aws

开发人员指南

AWS Lake Formation

AWS Lake Formation: 开发人员指南

Table of Contents

什么是 AWS Lake Formation？

欢迎阅读《AWS Lake Formation 开发人员指南》。

AWS Lake Formation 可帮助您集中管理、保护和全球共享数据，以便进行分析和机器学习。您可以对 Amazon Simple Storage Service (Amazon S3) 上的数据湖数据及其在 AWS Glue Data Catalog 中的元数据进行精细访问控制。

Lake Formation 提供了自己的权限模型，该模型增强了 IAM 权限模型。Lake Formation 权限模型支持通过简单的授予或撤销机制对存储在数据湖中的数据进行精细访问，这与关系数据库管理系统 (RDBMS) 非常相似。Lake Formation 权限是在 AWS 分析和机器学习服务（包括 Amazon Athena、Amazon QuickSight、Amazon Redshift Spectrum、Amazon EMR 和 AWS Glue）的列、行和单元格级别使用精细控制来强制实施的。

的 Lake Formation 混合访问模式AWS Glue Data Catalog允许您使用 Lake Formation 权限和 Amazon S3 和AWS Glue操作的 IAM 权限策略来保护和访问已编目的数据。借助混合访问模式，数据管理员可以有选择地以增量方式加载 Lake Formation 权限，一次专注于一个数据湖用例。

Lake Formation 还允许您在多个 AWS 账户、AWS 组织之间或直接与另一个账户中的 IAM 主体内部和外部共享数据，从而提供对 AWS Glue Data Catalog 元数据和基础数据的精细访问。

主题

Lake Formation 特征

Lake Formation 可帮助您打破数据孤岛，并将不同类型的结构化和非结构化数据合并到一个集中式存储库中。首先，确定 Amazon S3 或关系数据库和 NoSQL 数据库中的现有数据存储，然后将数据移动

到数据湖中。然后对数据进行抓取、编目和准备以供分析。接下来，通过用户选择的分析服务，为他们提供对数据的安全自助访问。

主题

数据摄取和管理

从 AWS 中已有的数据库导入数据

指定现有数据库的位置并提供访问凭证后，Lake Formation 就会读取数据及其元数据（架构）以了解数据来源的内容。然后，它会将数据导入您的新数据湖，并将元数据记录在中央目录中。借助 Lake Formation，您可以从在 Amazon RDS 中运行或托管在 Amazon EC2 中的 MySQL、PostgreSQL、SQL Server、MariaDB 和 Oracle 数据库导入数据。支持批量和增量数据加载。

从其他外部来源导入数据

您可以使用 Lake Formation 通过与 Java Database Connectivity (JDBC) 连接来从本地数据库移动数据。确定您的目标来源并在控制台中提供访问凭证，然后 Lake Formation 会读取您的数据并将其加载到数据湖中 要从上述数据库以外的数据库导入数据，您可以使用 AWS Glue 创建自定义 ETL 作业。

对数据进行编目和标记

您可以使用 AWS Glue 爬网程序读取您在 Amazon S3 中的数据，提取数据库和表架构，并将这些数据存储在可搜索 AWS Glue Data Catalog 中。然后，使用 Lake Formation Lake Formation 基于标签的访问控制 (TBAC) 管理对数据库、表和列的权限。有关将表添加到数据目录的更多信息，请参阅创建数据目录表和数据库。

安全管理

定义和管理访问控制

Lake Formation 提供了一个位置来管理数据湖中数据的访问控制。您可以定义安全策略，以限制对数据库、表、列、行和单元格级别的数据的访问。这些策略适用于 IAM 用户和角色，也适用于通过外部身份提供商进行联合身份验证时的用户和组。您可以使用精细控制来访问 Amazon Redshift

Spectrum、Athena、AWS Glue ETL 和 Amazon EMR for Apache Spark 中受 Lake Formation 保护的数据。每当您创建 IAM 身份时，请确保遵循 IAM 最佳实践。有关更多信息，请参阅《IAM 用户指南》中的安全最佳实践。

混合访问模式

Lake Formation 混合访问模式让您能够灵活地选择为 AWS Glue Data Catalog 中的数据库和表启用 Lake Formation 权限。在混合访问模式下，您现在有了增量路径，可您为一组特定的用户设置 Lake Formation 权限，而不会中断其他现有用户或工作负载的权限策略。有关更多信息，请参阅混合访问模式。

实施审计日志记录

Lake Formation 提供全面的审计日志， CloudTrail 用于监控访问情况并显示对集中定义策略的遵守情况。您可以跨分析和机器学习服务审核数据访问历史记录，这些服务通过 Lake Formation 读取数据湖中的数据。这使您可以查看哪些用户或角色尝试访问了哪些数据、使用了哪些服务以及何时访问了数据和使用了服务。您可以像使用 CloudTrail API 和控制台访问任何其他 CloudTrail 日志一样访问审核日志。有关 CloudTrail 日志的更多信息，请参阅使用 AWS CloudTrail 记录 AWS Lake Formation API 调用。

行和单元格级别安全功能

Lake Formation 提供了数据筛选条件，允许您限制对列和行组合的访问。使用行和单元格级别安全功能来保护敏感数据，例如个人身份信息 (PII)。有关行级别安全功能的更多信息，请参阅数据筛选概览。

基于标签的访问控制

使用基于 Lake Formation 标签的访问控制，通过创建名为 LF-Tags 的自定义标签来管理数百甚至数千个数据权限。现在，您可以定义 LF 标签并将其附加到数据库、表或列。然后，跨分析、机器学习 (ML) 和提取、转换、加载 (ETL) 服务共享受控访问权限以供使用。LF-tags 通过将数千个资源的策略定义替换为几个逻辑标签，确保可以轻松扩展数据治理。Lake Formation 提供了对这些元数据的基于文本的搜索，因此您的用户可以快速找到他们需要分析的数据。

跨账户访问

Lake Formation 权限管理功能通过集中式方法简化了跨多个 AWS 账户的分布式数据湖的保护和管理，从而提供了对数据目录和 Amazon S3 位置的精细访问控制。有关更多信息，请参阅Lake Formation 中的跨账户数据共享。

数据共享

数据共享功能使您能够对存储在不同数据来源（如 Amazon Redshift）中的数据集设置权限，而无需将数据或元数据迁移到 Amazon S3 或 AWS Glue Data Catalog。您可以使用以下方法在 Lake Formation 中共享数据：

有关更多信息，请参阅 Lake Formation 中的数据共享。

- 将 Lake Formation 与 Amazon Redshift 数据共享集成 – 使用 Lake Formation 集中管理 Amazon Redshift 数据共享的数据库、表、列和行级访问权限，并限制用户对数据共享内对象的访问。
- 将 AWS Glue Data Catalog 连接到外部元存储 – 使用 Lake Formation 将 AWS Glue Data Catalog 连接到外部元存储以管理对 Amazon S3 中数据集的访问权限。无需将元数据迁移到 AWS Glue Data Catalog。

  有关更多信息，请参阅管理对使用外部元存储的数据集的权限。
- 将 Lake Formation 和 AWS Data Exchange 集成 – Lake Formation 支持许可通过 AWS Data Exchange 对您的数据进行访问。如果您有兴趣获得 Lake Formation 数据的许可，请参阅《AWS Data Exchange 用户指南》中的什么是 AWS Data Exchange？。

AWS Lake Formation：工作方式

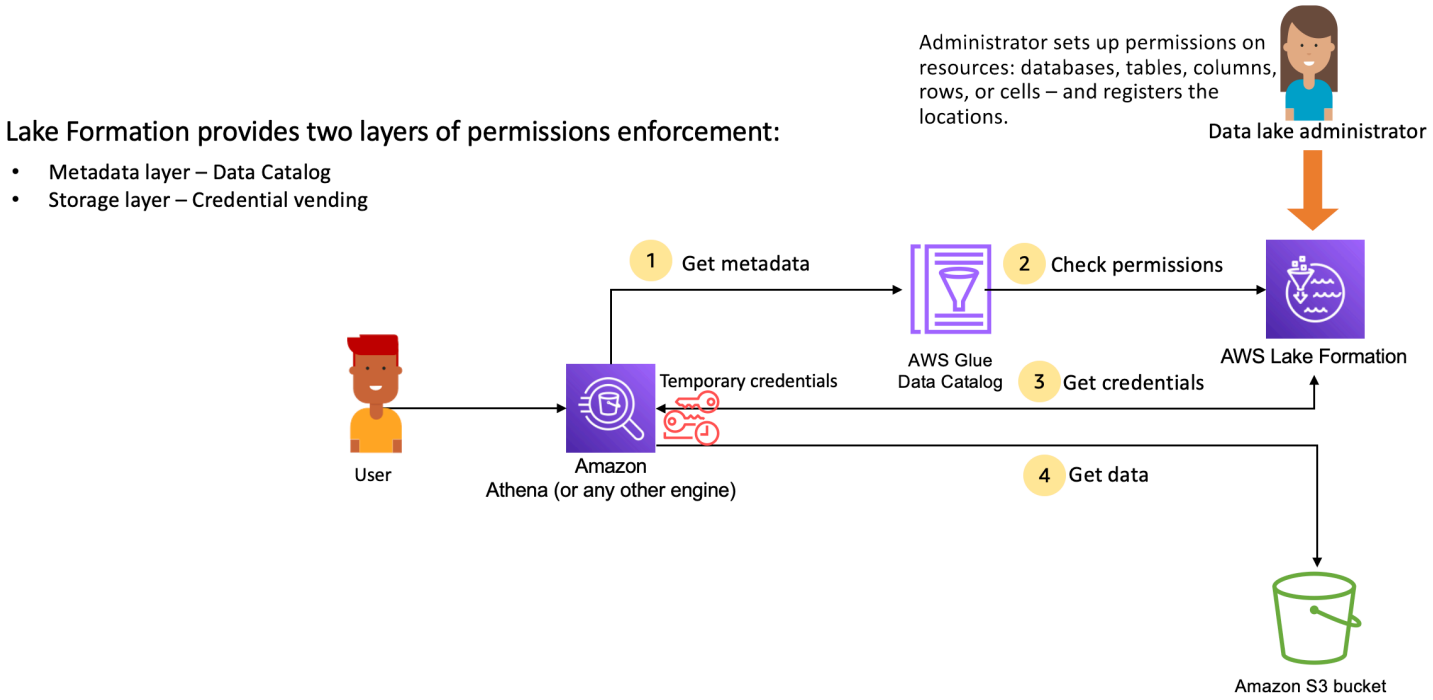
AWS Lake Formation 提供了关系数据库管理系统 (RDBMS) 权限模型，以用于授予或撤销对数据目录资源（例如，Amazon S3 中含基础数据的数据库、表和列）的访问权限。易于管理的 Lake Formation 权限取代了复杂的 Amazon S3 存储桶策略和相应的 IAM 策略。

在 Lake Formation 中，您可以在两个级别实施权限：

- 对数据目录资源（例如，数据库和表）强制实施元数据级别权限
- 代表集成引擎管理 Amazon S3 上存储的基础数据的存储访问权限

Lake Formation 权限管理工作流

Lake Formation 与分析引擎集成，以查询已在 Lake Formation 中注册的 Amazon S3 数据存储和元数据对象。下图说明了 Lake Formation 中权限管理的工作原理。



Lake Formation 权限管理主要步骤

数据湖管理员或具有管理权限的用户需设置单个数据目录表用户策略以允许或拒绝使用 Lake Formation 权限访问数据目录表，之后，Lake Formation 方可提供对数据湖中数据的访问控制。

然后，数据湖管理员或管理员委派的用户向各用户授予对数据目录数据库和表的 Lake Formation 权限，并将该表的 Amazon S3 位置注册到 Lake Formation 中。

1. 获取元数据 — 主体（用户）向集成分析引擎（例如 Amazon Athena、AWS Glue、Amazon EMR 或 Amazon Redshift Spectrum）提交查询或 ETL 脚本。集成分析引擎可识别正在请求的表，并将元数据请求发送到数据目录。
2. 检查权限 – 数据目录检查用户在 Lake Formation 中的权限，如果用户有权访问该表，则会将允许该用户查看的元数据返回给引擎。
3. 获取凭证 — 数据目录会告知引擎，该表是否由 Lake Formation 管理。如果基础数据已在 Lake Formation 中注册，则分析引擎会请求 Lake Formation 通过授予临时访问权限来提供数据访问权限。
4. 获取数据 – 如果用户有权访问该表，Lake Formation 将提供对集成分析引擎的临时访问权限。通过使用临时访问权限，分析引擎可从 Amazon S3 获取数据，并执行必要的筛选，例如列、行或单元格筛选。当引擎运行完作业后，它会将结果返回给用户。此过程称为凭证售卖。

   如果该表不是由 Lake Formation 管理的，则分析引擎会直接向 Amazon S3 进行第二次调用。系统会对相关 Amazon S3 存储桶策略和 IAM 用户策略进行评估以确定是否支持访问数据。

每当您使用 IAM policy 时，请确保遵循 IAM 最佳实践。有关更多信息，请参阅《IAM 用户指南》中的 IAM 安全最佳实践。

主题

元数据权限

Lake Formation 提供对数据目录的授权和访问控制。当 IAM 角色从任何系统进行数据目录 API 调用时，数据目录会验证用户的数据权限，并且仅返回用户有权访问的元数据。例如，如果某 IAM 角色只能访问数据库中的一个表，并且担任该角色的服务或用户执行 `GetTables` 操作，则无论数据库中有多少个表，响应都将只包含一个表。

默认设置 - **`IAMAllowedPrincipal`** 组权限

默认情况下，AWS Lake Formation 将对所有数据库和表的权限设置为一个名为 `IAMAllowedPrincipal` 的虚拟组。该组独一无二，只显示在 Lake Formation 中。`IAMAllowedPrincipal` 组包括所有通过 IAM 主体策略和 AWS Glue 资源策略访问数据目录资源的 IAM 主体。如果针对数据库或表存在此权限，则将授予所有主体访问该数据库或表的权限。

如果您想提供对数据库或表的更精细的权限，请删除 `IAMAllowedPrincipal` 权限，之后 Lake Formation 会强制实施与该数据库或表关联的所有其他策略。例如，如果存在允许用户 A 使用 `DESCRIBE` 权限访问数据库 A 的策略，并且 `IAMAllowedPrincipal` 中含有所有权限，则用户 A 将继续执行所有其他操作，直到 `IAMAllowedPrincipal` 权限被撤销。

此外，默认情况下，在创建所有新数据库和表时，`IAMAllowedPrincipal` 组便拥有对这些数据库和表的权限。有两种配置可以控制这种行为。第一种是可为新创建的数据库启用此功能的账户和区域级别配置，第二种是数据库级别配置。要修改默认设置，请参阅更改默认权限模式或使用混合访问模式。

授予权限

数据湖管理员可以向主体授予数据目录权限，以便主体可以创建和管理数据库和表，并且可以访问基础数据。

数据库和表级别权限

在 Lake Formation 中授予权限时，授予者必须指定要将权限授予的主体、要针对其授予权限的资源以及被授权者应有权执行的操作。对于 Lake Formation 中的大多数资源，主体列表和要针对其授予权限的资源都相似，但被授权者可以执行的操作因资源类型而异。例如，使用 SELECT 权限可以读取表，但对于数据库，则无法实施 SELECT 权限。CREATE_TABLE 权限可针对数据库实施，但不能针对表实施。

您可以使用两种方法授予 AWS Lake Formation 权限：

- 命名资源方法 – 允许您在向用户授予权限时选择数据库和表名。
- 基于 LF 标签的访问控制 (LF-TBAC) — 用户可创建 LF 标签，将其与数据目录资源关联，授予对 LF 标签的 Describe 权限，将权限关联到单个用户，以及使用 LF 标签为不同的用户编写 LF 权限策略。此类基于 LF 标签的策略应用于所有与这些 LF 标签值关联的数据目录资源。

> **Note**
>
> LF 标签是 Lake Formation 所独有的。它们仅在 Lake Formation 中显示，不应将其与 AWS 资源标签混淆。

LF-TBAC 是一项功能，允许用户将资源分组为用户定义的 LF 标签类别并对这些资源组应用权限。因此，这是跨大量数据目录资源扩展权限的最佳方式。

有关更多信息，请参阅Lake Formation 基于标签的访问控制。

当您向主体授予权限时，Lake Formation 会将权限作为该用户的所有策略的并集进行评估。例如，如果针对某主体有两项表策略，其中一项策略通过命名资源方法授予对列 col1、col2 和 col3 的权限，另一项策略通过 LF 标签向同一主体授予对同一表中 col5 和 col6 的权限，则有效权限将是对 col1、col2、col3、col5 和 col6 的权限的并集。这还包括数据筛选条件和行。

数据位置权限

数据位置权限使非管理用户能够在特定的 Amazon S3 位置创建数据库和表。如果用户尝试在他们无权创建数据库或表的位置创建数据库或表，则创建任务将失败。这是为了防止用户在数据湖中的任意位置创建表，并控制这些用户可以在何处读取和写入数据。在用于创建数据库的 Amazon S3 位置创建表时，存在隐式权限。有关更多信息，请参阅授予数据位置权限。

创建表和数据库权限

默认情况下，非管理用户无权创建数据库，也无权在数据库中创建表。对于数据库的创建，是使用 Lake Formation 设置在账户级别控制的，因此只有获得授权的主体才能创建数据库。有关更多信息，请参阅创建数据库。要创建表，主体需要对创建该表所在的数据库拥有 CREATE_TABLE 权限。有关更多信息，请参阅创建表。

隐式和显式权限

Lake Formation 根据角色和角色执行的操作提供隐式权限。例如，数据湖管理员可自动获得对数据目录中所有资源的 DESCRIBE 权限、对所有位置的数据位置权限、在所有位置创建数据库和表的权限以及对任何资源的 Grant 和 Revoke 权限。数据库创建者自动获得对他们创建的数据库的所有数据库权限，而表创建者则获得对他们创建的表的所有权限。有关更多信息，请参阅隐式 Lake Formation 权限。

可授予的权限

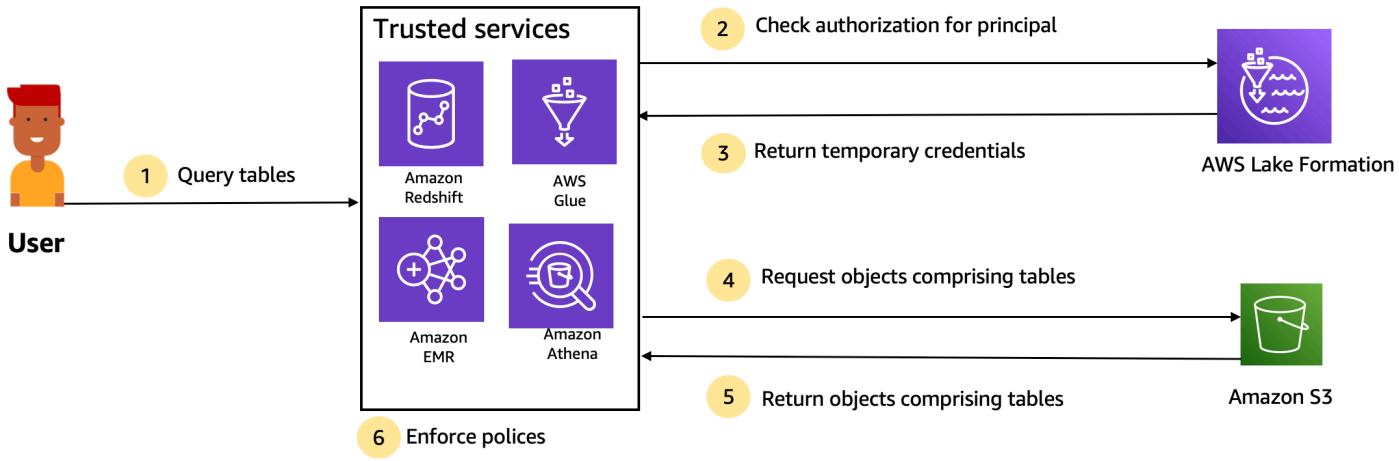
数据湖管理员能够通过提供可授予的权限将权限管理任务委派给非管理用户。当主体获得对资源的可授予权限和一组权限时，该主体能够向其他主体授予对该资源的权限。

存储访问管理

Lake Formation 使用凭证售卖功能提供对 Amazon S3 数据的临时访问权限。凭证售卖或令牌售卖是一种常见的模式，它向用户、服务或一些其他实体提供临时凭证，以授予其对资源的短期访问权限。

Lake Formation 利用这种模式提供对 Athena 等 AWS 分析服务的短期访问权限，以使被授予者能够代表调用主体访问数据。在授予权限时，用户无需更新其 Amazon S3 存储桶策略或 IAM 策略，也无需直接访问 Amazon S3。

下图显示了 Lake Formation 如何提供对注册位置的临时访问权限：



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

1. 主体（用户）通过可信的集成服务（例如 Athena、Amazon EMR、Redshift Spectrum 或 AWS Glue）输入对表数据的查询或请求。
2. 该集成服务会检查 Lake Formation 针对表和所请求的列提供的授权，并做出授权决定。如果用户未获得授权，Lake Formation 将拒绝用户访问数据，并且查询将失败。
3. 授权成功且为表和用户开启存储授权后，集成服务将从 Lake Formation 检索临时凭证以访问数据。
4. 该集成服务使用 Lake Formation 提供的临时凭证请求 Amazon S3 中的对象。
5. Amazon S3 向集成服务提供 Amazon S3 对象。Amazon S3 对象包含表中的所有数据。
6. 集成服务对 Lake Formation 策略执行必要的强制实施，例如列级别、行级别和/或单元格级别筛选。集成服务会处理查询并将结果返回给用户。

为数据目录表启用存储级别权限强制实施

默认情况下，不对数据目录中的表启用存储级别强制实施。要启用存储级别强制实施，您必须在 Lake Formation 中注册源数据的 Amazon S3 位置并提供 IAM 角色。将为与 Amazon S3 位置具有相同的表位置路径或前缀的所有表启用存储级别权限。

当集成服务代表用户请求访问数据位置时，Lake Formation 服务将担任此角色，将凭证返回给所请求的服务，并在确保可以访问数据的情况下缩小对资源的权限范围。已注册的 IAM 角色必须对 Amazon S3 位置（包括 AWS KMS 密钥）拥有所有必需访问权限。

有关更多信息，请参阅注册 Amazon S3 位置。

支持的 AWS 服务

Athena、Redshift Spectrum、Amazon EMR、AWS Glue、Amazon QuickSight 和 Amazon SageMaker 等 AWS 分析服务使用 Lake Formation 凭证售卖 API 操作与 AWS Lake Formation 集成。要查看与 Lake Formation 集成的 AWS 服务的完整列表以及它们支持的粒度级别和表格格式，请参阅使用其他 AWS 服务。

Lake Formation 中的跨账户数据共享

借助 Lake Formation，您可以使用命名资源方法或 LF 标签，通过简单设置在 AWS 账户内和跨账户共享数据目录资源（数据库和表）。您可以将整个数据库或数据库中的部分表共享给账户中的任何 IAM 主体（IAM 角色和用户）、账户级别的其他 AWS 账户，或者直接共享给另一个账户中的 IAM 主体。

您还可以利用数据筛选条件将数据目录表共享，以限制对行级别和单元格级别详细信息的访问。Lake Formation 使用 AWS Resource Access Manager (AWS RAM) 促进在账户之间授予权限的过程。在两个账户之间共享资源时，AWS RAM 会向接收方账户发送邀请。当用户接受 AWS RAM 共享邀请

时，AWS RAM 会向其提供必要的 Lake Formation 权限以确保数据目录资源可用并启用存储级别强制实施。有关更多信息，请参阅Lake Formation 中的跨账户数据共享。

当接收方账户的数据湖管理员接受 AWS RAM 共享时，共享资源将在接收方账户中可用。数据湖管理员会向接收方账户中的其他 IAM 主体授予对共享资源的更多 Lake Formation 权限（前提是管理员拥有对共享资源的 GRANTABLE 权限）。

但是，如果没有资源链接，主体就无法使用 Athena 或 Redshift Spectrum 来查询共享资源。资源链接是数据目录中的一个实体，类似于 Linux-Symlink 的概念。

接收方账户的数据湖管理员会针对共享资源创建资源链接。管理员会向其他用户授予对资源链接的 Describe 权限以及对原共享资源的必需权限。然后，接收方账户中的用户可以通过该资源链接，使用 Athena 和 Redshift Spectrum 查询共享资源。有关资源链接的更多信息，请参阅创建资源链接。

Lake Formation 组件

AWS Lake Formation 依赖于多个组件之间的交互来创建和管理数据湖。

Lake Formation 控制台

您可以使用 Lake Formation 控制台定义和管理数据湖，并授予和撤销 Lake Formation 权限；可以在该控制台上使用蓝图来发现、清理、转换和摄取数据；还可以启用或禁用单个 Lake Formation 用户对该控制台的访问。

Lake Formation API 和命令行界面

Lake Formation 通过多个特定于语言的 SDK 和 AWS Command Line Interface (AWS CLI) 提供 API 操作。Lake Formation 与 AWS Glue API 结合使用。Lake Formation API 主要侧重于管理 Lake Formation 权限，而 AWS Glue API 则提供数据目录 API 和托管式基础设施，用于定义、安排和运行对数据的 ETL 操作。

有关 AWS Glue API 的信息，请参阅 AWS Glue 开发人员指南。有关使用 AWS CLI 的信息，请参阅 AWS CLI 命令参考。

其他 AWS 服务

Lake Formation 使用以下服务：

- AWS Glue：用于编排作业和爬网程序以使用 AWS Glue 转换来转换数据。

- IAM：用于向 Lake Formation 主体授予权限策略。Lake Formation 权限模型增强了 IAM 权限模型，以保护您的数据湖。

Lake Formation 术语

以下是您将在本指南中遇到的一些重要术语。

数据湖

“数据湖”是存储在 Amazon S3 中并由 Lake Formation 使用数据目录管理的持久性数据。数据湖通常存储以下内容：

- 结构化数据和非结构化数据
- 原始数据和转换后的数据

要使 Amazon S3 路径位于数据湖内，必须向 Lake Formation 注册该路径。

数据访问

Lake Formation 通过增强 AWS Identity and Access Management (IAM) 策略的新授予/撤销权限模型，提供对数据的安全和精细访问。

分析师和数据科学家可以使用完整的 AWS 分析和机器学习服务组合（如 Amazon Athena）来访问数据。配置的 Lake Formation 安全策略有助于确保用户只能访问自己有权访问的数据。

混合访问模式

混合访问模式允许您使用 Lake Formation 权限以及 IAM 和 Amazon S3 权限来保护和访问已编目的数据。混合访问模式允许数据管理员有选择地以增量方式加载 Lake Formation 权限，一次专注于一个数据湖用例。

蓝图

“蓝图”是一种数据管理模板，可让您轻松地将数据摄取到数据湖中。Lake Formation 提供了多个蓝图，每个蓝图都适用于预定义的源类型，例如关系数据库或 AWS CloudTrail 日志。在蓝图中，您可以创建工作流。工作流由 AWS Glue 爬网程序、作业和触发器组成，生成它们是为了编排数据的加载和更新。蓝图将数据来源、数据目标和计划作为配置工作流的输入。

工作流

“工作流”是一组相关 AWS Glue 作业、爬网程序和触发器的容器。您可以在 Lake Formation 中创建工作流，然后在 AWS Glue 服务中执行。Lake Formation 可以将工作流作为单个实体跟踪其状态。

定义工作流时，您可以选择其所基于的蓝图。然后可以按需或按计划运行工作流。

您在 Lake Formation 中创建的工作流在 AWS Glue 控制台中显示为有向无环图 (DAG) 形式。使用 DAG，您可以跟踪工作流的进度并执行问题排查。

数据目录

“数据目录”是持久性元数据存储。它是一项托管式服务，可让您在 AWS 云中存储、注释和共享元数据，就像在 Apache Hive 元存储中一样。它提供了一个统一的存储库，不同的系统可以在其中存储和查找元数据来跟踪数据孤岛中的数据，然后使用该元数据来查询和转换数据。Lake Formation 使用 AWS Glue 数据目录来存储有关数据湖、数据来源、转换和目标的元数据。

有关数据来源和目标的元数据采用数据库和表的形式。表存储架构信息、位置信息等。数据库是表的集合。Lake Formation 提供权限层次结构来控制对数据目录中的数据库和表的访问权限。

每个 AWS 账户在每个 AWS 区域都有一个数据目录。

基础数据

“基础数据”是指数据目录表指向的数据湖中的源数据或数据。

主体

“主体”是 AWS Identity and Access Management (IAM) 用户或角色或者 Active Directory 用户。

数据湖管理员

“数据湖管理员”是可以向任何主体（包括自己）授予对任何数据目录资源或数据位置的任何权限的主体。将数据湖管理员指定为数据目录的第一个用户。然后，此用户可以向其他主体授予更精细的资源权限。

> **Note**
>
> IAM 管理用户（使用 `AdministratorAccess` AWS 托管式策略的用户）不会自动成为数据湖管理员。例如，他们无法授予 Lake Formation 对目录对象的权限，除非他们已获得相应权限。但是，他们可以使用 Lake Formation 控制台或 API 将自己指定为数据湖管理员。

有关数据湖管理员功能的信息，请参阅隐式 Lake Formation 权限。有关将用户指定为数据湖管理员的信息，请参阅创建数据湖管理员。

AWS 服务与 Lake Formation 的集成

您可以使用 Lake Formation 来管理对存储在 Amazon S3 中的数据的数据库、表和列级访问权限。当您在 Lake Formation 中注册数据时，可以使用 AWS Glue、Amazon Athena、Amazon Redshift Spectrum、Amazon EMR 等 AWS 分析服务来查询数据。以下 AWS 服务与 AWS Lake Formation 集成并遵循 Lake Formation 权限。

AWS 服务	集成详细信息
AWS Glue	参考主题：AWS Lake Formation 与一起使用 AWS Glue AWS Glue 与 Lake Formation 共享同一数据目录。对于控制台操作（例如查看表列表）和所有 API 操作，AWS Glue 用户只能访问自己对其具有 Lake Formation 权限的数据库和表。
Amazon Athena	参考主题：在亚马逊 A AWS Lake Formation thena 上使用 使用 Lake Formation 允许或拒绝读取 Amazon S3 中数据的权限。当 Amazon Athena 用户在查询编辑器中选择 AWS Glue 目录时，他们只能查询自己对其具有 Lake Formation 权限的数据库、表和列。不支持使用清单的查询。 目前，Lake Formation 不支持管理对采用开放表格格式的表的写入操作（例如 `VACUUM`、`MERGE`、`UPDATE` 和 `OPTIMIZE`）权限。 除了通过 AWS Identity and Access Management (IAM) 向 Athena 进行身份验证的主体之外，Lake Formation 还支持通过 JDBC 或 ODBC 驱动程序连接并通过 SAML 进行身份验证的 Athena 用户。支持的 SAML 提供商包括 Okta 和 Microsoft Active Directory 联合身份验证服务 (AD FS)。
Amazon Redshift Spectrum	参考主题：AWS Lake Formation 与亚马逊 Redshift Spectrum 一起使用

AWS 服务	集成详细信息
	当 Amazon Redshift 用户在 AWS Glue Data Catalog 中的数据库上创建外部架构时，他们只能查询该架构中自己对其具有 Lake Formation 权限的表和列。
亚马逊 QuickSight 企业版	参考：在 Amazon AWS Lake Formation 上使用 QuickSight 当亚马逊 QuickSight 企业版用户在亚马逊 S3 位置查询数据集时，该用户必须拥有该数据的 Lake Formation SELECT 权限。
Amazon EMR	参考：AWS Lake Formation 与 Amazon EMR 一起使用 创建具有运行时角色的 Amazon EMR 集群时，您可以集成 Lake Formation 权限。 运行时角色是您与 Amazon EMR 作业或查询关联的 IAM 角色，然后 Amazon EMR 会使用此角色访问 AWS 资源。

Lake Formation 与 AWS Key Management Service (AWS KMS) 集成，使您能够更轻松地设置这些集成服务，以加密和解密 Amazon Simple Storage Service (Amazon S3) 位置中的数据。

Lake Formation 的其他资源

主题

- 博客
- 技术讲座和网络研讨会
- 现代架构
- 数据网格资源
- 最佳实践指南

博客

- AWS Lake Formation 2022 年回顾
- 高弹性的多区域现代数据架构
- 使用 LF 标签进行跨账户共享，以引导 IAM 主体

- Lake Formation 权限清单控制面板
- 事件驱动型数据网格

技术讲座和网络研讨会

- re: Invent 2020 – 数据湖：轻松构建、保护安全以及和 AWS Lake Formation 共享
- re:Invent 2022 – Amazon S3 上构建和运行数据湖
- AWS 2022 年 SF 峰会 – 了解和实现现代数据架构
- AWS 2022 年 ATL 峰会 – 现代数据湖及 AWS Lake Formation、Amazon Redshift 和 AWS Glue
- AWS 2022 年 ANZ 峰会 – 数据湖、智能湖仓和数据网格：什么、为什么、如何？
- AWS在线技术讲座 – 简化数据湖中的权限和监管

现代架构

- 现代架构模式

数据网格资源

- 使用 AWS Lake Formation 基于标签的访问控制大规模构建现代数据架构和数据网格模式
- JPMorgan Chase 如何构建数据网格架构以大幅增加价值，从而增强其企业数据平台
- 在 AWS 上构建数据网格

最佳实践指南

- AWS Lake Formation 最佳实践指南

Lake Formation 入门

我们建议您首先阅读以下部分：

- AWS Lake Formation：工作方式 - 了解基本术语以及各个组件的交互方式。
- Lake Formation 入门 - 获取有关先决条件的信息，并完成重要的设置任务。
- 教程— 按照 step-by-step 教程学习如何使用 Lake Formation。

- AWS Lake Formation 中的安全性 - 了解如何帮助用户在 Lake Formation 中安全地访问数据。

Lake Formation 入门

如果您尚未注册 AWS 或需要入门帮助，请确保完成以下任务。

主题

完成初始 AWS 配置任务

要使用 AWS Lake Formation，必须首先完成以下任务：

主题

注册 AWS 账户

如果您没有AWS 账户，请完成以下步骤来创建一个。

要注册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。

   注册AWS 账户时，系统将会创建AWS 账户根用户。根用户有权访问该账户中的所有AWS 服务和资源。作为一种安全最佳实践，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。

AWS注册过程完成后，会发送一封确认电子邮件。任何时候您都可以通过转至 https://aws.amazon.com/ 并选择我的账户来查看当前的账户活动并管理账户。

创建管理用户

注册 AWS 账户 后，保护您的 AWS 账户根用户，启用 AWS IAM Identity Center，创建一个管理用户，以避免使用根用户执行日常任务。

保护AWS 账户根用户

1. 选择根用户并输入AWS 账户电子邮件地址，以账户拥有者身份登录AWS Management Console。在下一页上，输入密码。

   要获取使用根用户登录方面的帮助，请参阅《AWS 登录 用户指南》中的以根用户身份登录。

2. 对您的根用户启用多重身份验证（MFA）。

   有关说明，请参阅《IAM 用户指南》中的为 AWS 账户根用户启用虚拟 MFA 设备（控制台）。

创建管理用户

1. 启用 IAM Identity Center

   有关说明，请参阅《AWS IAM Identity Center 用户指南》中的启用 AWS IAM Identity Center。

2. 在 IAM Identity Center 中，为管理用户授予管理访问权限。

   有关将 IAM Identity Center 目录 用作身份源的教程，请参阅《AWS IAM Identity Center 用户指南》中的使用默认 IAM Identity Center 目录 配置用户访问权限。

作为管理用户登录

- 要使用 IAM Identity Center 用户身份登录，请使用在创建 IAM Identity Center 用户时发送到电子邮件地址的登录网址。

  要获取使用 IAM Identity Center 用户登录方面的帮助，请参阅《AWS 登录 用户指南》中的 登录 AWS 访问门户。

授权以编程方式访问

如果用户需要在 AWS Management Console 之外与 AWS 交互，则需要编程式访问权限。授予编程式访问权限的方法取决于访问 AWS 的用户类型。

要向用户授予编程式访问权限，请选择以下选项之一。

哪个用户需要编程式访问权限？	目的	方式
人力身份 （在 IAM Identity Center 中管理的用户）	使用临时凭证签署向 AWS CLI、AWS软件开发工具包或 AWS API 发出的编程请求。	按照您希望使用的界面的说明进行操作。 • 有关 AWS CLI 的更多信息，请参阅《AWS Command Line Interface 用户指南》中的配置 AWS CLI 以使用 AWS IAM Identity Center。 • 有关 AWS软件开发工具包、工具和 AWS API 的更多信息，请参阅《AWS软件开发工具包和工具参考指南》中的 IAM Identity Center 身份验证。
IAM	使用临时凭证签署向 AWS CLI、AWS软件开发工具包或 AWS API 发出的编程请求。	按照《IAM 用户指南》中将临时凭证用于 AWS 资源中的说明进行操作。
IAM	（不推荐使用） 使用长期凭证签署向 AWS CLI、AWS软件开发工具包或 AWS API 发出的编程请求。	按照您希望使用的界面的说明进行操作。 • 有关 AWS CLI 的更多信息，请参阅《AWS Command Line Interface 用户指南》中的使用 IAM 用户凭证进行身份验证。

哪个用户需要编程式访问权限？	目的	方式
		• 有关 AWS 软件开发工具包和工具的更多信息，请参阅《AWS 软件开发工具包和工具参考指南》中的使用长期凭证进行身份验证。 • 有关 AWS API 的更多信息，请参阅《IAM 用户指南》中的管理 IAM 用户的访问密钥。

设置 AWS Lake Formation

以下各节提供有关首次设置 Lake Formation 的信息。要开始使用 Lake Formation，并非本节中的所有主题都是必须学习的。您可以按照说明在 Amazon Simple Storage Service (Amazon S3) 中设置 Lake Formation 权限模型，以管理您的现有 AWS Glue Data Catalog 对象和数据位置。

1. 创建数据湖管理员
2. 更改默认权限模式或使用混合访问模式
3. the section called “为您的数据湖配置 Amazon S3 位置”
4. the section called “为 Lake Formation 用户分配权限”
5. the section called “集成 IAM Identity Center”
6. the section called “（可选）外部数据筛选设置”
7. the section called “（可选）授予对数据目录加密密钥的访问权限”
8. （可选）为工作流程创建 IAM 角色

本节将介绍如何通过两种不同的方式设置 Lake Formation 资源：

- 使用 AWS CloudFormation 模板
- 使用 Lake Formation 控制台

要使用 AWS 控制台设置 Lake Formation，请转至创建数据湖管理员。

使用 AWS CloudFormation 模板设置 Lake Formation 资源

> **Note**
>
> AWS CloudFormation堆栈执行上述步骤 1 到 6，但步骤 2 和 5 除外。在 Lake Formation 控制台中the section called “集成 IAM Identity Center”手动执行更改默认权限模式或使用混合访问模式和操作。

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/cloudformation 以 IAM 管理员身份登录 AWS CloudFormation 控制台。
2. 选择启动堆栈。
3. 在创建堆栈屏幕上，选择下一步。
4. 输入堆栈名称。
5. 对于DatalakeAdminName和 DatalakeAdminPassword，请输入您的数据湖管理员用户的用户名和密码。
6. 对于 DatalakeUser1Name 和 DatalakeUser1Passwor d，输入您的数据湖分析师用户的用户名和密码。
7. 对于 DataLakeBucketName，输入要创建的新存储桶名称。
8. 选择下一步。
9. 在下一页上，选择下一步。
10. 查看最后页面上的详细信息，然后选择我确认 AWS CloudFormation 可以创建 IAM 资源。
11. 选择创建。

    堆栈创建过程可能需要几分钟时间才能完成。

清理资源

如果您想清除 AWS CloudFormation 堆栈资源，请执行以下操作：

1. 取消注册您的堆栈创建并注册为数据湖位置的 Amazon S3 存储桶。
2. 删除 AWS CloudFormation 堆栈 这将删除堆栈创建的所有资源。

创建数据湖管理员

数据湖管理员最初是唯一可以向任何主体（包括自己）授予对数据位置和数据目录资源的 Lake Formation 权限的 AWS Identity and Access Management (IAM) 用户或角色。有关数据湖管理员功能的信息，请参阅隐式 Lake Formation 权限。默认情况下，Lake Formation 允许您最多创建 30 个数据湖管理员。

您可以使用 Lake Formation 控制台或 Lake Formation API 的 `PutDataLakeSettings` 操作来创建数据湖管理员。

创建数据湖管理员需要以下权限。`Administrator` 用户隐式拥有这些权限。

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

如果您向用户授予 `AWSLakeFormationDataAdmin` 策略，则该用户将无法创建其他 Lake Formation 管理员用户。

创建数据湖管理员（控制台）

1. 如果要成为数据湖管理员的用户尚不存在，请使用 IAM 控制台创建该用户。或者，请选择现有用户来担任数据湖管理员。

   > **Note**
   >
   > 我们建议您不要选择 IAM 管理用户（拥有 `AdministratorAccess` AWS 托管策略的用户）作为数据湖管理员。

   将以下 AWS 托管策略附加到用户：

策略	必需？	注意事项
`AWSLakeFormationDataAdmin`	强制性	基本数据湖管理员权限。此 AWS 托管策略包含显式拒绝 Lake Formation API 操作 `PutDataLakeSetting` ，该操作可限制用户创建新的数据湖管理员。

策略	必需？	注意事项
AWSGlueConsoleFullAccess ，CloudWatchLogsReadOnlyAccess	可选	如果数据湖管理员要对通过 Lake Formation 蓝图创建的工作流进行故障排除，请附加这些策略。这些策略使数据湖管理员能够在 AWS Glue 控制台和 Amazon CloudWatch Logs 控制台中查看故障排除信息。有关工作流的更多信息，请参阅the section called “使用工作流导入数据”。
AWSLakeFormationCrossAccountManager	可选	附加此策略可使数据湖管理员能够授予和撤销对数据目录资源的跨账户权限。有关更多信息，请参阅Lake Formation 中的跨账户数据共享。
AmazonAthenaFullAccess	可选	如果数据湖管理员要在 Amazon Athena 中运行查询，请附加此策略。

2. 附加以下内联策略，该策略向数据湖管理员授予创建 Lake Formation 服务相关角色的权限。建议将该策略命名为 LakeFormationSLR。

   服务相关角色使数据湖管理员能够更轻松地在 Lake Formation 中注册 Amazon S3 位置。有关 Lake Formation 服务相关角色的更多信息，请参阅the section called “使用服务相关角色”。

   > **Important**
   >
   > 在以下所有策略中，将 *<account-id>* 替换为有效的 AWS 账户编号。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
```

```
                    "iam:AWSServiceName": "lakeformation.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PutRolePolicy"
            ],
            "Resource": "arn:aws:iam::<account-id>:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess"
        }
    ]
}
```

3. （可选）将下面的 PassRole 内联策略附加到用户。此策略使数据湖管理员能够创建和运行工作流。iam:PassRole 权限使工作流能够担任创建爬网程序和作业的角色 LakeFormationWorkflowRole，并将该角色附加到所创建的爬网程序和作业。建议将该策略命名为 UserPassRole。

> **Important**
>
> 将 *<account-id>* 替换为有效的 AWS 账户编号。

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

4. （可选）如果您的账户将授予或接收跨账户 Lake Formation 权限，请附加此额外内联策略。此策略使数据湖管理员能够查看和接受 AWS Resource Access Manager (AWS RAM) 资源共享邀请。此外，对于 AWS Organizations 管理账户中的数据湖管理员，该策略还包括针对组织启用跨账户授权的权限。有关更多信息，请参阅Lake Formation 中的跨账户数据共享。

   建议将该策略命名为 RAMAccess。

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

5. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台，并以您在创建管理用户中创建的管理员用户或使用 AdministratorAccess 用户 AWS 托管策略的用户身份登录。
6. 如果显示欢迎使用 Lake Formation 窗口，请选择您在第 1 步创建或选择的 IAM 用户，然后选择开始使用。
7. 如果没有看到欢迎使用 Lake Formation 窗口，请执行以下步骤来配置 Lake Formation 管理员。
   a. 在导航窗格中的管理员下，选择管理角色和任务。在控制台页面的数据湖管理员部分，选择添加。
   b. 在添加管理员对话框中的“访问类型”下，选择数据湖管理员。
   c. 对于 IAM 用户和角色，选择您在第 1 步创建或选择的 IAM 用户，然后选择保存。

更改默认权限模式或使用混合访问模式

Lake Formation 一开始就启用了“仅使用 IAM 访问控制”设置，以便与现有 AWS Glue Data Catalog 行为兼容。此设置使您可以通过 IAM 策略和 Amazon S3 存储桶策略管理对数据湖中数据及其元数据的访问。

为了简化数据湖权限从 IAM 和 Amazon S3 模式向 Lake Formation 权限的过渡，我们建议您对数据目录使用混合访问模式。在混合访问模式下，您现在有了增量路径，允许您为一组特定的用户启用 Lake Formation 权限，而不会中断其他现有用户或工作负载。

有关更多信息，请参阅混合访问模式。

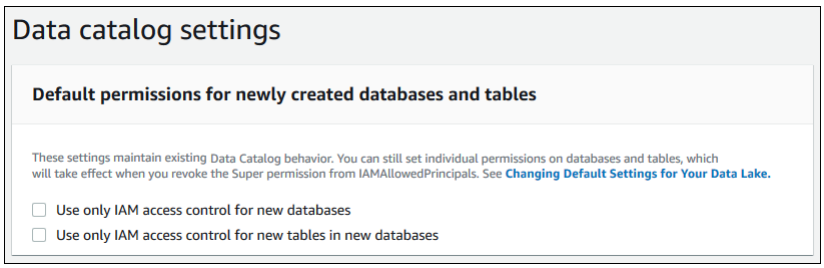
禁用默认设置，只需一步即可将表的所有现有用户移至 Lake Formation。

> **Important**
>
> 如果您已有 AWS Glue Data Catalog 数据库和表，请不要按照本节中的说明执行操作。而是应按照the section called “将 AWS Glue 数据权限升级为 Lake Formation 模型”中的说明操作。

> **Warning**
>
> 如果您实施自动化功能来在数据目录中创建数据库和表，则以下步骤可能会导致自动化和下游提取、转换、加载 (ETL) 作业失败。只有在修改现有流程或向所需主体授予显式 Lake Formation 权限后，才能继续操作。有关 Lake Formation 权限的信息，请参阅 the section called “Lake Formation 权限参考”。

更改默认数据目录设置

1. 通过 https://console.aws.amazon.com/lakeformation/ 继续在 Lake Formation 控制台中操作。确保以您在创建管理用户中创建的管理员用户或使用 `AdministratorAccess` AWS 托管策略的用户身份登录。
2. 修改数据目录设置：
   a. 在导航窗格中的管理下，选择数据目录设置。
   b. 清除两个复选框，然后选择保存。



3. 撤销数据库创建者的 `IAMAllowedPrincipals` 权限。

   a. 在导航窗格中的管理下，选择管理角色和任务。

   b. 在管理角色和任务控制台页面的数据库创建者部分，选择 `IAMAllowedPrincipals` 组，然后选择撤销。

      此时将出现撤销权限对话框，其中显示 `IAMAllowedPrincipals` 拥有创建数据库权限。

   c. 选择撤销。

为 Lake Formation 用户分配权限

创建用户以访问 AWS Lake Formation 中的数据湖。此用户拥有查询数据湖的最低权限。

有关创建用户或组的更多信息，请参阅《IAM 用户指南》中的 IAM 身份。

向非管理员用户附加用于访问 Lake Formation 数据的权限

1. 以您在创建管理用户中创建的管理员用户或使用 `AdministratorAccess` AWS 托管策略的用户身份通过 https://console.aws.amazon.com/iam 打开 IAM 控制台。
2. 选择用户或用户组。
3. 在列表中，请选择要在其中嵌入策略的用户或组的名称。

   选择权限。
4. 选择添加权限，然后选择直接附加策略。在筛选策略文本字段中输入 `Athena`。在结果列表中，选中 `AmazonAthenaFullAccess` 的复选框。
5. 选择创建策略按钮。在创建策略页面上，选择 JSON 选项卡。复制以下策略并将其粘贴到策略编辑器中。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
            ],
            "Resource": "*"
        }
    ]
}
```

6. 选择底部的下一步按钮，直到看到查看策略页面。输入策略名称，例如 `DatalakeUserBasic`。选择创建策略，然后关闭策略选项卡或浏览器窗口。

为您的数据湖配置 Amazon S3 位置

要使用 Lake Formation 管理和保护数据湖中的数据，您必须先注册一个 Amazon S3 位置。注册位置时，会注册 Amazon S3 路径和该路径下的所有文件夹，这使 Lake Formation 能够强制实施存储级别权限。当用户从 Amazon Athena 等集成引擎请求数据时，Lake Formation 会提供数据访问权限，而不是使用用户权限。

注册位置时，您需要指定一个 IAM 角色以授予对该位置的读/写权限。Lake Formation 担任该角色，向请求访问已注册 Amazon S3 位置的数据的集成 AWS 服务提供临时凭证。您可以指定 Lake Formation 服务相关角色 (SLR) 或创建自己的角色。

在以下情况下使用自定义角色：

- 您计划在 Amazon CloudWatch 日志中发布指标。除了 SLR 权限外，用户定义的角色还必须包括用于在 CloudWatch 日志中添加日志和发布指标的策略。有关授予必要 CloudWatch 权限的内联策略示例，请参阅用于注册位置的角色的要求。
- Amazon S3 位置位于其他账户中。有关更多信息，请参阅 the section called “在其他 AWS 账户中注册 Amazon S3 位置”。

- Amazon S3 位置包含使用 AWS 托管式密钥加密的数据。有关详细信息，请参阅 注册加密的 Amazon S3 位置 和 跨 AWS 账户注册加密的 Amazon S3 位置。
- 您计划使用 Amazon EMR 访问 Amazon S3 位置。有关角色要求的更多信息，请参阅《Amazon EMR 管理指南》中的适用于 Lake Formation 的 IAM 角色。

所选角色必须具有必需的权限，如用于注册位置的角色的要求中所述。有关如何注册 Amazon S3 位置的说明，请参阅向数据湖添加 Amazon S3 位置。

（可选）外部数据筛选设置

如果您打算使用第三方查询引擎分析和处理数据湖中的数据，则必须选择允许外部引擎访问由 Lake Formation 管理的数据。如果您不选择，则外部引擎将无法访问已在 Lake Formation 中注册的 Amazon S3 位置处的数据。

Lake Formation 支持使用列级别权限来限制对表中特定列的访问。Amazon Athena、Amazon Redshift Spectrum 和 Amazon EMR 等集成分析服务可从 AWS Glue Data Catalog 中检索未经过筛选的表元数据。集成服务负责对查询响应中的列进行实际筛选。第三方管理员负责妥善处理权限，以免有人未经授权便访问数据。

选择允许第三方引擎访问和筛选数据（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 继续在 Lake Formation 控制台中操作。确保您以拥有对 Lake Formation `PutDataLakeSettings` API 操作的 IAM 权限的主体身份登录。您在 注册 AWS 账户 中创建的 IAM 管理员用户拥有此权限。
2. 在导航窗格中的管理下，选择应用程序集成设置。
3. 在应用程序集成设置页面上，执行以下操作：
   a. 选中允许外部引擎筛选在 Lake Formation 中注册的 Amazon S3 位置处的数据复选框。
   b. 输入为第三方引擎定义的会话标签值。
   c. 对于 AWS 账户 ID，输入允许第三方引擎从中访问在 Lake Formation 中注册的位置的账户 ID。在输入每个账户 ID 之后按 Enter 键。
   d. 选择保存。

要允许外部引擎在不进行会话标签验证的情况下访问数据，请参阅集成应用程序以获取完整表访问权限

（可选）授予对数据目录加密密钥的访问权限

如果 AWS Glue Data Catalog 已加密，则应向需要授予对数据目录数据库和表的 Lake Formation 权限的所有主体授予对 AWS KMS 密钥的 AWS Identity and Access Management (IAM) 权限。

有关更多信息，请参见AWS Key Management Service 开发人员指南。

（可选）为工作流程创建 IAM 角色

借助 AWS Lake Formation，您可以使用 AWS Glue 爬网程序执行的工作流导入数据。工作流定义数据来源和计划以将数据导入到数据湖中。您可以使用 Lake Formation 提供的蓝图或模板轻松定义工作流。

创建工作流时，必须为其分配一个 AWS Identity and Access Management (IAM) 角色，该角色授予摄取数据所需的 Lake Formation 权限。

以下过程假定您熟悉 IAM。

为工作流创建 IAM 角色

1. 以您在创建管理用户中创建的管理员用户或使用 `AdministratorAccess` AWS 托管策略的用户身份通过 https://console.aws.amazon.com/iam 登录 IAM 控制台。
2. 在导航窗格中，选择角色，然后选择创建角色。
3. 在创建角色页面上，选择 AWS 服务，然后选择 Glue。选择下一步。
4. 在添加权限页面上，搜索AWSGlueServiceRole托管策略，然后选中列表中策略名称旁边的复选框。然后完成创建角色向导，命名角色 `LakeFormationWorkflowRole`。要完成操作，请选择创建角色。
5. 回到角色页面上，搜索 `LakeFormationWorkflowRole`，然后选择该角色名称。
6. 在角色摘要页面上的权限选项卡下，选择添加内联策略。在创建策略屏幕上，导航到 JSON 选项卡，然后添加以下内联策略。建议将该策略命名为 `LakeFormationWorkflow`。

> **Important**
>
> 在以下策略中，将 *<account-id>* 替换为有效的 AWS 账户 数字。

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                 "lakeformation:GetDataAccess",
                 "lakeformation:GrantPermissions"
             ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": ["iam:PassRole"],
            "Resource": [
                "arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

以下是关于该策略中权限的简要描述：

- `lakeformation:GetDataAccess` 使工作流创建的作业能够向目标位置写入数据。
- `lakeformation:GrantPermissions` 使工作流能够授予对目标表的 `SELECT` 权限。
- `iam:PassRole` 使服务能够担任创建爬网程序和作业（工作流实例）的角色 `LakeFormationWorkflowRole`，并将该角色附加到所创建的爬网程序和作业。

7. 验证角色 `LakeFormationWorkflowRole` 是否附加了两个策略。
8. 如果您要摄取数据湖位置之外的数据，请添加用于授予来源数据读取权限的内联策略。

将 AWS Glue 数据权限升级为 AWS Lake Formation 模型

AWS Lake Formation 权限支持对数据湖中的数据进行精细访问控制。您可以在 Amazon Simple Storage Service (Amazon S3) 中使用 Lake Formation 权限模型来管理您的现有 AWS Glue Data Catalog 对象和数据位置。

Lake Formation 权限模型使用粗粒度 AWS Identity and Access Management (IAM) 权限访问 API 服务。它限制了您的用户和这些服务通过 Lake Formation 功能可以访问的数据。相比之下，AWS Glue 模型通过精细访问控制 IAM 权限授予数据访问权限。要切换模型，请按照本指南中的步骤操作。

有关更多信息，请参阅 Lake Formation 权限概述 。

主题

关于升级为 Lake Formation 权限模型

为了保持与 AWS Glue 的向后兼容性，默认情况下，AWS Lake Formation 会向 `IAMAllowedPrincipals` 组授予对所有现有 AWS Glue 数据目录资源的 `Super` 权限，并且如果启用了仅使用 IAM 访问控制设置，则授予对新数据目录资源的 `Super` 权限。这实际上会导致对数据目录资源和 Amazon S3 位置的访问仅由 AWS Identity and Access Management (IAM) 策略控制。`IAMAllowedPrincipals` 组包括 IAM 策略允许访问数据目录对象的任何 IAM 用户和角色。`Super` 权限使主体能够对被授予该权限的数据库或表执行所有支持的 Lake Formation 操作。

您可以通过在 Lake Formation 中注册现有数据目录资源的位置或使用混合访问模式，开始使用 Lake Formation 来管理对数据的访问。当您在混合访问模式下注册 Amazon S3 位置时，您可以通过选择该位置下的数据库和表的主体来启用 Lake Formation 权限。

为了简化数据湖权限从 IAM 和 Amazon S3 模型向 Lake Formation 权限的过渡，我们建议您对数据目录使用混合访问模式。在混合访问模式下，您现在有了增量路径，允许您为一组特定的用户启用 Lake Formation 权限，而不会中断其他现有用户或工作负载。

有关更多信息，请参阅 混合访问模式。

禁用默认数据目录设置，只需一步即可将表的所有现有用户移至 Lake Formation。

要开始对现有 AWS Glue 数据目录数据库和表使用 Lake Formation 权限，您必须执行以下操作：

1. 确定用户具有的对每个数据库和表的现有 IAM 权限。
2. 在 Lake Formation 中复制这些权限。
3. 对于包含数据的每个 Amazon S3 位置：

a. 撤销 IAMAllowedPrincipals 组对引用该位置的每个数据目录资源的 Super 权限。

b. 向 Lake Formation 注册该位置。

4. 清理现有的精细访问控制 IAM 策略。

> **Important**
>
> 要在过渡数据目录的过程中添加新用户，您必须像之前一样在 IAM 中设置精细 AWS Glue 权限，还必须在 Lake Formation 中复制这些权限，如本部分所述。如果新用户具有本指南中所述的粗粒度 IAM 策略，则可以列出具有已向 IAMAllowedPrincipals 授予的 Super 权限的任何数据库或表。他们还可以查看这些资源的元数据。

按照本部分中的步骤升级为 Lake Formation 权限模型。首先是the section called “步骤 1：列出现有权限”。

步骤 1：列出用户和角色的现有权限

要开始将 AWS Lake Formation 权限用于您的现有 AWS Glue 数据库和表，必须先确定用户的现有权限。

> **Important**
>
> 在开始之前，请确保您已完成 开始使用中的任务。

主题

使用 API 操作

使用 AWS Identity and Access Management (IAM) ListPoliciesGrantingServiceAccess API 操作确定附加到每个主体（用户或角色）的 IAM 策略。根据结果中返回的策略，您可以确定授予主体的 IAM 权限。您必须分别为每个主体调用该 API。

Example

以下 AWS CLI 示例返回附加到用户 glue_user1 的策略。

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

该命令返回类似于以下内容的结果。

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

使用 AWS Management Console

您还可以在 AWS Identity and Access Management (IAM) 控制台的用户或角色摘要页面上的访问顾问选项卡中查看此信息：

1. 通过以下网址打开 IAM 控制台：https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择用户或角色。
3. 在列表中选择一个名称以打开其摘要页面，然后选择访问顾问选项卡。
4. 检查每个策略以确定每个用户有权访问的数据库、表和操作的组合。

请记住，在此过程中，除了检查用户之外，还要检查角色，因为数据处理作业可能会代入访问数据的角色。

使用 AWS CloudTrail

确定您的现有权限的另一种方法是在 AWS CloudTrail 中查找 AWS Glue API 调用，其中日志的 `additionaleventdata` 字段包含 `insufficientLakeFormationPermissions` 条目。此条目列出了用户需要对其具有 Lake Formation 权限才能执行相同操作的数据库和表。

这些是数据访问日志，因此不能保证它们会生成用户及其权限的完整列表。我们建议选择较宽的时间范围来捕获大多数用户的数据访问模式，例如几周或几个月。

有关更多信息，请参阅《AWS CloudTrail 用户指南》中的使用 CloudTrail 事件历史记录查看事件。

接下来，您可以设置 Lake Formation 权限以匹配 AWS Glue 权限。请参阅步骤 2：设置等效的 Lake Formation 权限。

步骤 2：设置等效的 Lake Formation 权限

使用在步骤 1：列出用户和角色的现有权限中收集的信息，授予 AWS Lake Formation 权限以匹配 AWS Glue 权限。使用以下任一方法进行授予：

- 使用 Lake Formation 控制台或 AWS CLI。

  请参阅the section called “授予和撤销数据目录权限”。
- 使用 `GrantPermissions` 和 `BatchGrantPermissions` API 操作。

  请参阅权限 API。

有关更多信息，请参阅 Lake Formation 权限概述 。

设置 Lake Formation 权限后，继续执行步骤 3：向用户授予 IAM 权限以使用 Lake Formation。

步骤 3：向用户授予 IAM 权限以使用 Lake Formation

要使用 AWS Lake Formation 权限模型，主体必须对 Lake Formation API 具有 AWS Identity and Access Management (IAM) 权限。

在 IAM 中创建以下策略，并将其附加到需要访问您的数据湖的每个用户。将该策略命名为 `LakeFormationDataAccess`。

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

接下来，升级为 Lake Formation 权限，一次升级一个数据位置。请参阅步骤 4：将数据存储切换到 Lake Formation 权限模型。

步骤 4：将数据存储切换到 Lake Formation 权限模型

升级为 Lake Formation 权限，一次升级一个数据位置。为此，请重复整个部分，直到您注册了数据目录引用的所有 Amazon Simple Storage Service (Amazon S3) 路径。

主题

- 验证 Lake Formation 权限
- 保护现有数据目录资源
- 为您的 Amazon S3 位置启用 Lake Formation 权限

验证 Lake Formation 权限

在注册位置之前，请执行验证步骤，以确保正确的主体具有所需的 Lake Formation 权限，并且不会向不应具有 Lake Formation 权限的主体授予任何这些权限。使用 Lake Formation `GetEffectivePermissionsForPath` API 操作，标识引用 Amazon S3 位置的数据目录资源，以及对这些资源具有权限的主体。

以下 AWS CLI 示例返回引用 Amazon S3 存储桶 `products` 的数据目录数据库和表。

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

请注意 profile 选项。我们建议您以数据湖管理员身份运行该命令。

以下是返回结果的摘录。

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

> **Important**
>
> 如果您的 AWS Glue 数据目录已加密，则 GetEffectivePermissionsForPath 仅返回在 Lake Formation 正式发布后创建或修改的数据库和表。

保护现有数据目录资源

接下来，撤销 IAMAllowedPrincipals 对您为该位置标识的每个表和数据库具有的 Super 权限。

> **Warning**
>
> 如果您实施自动化功能来在数据目录中创建数据库和表，则以下步骤可能会导致自动化和下游提取、转换和加载 (ETL) 作业失败。只有在修改现有流程或向所需主体授予显式 Lake

Formation 权限后，才能继续操作。有关 Lake Formation 权限的信息，请参阅 the section called “Lake Formation 权限参考”。

撤销 **IAMAllowedPrincipals** 对表具有的 **Super** 权限

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以数据湖管理员身份登录。
2. 在导航窗格中，选择表。
3. 在表页面上，选中所需表旁边的单选按钮。
4. 在操作菜单上，选择撤销。
5. 在撤销权限对话框的 IAM 用户和角色列表中，向下滚动到组标题，然后选择 IAMAllowedPrincipals。
6. 在表权限下，确保选中 Super 权限，然后选择撤销。

撤销 **IAMAllowedPrincipals** 对数据库具有的 **Super** 权限

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以数据湖管理员身份登录。
2. 在导航窗格中，选择数据库。
3. 在数据库页面上，选中所需数据库旁边的单选按钮。
4. 在操作 菜单上，选择编辑。
5. 在编辑数据库页面上，清除仅对此数据库中的新表使用 IAM 访问控制，然后选择保存。
6. 返回数据库页面，确保数据库仍处于选中状态，然后在操作菜单上选择撤销。
7. 在撤销权限对话框的 IAM 用户和角色列表中，向下滚动到组标题，然后选择 IAMAllowedPrincipals。
8. 在数据库权限下，确保选中 Super 权限，然后选择撤销。

为您的 Amazon S3 位置启用 Lake Formation 权限

接下来，向 Lake Formation 注册 Amazon S3 位置。为此，您可以使用向数据湖添加 Amazon S3 位置中所述的过程。或者，使用 `RegisterResource` API 操作，如凭证售卖 API 中所述。

> **Note**
>
> 如果注册了父位置，则无需注册子位置。

完成这些步骤并测试用户是否可以访问其数据后，您即已成功升级为 Lake Formation 权限。继续执行下一步：步骤 5：保护新的数据目录资源。

步骤 5：保护新的数据目录资源

接下来，通过更改默认数据目录设置来保护所有新的数据目录资源。关闭仅对新数据库和表使用 AWS Identity and Access Management (IAM) 访问控制的选项。

> **Warning**
>
> 如果您实施自动化功能来在数据目录中创建数据库和表，则以下步骤可能会导致自动化和下游提取、转换和加载 (ETL) 作业失败。只有在修改现有流程或向所需主体授予显式 Lake Formation 权限后，才能继续操作。有关 Lake Formation 权限的信息，请参阅 the section called “Lake Formation 权限参考”。

更改默认数据目录设置

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以 IAM 管理用户（用户 `Administrator` 或具有 `AdministratorAccess` AWS 托管式策略的其他用户）身份登录。
2. 在导航窗格中，选择设置。
3. 在数据目录设置页面上，清除这两个复选框，然后选择保存。

下一步是向用户授予将来访问其他数据库或表的权限。请参阅步骤 6：向用户提供新的 IAM 策略，以便将来访问数据湖。

步骤 6：向用户提供新的 IAM 策略，以便将来访问数据湖

要向您的用户授予将来访问其他数据目录数据库或表的权限，您必须向他们提供以下粗粒度 AWS Identity and Access Management (IAM) 内联策略。将该策略命名为 `GlueFullReadAccess`。

Important

如果您在撤销 IAMAllowedPrincipals 对您的数据目录中每个数据库和表具有的 Super 权限之前将此策略附加到用户，则该用户可以查看向 IAMAllowedPrincipals 授予对其的 Super 权限的任何资源的所有元数据。

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

Note

本步骤和前面步骤中指定的内联策略包含最低 IAM 权限。有关数据湖管理员、数据分析师和其他角色的建议策略，请参阅 the section called “Lake Formation 角色和 IAM 权限参考”。

接下来继续执行步骤 7：清理现有 IAM 策略。

步骤 7：清理现有 IAM 策略

设置 AWS Lake Formation 权限并创建并附加粗粒度访问控制 AWS Identity and Access Management (IAM) 策略后，请完成以下最后一步：

- 从用户、组和角色中移除您在 Lake Formation 中复制的旧的精细访问控制 IAM 策略。

这样，您可以确保这些主体不再能够直接访问 Amazon Simple Storage Service (Amazon S3) 中的数据。然后，您可以完全通过 Lake Formation 管理这些主体的数据湖访问权限。

AWS Lake Formation 和接口 VPC 端点 (AWS PrivateLink)

Amazon VPC 是一项 AWS 服务，可用来启动在虚拟网络中定义的 AWS 资源。借助 VPC，您可以控制您的网络设置，如 IP 地址范围、子网、路由表和网络网关。

如果您使用 Amazon Virtual Private Cloud (Amazon VPC) 托管 AWS 资源，则可以在您的 VPC 和 Lake Formation 之间建立私有连接。您可以使用此连接，以便 Lake Formation 可以与 VPC 中的资源进行通信，而无需访问公共 Internet。

您可以通过创建接口 VPC 端点在 VPC 和 AWS Lake Formation 之间建立私有连接。接口端点由 AWS PrivateLink 提供支持，该技术支持您通过私有连接访问 Lake Formation API，而无需采用互联网网关、NAT 设备、VPN 连接或 AWS Direct Connect 连接。VPC 中的实例不需要公有 IP 地址即可与 Lake Formation API 进行通信。您的 VPC 和 Lake Formation 之间的流量不会脱离 Amazon 网络。

每个接口端点均由子网中的一个或多个弹性网络接口表示。

有关更多信息，请参阅《Amazon VPC 用户指南》中的接口 VPC 端点 (AWS PrivateLink)。

Lake Formation VPC 端点的注意事项

请务必先查看《Amazon VPC 用户指南》中的接口端点属性和限制，然后再为 Lake Formation 设置接口 VPC 端点。

Lake Formation 支持从您的 VPC 调用其所有 API 操作。您可以在支持 Lake Formation 和 Amazon VPC 端点的所有 AWS 区域中将 Lake Formation 与 VPC 端点结合使用。

为 Lake Formation 创建接口 VPC 端点

您可以使用 Amazon VPC 控制台或 AWS Command Line Interface (AWS CLI) 为 Lake Formation 服务创建 VPC 端点。有关更多信息，请参阅《Amazon VPC 用户指南》中的创建接口端点。

使用以下服务名称为 Lake Formation 创建 VPC 端点：

- com.amazonaws.*region*.lakeformation

如果为端点启用私有 DNS，则可以使用该区域的默认 DNS 名称，向 Lake Formation 发送 API 请求，例如 `lakeformation.us-east-1.amazonaws.com`。

有关更多信息，请参阅《Amazon VPC 用户指南》中的通过接口端点访问服务。

为 Lake Formation 创建 VPC 端点策略

Lake Formation 支持 VPC 端点策略。VPC 端点策略是一种 AWS Identity and Access Management (IAM) 资源策略，您在创建或修改端点时可将该策略附加到端点。

您可以将端点策略附加到控制对 Lake Formation 的访问的 VPC 端点。该策略指定以下信息：

- 可执行操作的主体。
- 可执行的操作。
- 可对其执行操作的资源。

有关更多信息，请参阅《Amazon VPC 用户指南》中的使用 VPC 端点控制对服务的访问。

示例：Lake Formation 操作的 VPC 端点策略

以下 Lake Formation 的 VPC 端点策略示例允许使用 Lake Formation 权限进行凭证售卖。您可以使用此策略从位于私有子网中的 Amazon Redshift 集群或 Amazon EMR 集群使用 Lake Formation 权限运行查询。

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

> **Note**
>
> 如果在创建端点时未附加策略，则会附加允许对服务进行完全访问的默认策略。

有关更多信息，请参阅《Amazon VPC 文档》中的以下主题：

- 什么是 Amazon VPC？
- 创建接口端点
- 使用 VPC 端点策略

教程

以下教程分为三个部分，提供有关如何使用 AWS Lake Formation 构建数据湖、摄取数据、共享和保护数据湖的分步说明：

1. 构建数据湖并摄取数据：学习构建数据湖并使用蓝图移动、存储、编目、清除和整理数据。您还将学习设置受管控表。受管控表是一种新的 Amazon S3 表类型，它支持原子性、一致性、隔离性和持久性 (ACID) 事务。

   开始之前，请确保您已完成Lake Formation 入门中的步骤。

   - 从 AWS CloudTrail 源头创建数据湖

     使用您自己的 CloudTrail 日志作为数据来源，创建并加载您的第一个数据湖。

   - 在 Lake Formation 中从 JDBC 来源创建数据湖

     使用其中一个可访问 JDBC 的数据存储（例如关系数据库）作为数据来源来创建数据湖。

2. 保护数据湖：学习使用基于标签的和行级别访问控制来有效地保护和管理对数据湖的访问。

   - 在 Lake Formation 中为开放表存储格式设置权限

     本教程演示了如何在 Lake Formation 中设置对开源事务表格格式（Apache Iceberg、Apache Hudi 和 Linux Foundation Delta Lake 表）的权限。

   - 使用 Lake Formation 基于标签的访问控制管理数据湖

     学习在 Lake Formation 中使用基于标签的访问控制来管理对数据湖中数据的访问。

   - 使用行级别访问控制保护数据湖

     学习设置行级别权限，以便允许您根据 Lake Formation 中的数据合规性和监管策略限制对特定行的访问。

3. 共享数据：学习使用基于标签的访问控制 (TBAC) 安全地跨 AWS 账户共享您的数据，并管理对在 AWS 账户之间共享的数据集的精细权限。

   - 使用 Lake Formation 基于标签的访问控制和命名资源共享数据湖

     在本教程中，您将学习如何使用 Lake Formation 安全地跨 AWS 账户共享数据。

   - 使用 Lake Formation 细粒度访问控制共享数据湖

     在本教程中，您将学习如何在使用 AWS Organizations 管理多个 AWS 账户时使用 Lake Formation 快速轻松地共享数据集。

主题

从 AWS CloudTrail 源头创建数据湖

本教程将指导您完成在 Lake Formation 控制台上执行的操作，以便从 AWS CloudTrail 源头创建和加载您的第一个数据湖。

创建数据湖的主要步骤

1. 将 Amazon Simple Storage Service (Amazon S3) 路径注册为数据湖。
2. 授予向数据目录以及数据湖中的 Amazon S3 位置写入数据的 Lake Formation 权限。
3. 创建数据库以整理数据目录中的元数据表。
4. 使用蓝图创建工作流。运行工作流以从数据来源摄取数据。
5. 设置您的 Lake Formation 权限，以允许其他人管理数据目录和数据湖中的数据。
6. 设置 Amazon Athena，以查询您导入到 Amazon S3 数据湖中的数据。
7. 对于一些数据存储类型，设置 Amazon Redshift Spectrum 以查询您导入到 Amazon S3 数据湖中的数据。

主题

- 第 5 步：授予数据位置权限
- 第 6 步：在数据目录中创建数据库
- 第 7 步：授予数据权限
- 第 8 步：使用蓝图创建工作流
- 第 9 步：运行工作流
- 第 10 步：授予对表的选择权限
- 第 11 步：使用 Amazon Athena查询数据湖

目标受众

下表列出了本教程中用于创建数据湖的角色。

目标受众

角色	描述
IAM 管理员	有 AWS 托管策略:`AdministratorAccess` . 可以创建 IAM 角色和 Amazon S3 存储桶。
数据湖管理员	可以访问数据目录、创建数据库以及向其他用户授予 Lake Formation 权限的用户。拥有的 IAM 权限比 IAM 管理员少，但足以管理数据湖。
数据分析人员	可以对数据湖运行查询的用户。拥有的权限仅足以运行查询。
工作流角色	具有运行工作流所需的 IAM 策略的角色。有关更多信息，请参阅（可选）为工作流程创建 IAM 角色。

先决条件

开始前的准备工作：

- 请确保您已完成设置 AWS Lake Formation中的任务。
- 知道你的 CloudTrail 日志的位置。

- Athena 要求数据分析师角色在使用 Athena 之前创建一个 Amazon S3 存储桶来存储查询结果。

假设熟悉 AWS Identity and Access Management (IAM)。有关 IAM 的信息，请参阅 IAM 用户指南。

第 1 步：创建数据分析师用户

此用户拥有查询数据湖所需的一组最低权限。

1. 使用 https://console.aws.amazon.com/iam 打开 IAM 控制台。以您在中创建的管理员用户创建管理用户或使用AdministratorAccess AWS 托管策略的用户身份登录。
2. 使用以下设置创建名为 datalake_user 的用户：

   - 启用 AWS Management Console 访问权限。
   - 设置密码，不需要重置密码。
   - 附加AmazonAthenaFullAccess AWS 托管策略。
   - 附加下面的内联策略。将该策略命名为 DatalakeUserBasic。

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

步骤 2：向工作流程角色添加读取 AWS CloudTrail 日志的权限

1. 将下面的内联策略附加到角色 LakeFormationWorkflowRole。该策略授予读取您的 AWS CloudTrail 日志的权限。将该策略命名为 DatalakeGetCloudTrail。

   要创建 LakeFormationWorkflowRole 角色，请参阅（可选）为工作流程创建 IAM 角色。

   > **Important**
   >
   > <your-s3-cloudtrail-bucket>替换为 CloudTrail 数据所在的 Amazon S3 位置。

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

2. 确认向该角色附加了三项策略。

第 3 步：为数据湖创建 Amazon S3 存储桶

创建将作为数据湖根位置的 Amazon S3 存储桶。

1. 通过 https://console.aws.amazon.com/s3/ 打开 Amazon S3 控制台，并以您在创建管理用户中创建的管理员用户身份登录。
2. 选择创建存储桶，然后通过向导创建名为 *<yourName>*-datalake-cloudtrail 的存储桶，其中 *<yourName>* 是您的名字和姓氏。例如：jdoe-datalake-cloudtrail。

   有关创建 Amazon S3 存储桶的详细说明，请参阅创建存储桶。

第 4 步：注册 Amazon S3 路径

将 Amazon S3 路径注册为数据湖的根位置。

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。以数据湖管理员身份登录。
2. 在导航窗格中的注册和提取下，选择数据湖位置。
3. 选择注册位置，然后选择浏览。
4. 选择您之前创建的 *<yourName>*-datalake-cloudtrail 存储桶，接受默认 IAM 角色 AWSServiceRoleForLakeFormationDataAccess，然后选择注册位置。

   有关注册位置的更多信息，请参阅向数据湖添加 Amazon S3 位置。

第 5 步：授予数据位置权限

主体必须拥有对数据湖位置的数据位置权限，才能创建指向该位置的数据目录表或数据库。您必须向 IAM 角色授予对工作流的数据位置权限，这样工作流才能向数据摄取目标写入数据。

1. 在导航窗格中的权限下，选择数据位置。
2. 选择授权，然后在授予权限对话框中进行以下选择：
   a. 对于 IAM 用户和角色，请选择 LakeFormationWorkflowRole。
   b. 对于存储位置，选择您的 *<yourName>*-datalake-cloudtrail 存储桶。
3. 选择授权。

有关数据位置权限的更多信息，请参阅Underlying data access control。

第 6 步：在数据目录中创建数据库

Lake Formation 数据目录中的元数据表存储在数据库中。

1. 在导航窗格中的数据目录下，选择数据库。
2. 选择创建数据库，然后在数据库详细信息下输入名称 lakeformation_cloudtrail。
3. 将其他字段留空，然后选择创建数据库。

第 7 步：授予数据权限

您必须授予在数据目录中创建元数据表的权限。由于工作流将与角色 LakeFormationWorkflowRole 一起运行，因此您必须向该角色授予这些权限。

1. 在 Lake Formation 控制台的导航窗格中的数据目录下，选择数据库。
2. 选择 lakeformation_cloudtrail 数据库，然后从操作下拉列表中，选择“权限”标题下的授权。
3. 在授予数据权限对话框中进行以下选择：

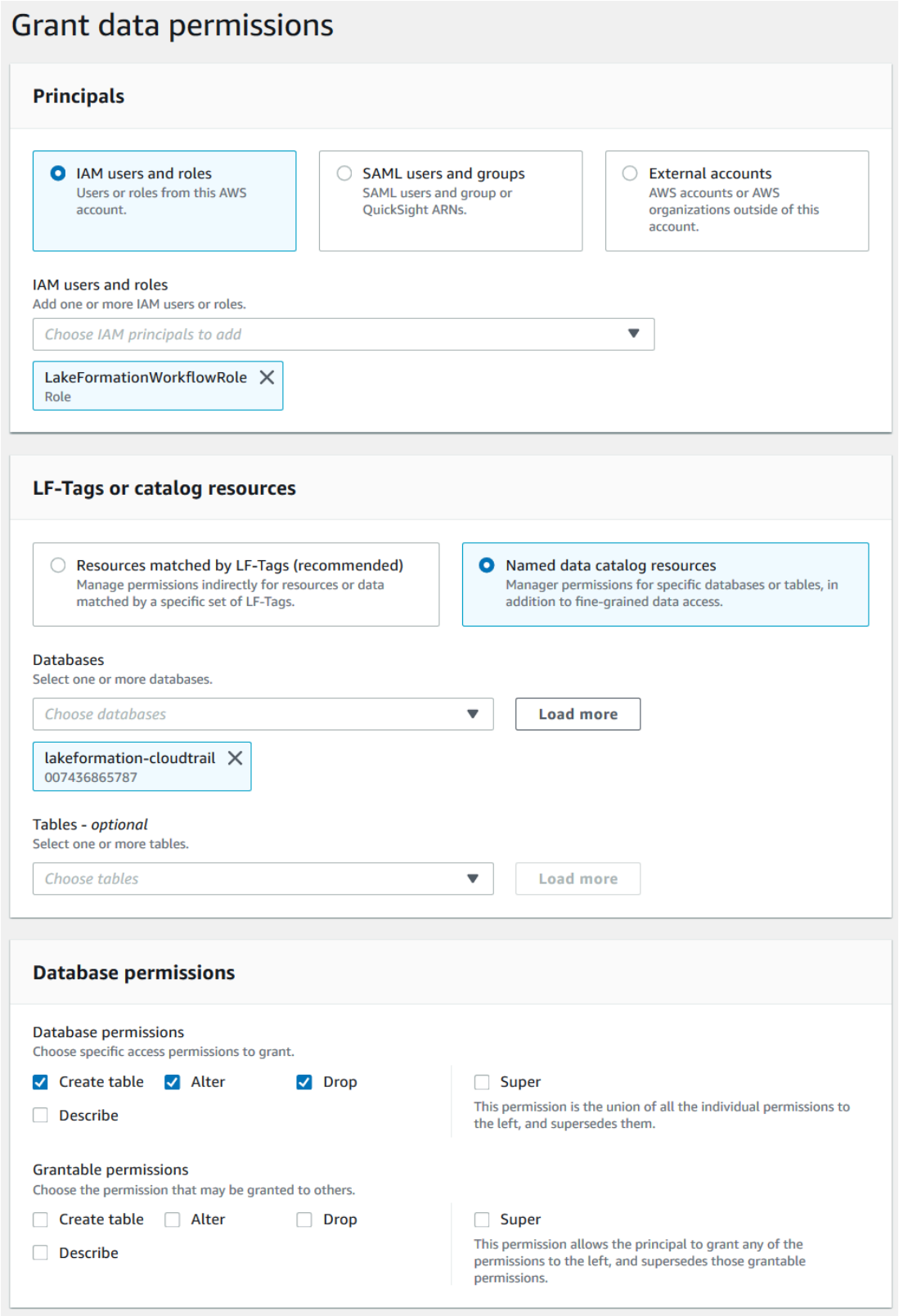
   a. 在主体下，对于 IAM 用户和角色，选择 LakeFormationWorkflowRole。

   b. 在 LF 标签或目录资源下，选择命名数据目录资源。

   c. 对于数据库，您应该看到已添加 lakeformation_cloudtrail 数据库。

   d. 在数据库权限下，选择创建表、更改和删除，如果已选中超级，请将其清除。

   现在，您的授予数据权限对话框看上去应类似于以下屏幕截图。


Grant data permissions
Principals
IAM users and roles
Users or roles from this AWS account.
SAML users and groups
SAML users and group or QuickSight ARNs.
External accounts
AWS accounts or AWS organizations outside of this account.
IAM users and roles
Add one or more IAM users or roles.
Choose IAM principals to add
LakeFormationWorkflowRole
Role
LF-Tags or catalog resources
Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.
Named data catalog resources
Manager permissions for specific databases or tables, in addition to fine-grained data access.
Databases
Select one or more databases.
Choose databases
Load more
lakeformation-cloudtrail
007436865787
Tables - optional
Select one or more tables.
Choose tables
Load more
Database permissions
Database permissions
Choose specific access permissions to grant.
Create table
Alter
Drop
Describe
Super
This permission is the union of all the individual permissions to the left, and supersedes them.
Grantable permissions
Choose the permission that may be granted to others.
Create table
Alter
Drop
Describe
Super
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

4. 选择授权。

有关授予 Lake Formation 权限的更多信息，请参阅管理 Lake Formation 权限。

第 8 步：使用蓝图创建工作流

为了读取 CloudTrail 日志、了解其结构、在数据目录中创建相应的表，我们需要设置一个由AWS Glue爬虫、作业、触发器和工作流组成的工作流程。Lake Formation 的蓝图简化了这一过程。

该工作流会生成作业、爬网程序和触发器，以用于发现数据并将其摄取到您的数据湖中。您可以基于其中一个预定义的 Lake Formation 蓝图创建工作流。

1. 在 Lake Formation 控制台的导航窗格中，选择蓝图，然后选择使用蓝图。
2. 在使用蓝图页面的蓝图类型下，选择AWS CloudTrail。
3. 在 “导入来源” 下，选择 CloudTrail 来源和开始日期。
4. 在导入目标下，指定以下参数：

目标数据库	`lakeformation_cloudtrail`
目标存储位置	`s3://<yourName> -datalake-cloudtrail`
Data format（数据格式）	Parquet

5. 对于导入频率，选择按需运行。
6. 在导入选项下，指定以下参数：

工作流名称	`lakeformationcloudtrailtest`
IAM 角色	`LakeFormationWorkflowRole`
表前缀	`cloudtrailtest` **Note** 必须小写。

7. 选择创建，然后等待控制台报告已成功创建工作流。

> **Tip**
>
> 您是否收到了以下错误消息？
> `User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...`
> 如果是，请检查您是否<account-id>将数据湖管理员用户的内联策略替换为有效的 AWS 账号。

第 9 步：运行工作流

由于您指定了工作流程 run-on-demand，因此必须手动启动该工作流程。

- 在蓝图页面上，选择工作流 `lakeformationcloudtrailtest`，然后在操作菜单中选择开始。

  当工作流运行时，您可以在上次运行状态列中查看其进度。时不时地选择刷新按钮。

  状态从正在运行依次变为正在发现、正在导入和已完成。

  工作流完成后：

  - 数据目录中将具有新的元数据表。
  - 您的 CloudTrail 日志将被提取到数据湖中。

  如果工作流失败，请执行以下操作：

  a. 选择该工作流，然后在操作菜单上选择查看图表。

     该工作流将在 AWS Glue 控制台中打开。

  b. 确保已选择工作流，然后选择历史记录选项卡。

  c. 在历史记录下，选择最近的运行，然后选择查看运行详细信息。

  d. 在动态（运行时）图表中选择失败的作业或爬网程序，然后查看错误消息。失败的节点为红色或黄色。

第 10 步：授予对表的选择权限

您必须授予对新数据目录表的 SELECT 权限，以便数据分析师可以查询这些表所指向的数据。

> **Note**
>
> 工作流会自动向运行它的用户授予对它创建的表的 SELECT 权限。由于数据湖管理员运行了该工作流，因此您必须向数据分析师授予 SELECT。

1. 在 Lake Formation 控制台的导航窗格中的数据目录下，选择数据库。
2. 选择 `lakeformation_cloudtrail` 数据库，然后从操作下拉列表中，选择“权限”标题下的授权。
3. 在授予数据权限对话框中进行以下选择：
   a. 在主体下，对于 IAM 用户和角色，选择 `datalake_user`。
   b. 在 LF 标签或目录资源下，选择命名数据目录资源。
   c. 对于数据库，应该已经选择了 `lakeformation_cloudtrail` 数据库。
   d. 对于表，选择 `cloudtrailtest-cloudtrail`。
   e. 在表和列权限 下，选择选择。
4. 选择授权。

以数据分析师的身份执行下一步。

第 11 步：使用 Amazon Athena查询数据湖

使用 Amazon Athena 控制台查询 CloudTrail 数据湖中的数据。

1. 通过 https://console.aws.amazon.com/athena/ 打开 Athena 控制台并以数据分析师用户 `datalake_user` 的身份登录。
2. 如有必要，请选择开始使用以继续使用 Athena 查询编辑器。
3. 对于 Data source (数据源)，选择 AwsDataCatalog。
4. 对于 Database (数据库)，请选择 `lakeformation_cloudtrail`。

   表列表即会填充。

5. 在表 `cloudtrailtest-cloudtrail` 旁边的溢出菜单（水平排列的 3 个点）上，选择预览表，然后选择运行。

   查询即会运行并显示 10 行数据。

   如果您以前没有使用过 Athena，则必须先在 Athena 控制台中配置 Amazon S3 位置以用于存储查询结果。`datalake_user` 必须拥有必需的权限，才能访问您选择的 Amazon S3 存储桶。

> **Note**
>
> 现在您已经完成了本教程的学习，请向组织中的主体授予数据权限和数据位置权限。

在 Lake Formation 中从 JDBC 来源创建数据湖

本教程将指导您完成在 AWS Lake Formation 控制台上使用 Lake Formation 从来源创建和加载您的第一个数据湖的步骤。

主题

目标受众

下表列出了本 AWS Lake Formation JDBC 教程中使用的角色。

角色	描述
IAM 管理员	可以创建 AWS Identity and Access Management (IAM) 用户和角色以及 Amazon Simple Storage Service (Amazon S3) 存储桶的用户。具有 `AdministratorAccess` AWS 托管策略。
数据湖管理员	可以访问数据目录、创建数据库以及向其他用户授予 Lake Formation 权限的用户。拥有的 IAM 权限比 IAM 管理员少，但足以管理数据湖。
数据分析人员	可以对数据湖运行查询的用户。拥有的权限仅足以运行查询。
工作流角色	具有运行工作流所需的 IAM 策略的角色。

有关完成教程的先决条件的信息，请参阅JDBC 教程的先决条件。

JDBC 教程的先决条件

在开始学习 AWS Lake Formation JDBC 教程之前，请确保您已完成以下操作：

- 完成Lake Formation 入门中所述的任务。
- 确定要在本教程中使用的可访问 JDBC 的数据存储。
- 收集创建 JDBC 类型的 AWS Glue 连接所需的信息。此数据目录对象包括数据存储的 URL 和登录凭证，如果数据存储是在 Amazon Virtual Private Cloud (Amazon VPC) 中创建的，则还包括其他特定于 VPC 的配置信息。有关更多信息，请参阅《AWS Glue 开发人员指南》中的在 AWS Glue 数据目录中定义连接。

本教程假定您熟悉 AWS Identity and Access Management (IAM)。有关 IAM 的信息，请参阅 IAM 用户指南。

要开始，请执行the section called “第 1 步：创建数据分析师用户”操作。

第 1 步：创建数据分析师用户

在这一步，您将在 AWS Lake Formation 中创建一个 AWS Identity and Access Management (IAM) 用户作为数据湖的数据分析师。

此用户拥有查询数据湖所需的一组最低权限。

1. 使用 https://console.aws.amazon.com/iam 打开 IAM 控制台。以您在创建管理用户中创建的管理员用户或使用 AdministratorAccess AWS 托管策略的用户身份登录。
2. 使用以下设置创建名为 datalake_user 的用户：

   - 启用 AWS Management Console 访问。
   - 设置密码，不需要重置密码。
   - 附加 AmazonAthenaFullAccess AWS 托管策略。
   - 附加下面的内联策略。将策略命名为 DatalakeUserBasic。

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

第 2 步：在 AWS Glue 中创建连接

> **Note**
>
> 如果您已经与 JDBC 数据来源建立了 AWS Glue 连接，请跳过这一步。

AWS Lake Formation 通过 AWS Glue 连接访问 JDBC 数据来源。连接是数据目录对象，其中包含连接到数据来源所需的所有信息。您可以使用 AWS Glue 控制台创建连接。

创建连接

1. 在 https://console.aws.amazon.com/glue/ 中打开 AWS Glue 控制台，并以您在创建管理用户中创建的管理员用户身份登录。
2. 在导航窗格的 Data catalog (数据目录) 下，选择 Connections (连接)。
3. 在 Connectors (连接器) 页面上，选择 Create custom connector (创建自定义连接器)。
4. 在连接器属性页面上，输入 **`datalake-tutorial`** 作为连接名称，然后选择 JDBC 作为连接类型。然后选择下一步。
5. 继续运行连接向导并保存连接。

   有关创建连接的信息，请参阅《AWS Glue 开发人员指南》中的 AWS Glue JDBC 连接属性。

第 3 步：为数据湖创建 Amazon S3 存储桶

在这一步，您要创建 Amazon Simple Storage Service (Amazon S3) 存储桶作为数据湖的根位置。

1. 通过 https://console.aws.amazon.com/s3/ 打开 Amazon S3 控制台，并以您在创建管理用户中创建的管理员用户身份登录。
2. 选择创建存储桶，然后通过向导创建名为 *`<yourName>`*`-datalake-tutorial` 的存储桶，其中 *`<yourName>`* 是您的名字和姓氏。例如：`jdoe-datalake-tutorial`。

   有关创建 Amazon S3 存储桶的详细说明，请参阅《Amazon Simple Storage Service 用户指南》中的如何创建 S3 存储桶？。

第 4 步：注册 Amazon S3 路径

在这一步，您要创建 Amazon Simple Storage Service (Amazon S3) 路径作为数据湖的根位置。

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。以数据湖管理员身份登录。
2. 在导航窗格中的注册和提取下，选择数据湖位置。
3. 选择注册位置，然后选择浏览。
4. 选择您之前创建的 *<yourName>*-datalake-tutorial 存储桶，接受默认 IAM 角色 AWSServiceRoleForLakeFormationDataAccess，然后选择注册位置。

   有关注册位置的更多信息，请参阅向数据湖添加 Amazon S3 位置。

第 5 步：授予数据位置权限

主体必须拥有对数据湖位置的数据位置权限，才能创建指向该位置的数据目录表或数据库。您必须向 IAM 角色授予对工作流的数据位置权限，这样工作流才能向数据摄取目标写入数据。

1. 在 Lake Formation 控制台的导航窗格中，在权限下选择数据位置。
2. 选择授权，然后在授予权限对话框中执行以下操作：
   a. 对于 IAM 用户和角色，请选择 LakeFormationWorkflowRole。
   b. 对于存储位置，选择您的 *<yourName>*-datalake-tutorial 存储桶。
3. 选择授权。

有关数据位置权限的更多信息，请参阅Underlying data access control。

第 6 步：在数据目录中创建数据库

Lake Formation 数据目录中的元数据表存储在数据库中。

1. 在 Lake Formation 控制台的导航窗格中，在数据目录下选择数据库。
2. 选择创建数据库，然后在数据库详细信息下输入名称 lakeformation_tutorial。
3. 将其他字段留空，然后选择创建数据库。

第 7 步：授予数据权限

您必须授予在数据目录中创建元数据表的权限。由于工作流与角色 LakeFormationWorkflowRole 一起运行，因此您必须向该角色授予这些权限。

1. 在 Lake Formation 控制台的导航窗格中，在权限下选择数据湖权限。
2. 选择授权，然后在授予数据权限对话框中执行以下操作：
   a. 在主体下，对于 IAM 用户和角色，选择 `LakeFormationWorkflowRole`。
   b. 在 LF 标签或目录资源下，选择命名数据目录资源。
   c. 对于数据库，请选择您之前创建的数据库 `lakeformation_tutorial`。
   d. 在数据库权限下，选择创建表、更改和删除，如果已选中超级，请将其清除。
3. 选择授权。

有关授予 Lake Formation 权限的更多信息，请参阅Lake Formation 权限概述 。

第 8 步：使用蓝图创建工作流

AWS Lake Formation 工作流会生成 AWS Glue 作业、爬网程序和触发器，以用于发现数据并将其摄取到您的数据湖中。您可以基于其中一个预定义的 Lake Formation 蓝图创建工作流。

1. 在 Lake Formation 控制台的导航窗格中，选择蓝图，然后选择使用蓝图。
2. 在使用蓝图页面的蓝图类型下，选择数据库快照。
3. 在导入源下的数据库连接中，选择您刚刚创建的连接 `datalake-tutorial`，或者为您的数据来源选择一个现有连接。
4. 对于源数据路径，以 *<database>*/*<schema>*/*<table>* 形式输入从中摄取数据的路径。

   您可以用百分比 (%) 字符替换架构或表。对于支持架构的数据库，请输入 *<database>*/*<schema>*/% 以匹配 *<database>* 中 *<schema>* 内的所有表。Oracle Database 和 MySQL 不支持路径中的架构，所以请改为输入 *<database>*/%。对于 Oracle 数据库，*<database>* 是系统标识符 (SID)。

   例如，如果 Oracle 数据库的 SID 为 `orcl`，则输入 `orcl/%` 以匹配在 JDCB 连接中指定的用户有权访问的所有表。

   > **Important**
   > 此字段区分大小写。

5. 在导入目标下，指定以下参数：

目标数据库	`lakeformation_tutorial`
目标存储位置	`s3://<yourName> -datalake-tutorial`
Data format（数据格式）	（选择 Parquet 或 CSV）

6. 对于导入频率，选择按需运行。
7. 在导入选项下，指定以下参数：

工作流名称	`lakeformationjdbctest`
IAM 角色	`LakeFormationWorkflowRole`
表前缀	`jdbctest` **Note** 必须小写。

8. 选择创建，然后等待控制台报告已成功创建工作流。

> **Tip**
> 您是否收到了以下错误消息？
> `User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...`
> 如果收到了，请检查您将是否将数据湖管理员用户的内联策略中的 *<account-id>* 替换为了有效的 AWS 帐号。

第 9 步：运行工作流

由于您指定工作流按需运行，因此必须在 AWS Lake Formation 中手动启动该工作流。

1. 在 Lake Formation 控制台的蓝图页面上，选择工作流 `lakeformationjdbctest`。
2. 选择操作，然后选择启动。
3. 当工作流运行时，您可以在上次运行状态列中查看其进度。时不时地选择刷新按钮。

   状态从正在运行依次变为正在发现、正在导入和已完成。

   工作流完成后：

   - 数据目录中将含有新的元数据表。
   - 您的数据已被摄取到数据湖中。

   如果工作流失败，请执行以下操作：

   a. 选择工作流。选择操作，然后选择查看图表。

      该工作流将在 AWS Glue 控制台中打开。
   b. 选择该工作流，然后选择历史记录选项卡。
   c. 选择最近的运行，然后选择查看运行详细信息。
   d. 在动态（运行时）图表中选择失败的作业或爬网程序，然后查看错误消息。失败的节点为红色或黄色。

第 10 步：授予对表的选择权限

您必须在 AWS Lake Formation 中授予对新数据目录表的 SELECT 权限，以便数据分析师可以查询这些表所指向的数据。

> **Note**
>
> 工作流会自动向运行它的用户授予对它创建的表的 SELECT 权限。由于数据湖管理员运行了该工作流，因此您必须向数据分析师授予 SELECT。

1. 在 Lake Formation 控制台的导航窗格中，在权限下选择数据湖权限。
2. 选择授权，然后在授予数据权限对话框中执行以下操作：

   a. 在主体下，对于 IAM 用户和角色，选择 `datalake_user`。
   b. 在 LF 标签或目录资源下，选择命名数据目录资源。

c. 对于数据库，选择 `lakeformation_tutorial`。

表列表即会填充。

d. 对于表，从数据来源中选择一个或多个表。

e. 在表和列权限 下，选择选择。

3. 选择授权。

以数据分析师的身份执行下一步。

第 11 步：使用 Amazon Athena 查询数据湖

使用 Amazon Athena 控制台查询数据湖中的数据。

1. 通过 https://console.aws.amazon.com/athena/ 打开 Athena 控制台并以数据分析师用户 `datalake_user` 的身份登录。
2. 如有必要，请选择开始使用以继续使用 Athena 查询编辑器。
3. 对于数据来源，选择 AwsDataCatalog。
4. 对于 Database (数据库)，请选择 `lakeformation_tutorial`。

表列表即会填充。

5. 在其中一个表旁边的弹出菜单中，选择预览表。

查询即会运行并显示 10 行数据。

第 12 步：使用 Amazon Redshift Spectrum 查询数据湖中的数据

您可以设置 Amazon Redshift Spectrum 以查询您导入到 Amazon Simple Storage Service (Amazon S3) 数据湖中的数据。首先，创建一个 AWS Identity and Access Management (IAM) 角色，该角色用于启动 Amazon Redshift 集群和查询 Amazon S3 数据。然后，向该角色授予对您要查询的表的 `Select` 权限。然后，授予该用户使用 Amazon Redshift 查询编辑器的权限。最后，创建一个 Amazon Redshift 集群并运行查询。

您要以管理员身份创建集群，并以数据分析师的身份查询集群。

有关 Amazon Redshift Spectrum 的更多信息，请参阅《Amazon Redshift 数据库开发人员指南》中的使用 Amazon Redshift Spectrum 查询外部数据。

设置运行 Amazon Redshift 查询的权限

1. 通过以下网址打开 IAM 控制台：https://console.aws.amazon.com/iam/。以您在创建管理用户中创建的管理员用户（用户名 Administrator）或使用 AdministratorAccess AWS 托管策略的用户身份登录。
2. 在导航窗格中，选择 Policies (策略)。

   如果这是您首次选择 Policies，则会显示 Welcome to Managed Policies 页面。选择开始使用。
3. 选择 Create policy (创建策略)。
4. 请选择 JSON 选项卡。
5. 粘贴以下 JSON 策略文档。

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

6. 完成后，选择审核对策略进行审核。策略验证程序将报告任何语法错误。
7. 在查看策略页面上，输入 **RedshiftLakeFormationPolicy** 作为您要创建的策略的名称。输入描述（可选）。查看策略摘要以查看您的策略授予的权限。然后，选择创建策略以保存您的工作。

8. 在 IAM 控制台的导航窗格中，选择 Roles，然后选择 Create role。
9. 对于选择可信实体，选择 AWS 服务。
10. 选择 Amazon Redshift 服务来代入此角色。
11. 为您的服务选择 Redshift Customizable (Redshift 可自定义)。然后选择 Next: Permissions (下一步：权限)。
12. 搜索刚创建的策略 `RedshiftLakeFormationPolicy`，然后在列表中选中该策略名称旁边的复选框。
13. 请选择下一步：标签。
14. 请选择下一步：审核。
15. 对于角色名称，输入名称 **`RedshiftLakeFormationRole`**。
16. （可选）对于 Role description (角色描述)，输入新角色的描述。
17. 检查该角色，然后选择创建角色。

授予对 Lake Formation 数据库中要查询的表的 **`Select`** 权限

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。以数据湖管理员身份登录。
2. 在导航窗格的权限下，选择数据湖权限，然后选择授权。
3. 提供以下信息：
   - 对于 IAM 用户和角色，选择您创建的 IAM 角色 `RedshiftLakeFormationRole`。运行 Amazon Redshift 查询编辑器时，它使用此 IAM 角色来获取数据权限。
   - 对于 Database (数据库)，请选择 `lakeformation_tutorial`。

     表列表即会填充。
   - 对于表，选择数据来源中要查询的表。
   - 选择选择表权限。
4. 选择授权。

设置 Amazon Redshift Spectrum 并运行查询

1. 通过以下网址打开 Amazon Redshift 控制台：https://console.aws.amazon.com/redshift。以用户 `Administrator` 的身份登录。

2. 选择创建集群。
3. 在创建集群页面上，输入 `redshift-lakeformation-demo` 作为集群标识符。
4. 对于节点类型，选择 dc2.large。
5. 向下滚动，在数据库配置下输入或接受以下参数：
   - 管理用户名称：`awsuser`
   - 管理用户密码：(*`Choose a password`*)
6. 展开集群权限，对于可用的 IAM 角色，选择 RedshiftLakeFormationRole。然后选择 Add IAM role (添加 IAM 角色)。
7. 如果您必须使用的端口与默认值 5439 不同，请关闭其他配置旁边的使用默认值选项。展开数据库配置部分，然后输入新的数据库端口号。
8. 选择创建集群。

   集群页面即会加载。
9. 等待到集群状态变为 可用。定期选择刷新图标。
10. 向数据分析师授予对集群运行查询的权限。为此，请完成以下步骤。
    a. 通过网址 https://console.aws.amazon.com/iam/ 打开 IAM 控制台，然后以 `Administrator` 用户身份登录。
    b. 在导航窗格中，选择用户，然后选择将以下托管策略附加到用户 `datalake_user`。
       - `AmazonRedshiftQueryEditor`
       - `AmazonRedshiftReadOnlyAccess`
11. 退出 Amazon Redshift 控制台，然后以用户 `datalake_user` 身份重新登录。
12. 在左侧垂直工具栏中，选择编辑器图标以打开查询编辑器并连接到集群。如果连接到数据库对话框出现，请选择集群名称 `redshift-lakeformation-demo`，然后输入数据库名称 **`dev`**、用户名 **`awsuser`** 和您创建的密码。然后，选择 Connect to database (连接到数据库)。

    > **Note**
    >
    > 如果系统没有提示您输入连接参数，并且已经在查询编辑器中选择了另一个集群，请选择更改连接打开连接到数据库对话框。

13. 在新查询 1 文本框中，输入并运行以下语句以将 Lake Formation 中的数据库 `lakeformation_tutorial` 映射到 Amazon Redshift 架构名称 `redshift_jdbc`：

> **Important**
>
> 将 *<account-id>* 替换为有效的 AWS 账户编号，并将 *<region>* 替换为有效的 AWS 区域名称（例如，us-east-1）。

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

14. 在选择架构下的架构列表中，选择 redshift_jdbc。

    表列表即会填充。查询编辑器仅显示您被授予了对其的 Lake Formation 数据湖权限的表。

15. 在表名称旁边的弹出菜单中，选择预览数据。

    Amazon Redshift 将返回前 10 行。

    现在，您可对您对其拥有权限的表和列运行查询。

第 13 步：使用 Amazon Redshift Spectrum 授予或撤销 Lake Formation 权限

Amazon Redshift 支持使用修改后的 SQL 语句授予和撤销对数据库和表的 Lake Formation 权限。这些语句与现有的 Amazon Redshift 语句相似。有关更多信息，请参阅《Amazon Redshift 数据库开发人员指南》中的授权和撤销。

在 Lake Formation 中为开放表存储格式设置权限

AWS Lake Formation 支持管理对诸如 Apache Iceberg、Apache Hudi 和 Linux Foundation Delta Lake 等开放表格格式 (OTF) 的访问权限。在本教程中，您将学习如何通过 AWS Glue 使用符号链接清单表在 AWS Glue Data Catalog 中创建 Iceberg、Hudi 和 Delta Lake，如何使用 Lake Formation 设置精细权限，以及如何使用 Amazon Athena 查询数据。

> **Note**
>
> AWS 分析服务并不支持所有事务表式。有关更多信息，请参阅使用其他 AWS 服务。本教程手动介绍如何仅使用AWS Glue作业在数据目录中创建新数据库和表。

本教程包括一个用于快速设置的AWS CloudFormation模板。您可以查看和自定义该模板来满足自己的需求。

主题

目标受众

本教程适用于 IAM 管理员、数据湖管理员和业务分析师。下表列出了本教程中使用 Lake Formation 创建受管控表时所使用的角色。

角色	描述
IAM 管理员	可以创建 IAM 用户和角色以及 Amazon S3 存储桶的用户。具有 `AdministratorAccess` AWS 托管策略。
数据湖管理员	可以访问数据目录、创建数据库以及向其他用户授予 Lake Formation 权限的用户。拥有的 IAM 权限比 IAM 管理员少，但足以管理数据湖。
业务分析师	可以对数据湖运行查询的用户。拥有运行查询的权限。

先决条件

在开始本教程之前，您必须拥有一个AWS 账户可以以具有正确权限的用户身份登录的。有关更多信息，请参阅 注册 AWS 账户 和 创建管理用户。

本教程假定您熟悉 IAM 角色和策略。有关 IAM 的信息，请参阅 IAM 用户指南。

要完成本教程，您需要设置以下 AWS 资源。

- 数据湖管理员用户
- Lake Formation 数据湖设置
- Amazon Athena 引擎版本 3

创建数据湖管理员

1. 以管理员用户身份登录 Lake For mation 控制台 https://console.aws.amazon.com/lakeformation/。您将在美国东部（弗吉尼亚北部）地区为本教程创建资源。
2. 在 Lake Formation 控制台上的导航窗格中，在权限下选择管理角色和任务。
3. 在数据湖管理员下选择数据湖管理员。
4. 在弹出窗口管理数据湖管理员中，在 IAM 用户和角色下，选择 IAM 管理用户。
5. 选择保存。

启用数据湖设置

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。在导航窗格中的数据目录下，选择设置。取消选中以下选项：
   - 仅对新数据库使用 IAM 访问控制。
   - 仅对新数据库中的新表使用 IAM 访问控制。
2. 在跨账户版本设置下，选择版本 3 作为跨账户版本。
3. 选择保存。

将 Amazon Athena 引擎升级到版本 3

1. 通过以下网址打开 Athena 控制台：https://console.aws.amazon.com/athena/。
2. 选择工作组，然后选择主工作组。

3. 确保工作组的最低版本为 3。如果不是，请编辑工作组，为升级查询引擎选择手动，然后选择版本 3。
4. 选择 保存更改。

第 1 步：调配资源

本节介绍如何使用 AWS CloudFormation 模板设置 AWS 资源。

使用 AWS CloudFormation 模板创建资源

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/cloudformation 以 IAM 管理员身份登录 AWS CloudFormation 控制台。
2. 选择启动堆栈。
3. 在创建堆栈屏幕上，选择下一步。
4. 输入堆栈名称。
5. 选择下一步。
6. 在下一页上，选择下一步。
7. 查看最后页面上的详细信息，然后选择我确认 AWS CloudFormation 可以创建 IAM 资源。
8. 选择创建。

   堆栈创建过程可能需要几分钟时间才能完成。

启动 Cloud Formation 堆栈可创建以下资源：

- lf-otf-datalake-123456789012 — 用于存储数据的亚马逊 S3 存储桶

> **Note**
>
> Amazon S3 存储桶名称中附加的账户 ID 将由您的账户 ID 替换。

- lf-otf-tutorial-123456789012 — 用于存储查询结果和任务脚本的 Amazon S3 存储桶 AWS Glue
- lficebergdb – AWS Glue Iceberg 数据库
- lfhudidb – AWS Glue Hudi 数据库
- lfdeltadb – AWS Glue Delta 数据库
- native-iceberg-create — 在数据目录中创建 Iceberg 表的作AWS Glue业

- native-hudi-create — 在数据目录中创建 Hudi 表的作AWS Glue业
- native-delta-create — 在数据目录中创建增量表的作AWS Glue业
- LF-OTF-GlueServiceRole — 你传递给它AWS Glue来运行任务的 IAM 角色。此角色附加了访问数据目录、Amazon S3 存储桶等资源所需的策略。
- LF-OTF-RegisterRole — IAM 角色向 Lake Formation 注册亚马逊 S3 地点。此角色已附加 `LF-Data-Lake-Storage-Policy`。
- lf-consumer-analystuser — IAM 用户使用 Athena 查询数据
- lf-consumer-analystuser-credentials — 存储在中的数据分析师用户的密码 AWS Secrets Manager

创建完堆栈后，导航至输出选项卡并记下以下各项的值：

- AthenaQueryResultLocation — Athena 查询输出的亚马逊 S3 位置
- BusinessAnalystUserCredentials — 数据分析师用户的密码

  检索密码值：

  1. 通过导航到 Secrets Manager 控制台来选择 `lf-consumer-analystuser-credentials` 值。
  2. 在密钥值部分中，选择检索密钥值。
  3. 记下密码的值。

