Amazon Lake Formation

开发人员指南

Amazon Lake Formation: 开发人员指南

Amazon Web Services 文档中描述的 Amazon Web Services 服务或功能可能因区域而异。要查看适用于中国区域的差异，请参阅中国的 Amazon Web Services 服务入门。

Table of Contents

什么是 Amazon Lake Formation？

欢迎阅读 Amazon Lake Formation 开发人员指南。

Amazon Lake Formation是一种完全托管服务，它让用户能够轻松地构建、保护和管理数据湖。Lake Formation 简化并自动化了创建数据湖通常所需的许多复杂的手动步骤。这些步骤包括收集、清理、移动和编目数据，以及安全地将这些数据用于分析和机器学习。

Lake Formation 提供了自己的权限模型，该模型增强了 IAM 权限模型。这种集中定义的权限模型允许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 (RDMS) 非常相似。Lake Formation 权限是在列、行和单元格级别使用精细控制强制执行的Amazon分析和机器学习服务，包括Amazon Athena、亚马逊 QuickSight和 Amazon Redshift。

主题

Lake Formation 功能

Lake Formation 可以打破数据孤岛，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，找出存储在 Amazon S3 或关系数据库和 NoSQL 数据库中的现有数据，然后将数据移动到您的数据湖中。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，为他们提供安全的自助数据访问权限。

主题

设置和数据管理

从已有的数据库导入数据Amazon

一旦您指定了现有数据库的位置并提供了访问凭证，Lake Formation 就会读取数据及其元数据（架构）以了解数据源的内容。然后，它将数据导入您的新数据湖，并将元数据记录在中央目录中。使用 Lake Formation，你可以从在亚马逊 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 通过连接 Java 数据库连接 (JDBC) 从本地数据库中移动数据。确定您的目标来源并在控制台中提供访问凭证，Lake Formation 会读取您的数据并将其加载到数据湖中。要从上面列出的数据库以外的数据库导入数据，您可以使用以下命令创建自定义 ETL 任务Amazon Glue.

对您的数据进行编目和标记

Lake Formation 会抓取和读取您的数据源以提取技术元数据，并创建一个可搜索的目录来为用户描述这些信息，以便他们能够发现可用的数据集。您还可以在数据（表和列级别）中添加自己的自定义标签，以定义属性，例如 “敏感信息” 和 “欧洲销售数据”。Lake Formation 针对这些元数据提供基于文本的搜索，因此您的用

户可以快速找到他们需要分析的数据。有关将表添加到数据目录的更多信息，请参阅管理数据目录表和数据库 (p. 113).

转换数据

Lake Formation 可以对您的数据进行转换，例如重写各种日期格式以保持一致性，以确保以便于分析的方式存储数据。Lake Formation 创建转换模板并安排作业以准备数据以供分析。您的数据通过以下方式进行转换Amazon Glue并以列式格式编写，例如 Parquet 和 ORC ）））。

清除和删除重复数据

Lake Formation 通过提供名为的机器学习转换，帮助清理和准备数据以供分析 FindMatches 用于重复数据删除和查找匹配记录。例如，请使用 FindMatches 在餐厅数据库中查找重复记录，例如当一条记录在 “121 Main St” 列出 “Joe's Pizza” 时 另一个是 “121 Main” 上的 “Joseph's Pizzeria”。 FindMatches 只会要求你将记录集标记为 “匹配” 或 “不匹配”。然后，系统将学习你将一对记录称为匹配项的标准，并将生成一个机器学习转换，你可以用它来查找数据库中的重复记录或两个数据库中的匹配记录。有关更多信息 FindMatches，请参阅将记录Amazon Lake Formation FindMatches中的Amazon Glue开发人员指南.

存储优化

许多小文件的存储效率低下可能会影响分析性能，这些小文件是在向数据湖写入新数据时自动创建的。处理这么多小文件会增加分析服务的开销，并导致查询响应变慢。Lake Formation 包含一个存储优化器，可自动将小文件合并成大文件，从而将查询速度提高多达 7 倍。此过程通常称为压缩，是在后台执行的，因此在此过程中不会对生产工作负载产生性能影响。有关 Lake Formation 的存储优化功能的更多信息，请参阅受管表的存储优化 (p. 118).

行级别和单元级别安全性

Lake Formation 提供数据筛选器，允许您限制对列和行组合的访问。使用行级和单元级安全保护敏感数据，例如个人身份信息 (PII)。有关行级别安全的更多信息，请参阅数据筛选概述 (p. 206).

安全管理

定义和管理访问控制

Lake Formation 为管理数据湖中数据的访问控制提供了一个单一位置。您可以定义安全策略，在数据库、表、列、行和单元格级别限制对数据的访问。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时的用户和群组。您可以使用精细的控制来访问Amazon Redshift Spectrum、Athena 中由 Lake Formation 保护的数据Amazon GlueETL 和 Apache Spark 的 Amazon EMR。

实施审计日志记录

Lake Formation 提供全面的审计日志 CloudTrail 监控访问并显示对集中定义的策略的遵守情况。您可以审核分析和机器学习服务的数据访问历史记录，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。您可以像访问其他任何日志一样访问审核日志 CloudTrail 使用对进行日志 CloudTrailAPI 和控制台。有关 的更多信息 CloudTrail 日志见日志系统AmazonLake Formation API 调用使用Amazon CloudTrail (p. 265).

基于标签的访问控制

您可以对数据进行分类并限制对敏感信息的访问。您还可以在表和列级别向数据添加自己的自定义标签（LF 标签）以定义属性，例如 “敏感信息” 或 “欧洲销售数据”。Lake Formation 针对这些元数据提供基于文本的搜索，因此您的用户可以快速找到他们需要分析的数据。您可以根据这些 LF 标签授予对数据的访问权限。有关基于标记的访问控制的更多信息，请参阅Lake Formation 标签访问控制 (p. 179).

跨账户访问

Lake Formation 权限管理功能简化了跨多个分布式数据湖的安全和管理Amazon通过集中式方法，提供对数据目录和 Amazon Simple Services 的细粒度访问控制。

受监管的表

数据湖需要随时向用户显示正确的数据视图，即使数据同时实时或频繁更新也是如此。加载流数据或合并来自多个源数据系统的更改需要并parallel 处理多个表中的插入和删除。如今，开发人员编写自定义应用程序代码或使用开源工具来管理这些更新。这些解决方案既复杂又难以扩展，因为编写能够在同时读取和写入相同数据时保持一致性的应用程序代码既繁琐又脆弱，而且容易出错。

Lake Formation 引入了新的 API，这些新的 API 使用一种新的数据湖表类型（称为受监管的表. 受管控的表允许多个用户使用清单同时在表中插入和删除数据，同时允许其他用户在相同的数据集上同时运行分析查询和机器学习模型，这些数据集返回一致和 up-to-date 结果。

有关如何使用 Lake Formation 的更多信息，请参阅以下主题：

- Lake Formation 中的受管桌子 (p. 114)
- 在事务中读取和写入数据湖 (p. 222)

Amazon与 Lake Formation 的服务集成

以下Amazon与集成的服务Amazon Lake Formation并尊重Lake Formation 的权限。

Amazon 服务	如何集成
Amazon Glue	Amazon Glue和 Lake Formation 共享相同的数据目录。对于控制台操作（例如查看表列表）和所有 API 操作，Amazon Glue用户只能访问他们拥有 Lake Formation 权限的数据库和表。 Note Amazon Glue不支持 Lake Formation 专栏权限。
Amazon Athena	何时Amazon Athena用户选择Amazon Glue在查询编辑器中，他们只能查询他们拥有 Lake Formation 权限的数据库、表和列。不支持使用清单的查询。 除了通过以下方式向 Athena 进行身份验证的校长之外Amazon Identity and Access Management(IAM)，Lake Formation 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 和微软 Active Directory 联合服务 (AD FS)。有关更多信息，请参阅将 Lake Formation 和 Athena JDBC 和 ODBC 驱动程序用于对 Athena 进行联合访问中的Amazon Athena 用户指南. Note 目前，以下区域不支持授权访问 Lake Formation 中的 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
Amazon Redshift Spectrum	当 Amazon Redshift 用户在数据库上创建外部架构时Amazon Glue目录，他们只能查询该架构中他们拥有 Lake Formation 权限的表和列。 不支持使用清单的查询。

Amazon 服务	如何集成
亚马逊 QuickSight 企业版	当亚马逊的时候 QuickSight 企业版用户在 Amazon S3 位置查询已注册到 Lake Formation 的数据集，用户必须有 Lake Formation`SELECT`对数据的许可。
Amazon EMR	使用 Apache Zeppelin 或 EMR 笔记本提交 Apache Spark 应用程序时，将强制执行 Lake Formation 权限。

Lake Formation 也适用于Amazon Key Management Service(Amazon KMS），使您能够更轻松地设置这些集成服务，在 Amazon Simple Storage Service (Amazon S3)) 中加密和解密数据。

支持的区域

对于Amazon Web Services 区域支持Amazon Lake Formation，请参阅Amazon Lake Formation定价.

有关每个区域的Lake Formation 服务终端节点以及Lake Formation 的配额的列表，请参阅Amazon Lake Formation终端节点和配额.

Lake Formation 的受管表格、事务支持、单元级安全和存储优化功能可在以下版本中找到Amazon Web Services 区域.

区域名称	区域参数
美国东部（弗吉尼亚州北部）	us-east-1
美国东部 (俄亥俄)	us-east-2
美国西部（俄勒冈州）	us-west-2
Asia Pacific (Mumbai)	ap-south-1
Asia Pacific (Seoul)	ap-northeast-2
亚太地区（新加坡）	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
欧洲（法兰克福）	eu-central-1
Europe (Ireland)	eu-west-1
欧洲（伦敦）	eu-west-2
欧洲（斯德哥尔摩）	eu-north-1
Canada (Central)	ca-central-1
South America (São Paulo)	sa-east-1

Lake Formation 入门

我们建议您从以下部分入手：

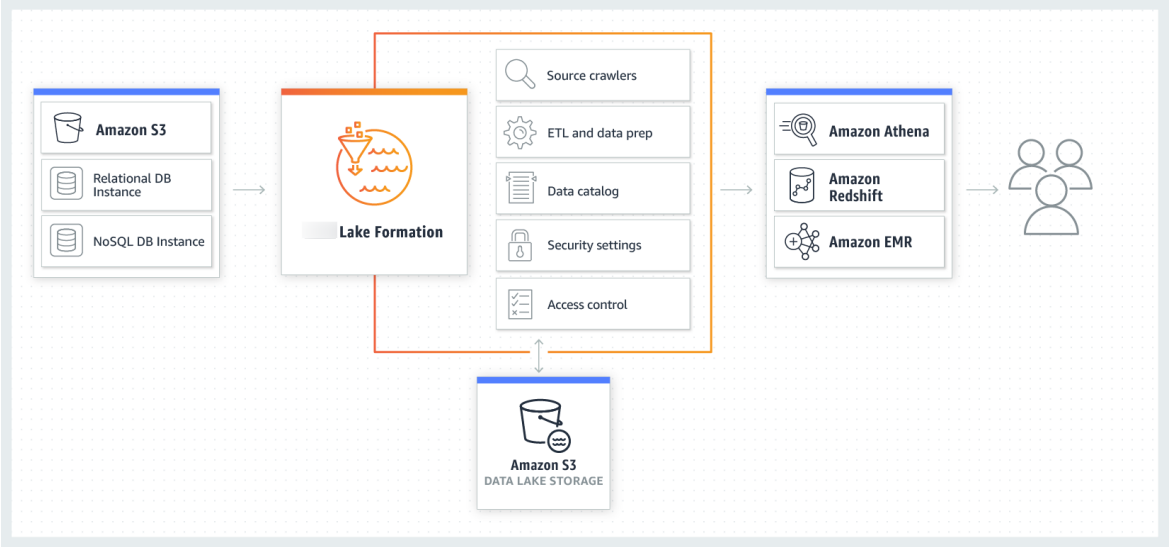
- Amazon Lake Formation：工作方式 (p. 6)— 了解基本术语以及各个组件如何相互作用。
- 设置 Amazon Lake Formation (p. 9)— 获取有关先决条件的信息，并完成重要的安装任务。
- 教程 (p. 29)— 跟随 step-by-step 教程，了解如何使用Lake Formation。
- Amazon Lake Formation 中的安全性 (p. 230)— 了解如何帮助安全访问 Lake Formation 中的数据。

Amazon Lake Formation：工作方式

Amazon Lake Formation使您能够更轻松地构建、保护和管理数据湖。Lake Formation 可以直接或通过其他方式帮助您完成以下操作Amazon服务：

- 注册 Amazon Simple Storage Service (Amazon S3) 存储桶和数据湖将驻留的路径。
- 协调收集、清理、转换和组织原始数据的数据流。
- 创建和管理包含有关数据源和数据湖中数据的元数据的数据目录。
- 通过授予/撤销权限模型定义对元数据和数据的精细数据访问策略。

下图说明了如何在 Lake Formation 中加载和保护数据。



如图所示，Lake Formation 管理Amazon Glue爬网程序，Amazon GlueETL 作业、数据目录、安全设置和访问控制。数据安全存储在数据湖中后，用户可以通过选择的分析服务（包括 Amazon Athena、Amazon Redshift 和 Amazon EMR）访问数据。

主题

Lake Formation

以下是您将在本指南中遇到的一些重要术语。

数据湖

这些区域有：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数据。数据湖通常存储以下内容：

- 结构化和非结构化数据

- 原始数据和转换后的数据

要使 Amazon S3 路径位于数据湖中，它必须是已注册与 Lake Formation

数据访问

Lake Formation 通过新的授予/撤销权限模型提供对数据的安全、精细访问，该模型可以增强Amazon Identity and Access Management(IAM) 策略。

分析师和数据科学家可以使用Amazon分析和机器学习服务，例如 Amazon Athena，以访问数据。配置的 Lake Formation 安全策略有助于确保用户只能访问已授权访问的数据。

Blueprint

一个蓝图是一个数据管理模板，使您能够轻松地将数据提取到数据湖中。Lake Formation 提供了几个蓝图，每个蓝图针对预定义的源类型，例如关系数据库或Amazon CloudTrail日志。从蓝图中，您可以创建工作流程。包含以下内容：Amazon Glue为协调数据的加载和更新而生成的爬虫、作业和触发器。蓝图将数据源、数据目标和计划作为输入来配置工作流程。

工作流程

一个工作流是一组相关的容器Amazon Glue作业、爬网程序和触发器。您在 Lake Formation 中创建工作流程，然后在Amazon Glue服务。Lake Formation 可以作为单个实体跟踪工作流的状态。

定义工作流时，您可以选择工作流所基于的蓝图。然后，您可以根据需要或计划运行工作流程。

您在 Lake Formation 中创建的工作流程可在Amazon Glue控制台作为有向无环图 (DAG)。利用 DAG，您可以跟踪工作流程的进度并执行故障排除。

数据目录

这些区域有：Data Catalog是您的持久性元数据存储。它是一项托管式服务，可让您在 Amazon 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Amazon Glue数据目录来存储有关数据湖、数据源、转换和目标的元数据。

有关数据源和目标的元数据采用数据库和表格的形式。表格存储架构信息、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Amazon每个账户都有一个数据目录Amazon区域。

底层数据

底层数据指的是数据目录表所指向的数据湖中的源数据或数据湖中的数据。

主体

一个校长是Amazon Identity and Access Management(IAM) 用户或角色或活动目录用户。

数据湖管理员

一个数据湖管理员是可以向任何委托人（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委托人。指定数据湖管理员作为数据目录的第一个用户。然后，该用户可以向其他委托人授予更精细的资源权限。

Note

IAM 管理用户 — 具有`AdministratorAccess` Amazon托管策略 — 不会自动成为数据湖管理员。例如，他们无法授予对目录对象的 Lake Formation 权限，除非他们获得了这样做的权限。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。

有关数据湖管理员的功能的信息，请参阅Lake Formation 的 (p. 140). 有关指定用户为数据湖管理员的信息，请参阅创建数据湖管理员 (p. 12).

Lake Formation

Amazon 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

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。您可以在控制台上使用蓝图来发现、清理、转换和摄取数据。您还可以启用或禁用单个 Lake Formation 用户对控制台的访问权限。

Lake Formation API 和命令行界面

Lake Formation 通过多个特定于语言的软件开发工具包和Amazon Command Line Interface(Amazon CLI)。Lake Formation API 与Amazon GlueAPI。Lake Formation API 主要关注管理 Lake Formation 权限，而Amazon GlueAPI 提供用于对您的数据定义、安排和运行 ETL 操作的数据目录 API 和托管基础设施。

有关的信息Amazon GlueAPI，请参阅Amazon Glue开发人员指南. 有关使用Amazon CLI，请参阅Amazon CLI命令参考.

其他 Amazon 服务

Lake Formation 使用以下服务：

- Amazon Glue编排作业和爬网程序以使用Amazon Glue转换。
- IAM向 Lake Formation 委托人授予权限策略。Lake Formation 权限模型增强了 IAM 权限模型以保护数据湖的安全。

设置 Amazon Lake Formation

完成以下任务以开始设置 Lake Formation：

1. 完成初始 Amazon 配置任务 (p. 9)
2. 为工作流创建 IAM 角色 (p. 11)
3. 创建数据湖管理员 (p. 12)
4. 更改默认权限模型 (p. 14)
5. the section called “创建更多 Lake Formation 用户” (p. 15)
6. the section called “为数据湖配置 AAmazon S3 ervice 位置” (p. 16)
7. the section called “为使用受管控表做好准备” (p. 17)
8. the section called “（可选）外部数据筛选设置” (p. 21)
9. the section called “（可选）授予对数据目录加密密钥的访问权限” (p. 21)

使用创建资源Amazon CloudFormation模板

你也可以使用Amazon CloudFormation模板用于在您的账户中执行最初的Lake Formation 设置。

> Note
>
> 这些区域有：Amazon CloudFormationstack 执行上述步骤 2 到 7，步骤 4 除外。执行更改默认权限模型 (p. 14)从 Lake Formation 控制台手动获取。

1. 登录到Amazon CloudFormation控制台处的https://console.aws.amazon.com/cloudformation作为美国东部（弗吉尼亚北部）区域的 IAM 用户。
2. 选择启动堆栈.
3. 选择下一步在创建堆栈屏幕。
4. 输入堆栈名称。
5. 适用于DatalakeAdminName和DatalakeAdminPassword，输入您的 IAM 用户名和密码。
6. 适用于DatalakeUser1 个名字和DatalakeUser1 个密码，输入您的 IAM 用户名和数据湖分析师用户的密码。
7. 适用于DataLakeBucketName，输入将要创建的新存储桶名称。
8. 选择 Next（下一步）。
9. 在下一页上，选择下一步.
10. 查看最后一页上的详细信息，然后选择我承认这一点Amazon CloudFormation可能会创建 IAM 资源。
11. 选择Create（创建）。

    创建堆栈可能需要长达两分钟。

清理资源

如果你想清理Amazon CloudFormation堆栈资源：

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. 删除存储在设备上的Amazon CloudFormation堆栈。这将删除堆栈创建的所有资源。

完成初始 Amazon 配置任务

使用Amazon Lake Formation您必须首先完成以下任务：

主题

注册Amazon

当您注册时Amazon，您的Amazon Web Services 账户已自动注册所有服务Amazon，包括Lake Formation. 您只需为使用的服务付费。

如果您已有Amazon Web Services 账户，请跳到下一个任务。如果您还没有 Amazon Web Services 账户，请使用以下步骤创建。

创建 Amazon Web Services 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，您将接到一通电话，要求您使用电话键盘输入一个验证码。

请记下您的 Amazon 账号，因为在下一个任务中您会用到它。

创建 IAM 管理员用户

中的服务Amazon（例如 Lake Formation）要求您在访问时提供凭证，以便服务可以确定您是否有权限访问其资源。我们建议不要访问Amazon使用您的证书Amazonaccount. 相反，我们建议您使用 Amazon Identity and Access Management (IAM)。您可以创建 IAM 用户，然后将该用户添加到具有管理权限的 IAM 组或授予此用户管理权限。然后你可以访问Amazon使用 IAM 用户的证书。

如果你注册了Amazon但尚未为自己创建管理 IAM 用户，您可以使用 IAM 控制台自行创建。如果您不熟悉如何使用控制台，请参阅使用 Amazon Web Services Management Console中的概述内容。

自行创建管理员用户并将该用户添加到管理员组（控制台）

1. 选择 Root user（根用户）并输入您的 Amazon Web Services 账户 电子邮件地址，以账户拥有者身份登录 IAM 控制台。在下一页上，输入您的密码。

   Note

   强烈建议您遵守以下使用 **`Administrator`** IAM 用户的最佳实践，妥善保存根用户凭证。只在执行少数账户和服务管理任务时才作为根用户登录。
2. 在导航窗格中，选择 Users（用户），然后选择 Add users（添加用户）。
3. 对于 User name（用户名），输入 **`Administrator`**。
4. 选中 Amazon Web Services Management Console access (Amazon Web Services Management Console 管理控制台访问）旁边的复选框。然后选择自定义密码，并在文本框中输入新密码。
5. （可选）默认情况下，Amazon要求新用户在首次登录时创建新密码。您可以清除 User must create a new password at next sign-in（用户必须在下次登录时创建新密码）旁边的复选框以允许新用户在登录后重置其密码。
6. 选择 Next:。Permissions (下一步：权限)。
7. 在设置权限下，选择将用户添加到组。
8. 选择创建组。
9. 在 Create group（创建组）对话框中，对于 Group name（组名称），输入 **`Administrators`**。
10. 选择 Filter policies（筛选策略），然后选择 Amazon managed - job function（Amazon 托管 – 工作职能）以筛选表内容。

11. 在策略列表中，选中AdministratorAccess. 然后选择 Create group（创建组）。

    Note

    您必须先激活 IAM 用户和角色对账单的访问权限，然后才能使用 `AdministratorAccess` 权限访问 Amazon Billing and Cost Management 控制台。为此，请按照"向账单控制台委派访问权限"教程第 1 步中的说明进行操作。
12. 返回到组列表中，选中您的新组所对应的复选框。如有必要，选择 Refresh（刷新）以在列表中查看该组。
13. 选择 Next:。标签。
14. (可选) 通过以键值对的形式附加标签来向用户添加元数据。有关在 IAM 中使用标签的更多信息，请参阅 IAM 用户指南中的标记 IAM 实体。
15. 选择 Next:。审核查看要添加到新用户的组成员资格的列表。如果您已准备好继续，请选择 Create user（创建用户）。

您可使用这一相同的流程创建更多组和用户，并允许您的用户访问 Amazon Web Services 账户 资源。要了解有关使用策略限制用户对特定 Amazon 资源的权限的信息，请参阅访问管理和示例策略。

作为 IAM 用户登录

通过选择 IAM user (IAM 用户) 并输入您的Amazon Web Services 账户 ID 或账户别名来登录 IAM 控制台。在下一页上，输入您的 IAM 用户名和密码。

Note

为方便起见，Amazon登录页面使用浏览器 Cookie 记住您的 IAM 用户名和账户信息。如果您之前以其他用户身份登录过，请选择此按钮下面的登录链接，返回登录主页。在此处，您可以输入要重新导向到您账户 IAM 用户登录页面的 Amazon Web Services 账户 ID 或账户别名。

为工作流创建 IAM 角色

与Amazon Lake Formation，你可以使用以下方法导入你的数据工作流由执行的Amazon Glue爬网程序。工作流程定义了将数据导入数据湖的数据源和计划。您可以使用以下命令轻松定义工作流程蓝图，或者是 Lake Formation 提供的模板。

在创建工作流程时，您必须为其指定Amazon Identity and Access Management(IAM) 角色，用于向 Lake Formation 授予摄取数据的必要权限。

以下过程假定您熟悉 IAM。

为工作流创建 IAM 角色

1. 使用打开 IAM 控制台https://console.amazonaws.cn/iam并以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者作为 IAM 用户使用`AdministratorAccess` Amazon托管策略。
2. 在导航窗格中，选择角色，那么创建角色.
3. 在存储库的创建角色页面，选择Amazon服务，然后选择Glue. 选择 Next（下一步）。
4. 在存储库的添加权限页面上，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，请选择创建角色.
5. Back角色页面，搜索`LakeFormationWorkflowRole`然后选择角色名称。
6. 关于角色摘要页面，在Permissions (权限)选项卡，选择添加内联策略，然后添加以下内联策略。该策略的建议名称为`LakeFormationWorkflow`.

    Important

    在以下策略中，替换*<account-id>*使用有效的Amazon Web Services 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"
            ]
        }
    ]
}
```

以下是此策略中权限的简要描述：

- `lakeformation:GetDataAccess`允许工作流创建的作业写入目标位置。
- `lakeformation:GrantPermissions`使工作流程能够授予`SELECT`对目标表的权限。
- `iam:PassRole`使服务能够代入该角色`LakeFormationWorkflowRole`创建搜寻器和作业（工作流实例），并将角色附加到创建的搜寻器和作业。

7. 验证该角色`LakeFormationWorkflowRole`附加了两种策略。
8. 如果您要摄取数据湖位置之外的数据，请添加一个内联策略，授予读取源数据的权限。

创建数据湖管理员

数据湖管理员最初是唯一的Amazon Identity and Access Management(IAM) 用户或角色，可以向任何委托人（包括自己）授予 Lake Formation 数据位置和数据目录资源的权限。有关数据湖管理员功能的更多信息，请参阅Lake Formation 的 (p. 140). 默认情况下，Lake Formation 允许您创建最多 30 个数据湖管理员。

您可以使用 Lake Formation 控制台或`PutDataLakeSettings`Lake Formation API 的操作。

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

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

如果您向用户授予`AWSLakeFormationDataAdmin`政策，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 IAM 用户尚不存在，请使用 IAM 控制台创建该用户。否则，请查看将担任数据湖管理员的现有 IAM 用户。

   Note

   建议您不要选择 IAM 管理用户 (用户)`AdministratorAccess` Amazon托管策略）成为数据湖管理员。

附加以下内容Amazon向用户提供托管策略：

策略	必需？	注意
AWSLakeFormationDataAdmin	必需	基本的数据湖管理员权限。
AWSGlueConsoleFullAccess, CloudWatchLogsReadOnlyAccess	可选	如果数据湖管理员要对根据 Lake Formation 蓝图创建的工作流进行故障排除，请附上这些策略。这些策略使数据湖管理员能够在Amazon Glue控制台和Amazon CloudWatch Logs控制台。有关工作流的信息，请参阅。使用工作流程导入数据 (p. 134).
AWSLakeFormationCrossAccountManager	可选	附加此策略以使数据湖管理员能够授予和撤销对 Data Catalog 资源的跨账户权限。有关更多信息，请参阅 the section called “跨账户访问权限” (p. 242)。
AmazonAthenaFullAccess	可选	如果数据湖管理员要在中运行查询，请附上此策略Amazon Athena.

2. 附上以下内联策略，该策略授予数据湖管理员创建 Lake Formation 服务相关角色的权限。该策略的建议名称为LakeFormationSLR.

   服务相关角色使数据湖管理员可以更轻松地向 Lake Formation 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” (p. 256).

   Important

   在以下所有政策中，替换*<account-id>*使用有效的Amazon账号。

```
{
    "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）附加以下内容PassRole将内联策略发送到用户。此策略允许数据湖管理员创建和运行工作流程。这些区域有：iam:PassRole权限使工作流程能够代入该角色LakeFormationWorkflowRole创建爬虫和作业，并将角色附加到创建的搜寻器和作业。该策略的建议名称为UserPassRole.

Important

Replace<*account-id*>使用有效的Amazon账号。

```
{
    "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，请附上此额外的内联政策。此策略允许数据湖管理员查看和接受Amazon Resource Access Manager(Amazon RAM) 资源共享邀请。另外，对于数据湖管理员来说Amazon Organizations管理账户，该政策包括允许向组织提供跨账户授予的权限。有关更多信息，请参阅 the section called “跨账户访问权限” (p. 242)。

   该策略的建议名称为RAMAccess.

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

5. 打开Amazon Lake Formation控制台处的https://console.aws.amazon.com/lakeformation/并以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者以任何 IAM 管理用户身份。
6. 如果欢迎来到Lake Formation窗口出现，选择您在步骤 1 中创建或选择的 IAM 用户，然后选择试用.
7. 如果您没有看到欢迎来到Lake Formation窗口，然后执行以下步骤来配置 Lake Formation 管理器。

   a. 在导航窗格中的下，Permissions (权限)，选择管理角色和任务. 在数据湖管理员控制台页面上的，选择选择管理员.

   b. 在管理数据湖管理员对话框，为IAM 用户和角色，选择您在步骤 1 中创建或选择的 IAM 用户，然后选择Save（保存）.

更改默认权限模型

Lake Formation 一开始启用 “仅使用 IAM 访问控制” 设置，以便与现有设置兼容Amazon Glue Data Catalog行为。我们建议您禁用这些设置，以启用具有 Lake Formation 权限的基于标签的精细访问控制。

有关更多信息，请参阅 the section called “更改数据湖的默认安全设置” (p. 253)。

Important

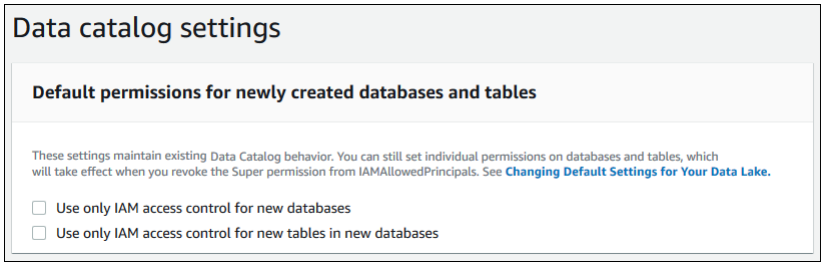
如果你有Amazon Glue Data Catalog数据库和表，请勿按照本节中的说明进行操作。而是应按照升级Amazon GlueLake Formation 模型的数据权限 (p. 22)中的说明操作。

Warning

如果您已实现自动化，可以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。只有在您修改了现有流程或向所需负责人授予明确的 Lake Formation 权限后才能继续。有关 Lake Formation 权限的更多信息，请参阅the section called “Lake Formation 权限参考” (p. 167).

更改原定设置数据目录设置

1. 在 Lake Formation 控制台中继续https://console.aws.amazon.com/lakeformation/. 确保您以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者作为 IAM 用户使用`AdministratorAccess` Amazon托管策略。
2. 修改数据目录设置：
   a. 在导航窗格中的下，Data Cat，选择设置.
   b. 清除两个复选框并选择Save（保存）.



3. REVO`IAMAllowedPrincipals`数据库创建者的权限。
   a. 在导航窗格中的下，Permissions (权限)，选择管理角色和任务.
   b. 在管理角色和任务控制台页面，在数据库创建者部分，选择`IAMAllowedPrincipals`分组，然后选择REVO.

      这些区域有：REVO出现权限对话框，显示`IAMAllowedPrincipals`有创建数据库权限。
   c. 选择REVO.

创建更多 Lake Formation 用户

创建 IAM 用户以访问中的数据湖Amazon Lake Formation. 此用户拥有查询数据湖的最低权限集。

创建有权访问 Lake ForMation 数据的非管理员用户

1. 使用打开 IAM 控制台https://console.amazonaws.cn/iam并以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者作为 IAM 用户使用`AdministratorAccess` Amazon管理的策略。
2. 选择用户，然后添加用户.
3. 输入用户的名称，然后选择密码-Amazon Web Services Management Console访问访问方法。配置用户密码要求。您还可选择启用访问密钥-编程访问对于这个用户。

   选择下一步：权限。
4. 在 Set permissions (设置权限) 下，选择 Attach existing policies directly (直接附加现有策略)。Enter`Athena`中的筛选策略文本字段。在结果列表中，选中`AmazonAthenaFullAccess`.

5. 选择创建策略按钮。在创建策略页面上，选择JSON选项卡。将下面的代码复制并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. 选择下一步按钮在底部，直到您看到查看策略页. 输入策略的名称，例如`DatalakeUserBasic`. 选择创建策略，然后关闭 “策略” 选项卡或浏览器窗口。
7. 回到 IAM添加用户窗口，输入`datalake`中的筛选策略搜索字段。如果新创建的策略未出现在结果列表中，请选择 “刷新” 或 “页面重新加载” 按钮。选中复选框中的`DatalakeUserBasic`政策。
8. 选择下一篇：标签和下一步：Review.
9. 在 Review (检查) 页面上，您应该看到两个`DatalakeUserBasic`政策和`AmazonAthenaFullAccess`为用户选择了策略。选择创建用户以完成设置。

为数据湖配置 AAmazon S3 ervice 位置

要使用 Lake Formation 来管理和保护数据湖中的数据，您必须先注册 Amazon S3 地点。当您注册位置时，该 Amazon S3 路径和该路径下的所有文件夹都会被注册，这使得 Lake Formation 能够强制执行存储级别的权限。当用户从像 Amazon Athena 这样的集成引擎请求数据时，Lake Formation 提供数据访问权限，而不是使用用户的权限。

注册位置时，需要指定一个 IAM 角色来授予该位置的读取/写入权限。Lake Formation 在向集成提供临时证书时扮演这个角色Amazon请求访问已注册的 Amazon S3 位置中的数据的服务。您可以指定 Lake Formation 服务相关角色 (SLR) 或自行创建。

在以下情况下使用自定义角色：

- 您计划在已注册的 Amazon S3 位置创建受管控的表。用户定义的角色必须包含用于添加日志的策略 CloudWatch 除 SLR 权限之外的日志和发布指标。以授予必要设置的内联策略为例 CloudWatch 权限，请参阅注册位置时使用的角色的要求 (p. 102).
- Amazon S3 位置存在于不同的账户中。有关详细信息，请参阅the section called “在另一个位置注册 Amazon S3Amazon帐户” (p. 108)。
- Amazon S3 位置包含使用加密的数据Amazon 托管式密钥. 有关详细信息，请参阅 注册加密的 Amazon S3 位置 (p. 105) 和 注册加密 Amazon S3 位置Amazon账户 (p. 109)。
- 您计划使用亚马逊 EMR 访问Amazon S3 ervice（ Amazon EMR ）位置。有关角色要求的更多信息，请参阅Lake Formation 的 IAM 角色中的Amazon EMR 管理指南.

您选择的角色必须具有必要的权限，如中所述注册位置时使用的角色的要求 (p. 102). 有关如何注册 Amazon Service 位置的说明，请参阅将 Amazon S3 位置添加到您的数据湖 (p. 102).

为使用受管控表和行级安全性做好准备

使用 Lake Formation 管理的表、行级筛选和存储优化功能需要额外配置。

主题

为使用受管控表做好准备

要在 Lake Formation 中创建受管控的表，您必须先在 Lake Formation 中注册一个 Amazon S3 位置，然后指定一个包含所有所需权限的角色，如前所述为数据湖配置 AAmazon S3 ervice 位置 (p. 16). 然后，您需要向将与受管控表交互的用户或角色授予权限。有关数据访问权限的更多信息，请参阅底层数据访问控制。

要创建受管理的表，用户必须是数据湖管理员或具有以下权限的用户：

- Lake FormationCREATE_TABLE对目标数据库的权限
  - 这些区域有：Amazon Identity and Access Management(IAM) 权限glue:CreateTable
  - 中的数据位置权限Lake Formation，如中所述授予数据位置权限 (p. 141). 数据位置权限控制创建或更改指向特定 Amazon S3 位置的数据目录资源的能力。

  要访问受管表中的数据，委托人需要对受管辖的表具有 SELECT 权限和 IAM 权限才能调用：

```
lakeformation:StartQueryPlanning
lakeformation:GetQueryState
lakeformation:GetWorkUnits

lakeformation:GetWorkUnitResults

lakeformation:StartTransaction

lakeformation:CommitTransaction

lakeformation:CancelTransaction
lakeformation:ExtendTransaction
```

为用户创建和使用受管理的表创建和分配角色

1. 使用打开 IAM 控制台https://console.amazonaws.cn/iam并以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者作为 IAM 用户使用 IAM 用户AdministratorAccess Amazon托管策略。
2. 在导航窗格中，选择角色，那么创建角色.
3. 在附加权限策略部分，选择创建策略. 在新打开的浏览器窗口中，创建用于您的角色的新策略。

   a. 在创建策略页面上，选择 JSON 选项卡。将以下 JSON 代码复制到策略编辑器字段中。

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

```
            "Effect": "Allow",
            "Action": [
                "glue:GetTable",
                "glue:GetPartitions",
                "glue:UpdateTable"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults",
                "lakeformation:GetQueryStatistics",
                "lakeformation:StartTransaction",
                "lakeformation:CommitTransaction",
                "lakeformation:CancelTransaction",
                "lakeformation:ExtendTransaction",
                "lakeformation:DescribeTransaction",
                "lakeformation:ListTransactions",
                "lakeformation:GetTableObjects",
                "lakeformation:UpdateTableObjects",
                "lakeformation:DeleteObjectsOnCancel"
            ],
            "Resource": "*"
        }
    ]
}
```

以下是此策略中权限的简要描述：

- `lakeformation:StartQueryPlanning`允许委托人提交请求以处理查询语句。
- `lakeformation:GetQueryState`允许委托人查看先前提交的查询的状态。
- `lakeformation:GetWorkUnits`允许负责人检索由生成的工作单元`StartQueryPlanning`operation.
- `lakeformation:GetWorkUnitResults`允许委托人查看查询生成的工作单元。
- `lakeformation:GetQueryStatistics`允许委托人检索有关计划和执行查询的统计信息。
- `lakeformation:StartTransaction`允许委托人和作业启动事务。
- `lakeformation:CommitTransaction`允许委托人和任务提交事务。
- `lakeformation:CancelTransaction`允许委托人和任务在提交之前停止事务。
- `lakeformation:ExtendTransaction`允许委托人和任务指明指定的交易仍处于活动状态，不应取消。
- `lakeformation:DescribeTransaction`允许委托人和任务列出有关交易的信息。
- `lakeformation:ListTransactions`允许委托人和作业查看有关事务及其状态的元数据。
- `lakeformation:GetTableObjects`允许负责人和作业列出存储在数据湖中的表对象。
- `lakeformation:UpdateTableObjects`允许委托人和作业更新存储在数据湖中的表对象。
- `lakeformation:DeleteObjectsOnCancel`允许委托人和任务指定 Amazon S3 对象的列表，这些对象将在当前事务期间写入，如果交易被取消，则可以自动删除。

b. 对于需要管理受管表的数据压缩和垃圾收集设置的用户，请向上述策略添加以下权限：

```
"lakeformation:UpdateTableStorageOptimizer",
"lakeformation:ListTableStorageOptimizers"
```

c. 选择 Next:Tags (下一步: 标签)。

d. 您可以选择性地添加标签，然后选择后续：审核。

e. 在 Review policy (查看策略) 页面上，输入策略的名称，例如`LakeFormationGovernedTables`，然后选择创建策略.

f. 您可以关闭此窗口，然后返回创建角色页.

4. 在存储库的创建角色页面，选择刷新按钮，然后搜索`LakeFormationGovernedTables`您在上一步中创建的策略。选中列表中策略名称旁边的复选框。
5. 完成创建角色通过选择向导下一步直到你到达审核页. 输入角色的名称，例如`LakeFormationTransactionsRole`. 要完成，请选择创建角色.
6. Back角色页面，搜索`LakeFormationTransactionsRole`然后选择角色名称。
7. 关于角色摘要页面，在Permissions (权限)选项卡，验证该角色是否具有`LakeFormationGovernedTables`附加了策略。

现在，您可以将此角色分配给使用受管表的委托人。

为使用受管控表的自动数据压缩做好准备

要为受管理的表配置数据压缩，委托人必须满足以下条件：

- 成为创建表的用户或成为数据湖管理员用户
- HAVOKE`glue:UpdateTable`、`glue:GetTable`和Lake Formation`ALTER`桌上的权限

此外，向 Lake Formation 注册 Amazon S3 数据湖位置时使用的角色必须包含以下权限才能使用数据压缩：

- `s3:PutObject` 和 `lakeformation:UpdateTableObjects`
- `lakeformation:StartTransaction`、`lakeformation:CommitTransaction`、`lakeformation:CancelTra`和`logs:CreateLogGroup`、`logs:CreateLogStream`、`logs:PutLogEvents`，到 `"arn:aws:logs:*:*<**ACCOUNT ID**>*:log-group:/aws-lakeformation-acceleration/compaction/logs:*"` 如以下示例所示。

```
{
            "Statement": [
        {
            "Effect":"Allow",
            "Action":[
                "s3:PutObject"
            ],
            "Resource": "arn:aws:s3:::<bucket>/<prefix>/*"
        },
        {
            "Effect":"Allow",
            "Action":[
                "lakeformation:StartTransaction",
                "lakeformation:CommitTransaction",
                "lakeformation:CancelTransaction",
                "lakeformation:DeleteObjectsOnCancel"
            ],
            "Resource": "*"
        },
        {
            "Effect":"Allow",
            "Action":[
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:*:*<**ACCOUNT ID**>*:log-group:/aws-lakeformation-
acceleration/compaction/logs:*"
        }
```

```
    ]
        }
```

- 如果数据目录已加密，Amazon KMS密钥策略必须包括信任关系`lakeformation.amazonaws.com`，例如以下示例。

```
{
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:DescribeKey"
            ],
            "Resource": "*"
        }
```

为使用行级别安全性做好准备

当您授予 Lake Formation 对数据目录表的权限时，可以加入数据筛选规范，以限制对查询结果中和查询结果中的某些数据的访问Amazon GlueETL 作业。Lake Formation 使用数据筛选来实现列级安全、行级安全和单元格级安全。您可以通过创建命名来实现列级、行级和单元格级安全性数据筛选器并在您授予数据过滤器时指定数据筛选器`SELECT`桌上有Lake Formation 许可。创建数据筛选器时，需要为需要包含的行提供一组列和一个筛选表达式。这允许您限制对查询结果中以及与 Lake Formation 集成的引擎（例如 Athena 和 Amazon GlueETL 作业。

有关数据筛选器的更多信息，请参阅Lake Formation 中的数据过滤和细胞级安全 (p. 206).

为表配置行级别安全性

1. 确定您要限制访问的内容并创建数据筛选器。有关创建数据筛选器的说明，请参阅the section called “创建数据筛选器” (p. 212).
2. Grant`DESCRIBE`向能够查看数据筛选器的用户提供数据筛选器的权限。

   创建数据筛选器时，只有您可以查看它。要允许其他委托人查看和使用数据筛选器，您可以授予`DESCRIBE`许可。
3. 在授予数据筛选器时指定数据筛选器`SELECT`允许校长在桌子上使用 Lake Formation。
4. 向将使用单元级筛选器查询表的委托人分配 IAM 权限。使用单元格级筛选查询表的委托人必须具有以下 IAM 权限：

```
lakeformation:StartQueryPlanning
                    lakeformation:GetQueryState
                    lakeformation:GetWorkUnits
                    lakeformation:GetWorkUnitResults
```

（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。如果您不选择加入，则外部引擎将无法访问已向 Lake Formation 注册的 Amazon S3 位置中的数据。

Lake Formation 支持列级权限，用于限制对表中特定列的访问。集成分析服务，例如Amazon Athena、Amazon Redshift Spectrum 和 Amazon EMR 从Amazon Glue Data Catalog. 实际筛选查询响应中的列是集成服务的责任。第三方管理员有责任正确处理权限，以避免未经授权访问数据。

选择允许第三方引擎访问和筛选数据（控制台）

1. 在 Lake Formation 控制台中继续https://console.aws.amazon.com/lakeformation/. 确保您以拥有 Lake Formation 的 IAM 权限的委托人身份登录`PutDataLakeSettings`API 操作。您在中创建的 IAM 管理员用户创建 IAM 管理员用户 (p. 10)有这个权限。
2. 在导航窗格中的下，Permissions (权限)，选择外部数据筛选.
3. 在存储库的外部数据筛选页面上，请执行以下操作：
   a. 选中复选框允许外部引擎筛选向 Lake Formation 注册的 Amazon S3 位置的数据.
   b. Enter会话标签值为第三方引擎定义。
   c. 适用于Amazon账户 ID，输入允许第三方引擎访问在 Lake Formation 注册的地点的账户 ID。按Enter在每个账户 ID 之后。
   d. 选择Save（保存）。

（可选）授予对数据目录加密密钥的访问权限

如果Amazon Glue Data Catalog已加密，授权Amazon Identity and Access Management上的 (IAM) 权限Amazon KMS对于任何需要授予 Lake Formation 对数据目录数据库和表的权限的校长来说都是密钥。

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

升级Amazon Glue的数据权限Amazon Lake Formation模型

Amazon Lake Formation权限支持对数据湖中数据的精细访问控制。你可以使用 Lake Formation 权限模型来管理你现有的Amazon Glue Data CatalogAmazon Simple Storage (Amazon S3) 中的对象和数据位置。

Lake Formation 权限模型使用粗粒度Amazon Identity and Access Management用于 API 服务访问的 (IAM) 权限。它限制了您的用户和这些服务可以通过Lake Formation功能访问的数据。相比之下，Amazon Glue模型通过以下方式授予数据访问权限访问权限的精访问权限的精访问控制IAM 权限。要进行切换，请按照本指南中的步骤操作。

有关更多信息，请参阅 the section called “Lake Formation 访问控制概述” (p. 234)。

主题

关于升级到 Lake Formation 权限模型

为了保持向后兼容性Amazon Glue，默认情况下，Amazon Lake FormationGRANT`Super`的许可`IAMAllowedPrincipals`对所有现有内容进行分组Amazon Glue数据目录资源，并授予`Super`对新数据目录资源的权限，如果仅使用 IAM 访问控制设置已启用。这实际上导致对数据目录资源和 Amazon S3 位置的访问完全由以下人员控制Amazon Identity and Access Management(IAM) 策略。这些区域有：`IAMAllowedPrincipals`群组包括您的 IAM 策略允许访问您的数据目录对象的所有 IAM 用户和角色。这些区域有：`Super`权限使委托人能够对被授予权限的数据库或表执行所有支持的 Lake Formation 操作。

您可以开始使用 Lake Formation 来管理对数据的访问，方法是在 Lake Formation 中注册现有数据目录资源的位置。开始使用现有的 Lake Formation 权限Amazon Glue数据目录数据库和表，您必须执行以下操作：

1. 确定您的用户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. 对于每个包含数据的 Amazon S3 位置：
   a. 撤OKEE`Super`来自的许可`IAMAllowedPrincipals`对引用该位置的每个数据目录资源进行分组。
   b. 在 Lake Formation 中注册该地点。
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，必须进行精细设置Amazon GlueIAM 中的权限和以前一样。您还必须按照本节所述在 Lake Formation 中复制这些权限。如果新用户拥有本指南中描述的粗略的 IAM 策略，则他们可以列出任何具有`Super`已访问权限`IAMAllowedPrincipals`. 他们还可

以查看这些资源的元数据。但是，除非您向 Lake Formation 注册 Amazon S3 位置，否则他们无法自行查询数据。

按照本节中的步骤升级到 Lake Formation 权限模型。从这里开始the section called “第 1 步：列出现有权限” (p. 23).

第 1 步：列出用户和角色的现有权限

要开始使用Amazon Lake Formation使用您现有的权限Amazon Glue数据库和表，必须首先确定用户的现有权限。

Important

在您开始之前，请确保您已完成中的任务设置 Amazon Lake Formation (p. 9).

主题

使用 API 操作

使用Amazon Identity and Access Management(IAM)ListPoliciesGrantingServiceAccessAPI 操作用于确定附加到每个委托人（用户或角色）的 IAM 策略。根据结果中返回的策略，您可以确定授予委托人的 IAM 权限。您必须分别为每个委托人调用 API。

Example

以下Amazon CLI示例返回附加到用户的策略`glue_user1`.

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

该命令返回的结果类似于下方内容。

```
{
    "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 Amazon Web Services Management Console

您还可在Amazon Identity and Access Management(IAM) 控制台，在访问顾问用户或角色上的选项卡摘要页：

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。

   在此过程中，请记住除了检查用户之外还要检查角色，因为您的数据处理任务可能会担任访问数据的角色。

使用 Amazon CloudTrail

确定现有权限的另一种方法是查看Amazon CloudTrail为了Amazon GlueAPI 调用其中`additionaleventdata`日志字段包含`insufficientLakeFormationPermissions`条目。此条目列出了用户需要 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完整的用户及其权限列表。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，请参阅使用查看事件 CloudTrail 事件历史记录中的Amazon CloudTrail用户指南.

接下来，你可以设置 Lake Formation 权限以匹配Amazon Glue权限。请参阅 第 2 步：设置等效的Lake Formation 权限 (p. 24)。

第 2 步：设置等效的Lake Formation 权限

使用在中收集的信息第 1 步：列出用户和角色的现有权限 (p. 23)，GRANTAmazon Lake Formation匹配的权限Amazon Glue权限。使用以下其中一种方法来执行授权：

- 使用 Lake Formation 控制台或Amazon CLI.

  请参阅 the section called “授予和撤消数据目录权限” (p. 145)。
- 使用`GrantPermissions`要么`BatchGrantPermissions`API 操作。

  请参阅 API 权限 (p. 271)。

有关更多信息，请参阅 FormLake Formation 权限概述 (p. 138)。

设置 Lake Formation 权限后，继续前往第 3 步：授予用户使用 Lake Formation 的 IAM 权限 (p. 24).

第 3 步：授予用户使用 Lake Formation 的 IAM 权限

使用Amazon Lake Formation权限模型，委托人必须具有Amazon Identity and Access Management(IAM) 对 Lake Formation API 的权限。

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。将策略命名为`LakeFormationDataAccess`。

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

接下来，一次升级到一个数据位置的 Lake Formation 权限。请参阅 第 4 步：将您的数据存储切换到 Lake Formation 权限模型 (p. 25)。

第 4 步：将您的数据存储切换到 Lake Formation 权限模型

一次升级到一个数据位置的 Lake Formation 权限。为此，请重复整个部分，直到您已注册数据目录引用的所有 Amazon Simple Storage Service (Amazon S3) 路径。

主题

验证Lake Formation

在注册地点之前，请执行验证步骤，确保正确的负责人拥有所需的Lake Formation 权限，并且不向不应拥有该权限的负责人授予任何Lake Formation 队权限。使用Lake Formation`GetEffectivePermissionsForPath`API 操作，识别引用 Amazon S3 位置的数据目录资源，以及对这些资源拥有权限的委托人。

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

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

注意`profile`选项。建议您以数据湖管理员的身份运行该命令。

下面是来自返回的结果的摘录。

```
{
        "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Amazon Glue数据目录已加密，`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`Super`来自的许可`IAMAllowedPrincipals`在您为该位置标识的每个表和数据库上。

Warning

如果您已实现自动化，可以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。只有在您修改了现有流程或向所需负责人授予明确的 Lake Formation 权限后才能继续。有关Lake Formation 权限的更多信息，请参阅the section called “Lake Formation 权限参考” (p. 167).

撤KEKE**`Super`**从**`IAMAllowedPrincipals`**在桌子上

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份登录。
2. 在导航窗格中，选择表。
3. 在存储库的表页面上，选择所需表格旁边的单选按钮。
4. 在存储库的操作菜单，选择撤KE.
5. 在撤访问权限对话框中IAM 用户和角色列表，向下滚动到Group标题，然后选择IAMAllowedPrincipals.
6. UNDER表权限，请确保超级已选中，然后选择撤KE.

撤KEKE**`Super`**从**`IAMAllowedPrincipals`**在数据库上

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份登录。
2. 在导航窗格中，选择 Databases（数据库）。
3. 在存储库的数据库页面上，选择所需数据库旁边的单选按钮。
4. 在 Actions 菜单上选择 Edit。
5. 在存储库的编辑数据库页面，清除仅对该数据库中的新表使用 IAM 访问控制，然后选择Save（保存）.
6. Back数据库页面上，确保数据库仍处于选中状态，然后在操作菜单，选择撤KE.
7. 在撤访问权限对话框中IAM 用户和角色列表，向下滚动到Group标题，然后选择IAMAllowedPrincipals.
8. UNDER数据库权限，请确保超级已选中，然后选择撤KE.

为您的 Amazon S3 位置开启 Lake Formation 权限

接下来，在 Lake Formation 中注册 Amazon S3 地点。要执行此操作，您可以使用中描述的流程将 Amazon S3 位置添加到您的数据湖 (p. 102). 或者，使用`RegisterResource`API 操作，如中所述凭据自动售货机 API (p. 283).

Note

如果已注册了上级地点，则无需注册子位置。

完成这些步骤并测试您的用户是否可以访问他们的数据后，您已成功升级到 Lake Formation 权限。继续下一步，第 5 步：保护新的数据目录资源 (p. 27).

第 5 步：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。关闭仅供使用的选项Amazon Identity and Access Management(IAM) 对新数据库和表的访问控制。

Warning

如果您已实现自动化，可以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。只有在您修改了现有流程或向所需负责人授予明确的 Lake Formation 权限后才能继续。有关Lake Formation 权限的更多信息，请参阅the section called “Lake Formation 权限参考” (p. 167).

更改默认数据目录设置

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 作为 IAM 管理用户（用户）登录`Administrator`或者其他用户使用`AdministratorAccess` Amazon管理的策略）。
2. 在导航窗格中，选择 Settings (设置)。
3. 在存储库的数据目录设置页面上，清除两个复选框，然后选择Save（保存）.

下一步是授予用户将来访问其他数据库或表的权限。请参阅 第 6 步：为用户提供新的 IAM 策略，以便future访问数据湖 (p. 27)。

第 6 步：为用户提供新的 IAM 策略，以便future 访问数据湖

要授予您的用户将来访问其他 Data Catalog 数据库或表的权限，您必须为他们提供粗略的权限Amazon Identity and Access Management以下是 (IAM) 内联策略。将策略命名为 `GlueFullReadAccess`。

Important

如果您在撤消之前将此策略附加到某个用户`Super`从`IAMAllowedPrincipals`在数据目录中的每个数据库和表上，该用户可以查看任何资源的所有元数据`Super`被授予了`IAMAllowedPrincipals`.

```
{
    "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

此步骤和之前步骤中指定的内联策略包含最低的 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Lake Formation 角色和 IAM 权限参考 (p. 322).

接下来，继续步骤 7：清理现有的 IAM 策略 (p. 28).

步骤 7：清理现有的 IAM 策略

在您设置之后Amazon Lake Formation权限，然后你创建并附加粗粒度访问控制Amazon Identity and Access Management(IAM) 策略，完成以下最后一步：

- 从用户、群组和角色中移除旧的访问权限的精访问权限的精访问控制您在 Lake Formation 中复制的 IAM 策略。

这样，您可以确保这些委托人不再能够直接访问Amazon Simple Storage Service (Amazon S3) 中的数据。然后，您可以完全通过 Lake Formation 管理这些负责人的数据湖访问权限。

教程

以下教程分为三个轨道并提供 step-by-step 有关如何使用构建数据湖、采集数据、共享和保护数据湖的说明 Amazon Lake Formation：

> Note
>
> 您浏览教程的顺序不重要。您在第一个教程中创建的用户可以在第二个教程中使用。其余教程包括 Amazon CloudFormation用于创建所需资源的模板。

1. 构建数据湖并采集数据：学习构建数据湖并使用蓝图移动、存储、编目、清理和组织数据。您还将学习如何设置受管理的表。受管理的表是一种新的 Amazon S3 表类型，支持原子、一致、隔离和持久 (ACID) 事务。

   开始之前，请确保您已完成设置 Amazon Lake Formation (p. 9)中的步骤。

   - 从创建数据湖Amazon CloudTrail资源 (p. 30)

     使用自己的数据湖创建和加载第一个数据湖 CloudTrail 将日志作为数据源。
   - 在 Lake Formation 中从 JDBC 源创建数据湖 (p. 37)

     通过使用 JDBC 可访问的数据存储之一（例如关系数据库）作为数据源来创建数据湖。
   - 在 Lake Formation 中创建受管控表 (p. 45)

     本教程演示如何在 Lake Formation 中设置受管理表。有关受管理表的更多信息，请参见Lake Formation 中的受管桌子 (p. 114).
2. 保护数据湖：学习如何使用基于标签和行级别的访问控制来有效地保护和管理对数据湖的访问。
   - 使用基于 Lake Formation 标签的访问控制来管理数据湖 (p. 60)

     学习如何在 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。
   - 使用行级访问控制保护数据湖 (p. 75)

     了解如何设置行级权限，允许您根据 Lake Formation 中的数据合规性和治理策略限制对特定行的访问。
3. 共享数据：学习如何安全地分享您的数据Amazon Web Services 账户使用基于标记的访问控制（TBAC）并管理对之间共享数据集的粒度权限Amazon Web Services 账户.
   - 使用基于 Lake Formation 标签的访问控制和命名资源共享数据湖 (p. 84)

     在本教程中，您将了解如何在跨安全共享数据Amazon Web Services 账户使用 Lake Formation。
   - 使用 Lake Formation 精细访问控制 (p. 97)

     在本教程中，您将了解如何在管理多个数据集时使用 Lake Formation 快速轻松共享数据集Amazon Web Services 账户和Amazon Organizations.

主题

从创建数据湖Amazon CloudTrail资源

本教程将指导你完成在 Lake Formation 控制台上执行的操作，以便从Amazon CloudTrail源。

创建数据湖的概括步骤

1. 将AmaSimple Storage Service (Amazon S3) 路径注册为数据湖。
2. 授予 Lake Formation 写入数据目录和数据湖中的 Amazon S3 位置的权限。
3. 创建数据库以组织数据目录中的元数据表。
4. 使用蓝图创建工作流程。运行工作流以从数据源提取数据。
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 Simple Storage S3 数据湖中的 Amazon S3 数据湖中的 Amazon S3 数据湖中的 Amazon S3
7. 对于某些数据存储类型，请设置 Amazon Redshift Spectrum 以查询您导入到 Amazon S3 数据湖中的数据。

主题

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	可以创建 IAM 用户和角色以及 Amazon S3 存储桶的用户。UNDAdministratorAccess Amazon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 权限少于 IAM 管理员，但足以管理数据湖。
数据分析人员	可以针对数据湖运行查询的用户。只有足够的权限运行查询。
工作流程	具有运行工作流程所需的 IAM 策略的角色。

先决条件

开始前的准备工作：

- 确保您已完成中的任务设置 Amazon Lake Formation (p. 9).
- 知道您的 CloudTrail 日志。

熟悉Amazon Identity and Access Management(IAM) 是假定的。有关 IAM 的信息，请参阅IAM 用户指南.

第 1 步：创建 IAM 用户作为数据分析师

此用户具有查询数据湖的最低权限集。

1. 使用 https://console.amazonaws.cn/iam 打开 IAM 控制台。以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者作为 IAM 用户使用`AdministratorAccess` Amazon托管策略。
2. 创建一个名为的用户`datalake_user`使用以下设置：
   - 启用Amazon Web Services Management Console访问。
   - 设置密码，不需要重置密码。
   - 将附加到`AmazonAthenaFullAccess` Amazon托管策略。
   - 附加以下内联策略。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：添加读取权限Amazon CloudTrail工作流角色的日志

1. 将下面的内联策略附加到角色`LakeFormationWorkflowRole`. 该政策授予读取您的Amazon CloudTrail日志。将策略命名为 `DatalakeGetCloudTrail`。

   Important

   Replace*<your-s3-cloudtrail-bucket>*使用您的 Amazon S3 位置 CloudTrail DATA

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

2. 确认已附加到角色的策略有三个。

第 3 步：为数据湖创建 Amazon S3 存储桶

创建要作为数据湖根位置的 Amazon S3 存储桶。

1. 在以下位置打开 Amazon S3 控制台https://console.aws.amazon.com/s3/以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10).
2. 选择创建存储桶，然后通过向导创建名为的存储桶*<yourName>*-datalake-cloudtrail，其中*<yourName>*是你的名字缩写和姓氏。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/ 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，在注册和摄取，选择数据湖位置.
3. 选择注册位置然后浏览.
4. 选择*<yourName>*-datalake-cloudtrail您之前创建的存储桶接受默认 IAM 角色AWSServiceRoleForLakeFormationDataAccess，然后选择注册位置.

   有关注册位置的更多信息，请参阅将 Amazon S3 位置添加到您的数据湖 (p. 102).

第 5 步：授予数据位置权限

校长必须具有数据位置权限在数据湖位置上创建指向该位置的数据目录表或数据库。您必须向 IAM 角色授予工作流的数据位置权限，以便工作流程可以写入数据摄取目标。

1. 在导航窗格中，在Permissions (权限)，选择数据位置.
2. 选择Grant，并在授予权限对话框中，进行以下选择：

   a. 适用于IM 用户和角色，选择LakeFormationWorkflowRole.

   b. 适用于存储位置，选择您的*<yourName>*-datalake-cloudtrail存储桶。
3. 选择 Grant（授权）。

有关数据位置权限的更多信息，请参阅Underlying Data Access Control (p. 239).

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，在数据atalog，选择数据库.
2. 选择创建数据库UN数据库详细信息，输入名称`lakeformation_cloudtrail`.
3. 将其他字段留空，然后选择创建数据库.

步骤 7：授予数据权限

您必须授予权限才能在数据目录中创建元数据表。因为工作流程将与角色一起运行`LakeFormationWorkflowRole`，您必须将这些权限授予角色。

1. 在Lake Formation 控制台的导航窗格中，在数据atalog，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Grant在权限标题下。
3. 在授予数据权限对话框中，进行以下选择：

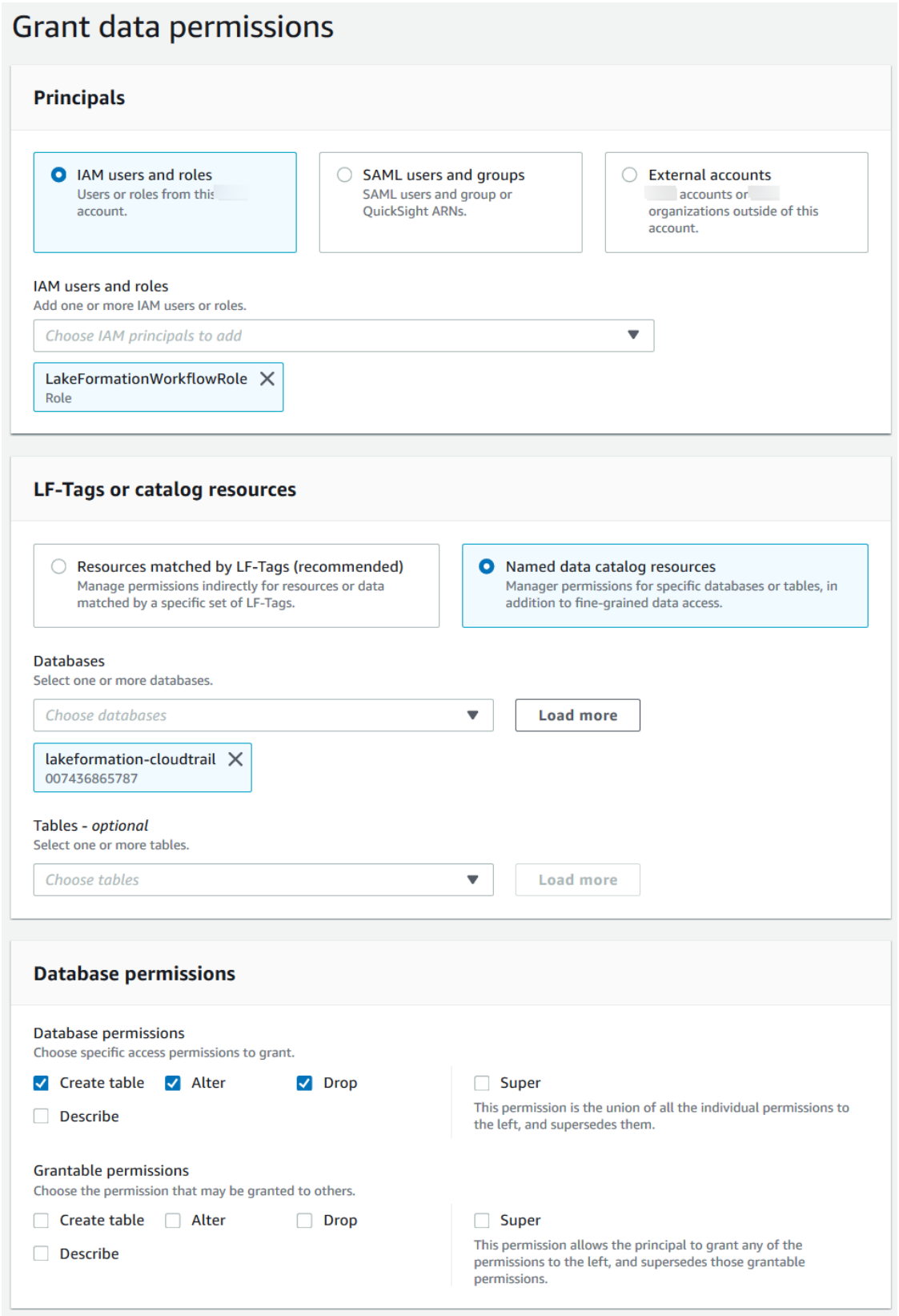
   a. UN委托人，对于IM 用户和角色，选择`LakeFormationWorkflowRole`.

   b. UNLF 标签或目录资源，选择命名数据目录资源.

   c. 适用于数据库，您应该会看到`lakeformation_cloudtrail`数据库已添加。

   d. UN数据库权限，Select创建表、更改, 和Drop，并且clear超级如果它被选中。

   您的授予数据权限对话框现在应该看上去是这样的。


Grant data permissions
Principals
IAM users and roles
Users or roles from this account.
SAML users and groups
SAML users and group or QuickSight ARNs.
External accounts
accounts or 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 Grant（授权）。

有关授予Lake Formation 权限的更多信息，请参阅安全和访问控制Lake Formation 中的元数据和数据 (p. 234).

步骤 8 使用蓝图创建工作流程

为了阅读 CloudTrail 日志，了解它们的结构，在数据目录中创建相应的表，我们需要设置一个工作流，其中包含Amazon Glue爬虫、作业、触发器和工作流程。Lake Formation 的蓝图简化了这一过程。

该工作流程会生成任务、爬网程序和触发器，用于发现数据并将其摄入数据湖中。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. 在存储库的使用蓝图页面，在蓝图类型，选择Amazon CloudTrail.
3. UN导入来源，请选择一个 CloudTrail 来源和开始日期。
4. UN导入目标，请指定以下参数：

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

5. 对于导入频率，选择按需运行.
6. UN导入选项，请指定以下参数：

工作流名称	`lakeformationcloudtrailtest`
IAM 角色	`LakeFormationWorkflowRole`
表格预留	`cloudtrailtest` Note 必须是小写字母。

7. 选择Create，然后等待控制台报告工作流已成功创建。

   Tip

   您是否收到以下错误消息？
   `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...`
   如果是这样，请检查您是否已更换`<account-id>`在数据湖管理员用户的内联策略中Amazon账号

步骤 9 运行工作流程

因为你指定工作流是 run-on-demand，则必须手动启动工作流。

- 在存储库的蓝图页面上，选择工作流`lakeformationcloudtrailtest`，然后在操作菜单启动.

  工作流程运行时，您可以在上次运行状态column. 偶尔选择 “刷新” 按钮。

  状态来自正在运行，至发现，至Importing，至已完成.

  工作流程完成后：

- 数据目录将有新的元数据表。
- 您的 CloudTrail 日志将被摄入到数据湖中。

如果工作流程失败，请执行以下操作：

a. 选择工作流程，然后在操作菜单查看图表.

工作流程将在Amazon Glue控制台。

b. 确保已选择工作流，然后选择 History (历史记录) 选项卡。
c. UN历史记录中，选择最近一次运行并选择查看运行详细信息.
d. 在动态（运行时）图中选择失败的作业或 Crawler，然后查看错误消息。出现故障的节点为红色或黄色。

步骤 在桌子上授予 SELECT

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

Note

工作流程会自动授予SELECT对它创建的表的权限授予运行它的用户。由于数据湖管理员运行了此工作流，因此您必须授予SELECT给数据分析师。

1. 在Lake Formation 控制台的导航窗格中，在数据atalog，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Grant在权限标题下。
3. 在授予数据权限对话框中，进行以下选择：

   a. UN委托人，对于IM 用户和角色，选择datalake_user.
   b. UNLF 标签或目录资源，选择命名数据目录资源.
   c. 适用于数据库，lakeformation_cloudtrail数据库应该已经处于选中状态。
   d. 适用于表，选择cloudtrailtest-cloudtrail.
   e. UN表和列权限，选择Select.
4. 选择 Grant（授权）。

下一步将以数据分析师的身份执行。

步骤 查询数据湖使用Amazon Athena

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

1. 在以下位置打开 Athena 控制台https://console.aws.amazon.com/athena/并以数据分析师、用户身份登录datalake_user.
2. 如有必要，选择开始使用继续使用 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。

   这些区域有：表列表填充。
5. 在桌子旁边的溢出菜单上（水平排列的 3 个点）cloudtrailtest-cloudtrail，选择Preview，然后选择运行.

   查询运行并显示 10 行数据。

如果您之前没有使用过 Athena，则必须先在 Athena 控制台中配置 Amazon S3 位置以存储查询结果。这些区域有：`datalake_user`必须具有必要的权限才能访问您选择的 Amazon S3 存储桶。

Note

现在，您已完成本教程，请向组织中的委托人授予数据权限和数据位置权限。

在 Lake Formation 中从 JDBC 源创建数据湖

本教程将指导您完成要执行的步骤Amazon Lake Formation控制台，使用 Lake Formation 从 JDBC 源创建和加载你的第一个数据湖。

主题

目标受众

下表列出了此操作中使用的角色Amazon Lake FormationJDBC 教程 (p. 37).

角色	描述
IAM 管理员	可以创建的用户Amazon Identity and Access Management(IM) 用户和角色以及亚马逊 Simple Storage Service (Amazon S3) 存储桶。Has`AdministratorAccess` Amazon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 权限少于 IAM 管理员，但足以管理数据湖。
数据分析人员	可以针对数据湖运行查询的用户。只有足够的权限运行查询。
工作流角色	具有运行工作流程所需的 IAM 策略的角色。

有关完成本教程的前提条件的信息，请参阅JDBC 教程先决条件 (p. 38).

JDBC 教程先决条件

开始前的准备工作Amazon Lake FormationJDBC 教程 (p. 37)，请确保您已完成以下操作：

- 完成设置 Amazon Lake Formation (p. 9)中所述的任务。
- 确定要用于本教程的 JDBC 可访问数据存储。
- 收集创建Amazon Glue类型为 JDBC 的连接。此数据目录对象包括数据存储的 URL、登录凭据，如果数据存储是在 Amazon 虚拟私有云 (Amazon VPC) 中创建的，还包括其他特定于 VPC 的配置信息。有关更多信息，请参阅 。在 中定义连接Amazon GlueData Catalog中的Amazon Glue开发人员指南.

该教程假定您已熟悉Amazon Identity and Access Management(我是)。有关 IAM 的信息，请参阅IAM 用户指南.

