，请参阅StartAudienceGenerationJob。

导出相似细分

在创建相似细分后，您可以将该数据导出到一个 Amazon S3 存储桶。

要在中导出相似的区段 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 在具有活跃成员身份选项卡上，选择一个协作。
4. 在机器学习建模选项卡上，选择一个相似细分，然后选择导出。
5. 对于导出相似模型，为导出相似模型详细信息输入名称和可选描述。
6. 对于细分大小，选择导出的细分所需的大小。
7. 选择导出。

有关相应的 API 操作，请参阅StartAudienceExportJob。

后续步骤

您现在创建了相似模型并导出了种子细分，您已准备好：

- Manage AWS Clean Rooms

Clean Rooms 加密计算

Clean Rooms(C3R) 的加密计算是一种除了分析规则之外还可以 AWS Clean Rooms 使用的功能。借助 C3R，组织可以将敏感数据整合在一起，从数据分析中获得新的见解，同时以加密方式限制任何一方在流程中可以了解到的信息。C3R 可供想要协作处理其敏感数据但只需要在云中使用加密数据的两方或多方使用。

C3R 加密客户端是一种客户端加密工具，您可以使用它来加密数据以供使用。 AWS Clean Rooms使用 C3R 加密客户端时，数据在 AWS Clean Rooms 协作中使用时仍会受到加密保护。与常规 AWS Clean Rooms 协作一样，输入数据是关系数据库表，计算以 SQL 查询表示。但是，C3R 仅支持对加密数据的有限 SQL 查询子集。

具体而言，C3R 对受加密保护的数据支持 SQL JOIN 和 SELECT 语句。输入表中的每列只能用于以下 SQL 语句类型之一：

- 在 JOIN 语句中使用的受加密保护的列称为 fingerprint 列。
- 在 SELECT 语句中使用的受加密保护的列称为 sealed 列。
- 在 JOIN 或 SELECT 语句中使用的受加密保护的列称为 cleartext 列。

在某些情况下，fingerprint 列上支持 GROUP BY 语句。有关更多信息，请参阅 Fingerprint 列。目前，C3R 不支持对加密数据使用其他 SQL 构造，例如 WHERE 子句或 SUM 和 AVERAGE 之类的聚合函数，即使相关分析规则允许使用这些构造。

C3R 旨在保护表中各个单元格中的数据。使用 C3R 的默认配置，当内容在 AWS Clean Rooms中使用时，客户通过协作向第三方提供的底层数据将保持加密。C3R 对所有 sealed 列使用行业标准 AES-GCM 加密，并使用行业标准伪随机函数（称为 HMAC 散列消息认证码）来保护 fingerprint 列。

尽管 C3R 会对表中的数据进行加密，但仍可以推断出以下信息：

- 有关表本身的信息，包括表中的列数、列名和行数。
- 与大多数标准加密形式一样，C3R 不会尝试隐藏加密值的长度。C3R 确实提供了填充加密值以隐藏明文确切长度的功能。但是，仍然可以向另一方揭示每列明文长度的上限。
- 日志级别的信息，例如何时将特定行添加到加密的 C3R 表中。

有关 C3R 的更多信息，请参阅以下主题。

主题

使用 Clean Rooms 加密计算时的注意事项

Clean Rooms 加密计算 (C3R) 旨在最大限度地保护数据。但是，某些使用案例可能会受益于较低级别的数据保护，以换取额外的功能。您可以通过修改 C3R 最安全的配置来做出这些特定的权衡。作为客户，您应该了解这些权衡，并确定它们是否适合您的使用案例。要考虑的权衡包括以下内容：

主题

- 允许在表中混合 cleartext 和加密数据
- 允许 fingerprint 列中有重复值
- 放宽对 fingerprint 列命名方式的限制
- 确定 NULL 值的表示方式

有关如何为这些场景设置参数的更多信息，请参阅加密计算参数。

允许在表中混合 cleartext 和加密数据

对所有数据进行客户端加密可最大限度地保护数据。但是，这限制了某些类型的查询（例如，SUM 聚合函数）。允许 cleartext 数据的风险在于，任何有权访问加密表的人都可以推断出一些有关加密值的信息。这可以通过对 cleartext 和关联数据进行统计分析来实现。

例如，假设您的列为 `City` 和 `State`。`City` 列为 cleartext，`State` 列加密。当您看到 `City` 列中的 `Chicago` 值时，这有助于您确定 `State` 很有可能是 `Illinois`。相比之下，如果一列是 `City`，另一列是 `EmailAddress`，则 cleartext `City` 不太可能揭示加密 `EmailAddress` 的任何信息。

有关此场景的参数的更多信息，请参阅允许 cleartext 列参数。

允许 fingerprint 列中有重复值

对于最安全的方法，我们假设任何 fingerprint 列都只包含一个变量实例。fingerprint 列中的任何项目都不能重复。C3R 加密客户端将这些 cleartext 值映射为与随机值无法区分的唯一值。因此，不可能从这些随机值中推断出 cleartext 信息。

fingerprint 列中有重复值的风险在于，重复的值会导致重复的随机值。因此，从理论上讲，任何有权访问加密表的人都可以对可能揭示 cleartext 值信息的 fingerprint 列进行统计分析。

同样，假设 fingerprint 列是 `State`，并且表中的每一行都对应一个美国家庭。通过频率分析，人们很有可能推断出哪个州是 `California`，哪个州是 `Wyoming`。这种推断是可能的，因为 `California` 的居民人数远远超过 `Wyoming`。相比之下，假设 fingerprint 列位于家庭标识符上，在包含数百万个条目的数据库中，每个家庭出现 1 到 4 次。频率分析不太可能揭示任何有用的信息。

有关此场景的参数的更多信息，请参阅“允许重复”参数。

放宽对 fingerprint 列命名方式的限制

默认情况下，我们假设当使用加密 fingerprint 列联接两个表时，这些列在每个表中的名称相同。此结果的技术原因是，默认情况下，我们派生出不同的加密密钥来加密每个 fingerprint 列。该密钥源自协作共享密钥和列名的组合。如果我们尝试联接具有不同列名的两列，则会派生出不同的密钥，并且无法计算出有效的联接。

要解决这个问题，可以关闭从每个列名派生密钥的功能。然后，C3R 加密客户端对所有 fingerprint 列使用一个派生密钥。风险在于可以进行另一种可能揭示信息的频率分析。

让我们再次以 `City` 和 `State` 为例。如果我们为每个 fingerprint 列派生相同的随机值（不包含列名）。`New York` 在 `City` 和 `State` 列中的随机值相同。纽约是美国为数不多的 `City` 名称与 `State` 名称相同的城市之一。相比之下，如果数据集的每一列都有完全不同的值，则不会泄露任何信息。

有关此场景的参数的更多信息，请参阅“允许对具有不同名称的列进行 JOIN”参数。

确定 NULL 值的表示方式

您可以选择是否像处理其他值一样对 NULL 值进行加密处理（加密和 HMAC）。如果您不像处理其他值一样处理 NULL 值，可能会揭示信息。

例如，假设 cleartext 中 `Middle Name` 列中的 NULL 表示没有中间名的人。如果您不加密这些值，则会泄露加密表中哪些行用于没有中间名的人。对于某些人群中的某些人来说，这些信息可能是一个识别信号。但是，如果您对 NULL 值进行加密处理，某些 SQL 查询的行为就会有所不同。例如，GROUP BY 子句不会将 fingerprint 列中的 fingerprintNULL 值分组在一起。

有关此场景的参数的更多信息，请参阅“保留 NULL 值”参数。

Clean Rooms 加密计算中支持的文件和数据类型

C3R 加密客户端可识别以下文件类型：

- CSV 文件
- Parquet 文件

您可以在 C3R 加密客户端中使用 `--fileFormat` 标志来明确指定文件格式。如果明确指定，则文件格式不取决于文件扩展名。

主题

CSV 文件

假定扩展名为 .csv 的文件采用 CSV 格式并包含 UTF-8 编码的文本。C3R 加密客户端将所有值视为字符串。

.csv 文件中支持的属性

C3R 加密客户端要求 .csv 文件具有以下属性：

- 可能包含也可能不包含唯一命名每列的初始标题行。
- 逗号分隔。（目前，不支持自定义分隔符。）
- UTF-8 编码的文本。

从 .csv 条目中修剪空格

.csv 条目中的前导和尾部空格都会被修剪。

.csv 文件的自定义 NULL 编码

.csv 文件可以使用自定义 NULL 编码。

使用 C3R 加密客户端，您可以使用 `--csvInputNULLValue=<csv-input-null>` 标志为输入数据中的 NULL 条目指定自定义编码。通过使用 `--csvOutputNULLValue=<csv-output-null>` 标志，C3R 加密客户端可以在生成的输出文件中为 NULL 条目使用自定义编码。

> **Note**
>
> NULL 条目被认为缺少 内容，特别是在 SQL 表等更丰富的表格格式的上下文中。尽管由于历史原因 .csv 并不明确支持这种描述，但通常的惯例是将仅包含空格的空条目视为 NULL。因此，这是 C3R 加密客户端的默认行为，可以根据需要对其进行自定义。

C3R 如何解释 .csv 条目

下表举例说明了如何根据为 `--csvInputNULLValue=<csv-input-null>` 和 `--csvOutputNULLValue=<csv-output-null>` 标志提供的值（如果有）编组 .csv 条目（为清楚起见，cleartext 至 cleartext）。在 C3R 解释任何值的含义之前，将修剪引号之外的前导和尾部空格。

<csv-input-null>	**<csv-output-null>**	输入条目	输出条目
无	无	,AnyProduct,	,AnyProduct,
无	无	, AnyProduct ,	,AnyProduct,
无	无	,"AnyProduct",	,AnyProduct,
无	无	, "AnyProduct" ,	,AnyProduct,
无	无	,,	,,
无	无	, ,	,,
无	无	,"",	,,
无	无	," ",	," ",
无	无	, " " ,	," ",
"AnyProduct"	"NULL"	,AnyProduct,	,NULL,

<csv-input-null>	**<csv-output-null>**	输入条目	输出条目
"AnyProduct"	"NULL"	, AnyProduct ,	,NULL,
"AnyProduct"	"NULL"	,"AnyProduct",	,NULL,
"AnyProduct"	"NULL"	, "AnyProduct" ,	,NULL,
无	"NULL"	,,	,NULL,
无	"NULL"	, ,	,NULL,
无	"NULL"	,"",	,NULL,
无	"NULL"	," ",	," ",
无	"NULL"	, " " ,	," ",
""	"NULL"	,,	,NULL,
""	"NULL"	, ,	,NULL,
""	"NULL"	,"",	,"",
""	"NULL"	," ",	," ",
""	"NULL"	, " " ,	," ",
"\"\""	"NULL"	,,	,,
"\"\""	"NULL"	, ,	,,
"\"\""	"NULL"	,"",	,NULL,
"\"\""	"NULL"	," ",	," ",
"\"\""	"NULL"	, " " ,	," ",

不带标题的 CSV 文件

源 .csv 文件不必在第一行中包含标题来唯一命名每列。但是，没有标题行的 .csv 文件需要位置加密架构。需要的是位置加密架构，而不是带标题行的 .csv 文件和 Parquet 文件使用的典型映射架构。

位置加密架构按位置而不是按名称指定输出列。映射加密架构将源列名映射到目标列名。有关更多信息，包括两种架构格式的详细讨论和示例，请参阅映射和定位表架构。

Parquet 文件

假定扩展名为 .parquet 的文件采用 Apache Parquet 格式。

支持的 Parquet 数据类型

C3R 加密客户端可以在表示 AWS Clean Rooms支持的数据类型的 Parquet 文件中处理任何非复杂（即基本类型）数据。

但是，只能将字符串列用于 sealed 列。

支持以下 Parquet 数据类型：

- 带以下逻辑注释的 Binary 基本类型：
  - 如果已设置 --parquetBinaryAsString（STRING 数据类型），则为 None
  - Decimal(scale, precision)（DECIMAL 数据类型）
  - String（STRING 数据类型）
- 不带逻辑注释的 Boolean 基本数据类型（BOOLEAN 数据类型）
- 不带逻辑注释的 Double 基本数据类型（DOUBLE 数据类型）
- 带 Decimal(scale, precision) 逻辑注释的 Fixed_Len_Binary_Array 基本类型（DECIMAL 数据类型）
- 不带逻辑注释的 Float 基本数据类型（FLOAT 数据类型）
- 带以下逻辑注释的 Int32 基本类型：
  - 无（INT 数据类型）
  - Date（DATE 数据类型）
  - Decimal(scale, precision)（DECIMAL 数据类型）
  - Int(16, true)（SMALLINT 数据类型）
  - Int(32, true)（INT 数据类型）

- 带以下逻辑注释的 Int64 基本数据类型：
  - 无（BIGINT 数据类型）
  - Decimal(scale, precision)（DECIMAL 数据类型）
  - Int(64, true)（BIGINT 数据类型）
  - Timestamp(isUTCAdjusted, TimeUnit.MILLIS)（TIMESTAMP 数据类型）
  - Timestamp(isUTCAdjusted, TimeUnit.MICROS)（TIMESTAMP 数据类型）
  - Timestamp(isUTCAdjusted, TimeUnit.NANOS)（TIMESTAMP 数据类型）

加密非字符串值

当前，sealed 列仅支持字符串值。

对于 .csv 文件，C3R 加密客户端会将所有值视为 UTF-8 编码的文本，并且在加密之前不会尝试对其进行不同的解释。

对于指纹列，类型被分为等价类。等价类是一组数据类型，可以通过代表性数据类型明确比较其是否相等。

等价类允许将相同的指纹分配给相同的语义值，而不管原始表示形式如何。但是，两个等价类中的相同值不会生成相同的指纹列。

例如，无论 INTEGRAL 值 42 最初是 SMALLINT、INT 还是 BIGINT，都将为其分配相同的指纹。此外，INTEGRAL 值 0 永远不会与 BOOLEAN 值 FALSE 匹配（由值 0 表示）。

指纹列支持以下等价类和相应 AWS Clean Rooms 的数据类型：

等价类	支持的 AWS Clean Rooms 数据类型
BOOLEAN	BOOLEAN
DATE	DATE
INTEGRAL	BIGINT, INT, SMALLINT
STRING	CHAR, STRING, VARCHAR

Clean Rooms 加密计算中的列名

默认情况下，在 Clean Rooms 加密计算中，列的名称很重要。

如果允许对具有不同名称的列进行 JOIN 参数的值为 false，则在加密 fingerprint 列时将使用列名。因此，默认情况下，协作者必须事先进行协调，并对将在查询中使用 JOIN 语句的数据使用相同的目标列名称。默认情况下，为 JOIN 加密的列如果名称不同，就不能成功地对任何值进行 JOIN。

如果允许对具有不同名称的列进行 JOIN 参数的值为 true，则跨加密为 fingerprint 列的列的 JOIN 语句会成功。使用此参数加密数据可能允许对 cleartext 值进行一些推断。例如，如果某行的 `City` 列和 `State` 列中都具有相同的 HMAC 散列消息认证码值，则该值可能为 `New York`。

列标题名称的标准化

列标题名称由 C3R 加密客户端进行标准化。所有前导和尾随的空格都将被删除，转换后的输出将列名改为小写。

标准化应用于可能受列名影响的所有其他计算或其他运算。发出的输出文件仅包含标准化名称。

Clean Rooms 加密计算中的列类型

本主题提供有关 Clean Rooms 加密计算中列类型的信息。

主题

Fingerprint 列

Fingerprint 列是在 JOIN 语句中使用的受加密保护的列。

fingerprint 列中的数据无法解密。只有密封列中的数据才能解密。

Fingerprint 列只能在以下 SQL 子句和函数中使用：

- JOIN (INNER, OUTER, LEFT, RIGHT, or FULL) 与其他 fingerprint 列对比：
  - 如果将 `allowJoinsOnColumnsWithDifferentNames` 参数的值设置为 `false`，则 JOIN 的两个 fingerprint 列的名称也必须相同。

- SELECT COUNT()
- SELECT COUNT(DISTINCT)
- GROUP BY（仅当协作将 preserveNulls 参数的值设置为 true 时才使用。）

违反这些限制条件的查询可能会生成不正确的结果。

密封列

密封列 是 SELECT 语句中使用的通过加密保护的列。

密封列只能在以下 SQL 子句和函数中使用：

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

> **Note**
>
> 不支持 SELECT COUNT(DISTINCT)。

违反这些限制条件的查询可能会生成不正确的结果。

加密前为 sealed 列填充数据

当您指定列应该是 sealed 列时，C3R 会询问您要选择哪种填充。加密前填充数据是可选的。如果不使用填充（填充类型为 none），则加密数据的长度表示 cleartext 的大小。在某些情况下，cleartext 的大小可能会暴露明文。如果使用填充（填充类型为 fixed 或 max），则先将所有值填充到常见大小，然后再加密。使用填充时，加密数据的长度除了给出其大小的上限外，不提供有关原始 cleartext 长度的信息。

如果要为列填充并且已知该列中数据的最大字节长度，请使用 fixed 填充。使用至少与该列中最长值的字节长度一样大的 length 值。

> **Note**
>
> 如果值长于提供的 length 值，则会发生错误并导致加密失败。

如果要为列填充，而该列中数据的最大字节长度尚不清楚，请使用max填充。这种填充模式将所有数据填充到最长值的长度加上额外的 length 字节。

> **Note**
>
> 您可能需要批量加密数据，或者定期使用新数据更新表。请注意，max 填充会将条目填充到给定批次中最长的明文条目的长度（加 length 字节）。这意味着加密文字长度可能因批次而异。因此，如果您知道列的最大字节长度，则应使用 fixed 而不是 max。

Cleartext 列

Cleartext列是未受加密保护的列，无法在JOIN或SELECT语句中使用。

Cleartext 列可以用于 SQL 查询的任何部分。

加密计算参数

在创建协作时，可使用 Clean Rooms 加密计算 (C3R) 为协作提供加密计算参数。您可以使用 AWS Clean Rooms 控制台或 CreateCollaboration API 操作创建协作。在控制台中，启用支持加密计算选项后，可以为加密计算参数中的参数设置值。有关更多信息，请参阅以下主题。

主题

- 允许 cleartext 列参数
- “允许重复”参数
- “允许对具有不同名称的列进行 JOIN”参数
- “保留 NULL 值”参数

允许 cleartext 列参数

在控制台中，您可以在创建协作时设置允许 cleartext 列参数，以指定是否允许在包含加密数据的表中包含 cleartext 数据。

下表描述了允许 cleartext 列参数的值。

参数值	描述
否	加密表中不允许有 Cleartext 列。所有数据都受到加密保护。
是	加密表中允许有 Cleartext 列。 Cleartext 列不受加密保护，包含为 cleartext。您应该注意行中的 cleartext 数据可能揭示了表中其他数据的哪些信息。 要在特定列上运行 SUM 或 AVG，这些列必须是 cleartext。

使用 `CreateCollaboration` API 操作，对于 `dataEncryptionMetadata` 参数，您可以将 `allowCleartext` 的值设置为 `true` 或 `false`。有关 API 操作的更多信息，请参阅 AWS Clean Rooms API 参考。

Cleartext 列对应于按表特定架构分类为 cleartext 的列。这些列中的数据未加密，可以以任何方式使用。如果数据不敏感和/或需要比加密 sealed 列或 fingerprint 列所允许的更大的灵活性，则列 Cleartext 可能很有用。

“允许重复”参数

在控制台中，您可以在创建协作时设置允许重复参数，以指定为 JOIN 查询加密的列是否可以包含重复的非 NULL 值。

> **Important**
>
> 允许重复、允许对不同名称列进行 JOIN 和保留 NULL 值参数具有单独但相关的效果。

下表描述了允许重复参数的值。

参数值	描述
否	fingerprint 列中不允许有重复的值。单个 fingerprint 列中的所有值都必须是唯一的。
是	fingerprint 列中允许有重复的值。

参数值	描述
	如果需要联接具有重复值的列，请将此值设置为是。如果设置为是，则 C3R 表或结果的 fingerprint 列中出现的频率模式可能意味着有关 cleartext 数据结构的一些其他信息。

使用 `CreateCollaboration` API 操作，对于 `dataEncryptionMetadata` 参数，您可以将 `allowDuplicates` 的值设置为 `true` 或 `false`。有关 API 操作的更多信息，请参阅 AWS Clean Rooms API 参考。

默认情况下，如果必须在 JOIN 查询中使用加密数据，则 C3R 加密客户端要求这些列没有重复值。此要求是为了加强数据保护。这种行为可以帮助确保无法观察到数据中的重复模式。但是，如果您想在 JOIN 查询中使用加密数据并且不担心重复值，则允许重复参数可以禁用此保守检查。

“允许对具有不同名称的列进行 JOIN”参数

在控制台中，您可以在创建协作时设置允许对具有不同名称的列进行 JOIN 参数，以指定是否支持具有不同名称的列之间的 JOIN 语句。

有关更多信息，请参阅 列标题名称的标准化 。

下表描述了允许对具有不同名称的列进行 JOIN 参数的值。

参数值	描述
否	不支持联接具有不同名称的 fingerprint 列。JOIN 语句仅在具有相同名称的列上提供准确的结果。 **Important** 否值可提高信息安全性，但要求协作参与者事先就列名达成共识。如果两列在加密为 fingerprint 列时具有不同的名称，并且允许对具有不同名称的列进行 JOIN 设置为否，则对这些列的 JOIN 语句不会生成任何结果。这是因为它们之间不共享加密后的值。

参数值	描述
是	支持联接具有不同名称的 fingerprint 列。为了提高灵活性，用户可以将此值设置为是，这样无论列名如何，都允许对列执行 JOIN 语句。 如果设置为是，则 C3R 加密客户端在保护 fingerprint 列时将不会考虑列名。因此，在 C3R 表中可以观察到不同 fingerprint 列的共同值。 例如，如果一行在 `City` 列和 `State` 列中都具有相同的加密 JOIN 值，则可以合理地推断出该值为 `New York`。

使用 `CreateCollaboration` API 操作，对于 `dataEncryptionMetadata` 参数，您可以将 `allowJoinsOnColumnsWithDifferentNames` 的值设置为 `true` 或 `false`。有关 API 操作的更多信息，请参阅 AWS Clean Rooms API 参考。

默认情况下，fingerprint 列加密受该列的 `targetHeader` 所影响，它在步骤 4：为表格文件生成加密架构 中设置。因此，同一个 cleartext 值在每个不同的 fingerprint 列中都有不同的加密表示。

在某些情况下，此参数可用于防止推断 cleartext 值。例如，在 fingerprint 列中看到相同的加密值时，可以通过 `City` 和 `State` 来合理地推断该值是 `New York`。但是，使用此参数需要事先进行额外的协调，以便查询中要联接的所有列都具有共享名称。

您可以使用允许对具有不同名称的列进行 JOIN 参数来放宽此限制。当参数值设置为 `Yes` 时，无论名称如何，都允许一起使用为 JOIN 加密的任何列。

“保留 NULL 值”参数

在控制台中，您可以在创建协作时设置保留 NULL 值参数，以指示该列不存在任何值。

下表描述了保留 NULL 值参数的值。

参数值	描述
否	NULL 值不会被保留。NULL 值在加密表中不会显示为 NULL。NULL 值在 C3R 表中显示为唯一的随机值。

参数值	描述
是	NULL 值会被保留。NULL 值在加密表中显示为 NULL。如果您需要 NULL 值的 SQL 语义，则可以将此值设置为是。因此，无论列是否加密以及允许重复的参数设置如何，NULL 条目在 C3R 表中都会显示为 NULL。

使用 `CreateCollaboration` API 操作，对于 `dataEncryptionMetadata` 参数，您可以将 `preserveNulls` 的值设置为 `true` 或 `false`。有关 API 操作的更多信息，请参阅 AWS Clean Rooms API 参考。

当协作的保留 NULL 值参数设置为否时：

1. `cleartext` 列中的 NULL 条目保持不变。
2. 加密 `fingerprint` 列中的 NULL 条目被加密为随机值以隐藏其内容。在 cleartext 中联接带有 NULL 条目的加密列不会生成任何 NULL 条目的任何匹配项。不会进行任何匹配，因为它们各自会收到自己的唯一随机内容。
3. 加密 `sealed` 列中的 NULL 条目已加密。

当协作的保留 NULL 值参数的值设置为是时，无论列是否加密，所有列中的 NULL 条目都将保持为 NULL。

保留 NULL 值参数在数据扩充等情况下非常有用，在这些情况下，您需要共享以 NULL 表示的信息缺失。在 fingerprint 或 HMAC 格式中，如果要进行 JOIN 或 GROUP BY 的列中有 NULL 值，保留 NULL 值参数也很有用。

如果允许重复和保留 NULL 值参数的值设置为否，则在 fingerprint 列中包含多个 NULL 条目会产生错误并停止加密。如果任一参数的值设置为是，则不会发生此类错误。

Clean Rooms 加密计算中的可选标志

以下各节描述了在使用 C3R 加密客户端加密数据以进行表格文件自定义和测试时可以设置的可选标志。

主题

- --csvInputNULLValue 标志
- --csvOutputNULLValue 标志

- --enableStackTraces 标志
- --dryRun 标志
- --tempDir 标志

--csvInputNULLValue 标志

使用 C3R 加密客户端加密数据时，您可以使用 --csvInputNULLValue 标志为输入数据中的 NULL 条目指定自定义编码。

下表总结了此标志的用法和参数。

用法	参数
可选。用户可以为输入数据中的 NULL 条目指定自定义编码。	输入 CSV 文件中 NULL 值的用户指定编码

NULL 条目是被视为缺少内容的条目，特别是在 SQL 表等更丰富的表格格式的上下文中。尽管由于历史原因 .csv 并不明确支持这种描述，但通常的惯例是将仅包含空格的空条目视为 NULL。因此，这是 C3R 加密客户端的默认行为，可以根据需要对其进行自定义。

--csvOutputNULLValue 标志

使用 C3R 加密客户端加密数据时，您可以使用 --csvOutputNULLValue 标志为输出数据中的 NULL 条目指定自定义编码。

下表总结了此标志的用法和参数。

用法	参数
可选。用户可以在生成的输出文件中为 NULL 条目指定自定义编码。	输入 CSV 文件中 NULL 值的用户指定编码

NULL 条目是被视为缺少内容的条目，特别是在 SQL 表等更丰富的表格格式的上下文中。尽管由于历史原因 .csv 并不明确支持这种描述，但通常的惯例是将仅包含空格的空条目视为 NULL。因此，这是 C3R 加密客户端的默认行为，可以根据需要对其进行自定义。

--enableStackTraces 标志

使用 C3R 加密客户端加密数据时，可以使用 --enableStackTraces 标志提供其他上下文信息，以便在 C3R 遇到错误时报告错误。

AWS 不收集错误。如果遇到错误，请使用堆栈跟踪自行解决错误，或者将堆栈跟踪发送到 AWS Support 寻求帮助。

下表总结了此标志的用法和参数。

用法	参数
可选。用于提供其他上下文信息，以便在 C3R 加密客户端遇到错误时报告错误。	无

--dryRun 标志

加密和解密 C3R 加密客户端命令包括一个可选的 --dryRun 标志。该标志采用用户提供的所有参数，并检查它们的有效性和一致性。

您可以使用 --dryRun 标志来检查您的架构文件是否有效且与其相应的输入文件一致。

下表总结了此标志的用法和参数。

用法	参数
可选。使 C3R 加密客户端解析参数和检查文件，但不执行加密或解密。	无

--tempDir 标志

您可能需要使用临时目录，因为加密文件有时可能比非加密文件大，具体取决于它们的设置。每个协作还必须对数据集进行加密才能正常工作。

使用 C3R 加密数据时，可以使用 --tempDir 标志来指定在处理输入时可以创建临时文件的位置。

下表总结了此标志的用法和参数。

用法	参数
用户可以指定在处理输入时可以创建临时文件的位置。	默认为系统临时目录。

使用 Clean Rooms 加密计算进行查询

本主题提供有关编写查询的信息，这些查询使用已使用 Clean Rooms 加密计算加密的数据表。

主题

在 NULL 上分支的查询

要在 NULL 语句上设置查询分支，需要使用 `IF x IS NULL THEN 0 ELSE 1` 这样的语法。

查询总是可以在 cleartext 列中的 NULL 语句上分支。

只有当保留 NULL 值参数 (`preserveNulls`) 的值设置为 `true` 时，查询才能在 sealed 列和 fingerprint 列中的 NULL 语句上分支。

违反这些限制条件的查询可能会生成不正确的结果。

将一个源列映射到多个目标列

将一个源列映射到多个目标列。例如，您可能希望在一列上同时进行 JOIN 和 SELECT。

有关更多信息，请参阅 在 JOIN 和 SELECT 查询中使用相同的数据。

在 JOIN 和 SELECT 查询中使用相同的数据

如果列中的数据不敏感，则它可以出现在 cleartext 目标列中，这样就可以用于任何目的。

如果列中的数据很敏感，必须同时用于 JOIN 和 SELECT 查询，则应将该源列映射到输出文件中的两个目标列。一列作为 fingerprint 列进行 `type` 加密，一列作为密封列进行 `type` 加密。C3R 加密客户

端的交互式架构生成建议标题后缀为 `_fingerprint` 和 `_sealed`。这些标题后缀可以成为快速区分此类列的有用惯例。

C3R 加密客户端指南

C3R 加密客户端是一种工具，它使组织能够将敏感数据整合在一起，从数据分析中获得新的洞察。该工具以加密方式限制了任何一方和在此过程中可以学到 AWS 的内容。尽管这一点至关重要，但以加密方式保护数据的过程可能会在计算和存储资源方面增加大量开销。因此，了解使用每种设置的利弊得失以及如何在保持所需的加密保障的同时优化设置非常重要。本主题重点介绍 C3R 加密客户端和架构中不同设置对性能的影响。

所有 C3R 加密客户端加密设置都提供不同的加密保障。默认情况下，协作级别的设置是最安全的。在创建协作时启用其他功能会削弱隐私保障，从而允许对加密文字进行频率分析等活动。有关如何使用这些设置及其影响的更多信息，请参阅加密计算。

主题

对列类型的性能影响

C3R 使用三种列类型：cleartext、fingerprint 和 sealed。每种列类型都提供不同的加密保障，并且具有不同的预期用途。在以下各节中，将讨论列类型的性能影响以及每种设置对性能的影响。

主题

Cleartext 列

Cleartext 列不会改变其原始格式，也不会以任何方式进行加密处理。此列类型无法配置，也不会影响存储或计算性能。

Fingerprint 列

Fingerprint 列旨在用于联接多个表中的数据。为此，生成的加密文字大小必须始终相同。但是，这些列受协作级别设置的影响。Fingerprint 列可能会对输出文件大小产生不同程度的影响，具体取决于输入中包含的 cleartext。

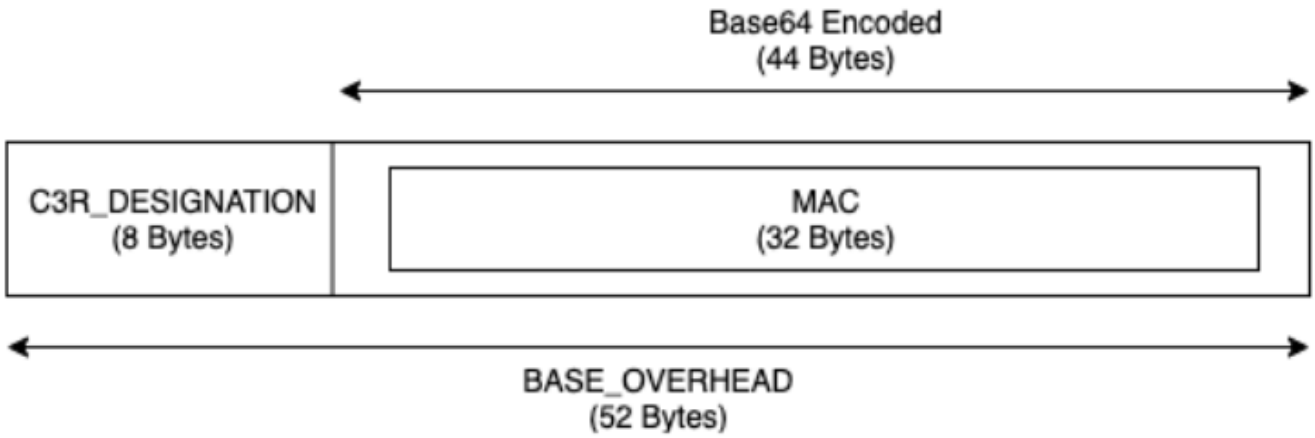
主题

fingerprint 列的基本开销

fingerprint 列有基本开销。这种开销是恒定的，与 cleartext 字节的大小无关。

fingerprint 列中的数据通过 HMAC 散列消息认证码函数进行加密处理，该函数将数据转换为 32 字节的消息认证码 (MAC)。然后通过 base64 编码器对这些数据进行处理，使字节大小增加约 33%。它前面有一个 8 字节的 C3R 标识，用于指定数据所属的列类型以及生成数据的客户端版本。最终结果为 52 字节。然后将此结果乘以行数得出总基本开销（如果 `preserveNulls` 设置为 true，则使用非 `null` 值总数）。

下图显示了如何操作 *`BASE_OVERHEAD = C3R_DESIGNATION + (MAC * 1.33)`*



fingerprint 列中的输出加密文字将始终为 52 字节。如果输入 cleartext 数据的平均值超过 52 字节（例如，完整的街道地址），则存储空间可能会显著减少。如果输入 cleartext 数据的平均值小于 52 字节（例如，客户年龄），则存储空间可能会显著增加。

fingerprint 列的协作设置

preserveNulls 设置

当协作级别设置 preserveNulls 为 false（默认值）时，每个 null 值都将替换为一个唯一的随机 32 字节，并被当作非 null 处理。结果是，现在每个 null 值都是 52 字节。与此设置为 true 且 null 值作为 null 传递时相比，对于包含非常稀疏的数据的表，这可能会增加大量存储需求。

如果您不需要此设置的隐私保障，并且希望在数据集中保留 null 值，请在创建协作时启用 preserveNulls 设置。创建协作后将无法更改 preserveNulls 设置。

fingerprint 列的示例数据

以下是带有要重现设置的 fingerprint 列的输入和输出数据的示例集。其他协作级别的设置（例如 allowCleartext 和 allowDuplicates）不会影响结果，如果尝试在本地重现，则可以设置为 true 或 false。

共享密钥示例：wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY

协作 ID 示例：a1b2c3d4-5678-90ab-cdef-EXAMPLE11111

allowJoinsOnColumnsWithDifferentNames：True 此设置不会影响性能或存储要求。但是，在重现下表中显示的值时，此设置使列名的选择变得无关紧要。

示例 1

输入	null
preserveNulls	TRUE
输出	null
确定性	Yes
输入字节	0
输出字节	0

示例 2

输入	null

preserveNulls	FALSE
输出	01:hmac:3lkFjthvV3IUu6mMvFc1a+XAHwgw/ElmOq4p3Yg25kk=
确定性	No
输入字节	0
输出字节	52

示例 3

输入	empty string
preserveNulls	-
输出	01:hmac:oKTgi3Gba+eUb3JteSz2EMgXUkF1WgM77UP0Ydw5kPQ=
确定性	Yes
输入字节	0
输出字节	52

示例 4

输入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
输出	01:hmac:kU/IqwG7FMmzzshr0B9scomE0UJUEE7j9keTctplGww=
确定性	Yes
输入字节	26

输出字节	52

示例 5

输入	`abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789`
`preserveNulls`	-
输出	`01:hmac:ks3htnQbw2vdhCRFF6JNzW5LMndJaHG57uvE26mBtSs=`
确定性	`Yes`
输入字节	62
输出字节	52

fingerprint 列疑难解答

为什么我的 fingerprint 列中的加密文字比进入它的 cleartext 大几倍？

fingerprint 列中的加密文字的长度始终为 52 字节。如果您的输入数据很小（例如，客户的年龄），则其大小将显著增加。如果将 `preserveNulls` 设置设置为 `false`，也可能发生这种情况。

为什么我的 fingerprint 列中的加密文字比进入它的 cleartext 小几倍？

fingerprint 列中的加密文字的长度始终为 52 字节。如果您的输入数据很大（例如，客户的完整街道地址），则其大小将显著减小。

我怎么知道我是否需要 **`preserveNulls`** 提供的加密保障？

遗憾的是，答案是“看情况”。至少，应查看 the section called “参数” 了解 `preserveNulls` 设置如何保护您的数据。但是，我们建议您参考组织的数据处理要求以及适用于相应协作的任何合同。

为什么我必须承担 base64 的开销？

为了与 CSV 等表格文件格式兼容，必须进行 base64 编码。尽管某些文件格式（例如 Parquet）可能支持数据的二进制表示，但重要的是，协作中的所有参与者都必须以相同的方式表示数据，以确保查询结果正确。

Sealed 列

Sealed 列用于在协作成员之间传输数据。这些列中的加密文字是不确定的，并且会根据列的配置方式对性能和存储产生重大影响。这些列可以单独配置，通常对 C3R 加密客户端的性能和由此产生的输出文件大小影响最大。