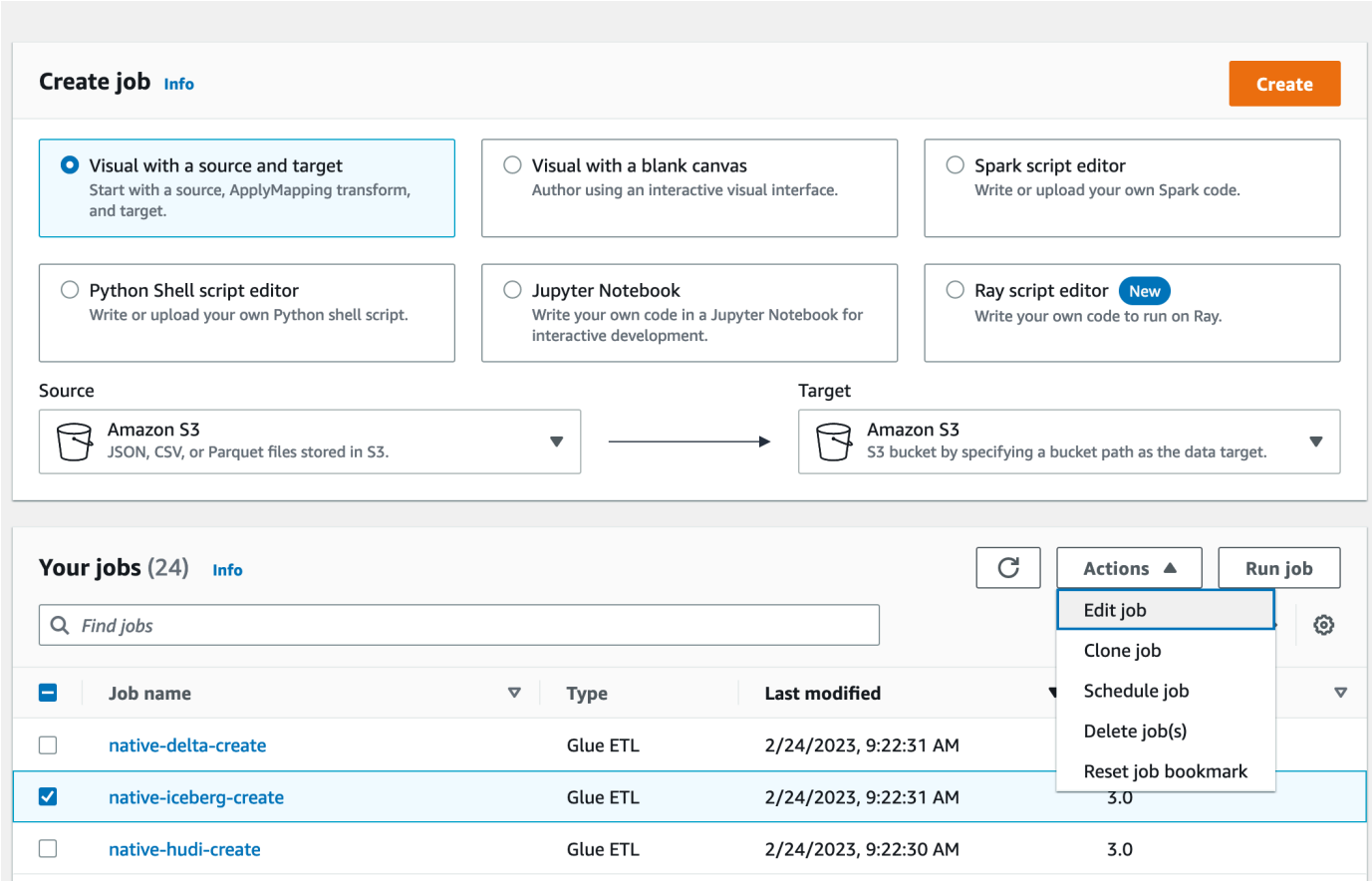
第 2 步：为 Iceberg 表设置权限

在本节中，您将学习如何在 AWS Glue Data Catalog 中创建 Iceberg 表、如何在 AWS Lake Formation 中设置数据权限以及如何使用 Amazon Athena 查询数据。

创建 Iceberg 表

在这一步，您将运行一个在数据目录中创建 Iceberg 事务表的 AWS Glue 作业。

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/glue/ 以数据湖管理员用户身份打开 AWS Glue 控制台。
2. 从左侧导航窗格中选择作业。
3. 选择 `native-iceberg-create`。



4. 在操作下，选择编辑作业。
5. 在作业详细信息下，展开高级属性，并选中使用 AWS Glue Data Catalog 作为 Hive 元存储旁边的框以在 AWS Glue Data Catalog 中添加表元数据。这将指定 AWS Glue Data Catalog 作为作业中使用的数据目录资源的元存储，并可在稍后对目录资源应用 Lake Formation 权限。
6. 选择保存。
7. 选择运行。您可以查看运行中作业的状态。

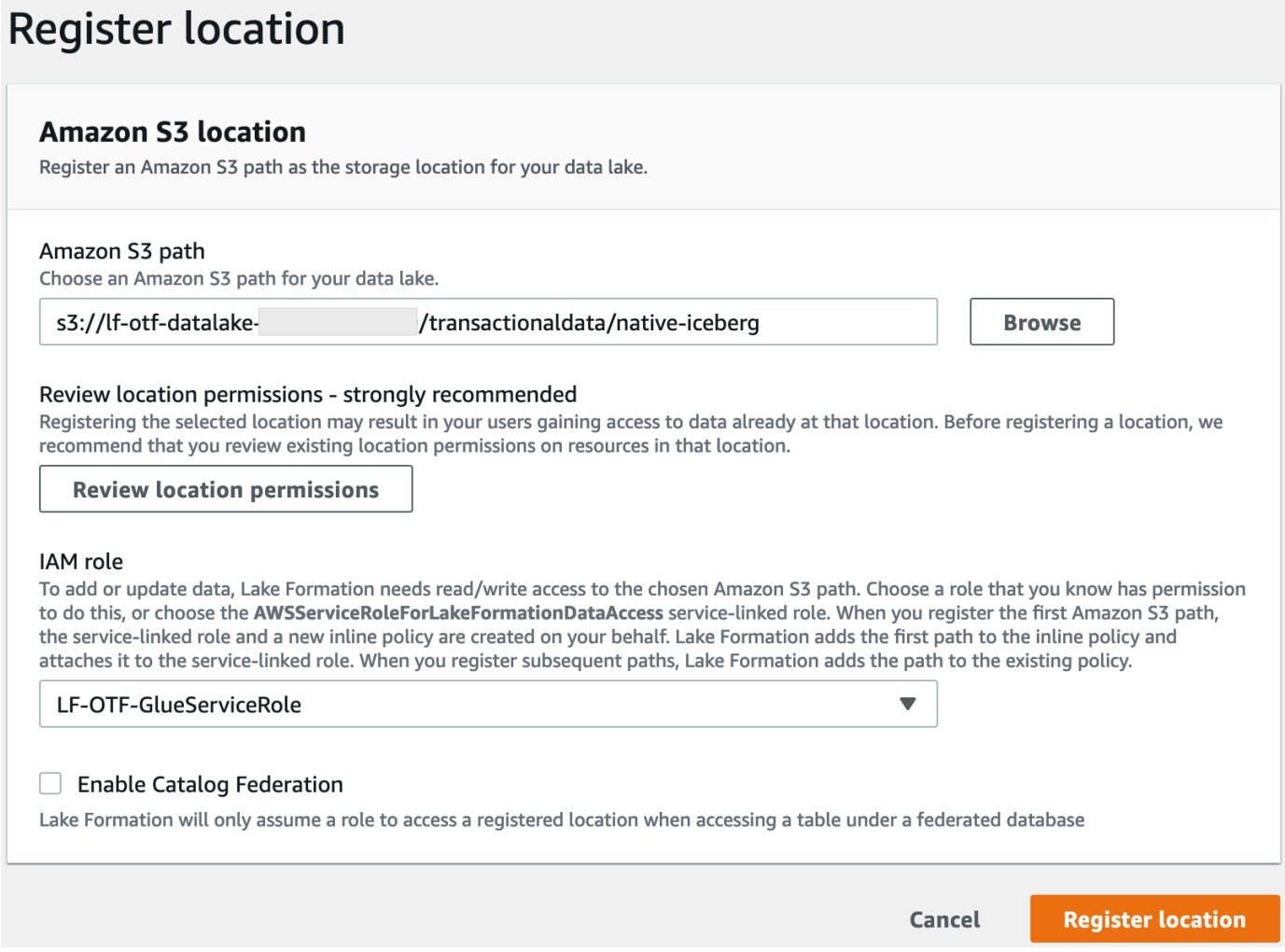
   有关 AWS Glue 作业的更多信息，请参阅《AWS Glue 开发人员指南》中的在 AWS Glue 控制台上处理作业。

   此作业将在 `lficebergdb` 数据库中创建名为 `product` 的 Iceberg 表。在 Lake Formation 控制台中验证 Product 表。

在 Lake Formation 中注册数据位置

接下来，将 Amazon S3 路径注册为数据湖的位置。

1. 通过 https://console.aws.amazon.com/lakeformation/ 以数据湖管理员用户身份打开 Lake Formation 控制台。
2. 在导航窗格中的注册和提取下，选择数据位置。
3. 在控制台右上角，选择注册位置。
4. 在注册位置页面上，输入以下内容：
   - Amazon S3 路径 – 选择浏览，然后选择 `lf-otf-datalake-123456789012`。单击 Amazon S3 根位置旁边的右箭头 (>) 以导航到 `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg` 位置。
   - IAM 角色 — 选择 `LF-OTF-RegisterRole` 作为 IAM 角色。
   - 选择注册位置。

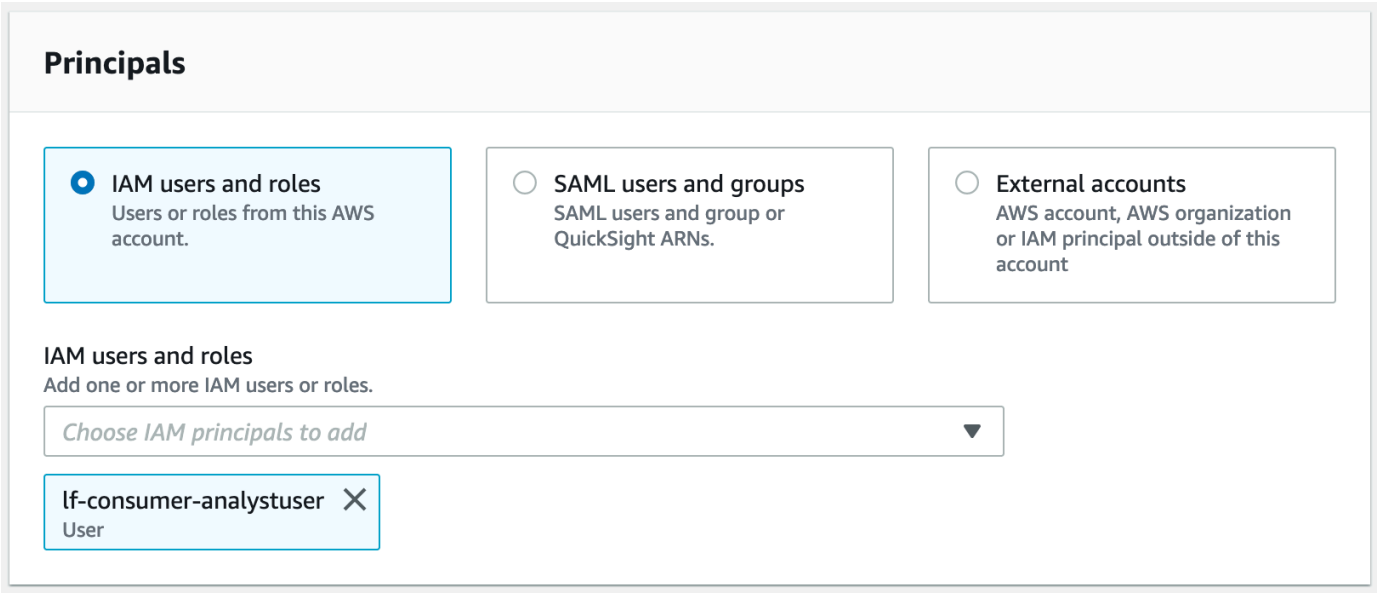


有关在 Lake Formation 中注册数据位置的更多信息，请参阅向数据湖添加 Amazon S3 位置。

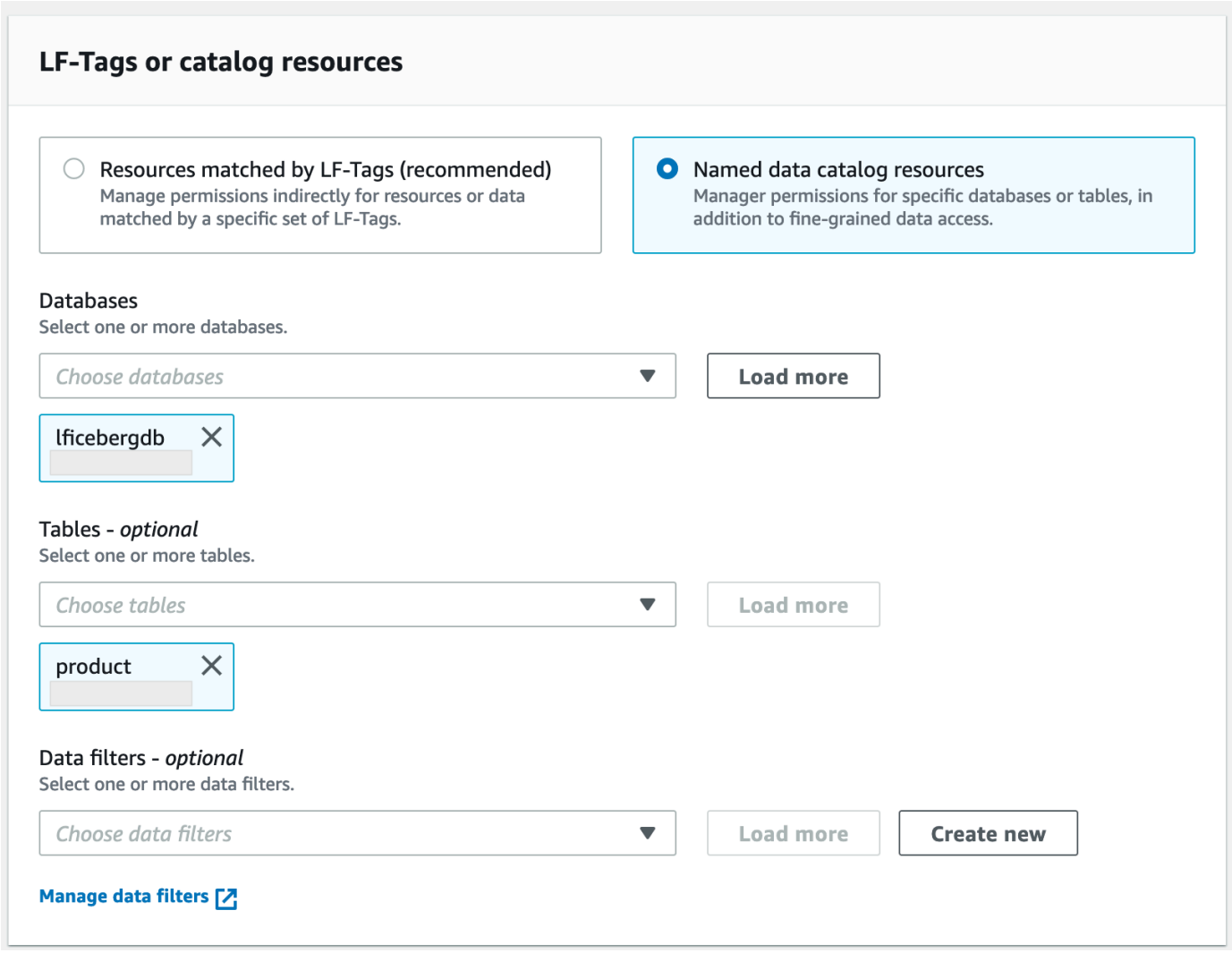
授予对 Iceberg 表的 Lake Formation 权限

在这一步，我们将向业务分析师用户授予数据湖权限。

1. 在数据湖权限下，选择授权。
2. 在授予数据权限屏幕上，选择 IAM 用户和角色。
3. 从下拉列表中选择 `lf-consumer-analystuser`。



4. 选择命名数据目录资源。
5. 对于数据库，选择 `lficebergdb`。
6. 对于表，选择 `product`。



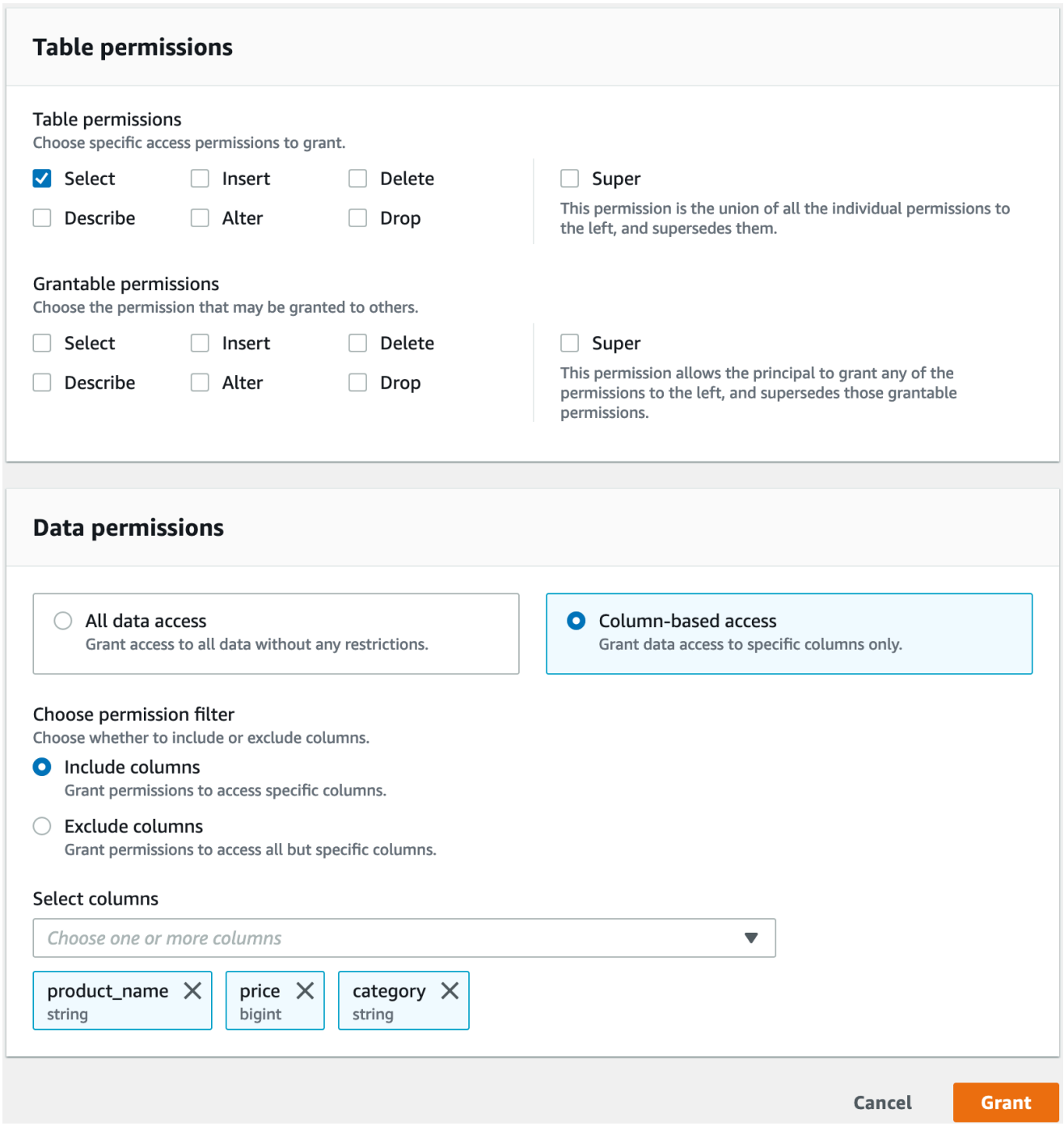
7. 接下来，您可以通过指定列来授予基于列的访问权限。

   a. 在表权限下，选择选择。

   b. 在数据权限下，选择基于列的访问权限，然后选择包括列。

   c. 依次选择列 `product_name`、`price` 和 `category`。

   d. 选择授权。



使用 Athena 查询 Iceberg 表

现在，您可以开始使用 Athena 查询您创建的 Iceberg 表。如果您是首次在 Athena 中运行查询，则需要配置查询结果位置。有关更多信息，请参阅指定查询结果位置。

1. 以数据湖管理员用户身份登录，然后使用AWS CloudFormation输出中前面注明的密码以美国东部（弗吉尼亚北部）区域的身份登录。`lf-consumer-analystuser`

2. 从 https://console.aws.amazon.com/athena/ 打开 Athena 控制台。
3. 选择设置，然后选择管理。
4. 在查询结果位置框中，输入 AWS CloudFormation 输出中您所创建的存储桶路径。复制 **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial-123456789012/athena-results/) 的值并选择“保存”。
5. 运行以下查询以预览存储在 Iceberg 表中的 10 条记录：

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

有关使用 Athena 查询 Iceberg 表的更多信息，请参阅《Amazon Athena 用户指南》中的查询 Iceberg 表。

第 3 步：为 Hudi 表设置权限

在本节中，您将学习如何在 AWS Glue Data Catalog 中创建 Hudi 表、如何在 AWS Lake Formation 中设置数据权限以及如何使用 Amazon Athena 查询数据。

创建 Hudi 表

在这一步，您将运行一个在数据目录中创建 Iceberg 事务表的 AWS Glue 作业。

1. 在美国东部（弗吉尼亚北部）地区通过 https://console.aws.amazon.com/glue/ 登录主AWS Glue机

   。
2. 从左侧导航窗格中选择作业。
3. 选择 `native-hudi-create`。
4. 在操作下，选择编辑作业。
5. 在作业详细信息下，展开高级属性，并选中使用 AWS Glue Data Catalog 作为 Hive 元存储旁边的框以在 AWS Glue Data Catalog 中添加表元数据。这将指定 AWS Glue Data Catalog 作为作业中使用的数据目录资源的元存储，并可在稍后对目录资源应用 Lake Formation 权限。
6. 选择保存。
7. 选择运行。您可以查看运行中作业的状态。

   有关 AWS Glue 作业的更多信息，请参阅《AWS Glue 开发人员指南》中的在 AWS Glue 控制台上处理作业。

此作业在 database:lfhudidb 中创建 Hudi(cow) 表。在 Lake Formation 控制台中验证 `product` 表。

在 Lake Formation 中注册数据位置

接下来，将 Amazon S3 路径注册为数据湖的根位置。

1. 以数据湖管理员用户身份登录 Lak e Formation 控制台 https://console.aws.amazon.com/lakeformation/。
2. 在导航窗格中的注册和提取下，选择数据位置。
3. 在控制台右上角，选择注册位置。
4. 在注册位置页面上，输入以下内容：
   - Amazon S3 路径 – 选择浏览，然后选择 `lf-otf-datalake-123456789012`。单击 Amazon S3 根位置旁边的右箭头 (>) 以导航到 `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi` 位置。
   - IAM 角色 — 选择 `LF-OTF-RegisterRole` 作为 IAM 角色。
   - 选择注册位置。

授予对 Hudi 表的数据湖权限

在这一步，我们将向业务分析师用户授予数据湖权限。

1. 在数据湖权限下，选择授权。
2. 在授予数据权限屏幕上，选择 IAM 用户和角色。
3. 从下拉列表中选择 `lf-consumer-analystuser`。
4. 选择命名数据目录资源。
5. 对于数据库，选择 `lfhudidb`。
6. 对于表，选择 `product`。
7. 接下来，您可以通过指定列来授予基于列的访问权限。
   a. 在表权限下，选择选择。
   b. 在数据权限下，选择基于列的访问权限，然后选择包括列。
   c. 依次选择列 `product_name`、`price` 和 `category`。
   d. 选择授权。

使用 Athena 查询 Hudi 表

现在开始使用 Athena 查询您创建的 Hudi 表。如果您是首次在 Athena 中运行查询，则需要配置查询结果位置。有关更多信息，请参阅指定查询结果位置。

1. 以数据湖管理员用户身份登录，然后使用AWS CloudFormation输出中前面注明的密码以美国东部（弗吉尼亚北部）区域的身份登录。`lf-consumer-analystuser`
2. 从 https://console.aws.amazon.com/athena/ 打开 Athena 控制台。
3. 选择设置，然后选择管理。
4. 在查询结果位置框中，输入 AWS CloudFormation 输出中您所创建的存储桶路径。复制 **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial-123456789012/athena-results/) 的值并保存。
5. 运行以下查询以预览存储在 Hudi 表中的 10 条记录：

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

   有关查询 Hudi 表的更多信息，请参阅《Amazon Athena 用户指南》中的查询 Hudi 表一节。

第 4 步：为 Delta Lake 表设置权限

在本节中，您将学习如何在 AWS Glue Data Catalog 中使用符号链接清单文件创建 Delta Lake 表、如何在 AWS Lake Formation 中设置数据权限以及如何使用 Amazon Athena 查询数据。

创建 Delta Lake 表

在这一步，您将运行一个在数据目录中创建 Delta Lake 事务表的 AWS Glue 作业。

1. 在美国东部（弗吉尼亚北部）地区通过 https://console.aws.amazon.com/glue/ 登录主AWS Glue机

   。
2. 从左侧导航窗格中选择作业。
3. 选择 `native-delta-create`。
4. 在操作下，选择编辑作业。
5. 在作业详细信息下，展开高级属性，并选中使用 AWS Glue Data Catalog 作为 Hive 元存储旁边的框以在 AWS Glue Data Catalog 中添加表元数据。这将指定 AWS Glue Data Catalog 作为作业中使用的数据目录资源的元存储，并可在稍后对目录资源应用 Lake Formation 权限。

6. 选择保存。
7. 在操作下选择运行。

   此作业将在 lfdeltadb 数据库中创建名为 product 的 Delta Lake 表。在 Lake Formation 控制台中验证 product 表。

在 Lake Formation 中注册数据位置

接下来，将 Amazon S3 路径注册为数据湖的根位置。

1. 通过 https://console.aws.amazon.com/lakeformation/ 以数据湖管理员用户身份打开 Lake Formation 控制台。
2. 在导航窗格中的注册和提取下，选择数据位置。
3. 在控制台右上角，选择注册位置。
4. 在注册位置页面上，输入以下内容：

   - Amazon S3 路径 – 选择浏览，然后选择 lf-otf-datalake-123456789012。单击 Amazon S3 根位置旁边的右箭头 (>) 以导航到 s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta 位置。
   - IAM 角色 — 选择 LF-OTF-RegisterRole 作为 IAM 角色。
   - 选择注册位置。

授予对 Delta Lake 表的数据湖权限

在这一步，我们将向业务分析师用户授予数据湖权限。

1. 在数据湖权限下，选择授权。
2. 在授予数据权限屏幕上，选择 IAM 用户和角色。
3. 从下拉列表中选择 lf-consumer-analystuser。
4. 选择命名数据目录资源。
5. 对于数据库，选择 lfdeltadb。
6. 对于表，选择 product。
7. 接下来，您可以通过指定列来授予基于列的访问权限。

   a. 在表权限下，选择选择。

   b. 在数据权限下，选择基于列的访问权限，然后选择包括列。

c. 依次选择列 product_name、price 和 category。

d. 选择授权。

使用 Athena 查询 Delta Lake 表

现在开始使用 Athena 查询您创建的 Delta Lake 表。如果您是首次在 Athena 中运行查询，则需要配置查询结果位置。有关更多信息，请参阅指定查询结果位置。

1. 以数据湖管理员用户身份注销，然后使用之前记下的 AWS CloudFormation 输出中的密码以 BusinessAnalystUser 身份在美国东部（弗吉尼亚州北部）区域登录。
2. 从 https://console.aws.amazon.com/athena/ 打开 Athena 控制台。
3. 选择设置，然后选择管理。
4. 在查询结果位置框中，输入 AWS CloudFormation 输出中您所创建的存储桶路径。复制 **AthenaQueryResultLocation** (s3://lf-otf-tutorial-123456789012/athena-results/) 的值并保存。
5. 运行以下查询来预览存储在 Delta Lake 表中的 10 条记录：

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

有关查询 Delta Lake 表的更多信息，请参阅《Amazon Athena 用户指南》中的查询 Delta Lake 表一节。

第 5 步：清除 AWS 资源

清理资源

为防止向你收取不必要的费用AWS 账户，请删除你在本教程中使用的AWS资源。

1. 以 IAM 管理员的身份登录AWS CloudFormation控制台，网址为 https://console.aws.amazon.com/cloudformation。
2. 删除 Cloud Formation 堆栈。您创建的表将自动与堆栈一起被删除。

使用 Lake Formation 基于标签的访问控制管理数据湖

成千上万的客户在 AWS 上构建 PB 级数据湖。这些客户中有许多使用 AWS Lake Formation 在整个组织中轻松构建和共享他们的数据湖。随着表和用户数量的增加，数据管家和管理员正在寻找方法来轻

松地大规模管理对数据湖的权限。Lake Formation 基于标签的访问控制 (LF-TBAC) 允许数据管家创建 LF 标签（基于其数据分类和本体）并在之后将其附加到资源，因而解决了这个问题。

LF-TBAC 是一种基于属性定义权限的授权策略。在 Lake Formation 中，这些属性被称为“LF 标签”。您可以将 LF 标签附加到数据目录资源和 Lake Formation 主体。数据湖管理员可以使用 LF 标签分配和撤销对 Lake Formation 资源的权限。有关更多信息，请参阅Lake Formation 基于标签的访问控制。

本教程演示如何使用 AWS 公共数据集创建 Lake Formation 基于标签的访问控制策略。此外，它还展示了如何查询具有相关的 Lake Formation 基于标签的访问策略的表、数据库和列。

您可以将 LF-TBAC 用于以下使用案例：

- 您有大量表和主体，数据湖管理员必须授予对它们的访问权限
- 您想基于本体对数据进行分类并基于分类授予权限
- 数据湖管理员想要以松耦合的方式动态分配权限

以下是使用 LF-TBAC 配置权限的主要步骤：

1. 数据管家使用以下两个 LF 标签定义标签本体：`Confidential` 和 `Sensitive`。带有 `Confidential=True` 的数据具有更严格的访问控制。带有 `Sensitive=True` 的数据需要分析师进行具体分析。
2. 数据管家为数据工程师分配不同级别的权限，以便他们使用不同 LF 标签构建表。
3. 数据工程师构建了两个数据库：`tag_database` 和 `col_tag_database`。`tag_database` 中的所有表都配有 `Confidential=True`。`col_tag_database` 中的所有表都配有 `Confidential=False`。`col_tag_database` 中表的一些列带有 `Sensitive=True` 标签，以满足特定的分析需求。
4. 数据工程师使用特定表达式条件 `Confidential=True` 以及 `Confidential=False`、`Sensitive=True` 向分析人员授予对表的读取权限。
5. 通过这种配置，数据分析师可以专注于使用正确的数据执行分析。

主题

- 第 3 步：创建 Lake Formation 数据库
- 第 4 步：授予表权限
- 第 5 步：在 Amazon Athena 中运行查询以验证权限
- 第 6 步：清除 AWS 资源

目标受众

本教程适用于数据管家、数据工程师和数据分析师。谈及在 Lake Formation 中管理 AWS Glue Data Catalog 和管理权限，生产账户中的数据管家会根据其支持的功能拥有职能所有权，并且可以向各种使用者、外部组织和账户授予访问权限。

下表列出了本教程中使用的角色：

角色	描述
数据管家（管理员）	`lf-data-steward` 用户拥有以下访问权限： • 对数据目录中所有资源的读取权限 • 可以创建 LF 标签并将其关联到数据工程师角色，以便向其他主体授予可授予权限
数据工程师	`lf-data-engineer` 用户具有以下访问权限： • 对数据目录中所有资源的完整读取、写入和更新权限 • 数据湖中的数据位置权限 • 可以关联 LF 标签并将其关联到数据目录 • 可以将 LF 标签附加到资源，从而根据数据管家创建的所有策略为主体提供访问权限
数据分析人员	`lf-data-analyst` 用户拥有以下访问权限： • 对通过 Lake Formation 基于标签的访问策略共享的资源的精细访问权限

先决条件

在开始学习本教程之前，您必须拥有 AWS 账户，以便能够以具有正确权限的管理用户身份登录。有关更多信息，请参阅完成初始 AWS 配置任务。

本教程假定您熟悉 IAM。有关 IAM 的信息，请参阅 IAM 用户指南。

第 1 步：调配资源

本教程中提供了一个用于进行快速设置的 AWS CloudFormation 模板。您可以查看和自定义该模板来满足自己的需求。该模板创建了三个不同的角色（列在中目标受众）来执行本练习，并将 nyc-taxi-data 数据集复制到您的本地 Amazon S3 存储桶。

- 一个 Amazon S3 存储桶
- 适当的 Lake Formation 设置
- 适当的 Amazon EC2 资源
- 三个具有凭证的 IAM 角色

创建您的资源

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/cloudformation 登录 AWS CloudFormation 控制台。
2. 选择启动堆栈。
3. 选择下一步。
4. 在用户配置部分，为以下三个角色输入密码：`DataStewardUserPassword`、`DataEngineerUserPassword` 和 `DataAnalystUserPassword`。
5. 查看最后页面上的详细信息，然后选择我确认 AWS CloudFormation 可以创建 IAM 资源。
6. 选择创建。

   堆栈创建可能最多需要五分钟时间才能完成。

> **Note**
>
> 完成本教程后，您可能需要删除 AWS CloudFormation 中的堆栈，以免继续产生费用。根据堆栈的事件状态验证是否已成功删除资源。

第 2 步：注册您的数据位置、创建 LF-Tag 本体并授予权限

在此步骤中，数据管家用户使用两个 LF 标签定义标签本体：`Confidential`和`Sensitive`，并允许特定的 IAM 委托人将新创建的 LF 标签附加到资源。

注册数据位置并定义 LF-Tag 本体

1. 以数据管家用户 (`lf-data-steward`) 的身份执行第一步，验证 Amazon S3 及 Lake Formation 中数据目录中的数据。

   a. 使用部署AWS CloudFormation堆栈时使用的密码登录 Lak `lf-data-steward` e Formation 控制台 https://console.aws.amazon.com/lakeformation/。

   b. 在导航窗格中的权限下，选择管理角色和任务。

   c. 在 “数据湖管理员” 部分选择 “添加”。

   d. 在添加管理员页面上，对于 IAM 用户和角色，选择用户`lf-data-steward`。

   e. 选择保存以添加 `lf-data-steward` 作为 Lake Formation 管理员。

2. 接下来，更新数据目录设置以使用 Lake Formation 权限来控制目录资源，而不是使用基于 IAM 的访问控制。

   a. 在导航窗格中的管理下，选择数据目录设置。

   b. 取消选中仅对新数据库使用 IAM 访问控制。

   c. 取消选中仅对新数据库中的新表使用 IAM 访问控制。

   d. 单击保存。

3. 接下来，注册数据湖的数据位置。

   a. 在导航窗格的管理下，选择数据湖位置。

   b. 选择注册位置。

   c. 在注册位置页面上，对于 Amazon S3 路径，输入`s3://lf-tagbased-demo-`*`Account-ID`*。

   d. 对于 IAM 角色，保留默认值 `AWSServiceRoleForLakeFormationDataAccess` 不变。

   e. 选择 Lake Formation 作为权限模式。

   f. 选择注册位置。

4. 接下来，通过定义 LF 标签来创建本体。

   a. 在导航窗格的 “权限” 下，选择 LF 标签和权限。 。

b. 选择添加 LF 标签。

c. 对于键，输入 Confidential。

d. 对于值，添加 True 和 False。

e. 选择添加 LF 标签。

f. 重复上述步骤，使用该值创建 LF-T **Sensitive** ag。True

您已经为本练习创建了所有必需的 LF 标签。

向 IAM 用户授予权限

1. 接下来，授权特定的 IAM 主体将新创建的 LF 标签附加到资源。

   a. 在导航窗格的 “权限” 下，选择 LF 标签和权限。

   b. 在 LF-tag 权限部分，选择授予权限。

   c. 对于权限类型，选择 LF-tag 键值对权限。

   d. 选择 IAM 用户和角色。

   e. 对于 IAM 用户和角色，搜索并选择 lf-data-engineer 角色。

   f. 在 LF-Tags 部分中，添加Confidential带有值True和的密钥和 w False ith 的keySensitive值。True

   g. 在 “权限” 下，为 “权限” 和 “可授予权限” 选择 “描述并关联”。

   h. 选择授权。

2. 接下来，授予 lf-data-engineer 在我们的数据目录和由 AWS CloudFormation 创建的底层 Amazon S3 存储桶上创建数据库的权限。

   a. 在导航窗格的 “管理” 下，选择 “管理角色和任务”。

   b. 在数据库创建者部分，选择授权。

   c. 对于 IAM 用户和角色，选择 lf-data-engineer 角色。

   d. 对于目录权限，选择创建数据库。

   e. 选择授权。

3. 接下来，向 lf-data-engineer 用户授予对 Amazon S3 存储桶 (s3://lf-tagbased-demo-*Account-ID*) 的权限。

   a. 在导航窗格中的权限下，选择数据位置。

b. 选择授权。

c. 选择我的账户。

d. 对于 IAM 用户和角色，选择 `lf-data-engineer` 角色。

e. 对于存储位置，输入通过 AWS CloudFormation 模板 (`s3://lf-tagbased-demo-`*`Account-ID`*) 创建的 Amazon S3 存储桶。

f. 选择授权。

4. 接下来，对与 LF- Tag 表达`lf-data-engineer`式关联的资源授予可授予权限。`Confidential=True`

a. 在导航窗格的权限下，选择数据湖权限。

b. 选择授权。

c. 选择 IAM 用户和角色。

d. 选择角色 `lf-data-engineer`。

e. 在 LF 标签或目录资源部分中，选择与 LF 标签匹配的资源。

f. 选择 “添加 LF-Tag 键值对”。

g. 添加值为 `True` 的键 `Confidential`。

h. 在数据库权限部分，为数据库权限和可授予的权限选择描述。

i. 在 “表权限” 部分中，为表权限和可授予权限选择描述、选择和更改。

j. 选择授权。

5. 接下来，对与 LF-Tag 表达`lf-data-engineer`式关联的资源授予可授予权限。`Confidential=False`

a. 在导航窗格的权限下，选择数据湖权限。

b. 选择授权。

c. 选择 IAM 用户和角色。

d. 选择角色 `lf-data-engineer`。

e. 选择通过 LF 标签匹配的资源。

f. 选择添加 LF 标签。

g. 添加值为 `False` 的键 `Confidential`。

h. 在数据库权限部分，为数据库权限和可授予的权限选择描述。

i. 在表和列权限 部分，请勿选择任何内容。

j. 选择授权。

6. 接下来，我们为与 LF-Tag 键值对关联的资源授`lf-data-engineer`予可授予的权限，以及。`Confidential=False Sensitive=True`

   a. 在导航窗格的权限下，选择数据权限。

   b. 选择授权。

   c. 选择 IAM 用户和角色。

   d. 选择角色 `lf-data-engineer`。

   e. 在 LF 标签或目录资源部分下，选择与 LF 标签匹配的资源。

   f. 选择添加 LF 标签。

   g. 添加值为 `False` 的键 `Confidential`。

   h. 选择 “添加 LF-Tag 键值对”。

   i. 添加值为 `True` 的键 `Sensitive`。

   j. 在数据库权限部分，为数据库权限和可授予的权限选择描述。

   k. 在 “表权限” 部分中，为表权限和可授予权限选择描述、选择和更改。

   l. 选择授权。

第 3 步：创建 Lake Formation 数据库

在此步骤中，您将创建两个数据库，并将 LF-Tag 附加到数据库和特定列以进行测试。

创建用于数据库级别访问的数据库和表

1. 首先，创建数据库`tag_database`、表`source_data`，并附加相应的 LF 标签。

   a. 在 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 的 “数据目录” 下，选择 “数据库”。

   b. 选择创建数据库。

   c. 在名称中，输入 `tag_database`。

   d. 对于位置，输入通过 AWS CloudFormation 模板 (`s3://lf-tagbased-demo-`*`Account-ID`*`/tag_database/`) 创建的 Amazon S3 位置。

   e. 取消选中仅对此数据库中的新表使用 IAM 访问控制。

   f. 选择创建数据库。

2. 接下来，在 `tag_database` 中创建一个新表。

a. 在数据库页面上，选择数据库 tag_database。

b. 选择查看表，然后单击创建表。

c. 在名称中，输入 source_data。

d. 对于 Database (数据库)，选择 tag_database 数据库。

e. 对于表格格式，请选择标准AWS Glue表格。

f. 对于数据位置，选择我的账户中的指定路径。

g. 在“包含路径”中，输入要通过 AWS CloudFormation 模板 (s3://lf-tagbased-demo*Account-ID*/tag_database/) 创建的 tag_database 的路径。

h. 对于数据格式，选择 CSV。

i. 在上传架构下，输入以下 JSON 列结构数组以创建架构：

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

              },
                 {
                   "Name": "store_and_fwd_flag",
                   "Type": "string"
              },
                 {
                   "Name": "ratecodeid",
                   "Type": "string"

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

              },
```

```
                {
                     "Name": "dolocationid",
                     "Type": "string"

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

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

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

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

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

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

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

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

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

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

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

                }
]
```

j. 选择上传。上传架构后，表架构看上去应如以下屏幕截图所示：

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

k. 选择提交。

3. 接下来，在数据库级别附加 LF 标签。

   a. 在数据库页面上，找到并选择 tag_database。
   b. 在 “操作” 菜单上，选择 “编辑 LF 标签”。
   c. 选择分配新的 LF 标签。
   d. 对于分配的密钥，请选择您之前创建的 Confidential LF-Tag。
   e. 对于值，选择 True。
   f. 选择保存。

   这样就完成了对 tag_database 数据库的 LF-Tag 赋值。

创建用于列级别访问的数据库和表

重复以下步骤创建数据库col_tag_database和表source_data_col_lvl，并在列级别附加 LF-Tag。

1. 在数据库页面上，选择创建数据库。
2. 在名称中，输入 col_tag_database。
3. 对于位置，输入通过 AWS CloudFormation 模板 (s3://lf-tagbased-demo-*Account-ID*/col_tag_database/) 创建的 Amazon S3 位置。
4. 取消选中仅对此数据库中的新表使用 IAM 访问控制。
5. 选择创建数据库。
6. 在数据库页面上，选择您的新数据库 (col_tag_database)。
7. 选择 “查看表”，然后单击 “创建表”。
8. 在名称中，输入 source_data_col_lvl。
9. 对于数据库，选择您的新数据库 (col_tag_database)。
10. 对于表格格式，请选择标准AWS Glue表格。
11. 对于数据位置，选择我的账户中的指定路径。
12. 输入 col_tag_database (s3://lf-tagbased-demo-*Account-ID*/col_tag_database/) 的 Amazon S3 路径。
13. 对于数据格式，选择 CSV。
14. 在Upload schema下方，输入以下架构 JSON：

```
[
                {
                     "Name": "vendorid",
                     "Type": "string"


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


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


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


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


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


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


                },
```

```
                {
                  "Name": "passenger_count",
                  "Type": "string"


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


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


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


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


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


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


              },
              {
                  "Name": "ehail_fee",
```

```
                    "Type": "string"


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


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


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


               }
]
```

15. 选择Upload。上传架构后，表架构看上去应如以下屏幕截图所示：

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

16. 选择提交以完成表的创建。
17. 现在，将 Sensitive=True LF-Tag 关联到列和vendorid。fare_amount

    a. 在表页面上，选择您创建的表 (source_data_col_lvl)。
    b. 在 “操作” 菜单上，选择 “架构”。
    c. 选择该列vendorid并选择 “编辑 LF-Tags”。
    d. 对于已分配的键，选择区分大小写。
    e. 对于值，选择 True。
    f. 选择保存。
18. 接下来，将 Confidential=False LF-Tag 关联到。col_tag_database这是在登录col_tag_database时lf-data-analyst能够描述数据库所必需的Amazon Athena。

    a. 在数据库页面上，找到并选择 col_tag_database。
    b. 在 “操作” 菜单上，选择 “编辑 LF 标签”。
    c. 选择分配新的 LF 标签。
    d. 对于分配的密钥，请选择您之前创建的 Confidential LF-Tag。
    e. 对于值，选择 False。
    f. 选择保存。

第 4 步：授予表权限

使用 LF 标签 Confidential 和 Sensitive 向数据分析师授予使用数据库 tag_database 和表 col_tag_database 的权限。

1. 按照以下步骤向lf-data-analyst用户授予对与 LF-Tag Confidential=True (database: tag_Database) 关联的对象的权限，使其拥有Describe数据库和对表的权限。Select

    a. 通过 https://console.aws.amazon.com/lakeformation/ 登录 Lake Formation 控制台 lf-data-engineer
    b. 在 “权限” 下，选择 “数据湖权限”。
    c. 选择授权。
    d. 在主体下，选择 IAM 用户和角色。
    e. 对于 IAM 用户和角色，选择 lf-data-analyst。
    f. 在 LF 标签或目录资源下，选择与 LF 标签匹配的资源。

g. 选择添加 LF 标签。

h. 对于键，选择 Confidential。

i. 对于值，选择 True。

j. 对于数据库权限，选择 Describe。

k. 对于表权限，选择选择和描述。

l. 选择授权。

2. 接下来，重复这些步骤，向数据分析师授予使用 LF-Tag 表达式的权限。Confidential=False当以 lf-data-analyst 身份从 Amazon Athena 登录时，使用此 LF 标签描述 col_tag_database 和表 source_data_col_lvl。

a. 通过 https://console.aws.amazon.com/lakeformation/ 登录 Lake Formation 控制台 lf-data-engineer

b. 在数据库页面上，选择数据库 col_tag_database。

c. 选择操作和授权。

d. 在主体下，选择 IAM 用户和角色。

e. 对于 IAM 用户和角色，选择 lf-data-analyst。

f. 选择与 LF 标签匹配的资源。

g. 选择添加 LF 标签。

h. 对于键，选择 Confidential。

i. 对于值，选择 False。

j. 对于数据库权限，选择 Describe。

k. 对于表权限，请勿选择任何内容。

l. 选择授权。

3. 接下来，重复上述步骤，向数据分析师授予和的 LF-Tag 表达式的Confidential=False权限。Sensitive=True当以 lf-data-analyst 身份从 Amazon Athena 登录时，使用此 LF 标签描述 col_tag_database 和表 source_data_col_lvl（列级别）。

a. 通过 https://console.aws.amazon.com/lakeformation/ 以 lf-data-engineer 身份登录 Lake Formation 控制台。

b. 在“数据库”页面上，选择数据库 col_tag_database。

c. 选择操作和授权。

d. 在主体下，选择 IAM 用户和角色。

e. 对于 IAM 用户和角色，选择 lf-data-analyst。
f. 选择与 LF 标签匹配的资源。
g. 选择添加 LF 标签。
h. 对于键，选择 Confidential。
i. 对于值，选择 False。
j. 选择添加 LF 标签。
k. 对于键，选择 Sensitive。
l. 对于值，选择 True。
m. 对于数据库权限，选择 Describe。
n. 对于表权限，选择 Select 和 Describe。
o. 选择授权。

第 5 步：在 Amazon Athena 中运行查询以验证权限

在这一步，使用 Amazon Athena 对两个表（source_data and source_data_col_lvl）运行 SELECT 查询。使用 Amazon S3 路径作为查询结果位置（s3://lf-tagbased-demo-*Account-ID*/athena-results/）。

1. 通过 https://console.aws.amazon.com/athena/ 以 lf-data-analyst 用户身份登录 Athena 控制台。
2. 在 Athena 查询编辑器中，在左侧面板中选择 tag_database。
3. 选择 source_data 旁边的其他菜单选项图标（三个竖点），然后选择预览表。
4. 选择运行查询。

   查询可能需要几分钟时间才能运行。查询显示了输出中的所有列，因为 LF 标签是在数据库级别关联的，并且 source_data 表会自动从数据库 tag_database 继承 LF-tag。
5. 使用col_tag_database 和 source_data_col_lvl 运行另一个查询。

   第二个查询返回标记为 Non-Confidential 和 Sensitive 的两个列。
6. 您还可以查看您未获得策略授权的列上的 Lake Formation 基于标签的访问策略行为。当从表 source_data_col_lvl 中选择未标记的列时，Athena 会返回错误。例如，您可以运行以下查询来选择未标记的列 geolocationid：

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

第 6 步：清除 AWS 资源

为帮助避免您的 AWS 账户产生不必要的费用，您可以删除您在本教程中使用的 AWS 资源。

1. 以 `lf-data-engineer` 身份登录 Lake Formation 控制台并删除数据库 `tag_database` 和 `col_tag_database`。
2. 接下来，以 `lf-data-steward` 身份登录并清除上面授予 `lf-data-engineer` 和 `lf-data-analyst.` 的所有 LF 标签权限、数据权限和数据位置权限。
3. 使用您用于部署 AWS CloudFormation 堆栈的 IAM 凭证，以账户所有者的身份登录 Amazon S3 控制台。
4. 删除以下存储桶：
   - lf-tagbased-demo-accesslogs-##
   - lf-tagbased-demo-##
5. 通过 https://console.aws.amazon.com/cloudformation 登录 AWS CloudFormation 控制台，然后删除您创建的堆栈。等待堆栈状态变为 `DELETE_COMPLETE`。

使用行级别访问控制保护数据湖

AWS Lake Formation 行级别权限允许您根据数据合规性和监管策略提供对表中特定行的访问权限。如果您有一些存储着数十亿条记录的大型表，则需要一种方法来使不同的用户和团队仅能访问允许他们查看的数据。行级别访问控制是一种简单而高效的数据保护方法，同时允许用户访问他们执行工作所需的数据。Lake Formation 通过识别哪些主体访问了哪些数据、何时以及通过哪些服务访问的，从而提供集中式审计与合规性报告。

在本教程中，您将了解行级别访问控制在 Lake Formation 中的工作原理，以及如何设置这种访问控制。

本教程中提供了一个用于快速设置所需资源的 AWS CloudFormation 模板。您可以查看和自定义该模板来满足自己的需求。

主题

- 第 3 步：设置数据筛选条件并授予权限
- 第 4 步：使用数据筛选条件进行查询
- 第 5 步：清除 AWS 资源

目标受众

本教程适用于数据管家、数据工程师和数据分析师。下表列出了数据所有者和数据使用者的角色和责任。

角色	描述
IAM 管理员	可以创建用户和角色以及 Amazon Simple Storage Service (Amazon S3) 存储桶的用户。具有 `AdministratorAccess` AWS 托管策略。
数据湖管理员	负责设置数据湖、创建数据筛选条件以及向数据分析师授予权限的用户。
数据分析人员	可以对数据湖运行查询的用户。居住在不同国家/地区（在我们的使用案例中，为美国和日本）的数据分析师只能为位于自己国家/地区的客户分析商品评论，并且，出于合规性原因，他们应该无法看到其他国家/地区的客户的数据。

先决条件

在开始学习本教程之前，您必须拥有 AWS 账户，以便能够以具有正确权限的管理用户身份登录。有关更多信息，请参阅完成初始 AWS 配置任务。

本教程假定您熟悉 IAM。有关 IAM 的信息，请参阅 IAM 用户指南。

更改 Lake Formation 设置

> **Important**
>
> 在启动 AWS CloudFormation 模板之前，请按照以下步骤禁用 Lake Formation 中的仅对新数据库/表使用 IAM 访问控制选项：

1. 在美国东部（弗吉尼亚州北部）区域或美国西部（俄勒冈州）区域通过 https://console.aws.amazon.com/lakeformation/ 登录 Lake Formation 控制台。
2. 在数据目录下，选择设置。
3. 取消选择仅对新数据库使用 IAM 访问控制和仅对新数据库中的新表使用 IAM 访问控制。
4. 选择 Save（保存）。

第 1 步：调配资源

本教程中提供了一个用于进行快速设置的 AWS CloudFormation 模板。您可以查看和自定义该模板来满足自己的需求。AWS CloudFormation 模板可生成以下资源：

- 以下用户及其策略：
  - DataLakeAdmin
  - DataAnalystUS
  - DataAnalystJP
- Lake Formation 数据湖设置和权限
- Lambda 函数（适用于 Lambda 支持的 AWS CloudFormation 自定义资源），用于将示例数据文件从公共 Amazon S3 存储桶复制到您的 Amazon S3 存储桶
- 用作数据湖的 Amazon S3 存储桶
- AWS Glue Data Catalog 数据库、表和分区

创建您的资源

要使用 AWS CloudFormation 模板创建这些资源，请按照以下步骤操作。

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/cloudformation 登录 AWS CloudFormation 控制台。

2. 选择启动堆栈。
3. 在创建堆栈屏幕上，选择下一步。
4. 输入堆栈名称。
5. 对于 DatalakeAdminUserName 和 DatalakeAdminUserPassword，输入您的 IAM 用户名和数据湖管理用户密码。
6. 对于 DataAnalystUsUserName 和 DataAnalystUsUserPassword，输入您想要的用户名和密码作为负责分析美国市场的数据分析师用户的用户名和密码。
7. 对于 DataAnalystJpUserName 和 DataAnalystJpUserPassword，输入您想要的用户名和密码作为负责分析日本市场的数据分析师用户的用户名和密码。
8. 对于 DataLakeBucketName，输入数据存储桶的名称。
9. 对于 DatabaseName 和 TableName，保留默认值。
10. 选择 Next（下一步）。
11. 在下一页上，选择下一步。
12. 查看最后页面上的详细信息，然后选择我确认 AWS CloudFormation 可以创建 IAM 资源。
13. 选择 Create（创建）。

    堆栈创建可能需要一分钟才能完成。

第 2 步：在不使用数据筛选条件的情况下进行查询

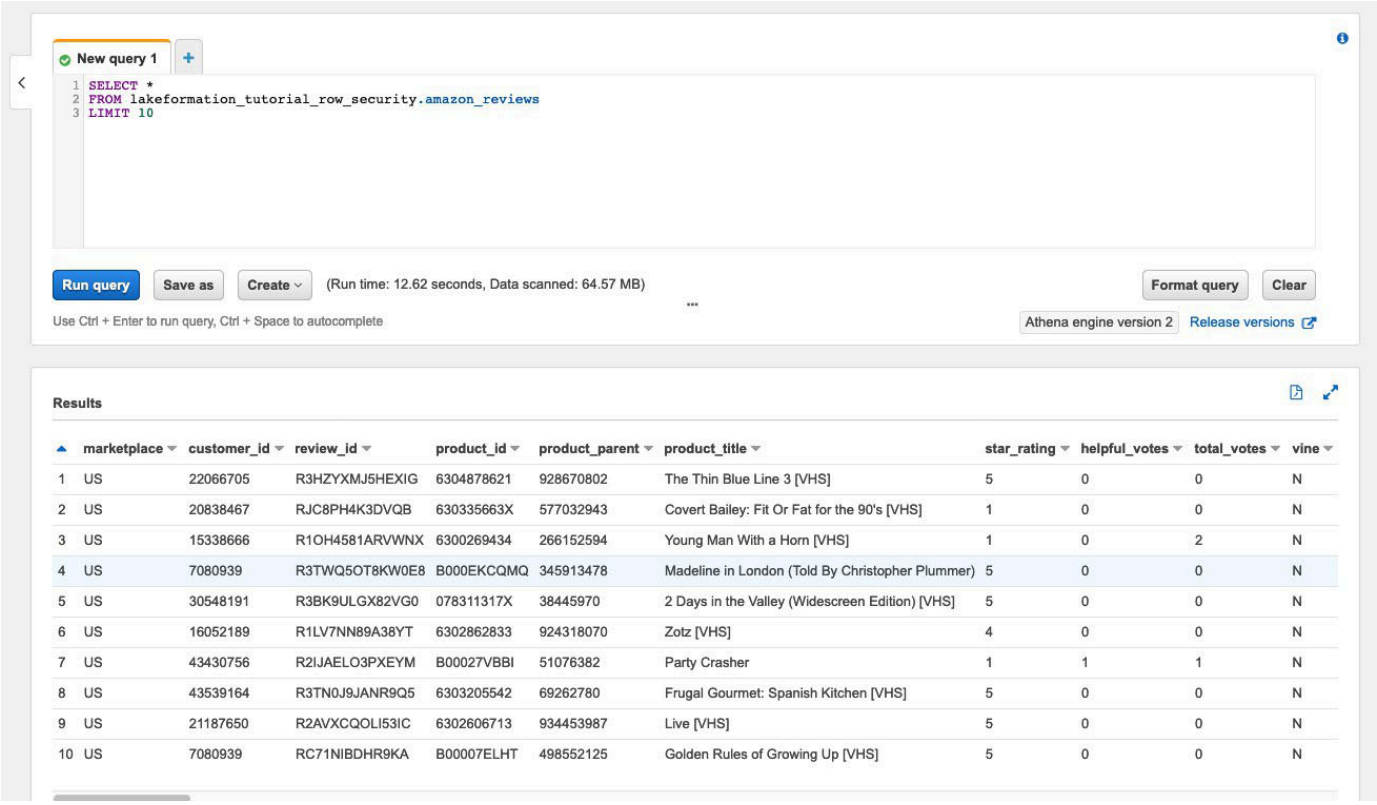
设置环境后，您可以查询商品评论表。首先在没有行级别访问控制的情况下查询表，以确保您可以看到数据。如果您是首次在 Amazon Athena 中运行查询，则需要配置查询结果位置。

在没有行级别访问控制的情况下查询表

1. 通过 https://console.aws.amazon.com/athena/ 以 `DatalakeAdmin` 用户身份登录 Athena 控制台，然后运行以下查询：

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

   以下屏幕截图显示了查询结果。此表只有一个分区 `product_category=Video`，因此每条记录都是对视频产品的评论。

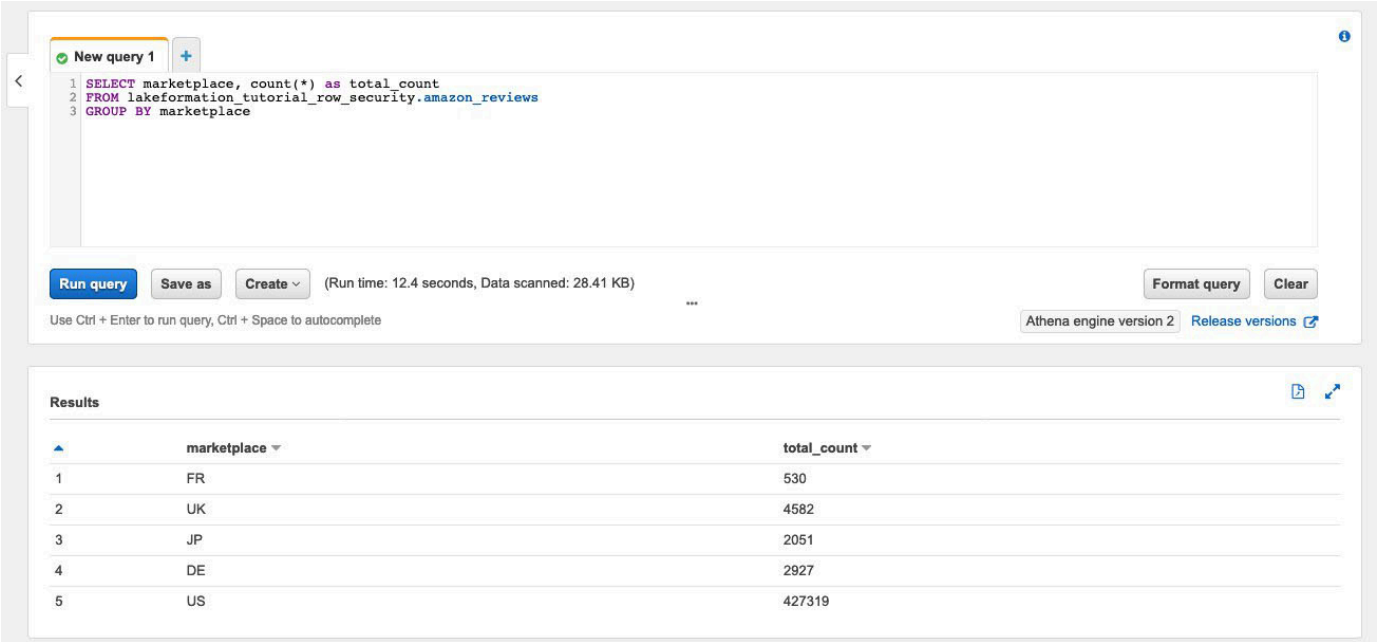


	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	N
3	US	15338666	R1OH4581ARVWNX	6300269434	266152594	Young Man With a Horn [VHS]	1	0	2	N
4	US	7080939	R3TWQ5OT8KW0E8	B000EKCQMQ	345913478	Madeline in London (Told By Christopher Plummer)	5	0	0	N
5	US	30548191	R3BK9ULGX82VG0	078311317X	38445970	2 Days in the Valley (Widescreen Edition) [VHS]	5	0	0	N
6	US	16052189	R1LV7NN89A38YT	6302862833	924318070	Zotz [VHS]	4	0	0	N
7	US	43430756	R2IJAELO3PXEYM	B00027VBBI	51076382	Party Crasher	1	1	1	N
8	US	43539164	R3TN0J9JANR9Q5	6303205542	69262780	Frugal Gourmet: Spanish Kitchen [VHS]	5	0	0	N
9	US	21187650	R2AVXCQOLI53IC	6302606713	934453987	Live [VHS]	5	0	0	N
10	US	7080939	RC71NIBDHR9KA	B00007ELHT	498552125	Golden Rules of Growing Up [VHS]	5	0	0	N

2. 接下来，运行聚合查询以检索每个 marketplace 的记录总数。

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

以下屏幕截图显示了查询结果。marketplace 列中有五个不同的值。在后续步骤中，您将使用 marketplace 列设置基于行的筛选条件。



第 3 步：设置数据筛选条件并授予权限

本教程使用了两位数据分析师：一位负责分析美国市场，另一位负责分析日本市场。每位分析师仅使用 Athena 分析其特定市场的客户评价。创建两个不同的数据筛选条件，一个供负责分析美国市场的分析师使用，另一个供负责分析日本市场的分析师使用。然后，授予分析师相应的权限。

创建数据筛选条件并授予权限

1. 创建筛选条件以限制对 USmarketplace 数据的访问。
   a. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/lakeformation/ 以 DatalakeAdmin 用户身份登录 Lake Formation 控制台。
   b. 选择数据筛选条件。
   c. 选择创建新筛选条件。
   d. 对于数据筛选条件名称，输入 amazon_reviews_US。
   e. 对于目标数据库，选择数据库 lakeformation_tutorial_row_security。
   f. 对于目标表，选择表 amazon_reviews。
   g. 对于列级别访问权限，保留默认值。
   h. 对于行筛选表达式，输入 marketplace='US'。
   i. 请选择 Create filter（创建筛选器）。
2. 创建筛选条件以限制对日本 marketplace 数据的访问。

a. 在数据筛选条件页面上，选择创建新筛选条件。

b. 对于数据筛选条件名称，输入 amazon_reviews_JP。

c. 对于目标数据库，选择数据库 lakeformation_tutorial_row_security。

d. 对于目标表，选择 table amazon_reviews。

e. 对于列级别访问权限，保留默认值。

f. 对于“行筛选表达式”，输入 marketplace='JP'。

g. 请选择 Create filter（创建筛选器）。

3. 接下来，使用这些数据筛选条件向数据分析师授予权限。按照以下步骤向美国数据分析师 (DataAnalystUS) 授予权限：

a. 在权限下，选择数据湖权限。

b. 在数据权限下，选择授权。

c. 对于主体，选择 IAM 用户和角色，然后选择角色 DataAnalystUS。

d. 对于 LF 标签或目录资源，选择命名数据目录资源。

e. 对于 Database (数据库)，请选择 lakeformation_tutorial_row_security。

f. 对于可选表，选择 amazon_reviews。

g. 对于数据筛选条件 — 可选，选择 amazon_reviews_US。

h. 对于数据筛选条件权限，选择选择。

i. 选择授权。

4. 按照以下步骤向日本数据分析师 (DataAnalystJP) 授予权限：

a. 在权限下，选择数据湖权限。

b. 在数据权限下，选择授权。

c. 对于主体，选择 IAM 用户和角色，然后选择角色 DataAnalystJP。

d. 对于 LF 标签或目录资源，选择命名数据目录资源。

e. 对于 Database (数据库)，请选择 lakeformation_tutorial_row_security。

f. 对于可选表，选择 amazon_reviews。

g. 对于数据筛选条件 — 可选，选择 amazon_reviews_JP。

h. 对于数据筛选条件权限，选择选择。

i. 选择授权。

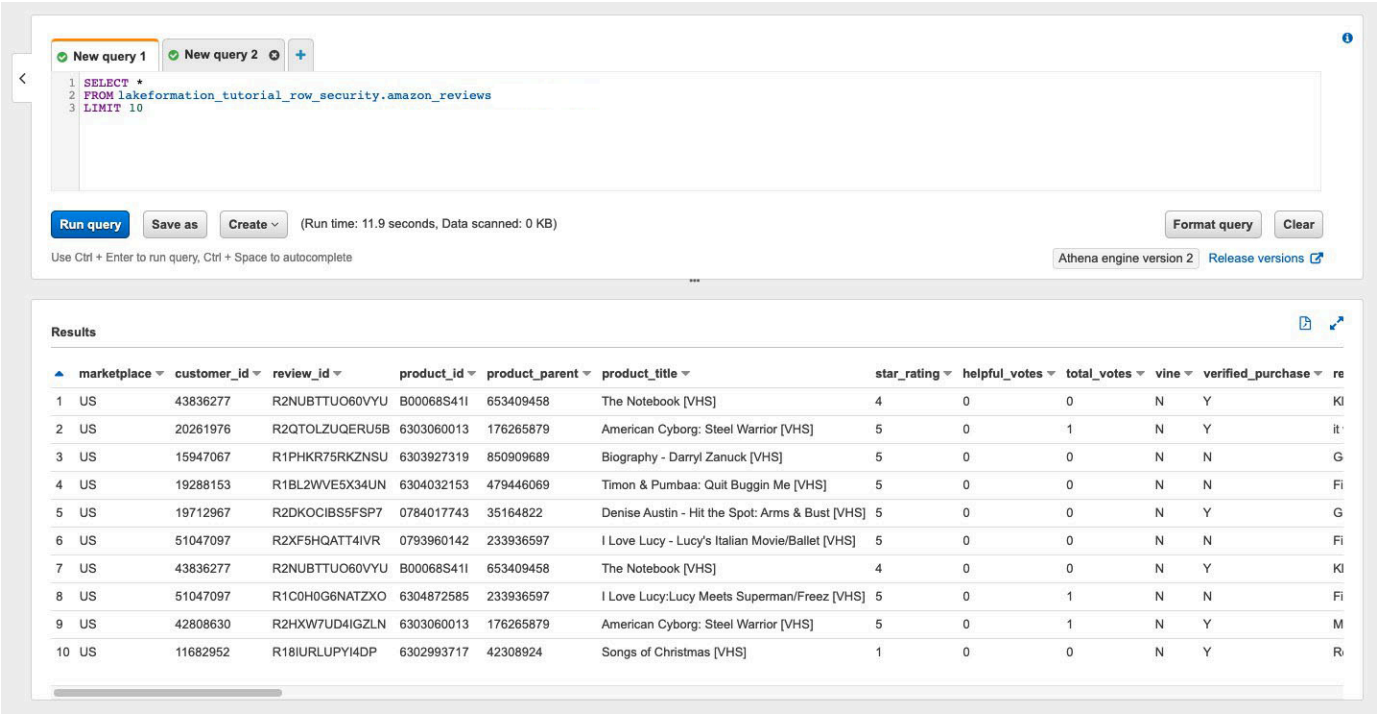
第 4 步：使用数据筛选条件进行查询

将数据筛选条件附加到商品评论表后，运行一些查询，看一看 Lake Formation 是如何强制实施权限的。

1. 通过 https://console.aws.amazon.com/athena/ 以 DataAnalystUS 用户身份登录 Athena 控制台。
2. 运行以下查询以检索一些记录，对于这些记录，是根据我们定义的行级别权限进行筛选的：

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

以下屏幕截图显示了查询结果。



	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine	verified_purchase	re
1	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	KI
2	US	20261976	R2QTOLZUQERU5B	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	it
3	US	15947067	R1PHKR75RKZNSU	6303927319	850909689	Biography - Darryl Zanuck [VHS]	5	0	0	N	N	G
4	US	19288153	R1BL2WVE5X34UN	6304032153	479446069	Timon & Pumbaa: Quit Buggin Me [VHS]	5	0	0	N	N	Fi
5	US	19712967	R2DKOCIBS5FSP7	0784017743	35164822	Denise Austin - Hit the Spot: Arms & Bust [VHS]	5	0	0	N	Y	G
6	US	51047097	R2XF5HQATT4IVR	0793960142	233936597	I Love Lucy - Lucy's Italian Movie/Ballet [VHS]	5	0	0	N	N	Fi
7	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	KI
8	US	51047097	R1C0H0G6NATZXO	6304872585	233936597	I Love Lucy:Lucy Meets Superman/Freez [VHS]	5	0	1	N	N	Fi
9	US	42808630	R2HXW7UD4IGZLN	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	M
10	US	11682952	R18IURLUPYI4DP	6302993717	42308924	Songs of Christmas [VHS]	1	0	0	N	Y	R

3. 同样，运行查询以统计每个市场的记录总数。

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

查询结果中仅显示 marketplace US。这是因为只允许用户查看 marketplace 列值等于 US 的行。

4. 切换到 DataAnalystJP 用户并运行相同的查询。

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

查询结果仅显示属于 JPmarketplace 的记录。

5. 运行查询来统计每个 marketplace 的记录总数。

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

查询结果仅显示属于 JPmarketplace 的行。

第 5 步：清除 AWS 资源

清理资源

为帮助避免您的 AWS 账户产生不必要的费用，您可以删除您在本教程中使用的 AWS 资源。

- 删除 Cloud Formation 堆栈。

使用 Lake Formation 基于标签的访问控制和命名资源共享数据湖

本教程演示如何将 AWS Lake Formation 配置为安全地与多个公司、组织或业务部门共享存储在数据湖中的数据，而不必复制整个数据库。通过使用 Lake Formation 跨账户访问控制，有两个选项可用于与其他 AWS 账户共享您的数据库和表：

- Lake Formation 基于标签的访问控制（推荐）

  Lake Formation 基于标签的访问控制是一种授权策略，它根据属性来定义权限。在 Lake Formation 中，这些属性被称为 LF 标签。有关详细信息，请参考 使用 Lake Formation 基于标签的访问控制管理数据湖。

- Lake Formation 命名资源

Lake Formation 命名资源方法是一种授权策略，它定义对资源的权限。资源包括数据库、表和列。数据湖管理员可以分配和撤销对 Lake Formation 资源的权限。有关详细信息，请参考 Lake Formation 中的跨账户数据共享。

如果数据湖管理员喜欢显式授予对单个资源的权限，我们建议使用命名资源。当您使用命名资源方法向外部账户授予对数据目录资源的 Lake Formation 权限时，Lake Formation 使用 AWS Resource Access Manager (AWS RAM) 来共享该资源。

主题

目标受众

本教程适用于数据管家、数据工程师和数据分析师。谈及共享来自 AWS Glue 的数据目录表和在 Lake Formation 中管理权限，生产账户中的数据管家根据其支持的功能拥有职能所有权，并且可以向各种使用者、外部组织和账户授予访问权限。下表列出了本教程中使用的角色：

角色	描述
DataLakeAdminProducer	数据湖管理员 IAM 用户具有以下访问权限： • 对数据目录中所有资源的完整读取、写入和更新权限 • 能够授予对资源的权限 • 可为共享表创建资源链接 • 可以将 LF 标签附加到资源，从而根据数据管家创建的所有策略为主体提供访问权限

角色	描述
DataLakeAdminConsumer	数据湖管理员 IAM 用户具有以下访问权限： • 对数据目录中所有资源的完整读取、写入和更新权限 • 能够授予对资源的权限 • 可为共享表创建资源链接 • 可以将 LF 标签附加到资源，从而根据数据管家创建的所有策略为主体提供访问权限
DataAnalyst	DataAnalyst 用户具有以下访问权限： • 对通过 Lake Formation 基于标签的访问策略或使用命名资源方法共享的资源的精细访问权限

在制作者账户中配置 Lake Formation 数据目录设置

在开始学习本教程之前，您必须拥有 AWS 账户，以便能够以具有正确权限的管理用户身份登录。有关更多信息，请参阅完成初始 AWS 配置任务。

本教程假定您熟悉 IAM。有关 IAM 的信息，请参阅 IAM 用户指南。

在制作者账户中配置 Lake Formation 数据目录设置

> **Note**
>
> 在本教程中，拥有源表的账户称为制作者账户，需要访问源表的账户称为使用者账户。

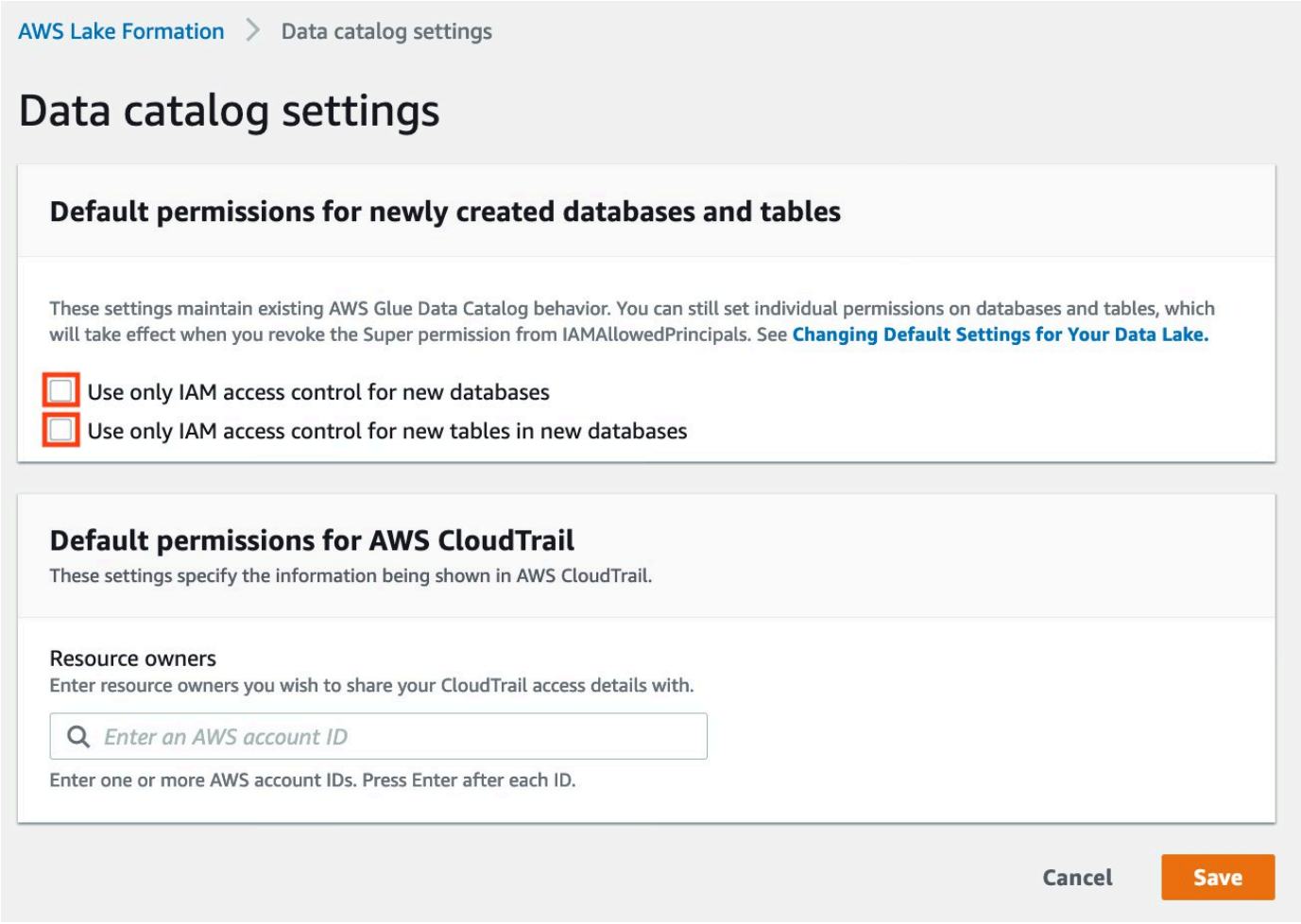
Lake Formation 提供了自己的权限管理模型。为了保持与 IAM 权限模型的向后兼容性，默认情况下会向组 `IAMAllowedPrincipals` 授予对所有现有 AWS Glue Data Catalog 资源的 `Super` 权限。此外，还为新的数据目录资源启用了仅使用 IAM 访问控制设置。本教程利用 Lake Formation 权限来使用细粒度访问控制，并使用 IAM 策略进行粗粒度访问控制。有关详细信息，请参阅精细访问控制的方法。因此，在使用 AWS CloudFormation 模板进行快速设置之前，需要在制作者账户中更改 Lake Formation 数据目录设置。

> **Important**
>
> 此设置会影响所有新创建的数据库和表，因此我们强烈建议使用非生产账户或新账户完成本教程。此外，如果您使用的是共享账户（例如贵公司的开发账户），请确保它不会影响其他资源。如果您希望保留默认安全设置，则在与其他账户共享资源时必须完成一个额外步骤，即撤销 `IAMAllowedPrincipals` 对数据库或表的默认 Super 权限。我们会在本教程的后面部分中讨论详细信息。

要在制作者账户中配置 Lake Formation 数据目录设置，请完成以下步骤：

1. 以管理用户或具有 Lake Formation `PutDataLakeSettings` API 权限的用户身份使用制作者账户登录 AWS Management Console。
2. 在 Lake Formation 控制台的导航窗格中，在数据目录下选择设置。
3. 取消选择仅对新数据库使用 IAM 访问控制和仅对新数据库中的新表使用 IAM 访问控制

   选择 Save（保存）。



此外，您可以在管理角色和任务的数据库创建者下删除 `IAMAllowedPrincipals` 的 `CREATE_DATABASE` 权限。只有这样，您才能通过 Lake Formation 权限控制谁可以创建新数据库。

第 1 步：使用 AWS CloudFormation 模板调配资源

制作者账户的 CloudFormation 模板可生成以下资源：

- 要用作数据湖的 Amazon S3 存储桶。
- Lambda 函数（用于 Lambda 支持的 AWS CloudFormation 自定义资源）。我们使用该函数将示例数据文件从公共 Amazon S3 存储桶复制到您的 Amazon S3 存储桶中。
- IAM 用户和策略：DataLakeAdminProducer。
- 适当的 Lake Formation 设置和权限，包括：
  - 在制作者账户中定义 Lake Formation 数据湖管理员

  - 将 Amazon S3 存储桶注册为 Lake Formation 数据湖位置（制作者账户）
- AWS Glue Data Catalog 数据库、表和分区。由于有两个跨 AWS 账户共享资源的选项，因此，此模板创建了两组单独的数据库和表。

使用者账户的 AWS CloudFormation 模板可生成以下资源：

- IAM 用户和策略：
  - DataLakeAdminConsumer
  - DataAnalyst
- AWS Glue Data Catalog 数据库。此数据库用于创建指向共享资源的资源链接。

在制作者账户中创建您的资源

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/cloudformation 登录 AWS CloudFormation 控制台。
2. 选择启动堆栈。
3. 选择 Next（下一步）。
4. 对于堆栈名称，为堆栈输入名称，如 `stack-producer`。
5. 在用户配置部分，输入 `ProducerDatalakeAdminUserName` 和 `ProducerDatalakeAdminUserPassword` 的用户名和密码。
6. 在 DataLakeBucketName 中，输入您的数据湖存储桶的名称。此名称需要全局唯一。
7. 对于 DatabaseName 和 TableName，保留默认值。
8. 选择 Next（下一步）。
9. 在下一页上，选择下一步。
10. 查看最后页面上的详细信息，然后选择我确认 AWS CloudFormation 可以创建 IAM 资源。
11. 选择 Create（创建）。

    堆栈创建过程可能最多需要一分钟。

在使用者账户中创建您的资源

1. 在美国东部（弗吉尼亚州北部）区域通过 https://console.aws.amazon.com/cloudformation 登录 AWS CloudFormation 控制台。

2. 选择启动堆栈。
3. 选择 Next（下一步）。
4. 对于堆栈名称，为堆栈输入名称，如 `stack-consumer`。
5. 在用户配置部分，输入 `ConsumerDatalakeAdminUserName` 和 `ConsumerDatalakeAdminUserPassword` 的用户名和密码。
6. 对于 `DataAnalystUserName` 和 `DataAnalystUserPassword`，为数据分析师 IAM 用户输入所需的用户名和密码。
7. 在 DataLakeBucketName 中，输入您的数据湖存储桶的名称。此名称需要全局唯一。
8. 对于 DatabaseName，保留默认值。
9. 对于 `AthenaQueryResultS3BucketName`，输入用于存储 Amazon Athena 查询结果的 Amazon S3 存储桶的名称。如果没有 Amazon S3 存储桶，请创建 Amazon S3 存储桶。
10. 选择 Next（下一步）。
11. 在下一页上，选择下一步。
12. 查看最后页面上的详细信息，然后选择我确认 AWS CloudFormation 可以创建 IAM 资源。
13. 选择 Create（创建）。

    堆栈创建过程可能最多需要一分钟。

> **Note**
>
> 完成本教程的学习后，删除 AWS CloudFormation 中的堆栈以免产生费用。根据堆栈的事件状态验证是否已成功删除资源。

第 2 步：Lake Formation 跨账户共享先决条件

在与 Lake Formation 共享资源之前，基于标签的访问控制方法和命名资源方法都需要满足一些先决条件。

完成基于标签的访问控制跨账户数据共享先决条件

- 有关跨账户数据共享要求的更多信息，请参阅“跨账户数据共享”一章中的先决条件一节。

  要在跨账户版本设置为版本 3 或更高版本的情况下共享数据目录资源，授予者需要拥有您账户的 AWS 托管策略 `AWSLakeFormationCrossAccountManager` 中定义的 IAM 权限。

如果您使用的跨账户版本设置为版本 1 或版本 2，则必须先将以下 JSON 权限对象添加到制作者账户中的数据目录资源策略中，然后才能使用基于标签的访问控制方法授予对资源的跨账户访问权限。当 `glue:EvaluatedByLakeFormationTags` 为 true 时，这可授予使用者账户访问数据目录的权限。此外，对于您使用 Lake Formation 权限标签向使用者账户授予其权限的资源，此条件也适用。您要向其授予权限的每个 AWS 账户都需要此策略。

以下策略必须位于 `Statement` 元素内。我们会在下一节中讨论完整的 IAM 策略。

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

完成命名资源方法跨账户共享先决条件

1. 如果您的账户中没有数据目录资源策略，则您执行的 Lake Formation 跨账户授权将照常进行。但是，如果存在数据目录资源策略，则必须在其中添加以下语句，以便能够成功使用命名资源方法完成跨账户授权。如果您计划仅使用命名资源方法或仅使用基于标签的访问控制方法，则可以跳过这一步。在本教程中，我们评估了这两种方法，并且我们需要添加以下策略。

   以下策略必须位于 `Statement` 元素内。我们会在下一节中讨论完整的 IAM 策略。

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

2. 接下来，使用 AWS Command Line Interface (AWS CLI) 添加 AWS Glue Data Catalog 资源策略。

   如果您同时使用基于标签的访问控制方法和命名资源方法授予跨账户权限，则在添加上述策略时必须将 EnableHybrid 参数设置为“true”。因为控制台目前不支持此选项，因此您必须使用 glue:PutResourcePolicy API 和 AWS CLI。

   首先，创建一个策略文档（例如 policy.json），然后添加上述两个策略。将 *consumer-account-id* 替换为获得授权的 AWS 账户的*## ID*，将 *region* 替换为包含您将授予对其权限的数据库和表的数据目录区域，并将 *account-id* 替换为制作者 AWS 账户ID。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "ram.amazonaws.com"
            },
            "Action": "glue:ShareResource",
            "Resource": [
                "arn:aws:glue:region:account-id:table/*/*",
                "arn:aws:glue:region:account-id:database/*",
                "arn:aws:glue:region:account-id:catalog"
            ]
```

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

输入以下 AWS CLI 命令。将 *glue-resource-policy* 替换为正确的值（例如 file://policy.json）。

```
aws glue put-resource-policy --policy-in-json glue-resource-policy --enable-hybrid
 TRUE
```

有关更多信息，请参阅 put-resource-policy。

第 3 步：使用基于标签的访问控制方法执行跨账户共享

在本节，我们将引导您完成以下主要步骤：

1. 定义 LF 标签。
2. 将 LF 标签分配给目标资源。
3. 向使用者账户授予 LF 标签权限。
4. 向使用者账户授予数据权限。
5. 或者，撤销 IAMAllowedPrincipals 对数据库、表和列的权限。
6. 创建指向共享表的资源链接。

7. 创建 LF 标签并将其分配给目标数据库。
8. 向使用者账户授予 LF 标签数据权限。

定义 LF 标签

> **Note**
>
> 如果您已登录制作者账户，请先退出，然后再完成以下步骤。

1. 通过 https://console.aws.amazon.com/lakeformation/ 以数据湖管理员的身份登录制作者账户。使用您在创建 AWS CloudFormation 堆栈期间指定的制作者账户编号、IAM 用户名（默认为 `DatalakeAdminProducer`）和密码。
2. 在 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 的导航窗格中，在权限下的管理角色和任务下，选择 LF 标签。
3. 选择添加 LF 标签。

将 LF 标签分配给目标资源

将 LF 标签分配给目标资源并向其他账户授予数据权限

作为数据湖管理员，您可以向资源附加标签。如果您计划使用单独的角色，则可能需要向单独的角色授予描述并附加权限。

1. 在导航窗格中的数据目录下，选择数据库。
2. 选择目标数据库（`lakeformation_tutorial_cross_account_database_tbac`），然后在操作菜单上选择编辑 LF 标签。

   在本教程中，您可为数据库分配 LF 标签，但也可为表和列分配 LF 标签。
3. 选择分配新的 LF 标签。
4. 添加键 `Confidentiality` 和值 `public`。
5. 选择 Save（保存）。

向使用者账户授予 LF 标签权限

留在制作者账户中，向使用者账户授予访问 LF 标签的权限。

1. 在导航窗格中，在权限的管理角色和任务中的 LF 标签权限下，选择授权。
2. 对于主体，选择外部账户。
3. 输入目标 AWS 账户 ID。

   同一个组织内的 AWS 账户会自动显示。否则，您必须手动输入 AWS 账户 ID。撰写本文时，Lake Formation 基于标签的访问控制并不支持向组织或组织单位授予权限。
4. 对于 LF 标签，请选择与使用者账户（键 `Confidentiality` 和值 `public`）共享的 LF 标签的键和值。
5. 对于权限，在 LF 标签权限中选择描述。

   LF 标签权限是授予使用者账户的权限。可授予的权限是指使用者账户可以向其他主体授予的权限。
6. 选择授权。

   此时，使用者数据湖管理员应该能够在 Lake Formation 控制台中权限下管理角色和任务的“LF 标签”中找到通过使用者账户共享的策略标签。

向使用者账户授予数据权限

现在，我们将通过指定 LF 标签表达式并授予使用者账户访问与该表达式匹配的任何表或数据库的权限，为使用者账户提供数据访问权限。

1. 在导航窗格中权限下的数据湖权限中，选择授权。
2. 对于主体，选择外部账户，然后输入目标 AWS 账户 ID。
3. 对于 LF 标签或目录资源，选择与使用者账户（键 `Confidentiality`和值 `public`）共享的 LF 标签的键和值。
4. 对于权限，在通过 LF 标签匹配的资源（推荐）下，选择添加 LF 标签。
5. 选择与使用者账户（键 `Confidentiality` 和值 `public`）共享的标签的键和值。
6. 对于数据库权限，选择数据库权限下的描述以授予数据库级别的访问权限。
7. 使用者数据湖管理员应该能够通过 https://console.aws.amazon.com/lakeformation/ 访问 Lake Formation 控制台，在权限下管理角色和任务的 LF 标签中找到通过使用者账户共享的策略标签。
8. 在可授予的权限下选择描述，这样使用者账户就可以向其用户授予数据库级别的权限。
9. 对于表和列权限，选择表权限下的选择和描述。
10. 在可授予的权限下选择选择和描述。
11. 选择授权。

撤销 **IAMAllowedPrincipals** 对数据库、表和列的权限（可选）。

在本教程的开头，您更改了 Lake Formation 数据目录设置。如果您跳过了该部分，则需要执行这一步。如果您更改了 Lake Formation 数据目录设置，则可以跳过这一步。

在这一步，我们需要撤销 IAMAllowedPrincipals 对数据库或表的默认 Super 权限。有关详细信息，请参阅步骤 4：将数据存储切换到 Lake Formation 权限模型。

在撤销 IAMAllowedPrincipals 的权限之前，请确保您已通过 Lake Formation 向现有 IAM 主体授予必要的权限。这包括三个步骤：

1. 通过 Lake Formation GetDataAccess 操作（使用 IAM 策略）向目标 IAM 用户或角色添加 IAM 权限。
2. 向目标 IAM 用户或角色授予 Lake Formation 数据权限（更改、选择等）。
3. 然后，撤销 IAMAllowedPrincipals 的权限。否则，在撤销 IAMAllowedPrincipals 的权限后，现有 IAM 主体可能无法再访问目标数据库或数据目录。

   如果您要应用 Lake Formation 权限模型（而不是 IAM 策略模型）来管理用户使用 Lake Formation 权限模型在单个账户内或多个账户之间进行的访问，则需要撤销 IAMAllowedPrincipals 的 Super 权限。对于要保留传统 IAM 策略模型的其他表，您不必撤销 IAMAllowedPrincipals 对这些表的权限。

   此时，使用者账户数据湖管理员应该能够通过 https://console.aws.amazon.com/lakeformation/ 访问 Lake Formation 控制台，在数据目录的“数据库”中找到通过使用者账户共享的数据库和表。如果不能，请确认以下配置是否正确。

   1. 为目标数据库和表分配了正确的策略标签和值。
   2. 为使用者账户分配了正确的标签权限和数据权限。
   3. 撤销 IAMAllowedPrincipals 对数据库或表的默认 Super 权限。

创建指向共享表的资源链接。

当在账户之间共享资源时，共享的资源未置于使用者账户的数据目录中。为了使它们可用，并使用 Athena 等服务查询共享表的基础数据，我们需要创建一个指向共享表的资源链接。资源链接是一个数据目录对象，它是指向本地或共享数据库或表的链接。有关详细信息，请参阅 创建资源链接。通过创建资源链接，您可以：

- 为数据库或表分配一个与您的数据目录资源命名策略相符的不同名称。
- 使用 Athena 和 Redshift Spectrum 等服务查询共享的数据库或表。

要创建资源链接，请完成以下步骤：

1. 如果您已登录使用者账户，请退出。
2. 以使用者账户数据湖管理员身份登录。使用您在创建 AWS CloudFormation 堆栈期间指定的使用者账户 ID、IAM 用户名（默认为 DataLakeAdminConsumer）和密码。
3. 在 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 的导航窗格中，在数据目录下的“数据库”中，选择共享数据库 `lakeformation_tutorial_cross_account_database_tbac`。

   如果看不到该数据库，请回顾前面的步骤，看一看是否所有配置均正确。
4. 选择查看详细信息。
5. 选择共享表 `amazon_reviews_table_tbac`。
6. 在操作菜单上，选择创建资源链接。
7. 对于资源链接名称，输入名称（在本教程中，为 `amazon_reviews_table_tbac_resource_link`）。
8. 在数据库下，选择在其中创建资源链接的数据库（在本博文中，为 AWS CloudFormation 堆栈创建的数据库 `lakeformation_tutorial_cross_account_database_consumer`）。
9. 选择 Create（创建）。

   资源链接显示在数据目录的表中。

创建 LF 标签并将其分配给目标数据库

Lake Formation 标签与资源位于同一个数据目录中。这意味着在向使用者账户中的资源链接授予访问权限时，无法使用在制作者账户中创建的标签。您需要在使用者账户中创建一组单独的 LF 标签，以便在共享使用者账户中的资源链接时使用基于 LF 标签的访问控制。

1. 在使用者账户中定义 LF 标签。在本教程中，我们使用键 `Division` 和值 `sales`、`marketing` 和 `analyst`。
2. 将 LF 标签键 `Division` 和值 `analyst` 分配给在其中创建资源链接的数据库 `lakeformation_tutorial_cross_account_database_consumer`。

向使用者授予 LF 标签数据权限

最后一步，向使用者授予 LF 标签数据权限。

1. 在导航窗格中权限下的数据湖权限中，选择授权。
2. 对于主体，选择 IAM 用户和角色，然后选择用户 DataAnalyst。
3. 对于 LF 标签或目录资源，选择通过 LF 标签匹配的资源（推荐）。
4. 选择键 Division 和值 analyst。
5. 对于数据库权限，选择数据库权限下的描述。
6. 对于表和列权限，选择表权限下的选择和描述。
7. 选择授权。
8. 对用户 DataAnalyst 重复这些步骤，其中 LF 标签键为 Confidentiality，值为 public。

   此时，使用者账户中的数据分析师用户应该能够找到数据库和资源链接并通过 Athena 控制台 (https://console.aws.amazon.com/athena/) 查询共享表。如果不能，请确认以下配置是否正确。

   - 已为共享表创建资源链接
   - 您已授予该用户访问通过制作者账户共享的 LF 标签的权限
   - 您已授予该用户访问与资源链接和在其中创建资源链接的数据库关联的 LF 标签的权限
   - 检查是否为资源链接和在其中创建资源链接的数据库分配了正确的 LF 标签

第 4 步：实施命名资源方法

为了使用命名资源方法，我们将引导您完成以下主要步骤：

1. （可选）撤销 IAMAllowedPrincipals 对数据库、表和列的权限。
2. 向使用者账户授予数据权限。
3. 接受 AWS Resource Access Manager 资源共享。
4. 为共享表创建资源链接。
5. 向使用者授予对共享表的数据权限。
6. 向使用者授予对资源链接的数据权限。

撤销 **IAMAllowedPrincipals** 对数据库、表和列的权限（可选）

- 在本教程的开头，我们更改了 Lake Formation 数据目录设置。如果您跳过了该部分，则需要执行这一步。有关说明，请参阅上一节中的可选步骤。

向使用者账户授予数据权限

1.