开始使用，前进到the section called “第 1 步：创建 IAM 用户作为数据分析师” (p. 38).

第 1 步：创建 IAM 用户作为数据分析师

在此步骤中，您将创建Amazon Identity and Access Management(IAM) 用户作为您的数据湖的数据分析师Amazon Lake Formation.

此用户具有查询数据湖的最低权限集。

1. 使用 https://console.amazonaws.cn/iam 打开 IAM 控制台。以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10)或者作为 IAM 用户使用`AdministratorAccess` Amazon托管策略。
2. 创建一个名为的用户`datalake_user`使用以下设置：
   - 启用Amazon Web Services Management Console访问。
   - 设置密码，不需要重置密码。
   - 将附加到`AmazonAthenaFullAccess` Amazon托管策略。
   - 附加以下内联策略。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：在中创建连接Amazon Glue

Note

如果您已有Amazon Glue连接您的 JDBC 数据源。

Amazon Lake Formation通过访问 JDBC 数据源Amazon Glue 连接. 连接是数据目录对象，其中包含了连接到数据源所需的所有信息。您可以使用Amazon Glue控制台。

创建连接

1. 打开Amazon Glue控制台位于https://console.amazonaws.cn/glue/，然后以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10).
2. 在导航窗格的 Data catalog (数据目录) 下，选择 Connections (连接)。
3. 在存储库的连接页面上，选择添加连接.
4. 在存储库的设置连接的属性页面，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. 继续执行连接向导并保存连接。

   要获取有关创建连接的帮助，请参阅在上处理连接Amazon Glue控制台中的Amazon Glue开发人员指南.

第 3 步：为数据湖创建 Amazon S3 存储桶

在此步骤中，您将创建 Amazon Simple Storage Service (Amazon S3) 存储桶，该存储桶将用作数据湖的根位置。

1. 以下网址打开 Amazon S3 控制台：https://console.aws.amazon.com/s3/并以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10).
2. 选择创建存储桶，然后通过向导创建名为的存储桶*<yourName>*`-datalake-tutorial`，其中*<yourName>*是你的名字缩写和姓氏。例如：`jdoe-datalake-tutorial`。

   有关创建 Amazon S3 存储桶的详细说明，请参阅如何创建 S3 存储桶？中的Amazon Simple Service.

第 4 步：注册 Amazon S3 路径

在此步骤中，您将注册一个 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. 在导航窗格中的下注册和摄取，选择数据湖位置.
3. 选择注册位置(容量预留)，然后选择浏览.
4. Select.*<yourName>*`-datalake-tutorial`您之前创建的存储桶接受默认 IAM 角色`AWSServiceRoleForLakeFormationDataAccess`(容量预留)，然后选择注册位置.

   有关注册位置的更多信息，请参阅将 Amazon S3 位置添加到您的数据湖 (p. 102).

第 5 步：授予数据位置权限

校长必须数据位置权限在数据湖位置上创建指向该位置的数据目录表或数据库。您必须向 IAM 角色授予工作流的数据位置权限，以便工作流程可以写入数据摄取目标。

1. 在 Lake Formation ity 控制台上，在导航窗格中，Permissions (权限)，选择数据位置.

2. 选择Grant，并在授予权限对话框中，执行以下操作：
   a. 适用于IAM 用户和角色，选择LakeFormationWorkflowRole.
   b. 适用于存储位置，选择您的*<yourName>*-datalake-tutorial存储桶。
3. 选择 Grant（授权）。

有关数据位置权限的更多信息，请参阅Underlying Data Access Control (p. 239).

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 ity 控制台上，在导航窗格中，数据目录，选择数据库.
2. 选择创建数据库，UNDER数据库详细信息，输入名称lakeformation_tutorial.
3. 将其他字段留空，然后选择创建数据库.

步骤 7：授予数据权限

您必须授予在 Data Catalog 中创建元数据表的权限。因为工作流程与角色一起运行LakeFormationWorkflowRole，您必须将这些权限授予该角色。

1. 在 Lake Formation ity 控制台上，在导航窗格中，Permissions (权限)，选择数据权限.
2. 选择Grant，并在授予数据权限对话框中，执行以下操作：
   a. UNDER委托人，对于IAM 用户和角色，选择LakeFormationWorkflowRole.
   b. UNDERLF 标签或目录资源，选择命名数据目录资源.
   c. 适用于数据库，选择您以前创建的数据库，lakeformation_tutorial.
   d. UNDER数据库权限，选择创建表、更改, 和Drop，而且很清除超级如果它被选中。
3. 选择 Grant（授权）。

有关授予Lake Formation 权限的更多信息，请参阅安全和访问控制Lake Formation 中的元数据和数据 (p. 234).

步骤 8：使用蓝图创建工作流程

这些区域有：Amazon Lake Formation工作流生成的Amazon Glue作业、爬网程序和触发器，用于发现数据并将其摄取到数据湖中。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 Lake Form蓝图(容量预留)，然后选择使用蓝图.
2. 在存储库的使用蓝图页面，在蓝图类型，选择数据库快照.
3. UNDER导入来源，对于数据库连接，选择您刚刚创建的连接，datalake-tutorial，或者为您的数据源选择一个现有连接。
4. 适用于源数据路径，输入要从中提取数据的路径，格式为*<database>*/*<schema>*/*<table>*.

   您可以用百分比 (%) 通配符替换 schema 或 table。对于支持模式的数据库，请输入*<database>*/*<schema>*/% 以匹配中的所有表*<schema>*之内*<database>*. Oracle Database 和 MySQL 不支持路径中的架构；相反，请输入*<database>*/%。对于Oracle数据库，*<database>*是系统标识符 (SID)。

   例如，如果 Oracle 数据库具有orcl作为其 SID，请输入orcl/%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. UNDER导入目标，请指定以下参数：

目标数据库	`lakeformation_tutorial`
目标存储位置	`s3://<yourName>-datalake-tutorial`
数据格式	（选择镶木地板或 CSV）

6. 对于导入频率，选择按需运行.
7. UNDER导入选项，请指定以下参数：

工作流名称	`lakeformationjdbctest`
IAM 角色	`LakeFormationWorkflowRole`
表prefi	`jdbctest` Note 必须是小写字母。

8. 选择Create，然后等待控制台报告工作流已成功创建。

Tip

您收到以下错误消息了吗？
`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...`
如果是这样，请检查您是否已更换`<account-id>`在数据湖管理员用户的内联策略中Amazon账号。

步骤 9：运行工作流

因为你指定工作流是 run-on-demand，则必须在中手动启动工作流Amazon Lake Formation.

1. 在 Lake Formation 控制台上，在蓝图页面上，选择工作流`lakeformationjdbctest`.
2. 选择操作(容量预留)，然后选择启动.
3. 工作流程运行时，请在上次运行状态column. 偶尔选择 “刷新” 按钮。

状态来自正在运行，至发现，至Importing，至已完成.

工作流程完成后：

- 数据目录有新的元数据表。
- 您的数据将被引入到数据湖中。

如果工作流失败，请执行以下操作：

a. 选择工作流。选择操作(容量预留)，然后选择视图.

工作流程将在Amazon Glue控制台。

b. 选择工作流，然后选择历史记录选项卡。
c. 选择最近的运行并选择查看运行详细信息.
d. 在动态（运行时）图中选择失败的作业或 Crawler，然后查看错误消息。出现故障的节点为红色或黄色。

步骤 10：在桌子上授予 SELECT

你必须授予`SELECT`对中的新数据目录表的权限Amazon Lake Formation以便数据分析师可以查询表所指向的数据。

Note

工作流程会自动授予`SELECT`对它创建的表的权限授予运行它的用户。由于数据湖管理员运行了此工作流，因此您必须授予`SELECT`给数据分析师。

1. 在 Lake Formation ity 控制台上，在导航窗格中，Permissions (权限)，选择数据权限.
2. 选择Grant，并在授予数据权限对话框中，执行以下操作：
   a. UNDER委托人，对于IAM 用户和角色，选择`datalake_user`.
   b. UNDERLF 标签或目录资源，选择命名数据目录资源.
   c. 适用于数据库，选择`lakeformation_tutorial`.

      这些区域有：表列表填充。
   d. 适用于表中，从数据源中选择一个或多个表。
   e. UNDER表和列权限，选择Select.
3. 选择 Grant（授权）。

下一步将作为数据分析师执行。

步骤 11：使用查询数据湖Amazon Athena

使用Amazon Athena控制台查询数据湖中的数据。

1. 通过以下网址打开 Athena 控制台：https://console.aws.amazon.com/athena/，然后以数据分析师、用户身份登录`datalake_user`.
2. 如有必要，选择开始使用以继续安全查询 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_tutorial`。

   这些区域有：表列表填充。
5. 在其中一个表格旁边的弹出菜单中，选择预览表.

   查询运行并显示 10 行数据。

步骤 12：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Amazon Identity and Access Management(IAM) 角色，用于启动 Amazon Redshift 集群和查询 Amazon S3 数据。然后，为该角色授予`Select`您想查询的表的权限。然后，授予用户使用 Amazon Redshift 查询编辑器的权限。最后，创建 Amazon Redshift 集群并运行查询。

您可以以管理员身份创建集群，然后以数据分析师身份查询集群。

有关 Amazon Redshift Spectrum 的更多信息，请参阅使用 Amazon Redshift Spectrum 查询外部数据中的Amazon Redshift 数据库开发人员指南.

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/。以您在中创建的 IAM 管理员用户身份登录创建 IAM 管理员用户 (p. 10) (user nameAdministrator) 或者作为 IAM 用户使用AdministratorAccess Amazon托管策略。
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. 在存储库的查看策略页面上，输入名称如同**RedshiftLakeFormationPolicy**用于您创建的策略。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. 适用于选择可信任的实体，选择Amazon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)。接下来，选择 Next (下一步)：Permissions (下一步：权限)。
12. 搜索您创建的权限策略，RedshiftLakeFormationPolicy，然后选中列表中策略名称旁的复选框。
13. 选择 Next:。标签。
14. 选择 Next:。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。

GRANT**Select**对表的权限以在 Lake Formation 数据库中进行查询

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. 在导航窗格中的下Permissions (权限)，选择数据湖权限(容量预留)，然后选择Grant.
3. 提供以下信息：
   - 适用于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`。

     表格列表随即填充。
   - 适用于表，选择数据源中要查询的表。
   - 选择Select表权限。
4. 选择 Grant（授权）。

设置Amazon Redshift Spectrum 并运行查询

1. 打开位于 Amazon Redshift 控制台https://console.amazonaws.cn/redshift. 作为用户登录`Administrator`.
2. 选择创建集群。
3. 在存储库的创建集群页面，输入`redshift-lakeformation-demo`(对于)集群标识符.
4. 对于节点类型，选择dc2.large.
5. 向下和向下滚动数据库配置，请输入或接受以下参数：
   - 管理员用户名称：`awsuser`
   - 管理员用户密码：(*`Choose a password`*)
6. Expand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. 如果必须使用不同于默认值 5439 的端口，请在其他配置，关闭使用原定设置选项。展开部分数据库配置，然后输入一个新的数据库端口数字。
8. 选择创建集群。

   这些区域有：集群页面加载。
9. 等到集群状态变成Available. 定期选择 “刷新” 图标。
10. 授予数据分析师对集群运行查询的权限。为此，请完成以下步骤。
    a. 通过以下网址打开 IAM 控制台：https://console.aws.amazon.com/iam/，然后以`Administrator`用户。
    b. 在导航窗格中，选择用户，并将下面的托管策略附加到用户`datalake_user`.
       - `AmazonRedshiftQueryEditor`
       - `AmazonRedshiftReadOnlyAccess`
11. 退出Amazon Redshift 控制台并以用户身份重新登录`datalake_user`.
12. 在左侧垂直工具栏中，选择编辑器图标，打开查询编辑器并连接到集群。如果连接到数据库对话框中，选择群集名称`redshift-lakeformation-demo`，然后输入数据库名称**`dev`**，用户名称**`awsuser`**，以及您创建的密码。然后，选择 Connect to database (连接到数据库)。

    Note

    如果系统未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，请选择更改连接以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映射数据库`lakeformation_tutorial`在 Lake Formation 中改为Amazon Redshift 模式名称`redshift_jdbc`：

    Important

    Replace*<account-id>*具有有效的Amazon账号，以及*<region>*具有有效的Amazon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. 在架构列表下选择架构，选择redshift_jdbc.

    表格列表随即填充。查询编辑器仅显示您被授予了 Lake Formation 数据湖权限的表。

15. 在表名旁边的弹出式菜单中，选择预览数据.

    Amazon Redshift 返回前 10 行。

    现在，您可以对您具有相应权限的表和列运行查询。

步骤 13：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 权限的功能。这些语句与现有的Amazon Redshift 语句类似。有关更多信息，请参阅 。授予和撤销中的Amazon Redshift 数据库开发人员指南.

在 Lake Formation 中创建受管控表

Amazon Lake Formation支持使用名为的新数据湖表类型的原子、一致、隔离和持久 (ACID) 事务受监管的表. Lake Formation 事务简化了 ETL 脚本和工作流程的开发，允许多个用户同时可靠地插入、删除和修改受管控表。Lake Formation 在后台自动压缩和优化受管控表的存储，以提高查询性能。

在本教程中，您将了解如何使用 Amazon Service（Amazon S3）上的现有数据创建新的受管控表。它还说明了如何查询受管控表以及如何使用 Amazon Athena 运行时空旅行查询。

Note

由于本教程使用公有 Amazon S3 存储桶，因此它仅包含只读使用案例。在现实世界中，您可能希望通过将对象放入您的 Amazon S3 存储桶、将它们添加到受管控表并启用压缩来做更多的事情。

本教程包括Amazon CloudFormation用于快速设置的模板。您可以查看和自定义它来满足要求。如果你更喜欢在上设置资源Amazon控制台，请参阅中的说明第 1 步：配置资源 (p. 46).

主题

目标受众

本教程面向 IAM 管理员、数据湖管理员和数据分析师。下表列出了本教程中使用 Lake Formation 创建受控表时使用的角色。

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。Has`AdministratorAccess` Amazon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 权限少于 IAM 管理员，但足以管理数据湖。
数据分析人员	可以针对数据湖运行查询的用户。只有足够的权限运行查询。

先决条件

开始本教程之前，您必须具有Amazon Web Services 账户您可以使用 IAM 用户身份登录的具有正确权限。有关更多信息，请参阅 注册Amazon (p. 10) 和 创建 IAM 管理员用户 (p. 10)。

此教程假设您已熟悉 IAM 角色和策略。有关 IAM 的信息，请参阅IAM 用户指南.

第 1 步：配置资源

您需要设置以下设置Amazon完成本教程的资源：

- IAM 用户、角色和策略
- Lake Formation 数据湖设置和权限

此部分将向您介绍如何设置Amazon资源有两种不同的方式：

1. 使用 Amazon CloudFormation 模板
2. 使用 Amazon 控制台

使用创建资源Amazon CloudFormation模板

完成以下步骤以使用Amazon CloudFormation模板：模板：模板

1. 登录到Amazon CloudFormation控制台位于控制台https://console.aws.amazon.com/cloudformation作为美国东部（弗吉尼亚北部）区域的 IAM 用户。
2. 选择启动堆栈.
3. 选择下一步在创建堆栈屏幕。
4. 输入堆栈名称。
5. 适用于DatalakeAdminUserName和DatalakeAdminUserPassword中，输入数据湖管理员用户的 IAM 用户名和密码。
6. 适用于DatalakeAnalystUserName和DatalakeAnalystUserPassword中，输入数据湖分析师用户的 IAM 用户名和密码。
7. 适用于DataLakeBucketName中，输入要创建的新存储桶名称。

8. 适用于DatabaseName，将保留为默认值。
9. 选择 Next（下一步）。
10. 在下一页上，选择下一步.
11. 查看最后一页的详细信息并选择我承认这一点Amazon CloudFormation可能会创建 IAM 资源。
12. 选择 Create（创建）。

    堆栈创建可能需要长达两分钟。

使用创建资源Amazon控制台

完成以下步骤以使用Amazon控制台：

1. 首先，您需要设置两个 IAM 角色；Amazon GlueETL 作业，另一个用于Lake Formation 数据湖位置。要创建 IAM 策略，请完成以下步骤：

   a. IAM 控制台 (IAM 控制台https://console.aws.amazon.com/iam/)，为 Amazon S3 创建新的策略。将以下策略另存为`S3DataLakePolicy`：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::your-datalake-bucket-name/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
        "Resource": [
                "arn:aws:s3:::your-datalake-bucket-name"
            ]
        }
    ]
}
```

   b. 创建名为的新 IAM 策略`LFLocationPolicy`使用以下语句：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LFtransactions",
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartTransaction",
                "lakeformation:CommitTransaction",
                "lakeformation:CancelTransaction",
                "lakeformation:GetTableObjects",
                "lakeformation:UpdateTableObjects"
            ],
            "Resource": "*"
        },
```

```
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetTableVersions",
                "glue:GetPartitions",
                "glue:GetTable",
                "glue:GetTables",
                "glue:UpdateTable"
            ],
            "Resource": "*"
        }
    ]
}
```

c. 创建名为的新 IAM 策略`LFQuery`使用以下语句：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LFtransactions",
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartTransaction",
                "lakeformation:CommitTransaction",
                "lakeformation:CancelTransaction",
                "lakeformation:ExtendTransaction",
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetTableObjects",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
         }
       ]
}
```

2. 接下来，完成以下步骤，为 Lake Formation 数据位置创建 IAM 角色：

   - 创建一个名为的新的Lake Formation 角色`LFRegisterLocationServiceRole`与 Lake Formation 建立信任关系：

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

     附加客户托管策略 (`S3DataLakePolicy`和`LFLocationPolicy`) 您在上一步中创建的。此角色用于向 Lake Formation 注册地点，后者反过来会在查询时为 Athena 执行凭证发售。

3. 接下来，按照以下步骤创建您的 IAM 用户：

a. 创建 IAM 用户被命名DatalakeAdmin并附上以下内容Amazon托管策略：托管策略

   1. AWSLakeFormationDataAdmin
   2. AmazonAthenaFullAccess
   3. IAMReadOnlyAccess

   适用于DataLakeBucketName中，输入要创建的新存储桶名称。

b. 附加客户管理的策略LFQueryPolicy.
c. 创建名为的 IAM 用户DataAnalyst可以使用 Athena 查询数据。
d. 将附加到Amazon管理的策略AmazonAthenaFullAccess.
e. 附上客户托管政策LFQueryPolicy.

4. 请按照以下步骤配置Lake Formation：

   a. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。UNDERPermissions (权限)，选择Admins和数据库创建者。
   b. 在数据湖管理员部分中，选择Grant.
   c. 适用于IAM 用户和角色，选择 IAM 用户DatalakeAdmin.
   d. 选择 Save（保存）。
   e. 在数据库创建者部分中，选择Grant.
   f. 适用于IAM 用户和角色，选择LFRegisterLocationServiceRole.
   g. Select创建数据库.
   h. 选择 Grant（授权）。
   i. UNDER注册并摄取，选择数据湖位置.
   j. 选择注册位置.
   k. Select数据库.
   l. 适用于Amazon S3 路径中，输入存储数据的 Amazon S3 存储桶位置。该存储桶必须与您在中列出的存储桶相同LFLocationPolicy. Lake Formation 使用此角色提供临时 Amazon S3 凭证，以查询需要对存储桶及其下所有前缀进行读/写访问的服务。
   m. 适用于IAM 角色，选择LFRegisterLocationServiceRole.
   n. 选择注册位置.
   o. UNDERData Catalog，选择设置.
   p. 确保两个复选框均为仅对新数据库使用 IAM 访问控制和仅对新数据库中的新表使用 IAM 访问控制已取消选择。
   q. UNDER数据目录，选择数据库.
   r. 选择 Create database（创建数据库）。
   s. 对于 Name (名称)，请输入 lakeformation_tutorial_amazon_reviews。
   t. 选择 Create database（创建数据库）。

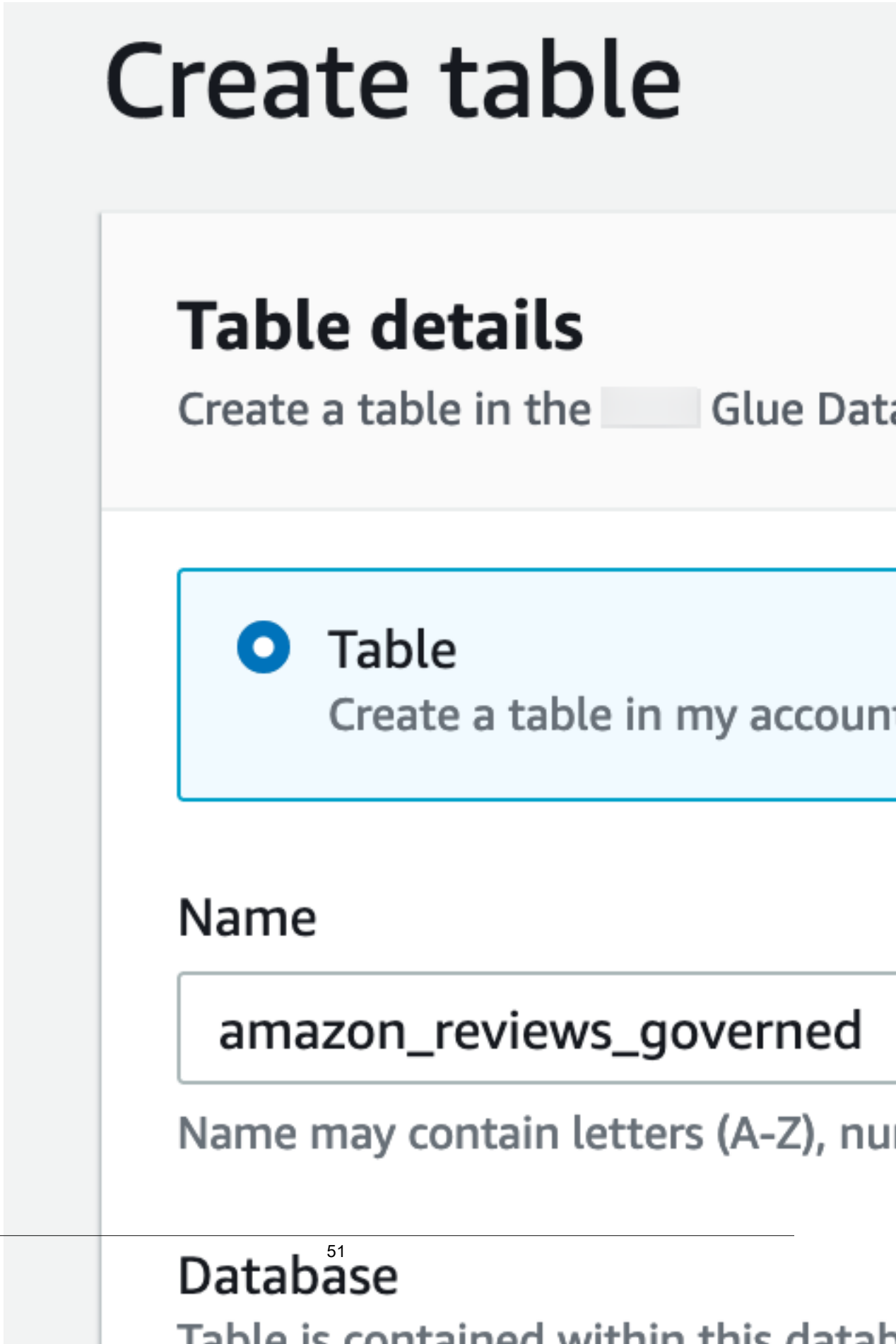
第 2 步：设置受管控表

现在，您可以在 Lake Formation 中创建和配置第一个受管控表。

创建受管控表

1. 登录到Lake Formation控制台位于控制台https://console.amazonaws.cn/lakeformation/作为DatalakeAdmin1用户。
2. 选择表。
3. 选择 Create Table（创建表）。

4. 对于名称，请输入`amazon_reviews_governed`.
5. 对于数据库，请输入`lakeformation_tutorial_amazon_reviews`.
6. Select启用受管控的数据访问和管理.


Create table
Table details
Create a table in the
Glue Data
Table
Create a table in my account
Name
amazon_reviews_governed
Name may contain letters (A-Z), nu
Database

7. 适用于数据位于，选择在我的账户中指定路径.
8. 输入路径s3://*your-datalake-bucket-name*/parquet/哪里*your-datalake-bucket-name*是您在中输入的存储桶名称Amazon CloudFormation模板。模板。模板
9. 适用于Classification，选择实木复合地板.
10. 选择上载架构.
11. 在文本框中输入以下 Json 数组。

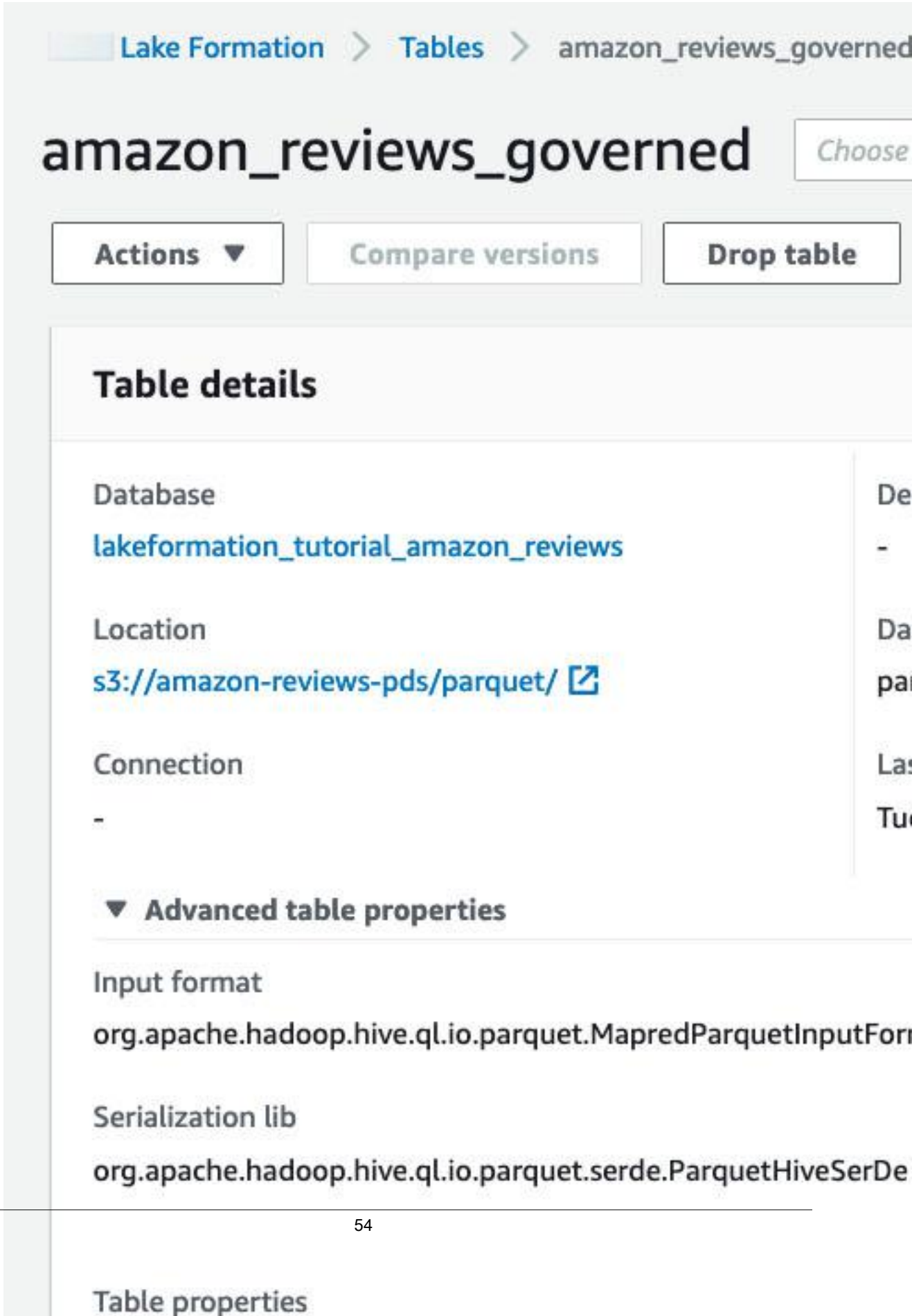
```
[
    {
        "Name": "marketplace",
        "Type": "string"
    },
    {
        "Name": "customer_id",
        "Type": "string"
    },
    {
        "Name": "review_id",
        "Type": "string"
    },
    {
        "Name": "product_id",
        "Type": "string"
    },
    {
        "Name": "product_parent",
        "Type": "string"
    },
    {
        "Name": "product_title",
        "Type": "string"
    },
    {
        "Name": "star_rating",
        "Type": "int"
    },
    {
        "Name": "helpful_votes",
        "Type": "int"
    },
    {
        "Name": "total_votes",
        "Type": "int"
    },
    {
        "Name": "vine",
        "Type": "string"
    },
    {
        "Name": "verified_purchase",
        "Type": "string"
    },
    {
        "Name": "review_headline",
        "Type": "string"
    },
    {
        "Name": "review_body",
        "Type": "string"
    },
    {
        "Name": "review_date",
        "Type": "bigint"
```

```
    },
    {
        "Name": "year",
        "Type": "int"
    }
]
```

12. 请选择 Upload（上传）。
13. 选择 Add column（添加列）。
14. 对于列名称，输入`product_category`.
15. 对于数据类型，选择字符串.
16. Select分区键.
17. 选择 Add（添加）。
18. 选择 Submit（提交）。

现在，您可以看到新的受管控表已创建。

选择表名后，可以看到受管控表的详细信息，还可以看到监管：监管 Enabled (已启用)在这个视图中。这意味着此表是Lake Formation 控制表。不受管控的表应显示为监管：监管 Disabled.


Lake Formation
Tables
amazon_reviews_governed
amazon_reviews_governed
Choose
Actions
Compare versions
Drop table
Table details
Database
lakeformation_tutorial_amazon_reviews
Location
s3://amazon-reviews-pds/parquet/
Connection
-
Advanced table properties
Input format
org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFor
Serialization lib
org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe
Table properties

默认情况下，对受管控表启用自动压缩。要禁用自动压缩Amazon Amazon CLI和 SDK，运行以下命令：

```
$ aws lakeformation update-table-storage-optimizer --database-name
 lakeformation_tutorial_amazon_reviews --table-name amazon_reviews_governed --storage-
optimizer-config '{"compaction": {"is_enabled":"false"}}'
```

第 3 步：配置Lake Formation

目前，您在中创建的受管控表the section called “第 2 步：设置受管控表” (p. 49)不包含任何数据或分区。在下一个步骤中，您将现有的 Amazon Service Service（Amazon S3）对象添加到清单。即使您的数据位于受管控表的表位置，在将数据添加到受管控表之前，数据也不会被识别。在将对象添加到受管控表之前，请配置 Lake Formation 权限以向用户授予所需的权限。

配置Lake Formation

要向用户授予对受管控表的必需的 Lake Formation 权限，请完成以下步骤：

1. 在以下位置登录Lake Formation 控制台https://console.aws.amazon.com/lakeformation/在美国东部（弗吉尼亚北部）区域为`DatalakeAdmin1`用户。
2. UNDERPermissions (权限)，选择数据湖权限.
3. UNDER数据权限，选择Grant.
4. 适用于委托人，选择IAM 用户和角色，然后选择角色`LFRegisterLocationServiceRole-CloudFormation ####`和用户`DatalakeAdmin1`.
5. 适用于策略标签或目录资源，选择命名数据目录资源.
6. 对于 Database (数据库)，请选择 `lakeformation_tutorial_amazon_reviews`。
7. 适用于表，选择`amazon_reviews_governed`.
8. 适用于表权限SSSS选择Select、描述、Insert、更改、Delete, 和Drop.
9. 适用于数据权限SSSS选择所有数据访问权限.
10. 选择 Grant（授权）。
11. UNDERPermissions (权限)，选择数据湖权限.
12. UNDER数据权限，选择Grant.
13. 适用于委托人，选择IAM 用户和角色，然后选择用户`DataAnalyst1`.
14. 适用于策略标签或目录资源，选择命名数据目录资源.
15. 对于 Database (数据库)，请选择 `lakeformation_tutorial_amazon_reviews`。
16. 适用于表，选择`amazon_reviews_governed`.
17. 适用于表权限SSSS选择Select和描述.
18. 适用于数据权限SSSS选择所有数据访问权限.
19. 选择 Grant（授权）。

第 4 步：将表对象添加到受管控表中

要将 Amazon S3 对象添加到受管控表，您需要调用`UpdateTableObjects`API。你可以用Amazon Command Line Interface(Amazon CLI) 和 SDK，还有Amazon GlueETL 库（API 在库中隐式调用）。在本教程中，我们使用Amazon CLI以解释 API 级别的行为。安装Amazon CLI，请参阅安装或更新最新版本的Amazon Command Line Interface.

为了简化示例，我们只添加一个带有一个文件的分区。在实际使用中，您可能想要添加所需的所有分区中的所有文件。

1. 使用运行以下命令`DatalakeAdmin1`用户的凭据。
2. 首先，使用启动新事务`StartTransaction`API。

```
$ aws lakeformation start-transaction
{
    "TransactionId": "396880372a0045dc9c8faff2d19dfeea"
}
```

3. 现在，您可以在事务中向此表中添加文件。要添加文件，请选择一个名为的示例分区product_category=Camera来自 的amazon-reviews-pds表，然后在此分区下选择一个文件。您需要了解的内容Uri、ETag, 和Size你添加的文件。要查找此信息，请输入以下Amazon CLI命令（替换*your-datalake-bucket-name*使用有效的存储桶名称）：

```
$ aws s3 ls s3://your-datalake-bucket-name/parquet/product_category=Camera/
2021-11-23 18:19:19   65227429 part-00004-495c48e6-96d6-4650-
aa65-3c36a3516ddd.c000.snappy.parquet

$ aws s3api head-object --bucketyour-datalake-bucket-name--key
 parquet/product_category=Camera/part-00004-495c48e6-96d6-4650-
aa65-3c36a3516ddd.c000.snappy.parquet
{
    "AcceptRanges": "bytes",
    "LastModified": "Mon, 09 Apr 2018 06:37:07 GMT",
    "ContentLength": 65227429,
    "ETag": "\"980669fcf6ccf31d2d686b9cccdd45e3-8\"",
    "ContentType": "binary/octet-stream",
    "Metadata": {}
}
```

4. 创建名为的新文件write-operations1.json并输入以下 JSON：(将Uri、ETag, 和Size使用你复制的值。)

```
[
    {
        "AddObject": {
            "Uri": "s3://your-datalake-bucket-name/parquet/product_category=Camera/
part-00000-495c48e6-96d6-4650-aa65-3c36a3516ddd.c000.snappy.parquet",
            "ETag": "d4c25c40f33071620fb31cf0346ed2ec-8",
            "Size": 65386769,
            "PartitionValues": [
                "Camera"
            ]
        }
    }
]
```

5. 要向受管控表中添加文件，请使用UpdateTableObjects使用调用 API 调用write-operations1.json您刚刚创建的文件。Replace####使用你在 start-transaction 命令中获得的交易 ID。

```
$ aws lakeformation update-table-objects --database-name
 lakeformation_tutorial_amazon_reviews --table-name amazon_reviews_governed --
transaction-id transaction-id --write-operations file://./write-operations1.json
```

6. 接下来，在提交事务之前通过调用GetTableObjects具有相同交易 ID 的 API :(替换####用你拿到的身份证start-transaction命令）。您将需要运行此命令两次，因为第一次上面的更改尚未应用于事务。第二次，UpdateTableObjects操作将成功。

```
$ aws lakeformation get-table-objects --database-name
 lakeformation_tutorial_amazon_reviews --table-name amazon_reviews_governed --
transaction-id transaction-id
```

```
{
    "Objects": [
        {
            "PartitionValues": [
                "Camera"
            ],
            "Objects": [
                {
                    "Uri": "s3://your-datalake-bucket-name/
parquet/product_category=Camera/part-00000-495c48e6-96d6-4650-
aa65-3c36a3516ddd.c000.snappy.parquet",
                    "ETag": "d4c25c40f33071620fb31cf0346ed2ec-8",
                    "Size": 65386769
                }
            ]
        }
    ]
}
```

7. 提交事务，以便其他用户可以在此事务之外使用此数据。为此，调用CommitTransactionAPI: (替换*###*使用你在启动交易命令中获得的交易 ID）。

```
$ aws lakeformation commit-transaction --transaction-id transaction-id
```

请注意提交更改后的当前日期时间UpdateTableObjectsAPI 调用。在本示例的后面部分中，我们将此时间戳用于时间旅行查询。

```
$ date -u
Wed May 26 08:12:00 UTC 2021
```

提交事务后，你可以在 Lake Formation 控制台上看到分区https://console.aws.amazon.com/lakeformation/.

使用以下命令，添加所有分区中的所有文件：

1. 调用StartTransaction用于启动另一个Lake Formation 事务的 API。

```
$ aws lakeformation start-transaction
{
"TransactionId": "c96cdbeab7e34b35a383faa75b372234"
}
```

2. 列出位于上的 Amazon S3 对象amazon-reviews-pdsbucket 来选择另一个示例文件。

```
s3://your-datalake-bucket-name/parquet/product_category=Books/
2018-04-09 15:35:58 1094842361 part-00000-495c48e6-96d6-4650-
aa65-3c36a3516ddd.c000.snappy.parquet
```

3. 调用HeadObject针对一个示例文件的 API 以进行复制ETag和Size.

```
$ aws s3api head-object --bucket your-datalake-bucket-name --key
 parquet/product_category=Books/part-00000-495c48e6-96d6-4650-
aa65-3c36a3516ddd.c000.snappy.parquet
{
   "AcceptRanges": "bytes",
   "LastModified": "Mon, 09 Apr 2018 06:35:58 GMT",
   "ContentLength": 1094842361,
   "ETag": "\"9805c2c9a0459ccf337e01dc727f8efc-131\"",
   "ContentType": "binary/octet-stream",
```

```
   "Metadata": {}
 }
```

4. 创建名为的新文件`write-operations2.json`并输入以下 JSON：(将*URI*、*ETag*, 和*Size*使用你复制的值。)

```
[
    {
            "AddObject": {
            "Uri": "s3://your-datalake-bucket-name/parquet/product_category=Books/
part-00000-495c48e6-96d6-4650-aa65-3c36a3516ddd.c000.snappy.parquet",
            "ETag": "9805c2c9a0459ccf337e01dc727f8efc-131",
            "Size": 1094842361,
            "PartitionValues": [
                "Books"
            ]
        }
    }
]
```

5. 调用`UpdateTableObjects`使用写操作的 API 2.json: (替换*####*使用你在启动交易命令中获得的交易 ID)。

```
$ aws lakeformation update-table-objects --database-name
 lakeformation_tutorial_amazon_reviews --table-name amazon_reviews_governed --
transaction-id transaction-id --write-operations file://./write-operations2.json
```

6. 调用`CommitTransaction`API: (替换*####*使用你在 start-transaction 命令中获得的事务 ID)。

```
$ aws lakeformation commit-transaction --transaction-id transaction-id
```

现在 Lake Formation 控制台上可见这两个分区https://console.aws.amazon.com/lakeformation/.

第 5 步：使用 Amazon Athena 查询受管控表

现在开始查询您使用 Amazon Athena 创建的受管控表。

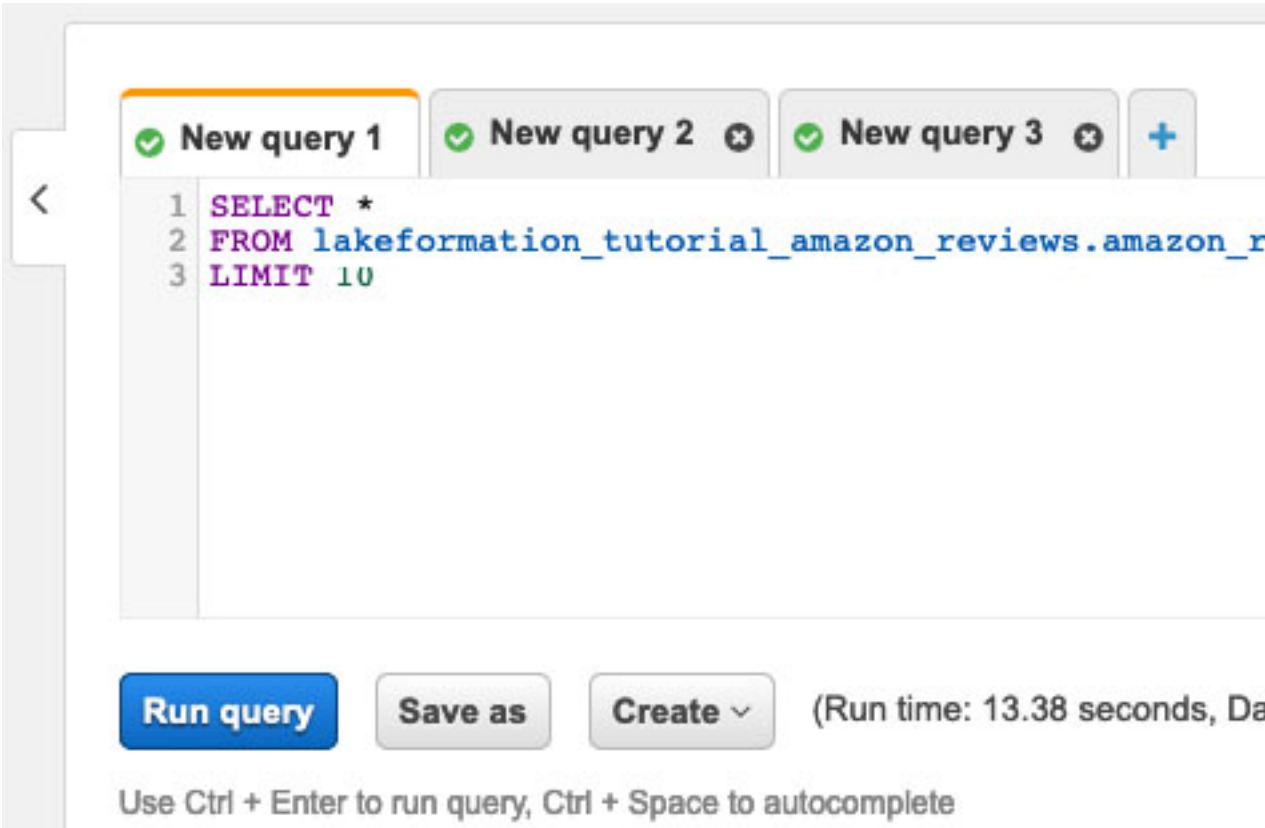
如果您是第一次在 Athena 中运行查询，则需要配置查询结果位置。有关更多信息，请参阅 。指定查询结果位置.

运行简单查询

- 要运行简单查询，请登录 Athena 控制台https://console.aws.amazon.com/athena/在美国东部（弗吉尼亚北部）区域为`DataAnalyst1`用户。运行以下查询以预览存储在受管控表中的 10 条记录：

```
SELECT *
FROM lakeformation_tutorial_amazon_reviews.amazon_reviews_governed
LIMIT 10
```

结果如下所示：



Results

	marketplace	customer_id	review_id	product_id
1	US	12115179	R249G3JHC13PYV	0765356155
2	US	30255545	R17L9M1OUKCWY0	0764209833
3	US	10609904	R3FAAKHI4W47AK	1634154797
4	US	32941543	R2VWIDVTOLEYCP	1616385944
5	US	13909031	R2S5U9H76YLWKT	0736062785
6	US	9911218	RM0CW1UGZTCZK	0312538367
7	US	17033727	RXEZMN6LBLSBU	1581060629
8	US	45799330	RN3F4XOK39ZS0	0240812581
9	US	51455130	R26RXQ0QNNI9CN	1563921197
10	US	47323097	RUZTLC3RWK0PB	0135026961

运行分析查询

- 运行以下脚本以运行带聚合的分析查询，以模拟实际用例：

```
SELECT product_category, count(*) as TotalReviews, avg(star_rating) as AverageRating
FROM lakeformation_tutorial_amazon_reviews.amazon_reviews_governed
GROUP BY product_category
```

此查询返回每个产品类别的评论总数和平均评分。

使用时空旅行运行分析查询

受管控表启用时空旅行-您可以查询上一次的表。

Note

为了在 Athena 中运行时空旅行查询，你需要使用 Athena 引擎版本 2。如果您的工作组仍在使用 Athena 引擎版本 1，请更新您的工作组以使用 Athena 引擎版本 2。

- 要提交时空旅行查询，请使用`FOR SYSTEM_TIME AS OF`在中位于表名称之后的时间戳`SELECT`语句，如以下示例语法所示：

```
SELECT *
FROM database.table
FOR SYSTEM_TIME AS OF timestamp
```

timestamp参数可以是时间戳，也可以是带时区的时间戳。如果未指定，Athena 将该值视为以 UTC 时间表示的时间戳。运行时空旅行查询以检索截至世界标准时间 2021-05-26 08:15:00 的数据：

```
SELECT product_category, count(*) as TotalReviews, avg(star_rating) as AverageRating
FROM lakeformation_tutorial_amazon_reviews.amazon_reviews_governed
FOR SYSTEM_TIME AS OF TIMESTAMP '2021-05-26 08:15:00 UTC'
GROUP BY product_category
```

这些区域有：结果屏幕包含以下内容的记录`product_category=Camera`. 这是因为该文件在`product_category=Books`是在时间戳之后添加的 (*2021-05-26 ###### 08:15:00*)，已在中指定`FOR SYSTEM_TIME AS OF`.

第 6 步：清理 Amazon资源

清理 资源

为了防止向您收取不必要的费用Amazon Web Services 账户，您可以删除Amazon您在本教程中使用的资源。

- 删除云层堆栈. 您创建的受管控表将随堆栈一起自动删除。

使用基于 Lake Formation 标签的访问控制来管理数据湖

成千上万的客户正在构建 PB 级的数据湖Amazon. 这些客户中有许多使用Amazon Lake Formation以便在整个组织内轻松构建和共享他们的数据湖。随着表和用户数量的增加，数据管理员和管理员正在寻求轻松地大

规模管理数据湖权限的方法。基于 Lake Formation 标签的访问控制 (LF-TBAC) 通过允许数据管理员创建LF 标签（基于其数据分类和本体），然后可以将其附加到资源。

LF-TBAC 是一种授权策略，基于属性来定义权限。在 Lake Formation 中，这些属性称为 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 标签分配和撤消对Lake Formation 资源的权限。有关的更多信息，请参阅。Lake Formation 标签访问控制 (p. 179).

本教程演示如何使用创建基于 Lake Formation 标签的访问控制策略Amazon公有数据集。此外，它还演示了如何查询具有与之关联的基于 Lake Formation 标记的访问策略的表、数据库和列。

您可以将 LF-TBAC 用于以下使用案例：

- 您有大量表和委托人需要数据湖管理员授予访问权限
- 您希望根据本体对数据进行分类，并根据分类授予权限
- 数据湖管理员希望以松散耦合的方式动态分配权限

以下是使用 LF-TBAC 配置权限的概要步骤：

1. 数据管家用两个 LF 标签来定义标签本体：`Confidential`和`Sensitive`. DATA`Confidential=True`具有更严格的访问控制。DATA`Sensitive=True`需要分析师进行具体分析。
2. 数据管理员为数据工程师分配不同的权限级别，以构建具有不同 LF 标签的表。
3. 数据工程师建立了两个数据库：`tag_database`和`col_tag_database`. 中的所有表`tag_database`配置为`Confidential=True`. 中的所有表`col_tag_database`配置为`Confidential=False`. 表的某些列位于`col_tag_database`被标记为`Sensitive=True`以满足特定的分析需求。
4. 数据工程师向分析师授予具有特定表达式条件的表的读取权限`Confidential=True`和`Confidential=False`、`Sensitive=True`.
5. 通过这种配置，数据分析师可以专注于使用正确的数据进行分析。

主题

目标受众

本教程面向数据管理员、数据工程师和数据分析师。说到管理Amazon Glue Data Catalog和 Lake Formation 中的管理权限，生产账户中的数据管理员根据其支持的功能拥有职能所有权，并且可以向各种消费者、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）	这些区域有：`lf-data-steward`用户具有以下访问权限：

角色	描述
	• 针对 Data Catalog 中所有资源的读权限 • 可以创建 LF 标签并关联到数据工程师角色，以便向其他委托人授予权限
数据工程师	`lf-data-engineer`用户具有以下访问权限： • 对 Data Catalog 中所有资源的完全读取、写入和更新访问权限 • 数据湖中的数据位置权限 • 可以关联 LF 标签并关联到数据目录 • 可以将 LF 标签附加到资源，这可以根据数据管理员创建的任何策略提供对委托人的访问权限
数据分析人员	这些区域有：`lf-data-analyst`用户具有以下访问权限： • 精细访问由基于标签的 Lake Formation 访问策略共享的资源

先决条件

在开始本教程之前，您必须具有Amazon Web Services 账户您可以使用 IAM 用户身份登录的具有正确权限。有关更多信息，请参阅 注册Amazon (p. 10) 和 创建 IAM 管理员用户 (p. 10)。

本教程假设您已熟悉 IAM。有关 IAM 的信息，请参阅IAM 用户指南.

第 1 步：配置资源

本教程包含Amazon CloudFormation用于快速设置的模板。您可以查看和自定义它来满足要求。该模板创建了三个不同的角色（在目标受众 (p. 61)) 来执行此练习并复制 nyc-taxi-data 数据集到您本地 Amazon S3 存储桶。

- 一个 Amazon S3 存储桶
- 相应的Lake Formation 设置
- 相应的 Amazon EC2 资源
- 三个具有用户 ID 凭证的 IAM 用户角色

创建您的资源

1. 登录到Amazon CloudFormation控制台位于https://console.aws.amazon.com/cloudformation位于美国东部（弗吉尼亚北部）区域。
2. 选择启动堆栈.
3. 选择 Next（下一步）。
4. 在用户配置部分中，输入三个角色的密码：`DataStewardUserPassword`、`DataEngineerUserPassword`和`DataAnalystUserPassword`.
5. 查看最后一页的详细信息并选择我承认这一点Amazon CloudFormation可能会创建 IAM 资源.
6. 选择Create（创建）。

   创建堆栈可能需要长达五分钟。

Note

完成本教程后，您可能需要删除堆栈。Amazon CloudFormation以避免继续产生费用。在堆栈的事件状态下验证资源是否已成功删除。

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 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. 以数据管家用户身份执行第一步（`lf-data-steward`) 以验证 Amazon S3 中的数据和Lake Formation 中的数据目录。
   a. 通过以下方式登录到 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/如同`lf-data-steward`使用部署时使用的密码Amazon CloudFormation堆。
   b. 在导航窗格中的下Permissions (权限)选择管理角色和任务.
   c. 适用于IAM 用户和角色，选择用户`lf-data-steward`.
   d. 选择Save（保存）添加`lf-data-steward`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，而不是基于 IAM 的访问控制。
   a. 在导航窗格中的下Data Catalog，选择设置.
   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. 在导航窗格中的下注册和摄取，选择数据湖位置.
   b. 适用于Amazon S3 路径，输入`s3://lf-tagbased-demo-`*`Account-ID`*.
   c. 适用于IAMrole保留默认值`AWSServiceRoleForLakeFormationDataAccess`照原样。
   d. 选择注册位置.
4. 接下来，通过定义一个 LF-tag 来创建本体。
   a. UNDERPermissions (权限)在导航窗格中的下管理角色，选择LF 标签。.
   b. 选择添加 LF 标签.
   c. 对于 Key（键），输入 `Confidential`。
   d. 适用于值，添加`True`和`False`.
   e. 选择添加 LF-tag.
   f. 重复这些步骤以创建lf-tag `Sensitive`有值`True`.

   你已经创建了所有必需的LF 标签本练习。

向 IAM 用户授予权限

1. 接下来，让特定 IAM 委托人能够将新创建的 LF 标签附加到资源。
   a. UNDERPermissions (权限)在导航窗格中的下管理角色，选择LF-Tag 权限.
   b. 选择 Grant（授权）。

c. SelectIAM 用户和角色.
d. 适用于IAM 用户和角色搜索并选择`lf-data-engineer`角色。
e. 在LF-tag 权限范围部分，添加密钥`Confidential`有价值观`True`和`False`，以及key `Sensitive`有价值的`True`.
f. UNDERPermissions (权限)选择描述和AAs为了LF-Tag 权限和可授予权限。
g. 选择 Grant（授权）。

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

a. UNDERPermissions (权限)在导航窗格中，选择管理角色.
b. 在数据库创建者部分，选择Grant.
c. 适用于IAM 用户和角色，选择`lf-data-engineer`角色。
d. 适用于目录权限，选择创建数据库.
e. 选择 Grant（授权）。

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

a. 在导航窗格中，选择数据位置.
b. 选择 Grant（授权）。
c. Select我的账户.
d. 适用于IAM 用户和角色，选择`lf-data-engineer`角色。
e. 适用于存储位置，输入 Amazon S3 存储桶。Amazon CloudFormation模板(`s3://lf-tagbased-demo-`*`Account-ID`*).
f. 选择 Grant（授权）。

4. 下一步：Grant`lf-data-engineer`与关联的资源的可授予权限lf-tag表情`Confidential=True`.

a. 在导航窗格中，选择数据权限.
b. 选择 Grant（授权）。
c. SelectIAM 用户和角色.
d. 选择角色`lf-data-engineer`.
e. 在LF-tag 或目录资源部分，选择LF 标签匹配的资源.
f. 选择添加 LF-tag.
g. 添加关键帧`Confidential`有有效值`True`.
h. 在数据库权限部分，选择描述为了数据库权限和可授予权限.
i. 在表和列权限部分，选择描述、Select, 和更改代表这两者的表权限和可授予权限.
j. 选择 Grant（授权）。

5. 下一步：Grant`lf-data-engineer`对与 LF-tag 表达式关联的资源的可授予权限`Confidential=False`.

a. 在导航窗格中，选择数据权限.
b. 选择 Grant（授权）。
c. SelectIAM 用户和角色.
d. 选择角色`lf-data-engineer`.
e. SelectLF 标签匹配的资源.
f. 选择添加 LF-tag.
g. 添加关键帧`Confidential`有值`False`.
h. 在数据库权限部分，选择描述为了数据库权限和可授予权限.
i. 在表和列权限部分中，不要选择任何内容。
j. 选择 Grant（授权）。

6. 接下来，我们授予`lf-data-engineer`与关联的资源的可授予权限lf-tag表情`Confidential=False`和`Sensitive=True`.

   a. 在导航窗格中，选择数据权限.
   b. 选择 Grant（授权）。
   c. SelectIAM 用户和角色.
   d. 选择角色`lf-data-engineer`.
   e. SelectLF 标签匹配的资源.
   f. 选择添加 LF-tag.
   g. 添加关键帧`Confidential`有值`False`.
   h. 选择添加 LF-tag.
   i. 添加关键帧`Sensitive`有值`True`.
   j. 在数据库权限部分，选择描述为了数据库权限和可授予权限.
   k. 在表和列权限部分，选择描述、Select, 和更改代表这两者的表权限和可授予权限.
   l. 选择 Grant（授权）。

第 3 步：创建Lake Formation

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`，桌子`source_data`，并附上相应的 LF 标签。

   a. 在 Lake Formation 控制台上 (https://console.aws.amazon.com/lakeformation/)，选择数据库.
   b. 选择 Create database（创建数据库）。
   c. 对于 Name (名称)，请输入 `tag_database`。
   d. 适用于位置，输入由创建的 Amazon S3 位置Amazon CloudFormation模板(`s3://lf-tagbased-demo-`*`Account-ID`*`/tag_database/`).
   e. 取消选择仅对数据库中的新表使用 IAM 访问控制.
   f. 选择 Create database（创建数据库）。
2. 接下来，在中创建一个新表`tag_database`.

   a. 在存储库的数据库页面上，选择数据库`tag_database`.
   b. 选择查看表然后点击创建表.
   c. 对于 Name (名称)，请输入 `source_data`。
   d. 对于 Database (数据库)，选择 `tag_database` 数据库。
   e. 适用于数据位于，选择在我的账户中指定路径.
   f. 对于 Include path，输入`tag_database`由创建Amazon CloudFormation模板(`s3://lf-tagbased-demo`*`Account-ID`*`/tag_database/`).
   g. 适用于数据格式，选择CSV.
   h. UNDER上传架构中，输入以下 JSON 列结构数组来创建架构：

```
[
              {
                   "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"

               }
]
```

i. 请选择 Upload（上传）。上传架构后，表架构应类似于以下屏幕截图：

#	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

j. 选择 Submit（提交）。

3. 接下来，在数据库级别附加 LF 标签。

   a. 在存储库的数据库页面上，查找并选择tag_database.
   b. 在存储库的操作菜单中，选择编辑 LF 标签.
   c. 选择分配新的 LF-tag.
   d. 适用于已分配密钥选择Confidential LF-tag您之前创建的。
   e. 适用于值，选择True.
   f. 选择Save（保存）。

   这样就完成了 tag_database 数据库的 LF-tag 分配。

创建您的数据库和表以进行列级访问

重复以下步骤以创建数据库col_tag_database和桌子source_data_col_lvl，并在列级别附加 LF 标签。

1. 在存储库的数据库页面上，选择创建数据库.
2. 对于 Name (名称)，请输入 col_tag_database。
3. 适用于位置，输入由创建的 Amazon S3 位置Amazon CloudFormation模板(s3://lf-tagbased-demo-*Account-ID*/col_tag_database/).
4. 取消选择仅对数据库中的新表使用 IAM 访问控制.
5. 选择 Create database（创建数据库）。
6. 在存储库的数据库页面上，选择新数据库(col_tag_database).
7. 选择查看表然后舔创建表.
8. 对于 Name (名称)，请输入 source_data_col_lvl。
9. 适用于数据库，选择新数据库(col_tag_database).
10. 适用于数据位于，选择在我的账户中指定路径.
11. 输入AthAmazon S3 路径col_tag_database (s3://lf-tagbased-demo-*Account-ID*/col_tag_database/).
12. 适用于数据格式，选择CSV.
13. UNDERUpload schema中，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"


                }
]
```

14. 选择 Upload。上传架构后，表架构应类似于以下屏幕截图。

#	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

15. 选择SITT以完成表的创建。
16. 现在，关联Sensitive=TrueLF-tag 到列中vendorid和fare_amount.
    a. 在存储库的表页面上，选择您创建的表(source_data_col_lvl).
    b. 在存储库的操作菜单中，选择架构.
    c. 选择列vendorid然后选择编辑 LF 标签.
    d. 适用于已分配密钥，选择SITI.
    e. 适用于值，选择True.
    f. 选择Save（保存）。
17. 接下来，关联Confidential=Falself-tag tocol_tag_database. 该项为必填项lf-data-analyst能够描述数据库col_tag_database从 Athena 登录时。
    a. 在存储库的数据库页面上，查找并选择col_tag_database.
    b. 在存储库的操作菜单中，选择编辑 LF 标签.
    c. 选择分配新的 LF-tag.
    d. 适用于已分配密钥选择Confidential您之前创建的 LF 标签。
    e. 适用于值，选择False.
    f. 选择Save（保存）。

第 4 步：授予表权限

授予数据分析师使用数据库的权限tag_database还有桌子col_tag_database使用 LF 标签Confidential和Sensitive.

1. 按照以下步骤将权限授予lf-data-analyst用户正在访问与 LF-tag 关联的对象Confidential=True（数据库：tag_database）Describe数据库和Select对表的权限。
    a. 通过以下方式登录到 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/如同lf-data-engineer.
    b. 在存储库的Permissions (权限)页面上，选择数据权限.
    c. 选择 Grant（授权）。
    d. UNDER委托人，选择IAM 用户和角色.
    e. 适用于IAM 用户和角色，选择lf-data-analyst.
    f. SelectLF 标签匹配的资源.
    g. 选择添加 LF-tag.
    h. 适用于密钥，选择Confidential.
    i. 适用于值选择True.
    j. 适用于数据库权限，选择Describe.
    k. 适用于表权限，选择Select和描述.
    l. 选择 Grant（授权）。
2. 接下来，重复上述步骤，向数据分析师授予 LF-tag 表达式的权限Confidential=False. 该lf-tag用于描述col_tag_database还有桌子source_data_col_lvl以登录身份登录时lf-data-analyst来自Amazon Athena
    a. 通过以下方式登录到 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/如同lf-data-engineer.
    b. 在存储库的数据库页面上，选择数据库col_tag_database.
    c. 选择操作和Grant.
    d. UNDER委托人，选择IAM 用户和角色.

e. 适用于IAM 用户和角色，选择`lf-data-analyst`.
f. SelectLF 标签匹配的资源.
g. 选择添加 LF-tag.
h. 适用于密钥，选择`Confidential`.
i. 适用于值选择`False`.
j. 适用于数据库权限，选择`Describe`.
k. 适用于表权限，请勿选择任何内容。
l. 选择 Grant（授权）。

3. 接下来，重复上述步骤，向数据分析师授予 LF-tag 表达式的权限`Confidential=False`和`Sensitive=True`. 这个 LF 标签用于描述`col_tag_database`还有桌子`source_data_col_lvl`（列级）登录时以`lf-data-analyst`来自Amazon Athena

a. 通过以下方式登录到 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/如同`lf-data-engineer`.
b. 在 “数据库” 页上，选择数据库`col_tag_database`.
c. 选择操作和Grant.
d. UNDER委托人，选择IAM 用户和角色.
e. 适用于IAM 用户和角色，选择`lf-data-analyst`.
f. SelectLF 标签匹配的资源.
g. 选择添加 LF-tag.
h. 适用于密钥，选择`Confidential`.
i. 适用于值选择`False`.
j. 选择添加 LF-tag.
k. 适用于密钥，选择`Sensitive`.
l. 适用于值选择`True`.
m. 适用于数据库权限，选择`Describe`.
n. 适用于表权限，选择`Select`和`Describe`.
o. 选择 Grant（授权）。

第 5 步：在 Amazon Athena 中运行查询以验证权限

对于此步骤，请使用 Amazon Athena 运行`SELECT`对两个表的查询(`source_data and source_data_col_lvl`). 使用 Amazon S3 路径作为查询结果位置(`s3://lf-tagbased-demo-`*`Account-ID`*`/athena-results/`).

1. 通过以下网址登录 Athena 控制台https://console.aws.amazon.com/athena/如同`lf-data-analyst`.
2. 在 Athena 查询编辑器中，选择`tag_database`在左侧面板中。
3. 选择旁边的附加菜单选项图标（三个垂直点）`source_data`然后选择Previet.
4. 选择 Run query (运行查询)。

   此查询应需要几分钟才能运行。查询会显示输出中的所有列，因为 LF-tag 在数据库级别关联，而`source_data`表自动继承了`LF-tag`来自数据库`tag_database`.
5. 使用运行另一个查询`col_tag_database`和`source_data_col_lvl`.

   第二个查询返回被标记为的两列`Non-Confidential`和`Sensitive`.
6. 您还可以查看在没有策略授权的列上查看基于 Lake Formation 标记的访问策略行为。从表中选择未加标签的列时`source_data_col_lvl`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：清理 Amazon资源

为了防止向您收取不必要的费用Amazon Web Services 账户，您可以删除Amazon您在本教程中使用的资源。

1. 以身份登录到 Lake Formation 控制台`lf-data-engineer`并删除数据库`tag_database`和`col_tag_database`.
2. 接下来，以身份登录`lf-data-steward`然后把所有的LF-tag 权限、数据权限和数据位置权限在上面授予的那些被授予的`lf-data-engineer`和`lf-data-analyst`..
3. 使用 IAM 凭证登录到 Amazon S3 控制台，以账户拥有者身份登录 Amazon S3 控制台。Amazon CloudFormation堆。
4. 删除以下存储桶：
   - lf-tagbased-demo-accesslogs-*acct-id*
   - lf-tagbased-demo-*acct-id*
5. 登录Amazon CloudFormation控制台位于https://console.aws.amazon.com/cloudformation，然后删除您创建的堆栈。等待堆栈状态更改为`DELETE_COMPLETE`.

使用行级访问控制保护数据湖

Amazon Lake Formation行级权限允许您根据数据合规性和治理策略提供对表中特定行的访问权限。如果您有存储数十亿条记录的大型表，则需要一种方法来允许不同的用户和团队仅访问允许他们查看的数据。行级访问控制是保护数据的一种简单而高效的方法，同时允许用户访问执行工作所需的数据。Lake Formation 通过确定哪些委托人访问哪些服务、时间和通过哪些服务访问哪些服务，提供集中式审计

在本教程中，您将学习行级访问控制在 Lake Formation (Lake Formation) 中的工作原理，以及如何设置它们。

本教程包括Amazon CloudFormation用于快速设置所需资源的模板。您可以根据需要对其进行查看和自定义。

主题

目标受众

本教程面向数据管理员、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	可以创建的用户Amazon Identity and Access Management(IAM) 用户和角色以及 Amazon Simple Storage Service (Amazon S3) 存储桶。有`AdministratorAccess` Amazon托管策略。
数据湖管理员	负责设置数据湖、创建数据筛选器和向数据分析师授予权限的用户。
数据分析人员	针对数据湖运行查询的用户。居住在不同国家（对于我们的使用案例，美国和日本）的数据分析师只能分析位于自己国家/地区的买家的商品评论，出于合规性原因，不应该能够看到位于其他国家/地区的客户数据。

先决条件

开始本教程前，您必须具有Amazon Web Services 账户您可以登录的身份Amazon Identity and Access Management具有正确权限的 (IAM) 用户。有关更多信息，请参阅 注册Amazon (p. 10) 和 创建 IAM 管理员用户 (p. 10)。

本教程假设您已熟悉 IAM。有关 IAM 的信息，请参阅IAM 用户指南.