主题

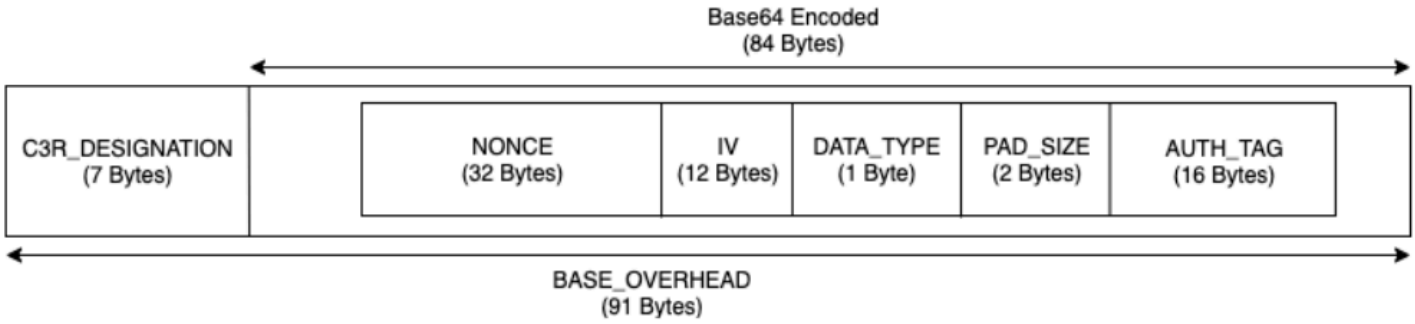
- sealed 列的基本开销
- sealed 列的协作设置
- 架构设置 sealed 列：填充类型
- sealed 列的示例数据
- sealed 列疑难解答

sealed 列的基本开销

sealed 列有基本开销。此开销是恒定的，是 cleartext 和填充（如有）字节大小之外的开销。

在进行任何加密之前，在 sealed 列中的数据前面加上一个 1 字节的字符，表示所包含的数据类型。如果选择了填充，则会对数据进行填充并追加 2 个字节，说明填充大小。添加这些字节后，使用 AES-GCM 对数据进行加密处理，并以 IV（12 字节）、nonce（32 字节）和 Auth Tag（16 字节）进行存储。然后通过 base64 编码器对这些数据进行处理，使字节大小增加约 33%。数据前面有一个 7 字节的 C3R 标识，用于指定数据属于哪种类型的列以及用于生成数据的客户端版本。结果是 91 字节的最终基本开销。然后将此结果乘以行数得出总基本开销（如果 `preserveNulls` 设置为 true，则使用非 null 值总数）。

下图显示了如何操作 *`BASE_OVERHEAD = C3R_DESIGNATION + ((NONCE + IV + DATA_TYPE + PAD_SIZE + AUTH_TAG) * 1.33)`*



sealed 列的协作设置

`preserveNulls` 设置

当协作级别设置 `preserveNulls` 为 `false`（默认值）时，每个 `null` 值都将是唯一的随机 32 字节，并被当作非 `null` 处理。结果是，现在每个 `null` 值都是 91 字节（如果填充，则更多）。与此设置为 `true` 且 `null` 值作为 `null` 传递时相比，对于包含非常稀疏的数据的表，这可能会增加大量存储需求。

如果您不需要此设置的隐私保障，并且希望在数据集中保留 `null` 值，请在创建协作时启用 `preserveNulls` 设置。创建协作后将无法更改 `preserveNulls` 设置。

架构设置 sealed 列：填充类型

主题

- none 的填充类型
- fixed 的填充类型
- max 的填充类型

`none` 的填充类型

选择 `none` 的填充类型不会向 cleartext 增加任何填充，也不会在前面描述的基本开销之外增加额外的开销。没有填充会产生最节省空间的输出大小。但是，它不提供与 `fixed` 和 `max` 填充类型相同的隐私保障。这是因为底层 cleartext 的大小可以从加密文字的大小中分辨出来。

`fixed` 的填充类型

选择 `fixed` 的填充类型是一种隐私保护措施，用于隐藏列中包含的数据的长度。这是通过在加密之前将所有 `pad_length` 填充到提供的 cleartext 来完成的。任何超过该大小的数据都会导致 C3R 加密客户端失败。

假设填充是在加密之前添加到 cleartext 的，因此 AES-GCM 具有 cleartext 到加密文字字节的 1:1 映射。base64 编码将增加 33%。填充的额外存储开销可以通过从 pad_length 的值中减去 cleartext 的平均长度，然后乘以 1.33 来计算。结果就是每条记录的平均填充开销。然后将此结果乘以行数得出总填充开销（如果 preserveNulls 设置为 true，则使用非 null 值总数）。

*PADDING_OVERHEAD = (PAD_LENGTH - AVG_CLEARTEXT_LENGTH) * 1.33 * ROW_COUNT*

我们建议您选择包含列中最大值的最小 pad_length。例如，如果最大值为 50 字节，则 pad_length 为 50 就足够了。大于该值的值只会增加额外的存储开销。

固定填充不会增加任何显著的计算开销。

max 的填充类型

选择 max 的填充类型是一种隐私保护措施，用于隐藏列中包含的数据的长度。这是通过将所有 cleartext 填充到列中的最大值再加上加密之前的额外 pad_length 来完成的。通常，max 填充提供的保障与单个数据集的 fixed 填充相同，同时允许不知道列中的最大 cleartext 值。但是，max 填充可能无法提供与跨更新的 fixed 填充相同的隐私保障，因为各个数据集中的最大值可能有所不同。

我们建议您在使用 max 填充时，额外选择 0 的 pad_length。此长度将所有值填充到与列中最大值的大小相同。大于该值的值只会增加额外的存储开销。

如果已知给定列的最大 cleartext 值，我们建议您改用 fixed 填充类型。使用 fixed 填充可在更新的数据集之间实现一致性。使用 max 填充会使每个数据子集填充到该子集中的最大值。

sealed 列的示例数据

以下是带有要重现设置的 sealed 列的输入和输出数据的示例集。其他协作级别的设置（例如 allowCleartext、allowJoinsOnColumnsWithDifferentNames 和 allowDuplicates）不会影响结果，如果尝试在本地重现，则可以设置为 true 或 false。尽管这些是要重现的基本设置，但 sealed 列是不确定的，值每次都会更改。目的是显示输入字节与输出字节的对比。示例 pad_length 值是故意选择的。它们表明，使用推荐的最低 pad_length 设置或需要额外的填充时，fixed 填充产生的值与 max 填充相同。

共享密钥示例：wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY

协作 ID 示例：a1b2c3d4-5678-90ab-cdef-EXAMPLE11111

主题

- none 的填充类型

- fixed 的填充类型（示例 1）
- fixed 的填充类型（示例 2）
- max 的填充类型（示例 1）
- max 的填充类型（示例 2）

none 的填充类型

示例 1

输入	null
preserveNulls	TRUE
输出	null
确定性	Yes
输入字节	0
输出字节	0

示例 2

输入	null
preserveNulls	FALSE
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfssGSPbNIJfG3iXmu6cbCUrizuV
确定性	No
输入字节	0
输出字节	91

示例 3

输入	empty string
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfstGSPEM6qR8DWC2PB2GMlX41YK
确定性	No
输入字节	0
输出字节	91

示例 4

输入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfsteEE1GKEPiRzyh0h7t6OmWMLTWCvO2ckr6pkx9sGL5VLDQeHzh6DmPpyWNuI=
确定性	No
输入字节	26
输出字节	127

示例 5

输入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t6OmWMLTWCvO2ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ QOQ3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
确定性	No
输入字节	62
输出字节	175

fixed 的填充类型（示例 1）

在此示例中，pad_length 为 62，最大输入为 62 字节。

示例 1

输入	null
preserveNulls	TRUE
输出	null
确定性	Yes
输入字节	0
输出字节	0

示例 2

输入	null
preserveNulls	FALSE
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfssGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcg1yOPvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcoNpATs0GzbnLkor4L+/aSuA=
确定性	No
输入字节	0
输出字节	175

示例 3

输入	empty string
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcg1yOPvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcoLB53l07VZpA6OwkuXu29CA=
确定性	No
输入字节	0
输出字节	175

示例 4

输入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfsteEE1GKEPiRzyh0h7t6OmWMLTWCvO2ckr6pkx9jy48Fcg1yOPvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcutBAcO+Mb9tuU2KIHH31AWg=
确定性	No
输入字节	26
输出字节	175

示例 5

输入	abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfsteEE1GKEPiRzyh0h7t6OmWMLTWCvO2ckr6plwtH/8tRFnn2rF91bcB9G4+n8GiRfJNmqdP4/QOQ3cXb/pbvPcnnohrHIGSX54ua+1/JfcVjc=
确定性	No
输入字节	62

输出字节	175

fixed 的填充类型（示例 2）

在此示例中，pad_length 为 162，最大输入为 62 字节。

示例 1

输入	null
preserveNulls	TRUE
输出	null
确定性	Yes
输入字节	0
输出字节	0

示例 2

输入	null
preserveNulls	FALSE
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKLOhK1+7r75Tk+Mx9jy48 Fcg1yOPvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv/xAySX+xcntotL703aBTBb

确定性	No
输入字节	0
输出字节	307

示例 3

输入	empty string
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKLOhK1+7r75Tk+Mx9jy48 Fcg1yOPvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv84lVaT9Yd+6oQx65/+gdVT
确定性	No
输入字节	0
输出字节	307

示例 4

输入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY

	Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t6OmWMLTWCvO2ckr6pkx9jy48 Fcg1yOPvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwtX5Hnl+WyfO6ks3QMaRDGSf
确定性	No
输入字节	26
输出字节	307

示例 5

输入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t6OmWMLTWCvO2ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/QOQ3cXb/pbvPcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwjkJXQZOgPdeFX9Yr/8alV5i
确定性	No

输入字节	62
输出字节	307

max 的填充类型（示例 1）

在此示例中，pad_length 为 0，最大输入为 62 字节。

示例 1

输入	null
preserveNulls	TRUE
输出	null
确定性	Yes
输入字节	0
输出字节	0

示例 2

输入	null
preserveNulls	FALSE
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKLOhK1+7r75Tk+Mx9jy48 Fcg1yOPvBqRSZ7oqy1V3UKfYTLEZb/ hCz7oaIneVsrcoNpATs0GzbnLkor4L+/ aSuA=
确定性	No
输入字节	0

输出字节	175

示例 3

输入	empty string
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfstGSNWfMRp7nSb7SMX2s3JKLOhK1+7r75Tk+Mx9jy48Fcg1yOPvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcoLB53l07VZpA6OwkuXu29CA=
确定性	No
输入字节	0
输出字节	175

示例 4

输入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5MG5vbmNlMDEyMzQ1Njc4OTBqfRYZ98t5KU6aWfsteEE1GKEPiRzyh0h7t6OmWMLTWCvO2ckr6pkx9jy48Fcg1yOPvBqRSZ7oqy1V3UKfYTLEZb/hCz7oaIneVsrcutBAcO+Mb9tuU2KIHH31AWg=
确定性	No

输入字节	26
输出字节	175

示例 5

输入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t6OmWMLTWCvO2ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqdP4/ QOQ3cXb/pbvPcnnohrHIGSX54ua+1/ JfcVjc=
确定性	No
输入字节	62
输出字节	175

max 的填充类型（示例 2）

在此示例中，pad_length 为 100，最大输入为 62 字节。

示例 1

输入	null
preserveNulls	TRUE
输出	null
确定性	Yes

输入字节	0
输出字节	0

示例 2

输入	null
preserveNulls	FALSE
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfssGSNWfMRp7nSb7S MX2s3JKLOhK1+7r75Tk+Mx9jy48 Fcg1yOPvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv/xAySX+xcntotL703aBTBb
确定性	No
输入字节	0
输出字节	307

示例 3

输入	empty string
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfstGSNWfMRp7nSb7S MX2s3JKLOhK1+7r75Tk+Mx9jy48

	Fcg1yOPvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwv84lVaT9Yd+6oQx65/+gdVT
确定性	No
输入字节	0
输出字节	307

示例 4

输入	abcdefghijklmnopqrstuvwxyz
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t6OmWMLTWCvO2ckr6pkx9jy48 Fcg1yOPvBqRSZ7oqy1V3UKfYTLE Zb/hCz7oaIneVsrcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwtX5Hnl+WyfO6ks3QMaRDGSf
确定性	No
输入字节	26
输出字节	307

示例 5

输入	abcdefghijklmnopqrstuvwxyzA BCDEFGHIJKLMNOPQRSTUVWXYZ01 23456789
preserveNulls	-
输出	01:enc:bm9uY2UwMTIzNDU2Nzg5 MG5vbmNlMDEyMzQ1Njc4OTBqfRY Z98t5KU6aWfsteEE1GKEPiRzyh0 h7t6OmWMLTWCvO2ckr6plwtH/8t RFnn2rF91bcB9G4+n8GiRfJNmqd P4/QOQ3cXb/pbvPcnkB0xbLWD7z NdAqQGR0rXoSESdW0I0vpNoGcBf v4cJbG0A3h1DvtkSSVc2B80OOGp pzdDqhrUVN5wFNyn8vgfPMqDaeJk5bn +8o4WtG/ClipNcjDXvXVtK4vfCohcCA6 uwrmwjkJXQZOgPdeFX9Yr/8alV5i
确定性	No
输入字节	62
输出字节	307

sealed 列疑难解答

为什么我的 sealed 列中的加密文字比进入它的 cleartext 大几倍？

这取决于多个因素。首先，Cleartext 列中的加密文字的长度始终至少为 91 字节。如果您的输入数据很小（例如，客户的年龄），则其大小将显著增加。其次，如果 preserveNulls 设置为 false，并且您的输入数据包含很多 null 值，则每个 null 值都将变成 91 字节的加密文字。最后，如果您使用填充，则根据定义，在加密 cleartext 数据之前会将字节添加到该数据中。

我的 sealed 列中的大部分数据都非常小，我需要使用填充。我可以删除大值并单独处理它们以节省空间吗？

我们不建议删除大值并单独处理。这样做会改变 C3R 加密客户端提供的隐私保证。作为威胁模型，假设观察者可以看到两个加密的数据集。如果观察者发现一个数据子集的列填充明显大于或小于另一个子集，则他们可以推断每个子集中数据的大小。例如，假设 `fullName` 列在一个文件中填充到总共 40 字节，而在另一个文件中填充到 800 字节。观察者可能会认为，其中一个数据集包含了世界上最长的名字（747 字节）。

使用 **`max`** 填充类型时，我需要提供额外的填充吗？

不需要。使用 `max` 填充时，我们建议将 `pad_length`（也称为列中最大值之外的额外填充）设置为 0。

我能否在使用 **`fixed`** 填充时选择大的 **`pad_length`** 来避免担心最大的值是否合适？

能，但是大的填充长度效率低下，并且占用的存储空间超出了必要的范围。我们建议您查看最大值有多大，并将 `pad_length` 设置为该值。

我怎么知道我是否需要 **`preserveNulls`** 提供的加密保障？

遗憾的是，答案是“看情况”。至少，应查看 Clean Rooms 加密计算 了解 `preserveNulls` 设置如何保护您的数据。但是，我们建议您参考组织的数据处理要求以及适用于相应协作的任何合同。

为什么我必须承担 base64 的开销？

为了与 CSV 等表格文件格式兼容，必须进行 base64 编码。尽管某些文件格式（例如 Parquet）可能支持数据的二进制表示，但重要的是，协作中的所有参与者都必须以相同的方式表示数据，以确保查询结果正确。

加密文字大小意外增加疑难解答

假设您加密了数据，结果数据的大小出人意料地大。以下步骤可以帮助您确定大小增加的位置以及可以采取的措施（如果有）。

确定大小增加的位置

在排查加密数据明显大于 cleartext 数据的原因之前，必须先确定大小的增加位置。可以放心地忽略 Cleartext 列，因为它们没有变化。查看其余的 fingerprint 和 sealed 列，然后选择一个看起来很重要的列。

确定大小增加的原因

fingerprint 列或 sealed 列可能会导致大小增加。

主题

大小增加是否来自 fingerprint 列？

如果对存储增加贡献最大的列是 fingerprint 列，则可能是因为 cleartext 数据很小（例如，客户年龄）。生成的每个 fingerprint 加密文字的长度为 52 字节。不幸的是，在这个问题上无能为 column-by-column 力。有关更多信息，请参阅 fingerprint 列的基本开销 了解有关此列的详细信息，包括它如何影响存储要求。

导致 fingerprint 列中大小增加的另一个可能原因是协作设置 `preserveNulls`。如果禁用了 `preserveNulls` 的协作设置（默认设置），则 fingerprint 列中的所有 `null` 值都将变为 52 字节的加密文字。目前的协作对此无能为力。`preserveNulls` 设置是在创建协作时设置的，所有协作者必须使用相同的设置以确保查询结果正确。有关 `preserveNulls` 设置以及启用它会如何影响数据隐私保障的更多信息，请参阅 加密计算。

大小增加是否来自 sealed 列？

如果对存储增加贡献最大的列是 sealed 列，那么有一些细节可能会导致大小增加。

如果 cleartext 数据很小（例如，客户年龄），则生成的每个 sealed 加密文字的长度至少为 91 字节。遗憾的是，我们对这个问题无能为力。有关更多信息，请参阅 sealed 列的基本开销 了解有关此列的详细信息，包括它如何影响存储要求。

sealed 列存储增加的第二个主要原因是填充。填充会在加密 cleartext 之前向其添加额外的字节，以隐藏数据集中各个值的大小。我们建议您将数据集的填充设置为可能的最小值。至少必须将 `fixed` 填充的 `pad_length` 设置为包含列中可能的最大值。任何高于此值的设置都不会增加额外的隐私保障。例如，如果您知道列中可能的最大值可能为 50 字节，我们建议您将 `pad_length` 设置为 50 字节。但是，如果 sealed 列使用 `max` 填充，我们建议您将 `pad_length` 设置为 0 字节。这是因为 `max` 填充是指列中最大值之外的额外填充。

导致 sealed 列中大小增加的最后一个可能原因是协作设置 `preserveNulls`。如果禁用了 `preserveNulls` 的协作设置（默认设置），则 sealed 列中的所有 `null` 值都将变为 91 字节的加密文字。目前的协作对此无能为力。`preserveNulls` 设置是在创建协作时设置的，所有协作者必须使用相同的设置以确保查询结果正确。有关此设置以及启用它会如何影响数据隐私保障的更多信息，请参阅 加密计算。

查询登录 AWS Clean Rooms

查询日志是中的一项功能 AWS Clean Rooms。当您创建协作并开启查询日志时，成员可以在 Amazon Logs 中存储与他们相关的查询 CloudWatch 日志。

通过查询日志，成员可以确定查询是否符合分析规则并符合协作协议。此外，查询日志有助于支持审计。

在 AWS Clean Rooms 控制台中打开查询日志选项后，查询日志包括以下内容：

- `analysisRule` — 已配置表的分析规则。
- `analysisTemplateArn` — 已运行的分析模板（根据分析规则显示）。
- `collaborationId` — 运行查询的协作的唯一标识符。
- `configuredTableID` — 查询中引用的已配置表的唯一标识符。
- `directQueryAnalysisRulePolicy.custom.allowedAnalysis` — 允许在配置表上运行的分析模板（根据分析规则显示）。
- `directQueryAnalysisRulePolicy.v1.custom.allowedAnalysisProviders` — 允许创建查询的查询提供者（根据分析规则显示）。
- `eventID` — 查询运行的唯一标识符。2023 年 8 月 31 日之后，唯一标识符与 `protectedQueryID` 相同。
- `eventTimestamp` — 查询运行时间。
- `parameters.parametervalue` — 参数值（根据查询文本显示）。
- `queryText` — 查询运行的 SQL 定义。如果有参数，则会将其标记为 `:parametervalue`。
- `queryValidationErrors` — 查询验证时的查询错误。
- `schemaName` — 查询中引用的已配置表关联的名称。

接收查询日志

您无需在之外执行任何操作 AWS Clean Rooms 即可设置查询日志。 AWS Clean Rooms 在每个协作成员创建成员资格后，为协作创建日志组。

可以查询的成员、可以接收结果的成员，以及在查询中引用了已配置表的成员，都将收到查询日志。

可以查询的成员、可以接收结果的成员将收到查询中引用的每个已配置表的查询日志。如果他们不拥有配置表，将无法查看配置表 ID (`configuredTableID`)。

如果成员在查询中引用了多个配置表关联，则他们将收到每个配置表的查询日志。

将为包含 AWS Clean Rooms中不支持和支持的 SQL 的查询创建日志。有关详细信息，请参阅 AWS Clean Rooms SQL 参考。

当查询引用与协作无关的已配置表时，也会创建日志。

中没有为不正确的 SQL 创建日志 AWS Clean Rooms。

查询日志不表明查询成功且查询输出已送达。它们确认查询是由可以查询的成员提交的。查询日志还确认查询中包含支持的 SQL， AWS Clean Rooms 并引用了与协作关联的已配置表。

Example

例如，如果在 AWS Clean Rooms 验证查询是否符合分析规则之后并在查询处理期间取消查询，则不会生成日志。

如果删除日志组，则必须使用相同的日志组名称（协作的协作 ID）手动重新创建日志组。或者，您也可以在成员身份中禁用和启用日志记录。

有关如何启用查询日志记录的更多信息，请参阅在 AWS Clean Rooms 中创建协作。

有关 Amazon CloudWatch 日志的更多信息，请参阅 Amazon CloudWatch 日志用户指南。

使用查询日志

我们建议成员定期采取以下行动：

- 要验证查询是否与协作商定的使用案例或查询相匹配，请查看协作中运行的查询。

  有关如何查看最近查询的更多信息，请参阅查看最近的查询。
- 要验证配置表列是否与协作商定的内容相匹配，请查看协作成员分析规则和查询中使用的配置表列。

  有关如何查看已配置列的更多信息，请参阅查看表和分析规则。

设置 AWS Clean Rooms

以下主题说明了如何设置 AWS Clean Rooms。

主题

报名参加 AWS

在使用任何 AWS 服务（包括）之前 AWS Clean Rooms，您必须先注册 AWS。

如果您没有 AWS 账户，请完成以下步骤来创建一个。

报名参加 AWS 账户

1. 打开 https://portal.aws.amazon.com/billing/signup。
2. 按照屏幕上的说明进行操作。

   在注册时，将接到一通电话，要求使用电话键盘输入一个验证码。
3. 当您注册时 AWS 账户，将创建一个 AWS 账户 root 用户。根用户有权访问该账户中的所有 AWS 服务 和资源。作为安全最佳实践，请为管理用户分配管理访问权限，并且只使用根用户执行需要根用户访问权限的任务。

为设置服务角色 AWS Clean Rooms

主题

创建管理员用户

要使用 AWS Clean Rooms，您需要为自己创建一个管理员用户，并将该管理员用户添加到管理员组中。

要创建管理员用户，请选择以下选项之一。

选择一种方法来管理您的管理员	目的	方式	您也可以
在 IAM Identity Center 中（建议）	使用短期凭证访问 AWS。 这符合安全最佳实操。有关最佳实践的信息，请参阅《IAM 用户指南》中的 IAM 中的安全最佳实践。	有关说明，请参阅《AWS IAM Identity Center 用户指南》中的入门。	通过在《AWS Command Line Interface 用户指南》 AWS IAM Identity Center中配置 AWS CLI 要使用的来配置编程访问权限。
在 IAM 中（不推荐使用）	使用长期凭证访问 AWS。	按照《IAM 用户指南》中的创建您的首个 IAM 管理员用户和组的说明操作。	按照《IAM 用户指南》中的管理 IAM 用户的访问密钥，配置编程式访问。

为协作成员创建 IAM 角色

成员是 AWS 指参与协作的客户。

为协作成员创建 IAM 角色

1. 按照用户指南中的创建向 IAM 用户委派权限的角色过程进行AWS Identity and Access Management 操作。

2. 在“创建策略”步骤中，在策略编辑器中选择 JSON 选项卡，然后根据授予协作成员的权限添加策略。

   AWS Clean Rooms 根据常见用例提供以下托管策略：

如果要...	然后使用...
查看资源和元数据	AWS 托管策略：AWSCleanRoomsReadOnlyAccess
查询	AWS 托管策略：AWSCleanRoomsFullAccess
查询和接收结果	AWS 托管策略：AWSCleanRoomsFullAccess
管理协作资源但不要查询	AWS 托管策略：AWSCleanRoomsFullAccessNoQuerying

   有关提供的不同托管策略的信息 AWS Clean Rooms，请参阅 AWS 的托管策略 AWS Clean Rooms

创建服务角色来读取数据

AWS Clean Rooms 使用服务角色读取数据。

有两种方法可以创建此服务角色：

如果...	那么
您拥有创建服务角色所必需的 IAM 权限	使用 AWS Clean Rooms 控制台创建服务角色。
你没有iam:CreateRole ，iam:CreatePolicy 和iam:AttachRolePolicy 权限 或者 你想手动创建 IAM 角色	请执行以下操作之一： • 使用以下步骤创建服务角色。 • 请您的管理员使用以下步骤创建服务角色。

创建服务角色来读取数据

> **Note**
>
> 只有在您没有使用 AWS Clean Rooms 控制台创建服务角色的必要权限时，您或您的 IAM 管理员才应遵循此步骤。

1. 按照AWS Identity and Access Management 用户指南中的使用自定义信任策略创建角色（控制台）过程进行操作。
2. 根据使用自定义信任策略创建角色（控制台）过程使用以下自定义信任策略。

> **Note**
>
> 如果您想确保该角色只能在特定的协作成员身份环境中使用，则可以进一步缩小信任策略的范围。有关更多信息，请参阅 防止跨服务混淆代理。

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

3. 根据使用自定义信任策略创建角色（控制台）过程使用以下权限策略。

> **Note**
>
> 以下示例策略支持读取 AWS Glue 元数据及其相应的 Amazon S3 数据所需的权限。但是，您可能需要修改此策略，具体取决于您设置 S3 数据的方式。例如，如果您为 S3 数据设置了自定义 KMS 密钥，则可能需要修改此策略，使其具有额外的 AWS KMS 权限。

您的 AWS Glue 资源和底层 Amazon S3 资源必须与 AWS Clean Rooms 协作 AWS 区域相同。

```
{
    "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:aws-region:accountId:database/database",
                "arn:aws:glue:aws-region:accountId:table/table",
                "arn:aws:glue:aws-region:accountId: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":[
                        "s3BucketOwnerAccountId"
                    ]
                }
            }
        },
        {
            "Sid": "NecessaryS3ObjectPermissions",
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3::bucket/prefix/*"
            ],
            "Condition":{
                "StringEquals":{
                    "s3:ResourceAccount":[
                        "s3BucketOwnerAccountId"
                    ]
                }
            }
        }

    ]
}
```

4. 用您自己的信息替换每个*###*。
5. 继续按照使用自定义信任策略创建角色（控制台）过程创建角色。

创建服务角色以接收结果

> **Note**
>
> 如果您是只能接收结果的成员（在控制台中，您的成员权限仅为接收结果），请按照以下步骤操作。

如果您是同时可以查询和接收结果的成员（在控制台中，您的成员权限是查询和接收结果），则可以跳过此过程。

对于只能接收结果的协作成员，AWS Clean Rooms 使用服务角色将协作中查询数据的结果写入指定的 Amazon S3 存储桶。

有两种方法可以创建此服务角色：

如果...	那么
您拥有创建服务角色所必需的 IAM 权限	使用 AWS Clean Rooms 控制台创建服务角色。
你没有`iam:CreateRole`，`iam:CreatePolicy` 和`iam:AttachRolePolicy` 权限 或者 你想手动创建 IAM 角色	请执行以下操作之一： • 使用以下步骤创建服务角色。 • 请您的管理员使用以下步骤创建服务角色。

创建用于接收结果的服务角色

Note

只有在您没有使用 AWS Clean Rooms 控制台创建服务角色的必要权限时，您或您的 IAM 管理员才应遵循此步骤。

1. 按照AWS Identity and Access Management 用户指南中的使用自定义信任策略创建角色（控制台）过程进行操作。
2. 根据使用自定义信任策略创建角色（控制台）过程使用以下自定义信任策略。

```
{
    "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-EXAMPLEaaaaa*"
                }
            }
        },
        {
            "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. 根据使用自定义信任策略创建角色（控制台）过程使用以下权限策略。

> **Note**
>
> 以下示例策略支持读取 AWS Glue 元数据及其相应的 Amazon S3 数据所需的权限。但是，您可能需要修改此策略，具体取决于您设置 S3 数据的方式。
> 您的 AWS Glue 资源和底层 Amazon S3 资源必须与 AWS Clean Rooms 协作 AWS 区域相同。

```
{
    "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. 用您自己的信息替换每个*###*：

   - *region* - AWS 区域的名称。例如，**us-east-1**。
   - *a1b2c3d4-5678-90ab-cdef-EXAMPLEaaaaa* - 可以查询的成员的成员身份 ID。可以在协作的详细信息选项卡上找到成员身份 ID。这样可以确保 AWS Clean Rooms 只有当该成员在此协作中运行分析时才担任该角色。

- *arn: aws: cleanrooms: us-east-1:555555555555: membership/ a1b2c3d4-5678-* 90ab-cdef-exampleaaaa — 可以查询的会员的单一会员 ARN。可以在协作的详细信息选项卡上找到成员身份 ARN。这样可以确保 AWS Clean Rooms 只有当该成员在此协作中运行分析时才担任该角色。
- *bucket_name* - S3 存储桶的 Amazon 资源名称 (ARN)。Amazon 资源名称 (ARN) 可在 Amazon S3 存储桶的属性选项卡上找到。
- *## ID* — AWS 账户 S3 存储桶所在的 ID。

  bucket_name/optional_key_prefix - S3 中结果目标的 Amazon 资源名称 (ARN)。Amazon 资源名称 (ARN) 可在 Amazon S3 存储桶的属性选项卡上找到。

5. 继续按照使用自定义信任策略创建角色（控制台）过程创建角色。

为 AWS Clean Rooms ML 设置服务角色

主题

- 创建服务角色以读取训练数据
- 创建服务角色以写入相似细分
- 创建服务角色以读取种子数据

创建服务角色以读取训练数据

AWS Clean Rooms 使用服务角色读取训练数据。如果您具有必要的 IAM 权限，则可以使用控制台创建此角色。如果您没有`CreateRole`权限，请要求您的管理员创建服务角色。

创建服务角色以训练数据集

1. 使用您的管理员账户登录到 IAM 控制台 (https://console.aws.amazon.com/iam/)。
2. 在访问管理下，选择策略。
3. 选择创建策略。
4. 在策略编辑器中，选择 JSON 选项卡，然后复制粘贴以下策略。

> **Note**
>
> 以下示例策略支持读取 AWS Glue 元数据及其相应的 Amazon S3 数据所需的权限。但是，您可能需要修改此策略，具体取决于您设置 S3 数据的方式。此策略不包括用于解密数据的 KMS 密钥。
>
> 您的 AWS Glue 资源和底层 Amazon S3 资源必须与 AWS Clean Rooms 协作 AWS 区域相同。

```
{
    "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:region:accountId:database/databases",
                "arn:aws:glue:region:accountId:table/databases/tables",
                "arn:aws:glue:region:accountId:catalog",
                "arn:aws:glue:region:accountId:database/default"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateDatabase"
            ],
            "Resource": [
                "arn:aws:glue:region:accountId:database/default"
            ]
        },
        {
```

```
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket",
                "s3:GetBucketLocation"
            ],
            "Resource": [
                "arn:aws:s3:::bucket"
            ],
            "Condition":{
                "StringEquals":{
                    "s3:ResourceAccount":[
                        "accountId"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucketFolders/*"
            ],
            "Condition":{
                "StringEquals":{
                    "s3:ResourceAccount":[
                        "accountId"
                    ]
                }
            }
        }
    ]
}
```

如果您需要使用 KMS 密钥来解密数据，请将以下 AWS KMS 语句添加到之前的模板中：

```
{
            "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. 选择下一步。
6. 对于查看并创建，输入策略名称和描述，然后查看摘要。
7. 选择 创建策略。

   您已经为创建了策略 AWS Clean Rooms。
8. 在 Access management（访问管理）下，请选择 Roles（角色）。

   通过使用角色，您可以创建短期凭证，建议这样做以提高安全性。您也可以选择用户来创建长期凭证。
9. 选择 创建角色。
10. 在创建角色向导中，对于可信实体类型，选择自定义信任策略。
11. 将以下自定义信任策略复制粘贴到 JSON 编辑器中。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowAssumeRole",
            "Effect": "Allow",
            "Principal": {
                "Service": "cleanrooms-ml.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEqualsIfExists": {
                    "aws:SourceAccount": ["accountId"]
                },
                "StringLikeIfExists": {
```

```
                    "aws:SourceArn": "arn:aws:cleanrooms-
ml:region:account:training-dataset/*"
                }
            }
        }
    ]
}
```

永远SourceAccount是你的 AWS 账户。可以将 SourceArn 限制为特定的训练数据集，但仅在创建该数据集之后。由于您无法预先知道训练数据集 ARN，因此，此处指定了通配符。

12. 选择下一步，在添加权限下面，输入您刚刚创建的策略的名称。（您可能需要重新加载页面。）
13. 选中您创建的策略旁边的复选框，然后选择下一步。
14. 对于命名、查看和创建，输入角色名称和描述。

> **Note**
>
> 角色名称必须与授予可以查询和接收结果的成员和成员角色的 passRole 权限中的模式相匹配。

    a. 查看选择受信任的实体，并在必要时进行编辑。
    b. 在添加权限中查看权限，并在必要时进行编辑。
    c. 查看标签，并在必要时添加标签。
    d. 选择 创建角色。
15. 的服务角色 AWS Clean Rooms 已创建。

创建服务角色以写入相似细分

AWS Clean Rooms 使用服务角色将相似的区段写入存储桶。如果您具有必要的 IAM 权限，则可以使用控制台创建此角色。如果您没有CreateRole权限，请要求您的管理员创建服务角色。

创建服务角色以写入相似细分

1. 使用您的管理员账户登录到 IAM 控制台 (https://console.aws.amazon.com/iam/)。
2. 在访问管理下，选择策略。
3. 选择创建策略。

4. 在策略编辑器中，选择 JSON 选项卡，然后复制粘贴以下策略。

> **Note**
>
> 以下示例策略支持读取 AWS Glue 元数据及其相应的 Amazon S3 数据所需的权限。但是，您可能需要修改此策略，具体取决于您设置 S3 数据的方式。此策略不包括用于解密数据的 KMS 密钥。
> 您的 AWS Glue 资源和底层 Amazon S3 资源必须与 AWS Clean Rooms 协作 AWS 区域相同。

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

如果您需要使用 KMS 密钥来加密数据，请将以下 AWS KMS 语句添加到模板中：

```
{
            "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*"
                }
            }
        }
    ]
}
```

如果您需要使用 KMS 密钥来解密数据，请将以下 AWS KMS 语句添加到模板中：

```
{
            "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. 选择下一步。
6. 对于查看并创建，输入策略名称和描述，然后查看摘要。
7. 选择 创建策略。

   您已经为创建了策略 AWS Clean Rooms。
8. 在 Access management（访问管理）下，请选择 Roles（角色）。

   通过使用角色，您可以创建短期凭证，建议这样做以提高安全性。您也可以选择用户来创建长期凭证。
9. 选择 创建角色。
10. 在创建角色向导中，对于可信实体类型，选择自定义信任策略。
11. 将以下自定义信任策略复制粘贴到 JSON 编辑器中。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowAssumeRole",
            "Effect": "Allow",
            "Principal": {
                "Service": "cleanrooms-ml.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEqualsIfExists": {
                    "aws:SourceAccount": ["accountId"]

                },
                "StringLikeIfExists": {
                    "aws:SourceArn": "arn:aws:cleanrooms-
ml:region:account:configured-audience-model/*"
                }
```

```
            }
        }
    ]
}
```

永远SourceAccount是你的 AWS 账户。可以将 SourceArn 限制为特定的训练数据集，但仅在创建该数据集之后。由于您无法预先知道训练数据集 ARN，因此，此处指定了通配符。

12. 选择下一步。
13. 选中您创建的策略旁边的复选框，然后选择下一步。
14. 对于命名、查看和创建，输入角色名称和描述。

> **Note**
>
> 角色名称必须与授予可以查询和接收结果的成员和成员角色的 passRole 权限中的模式相匹配。

    a. 查看选择受信任的实体，并在必要时进行编辑。
    b. 在添加权限中查看权限，并在必要时进行编辑。
    c. 查看标签，并在必要时添加标签。
    d. 选择 创建角色。
15. 的服务角色 AWS Clean Rooms 已创建。

创建服务角色以读取种子数据

AWS Clean Rooms 使用服务角色读取种子数据。如果您具有必要的 IAM 权限，则可以使用控制台创建此角色。如果您没有CreateRole权限，请要求您的管理员创建服务角色。

创建服务角色以读取种子数据

1. 使用您的管理员账户登录到 IAM 控制台 (https://console.aws.amazon.com/iam/)。
2. 在访问管理下，选择策略。
3. 选择创建策略。
4. 在策略编辑器中，选择 JSON 选项卡，然后复制粘贴以下策略。

> **Note**
>
> 以下示例策略支持读取 AWS Glue 元数据及其相应的 Amazon S3 数据所需的权限。但是，您可能需要修改此策略，具体取决于您设置 S3 数据的方式。此策略不包括用于解密数据的 KMS 密钥。
> 您的 AWS Glue 资源和底层 Amazon S3 资源必须与 AWS Clean Rooms 协作 AWS 区域相同。

```
{
    "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"
                    ]
```

```
                }
            }
        }
    ]
}
```

如果您需要使用 KMS 密钥来解密数据，请将以下 AWS KMS 语句添加到模板中：

```
{
            "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. 选择下一步。
6. 对于查看并创建，输入策略名称和描述，然后查看摘要。
7. 选择 创建策略。

   您已经为创建了策略 AWS Clean Rooms。
8. 在 Access management（访问管理）下，请选择 Roles（角色）。

   通过使用角色，您可以创建短期凭证，建议这样做以提高安全性。您也可以选择用户来创建长期凭证。
9. 选择 创建角色。
10. 在创建角色向导中，对于可信实体类型，选择自定义信任策略。
11. 将以下自定义信任策略复制粘贴到 JSON 编辑器中。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowAssumeRole",
            "Effect": "Allow",
            "Principal": {
                "Service": "cleanrooms-ml.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEqualsIfExists": {
                    "aws:SourceAccount": ["accountId"]

                },
                "StringLikeIfExists": {
                    "aws:SourceArn": "arn:aws:cleanrooms-
ml:region:account:audience-generation-job/*"
                }
            }
        }
    ]
}
```