> **Note**
>
> 如果您以其他用户身份登录了制作者账户，请先退出。

以制作者账户数据湖管理员身份使用创建 AWS CloudFormation 堆栈期间指定的 AWS 账户 ID、IAM 用户名（默认为 `DatalakeAdminProducer`）和密码登录 Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。

2. 在权限页面上的数据湖权限下，选择授权。
3. 在主体下，选择外部账户，然后输入一个或多个 AWS 账户 ID 或 AWS 组织 ID。有关更多信息，请参阅 AWS 组织。

   制作者账户所属的组织和同一组织内的 AWS 账户会自动显示。否则，请手动输入账户 ID 或组织 ID。
4. 对于 LF 标签或目录资源，选择`Named data catalog resources`。
5. 在数据库下，选择数据库 `lakeformation_tutorial_cross_account_database_named_resource`。
6. 选择添加 LF 标签。
7. 在表下，选择所有表。
8. 对于表列权限，在表权限下选择选择和描述。
9. 在可授予的权限下选择选择和描述。
10. 或者，对于数据权限，如果需要进行列级别权限管理，请选择基于列的简单访问权限。
11. 选择授权。

如果您尚未撤销 `IAMAllowedPrincipals` 的权限，则会遇到授予权限失败错误。此时，您应该可以在权限下的“数据权限”中看到目标表是通过 AWS RAM 与使用者账户共享的。

接受 AWS RAM 资源共享

> **Note**
>
> 仅对基于 AWS 账户的共享（而非基于组织的共享），需要执行此步骤。

1. 以使用者账户数据湖管理员身份使用创建 AWS CloudFormation 堆栈期间指定的 IAM 用户名（默认为 DatalakeAdminConsumer）和密码登录 AWS 控制台，网址为 https://console.aws.amazon.com/connect/。
2. 在 AWS RAM 控制台的导航窗格中，在已和我分享的“资源共享”下选择共享的 Lake Formation 资源。状态应为待处理。
3. 选择操作和授权。
4. 确认资源详细信息，然后选择接受资源共享。

   此时，使用者账户数据湖管理员应该能够通过 (https://console.aws.amazon.com/lakeformation/) 访问 Lake Formation 控制台，在数据目录下的数据库中找到共享资源。

为共享表创建资源链接

- 按照第 3 步：使用基于标签的访问控制方法执行跨账户共享（第 6 步）中的说明为共享表创建资源链接。命名资源链接 `amazon_reviews_table_named_resource_resource_link`。在数据库 `lakeformation_tutorial_cross_account_database_consumer` 中创建资源链接。

向使用者授予对共享表的数据权限

要向使用者授予对共享表的数据权限，请完成以下步骤：

1. 在 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 上权限下的数据湖权限中，选择授权。
2. 对于主体，选择 IAM 用户和角色，然后选择用户 `DataAnalyst`。
3. 对于 LF 标签或目录资源，选择命名数据目录资源。
4. 在数据库下，选择数据库 `lakeformation_tutorial_cross_account_database_named_resource`。如果在下拉列表中看不到该数据库，请选择加载更多。
5. 在表下，选择表 `amazon_reviews_table_named_resource`。
6. 对于表和列权限，选择表权限下的选择和描述。
7. 选择授权。

向使用者授予对资源链接的数据权限

除了向数据湖用户授予访问共享表的权限外，您还需要向数据湖用户授予访问资源链接的权限。

1. 在 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 上权限下的数据湖权限中，选择授权。
2. 对于主体，选择 IAM 用户和角色，然后选择用户 DataAnalyst。
3. 对于 LF 标签或目录资源，选择命名数据目录资源。
4. 在数据库下，选择数据库 lakeformation_tutorial_cross_account_database_consumer。如果在下拉列表中看不到该数据库，请选择加载更多。
5. 在表下，选择表 amazon_reviews_table_named_resource_resource_link。
6. 对于资源链接权限，在资源链接权限下选择描述。
7. 选择授权。

   此时，使用者账户中的数据分析师用户应该能够找到数据库和资源链接，并通过 Athena 控制台查询共享表。

   如果不能，请确认以下配置是否正确。

   - 已为共享表创建资源链接
   - 您已授予用户访问通过制作者账户共享的表的权限
   - 您已授予用户访问资源链接和为其创建资源链接的数据库的权限

第 5 步：清除 AWS 资源

为帮助避免您的 AWS 账户产生不必要的费用，您可以删除您在本教程中使用的 AWS 资源。

1. 通过制作者账户登录 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 并删除或更改以下内容：

   - AWS Resource Access Manager 资源共享
   - Lake Formation 标签
   - AWS CloudFormation 堆栈
   - Lake Formation 设置
   - AWS Glue Data Catalog
2. 通过使用者账户登录 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 并删除或更改以下内容：

   - Lake Formation 标签

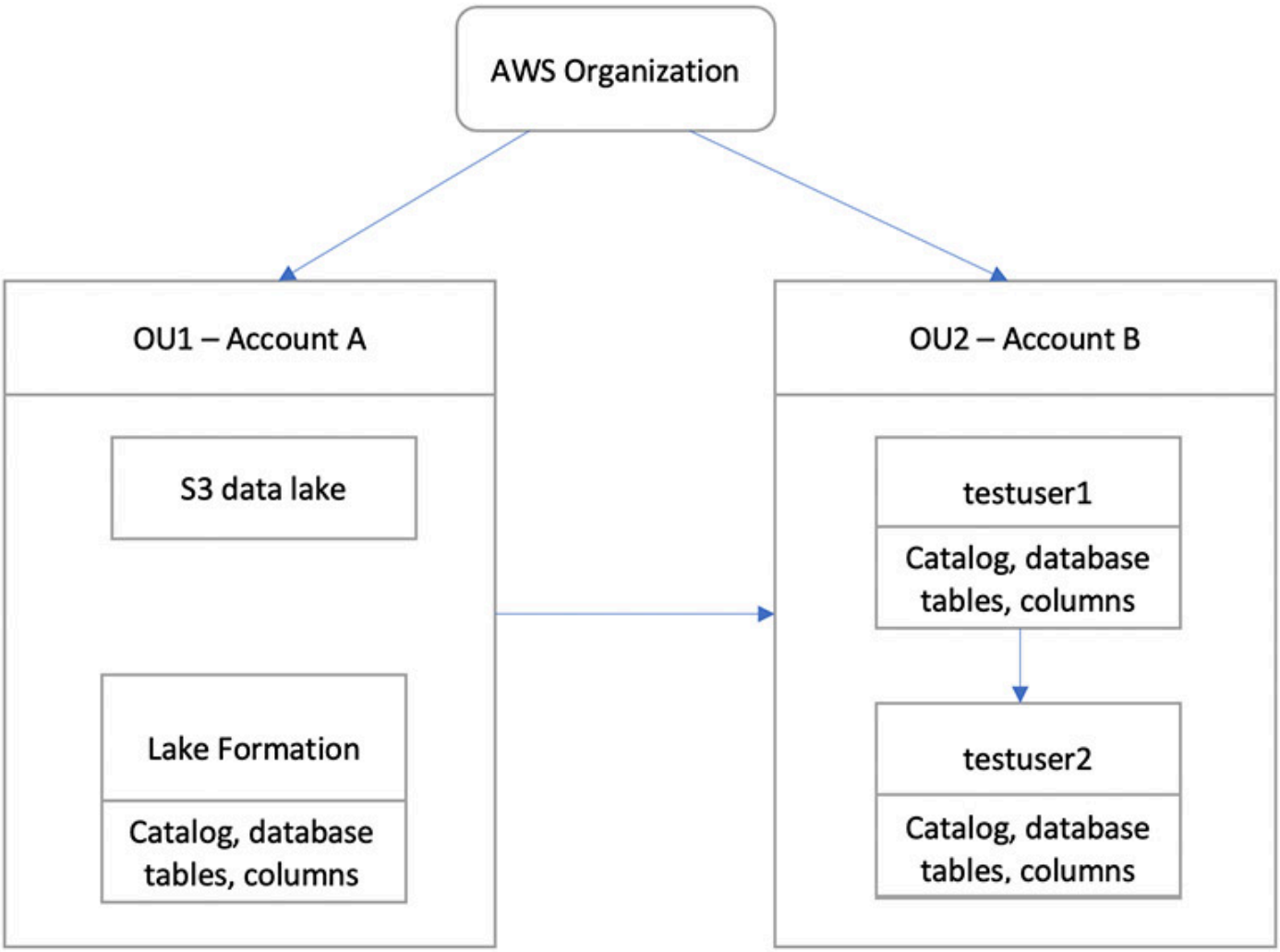
- AWS CloudFormation 堆栈

使用 Lake Formation 细粒度访问控制共享数据湖

本教程提供了有关如何在使用 AWS Organizations 管理多个 AWS 账户时使用 Lake Formation 快速轻松地共享数据集的分步说明。您可以定义精细权限来控制对敏感数据的访问。

以下过程还介绍了账户 A 的数据湖管理员如何为账户 B 提供细粒度访问权限，以及作为数据管家的账户 B 中用户如何向其账户中的其他用户授予对共享表的细粒度访问权限。每个账户中的数据管家可以独立地将访问权限委派给自己的用户，从而赋予每个团队或业务线 (LOB) 自主权。

该使用案例假定您在使用 AWS Organizations 管理自己的 AWS 账户。一个组织单位 (OU1) 中的账户 A 的用户向 OU2 中账户 B 的用户授予访问权限。当不使用组织时（例如，当您只有几个账户时），您可以使用同样的方法。下图说明了对数据湖中数据集的细粒度访问控制。账户 A 中提供了数据湖。账户 A 的数据湖管理员为账户 B 提供了细粒度访问权限。该图表还显示，账户 B 的某用户向账户 B 中的其他用户提供对账户 A 数据湖表的列级别访问权限。



主题

目标受众

本教程适用于数据管家、数据工程师和数据分析师。下表列出了本教程中使用的角色：

角色	描述
IAM 管理员	具有 AWS 托管策略 AdministratorAccess 的用户。
数据湖管理员	其角色附加了 AWS 托管策略 AWSLakeFormationDataAdmin 的用户。
数据分析人员	附加了 AWS 托管策略 AmazonAthenaFullAccess 的用户。

先决条件

在开始学习本教程之前，您必须拥有 AWS 账户，以便能够以具有正确权限的管理用户身份登录。有关更多信息，请参阅完成初始 AWS 配置任务。

本教程假定您熟悉 IAM。有关 IAM 的信息，请参阅 IAM 用户指南。

在此教程中，您需要以下资源：

- 两个组织单位：
  - OU1 — 包含账户 A
  - OU2 — 包含账户 B
- 账户 A 中的 Amazon S3 数据湖位置（存储桶）。
- 账户 A 中的数据湖管理员用户。您可以使用 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 或 Lake Formation API 的 PutDataLakeSettings 操作来创建数据湖管理员。
- 在账户 A 中配置的 Lake Formation 和在账户 A 中向 Lake Formation 注册的 Amazon S3 数据湖位置。
- 账户 B 中的两个用户使用以下 IAM 托管策略：
  - testuser1 — 附加了 AWS 托管策略 AWSLakeFormationDataAdmin。
  - testuser2 — 附加了 AWS 托管策略 AmazonAthenaFullAccess。
- 账户 B 的 Lake Formation 数据库中的数据库 testdb。

第 1 步：提供对其他账户的细粒度访问权限

了解账户 A 的数据湖管理员如何为账户 B 提供细粒度访问权限。

授予对其他账户的细粒度访问权限

1. 以数据湖管理员身份通过账户 A 登录 AWS Management Console (https://console.aws.amazon.com/connect/)。
2. 打开 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/)，然后选择开始使用。
3. 在导航窗格中，选择数据库。
4. 选择 Create database（创建数据库）。
5. 在数据库详细信息部分，选择数据库。
6. 对于名称，输入名称（在本教程中，我们使用 `sampledb01`）。
7. 确保未选中仅对此数据库中的新表使用 IAM 访问控制。如果不选择此选项，我们便可以从 Lake Formation 控制访问。
8. 选择 Create database（创建数据库）。
9. 在数据库页面上，选择您的数据库 `sampledb01`。
10. 在操作菜单上选择授权。
11. 在授予权限部分，选择外部账户。
12. 对于 AWS 账户 ID 或 AWS 组织 ID，请在 OU2 中输入账户 B 的账户 ID。
13. 对于表，选择您希望账户 B 有权访问的表（在本博文中，我们使用表 `acc_a_area`）。（可选）您可以授予对表中列的访问权限，我们在本文中便执行了此操作。
14. 对于包括列，选择您希望账户 B 有权访问的列（在博文中，我们授予对类型、名称和标识符的权限）。
15. 对于列，选择包括列。
16. 对于表权限，选择选择。
17. 对于可授予的权限，选择选择。账户 B 中的管理用户需要拥有可授予的权限，这样才能向账户 B 中的其他用户授予权限。
18. 选择授权。
19. 在导航窗格中，选择表。
20. 您可以在“具有访问权限的 AWS 账户 和 AWS 组织”部分中看到一个活动连接。

创建资源链接

Amazon Athena 等集成式服务无法跨账户直接访问数据库或表。因此，您需要创建一个资源链接，这样 Athena 才能访问您账户中指向其他账户中数据库和表的资源链接。创建指向表 (acc_a_area) 的资源链接，以便账户 B 用户可以使用 Athena 查询其数据。

1. 在账户 B 中以 testuser1 身份登录 AWS 控制台 (https://console.aws.amazon.com/connect/)。
2. 在 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/) 上的导航窗格中，选择表。您应该会看到账户 A 已提供对其的访问权限的表。
3. 选择 acc_a_area 表。
4. 在操作菜单上，选择创建资源链接。
5. 对于资源链接名称，输入名称（在本教程中，为 acc_a_area_rl）。
6. 对于数据库，选择您的数据库 (testdb)。
7. 选择 Create（创建）。
8. 在导航窗格中，选择表。
9. 选择 acc_b_area_rl 表。
10. 在操作菜单上，选择查看数据。

    您将被重定向到 Athena 控制台，在该控制台中，您应该可以看到数据库和表。

    现在，您可以对该表运行查询，以查看向账户 B 中 testuser1 提供了对其的访问权限的列值。

第 2 步：为同一账户中的用户提供细粒度访问权限

本节介绍账户 B 中的用户 (testuser1) 作为数据管家，如何向同一账户中的其他用户 (testuser2) 提供对共享表 aac_b_area_rl 中列名的细粒度访问权限。

向同一个账户中的用户授予细粒度访问权限

1. 在账户 B 中以 testuser1 身份登录 AWS 控制台 (https://console.aws.amazon.com/connect/)。
2. 在 Lake Formation 控制台的导航窗格中，选择表。

    您可以通过表的资源链接授予对表的权限。为此，在表页面上选择资源链接 acc_b_area_rl，然后在操作菜单上选择对目标的授权。
3. 在授予权限部分，选择我的账户。
4. 对于 IAM 用户和角色，选择用户 testuser2。

5. 对于列，选择列名。
6. 对于表权限，选择选择。
7. 选择授权。

   创建资源链接时，只有您可以查看和访问它。要允许您账户中的其他用户访问资源链接，您需要授予对资源链接本身的权限。您需要授予描述或删除权限。在表页面上，再次选择您的表，然后在操作菜单上选择授权。
8. 在授予权限部分，选择我的账户。
9. 对于 IAM 用户和角色，选择用户 `testuser2`。
10. 对于资源链接权限，选择描述。
11. 选择授权。
12. 以 `testuser2` 身份使用账户 B 登录 AWS 控制台。

    在 Athena 控制台 (https://console.aws.amazon.com/athena/) 上，您应该会看到数据库和表 `acc_b_area_rl`。现在，您可以对表运行查询以查看 `testuser2` 有权访问的列值。

加入 Lake Formation 权限

AWS Lake Formation 使用 AWS Glue Data Catalog 以数据库和表的形式存储 Amazon S3 数据的元数据。表存储有关基础数据的信息，包括架构信息、分区信息和数据位置。数据库是表的集合。数据目录还包含资源链接，这些链接是指向外部账户中共享数据库和表的链接，用于跨账户访问数据湖中的数据。每个 AWS 账户在每个 AWS 区域都有一个数据目录。

Lake Formation 提供了关系数据库管理系统 (RDBMS) 权限模型，用于授予或撤销对数据目录中的数据库、表和列以及 Amazon S3 中的基础数据的访问权限。

在了解 Lake Formation 权限模型的详细信息之前，查看以下背景信息会很有帮助：

- Lake Formation 管理的数据湖位于 Amazon Simple Storage Service (Amazon S3) 中的指定位置。
- Lake Formation 维护一个数据目录，其中包含有关要导入数据湖的源数据（例如日志和关系数据库中的数据）以及有关 Amazon S3 中数据湖中的数据的元数据。元数据以数据库和表的形式进行组织。元数据表包含架构、位置、分区以及有关它们所表示的数据的其他信息。元数据数据库是表的集合。
- Lake Formation 数据目录与 AWS Glue 使用的数据目录相同。您可以使用 AWS Glue 爬网程序创建数据目录表，也可以使用 AWS Glue 提取、转换、加载 (ETL) 作业来填充数据湖中的基础数据。
- 数据目录中的数据库和表称为“数据目录资源”。数据目录中的表称为“元数据表”，以区别于数据来源中的表或 Amazon S3 中的表格数据。元数据表在 Amazon S3 或数据来源中指向的数据称为“基础数据”。
- 主体是指用户或角色、Amazon QuickSight 用户或组、通过 SAML 提供商向 Lake Formation 进行身份验证的用户或组，或者用于跨账户访问控制的 AWS 账户 ID、组织 ID 或组织单位 ID。
- AWS Glue 爬网程序创建元数据表，但您也可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 手动创建元数据表。创建元数据表时，必须指定一个位置。创建数据库时，位置是可选的。表位置可以是 Amazon S3 位置或数据来源位置，例如 Amazon Relational Database Service (Amazon RDS) 数据库。数据库位置始终是 Amazon S3 位置。
- 与 Lake Formation 集成的服务（如 Amazon Athena 和 Amazon Redshift）可以访问数据目录以获取元数据并检查运行查询的授权。有关集成服务的完整列表，请参阅 AWS 服务与 Lake Formation 的集成。

主题

- 更改数据湖的默认设置
- 隐式 Lake Formation 权限
- Lake Formation 权限参考
- 集成 IAM Identity Center
- 向数据湖添加 Amazon S3 位置
- 混合访问模式
- 创建数据目录表和数据库
- 在 Lake Formation 中使用工作流导入数据

Lake Formation 权限概述

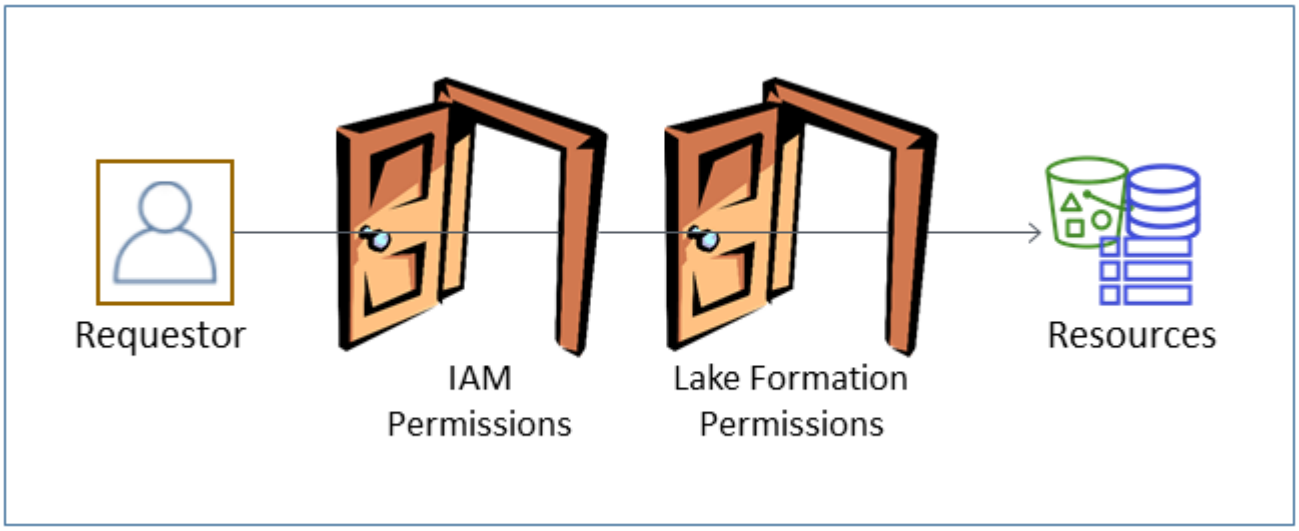
AWS Lake Formation 中有两种主要类型的权限：

- 元数据访问权限 - 对数据目录资源的权限（数据目录权限）。

  这些权限使主体能够创建、读取、更新和删除数据目录中的元数据数据库和表。
- 基础数据访问 - 对 Amazon Simple Storage Service (Amazon S3) 中位置的权限（数据访问权限和数据位置权限）。
  - 数据湖权限使主体能够在基础 Amazon S3 位置中读取和写入数据，即数据目录资源指向的数据。
  - 数据位置权限使主体能够创建和更改指向特定 Amazon S3 位置的元数据数据库和表。

对于这两个区域，Lake Formation 结合使用 Lake Formation 权限和 AWS Identity and Access Management (IAM) 权限。IAM 权限模型由 IAM 策略组成。Lake Formation 权限模型是作为 DBMS 样式的 GRANT/REVOKE 命令实现的，例如 `Grant SELECT on tableName to userName`。

当主体请求访问数据目录资源或基础数据时，请求必须通过 IAM 和 Lake Formation 的权限检查才能成功。



Lake Formation 权限控制对数据目录资源、Amazon S3 位置以及这些位置的基础数据的访问。IAM 权限控制对 Lake Formation 和 AWS Glue API 及资源的访问。因此，尽管您可能具有在数据目录 (CREATE_TABLE) 中创建元数据表的 Lake Formation 权限，但如果您不具有对 `glue:CreateTable` API 的 IAM 权限，您的操作将失败。（为什么需要 `glue:` 权限？因为 Lake Formation 使用 AWS Glue Data Catalog。）

> **Note**
>
> Lake Formation 权限仅适用于被授予这些权限的区域。

AWS Lake Formation 要求授权每个主体（用户或角色）对 Lake Formation 管理的资源执行操作。主体由数据湖管理员或其他有权授予 Lake Formation 权限的主体授予必要的授权。

当您向主体授予 Lake Formation 权限时，您可以选择授予将该权限传递给其他主体的能力。

您可以使用 Lake Formation API、AWS Command Line Interface (AWS CLI) 或 Lake Formation 控制台的数据权限和数据位置页面来授予和撤销 Lake Formation 权限。

精细访问控制的方法

使用数据湖时，目标是对数据进行精细访问控制。在 Lake Formation 中，这意味着对数据目录资源和 Amazon S3 位置进行精细访问控制。您可以使用以下方法之一实现精细访问控制。

方法	Lake Formation 权限	IAM 权限	注释
方法 1	Open（打开）	精细	这是默认方法，用于向后兼容 AWS Glue。 • “开放”是指将特殊权限 Super 授予组 IAMAllowedPrincipals ，其中 IAMAllowedPrincipals 会自动创建并包括 IAM 策略允许访问您的数据目录资源的任何 IAM 用户和角色，并且 Super 权限使主体能够对授予该权限的数据库或表执行所有受支持的 Lake Formation 操作。这实际上会导致对数据目录资源和 Amazon S3 位置的访问仅由 IAM 策略控制。有关更多信息，请参阅更改数据湖的默认设置和将 AWS Glue 数据权限升级为 AWS Lake Formation 模型。 • “精细”是指 IAM 策略控制对数据目录资源和单个 Amazon S3 存储桶的所有访问。 在 Lake Formation 控制台上，此方法显示为仅使用 IAM 访问控制。
方法 2	精细	粗粒度	这是推荐的方法。 • “精细”访问是指向各个主体授予对数据目录资源、Amazon S3 位置以及这些位置中的基础数据的有限 Lake Formation 权限。 • “粗粒度”是指对各项操作和 Amazon S3 位置的访问具有更广泛的权限。例如，粗粒度 IAM 策略可能包括 "glue:*" 或 "glue:Create*" ，而不是 "glue:CreateTables" ，从而保留 Lake Formation 权限来控制主体是否可

方法	Lake Formation 权限	IAM 权限	注释
			以创建目录对象。这也意味着向主体授予他们完成工作所需的 API 的访问权限，但会锁定其他 API 和资源。例如，您可以创建一个 IAM 策略，使主体能够创建数据目录资源以及创建和运行工作流，但不允许创建 AWS Glue 连接或用户定义的函数。请参阅本部分后面的相关示例。

Important

请注意以下事项：

- 默认情况下，Lake Formation 启用了仅使用 IAM 访问控制设置，以便与现有 AWS Glue Data Catalog 行为兼容。我们建议您在过渡到使用 Lake Formation 权限后禁用这些设置。有关更多信息，请参阅更改数据湖的默认设置。
- 数据湖管理员和数据库创建者具有您必须了解的隐式 Lake Formation 权限。有关更多信息，请参阅隐式 Lake Formation 权限。

元数据访问控制

对于数据目录资源的访问控制，以下讨论假定使用 Lake Formation 权限进行精细访问控制，并使用 IAM 策略进行粗粒度访问控制。

有两种不同的方法可以授予对数据目录资源的 Lake Formation 权限：

- 命名资源访问控制 - 使用此方法，您可以通过指定数据库名称或表名称来授予对特定数据库或表的权限 授予形式如下：

  [使用授予选项] 向主体授予对 的权限。

  使用授予选项，您可以允许被授权者向其他主体授予权限。
- 基于标签的访问控制 - 使用此方法，您可以为数据目录数据库、表和列分配一个或多个 LF 标签，并向主体授予对一个或多个 LF 标签的权限。每个 LF 标签都是一个键值对，例如 `department=sales`。具有与数据目录资源上的 LF 标签匹配的主体可以访问该资源。对于包含大

量数据库和表的数据湖，建议使用此方法。Lake Formation 基于标签的访问控制中对此进行了详细说明。

主体对资源具有的权限是这两种方法授予的权限的联合。

下表汇总了对数据目录资源可用的 Lake Formation 权限。列标题表示被授予权限的资源。

目录	数据库	表
CREATE_DATABASE	CREATE_TABLE	ALTER
	ALTER	DROP
	DROP	DESCRIBE
	DESCRIBE	SELECT*
		INSERT*
		DELETE*

例如，CREATE_TABLE 权限是针对数据库授予的。例如，这意味着允许主体在该数据库中创建表。

带星号 (*) 的权限是针对数据目录资源授予的，但它们适用于基础数据。例如，通过对元数据表的 DROP 权限，您可以从数据目录中删除该表。但是，通过对同一表授予的 DELETE权限，您可以使用 SQL DELETE 等语句删除 Amazon S3 中表的基础数据。通过这些权限，您还可以在 Lake Formation 控制台上查看表，并使用 AWS Glue API 检索有关表的信息。因此，SELECT、INSERT 和 DELETE 既是数据目录权限，也是数据访问权限。

对表授予 SELECT 权限时，您可以添加包含或排除一列或多列的筛选条件。这允许对元数据表列进行精细访问控制，从而限制集成服务的用户在运行查询时可以看到的列。仅使用 IAM 策略时无法使用此功能。

还有一个名为 Super 的特殊权限。Super 权限使主体能够对被授予该权限的数据库或表执行所有支持的 Lake Formation 操作。此权限可以与其他 Lake Formation 权限共存。例如，您可以授予对元数据表的 Super、SELECT、和 INSERT 权限。主体可以对表执行所有支持的操作，当您撤销 Super 时，SELECT 和 INSERT 权限将保留。

有关每项权限的详细信息，请参阅 Lake Formation 权限参考。

> **Important**
>
> 为了能够查看其他用户创建的数据目录表，您必须至少被授予对该表的一项 Lake Formation 权限。如果您被授予对该表的至少一项权限，则还可以查看该表中包含的数据库。

您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予或撤销数据目录权限。以下是 AWS CLI 命令的示例，该命令向用户授予在 `retail` 数据库中创建表的 `datalake_user1` 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

以下是粗粒度访问控制 IAM 策略的示例，该策略使用 Lake Formation 权限对精细访问控制进行了补充。它允许对任何元数据数据库或表执行所有操作。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:*Database*",
                "glue:*Table*",
                "glue:*Partition*"
            ],
            "Resource": "*"
        }
    ]
}
```

下一个示例也是粗粒度访问控制，但在某种程度上更具限制性。它允许对指定账户和区域中数据目录中的所有元数据数据库和表进行只读操作。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
```

```
            "Action": [
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetDatabase",
                "glue:GetDatabases"
            ],
            "Resource": "arn:aws:glue:us-east-1:111122223333:*"
        }
    ]
}
```

将这些策略与以下策略进行比较，后者实现了基于 IAM 的精细访问控制。它仅授予对指定账户和区域中客户关系管理 (CRM) 元数据数据库中的部分表的权限。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetDatabase",
                "glue:GetDatabases"
            ],
            "Resource": [
                "arn:aws:glue:us-east-1:111122223333:catalog",
                "arn:aws:glue:us-east-1:111122223333:database/CRM",
                "arn:aws:glue:us-east-1:111122223333:table/CRM/P*"
            ]
        }
    ]
}
```

有关粗粒度访问控制策略的更多示例，请参阅 Lake Formation 角色和 IAM 权限参考。

基础数据访问控制

当集成 AWS 服务请求访问由 AWS Lake Formation 访问控制的 Amazon S3 位置中的数据时，Lake Formation 会提供临时凭证来访问数据。

要使 Lake Formation 能够控制对 Amazon S3 位置基础数据的访问，请向 Lake Formation 注册该位置。

注册 Amazon S3 位置后，您即可开始授予以下 Lake Formation 权限：

- 对指向该位置的数据目录表的数据访问权限（SELECT、INSERT 和 DELETE））。
- 对该位置的数据位置权限。

Lake Formation 数据位置权限可控制创建指向特定 Amazon S3 位置的数据目录资源的能力。数据位置权限为数据湖中的位置提供了一层额外的安全保护。当您向主体授予 CREATE_TABLE 或 ALTER 权限时，您还会授予数据位置权限，以限制主体可以为其创建或更改元数据表的位置。

Amazon S3 位置是存储桶下的存储桶或前缀，但不是单个 Amazon S3 对象。

您可以使用 Lake Formation 控制台、API 或 AWS CLI 来向主体授予数据位置权限。授予的一般形式如下：

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

如果包含 `with grant option`，则被授权者可以向其他主体授予权限。

回想一下，Lake Formation 权限始终与 AWS Identity and Access Management (IAM) 权限结合使用，以实现精细访问控制。对于基础 Amazon S3 数据的读/写权限，将按如下方式授予 IAM 权限：

注册位置时，您需要指定一个 IAM 角色来授予对该位置的读/写权限。Lake Formation 在向集成 AWS 服务提供临时凭证时会代入该角色。典型角色可能附加了以下策略，其中注册位置为存储桶 awsexamplebucket。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket/*"
```

```
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket"
            ]
        }
    ]
}
```

Lake Formation 提供了一个服务相关角色，您可以在注册期间使用该角色自动创建此类策略。有关更多信息，请参阅[在 Lake Formation 中使用服务相关角色]。

因此，注册 Amazon S3 位置会授予对该位置所需的 IAM `s3:` 权限，其中这些权限由用于注册该位置的角色指定。

> **Important**
>
> 避免注册启用了请求者付费的 Amazon S3 存储桶。对于在 Lake Formation 中注册的存储桶，用于注册存储桶的角色始终被视为请求者。如果存储桶由其他 AWS 账户访问，则当该角色与存储桶拥有者属于同一个账户时，存储桶拥有者需要支付数据访问费用。

要获得基础数据的读/写访问权限，除了 Lake Formation 权限外，主体还需要以下 IAM 权限：

`lakeformation:GetDataAccess`

获得此权限后，Lake Formation 将授权访问数据的临时凭证请求。

> **Note**
>
> Amazon Athena 要求用户具有 `lakeformation:GetDataAccess` 权限。其他集成服务要求其基础执行角色具有 `lakeformation:GetDataAccess` 权限。

此权限包含在 [Lake Formation 角色和 IAM 权限参考]中的建议策略中。

总而言之，要让 Lake Formation 主体能够通过由 Lake Formation 权限控制的访问权限读取和写入基础数据，请执行以下操作：

- 向 Lake Formation 注册包含此类数据的 Amazon S3 位置。
- 创建指向基础数据位置的数据目录表的主体必须具有数据位置权限。
- 读取和写入基础数据的主体必须对指向基础数据位置的数据目录表具有 Lake Formation 数据访问权限。
- 在向 Lake Formation 注册基础数据位置时，读取和写入基础数据的主体必须具有 `lakeformation:GetDataAccess` IAM 权限。

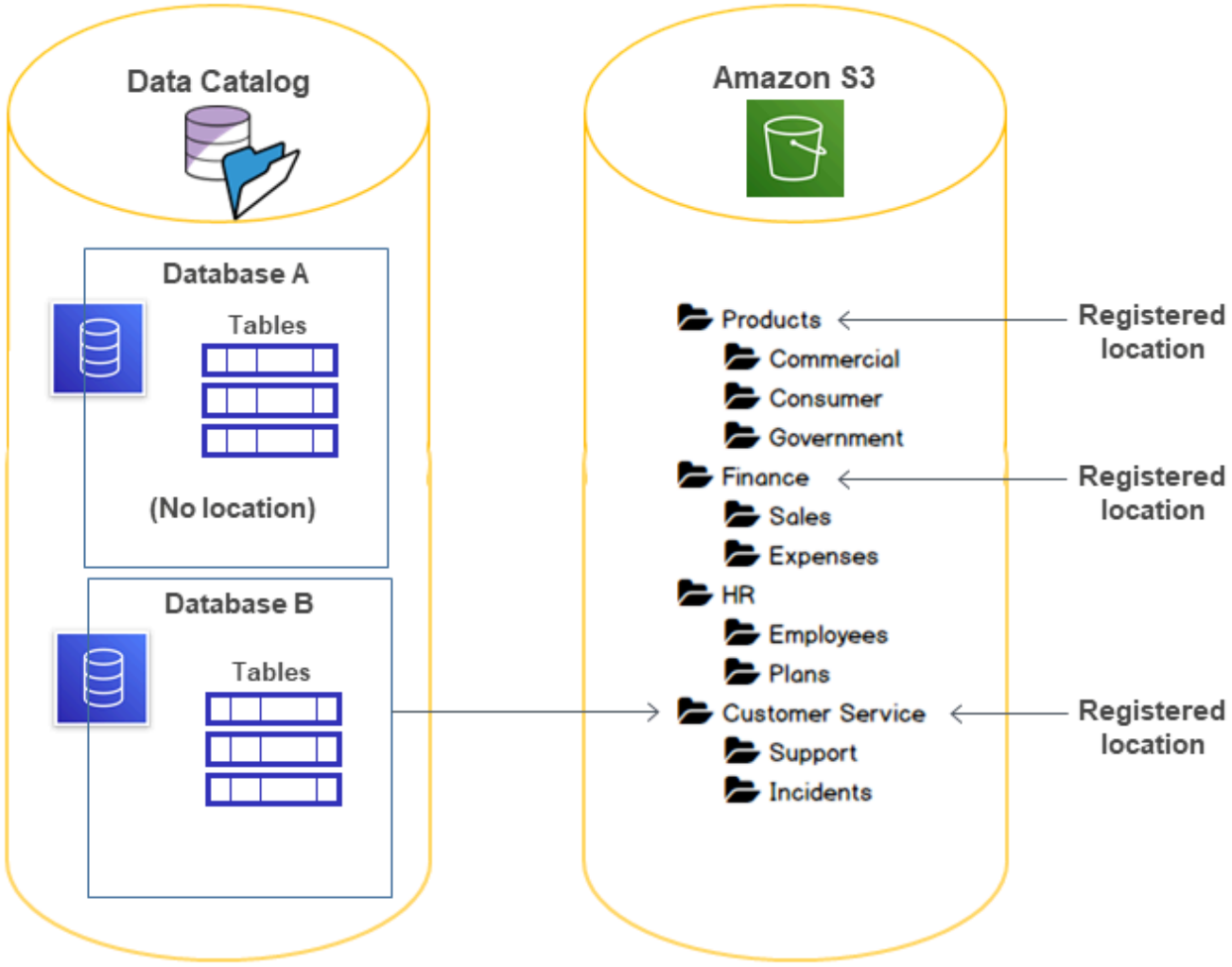
> **Note**
>
> 如果您可以通过 IAM 或 Amazon S3 策略访问 Amazon S3 位置，则 Lake Formation 权限模型不会阻止通过 Amazon S3 API 或控制台访问 Amazon S3 位置。您可以将 IAM 策略附加到主体以阻止此访问。

有关数据位置权限的更多信息

数据位置权限控制对数据目录数据库和表的创建和更新操作的结果。规则如下：

- 主体必须对 Amazon S3 位置具有显式或隐式数据位置权限，才能创建或更新指定该位置的数据库或表。
- 显式权限 `DATA_LOCATION_ACCESS` 使用控制台、API 或 AWS CLI 进行授予。
- 当数据库具有指向注册位置的位置属性、主体具有对数据库的 `CREATE_TABLE` 权限以及主体尝试在该位置或子位置创建表时，将授予隐式权限。
- 如果某个主体被授予对某个位置的数据位置权限，则该主体对所有子位置都具有数据位置权限。
- 主体不需要数据位置权限即可对基础数据执行读/写操作。具有 `SELECT` 或 `INSERT` 数据访问权限就足够了。数据位置权限仅适用于创建指向该位置的数据目录资源。

请考虑下图中所示的场景。



在此示意图中：

- Amazon S3 存储桶 Products、Finance 和 Customer Service 已向 Lake Formation 注册。
- Database A 不具有位置属性，而 Database B 具有指定 Customer Service 存储桶的位置属性。
- 用户 datalake_user 具有对这两个数据库的 CREATE_TABLE 权限。
- 用户 datalake_user 仅被授予对 Products 存储桶的数据位置权限。

以下是用户 datalake_user 尝试在特定位置的特定数据库中创建目录表时的结果。

datalake_user 在其中尝试创建表的位置

数据库和位置	成功或失败	Reason
位于 Finance/Sales 的数据库 A	失败	无数据位置权限
位于 Products 的数据库 A	成功	具有数据位置权限
位于 HR/Plans 的数据库 A	成功	位置未进行注册
位于 Customer Service/Incidents 的数据库 B	成功	数据库在 Customer Service 处具有位置属性。

有关更多信息，请参阅下列内容：

- 向数据湖添加 Amazon S3 位置
- Lake Formation 权限参考
- Lake Formation 角色和 IAM 权限参考

Lake Formation 角色和 IAM 权限参考

本部分列出了一些建议的 Lake Formation 角色及其建议的 AWS Identity and Access Management (IAM) 权限。有关 Lake Formation 权限的信息，请参阅 the section called “Lake Formation 权限参考”。

AWS Lake Formation 人物角色

下表列出了建议 AWS Lake Formation 的角色。

Lake Formation 角色

角色	描述
IAM 管理员（超级用户）	（必填）可以创建 IAM 用户和角色的用户。有AdministratorAccess AWS 托管策略。具有对所有 Lake Formation 资源的所有权限。可以添加数据湖管理员。如果未同时指定数据湖管理员，则无法授予 Lake Formation 权限。

角色	描述
数据湖管理员	（必填）可以注册 Amazon S3 地点、访问数据目录、创建数据库、创建和运行工作流程、向其他用户授予 Lake Formation 权限以及查看 AWS CloudTrail 日志的用户。与 IAM 管理员相比，具有的 IAM 权限较少，但足以管理数据湖。无法添加其他数据湖管理员。
只读管理员	（可选）可以查看主体、数据目录资源、权限和 AWS CloudTrail 日志但无权进行更新的用户。
数据工程师	（可选）可以创建数据库、创建和运行爬网程序和工作流，以及授予对爬网程序和工作流创建的数据目录表的 Lake Formation 权限的用户。我们建议您将所有数据工程师设置为数据库创建者。有关更多信息，请参阅创建数据库。
数据分析人员	（可选）可以使用 Amazon Athena等对数据湖运行查询的用户。只有足够的权限来运行查询。
工作流角色	（必需）代表用户运行工作流的角色。您可以在从蓝图创建工作流时指定此角色。

AWS Lake Formation 的托管策略

您可以使用 AWS 托管策略和内联策略授予使用所需 AWS Lake Formation 的 AWS Identity and Access Management (IAM) 权限。以下 AWS 托管策略适用于 Lake Formation。

AWS 托管策略：AWSLakeFormationDataAdmin

AWSLakeFormationDataAdmin策略授予对管理数据湖 AWS Lake Formation 等 AWS Glue 相关服务的管理权限。

您可以将 `AWSLakeFormationDataAdmin` 附加到您的用户、组和角色。

权限详细信息

- `CloudTrail`— 允许委托人查看 AWS CloudTrail 日志。这是检查数据湖设置中的任何错误所必需的权限。

- Glue - 允许主体查看、创建和更新数据目录中的元数据表和数据库。这包括以 Get、List、Create、Update、Delete 和 Search 开头的 API 操作。这是管理数据湖表的元数据所必需的权限。
- IAM - 允许主体检索有关 IAM 用户、角色和附加到角色的策略的信息。这是数据管理员查看和列出 IAM 用户和角色以授予 Lake Formation 权限所必需的。
- Lake Formation - 向数据湖管理员授予管理数据湖所需的 Lake Formation 权限。
- S3 - 允许主体检索有关 Amazon S3 存储桶及其位置的信息，以便为数据湖设置数据位置。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:*",
                "cloudtrail:DescribeTrails",
                "cloudtrail:LookupEvents",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTableVersions",
                "glue:GetPartitions",
                "glue:GetTables",
                "glue:GetWorkflow",
                "glue:ListWorkflows",
                "glue:BatchGetWorkflows",
                "glue:DeleteWorkflow",
                "glue:GetWorkflowRuns",
                "glue:StartWorkflowRun",
                "glue:GetWorkflow",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
```

```
                "s3:GetBucketAcl",
                "iam:ListUsers",
                "iam:ListRoles",
                "iam:GetRole",
                "iam:GetRolePolicy"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Deny",
            "Action": [
                "lakeformation:PutDataLakeSettings"
            ],
            "Resource": "*"
        }
    ]
}
```

> **Note**
>
> AWSLakeFormationDataAdmin 策略不会向数据湖管理员授予所有必需的权限。需要额外的权限才能创建和运行工作流并向服务相关角色 AWSServiceRoleForLakeFormationDataAccess 注册位置。有关更多信息，请参阅 创建数据湖管理员 和 在 Lake Formation 中使用服务相关角色。

AWS 托管策略：AWSLakeFormationCrossAccountManager

AWSLakeFormationCrossAccountManager策略提供通过 Lake Formation 对 AWS Glue 资源的跨账户访问权限，并授予对其他必需服务（例如 AWS Organizations 和）的读取权限 AWS RAM。

您可以将 AWSLakeFormationCrossAccountManager 附加到您的用户、组和角色。

权限详细信息

该策略包含以下权限。

- Glue - 允许主体设置或删除用于访问控制的数据目录资源策略。
- Organizations - 允许主体检索组织的账户和组织单位 (OU) 信息。
- ram:CreateResourceShare - 允许主体创建资源共享。

- ram:UpdateResourceShare - 允许主体修改指定资源共享的某些属性。
- ram:DeleteResourceShare - 允许主体删除指定的资源共享。
- ram:AssociateResourceShare - 允许主体将指定的主体列表和资源列表添加到资源共享。
- ram:DisassociateResourceShare - 允许主体将指定的主体或资源从参与指定资源共享中移除。
- ram:GetResourceShares – 允许主体检索有关您拥有的或与您共享的资源共享的详细信息。
- ram:RequestedResourceType - 允许主体检索资源类型（数据库、表或目录）。
- AssociateResourceSharePermission— 允许委托人添加或替换资源共享中包含的资源类型的 AWS RAM 权限。您只能将一个权限与资源共享中包含的每种资源类型相关联。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ram:CreateResourceShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringLikeIfExists": {
                    "ram:RequestedResourceType": [
                        "glue:Table",
                        "glue:Database",
                        "glue:Catalog"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "ram:UpdateResourceShare",
                "ram:DeleteResourceShare",
                "ram:AssociateResourceShare",
                "ram:DisassociateResourceShare",
                "ram:GetResourceShares"
            ],
            "Resource": "*",
```

```
            "Condition": {
                "StringLike": {
                    "ram:ResourceShareName": [
                        "LakeFormation*"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "ram:AssociateResourceSharePermission"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "ram:PermissionArn": [
                        "arn:aws:ram::aws:permission/AWSRAMLFEnabled*"
                    ]
                }
            }
       },
        {
            "Effect": "Allow",
            "Action": [
                "glue:PutResourcePolicy",
                "glue:DeleteResourcePolicy",
                "organizations:DescribeOrganization",
                "organizations:DescribeAccount",
                "ram:Get*",
                "ram:List*"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "organizations:ListRoots",
                "organizations:ListAccountsForParent",
                "organizations:ListOrganizationalUnitsForParent"
            ],
            "Resource": "*"
        }
    ]
```

```
}
```

AWS 托管策略：AWSGlueConsoleFullAccess

AWSGlueConsoleFullAccess当策略所关联的身份使用时，策略会授予对 AWS Glue 资源的完全访问权限 AWS Management Console。如果遵循此策略中指定的资源的命名约定，则用户具有完全控制台功能。此策略通常附加到 AWS Glue 控制台的用户。

此外，AWS GlueLake Formation 还担任服务角色，AWSGlueServiceRole允许访问相关服务，包括亚马逊弹性计算云 (Amazon EC2)、亚马逊简单存储服务 (Amazon S3) Simple Storage S3 和亚马逊。CloudWatch

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

此策略附加到名为的服务相关角色ServiceRoleForLakeFormationDataAccess，该角色允许服务根据您的请求对资源执行操作。您不能将此策略附加到您的 IAM 身份。

该政策允许 Lake Formation 集成 AWS 服务（例如 Amazon Athena 或 Amazon Redshift）使用服务相关角色来发现 Amazon S3 资源。

有关更多信息，请参阅在 Lake Formation 中使用服务相关角色。

权限详细信息

此政策包括以下权限。

- s3:ListAllMyBuckets— 返回请求的经过身份验证的发送者拥有的所有存储桶的列表。

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "LakeFormationDataAccessServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": [
    "arn:aws:s3:::*"
   ]
  }
```

```
 ]
}
```

Lake Formation 更新 AWS 了托管政策

查看自该服务开始追踪 Lake Formation AWS 托管政策变更以来这些更新的详细信息。

更改	描述	日期
Lake Formation 更新了 `LakeFormationDataAccessServiceRolePolicy` 策略。	Lake Formation在LakeFormationDataAccessServiceRolePolicy政策声明中添加了Sid元素，从而增强了该政策。	2024年2月
Lake Formation 更新了 `AWSLakeFormationCrossAccountManager` 策略。	Lake Formation 增加了在混合访问模式下启用跨账户数据共享的新权限，从而增强了该AWSLakeFormationCrossAccountManager政策。	2023 年 10 月
Lake Formation 更新了 `AWSLakeFormationCrossAccountManager` 策略。	Lake Formation 增强了AWSLakeFormationCrossAccountManager政策，在首次共享资源时，每个接收者账户仅创建一个资源共享。此后与同一账户共享的所有资源都将附加到同一资源共享。	2022 年 5 月 6 日
Lake Formation 开启了跟踪更改。	Lake Formation开始跟踪其 AWS 托管政策的变更。	2022 年 5 月 6 日

角色建议的权限

以下是针对每个角色建议的权限。IAM 管理员不包括在内，因为该类用户具有对所有资源的所有权限。

主题

- 数据湖管理员权限
- 只读管理员权限

- 数据工程师权限
- 数据分析师权限
- 工作流角色权限

数据湖管理员权限

> **Important**
>
> 在以下策略中，<account-id>替换为有效的 AWS 账号，然后<workflow_role>替换为有权运行工作流程的角色的名称，如中所定义工作流角色权限。

策略类型	Policy
AWS 托管策略	• AWSLakeFormationDataAdmin • LakeFormationDataAccessServiceRolePolicy （服务相关角色策略） • AWSGlueConsoleFullAccess （可选） • CloudWatchLogsReadOnlyAccess （可选） • AWSLakeFormationCrossAccountManager （可选） • AmazonAthenaFullAccess （可选） 有关可选 AWS 托管策略的信息，请参阅the section called “创建数据湖管理员”。

内联策略（用于创建 Lake Formation 服务相关角色）	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Effect": "Allow",
	            "Action": "iam:CreateServiceLinkedRole",
	            "Resource": "*",
	            "Condition": {
	                "StringEquals": {</pre>

策略类型	Policy

	"iam:AWSServiceName": "lakeformation.amazonaws.com"
	}
	}
	},
	{
	"Effect": "Allow",
	"Action": [
	"iam:PutRolePolicy"
	],
	"Resource": "arn:aws:iam:: *<account-id>* :role/aws-service-role/lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess"
	}
	]
	}

（可选）内联策略（工作流角色的 passrole 策略）。仅当数据湖管理员创建并运行工作流时，才需要此策略。	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow_role>* "
	]
	}
	]
	}

策略类型	Policy

（可选）内联策略（如果您的账户要授予或接收跨账户 Lake Formation 权限）。此政策用于接受或拒绝 AWS RAM 资源共享邀请，以及允许向组织授予跨账户权限。 `ram:EnableSharingWithAwsOrganization` 只有 AWS Organizations 管理账户中的数据湖管理员才需要填写。	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"ram:AcceptResourceShareInvitation",
	"ram:RejectResourceShareInvitation",
	"ec2:DescribeAvailabilityZones",
	"ram:EnableSharingWithAwsOrganization"
	],
	"Resource": "*"
	}
	]
	}

只读管理员权限

策略类型	Policy
内联策略（基本）	{ "Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Action":["lakeformation:GetEffectivePermissionsForPath", "lakeformation:ListPermissions", "lakeformation:ListDataCellsFilter", "lakeformation:GetDataCellsFilter", "lakeformation:SearchDatabasesByLFTags", "lakeformation:SearchTablesByLFTags", "lakeformation:GetLFTag",

策略类型	Policy

```
            "lakeformation:ListLFTags",
            "lakeformation:GetResourceLFTags",
            "lakeformation:ListLakeFormationOpti
n",
            "cloudtrail:DescribeTrails",
            "cloudtrail:LookupEvents",
            "glue:GetDatabase",
            "glue:GetDatabases",
            "glue:GetConnections",
            "glue:SearchTables",
            "glue:GetTable",
            "glue:GetTableVersions",
            "glue:GetPartitions",
            "glue:GetTables",
            "glue:GetWorkflow",
            "glue:ListWorkflows",
            "glue:BatchGetWorkflows",
            "glue:GetWorkflowRuns",
            "glue:GetWorkflow",
            "s3:ListBucket",
            "s3:GetBucketLocation",
            "s3:ListAllMyBuckets",
            "s3:GetBucketAcl",
            "iam:ListUsers",
            "iam:ListRoles",
            "iam:GetRole",
            "iam:GetRolePolicy"
         ],
         "Resource":"*"
      },
      {
         "Effect":"Deny",
         "Action":[
            "lakeformation:PutDataLakeSettings"
         ],
         "Resource":"*"
      }
   ]
}
```

数据工程师权限

> **Important**
>
> 在以下策略中，<account-id>替换为有效的 AWS 账号，然后<workflow_role>替换为工作流程角色的名称。

策略类型	Policy
AWS 托管策略	`AWSGlueConsoleFullAccess`
内联策略（基本）	

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "lakeformation:GrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:BatchRevokePermissions",
                "lakeformation:ListPermissions",
                "lakeformation:AddLFTagsToResource",
                "lakeformation:RemoveLFTagsFromResou
rce",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTa
gs",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults",
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetQueryStatistics"
            ],
            "Resource": "*"
        }
```

策略类型	Policy
	] }
内联策略（用于对受管控表的操作，包括事务中的操作）	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction", "lakeformation:DescribeTransaction", "lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel"], "Resource": "*" }] }

策略类型	Policy

内联策略（用于使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法进行元数据访问控制）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:AddLFTagsToResource",
	"lakeformation:RemoveLFTagsFromResou
	rce",
	"lakeformation:GetResourceLFTags",
	"lakeformation:ListLFTags",
	"lakeformation:GetLFTag",
	"lakeformation:SearchTablesByLFTags",
	"lakeformation:SearchDatabasesByLFTags"
	],
	"Resource": "*"
	}
	]
	}

内联策略（工作流角色的 passrole 策略）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow
	_role>* "
	]
	}
	]
	}

数据分析师权限

策略类型	Policy
AWS 托管策略	AmazonAthenaFullAccess

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

（可选）内联策略（用于对受管控表的操作，包括事务中的操作）	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction", "lakeformation:DescribeTransaction",

策略类型	Policy
	"lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel"], "Resource": "*" }] }

工作流角色权限

该角色具有运行工作流所需的权限。创建工作流时，您可以指定具有这些权限的角色。

> **Important**
>
> 在以下策略中，<region>使用有效的 AWS 区域标识符（例如us-east-1）、<account-id>有效的 AWS 账号、<workflow_role>工作流程角色的名称以及 *AWS CloudTrail* <your-s3-cloudtrail-bucket>日志的 Amazon S3 路径替换。

策略类型	Policy
AWS 托管策略	AWSGlueServiceRole

内联策略（数据访问）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "Lakeformation",
	"Effect": "Allow",
	"Action": [
	"lakeformation:GetDataAccess",
	"lakeformation:GrantPermissions"
	],
	"Resource": "*"
	}
	]

策略类型	Policy
	} }

内联策略（工作流角色的 passrole 策略）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow_role>* "
	]
	}
	]
	}

内联策略（用于在数据湖之外提取数据，AWS CloudTrail 例如日志）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": ["s3:GetObject", "s3:ListBucket"],
	"Resource": ["arn:aws:s3::: *<your-s3-cloudtrail-bucket>* /*"]
	}
	]
	}

更改数据湖的默认设置

要保持与 AWS Glue 的向后兼容性，AWS Lake Formation 应具有以下初始安全设置：

- 向组 `IAMAllowedPrincipals` 授予对所有现有 AWS Glue 数据目录资源的 `Super` 权限。
- 为新的数据目录资源启用“仅使用 IAM 访问控制”设置。

这些设置实际上使对数据目录资源和 Amazon S3 位置的访问只能由 AWS Identity and Access Management (IAM) 策略控制。单个 Lake Formation 权限无效。

`IAMAllowedPrincipals` 组包括 IAM 策略允许访问数据目录资源的任何 IAM 用户和角色。`Super` 权限使主体能够对被授予该权限的数据库或表执行所有支持的 Lake Formation 操作。

要更改安全设置，以便对数据目录资源（数据库和表）的访问由 Lake Formation 权限管理，请执行以下操作：

1. 更改新资源的默认安全设置。有关说明，请参阅更改默认权限模式或使用混合访问模式。
2. 更改现有数据目录资源的设置。有关说明，请参阅将 AWS Glue 数据权限升级为 AWS Lake Formation 模型。

使用 Lake Formation **`PutDataLakeSettings`** API 操作更改默认安全设置

您也可以使用 Lake Formation PutDataLakeSettings API 操作来更改默认安全设置。此操作采用可选目录 ID 和 DataLakeSettings 结构作为参数。

要通过 Lake Formation 对新数据库和表实施元数据和基础数据访问控制，请按如下方式对 `DataLakeSettings` 结构进行编码。

> **Note**
>
> 将 *<AccountID>* 替换为有效的 AWS 账户 ID，并将 *<Username>* 替换为有效的 IAM 用户名。您可以将多个用户指定为数据湖管理员。

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [],
        "CreateTableDefaultPermissions": []
    }
}
```

您也可以按如下方式对该结构进行编码。省略 CreateDatabaseDefaultPermissions 或 CreateTableDefaultPermissions 参数等同于传递空列表。

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ]
    }
}
```

此操作实际上会撤销 IAMAllowedPrincipals 组对新数据库和表的所有 Lake Formation 权限。创建数据库时，您可以覆盖此设置。

要仅通过 IAM 对新数据库和表实施元数据和基础数据访问控制，请按如下方式对 DataLakeSettings 结构进行编码。

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": [
                    "ALL"
                ]
            }
        ],
        "CreateTableDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
```

```
                },
                "Permissions": [
                    "ALL"
                ]
            }
        ]
    }
}
```

这将向 IAMAllowedPrincipals 组授予对新数据库和表的 Super Lake Formation 权限。创建数据库时，您可以覆盖此设置。

> **Note**
>
> 在前面的 DataLakeSettings 结构中，DataLakePrincipalIdentifier 的唯一允许值是 IAM_ALLOWED_PRINCIPALS，Permissions 的唯一允许值是 ALL。

隐式 Lake Formation 权限

AWS Lake Formation 向数据湖管理员、数据库创建者和表创建者授予以下隐式权限。

数据湖管理员

- 对数据目录中的所有资源具有 Describe 访问权限，但从另一个账户直接共享到其他主体的资源除外。管理员无法撤销此访问权限。
- 在数据湖中的任何位置都具有数据位置权限。
- 可以向任何主体（包括自身）授予或撤销对数据目录中任何资源的访问权限。管理员无法撤销此访问权限。
- 可以在数据目录中创建数据库。
- 可以向其他用户授予创建数据库的权限。

> **Note**
>
> 数据湖管理员只有在具有 IAM 权限的情况下才能注册 Amazon S3 位置。本指南中建议的数据湖管理员策略可以授予这些权限。此外，数据湖管理员没有删除数据库或更改/删除其他人创建的表的隐式权限。但是，他们可以授予自己执行此操作的权限。

有关数据湖管理员的更多信息，请参阅创建数据湖管理员。

数据库创建者

- 对自己创建的数据库具有所有数据库权限，对自己在数据库中创建的表具有权限，并且可以向同一 AWS 账户中的其他主体授予在数据库中创建表的权限。同时拥有 `AWSLakeFormationCrossAccountManager` AWS 托管式策略的数据库创建者可以向其他 AWS 账户或组织授予对数据库的权限。

  数据湖管理员可以使用 Lake Formation 控制台或 API 来指定数据库创建者。

  > **Note**
  >
  > 数据库创建者对其他人在数据库中创建的表不具有隐式权限。

有关更多信息，请参阅创建数据库。

表创建者

- 具有对自己创建的表的所有权限。
- 可以向同一 AWS 账户中的主体授予对他们创建的所有表的权限。
- 如果其他 AWS 账户或组织具有 `AWSLakeFormationCrossAccountManager` AWS 托管式策略，则可以向他们授予对自己创建的所有表的权限。
- 可以查看包含自己创建的表的数据库。

Lake Formation 权限参考

要执行 AWS Lake Formation 操作，委托人需要 Lake Formation 权限和 AWS Identity and Access Management (IAM) 权限。您通常使用粗粒度访问控制策略授予 IAM 权限，如 the section called “Lake Formation 权限概述 ”中所述。您可以使用控制台、API 或 AWS Command Line Interface (AWS CLI) 授予 Lake Formation 权限。

要了解如何授予或撤销 Lake Formation 权限，请参阅the section called “授予和撤销数据目录权限”和the section called “授予数据位置权限”。

> **Note**
>
> 本部分中的示例说明如何向同一 AWS 账户中的主体授予权限。有关跨账户授权的示例，请参阅 the section called “跨账户数据共享”。

每种资源类型的 Lake Formation 权限

以下是适用于每种资源类型的有效 Lake Formation 权限：

资源	权限
Database	ALL (Super)
	ALTER
	CREATE_TABLE
	DESCRIBE
	DROP
Table	ALL (Super)
	ALTER
	DELETE
	DESCRIBE
	DROP
	INSERT
	SELECT
View	ALL (Super)
	SELECT
	DESCRIBE
	DROP
Data Catalog	CREATE_DATABASE
Amazon S3 location	DATA_LOCATION_ACCESS

资源	权限	
LF-Tags	DROP	
	ALTER	
LF-Tag values	ASSOCIATE	
	DESCRIBE	
	GrantWithLFTagExpression	
LF-Tag policy - Database	ALL (Super)	
	ALTER	
	CREATE_TABLE	
	DESCRIBE	
	DROP	
LF-Tag policy - Table	ALL (Super)	
	ALTER	
	DESCRIBE	
	DELETE	
	DROP	
	INSERT	
	SELECT	
Resource link - Database or Table	DESCRIBE	
	DROP	

资源	权限
Table with data filters	DESCRIBE
	DROP
	SELECT
Table with column filter	SELECT

主题

- Lake Formation 授予和撤销命令 AWS CLI
- Lake Formation 权限

Lake Formation 授予和撤销命令 AWS CLI

本节中的每个权限描述都包括使用 AWS CLI 命令授予权限的示例。以下是 Lake Formation grant-permissions 和revoke-permissions AWS CLI 命令的提要。

```
grant-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

```
revoke-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

有关这些命令的详细说明，请参阅《AWS CLI 命令参考》中的 grant-permissions 和 revoke-permissions。本部分提供有关 `--principal` 选项的更多信息。

`--principal` 选项的值为以下值之一：

- (IAM) 用户或角色的亚马逊资源名称 AWS Identity and Access Management (ARN)
- 通过 SAML 提供商（例如 Microsoft Active Directory 联合身份验证服务 (AD FS) 进行身份验证的用户或组的 ARN
- 亚马逊 QuickSight 用户或群组的 ARN
- 对于跨账户权限，需要 AWS 账户 ID、组织 ID 或组织单位 ID

以下是所有 `--principal` 类型的语法和示例。

主体为 IAM 用户

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

主体为 IAM 角色

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

主体为通过 SAML 提供商进行身份验证的用户

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:user/<user-name>
```

示例：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:user/datalake_user1
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:user/athena-user@example.com
```

主体为通过 SAML 提供商进行身份验证的组

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:group/<group-name>
```

示例：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:group/data-scientists
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:group/my-group
```

校长是亚马逊 QuickSight 企业版用户

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:user/<namespace>/<user-name>
```

> **Note**
>
> 对于 *<namespace>*，必须指定 `default`。

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:user/default/bi_user1
```

校长是亚马逊 QuickSight 企业版群组

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:group/<namespace>/<group-name>
```

> **Note**
>
> 对于 *<namespace>*，必须指定 `default`。

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:group/default/data_scientists
```

本金是一个 AWS 账户

语法：

```
--principal DataLakePrincipalIdentifier=<account-id>
```

例如：

```
--principal DataLakePrincipalIdentifier=111122223333
```

主体为组织

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:organization/<organization-id>
```

例如：

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl
```

主体为组织单位

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:ou/<organization-id>/<organizational-unit-id>
```

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-
abcdefghijkl/ou-ab00-cdefghij
```

委托人是 IAM 身份中心身份用户或群组

示例：用户

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

示例：群组：

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

校长是一个 IAM 群组-**IAMAllowedPrincipals**

Lake Formation 将数据目录中所有数据库和表的Super权限设置为一个IAMAllowedPrincipals默认名为的组。如果数据库或表上存在此群组权限，则您账户中的所有委托人都可以通过 IAM 委托人策略访问该资源。 AWS Glue当您开始使用 Lake Formation 权限来保护之前受 IAM 策略保护的数据目录资源时，它提供了向后兼容性 AWS Glue。

使用 Lake Formation 管理数据目录资源的权限时，需要先撤消对资源的IAMAllowedPrincipals权限，或者将委托人和资源选择为混合访问模式，Lake Formation 权限才能生效。

例如：

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

校长是一个 IAM 群组-**ALLIAMPrincipals**

当您授予对数据目录资源的ALLIAMPrincipals分组权限时，账户中的每位委托人都可以使用 Lake Formation 权限和 IAM 权限访问数据目录资源。

例如：

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Lake Formation 权限

本部分包含您可以向主体授予的可用的 Lake Formation 权限。

ALTER

权限	针对此项资源授予的权限	被授权者还需要
ALTER	DATABASE	glue:UpdateDatabase
ALTER	TABLE	glue:UpdateTable
ALTER	LF-Tag	lakeformation:UpdateLFTag

具有此权限的主体可以更改数据目录中数据库或表的元数据。对于表，您可以更改列架构并添加列参数。您无法更改元数据表指向的基础数据中的列。

如果要更改的属性是已注册的 Amazon Simple Storage Service (Amazon S3) 位置，则主体必须对新位置具有数据位置权限。

Example

以下示例向 AWS 账户 1111-2222-33 datalake_user1 33 retail 中的用户授予数据库ALTER权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"}}'
```

Example

以下示例向用户 datalake_user1 授予对数据库 retail 中 inventory 表的 ALTER 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

CREATE_DATABASE

权限	针对此项资源授予的权限	被授权者还需要
CREATE_DATABASE	数据目录	glue:CreateDatabase

具有此权限的主体可以在数据目录中创建元数据数据库或资源链接。主体还可以在数据库中创建表。

Example

以下示例CREATE_DATABASE向 AWS 账户 1111-2222-33 datalake_user1 33 中的用户授权。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

当主体在数据目录中创建数据库时，不会授予对基础数据的权限。将授予以下其他元数据权限（以及将这些权限授予其他人的能力）：

- 数据库中的 CREATE_TABLE
- ALTER 数据库
- DROP 数据库

创建数据库时，主体可以选择指定一个 Amazon S3 位置。根据主体是否具有数据位置权限，CREATE_DATABASE 权限可能不足以在所有情况下创建数据库。请务必牢记以下三种情况。

创建数据库用例	所需权限
未指定位置属性。	CREATE_DATABASE 已足够。
指定了位置属性，并且该位置不由 Lake Formation 管理（未注册）。	CREATE_DATABASE 已足够。
指定了位置属性，并且该位置由 Lake Formation（已注册）管理。	需要 CREATE_DATABASE 以及指定位置的数据位置权限。

CREATE_TABLE

权限	针对此项资源授予的权限	被授权者还需要
CREATE_TABLE	DATABASE	glue:CreateTable

具有此权限的主体可以在指定数据库的数据目录中创建元数据表或资源链接。

Example

以下示例授予用户使用 AWS 账户 1111-2222-3333 在retail数据库中创建表的datalake_user1权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

当某个主体在数据目录中创建表时，系统会将表上的所有 Lake Formation 权限授予给该主体，并且该主体能够将这些权限授予其他主体。

跨账户资助

如果数据库拥有者账户向接收者账户授予 CREATE_TABLE 权限，并且该接收者账户中的用户在拥有者账户的数据库中成功创建了表，则以下规则适用：

- 接收者账户中的用户和数据湖管理员具有对表的所有 Lake Formation 权限。他们可以将对表的权限授予给其账户中的其他主体；但无法向拥有者账户或任何其他账户中的主体授予权限。
- 拥有者账户中的数据湖管理员可以向其账户中的其他主体授予对表的权限。

数据位置权限

当您尝试创建指向 Amazon S3 位置的表时，根据您是否具有数据位置权限，CREATE_TABLE 权限可能不足以创建表。请务必牢记以下三种情况。

创建表格用例	所需权限
指定位置不受 Lake Formation 管理（未注册）。	CREATE_TABLE 已足够。
指定位置由 Lake Formation（已注册）管理，并且包含的数据库没有位置属性或具有不是表位置的 Amazon S3 前缀的位置属性。	需要 CREATE_TABLE 以及指定位置的数据位置权限。
指定位置由 Lake Formation（已注册）管理，并且包含的数据库具有一个位置属性，该属性指向已注册的位置且是表位置的 Amazon S3 前缀。	CREATE_TABLE 已足够。

DATA_LOCATION_ACCESS

权限	针对此项资源授予的权限	被授权者还需要
DATA_LOCATION_ACCESS	Amazon S3 位置	（Amazon S3 对位置的权限，必须由用于注册位置的角色指定。）

这是唯一的数据位置权限。具有此权限的主体可以创建指向指定 Amazon S3 位置的元数据数据库或表。必须注册该位置。对某个位置具有数据位置权限的主体也对子位置具有位置权限。

Example

以下示例在 AWS 账户 1111-2222-3333 中向用户 datalake_user1 授予对 s3://products/retail 的数据位置权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::products/retail"}}'
```

查询或更新基础数据不需要 DATA_LOCATION_ACCESS 权限。此权限仅适用于创建数据目录资源。

有关数据位置权限的更多信息，请参阅Underlying data access control。

DELETE

权限	针对此项资源授予的权限	被授权者还需要
DELETE	TABLE	（如果注册了位置，则无需其他 IAM 权限。）

具有此权限的主体可以删除表指定的 Amazon S3 位置的基础数据。主体还可以在 Lake Formation 控制台上查看表，并使用 AWS Glue API 检索有关表的信息。

Example

以下示例向 AWS 账户 1111-2222-33 datalake_user1 33 inventory 中的用户授予对数据库retail中表的DELETE权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

此权限仅适用于 Amazon S3 中的数据，不适用于 Amazon Relational Database Service (Amazon RDS) 等其他数据存储中的数据。

DESCRIBE

权限	针对此项资源授予的权限	被授权者还需要
DESCRIBE	表资源链接	glue:GetTable
	数据库资源链接	glue:GetDatabase
DESCRIBE	DATABASE	glue:GetDatabase
DESCRIBE	TABLE	glue:GetTable
DESCRIBE	LF-Tag	glue:GetTable

权限	针对此项资源授予的权限	被授权者还需要
		glue:GetDatabase lakeformation:GetResourceLFTags lakeformation:ListLFTags lakeformation:GetLFTag lakeformation:SearchTablesByLFTags lakeformation:SearchDatabasesByLFTags

具有此权限的主体可以查看指定的数据库、表或资源链接。不会隐式授予任何其他数据目录权限，也不会隐式授予任何数据访问权限。数据库和表显示在集成服务的查询编辑器中，但除非授予其他 Lake Formation 权限（例如 SELECT），否则无法对它们进行查询。

例如，对数据库具有 DESCRIBE 权限的用户可以查看数据库和所有数据库元数据（描述、位置等）。但是，用户无法找出数据库包含哪些表，也无法删除、更改或创建数据库中的表。同样，对表具有 DESCRIBE 权限的用户可以查看表和表元数据（描述、架构、位置等），但无法删除、更改或运行对表的查询。

以下是 DESCRIBE 的一些附加规则：

- 如果用户对数据库、表或资源链接具有其他 Lake Formation 权限，则会隐式授予 DESCRIBE 权限。
- 如果用户仅对表的列子集具有 SELECT 权限（部分 SELECT），则用户只能查看这些列。
- 您无法向对表具有部分 SELECT 权限的用户授予 DESCRIBE 权限。反之，您无法为被授予了 DESCRIBE 权限的表指定列包含或排除列表。

Example

以下示例向 AWS 账户 1111-2222-33 datalake_user1 33 中的用户授予对数据库inventory-linkretail中表资源链接的DESCRIBE权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory-link"}}'
```

DROP

权限	针对此项资源授予的权限	被授权者还需要
DROP	DATABASE	glue:DeleteDatabase
DROP	TABLE	glue:DeleteTable
DROP	LF-Tag	lakeformation:DeleteLFTag
DROP	数据库资源链接 表资源链接	glue:DeleteDatabase glue:DeleteTable

具有此权限的主体可以在数据目录中删除数据库、表或资源链接。您无法向外部账户或组织授予对数据库的 DROP 权限。

> **Warning**
>
> 删除数据库会删除数据库中的所有表。

Example

以下示例向 AWS 账户 1111-2222-33 datalake_user1 33 retail 中的用户授予数据库DROP权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Example

以下示例向用户 datalake_user1 授予对数据库 retail 中 inventory 表的 DROP 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Example

以下示例向用户 datalake_user1 授予对数据库 retail 中表资源链接 inventory-link 的 DROP 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

权限	针对此项资源授予的权限	被授权者还需要
INSERT	TABLE	（如果注册了位置，则无需其他 IAM 权限。）

具有此权限的主体可以在表指定的 Amazon S3 位置插入、更新和读取基础数据。主体还可以在 Lake Formation 控制台中查看表，并使用 AWS Glue API 检索有关表的信息。

Example

以下示例向 AWS 账户 1111-2222-33 datalake_user1 33 inventory 中的用户授予对数据库retail中表的INSERT权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

此权限仅适用于 Amazon S3 中的数据，不适用于 Amazon RDS 等其他数据存储中的数据。

SELECT

权限	针对此项资源授予的权限	被授权者还需要
SELECT	• TABLE	（如果注册了位置，则无需其他 IAM 权限。）

具有此权限的主体可以查看数据目录中的表，并可以在表指定的位置查询 Amazon S3 中的基础数据。主体可以在 Lake Formation 控制台中查看表，并使用 AWS Glue API 检索有关表的信息。如果在授予此权限时应用了列筛选，则主体只能查看所包含列的元数据，并且只能从所包含的列中查询数据。

> **Note**
>
> 集成分析服务负责在处理查询时应用列筛选。

Example

以下示例向 AWS 账户 1111-2222-33 `datalake_user1` 33 `inventory` 中的用户授予对数据库`retail`中表的SELECT权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

此权限仅适用于 Amazon S3 中的数据，不适用于 Amazon RDS 等其他数据存储中的数据。

您可以使用可选的包含列表或排除列表来筛选特定列（限制访问权限）。包含列表指定可以访问的列。排除列表指定无法访问的列。如果没有包含列表或排除列表，则所有表列均可访问。

`glue:GetTable` 的结果仅返回调用方有权查看的列。Amazon Athena 和 Amazon Redshift 等集成服务支持列包含和排除列表。

Example

以下示例使用包含列表向用户 `datalake_user1` 授予对表 `inventory` 的 SELECT 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
```

```
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnNames": ["prodcode","location","period","withdrawals"]}}'
```

Example

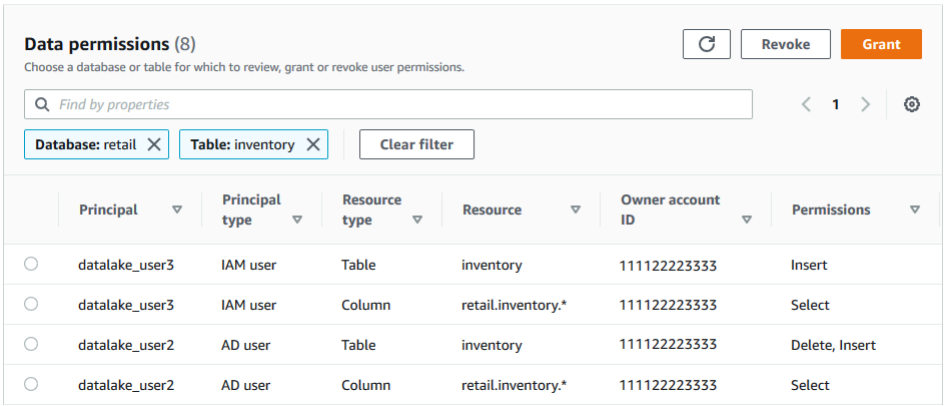
下一个示例使用排除列表授予对 inventory 表的 SELECT 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnWildcard": {"ExcludedColumnNames": ["intkey",
 "prodcode"]}}}'
```

以下限制适用于 SELECT 权限：

- 授予 SELECT 时，如果应用了列筛选，则不能包含授权选项。
- 不能限制对作为分区键的列的访问控制。
- 不能向对表中的列子集具有 SELECT 权限的主体授予对该表的 ALTER、DROP、DELETE 或 INSERT 权限。同样，不能通过列筛选向对表具有 ALTER、DROP、DELETE 或 INSERT 权限的主体授予 SELECT 权限。

SELECT 权限始终作为单独的行显示在 Lake Formation 控制台的数据权限页面上。下图显示向用户 datalake_user2 和 datalake_user3 授予对 inventory 表中所有列的 SELECT 权限。



Super

权限	针对此项资源授予的权限	被授权者还需要
Super	DATABASE	glue:*Database*

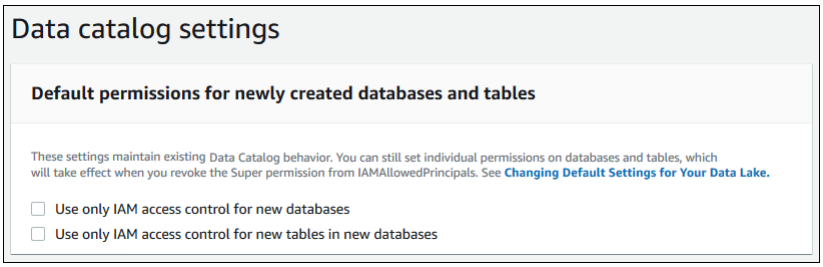
权限	针对此项资源授予的权限	被授权者还需要
Super	TABLE	glue:*Table*, glue:*Partition*

此权限允许主体对数据库或表执行所有支持的 Lake Formation 操作。您无法向外部账户授予对数据库的 Super 权限。

此权限可以与其他 Lake Formation 权限共存。例如，您可以授予对元数据表的 Super、SELECT、和 INSERT 权限。然后，主体可以对表执行所有受支持的操作。撤销 Super 后，SELECT 和 INSERT 权限将保留，且主体只能执行 select 和 insert 操作。

您可以将 Super 授予组 IAMAllowedPrincipals，而不是将其授予单个主体。IAMAllowedPrincipals 组是自动创建的，其中包括 IAM 策略允许访问数据目录资源的所有 IAM 用户和角色。当向 IAMAllowedPrincipals 授予对数据目录资源的 Super 权限时，对该资源的访问实际上完全由 IAM 策略控制。

您可以利用 Lake Form IAMAllowedPrincipals ation 控制台的 “设置” 页面上的选项，自动获得新目录资源的Super权限。



- 要向 IAMAllowedPrincipals 授予对所有新数据数据库的 Super 权限，请选择仅对新数据库使用 IAM 访问控制。
- 要向 IAMAllowedPrincipals 授予对新数据数据库中所有新表的 Super 权限，请选择仅对新数据库中的新表使用 IAM 访问控制。

> **Note**
>
> 此选项会导致默认情况下选中创建数据库对话框中的仅对此数据库中的新表使用 IAM 访问控制复选框。它的作用仅此而已。它是创建数据库对话框中的复选框，用于向 IAMAllowedPrincipals 授予 Super 权限。

默认情况下，将启用这些设置页面选项。有关更多信息，请参阅下列内容：

- the section called “更改数据湖的默认设置”
- the section called “将 AWS Glue 数据权限升级为 Lake Formation 模型”

ASSOCIATE

权限	针对此项资源授予的权限	被授权者还需要
ASSOCIATE	LF-Tag	glue:GetDatabase glue:GetTable lakeformation:AddLFTagsToResource" lakeformation:RemoveLFTagsFromResource" lakeformation:GetResourceLFTags lakeformation:ListLFTags lakeformation:GetLFTag lakeformation:SearchTablesByLFTags lakeformation:SearchDatabasesByLFTags

对 LF 标签具有此权限的主体可以将 LF 标签分配给数据目录资源。授予 ASSOCIATE 会隐式授予 DESCRIBE 权限。

Example

此示例向用户 datalake_user1 授予对带有键 module 的 LF 标签的 ASSOCIATE 权限。它授予查看和分配该键的所有值的权限，如星号 (*) 所示。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

集成 IAM Identity Center

借 AWS IAM Identity Center助，您可以连接到身份提供商 (IdPs)，并集中管理 AWS 分析服务中用户和群组的访问权限。您可以将 Okta、Ping 和 Microsoft Entra ID（以前称为 Azure Active Directory）等身份提供者与 IAM Identity Center 集成，以便您组织中的用户使用单点登录体验访问数据。IAM Identity Center 还支持连接其他第三方身份提供者。

有关更多信息，请参阅《 AWS IAM Identity Center 用户指南》中支持的身份提供商。

您可以在 IAM Identity Center 中配置 AWS Lake Formation 为已启用的应用程序，数据湖管理员可以向授权用户和群组授予对 AWS Glue Data Catalog 资源的精细权限。

您组织中的用户可以使用组织的身份提供程者登录到任何启用了 Identity Center 的应用程序，并查询应用 Lake Formation 权限的数据集。通过此集成，您无需创建多个 IAM 角色即可管理对 AWS 服务的访问权限。

> **Note**
>
> 可信身份传播允许用户现有的用户和群组成员资格访问所有 AWS 分析服务的数据。通过可信身份传播，用户可以登录应用程序，应用程序可以在请求访问 AWS 服务中的数据时传递用户的身份。您无需执行任何特定于服务的身份提供商配置或 IAM 角色设置。有关更多信息，请参阅《 AWS IAM Identity Center 用户指南》中的跨应用程序的可信身份传播。

有关限制，请参阅IAM Identity Center 集成限制。

主题

- 先决条件

- 将 Lake Formation 与 IAM Identity Center 连接
- 更新 IAM Identity Center 集成
- 删除 Lake Formation 与 IAM Identity Center 的连接
- 向用户和组授予权限

先决条件

以下是将 IAM Identity Center 与 Lake Formation 集成的先决条件。

1. 启用 IAM Identity Center – 启用 IAM Identity Center 是支持身份验证和身份传播的先决条件。
2. 选择您的身份源 – 启用 IAM Identity Center 后，您必须有身份提供者来管理用户和组。您可以使用内置的 Identity Center 目录作为身份源，也可以使用外部 IdP，例如 Microsoft Entra ID 或 Okta。

   有关更多信息，请参阅《AWS IAM Identity Center 用户指南》中的 “管理您的身份源” 和 “Connect 到外部身份提供商”。
3. 创建 IAM 角色 - 创建 IAM Identity Center 连接的角色需要具有在 Lake Formation 和 IAM Identity Center 中创建和修改应用程序配置的权限，如以下内联策略所示。

   您需要按照 IAM 最佳实践添加权限。后面的步骤将会详细介绍具体权限。有关更多信息，请参阅开始使用 IAM Identity Center。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
                "sso:CreateApplication",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
```

```
}
```

以下内联策略包含查看、更新和删除 Lake Formation 与 IAM Identity Center 集成的属性所需的特定权限。

- 使用以下内联策略允许 IAM 角色查看 Lake Formation 与 IAM Identity Center 的集成。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- 使用以下内联策略允许 IAM 角色更新 Lake Formation 与 IAM Identity Center 的集成。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication",
                "sso:UpdateApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
```

```
}
```

- 使用以下内联策略允许 IAM 角色删除 Lake Formation 与 IAM Identity Center 的集成。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
                "sso:DeleteApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- 有关授予或撤销 IAM Identity Center 用户和组的数据湖权限所需的 IAM 权限，请参阅[授予或撤销 Lake Formation 权限所需的 IAM 权限]。

权限描述

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration` – 创建 Lake Formation IdC 配置。
- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration` – 描述现有的 IdC 配置。
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration` – 允许删除现有的 Lake Formation IdC 配置。
- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration` – 用于更改现有的 Lake Formation 配置。
- `sso:CreateApplication` – 用于创建 IAM Identity Center 应用程序。
- `sso:DeleteApplication` – 用于删除 IAM Identity Center 应用程序。
- `sso:UpdateApplication` – 用于更新 IAM Identity Center 应用程序。
- `sso:PutApplicationGrant` – 用于更改可信令牌发布者信息。

- `sso:PutApplicationAuthenticationMethod` – 授予 Lake Formation 身份验证访问权限。
- `sso:GetApplicationGrant` – 用于列出可信令牌发布者信息。
- `sso:DeleteApplicationGrant` – 删除可信令牌颁发者信息。
- `sso:PutApplicationAccessScope` – 添加或更新应用程序的 IAM Identity Center 访问范围的授权目标列表。
- `sso:PutApplicationAssignmentConfiguration` – 用于配置用户访问应用程序的方式。

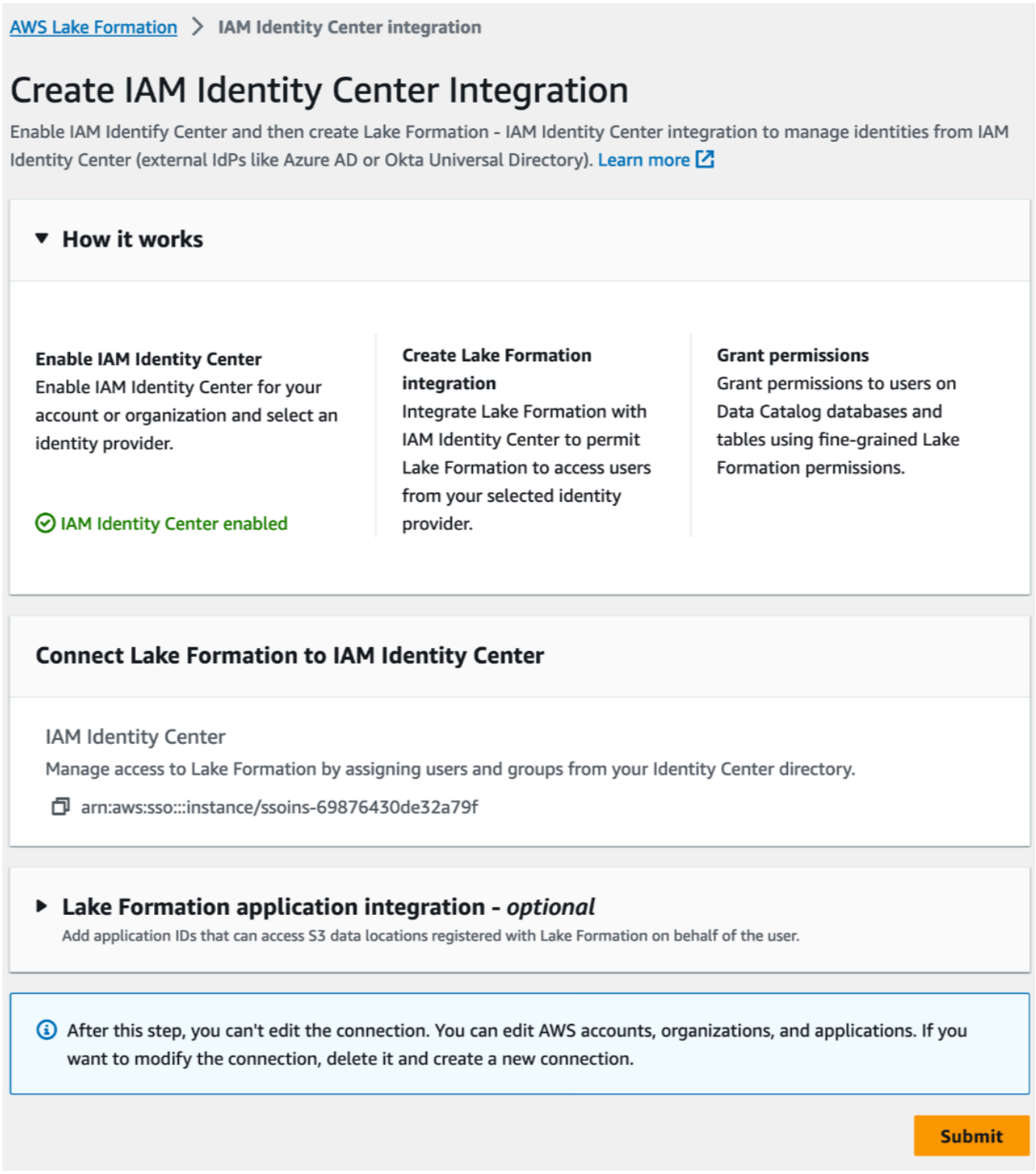
将 Lake Formation 与 IAM Identity Center 连接

您必须先完成以下步骤，然后才能使用 IAM Identity Center 管理身份以使用 Lake Formation 授予对数据目录资源的访问权限。您可以使用 Lake Formation 控制台或 AWS CLI创建 IAM Identity Center 集成。

AWS Management Console

将 Lake Formation 与 IAM Identity Center 连接

1. 登录并打开 Lake AWS Management Console Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。
2. 在左侧导航窗格中，选择 IAM Identity Center 集成。



3. （可选）在创建 Lake Formation 集成屏幕上，指定第三方应用程序的 ARN，这些应用程序可以访问已向 Lake Formation 注册的 Amazon S3 位置中的数据。Lake Formation 根据有效权限以 AWS STS 令牌的形式向已注册的 Amazon S3 地点出售限定范围的临时证书，以便授权的应用程序可以代表用户访问数据。

4. 选择提交。

   Lake Formation 管理员完成这些步骤并创建集成后，IAM Identity Center 属性将显示在 Lake Formation 控制台中。完成这些任务后，Lake Formation 将成为启用了 IAM Identity Center 的应用程序。控制台中的属性包括集成状态。集成完成后，状态显示为 `Success`。此状态指示 IAM Identity Center 配置是否已完成。

AWS CLI

- 以下示例演示如何创建 Lake Formation 与 IAM Identity Center 的集成。您还可以指定该应用程序的 `Status`（`ENABLED`、`DISABLED`）。

```
aws lakeformation create-lake-formation-identity-center-configuration \
    --catalog-id <123456789012> \
    --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
    --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"],
 "Status": "ENABLED"}'
```

- 以下示例演示如何查看 Lake Formation 与 IAM Identity Center 的集成。

```
aws lakeformation describe-lake-formation-identity-center-configuration
 --catalog-id <123456789012>
```

更新 IAM Identity Center 集成

创建连接后，您可以为 IAM Identity Center 集成添加第三方应用程序，以便与 Lake Formation 集成，并可以代表用户访问 Amazon S3 数据。您也可以从 IAM Identity Center 集成中移除现有应用程序。您可以使用 Lake Formation 控制台和UpdateLakeFormationIdentityCenterConfiguration操作添加或删除应用程序。 AWS CLI

> **Note**
>
> 创建 IAM Identity Center 集成后，您无法更新实例 ARN。

AWS Management Console

更新 IAM Identity Center 与 Lake Formation 的现有连接

1. 登录并打开 Lake AWS Management Console Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。
2. 在左侧导航窗格中，选择 IAM Identity Center 集成。
3. 在 IAM Identity Center 集成页面上选择添加。
4. 在添加应用程序屏幕上，输入要与 Lake Formation 集成的第三方应用程序的应用程序 ID。
5. 选择添加。

AWS CLI

您可以通过运行以下 AWS CLI 命令为 IAM Identity Center 集成添加或删除第三方应用程序。当您将外部筛选状态设置为 ENABLED 时，它使 IAM Identity Center 能够为第三方应用程序提供身份管理，以便访问 Lake Formation 管理的数据。您还可以通过设置应用程序状态来启用或禁用 IAM Identity Center 集成。

```
aws lakeformation update-lake-formation-identity-center-configuration \
 --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"], "Status":
 "ENABLED"}'\
 --application-status ENABLED
```

删除 Lake Formation 与 IAM Identity Center 的连接

如果您想删除现有的 IAM 身份中心集成，可以使用 Lake Formation 控制台或DeleteLakeFormationIdentityCenterConfiguration操作来删除。 AWS CLI

AWS Management Console

删除 IAM Identity Center 与 Lake Formation 的现有连接

1. 登录并打开 Lake AWS Management Console Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。
2. 在左侧导航窗格中，选择 IAM Identity Center 集成。
3. 在 IAM Identity Center 集成页面上选择删除。
4. 在确认集成屏幕上，确认该操作，然后选择删除。

AWS CLI

您可以通过运行以下 AWS CLI 命令删除 IAM 身份中心集成。

```
aws lakeformation delete-lake-formation-identity-center-configuration \
    --catalog-id <123456789012>
```

向用户和组授予权限

您的数据湖管理员可以向 IAM Identity Center 用户和组授予对数据目录资源（数据库、表和视图）的权限，以便轻松访问数据。要授予或撤销数据湖权限，授予者需要具有执行以下 IAM Identity Center 操作的权限。

- DescribeUser
- DescribeGroup
- DescribeInstance

您可以使用 Lake Formation 控制台、API 或 AWS CLI来授予权限。

有关授予权限的更多信息，请参阅the section called “授予和撤销数据目录权限”。

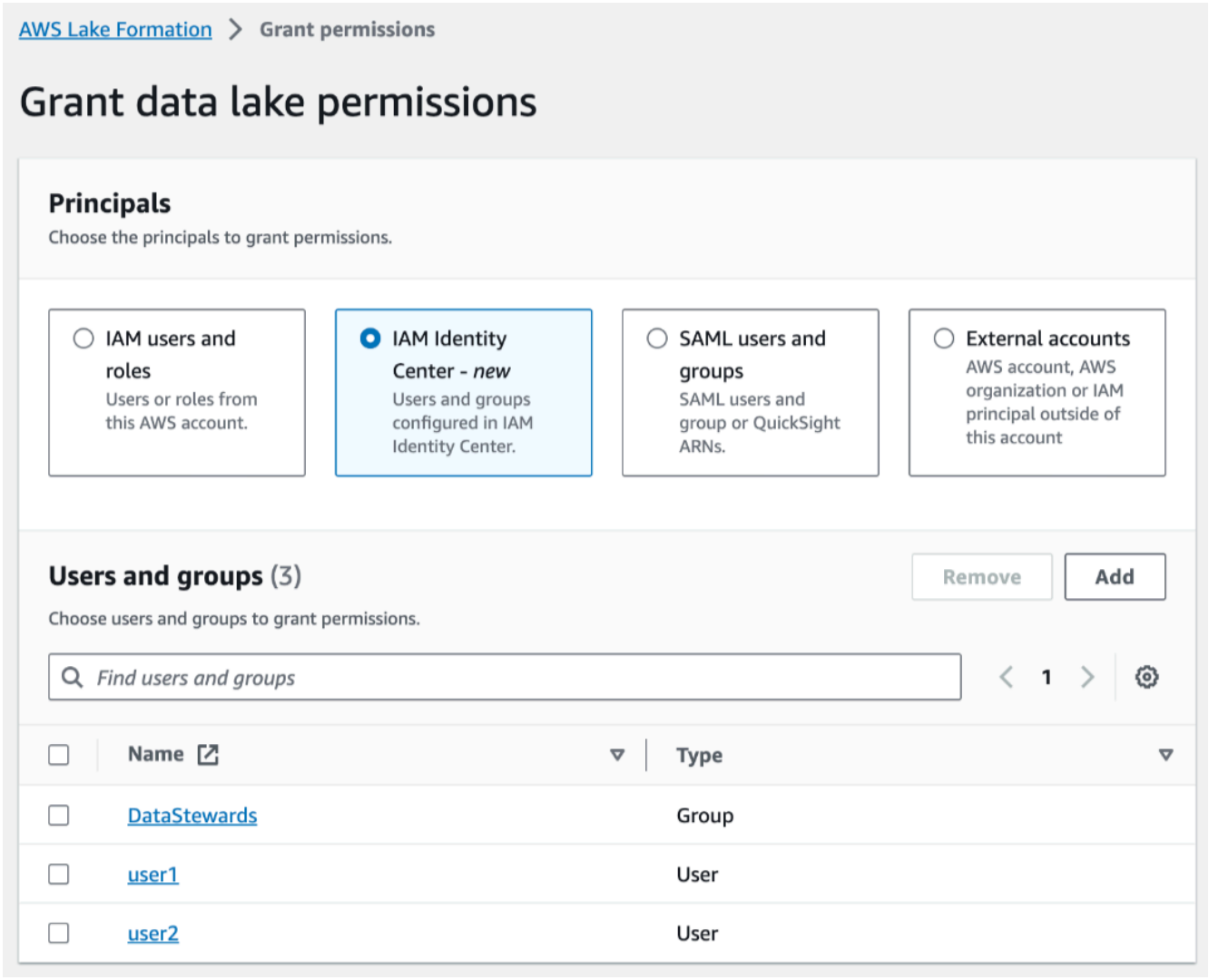
> **Note**
>
> 您只能授予对账户中资源的权限。要将权限级联到用户和群组对与您共享的资源，您必须使用 AWS RAM 资源共享。

AWS Management Console

向用户和组授予权限

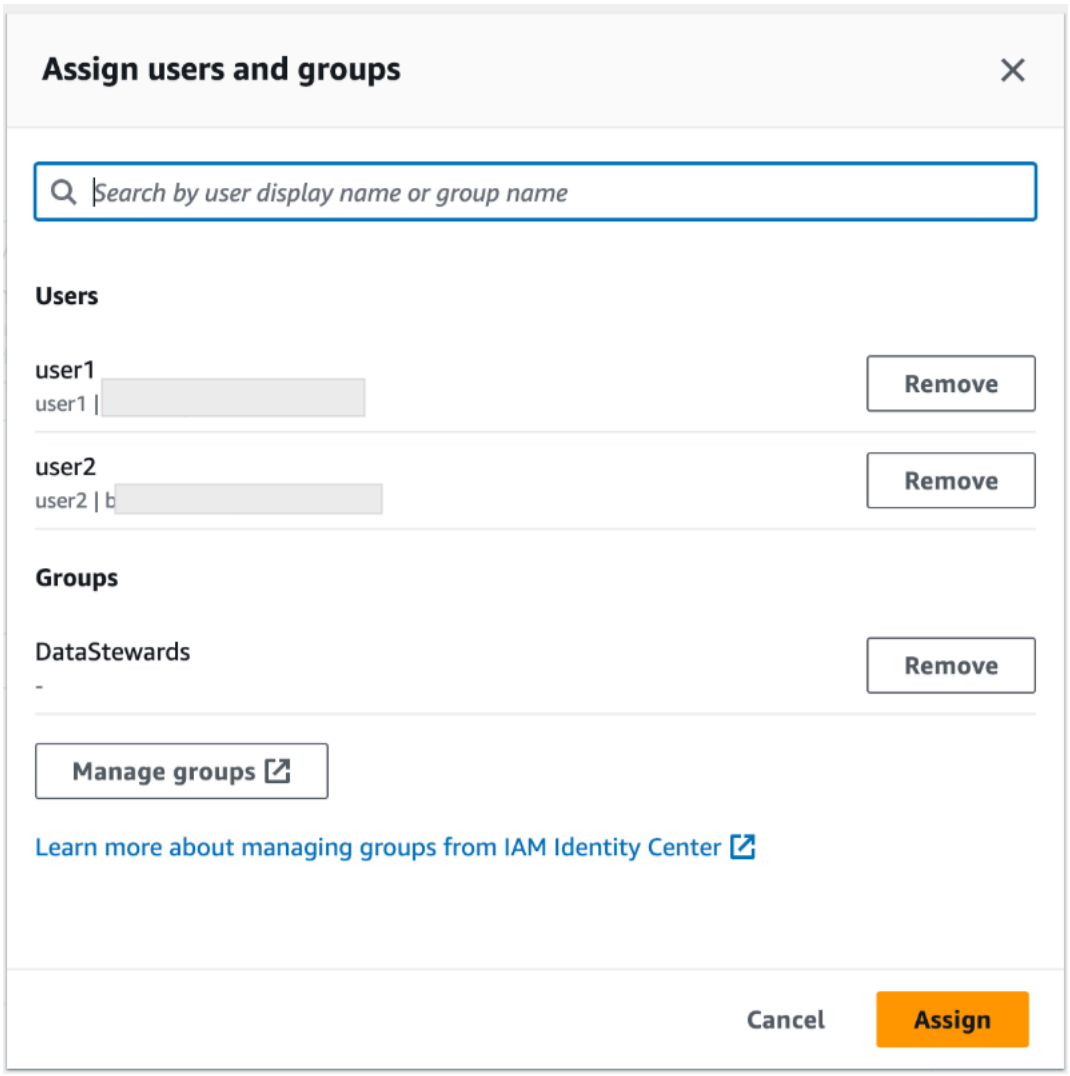
1. 登录并打开 Lake AWS Management Console Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。
2. 在 Lake Formation 控制台的权限下，选择数据湖权限。
3. 选择授予。
4. 在授予数据湖权限页面上，选择 SSM 用户和组。

5. 选择添加以选择要授予权限的用户和组。



6. 在分配用户和组屏幕上，选择要授予权限的用户和/或组。

   选择分配。



7. 接下来，选择授予权限的方法。

   有关使用命名资源方法授予权限的说明，请参阅使用命名资源方法授予数据湖权限。

   有关使用 LF 标签授予权限的说明，请参阅使用 LF-TBAC 方法授予数据湖权限。
8. 选择要授予其权限的数据目录资源。
9. 选择要授予的数据目录权限。
10. 选择授予。

AWS CLI

以下示例演示如何向 IAM Identity Center 用户授予对表的 SELECT 权限。

```
aws lakeformation grant-permissions \
```

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \
--permissions "SELECT" \
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

要UserId从 IAM 身份中心检索，请参阅 IAM 身份中心 API 参考中的GetUserId操作。

向数据湖添加 Amazon S3 位置

要在数据湖中添加亚马逊简单存储服务 (Amazon S3) Service 位置作为存储空间，请向注册该位置。 AWS Lake Formation然后，您可以使用 Lake Formation 权限对指向该位置的 AWS Glue Data Catalog 对象以及该位置中的基础数据进行精细的访问控制。

Lake Formation 还允许在混合访问模式下注册数据位置，并让您能够灵活地选择为数据目录中的数据库和表启用 Lake Formation 权限。在混合访问模式下，您现在有了增量路径，允许您为一组特定的用户设置 Lake Formation 权限，而不会中断其他现有用户或工作负载的权限策略。

有关设置混合访问模式的更多信息，请参阅混合访问模式

当您注册某个位置时，将注册 Amazon S3 路径以及该路径下的所有文件夹。

例如，假设您有一个如下所示的 Amazon S3 路径组织：

```
/mybucket/accounting/sales/
```

如果您注册了 S3://mybucket/accounting，则 sales 文件夹也会被注册并由 Lake Formation 管理。

有关注册位置的更多信息，请参阅Underlying data access control。

> **Note**
>
> 建议对结构化数据（按包含行和列的表进行排列）使用 Lake Formation 权限。如果您的数据包含基于对象的非结构化数据，请考虑使用 Amazon S3 的 IAM 权限来管理数据访问。

主题

- 用于注册位置的角色的要求

- 注册 Amazon S3 位置
- 注册加密的 Amazon S3 位置
- 在其他 AWS 账户中注册 Amazon S3 位置
- 跨 AWS 账户注册加密的 Amazon S3 位置
- 取消注册 Amazon S3 位置

用于注册位置的角色的要求

注册亚马逊简单存储服务 AWS Identity and Access Management (Amazon S3) 位置时，必须指定 (IAM) 角色。 AWS Lake Formation 在访问该位置的数据时担任该角色。

您可以使用以下角色类型之一来注册位置：

- Lake Formation 服务相关角色。此角色授予对该位置的所需权限。使用此角色是注册位置的最简单方法。有关更多信息，请参阅在 Lake Formation 中使用服务相关角色。
- 用户定义的角色。当您需要授予的权限多于服务相关角色提供的权限时，请使用用户定义的角色。

  在以下情况下，您必须使用用户定义的角色：

  - 在其他账户中注册位置时。

    有关更多信息，请参阅 the section called “在其他 AWS 账户中注册 Amazon S3 位置” 和 the section called “跨 AWS 账户注册加密的 Amazon S3 位置”。
  - 如果您使用 AWS 托管 CMK (aws/s3) 对 Amazon S3 位置进行加密。

    有关更多信息，请参阅注册加密的 Amazon S3 位置。
  - 如果您计划使用 Amazon EMR。

    如果您已使用服务相关角色注册了某个位置，并且想要开始使用 Amazon EMR 访问该位置，则必须取消注册该位置，然后使用用户定义的角色重新注册该位置。有关更多信息，请参阅the section called “取消注册 Amazon S3 位置”。

在 Lake Formation 中使用服务相关角色

AWS Lake Formation 使用 AWS Identity and Access Management (IAM) 服务相关角色。服务相关角色是一种独特类型的 IAM 角色，它与 Lake Formation 直接相关。服务相关角色是由 Lake Formation 预定义的，包含该服务代表您调用其他 AWS 服务所需的所有权限。

服务相关角色可让您更轻松地设置 Lake Formation，因为您不必创建角色并手动添加必要的权限。Lake Formation 定义其服务相关角色的权限，除非另有定义，否则只有 Lake Formation 可以代入该角色。定义的权限包括信任策略和权限策略，以及不能附加到任何其他 IAM 实体的权限策略。

此服务相关角色仅信任以下服务来代入该角色：

- `lakeformation.amazonaws.com`

Lake Formation 的服务相关角色权限

Lake Formation 使用名为 `AWSServiceRoleForLakeFormationDataAccess` 的服务相关角色。此角色提供了一组 Amazon Simple Storage Service (Amazon S3) 权限，使 Lake Formation 集成服务（如 Amazon Athena）能够访问注册位置。注册数据湖位置时，您必须提供对该位置具有所需 Amazon S3 读/写权限的角色。您可以使用此服务相关角色，而不是创建具有所需 Amazon S3 权限的角色。

首次将服务相关角色命名为用于注册路径的角色时，将代表您创建服务相关角色和新的 IAM 策略。Lake Formation 将路径添加到内联策略，并将其附加到服务相关角色。当您向服务相关角色注册后续路径时，Lake Formation 会将该路径添加到现有策略中。

以数据湖管理员身份登录后，注册数据湖位置。然后，在 IAM 控制台中，搜索角色 `AWSServiceRoleForLakeFormationDataAccess` 并查看其附加的策略。

例如，在您注册位置 `s3://my-kinesis-test/logs` 后，Lake Formation 会创建以下内联策略并将其附加到 `AWSServiceRoleForLakeFormationDataAccess`。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3",
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject",
                "s3:AbortMultipartUpload",
                "s3:ListMultipartUploadParts"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test/logs/*"
            ]
```

```
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test"
            ]
        }
    ]
}
```

需要以下权限才能使用此服务相关角色注册位置：

- `iam:CreateServiceLinkedRole`
- `iam:PutRolePolicy`

数据湖管理员通常具有这些权限。

以下是用户定义的角色的要求：

- 创建新角色时，在 IAM 控制台的创建角色页面上，选择 AWS 服务，然后在选择一个使用案例下选择 Lake Formation。

  如果使用其他路径创建角色，请确保该角色与 `lakeformation.amazonaws.com` 具有信任关系。有关更多信息，请参阅修改角色信任策略（控制台）。
- 该角色必须与以下实体建立信任关系：
  - `glue.amazonaws.com`
  - `lakeformation.amazonaws.com`

  有关更多信息，请参阅修改角色信任策略（控制台）。
- 该角色必须具有授予 Amazon S3 对该位置的读/写权限的内联策略。以下是典型的策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket"
            ]
        }
    ]
}
```

- 注册该位置的数据湖管理员必须具有对该角色的 `iam:PassRole` 权限。

  以下是授予此权限的内联策略。<account-id>替换为有效的 AWS 账号，然后<role-name>替换为角色的名称。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::<account-id>:role/<role-name>"
            ]
        }
    ]
}
```

- 要允许 Lake Formati CloudWatch on 在日志中添加日志并发布指标，请添加以下内联策略。

> **Note**
>
> 写入 CloudWatch 日志会产生费用。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Sid1",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents"
            ],
            "Resource": [
                 "arn:aws:logs:<region>:<account-id>:log-group:/aws-lakeformation-
acceleration/*",
                 "arn:aws:logs:<region>:<account-id>:log-group:/aws-lakeformation-
acceleration/*:log-stream:*"
            ]
        }
    ]
}
```

注册 Amazon S3 位置

注册亚马逊简单存储服务 AWS Identity and Access Management (Amazon S3) 位置时，必须指定 (IAM) 角色。Lake Formation 在向访问该位置数据的集成 AWS 服务授予临时证书时担任该角色。

> **Important**
>
> 避免注册启用了请求者付费的 Amazon S3 存储桶。对于在 Lake Formation 中注册的存储桶，用于注册存储桶的角色始终被视为请求者。如果存储桶被其他 AWS 账户访问，则如果该角色与存储桶拥有者属于同一个账户，则该存储桶拥有者需要支付数据访问费用。

您可以使用 AWS Lake Formation 控制台、Lake Formation API 或 AWS Command Line Interface (AWS CLI) 注册亚马逊 S3 地点。

开始前的准备工作

查看用于注册位置的角色的要求。

注册位置（控制台）

> **Important**
>
> 以下过程假设 Amazon S3 位置与数据目录位于同一个 AWS 账户中，并且该位置中的数据未加密。本章中的其他部分介绍了跨账户注册和加密位置的注册。

1. 打开 AWS Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。以数据湖管理员或具有 `lakeformation:RegisterResource` IAM 权限的用户身份登录。
2. 在导航窗格的注册和提取下，选择数据湖位置。
3. 选择注册位置，然后选择浏览以选择 Amazon Simple Storage Service (Amazon S3) 路径。
4. （可选，但强烈推荐）选择查看位置权限以查看所选 Amazon S3 位置中所有现有资源及其权限的列表。

   注册所选位置可能会导致 Lake Formation 用户可以访问该位置已有的数据。查看此列表有助于确保现有数据保持安全。
5. 对于 IAM 角色，选择 `AWSServiceRoleForLakeFormationDataAccess` 服务相关角色（默认）或符合the section called “用于注册位置的角色的要求”中要求的自定义 IAM 角色。

   只有在使用自定义 IAM 角色注册已注册位置时，您才能更新该位置或其他详细信息。要编辑使用服务相关角色注册的位置，应取消注册该位置，然后重新注册。
6. 选择 “启用数据目录联合” 选项，允许 Lake Formation 代入角色并向集成 AWS 服务提供临时凭证，以访问联合数据库下的表。如果某个位置已注册到 Lake Formation，并且您希望对联合数据库下的表使用同一位置，则需要使用启用数据目录联合身份验证选项注册该同一位置。
7. 选择混合访问模式以默认不启用 Lake Formation 权限。当您在混合访问模式下注册 Amazon S3 位置时，您可以通过选择该位置下的数据库和表的主体来启用 Lake Formation 权限。

   有关设置混合访问模式的更多信息，请参阅混合访问模式。
8. 选择注册位置。

注册位置 (AWS CLI)

1. 向 Lake Formation 注册新位置

   此示例使用服务相关角色注册位置。您可以改用 `--role-arn` 参数来提供自己的角色。

   <s3-path>替换为有效的 Amazon S3 路径、<s3-access-role>使用有效 AWS 账户的账号以及有权注册数据位置的 IAM 角色。

   > **Note**
   >
   > 如果已注册位置是使用服务相关角色注册的，则无法编辑该位置的属性。

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --use-service-linked-role
   ```

   以下示例使用自定义角色注册位置。

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
   ```

2. 更新向 Lake Formation 注册的位置

   仅当使用自定义 IAM 角色注册已注册位置时，您才能对其进行编辑。对于使用服务相关角色注册的位置，应取消注册该位置并重新注册。有关更多信息，请参阅the section called “取消注册 Amazon S3 位置”。

   ```
   aws lakeformation update-resource \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>\
    --resource-arn arn:aws:s3:::<s3-path>
   ```

   ```
   aws lakeformation update-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --use-service-linked-role
   ```

3. 在混合访问模式下使用联合身份验证注册数据位置

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --hybrid-access-enabled
```

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --with-federation
```

```
aws lakeformation update-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --hybrid-access-enabled
```

有关更多信息，请参阅 RegisterResourceAPI 操作。

> **Note**
>
> 注册了 Amazon S3 位置后，任何指向该地点（或其任何子位置）的 AWS Glue 表都将返回`GetTable`调用`true`中`IsRegisteredWithLakeFormation`参数的值。存在一个已知限制，即数据目录 API 操作（如 `GetTables` 和 `SearchTables`）不会更新 `IsRegisteredWithLakeFormation` 参数的值，并返回默认值 false。建议使用 `GetTable` API 查看 `IsRegisteredWithLakeFormation` 参数的正确值。

注册加密的 Amazon S3 位置

Lake Formation 与 AWS Key Management Service (AWS KMS) 集成，使您能够更轻松地设置其他集成服务，以加密和解密 Amazon Simple Storage Service (Amazon S3) 位置中的数据。

既由客户管理 AWS KMS keys ，又 AWS 托管式密钥 受支持。目前，只有 Athena 支持客户端加密/解密。

在注册 Amazon S3 营业地点时，您必须指定 AWS Identity and Access Management (IAM) 角色。对于加密的 Amazon S3 位置，要么角色必须有权使用加密和解密数据 AWS KMS key，要么是 KMS 密钥策略必须向该角色授予对密钥的权限。

> **Important**
>
> 避免注册启用了请求者付费的 Amazon S3 存储桶。对于在 Lake Formation 中注册的存储桶，用于注册存储桶的角色始终被视为请求者。如果存储桶被其他 AWS 账户访问，则如果该角色与存储桶拥有者属于同一个账户，则该存储桶拥有者需要支付数据访问费用。

注册位置的最简单方法是使用 Lake Formation 服务相关角色。此角色授予对该位置的所需读/写权限。您也可以使用一个自定义角色注册位置，前提是该角色符合the section called “用于注册位置的角色的要求”中的要求。

> **Important**
>
> 如果您使用 AWS 托管式密钥 (aws/s3) 对 Amazon S3 位置进行加密，则无法使用 Lake Formation 服务相关角色。您必须使用一个自定义角色，并向该角色添加对密钥的 IAM 权限。本部分后面将提供相关详细信息。

以下过程说明如何注册使用客户自主管理型密钥或 AWS 托管式密钥加密的 Amazon S3 位置。

- 注册使用客户自主管理型密钥加密的位置
- 注册使用加密的地点 AWS 托管式密钥

开始前的准备工作

查看用于注册位置的角色的要求。

注册使用客户自主管理型密钥加密的 Amazon S3 位置

> **Note**
>
> 如果 KMS 密钥或 Amazon S3 位置与数据目录不在同一个 AWS 账户中，请the section called “跨 AWS 账户注册加密的 Amazon S3 位置”改为按照中的说明进行操作。

1. 打开 AWS KMS 控制台 https://console.aws.amazon.com/kms 并以 AWS Identity and Access Management (IAM) 管理用户或可以修改用于加密位置的 KMS 密钥策略的用户身份登录。
2. 在导航窗格中，选择客户自主管理型密钥，然后选择所需的 KMS 密钥的名称。
3. 在 KMS 密钥详细信息页面上，选择密钥策略选项卡，然后执行以下任一操作将您的自定义角色或 Lake Formation 服务相关角色添加为 KMS 密钥用户：

   - 如果显示默认视图（包括“密钥管理员”、“密钥删除”、“密钥用户”和“其他 AWS 账户”部分），则在“密钥用户”部分下，添加您的自定义角色或 Lake Formation 服务相关角色`AWSServiceRoleForLakeFormationDataAccess`。
   - 如果显示密钥策略 (JSON) – 编辑策略以将您的自定义角色或 Lake Formation 服务相关角色 `AWSServiceRoleForLakeFormationDataAccess` 添加到“Allow use of the key”对象，如以下示例所示。

   > **Note**
   >
   > 如果缺少该对象，请使用示例中显示的权限添加该对象。该示例使用服务相关角色。

```
        ...
        {
            "Sid": "Allow use of the key",
            "Effect": "Allow",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess",
                    "arn:aws:iam::111122223333:user/keyuser"
                ]
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "*"
        },
        ...
```

4. 打开 AWS Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。以数据湖管理员或具有 `lakeformation:RegisterResource` IAM 权限的用户身份登录。
5. 在导航窗格的注册和提取下，选择数据湖位置。
6. 选择注册位置，然后选择浏览以选择 Amazon Simple Storage Service (Amazon S3) 路径。
7. （可选，但强烈推荐）选择查看位置权限以查看所选 Amazon S3 位置中所有现有资源及其权限的列表。

   注册所选位置可能会导致 Lake Formation 用户可以访问该位置已有的数据。查看此列表有助于确保现有数据保持安全。
8. 对于 IAM 角色，选择 `AWSServiceRoleForLakeFormationDataAccess` 服务相关角色（默认）或符合the section called “用于注册位置的角色的要求”的自定义角色。
9. 选择注册位置。

有关服务相关角色的更多信息，请参阅 Lake Formation 的服务相关角色权限。

注册使用加密的 Amazon S3 地点 AWS 托管式密钥

> **Important**
>
> 如果 Amazon S3 位置与数据目录不在同一个 AWS 账户中，请the section called “跨 AWS 账户注册加密的 Amazon S3 位置”改为按照中的说明进行操作。

1. 创建用于注册位置的 IAM 角色。确保该角色符合the section called “用于注册位置的角色的要求”中列出的要求。
2. 将下面的内联策略附加到该角色。该策略会向该角色授予对密钥的权限。`Resource` 规范必须指定 AWS 托管式密钥的 Amazon 资源名称 (ARN)。您可以从控制台获取 ARN。 AWS KMS 要获得正确的 ARN，请确保使用与加密该位置相同的 AWS 账户和区域登录 AWS KMS 控制台。 AWS 托管式密钥

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
```

```
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "<AWS ##### ARN>"
    }
  ]
}
```

3. 打开 AWS Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。以数据湖管理员或具有 `lakeformation:RegisterResource` IAM 权限的用户身份登录。
4. 在导航窗格的注册和提取下，选择数据湖位置。
5. 选择注册位置，然后选择浏览以选择 Amazon S3 路径。
6. （可选，但强烈推荐）选择查看位置权限以查看所选 Amazon S3 位置中所有现有资源及其权限的列表。

   注册所选位置可能会导致 Lake Formation 用户可以访问该位置已有的数据。查看此列表有助于确保现有数据保持安全。
7. 对于 IAM 角色，选择您在步骤 1 中创建的角色。
8. 选择注册位置。

在其他 AWS 账户中注册 Amazon S3 位置

AWS Lake Formation 允许您跨账户注册亚马逊简单存储服务 (Amazon S3) Service 地点 AWS 。例如，如果在账户 A 中， AWS Glue Data Catalog 则账户 A 中的用户可以在账户 B 中注册 Amazon S3 存储桶。

使用账户 A 中的 AWS Identity and Access Management (IAM) 角色在 AWS 账户 B 中 AWS 注册 Amazon S3 存储桶需要以下权限：

- 账户 A 中的角色必须授予对账户 B 中存储桶的权限。
- 账户 B 中的存储桶策略必须向账户 A 中的角色授予访问权限。

> **Important**
>
> 避免注册启用了请求者付费的 Amazon S3 存储桶。对于在 Lake Formation 中注册的存储桶，用于注册存储桶的角色始终被视为请求者。如果存储桶被其他 AWS 账户访问，则如果该角色与存储桶拥有者属于同一个账户，则该存储桶拥有者需要支付数据访问费用。
> 您不能使用 Lake Formation 服务相关角色在其他账户中注册位置。您必须改用用户定义的角色。该角色必须符合the section called “用于注册位置的角色的要求”中的要求。有关服务相关角色的更多信息，请参阅 Lake Formation 的服务相关角色权限。

开始前的准备工作

查看用于注册位置的角色的要求。

在其他 AWS 账户中注册营业地点

> **Note**
>
> 如果该位置已加密，请改为按照the section called “跨 AWS 账户注册加密的 Amazon S3 位置”中的说明进行操作。

以下过程假定包含数据目录的账户 1111-2222-3333 中的主体想要注册位于账户 1234-5678-9012 中的 Amazon S3 存储桶 `awsexamplebucket1`。

1. 在账户 1111-2222-3333 中，登录 AWS Management Console 并打开 IAM 控制台，网址为。https://console.aws.amazon.com/iam/
2. 创建新角色或查看符合the section called “用于注册位置的角色的要求”中要求的现有角色。确保该角色授予对 `awsexamplebucket1` 的 Amazon S3 权限。
3. 打开 Amazon S3 控制台，网址为：https://console.aws.amazon.com/s3/。使用账户 1234-5678-9012 登录。
4. 在存储桶名称列表中，选择存储桶名称 `awsexamplebucket1`。
5. 选择权限。
6. 在权限页面上，选择存储桶策略。
7. 在存储桶策略编辑器中，粘贴以下策略。将 *`<role-name>`* 替换为您的角色的名称。

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::awsexamplebucket1"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::awsexamplebucket1/*"
        }
    ]
}
```

8. 选择保存。
9. 打开 AWS Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。以数据湖管理员或具有足够权限注册位置的用户身份登录账户 1111-2222-3333。
10. 在导航窗格的管理下，选择数据湖位置。
11. 在数据湖位置页面上，选择注册位置。
12. 在注册位置页面上，对于 Amazon S3 路径，输入存储桶名称 `s3://awsexamplebucket1`。

> **Note**
>
> 您必须键入存储桶名称，因为当您选择浏览时，跨账户存储桶不会显示在列表中。

13. 对于 IAM 角色，选择您的角色。
14. 选择注册位置。

跨 AWS 账户注册加密的 Amazon S3 位置

AWS Lake Formation 与 AWS Key Management Service(AWS KMS) 集成，使您能够更轻松地设置其他集成服务，以加密和解密亚马逊简单存储服务 (Amazon S3) 中的数据。

同时支持客户管理 AWS 托管式密钥 的密钥和。不支持客户端加密/解密。

> **Important**
>
> 避免注册启用了请求者付费的 Amazon S3 存储桶。对于在 Lake Formation 中注册的存储桶，用于注册存储桶的角色始终被视为请求者。如果存储桶被其他 AWS 账户访问，则如果该角色与存储桶拥有者属于同一个账户，则该存储桶拥有者需要支付数据访问费用。

本部分介绍如何在以下情况下注册 Amazon S3 位置：

- Amazon S3 位置中的数据使用在 AWS KMS中创建的 KMS 密钥进行加密。
- Amazon S3 的营业地点与不在同一个 AWS 账户中 AWS Glue Data Catalog。
- KMS 密钥与数据目录位于或不在同一个 AWS 账户中。

使用账户 A 中的 AWS Identity and Access Management (IAM) 角色在 AWS 账户 B 中 AWS 注册 AWS KMS加密的 Amazon S3 存储桶需要以下权限：

- 账户 A 中的角色必须授予对账户 B 中存储桶的权限。
- 账户 B 中的存储桶策略必须向账户 A 中的角色授予访问权限。
- 如果 KMS 密钥位于账户 B 中，则密钥策略必须向账户 A 中的角色授予访问权限，并且账户 A 中的角色必须授予对 KMS 密钥的权限。

在以下步骤中，您将在包含数据目录的 AWS 账户（前面讨论中的账户 A）中创建一个角色。然后，使用此角色注册位置。Lake Formation 在访问 Amazon S3 中的基础数据时代入此角色。代入的角色具有对 KMS 密钥的所需权限。因此，您不必向使用 ETL 作业或集成服务（如 Amazon Athena）访问基础数据的主体授对 KMS 密钥的权限。

> **Important**
>
> 您不能使用 Lake Formation 服务相关角色在其他账户中注册位置。您必须改用用户定义的角色。该角色必须符合the section called “用于注册位置的角色的要求”中的要求。有关服务相关角色的更多信息，请参阅 Lake Formation 的服务相关角色权限。

开始前的准备工作

查看用于注册位置的角色的要求。

跨 AWS 账户注册加密的 Amazon S3 营业地点

1. 使用与数据目录相同的 AWS 账户，登录 AWS Management Console 并打开 IAM 控制台，网址为https://console.aws.amazon.com/iam/。
2. 创建新角色或查看符合the section called “用于注册位置的角色的要求”中要求的现有角色。确保该角色包含授予对该位置的 Amazon S3 权限的策略。
3. 如果 KMS 密钥与数据目录不在同一账户中，请向该角色附加一个内联策略，该策略授予对 KMS 密钥的所需权限。以下是示例策略。将<cmk-region>和 *< cmk-account-id >* 替换为 KMS 密钥的区域和账号。将 *<key-id>* 替换为密钥 ID。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
        "Effect": "Allow",
        "Action": [
            "kms:Encrypt",
            "kms:Decrypt",
            "kms:ReEncrypt*",
            "kms:GenerateDataKey*",
            "kms:DescribeKey"
         ],
        "Resource": "arn:aws:kms:<cmk-region>:<cmk-account-id>:key/<key-id>"
        }
    ]
}
```

4. 在 Amazon S3 控制台上，添加存储桶策略，该策略向该角色授予所需的 Amazon S3 权限。下面是一个示例存储桶策略。将 *< catalog-account-id >* 替换为数据目录的 AWS 账号、<role-name>您的角色<bucket-name>名称和存储桶的名称。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::<catalog-account-id>:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::<bucket-name>"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::<catalog-account-id>:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::<bucket-name>/*"
        }
    ]
}
```

5. 在中 AWS KMS，将角色添加为 KMS 密钥的用户。

   a. 打开 AWS KMS 控制台，网址为 https://console.aws.amazon.com/kms。然后，以管理员用户或可以修改用于加密位置的 KMS 密钥策略的用户身份登录。

   b. 在导航窗格中，选择客户自主管理型密钥，然后选择 KMS 密钥的名称。

   c. 在“KMS 密钥详细信息”页面的密钥策略选项卡下，如果未显示密钥策略的 JSON 视图，请选择切换到策略视图。

   d. 在密钥策略部分中，选择编辑，然后将该角色的 Amazon 资源名称 (ARN) 添加到 `Allow use of the key` 对象，如以下示例所示。

> **Note**
>
> 如果缺少该对象，请使用示例中显示的权限添加该对象。

```
...
{
    "Sid": "Allow use of the key",
    "Effect": "Allow",
    "Principal": {
        "AWS": [
            "arn:aws:iam::<catalog-account-id>:role/<role-name>"
        ]
    },
    "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
    ],
    "Resource": "*"
},
...
```

有关更多信息，请参阅《AWS Key Management Service 开发人员指南》中的允许其他账户中的用户使用 KMS 密钥。

6. 打开 AWS Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。以数据湖管理员身份登录数据目录 AWS 账户。
7. 在导航窗格的注册和提取下，选择数据湖位置。
8. 选择注册位置。
9. 在注册位置页面上，对于 Amazon S3 路径，输入位置路径 **s3://*<bucket>*/*<prefix>***。将 *<bucket>* 替换为存储桶的名称，并将 *<prefix>* 替换为该位置路径的其余部分。

> **Note**
>
> 您必须键入该路径，因为当您选择浏览时，跨账户存储桶不会显示在列表中。

10. 对于 IAM 角色，从步骤 2 中选择角色。
11. 选择注册位置。

取消注册 Amazon S3 位置

如果您不再希望 Amazon Simple Storage Service (Amazon S3) 位置由 Lake Formation 管理，则可以取消注册该位置。取消注册某个位置不会影响对该位置授予的 Lake Formation 数据位置权限。您可以重新注册已取消注册的位置，并且数据位置权限仍然有效。您可以使用其他角色重新注册该位置。