更改 Lake Formation 设置

> **Important**
>
> 在启动之前Amazon CloudFormation模板，禁用该选项仅对新数据库/表使用IAM 访问控制器在 Lake Formation 中，按以下步骤操作：

1. 通过以下网址登录 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/在美国东部(弗吉尼亚北部) 区域或美国西部 (俄勒冈) 区域。
2. 在 “数据目录” 下，选择设置.
3. 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：配置资源

本教程包括Amazon CloudFormation快速设置的模板。您可以根据需要对其进行查看和自定义。这些区域有：Amazon CloudFormation模板生成以下资源：

- IAM 用户和策略适用于：
  - DataLake 管理员
  - dataAnalyst
  - dataAnalystJP
- Lake Formation 数据湖设置和权限
- Lambda 函数（适用于 Lambda 支持Amazon 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 Amazon S3 存储桶
- 作为数据湖的 Amazon S3 存储桶
- 网络 ACL 和安全组都允许 (因此可到达您的实例) 的发起 ping 的Amazon Glue Data Catalog数据库、表和分区

创建资源

按照以下步骤，使用Amazon CloudFormationTemplate。

1. 登录登录到Amazon CloudFormation控制台https://console.aws.amazon.com/cloudformation在美国东部(弗吉尼亚北部) 区域。
2. 选择启动堆栈.
3. 选择下一步在创建堆栈屏幕。
4. 输入堆栈名称。
5. 适用于DataLake 管理员用户名和DataLake 管理员用户密码中，输入数据湖管理员用户的 IAM 用户名和密码。
6. 适用于数据分析师/用户名称和Data Analyst 用户密码中，为负责美国商城的数据分析师用户输入所需的用户名和密码的用户名和密码。
7. 适用于dataAnalyst JP 用户名和dataAnalyst JPUSERPUSER 密码中，为负责日本商城的数据分析师用户输入所需的用户名和密码的用户名和密码。
8. 适用于DataLakeBucketName中，输入数据存储桶的名称。
9. 适用于DatabaseName, 和TableName保留为默认值。
10. 选择 Next（下一步）
11. 在下一页上，选择下一步.
12. 查看最后一页上的详细信息并选择我承认Amazon CloudFormation可能会创建 IAM 资源。
13. 选择 Create（创建）。

    堆栈创建可能需要一分钟时间才能完成。

第 2 步：没有数据过滤器的查

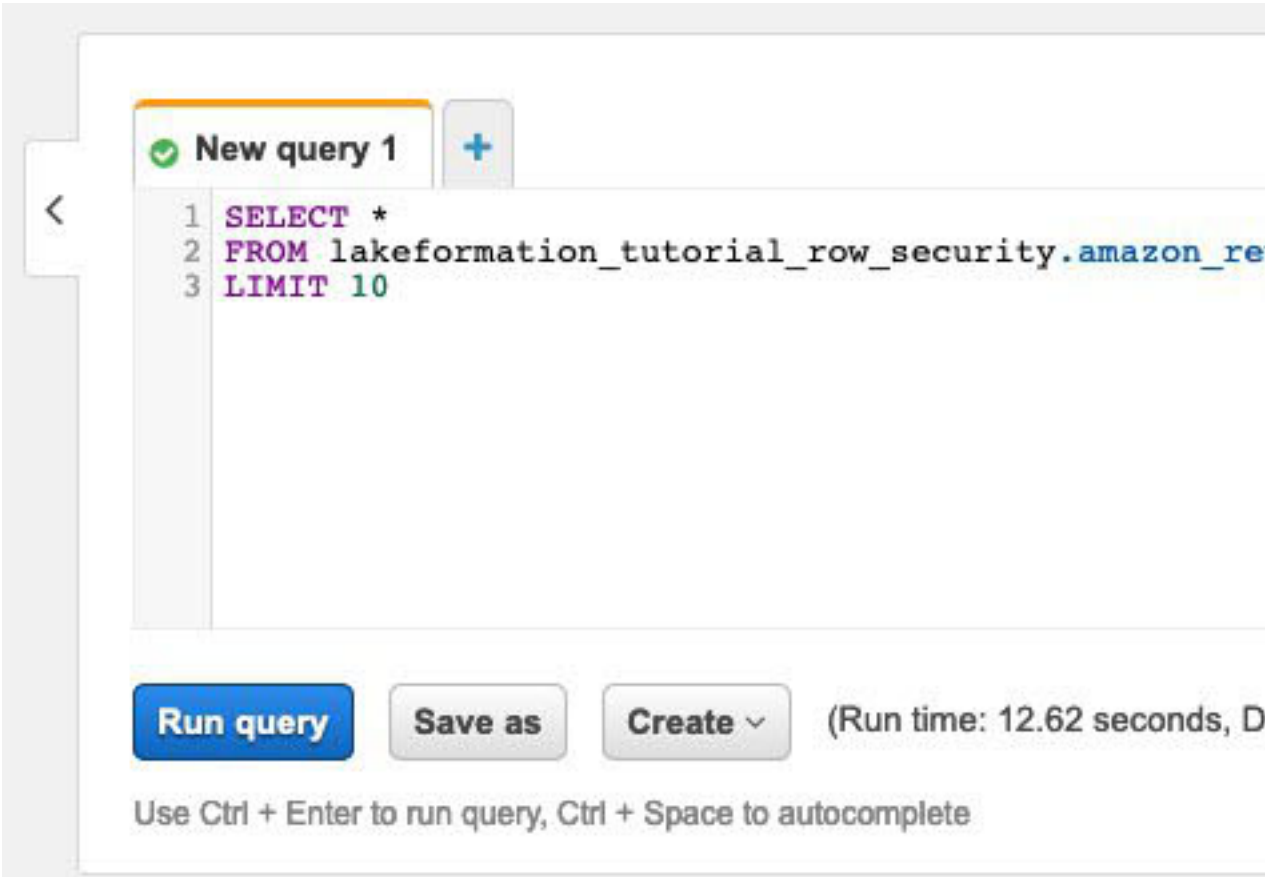
设置环境后，您可以查询产品评论表。首先在没有行级别访问控制的情况下查询表，以确保您可以看到数据。如果您首次在 Amazon Athena 中运行查询，则需要配置查询结果位置。

在没有行级访问控制的情况下查询表

1. 登录登录Athena控制台https://console.aws.amazon.com/athena/作为`DatalakeAdmin`用户，并运行以下查询：

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

   以下屏幕截图显示查询结果。这张表只有一个分区，`product_category=Video`，因此每条记录都是视频产品的评论评论。



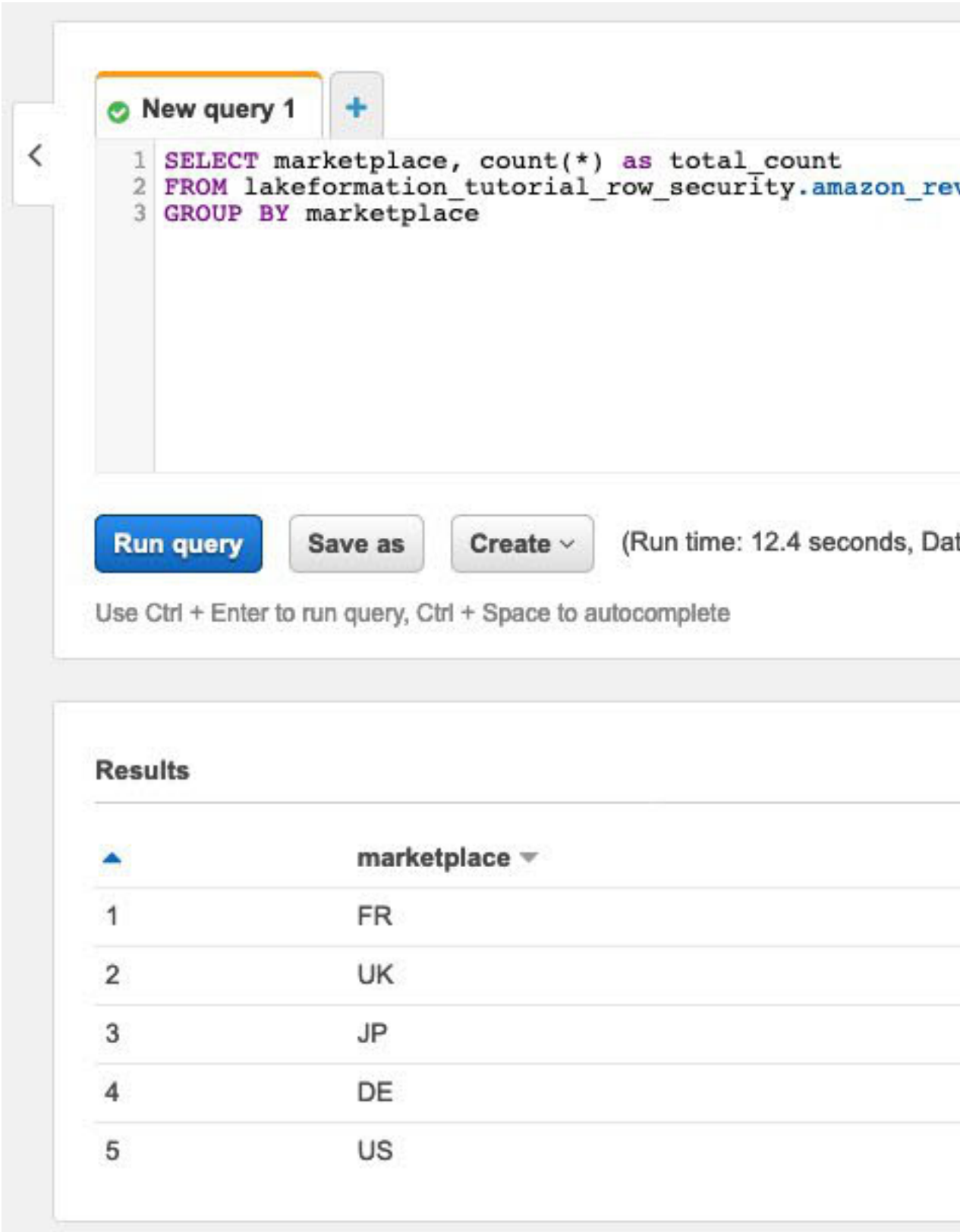
Results

	marketplace	customer_id	review_id	product_i
1	US	22066705	R3HZYXMJ5HEXIG	630487862
2	US	20838467	RJC8PH4K3DVQB	630335663
3	US	15338666	R1OH4581ARVWNX	630026943
4	US	7080939	R3TWQ5OT8KW0E8	B000EKC
5	US	30548191	R3BK9ULGX82VG0	078311317
6	US	16052189	R1LV7NN89A38YT	630286283
7	US	43430756	R2IJAELO3PXEYM	B00027VB
8	US	43539164	R3TN0J9JANR9Q5	630320554
9	US	21187650	R2AVXCQOLI53IC	630260671

2. 接下来，运行聚合查询以检索每个记录的总数marketplace.

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

以下屏幕截图显示查询结果。这些区域有：marketplace列有五个不同的值。在后续步骤中，您将使用marketplacecolumn.


New query 1
1 SELECT marketplace, count(*) as total_count
2 FROM lakeformation_tutorial_row_security.amazon_rev
3 GROUP BY marketplace
Run query
Save as
Create
(Run time: 12.4 seconds, Dat
Use Ctrl + Enter to run query, Ctrl + Space to autocomplete
Results
marketplace
1 FR
2 UK
3 JP
4 DE
5 US

第 3 步：设置数据过滤器并授予权限

本教程使用两位数据分析师：一位负责美国市场，另一位负责日本市场。每位分析师只使用 Athena 来分析其特定市场的客户评论。创建两个不同的数据筛选器，一个用于负责美国商城的分析师，另一个用于负责日本商城的分析师。然后，向分析师授予各自的权限。

创建数据过滤器并授予权限

1. 创建过滤器以限制对US marketplace数据。
   a. 通过以下网址登录 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/在美国东部 (弗吉尼亚北部) 区域中，DatalakeAdmin用户。
   b. 选择筛选数据.
   c. 选择创建新筛选条件.
   d. 适用于数据筛选器名称输入amazon_reviews_US.
   e. 适用于目标数据库，选择数据库lakeformation_tutorial_row_security.
   f. 适用于目标表，选择表amazon_reviews.
   g. 适用于访问列级访问保留为默认值。
   h. 适用于行筛选表达式输入marketplace='US'.
   i. 请选择 Create filter (创建筛选器)。
2. 创建过滤器以限制对日语的访问marketplace数据。
   a. 在存储库的筛选数据页面上，选择创建新筛选条件.
   b. 适用于数据筛选器名称输入amazon_reviews_JP.
   c. 适用于目标数据库，选择数据库lakeformation_tutorial_row_security.
   d. 适用于目标表，选择table amazon_reviews.
   e. 适用于访问列级访问保留为默认值。
   f. 对于行筛选器表达式，请输入marketplace='JP'.
   g. 请选择 Create filter (创建筛选器)。
3. 接下来，使用这些数据过滤器向数据分析师授予权限。按照以下步骤向美国数据分析师授予权限 (DataAnalystUS):
   a. UNDERPermissions (权限)，选择数据湖权限.
   b. UNDER数据权限，选择Grant.
   c. 适用于委托人，选择IAM 用户和角色，然后选择角色DataAnalystUS.
   d. 适用于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. 适用于表-可选，选择amazon_reviews.
   g. 适用于数据过滤器 — 可选DER Selectamazon_reviews_US.
   h. 适用于数据筛选权限，选择Select.
   i. 选择 Grant (授权) 。
4. 按照以下步骤向日本数据分析师授予权限 (DataAnalystJP):
   a. UNDERPermissions (权限)，选择数据湖权限.
   b. UNDER数据权限，选择Grant.
   c. 适用于委托人，选择IAM 用户和角色，然后选择角色DataAnalystJP.
   d. 适用于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. 适用于表-可选，选择amazon_reviews.
   g. 适用于数据过滤器 — 可选DER Selectamazon_reviews_JP.

h. 适用于数据筛选权限，选择Select.

i. 选择 Grant（授权）。

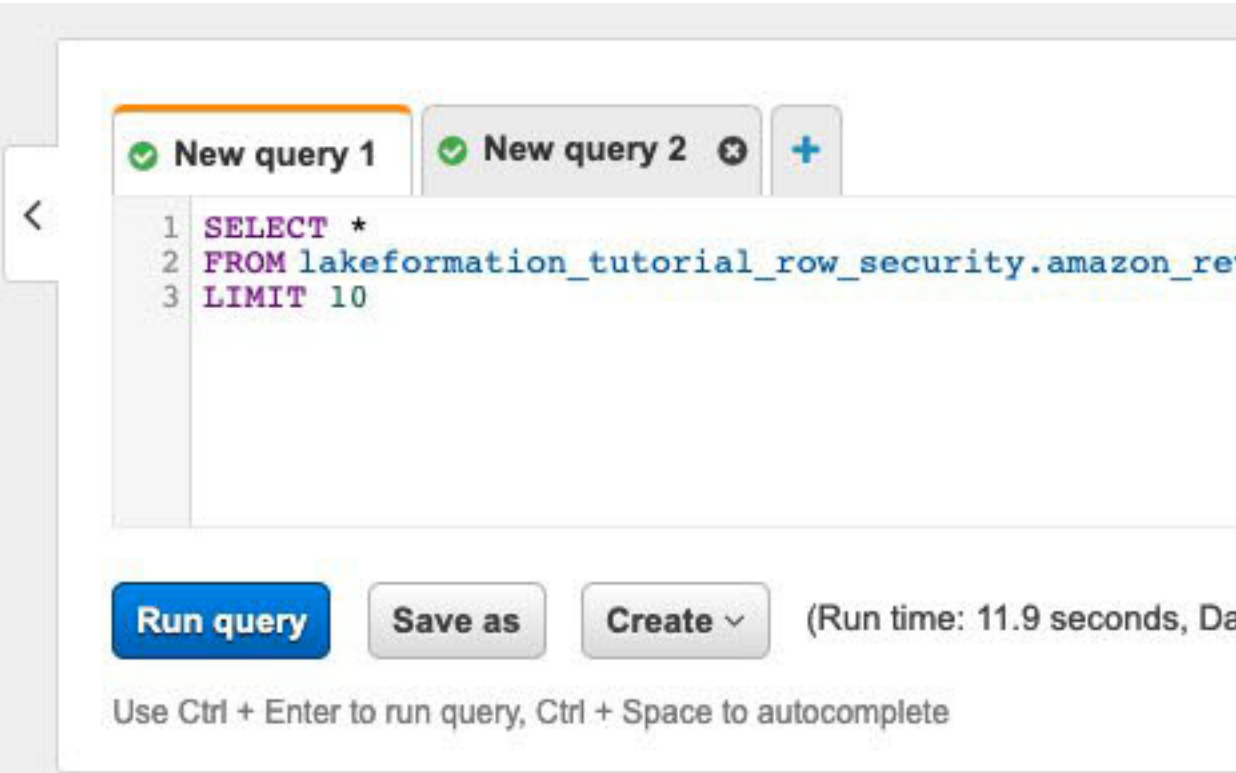
第 4 步：使用数据过滤器查询

使用附加到产品评论表的数据筛选器，运行一些查询并了解 Lake Formation 如何强制执行权限。

1. 登录位于的 Athena 控制台https://console.aws.amazon.com/athena/作为`DataAnalystUS`用户。
2. 运行以下查询以检索一些记录，这些记录将根据我们定义的行级权限进行筛选：

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

以下屏幕截图显示查询结果。



Results

	marketplace	customer_id	review_id	product_i
1	US	43836277	R2NUBTTUO60VYU	B00068S4
2	US	20261976	R2QTOLZUQERU5B	63030600
3	US	15947067	R1PHKR75RKZNSU	63039273
4	US	19288153	R1BL2WVE5X34UN	63040321
5	US	19712967	R2DKOCIBS5FSP7	07840177
6	US	51047097	R2XF5HQATT4IVR	07939601
7	US	43836277	R2NUBTTUO60VYU	B00068S4
8	US	51047097	R1C0H0G6NATZXO	63048725
9	US	42808630	R2HXW7UD4IGZLN	63030600
10	US	11682952	R18IURLUPYI4DP	63029937

3. 同样，运行查询以计算每个商城的记录总数。

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

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

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
```

查询结果只显示记录属于JP marketplace.

5. 运行查询来计算每条记录的总数。marketplace.

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

查询结果只显示属于JP marketplace.

第 5 步：清理 Amazon资源

清理 资源

为了防止对你的不必要的费用Amazon Web Services 账户，您可以删除Amazon你用于本教程的资源。

- 删除云形成堆栈.

使用基于 Lake Formation 标签的访问控制和命名资源共享数据湖

本教程演示如何配置Amazon Lake Formation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，而无需拷贝整个数据库。有两种方法可以与他人共享数据库和表Amazon Web Services 账户通过使用 Lake Formation 跨账户访问控制：

- 基于Lake Formation 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，该策略基于属性来定义权限。在 Lake Formation 中，这些属性被称为LF 标签. 有关详细信息，请参考 使用基于 Lake Formation 标签的访问控制来管理数据湖 (p. 60)。
- Lake Formation (

  Lake Formation 命名资源方法是一种授权策略，用于定义资源权限。资源包括数据库、表和列。数据湖管理员可以分配和撤消对Lake Formation 资源的权限。有关详细信息，请参考 Lake Formation 中的跨账户访问 (p. 242)。

  如果数据湖管理员倾向于明确授予单个资源权限，我们建议使用命名资源。当您使用 named resource 方法向外部账户授予 Lake Formation 对数据目录资源的权限时，Lake Formation 会使用Amazon Resource Access Manager(Amazon RAM) 以共享资源。

主题

目标受众

本教程面向数据管理员、数据工程师和数据分析师。当谈到从共享数据目录表时Amazon Glue和 Lake Formation 中的管理权限，生产账户中的数据管理员根据其支持的功能拥有职能所有权，并且可以向各种消费者、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全读取、写入和更新访问权限 • 能够授予资源权限 • 可以为共享表创建资源链接 • 可以将 LF 标签附加到资源，这可以根据数据管理员创建的任何策略提供对委托人的访问权限
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全读取、写入和更新访问权限 • 能够授予资源权限 • 可以为共享表创建资源链接 • 可以将 LF 标签附加到资源，这可以根据数据管理员创建的任何策略提供对委托人的访问权限
DataAnalyst	这些区域有： DataAnalyst 用户具有以下访问权限： • 精细访问由 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

在生产者账户中配置Lake Formation 数据目录设置

开始本教程之前，您必须具有Amazon Web Services 账户您可以使用登录的Amazon Identity and Access Management具有正确权限的 (IAM) 用户。有关更多信息，请参阅 注册Amazon (p. 10) 和 创建 IAM 管理员用户 (p. 10)。

本教程假定您已熟悉 IAM。有关 IAM 的信息，请参阅IAM 用户指南.

在生产者账户中配置Lake Formation 数据目录设置

> **Note**
>
> 在本教程中，具有源表的账户称为创建者账户，需要访问源表的账户称为使用者账户。

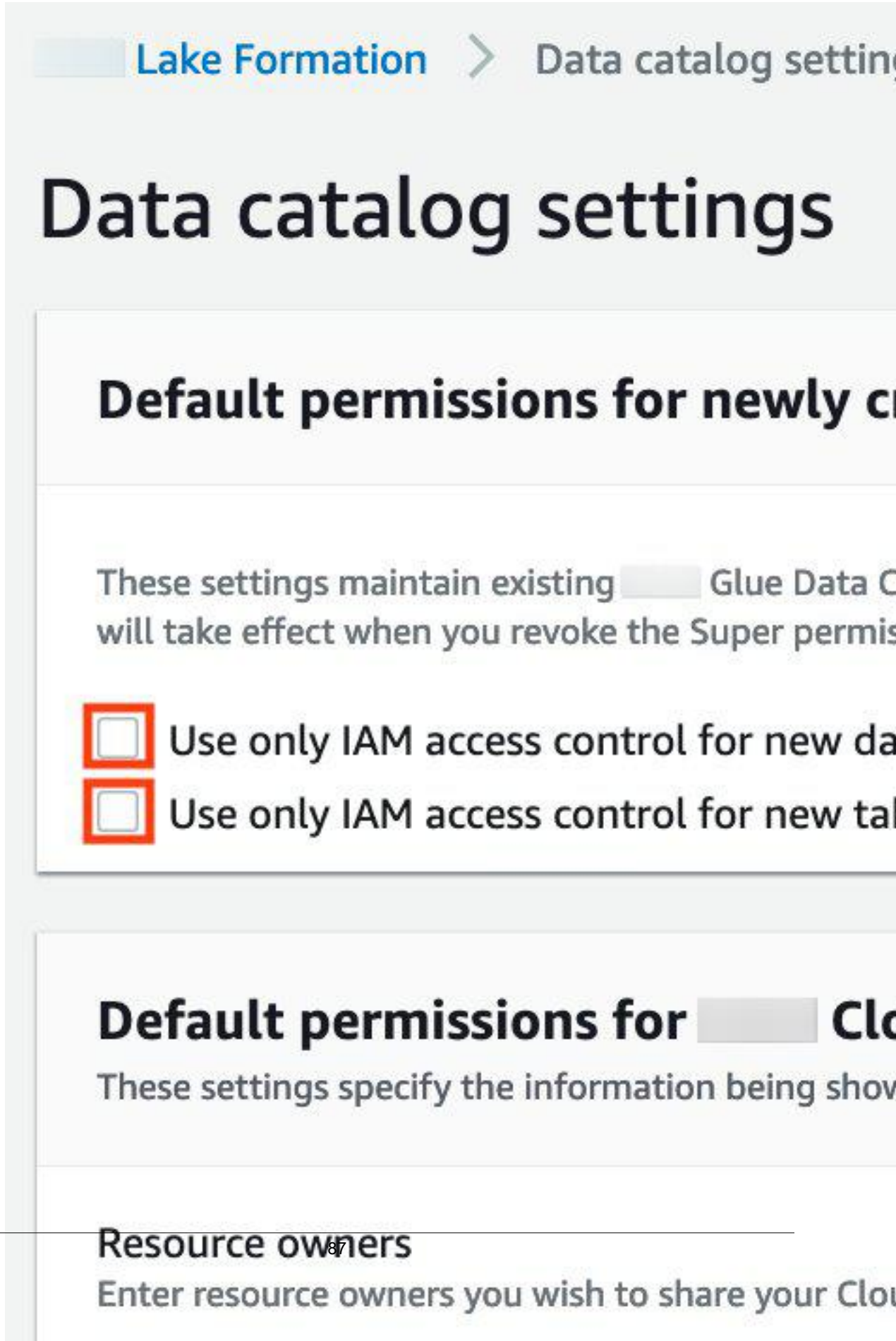
Lake Formation 提供了自己的权限管理模型。为了保持与 IAM 权限模型的向后兼容性，`Super`已向该组授予权限`IAMAllowedPrincipals`所有现有的Amazon Glue Data Catalog默认情况下是资源。另外，仅使用 IAM 访问控制设置已为新的数据目录资源启用。本教程使用 Lake Formation 权限进行细粒度访问控制，并使用 IAM 策略进行粗粒度访问控制。有关详细信息，请参阅 用于精细访问控制的方法 (p. 235)。因此，在使用Amazon CloudFormation模板进行快速设置，你需要在生产者账户中更改 Lake Formation 数据目录设置。

> **Important**
>
> 此设置会影响所有新创建的数据库和表，因此我们强烈建议您在非生产帐户或新帐户中完成本教程。此外，如果您使用的是共享帐户（例如公司的开发账户），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，在该步骤中撤消默认设置超级权限来自`IAMAllowedPrincipals`在数据库或表上。我们将在本教程的后面部分中讨论细节。

要在创建者账户中配置 Lake Foration 数据目录设置，请完成以下步骤：

1. 登录到Amazon Web Services Management Console以管理员用户身份或作为 Lake Formation 用户使用制作者账户`PutDataLakeSettings`API 权限。
2. 在 Lake Formation 控制台上的导航窗格中的Data Catalog，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）。


Lake Formation
Data catalog settin
Data catalog settings
Default permissions for newly c
These settings maintain existing Glue Data C
will take effect when you revoke the Super permis
Use only IAM access control for new da
Use only IAM access control for new tal
Default permissions for Clo
These settings specify the information being shov
Resource owners
Enter resource owners you wish to share your Clou

此外，您可以删除`CREATE_DATABASE`的权限`IAMAllowedPrincipals`下管理角色和任务、数据库创建者. 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：使用预配置您的资源Amazon CloudFormation模板

这些区域有： CloudFormation 创建者账户的模板生成以下资源：

- 用作数据湖的 Amazon S3 存储桶。
- 一个 Lambda 函数（用于 Lambda 支持的Amazon CloudFormation自定义资源)。我们使用函数将示例数据文件从公有 Simple Storage S3 存储桶复制到您的 Simple StoragAmazon S3 存储桶中。
- IAM 用户和策略： DataLakeAdminProducer.
- 相应的Lake Formation 设置和权限包括：
  - 在生产者账户中定义 Lake Formation 数据湖管理员
  - 将 Amazon S3 存储桶注册为Lake Formation 数据湖位置（生产者账户）
- 网络 ACL 和安全组都允许 (因此可到达您的实例) 的发起 ping 的Amazon Glue Data Catalog数据库、表和分区。因为共享资源有两个选项Amazon Web Services 账户，此模板创建了两组单独的数据库和表。

这些区域有：Amazon CloudFormation消费者账户模板生成以下资源：

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

在生产者账户中创建你的资源

1. 登录到Amazon CloudFormation在处登录https://console.aws.amazon.com/cloudformation在美国东部(弗吉尼亚北部) 区域。
2. 选择启动堆栈.
3. 选择 Next（下一步）。
4. 适用于堆栈名称，请输入堆栈名称，例如`stack-producer`.
5. 在用户配置部分，输入用户名和密码`ProducerDatalakeAdminUserName`和`ProducerDatalakeAdminUserPassword`.
6. 适用于DataLakeBucketName下，输入您的数据湖存储桶的名称。此名称必须全局唯一。
7. 适用于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. 查看最后一页的详细信息并选择我承认这一点Amazon CloudFormation可能会创建 IAM 资源.
11. 选择 Create（创建）。

    堆栈创建可能需要长达一分钟。

在使用者账户中创建您的资源

1. 登录到Amazon CloudFormation在处登录https://console.aws.amazon.com/cloudformation在美国东部(弗吉尼亚北部) 区域。

2. 选择启动堆栈.
3. 选择 Next（下一步）。
4. 适用于堆栈名称，请输入堆栈名称，例如stack-consumer.
5. 在用户配置部分，输入用户名和密码ConsumerDatalakeAdminUserName和ConsumerDatalakeAdminUserPassword.
6. 适用于DataAnalystUserName和DataAnalystUserPassword，请输入数据分析师 IAM 用户所需的用户名和密码。
7. 适用于DataLakeBucketName下，输入您的数据湖存储桶的名称。此名称必须全局唯一。
8. 适用于DatabaseName，保留默认值。
9. 适用于AthenaQueryResultS3BucketName，请输入用于存储 Amazon S3 存储桶的名称。如果你没有创建 Amazon S3 存储桶.
10. 选择 Next（下一步）。
11. 在下一页上，选择下一步.
12. 查看最后一页的详细信息并选择我承认这一点Amazon CloudFormation可能会创建 IAM 资源.
13. 选择 Create（创建）。

    堆栈创建可能需要长达一分钟。

    Note

    完成教程后，在中删除堆栈Amazon CloudFormation以避免产生费用。在堆栈的事件状态下验证资源是否已成功删除。

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有先决条件。

完整的基于标签的访问控制跨账户共享先决条件

- 在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之前，您必须添加以下内容JSON权限对象添加到创建者账户中的数据目录资源策略。这会授予消费者账户在以下情况下访问数据目录的权限glue:EvaluatedByLakeFormationTagsis true。此外，对于您使用 Lake Formation 权限标签向使用者账户授予权限的资源，此条件也成立。此策略对于每个Amazon Web Services 账户向其授予权限。

  以下策略必须在Statement元素。我们将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 Lake Formation 跨账户授予您照常进行的授权。但是，如果存在 Data Catalog 资源策略，则必须向其添加以下语句，以允许跨账户授予成功（如果这些授权是使用命名资源方法进行的）。如果您计划仅使用命名资源方法，或者仅使用基于标签的访问控制方法，则可以跳过此步骤。在本教程中，我们将评估这两种方法，我们需要添加以下策略。

   以下策略必须在`Statement`元素。我们将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. 下一步，在处登录，在Amazon Glue Data Catalog; 资源策略使用Amazon Command Line Interface(Amazon CLI)。

   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，则必须将`EnableHybrid`添加上述策略时，参数变为 “true”。因为控制台当前不支持此选项，因此您必须使用`glue:PutResourcePolicy`API 和Amazon CLI.

   首先，创建一个策略文档（比如 policy.json），然后添加前面的两个策略。Replace*consumer-account-id*用*## ID*的Amazon Web Services 账户领取补助金，*##*数据目录的区域包含您正在授予权限的数据库和表，以及*account-id*和制片人在一起Amazon Web Services 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"
                }
            }
        }
    ]
}
```

输入以下信息Amazon CLI命令。Replace*glue-resource-policy*使用正确的值（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实现跨账户共享

在本部分中，我们将引导您完成以下概括步骤：

1. 定义一个 LF 标签。
2. 将 LF-tag 分配给目标资源。
3. 向消费者账户授予 LF-tag 权限。
4. 向消费者账户授予数据权限。
5. （可选择）撤消对的权限`IAMAllowedPrincipals`在数据库、表和列上。
6. 创建指向共享表的资源链接。
7. 创建一个 LF-tag 并将其分配给目标数据库。
8. 向消费者账户授予 LF-tag 数据权限。

定义一个 LF-tag

> **Note**
>
> 如果您已登录制作人账户，请先注销，然后再完成以下步骤。

1. 以数据湖管理员身份登录生产者账户，网址为https://console.aws.amazon.com/lakeformation/. 使用生产者账号，IAM 用户名（默认为`DatalakeAdminProducer`），以及您在期间指定的密码Amazon CloudFormation堆栈创建。
2. 在 Lake Formation 控制台上 (https://console.aws.amazon.com/lakeformation/)，在导航窗格中的下，在Permissions (权限)，在下选择管理角色和任务，选择LF 标签.
3. 选择添加 LF-tag.

将 LF-tag 分配给目标资源

将 LF-tag 分配给目标资源并将数据权限授予其他账户

作为数据湖管理员，您可以将标签附加到资源。如果您计划使用单独的角色，则可能需要为单独的角色授予描述权限和附加权限。

1. 在导航窗格中的下，在Data Catalog，选择数据库.
2. 选择目标数据库(lakeformation_tutorial_cross_account_database_tbac)然后在操作菜单中，选择，选择编辑 LF 标签.

   在本教程中，您可以为数据库指定 LF-tag，但也可以为表和列分配 LF 标签。
3. 选择新分配lf-tag.
4. 添加关键帧Confidentiality和Valuepublic.
5. 选择 Save（保存）。

Grantlf-tag消费者账户的权限

仍然在生产者账户中，向消费者账户授予访问 LF-tag 的权限。

1. 在导航窗格中的下，在Permissions (权限)、管理角色和任务、LF-tag 权限，选择Grant.
2. 适用于委托人，选择外部账户.
3. 输入目标Amazon Web Services 账户ID.

   Amazon Web Services 账户在同一组织内自动显示。否则，您必须手动输入Amazon Web Services 账户ID。截至撰写本文时，Lake Formation基于标签的访问控制不支持向组织或组织单位授予权限。
4. 对于 LF 标签，请选择密钥和价值正在与消费者账户共享的 LF-tag（密钥 Confidentiality和值 public)。
5. 适用于Permissions (权限)，选择描述为了LF-tag 权限.

   LF-tag 权限是授予使用者账户的权限。可授予的权限是消费者账户可以授予其他委托人的权限。
6. 选择 Grant（授权）。

   此时，消费者数据湖管理员应该能够找到通过消费者账户 Lake Formation 控制台共享的策略标签，位于Permissions (权限)、管理角色和任务、LF 标签.

向消费者账户授予数据权限

现在，我们将通过指定 LF-tag 表达式并授予消费者账户访问与该表达式匹配的任何表或数据库的权限，来提供对消费者账户的数据访问。

1. 在导航窗格中的下，在Permissions (权限)、数据Lake 权限，选择Grant.
2. 适用于委托人，选择外部账户，然后输入目标Amazon Web Services 账户ID。
3. 适用于LF 标签或目录资源，选择密钥和价值的lf-tag正在与消费者账户共享（密钥 Confidentiality和值 public)。
4. 适用于Permissions (权限)，在下选择LF 标签匹配的资源（推荐）选择添加 LF-tag.
5. 在处登录密钥和值正在与消费者账户（key）共享的标签Confidentiality和Valuepublic)。
6. 适用于数据库权限，选择描述下数据库权限以授予数据库级别的访问权限。
7. 消费者数据湖管理员应该能够在 Lake Formation 控制台上找到通过消费者账户共享的策略标签https://console.aws.amazon.com/lakeformation/，在下选择Permissions (权限)、管理角色和任务、LF 标签.
8. Select描述下可授予的权限因此，消费者账户可以向其用户授予数据库级别的权限。
9. 适用于表和列权限，选择Select和描述下表权限.
10. SelectSelect和描述下可授予的权限.
11. 选择 Grant（授权）。

撤销对的权限IAMAllowedPrincipals在数据库、表和列上（可选）。

在本教程开始时，您更改了Lake Formation 数据目录设置。如果您跳过该部分，则需要执行此步骤。如果您更改了 Lake Foration 数据目录设置，则可以跳过此步骤。

在这一步中，我们需要撤消默认值超级权限来自IAMAllowedPrincipals在数据库或表上。有关详细信息，请参阅 第 4 步：将您的数据存储切换到 Lake Formation 权限模型 (p. 25)。

在撤销的权限之前IAMAllowedPrincipals，请确保您通过 Lake Formation 向现有 IAM 委托人授予了必要的权限。这包括三个步骤：

1. 使用 Lake Formation 向目标 IAM 用户或角色添加 IAM 权限GetDataAccess操作（使用 IAM 策略）。
2. 授予目标 IAM 用户或角色具有 Lake Formation 数据权限（更改、选择等）。
3. 然后，撤消的权限IAMAllowedPrincipals. 否则，在撤消权限后IAMAllowedPrincipals，则现有 IAM 委托人可能无法再访问目标数据库或数据目录。

   正在撤销超级的权限IAMAllowedPrincipals当您想要应用 Lake Formation 权限模型（而不是 IAM 策略模型）来管理单个账户内或使用 Lake Formation 权限模型的多个账户之间的用户访问权限时，必须使用。您不必撤消的权限IAMAllowedPrincipals对于您希望保留传统 IAM 策略模型的其他表。

   此时，消费者账户数据湖管理员应该能够在 Lake Formation 控制台上找到通过消费者账户共享的数据库和表，网址为https://console.aws.amazon.com/lakeformation/，在下选择数据目录、数据库. 如果没有，请确认是否正确配置了以下各项：

   1. 将正确的策略标记和值分配给目标数据库和表。
   2. 将正确的标签权限和数据权限分配给消费者账户。
   3. 撤销默认的超级权限IAMAllowedPrincipals在数据库或表上。

创建指向共享表的资源链接

当资源在账户之间共享时，共享的资源未放入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，我们需要创建指向共享表的资源链接。资源链接是一种数据目录对象，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 (p. 126)通过创建资源链接，您可以：

- 为符合数据目录资源命名策略的数据库或表分配不同的名称。
- 使用 Athena 和 Redshift Spectrum 等服务查询共享数据库或表。

要创建资源链接，请完成以下步骤：

1. 如果您已登录消费者账户，请注销。
2. 以消费者账户数据湖管理员身份登录。使用消费者账户 ID、IAM 用户名（默认）DatalakeAdminConsumer) 和您在期间指定的密码Amazon CloudFormation堆栈创建。
3. 在 Lake Formation 控制台上 (https://console.aws.amazon.com/lakeformation/)，在导航窗格中的下，在数据目录、数据库中，选择共享数据库lakeformation_tutorial_cross_account_database_tbac.

   如果看不到数据库，请重新访问前面的步骤，看看是否所有内容都配置正确。
4. 选择查看表.
5. 选择共享表amazon_reviews_table_tbac.
6. 在存储库的操作菜单中，选择，选择创建资源链接.
7. 适用于资源链接名称，请输入名称（在本教程中，amazon_reviews_table_tbac_resource_link)。
8. UNDER数据库，选择在其中创建资源链接的数据库（对于这篇文章，Amazon CloudFormationn 堆栈创建了数据库lakeformation_tutorial_cross_account_database_consumer)。
9. 选择 Create（创建）。

   资源链接显示在Datog、表.

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。这意味着在向消费者账户中的资源链接授予访问权限时，在生产者账户中创建的标签不可用。在使用者账户中共享资源链接时，您需要在消费者账户中创建一组单独的 LF 标签，以便使用基于 LF 标签的访问控制。

1. 在使用者账户中定义 LF-tag。在本教程中，我们使用 key`Division`和值`sales`、`marketing`, 和`analyst`.
2. 分配 LF-tag 键`Division`和Value`analyst`到数据库中`lakeformation_tutorial_cross_account_database_consumer`，其中创建资源链接。

向消费者授予 LF-tag 数据权限

最后一步，向消费者授予 LF-tag 数据权限。

1. 在导航窗格中的下，在Permissions (权限)、数据Lake 权限，选择Grant.
2. 适用于委托人，选择IAM 用户和角色，然后选择用户`DataAnalyst`.
3. 适用于LF 标签或目录资源，选择LF 标签匹配的资源(建议).
4. 选择密钥除值分析人员。
5. 适用于数据库权限，选择描述下数据库权限.
6. 适用于表和列权限，选择Select和描述下表权限.
7. 选择 Grant（授权）。
8. 对用户重复这些步骤`DataAnalyst`，其中 LF-tag 密钥在哪里`Confidentiality`和有效值`public`.

   此时，消费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https://console.aws.amazon.com/athena/. 如果没有，请确认是否正确配置了以下各项：

   - 已为共享表创建资源链接
   - 您已授予用户访问由制作者账户共享的 LF-tag 的权限
   - 您已授予用户访问与创建资源链接时所在的资源链接和数据库关联的 LF-tag 的访问权限
   - 检查是否为资源链接以及创建资源链接的数据库分配了正确的 LF-tag

第 4 步：实现命名资源方法

要使用 named 资源方法，我们将引导您完成以下概括步骤：

1. （可选）撤消对的权限`IAMAllowedPrincipals`在数据库、表和列上。
2. 向消费者账户授予数据权限。
3. 接受来自的资源共享Amazon Resource Access Manager.
4. 为共享表创建资源链接。
5. 将共享表的数据权限授予使用者。
6. 将资源链接的数据权限授予使用者。

撤销对的权限`IAMAllowedPrincipals`在数据库、表和列上（可选）

- 在本教程开始时，我们更改了Lake Formation 数据目录设置。如果您跳过该部分，则需要执行此步骤。有关说明，请参阅前一部分中的可选步骤。

向消费者账户授予数据权限

1. Note

   如果您以其他用户身份登录到生产者账户，请先注销。

   登录Lake Formation 控制台https://console.aws.amazon.com/lakeformation/使用生产者账户数据湖管理员使用Amazon Web Services 账户ID、IAM 用户名（默认为`DatalakeAdminProducer`），并在期间指定的密码Amazon CloudFormation堆栈创建。
2. 在存储库的Permissions (权限)页面，位于数据Lake 权限选择Grant.
3. UNDER委托人，选择外部账户，然后输入一个或多个Amazon Web Services 账户IdAmazon组织 ID。有关更多信息，请参阅：AmazonOrganizations.

   生产者账户所属的Organizations 和Amazon Web Services 账户在同一组织内自动显示。否则，请手动输入账户 ID 或组织 ID。
4. 适用于LF 标签或目录资源，选择`Named data catalog resources`.
5. UNDER数据库，选择数据库`lakeformation_tutorial_cross_account_database_named_resource`.
6. 选择添加 LF-tag.
7. UNDER表，选择所有桌子.
8. 适用于表列权限选择Select, 和描述下表权限.
9. SelectSelect和描述，在下选择可授予权限.
10. （可选）对于数据权限，选择基于列的简单访问如果需要列级权限管理。
11. 选择 Grant（授权）。

如果您尚未撤销对的权限`IAMAllowedPrincipals`，你得到一个授予权限错误失败。此时，您应该会看到目标表正在通过共享Amazon RAM消费者账户低于权限、数据权限.

接受来自的资源共享Amazon RAM

Note

这个步骤仅在以下情况下为必填项Amazon Web Services 账户基于组织的共享，不适用于基于组织的共享。

1. 登录到Amazon在处登录https://console.aws.amazon.com/connect/使用消费者账户数据湖管理员使用 IAM 用户名（默认为 DatalakeAdminConsumer) 和期间指定的密码Amazon CloudFormation堆栈创建。
2. 在存储库的Amazon RAM控制台，在导航窗格中的下，在与我共享和 Resource shares中，选择共享的 Lake Formation 资源。这些区域有：状态应为Pending.
3. 选择操作和Grant.
4. 确认资源详细信息，然后选择接受资源共享.

   此时，消费者账户数据湖管理员应该能够在 Lake Formation 控制台上找到共享资源（https://console.aws.amazon.com/lakeformation/) UNDERData Catalog、数据库.

为共享表创建资源链接

- 按照中的说明进行操作第 3 步：使用基于标签的访问控制方法实现跨账户共享 (p. 91)（步骤 6）为共享表创建资源链接。命名资源链接`amazon_reviews_table_named_resource_resource_link`. 在数据库中创建资源链接`lakeformation_tutorial_cross_account_database_consumer`.

将共享表的数据权限授予使用者

要向使用者授予共享表的数据权限，请完成以下步骤：

1. 在 Lake 编队控制台上 (https://console.aws.amazon.com/lakeformation/)，在下选择Permissions (权限)、数据Lake 权限，选择Grant.
2. 适用于委托人，选择IAM 用户和角色，然后选择用户`DataAnalyst`.
3. 适用于LF 标签或目录资源，选择命名数据目录资源.
4. UNDER数据库，选择数据库`lakeformation_tutorial_cross_account_database_named_resource`. 如果您在下拉列表中看不到数据库，请选择加载更多.
5. UNDER表，选择表`amazon_reviews_table_named_resource`.
6. 适用于表和列权限，选择Select和描述下表权限.
7. 选择 Grant（授权）。

将资源链接的数据权限授予使用者

除了授予数据湖用户访问共享表的权限外，您还需要授予数据湖用户访问资源链接的权限。

1. 在 Lake Formation 控制台上 (https://console.aws.amazon.com/lakeformation/)，在下选择Permissions (权限)、数据Lake 权限，选择Grant.
2. 适用于委托人，选择IAM 用户和角色，然后选择用户`DataAnalyst`.
3. 适用于LF 标签或目录资源，选择命名数据目录资源.
4. UNDER数据库，选择数据库`lakeformation_tutorial_cross_account_database_consumer`. 如果您在下拉列表中看不到数据库，请选择加载更多.
5. UNDER表，选择表`amazon_reviews_table_named_resource_resource_link`.
6. 适用于资源链接权限，选择描述下资源链接权限.
7. 选择 Grant（授权）。

   此时，消费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。

   如果没有，请确认是否正确配置了以下内容：

   - 已为共享表创建资源链接
   - 您授予了用户对创建者账户共享的表的访问权限
   - 您已授予用户访问为其创建资源链接的资源链接和数据库的访问权限

第 5 步：清理 Amazon资源

为了防止向您收取不必要的费用Amazon Web Services 账户，您可以删除Amazon您在本教程中使用的资源。

1. 登录到Lake Formation 控制台https://console.aws.amazon.com/lakeformation/使用生产者账户并删除或更改以下内容：

   - Amazon Resource Access Manager资源共享
   - Lake Formation
   - Amazon CloudFormation 堆栈
   - Lake Formation
   - Amazon Glue Data Catalog
2. 登录到Lake Formation 控制台https://console.aws.amazon.com/lakeformation/使用消费者账户并删除或更改以下内容：

   - Lake Formation

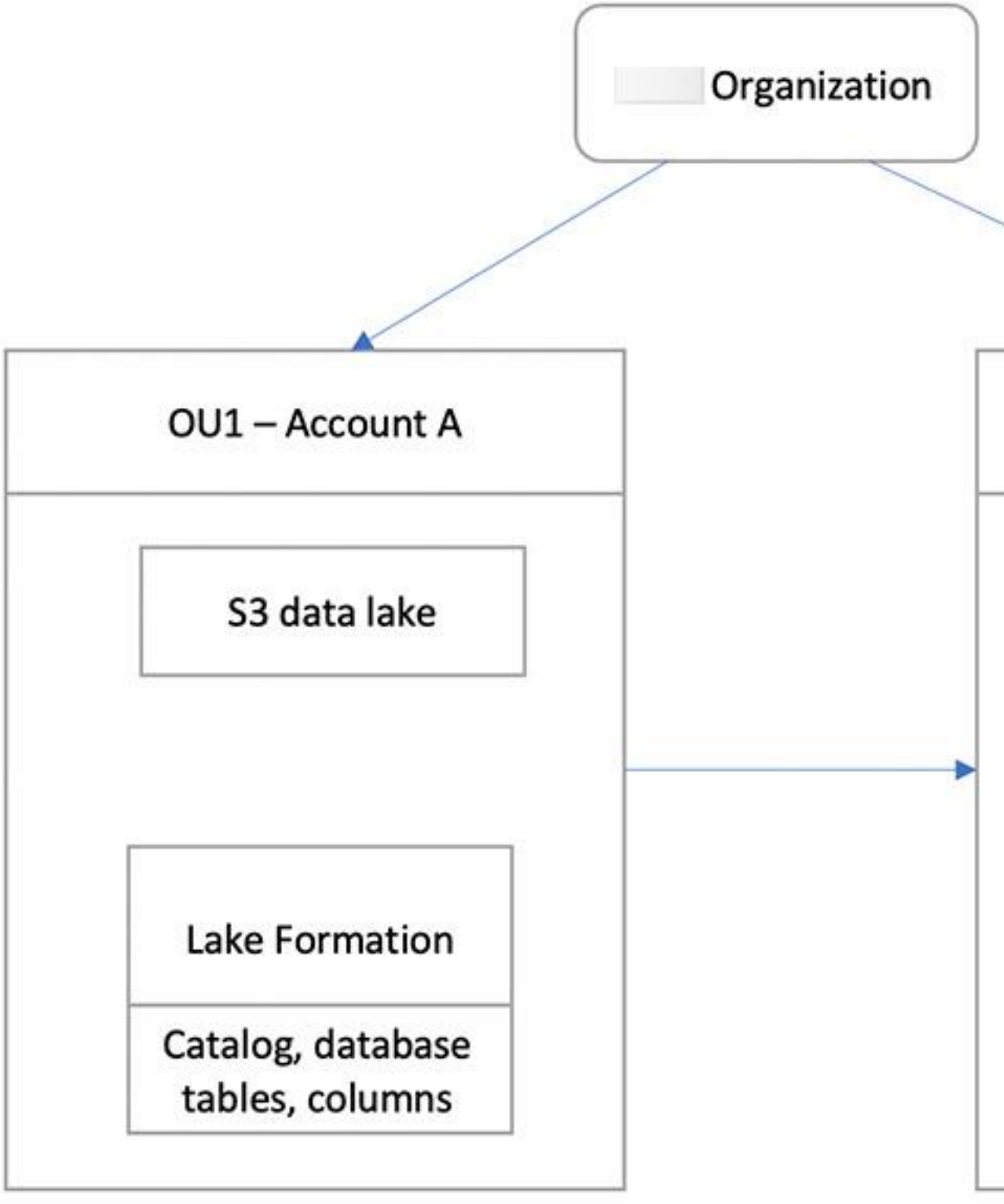
- Amazon CloudFormation 堆栈

使用 Lake Formation 精细访问控制

本教程提供 step-by-step 关于在管理多个数据集时如何使用 Lake Formation 快速轻松地共享数据集的说明Amazon Web Services 账户和Amazon Organizations. 您可以定义精细权限来控制对敏感数据的访问。

以下步骤还显示了账户 A 的数据湖管理员如何为账户 B 提供精细访问权限，以及账户 B 中的用户如何为其账户中的其他用户授予对共享表的精细访问权限。每个账户中的数据管理员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。

用例假设你正在使用Amazon Organizations来管理您的Amazon Web Services 账户. 一个组织单位 (OU1) 中账户 A 的用户向 OU2 中账户 B 的用户授予访问权限。在不使用 Organizations 时，例如只有几个账户时，您可以使用相同的方法。下图说明了数据湖中数据集的精细访问控制。账户 A 中的数据湖可用。账户 A 的数据湖管理员为账户 B 提供精细访问权限。该图还显示，账户 B 的用户向账户 B 中的其他用户提供账户 A 数据湖表的列级访问权限。


Organization
OU1 – Account A
S3 data lake
Lake Formation
Catalog, database
tables, columns

主题

目标受众

本教程适用于数据管理员、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	可以创建 IAM 用户和角色以及 Amazon S3 存储桶的用户。Has`AdministratorAccess` Amazon管理的策略。
数据湖管理员	拥有的用户Amazon托管策略`AWSLakeFormationDataAdmin`将附加。
数据分析人员	拥有的用户Amazon管理的策略`AmazonAthenaFullAccess`将附加。

先决条件

在开始本教程之前，您必须具有Amazon Web Services 账户您可以登录到Amazon Identity and Access Management(IAM) 具有正确权限的用户。有关更多信息，请参阅 注册Amazon (p. 10) 和 创建 IAM 管理员用户 (p. 10)。

本教程假设您已熟悉 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 Simple Storage S3（Amazon S3）数据湖位置
- 账户 A 中的数据湖管理员用户您可以使用 Lake Formation 控制台创建数据湖管理员 (https://console.aws.amazon.com/lakeformation/) 或者`PutDataLakeSettings`Lake Formation API 的操作。
- 在账户 A 中配置了 Lake Formation，Amazon S3 数据湖位置在账户 A 中注册了 Lake Formation
- 账户 B 中具有以下 IAM 托管策略的两个用户：
  - testuser1 — 有Amazon托管策略`AWSLakeFormationDataAdmin`将附加。
  - testuser2 — 有Amazon管理的策略`AmazonAthenaFullAccess`将附加。
- 账户 B 的 Lake Formation 数据库中的数据库 testdb

第 1 步：提供针对其他账户的精细访问权限

了解账户 A 的数据湖管理员如何为账户 B 提供精细访问权限。

授予另一个账户精细访问权限

1. 登录Amazon Web Services Management Console在https://console.aws.amazon.com/connect/以数据湖管理员的身份进入账户 A。
2. 打开 Lake Formation 控制台 (https://console.aws.amazon.com/lakeformation/ ），然后选择试用.
3. 在导航窗格中，选择数据库.
4. 选择创建数据库。
5. 在数据库详细信息部分，选择数据库.
6. 适用于名称，请输入名称（在本教程中，我们将使用`sampledb01`)。
7. 确保仅对该数据库中的新表使用 IAM 访问控制未被选中。不选中此选项允许我们控制来自Lake Formation 的进入。
8. 选择 Create database（创建数据库）。
9. 在存储库的数据库页面，选择您的数据库`sampledb01`.
10. 在存储库的操作菜单上，选择Grant.
11. 在授予权限部分，选择External 账户.
12. 适用于Amazon Web Services 账户ID 或Amazon组织 ID，输入 OU2 中账户 B 的账户 ID。
13. 适用于表，选择你希望账户 B 有权访问的表（在这篇文章中，我们使用表`acc_a_area`)。或者，您可以授予对表中列的访问权限，我们在本文中就是这样做的。
14. 适用于包括列选择您希望账户 B 有权访问的列（对于这篇文章，我们授予了键入、名称和标识符的权限）。
15. 适用于列，选择包括列.
16. 适用于表权限，选择Select.
17. 适用于可授予权限，选择Select. 需要可授予的权限，这样账户 B 中的管理员用户才能向账户 B 中的其他用户授予权限。
18. 选择 Grant（授权）。
19. 在导航窗格中，选择表。
20. 你可以在里面看到一个活跃的连接Amazon Web Services 账户和Amazon有访问权限的组织。

创建资源链接

像 Amazon Athena 这样的集成服务无法直接跨账户访问数据库或表。因此，您需要创建资源链接，以便 Athena 可以访问您账户中的资源链接，指向其他账户中的数据库和表。创建指向表格的资源链接 (`acc_a_area`）因此账户 B 用户可以向 Athena 查询其数据。

1. 登录到Amazon控制台处登录https://console.aws.amazon.com/connect/在账户 B 中`testuser1`.
2. 在 Lake Formation 控制台上 (https://console.aws.amazon.com/lakeformation/ ），在导航窗格中，选择表. 您应该看到账户 A 提供访问权限的表格。
3. 选择 `acc_a_area` 表。
4. 在存储库的操作菜单上，选择创建资源链接.
5. 适用于资源链接名称，请输入名称（在本教程中，请输入名称（在本教程中，`acc_a_area_rl`)。
6. 适用于数据库，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. 在存储库的操作菜单上，选择查看数据.

    您将被重定向到 Athena 控制台，在其中您应看到数据库和表。

    现在，您可以对表运行查询，查看账户 B 向 testuser1 提供了访问权限的列值。

第 2 步：为同一账户的用户提供精细访问权限

本节显示了账户 B 中的用户 (testuser1），充当数据管理员，为同一账户中的其他用户提供精细访问权限（testuser2) 到共享表中的列名aac_b_area_rl.

授予同一个账户的用户精细访问权限

1. 登录到Amazon控制台处登录https://console.aws.amazon.com/connect/在账户 B 中testuser1.
2. 在 Lake Formation 控制台上，在导航窗格中选择表.

   您可以通过表的资源链接授予表的权限。为此，请在处登录表页面，选择资源链接acc_b_area_rl，然后在操作菜单上，选择在目标上授予权限.
3. 在授予权限部分，选择我的账户.
4. 适用于IAM 用户和角色选择用户testuser2.
5. 适用于列，选择列名。
6. 适用于表权限，选择Select.
7. 选择 Grant（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，您需要授予资源链接本身的权限。你需要授予描述要么下降权限。在存储库的表表页面，再次选择你的桌子，然后在操作菜单上，选择Grant.
8. 在授予权限部分，选择我的账户.
9. 适用于IAM 用户和角色，选择用户testuser2.
10. 适用于资源链接权限或者选择描述.
11. 选择 Grant（授权）。
12. 登录到Amazon账户 B 中的控制台为testuser2.

    在 Athena 控制台上 (https://console.aws.amazon.com/athena/)，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，查看该列的值testuser2有权访问。

将 Amazon S3 位置添加到您的数据湖

要将 Amazon Simple Storage Service (Amazon S3) 位置添加为数据湖中的存储，注册使用的位置Amazon Lake Formation. 然后，您可以使用 Lake Formation 权限进行精细的访问控制，Amazon Glue Data Catalog 指向此位置以及该位置中的基础数据的对象。

当您注册某个位置时，该 Amazon S3 路径和该路径下的所有文件夹都将被注册。

例如，假设您具有如下所示的 Amazon S3 路径组织：

```
/mybucket/accounting/sales/
```

如果您注册`S3://mybucket/accounting`，`sales`文件夹也已注册并在 Lake Formation 管理下。

有关注册位置的更多信息，请参阅Underlying Data Access Control (p. 238).

主题

注册位置时使用的角色的要求

您必须指定Amazon Identity and Access Management注册 Amazon Simple Storage Service (Amazon S3) 位置时的角色。Amazon在访问位置的数据时，Lake Formation 将代入该角色。

您可以使用以下角色类型之一注册位置：

- Lake Formation 服务相关角色。此角色授予该位置所需的权限。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 对 Lake Formation 使用服务相关角色 (p. 256)。
- 用户定义的角色。如果您需要授予超过服务相关角色所提供的权限，则可以使用用户定义的角色。

  在下列情况下，您必须使用用户定义的角色：

  - 注册受管理表指向的位置时。

    有关更多信息，请参阅 管理受管表 (p. 114)。
  - 在另一个账户中注册位置时。

    有关更多信息，请参阅 the section called “在另一个位置注册 Amazon S3Amazon帐户” (p. 108) 和 the section called “注册加密 Amazon S3 位置Amazon账户” (p. 109)。
  - 如果您使用Amazon托管 CMK (`aws/s3`) 以加密 Amazon S3 位置。

    有关更多信息，请参阅 注册加密的 Amazon S3 位置 (p. 105)。

- 如果您计划使用亚马逊 EMR 访问该位置。

  如果您已经使用服务相关角色注册了一个位置，并希望开始使用 Amazon EMR 访问该位置，则必须注销该位置并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” (p. 112)。

以下是用户定义角色的要求：

- 创建新角色时，请在创建角色在 IAM 控制台的页面中，选择Amazon服务，然后在选择使用案例，选择Lake Formation.

  如果您使用不同的路径创建角色，请确保角色与之建立信任关系`lakeformation.amazonaws.com`. 有关更多信息，请参阅 。修改角色信任策略（控制台）.
- 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）.
- 角色必须具有内联策略，该策略授予 Amazon S3 对该位置的读/写权限。以下是一项典型策略。

```
{
    "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拥有`iam:PassRole`角色的权限。

  以下是授予此权限的内联策略。Replace*<account-id>*拥有有效的Amazon账号，然后替换*<role-name>*与角色的名称相结合使用。

```
{
    "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 Formation 在 CloudWatch Logs 中添加日志并发布指标，请添加以下内联策略。

  Note

  写入 CloudWatch Logs 将产生一定的费用。

```
{
    "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

您必须指定Amazon Identity and Access Management注册 Amazon Simple Storage Service (Amazon S3) 位置时的 (IAM) 角色。Lake Formation 在向集成版授予临时证书时担任该角色Amazon访问该位置中数据的服务。

Important

避免注册具有 Amazon S3 存储桶申请方付款已启用. 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。如果存储桶被另一个人访问Amazon账户，如果该角色与存储桶拥有者属于同一账户，则该存储桶拥有者需要支付数据访问费用。

您可以使用Amazon Lake Formation控制台、Lake Formation API 或Amazon Command Line Interface(Amazon CLI注册 Amazon S3 位置。

开始前的准备工作

查看用于注册位置的角色的要求 (p. 102).

注册位置（控制台）

Important

以下过程假设 Amazon S3 位置位于同一个位置Amazon帐户作为数据目录，而且该位置中的数据未加密。本章的其他章节介绍了跨账户注册和加密地点的注册。

1. 打开Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份或以用户身份登录`lakeformation:RegisterResource`IAM 权限。
2. 在导航窗格中的下，注册并提取，选择数据湖位置.
3. 选择注册位置，然后选择浏览选择 Amazon Simple Storage Service (Amazon S3) 路径。
4. （可选，但强烈建议）选择查看位置权限查看所选 Amazon S3 位置中所有现有资源及其权限的列表。

   注册选定的位置可能会导致您的 Lake Formation 用户获得该位置已存在的数据的访问权限。查看此列表有助于确保现有数据保持安全。
5. 适用于IAM 角色选择，选择`AWSServiceRoleForLakeFormationDataAccess`服务相关角色（默认）或符合中要求的自定义 IAM 角色the section called “注册位置时使用的角色的要求” (p. 102).
6. 选择注册位置.

要注册位置（Amazon CLI）

- 输入以下 CLI 命令。Replace*<s3-path>*使用有效的 Amazon S3 路径。

```
aws lakeformation register-resource --resource-arn arn:aws:s3:::<s3-path> --use-
service-linked-role
```

此命令使用服务相关角色注册位置。您可以使用`--role-arn`而是为了提供你自己的角色。

有关更多信息，请参阅 RegistRegerResource 操作（Python：register _resource）(p. 284)。

> **Note**
>
> 注册 Amazon S3 位置后，Amazon Glue指向该位置（或其任何子位置）的表格将返回`IsRegisteredWithLakeFormation`参数为`true`中的`GetTable`调用。数据目录 API 操作有一个已知的限制，例如`GetTables`和`SearchTables`不要更新的值`IsRegisteredWithLakeFormation`参数，然后返回默认值，为 false。建议使用`GetTable`用于查看的正确值的 API`IsRegisteredWithLakeFormation`参数。

注册加密的 Amazon S3 位置

Lake FormationAmazon Key Management Service(Amazon KMS），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。

两位客户管理Amazon KMS keys和Amazon 托管式密钥支持。不支持客户端加密/解密。

您必须指定Amazon Identity and Access Management注册 Amazon S3 位置时的 (IAM) 角色。对于加密 Amazon S3 位置，角色必须具有加密和解密数据的权限。Amazon KMS key，或者 KMS 密钥策略必须向角色授予密钥的权限。

> **Important**
>
> 避免注册具有 Amazon S3 存储桶申请方付款已启用. 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。如果存储桶被另一个人访问Amazon账户，如果该角色与存储桶拥有者属于同一账户，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。此角色授予该位置所需的读/写权限。您也可以使用自定义角色来注册该位置，前提是该位置符合the section called “注册位置时使用的角色的要求” (p. 102).

Important

如果您使用Amazon 托管式密钥(aws/s3) 要加密 Amazon S3 位置，您不能使用 Lake Formation 服务相关角色。您必须使用自定义角色并向该角色添加密钥的 IAM 权限。此部分的后文提供了详细信息。

以下过程说明了如何注册使用客户托管密钥或使用客户托管密钥加密的 Amazon S3 位置Amazon 托管式密钥.

- 注册使用客户托管式密钥加密的位置 (p. 106)
- 注册使用加密的位置Amazon 托管式密钥 (p. 107)

开始前的准备工作

查看用于注册位置的角色的要求 (p. 102).

要注册使用客户托管式密钥加密的 Amazon S3 位置

Note

如果 KMS 密钥或 Amazon S3 位置不在同一个位置Amazon将帐户作为数据目录，请按照中的说明操作the section called “注册加密 Amazon S3 位置Amazon账户” (p. 109)相反。

1. 打开Amazon KMS控制台https://console.aws.amazon.com/kms然后以Amazon Identity and Access Management(IAM) 管理用户或可以修改用于加密位置的 KMS 密钥的密钥策略的用户。
2. 在导航窗格中，选择客户托管密钥，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，然后执行以下操作之一以 KMS 密钥用户身份添加自定义角色或 Lake Formation 服务相关角色：

   - 如果显示默认视图（带关键管理员、删除密钥、关键用户, 和其他Amazon账户部分）— 在关键用户部分中，添加自定义角色或 Lake Formation 服务相关角色AWSServiceRoleForLakeFormationDataAccess.
   - 如果显示密钥策略 (JSON)— 编辑策略以添加自定义角色或 Lake Formation 服务相关角色AWSServiceRoleForLakeFormationDataAccess转到对象 “允许使用密钥”，如以下示例所示。

     Note

     如果该对象丢失，请使用示例中显示的权限添加它。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Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份或以用户身份登录`lakeformation:RegisterResource`IAM 权限。
5. 在导航窗格中的下，注册并提取，选择数据湖位置.
6. 选择注册位置，然后选择浏览选择 Amazon Simple Storage Service (Amazon S3) 路径。
7. （可选，但强烈建议）选择查看位置权限查看所选 Amazon S3 位置中所有现有资源及其权限的列表。

   注册选定的位置可能会导致您的 Lake Formation 用户获得该位置已存在的数据的访问权限。查看此列表有助于确保现有数据保持安全。
8. 适用于IAM 角色选择，选择`AWSServiceRoleForLakeFormationDataAccess`服务相关角色（默认）或符合the section called “注册位置时使用的角色的要求” (p. 102).
9. 选择注册位置.

有关 service-linked role 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 (p. 257)。

要注册使用加密的 Amazon S3 位置Amazon 托管式密钥

> Important
>
> 如果 Amazon S3 位置不在同一位置Amazon将帐户作为数据目录，请按照中的说明操作the section called “注册加密 Amazon S3 位置Amazon账户” (p. 109)相反。

1. 创建用于注册位置的 IAM 角色。确保它满足中列出的要求the section called “注册位置时使用的角色的要求” (p. 102).
2. 将下面的内联策略添加到角色。它向角色授予对密钥的权限。这些区域有：`Resource`规范必须指定的 Amazon 资源名称 (ARN)Amazon 托管式密钥. 您可 ARN 从Amazon KMS控制台。要获得正确的 ARN，请确保您登录到Amazon KMS具有相同的控制台Amazon账户和区域为Amazon 托管式密钥该位置被用来加密位置。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "<Amazon ##### ARN>"
    }
  ]
}
```

3. 打开Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份或以用户身份登录`lakeformation:RegisterResource`IAM 权限。
4. 在导航窗格中的下，注册并提取，选择数据湖位置.
5. 选择注册位置，然后选择浏览选择 Amazon S3 路径。
6. （可选，但强烈建议）选择查看位置权限查看所选 Amazon S3 位置中所有现有资源及其权限的列表。

   注册选定的位置可能会导致您的 Lake Formation 用户获得该位置已存在的数据的访问权限。查看此列表有助于确保现有数据保持安全。
7. 适用于IAM 角色，选择您在步骤 1 中创建的角色。

8. 选择注册位置.

在另一个位置注册 Amazon S3Amazon帐户

Amazon Lake Formation使您能够注册 Amazon Simple Storage Service (Amazon S3)Amazon账户。例如，如果Amazon Glue Data Catalog在账户 A 中，账户 A 中的用户可以在账户 B 中注册 Amazon S3 存储桶

在中注册 Amazon S3 存储桶Amazon账户 B 使用Amazon Identity and Access Management(IAM) 角色Amazon账户 A 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**
>
> 避免注册具有 Amazon S3 存储桶申请方付款已启用. 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。如果存储桶被另一个人访问Amazon账户，如果该角色与存储桶拥有者属于同一账户，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 服务相关角色在另一个账户中注册位置。而必须使用用户定义的角色。该角色必须满足中的要求the section called “注册位置时使用的角色的要求” (p. 102). 有关 service-linked role 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 (p. 257)。

开始前的准备工作

查看用于注册位置的角色的要求 (p. 102).

在另一个地点注册Amazon帐户

> **Note**
>
> 如果该位置已加密，请按照中的说明操作the section called “注册加密 Amazon S3 位置Amazon账户” (p. 109)相反。

以下过程假定账户 1111-2222-3333（包含数据目录）中的委托人想注册 Amazon S3 存储桶`awsexamplebucket1`在账户 1234-5678-9012 中。

1. 在账户 1111-2222-3333 中，登录Amazon Web Services Management Console在以下网址打开 IAM 控制台：https://console.amazonaws.cn/iam/.
2. 创建新角色或查看符合中要求的现有角色the section called “注册位置时使用的角色的要求” (p. 102). 确保该角色向 Amazon S3 授予以下权限：`awsexamplebucket1`.
3. 通过以下网址打开 Amazon S3 控制台：https://console.aws.amazon.com/s3/。使用账户 1234-5678-9012 登录。
4. 在Bucket name列表中，选择存储桶名称`awsexamplebucket1`.
5. 选择权限。
6. 在存储库的Permissions (权限)在页面上，选择存储桶策略.
7. 在存储桶策略编辑，粘贴以下策略。Replace*<role-name>*以你的角色的名称。

```
{
    "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Amazon Lake Formation控制台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帐户 1111-2222-3333。
10. 在导航窗格中的下注册并提取，选择数据湖位置.
11. 选择注册位置.
12. 在存储库的注册位置页，对于Amazon S3 路径中，输入存储桶名称`s3://awsexamplebucket1`.

    Note

    您必须键入存储桶名称，因为当您选择时跨账户存储桶不会显示在列表中浏览.
13. 适用于IAM 角色选择你的角色。
14. 选择注册位置.

注册加密 Amazon S3 位置Amazon账户

Amazon Lake Formation已与 集成Amazon Key Management Service(Amazon KMS），以便您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。

客户托管密钥和Amazon 托管式密钥支持。不支持客户端加密/解密。

Important

避免注册 Amazon S3 存储桶中的申请方付款已启用. 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。如果存储桶被另一个人访问Amazon账户，如果该角色与存储桶拥有者属于同一账户，则该存储桶拥有者需要支付数据访问费用。

本部分介绍了在以下情况下如何注册 Amazon S3 位置：

- Amazon S3 位置中的数据使用中创建的 KMS 密钥进行加密Amazon KMS.
- Amazon S3 位置不同的Amazon帐户作为Amazon Glue Data Catalog.
- KMS 密钥要么是要么不一样Amazon账户为数据目录。

注册Amazon KMS— 加密 Amazon S3 存储桶Amazon账户 B 使用Amazon Identity and Access Management(IAM) 角色Amazon账户 A 需要以下权限：

- 账户 A 中的角色必须授予账户 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 密钥在账户 B 中，则密钥策略必须授予对账户 A 中角色的访问权限，而账户 A 中的角色必须授予对 KMS 密钥的权限。

在以下过程中，您可以在Amazon包含数据目录的帐户（上一讨论中的账户 A）。然后，您可以使用此角色注册该位置。在访问 Amazon S3 中的底层数据时，Lake Formation 担任此角色。代入的角色具有 KMS 密钥所需的权限。因此，您不必将 KMS 密钥的权限授予委托人使用 ETL 作业或集成服务（例如）访问底层数据的委托人Amazon Athena.

> **Important**
>
> 您不能使用 Lake Formation 服务相关角色在另一个账户中注册位置。而必须使用户定义的角色。该角色必须满足中的要求the section called “注册位置时使用的角色的要求” (p. 102). 有关 service-linked role 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 (p. 257)。

开始前的准备工作

查看用于注册位置的角色的要求 (p. 102).

注册 Amazon S3 加密位置Amazon账户

1. 同样的Amazon帐户作为数据目录，登录Amazon Web Services Management Console在以下网址打开 IAM 控制台：https://console.amazonaws.cn/iam/.
2. 创建新角色或查看符合中要求的现有角色the section called “注册位置时使用的角色的要求” (p. 102). 确保角色包括向 Amazon S3 授予对该位置的权限的策略。
3. 如果 KMS 密钥与数据目录不在同一个账户中，请向角色添加一个内联策略，以授予对 KMS 密钥的所需权限。以下是示例 策略。Replace*<cmk-region>*和*<cmk-account-id>*其中包含 KMS 密钥的地区和账号。Replace*<key-id>*带密钥 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，添加一个存储桶策略，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。Replace*<catalog-account-id>*使用Amazon数据目录的账号，*<role-name>*使用你的角色的名称，和*<bucket-name>*使用存储桶的名称。

```
{
    "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. InAmazon KMS中，将角色添加为 KMS 密钥的用户。

   a. 访问 https://console.aws.amazon.com/kms，打开 Amazon KMS 控制台。然后，以 IAM 管理员或可以修改用于加密位置的 KMS 密钥的密钥策略的用户身份登录。

   b. 在导航窗格中，选择客户托管密钥，然后选择 KMS 密钥的名称。

   c. 在 KMS 密钥详细信息页面上的密钥策略选项卡中，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，选择编辑将角色的 Amazon 资源名称 (ARN) 添加到`Allow use of the key`对象，如以下示例所示。

      Note

      如果该对象丢失，请使用示例中显示的权限添加它。

```
        ...
        {
            "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 。允许其他账户中的用户使用 KMS 密钥中的Amazon Key Management Service开发人员指南.

6. 打开Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/. 登录到数据目录Amazon帐户作为数据湖管理员。
7. 在导航窗格中的下注册和摄取，选择数据湖位置.
8. 选择注册位置.
9. 在存储库的注册位置页面，对于Amazon S3 路径中，输入位置路径为**s3://<bucket>/<prefix>**. Replace<bucket>和存储桶的名称和<prefix>还有该位置的其余路径。

   Note

   您必须键入路径，因为当您选择时跨账户存储桶不会出现在列表中浏览.

10. 适用于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

如果您不想再由 Lake Formation 管理，则可以取消注册 Amazon Simple Storage Service (Amazon S3) 位置。取消注册位置不会影响 Lake Formation 在该位置上授予的数据位置权限。您可以重新注册已取消注册的位置，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。

要取消注册位置，请执行以下操作：

1. 打开AmazonLake Formation 控制台在https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份或以用户身份登录`lakeformation:RegisterResource`IAM 权限。
2. 在导航窗格中的下注册并采集，选择数据湖位置.
3. 选择一个位置，然后在操作菜单中，选择Remove.
4. 当系统提示进行确认时，选择Remove.

管理数据目录表和数据库

Amazon Lake Formation使用Amazon Glue数据目录来存储有关数据湖、数据源、转换和目标的元数据。有关数据源和目标的元数据采用数据库和表格的形式。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帐户中的共享数据库和表的链接，用于跨账户访问数据湖中的数据。

EAERAmazon每个账户都有一个数据目录Amazon区域。

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，也可以为每个数据库授予不同的 Lake Formation 权限。

数据库可以有可选的位置属性。该位置通常位于在 Lake Formation 中注册的 Simple Storage Service (Amazon S3) Storage Storage Storage Storage Storage Storage S 指定位置时，承担者不需要数据位置权限即可创建指向数据库位置内的位置的数据目录表。有关更多信息，请参阅 Underlying Data Access Control (p. 239)。

要使用 Lake Formation 控制台创建数据库，您必须以数据湖管理员身份登录或数据库创建. 数据库创建者是被授予 Lake Formation 的委托人`CREATE_DATABASE`权限。您可以在上看到数据库创建者的列表。管理员和数据库创建者Lake Formation 控制台的页面。要查看此列表，您必须拥有`lakeformation:ListPermissions`IAM 权限并以数据湖管理员或数据库创建者身份登录，并在`CREATE_DATABASE`权限。

创建数据库

1. 打开Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/，并以数据湖管理员或数据库创建者身份登录。
2. 在导航窗格中的下，数据目录，选择数据库.
3. 选择 Create database（创建数据库）。
4. 在创建数据库对话框中，输入数据库名称、可选位置和可选说明。
5. 可选择选择仅对此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 “更改数据湖的默认安全设置” (p. 253)。
6. 选择 Create database（创建数据库）。

创建表

Amazon 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Amazon Glue数据目录。您可以使用它们访问数据湖中的底层数据，并使用 Lake Formation 权限管理这些数据。数据目录中的数据库存储元数据表。

创建数据目录表有多种方式：

- 运行爬网程序Amazon Glue. 请参阅定义爬网程序中的Amazon Glue开发人员指南.
- 创建并运行工作流。请参阅 使用工作流程导入数据 (p. 134)。
- 使用 Lake Formation 控制台手动创建表格，Amazon GlueAPI，或者Amazon Command Line Interface(Amazon CLI)。
- 在外部账户中创建指向表的资源链接。请参阅 the section called “创建资源链接” (p. 126)。

管理受管表

受管控的表格在中提供了多种高级功能Amazon Lake Formation，包括对 ACID（原子、一致、隔离和持久）事务的支持、自动数据压缩和时间旅行查询。

主题

Lake Formation 中的受管桌子

中的元数据表Amazon Glue Data Catalog存储有关数据源和目标的信息，包括架构信息、分区信息、数据位置等。

数据目录支持两种类型的元数据表：受管的表和不受管控的表。受管控的表格是独一无二的Amazon Lake Formation. 创建表时，您可以指定表是否受管表。

受管控的表格提供以下高级功能：

ACID 事务

ACID（原子、一致、隔离和持久）事务保护创建或更新表等数据目录操作的完整性。它们还允许多个用户同时可靠地在 Amazon S3 数据湖中添加和删除对象，同时允许其他用户在相同的数据集上同时运行分析查询和机器学习 (ML) 模型，这些数据集返回一致的和 up-to-date 结果。当受管控表参与对 Amazon S3 上数据湖的读取或写入时，这些操作发生在事务中。

事务保护受管理的表元数据的完整性，包括表现— 用于在表的基础数据中定义 Amazon S3 对象的元数据。集成Amazon服务，例如Amazon Athena支持受管控的表以在查询中提供一致的读取。要在您的账户中使用交易Amazon GlueETL 任务，在对数据湖执行任何读取或写入操作之前先开始事务，并在事务完成后提交事务。

有关事务的更多信息，请参阅 在事务中读取和写入数据湖 (p. 222)。

自动数据压缩

为了提高所管表中的小型 Amazon S3 对象的性能，Lake Formation 自动将受管表中的小型 Amazon S3 对象压缩为更大的对象。

默认情况下，对受管辖的表启用压缩。您可以对单个受管控的表禁用压缩。有关更多信息，请参阅 受管表的存储优化 (p. 118)。

时间旅行查询

如前所述，每个受管表都维护所包含 Amazon S3 对象的版本控制清单。清单的早期版本可用于时间旅行查询。您对 Athena 和 Athena 中受管控表的查询Amazon GlueETL 任务可以包含时间戳，表示您要发现数据在特定日期和时间的状态。

要在 Athena 中提交时空旅行查询，请使用语法`FOR SYSTEM_TIME AS OF` *`timestamp`*要么`FOR SYSTEM_VERSION AS OF` *`version`*.