永远SourceAccount是你的 AWS 账户。可以将 SourceArn 限制为特定的训练数据集，但仅在创建该数据集之后。由于您无法预先知道训练数据集 ARN，因此，此处指定了通配符。

12. 选择下一步。
13. 选中您创建的策略旁边的复选框，然后选择下一步。
14. 对于命名、查看和创建，输入角色名称和描述。

> **Note**
>
> 角色名称必须与授予可以查询和接收结果的成员和成员角色的 passRole 权限中的模式相匹配。

a. 查看选择受信任的实体，并在必要时进行编辑。

b. 在添加权限中查看权限，并在必要时进行编辑。

c. 查看标签，并在必要时添加标签。

d. 选择 创建角色。

15. 的服务角色 AWS Clean Rooms 已创建。

在 AWS Clean Rooms 中创建协作

协作是 AWS Clean Rooms 中的安全逻辑边界，成员可以在其中对配置表执行 SQL 查询。

AWS Clean Rooms 中的任何成员都可以创建协作。

协作创建者可以指定一个成员来查询和接收结果。但是，协作创建者可能希望阻止可以查询的成员访问查询结果。在这种情况下，协作创建者可以指定一名可以查询的成员，另一名可以接收结果的成员。

在大多数情况下，可以查询的成员也是为查询计算费用付费的成员。但是，协作创建者可以将其他成员配置为负责为查询计算费用付费。

有关如何使用 AWS SDK 创建协作的信息，请参阅 AWS Clean Rooms API 参考。

主题

创建协作

在开始之前，请确保您已完成以下先决条件：

- 您拥有要邀请参与协作的每位成员的姓名和 AWS 账户 ID。
- 您有权与协作的所有成员共享每个成员的姓名和 AWS 账户 ID。

> **Note**
>
> 创建协作后，您无法添加更多成员。

使用 AWS Clean Rooms 控制台创建协作

1. 登录 AWS Management Console并打开 AWS Clean Rooms 控制台，AWS 账户作为协作创建者。
2. 在左侧导航窗格中，选择协作。
3. 在右上角，选择创建协作。
4. 对于步骤 1: 定义协作，请执行以下操作：

a. 在详细信息中，输入协作的名称和描述。

受邀参与协作的协作成员将可以看到这些信息。名称和描述可帮助他们了解协作的意义。

b. 对于成员：

i. 对于成员 1: 您，输入您希望在协作中显示的成员显示名称。

> **Note**
>
> 成员 AWS 账户 ID 会自动包含您的 AWS 账户 ID。

ii. 在成员 2 中，输入要邀请参与协作的成员的成员显示名称和成员 AWS 账户 ID。

所有受邀参与协作的人都可以看到成员显示名称和成员 AWS 账户 ID。输入并保存这些字段的值后，它们不可编辑。

> **Note**
>
> 您必须告知协作成员，协作中所有受邀和活跃的协作者都将看到他们的成员 AWS 账户 ID 和成员显示名称。

iii. 如果要添加其他成员，请选择添加其他成员。然后输入要邀请参与协作以贡献数据的每个成员的成员显示名称和成员 AWS 账户 ID。

c. 对于成员能力，请选择以下选项之一，

如果要...	操作...
查询协作中的数据并接收结果	1. 将自己选为可以运行查询的成员。 2. 将可以接收结果的成员的默认设置保留为与运行查询的人相同。
查询协作中的数据并分配其他成员来接收结果	1. 将自己选为可以运行查询的成员。 2. 从下拉列表中选择可以接收结果的成员。

如果要...	操作...
接收协作中的查询结果并分配其他成员来查询数据	1. 从下拉列表中选择可以运行查询的成员。 2. 从下拉列表中将自己选为可以接收结果的成员。
创建和管理协作，分配其他成员来查询数据，并分配其他成员来接收结果	1. 从下拉列表中选择可以运行查询的成员。 2. 从下拉列表中选择可以接收结果的成员。

d. 对于付款配置，选择下列选项之一：

如果要...	操作...
将可以运行查询的成员指定为支付查询计算费用的成员	将为查询付费的成员的默认设置保留为与运行查询的人相同。
分配其他成员来支付查询计算费用	从下拉列表中选择将为查询付费的成员。

e. 如果要启用查询日志记录，请选中支持对此协作进行查询日志记录复选框。

f. 如果要启用加密计算功能，请选中支持在此协作中进行加密计算复选框，然后选择以下加密计算参数：

- 允许 cleartext 列

  如果您不希望在加密表中允许 cleartext 列，请选择否。

  如果您希望在加密表中允许 cleartext 列，请选择是。

  要在特定列上运行 SUM 或 AVG，这些列必须是 cleartext。

- 允许重复

  如果您不希望 fingerprint 列中允许重复条目，请选择否。

  如果您希望 fingerprint 列中允许重复条目，请选择是。

- 允许对具有不同名称的列进行 JOIN

如果您不希望对具有不同名称的 fingerprint 列进行联接，请选择否。

如果您希望对具有不同名称的 fingerprint 列进行联接，请选择是。

- 保留 NULL 值

  如果您不希望保留 NULL 值，请选择否。NULL 值不会在加密表中显示为 NULL。

  如果您希望保留 NULL 值，请选择是。NULL 值将在加密表中显示为 NULL。

有关加密计算参数的更多信息，请参阅加密计算参数。

有关如何解密数据以在 AWS Clean Rooms 中使用的更多信息，请参阅使用 Clean Rooms 加密计算准备加密的数据表。

> **Note**
>
> 在完成下一步之前，请仔细验证这些配置。创建协作后，您只能编辑协作名称、描述以及查询日志是否存储在 Amazon CloudWatch Logs 中。

g. 如果要为协作资源启用标签，请选择添加新标签，然后输入键和值对。

h. 选择 Next（下一步）。

5. 对于步骤 2: 配置成员身份，执行以下操作：

   a. 选择一个选项：

如果选择...	操作...
是，立即通过创建成员身份来加入	同时创建协作和您的成员身份。 您在协作中的状态为活跃。
不，我将稍后创建成员身份	仅创建协作。 您在协作中的状态为非活跃。

   b. 如果您是可以接收结果的成员，请在查询结果设置默认值下选择一个选项：

如果您...	操作...
保持立即设置默认设置复选框处于选中状态。（默认处于选中状态。）	1. 对于 Amazon S3 中的结果目标，请输入 Amazon S3 目标。 2. 对于查询结果格式，请选择 CSV 或 PARQUET。
清除立即设置默认设置复选框	仅创建协作。 您在协作中的状态为非活跃。

c. 如果您在步骤 4.e 中选择启用查询日志记录，请为 Amazon CloudWatch Logs 中的日志存储选择以下选项之一：

如果选择...	操作...
打开	与您相关的查询日志存储在 Amazon CloudWatch Logs 中。 每个成员只能接收他们发起的查询或包含其数据的查询的日志。 可以接收结果的成员还会收到协作中运行的所有查询的日志，即使查询中未访问他们的数据也是如此。
关闭	与您相关的查询日志存储在 Amazon CloudWatch Logs 账户中。

> **Note**
>
> 打开查询日志记录后，可能需要几分钟才能设置日志存储并开始在 Amazon CloudWatch Logs 中接收日志。在这段短暂的时间内，可以查询的成员可能会运行实际上并未发送日志的查询。

d. 如果要为成员身份资源启用标签，请选择添加新标签，然后输入键和值对。

e. 如果您是支付查询费用的成员，请选中我同意支付此协作中的查询计算费用复选框以表示您接受。

> **Note**
>
> 必须选中此复选框才能继续。
> 有关如何计算费用的更多信息，请参阅的定价 AWS Clean Rooms。

如果您是为查询计算费用付费的成员，但不是可以查询的成员，则建议您使用 AWS Budgets 来配置 AWS Clean Rooms 预算，并在达到预算上限时接收通知。有关设置预算的更多信息，请参阅《AWS Cost Management 用户指南》中的使用 AWS Budgets 管理成本。有关设置通知的更多信息，请参阅《AWS Cost Management 用户指南》中的针对预算通知创建 Amazon SNS 主题。如果已达到预算上限，您可以联系可以查询的成员或退出协作。如果您退出协作，将不再允许运行查询，因此将不再向您收取查询计算费用。

f. 选择 Next（下一步）。

6. 对于步骤 3: 查看并创建，执行以下操作：

a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。

b. 选择以下选项之一：

如果您选择了...	则选择...
同步创建成员身份和协作（是，立即通过创建成员身份来加入）	创建协作和成员身份
创建协作，此时不创建成员身份（不，我将稍后创建成员身份）	创建协作

成功创建协作后，您可以在协作下看到协作详细信息页面。

后续步骤

您现在已准备好执行以下操作：

- 准备要在 AWS Clean Rooms 中查询的数据表（如果您想查询自己的数据，则是可选的。）

- 将配置表与协作关联。（如果您想查询自己的数据，则是可选的。）
- 为配置表配置分析规则。（如果您想查询自己的数据，则是可选的。）
- 创建成员身份并加入协作。
- 管理协作。

创建成员身份并加入协作

成员身份是成员在 AWS Clean Rooms中加入协作时创建的资源。

您可以作为可以查询数据的成员和/或可以接收查询结果的成员的身份加入协作。您也可以作为为查询计算费用付费的成员的身份加入协作。所有成员都可以贡献数据。

有关如何使用 AWS SDK 创建成员身份和加入协作的信息，请参阅 AWS Clean Rooms API 参考。

主题

创建成员身份并加入协作

创建成员身份并加入协作

1. 登录 AWS Management Console 并与您的成员一起打开AWS Clean Rooms 控制台 AWS 账户。
2. 在左侧导航窗格中，选择协作。
3. 在可加入选项卡上，对于可供加入的协作，选择协作的名称。
4. 在协作详细信息页面上，查看协作详细信息，包括您的成员详细信息和其他成员列表。

   验证协作中每个成员的 AWS 账户 ID 是否都是您打算与之一起参与协作的 ID。
5. 选择创建成员身份。
6. 在创建成员资格页面的概述中，查看协作名称、协作描述、协作创建者的 AWS 账户 ID、您的成员权限以及将为查询付费的成员的 AWS 账户 ID。
7. 如果协作创建者选择启用查询日志，请选择以下选项之一作为 Amazon Logs 中的 CloudWatch 日志存储：

如果选择...	操作...
打开	与您相关的查询日志存储在 Amazon CloudWatch 日志中。 每个成员只能接收他们发起的查询或包含其数据的查询的日志。

如果选择...	操作...
	能够接收结果的成员还会收到协作中运行的所有查询的日志，即使查询中未访问他们的数据。
关闭	与您相关的查询日志不会存储在您的 Amazon CloudWatch Logs 账户中。

> **Note**
>
> 开启查询日志后，可能需要几分钟才能设置日志存储并开始在 Amazon Logs 中接收 CloudWatch 日志。在这段短暂的时间内，可以查询的成员可能会运行实际上并未发送日志的查询。

8. 如果您的成员能力包括接收结果：

   a. 对于查询结果设置，

      i. 通过输入 S3 目标指定 Amazon S3 中的结果目标，或者选择浏览 S3 从可用 S3 存储桶列表中进行选择。

         Example

         例如：**`s3://bucket/prefix`**

      ii. 对于查询结果格式（CSV 或 PARQUET）。

   b. 对于服务访问，选择创建并使用新的服务角色或使用现有服务角色。

> **Note**
>
> 您必须选择现有的服务角色或拥有创建新服务角色的权限。有关更多信息，请参阅 创建服务角色以接收结果。

9. 如果要为成员身份资源启用标签，请选择添加新标签，然后输入键和值对。
10. 如果协作创建者已将您指定为为查询付费的成员，请选中我同意支付此协作中的查询计算费用复选框以表示您接受。

> **Note**
>
> 必须选中此复选框才能继续。
> 有关如何计算费用的更多信息，请参阅的定价 AWS Clean Rooms。

如果您是支付查询计算费用的会员，但不是可以查询的成员，则建议您使用 AWS Budgets 来配置预算， AWS Clean Rooms 并在达到最高预算后接收通知。有关设置预算的更多信息，请参阅《AWS Cost Management 用户指南》中的使用 AWS Budgets管理成本。有关设置通知的更多信息，请参阅《AWS Cost Management 用户指南》中的针对预算通知创建 Amazon SNS 主题。如果已达到预算上限，您可以联系可以查询的成员或退出协作。如果您退出协作，将不再允许运行查询，因此将不再向您收取查询计算费用。

11. 如果您确定要创建成员身份并加入协作，请选择创建成员身份。

授予您对协作元数据的读取权限。这包括协作的显示名称和描述等信息，以及其他成员的所有姓名和 AWS 账户 ID。

有关如何退出协作的信息，请参阅退出协作。

后续步骤

您现在已准备好执行以下操作：

- 准备好要查询的数据表。 AWS Clean Rooms（如果您想查询自己的数据，则是可选的。）
- 将配置表与协作关联。
- 为配置表配置分析规则。

在中为查询准备数据表 AWS Clean Rooms

> **Note**
>
> 准备数据表可以在您加入协作之前或之后进行。准备好表格后，只要您对该表格的隐私需求相同，您就可以在多个协作中重复使用该表格。

作为协作成员，您必须先准备好数据表，然后才能 AWS Clean Rooms 由可以查询的协作成员进行查询。

如果您的用例不需要您自带数据，则可以跳过此过程。

如果您的数据表已在中编目 AWS Glue，请跳至。在 AWS Clean Rooms中创建配置表

准备数据表涉及以下步骤：

- 步骤 1：完成先决条件
- 步骤 2：（可选）准备用于加密计算的数据
- 步骤 3：将数据表上传到 Amazon S3
- 步骤 4：创建 AWS Glue 表
- 后续步骤

有关可用于查询的数据格式的更多信息，请参阅的数据格式 AWS Clean Rooms。

步骤 1：完成先决条件

要准备数据表以供使用 AWS Clean Rooms，必须满足以下先决条件：

- 您的数据集必须另存为 AWS Clean Rooms支持的数据格式之一。
- 您的数据表必须编入目录 AWS Glue 并使用支持的数据类型。 AWS Clean Rooms
- 您的所有数据表都必须存储在亚马逊简单存储服务 (Amazon S3) 中， AWS 区域 与创建协作时相同。
- AWS Glue Data Catalog 必须位于创建协作的同一区域。
- AWS Glue Data Catalog 必须与成员资格 AWS 账户 相同。

- 无法向注册 Amazon S3 存储桶 AWS Lake Formation。
- 协作创建者已在 AWS Clean Rooms中建立了协作。有关更多信息，请参阅 在 AWS Clean Rooms 中创建协作。
- 协作创建者已将协作 ID 发送给作为协作参与者的您。

步骤 2：（可选）准备用于加密计算的数据

（可选）如果您使用的是加密计算，并且您的数据表包含要加密的敏感信息，则必须使用 C3R 加密客户端对数据表进行加密。

要为加密计算准备数据，请按照使用 Clean Rooms 加密计算准备加密的数据表中的步骤操作。

步骤 3：将数据表上传到 Amazon S3

> **Note**
>
> 如果您打算在协作中使用加密的数据表，则必须先加密数据以进行加密计算，然后再将数据表上传到 Amazon S3。有关更多信息，请参阅 使用 Clean Rooms 加密计算准备加密的数据表。

将数据表上传到 Amazon S3

1. 登录 AWS Management Console 并打开 Amazon S3 控制台，网址为 https://console.aws.amazon.com/s3/。
2. 选择桶，然后选择您想要用于存储数据表的桶。
3. 选择上传，然后按照提示进行操作。
4. 选择对象选项卡，查看存储数据的前缀。记下文件夹的名称。

   您可以选择用于查看数据的文件夹。

步骤 4：创建 AWS Glue 表

如果您已经有 AWS Glue 数据表，则可以跳过此步骤。

在此步骤中，您将在中设置一个抓取器 AWS Glue ，用于抓取 S3 存储桶中的所有文件并创建 AWS Glue 表。有关更多信息，请参阅《AWS Glue 用户指南》 AWS Glue中的定义抓取工具。

有关支持 AWS Glue Data Catalog 的数据类型的更多信息，请参阅支持的数据类型。

> **Note**
>
> AWS Clean Rooms 目前不支持向注册的 S3 存储桶。 AWS Lake Formation

以下过程描述了如何创建 AWS Glue 表。如果要使用带有 AWS Key Management Service (AWS KMS) 密钥的加密 AWS Glue Data Catalog 对象，则需要配置 KMS 密钥权限策略以允许访问该加密表。有关更多信息，请参阅《AWS Glue 开发人员指南》中的在 AWS Glue 中设置加密。

创建 AWS Glue 表

1. 按照《AWS Glue 用户指南》中的 “在AWS Glue 控制台上使用抓取工具” 步骤进行操作。
2. 记下 AWS Glue 数据库名称和 AWS Glue 表名。

后续步骤

现在，您已经准备好了数据表，您已准备好：

- 创建已配置的表
- 创建 ML 模型

的数据格式 AWS Clean Rooms

您在中用于查询的数据集 AWS Clean Rooms 通常与用于其他应用程序的数据集类型相同。例如，亚马逊 Athena、亚马逊 EMR、亚马逊 Redshift Spectrum 和亚马逊使用相同类型的数据集。 QuickSight 您可以直接从 Amazon Simple Storage Service (Amazon S3) 以数据的原始格式查询数据。

要查询数据，数据集必须采用 AWS Clean Rooms 支持的格式。包含数据集的 Amazon S3 存储桶和集 AWS Clean Rooms 群必须位于同一存储桶中 AWS 区域。

支持的数据格式

AWS Clean Rooms 支持以下结构化格式：

- Apache Iceberg 表
- Parquet

- RCFile
- TextFile
- SequenceFile
- RegexSerde
- OpenCSV
- AVRO
- JSON

> **Note**
>
> 文本文件中的 `timestamp` 值必须采用 `yyyy-MM-dd HH:mm:ss.SSSSSS` 格式。例如：`2017-05-01 11:30:59.000000`。

我们建议使用列式存储文件格式（例如 Apache Parquet）。利用列式存储文件格式，您可以通过仅选择所需的列来最大程度地减少 Amazon S3 外部的数据传输。为了获得最佳性能，应将大型对象拆分为 100 MB - 1 GB 的对象。

支持的数据类型

为了获得最佳的使用体验 AWS Clean Rooms，必须将您的所有数据编入其中。 AWS Glue有关更多信息，请参阅《AWS Glue 开发人员指南》中的 AWS Glue Data Catalog入门。

AWS Clean Rooms 支持以下 AWS Glue Data Catalog 数据类型：

- bigint
- boolean
- char
- date
- decimal
- double
- float
- int
- 嵌套数据类型，例如：

  - array
  - map
  - struct
- smallint
- string
- timestamp
- varchar

AWS Clean Rooms 不支持：

- binary
- interval

的文件压缩类型 AWS Clean Rooms

要减少存储空间、提高性能和最大程度地降低成本，我们强烈建议您压缩数据集。

AWS Clean Rooms 根据文件扩展名识别文件压缩类型，并支持下表所示的压缩类型和扩展名。

压缩算法	文件扩展名
GZIP	.gz
Bzip2	.bz2
Snappy	.snappy

可以在不同的级别应用压缩。最常见的情况是，压缩整个文件或压缩文件中的单个块。在文件级压缩列格式不会产生性能优势。

服务器端加密 AWS Clean Rooms

> **Note**
>
> 对于需要加密计算的使用案例，服务器端加密并不能取代加密计算。

AWS Clean Rooms 透明地解密使用以下加密选项加密的数据集：

- SSE-S3 — 使用由 Amazon S3 管理的 AES-256 加密密钥的服务器端加密
- SSE-KMS — 使用由管理的密钥进行服务器端加密 AWS Key Management Service

要使用 SSE-S3，用于将配置的表与协作关联的 AWS Clean Rooms 服务角色必须具有 KMS-Decrypt 权限。要使用 SSE-KMS，KMS 密钥策略还必须允许 AWS Clean Rooms 服务角色解密。

AWS Clean Rooms 不支持 Amazon S3 客户端加密。有关服务器端加密的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的使用服务器端加密保护数据。

在中使用Apache Iceberg表格 AWS Clean Rooms

Apache Iceberg是一种用于数据湖的开源表格式。 AWS Clean Rooms 可以使用存储在Apache Iceberg元数据中的统计信息来优化查询计划并减少无尘室查询处理期间的文件扫描。有关更多信息，请参阅 Apache Iceberg 文档。

AWS Clean Rooms 与 Iceberg 表一起使用时，请考虑以下几点：

- AWS Glue Data Catalog 唯一的表Apache Iceberg格 — 表必须在 AWS Glue Data Catalog 基于开源胶水目录实现的中定义。
- Parquet 文件格式 — AWS Clean Rooms 仅支持 Parquet 数据文件格式的 Iceberg 表。
- GZIP 和 Snappy 压缩 — AWS Clean Rooms 支持带有 GZIP 和压缩功能的 Parquet。Snappy
- Iceberg 版本 — AWS Clean Rooms 支持对版本 1 和版本 2 的 Iceberg 表运行查询。
- 分区-您无需在中为Apache Iceberg表手动添加分区 AWS Glue。 AWS Clean Rooms 自动检测 Apache Iceberg表中的新分区，无需手动操作即可更新表定义中的分区。Iceberg 分区在 AWS Clean Rooms 表架构中显示为常规列，而不是在配置表架构中单独显示为分区键。
- 限制
  - 仅限全新 Iceberg 表

    不支持从 Apache Parquet 表转换的 Apache Iceberg 表。
  - 时间旅行查询

    AWS Clean Rooms 不支持使用Apache Iceberg表进行时空旅行查询。
  - Athena 引擎版本 2

    不支持使用 Athena 引擎版本 2 创建的 Iceberg 表。

- 文件格式

  不支持 Avro 和优化行列式 (ORC) 文件格式。
- 压缩

  不支持 Parquet 的 Zstandard (Zstd) 压缩。

支持的 Iceberg 表数据类型

AWS Clean Rooms 可以查询包含以下数据类型的Iceberg表：

- boolean
- date
- decimal
- double
- float
- int
- list
- long
- map
- string
- struct
- timestamp without time zone

有关 Iceberg 数据类型的更多信息，请参阅 Apache 文档中的 Schemas for Iceberg。

使用 Clean Rooms 加密计算准备加密的数据表

Clean Rooms(C3R) 的加密计算是一项功能。 AWS Clean Rooms您可以使用 C3R 以加密方式限制任何一方和协作 AWS 中可以学到的内容。 AWS Clean Rooms

在将数据表上传到 Amazon Simple Storage Service (Amazon S3) 之前，您可以使用 C3R 加密客户端（一种客户端加密工具）对数据表进行加密。

有关更多信息，请参阅 Clean Rooms 加密计算。

使用 C3R 准备加密的数据表涉及以下步骤：

步骤

- 步骤 1：完成先决条件
- 步骤 2：下载 C3R 加密客户端
- （可选）步骤 3：查看 C3R 加密客户端中的可用命令。
- 步骤 4：为表格文件生成加密架构
- 步骤 5：创建共享密钥
- 步骤 6：将共享密钥存储在 环境变量中。
- 步骤 7：加密数据
- 步骤 8：验证数据加密
- （可选）创建架构（高级用户）

步骤 1：完成先决条件

要准备数据表以供 C3R 使用，您必须满足以下先决条件：

- 您可以通过以下网址访问Clean Rooms存储库的加密计算： GitHub

  https://github.com/aws/c3r
- 您已经设置了使用 C3R 加密客户端的 AWS 凭据。C3R 加密客户端使用这些凭据进行只读 API 调用， AWS Clean Rooms 以检索协作元数据。有关更多信息，请参阅《AWS Command Line Interface 用户指南版本 2》中的配置 AWS CLI。
- 您的计算机上安装了 Java Runtime Environment (JRE) 11 或更高版本。

  - 推荐的 Java Runtime Environment、Amazon Corretto 11 或更高版本可以从 https://aws.amazon.com/corretto 下载。
  - Java Development Kit(JDK) 包括同一个版本的对应 JRE。但是，运行 Clean Rooms 加密计算 (C3R) 加密客户端不需要 JDK 的附加功能。
- 您的表格数据文件 (.csv) 或 Parquet 文件 (.parquet) 保存在本地。
- 您或协作中的其他成员可以创建共享密钥。有关更多信息，请参阅 步骤 5：创建共享密钥。
- 协作创建者创建了一种协作， AWS Clean Rooms 其中启用了加密计算，以实现协作。有关更多信息，请参阅 在 AWS Clean Rooms 中创建协作。
- 协作创建者已将协作 ID 发送给作为协作参与者的您。发送的邀请中包含协作的 Amazon 资源名称 (ARN)，其中包含协作 ID。

步骤 2：下载 C3R 加密客户端

要从以下网址下载 C3R 加密客户端 GitHub

1. 前往密码计算获取Clean RoomsAWSGitHub存储库：https://github.com/aws/c3r
2. 选择并下载文件。

   源代码、许可证和相关资料可以从 GitHub 存储库的登陆页面克隆或下载为 .zip 文件。（参见存储库内容列表右上角的代码按钮）。

   最新签名的 C3R 加密客户端 Java Executable File（即命令行界面应用程序）位于 GitHub 存储库的发布页面上。

   Apache Spark 的 C3R 加密客户端包 (`c3r-cli-spark`) 是 c3r-cli 的一个版本，必须作为作业提交给正在运行的 Apache Spark 服务器。有关更多信息，请参阅在 Apache Spark 上运行 C3R。

（可选）步骤 3：查看 C3R 加密客户端中的可用命令。

使用此过程熟悉 C3R 加密客户端中的可用命令。

查看 C3R 加密客户端中的所有可用命令。

1. 从命令行界面 (CLI)，导航到包含已下载 c3r-cli.jar 文件的文件夹。
2. 运行以下命令：`java -jar c3r-cli.jar`
3. 查看可用命令和选项的列表。

步骤 4：为表格文件生成加密架构

要加密数据，需要一个描述如何使用数据的加密架构。本节介绍 C3R 加密客户端如何帮助为带有标题行的 CSV 文件或 Parquet 文件生成加密架构。

每个文件只需进行一次该操作。架构存在后，可以重复使用它来加密同一个文件（或任何具有相同列名的文件）。如果列名或所需的加密架构发生变化，则必须更新架构文件。有关更多信息，请参阅（可选）创建架构（高级用户）。

> **Important**
>
> 所有协作方都必须使用相同的共享密钥。如果要在查询中对列名进行 JOIN 或以其他方式比较列名是否相等，各协作方还应协调列名，使其相匹配。否则，SQL 查询可能会产生意外或不正确的结果。但是，如果协作创建者在创建协作期间启用了 `allowJoinsOnColumnsWithDifferentNames` 加密设置，则无需这样做。有关加密相关设置的更多信息，请参阅加密计算参数。

在架构模式下运行时，C3R 加密客户端会逐列浏览输入文件，提示您是否以及如何处理该列。如果文件中包含许多加密输出不需要的列，交互式模式生成可能会变得繁琐，因为您必须跳过每个不需要的列。为避免这种情况，您可以手动编写架构，或者创建仅包含所需列的输入文件的简化版本。然后，可以在该简化的文件上运行交互式架构生成器。C3R 加密客户端会输出有关架构文件的信息，并询问您应如何在目标输出中包含或加密源列（如果有）。

对于输入文件中的每个源列，系统会提示：

1. 应该生成多少个目标列
2. 应如何加密每个目标列（如果有）
3. 每个目标列的名称
4. 如果将列作为 sealed 列进行加密，在加密之前应如何填充数据。

> **Note**
>
> 对已加密为 sealed 列的列的数据进行加密时，必须确定哪些数据需要填充。C3R 加密客户端建议在架构生成期间使用默认填充，将列中的所有条目填充到相同的长度。
> 在确定 `fixed` 的长度时，请注意填充以字节为单位，而不是以位为单位。

以下是创建架构的决定表。

架构决定表

决策	源列中的目标列数 <' name-of-column '> ?	目标列类型：[c] cleartext 、[f] fingerprint 或 [s] sealed ?	目标列标题名称 <default 'name-of-column'>	在标题中添加 <suffix> 后缀以指示它是如何加密的，[y] 是或 [n] 否 <default 'yes'>	<' name-of-column _sealed'> 填充类型：[n] 一、[f] 固定或 [m] 最大 <default 'max'>
保持列未加密。	1	c	不适用	不适用	不适用
将列加密为 fingerprint 列。	1	f	选择默认值或输入新的标题名称。	输入 y 选择默认值 (_fingerprint) 或输入 n。	不适用
将列加密为 sealed 列。	1	s	选择默认值或输入新的标题名称。	输入 y 选择默认值 (_sealed) 或输入 n。	选择填充类型。 有关更多信息，请参阅（可选）创建架构（高级用户）。
将列同时加密为 fingerprint 和 sealed。	2	输入第一个目标列：f。 输入第二个目标列：s。	为每个目标列选择目标标题。	输入 y 选择默认值或输入 n.	选择填充类型（仅适用于 sealed 列）。 有关更多信息，请参阅（可选）创建

决策	源列中的目标列数 <' name-of-column '> ?	目标列类型：[c] cleartext 、[f] fingerprint 或 [s] sealed ?	目标列标题名称 <default 'name-of-column'>	在标题中添加 <suffix> 后缀以指示它是如何加密的，[y] 是或 [n] 否 <default 'yes'>	<' name-of-column _sealed'> 填充类型：[n] 一、[f] 固定或 [m] 最大 <default 'max'>
					架构（高级用户）。

以下是如何创建加密架构的两个示例。交互的具体内容取决于输入文件和您提供的响应。

示例

- 示例：为 fingerprint 列和 cleartext 列生成加密架构
- 示例：生成带有 sealed、fingerprint 和 cleartext 列的加密架构

示例：为 fingerprint 列和 cleartext 列生成加密架构

在此示例中，对于 `ads.csv`，只有两列：`username` 和 `ad_variant`。对于这些列，我们需要以下内容：

- 将 `username` 列加密为 `fingerprint` 列
- 将 `ad_variant` 列加密为 `cleartext` 列

为 fingerprint 列和 cleartext 列生成加密架构

1. （可选）要确保 c3r-cli.jar 文件和要加密的文件存在，请执行以下操作：

   a. 导航到所需的目录并运行 `ls`（如果使用 Mac 或 Unix/Linux）或 `dir`（如果使用 Windows）。

   b. 查看表格数据文件（例如 .csv）列表，然后选择要加密的文件。

      在此示例中，`ads.csv` 是我们要加密的文件。

2. 在 CLI 中，运行以下命令以交互方式创建架构。

```
java -jar c3r-cli.jar schema ads.csv --interactive --output=ads.json
```

> **Note**
>
> - 您可以运行 `java --jar PATH/TO/c3r-cli.jar`。或者，如果您已将 `PATH/TO/c3r-cli.jar` 添加到 CLASSPATH 环境变量中，也可以运行该类名。C3R 加密客户端将在 CLASSPATH 中查找以找到它（例如 `java com.amazon.psion.cli.Main`）。
> - `--interactive` 标志选择开发架构的交互模式。这将引导用户完成创建架构的向导。具有高级技能的用户无需使用向导即可创建自己的架构 JSON。有关更多信息，请参阅（可选）创建架构（高级用户）。
> - `--output` 标志设置输出名称。如果不包含 `--output` 标志，则 C3R 加密客户端会尝试选择默认输出名称（例如 `<input>.out.csv` 或架构的 `<input>.json`）。

3. 对于 `Number of target columns from source column 'username'?`，输入 **1**，然后按 Enter。
4. 对于 `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`，输入 **f**，然后按 Enter。
5. 对于 `Target column headername <default 'username'>`，按 Enter。

   默认名称为“username”。
6. 对于 `Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'>`，输入 **y**，然后按 Enter。

> **Note**
>
> 交互模式建议在加密的列标题中添加的后缀（fingerprint 列添加 `_fingerprint`，sealed 列添加 `_sealed`）。当您执行诸如将数据上传到 AWS 服务 或创建 AWS Clean Rooms 协作之类的任务时，这些后缀可能会有所帮助。这些后缀可以帮助指示对每列中的加密数据可以做些什么。例如，如果您将列加密为 sealed 列 (`_sealed`) 并尝试对其进行 JOIN 或尝试反向操作，则会出现问题。

7. 对于 `Number of target columns from source column 'ad_variant'?`，输入 **1**，然后按 Enter。

8. 对于 `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`，输入 **c**，然后按 Enter。
9. 对于 `Target column headername <default 'username'>`，按 Enter。

   默认名称为“`ad_variant`”。

   架构被写入名为 `ads.json` 的新文件。

   > **Note**
   >
   > 您可以通过在任何文本编辑器（例如 Windows 上的 Notepad 或 macOS 上的 TextEdit）中打开架构来查看架构。

10. 现在，您可以加密数据了。

示例：生成带有 sealed、fingerprint 和 cleartext 列的加密架构

在此示例中，对于 `sales.csv`，有三列：`username`、`purchased` 和 `product`。对于这些列，我们需要以下内容：

- 将 `product` 列加密为 `sealed` 列
- 将 `username` 列加密为 `fingerprint` 列
- 将 `purchased` 列加密为 `cleartext` 列

生成带有 sealed、fingerprint 和 cleartext 列的加密架构

1. （可选）要确保 c3r-cli.jar 文件和要加密的文件存在，请执行以下操作：
   a. 导航到所需的目录并运行 `ls`（如果使用 Mac 或 Unix/Linux）或 `dir`（如果使用 Windows）。
   b. 查看表格数据文件 (.csv) 列表并选择要加密的文件。

      在此示例中，`sales.csv` 是我们要加密的文件。
2. 在 CLI 中，运行以下命令以交互方式创建架构。

```
java -jar c3r-cli.jar schema sales.csv --interactive --
output=sales.json
```

> **Note**
>
> - `--interactive` 标志选择开发架构的交互模式。这将引导用户完成创建架构的指导性工作流程。
> - 如果您是高级用户，则无需使用指导性工作流程即可创建自己的架构 JSON。有关更多信息，请参阅 （可选）创建架构（高级用户）。
> - 对于没有列标题的 .csv 文件，请参阅 CLI 中可用的架构命令的 `--noHeaders` 标志。
> - `--output` 标志设置输出名称。如果不包含 `--output` 标志，则 C3R 加密客户端会尝试选择默认输出名称（例如 `<input>.out` 或架构的 `<input>.json`）。

3. 对于 `Number of target columns from source column 'username'?`，输入 **1**，然后按 Enter。
4. 对于 `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`，输入 **f**，然后按 Enter。
5. 对于 `Target column headername <default 'username'>`，按 Enter。

   默认名称为“username”。
6. 对于 `Add suffix '_fingerprint' to header to indicate how it was encrypted, [y]es or [n]o <default 'yes'>`，输入 **y**，然后按 Enter。
7. 对于 `Number of target columns from source column 'purchased'?`，输入 **1**，然后按 Enter。
8. 对于 `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`，输入 **c**，然后按 Enter。
9. 对于 `Target column headername <default 'purchased'>`，按 Enter。

   默认名称为“purchased”。
10. 对于 `Number of target columns from source column 'product'?`，输入 **1**，然后按 Enter。
11. 对于 `Target column type: [c]leartext, [f]ingerprint, or [s]ealed?`，输入 **s**，然后按 Enter。
12. 对于 `Target column headername <default 'product'>`，按 Enter。

    默认名称为“product”。

13. 对于 'product_sealed' padding type: [n]one, [f]ixed, or [m]ax <default 'max'?>，按 Enter 选择默认值。
14. 对于 Byte-length beyond max length to pad cleartext to in 'product_sealed' <default '0'>?，按 Enter 选择默认值。

    架构被写入名为 sales.json 的新文件。
15. 现在，您可以加密数据了。

步骤 5：创建共享密钥

要加密数据表，协作参与者必须同意并安全地共享共享密钥。

共享密钥必须至少为 256 位（32 字节）。您可以指定更大的密钥，但它不会为您提供任何额外的安全性。

> **Important**
>
> 请记住，所有协作参与者用于加密和解密的密钥和协作 ID 必须相同。

以下各节提供了控制台命令的示例，这些命令用于生成作为 secret.key 保存在相应终端当前工作目录中的共享密钥。

主题

- 示例：使用 OpenSSL 生成密钥
- 示例：使用 PowerShell 在 Windows 上生成密钥

示例：使用 OpenSSL 生成密钥

对于常见的通用密码库，请运行以下命令创建共享密钥。

```
openssl rand 32 > secret.key
```

如果您使用 Windows 但尚未安装 OpenSSL，则可以使用示例：使用 PowerShell 时在 Windows 上生成密钥中描述的示例生成密钥。

示例：使用 PowerShell 在 Windows 上生成密钥

对于 Windows 上可用的终端应用程序 PowerShell，请运行以下命令来创建共享密钥。

```
$bs = New-Object Byte[](32);
[Security.Cryptography.RandomNumberGenerator]::Create().GetBytes($bs); Set-Content 'secret.key' -Encoding Byte -Value $bs
```

步骤 6：将共享密钥存储在 环境变量中。

环境变量是一种方便且可扩展的方式，用户可以从各种密钥存储库中提供密钥，例如 AWS Secrets Manager 并将其传递给 C3R 加密客户端。

AWS 服务 如果您使用将这些密钥存储在相关的环境变量中，则 C3R 加密客户端可以使用中存储的密钥。 AWS CLI 例如，C3R 加密客户端可以使用中的密钥。 AWS Secrets Manager有关更多信息，请参阅《AWS Secrets Manager 用户指南》中的使用 AWS Secrets Manager创建和管理密钥。