取消注册位置（控制台）

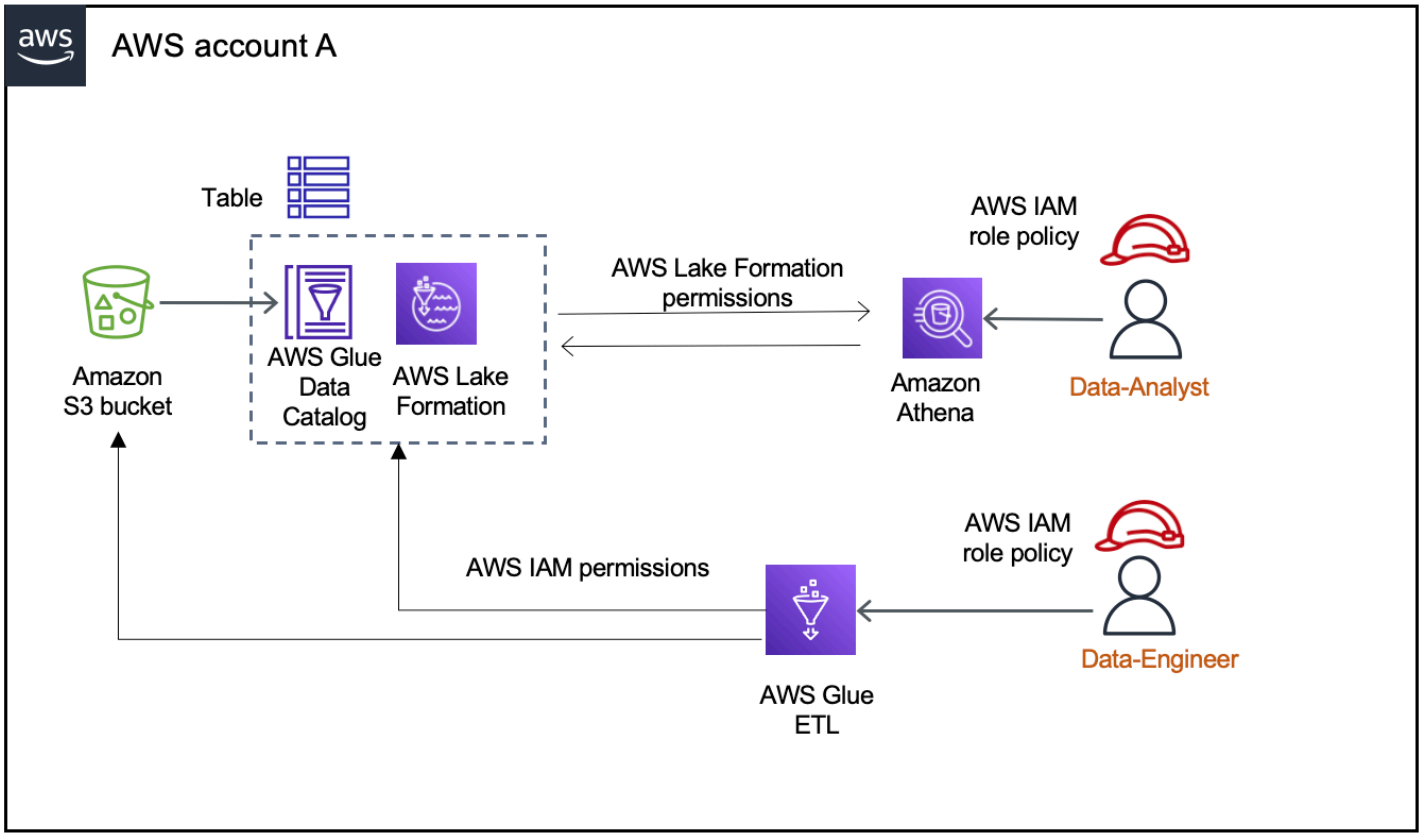
1. 打开 AWS Lake Formation 控制台，网址为 https://console.aws.amazon.com/lakeformation/。以数据湖管理员或具有 `lakeformation:RegisterResource` IAM 权限的用户身份登录。
2. 在导航窗格的注册和提取下，选择数据湖位置。
3. 选择一个位置，然后在操作菜单上选择删除。
4. 当系统提示进行确认时，选择删除。

混合访问模式

AWS Lake Formation 混合访问模式支持访问相同 AWS Glue Data Catalog 数据库和表的两种权限路径。
在第一种途径中，Lake Formation 允许您选择特定的主体，并通过选择加入向他们授予 Lake Formation 访问数据库和表格的权限。第二种途径允许所有其他委托人通过 Amazon S3 的默认 IAM 委托人策略和 AWS Glue 操作访问这些资源。

在 Lake Formation 中注册 Amazon S3 位置时，您可以选择对该位置的所有资源强制实施 Lake Formation 权限，也可以选择使用混合访问模式。默认情况下，混合访问模式仅强制实施 `CREATE_TABLE`、`CREATE_PARTITION`、`UPDATE_TABLE` 权限。当 Amazon S3 位置处于混合模式时，您可以通过为该位置下的数据库和表选择主体来启用 Lake Formation 权限。

因此，混合访问模式使您可以灵活地且有选择性地为一组特定用户针对数据目录中的数据库和表启用 Lake Formation，而不会中断其他现有用户或工作负载的访问。



有关注意事项和限制，请参阅混合访问模式注意事项和限制 。

术语和定义

以下是基于您对访问权限的设置方式的数据目录资源定义：

Lake Formation 资源

已向 Lake Formation 注册的资源。用户需要 Lake Formation 权限才能访问该资源。

AWS Glue 资源

未向 Lake Formation 注册的资源。用户只需 IAM 权限即可访问该资源，因为它具有 `IAMAllowedPrincipals` 组权限。Lake Formation 权限不会被强制实施。

有关 `IAMAllowedPrincipals` 组权限的更多信息，请参阅元数据权限。

混合资源

在混合访问模式下注册的资源。根据访问资源的用户，该资源会在充当 Lake Formation 资源或 AWS Glue 资源之间动态切换。

常见的混合访问模式使用案例

您可以在单账户和跨账户数据共享场景中使用混合访问模式提供访问权限：

单账户场景

- 将@@ AWS Glue 资源转换为混合资源-在这种情况下，您当前未使用 Lake Formation，但希望对数据目录数据库和表采用 Lake Formation 权限。当您在混合访问模式下注册 Amazon S3 位置时，您可以向选择使用指向该位置的特定数据库和表的用户授予 Lake Formation 权限。
- 将 Lake Formation 资源转换为混合资源 — 目前，您正在使用 Lake Formation 权限来控制对数据目录数据库的访问权限，但希望在 AWS Glue 不中断现有的 Lake Formation 权限的情况下使用适用于 Amazon S3 的 IAM 权限向新委托人提供访问权限。

  当您将数据位置注册更新为混合访问模式时，新的主体可以使用 IAM 权限策略访问指向 Amazon S3 位置的数据目录数据库，而不会中断现有用户的 Lake Formation 权限。

  在更新数据位置注册以启用混合访问模式之前，您需要先选择当前使用 Lake Formation 权限访问资源的主体。
  这是为了防止当前工作流可能出现中断。
  您还需要向 `IAMAllowedPrincipal` 组授予对数据库中表的 `Super` 权限。

跨账户数据共享场景

- 使用混合访问模式共享 AWS Glue 资源-在这种情况下，创建者账户在数据库中拥有表，这些表当前使用针对 Amazon S3 和 AWS Glue 操作的 IAM 权限策略与消费者账户共享。数据库的数据位置未在 Lake Formation 中注册。

  在混合访问模式下注册数据位置之前，您需要将跨账户版本设置更新为版本 4。版本 4 提供了`IAMAllowedPrincipal`群组拥有资源 AWS RAM 权限时跨账户共享所需的新`Super`权限策略。对于那些具有 `IAMAllowedPrincipal` 组权限的资源，您可以向外部账户授予 Lake Formation 权限，并选择他们以使他们可以使用 Lake Formation 权限。接收方账户中的数据湖管理员可以向账户中的主体授予 Lake Formation 权限，并选择他们以强制实施 Lake Formation 权限。
- 使用混合访问模式共享 Lake Formation 资源 — 当前，制作者账户拥有与强制实施 Lake Formation 权限的使用者账户共享的数据库中的表。数据库的数据位置在 Lake Formation 中注册。

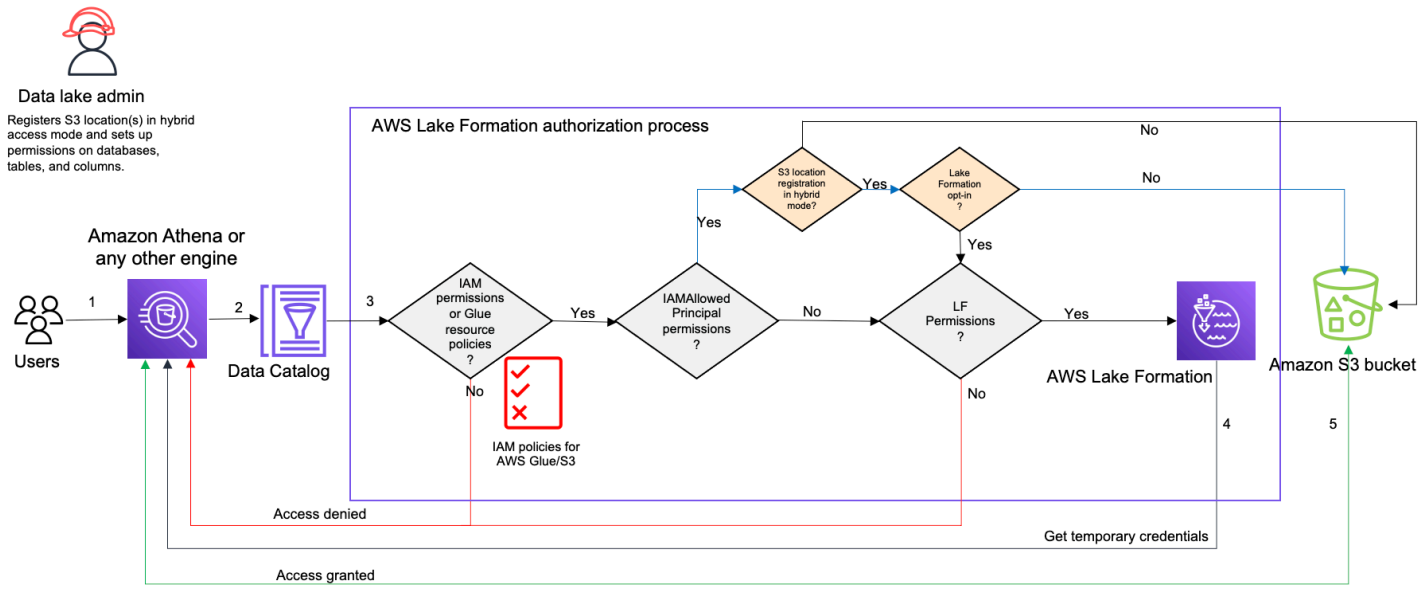
  在这种情况下，您可以将 Amazon S3 位置注册更新为混合访问模式，并针对使用者账户中的主体使用 Amazon S3 存储桶策略和数据目录资源策略来共享来自 Amazon S3 的数据和来自数据目录的

元数据。在更新 Amazon S3 位置注册之前，您需要重新授予现有的 Lake Formation 权限并选择主体。此外，您还需要向 `IAMAllowedPrincipals` 组授予对数据库中表的 `Super` 权限。

主题

混合访问模式的工作原理

下图显示了查询数据目录资源时 Lake Formation 授权在混合访问模式下的工作原理。



在访问数据湖中的数据之前，数据湖管理员或具有管理权限的用户会设置单个数据目录表用户策略以允许或拒绝访问数据目录中的表。然后，有权执行 `RegisterResource` 操作的主体在混合访问模式下向 Lake Formation 中注册表的 Amazon S3 位置。管理员向特定用户授予对数据目录数据库和表的 Lake Formation 权限，并选择他们以使他们能够在混合访问模式下对这些数据库和表使用 Lake Formation 权限。

1. 提交查询-委托人使用集成服务（例如亚马逊 Athena、Amazon EMR 或 AWS Glue Amazon Redshift Spectrum）提交查询或 ETL 脚本。

2. 请求数据 - 集成分析引擎可识别正在请求的表，并向数据目录发送元数据请求（`GetTable`、`GetDatabase`）。
3. 检查权限 - 数据目录可通过 Lake Formation 验证查询主体的访问权限。
   a. 如果该表未附加 `IAMAllowedPrincipals` 组权限，则会强制实施 Lake Formation 权限。
   b. 如果主体选择在混合访问模式下使用 Lake Formation 权限，并且该表附加了 `IAMAllowedPrincipals` 组权限，则会强制实施 Lake Formation 权限。查询引擎应用从 Lake Formation 接收的筛选条件，并将数据返回给用户。
   c. 如果表位置未在 Lake Formation 中注册，并且主体未选择在混合访问模式下使用 Lake Formation 权限，则数据目录将检查该表是否附加了 `IAMAllowedPrincipals` 组权限。如果针对该表存在此权限，则该账户中的所有主体均将获得对该表的 `Super` 或 `All` 权限。
4. 获取凭证 - 数据目录会检查表位置是否已在 Lake Formation 中注册并将结果告知引擎。如果基础数据已在 Lake Formation 中注册，则分析引擎会请求 Lake Formation 提供临时凭证，以访问 Amazon S3 存储桶中的数据。
5. 获取数据 - 如果主体有权访问表数据，Lake Formation 将提供对集成分析引擎的临时访问权限。通过使用临时访问权限，分析引擎可从 Amazon S3 获取数据，并执行必要的筛选，例如列、行或单元格筛选。当引擎运行完作业后，它会将结果返回给用户。此过程称为凭证售卖。有关更多信息，请参阅与 Lake Formation 集成。
6. 如果表的数据位置未在 Lake Formation 中注册，则分析引擎将直接向 Amazon S3 进行第二次调用。系统会对相关 Amazon S3 存储桶策略和 IAM 用户策略进行评估以确定是否支持访问数据。每当您使用 IAM policy 时，请确保遵循 IAM 最佳实践。有关更多信息，请参阅《IAM 用户指南》中的 IAM 中的安全最佳实践。

设置混合访问模式 - 常见场景

与 Lake Formation 权限一样，您通常有两种类型的场景可以使用混合访问模式来管理数据访问权限：在一个场景中提供对委托人的访问权限， AWS 账户 以及提供对外部 AWS 账户 或委托人的访问权限。

本节提供有关在以下场景下设置混合访问模式的说明：

在混合访问模式下在一个模式下管理权限 AWS 账户

- 将 AWS Glue 资源转换为混合资源 — 您目前正在使用 Amazon S3 的 IAM 权限为账户中的所有委托人提供数据库中表的访问权限， AWS Glue 但希望采用 Lake Formation 来逐步管理权限。

- 将 Lake Formation 资源转换为混合资源 — 您目前正在使用 Lake Formation 来管理账户中所有委托人对数据库中表的访问权限，但只想对特定的委托人使用 Lake Formation。您想通过对同一数据库和表使用 IAM 权限 AWS Glue 和 Amazon S3 来提供对新委托人的访问权限。

在混合访问模式下管理跨 AWS 账户的权限

- 使用混合访问模式共享 AWS Glue 资源— 你目前没有使用 Lake Formation 来管理表格的权限，但想应用 Lake Formation 权限为其他账户中的委托人提供访问权限。
- 使用混合访问模式共享 Lake Formation 资源— 您正在使用 Lake Formation 管理表的访问权限，但希望通过对同一数据库 AWS Glue 和表使用 Amazon S3 的 IAM 权限为其他账户中的委托人提供访问权限。

设置混合访问模式 – 主要步骤

1. 通过选择混合访问模式，在 Lake Formation 中注册 Amazon S3 数据位置。
2. 主体必须拥有对数据湖位置的 DATA_LOCATION 权限才能创建指向该位置的数据目录表或数据库。
3. 将跨账户版本设置设置为版本 4。
4. 向特定 IAM 用户或角色授予对数据库和表的精细权限。同时，确保为 IAMAllowedPrincipals 组设置对数据库以及数据库中所有或部分表的 Super 或 All 权限。
5. 选择主体和资源。账户中的其他委托人可以使用针对和 Amazon S3 操作的 IAM 权限策略继续访问数据库 AWS Glue 和表。
6. （可选）为选择使用 Lake Formation 权限的主体清除适用于 Amazon S3 的 IAM 权限策略。

设置混合访问模式的先决条件

以下是设置混合访问模式的先决条件：

> **Note**
>
> 我们建议 Lake Formation 管理员在混合访问模式下注册 Amazon S3 位置，并选择主体和资源。

1. 授予数据位置权限 (DATA_LOCATION_ACCESS)，以创建指向 Amazon S3 位置的数据目录资源。数据位置权限可控制创建指向特定 Amazon S3 位置的数据目录数据库和表的能力。

2. 要在混合访问模式下与其他账户共享数据目录资源（无需从资源中删除 IAMAllowedPrincipals 组权限），您需要将跨账户版本设置更新为版本 4。要使用 Lake Formation 控制台更新版本，请在数据目录设置页面上的跨账户版本设置下选择版本 4。

   您也可以使用put-data-lake-settings AWS CLI 命令将CROSS_ACCOUNT_VERSION参数设置为版本 4：

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
"DataLakeAdmins": [
        {
"DataLakePrincipalIdentifier": "arn:aws:iam::<111122223333>:user/<user-name>"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
"CROSS_ACCOUNT_VERSION": "4"
    }
}
```

3. 要在混合访问模式下授予跨账户权限，授予者必须拥有 AWS Glue 和 AWS RAM 服务所需的 IAM 权限。 AWS 托管策略AWSLakeFormationCrossAccountManager授予所需的权限。为了在混合访问模式下启用跨账户数据共享，我们通过添加两项新的 IAM 权限更新了 AWSLakeFormationCrossAccountManager 托管策略：

   - 内存:ListResourceSharePermissions
   - 内存:AssociateResourceSharePermission

> **Note**
>
> 如果您未使用授予者角色的 AWS 托管策略，请将上述策略添加到您的自定义策略中。

将 AWS Glue 资源转换为混合资源

按照以下步骤在混合访问模式下注册 Amazon S3 位置，并在不中断现有数据目录用户数据访问的情况下引导新的 Lake Formation 用户。

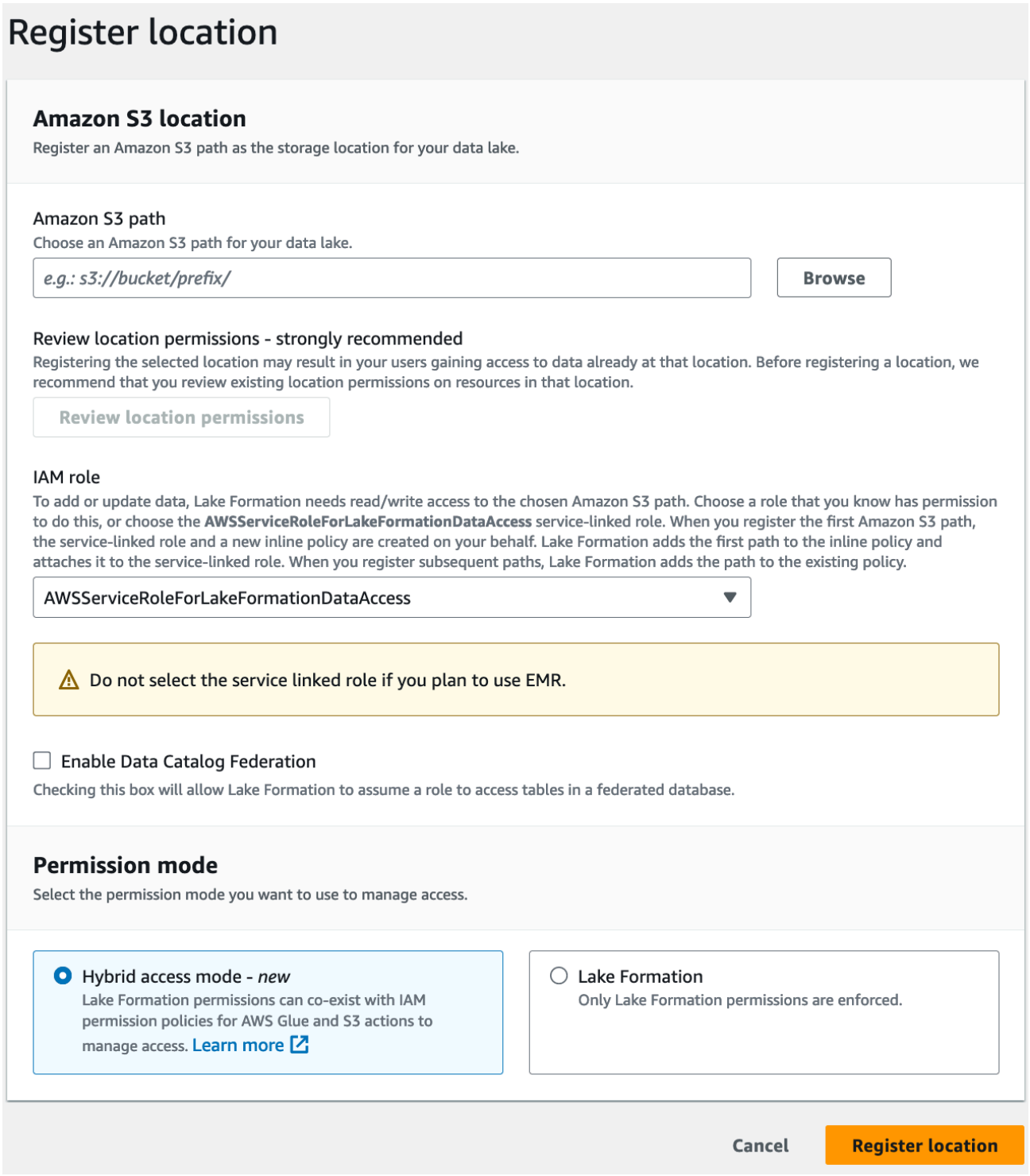
场景描述 - 数据位置未在 Lake Formation 中注册，用户对数据目录数据库和表的访问权限由适用于 Amazon S3 和 AWS Glue 操作的 IAM 权限策略决定。
默认情况下，`IAMAllowedPrincipals` 组拥有对数据库中所有表的 `Super` 权限。

为未在 Lake Formation 中注册的数据位置启用混合访问模式

1. 注册 Amazon S3 位置，以启用混合访问模式。

   Console

   1. 以数据湖管理员的身份登录 Lake Formation 控制台。
   2. 在导航窗格中，选择管理下的数据湖位置。
   3. 选择注册位置。



4. 在注册位置窗口中，选择要在 Lake Formation 中注册的 Amazon S3 路径。
5. 对于 IAM 角色，选择 `AWSServiceRoleForLakeFormationDataAccess` 服务相关角色（默认）或符合用于注册位置的角色的要求中要求的自定义 IAM 角色。
6. 选择混合访问模式以应用精细的 Lake Formation 访问控制策略来选择主体以及指向注册位置的数据目录数据库和表。

选择 Lake Formation 以允许 Lake Formation 授权对注册位置的访问请求。

7. 选择注册位置。

AWS CLI

以下是使用:true/false 向 Lake Formation 注册数据位置的 HybridAccessEnabled示例。HybridAccessEnabled 参数的默认值为 false。将 Amazon S3 路径、角色名称和 AWS 账户 ID 替换为有效值。

```
aws lakeformation register-resource --cli-input-json file:file path
json:
    {
        "ResourceArn": "arn:aws:s3:::s3-path",
        "UseServiceLinkedRole": false,
        "RoleArn": "arn:aws:iam::<123456789012>:role/<role-name>",
        "HybridAccessEnabled": true
    }
```

2. 授予权限并选择主体以在混合访问模式下对资源使用 Lake Formation 权限

在混合访问模式下选择主体和资源之前，请确认在混合访问模式下向 IAMAllowedPrincipals 组授予了对在 Lake Formation 中注册的位置处的数据库和表的 Super 或 All 权限。

> **Note**
>
> 您不能向 IAMAllowedPrincipals 组授予对数据库中 All tables 的权限。您需要从下拉菜单中分别选择每个表并授予权限。此外，在数据库中创建新表时，可以选择使用数据目录设置中的。Use only IAM access control for new tables in new databases当您在数据库中创建新表时，此选项会自动向 IAMAllowedPrincipals 组授予 Super 权限。

Console

1. 在 Lake Formation 控制台的数据目录下，选择数据库或表。
2. 从列表中选择数据库或表，然后从 “操作” 菜单中选择 “授权”。
3. 选择要使用命名资源方法或 LF 标签授予其对数据库、表和列的权限的主体。

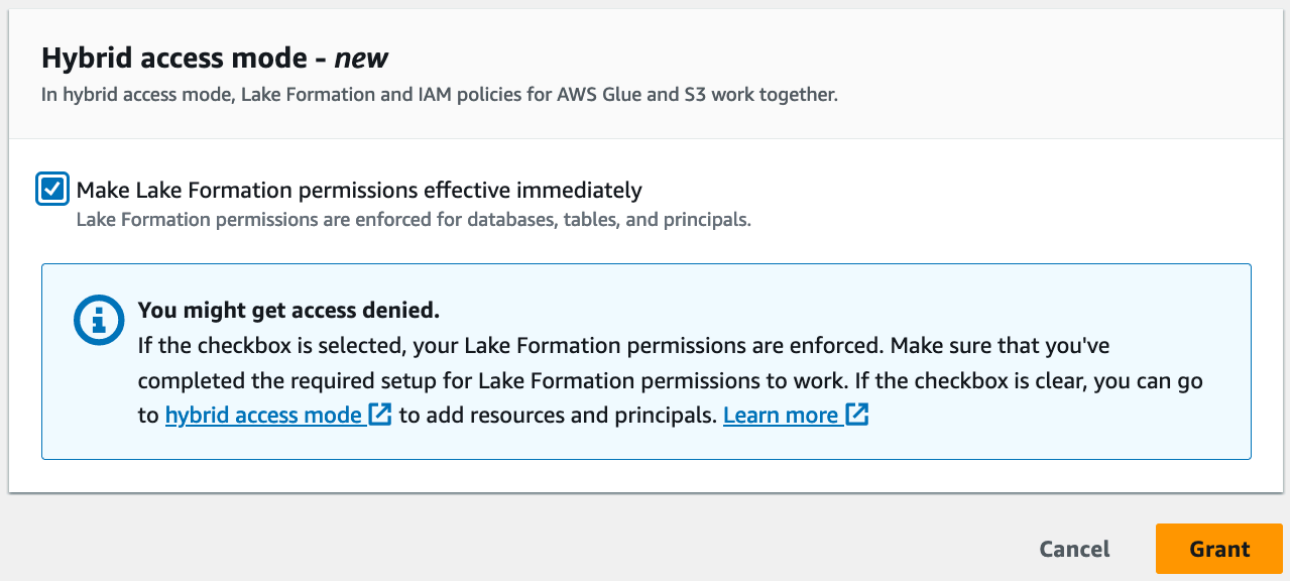
或者，选择数据湖权限，从列表中选择要向其授予权限的主体，然后选择授权。

有关授予数据权限的更多详细信息，请参阅授予和撤销对数据目录资源的权限。

> **Note**
>
> 如果您要向主体授予创建表的权限，则还需要向主体授予数据位置权限 (DATA_LOCATION_ACCESS)。更新表不需要此权限。
> 有关更多信息，请参阅授予数据位置权限。

4. 当您使用命名资源方法授予权限时，授予数据权限页面的下半部分提供了用于选择主体和资源的选项。

   选择让 Lake Formation 权限立即生效，为主体和资源启用 Lake Formation 权限。



5. 选择授权。

   当您在指向数据位置的表 A 上选择主体 A 时，如果数据位置是在混合模式下注册的，则为主体 A 提供对该表的位置的 Lake Formation 权限。

AWS CLI

以下是在混合访问模式下选择主体和表的示例。将角色名、 AWS 账户 ID、数据库名称和表名替换为有效值。

```
aws lakeformation create-lake-formation-opt-in --cli-input-json file://file path
```

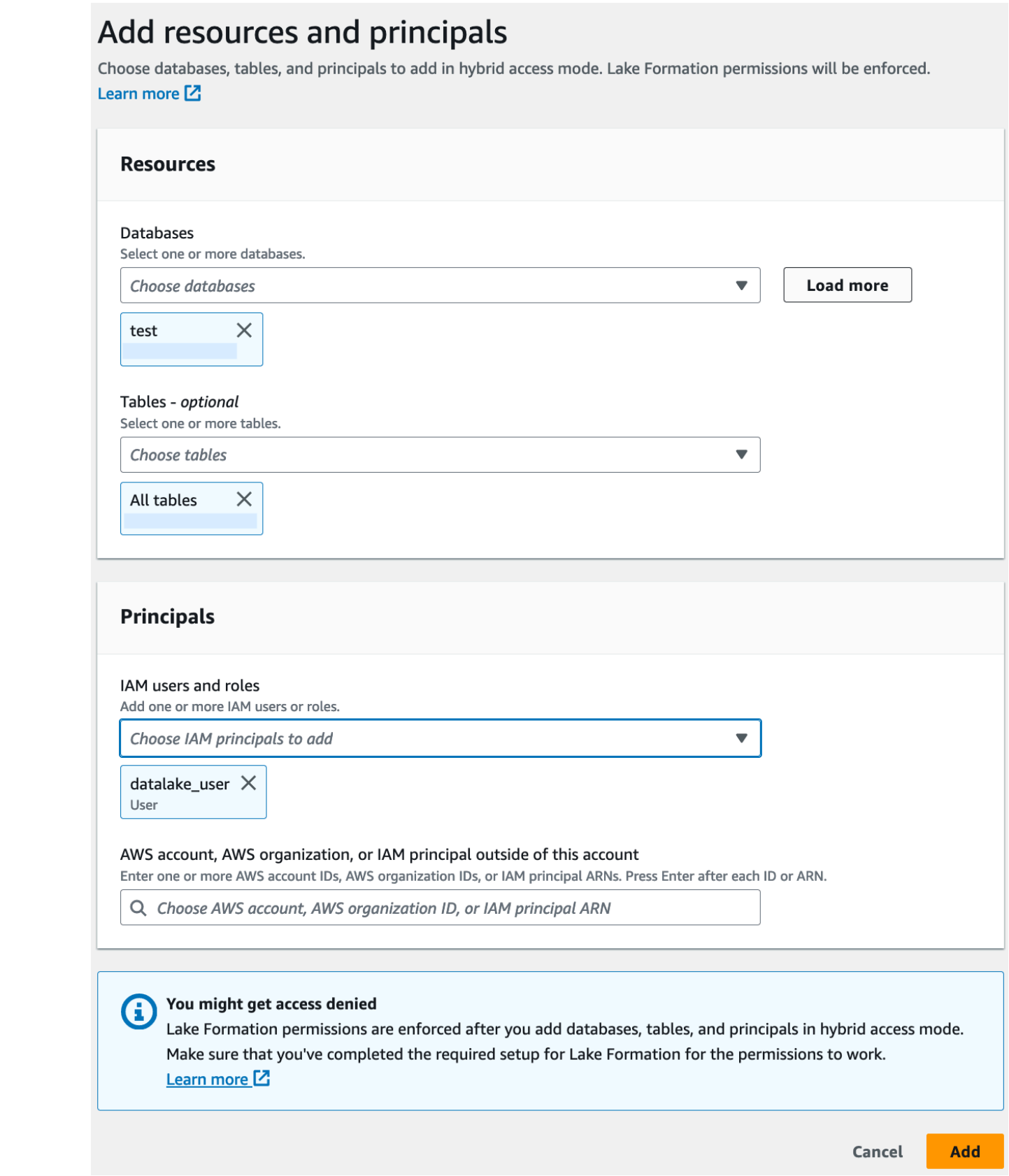
```
json:
  {
        "Principal": {
            "DataLakePrincipalIdentifier":
 "arn:aws:iam::<123456789012>:role/<hybrid-access-role>"
        },
        "Resource": {
            "Table": {
                "CatalogId": "<123456789012>",
                "DatabaseName": "<hybrid_test>",
                "Name": "<hybrid_test_table>"
            }
        }
    }
```

a. (Optional) 如果您选择 LF 标签来授予权限，则可以在单独的步骤中选择主体来使用 Lake Formation 权限。您可以通过在左侧导航栏的权限下选择混合访问模式来执行此操作。

b. 在混合访问模式页面的下半部分，选择添加将资源和主体添加到混合访问模式下。

c. 在添加资源和主体页面上，选择在混合访问模式下注册的数据库和表。选择要选择在混合访问模式下使用 Lake Formation 权限的主体。

   您可以选择数据库下的 `All tables` 以授予主体访问权限。


Add resources and principals
Choose databases, tables, and principals to add in hybrid access mode. Lake Formation permissions will be enforced.
Learn more
Resources
Databases
Select one or more databases.
Choose databases
Load more
test
Tables - optional
Select one or more tables.
Choose tables
All tables
Principals
IAM users and roles
Add one or more IAM users or roles.
Choose IAM principals to add
datalake_user
User
AWS account, AWS organization, or IAM principal outside of this account
Enter one or more AWS account IDs, AWS organization IDs, or IAM principal ARNs. Press Enter after each ID or ARN.
Choose AWS account, AWS organization ID, or IAM principal ARN
You might get access denied
Lake Formation permissions are enforced after you add databases, tables, and principals in hybrid access mode.
Make sure that you've completed the required setup for Lake Formation for the permissions to work.
Learn more
Cancel
Add

将 Lake Formation 资源转换为混合资源

如果您当前正对数据目录数据库和表使用 Lake Formation 权限，则可以编辑位置注册属性以启用混合访问模式。这样，您就可以在不中断现有 Lake Formation 权限的情况下使用适用于 Amazon S3 的 IAM 权限策略和 AWS Glue 操作为新的委托人提供对相同资源的访问权限。

场景描述 - 以下步骤假定您已在 Lake Formation 中注册了一个数据位置，并且您已为主体设置了对指向该位置的数据库、表或列的权限。如果该位置是通过服务相关角色注册的，则无法更新位置参数和启用混合访问模式。默认情况下，`IAMAllowedPrincipals` 组对数据库及其所有表拥有 Super 权限。

> **Important**
>
> 如果不选择访问该位置数据的委托人，请勿将位置注册更新为混合访问模式。

为在 Lake Formation 中注册的数据位置启用混合访问模式

1.
   > **Warning**
   >
   > 我们不建议将 Lake Formation 托管数据位置转换为混合访问模式，以免中断其他现有用户或工作负载的权限策略。

   选择拥有 Lake Formation 权限的现有主体。

   1. 列出并查看您授予主体的对数据库和表的权限。有关更多信息，请参阅[在 Lake Formation 中查看数据库和表权限]。
   2. 在左侧导航栏中的权限下选择混合访问模式，然后选择添加。
   3. 在添加主体和资源页面上，从 Amazon S3 数据位置中选择要在混合访问模式下使用的数据库和表。选择已经拥有 Lake Formation 权限的主体。
   4. 选择添加以选择要在混合访问模式下使用 Lake Formation 权限的主体。
2. 通过选择混合访问模式选项来更新 Amazon S3 存储桶/前缀注册。

   Console

   1. 以数据湖管理员身份登录 Lake Formation 控制台。
   2. 在导航窗格中的注册和提取下，选择数据湖位置。
   3. 选择一个位置，然后在操作菜单上选择编辑。

4. 选择混合访问模式。
5. 选择保存。
6. 在数据目录下，选择数据库或表，然后向名为 `IAMAllowedPrincipals` 的虚拟组授予 `Super` 或 `All` 权限。
7. 确认在更新位置注册属性时，您的现有 Lake Formation 用户的访问没有中断。以 Lake Formation 主体身份登录 Athena 控制台，然后对指向更新后位置的表运行示例查询。

   同样，验证使用 IAM 权限策略访问数据库和表的 AWS Glue 用户的访问权限。

AWS CLI

以下是使用:true/false 向 Lake Formation 注册数据位置的 HybridAccessEnabled示例。`HybridAccessEnabled` 参数的默认值为 false。将 Amazon S3 路径、角色名称和 AWS 账户 ID 替换为有效值。

```
aws lakeformation update-resource --cli-input-json file://file path
json:
{
    "ResourceArn": "arn:aws:s3:::<s3-path>",
    "RoleArn": "arn:aws:iam::<123456789012>:role/<test>",
    "HybridAccessEnabled": true
}
```

使用混合访问模式共享 AWS Glue 资源

在不中断现有数据目录用户基于 IAM 的访问的情况下，与 AWS 账户 执行 Lake Formation 权限的其他人 AWS 账户 或委托人共享数据。

场景描述-创建者账户有一个数据目录数据库，该数据库的访问权限使用适用于 Amazon S3 的 IAM 委托人策略和 AWS Glue 操作进行控制。数据库的数据位置未在 Lake Formation 中注册。默认情况下，该`IAMAllowedPrincipals`组拥有数据库及其所有表的`Super`权限。

在混合访问模式下授予跨账户 Lake Formation 权限

1. 制作者账户设置

   1. 使用具有 `lakeformation:PutDataLakeSettings` IAM 权限的角色登录 Lake Formation 控制台。
   2. 前往数据目录设置，然后在跨账户版本设置中选择`Version 4`。

      如果您当前使用的是版本 1 或 2，请参阅有关更新至版本 3 的更新跨账户数据共享版本设置说明。

      从版本 3 升级到 4 时，无需更改权限策略。
   3. 注册您计划在混合访问模式下共享的数据库或表的 Amazon S3 位置。
   4. 验证 `IAMAllowedPrincipals` 组对在上述步骤中在混合访问模式下注册其数据位置的数据库和表是否拥有 `Super` 权限。
   5. 向 AWS 组织、组织单位 (OU) 授予 Lake Formation 权限，或者直接向其他账户中的 IAM 委托人授予 Lake Formation 权限。
   6. 如果您要直接向 IAM 主体授予权限，请选择使用者账户中的主体，通过启用让 Lake Formation 权限立即生效选项，在混合访问模式下强制实施 Lake Formation 权限。

      如果您向其他账户授予跨账户权限，则当您选择加入该 AWS 账户时，Lake Formation 权限仅适用于该账户的管理员。接收方账户数据湖管理员需要向下级联权限并选择账户中的主体，才能对处于混合访问模式的共享资源强制实施 Lake Formation 权限。

      如果您选择通过 LF 标签匹配的资源选项来授予跨账户权限，则需要先完成权限授予步骤。您可以单独执行一步，通过在 Lake Formation 控制台左侧导航栏的“权限”下选择混合访问模式来选择要置于混合访问模式下的主体和资源。然后选择添加以添加资源和主体来强制实施 Lake Formation 权限。

2. 使用者账户设置

   1. 通过 https://console.aws.amazon.com/lakeformation/ 以数据湖管理员身份登录 Lake Formation 控制台。
   2. 前往 https://console.aws.amazon.com/ram 并接受资源共享邀请。 AWS RAM 控制台中的 “与我共享” 选项卡显示与您的账户共享的数据库和表。
   3. 在 Lake Formation 中创建指向共享数据库和/或表的资源链接。
   4. 向您（使用者）账户中的 IAM 主体授予对资源链接的 `Describe` 权限和对原共享资源的 `Grant on target` 权限。

5. 向账户中的主体授予对与您共享的数据库或表的 Lake Formation 权限。通过启用让 Lake Formation 权限立即生效选项，选择主体和资源以在混合访问模式下强制实施 Lake Formation 权限。
6. 通过运行示例 Athena 查询来测试主体的 Lake Formation 权限。使用 Amazon S3 的 IAM 委托人策略和 AWS Glue 操作测试 AWS Glue 用户的现有访问权限。

   （可选）为您配置为使用 Lake Formation 权限的主体删除关于数据访问的 Amazon S3 存储桶策略以及适用于 AWS Glue 和 Amazon S3 数据访问的 IAM 主体策略。

使用混合访问模式共享 Lake Formation 资源

允许外部账户中的新数据目录用户使用基于 IAM 的策略访问数据目录数据库和表，而无需中断现有的 Lake Formation 跨账户共享权限。

场景描述 - 制作者账户拥有在账户级别或 IAM 主体级别与外部（使用者）账户共享的 Lake Formation 托管数据库和表。数据库的数据位置在 Lake Formation 中注册。`IAMAllowedPrincipals` 组没有对数据库及其表的 `Super` 权限。

在不中断现有 Lake Formation 权限的情况下，通过基于 IAM 的策略向新的数据目录用户授予跨账户访问权限

1. 制作者账户设置

   1. 以拥有以下权限的角色登录 Lake Formation 控制台：`lakeformation:PutDataLakeSettings`。
   2. 前往数据目录设置，然后在跨账户版本设置中选择`Version 4`。

      如果您当前使用的是版本 1 或 2，请参阅有关更新至版本 3 的更新跨账户数据共享版本设置说明。

      从版本 3 升级到 4 无需更改权限策略。
   3. 列出您已授予主体的对数据库和表的权限。有关更多信息，请参阅在 Lake Formation 中查看数据库和表权限。
   4. 通过选择主体和资源来重新授予现有的 Lake Formation 跨账户权限。

> **Note**
>
> 在将数据位置注册更新为混合访问模式以授予跨账户权限之前，您需要为每个账户重新授予至少一个跨账户数据共享。要更新附加到 AWS RAM 资源共享的 AWS RAM 托管权限，必须执行此步骤。
> 2023 年 7 月，Lake Formation 更新了用于共享数据库和表格的 AWS RAM 托管权限：
>
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase`（数据库级别共享策略）
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite`（表级别共享策略）
>
> 2023 年 7 月之前授予的跨账户权限没有这些更新的 AWS RAM 权限。
> 如果您已直接向主体授予跨账户权限，则需要单独向主体重新授予这些权限。如果跳过此步骤，访问共享资源的主体可能会遇到非法组合错误。

5. 前往 https://console.aws.amazon.com/ram。
6. AWS RAM 控制台中的 “由我共享” 选项卡显示您与外部账户或委托人共享的数据库和表名。

   确保附加到共享资源的权限具有正确的 ARN。
7. 验证 AWS RAM 共享中的资源是否处于`Associated`状态。如果状态显示为 `Associating`，请等待，直到它们进入 `Associated` 状态。如果状态变为 `Failed`，请停下来联系 Lake Formation 服务团队。
8. 在左侧导航栏中的权限下选择混合访问模式，然后选择添加。
9. 添加主体和资源页面显示数据库和/或表以及有权访问的主体。您可以通过添加或删除主体和资源来进行所需的更新。
10. 选择对要更改为混合访问模式的数据库和表拥有 Lake Formation 权限的主体。选择数据库和表。
11. 选择添加以选择要在混合访问模式下强制实施 Lake Formation 权限的主体。
12. 向虚拟组 `IAMAllowedPrincipals` 授予对您的数据库和所选表的 `Super` 权限。
13. 将 Amazon S3 位置的 Lake Formation 注册编辑为混合访问模式。
14. 使用 IAM 权限策略为外部（消费者）账户中的 AWS Glue 用户授予对 Amazon S3 AWS Glue 操作的权限。

2. 使用者账户设置

   1. 通过 https://console.aws.amazon.com/lakeformation/ 以数据湖管理员身份登录 Lake Formation 控制台。
   2. 前往 https://console.aws.amazon.com/ram 并接受资源共享邀请。 AWS RAM 页面中的 “与我共享的资源” 选项卡显示与您的账户共享的数据库和表名。

      对于 AWS RAM 共享，请确保附加的权限具有共享 AWS RAM 邀请的正确 ARN。检查 AWS RAM 共享中的资源是否处于`Associated`状态。如果状态显示为 `Associating`，请等待，直到它们进入 `Associated` 状态。如果状态变为 `Failed`，请停下来联系 Lake Formation 服务团队。
   3. 在 Lake Formation 中创建指向共享数据库和/或表的资源链接。
   4. 向您（使用者）账户中的 IAM 主体授予对资源链接的 `Describe` 权限和对原共享资源的 `Grant on target` 权限。
   5. 接下来，为您账户中的主体设置对共享数据库或表的 Lake Formation 权限。

      在左侧导航栏中的权限下选择混合访问模式。
   6. 在混合访问模式页面的下半部分选择添加，从制作者账户中选择主体以及与您共享的数据库或表。
   7. 使用 IAM 权限策略向账户中的 AWS Glue 用户授予对 Amazon S3 AWS Glue 操作的权限。
   8. 使用 Athena 在表格上运行单独的示例查询，测试用户的 Lake Formation AWS Glue 权限和权限

      （可选）为处于混合访问模式的主体清除 Amazon S3 的 IAM 权限策略。

从混合访问模式下删除主体和资源

按照以下步骤将数据库、表和主体从混合访问模式下删除。

Console

1. 通过以下网址登录 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在权限下，选择混合访问模式。
3. 在混合访问模式页面上，选中数据库或表名称旁边的复选框，然后选择 `Remove`。
4. 此时将显示一条警告消息，提示您确认此操作。选择移除。

Lake Formation 不再对这些资源强制执行权限，对该资源的访问将使用 IAM 和 AWS Glue 权限进行控制。如果用户没有适当的 IAM 权限，则可能会导致他们无法再访问这些资源。

AWS CLI

以下示例显示了如何从混合访问模式下删除资源。

```
aws lakeformation delete-lake-formation-opt-in --cli-input-json file://file path

json:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::<123456789012>:role/role name"
    },
    "Resource": {
        "Table": {
            "CatalogId": "<123456789012>",
            "DatabaseName": "<database name>",
            "Name": "<table name>"
          }
    }
}
```

在混合访问模式下查看主体和资源

按照以下步骤在混合访问模式下查看数据库、表和主体。

Console

1. 通过以下网址登录 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在权限下，选择混合访问模式。
3. 混合访问模式页面显示当前处于混合访问模式的资源和主体。

AWS CLI

以下示例说明如何列出处于混合访问模式的全部所选主体和资源。

```
aws lakeformation list-lake-formation-opt-ins
```

以下示例说明了如何列出所选择的特定主体-资源对。

```
aws lakeformation list-lake-formation-opt-ins --cli-input-json file://file path

json:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::<account-id>:role/<role name>"
    },
    "Resource": {
        "Table": {
            "CatalogId": "<account-id>",
            "DatabaseName": "<database name>",
            "Name": "<table name>"
          }
    }
}
```

其他 资源

在以下博客文章中，我们将带您查看有关如何在其他用户已经可以通过 IAM 和 Amazon S3 权限访问数据库时在混合访问模式下为选定用户加载 Lake Formation 权限的说明。我们将查看在一个账户内和两个 AWS 账户之间设置混合访问模式的说明。

- 引入混合访问模式，使用 Lake Formation、IAM 和 Amazon S3 策略进行安全访问。 AWS Glue Data Catalog

创建数据目录表和数据库

AWS Lake Formation 使用 AWS Glue 数据目录来存储有关数据湖、数据来源、转换和目标的元数据。有关数据来源和目标的元数据采用数据库和表的形式。表存储有关基础数据的信息，包括架构信息、分区信息和数据位置。数据库是表的集合。数据目录还包含资源链接，这些链接是指向外部账户中共享数据库和表的链接，用于跨账户访问数据湖中的数据。

每个 AWS 账户在每个 AWS 区域都有一个数据目录。

主题

创建数据库

数据目录中的元数据表存储在数据库中。您可以根据需要创建任意数量的数据库，并且可以为每个数据库授予不同的 Lake Formation 权限。

数据库可以具有可选的位置属性。此位置通常位于向 Lake Formation 注册的Amazon Simple Storage (Amazon S3) 位置内。指定位置时，主体不需要数据位置权限即可创建指向数据库位置内位置的数据目录表。有关更多信息，请参阅Underlying data access control。

要使用 Lake Formation 控制台创建数据库，您必须以数据湖管理员或“数据库创建者”身份登录。数据库创建者是已被授予 Lake Formation CREATE_DATABASE 权限的主体。您可以在 Lake Formation 控制台的管理角色和任务页面上查看数据库创建者列表。要查看此列表，您必须具有 `lakeformation:ListPermissions` IAM 权限，并以数据湖管理员或通过授予选项获得 CREATE_DATABASE 权限的数据库创建者身份登录。

创建数据库

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台，然后以数据湖管理员或数据库创建者身份登录。
2. 在导航窗格的数据目录下，选择数据库。
3. 选择创建数据库。
4. 在创建数据库对话框中，输入数据库名称、可选位置和可选描述。
5. 取消选中仅对新数据库中的新表使用 IAM 访问控制。

   有关此选项的更多信息，请参阅the section called “更改数据湖的默认设置”。
6. 选择创建数据库。

创建表

AWS Lake Formation 元数据表包含有关数据湖中数据的信息，包括架构信息、分区信息和数据位置。这些表存储在 AWS Glue 数据目录中。您可以使用它们访问数据湖中的基础数据，并使用 Lake Formation 权限管理这些数据。表存储在数据目录的数据库中。

有以下几种方法可以创建数据目录表：

- 在 AWS Glue 中运行爬网程序。请参阅《AWS Glue 开发人员指南》中的定义爬网程序。
- 创建并运行工作流。请参阅 the section called “使用工作流导入数据”。
- 使用 Lake Formation 控制台、AWS Glue API 或 AWS Command Line Interface（AWS CLI）手动创建表。
- 使用 Amazon Athena 创建表。
- 创建指向外部账户中的表的资源链接。请参阅 the section called “创建资源链接”。

创建 Apache Iceberg 表

AWS Lake Formation 支持在 AWS Glue Data Catalog 中创建使用 Apache Parquet 数据格式的 Apache Iceberg 表，这些表的数据驻留在 Amazon S3 中。该数据目录中的表是表示数据存储中数据的元数据定义。默认情况下，Lake Formation 会创建 Iceberg v2 表。有关 v1 和 v2 表之间的区别，请参阅 Apache Iceberg 文档中的格式版本更改。

Apache Iceberg 是适用于超大型分析数据集的开放表格式。Iceberg 允许轻松更改架构，也称为架构发展，这意味着用户可以在不破坏基础数据的情况下添加、重命名或删除数据表中的列。Iceberg 还支持数据版本控制，允许用户跟踪数据随时间的变化。这将启用时间旅行功能，该功能允许用户访问和查询数据的历史版本，并分析更新和删除之间的数据更改。

您可以使用 Lake Formation 控制台或 AWS Glue API 中的 `CreateTable` 操作在数据目录中创建 Iceberg 表。有关更多信息，请参阅CreateTable 操作（Python：create_table）。

在数据目录中创建 Iceberg 表时，您必须在 Amazon S3 中指定表格式和元数据文件路径，以便能够执行读取和写入操作。

当您向 AWS Lake Formation 注册 Amazon S3 数据位置时，您可以使用 Lake Formation 通过精细访问控制权限来保护 Iceberg 表。对于 Amazon S3 中的源数据和未向 Lake Formation 注册的元数据，访问权限由 Amazon S3 和 AWS Glue 操作的 IAM 权限策略决定。有关更多信息，请参阅管理 Lake Formation 权限。

> **Note**
>
> 数据目录不支持创建分区和添加 Iceberg 表属性。

主题

先决条件

要在数据目录中创建 Iceberg 表并设置 Lake Formation 数据访问权限，您需要完成以下要求：

1. 在没有向 Lake Formation 注册数据的情况下创建 Iceberg 表所需的权限。

   除了在数据目录中创建表所需的权限外，表创建者还需要以下权限：

   - 针对资源 arn:aws:s3:::{bucketName} 的 `s3:PutObject`
   - 针对资源 arn:aws:s3:::{bucketName} 的 `s3:GetObject`
   - 针对资源 arn:aws:s3:::{bucketName} 的 `s3:DeleteObject`
2. 使用向 Lake Formation 注册的数据创建 Iceberg 表所需的权限：

   要使用 Lake Formation 管理和保护数据湖中的数据，请向 Lake Formation 注册包含表数据的 Amazon S3 位置。这样，Lake Formation 就可以向 Athena、Redshift Spectrum 和 Amazon EMR 等 AWS 分析服务提供凭证以访问数据。有关注册 Amazon S3 位置的更多信息，请参阅向数据湖添加 Amazon S3 位置。

   读取和写入向 Lake Formation 注册的基础数据的主体需要以下权限：

   - `lakeformation:GetDataAccess`
   - `DATA_LOCATION_ACCESS`

     对某个位置具有数据位置权限的主体也对所有子位置具有位置权限。

     有关数据位置权限的更多信息，请参阅基础数据访问控制。

要启用压缩，该服务需要代入有权更新数据目录中的表的 IAM 角色。有关详细信息，请参阅 表优化的先决条件

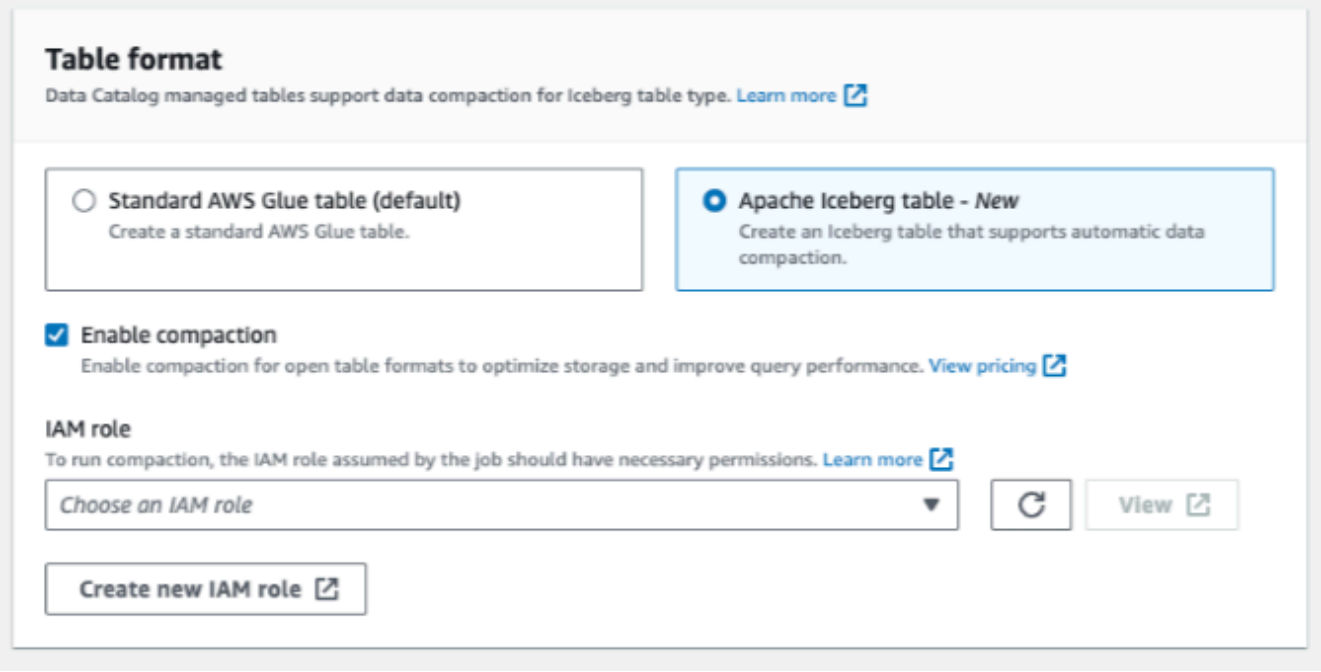
创建 Iceberg 表

您可以使用 Lake Formation 控制台或 AWS Command Line Interface创建 Iceberg v1 和 v2 表，如本页所述。您也可以使用 AWS Glue 控制台或 AWS Glue 爬网程序创建 Iceberg 表。有关更多信息，请参阅《AWS Glue 开发人员指南》中的数据目录和爬网程序。

创建 Iceberg 表

Console

1. 登录 AWS Management Console，然后通过以下网址打开 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在数据目录下，选择表，然后使用创建表按钮指定以下属性：
   - 表名称：输入表的唯一名称。如果您使用 Athena 访问表，请使用《Amazon Athena 用户指南》中的这些命名提示。
   - 数据库：选择现有数据库或创建新数据库。
   - 描述：表的描述。您可以编写描述以帮助您了解表的内容。
   - 表格式：对于表格式，请选择 Apache Iceberg。



   - 启用压缩：选择启用压缩，将表中的小 Amazon S3 对象压缩成较大对象。
   - IAM 角色：为了运行压缩，该服务会代表您代入一个 IAM 角色。您可以使用下拉列表选择一个 IAM 角色。确保该角色具有启用压缩所需的权限。

     要了解有关所需权限的更多信息，请参阅 表优化的先决条件 。

- 位置：指定 Amazon S3 中存储元数据表的文件夹的路径。Iceberg 需要数据目录中的元数据文件和位置才能执行读取和写入。
- 架构：选择添加列以添加列和列的数据类型。您可以选择创建一个空表，然后稍后更新架构。数据目录支持 Hive 数据类型。有关更多信息，请参阅 Hive 数据类型。

  Iceberg 允许您在创建表后演变架构和分区。您可以使用 Athena 查询更新表架构，使用 Spark 查询更新分区。

AWS CLI

```
aws glue create-table \
    --database-name iceberg-db \
    --region us-west-2 \
    --open-table-format-input '{
      "IcebergInput": {
           "MetadataOperation": "CREATE",
           "Version": "2"
        }
      }' \
    --table-input '{"Name":"test-iceberg-input-demo",
            "TableType": "EXTERNAL_TABLE",
            "StorageDescriptor":{
               "Columns":[
                   {"Name":"col1", "Type":"int"},
                   {"Name":"col2", "Type":"int"},
                   {"Name":"col3", "Type":"string"}
                ],
               "Location":"s3://DOC_EXAMPLE_BUCKET_ICEBERG/"
            }
       }'
```

优化 Iceberg 表

使用 Apache Iceberg 等开放表格式的 Amazon S3 数据湖会将数据存储为 Amazon S3 对象。如果数据湖表中包含成千上万个小 Amazon S3 对象，则会增加 Iceberg 表的元数据开销并影响读取性能。为提高 AWS 分析服务（例如 Amazon Athena 和 Amazon EMR）和 AWS Glue ETL 作业的读取性能，AWS Glue Data Catalog 为数据目录中的 Iceberg 表提供了托管式压缩功能（一种将小 Amazon

S3 对象压缩成较大对象的进程）。您可以使用 Lake Formation 控制台、AWS Glue 控制台、AWS CLI 或 AWS API 为数据目录中的单个 Iceberg 表启用或禁用压缩。

表优化器会持续监控表分区，并在超过文件数量和文件大小阈值时启动压缩进程。在数据目录中，启动压缩的默认阈值设置为 384MB，而在 Iceberg 库中，压缩阈值约为目标文件大小的 75%。数据目录会在不干扰并发查询的情况下执行压缩。数据目录仅支持对 Parquet 格式的表进行数据压缩。

有关支持的数据类型、压缩格式和限制，请参阅托管式数据压缩的支持的格式和限制 。

主题

表优化的先决条件

表优化器会代入您在为表启用压缩时指定的 AWS Identity and Access Management (IAM) 角色的权限。该 IAM 角色必须具有读取数据和更新数据目录中元数据的权限。您可以创建一个 IAM 角色并附加以下内联策略：

- 添加以下内联策略，以向 Amazon S3 授予对未注册到 Lake Formation 的数据位置的读/写权限。此策略还包括更新数据目录中表的权限，以及允许 AWS Glue 在 Amazon CloudWatch 日志中添加日志并发布指标的权限。对于 Amazon S3 中未注册到 Lake Formation 的源数据，访问权限由 Amazon S3 和 AWS Glue 操作的 IAM 权限策略决定。

  请将以下内联策略中的 `bucket-name` 替换为您的 Amazon S3 存储桶名称，请将 `aws-account-id` 和 `region` 替换为有效的 AWS 账户和数据目录所在的区域，将 `database_name` 替换为数据库的名称，并将 `table_name` 替换为表的名称。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
```

```
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::<bucket-name>/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::<bucket-name>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "glue:UpdateTable",
                "glue:GetTable"
            ],
            "Resource": [
                "arn:aws:glue:<region>:<aws-account-id>:table/<database-name>/<table-
name>",
                "arn:aws:glue:<region>:<aws-account-id>:database/<database-name>",
                "arn:aws:glue:<region>:<aws-account-id>:catalog"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:<region>:<aws-account-id>:log-group:/aws-glue/
iceberg-compaction/logs:*"
        }
    ]
}
```

- 使用以下策略为注册到 Lake Formation 的数据启用压缩功能。

如果该压缩角色不具有对表的 IAM_ALLOWED_PRINCIPALS 组权限，则该角色需要具有对该表的 Lake Formation ALTER、DESCRIBE、INSERT 和 DELETE 权限。

有关向 Lake Formation 注册 Amazon S3 存储桶的更多信息，请参阅向数据湖添加 Amazon S3 位置。

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

        "Resource": [
            "arn:aws:glue:<region>:<aws-account-
id>:table/<databaseName>/<tableName>",
            "arn:aws:glue:<region>:<aws-account-id>:database/<database-name>",
            "arn:aws:glue:<region>:<aws-account-id>:catalog"
        ]
    },
    {
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:CreateLogStream",
            "logs:PutLogEvents"
        ],
        "Resource": "arn:aws:logs:<region>:<aws-account-id>:log-group:/aws-glue/
iceberg-compaction/logs:*"
        }
    ]
 }
```

- （可选）如果要压缩的 Iceberg 表包含使用服务器端加密进行加密的 Amazon S3 存储桶中数据，该压缩角色需要具有解密 Amazon S3 对象并生成新数据密钥以将对象写入加密存储桶的权限。将以下策略添加到需要的 AWS KMS 密钥。我们仅支持在存储桶级加密。

```
{
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::<aws-account-id>:role/<compaction-role-name>"
    },
    "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey"
    ],
    "Resource": "*"
}
```

- （可选）对于注册到 Lake Formation 的数据位置，用于注册该位置的角色需要具有解密 Amazon S3 对象并生成新数据密钥以将对象写入加密存储桶的权限。有关更多信息，请参阅注册加密的 Amazon S3 位置。
- （可选）如果 AWS KMS 密钥存储在其他 AWS 账户中，则需要为该压缩角色添加以下权限。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:GenerateDataKey"
            ],
            "Resource": ["arn:aws:kms:<REGION>:<KEY_OWNER_ACCOUNT_ID>:key/<KEY_ID>" ]
        }
    ]
}
```

- 用于运行压缩的角色必须拥有该角色的 `iam:PassRole` 权限。

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::<account-id>:role/<compaction-role-name>"
            ]
        }
    ]
}
```

- 将以下信任策略添加到该角色，以便 AWS Glue 服务代入该 IAM 角色来运行压缩进程。

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

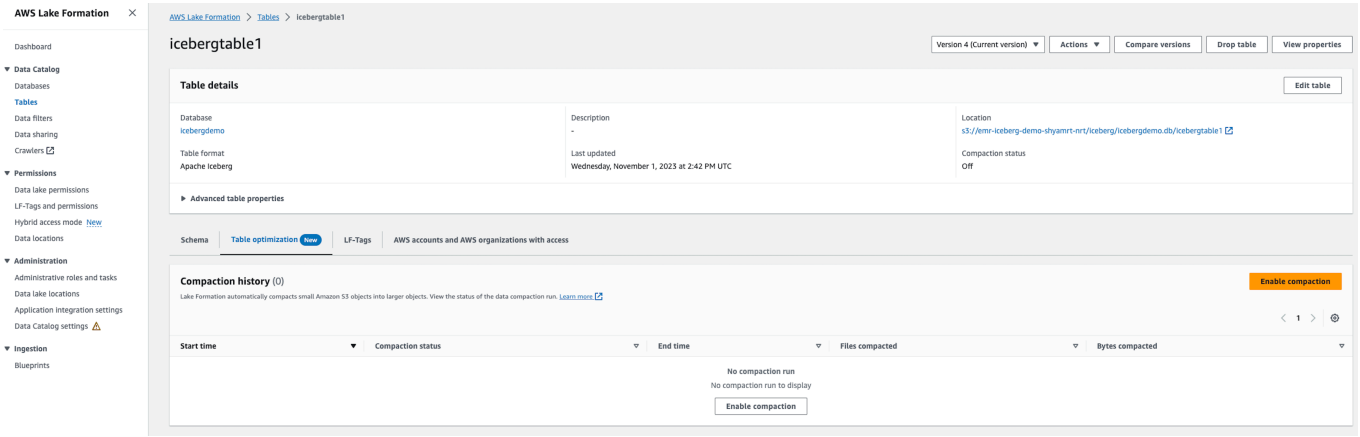
启用压缩

您可以使用 Lake Formation 控制台、AWS Glue 控制台、AWS CLI 或 AWS API 为数据目录中的 Apache Iceberg 表启用压缩。对于新表，您可以在创建表时选择 Apache Iceberg 表格式并启用压缩。新表会默认禁用压缩。

Console

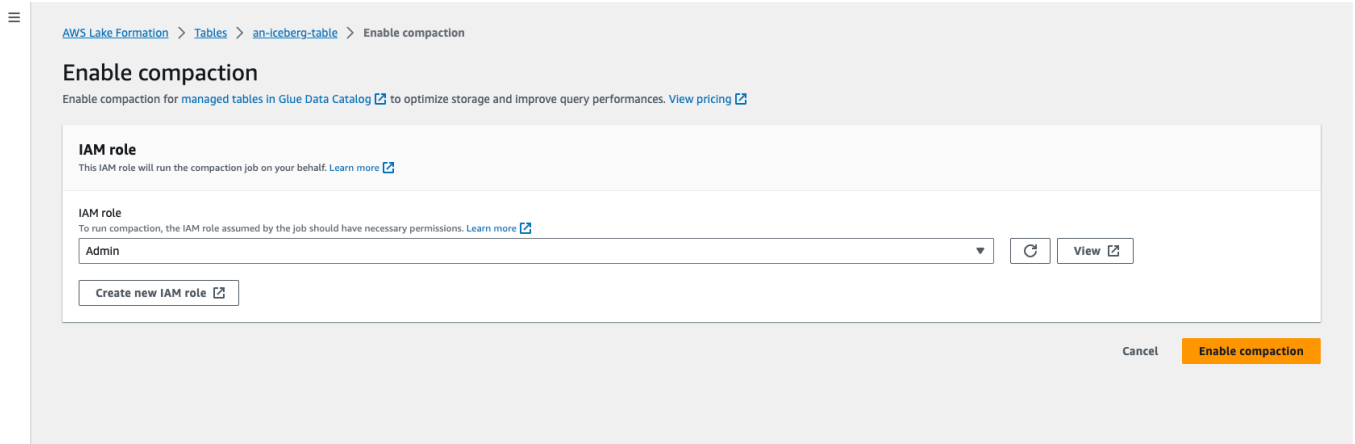
启用压缩

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台，然后以数据湖管理员、表创建者或已被授予对表的 `glue:UpdateTable` 和 `lakeformation:GetDataAccess` 权限的用户身份登录。
2. 在导航窗格的数据目录下，选择表。
3. 在表页面上，选择要启用压缩的开放表格式的表，然后在操作菜单下，选择启用压缩。
4. 您也可以通过选中该表并打开表详细信息页面来启用压缩。选择页面下半部的表优化选项卡，然后选择启用压缩。



5. 然后从下拉列表中选择一个具有 表优化的先决条件 部分所示权限的现有 IAM 角色。

   选择创建新的 IAM 角色选项后，服务会创建一个具有运行压缩所需权限的自定义角色。



   按照以下步骤更新一个现有的 IAM 角色：

   a. 要更新 IAM 角色的权限策略，请在 IAM 控制台中转到用于运行压缩的 IAM 角色。

b. 在“添加权限”部分中，选择“创建策略”。在新打开的浏览器窗口中，创建将用于您的角色的新策略。

c. 在“创建策略”页面上，选择 JSON 选项卡。将“先决条件”中显示的 JSON 代码复制到策略编辑器字段中。

AWS CLI

以下示例演示如何启用压缩。将账户 ID 替换为有效的 AWS 账户 ID。将数据库名称和表名称替换为实际的 Iceberg 表名称和数据库名称。将 `roleArn` 替换为 IAM 角色的 AWS 资源名称 (ARN) 以及具有运行压缩所需权限的 IAM 角色的名称。

```
aws glue create-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --table-optimizer-configuration
 '{"roleArn":"arn:aws:iam::123456789012:role/compaction_role", "enabled":'true'}' \
  --type compaction
```

AWS API

调用 `CreateTableOptimizer` 操作为表启用压缩。

启用压缩后，表优化选项卡会显示以下压缩详细信息（大约 15–20 分钟后）：

开始时间

在 Lake Formation 中启动压缩进程的时间。该值是一个采用 UTC 时间格式的时间戳。

结束时间

数据目录中压缩进程结束的时间。该值是一个采用 UTC 时间格式的时间戳。

Status

压缩运行的状态。值为成功或失败。

已压缩的文件数

已压缩的文件总数。

已压缩的字节数

已压缩的字节总数。

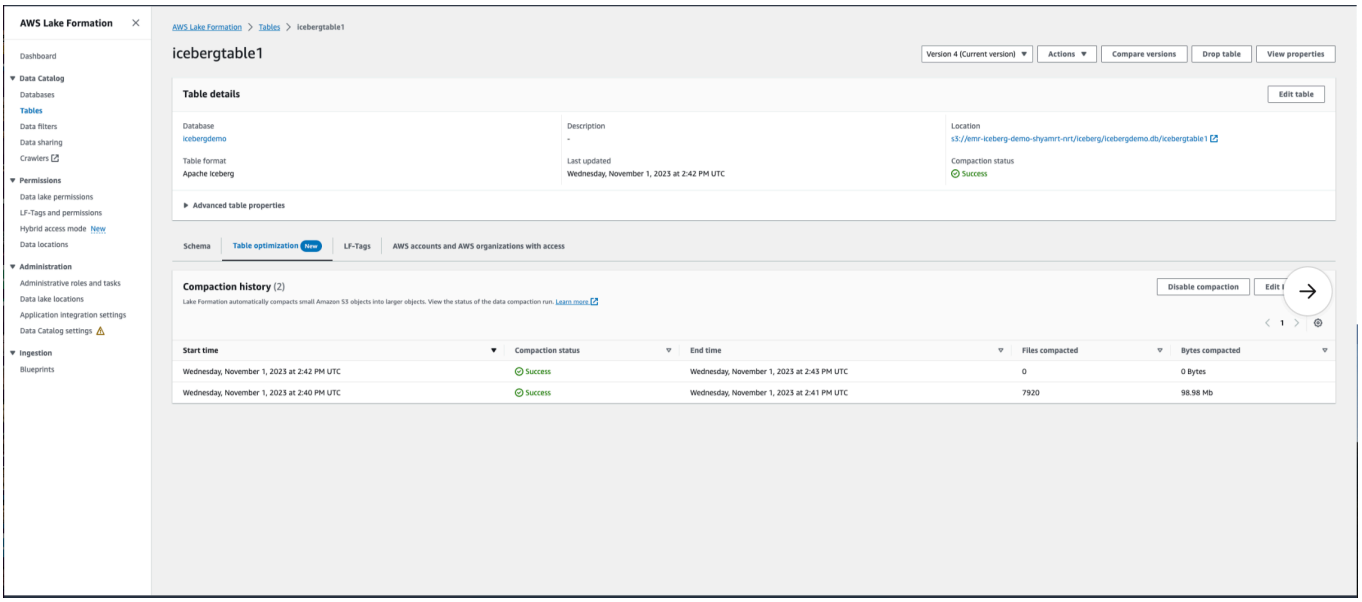
禁用压缩

您可以使用 AWS Glue 控制台或 AWS CLI 来为特定 Apache Iceberg 表禁用自动压缩。

Console

1. 选择数据目录，然后选择表。从表列表中，选择要禁用压缩的开放表格式的表。
2. 您可以选择一个 Iceberg 表，然后在操作下选择禁用压缩。

   您也可以选择表详细信息页面下半部分的禁用压缩，从而为表禁用压缩。



3. 在确认消息页面选择禁用压缩。您可以在以后重新启用压缩。

   确认后，压缩将被禁用，并且表的压缩状态将恢复为 `Off`。

AWS CLI

将以下示例中的账户 ID 替换为有效的 AWS 账户 ID。将数据库名称和表名称替换为实际的 Iceberg 表名称和数据库名称。将 `roleArn` 替换为 IAM 角色的 AWS 资源名称 (ARN) 以及具有运行压缩所需权限的 IAM 角色的实际名称。

```
aws glue update-table-optimizer \
```

```
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --table-optimizer-configuration
'{"roleArn":"arn:aws:iam::123456789012:role/compaction_role", "enabled":'false'}'\
  --type compaction
```

AWS API

调用 UpdateTableOptimizer 操作以禁用特定表的压缩。

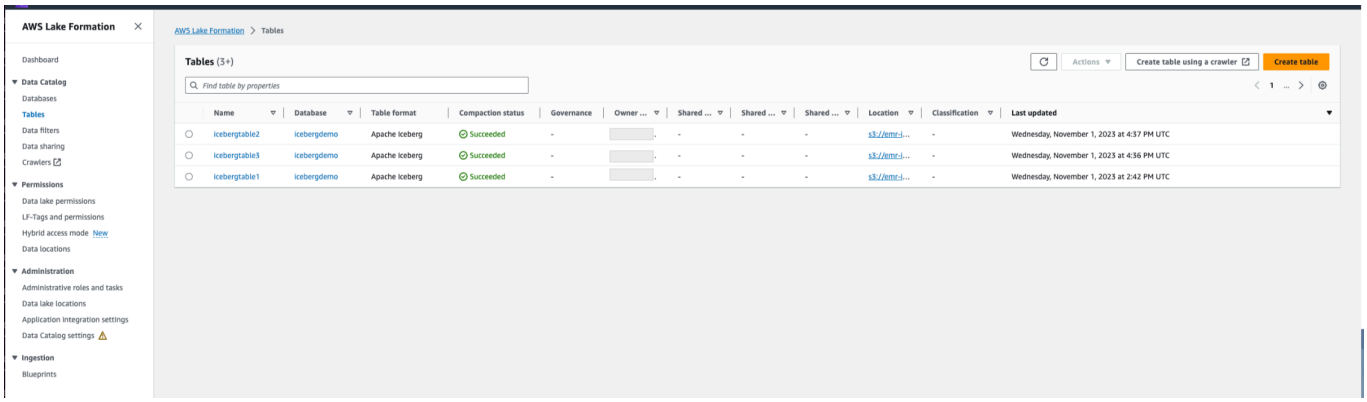
查看压缩详细信息

您可以使用 Lake Formation 控制台、AWS CLI 或使用 AWS API 操作查看 Apache Iceberg 的压缩状态。

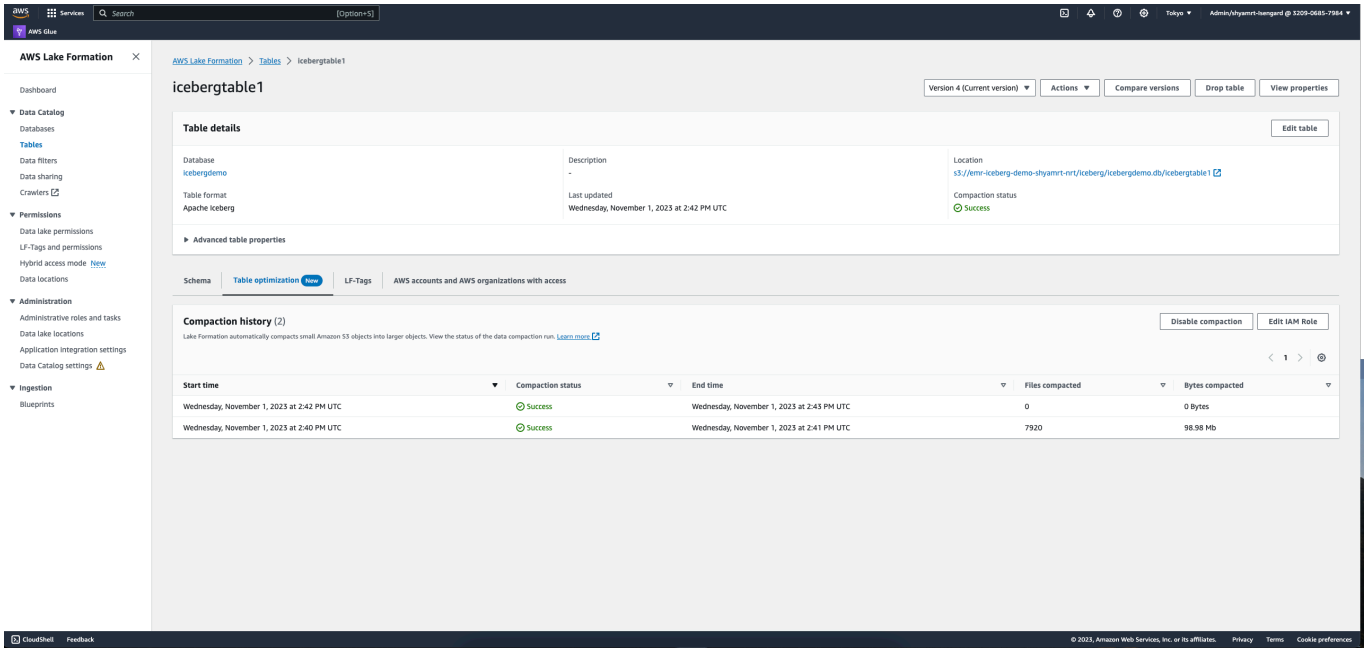
Console

查看 Iceberg 表的压缩状态（控制台）

- 您可以通过在 Lake Formation 控制台中选择数据目录下的表来查看 Iceberg 表的压缩状态。压缩状态字段将显示压缩运行的状态。您可以使用表首选项来显示表的格式和压缩状态。



- 要查看特定表的压缩运行历史记录，请选择 AWS Glue Data Catalog 下的表，然后选择一个表来查看该表的详细信息。表优化选项卡将显示表的压缩历史记录。



AWS CLI

您可以使用 AWS CLI 查看压缩详细信息。

请将以下示例中的账户 ID 替换为有效的 AWS 账户 ID，将数据库名称和表名称替换为实际的 Iceberg 表名称。

- 获取表的上次压缩详细信息

```
aws get-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

- 使用以下示例来检索特定表的优化器历史记录。

```
aws list-table-optimizer-runs \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

- 以下示例演示了如何检索多个优化器的压缩运行和配置详细信息。您最多可以指定 20 个优化器。

```
aws glue batch-get-table-optimizer \
--entries '[{"catalogId":"123456789012", "databaseName":"iceberg_db",
 "tableName":"iceberg_table", "type":"compaction"}]'
```

AWS API

- 使用 GetTableOptimizer 操作检索优化器的上次运行详细信息。
- 使用 ListTableOptimizerRuns 操作检索特定表上给定优化器的历史记录。您可以在单个 API 调用中指定 20 个优化器。
- 使用 BatchGetTableOptimizer 操作检索您账户中多个优化器的配置详细信息。此操作不支持跨账户调用。

查看 Amazon CloudWatch 指标

成功运行压缩后，服务会创建有关压缩作业性能的 Amazon CloudWatch 指标。您可以前往CloudWatch 指标并选择指标、所有指标。您可以按特定的命名空间（例如 AWS Glue）、表名称或数据库名称筛选指标。

有关更多信息，请参阅《Amazon CloudWatch 用户指南》中的查看可用的指标。

- 已压缩的字节数
- 已压缩的文件数
- 分配给作业的 DPU 数
- 作业持续时间（小时）

删除优化器

您可以使用 AWS CLI 或 AWS API 操作来删除表的优化器和关联的元数据。

运行以下 AWS CLI 命令删除表的压缩历史记录。

```
aws glue delete-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
```

```
  --table-name iceberg_table \
  --type compaction
```

使用 DeleteTableOptimizer 操作删除表的优化器。

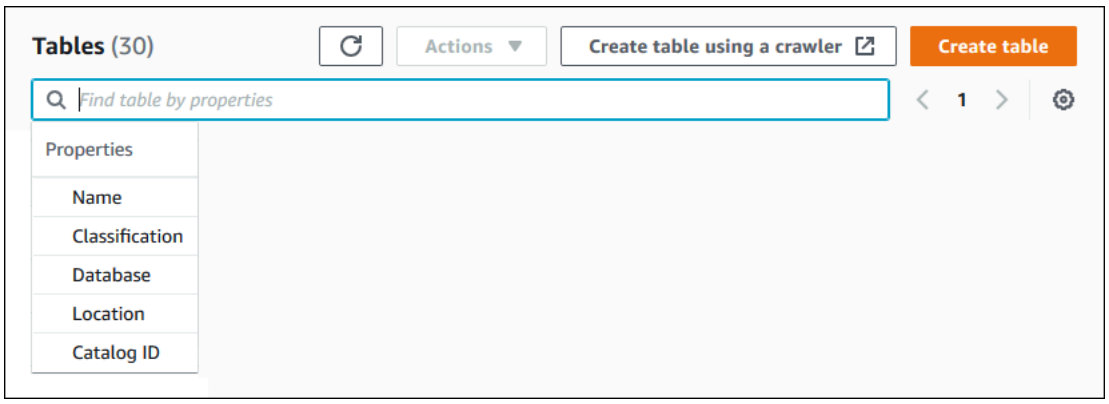
搜索表

您可以使用 AWS Lake Formation 控制台按名称、位置、包含的数据库等搜索数据目录表。搜索结果仅显示您具有 Lake Formation 权限的表。

搜索表（控制台）

1. 登录 AWS Management Console，然后通过以下网址打开 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在导航窗格中，选择表。
3. 将光标置于页面顶部的搜索字段中。该字段具有占位符文本“按属性查找表”。

   此时将显示属性菜单，其中显示了要搜索的各种表属性。

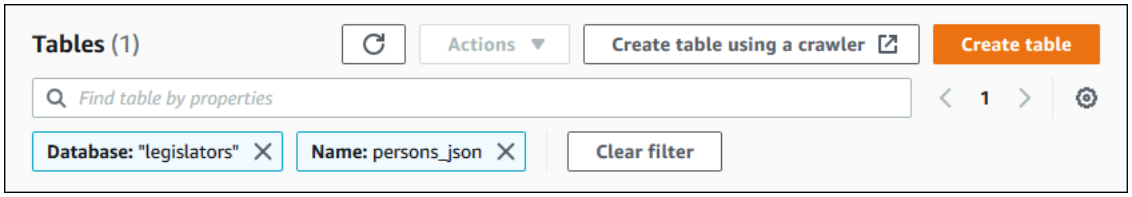


4. 请执行以下操作之一：

   - 按包含的数据库进行搜索。
     1. 从属性菜单中选择数据库，然后从显示的数据库菜单中选择一个数据库，或者键入数据库名称并按 Enter。

        此时将列出您在数据库中具有权限的表。
     2. （可选）要将列表范围缩小到数据库中的单个表，请再次将光标置于搜索字段中，从属性菜单中选择名称，然后从显示的表菜单中选择表名称，或者键入表名称并按 Enter。

        此时将列出单个表，并且数据库名称和表名称都显示为搜索字段下的磁贴。



要调整筛选条件，请关闭任一磁贴或选择清除筛选条件。

- 按其他属性进行搜索。

  1. 从属性菜单中选择搜索属性。

     要按 AWS 账户 ID 进行搜索，请从属性菜单中选择目录 ID，输入有效的 AWS 账户 ID（例如，111122223333），然后按 Enter。

     要按位置进行搜索，请从属性菜单中选择位置，然后从显示的位置菜单中选择一个位置。此时将返回所选位置（例如，Amazon S3）的根位置中的所有表。

跨 AWS 账户共享数据目录表和数据库

通过向外部 AWS 账户授予对数据目录资源（数据库和表）的 Lake Formation 权限，您可以与这些外部账户共享这些资源。然后，用户可以运行跨多个账户联接和查询表的查询和作业。但有一些限制，当您与其他账户共享数据目录资源时，该账户中的主体可以对该资源进行操作，就像该资源位于其数据目录中一样。

您不会与外部 AWS 账户中的特定主体共享资源，而是与 AWS 账户或组织共享资源。当您与 AWS 组织共享资源时，就是与该组织中所有级别的所有账户共享资源。然后，每个外部账户中的数据湖管理员必须向其账户中的主体授予共享资源的权限。

有关更多信息，请参阅 Lake Formation 中的跨账户数据共享 和 授予和撤销对数据目录资源的权限。

另请参见：

- 访问和查看共享数据目录表和数据库
- 先决条件

使用视图

该功能为预览版，可能会发生变化。有关更多信息，请参阅 AWS 服务条款文档中的“测试版和预览”部分。

在 AWS Glue Data Catalog 中，“视图”是一个虚拟表，其中的内容由引用一个或多个表的查询定义。您可以使用适用于 Amazon Athena、Amazon Redshift 或 Amazon EMR 的 SQL 编辑器创建一个最多引用 10 个表的视图。视图的基础引用表可以属于同一数据库，也可以属于同一 AWS 账户内的不同数据库。

SQL 是一种用于查询表的编程语言，每个 AWS 分析引擎都使用自己的 SQL 变体或 SQL 方言。数据目录支持使用不同的 SQL 方言创建视图，前提是每种方言引用一组相同的表、列和数据类型。通过定义可从多个引擎查询的通用视图架构和元数据对象，数据目录视图使您能够在整个数据湖中使用统一视图。

在数据目录中管理视图时，您可以使用 AWS Lake Formation 通过命名资源方法或使用 LF 标签授予精细权限，并在 AWS 账户、AWS Organizations 和组织单位之间共享这些权限。您也可以跨 AWS 区域共享数据目录视图。这允许用户跨 AWS 区域提供数据访问，而无需复制数据来源。

有关跨账户数据共享和跨区域数据访问的更多信息，请参阅：

- Lake Formation 中的跨账户数据共享
- 跨区域访问表

您可以使用数据目录视图执行以下操作：

- 创建和管理对单个视图架构的权限。这有助于避免对在多个引擎中创建的重复视图的权限存在不一致的风险。
- 向用户授予对引用多个表的视图的权限，而无需直接授予对基础引用表的权限。

有关限制，请参阅数据目录视图注意事项和限制。

主题

- 创建视图的先决条件
- 创建视图
- 授予对数据目录视图的权限

创建视图的先决条件

- 要在数据目录中创建视图，您必须向 Lake Formation 注册引用表的基础 Amazon S3 数据位置。

  有关向 Lake Formation 注册数据的详细信息，请参阅向数据湖添加 Amazon S3 位置。
- 视图定义者必须是 IAM 角色。其他 IAM 身份无法创建数据目录视图。
- 定义视图的 IAM 角色必须具有以下权限：
  - 对所有引用表具有 `Grantable` 选项的完整 `SELECT` 权限。
  - Lake Formation 和 AWS Glue 服务代入角色的信任策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

  - AWS Glue 和 Lake Formation 的 iam:PassRole 权限。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerPassRole1",
            "Action": [
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
```

```
                "StringEquals": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com",
                        "lakeformation.amazonaws.com"
                      ]
                }
            }
        }
    ]
}
```

- AWS Glue 和 Lake Formation 权限。

```
{
    "Version": "2012-10-17",
                 "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "Glue:GetDatabase",
                "Glue:GetDatabases",
                "Glue:CreateTable",
                "Glue:GetTable",
                "Glue:UpdateTable",
                "Glue:DeleteTable",
                "Glue:GetTables",
                "Glue:SearchTables",
                "Glue:BatchGetPartition",
                "Glue:GetPartitions",
                "Glue:GetPartition",
                "Glue:GetTableVersion",
                "Glue:GetTableVersions",
                "lakeFormation:GetDataAccess",
                "lakeFormation:GetTemporaryTableCredentials",
                "lakeFormation:GetTemporaryGlueTableCredentials",
                "lakeFormation:GetTemporaryUserCredentialsWithSAML"
            ],
            "Resource": "*"
        }
    ]
}
```

- 如果要创建视图的数据库具有已向 IAMAllowedPrincipals 组授予的 Super 或 ALL 权限，则无法创建视图。要撤销 IAMAllowedPrincipals 组对数据库的 Super 权限，请参阅步骤 4：将数据存储切换到 Lake Formation 权限模型。

  如果现有数据湖设置不允许将 IAMAllowedPrincipals 组的 CreateTableDefaultPermissions 设置为空，则可以创建新数据库并使用以下结构对数据湖设置进行编码。

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        CreateTableDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": []
            }
        ]
 }
```

创建视图

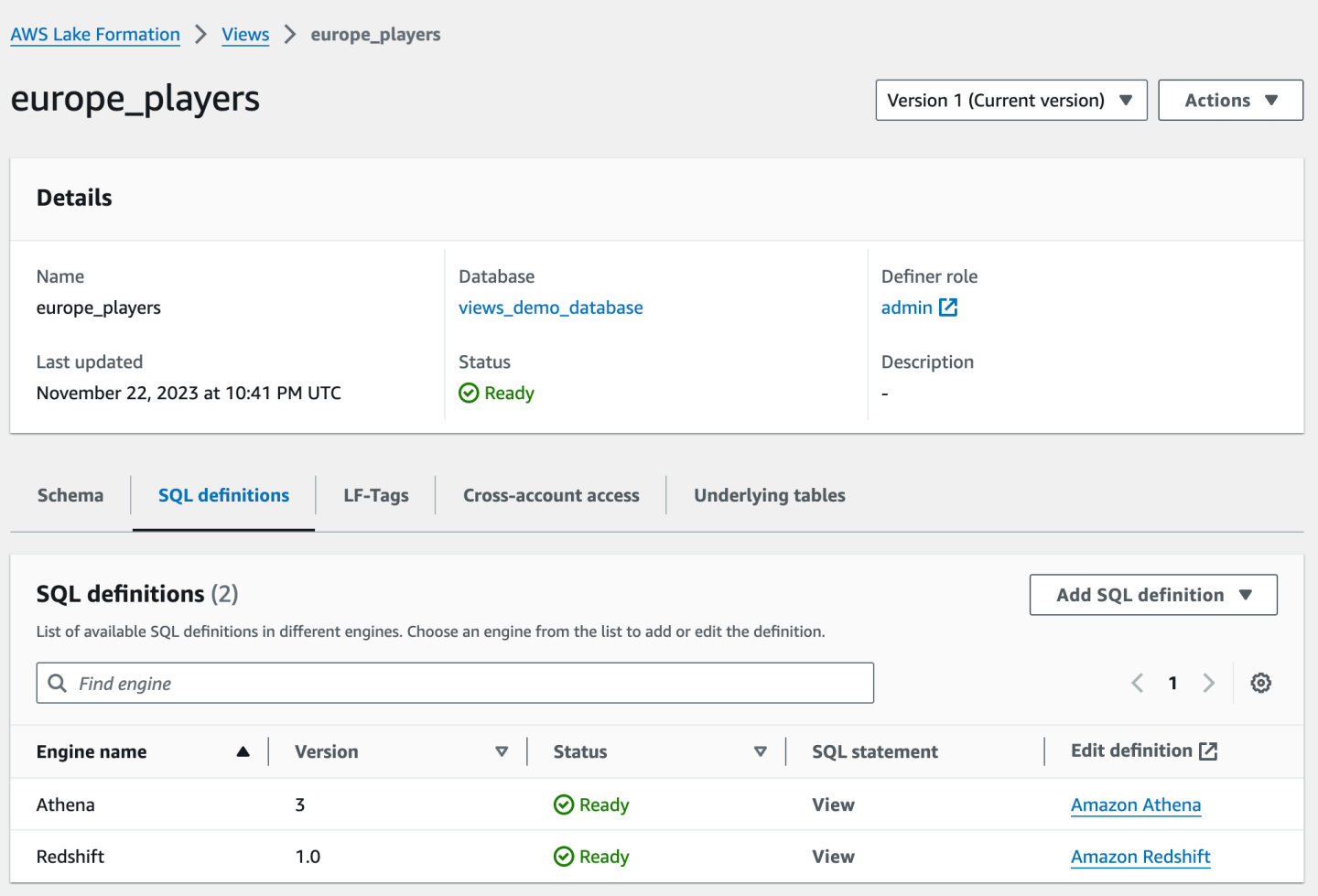
您可以使用适用于 Athena、Amazon Redshift 或 Amazon EMR 的 SQL 编辑器在 AWS Glue Data Catalog 中创建视图。

有关用于创建和管理数据目录视图的语法的更多信息，请参阅：

- 《Amazon Athena 用户指南》中的使用 AWS Glue Data Catalog 视图。
- 《Amazon Redshift 数据库开发人员指南》中的在 AWS Glue Data Catalog 中创建视图。
- 《Amazon EMR 管理指南》中的使用 AWS Glue Data Catalog 视图。

创建数据目录视图后，该视图的详细信息将显示在 Lake Formation 控制台中。

1. 在 Lake Formation 控制台中选择“数据目录”下的视图。
2. 可用视图列表将显示在“视图”页面上。
3. 从列表中选择一个视图，详细信息页面将显示该视图的属性。



架构

选择一个 Column 行，然后选择编辑 LF 标签，以更新标签值或分配新的 LF 标签。

SQL 定义

您可以查看可用 SQL 定义的列表。选择添加 SQL 定义，然后选择要添加 SQL 定义的查询引擎。在 Edit definition 列下选择查询引擎（Athena 或 Amazon Redshift）以更新 SQL 定义。

LF 标签

选择编辑 LF 标签以编辑标签的值或分配新标签。您可以使用 LF 标签来授予对视图的权限。

跨账户存取

您可以查看已共享数据目录视图的 AWS 账户、Organizations 和组织单位（OU）的列表。

基础表

用于创建视图的 SQL 定义中引用的基础表显示在此选项卡下。

授予对数据目录视图的权限

创建视图后，您可以跨 AWS 账户、Organizations 和组织单位向主体授予数据湖对视图的权限。有关授予权限的更多信息，请参阅 使用命名资源方法授予对视图的权限 。

在 Lake Formation 中使用工作流导入数据

通过 AWS Lake Formation，您可以使用工作流导入数据。工作流定义数据来源和计划以将数据导入到数据湖中。它是一个容器，用于存放 AWS Glue 爬网程序、作业和触发器，用于编排加载和更新数据湖的流程。

主题

- Lake Formation 中的蓝图和工作流
- 创建工作流
- 运行工作流

Lake Formation 中的蓝图和工作流

工作流封装了复杂的多作业提取、转换、加载 (ETL) 活动。工作流生成 AWS Glue 爬网程序、作业和触发器，以编排数据的加载和更新。Lake Formation 将工作流作为单个实体来执行和跟踪。您可以将工作流配置为按需或按计划运行。

您在 Lake Formation 中创建的工作流在 AWS Glue 控制台中显示为有向无环图 (DAG) 形式。每个 DAG 节点都是一个作业、爬网程序或触发器。要监控进度并进行故障排除，您可以跟踪工作流中每个节点的状态。

Lake Formation 工作流完成后，运行该工作流的用户将获得对该工作流创建的数据目录表的 Lake Formation SELECT 权限。

您也可以在 AWS Glue 中创建工作流。但是，由于 Lake Formation 允许您从蓝图创建工作流，因此在 Lake Formation 中创建工作流要简单得多，自动化程度也更高。Lake Formation 提供以下类型的蓝图：

- 数据库快照 – 将所有表中的数据从 JDBC 源加载或重新加载到数据湖中。您可以根据排除模式从该源中排除某些数据。
- 增量数据库 - 根据先前设置的书签，仅将新数据从 JDBC 源加载到数据湖中。您可以指定 JDBC 源数据库中要包含的各个表。对于每个表，您可以选择书签列和书签排序顺序，以跟踪之前加载的数据。首次对一组表运行增量数据库蓝图时，工作流会加载表中的所有数据，并为下一次增量数据库蓝图运行设置书签。因此，您可以使用增量数据库蓝图（而不是数据库快照蓝图）来加载所有数据，前提是将数据来源中的每个表指定为参数。
- 日志文件 - 从日志文件来源（包括 AWS CloudTrail、Elastic Load Balancing 日志和应用程序负载均衡器日志）批量加载数据。

使用下表可帮助确定是使用数据库快照蓝图还是增量数据库蓝图。

在以下情况下使用数据库快照...	在以下情况下使用增量数据库...
• 架构演变是灵活的。（将重命名列，删除以前的列，并在其位置添加新列。） • 源和目标之间需要完全一致。	• 架构演变是增量的。（只有连续添加列。） • 仅添加新行；不更新以前的行。

> **Note**
>
> 用户无法编辑 Lake Formation 创建的蓝图和工作流。

创建工作流

在开始之前，请确保已向角色 `LakeFormationWorkflowRole` 授予所需的数据权限和数据位置权限。这样，工作流就可以在数据目录中创建元数据表，并将数据写入 Amazon S3 中的目标位置。有关更多信息，请参阅（可选）为工作流程创建 IAM 角色和 Lake Formation 权限概述。

通过蓝图创建工作流

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以数据湖管理员或具有数据工程师权限的用户身份登录。有关更多信息，请参阅 Lake Formation 角色和 IAM 权限参考。
2. 在导航窗格中，选择蓝图，然后选择使用蓝图。
3. 在使用蓝图页面上，选择一个磁贴以选择蓝图类型。
4. 在导入来源下，指定数据来源。

   如果要从 JDBC 源导入，请指定以下内容：

   - 数据库连接 - 从列表中选择一个连接。使用 AWS Glue 控制台创建其他连接。连接中的 JDBC 用户名和密码决定了工作流有权访问的数据库对象。
   - 来源数据路径 - 输入 *<database>*/*<schema>*/*<table>* 或 *<database>*/*<table>*，具体取决于数据库产品。Oracle Database 和 MySQL 不支持路径中的架构。您可以用百分比 (%) 字符替换 *<schema>* 或 *<table>*。例如，对于系统标识符 (SID) 为 `orcl` 的 Oracle 数据库，输入 `orcl/%` 以导入连接中指定的用户有权访问的所有表。

     > **Important**
     >
     > 此字段区分大小写。如果任何组件存在大小写不匹配的情况，则工作流将失败。

   如果指定 MySQL 数据库，则 AWS Glue ETL 默认使用 Mysql5 JDBC 驱动程序，因此 MySQL8 不受原生支持。您可以将 ETL 作业脚本编辑为使用 `customJdbcDriverS3Path` 参数，如《AWS Glue 开发人员指南》中的 JDBC connectionType 值中所述，以使用支持 MySQL8 的其他 JDBC 驱动程序。

   如果要从日志文件导入，请确保您为工作流指定的角色（“工作流角色”）具有访问数据来源所需的 IAM 权限。例如，要导入 AWS CloudTrail 日志，用户必须具有 `cloudtrail:DescribeTrails` 和 `cloudtrail:LookupEvents` 权限才能在创建工作流时查看 CloudTrail 日志列表，并且工作流角色必须具有对 Amazon S3 中 CloudTrail 位置的权限。
5. 请执行下列操作之一：

   - 对于数据库快照蓝图类型，可以选择通过指定一个或多个排除模式来识别要导入的一部分数据。这些排除模式是 Unix 风格的 `glob` 模式。它们将存储为由工作流创建的表的属性。

有关可用排除模式的详细信息，请参阅《AWS Glue 开发人员指南》中的包含和排除模式。

- 对于增量数据库蓝图类型，请指定以下字段。为每个要导入的表添加一行。

  表名称

  要导入的表。必须全部小写。

  书签键

  定义书签键的列名的逗号分隔列表。如果为空，则使用主键确定新数据。每列的大小写必须与数据来源中定义的大小写匹配。

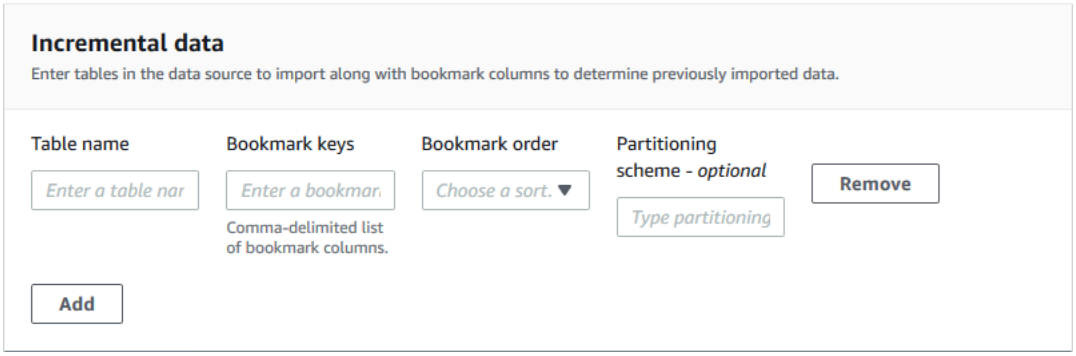
  > **Note**
  >
  > 仅当主键按顺序递增或递减（无间隙）时，主键才有资格成为默认书签键。如果要使用主键作为书签键并且它有间隙，则必须将主键列命名为书签键。

  书签顺序

  选择升序时，值大于书签值的行将被标识为新行。选择降序时，值小于书签值的行将被标识为新行。

  分区方案

  （可选）分区键列列表，用斜杠 (/) 分隔。示例： `year/month/day`。



有关更多信息，请参阅《AWS Glue 开发人员指南》中的使用作业书签来跟踪已处理的数据。

6. 在导入目标下，指定目标数据库、目标 Amazon S3 位置和数据格式。

确保工作流角色对数据库和 Amazon S3 目标位置具有所需的 Lake Formation 权限。

> **Note**
>
> 目前，蓝图不支持加密目标位置的数据。

7. 选择导入频率。

   您可以使用自定义选项指定 cron 表达式。

8. 在导入选项下：

   a. 输入工作流名称。

   b. 对于角色，选择您在（可选）为工作流程创建 IAM 角色中创建的角色 LakeFormationWorkflowRole。

   c. （可选）指定表前缀。该前缀位于工作流创建的数据目录表的名称之前。

9. 选择创建，然后等待控制台报告已成功创建工作流。

> **Tip**
>
> 您是否收到了以下错误消息？
> `User: arn:aws:iam::<account-id>:user/<username> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/<rolename>...`
> 如果是，请检查您是否已在所有策略中将 *<account-id>* 替换为有效的 AWS 账号。

> **另请参见：**
>
> - Lake Formation 中的蓝图和工作流

运行工作流

您可以使用 Lake Formation 控制台、AWS Glue 控制台、AWS Glue 命令行界面 (AWS CLI) 或 API 运行工作流。

运行工作流（Lake Formation 控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以数据湖管理员或具有数据工程师权限的用户身份登录。有关更多信息，请参阅 Lake Formation 角色和 IAM 权限参考。
2. 在导航窗格中，选择蓝图。
3. 在蓝图页面上，选择工作流。然后，在操作菜单上，选择开始。
4. 当工作流运行时，在上次运行状态列中查看其进度。时不时地选择刷新按钮。

   状态从正在运行依次变为正在发现、正在导入和已完成。

   工作流完成后：

   - 数据目录中将含有新的元数据表。
   - 您的数据已被摄取到数据湖中。

   如果工作流失败，请执行以下操作：

   a. 选择工作流。选择操作，然后选择查看图表。

      该工作流将在 AWS Glue 控制台中打开。

   b. 确保已选择工作流，然后选择历史记录选项卡。

   c. 在历史记录下，选择最近的运行，然后选择查看运行详细信息。

   d. 在动态（运行时）图表中选择失败的作业或爬网程序，然后查看错误消息。失败的节点为红色或黄色。

另请参见：

- Lake Formation 中的蓝图和工作流

管理 Lake Formation 权限

Lake Formation 为数据湖中的数据提供集中访问控制。您可以在 Lake Formation 中按角色为用户和应用程序定义基于安全策略的规则，并且与 AWS Identity and Access Management 的集成可对这些用户和角色进行身份验证。定义规则后，Lake Formation 将对 Amazon Redshift Spectrum 和 Amazon Athena 的用户实施表级和列级粒度的访问控制。

主题

授予数据位置权限

AWS Lake Formation 中的数据位置权限使主体能够创建和更改指向指定已注册的 Amazon S3 位置的数据目录资源。数据位置权限与 Lake Formation 数据权限搭配使用，共同保护数据湖中的信息。

Lake Formation 不使用 AWS Resource Access Manager (AWS RAM) 服务授予数据位置权限，因此您无需接受资源共享邀请即可获得数据位置权限。

您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 授予数据位置权限。

> **Note**
>
> 要使授权成功，您必须先向 Lake Formation 注册数据位置。

另请参见：

- Underlying data access control

主题

授予数据位置权限（同一账户）

按照以下步骤向 AWS 账户中的主体授予数据位置权限。您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 授予权限。

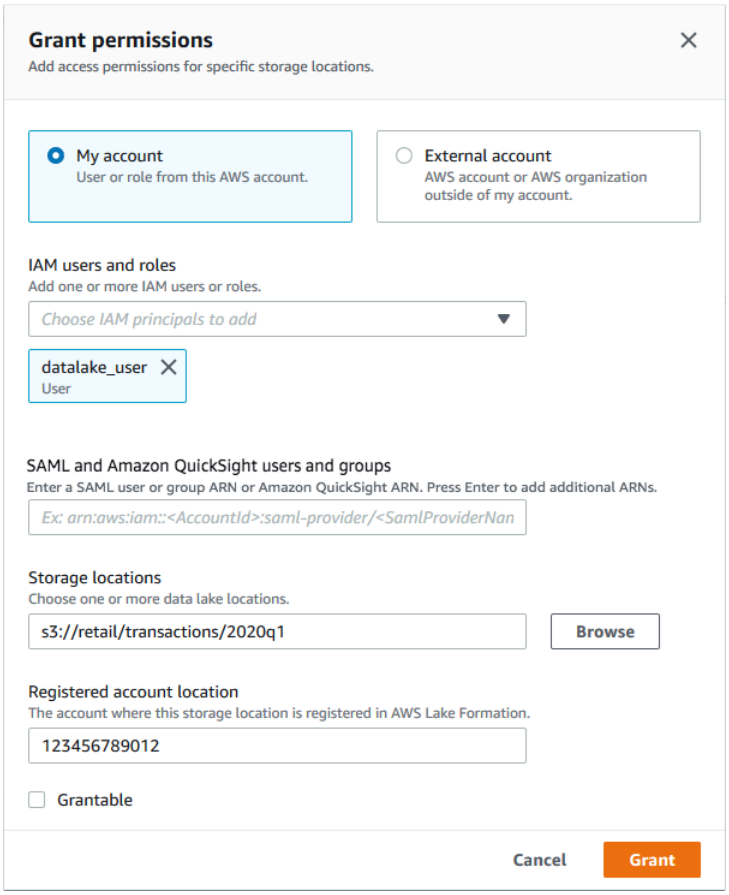
授予数据位置权限（同一账户）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以数据湖管理员或对所需数据位置具有授予权限的主体的身份登录。
2. 在导航窗格中，选择数据位置。
3. 选择授予。
4. 在授予权限对话框中，确保选中我的账户磁贴。然后提供以下信息：

   - 对于 IAM 用户和角色，请选择一个或多个主体。
   - 对于 SAML 和 Amazon QuickSight 用户和组，为通过 SAML 联合的用户或组输入一个或多个 Amazon 资源名称 (ARN)，或者为 Amazon QuickSight 用户或组输入 ARN。

     一次输入一个 ARN，然后在每个 ARN 后按 Enter。有关如何构建 ARN 的信息，请参阅 Lake Formation 授予和撤销命令 AWS CLI。
   - 对于存储位置，选择浏览，然后选择一个 Amazon Simple Storage Service (Amazon S3) 存储位置。该位置必须在 Lake Formation 中注册。再次选择浏览以添加其他位置。您也可以键入位置，但请确保在位置前面加上 `s3://`。
   - 对于注册的账户位置，输入注册该位置的 AWS 账户 ID。这默认为您的账户 ID。在跨账户场景中，接收者账户中的数据湖管理员在向接收者账户中的其他主体授予数据位置权限时，可以在此处指定拥有者账户。

- （可选）要使所选主体能够授予对所选位置的数据位置权限，请选择可授予。



5. 选择授权。

授予数据位置权限（同一账户，使用 AWS CLI）

- 运行 `grant-permissions` 命令，向主体授予 `DATA_LOCATION_ACCESS` 权限，并将 Amazon S3 路径指定为资源。

  Example

  以下示例向用户 `datalake_user1` 授予对 `s3://retail` 的数据位置权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail"}}'
```

Example

以下示例向 ALLIAMPrincipals 组授予对 s3://retail 的数据位置权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

另请参见：

- Lake Formation 权限参考

授予数据位置权限（外部账户）

按照以下步骤向外部 AWS 账户或组织授予数据位置权限。

您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 授予权限。

开始之前

确保满足所有跨账户访问先决条件。有关更多信息，请参阅先决条件。

授予数据位置权限（外部账户，使用控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台。以数据湖管理员身份登录。
2. 在导航窗格中，选择数据位置，然后选择授予。
3. 在授予权限对话框中，选择外部账户磁贴。
4. 提供以下信息：

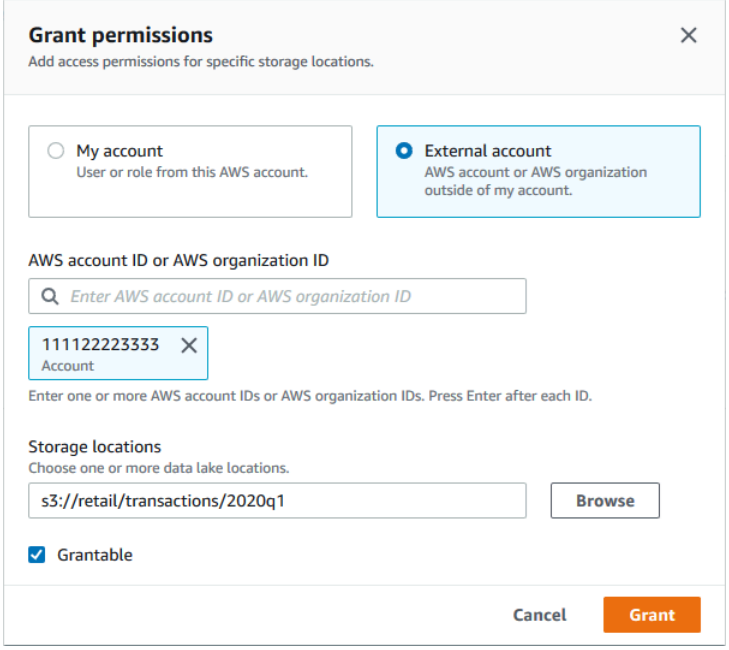
   - 对于 AWS 账户 ID 或 AWS 组织 ID，输入有效的 AWS 账号、组织 ID 或组织单位 ID。

     在每个 ID 后按 Enter。

     组织 ID 由“o-”后跟 10 到 32 个小写字母或数字组成。

组织单位 ID 由“ou-”后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）组成。此字符串后跟第二个“-”（连字符）和 8 到 32 个额外的小写字母或数字。

- 在存储位置下，选择浏览，然后选择一个 Amazon Simple Storage Service (Amazon S3) 存储位置。该位置必须在 Lake Formation 中注册。



5. 选择可授予。
6. 选择授权。

授予数据位置权限（外部账户，使用 AWS CLI）

- 要向外部 AWS 账户授予权限，请输入类似如下的命令。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333  --permissions "DATA_LOCATION_ACCESS"
 --permissions-with-grant-option "DATA_LOCATION_ACCESS" --resource
 '{ "DataLocation": {"CatalogId":"123456789012","ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

此命令使用授予选项向账户 1111-2222-3333 授予对 Amazon S3 位置 `s3://retail/transactions/2020q1` 的 `DATA_LOCATION_ACCESS` 权限，该位置由账户 1234-5678-9012 拥有。

要向组织授予权限，请输入类似如下的命令。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "DATA_LOCATION_ACCESS" --permissions-
with-grant-option "DATA_LOCATION_ACCESS" --resource '{"DataLocation":
 {"CatalogId":"123456789012","ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