```
SELECT *
FROM cloudtraildb.cloudtraildata
FOR SYSTEM_TIME AS OF TIMESTAMP '2021-09-30 10:00:00'
```

有关 Athena 对受管表进行时空旅行查询的更多示例，请参阅查询受管表中的Amazon Athena 用户指南.

在 ETL 作业脚本中，要使用时空旅行将数据读入动态框架，请包含与以下内容类似的代码。

Python

```
dynamic_frame = glueContext.create_dynamic_frame_from_catalog(database =
 'cloudtraildb, table_name = 'cloudtraildata',
   additional_options = {"asOfTime": "2021-09-30 10:00:00"})
```

Scala

```
val persons: DynamicFrame = glueContext.getCatalogSource(database = "cloudtraildb",
 tableName = "cloudtraildata",
 additional_options = JsonOptions("""{"asOfTime": "2021-09-30
 10:00:00"}""") ).getDynamicFrame()
```

Note

Lake Formation 权限不受版本控制。时空旅行查询始终遵守当前权限。例如，如果时间 T1 的权限限制了对表列的访问权限，而当前权限（时间 T2）授予了对所有列的访问权限，则针对时间 T1 的数据进行的时间旅行查询将返回所有列。

对受管表和行筛选器进行性能优化

受管控的表和表具有行级和单元级安全性，使用访问数据Lake Formation 存储 (p. 310). 存储 API 一致地强制执行权限并返回一致的数据视图。您应注意以下注意事项。

- 存储 API 允许引擎将谓词推送到 Lake Formation，以优化数据的扫描方式。这些谓词可以减少 Lake Formation 和查询引擎之间需要读取和传输的数据量，从而提高查询的性能。
- 不支持的谓词在没有优化的情况下传递。结果返回到查询引擎，可以在其中执行进一步的优化。不支持以下谓词：
  - `LIKE`运算符 — 例如，`SELECT column1 FROM table1 where column2 LIKE '%some_value %'`
  - `OR`对同一个表中的两个不同列执行运算符-例如，`SELECT column1 FROM table1 where column2 = 'value1' OR column2 = 'value2'`
  - `LIMIT`—`SELECT column1 from table1 LIMIT 100`
- 目前不支持向下推聚合。

受管表的先决条件

以下是受管表的先决条件：

- 仅支持 Amazon S中的数据和。Amazon S3 的位置必须向 Lake Formation 注册。有关注册位置的信息将 Amazon S3 位置添加到您的数据湖 (p. 102).
- 默认情况下，受管控的表会启用自动数据压缩。有关自动压缩的更多信息。Automatic data compaction (p. 114).
- 校长们称之为交易 API 操作 (p. 296)和受管控的表对象 API 操作 (p. 303)必须附加了以下 IAM 策略。

  在以下策略中，第一个权限块创建和管理事务。第二个模块允许从受管控的表中读取数据。第三个区块允许委托人与受管控表的清单进行交互。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartTransaction",
                "lakeformation:CommitTransaction",
                "lakeformation:CancelTransaction",
                "lakeformation:ExtendTransaction",
                "lakeformation:DescribeTransaction",
                "lakeformation:ListTransactions",

                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnitResults",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetQueryStatistics",

                "lakeformation:GetTableObjects",
                "lakeformation:UpdateTableObjects",
                "lakeformation:DeleteObjectsOnCancel"

            ],
            "Resource": "*"
        }
    ]
}
```

- 如果您计划在 VPC 中运行的任务中使用事务读取或写入受管控表，则必须先配置 Lake Formation VPC 终端节点。有关 VPC 端点的更多信息Amazon Lake Formation 和接口 VPC 终端节点 (Amazon PrivateLink) (p. 231).

有关 IAM 权限和策略的更多信息Lake Formation 权限参考 (p. 167).

创建受管表

受管控的表格在中提供了多种高级功能Amazon Lake Formation，包括对 ACID（原子、一致、隔离和持久）事务的支持、自动数据压缩和时间旅行查询。受管表仅支持存储在 Amazon S3 中的数据。

您可以使用创建受管控的表Amazon Lake Formation控制台Amazon GlueAPI，或Amazon Command Line Interface(Amazon CLI)。

有关受管表的更多信息the section called “Lake Formation 中的受管桌子” (p. 114).

Console

1. 确保满足创建受管表的所有先决条件。有关更多信息，请参阅 the section called “受管表的先决条件” (p. 115)。
2. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

以数据湖管理员或拥有 Lake Formation 的用户身份登录`CREATE_TABLE`目标数据库的权限以及谁拥有`glue:CreateTable` Amazon Identity and Access Management(IAM) 权限。

3. 在导航窗格的 Data catalog (数据目录) 中，请选择 Tables (表)。然后选择 。创建表.
4. 填写创建表表单，并指定以下内容：

   - UNDER数据管理和安全，启用启用受管控的数据访问和管理.
   - UNDER数据存储，指定一个向 Lake Formation 注册的 Amazon S3 位置。

Amazon Glue CLI/SDK

1. 确保满足创建受管表的所有先决条件。有关更多信息，请参阅 the section called “受管表的先决条件” (p. 115)。
2. 在`TableInput`你提供给的结构`CreateTable`操作`TableType=GOVERNED`，然后指定一个向 Lake Formation 注册的 Amazon S3 位置。

   有关更多信息，请参阅TableInput 结构中的Amazon Glue开发人员指南.

以下是 Python 中的一个示例：

```
aws glue create-table --database-name default --table-input '{
   "Name":"product_table",
   "Description":"product table for manual entries.",
   "StorageDescriptor":{
      "Columns":[
         { "Name":"product_id", "Type":"string" },
         { "Name":"product_name", "Type":"string" }
      ],
      "Location":"s3://my_bucket_name/dbs/product_table/",
      "InputFormat":"org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat",
      "OutputFormat":"org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat",
      "Compressed":false,
      "NumberOfBuckets":0,
      "SerdeInfo":{

 "SerializationLibrary":"org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe",
         "Parameters":{
         }
      },
      "SortColumns":[
      ],
      "StoredAsSubDirectories":false
   },
   "TableType":"GOVERNED",
   "Parameters":{
      "classification":"parquet",
      "lakeformation.aso.status":"true"
   }
}'
```

Athena

有关使用 Athena 创建受管表的说明，请参阅开始使用中的Amazon Athena 用户指南.

Following is an example:

```
CREATE TABLE productsdb.product_table(
product_id string,
product_name string)
  STORED AS PARQUET
  LOCATION 's3://DOC-EXAMPLE-BUCKET/governed-folder'
  TBLPROPERTIES (
  'table_type'='LAKEFORMATION_GOVERNED',
  'classification'='parquet'
)
```

在 Lake Formation 中阅读和写受管制的表

受管理的表格有表现在 Lake Formation 中跟踪构成表数据的 Amazon S3 对象。当您向数据添加新对象时，Lake Formation 会更新清单。清单通过以下方式参与受管理的表读取和写入操作：

- 写入 — 当您将表数据添加为新的 Amazon S3 对象时，您可以调用`UpdateTableObjects`API 操作将这些对象添加到受管理的表清单中。您通常在交易中进行一个或多个这些调用，如果出现故障，可以回退这些调用。调用`UpdateTableObjects`API，主要需要`INSERT`和`DELETE`受管理表上的 Lake Formation 权限。
- 读取 — 您可以使用 Lake Formation 的查询受管理表的数据查询 API(`StartPlanQuery`，例如）。查询 API 强制执行行级和列级权限。要调用 API，委托人需要`SELECT`受管理表上的 Lake Formation 权限。您可以选择传递交易 ID 或`QueryAsOfTime`从某个时间起检索数据的时间戳。

在 Amazon S3 中创建清单中的对象并将其添加到受管理的表中之后，它应该不被修改。尽管 Amazon S3 允许多次写入对象，但 Lake Formation 假设数据湖中的对象已写入一次。

有关针对受管理表的事务的详细信息，包括演示使用的示例代码`UpdateTableObjects`和`GetTableObjects`请参阅在事务中读取和写入数据湖 (p. 222)和对象 API.

受监管的表的 CloudTrail 数据事件

当你的应用程序调用`GetTableObjects`、`UpdateTableObjects`，或者`DeleteObjectsOnCancel`，以下权限Amazon CloudTrail将生成事件。

- 对应于的数据事件`GetTableObjects`对于资源链接。
- 对应于的管理事件`GetTable`对于资源链接。
- 对应于的管理事件`GetTable`对于受管理的表格。

默认情况下不会启用 CloudTrail 数据事件。有关 CloudTrail 中数据事件的更多信息，请参阅记录跟踪记录的数据事件中的Amazon CloudTrail用户指南.

如果您是否已启用 CloudTrail 数据事件并使用 Lake Formation 拥有共享资源资源链接，您可能看不到 CloudTrail 中的预期事件。对于跨账户调用，无论是在使用资源链接时，还是在 AP 调用中指定目录 ID 时，Lake Formation 都会发出 CloudTrail 事件仅在呼叫者方面进行。资源所有者没有获得事件的副本。

受管表的存储优化

默认情况下，创建受管控表时所有存储优化功能均处于启用状态。其中包括数据压缩和垃圾收集。

压缩数据

受管表的一个重要用例是流式传输数据或其他应用程序，在这些应用程序中，小块数据会持续进入 Amazon S3 数据湖。一个表可能会增长到数千个 Amazon S3 对象。每个受管控的表都保留一个表现它标识表中包含的所有 Amazon S3 对象。该清单经过版本控制和自动更新，以确保您始终看到一致的表视图。

Note

数据压缩优化器会持续监控您的表分区，并在超过文件数量和文件大小的阈值时启动。Lake Formation 在不干扰并发查询的情况下进行压缩。目前，仅对 Parquet 格式的分区表支持压缩。

垃圾回收

受管表的另一项存储优化功能通过删除不再属于受管表的 Amazon S3 对象来帮助降低存储成本。如果在将对象添加到受管表的清单时取消事务，则不会自动清理这些对象。这是为了允许在无需重新生成数据的情况下重试交易。在某些情况下，最好从已取消的事务中移除对象。

要使用此功能`DeleteObjectsOnCancel`在调用 S3 之前`PutObject`. 这告诉 Lake Formation 异步删除这些文件以帮助节省成本。调用`DeleteObjectsOnCancel`提供在 Amazon S3 中止时从 Amazon S3 中删除对象的授权。此功能无法手动禁用。有关中止事务和移除不需要的对象的更多信息，请参阅回滚Amazon S3写入 (p. 223).

Note

数据压缩仅适用于 Parquet 分区表。

使用存储优化的先决条件

必须先完成中所述的设置说明，然后才能使用数据压缩为使用受管控表的自动数据压缩做好准备 (p. 19).

为受管控的表禁用和重新启用数据压缩

为了提高 ETL 任务和分析服务的性能，Lake Formation 会自动将受管表中的小型 Amazon S3 数据元压缩为更大的对象。默认情况下，对受管控的表启用数据压缩。您可以对单个受管控的表禁用压缩，稍后再重新启用压缩。

可以使用 Lake Formation 控制台或 Lake Formation 控制台启用和禁用数据压缩Amazon CLI.

Console

为受管控的表禁用或重新启用数据压缩

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、表创建者或被授予以下权限的用户身份登录`glue:UpdateTable`许可和Lake Formation`ALTER`桌上的权限。
2. 在导航窗格中，选择表。
3. 选择表名旁边的选项按钮操作菜单中编辑.

   Note

   确保选择受管控的表。受管表的第一个版本Enabled (已启用)中的治理column.
4. 在存储库的编辑表页面上，执行以下操作之一

   - UNDER数据管理和安全，请选择或清除自动压缩选项。
5. 选择Save（保存）。

Amazon CLI

例如，要禁用压缩功能Amazon CLI命令。

```
aws update-table-storage-optimizer --database-name database-name --table-name table-name
```

```
--storage-optimizer-config '{"compaction" :{"is_enabled": "false"}}')
```

要为表重新启用数据压缩，请使用相似的代码，但将值设置为`is_enabled`到`true`.

检查受管理的表压缩状态

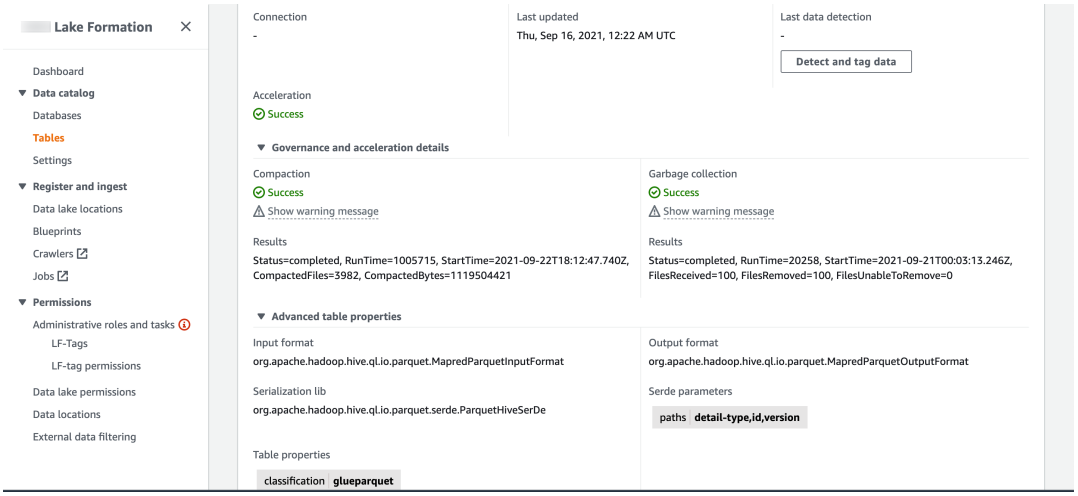
对于受管控的表，您可以查看已取消事务优化器的数据压缩状态和删除对象，方法是在控制台中查看该表，或者运行Amazon CLI命令。

Console

检查受管表的压缩状态

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、表创建者或被授予以下权限的用户身份登录`glue:GetTable`许可和桌面上的任何 Lake Formation 许可。
2. 在导航窗格中，选择表。
3. 在存储库的表页面上。
4. UNDER表详细信息，向下滚动到治理和加速详细信息部分。



Amazon CLI

使用与以下类似的命令查看与特定表关联的所有加速的配置和上次运行状态。

```
aws list-table-storage-optimizers --database-name database-name --table-name table-name
```

以下是此命令的响应示例。

```
[
    {
        StorageOptimizerType: "compaction",
        config: {
            "state": "enabled"
        },
        errorMessage: "",
        lastRunDetails: "lastRunTime: December 14, 2021. Compacted 1000 objects"
```

```
    },
    {
        StorageOptimizerType: "garbage_collection",
        config: {
           "state": "disabled"
        },
        errorMessage: "IAM role is missing DeleteObject permissions",
        lastRunDetails: "lastRunTime: December 14, 2021. Collected 1000 objects"
    }
]
```

受管表的注释和限制

请记住以下受管表的注释和限制：

- 目前仅Amazon Athena、Amazon Redshift SpectrumAmazon GlueETL 脚本支持查询受管控的表。Athena查询仅限于只读。
- 有关从 Amazon Redshift Spectrum 查询Lake Formation 表的信息，请参阅将 Redshift 与结合使用 Amazon Lake Formation中的Amazon Redshift 开发人员指南。
- 对于静态加密的数据，受管控表照常工作，其中Amazon Glue管理加密密钥。与受管表所在的 Amazon S3 位置关联的 IAM 角色需要具有Amazon KMS权限。
- 启用数据目录元数据加密后，受管控的表可以照常运行。与受管表所在的 Amazon S3 位置关联的 IAM 角色需要具有Amazon KMS权限。此外，您需要向 IAM 角色和 Lake Formation 服务授予加密或解密密钥的权限。
- 默认的 Lake Formation SLR 角色不能用于加密的受管控表。您必须在 Amazon S3 中使用自定义 IAM 角色Amazon KMS和 CloudWatch 策略。
- 使用以下命令创建受管控表Amazon Web Services Management Console，你必须使用 Lake Formation控制台。您不能使用Amazon Glue控制台。
- 仅支持包含 Parquet 格式文件的分区表进行数据压缩。
- 您无法将现有的非管控表转换为受管控表，也不能将现有受管控表转换为不受管控的表。
- Amazon Glue搜寻器不支持受管控的表。
- 你不能使用 Apache Spark DataFrames 从受管表中读取以及向其写入数据。
- 中不支持向下推谓词Amazon GlueETL。
- 如果您在 30 分钟内主动写入超过 250 个分区，则数据压缩所需的时间可能比平时长。
- 使用动态框架读取受管控表时，不支持以下功能Amazon GlueETL
  - 作业书签
  - 限期执行
  - 下推谓词
  - 服务器端目录分区谓词
  - enableUpdateCatalog
- 以下Amazon Glue不允许在受管控的表上进行 API 操作：
  - `CreatePartition`
  - `BatchCreatePartition`
  - `UpdatePartition`
  - `BatchUpdatePartitions`
  - `DeletePartition`
  - `BatchDeletePartition`
  - `GetPartition`
  - `BatchGetPartition`

这些限制的原因是，必须使用支持事务的 API 操作对受管表执行分区操作。有关更多信息，请参阅受管控的表格对象 API (p. 303).

此外，还有一些限制`UpdateTable`API 操作。您无法更新表类型、更改分区键或更改表位置。

- 在 Amazon S3 中创建受管表清单中的对象并将其添加到受管控表后，它应该不被修改。尽管 Amazon S3 允许多次写入一个对象，但 Lake Formation 假设数据湖中的对象只写入一次。
- Amazon S3 对象一次只能添加到一个受管控的表中。强烈建议不将多个活动事务中的同一 Amazon S3 对象添加到多个受管控的表中。

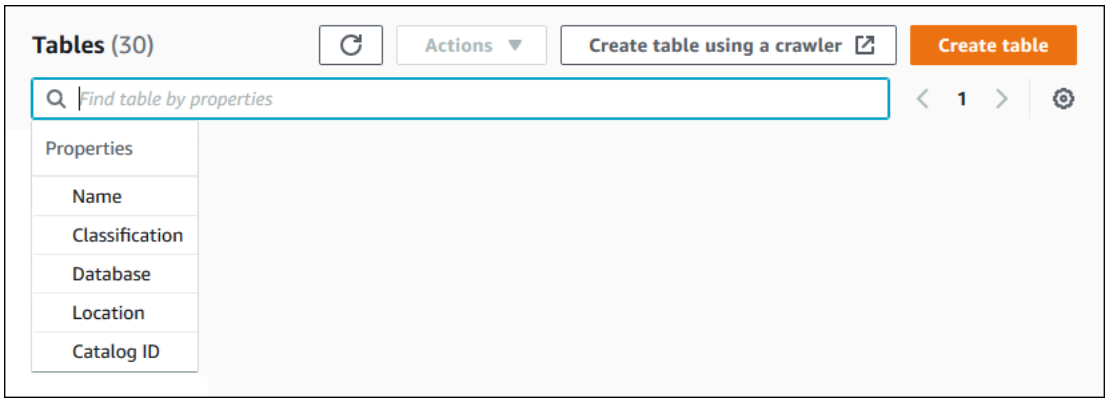
搜索表

您可以使用Amazon Lake Formation控制台可按名称、位置、包含数据库等搜索数据目录表。搜索结果只显示您对 Lake Formation 拥有权限的表格。

要搜索表（控制台）

1. 登录到Amazon Web Services Management Console然后在处打开 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。该字段包含占位符文本按属性查找表.

   这些区域有：属性菜单显示了要搜索的各种表格属性。

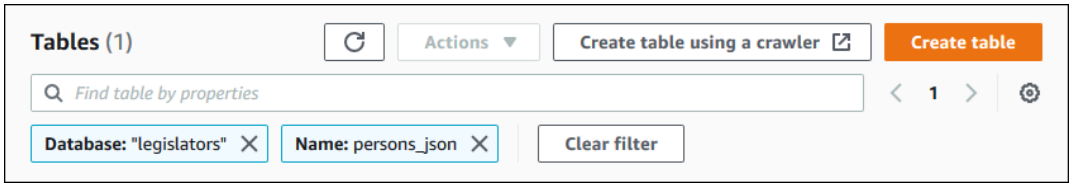


4. 请执行下列操作之一：

   - 通过包含数据库来搜索。
     1. 选择数据库来自 的属性菜单，然后从数据库出现的菜单或键入数据库名称，然后按Enter.

        将列出您在数据库中拥有权限的表。
     2. （可选）要将列表缩小为数据库中的单个表，请再次将光标置于搜索字段中，选择名称来自 的属性菜单，然后从表出现的菜单或键入表名称然后按Enter.

        此时将列出单个表，数据库名称和表名称都以切片的形式显示在搜索字段下。



        要调整过滤器，请关闭其中一个图块或选择清除筛选.

- 按其他房产进行搜索。
  1. 从属性菜单。

     要搜索Amazon账户 ID，选择目录 ID来自 的属性菜单中，输入有效Amazon账户 ID（例如 111122223333），然后按Enter.

     要按位置搜索，请选择位置来自 的属性菜单，然后从Locations出现的菜单。返回所选位置（例如 Amazon S3）根位置的所有表格。

跨共享数据目录表和数据库Amazon账户

您可以与外部共享数据目录资源（数据库和表）Amazon通过向外部账户授予 Lake Formation 对资源的权限来实现账户。然后，用户可以在多个账户中运行加入和查询表的查询和作业。有一些限制，当您与另一个账户共享数据目录资源时，该账户中的承担者可以像资源在其数据目录中一样对该资源进行操作。

你不会与外部的特定委托人共享资源Amazon账户-您将资源与Amazon账户或组织。与共享资源时Amazon组织，您正在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委托人授予对共享资源的权限。

有关更多信息，请参阅跨账户访问：如何运作 (p. 242)和授予和撤消对数据目录资源的权限 (p. 145)。

另请参见：

- 访问和查看共享数据目录表和数据库 (p. 123)
- 跨账户访问权限前提条件 (p. 243)

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授予权限的委托人，与您共享的资源Amazon帐户在数据目录中显示，就像它们是您账户中的资源一样。控制台显示拥有资源的账户。

您可以使用 Lake Formation 控制台查看与账户共享的资源。您也可以使用Amazon Resource Access Manager(Amazon RAM) 控制台可查看与您的账户共享的资源以及与其他人共享的资源Amazon使用命名资源方法进行帐户。

Important

当有人使用指定资源方法向您的账户授予对数据目录资源的跨账户权限时，或者Amazon组织，Lake Formation 使用Amazon Resource Access Manager(Amazon RAM) 共享资源的服务。如果你的账户是相同的Amazon组织作为授权账户，您可以立即使用共享资源。
但是，如果你的账户不在同一个组织中，Amazon RAM向您的账户发送邀请以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Amazon RAM控制台或 CLI 以接受邀请。
如果存在 Lake Formation 控制台会显示警报Amazon RAM资源共享邀请等待接受。只有授权查看的用户Amazon RAM邀请会收到警报。
共享资源的基于 Lake Formation 标签的访问方法（LF-TBAC）不使用Amazon RAM. 因此，使用 LF-TBAC 方法跨账户共享的资源立即可用。

另请参见：

- 跨共享数据目录表和数据库Amazon账户 (p. 123)
- 跨账户访问：如何运作 (p. 242)

- 访问共享表的基础数据 (p. 246)
- 元数据访问控制 (p. 236) (有关指定资源方法与用于共享资源的 LF-TBAC 方法的信息。)

主题

- 接受来自的资源共享邀请Amazon RAM (p. 124)
- 查看共享数据目录表和数据库 (p. 125)

接受来自的资源共享邀请Amazon RAM

如果与您的数据目录资源共享Amazon账户和你的账户不一样Amazon组织作为共享帐户，在接受来自的资源共享邀请之前，您无权访问共享资源Amazon Resource Access Manager(Amazon RAM)。作为数据湖管理员，必须先查询Amazon RAM对于待处理的邀请，然后接受邀请。

您可以使用Amazon RAM控制台、API 或Amazon Command Line Interface(Amazon CLI) 查看和接受邀请。

查看和接受来自的资源共享邀请Amazon RAM(console)

1. 确保您具有所需的Amazon Identity and Access Management(IAM) 查看和接受资源共享邀请的权限。

   有关针对数据湖管理员的建议 IAM 策略的信息，请参阅the section called “数据湖管理员权限” (p. 322).
2. 按照中的说明进行操作接受和拒绝邀请中的Amazon RAM用户指南.

查看和接受来自的资源共享邀请Amazon RAM(AmazonCLI)

1. 确保您具有所需的Amazon Identity and Access Management(IAM) 查看和接受资源共享邀请的权限。

   有关针对数据湖管理员的建议 IAM 策略的信息，请参阅the section called “数据湖管理员权限” (p. 322).
2. 输入以下命令以查看待处理的资源共享邀请。

   ```
   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"
           }
       ]
   }
   ```

   记下的状态`PENDING`.
3. 复制的值`resourceShareInvitationArn`键到剪贴板中。
4. 将值粘贴到以下命令中，替换*<invitation-arn>*，然后输入命令。

```
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"
        }
    ]
}
```

记下的状态ACCEPTED.

查看共享数据目录表和数据库

您可以使用 Lake Formation 控制台查看与您的帐户共享的资源，或者AmazonCLI。您也可以使用Amazon Resource Access Manager(Amazon RAM) 控制台或 CLI 来查看与您的账户共享的资源以及与其他人共享的资源Amazon账户。

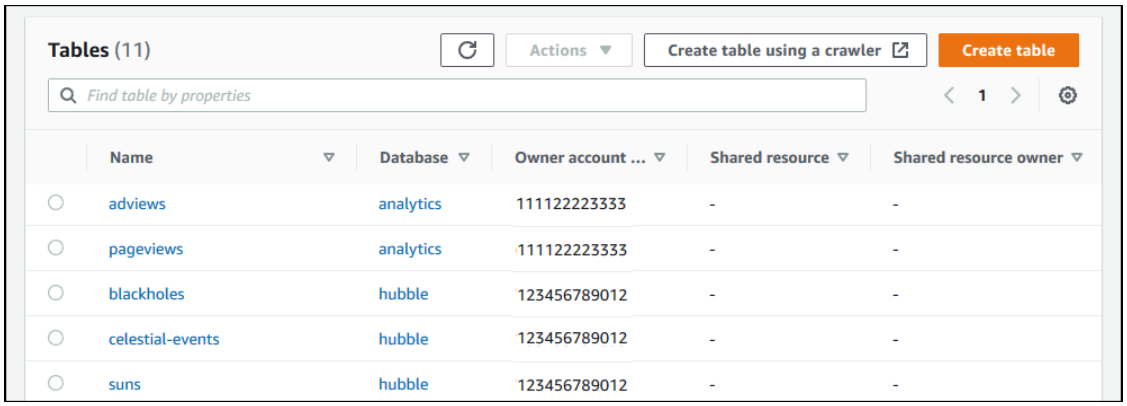
使用 Lake Formation 控制台查看共享资源

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员或已获得共享表权限的用户身份登录。
2. 查看与您共享的资源Amazon请执行以下操作之一：

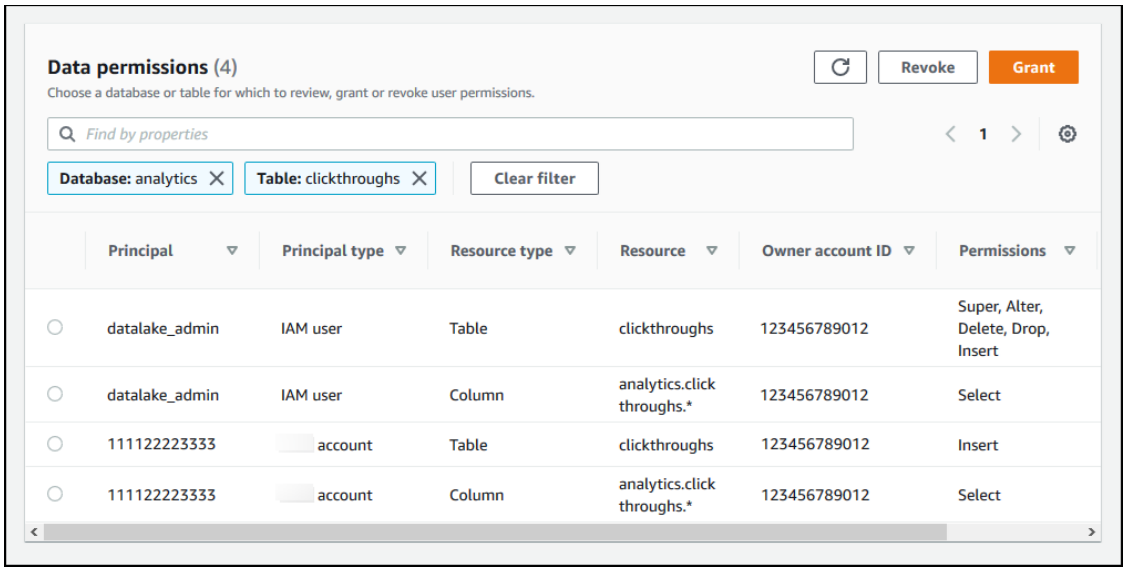
   - 要查看与您的账户共享的表，请在导航窗格中选择。表.
   - 要查看与您的帐户共享的数据库，请在导航窗格中选择。数据库.

   控制台将显示您账户中的数据库或表列表以及与您的账户共享的数据库或表。对于与您的账户共享的资源，控制台会显示所有者的Amazon下面的账户 ID拥有者账户 ID列（以下屏幕截图中的第三列）。



3. 查看与其他人共享的资源Amazon在导航窗格中，选择帐户或组织。数据权限.

   您共享的资源列在数据权限页面中显示了外部账号委托人列，如下图所示。



使用 查看共享资源Amazon RAM控制台

1. 确保您具有所需的Amazon Identity and Access Management使用查看共享资源的 (IAM) 权限Amazon RAM.

   至少，您必须具有`ram:ListResources`权限。此权限包含在 AWS 托管策略中。`AWSLakeFormationCrossAccountManager`.
2. 登录到Amazon Web Services Management Console然后打开Amazon RAM控制台在https://console.aws.amazon.com/ram.
3. 请执行下列操作之一：

   - 要查看共享的资源，请在导航窗格中的下由我共享，选择共享资源.
   - 要查看与您共享的资源，请在导航窗格中的下与我共享，选择共享资源.

创建资源链接

资源链接是数据目录对象，它们是指向元数据库和表的链接，通常是指向共享数据库和其他表的链接Amazon账户。它们有助于实现跨账户访问数据湖中的数据。

主题

资源链接在 Lake Formation 中的工作原理

一个资源链接是指向本地或共享数据库或表的数据库或表的数据库或表的数据库目录对象。创建到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。与您拥有的表或与您共

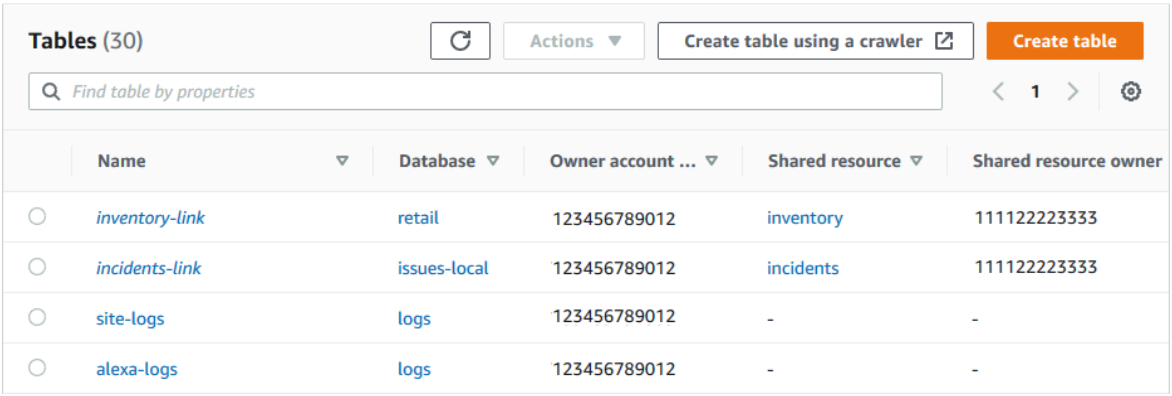
享的表一起，表资源链接由`glue:GetTables()`并作为条目显示在表Lake Formation 控制台的页面。数据库的资源链接的行为方式类似。

通过创建指向数据库或表的资源链接，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，这将特别有用Amazon帐户共享具有相同名称的数据库或表，或者如果您的账户中的多个数据库具有相同名称的表。
- 使用集成Amazon例如服务Amazon Athena和 Amazon Redshift Spectrum 来运行访问共享数据库或表的查询。某些集成服务无法跨账户直接访问数据库或表。但是，他们可以访问您账户中的资源链接到其他账户中的数据库和表。

> **Note**
>
> 在中，您无需创建资源链接即可在中引用共享数据库或表。Amazon Glue提取、转换和加载 (ETL) 脚本。但是，为了避免在多个时出现模糊Amazon帐户共享具有相同名称的数据库或表，您可以在调用 ETL 操作时创建和使用资源链接或指定目录 ID。

以下示例显示了 Lake Formation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。每个资源链接随其链接的共享资源的名称和所有者一起显示。在此示例中，中的数据湖管理员Amazon账户1111-2222-3333`inventory`和`incidents`表拥有 1234-5678-9012。然后，该账户中的用户创建了指向这些共享表的资源链接。



以下是资源链接的注释和限制：

- 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底层数据。这些集成服务中的查询是根据资源链接名称构建的。
- 假设设置仅对此数据库中的新表使用 IAM 访问控制已关闭包含的数据库，只有创建资源链接的委托人才能查看和访问该资源链接。要使您账户中的其他委托人能够访问资源链接，请授予`DESCRIBE`允许它。要使其他人能够删除资源链接，请授予`DROP`允许它。数据湖管理员可以访问账户中的所有资源链接。要删除另一个委托人创建的资源链接，数据湖管理员必须首先授予自己`DROP`资源链接的权限。有关更多信息，请参阅 Lake Formation 权限参考 (p. 167)。

  > **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。您必须对目标单独授予权限。

- 要创建资源链接，你需要 Lake Formation`CREATE_TABLE`要么`CREATE_DATABASE`许可，以及`glue:CreateTable`要么`glue:CreateDatabase` Amazon Identity and Access Management(IAM) 权限。
- 您可以创建指向本地（拥有）数据目录资源的资源链接，以及与您的Amazonaccount.
- 创建资源链接时，不会执行任何检查来查看目标共享资源是否存在，或者您是否对该资源具有跨账户权限。这使您可以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。如果删除共享资源，则不会删除指向该资源的资源链接。
- 可以创建资源链接链。但是，这样做没有价值，因为 API 只遵循第一个资源链接。

另请参见：

- 授予和撤消对数据目录资源的权限 (p. 145)

创建指向共享数据目录表的资源链接

您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

创建指向共享表的资源链接（控制台）

1. 打开Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/. 以拥有 Lake Formation 的校长身份登录`CREATE_TABLE`数据库中包含资源链接的权限。
2. 在导航窗格中，选择 Tables (表)，然后选择 Create table (创建表)。
3. 在存储库的创建表页面，选择资源链接，然后提供以下信息：

   资源链接

   输入与表名遵守相同规则的名称。名称可以与目标共享表相同。

   数据库。

   本地数据目录中要包含资源链接的数据库。

   共享表

   从列表中选择共享表，或输入本地（拥有）或共享表名称。

   该列表包含与您的账户共享的所有表格。请注意每个表中列出的数据库和所有者帐户 ID。如果您未看到已与账户共享的表，请检查以下内容：

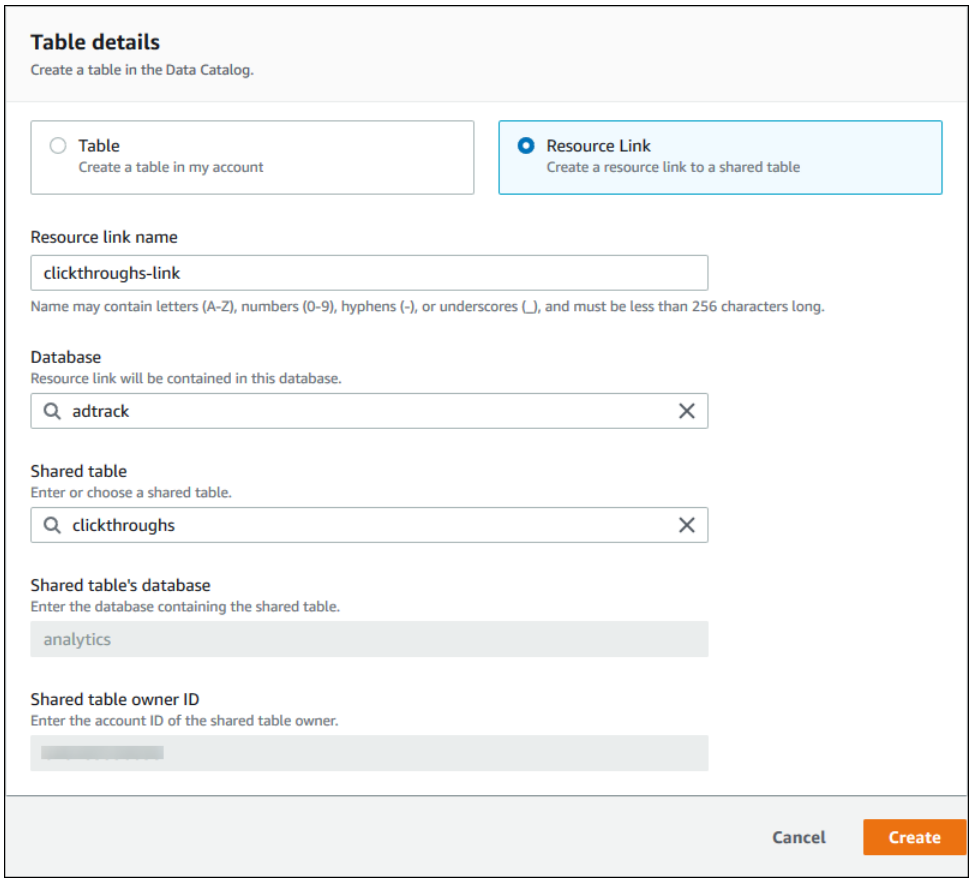
   - 如果您不是数据湖管理员，请检查数据湖管理员是否授予了您对表格的 Lake Formation 权限。
   - 如果你是数据湖管理员，并且你的帐户不同Amazon组织作为授权账户，请确保您已接受Amazon Resource Access Manager(Amazon RAM) 表的资源共享邀请。有关更多信息，请参阅 接受来自的资源共享邀请Amazon RAM (p. 124)。

   共享表的数据库

   如果从列表中选择了共享表，则此字段将使用外部帐户中的共享表的数据库填充。否则，请在外部帐户中输入本地数据库（用于指向本地表的资源链接）或共享表的数据库。

   共享表所有者

   如果从列表中选择了共享表，则此字段将使用共享表的所有者帐户 ID 填充。否则，请输入Amazon账户 ID（用于指向本地表的资源链接）或Amazon共享该表的账户。



4. 选择Create以创建资源链接。

   然后，您可以在名称 column on the 表页.

5. （可选）授予 Lake FormationDESCRIBE授予必须能够查看链接并通过链接访问链接目标的承担者的权限。

创建指向共享表的资源链接 (Amazon CLI)

1. 输入类似以下的命令。

```
aws glue create-table --database-name myissues --
table-input '{"Name":"mycustomers","TargetTable":
{"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
```

   完成后，此命令将创建一个名为mycustomers转到共享表customers，它在数据库中issues中的Amazon账户 1111-2222-3333。资源链接存储在本地数据库中myissues.

2. （可选）授予 Lake FormationDESCRIBE授予必须能够查看链接并通过链接访问链接目标的承担者的权限。

   另请参见：

   - 资源链接在 Lake Formation 中的工作原理 (p. 126)
   - DESCRIBE (p. 174)

创建指向共享数据目录数据库的资源链接

您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

创建指向共享数据库的资源链接（控制台）

1. 打开Amazon Lake Formation控制台https://console.aws.amazon.com/lakeformation/. 以数据湖管理员或数据库创建者身份登录。

   数据库创建者是被授予 Lake Formation 的委托人CREATE_DATABASE权限。
2. 在导航窗格中，选择数据库，然后选择创建数据库.
3. 在存储库的创建数据库页面，选择资源链接，然后提供以下信息：

   资源链接

   输入与数据库名称遵守相同规则的名称。名称可以与目标共享数据库相同。

   共享数据库

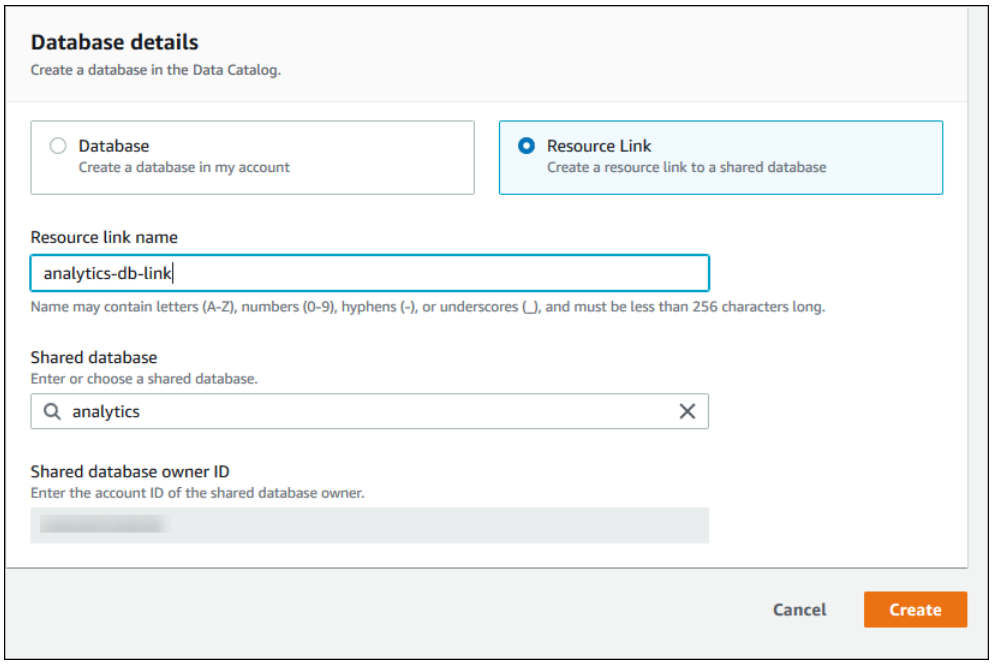
   从列表中选择数据库，或输入本地（拥有）或共享数据库名称。

   该列表包含与您的账户共享的所有数据库。请注意每个数据库中列出的所有者账户 ID。如果您未看到已与账户共享的数据库，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授予了您对数据库的 Lake Formation 权限。
   - 如果你是数据湖管理员，并且你的帐户不同Amazon组织作为授权账户，请确保您已接受Amazon Resource Access Manager(Amazon RAM) 数据库的资源共享邀请。有关更多信息，请参阅 接受来自的资源共享邀请Amazon RAM (p. 124)。

   共享数据库所有

   如果从列表中选择了共享数据库，则此字段将使用共享数据库的所有者帐户 ID 填充。否则，请输入Amazon账户 ID（用于指向本地数据库的资源链接）或Amazon共享数据库的帐户。



4. 选择Create以创建资源链接。

然后，您可以在名称 column on the 数据库页.

5. （可选）授予 Lake FormationDESCRIBE授予必须能够查看链接并通过链接访问链接目标的承担者的权限。

创建指向共享数据库的资源链接（Amazon CLI）

1. 输入类似以下的命令。

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

完成后，此命令将创建一个名为myissues转到共享数据库issues，在Amazon账户1111-2222-3333。

2. （可选）授予 Lake FormationDESCRIBE授予必须能够查看链接并通过链接访问链接目标的承担者的权限。

另请参见：

- 资源链接在 Lake Formation 中的工作原理 (p. 126)
- DESCRIBE (p. 174)

中的资源链接处理Amazon GlueAPI

下表说明了如何Amazon Glue数据目录 API 处理数据库和表资源链接。对于所有Get*API 操作，只返回调用者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，必须同时拥有两者Amazon Identity and Access Management(IAM) 和 Lake Formation 对目标和资源链接的权限。资源链接所需的 Lake Formation 许可是DESCRIBE. 有关更多信息，请参阅 DESCRIBE (p. 174)。

数据库 API 操作

API 操作	处理资源链接
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，请跟随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，则您必须将其删除，然后创建新的链接。
DeleteDatabase	删除资源链接。它不会删除链接的（目标）数据库。
GetDatabase	如果调用者对目标具有权限，请点击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，操作跟随链接以获取链接目标的属性。您必须指定ResourceShareType=ALL以查看与您的账户共享的数据库。

表 API 操作

API 操作	处理资源链接
CreateTable	如果数据库是资源链接，请跟随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	如果表或指定的数据库是资源链接，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，请跟随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，则操作将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，请跟随数据库链接并从目标数据库返回表或表资源链接。否则，如果表是资源链接，则操作将跟随链接并返回目标表属性。
GetTables	如果指定的数据库是资源链接，请跟随数据库链接并从目标数据库返回表和表资源链接。如果目标数据库是另一个数据库的共享数据库Amazon账户，该操作只返回该数据库中的共享表。它不跟随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（拥有）数据库，则操作将返回本地数据库中的所有表，并跟随每个表资源链接返回目标表属性。
SearchTables	返回表格和表资源链接。它不跟随链接返回目标表属性。您必须指定ResourceShareType=ALL查看与您的账户共享的表格。
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	处理资源链接
CreatePartition	如果指定的数据库是资源链接，请跟随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，则操作将跟随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，请跟随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，则操作将跟随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更新分区。
DeletePartition	如果指定的数据库是资源链接，请跟随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，则操作将跟随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删除分区。
BatchDeletePartition	与 DeletePartition 相同。

API 操作	处理资源链接
`GetPartition`	如果指定的数据库是资源链接，请跟随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，则操作将跟随链接并返回分区信息。如果表和数据库都是资源链接，则返回一个空的结果集。
`GetPartitions`	如果指定的数据库是资源链接，请跟随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，则操作将跟随链接并返回分区信息。如果表和数据库都是资源链接，则返回一个空的结果集。
`BatchGetPartition`	与 `GetPartition` 相同。

自定义的函数 API 操作

API 操作	处理资源链接
（所有 API 操作）	如果数据库是资源链接，请跟随资源链接并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 (p. 126)

使用 Lake Formation 中的工作流导入数据

与Amazon Lake Formation，你可以使用导入你的数据工作流. 工作流定义了将数据导入数据湖的数据源和计划。它是一个容器Amazon Glue搜寻器、作业和触发器，用于编排加载和更新数据湖的进程。

主题

Lake Formation 中的蓝图和工作流程

工作流封装了复杂的多任务提取、转换和加载 (ETL) 活动。工作流程生成Amazon Glue搜寻器、作业和触发器来协调数据的加载和更新。Lake Formation 将工作流作为单个实体执行和跟踪。您可以将工作流程配置为根据需要或计划运行。

您在 Lake Formation 中创建的工作流在Amazon Glue控制台作为有向无环图 (DAG) (DAG)。每个 DAG 节点都是一个作业、爬网程序或触发器。要监视进度并进行故障排除，您可跟踪工作流中各节点的状态。

当 Lake Formation 工作流完成后，运行该工作流的用户将被授予Lake FormationSELECT对工作流创建的数据目录表的权限。

您也可以在中创建工作流程Amazon Glue. 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— 从 JDBC 源将所有表中的数据加载或重新加载到数据湖中。您可以根据排除模式从源中排除某些数据。
- 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。在 JDBC 源数据库中指定要包含的各个表。对于每个表，您可以选择书签列和书签排序顺序，以跟踪以前加载的数据。首次针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，前提是您将数据源中的每个表指定为参数。
- 日志文件— 从日志文件源批量加载数据，包括Amazon CloudTrail、Elastic Load Balancing and Load LApplication Load Balancer anand B

使用下表帮助决定是使用数据库快照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...
• 模式演变是灵活多变的。（重新命名列，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。	• 架构演变是渐进的。（只有连续添加列。） • 只添加新行；不更新之前的行。

Note

用户无法编辑 Lake Formation 创建的蓝图和工作流。

创建工作流程

在开始之前，请确保您已向角色授予所需的数据权限和数据位置权限`LakeFormationWorkflowRole`. 这样，工作流程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 为工作流创建 IAM 角色 (p. 11) 和 FormLake Formation 权限概述 (p. 138)。

从蓝图创建工作流程

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 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 (p. 322)。
2. 在导航窗格中，选择蓝图，然后选择使用蓝图.
3. 在存储库的使用蓝图页面上，选择一个磁贴以选择蓝图类型。
4. UNUNUN导入来源中，指定数据源。

   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连接。使用创建其他连接Amazon Glue控制台。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。
   - 源数据路径—输入*<database>*/*<schema>*/*<table>*要么*<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) 为的 Oracle 数据库`orcl`，输入`orcl/%`导入连接中指定的用户有权访问的表。

     Important

     此字段区分大小写。如果任何组件的大小写不匹配，工作流程将失败。

     如果您指定了 MySQL 数据库，Amazon Glue默认情况下，ETL 使用 Mysql5 JDBC 驱动程序，因此本机不支持 MySQL8。您可以编辑 ETL 作业脚本以使用`customJdbcDriverS3Path`参数，如中所述JDBC connectionType 值中的Amazon Glue开发人员指南使用支持 MySQL8 的其他 JDBC 驱动程序。

   如果要从日志文件导入，请确保您为工作流程指定的角色（“工作流程角色”）具有访问数据源所需的 IAM 权限。例如，要导入Amazon CloudTrail日志，用户必须具有`cloudtrail:DescribeTrails`和`cloudtrail:LookupEvents`查看列表的权限 CloudTrail 日志，并且工作流角色必须具有对 CloudTrail Amazon S3 中的位置。
5. 请执行下列操作之一：

   - 对于数据库快照blueprint type，可以选择通过指定一个或多个排除模式来标识要导入的数据子集。这些排除模式是 Unix 样式的`glob`模式。它们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，请参阅包含和排除模式中的Amazon Glue开发人员指南.
   - 对于增量数据库蓝图类型，请指定以下字段。为要导入的表添加一行。

     表名称

     要导入的表。必须全部为小写。

     书签密钥

     以逗号分隔的定义书签键的列名列表。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源中定义的大小写相匹配。

     Note

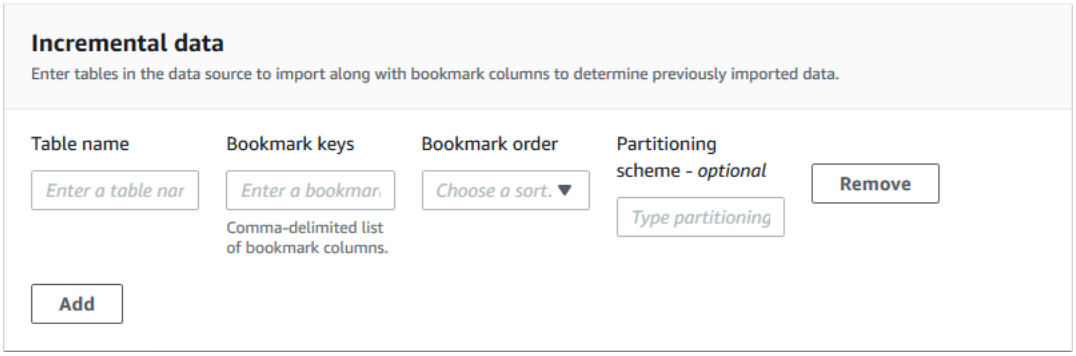
     只有在按顺序递增或递减（没有间隙）时，它才有资格作为默认书签键。如果要将主键用作书签键并且它有间隙，则必须将主键列命名为书签键。

书签订单

在选择时升序，值大于书签值的行将被标识为新行。在选择时降序，值小于书签值的行将被标识为新行。

分区方案

（可选）分区键列的列表，用斜杠 (/) 分隔。示例： `year/month/day`.



有关更多信息，请参阅 。使用作业书签来跟踪已处理的数据中的Amazon Glue开发人员指南.

6. UNUNUN导入目标中，指定目标数据库、目标 Amazon S3 位置和数据格式。

   确保工作流程角色对数据库和 Amazon S3 目标位置具有所需的Lake Formation 权限。

   Note

   目前，蓝图不支持在目标位置加密数据。

7. 选择导入频率。

   您可以指定一个`cron`用表达式Custom (自定义)选项。

8. UNUNUN导入选项：

   a. 输入工作流程名称。
   b. 对于角色，请选择角色`LakeFormationWorkflowRole`，这是你在中创建的为工作流创建 IAM 角色 (p. 11).
   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Create，然后等待控制台报告工作流已成功创建。

   Tip

   您收到以下错误消息吗？
   `User: arn:aws:iam::<account-id>:user/<username> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/<rolename>...`
   如果是这样，请检查您是否更换了`<account-id>`使用有效的Amazon所有保单中的账号。

另请参见：

- Lake Formation 中的蓝图和工作流程 (p. 134)

运行工作流程

你可以使用 Lake Formation 控制台运行工作流Amazon Glue控制台，或者Amazon Glue命令行界面 (Amazon CLI) 或 API。

运行工作流程（Lake Formation 控制台）

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 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 (p. 322)。
2. 在导航窗格中，选择 Blueprints (蓝图)。
3. 在存储库的蓝图页面上，选择工作流程。然后在操作菜单中，选择启动.
4. 工作流程运行时，请在上次运行状态column. 偶尔选择刷新按钮。

   状态来自正在运行，至发现，至Importing，至已完成.

   工作流程完成后：

   - 数据目录有新的元数据表。
   - 您的数据会被摄入到数据湖中。

   如果工作流程失败，请执行以下操作：

   a. 选择工作流。选择操作，然后选择视图图.

      工作流程将在Amazon Glue控制台。

   b. 确保已选择工作流，然后选择 History (历史记录) 选项卡。
   c. UNUNUN历史记录中，选择最近一次运行并选择查看运行详细信息.
   d. 在动态（运行时）图中选择失败的作业或 Crawler，然后查看错误消息。出现故障的节点为红色或黄色。

   另请参见：

   - Lake Formation 中的蓝图和工作流程 (p. 134)

管理Lake Formation

Lake Formation 为数据湖中的数据提供中央访问控制。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，并与Amazon Identity and Access Management对这些用户和角色进行身份验证。定义规则后，Lake Formation 将对亚马逊 Redshift Spectrum 和 Amazon Athena 的用户强制执行表和列级别的访问控制。

主题

FormLake Formation 权限概述

中主要有两种权限类型Amazon Lake Formation：

- 元数据访问权限 — 数据目录资源的权限 (数据目录权限)。

  这些权限使主体能够在数据目录中创建、读取、更新和删除元数据数据库和表。
- 基础数据访问权限 — 对中位置的权限Amazon Simple Storage Service(Amazon S3)数据访问权限和数据位置权限)。
  - 数据访问权限使委托人能够读取和写入数据隐含的Amazon S3 位置-数据目录资源所指向的数据。
  - 数据位置权限使委托人能够创建和更改指向特定 Amazon S3 位置的元数据数据库和表。

对于这两种类型，Lake Formation 使用 Lake Formation 权限和 IAM 权限的组合。IAM 权限模型由 IAM 策略组成。Lake Formation 权限模型以 DBMS 风格实现`GRANT`/`REVOKE`命令，例如：

```
Grant SELECT on tableName to userName
```

当委托人请求访问数据目录资源或基础数据时，为了使请求成功，它必须通过 IAM 和 Lake Formation 的权限检查。

Amazon Lake Formation要求授权每个委托人（用户或角色）对 Lake Formation 管理的资源执行操作。委托人由数据湖管理员或其他有权授予 Lake Formation 权限的委托人授予必要的授权。

当您向委托人授予 Lake Formation 权限时，您可以选择授予将该权限传递给其他委托人的能力。

你可以使用Lake Formation API，Amazon Command Line Interface(Amazon CLI)，或者数据权限和数据位置Lake Formation 控制台的页面，用于授予和撤销 Lake Formation 权限。

授予或撤消 Lake Formation 权限所需的 IAM 权限

所有委托人，包括数据湖管理员，都需要以下内容Amazon Identity and Access Management(IAM) 授予或撤销的权限Amazon Lake Formation使用 Lake Formation API 的数据目录权限或数据位置权限或Amazon CLI：

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable`要么`glue:GetDatabase`对于您正在授予权限的表或数据库n 使用命名资源方法

> **Note**
>
> 数据湖管理员具有隐含的 Lake Formation 权限，可以授予和撤消Lake Formation 权限。但是他们仍然需要获得 Lake Formation 授予的 IAM 权限并撤销 API 操作。

对于不是数据湖管理员且想要使用 Lake Formation 控制台授予或撤消权限的委托人，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"
            ],
            "Resource": "*"
        }
    ]
}
```

所有`glue:`和`iam:`此策略中的权限在Amazon管理的策略`AWSGlueConsoleFullAccess`.

要使用基于Lake Formation 标签的访问控制 (LF-TBAC) 授予权限，委托人需要额外的 IAM 权限。有关更多信息，请参阅 基于Lake Formation 标签的访问控制权限模型 (p. 184) 和 Lake Formation 角色和 IAM 权限参考 (p. 322)。

跨账户 权限

想要授予跨账户 Lake Formation 权限的用户通过使用命名资源方法还必须具有以下权限`AWSLakeFormationCrossAccountManager` Amazon托管策略。

数据湖管理员需要同样的权限才能授予跨账户权限，再加上Amazon Resource Access Manager(Amazon RAM) 权限以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 (p. 322)。

管理用户

具有 IAM 管理权限的委托人，例如，拥有`AdministratorAccess` Amazon托管策略-具有授予 Lake Formation 权限和创建数据湖管理员的权限。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`Deny`管理员 API 操作的语句。

```
{
```

```
    "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。

另请参阅

- Lake Formation 中的跨账户访问 (p. 242)

Lake Formation 的

Amazon 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Lake 管理员

- 具有对数据目录中所有资源的完全读取权限。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。
- 可以向任何委托人（包括自己）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。
- 可以授予其他用户创建数据库的权限。

  Note

  数据湖管理员只有在拥有 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会授予这些权限。此外，数据湖管理员没有删除他人创建的数据库或更改/删除表的隐式权限。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 (p. 12).

数据库创建者

- 对自己创建的数据库拥有所有数据库权限，对他们在数据库中创建的表具有权限，并且可以向同一数据库中的其他主体授予权限Amazon帐户在数据库中创建表的权限。数据库创建者同时拥有`AWSLakeFormationCrossAccountManager` Amazon托管策略可以将数据库的权限授予其他Amazon账户或组织。

  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没有隐式权限。

有关更多信息，请参阅 创建数据库 (p. 113)。

表格创作者

- 对他们创建的表拥有所有权限。

- 可以将他们创建的所有表的权限授予同一个中的委托人Amazonaccount.
- 可以将他们创建的所有表的权限授予其他Amazon账户或组织，如果他们有`AWSLakeFormationCrossAccountManager` Amazon托管策略。
- 可以查看包含他们创建的表的数据库。

授予数据位置权限

中的数据位置权限Amazon Lake Formation允许委托人创建和更改指向指定注册 Amazon S3 位置的数据目录资源。除了 Lake Formation 数据权限外，数据位置权限还可以用于保护数据湖中的信息。

Lake Formation 不使用Amazon Resource Access Manager(Amazon RAM) 服务以授予数据位置权限，因此您无需接受数据位置权限的资源共享邀请。

您可以使用 Lake Formation 控制台、API 或Amazon Command Line Interface(Amazon CLI)。

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 (p. 239)

主题

授予数据位置权限（同一账户）

请按照以下步骤向您的中的委托人授予数据位置权限Amazonaccount. 你可以使用 Lake Formation 控制台、API 或Amazon Command Line Interface(Amazon CLI)。

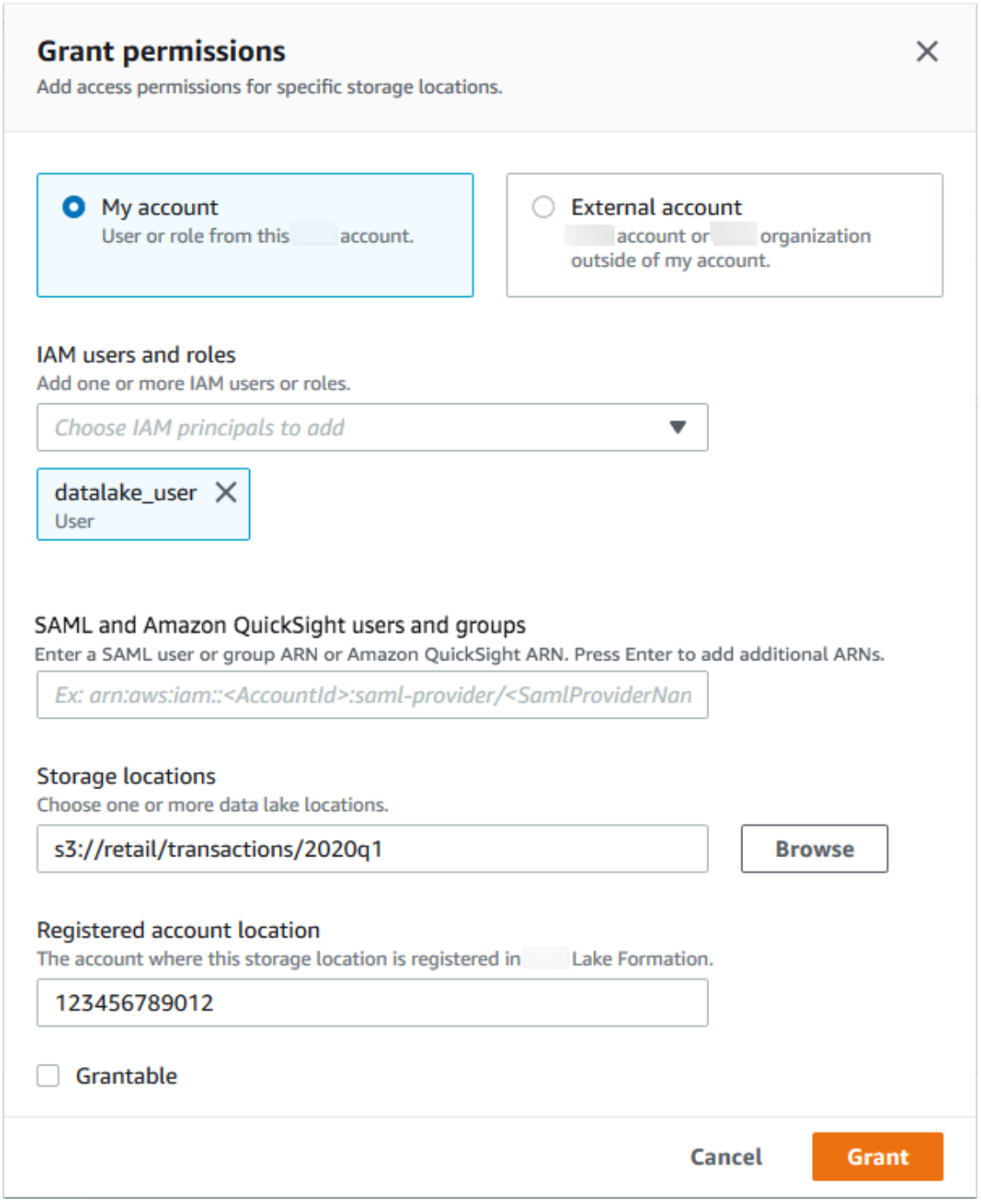
授予数据位置权限（同一账户、控制台）

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 以数据湖管理员或拥有所需数据位置授予权限的委托人身份登录。
2. 在导航窗格中，选择数据位置.
3. 选择 Grant（授权）。
4. 在授予权限对话框中，确保我的账户磁贴处于选中状态。然后提供以下信息：

   - 适用于IAM 用户和角色中，选择一个或多个承担者。
   - 适用于SAML 和Amazon QuickSight 用户和组中，为通过 SAML 联合的用户或组输入一个或多个 Amazon 资源名称 (ARN)，或者为亚马逊的 ARN 输入一个或多个亚马逊资源名称 (ARN) QuickSight 用户或组。

     一次输入一个 ARN，然后按Enter在每个 ARN 之后。有关如何构建 ARN 的信息，请参阅Lake Formation 补助金和撤销Amazon CLI命令 (p. 168).
   - 适用于存储位置，选择浏览，然后选择Amazon S3 的存储位置。该地点必须向 Lake Formation 登记。选择浏览再次添加另一个位置。您也可以键入位置，但请确保在位置前面加上`s3://`.
   - 适用于注册账户所在地中，输入Amazon注册位置的账户 ID。默认为您的账户 ID。在跨账户方案中，当向收件人账户中的其他委托人授予数据位置权限时，收件人账户中的数据湖管理员可以在此处指定所有者账户。

- （可选）要允许选定的承担者授予对所选位置的数据位置权限，请选择可授予.