> **Note**
>
> 但是，在使用 AWS 服务 诸如 AWS Secrets Manager 来保存 C3R 密钥之前，请验证您的用例是否允许。某些用例可能需要隐瞒 AWS密钥。这是为了确保加密的数据和密钥永远不会由同一个第三方持有。

共享密钥的唯一要求是，共享密钥必须经过 base64 编码并存储在环境变量 `C3R_SHARED_SECRET` 中。

以下各节介绍用于将 `secret.key` 文件转换为 base64 并将其存储为环境变量的控制台命令。`secret.key` 文件可能由 步骤 5：创建共享密钥 中列出的任何命令生成，并且只是一个示例源。

在 Windows 上使用 PowerShell 将密钥存储到环境变量中

要在 Windows 上使用 PowerShell 转换为 base64 并设置环境变量，请运行以下命令。

```
$Bytes=[IO.File]::ReadAllBytes((Get-Location).ToString()+'\secret.key');
 $env:C3R_SHARED_SECRET=[Convert]::ToBase64String($Bytes)
```

在 Linux 或 macOS 上将密钥存储到环境变量中

要在 Linux 或 macOS 上转换为 base64 并设置环境变量，请运行以下命令。

```
export C3R_SHARED_SECRET="$(cat secret.key | base64)"
```

步骤 7：加密数据

要执行此步骤，您必须获取协 AWS Clean Rooms 作 ID 和共享密钥。有关更多信息，请参阅先决条件。

在以下示例中，我们使用我们创建的名为 ads.json 的架构在 ads.csv 上运行加密。

加密数据

1. 将协作的共享密钥存储在 步骤 6：将共享密钥存储在 环境变量中。 中。
2. 在命令行中，输入以下命令。

   ```
   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. 对于 *<name of input .csv file>*，输入输入 .csv 文件的名称。
4. 对于 schema=，输入 .json 加密架构文件的名称。
5. 对于 id=，输入协作 ID。
6. 对于 output=，输入输出文件的名称（例如，ads-output.csv）。
7. 包括 加密计算参数 和 Clean Rooms 加密计算中的可选标志 中描述的任何命令行标志。
8. 运行命令。

在 ads.csv 的示例中，我们运行以下命令。

java -jar c3r-cli.jar encrypt ***ads.csv*** --schema=***ads.json*** --id=***123e4567-e89b-42d3-a456-556642440000*** --output=***ads-output.csv***

在 sales.csv 的示例中，我们运行以下命令。

java -jar c3r-cli.jar encrypt ***sales.csv*** --schema=***sales.json*** --id=***123e4567-e89b-42d3-a456-556642440000***

> **Note**
>
> 在此示例中，我们没有指定输出文件名 (`--output=sales-output.csv`)。结果，生成了默认的输出文件名 `name-of-file.out.csv`。

现在，您可以验证加密的数据了。

步骤 8：验证数据加密

验证数据是否已加密

1. 查看加密的数据文件（例如 `sales-output.csv`）。
2. 验证以下列：
   a. 列 1 — 已加密（例如 `username_fingerprint`）。

      对于 fingerprint 列 (HMAC)，在版本和类型前缀（例如 `01:hmac:`）之后，有 44 个字符的 base64 编码数据。

   b. 列 2 — 未加密（例如 `purchased`）。
   c. 列 3 — 已加密（例如 `product_sealed`）。

      对于已加密 (SELECT) 列，cleartext 的长度加上版本和类型前缀（例如 `01:enc:`）后的任何填充，与加密后的 cleartext 的长度成正比。也就是说，长度等于输入的大小加上大约 33% 的编码开销。

您现在已准备好执行以下操作：

1. 将加密的数据上传到 S3。
2. 创建 AWS Glue 表。
3. 在 AWS Clean Rooms中创建配置表。

C3R 加密客户端将创建不包含未加密数据的临时文件（除非这些数据在最终输出中也未加密）。但是，某些加密值可能无法正确填充。即使协作设置 `allowRepeatedFingerprintValue` 为 `false`，指纹列也可能包含重复的值。之所以出现此问题，是因为临时文件是在检查正确的填充长度和重复项删除属性之前写入的。

如果 C3R 加密客户端失败或在加密过程中中断，则它可能会在写入临时文件之后但在检查这些属性和删除临时文件之前停止。因此，这些临时文件可能仍在磁盘上。在这种情况下，这些文件中的内容对明文数据的保护程度将不如输出文件。特别是，这些临时文件可能会向统计分析揭示明文数据，而这些数据不会对最终输出产生影响。用户应删除这些文件（尤其是 SQLite 数据库），以防止这些文件落入未经授权的人手中。

（可选）创建架构（高级用户）

手动创建架构适用于高级用户。

以下是对带或不带列标题的输入文件的 JSON 架构文件格式的描述。如果需要，高级用户可以直接编写或修改架构。

> **Note**
>
> C3R 加密客户端可通过 示例：生成带有 sealed、fingerprint 和 cleartext 列的加密架构 中描述的交互式流程或通过创建存根模板协助您创建架构。

映射和定位表架构

以下部分描述了两种表架构：

- 映射表架构 — 此架构用于加密带有标题行的 .csv 文件和 Apache Parquet 文件。
- 位置表架构 — 此架构用于加密没有标题行的 .csv 文件。

C3R 加密客户端可以加密表格文件以进行协作。为此，它必须具有相应的架构文件，该文件指定应如何从输入中导出加密输出。

C3R 加密客户端可以通过在命令行运行 C3R 加密客户端架构命令来帮助为 `INPUT` 文件生成架构。命令的一个示例是 `java -jar c3r-cli.jar schema --interactive INPUT`。

架构指定以下信息：

1. 哪些源列通过标题名称（映射架构）或位置（位置架构）映射到输出文件中哪些已转换的列
2. 哪些目标列要保留 cleartext
3. 要对哪些目标列进行加密以进行 SELECT 查询

4. 要对哪些目标列进行加密以进行 JOIN 查询

这些信息在特定于表的 JSON 架构文件中编码，该文件由一个对象组成，其 `headerRow` 字段是一个布尔值。对于有标题行的 Parquet 文件和 .csv 文件，该值必须为 `true`，否则为 `false`。

映射表架构

映射架构具有以下形状。

```
{
  "headerRow": true,
  "columns": [
    {
      "sourceHeader": STRING,
      "targetHeader": STRING,
      "type": TYPE,
      "pad": PAD
    },
    ...
  ]
}
```

如果 `headerRow` 为 `true`，对象中的下一个字段就是 `columns`，其中包含一个将源标题映射到目标标题的列架构（即描述输出列应包含内容的 JSON 对象）数组。

- `sourceHeader` — 数据来源于的源列的 `STRING` 标题名称。

> **Note**
>
> 同一个源列可以用于多个目标列。
>
> 输入文件中未列为架构中任意位置的 `sourceHeader` 列不会出现在输出文件中。

- `targetHeader` — 输出文件中相应列的 `STRING` 标题名称。

> **Note**
>
> 对于映射架构，此字段为可选项。如果省略此字段，输出中的标题名称将重复使用 `sourceHeader`。如果输出列分别为 fingerprint 列或 sealed 列，则附加 `_fingerprint` 或 `_sealed`。

- type — 输出文件中目标列的 TYPE。即 cleartext、sealed 或 fingerprint 其中之一，具体取决于该列在协作中的使用方式。
- pad — 列架构对象的字段，仅当 TYPE 为 sealed 才存在。其对应的 PAD 值是一个对象，用于描述数据在加密前应如何填充。

```
{
  "type": PAD_TYPE,
  "length": INT
}
```

要指定加密前的填充，请按如下方式使用 type 和 length：

- PAD_TYPE 为 none — 不对列的数据进行填充，length 字段不适用（即省略）。
- PAD_TYPE 为 fixed — 将列的数据填充到指定的字节 length。
- PAD_TYPE 为 max — 列的数据填充到最长值的字节长度加上额外的 length 字节。

以下是映射架构的示例，每个类型都有一列。

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

作为一个更复杂的示例，以下是带有标题的 .csv 文件示例。

```
FirstName,LastName,Address,City,State,PhoneNumber,Title,Level,Notes
Jorge,Souza,12345 Mills Rd,Anytown,SC,703-555-1234,CEO,10,
Paulo,Santos,0 Street,Anytown,MD,404-555-111,CIO,9,This is a really long note that
 could really be a paragraph
Mateo,Jackson,1 Two St,Anytown,NY,304-555-1324,COO,9,""
Terry,Whitlock4 N St,Anytown,VA,407-555-8888,EA,7,Secret notes
Diego,Ramirez,9 Hollows Rd,Anytown,VA,407-555-1222,SDE I,4,null
John,Doe,8 Hollows Rd,Anytown,VA,407-555-4321,SDE I,4,Jane's younger brother
Jane,Doe,8 Hollows Rd,Anytown,VA,407-555-4322,SDE II,5,John's older sister
```

在以下映射架构示例中，列 `FirstName` 和 `LastName` 是 `cleartext` 列。`State` 列作为 `fingerprint` 列和 `sealed` 列进行加密，填充为 none。其余列均省略。

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

以下是映射架构生成的 .csv 文件。

```
givenname,surname,state_fingerprint,state
John,Doe,01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:FQ3n3Ahv9BQQNWQGcugeHzHYzEZE1vapHa2Uu4SRgSAtZ3qObjPA4TcsHt
+BOkMKBcnHWI13BeGG/SBqmj7vKpI=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:KZ5n5GtaXACco65AXk48BQO2durDNR2ULc4YxmMC8NaZZKKJiksU1IwFadAvV4iBQ1
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:mLKpS5HIOSgphdEsrzhEd
eN9nBO2gAbIygt4OFn4LalYn9Xyj/XUWXlmn8zFe2T4kyDTD8kGOvpQEUGxAUFk=
Diego,Ramirez,01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:rmZhT98Zm
+IIGw1UTjMIJP4IrW/AAltBLMXcHvnYfRgmWP623VFQ6aUnhsb2MDqEw4G5Uwg5rKKZepUxx5uKbfk=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhNOd7s2ZiKUe7QiTyo8=,01:enc:vVaqWC1VRbhvkf8gnuR7q0z
Terry,Whitlock01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:3c9VEWbODO/
xbQjdGuccLvI7oZTBdPU+SyrJIyr2kudfAxbuMQ2uRdU/q7rbgyJjxZS8M2U35ILJf/lDgTyg7cM=
Jane,Doe,01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:9RWv46YLveykeNZ/
G0NdlYFg+AVdOnu05hHyAYTQkPLHnyX+0/jbzD/g9ZT8GCgVE9aB5bV4ooJIXHGBVMXcjrQ=
```

位置表架构

映射架构具有以下形状。

```
{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": STRING,
        "type": TYPE,
        "pad": PAD
```

```
      },
      {
        "targetHeader": STRING,
        "type": TYPE,
        "pad": PAD
      }
    ],
    [],
    ...
  ]
}
```

如果 headerRow 是 false，则对象中的下一个字段是 columns，其中包含一个条目数组。每个条目本身就是一个由零个或多个位置列架构（无 sourceHeader 字段）组成的数组，这些架构是描述输出应包含内容的 JSON 对象。

- sourceHeader — 数据来源于的源列的 STRING 标题名称。

  > **Note**
  >
  > 在位置架构中必须省略此字段。在位置架构中，源列由架构文件中该列的相应索引推断出来。

- targetHeader — 输出文件中相应列的 STRING 标题名称。

  > **Note**
  >
  > 对于位置架构，此字段为必填字段。

- type — 输出文件中目标列的 TYPE。即 cleartext、sealed 或 fingerprint 其中之一，具体取决于该列在协作中的使用方式。
- pad — 列架构对象的字段，仅当 TYPE 为 sealed 才存在。其对应的 PAD 值是一个对象，用于描述数据在加密前应如何填充。

  ```
  {
    "type": PAD_TYPE,
    "length": INT
  }
  ```

  要指定加密前的填充，请按如下方式使用 type 和 length：

- PAD_TYPE 为 none — 不对列的数据进行填充，length 字段不适用（即省略）。
- PAD_TYPE 为 fixed — 将列的数据填充到指定的字节 length。
- PAD_TYPE 为 max — 列的数据填充到最长值的字节长度加上额外的 length 字节。

> **Note**
>
> 如果您提前知道列数据的字节大小的上限，则 fixed 很有用。如果该列中的任何数据长于指定的 length，则会引发错误。
> 当输入数据的确切大小未知时，max 很方便，因为无论数据的大小如何，它都能正常工作。但是，由于它会对数据进行两次加密，因此 max 需要额外的处理时间。max 在读入临时文件时对数据进行一次加密，在已知列中最长的数据条目之后加密一次。
> 此外，最长值的长度不会在两次调用客户端之间保存。如果您计划分批加密数据或定期加密新数据，请注意生成的加密文字长度可能因批次而异。

以下是位置架构的示例。

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

举一个复杂的示例，以下是一个 .csv 文件示例，前提是它的第一行没有标题。

```
Jorge,Souza,12345 Mills Rd,Anytown,SC, 703 -555 -1234,CEO, 10,
Paulo,Santos, 0 Street,Anytown,MD, 404-555-111,CIO, 9,This is a really long note that
 could really be a paragraph
Mateo,Jackson, 1 Two St,Anytown,NY, 304-555-1324,COO, 9, ""
Terry,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
```

映射架构具有以下形式。

```
{
  "headerRow": false,
  "columns": [
    [
      {
        "targetHeader": "GivenName",
        "type": "cleartext"
      }
    ],
    [
      {
        "targetHeader": "Surname",
        "type": "cleartext"
      }
    ],
    [],
    [],
```

```
    [
      {
        "targetHeader": "State_Join",
        "type": "fingerprint"
      },
      {
        "targetHeader": "State",
        "type": "sealed",
        "pad": {
          "type": "none"
        }
      }
    ],
    [],
    [],
    [],
    []
  ]
}
```

上述架构生成以下输出文件，该文件具有包含指定目标标题的标题行。

```
givenname,surname,state_fingerprint,state
Mateo,Jackson,01:hmac:iIRnjfNBzryusIJ1w35lgNzeY1RQ1bSfq6PDHW8Xrbk=,01:enc:ENS6QD3cMVl9vQEGfe9MN
Q8m/Y5SA89dJwKpT5rGPp8e36h6klwDoslpFzGvU0=
Jorge,Souza,01:hmac:3BxJdXiFFyZ8HBbYNqqEhBVqhNOd7s2ZiKUe7QiTyo8=,01:enc:LKo0zirq2+
+XEIIIMNRjAsGMdyWUDwYaum0B+IFP+rUf1BNeZDJjtFe1Z+zbZfXQWwJy52Rt7HqvAb2WIK1oMmk=
Paulo,Santos,01:hmac:CHF4eIrtTNgAooU9v4h9Qjc
+txBnMidQTjdjWuaDTTA=,01:enc:MyQKyWxJ9kvK1xDQQtXlUNwv3F+yrBRr0xrUY/1BGg5KFgOn9pK+MZ7g
+ZNqZEPcPz4lht1u0t/wbTaqzOCLXFQ=
Jane,Doe,01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:Pd8sbITBfb0/
ttUB4svVsgoYkDfnDvgkvxzeci0Yxq54rLSwccy1o3/B50C3cpkkn56dovCwzgmmPNwrmCmYtb4=
Terry,Whitlock01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv
+1Mk=,01:enc:Qmtzu3B3GAXKh2KkRYTiEAaMopYedsSdF2e/
ADUiBQ9kv2CxKPzWyYTD3ztmKPMka19dHre5VhUHNpO3O+j1AQ8=
Diego,Ramirez,01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:ysdg
+GHKdeZrS/geBIooOEPLHG68MsWpx1dh3xjb+fG5rmFmqUcJLNuuYBHhHAlxchM2WVeV1fmHkBX3mvZNvkc=
John,Doe,01:hmac:UK8s8Cn/WR2JO/To2dTxWD73aDEe2ZUXeSHy3Tv+1Mk=,01:enc:9uX0wZuO7kAPAx
+Hf6uvQownkWqFSKtWS7gQIJSe5aXFquKWCK6yZN0X5Ea2N3bn03Uj1kh0agDWoiP9FRZGJA4=
```

在 AWS Clean Rooms中创建配置表

配置表是对 AWS Glue Data Catalog中现有表的引用。它包含一个分析规则，用于确定如何在 AWS Clean Rooms中查询数据。配置表可以与一个或多个协作关联。有关更多信息 AWS Glue，请参阅 AWS Glue 开发人员指南。

使用提供的统计数据生成 AWS Glue 来计算表的列级统计数据。 AWS Glue Data Catalog 为数据目录中的表 AWS Glue 生成统计数据后，Amazon Redshift Spectrum 会自动使用这些统计数据来优化查询计划。有关使用计算列级统计信息的更多信息 AWS Glue，请参阅《使用列统计信息指南》。

创建配置表

在此步骤中，您将在中创建 AWS Clean Rooms 要在协作中使用的已配置表。

要在中创建已配置的表 AWS Clean Rooms

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 在右上角，选择配置新表。
4. 对于配置新表的选择 AWS Glue 表：
   a. 从下拉列表中选择要配置的数据库。
   b. 从下拉列表中选择要配置的表。

> **Note**
>
> 要验证是否是正确的表，请执行以下任一操作：
>
> - 选择 “在” 中查看 AWS Glue。
> - 打开查看架构以查看架构。

5. 对于协作中允许的列，选择所有列或自定义列表。

如果选择...	操作...
所有列	允许在中使用所有列 AWS Clean Rooms（视分析规则而定）。
自定义列表	从指定允许的列下拉列表中选择要允许的一列或多列。

6. 对于已配置表的详细信息，

   a. 为已配置的表输入名称。

      您可以使用默认名称或重命名此表。

   b. 输入表的描述。

      该描述有助于区分其他具有相似名称的已配置表。

   c. 如果要为已配置的表资源启用标签，请选择添加新标签，然后输入键和值对。

7. 选择配置新表。

后续步骤

现在您已经创建了一个配置表，您已准备好：

- 为配置表配置分析规则
- 将配置表与协作关联

为配置表配置分析规则

以下各节介绍如何为您的配置表配置分析规则。通过定义分析规则，您可以授权可以查询的成员运行与 AWS Clean Rooms 支持的特定分析规则匹配的查询。

AWS Clean Rooms 支持以下分析规则类型：聚合、列表和自定义。

每个配置表只能有一个分析规则。

> **Important**
>
> 如果您在协作中使用 Clean Rooms 加密计算且有加密数据表，则添加到加密配置表的分析规则应与数据的加密方式一致。例如，如果您为 SELECT（聚合分析规则）加密了数据，则不应添加 JOIN（列表分析规则）的分析规则。

要了解 AWS Clean Rooms 中可用的分析规则类型，请参阅 中的分析规则 AWS Clean Rooms。

有关聚合分析规则的更多信息，请参阅聚合分析规则。

有关列表分析规则的更多信息，请参阅列表分析规则。

有关自定义分析规则的更多信息，请参阅中的自定义分析规则 AWS Clean Rooms。

在查看并理解了这些章节之后，您可以执行以下过程：

主题

为表配置聚合分析规则（引导流程）

聚合分析规则允许使用 COUNT、SUM 和 AVG 函数按可选维度聚合统计数据的查询，而不会泄露行级信息。

此过程描述了使用 AWS Clean Rooms 控制台中的引导流程选项为配置表添加聚合分析规则的过程。

为表添加聚合分析规则（引导流程）

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择配置表。
4. 在配置表详细信息页面上，选择配置分析规则。
5. 在步骤 1: 选择类型下，在类型下，保持聚合选项的默认选中状态。
6. 在创建方法下，选择引导流程，然后选择下一步。
7. 在步骤 2: 指定查询控制下，对于聚合函数：

   a. 从下拉列表中选择一个聚合函数：

      - COUNT
      - COUNT DISTINCT
      - SUM
      - SUM DISTINCT
      - AVG

   b. 从列下拉列表中选择哪些列可以用于聚合函数。

   c. （可选）选择添加其他函数以添加另一个聚合函数，并将一个或多个列与该函数相关联。

      > **Note**
      >
      > 至少需要一个聚合函数。

   d. （可选）选择移除以删除聚合函数。
8. 对于联接控制，

   a. 为允许单独查询表选择一个选项：

如果选择...	操作...
否，只能查询重叠	只有在联接到可以查询的成员拥有的表时，才能对表进行查询。

如果选择...	操作...
是	表可以单独查询，也可以在与其他表联接后进行查询。

b. 在指定联接列下，选择要允许在 INNER JOIN 语句中使用的列。

如果您在上一步中选择了是，则这是可选的。

c. 在指定允许的匹配运算符下，选择哪些运算符（如果有）可用于在多个联接列上进行匹配。如果您选择两列或更多 JOIN 列，则需要其中一个运算符。

如果选择...	操作...
AND	您可以在 `INNER JOIN` 匹配条件中包含 `AND`，在表之间将一列联接到另一列。
或者	您可以在 `INNER JOIN` 匹配条件中包含 `OR`，在表之间将一列与另一列进行匹配。此逻辑运算符对于获得更高的匹配率很有用。

9. （可选）对于维度控制，在指定维度列下拉列表中，选择要允许在 SELECT 语句中使用的列，以及查询的 WHERE、GROUP BY 和 ORDER BY 部分。

> **Note**
>
> 聚合函数或联接列不能用作维度列。

10. 对于标量函数，请为要允许哪些标量函数？选择一个选项。

如果选择...	操作...
AWS Clean Rooms 当前支持的全部	您允许 AWS Clean Rooms 当前支持的所有标量函数。 • 您可以选择查看列表以查看 AWS Clean Rooms 中支持的标量函数的完整列表。
自定义列表	您可以自定义允许哪些标量函数。

如果选择...	操作...
	• 从指定允许的标量函数下拉列表中选择一个或多个选项。
无	您不想允许任何标量函数。

有关更多信息，请参阅标量函数。

11. 选择下一步。
12. 在步骤 3: 指定查询结果控制下，为聚合约束：
    a. 选择每个列名称的下拉列表。
    b. 选择应用 COUNT DISTINCT 函数后返回的每个输出行必须满足的每个不同值的最小数量的下拉列表。
    c. 选择添加约束，添加更多聚合约束。
    d. （可选）选择移除以删除聚合约束。
13. 选择下一步。
14. 在步骤 4: 查看并配置下，查看您在之前的步骤中所做的选择，必要时进行编辑，然后选择配置分析规则。

您将看到一条确认消息，指出您成功为表配置了聚合分析规则。

为配置列表分析规则（引导流程）

列表分析规则允许输出关联表与可以查询的成员的表之间重叠的行级列表的查询。

此过程描述了使用 AWS Clean Rooms 控制台中的引导流程选项将列表分析规则添加到配置表中的过程。

为表添加列表分析规则（引导流程）

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择配置表。

4. 在配置表详细信息页面上，选择配置分析规则。
5. 在步骤 1: 选择类型下，在类型下，选择列表选项。
6. 在创建方法下，选择引导流程，然后选择下一步。
7. 在步骤 2: 指定查询控制下，对于联接控制：
   a. 在指定联接列下，选择要允许在 INNER JOIN 语句中使用的列。
   b. 在指定允许的匹配运算符下，选择哪些运算符（如果有）可用于在多个联接列上进行匹配。如果您选择两列或更多 JOIN 列，则需要其中一个运算符。

如果选择...	操作...
AND	您可以在 INNER JOIN 匹配条件中包含 AND，在表之间将一列联接到另一列。
或者	您可以在 INNER JOIN 匹配条件中包含 OR，在表之间将一列与另一列进行匹配。此逻辑运算符对于获得更高的匹配率很有用。

8. （可选）对于列表控制，在指定列表列下拉列表中，选择要允许在查询输出中使用（即在 SELECT 语句中使用）或用于筛选结果（即 WHERE 语句）的列。
9. 选择下一步。
10. 在步骤 3: 查看并配置下，查看您在之前的步骤中所做的选择，必要时进行编辑，然后选择配置分析规则。

您将看到一条确认消息，指出您成功为表配置了列表分析规则。

为表配置自定义分析规则（引导流程）

自定义分析规则允许对配置表进行自定义 SQL 查询。如果使用分析模板或差别隐私，则需要使用自定义分析规则。

此过程描述了使用 AWS Clean Rooms 控制台中的引导流程选项将自定义分析规则添加到配置表中的过程。

为表添加自定义分析规则（引导流程）

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择配置表。
4. 在配置表详细信息页面上，选择配置分析规则。
5. 在步骤 1: 选择类型下，在类型下，选择自定义选项。
6. 在创建方法下，选择引导流程，然后选择下一步。
7. 在步骤 2：设置差别隐私下面，确定是要开启还是关闭差别隐私。差别隐私是一种经过数学验证的技术，可以保护您的数据以免受到重新识别攻击。

   a. 对于差别隐私：

如果您...	则选择...
具有用户级数据，并希望防范重新识别尝试	打开
没有用户级数据，或不需要防范重新识别尝试	关闭

   b. 如果您已选择开启差别隐私，请选择包含要保护隐私的用户的唯一标识符的用户标识符列，例如 `user_id` 列。如果要为协作中的两个或更多表开启差别隐私，您必须在两个分析规则中配置与用户标识符列相同的列，以在表之间保持一致的用户定义。如果未正确进行配置，可以查询的成员将收到一条错误消息，指出具有两列可供选择，以便在运行查询时计算用户贡献数量（例如，用户生成的广告展示次数）。

   c. 选择下一步。

8. 在步骤 3：指定查询控制下面，

   a. 对于控制类型：

如果要...	则选择...
在配置表上运行每个新的分析模板之前，先对其进行审核	在允许在此表上运行每项新分析之前，先对其进行审核

如果要...	则选择...
允许对配置表执行任何分析模板或直接查询	允许特定协作者创建的任何查询无需审核即可在此表上运行

b. 选择以下操作之一：

如果您选择了...	操作...
在允许在此表上运行每项新分析之前，先对其进行审核	在允许运行分析模板下，选择添加分析模板，然后从下拉列表中选择相应的协作和分析模板。
允许特定协作者创建的任何查询无需审核即可在此表上运行	在允许使用 AWS 账户创建任何查询下，选择添加 AWS 账户，然后选择相应的 AWS 账户 ID。

9. 选择下一步。
10. 在步骤 4: 查看并配置下，查看您在之前的步骤中所做的选择，必要时进行编辑，然后选择配置分析规则。

您将看到一条确认消息，指出您成功为表配置了自定义分析规则。

为表配置分析规则（JSON 编辑器）

以下过程说明如何使用 AWS Clean Rooms 控制台中的 JSON 编辑器选项为表添加分析规则。

为表配置聚合、列表或自定义分析规则（JSON 编辑器）

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择配置表。
4. 在配置表详细信息页面上，选择配置分析规则。
5. 在步骤 1: 选择类型下，在类型下，选择聚合、列表或自定义选项。
6. 在创建方法下，选择 JSON 编辑器，然后选择下一步。
7. 在步骤 2: 指定控制下，您可以选择插入查询结构（插入模板）或插入文件（从文件导入）。

如果选择...	操作...
插入模板	1. 在分析规则定义中为所选分析规则指定参数。 2. 您可以按 Ctrl + 空格键启用自动完成。 有关聚合分析规则参数的更多信息，请参阅聚合分析规则 — 查询控制。 有关列表分析规则参数的更多信息，请参阅列表分析规则 — 查询控制。
从文件导入	1. 从本地驱动器中选择您的 JSON 文件。 2. 选择打开。 分析规则定义显示上传文件中的分析规则。

8. 选择下一步。
9. 在步骤 3: 查看并配置下，查看您在之前的步骤中所做的选择，必要时进行编辑，然后选择配置分析规则。

您将收到一条确认消息，指出您成功为表配置了分析规则。

后续步骤

现在，您已经为配置表配置了分析规则，您已准备好：

- 将配置表与协作关联
- 查询数据表（以可以查询的成员身份）

将配置表与协作关联

创建配置表并向其添加分析规则后，可以将其与协作关联。

> **Important**
>
> 在将配置的表与协作关联之前， AWS Glue 表的位置必须指向亚马逊简单存储服务 (Amazon S3) Simp AWS Glue le Service 文件夹，而不是指向单个文件。您可以通过查看 AWS Glue 控制台中的表格来验证此位置，网址为 https://console.aws.amazon.com/glue/。

> **Note**
>
> 如果您已在中配置加密 AWS Glue 并创建了服务角色，则必须向该角色授予访问权限 AWS KMS keys 才能用于解密表 AWS Glue 。
> 如果您关联了由 AWS KMS加密的 Amazon S3 数据集支持的已配置表，则必须向该角色授予访问权限，才能使用 KMS 密钥解密 Amazon S3 数据。
> 有关更多信息，请参阅《AWS Glue 开发人员指南》中的在 AWS Glue中设置加密。

以下主题介绍如何使用 AWS Clean Rooms 控制台将已配置的表格与协作关联起来：

主题

有关如何使用 AWS SDK 将配置表与协作关联的信息，请参阅 AWS Clean Rooms API 参考。

从配置表详细信息页面关联配置表

从已配置的 AWS Glue 表格详细信息页面将表格与协作关联

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。

2. 在左侧导航窗格中，选择配置表。
3. 选择配置表。
4. 在配置表详细信息页面上，选择与协作关联。
5. 在将表与协作关联对话框中，从下拉列表中选择协作。
6. 选择选择协作。

   在关联表页面上，您选择的配置表的名称会出现在选择配置表部分下。
7. 对于选择配置表，执行以下操作：

如果要...	操作...
配置新表	选择配置表，然后按照配置表页面上的提示进行操作。
查看配置表的架构和分析规则。	启用查看架构和分析规则。

8. 通过选择创建并使用新的服务角色或使用现有服务角色来指定服务访问权限。

如果选择...	操作...
创建并使用新的服务角色	• AWS Clean Rooms 使用此表所需的策略创建服务角色。 • 默认服务角色名称为 `cleanrooms-<timestamp>` 。 • 您必须拥有创建角色并附加策略的权限。 • 如果您的输入数据已加密，则可以选择 “此数据是使用 KMS 密钥加密的” AWS KMS key ，然后输入将用于解密您输入的数据的。
使用现有服务角色	1. 从下拉列表中选择一个现有服务角色名称。 如果您有列出角色的权限，则会显示角色列表。 如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。

如果选择...	操作...
	2. 通过选择 “在 IAM 中查看” 外部链接来查看服务角色。 如果没有现有的服务角色，则使用现有服务角色选项不可用。 默认情况下， AWS Clean Rooms 不会尝试更新现有角色策略以添加必要的权限。 3. （可选）选中为该角色添加具有必要权限的预配置策略复选框以向该角色添加必要的附加权限。您必须拥有修改角色并创建策略的权限。

> **Note**
>
> - AWS Clean Rooms 需要权限才能根据分析规则进行查询。有关权限的更多信息 AWS Clean Rooms，请参阅AWS 的托管策略 AWS Clean Rooms。
> - 如果该角色没有足够的权限 AWS Clean Rooms，则会收到一条错误消息，指出该角色没有足够的权限 AWS Clean Rooms。必须先添加角色策略，然后才能继续。
> - 如果您无法修改角色策略，则会收到一条错误消息，指出 AWS Clean Rooms 找不到该服务角色的策略。

9. 如果要为配置表关联资源启用标签，请选择添加新标签，然后输入键和值对。
10. 选择关联表。

从协作详细信息页面关联配置表

从协作详情页面将 AWS Glue 表格与协作关联起来

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。

4. 在表选项卡上，选择关联表。
5. 对于选择配置表，执行以下操作：

如果要...	操作...
选择一个现有的配置表	从下拉列表中选择要与协作关联的配合表名。
配置新表	选择配置表，然后按照配置表页面上的提示进行操作。
查看配置表的架构和分析规则。	启用查看架构和分析规则。

6. 对于表关联详细信息，
   a. 输入关联表的名称。

      您可以使用默认名称或重命名此表。

   b. （可选）输入表的描述。

      该描述有助于编写查询。
7. 通过选择创建并使用新的服务角色或使用现有服务角色来指定服务访问权限。

如果选择...	操作...
创建并使用新的服务角色	• AWS Clean Rooms 使用此表所需的策略创建服务角色。 • 默认服务角色名称为 `cleanrooms-<timestamp>` 。 • 您必须拥有创建角色并附加策略的权限。 • 如果您的输入数据已加密，则可以选择 “此数据是使用 KMS 密钥加密的” AWS KMS key ，然后输入将用于解密您输入的数据的。
使用现有服务角色	1. 从下拉列表中选择一个现有服务角色名称。 如果您有列出角色的权限，则会显示角色列表。

如果选择...	操作...
	如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。 2. 通过选择 “在 IAM 中查看” 外部链接来查看服务角色。 如果没有现有的服务角色，则使用现有服务角色选项不可用。 默认情况下， AWS Clean Rooms 不会尝试更新现有角色策略以添加必要的权限。 3. （可选）选中为该角色添加具有必要权限的预配置策略复选框以向该角色添加必要的附加权限。您必须拥有修改角色并创建策略的权限。

> **Note**
>
> - AWS Clean Rooms 需要权限才能根据分析规则进行查询。有关权限的更多信息 AWS Clean Rooms，请参阅AWS 的托管策略 AWS Clean Rooms。
> - 如果该角色没有足够的权限 AWS Clean Rooms，则会收到一条错误消息，指出该角色没有足够的权限 AWS Clean Rooms。必须先添加角色策略，然后才能继续。
> - 如果您无法修改角色策略，则会收到一条错误消息，指出 AWS Clean Rooms 找不到该服务角色的策略。

8. 如果要为配置表关联资源启用标签，请选择添加新标签，然后输入键和值对。
9. 选择关联表。

后续步骤

现在，您已将配置数据表与协作关联，您已准备好：

- 编辑协作（如果您是协作创建者）
- 查询数据表（以可以查询的成员身份）

配置差别隐私策略

此过程描述了使用 AWS Clean Rooms 控制台中的 “引导流程” 选项在协作中配置差异隐私策略的过程。对于所有具有差别隐私保护的表来说，这是一次性步骤。

配置差别隐私设置（引导流程）

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在协作页面的表选项卡上，选择配置差别隐私策略。
5. 在配置差别隐私策略页面上，选择以下属性的值：

   - 隐私预算
   - 每月刷新隐私预算
   - 每个查询添加的噪声

   您可以使用默认值，或输入支持您的特定使用案例的自定义值。在选择隐私预算和每个查询添加的噪声值后，您可以根据数据的所有查询中可能进行的聚合数量预览产生的效用。
6. 选择 配置。

您将看到一条确认消息，指出您成功为协作配置了差别隐私策略。

后续步骤

您现在配置了差别隐私，您已准备好：

- 查询数据表（以可以查询的成员身份）
- 管理协作（如果您是协作创建者）

使用分析模板

分析模板可与 中的自定义分析规则 AWS Clean Rooms 配合使用。使用分析模板，您可以定义参数来帮助您重复使用相同的查询。 AWS Clean Rooms 支持带有字面值的参数化子集。

分析模板针对协作。对于每个协作，成员只能看到该协作中的查询。如果您计划在协作中使用差别隐私，应确保您的分析模板与 AWS Clean Rooms Differential Privacy 的通用查询结构兼容。

主题

创建分析模板

有关如何使用 AWS 软件开发工具包创建分析模板的信息，请参阅 AWS Clean Rooms API 参考。

使用 AWS Clean Rooms 控制台创建分析模板

1. 登录 AWS Management Console 并打开AWS Clean Rooms 控制台，该控制台将充当协作创建者。 AWS 账户
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在模板选项卡上，转到您创建的分析模板部分。
5. 选择创建分析模板。
6. 在创建分析模板页面上，在详细信息中，输入名称和可选描述。
7. 对于表，查看与协作关联的配置表。
8. 对于定义，

    a. 输入分析模板的定义。

    b. 选择导入自以导入定义。

    c. （可选）在 SQL 编辑器中通过在参数名称前输入冒号 (:) 来指定参数。

    例如：