此命令使用授予选项向组织 o-abcdefghijkl 授予对 Amazon S3 位置 s3://retail/transactions/2020q1 的 DATA_LOCATION_ACCESS 权限，该位置由账户 1234-5678-9012 拥有。

要向外部 AWS 账户中的主体授予权限，请输入类似如下的命令。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3::retail/transactions/2020q1", "CatalogId":
 "123456789012"}}'
```

此命令使用授予选项向账户 1111-2222-3333 中的主体授予对 Amazon S3 位置 s3://retail/transactions/2020q1 的 DATA_LOCATION_ACCESS 权限，该位置由账户 1234-5678-9012 拥有。

Example

以下示例向外部账户中的 ALLIAMPrincipals 组授予对 s3://retail 的数据位置权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}}'
```

另请参见：

- Lake Formation 权限参考

授予对与您的账户共享的数据位置的权限

与您的 AWS 账户共享数据目录资源后，作为数据湖管理员，您可以向账户中的其他主体授予对该资源的权限。如果授予了对共享表的 ALTER 权限，并且该表指向已注册的 Amazon S3 位置，则您还必须授予对该位置的数据位置权限。同样，如果对共享数据库授予了 CREATE_TABLE 或 ALTER 权限，并且该数据库具有指向已注册位置的位置属性，则您还必须授予对该位置的数据位置权限。

要向您账户中的主体授予对共享位置的数据位置权限，您的账户必须已通过授予选项被授予对该位置的 DATA_LOCATION_ACCESS 权限。当您向账户中的其他主体授予 DATA_LOCATION_ACCESS 权限，必须包括拥有者账户的数据目录 ID（AWS 账户 ID）。拥有者账户是注册了该位置的账户。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 授予数据位置权限。

授予对与您的账户共享的数据位置的权限（控制台）

- 按照授予数据位置权限（同一账户）中的步骤操作。

  对于存储位置，必须键入存储位置。对于注册的账户位置，输入拥有者账户的 AWS 账户 ID。

授予对与您的账户共享的数据位置的权限 (AWS CLI)

- 输入以下命令之一以向用户或角色授予权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
```

授予和撤销对数据目录资源的权限

您可以向 AWS Lake Formation 中的主体授予数据湖权限，以便主体可以创建和管理数据目录资源，并且可以访问基础数据。您可以授予对数据库、表和视图的数据湖权限。授予对表的权限时，您可以限制对特定表列或行的访问权限，从而实现更精细的访问控制。

您可以授予对单个表或视图的权限，也可以通过一次授予操作来授予对数据库中所有表和视图的权限。如果您授予对数据库中所有表的权限，则将隐式授予对数据库的 DESCRIBE 权限。然后，数据库将显示在控制台的数据库页面上，由 GetDatabases API 操作返回。

您可以使用命名资源方法或 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法来授予权限。

您可以向同一 AWS 账户中的主体或外部账户或组织授予权限。当您向外部账户或组织授予权限时，将与这些账户或组织共享您拥有的资源。然后，这些账户或组织中的主体可以访问您拥有的数据目录资源和基础数据。

> **Note**
>
> 目前，LF-TBAC 方法支持向 IAM 主体、AWS 账户、Organizations 和组织单位（OU）授予跨账户权限。

向外部账户或组织授予权限时，您必须包括授予选项。只有外部账户中的数据湖管理员才能访问共享资源，直到管理员向外部账户中的其他主体授予对共享资源的权限。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予数据目录权限。

> **Note**
>
> 删除数据目录资源时，与该资源关联的所有权限都将失效。重新创建同名的相同资源将无法恢复 Lake Formation 权限。用户必须重新设置新权限。

另请参阅：

- 跨 AWS 账户共享数据目录表和数据库
- 元数据访问控制
- Lake Formation 权限参考

授予或撤销 Lake Formation 权限所需的 IAM 权限

所有主体（包括数据湖管理员）都需要以下 AWS Identity and Access Management (IAM) 权限才能使用 Lake Formation API 或 AWS CLI 授予或者撤销 AWS Lake Formation 数据目录权限或数据位置权限：

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable` 或 `glue:GetDatabase`，用于使用命名资源方法授予权限的表或数据库。

> **Note**
>
> 数据湖管理员具有隐式 Lake Formation 权限，可以授予和撤销 Lake Formation 权限。但是他们仍然需要对 Lake Formation 授予和撤销 API 操作的 IAM 权限。
> 具有 `AWSLakeFormationDataAdmin` AWS 托管式策略的 IAM 角色无法添加新的数据湖管理员，因为此策略包含对 Lake Formation API 操作 `PutDataLakeSetting` 的显式拒绝。

对于非数据湖管理员且想要使用 Lake Formation 控制台授予或撤销权限的主体，建议使用以下 IAM 策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:ListPermissions",
                "lakeformation:GrantPermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchRevokePermissions",
                "glue:GetDatabases",
                "glue:SearchTables",
                "glue:GetTables",
                "glue:GetDatabase",
```

```
                "glue:GetTable",
                "iam:ListUsers",
                "iam:ListRoles",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "sso:DescribeInstance"
            ],
            "Resource": "*"
        }
    ]
}
```

此策略中的所有 `glue:` 和 `iam:` 权限均可在 AWS 托管式策略 `AWSGlueConsoleFullAccess` 中使用。

要使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 授予权限，主体需要额外的 IAM 权限。有关更多信息，请参阅 Lake Formation 基于标签的访问控制最佳实践和注意事项 和 Lake Formation 角色和 IAM 权限参考：

跨账户权限

想要使用命名资源方法授予跨账户 Lake Formation 权限的用户还必须具有 `AWSLakeFormationCrossAccountManager` AWS 托管式策略中的权限。

数据湖管理员需要相同的权限才能授予跨账户权限，还需要 AWS Resource Access Manager (AWS RAM) 权限才能允许向组织授予权限。有关更多信息，请参阅数据湖管理员权限。

管理用户

具有管理权限（例如采用 `AdministratorAccess` AWS 托管式策略）的主体有权授予 Lake Formation 权限和创建数据湖管理员。要拒绝用户或角色访问 Lake Formation 管理员操作，请在其策略中附加或添加管理员 API 操作的 `Deny` 语句。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "lakeformation:GetDataLakeSettings",
                "lakeformation:PutDataLakeSettings"
            ],
            "Effect": "Deny",
```

```
            "Resource": [
                "*"
            ]
        }
    ]
}
```

> **Important**
>
> 要防止用户使用提取、转换和加载 (ETL) 脚本将自己添加为管理员，请确保拒绝所有非管理员用户和角色访问这些 API 操作。`AWSLakeFormationDataAdmin` AWS 托管式策略包含对 Lake Formation API 操作 `PutDataLakeSetting` 的显式拒绝，该操作可阻止用户添加新的数据湖管理员。

使用命名资源方法授予数据湖权限

您可以使用命名资源方法授予 Lake Formation 对特定数据目录数据库、表和视图的权限。您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予权限。

主题

使用命名资源方法授予数据库权限

以下步骤说明如何使用命名资源方法授予数据库权限。

Console

使用 Lake Formation 控制台上的授予数据湖权限页面。该页面分为以下几个部分：

- 主体 – 授予权限的 IAM 用户、角色、IAM Identity Center 用户和组、SAML 用户和组、AWS 账户、Organizations 或组织单位。
- LF 标签或目录资源 – 要对其授予权限的数据库、表、视图或资源链接。
- 权限 – 要授予的 Lake Formation 权限。

> **Note**
>
> 要授予对数据库资源链接的权限，请参阅授予资源链接权限。

1. 打开授予数据湖权限页面。

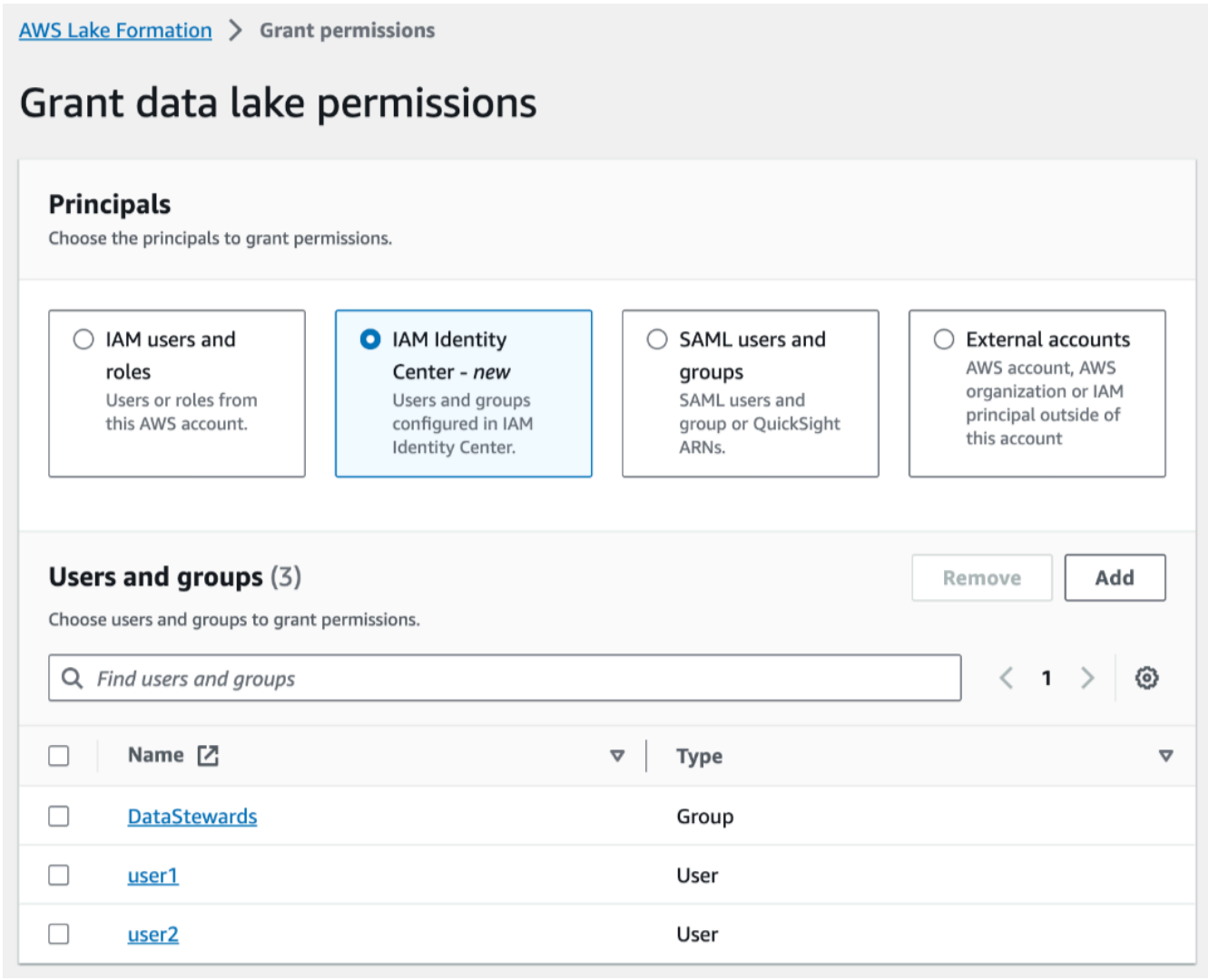
   通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台，然后以数据湖管理员、表创建者或对数据库具有可授予的权限的 IAM 用户身份登录。

   请执行以下操作之一：

   - 在导航窗格的权限下，选择数据湖权限。然后选择授予。
   - 在导航窗格的数据目录下，选择数据库。然后在数据库页面上选择一个数据库，并在操作菜单的权限下选择授予。

   > **Note**
   >
   > 您可以通过数据库的资源链接授予对数据库的权限。为此，在数据库页面上选择一个资源链接，然后在操作菜单上选择对目标的授权。有关更多信息，请参阅资源链接在 Lake Formation 中的工作原理。

2. 接下来，在主体部分中，选择主体类型，然后指定要授予权限的主体。



IAM 用户和角色

从 IAM 用户和角色列表中选择一个或多个用户或角色。

IAM Identity Center

从用户和组列表中选择一个或多个用户或组。选择添加以添加更多用户或组。

SAML 用户和组

对于 SAML 和 Amazon QuickSight 用户和组，为通过 SAML 联合的用户或组输入一个或多个 Amazon 资源名称 (ARN)，或者为 Amazon QuickSight 用户或组输入 ARN。在每个 ARN 后按 Enter。

有关如何构建 ARN 的信息，请参阅 Lake Formation 授予和撤销命令 AWS CLI。

> **Note**
>
> 只有 Amazon QuickSight 企业版支持 Lake Formation 与 Amazon QuickSight 的集成。

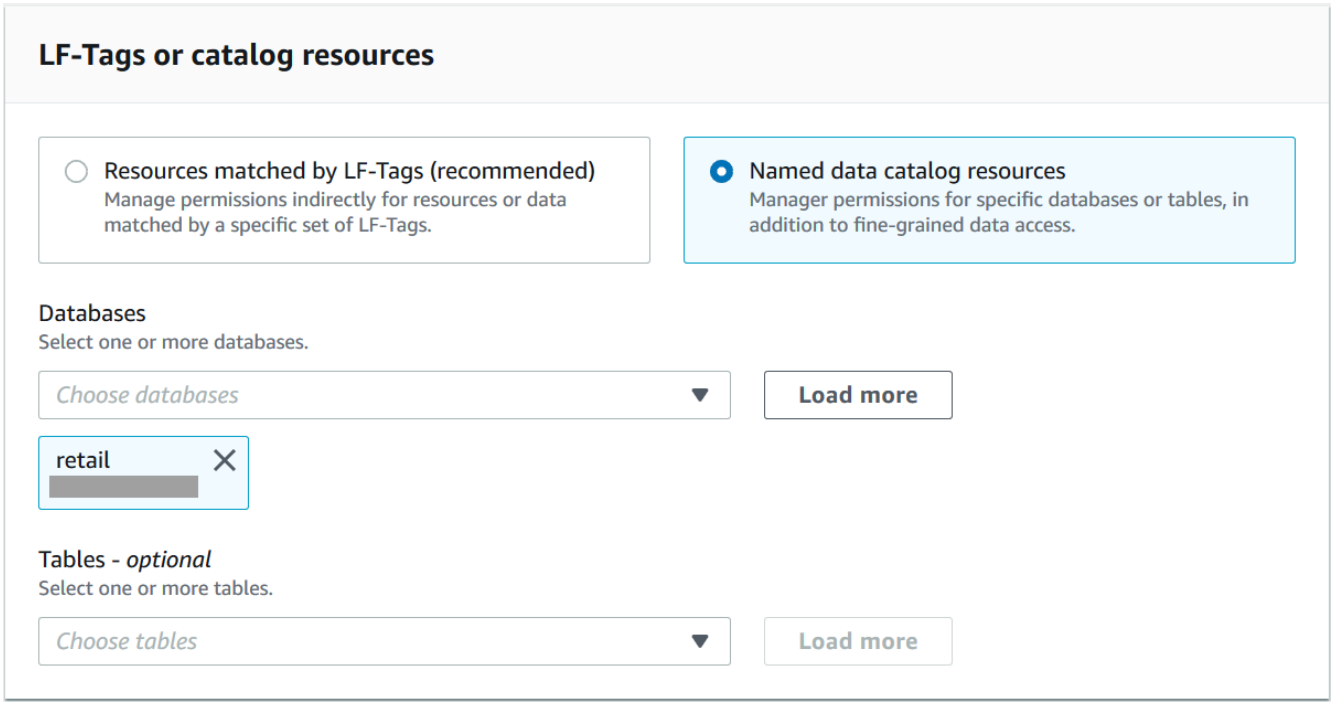
外部账户

对于 AWS 账户、AWS Organizations 或 IAM 主体，为 IAM 用户或角色输入一个或多个有效的 AWS 账户 ID、组织 ID、组织单位 ID 或 ARN。在每个 ID 后按 Enter。

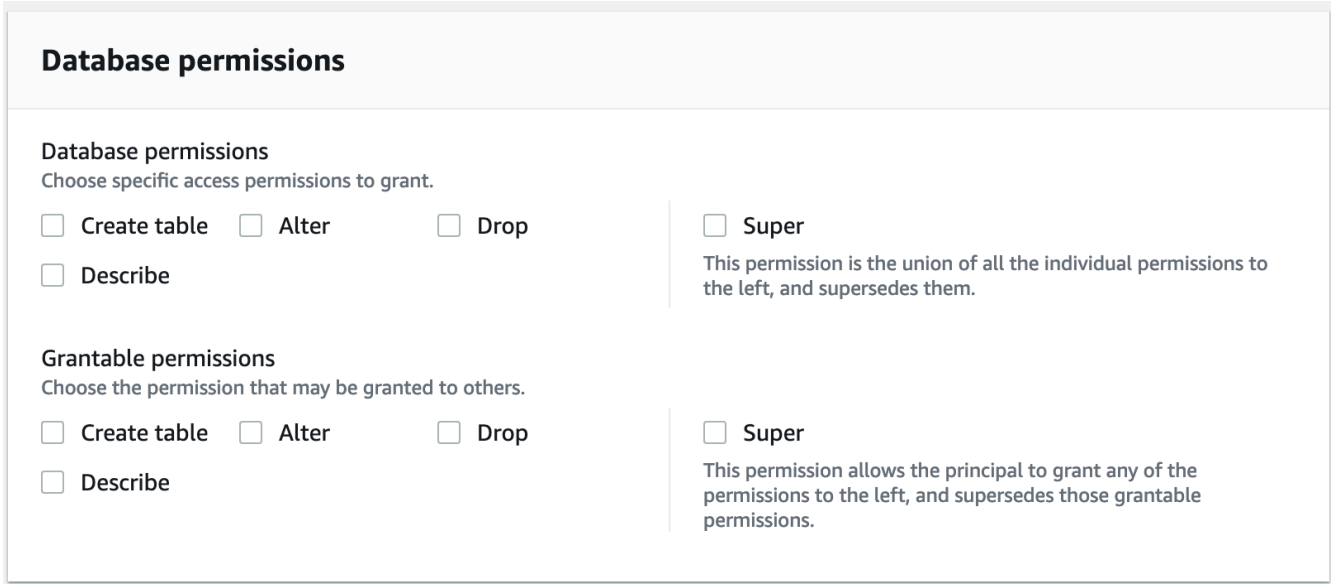
组织 ID 由“o-”后跟 10 到 32 个小写字母或数字组成。

单位 ID 以“ou-”开头，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。该字符串后跟第二个“-”短横线和 8 到 32 个额外的小写字母或数字。

3. 在 LF 标签或目录资源部分下，选择已命名数据目录资源。



4. 从数据库列表中选择一个或多个数据库。您还可以选择一个或多个表和/或数据筛选条件。
5. 在权限部分中，选择权限和可授予的权限。在数据库权限下，选择一项或多项要授予的权限。



> **Note**
>
> 在授予对具有指向注册位置的位置属性的数据库的 `Create Table` 或 `Alter` 权限后，请确保还向主体授予对该位置的数据位置权限。有关更多信息，请参阅[授予数据位置权限]。

6. （可选）在可授予的权限下，选择授予接收人可以向其 AWS 账户中的其他主体授予的权限。当您从外部账户向 IAM 主体授予权限时，不支持此选项。
7. 选择授予。

AWS CLI

您可以使用命名资源方法和 AWS Command Line Interface (AWS CLI) 来授予数据库权限。

使用 AWS CLI 授予数据库权限

- 运行 `grant-permissions` 命令，并根据所授予的权限将数据库或数据目录指定为资源。

  将以下示例中的 *<account-id>* 替换为有效的 AWS 账户 ID。

  Example – 授予创建数据库的权限

  此示例向用户 `datalake_user1` 授予 `CREATE_DATABASE` 权限。由于被授予此权限的资源是数据目录，因此该命令会将空的 `CatalogResource` 结构指定为 `resource` 参数。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Example – 授予在指定数据库中创建表的权限

下一个示例向用户 datalake_user1 授予对数据库 retail 的 CREATE_TABLE 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example – 使用“授予”选项向外部 AWS 账户授予权限

下一个示例使用授予选项向外部账户 1111-2222-3333 授予对数据库 retail 的 CREATE_TABLE 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
 --permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
 {"Name":"retail"}}'
```

Example – 向组织授予权限

下一个示例使用授予选项向组织 o-abcdefghijkl 授予对数据库 issues 的 ALTER 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "ALTER" --permissions-with-grant-option "ALTER" --
resource '{ "Database": {"Name":"issues"}}'
```

Example - 向同一账户中的 **ALLIAMPrincipals** 授予权限

下一个示例向同一账户中的所有主体授予对数据库 retail 的 CREATE_TABLE 权限。选择此选项后，账户中的每个主体即可在数据库中创建表并创建表资源链接，从而允许集成查询引擎访问共享数据库和表。当主体获得跨账户授权但无权创建资源链接时，此选项尤为有用。在这种情况下，数据湖管理员可以创建占位符数据库并向 ALLIAMPrincipal 组授予 CREATE_TABLE 权限，从而使账户中的每个 IAM 主体都能在占位符数据库中创建资源链接。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"temp","CatalogId":"111122223333"}}'
```

Example - 向外部账户中的 **ALLIAMPrincipals** 授予权限

下一个示例向外部账户中的所有主体授予对数据库 `retail` 的 CREATE_TABLE 权限。选择此选项后，账户中的每个主体即可在数据库中创建表。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"retail","CatalogId":"123456789012"}}'
```

> **Note**
>
> 在授予对具有指向注册位置的位置属性的数据库的 CREATE_TABLE 或 ALTER 权限后，请确保还向主体授予对该位置的数据位置权限。有关更多信息，请参阅授予数据位置权限。

> **另请参阅**
>
> - Lake Formation 权限参考
> - 授予对与您的账户共享的数据库或表的权限
> - 访问和查看共享数据目录表和数据库

使用命名资源方法授予表权限

您可以使用 Lake Formation 控制台或 AWS CLI 授予对数据目录表的 Lake Formation 权限。您可以授予对单个表的权限，也可以通过一次授予操作来授予对数据库中所有表的权限。

如果您授予对数据库中所有表的权限，则将隐式授予对数据库的 DESCRIBE 权限。然后，数据库将显示在控制台的数据库页面上，由 GetDatabases API 操作返回。

当您选择 SELECT 作为要授予的权限时，可以选择应用列筛选条件、行筛选条件或单元格筛选条件。

Console

以下步骤说明如何使用命名资源方法和 Lake Formation 控制台上的授予数据湖权限页面来授予表权限。该页面分为以下几个部分：

- 主体 – 要向其授予权限的用户、角色、AWS 账户、组织或组织单位。
- LF 标签或目录资源 – 要对其授予权限的数据库、表或资源链接。
- 权限 – 要授予的 Lake Formation 权限。

> **Note**
>
> 要授予对表资源链接的权限，请参阅授予资源链接权限。

1. 打开“授予数据湖权限”页面。

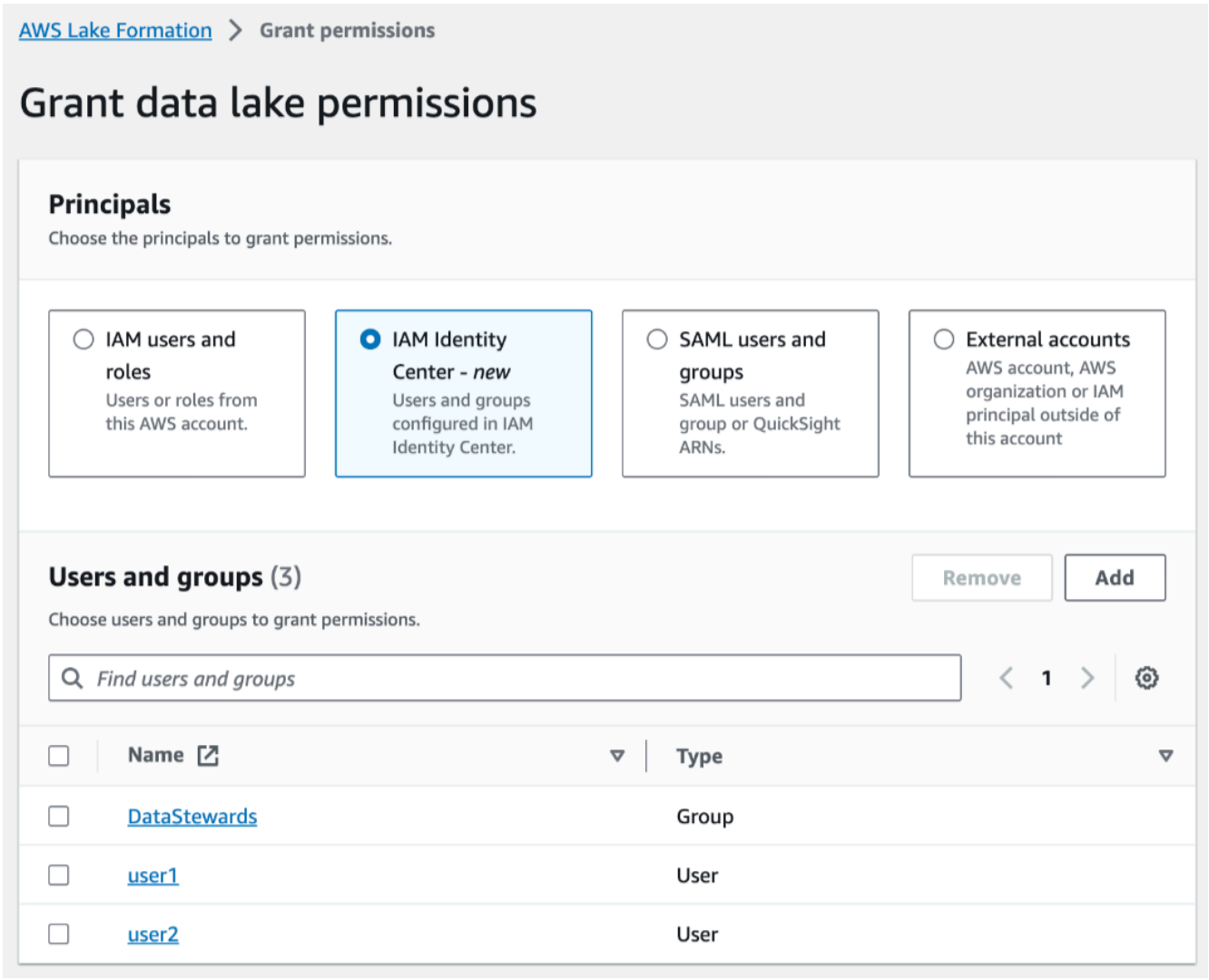
   通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台，然后以数据湖管理员、表创建者或已通过授予选项获得对表的权限的用户身份登录。

   请执行以下操作之一：

   - 在导航窗格的权限下，选择数据湖权限。然后选择授予。
   - 在导航窗格中，选择表。然后在表页面上选择一个表，并在操作菜单的权限下选择授予。

   > **Note**
   >
   > 您可以通过表的资源链接授予对表的权限。为此，在表页面上选择一个资源链接，然后在操作菜单上选择对目标的授权。有关更多信息，请参阅资源链接在 Lake Formation 中的工作原理。

2. 接下来，在主体部分中，选择主体类型，然后指定要向其授予权限的主体。



IAM 用户和角色

从 IAM 用户和角色列表中选择一个或多个用户或角色。

IAM Identity Center

从用户和组列表中选择一个或多个用户或组。

SAML 用户和组

对于 SAML 和 Amazon QuickSight 用户和组，为通过 SAML 联合的用户或组输入一个或多个 Amazon 资源名称 (ARN)，或者为 Amazon QuickSight 用户或组输入 ARN。在每个 ARN 后按 Enter。

有关如何构建 ARN 的信息，请参阅 Lake Formation 授予和撤销命令 AWS CLI。

> **Note**
>
> 只有 Amazon QuickSight 企业版支持 Lake Formation 与 Amazon QuickSight 的集成。

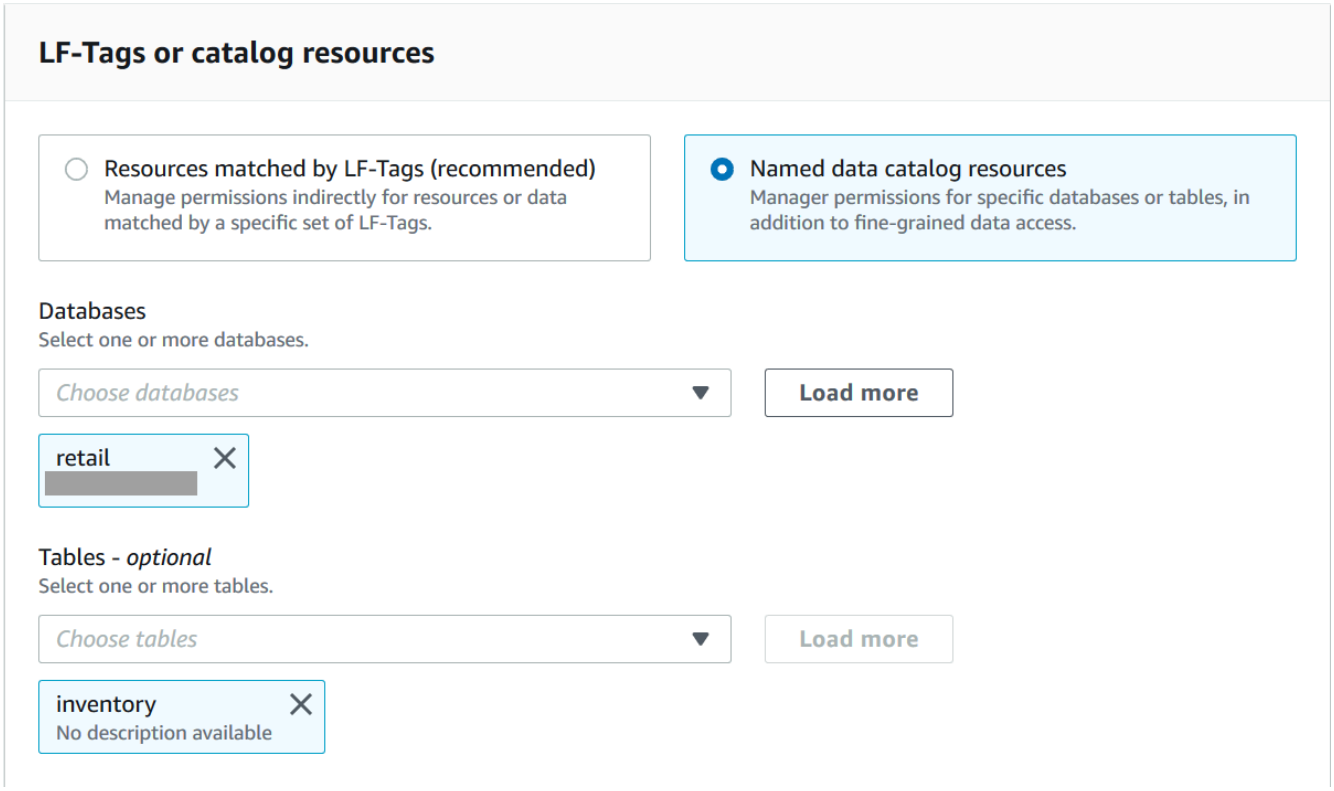
外部账户

对于 AWS 账户、AWS Organizations 或 IAM 主体，为 IAM 用户或角色输入一个或多个有效的 AWS 账户 ID、组织 ID、组织单位 ID 或 ARN。在每个 ID 后按 Enter。

组织 ID 由“o-”后跟 10 到 32 个小写字母或数字组成。

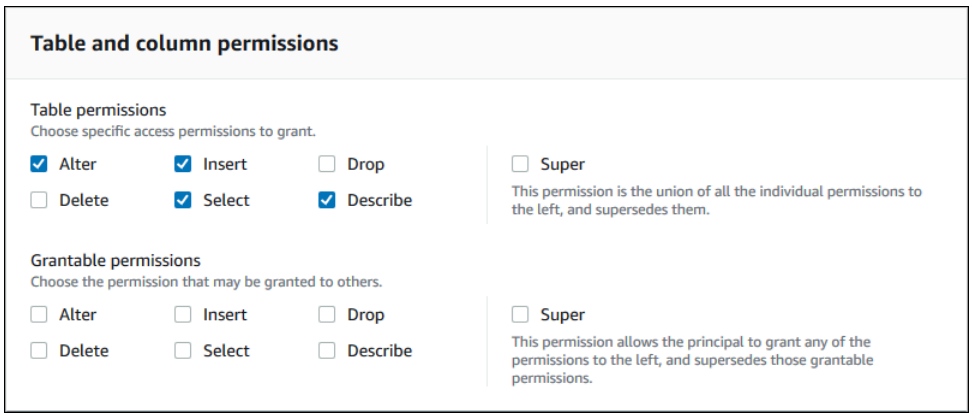
单位 ID 以“ou-”开头，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。该字符串后跟第二个“-”字符和 8 到 32 个额外的小写字母或数字。

3. 在 LF 标签或目录资源部分中，选择一个数据库。然后选择一个或多个表或所有表。

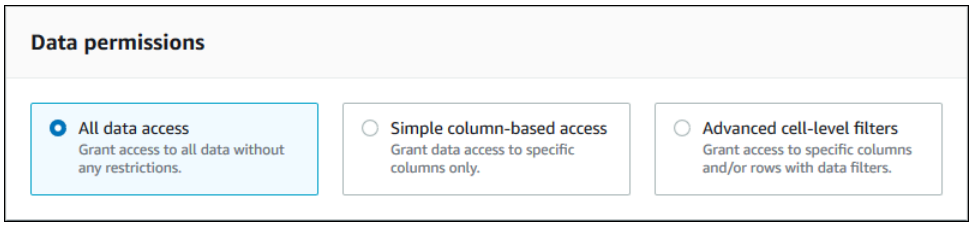


4. 指定权限（无数据筛选）

在权限部分中，选择要授予的表权限，也可以选择可授予的权限。



如果授予 Select 权限，则数据权限部分将显示在表和列权限部分下方，其中所有数据访问选项默认处于选中状态。接受默认设置。



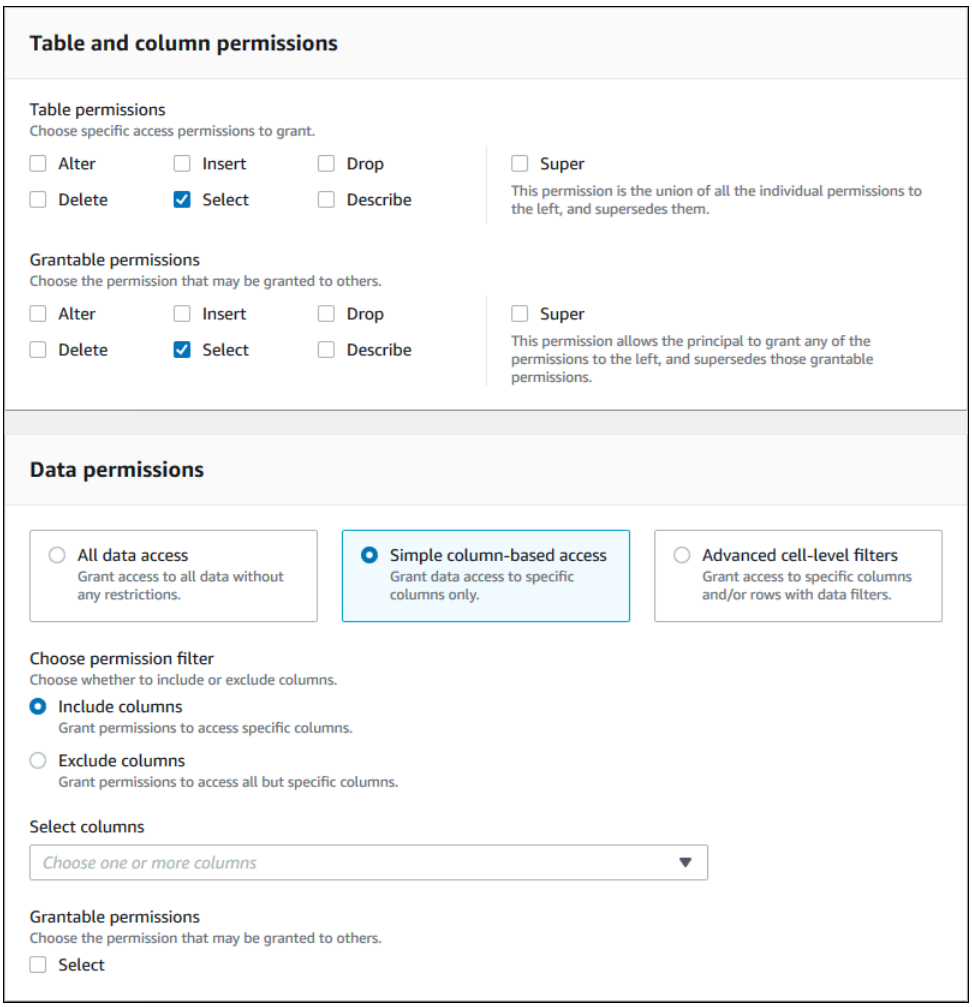
5. 选择授予。
6. 通过数据筛选指定 Select 权限

   选择 Select 权限。不要选择任何其他权限。

   数据权限部分将显示在表和列权限部分下方。
7. 请执行以下操作之一：

   - 仅应用简单列筛选。

     1. 选择基于列的简单访问。



2. 选择是包括还是排除列，然后选择要包括或排除的列。

   向外部 AWS 账户或组织授予权限时，仅支持包括列表。

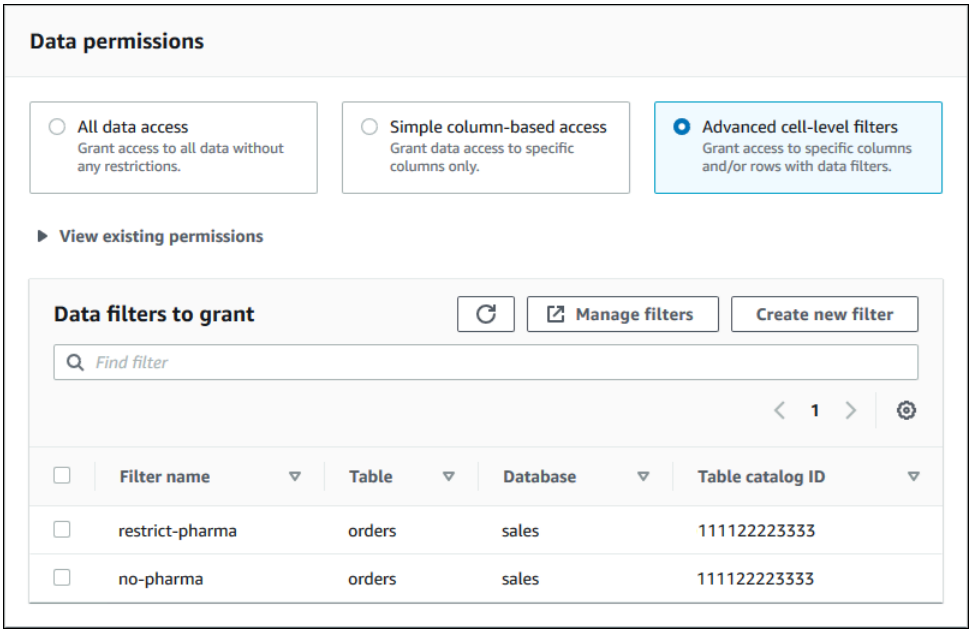
3. （可选）在可授予的权限下，开启“Select”权限的授予选项。

   如果包括授予选项，则授予接收人只能授予对您授予给他们的列的权限。

> **Note**
>
> 您也可以通过创建指定列筛选条件并将所有行指定为行筛选条件的数据筛选条件来应用列筛选。但是，这需要更多的步骤。

- 应用列、行或单元格筛选。

  1. 选择高级单元格级筛选条件。



2. （可选）展开查看现有权限。
3. （可选）选择创建新筛选条件。
4. （可选）要查看列出的筛选条件的详细信息，或者要创建新筛选条件或删除现有筛选条件，请选择管理筛选条件。

   此时将在新的浏览器窗口中打开数据筛选条件页面。

   完成数据筛选条件页面上完成操作后，返回到授予权限页面，如有必要，请刷新该页面以查看创建的任何新的数据筛选条件。
5. 选择一个或多个要应用于授予的数据筛选条件。

> **Note**
>
> 如果列表中没有数据筛选条件，则表示未为所选表创建任何数据筛选条件。

8. 选择授权。

AWS CLI

您可以使用命名资源方法和 AWS Command Line Interface (AWS CLI) 来授予表权限。

使用 AWS CLI 授予表权限

- 运行 `grant-permissions` 命令并指定表作为资源。

Example – 对单个表授予权限 - 无筛选

以下示例向 AWS 账户 1111-2222-3333 中的用户 datalake_user1 授予对数据库 retail 中表 inventory 的 SELECT 和 ALTER 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

> **Note**
>
> 如果您授予对其基础数据位于注册位置的表的 ALTER 权限，请确保还向主体授予对该位置的数据位置权限。有关更多信息，请参阅授予数据位置权限。

Example – 使用“授予”选项对所有表授予权限 - 无筛选

下一个示例使用授予选项授予对数据库 retail 中所有表的 SELECT 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --permissions-with-grant-option "SELECT" --resource '{ "Table":
 { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Example – 通过简单列筛选授予权限

下一个示例授予对表 persons 中一部分列的 SELECT 权限。它使用简单列筛选。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"hr",
 "Name":"persons", "ColumnNames":["family_name", "given_name", "gender"]}}'
```

Example – 使用数据筛选条件授予权限

此示例授予对 orders 表的 SELECT 权限并应用 restrict-pharma 数据筛选条件。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是 grant-params.json 文件的内容。

```
{
    "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["SELECT"],
    "PermissionsWithGrantOption": ["SELECT"]
}
```

另请参阅

- Lake Formation 权限概述
- Lake Formation 中的数据筛选和单元格级别安全性
- Lake Formation 角色和 IAM 权限参考
- 授予资源链接权限
- 访问和查看共享数据目录表和数据库

使用命名资源方法授予对视图的权限

以下步骤说明如何使用命名资源方法和授予数据湖权限页面来授予对视图的权限。该页面分为以下几个部分：

- 主体 – 授予权限的 IAM 用户、角色、IAM Identity Center 用户和组、AWS 账户、Organizations 或组织单位。
- LF 标签或目录资源 – 要对其授予权限的数据库、表、视图或资源链接。
- 权限 – 要授予的数据湖权限。

打开授予数据湖权限页面

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台，然后以数据湖管理员、表创建者或对数据库具有可授予的权限的 IAM 用户身份登录。
2. 请执行以下操作之一：

   - 在导航窗格的权限下，选择数据湖权限。然后选择授予。
   - 在导航窗格的数据目录下，选择视图。然后在视图页面上选择一个表，并在操作菜单的权限下选择授予。

> **Note**
>
> 您可以通过视图的资源链接授予对视图的权限。为此，在视图页面上选择一个资源链接，然后在操作菜单上选择对目标的授权。有关更多信息，请参阅资源链接在 Lake Formation 中的工作原理。

指定主体

在主体部分中，选择主体类型，然后指定要授予权限的主体。

IAM 用户和角色

从 IAM 用户和角色列表中选择一个或多个用户或角色。

IAM Identity Center

从用户和组列表中选择一个或多个用户或组。

SAML 用户和组

对于 SAML 和 Amazon QuickSight 用户和组，为通过 SAML 联合的用户或组输入一个或多个 Amazon 资源名称 (ARN)，或者为 Amazon QuickSight 用户或组输入 ARN。在每个 ARN 后按 Enter。

有关如何构建 ARN 的信息，请参阅 Lake Formation 授予和撤销命令 AWS CLI。

> **Note**
>
> 只有 Amazon QuickSight 企业版支持 Lake Formation 与 Amazon QuickSight 的集成。

外部账户

对于 AWS 账户、AWS Organizations 或 IAM 主体，为 IAM 用户或角色输入一个或多个有效的 AWS 账户 ID、组织 ID、组织单位 ID 或 ARN。在每个 ID 后按 Enter。

组织 ID 由“o-”后跟 10 到 32 个小写字母或数字组成。

单位 ID 以“ou-”开头，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。该字符串后跟第二个“-”短横线和 8 到 32 个额外的小写字母或数字。

> 另请参阅
>
> - 访问和查看共享数据目录表和数据库

指定视图

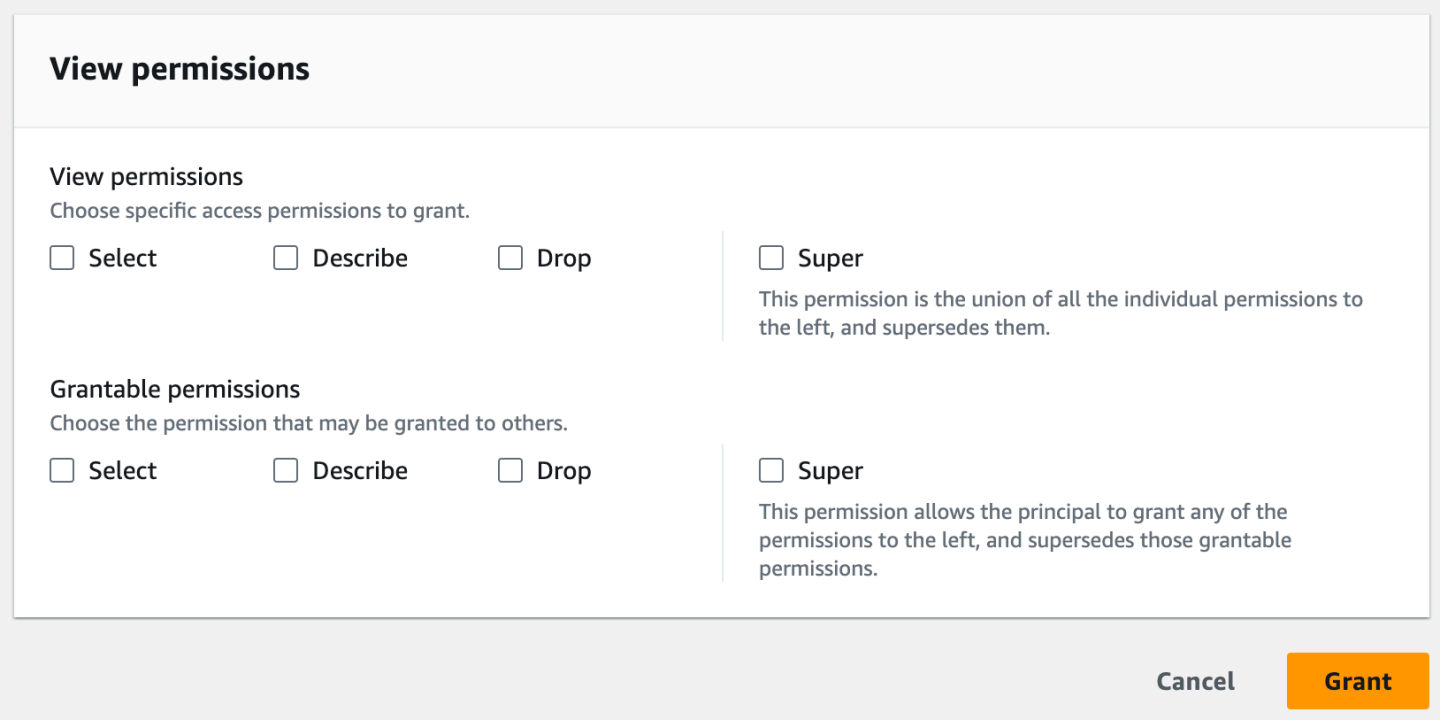
在 LF 标签或目录资源部分中，选择一个或多个要对其授予权限的视图。

1. 选择命名数据目录资源。
2. 从视图列表中选择一个或多个视图。您还可以选择一个或多个“数据库”、“表”和/或“数据筛选条件”。

   向数据库内的 `All views` 授予数据湖权限将导致被授权者具有对数据库内所有表和视图的权限。

指定权限

在权限部分中，选择权限和可授予的权限。



1. 在视图权限下，选择一项或多项要授予的权限。
2. （可选）在可授予的权限下，选择授予接收人可以向其 AWS 账户中的其他主体授予的权限。当您从外部账户向 IAM 主体授予权限时，不支持此选项。
3. 选择授予。

> **另请参阅**
>
> - Lake Formation 权限参考
> - 授予对与您的账户共享的数据库或表的权限

Lake Formation 基于标签的访问控制

Lake Formation 基于标签的访问控制 (LF-TBAC) 是一种基于属性定义权限的授权策略。在 Lake Formation 中，这些属性被称为“LF 标签”。您可以将 LF 标签附加到数据目录资源，并使用这些 LF 标签向 Lake Formation 委托人授予对这些资源的权限。当委托人的标签值与资源标签值匹配时，Lake Formation 允许对这些资源进行操作。LF-TBAC 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。

当有大量数据目录资源时，建议使用 LF-TBAC 来授予 Lake Formation 权限。LF-TBAC 比命名资源方法更具可扩展性，并且需要的权限管理开销更少。

> **Note**
>
> IAM 标签与 LF 标签不同。这些标签不可互换。LF 标签用于授予 Lake Formation 权限，而 IAM 标签用于定义 IAM 策略。

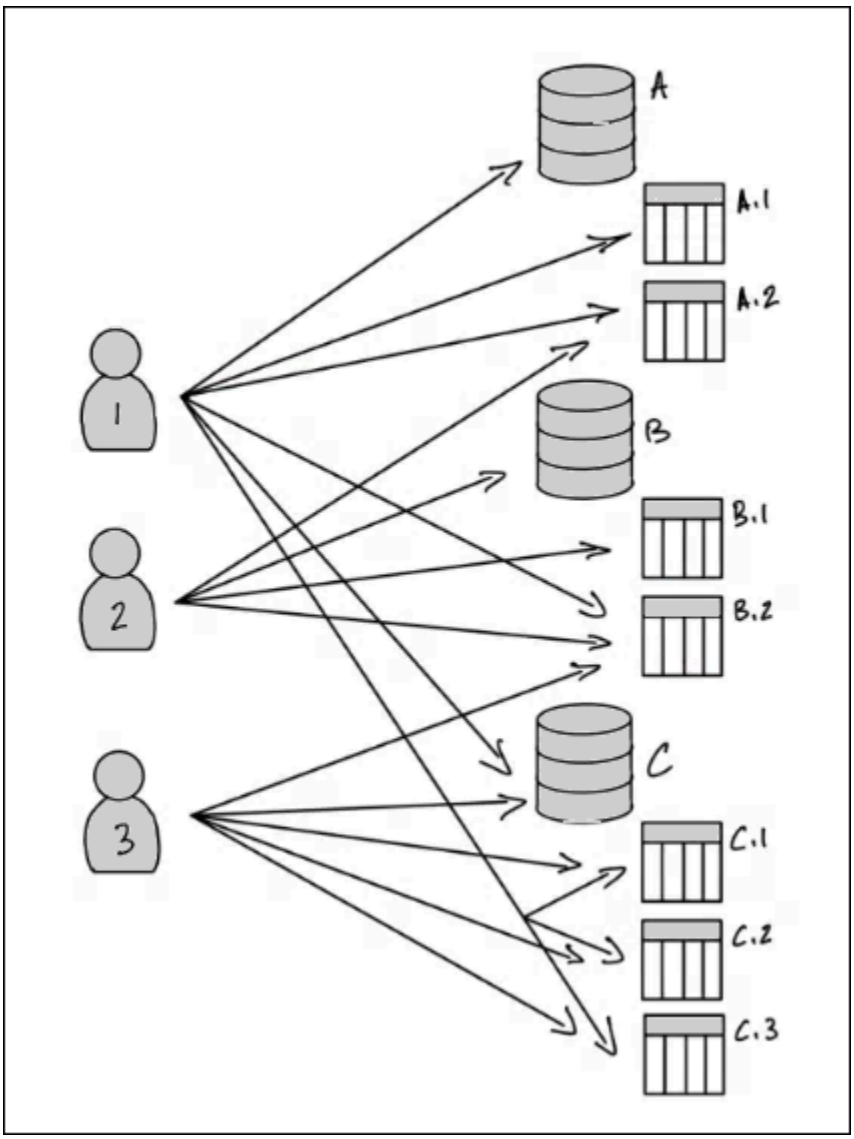
Lake Formation 基于标签的访问控制的工作原理

每个 LF 标签都是一个键值对，例如 `department=sales` 或 `classification=restricted`。一个键可以有多个定义的值，例如 `department=sales,marketing,engineering,finance`。

要使用 LF-TBAC 方法，数据湖管理员和数据工程师需要执行以下任务。

任务	任务详细信息
1. 定义 LF 标签的属性和关系。	-
2. 在 Lake Formation 中创建 LF 标签创建者。	添加 LF 标签创建者
3. 在 Lake Formation 中创建 LF 标签。	创建 LF 标签
4. 将 LF 标签分配给数据目录资源。	将 LF 标签分配给数据目录资源
5. 向其他主体授予权限以将 LF 标签分配给资源，也可以使用授予选项。	授予、撤销和列出 LF 标签值权限
6. 向主体授予 LF 标签表达式，也可以使用授予选项。	使用 LF-TBAC 方法授予数据湖权限
7.（推荐）通过 LF-TBAC 方法验证主体是否有权访问正确的资源后，撤销使用命名资源方法授予的权限。	-

请考虑以下情况：您必须向三个主体授予对三个数据库和七个表的权限。



要使用命名资源方法实现上图中指示的权限，您必须进行 17 次授予，如下所示（使用伪代码）。

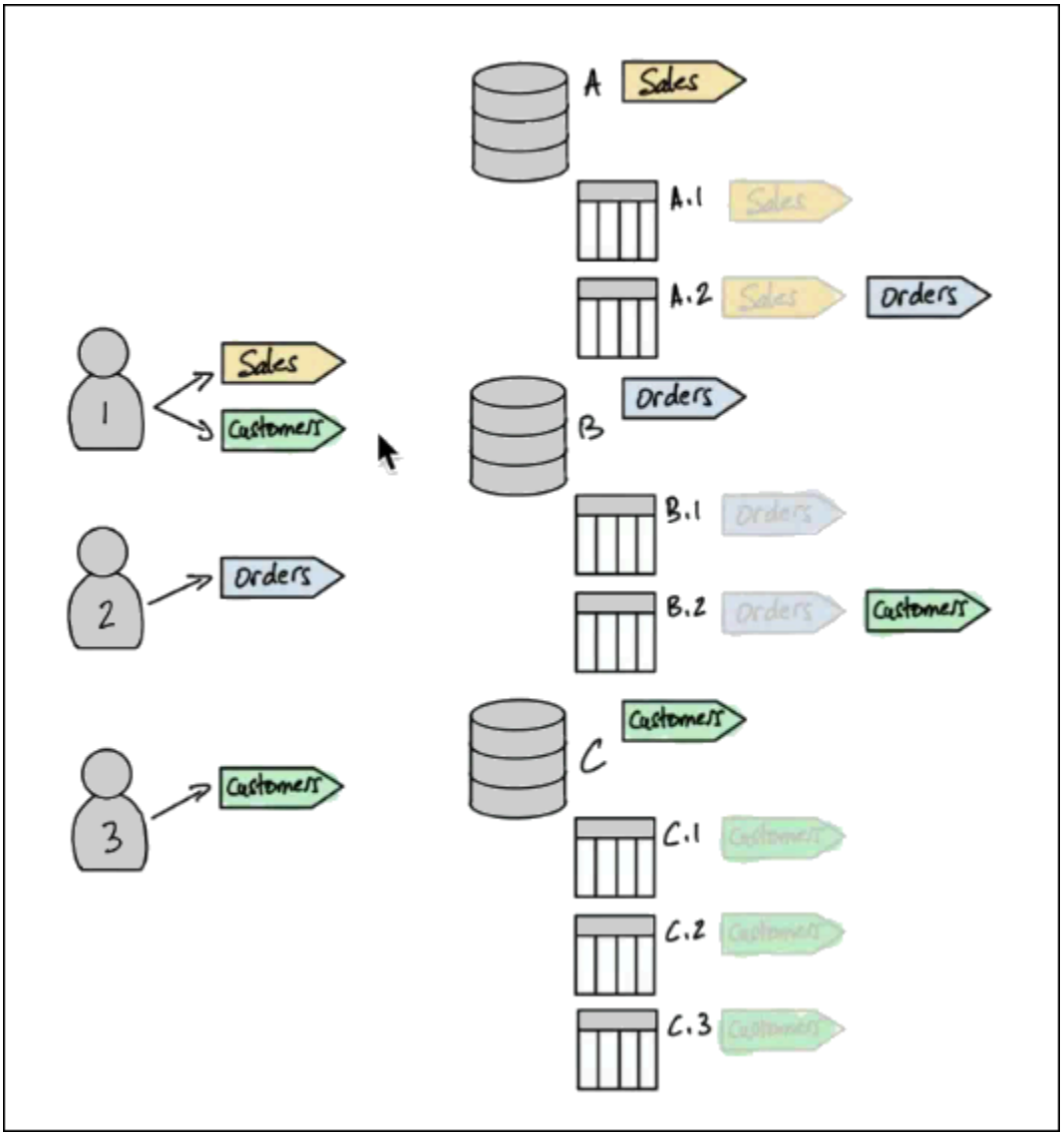
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

现在考虑如何使用 LF-TBAC 授予权限。下图表明您已将 LF 标签分配给数据库和表，并已向主体授予对 LF 标签的权限。

在此示例中，LF 标签表示数据湖中包含企业资源规划 (ERP) 应用程序套件不同模块的分析的区域。您可以控制对各个模块的分析数据的访问。所有 LF 标签均具有键 `module` 和可能的值 `Sales`、`Orders` 和 `Customers`。LF 标签示例如下所示：

```
module=Sales
```

该图仅显示 LF 标签值。



对数据目录资源和继承的标签分配

表从数据库继承 LF 标签，列从表继承 LF 标签。继承的值可以被覆盖。在上图中，继承了灰显 LF 标签。

由于继承，数据湖管理员只需对资源进行以下五个 LF 标签分配（使用伪代码）。

```
ASSIGN TAGS module=Sales TO database A
ASSIGN TAGS module=Orders TO table A.2
ASSIGN TAGS module=Orders TO database B
ASSIGN TAGS module=Customers TO table B.2
ASSIGN TAGS module=Customers TO database C
```

标记向主体的授予

将 LF 标签分配给数据库和表后，数据湖管理员必须仅向主体授予 4 个 LF 标签，如下所示（使用伪代码）。

```
GRANT TAGS module=Sales TO Principal 1
GRANT TAGS module=Customers TO Principal 1
GRANT TAGS module=Orders TO Principal 2
GRANT TAGS module=Customers TO Principal 3
```

现在，具有 module=Sales LF 标签的主体可以访问带有 module=Sales LF 标签的数据目录资源（例如，数据库 A），具有 module=Customers LF 标签的主体可以访问带有 module=Customers LF 标签的资源，依此类推。

上述授予命令不完整。这是因为，尽管它们通过 LF 标签指示主体对其具有权限的数据目录资源，但它们并未准确指示主体对这些资源具有哪些 Lake Formation 权限（例如 SELECT、ALTER）。因此，以下伪代码命令更准确地表示了如何通过 LF 标签授予对数据目录资源的 Lake Formation 权限。

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Sales TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Sales TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Orders TO Principal 2
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Orders TO Principal 2
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 3
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 3
```

综上所述，由此授予对资源的权限

根据上图中分配给数据库和表的 LF 标签以及图中向主体授予的 LF 标签，下表列出了主体对数据库和表具有的 Lake Formation 权限。

主体	通过 LF 标签授予的权限
主体 1	• 对数据库 A 的 CREATE_TABLE 权限

主体	通过 LF 标签授予的权限
	• 对表 A.1 的 SELECT、INSERT 权限 • 对表 B.2 的 SELECT、INSERT 权限 • 对数据库 C 的 CREATE_TABLE 权限 • 对表 C.1 的 SELECT、INSERT 权限 • 对表 C.2 的 SELECT、INSERT 权限 • 对表 C.3 的 SELECT、INSERT 权限
主体 2	• 对表 A.2 的 SELECT、INSERT 权限 • 对数据库 B 的 CREATE_TABLE 权限 • 对表 B.1 的 SELECT、INSERT 权限 • 对表 B.2 的 SELECT、INSERT 权限
主体 3	• 对表 B.2 的 SELECT、INSERT 权限 • 对数据库 C 的 CREATE_TABLE 权限 • 对表 C.1 的 SELECT、INSERT 权限 • 对表 C.2 的 SELECT、INSERT 权限 • 对表 C.3 的 SELECT、INSERT 权限

总结

在这个简单的示例中，使用 5 项分配操作和 8 项授予操作，数据湖管理员能够指定 17 项权限。当有数十个数据库和数百个表时，LF-TBAC 方法相对于命名资源方法的优势就显而易见了。假设需要向每个主体授予对每种资源的访问权限，其中 n(P) 是主体的数量，n(R) 是资源的数量：

- 使用命名资源方法时，所需的授予数为 n(P) × n(R)。
- 通过 LF-TBAC 方法，使用单个 LF 标签，向主体的授予和对资源的分配总数为 n(P) + n(R)。

另请参阅

- 管理 LF 标签以实现元数据访问控制
- 使用 LF-TBAC 方法授予数据湖权限

主题

管理 LF 标签以实现元数据访问控制

要使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法来保护数据目录资源（数据库、表和列），您可以创建 LF 标签，将其分配给资源，并向主体授予 LF 标签权限。

在将 LF 标签分配给数据目录资源或向主体授予权限之前，需要先定义 LF 标签。只有数据湖管理员或具有 LF 标签创建者权限的主体才能创建 LF 标签。

LF 标签创建者

LF 标签创建者是有权创建和管理 LF 标签的非管理员主体。数据湖管理员可以使用 Lake Formation 控制台或 CLI 添加 LF 标签创建者。LF 标签创建者具有隐式 Lake Formation 权限，可以更新和删除 LF 标签、将 LF 标签分配给资源，以及向其他主体授予 LF 标签权限和 LF 标签值权限。

通过 LF 标签创建者角色，数据湖管理员可以将标签管理任务（例如创建和更新标签键和值）委派给非管理员主体。数据湖管理员还可以向 LF 标签创建者授予可授予的 `Create LF-Tag` 权限。然后，LF 标签创建者可以向其他主体授予创建 LF 标签的权限。

您可以授予对 LF 标签的两种权限：

- LF 标签权限 - `Create LF-Tag`、`Alter` 和 `Drop`。创建、更新和删除 LF 标签需要这些权限。

  数据湖管理员和 LF 标签创建者隐式具有对其创建的 LF 标签的这些权限，并且可以向主体显式授予这些权限，以管理数据湖中的标签。
- LF 标签键值对权限 - `Assign`、`Describe` 和 `Grant with LF-Tag expressions`。需要这些权限才能将 LF 标签分配给数据目录数据库、表和列，以及使用 Lake Formation 基于标签的访问控制向主体授予对资源的权限。LF 标签创建者在创建 LF 标签时会隐式获得这些权限。

在获得 `Create LF-Tag` 权限并成功创建 LF 标签后，LF 标签创建者可以将 LF 标签分配给资源，并向其他非管理员主体授予 LF 标签权限（`Create LF-Tag`、`Alter` 和 `Drop`），以管理数据湖中的标签。您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来管理 LF 标签。

> **Note**
>
> 数据湖管理员具有隐式 Lake Formation 权限，可以创建、更新和删除 LF 标签、将 LF 标签分配给资源，以及向其他主体授予 LF 标签权限。

有关最佳实践和注意事项，请参阅 Lake Formation 基于标签的访问控制最佳实践和注意事项

主题

> **另请参阅**
>
> - 授予、撤销和列出 LF 标签值权限
> - 使用 LF-TBAC 方法授予数据湖权限
> - Lake Formation 基于标签的访问控制

添加 LF 标签创建者

默认情况下，数据湖管理员可以创建、更新和删除 LF 标记，将标签分配给数据目录资源，以及向主体授予标签权限。如果您希望将标签创建和管理操作委托给非管理员主体，则数据湖管理员可以创建 LF 标签创建者角色并向这些角色授予 Lake Formation `Create LF-Tag` 权限。通过可授予的 `Create LF-Tag` 权限，LF 标签创建者可以将标签创建和维护任务委托给其他非管理员主体。

> **Note**
>
> 跨账户权限授予只能包括 `Describe` 和 `Associate` 权限。您无法向其他账户中的主体授予 `Create LF-Tag`、`Drop`、`Alter` 和 `Grant with LFTag expressions` 权限。

主题

- 创建 LF 标签所需的 IAM 权限
- 添加 LF 标签创建者

> **另请参阅**
>
> - 授予、撤销和列出 LF 标签值权限
> - 使用 LF-TBAC 方法授予数据湖权限
> - Lake Formation 基于标签的访问控制

创建 LF 标签所需的 IAM 权限

您必须配置权限以允许 Lake Formation 主体创建 LF 标签。将以下语句添加到需要成为 LF 标签创建者的主体的权限策略中。

> **Note**
>
> 尽管数据湖管理员具有隐式 Lake Formation 权限，可以创建、更新和删除 LF 标签、将 LF 标签分配给资源，以及向主体授予 LF 标签，但数据湖管理员还需要以下 IAM 权限。

有关更多信息，请参阅 Lake Formation 角色和 IAM 权限参考。

```
{
"Sid": "Transformational",
"Effect": "Allow",
    "Action": [
        "lakeformation:AddLFTagsToResource",
        "lakeformation:RemoveLFTagsFromResource",
```

```
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:CreateLFTag",
        "lakeformation:GetLFTag",
        "lakeformation:UpdateLFTag",
        "lakeformation:DeleteLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
    ]
}
```

将 LF 标签分配给资源并向主体授予 LF 标签的主体必须具有相同的权限，但 `CreateLFTag`、`UpdateLFTag` 和 `DeleteLFTag` 权限除外。

添加 LF 标签创建者

LF 标签创建者可以创建 LF 标签、更新标签键和值、删除标签、将标签与数据目录资源关联，以及使用 LF-TBAC 方法向主体授予对数据目录资源的权限。LF 标签创建者还可以向主体授予这些权限。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来创建 LF 标签创建者角色。

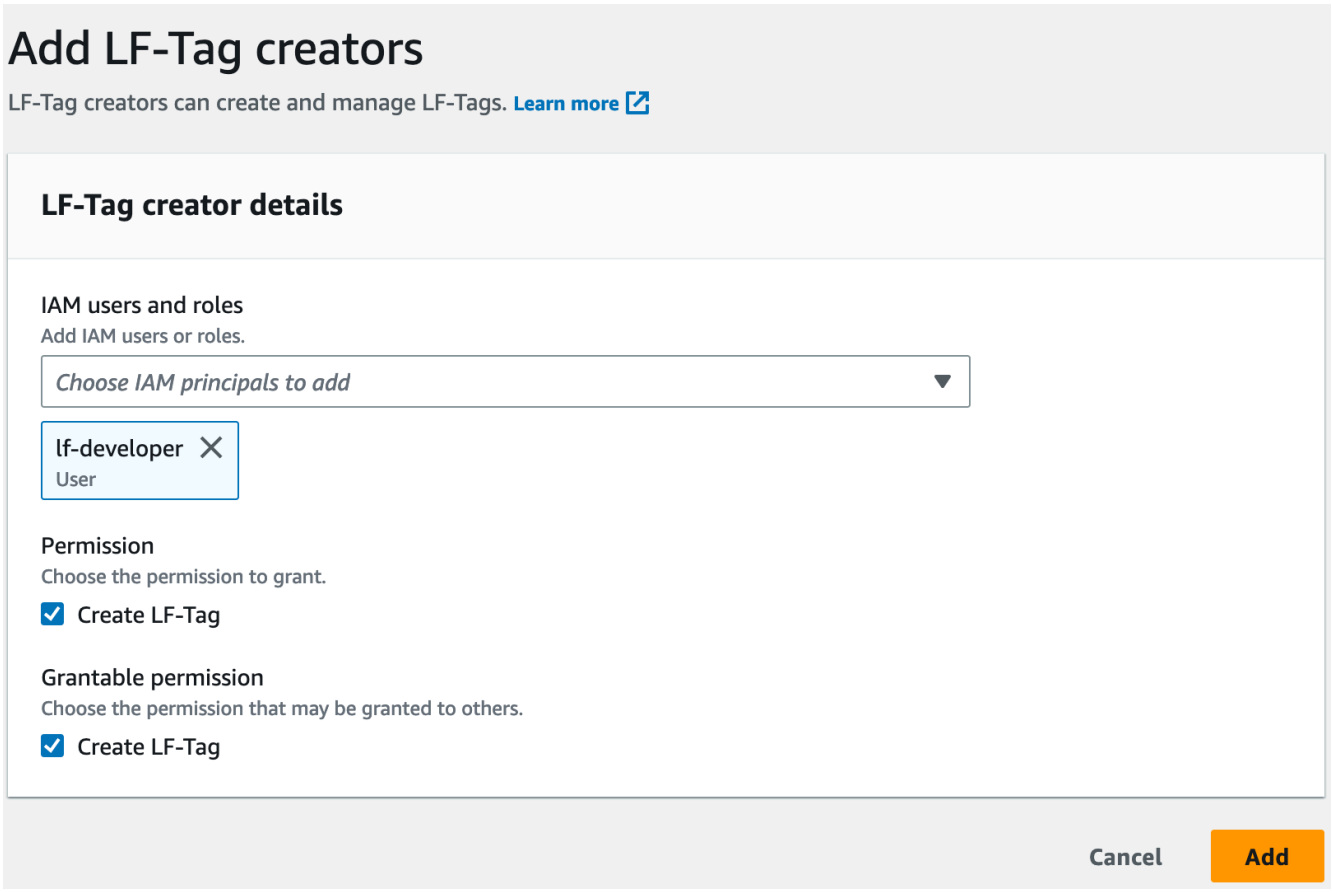
console

添加 LF 标签创建者

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员身份登录。
2. 在导航窗格的权限下，选择 LF 标签和权限。

   在 LF 标签和权限页面上，选择 LF 标签创建者部分，然后选择添加 LF 标签创建者。



3. 在添加 LF 标签创建者页面上，选择具有创建 LF 标签所需权限的 IAM 角色或用户。
4. 选中 `Create LF-Tag` 权限复选框。
5. （可选）要使所选主体能够向主体授予 `Create LF-Tag` 权限，请选择可授予的 `Create LF-Tag` 权限。
6. 选择添加。

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
    },
    "Resource": {
        "Catalog": {}
    },
    "Permissions": [
        "CreateLFTag"
```

```
    ],
    "PermissionsWithGrantOption": [
        "CreateLFTag"
    ]
}
```

以下是 LF 标签创建者角色可用的权限：

权限	描述
Drop	对 LF 标签具有此权限的主体可以从数据湖中删除 LF 标签。主体会获取对 LF 标签资源的所有标签值的隐式 Describe 权限。
Alter	对 LF 标签具有此权限的主体可以在 LF 标签中添加或移除标签值。主体会获取对 LF 标签的所有标签值的隐式 Alter 权限。
Describe	对 LF 标签具有此权限的主体在将 LF 标签分配给资源或授予对 LF 标签的权限时可以查看 LF 标签及其值。您可以授予对所有键值或特定值的 Describe 权限。
Associate	对 LF 标签具有此权限的主体可以将 LF 标签分配给数据目录资源。授予 Associate 会隐式授予 Describe 权限。
Grant with LF-Tag expression	对 LF 标签具有此权限的主体可以使用 LF 标签键和值授予对数据目录资源的权限。授予 Grant with LF-Tag expression 会隐式授予 Describe 权限。

这些权限是可以授予的。已通过授予选项被授予这些权限的主体可以将这些权限授予给其他主体。

创建 LF 标签

所有 LF 标签都必须先在 Lake Formation 中定义，然后才能使用。LF 标签由一个键和一个或多个可能的键值组成。

数据湖管理员为 LF 标签创建者角色设置所需的 IAM 权限和 Lake Formation 权限后，主体即可创建 LF 标签。LF 标签创建者获得隐式权限，可以更新或删除 LF 标签中的任何标签值以及删除 LF 标签。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来创建 LF 标签。

Console

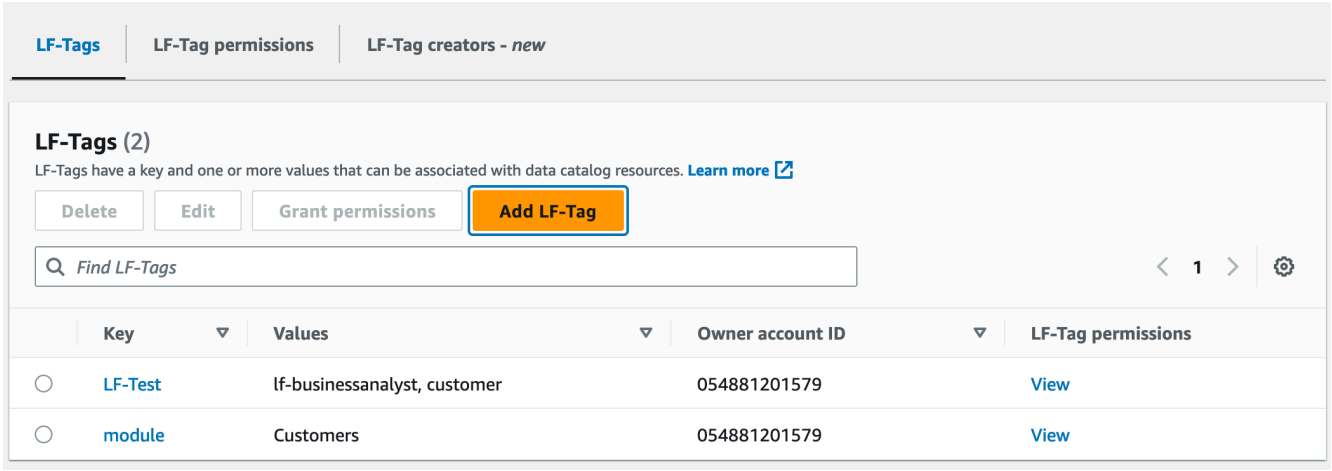
创建 LF 标签

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以具有 LF 标签创建者权限的主体或数据湖管理员身份登录。

2. 在导航窗格的 LF 标签和权限下，选择 LF 标签。

   此时将显示 LF 标签页面。



3. 选择添加 LF 标签。
4. 在添加 LF 标签对话框中，输入键和一个或多个值。

   每个键必须具有至少一个值。要输入多个值，请输入逗号分隔的列表，然后按 Enter，或者一次输入一个值，然后在每个值后选择添加。允许的值的最大数量为 1000。

5. 选择添加标签。

AWS CLI

创建 LF 标签

- 输入 `create-lf-tag` 命令。

  以下示例创建键为 `module` 以及值为 `Customers` 和 `Orders` 的 LF 标签。

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

作为标签创建者，主体会获得对此 LF 标签的 `Alter` 权限，并可以更新或移除此 LF 标签中的任何标签值。LF 标签创建者主体还可以向其他主体授予 `Alter` 权限，以更新和删除此 LF 标签上的标签值。

更新 LF 标签

您可以通过添加或删除允许的键值来更新您对其具有 `Alter` 权限的 LF 标签。您无法更改 LF 标签键。要更改键，请删除 LF 标签，然后添加一个具有所需键的 LF 标签。除了 `Alter` 权限外，您还需要 `lakeformation:UpdateLFTag` IAM 权限才能更新值。

删除 LF 标签值时，不会检查任何数据目录资源上是否存在该 LF 标签值。如果已删除的 LF 标签值与资源关联，则该值对资源不再可见，并且任何被授予对该键值对的权限的主体都将不再具有该权限。

在删除 LF 标签值之前，您可以选择使用 `remove-lf-tags-from-resource` 命令从具有要删除的值的数据目录资源中删除 LF 标签，然后使用要保留的值为该资源重新添加标签。

只有数据湖管理员、LF 标签创建者和具有对 LF 标签的 `Alter` 权限的主体才能更新 LF 标签。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来更新 LF 标签。

Console

更新 LF 标签（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员、LF 标签创建者或对 LF 标签具有 `Alter` 权限的主体身份登录。
2. 在导航窗格的 LF 标签和权限下，选择 LF 标签。
3. 在 LF 标签页面上，选择一个 LF 标签，然后选择编辑。
4. 在编辑 LF 标签对话框中，添加或移除 LF 标签值。

   要添加多个值，请在值字段中输入逗号分隔的列表并按 Enter，或者一次输入一个值，或者在每个值后选择添加。
5. 选择 Save（保存）。

AWS CLI

更新 LF 标签 (AWS CLI)

- 输入 `update-lf-tag` 命令。提供以下一个或两个参数：
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

以下示例将值 `vp` 替换为 LF 标签键 `level` 的值 `vice-president`。

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

删除 LF 标签

您可以删除不再使用的 LF 标签。不会检查数据目录资源上是否存在 LF 标签。如果已删除的 LF 标签与资源关联，则该标签对资源不再可见，并且任何被授予对该 LF 标签的权限的主体都将不再具有该权限。

在删除 LF 标签之前，您可以选择使用 `remove-lf-tags-from-resource` 命令来从所有资源中删除 LF 标签。

只有数据湖管理员、LF 标签创建者或具有对 LF 标签的 `Drop` 权限的主体才能删除 LF 标签。除了 `Drop` 权限外，主体还需要 `lakeformation:DeleteLFTag` IAM 权限才能删除 LF 标签。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来删除 LF 标签。

Console

删除 LF 标签（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员身份登录。
2. 在导航窗格的 LF 标签和权限下，选择 LF 标签。
3. 在 LF 标签页面上，选择一个 LF 标签，然后选择删除。

4. 在删除标签环境？对话框中，要确认删除，请在指定字段中输入 LF 标签键值，然后选择删除。

AWS CLI

删除 LF 标签 (AWS CLI)

- 输入 `delete-lf-tag` 命令。提供要删除的 LF 标签的键。

  Example

  以下示例删除具有键 `region` 的 LF 标签。

```
aws lakeformation delete-lf-tag --tag-key region
```

列出 LF 标签

您可以列出您对其具有 `Describe` 或 `Associate` 权限的 LF 标签。每个 LF 标签键中列出的值都是您对其具有权限的值。

LF 标签创建者具有查看他们创建的 LF 标签的隐式权限。

数据湖管理员可以查看本地 AWS 账户中定义的所有 LF 标签，以及已从外部账户向本地账户授予 `Describe` 和 `Associate` 权限的所有 LF 标签。数据湖管理员可以查看所有 LF 标签的所有值。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来列出 LF 标签。

Console

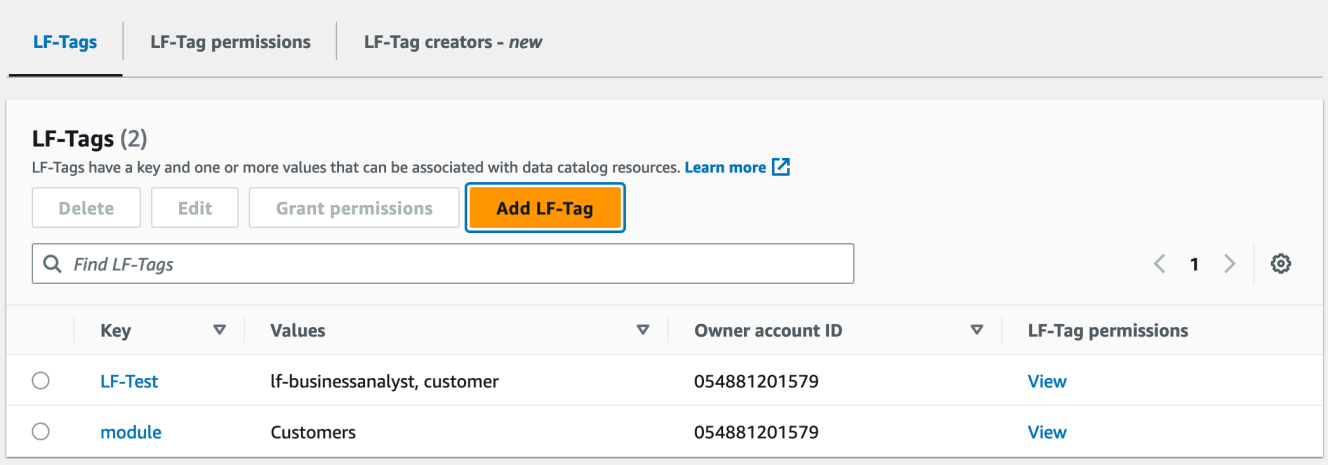
列出 LF 标签（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以 LF 标签创建者、数据湖管理员或已被授予对 LF 标签的权限且具有 `lakeformation:ListLFTags` IAM 权限的主体身份登录。

2. 在导航窗格的 LF 标签和权限下，选择 LF 标签。

   此时将显示 LF 标签页面。



查看拥有者账户 ID 列，以确定从外部账户与您的账户共享的 LF 标签。

AWS CLI

列出标签 (AWS CLI)

- 以数据湖管理员或已被授予对 LF 标签的权限且具有 `lakeformation:ListLFTags` IAM 权限的主体身份运行以下命令。

```
aws lakeformation list-lf-tags
```

输出类似于以下内容。

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
```

```
                "Sales",
                "Customers"
            ]
        }
    ]
}
```

要同时查看从外部账户授予的 LF 标签，请添加命令选项 `--resource-share-type ALL`。

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

输出类似于以下内容。请注意 `NextToken` 键，该键表示还有更多要列出的内容。

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ],
    "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}
```

重复该命令，然后添加 `--next-token` 参数以查看任何剩余的本地 LF 标签和从外部账户授予的 LF 标签。来自外部账户的 LF 标签始终位于单独的页面上。

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

```
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==
```

```
{
    "LFTags": [
        {
            "CatalogId": "123456789012",
            "TagKey": "region",
            "TagValues": [
                "central",
                "south"
            ]
        }
    ]
}
```

API

您可以使用可用于 Lake Formation 的 SDK 列出请求者有权查看的标签。

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50'
)
```

此命令返回具有以下结构的 dict 对象：

```
{
    'LFTags': [
        {
            'CatalogId': 'string',
            'TagKey': 'string',
            'TagValues': [
                'string',
            ]
        },
```

```
    ],
    'NextToken': 'string'
}
```

有关所需权限的更多信息，请参阅 Lake Formation 角色和 IAM 权限参考。

将 LF 标签分配给数据目录资源

您可以将 LF 标签分配给数据目录资源（数据库、表和列）以控制对这些资源的访问。只有被授予匹配 LF 标签的主体（以及使用命名资源方法被授予访问权限的主体）才能访问这些资源。

如果表继承了数据库中的 LF 标签，或者某列继承了表中的 LF 标签，则可以通过向 LF 标签键分配新值来覆盖继承的值。

您可以分配给资源的 LF 标签的最大数量为 50。

主题

- 管理分配给资源的标签的要求
- 将 LF 标签分配给表列
- 将 LF 标签分配给数据目录资源
- 更新资源的 LF 标签
- 从资源中删除 LF 标签

管理分配给资源的标签的要求

要将 LF 标签分配给数据目录资源，您必须：

- 具有对 LF 标签的 Lake Formation ASSOCIATE 权限。
- 具有 IAM `lakeformation:AddLFTagsToResource` 权限。
- 具有对 Glue 数据库的 glue:GetDatabase 权限。
- 是资源拥有者（创建者），通过 GRANT 选项具有对资源的 Super Lake Formation 权限，或通过 GRANT 选项具有以下权限：
  - 对于同一 AWS 账户中的数据库：DESCRIBE、CREATE_TABLE、ALTER 和 DROP 权限
  - 对于外部账户中的数据库：DESCRIBE、CREATE_TABLE 和 ALTER 权限
  - 对于表（和列）：DESCRIBE、ALTER、DROP、INSERT、SELECT 和 DELETE 权限

此外，LF 标签和要分配给它的资源必须位于同一个 AWS 账户中。

要从数据目录资源中删除 LF 标签，您必须满足这些要求并具有 `lakeformation:RemoveLFTagsFromResource` IAM 权限。

将 LF 标签分配给表列

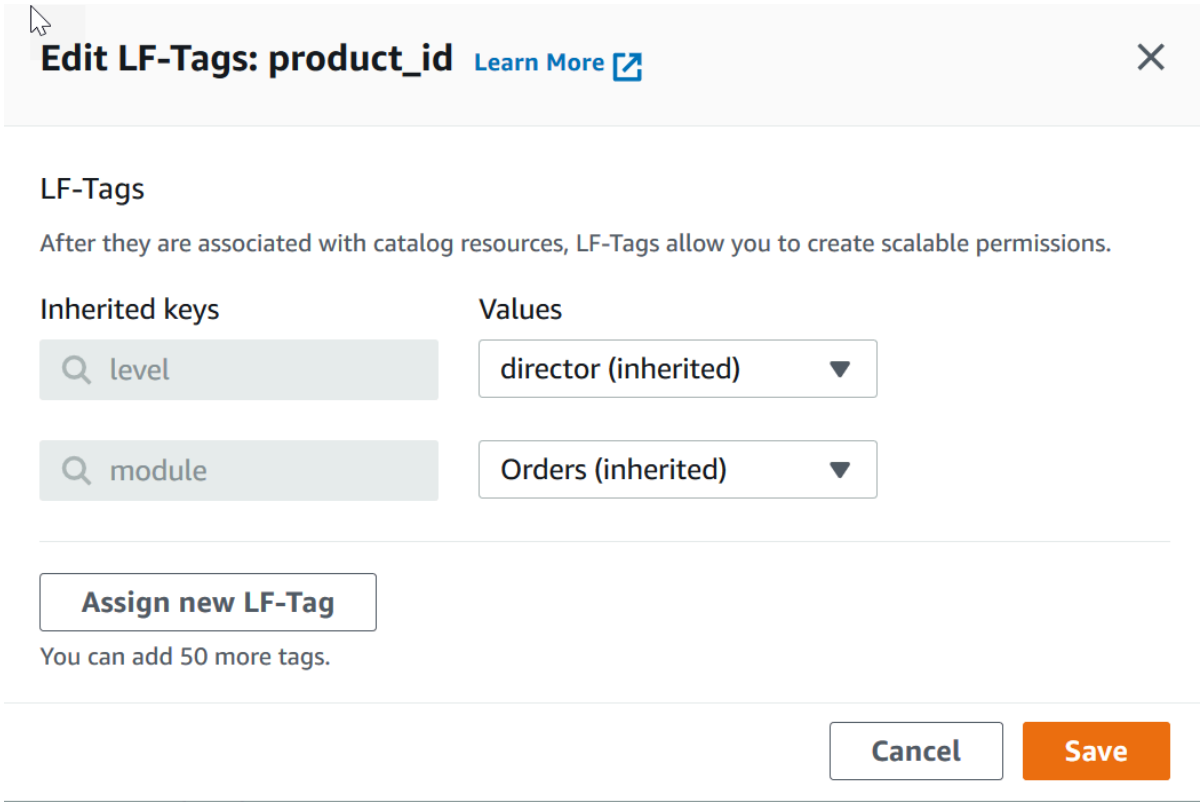
将 LF 标签分配给表列（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

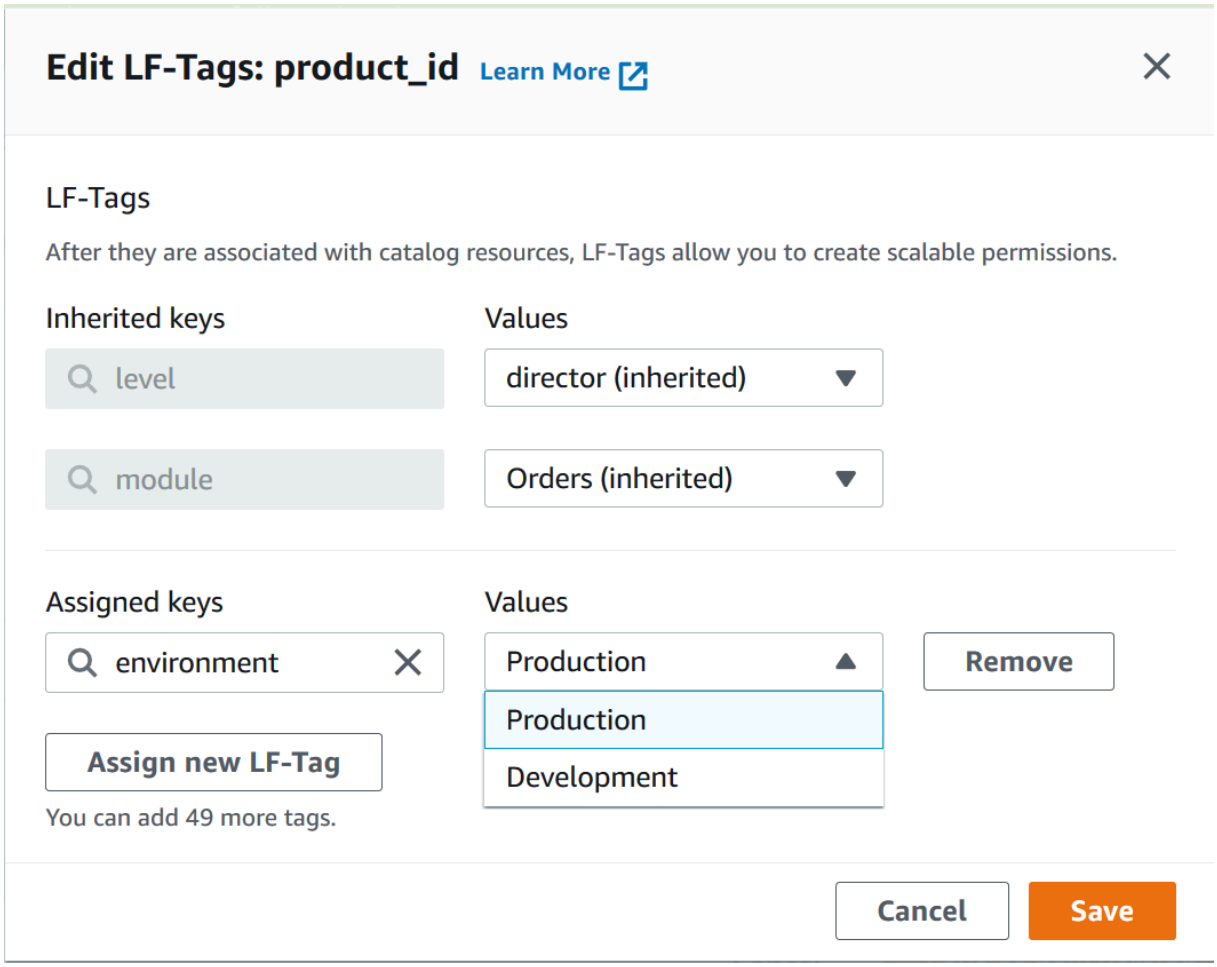
   以满足上述要求的用户身份登录。
2. 在导航窗格中，选择表。
3. 选择表名称（而不是表名称旁边的选项按钮）。
4. 在表详细信息页面的架构部分，选择编辑架构。
5. 在编辑架构页面上，选择一列或多列，然后选择编辑标签。

   > **Note**
   >
   > 如果要添加或删除列并保存新版本，请先执行此操作。然后编辑 LF 标签。

   此时将出现编辑 LF 标签对话框，并显示从该表继承的所有 LF 标签。



6. （可选）在继承的键字段旁边的值列表中，选择一个值来覆盖继承的值。
7. （可选）选择分配新的 LF 标签。然后，在分配的键中，选择一个键，然后对于值，为该键选择一个值。



8. （可选）再次选择分配新的 LF 标签以添加其他 LF 标签。
9. 选择 Save（保存）。

将 LF 标签分配给数据目录资源

Console

将 LF 标签分配给数据目录数据库或表

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

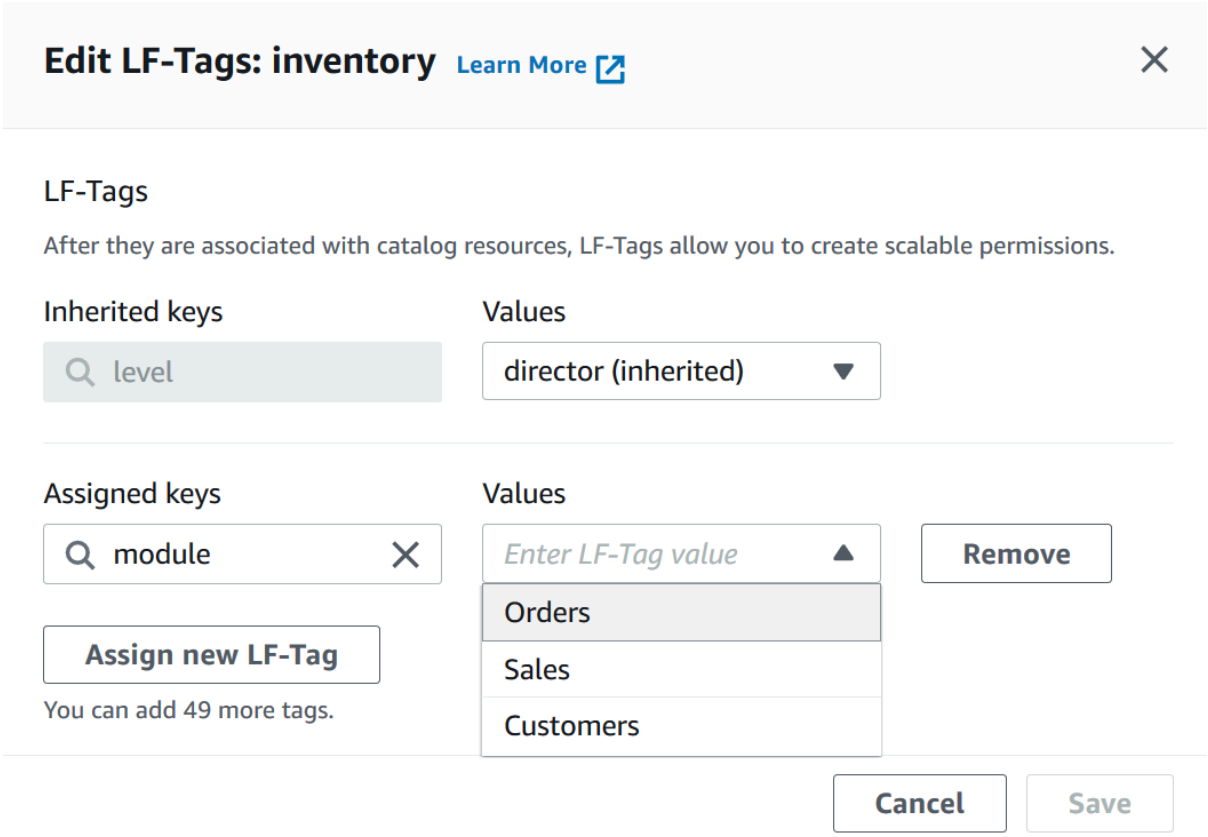
   以满足前面列出的要求的用户身份登录。
2. 在导航窗格的数据目录下，执行下列操作之一：

   - 要将 LF 标签分配给数据库，请选择数据库。
   - 要将 LF 标签分配给表，请选择表。

3. 选择数据库或表，然后在操作菜单上，选择编辑标签。

   此时将显示编辑 LF 标签：***resource-name*** 对话框。

   如果表从其包含的数据库继承 LF 标签，则窗口将显示继承的 LF 标签。否则，窗口将显示文本“没有与该资源关联的继承的 LF 标签”。



4. （可选）如果表继承了 LF 标签，则对于继承的键字段旁边的值列表，您可以选择一个值来覆盖继承的值。
5. 要分配新的 LF 标签，请执行以下步骤：

   a. 选择分配新的 LF 标签。

   b. 在分配的键字段中，选择一个 LF 标签键，然后在值字段中选择一个值。

   c. （可选）再次选择分配新的 LF 标签以分配额外的 LF 标签。
6. 选择 Save（保存）。

AWS CLI

将 LF 标签分配给数据目录资源

- 运行 add-lf-tags-to-resource命令。

  以下示例将 LF 标签 module=orders 分配给数据库 erp 中的表 orders。它使用 --lf-tags 参数的快捷语法。--lf-tags 的 CatalogID 属性是可选的。如果未提供该属性，则假定为资源（在本例中为表）的目录 ID。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags
 CatalogId=111122223333,TagKey=module,TagValues=orders
```

  以下是命令成功时的输出。

```
{
    "Failures": []
}
```

  下一个示例将 LF 标签分配给 sales 表，并对 --lf-tags 参数使用 JSON 语法。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":
 "module","TagValues": ["sales"]},{"TagKey": "environment","TagValues":
 ["development"]}]'
```

  下一个示例将 LF 标签 level=director 分配给表 sales 的 total 列。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":
 {"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags
 TagKey=level,TagValues=director
```

更新资源的 LF 标签

更新数据目录资源的 LF 标签 (AWS CLI)

- 如前面的步骤所述，使用 add-lf-tags-to-resource 命令。

添加与现有 LF 标签具有相同键但具有不同值的 LF 标签会更新现有值。

从资源中删除 LF 标签

删除数据目录资源的 LF 标签 (AWS CLI)

- 运行 `remove-lf-tags-from-resource`命令。

  如果表的 LF 标签值覆盖了从父数据库继承的值，则从表中删除该 LF 标签将还原继承的值。此行为也适用于覆盖从表中继承的键值的列。

  以下示例从 `sales` 表的 `total` 列中删除 LF 标签 `level=director`。`--lf-tags` 的 `CatalogID` 属性是可选的。如果未提供该属性，则假定为资源（在本例中为表）的目录 ID。

```
aws lakeformation remove-lf-tags-from-resource
--resource ' { "TableWithColumns":
{ "DatabaseName": "erp",  "Name": "sales",  "ColumnNames":[ "total"]}}'
--lf-tags  CatalogId=111122223333,TagKey=level,TagValues=director
```

查看分配给资源的 LF 标签

您可以查看分配给数据目录资源的 LF 标签。您必须对 LF 标签具有 `DESCRIBE` 或 `ASSOCIATE` 权限才能查看它。

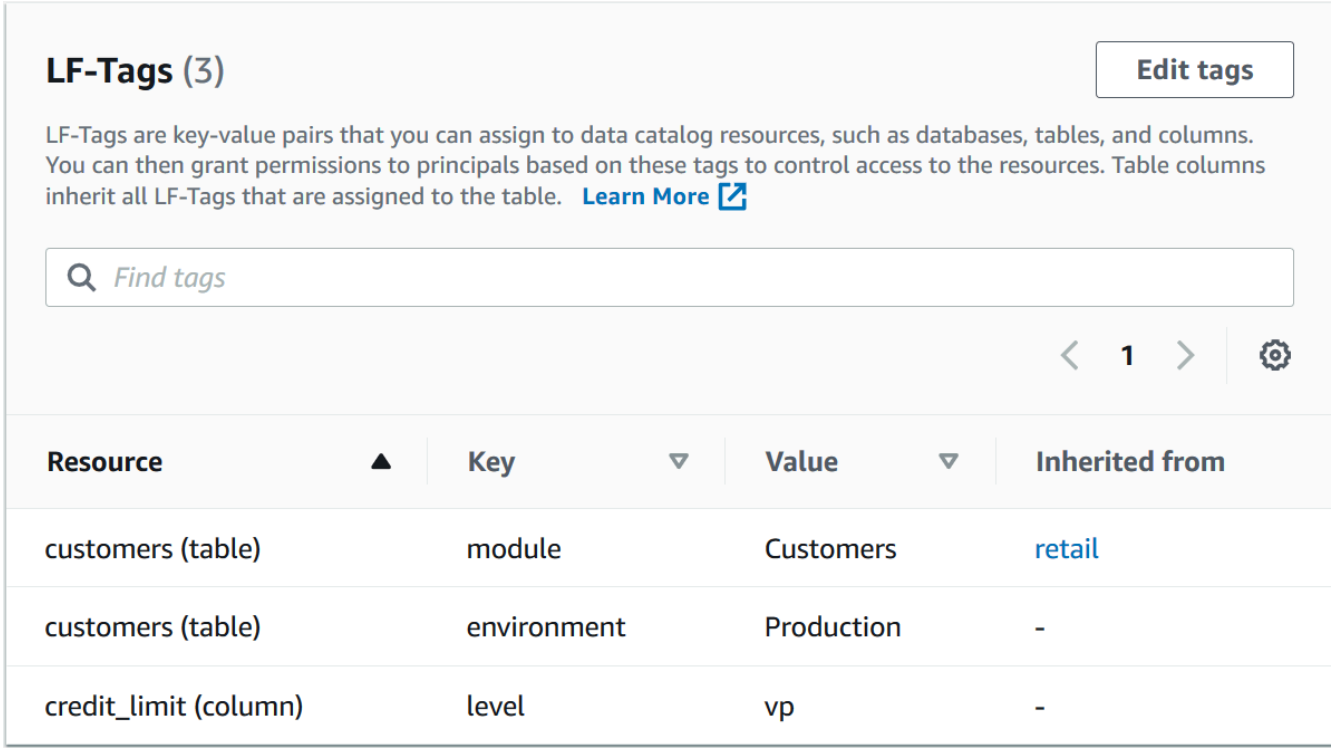
Console

查看分配给资源的 LF 标签（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员、资源拥有者或已被授予对资源的 Lake Formation 权限的用户身份登录。
2. 在导航窗格的数据目录标题下，执行下列操作之一：
   - 要查看分配给数据库的 LF 标签，请选择数据库。
   - 要查看分配给表的 LF 标签，请选择表。
3. 在表或数据库页面上，选择数据库或表的名称。然后在详细信息页面上，向下滚动到 LF 标签部分。

以下屏幕截图显示了分配给 customers 表的 LF 标签，该表包含在 retail 数据库中。module LF 标签继承自该数据库。credit_limit 列分配了 level=vp LF 标签。



AWS CLI

查看分配给资源的 LF 标签 (AWS CLI)

- 输入类似以下的命令。

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
 "Name":"sales"}}'
```

该命令将返回以下输出。

```
{
    "TableTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "sales"
```

```
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "environment",
            "TagValues": [
                "development"
            ]
        }
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}
```

此输出仅显示显式分配的 LF 标签，而不是继承的 LF 标签。如果要查看所有列上的所有 LF 标签，包括继承的 LF 标签，请忽略 `--show-assigned-lf-tags` 选项。

查看 LF 标签分配到的资源

您可以查看特定 LF 标签键分配到的所有数据目录资源。为此，您需要以下 Lake Formation 权限：

- 对 LF 标签的 `Describe` 或 `Associate` 权限。
- 对资源的 `Describe` 或任何其他 Lake Formation 权限。

此外，您还需要以下 AWS Identity and Access Management (IAM) 权限：

- `lakeformation:SearchDatabasesByLFTags`
- `lakeformation:SearchTablesByLFTags`

Console

查看 LF 标签分配到的资源（控制台）

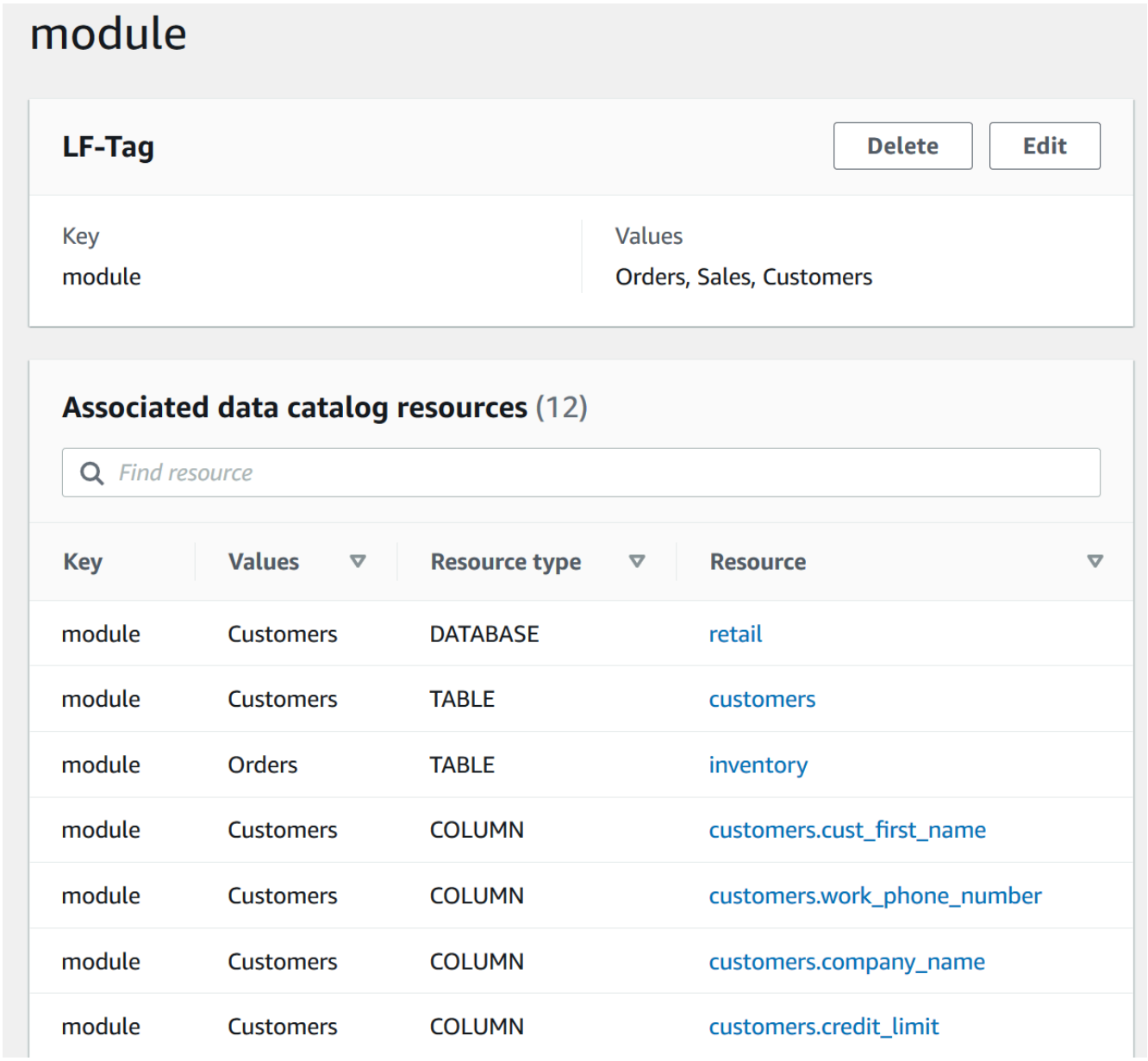
1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员或满足前面列出的要求的用户身份登录。

2. 在导航窗格的权限和管理角色和任务标题下，选择 LF 标签。

3. 选择 LF 标签键（而不是键名称旁边的选项按钮）。

   LF 标签详细信息页面显示 LF 标签已分配到的资源的列表。



AWS CLI

查看 LF 标签分配到的资源

- 运行 `search-tables-by-lf-tags` 或 `search-databases-by-lf-tags` 命令。

  Example

  以下示例列出分配了 `level=vp` LF 标签的表和列。对于列出的每个表和列，将输出为该表或列分配的所有 LF 标签，而不仅仅是搜索表达式。