5. 选择 Grant（授权）。

要授予数据位置权限（同一账户，Amazon CLI）

- Run `grant-permissions`命令和授予`DATA_LOCATION_ACCESS`到委托人，将 Amazon S3 路径指定为资源。

Example

以下示例向上授予数据位置权限`s3://retail`给用户`datalake_user1`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail"}}'
```

另请参见：

- Lake Formation 权限参考 (p. 167)

授予数据位置权限（外部账户）

请按照以下步骤将数据位置权限授予外部Amazon账户或组织。

你可以使用 Lake Formation 控制台、API 或Amazon Command Line Interface(Amazon CLI)。

开始前的准备工作

确保满足所有跨账户访问先决条件。有关更多信息，请参阅 跨账户访问权限前提条件 (p. 243)。

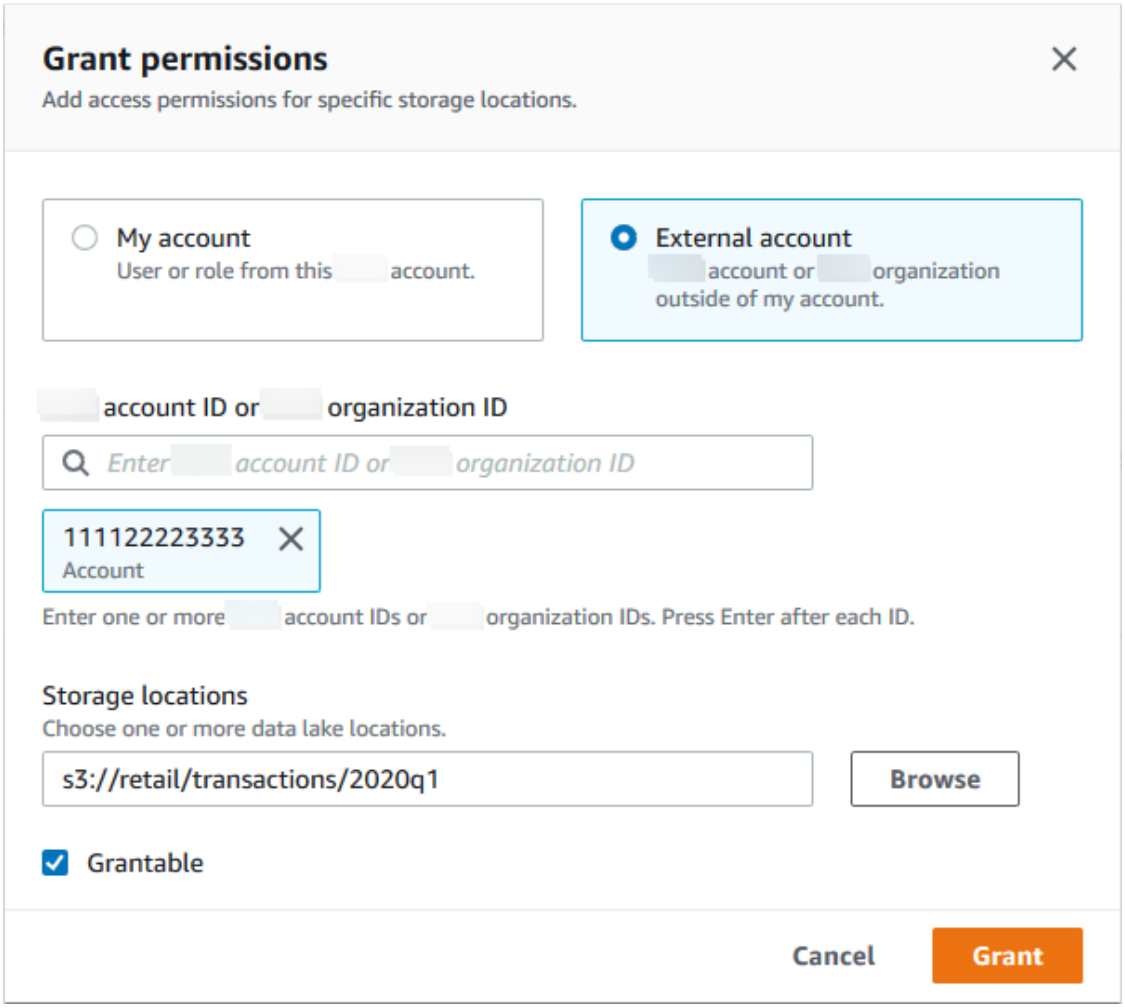
授予数据位置权限（外部账户、控制台）

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/. 以数据湖管理员身份登录。
2. 在导航窗格中，选择数据位置中，然后选择Grant.
3. 在授予权限对话框中，选择外部账户Tile。
4. 提供以下信息：
   - 适用于Amazon账户 ID 或Amazon组织 ID，请输入有效的Amazon账号、组织 ID 或组织单位 ID。

     按Enter在每个身份证之后。

     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-” 组成，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。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。
   - UNDER存储位置，选择浏览，然后选择Amazon S3 的存储位置。该地点必须向 Lake Formation 登记。



5. Select可授予.
6. 选择 Grant（授权）。

要授予数据位置权限（外部账户、Amazon CLI）

- 向外部授予权限Amazonaccount，请输入与以下内容类似的命令。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333  --permissions "DATA_LOCATION_ACCESS" --
permissions-with-grant-option "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"123456789012","ResourceArn":"s3:arn:aws:s3::retail/
transactions/2020q1"}}'
```

此命令授予DATA_LOCATION_ACCESS可以选择授予 Amazon S3 位置上的账户 1111-2222-3333s3://retail/transactions/2020q1，此账户由账户 1234-5678-9012 拥有。

要授予组织权限，请输入与以下内容类似的命令。

```
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":"s3:arn:aws:s3::retail/
transactions/2020q1"}}'
```

此命令授予DATA_LOCATION_ACCESS为组织提供授予选择权o-abcdefghijkl在Amazon S3 位置s3://retail/transactions/2020q1，此账户由账户 1234-5678-9012 拥有。

另请参见：

- Lake Formation 权限参考 (p. 167)

授予与您的账户共享的数据位置的权限

与共享数据目录资源后Amazon账户，作为数据湖管理员，您可以向账户中的其他委托人授予对资源的权限。如果ALTER对共享表授予权限，并且该表指向已注册的 Amazon S3 位置，您还必须授予该位置的数据位置权限。同样，如果CREATE_TABLE要么ALTER在共享数据库上授予权限，并且该数据库具有指向已注册位置的 location 属性，您还必须授予该位置的数据位置权限。

要向账户中的委托人授予共享位置的数据位置权限，您的账户必须已获得DATA_LOCATION_ACCESS具有授予选项的位置的权限。当你授予DATA_LOCATION_ACCESS您账户中的其他委托人，则必须包含数据目录 ID (Amazon账户 ID）的所有者账户。所有者账户是注册该地点的账户。

您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI以授予数据位置权限。

授予与您的账户共享的数据位置的权限（控制台）

- 按 授予数据位置权限（同一账户） (p. 141) 中的步骤操作。

  适用于存储位置，则必须键入位置。适用于注册账户所在地中，输入Amazon拥有者账户户 ID。

要授予与您的账户共享的数据位置的权限（Amazon CLI）

- 输入以下命令之一以授予用户或角色权限。

```
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>"}}'
```

授予和撤消对数据目录资源的权限

您可以在中向委托人授予数据目录权限Amazon Lake Formation以便委托人可以创建和管理数据目录资源，并可以访问基础数据。

您可以授予对元数据数据库和元数据表的 “数据目录” 权限。当您授予对表的权限时，您可以限制对特定表列的访问权限，以实现更精细的访问控制。

您可以授予对单个表的权限，也可以通过单个授予操作授予对数据库中所有表的权限。如果授予对数据库中所有表的权限，则会隐式授予DESCRIBE数据库的权限。然后，该数据库将显示在数据库页面，并由GetDatabasesAPI 操作。

您可以使用命名资源访问控制方法或基于 Lake Formation 标签的访问控制 (LF-TBAC) 方法来授予权限。

您可以在同一个中向委托人授予权限Amazon账户，或外部账户或组织。当您向外部账户或组织授予权限时，您正在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委托人可以访问您拥有的数据目录资源和基础数据。

Note

目前，LF-TBAC 方法支持授予跨账户权限Amazon仅企业或企业部门。

向外部账户或组织授予权限时，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，直到该管理员将共享资源的权限授予外部账户中的其他委托人。

您可以使用以下命令授予数据目录权限Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

主题

另请参见：

- 跨共享数据目录表和数据库Amazon账户 (p. 123)
- 元数据访问控制 (p. 236)
- Lake Formation 权限参考 (p. 167)

使用命名资源方法授予数据目录权限

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和表的权限。您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

主题

使用 Lake Formation 控制台和命名资源方法授予数据库权限

以下步骤说明了如何使用 named resource 方法和授予权限Lake Formation 控制台上的页面。此页面分为以下几个部分：

- 委托人— 用户、角色、Amazon要向其授予权限的账户、企业或企业部门。
- LF 标签或目录资源— 要授予权限的数据库、表或资源链接。
- Permissions (权限)-Lake Formation 权限授予。

Note

要授予对数据库资源链接的权限，请参见授予资源链接权限 (p. 166).

打开 “授予权限” 页面

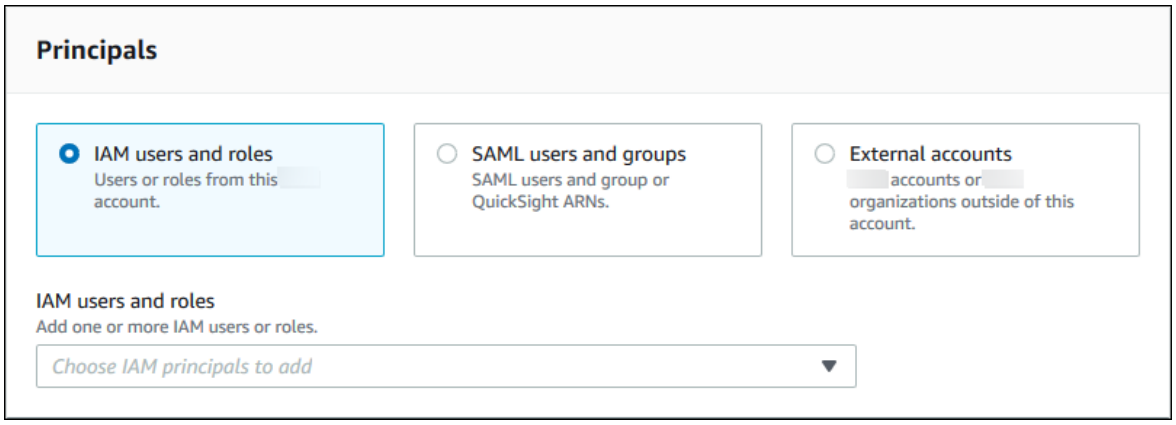
1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/，然后以数据湖管理员、数据库创建者或使用授权选项被授予 Lake Formation 数据库权限的用户身份登录。
2. 请执行下列操作之一：
   - 在导航窗格中，选择数据湖权限. 然后选择 。Grant.
   - 在导航窗格中，选择 Databases（数据库）。然后，在数据库页上，选择一个数据库，然后在操作菜单，在Permissions (权限)，选择Grant.

   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，请在数据库页面上，选择一个资源链接，然后在操作菜单中，选择在目标上授予权限. 有关更多信息，请参阅 资源链接在 Lake Formation 中的工作原理 (p. 126)。

指定委托人

在委托人部分中，选择承担者类型，然后指定要向其授予权限的委托人。



IAM 用户和角色

从IAM 用户和角色list.

SAML 用户和组

适用于SAML 和亚马逊 QuickSight 用户和组中，为通过 SAML 联合的用户或组输入一个或多个亚马逊资源名称 (ARN)，或为亚马逊输入一个或多个亚马逊资源名称 (ARN) QuickSight 用户或组。在每个 ARN 之后按 Enter 键。

有关如何构建 ARN 的信息，请参阅Lake Formation 补助金和撤销Amazon CLI命令 (p. 168).

Note

Lake Formation 与亚马逊的整合 QuickSight Amazon 支持 QuickSight 仅企业版。

External

适用于Amazon账户或Amazon组织，输入一个或多个有效的Amazon账户 ID、组织 ID 或组织单位 ID。按Enter在每个身份证之后。

组织 ID 由 “o-” 后跟 10-32 个小写字母或数字组成。

组织单位 ID 以 “ou-” 开头，后跟 4-32 个小写字母或数字（包含 OU 的根的 ID）。此字符串后跟第二个 “-” 短划线和 8 到 32 个额外的小写字母或数字。

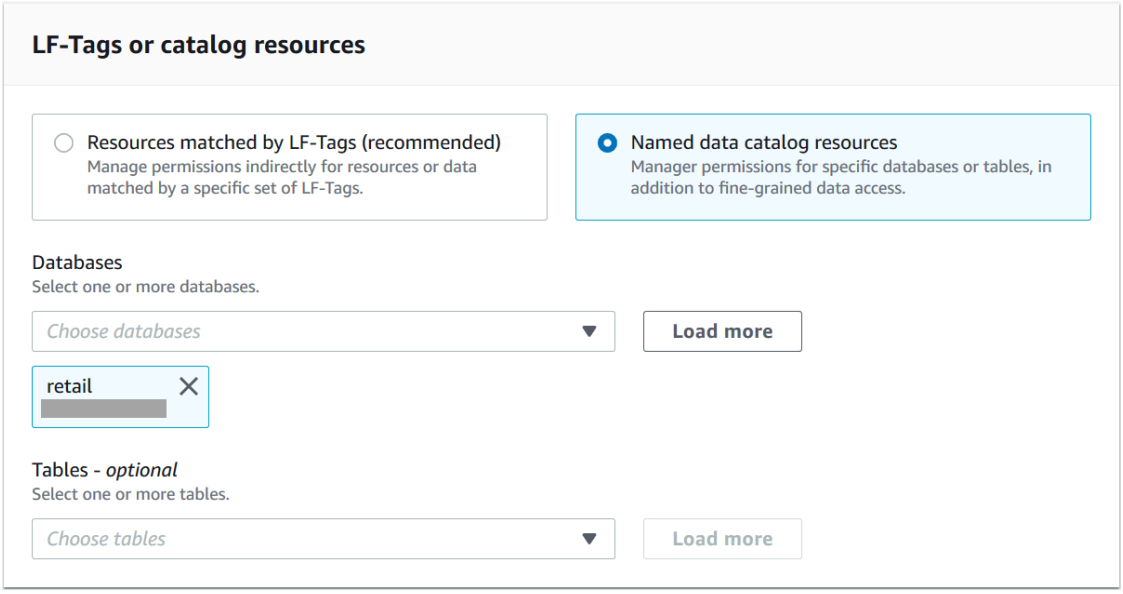
另请参阅

- 访问和查看共享数据目录表和数据库 (p. 123)

指定数据库

在LF 标签或目录资源部分中，选择一个或多个要授予权限的数据库。

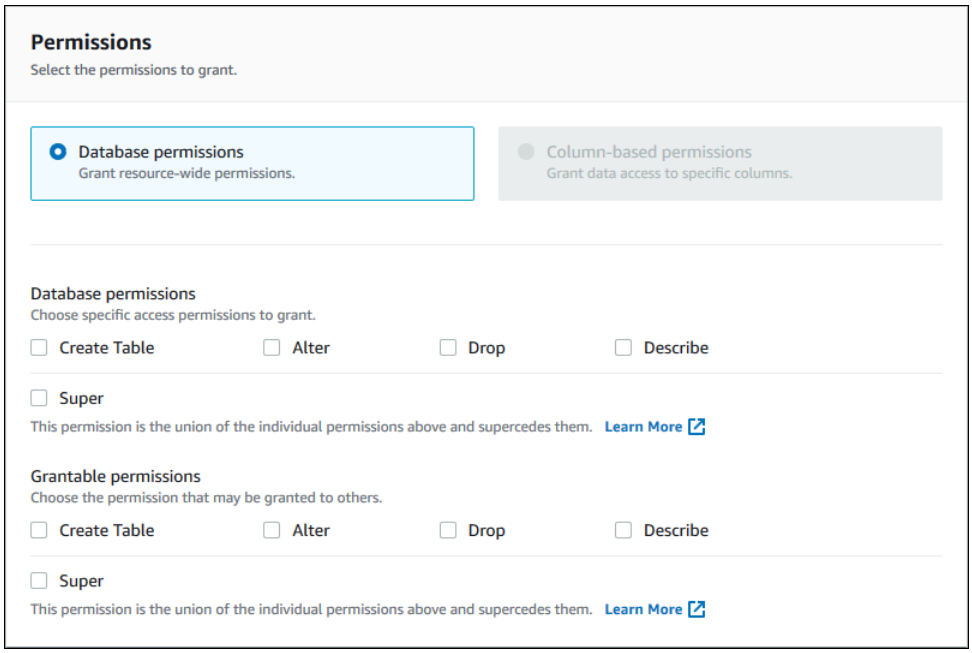
1. 选择命名数据目录资源.



2. 从数据库list.

指定权限

在Permissions (权限)部分，选择权限和可授予的权限。



1. UNDER数据库权限中，选择一个或多个要授予的权限。

   Note

   授予后`Create Table`要么`Alter`对于具有指向已注册位置的 location 属性的数据库，请确保还向承担者授予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 (p. 141)。
2. （可选）在可授予权限中，选择授权接受者可以向其中的其他委托人授予的权限Amazonaccount.
3. 选择 Grant（授权）。

另请参阅

- Lake Formation 权限参考 (p. 167)
- 授予对与您的账户共享的数据库或表的权限 (p. 164)

使用授予数据库权限Amazon CLI和命名资源方法

您可以通过使用 named resource 方法和Amazon Command Line Interface(Amazon CLI)。

使用授予数据库权限Amazon CLI

- 运行`grant-permissions`命令，并将数据库或数据目录指定为资源，具体取决于所授予的权限。

  在以下示例中，将*<account-id>*使用有效的Amazon账户 ID。

  Example — 授予权限以创建数据库

  这个例子授予`CREATE_DATABASE`给用户`datalake_user1`. 由于授予此权限的资源是 Data Catalog，因此该命令指定了一个空`CatalogResource`结构作为`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

接下来的例子是补助金CREATE_TABLE在数据库上retail给用户datalake_user1.

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

Example — 拨款给外部Amazon具有 “授予” 选项的账户

接下来的例子是补助金CREATE_TABLE在数据库上使用 grant 选项retail转至外部账户 11-2222-33333333。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE" --permissions-
with-grant-option "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example — 为组织授予补助金

接下来的例子是补助金ALTER在数据库上使用 grant 选项issues给组织o-abcdefghijkl.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl --permissions "ALTER" --permissions-with-grant-option "ALTER" --resource
 '{ "Database": {"Name":"issues"}}'
```

Note

授予后CREATE_TABLE要么ALTER对于具有指向已注册位置的 location 属性的数据库，请确保还向承担者授予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 (p. 141)。

另请参阅

- Lake Formation 权限参考 (p. 167)
- 授予对与您的账户共享的数据库或表的权限 (p. 164)

使用 Lake Formation 控制台和命名资源方法授予表权限

您可以使用 Lake Formation 控制台和命名资源方法授予 Lake Formation 对数据目录表的权限。您可以授予对单个表的权限，也可以通过单个授予操作授予对数据库中所有表的权限。

如果授予对数据库中所有表的权限，则会隐式授予DESCRIBE数据库的权限。然后，该数据库将显示在数据库页面，并由GetDatabasesAPI 操作。

在选择时SELECT作为要授予的权限，您可以选择应用列筛选器、行筛选器或单元格筛选器。

以下步骤说明了如何使用 named resource 方法和授予权限Lake Formation 控制台上的页面。此页面分为以下几个部分：

- 委托人— 用户、角色、Amazon要向其授予权限的账户、企业或企业部门。
- LF 标签或目录资源— 要授予权限的数据库、表或资源链接。
- Permissions (权限)-Lake Formation 权限授予。

Note

要授予对表资源链接的权限，请参见授予资源链接权限 (p. 166).

打开 “授予权限” 页面

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/，然后以数据湖管理员、表创建者或使用授权选项被授予表权限的用户身份登录。
2. 请执行下列操作之一：

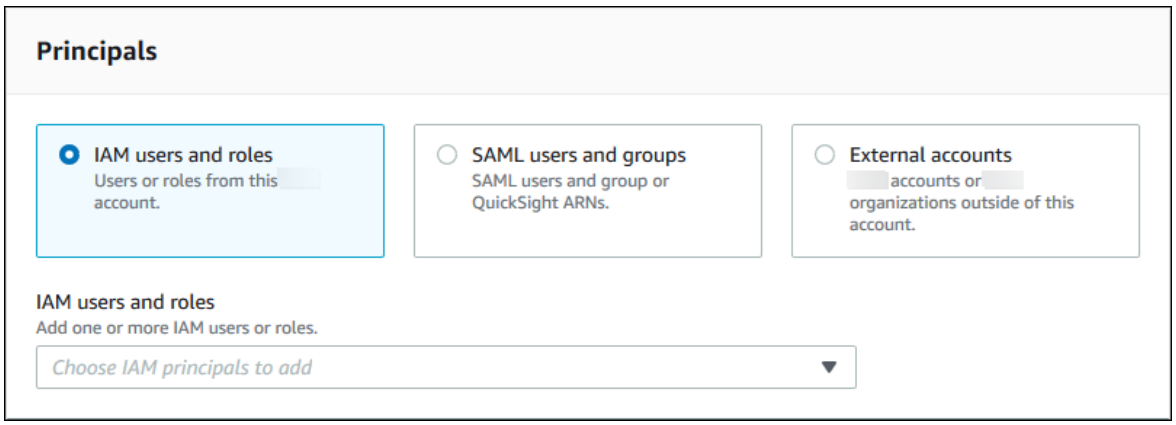
   - 在导航窗格中，选择数据权限. 然后选择 。Grant.
   - 在导航窗格中，选择表。然后，在表页面上，选择一个表，然后在操作菜单，在Permissions (权限)，选择Grant.

   Note

   您可以通过表的资源链接授予对表的权限。为此，请在表页面上，选择一个资源链接，然后在操作菜单中，选择在目标上授予权限. 有关更多信息，请参阅 资源链接在 Lake Formation 中的工作原理 (p. 126)。

指定委托人

在委托人部分中，选择承担者类型并指定要向其授予权限的委托人。



IAM 用户和角色

从IAM 用户和角色list.

SAML 用户和组

适用于SAML 和亚马逊 QuickSight 用户和组中，为通过 SAML 联合的用户或组输入一个或多个亚马逊资源名称 (ARN)，或为亚马逊输入一个或多个亚马逊资源名称 (ARN) QuickSight 用户或组。在每个 ARN 之后按 Enter 键。

有关如何构建 ARN 的信息，请参阅Lake Formation 补助金和撤销Amazon CLI命令 (p. 168).

Note

Lake Formation 与亚马逊的整合 QuickSight Amazon 支持 QuickSight 仅企业版。

External

适用于Amazon账户或Amazon组织，输入一个或多个有效的Amazon账户 ID、组织 ID 或组织单位 ID。按Enter在每个身份证之后。

组织 ID 由 “o-” 后跟 10-32 个小写字母或数字组成。

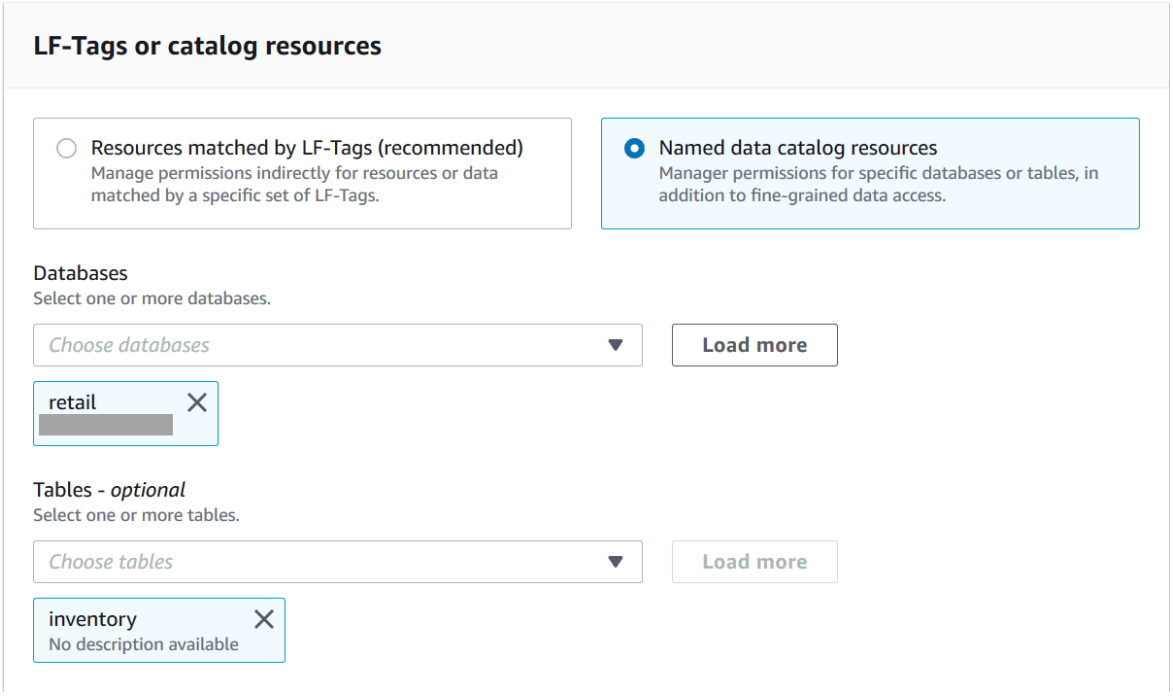
组织单位 ID 以 “ou-” 开头，后跟 4-32 个小写字母或数字（包含 OU 的根的 ID）。此字符串后跟第二个 “-” 字符和 8 到 32 个附加的小写字母或数字。

另请参见：

- 访问和查看共享数据目录表和数据库 (p. 123)

指定表

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，或者所有桌子.



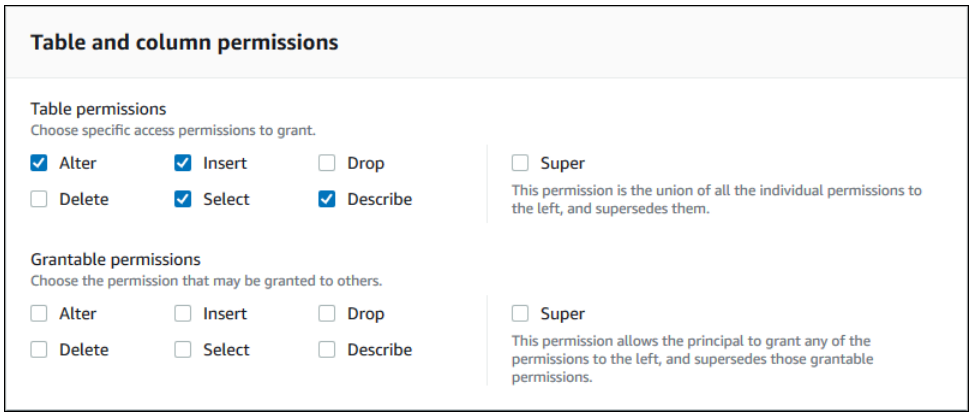
指定权限

在Permissions (权限)部分，执行以下操作之一以选择权限和可授予的权限：

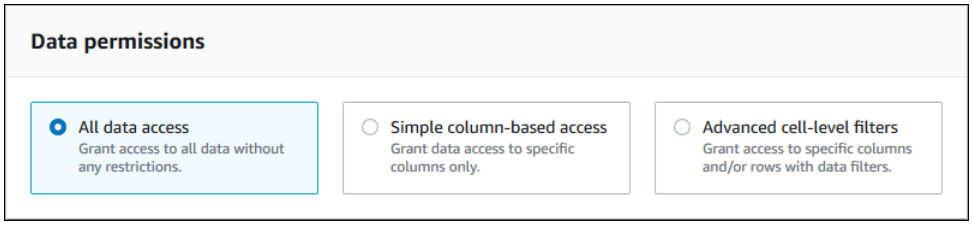
- 指定表权限（无数据筛选）(p. 152)
- 指定Select数据筛选权限 (p. 153)

指定表权限（无数据筛选）

1. 选择要授予的表权限，也可以选择可授予的权限。



如果你同意Select，数据权限部分显示在表和列权限部分，使用所有数据访问选项默认处于选中状态。接受默认值。



2. 选择 Grant（授权）。

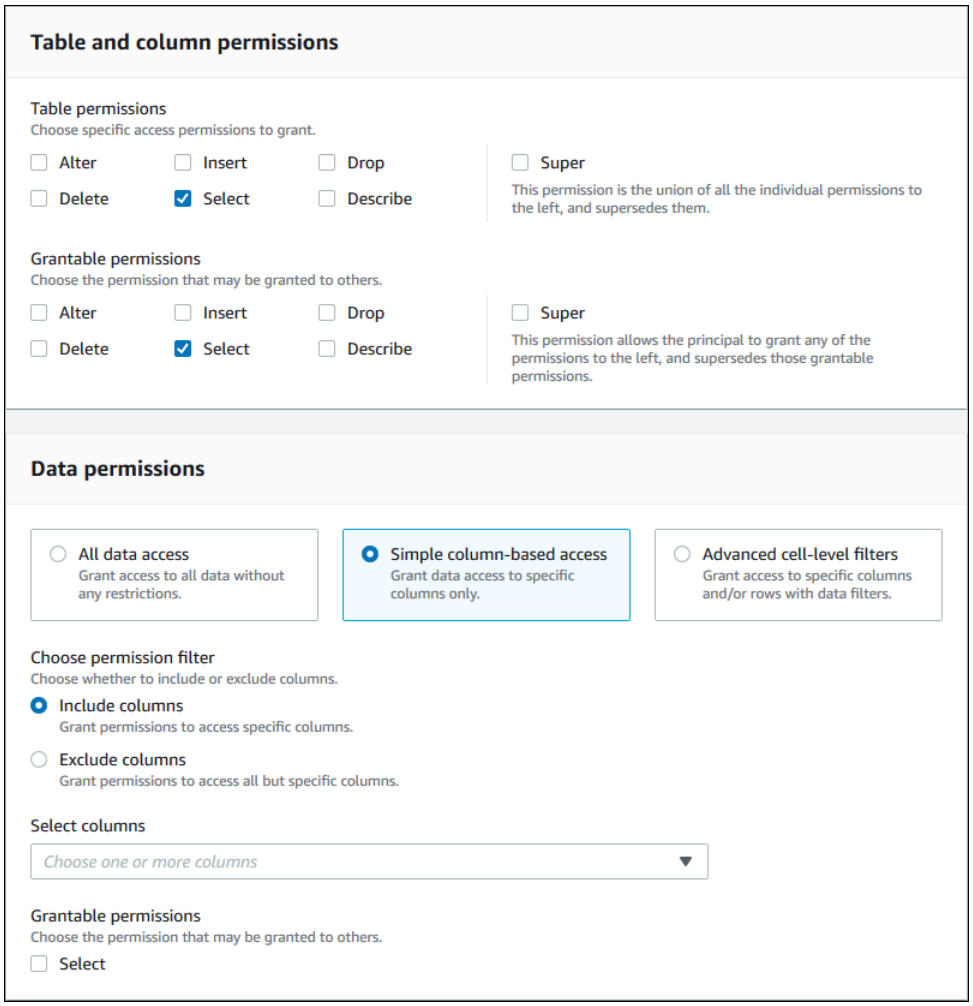
指定Select数据筛选权限

1. SelectSelect授权。不要选择任何其他权限。

   这些区域有：数据权限部分显示在表和列权限部分。
2. 请执行下列操作之一：

   - 仅应用简单列过滤。

     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。

   向外部授予权限时，仅支持包含列表Amazon账户或组织。

3. （可选）在可授予权限中，打开 “选择” 权限的授予选项。

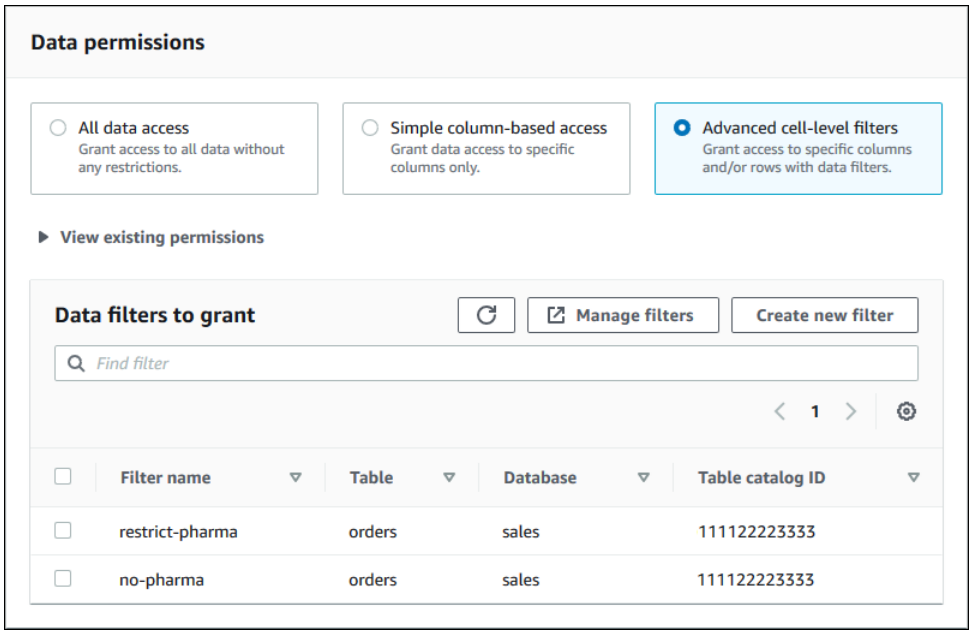
   如果包含 grant 选项，则授予接收者只能对您授予他们的列授予权限。

   Note

   您也可以仅通过创建指定列筛选器并将所有行指定为行筛选器的数据筛选器来应用列筛选。但是，这需要更多步骤。

- 应用列、行或单元格筛选。

  1. 选择高级单元格级过滤器.



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）要查看列出的筛选器的详细信息，或者要创建新筛选器或删除现有筛选器，请选择管理筛选条件.

   这些区域有：数据筛选条件页面将在新的浏览器窗口中打开。

   完成后数据筛选条件页面，返回授予权限页面，如有必要，刷新该页面以查看您创建的任何新数据筛选器。
5. 选择一个或多个要应用于补助金的数据筛选器。

   Note

   如果列表中没有数据筛选器，则表示没有为所选表创建数据筛选器。

3. 选择 Grant（授权）。

另请参阅

- Lake Formation 访问控制概述 (p. 234)
- Lake Formation 中的数据过滤和细胞级安全 (p. 206)
- Lake Formation 角色和 IAM 权限参考 (p. 322)

使用授予表权限Amazon CLI和命名的资源方法

您可以通过使用 named resource 方法和Amazon Command Line Interface(Amazon CLI)。

使用授予表权限Amazon CLI

- 运行`grant-permissions`命令，并指定一个表作为资源。

Example — 对单个表格授予-无过滤

以下几个例子授予SELECT和ALTER给用户datalake_user1在Amazon账户 11-2222-3333333 在桌子上inventory在数据库中retail.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --permissions
 "SELECT" "ALTER" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory"}}'
```

Note

如果你授予ALTER对在注册位置具有基础数据的表的权限，请务必同时向主体授予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 (p. 141)。

Example — 使用 “授予” 选项对所有表格授予-无过滤

接下来的例子是补助金SELECT在数据库中的所有表上使用 grant 选项retail.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --permissions
 "SELECT" --permissions-with-grant-option "SELECT" --resource '{ "Table": { "DatabaseName":
 "retail", "TableWildcard": {} } }'
```

Example — 通过简单的列筛选授予

接下来的例子授予SELECT在表的列子集上persons. 它使用简单的列过滤。

```
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 — 使用数据筛选器授予

这个例子授予SELECT在orders表并应用restrict-pharma数据筛选条件。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是文件的内容grant-params.json.

```
{
    "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["SELECT"],
    "PermissionsWithGrantOption": ["SELECT"]
}
```

另请参阅

- Lake Formation 访问控制概述 (p. 234)
- Lake Formation 中的数据过滤和细胞级安全 (p. 206)
- Lake Formation 权限参考 (p. 167)

使用 LF-TBAC 方法授予数据目录权限

您可以使用基于 Lake Formation 标记的访问控制 (LF-TBAC) 方法授予 Lake Formation 对数据目录数据库、表和列的权限。

您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

主题

- 使用 Lake Formation 控制台和 LF-TBAC 方法授予数据目录权限 (p. 157)
- 使用授予数据目录权限Amazon CLI和 LF-TBAC 方法 (p. 160)

另请参阅

- 授予、撤销和列出 LF-tag 权限 (p. 199)
- 管理用于元数据访问控制的 LF 标签 (p. 186)
- Lake Formation 标签访问控制 (p. 179)

使用 Lake Formation 控制台和 LF-TBAC 方法授予数据目录权限

以下步骤说明了如何使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法和授予权限Lake Formation控制台上的页面。此页面分为以下几个部分：

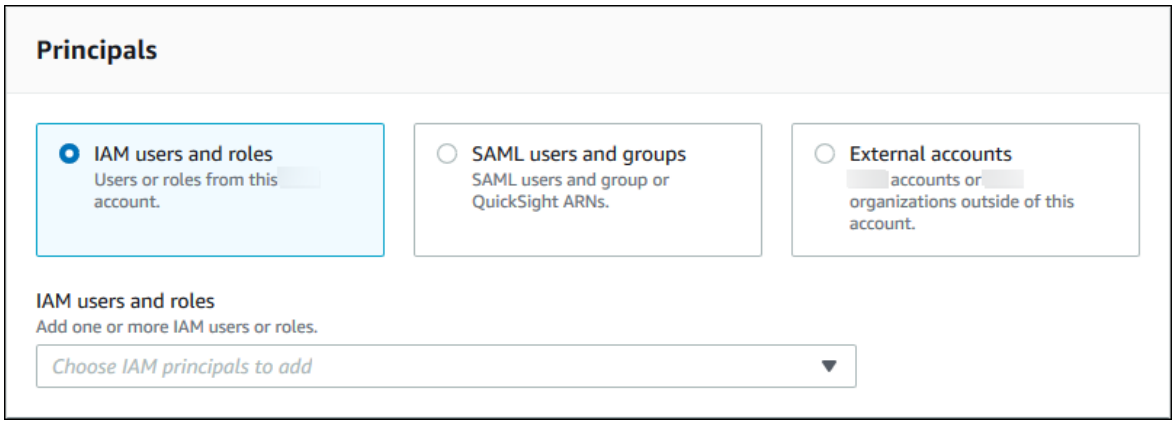
- 委托人— 用户、角色和Amazon Web Services 账户以向授予权限。
- LF 标签或目录资源— 要授予权限的数据库、表或资源链接。
- Permissions (权限)-Lake Formation 权限授予。

打开 “授予权限” 页面

1. 打开Amazon Lake Formation控制台位于https://console.aws.amazon.com/lakeformation/，然后以数据湖管理员身份登录，或者通过授予选项通过 LF-TBAC 被授予对数据目录资源的 Lake Formation 权限的用户登录。
2. 在导航窗格中，选择数据权限. 然后选择 。Grant.

指定委托人

在委托人部分中，选择承担者类型，然后指定要向其授予权限的委托人。



IAM 用户和角色

从IAM 用户和角色list.

SAML 用户和组

适用于SAML 和亚马逊 QuickSight 用户和组中，为通过 SAML 联合的用户或组输入一个或多个亚马逊资源名称 (ARN)，或为亚马逊输入一个或多个亚马逊资源名称 (ARN) QuickSight 用户或组。在每个 ARN 之后按 Enter 键。

有关如何构建 ARN 的信息，请参阅Lake Formation 补助金和撤销Amazon CLI命令 (p. 168).

Note

Lake Formation 与亚马逊的整合 QuickSight Amazon 支持 QuickSight 仅企业版。

External

适用于Amazon Web Services 账户要么Amazon组织，输入一个或多个有效的Amazon Web Services 账户ID、组织 ID 或组织单位 ID。按Enter在每个身份证之后。

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 以 “ou-” 开头，后跟 4 到 32 个小写字母或数字（包含 OU 的根的 ID）。此字符串后跟第二个 “-” 短划线和 8 到 32 个额外的小写字母或数字。

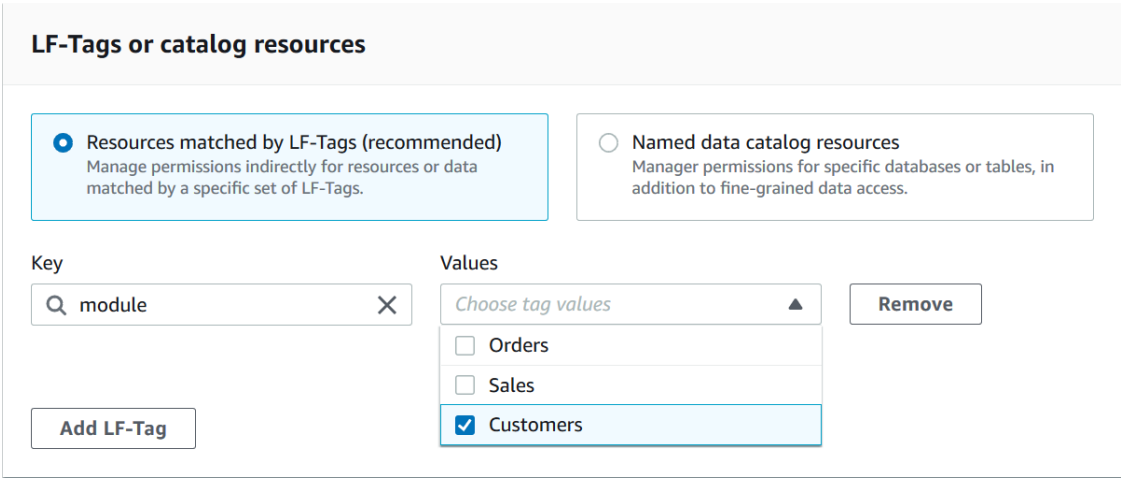
另请参阅

- 访问和查看共享数据目录表和数据库 (p. 123)

指定 LF 标签

1. 确保LF 标签匹配的资源已选择。
2. 选择添加 LF-tag.
3. 选择一个 LF-tag 键和数值。

如果您选择了多个值，则您正在创建一个 LF-tag 表达式OR运算符。这意味着，如果任何 LF-tag 值与分配给数据目录资源的 LF-tag 相匹配，则会授予您对该资源的权限。



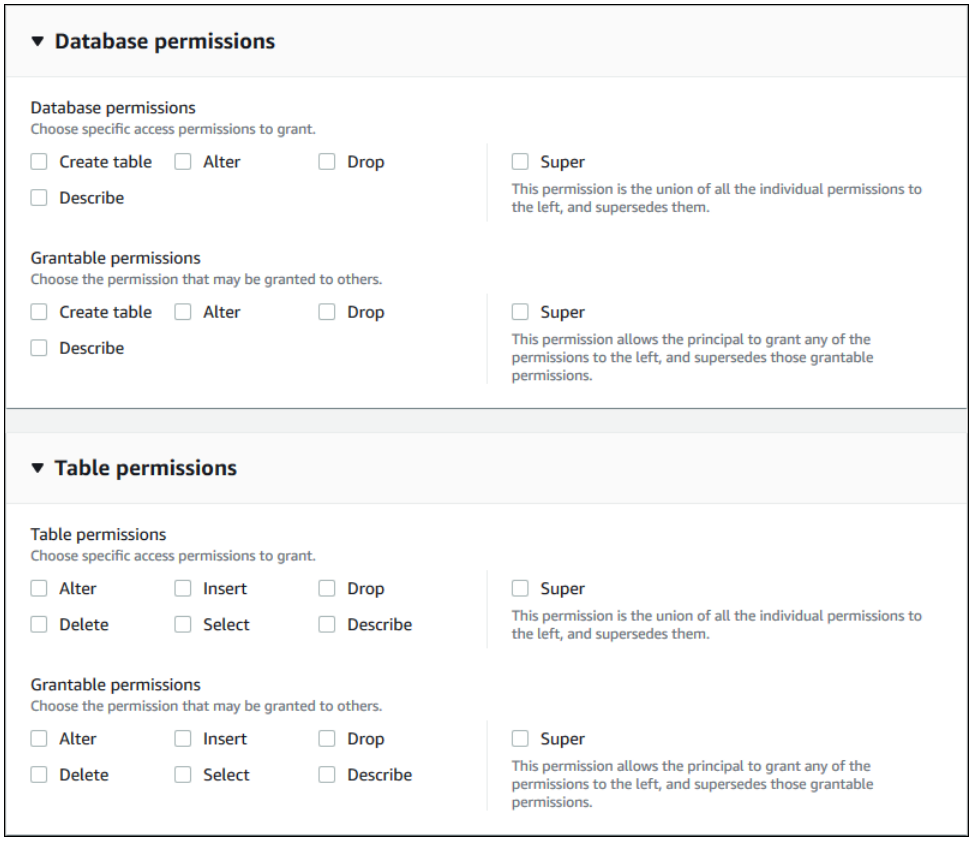
4. (可选) 选择添加 LF-tag再次指定另一个 LF 标签。

   如果您指定了多个 LF-tag，则会创建一个 LF-tag 表达式AND运算符。只有在为 LF-tag 表达式中的每个 LF-tag 分配了匹配的 LF-tag 时，才会向委托人授予对数据目录资源的权限。

指定权限

指定向委托人授予执行匹配数据目录资源的权限。匹配资源是指分配给委托人的 LF-tag 与授予主体的 LF-tag 表达式之一相匹配的资源。

您可以指定要对匹配的数据库、匹配表或两者都授予的权限。



1. UNDER数据库权限中，选择要在匹配的数据库上向承担者授予的数据库权限。
2. UNDER表权限中，选择要授予承担者对匹配表的表权限。
3. 选择 Grant（授权）。

使用授予数据目录权限Amazon CLI和 LF-TBAC 方法

您可以使用Amazon Command Line Interface(Amazon CLI）和 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法，用于授予 Lake Formation 对数据目录数据库、表和列的权限。

使用 LF-TBAC 授予数据目录权限 (Amazon CLI）

- 使用 `grant-permissions` 命令。

  Example

  以下示例授予 LF-tag 表达式"`module=*`"（LF-tag 键的所有值`module`) 发送给用户`datalake_user1`. 该用户将拥有`CREATE_TABLE`对所有匹配数据库的权限-已为其分配了 LF-tag 和密钥的数据库`module`，带有任何值。

```
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

  下一个示例授予 LF-tag 表达式"`(level=director) AND (region=west OR region=south)`"给用户`datalake_user1`. 该用户将拥有`SELECT`、`ALTER`, 和`DROP`对匹配表（已同时分配了两个表的表）使用授予选项的权限`level=director`和 (`region=west`要么`region=south`)。

```
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

  下一个示例授予 LF-tag 表达式"`module=orders`"到Amazon账户 1234-5678-9012。然后，该账户中的数据湖管理员可以授予"`module=orders`"在他们的账户中向委托人表达。然后那些校长会有`CREATE_TABLE`对账户 1111-2222-3333 拥有、使用命名资源方法或 LF-TBAC 方法与账户 1234-5678-9012 共享的匹配数据库的权限。

```
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"]}]}}'
```

在 Lake Formation 中查看数据库和表权限

您可以查看在数据目录数据库或表上授予的 Lake Formation 权限。您可以使用 Lake Formation 控制台、API 或Amazon Command Line Interface(Amazon CLI)。

从控制台开始，您可以查看权限数据库要么表页面，或者来自数据权限页.

> Note
>
> 如果您不是数据库管理员或资源所有者，则只有在具有授予选项的 Lake Formation 权限的情况下，才能查看其他委托人对资源拥有的权限。
> 除了所需的 Lake Formation 权限之外，您还需要Amazon Identity and Access Management(IAM) 权限`glue:GetDatabases`、`glue:GetDatabase`、`glue:GetTables`、`glue:GetTable`, 和`glue:ListPermissions`.

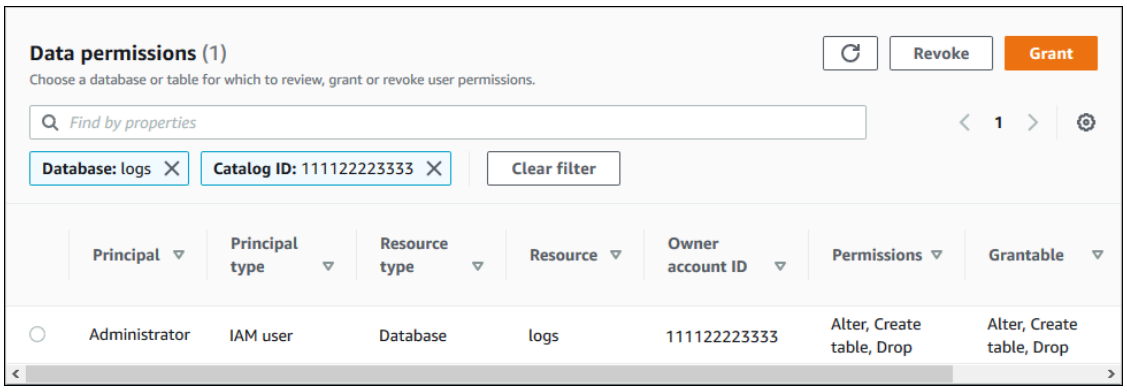
查看数据库的权限（控制台，从数据库页面开始）

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、数据库创建者或使用授予选项对数据库具有任何 Lake Formation 权限的用户身份登录。
2. 在导航窗格中，选择 Databases（数据库）。
3. 选择一个数据库，然后在操作选择菜单，选择查看权限.

   > Note
   >
   > 如果选择数据库资源链接，Lake Formation 将显示对资源链接的权限，而不是资源链接的目标数据库上的权限。

   这些区域有：数据权限页面列出了该数据库的所有 Lake Formation 权限。数据库名称和目录 ID (Amazon数据库所有者的账户 ID）在搜索框下显示为标签。磁贴表示筛选器已应用于仅列出该数据库的权限。您可以通过关闭磁贴或选择来调整过滤器清除过滤器.



查看数据库的权限（控制台，从 “数据权限” 页面开始）

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、数据库创建者或使用授予选项对数据库具有任何 Lake Formation 权限的用户身份登录。
2. 在导航窗格中，选择和。数据权限.
3. 将光标放在页面顶部的搜索框中和属性出现的菜单，选择数据库.
4. 在存储库的数据库菜单中，选择一个数据库。

Note

如果选择数据库资源链接，Lake Formation 将显示对资源链接的权限，而不是资源链接的目标数据库上的权限。

这些区域有：数据权限页面列出了该数据库的所有 Lake Formation 权限。数据库名称显示为搜索框下方的磁贴。磁贴表示筛选器已应用于仅列出该数据库的权限。您可以通过关闭磁贴或选择来删除过滤器清除过滤器.

查看表的权限（控制台，从表页面开始）

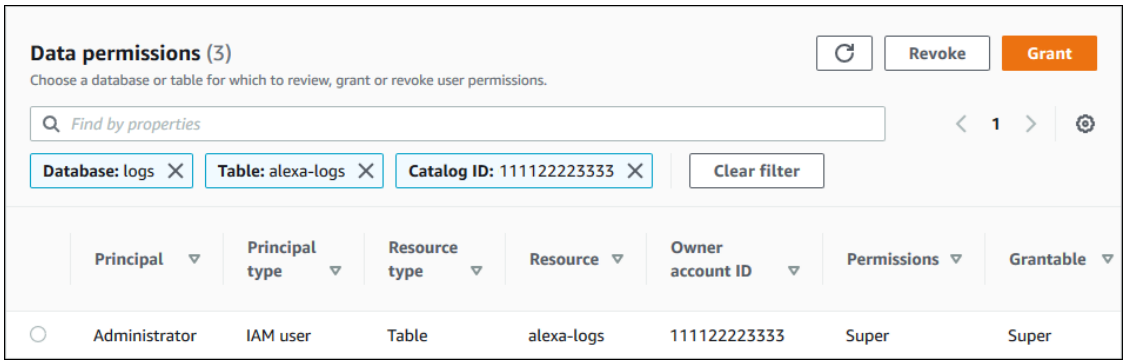
1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、表格创建者或使用授予选项对表具有任何 Lake Formation 权限的用户身份登录。
2. 在导航窗格中，选择表。
3. 选择一张桌子，然后在操作选择菜单，选择查看权限.

   Note

   如果选择表资源链接，Lake Formation 将显示对资源链接的权限，而不是在资源链接的目标表上显示权限。

   这些区域有：数据权限页面列出了该表的所有 Lake Formation 权限。表名、包含表的数据库的数据库名称以及目录 ID (Amazon表所有者的账户 ID）在搜索框下显示为标签。这些标签表明筛选器已应用于仅列出该表的权限。您可以通过关闭标签或选择来调整过滤器清除过滤器.



查看表格上的权限（控制台，从“数据权限”页面开始）

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、表格创建者或使用授予选项对表具有任何 Lake Formation 权限的用户身份登录。
2. 在导航窗格中，选择和。数据权限.
3. 将光标放在页面顶部的搜索框中和属性出现的菜单，选择数据库.
4. 在存储库的数据库菜单中，选择一个数据库。

   Important

   如果你想查看与你共享的表的权限Amazon帐户来自外部帐户，则必须在包含表的外部帐户中选择数据库，而不是指向数据库的资源链接。

   这些区域有：数据权限页面列出了该数据库的所有 Lake Formation 权限。
5. 再次将光标放在搜索框中，然后在属性出现的菜单，选择表.
6. 在存储库的表出现的菜单中，选择一个表格。

这些区域有：数据权限页面列出了该表的所有 Lake Formation 权限。包含该表的数据库的表名和数据库名称在搜索框下显示为切片。磁贴表示筛选器已应用于仅列出该表的权限。您可以通过关闭磁贴或选择来调整过滤器清除过滤器.

查看表格上的权限 (Amazon 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。这些区域有：`CatalogId`属性是Amazon外部账户的帐户 ID，数据库名称是指包含该表的外部帐户中的数据库。

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

授予跨账户资源的权限

要启用跨账户访问，您可以将 Lake Formation 权限与数据目录表和数据库（数据目录资源）的授予选项授予外部Amazon账户、组织或组织单位。授权操作会自动共享这些资源。

您不与外部特定委托人共享资源Amazon账户-仅与账户共享资源。向组织或组织单位授予 Lake Formation 权限相当于向每个组织或组织单位授予权限Amazon该组织或组织单位的账户。

Amazon Lake Formation现在提供了新版本的跨账户补助金，可以最佳利用Amazon RAM容量以最大限度地提高跨账户使用率。当你与外部共享资源时Amazon Web Services 账户，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 减少了消费者需要接受的资源共享邀请的数量。

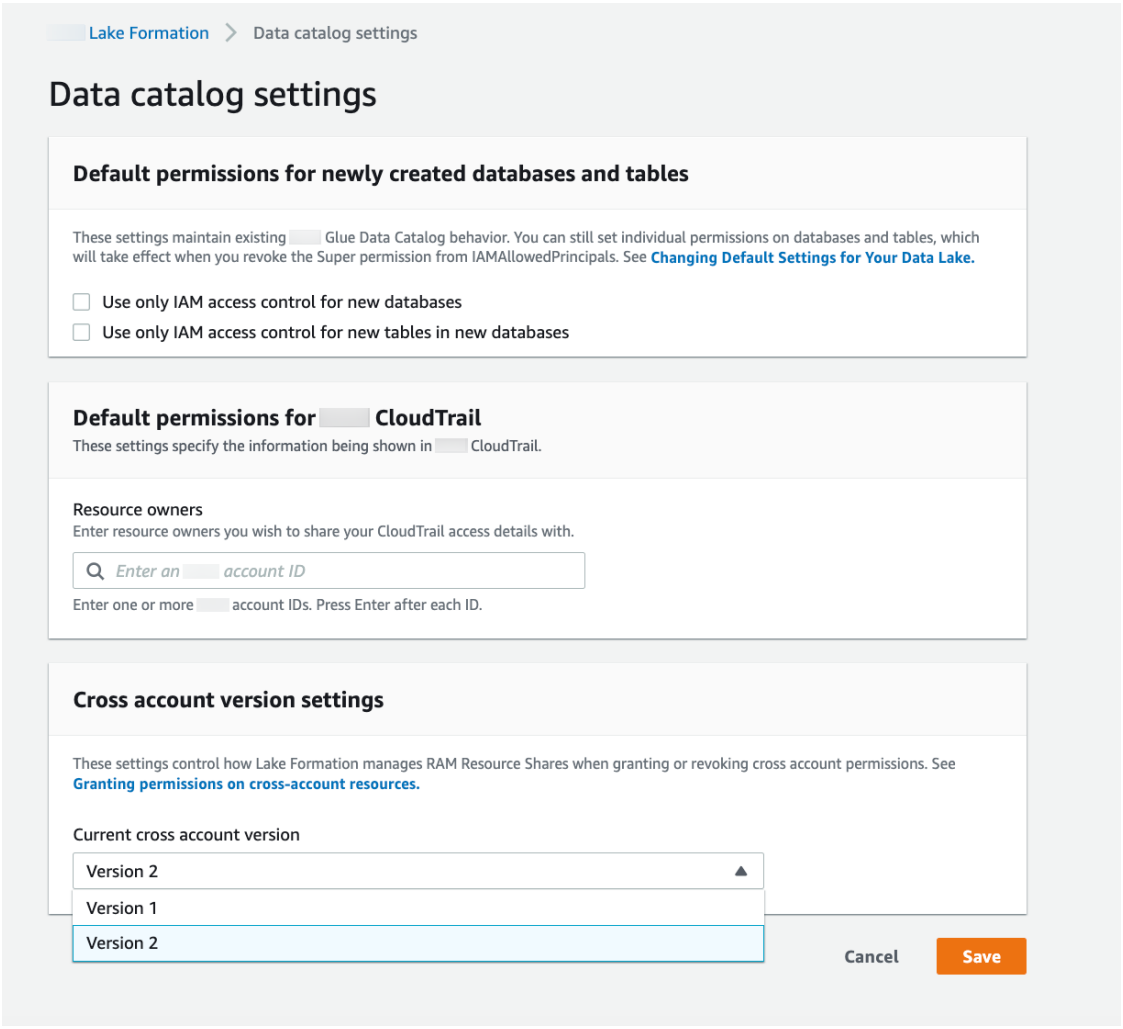
有关 的更多信息Amazon RAM资源限制，请参阅跨账户最佳实践和限制 (p. 245).

启用新版本

1. 如果跨账户权限授予者具有`AWSLakeFormationCrossAccountManager`托管 IAM 策略权限，则跨账户权限授予者角色或委托人不需要额外的权限设置。但是，如果跨账户授予者未使用托管策略，则授予者角色或委托人应具有以下 IAM 权限，新版本的跨账户授予才能成功。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
         "ram:ResourceShareName": "LakeFormation*"
        }
      }
    }
  ]
}
```

2. 选择版本 2下跨账户版本设置在数据目录设置页. 如果你选择版本 1，Lake Formation 将使用默认的资源共享模式。



Important

一旦你选择了版本 2，所有新的授权都将通过新的跨账户授予模式。以最佳方式使用Amazon RAM现有跨账户共享的容量，我们建议您撤销使用旧版本发出的授权，然后在新模式下重新授予。

Amazon 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 (p. 126).

授予对与您的账户共享的数据库或表的权限

在属于另一个数据目录资源之后Amazon账户已与你共享Amazon账户，作为数据湖管理员，您可以将共享资源的权限授予账户中的其他委托人。但是，您不能将资源的权限授予其他人Amazon账户或组织。

您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI) 以授予权限。

授予对共享数据库的权限（命名资源方法，控制台）

- 按照中的说明进行操作使用 Lake Formation 控制台和命名资源方法授予数据库权限 (p. 146) 在数据库在下面列出LF 标签或目录资源，请确保在外部帐户中选择数据库，而不是数据库的资源链接。

  如果您在数据库列表中没有看到该数据库，请确保您已接受Amazon Resource Access Manager(Amazon RAM) 数据库的资源共享邀请。有关更多信息，请参阅 接受来自的资源共享邀请Amazon RAM (p. 124)。

  另外，对于`CREATE_TABLE`和`ALTER`权限，请遵循中的说明授予数据位置权限（同一账户）(p. 141)，请务必在注册账户所在地字段中返回的子位置类型。

授予对共享表的权限（命名资源方法，控制台）

- 按照中的说明进行操作使用 Lake Formation 控制台和命名资源方法授予表权限 (p. 150) 在数据库在下面列出LF 标签或目录资源，请确保在外部帐户中选择数据库，而不是数据库的资源链接。

  如果您在表列表中没有看到该表，请确保您已接受Amazon RAM表的资源共享邀请。有关更多信息，请参阅 接受来自的资源共享邀请Amazon RAM (p. 124)。

  另外，对于`ALTER`权限，请按照中的说明操作授予数据位置权限（同一账户）(p. 141)，请务必在注册账户所在地字段中返回的子位置类型。

授予对共享资源的权限（LF-TBAC 方法，控制台）

- 按照中的说明进行操作使用 Lake Formation 控制台和 LF-TBAC 方法授予数据目录权限 (p. 157) 在LF 标签或目录资源部分中，授予外部账户授予您的账户的精确 LF-tag 表达式或该表达式的子集。

  例如，如果外部账户授予了 LF-tag 表达式`module=customers AND environment=production`使用授权选项添加到您的账户，作为数据湖管理员，您可以授予相同的表达式，或者`module=customers`要么`environment=production`给您账户中的委托人。你只能授予相同或部分Lake Formation 权限（例如`SELECT`、`ALTER`，依此类推），这些是通过 LF-tag 表达式授予资源的。

要授予对共享表（名为 resource 方法）的权限，Amazon CLI）

- 输入类似以下的命令。在这个示例中：
  - 您的Amazon账户 ID 为 112222-322-322-322-3322-
  - 拥有该表并将其授予您的账户的账户是 1234-5678-9012。
  - 这些区域有：`SELECT`正在授予对共享表的权限`pageviews`给用户`datalake_user1`. 该用户是您账户的委托人。
  - 这些区域有：`pageviews`表位于`analytics`数据库，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，拥有账户必须在`CatalogId`属性位于`resource`参数。

授予资源链接权限

请按照以下步骤授予Amazon Lake Formation对一个或多个资源的权限链接到您的中的委托人Amazonaccount.

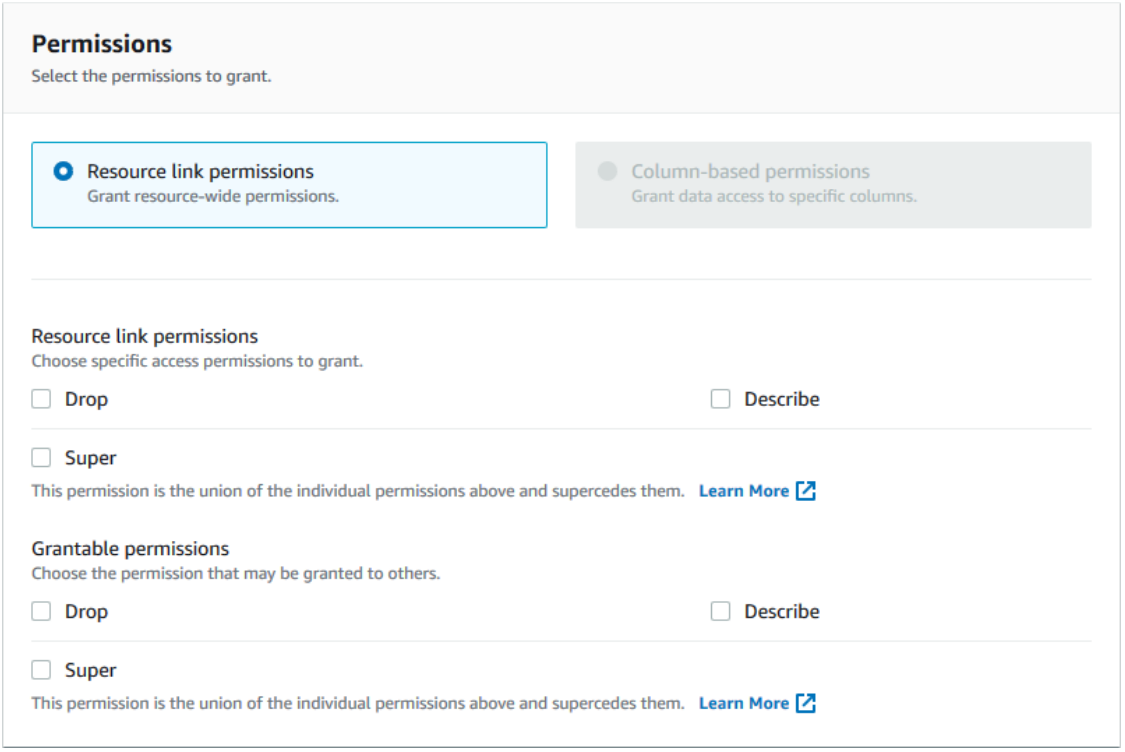
创建资源链接后，只有您可以查看和访问它。（这假设仅对数据库中的新表使用 IAM 访问控制未为数据库启用。）要允许账户中的其他委托人访问资源链接，请至少授予DESCRIBE权限。

> **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。您必须对目标单独授予权限。

你可以使用 Lake Formation 控制台、API 或Amazon Command Line Interface(Amazon CLI)。

授予资源链接权限（控制台）

1. 请执行下列操作之一：
   - 对于数据库资源链接，请按照中的步骤操作使用 Lake Formation 控制台和命名资源方法授予数据库权限 (p. 146)执行以下操作：
     1. 打开 “授予权限” 页面 (p. 147).
     2. 指定数据库 (p. 148). 指定一个或多个数据库资源链接。
     3. 指定委托人 (p. 147).
   - 有关表资源链接，请按照中的步骤操作使用 Lake Formation 控制台和命名资源方法授予表权限 (p. 150)执行以下操作：
     1. 打开 “授予权限” 页面 (p. 151).
     2. 指定表 (p. 152). 指定一个或多个表资源链接。
     3. 指定委托人 (p. 151).
2. UNDERPermissions (权限)中，选择要授予的权限。（可选）选择可授予权限。



3. 选择 Grant（授权）。

要授予资源链接权限 (Amazon CLI)

- 运行grant-permissions命令，将资源链接指定为资源。

  Example

  这个例子授予DESCRIBE给用户datalake_user1在表格资源链接上incidents-link在数据库中issues在Amazon账户 112222-322-322-3322-3

```
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 (p. 126)
  - Lake Formation 权限参考 (p. 167)

使用Lake Formation 控制台撤销权限

您可以使用控制台撤消所有类型的 Lake Formation 权限-数据目录权限、策略标记权限、数据筛选器权限和位置权限。

撤消对资源的 Lake Formation 权限（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员身份登录，或者以已通过资源的授权选项被授予权限的用户身份登录。
2. 在导航窗格中的Permissions (权限)，选择标签权限、数据权限，或者数据位置.
3. 选择权限或位置，然后选择REVOKE.
4. 在打开的对话框中，选择REVOKE.

Lake Formation 权限参考

执行Amazon Lake Formation操作，委托人需要 Lake Formation 权限和Amazon Identity and Access Management(IAM) 权限。您通常使用粗粒度访问控制策略，如中所述the section called “Lake Formation 访问控制概述” (p. 234). 您可 Lake Formation 使用控制台、API 或Amazon Command Line Interface(Amazon CLI)。

要了解如何通过 Lake Formation 授予或撤消权限的信息，请参阅the section called “授予和撤消数据目录权限” (p. 145)和the section called “授予数据位置权限” (p. 141).

Note

本节中的示例演示如何向委托人授予权限的信息。Amazonaccount. 有关跨账户授权的示例，请参阅the section called “FormLake Formation 权限概述 ” (p. 138).

主题

Lake Formation 补助金和撤销Amazon CLI命令

本节中的每个权限描述都包括使用Amazon CLI命令。以下是 Lake Formation 的大纲grant-permissions和revoke-permissions Amazon CLI命令。

```
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描述，请参阅授予权限和撤消权限中的Amazon CLI命令参考. 本节提供了有关`--principal`选项。

的价值`--principal`选项是以下任一：

- ARN 于Amazon Identity and Access Management(IAM) 用户或角色
- 针对通过 SAML 提供商进行身份验证的用户或组的 ARN，例如 Microsoft Active Directory 联合身份验证服务 (AD FS)
- 针对 Amazon QuickSight 用户或组的 ARN
- 对于跨账户权限，请将 ARN 用于Amazon账户 ID、组织 ID 或组织单位 ID

以下是所有人的语法和示例`--principal`类型。

委托人是 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
```

委托人是 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
```

主体是通过 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
```

负责人是通过 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
```

负责人是 Amazon QuickSight 企业版用户

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:user/<namespace>/<user-name>
```

Note

适用于*<namespace>*，则必须指定`default`.

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-east-1:111122223333:user/
default/bi_user1
```

负责人是 Amazon QuickSight 企业版组

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:group/<namespace>/<group-name>
```

Note

适用于*<namespace>*，则必须指定`default`.

例如：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:group/default/data_scientists
```

委托人是Amazon帐户

语法：

```
--principal DataLakePrincipalIdentifier=<account-id>
```

例如：

```
--principal DataLakePrincipalIdentifier=111122223333
```

委托人是组织

语法：

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:organization/<organization-id>
```

例如：

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl
```

委托人是组织单位

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
```

ALTER

权限	在此资源上授予	被授权者还需要
ALTER	DATABASE	glue:UpdateDatabase
ALTER	TABLE	glue:UpdateTable

具有此权限的委托人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。不能更改元数据表所指向的基础数据中的列。

如果正在更改的属性是注册的 Amazon Simple Storage Service (Amazon S3) 位置，则委托人必须对新位置具有数据位置权限。

Example

以下示例将ALTER向用户授权datalake_user1在数据库retail在Amazon账户 1111-2222-3333。

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

Example

以下示例将授予。ALTER给用户datalake_user1在桌子上inventory中的数据库retail.

```
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	在此资源上授予	被授权者还需要
CREATE_DATABASE	数据目录	glue:CreateDatabase

具有此权限的委托人可以在数据目录中创建元数据库或资源链接。委托人还可以在数据库中创建表。

Example

以下示例将授予。CREATE_DATABASE给用户datalake_user1在Amazon账户 1111-2222-3333。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --permissions
 "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

当委托人在数据目录中创建数据库时，不会授予对基础数据的权限。授予以下额外的元数据权限（以及向其他人授予这些权限的能力）：

- CREATE_TABLE中的数据库
- ALTER 数据库
- DROP 数据库

创建数据库时，委托人可以选择指定 Amazon S3 位置。根据委托人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。请务必记住以下三种情况。

创建数据库使用案例	需要权限
位置属性未指定。	CREATE_DATABASE已足够。
位置属性已指定，该位置不由 Lake Formation 管理（未注册）。	CREATE_DATABASE已足够。
位置属性已指定，该位置由 Lake Formation（已注册）管理。	CREATE_DATABASE是必需的，以及指定位置上的数据位置权限。

CREATE_TABLE

权限	在此资源上授予	被授权者还需要
CREATE_TABLE	DATABASE	glue:CreateTable

具有此权限的委托人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

以下示例向用户授予。datalake_user1在中创建表的权限retail中的数据库Amazon账户1111-2222-3333。

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

当委托人在数据目录中创建表时，表格上的所有 Lake Formation 权限都将授予委托人，并能够将这些权限授予其他人。

跨账户资助

如果数据库所有者帐户授予CREATE_TABLE对于收件人帐户，并且收件人账户中的用户成功在所有者帐户的数据库中创建表，则适用以下规则：

- 收件人账户中的用户和数据湖管理员拥有表格上的所有 Lake Formation 权限。他们可以向账户中的其他委托人授予对表的权限。他们不能向所有者账户或任何其他账户中的委托人授予权限。
- 所有者帐户中的数据湖管理员可以向其账户中的其他委托人授予对表的权限。

数据位置权限

当您尝试创建指向 Amazon S3 位置的表时，具体取决于您是否拥有数据位置权限，CREATE_TABLE权限可能不足以创建表。牢记以下三种情况非常重要。

创建表格用例	需要权限
指定的位置不由 Lake Formation 管理（未注册）。	CREATE_TABLE已足够。
指定的位置由 Lake Formation 管理（已注册），并且包含的数据库没有位置属性或者位置属性不是表位置位置的 Amazon S3 前缀。	CREATE_TABLE是必需的，以及指定位置上的数据位置权限。
指定的位置由 Lake Formation 管理（已注册），包含的数据库具有指向已注册位置的位置属性，并且是表位置的 Amazon S3 前缀。	CREATE_TABLE已足够。

DATA_LOCATION_ACCESS

权限	在此资源上授予	被授权者还需要
DATA_LOCATION_ACCESS	Amazon S3 位置	（Amazon S3 对该位置的权限，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。拥有此权限的委托人可以创建指向指定 Amazon S3 位置的元数据库或表。必须注册该位置。对某个位置拥有数据位置权限的委托人也具有子位置的位置权限。

Example

以下示例将数据位置授予权限s3://products/retail给用户datalake_user1在Amazon账户1111-2222-3333。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::products/retail"}}'
```

DATA_LOCATION_ACCESS不需要查询或更新基础数据。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 (p. 239).