```
    WHERE table1.date + :date_period > table1.date
```

9. 如果您之前添加了参数，请在参数 - 可选下，为每个参数名称选择类型和默认值（可选）。
10. 如果要为配置表资源启用标签，请选择添加新标签，然后输入键和值对。
11. 选择创建。

您现在已准备好执行以下操作：

- 告知您的协作成员他们可以审核分析模板。（如果您想查询自己的数据，则是可选的。）

审核分析模板

协作成员创建分析模板后，您可以对其进行审核和批准。在分析模板获得批准后，它可以在查询中进行查询 AWS Clean Rooms。

使用 AWS Clean Rooms 控制台查看分析模板

1. 登录 AWS Management Console 并打开AWS Clean Rooms 控制台，该控制台将充当协作创建者。 AWS 账户
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在模板选项卡上，转到其他成员创建的分析模板部分。
5. 选择 “可以运行” 状态为 “否” 的分析模板需要您审核。
6. 选择审核。
7. 审核分析规则概述、定义和参数（如果有）。
8. 审核定义中引用的表格下列出的已配置表。

   每个表旁边的状态将显示为不允许使用模板。
9. 选择一个 表。

如果您	则选择...
批准分析模板	桌子上的模板。通过选择来确认您的批准。
不批准分析模板	不允许

现在，您已准备好使用分析模板来查询数据表（作为可以查询的成员）。

使用分析模板查询已配置的表

此过程演示如何使用 AWS Clean Rooms 控制台中的分析模板通过自定义分析规则查询已配置的表。

使用分析模板通过自定义分析规则查询配置表

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员能力状态为查询的协作。
4. 在查询选项卡的表下，查看表及其关联的分析规则类型（自定义分析规则）。

   > **Note**
   >
   > 如果您没有在列表中看到所期望的表，可能是由于以下原因：
   >
   > - 这些表尚未关联。
   > - 这些表没有配置分析规则。

5. 在分析部分下，从下拉列表中选择分析模板。
6. 输入要在查询中使用的分析模板中的参数值。该值必须是参数的指定数据类型。每次运行分析模板时都可以使用不同的值。不支持该参数的空NULL值或值。也不支持在LIMIT子句中使用参数。
7. 选择运行。

   > **Note**
   >
   > 如果可以接收结果的成员尚未配置查询结果设置，您将无法运行查询。

8. 继续调整参数并再次运行查询，或者选择 + 按钮在新选项卡中开始新查询。

在协作中查询数据

作为可以查询的成员，您可以执行以下操作之一：

- 使用 SQL 代码编辑器手动构建 SQL 查询。
- 使用分析构建器用户界面无需编写 SQL 代码即可生成查询。
- 使用经批准的分析模板。

当可以查询的成员对协作中的表运行 SQL 查询时， AWS Clean Rooms 将扮演相关角色来代表他们访问这些表。 AWS Clean Rooms 根据需要将分析规则应用于输入查询及其输出。

AWS Clean Rooms 支持可能与其他查询引擎不同的 SQL 查询。有关规范，请参阅 AWS Clean Rooms SQL 参考。如果要对受差别隐私保护的数据表运行查询，您应该确保查询与 AWS Clean Rooms Differential Privacy 的通用查询结构兼容。

> **Note**
>
> 使用 Clean Rooms 加密计算时，并非所有 SQL 操作都会生成有效的结果。例如，您可以对加密列执行 COUNT，但是对加密的数字执行 SUM 会导致错误。此外，查询还可能产生错误的结果。例如，SUM 密封列的查询会产生错误。但是，对密封列的 GROUP BY 查询似乎成功了，但生成的组与通过对 cleartext 的 GROUP BY 查询生成的组不同。

以下主题介绍如何使用 AWS Clean Rooms 控制台在协作中查询数据。

主题

- 使用 SQL 代码编辑器
- 使用分析构建器
- 查询具有差别隐私的数据
- 查看最近的查询
- 查看查询详细信息

有关如何通过直接调用 AWS Clean Rooms `StartProtectedQuery` API 操作或使用 AWS 软件开发工具包来查询数据或查看查询的信息，请参阅 AWS Clean Rooms API 参考。

有关查询日志记录的信息，请参阅查询登录 AWS Clean Rooms。

> **Note**
>
> 如果您对加密数据表运行查询，则加密列的结果将被加密。

有关接收查询结果的信息，请参阅接收查询结果。

使用 SQL 代码编辑器

作为可以查询的成员，您可以通过在 SQL 代码编辑器中编写 SQL 代码来手动生成查询。SQL 代码编辑器位于 AWS Clean Rooms 控制台中 “查询” 选项卡的 “分析” 部分。

默认情况下显示 SQL 代码编辑器。如果要使用分析构建器来生成查询，请参阅使用分析构建器。

> **Important**
>
> 如果您开始在代码编辑器中编写 SQL 查询，然后打开分析构建器用户界面，则不会保存您的查询。

AWS Clean Rooms 支持许多 SQL 命令、函数和条件。有关更多信息，请参阅 AWS Clean Rooms SQL 参考。

> **Tip**
>
> 如果查询运行时发生计划的维护，查询会终止并回滚。必须重新开始查询。

使用 SQL 代码编辑器手动构建 SQL 查询。

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员能力状态为查询的协作。
4. 在查询选项卡上，转到分析部分。

> **Note**
>
> 只有在可以接收结果的成员和负责支付查询计算费用的成员作为活跃成员加入协作时，才会显示分析部分。

5. 在查询选项卡的表下，查看表列表及其关联的分析规则类型（聚合分析规则、列表分析规则或自定义分析规则）。

> **Note**
>
> 如果您没有在列表中看到所期望的表，可能是由于以下原因：
>
> - 这些表尚未关联。
> - 这些表没有配置分析规则。

6. （可选）要查看表的架构和分析规则控制，请选择加号图标 (+) 展开表。
7. 通过在 SQL 代码编辑器中键入查询来构建查询。

（可选）如果要使用示例查询	（可选）如果要插入列名或函数
1. 选择表名称旁边的三个垂直点。 2. 在在编辑器中插入下，选择查询示例。 **Note** 插入查询示例会追加编辑器中已有的查询。 此时将显示查询示例。表下列出的所有表都包含在查询中。 3. 编辑查询中的占位符值。	1. 选择列旁边的三个垂直点。 2. 在在编辑器中插入下，选择列名。 3. 要手动插入列上允许的函数，请选择列旁边的三个垂直点，选择在编辑器中插入，然后选择允许的函数的名称（例如 INNER JOIN、SUM、SUM DISTINCT 或 COUNT）。 4. 按 Ctrl + 空格键可在代码编辑器中查看表架构。 **Note** 可以查询的成员可以查看和使用每个配置表关联中

（可选）如果要使用示例查询	（可选）如果要插入列名或函数
	的分区列。确保将分区列标记为已配置 AWS Glue 表下方的表中的分区列。 5. 编辑查询中的占位符值。

8. 选择运行。

> **Note**
>
> 如果可以接收结果的成员尚未配置查询结果设置，您将无法运行查询。

9. 继续调整参数并再次运行查询，或者选择 + 按钮在新选项卡中开始新查询。

> **Note**
>
> AWS Clean Rooms 旨在提供清晰的错误消息。如果错误消息中没有足够的详细信息来帮助您进行故障排除，请联系客户团队。向他们说明错误情况和错误信息（包括任何标识符）。有关更多信息，请参阅 故障排除 AWS Clean Rooms。

使用分析构建器

您无需编写 SQL 代码即可使用分析构建器来构建查询。使用分析构建器，您可以为具有以下特征的协作构建查询：

- 单个使用聚合分析规则且不需要 JOIN 的表
- 两个使用聚合分析规则的表（每个成员一个）
- 两个使用列表分析规则的表（每个成员一个）
- 两个使用聚合分析规则的表（每个成员一个）和两个使用列表分析规则的表（每个成员一个）

如果要手动编写 SQL 查询，请参阅使用 SQL 代码编辑器。

分析构建器在 AWS Clean Rooms 控制台的查询选项卡的分析部分中显示为分析构建器用户界面选项。

> **Important**
>
> 如果您打开分析构建器用户界面，开始在分析构建器中构建查询，然后关闭分析构建器用户界面，则不会保存您的查询。

> **Tip**
>
> 如果查询运行时发生计划的维护，查询会终止并回滚。必须重新开始查询。

以下主题介绍分析构建器的使用。

主题

使用分析构建器查询单个表（聚合）

此过程演示如何使用 AWS Clean Rooms 控制台中的 Analys is Builder 用户界面来生成查询。该查询适用于具有单个表的协作，该表使用聚合分析规则且无需 JOIN。

使用分析构建器查询单个表

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员能力状态为查询的协作。
4. 在查询选项卡的表下，查看表及其关联的分析规则类型。（分析规则类型应为聚合分析规则。）

> **Note**
>
> 如果您没有看到所期望的表，可能是由于以下原因：
>
> - 该表尚未关联。
> - 该表没有配置分析规则。

5. 在分析部分下，打开分析构建器用户界面。
6. 构建查询。

   如果要查看所有聚合指标，请跳至步骤 9。

   a. 对于选择指标，请查看默认情况下预先选择的聚合指标，并在需要时删除任何指标。

   b. （可选）对于添加分段 - 可选，请选择一个或多个参数。

   > **Note**
   >
   > 只有在为表指定维度时才会显示添加分段 - 可选。

   c. （可选）对于添加筛选条件 - 可选，请选择添加筛选条件，然后选择参数、运算符和值。

   要添加更多筛选条件，请选择再添加一个筛选条件。

   要删除筛选条件，请选择移除。

   > **Note**
   >
   > ORDER BY 不支持聚合查询。
   > 筛选条件仅支持 AND 运算符。

   d. （可选）对于添加描述 - 可选，请输入描述以帮助识别查询列表中的查询。
7. 展开预览 SQL 代码。

   a. 查看分析构建器生成的 SQL 代码。

   b. 要复制 SQL 代码，请选择复制。

   c. 要编辑 SQL 代码，请选择在 SQL 代码编辑器中编辑。
8. 选择运行。

   > **Note**
   >
   > 如果可以接收结果的成员尚未配置查询结果设置，您将无法运行查询。
9. 继续调整参数并再次运行查询，或者选择 + 按钮在新选项卡中开始新查询。

> **Note**
>
> AWS Clean Rooms 旨在提供清晰的错误消息。如果错误消息中没有足够的详细信息来帮助您进行故障排除，请联系客户团队。向他们说明错误情况和错误信息（包括任何标识符）。有关更多信息，请参阅 故障排除 AWS Clean Rooms。

使用分析构建器查询两个表（聚合或列表）

此过程介绍如何使用 AWS Clean Rooms 控制台中的分析生成器为具有以下特征的协作生成查询：

- 两个使用聚合分析规则的表（每个成员一个）
- 两个使用列表分析规则的表（每个成员一个）
- 两个使用聚合分析规则的表（每个成员一个）和两个使用列表分析规则的表（每个成员一个）

使用分析构建器查询两个表

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员能力状态为查询的协作。
4. 在查询选项卡的表下，查看两个表及其关联的分析规则类型（聚合分析规则或列表分析规则）。

   > **Note**
   >
   > 如果您没有在列表中看到所期望的表，可能是由于以下原因：
   >
   > - 这些表尚未关联。
   > - 这些表没有配置分析规则。

5. 在分析部分下，打开分析构建器用户界面。
6. 构建查询。

   如果协作包含两个使用聚合分析规则的表和两个使用列表分析规则的表，请先选择聚合或列表，然后根据所选分析规则按照提示进行操作。

如果两个表使用聚合分析规则	如果两个表使用列表分析规则
1. 对于选择指标，请查看默认情况下预先选择的聚合指标，并在需要时删除任何指标。 2. 对于匹配记录，请选择一个或多个记录。 **Note** 使用分析构建器时，只能对一对列进行匹配。 3. （可选）对于添加分段 - 可选，请选择一个或多个参数。 **Note** 只有在为表指定维度时才会显示添加分段 - 可选。 4. （可选）对于添加筛选条件 - 可选，请选择添加筛选条件，然后选择参数、运算符和值。 要添加更多筛选条件，请选择再添加一个筛选条件。 要删除筛选条件，请选择移除。 **Note** ORDER BY 不支持聚合查询。	1. 对于选择属性，请查看默认情况下预先选择的列表属性，并在需要时删除任何指标。 2. 对于匹配记录，请选择一个或多个记录。 **Note** 使用分析构建器时，只能对一对列进行匹配。 3. （可选）对于添加筛选条件 - 可选，请选择添加筛选条件，然后选择参数、运算符和值。 要添加更多筛选条件，请选择再添加一个筛选条件。 要删除筛选条件，请选择删除。 **Note** LIMIT 不支持列表查询。 筛选条件仅支持 AND 运算符。 4. （可选）对于添加描述 - 可选，请输入描述以帮助识别最近查询列表中的查询。

如果两个表使用聚合分析规则	如果两个表使用列表分析规则
筛选条件仅支持 AND 运算符。 5. （可选）对于添加描述 - 可选，请输入描述以帮助识别最近查询列表中的查询。	

7. 展开预览 SQL 代码。
   a. 查看分析构建器生成的 SQL 代码。
   b. 要复制 SQL 代码，请选择复制。
   c. 要编辑 SQL 代码，请选择在 SQL 代码编辑器中编辑。
8. 选择运行。

> **Note**
>
> 如果可以接收结果的成员尚未配置查询结果设置，您将无法运行查询。

9. 继续调整参数并再次运行查询，或者选择 + 按钮在新选项卡中开始新查询。

> **Note**
>
> AWS Clean Rooms 旨在提供清晰的错误消息。如果错误消息中没有足够的详细信息来帮助您进行故障排除，请联系客户团队。向他们说明错误情况和错误信息（包括任何标识符）。有关更多信息，请参阅 故障排除 AWS Clean Rooms。

查询具有差别隐私的数据

一般来说，在开启差别隐私后，查询编写和运行不会发生变化。不过，如果剩余的隐私预算不足，则无法运行查询。在您运行查询并使用隐私预算时，您可以大致了解可以运行的聚合数量以及这可能会如何影响将来的查询。

查看差别隐私在协作中的影响

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员详细信息状态为运行查询的协作。
4. 在查询选项卡上的表下面，查看剩余的隐私预算。这显示为估计的剩余聚合函数数量和使用的效用（显示为百分比）。

   > **Note**
   >
   > 仅为可以查询的成员显示估计的剩余聚合函数数量和使用的效用百分比。

5. 选择查看影响以查看在结果中注入了多少噪声以及您大约可以运行多少个聚合函数。

查看最近的查询

您可以在最近的查询选项卡上查看过去 90 天内运行的查询。

> **Note**
>
> 如果您唯一的成员能力是贡献数据，并且您不是为查询计算费用付费的成员，则控制台上不会显示查询选项卡。

查看最近的查询

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在查询选项卡的查询下，查看过去 90 天内运行的查询。
5. 要按状态对最近的查询进行排序，请从所有状态下拉列表中选择一个状态。

   状态为：已提交、已开始、已取消、成功、失败和超时。

查看查询详细信息

您可以以能够运行查询的成员或能够接收结果的成员的身份查看查询的详细信息。

查看查询的详细信息

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （ 如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在查询选项卡上，执行以下操作之一：
   - 选择要查看的特定查询对应的选项按钮，然后选择查看详细信息。
   - 选择受保护的查询 ID。
5. 在查询详细信息页面上，
   - 如果您是可以运行查询的成员，请查看查询详细信息、SQL 文本和结果。

     您会看到一条消息，确认查询结果已发送给可以接收结果的成员。
   - 如果您是可以接收结果的成员，请查看查询详细信息和结果。

接收查询结果

作为可以接收结果的成员，您可以将 AWS Clean Rooms 的查询输出接收到您加入协作时指定的 Amazon S3 存储桶中。

以下主题介绍如何使用 AWS Clean Rooms 控制台接收查询结果。

主题

- 接收查询结果
- 编辑查询结果设置的默认值
- 在其他 AWS 服务 中使用查询输出

有关如何通过直接调用 AWS Clean Rooms API 操作或使用 AWS SDK 来查询数据或查看查询的信息，请参阅 AWS Clean Rooms API 参考。

有关查询日志记录的信息，请参阅查询登录 AWS Clean Rooms。

> **Note**
>
> 如果您对加密数据表运行查询，则加密列的结果将被加密。

接收查询结果

查询结果位于 AWS Clean Rooms 控制台中查询选项卡的查询结果设置默认值部分和查询部分。

接收查询结果

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员能力状态为接收结果的协作。
4. 要直接从 AWS Clean Rooms 中接收查询结果，请在查询选项卡的查询下的受保护的查询 ID 列下，选择查询。
5. 在查询详细信息页面的结果下，执行以下任一操作：

如果要...	则选择...
复制结果。	复制
下载结果。	下载 **Note** 默认情况下，下载的文件的名称是在 AWS Clean Rooms 中运行查询时显示的相应 `Query id`。
在 Amazon S3 中查看结果。	在 Amazon S3 中查看 这将在单独的选项卡中打开 Amazon S3 控制台。

6. 如果您使用的是加密数据，则现在可以解密数据表。

   有关更多信息，请参阅使用 C3R 加密客户端解密数据表。

编辑查询结果设置的默认值

作为可以接收结果的成员，您可以在 AWS Clean Rooms 控制台中编辑查询结果设置的默认值。

编辑查询结果设置的默认值

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您的成员能力状态为接收结果的协作。
4. 在查询选项卡的查询结果设置下，选择编辑。
5. 在编辑查询结果设置默认值页面上，根据需要修改以下任意内容：

   a. 在查询结果设置下，修改 Amazon S3 中的结果目标或结果格式。

b. 在服务访问下，修改授权 AWS Clean Rooms 的方法，以写入指定的 Amazon S3 存储桶和格式。

更新后的查询结果设置显示在协作详细信息页面上。

在其他 AWS 服务 中使用查询输出

AWS Clean Rooms 的查询输出可在控制台上找到（如果使用控制台运行查询），并下载到指定的 Amazon S3 存储桶中。然后，您可以在其他 AWS 服务 中使用查询输出，例如 Amazon QuickSight 和 Amazon SageMaker，具体取决于这些服务如何使用来自 Amazon S3 的数据。

有关 Amazon QuickSight 的更多信息，请参阅 Amazon QuickSight 文档。

有关 Amazon SageMaker 的更多信息，请参阅 Amazon SageMaker 文档。

使用 C3R 加密客户端解密数据表

对于使用 Clean Rooms 加密计算和 C3R 加密客户端加密数据表的协作，请按照以下过程操作。在协作中查询数据后，请使用此过程。

此过程需要共享密钥和协作 ID。

能够接收结果的成员使用用于加密协作数据的共享密钥和协作 ID 来解密数据。

> **Note**
>
> AWS Clean Rooms 协作已经限制了谁可以执行和查看查询结果。要执行解密，任何有权访问这些结果的人都需要使用与加密数据相同的共享密钥和协作 ID。

解密已加密的数据表

1. （可选）在 C3R 加密客户端中查看可用命令。
2. （可选）导航到所需的目录并运行 `ls` (macOS) 或 `dir` (Windows)。
   - 确认 c3r-cli.jar 文件和加密的查询结果数据文件位于所需目录中。

     > **Note**
     >
     > 如果查询结果是从 AWS Clean Rooms 控制台界面下载的，则可能位于您的用户账户的下载文件夹中。（例如，在 Windows 和 macOS 上，则位于您的用户目录中的下载文件夹中。）我们建议您将查询结果文件移到与 c3r-cli.jar 相同的文件夹。
3. 将共享密钥存储在 `C3R_SHARED_SECRET` 环境变量中。有关更多信息，请参阅步骤 6：将共享密钥存储在 环境变量中。。
4. 从 AWS Command Line Interface (AWS CLI) 运行以下命令。

   java -jar c3r-cli.jar decrypt ***<name of input .csv file>*** --id=***<collaboration id>*** --output=***<output file name>***
5. 将每个`#######`替换为您自己的信息：
   a. 对于 `id=`，输入协作 ID。
   b. 对于 `output=`，输入输出文件的名称（例如，`results-decrypted.csv`）。

如果不指定输出名称，则终端中会显示默认名称。

c. 使用首选 CSV 或 Parquet 查看应用程序（例如 MicrosoftExcel、文本编辑器或其他应用程序）查看指定输出文件中的解密数据。

管理 AWS Clean Rooms

以下主题介绍如何 AWS Clean Rooms 使用 AWS Clean Rooms 控制台管理协作、成员和已配置的表。

有关如何 AWS Clean Rooms 使用 AWS 软件开发工具包进行管理的信息，请参阅 AWS Clean Rooms API 参考。

主题

- 在 AWS Clean Rooms 中管理协作
- 管理中配置的表 AWS Clean Rooms

在 AWS Clean Rooms 中管理协作

以下主题介绍协作创建者如何使用 AWS Clean Rooms 控制台在 AWS Clean Rooms 中管理协作。

有关如何使用 AWS SDK 管理协作的信息，请参阅 AWS Clean RoomsAPI 参考。

主题

- 编辑协作
- 删除协作
- 查看协作
- 查看表格和分析规则
- 查看差别隐私使用情况日志
- 监控成员状态
- 从协作中删除成员
- 退出协作
- 编辑配置表关联
- 取消关联已配置的表
- 编辑差别隐私策略
- 删除差别隐私策略
- 查看计算的差别隐私参数

编辑协作

了解如何编辑协作的不同部分。

主题

编辑协作名称和描述

创建协作后，您只能编辑协作名称和描述。

> **Note**
>
> 如果您启用了查询日志记录，则可以编辑是否将查询日志存储在您的 Amazon CloudWatch Logs 账户中。

编辑协作名称和描述

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 在协作详细信息页面上，选择操作，然后选择编辑协作。
5. 在详细信息中，编辑协作的名称和描述。
6. 选择保存更改。

编辑协作标签

作为协作创建者，在创建协作后，您可以管理协作资源上的标签。

编辑协作标签

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 选择以下操作之一：

如果您是...	操作...
协作的成员	选择详细信息选项卡。
协作创建者但不是协作的成员	在页面中向下滚动到标签部分。

5. 有关协作详细信息，请选择管理标签。
6. 在管理标签页面上，可以执行以下操作：
   - 要删除标签，请选择移除。
   - 要添加标签，请选择添加新标签。
   - 要保存您的更改，请选择保存更改。

编辑成员资身份标签

作为协作创建者，在创建协作后，您可以管理成员身份资源上的标签。

编辑成员身份标签

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 选择详细信息选项卡。
5. 对于成员身份详细信息，选择管理标签。
6. 在管理成员身份标签页面上，可以执行以下操作：
   - 要删除标签，请选择移除。

- 要添加标签，请选择添加新标签。
- 要保存您的更改，请选择保存更改。

编辑关联表标签

作为协作创建者，在将表与一个协作关联后，您可以管理关联的表资源上的标签。

编辑关联表标签

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 选择 Tables (表) 选项卡。
5. 对于由您关联的表，请选择一个表。
6. 在已配置表的详细信息页面上，对于标签，选择管理标签。

   在管理标签页面上，可以执行以下操作：

   - 要删除标签，请选择移除。
   - 要添加标签，请选择添加新标签。
   - 要保存您的更改，请选择保存更改。

编辑分析模板标签

作为协作创建者，在创建协作后，您可以管理分析模板资源上的标签。

编辑成员身份标签

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 选择 Templates 选项卡。
5. 在您创建的分析模板部分，选择分析模板。

6. 在分析模板表详细信息页面上，向下滚动到标签部分。
7. 选择管理标签。
8. 在管理标签页面上，可以执行以下操作：
   - 要删除标签，请选择移除。
   - 要添加标签，请选择添加新标签。
   - 要保存您的更改，请选择保存更改。

编辑差别隐私策略标签

作为协作创建者，在创建协作后，您可以管理分析模板资源上的标签。

编辑成员身份标签

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择包含您要编辑的差别隐私策略的协作。
4. 选择 Tables (表) 选项卡。
5. 在表选项卡上，选择管理标签。
6. 在管理标签页面上，可以执行以下操作：
   - 要删除标签，请选择移除。
   - 要添加标签，请选择添加新标签。
   - 要保存您的更改，请选择保存更改。

删除协作

作为协作创建者，您可以删除您创建的协作。

> **Note**
>
> 在删除协作时，您和所有成员无法运行查询，接收结果或贡献数据。每个协作成员根据其成员身份继续访问自己的数据。

删除协作

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择要删除的协作。
4. 在操作下，选择删除协作。
5. 确认删除，然后选择删除。

查看协作

作为协作创建者，您可以查看自己创建的所有协作。

查看协作

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 在协作页面的最后使用下，查看最近使用的 5 个协作。
4. 在具有活跃成员身份选项卡上，查看拥有活跃成员身份的协作列表。

   您可以按名称、成员身份创建日期和您的成员详细信息进行排序。

   您可以使用搜索栏搜索协作。
5. 在可供加入选项卡上，查看可供加入的协作列表。
6. 在不再可用选项卡上，查看已删除的协作列表和不再可用的协作成员身份（已删除的成员身份）。

查看表格和分析规则

查看与协作和分析规则关联的表

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。

4. 选择 Tables (表) 选项卡。
5. 选择以下操作之一：
   a. 要查看协作中关联的表，请针对由您关联的表，选择一个表（蓝色文本）。
   b. 要查看协作中关联的其他表，请针对由协作者关联的表，选择一个表（蓝色文本）。
6. 在表详细信息页面查看表的详细信息和分析规则。

查看差别隐私使用情况日志

作为使用差别隐私保护数据的协作成员，在创建具有差别隐私的协作后，您可以监控隐私预算的使用情况。

查看运行了多少聚合以及使用了多少隐私预算

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 选择 Tables (表) 选项卡。
5. 选择查看使用情况日志（蓝色文本）。
6. 查看使用情况详细信息，包括隐私预算和提供了多少效用。

监控成员状态

作为协作创建者，创建协作后，您可以在成员选项卡上监控所有成员的状态。

检查成员的状态

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 选择成员选项卡。
5. 查看每个成员的成员状态。

从协作中删除成员

> **Note**
>
> 如果删除成员，还会从协作中删除成员的所有关联数据集。

从协作中删除成员

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择您创建的协作。
4. 选择成员选项卡。
5. 选择要删除的成员旁边的选项按钮。

   > **Note**
   >
   > 协作创建者无法选择自己的账户 ID。

6. 选择移除。
7. 在对话框中，通过在文本输入字段中键入 **`confirm`** 来确认删除成员的决定。

   > **Note**
   >
   > 如果您删除支付查询计算费用的成员，则不允许在协作中运行其他查询。

退出协作

作为协作成员，您可以通过删除成员身份来退出协作。如果您是协作创建者，则只能通过删除协作来退出协作。

> **Note**
>
> 当您删除成员身份时，您将退出协作且无法重新加入。如果您是支付查询计算费用的成员并且删除了您的成员身份，则不允许运行其他查询。

退出协作

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 对于具有活跃成员身份，请选择您所属的协作。
4. 选择操作。
5. 选择删除成员身份。
6. 在对话框中，通过在文本输入字段中键入 **confirm** 来确认退出协作的决定，然后选择清空并删除成员身份。

   您会在控制台上看到一条消息，指出成员身份已删除。

   协作创建者将看到成员状态为退出。

编辑配置表关联

作为协作成员，您可以编辑已创建的已配置表关联。

编辑配置表关联

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 选择表选项卡。
5. 对于由您关联的表，请选择一个表。
6. 在表详细信息页面上，向下滚动以查看表关联详细信息。
7. 选择编辑。
8. 在编辑已配置的表关联页面上，更新描述或服务访问信息。
9. 选择保存更改。

取消关联已配置的表

作为协作成员，您可以取消已配置的表与协作的关联。此操作可阻止可以查询的成员查询表。

取消关联配置表

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 选择表选项卡。
5. 对于由您关联的表，选择要取消关联的表旁边的选项按钮。
6. 选择取消关联。
7. 在对话框中，确认取消关联配置表的决定，并通过选择取消关联来阻止可以查询的成员查询该表。

编辑差别隐私策略

在配置差别隐私策略后，您可以随时更新该策略以更好地反映您的隐私需求。

编辑差别隐私策略

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在协作页面的表选项卡上，在由您关联的表下面选择编辑。
5. 在编辑差别隐私页面上，为以下属性选择新的值：
   - 隐私预算 - 移动滑块以在协作期间随时增加或减少预算。在可以查询的成员开始查询您的数据后，您无法减少预算。如果增加隐私预算，AWS Clean Rooms 将继续使用现有预算，直到用完现有预算，然后再使用新添加的隐私预算。
   - 每个查询添加的噪声 - 移动滑块以在协作期间随时增加或减少每个查询添加的噪声。

> **Note**
>
> 您可以选择交互式示例以了解隐私预算和每个查询添加的噪声的不同值如何影响您可以运行的聚合函数数量。

您无法更改隐私预算刷新的值。要更改您选择的值，您必须删除差别隐私策略并创建一个新策略。

6. 选择保存更改。

您会看到一条确认消息，指出您已成功编辑差别隐私策略。

删除差别隐私策略

您可以从协作的表选项卡中删除差别隐私策略。

删除差别隐私策略

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在协作页面的表选项卡上的差别隐私策略旁边，选择删除。
5. 如果您确定要删除差别隐私策略，请选择删除。

在删除差别隐私策略后，您无法访问该策略的隐私预算使用情况日志。如果删除了差别隐私策略，将无法查询开启了差别隐私的表。

查看计算的差别隐私参数

对于具有差别隐私专业知识的用户，您可以从协作的查询选项卡中查看计算的差别隐私参数。

查看计算的差别隐私参数

1. 登录 AWS Management Console并使用您的 AWS 账户打开 AWS Clean Rooms 控制台（如果您尚未这样做）。
2. 在左侧导航窗格中，选择协作。
3. 选择协作。
4. 在查询选项卡的结果部分中，选择查看计算得出的差别隐私参数。

在计算得出的差别隐私参数表中，您可以看到聚合函数的灵敏度值，该值定义为在添加、删除或修改单个用户的记录时函数结果可能发生的最大变化量。该列表包括以下差别隐私参数：

- 用户贡献限制 (UCL) 是用户在 SQL 查询中贡献的最大行数。例如，如果要计算指定活动中的总匹配展示次数，其中每个用户可以具有多次展示，AWS Clean Rooms Differential Privacy 需要限制单个用户的展示次数，以确保差别隐私计算是准确的。换句话说，如果任何用户的展示次数超过限制，则 AWS Clean Rooms 根据计算的 UCL 值自动对该用户的展示进行统一随机采样，并在执行查询时排除该用户的剩余展示。如果您计算唯一用户数，则 UCL 值等于 1。这是因为添加、删除或修改单个用户最多可以将不同用户的计数更改 1。
- 最小值是聚合函数（例如 `sum()`）中使用的表达式的下限。例如，如果表达式是名为 `purchase_value` 的列，则最小值是该列的下限。
- 最大值是聚合函数（例如 `sum()`）中使用的表达式的上限。例如，如果表达式是名为 `purchase_value` 的列，则最大值是该列的上限。

在计算得出的差别隐私参数表中，您可以使用这些参数更好地了解查询结果中的总噪声量。例如，如果配置的每个查询添加的噪声为 30 个用户并运行 `COUNT DISTINCT (user_id)` 查询，则 AWS Clean Rooms Differential Privacy 添加的随机噪声在很大概率上位于 -30 到 30 之间，因为 `COUNT DISTINCT` 的灵敏度为 1。对于具有相同配置的 `COUNT` 查询，AWS Clean Rooms Differential Privacy 添加按用户贡献限制扩展的统计噪声，因为单个用户可能为查询结果贡献多个行。对于 `SUM` 查询（如 `SUM (purchase_value)`），所有列值均为正值，总噪声按用户贡献限制乘以最大值进行扩展。AWS Clean RoomsDifferential Privacy 自动计算灵敏度参数，以在查询运行时执行噪声添加并耗尽隐私预算。由于灵敏度参数依赖于数据，因此，需要耗尽隐私预算。

管理中配置的表 AWS Clean Rooms

以下主题介绍如何 AWS Clean Rooms 使用 AWS Clean Rooms 控制台管理配置的表。

有关如何使用 AWS 软件开发工具包管理已配置表的信息，请参阅 AWS Clean Rooms API 参考。

主题

编辑配置表的详细信息

作为协作成员，您可以编辑配置表的详细信息。

编辑配置表的详细信息

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择您创建的配置表。
4. 在配置表详细信息页面上，向下滚动到已配置表的详细信息。
5. 选择编辑。
6. 更新配置表的名称或描述。
7. 选择保存更改。

编辑配置表标签

作为协作成员，在创建已配置表后，您可以在配置表选项卡上管理配置表资源上的标签。

编辑配置表标签

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择您创建的配置表。
4. 在配置表详细信息页面上，向下滚动到标签部分。
5. 选择管理标签。
6. 在管理标签页面上，可以执行以下操作：
   - 要删除标签，请选择移除。
   - 要添加标签，请选择添加新标签。
   - 要保存您的更改，请选择保存更改。

编辑配置表的分析规则

编辑配置表的分析规则

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户 （如果您尚未这样做）。

2. 在左侧导航窗格中，选择配置表。
3. 选择您创建的配置表。
4. 在配置表详细信息页面上，向下滚动到聚合分析规则、列表分析规则或自定义分析规则部分。（您的选择取决于您为配置表选择的分析规则类型。）
5. 选择编辑。
6. 在编辑分析规则页面上，您可以：
   - 通过以下方式修改分析规则定义：
     - 修改 JSON 编辑器。
     - 选择从文件导入以上传新的分析规则定义。
   - 从以下选项中进行选择，预览成员将在协作中看到的内容：
     - 表视图
     - JSON
     - 查询示例
7. 选择保存更改以保存您的更改。

删除配置表分析规则

> **Warning**
>
> 此操作无法撤消，并且会影响所有相关资源。

删除配置表分析规则

1. 登录 AWS Management Console 并使用您的AWS Clean Rooms 主机打开主机 AWS 账户（如果您尚未这样做）。
2. 在左侧导航窗格中，选择配置表。
3. 选择您创建的配置表。
4. 在配置表详细信息页面上，向下滚动到聚合分析规则、列表分析规则或自定义分析规则部分。（您的选择取决于您为配置表选择的分析规则类型。）
5. 选择删除。
6. 如果您确定要删除分析规则，请选择删除。

故障排除 AWS Clean Rooms

本节介绍使用时可能出现的一些常见问题 AWS Clean Rooms 以及如何解决这些问题。

问题

查询所引用的一个或多个表不能由其关联的服务角色访问。表/角色所有者必须向服务角色授予对表的访问权限。

- 验证服务角色的权限是否已按要求设置。有关更多信息，请参阅设置 AWS Clean Rooms。

其中一个底层数据集的文件格式不受支持。

- 确保您的数据集采用支持的文件格式之一：
  - Parquet
  - RCFile
  - TextFile
  - SequenceFile
  - RegexSerde
  - OpenCSV
  - AVRO
  - JSON

  有关更多信息，请参阅 的数据格式 AWS Clean Rooms。

使用 Clean Rooms 加密计算时，查询结果不如预期。

如果您使用 Clean Rooms 加密计算 (C3R)，请验证您的查询是否正确使用了加密列：

- sealed 列仅用于 SELECT 子句。
- fingerprint 列仅用于 JOIN 子句（以及某些条件下的 GROUP BY 子句）。
- 只有在协作设置要求的情况下，才 JOINing 具有相同名称的 fingerprint 列。

有关更多信息，请参阅 加密计算 和 the section called “列类型”。

安全性 AWS Clean Rooms

云安全 AWS 是重中之重。作为 AWS 客户，您可以受益于专为满足大多数安全敏感型组织的要求而构建的数据中心和网络架构。

安全是双方共同承担 AWS 的责任。责任共担模式将此描述为云的安全性和云中的安全性：

- 云安全 — AWS 负责保护在 AWS 云中运行 AWS 服务的基础架构。 AWS 还为您提供可以安全使用的服务。作为AWS 合规计划合规计划合规计划合的一部分，第三方审计师定期测试和验证我们安全的有效性。要了解适用于的合规计划 AWS Clean Rooms，请参阅按合规计划划分的 AWS 范围内服务 AWS 按合规计划。
- 云端安全-您的责任由您使用的 AWS 服务决定。您还需要对其他因素负责，包括您的数据的敏感性、您公司的要求以及适用的法律法规。

本文档可帮助您了解在使用时如何应用分担责任模型 AWS Clean Rooms。它向您展示了如何进行配置 AWS Clean Rooms 以满足您的安全和合规性目标。您还将学习如何使用其他 AWS 服务来帮助您监控和保护您的 AWS Clean Rooms 资源。

内容

中的数据保护 AWS Clean Rooms

分 AWS 担责任模型适用于中的数据保护 AWS Clean Rooms。如本模型所述 AWS ，负责保护运行所有内容的全球基础架构 AWS Cloud。您负责维护对托管在此基础架构上的内容的控制。您还负责您所使用的 AWS 服务 的安全配置和管理任务。有关数据隐私的更多信息，请参阅数据隐私常见问

题。有关欧洲数据保护的信息，请参阅 AWS Security Blog 上的 AWS Shared Responsibility Model and GDPR 博客文章。

出于数据保护目的，我们建议您保护 AWS 账户 凭证并使用 AWS IAM Identity Center 或 AWS Identity and Access Management (IAM) 设置个人用户。这样，每个用户只获得履行其工作职责所需的权限。我们还建议您通过以下方式保护数据：

- 对每个账户使用多重身份验证（MFA）。
- 使用 SSL/TLS 与资源通信。 AWS 我们要求使用 TLS 1.2，建议使用 TLS 1.3。
- 使用设置 API 和用户活动日志 AWS CloudTrail。
- 使用 AWS 加密解决方案以及其中的所有默认安全控件 AWS 服务。
- 使用高级托管安全服务（例如 Amazon Macie），它有助于发现和保护存储在 Amazon S3 中的敏感数据。
- 如果您在 AWS 通过命令行界面或 API 进行访问时需要经过 FIPS 140-2 验证的加密模块，请使用 FIPS 端点。有关可用的 FIPS 端点的更多信息，请参阅《美国联邦信息处理标准 (FIPS) 第 140-2 版》。

我们强烈建议您切勿将机密信息或敏感信息（如您客户的电子邮件地址）放入标签或自由格式文本字段（如名称字段）。这包括您使用控制台、API AWS Clean Rooms 或 SDK 或以其他 AWS 服务 方式使用控制台 AWS CLI、API 或 AWS SDK 的情况。在用于名称的标签或自由格式文本字段中输入的任何数据都可能会用于计费或诊断日志。如果您向外部服务器提供网址，强烈建议您不要在网址中包含凭证信息来验证对该服务器的请求。

静态加密

AWS Clean Rooms 始终对所有静态服务元数据进行加密，无需任何其他配置。使用时会自动进行加密 AWS Clean Rooms。

Clean Rooms ML 对存储在服务中的所有静态数据进行加密。 AWS KMS如果您选择提供自己的 KMS 密钥，则相似模型和相似细分生成作业内容将使用您的 KMS 密钥进行静态加密。

> **Note**
>
> 您可以利用 Amazon S3 中的加密选项来保护静态数据。
> 有关更多信息，请参阅《Amazon S3 用户指南》中的指定 Amazon S3 加密。

传输中加密

AWS Clean Rooms 使用传输层安全 (TLS) 和客户端加密进行传输中的加密。与的通信 AWS Clean Rooms 始终通过 HTTPS 完成，因此您的数据在传输过程中始终处于加密状态。这包括使用 Clean Rooms ML 时传输的所有数据。

加密底层数据

有关如何解密您的底层数据的更多信息，请参阅Clean Rooms 加密计算。

数据保留在 AWS Clean Rooms

创建外观相似的模型时，Clean Rooms ML 会读取您的训练数据，将其转换为适合我们的 ML 模型的格式，并将经过训练的模型参数存储在 Clean Rooms ML 中。Clean Rooms ML 不会保留您的训练数据的副本。 AWS Clean Rooms 查询运行后，SQL 查询不会保留您的任何数据。然后，Clean Rooms ML 使用经过训练的模型来总结所有用户的行为。只要你的相似模型处于活动状态，Clean Rooms ML 就会为你的数据中的每个用户存储一个用户级别的数据集。

当您启动相似区段生成作业时，Clean Rooms ML 会读取种子数据，从关联的相似模型中读取行为摘要，然后创建存储在服务中的相似区段。 AWS Clean Rooms Clean Rooms ML 不会保留您的种子数据的副本。只要作业处于活动状态，Clean Rooms ML 就会存储作业的用户级输出。

如果要删除相似模型或相似细分生成作业数据，请使用 API 将其删除。Clean Rooms ML 会异步删除与模型或作业关联的所有数据。此过程完成后，Clean Rooms ML 会删除模型或作业的元数据，并且该元数据在 API 中不再可见。为了防止灾难恢复，Clean Rooms ML 会将已删除的数据保留 3 天。在 API 中不再显示作业或模型并且经过 3 天后，将永久删除与模型或作业关联的所有数据。

中数据协作的最佳实践 AWS Clean Rooms

本主题介绍在 AWS Clean Rooms中开展数据协作的最佳实践。

AWS Clean Rooms 遵循AWS 分担责任模型。 AWS Clean Rooms 提供了分析规则，您可以配置这些规则以增强在协作中保护敏感数据的能力。您在中配置的分析规则 AWS Clean Rooms 将强制执行您配置的限制（查询控件和查询输出控件）。您负责确定限制并相应地配置分析规则。

数据协作可能涉及的不仅仅是您的使用。 AWS Clean Rooms为了帮助您最大限度地发挥数据协作的优势，我们建议您在使用分析规则时执行以下最佳实践 AWS Clean Rooms ，特别是分析规则。

主题

最佳实践 AWS Clean Rooms

您负责评估每个数据协作的风险，并将其与您的隐私要求（例如外部和内部合规性计划和策略）进行比较。我们建议您在使用时采取其他措施 AWS Clean Rooms。这些操作可能有助于进一步管理风险，并有助于防范第三方试图重新识别您的数据（例如，差异攻击或侧信道攻击）。

例如，考虑对您的其他协作者进行尽职调查，并在进行协作之前 与他们签订法律协议。要监控数据的使用情况，还要考虑在使用 AWS Clean Rooms时采用其他审计机制。

在 AWS Clean Rooms中使用分析规则的最佳实践

中的分析规则 AWS Clean Rooms 允许您通过在已配置的表上设置查询控件来限制可以运行的查询。例如，您可以设置查询控制，以确定如何联接配置表以及可以选择哪些列。您还可以通过设置查询结果控制（例如输出行的聚合阈值）来限制查询输出。该服务拒绝任何查询，并删除不符合成员在查询中配置表上设置的分析规则的行。

对于在配置表上使用分析规则，我们推荐以下 10 种最佳实践：

- 为不同的查询使用案例（例如受众规划或归因）创建单独的配置表。您可以使用同一底层 AWS Glue 表创建多个配置表。
- 在分析规则中指定协作中查询所必需的列（例如维度列、列表列、联接列）。这可能有助于降低差异攻击的风险或使其他成员能够对您的数据进行逆向工程。使用允许列表列功能记下将来可能要设置为可查询的其他列。要自定义可用于特定协作的列，请使用相同的基础表创建其他已配置 AWS Glue 表。
- 在分析规则中指定协作中分析所必需的函数。这有助于降低因罕见的函数错误而带来的风险，这些错误可能会显示单个数据点的信息。要自定义可用于特定协作的函数，请使用同一底层 AWS Glue 表创建其他配置表。
- 对行级值敏感的任何列添加聚合约束。这包括您的配置表中的列，这些列也存在于其他协作成员的表中，并且有分析规则作为聚合约束。这也包括您的配置表中不可查询的列，即配置表中有但不在分析规则中的列。聚合约束可以帮助降低将查询结果与协作之外的数据关联起来的风险。
- 创建测试协作和分析规则，以测试使用指定分析规则创建的限制。
- 查看协作者配置表和成员对配置表的分析规则，以检查它们是否符合协作商定的内容。这可以帮助降低其他成员设计自己的数据以运行未商定的查询所带来的风险。

- 查看提供的示例查询（仅限控制台），该查询在设置分析规则后在配置表上启用。

> **Note**
>
> 除了提供的示例查询外，还可以根据分析规则和其他协作成员表和分析规则进行其他查询。

- 您可以为协作中的配置表添加或更新分析规则。完成后，请查看与配置表关联的所有协作及其产生的影响。这有助于确保任何协作都不会使用过时的分析规则。
- 审核协作中运行的查询，检查查询是否与协作中商定的使用案例或查询相匹配。（打开查询日志记录功能后，可在查询日志中查看查询）。这可以帮助降低成员运行未商定的分析和潜在攻击（例如侧信道攻击）带来的风险。
- 审核协作成员分析规则和查询中使用的配置表列，检查它们是否与协作中商定的内容相匹配。（打开该功能后，可在查询日志中查看查询。）这可以帮助降低其他成员设计自己的数据以进行未商定的查询所带来的风险。

适用于 Identity and Access Managem AWS Clean Rooms

AWS Identity and Access Management (IAM) AWS 服务 可帮助管理员安全地控制对 AWS 资源的访问权限。IAM 管理员控制谁可以进行身份验证（登录）和授权（有权限）使用 AWS Clean Rooms 资源。您可以使用 IAM AWS 服务 ，无需支付额外费用。

主题

受众

您的使用方式 AWS Identity and Access Management (IAM) 会有所不同，具体取决于您所做的工作 AWS Clean Rooms。

服务用户-如果您使用该 AWS Clean Rooms 服务完成工作，则您的管理员会为您提供所需的凭证和权限。当你使用更多 AWS Clean Rooms 功能来完成工作时，你可能需要额外的权限。了解如何管理访问权限有助于您向管理员请求适合的权限。如果您无法访问 AWS Clean Rooms中的特征，请参阅 对 AWS Clean Rooms 身份和访问进行故障排除。

服务管理员-如果您负责公司的 AWS Clean Rooms 资源，则可能拥有完全访问权限 AWS Clean Rooms。您的工作是确定您的服务用户应访问哪些 AWS Clean Rooms 功能和资源。然后，您必须向 IAM 管理员提交请求以更改服务用户的权限。请查看该页面上的信息以了解 IAM 的基本概念。要详细了解您的公司如何将 IAM 与配合使用 AWS Clean Rooms，请参阅如何 AWS Clean Rooms 与 IAM 配合使用。

IAM 管理员：如果您是 IAM 管理员，您可能希望了解如何编写策略以管理对 AWS Clean Rooms的访问权限的详细信息。要查看您可以在 IAM 中使用的 AWS Clean Rooms 基于身份的策略示例，请参阅。基于身份的策略示例 AWS Clean Rooms

使用身份进行身份验证

身份验证是您 AWS 使用身份凭证登录的方式。您必须以 IAM 用户身份或通过担 AWS 账户根用户任 IAM 角色进行身份验证（登录 AWS）。

您可以使用通过身份源提供的凭据以 AWS 联合身份登录。 AWS IAM Identity Center （ IAM Identity Center ）用户或贵公司的单点登录身份验证就是联合身份的示例。当您以联合身份登录时，您的管理员以前使用 IAM 角色设置了身份联合验证。当你使用联合访问 AWS 时，你就是在间接扮演一个角色。

根据您的用户类型，您可以登录 AWS Management Console 或 AWS 访问门户。有关登录的更多信息 AWS，请参阅《AWS 登录 用户指南》中的如何登录到您 AWS 账户的。

如果您 AWS 以编程方式访问，则会 AWS 提供软件开发套件 (SDK) 和命令行接口 (CLI)，以便使用您的凭据对请求进行加密签名。如果您不使用 AWS 工具，则必须自己签署请求。有关使用推荐的方法自行对请求签名的更多信息，请参阅《AWS 一般参考》中的 签名版本 4 签名流程。

无论使用何种身份验证方法，您可能需要提供其他安全信息。例如， AWS 建议您使用多重身份验证 (MFA) 来提高账户的安全性。要了解更多信息，请参阅《AWS IAM Identity Center 用户指南》中的 多重身份验证和《IAM 用户指南》中的 在 AWS中使用多重身份验证（ MFA ）。

AWS 账户 root 用户

创建时 AWS 账户，首先要有一个登录身份，该身份可以完全访问账户中的所有资源 AWS 服务 和资源。此身份称为 AWS 账户 根用户，使用您创建账户时所用的电子邮件地址和密码登录，即可获得该身份。强烈建议您不要使用根用户执行日常任务。保护好根用户凭证，并使用这些凭证来执行仅根用户可以执行的任务。有关需要您以根用户身份登录的任务的完整列表，请参阅《AWS 一般参考》中的AWS 账户根用户 凭证和 IAM 身份。

联合身份

作为最佳实践，要求人类用户（包括需要管理员访问权限的用户）使用与身份提供商的联合身份验证 AWS 服务 通过临时证书进行访问。

联合身份是指您的企业用户目录、Web 身份提供商、Identity C enter 目录中的用户，或者任何使用 AWS 服务 通过身份源提供的凭据进行访问的用户。 AWS Directory Service当联合身份访问时 AWS 账户，他们将扮演角色，角色提供临时证书。

要集中管理访问权限，建议您使用 AWS IAM Identity Center。您可以在 IAM Identity Center 中创建用户和群组，也可以连接并同步到您自己的身份源中的一组用户和群组，以便在您的所有 AWS 账户 和应用程序中使用。有关 IAM Identity Center 的信息，请参阅《AWS IAM Identity Center 用户指南》中的什么是 IAM Identity Center？

IAM 用户和群组

I AM 用户是您 AWS 账户 内部对个人或应用程序具有特定权限的身份。在可能的情况下，我们建议使用临时凭证，而不是创建具有长期凭证（如密码和访问密钥）的 IAM 用户。但是，如果您有一些特定的使用场景需要长期凭证以及 IAM 用户，建议您轮换访问密钥。有关更多信息，请参阅《IAM 用户指南》中的 对于需要长期凭证的使用场景定期轮换访问密钥。

IAM 组是一个指定一组 IAM 用户的身份。您不能使用组的身份登录。您可以使用组来一次性为多个用户指定权限。如果有大量用户，使用组可以更轻松地管理用户权限。例如，您可能具有一个名为 IAMAdmins 的组，并为该组授予权限以管理 IAM 资源。

用户与角色不同。用户唯一地与某个人员或应用程序关联，而角色旨在让需要它的任何人代入。用户具有永久的长期凭证，而角色提供临时凭证。要了解更多信息，请参阅《IAM 用户指南》中的何时创建 IAM 用户（而不是角色）。

IAM 角色

I AM 角色是您内部具有特定权限 AWS 账户 的身份。它类似于 IAM 用户，但与特定人员不关联。您可以 AWS Management Console 通过切换角色在中临时担任 IAM 角色。您可以通过调用 AWS CLI 或

AWS API 操作或使用自定义 URL 来代入角色。有关使用角色的方法的更多信息，请参阅《IAM 用户指南》中的使用 IAM 角色。

具有临时凭证的 IAM 角色在以下情况下很有用：

- 联合用户访问 – 要向联合身份分配权限，请创建角色并为角色定义权限。当联合身份进行身份验证时，该身份将与角色相关联并被授予由此角色定义的权限。有关联合身份验证的角色的信息，请参阅《IAM 用户指南》中的为第三方身份提供商创建角色。如果您使用 IAM Identity Center，则需要配置权限集。为控制您的身份在进行身份验证后可以访问的内容，IAM Identity Center 将权限集与 IAM 中的角色相关联。有关权限集的信息，请参阅《AWS IAM Identity Center 用户指南》中的 权限集。
- 临时 IAM 用户权限 – IAM 用户可代入 IAM 用户或角色，以暂时获得针对特定任务的不同权限。
- 跨账户存取 – 您可以使用 IAM 角色以允许不同账户中的某个人（可信主体）访问您的账户中的资源。角色是授予跨账户访问权限的主要方式。但是，对于某些资源 AWS 服务，您可以将策略直接附加到资源（而不是使用角色作为代理）。要了解角色和基于资源的跨账户访问策略之间的区别，请参阅 IA M 用户指南中的跨账户资源访问。
- 跨服务访问 — 有些 AWS 服务 使用其他 AWS 服务服务中的功能。例如，当您在某个服务中进行调用时，该服务通常会在 Amazon EC2 中运行应用程序或在 Simple Storage Service (Amazon S3) 中存储对象。服务可能会使用发出调用的主体的权限、使用服务角色或使用服务相关角色来执行此操作。
  - 转发访问会话 (FAS) — 当您使用 IAM 用户或角色在中执行操作时 AWS，您被视为委托人。使用某些服务时，您可能会执行一个操作，然后此操作在其他服务中启动另一个操作。FAS 使用调用委托人的权限以及 AWS 服务 向下游服务发出请求的请求。 AWS 服务只有当服务收到需要与其他 AWS 服务 或资源交互才能完成的请求时，才会发出 FAS 请求。在这种情况下，您必须具有执行这两个操作的权限。有关发出 FAS 请求时的策略详情，请参阅转发访问会话。
  - 服务角色 - 服务角色是服务代表您在您的账户中执行操作而分派的 IAM 角色。IAM 管理员可以在 IAM 中创建、修改和删除服务角色。有关更多信息，请参阅《IAM 用户指南》中的创建向 AWS 服务委派权限的角色。
  - 服务相关角色-服务相关角色是一种与服务相关联的服务角色。 AWS 服务服务可以代入代表您执行操作的角色。服务相关角色出现在您的中 AWS 账户 ，并且归服务所有。IAM 管理员可以查看但不能编辑服务相关角色的权限。
- 在 A@@ mazon EC2 上运行的应用程序 — 您可以使用 IAM 角色管理在 EC2 实例上运行并发出 AWS CLI 或 AWS API 请求的应用程序的临时证书。这优先于在 EC2 实例中存储访问密钥。要向 EC2 实例分配 AWS 角色并使其可供其所有应用程序使用，您需要创建附加到该实例的实例配置文件。实例配置文件包含角色，并使 EC2 实例上运行的程序能够获得临时凭证。有关更多信息，请参阅《IAM 用户指南》中的使用 IAM 角色为 Amazon EC2 实例上运行的应用程序授予权限。

要了解是使用 IAM 角色还是 IAM 用户，请参阅IAM 用户指南中的何时创建 IAM 角色（而不是用户）。

使用策略管理访问

您可以 AWS 通过创建策略并将其附加到 AWS 身份或资源来控制中的访问权限。策略是其中的一个对象 AWS ，当与身份或资源关联时，它会定义其权限。 AWS 在委托人（用户、root 用户或角色会话）发出请求时评估这些策略。策略中的权限确定是允许还是拒绝请求。大多数策略都以 JSON 文档的 AWS 形式存储在中。有关 JSON 策略文档的结构和内容的更多信息，请参阅 IAM 用户指南中的 JSON 策略概览。

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体 可以对什么资源 执行操作，以及在什么条件 下执行。

每个 IAM 实体（用户或角色）最初没有任何权限。原定设置情况下，用户什么都不能做，甚至不能更改他们自己的密码。要为用户授予执行某些操作的权限，管理员必须将权限策略附加到用户。或者，管理员可以将用户添加到具有预期权限的组中。当管理员为某个组授予访问权限时，该组内的全部用户都会获得这些访问权限。

IAM 策略定义操作的权限，无关乎您使用哪种方法执行操作。例如，假设您有一个允许 `iam:GetRole` 操作的策略。拥有该策略的用户可以从 AWS Management Console AWS CLI、或 AWS API 获取角色信息。

基于身份的策略

基于身份的策略是可附加到身份（如 IAM 用户、用户组或角色）的 JSON 权限策略文档。这些策略控制用户和角色可在何种条件下对哪些资源执行哪些操作。要了解如何创建基于身份的策略，请参阅《IAM 用户指南》中的创建 IAM policy。

基于身份的策略可以进一步归类为内联策略或托管策略。内联策略直接嵌入单个用户、组或角色中。托管式策略是可以附加到 AWS 账户中的多个用户、组和角色的独立策略。托管策略包括 AWS 托管策略和客户托管策略。要了解如何在托管式策略和内联策略之间进行选择，请参阅 IAM 用户指南中的在托管式策略与内联策略之间进行选择。

基于资源的策略

基于资源的策略是附加到资源的 JSON 策略文档。基于资源的策略的示例包括 IAM 角色信任策略和 Amazon S3 存储桶策略。在支持基于资源的策略的服务中，服务管理员可以使用它们来控制对特定资

源的访问。对于在其中附加策略的资源，策略定义指定主体可以对该资源执行哪些操作以及在什么条件下执行。您必须在基于资源的策略中指定主体。委托人可以包括账户、用户、角色、联合用户或 AWS 服务。

基于资源的策略是位于该服务中的内联策略。您不能在基于资源的策略中使用 IAM 中的 AWS 托管策略。

其他策略类型

AWS 支持其他不太常见的策略类型。这些策略类型可以设置更常用的策略类型向您授予的最大权限。

- 权限边界：权限边界是一个高级特征，用于设置基于身份的策略可以为 IAM 实体（IAM 用户或角色）授予的最大权限。您可为实体设置权限边界。这些结果权限是实体的基于身份的策略及其权限边界的交集。在 `Principal` 中指定用户或角色的基于资源的策略不受权限边界限制。任一项策略中的显式拒绝将覆盖允许。有关权限边界的更多信息，请参阅《IAM 用户指南》中的 IAM 实体的权限边界。
- 服务控制策略 (SCP)-SCP 是 JSON 策略，用于指定组织或组织单位 (OU) 的最大权限。 AWS Organizations AWS Organizations 是一项用于对您的企业拥有的多 AWS 账户 项进行分组和集中管理的服务。如果在组织内启用了所有功能，则可对任意或全部账户应用服务控制策略 (SCP)。SCP 限制成员账户中的实体（包括每个 AWS 账户根用户实体）的权限。有关 Organizations 和 SCP 的更多信息，请参阅《AWS Organizations 用户指南》中的 SCP 的工作原理。
- 会话策略 – 会话策略是当您以编程方式为角色或联合用户创建临时会话时作为参数传递的高级策略。结果会话的权限是用户或角色的基于身份的策略和会话策略的交集。权限也可以来自基于资源的策略。任一项策略中的显式拒绝将覆盖允许。有关更多信息，请参阅《IAM 用户指南》中的会话策略。

多个策略类型

当多个类型的策略应用于一个请求时，生成的权限更加复杂和难以理解。要了解在涉及多种策略类型时如何 AWS 确定是否允许请求，请参阅 IAM 用户指南中的策略评估逻辑。

如何 AWS Clean Rooms 与 IAM 配合使用

在使用 IAM 管理访问权限之前 AWS Clean Rooms，请先了解有哪些 IAM 功能可供使用 AWS Clean Rooms。

您可以搭配使用的 IAM 功能 AWS Clean Rooms

IAM 功能	AWS Clean Rooms 支持
基于身份的策略	是
基于资源的策略	部分
策略操作	是
策略资源	是
策略条件键（特定于服务）	部分
ACL	否
ABAC（策略中的标签）	是
临时凭证	是
转发访问会话（FAS）	是
服务角色	是
服务相关角色	否

要全面了解大多数 IAM 功能的使用 AWS 服务 方式 AWS Clean Rooms 和其他功能，请参阅AWS 服务 IAM 用户指南中的与 IAM 配合使用的内容。

基于身份的策略 AWS Clean Rooms

支持基于身份的策略	是

基于身份的策略是可附加到身份（如 IAM 用户、用户组或角色）的 JSON 权限策略文档。这些策略控制用户和角色可在何种条件下对哪些资源执行哪些操作。要了解如何创建基于身份的策略，请参阅《IAM 用户指南》中的创建 IAM policy。

通过使用 IAM 基于身份的策略，您可以指定允许或拒绝的操作和资源以及允许或拒绝操作的条件。您无法在基于身份的策略中指定主体，因为它适用于其附加的用户或角色。要了解可在 JSON 策略中使用的所有元素，请参阅《IAM 用户指南》中的 IAM JSON 策略元素引用。

基于身份的策略示例 AWS Clean Rooms

要查看 AWS Clean Rooms 基于身份的策略的示例，请参阅。基于身份的策略示例 AWS Clean Rooms

内部基于资源的政策 AWS Clean Rooms

支持基于资源的策略	部分

基于资源的策略是附加到资源的 JSON 策略文档。基于资源的策略的示例包括 IAM 角色信任策略和 Amazon S3 存储桶策略。在支持基于资源的策略的服务中，服务管理员可以使用它们来控制对特定资源的访问。对于在其中附加策略的资源，策略定义指定主体可以对该资源执行哪些操作以及在什么条件下执行。您必须在基于资源的策略中指定主体。委托人可以包括账户、用户、角色、联合用户或 AWS 服务。

要启用跨账户存取，您可以将整个账户或其他账户中的 IAM 实体指定为基于资源的策略中的主体。将跨账户主体添加到基于资源的策略只是建立信任关系工作的一半而已。当委托人和资源处于不同位置时 AWS 账户，可信账户中的 IAM 管理员还必须向委托人实体（用户或角色）授予访问资源的权限。他们通过将基于身份的策略附加到实体以授予权限。但是，如果基于资源的策略向同一个账户中的主体授予访问权限，则不需要额外的基于身份的策略。有关更多信息，请参阅 IAM 用户指南中的跨账户在 IAM 中访问资源。

该 AWS Clean Rooms 服务仅支持一种基于资源的策略，称为配置的相似模型托管资源策略，该策略附加到已配置的相似模型上。此策略定义了哪些委托人可以对配置的相似模型执行操作。

要了解如何将基于资源的策略附加到已配置的相似模型，请参阅。AWS Clean Rooms 机器学习的 IAM 行为

的政策行动 AWS Clean Rooms

支持策略操作	是

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

JSON 策略的 `Action` 元素描述可用于在策略中允许或拒绝访问的操作。策略操作通常与关联的 AWS API 操作同名。有一些例外情况，例如没有匹配 API 操作的仅限权限 操作。还有一些操作需要在策略中执行多个操作。这些附加操作称为相关操作。

在策略中包含操作以授予执行关联操作的权限。

要查看 AWS Clean Rooms 操作列表，请参阅《服务授权参考》 AWS Clean Rooms中定义的操作。

正在执行的策略操作在操作前 AWS Clean Rooms 使用以下前缀。