```
aws lakeformation search-tables-by-lf-tags --expression
 TagKey=level,TagValues=vp
```

有关所需权限的更多信息，请参阅 Lake Formation 角色和 IAM 权限参考。

LF 标签的生命周期

1. LF 标签创建者 Michael 创建了一个 LF 标签 `module=Customers`。
2. Michael 向数据工程师 Eduardo 授予了对 LF 标签的 `Associate` 权限。授予 `Associate` 会隐式授予 `Describe` 权限。
3. Michael 使用授予选项向 Eduardo 授予了对表 `Custs` 的 `Super` 权限，因此 Eduardo 可以将 LF 标签分配给该表。有关更多信息，请参阅将 LF 标签分配给数据目录资源。
4. Eduardo 将 LF 标签 `module=customers` 分配给表 `Custs`。
5. Michael 向数据工程师 Sandra 进行了以下授予（使用伪代码）。

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra 向数据分析师 Maria 进行了以下授予。

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Maria 现在可以对 `Custs` 表运行查询。

另请参阅

- 元数据访问控制

比较 Lake Formation 基于标签的访问控制与 IAM 基于属性的访问控制

基于属性的访问控制 (ABAC) 是一种基于属性定义权限的授权策略。在 AWS 中，这些属性被称为“标签”。您可以将标签附加到 IAM 资源（包括 IAM 实体（用户和角色））以及 AWS 资源。您可以为 IAM 委托人创建单个 ABAC 策略或者一小组策略。这些 ABAC 策略可设计为在主体的标签与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。

云安全和治理团队使用 IAM 为所有资源定义访问策略和安全权限，包括 Amazon S3 存储桶、Amazon EC2 实例以及您可以使用 ARN 引用的任何资源。IAM 策略定义了对数据湖资源的广泛（粗粒度）权限，例如，允许或拒绝 Amazon S3 存储桶或前缀级别或数据库级别的访问。有关 IAM ABAC 的更多信息，请参阅《IAM 用户指南》中的什么是适用于 AWS 的 ABAC？。

例如，您可以创建具有 `project-access` 标签键的三个角色。将第一个角色的标签值设置为 `Dev`，第二个为 `Marketing`，第三个为 `Support`。向资源分配具有适当值的标签。然后，您可以使用单个策略，在角色和资源标记了 `project-access` 的相同值时允许访问。

数据治理团队使用 Lake Formation 来定义对特定数据湖资源的精细权限。LF 标签分配给数据目录资源（数据库、表和列）并授予给主体。具有与资源的 LF 标签匹配的 LF 标签的主体可以访问该资源。Lake Formation 权限次要于 IAM 权限。例如，如果 IAM 权限不允许用户访问数据湖，则 Lake Formation 不会向该用户授予对该数据湖中任何资源的访问权限，即使主体和资源具有匹配的 LF 标签也是如此。

Lake Formation 基于标签的访问控制 (LF-TBAC) 与 IAM ABAC 结合使用，为您的 Lake Formation 数据和资源提供更多级别的权限。

- Lake Formation TBAC 权限随着创新而扩展。它不再需要管理员更新现有策略以允许对新资源的访问。例如，假设您使用带有 `project-access` 标签的 IAM ABAC 策略来提供对 Lake Formation 中特定数据库的访问权限。使用 LF-TBAC，将 LF 标签 `Project=SuperApp` 分配给特定的表或列，并向该项目的开发人员授予相同的 LF 标签。通过 IAM，开发人员可以访问数据库，而 LF-TBAC 权限授予开发人员对特定表或表中列的进一步访问权限。如果将新表添加到项目中，则 Lake Formation 管理员只需将标签分配给新表，开发人员即可获得对该表的访问权限。
- Lake Formation TBAC 需要较少的 IAM 策略。由于您使用 IAM 策略来授予对 Lake Formation 资源的高级访问权限，并使用 Lake Formation TBAC 来管理更精确的数据访问，因此您创建的 IAM 策略更少。

- 使用 Lake Formation TBAC，团队可以快速变化和成长。这是因为新资源的权限根据属性自动授予。例如，如果新开发人员加入项目，则通过将 IAM 角色与用户关联，然后将所需的 LF 标签分配给用户，即可轻松授予该开发人员访问权限。您无需更改 IAM 策略即可支持新项目或创建新的 LF 标签。
- 使用 Lake Formation TBAC 可以获得更精细的权限。IAM 策略授予对顶级资源（例如数据目录数据库或表）的访问权限。使用 Lake Formation TBAC，您可以授予对包含特定数据值的特定表或列的访问权限。

> **Note**
>
> IAM 标签与 LF 标签不同。这些标签不可互换。LF 标签用于授予 Lake Formation 权限，而 IAM 标签用于定义 IAM 策略。

授予、撤销和列出 LF 标签值权限

您可以向主体授予对 LF 标签的 `Drop`、`Alter`权限，以管理 LF 标签值表达式。您还可以向主体授予对 LF 标签的 `Describe`、`Associate` 和 `Grant with LF-Tag expressions` 权限，以查看 LF 标签并将其分配给数据目录资源（数据库、表和列）。将 LF 标签分配给数据目录资源时，您可以使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法来保护这些资源。有关更多信息，请参阅 Lake Formation 基于标签的访问控制。

您可以使用授予选项授予这些权限，以便其他主体可以授予这些权限。添加 LF 标签创建者中介绍了 `Grant with LF-Tag expressions`、`Describe` 和 `Associate` 权限。

您可以向外部 AWS 账户授予 LF-Tag 的`Describe`和`Associate`权限。然后，该账户中的数据湖管理员可以将这些权限授予给该账户中的其他主体。接着，外部账户中的数据湖管理员向其授予 `Associate` 权限的主体可以将 LF 标签分配给您与其账户共享的数据目录资源。

向外部账户授予权限时，您必须包括授予选项。

您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予对 LF 标签的权限。

主题

- 使用控制台列出 LF 标签权限
- 使用控制台授予 LF 标签权限

- 使用授予、撤消和列出 LF-Tag 权限 AWS CLI

有关更多信息，请参阅管理 LF 标签以实现元数据访问控制和Lake Formation 基于标签的访问控制。

使用控制台列出 LF 标签权限

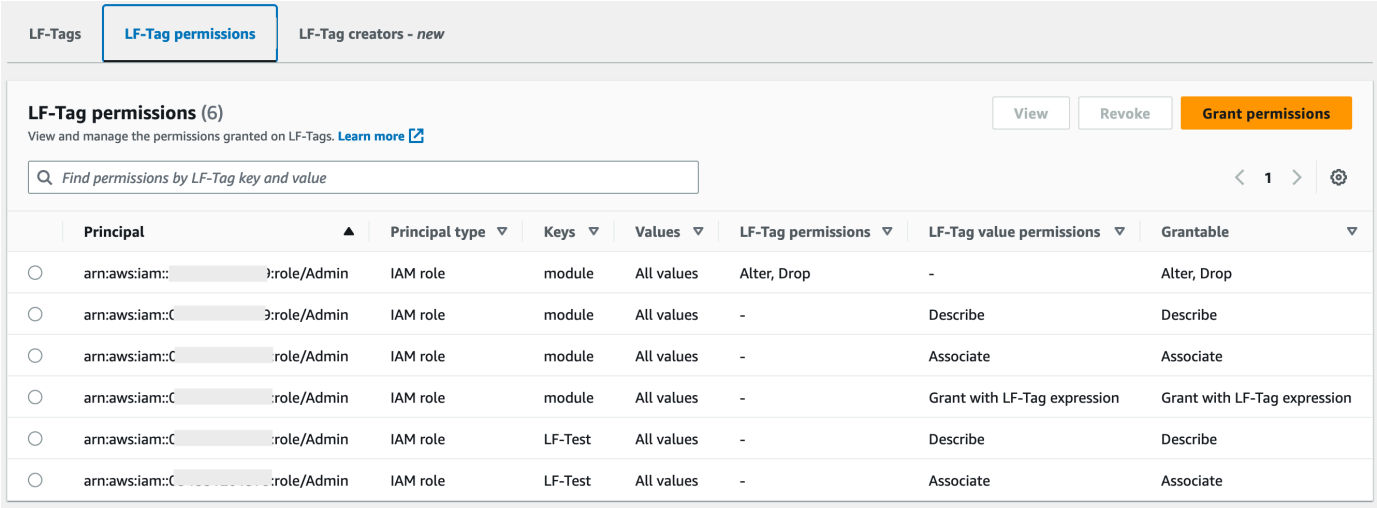
您可以使用 Lake Formation 控制台查看对 LF 标签授予的权限。您必须是 LF 标签创建者、数据湖管理员，或者具有对 LF 标签的 `Describe` 或 `Associate` 权限才能查看它。

列出 LF 标签权限（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以 LF 标签创建者、数据湖管理员或已被授予对 LF 标签的 `Drop`、`Alter`、`Associate` 或 `Describe` 权限的用户身份登录。
2. 在导航窗格的权限下，选择 LF 标签和权限，然后选择 LF 标签权限部分。

   LF 标签权限部分显示了一个包含主体、标签键、标签值和权限的表。



使用控制台授予 LF 标签权限

以下步骤说明如何使用 Lake Formation 控制台上的授予 LF 标签权限页面来授予对 LF 标签的权限。该页面分为以下几个部分：

- 权限类型 - 要授予的权限的类型。
- 委托人-要向其授予权限的用户、角色或 AWS 账户。
- LF 标签 - 要对其授予权限的 LF 标签。

- 权限 - 要授予的权限。

打开授予 LF 标签权限页面

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以 LF 标签创建者、数据湖管理员或已通过 `Grant` 选项被授予 LF 标签权限或对 LF 标签的 LF 标签键值对权限的用户身份登录。
2. 在导航窗格中，选择 LF 标签和权限，然后选择 LF 标签权限部分。
3. 选择 Grant permissions（授予权限）。

指定权限类型

在权限类型部分中，选择权限类型。

LF 标签权限

选择 LF 标签权限以允许主体更新 LF 标签值或删除 LF 标签。

LF 标签键值对权限

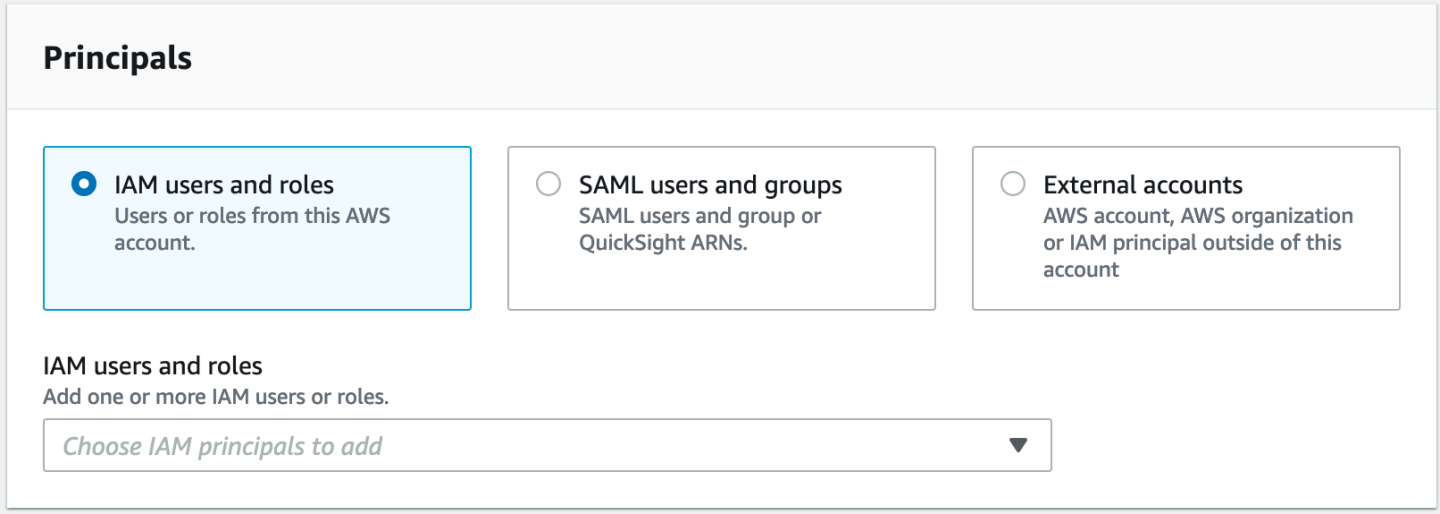
选择 LF 标签键值对权限以允许主体将 LF 标签分配给数据目录资源、查看 LF 标签和值，并向主体授予对数据目录资源的基于 LF 标签的权限。

以下各部分中可用的选项取决于权限类型。

指定主体

> **Note**
>
> 您不能向外部账户或其他账户中的主体授予 LF 标签权限（`Alter` 和 `Drop`）。

在主体部分中，选择主体类型，然后指定要向其授予权限的主体。



IAM 用户和角色

从 IAM 用户和角色列表中选择一个或多个用户或角色。

SAML 用户和组

对于 SAML 和 Amazon QuickSight 用户和群组，请为通过 SAML 联合的用户或群组输入一个或多个亚马逊资源名称 (ARN)，为亚马逊用户或群组输入一个或多个 QuickSight亚马逊资源名称 (ARN)。在每个 ARN 后按 Enter。

有关如何构建 ARN 的信息，请参阅 Lake Formation 授予和撤销命令 AWS CLI。

> **Note**
>
> 仅亚马逊 QuickSight 企业版支持 Lake Format QuickSight ion 与亚马逊的集成。

外部账户

对于AWS 账户，请输入一个或多个有效的 AWS 账户 ID。在每个 ID 后按 Enter。

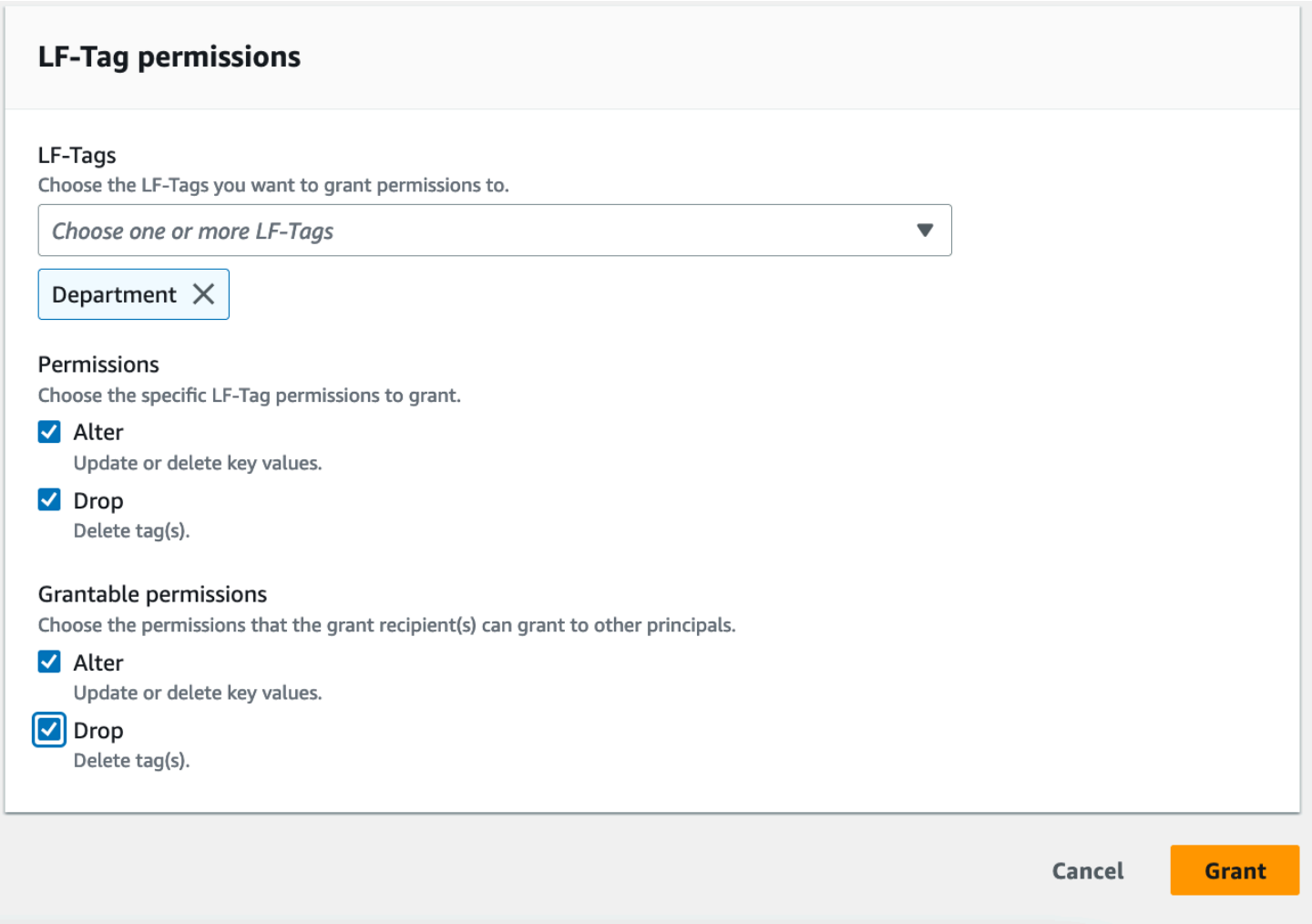
组织 ID 由“o-”后跟 10 到 32 个小写字母或数字组成。

单位 ID 以“ou-”开头，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。该字符串后跟第二个“-”短横线和 8 到 32 个额外的小写字母或数字。

对于 IAM 主体，输入 IAM 用户或角色的 ARN。

指定 LF 标签

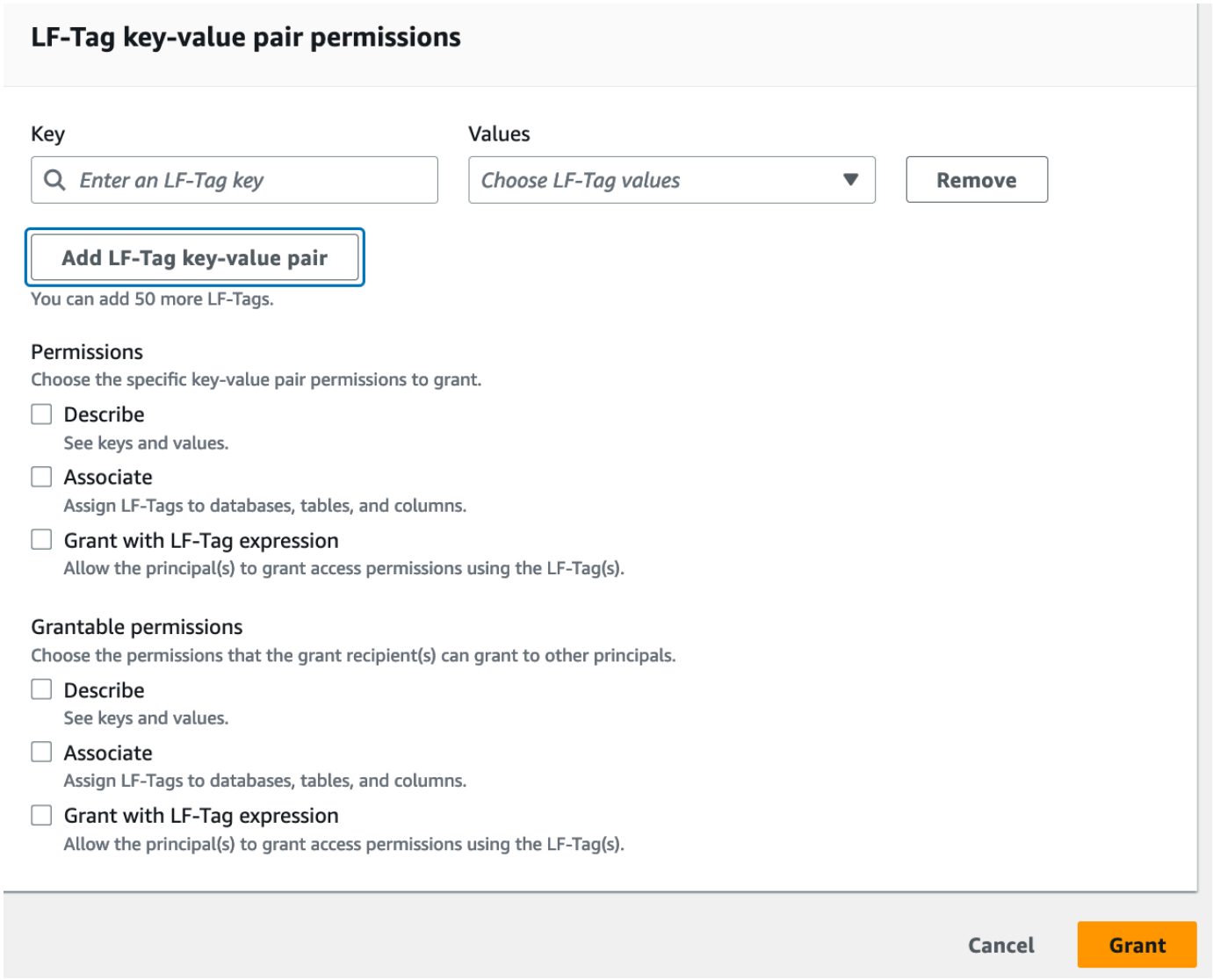
要授予对 LF 标签的权限，请在 LF 标签权限部分中，指定要对其授予权限的 LF 标签。



- 使用下拉列表选择一个或多个 LF 标签。

指定 LF 标签键值对

1. 要授予对 LF 标签键值对的权限，（您需要先选择 LF 标签键值对权限作为权限类型）选择添加 LF 标签键值对以显示用于指定 LF 标签键和值的第一行字段。



2. 将光标置于键字段中，可以选择开始键入以缩小选择列表范围，然后选择 LF 标签键。
3. 在值列表中，选择一个或多个值，然后按 Tab 键或在字段外单击或点按以保存所选值。

> **Note**
>
> 如果值列表中的某一行具有焦点，则按 Enter 键可选中或清除相应的复选框。

所选值显示为值列表下方的磁贴。选择 ✖ 以删除值。选择删除以删除整个 LF 标签。

4. 要添加另一个 LF 标签，请再次选择添加 LF 标签，然后重复前两个步骤。

指定权限

本部分根据您在上一步中选择的权限类型显示 LF 标签权限或 LF 标签值权限。

根据您选择授予的权限类型，选择 LF 标签权限或 LF 标签键值对权限和可授予的权限。

1. 在 LF 标签权限下，选择要授予的权限。

   授予 Drop 和 Alter 权限会隐式授予 Describe 权限。

   您需要授予对所有标签值的 Alter 和 Drop 权限。
2. 在 LF 标签键值对权限下，选择要授予的权限。

   授予 Associate 权限会隐式授予 Describe 权限。选择使用 LF 标签表达式授予权限，以允许授予接收人使用 LF-TBAC 方法授予或撤销对数据目录资源的访问权限。
3. （可选）在“可授予权限”下，选择授予接受者可以向其 AWS 账户中的其他委托人授予的权限。
4. 选择授权。

使用授予、撤消和列出 LF-Tag 权限 AWS CLI

您可以使用 AWS Command Line Interface (AWS CLI) 授予、撤销和列出对 LF 标签的权限。

列出 LF 标签权限 (AWS CLI)

- 输入 `list-permissions` 命令。您必须是 LF 标签创建者、数据湖管理员，或者具有对 LF 标签的 `Drop`、`Alter`、`Describe`、`Associate`、`Grant with LF-Tag permissions` 权限才能查看它。

  以下命令请求您有权访问的所有 LF 标签。

  ```
  aws lakeformation list-permissions --resource-type LF_TAG
  ```

  以下是数据湖管理员的示例输出，该人员可以查看向所有主体授予的所有 LF 标签。非管理员用户只能查看向其授予的 LF 标签。从外部账户授予的 LF 标签权限将显示在单独的结果页面上。要查看它们，请重复该命令并向 `--next-token` 参数提供上一次命令运行时返回的令牌。

  ```
  {
      "PrincipalResourcePermissions": [
          {
              "Principal": {
  ```

```
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_admin"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "environment",
                    "TagValues": [
                        "*"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": [
                "ASSOCIATE"
            ]
        },
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "DESCRIBE"
            ],
            "PermissionsWithGrantOption": []
        },
...
    ],
    "NextToken": "eyJzaG91bGRRdWVy...Wlzc2lvbnMiOnRydWV9"
}
```

您可以列出特定 LF 标签键的所有授予。以下命令返回对 LF 标签 module 授予的所有权限。

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

您还可以列出向特定主体授予的针对特定 LF 标签的 LF 标签值。提供 --principal 参数时，必须提供 --resource 参数。因此，该命令实际只能请求向特定主体授予的针对特定 LF 标签键的值。以下命令显示如何针对主体 datalake_user1 和 LF 标签键 module 执行此操作。

```
aws lakeformation list-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

下面是示例输出。

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": []
        }
    ]
```

```
}
```

授予对 LF 标签的权限 (AWS CLI)

1. 输入类似以下的命令。此示例向用户 datalake_user1 授予对带有键 module 的 LF 标签的 Associate 权限。它授予查看和分配该键的所有值的权限，如星号 (*) 所示。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

授予 Associate 权限会隐式授予 Describe 权限。

下一个示例使用密钥和授权选项Associate向外部 AWS 账户 1234-5678-9012 授予LF-tag。module它仅授予查看和分配值 sales 和 orders 的权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=123456789012 --permissions "ASSOCIATE"
 --permissions-with-grant-option "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
```

2. 授予 GrantWithLFTagExpression 权限会隐式授予 Describe 权限。

下一个示例使用授予选项向用户授予对带有键 module 的 LF 标签的 GrantWithLFTagExpression 权限。它仅使用值 sales 和 orders 授予查看和授予对数据目录资源的权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "GrantWithLFTagExpression"
 --permissions-with-grant-option "GrantWithLFTagExpression" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
```

3. 下一个示例使用授予选项向用户授予对带有键 module 的 LF 标签的 Drop 权限。它会授予删除 LF 标签的权限。要删除 LF 标签，您需要具有对该键的所有值的权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
```

```
 --permissions-with-grant-option "DROP" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

4. 下一个示例使用授予选项向用户授予对带有键 module 的 LF 标签的 Alter 权限。它会授予删除 LF 标签的权限。要更新 LF 标签，您需要具有对该键的所有值的权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
 --permissions-with-grant-option "ALTER" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

撤销对 LF 标签的权限 (AWS CLI)

- 输入类似以下的命令。此示例向用户 datalake_user1 授予对带有键 module 的 LF 标签的 Associate 权限。

```
aws lakeformation revoke-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

使用 LF-TBAC 方法授予数据湖权限

您可以向主体授予对 LF 标签的 DESCRIBE 和 ASSOCIATE Lake Formation 权限，以便他们可以查看 LF 标签并将其分配给数据目录资源（数据库、表、视图和列）。将 LF 标签分配给数据目录资源时，您可以使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法来保护这些资源。有关更多信息，请参阅 Lake Formation 基于标签的访问控制。

首先，只有数据湖管理员才能授予这些权限。如果数据湖管理员通过授予选项授予这些权限，则其他主体可以授予这些权限。Lake Formation 基于标签的访问控制最佳实践和注意事项中介绍了 DESCRIBE 和 ASSOCIATE 权限。

您可以向外部 AWS 账户授予对 LF 标签的 DESCRIBE 和 ASSOCIATE 权限。然后，该账户中的数据湖管理员可以将这些权限授予给该账户中的其他主体。接着，外部账户中的数据湖管理员向其授予 ASSOCIATE 权限的主体可以将 LF 标签分配给您与其账户共享的数据目录资源。

向外部账户授予权限时，您必须包括授予选项。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予对 LF 标签的权限。

主题

另请参阅

- 授予、撤销和列出 LF 标签值权限
- 管理 LF 标签以实现元数据访问控制
- Lake Formation 基于标签的访问控制

授予数据目录权限

使用 Lake Formation 控制台或 AWS CLI 利用 Lake Formation 基于标签的访问控制（LF-TBAC）方法授予对数据目录数据库、表、视图和列的 Lake Formation 权限。

Console

以下步骤说明如何使用 Lake Formation 基于标签的访问控制（LF-TBAC）方法和 Lake Formation 控制台上的授予数据湖权限页面来授予权限。该页面分为以下几个部分：

- 主体 – 要向其授予权限的用户、角色和 AWS 账户。
- LF 标签或目录资源 – 要对其授予权限的数据库、表或资源链接。
- 权限 – 要授予的 Lake Formation 权限。

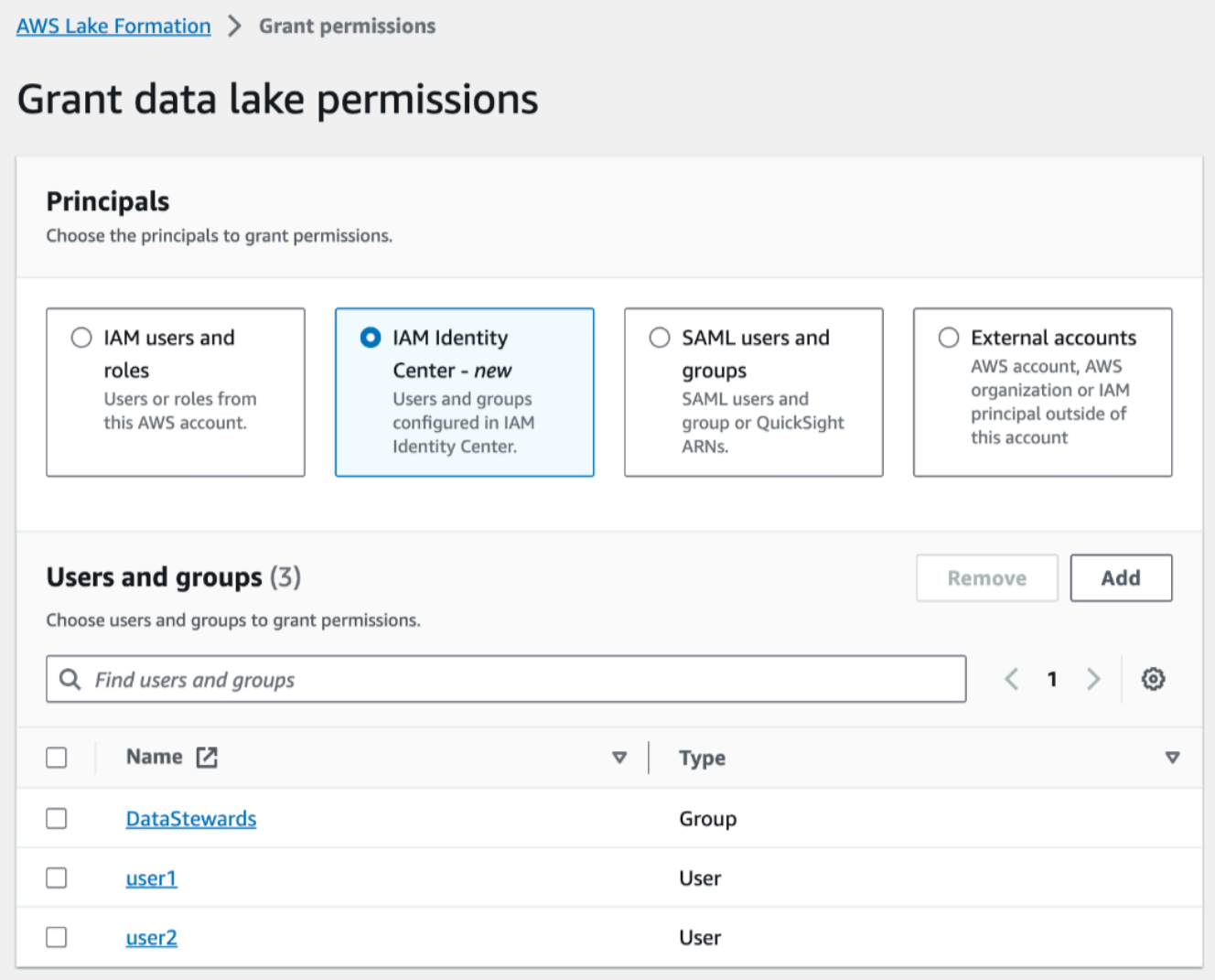
1. 打开“授予数据湖权限”页面。

   通过 https://console.aws.amazon.com/lakeformation/ 打开 AWS Lake Formation 控制台，然后以数据湖管理员或已使用授予选项通过 LF-TBAC 获得对数据目录资源的 Lake Formation 权限的用户身份登录。

   在导航窗格的权限下，选择数据湖权限。然后选择授予。

2. 指定主体。

   在主体部分中，选择主体类型，然后指定要向其授予权限的主体。



IAM 用户和角色

从 IAM 用户和角色列表中选择一个或多个用户或角色。

IAM Identity Center

从用户和组列表中选择一个或多个用户或组。

SAML 用户和组

对于 SAML 和 Amazon QuickSight 用户和组，为通过 SAML 联合的用户或组输入一个或多个 Amazon 资源名称 (ARN)，或者为 Amazon QuickSight 用户或组输入 ARN。在每个 ARN 后按 Enter。

有关如何构建 ARN 的信息，请参阅 Lake Formation 授予和撤销命令 AWS CLI。

> **Note**
>
> 只有 Amazon QuickSight 企业版支持 Lake Formation 与 Amazon QuickSight 的集成。

外部账户

对于 AWS 账户、AWS 组织或 IAM 主体，为 IAM 用户或角色输入一个或多个有效 AWS 账户 ID、组织 ID、组织单位 ID 或 ARN。在每个 ID 后按 Enter。

组织 ID 由“o-”后跟 10 到 32 个小写字母或数字组成。

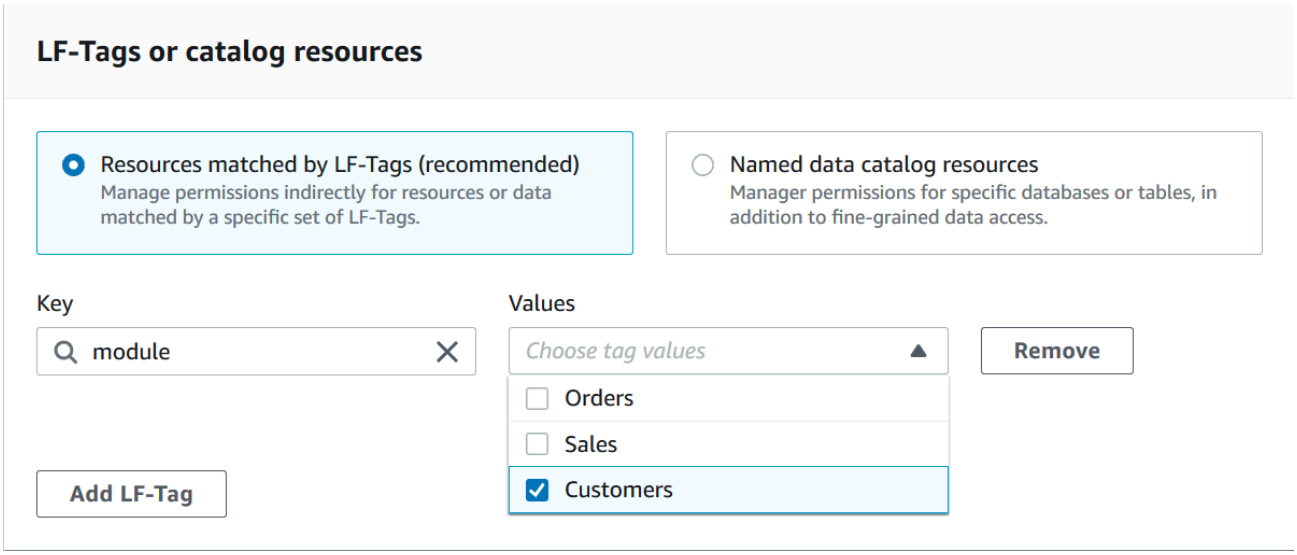
单位 ID 以“ou-”开头，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。该字符串后跟第二个“-”短横线和 8 到 32 个额外的小写字母或数字。

3. 指定 LF 标签。

确保选择与 LF 标签匹配的资源选项。选择添加 LF 标签。

1. 选择 LF 标签键和值。

如果选择多个值，则将使用 OR 运算符创建 LF 标签表达式。这意味着，如果任何 LF 标签值与分配给数据目录资源的 LF 标签相匹配，则将授予您对该资源的权限。



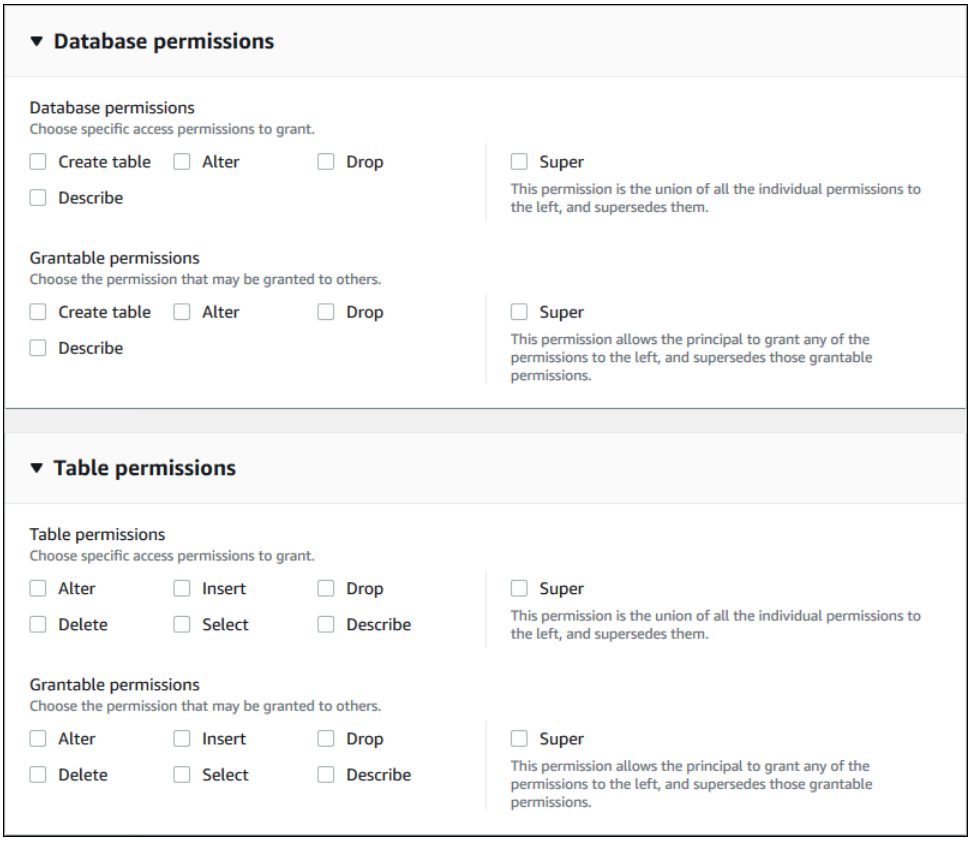
2. （可选）再次选择添加 LF 标签以指定其他 LF 标签。

如果指定多个值，则将使用 AND 运算符创建 LF 标签表达式。仅在为 LF 标签表达式中的每个 LF 标签分配了匹配的 LF 标签时，才会向主体授予对数据目录资源的权限。

4. 指定权限。

指定向主体授予的对匹配数据目录资源的权限。匹配的资源是指那些分配了 LF 标签的资源，这些标签与向主体授予的其中一个 LF 标签表达式相匹配。

您可以指定要授予的对匹配数据库、匹配表和匹配视图的权限。



在数据库权限下，选择要向主体授予的对匹配数据库的数据库权限。

在表权限下，选择要向主体授予的对匹配表和匹配视图的表或视图权限。

您也可以从表权限中选择要对视图应用的 `Select`、`Describe` 和 `Drop` 权限。

5. 选择授权。

AWS CLI

您可以使用 AWS Command Line Interface (AWS CLI) 和 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法授予对数据目录数据库、表和列的 Lake Formation 权限。

使用 AWS CLI 和 LF-TBAC 方法授予数据湖权限

- 使用 `grant-permissions` 命令。

  Example

  以下示例向用户 `datalake_user1` 授予 LF 标签表达式“`module=*`”（LF 标签键 `module` 的所有值）。该用户将具有对所有匹配数据库（即分配了带有键 `module` 及任何值的 LF 标签的数据库）的 `CREATE_TABLE` 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "CREATE_TABLE" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["*"]}]}}'
```

  Example

  下一个示例向用户 `datalake_user1` 授予 LF 标签表达式“`(level=director) AND (region=west OR region=south)`”。该用户将通过授予选项具有对匹配表（即分配了 `level=director` 和 `region=west` 或`region=south` 的表）的 `SELECT`、`ALTER` 和 `DROP` 权限。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "SELECT" "ALTER" "DROP" --permissions-
with-grant-option "SELECT" "ALTER" "DROP" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"TABLE","Expression": [{"TagKey":
 "level","TagValues": ["director"]},{"TagKey": "region","TagValues": ["west",
 "south"]}]}}'
```

  Example

  下一个示例向账户 1234-5678-9012 中的 AWS 授予 LF 标签表达式“`module=orders`”。然后，该账户中的数据湖管理员可以向其账户中的主体授予“`module=orders`”表达式。接着，这

些主体将具有对匹配数据库的 CREATE_TABLE 权限，这些数据库由账户 1111-2222-3333 拥有并通过使用命名资源方法或 LF-TBAC 方法与账户 1234-5678-9012 共享。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=123456789012 --permissions "CREATE_TABLE" --
permissions-with-grant-option "CREATE_TABLE" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["orders"]}]}}'
```

权限示例场景

以下场景有助于演示如何设置权限以保护对 AWS Lake Formation 中数据的访问。

Shirley 是一名数据管理员。她想为自己的公司 AnyCompany 建立一个数据湖。目前，所有数据都存储在 Amazon S3 中。John 是一名营销经理，需要对客户购买信息（包含在 s3://customerPurchases 中）进行写入访问。今年夏天，市场分析师 Diego 加入了 John 的团队。John 需要能够授予 Diego 访问权限，以便在不涉及 Shirley 的情况下对数据执行查询。

财务部的 Mateo 需要访问查询会计数据（例如，s3://transactions）。他想查询财务团队使用的数据库 (Finance_DB) 的表中的交易数据。他的经理 Arnav 可以允许他访问 Finance_DB。尽管他不应该能够修改会计数据，但他需要能够将数据转换为适合预测的格式（架构）。此类数据将存储在他可以修改的单独存储桶 (s3://financeForecasts) 中。

总结一下：

- Shirley 是数据湖管理员。
- John 需要 CREATE_DATABASE 和 CREATE_TABLE 权限才能在数据目录中创建新的数据库和表。
- 还需要对自己创建的表具有 SELECT、INSERT 和 DELETE 权限。
- Diego 需要对表具有 SELECT 权限才能运行查询。

AnyCompany 的员工需要执行以下操作来设置权限。为清楚起见，本场景中显示的 API 操作显示了简化的语法。

1. Shirley 向 Lake Formation 注册包含客户购买信息的 Amazon S3 路径。

   ```
   RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
   ```

2. Shirley 向 John 授予访问包含客户购买信息的 Amazon S3 路径的权限。

```
GrantPermissions(John, S3Location("s3://customerPurchases"),
 [DATA_LOCATION_ACCESS]) )
```

3. Shirley 向 John 授予创建数据库的权限。

```
GrantPermissions(John, catalog, [CREATE_DATABASE])
```

4. John 创建数据库 John_DB。John 自动具有对该数据库的 CREATE_TABLE 权限，因为他创建了该数据库。

```
CreateDatabase(John_DB)
```

5. John 创建指向 s3://customerPurchases 的表 John_Table。由于他创建了该表，因此他具有对该表的所有权限，并且可以授予对该表的权限。

```
CreateTable(John_DB, John_Table)
```

6. John 允许他的分析师 Diego 访问表 John_Table。

```
 GrantPermissions(Diego, John_Table, [SELECT])
```

7. John 允许他的分析师 Diego 访问 s3://customerPurchases/London/。由于 Shirley 已经注册 s3://customerPurchases，因此其子文件夹已在 Lake Formation 中注册。

```
 GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],
 S3Location("s3://customerPurchases/London/") )
```

8. John 允许他的分析师 Diego 在数据库 John_DB 中创建表。

```
 GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],
 [] )
```

9. Diego 在 s3://customerPurchases/London/ 处的 John_DB 中创建表，并自动获取 ALTER、DROP、SELECT、INSERT 和 DELETE 权限。

```
 CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Lake Formation 中的数据筛选和单元格级别安全性

当您授予对数据目录表的 Lake Formation 权限时，可以包括数据筛选规范，以限制对查询结果中以及与 Lake Formation 集成的引擎中某些数据的访问。Lake Formation 使用数据筛选来实现列级别安全性、行级别安全性以及单元格级别安全性。如果源数据包含嵌套结构，则可以对嵌套列定义和应用数据筛选条件。

主题

数据筛选概览

借助 Lake Formation 的数据筛选功能，您可以实现以下级别的数据安全性。

列级别安全性

授予对具有列级别安全性（列筛选）的数据目录表的权限将仅允许用户查看表中他们有权访问的特定列和嵌套列。以一家大型多区域通信公司在多个应用程序中使用的 `persons` 表为例。通过列筛选来授予对数据目录表的权限可以限制非人力资源部门的用户查看社会保险号或出生日期等个人身份信息 (PII)。您还可以定义安全策略并仅授予对嵌套列的部分子结构的访问权限。

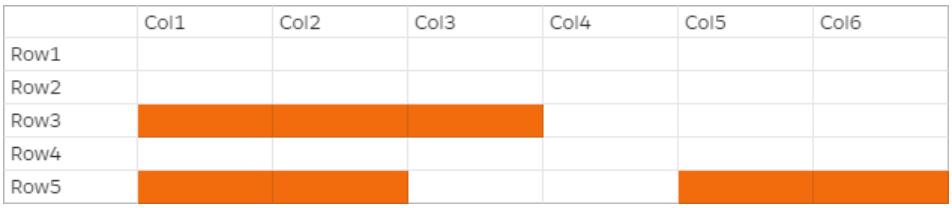
行级别安全性

授予对具有行级别安全性（行筛选）的数据目录表的权限将仅允许用户查看表中他们有权访问的特定行。筛选基于一列或多列值。定义行筛选条件表达式时，可以包含嵌套列结构。例如，如果通信公司的不同地区办事处都有自己的人力资源部门，则可以限制人力资源部员工可以查看的人员记录，只允许他们查看他们所在区域的员工的记录。

单元格级别安全性

单元格级别安全功能将行筛选和列筛选相结合，打造出了高度灵活的权限模型。如果您以网格形式查看表的行和列，则通过使用单元格级别安全功能，可以从两个方面限制从任意位置对网格中各个元素（单

元格）的访问。也就是说，您可以根据行来限制对不同列的访问。下图阐明了这一点，其中受限制的列是带阴影的。



继续来看 Persons 表示例，您可以在单元格级别创建数据筛选条件，如果行的国家/地区列设置为“英国”，则限制对街道地址列的访问，但如果行的国家/地区列设置为“美国”，则允许访问街道地址列。

筛选条件仅适用于读取操作。因此，您只能通过筛选条件授予 SELECT Lake Formation 权限。

嵌套列的单元格级别安全性

Lake Formation 允许您对嵌套列定义和应用具有单元格级别安全性的数据筛选条件。但是，Amazon Athena、Amazon EMR 和 Amazon Redshift Spectrum 等集成分析引擎支持对具有行级别和列级别安全性的 Lake Formation 托管嵌套表执行查询。

有关限制，请参阅数据筛选限制。

Lake Formation 中的数据筛选条件

您可以通过创建数据筛选条件来实现列级别、行级别和单元格级别安全性。在授予对表的 SELECT Lake Formation 权限时，您可以选择数据筛选条件。如果您的表包含嵌套列结构，则可以通过包含或排除子列来定义数据筛选条件，并针对嵌套属性定义行级别筛选表达式。

每个数据筛选条件属于数据目录中的一个特定表。数据筛选条件包含以下信息。

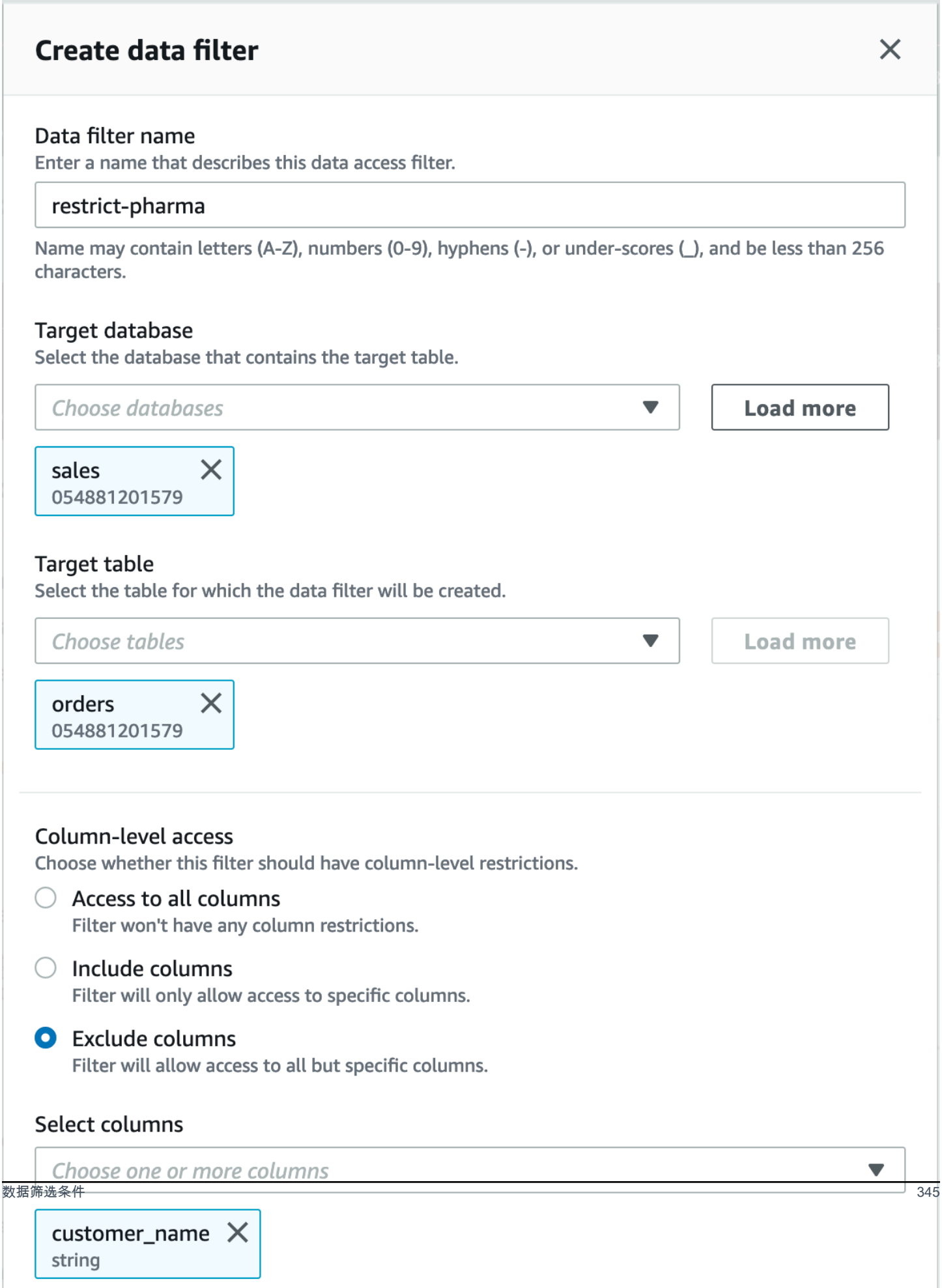
- 筛选条件名称
- 与筛选条件关联的表的目录 ID
- 表名称
- 包含表的数据库的名称
- 列规范 – 要在查询结果中包含或排除的列和嵌套列的列表（使用 `struct` 数据类型）。
- 行筛选表达式 – 用于指定要包含在查询结果中的行的表达式。由于实施一些限制，该表达式的语法与 PartiQL 语言中 `WHERE` 子句的语法相同。要指定所有行，请选择控制台的行级别访问下的访问所有行，或者在 API 调用中使用 `AllRowsWildcard`。

有关行筛选表达式支持哪些内容的更多信息，请参阅行筛选表达式支持 PartiQL。

您获得的筛选级别取决于填充数据筛选器的方式。

- 如果指定“所有列”通配符并提供行筛选条件表达式，则仅建立行级安全性（行筛选）。
- 如果包含或排除特定列和嵌套列并使用“所有行”通配符指定所有行，则仅会建立列级别安全性（列筛选）。
- 如果包含或排除特定列，同时提供行筛选条件表达式，则建立单元格级别的安全性（单元格筛选）。

以下来自 Lake Formation 控制台的屏幕截图显示了执行单元格级别筛选的数据筛选条件。对于针对 `orders` 表的查询，它会限制对 `customer_name` 列的访问，并且查询结果仅返回 `product_type` 列包含“pharma”的行。


Create data filter
Data filter name
Enter a name that describes this data access filter.
restrict-pharma
Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.
Target database
Select the database that contains the target table.
Choose databases
Load more
sales
054881201579
Target table
Select the table for which the data filter will be created.
Choose tables
Load more
orders
054881201579
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Select columns
Choose one or more columns
customer_name
string

请注意使用单引号将字符串 'pharma' 引起来。

您可以使用 Lake Formation 控制台创建此数据筛选条件，也可以向 CreateDataCellsFilter API 操作提供以下请求对象。

```
{
     "Name": "restrict-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type='pharma'"},
     "ColumnWildcard": {
         "ExcludedColumnNames": ["customer_name"]
     }
}
```

您可以根据需要为表创建许多数据筛选条件。为此，您需要使用授权选项授予对表的 SELECT 权限。默认情况下，数据湖管理员有权在该账户中的所有表上创建数据筛选条件。在向主体授予对表的权限时，通常只使用一部分可能的数据筛选条件。例如，您可以为 orders 表创建第二个数据筛选条件，即 row-security-only 数据筛选条件。根据前面的屏幕截图，您可以选择访问所有列选项，并包括行筛选表达式 product_type<>pharma。此数据筛选条件的名称可能是 no-pharma。它限制对 product_type 列设置为“pharma”的所有行的访问。

此数据筛选条件的 CreateDataCellsFilter API 操作的请求对象如下。

```
{
     "Name": "no-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type<>'pharma'"},
     "ColumnNames": ["customer_id", "customer_name", "order_num"
          "product_id", "purchase_date", "product_type",
          "product_manufacturer", "quantity", "price"]
}
```

然后，您可以使用 restrict-pharma 数据筛选条件向管理用户授予对 orders 表的 SELECT，并使用 no-pharma 数据筛选条件向非管理用户授予对 orders 表的 SELECT。对于医疗保健业的用户，您可以授予对 orders 表的 SELECT 以及对所有行和列的完整访问权限（没有数据筛选条件），或者还可以使用数据筛选条件来限制对定价信息的访问。

在数据筛选条件中指定列级别和行级别安全性时，可以包含或排除嵌套列。在以下示例中，使用限定列名称（用双引号括起来）指定对该 `product.offer` 字段的访问权限。这对于嵌套字段很重要，这样可以避免在列名称包含特殊字符时发生错误，并保持与顶级列级别安全性定义的向后兼容性。

```
{
     "Name": "example_dcf",
     "DatabaseName": "example_db",
     "TableName": "example_table",
     "TableCatalogId": "111122223333",
     "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
     "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

另请参阅

- 管理数据筛选条件

行筛选表达式支持 PartiQL

您可以使用一部分 PartiQL 数据类型、运算符和聚合来构造行筛选表达式。Lake Formation 不允许在筛选表达式中使用任何用户定义的或标准的 PartiQL 函数。您可以使用比较运算符将列与常量（例如，`views >= 10000`）进行比较，但不能将列与其他列进行比较。

行筛选表达式可以是简单表达式或复合表达式。表达式的总长度必须少于 2048 个字符。

简单表达式

简单表达式的格式为 `<column name > <comparison operator ><value >`

- 列名称

  它可以是表架构中的顶层数据列、分区列或嵌套列，并且必须属于下面列出的支持的 数据类型。
- 比较运算符

  以下是支持的运算符：`=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE, NOT, IS [NOT] NULL`
- 所有字符串比较和 `LIKE` 模式匹配项均区分大小写。不能对分区列使用 IS [NOT] NULL 运算符。
- 列值

列值必须匹配列名称的数据类型。

复合表达式

复合表达式的格式为 (<simple expression >) <AND/OR >(<simple expression >)。可以使用逻辑运算符 AND/OR 进一步组合复合表达式。

支持的 数据类型

如果行筛选条件引用的 AWS Glue Data Catalog 表包含不受支持的数据类型，则会导致错误。以下是表列和常量支持的数据类型，它们映射到 Amazon Redshift 数据类型：

- STRING, CHAR, VARCHAR
- INT, LONG, BIGINT, FLOAT, DECIMAL, DOUBLE
- BOOLEAN
- STRUCT

有关 Amazon Redshift 中数据类型的更多信息，请参阅《Amazon Redshift 数据库开发人员指南》中的数据类型。

行筛选表达式

Example

例如，以下便是包含列的表的有效行筛选表达式： country (String), id (Long), year (partition column of type Integer), month (partition column of type Integer)

- year > 2010 and country != 'US'
- (year > 2010 and country = 'US') or (month < 8 and id > 23)
- (country between 'Z' and 'U') and (year = 2018)
- (country like '%ited%') and (year > 2000)

Example

以下是包含嵌套列的表的行筛选条件表达式的有效示例：year > 2010 and customer.customerId <> 1

定义嵌套的行级别表达式时，不应引用分区列下的嵌套字段。

字符串常量必须用单引号引起来。

保留关键字

如果您的行筛选表达式包含 PartiQL 关键字，则您会收到解析错误，因为列名可能与关键字冲突。发生这种情况时，使用双引号对列名进行转义。例如，下面便是一些保留关键字：“first”、“last”、“asc”、“missing”。有关保留关键字的列表，请参阅 PartiQL 规范。

PartiQL 参考

有关 PartiQL 的更多信息，请参阅https://partiql.org/。

有关列级别筛选的注意事项和限制

指定列筛选的方法有三种：

- 通过使用数据筛选条件（如前所述）。
- 通过使用简单的列筛选或嵌套列筛选。
- 通过使用 TAG。

简单列筛选仅指定要包含或排除的列的列表。Lake Formation 控制台、API 和 AWS CLI 均支持简单列筛选。有关示例，请参阅Grant with Simple Column Filtering。

以下注意事项和限制适用于列筛选：

- AWS Glue ETL 作业仅支持使用数据筛选条件进行列筛选（单元格级别安全性）。如果对作业引用的任何表应用简单列筛选，则作业将失败。如果您只想进行列筛选，请使用数据筛选条件授予对表的访问权限，然后在控制台中为行筛选表达式输入 `true`，或者在 API 调用中使用 `AllRowsWildcard`。
- 要使用授权选项和列筛选来授予 SELECT，必须使用包含列表，而不是排除列表。如果没有授权选项，则可以使用包含列表或排除列表。
- 要使用列筛选来授予对表的 SELECT，您必须已通过授权选项获得了对该表的 SELECT，且没有任何行限制。您必须有权访问所有行。
- 如果您使用授权选项和列筛选向您账户中的主体授予 SELECT，则该主体在向其他主体授予权限时，必须为相同列指定列筛选，或指定部分已授权列。如果您使用授权选项和列筛选向外部账户授予

SELECT，则外部账户中的数据湖管理员可以将对所有列的 SELECT 授予其账户中的其他主体。但是，即使对所有列均具有 SELECT，该主体也只能查看授权外部账户查看的列。

- 您不能对分区键应用列筛选。
- 不能向对表中部分列拥有 SELECT 权限的主体授予对该表的 ALTER、DROP、DELETE 或 INSERT 权限。对于对表拥有 ALTER、DROP、DELETE 或 INSERT 权限的主体，如果您使用列筛选来授予 SELECT 权限，则授予无效。

以下注意事项和限制适用于嵌套列筛选：

- 您可以在数据筛选条件中包含或排除五个级别的嵌套字段。

  Example

  Col1.Col1_1.Col1_1_1.Col1_1_1_1.Col1_1_1_1_1
- 您不能对分区列中的嵌套字段应用列筛选。
- 如果您的表架构包含顶级列名称 ("customer"."address")，在数据筛选条件中具有相同的嵌套字段表示形式模式（具有顶级列名称 customer 和嵌套字段名称 address 的嵌套列在数据筛选条件中指定为 "customer"."address"），则您无法明确指定对顶级列或嵌套字段的访问权限，因为两者在包含/排除列表中使用相同的模式表示形式。这是不明确的，如果您指定顶级列或嵌套字段，Lake Formation 将无法解析。
- 如果顶级列或嵌套字段的名称中包含双引号，则在数据单元格筛选条件的包含和排除列表中指定对嵌套字段的访问权限时，必须包含第二个双引号。

  Example

  带双引号的嵌套列名称示例 – a.b.double"quote

  Example

  数据筛选条件中的嵌套列表示形式示例 – "a"."b"."double""quote"

使用单元格级别筛选对表进行查询所需的权限

使用单元格级别筛选对表运行查询需要以下 AWS Identity and Access Management (IAM) 权限。

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
        }
    ]
}
```

有关 Lake Formation 权限的更多信息，请参阅Lake Formation 角色和 IAM 权限参考。

管理数据筛选条件

要实现列级、行级和单元格级安全性，您可以创建和维护数据筛选条件。每个数据筛选条件都属于一个数据目录表。您可以为一个表创建多个数据筛选条件，然后在授予对表的权限时使用其中的一个或多个筛选条件。您还可以对嵌套列定义和应用数据筛选条件，这些嵌套列具有 `struct` 数据类型，允许用户仅访问嵌套列的子结构。

您需要具有授予选项的 `SELECT` 权限才能创建或查看数据筛选条件。要允许您账户中的主体查看和使用某个数据筛选条件，您可以授予对该筛选条件的 `DESCRIBE` 权限。

> **Note**
>
> Lake Formation 不支持授予对某个数据筛选条件的 `Describe` 权限，该筛选条件是从另一个账户共享的。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 管理数据筛选条件。

有关数据筛选条件的信息，请参阅 Lake Formation 中的数据筛选条件。

创建数据筛选条件

您可以为每个数据目录表创建一个或多个数据筛选条件。

为数据目录表创建数据筛选条件（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员、目标表拥有者或对目标表具有 Lake Formation 权限的主体的身份进行登录。
2. 在导航窗格的数据目录下，选择数据筛选条件。
3. 在数据筛选条件页面上，选择创建新筛选条件。
4. 在创建数据筛选条件对话框中，输入以下信息：

   - 数据筛选条件名称
   - 目标数据库 - 指定包含表的数据库。
   - 目标表
   - 列级访问 - 将此设置保留为访问所有列以仅指定行筛选。选择包括列或排除列以指定列或单元格筛选，然后指定要包含或排除的列。

     嵌套列 - 如果要对包含嵌套列的表应用筛选条件，则可以在数据筛选条件中显式指定嵌套结构列的子结构。

     当您向某个主体授予对此筛选条件的 SELECT 权限时，执行以下查询的主体将只能看到 `customer.customerName` 的数据，而看不到 `customer.customerId` 的数据。

     ```
     SELECT "customer" FROM "example_db"."example_table";
     ```

Column-level access

Choose whether this filter should have column-level restrictions.

Column-level access

Choose whether this filter should have column-level restrictions.

- ○ Access to all columns
  Filter won't have any column restrictions.
- ◉ Include columns
  Filter will only allow access to specific columns.
- ○ Exclude columns
  Filter will allow access to all but specific columns.

Included columns (4/11)

Choose the columns for column-level access

Find column

1

Name	Type
☐ customer	struct
☐ customerId	string
☑ customerName	string
☑ customerapplication	struct
☐ appId	string
☑ product	struct
☐ offer	struct
☐ listingId	string
☐ prodId	string
☐ type	string
☑ purchaseid	string

Row-level access

Choose whether this filter should have row-level restrictions.

- ○ Access to all rows
- ◉ Filter rows

Row filter expression

Enter the rest of the following query statement SELECT * FROM nested-table WHERE...
Please see the documentation for examples of filter expressions.

```
customer.customerName <> 'John'
```

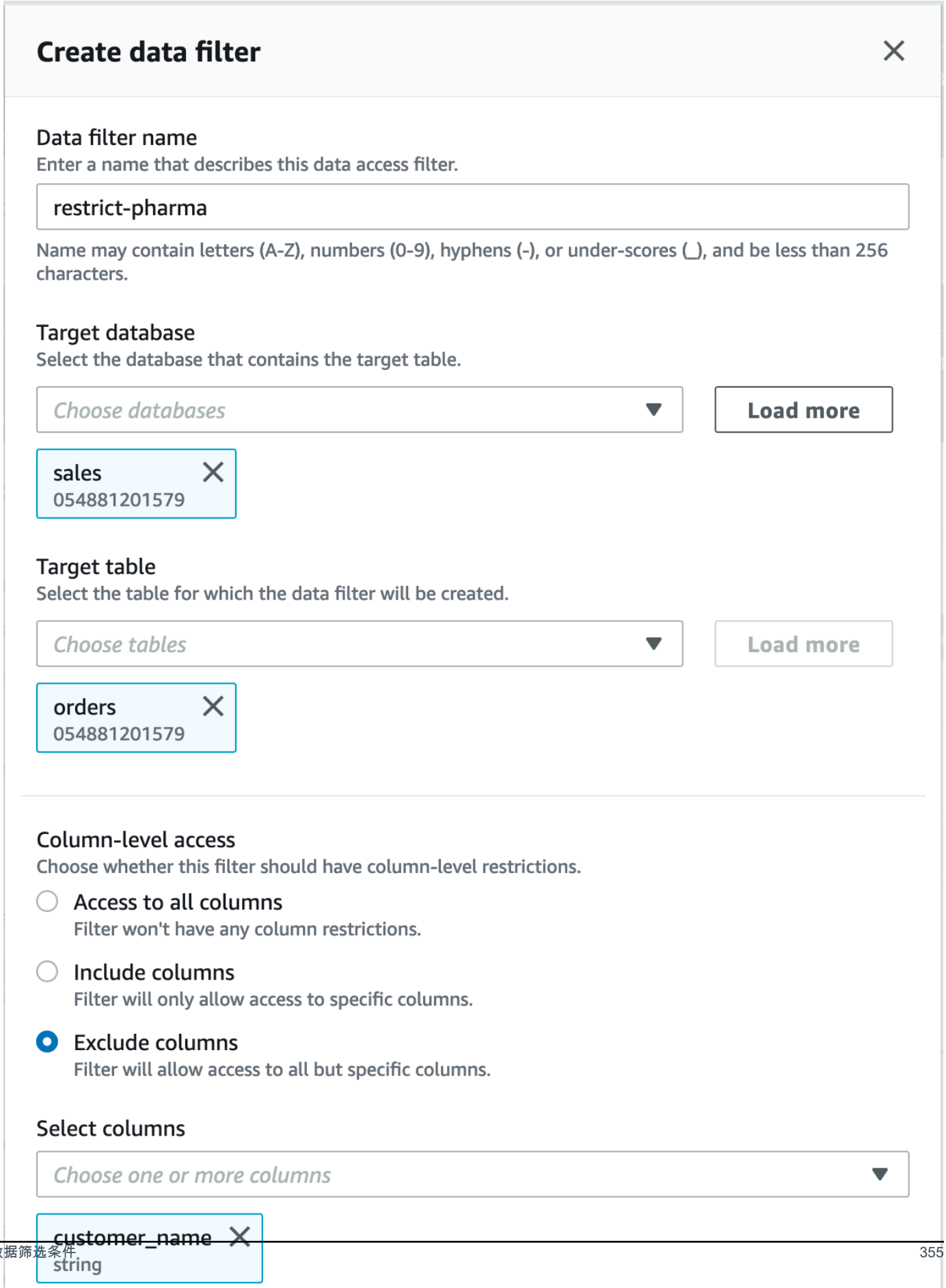
当您授予对 customer 列的权限时，该主体将获得对该列和该列下的嵌套字段（customerName 和 customerID）的访问权限。

- 行筛选表达式 – 输入筛选表达式以指定行或单元格筛选。有关支持的数据类型和运算符，请参阅行筛选表达式支持 PartiQL。选择访问所有行以授予对所有行的访问权限。

  您可以在行筛选条件表达式中包含嵌套列中的部分列结构，以筛选包含特定值的行。

  当主体被授予对带有行筛选条件表达式 `Select * from example_nestedtable where customer.customerName <>'John'` 的表的权限并且列级别访问权限设置为访问所有列时，查询结果仅会显示 customerName <>'John' 的计算结果为 true 的行。

以下屏幕截图显示了实现单元格筛选的数据筛选条件。在针对 orders 表的查询中，它拒绝对 customer_name 列的访问，并且仅显示 product_type 列中包含“pharma”的行。


Create data filter
Data filter name
Enter a name that describes this data access filter.
restrict-pharma
Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.
Target database
Select the database that contains the target table.
Choose databases
Load more
sales
054881201579
Target table
Select the table for which the data filter will be created.
Choose tables
Load more
orders
054881201579
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Select columns
Choose one or more columns
customer_name
string

5. 请选择 Create filter（创建筛选器）。

对嵌套字段使用单元格筛选条件策略创建数据筛选条件

本部分使用以下示例架构来演示如何创建数据单元格筛选条件：

```
[
    { name: "customer", type: "struct<customerId:string,customerName:string>" },
    { name: "customerApplication", type: "struct<appId:string>" },
    { name: "product", type:
 "struct<offer:struct<prodId:string,listingId:string>,type:string>" },
    { name: "purchaseId", type: "string" },
]
```

1. 在创建数据筛选条件页面上，输入数据筛选条件的名称。
2. 接下来，使用下拉列表选择数据库名称和表名称。
3. 在列级别访问部分中，选择包含的列，然后选择嵌套列 (customer.customerName)。
4. 在行级别访问部分中，选择访问所有行选项。
5. 请选择 Create filter（创建筛选器）。

   当您授予对此筛选条件的 SELECT 权限时，主体可以访问 customerName 列中的所有行。
6. 接下来，为同一个数据库/表定义另一个数据筛选条件。
7. 在列级别访问部分中，选择包含的列，然后选择另一个嵌套列 (customer.customerid)。
8. 在行级别访问部分中，选择筛选行，然后输入行筛选条件表达式 (customer.customerid <> 5)。
9. 请选择 Create filter（创建筛选器）。

   当您授予对此筛选条件的 SELECT 权限时，主体将获得对 customerName 和 customerId 字段中所有行的访问权限，但 customerId 列中值为 5 的单元格除外。

授予数据筛选条件权限

您可以向主体授予对数据筛选条件的 SELECT、DESCRIBE 和 DROP Lake Formation 权限。

首先，只有您可以查看为表创建的数据筛选条件。要使其他主体能够查看数据筛选条件并向数据筛选条件授予数据目录权限，您必须执行以下任一操作：

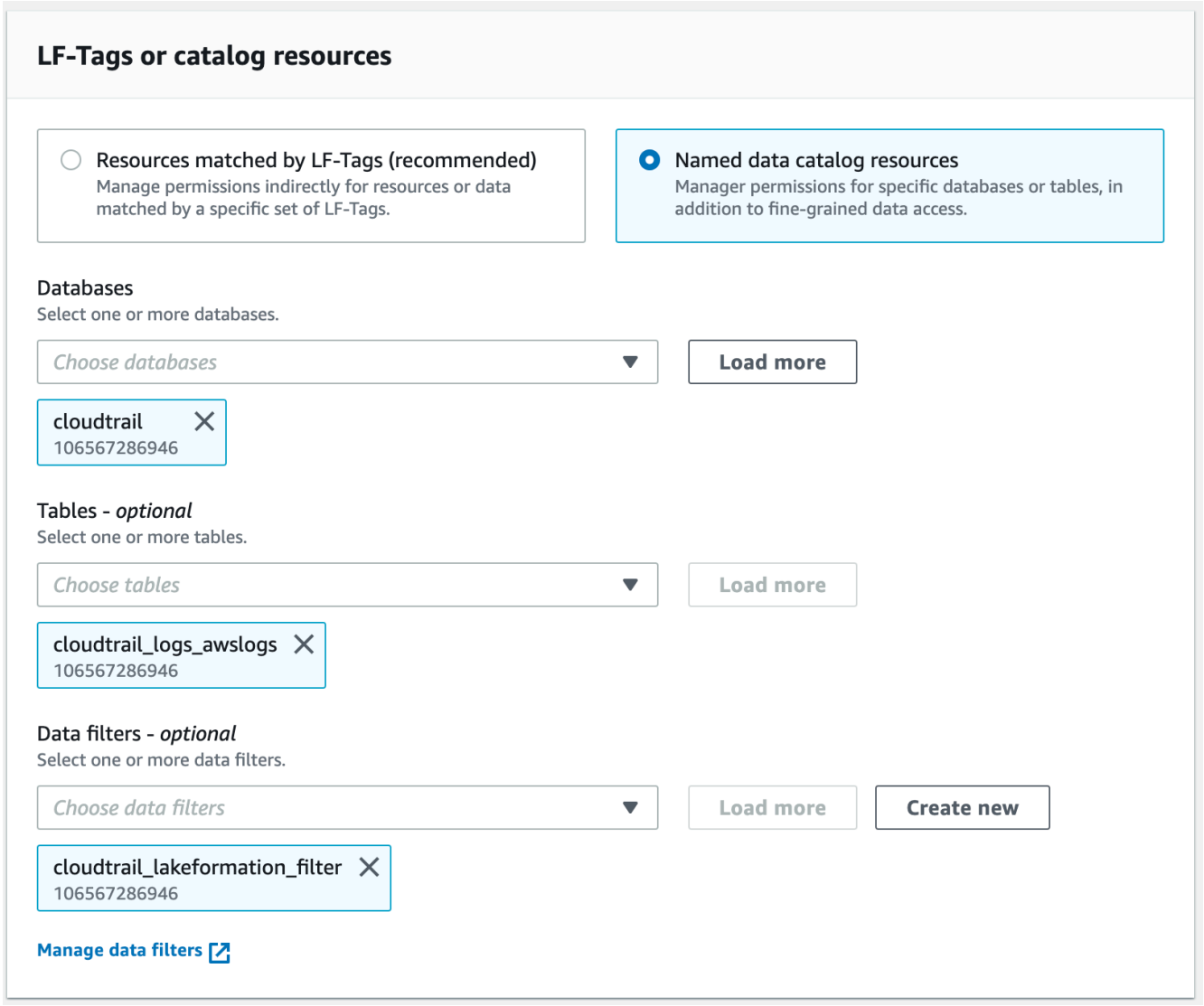
- 使用授予选项向主体授予对表的 SELECT 权限，然后将数据筛选条件应用于授权。
- 向主体授予对数据筛选条件的 DESCRIBE 或 DROP 权限。

您可以向外部 AWS 账户授予 SELECT 权限。然后，该账户中的数据湖管理员可以将该权限授予给该账户中的其他主体。向外部账户授予权限时，必须包括授予选项，以便外部账户的管理员可以进一步将权限级联到其账户中的其他用户。向账户中的主体授权时，可以使用授予选项进行授权。

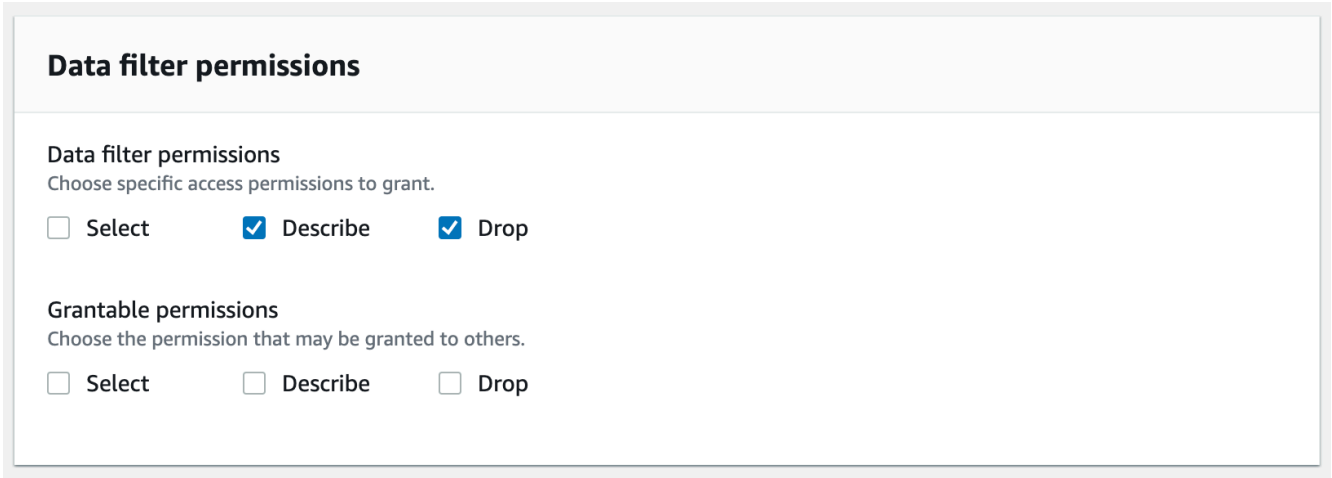
您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 授予和撤销对数据筛选条件的权限。

Console

1. 登录 AWS Management Console，然后通过以下网址打开 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在导航窗格的权限下，选择数据湖权限。
3. 在权限页面的数据权限部分中，选择授予。
4. 在授予数据权限页面上，选择要向其授予权限的主体。
5. 在“LF 标签或目录资源”部分下，选择已命名数据目录资源。然后选择要为其授予权限的数据库、表和数据筛选条件。



6. 在数据筛选条件权限部分中，选择要向所选主体授予的权限。



AWS CLI

- 输入 grant-permissions 命令。为 resource 参数指定 DataCellsFilter，并为 Permissions 参数和 PermissionsWithGrantOption 参数（可选）指定 DESCRIBE 或 DROP。

  以下示例在 AWS 账户 1111-2222-3333 中使用授予选项向用户 datalake_user1 授予对数据筛选条件 restrict-pharma 的 DESCRIBE 权限，该筛选条件属于 sales 数据库中的 orders 表。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

  以下是 grant-params.json 文件的内容。

```
{
    "Principal": {"DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["DESCRIBE"],
    "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

授予数据筛选条件提供的数据权限

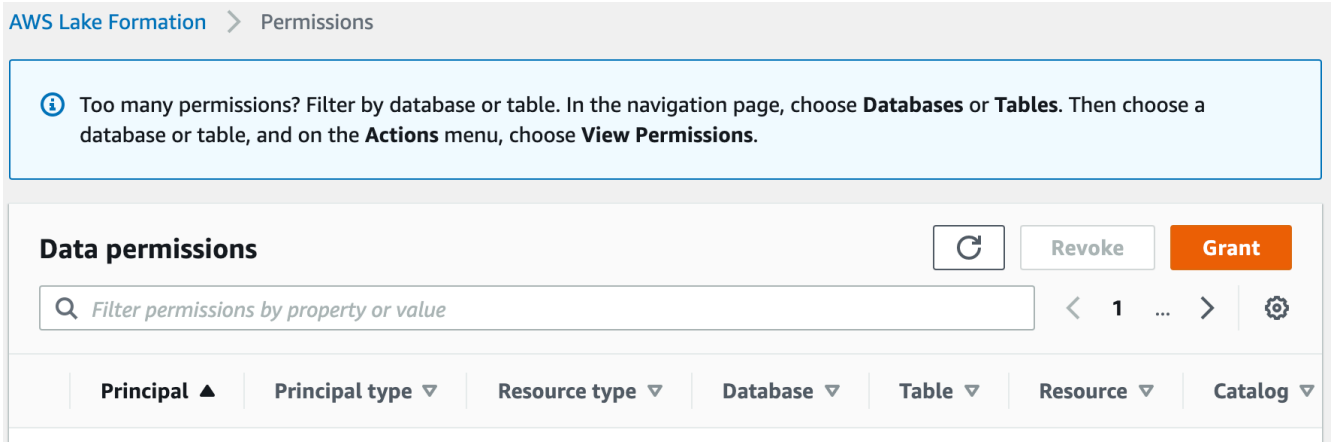
数据筛选条件表示表中数据的子集。要向主体提供数据访问权限，需要向这些主体授予 SELECT 权限。有了这项权限，主体可以：

- 在与其账户共享的表列表中查看实际的表名。
- 针对共享表创建数据筛选条件，并向其用户授予对这些数据筛选条件的权限。

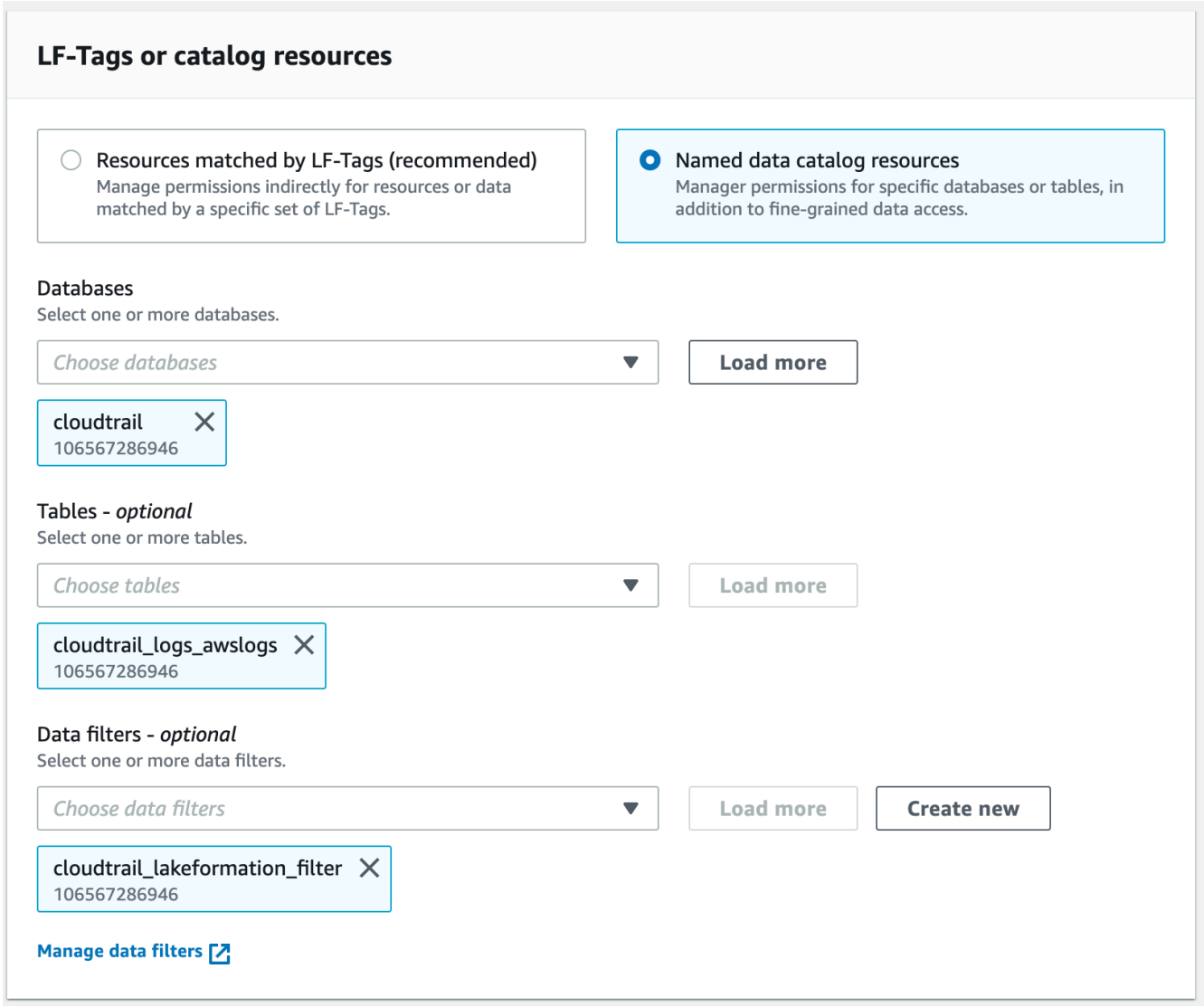
Console

授予 SELECT 权限

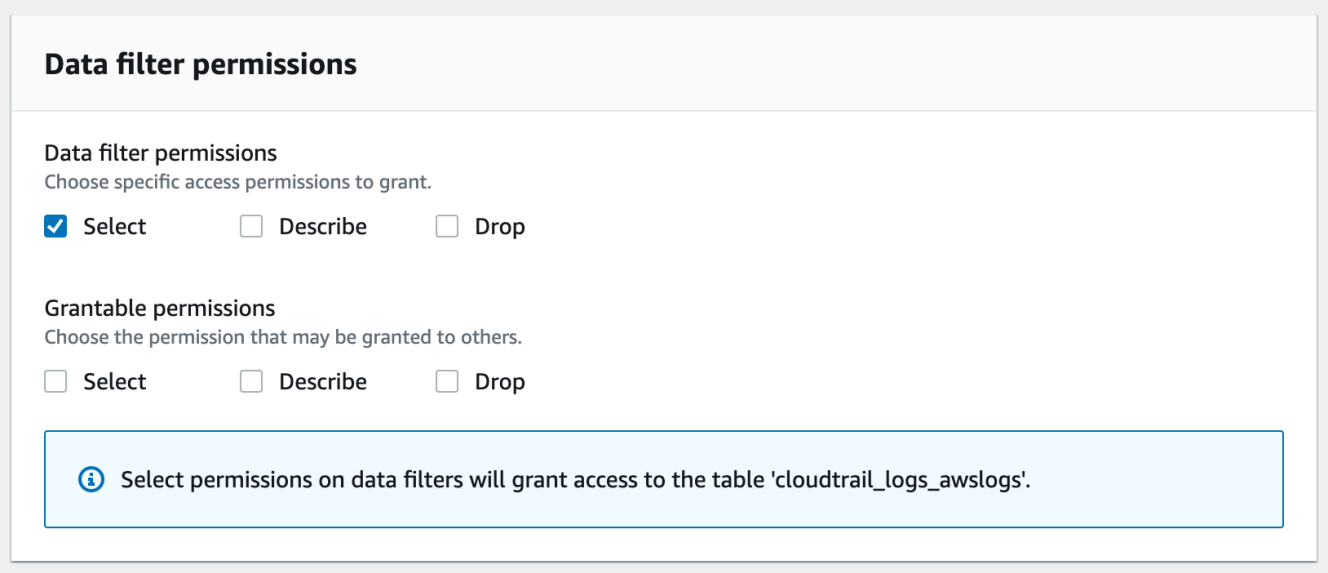
1. 前往 Lake Formation 控制台中的权限页面，然后选择授予。



2. 选择您要向其提供访问权限的主体，然后选择已命名数据目录资源。



3. 要提供对筛选条件所代表数据的访问权限，请在数据筛选条件权限下选择 Select。



CLI

输入 grant-permissions 命令。为资源参数指定 DataCellsFilter，并为“权限”参数指定 SELECT。

以下示例在 AWS 账户 1111-2222-3333 中使用授予选项向用户 datalake_user1 授予对数据筛选条件 restrict-pharma 的 SELECT 权限，该筛选条件属于 sales 数据库中的 orders 表。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是 grant-params.json 文件的内容。

```
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
    },
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["SELECT"]
```

```
}
```

查看数据筛选条件

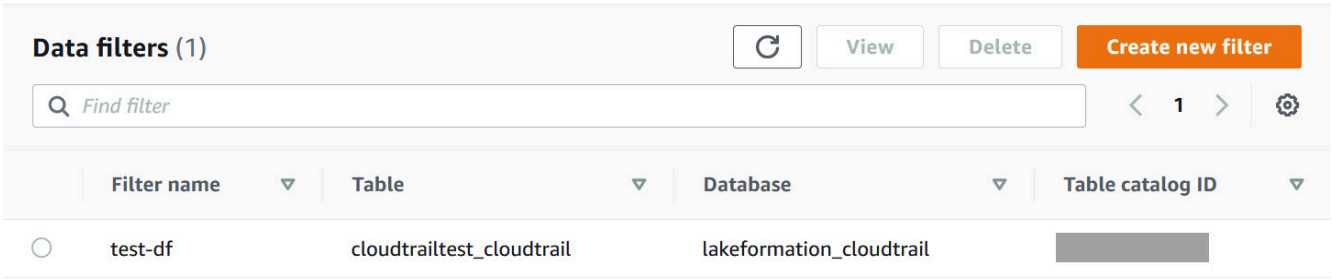
您可以使用 Lake Formation 控制台、AWS CLI 或 Lake Formation API 查看数据筛选条件。

要查看数据筛选条件，您必须是数据湖管理员或具有对数据筛选条件所需的权限。

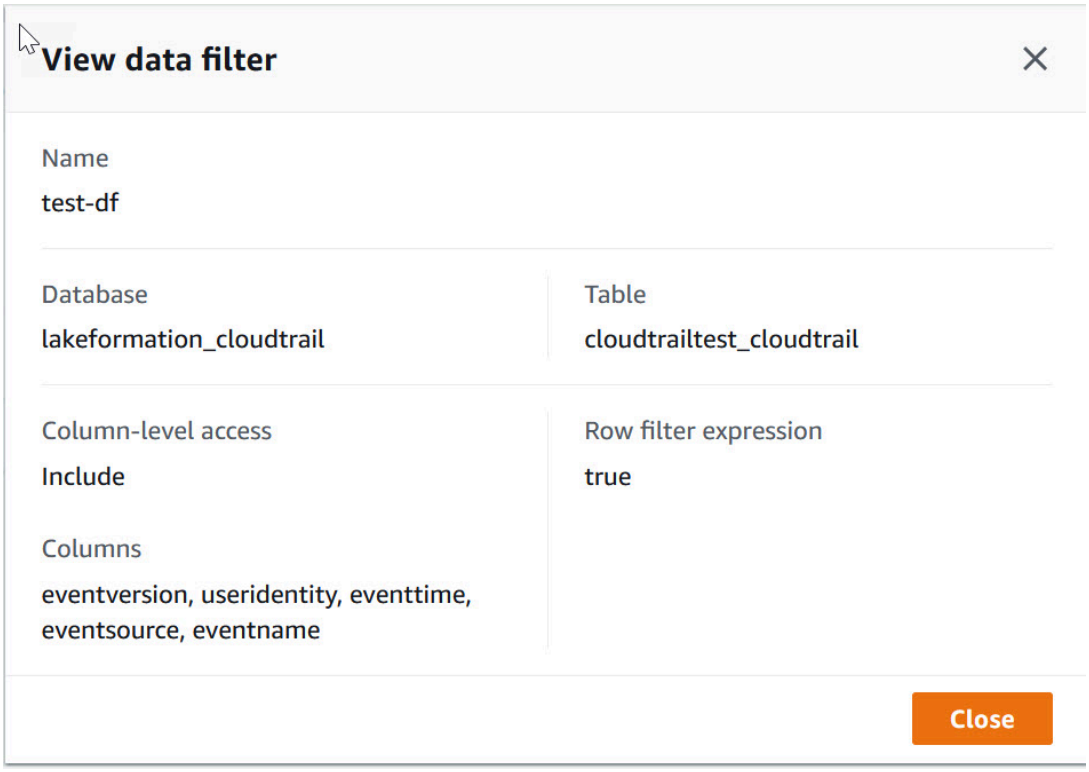
Console

1. 登录 AWS Management Console，然后通过以下网址打开 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在导航窗格的数据目录下，选择数据筛选条件。

   该页面显示了您可以访问数据筛选条件。



3. 要查看数据筛选条件的详细信息，请选择数据筛选条件，然后选择“查看”。此时将显示一个包含数据筛选条件详细信息的新窗口。



AWS CLI

输入 `list-data-cells-filter` 命令并指定表资源。

以下示例列出了 `cloudtrailtest_cloudtrail` 表的数据筛选条件。

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}'
```

API/SDK

使用 `ListDataCellsFilter` API 并指定表资源。

以下示例使用 Python 列出了 `myTable` 表的前 20 个数据筛选条件。

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
    },
```

```
    MaxResults=20
)
```

列出数据筛选条件权限

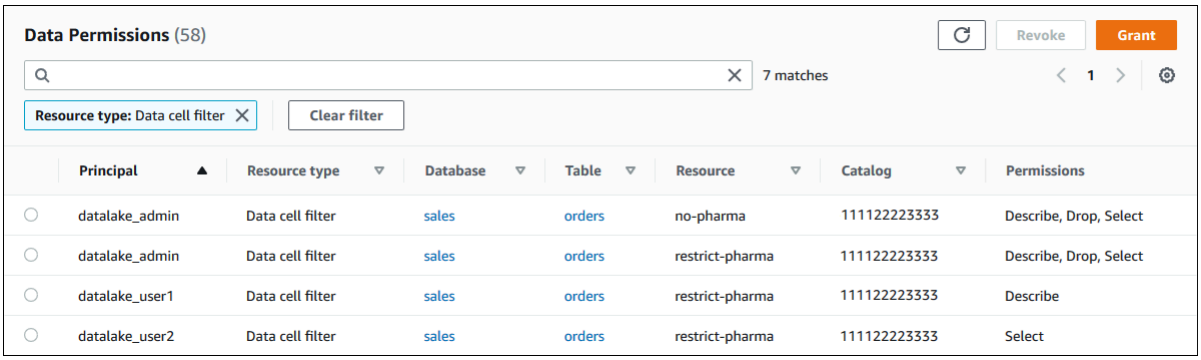
您可以使用 Lake Formation 控制台查看对数据筛选条件授予的权限。

要查看对某个数据筛选条件的权限，您必须是 Data Lake 管理员或具有对该数据筛选条件所需的权限。

Console

1. 登录 AWS Management Console，然后通过以下网址打开 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在导航窗格的权限下，选择数据权限。
3. 在数据权限页面上，单击或点按搜索字段，然后在属性菜单上选择资源类型。
4. 在资源类型菜单上，选择资源类型：数据单元格筛选条件。

   此时将列出您有权访问的数据筛选条件。您可能需要水平滚动才能看到权限和可授予列。



AWS CLI

- 输入 `list-permissions` 命令。为 `resource` 参数指定 `DataCellsFilter`，并为 `Permissions` 参数和 `PermissionsWithGrantOption` 参数（可选）指定 `DESCRIBE` 或 `DROP`。

  以下示例列出了使用授予选项授予的对数据筛选条件 `restrict-pharma` 的 `DESCRIBE` 权限。结果仅限于在 AWS 账户 1111-2222-3333 中向主体 `datalake_user1` 和 `sales` 数据库中的 `orders` 表授予的权限。

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

以下是 grant-params.json 文件的内容。

```
{
    "Principal": {"DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["DESCRIBE"],
    "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

在 Lake Formation 中查看数据库和表权限

您可以查看对数据目录数据库或表授予的 Lake Formation 权限。为此，您可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI)。

使用控制台，您可以从数据库或表页面或者从数据权限页面开始查看权限。

> **Note**
>
> 如果您不是数据库管理员或资源拥有者，则仅当您通过授予选项具有对资源的 Lake Formation 权限时，才能查看其他主体对资源具有的权限。
> 除了所需的 Lake Formation 权限外，您还需要 AWS Identity and Access Management (IAM) 权限 `glue:GetDatabases`、`glue:GetDatabase`、`glue:GetTables`、`glue:GetTable` 和 `glue:ListPermissions`。

查看对数据库的权限（使用控制台，从“数据库”页面开始）

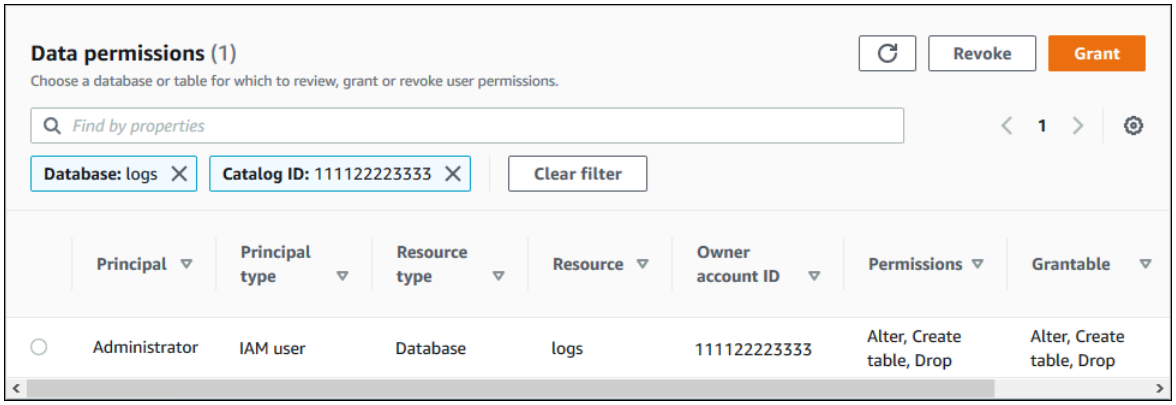
1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

以数据湖管理员、数据库创建者或通过授予选项具有对数据库的任何 Lake Formation 权限的用户身份登录。

2. 在导航窗格中，选择数据库。
3. 选择一个数据库，然后在操作菜单上选择查看权限。

> **Note**
>
> 如果选择数据库资源链接，则 Lake Formation 将显示对该资源链接的权限，而不是对该资源链接的目标数据库的权限。

数据权限页面列出了该数据库的所有 Lake Formation 权限。数据库拥有者的数据库名称和目录 ID（AWS 账户 ID）在搜索框下显示为标签。这些磁贴指示已应用筛选条件来仅列出该数据库的权限。您可以通过关闭磁贴或选择清除筛选条件来调整该筛选条件。



查看对数据库的权限（使用控制台，从“数据权限”页面开始）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员、数据库创建者或通过授予选项具有对数据库的任何 Lake Formation 权限的用户身份登录。
2. 在导航窗格中，选择数据权限。
3. 将光标置于页面顶部的搜索框中，然后在显示的属性菜单上选择数据库。
4. 在显示的数据库菜单上选择一个数据库。

> **Note**
>
> 如果选择数据库资源链接，则 Lake Formation 将显示对该资源链接的权限，而不是对该资源链接的目标数据库的权限。

数据权限页面列出了该数据库的所有 Lake Formation 权限。数据库名称在搜索框下显示为磁贴。该磁贴指示已应用筛选条件来仅列出该数据库的权限。您可以通过关闭磁贴或选择清除筛选条件来移除该筛选条件。

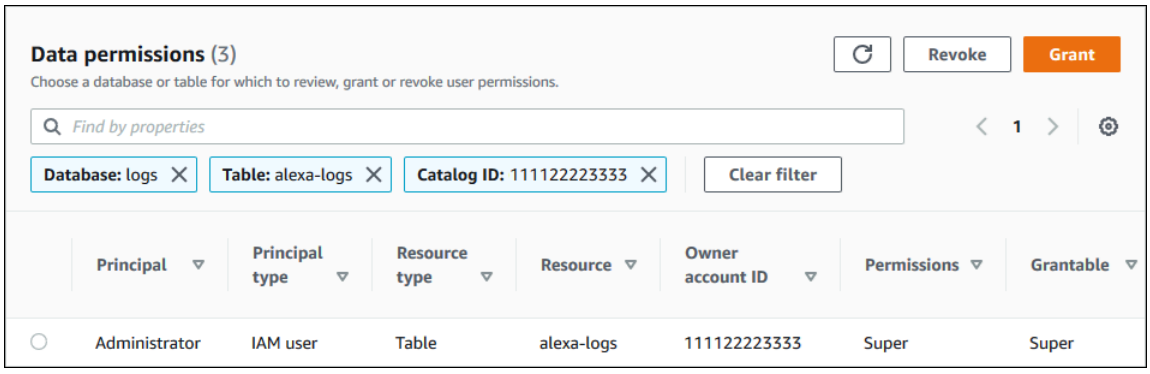
查看对表的权限（使用控制台，从“表”页面开始）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员、表创建者或通过授予选项具有对表的任何 Lake Formation 权限的用户身份登录。
2. 在导航窗格中，选择表。
3. 选择一个表，然后在操作菜单上选择查看权限。

> **Note**
>
> 如果选择表资源链接，则 Lake Formation 将显示对该资源链接的权限，而不是对该资源链接的目标表的权限。

数据权限页面列出了该表的所有 Lake Formation 权限。表拥有者的表名称、包含该表的数据库的数据库名称和目录 ID（AWS 账户 ID）在搜索框下显示为标签。这些标签指示已应用筛选条件来仅列出该表的权限。您可以通过关闭标签或选择清除筛选条件来调整该筛选条件。



查看对表的权限（使用控制台，从“数据权限”页面开始）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员、表创建者或通过授予选项具有对表的任何 Lake Formation 权限的用户身份登录。

2. 在导航窗格中，选择数据权限。
3. 将光标置于页面顶部的搜索框中，然后在显示的属性菜单上选择数据库。
4. 在显示的数据库菜单上选择一个数据库。

   > **Important**
   >
   > 如果要查看对从外部账户与您的 AWS 账户共享的表的权限，您必须选择包含该表的外部账户中的数据库，而不是指向该数据库的资源链接。

   数据权限页面列出了该数据库的所有 Lake Formation 权限。
5. 将光标再次置于搜索框中，然后在显示的属性菜单上选择表。
6. 在显示的表菜单上选择一个表。

   数据权限页面列出了该表的所有 Lake Formation 权限。表名称和包含该表的数据库的数据库名称在搜索框下显示为磁贴。这些磁贴指示已应用筛选条件来仅列出该表的权限。您可以通过关闭磁贴或选择清除筛选条件来调整该筛选条件。

查看对表的权限 (AWS CLI)

- 输入 `list-permissions` 命令。

  以下示例列出了对从外部账户共享的表的权限。`CatalogId` 属性是外部账户的 AWS 账户 ID，数据库名称是指外部账户中包含该表的数据库。

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

使用 Lake Formation 控制台撤销权限

您可以使用控制台撤销所有类型的 Lake Formation 权限，包括数据目录权限、策略标签权限、数据筛选条件权限和位置权限。

撤消对资源的 Lake Formation 权限（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以数据湖管理员或以已通过授予选项被授予对该资源的权限的用户身份登录。
2. 在导航窗格的权限下，选择数据湖权限、LF 标签和权限或数据位置。
3. 选择权限或位置，然后选择撤销。
4. 在打开的对话框中，选择撤销。

Lake Formation 中的跨账户数据共享

Lake Formation 跨账户功能允许用户安全地跨多个 AWS 组织共享分布式数据湖 AWS 账户，或者直接与其他账户中的 IAM 委托人共享分布式数据湖，从而提供对数据目录元数据和底层数据的精细访问权限。大型企业通常使用多个账户 AWS 账户，其中许多账户可能需要访问由单个账户管理的数据湖 AWS 账户。用户和 AWS Glue 提取、转换和加载 (ETL) 作业可以跨多个账户查询和联接表，但仍然可以利用 Lake Formation 表级和列级数据保护。

当您将数据目录资源的 Lake Formation 权限授予外部账户或直接授予其他账户中的 IAM 委托人时，Lake Formation 会使用 AWS Resource Access Manager (AWS RAM) 服务共享该资源。如果被授权者账户与授予者账户在同一个组织中，则被授权者立即可以使用共享资源。如果被授权者账户不在同一个组织中，则会向被授权者账户 AWS RAM 发送邀请，要求其接受或拒绝资源授予。然后，要使共享资源可用，被授权者账户中的数据湖管理员必须使用 AWS RAM 控制台或接受 AWS CLI 邀请。

Lake Formation 支持在混合访问模式下与外部账户共享数据目录资源。混合访问模式使您可以灵活地且有选择性地为 AWS Glue Data Catalog中的数据库和表启用 Lake Formation 权限。
在混合访问模式下，您现在有了增量路径，可您为一组特定的用户设置 Lake Formation 权限，而不会中断其他现有用户或工作负载的权限策略。

有关更多信息，请参阅混合访问模式。

直接跨账户共享

授权主体可以与外部账户中的 IAM 主体显式共享资源。当账户所有者想要控制外部账户中谁可以访问资源时，此功能非常有用。IAM 主体获得的权限将是直接授权和向下级联到主体的账户级别授予的并

集。接收方账户的数据湖管理员可以查看直接跨账户授权，但无法撤销权限。获得资源共享的主体不能与其他主体共享该资源。

共享数据目录资源的方法

通过单个 Lake Formation 授权操作，您可以授予对以下数据目录资源的跨账户权限。

- 数据库
- 单个表（带有可选列筛选功能）
- 一些选定表
- 数据库中的所有表（通过使用“所有表”通配符）

您可以通过两种方式与另一账户中的其他用户 AWS 账户 或另一账户中的 IAM 委托人共享您的数据库和表。

- Lake Formation 基于标签的访问控制 (LF-TBAC)（推荐）

  Lake Formation 基于标签的访问控制是一种授权策略，它根据属性来定义权限。您可以使用基于标签的访问控制与外部 IAM 委托人、Organizations 和组织单位 (OU) 共享数据目录资源（数据库、AWS 账户表和列）。在 Lake Formation 中，这些属性被称为“LF 标签”。有关更多信息，请参阅使用 Lake Formation 基于标签的访问控制管理数据湖。

  > **Note**
  >
  > 授予数据目录权限的 LF-TBAC 方法用于 AWS Resource Access Manager 跨账户授权。Lake Formation 现在支持使用 LF-TBAC 方法向组织和组织单位授予跨账户权限。
  > 要启用此功能，您需要将跨账户版本设置更新为版本 3。
  > 有关更多信息，请参阅更新跨账户数据共享版本设置。

- Lake Formation 命名资源

  使用命名资源方法的 Lake Formation 跨账户数据共享允许您向外部 AWS 账户、IAM 委托人、组织或组织单位授予 Lake Formation 权限以及数据目录表和数据库的授予选项。授权操作可自动将这些资源共享。

> **Note**
>
> 您还可以允许 AWS Glue 爬网程序使用 Lake Formation 凭据访问其他账户中的数据存储。有关更多信息，请参阅 AWS Glue 开发者指南中的跨账户抓取。

Athena 和 Amazon Redshift Spectrum 等集成服务需要资源链接才能在查询中包含共享资源。有关资源链接的更多信息，请参阅资源链接在 Lake Formation 中的工作原理。

有关注意事项和限制，请参阅跨账户数据共享最佳实践和注意事项。

主题

> **相关 主题**
>
> - Lake Formation 权限概述
> - 访问和查看共享数据目录表和数据库
> - 创建资源链接
> - 对跨账户访问问题进行故障排除

先决条件

您的 AWS 账户必须满足以下先决条件，然后才能与其他账户中的其他账户或委托人共享数据目录资源（数据库和表），然后才能访问与您的账户共享的资源。

跨账户数据共享的一般要求

- 要在混合访问模式下共享数据目录数据库和表，您需要将跨账户版本设置更新为版本 4。
- 在向数据目录资源授予跨账户权限之前，必须撤销该 IAMAllowedPrincipals 组对该资源的所有 Lake Formation 权限。如果调用主体拥有跨账户访问资源的权限，并且对于该资源存在 IAMAllowedPrincipals 权限，则 Lake Formation 会引发 AccessDeniedException。

  仅当在 Lake Formation 模式下注册基础数据位置时，此要求才适用。如果在混合模式下注册数据位置，则对于共享数据库或表可以存在 IAMAllowedPrincipals 组权限。
- 对于包含要共享的表的数据库，必须防止新表的默认授权 Super 为 IAMAllowedPrincipals。在 Lake Formation 控制台上，编辑数据库并关闭 “仅对该数据库中的新表使用 IAM 访问控制”，或者输入以下 AWS CLI 命令（**database**替换为数据库名称）。如果基础数据位置是在混合访问模式下注册的，则无需更改此默认设置。在混合访问模式下，Lake Formation 允许您有选择地对亚马逊 S3 和同一资源强制执行 Lake Formati AWS Glue on 权限和 IAM 权限策略。

```
aws glue update-database --name database --database-input
 '{"Name":"database","CreateTableDefaultPermissions":[]}'
```

- 要授予跨账户权限，授予者必须对AWS Glue和 AWS RAM 服务具有必需的 AWS Identity and Access Management (IAM) 权限。 AWS 托管策略AWSLakeFormationCrossAccountManager授予所需的权限。

  使用接收资源共享的账户中的数据湖管理员 AWS RAM 必须遵守以下附加策略。它允许管理员接受 AWS RAM 资源共享邀请。它还允许管理员启用与组织的资源共享。

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

```
    ]
}
```

- 如果要与 AWS Organizations 或组织单位共享数据目录资源，则必须在中启用与组织共享 AWS RAM。

  有关如何启用与组织共享的信息，请参阅《AWS RAM 用户指南》中的启用与 AWS 组织共享。

  您必须拥有用于启用与组织的共享所需的 `ram:EnableSharingWithAwsOrganization` 权限。
- 要直接与其他账户中的 IAM 主体共享资源，您需要将跨账户版本设置更新为版本 3。此设置在数据目录设置页面上提供。如果您使用的是版本 1，请参阅有关更新设置 更新跨账户数据共享版本设置 的说明。
- 您不能与其他账户共享使用 AWS Glue 服务托管密钥加密的数据目录资源。您只能共享使用客户加密密钥加密的数据目录资源，并且接收资源共享的账户必须对数据目录加密密钥拥有相应权限才能解密对象。

使用 LF-TBAC 要求进行跨账户数据共享

- 要 AWS Organizations 与组织单位 (OU) 共享数据目录资源，您需要将跨账户版本设置更新为版本 3。
- 要使用“版本 3”这一跨账户版本设置共享数据目录资源，授予者需要拥有在您的账户中的 AWS 托管策略 `AWSLakeFormationCrossAccountManager` 中定义的 IAM 权限。
- 如果您使用的跨账户版本设置是“版本 1”或“版本 2”，则必须具有启用 LF-TBAC 的数据目录资源策略 (`glue:PutResourcePolicy`)。有关更多信息，请参阅使用 AWS Glue 和 Lake Formation 管理跨账户权限。。
- 如果您当前在使用 AWS Glue 数据目录资源策略共享资源，并且想要使用“版本 3”这一跨账户版本设置授予跨账户权限，则必须使用 `glue:PutResourcePolicy` API 操作在数据目录设置中添加 `glue:ShareResource` 权限，如使用 AWS Glue 和 Lake Formation 管理跨账户权限。一节所示。如果您的账户未进行跨账户授权，即未使用 AWS Glue 数据目录资源策略（版本 1 和版本 2 使用 `glue:PutResourcePolicy` 权限）授予跨账户访问权限，则无需此策略。

```
{
      "Effect": "Allow",
      "Action": [
        "glue:ShareResource"
      ],
      "Principal": {"Service": [
        "ram.amazonaws.com"
```

```
    ]},
    "Resource": [
      "arn:aws:glue:<region>:<account-id>:table/*/*",
      "arn:aws:glue:<region>:<account-id>:database/*",
      "arn:aws:glue:<region>:<account-id>:catalog"
    ]
  }
```

- 如果您的账户已使用 AWS Glue 数据目录资源策略进行跨账户共享，并且您当前在使用命名资源方法或跨账户设置为“版本 3”的 LF-TBAC 来共享资源（使用 AWS RAM 共享资源），则在调用 `glue:PutResourcePolicy` API 操作时必须将 `EnableHybrid` 参数设置为 `'true'`。有关更多信息，请参阅使用 AWS Glue 和 Lake Formation 管理跨账户权限。。

访问共享资源的每个账户所需的设置

- 如果您与共享资源 AWS 账户，则消费者账户中必须至少有一个用户是数据湖管理员才能查看共享资源。有关如何创建数据湖管理员的信息，请参阅创建数据湖管理员。

  数据湖管理员可以向账户中的其他主体授予对共享资源的 Lake Formation 权限。在数据湖管理员向其他主体授予对共享资源的权限之前，这些主体无法访问共享资源。
- Athena 和 Redshift Spectrum 等集成服务需要资源链接才能在查询中包含共享资源。主体需要在其数据目录中创建指向其他 AWS 账户中共享资源的资源链接。有关资源链接的更多信息，请参阅资源链接在 Lake Formation 中的工作原理。
- 当直接与 IAM 主体共享资源时，要使用 Athena 查询表，则主体需要创建资源链接。要创建资源链接，主体需要 Lake Formation `CREATE_TABLE` 或 `CREATE_DATABASE` 权限，以及 `glue:CreateTable` 或 `glue:CreateDatabase` IAM 权限。

  如果制作者账户与同一主体或其他主体共享同一数据库下的不同表，则该主体立即可以查询该表。

> **Note**
>
> 对于数据湖管理员和数据湖管理员已向其授予权限的主体，共享资源在数据目录中显示为本地（自有）资源。提取、转换、加载 (ETL) 作业可以访问共享资源的基础数据。
> 对于共享资源，Lake Formation 控制台上的表和数据库页面会显示所有者的账户 ID。

访问共享资源的底层数据时，将在共享资源接收者的账户和资源所有者的账户中生成 CloudTrail 日志事件。 CloudTrail 事件可以包含访问数据的委托人的 ARN，但前提是接收方账户选择在日志中包含委托人 ARN。有关更多信息，请参阅跨账户 CloudTrail 日志。

更新跨账户数据共享版本设置

不时 AWS Lake Formation 更新跨账户数据共享设置，以区分对 AWS RAM 使用情况所做的更改，并支持对跨账户数据共享功能所做的更新。当 Lake Formation 执行此操作时，它会创建新版本的跨账户版本设置。

跨账户版本设置之间的主要区别

有关如何在不同跨账户版本设置下进行跨账户数据共享的更多信息，请参阅以下章节。

> **Note**
>
> 要与其他账户共享数据，授予者必须拥有 `AWSLakeFormationCrossAccountManager` 托管 IAM 策略权限。这是所有版本的先决条件。
> 更新跨账户版本设置不会影响接收方对共享资源的权限。这适用于从版本 1 更新为版本 2、从版本 2 更新为版本 3 以及从版本 1 更新为版本 3 的情况。更新版本时，请参阅下面列出的注意事项。