DELETE

权限	在此资源上授予	被授权者还需要
DELETE	TABLE	（如果该位置已注册，则不需要额外的 IAM 权限。）

拥有此权限的委托人可以删除表指定的 Amazon S3 位置的底层数据。委托人还可以在 Lake Formation 控制台上查看表格，并使用Amazon GlueAPI。

Example

以下示例将DELETE向用户授权datalake_user1在桌子上inventory中的数据库retail在Amazon账户1111-2222-3333。

```
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，而不适用于 Amazon RetAmazon Relational Database Service (Amazon RDS) 等其他数据存储中的数据。

DESCRIBE

权限	在此资源上授予	被授权者还需要
DESCRIBE	表资源链接 数据库资源链接	glue:GetTable glue:GetDatabase
DESCRIBE	DATABASE	glue:GetDatabase
DESCRIBE	TABLE	glue:GetTable

具有此权限的委托人可以查看指定的数据库、表或资源链接。没有隐式授予任何其他数据目录权限，也没有隐式授予任何数据访问权限。数据库和表格出现在综合服务的查询编辑器中，但除非其他 Lake Formation 权限（例如，SELECT) 被授予。

例如，具有DESCRIBE在数据库上可以看到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，也无法在数据库中删除、更改或创建表。同样，拥有DESCRIBE在表上可以看到表和表元数据（描述、架构、位置等），但不能删除、更改或运行对表的查询。

以下是适用于的一些额外规则DESCRIBE：

- 如果用户对数据库、表或资源链接具有其他 Lake Formation 权限，DESCRIBE是隐式授予的。
- 如果用户有SELECT仅针对表的一部分列（部分）SELECT），用户只能看到那些列。
- 您不能授予DESCRIBE给在表格上有部分选择的用户。相反，您无法为以下表指定列包含列或排除列表DESCRIBE已被授予。

Example

以下示例将DESCRIBE向用户授权datalake_user1在表资源链接上inventory-link中的数据库retail在Amazon账户 1111-2222-3333。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --permissions
 "DESCRIBE" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-link"}}'
```

DROP

权限	在此资源上授予	被授权者还需要
DROP	DATABASE	glue:DeleteDatabase
DROP	TABLE	glue:DeleteTable
DROP	数据库资源链接	glue:DeleteDatabase

权限	在此资源上授予	被授权者还需要
	表资源链接	glue:DeleteTable

具有此权限的委托人可以删除数据目录中的数据库、表或资源链接。您不能将数据库上的 DROP 授予外部帐户或组织。

> Warning
>
> 删除数据库会删除数据库中的所有表。

Example

以下示例将DROP向用户授权datalake_user1在数据库retail在Amazon账户 1111-2222-3333。

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

Example

以下示例将授予。DROP给用户datalake_user1在桌子上inventory中的数据库retail.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --permissions
 "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory"}}'
```

Example

以下示例将授予。DROP给用户datalake_user1在表资源链接上inventory-link中的数据库retail.

```
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	在此资源上授予	被授权者还需要
INSERT	TABLE	（如果该位置已注册，则不需要额外的 IAM 权限。）

拥有此权限的委托人可以在表指定的 Amazon S3 位置插入、更新和读取基础数据。委托人还可以在 Lake Formation 控制台中查看表格，并使用Amazon GlueAPI。

Example

以下示例将INSERT向用户授权datalake_user1在桌子上inventory中的数据库retail在Amazon账户 1111-2222-3333。

```
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	在此资源上授予	被授权者还需要
SELECT	• TABLE	（如果该位置已注册，则不需要额外的 IAM 权限。）

拥有此权限的委托人可以在数据目录中查看表，并可以在该表指定的位置查询 Amazon S3 中的底层数据。委托人可以在 Lake Formation 控制台中查看表格，然后使用Amazon GlueAPI。如果在授予此权限时应用了列筛选，则委托人只能查看所包含列的元数据，并且只能查询所包含列的数据。

Note

集成分析服务有责任在处理查询时应用列筛选。

Example

以下示例将SELECT向用户授权datalake_user1在桌子上inventory中的数据库retail在Amazon账户1111-2222-3333。

```
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筛选（限制访问）特定列。包含列表指定可以访问的列。排除列表指定无法访问的列。在没有包含列表或排除列表的情况下，所有表列都可以访问。

的结果glue:GetTable仅返回调用方有权查看的列。集成服务，例如 Amazon Athena 和 Amazon Redshift 荣誉栏包含和排除名单。

Example

以下示例将授予。SELECT给用户datalake_user1在桌子上inventory使用包含列表。

```
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

下一个示例授予SELECT在inventory表使用排除列表。

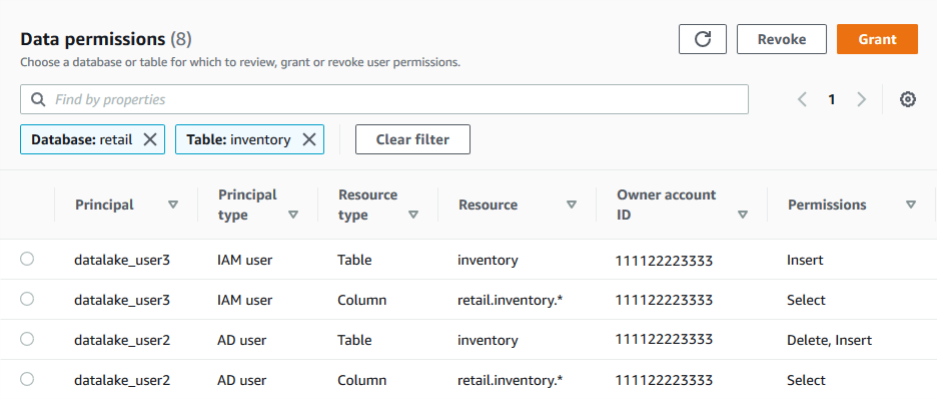
```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --permissions
 "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail", "Name":"inventory",
 "ColumnWildcard": {"ExcludedColumnNames": ["intkey", "prodcode"]}}}'
```

以下限制适用于SELECT权限：

- 授予时SELECT，如果应用了列筛选，则无法包含授权选项。
- 不能限制对属于分区键的列的访问控制。

- 拥有SELECT不能授予表中某个列子集的权限ALTER、DROP、DELETE，或者INSERT在那张桌子上的许可。同样，拥有ALTER、DROP、DELETE，或者INSERT无法授予对表的权限SELECT具有列筛选的权限。

这些区域有：SELECT权限始终显示在数据权限将 Lake Formation 控制台的页面作为单独的行。下图显示了SELECT被授予用户datalake_user2和datalake_user3在中的所有列inventory表。



Super

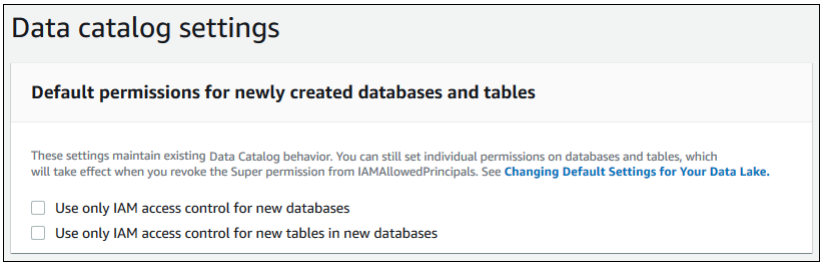
权限	在此资源上授予	被授权者还需要
Super	DATABASE	glue:*Database*
Super	TABLE	glue:*Table*, glue:*Partition*

此权限允许委托人对数据库或表执行每个受支持的 Lake Formation 操作。您不能授予Super在数据库上转移到外部账户。

此权限可以与其他 Lake Formation 权限共存。例如，您可以授予Super、SELECT, 和INSERT元数据表的权限。然后，委托人可以在表上执行所有受支持的操作。当你撤销时Super，SELECT和INSERT权限仍然存在，委托人只能执行选择和插入操作。

而不是授予Super对于个人委托人，你可以将其授予团体IAMAllowedPrincipals. 这些区域有：IAMAllowedPrincipals组将自动创建，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。何时Super被授予IAMAllowedPrincipals对于数据目录资源，对资源的访问权限完全由 IAM 策略进行有效控制。

你可以导致Super自动授予的权限IAMAllowedPrincipals通过利用设置Lake Formation 控制台的页面。



- 授予Super到IAMAllowedPrincipals对于所有新数据库，选择仅对新数据库使用 IAM 访问控制.
- 授予Super到IAMAllowedPrincipals对于新数据库中的所有新表，选择仅对新数据库中的新表使用 IAM 访问控制.

> **Note**
>
> 此选项将复选框导致复选框仅对此数据库中的新表使用 IAM 访问控制中的创建数据库默认处于选中状态的对话框。它只能做到这一点。这是中的复选框创建数据库启用授予`Super`到`IAMAllowedPrincipals`.

这些设置默认情况下，将启用页面选项。有关更多信息，请参阅下列内容：

- the section called “更改数据湖的默认安全设置” (p. 253)
- 升级Amazon GlueLake Formation 模型的数据权限 (p. 22)

Lake Formation 标签访问控制

当有大量数据目录资源时，推荐使用基于 Lake Formation 标签的访问控制 (LF-TBAC) 来授予 Lake Formation 权限的方法。LF-TBAC 比命名资源方法更具可扩展性，并且所需的权限管理开销更少。

主题

Lake Formation 标签访问控制概述

Lake Formation 基于标签的访问控制 (LF-TBAC) 与 IAM 的基于属性的访问控制 (ABAC) 配合使用，提供对数据湖资源和数据的精细访问。

> Note
>
> IAM 标签与 LF 标签不同。这些标签不可互换。LF-tags 用于授予 Lake Formation 权限，IAM 标签用于定义 IAM 策略。

什么是基于Lake Formation 标签的访问控制？

Lake Formation (Lake Formation) 是一种授权策略，该策略基于属性来定义权限。在 Lake Formation 中，这些属性被称为LF 标签. 您可以将 LF-Tags 附加到数据目录资源、Lake Formation 主体和表列。你可以使用这些 LF 标签分配和撤销对 Lake Formation 资源的权限。Lake Formation 允许在委托人的标签与资源标签匹配时对这些资源执行操作。LFEC 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。

基于 Lake Formation 标签的访问控制与基于 IAM 属性的访问控制的比较

基于属性的访问控制 (ABAC) 是一种授权策略，该策略基于属性来定义权限。在 Amazon 中，这些属性称为标签。您可以将标签附加到 IAM 资源（包括 IAM 实体（用户和角色））以及 Amazon 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 策略为您的数据湖资源定义了广泛（粗粒度）权限，例如，允许或拒绝在 Amazon S3 存储桶、前缀级别或数据库级别进行访问。有关 IABAC 的更多信息，请参阅什么是适用于的 ABAC？Amazon?中的IAM 用户指南.

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`。为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 定义对特定数据湖资源的精细权限。LFG 分配给数据目录资源（数据库、表和列），并授予委托人。具有与资源的 LF 标签相匹配的 LF-tag 的主体可以访问该资源。Lake Formation 权限是 IAM 权限的次要权限。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。

基于 Lake Formation 标签的访问控制 (LF-TBAC) 与 IAM ABAC 配合使用，为您的Lake Formation 数据和资源提供更高级别的权限。

- Lake Formation 权限随着创新扩展。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 IAB 策略与`project-access`标签提供对 Lake Formation 中特定数据库的访问权限。使用 LF-TBAC，LF-tag`Project=SuperApp`被分配给特定的表或列，并且向该项目的开发人员授予相同的 LF-tag。通过 IAM，开发人员可以访问数据库，而 LF-TBAC 权限则授予开发人员对表中特定表或列的进一步访问权限。如果向项目中添加了新表，Lake Formation 管理员只需要为新表分配标签，开发者就可以访问该表。
- Lake Formation TBAC 需要较少的 IAM 由于您使用 IAM 策略授予对 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权限，因此创建的 IAM 策略较少。
- 使用 Lake Formation TBAC，团队可以进行更改和扩展。这是因为新资源的权限根据属性自动授予。例如，如果有新开发人员加入项目，则通过将 IAM 角色与用户关联，然后向该用户分配所需的 LF 标签，可以很容易地向该开发人员授予访问权限。您无需更改 IAM 策略即可支持新项目或创建新的 LF-Tags。
- 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的访问权限，例如数据目录数据库或表。使用Lake Formation，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-tags 用于授予 Lake Formation 权限，IAM 标签用于定义 IAM 策略。

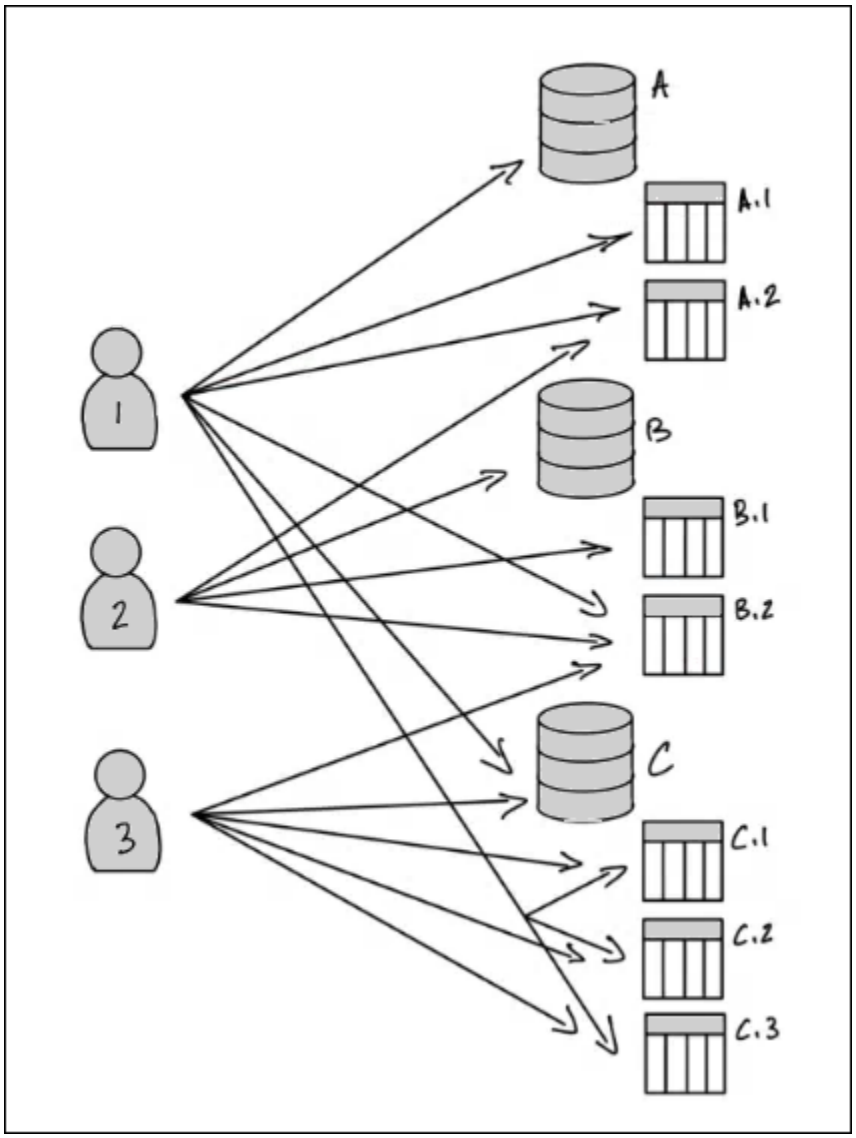
Lake Formation 标签的访问控制的工作原理

每 LFG 都是一个键/值对，例如`department=sales`要么`classification=restricted`. 一个键可以有多个已定义的值，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	任务详情
1. 定义 LF-tags 的属性和关系。	-
2. 在 Lake Formation 中创建 LF 标签。	创建 LF 标签 (p. 187)
3. 为数据目录资源分配 LF 标签。	为数据目录资源分配 LF 标签 (p. 192)
4. 授予其他委托人为资源分配 LF-tags 的权限，可以选择使用授予选项。	授予、撤销和列出 LF-tag 权限 (p. 199)
5. 向委托人授予 LF-tag 表达式，可以选择使用授予选项。	使用 LF-TBAC 方法授予数据目录权限 (p. 157)
6.（推荐）确认委托人有权通过 LF-TBAC 方法访问正确的资源后，撤消使用命名资源方法授予的权限。	-

假设数据湖管理员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获得上图所示的权限，数据湖管理员必须发出 17 次授权，如下所示（使用伪代码）。

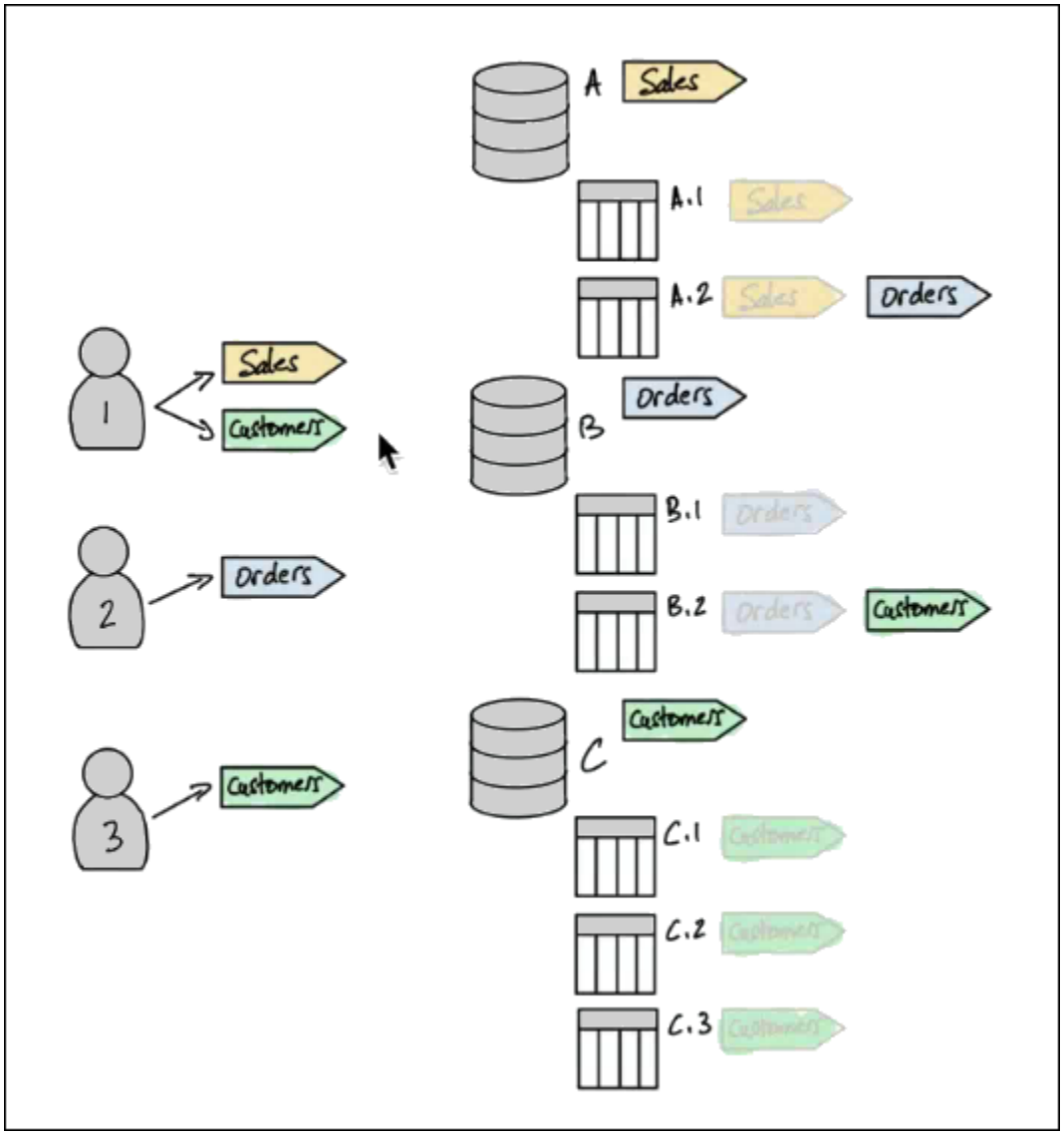
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

现在考虑一下数据湖管理员将如何使用 LF-TBAC 授予权限。下图表明数据湖管理员已为数据库和表分配了 LF-tag，并已向委托人授予了 LF-tag 的权限。

在此示例中，LF 标签表示数据湖的区域，其中包含对企业资源规划 (ERP) 应用程序套件的不同模块的分析。数据湖管理员想要控制对各种模块分析数据的访问权限。所有 LF 标签都有密钥`module`和可能的值`Sales`、`Orders`, 和`Customers`. LFormation 标签的示例如下：

```
module=Sales
```

该图仅显示了 LF-tag 值。



数据目录资源的标签分配和继承

表继承数据库的 LF 标签，列继承表中的 LF 标签。可以覆盖继承的值。在前面的示意图中，将继承变暗的 LFormation 标签。

由于继承的原因，数据湖管理员只需要对资源进行以下五个 LF-tag 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
```

标签：向校长发放补助金

为数据库和表分配 LF-tag 后，数据湖管理员只能向委托人授予四个 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，一位校长有`module=Sales`LF-tag 可以使用以下命令访问数据目录资源`module=Sales`LF-tag（例如，数据库 A），主体具有`module=Customers`LF-tag 可以通过以下方式访问资源`module=Customers`LF-tag，依此类推。

前面的授权命令不完整。这是因为尽管它们通过 LF-tags 指明了委托人拥有权限的数据目录资源，但它们并不能准确指明哪个 Lake Formation 权限（例如`SELECT`、`ALTER`）校长们拥有这些资源。因此，以下伪代码命令可以更准确地表示如何通过 LF-Tags 向数据目录资源授予 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
```

合而为一-由此产生的资源主体权限

考虑到上图中分配给数据库和表的 LF-tag，以及图中授予委托人的 LF 标签，下表列出了委托人对数据库和表拥有的 Lake Formation 权限。

主体	通过 LF 标签授予的权限
委托人 1	• `CREATE_TABLE`在数据库 A 上 • `SELECT`、`INSERT`在表 A.1 上 • `SELECT`、`INSERT`在表 A.2 上 • `SELECT`、`INSERT`在表 B.2 上 • `CREATE_TABLE`在数据库 C 上 • `SELECT`、`INSERT`在表 C.1 上 • `SELECT`、`INSERT`在表 C.2 上 • `SELECT`、`INSERT`在表 C.3 上
委托人 2	• `SELECT`、`INSERT`在表 A.2 上 • `CREATE_TABLE`在数据库 B 上 • `SELECT`、`INSERT`在表 B.1 上 • `SELECT`、`INSERT`在表 B.2 上
委托人 3	• `SELECT`、`INSERT`在表 B.2 上 • `CREATE_TABLE`在数据库 C 上 • `SELECT`、`INSERT`在表 C.1 上 • `SELECT`、`INSERT`在表 C.2 上 • `SELECT`、`INSERT`在表 C.3 上

底线

在这个简单的示例中，使用五个分配操作和八个授权操作，数据湖管理员能够指定 17 个权限。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。假设需要向每位委托人授予对每项资源的访问权限，以及在哪里n(P)是校长人数，n(R)是资源数：

- 使用命名资源方法时，所需的授权数量为n(P)×n(R).
- 使用 LF-TBAC 方法，使用单个 LF-tag，向委托人授予的权限和分配给资源的总数为n(P)+n(R).

另请参阅

- 管理用于元数据访问控制的 LF 标签 (p. 186)
- 使用 LF-TBAC 方法授予数据目录权限 (p. 157)

基于Lake Formation 标签的访问控制权限模型

要有效使用基于 Lake Formation 标签的访问控制 (LF-TBAC) 方法来保护数据湖，您必须了解以下规则和权限。

- 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委托人。

  数据工程师和分析师决定 LF-tag 的特征和关系。然后，数据湖管理员在 Lake Formation 中创建并维护 LF 标签。只有数据Lake 管理员可以对 LFED 标签执行创建、更新和删除操作。
- 您可以为数据目录资源分配多个 LF 标签。只能为特定资源分配一个特定密钥的值。

  例如，您可以分配module=Orders、region=West、division=Consumer，依此类推到数据库、表或列。无法分配module=Orders,Customers.
- 在创建资源时，您不能将LFormation 分配给资源。您只能向现有资源添加 LF-tag。
- 您可以向委托人授予 LF-tag 表达式，而不仅仅是单个 LF-tag。

  LFeg 表达式将类似于下文（在伪代码中）。

```
module=sales AND division=(consumer OR commercial)
```

  被授予此 LF-tag 表达式的委托人只能访问分配的数据目录资源（数据库、表和列）module=sales 和或者division=consumer要么division=commercial. 如果你希望委托人能够访问具有以下资源的资源module=sales 要么 division=commercial，不要将两者都包括在同一个补助金中。发放两笔补助金，一笔用于module=sales还有一个用于division=commercial.

  最简单的 LF-tag 表达式仅由一个 LF-tag 组成，例如module=sales.
- 被授予具有多个值的 LF-tag 权限的委托人可以使用其中任何一个值访问数据目录资源。例如，如果向用户授予了带有 key= 的 LF-tagmodule和有效值orders,customers，则用户有权访问分配给以下任一条件的资源module=orders要么module=customers.
- 起初，只有数据湖管理员可以为数据目录资源分配 LF 标签。数据湖管理员可以授予DESCRIBE和ASSOCIATE向委托人授予 LF-Tags 的权限，以便这些委托人可以查看和分配 LF-Tags。下表介绍这些权限。

权限	描述
DESCRIBE	对 LF-tag 拥有此权限的委托人可以在为资源分配 LF-tag 或授予 LF-Tag 权限时查看 LF-tag 及其值。你可以授予DESCRIBE在所有键值或特定值上。
ASSOCIATE	在 LFG 上拥有此权限的委托人可以将 LFG 分配给数据目录资源。授权ASSOCIATE隐含地授予DESCRIBE.

这些权限是可授权。通过授予选项被授予这些权限的委托人可以将这些权限授予其他委托人。

- 首先，数据湖管理员是唯一可以使用 LF-TBAC 方法授予数据目录资源权限（数据权限）的委托人。如果数据湖管理员使用授予选项通过 LF-TBAC 向其账户中的委托人授予数据权限，则授权接受者可以通过以下两种方式之一授予资源的数据权限：
  - 使用命名的资源方法。
  - 使用 LF-TBAC 方法，但仅使用相同的 LF-tag 表达式。

    例如，假设数据湖管理员发放了以下授权（使用伪代码）。

```
GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
 GRANT OPTION
```

    在本例中为`user1`可以授予`SELECT`使用 LF-TBAC 方法在表上传给其他主体，但只能使用完整的 LF-tag 表达式`module=customers, region=west,south`.
- 尽管数据湖管理员拥有创建、更新和删除 LF-Tags、为资源分配 LF-Tags 以及向负责人授予 LF-Tag 的隐含权限，但数据湖管理员还需要以下与 LF-TBAC 相关的权限Amazon Identity and Access Management(IAM) 权限。

```
"lakeformation:AddLFTagsToResource",
"lakeformation:RemoveLFTagsFromResource",
"lakeformation:GetResourceLFTags",
"lakeformation:ListLFTags",
"lakeformation:CreateLFTag",
"lakeformation:GetLFTag",
"lakeformation:UpdateLFTag",
"lakeformation:DeleteLFTag",
"lakeformation:SearchTablesByLFTags",
"lakeformation:SearchDatabasesByLFTags"
```

    为资源分配 LF-Tags 以及向委托人授予 LF-Tags 的委托人必须具有相同的权限，除了`CreateLFTag`、`UpdateLFTag`, 和`DeleteLFTag`权限。

    有关更多信息，请参阅 Lake Formation 角色和 IAM 权限参考 (p. 322)。
- 如果委托人同时使用 LF-TBAC 方法和命名资源方法被授予对资源的权限，则委托人对该资源的权限是两种方法授予的权限的组合。
- Lake Formation`DESCRIBE`和`ASSOCIATE`在跨账户的 LF-Tags 上，使用 LF-TBAC 方法跨账户授予对数据目录资源的权限。在两种情况下，委托人都是Amazon账户 ID。

    Note

    目前，不支持向组织和组织单位提供 LF-TBAC 跨账户补助。

    有关更多信息，请参阅 Lake Formation 中的跨账户访问 (p. 242)。

Example — LFormation 标签的生命周期

1. 数据湖管理员 Michael 创建了一个 LF-tag`module=Customers`.
2. 迈克尔·Gran`ASSOCIATE`在数据工程师爱德华多的 LF-tag 上。授权`ASSOCIATE`隐含地授予`DESCRIBE`.
3. 迈克尔·Gran`Super`在桌子上`Custs`使用授予选项给爱德华多，这样爱德华多就可以为表格分配 LF-tags。有关更多信息，请参阅 为数据目录资源分配 LF 标签 (p. 192)。
4. 爱德华多分配 LF-tag`module=customers`到桌子上`Custs`.
5. 迈克尔向数据工程师桑德拉（使用伪代码）提供以下补助。

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. 桑德拉向数据分析师玛丽亚提供以下资助。

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Maria 现在可以运行查询`Custs`桌子。

另请参阅

- 元数据访问控制 (p. 236)

Lake Formation Tagion 访问控制说明和限制

以下是基于 Lake Formation 标签的访问控制的注意事项和限制：

- 使用基于 Lake Formation 标签的访问控制 (LF-TBAC) 授予对数据目录资源的跨账户访问权限需要为您的数据目录资源策略添加内容Amazonaccount. 有关更多信息，请参阅 Lake Formation 基于标签的访问控制跨账户先决条件 (p. 244)。
- LF-tag 密钥和 LF-tag 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 LFG 标签的最大数量为 1000。
  - 可以为 LF-tag 定义的最大值数为 1000。
- 标签密钥和值在存储时全部转换为小写。
- 只能将 LF-tag 的一个值分配给特定资源。
- 如果通过一次授权向委托人授予多个 LF-Tag，则该委托人只能访问具有所有 LF 标签的数据目录资源。
- Amazon GlueETL 作业需要完整的表访问权限。如果出现以下情况，则作业将失败Amazon GlueETL 角色无权访问表中的所有列。可以在列级别应用 LF-tags，但这可能会导致Amazon GlueETL 角色将失去表的完整访问权限并导致作业失败。使用数据筛选器进行列和/或行筛选不受此限制的影响。
- 如果 LF-tag 表达式求值结果只能访问表列的子集，但在存在匹配项时授予的 Lake Formation 权限是需要完整列访问权限的权限之一，即`ALTER`、`DROP`、`INSERT`，或者`DELETE`，则这些权限均未被授予。相反，只有`DESCRIBE`GRANTED 如果授予的权限是`ALL`(`Super`)，那么只有`SELECT`和`DESCRIBE`GRANTED
- 支持对 LF-tags 授予跨账户权限，但仅限于向外部授予Amazon账户，不分配给企业或企业单位。

管理用于元数据访问控制的 LF 标签

要使用基于 Lake F-tag，您可以创建 Lake F-tag，您可以创建 LF-tag，分配给资源和授予委托人，您可以创建 LF-tag，分配给资源和授予委托人。

只有数据湖管理员才能创建、更新和删除 LF-tag。起初，只有数据湖管理员可以为数据目录资源分配 LF 标签。数据湖管理员可以向其他委托人授予权限，以便为资源分配 LF 标签。

您可以使用管理 LF-tagAmazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

主题

另请参阅

创建 LF 标签

所有 LF-tags 都必须在 Lake Formation 中定义，然后才能使用。LF-tag 由一个键和一个或多个可能的键值组成。只有数据湖管理员才能创建 LF-tag。

您可以使用下的 LF-tagAmazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

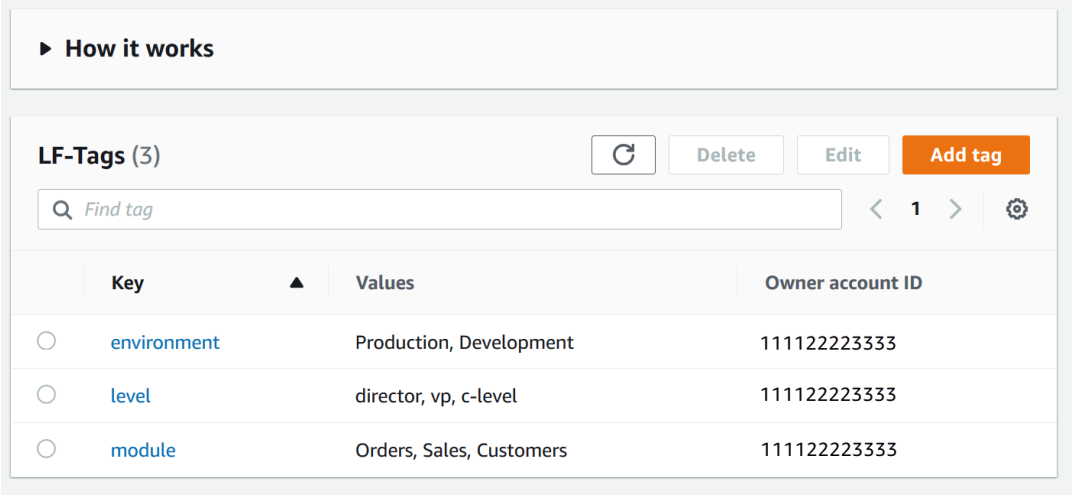
Console

创建 LF-tag

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. 在导航窗格中的下Permissions (权限)，选择LF 标签.

   这些区域有：LF 标签此时将显示页面。



3. 选择 Add tag (添加标签)。
4. 在添加 LF-tag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，请输入以逗号分隔的列表，然后按Enter，或者一次输入一个值然后选择Add在每一个之后。允许的最大值为 15。

5. 选择 Add tag (添加标签)。

Amazon CLI

创建 LF-tag

- 输入`create-lf-tag`命令。

  以下示例创建带键的 LF-tag。`module`和有效值`Customers`和`Orders`.

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

更新 LF 标签

您可以通过添加或删除允许的密钥值来更新 LF-tag。您不能更改 LF-tag 键。要更改密钥，请删除 LF-tag，然后使用所需密钥添加一个。

删除 LF-tag 值时，不会检查任何数据目录资源上是否存在该 LF-tag 值。如果删除的 LF-tag 值与某个资源相关联，则该值在该资源中不再可见，并且被授予该键值对权限的任何委托人也不再拥有该权限。

在删除 LF-tag 值之前，您可以选择使用`remove-lf-tags-from-resource`命令 (p. 196)命令从具有要删除的值的 Data Catalog 资源中移除 LF-tag，然后使用要保留的值重新标记该资源。

只有数据湖管理员才能更新 LF-tag。

您可以使用以下命令更新 LF-tagAmazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

Console

更新 LF-tag (控制台)

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. 在导航窗格中的下Permissions (权限)，选择LF 标签.
3. 在存储库的LF 标签页面，选择 LF-tag，然后选择编辑.
4. 在编辑 LF-tag对话框中，添加或删除 LF-tag 值。

   要添加多个值，请在值字段中，要么输入逗号分隔列表，然后按Enter，或者一次输入一个值或者选择Add在每一个之后。
5. 选择Save（保存）。

Amazon CLI

更新 LF-tag (Amazon CLI)

- 输入`update-lf-tag`命令。提供以下参数之一或提供两个参数：
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

以下示例替换了值vp带上的vice-president对于 LF-tag 键level.

```
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。不检查数据目录资源上是否存在 LF-tag。如果已删除的 LF-tag 与某个资源相关联，则该资源不再可见该标签，任何被授予该 LF-tag 权限的委托人也不再拥有该权限。

在删除 LF-tag 之前，您可以选择使用remove-lf-tags-from-resource (p. 196)命令从所有资源中删除 LF-tag。

只有数据湖管理员才能删除 LF-tag。

您可以使用以下命令删除 LF-tagAmazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

Console

删除 LF-tag (控制台)

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. 在导航窗格中的下Permissions (权限)，选择LF 标签.
3. 在存储库的LF 标签页面，选择 LF-tag，然后选择Delete.
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-tag 密钥值，然后选择Delete.

Amazon CLI

删除 LF-tag (Amazon CLI)

- 输入delete-lf-tag命令。提供要删除的 LF-tag 的密钥。

  Example

  以下示例删除带有键的 LF-tag。region.

  ```
  aws lakeformation delete-lf-tag --tag-key region
  ```

列出 LF 标签

你可以列出你拥有的 LF 标签DESCRIBE要么ASSOCIATE权限开的。每个 LF-tag 密钥列出的值是您拥有权限的值。

数据湖管理员可以看到在本地定义的所有 LF-tagAmazon账户和所有 LF 标签DESCRIBE和ASSOCIATE已从外部账户向本地账户授予权限。数据湖管理员可以看到所有 LF 标签的所有值。

您可以使用以下命令列出 LF 标签Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

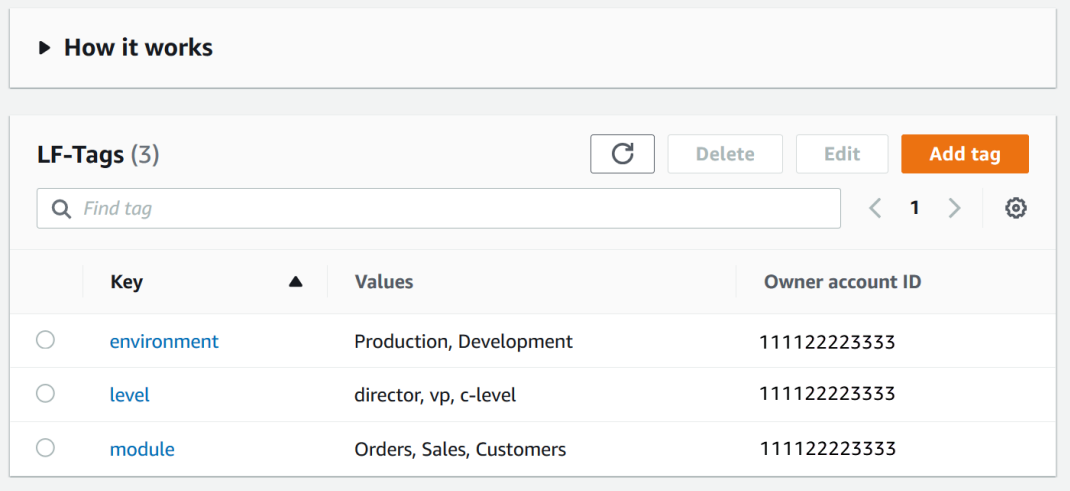
Console

列出 LF-tag (控制台)

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员或已获得 LF-Tags 权限且具有`lakeformation:ListLFTags`IAM 权限。
2. 在导航窗格中的下Permissions (权限)，选择LF 标签.

   这些区域有：LF 标签此时将显示页面。



   Check所有者账户 ID列以确定从外部账户与您的账户共享的 LF 标签。

Amazon CLI

列出 LF 标签 (Amazon CLI)

- 以数据湖管理员或已获得 LF-Tags 权限且具有`lakeformation:ListLFTags`IAM 权限。

```
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。注意`NextToken`key，表示还有更多要列出。

```
{
    "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`参数用于查看从外部账户授予的所有剩余本地 LF-tag 和 LF 标签。来自外部账户的 LF-tag 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 来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'
)
```

该命令返回一个dict具有以下结构的对象：

```
{
    '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 (p. 322)。

为数据目录资源分配 LF 标签

您可以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 标签的委托人（以及使用命名资源方法被授予访问权限的委托人）才能访问资源。

如果表继承了数据库的 LF-tag，或者某列继承了表中的 LF-tag，则可以通过为 LF-tag 键分配新值来覆盖继承的值。

您可以分配给资源的最大 LF-tag。

主题

- 管理分配给资源的标签的要求 (p. 192)
- 为表列分配 LF 标签 (p. 193)
- 分配给数据目录资源 LF-tag 分配给数据目录资源 (p. 194)
- 更新资源的 LF 标签 (p. 196)
- 从资源中删除 LF-tag (p. 196)

管理分配给资源的标签的要求

要为数据目录资源分配 LF-tag，您必须：

- Lake FormationASSOCIATELF-tag 的权限。
- 有 IAMlakeformation:AddLFTagsToResource权限。
- 有胶水：GetDatabase Glue 数据库的权限。
- 成为资源所有者（创建者），拥有Super使用 Lake Formation 获得资源许可GRANT选项，或者拥有以下权限GRANT选项：

- 对于相同的数据库Amazon账户:`DESCRIBE`、`CREATE_TABLE`、`ALTER`, 和`DROP`
- 对于外部账户中的数据库：`DESCRIBE`、`CREATE_TABLE`和`ALTER`
- 对于表（和列）：`DESCRIBE`、`ALTER`、`DROP`、`INSERT`、`SELECT`, 和`DELETE`

此外，LF-tag 和分配给它的资源必须相同Amazonaccount.

要从数据目录资源中删除 LF-tag，您必须满足这些要求，并具有`lakeformation:RemoveLFTagsFromResource`IAM 权限。

为表列分配 LF 标签

为表列分配 LF 标签（控制台）

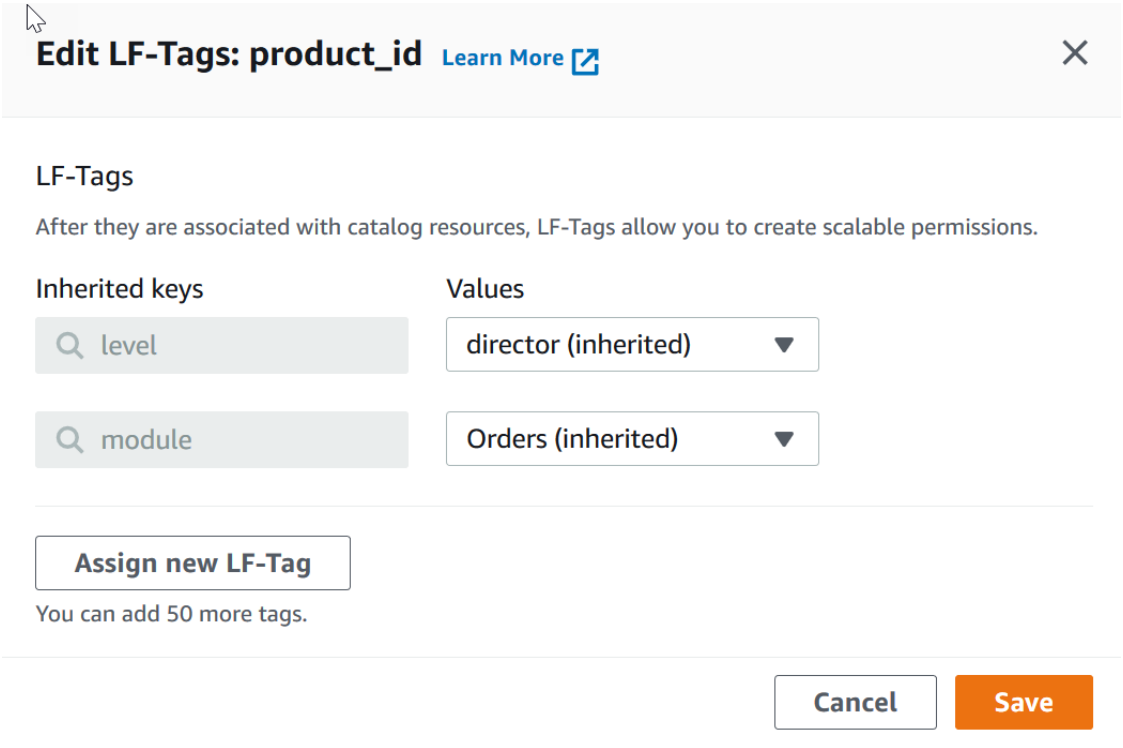
1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以符合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（不是表名称旁边的选项按钮）。
4. 在表详细信息页面上的下，架构部分，选择编辑架构.
5. 在存储库的编辑架构页面，选择一列或多列，然后选择编辑标签.

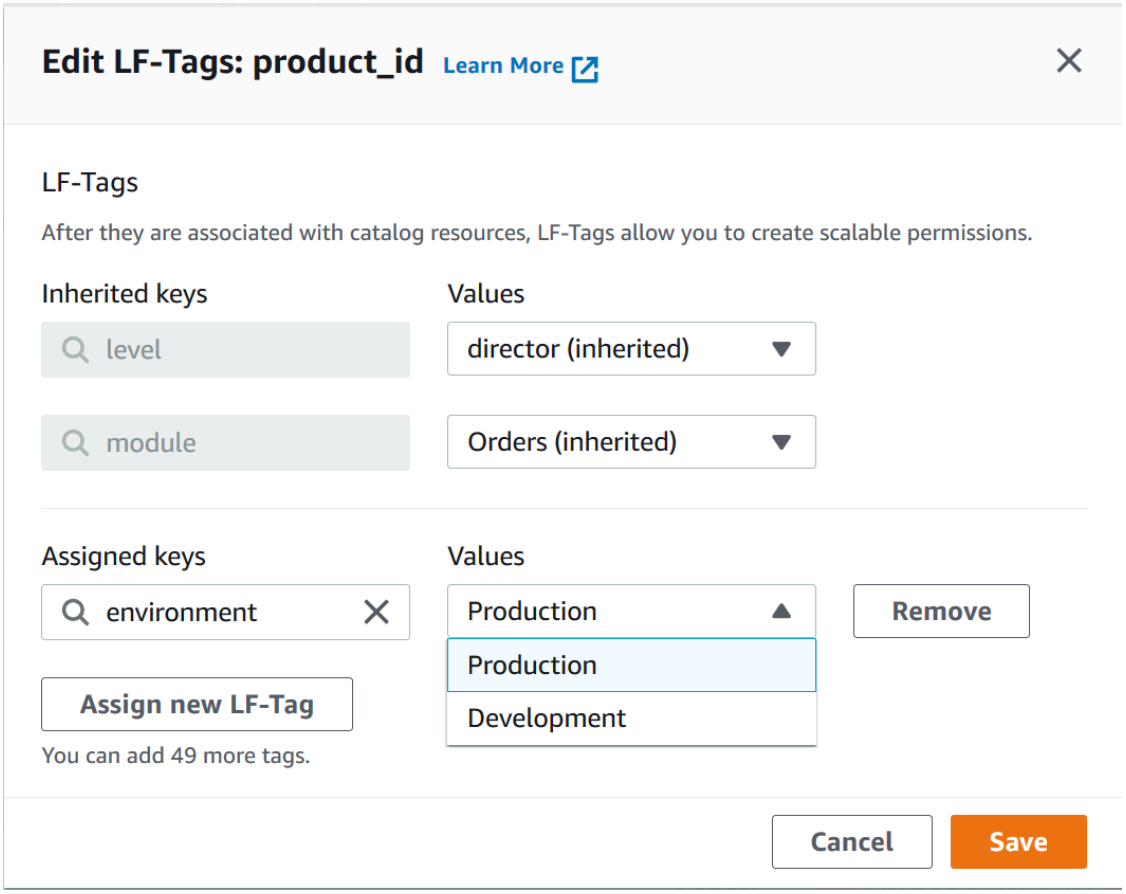
   Note

   如果您打算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。

   这些区域有：编辑 LF 标签出现对话框，显示从表中继承的所有 LF 标记。



6. （可选）对于值旁边的列表继承的键字段中，选择一个值来覆盖继承的值。
7. (可选)分配新的 LF-tag. 然后对于分配的密钥，选择一个键，对于值，为键选择一个值。



8. (可选)分配新的 LF-tag再次添加另一个 LF-tag。
9. 选择Save（保存）。

分配给数据目录资源 LF-tag 分配给数据目录资源

Console

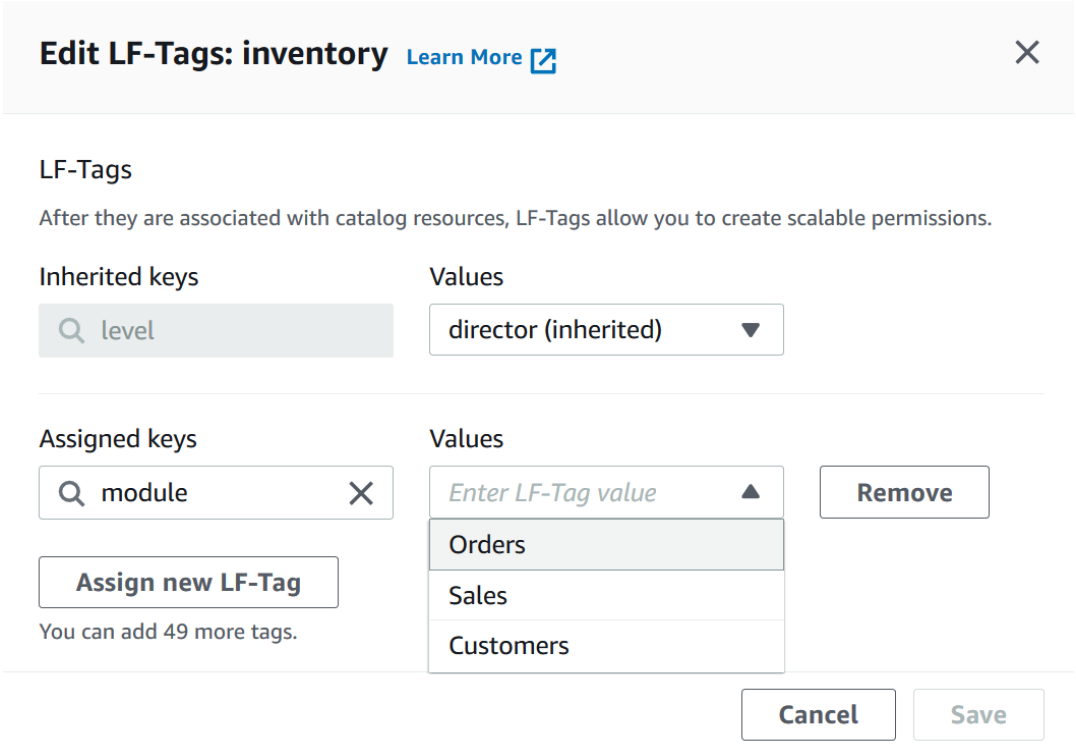
为数据目录数据库或表分配 LF 标签

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以符合前面列出的要求的用户身份登录。
2. 在导航窗格中的下Data Cat，请执行以下操作之一：
   - 要为数据库分配 LF 标签，请选择数据库.
   - 要为表分配 LF 标签，请选择表.
3. 选择一个数据库或表，然后在操作菜单，选择编辑标签.

   这些区域有：编辑 LF 标签：####此时将显示对话框。

   如果表继承了其包含数据库的 LF-tags，则该窗口将显示继承的 LF 标签。否则，它会显示文本 “没有与资源关联的继承的 LF 标签”。



4. （可选）如果表继承了 LF 标签，对于值旁边的列表继承的键字段中，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-tag.
   b. 在分配的密钥字段中，选择一个 LF-tag 密钥，然后在值字段中，选择一个值。
   c. (可选)分配新的 LF-tag再次分配一个额外的 LF-tag。
6. 选择Save（保存）。

Amazon CLI

为数据目录资源分配 LF 标签

- 运行 `add-lf-tags-to-resource` 命令。

  以下示例分配 LF-tag`module=orders`到桌子上`orders`在数据库中`erp`. 它使用快捷语法`--lf-tags`参数。这些区域有：`CatalogID`的财产`--lf-tags`为可选项。如果未提供，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
```

  如果此命令成功，则输出如下。

```
{
    "Failures": []
}
```

下一个示例将两个 LF 标签分配给sales表，并使用 JSON 语法作为--lf-tags参数。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":
 "module","TagValues": ["sales"]},{"TagKey": "environment","TagValues":
 ["development"]}]'
```

下一个示例分配 LF-taglevel=director到total表中的列sales.

```
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

要更新数据目录资源的 LF-tag，请执行以下操作，请执行以下操作：Amazon CLI）

- 使用add-lf-tags-to-resource命令，如前面的步骤所述，如前面的步骤所述。

  添加具有与现有 LF-tag，以更新现有值。

从资源中删除 LF-tag

删除数据目录资源的 LF-tag，以删除数据目录资源的 LF-tagAmazon CLI）

- 运行 remove-lf-tags-from-resource 命令。

  如果表的 LF-tag 值覆盖从父数据库继承的值，则从表中删除 thatLF-tag 将恢复继承的值。此行为也适用于覆盖从表继承的键值的列。

  以下示例删除了 LF-tag。level=director来自 的totalcolumnsales桌子。这些区域有：CatalogID的财产--lf-tags为可选项。如果未提供，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
```

查看分配给资源分配给资源的 LF-tag

您可以查看分配给数据目录资源的 LF 标签。您必须具有DESCRIBE要么ASSOCIATE在 LF-tag 上查看它的权限。

Console

查看分配给资源的 LF-tag，以查看分配给资源的 LF-tag

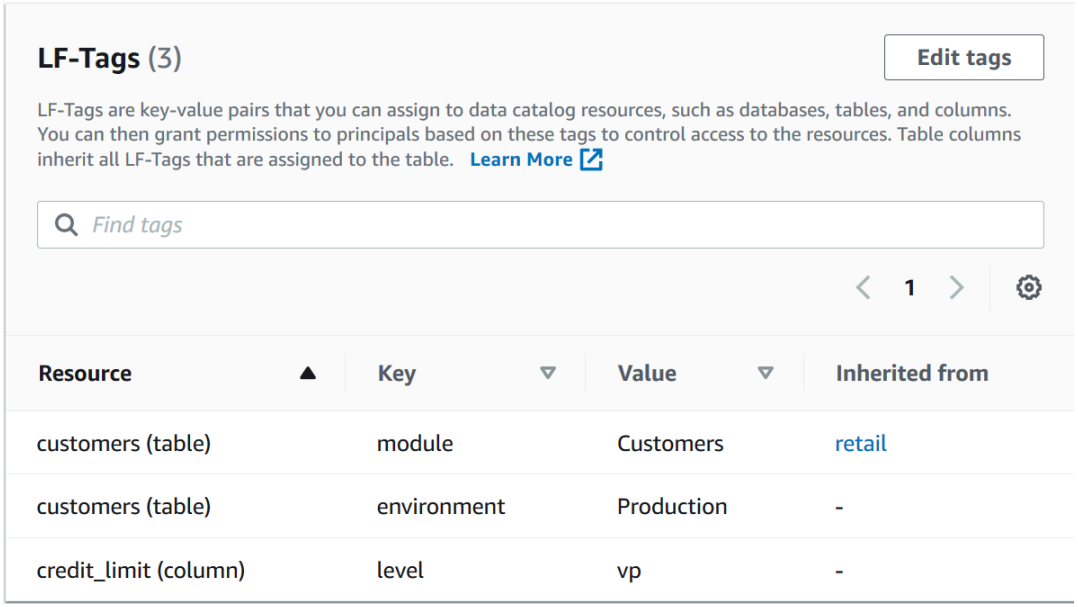
1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员、资源所有者或已获得 Lake Formation 资源权限的用户身份登录。
2. 在导航窗格中的标题下，位于标题下，在Data Cat，请执行以下操作之一：

- 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. 在存储库的表要么数据库页面上，请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。

   以下屏幕截图显示分配给 LF-tag 分配给`customers`表，它包含在`retail`数据库。这些区域有：`module`LF-tag 继承自数据库。这些区域有：`credit_limit`列中有`level=vp`已分配 LF-tag。



Amazon CLI

要查看分配给资源的 LF-tag，请执行以下操作：Amazon 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，不显示继承的 LF 标签。如果你想查看所有列上的所有 LF 标签，包括继承的 LF 标签，请省略`--show-assigned-lf-tags`选项。

查看 LF-tag 分配给的资源

您可以查看分配特定 LF-Tag 密钥的所有数据目录资源。要执行该操作，您需要以下Lake Formation 权限：

- `DESCRIBE`要么`ASSOCIATE`在LF-tag 上。
- `DESCRIBE`或资源上的任何其他Lake Formation 许可。

此外，您需要满足以下条件Amazon Identity and Access Management(IAM) 权限：

- `lakeformation:SearchDatabasesByLFTags`
- `lakeformation:SearchTablesByLFTags`

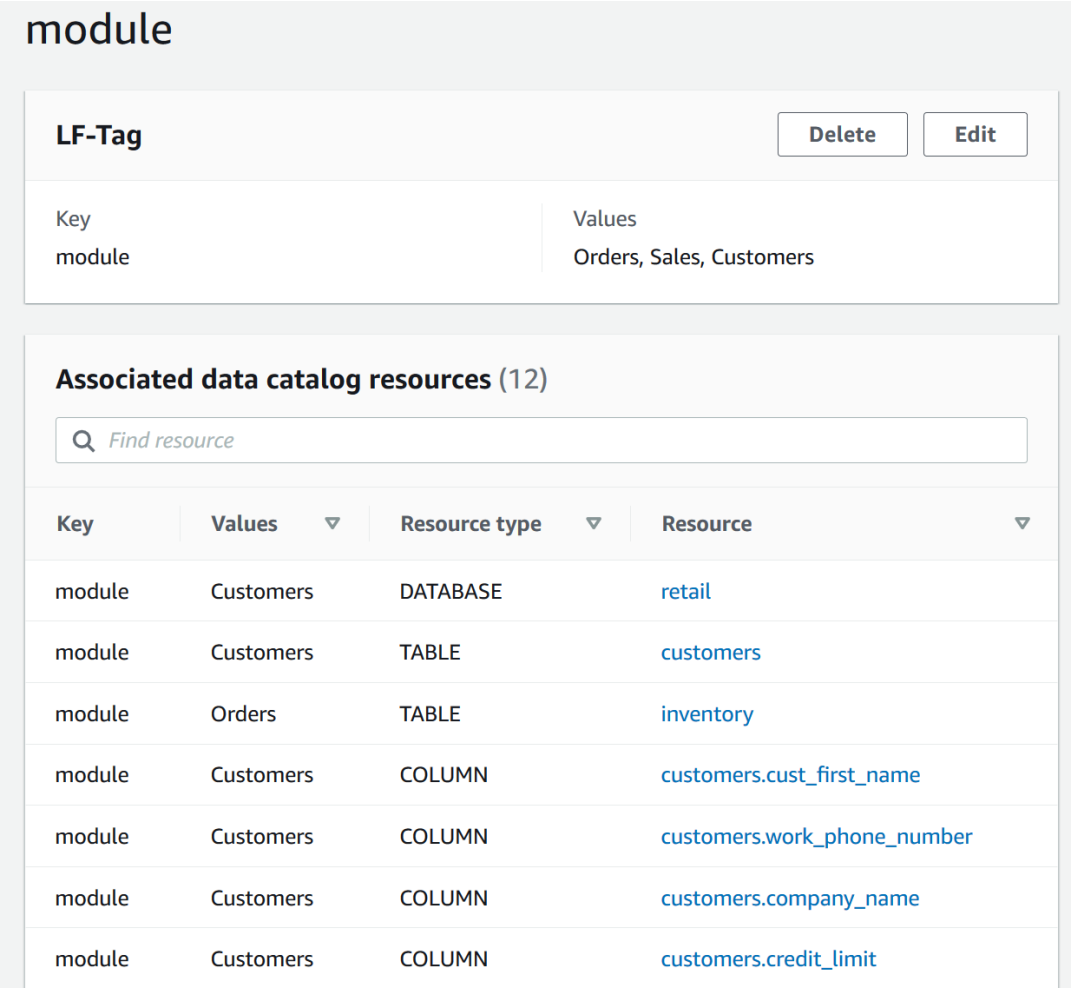
Console

查看 LF-tag 分配给的资源（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 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. 在导航窗格中的标题下，位于标题下，位于标题下，Permissions (权限)和管理角色和任务，选择LF标签.
3. 选择 LF-tag 密钥（不是密钥名称旁边的选项按钮）。

   LF-tag 详细信息页面显示已分配 LF-tag 的资源列表。



Amazon CLI

查看 LF-tag 分配给的资源

- 运行`search-tables-by-lf-tags`要么`search-databases-by-lf-tags`命令。

  Example

  以下示例列出了具有以下特点的表和列`level=vp`已分配 LF-tag。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 (p. 322)。

授予、撤销和列出 LF-tag 权限

你可以授予`DESCRIBE`和`ASSOCIATE`Lake Formation 对 Lf-Tags 的权限授予委托人，以便他们可以查看 LF-Tags 并将其分配给数据目录资源（数据库、表和列）。将 LF-Tags 分配给数据目录资源时，您可以使

用基于 Lake Formation 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 标签访问控制 (p. 179)。

起初，只有数据湖管理员可以授予这些权限。如果数据湖管理员使用授予选项授予这些权限，则其他委托人可以授予这些权限。这些区域有：DESCRIBE和ASSOCIATE权限在中进行了解释基于Lake Formation 标签的访问控制权限模型 (p. 184).

你可以授予DESCRIBE和ASSOCIATE对外部的 LF-tag 的权限Amazonaccount. 然后，该账户中的数据湖管理员可以将这些权限授予账户中的其他委托人。外部账户中的数据湖管理员向其授予的委托人ASSOCIATE然后，权限可以将 LF-Tags 分配给您与其帐户共享的数据目录资源。

授予外部账户时，必须包括授予选项。

您可以通过使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

主题

- 使用控制台列出 lf-tag 权限 (p. 200)
- 使用控制台授予 lf-tag 权限 (p. 200)
- 使用Amazon CLI (p. 203)

有关更多信息，请参阅 管理用于元数据访问控制的 LF 标签 (p. 186)和Lake Formation 标签访问控制 (p. 179).

使用控制台列出 lf-tag 权限

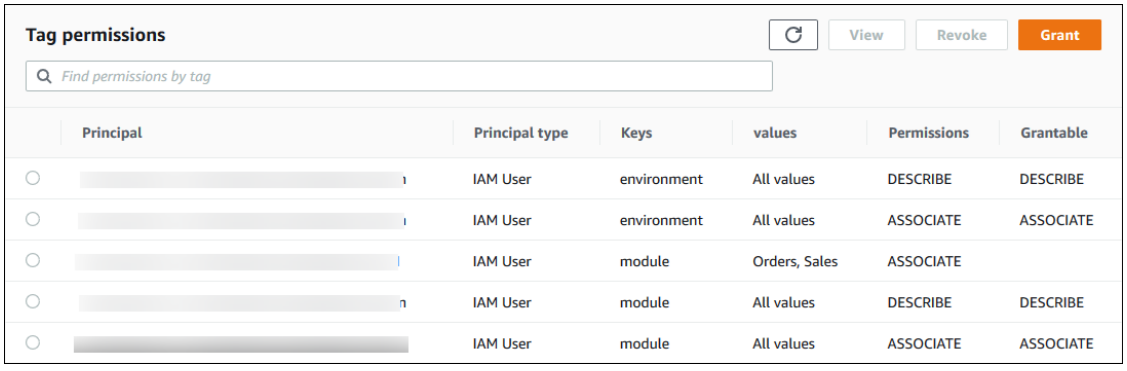
您可以使用 Lake Formation 控制台查看授予的 LF-Tags 权限。你必须是数据湖管理员或者拥有DESCRIBE要么ASSOCIATE允许 LF-tag 查看它。

列出 LF-tag 权限（控制台）

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员或用户身份登录ASSOCIATE要么DESCRIBE已授予 LF-Tags 的权限。
2. 在导航窗格中的下Permissions (权限)，选择标签权限.

   这些区域有：标签权限此时将显示页。



使用控制台授予 lf-tag 权限

以下步骤说明如何使用授予 LF-Tags 权限：授予标签权限在 Lake Formation 控制台上显示页面。该页面分为以下部分：

- 委托人— 用户、角色或Amazon要向其授予权限的帐户。

- LF-标签— 要授予权限的 LF-Tags。
- Permissions (权限)— 授予的权限。

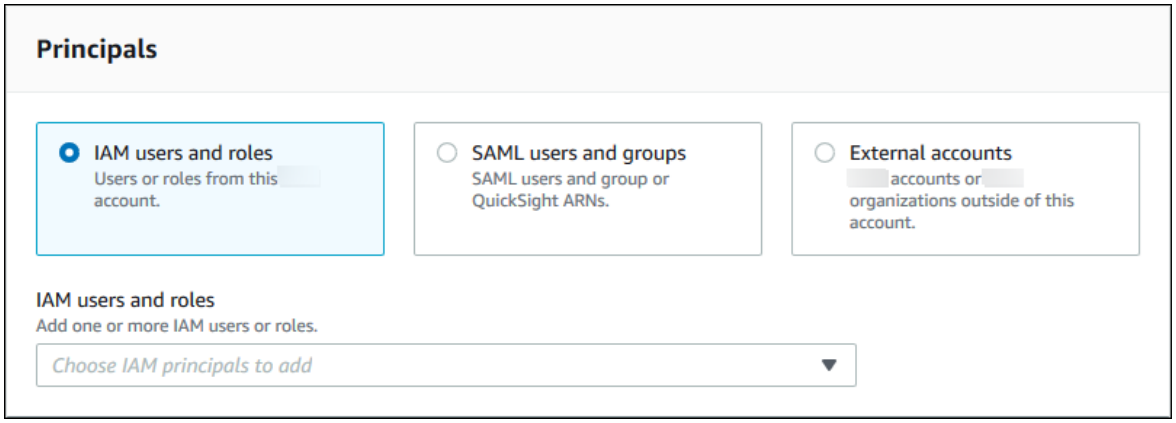
打开授予标签权限页

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以数据湖管理员或用户身份登录ASSOCIATE要么DESCRIBE权限 OnLF-Tags 已通过GRANT选项。
2. 在导航窗格中的下Permissions (权限)，选择标签权限.
3. 选择 Grant（授权）。

指定委托人

在委托人部分中，选择委托人类型并指定要向其授予权限的承担者。



IAM 用户和角色

从中选择一个或多个用户或角色IAM 用户和角色列表。

SAML 用户和组

适用于SAML 和 Amazon QuickSight 用户和组中，为通过 SAML 联合的用户或组输入一个或多个亚马逊资源名称 (ARN)，或者为 Amazon QuickSight 用户或组输入 ARN。按Enter在每个 ARN 之后。

有关如何构建 ARN 的信息，请参阅Lake Formation 补助金和撤销Amazon CLI命令 (p. 168).

Note

仅 Amazon QuickSight 企业版支持 Lake Formation 与 Amazon QuickSight 的集成。

外部账户

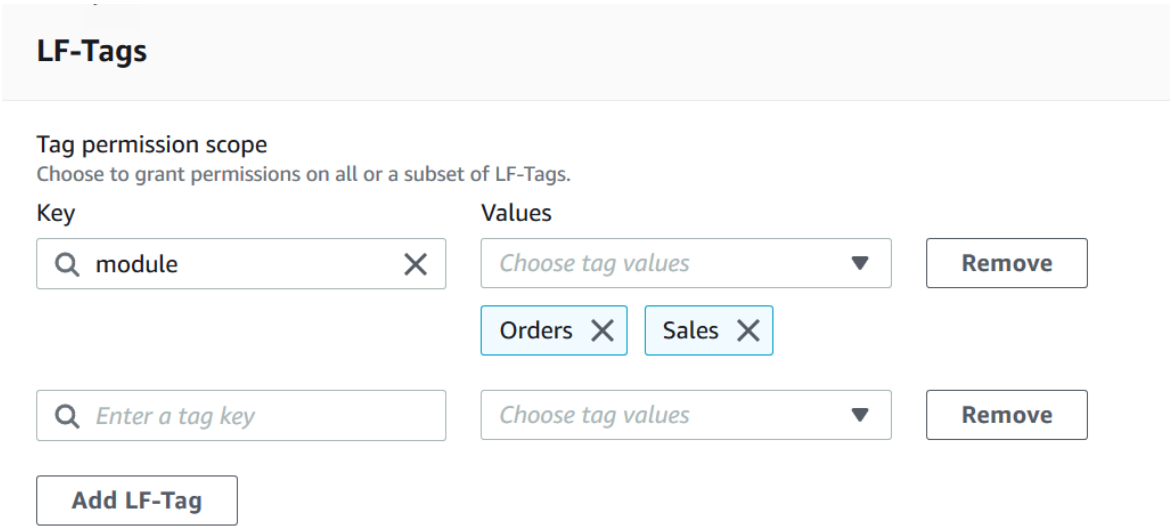
适用于Amazon帐户，输入一个或多个有效Amazon账户 ID。按Enter在每个身份证之后。

Note

目前，LF-TBAC 不支持向组织和组织单位授予 LF-Tag 权限。

指定 LF 标签

在LF-标签部分中，指定要授予权限的 LF-Tags。



1. 选择添加 lf-tag以显示用于指定 LF-Tag 的第一行字段。
2. 将光标放在密钥字段中，可选择开始键入以缩小选择列表范围，然后选择 LF-tag 键。
3. 在值列表中，选择一个或多个值，然后按Tab或者单击或点击字段外部以保存选定的值。

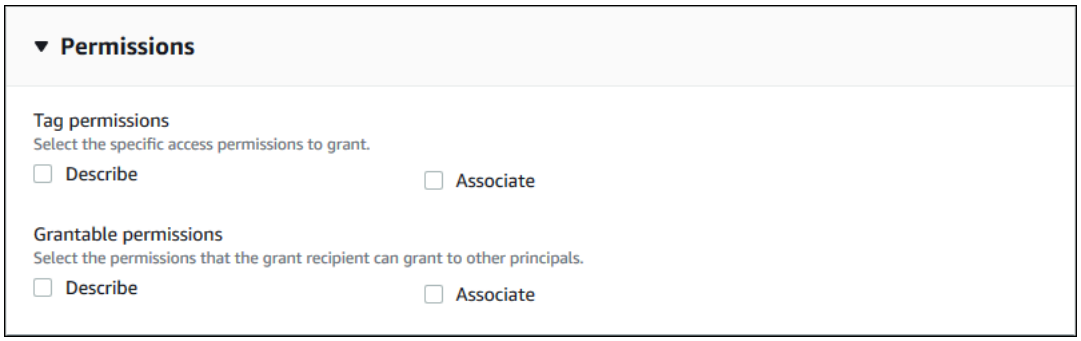
   Note

   如果中的其中一行值列表有焦点，按Enter选中或清除复选框。

   选定的值显示为磁贴下方值列表。选择 ✖ 以删除一个值。选择Remove删除整个 LF 标签。
4. 要添加另一个 LF 标签，请选择添加 lf-tag再说一遍，然后重复前两个步骤。

指定权限

在Permissions (权限)部分中，选择权限和可授权的权限。



1. 下标签权限中，选择要授予的权限。

   授予关联性隐含地授予描述.
2. （可选）下可授权限中，选择授权收件人可以在其中授予其他委托人的权限Amazonaccount.
3. 选择 Grant（授权）。

使用Amazon CLI

您可以通过使用Amazon Command Line Interface(Amazon CLI)。

要列出 lf-tag 权限 (Amazon CLI)

- 输入`list-permissions`命令。你必须是数据湖管理员或者拥有`DESCRIBE`要么`ASSOCIATE`允许 LF-tag 查看它。

  以下命令请求您拥有权限的所有 LF-Tags。

```
aws lakeformation list-permissions --resource-type LF_TAG
```

  以下是数据湖管理员的示例输出，该管理员可以看到授予所有委托人的所有 LF-Tag。非管理员用户只能看到授予他们的 LF 标签。从外部账户授予的 LF-tag 权限将显示在单独的结果页面上。要查看它们，请重复该命令并提供`--next-token`参数中返回的标记从上次命令运行中返回的令牌。

```
{
    "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-tag 密钥的所有授权。以下命令返回对 lf-tag 授予的所有权限module.

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

您还可以列出授予特定 LF-Tag 的特定委托人的 LF-Tag 值。当提供--principal参数，你必须提供--resource参数。因此，该命令只能有效地请求授予特定委托人的值以获得特定 LF-tag 密钥。以下命令说明如何为委托人执行此操作：datalake_user1和 lf-tag 密钥module.