```
cleanrooms
```

要在单个语句中指定多项操作，请使用逗号将它们隔开。

```
"Action": [
      "cleanrooms:action1",
      "cleanrooms:action2"
         ]
```

要查看 AWS Clean Rooms 基于身份的策略的示例，请参阅。基于身份的策略示例 AWS Clean Rooms

的政策资源 AWS Clean Rooms

支持策略资源	是

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体 可以对什么资源执行操作，以及在什么条件下执行。

`Resource` JSON 策略元素指定要向其应用操作的一个或多个对象。语句必须包含 `Resource` 或 `NotResource` 元素。作为最佳实践，请使用其 Amazon 资源名称 (ARN) 指定资源。对于支持特定资源类型（称为资源级权限）的操作，您可以执行此操作。

对于不支持资源级权限的操作（如列出操作），请使用通配符 (*) 指示语句应用于所有资源。

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

要查看 AWS Clean Rooms 资源类型及其 ARN 的列表，请参阅《服务授权参考》 AWS Clean Rooms 中定义的资源。要了解可以在哪些操作中指定每个资源的 ARN，请参阅 AWS Clean Rooms定义的操作。

要查看 AWS Clean Rooms 基于身份的策略的示例，请参阅。基于身份的策略示例 AWS Clean Rooms

的策略条件密钥 AWS Clean Rooms

支持特定于服务的策略条件密钥	部分

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

在 Condition 元素（或 Condition 块）中，可以指定语句生效的条件。Condition 元素是可选的。您可以创建使用条件运算符（例如，等于或小于）的条件表达式，以使策略中的条件与请求中的值相匹配。

如果您在一个语句中指定多个 Condition 元素，或在单个 Condition 元素中指定多个键，则 AWS 使用逻辑 AND 运算评估它们。如果您为单个条件键指定多个值，则使用逻辑OR运算来 AWS 评估条件。在授予语句的权限之前必须满足所有的条件。

在指定条件时，您也可以使用占位符变量。例如，只有在使用 IAM 用户名标记 IAM 用户时，您才能为其授予访问资源的权限。有关更多信息，请参阅《IAM 用户指南》中的 IAM policy 元素：变量和标签。

AWS 支持全局条件密钥和特定于服务的条件密钥。要查看所有 AWS 全局条件键，请参阅 IAM 用户指南中的AWS 全局条件上下文密钥。

要了解 AWS Clean Rooms ML 如何使用策略条件密钥，请参阅AWS Clean Rooms 机器学习的 IAM 行为。

输入的 ACL AWS Clean Rooms

支持 ACL	否

访问控制列表 (ACL) 控制哪些主体（账户成员、用户或角色）有权访问资源。ACL 与基于资源的策略类似，尽管它们不使用 JSON 策略文档格式。

ABAC with AWS Clean Rooms

支持 ABAC（策略中的标签）	是

基于属性的访问控制 (ABAC) 是一种授权策略，该策略基于属性来定义权限。在中 AWS，这些属性称为标签。您可以将标签附加到 IAM 实体（用户或角色）和许多 AWS 资源。标记实体和资源是 ABAC 的第一步。然后设计 ABAC 策略，以在主体的标签与他们尝试访问的资源标签匹配时允许操作。

ABAC 在快速增长的环境中非常有用，并在策略管理变得繁琐的情况下可以提供帮助。

要基于标签控制访问，您需要使用 `aws:ResourceTag/key-name`、`aws:RequestTag/key-name` 或 `aws:TagKeys` 条件键在策略的条件元素中提供标签信息。

如果某个服务对于每种资源类型都支持所有这三个条件键，则对于该服务，该值为是。如果某个服务仅对于部分资源类型支持所有这三个条件键，则该值为部分。

有关 ABAC 的更多信息,请参阅《IAM 用户指南》中的什么是 ABAC？。要查看设置 ABAC 步骤的教程，请参阅《IAM 用户指南》中的使用基于属性的访问权限控制（ABAC）。

将临时证书与 AWS Clean Rooms

支持临时凭证	是

当你使用临时证书登录时，有些 AWS 服务 不起作用。有关更多信息，包括哪些 AWS 服务 适用于临时证书，请参阅 IAM 用户指南中的AWS 服务 与 IA M 配合使用的信息。

如果您使用除用户名和密码之外的任何方法登录，则 AWS Management Console 使用的是临时证书。例如，当您 AWS 使用公司的单点登录 (SSO) 链接进行访问时，该过程会自动创建临时证书。当您以用户身份登录控制台，然后切换角色时，您还会自动创建临时凭证。有关切换角色的更多信息，请参阅《IAM 用户指南》中的 切换到角色（控制台）。

您可以使用 AWS CLI 或 AWS API 手动创建临时证书。然后，您可以使用这些临时证书进行访问 AWS。 AWS 建议您动态生成临时证书，而不是使用长期访问密钥。有关更多信息，请参阅 IAM 中的临时安全凭证。

转发访问会话 AWS Clean Rooms

支持转发访问会话 (FAS)	是

当您使用 IAM 用户或角色在中执行操作时 AWS，您被视为委托人。使用某些服务时，您可能会执行一个操作，然后此操作在其他服务中启动另一个操作。FAS 使用调用委托人的权限以及 AWS 服务 向下游服务发出请求的请求。 AWS 服务只有当服务收到需要与其他 AWS 服务 或资源交互才能完成的请求时，才会发出 FAS 请求。在这种情况下，您必须具有执行这两个操作的权限。有关发出 FAS 请求时的策略详细信息，请参阅转发访问会话。

AWS Clean Rooms的服务角色

支持服务角色	是

服务角色是由一项服务担任、代表您执行操作的 IAM 角色。IAM 管理员可以在 IAM 中创建、修改和删除服务角色。有关更多信息，请参阅《IAM 用户指南》中的创建向 AWS 服务委派权限的角色。

> **Warning**
>
> 更改服务角色的权限可能会中断 AWS Clean Rooms 功能。只有在 AWS Clean Rooms 提供操作指导时才编辑服务角色。

的服务相关角色 AWS Clean Rooms

支持服务相关角色	否

服务相关角色是一种链接到的服务角色。 AWS 服务服务可以代入代表您执行操作的角色。服务相关角色出现在您的中 AWS 账户 ，并且归服务所有。IAM 管理员可以查看但不能编辑服务相关角色的权限。

有关创建或管理服务相关角色的详细信息，请参阅能够与 IAM 搭配使用的AWS 服务。在表中查找服务相关角色列中包含 Yes 的表。选择是链接以查看该服务的服务相关角色文档。

基于身份的策略示例 AWS Clean Rooms

默认情况下，用户和角色没有创建或修改 AWS Clean Rooms 资源的权限。他们也无法使用 AWS Management Console、 AWS Command Line Interface (AWS CLI) 或 AWS API 执行任务。要授予用户对所需资源执行操作的权限，IAM 管理员可以创建 IAM policy。管理员随后可以向角色添加 IAM 策略，用户可以代入角色。

要了解如何使用这些示例 JSON 策略文档创建基于 IAM 身份的策略，请参阅 IAM 用户指南中的 创建 IAM 策略。

有关由定义的操作和资源类型的详细信息 AWS Clean Rooms，包括每种资源类型的 ARN 格式，请参阅《服务授权参考》 AWS Clean Rooms中的操作、资源和条件密钥。

主题

- 策略最佳实践
- 使用 AWS Clean Rooms 控制台
- 允许用户查看他们自己的权限

策略最佳实践

基于身份的策略决定了某人是否可以在您的账户中创建、访问或删除 AWS Clean Rooms 资源。这些操作可能会使 AWS 账户产生成本。创建或编辑基于身份的策略时，请遵循以下指南和建议：

- 开始使用 AWS 托管策略并转向最低权限权限 — 要开始向用户和工作负载授予权限，请使用为许多常见用例授予权限的AWS 托管策略。它们在你的版本中可用 AWS 账户。我们建议您通过定义针对您的用例的 AWS 客户托管策略来进一步减少权限。有关更多信息，请参阅《IAM 用户指南》中的 AWS 托管策略 或 工作职能的AWS 托管策略。
- 应用最低权限 – 在使用 IAM policy 设置权限时，请仅授予执行任务所需的权限。为此，您可以定义在特定条件下可以对特定资源执行的操作，也称为最低权限许可。有关使用 IAM 应用权限的更多信息，请参阅《IAM 用户指南》中的 IAM 中的策略和权限。
- 使用 IAM policy 中的条件进一步限制访问权限 – 您可以向策略添加条件来限制对操作和资源的访问。例如，您可以编写策略条件来指定必须使用 SSL 发送所有请求。如果服务操作是通过特定的方式使用的，则也可以使用条件来授予对服务操作的访问权限 AWS 服务，例如 AWS CloudFormation。有关更多信息，请参阅《IAM 用户指南》中的 IAM JSON 策略元素：条件 。

- 使用 IAM Access Analyzer 验证您的 IAM policy，以确保权限的安全性和功能性 – IAM Access Analyzer 会验证新策略和现有策略，以确保策略符合 IAM policy语言 (JSON) 和 IAM 最佳实践。IAM Access Analyzer 提供 100 多项策略检查和可操作的建议，以帮助您制定安全且功能性强的策略。有关更多信息，请参阅《IAM 用户指南》中的 IAM Acess Analyzer 策略验证。
- 需要多重身份验证 (MFA)-如果 AWS 账户您的场景需要 IAM 用户或根用户，请启用 MFA 以提高安全性。若要在调用 API 操作时需要 MFA，请将 MFA 条件添加到您的策略中。有关更多信息，请参阅《IAM 用户指南》中的 配置受 MFA 保护的 API 访问。

有关 IAM 中的最佳实操的更多信息，请参阅《IAM 用户指南》中的 IAM 中的安全最佳实操。

使用 AWS Clean Rooms 控制台

要访问 AWS Clean Rooms 控制台，您必须拥有一组最低权限。这些权限必须允许您列出和查看有关您的 AWS Clean Rooms 资源的详细信息 AWS 账户。如果创建比必需的最低权限更为严格的基于身份的策略，对于附加了该策略的实体（用户或角色），控制台将无法按预期正常运行。

对于仅调用 AWS CLI 或 AWS API 的用户，您无需为其设置最低控制台权限。相反，只允许访问与其尝试执行的 API 操作相匹配的操作。

为确保用户和角色仍然可以使用 AWS Clean Rooms 控制台，还需要将 AWS Clean Rooms *FullAccess*或*ReadOnly* AWS 托管策略附加到实体。有关更多信息，请参阅《IAM 用户指南》中的为用户添加权限。

允许用户查看他们自己的权限

该示例说明了您如何创建策略，以允许 IAM 用户查看附加到其用户身份的内联和托管策略。此策略包括在控制台上或使用 AWS CLI 或 AWS API 以编程方式完成此操作的权限。

```
{
    "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 的托管策略 AWS Clean Rooms

AWS 托管策略是由创建和管理的独立策略 AWS。 AWS 托管策略旨在为许多常见用例提供权限，以便您可以开始为用户、组和角色分配权限。

请记住， AWS 托管策略可能不会为您的特定用例授予最低权限权限，因为它们可供所有 AWS 客户使用。我们建议通过定义特定于您的使用场景的客户管理型策略来进一步减少权限。

您无法更改 AWS 托管策略中定义的权限。如果 AWS 更新 AWS 托管策略中定义的权限，则更新会影响该策略所关联的所有委托人身份（用户、组和角色）。 AWS 最有可能在启动新的 API 或现有服务可以使用新 AWS 服务 的 API 操作时更新 AWS 托管策略。

有关更多信息，请参阅《IAM 用户指南》中的 AWS 托管式策略。

AWS 托管策略：**AWSCleanRoomsReadOnlyAccess**

您可以将 AWSCleanRoomsReadOnlyAccess 附加到 IAM 主体。

该策略授予 AWSCleanRoomsReadOnlyAccess 协作中的资源和元数据的只读权限。

权限详细信息

该策略包含以下权限：

- CleanRoomsRead - 允许主体对服务进行只读访问。
- ConsoleDisplayTables— 允许委托人对在控制台上显示有关基础 AWS Glue 表的数据所需的 AWS Glue 元数据的只读访问权限。
- ConsoleLogSummaryQueryLogs - 允许主体查看查询日志。
- ConsoleLogSummaryObtainLogs - 允许主体检索日志结果。

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "CleanRoomsRead",
   "Effect": "Allow",
   "Action": [
    "cleanrooms:BatchGet*",
    "cleanrooms:Get*",
    "cleanrooms:List*"
   ],
   "Resource": "*"
  },
  {
   "Sid": "ConsoleDisplayTables",
   "Effect": "Allow",
   "Action": [
    "glue:GetDatabase",
    "glue:GetDatabases",
    "glue:GetTable",
    "glue:GetTables",
    "glue:GetPartition",
    "glue:GetPartitions",
    "glue:GetSchema",
    "glue:GetSchemaVersion",
    "glue:BatchGetPartition"
   ],
   "Resource": "*"
  },
  {
   "Sid": "ConsoleLogSummaryQueryLogs",
```