版本 1

命名资源方法：将授予的每个跨账户 Lake Formation 权限映射到一个 AWS RAM 资源共享。用户（授予者角色或主体）不需要其他权限。

LF-TBAC 方法：跨账户 Lake Formation 权限授予不 AWS RAM 用于共享数据。您必须拥有 `glue:PutResourcePolicy` 权限。

更新版本的好处：初始版本 - 不适用。

更新版本时的注意事项：初始版本 - 不适用

版本 2

命名资源方法：通过将多个跨账户权限授予映射到一个 AWS RAM 资源共享来优化 AWS RAM 资源共享的数量。用户不需要额外的权限。

LF-TBAC 方法：跨账户 Lake Formation 权限授予不 AWS RAM 用于共享数据。您必须拥有 `glue:PutResourcePolicy` 权限。

更新版本的好处：通过优化 AWS RAM 容量利用率实现可扩展的跨账户设置。

更新版本时的注意事项：想要授予跨账户 Lake Formation 权限的用户必须拥有`AWSLakeFormationCrossAccountManager` AWS 托管策略中的权限。否则，您需要拥有 `ram:AssociateResourceShare` 和 `ram:DisassociateResourceShare` 权限才能成功与其他账户共享资源。

版本 3

命名资源方法：通过将多个跨账户权限授予映射到一个 AWS RAM 资源共享来优化 AWS RAM 资源共享的数量。用户不需要额外的权限。

LF-TBAC 方法：Lake Formation 用于 AWS RAM 跨账户拨款。用户必须在`glue:PutResourcePolicy`权限中添加 glue: ShareResource 语句。收件人必须接受来自的资源共享邀请 AWS RAM。

更新版本的好处：支持以下功能：

- 允许与外部账户中的 IAM 主体显式共享资源。

  有关更多信息，请参阅授予和撤销对数据目录资源的权限。
- 使用 LF-TBAC 方法为组织或组织单位 (OU) 启用跨账户共享。
- 消除了维护跨账户授予额外 AWS Glue 政策的开销。

更新版本时的注意事项：如果授予者使用的版本低于版本 3，而接收方使用的是版本 3 或更高版本，则授予者会收到以下错误消息：“跨账户授权请求无效。使用者账户可以选择使用跨账户版本：v3。请更新`CrossAccountVersionDataLakeSetting`至最低版本 v3（服务：AmazonDataCatalog；状态码：400；错误代码： InvalidInputException）”。但是，如果授予者使用版本 3，而接收方使用的是版本 1 或版本 2，则跨账户授权将成功完成。

要直接与其他账户中的 IAM 主体共享资源，只有授予者需要使用版本 3。

使用 LF-TBAC 方法进行的跨账户授权要求用户的账户中具有 AWS Glue Data Catalog 资源策略。当您更新为版本 3 时，LF-TBAC 会授权使用 AWS RAM。要允许 AWS RAM 基于跨账户的授予成功，您必须将该`glue:ShareResource`声明添加到现有的 Data Catalog 资源策略中，如使用 AWS Glue 和 Lake Formation 管理跨账户权限。部分所示。

版本 4

授予者需要版本 4 或更高版本才能在混合访问模式下共享数据目录资源。

优化 AWS RAM 资源共享

跨账户赠款的新版本（第 2 版及更高版本）以最佳方式利用 AWS RAM 容量来最大限度地提高跨账户使用率。当您与外部 AWS 账户 或 IAM 委托人共享资源时，Lake Formation 可能会创建新的资源共享或将该资源与现有共享关联。通过与现有共享关联，Lake Formation 减少了用户需要接受的资源共享邀请的数量。

通过 TBAC 启用 AWS RAM 共享或直接与委托人共享资源

要直接与其他账户中的 IAM 主体共享资源或者为组织或组织单位启用 TBAC 跨账户共享，您需要将跨账户版本设置更新为版本 3。有关 AWS RAM 资源限制的更多信息，请参阅跨账户数据共享最佳实践和注意事项。

更新跨账户版本设置所需的权限

如果跨账户权限授予者拥有 AWSLakeFormationCrossAccountManager 托管 IAM 策略权限，则无需为跨账户权限授予者角色或主体进行额外的权限设置。但是，如果跨账户授予者未使用托管策略，则授予者角色或主体应授予以下 IAM 权限，新版本的跨账户授权才能成功完成。

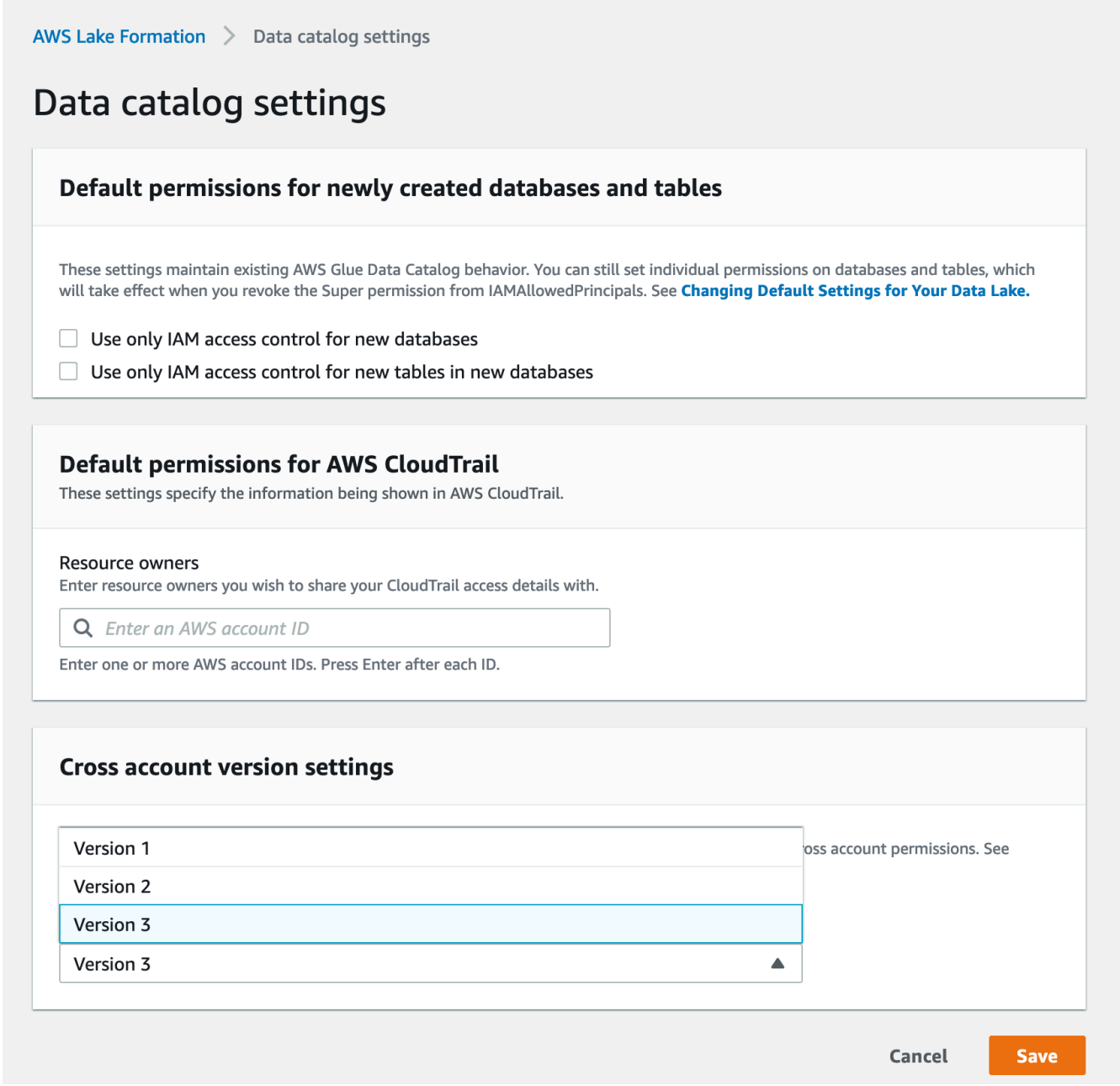
```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
         "ram:ResourceShareName": "LakeFormation*"
        }
      }
    }
  ]
}
```

启用新版本

按照以下步骤通过 AWS Lake Formation控制台或更新跨账户版本设置 AWS CLI。

Console

1. 在数据目录设置页面的跨账户版本设置下选择版本 2、版本 3 或版本 4。如果您选择版本 1，Lake Formation 将使用默认的资源共享模式。



2. 选择保存。

AWS Command Line Interface (AWS CLI)

使用`put-data-lake-settings` AWS CLI 命令设置`CROSS_ACCOUNT_VERSION`参数。可接受的值为 1、2、3 和 4。

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
    "DataLakeAdmins": [
        {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/test"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
        "CROSS_ACCOUNT_VERSION": "3"
    }
}
```

> **Important**
>
> 选择版本 2 或版本 3 后，所有新的命名资源授权都将通过新的跨账户授权模式进行。为了以最佳方式使用现有跨账户共享的 AWS RAM 容量，我们建议您撤销旧版本的授权，并在新模式下重新授权。

跨 AWS 账户 或来自外部账户的 IAM 主体共享数据目录表和数据库

本节包括有关如何为外部账户、IAM 委托人、组织或组织单位启用对数据目录表和数据库的跨 AWS 账户权限的说明。授权操作可自动将这些资源共享。

主题

- 使用基于标签的访问控制共享数据
- 使用命名资源方法进行跨账户数据共享

使用基于标签的访问控制共享数据

需要在制作者/授予者账户上进行的设置

1. 定义 LF 标签。有关创建 LF 标签的说明，请参阅创建 LF 标签。
2. 将 LF 标签分配给目标资源。有关更多信息，请参阅将 LF 标签分配给数据目录资源。
3. 向外部账户授予 LF 标签权限。有关更多信息，请参阅使用控制台授予 LF 标签权限。

   此时，使用者数据湖管理员应该能够在权限、管理角色和任务、LF 标签下找到通过被授权者账户 Lake Formation 控制台共享的策略标签。
4. 您可以向外部/被授予者账户授予数据权限。
   a. 在导航窗格中权限下的数据湖权限中，选择授权。
   b. 对于委托人，选择外部账户，然后输入委托人的目标 AWS 账户 ID 或 IAM 角色或委托人（委托人 ARN）的 Amazon 资源名称 (ARN)。
   c. 对于 LF 标签或目录资源，选择与使用者账户（键 `Confidentiality`和值 `public`）共享的 LF 标签的键和值。
   d. 对于权限，在通过 LF 标签匹配的资源（推荐）下，选择添加 LF 标签。
   e. 选择与被授权者账户（键 `Confidentiality` 和值 `public`）共享的标签的键和值。
   f. 对于数据库权限，选择数据库权限下的描述以授予数据库级别的访问权限。
   g. 使用者数据湖管理员应该能够通过 https://console.aws.amazon.com/lakeformation/ 访问 Lake Formation 控制台，在权限下管理角色和任务的 LF 标签中找到通过使用者账户共享的策略标签。
   h. 在可授予的权限下选择描述，这样使用者账户就可以向其用户授予数据库级别的权限。

      由于数据湖管理员必须向被授权者账户中的主体授予对共享资源的权限，因此必须始终使用授权选项授予跨账户权限。

      > **Note**
      >
      > 对于获得直接跨账户授权的主体，将不提供可授予的权限选项。

   i. 对于表和列权限，选择表权限下的选择和描述。
   j. 在可授予的权限下选择选择和描述。
   k. 选择授权。

需要在接收/被授予者账户上进行的设置

1. 当您与其他账户共享资源时，该资源仍属于制作者账户，并且在 Athena 控制台中不可见。要使资源在 Athena 控制台中可见，您需要创建一个指向共享资源的资源链接。有关创建资源链接的说明，请参阅创建指向共享数据目录表的资源链接和创建指向共享数据目录数据库的资源链接
2. 您需要在使用者账户中创建一组单独的 LF 标签，以便在共享资源链接时使用基于 LF 标签的访问控制。创建所需的 LF 标签并将其分配给共享数据库/表和资源链接。
3. 向被授权者账户中的 IAM 主体授予对这些 LF 标签的权限。

使用命名资源方法进行跨账户数据共享

您可以直接向其他 AWS 账户中的委托人授予权限，也可以向外部 AWS 账户 或 AWS Organizations授予权限。向组织或组织单位授予 Lake Formation 权限等同于向该组织或组织单位 AWS 账户 中的每个人授予权限。

向外部账户或组织授予权限时，必须包括可授予的权限选项。只有外部账户中的数据湖管理员才能访问共享资源，直到管理员向外部账户中的其他主体授予对共享资源的权限。

> **Note**
>
> 直接向外部账户中的 IAM 主体授予权限时，可授予的权限选项不受支持。

请按照使用命名资源方法授予数据库权限中的说明，使用命名资源方法来授予跨账户权限。

授予对与您的账户共享的数据库或表的权限

将属于另一个 AWS 账户的数据目录资源与您的 AWS 账户共享后，作为数据湖管理员，您可以向账户中的其他委托人授予共享资源的权限。但是，您不能向其他 AWS 账户或组织授予对该资源的权限。

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予权限。

授予对共享数据库的权限（命名资源方法、控制台）

- 按照使用命名资源方法授予数据库权限中的说明进行操作。在 LF 标签或目录资源下的数据库列表中，确保选择的是外部账户中的数据库，而不是数据库的资源链接。

如果您在数据库列表中看不到该数据库，请确保您已接受关于该数据库的 AWS Resource Access Manager (AWS RAM) 资源共享邀请。有关更多信息，请参阅接受来自 AWS RAM 的资源共享邀请。

另外，对于 CREATE_TABLE 和 ALTER 权限，请按照授予数据位置权限（同一账户）中的说明进行操作，并确保在注册账户位置字段中输入所有者账户 ID。

授予对共享表的权限（命名资源方法、控制台）

- 按照使用命名资源方法授予表权限中的说明进行操作。在 LF 标签或目录资源下的数据库列表中，确保选择的是外部账户中的数据库，而不是数据库的资源链接。

  如果您在表列表中看不到该表，请确保您已接受关于该表的 AWS RAM 资源共享邀请。有关更多信息，请参阅接受来自 AWS RAM 的资源共享邀请。

  另外，对于 ALTER 权限，请按照授予数据位置权限（同一账户）中的说明进行操作，并确保在注册账户位置字段中输入所有者账户 ID。

授予对共享资源的权限（LF-TBAC 方法、控制台）

- 按照授予数据目录权限 中的说明进行操作。在 LF 标签或目录资源部分，授予外部账户向您的账户授予的精确 LF 标签表达式或该表达式的子集。

  例如，如果外部账户使用授权选项向您的账户授予了 LF 标签表达式 `module=customers AND environment=production`，则作为数据湖管理员，您可以向您账户中的主体授予相同的表达式，即 `module=customers` 或 `environment=production`。您只能授予通过 LF 标签表达式授予的对资源的相同或部分 Lake Formation 权限（例如 SELECT、ALTER 等）。

要授予共享表（命名资源方法 AWS CLI）的权限

- 输入类似以下的命令。在本示例中：
  - 你的 AWS 账户编号是 1111-2222-3333。
  - 拥有该表并将其授予您的账户的账户是 1234-5678-9012。
  - 正在向用户 `datalake_user1` 授予对共享表 `pageviews` 的 SELECT 权限。该用户是您账户中的主体。
  - `pageviews` 表位于 `analytics` 数据库中，该数据库归账户 1234-5678-9012 所有。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"CatalogId":"123456789012",
 "DatabaseName":"analytics", "Name":"pageviews"}}'
```

请注意，必须在 `resource` 参数的 `CatalogId` 属性中指定所有者账户。

授予资源链接权限

按照以下步骤向 AWS 账户中的委托人授予对一个或多个资源链接的 AWS Lake Formation 权限。

创建资源链接时，只有您可以查看和访问它。（这假定没有为该数据库启用仅对此数据库中的新表使用 IAM 访问控制。）要允许您账户中的其他主体访问资源链接，请至少授予 `DESCRIBE` 权限。

> **Important**
>
> 授予对资源链接的权限不会授予对目标（链接）数据库或表的权限。您必须单独授予对目标的权限。

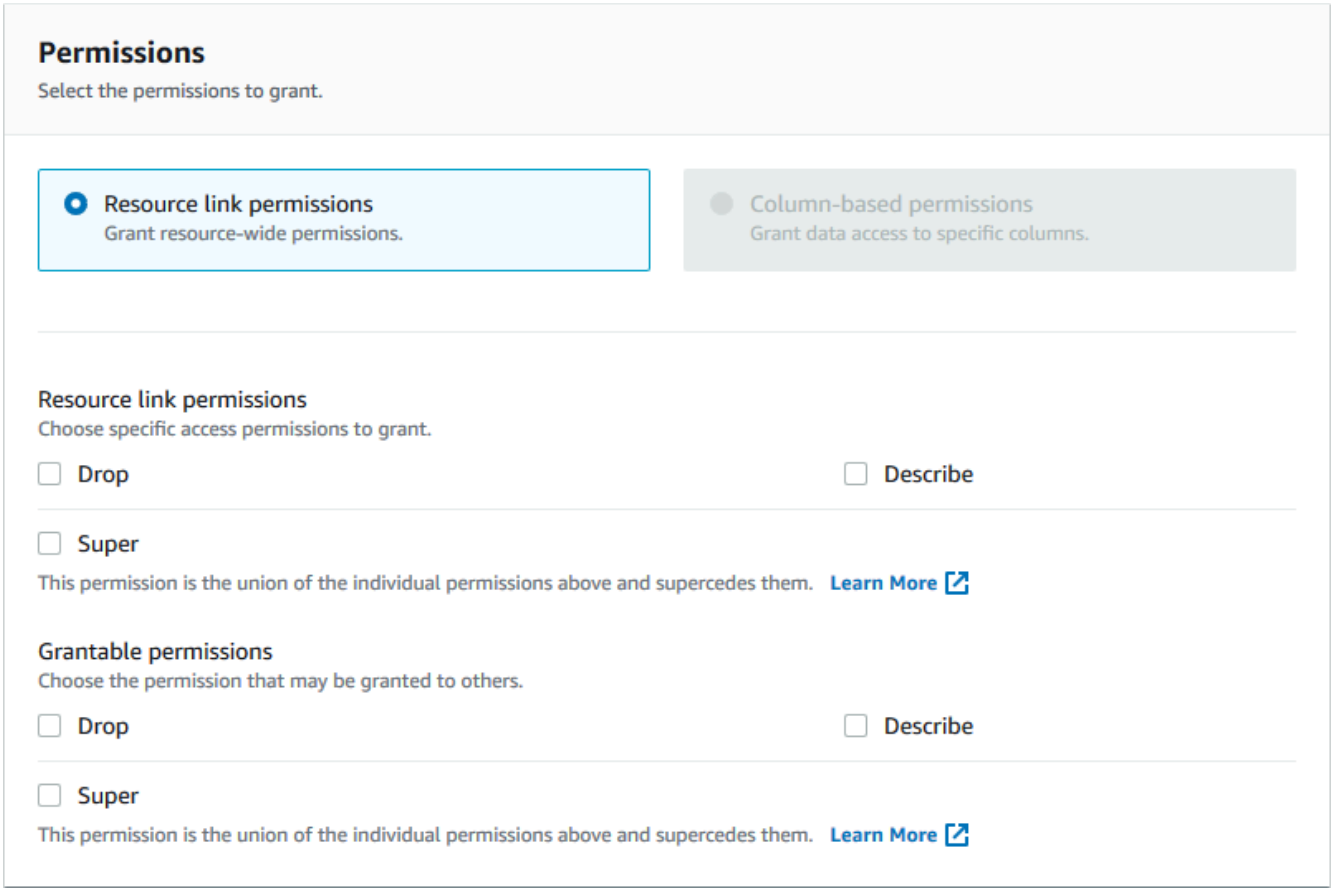
你可以使用 Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 来授予权限。

console

使用 Lake Formation 控制台授予资源链接权限

1. 请执行以下操作之一：
   - 对于数据库资源链接，请按照使用命名资源方法授予数据库权限中的步骤执行以下操作：
     1. 打开授予数据湖权限页面。
     2. 指定数据库。指定一个或多个数据库资源链接。
     3. 指定主体。
   - 对于表资源链接，请按照使用命名资源方法授予表权限中的步骤执行以下操作：
     1. 打开授予数据湖权限页面。
     2. 指定表。指定一个或多个表资源链接。
     3. 指定主体。

2. 在权限下，选择要授予的权限。（可选）选择可授予的权限。



3. 选择授权。

AWS CLI

要授予资源链接权限，请使用 AWS CLI

- 运行 `grant-permissions` 命令，指定资源链接作为资源。

  Example

  此示例`DESCRIBE`向 AWS 账户 1111-2222-3333 中的数据库`incidents-linkissues`中的表资源链接`datalake_user1`上的用户授权。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"issues",
 "Name":"incidents-link"}}'
```

另请参见：

- 创建资源链接
- Lake Formation 权限参考

访问共享表的基础数据

假设 AWS 账户 A 与账户 B 共享一个数据目录表，例如，通过向SELECT账户 B 授予表上的授予选项。要使账户 B 中的委托人能够读取共享表的基础数据，必须满足以下条件：

- 账户 B 中的数据湖管理员必须接受共享。（如果账户 A 和账户 B 在同一个组织中，或者如果使用 Lake Formation 基于标签的访问控制方法进行授权，则无需这样做。）
- 数据湖管理员必须向主体重新授予账户 A 被授予的对共享表的 Lake Formation SELECT 权限。
- 主体必须对表、包含该表的数据库和账户 A 数据目录拥有以下 IAM 权限。

Note

在以下 IAM 策略中：

- <account-id-A>替换为 AWS 账户 A 的账户 ID
- 将 *<region>* 替换为有效的区域。
- 将 *<database>* 替换为账户 A 中包含共享表的数据库的名称。
- 将 *<table>* 替换为共享表的名称。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
          "Effect": "Allow",
          "Action": [
            "glue:GetTable",
            "glue:GetTables",
            "glue:GetPartition",
            "glue:GetPartitions",
            "glue:BatchGetPartition",
            "glue:GetDatabase",
```

```
            "glue:GetDatabases"
           ],
           "Resource": [
            "arn:aws:glue:<region>:<account-id-A>:table/<database>/<table>",
            "arn:aws:glue:<region>:<account-id-A>:database/<database>",
            "arn:aws:glue:<region>:<account-id-A>:catalog"
           ]
        },
        {
          "Effect": "Allow",
          "Action": [
            "lakeformation:GetDataAccess"
           ],
          "Resource": [
            "*"
           ],
          "Condition": {
            "StringEquals": {
              "lakeformation:GlueARN":"arn:aws:glue:<region>:<account-id-
A>:table/<database>/<table>"
            }
        }
    }
   ]
}
```

另请参见：

- 接受来自 AWS RAM 的资源共享邀请

跨账户 CloudTrail 日志

Lake Formation 提供了有关对数据湖中数据的所有跨账户访问的集中审计跟踪记录。当接收者 AWS 账户访问共享表中的数据时，Lake Formation CloudTrail 会将事件复制到拥有该账户的 CloudTrail 日志中。复制的事件包括通过集成服务（例如 Amazon Athena 和 Amazon Redshift Spectrum）对数据的查询，以及按任务访问AWS Glue的数据。

CloudTrail 对数据目录资源进行跨账户操作的事件同样会被复制。

作为资源所有者，如果您在 Amazon S3 中启用对象级日志记录，则可以运行将 S3 CloudTrail 事件与 Lake Formation 事件关联的查询，以确定访问了您的 S3 存储桶的账户。 CloudTrail

主题

在跨账户 CloudTrail 日志中包含主体身份

默认情况下，添加到共享资源接收者日志并复制到资源所有者日志的跨账户 CloudTrail 事件仅包含外部账户委托人的委托人 ID，而不是 AWS 委托人（委托人 ARN）的人类可读的 Amazon 资源名称 (ARN)。在可信范围内（例如在同一个组织或团队内）共享资源时，您可以选择在活动中包括委托人 ARN。 CloudTrail 然后，资源所有者账户可以对访问他们拥有的资源的接收方账户中主体进行跟踪。

> **Important**
>
> 作为共享资源接收者，要在自己的 CloudTrail 日志中查看事件中的委托人 ARN，您必须选择与所有者账户共享委托人 ARN。
> 如果数据访问是通过资源链接进行的，则共享资源接收方账户中会记录两个事件：一个是资源链接访问事件，另一个是目标资源访问事件。资源链接访问事件确实包括主体 ARN。目标资源访问事件不包括主体 ARN（不含选择服务）。资源链接访问事件不会被复制到所有者账户。

以下是默认跨账户 CloudTrail 事件的摘录（没有选择加入）。执行数据访问的账户为 1111-2222-3333。这是同时显示在调用账户和资源所有者账户中的日志。在跨账户案例中，Lake Formation 会在两个账户中填充日志。

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
```

```
        "requesterService": "GLUE_JOB",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

作为共享资源使用者，当您选择包括主体 ARN 时，摘录将变为以下内容。`lakeFormationPrincipal` 字段表示通过 Amazon Athena、Amazon Redshift Spectrum 或 AWS Glue 作业执行查询的最终角色或用户。

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

选择在跨 CloudTrail 账户日志中包含委托人 ARN

1. 通过 https://console.aws.amazon.com/lakeformation/ 打开 Lake Formation 控制台。

   以 `Administrator` 用户或具有 `Administrator Access` IAM 策略的用户身份登录。
2. 在导航窗格中，选择 Settings（设置）。
3. 在数据目录设置页面的默认权限 AWS CloudTrail部分中，为资源所有者输入一个或多个 AWS 资源所有者账户 ID。

   在输入每个账户 ID 之后按 Enter 键。
4. 选择保存。

现在，存储在共享资源接收者和资源所有者的日志中的跨账户 CloudTrail 事件包含委托人 ARN。

查询 Amazon S3 跨账户访问 CloudTrail 日志

作为共享资源所有者，您可以查询 S3 CloudTrail 日志以确定访问过您的 Amazon S3 存储桶的账户（前提是您在 Amazon S3 中启用了对象级日志记录）。这仅适用于您在 Lake Formation 中注册的 S3 位置。如果共享资源使用者选择在 Lake Formation CloudTrail 日志中包含委托人 Ran，则可以确定访问存储桶的角色或用户。

使用运行查询时 Amazon Athena，您可以通过会话名称属性加入 Lake Formation CloudTrail CloudTrail 事件和 S3 事件。查询还可以筛选关于 `eventName="GetDataAccess"` 的 Lake Formation 事件以及关于 `eventName="Get Object"` 或 `eventName="Put Object"` 的 S3 事件。

以下是 Lake Formation 跨账户 CloudTrail 事件的摘录，在该事件中，已注册 S3 位置的数据被访问。

```
{
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GetDataAccess",
  ..............
  ..............
  "additionalEventData": {
    "requesterService": "GLUE_JOB",
    "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
    "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-B8JSAjo5QA"
   }
}
```

`lakeFormationRoleSessionName`密钥值可以与 S3 CloudTrail 事件`principalId`密钥中的会话名称合并。`AWSLF-00-GL-111122223333-B8JSAjo5QA`以下是 S3 CloudTrail 活动的摘录。它显示会话名称的位置。

```
{
   "eventSource": "s3.amazonaws.com",
   "eventName": "Get Object"
   ..............
   ..............
   "principalId": "AROAQSOX5XXUR7D6RMYLR:AWSLF-00-GL-111122223333-B8JSAjo5QA",
```

```
   "arn": "arn:aws:sets::111122223333:assumed-role/Deformationally/AWSLF-00-
GL-111122223333-B8JSAjo5QA",
   "session Context": {
     "session Issuer": {
       "type": "Role",
       "principalId": "AROAQSOX5XXUR7D6RMYLR",
       "arn": "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/Deformationally",
       "accountId": "111122223333",
       "user Name": "Deformationally"
     },
   ..............
   ..............
}
```

会话名称的格式如下：

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

version-number

此格式的版本，当前为 00。如果会话名称格式发生变化，则下一个版本将是 01。

query-engine-code

表示访问数据的实体。当前值如下：

GL	AWS Glue ETL 作业
AT	Athena
RE	Amazon Redshift Spectrum

account-id

向 Lake Formation 请求凭证的 AWS 账户 ID。

suffix

随机生成的字符串。

使用 AWS Glue 和 Lake Formation 管理跨账户权限。

可以使用 AWS Glue 或 AWS Lake Formation授予对数据目录资源和基础数据的跨账户访问权限。

在中AWS Glue，您可以通过创建或更新数据目录资源策略来授予跨账户权限。在 Lake Formation 中，您可以使用 Lake Formation GRANT/REVOKE 权限模型和 `Grant Permissions` API 操作来授予跨账户权限。

> **Tip**
>
> 我们建议仅依靠 Lake Formation 权限来保护您的数据湖。

你可以使用 Lake Formation 控制台或 AWS Resource Access Manager (AWS RAM) 控制台查看 Lake Formation 跨账户授权。但是，这些控制台页面不显示通过 AWS Glue 数据目录资源策略授予的跨账户权限。同样，您可以使用 AWS Glue 控制台的设置页面在数据目录资源策略中查看跨账户授权，但该页面不显示使用 Lake Formation 授予的跨账户权限。

为了确保您在查看和管理跨账户权限时不会错过任何授权，Lake Formation 和 AWS Glue 会要求您执行以下操作，以表明您知道并将允许 Lake Formation 和 AWS Glue 进行跨账户授权。

使用 AWS Glue 数据目录资源策略授予跨账户权限时

如果您的账户（授权人账户或创建者账户）没有进行用于 AWS RAM 共享资源的跨账户授权，则可以照常在中保存数据目录资源策略。AWS Glue但是，如果已经进行了涉及 AWS RAM 资源共享的授权，则必须执行以下操作之一，以确保成功保存资源策略：

- 当您在 AWS Glue 控制台的设置 页面上保存资源策略时，控制台会发出警报，指出除了使用 Lake Formation 控制台授予的任何权限之外，还将授予该策略中的权限。必须选择继续才能保存该策略。
- 使用 `glue:PutResourcePolicy` API 操作保存资源策略时，必须将 `EnableHybrid` 字段设置为“TRUE”（类型 = 字符串）。以下代码示例演示如何在 Python 中执行此操作。

```
import boto3
import json

REGION = 'us-east-2'
PRODUCER_ACCOUNT_ID = '123456789012'
CONSUMER_ACCOUNT_IDs = ['111122223333']

glue = glue_client = boto3.client('glue')
```

```
policy = {
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Cataloguers",
            "Effect": "Allow",
            "Action": [
                "glue:*"
            ],
            "Principal": {
                "AWS": CONSUMER_ACCOUNT_IDs
            },
            "Resource": [
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:catalog",
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:database/*",
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:table/*/*"
            ]
        }
    ]
}

policy = json.dumps(policy)
glue.put_resource_policy(PolicyInJson=policy, EnableHybrid='TRUE')
```

有关更多信息，请参阅PutResourcePolicy 《开发者指南》中的操作（Python：put_resource_policy）。AWS Glue

使用 Lake Formation 命名资源方法授予跨账户权限时

如果您的账户中没有数据目录资源策略，您进行的 Lake Formation 跨账户授权将照常进行。但是，如果存在数据目录资源策略，则必须在其中添加以下语句，以确保使用命名资源方法进行的跨账户授权成功完成。<region>替换为有效的地区<account-id>名称和您的 AWS 账户 ID。

```
  {
    "Effect": "Allow",
    "Action": [
      "glue:ShareResource"
    ],
    "Principal": {"Service": [
      "ram.amazonaws.com"
    ]},
    "Resource": [
```

```
        "arn:aws:glue:<region>:<account-id>:table/*/*",
        "arn:aws:glue:<region>:<account-id>:database/*",
        "arn:aws:glue:<region>:<account-id>:catalog"
      ]
    }
```

如果没有这份额外的声明，Lake Formation 授权就会成功 AWS RAM，但会被封锁，接收者账户将无法访问授予的资源。

> **Important**
>
> 使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法进行跨账户授权时，您的数据目录资源策略必须至少含有先决条件中指定的权限。

> **另请参见：**
>
> - 元数据访问控制（用于讨论命名资源方法与 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法）。
> - 查看共享数据目录表和数据库
> - 《AWS Glue 开发人员指南》中的在 AWS Glue 控制台中使用数据目录设置
> - 《AWS Glue 开发人员指南》中的授予跨账户访问权限（适用于数据目录资源策略示例）

使用 GetResourceShares API 操作查看所有跨账户授权

如果您的企业同时使用 AWS Glue Data Catalog 资源策略和 Lake Formation 授权来授予跨账户权限，那么在一个地方查看所有跨账户授权的唯一方法就是使用 `glue:GetResourceShares` API 操作。

当您使用指定资源方法向账户授予 Lake Formation 权限时， AWS Resource Access Manager (AWS RAM) 会创建一个 AWS Identity and Access Management (IAM) 资源策略并将其存储在您的 AWS 账户中。该策略授予访问资源所需的权限。 AWS RAM 为每项跨账户授予创建单独的资源策略。您可以使用 `glue:GetResourceShares` API 操作查看所有这些策略。

> **Note**
>
> 此操作还会返回数据目录资源策略。但是，如果您在数据目录设置中启用了元数据加密，并且您没有 AWS KMS 密钥权限，则该操作将不会返回数据目录资源策略。

查看所有跨账户授权

- 输入以下 AWS CLI 命令。

```
aws glue get-resource-policies
```

以下是一个资源策略示例，当您向 AWS 账户 1111-2222-3333 授予数据库t中表的权限时，db1该策略 AWS RAM 会创建和存储。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
         "glue:GetTable",
         "glue:GetTables",
         "glue:GetTableVersion",
         "glue:GetTableVersions",
         "glue:GetPartition",
         "glue:GetPartitions",
         "glue:BatchGetPartition",
         "glue:SearchTables"
       ],
      "Principal": {"AWS": [
        "111122223333"
      ]},
      "Resource": [
        "arn:aws:glue:<region>:111122223333:table/db1/t"
     ]
    }
  ]
}
```

另请参阅：

- GetResourceShares 开发者指南中的@@ 操作（Python：get_resource_policies）AWS Glue

访问和查看共享数据目录表和数据库

对于数据湖管理员和已被授予权限的主体来说，与您的 AWS 账户共享的资源将显示在数据目录中，就像它们是您账户中的资源一样。控制台显示拥有该资源的账户。

您可以使用 Lake Formation 控制台查看与您的账户共享的资源。您还可以使用 AWS Resource Access Manager (AWS RAM) 控制台查看与您的账户共享的资源以及使用命名资源方法与其他 AWS 账户共享的资源。

Important

当有人使用命名资源方法向您的账户或 AWS 组织授予对数据目录资源的跨账户权限时，Lake Formation 会使用 AWS Resource Access Manager (AWS RAM) 服务共享该资源。如果您的账户与授予账户位于同一 AWS 组织中，则共享资源将立即可供您使用。
但是，如果您的账户不在同一组织中，则 AWS RAM 会向您的账户发送邀请，以便您接受或拒绝资源共享。然后，要使共享资源可用，您账户中的数据湖管理员必须使用 AWS RAM 控制台或 CLI 接受该邀请。
如果有 AWS RAM 资源共享邀请等待接受，Lake Formation 控制台会显示提醒。只有有权查看 AWS RAM 邀请的用户才会收到提醒。

另请参见：

- 跨 AWS 账户共享数据目录表和数据库
- Lake Formation 中的跨账户数据共享
- 访问共享表的基础数据
- 元数据访问控制（有关使用命名资源方法与 LF-TBAC 方法共享资源的信息。）

主题

接受来自 AWS RAM 的资源共享邀请

如果与您的 AWS 账户共享数据目录资源，并且您的账户与共享账户不在同一 AWS 组织中，则在接受来自 AWS Resource Access Manager (AWS RAM) 的资源共享邀请之前，您无法访问该共享资源。作为数据湖管理员，您必须先查询 AWS RAM 以查找待处理的邀请，然后接受邀请。

您可以使用 AWS RAM 控制台、API 或 AWS Command Line Interface (AWS CLI) 来查看和接受邀请。

查看和接受来自 AWS RAM 的资源共享邀请（控制台）

1. 确保您具有查看和接受资源共享邀请所需的 AWS Identity and Access Management (IAM) 权限。

   有关为数据湖管理员建议的 IAM 策略的信息，请参阅the section called “数据湖管理员权限”。
2. 按照《AWS RAM 用户指南》中的接受和拒绝邀请部分中的说明进行操作。

查看和接受来自 AWS RAM 的资源共享邀请 (AWS CLI)

1. 确保您具有查看和接受资源共享邀请所需的 AWS Identity and Access Management (IAM) 权限。

   有关为数据湖管理员建议的 IAM 策略的信息，请参阅the section called “数据湖管理员权限”。
2. 输入以下命令以查看待处理资源共享邀请。

```
aws ram get-resource-share-invitations
```

该输出值应该类似于以下内容。

```
{
    "resourceShareInvitations": [
        {
            "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
            "resourceShareName": "111122223333-123456789012-uswuU",
            "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
```

```
            "senderAccountId": "111122223333",
            "receiverAccountId": "123456789012",
            "invitationTimestamp": 1589576601.79,
            "status": "PENDING"
        }
    ]
}
```

请注意 PENDING 的状态。

3. 将 resourceShareInvitationArn 键的值复制到剪贴板。
4. 将该值粘贴到以下命令中，替换 *<invitation-arn>*，然后输入该命令。

```
aws ram accept-resource-share-invitation --resource-share-invitation-
arn <invitation-arn>
```

该输出值应该类似于以下内容。

```
{
    "resourceShareInvitations": [
        {
            "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
            "resourceShareName": "111122223333-123456789012-uswuU",
            "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
            "senderAccountId": "111122223333",
            "receiverAccountId": "123456789012",
            "invitationTimestamp": 1589576601.79,
            "status": "ACCEPTED"
        }
    ]
}
```

请注意 ACCEPTED 的状态。

查看共享数据目录表和数据库

您可以使用 Lake Formation 控制台或 AWS CLI 查看与您的账户共享的资源。您还可以使用 AWS Resource Access Manager (AWS RAM) 控制台或 CLI 查看与您的账户共享的资源以及您已与其他 AWS 账户共享的资源。

使用 Lake Formation 控制台查看共享资源

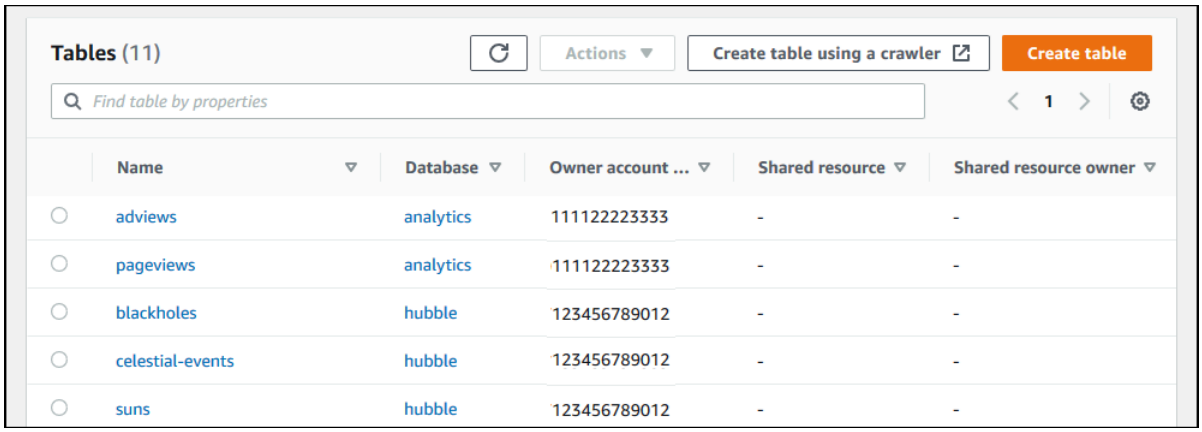
1. 打开 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。

   以数据湖管理员或已被授予对共享表的权限的用户身份登录。

2. 要查看与您的 AWS 账户共享的资源，请执行下列操作之一：

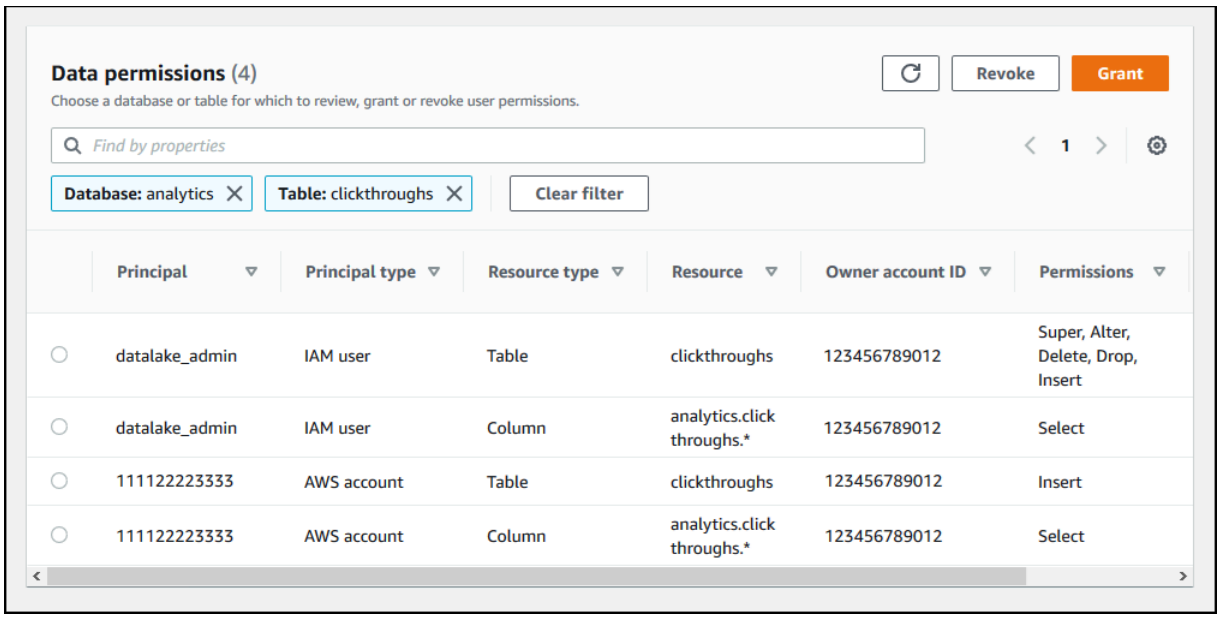
   - 要查看与您的账户共享的表，请在导航窗格中选择表。
   - 要查看与您的账户共享的数据库，请在导航窗格中选择数据库。

   控制台会显示您账户中以及与您的账户共享的数据库或表的列表。对于与您的账户共享的资源，控制台会在拥有者账户 ID 列（以下屏幕截图中的第三列）下显示该拥有者的 AWS 账户 ID。



3. 要查看您与其他 AWS 账户或组织共享的资源，请在导航窗格中选择数据权限。

   您共享的资源列在数据权限页面上，其中外部账号显示在主体列中，如下图所示。



使用 AWS RAM 控制台查看共享资源

1. 确保您具有使用 AWS RAM 查看共享资源所需的 AWS Identity and Access Management (IAM) 权限。

   您必须至少具有 `ram:ListResources` 权限。此权限包含在 AWS 托管式策略 `AWSLakeFormationCrossAccountManager` 中。

2. 登录到 AWS Management Console，然后通过以下网址打开 AWS RAM 控制台：https://console.aws.amazon.com/ram。
3. 请执行下列操作之一：

   - 要查看您共享的资源，请在导航窗格的由我共享下，选择共享资源。
   - 要查看与您共享的资源，请在导航窗格的与我共享下，选择共享资源。

创建资源链接

资源链接是数据目录对象，它们是指向元数据的数据库和表的链接，通常指向其他 AWS 账户中的共享数据库和表。它们有助于在所有 AWS 区域跨账户访问数据湖中的数据。

> **Note**
>
> Lake Formation 支持跨 AWS 区域查询数据目录表。您可以从任何 AWS 区域访问数据目录数据库和表，方法是在这些区域中创建指向不同区域中共享数据库和表的资源链接。

主题

资源链接在 Lake Formation 中的工作原理

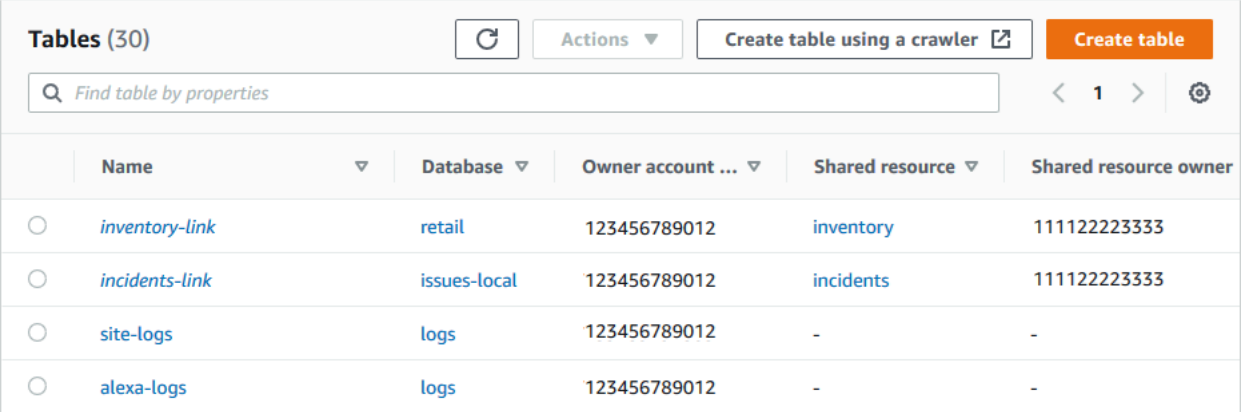
资源链接是一个数据目录对象，它是指向本地或共享数据库或表的链接。创建指向数据库或表的资源链接后，您可以在需要使用数据库或表名称的任何位置使用资源链接名称。表资源链接将与您拥有的或与您共享的表一起由 `glue:GetTables()` 返回，并在 Lake Formation 控制台的表页面上作为条目显示。数据库的资源链接的作用方式相似。

创建指向数据库或表的资源链接使您可以执行以下操作：

- 为数据目录中的数据库或表指定不同的名称。如果不同的 AWS 账户共享同名的数据库或表，或者如果您的账户中的多个数据库含有同名的表，则此功能特别有用。
- 通过在任何 AWS 区域中创建指向其他区域中数据库和表的资源链接，在这些区域中访问数据目录数据库和表。您可以使用 Athena、Amazon EMR 在任何区域通过这些资源链接运行查询，并运行 AWS Glue ETL Spark 作业，而无需复制源数据或 Glue Data Catalog 中的元数据。
- 使用 Amazon Athena 和 Amazon Redshift Spectrum 等集成 AWS 服务来运行查询，以访问共享数据库或表。Amazon Athena 等集成服务无法直接跨账户访问数据库或表。但是，他们可以访问您账户中指向其他账户中数据库和表的资源链接。

> **Note**
>
> 无需创建资源链接即可在 AWS Glue 提取、转换、加载 (ETL) 脚本中引用共享数据库或表。但是，为了避免在多个 AWS 账户共享同名的数据库或表时出现歧义，您可以创建并使用资源链接，或者在调用 ETL 操作时指定目录 ID。

以下示例显示了 Lake Formation 控制台的表页面，其中列出了两个资源链接。资源链接名称始终以斜体显示。每个资源链接会与它链接的共享资源的名称和所有者一起显示。在此示例中，AWS 账户 1111-2222-3333 中的数据湖管理员与账户 1234-5678-9012 共享了表 `inventory` 和 `incidents`。然后，该账户中的用户创建了指向这些共享表的资源链接。



以下是有关资源链接的注意事项和限制：

- 需要资源链接才能使 Athena 和 Redshift Spectrum 等集成服务能够查询共享表的基础数据。这些集成服务中的查询是针对资源链接名称构建的。
- 假定所含数据库的仅对此数据库中的新表使用 IAM 访问控制设置已关闭，则只有创建资源链接的主体才能查看和访问该资源链接。要使您账户中的其他主体能够访问资源链接，请向其授予对它的 DESCRIBE 权限。要使其他人能够删除资源链接，请向其授予对它的 DROP 权限。数据湖管理员可以访问账户中的所有资源链接。要删除其他主体创建的资源链接，数据湖管理员必须先向自己授予对资源链接的 DROP 权限。有关更多信息，请参阅Lake Formation 权限参考。

  > **Important**
  >
  > 授予对资源链接的权限不会授予对目标（链接）数据库或表的权限。您必须单独授予对目标的权限。

- 要创建资源链接，您需要拥有 Lake Formation CREATE_TABLE 或 CREATE_DATABASE 权限，以及 `glue:CreateTable` 或 `glue:CreateDatabase` AWS Identity and Access Management (IAM) 权限。
- 您可以创建指向本地（自有）数据目录资源以及与您的 AWS 账户共享的资源的资源链接。
- 创建资源链接时，不会检查目标共享资源是否存在或您是否拥有对该资源的跨账户权限。这使您能够按任意顺序创建资源链接和共享资源。
- 如果删除资源链接，则不会删除链接的共享资源。如果您删除共享资源，则指向该资源的资源链接不会被删除。

- 您可以创建资源链接链。但是，这样做没有任何价值，因为 API 只访问第一个资源链接。

> 另请参见：
>
> - 授予和撤销对数据目录资源的权限

创建指向共享数据目录表的资源链接

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 创建指向任何 AWS 区域中共享表的资源链接。

创建指向共享表的资源链接（控制台）

1. 访问 https://console.aws.amazon.com/lakeformation/ 并打开 AWS Lake Formation 控制台。以对要包含资源链接的数据库拥有 Lake Formation CREATE_TABLE 权限的主体身份登录。
2. 在导航窗格中，选择 Tables (表)，然后选择 Create table (创建表)。
3. 在创建表页面上，选择资源链接磁贴，然后提供以下信息：

   资源链接名称

   输入一个与表名遵循相同规则的名称。该名称可以与目标共享表的名称相同。

   数据库

   本地数据目录中要包含资源链接的数据库。

   共享表所有者区域

   如果您要在其他区域创建资源链接，请选择目标共享表的区域。

   共享表

   从列表中选择共享表，或者输入本地（自有）或共享表的名称。

   列表中列出了与您的账户共享的所有表。记下与每个表一起列出的数据库和所有者账户 ID。如果您未看到您知道与您的账户共享的表，请检查以下内容：

   - 如果您不是数据湖管理员，请检查数据湖管理员是否向您授予了对该表的 Lake Formation 权限。

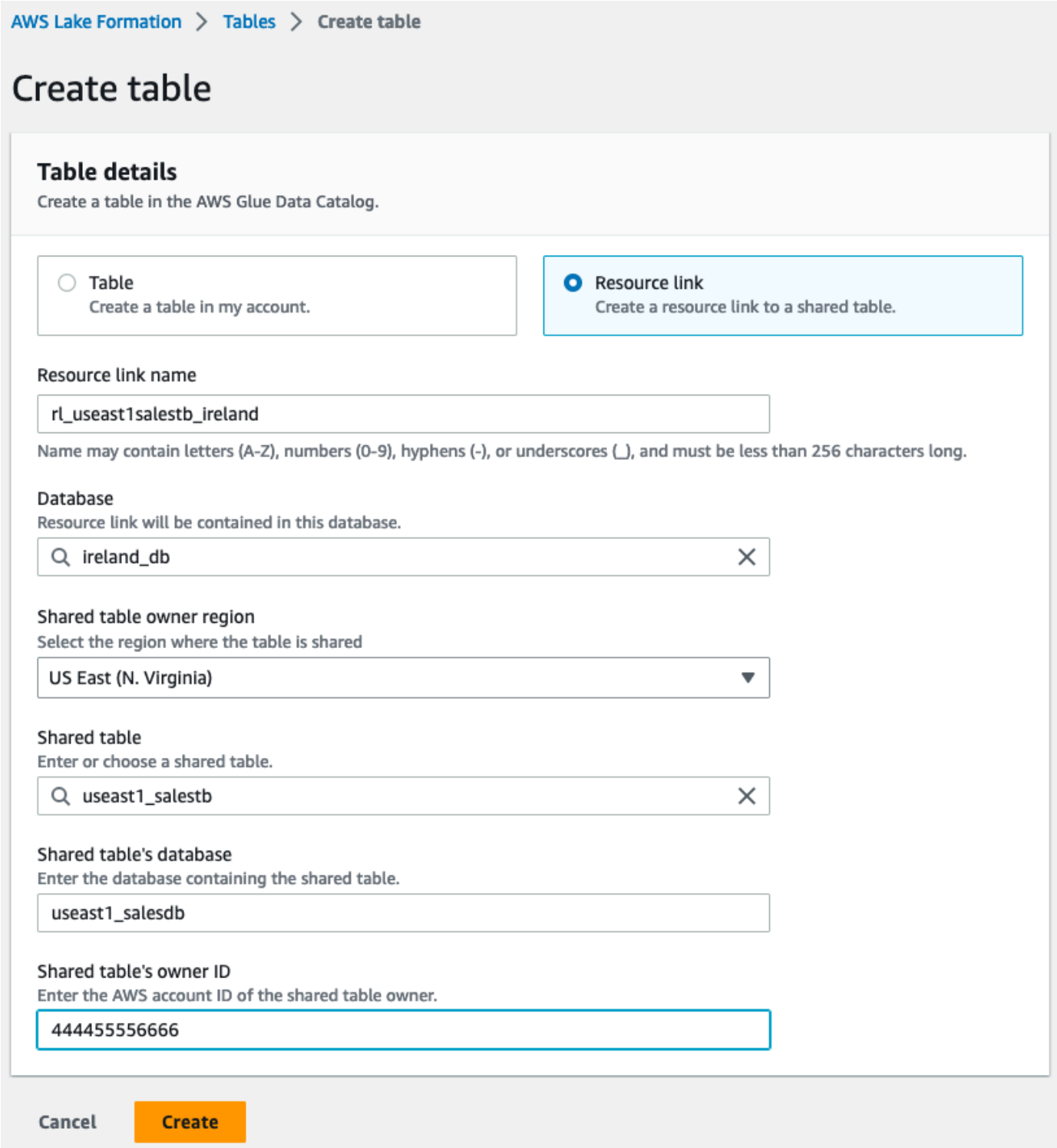
- 如果您是数据湖管理员，并且您的账户与授权账户不在同一个 AWS 组织中，请确保您已接受关于该表的 AWS Resource Access Manager (AWS RAM) 资源共享邀请。有关更多信息，请参阅接受来自 AWS RAM 的资源共享邀请。

共享表的数据库

如果您从列表中选择了共享表，则此字段中将填充外部账户中共享表的数据库。否则，请输入本地数据库（以获取指向本地表的资源链接）或外部账户中共享表的数据库。

共享表的所有者

如果您从列表中选择了共享表，则此字段中将填充该共享表的所有者账户 ID。否则，请输入您的 AWS 账户 ID（以获取指向本地表的资源链接）或共享该表的 AWS 账户的 ID。



4. 选择创建以创建资源链接。

   然后，您可以在表页面的名称列下查看资源链接名称。

5. （可选）向必须能够查看链接和访问目标表的主体授予对资源链接的 Lake Formation DESCRIBE 权限。

创建指向同一区域 (AWS CLI) 中共享表的资源链接

1. 输入类似以下的命令。

```
aws glue create-table --database-name myissues --table-input
 '{"Name":"my_customers","TargetTable":
{"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
```

此命令会创建一个指向共享 customers 的名为 my_customers 的资源链接，该共享表位于 AWS 账户 1111-2222-3333 中的数据库 issues 中。资源链接存储在本地数据库 myissues 中。

2. （可选）向必须能够查看链接和访问目标表的主体授予对资源链接的 Lake Formation DESCRIBE 权限。

创建指向不同区域 (AWS CLI) 中共享表的资源链接

1. 输入类似以下的命令。

```
aws glue create-table –region eu-west-1 –cli-input-json '{
    "CatalogId": "111122223333",
    "DatabaseName": "ireland_db",
    "TableInput": {
        "Name": "rl_useast1salestb_ireland",
        "TargetTable": {
            "CatalogId": "444455556666",
            "DatabaseName": "useast1_salesdb",
            "Region": "us-east-1",
            "Name":"useast1_salestb"
        }
    }
}‘
```

此命令在欧洲地区（爱尔兰）区域创建一个指向共享表 useast1_salestb 的名为 rl_useast1salestb_ireland 的资源链接，该共享表位于美国东部（弗吉尼亚州北部）区域的 AWS 账户 444455556666 中的数据库 useast1_salesdb 中。资源链接存储在本地数据库 ireland_db 中。

2. 向必须能够查看链接并通过链接访问链接目标的主体授予 Lake Formation DESCRIBE 权限。

> 另请参见：
>
> - 资源链接在 Lake Formation 中的工作原理
> - DESCRIBE

创建指向共享数据目录数据库的资源链接

您可以使用 AWS Lake Formation 控制台、API 或 AWS Command Line Interface (AWS CLI) 创建指向共享数据库的资源链接。

创建指向共享数据库的资源链接（控制台）

1. 访问 https://console.aws.amazon.com/lakeformation/ 并打开 AWS Lake Formation 控制台。以数据湖管理员或数据库创建者的身份登录。

   数据库创建者是已获得 Lake Formation CREATE_DATABASE 权限的主体。
2. 在导航窗格中，选择数据库，然后选择创建数据库。
3. 在创建数据库页面上，选择资源链接磁贴，然后提供以下信息：

   资源链接名称

   输入一个与数据库名称遵循相同规则的名称。该名称可以与目标共享数据库的名称相同。

   共享数据库所有者区域

   如果您要在其他区域创建资源链接，请选择目标共享数据库的区域。

   共享数据库

   从列表中选择数据库，或输入本地（自有）或共享数据库名称。

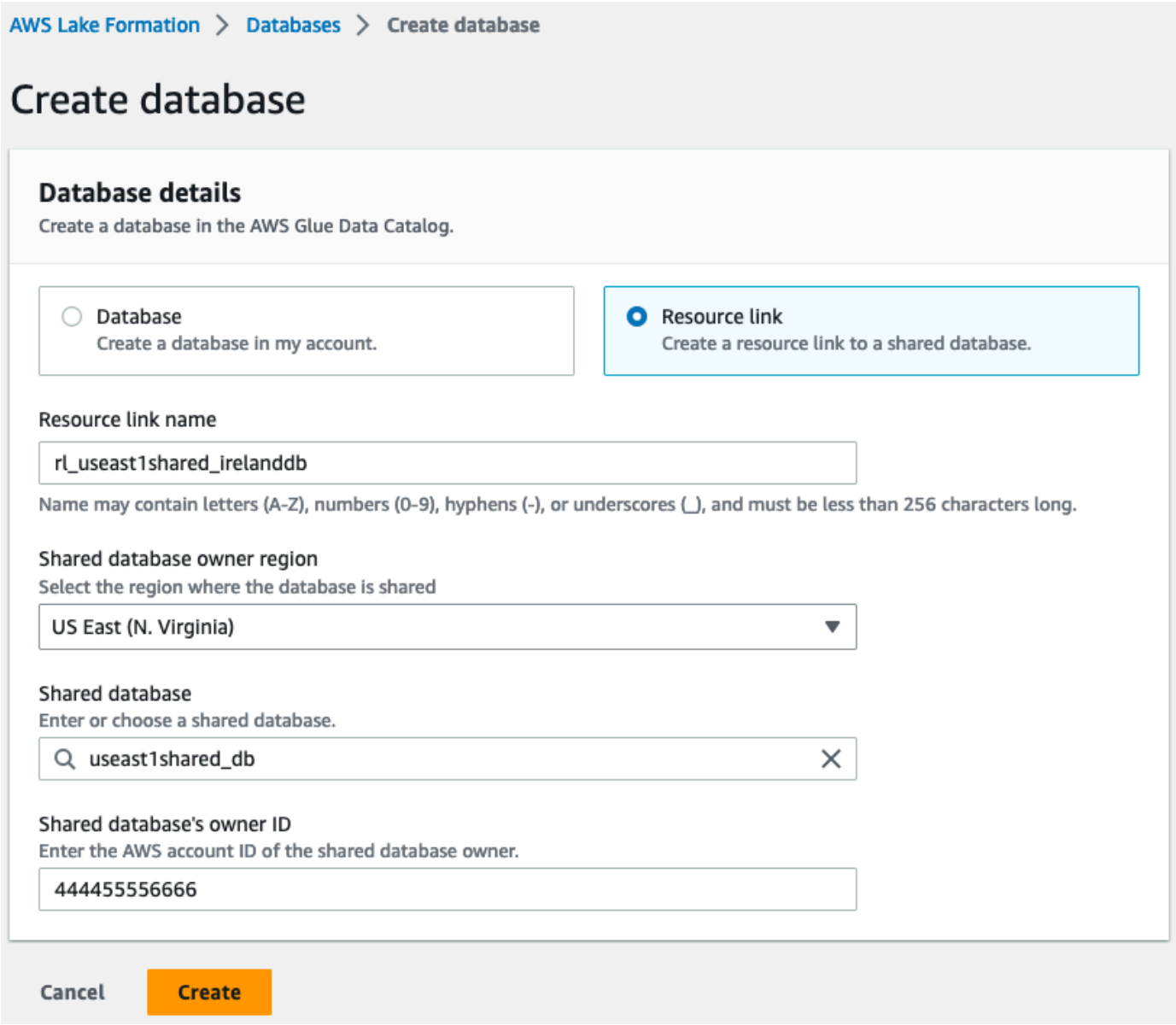
   列表中包含与您的账户共享的所有数据库。记下与每个数据库一起列出的所有者账户 ID。如果您未看到您知道与您的账户共享的数据库，请检查以下内容：

   - 如果您不是数据湖管理员，请检查数据湖管理员是否向您授予了对该数据库的 Lake Formation 权限。

- 如果您是数据湖管理员，并且您的账户与授权账户不在同一个 AWS 组织中，请确保您已接受关于该数据库的 AWS Resource Access Manager (AWS RAM) 资源共享邀请。有关更多信息，请参阅接受来自 AWS RAM 的资源共享邀请。

共享数据库所有者

如果您从列表中选择了共享数据库，则此字段中将填充该共享数据库的所有者账户 ID。否则，请输入您的 AWS 账户 ID（以获取指向本地数据库的资源链接）或共享该数据库的 AWS 账户的 ID。



4. 选择创建以创建资源链接。

然后，您可以在数据库页面的名称列下查看资源链接名称。

5. （可选）向来自欧洲地区（爱尔兰）区域的主体授予对资源链接的 Lake Formation DESCRIBE 权限，这些主体必须能够查看该链接并访问目标数据库。

创建指向同一区域中共享数据库的资源链接 (AWS CLI)

1. 输入类似以下的命令。

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

此命令创建一个指向 AWS 账户 1111-2222-3333 中的共享数据库 issues 的名为 myissues 的资源链接。

2. （可选）向主体授予对资源链接的 Lake Formation DESCRIBE 权限，这些主体必须能够查看该链接并访问目标数据库。

创建指向不同区域中共享数据库的资源链接 (AWS CLI)

1. 输入类似以下的命令。

```
aws glue create-database --region eu-west-1 --cli-input-json '{
    "CatalogId": "111122223333",
    "DatabaseInput": {
      "Name": "rl_useast1shared_irelanddb",
      "TargetDatabase": {
          "CatalogId": "444455556666",
          "DatabaseName": "useast1shared_db",
          "Region": "us-east-1"
       }
    }
}'
```

此命令在欧洲地区（爱尔兰）区域的 AWS 账户 111122223333 中创建指向共享数据库 useast1shared_db 的名为 rl_useast1shared_irelanddb 的资源链接，该数据库位于美国东部（弗吉尼亚州北部）区域的 AWS 账户 444455556666 中。

2. 向来自欧洲地区（爱尔兰）区域的主体授予 Lake Formation DESCRIBE 权限，这些主体必须能够查看该链接并通过该链接访问链接目标。

另请参见：

- 资源链接在 Lake Formation 中的工作原理
- DESCRIBE

AWS Glue API 中的资源链接处理

下表说明了 AWS Glue 数据目录 API 如何处理数据库和表资源链接。对于所有 Get* API 操作，仅返回调用者对其拥有权限的数据库和表。此外，通过资源链接访问目标数据库或表时，您必须同时拥有对目标和资源链接的 AWS Identity and Access Management (IAM) 和 Lake Formation 权限。您需要对资源链接拥有的 Lake Formation 权限是 DESCRIBE。有关更多信息，请参阅DESCRIBE。

数据库 API 操作

API 操作	资源链接处理
CreateDatabase	如果数据库是资源链接，则创建指向指定目标数据库的资源链接。
UpdateDatabase	如果指定的数据库是资源链接，则访问该链接并更新目标数据库。如果必须修改资源链接以链接到其他数据库，则必须将其删除，然后再创建一个新的资源链接。
DeleteDatabase	删除资源链接。这不会删除所链接的（目标）数据库。
GetDatabase	如果调用者对目标拥有权限，请单击该链接返回目标的属性。否则，它将返回链接的属性。
GetDatabases	返回数据库列表，包括资源链接。对于结果集中的每个资源链接，操作都会访问该链接以获取链接目标的属性。必须指定 ResourceShareType = ALL 才能查看与您的账户共享的数据库。

表 API 操作

API 操作	资源链接处理
CreateTable	如果数据库是资源链接，则访问该数据库链接并在目标数据库中创建表。如果表是资源链接，则该操作将在指定的数据库中创建资源链接。不支持通过数据库资源链接创建表资源链接。
UpdateTable	如果表或指定数据库是资源链接，则更新目标表。如果表和数据库都是资源链接，则操作将失败。
DeleteTable	如果指定的数据库是资源链接，则访问该链接并删除目标数据库中的表或表资源链接。如果表是资源链接，则该操作会删除指定数据库中的表资源链接。删除表资源链接不会删除目标表。
BatchDeleteTable	与 DeleteTable 相同。
GetTable	如果指定的数据库是资源链接，则访问该数据库链接并返回目标数据库中的表或表资源链接。否则，如果表是资源链接，则该操作访问该链接并返回目标表的属性。
GetTables	如果指定的数据库是资源链接，则访问该数据库链接并返回目标数据库中的表和表资源链接。如果目标数据库是来自另一个 AWS 账户的共享数据库，则该操作仅返回该数据库中的共享表。它不访问目标数据库中的表资源链接。否则，如果指定的数据库是本地（自有）数据库，则该操作将返回本地数据库中的所有表，并访问每个表资源链接以返回目标表的属性。
SearchTables	返回表和表资源链接。这不会访问链接以返回目标表的属性。必须指定 ResourceShareType = ALL 才能查看与您的账户共享的表。
GetTableVersion	与 GetTable 相同。
GetTableVersions	与 GetTable 相同。
DeleteTableVersion	与 DeleteTable 相同。
BatchDeleteTableVersion	与 DeleteTable 相同。

分区 API 操作

API 操作	资源链接处理
CreatePartition	如果指定的数据库是资源链接，则访问该数据库链接并在目标数据库的指定表中创建一个分区。如果表是资源链接，则该操作访问该资源链接并在目标表中创建分区。不支持通过表资源链接和数据库资源链接创建分区。
BatchCreatePartition	与 CreatePartition 相同。
UpdatePartition	如果指定的数据库是资源链接，则访问该数据库链接并更新目标数据库中指定表中的分区。如果表是资源链接，则该操作访问该资源链接并更新目标表中的分区。不支持通过表资源链接和数据库资源链接创建分区。
DeletePartition	如果指定的数据库是资源链接，则访问该数据库链接并删除目标数据库中指定表中的分区。如果表是资源链接，则该操作访问该资源链接并删除目标表中的分区。不支持通过表资源链接和数据库资源链接创建分区。
BatchDeletePartition	与 DeletePartition 相同。
GetPartition	如果指定的数据库是资源链接，则访问该数据库链接并返回指定表中的分区信息。否则，如果表是资源链接，则该操作访问该链接并返回分区信息。如果表和数据库都是资源链接，则返回一个空结果集。
GetPartitions	如果指定的数据库是资源链接，则访问该数据库链接并返回指定表中所有分区的分区信息。否则，如果表是资源链接，则该操作访问该链接并返回分区信息。如果表和数据库都是资源链接，则返回一个空结果集。
BatchGetPartition	与 GetPartition 相同。

用户定义的函数 API 操作

API 操作	资源链接处理
（所有 API 操作）	如果数据库是资源链接，则访问该资源链接并针对目标数据库执行操作。

另请参见：

- 资源链接在 Lake Formation 中的工作原理

跨区域访问表

Lake Formation 支持跨 AWS 区域查询数据目录表。您可以使用 Amazon Athena、Amazon EMR AWS Glue 和 ETL 从其他区域访问某个区域的数据，方法是在其他区域创建指向源数据库和表的资源链接。通过跨区域表访问功能，您可以跨区域访问数据，而无需将基础数据或元数据复制到数据目录中。

例如，您可以将制作者账户中的数据库或表共享给区域 A 中的使用者账户。接受区域 A 中的资源共享邀请后，使用者账户的数据湖管理员可以创建指向区域 B 中共享资源的资源链接。使用者账户管理员可以向区域 A 中该账户中的 IAM 主体授予对共享资源的权限，并可以授予区域 B 中的资源链接权限。使用资源链接，使用者账户中的主体可以查询区域 B 中的共享数据。

您还可以在制作者账户中托管 A 区域的 Amazon S3 数据来源，并在区域 B 的中央账户中注册数据位置。您可以在中央账户中创建数据目录资源，设置 Lake Formation 权限，并与您账户中的使用者或区域 B 中的外部账户共享数据。跨区域功能让用户可以使用资源链接从区域 C 访问这些数据目录表。

使用此功能，您可以跨区域查询 Apache Hive 元存储中的联合数据库，还可以在运行查询时将本地区域中的表与其他区域的表联接起来。

Lake Formation 通过跨区域表访问功能支持以下功能：

- 基于 LF 标签的访问控制
- 细粒度访问控制权限
- 使用适当的权限对共享数据库或表执行写入操作
- 在账户级别以及直接与 IAM 主体跨账户共享数据

具有 `Create_Database` 和 `Create_Table` 权限的非管理用户可以创建跨区域资源链接。

> **Note**
>
> 您可以在任何区域创建跨区域资源链接并访问数据，而无需应用 Lake Formation 权限。对于未向 Lake Formation 注册的 Amazon S3 中的源数据，访问权限由适用于 Amazon S3 和 AWS Glue 操作的 IAM 权限策略决定。

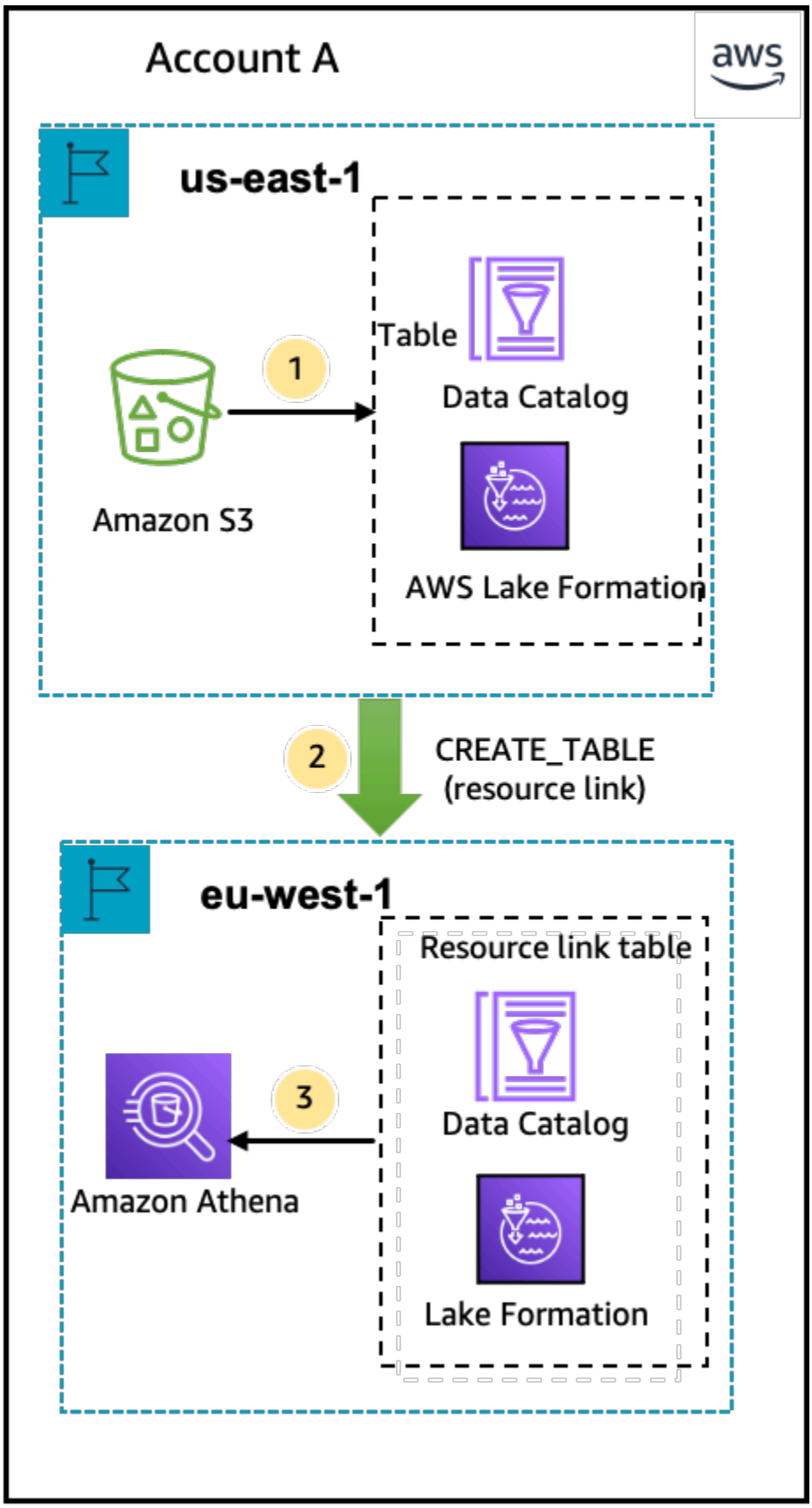
有关限制，请参阅跨区域数据访问限制。

工作流

下图显示了使用同一 AWS 账户和外部账户跨 AWS 区域访问数据的工作流。

用于访问同一个 AWS 账户内共享表的工作流

在下图中，与美国东部（弗吉尼亚州北部）区域中同一 AWS 账户中的用户共享数据，并且用户从欧洲地区（爱尔兰）区域查询共享数据。


Account A
aws
us-east-1
Table
Data Catalog
1
Amazon S3
AWS Lake Formation
2
CREATE_TABLE
(resource link)
eu-west-1
Resource link table
Data Catalog
3
Amazon Athena
Lake Formation

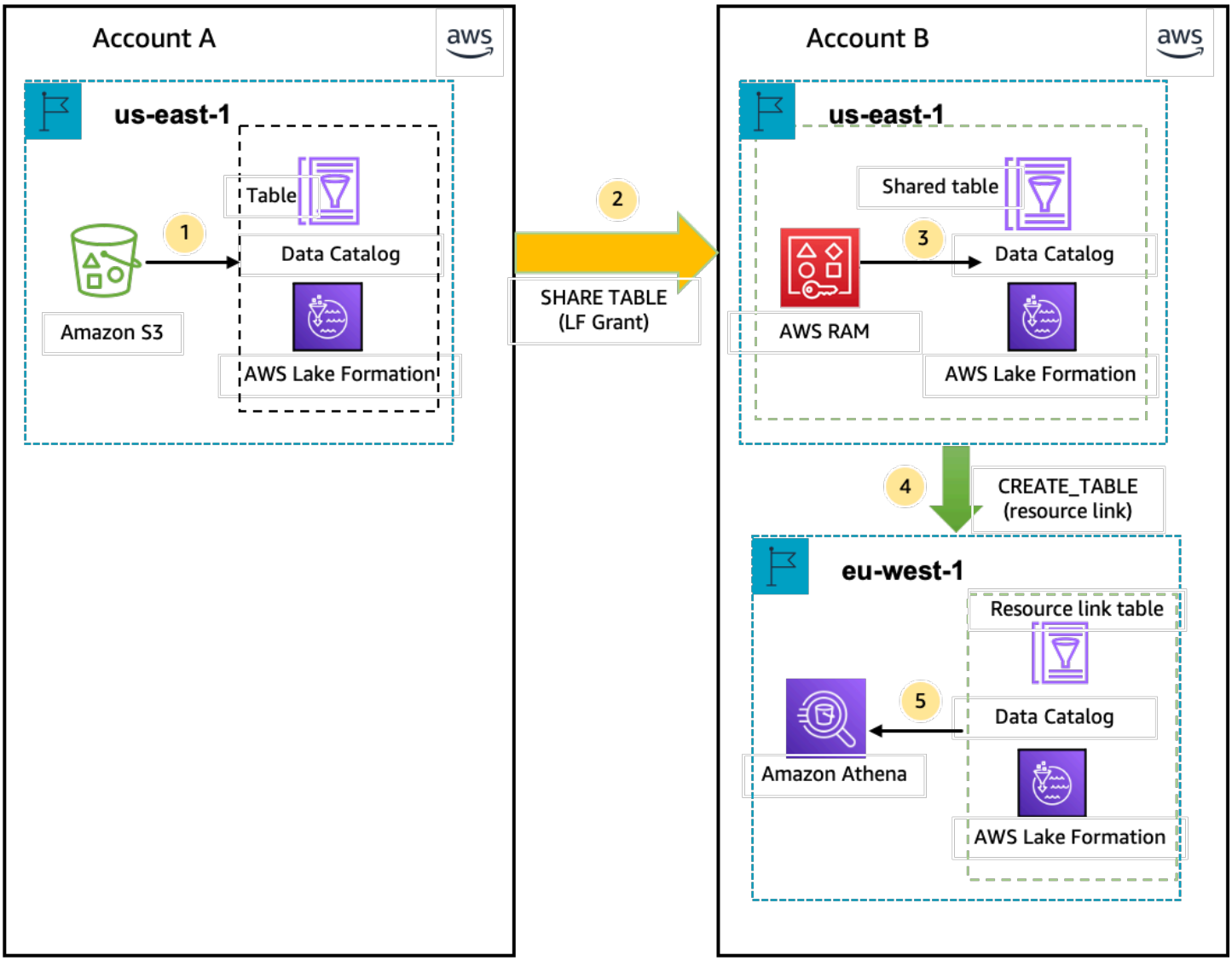
数据湖管理员执行以下活动（第 1-2 步）：

1. 数据湖管理员使用数据目录数据库和表设置 AWS 账户，并在美国东部（弗吉尼亚州北部）区域向 Lake Formation 中注册 Amazon S3 数据位置。

   向同一账户中的主体（用户）授予对数据目录资源（图中为 Product 表）的 `Select` 权限。
2. 在欧洲地区（爱尔兰）区域中创建指向美国东部（弗吉尼亚州北部）区域中源表的资源链接。向主体授予对欧洲地区（爱尔兰）区域资源链接的 `DESCRIBE` 权限。
3. 用户从欧洲地区（爱尔兰）区域使用 Athena 查询该表。

用于访问与外部 AWS 账户共享的表的工作流

在下图中，制作者账户（账户 A）托管 Amazon S3 存储桶，注册数据位置，并与美国东部（弗吉尼亚州北部）区域的使用者账户（账户 B）共享数据目录表，使用者账户（账户 B）中的用户从欧洲地区（爱尔兰）区域查询该表。



1. 数据湖管理员使用数据目录资源和在美国东部（弗吉尼亚州北部）区域向 Lake Formation 注册的 Amazon S3 数据位置来设置 AWS 账户（制作者账户）。
2. 制作者账户的数据湖管理员与使用者账户共享数据目录表。
3. 使用者账户的数据湖管理员接受美国东部（弗吉尼亚州北部）区域的数据共享邀请，并向来自同一区域的主体授予对共享表的 `Select` 权限。
4. 使用者账户的数据湖管理员在欧洲地区（爱尔兰）区域创建指向美国东部（弗吉尼亚州北部）区域的目标共享表的资源链接，并向用户授予对欧洲地区（爱尔兰）区域中该资源链接的 `DESCRIBE` 权限。
5. 用户从欧洲地区（爱尔兰）区域使用 Athena 查询数据。

设置跨区域表访问权限

要从其他区域访问数据，您需要先在注册 Amazon S3 数据位置的区域中设置数据目录数据库和表。您可以与您的账户中或其他账户中的主体共享数据目录数据库和表。然后，您需要创建数据湖管理员，他们可以创建指向用户查询数据的区域中目标共享数据位置的资源链接。

从不同区域查询同一个账户内的共享数据

在本节中，目标共享表区域是区域 A，用户从区域 B 运行查询。

1. 区域 A 中的账户设置（您可以在该区域中创建和共享数据）

   数据湖管理员需要完成以下操作：

   a. 注册 Amazon S3 位置。

      有关更多信息，请参阅向数据湖添加 Amazon S3 位置。

   b. 在账户中创建数据库和表。这也可以由有权创建数据库和表的非管理用户来完成。

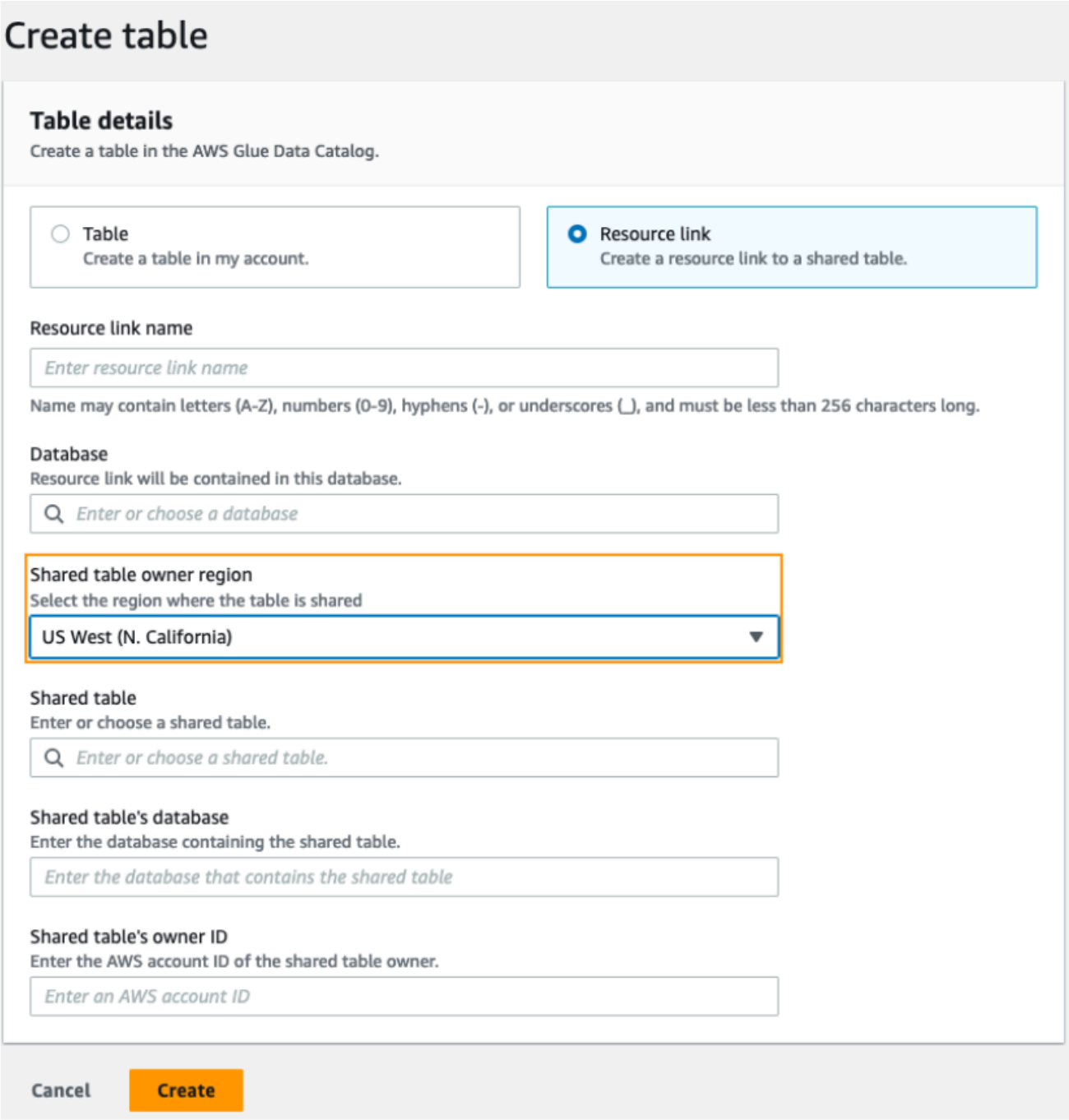
   c. 使用 `Grantable permissions` 向主体授予对表的数据权限。

      有关更多信息，请参阅授予和撤销对数据目录资源的权限。

2. 在区域 B（您访问数据的位置）中设置账户

   数据湖管理员需要完成以下操作：

   a. 在区域 B 中创建指向区域 A 中目标共享表的资源链接。在创建表屏幕上指定共享表所有者区域。



有关创建指向数据库和表的资源链接的说明，请参阅创建资源链接。

b. 向 IAM 主体授予对区域 B 中资源链接的 `Describe` 权限。

有关授予资源链接权限的更多信息，请参阅授予资源链接权限。

区域 B 中的 IAM 主体可以使用 Athena 通过链接查询目标表。

访问来自不同区域的跨账户数据

1. 制作者/授予者账户设置

   数据湖管理员需要完成以下操作：

   a. 在区域 A 中设置制作者/授予者账户
   b. 在区域 A 中注册 Amazon S3 数据位置
   c. 创建数据库和表。这可以由有权创建表的非管理用户完成。
   d. 使用 `Grantable permissions` 向使用者/被授权者账户授予对区域 A 中表的数据权限。

   有关更多信息，请参阅跨 AWS 账户 或来自外部账户的 IAM 主体共享数据目录表和数据库。

2. 使用者/被授权者账户设置

   数据湖管理员需要完成以下操作：

   a. 接受区域 A 中 AWS RAM 的资源共享邀请。
   b. 在区域 B 中创建指向共享表的资源链接。用户需要在区域 B 中查询表。
   c. 向区域 A 中的 IAM 主体授予对共享表的数据权限。

   > **Note**
   > 您必须在共享表的同一区域中授予对该共享表的权限。

   d. 向主体授予对区域 B 中资源链接的权限。

   然后，区域 B 中使用者账户中的主体使用 Athena 查询区域 B 中的共享表。

AWS Lake Formation 中的数据共享

您可以使用AWS Lake Formation数据共享功能来授予和管理存储在 Amazon S3 以外位置的数据以及存储在 Amazon S3 以外位置的元数据的权限AWS Glue Data Catalog。借助数据共享功能，您可以在 Amazon Redshift 中设置和管理数据集的权限，而无需将数据迁移到 Amazon S3 中。您也可以使用数据目录联合功能连接到外部元数据仓。

之后，您可以使用 Lake Formation 通过定义精细的访问控制策略来管理中央数据目录中的数据和访问权限。数据湖管理员可以向账户内的其他 IAM 主体或跨账户授予对数据目录资源的权限。IAM 主体可以使用 Amazon Redshift Spectrum 和 Amazon Athena 查询共享数据。

Lake Formation 提供了以下方法来共享数据并管理对外部数据集和外部元存储的权限：

- 将 Lake Formation 与 Amazon Redshift 数据共享集成 – 使用 Lake Formation 集中管理 Amazon Redshift 数据共享的数据库、表、列和行级别访问权限，并限制用户对数据共享内对象的访问。
- AWS Glue Data Catalog连接到外部元数据仓库-使用 Lake Formation 将AWS Glue Data Catalog连接到外部元数据仓库，以管理 Amazon S3 中数据集的访问权限。无需将元数据迁移到 AWS Glue Data Catalog。
- 将 Lake Formation 和 AWS Data Exchange 集成 – Lake Formation 支持许可通过 AWS Data Exchange 对您的数据进行访问。如果您有兴趣获得 Lake Formation 数据的许可，请参阅《AWS Data Exchange 用户指南》中的什么是 AWS Data Exchange？。

主题

管理对 Amazon Redshift 数据共享中数据的权限

借助AWS Lake Formation，您可以在 Amazon Redshift 的数据共享中安全地管理数据。Amazon Redshift 是 AWS Cloud 中的一种完全托管的 PB 级数据仓库服务。使用数据共享功能，Amazon Redshift 可帮助您跨 AWS 账户共享数据。有关 Amazon Redshift 数据共享的更多信息，请参阅 Amazon Redshift 中的数据共享概述。

在 Amazon Redshift 中，生产者集群管理员创建一个数据共享，并将其与数据湖管理员共享。有关创建数据湖管理员的 step-by-step 说明，请参阅创建数据湖管理员。

在您（数据湖管理员）接受数据共享后，必须为特定数据共享创建 AWS Glue Data Catalog 数据库。这样您就可以使用 Lake Formation 权限来控制对它的访问。Lake Formation 将每个数据共享映射到相应的数据目录数据库。它们在数据目录中显示为联合数据库。

当数据库指向数据目录之外的实体时，该数据库被称为联合数据库。Amazon Redshift 数据共享中的表和视图在数据目录中作为单个表列出。您可以通过 Lake Formation 与同一个账户或其他账户中的选定 IAM 主体和 SAML 用户共享联合数据库。您还可以添加行和列筛选表达式，以限制对某些数据的访问。有关更多信息，请参阅数据筛选概览。

要向用户提供访问 Amazon Redshift 数据共享的权限，您必须执行以下操作：

1. 更新数据目录设置以启用 Lake Formation 权限。
2. 接受 Amazon Redshift 制作者集群管理员发出的数据共享邀请，然后在 Lake Formation 中注册数据共享。

   完成此步骤后，您可以在 Lake Formation 数据目录中管理数据共享。
3. 创建联合数据库并定义对该数据库的权限。
4. 向用户授予数据库和表的权限。您可以与同一账户或其他账户中的用户共享整个数据库或表的子集。

有关限制，请参阅Amazon Redshift 数据共享限制。

主题

在 Amazon Redshift 数据共享上设置权限的先决条件

更新默认数据目录设置

要为数据目录资源启用 Lake Formation 权限，我们建议您禁用 Lake Formation 中的默认数据目录设置。有关更多信息，请参阅更改默认权限模式或使用混合访问模式。

更新权限

除了数据湖管理员权限 (`AWSLakeFormationDataAdmin`) 之外，还需要以下权限才能在 Lake Formation 中接受 Amazon Redshift 数据共享：

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`
- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

数据湖管理员 IAM 用户隐式拥有以下权限。

- data_location_access
- create_database
- lakefomation:registerResource

设置 Amazon Redshift 数据共享权限

本主题介绍接受数据共享邀请、创建联合数据库和授予权限所需执行的步骤。您可以使用 Lake Formation 控制台或 AWS Command Line Interface (AWS CLI)。本主题中的示例显示了同一个账户中的制作者集群、数据目录和数据使用者。

要了解有关 Lake Formation 跨账户功能的更多信息，请参阅Lake Formation 中的跨账户数据共享。

设置数据共享权限

1. 查看数据共享邀请并接受。

   Console

   1. 通过 https://console.aws.amazon.com/lakeformation/ 以数据湖管理员的身份登录 Lake Formation 控制台。导航到数据共享页面。
   2. 查看您有权访问的数据共享。状态列表示您当前参与数据共享的状态。待处理状态表示您已被添加到数据共享，但尚未接受或已拒绝邀请。
   3. 要回复数据共享邀请，请选择数据共享名称并选择查看邀请。在接受或拒绝数据共享中，查看邀请详细信息。选择接受接受邀请，或选择拒绝拒绝邀请。如果您拒绝邀请，则无法访问数据共享。

   AWS CLI

   以下示例显示如何查看、接受和注册邀请。将AWS 账户身份证替换为有效的AWS 账户身份证。将 `data-share-arn` 替换为引用数据共享的实际 Amazon 资源名称 (ARN)。

1. 查看待处理的邀请。

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
```

2. 接受数据共享。

```
 aws redshift associate-data-share-consumer \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
 --consumer-arn 'arn:aws:glue:us-east-1:111122223333:catalog
```

3. 在 Lake Formation 账户中注册数据共享。使用 RegisterResourceAPI 操作在 Lake Formation 中注册数据共享。 `DataShareArn`是的输入参数`ResourceArn`。

> **Note**
>
> 此步骤为必需步骤。

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds'
```

2. 创建数据库。

接受数据共享邀请后，您需要创建一个指向与数据共享关联的 Amazon Redshift 数据库的数据库。您必须是数据湖管理员才能创建数据库。

Console

1. 从邀请窗格中选择数据共享，然后选择设置数据库详细信息。
2. 在设置数据库详细信息中，输入数据共享的唯一名称和标识符。您可以使用此标识符在元数据层次结构 (dbname.schema.Table) 内部映射数据共享。
3. 选择下一步，向其他用户授予对共享数据库和表的权限。

AWS CLI

使用以下示例代码创建指向使用与 Lake Formation 共享的 Amazon Redshift 数据库的数据库。AWS CLI

```
aws glue create-database --cli-input-json \

'{
 "CatalogId": "111122223333",
 "DatabaseInput": {
  "Name": "tahoedb",
  "FederatedDatabase": {
       "Identifier": "arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/federatedds",
       "ConnectionName": "aws:redshift"
   }
 }
 }'
```

3. 授予权限

创建数据库后，您可以向账户中的用户或外部AWS 账户和组织授予权限。您将无法对映射到 Amazon Redshift 数据共享的联合数据库授予写入数据权限（插入、删除）和元数据权限（更改、删除、创建）。有关授予权限的更多信息，请参阅管理 Lake Formation 权限。

> **Note**
>
> 作为数据湖管理员，您只能查看联合数据库中的表。要执行任何其他操作，您需要授予自己对这些表的更多权限。

Console

1. 在授予权限屏幕上，选择要向其授予权限的用户。
2. 选择授权。

AWS CLI

在 AWS CLI 中使用以下示例通过以下方式授予数据库和表权限：

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
   "Principal": {
           "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/non-
admin"
   },
   "Resource": {
          "Database": {
                "CatalogId": "111122223333",
                 "Name": "tahoedb"
           }
    },
    "Permissions": [
             "DESCRIBE"
     ],
    "PermissionsWithGrantOption": [

     ]
 }
```

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
                   "Principal": {
                          "DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/non-admin"
                    },
                   "Resource": {
                          "Table": {
                               "CatalogId": "111122223333",
                               "DatabaseName": "tahoedb",
                               "Name": "public.customer"
                           }
                    },
                   "Permissions": [
                          "SELECT"
                    ],
                   "PermissionsWithGrantOption": [
                          "SELECT"
                     ]
```

```
}
```

查询联合数据库

在您授予权限后，用户可以登录并开始使用 Amazon Redshift 查询联合数据库。用户现在可以在 SQL 查询中使用本地数据库名称引用 Amazon Redshift 数据共享。在 Amazon Redshift 中，通过数据共享来共享的公共架构中的客户表将在数据目录中创建相应的表，即 `public.customer`。

1. 在使用 Amazon Redshift 查询联合数据库之前，集群管理员使用以下命令通过数据目录数据库创建一个数据库：

```
CREATE DATABASE sharedcustomerdb FROM ARN
 'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA
 tahoedb
```

2. 集群管理员授予对数据库的使用权限。

```
GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
```

3. 您（联合用户）现在可以登录 SQL 工具来查询表。

```
Select * from sharedcustomerdb.public.customer limit 10;
```

有关更多信息，请参阅《Amazon Redshift 管理指南》中的查询 AWS Glue Data Catalog。

管理对使用外部元存储的数据集的权限

借助 AWS Glue Data Catalog 元数据联合身份验证（数据目录联合身份验证），您可以将数据目录连接到存储您 Amazon S3 数据的元数据的外部元存储，并使用 AWS Lake Formation 安全地管理数据访问权限。您不必将元数据从外部元存储迁移到数据目录中。

数据目录提供了一个集中的元数据存储库，可以更轻松地管理和发现不同系统的数据。当您的组织管理数据目录中的数据时，您可以使用 AWS Lake Formation 控制对 Amazon S3 中数据集的访问。

> **Note**
>
> 目前，我们仅支持 Apache Hive（版本 3 及更高版本）元存储联合身份验证。

为了设置数据目录联合，我们在HiveMetastore中提供了一个名为 GlueDataCatalogFederation- 的 AWS Serverless Application Model (AWS SAM) 应用程序AWS Serverless Application Repository。

该参考实现 GitHub 作为开源项目在 Federation -Hive Meta store 上提供。AWS Glue Data Catalog

AWS SAM 应用程序创建并部署将数据目录连接到 Hive 元存储所需的以下资源：

- AWS Lambda函数 — 托管联合身份验证服务的实现，该服务在数据目录和 Hive 元数据仓之间进行通信。 AWS Glue调用此 Lambda 函数从 Hive 元数据存储中检索元数据对象。
- Amazon API Gateway – Hive 元存储的连接端点，该存储充当将所有调用路由到 Lambda 函数的代理。
- IAM 角色 — 具有在数据目录和 Hive 元数据仓之间创建连接的必要权限的角色。
- AWS Glueconnec tion — Amazon API Gateway 一种存储Amazon API Gateway终端节点和用于调用它的 IAM 角色的AWS Glue连接类型。

当您查询表时，AWS Glue 服务会对 Hive 元存储进行运行时调用并获取元数据。Lambda 函数充当 Hive 元存储和数据目录之间的转换器。

建立连接后，为了将 Hive 元存储中的元数据与数据目录同步，您需要使用 Hive 元存储连接详细信息在数据目录中创建联合数据库，并将该数据库映射到 Hive 数据库。当数据库指向数据目录外的实体时，该数据库被称为“联合数据库”。

您可以使用基于标签的访问控制和联合数据库上的命名资源方法来应用 Lake Formation 权限，并在多个AWS 账户AWS Organizations、和组织单位 (OU) 之间共享该权限。您也可以直接与其他账户的 IAM 主体共享联合数据库。

您可以使用外部 Hive 表上的 Lake Formation 数据筛选器在列级别、行级别和单元级别定义精细权限。你可以使用亚马逊 Athena、Amazon Redshift 或亚马逊 EMR 来查询 Lake Formation 托管的外部 Hive 表。

有关跨账户数据共享和数据筛选的更多信息，请参阅：

- Lake Formation 中的跨账户数据共享
- Lake Formation 中的数据筛选和单元格级别安全性

数据目录元数据联合身份验证主要步骤

1. 您可以创建具有部署AWS SAM应用程序和创建联合数据库的相应权限的 IAM 用户和角色。

2. 通过为使用外部 Hive 元存储的数据集的选择 `Enable Data Catalog federation` 选项，您可以在 Lake Formation 中注册 Amazon S3 数据位置。
3. 您可以配置 AWS SAM 应用程序设置（AWS Glue 连接名称、Hive 元存储 URL 和 Lambda 函数参数）并部署 AWS SAM 应用程序。
4. AWS SAM 应用程序部署将外部 Hive 元存储与数据目录连接所需的资源：
5. 要对 Hive 数据库和表应用 Lake Formation 权限，请使用 Hive 元数据仓连接详细信息在数据目录中创建一个数据库，然后将此数据库映射到 Hive 数据库。
6. 向您的账户或其他账户中的主体授予对联合数据库的权限。

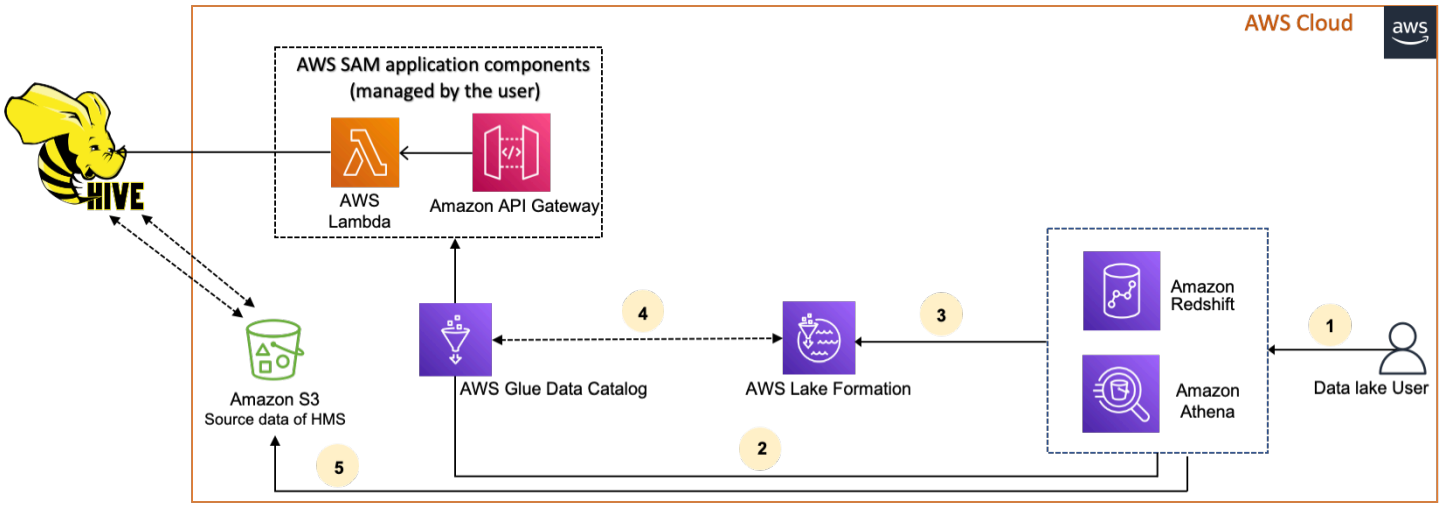
> **Note**
>
> 无需应用 Lake Formation 权限，即可将数据目录连接到外部 Hive 元存储、创建联合数据库以及在 Hive 数据库和表上运行查询和 ETL 脚本。对于未向 Lake Formation 注册的 Amazon S3 中的源数据，访问权限由适用于 Amazon S3 和 AWS Glue 操作的 IAM 权限策略决定。

有关限制，请参阅Hive 元数据存储数据共享注意事项和限制。

主题

工作流

以下图表显示了将 AWS Glue Data Catalog 连接到外部 Hive 元存储的工作流。



1. 主体使用 Athena 或 Redshift Spectrum 等集成服务提交查询。
2. 集成服务调用数据目录获取元数据，然后数据目录会调用后面可用的 Hive 元数据仓端点Amazon API Gateway，并接收对元数据请求的响应。
3. 集成服务向 Lake Formation 发送请求，以验证表信息和用于访问表的凭证。
4. Lake Formation 对请求进行授权，并将临时凭证售卖给集成应用程序以允许其访问数据。
5. 该集成服务使用从 Lake Formation 收到的临时凭证从 Amazon S3 读取数据并将结果共享给主体。

将数据目录连接到 Hive 元存储的先决条件

要将 AWS Glue Data Catalog 连接到外部 Apache Hive 元存储并设置数据访问权限，您需要完成以下要求：

> **Note**
>
> 我们建议 Lake Formation 管理员部署 AWS SAM 应用程序，并且只有特权用户才能使用 Hive 元存储连接来创建相应的联合数据库。

1. 创建 IAM 角色。

   部署 AWS SAM 应用程序

   - 创建具有部署资源（Lambda 函数、Amazon API Gateway、IAM 角色和 AWS Glue 连接）所需的权限的角色，以创建与 Hive 元存储的连接。

创建联合数据库

资源需要以下权限：

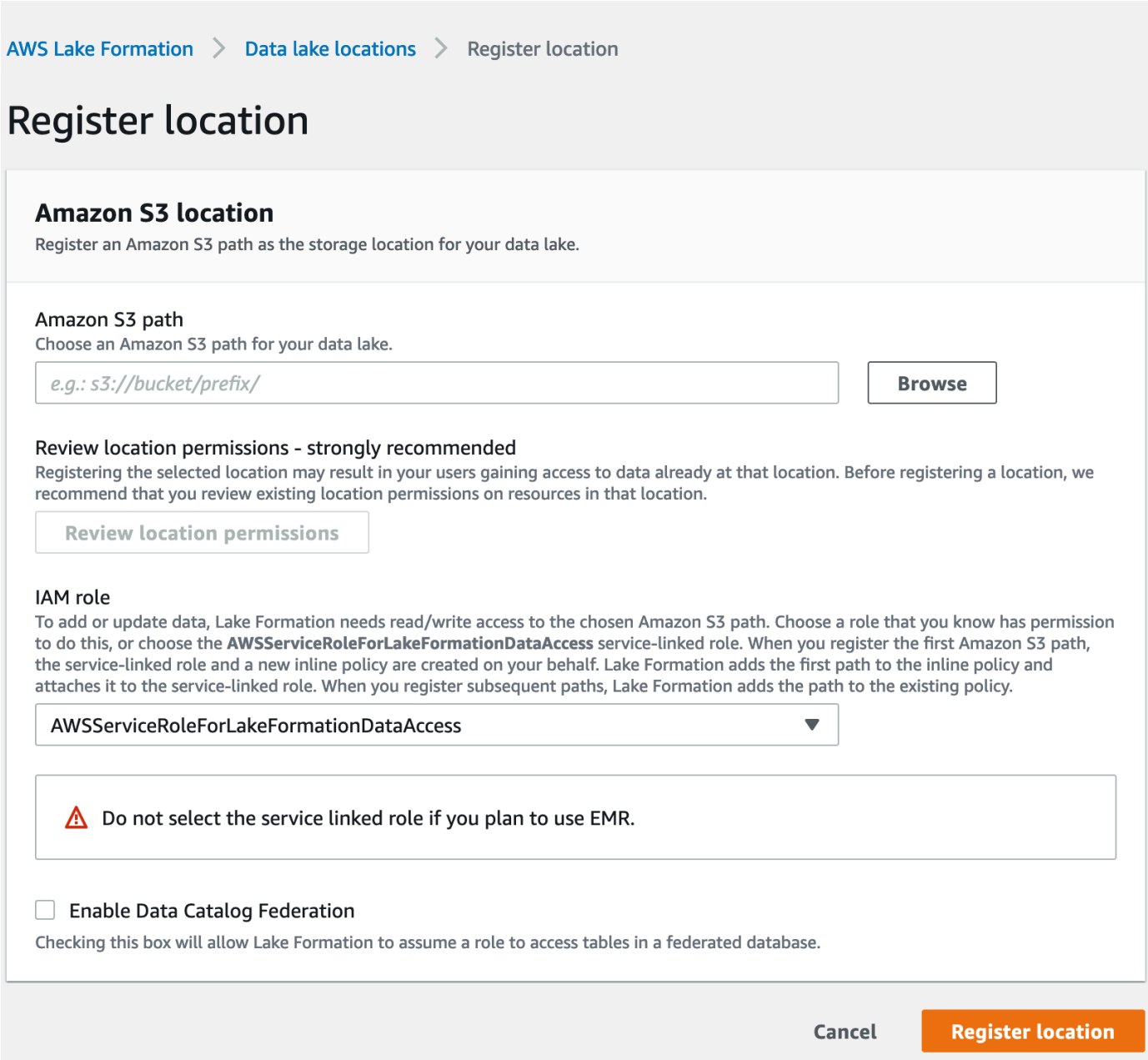
- `glue:CreateDatabase on resource arn:aws:glue:region:account-id:database/gluedatabasename`
- `glue:PassConnection on resource arn:aws:glue:region:account-id:connection/hms_connection`

2. 在 Lake Formation 中注册 Amazon S3 位置。

   要使用 Lake Formation 管理和保护数据湖中的数据，您必须在 Lake Formation 中注册含有 Hive 元存储中表数据的 Amazon S3 位置。通过这样做，Lake Formation可以向Athena、Redshift Spectrum和Amazon EMR等AWS分析服务机构出售证书。

   有关注册 Amazon S3 位置的更多信息，请参阅向数据湖添加 Amazon S3 位置。

   注册 Amazon S3 位置时，选中 “启用数据目录联合” 复选框以允许 Lake Formation 代入访问联合数据库中表的角色。



有关在 Lake Formation 中注册数据位置的更多信息，请参阅[为您的数据湖配置 Amazon S3 位置](#)。

3. 使用正确的亚马逊 EMR 版本。

   要将 Amazon EMR 与联合 Hive 元数据仓数据库配合使用，您需要拥有 Hive 3.x 或更高版本和亚马逊 EMR 版本 6.x 或更高版本。

将数据目录连接到外部 Hive 元存储

要连接AWS Glue Data Catalog到 Hive 元数据仓，您需要部署一个名为 GlueDataCatalogFederation-的AWS SAM应用程序。HiveMetastore它创建将外部 Hive 元存储与数据目录连接起来所需的资源。您可以在 AWS Serverless Application Repository 中访问 AWS SAM 应用程序。

AWS SAM 应用程序使用 Lambda 函数为 Amazon API Gateway 后面的 Hive 元存储创建连接。AWS SAM 应用程序使用统一资源标识符 (URI) 作为用户的输入，并将外部 Hive 元存储连接到数据目录。当用户对 Hive 表运行查询时，数据目录会调用 API Gateway 终端节点。该端点调用 Lambda 函数来检索 Hive 表的元数据。

将数据目录连接到 Hive 元存储并设置权限

1. 部署 AWS SAM 应用程序。

    1. 登录到 AWS Management Console并打开 AWS Serverless Application Repository。
    2. 在导航窗格中，选择 Available applications (可用应用程序)。
    3. 选择公用应用程序。
    4. 选择 Show apps that create custom IAM roles or resource policies（显示创建自定义 IAM 角色或资源策略的应用程序）选项。
    5. 在搜索框中输入名称 GlueDataCatalogFederation-HiveMetastore。
    6. 选择 GlueDataCatalogFederation-HiveMetastore 应用程序。
    7. 在应用程序设置下，为您的 Lambda 函数输入以下必需的最低设置：
        - 应用程序名称 - AWS SAM 应用程序的名称。
        - GlueConnectionName-连接的名称。
        - HiveMetastoreURI-您的 Hive 元数据仓主机的 URI。
        - LambdaMemory-从 128-10240 开始的 Lambda 内存量（以 MB 为单位）。默认值为 1024。
        - LambdaTimeout-Lambda 调用的最大运行时间（以秒为单位）。默认值为 30。
        - VPC SecurityGroupIds 和 VPC SubnetIds-存在 Hive 元数据仓的 VPC 的信息。
    8. 选中 I acknowledge that this app creates custom IAM roles and resource policies（我确认此应用程序创建自定义 IAM 角色和资源策略）。有关更多信息，请选择 Info (信息) 链接。
    9. 在 Application settings（应用程序设置）部分的右下角，选择 Deploy（部署）。部署完成后，Lambda 函数将显示在 Lambda 控制台中的 Resource（资源）部分。

该应用程序已部署到 Lambda。它的名称前面带有 serverlessrepo-，表示该应用程序是从中部署的。AWS Serverless Application Repository选择该应用程序会将您带到资源页面，该页面中列出了已部署的应用程序的每个资源。这些资源包括允许在数据目录和 Hive 元存储之间进行通信的 Lambda 函数、AWS Glue 连接以及数据库联合身份验证所需的其他资源。

2. 在数据目录中创建数据库。

   创建与 Hive 元数据仓的连接后，可以在数据目录中创建指向外部 Hive 元数据仓数据库的联合数据库。您需要在数据目录中为连接到数据目录的每个 Hive 元数据仓数据库创建相应的数据库。

   Lake Formation console

   1. 在数据共享页面上，选择共享数据库选项卡，然后选择创建数据库。
   2. 在 “连接名称” 中，从下拉菜单中选择 Hive 元数据仓连接的名称。
   3. 输入唯一数据库名称和数据库的联合身份验证源标识符。这是您在查询表时在 SQL 语句中使用的名称。该名称最多可包含 255 个字符，并且在您的账户中必须是唯一的。
   4. 选择创建数据库。

   AWS CLI

   ```
   aws glue create-database \
   '{
    "CatalogId": "<111122223333>",
     "database-input": {
       "Name":"<fed_glue_db>",
       "FederatedDatabase":{
           "Identifier":"<hive_db_on_emr>",
           "ConnectionName":"<hms_connection>"
        }
      }
    }'
   ```

3. 查看联合数据库中的表。

   创建联合数据库后，您可以使用 Lake Formation 控制台或 AWS CLI 查看 Hive 元存储中表的列表。

Lake Formation console

1. 从共享数据库选项卡中选择数据库名称。
2. 在数据库页面上，选择查看表。

AWS CLI

以下示例说明如何检索连接定义、数据库名称以及数据库中的部分或全部表。将数据目录的 ID 替换为创建数据库时使用的有效 AWS 账户 ID。将 hms_connection 替换为连接名称。