```
aws lakeformation list-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

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": []
        }
    ]
}
```

要授予 LF-Tags 的权限 (Amazon CLI)

- 输入类似以下的命令。此示例授予用户datalake_user1这ASSOCIATE使用密钥对 LF-tag 的权限module. 它授予查看和分配该密钥的所有值的权限，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权限隐式授予DESCRIBE权限。

下一个示例授予ASSOCIATE到外部Amazon带钥匙的 LF 标签上的账户 1234-5678-9012module，带有赠款选项。它授予仅查看和分配值的权限sales和orders.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=123456789012 --permissions "ASSOCIATE"
 --permissions-with-grant-option "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
```

撤消对 LF-Tags 的权限 (Amazon CLI)

- 输入类似以下的命令。此示例撤消了ASSOCIATE使用密钥对 LF-tag 的权限module来自用户datalake_user1.

```
aws lakeformation revoke-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Lake Formation 中的数据过滤和细胞级安全

授予 Lake Formation 对数据目录表的权限时，可以在查询结果和与 Lake Formation 集成的引擎中加入数据筛选规范以限制对某些数据的访问。Lake Formation 使用数据筛选来实现列级安全性、行级安全性和单元级安全性。

主题

数据筛选概述

利用 Lake Formation 的数据筛选功能，您可以实现以下级别的数据安全。

列级安全性

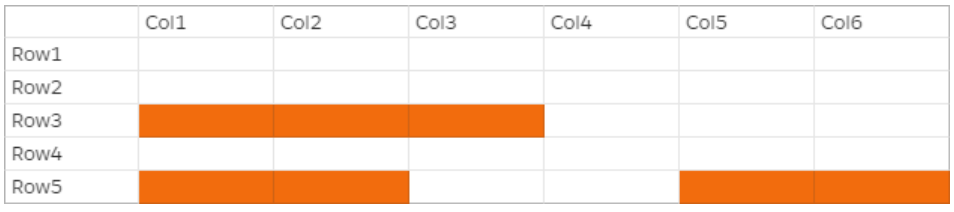
授予对具有列级安全性（列筛选）的 Data Catalog 表的权限，允许用户仅查看他们在表中具有访问权限的特定列。考虑使用一个`persons`表，在一家大型多区域通信公司的多个应用程序中使用。通过列筛选授予对数据目录表的权限可以限制不在 HR 部门工作的用户查看个人身份信息 (PII)，例如社会安全号码或出生日期。

行级别安全性

授予对具有行级安全性（行筛选）的数据目录表的权限，允许用户仅查看他们在表中有权访问的特定数据行。筛选基于一个或多个列的值进行筛选。例如，如果通信公司的不同地区办事处有自己的 HR 部门，则可以将 HR 员工可以查看的人员记录限制为仅限其所在区域的员工记录。

单元级安全性

单元格级别的安全性结合了行筛选和列筛选，以实现高度灵活的权限模型。如果以网格形式查看表的行和列，则通过使用单元格级别的安全性，可以限制对两个维度中任何位置的网格单个元素（单元格）的访问。也就是说，您可以根据行限制对不同列的访问。下方的关系图阐述了这一点，其中受限制的列进行阴影处理。



继续以 persons 表为例，您可以创建数据筛选器位于单元格级别，如果行的 country 列设置为 “UK”，则限制访问街道地址列；但如果行的 country 列设置为 “US”，则允许访问街道地址列。

筛选器仅适用于读取操作。因此，您只能授予SELECT带过滤器的Lake Formation 许可。

Lake Formation 中的数据过滤器

您可以通过以下方式实现列级、行级和单元级安全性：数据筛选器. 在授予SELECT桌子上的Lake Formation 许可。

每个数据筛选器都属于数据目录中的特定表。数据筛选器包括以下信息：

- 筛选器名称
- 与筛选器关联的表的目录 ID
- 表名称
- 包含该表的数据库的名称
- 列规范 — 要在查询结果中包括或排除的列的列表。
- 行筛选器表达式 — 指定要包含在查询结果中的行的表达式。有一些限制，表达式的语法为WHEREPartiQL 语言中的子句。要指定所有行，请输入true使用控制台中的AllRowsWildcard在 API 调用中。

  有关行筛选器表达式中支持的内容的更多信息，请参阅行筛选器表达式中的 PartiQL 支持 (p. 209).

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 “all columns” 通配符并提供行筛选器表达式，则仅建立行级安全性（行筛选）。
- 如果包括或排除特定列，并使用全行通配符指定 “所有行”，则仅建立列级安全性（列筛选）。
- 如果包括或排除特定列，同时提供行筛选器表达式，则是在建立单元格级别的安全性（单元格过滤）。

以下来自 Lake Formation 控制台的屏幕截图显示了一个执行单元级过滤的数据筛选器。针对的查询orders表，它限制了对customer_name列在行中product_type列包含 “pharma”。在受限行中，查询结果返回NULL(对于)customer_namecolumn.

Create data filter

Data filter name
Enter a name that describes this data access filter.

restrict-pharma

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.

Target database
Select the database that contains the target table.

Choose databases | Load more

sales
054881201579

Target table
Select the table for which the data filter will be created.

Choose tables | Load more

orders
054881201579

Column-level access
Choose whether this filter should have column-level restrictions.

- ○ Access to all columns
  Filter won't have any column restrictions.
- ○ Include columns
  Filter will only allow access to specific columns.
- ◉ Exclude columns
  Filter will allow access to all but specific columns.

Select columns

Choose one or more columns

customer_name
string

Row filter expression

请注意使用单引号将字符串文字括起来，`'pharma'`.

您可以使用 Lake Formation 控制台创建此数据过滤器，也可以将以下请求对象提供给`CreateDataCellsFilter`API 操作。

```
{
     "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进行任意数量的数据筛选器。要执行该操作，您需要`SELECT`在表上使用授予选项的权限。默认情况下，数据湖管理员有权创建数据筛选器在该账户的所有表上。在向委托人授予对表的权限时，通常只使用可能的数据筛选器子集。例如，您可以为`orders`那是一张表 row-security-only 数据筛选器。参考前面的屏幕截图，你可以选择访问所有列选项并包含一个行筛选器表达式`product_type<>pharma`. 此数据筛选器的名称可能是`no-pharma`. 它限制对所有具有`product_type`列设置为 “药房”。

的请求对象`CreateDataCellsFilter`此数据筛选器的 API 操作如下。

```
{
     "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"]
}
```

然后你可以授予`SELECT`在`orders`表中带有`restrict-pharma`管理用户的数据筛选器，以及`SELECT`在`orders`表中带有`no-pharma`非管理用户的数据筛选器。对于医疗保健领域的用户，你可以授予`SELECT`在`orders`表具有对所有行和列的完全访问权限（无数据筛选器），或者可能使用另一个限制访问定价信息的数据筛选器。

另请参阅

- 管理数据筛选器 (p. 212)

行筛选器表达式中的 PartiQL 支持

您可以使用 PartiQL 数据类型、运算符和聚合的子集来构造行筛选器表达式。Lake Formation 不允许在过滤器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进行比较（例如`views >= 10000`），但您无法将列与其他列进行比较。

行筛选器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小于 2048 个字符。

简单表达式

一个简单的表达式将采用以下格式： `<column name > <comparison operator ><value >`

- 列名称

  它必须是表架构中存在的顶级数据列或分区列，并且必须属于支持的数据类型 (p. 210)下面列出。

- 比较运算符

  以下是支持的运算符：=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE所有字符串比较和 LIKE 模式匹配项均区分大小写。
- 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进一步组合复合表达式AND/OR.

支持的数据类型

引用Amazon Glue Data Catalog包含不受支持的数据类型的表将导致错误。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

有关 Amazon Redshift 中数据类型的更多信息，请参阅数据类型在Amazon Redshift 数据库开发人员指南.

行筛选器表达式

Example

以下是包含列的表的有效行筛选器表达式的示例： 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)

字符串常量必须用单引号引起来。

保留关键字

如果行筛选器表达式包含 PartiQL 关键字，则会收到解析错误，因为列名可能与关键字冲突。发生这种情况时，请使用双引号对列名进行转义。保留关键字的一些示例有 “first”、“last”、“asc”、“缺失”。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/.

列级筛选的注意事项和限制

您可以通过三种方式指定列筛选：

- 通过使用数据筛选器，如前所述。
- 通过使用简单的列过滤。
- 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Amazon CLI支持简单列过滤。有关示例，请参阅 Grant with Simple Column Filtering (p. 156)。

列筛选适用列筛选的注意事项和限制：

- Amazon GlueETL 作业仅支持使用数据筛选器进行列筛选（单元级安全性）。如果将简单列筛选应用于作业引用的任何表，则作业将失败。如果只需要列筛选，请使用数据筛选器授予对表的访问权限，然后输入`true`用于控制台中的行筛选器表达式，或者使用`AllRowsWildcard`在你的 API 调用中。
- 要授予`SELECT`使用 grant 选项和列筛选时，必须使用包含列表，而不是排除列表。如果没有授予选项，则可以使用包括列表或排除列表。
- 要授予`SELECT`在具有列筛选功能的表上，您必须已获得授权`SELECT`在表中使用 grant 选项且没有任何行限制。您必须拥有对所有行的访问权限。
- 如果你同意`SELECT`对账户中的委托人使用授予选项和列筛选时，该委托人必须在授予其他委托人时为相同的列或授权列的子集指定列筛选。如果你同意`SELECT`使用授予选项和列筛选到外部账户，外部账户中的数据湖管理员可以授予`SELECT`在所有列上转移到他们账户中的另一个委托人。但是，即使`SELECT`在所有列上，该委托人将仅在授予外部账户的列上可见。
- 您无法对分区键应用列筛选。
- 一位校长`SELECT`无法授予对表中列子集的权限`ALTER`、`DROP`、`DELETE`，或者`INSERT`对该表的权限。对于拥有`ALTER`、`DROP`、`DELETE`，或者`INSERT`对表的权限，如果您授予`SELECT`列筛选权时，会没有效果。

行级和单元格级别筛选的注意事项和限制

请记住以下针对行级筛选的注意事项和限制：

- 不支持 `SELECT INTO` 语句。
- 这些区域有：`struct`、`array`，和`map`行筛选器表达式中不支持行筛选器表达式中的数据类型。
- 可以对表定义的数据筛选器数量没有限制，但数据筛选器限制为 100 个`SELECT`表中单个主体的权限。
- 要使用行筛选器表达式应用数据筛选器，您必须`SELECT`所有表列上都有 grant 选项。向外部账户授予权限时，此限制不适用于外部账户中的管理员。
- 如果委托人是组的成员，并且委托人和组都被授予对行子集的权限，则委托人的有效行权限是委托人的权限和组权限的联合。
- 以下列名在行级和单元格级别筛选的表中受到限制：
  - ctid
  - oid
  - xmin
  - cmin
  - xmax
  - cmax
  - 小说
  - insertxid
  - deletexid
  - importoid
  - redcatuniqueid

- 如果将所有行筛选器表达式与其他带谓词的筛选器表达式同时应用于表，则所有行表达式将优先于所有其他筛选器表达式。
- 将行子集的权限授予外部Amazon账户和外部账户的数据湖管理员将这些权限授予该账户中的委托人，委托人的有效筛选谓词是账户谓词与直接授予委托人的任何谓词的交集。

  例如，如果账户具有谓词的行权限`dept='hr'`并且委托人被单独授予权限`country='us'`，则委托人只能访问具有`dept='hr'`和`country='us'`.

有关单元格列筛选的更多信息，请参阅Lake Formation 中的数据过滤和细胞级安全 (p. 206).

使用单元格级别筛选查询表所需的权限

以下Amazon Identity and Access Management(IAM) 权限是针对具有单元格级别筛选的表运行查询所必需的。

```
{
    "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 (p. 322).

管理数据筛选器

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器。每个数据筛选器都属于数据目录表。您可以为一个表创建多个数据筛选器，然后在授予对表的权限时使用其中的一个或多个过滤器。

你需要`SELECT`具有创建或查看数据筛选器的授予选项的权限。允许其他委托人或其他Amazon要查看和使用数据筛选器的帐户，您可以授予`DESCRIBE`允许它。

您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

有关数据筛选器的信息，请参阅Lake Formation 中的数据过滤器 (p. 207)

创建数据筛选器

您可以为每个数据目录表创建一个或多个数据筛选器。

为数据目录表创建数据筛选器（控制台）

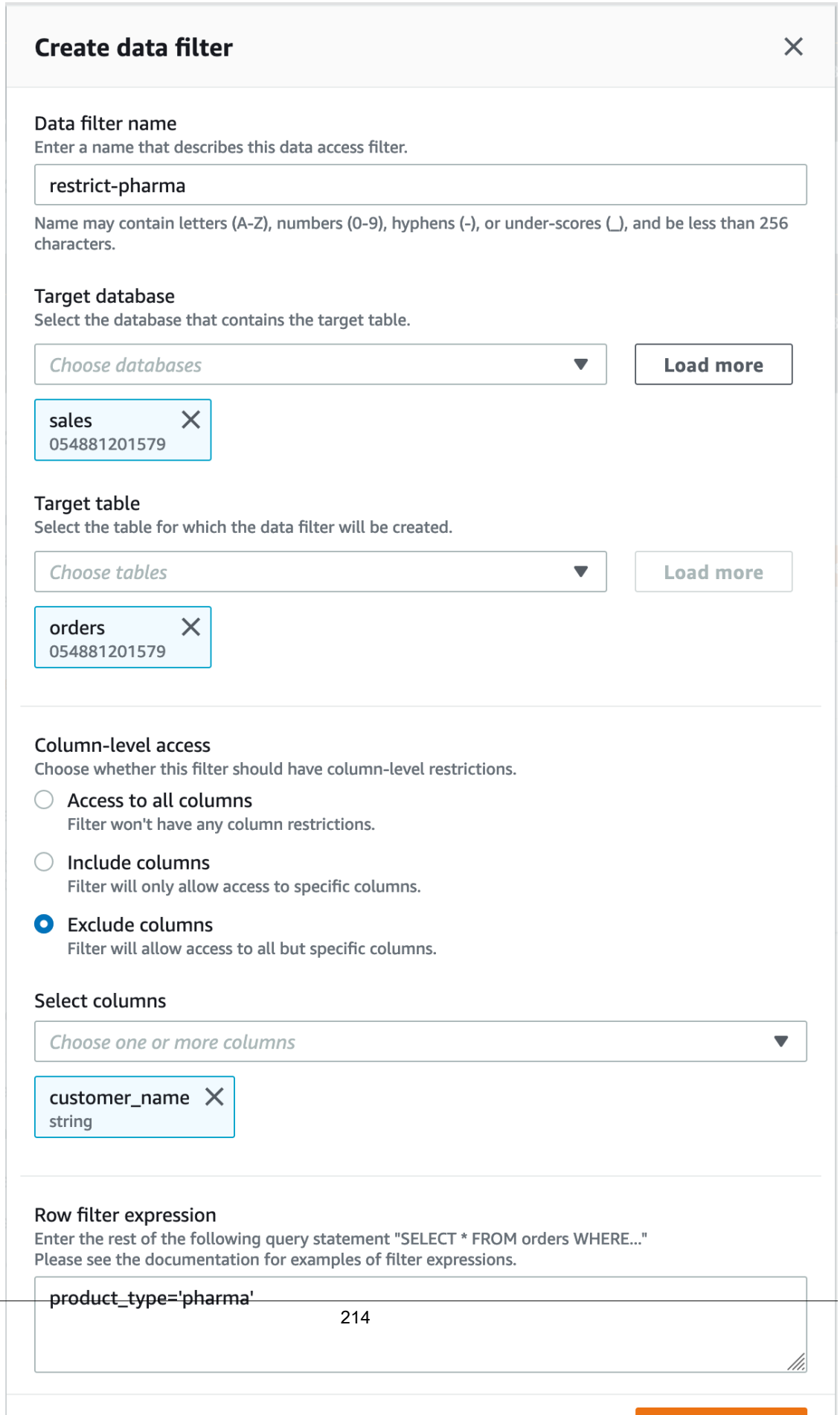
1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

以数据湖管理员、目标表所有者或对目标表具有 Lake Formation 权限的委托人身份进行签名。

2. 在导航窗格中的下，Data Catalog，选择筛选数据.
3. 在存储库的筛选数据页面上，选择创建新筛选条件.
4. 在创建数据筛选器对话框中，输入以下信息：

   - 数据筛选器名称
   - 目标数据库 — 指定包含表的数据库。
   - 目标表
   - 将此设置保留为访问所有列以仅指定行筛选。选择包括列要么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。
   - 行筛选器表达式 — 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，请参行筛选器表达式中的 PartiQL 支持 (p. 209). 如果您不想使用行筛选器表达式，请输入`true`在外地。

   以下屏幕截图显示了实现单元格过滤的数据筛选器。在针对`orders`表格中，它拒绝访问`customer_name`列中包含 “药品” 的任何行`product_type`column.


Create data filter
Data filter name
Enter a name that describes this data access filter.
restrict-pharma
Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.
Target database
Select the database that contains the target table.
Choose databases
Load more
sales
054881201579
Target table
Select the table for which the data filter will be created.
Choose tables
Load more
orders
054881201579
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Select columns
Choose one or more columns
customer_name
string
Row filter expression
Enter the rest of the following query statement "SELECT * FROM orders WHERE..."
Please see the documentation for examples of filter expressions.
product_type='pharma'

5. 请选择 Create filter (创建筛选器)。

授予数据过滤器权限

你可以授予SELECT、DESCRIBE和DROPLake Formation 对委托人的数据过滤器权限。

起初，只有您可以查看为表创建的数据筛选器。要使另一个委托人能够查看数据筛选器并授予数据目录权限，您必须执行以下任一操作：

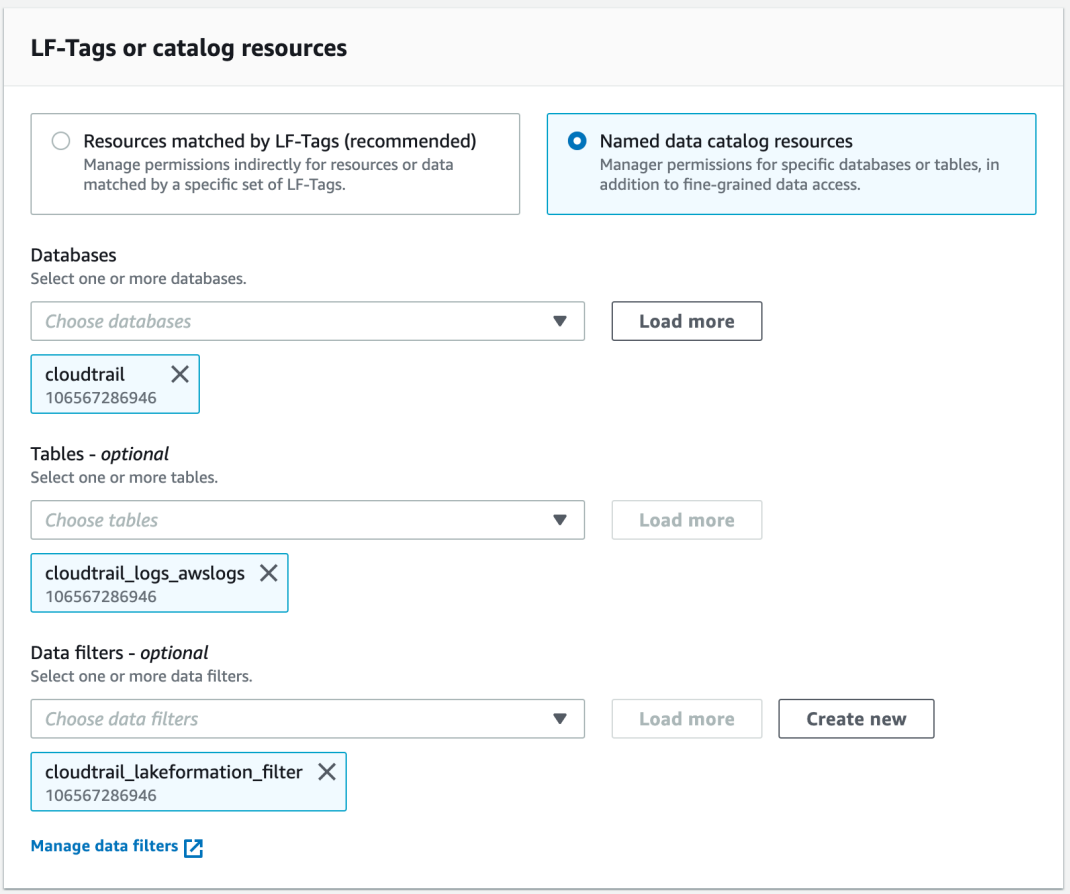
- GrantSELECT在带有授权选项的委托人的表格上，然后将数据筛选器应用于授权。
- 授予DESCRIBE要么DROP向委托人授予数据筛选权限。

你可以授予SELECT对外部的许可Amazonaccount. 然后，该账户中的数据湖管理员可以将该权限授予账户中的其他委托人。授予外部帐户时，您必须包括授予选项，以便外部账户的管理员可以进一步将权限级联给他/她账户中的其他用户。授予账户中的委托人时，使用授权选项授予权限是可选的。

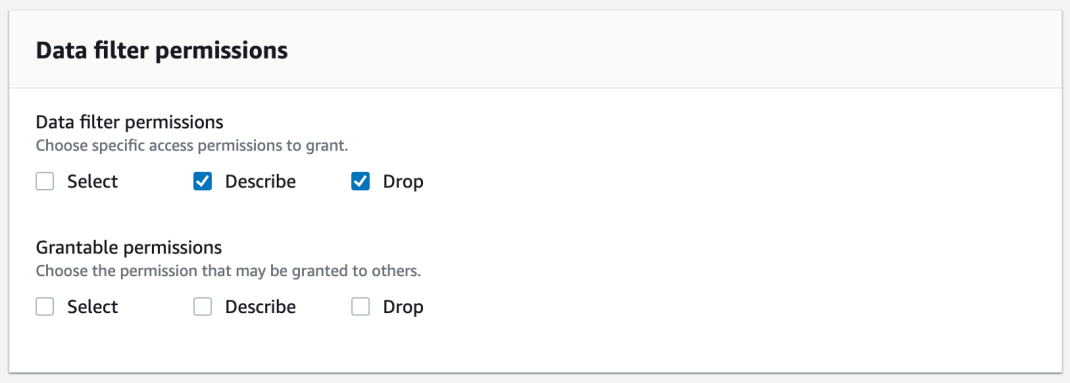
您可以使用Amazon Lake Formation控制台、API 或Amazon Command Line Interface(Amazon CLI)。

Console

1. 登录到Amazon Web Services Management Console在处打开 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/.
2. 在导航窗格中的下，Permissions (权限)，选择数据湖权限.
3. 在存储库的Permissions (权限)在页面中，数据权限部分中，选择Grant.
4. 在存储库的授予数据权限页面中，选择要向其授予权限的委托人。
5. 在 LF-Tags 或目录资源部分中，选择命名数据目录资源. 然后选择要授予权限的数据库、表和数据筛选器。



6. 在数据筛选权限部分中，选择要授予所选承担者的权限。



Amazon CLI

- 输入`grant-permissions`命令。指定`DataCellsFilter`(对于)`resource`参数，然后指定`DESCRIBE`要么`DROP`(对于)`Permissions`参数以及 (可选) `PermissionsWithGrantOption`参数。

  以下示例授予：`DESCRIBE`向用户提供授予选项`datalake_user1`在数据过滤器上`restrict-pharma`，它属于`orders`表格中的`sales`中的数据库Amazon账户 1111-2222-3333。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是文件内容grant-params.json.

```
{
    "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["DESCRIBE"],
    "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

授予数据过滤器提供的数据权限

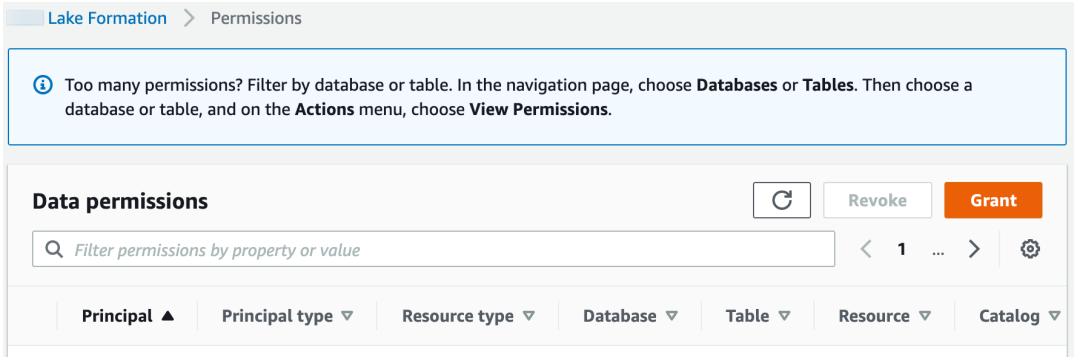
数据筛选器表示表中数据的子集。要向委托人提供数据访问权限，SELECT需要向这些委托人授予权限。有了这个权限，委托人可以：

- 在与其账户共享的表列表中查看实际表名称。
- 在共享表上创建数据筛选器，并向其用户授予对这些数据筛选器的权限。

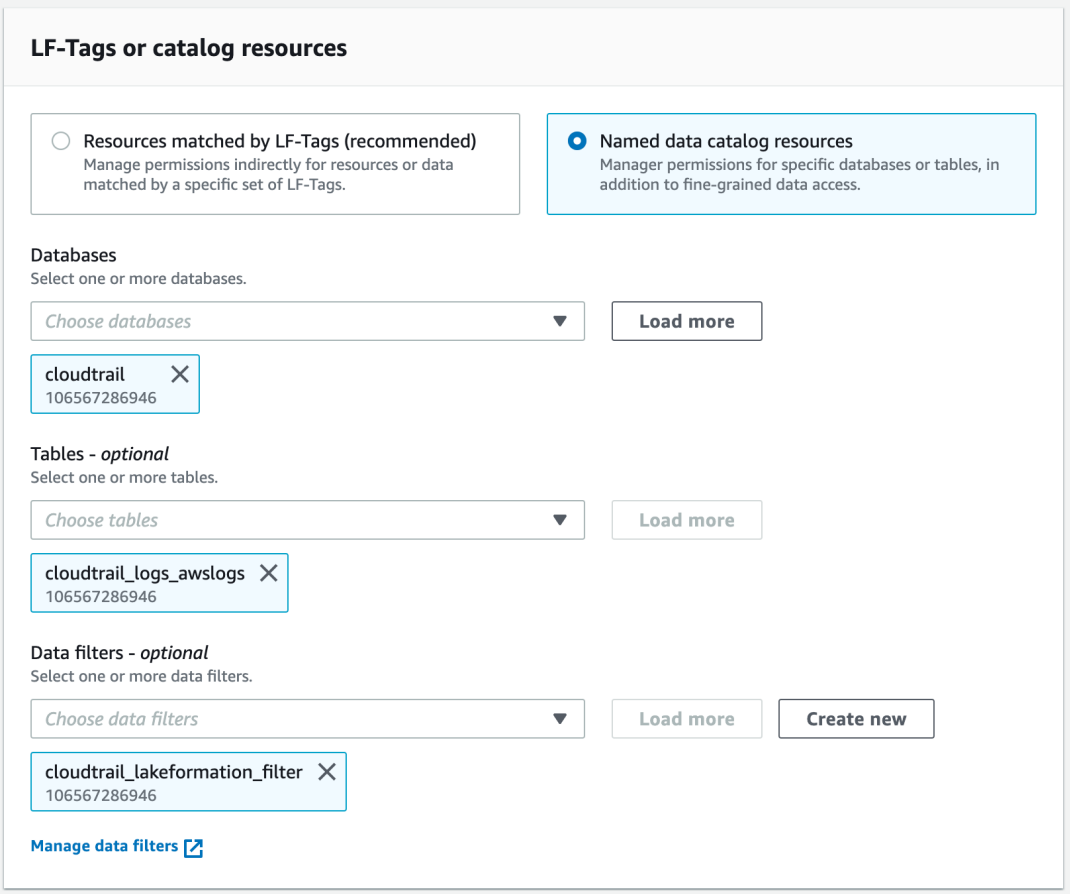
Console

授予 SELECT 权限

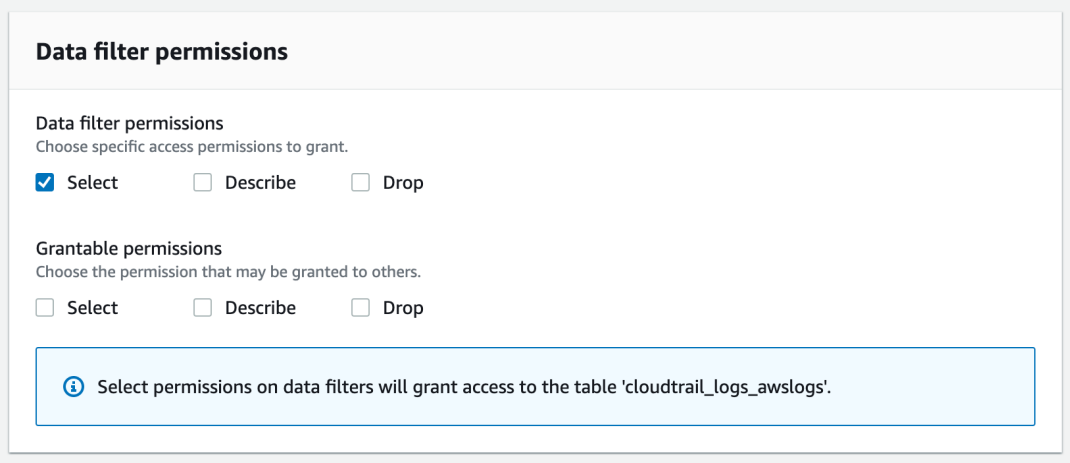
1. 转至Permissions (权限)在 Lake Formation 控制台中的页面，然后选择Grant.



2. 选择要提供访问权限的承担者，然后选择命名数据目录资源.



3. 要提供对筛选器所表示的数据的访问权限，请选择Select下数据筛选权限.



CLI

输入`grant-permissions`命令。指定`DataCellsFilter`对于资源参数，然后指定`SELECT`对于权限参数。

以下示例授予：`SELECT`向用户提供授予选项`datalake_user1`在数据过滤器上`restrict-pharma`，它属于`orders`表格中的`sales`中的数据库Amazon Web Services 账户 `1111-2222-3333`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是文件内容grant-params.json.

```
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/datalake_user1"
    },
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["SELECT"]
}
```

查看数据筛选器

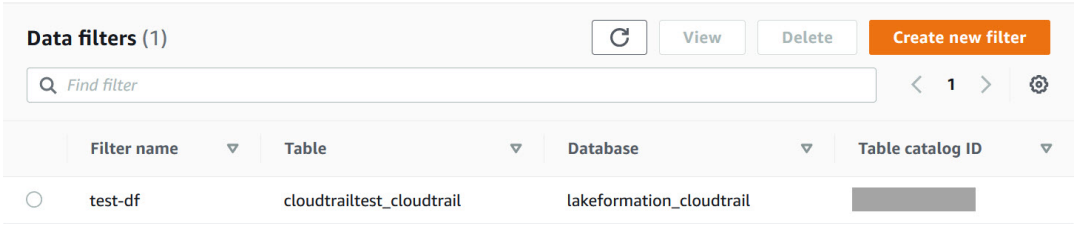
你可以使用 Lake Formation 控制台，Amazon CLI，或者使用 Lake Formation API 来查看数据过滤器。

要查看数据筛选器，您必须是 Data Lake 管理员，或者对数据筛选器具有所需的权限。

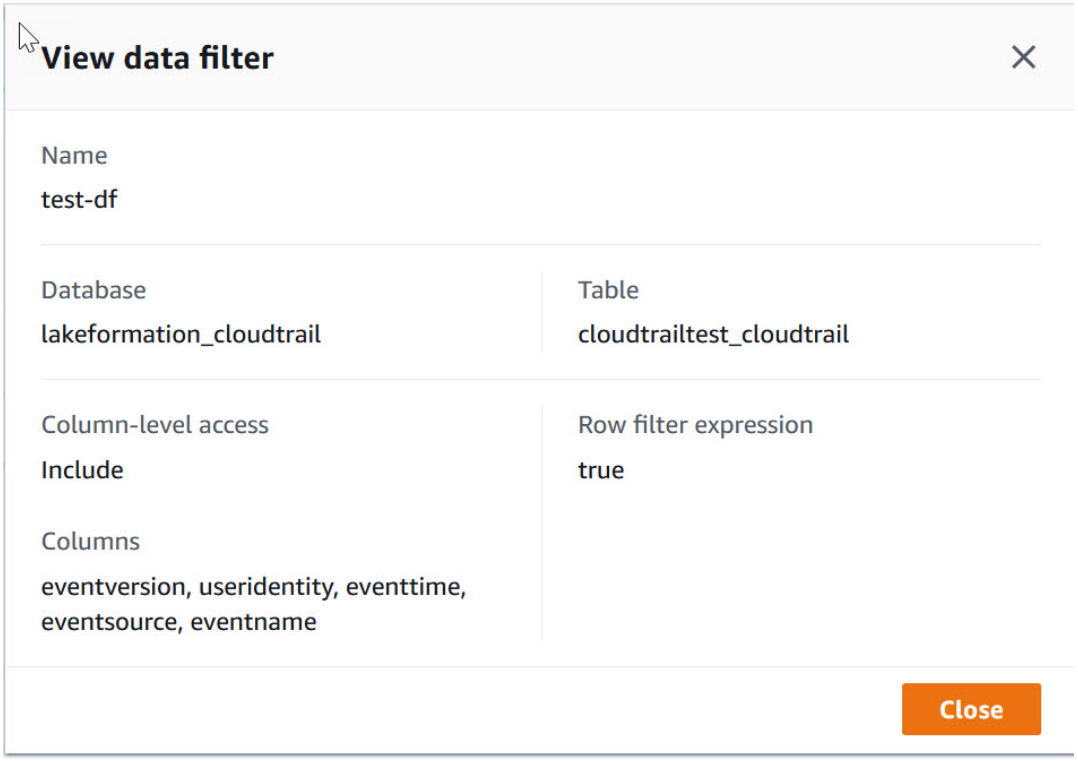
Console

1. 登录到Amazon Web Services Management Console在处打开 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/.
2. 在导航窗格中的下，Data Catalog，选择筛选数据.

   该页面显示您有权访问的数据筛选器。



3. 要查看数据筛选器详细信息，请选择数据筛选器，然后选择 “查看”。出现一个新窗口，其中包含数据筛选器的详细信息



Amazon 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。

```
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 列出的前 20 个数据筛选器`myTable`表。

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
    },
    MaxResults=20
)
```

列出数据过滤器权限

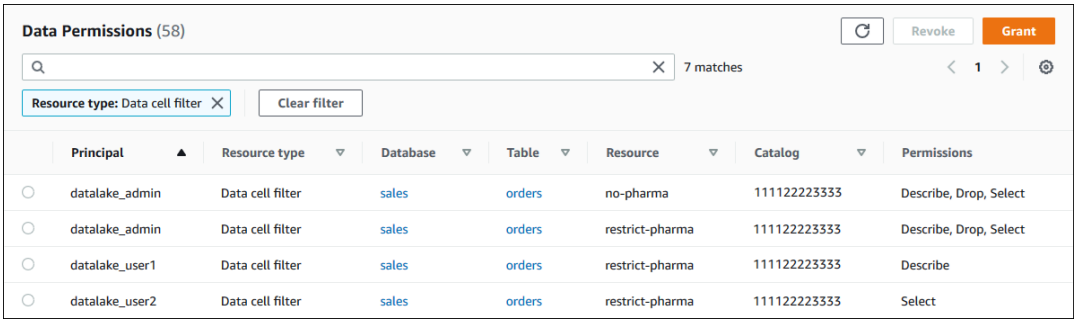
您可以使用 Lake Formation 控制台查看对数据过滤器授予的权限。

要查看数据筛选器的权限，您必须是 Data Lake 管理员，或者对数据筛选器具有所需的权限。

Console

1. 登录到Amazon Web Services Management Console在处打开 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/.
2. 在导航窗格中的下，Permissions (权限)，选择数据权限.
3. 在存储库的数据权限页面，单击或点按搜索字段中的属性菜单中，选择。资源类型.
4. 在存储库的资源类型菜单中，选择。资源类型：筛选数据单元格.

   列出了您具有权限的数据筛选器。您可能需要水平滚动才能看到Permissions (权限)和GRANTABLEColumn。



Amazon CLI

- 输入`list-permissions`命令。指定`DataCellsFilter`(对于)`resource`参数，然后指定`DESCRIBE`要么`DROP`(对于)`Permissions`参数以及 (可选) `PermissionsWithGrantOption`参数。

  以下示例列出了`DESCRIBE`在数据筛选器上使用授予选项的权限`restrict-pharma`. 结果仅限于授予委托人的权限`datalake_user1`和`orders`表格中的`sales`中的数据库Amazon账户1111-2222-3333。

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

  以下是文件内容`grant-params.json`.

```
{
    "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["DESCRIBE"],
    "PermissionsWithGrantOption": ["DESCRIBE"]
}
```

在事务中读取和写入数据湖

Amazon Lake Formation读取和写入 Amazon S3 对象的写入 Amazon S3 对象以及创建和更新数据目录中的表元数据时，支持 ACID（原子、一致、隔离和持久）事务支持 ACID（原子、一致、隔离和持久）事务支持 ACID（原子、一致 事务保持受管理的表清单的完整性 (交易数据操作) 以及其他表元数据，例如架构 (事务元数据操作)。以下是针对受管表进行事务的典型用例：

- ETL 将传递到新表— 在此用例中，您可能有Amazon Glue提取、转换和加载 (ETL) 任务，用于启动事务，从数据源读取，写入数据湖中已注册 Amazon S3 位置的数据接收器，并在数据目录中为数据接收器创建受管控表。如果 ETL 脚本在某个时候检测到故障，则该脚本可以取消事务，结果会发生以下情况：
  - 受管辖的表已从目录中删除。
  - 如果脚本调用了Lake Formation`DeleteObjectsOnCancel`在将每个新对象写入 Amazon S3 之前，API 操作，然后 Lake Formation 还会删除事务中写入到 Amazon S3 的所有对象。有关更多信息，请参阅 回滚Amazon S3 写入 (p. 223)。
- 表更新— 假设对于现有的受管控表，您的 ETL 任务会启动事务，将新对象写入 Amazon S3 并使用更新表清单`UpdateTableObjects`API 操作。如果脚本检测到故障，它可以取消事务，结果会发生以下情况：
  - 表清单将恢复到事务开始之前的状态。
  - 如果脚本调用了Lake Formation`DeleteObjectsOnCancel`API 操作在将每个新对象写入 Amazon S3 之前，Lake Formation 还会删除事务中写入到 Amazon S3 的所有对象。
- 架构更新— 对于 Amazon S3 中具有流式数据接收器的现有受管控表，如果流式处理 ETL 任务确定数据中还有其他表列，则它可以在事务中更新表架构。如果出现故障，任务可以取消事务，在这种情况下，表架构将恢复到事务开始之前的状态。
- 时间旅行查询— Lake Formation 维护多个版本（快照）表元数据，因为数据湖中的数据会发生变化。即使架构已更改，您也可以时光倒流并查询数据。

有关受管控表的更多信息，请参阅Lake Formation 中的受管桌子 (p. 114).

主题

事务数据操作

Lake Formation ACID 事务提供快照隔离，通过添加和删除 Amazon S3 对象，使多个任务能够同时可靠地更新受管控的表，同时保持针对数据湖的每次查询的读取一致性。

ACID 事务提供的读取一致性的一个示例是用户在中启动查询时Amazon Athena它会扫描表的所有分区以汇总销售数据。在扫描进行期间和扫描完成之前，ETL 任务会向表中添加一个分区。如果用户在事务中执行查询，他们将在查询开始时看到反映表状态的查询结果。结果将不包括作业创建的新分区中的数据。

参与交易访问和修改受管表的 Lake Formation 操作Amazon Glue Data Catalog. 一个事务可能涉及多个受管控的表。

事务还为涉及元数据的更改提供 ACID 属性，元数据定义了构成受管表的 Amazon S3 对象（表现）以及表架构。有关受管控表的更多信息，请参阅Lake Formation 中的受管桌子 (p. 114).

集成Amazon当涉及受管理的表时，诸如 Athena 之类的服务会自动在事务中执行查询。要在您的账户中使用交易Amazon GlueETL 任务，作业脚本在对数据湖执行任何读取或写入操作之前开始事务，引用事务内任何操作的交易 ID，并在完成时提交事务。如果读取或写入操作失败，作业脚本可以重试该操作，也可以选择取消事务。您还可以通过指定要读取的过去时间点来在查询中使用事务。

如果 Lake Formation 检测到某些故障（例如写入冲突），Lake Formation 可能会自动取消交易。取消事务会导致所有操作被回退，如中所述在事务中读取和写入数据湖 (p. 222).

为了让 Lake Formation 能够区分长时间运行的事务（例如，运行数小时的 Spark ETL 作业）和由于崩溃而放弃的事务，长时间运行的写入事务应调用 Heartbeat API 操作`ExtendTransaction`定期地。对于只读交易，这不是必需的。Lake Formation 会自动取消闲置时间过长的交易。

受管表中的提交流程

对受管控表的修改必须在事务的上下文中进行。如果您的 ETL 任务在未明确提供事务 ID 的情况下对受管表执行操作，则 Lake Formation 会自动启动事务并在操作结束时提交（或取消）该事务。这被称为单语句交易。

对于具有写入操作的交易，调用`CommitTransaction`会将交易移至`COMMIT_IN_PROGRESS`状态。内部后台进程负责将事务中的更改应用到受管控表中，然后再将事务移至受管控表`COMMITTED`status。因此，在调用后立即执行读取操作`CommitTransaction`可能反映也可能不反映写入操作的结果。为了确定性地读取写入操作的结果，客户应等到交易状态更改为`COMMITTED`. 这可以通过以下任一调用来检查`CommitTransaction`要么`DescribeTransaction`API 操作。单语句事务的读取操作也演示了相同的行为。

回滚Amazon S3 写入

当交易被取消时，可以是自动取消的，也可以是通过调用CancelTransaction，未经您的许可，Lake Formation 绝不会删除写入 Amazon S3 的数据。要授予 Lake Formation 回滚交易期间写入的权限，你的代码必须调用DeleteObjectsOnCancel API 操作，其中列出了取消交易后可以删除的 Amazon S3 对象。建议你打电话`DeleteObjectsOnCancel`在写作之前。

这些区域有：Amazon GlueETL 库函数`write_dynamic_frame.from_catalog()`包括自动拨打电话的选项`DeleteObjectsOnCancel`在写入之前。在下面的示例中，`callDeleteObjectsOnCancel`选项包含在`additional_options`参数。因为价值`False`传递到`read_only`的参数`start_transaction`、事务不是只读事务。

```
transactionId = glueContext.start_transaction(False)

try:
    datasink0 = glueContext.write_dynamic_frame.from_catalog(
        frame = datasource0,
        database="MyDatabase",
        table_name="MyGovernedTable",
        additional_options={
            "partitionKeys":["key1", "key2"],
            "transactionId":transactionId,
            "callDeleteObjectsOnCancel":"true"
        }
    )
    glueContext.commit_transaction(transactionId)

except:
```

```
glueContext.cancel_transaction(transactionId)
```

交易元数据操作

受管表的元数据包括以下内容：

- 表架构
- Amazon S3 表对象清单
- 版本历史记录
- 权限

在更改受管表的清单时，应用程序还可以以事务方式更改数据目录中的架构。通过提供元数据交易，您可以对数据进行时空旅行查询，也可以查询以前存在的数据。

例如，Amazon Glue表 API 操作采用可选事务 ID，以允许应用程序创建、更新或删除受管表的元数据。这是由返回的交易 IDStartTransactionAPI。

当交易提交时，快照由表元数据创建并与事务 ID 相关联。API 操作可以通过传递事务 ID 来引用特定的快照。例如，您对UpdateTable使用事务 IDabc然后是另一个更新xyz更改架构。然后你可以打电话GetTable使用事务 IDabc，它接收的架构与事务 ID 时相同abc已提交。

交易元数据的另一个用例是执行时空旅行查询，以读取截至指定时间的表的元数据GetTable要么GetTablesoperation.

支持事务元数据操作的 API 操作

以下Amazon GlueAPI 操作支持事务元数据操作。对于每种描述，都假设交易 ID（返回者为StartTransaction要么start_transaction) 被传递给 API 调用，并且在某个时候会有提交。

API 操作	描述
CreateTable	当您提交事务创建数据目录中创建表（当您调用CommitTransaction带有交易 ID 的 API）。
UpdateTable	更新表的元数据。提交事务后，创建表元数据的新快照。快照按事务 ID 编制索引。
DeleteTable, BatchDeleteTable	提交事务时删除表。所有快照都被删除。
GetTable	（可选）采用TransactionId或者一个QueryAsOfTime参数。 如果你通过了TransactionId，该操作将检索匹配的快照并返回截至该事务提交时的表元数据。 如果你通过了QueryAsOfTime参数，返回截至指定时间的表元数据。
GetTables	（可选）采用TransactionId或者一个QueryAsOfTime参数。 如果两个参数均未传递，则该操作将返回受管表和未受管控表。如果你通过了TransactionId或者一个QueryAsOfTime参数，仅返回受管控的表。 如果你通过了TransactionId，返回截至交易提交时的表元数据。 如果你通过了QueryAsOfTime参数，该操作返回截至指定时间的表的元数据。

交易元数据操作的限制

以下是事务元数据操作的当前限制：

- 例如，在数据目录中添加或修改数据库时，Lake Formation 不支持事务操作，`CreateDatabase`、`UpdateDatabase`、`GetDatabase`、等
- Lake Formation 不支持在数据目录中修改数据库时进行时空旅行读取。
- 例如，在添加或修改表分区时，Lake Formation 不支持事务操作`CreatePartition`、`BatchUpdatePartition`、等 分区键可以通过查看表对象来识别。

支持交易的 Lake Formation API 操作

Lake Formation 提供以下 API 操作来支持交易：

API 操作	描述
`StartTransaction`	启动新交易并返回其交易 ID。
`CommitTransaction`	尝试使用指定内容提交事务`TransactionId`. 如果成功，则事务期间所做的所有修改都将保留。
`CancelTransaction`	停止与指定的关联事务`TransactionId`. 交易期间所做的所有修改都将回退。
`ExtendTransaction`	指示指定的事务 (`TransactionId`) 仍处于活动状态，不应取消。空闲事务的当前超时时间为 30 秒。只读事务不需要。
`DescribeTransaction`	返回由其标识的事务返回其状态`TransactionId`(`ACTIVE`、`COMMITTED`，或者`ABORTED`)。
`ListTransactions`	返回未提交的交易和可用于时空旅行的交易 ID、状态、开始和结束时间。
`DeleteObjectsOnCance`	返回将在当前事务中写入 Amazon S3 对象的列表（已授权由`TransactionId`)。如果事务被取消，Lake Formation 会自动删除这些对象。
`GetTableObjects`	返回构成指定受管控表的 Amazon S3 对象集。可以为时空旅行查询指定事务 ID 或时间戳。
`UpdateTableObjects`	更新构成指定受管控表的 Amazon S3 对象清单。

事务编码的最佳实践

以下是编码的一些最佳实践Amazon Glue用于交易的 ETL 脚本。

- 启动事务时，请确保存在异常处理，以便在出现任何异常时取消事务。有关示例，请参阅回滚Amazon S3 写入 (p. 223)。
- 如果你的任务使用了中列出的任何 API支持事务元数据操作的 API 操作 (p. 224)，请指定事务 ID，以确保在事务失败或被取消时可以清理该事务。
- 请记住，系统可能出于多种原因取消您的交易。使用`ExtendTransaction`用于防止长时间运行的交易被取消的 API。
- 如果您尝试在已取消的交易中进行操作，则会得到`TransactionCanceledException`，所以你的代码应该处理这个异常。您可以先检查事务的状态，然后再尝试使用`DescribeTransaction`API 操作。

- 用于管理内部的交易Amazon GlueETL 任务脚本，请参阅交易函数`GlueContext`或者Python要么斯卡拉.

数据湖交易代码示例

以下 PySpark 和 Java 代码示例演示了如何启动和提交事务，以及在出现异常时如何停止和回滚事务。

Note

对于写入 Apache Parquet，Amazon Glue ETL 仅支持为针对动态帧进行优化的自定义 Parquet 编写器类型指定选项来写入受管表。使用 `parquet` 格式写入受管表时，应在表参数中添加值为 `true` 的键 `useGlueParquetWriter`。

PySpark

此示例显示了 ETL 脚本Amazon Glue工作。它将数据从不受管控的表复制到受管表中。需要按照中的说明创建受管控表创建受管表 (p. 116)或者使用Amazon Athena。

```
import sys
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job

## @params: [JOB_NAME]
args = getResolvedOptions(sys.argv, ["JOB_NAME"])
sc = SparkContext.getOrCreate()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
job.init(args["JOB_NAME"], args)

db = "retail"
tbl = "inventory"
tx_id = glueContext.start_transaction(False)

datasource0 = glueContext.create_dynamic_frame.from_catalog(
    database = db, table_name = tbl,
    transformation_ctx = "datasource0")
datasource0.show()

dest_path = "s3://path_to_sales_retail_data/"

try:
    glueContext.write_dynamic_frame.from_catalog(
        frame = datasource0,
        database = "retail",
        table_name = "inventory-governed",
        additional_options = {
            "transactionId":tx_id
      }
    )
    glueContext.commit_transaction(tx_id)
except Exception:
    glueContext.cancel_transaction(tx_id)
    raise
job.commit()
```

下一个示例演示了事务的用法Amazon Glue直播 ETL 作业。

```
import sys
from awsglue.transforms import ApplyMapping
```

```
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job
from awsglue.dynamicframe import DynamicFrame

args = getResolvedOptions(sys.argv, ['JOB_NAME'])

sc = SparkContext.getOrCreate()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
job.init(args['JOB_NAME'], args)

data_frame_DataSource0 = glueContext.create_data_frame.from_catalog(
    database = "demo",
    table_name = "kinesis_cloudtrail_demo",
    transformation_ctx = "DataSource0",
    additional_options = {
        "startingPosition": "TRIM_HORIZON",
        "inferSchema": "true"
    }
)

def processBatch(data_frame, batchId):
    if (data_frame.count() > 0):
        dynamic_frame = DynamicFrame.fromDF(data_frame, glueContext, "from_data_frame")
        dynamic_frame = ApplyMapping.apply(
            frame = dynamic_frame,
            mappings = [
                ("eventversion", "string", "eventversion", "string"),
                ("eventtime", "string", "eventtime", "string"),
                ("eventsource", "string", "eventsource", "string"),
                ("eventname", "string", "eventname", "string"),
                ("awsregion", "string", "awsregion", "string"),
                ("sourceipaddress", "string", "sourceipaddress", "string"),
                ("useragent", "string", "useragent", "string"),
                ("errorcode", "string", "errorcode", "string"),
                ("errormessage", "string", "errormessage", "string"),
                ("requestid", "string", "requestid", "string"),
                ("eventid", "string", "eventid", "string"),
                ("eventtype", "string", "eventtype", "string"),
                ("apiversion", "string", "apiversion", "string"),
                ("readonly", "boolean", "readonly", "string"),
                ("recipientaccountid", "string", "recipientaccountid", "string"),
                ("sharedeventid", "string", "sharedeventid", "string"),
                ("vpcendpointid", "string", "vpcendpointid", "string")
            ],
            transformation_ctx = "ApplyMapping"
        )

        table_name = "cloudtrail_demo_sample"
        txId = glueContext.begin_transaction(False)
        try:
            glueContext.write_dynamic_frame.from_catalog(
                frame = dynamic_frame,
                database = "demo",
                table_name = table_name,
                additional_options = {
                    "transactionId":txId
                }
            )
            glueContext.commit_transaction(txId)
        except Exception:
            glueContext.cancel_transaction(txId)
            raise
```

```
glueContext.forEachBatch(
    frame = data_frame_DataSource0,
    batch_function = processBatch,
    options = {
        "windowSize": "100 seconds",
        "checkpointLocation": "s3://my_checkpoint_bucket/cloudtrail_demo_sample/"
    }
)
job.commit()
```

Java

以下两个 Java 示例使用了`getTableObjects`和`updateTableObjects`事务中受管控表的对象 API 操作。有关这些 API 操作的信息，请参阅API 文档.

第一个示例将有关新分区的信息添加到现有的受管控表中。

导入

```
import com.amazonaws.services.lakeformation.AWSLakeFormation;
import com.amazonaws.services.lakeformation.AWSLakeFormationClientBuilder;
import com.amazonaws.services.lakeformation.model.CancelTransactionRequest;
import com.amazonaws.services.lakeformation.model.AddObjectInput;
import com.amazonaws.services.lakeformation.model.StartTransactionRequest;
import com.amazonaws.services.lakeformation.model.StartTransactionResult;
import com.amazonaws.services.lakeformation.model.CommitTransactionRequest;
import com.amazonaws.services.lakeformation.model.UpdateTableObjectsRequest;
import com.amazonaws.services.lakeformation.model.WriteOperation;
```

代码

```
AWSLakeFormation lake = AWSLakeFormationClientBuilder.standard().build();
// Begin transaction
BeginTransactionResult startTransactionResult = lake.startTransaction(new
 StartTransactionRequest());
String transactionId = startTransactionResult.getTransactionId();
// Construct a write operation
WriteOperation write = new WriteOperation()
        .withAddObject(new AddObjectInput()
                .withPartitionValues("par0", "par1")
                .withUri("s3://bucket/prefix")
                .withETag("eTag")
                .withSize(100L));
// Save a partition table object
try {
    lake.updateTableObjects(new UpdateTableObjectsRequest()
            .withDatabaseName(databaseName)
            .withTableName(tableName)
            .withTransactionId(transactionId)
            .withWriteOperations(write));
    // Commit transaction
    lake.commitTransaction(new
 CommitTransactionRequest().withTransactionId(transactionId));
} catch (Exception e) {
    // Abort transaction
    lake.cancelTransaction(new
 CancelTransactionRequest().withTransactionId(transactionId));
}
```

下一个示例检索与受管表关联的所有 Amazon S3 对象并对其进行处理。

导入

```
import com.amazonaws.services.lakeformation.AWSLakeFormation;
import com.amazonaws.services.lakeformation.AWSLakeFormationClientBuilder;
import com.amazonaws.services.lakeformation.model.StartTransactionRequest;
import com.amazonaws.services.lakeformation.model.StartTransactionResult;
import com.amazonaws.services.lakeformation.model.CommitTransactionRequest;
import com.amazonaws.services.lakeformation.model.GetTableObjectsRequest;
import com.amazonaws.services.lakeformation.model.GetTableObjectsResult;
import com.amazonaws.services.lakeformation.model.PartitionObjects;
import com.amazonaws.services.lakeformation.model.TableObject;
```

代码

```
AWSLakeFormation lake = AWSLakeFormationClientBuilder.standard().build();
// Start read only transaction
StartTransactionResult startTransactionResult = lake.startTransaction(new
 StartTransactionRequest().withReadOnly(true));
String transactionId = startTransactionResult.getTransactionId();
// Read all table objects from a table
GetTableObjectsResult getTableObjectsResult = lake.getTableObjects(
        new GetTableObjectsRequest()
                .withTransactionId(transactionId)
                .withDatabaseName(databaseName)
                .withTableName(tableName));
for (PartitionObjects partitionObjects: getTableObjectsResult.getObjects()) {
    for (TableObject tableObject: partitionObjects.getObjects()) {
        // do something with the data
    }
}
// Commit transaction
lake.commitTransaction(new
 CommitTransactionRequest().withTransactionId(transactionId));
```

Amazon Lake Formation 中的安全性

Amazon 的云安全性的优先级最高。作为 Amazon 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 Amazon 和您的共同责任。责任共担模型将其描述为云的安全性和云中的安全性：

- 云的安全性 – Amazon负责保护在Amazon云中运行Amazon服务的基础设施。Amazon还向您提供可安全使用的服务。作为 Amazon 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 Amazon Lake Formation 的合规性计划，请参阅合规性计划范围内的 Amazon 服务。
- 云中的安全性 - 您的责任由您使用的 Amazon 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。

该文档帮助您了解如何在使用 Lake Formation 时应用责任共担模型。以下主题说明如何配置以 Lake Formation 现您的安全性和合规性目标。您还将了解如何使用其他Amazon帮助您监控和保护您的 Lake Formation 资源的服务。

主题

Lake Formation 中的数据保护

这些区域有：Amazon 责任共担模式适用于中的数据保护AmazonLake Formation。如该模式中所述，Amazon 负责保护运行所有 Amazon Web Services 云 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。此内容包括您所使用的Amazon Web Services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。

出于数据保护目的，我们建议您保护 Amazon Web Services 账户 凭证并使用 Amazon Identity and Access Management (IAM) 设置单独的用户账户。这仅向每个用户授予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您的数据：

- 对每个账户使用 Multi-Factor Authentication (MFA)。
- 使用 SSL/TLS 与 Amazon 资源进行通信。建议使用 TLS 1.2 或更高版本。
- 使用 Amazon CloudTrail 设置 API 和用户活动日志记录。
- 使用 Amazon 加密解决方案以及 Amazon 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 Simple Storage Service（Amazon S3）中的个人数据。
- 如果在通过命令行界面或 API 访问 Amazon 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（例如您客户的电子邮件地址）放入标签或自由格式字段（例如名称字段）。这包括使用 Lake Formation 或其他工作时Amazon使用控制台、API、Amazon CLI，或者Amazon开发工具包。您在用于名称的标签或自由格式字段中输入的任何数据都可能会用于计费或诊断日志。当您向外部服务器提供 URL 时，强烈建议您不要在 URL 中包含凭证信息来验证您对该服务器的请求。

静态加密

Amazon Lake Formation支持以下方面的数据加密：

- Amazon Simple Simple Storage Service (Amazon S3) 数据湖中的数据。

  Lake Formation 支持数据加密Amazon Key Management Service(Amazon KMS)。数据通常通过以下方式写入数据湖Amazon Glue提取、转换和加载 (ETL) 作业。了解如何对写入的数据进行加密的信息Amazon Glue工作，请参阅加密爬网程序、作业和开发终端节点写入的数据中的Amazon Glue开发人员指南.
- 这些区域有：Amazon Glue Data Catalog，Lake Formation 是 Lake Formation 存储描述数据湖中数据的元数据表的地方。

  有关更多信息，请参阅 。加密数据目录中的Amazon Glue开发人员指南.

要在数据湖中添加 Amazon S3 位置作为存储空间，您注册有的位置Amazon Lake Formation. 然后，您可以使用 Lake Formation 权限进行精细的访问控制，Amazon Glue Data Catalog指向此位置以及该位置中的基础数据的对象。

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 (p. 105)。

Amazon Lake Formation 中的基础设施安全性

作为托管服务，Amazon Lake Formation受保护Amazon中描述的全局网络安全程序Amazon Web Services：安全过程概述白皮书。

你用Amazon发布 API 调用以通过网络访问 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 Amazon Security Token Service (Amazon STS) 生成临时安全凭证来对请求进行签名。

主题

Amazon Lake Formation 和接口 VPC 终端节点 (Amazon PrivateLink)

Amazon VPC 是一项 Amazon 服务，可用来启动在虚拟网络中定义的 Amazon 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) 托管您的Amazon资源，则您可以在 VPC 和 Lake Formation 之间建立私有连接。您使用此连接以便 Lake Formation 可以与 VPC 中的资源通信而无需通过公共互联网访问。

您可以通过创建接口 VPC 终端节点在 VPC 和 Amazon Lake Formation 之间建立私有连接。接口终端节点由以下公司提供提供支持Amazon PrivateLink，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Amazon Direct Connect连接。VPC 中的实例即使没有公有 IP 地址也可与 Lake Formation API 进行通信。VPC 和 Lake Formation 之间的流量不会脱离 Amazon 网络。

每个接口终端节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 终端节点 (Amazon PrivateLink)。

Lake Formation VPC 端点的注意事项

在为 Lake Formation 设置接口 VPC 终端节点之前，请务必查看接口终端节点属性和限制中的Amazon VPC User Guide.

Lake Formation 支持从 VPC 调用它的所有 API 操作。您可以将 Lake Formation 与 VPC 终端节点一起使用。Amazon Web Services 区域同时支持 Lake Formation 和 Amazon VPC 终端节点。

为 Lake Formation 创建接口 VPC 终端节点

您可以使用 Amazon VPC 控制台或使用 Amazon VPC 控制台或为 Lake Formation Services 创建 VPC 终端节点。Amazon Command Line Interface(Amazon CLI)。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 终端节点：

- com.amazonaws.*region*.lakeformation

如果为终端节点启用私有 DNS，则可以使用针对区域的默认 DNS 名称向 Lake Formation 发出 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 终端节点策略

Lake Formation 支持 VPC 终端节点策略 VPC 终端节点策略是Amazon Identity and Access Management在创建或修改终端节点时可附加到终端节点的 (IAM) 资源策略。

您可以为 VPC 终端节点附加控制对 Lake Formation 的访问的终端节点策略。该策略指定以下信息：

- 可执行操作的委托人。
- 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 终端节点控制对服务的访问。

示例：Lake Formation 操作的 VPC 端点策略

下面的示例 VPC 终端节点策略允许使用 Lake Formation 权限进行凭据自动售货。您可以使用此策略使用来自 Amazon Redshift 集群或使用 Lake Formation 权限运行查询Amazon EMR位于私有子网中的集群。

```
{
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "lakeformation:GetDataAccess",
```

```
            "Resource": "*",
            "Principal": "*"
        }
    ]
}
```

Note

如果您在创建终端节点时未附加策略，则会附加一个默认策略，该策略允许对服务的完全访问。

有关更多信息，请参阅 Amazon VPC 文档中的以下主题：

- 什么是 Amazon VPC？
- 创建接口终端节点
- 使用 VPC 终端节点策略

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 Amazon 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，Amazon 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。

我们建议使用资源策略中的 `aws:SourceArn` 和 `aws:SourceAccount` 全局条件上下文键，限制 Amazon Lake Formation 为另一项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 只支持`aws:SourceArn`采用以下格式：

```
arn:aws:lakeformation:aws-region:account-id:*
```

以下示例显示了如何使用`aws:SourceArn`和`aws:SourceAccount`Lake Formation 中的全局条件上下文键，以防止混淆的副手问题。

```
{
  "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:*"
        }
      }
    }
```

```
  ]
}
```

安全和访问控制Lake Formation 中的元数据和数据

Amazon Lake Formation提供了基于简单授予/撤销机制的权限模型。Lake Formation 权限与Amazon Identity and Access Management(IAM) 权限来控制对存储在数据湖中的数据的访问以及对描述该数据的元数据的访问。

在了解 Lake Formation 权限模型的详细信息之前，请先查看以下背景信息：

- 由 Lake Formation 管理的数据湖位于Amazon Simple Storage Service Service Service
- Lake Formation 维护一个数据目录，其中包含有关要导入到数据湖中的源数据（例如日志和关系数据库中的数据）的元数据，以及有关 Amazon S3 中数据湖中的数据的元数据。元数据被组织为数据库和表。元数据表包含架构、位置、分区以及与其所代表的数据有关的其他信息。元数据数据库是表的集合。
- Lake Formation 数据目录与所使用的数据目录相同Amazon Glue. 您可以使用Amazon Gluecrawler 来创建数据目录表，你可以使用Amazon Glue提取、转换和加载 (ETL) 作业来填充数据湖中的底层数据。
- 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将它们与数据源中的表或 Amazon S3 中的表格数据区分开来。元数据表在 Amazon S3 或数据源中指向的数据称为基础数据.
- 一个校长是 IAM 用户或角色、Amazon 用户或角色，是亚马逊 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，或者对于跨账户访问控制，Amazon账户 ID、组织 ID 或组织单位 ID。
- Amazon 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Amazon Command Line Interface(Amazon CLI)。创建元数据表时，您必须指定位置。在创建数据库时，该位置是可选的。表位置可以是 Amazon S3 位置或数据源位置，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（例如 Amazon Athena 和 Amazon Redshift）可以访问数据目录以获取元数据并检查正在运行的查询的授权。有关集成服务的完整列表，请参阅Amazon与 Lake Formation 的服务集成 (p. 3).

主题

Lake Formation 访问控制概述

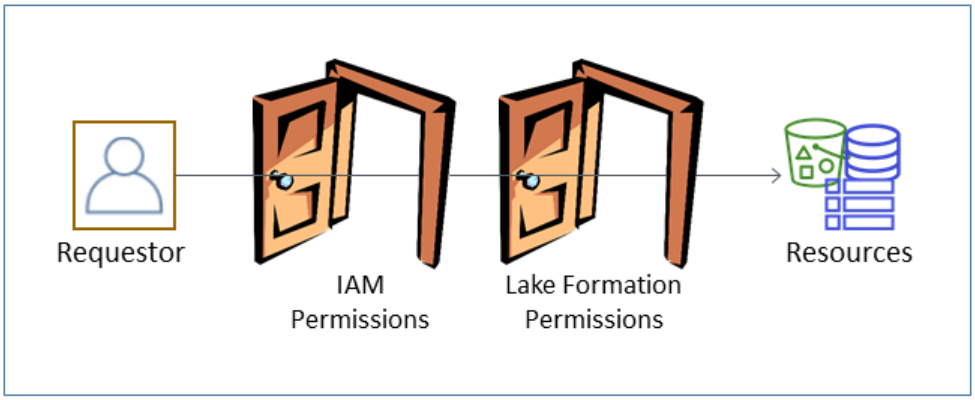
中的访问控制Amazon Lake Formation分为以下两个区域：

- 元数据访问控制— 数据目录资源的权限 (数据目录权限)。

  使用这些权限，主体可以创建、读取、更新和删除数据目录中的元数据数据库和表。
- 底层数据访问控制— 使用 Amazon SSimple Storage Service (Amazon S3) vic数据访问权限和数据位置权限)。
  - 数据访问权限使主体能够读取和写入数据隐含的Amazon S3 位置-数据目录资源所指向的数据。
  - 数据位置位置权限支持委托人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Amazon 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 风格的 GRANT/REVOKE 命令实现的，例如`Grant SELECT on tableName to userName`.

当委托人请求访问数据目录资源或基础数据时，为了使请求成功，它必须通过 IAM 和 Lake Formation 的权限检查。



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 的访问和Amazon GlueAPI 和资源。因此，尽管你可能拥有在数据目录中创建元数据表的 Lake Formation 权限（`CREATE_TABLE`），如果您没有对 IAM 权限，则您的操作将失败`glue:CreateTableAPI`。（为什么选择 `aglue:`权限？ 因为Lake Formation 使用Amazon Glue数据目录。）

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

主题

用于精细访问控制的方法

有了数据湖，目标是实现对数据的精细访问控制。在 Lake Formation 中，这意味着对数据目录资源和 Amazon S3 位置进行精细的访问控制。您可以使用以下方法之一实现了精细的访问控制。

方法	Lake Formation mat	IAM 权限	注释
方法 1	Open（打开）	精密码策略	这是默认方法用于向后兼容Amazon Glue. • 打开意味着特殊权限`Super`被授予组`IAMAllowedPrincipals`其中，其中，`IAMAllowedPrincipals`是自动创建的，包括您的 IAM 策略允许访问您的数据目录资源的任何 IAM 用户和角色，`Super`权限使主体能够对授予权限的数据库或表执行所有受支持的 Lake Formation 操作。这实际上导致对数据目录资源和 Amazon S3 位置的访问完全由 IAM 策略控制。有关更多信息，请参阅 更改数据湖的默认安全

方法	Lake Formation mat	IAM 权限	注释
			设置 (p. 253) 和 升级Amazon Glue的数据权限 Amazon Lake Formation模型 (p. 22)。 • 精密码策略意味着 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	精密码策略	粗粒度	这是推荐方法。 • 精密码策略访问权限是指向数据目录资源、Amazon S3 位置以及这些位置中的基础数据的个别委托人授予有限的 Lake Formation 权限。 • 粗粒度意味着对单个操作和 Amazon S3 位置的访问权限更广。例如，粗粒度的 IAM 策略可能包括`"glue:*"`要么`"glue:Create*"`而不是`"glue:CreateTables"`，保留 Lake Formation 权限来控制委托人是否可以创建目录对象。这也意味着授予委托人访问他们完成工作所需的 API 的权限，但要锁定其他 API 和资源。例如，您可以创建一个 IAM 策略，该策略允许委托人创建 Data Catalog 资源以及创建和运行工作流程，但不允许创建Amazon Glue连接或用户定义的函数。请参阅此部分后面的示例。

Important

请注意以下事项：

- 默认情况下，Lake Formation 具有仅使用 IAM 访问控制已启用与现有设置兼容Amazon Glue数据目录行为。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 更改数据湖的默认安全设置 (p. 253)。
- 数据湖管理员和数据库创建者具有隐含的 Lake Formation 权限，您必须了解这些权限。有关更多信息，请参阅 Lake Formation 的 (p. 140)。

元数据访问控制

对于数据目录资源的访问控制，以下讨论假设具有 Lake Formation 权限的精细访问控制和使用 IAM 策略的粗粒度访问控制。

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

- 命名资源访问控制— 使用此方法，您可以通过指定数据库或表名来授予对特定数据库或表的权限。补助金有这样的形式：

  Grant许可到校长上资源[有授予选项]。

  使用授予选项，您可以允许被授权者将权限授予其他委托人。
- 基于标签的访问控制— 使用此方法，您可以分配一个或多个LF 标签访问数据目录数据库、表和列，并将一个或多个 LF 标记的权限授予主体。每个 LF-Tag 都是一个键-值对，如`department=sales`. 拥有与数据目录资源上的 LF 标签匹配的 LF 标签的委托人可以访问该资源。对于具有大量数据库和表的数据湖，建议使用此方法。在中详细解释了它Lake Formation 标签访问控制概述 (p. 179).

委托人对资源拥有的权限是这两种方法所授予的权限的联合。

下表汇总了数据目录资源的可用Lake Formation 权限。列标题指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已授予对数据库的权限。这意味着主体可以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，但这些权限适用于基础数据。例如，DROP元数据表的权限允许您从数据目录中删除该表。但是，DELETE授予对同一表的权限使您能够在 Amazon S3 中删除该表的基础数据，例如，使用 SQLDELETE网页。有了这些权限，您还可以在 Lake Formation 控制台上查看该表，并使用Amazon GlueAPI。因此，SELECT、INSERT, 和DELETE既是数据目录权限，又是数据访问权限。

授予时SELECT在表中，可以添加包含或排除一列或多列的筛选器。这允许对元数据表列进行精细的访问控制，从而限制集成服务的用户在运行查询时可以看到的列。此功能在使用 IAM 策略时，无法使用 IAM 策略。

还有一个名为的特殊权限Super. 这些区域有：Super权限使主体能够对授予权限的数据库或表执行所有受支持的 Lake Formation 操作。此权限可以与其他 Lake Formation 权限共存。例如，您可以授予权限Super、SELECT, 和INSERT在元数据表上。委托人可以对表执行所有受支持的操作，并且在您撤消时Super，SELECT和INSERT权限仍然存在。

有关每个权限的详细信息，请参阅Lake Formation 权限参考 (p. 167).