```
    "Effect": "Allow",
    "Action": [
     "logs:StartQuery"
    ],
    "Resource": "arn:aws:logs:*:*:log-group:/aws/cleanrooms*"
   },
   {
    "Sid": "ConsoleLogSummaryObtainLogs",
    "Effect": "Allow",
    "Action": [
     "logs:GetQueryResults"
    ],
    "Resource": "*"
   }
  ]
}
```

AWS 托管策略：**AWSCleanRoomsFullAccess**

您可以将 AWSCleanRoomsFullAccess 附加到 IAM 主体。

此策略授予管理权限，允许对 AWS Clean Rooms 协作中的资源和元数据进行完全访问（读取、写入和更新）。此策略包括执行查询的权限。

权限详细信息

该策略包含以下权限：

- CleanRoomsAccess— 授予对所有资源执行所有操作的完全访问权限 AWS Clean Rooms。
- PassServiceRole - 仅授予将服务角色传递给名称中带有“cleanrooms”的服务（PassedToService 条件）的访问权限。
- ListRolesToPickServiceRole— 允许委托人列出其所有角色以便在使用 AWS Clean Rooms时选择服务角色。
- GetRoleAndListRolePoliciesToInspectServiceRole - 允许主体在 IAM 中查看服务角色和相应的策略。
- ListPoliciesToInspectServiceRolePolicy - 允许主体在 IAM 中查看服务角色和相应的策略。
- GetPolicyToInspectServiceRolePolicy - 允许主体在 IAM 中查看服务角色和相应的策略。
- ConsoleDisplayTables— 允许委托人对在控制台上显示有关基础 AWS Glue 表的数据所需的 AWS Glue 元数据的只读访问权限。

- ConsolePickQueryResultsBucketListAll - 允许主体从查询结果写入的所有可用 S3 存储桶的列表中选择一个 Amazon S3 存储桶。
- SetQueryResultsBucket - 允许主体选择查询结果写入的 S3 存储桶。
- ConsoleDisplayQueryResults - 允许主体向客户显示从 S3 存储桶读取的查询结果。
- WriteQueryResults - 允许主体将查询结果写入客户拥有的 S3 存储桶。
- EstablishLogDeliveries— 允许委托人将查询日志传送到客户的 Amazon Lo CloudWatch gs 日志组。
- SetupLogGroupsDescribe— 允许委托人使用 Amazon Logs CloudWatch 日志组的创建流程。
- SetupLogGroupsCreate— 允许委托人创建 Amazon CloudWatch 日志组。
- SetupLogGroupsResourcePolicy— 允许委托人在 Amazon Logs CloudWatch 日志组上设置资源策略。
- ConsoleLogSummaryQueryLogs - 允许主体查看查询日志。
- ConsoleLogSummaryObtainLogs - 允许主体检索日志结果。

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "CleanRoomsAccess",
   "Effect": "Allow",
   "Action": [
    "cleanrooms:*"
   ],
   "Resource": "*"
  },
  {
   "Sid": "PassServiceRole",
   "Effect": "Allow",
   "Action": [
    "iam:PassRole"
   ],
   "Resource": "arn:aws:iam::*:role/service-role/*cleanrooms*",
   "Condition": {
    "StringEquals": {
     "iam:PassedToService": "cleanrooms.amazonaws.com"
    }
   }
  },
```

```
{
 "Sid": "ListRolesToPickServiceRole",
 "Effect": "Allow",
 "Action": [
  "iam:ListRoles"
 ],
 "Resource": "*"
},
{
 "Sid": "GetRoleAndListRolePoliciesToInspectServiceRole",
 "Effect": "Allow",
 "Action": [
  "iam:GetRole",
  "iam:ListRolePolicies",
  "iam:ListAttachedRolePolicies"
 ],
 "Resource": "arn:aws:iam::*:role/service-role/*cleanrooms*"
},
{
 "Sid": "ListPoliciesToInspectServiceRolePolicy",
 "Effect": "Allow",
 "Action": [
  "iam:ListPolicies"
 ],
 "Resource": "*"
},
{
 "Sid": "GetPolicyToInspectServiceRolePolicy",
 "Effect": "Allow",
 "Action": [
  "iam:GetPolicy",
  "iam:GetPolicyVersion"
 ],
 "Resource": "arn:aws:iam::*:policy/*cleanrooms*"
},
{
 "Sid": "ConsoleDisplayTables",
 "Effect": "Allow",
 "Action": [
  "glue:GetDatabase",
  "glue:GetDatabases",
  "glue:GetTable",
  "glue:GetTables",
  "glue:GetPartition",
```

```
    "glue:GetPartitions",
    "glue:GetSchema",
    "glue:GetSchemaVersion",
    "glue:BatchGetPartition"
   ],
   "Resource": "*"
  },
  {
   "Sid": "ConsolePickQueryResultsBucketListAll",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": "*"
  },
  {
   "Sid": "SetQueryResultsBucket",
   "Effect": "Allow",
   "Action": [
    "s3:GetBucketLocation",
    "s3:ListBucketVersions"
   ],
   "Resource": "arn:aws:s3:::cleanrooms-queryresults*"
  },
  {
   "Sid": "WriteQueryResults",
   "Effect": "Allow",
   "Action": [
    "s3:ListBucket",
    "s3:PutObject"
   ],
   "Resource": "arn:aws:s3:::cleanrooms-queryresults*",
   "Condition": {
    "ForAnyValue:StringEquals": {
     "aws:CalledVia": "cleanrooms.amazonaws.com"
    }
   }
  },
  {
   "Sid": "ConsoleDisplayQueryResults",
   "Effect": "Allow",
   "Action": [
    "s3:GetObject"
   ],
```

```
    "Resource": "arn:aws:s3:::cleanrooms-queryresults*"
   },
   {
    "Sid": "EstablishLogDeliveries",
    "Effect": "Allow",
    "Action": [
     "logs:CreateLogDelivery",
     "logs:GetLogDelivery",
     "logs:UpdateLogDelivery",
     "logs:DeleteLogDelivery",
     "logs:ListLogDeliveries"
    ],
    "Resource": "*",
    "Condition": {
     "ForAnyValue:StringEquals": {
      "aws:CalledVia": "cleanrooms.amazonaws.com"
     }
    }
   },
   {
    "Sid": "SetupLogGroupsDescribe",
    "Effect": "Allow",
    "Action": [
     "logs:DescribeLogGroups"
    ],
    "Resource": "*",
    "Condition": {
     "ForAnyValue:StringEquals": {
      "aws:CalledVia": "cleanrooms.amazonaws.com"
     }
    }
   },
   {
    "Sid": "SetupLogGroupsCreate",
    "Effect": "Allow",
    "Action": [
     "logs:CreateLogGroup"
    ],
    "Resource": "arn:aws:logs:*:*:log-group:/aws/cleanrooms*",
    "Condition": {
     "ForAnyValue:StringEquals": {
      "aws:CalledVia": "cleanrooms.amazonaws.com"
     }
    }
```

```
  },
  {
   "Sid": "SetupLogGroupsResourcePolicy",
   "Effect": "Allow",
   "Action": [
     "logs:DescribeResourcePolicies",
    "logs:PutResourcePolicy"
   ],
   "Resource": "*",
   "Condition": {
    "ForAnyValue:StringEquals": {
     "aws:CalledVia": "cleanrooms.amazonaws.com"
    }
   }
  },
  {
   "Sid": "ConsoleLogSummaryQueryLogs",
   "Effect": "Allow",
   "Action": [
    "logs:StartQuery"
   ],
   "Resource": "arn:aws:logs:*:*:log-group:/aws/cleanrooms*"
  },
  {
   "Sid": "ConsoleLogSummaryObtainLogs",
   "Effect": "Allow",
   "Action": [
    "logs:GetQueryResults"
   ],
   "Resource": "*"
  }
 ]
}
```

AWS 托管策略：**AWSCleanRoomsFullAccessNoQuerying**

您可以将 AWSCleanRoomsFullAccessNoQuerying 附加到 IAM principals。

此策略授予管理权限，允许对 AWS Clean Rooms 协作中的资源和元数据进行完全访问（读取、写入和更新）。此策略不包括执行查询的权限。

权限详细信息

该策略包含以下权限：

- CleanRoomsAccess— 授予对所有资源执行所有操作的完全访问权限 AWS Clean Rooms，协作中查询除外。
- CleanRoomsNoQuerying - 明确拒绝 StartProtectedQuery 和 UpdateProtectedQuery，阻止查询。
- PassServiceRole - 仅授予将服务角色传递给名称中带有“cleanrooms”的服务（PassedToService 条件）的访问权限。
- ListRolesToPickServiceRole— 允许委托人列出其所有角色以便在使用 AWS Clean Rooms时选择服务角色。
- GetRoleAndListRolePoliciesToInspectServiceRole - 允许主体在 IAM 中查看服务角色和相应的策略。
- ListPoliciesToInspectServiceRolePolicy - 允许主体在 IAM 中查看服务角色和相应的策略。
- GetPolicyToInspectServiceRolePolicy - 允许主体在 IAM 中查看服务角色和相应的策略。
- ConsoleDisplayTables— 允许委托人对在控制台上显示有关基础 AWS Glue 表的数据所需的 AWS Glue 元数据的只读访问权限。
- EstablishLogDeliveries— 允许委托人将查询日志传送到客户的 Amazon Lo CloudWatch gs 日志组。
- SetupLogGroupsDescribe— 允许委托人使用 Amazon Logs CloudWatch 日志组的创建流程。
- SetupLogGroupsCreate— 允许委托人创建 Amazon CloudWatch 日志组。
- SetupLogGroupsResourcePolicy— 允许委托人在 Amazon Logs CloudWatch 日志组上设置资源策略。
- ConsoleLogSummaryQueryLogs - 允许主体查看查询日志。
- ConsoleLogSummaryObtainLogs - 允许主体检索日志结果。
- cleanrooms— 管理服务中的协作、分析模板、配置表、成员资格和关联资源。 AWS Clean Rooms 执行各种操作，例如创建、更新、删除、列出和检索有关这些资源的信息。
- iam— 将名称包含 “cleanrooms” 的服务角色传递给 AWS Clean Rooms 服务。列出角色、策略，并检查服务角色和与 AWS Clean Rooms 服务相关的策略。
- glue— 从中检索有关数据库、表、分区和架构的信息。 AWS Glue这是 AWS Clean Rooms 服务显示底层数据源并与之交互所必需的。
- logs— 管理日志传送、日志组和 CloudWatch 日志资源策略。查询和检索与 AWS Clean Rooms 服务相关的日志。这些权限对于在服务中进行监控、审计和故障排除是必需的。

该政策还明确拒绝这些操作，cleanrooms:StartProtectedQuerycleanrooms:UpdateProtectedQuery并防止用户直接执行或更新受保护的查询，这应通过 AWS Clean Rooms 受控机制完成。

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "CleanRoomsAccess",
   "Effect": "Allow",
   "Action": [
    "cleanrooms:BatchGetCollaborationAnalysisTemplate",
    "cleanrooms:BatchGetSchema",
    "cleanrooms:BatchGetSchemaAnalysisRule",
    "cleanrooms:CreateAnalysisTemplate",
    "cleanrooms:CreateCollaboration",
    "cleanrooms:CreateConfiguredTable",
    "cleanrooms:CreateConfiguredTableAnalysisRule",
    "cleanrooms:CreateConfiguredTableAssociation",
    "cleanrooms:CreateMembership",
    "cleanrooms:DeleteAnalysisTemplate",
    "cleanrooms:DeleteCollaboration",
    "cleanrooms:DeleteConfiguredTable",
    "cleanrooms:DeleteConfiguredTableAnalysisRule",
    "cleanrooms:DeleteConfiguredTableAssociation",
    "cleanrooms:DeleteMember",
    "cleanrooms:DeleteMembership",
    "cleanrooms:GetAnalysisTemplate",
    "cleanrooms:GetCollaboration",
    "cleanrooms:GetCollaborationAnalysisTemplate",
    "cleanrooms:GetConfiguredTable",
    "cleanrooms:GetConfiguredTableAnalysisRule",
    "cleanrooms:GetConfiguredTableAssociation",
    "cleanrooms:GetMembership",
    "cleanrooms:GetProtectedQuery",
    "cleanrooms:GetSchema",
    "cleanrooms:GetSchemaAnalysisRule",
    "cleanrooms:ListAnalysisTemplates",
    "cleanrooms:ListCollaborationAnalysisTemplates",
    "cleanrooms:ListCollaborations",
    "cleanrooms:ListConfiguredTableAssociations",
    "cleanrooms:ListConfiguredTables",
    "cleanrooms:ListMembers",
```

```
    "cleanrooms:ListMemberships",
    "cleanrooms:ListProtectedQueries",
    "cleanrooms:ListSchemas",
    "cleanrooms:UpdateAnalysisTemplate",
    "cleanrooms:UpdateCollaboration",
    "cleanrooms:UpdateConfiguredTable",
    "cleanrooms:UpdateConfiguredTableAnalysisRule",
    "cleanrooms:UpdateConfiguredTableAssociation",
    "cleanrooms:UpdateMembership",
    "cleanrooms:ListTagsForResource",
    "cleanrooms:UntagResource",
    "cleanrooms:TagResource"
   ],
   "Resource": "*"
  },
  {
   "Sid": "CleanRoomsNoQuerying",
   "Effect": "Deny",
   "Action": [
    "cleanrooms:StartProtectedQuery",
    "cleanrooms:UpdateProtectedQuery"
   ],
   "Resource": "*"
  },
  {
   "Sid": "PassServiceRole",
   "Effect": "Allow",
   "Action": [
    "iam:PassRole"
   ],
   "Resource": "arn:aws:iam::*:role/service-role/*cleanrooms*",
   "Condition": {
    "StringEquals": {
     "iam:PassedToService": "cleanrooms.amazonaws.com"
    }
   }
  },
  {
   "Sid": "ListRolesToPickServiceRole",
   "Effect": "Allow",
   "Action": [
    "iam:ListRoles"
   ],
   "Resource": "*"
```

```
  },
  {
   "Sid": "GetRoleAndListRolePoliciesToInspectServiceRole",
   "Effect": "Allow",
   "Action": [
    "iam:GetRole",
    "iam:ListRolePolicies",
    "iam:ListAttachedRolePolicies"
   ],
   "Resource": "arn:aws:iam::*:role/service-role/*cleanrooms*"
  },
  {
   "Sid": "ListPoliciesToInspectServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "iam:ListPolicies"
   ],
   "Resource": "*"
  },
  {
   "Sid": "GetPolicyToInspectServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "iam:GetPolicy",
    "iam:GetPolicyVersion"
   ],
   "Resource": "arn:aws:iam::*:policy/*cleanrooms*"
  },
  {
   "Sid": "ConsoleDisplayTables",
   "Effect": "Allow",
   "Action": [
    "glue:GetDatabase",
    "glue:GetDatabases",
    "glue:GetTable",
    "glue:GetTables",
    "glue:GetPartition",
    "glue:GetPartitions",
    "glue:GetSchema",
    "glue:GetSchemaVersion",
    "glue:BatchGetPartition"
   ],
   "Resource": "*"
  },
```

```
  {
   "Sid": "EstablishLogDeliveries",
   "Effect": "Allow",
   "Action": [
    "logs:CreateLogDelivery",
    "logs:GetLogDelivery",
    "logs:UpdateLogDelivery",
    "logs:DeleteLogDelivery",
    "logs:ListLogDeliveries"
   ],
   "Resource": "*",
   "Condition": {
    "ForAnyValue:StringEquals": {
     "aws:CalledVia": "cleanrooms.amazonaws.com"
    }
   }
  },
  {
   "Sid": "SetupLogGroupsDescribe",
   "Effect": "Allow",
   "Action": [
    "logs:DescribeLogGroups"
   ],
   "Resource": "*",
   "Condition": {
    "ForAnyValue:StringEquals": {
     "aws:CalledVia": "cleanrooms.amazonaws.com"
    }
   }
  },
  {
   "Sid": "SetupLogGroupsCreate",
   "Effect": "Allow",
   "Action": [
    "logs:CreateLogGroup"
   ],
   "Resource": "arn:aws:logs:*:*:log-group:/aws/cleanrooms*",
   "Condition": {
    "ForAnyValue:StringEquals": {
     "aws:CalledVia": "cleanrooms.amazonaws.com"
    }
   }
  },
  {
```

```
    "Sid": "SetupLogGroupsResourcePolicy",
    "Effect": "Allow",
    "Action": [
      "logs:DescribeResourcePolicies",
     "logs:PutResourcePolicy"
    ],
    "Resource": "*",
    "Condition": {
     "ForAnyValue:StringEquals": {
      "aws:CalledVia": "cleanrooms.amazonaws.com"
     }
    }
   },
   {
    "Sid": "ConsoleLogSummaryQueryLogs",
    "Effect": "Allow",
    "Action": [
     "logs:StartQuery"
    ],
    "Resource": "arn:aws:logs:*:*:log-group:/aws/cleanrooms*"
   },
   {
    "Sid": "ConsoleLogSummaryObtainLogs",
    "Effect": "Allow",
    "Action": [
     "logs:GetQueryResults"
    ],
    "Resource": "*"
   }
  ]
}
```

AWS 托管策略：**AWSCleanRoomsMLReadOnlyAccess**

您可以将 AWSCleanRoomsMLReadOnlyAccess 附加到 IAM 主体。

该策略授予 AWSCleanRoomsMLReadOnlyAccess 协作中的资源和元数据的只读权限。

该策略包含以下权限：

- CleanRoomsConsoleNavigation— 授予查看 AWS Clean Rooms 控制台屏幕的权限。
- CleanRoomsMLRead— 允许委托人以只读方式访问 Clean Rooms 机器学习服务。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "CleanRoomsConsoleNavigation",
            "Effect": "Allow",
            "Action": [
                "cleanrooms:GetCollaboration",
                "cleanrooms:GetConfiguredAudienceModelAssociation",
                "cleanrooms:GetMembership",
                "cleanrooms:ListAnalysisTemplates",
                "cleanrooms:ListCollaborationAnalysisTemplates",
                "cleanrooms:ListCollaborationConfiguredAudienceModelAssociations",
                "cleanrooms:ListCollaborations",
                "cleanrooms:ListConfiguredTableAssociations",
                "cleanrooms:ListConfiguredTables",
                "cleanrooms:ListMembers",
                "cleanrooms:ListMemberships",
                "cleanrooms:ListProtectedQueries",
                "cleanrooms:ListSchemas",
                "cleanrooms:ListTagsForResource"
            ],
            "Resource": "*"
        },
        {
            "Sid": "CleanRoomsMLRead",
            "Effect": "Allow",
            "Action": [
                "cleanrooms-ml:Get*",
                "cleanrooms-ml:List*"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS 托管策略：**AWSCleanRoomsMLFullAccess**

您可以将 AWSCleanRoomsMLFullAcces 附加到 IAM 主体。此策略授予管理权限，允许对 Clean Rooms ML 所需的资源和元数据进行完全访问权限（读取、写入和更新）。

权限详细信息

该策略包含以下权限：

- CleanRoomsMLFullAccess— 授予对所有 Clean Rooms 机器学习操作的访问权限。
- PassServiceRole - 仅授予将服务角色传递给名称中带有“cleanrooms-ml”的服务（PassedToService 条件）的访问权限。
- CleanRoomsConsoleNavigation— 授予查看 AWS Clean Rooms 控制台屏幕的权限。
- CollaborationMembershipCheck— 当您在协作中启动受众生成（相似区段）工作时，Clean Rooms ML 服务会调用ListMembers以检查协作是否有效，来电者是否为活跃成员，配置的受众模型所有者是否为活跃成员。始终需要该权限；仅控制台用户需要控制台导航 SID。
- AssociateModels— 允许负责人将 Clean Rooms 机器学习模型与您的协作关联起来。
- TagAssociations - 允许主体将标签添加到相似模型和协作之间的关联中。
- ListRolesToPickServiceRole— 允许委托人列出其所有角色以便在使用 AWS Clean Rooms时选择服务角色。
- GetRoleAndListRolePoliciesToInspectServiceRole - 允许主体在 IAM 中查看服务角色和相应的策略。
- ListPoliciesToInspectServiceRolePolicy - 允许主体在 IAM 中查看服务角色和相应的策略。
- GetPolicyToInspectServiceRolePolicy - 允许主体在 IAM 中查看服务角色和相应的策略。
- ConsoleDisplayTables— 允许委托人对在控制台上显示有关基础 AWS Glue 表的数据所需的 AWS Glue 元数据的只读访问权限。
- ConsolePickOutputBucket - 允许主体为配置的受众模型输出选择 Amazon S3 存储桶。
- ConsolePickS3Location - 允许主体为配置的受众模型输出选择存储桶中的位置。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "CleanRoomsMLFullAccess",
            "Effect": "Allow",
            "Action": [
                "cleanrooms-ml:*"
            ],
            "Resource": "*"
        },
        {
            "Sid": "PassServiceRole",
```

```
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::*:role/cleanrooms-ml*"
            ],
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": "cleanrooms-ml.amazonaws.com"
                }
            }
        },
        {
            "Sid": "CleanRoomsConsoleNavigation",
            "Effect": "Allow",
            "Action": [
                "cleanrooms:GetCollaboration",
                "cleanrooms:GetConfiguredAudienceModelAssociation",
                "cleanrooms:GetMembership",
                "cleanrooms:ListAnalysisTemplates",
                "cleanrooms:ListCollaborationAnalysisTemplates",
                "cleanrooms:ListCollaborationConfiguredAudienceModelAssociations",
                "cleanrooms:ListCollaborations",
                "cleanrooms:ListConfiguredTableAssociations",
                "cleanrooms:ListConfiguredTables",
                "cleanrooms:ListMembers",
                "cleanrooms:ListMemberships",
                "cleanrooms:ListProtectedQueries",
                "cleanrooms:ListSchemas",
                "cleanrooms:ListTagsForResource"
            ],
            "Resource": "*"
        },
        {
            "Sid": "CollaborationMembershipCheck",
            "Effect": "Allow",
            "Action": [
                "cleanrooms:ListMembers"
            ],
            "Resource": "*",
            "Condition": {
                "ForAnyValue:StringEquals": {
                    "aws:CalledVia": ["cleanrooms-ml.amazonaws.com"]
```

```
                }
            }
        },
        {
            "Sid": "AssociateModels",
            "Effect": "Allow",
            "Action": [
                "cleanrooms:CreateConfiguredAudienceModelAssociation"
            ],
            "Resource": "*"
        },
        {
            "Sid": "TagAssociations",
            "Effect": "Allow",
            "Action": [
                "cleanrooms:TagResource"
            ],
            "Resource": "arn:aws:cleanrooms:*:*:membership/*/
configuredaudiencemodelassociation/*"
        },
        {
            "Sid": "ListRolesToPickServiceRole",
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        },
        {
            "Sid": "GetRoleAndListRolePoliciesToInspectServiceRole",
            "Effect": "Allow",
            "Action": [
                "iam:GetRole",
                "iam:ListRolePolicies",
                "iam:ListAttachedRolePolicies"
            ],
            "Resource": [
                "arn:aws:iam::*:role/service-role/cleanrooms-ml*",
                "arn:aws:iam::*:role/role/cleanrooms-ml*"
            ]
        },
        {
            "Sid": "ListPoliciesToInspectServiceRolePolicy",
            "Effect": "Allow",
```

```
            "Action": [
                "iam:ListPolicies"
            ],
            "Resource": "*"
        },
        {
            "Sid": "GetPolicyToInspectServiceRolePolicy",
            "Effect": "Allow",
            "Action": [
                "iam:GetPolicy",
                "iam:GetPolicyVersion"
            ],
            "Resource": "arn:aws:iam::*:policy/*cleanroomsml*"
        },
        {
            "Sid": "ConsoleDisplayTables",
            "Effect": "Allow",
            "Action": [
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:GetSchema",
                "glue:GetSchemaVersion",
                "glue:BatchGetPartition"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ConsolePickOutputBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListAllMyBuckets"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ConsolePickS3Location",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket",
                "s3:GetBucketLocation"
```

```
            ],
            "Resource": "arn:aws:s3:::*cleanrooms-ml*"
        }
    ]
}
```

AWS Clean RoomsAWS 托管策略的更新

查看 AWS Clean Rooms 自该服务开始跟踪这些更改以来 AWS 托管策略更新的详细信息。要获得有关此页面变更的自动提醒，请订阅 “ AWS Clean Rooms 文档历史记录” 页面上的 RSS feed。

更改	描述	日期
AWSCleanRoomsFullAccessNoQuerying – 对现有策略的更新	已将 cleanrooms:BatchGetSchemaAnalysisRule 添加到 CleanRoomsAccess。	2024年5月13日
AWSCleanRoomsFullAccess - 对现有策略的更新	将此策略中的 Stat AWSCleanRoomsFullAccess ement ID 从ConsolePickQueryResultsBucket更新为以更好地表示权限，因为无论使用控制台还是不使用控制台，都需要这些权限来设置查询结果存储桶。SetQueryResultsBucket	2024 年 3 月 21 日
AWSCleanRoomsMLReadOnlyAccess - 新策略 AWSCleanRoomsMLFullAccess - 新策略	添加了 AWSCleanRoomsMLReadOnlyAccess 和 AWSCleanRoomsMLFullAccess 以支持 AWS Clean Rooms 机器学习。	2023 年 11 月 29 日
AWSCleanRoomsFullAccessNoQuerying - 对现有策略的更新	添加了cleanrooms:CreateAnalysisTemplatecleanrooms:GetAnalysisTemplate、cleanrooms:UpdateAnalysisTemplate、 cleanrooms:DeleteAnalysisTemplate、cleanrooms:ListAnalysisTemplates、cleanrooms:GetCollaborationAnalysisTemplate、cleanrooms:BatchGetCollaborationAnalysis	2023 年 7 月 31 日

更改	描述	日期
	Template、和，cleanrooms:ListCollaborationAnalysisTemplatesCleanRoomsAccess以启用新的分析模板功能。	
AWSCleanRoomsFullAccessNoQuerying - 对现有策略的更新	向 CleanRoomsAccess 添加了 cleanrooms:ListTagsForResource、cleanrooms:UntagResource 和 cleanrooms:TagResource，以启用资源标记。	2023 年 3 月 21 日
AWS Clean Rooms 开始跟踪更改	AWS Clean Rooms 开始跟踪其 AWS 托管策略的更改。	2023 年 1 月 12 日

对 AWS Clean Rooms 身份和访问进行故障排除

使用以下信息来帮助您诊断和修复在使用 AWS Clean Rooms 和 IAM 时可能遇到的常见问题。

主题

- 我无权在以下位置执行操作 AWS Clean Rooms
- 我无权执行 iam：PassRole
- 我想允许我以外的人 AWS 账户 访问我的 AWS Clean Rooms 资源

我无权在以下位置执行操作 AWS Clean Rooms

如果您收到错误提示，表明您无权执行某个操作，则您必须更新策略以允许执行该操作。

当 `mateojackson` IAM 用户尝试使用控制台查看有关虚构 *`my-example-widget`* 资源的详细信息，但不拥有虚构 `cleanrooms:`*`GetWidget`* 权限时，会发生以下示例错误。

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 cleanrooms:GetWidget on resource: my-example-widget
```

在此情况下，Mateo 的策略必须更新以允许其使用 `cleanrooms:`*`GetWidget`* 操作访问 *`my-example-widget`* 资源。

如果您需要帮助，请联系您的 AWS 管理员。您的管理员是提供登录凭证的人。

我无权执行 iam：PassRole

如果您收到一个错误，表明您无权执行 `iam:PassRole` 操作，则必须更新策略以允许您将角色传递给 AWS Clean Rooms。

有些 AWS 服务 允许您将现有角色传递给该服务，而不是创建新的服务角色或服务相关角色。为此，您必须具有将角色传递到服务的权限。

当名为 `marymajor` 的 IAM 用户尝试使用控制台在 AWS Clean Rooms中执行操作时，会发生以下示例错误。但是，服务必须具有服务角色所授予的权限才可执行此操作。Mary 不具有将角色传递到服务的权限。

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

在这种情况下，必须更新 Mary 的策略以允许她执行 `iam:PassRole` 操作。

如果您需要帮助，请联系您的 AWS 管理员。您的管理员是提供登录凭证的人。

我想允许我以外的人 AWS 账户 访问我的 AWS Clean Rooms 资源

您可以创建一个角色，以便其他账户中的用户或您组织外的人员可以使用该角色来访问您的资源。您可以指定谁值得信赖，可以担任角色。

要了解更多信息，请参阅以下内容：

- 要了解是否 AWS Clean Rooms 支持这些功能，请参阅如何 AWS Clean Rooms 与 IAM 配合使用。
- 要了解如何提供对您拥有的资源的访问权限 AWS 账户 ，请参阅 IAM 用户指南中的向您拥有 AWS 账户 的另一个 IAM 用户提供访问权限。
- 要了解如何向第三方提供对您的资源的访问权限 AWS 账户，请参阅 IAM 用户指南中的向第三方提供访问权限。 AWS 账户
- 要了解如何通过联合身份验证提供访问权限，请参阅《IAM 用户指南》中的为经过外部身份验证的用户（联合身份验证）提供访问权限。
- 要了解使用角色和基于资源的策略进行跨账户存取之间的差别，请参阅《IAM 用户指南》中的 IAM 角色与基于资源的策略有何不同。

防止跨服务混淆代理

混淆代理问题是一个安全性问题，即不具有操作执行权限的实体可能会迫使具有更高权限的实体执行该操作。在中 AWS，跨服务模仿可能会导致混乱的副手问题。一个服务（呼叫服务）调用另一项服务（所谓的服务）时，可能会发生跨服务模拟。可以操纵调用服务，使用其权限以在其他情况下该服务不应有访问权限的方式对另一个客户的资源进行操作。为防止这种情况， AWS 提供可帮助您保护所有服务的数据的工具，而这些服务中的服务主体有权限访问账户中的资源。

我们建议在资源策略中使用 `aws:SourceArn` 全局条件上下文键，以限制 AWSClean Rooms 授予其他服务对资源的权限。如果您只希望将一个资源与跨服务访问相关联，请使用 `aws:SourceArn`。

防范混淆代理问题最有效的方法是使用 `aws:SourceArn` 全局条件上下文键和资源的完整 ARN。在中 AWSClean Rooms，您还必须与`sts:ExternalId`条件键进行比较。

`aws:SourceArn` 的值必须设置为所担任角色的成员身份的 ARN。

以下示例演示如何使用 AWSClean Rooms 中的 `aws:SourceArn` 全局条件上下文键来防范混淆代理问题。

> **Note**
>
> 示例策略适用于 AWSClean Rooms 用于访问客户数据的服务角色的信任策略。*`MembershipID`* 的值就是您在协作中的 AWSClean Rooms 成员身份 ID。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowIfExternalIdMatches",
            "Effect": "Allow",
            "Principal": {
                "Service": "cleanrooms.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringLike": {
                    "sts:ExternalId": "arn:aws:*:aws-region:*:dbuser:*/membershipID*"
                }
            }
        },
```

```
        {
            "Sid": "AllowIfSourceArnMatches",
            "Effect": "Allow",
            "Principal": {
                "Service": "cleanrooms.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "ForAnyValue:ArnEquals": {
                    "aws:SourceArn": "arn:aws:cleanrooms:aws-
region:123456789012:membership/membershipID"
                }
            }
        }
    ]
}
```

AWS Clean Rooms 机器学习的 IAM 行为

跨账户作业

Clean Rooms ML AWS 账户 允许一个人在其帐户中安全地访问另一个人创建的某些资源 AWS 账户。当 AWS 账户 A 中的客户端调用 AWS 账户 B 拥有StartAudienceGenerationJob的ConfiguredAudienceModel资源时，Clean Rooms ML 会为该任务创建两个 ARN。一个 ARN 在 AWS 账户 A 中，另一个 ARN 在 B 中 AWS 账户 除了它们之外，ARN 完全相同。 AWS 账户

Clean Rooms ML 为任务创建两个 ARN，以确保两个账户都可以将自己的 IAM 策略应用于任务。例如，两个账户都可以使用基于标签的访问控制并应用其 AWS 组织的策略。作业处理来自两个账户的数据，因此，两个账户都可以删除作业及其关联数据。两个账户都不能阻止另一个账户删除作业。

只能执行一个作业，两个账户可以在调用 ListAudienceGenerationJobs 时看到该作业。两个账户都可以在工作中Get使用带有自己 AWS 账户 的 ID 的 ARN 调用Delete、和 Export API。

使用带有另一 AWS 账户 个 ID 的 ARN 时，两者都 AWS 账户 无法访问任务。

作业的名称在 AWS 账户中必须是唯一的。 AWS 账户 B 中的名字是 *$accounta-$* name。在 B 中查看作业时 AWS 账户 ， AWS 账户 A 选择的名称以 A 为前缀 AWS 账户

为了使跨账户StartAudienceGenerationJob成功， AWS 账户 B 必须使用类似于以下示例的资源策略允许对 AWS 账户 B 中的新任务和 AWS 账户 B ConfiguredAudienceModel 中的新任务执行该操作：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Clean-Rooms-<CAMA ID>",
            "Effect": "Allow",
            "Principal": {
                "AWS": [
                    "accountA"
                ]
            },
            "Action": [
                "cleanrooms-ml:StartAudienceGenerationJob"
            ],
            "Resource": [
                "arn:aws:cleanrooms-ml:us-west-1:AccountB:configured-audience-
model/id",
                "arn:aws:cleanrooms-ml:us-west-1:AccountB:audience-generation-job/*"
            ],
            // optional - always set by AWS Clean Rooms
"Condition":{"StringEquals":{"cleanrooms-ml:CollaborationId":"UUID"}}
        }
    ]
}
```

如果您使用 AWS Clean Rooms ML API 创建`manageResourcePolicies`设置为 true 的配置相似模型，则会为您 AWS Clean Rooms 创建此策略。

此外，AWS 账户 A 中来电者的身份策略需要获得`StartAudienceGenerationJob`许可`arn:aws:cleanrooms-ml:us-west-1:AccountA:audience-generation-job/*`。因此，有三个 IAM 资源可供操作`StartAudienceGenerationJob`：AWS 账户 A 作业、AWS 账户 B 作业和 AWS 账户 B `ConfiguredAudienceModel`。

> **Warning**
>
> 启动 AWS 账户 该作业的用户会收到有关该作业的 AWS CloudTrail 审核日志事件。拥有 `ConfiguredAudienceModel` 的 AWS 账户 不会收到 AWS CloudTrail 审核日志事件。

标记作业

在您设置 CreateConfiguredAudienceModel 的 childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE 参数时，您的账户中通过该配置的相似模型创建的所有相似细分生成作业默认具有与配置的相似模型相同的标签。配置的相似模型是父模型，相似细分生成作业是子模型。

如果您在自己的账户中创建作业，作业的请求标签将覆盖父标签。其他账户创建的作业绝不会在您的账户中创建标签。如果您设置 childResourceTagOnCreatePolicy=FROM_PARENT_RESOURCE 并且另一个账户创建作业，则作业具有两个副本。您的账户中的副本具有父资源标签，作业提交者账户中的副本具有来自请求的标签。

验证协作者

向 AWS Clean Rooms 协作中的其他成员授予权限时，资源策略应包含条件键cleanrooms-ml:CollaborationId。这会强制要求collaborationId参数包含在StartAudienceGenerationJob请求中。当collaborationId参数包含在请求中时，Clean Rooms ML 会验证协作是否存在，作业提交者是协作的活跃成员，配置的相似模型所有者是协作的活跃成员。

AWS Clean Rooms 管理您配置的相似模型资源策略（manageResourcePolicies参数在CreateConfiguredAudienceModelAssociation 请求TRUE中）时，将在资源策略中设置此条件密钥。因此，必须在 in collaborationId 中指定StartAudienceGenerationJob。

跨账户存取

只能跨账户调用 StartAudienceGenerationJob。所有其他 Clean Rooms ML API 只能与您自己账户中的资源一起使用。这可确保您的训练数据、相似模型配置和其他信息保持私密。

Clean Rooms ML 永远不会透露 Amazon S3 或各个账户 AWS Glue 的位置。训练数据位置、配置的相似模型输出位置和相似细分生成作业种子位置绝不会在账户之间可见。如果您获取 (Get) 另一个账户提交的受众生成作业，该服务不会显示种子位置。

合规性验证 AWS Clean Rooms

要了解是否属于特定合规计划的范围，请参阅AWS 服务 “按合规计划划分的范围” ”，然后选择您感兴趣的合规计划。 AWS 服务 有关一般信息，请参阅AWS 合规计划AWS。

您可以使用下载第三方审计报告 AWS Artifact。有关更多信息，请参阅中的 “下载报告” 中的 “ AWS Artifact。

您在使用 AWS 服务 时的合规责任取决于您的数据的敏感性、贵公司的合规目标以及适用的法律和法规。 AWS 提供了以下资源来帮助实现合规性：

- 安全与合规性快速入门指南 — 这些部署指南讨论了架构注意事项，并提供了在这些基础上 AWS 部署以安全性和合规性为重点的基准环境的步骤。
- 在 A@@ mazon Web Services 上构建 HIPAA 安全与合规性 — 本白皮书描述了各公司如何使用 AWS 来创建符合 HIPAA 资格的应用程序。

> **Note**
>
> 并非所有 AWS 服务 人都符合 HIPAA 资格。有关更多信息，请参阅符合 HIPAA 要求的服务参考。

- AWS 合规资源AWS — 此工作簿和指南集可能适用于您所在的行业和所在地区。
- AWS 客户合规指南 — 从合规角度了解责任共担模式。这些指南总结了保护的最佳实践， AWS 服务并将指南映射到跨多个框架（包括美国国家标准与技术研究院 (NIST)、支付卡行业安全标准委员会 (PCI) 和国际标准化组织 (ISO)）的安全控制。
- 使用AWS Config 开发人员指南中的规则评估资源 — 该 AWS Config 服务评估您的资源配置在多大程度上符合内部实践、行业准则和法规。
- AWS Security Hub— 这 AWS 服务 提供了您内部安全状态的全面视图 AWS。Security Hub 通过安全控件评估您的 AWS 资源并检查其是否符合安全行业标准和最佳实践。有关受支持服务及控件的列表，请参阅 Security Hub 控件参考。
- Amazon GuardDuty — 它通过监控您的 AWS 账户环境中是否存在可疑和恶意活动，来 AWS 服务检测您的工作负载、容器和数据面临的潜在威胁。 GuardDuty 通过满足某些合规性框架规定的入侵检测要求，可以帮助您满足各种合规性要求，例如 PCI DSS。
- AWS Audit Manager— 这 AWS 服务 可以帮助您持续审计 AWS 使用情况，从而简化风险管理以及对法规和行业标准的合规性。

韧性在 AWS Clean Rooms

AWS 全球基础设施是围绕 AWS 区域和可用区构建的。区域提供多个在物理上独立且隔离的可用区，这些可用区通过延迟低、吞吐量高且冗余性高的网络连接在一起。利用可用区，您可以设计和操作在可用区之间无中断地自动实现故障转移的应用程序和数据库。与传统的单个或多个数据中心基础设施相比，可用区具有更高的可用性、容错性和可扩展性。

有关 AWS 区域和可用区的更多信息，请参阅AWS 全球基础设施。

中的基础设施安全 AWS Clean Rooms

作为一项托管服务 AWS Clean Rooms ，受 AWS 全球网络安全的保护。有关 AWS 安全服务以及如何 AWS 保护基础设施的信息，请参阅AWS 云安全。要使用基础设施安全的最佳实践来设计您的 AWS 环境，请参阅 S AWS ecurity Pillar Well-Architected Fram ework 中的基础设施保护。

您可以使用 AWS 已发布的 API 调用 AWS Clean Rooms 通过网络进行访问。客户端必须支持以下内容：

- 传输层安全性协议 (TLS) 我们要求使用 TLS 1.2，建议使用 TLS 1.3。
- 具有完全向前保密 (PFS) 的密码套件，例如 DHE（临时 Diffie-Hellman）或 ECDHE（临时椭圆曲线 Diffie-Hellman）。大多数现代系统（如 Java 7 及更高版本）都支持这些模式。

此外，必须使用访问密钥 ID 和与 IAM 主体关联的秘密访问密钥来对请求进行签名。或者，您可以使用 AWS Security Token Service（AWS STS）生成临时安全凭证来对请求进行签名。

网络安全

在查询执行期间从 S3 存储桶 AWS Clean Rooms 读取数据时，与 Amazon S3 AWS Clean Rooms 之间的流量将通过 AWS 私有网络安全路由。使用 Amazon 签名版本 4 协议 (SIGv4) 签署正在传输的流量并使用 HTTPS 对该流量加密。此流量根据您为配置表设置的 IAM 服务角色进行授权。

您可以 AWS Clean Rooms 通过终端节点以编程方式连接到。有关服务端点的列表，请参阅《AWS 一般参考》中的 AWS Clean Rooms 端点和配额。

所有服务端点都只支持 HTTPS。如果您想从 VPC 连接但又不想连接互联 AWS Clean Rooms 网，则可以使用亚马逊虚拟私有云 (VPC) 终端节点。有关更多信息，请参阅AWS PrivateLink 指南 AWS PrivateLink中的通过访问 AWS 服务。

您可以为您的 IAM 委托人分配 IAM 策略，这些委托人使用 aws: SourceVpce 上下文密钥来限制您的 IAM 委托人只能通过 VPC 终端节点进行调用，而不能 AWS Clean Rooms 通过互联网进行调用。

使用接口端点进行访问 AWS Clean Rooms 或 AWS Clean Rooms ML (AWS PrivateLink)

您可以使用 AWS PrivateLink 在您的虚拟私有云 (VPC) 和/ AWS Clean Rooms 或 AWS Clean Rooms ML 之间创建私有连接。您可以像在 VPC 中一样访问 AWS Clean Rooms 或 AWS Clean Rooms ML，

无需使用互联网网关、NAT 设备、VPN 连接或 AWS Direct Connect 连接。VPC 中的实例不需要公有 IP 地址即可访问 AWS Clean Rooms。

您可以通过创建由 AWS PrivateLink提供支持的接口端点来建立此私有连接。我们将在您为接口端点启用的每个子网中创建一个端点网络接口。这些是请求者托管的网络接口，用作发往 AWS Clean Rooms 的流量的入口点。

有关更多信息，请参阅《AWS PrivateLink 指南》中的通过 AWS PrivateLink访问 AWS 服务。

的注意事项 AWS Clean Rooms

在为设置接口终端节点之前 AWS Clean Rooms，请查看AWS PrivateLink 指南中的注意事项。

AWS Clean Rooms 而且 AWS Clean Rooms ML 支持通过接口端点调用其所有 API 操作。

AWS Clean Rooms 或 AWS Clean Rooms ML 不支持 VPC 终端节点策略。默认情况下，允许通过接口端点对 AWS Clean Rooms 和 AWS Clean Rooms ML 进行完全访问。或者，您可以将安全组与端点网络接口关联，以控制通过接口终端节点 AWS Clean Rooms 传入或 AWS Clean Rooms 机器学习的流量。

为创建接口终端节点 AWS Clean Rooms

您可以使用 Amazon VPC 控制台 AWS Clean Rooms 或 AWS Command Line Interface (AWS CLI) 为或 AWS Clean Rooms ML 创建接口终端节点。有关更多信息，请参阅《AWS PrivateLink 指南》中的创建接口端点。

AWS Clean Rooms 使用以下服务名称创建接口终端节点。