```
aws glue get-connection \
--name <hms_connection>  \
--catalog-id 111122223333
```

```
aws glue get-database \
--name <fed_glu_db> \
--catalog-id 111122223333
```

```
aws glue get-tables \
--database-name <fed_glue_db> \
--catalog-id 111122223333
```

```
aws glue get-table \
--database-name <fed_glue_db> \
--name <hive_table_name> \
--catalog-id 111122223333
```

4. 授予权限

创建数据库后，您可以向账户中的其他 IAM 用户和角色或外部AWS 账户和组织授予权限。您将无法授予对联合数据库的写入数据权限（插入、删除）和元数据权限（更改、删除、创建）。有关授予权限的更多信息，请参阅管理 Lake Formation 权限。

5. 查询联合数据库。

在您授予权限后，用户可以使用 Athena 和 Amazon Redshift 登录并开始查询联合数据库。用户现在可以在 SQL 查询中使用本地数据库名称引用 Hive 数据库。

Amazon Athena查询语法示例

fed_glue_db替换为之前创建的本地数据库名称。

```
Select * from fed_glue_db.customers limit 10;
```

其他 资源

以下博客文章包含有关在 Hive 元存储数据库和表上设置 Lake Formation 权限以及使用 Athena 查询它们的详细说明。我们还演示了一个跨账户共享用例，其中生产者账户 A 中的 Lake Formation 委托人与消费者账户 B 共享一个联合 Hive 数据库和使用 LF-Tag 的表。

- 使用 AWS Lake Formation 权限查询您的 Apache Hive 元存储

AWS Lake Formation 中的安全性

AWS 十分重视云安全性。作为 AWS 客户，您将从专为满足大多数安全敏感型企业的要求而打造的数据中心和网络架构中受益。

安全性是 AWS 和您的共同责任。责任共担模式将其描述为云的安全性和云中的安全性：

- 云的安全性 – AWS负责保护在AWS云中运行AWS服务的基础设施。AWS还向您提供可安全使用的服务。作为 AWS 合规性计划的一部分，第三方审核人员将定期测试和验证安全性的有效性。要了解适用于 AWS Lake Formation 的合规性计划，请参阅合规性计划范围内的 AWS 服务。
- 云中的安全性 - 您的责任由您使用的 AWS 服务决定。您还需要对其它因素负责，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。

此文档将帮助您了解如何在使用 Lake Formation 时应用责任共担模式。以下主题说明如何配置 Lake Formation 以实现您的安全性和合规性目标。您还将了解如何使用其他 AWS 服务来帮助您监控和保护 Lake Formation 资源。

主题

Lake Formation 中的数据保护

AWS 责任共担模式适用于 AWS Lake Formation 中的数据保护。如该模式中所述，AWS 负责保护运行所有 AWS Cloud 的全球基础设施。您负责维护对托管在此基础设施上的内容的控制。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 安全性博客 上的博客文章 AWS Shared Responsibility Model and GDPR。

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) 设置单个用户。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 AWS 资源进行通信。我们要求使用 TLS 1.2，建议使用 TLS 1.3。
- 使用 AWS CloudTrail 设置 API 和用户活动日志记录。
- 使用 AWS 加密解决方案以及 AWS 服务 中的所有默认安全控制。
- 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。
- 如果您在通过命令行界面或 API 访问 AWS 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）。这包括使用控制台、API、AWS CLI 或 AWS SDK 处理 Lake Formation 或其他 AWS 服务时。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，我们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 Lake Formation 支持以下方面的数据加密：

- Amazon Simple Storage Service (Amazon S3) 数据湖中的数据。

  Lake Formation 支持使用 AWS Key Management Service (AWS KMS) 进行数据加密。数据通常通过 AWS Glue 提取、转换、加载 (ETL) 作业写入数据湖。有关如何加密 AWS Glue 作业写入的数据的信息，请参阅《AWS Glue 开发人员指南》中的加密爬网程序、作业和开发端点写入的数据。
- AWS Glue Data Catalog，这是 Lake Formation 存储描述数据湖中数据的元数据表的位置。

  有关更多信息，请参阅《AWS Glue 开发人员指南》中的加密数据目录。

要在数据湖中将 Amazon S3 位置添加为存储，请向 AWS Lake Formation 注册该位置。然后，您可以使用 Lake Formation 权限对指向此位置的 AWS Glue Data Catalog 对象以及该位置中的基础数据进行精细访问控制。

Lake Formation 支持注册包含加密数据的 Amazon S3 位置。有关更多信息，请参阅注册加密的 Amazon S3 位置。

AWS Lake Formation 中的基础设施安全性

作为一项托管式服务，AWS Lake Formation 由Amazon Web Services：安全流程概览白皮书中所述的AWS全球网络安全程序提供保护。

您可以使用 AWS 发布的 API 调用通过网络访问 Lake Formation。客户端必须支持传输层安全性 (TLS) 1.0 或更高版本。建议使用 TLS 1.2 或更高版本。客户端还必须支持具有完全向前保密 (PFS) 的密码套件，例如 Ephemeral Diffie-Hellman (DHE) 或 Elliptic Curve Ephemeral Diffie-Hellman (ECDHE)。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 委托人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。

跨服务混淆代理问题防范

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 AWS 中，跨服务模拟可能会导致混淆代理问题。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委托人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 和 aws:SourceAccount 全局条件上下文键，以限制 AWS Lake Formation 为其他服务提供的资源访问权限。如果使用两个全局条件上下文键，在同一策略语句中使用时，aws:SourceAccount 值和 aws:SourceArn 值中的账户必须使用相同的账户 ID。

目前，Lake Formation 仅支持以下格式的 aws:SourceArn：

```
arn:aws:lakeformation:aws-region:account-id:*
```

以下示例演示如何使用 Lake Formation 中的 aws:SourceArn 和 aws:SourceAccount 全局条件上下文键来防范代理混淆问题。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ConfusedDeputyPreventionExamplePolicy",
      "Effect": "Allow",
      "Principal": {
```

```
        "Service": "lakeformation.amazonaws.com"
      },
      "Action": [
        "sts:AssumeRole"
      ],
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "account-id"
        },
        "ArnEquals": {
          "aws:SourceArn": "arn:aws:lakeformation:aws-region:account-id:*"
        }
      }
    }
  ]
}
```

AWS Lake Formation 中的安全事件日志记录

AWS Lake Formation 与 AWS CloudTrail 集成，后者是一项服务，可用于记录 Lake Formation 中由用户、角色或 AWS 服务所执行的操作。CloudTrail 将 Lake Formation 的所有 API 调用作为事件捕获。捕获的调用包括来自 Lake Formation 控制台的调用、来自 AWS Command Line Interface的调用，以及对 Lake Formation API 操作的代码调用。

有关 Lake Formation 中的事件日志记录的更多信息，请参阅使用 AWS CloudTrail 记录 AWS Lake Formation API 调用。

> **Note**
>
> `GetTableObjects`、`UpdateTableObjects` 和 `GetWorkUnitResults` 大容量是数据面板操作。对这些 API 的调用当前未记录到 CloudTrail。有关 CloudTrail 中的数据面板操作的更多信息，请参阅《AWS CloudTrail 用户指南》中的记录跟踪的数据事件。
> Lake Formation 中为支持其他 CloudTrail 事件而进行的更改将记录 AWS Lake Formation 的文档历史记录中。

将第三方服务与 Lake Formation 集成

与 AWS Lake Formation 集成使第三方服务能够安全地访问其基于 Amazon S3 的数据湖中的数据。您可以使用 Lake Formation 作为授权引擎，以通过集成的 AWS 服务（例如 Amazon Athena、Amazon EMR 和 Redshift Spectrum）管理或强制实施对数据湖的权限。Lake Formation 提供了两个用于集成服务的选项：

1. Lake Formation 应用程序集成设置：Lake Formation 可以基于有效权限将范围缩小的临时凭证以 AWS STS 令牌的形式售卖给已注册的 Amazon S3 位置，以便经授权的应用程序可以代表用户访问数据。
2. 集中强制实施：Lake Formation 查询 API 操作从 Amazon S3 检索数据并基于有效权限筛选结果。与查询 API 操作集成的引擎或应用程序可以依赖 Lake Formation 来评估调用身份的权限，并基于这些权限安全地筛选数据。第三方查询引擎只能查看和操作筛选后的数据。

主题

使用 Lake Formation 应用程序集成

Lake Formation 允许第三方服务与 Lake Formation 集成，并通过使用GetTemporaryGlueTableCredentials和GetTemporaryGluePartitionCredentials操作代表其用户临时访问亚马逊 S3 数据。这允许第三方服务使用其他 AWS 分析服务使用的相同授权和凭证售卖功能。本节介绍如何使用这些 API 操作将第三方查询引擎与 Lake Formation 集成。

默认情况下，这些 API 操作处于禁用状态。有两个选项可以授权 Lake Formation 集成应用程序：

- 配置每次调用应用程序集成 API 操作时都要验证的 IAM 会话标签

  有关更多信息，请参阅为第三方查询引擎启用调用应用程序集成 API 操作所需的权限。
- 启用允许外部引擎以完整表访问权限访问 Amazon S3 位置的数据选项

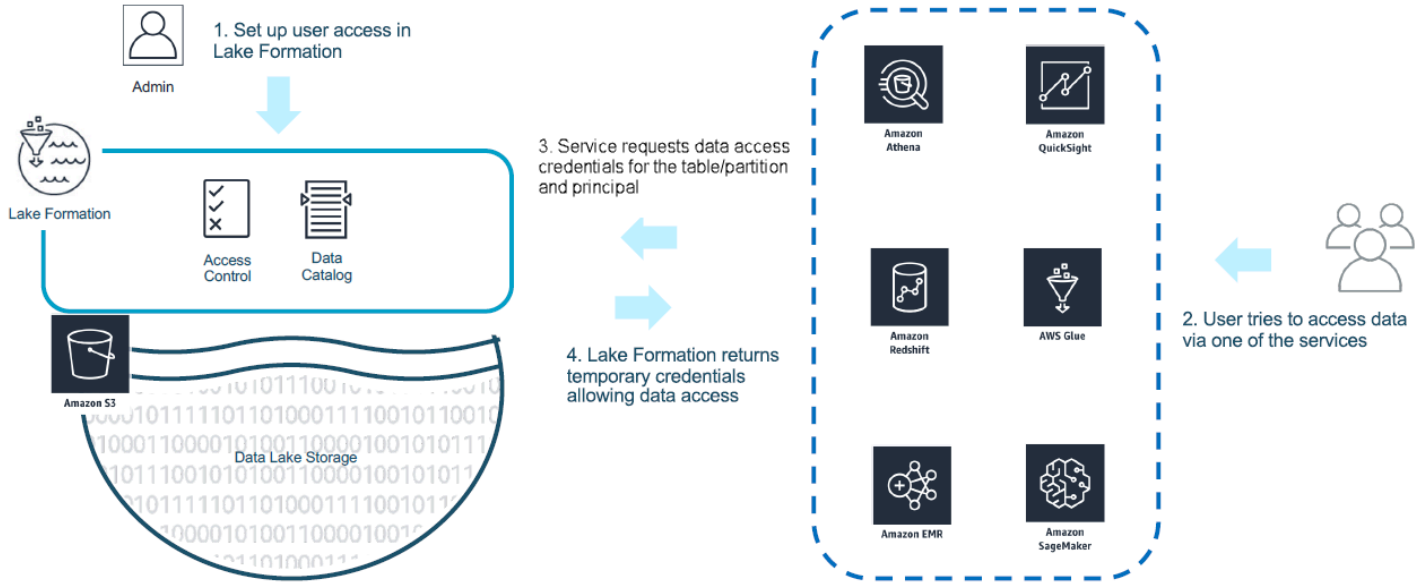
  如果用户拥有完整的表访问权限，则此选项允许查询引擎和应用程序在没有 IAM 会话标签的情况下获取凭证。它为查询引擎和应用程序提供了性能优势，并简化了数据访问。Amazon EC2 上的 Amazon EMR 能够利用此设置。

  有关更多信息，请参阅集成应用程序以获取完整表访问权限 。

主题

Lake Formation 应用程序集成的工作原理

本节介绍如何使用应用程序集成 API 操作将第三方应用程序（查询引擎）与 Lake Formation 集成。



1. Lake Formation 管理员执行以下活动：
   - 通过提供具有访问 Amazon S3 位置处数据的适当权限的 IAM 角色（用于售卖凭证）向 Lake Formation 注册该位置
   - 注册第三方应用程序，使其能够调用 Lake Formation 的凭证售卖 API 操作。请参阅 the section called “注册第三方查询引擎”。
   - 向用户授予访问数据库和表的权限

     例如，如果您要发布的用户会话数据集中的一些列包含个人身份信息 (PII) 以限制访问，则您可以为这些列分配一个名为“分类”，值为“敏感”的 LF-TBAC 标签。接下来，您可以定义一种权限，允许业务分析师访问用户会话数据，但那些标有分类 = 敏感的列除外。

2. 主体（用户）向集成服务提交查询。
3. 集成应用程序向 Lake Formation 发送请求，要求它提供表信息和用于访问表的凭证。
4. 如果查询主体有权访问该表，Lake Formation 会将凭证返回到允许访问数据的集成应用程序。

> **Note**
>
> 出售凭据时，Lake Formation 无法访问底层数据。

5. 该集成服务读取来自 Amazon S3 的数据，根据它收到的策略筛选列，然后将结果返回给主体。

> **Important**
>
> Lake Formation凭证售卖 API 操作支持分布式强制实施，并采用失败时显式拒绝（失败关闭）模型。这在客户、第三方服务和 Lake Formation 之间引入了一个三方安全模型。集成服务值得信赖，可以正确强制实施 Lake Formation 权限（分布式强制实施）。

集成服务负责根据 Lake Formation 返回的策略筛选从 Amazon S3 读取的数据，然后再将筛选后的数据返回给用户。集成服务遵循失败关闭模型，这意味着如果它们无法强制实施所需的 Lake Formation 权限，则必定会使查询失败。

Lake Formation 应用程序集成中的角色和责任

角色	责任
客户	• 启用 Lake Formation 应用程序集成设置（请参阅the section called “注册第三方查询引擎”）。 • 在 Lake Formation 中显式注册经批准的第三方（请参阅the section called “注册第三方查询引擎”）。 • 使用 Lake Formation 权限测试和验证第三方解决方案。 • 监控和审计第三方对 Lake Formation 凭证售卖 API 操作的使用情况。
第三方	• 公开记录每个软件版本的支持的功能，并提供有关正确启用该功能的说明。 • 准确公布在调用 Lake Formation 凭证售卖 API 操作时支持的功能（根据文档）。

角色	责任
	• 安全地存储和处理所售卖的凭证，以避免凭证泄露和权限提升。 • 根据支持的功能强制实施权限并仅向用户返回筛选后的数据 • 无法正确强制实施所需权限时使查询失败
AWS Lake Formation	• 为给定主体正确派生并返回有效权限。 • 以 API 操作为 call-by-call 基础验证第三方支持的功能。 • 只有当引擎公布的功能与目录资源上定义的功能相匹配时，才会返回范围缩小的 IAM 凭证，否则会返回错误。

用于应用程序集成 API 操作的 Lake Formation 工作流

以下是用于应用程序集成 API 操作的工作流：

1. 某用户使用集成的第三方查询引擎提交数据查询或请求。查询引擎担任一个代表该用户或一组用户的 IAM 角色，并检索可信凭证，以便在调用应用程序集成 API 操作时使用。
2. 查询引擎会调用 `GetUnfilteredTableMetadata`，如果它是分区表，则查询引擎会调用 `GetUnfilteredPartitionsMetadata` 以从数据目录中检索元数据和策略信息。
3. Lake Formation 为请求执行授权。如果用户对该表没有适当的权限，则会AccessDeniedException被抛出。
4. 作为请求的一部分，查询引擎会发送它支持的筛选。可以在数组中发送以下两个标志：COLUMN_PERMISSIONS 和 CELL_FILTER_PERMISSION。如果查询引擎不支持这些功能中的任何一个，并且表上存在该功能的策略，则会抛出，查询失败。PermissionTypeMismatchException这是为了避免数据泄露。
5. 返回的响应包含以下内容：
   - 表的整个架构，以便查询引擎可以使用它来解析存储中的数据。
   - 用户有权访问的授权列的列表。如果授权列的列表为空，则表示用户拥有 `DESCRIBE` 权限，但没有 `SELECT` 权限，查询将失败。
   - 一个标志 `IsRegisteredWithLakeFormation`，用于指示 Lake Formation 是否可以针对此资源数据售卖凭证。如果返回 false，则应使用客户的凭证来访问 Amazon S3。
   - 应该应用于各行数据的 `CellFilters` 的列表。此列表包含列和用于评估每一行的表达式。只有在将 CCELL_FILTER_PERMISSION 作为请求的一部分发送，并且有适用于调用用户的表数据筛选条件时，才会填充此字段。

6. 检索到元数据后，查询引擎会调用 `GetTemporaryGlueTableCredentials` 或 `GetTemporaryGluePartitionCredentials` 以获取 AWS 凭证，从而从 Amazon S3 位置检索数据。
7. 查询引擎从 Amazon S3 读取相关对象，根据在第 2 步中收到的策略筛选数据，并将结果返回给用户。

Lake Formation 的应用程序集成 API 操作包含用于配置与第三方查询引擎的集成的其他内容。您可以在凭证售卖 API 操作一节查看操作详细信息。

注册第三方查询引擎

在第三方查询引擎使用应用程序集成 API 操作之前，您需要显式启用相应的权限，以便查询引擎能够代表您调用 API 操作。只需几个步骤即可完成此操作：

1. 您需要指定 AWS 账户和 IAM 会话标签，它们需要相应的权限以便通过 AWS Lake Formation 控制台、AWS CLI 或 API/SDK 调用应用程序集成 API 操作。
2. 当第三方查询引擎在您的账户中担任执行角色时，查询引擎必须附加一个在 Lake Formation 中注册的代表第三方引擎的会话标签。Lake Formation 使用该标签验证请求是否来自经批准的引擎。有关会话标签的更多信息，请参阅《IAM 用户指南》中的会话标签。
3. 设置第三方查询引擎执行角色时，您必须拥有 IAM 策略中的以下一组最低权限：

```
{
  "Version": "2012-10-17",
  "Statement": {"Effect": "Allow",
    "Action": [
      "lakeformation:GetDataAccess",
      "glue:GetTable",
      "glue:GetTables",
      "glue:GetDatabase",
      "glue:GetDatabases",
      "glue:CreateDatabase",
      "glue:GetUserDefinedFunction",
      "glue:GetUserDefinedFunctions",
      "glue:GetPartition",
      "glue:GetPartitions"
    ],
    "Resource": "*"
  }
}
```

4. 针对查询引擎执行角色设置角色信任策略，以便该角色可以对将哪个会话标签键值对附加到该角色进行精细访问控制。在以下示例中，只允许该角色附加会话标签键 "LakeFormationAuthorizedCaller" 和会话标签值 "engine1"，不允许附加其他会话标签键值对。

```
{
    "Sid": "AllowPassSessionTags",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
        "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"        }
    }
}
```

当LakeFormationAuthorizedCaller调用 STS: AssumeRole API 操作来获取证书供查询引擎使用时，AssumeRole 请求中必须包含会话标签。返回的临时凭证可用于发出 Lake Formation 应用程序集成 API 请求。

Lake Formation 应用程序集成 API 操作要求调用主体担任 IAM 角色。IAM 角色必须包含已在 Lake Formation 中注册的具有预定值的会话标签。该标签使 Lake Formation 可以验证用于调用应用程序集成 API 操作的角色是否被允许这样做。

为第三方查询引擎启用调用应用程序集成 API 操作所需的权限

按照以下步骤操作，以便允许第三方查询引擎通过 AWS Lake Formation 控制台、AWS CLI 或 API/SDK 调用应用程序集成 API 操作。

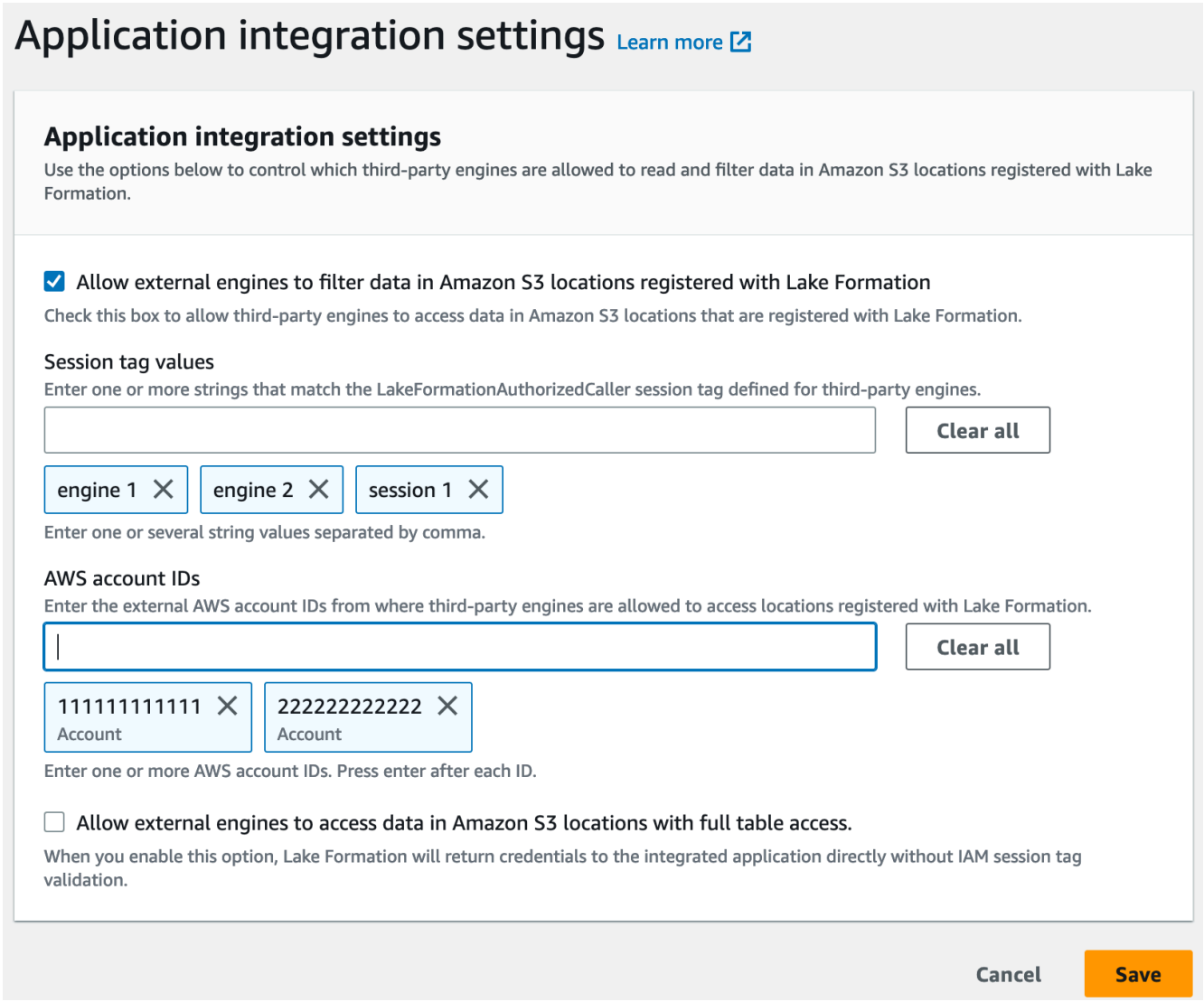
Console

注册您的账户以进行外部数据筛选：

1. 登录 AWS Management Console 并打开 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/)。
2. 在左侧导航栏中，展开权限，然后选择应用程序集成设置。

3. 在应用程序集成设置页面上，选择允许外部引擎筛选在 Lake Formation 中注册的 Amazon S3 位置处的数据。
4. 输入您为第三方引擎创建的会话标签。有关会话标签的信息，请参阅《AWS Identity and Access Management 用户指南》中的在 AWS STS 中传递会话标签。
5. 输入可使用第三方引擎访问未经过筛选的元数据信息的用户的账户 ID 以及当前账户中资源的数据访问凭证。

   您也可以使用 AWS 账户 ID 字段配置跨账户访问权限。



CLI

使用 `put-data-lake-settings` CLI 命令可设置以下参数。

使用此 AWS CLI 命令时，需要配置以下三个字段：

- allow-external-data-filtering – (布尔值）表示第三方引擎可以访问未经过筛选的元数据信息和当前账户中资源的数据访问凭证。
- external-data-filtering-allow-list – (数组）可以使用第三方引擎访问未经过筛选的元数据信息和当前账户中资源的数据访问凭证的账户 ID 的列表。
- authorized-sessions-tag-value-list – (数组）授权会话标签值（字符串）的列表。如果 IAM 角色凭证附加了授权键值对，那么，如果会话标签包含在列表中，则会授予会话访问未经过筛选的元数据信息和已配置账户中资源的数据访问凭证的权限。授权会话标签键为 *LakeFormationAuthorizedCaller*。
- AllowFullTableExternalDataAccess – (布尔值）当调用者拥有完整的数据访问权限时，是否允许第三方查询引擎在没有会话标签的情况下获取数据访问凭证。

例如：

```
aws lakeformation put-data-lake-settings --cli-input-json file://
datalakesettings.json

{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/lakeAdmin"
      }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "TrustedResourceOwners": [],
    "AllowExternalDataFiltering": true,
    "ExternalDataFilteringAllowList": [
        {"DataLakePrincipalIdentifier": "111111111111"}
        ],
    "AuthorizedSessionTagValueList": ["engine1"]
    }
    "AllowFullTableExternalDataAccess": false

}
```

API/SDK

使用 PutDataLakeSetting API 操作可设置以下参数。

使用此 API 操作时，需要配置以下三个字段：

- AllowExternalDataFiltering – （布尔值）表示第三方引擎可以访问未经过筛选的元数据信息和当前账户中资源的数据访问凭证。
- ExternalDataFilteringAllowList – （数组）可以使用第三方引擎访问未经过筛选的元数据信息和当前账户中资源的数据访问凭证的账户 ID 的列表。
- AuthorizedSectionsTagValueList – （数组）授权标签值（字符串）的列表。如果 IAM 角色凭证附加了授权标签，则会向会话授予访问未经过筛选的元数据信息和已配置账户中资源的数据访问凭证的权限。授权会话标签键为 *LakeFormationAuthorizedCaller*。
- AllowFullTableExternalDataAccess – （布尔值）当调用者拥有完整的数据访问权限时，是否允许第三方查询引擎在没有会话标签的情况下获取数据访问凭证。

例如：

```
//Enable session tag on existing data lake settings
public void sessionTagSetUpForExternalFiltering(AWSLakeFormationClient
 lakeformation) {
    GetDataLakeSettingsResult getDataLakeSettingsResult =
 lfClient.getDataLakeSettings(new GetDataLakeSettingsRequest());
    DataLakeSettings dataLakeSettings =
 getDataLakeSettingsResult.getDataLakeSettings();

    //set account level flag to allow external filtering
    dataLakeSettings.setAllowExternalDataFiltering(true);

    //set account that are allowed to call credential vending or Glue
 GetFilteredMetadata API
    List<DataLakePrincipal> allowlist = new ArrayList<>();
    allowlist.add(new
 DataLakePrincipal().withDataLakePrincipalIdentifier("111111111111"));
    dataLakeSettings.setWhitelistedForExternalDataFiltering(allowlist);

    //set registered session tag values
    List<String> registeredTagValues = new ArrayList<>();
    registeredTagValues.add("engine1");
    dataLakeSettings.setAuthorizedSessionTagValueList(registeredTagValues);
```

```
    lakeformation.putDataLakeSettings(new
 PutDataLakeSettingsRequest().withDataLakeSettings(dataLakeSettings));
}
```

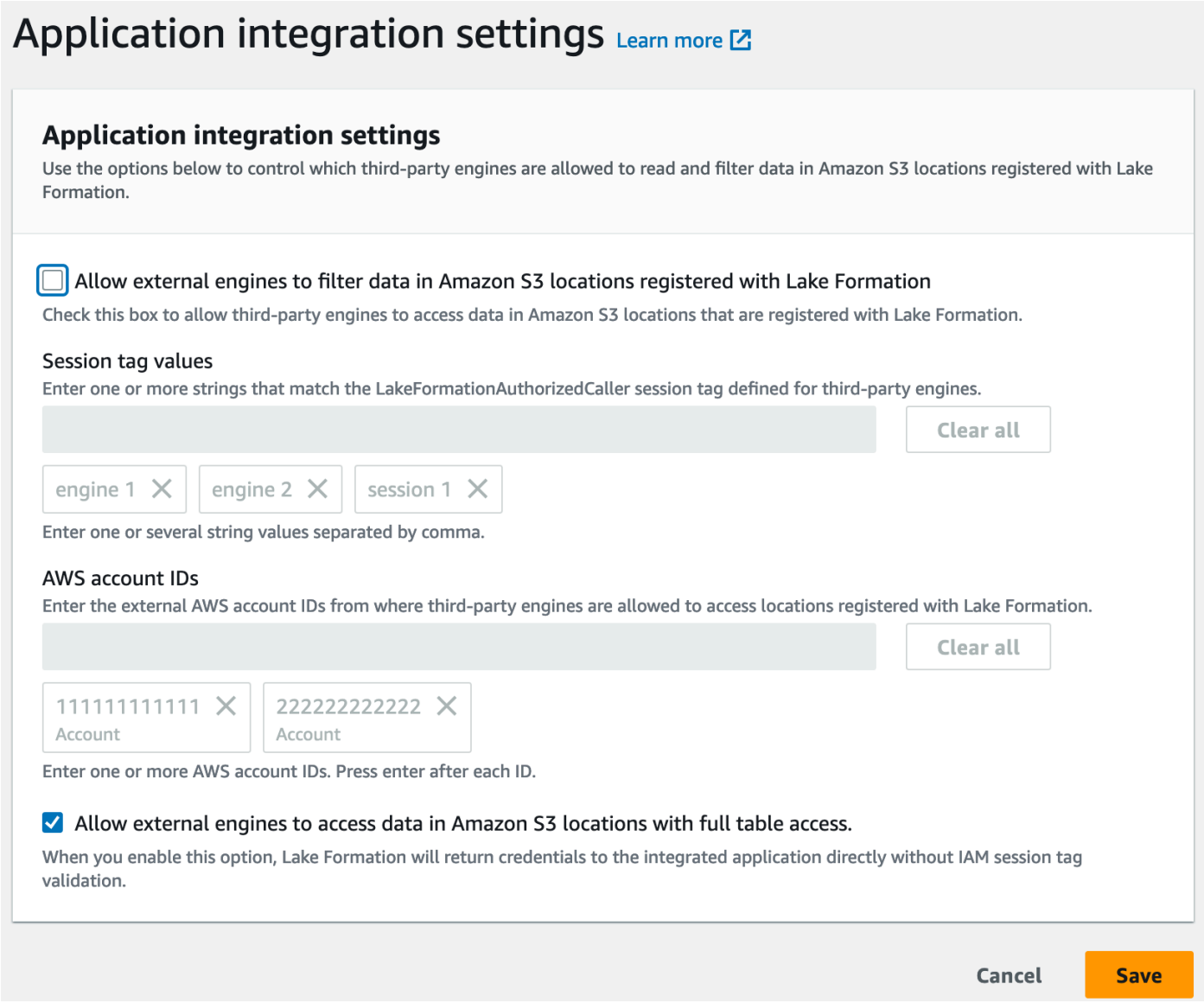
集成应用程序以获取完整表访问权限

按照以下步骤操作，使第三方查询引擎无需验证 IAM 会话标签即可访问数据：

Console

1. 通过以下网址登录 Lake Formation 控制台：https://console.aws.amazon.com/lakeformation/。
2. 在左侧导航栏中，展开权限，然后选择应用程序集成设置。
3. 在应用程序集成设置页面上，选中允许外部引擎以完整表访问权限访问 Amazon S3 位置的数据。

   启用此选项后，Lake Formation 将直接向查询应用程序返回凭证，而无需验证 IAM 会话标签。



AWS CLI

使用 `put-data-lake-settings` CLI 命令设置 `AllowFullTableExternalDataAccess` 参数。

```
aws lakeformation put-data-lake-settings –cli-input-json file://put-data-lake-
settings.json –region ap-northeast-1
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/
lakeAdmin"
            }
```

```
        ],
        "AllowFullTableExternalDataAccess": true
    }
}
```

使用其他 AWS 服务

AWS 亚马逊 Athena AWS Glue、Amazon Redshift Spectrum 和亚马逊 EMR 等服务可以使用 Lake Formation 安全地访问在 Lake Formation 注册的亚马逊 S3 地点的数据。使用 Lake Formation，您可以为 AWS Glue Data Catalog中的数据定义和管理细粒度访问控制 (FGAC) 权限。这些 AWS 服务中的每一项都是 Lake Formation 的可信调用者，而 Lake Formation 通过临时证书提供对存储在 Amazon S3 中的数据的访问权限。有关更多信息，请参阅Lake Formation 应用程序集成的工作原理。

为了利用这些功能，Lake Formation 要求您首先注册 Amazon S3 位置，然后向 IAM 主体分配用于访问表、数据库和 Amazon S3 位置的相应权限。有关更多信息，请参阅管理 Lake Formation 权限。

主题

在亚马逊 A AWS Lake Formation thena 上使用

Amazon Athena 是一种无服务器查询服务，可帮助您分析存储在 Amazon S3 中的结构化、半结构化和非结构化数据。Athena 支持以 CSV、JSON、Parquet 和 Avro 数据格式查询数据。Athena 还支持 Apache Hive、Apache Hudi、Apache Iceberg 和 Lake Formation 受管控表等表格格式。Athena 与集成，可将您的数据集 AWS Glue Data Catalog 的元数据存储在 Amazon S3 中。Athena 可以使用 Lake Formation 来定义和维护对这些数据集的访问控制策略。

以下是一些常见使用案例，您可以在其中将 Lake Formation 与 Athena 一起使用。

- 从 Athena 中使用 Lake Formation 权限访问数据目录资源（数据库和表）。您可以使用命名资源方法或 LF 标签来定义对数据库和表的权限。有关更多信息，请参阅：
  - 使用命名资源方法授予数据库权限
  - Lake Formation 基于标签的访问控制

> **Note**
>
> Lake Formation 权限仅在使用 Athena 查询 Amazon S3 中源数据和数据目录中元数据时应用。

Lake Formation 权限支持对数据库和表执行读取和写入操作。

> **Note**
>
> 当您使用 LF 标签管理对数据目录资源的权限时，无法应用数据筛选条件。

- 通过在列、行和单元格级别授予权限来使用 Lake Formation 中的数据筛选条件 保护 Amazon S3 数据湖中的表，从而控制查询结果。请参阅《Amazon Athena 用户指南》中的分区投影限制。
- 运行联合查询时，对基于 SAML 的 Athena 用户可以使用的数据实施细粒度访问控制。

  Athena JDBC 和 ODBC 驱动程序支持使用基于 SAML 的身份提供者 (IdP) 配置对数据来源的联合访问权限。使用与 Lake Formation QuickSight 集成的亚马逊与您的现有 IAM 角色或 SAML 用户或群组，以可视化 Athena 的查询结果。

  > **Note**
  >
  > 仅当您使用 JDBC 或 ODBC 驱动程序向 Athena 提交查询时，才应用 SAML 用户和组的 Lake Formation 权限。

  有关更多信息，请参阅使用 Lake Formation 以及 Athena JDBC 和 ODBC 驱动程序对 Athena 进行联合访问。

  > **Note**
  >
  > 目前，以下区域不支持在 Lake Formation 中授权访问 SAML 身份：
  >
  > - 中东（巴林）- me-south-1
  > - 亚太地区（香港）- ap-east-1
  > - 非洲（开普敦）- af-south-1
  > - 中国（宁夏）- cn-northwest-1

- 亚太地区（大阪）– ap-northeast-3

- 使用Lake Formation 中的跨账户数据共享查询其他账户中的表。

> **Note**
>
> 有关使用 `Views` 这一 Lake Formation 权限时的限制的更多信息，请参阅注意事项和限制。

支持事务表格格式

通过应用 Lake Formation 权限，您可以保护基于 Amazon S3 的数据湖中的事务数据。下表列出了 Athena 和 Lake Formation 权限支持的事务表格格式。当 Athena 用户运行查询时，Lake Formation 会强制实施这些权限。

表格式	描述和允许的操作	Athena 支持的 Lake Formation 权限
Apache Hudi	一种用于简化增量数据处理和数据管线开发的格式。 Athena 支持使用 Apache Hudi 表格格式对写时复制 (CoW) 和读时合并 (MoR) Hudi 表格类型的 Amazon S3 数据集执行创建和读取操作。Athena 不支持对 Hudi 表执行写入操作。 使用 Athena 查询 Hudi 数据集。	通过Lake Formation 中的数据筛选和单元格级别安全性使用表、列、行和单元格级别权限保护 Hudi 表。
Apache Iceberg	一种开放表格格式，它将大量文件作为表进行管理，并且支持现代分析数据湖操作，例如记录级别插入、更新、删除和时间旅行查询。	支持表、列、行和单元格级别权限。目前，Lake Formation 不支持管理对采用开放表格格式的表的写入操作（例如 `VACUUM`、`MERGE`、`UPDATE` 和 `OPTIMIZE`）权限。

表格式	描述和允许的操作	Athena 支持的 Lake Formation 权限
	有关 Athena 对 Iceberg 表的支持的更多信息，请参阅使用 Iceberg 表。	
Linux Foundation Delta Lake	Delta Lake 是一个开源项目，可帮助实施通常在 Amazon S3 或 Hadoop Distributed File System (HDFS) 上构建的现代数据湖架构。 Athena 支持在 Delta Lake 表上使用基于符号链接的清单表定义 AWS Glue Data Catalog 创建的 Delta 湖表。 有关更多信息，请参阅使用 AWS Glue 爬虫抓取 Delta Lake 表。 Athena（引擎版本 3）支持读取原生 Delta Lake 表。 有关更多信息，请参阅通过 AWS Glue 爬虫引入原生 Delta Lake 表格支持。	符号链接表和原生 Delta Lake 表支持表、列、行和单元格级别权限。

其他 资源

博客文章、视频和研讨会

- 使用 Amazon Athena 查询 Amazon S3 数据湖中的 Apache Hudi 数据集
- 使用亚马逊 Athena、亚马逊 EMR 和 AWS Glue
- 使用 Athena 和 Apache Iceberg 在 Amazon S3 上插入、更新、删除
- 基于 LF 标签的访问控制 有关查询数据湖的 Lake Formation 研讨会。

AWS Lake Formation 与亚马逊 Redshift Spectrum 一起使用

Amazon Redshift Spectrum 使您可以查询和检索 Amazon S3 数据湖中的数据，而不必将数据加载到 Amazon Redshift 集群节点中。

Redshift Spectrum 支持两种注册启用了 Lake Formation 的外部 AWS Glue 数据目录的方法。

- 使用附加了集群的且有权访问数据目录的 IAM 角色

  要创建 IAM 角色，请按照以下过程中概述的步骤操作。

  使用启用的为 Amazon Redshift 创建 IA AWS Glue Data Catalog M 角色 AWS Lake Formation
- 使用为管理对外部 AWS Glue Data Catalog 资源的访问而配置的联合 IAM 身份。

  Redshift Spectrum 支持使用联合 IAM 身份查询 Lake Formation 表。IAM 身份可以是 IAM 用户或 IAM 角色。有关 Redshift Spectrum 中 IAM 身份联合验证的更多信息，请参阅使用联合身份管理 Amazon Redshift 对本地资源和 Amazon Redshift 外部表的访问权限。

利用 Lake Formation 与 Redshift Spectrum 的集成，您可以在将数据注册到 Lake Formation 后定义对表的行、列和单元格级别访问控制权限。

有关更多信息，请参阅将 Redshift 频谱与配合使用。 AWS Lake Formation

Redshift Spectrum 支持对 Lake Formation 托管的外部架构表执行读取或 SELECT 查询。

有关更多信息，请参阅为 Redshift Spectrum 创建外部架构。

支持事务表类型

下表列出了 Redshift Spectrum 中支持的事务表格格式以及适用的 Lake Formation 权限。

支持的表格格式

表格式	描述和允许的操作	Redshift Spectrum 支持的 Lake Formation 权限
Apache Hudi	一种用于简化增量数据处理和数据管线开发的格式。 Redshift Spectrum 支持在 Amazon S3 上使用 Apache	通过Lake Formation 中的数据筛选和单元格级别安全性使用表、列、行和单元格级别权限保护 Hudi 表。

表格式	描述和允许的操作	Redshift Spectrum 支持的 Lake Formation 权限
	Hudi 写时复制 (CoW) 表格格式执行插入、删除和更新写入操作。 有关更多信息，请参阅为 Apache Hudi 中管理的数据创建外部表。	
Apache Iceberg	一种开放表格格式，它将大量文件作为表进行管理，并且支持现代分析数据湖操作，例如记录级别插入、更新、删除和时间旅行查询。 有关更多信息，请参阅将 Apache Iceberg 表与 Amazon Redshift 搭配使用。	Redshift Spectrum 支持使用 Apache Iceberg 表进行查询。
Linux Foundation Delta Lake	Delta Lake 是一个开源项目，可帮助实施通常在 Amazon S3 或 Hadoop Distributed File System (HDFS) 上构建的现代数据湖架构。 Redshift Spectrum 支持查询 Delta Lake 表。有关更多信息，请参阅为 Apache Hudi 中托管的数据创建外部表。	支持表、列、行和单元格级别权限。

其他 资源

博客文章和研讨会

- 使用 Amazon Redshift Spectrum 集中管理您的数据湖， AWS Lake Formation 同时启用现代数据架构

- 使用 Redshift Spectrum 查询 Amazon S3 数据湖中的 Apache HUDI 写时复制 (CoW) 表

AWS Lake Formation 与一起使用 AWS Glue

数据工程师和 DevOps 专业人员使用 AWS Glue 带有 Apache Spark 的提取、转换和加载 (ETL)，在 Amazon S3 中对其数据集进行转换，并将转换后的数据加载到数据湖和数据仓库中，用于分析、机器学习和应用程序开发。由于会有不同的团队访问 Amazon S3 中的相同数据集，因此必须根据其角色授予和限制权限。

AWS Lake Formation 是在此基础上构建的 AWS Glue，并且服务通过以下方式进行交互：

- Lake Formation 和 AWS Glue 共享同一数据目录。
- 以下 Lake Formation 控制台功能可以调用 AWS Glue 控制台：
  - 作业 – 有关更多信息，请参阅《AWS Glue 开发人员指南》中的添加作业。
  - 爬网程序 – 有关更多信息，请参阅《AWS Glue 开发人员指南》中的使用爬网程序编录数据。
- 使用 Lake Formation 蓝图时生成的工作流是 AWS Glue 工作流。您可以在 Lake Formation 控制台和 AWS Glue 控制台中查看和管理这些工作流。
- 机器学习转换功能在 Lake Formation 中提供，并且是针对 AWS Glue API 操作构建的。您可以在 AWS Glue 控制台上创建和管理机器学习转换功能。有关更多信息，请参阅《AWS Glue 开发人员指南》中的机器学习转换。

您可以使用 Lake Formation 细粒度访问控制来管理现有的数据目录资源和 Amazon S3 数据位置。

> **Note**
>
> AWS Glue 在从底层 Amazon S3 位置获取数据时，ETL 需要对整个表具有完全访问权限。AWS Glue 如果您对表应用列级权限，ETL 作业就会失败。但是，您可以通过定义数据筛选条件来创建列级别和行级别安全性。有关更多信息，请参阅 有关列级别筛选的注意事项和限制 Lake Formation 评估表中定义的数据筛选器，并仅从 Amazon S3 中检索 AWS Glue ETL 任务所需的筛选数据。

支持事务表类型

通过应用 Lake Formation 权限，您可以保护基于 Amazon S3 的数据湖中的事务数据。下表列出了中支持的交易表格式 AWS Glue 和 Lake Formation 权限。Lake Formation 强制执行这些 AWS Glue 操作权限。

支持的表格格式

表格式	描述和允许的操作	中支持 Lake Formation 权限 AWS Glue
Apache Hudi	一种开放表格格式，用于简化增量数据处理和数据管线开发。 有关示例，请参阅中的 “使用 Hudi 框架”。 AWS Glue	表级权限可用于 Hudi 表。 有关更多信息，请参阅限制。
Apache Iceberg	一种开放表格格式，可将大量文件作为表进行管理。 有关示例，请参阅中的使用 Iceberg 框架。 AWS Glue	表级权限可用于 Iceberg 表。 有关更多信息，请参阅限制。
Linux Foundation Delta Lake	Delta Lake 是一个开源项目，可帮助实施通常在 Amazon S3 或 Hadoop Distributed File System (HDFS) 上构建的现代数据湖架构。 有关示例，请参阅中的使用 Delta Lake 框架 AWS Glue。	表级权限可用于 Delta Lake 表。 有关更多信息，请参阅限制。

其他 资源

博客文章和存储库

- 使用 AWS Glue 连接器读写带有 ACID 事务的 Apache Iceberg 表，并执行时空旅行
- 使用 AWS Glue 自定义连接器写入 Apache Hudi 表

- AWS Cloudformation 模板和 pyspark 代码示例存储库，用于使用 AWS Glue Apache Hudi 和 Amazon S3 分析流数据。

AWS Lake Formation 与 Amazon EMR 一起使用

Amazon EMR 是一个灵活的 AWS 托管集群平台，您可以在支持的大数据框架（例如 Hadoop Map-Reduce、Spark、Hive、Presto 等）上运行任何自定义代码。组织还可以使用 Amazon EMR 在高度分布式集群中运行批处理和流式数据处理应用程序。使用 Amazon EMR，您可以在其权限由 Lake Formation 管理的数据库和表上运行数据转换和自定义代码。

Amazon EMR 部署选项有三个：

- EC2 上的 EMR
- EMR Serverless
- Amazon EMR on EKS

有关更多信息，请参阅将 Amazon EMR 与 Lake Formation 集成或将 EMR Serverles s 与 Lake Formation 集成，实现精细的访问控制 AWS Lake Formation

支持事务表格格式

当您使用 Spark SQL 读取和写入数据时，Amazon EMR 发行版 6.15.0 及更高版本支持对 Apache Hudi、Apache Iceberg 和 Delta Lake 表格式的 Lake Formation 表、行、列和单元格级别的访问控制权限。

支持的表格格式

表格式	描述和允许的操作	Amazon EMR 中支持的 Lake Formation 权限
Apache Hudi	一种开放表格格式，用于简化增量数据处理和数据管线开发。 有关支持的操作列表，请参阅 Apache Hudi 和 Lake Formation。	Amazon EMR 支持使用 Apache Hudi 进行表、行、列和单元格级别的访问控制。

表格式	描述和允许的操作	Amazon EMR 中支持的 Lake Formation 权限
Apache Iceberg	一种开放表格格式，可将大量文件作为表进行管理。 有关支持的操作列表，请参阅 Apache Iceberg 和 Lake Formation。	Amazon EMR 支持使用 Apache Iceberg 进行表、行、列和单元格级别的访问控制。
Linux Foundation Delta Lake	Delta Lake 是一个开源项目，可帮助实施通常在 Amazon S3 或 Hadoop Distributed File System (HDFS) 上构建的现代数据湖架构。 有关支持的操作列表，请参阅 Delta Lake 和 Lake Formation。	Amazon EMR 支持使用 Delta Lake 表进行表、行、列和单元级别的访问控制。

其他 资源

用户指南、博客文章和研讨会

- 使用运行时角色与 Amazon EMR 集成
- 利用 EKS 上的 Amazon EMR 快速开始使用 Apache Hudi、Apache Iceberg 和 Delta Lake
- 将 Delta Lake OSS 与 EMR Serverless 结合使用

在 Amazon AWS Lake Formation 上使用 QuickSight

亚马逊 QuickSight 支持使用 Athena 在亚马逊 S3 中浏览由 Lake Formation 权限管理的数据集。

亚马逊的标准版和企业版用户都 QuickSight 与 Lake Formation 集成，但略有不同。

- 企业版 — 向个人 Amazon QuickSight 用户、群组和 IAM 角色授予访问数据库和表格的精细访问控制 (FGAC) 权限。
- 标准版 — 向 IAM 角色授予访问数据库和表的权限。

> **Note**
>
> 默认情况下，Amazon QuickSight 使用名为的角色`aws-quicksight-service-role-v0`。您还可以定义具有所需权限的自定义角色，使亚马逊 QuickSight 能够访问 Athena。

有关更多信息，请参阅通过以下方式对连接进行授权 AWS Lake Formation

其他 资源

博客文章

- 在中为 Ama QuickSight zon 作者启用细粒度权限 AWS Lake Formation
- 使用 AWS Lake Formation 和 Amazon 安全地分析您的数据 QuickSight

与 La AWS CloudTrail ke AWS Lake Formation 一起使用

AWS CloudTrail Lake 支持在中使用精细 Amazon Athena 权限浏览事件数据存储。 AWS Lake Formation

> **Note**
>
> CloudTrail 只能通过 Amazon Athena查询湖泊。

要向 CloudTrail Lake Formation 注册您的 Lake 事件数据存储，请参阅合并事件数据存储。

使用 AWS CloudTrail 记录 AWS Lake Formation API 调用

AWS Lake Formation 与 AWS CloudTrail 集成，后者是一项服务，可用于记录 Lake Formation 中由用户、角色或 AWS 服务所执行的操作。CloudTrail 将所有 Lake Formation API 调用作为事件进行捕获。捕获的调用包括来自 Lake Formation 控制台的调用、来自 AWS Command Line Interface 的调用，以及针对Lake Formation API 操作的代码调用。如果您创建跟踪记录，则可以使 CloudTrail 事件能够持续传输到 Amazon S3 存储桶（包括 Lake Formation 的事件）。如果您不配置跟踪，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。使用通过 CloudTrail 收集的信息，您可以确定向 Lake Formation 发出了什么请求、发出请求的 IP 地址、请求方、请求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 Lake Formation 信息

当您创建新的 AWS 账户时，将默认启用 CloudTrail。当 Lake Formation 中发生活动时，会将该活动作为 CloudTrail 事件与其他 AWS 服务事件一起记录在事件历史记录中。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、操作的日期和时间以及请求参数等方面的信息。此外，每个事件或日志条目都包含有关生成请求的人员的信息。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 AWS Identity and Access Management（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。
- 请求是否由其它 AWS 服务发出。

有关更多信息，请参阅 CloudTrail userIdentity 元素。

您可以查看、搜索和下载 AWS 账户的最新事件。有关更多信息，请参阅使用 CloudTrail 事件历史记录查看事件。

要持续记录 AWS 账户中的事件（包括 CloudFront 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。在控制台创建跟踪时，跟踪默认应用于所有 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（如 Amazon Athena），进一步分析在 CloudTrail 日志中收集的事件数据并采取行动。CloudTrail 还可以将日志文件传输到 Amazon CloudWatch Logs 和 CloudWatch Events。

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 日志文件和从多个账户接收 CloudTrail 日志文件

了解 Lake Formation 事件

所有 Lake Formation API 操作都会被 CloudTrail 记录下来，详见《AWS Lake Formation 开发人员指南》。例如，对 PutDataLakeSettings、GrantPermissions 和 RevokePermissions 操作的调用会在 CloudTrail 日志文件中生成条目。

以下示例显示了关于 GrantPermissions 操作的 CloudTrail 事件。该条目包括已授予权限的用户 (datalake_admin)、被授予权限的主体 (datalake_user1) 以及所授予的权限 (CREATE_TABLE)。该条目还显示授权失败，原因是 resource 参数中未指定目标数据库。


```
{
    "eventVersion": "1.08",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "AIDAZKE67KM3P775X74U2",
        "arn": "arn:aws:iam::111122223333:user/datalake_admin",
        "accountId": "111122223333",
        "accessKeyId": "...",
        "userName": "datalake_admin"
    },
    "eventTime": "2021-02-06T00:43:21Z",
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GrantPermissions",
    "awsRegion": "us-east-1",
    "sourceIPAddress": "72.21.198.65",
    "userAgent": "aws-cli/1.19.0 Python/3.6.12
 Linux/4.9.230-0.1.ac.223.84.332.metal1.x86_64 botocore/1.20.0",
    "errorCode": "InvalidInputException",
    "errorMessage": "Resource must have one of the have either the catalog, table or
 database field populated.",
    "requestParameters": {
        "principal": {
            "dataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
        },
```

```
        "resource": {},
        "permissions": [
            "CREATE_TABLE"
        ]
    },
    "responseElements": null,
    "requestID": "b85e863f-e75d-4fc0-9ff0-97f943f706e7",
    "eventID": "8d2ccef0-55f3-42d3-9ede-3a6faedaa5c1",
    "readOnly": false,
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "eventCategory": "Management",
    "recipientAccountId": "111122223333"
}
```

以下示例显示了有关 GetDataAccess 操作的 CloudTrail 日志条目。主体不会直接调用此 API。相反，每当主体或集成 AWS 服务请求临时凭证以访问已注册到 Lake Formation 的数据湖位置中的数据时，都会将 GetDataAccess 记录在日志中。

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

另请参阅

- 跨账户 CloudTrail 日志

Lake Formation 最佳实践、注意事项和限制

使用本部分可快速查找 AWS Lake Formation中的最佳实践、注意事项和限制。

请参阅服务限额，了解您的 AWS 账户使用的服务资源或操作的最大数量。

主题

跨账户数据共享最佳实践和注意事项

Lake Formation 跨账户功能允许用户安全地跨多个 AWS 组织共享分布式数据湖 AWS 账户，或者直接与其他账户中的 IAM 委托人共享分布式数据湖，从而提供对数据目录元数据和底层数据的精细访问权限。

使用 Lake Formation 跨账户数据共享时，请考虑以下最佳实践：

- 您可以向自己 AWS 账户中的委托人授予的 Lake Formation 权限数量没有限制。但是，Lake For AWS Resource Access Manager mation 使用 (AWS RAM) 容量进行跨账户授权，您的账户可以使用指定的资源方法进行跨账户授权。要最大限度地提高 AWS RAM 容量，请遵循指定资源方法的以下最佳实践：
  - 使用新的跨账户授予模式（跨账户版本设置下的版本 3 及更高版本）与外部 AWS 账户用户共享资源。有关更多信息，请参阅更新跨账户数据共享版本设置。
  - 将 AWS 账户整理到组织中，并向组织或组织单位授予权限。每向组织或组织单位进行一次授权都计为一次授权。

授予组织或组织单位还无需接受 AWS Resource Access Manager (AWS RAM) 资源共享邀请即可获得授权。有关更多信息，请参阅访问和查看共享数据目录表和数据库。

- 可以使用特殊的所有表 通配符来授予对数据库中所有表的权限，而不必授予对数据库中许多单独表的权限。每针对所有表授予一次权限都计为一次授权。有关更多信息，请参阅授予和撤销对数据目录资源的权限。

> **Note**
>
> 有关请求提高资源共享数量限制的更多信息 AWS RAM，请参阅中的AWS 服务配额AWS 一般参考。

- 您必须创建指向共享数据库的资源链接，该数据库才能出现在 Amazon Athena 和 Amazon Redshift Spectrum 查询编辑器中。同样，为了能够使用 Athena 和 Redshift Spectrum 查询共享表，您必须创建指向这些表的资源链接。之后，资源链接便会出现在查询编辑器的表列表中。

  您可以使用所有表通配符来授予对数据库中所有表的权限，而不必为许多要查询的单独表创建资源链接。然后，当您为该数据库创建资源链接并在查询编辑器中选择该数据库资源链接时，您将可以访问该数据库中的所有表进行查询。有关更多信息，请参阅创建资源链接。
- 当您直接与其他账户中的主体共享资源时，接收方账户中的 IAM 主体可能无权创建资源链接，因而无法使用 Athena 和 Amazon Redshift Spectrum 查询共享表。数据湖管理员可以创建占位符数据库并向 `ALLIAMPrincipal` 组授予 `CREATE_TABLE` 权限，而不必为每个共享的表创建资源链接。然后，接收方账户中的所有 IAM 主体均可以在占位符数据库中创建资源链接并开始查询共享表。

  请参阅使用命名资源方法授予数据库权限中有关向 `ALLIAMPrincipals` 授予权限的 CLI 命令示例。
- Athena 和 Redshift Spectrum 支持列级别访问控制，但仅限于包含，不支持排除。AWS Glue ETL 作业不支持列级别访问控制。
- 当资源与您的 AWS 账户共享时，您只能向账户中的用户授予该资源的权限。您不能向其他 AWS 账户、组织（甚至不是您自己的组织）或`IAMAllowedPrincipals`群组授予该资源的权限。
- 您不能向外部账户授予对数据库的 `DROP` 或 `Super` 权限。
- 在删除数据库或表之前，请撤销跨账户权限。否则，您必须删除中的孤立资源共享。 AWS Resource Access Manager

另请参阅

- Lake Formation 基于标签的访问控制最佳实践和注意事项
- 有关跨账户访问的更多规则和限制，请参阅Lake Formation 权限参考中的CREATE_TABLE。

跨区域数据访问限制

Lake Formation 支持跨 AWS 区域查询数据目录表。您可以使用 Amazon Athena Amazon EMR 和 AWS Glue ETL 访问其他区域中的数据，方法是在其他区域中创建指向源数据库和表的资源链接。通过跨区域表访问功能，您可以跨区域访问数据，而无需将基础数据或元数据复制到数据目录中。

以下限制适用于跨区域表访问。

- Lake Formation 不支持使用 Amazon Redshift Spectrum 从其他地区查询数据目录表。
- 在 Lake Formation 控制台中，数据库和表视图不显示源区域数据库/表名称。
- 要从其他区域查看共享数据库下表的列表，您需要先创建指向共享数据库的资源链接，选择该资源链接，然后选择查看表。
- 当您在其中创建指向共享数据库和在 opt in Regions 中创建的表的资源链接时 AWS 区域 ，跨区域表访问功能不起作用。

  有关更多信息，请参阅支持 AWS 区域 和服务页面上的选择加入区域。
- Lake Formation 不支持 SAML 用户进行的跨区域资源链接调用。

数据目录视图注意事项和限制

在中 AWS Glue Data Catalog，视图是一个虚拟表，其中的内容由引用一个或多个表的查询定义。您可以使用适用于 Amazon Athena 或 Amazon Redshift 的 SQL 编辑器创一个最多引用 10 个表的视图。视图的基础引用表可以属于同一数据库，也可以属于同一 AWS 账户内的不同数据库。

以下注意事项和限制适用于数据目录视图。

- Amazon Redshift 始终从带有字符串的表创建包含 varchar 列的视图。从其他引擎添加方言时，您必须将字符串列转换为具有显式长度的 varchar。
- 向数据库内的 `All views` 授予数据湖权限将导致被授权者具有对数据库内所有表和视图的权限。

- 您无法创建视图：
  - 这引用了其他视图。
  - 当引用表是资源链接时。
  - 当引用表具有 IAM_ALLOWED_GROUP 主体权限时。
  - 当引用表位于另一个账户中时。
  - 来自外部 Hive 元存储。

数据筛选限制

当您授予对数据目录表的 Lake Formation 权限时，可以包括数据筛选规范，以限制对查询结果中以及与 Lake Formation 集成的引擎中某些数据的访问。Lake Formation 使用数据筛选来实现列级别安全性、行级别安全性以及单元格级别安全性。如果源数据包含嵌套结构，则可以对嵌套列定义和应用数据筛选条件。

请牢记以下关于行级别筛选和单元格级别筛选的注意事项和限制。

- 嵌套列不支持单元格级别安全性。
- 嵌套列也支持顶级列支持的所有表达式。但是，定义嵌套的行级别表达式时，不应引用分区列下的嵌套字段。
- 使用 Athena 引擎版本 3 或 Amazon Redshift Spectrum 时，所有区域均可提供单元格级别安全性。对于其他服务，只有支持的区域上提及的区域提供单元格级别安全性。
- 不支持 SELECT INTO 语句。
- 行筛选条件表达式不支持 array 和 map 数据类型。支持 struct 数据类型。
- 要对使用行级别和单元级别筛选的表运行查询操作，必须使用名为 AmazonAthenaLakeFormation 的特殊工作组。有关 Athena 中工作组的信息，请参阅《Amazon Athena 用户指南》中的使用工作组运行查询。
- 可以针对表定义的数据筛选条件数量没有限制，但是对于单个主体，最多只能针对一个表定义 100 个数据筛选条件 SELECT 权限。
- 可包含在对一个表的授权中的最大数据筛选条件数量为 10。
- 要使用行筛选表达式应用数据筛选条件，必须使用授权选项获得对所有表列的 SELECT。当向外部账户授予权限时，此限制不适用于外部账户的管理员。
- 如果主体是组的成员，并且主体和该组都被授予了对部分行的权限，则主体的有效行权限是主体权限和组权限的并集。

- 在行级别筛选和单元格级别筛选中，以下列名受限：
  - ctid
  - oid
  - xmin
  - cmin
  - xmax
  - cmax
  - tableoid
  - insertxid
  - deletexid
  - importoid
  - redcatuniqueid
- 如果将 all-rows 筛选表达式与其他带有谓词的筛选表达式同时应用于表，则 all-rows 表达式将优先于所有其他筛选表达式。
- 当向外部账户授予行子集的权限，而外部 AWS 账户的数据湖管理员向该账户中的委托人授予这些权限时，委托人的有效筛选谓词是该账户的谓词和直接授予委托人的任何谓词的交集。

  例如，如果账户通过谓词 `dept='hr'` 获得了行权限，而主体被单独授予了对 `country='us'` 的该权限，则主体只能访问带有 `dept='hr'` 和 `country='us'` 的行。

有关单元格级别筛选的更多信息，请参阅Lake Formation 中的数据筛选和单元格级别安全性。

混合访问模式注意事项和限制

混合访问模式使您可以灵活地且有选择性地为 AWS Glue Data Catalog中的数据库和表启用 Lake Formation 权限。
在混合访问模式下，您现在有了增量路径，允许您为一组特定的用户设置 Lake Formation 权限，而不会中断其他现有用户或工作负载的权限策略。

以下注意事项和限制适用于混合访问模式。

限制

- 更新 Amazon S3 位置注册 – 您无法编辑使用服务相关角色在 Lake Formation 中注册的位置的参数。

- 使用 LF 标签时选择选项 – 当您可以使用 LF 标签授予 Lake Formation 权限时，您可以执行连续的一步，选择主体和附加了 LF 标签的数据库和表来强制实施 Lake Formation 权限。
- 选择主体 – 目前，只有数据湖管理员角色可以针对资源选择主体。
- 选择数据库中的所有表 – 在跨账户授权中，当您授予权限并选择数据库中的所有表时，需要选择数据库才能使权限生效。

注意事项

- 将在 Lake Formation 中注册的 Amazon S3 位置更新为混合访问模式 — 尽管可以将已经在 Lake Formation 中注册的 Amazon S3 数据位置转换为混合访问模式，但我们不建议这样做。
- 在混合访问模式下注册数据位置时的 API 行为
  - CreateTable — 无论混合访问模式旗帜和选择加入状态如何，该位置都被视为已在 Lake Formation 中注册。因此，用户需要拥有数据位置权限才能创建表。
  - CreatePartition/BatchCreatePartitions/UpdatePartitions（当分区位置更新为指向混合模式注册的位置时）— 无论混合访问模式标志和选择加入状态如何，Amazon S3 位置都被视为已在 Lake Formation 中注册。因此，用户需要拥有数据位置权限才能创建或更新数据库。
  - CreateDatabase/UpdateDatabase（当数据库位置更新为指向在混合访问模式下注册的位置时）— 无论混合访问模式标志和选择加入状态如何，该位置都被视为已在 Lake Formation 中注册。因此，用户需要拥有数据位置权限才能创建或更新数据库。
  - UpdateTable（当表格位置更新为指向在混合访问模式下注册的位置时）— 无论混合访问模式标志和选择加入状态如何，该位置都被视为已在 Lake Formation 中注册。因此，用户需要拥有数据位置权限才能更新表。如果表位置未更新或指向未在 Lake Formation 中注册的位置，则用户无需数据位置权限即可更新表。

Hive 元数据存储数据共享注意事项和限制

借助 AWS Glue Data Catalog 元数据联合（数据目录联合），您可以将数据目录连接到存储您的 Amazon S3 数据元数据的外部元数据存储，并使用 AWS Lake Formation安全地管理数据访问权限。

以下注意事项和限制适用于从 Hive 数据库创建的联合数据库：

注意事项

- AWS SAM 应用程序支持 — 您负责 AWS SAM 部署的应用程序资源（Amazon API Gateway 以及 Lambda 函数）的可用性。当用户运行查询时，请确保 AWS Glue Data Catalog 和 Hive 元数据仓之间的连接正常。

- Hive 元存储版本要求 – 您只能使用 Apache Hive 版本 3 及更高版本创建联合数据库。
- 映射数据库要求 — 每个 Hive 数据库都必须映射到 Lake Formation 中的新数据库。
- 数据库级别联合身份验证支持 – 您只能在数据库级别连接到 Hive 元存储。
- 对联合数据库的权限 – 即使删除了源表或数据库，应用于联合数据库或联合数据库下的表的权限也将保留。重新创建源数据库或表时，您无需重新授予权限。当在来源中删除具有 Lake Formation 权限的联合表时，Lake Formation 权限仍然显示，您可以根据需要撤销这些权限。

  如果用户删除联合数据库，则其所有相应权限都将丢失。重新创建同名的相同数据库将无法恢复 Lake Formation 权限。用户必须重新设置新权限。
- 联合数据库上@@ 的 IAM AllowedPrincipal 群组权限 — 基于此`DataLakeSettings`，Lake Formation 可能会将所有数据库和表的权限设置为名为的虚拟组`IAMAllowedPrincipal`。是`IAMAllowedPrincipal`指通过 IAM 委托人策略和 AWS Glue 资源策略访问数据目录资源的所有 IAM 委托人。如果对某数据库或表存在这些权限，则所有主体均被授予对该数据库或表的访问权限。

  但是，Lake Formation 不允许对联合数据库下的表实施 `IAMAllowedPrincipal` 权限。创建联合数据库时，请确保将 `CreateTableDefaultPermissions` 参数作为空列表传递。

  有关更多信息，请参阅更改数据湖的默认设置。
- 在查询中联接表 – 您可以将 Hive 元存储表与数据目录本机表联接以运行查询。

限制

- 在 AWS Glue Data Catalog 和 Hive 元数据仓之间同步元数据的限制 — 建立 Hive 元数据仓连接后，您需要创建一个联合数据库，以便将 Hive 元数据仓中的元数据与同步。 AWS Glue Data Catalog当用户运行查询时，联合数据库下的表将在运行时同步。
- 有关在联合数据库下创建新表的限制 – 您将无法在联合数据库下创建新表。
- 数据权限的限制 — 不支持对 Hive 元存储表视图的权限。

Amazon Redshift 数据共享限制

AWS Lake Formation 允许您安全地管理来自亚马逊 Redshift 的数据共享中的数据。Amazon Redshift 是一项完全托管的 PB 级云端数据仓库服务。 AWS 使用数据共享功能，Amazon Redshift 可帮助您跨 AWS 账户共享数据。有关 Amazon Redshift 数据共享的更多信息，请参阅 Amazon Redshift 中的数据共享概述。

以下注意事项和限制适用于通过 Amazon Redshift 数据共享创建的联合数据库：

- 映射的数据库要求 — 每个 Amazon Redshift 数据共享都必须映射到 Lake Formation 中的一个新数据库。当数据目录数据库中的数据共享对象表示形式精简化时，这是保持表名的唯一性所必需的。
- 有关在联合数据库下创建新表的限制 – 您将无法在联合数据库下创建新表。
- 对联合数据库的权限 – 即使删除了源表或数据库，应用于联合数据库或联合数据库下的表的权限也仍然存在。重新创建源数据库或表时，无需重新授予权限。当从来源中删除具有 Lake Formation 权限的联合表时，Lake Formation 权限仍然显示，并且您可以根据需要撤销这些权限。

  如果用户删除联合数据库，则其所有相应权限都将丢失。重新创建同名的相同数据库将无法恢复 Lake Formation 权限。用户必须重新设置新权限。
- 联合数据库上的 IAM AllowedPrincipal 群组权限 — 基于此`DataLakeSettings`，Lake Formation 可能会将所有数据库和表的权限设置为名为的虚拟组`IAMAllowedPrincipal`。是`IAMAllowedPrincipal`指通过 IAM 委托人策略和 AWS Glue 资源策略访问数据目录资源的所有 IAM 委托人。如果对某数据库或表存在这些权限，则所有主体均被授予对该数据库或表的访问权限。

  但是，Lake Formation 不允许对联合数据库下的表实施 `IAMAllowedPrincipal` 权限。创建联合数据库时，请确保将 `CreateTableDefaultPermissions` 参数作为空列表传递。

  有关更多信息，请参阅更改数据湖的默认设置。
- 数据筛选 – 在 Lake Formation 中，您可以使用列级别和行级别筛选来授予对联合数据库下表的权限。但是，您不能结合使用列级别和行级别筛选来限制以单元格级别粒度对联合数据库下的表进行访问。
- 区分大小写的标识符 – 由 Lake Formation 管理的 Amazon Redshift 数据共享对象仅支持小写的表名和列名。如果将使用 Lake Formation 共享和管理 Amazon Redshift 数据共享中的数据库、表和列，请勿针对它们启用区分大小写的标识符。

有关在 Amazon Redshift 中使用数据共享的限制的更多信息，请参阅《Amazon Redshift 数据库开发人员指南》中的数据共享限制。

IAM Identity Center 集成限制

借 AWS IAM Identity Center助，您可以连接到身份提供商 (IdPs)，并集中管理 AWS 分析服务中用户和群组的访问权限。您可以在 IAM Identity Center 中配置 AWS Lake Formation 为已启用的应用程序，数据湖管理员可以向授权用户和群组授予对 AWS Glue Data Catalog 资源的精细权限。

以下限制适用于 Lake Formation 与 IAM Identity Center 的集成：

- 您无法在 Lake Formation 中将 IAM Identity Center 用户和组分配为数据湖管理员或只读管理员。
- IAM Identity Center 用户和群组无法查询使用 AWS Key Management Service (AWS KMS) 密钥加密的数据目录表。 AWS KMS 不支持可信身份传播。
- IAM Identity Center 用户和组只能调用 IAM Identity Center 提供的 `AWSIAMIdentityCenterAllowListForIdentityContext` 策略中列出的 API 操作。

Lake Formation 基于标签的访问控制最佳实践和注意事项

您可以创建、维护和分配 LF 标签以控制对数据目录数据库、表和列的访问。

使用 Lake Formation 基于标签的访问控制时，请考虑以下最佳实践：

- 必须先预定义所有 LF 标签，然后才能将其分配给数据目录资源 或授予给主体。

  数据湖管理员可以通过创建具有所需 IAM 权限的 LF 标签创建者来委派标签管理任务。数据工程师和分析师决定 LF 标签的特征和关系。然后，LF 标签创建者在 Lake Formation 中创建和维护 LF 标签。
- 您可以将多个 LF 标签分配给数据目录资源。只能将特定键的一个值分配给特定资源。

  例如，您可以将 `module=Orders`、`region=West` 和 `division=Consumer` 等分配给数据库、表或列。您无法分配 `module=Orders,Customers`。
- 创建资源时，您无法将 LF 标签分配给资源。您只能将 LF 标签添加到现有资源。
- 您可以向主体授予 LF 标签表达式，而不仅仅是单个 LF 标签。

  LF 标签表达式如下所示（使用伪代码）。

  ```
  module=sales AND division=(consumer OR commercial)
  ```

  被授予此 LF 标签表达式的主体只能访问分配了 `module=sales` 和 `division=consumer` 或 `division=commercial` 的数据目录资源（数据库、表和列）。如果您希望主体能够访问具有 `module=sales` 或 `division=commercial` 的资源，请不要将两者都包含在同一个授予中。进行两次授予，一次用于 `module=sales`，一次用于 `division=commercial`。

  最简单的 LF 标签表达式仅包含一个 LF 标签，例如 `module=sales`。

- 被授予对具有多个值的 LF 标签的权限的主体可以使用其中任一值访问数据目录资源。例如，如果用户被授予 key=module 和 values=orders,customers 的 LF 标签，则该用户有权访问分配了 module=orders 或 module=customers 的资源。
- 您需要具有 Grant with LF-Tag expressions 权限，才能使用 LF-TBAC 方法授予对数据目录资源的数据权限。数据湖管理员和 LF 标签创建者会隐式接收此权限。具有 Grant with LFTag expressions 权限的主体可以使用以下方式授予对资源的数据权限
  - 命名资源方法
  - LF-TBAC 方法，但只能使用相同的 LF 标签表达式

  例如，假设数据湖管理员进行以下授予（使用伪代码）。

  ```
  GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
   GRANT OPTION
  ```

  在这种情况下，user1 可以使用 LF-TBAC 方法向其他主体授予对表的 SELECT 权限，但只能使用完整的 LF 标签表达式 module=customers, region=west,south。
- 如果同时使用 LF-TBAC 方法和命名资源方法向主体授予对资源的权限，则主体对资源具有的权限是这两种方法授予的权限的联合。
- Lake Formation 支持跨账户授予对 LF 标签的 DESCRIBE 和 ASSOCIATE 权限，以及支持使用 LF-TBAC 方法跨账户授予对数据目录资源的权限。在这两种情况下，本金都是 AWS 账户 ID。

  > **Note**
  >
  > Lake Formation 支持使用 LF-TBAC 方法向组织和组织单位进行跨账户授予。要使用此功能，您需要将跨账户版本设置更新为版本 3。

  有关更多信息，请参阅Lake Formation 中的跨账户数据共享。
- 在一个账户中创建的数据目录资源只能使用在同一账户中创建的 LF 标签进行标记。在一个账户中创建的 LF 标签不能与另一个账户中的共享资源关联。
- 使用基于 Lake Formation 标签的访问控制 (LF-TBAC) 授予跨账户访问数据目录资源的权限，需要为您的账户添加数据目录资源策略。 AWS 有关更多信息，请参阅先决条件。
- LF 标签键和 LF 标签值的长度不能超过 50 个字符。
- 可以分配给数据目录资源的 LF 标签的最大数量为 50。
- 以下限制是软限制：

  - 可以创建的 LF 标签的最大数量为 1000。
  - 可以为 LF 标签定义的值的最大数量为 1000。
- 标签、键和值在存储时将全部转换为小写。
- 只能将 LF 标签的一个值分配给特定资源。
- 如果通过一次授予向主体授予多个 LF 标签，则该主体只能访问具有所有 LF 标签的数据目录资源。
- AWS Glue ETL 作业需要完整表访问权限。如果 AWS Glue ETL 角色无法访问表中的所有列，则作业将失败。可以在列级别应用 LF-Tag，但这可能会导致 AWS Glue ETL 角色失去对表的访问权限并使作业失败。使用数据筛选条件进行列和/或行筛选不受此限制的影响。
- 如果 LF 标签表达式计算仅导致访问表列的子集，但在匹配时授予的 Lake Formation 权限是需要完整列访问权限的权限之一，即 `Alter`、`Drop`、`Insert` 或 `Delete`，则不会授予这些权限。相反，仅会授予 `Describe` 权限。如果授予的权限为 `All` (`Super`)，则仅会授予 `Select` 和 `Describe` 权限。
- 通配符不与 LF 标签一起使用。要将 LF 标签分配给某个表的所有列，请将 LF 标签分配给该表，并且表中的所有列都将继承该 LF 标签。要将 LF 标签分配给某个数据库中的所有表，请将 LF 标签分配给该数据库，并且数据库中的所有表都将继承该 LF 标签。

托管式数据压缩的支持的格式和限制

为了提高诸如 Amazon Athena、Amazon EMR 和 ETL AWS Glue Data Catalog 任务之类的 AWS 分析服务的读取性能 AWS Glue ，为数据目录中的 Iceberg 表提供了托管压缩（一种将小 Amazon S3 对象压缩成较大对象的过程）。

数据压缩支持多种用于读取和写入数据的压缩格式，例如从加密表中读取数据。

数据压缩支持：

- 数据类型：布尔值、整型、长整型、浮点数、双精度、字符串、十进制、日期、时间、时间戳、字符串、UUID、二进制
- 压缩：zstd、gzip、snappy、未压缩
- 加密：数据压缩仅支持默认的 Amazon S3 加密 (SSE-S3) 和服务器端 KMS 加密 (SSE-KMS)。
- 资源装箱压缩
- 架构演变
- 具有目标文件大小的表（写入。 target-file-size-bytes 冰山配置中的属性）在 128MB 到 512 MB 的包含范围内。

- 区域
  - 亚太地区（东京)
  - 亚太地区（首尔）
  - 亚太地区（孟买）
  - 欧洲地区（爱尔兰）
  - 欧洲地区（法兰克福）
  - 美国东部（弗吉尼亚州北部）
  - 美国东部（俄亥俄州）
  - 美国西部（北加利福尼亚）
  - South America（São Paulo）
- 当存储基础数据的 Amazon S3 存储桶位于另一个账户中时，您可以从数据目录所在的账户运行压缩。要实现此目的，压缩角色需要具有访问 Amazon S3 存储桶的权限。

数据压缩目前不支持：

- 数据类型：固定
- 压缩：brotli、lz4
- 随分区规格的演变压缩文件。
- 常规排序或 Z-Order 排序
- 合并或删除文件：压缩进程会跳过拥有与之关联的删除文件的数据文件。
- 对跨账户表进行压缩：您无法对跨账户表进行压缩。
- 对跨区域表进行压缩：您无法对跨区域表进行压缩。
- 针对资源链接启用压缩
- Amazon S3 存储桶的 VPC 端点

对 Lake Formation 问题进行故障排除

如果您在使用 AWS Lake Formation 时遇到问题，请查询本部分中的相关主题。

主题

一般故障排除

使用此处的信息可帮助您诊断和修复各种 Lake Formation 问题。

错误：对 <Amazon S3 位置> 的 Lake Formation 权限不足

试图创建或更改数据目录资源，但不具有对该资源指向的 Amazon S3 位置的数据位置权限。

如果数据目录数据库或表指向 Amazon S3 位置，则在授予 Lake Formation 权限 `CREATE_TABLE` 或 `ALTER` 时，您还必须授予对该位置的 `DATA_LOCATION_ACCESS` 权限。如果要向外部账户或组织授予这些权限，您必须包含授予选项。

向外部账户授予这些权限后，该账户中的数据湖管理员必须向该账户中的主体（用户或角色）授予这些权限。在授予从其他账户获得的 `DATA_LOCATION_ACCESS` 权限时，您必须指定拥有者账户的目录 ID（AWS 账户 ID）。拥有者账户是注册了该位置的账户。

有关更多信息，请参阅基础数据访问控制和授予数据位置权限。

错误：“Glue API 的加密密钥权限不足”

试图授予对加密数据目录的 AWS KMS 加密密钥的 Lake Formation 权限，但没有 AWS Identity and Access Management (IAM) 权限。

使用清单的 Amazon Athena 或 Amazon Redshift 查询失败

Lake Formation 不支持使用清单的查询。

错误：“Lake Formation 权限不足：需要在目录上创建标签”

用户/角色必须是数据湖管理员。

删除无效的数据湖管理员时出错

您应同时删除所有无效的数据湖管理员（已删除定义为数据湖管理员的 IAM 角色）。如果尝试单独删除无效的数据湖管理员，Lake Formation 会引发无效主体错误。

对跨账户访问问题进行故障排除

使用此处的信息可帮助您诊断和修复跨账户访问问题。

主题

我授予了跨账户 Lake Formation 权限，但接收者看不到资源。

- 接收者账户中的用户是否为数据湖管理员？ 只有数据湖管理员才能在共享时看到资源。
- 您是否在使用命名资源方法与组织外部的账户共享？ 如果是，则接收者账户的数据湖管理员必须接受 AWS Resource Access Manager (AWS RAM) 中的资源共享邀请。

  有关更多信息，请参阅the section called “接受 AWS RAM 资源共享邀请”。
- 您是否在 AWS Glue 中使用账户级（数据目录）资源策略？ 如果是，则如果您使用命名资源方法，则必须在策略中包含一条特殊语句，以授权 AWS RAM 代表您共享策略。

  有关更多信息，请参阅the section called “使用 AWS Glue 和 Lake Formation 管理跨账户权限。”。

- 您是否具有授予跨账户访问所需的 AWS Identity and Access Management (IAM) 权限？

  有关更多信息，请参阅the section called “先决条件”。
- 您已对其授予权限的资源不得具有向 `IAMAllowedPrincipals` 组授予的任何 Lake Formation 权限。
- 账户级策略中是否有关于资源的 `deny` 语句？

接收者账户中的主体可以看到数据目录资源，但无法访问基础数据

接收者账户中的主体必须具有必需的 AWS Identity and Access Management (IAM) 权限。有关详细信息，请参阅访问共享表的基础数据。

接受 AWS RAM 资源共享邀请时出现错误：“由于调用方未获得授权，关联失败”

在向其他账户授予对资源的访问权限后，当接收账户尝试接受资源共享邀请时，操作将失败。

```
$ aws ram get-resource-share-associations --association-type PRINCIPAL --resource-
share-arns arn:aws:ram:aws-region:444444444444:resource-share/e1d1f4ba-xxxx-xxxx-xxxx-
xxxxxxxx5d8d
{
    "resourceShareAssociations": [
        {
            "resourceShareArn": "arn:aws:ram:aws-region:444444444444:resource-share/
e1d1f4ba-xxxx-xxxx-xxxx-xxxxxxxx5d8d
",
            "resourceShareName": "LakeFormation-MMCC0XQBH3Y",
            "associatedEntity": "5815803XXXXX",
            "associationType": "PRINCIPAL",
            "status": "FAILED",
            "statusMessage": "Association failed because the caller was not
 authorized.",
            "creationTime": "2021-07-12T02:20:10.267000+00:00",
            "lastUpdatedTime": "2021-07-12T02:20:51.830000+00:00",
            "external": true
        }
    ]
}
```

发生此错误的原因是，当接收账户接受资源共享邀请时，AWS Glue 会调用 `glue:PutResourcePolicy`。要解决此问题，请允许生产者/授予者账户使用的代入角色执行 `glue:PutResourcePolicy` 操作。

错误：“无权授予对资源的权限”

试图授予对另一个账户拥有的数据库或表的跨账户权限。当与您的账户共享数据库或表时，作为数据湖管理员，您只能向账户中的用户授予对数据库或表的权限。

错误：“检索 AWS Organizations 信息的访问被拒绝”

您的账户是 AWS Organizations 管理账户，您没有检索组织信息（如账户中的组织单位）所需的权限。

有关更多信息，请参阅Required permissions for cross-account grants。

错误：“未找到组织 <organization-ID>”

尝试与组织共享资源，但未启用与 Organizations 的共享。启用与 Organizations 的资源共享。

有关更多信息，请参阅《AWS RAM 用户指南》中的启用与 AWS Organizations 的共享。

错误：“Lake Formation 权限不足：非法组合”

用户共享了数据目录资源，同时向 `IAMAllowedPrincipals` 组授予了该资源的 Lake Formation 权限。在共享资源之前，用户必须撤销 `IAMAllowedPrincipals` 的所有 Lake Formation 权限。

向外部账户授予/撤销请求时出现 ConcurrentModificationException 异常

当用户根据 LF 标签策略对主体提出多个并发授予和/或撤销权限请求时，Lake Formation 会引发 ConcurrentModificationException 异常。用户需要捕获异常，然后重试失败的授予/撤销请求。使用 `GrantPermissions/RevokePermissions` API 操作的批处理版本 - BatchGrantPermissions 和 BatchRevokePermissions 通过减少并发授予/撤销请求的数量，在一定程度上缓解了此问题。

使用 Amazon EMR 访问通过跨账户共享的数据时出错

使用 Amazon EMR 从其他账户访问与您共享的数据时，某些 Spark 库会尝试调用 `Glue:GetUserDefinedFunctions` API 操作。由于 AWS RAM 托管权限的版本 1 和 2 不支持此操作，您会收到以下错误消息：

```
"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-
spark-role/i-06ab8c2b59299508a is not authorized to perform:
glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource
because no resource-based policy allows the glue:GetUserDefinedFunctions
action"
```

要解决此错误，创建资源共享的数据湖管理员必须更新附加到资源共享的 AWS RAM 托管权限。托管权限的版本 3 AWS RAM 允许主体执行 `glue:GetUserDefinedFunctions` 操作。

如果您创建了新的资源共享，Lake Formation 会默认应用最新版 AWS RAM 托管权限，您无需执行任何操作。要为现有资源共享启用跨账户数据访问，您需要将 AWS RAM 托管权限更新到版本 3。

您可以在 AWS RAM 中查看分配给与您共享的资源的 AWS RAM 权限。版本 3 中包含以下权限：

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

更新现有资源共享的 AWS RAM 托管权限版本

您（数据湖管理员）可以按照《AWS RAM 用户指南》中的说明将 AWS RAM 托管权限更新到较新版本，也可以撤消资源类型的所有现有权限然后重新授予。如果您撤消权限，则 AWS RAM 会删除与资源类型关联的 AWS RAM 资源共享。如果您重新授予权限，AWS RAM 会创建新的资源共享并附加最新版本的 AWS RAM 托管权限。

对蓝图和工作流问题进行故障排除

使用此处的信息可帮助您诊断和修复蓝图和工作流问题。

主题

- 我的蓝图失败，并显示“User: <user-ARN> is not authorized to perform: iam:PassRole on resource: <role-ARN>”

我的蓝图失败，并显示“User: <user-ARN> is not authorized to perform: iam:PassRole on resource: <role-ARN>”

用户尝试创建蓝图，但该用户没有足够的权限来传递所选角色。

更新用户的 IAM 策略以便能够传递角色，或者要求用户选择具有所需 passrole 权限的其他角色。

有关更多信息，请参阅 the section called “Lake Formation 角色和 IAM 权限参考”。

我的工作流失败，并显示“User: <user-ARN> is not authorized to perform: iam:PassRole on resource: <role-ARN>”

您为工作流指定的角色没有允许角色自行传递的内联策略。

有关更多信息，请参阅the section called “（可选）为工作流程创建 IAM 角色”。

我的工作流中的爬网程序失败，并显示“资源不存在或请求者无权访问请求的权限”

一个可能的原因是传递的角色没有足够的权限在目标数据库中创建表。向该角色授予对数据库 `CREATE_TABLE` 权限。

我的工作流中的爬网程序失败，并显示“调用 CreateTable 操作时出错 (AccessDeniedException)...”

一个可能的原因是工作流角色对目标存储位置没有数据位置权限。向该角色授予数据位置权限。

有关更多信息，请参阅the section called “`DATA_LOCATION_ACCESS`”。

AWS Lake Formation 的已知问题

查看 AWS Lake Formation 的以下已知问题。

主题

对表元数据筛选的限制

AWS Lake Formation 列级权限可用于限制对表中特定列的访问。当用户使用控制台或 API（如 `glue:GetTable`）检索有关表的元数据时，表对象中的列列表仅包含他们有权访问的字段。请务必要了解这种元数据筛选的相关限制。

尽管 Lake Formation 为集成服务提供了有关列权限的元数据，但实际筛选查询响应中的列是相关集成服务的责任。支持列级筛选的 Lake Formation 客户端（包括 Amazon Athena、Amazon Redshift Spectrum 和 Amazon EMR）根据向 Lake Formation 注册的列权限筛选数据。用户将无法读取他们不应访问的任何数据。目前，AWS Glue ETL 不支持列筛选。

> **Note**
>
> EMR 集群并非完全由 AWS 管理。因此，EMR 管理员有责任妥善保护集群，以避免未经授权的数据访问。

某些应用程序或格式可能会将其他元数据（包括列名和类型）作为表属性存储在 `Parameters` 映射中。这些属性将未经修改返回，并且可由对任何列具有 `SELECT` 权限的任何用户访问。

例如，Avro Serde 将表架构的 JSON 表示形式存储在名为 `avro.schema.literal` 的表属性中，该属性可供有权访问该表的所有用户使用。建议避免在表属性中存储敏感信息，并注意用户可以了解 Avro 格式表的完整架构。此限制特定于表的元数据。

如果调用方对表中的所有列没有 `SELECT` 权限，则 AWS Lake Formation 在响应 `glue:GetTable` 或类似请求时会删除任何以 `spark.sql.sources.schema` 开头的表属性。这样可以防止用户访问有关

使用 Apache Spark 创建的表的其他元数据。在 Amazon EMR 上运行时，Apache Spark 应用程序仍然可以读取这些表，但可能不会应用某些优化，并且不支持区分大小写的列名。如果用户有权访问表中的所有列，则 Lake Formation 将返回未修改的表以及所有表属性。

重命名排除列时出现问题

如果使用列级权限排除列，然后重命名该列，则该列将不再被排除在查询之外，例如 `SELECT *`。

删除 CSV 表中的列时出现问题

如果您使用 CSV 格式创建数据目录表，然后从架构中删除列，则查询可能会返回错误数据，并且可能无法遵守列级权限。

解决方法：改为创建新表。

必须在公共路径下添加表分区

Lake Formation 期望表的所有分区都位于表的位置字段中设置的公共路径下。当您使用爬网程序将分区添加到目录时，这会无缝工作。但是，如果手动添加分区，并且这些分区不在父表中设置的位置下，则无法访问数据。

在工作流创建期间创建数据库时出现问题

使用 Lake Formation 控制台从蓝图创建工作流时，您可以创建目标数据库（如果该数据库不存在）。执行此操作时，登录的用户将获得对所创建数据库的 `CREATE_TABLE` 权限。但是，工作流生成的爬网程序在尝试创建表时会承担工作流的角色。由于该角色不具有对数据库的 `CREATE_TABLE` 权限，因此此操作会失败。

解决方法：如果您在工作流设置期间通过控制台创建数据库，则在运行工作流之前，必须向与工作流关联的角色授予对刚刚创建的数据库的 `CREATE_TABLE` 权限。

删除然后重新创建用户时出现问题

以下情况会导致 `lakeformation:ListPermissions` 返回错误的 Lake Formation 权限：

1. 创建用户并授予 Lake Formation 权限。
2. 删除用户。
3. 重新创建具有相同名称的用户。

ListPermissions 返回两个条目，一个用于旧用户，一个用于新用户。如果您尝试撤销授予旧用户的权限，则系统会撤消新用户的权限。

GetTables 和 SearchTables API 不会更新 IsRegisteredWithLakeFormation 参数的值

存在一个已知限制，即数据目录 API 操作（如 GetTables 和 SearchTables）不会更新 IsRegisteredWithLakeFormation parameter的值，并返回默认值 false。建议使用 GetTable API 查看 IsRegisteredWithLakeFormation parameter的正确值。

数据目录 API 操作不会更新 IsRegisteredWithLakeFormation 参数的值

存在一个已知限制，即数据目录 API 操作（如 GetTables 和 SearchTables）不会更新 IsRegisteredWithLakeFormation 参数的值，并返回默认值 false。建议使用 GetTable API 查看 IsRegisteredWithLakeFormation 参数的正确值。

Lake Formation 操作不支持 AWS Glue 架构注册表

Lake Formation 操作不支持在 StorageDescriptor 中包含 SchemaReference 的 AWS Glue 表，以便在架构注册表中使用。

更新了错误消息

AWS Lake Formation 已将以下 API 操作的特定资源异常更新为一般 EntityNotFound 错误消息，以满足安全性和合规目标。

- RevokePermissions
- GrantPermissions
- GetResourceLFTags
- GetTable
- GetDatabase

AWS Lake Formation API

> **Note**
>
> AWS Lake Formation 服务的更新版 API 参考现已推出。

目录

- 字符串模式

权限 API

“权限 API”介绍了在 AWS Lake Formation 中授予和撤消权限所需的操作和数据类型。有关所有 AWS Lake Formation API 操作和数据类型，请参阅 Lake Formation API 参考指南。

操作

- GrantPermissions
- RevokePermissions
- BatchGrantPermissions
- BatchRevokePermissions
- GetEffectivePermissionsForPath
- ListPermissions

数据类型

- 资源
- DatabaseResource
- TableResource
- TableWithColumnsResource
- DataCellsFilterResourcee
- DataLocationResource
- DataLakePrincipal
- PrincipalPermissions
- PrincipalResourcePermissions
- DetailsMap
- ColumnWildcard
- BatchPermissionsRequestEntry
- BatchPermissionsFailureEntry

数据湖设置 API

本节包含用于管理数据湖管理员的数据湖设置 API 操作和数据类型。

操作

- GetDataLakeSettings
- PutDataLakeSettings

数据类型

- DataLakeSettings

IAM Identity Center 集成 API

本部分包含用于创建和管理 Lake Formation 与 IAM Identity Center 集成的操作。

操作

- CreateLakeFormationIdentityCenterConfiguration
- DeleteLakeFormationIdentityCenterConfiguration
- DescribeLakeFormationIdentityCenterConfiguration
- UpdateLakeFormationIdentityCenterConfiguration

数据类型

- ExternalFilteringConfiguration

混合访问模式 API

“混合访问模式 API”一节介绍了在 AWS Lake Formation 中设置混合访问模式所需的操作和数据类型。有关所有 AWS Lake Formation API 操作和数据类型，请参阅 Lake Formation API 参考指南。

操作

- CreateLakeFormationOptIn

- DeleteLakeFormationOptIn
- ListLakeFormationOptIns

数据类型

- 资源
- DatabaseResource
- TableResource
- 资源信息
- LakeFormationOptInsInfo
- DataLocationResource

凭证售卖 API

“凭证售卖 API”一节介绍了与使用 AWS Lake Formation 服务来售卖凭据以及注册和管理数据湖资源相关的操作和数据类型。

操作

- RegisterResource
- DeregisterResource
- ListResources
- GetUnfilteredTableMetadata
- GetUnfilteredPartitionsMetadata
- GetTemporaryGluePartitionCredentials
- GetTemporaryGlueTableCredentials
- UpdateResource

数据类型

- FilterCondition
- RowFilter
- ResourceInfo

标记 API

“标记 API”一节介绍了与授权策略相关的操作和数据类型，该策略针对属性或键值对标签定义了权限模型。

操作

- AddLFTagsToResource
- RemoveLFTagsFromResource
- GetResourceLFTags
- ListLFTags
- CreateLFTag
- GetLFTag
- UpdateLFTag
- DeleteLFTag
- SearchTablesByLFTags
- SearchDatabasesByLFTags

数据类型

- LFTagKeyResource
- LFTagPolicyResource
- TaggedTable
- TaggedDatabase
- LFTag
- LFTagPair
- LFTagError
- ColumnLFTag

数据筛选条件 API

“数据筛选条件 API”介绍了如何在 AWS Lake Formation 中管理数据单元格筛选条件。

操作

- CreateDataCellsFilter
- DeleteDataCellsFilter
- ListDataCellsFilter
- GetDataCellsFilter
- UpdateDataCellsFilter

数据类型

- DataCellsFilter
- RowFilter

常见数据类型

常见数据类型介绍 AWS Lake Formation 中的各种常见的数据类型。

ErrorDetail 结构

包含有关错误的详细信息。

字段

- `ErrorCode` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern 匹配。

  与此错误关联的代码。
- `ErrorMessage` – 描述字符串，长度不超过 2048 个字节，与 URI address multi-line string pattern 匹配。

  描述错误的消息。

字符串模式

API 使用以下正则表达式来定义对于各种字符串参数和成员有效的内容：

- 单行字符串模式 -“`[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*`”

- URI 地址多行字符串模式 -“[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*”
- 自定义字符串模式 3 –“^\w+\.\w+\.\w+$”
- 自定义字符串模式 4 –“^\w+\.\w+$”
- 自定义字符串模式 5 –“arn:aws:iam::[0-9]*:role/.*”
- 自定义字符串模式 6 –“arn:aws:iam::[0-9]*:user/.*”
- 自定义字符串模式 7 –“arn:aws:iam::[0-9]*:group/.*”
- 自定义字符串模式 #8 –“arn:aws:iam::[0-9]*:saml-provider/.*”
- 自定义字符串模式 #9 –“^([\p{L}\p{Z}\p{N}_.:\/=+\-@%]*)$”
- 自定义字符串模式 #10 –“^([\p{L}\p{Z}\p{N}_.:*\/=+\-@%]*)$”
- 自定义字符串模式 #11 –“[\p{L}\p{N}\p{P}]*”

支持的区域

本节包含有关 Lake Formation 的支持 AWS 区域 和功能的信息。

正式发布

有关 AWS 区域 支持的服务 AWS Lake Formation，请参阅按地区划分的可用 AWS 服务列表。

有关每个区域的 Lake Formation 服务端点和 Lake Formation 服务限额的列表，请参阅 AWS Lake Formation 端点和限额。

AWS GovCloud (US)

有关 AWS GovCloud (US) 地区与标准版之间差异的概述 AWS 区域，请参阅 “有何 AWS Lake Formation 区别” AWS GovCloud (US)。

事务和存储优化

Lake Formation 的受管控表、事务支持和存储优化功能可在以下 AWS 区域版本中找到：

区域名称	区域参数	终端节点
美国东部（弗吉尼亚州北部）	us-east-1	lakeformation.us-east-1.amazonaws.com lakeformation-fips.us-east-1.amazonaws.com
美国东部（俄亥俄州）	us-east-2	lakeformation.us-east-2.amazonaws.com lakeformation-fips.us-east-2.amazonaws.com
美国西部（俄勒冈州）	us-west-2	lakeformation.us-west-2.amazonaws.com

区域名称	区域参数	终端节点
		lakeformation-fips.us-west-2.amazonaws.com
亚太地区（孟买）	ap-south-1	lakeformation.ap-south-1.amazonaws.com
亚太地区（首尔）	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
亚太地区（新加坡）	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
亚太地区（悉尼）	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
亚太地区（东京)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
欧洲地区（法兰克福）	eu-central-1	lakeformation.eu-central-1.amazonaws.com
欧洲地区（爱尔兰）	eu-west-1	lakeformation.eu-west-1.amazonaws.com
欧洲（伦敦）	eu-west-2	lakeformation.eu-west-2.amazonaws.com
欧洲地区（斯德哥尔摩）	eu-north-1	lakeformation.eu-north-1.amazonaws.com
加拿大（中部）	ca-central-1	lakeformation.ca-central-1.amazonaws.com

区域名称	区域参数	终端节点
South America（São Paulo）	sa-east-1	lakeformation.sa-east-1.amazonaws.com

AWS Lake Formation 的文档历史记录

下表介绍对 AWS Lake Formation 文档的一些重要更改。

变更	说明	日期
更新了 Lake Formation 的设置	更新了设置AWS Lake Formation部分中的步骤。	2024年2月7日
更新了政策变更	为服务相关角色的内联策略添加了新权限。有关更多信息，请参阅为 La ke Formation 使用服务相关角色。	2024年2月7日
更新了政策变更	记录了LakeFormationDataAccessServiceRolePolicy政策的变更。	2024 年 2 月 2 日
整合了 Lake Formation 限制	针对 Lake Formation 限制和注意事项创建了一个统一的部分。有关更多信息，请参阅 Lake Formation 限制。	2023 年 12 月 15 日
添加了有关 Iceberg 压缩的文档	为提高 AWS 分析服务（例如 Athena 和 Amazon EMR）和 AWS Glue ETL 作业的读取性能，AWS Glue Data Catalog 为数据目录中的 Iceberg 表提供了托管式压缩功能（一种将小 Amazon S3 对象压缩成较大对象的进程）。有关更多信息，请参阅优化 Iceberg 表。	2023 年 11 月 25 日
添加了有关 IAM Identity Center 集成的文档	IAM Identity Center 集成允许用户和组访问数据目录资源，从而强制执行 Lake Formation	2023 年 11 月 25 日

	权限。有关更多信息，请参阅 IAM Identity Center 集成。	
添加了有关数据目录视图的文档	您可以使用适用于 Amazon Athena 或 Amazon Redshift 的 SQL 编辑器在 AWS Glue Data Catalog 中创建最多引用 10 个表的视图。有关更多信息，请参阅创建视图。	2023 年 11 月 25 日
更新了策略更改	记录了AWSLakeFormationCrossAccountManager政策的变更。	2023 年 10 月 25 日
添加了有关混合访问模式的文档	混合访问模式使您可以灵活地且有选择性地为 AWS Glue Data Catalog 中的数据库和表启用 Lake Formation 权限。在混合访问模式下，您现在有了增量路径，可为一组特定的用户设置 Lake Formation 权限，而不会中断其他现有用户或工作负载的权限策略。有关更多信息，请参阅混合访问模式。	2023 年 9 月 26 日
添加了有关创建 Apache Iceberg 表的文档	现在，您可以使用驻留在 Amazon S3 中的数据在 AWS Glue Data Catalog 中创建使用 Apache Parquet 数据格式的 Apache Iceberg 表。有关更多信息，请参阅创建 Iceberg 表。	2023 年 8 月 16 日

添加了有关跨区域数据访问的文档	Lake Formation 支持跨 AWS 区域查询数据目录表。您可以使用 Athena、Amazon EMR 访问其他区域中的数据，AWS Glue并通过在其他区域创建指向源数据库和表的资源链接来运行 ETL。您可以将数据目录连接到存储您的 Amazon S3 数据元数据的外部元存储，并使用 AWS Lake Formation 安全地管理数据访问权限。有关更多信息，请参阅跨区域访问表。	2023 年 6 月 30 日
重新整理了内容	重新整理了本指南中的章节，以匹配 Lake Formation 用户旅程。	2023 年 5 月 15 日
添加了有关 HMS 联合身份验证的文档	您可以将数据目录连接到存储您的 Amazon S3 数据元数据的外部元存储，并使用 AWS Lake Formation 安全地管理数据访问权限。有关更多信息，请参阅管理对使用外部元存储的数据集的权限。	2023 年 4 月 15 日
添加了有关 Amazon Redshift 数据共享的文档	现在，您可以使用 Lake Formation 权限安全地管理来自 Amazon Redshift 的数据共享中的数据。Lake Formation 支持通过 AWS Data Exchange 许可访问您的数据。有关更多信息，请参阅 AWS Lake Formation 中的数据共享。	2022 年 11 月 30 日

支持直接跨账户与主体共享数据	添加了有关直接与其他账户中的 IAM 主体共享数据的信息。有关更多信息，请参阅 AWS Lake Formation 中的跨账户数据共享。	2022 年 11 月 10 日
支持 AWS RAM 使用 TBAC 启用数据共享	添加了有关 LF-TBAC 这一数据目录权限授予方法的信息，该方法使用 AWS Resource Access Manager 进行跨账户授权。	2022 年 11 月 10 日
添加了有关使用其他服务的章节	添加了有关 Athena、AWS Glue、Redshift Spectrum 和 Amazon EMR 等 AWS 服务如何使用 Lake Formation 安全地访问在 Lake Formation 中注册的 Amazon S3 位置处的数据的信息。有关更多信息，请参阅使用其他 AWS 服务。	2022 年 11 月 10 日
???	添加了有关在使用 Amazon EMR 访问跨账户数据时对错误进行故障排除的信息。有关更多信息，请参阅使用 Amazon EMR 访问通过跨账户共享的数据时出错。	2022 年 11 月 7 日
更新了跨账户资源共享	添加了有关跨账户资源共享在 Lake Formation 中的工作原理的描述。记录了AWSLakeFormationCrossAccountManager政策的变更。	2022 年 5 月 6 日

新教程	添加了有关创建受管控表、保护数据湖和共享数据湖的新教程。有关更多详细信息，请参阅开始使用一节。	2022 年 4 月 20 日
全新 Lake Formation 登录页面	更新了 Lake Formation 登录页面，增加了教程链接，这些链接提供了 step-by-step有关如何使用 Lake Formation 构建数据湖、采集数据、共享和保护数据湖的说明。	2022 年 4 月 20 日
支持凭证售卖	添加了有关凭证售卖的信息，该功能支持 Lake Formation 允许第三方服务使用凭证售卖 API 操作与 Lake Formation 集成。有关更多信息，请参阅 Lake Formation 中凭证售卖功能的工作原理。	2022 年 2 月 28 日
支持受管控表和高级数据筛选	添加了关于支持 ACID 事务、自动数据压缩和时间旅行查询的受管控表的信息。添加了有关创建数据筛选条件以支持实现列级别安全性、行级别安全性和单元格级别安全性的信息。有关更多信息，请参阅 Lake Formation 中的受管控表和 Lake Formation 中的数据筛选和单元格级别安全性。	2021 年 11 月 30 日

支持 VPC 接口端点	添加了有关为 Lake Formation 创建虚拟私有云 (VPC) 接口端点的信息，这样您的 VPC 和 Lake Formation 之间的通信就可以在 AWS 网络内完整且安全地进行。有关更多信息，请参阅将 Lake Formation 与 VPC 端点搭配使用。	2021 年 10 月 11 日
支持 VPC 终端节点策略	添加了有关在 Lake Formation 中支持虚拟私有云 (VPC) 端点策略的信息。有关更多信息，请参阅将 Lake Formation 与 VPC 端点搭配使用。	2021 年 10 月 11 日
支持基于标签的访问控制	Lake Formation 基于标签的访问控制提供了一种更具可扩展性的新方式，通过使用 LF 标签来管理对数据目录资源和基础数据的访问。有关更多信息，请参阅 Lake Formation 基于标签的访问控制。	2021 年 5 月 7 日
关于在 Amazon EMR 上筛选数据的全新选择要求	添加了有关如何进行选择以允许 Amazon EMR 筛选 Lake Formation 托管数据的要求的信息。有关更多信息，请参阅允许在 Amazon EMR 上筛选数据。	2020 年 10 月 9 日
支持授予对数据目录数据库的完整跨账户权限	添加了有关跨 AWS 账户（包括 CREATE_TABLE ）授予对数据目录数据库的完整 Lake Formation 权限的信息。有关更多信息，请参阅共享数据目录数据库。	2020 年 10 月 1 日

支持 Amazon Athena 用户通过 SAML 进行身份验证。	添加了有关支持 Athena 用户通过 JDBC 或 ODBC 驱动程序进行连接并通过 Okta 和 Microsoft Active Directory 联合身份验证服务 (AD FS) 等 SAML 身份提供者进行身份验证的信息。有关更多信息，请参阅 AWS 服务与 Lake Formation 集成。	2020 年 9 月 30 日
支持使用加密数据目录进行跨账户访问	添加了有关在数据目录加密时授予跨账户权限的信息。有关更多信息，请参阅跨账户访问先决条件。	2020 年 7 月 30 日
支持跨账户访问数据湖	添加了有关向外部 AWS 账户和组织授予对数据目录数据库和表的 AWS Lake Formation 权限以及从外部账户访问共享数据目录对象的信息。有关更多信息，请参阅跨账户访问。	2020 年 7 月 7 日
与 Amazon 集成 QuickSight	添加了有关如何向亚马逊 QuickSight 企业版用户授予 Lake Formation 权限，以便他们可以访问位于已注册的 Amazon S3 位置的数据集的信息。有关更多信息，请参阅授予数据目录权限。	2020 年 6 月 29 日
更新了“设置”和“入门”章节	重新整理并改进了“设置”和“入门”章节。更新了为数据湖管理员建议的 AWS Identity and Access Management (IAM) 权限。	2020 年 2 月 27 日

支持 AWS Key Management Service	添加了有关 Lake Formation 对 AWS Key Management Service (AWS KMS) 的支持如何简化集成服务的设置，以便在已注册的 Amazon Simple Storage Service (Amazon S3) 位置读取和写入加密数据的信息。添加了有关如何注册已使用 AWS KMS keys 加密的 Amazon S3 位置的信息。有关更多信息，请参阅the section called “向数据湖添加 Amazon S3 位置”。	2020 年 2 月 27 日
更新了蓝图和数据湖管理员 IAM 策略	阐明了增量数据库蓝图的输入参数。更新了数据湖管理员所需的 IAM 策略。	2019 年 12 月 20 日
重写了“安全性”章节并修订了“升级”章节	改进了“安全性”和“升级”章节。	2019 年 10 月 29 日
Super 权限取代了 All 权限	更新了“安全性”和“升级”章节，说明了已将权限 `All` 替换为 `Super`。	2019 年 10 月 10 日
补充、更正和澄清说明	根据反馈进行了补充、更正和澄清说明。修订了“安全性”章节。更新了“安全性”和“升级”章节，说明了已将组 `Everyone` 替换为 `IAMAllowedPrincipals` 。	2019 年 9 月 11 日
新指南	这是 AWS Lake Formation 开发人员指南的初始版本。	2019 年 8 月 8 日

AWS 术语表

有关最新的 AWS 术语，请参阅《AWS 词汇表参考》中的 AWS 词汇表。

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。