```
com.amazonaws.region.cleanrooms
```

使用以下服务名称为 AWS Clean Rooms ML 创建接口终端节点。

```
com.amazonaws.region.cleanrooms-ml
```

如果为接口端点启用私有 DNS，则可使用其默认区域 DNS 名称向 AWS Clean Rooms 发出 API 请求。例如，`cleanrooms-ml.us-east-1.amazonaws.com`。

监控 AWS Clean Rooms

监控是维护和其他 AWS 解决方案的可靠性、可用性和性能的重要组成部分。 AWS Clean Rooms AWS 提供以下监控工具 AWS Clean Rooms，供您监视、报告问题并在适当时自动采取措施：

- Amazon CloudWatch Logs 允许您监控、存储和访问来自 Amazon EC2 实例和其他来源的日志文件。 AWS CloudTrail Amazon Log CloudWatch s 可以监控日志文件中的信息，并在达到特定阈值时通知您。您还可以在高持久性存储中检索您的日志数据。有关更多信息，请参阅 Amazon CloudWatch 日志用户指南。

  Clean Rooms ML 允许跨账户任务执行某些 API 操作。启动 AWS 账户 该作业的将收到该作业的 AWS CloudTrail 审核日志事件。有关更多信息，请参阅 AWS Clean Rooms 机器学习的 IAM 行为。
- AWS CloudTrail捕获由您或代表您发起的 API 调用和相关事件， AWS 账户 并将日志文件传输到您指定的 Amazon S3 存储桶。您可以识别哪些用户和帐户拨打了电话 AWS、发出呼叫的源 IP 地址以及呼叫发生的时间。有关更多信息，请参阅 AWS CloudTrail 用户指南。

使用 AWS CloudTrail 记录 AWS Clean Rooms API 调用

AWS Clean Rooms 与 AWS CloudTrail 集成，后者是在 AWS Clean Rooms 中记录用户、角色或 AWS 服务 服务所执行操作的服务。CloudTrail 将 AWS Clean Rooms 的所有 API 调用作为事件捕获。捕获的调用包含来自 AWS Clean Rooms 控制台和代码的 AWS Clean Rooms API 操作调用。如果您创建跟踪，则可以使 CloudTrail 事件持续传送到 Amazon S3 存储桶（包括 AWS Clean Rooms 的事件）。如果您不配置跟踪记录，则仍可在 CloudTrail 控制台中的 Event history（事件历史记录）中查看最新事件。使用 CloudTrail 收集的信息，您可以确定向 AWS Clean Rooms 发出了什么请求、发出请求的 IP 地址、何人发出的请求、请求的发出时间以及其他详细信息。

要了解有关 CloudTrail 的更多信息，请参阅《AWS CloudTrail 用户指南》。

CloudTrail 中的 AWS Clean Rooms 信息

在您创建 AWS 账户 时，将在该账户上启用 CloudTrail。当 AWS Clean Rooms 中发生活动时，该活动将记录在 CloudTrail 事件中，并与其他 AWS 服务 事件一同保存在事件历史记录中。您可以在 AWS 账户中查看、搜索和下载最新事件。有关更多信息，请参阅使用 CloudTrail 事件历史记录查看事件。

对于AWS 账户中的事件的持续记录（包括 AWS Clean Rooms 的事件），请创建跟踪记录。通过跟踪记录，CloudTrail 可将日志文件传送至 Simple Storage Service（Amazon S3）存储桶。预设情况下，

在控制台中创建跟踪记录时，此跟踪记录应用于所有AWS 区域。此跟踪记录在 AWS 分区中记录所有区域中的事件，并将日志文件传送至您指定的 Simple Storage Service（Amazon S3）桶。此外，您可以配置其他 AWS 服务，进一步分析在 CloudTrail 日志中收集的事件数据并采取行动。有关更多信息，请参阅下列内容：

- 创建跟踪概览
- CloudTrail 支持的服务和集成
- 为 CloudTrail 配置 Amazon SNS 通知
- 从多个区域接收 CloudTrail 日志文件
- 从多个账户接收 CloudTrail 日志文件

CloudTrail 记录所有 AWS Clean Rooms 操作，AWS Clean Rooms API 参考中介绍了这些操作。

每个事件或日志条目都包含有关生成请求的人员信息。身份信息可帮助您确定以下内容：

- 请求是使用根用户凭证还是 IAM 用户凭证发出的。
- 请求是使用角色还是联合身份用户的临时安全凭证发出的。
- 请求是否由其他 AWS 服务 发出。

有关更多信息，请参阅 CloudTrail userIdentity 元素。

了解 AWS Clean Rooms 日志文件条目

跟踪是一种配置，可用于将事件作为日志文件传送到您指定的 Amazon S3 桶。CloudTrail 日志文件包含一个或多个日志条目。一个事件表示来自任何源的一个请求，包括有关所请求的操作、操作的日期和时间、请求参数等方面的信息。CloudTrail 日志文件不是公用 API 调用的有序堆栈跟踪，因此它们不会按任何特定顺序显示。

AWS Clean Rooms CloudTrail 事件示例

以下示例展示了 CloudTrail 事件：

主题

- StartProtectedQuery（成功）
- StartProtectedQuery（失败）

StartProtectedQuery（成功）


```
{
    "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（失败）


```
{
    "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"
}
```

使用创建 AWS Clean Rooms 资源 AWS CloudFormation

AWS Clean Rooms 已与 AWS CloudFormation一项服务集成，该服务可帮助您对 AWS 资源进行建模和设置。通过这种集成，您可以花费更少的时间来创建和管理您的资源和基础设施。您可以创建一个描述所需所有 AWS 资源的模板，并为您预 AWS CloudFormation 置和配置这些资源。资源的示例包括协作、配置表、配置表关联和成员身份。

使用时 AWS CloudFormation，您可以重复使用模板来一致且重复地设置 AWS Clean Rooms 资源。描述一次您的资源，然后在多个 AWS 账户 和（或）中一遍又一遍地配置相同的资源 AWS 区域。

AWS Clean Rooms 和 AWS CloudFormation 模板

要为和相关服务配置 AWS Clean Rooms 和配置资源，必须了解AWS CloudFormation 模板。模板是 JSON 或 YAML 格式的文本文件。这些模板描述了您要在 AWS CloudFormation 堆栈中配置的资源。如果你不熟悉 JSON 或 YAML，可以使用 D AWS CloudFormation esigner 来帮助你开始使用 AWS CloudFormation 模板。有关更多信息，请参阅《AWS CloudFormation 用户指南》中的什么是 AWS CloudFormation Designer？。

AWS Clean Rooms 支持在中创建协作、配置的表、配置的表关联和成员资格。 AWS CloudFormation 有关更多信息（包括协作的 JSON 和 YAML 模板示例、配置表关联和成员身份），请参阅《AWS CloudFormation 用户指南》中的 AWS Clean Rooms 资源类型参考。

可用模板如下：

- 分析模板

  指定 AWS Clean Rooms 分析模板，包括名称、描述、格式、来源、参数和标签。

  有关更多信息，请参阅以下主题：

  《AWS Clean Rooms 用户指南》中的 AWS::CleanRooms::AnalysisTemplate

  《AWS Clean Rooms API 参考》中的 CreateAnalysisTemplate
- 协作

  指定 AWS Clean Rooms 协作，包括名称、描述、类型、参数和标签。

  有关更多信息，请参阅以下主题：

  《AWS CloudFormation 用户指南》中的 AWS::CleanRooms::Collaboration

《AWS Clean Rooms API 参考》中的 CreateCollaboration

- 配置表

  在中指定已配置的表 AWS Clean Rooms，包括允许的列、分析方法、描述、名称、表格引用、隐私预算和标签。已配置的表表示对中已配置为在中 AWS Glue Data Catalog 使用的现有表的引用 AWS Clean Rooms。配置表包含用于确定如何使用数据的分析规则。

  有关更多信息，请参阅以下主题：

  《AWS CloudFormation 用户指南》中的 AWS::CleanRooms::ConfiguredTable

  《AWS Clean Rooms API 参考》中的 CreateConfiguredTable

- 配置表关联

  在中指定已配置的表关联 AWS Clean Rooms，包括 ID、描述、会员 ID、名称、角色、Amazon 资源名称 (ARN) 和标签。配置表关联将配置表与协作关联。

  有关更多信息，请参阅以下主题：

  《AWS CloudFormation 用户指南》中的 AWS::CleanRooms::ConfiguredTableAssociation

  《AWS Clean Rooms API 参考》中的 CreateConfiguredTableAssociation

- 成员身份

  在 AWS Clean Rooms为特定协作标识符指定成员身份并加入协作。

  有关更多信息，请参阅以下主题：

  《AWS CloudFormation 用户指南》中的 AWS::CleanRooms::Membership

  《AWS Clean Rooms API 参考》中的 CreateMembership

- 隐私预算模板

  指定 AWS Clean Rooms 隐私预算模板，包括隐私预算、每次查询添加的噪音以及每月隐私预算刷新。

  有关更多信息，请参阅以下主题：

  《AWS CloudFormation 用户指南》中的 AWS::CleanRooms::PrivacyBudgetTemplate

《AWS Clean Rooms API 参考》中的 CreatePrivacyBudgetTemplate

- 创建训练数据集

  从 AWS Glue 表中为 Clean Rooms 机器学习模型指定训练数据集。

  有关更多信息，请参阅以下主题：

  《AWS CloudFormation 用户指南》中的 AWS::CleanRoomsML::TrainingDataset

  CreateTrainingDataset在 C lean Rooms 中 ML API 参考资料

了解更多关于 AWS CloudFormation

要了解更多信息 AWS CloudFormation，请参阅以下资源：

- AWS CloudFormation
- AWS CloudFormation 用户指南
- AWS CloudFormation API 引用
- AWS CloudFormation 命令行界面用户指南

的配额 AWS Clean Rooms

您的每个配额 AWS 账户 都有默认配额，以前称为限制 AWS 服务。除非另有说明，否则每个配额都特定于 AWS 区域。您可以请求增加某些配额，但其他一些配额无法增加。

要查看的配额 AWS Clean Rooms，请打开 S ervice Quotas 控制台。在导航窗格中，选择 AWS 服务，然后选择 AWS Clean Rooms。

要请求提高配额，请参阅《Service Quotas 用户指南》中的请求提高配额。如果配额在 Service Quotas 中尚不可用，请使用提高服务限额表单。

您的 AWS 账户 配额与以下有关 AWS Clean Rooms。

资源	默认值	描述
每个协作邀请的成员数	5	每个协作邀请的最大成员数
每个账户的成员数	100	一个账户的最大成员身份数
每个账户创建的协作数	10	每个账户创建的最大协作数
每个账户的配置表数	60	一个账户可以创建的最大已配置表数
每个成员资格的表关联数	25	每个活跃成员身份关联的最大表数
每个成员正在进行的并发查询数	5	每个成员正在进行的最大并发查询数
每个配置允许列表的列数	100	每个已配置表可以列入许可名单的最大列数
每个受保护的查询的配置表数	15	受保护查询中的最大已配置表数
每个成员身份的分析模板数	25	每个成员身份的最大分析模板数

资源	默认值	描述
为每个成员配置的相似模型（受众模型）关联	5	为每个成员配置的最大相似模型关联数量。

资源参数限制

资源	默认值	描述
分析规则大小	100KB	分析规则的最大 JSON 大小
查询文本长度	90 KB（8 KB 用于差别隐私查询）	SQL 查询语句的最大文本长度
查询运行时间	12 小时	查询在超时前运行的最长持续时间
查询数据文件输出大小	6.2 GB	受保护查询的输出文件的最大大小

AWS 账户 每个终端节点配额中每个账户的每秒 API 交易量 (TPS) 如下。

API 节流配额

资源	速率限制	描述
`BatchGetCollaborationAnalysisTemplate` 请求速率	5 TPS	每秒的最大 `BatchGetCollaborationAnalysisTemplate` API 调用数
`BatchGetSchema` 请求速率	5 TPS	每秒的最大 `BatchGetSchema` API 调用数
`CreateAnalysisTemplate` 请求速率	5 TPS	每秒的最大 `CreateAnalysisTemplate` API 调用数

资源	速率限制	描述
CreateCollaboration 请求速率	5 TPS	每秒的最大 CreateCollaboration API 调用数
CreateConfiguredAudienceModelAssociation 请求速率	5 TPS	每秒最大 CreateConfiguredAudienceModelAssociation 调用数
CreateConfiguredTable 请求速率	5 TPS	每秒最大 CreateConfiguredTable 调用数
CreateConfiguredTableAnalysisRule 请求速率	5 TPS	每秒最大 CreateConfiguredTableAnalysisRule 调用数
CreateConfiguredTableAssociation 请求速率	5 TPS	每秒最大 CreateConfiguredTableAssociation 调用数
CreateMembership 请求速率	5 TPS	每秒最大 CreateMembership 调用数
CreatePrivacyBudgetTempate 请求速率	5 TPS	每秒最大 CreatePrivacyBudgetTemplate 调用数
DeleteAnalysisTemplate 请求速率	5 TPS	每秒最大 DeleteAnalysisTemplate 调用数
DeleteCollaboration 请求速率	5 TPS	每秒最大 DeleteCollaboration 调用数
DeleteConfiguredAudienceModelAssociation 请求速率	5 TPS	每秒最大 DeleteConfiguredAudienceModelAssociation 调用数
DeleteConfiguredTable 请求速率	5 TPS	每秒最大 DeleteConfiguredTable 调用数

资源	速率限制	描述
DeleteConfiguredTableAnalysisRule 请求速率	5 TPS	每秒最大 DeleteConfiguredTableAnalysisRule 调用数
DeleteConfiguredTableAssociation 请求速率	5 TPS	每秒最大 DeleteConfiguredTableAssociation 调用数
DeleteMember 请求速率	5 TPS	每秒最大 DeleteMember 调用数
DeleteMembership 请求速率	5 TPS	每秒最大 DeleteMembership 调用数
DeletePrivacyBudgetTemplate 请求速率	5 TPS	每秒最大 DeletePrivacyBudgetTemplate 调用数
GetAnalysisTemplate 请求速率	5 TPS	每秒最大 GetAnalysisTemplate 调用数
GetCollaboration 请求速率	5 TPS	每秒最大 GetCollaboration 调用数
GetCollaborationConfiguredAudienceModelAssociation 请求速率	5 TPS	每秒最大 GetCollaborationConfiguredAudienceModelAssociation 调用数
GetCollaborationPrivacyBudgetTemplate 请求速率	5 TPS	每秒最大 GetCollaborationPrivacyBudgetTemplate 调用数
GetConfiguredAudienceModelAssociation 请求速率	5 TPS	每秒最大 GetConfiguredAudienceModelAssociation 调用数

资源	速率限制	描述
GetConfiguredTable 请求速率	5 TPS	每秒最大 GetConfiguredTable 调用数
GetConfiguredTableAnalysisRule 请求速率	5 TPS	每秒最大 GetConfiguredTableAnalysisRule 调用数
GetConfiguredTableAssociation 请求速率	20 TPS	每秒最大 GetConfiguredTableAssociation 调用数
GetMembership 请求速率	5 TPS	每秒最大 GetMembership 调用数
GetPrivacyBudgetTemplate 请求速率	5 TPS	每秒最大 GetPrivacyBudgetTemplate 调用数
GetProtectedQuery 请求速率	20 TPS	每秒最大 GetProtectedQuery 调用数
GetSchema 请求速率	5 TPS	每秒最大 GetSchema 调用数
GetSchemaAnalysisRule 请求速率	5 TPS	每秒最大 GetSchemaAnalysisRule 调用数
ListAnalysisTemplates 请求速率	5 TPS	每秒最大 ListAnalysisTemplates 调用数
ListCollaborationConfiguredAudienceModelAssociations 请求速率	5 TPS	每秒最大 ListCollaborationConfiguredAudienceModelAssociations 调用数
ListCollaborationPrivacyBudgets 请求速率	5 TPS	每秒最大 ListCollaborationPrivacyBudgets 调用数

资源	速率限制	描述
ListCollaborationPrivacyBudgetTemplates 请求速率	5 TPS	每秒最大 ListCollaborationPrivacyBudgetTemplates 调用数
ListCollaborations 请求速率	5 TPS	每秒最大 ListCollaborations 调用数
ListConfiguredAudienceModelAssociations 请求速率	5 TPS	每秒最大 ListConfiguredAudienceModelAssociations 调用数
ListConfiguredTableAssociations 请求速率	5 TPS	每秒最大 ListConfiguredTableAssociations 调用数
ListConfiguredTables 请求速率	5 TPS	每秒最大 ListConfiguredTables 调用数
ListMembers 请求速率	5 TPS	每秒最大 ListMembers 调用数
ListMemberships 请求速率	5 TPS	每秒最大 ListMemberships 调用数
ListPrivacyBudgets 请求速率	5 TPS	每秒最大 ListPrivacyBudgets 调用数
ListPrivacyBudgetTemplates 请求速率	5 TPS	每秒最大 ListPrivacyBudgetTemplates 调用数
ListProtectedQueries 请求速率	5 TPS	每秒最大 ListProtectedQueries 调用数
ListSchemas 请求速率	5 TPS	每秒最大 ListSchemas 调用数

资源	速率限制	描述
StartProtectedQuery 请求速率	5 TPS	每秒最大 StartProtectedQuery 调用数
UpdateAnalysisTemplate 请求速率	5 TPS	每秒最大 UpdateAnalysisTemplate 调用数
UpdateCollaboration 请求速率	5 TPS	每秒最大 UpdateCollaboration 调用数
UpdateConfiguredAudienceModelAssociation 请求速率	5 TPS	每秒最大 UpdateConfiguredAudienceModelAssociation 调用数
UpdateConfiguredTable 请求速率	5 TPS	每秒最大 UpdateConfiguredTable 调用数
UpdateConfiguredTableAnalysisRule 请求速率	5 TPS	每秒最大 UpdateConfiguredTableAnalysisRule 调用数
UpdateConfiguredTableAssociation 请求速率	5 TPS	每秒最大 UpdateConfiguredTableAssociation 调用数
UpdatePrivacyBudgetTemplate 请求速率	5 TPS	每秒最大 UpdatePrivacyBudgetTemplate 调用数

AWS Clean Rooms 机器学习 API 限制配额

资源	速率限制	描述
CreateAudienceModel 请求速率	1 TPS 速率，3 TPS 突增	每秒的最大 CreateAudienceModel API 调用数

资源	速率限制	描述
CreateConfiguredAudienceModel 请求速率	10 TPS	每秒的最大 CreateConfiguredAudienceModel API 调用数
CreateTrainingDataset 请求速率	10 TPS	每秒的最大 CreateTrainingDataset API 调用数
DeleteAudienceGenerationJob 请求速率	2 TPS 速率，10 TPS 突增	每秒的最大 DeleteAudienceGenerationJob API 调用数
DeleteAudienceModel 请求速率	2 TPS 速率，10 TPS 突增	每秒的最大 DeleteAudienceModel API 调用数
DeleteConfiguredAudienceModel 请求速率	10 TPS	每秒的最大 DeleteConfiguredAudienceModel API 调用数
DeleteConfiguredAudienceModelPolicy 请求速率	25 TPS	每秒的最大 DeleteConfiguredAudienceModelPolicy API 调用数
DeleteTrainingDataset 请求速率	10 TPS	每秒的最大 DeleteTrainingDataset API 调用数
GetAudienceGenerationJob 请求速率	50 TPS	每秒的最大 GetAudienceGenerationJob API 调用数
GetAudienceModel 请求速率	50 TPS	每秒的最大 GetAudienceModel API 调用数
GetConfiguredAudienceModel 请求速率	50 TPS	每秒的最大 GetConfiguredAudienceModel API 调用数

资源	速率限制	描述
GetConfiguredAudienceModelPolicy 请求速率	50 TPS	每秒的最大 GetConfiguredAudienceModelPolicy API 调用数
GetTrainingDataset 请求速率	50 TPS	每秒的最大 GetTrainingDataset API 调用数
ListAudienceExportJobs 请求速率	50 TPS	每秒的最大 ListAudienceExportJobs API 调用数
ListAudienceGenerationJobs 请求速率	50 TPS	每秒的最大 ListAudienceGenerationJobs API 调用数
ListAudienceModels 请求速率	50 TPS	每秒的最大 ListAudienceModels API 调用数
ListConfiguredAudienceModels 请求速率	50 TPS	每秒的最大 ListConfiguredAudienceModels API 调用数
ListTagsForResource 请求速率	50 TPS	每秒的最大 ListTagsForResource API 调用数
ListTrainingDatasets 请求速率	50 TPS	每秒的最大 ListTrainingDatasets API 调用数
PutConfiguredAudienceModelPolicy 请求速率	25 TPS	每秒的最大 PutConfiguredAudienceModelPolicy API 调用数
StartAudienceExportJob 请求速率	1 TPS 速率，3 TPS 突增	每秒的最大 StartAudienceExportJob API 调用数

资源	速率限制	描述
StartAudienceGenerationJob 请求速率	1 TPS 速率，5 TPS 突增	每秒的最大 StartAudienceGenerationJob API 调用数
TagResource 请求速率	10 TPS	每秒的最大 TagResource API 调用数
UntagResource 请求速率	50 TPS	每秒的最大 UntagResource API 调用数
UpdateConfiguredAudienceModel 请求速率	10 TPS	每秒的最大 UpdateConfiguredAudienceModel API 调用数

名称	默认值	可调整	描述
每个受众生成任务的活跃受众导出职位	每个受支持的区域：25 个	否	受众生成作业的活跃受众导出任务的最大数量
每位客户的待定/进行中的受众导出任务	每个受支持的区域：20 个	否	每位客户待处理/进行中的受众导出任务的最大数量
每位客户的待定/正在进行的受众群体生成工作	每个受支持的区域：10 个	是	每位客户的最大待处理/正在进行的受众群体生成工作数量
每位客户的待定/正在进行的受众模型	每个受支持的区域：2 个	是	每位客户待处理/正在进行的受众模型培训作业的最大数量

无尘室机器学习配额

资源	默认值	描述
数据集	每个作业	
最大交互次数	200 亿	训练数据中允许的最大交互次数。对于较大的输入，将缩减采样。
最小交互次数	100 万	
用于相似模型训练的最大不同用户数	100 万	如果包含更多用户，则仅使用前 1 亿个用户（按交互次数排名）。
用于相似模型训练的最小不同用户数	100000	
导出相似区段（受众）作业的最大用户数	10000	
用于模型训练的最大不同项目数。	100 万	您最多可以包含 5000 万个项目，但仅使用最常用的 100 万个项目。
训练数据集中的最大特征列数。	10	
每位用户的最小不同项目数	2	AWS Clean Rooms 机器学习要求每行或每用户都有两个或更多项目，包括重复的项目。
种子受众的最大规模	500,000	
种子受众的最小规模	500	训练数据提供者可以将此值设置为低至 25。
API	每个客户	

资源	默认值	描述
活跃训练数据集总数	500	
活跃相似模型（受众模型）总数	500	
配置的活跃相似模型（受众模型）总数	10000	
完成的相似细分（受众）生成作业总数	无限制	
完成的导出相似细分（受众）作业总数	无限制	
相似模型（受众模型）生成作业的最长持续时间	1 天（24 小时）	
相似细分（受众）生成作业的最长持续时间	10 小时	在你提供种子后，Clean Rooms ML 最多需要 10 个小时才能生成长相相似的区段。
细分（受众）大小区间的最小百分比	1%	
细分（受众）大小区间的最大百分比	20%	
细分（受众）大小区间的最小绝对大小	不同用户数量的 1%	
细分（受众）大小区间的最大绝对大小	不同用户数量的 20%	

《 AWS Clean Rooms 用户指南》的文档历史记录

下表描述了的文档版本 AWS Clean Rooms。

如需对此文档更新的通知，您可以订阅 RSS 源。要订阅 RSS 更新，您必须为当前使用的浏览器启用 RSS 插件。

变更	说明	日期
更新现有策略	在 `AWSCleanRoomsFullAccessNoQuerying` 托管策略添加了以下新权限：`cleanrooms:BatchGetSchemaAnalysisRule` 。	2024年5月13日
AWS Clean Rooms ML 现已完全可用	AWS Clean Rooms 机器学习为双方提供了一种增强隐私的方法，可以识别其数据中的相似用户，而无需彼此共享数据。	2024 年 4 月 3 日
更新现有策略	`AWSCleanRoomsFullAccess` 托管策略中的声明 ID 已从更新ConsolePickQueryResultsBucket为，SetQueryResultsBucket以更好地代表自获得权限以来的权限。	2024 年 3 月 21 日
AWS Clean Rooms 机器学习的新托管策略	添加了两个新的托管式策略：`AWSCleanRoomsMLReadOnlyAccess` 和 `AWSCleanRoomsMLFullAccess` 。	2023 年 11 月 29 日
AWS Clean Rooms 机器学习（预览）	AWS Clean Rooms 机器学习为双方提供了一种增强隐私	2023 年 11 月 29 日

	的方法，可以识别其数据中的相似用户，而无需彼此共享数据。	
AWS Clean Rooms 差异隐私（预览版）	客户现在可以使用 AWS Clean Rooms 差分隐私来帮助保护其用户的隐私。	2023 年 11 月 29 日
付款配置	协作创建者现在可以配置可以运行查询的成员或协作中的其他成员，以收取查询计算费用。	2023 年 11 月 14 日
查询运行时间 - 更新	超时前运行查询的最长时间从 4 小时更新为 12 小时。	2023 年 10 月 6 日
AWS CloudFormation 资源-更新	AWS Clean Rooms 添加了以下新资源：`AWS::CleanRooms::Membership ProtectedQueryOutputConfiguration AWS::CleanRooms::Membership ProtectedQueryResultConfiguration` 、和`AWS::CleanRooms::Membership ProtectedQueryS3OutputConfiguration` 。	2023 年 9 月 7 日
AWS CloudFormation 资源-更新	AWS Clean Rooms 添加了以下新资源：`AWS::CleanRooms::AnalysisTemplate` 和`AWS::CleanRooms::ConfiguredTable AnalysisRuleCustom` 。	2023 年 8 月 31 日

成员能力分开	协作创建者现在可以指定一名成员负责查询，另一名成员负责接收结果。这样，协作创建者就能确保可以查询的成员无法访问查询结果。	2023 年 8 月 30 日
AWS Clean Rooms 术语表	仅限文档的更新以添加术语表。 AWS Clean Rooms	2023 年 8 月 30 日
对 Apache Iceberg 表的支持（预览版）	AWS Clean Rooms 现在支持Apache Iceberg表格（预览）。	2023 年 8 月 25 日
配额更新	更新了配额部分，以反映每个账户成员身份的新默认配额。	2023 年 8 月 9 日

对现有策略的更新	在 AWSCleanRoomsFullAccessNoQuerying 托管式策略中新增了以下新权限：cleanrooms:CreateAnalysisTemplate 、cleanrooms:GetAnalysisTemplate 、cleanrooms:UpdateAnalysisTemplate 、 cleanrooms:DeleteAnalysisTemplate 、cleanrooms:ListAnalysisTemplates cleanrooms:GetCollaborationAnalysisTemplate 、cleanrooms:BatchGetCollaborationAnalysisTemplate 和 cleanrooms:ListCollaborationAnalysisTemplates 。	2023 年 7 月 31 日
分析模板和自定义分析规则	AWS Clean Rooms 现在支持分析模板和自定义分析规则。分析模板使协作者能够构建或导入自己的自定义 SQL 查询，以便在协作中使用。使用自定义分析规则，表所有者可以批准对其配置表进行自定义 SQL 查询。	2023 年 7 月 31 日
分析规则支持 OR 逻辑条件	AWS Clean Rooms 分析规则现在支持子JOIN句中的OR逻辑条件。	2023 年 6 月 29 日

CloudFormation 整合	AWS Clean Rooms 现在与集成 AWS CloudFormation。	2023 年 6 月 15 日
分析构建器	现在，能够查询和接收结果的成员可以使用分析构建器用户界面对某些表运行查询，而无需编写 SQL 代码。	2023 年 6 月 15 日
SQL 函数	仅限文档的更新，阐明支持的 SQL 函数。	2023 年 5 月 5 日
故障排除	仅限文档的更新，为常见问题添加了“疑难解答”一节。	2023 年 4 月 27 日
支持的数据类型 AWS Clean Rooms	仅限文档的更新以添加列出支持 AWS Glue Data Catalog 的数据类型的新部分。	2023 年 4 月 26 日
AWS CloudTrail 事件示例	仅限文档的更新，添加了 StartProtectedQuery（成功）和StartProtectedQuery（失败）CloudTrail 的事件示例。	2023 年 4 月 20 日
对现有策略的更新	在 `AWSCleanRoomsFullAccessNoQuerying` 托管式策略中新增了以下新权限：`cleanrooms:ListTagsForResource` 、`cleanrooms:UntagResource` 和 `cleanrooms:TagResource` 。有关更多信息，请参阅 AWS 托管式策略。	2023 年 3 月 21 日
正式发布	AWS Clean Rooms 现已正式上市。	2023 年 3 月 21 日

预览版	《AWS Clean Rooms 用户指南》的预览版	2023 年 1 月 12 日

AWS Clean Rooms 词汇表

请查阅此词汇表，熟悉 AWS Clean Rooms所用的术语。

聚合分析规则

查询限制，允许使用 COUNT、SUM 或 AVG 函数沿可选维度进行聚合分析的查询。这些查询不会泄露行级信息。

支持活动规划、媒体覆盖面、频率和换算测量值等使用案例。

其他类型的分析规则包括自定义和列表。

分析规则

授权特定类型查询的查询限制。

分析规则类型决定了可以在配置表上运行哪种分析。每种类型都有预定义的查询结构。您可以通过查询控制来控制如何在结构中使用表列。

分析规则类型包括聚合、列表和自定义。

分析模板

特定于协作的预先批准的查询，可以重复使用。

支持中支持的自定义 SQL 查询 AWS Clean Rooms。

可在 SQL 查询中出现字面值的任何地方包含参数。有关支持的参数类型的更多信息，请参阅《AWS Clean Rooms SQL 参考》中的数据类型。

分析模板仅适用于自定义分析规则。

C3R 加密客户端

Clean Rooms 计算加密 (C3R) 加密客户端。

C3R 是一个具有命令行界面的客户端加密 SDK，用于加密和解密数据。

cleartext 列

在 JOIN 或 SELECT SQL 构造中未受加密保护的列。

cleartext 列可以用于 SQL 查询的任何部分。

协作

一种安全的逻辑边界， AWS Clean Rooms 成员可以在其中对已配置的表执行 SQL 查询。

协作由协作创建者创建。

只有受邀参与协作的成员才能加入协作。

一个协作只能有一个成员可以查询数据，一个成员可以接收结果，一个成员可为查询计算费用付费。

所有成员在加入协作之前都可以看到协作的受邀参与者列表。

协作创建者

创建协作的成员。

每个协作只有一个协作创建者。

只有协作创建者才能从协作中删除成员或删除协作。

配置表

每个已配置的表都表示对中已配置为在 AWS Glue Data Catalog 中使用的现有表的引用 AWS Clean Rooms。配置表包含用于确定如何使用数据的分析规则。

目前， AWS Clean Rooms 支持关联存储在亚马逊简单存储服务 (Amazon S3) 中的数据，这些数据是通过编目的。 AWS Glue

有关的更多信息 AWS Glue，请参阅《AWS Glue 开发人员指南》。

配置表可以与一个或多个协作关联。

> **Note**
>
> AWS Clean Rooms 目前不支持注册到的 Amazon S3 存储桶位置 AWS Lake Formation。

自定义分析规则

查询限制，允许一组特定的预先批准的查询（分析模板），或者允许一组特定的账户来提供使用您的数据的查询。

支持首触归因、增量分析和受众发现分析等使用案例。

支持差别隐私。

解密

将加密数据转换回其原始形式的过程。只有获得密钥，才能进行解密。

差别隐私

一种在数学上非常严格的技术，可以保护协作数据以防止可以接收结果的成员了解特定个人的数据。

加密

使用称为密钥的机密值将数据编码成看似随机的形式的过程。如果无法访问密钥，就无法确定原始明文。

指纹列

在 JOIN SQL 构造中未受加密保护的列。

列表分析规则

查询限制，允许对该表和可查询成员表之间的重叠情况输出行级属性分析的查询。

支持扩充以及受众拓展或抑制等使用案例。

成员

作为协作参与者的 AWS 客户。

使用 AWS 账户识别成员身份。

所有成员都可以贡献数据。

可以查询的成员

可以在协作中查询数据的成员。

每个协作中只有一个成员可以查询，而且该成员是不可变的。

管理用户可以使用 AWS Identity and Access Management (IAM) 权限来控制其哪些 IAM 委托人（例如用户或角色）可以查询协作中的数据。有关更多信息，请参阅 创建服务角色来读取数据。

可以接收结果的成员

可以接收查询结果的成员。可以接收结果的成员指定 Amazon S3 目标的查询结果设置和查询结果格式。

每个协作中只有一个成员可以接收结果，而且该成员是不可变的。

支付查询计算费用的成员

负责支付查询计算费用的成员。

只有一个成员负责支付每个协作的查询计算费用，而且该成员是不可变的。

如果协作创建者未将任何人指定为支付查询计算费用的成员，则可以查询的成员为默认付款人。

支付查询计算费用的成员会收到协作中已运行的查询的账单。

成员身份

成员加入协作时创建的资源。

成员关联到协作的所有资源都是成员身份的一部分，或与成员身份相关联。

只有拥有该成员身份的成员才能在该成员身份中添加、删除或编辑资源。

密封列

在 SELECT SQL 构造中未受加密保护的列。

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。