> **Important**
>
> 要能够查看由其他用户创建的数据目录表，您必须至少被授予一个对该表的 Lake Formation 权限。如果您被授予对该表的至少一项权限，则还可以查看该表的包含数据库。

您可以使用 Lake Formation 控制台、API 或Amazon Command Line Interface(Amazon CLI)。以下是的示例Amazon CLI授予用户权限的命令datalake_user1在中创建表的权限retail数据库。

```
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 权限补充了细粒度的访问控制。它允许对任何元数据数据库或表进行所有操作。

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

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

下一个例子也是粗粒度的，但限制性更强。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执行只读操作。

```
{
    "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) 元数据数据库中表子集的权限。

```
{
    "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 (p. 322).

底层数据访问控制

当一个集成Amazon服务请求访问由 Amazon S3 访问控制的 Amazon S3 位置中的数据Amazon Lake Formation，Lake Formation 提供临时凭证来访问数据。

要使 Lake Formation 能够控制对 Amazon S3 位置底层数据的访问，您可以注册Lake Formation Formation

注册 Amazon S3 位置后，您可以开始授予以下Lake Formation 权限：

- 数据访问权限 (`SELECT`、`INSERT`, 和`DELETE`)在指向该位置的数据目录表上。
- 该位置的数据位置权限。

Lake Formation 数据位置权限控制创建或更改指向特定 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额外的安全层。当你授予`CREATE_TABLE`要么`ALTER`权限时，您还可以授予数据位置权限，以限制承担者可以创建或更改元数据表的位置。

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

您可以使用 Lake Formation 控制台、API 或Amazon CLI. 授权的一般形式如下所示：

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

如果你包含`with grant option`，被授权者可以将权限授予其他委托人。

回想一下，Lake Formation 权限始终与Amazon Identity and Access Management用于精细访问控制的 (IAM) 权限。对于基础 Amazon S3 数据的读/写权限，IAM 权限按如下方式授予：

注册位置时，您需要指定一个 IAM 角色，该角色授予对该位置的读/写权限。Lake Formation 在为集成版提供临时证书时担任该角色Amazon服务。典型角色可能附加了以下策略，其中注册的位置是存储桶`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来自动创建类似这样的策略。有关更多信息，请参阅 对 Lake Formation 使用服务相关角色 (p. 256)。

因此，注册 Amazon S3 位置将授予所需的 IAM 权限`s3:`权限，其中权限由用于注册位置的角色指定。

> **Important**
>
> 避免注册具有以下功能的 Amazon S3 存储桶申请方付款已启用. 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。如果存储桶被其他人访问Amazon账户，如果该角色与存储桶拥有者属于同一账户，则向存储桶拥有者收取数据访问费用。

对于基础数据的读/写访问权限，除了 Lake Formation 权限外，委托人还需要以下 IAM 权限：

`lakeformation:GetDataAccess`

有了此权限，Lake Formation 会批准访问数据的临时凭证请求。

> **Note**
>
> 仅限Amazon Athena要求用户具有`lakeformation:GetDataAccess`权限。对于其他集成服务，代入的角色必须具有权限。

此权限包含在建议策略中的Lake Formation 角色和 IAM 权限参考 (p. 322).

总而言之，要使 Lake Formation 负责人能够读取和写入基础数据，其访问权限受 Lake Formation 权限控制：

- 包含数据的 Amazon S3 位置必须向 Lake Formation 注册。
- 创建指向基础数据位置的数据目录表的委托人必须具有数据位置权限。
- 读取和写入基础数据的委托人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。
- 读取和写入基础数据的委托人必须具有`lakeformation:GetDataAccess`IAM 权限。

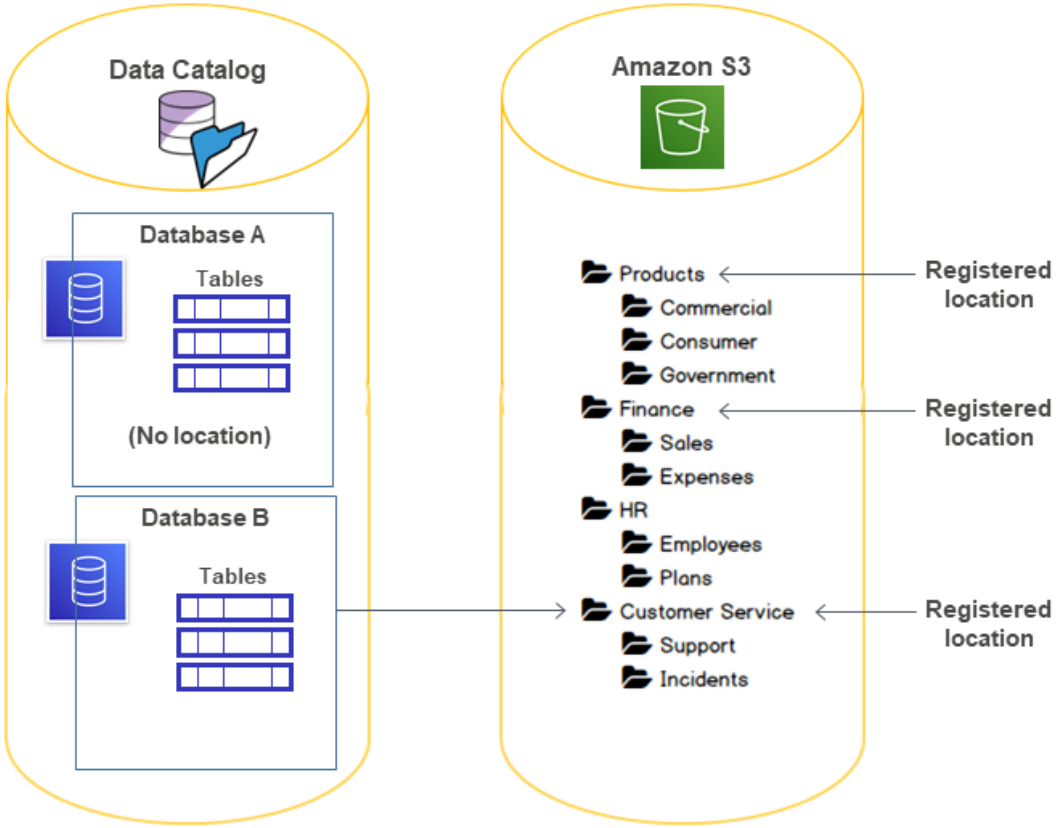
> **Note**
>
> 如果您通过 IAM 或 Amazon S3 策略可以访问 Amazon S3 位置，则Lake Formation 权限模型不会阻止通过 Amazon S3 API 或控制台访问 Amazon S3 位置。您可以将 IAM 策略附加到委托人以阻止此访问。

详细了解数据位置权限

数据位置权限控制对 Data Catalog 数据库和表执行的创建和更新操作的结果。规则如下所示：

- 委托人必须对 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Amazon CLI.
- 如果数据库具有指向注册位置的 location 属性，则会授予隐式权限，而主体具有`CREATE_TABLE`权限，并且委托人尝试在该位置或子位置创建表。
- 如果委托人被授予对某个位置的数据位置权限，则承担者对所有子位置都具有数据位置权限。
- 委托人不需要数据位置权限即可对基础数据执行读/写操作。只要有`SELECT`要么`INSERT`数据访问权限。数据位置权限仅适用于创建指向该位置的数据目录资源。

考虑下图中显示的场景。



在此示意图中：

- Amazon S3 iceProducts、Finance, 和Customer Service已向Lake Formation 注册。
- Database A没有位置属性，并且Database B有一个位置属性指向Customer Service存储桶。
- 用户datalake_user有CREATE_TABLE在两个数据库上。
- 用户datalake_user已被授予数据位置权限仅在Products存储桶。

以下是用户的结果datalake_user尝试在特定数据库中的特定位置创建目录表。

位置在哪里**datalake_user**尝试创建表

数据库和位置	成功或失败	Reason
数据库 AFinance/Sales	失败	没有数据定位权限
数据库 A，位于 A，Products	Succeed	具有数据定位权限
数据库 A，位于 A，HR/Plans	Succeed	位置未注册
数据库 B，位于 B，Customer Service/Incidents	Succeed	数据库的位置属性位于Customer Service

有关更多信息，请参阅下列内容：

- 将 Amazon S3 位置添加到您的数据湖 (p. 102)
- Lake Formation 权限参考 (p. 167)
- Lake Formation 角色和 IAM 权限参考 (p. 322)

Lake Formation 中的跨账户访问

Amazon Lake Formation允许对数据目录元数据和底层数据的跨账户访问。大型企业通常使用多个Amazon账户，其中许多账户可能需要访问由单个Amazonaccount. 用户和Amazon Glue提取、转换和加载 (ETL) 任务可以跨多个账户查询和联接表，并且仍然可以利用 Lake Formation 表级别和列级的数据保护。

主题

- 跨账户访问：如何运作 (p. 242)
- 跨账户访问权限前提条件 (p. 243)
- Lake Formation 基于标签的访问控制跨账户先决条件 (p. 244)
- 跨账户最佳实践和限制 (p. 245)
- 访问共享表的基础数据 (p. 246)
- 跨账户 CloudTrail 记录 (p. 247)
- 使用这两者来管理跨账户权限Amazon GlueLake Format (p. 250)
- 使用查看所有跨账户授权数 GetResourcePolicies API 操作 (p. 252)

相关主题

- 跨共享数据目录表和数据库Amazon账户 (p. 123)
- FormLake Formation 权限概述 (p. 138)
- 访问和查看共享数据目录表和数据库 (p. 123)
- 创建资源链接 (p. 126)
- 跨账户访问问题排除 (p. 328)

跨账户访问：如何运作

要启用跨账户访问，您可以将 Lake Formation 权限与数据目录表和数据库（数据目录资源）的授予选项授予外部Amazon账户、组织或组织单位。授权操作会自动共享这些资源。

您不与外部特定委托人共享资源Amazon账户-仅与账户共享资源。向组织或组织单位授予 Lake Formation 权限相当于向每个组织或组织单位授予权限Amazon该组织或组织单位的账户。

有关与共享数据的更多信息Amazon使用Lake Formation 组织，请参阅使用简化跨账户数据共享Amazon Lake Formation和Amazon起源.

Note

Lake Formation 基于标签的访问控制 (LF-TBAC) 不支持向组织和组织单位授予跨账户权限。

当您使用命名资源方法向外部账户授予 Lake Formation 对数据目录资源的权限时，Lake Formation 会使用 Amazon Resource Access Manager(Amazon RAM) 服务来共享资源。如果被授权者账户与授予者账户位于同一个组织中，则共享资源立即可供被授权者使用。如果被授权者账户不在同一个组织中，Amazon RAM向被授权者账户发送接受或拒绝资源授予的邀请。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Amazon RAM控制台或Amazon CLI以接受邀请。

Note

授予数据目录权限的 Lake Formation 基于标记的访问控制 (LF-TBAC) 方法不使用Amazon RAM用于跨账户授权。因此，跨账户授予可立即使用。有关更多信息，请参阅 Lake Formation 标签访问控制 (p. 179)。

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（带有可选的列筛选）
- 几个精选的桌子
- 数据库中的所有表（通过使用`All Tables`通配符）

在访问共享资源的每个账户中：

- 必须至少有一个用户是数据湖管理员。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 (p. 12).
- 数据湖管理员可以查看共享资源并将共享资源的 Lake Formation 权限授予账户中的其他委托人。其他委托人只有在数据湖管理员向他们授予对共享资源的权限后才能访问共享资源。由于数据湖管理员必须向被授权者账户中的委托人授予共享资源的权限，因此必须始终使用 grant 选项授予跨账户权限。
- 对于数据湖管理员和数据湖管理员已向其授予权限的委托人，共享资源在数据目录中显示为本地（拥有）资源。提取、转换和加载 (ETL) 作业可以访问共享资源的底层数据。
- 对于共享资源，表和数据库Lake Formation 控制台上的页面会显示所有者的账户 ID。
- 校长可以创建资源链接在他们的数据目录中转移到另一个共享资源中Amazonaccount. 集成服务，如 Amazon Athena和 Amazon Redshift Spectrum 要求资源链接能够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 (p. 126).
- 访问共享资源的基础数据时，Amazon CloudTrail日志事件在共享资源收件人的账户和资源所有者的账户中生成。这些区域有： CloudTrail 事件可以包含访问数据的委托人的 ARN，但前提是收件人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 记录 (p. 247)。

另请参见：

- 跨共享数据目录表和数据库Amazon账户 (p. 123)
- 访问和查看共享数据目录表和数据库 (p. 123)
- 授予和撤消对数据目录资源的权限 (p. 145)
- 创建资源链接 (p. 126)
- 是什么AmazonOrganizations中的AmazonOrganizations 用户指南

跨账户访问权限前提条件

在您的Amazon帐户可以共享数据目录数据库和表（数据目录资源），并且必须先满足以下先决条件，然后才能访问与您的帐户共享的资源：

- 如果您当前使用的是Amazon GlueData Catalog 资源策略如果要使用命名资源方法授予跨账户权限，则必须删除该策略或向其添加跨账户授予所需的新权限。如果您打算使用 Lake Formation 基于标记的访问控制 (LF-TBAC) 方法，则必须具有启用 LF-TBAC 的数据目录资源策略。有关更多信息，请参阅 使用这两者来管理跨账户权限Amazon GlueLake Format (p. 250) 和 Lake Formation 基于标签的访问控制跨账户先决条件 (p. 244)。
- 在授予对数据目录资源的跨账户权限之前，您必须撤消所有 Lake Formation 权限`IAMAllowedPrincipals`资源的组。
- 对于包含要共享的表的 Data Catalog 数据库，必须阻止新表的默认授予`Super`到`IAMAllowedPrincipals`. 在 Lake Formation 控制台上，编辑数据库并关闭仅对数据库中的

新表使用 IAM 访问控制. 或者，输入以下内容Amazon CLI命令，替换*<database>*与数据库的名称结合使用。

```
aws glue update-database --name <database> --database-input
 '{"Name":"<database>","CreateTableDefaultPermissions":[]}'
```

此外，对于希望外部帐户创建表的数据库，请确保禁用此设置。

- 如果要使用 named resources 方法与组织或组织单位共享数据目录资源，则必须在Amazon RAM.

  有关如何启用与组织共享的信息，请参阅启用与 共享AmazonOrganizations中的Amazon RAM用户指南.

  您必须具有`ram:EnableSharingWithAwsOrganization`允许与组织共享的权限。
- 如果您没有数据目录加密密钥的权限，则无法授予跨账户访问加密的数据库或表（在加密数据目录中创建）的权限。
- 想要使用 named resources 方法授予跨账户权限的用户必须具有必需的Amazon Identity and Access Management(IAM) 权限Amazon Glue和Amazon Resource Access Manager(Amazon RAM) 服务。这些区域有：Amazon管理的策略`AWSLakeFormationCrossAccountManager`授予所需的权限。

  如果账户中的数据湖管理员接收与命名资源方法共享的资源，则必须具有以下附加策略。它允许管理员接受Amazon RAM资源共享邀请。它还允许管理员启用与组织共享资源。

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

- 接收跨账户共享的账户必须具有`glue:PutResourcePolicy`接受的权限接受Amazon RAM资源共享邀请.
- 当你使用Amazon Gluecrawler 抓取另一个账户中的 Amazon S3 位置，并将生成的表保存在另一个账户的数据库中。S3 存储桶必须具有为 Crawler 角色授予对存储桶的权限的存储桶策略。

另请参阅

- Lake Formation 基于标签的访问控制跨账户先决条件 (p. 244)

Lake Formation 基于标签的访问控制跨账户先决条件

在使用 Lake Formation 基于标签的访问控制 (LF-TBAC) 方法授予对数据目录资源的跨账户访问权限之前，必须将以下 JSON 权限对象添加到Amazon Glue Data Catalog资源策略。你必须为每个Amazon您正在向其授予权限的账户。

要添加此代码，可以使用设置的 “” 页面Amazon Glue控制台，或者`glue:PutResourcePolicy`API 操作。

Replace*<recipient-account-id>*与Amazon收到补助金的账户，*<region>*数据目录的区域包含您正在授予权限的数据库和表，以及*<account-id>*使用您的Amazon账户 ID。

```
{
  "Effect": "Allow",
  "Action": [
    "glue:*"
  ],
  "Principal": {
    "AWS": [
      "<recipient-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
    }
  }
}
```

Note

资源策略中的所有代码必须位于`Statement`.

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

Important

如果您当前还通过使用 named resource 方法授予跨账户权限，则必须将`EnableHybrid`参数`'true'`当你调用`glue:PutResourcePolicy`API 操作。有关更多信息，请参阅 使用这两者来管理跨账户权限Amazon GlueLake Format (p. 250)。

另请参阅

- 使用这两者来管理跨账户权限Amazon GlueLake Format (p. 250)
- 元数据访问控制 (p. 236)

跨账户最佳实践和限制

以下是跨账户访问的最佳实践和限制：

- 你可以自己向校长发放的 Lake Formation 许可授予的数量没有限制Amazonaccount. 但是，Lake Formation 使用Amazon Resource Access Manager(Amazon RAM) 跨账户授予的容量，您的账户可以使用命名资源方法进行授权。为了最大限度地提高Amazon RAMcapacity 时，请针对命名资源方法遵循以下最佳实践：
  - 使用新的跨账户授权模式（版本 2下跨账户版本设置) 与外部共享资源Amazon Web Services 账户. 有关更多信息，请参阅 授予跨账户资源的权限 (p. 163)。
  - ArrangeAmazon帐户到组织，并向组织或组织单位授予权限。对组织或组织单位的补助金算作一次补助金。

    授予组织或组织单位也消除了接受Amazon Resource Access Manager(Amazon RAM) 授予的资源共享邀请。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 (p. 123)。

- 不要授予对数据库中许多单个表的权限，而是使用特殊的所有桌子通配符可授予对数据库中所有表的权限。授予所有桌子算作单笔补助金。有关更多信息，请参阅 授予和撤消对数据目录资源的权限 (p. 145)。

  Note

  有关请求提高资源共享数量限制的更多信息Amazon RAM，请参阅Amazon服务配额中的Amazon一般参考.

- 您必须创建指向共享数据库的资源链接，才能使该数据库显示在Amazon Athena和亚马逊 Redshift Spectrum 查询编辑器 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。然后，资源链接将出现在查询编辑器的表列表中。

  不必为许多单独的表创建资源链接以进行查询，您可以使用所有桌子通配符可授予对数据库中所有表的权限。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用于查询的所有表。有关更多信息，请参阅 创建资源链接 (p. 126)。
- Athena 和 Redshift Spectrum 支持列级访问控制，但仅限于包含，不排除。在中不支持列级别访问控制Amazon GlueETL 个作业。
- 当资源与您的共享时Amazon帐户，您可以仅向您账户中的用户授予对该资源的权限。您无法将资源的权限授予其他用户Amazon账户、组织（甚至不是你自己的组织）或`IAMAllowedPrincipals`组中)。
- 您不能授予`DROP`要么`Super`在数据库上转移到外部账户。
- 在删除数据库或表之前撤消跨账户权限。否则，您必须删除中的孤立资源共享Amazon Resource Access Manager.

  另请参阅

  - Lake Formation Tagion 访问控制说明和限制 (p. 186)
  - `CREATE_TABLE` (p. 172)中的Lake Formation 权限参考 (p. 167)了解更多跨账户访问规则和限制。

访问共享表的基础数据

假定选择Amazon账户 A 与账户 B 共享数据目录表-例如，通过授予`SELECT`将表上的授予选项授予账户 B。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该共享。（如果账户 A 和 B 位于同一个组织中，或者如果授予是使用基于 Lake Formation 标签的访问控制方法进行的，则不需要这样做。）
- 数据湖管理员必须将Lake Formation 重新授予委托人`SELECT`账户 A 在共享表上授予的权限。
- 委托人必须对表、包含它的数据库和账户 A Data Catalog 具有以下 IAM 权限。

  Note

  在以下 IAM 策略中：

  - Replace*<account-id-A>*用Amazon账户 A 的账户 ID
  - Replace*<region>*具有有效的区域。
  - Replace*<database>*使用账户 A 中包含共享表的数据库的名称。
  - Replace*<table>*与 Policy 共享表的名称为共享表名。

```
{
    "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": [
            "arn:aws:lakeformation:<region>:<account-id-A>:catalog:<account-id-A>"
           ],
          "Condition": {
            "StringEquals": {
              "lakeformation:GlueARN":"arn:aws:glue:<region>:<account-id-
A>:table/<database>/<table>"
            }
        }
    }
   ]
}
```

另请参见：

- 接受来自的资源共享邀请Amazon RAM (p. 124)

跨账户 CloudTrail 记录

Lake Formation 提供对数据湖中数据的所有跨账户访问的集中审核跟踪。当收件人Amazon账户访问共享表中的数据，Lake Formation 会复制 CloudTrail 事件到拥有账户的 CloudTrail 日志。复制的事件包括集成服务对数据的查询，例如Amazon Athena和 Amazon Redshift Spectrum，以及数据访问权限Amazon Glue个作业。

CloudTrail 类似地复制数据目录资源上跨账户操作的事件。

作为资源所有者，如果您在 Amazon S3 中启用了对象级日志记录，则可以运行加入 S3 的查询 CloudTrail Lake Formation 事件 CloudTrail 事件，以确定已访问您的 S3 存储桶的账户。

主题

- 在跨账户中包含委托人身份 CloudTrail 圆木 (p. 247)
- 查询 CloudTrail Amazon S3 跨账户访问的日志 (p. 249)

在跨账户中包含委托人身份 CloudTrail 圆木

默认情况下，跨账户 CloudTrail 添加到共享资源收件人日志并复制到资源所有者日志中的事件仅包含Amazon外部账户委托人 ID，而不是委托人（委托人 ARN）的 Amazon Resource Name (ARN)。在受信任的边界内（例如在同一个组织或团队中）共享资源时，ARN 选择在 CloudTrail 事件. 然后，资源所有者账户可以跟踪访问其自有资源的接收者账户中的委托人。

Important

作为共享资源接收者，在您自己的事件中查看委托人 ARN CloudTrail 日志，您必须选择与所有者账户共享委托人 ARN。
如果通过资源链接进行数据访问，则在共享资源接收者帐户中记录两个事件：一个用于资源链接访问，另一个用于目标资源访问。资源链接访问的事件确实选择包括委托人 ARN。目标资源访问的事件不包括没有选择加入的委托人 ARN。资源链接访问事件不会复制到所有者账户。

以下是来自默认跨账户的摘录 CloudTrail 事件（没有选择加入）。执行数据访问权限的账户为 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，当您选择包含委托人 ARN 时，摘录将变为以下内容。这些区域有：`lakeFormationPrincipal`字段表示通过亚马逊 Athena、Amazon Redshift Spectrum 或Amazon Glue个作业。

```
{
    "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"
    },
...
}
```

选择在跨账户中包含主要 ARN CloudTrail 圆木

1. 通过 https://console.aws.amazon.com/lakeformation/ 中打开 Lake Formation 控制台。

   以登录`Administrator`用户，或具有`AdministratorAccess`IAM 策略。
2. 在导航窗格中，选择 Settings (设置)。
3. 在存储库的数据目录设置页, 在的默认权限Amazon CloudTrail部分, 对于资源拥有者，输入一个或多个Amazon资源所有者账户 ID。

按Enter在每个账户 ID 之后。

4. 选择Save（保存）。

现在跨账户 CloudTrail 存储在共享资源接收者和资源所有者的日志中的事件包含委托人 ARN。

查询 CloudTrail Amazon S3 跨账户访问的日志

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 注册的 S3 地点。如果共享资源使用者选择将主要 ARN 包含在 Lake Formation 中 CloudTrail 日志，您可以确定访问存储桶的角色或用户。

使用运行查询时Amazon Athena，你可以加入Lake Formation CloudTrail 事件和 S3 CloudTrail 会话名称属性上的事件。查询还可以在上过滤Lake Formation 事件`eventName="GetDataAccess"`，S3 事件开启`eventName="GetObject"`要么`eventName="PutObject"`.

以下是来自Lake Formation 交叉账户的摘录 CloudTrail 访问已注册 S3 位置中的数据的事件。

```
{
  "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"
   }
}
```

这些区域有：`lakeFormationRoleSessionName`键值，`AWSLF-00-GL-111122223333-B8JSAjo5QA`，可以使用会话名称在`principalId`S3 的键 CloudTrail event. 下面是来自 S3 的摘录 CloudTrail event. 它显示了会话名称的位置。

```
{
   "eventSource": "s3.amazonaws.com",
   "eventName": "GetObject"
   ..............
   ..............
   "principalId": "AROAQSOX5XXUR7D6RMYLR:AWSLF-00-GL-111122223333-B8JSAjo5QA",
   "arn": "arn:aws:sts::111122223333:assumed-role/AWSServiceRoleForLakeFormationDataAccess/
AWSLF-00-GL-111122223333-B8JSAjo5QA",
   "sessionContext": {
     "sessionIssuer": {
       "type": "Role",
       "principalId": "AROAQSOX5XXUR7D6RMYLR",
       "arn": "arn:aws:iam::111122223333:role/aws-service-role/lakeformation.amazonaws.com/
AWSServiceRoleForLakeFormationDataAccess",
       "accountId": "111122223333",
       "userName": "AWSServiceRoleForLakeFormationDataAccess"
     },
   ..............
   ..............
}
```

会话名称的格式如下：

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

version-number

该格式的版本，目前00. 如果会话名称格式发生变化，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	Amazon GlueETL 作业
AT	Athena
RE	Amazon Redshift Spectrum

account-id

这些区域有：Amazon从 Lake Formation 请求凭证的账户 ID。

suffix

随机生成的字符串。

使用这两者来管理跨账户权限Amazon GlueLake Format

您可以使用以下任一方法来授予对数据目录资源和底层数据的跨账户访问权限Amazon Glue要么Amazon Lake Formation.

InAmazon 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，你可以使用Lake Formation 授予跨账户权限GRANT/REVOKE权限模型和GrantPermissionsAPI 操作。

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 控制台查看 Lake Formation 跨账户授予；对于使用命名资源方法发放的授予，你可以使用Amazon Resource Access Manager(Amazon RAM) 控制台。但是，这些控制台页面不显示由 Amazon GlueData Catalog 资源策略。同样，您可以使用 “数据目录” 资源策略查看跨账户授权设置的 “” 页面Amazon Glue控制台，但该页面不显示使用 Lake Formation 授予的跨账户权限。

为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Amazon Glue要求你执行以下操作，以表明你知道并允许 Lake Formation 和Amazon Glue.

使用Amazon Glue数据目录资源策略

如果您的账户没有使用 named resources 方法进行跨账户授权、哪个用s Amazon RAM要共享资源，您可以像往常一样将数据目录资源策略保存在Amazon Glue. 但是，如果补助金涉及Amazon RAM已创建资源共享，您必须执行以下操作之一，以确保成功保存资源策略：

- 当您将资源策略保存在设置的 “” 页面Amazon Glue控制台时，控制台会发出提示，指明策略中的权限不属于使用这Lake Formation控制台. 你必须选择继续以保存策略。
- 当您使用保存资源策略时glue:PutResourcePolicyAPI 操作时，您必须设置EnableHybrid的成员PutResourcePolicyRequest结构改为 'TRUE'（类型 = 字符串）。以下代码示例演示了如何在 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": "AllowConsumerFullCatalogAccess",
            "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_str = json.dumps(policy)
glue.put_resource_policy(PolicyInJson=policy_str, EnableHybrid='TRUE')
```

有关更多信息，请参阅 。PutResourcePolicy 操作（Python：put_pout_pout_policy）中的Amazon Glue开发人员指南.

使用这Lake Formation命名资源方法

如果你的账户中没有数据目录资源策略，Lake Formation 跨账户授予你照常进行的授权。但是，如果 Data Catalog 资源策略存在，则必须向其中添加以下语句以允许跨账户授权才能成功如果它们是用命名的资源方法创建的. Replace*<region>*使用有效的区域名称和*<account-id>*使用您的Amazon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"
    ]
  }
```

如果没有此附加声明，Lake Formation 补助金将成功，但会被封锁Amazon RAM，并且收件人账户无法访问授予的资源。

> **Important**
>
> 如果您还使用基于 Lake Formation 标记的访问控制 (LF-TBAC) 方法进行跨账户授权，则必须具有至少具有中指定的权限的数据目录资源策略Lake Formation 基于标签的访问控制跨账户先决条件 (p. 244).

另请参见：

- 元数据访问控制 (p. 236)（有关命名资源方法与基于Lake Formation 标签的访问控制（LF-TBAC）方法的讨论）。
- 查看共享数据目录表和数据库 (p. 125)
- 在Amazon Glue控制台中的Amazon Glue开发人员指南
- 授予跨账户访问权限中的Amazon Glue开发人员指南（对于示例数据目录资源策略）

使用查看所有跨账户授权数 GetResourcePolicies API 操作

如果您的企业同时使用Amazon Glue Data Catalog资源策略和 Lake Formation 补助金，在一个地方查看所有跨账户补助金的唯一方法是使用`glue:GetResourcePolicies`API 操作。

当你跨账户授予 Lake Formation 权限时通过使用命名资源方法、Amazon Resource Access Manager(Amazon RAM) 创建一个Amazon Identity and Access Management(IAM) 资源策略并将其存储在您的Amazonaccount. 该策略授予访问资源所需的权限。Amazon RAM为每个跨账户授权创建单独的资源策略。您可以使用以下方法查看所有这些策略`glue:GetResourcePolicies`API 操作。

Note

此操作还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Amazon KMS密钥时，操作将不会返回数据目录资源策略。

要查看所有跨账户授权数

- 输入以下内容Amazon CLI命令。

```
aws glue get-resource-policies
```

以下是资源策略的示例Amazon RAM在您授予对表的权限时创建和存储`t`在数据库中`db1`到Amazon账户1111-22333333。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
         "glue:GetTable",
         "glue:GetTableVersion",
         "glue:GetTableVersions",
         "glue:GetPartition",
         "glue:GetPartitions",
         "glue:BatchGetPartition",
         "glue:GetTables",
         "glue:SearchTables"
       ],
      "Principal": {"AWS": [
        "111122223333"
      ]},
      "Resource": [
        "arn:aws:glue:<region>:111122223333:table/db1/t"
     ]
    }
  ]
}
```

另请参见：

- GetResourcePolicies 操作（Python：get_policy）中的Amazon Glue开发人员指南

AmazonLake Formation 托管策略

你可以授予Amazon Identity and Access Management(IAM) 权限（IAM）权限（IAM）权限Amazon Lake Formation通过使用Amazon托管策略与内联策略。以下Amazon托管策略可用于 Lake Formation 策略。

主体	Amazon 托管策略	描述
Lake Formation 用户，包括数据湖管理员	`AWSGlueConsoleFullAccess`	允许主体在 Lake Formation 控制台上执行各种操作。
数据湖管理员	`AWSLakeFormationDataAdmin`	允许数据湖管理员执行管理操作和查看Amazon CloudTrail日志。
Lake Formation 用户，包括数据湖管理员	`AWSLakeFormationCrossAccountManager`	允许委托人向外部授予 Lake Formation 权限Amazon账户、组织或组织单位。

Note

这些区域有：`AWSLakeFormationDataAdmin`策略不会向数据湖管理员授予所有必需的权限。创建和运行工作流以及向服务关联角色注册位置需要其他权限`AWSServiceRoleForLakeFormationDataAccess`. 有关更多信息，请参阅 创建数据湖管理员 (p. 12) 和 对 Lake Formation 使用服务相关角色 (p. 256)。

此外，Amazon GlueLake Formation 承担了服务角色`AWSGlueServiceRole`允许访问相关服务，包括Amazon EElastic Compute Cloud (Amazon EC2) Elastic Compute Compute Cloud)、Amazon S CloudWatch.

Lake Formation 更新Amazon托管策略

查看有关更新的详细信息Amazon从该服务开始跟踪这些更改开始，该服务开始为 Lake Formation 策略。

更改	说明	日期
Lake Formation 更新`AWSLakeFormationCrossAccountManager`策略。	Lake Formation 对AWSLakeFormationCrossAccountManager策略在首次共享资源时，每个收件人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个资源共享。	2022 年 5 月 6 日
Lake Formation 已开始跟踪变化。	Lake Formation 开始为其Amazon托管策略。	2022 年 5 月 6 日

更改数据湖的默认安全设置

保持向后兼容性Amazon Glue、Amazon Lake Formation具有以下初始安全设置：

- 这些区域有：`Super`已向该组授予权限`IAMAllowedPrincipals`在所有现有的Amazon Glue数据目录资源。

- 已为新的数据目录资源启用 “仅使用 IAM 访问控制” 设置。

这些设置实际上导致对数据目录资源和 Amazon S3 位置的访问仅由以下人员控制Amazon Identity and Access Management(IAM) 策略。单个Lake Formation 权限无效。

这些区域有：`IAMAllowedPrincipals`组包括您的 IAM 策略允许访问您的数据目录资源的任何 IAM 用户和角色。这些区域有：`Super`权限使主体能够对授予权限的数据库或表执行所有受支持的 Lake Formation 操作。

要更改安全设置，以便通过 Lake Formation 权限管理对数据目录资源（数据库和表）的访问，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 更改默认权限模型 (p. 14)。
2. 更改现有数据目录资源的设置。有关说明，请参阅 升级Amazon Glue的数据权限Amazon Lake Formation 模型 (p. 22)。

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

也可以使用Lake Formation 更改默认安全设置PutDataLakeSettings操作（Python：put_data_lake_settings）(p. 282). 此操作将可选的目录 ID 和DataLake设置结构 (p. 281).

要强制执行 Lake Formation 对新数据库和表的元数据和基础数据访问控制，请将`DataLakeSettings`结构如下。

> **Note**
>
> Replace*`<AccountID>`*使用有效的Amazon账户 ID 和*`<Username>`*具有有效的 IAM 用户名。您可以指定多个用户作为数据湖管理员。

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

您也可以按如下方式对结构进行编码。忽略`CreateDatabaseDefaultPermissions`要么`CreateTableDefaultPermissions`参数等效于传递空列表。

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

此操作会有效地撤Lake Formation 来自`IAMAllowedPrincipals`对新数据库和表进行分组。在创建数据库时，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，请将`DataLakeSettings`结构如下。

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

这授予SuperLake Formation 许可IAMAllowedPrincipals对新数据库和表进行分组。在创建数据库时，您可以覆盖此设置。

> Note
>
> 在以上的DataLakeSettingsstructure，唯一允许的值DataLakePrincipalIdentifier是IAM_ALLOWED_PRINCIPALS，并且是唯一允许的值Permissions是ALL.

权限示例方案

以下场景有助于演示如何设置权限以保护对中数据的访问Amazon Lake Formation.

Shirley 是一名数据管理员。她想为自己的公司建立一个数据湖， AnyCompany. 目前，所有数据都存储在 Amazon S3 中。John 是一名营销经理，需要写入客户购买信息（包含在s3://customerPurchases)。营销分析师迭戈今年夏天加入了约翰的行列。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。

总结一下：

- Shirley 是数据湖管理员。
- John 要求CREATE_DATABASE和CREATE_TABLE在数据目录中创建新数据库和表的权限。
- John 还需要SELECT、INSERT, 和DELETE他创建的表的权限。
- Dieg 需要SELECT对表的权限以运行查询。

的员工 AnyCompany 执行以下操作以设置权限。为了清楚起见，本场景中显示的 API 操作显示了简化的语法。

1. Shirley 向 Lake Formation 注册了包含买家购买信息的 Amazon S3 路径。

```
RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
```

2. Shirley 授予了 John 访问包含买家购买信息的 Amazon S3 路径的访问权限。

```
GrantPermissions(John, S3Location("s3://customerPurchases"), [DATA_LOCATION_ACCESS]) )
```

3. Shirley 授予 John 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`CREATE_TABLE`因为他创建了这个数据库。

```
CreateDatabase(John_DB)
```

5. John 创建表`John_Table`指向`s3://customerPurchases`. 因为他创建了表，所以他拥有该表的所有权限，并且可以授予对表的权限。

```
CreateTable(John_DB, John_Table)
```

6. John 允许他的分析师 Diego 进入桌子`John_Table`.

```
 GrantPermissions(Diego, John_Table, [SELECT])
```

安全事件登录Amazon Lake Formation

AmazonLake Formation 与Amazon CloudTrail，提供用户、角色或者执行操作的记录的服务Amazon在 Lake Formation 服务。CloudTrail 将所有对 Lake Formation 的 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，Amazon Command Line Interface，并且包含 Lake Formation API 操作的代码调用。

有关 Lake Formation 中事件记录的更多信息，请参阅日志系统AmazonLake Formation API 调用使用 Amazon CloudTrail (p. 265).

> Note
>
> `GetTableObjects`、`UpdateTableObjects`, 和`GetWorkUnitResults`是大量数据层面操作。对这些 API 的调用目前尚未记录到 CloudTrail。有关 CloudTrail 中数据层面操作的更多信息，请参阅记录跟踪的数据事件中的Amazon CloudTrail用户指南.
> 为支持其他 CloudTrail 活动而进行的 Lake Formation 变将记录在Amazon Lake Formation 的文档历史记录 (p. 335).

对 Lake Formation 使用服务相关角色

Amazon Lake Formation使用Amazon 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。服务相关角色由 Lake Formation 预定义，并具有该服务调用其他服务所需的一切权限。Amazon服务代表您。

通过使用服务相关角色，您可以更轻松地设置 Lake Formation，因为您不必创建角色和手动添加所需的权限。Lake Formation 定义其服务相关角色的权限，除非另外定义，否则，仅 Lake Formation 可以担任其角色。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。

此服务相关角色信任以下服务代入该角色：

- lakeformation.amazonaws.com

Lake Formation 的服务相关角色权限

Lake Formation 使用名为的服务相关角色AWSServiceRoleForLakeFormationDataAccess. 此角色提供一组 Amazon Simple Storage Service (Amazon S3) 权限，这些权限可启用 Lake Formation 集成服务（例如）。Amazon Athena访问注册的位置。注册数据湖位置时，您必须提供一个角色，该角色在该位置具有所需的 Amazon S3 读/写权限。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 为内联策略添加了路径，并将其附加到服务相关角色。当您使用服务相关角色注册后续路径时，Lake Formation 会将路径添加到现有策略中。

以数据湖管理员身份登录时，请注册数据湖位置。然后，在 IAM 控制台中搜索角色AWSServiceRoleForLakeFormationDataAccess然后查看其附加的政策。

例如，在注册该位置之后s3://my-kinesis-test/logs，Lake Formation 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"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test/logs/*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test"
            ]
        }
    ]
}
```

要使用此服务相关角色注册位置，需要以下权限：

- iam:CreateServiceLinkedRole
- iam:PutRolePolicy

数据湖管理员通常具有这些权限。

将第三方服务与集成Lake Formation

与 集成Amazon Lake Formation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 作为授权引擎来管理或强制执行对集成数据湖的权限Amazon例如Amazon、Amazon Athena、Amazon SpectRedshift m 等服务。Lake Formation 为整合服务提供了两种选择：

1. Lake Formation 证书自动售货引擎：Lake Formation 可以以以下形式出售范围缩小的临时证书Amazon根据有效权限将令牌发送到已注册的 Amazon S3 位置，以便授权引擎可以代表用户访问数据。
2. 中央强制执行：Lake Formation查询 API (p. 310)操作从 Amazon S3 检索数据并筛选结果。与查询 API 操作集成的引擎或应用程序可以依靠 Lake Formation 来评估调用身份的权限，并根据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过滤后的数据。

主题

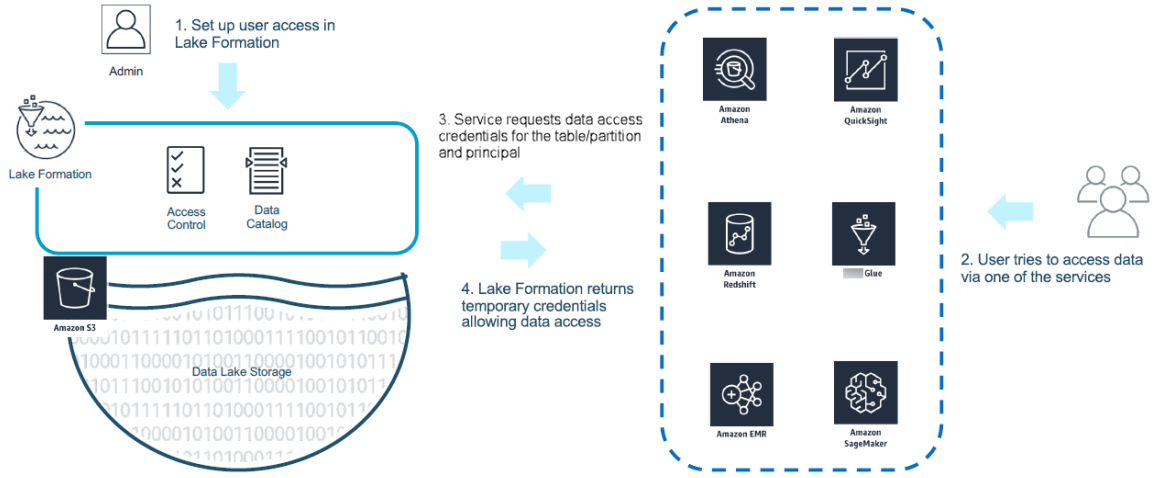
使用Lake Formation 凭证自动售货机

Lake Formation 允许第三方服务通过使用凭证自动售出 API 操作与 Lake Formation 集成。这允许第三方服务使用与其他服务相同的授权和凭证自动售货引擎Amazon分析服务的使用。本节介绍如何使用凭证售卖 API 操作将第三方查询引擎与Lake Formation.

主题

Lake Formation 证书自动售的工作原

本节介绍如何使用凭证售卖 API 操作将第三方应用程序（查询引擎）与Lake Formation.



1. 这些区域有：Lake Formation管理员执行以下活动：
   - 在 Lake Formation 中注册 Amazon S3 地点，方法是提供一个 IAM 角色（用于出售证书），该角色具有访问该 Amazon S3 位置内的数据的相应权限
   - 注册第三方应用程序，以便能够调用 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” (p. 260)。
   - 授予用户访问数据库和表的权限

     例如，如果您想发布用户会话日期集，其中包含一些包含个人身份信息 (PII) 的列，为了限制访问权限，您可以为这些列分配一个LF-TBAC名为 “分类” 的标签，值为 “敏感”。接下来，定义一个权限，允许业务分析师访问用户会话数据，但排除那些标有用户会话数据的列分类 = 敏感.
2. 委托人（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，要求提供表格信息和访问表的凭证。
4. 如果查询主体被授权访问表，则 Lake Formation 会将凭证返回给允许数据访问的集成应用程序。
5. 集成服务从 Amazon S3 读取来自 Amazon S3 的数据，然后将结果返回给委托人。

> **Important**
>
> Lake Formation凭证售卖 API 操作启用分布式强制执行，采用显式故障拒绝（失效关闭）模型。这引入了客户、第三方服务和Lake Formation之间的三方安全模式。集成服务值得信赖，可以正确执行Lake Formation权限（分布式强制执行）。

集成服务负责根据从 Amazon S3 返回的策略，筛选从 Amazon S3 读取的数据Lake Formation在筛选的数据返回给用户之前。集成服务遵循失效关闭模式，这意味着如果它们无法强制执行必需的操作，则必须无法通过查询Lake Formation权限。

Lake Formation 凭证售卖中的角色和职责

角色	责任
客户	• 启用 Lake Formation 凭证自动售卖the section called “注册第三方查询引擎” (p. 260))。 • 向 Lake Formation 明确注册经批准的第三方（参见the section called “注册第三方查询引擎” (p. 260))。 • 测试和验证具有 Lake Formation 权限的第三方解决方案。 • 监控和审核第三方对 Lake Formation 凭证自动售货 API 操作的使用。
第三方	• 公开记录每个软件修订版支持的功能，并提供正确启用该功能的说明。 • 在调用 Lake Formation 凭证自动售货 API 操作时，准确宣传支持的功能（根据文档）。 • 安全地存储和处理已发放的证书，以避免凭证泄露和权限升级。 • 根据支持的功能强制执行权限，仅向用户返回筛选后的数据 • 无法正确执行所需权限时查询失败
`Amazon Lake Formation`	• 正确派生并返回给定委托人的有效权限。 • 验证 API 操作中第三方支持的功能 call-by-call 基础。 • 仅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 证书，否则返回错误。

Lake Formation凭证自动售卖 API 操作的工作流程

以下是凭证自动售卖 API 操作的工作流程：

1. 用户使用集成的第三方查询引擎提交查询或数据请求。查询引擎担任代表用户或一组用户的 IAM 角色，并检索调用凭证售卖 API 操作时使用的可信证书。
2. 查询引擎调用`GetUnfilteredTableMetadata`，如果是分区表，则查询引擎会调用`GetUnfilteredPartitionsMetadata`从数据目录中检索元数据和策略信息。
3. Lake Formation 对请求进行授权。如果用户在表上没有相应的权限，那么AccessDeniedException被抛出。
4. 作为请求的一部分，查询引擎发送其支持的筛选条件。在数组内可以发送两个标志：COLUMN_权限和CELL_FILTER_权限. 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，那么PermissionTypeMismatchException被抛出，查询失败。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。如果授权列列表为空，则表示用户有`DESCRIBE`权限，但没有`SELECT`权限，查询失败。
   - 一面旗帜，`IsRegisteredWithLakeFormation`，这表明 Lake Formation 能否向该资源数据发送证书。如果返回 false，则应使用客户的凭证访问 Amazon S3。
   - 列表`CellFilters`如果有的话，应该应用于数据行。此列表包含列和用于计算每行的表达式。只有在以下情况下才应填充此值CELL_FILTER_权限是作为请求的一部分发送的，并且有一个针对表的数据过滤器，供调用用户使用。
6. 检索到元数据后，查询引擎调用`GetTemporaryGlueTableCredentials`要么`GetTemporaryGluePartitionCredentials`来获取Amazon用于从 Amazon S3 位置检索数据的证书。
7. 查询引擎从 Amazon S3 读取相关对象，根据在步骤 2 中收到的策略筛选数据，并将结果返回给用户。

的凭证自动售卖 API 操作Lake Formation包含用于配置与第三方查询引擎集成的其他内容。您可以在中查看操作详情凭证售卖 API 操作部分。 (p. 283)

注册第三方查询引擎

在第三方查询引擎可以使用凭证售卖 API 操作之前，您需要明确启用查询引擎代表您调用 API 操作的权限。这是通过几个步骤完成的：

1. 你需要指定Amazon账户和 IAM 会话标签，需要获得权限才能通过以下方式调用证书销售 API 操作Amazon Lake Formation控制台、Amazon CLI或 API/开发工具包。
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代表第三方引擎的 Lake Formation 中注册的会话标签。Lake Formation使用此标签来验证请求是否来自已批准的引擎。有关会话标签的更多信息，请参阅会话标签(在 IAM 用户指南中)。
3. 在设置第三方查询引擎执行角色时，您必须在 IAM 策略中拥有以下最低权限集：

```
{
  "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. 在查询引擎执行角色上设置角色信任策略，以便对可以将哪个会话标签密钥值对附加到该角色进行精细的访问控制。在以下示例中，仅允许此角色拥有会话标签密钥"LakeFormationAuthorizedCaller"和会话标签值"engine1"待附加，并且不允许其他会话标签密钥值对。

```
{
    "Sid": "AllowPassSessionTags",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
        "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"        }
    }
}
```

何时LakeFormationAuthorizedCaller称之为 STS：AssumeRole API 操作要获取证书供查询引擎使用，会话标签必须包含在 AssumeRole 请求. 返回的临时证书可用于制作Lake Formation凭证售卖 API 请求。

Lake Formation凭证销售 API 操作要求调用委托人成为 IAM 角色。IAM 角色必须包含已注册的具有预定值的会话标签Lake Formation. 这个标签允许Lake Formation以验证用于调用证书发放 API 操作的角色是否被允许这样做。

为第三方查询引擎启用权限以调用凭证自动售卖 API 操作

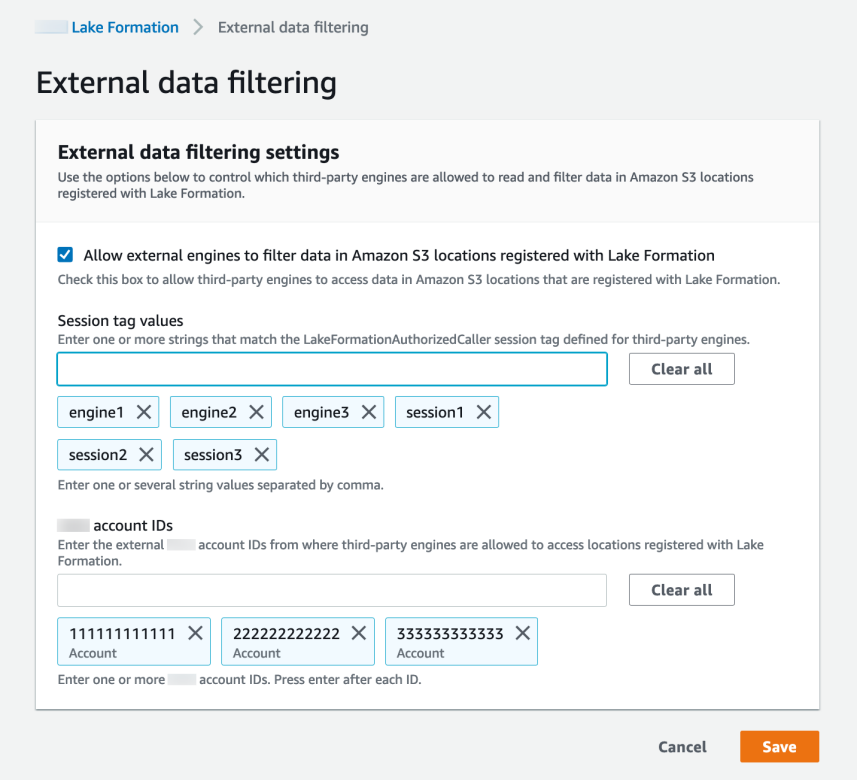
按照以下步骤允许第三方查询引擎通过调用凭证售卖 API 操作Amazon Lake Formation控制台、Amazon CLI 或 API/开发工具包。

Console

要注册您的账户以进行外部数据筛选，请执行以下操作：

1. 登录到Amazon Web Services Management Console，然后在处打开 Lake Formation 控制台https://console.aws.amazon.com/lakeformation/.
2. 在左侧导航中，展开Permissions (权限)，然后选择外部数据筛选.
3. 在存储库的外部数据筛选页面，选择选项允许外部引擎筛选已注册的 Amazon S3 位置的数据Lake Formation.
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，请参阅在 中传递会话标签AmazonSTS中的Amazon Identity and Access Management用户指南.
5. 输入可以使用第三方引擎访问未经过滤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Amazon用于配置跨账户访问权限的账户 ID 字段。



CLI

使用`put-data-lake-settings`CLI 命令用于设置以下参数。

使用此字段时需要配置三个字段Amazon CLI命令：

- `allow-external-data-filtering` —（布尔值）表示第三方引擎可以访问当前账户中资源的未经过滤的元数据信息和数据访问凭证。
- `external-data-filtering-allow-list`—（数组）账户 ID 列表，在使用第三方引擎时，这些账户可以访问当前账户中资源的未过滤元数据信息和数据访问凭证。
- `authorized-sessions-tag-value-list`—（数组）授权会话标签值（字符串）的列表。如果 IAM 角色证书已附加了授权的键值对，则如果该会话标签包含在列表中，则会话被授予访问已配置账户中资源的未筛选元数据信息和数据访问凭证的权限。授权会话标签密钥定义为`*LakeFormationAuthorizedCaller*`.

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"]
    }
}
```

API/SDK

使用`PutDataLakeSetting`API 操作用于设置以下参数。

使用此 API 操作时，需要配置三个字段：

- `AllowExternalDataFiltering`— (布尔值) 表示第三方引擎是否可以访问当前账户中资源的未过滤元数据信息和数据访问凭证。
- `ExternalDataFilteringAllowList`— (数组) 账户 ID 列表，可使用第三方引擎访问当前账户中未经过滤的元数据信息和资源的数据访问凭证。
- `AuthorizedSectionsTagValueList`— (数组) 授权标签值列表 (字符串)。如果 IAM 角色证书已附加了授权标签，则会话被授予访问未经过滤的元数据信息的权限，以及配置账户中资源的数据访问证书。授权会话标签密钥定义为`*LakeFormationAuthorizedCaller*`.

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));
}
```

Amazon GlueLake Formation 中的功能

Amazon Lake Formation是基于构建的。Amazon Glue，并且服务通过以下方式进行交互：

- Lake FormationAmazon Glue共享同一数据目录。
- 以下 Lake Formation 控制台功能调用Amazon Glue控制台：
  - 作业-有关更多信息，请参阅。添加作业中的Amazon Glue开发人员指南.
  - Crawers-有关更多信息，请参阅。使用爬网程序编目表中的Amazon Glue开发人员指南.
- 使用 Lake Formation 蓝图时生成的工作流如下：Amazon Glue工作流。您可以在 Lake Formation 控制台和Amazon Glue控制台。
- Lake Formation 一起提供机器学习转型，并基于Amazon GlueAPI 操作。您可以在上创建和管理机器学习转换。Amazon Glue控制台。有关更多信息，请参阅 。机器学习转换中的Amazon Glue开发人员指南.

日志系统AmazonLake Formation API调用使用Amazon CloudTrail

AmazonLake Formation 与Amazon CloudTrail，提供用户、角色或用户所执行操作的记录的服务。Amazon在 Lake Formation 服务。CloudTrail 将所有 Lake Formation API 调用捕获为事件。捕获的调用包括来自 Lake Formation 控制台的调用。Amazon Command Line Interface，以及对 Lake Formation API 操作的代码调用。如果您创建跟踪记录，则可以使 CloudTrail 事件持续传送到 Amazon S3 存储桶（包括 Lake Formation）的事件。如果您不配置跟踪记录，则仍可在 CloudTrail 控制台中的 Event history（事件历史记录）中查看最新事件。使用 CloudTrail 收集的信息，您可以确定向 Lake Formation 发出了什么请求、发出请求的 IP 地址、请求的发出时间以及其他详细信息。

要了解有关 CloudTrail 的更多信息，请参阅 Amazon CloudTrail 用户指南。

CloudTrail 中的 Lake Formation 信息

当您创建新的 CloudTrail 时，默认处于启用状态。Amazonaccount. 当 Lake Formation 中发生活动时，该活动将记录为 CloudTrail 事件和其他事件。Amazon中的服务事件历史记录. 一个事件表示来自任何源的一个请求，包括有关所请求的操作、操作的日期和时间、请求参数等方面的信息。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 Amazon 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Amazon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 Amazon 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。

要持续记录您的事件Amazon账户（包括 Lake Formation）的事件。通过跟踪记录，CloudTrail 可将日志文件传送至 Amazon S3 存储桶。在控制台创建跟踪时，跟踪默认应用于所有 Amazon 区域。此跟踪记录在 Amazon 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 Amazon S3 存储桶。此外，您还可以配置其他Amazon服务，例如Amazon Athena，进一步分析在 CloudTrail 日志中收集的事件数据并采取措施。CloudTrail 还可以将日志文件传送到 Amazon CloudWatch Logs 和 CloudWatch 事件。

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

CloudTrail 记录所有 Lake Formation API 操作，并记录在Amazon Lake Formation开发人员指南 的第一个版本。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 CloudTrail 了用于GrantPermissionsaction. 该条目包括授予权限的用户（datalake_admin），授予该权限的委托人（datalake_user1），以及授予的许可（CREATE_TABLE)。该条目还显示授权失败是因为目标数据库未在resource参数。


```
{
    "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"
}
```


下一个示例显示的是用于GetDataAccessaction. 委托人不会直接调用此 API。相反，GetDataAccess每当委托人或集成时都会记录Amazon服务请求临时证书以访问已注册到 Lake Formation 的数据湖位置中的数据。


```
{
    "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 记录 (p. 247)

Amazon Lake Formation API

目录

API 权限

权限 API 介绍与授予和撤消权限有关的数据类型和操作Amazon Lake Formation.

数据类型

资源结构

资源的结构。

字段

- `Catalog`— 一个名为的空结构`CatalogResource`.

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。
- `Database` – 一个 DatabaseResource (p. 272) 对象。

  资源的数据库。对数据目录是唯一的。数据库是一组关联的表定义，这些定义组织到一个逻辑组内。您可以向委托人授予和撤消数据库权限。
- `Table` – 一个 TableResource (p. 272) 对象。

  资源的表格。表是代表您的数据的元数据定义。您可以向委托人授予和撤消表权限。
- `TableWithColumns` – 一个 TableWithColumnsResource (p. 272) 对象。

  包含资源列的表。对此资源具有权限的委托人可以从数据目录中表的列中选择元数据以及中的基础数据Amazon S3.
- `DataLocation` – 一个 DataLocationResource (p. 273) 对象。

  一个Amazon S3授予或撤销权限的路径。
- `DataCellsFilter` – 一个 DataCellsFilter 资源 (p. 273) 对象。

数据单元格过滤器。

- `LFTag` – 一个 lftagkey 资源 (p. 287) 对象。

  附加到资源的 LF-tag 键值。
- `LFTagPolicy` – 一个 lftag 策略资源 (p. 287) 对象。

  定义资源的 LF-tag 策略的 LF-tag 条件列表。

DatabaseResource 结构

数据库对象的结构。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，它是调用者的账户 ID。
- `Name`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  数据库资源的名称。对数据目录是唯一的。

TableResource 结构

表对象的结构。表是代表您的数据的元数据定义。您可以向委托人授予和撤消表权限。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，它是调用者的账户 ID。
- `DatabaseName`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  表的数据库名称。对数据目录是唯一的。数据库是一组关联的表定义，这些定义组织到一个逻辑组内。您可以向委托人授予和撤消数据库权限。
- `Name` – 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 (p. 320) 匹配。

  表的名称。
- `TableWildcard`— 一个名为的空结构`TableWildcard`.

  表示数据库下每个表的通配符对象。

  至少需要 `TableResource$Name` 或 `TableResource$TableWildcard` 之一。

TableWithColumnsResource 结构

具有列对象的表的结构。此对象仅在授予 SELECT 权限时使用。

此对象必须至少从 `ColumnsNames`、`ColumnsIndexes` 或 `ColumnsWildcard` 之一获取值。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，它是调用者的账户 ID。
- `DatabaseName`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  具有列资源的表的数据库名称。对数据目录是唯一的。数据库是一组关联的表定义，这些定义组织到一个逻辑组内。您可以向委托人授予和撤消数据库权限。
- `Name`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  表资源的名称。表是代表您的数据的元数据定义。您可以向委托人授予和撤消表权限。
- `ColumnNames` – UTF-8 字符串数组。

  表的列名的列表。至少需要 `ColumnNames` 或 `ColumnWildcard` 之一。
- `ColumnWildcard` – 一个 ColumnWildcard (p. 275) 对象。

  由 `ColumnWildcard` 对象指定的通配符。至少需要 `ColumnNames` 或 `ColumnWildcard` 之一。

DataCellsFilterResource 结构

数据单元格筛选资源的结构。

字段

- `TableCatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  表所属目录的 ID。
- `DatabaseName` – 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 (p. 320) 匹配。

  中的数据库Amazon Glue数据目录。
- `TableName` – 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 (p. 320) 匹配。

  表的名称。
- `Name` – 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 (p. 320) 匹配。

  数据单元格筛选器的名称。

DataLocationResource 结构

授予或撤销权限的数据位置对象的结构。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  向其注册位置的数据目录的标识符。Amazon Lake Formation. 默认情况下，它是调用者的账户 ID。

- `ResourceArn`–：必需 UTF-8 字符串。

  唯一标识数据位置资源的 Amazon 资源名称 (ARN)。

DataLakePrincipal 结构

Amazon Lake Formation 委托人。受支持的委托人IAM用户或IAM角色。

字段

- `DataLakePrincipalIdentifier` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节。

  Amazon Lake Formation 委托人的标识符。

ResourcePercount 结构

向资源授予或撤销的权限。

字段

- `Resource` – 一个 资源 (p. 271) 对象。

  要授予或撤销权限的资源。
- `Permissions` – UTF-8 字符串数组。

  要对资源授予或撤销的权限。
- `PermissionsWithGrantOption` – UTF-8 字符串数组。

  指示是否提供授予权限的能力（作为授予的权限的子集）。

Resource PerperCurterCurterCurterC

表示在授予或撤消对资源的权限时出现错误的结构。

字段

- `ResourcePermissions` – 一个 ResourcePercount (p. 274) 对象。

  出现错误的资源权限列表。
- `Error` – 一个 ErrorDetail (p. 320) 对象。

  与尝试授予或撤消对资源的权限相关的错误。

PrincipalResource 权限结构

向资源授予或撤销的权限。

字段

- `Principal` – 一个 DataLakePrincipal (p. 274) 对象。

  要授予或撤销权限的数据湖委托人。
- `Resource` – 一个 资源 (p. 271) 对象。

要授予或撤销权限的资源。

- `Permissions` – UTF-8 字符串数组。

  要对资源授予或撤销的权限。

- `PermissionsWithGrantOption` – UTF-8 字符串数组。

  指示是否提供授予权限的能力（作为授予的权限的子集）。

- `AdditionalDetails` – 一个 DetailsMap (p. 275) 对象。

  此属性可用于返回任何其他详细信息`PrincipalResourcePermissions`. 目前只能作为Amazon RAM资源共享 ARN。

Environment 结构

包含要返回的其他详细信息的结构`AdditionalDetails`的属性`PrincipalResourcePermissions`.

如果通过共享目录资源Amazon Resource Access Manager(Amazon RAM)，那么就会存在一个相应的Amazon RAM资源共享 ARN。

字段

- `ResourceShare` – UTF-8 字符串数组。

  通过共享的目录资源的资源共享 ARNAmazon RAM.

PrincipalResource 权限错误结构

表示在向委托人授予或撤消权限时出现错误的结构。

字段

- `PrincipalResourcePermissions` – 一个 主要资源权限 (p. 274) 对象。

  要授予或撤销的委托人权限。

- `Error` – 一个 ErrorDetail (p. 320) 对象。

  尝试授予或撤消权限的错误消息。

ColumnWildcard 结构

一个通配符对象，由排除的列名或索引的可选列表组成。

字段

- `ExcludedColumnNames` – UTF-8 字符串数组。

  排除列名称。将排除具有此名称的任何列。

BatchPerperCurperCurperCourterCur

批处理操作向委托人授予对资源的权限。

字段

- `Id–`：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节。

  批处理权限请求条目的唯一标识符。
- `Principal` – 一个 DataLakePrincipal (p. 274) 对象。

  要获得许可的委托人。
- `Resource` – 一个 资源 (p. 271) 对象。

  授予委托人权限的资源。
- `Permissions` – UTF-8 字符串数组。

  要授予的权限。
- `PermissionsWithGrantOption` – UTF-8 字符串数组。

  指示是否授予了传递权限的选项。

BatchPatch许可失败输入结构

执行批量授予或批量撤销操作时失败的列表。

字段

- `RequestEntry` – 一个 批处理权限请求输入 (p. 275) 对象。

  批处理请求条目的标识符。
- `Error` – 一个 ErrorDetail (p. 320) 对象。

  适用于条目失败的错误消息。

PrincipalPermissions 结构

向委托人授予的权限。

字段

- `Principal` – 一个 DataLakePrincipal (p. 274) 对象。

  被授予权限的委托人。
- `Permissions` – UTF-8 字符串数组。

  向委托人授予的权限。

操作

- 授予权限操作（Python：grant_permissions）(p. 277)
- LevkeLevkePallget 操作（Python：revke_permissions）(p. 277)
- BatchGrantGrant 权限操作（Python：batch_grant_permissions）(p. 278)
- BatchRevkeLevkePatchRevkeEpermissions 操作（Python：batch_revke_permissions）(p. 279)
- 获取路径操作的有效权限 (Python: get_效e_permissions_for_path) (p. 279)

- ListListListListListList_permissions 操作（Python (p. 280)

授予权限操作（Python：grant_permissions）

授予委托人访问数据目录中的元数据和基础数据存储中组织的数据的权限，例如Amazon S3.

有关权限的更多信息，请参阅元数据和数据的安全性和访问控制.

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。
- `Principal`–：必需 一个 DataLakePrincipal (p. 274) 对象。

  要授予资源权限的委托人。受支持的委托人IAM用户或IAM角色，它们由主体类型和 ARN 定义。

  请注意，如果您使用特定 ARN 定义资源，然后再删除并重新创建具有相同 ARN 的资源，则该资源将保留已授予的权限。
- `Resource`–：必需 一个 资源 (p. 271) 对象。

  向其授予权限的资源。中的 资源Amazon Lake Formation是数据目录、数据库和表。
- `Permissions`–：必需 UTF-8 字符串数组。

  向委托人授予对资源的权限。Amazon Lake Formation定义了授予和撤销对数据目录中元数据的访问权限以及在基础数据存储中组织的数据的权限，例如Amazon S3.Lake Formation要求授权每个委托人执行特定任务Lake Formation资源的费用。
- `PermissionsWithGrantOption` – UTF-8 字符串数组。

  表示委托人可以传递给其他用户的授予权限的列表。这些权限可能只是中授予的权限的子集`Privileges`.

响应

- 无响应参数。

错误

- `ConcurrentModificationException`
- `EntityNotFoundException`
- `InvalidInputException`

LevkeLevkePallget 操作（Python：revke_permissions）

撤销委托人访问数据目录中的元数据和基础数据存储中组织的数据的权限，例如Amazon S3.

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。

- `Principal–`：必需 一个 DataLakePrincipal (p. 274) 对象。

  要撤销对资源的权限的委托人。

- `Resource–`：必需 一个 资源 (p. 271) 对象。

  将撤销权限的资源。

- `Permissions–`：必需 UTF-8 字符串数组。

  撤销了对资源的委托人的权限。有关权限的更多信息，请参阅元数据和数据的安全性和访问控制.

- `PermissionsWithGrantOption` – UTF-8 字符串数组。

  指示要撤消授权选项允许委托人将权限传递给其他委托人的权限列表。

响应

- 无响应参数。

错误

- `ConcurrentModificationException`
- `EntityNotFoundException`
- `InvalidInputException`

BatchGrantGrant 权限操作（Python：batch_grant_permissions）

向委托人授予权限的 Batch 操作。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。

- `Entries–`：必需 批处理权限请求输入 (p. 275) 对象数组。

  最多 20 个条目的列表，用于通过批处理操作向委托人授予资源权限。

响应

- `Failures` – 批处理权限失败输入 (p. 276) 对象的数组。

  向资源授予权限失败的列表。

错误

- `InvalidInputException`

- `OperationTimeoutException`

BatchRevkeLevkePatchRevkeEpermissions 操作（Python：batch_revke_permissions）

用于撤销委托人的权限的 Batch 操作。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。
- `Entries–`：必需 批处理权限请求输入 (p. 275) 对象数组。

  最多 20 个条目的列表，用于通过对委托人的批处理操作撤销的资源权限。

响应

- `Failures` – 批处理权限失败输入 (p. 276) 对象的数组。

  撤销资源权限的失败列表。

错误

- `InvalidInputException`
- `OperationTimeoutException`

获取路径操作的有效权限 (Python: get_效e_permissions_for_path)

返回Lake Formation对位于中路径中的指定表或数据库资源的权限Amazon S3.`GetEffectivePermissionsForPath`如果目录已加密，则不会返回数据库和表。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。
- `ResourceArn–`：必需 UTF-8 字符串。

  您要获得权限的资源的 Amazon 资源名称 (ARN)。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索此清单的第一个调用)。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  要返回的最大结果数量。

响应

- `Permissions` – 主要资源权限 (p. 274) 对象的数组。

  位于中的路径中的指定表或数据库资源的权限列表Amazon S3.
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索此清单的第一个调用)。

错误

- `InvalidInputException`
- `EntityNotFoundException`
- `OperationTimeoutException`
- `InternalServiceException`

ListListListListListList_permissions 操作 (Python

返回资源的委托人权限列表，该列表由调用者的权限筛选。例如，如果您被授予 ALTER 权限，则只能看到 ALTER 的委托人权限。

此操作仅返回那些明确授予的权限。

有关权限的更多信息，请参阅元数据和数据的安全性和访问控制.

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake FormationEnvironment。
- `Principal` – 一个 DataLakePrincipal (p. 274) 对象。

  指定一个委托人以筛选返回的权限。
- `ResourceType` – UTF-8 字符串 (有效值：`CATALOG` | `DATABASE` | `TABLE` | `DATA_LOCATION` | `LF_TAG` | `LF_TAG_POLICY` | `LF_TAG_POLICY_DATABASE` | `LF_TAG_POLICY_TABLE`) 。

  指定资源类型以筛选返回的权限。
- `Resource` – 一个 资源 (p. 271) 对象。

  一种资源，您将在其中获得委托人权限列表。

  此操作不支持获取具有列的表的权限。相反，请在表格上调用此操作，然后该操作返回表格和表 w 列。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索此清单的第一个调用)。
- `MaxResults` – 数字 (整数) ，不小于 1 或大于 1000。

  要返回的最大结果数量。
- `IncludeRelated`— UTF-8 字符串，长度不少于 1 个字节或超过 5 个字节，与匹配Single-line string pattern (p. 320).

  表示结果中应包括相关权限。

响应

- `PrincipalResourcePermissions` – 主要资源权限 (p. 274) 对象的数组。

  委托人及其对指定委托人和资源类型的资源权限的列表。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索此清单的第一个调用)。

错误

- `InvalidInputException`
- `OperationTimeoutException`
- `InternalServiceException`

数据湖设置 API

数据湖设置 API 描述了用于管理数据湖管理员的数据类型和操作。

数据类型

- DataLake设置结构 (p. 281)

DataLake设置结构

表示列表的结构Amazon Lake Formation指定为数据湖管理员的承担者以及默认创建数据库和默认创建表权限的委托人权限条目列表。

字段

- `DataLakeAdmins` – DataLakePrincipal (p. 274) 对象的数组。

  列表Amazon Lake Formation委托人。受支持委托人IAM用户或IAM角色。
- `CreateDatabaseDefaultPermissions` – PrincipalPermissions (p. 276) 对象的数组。

  指定是否由管理新创建的数据库的访问控制Lake Formation权限或完全由IAM权限。可以在创建数据库时覆盖此默认设置。

  空值表示访问控制Lake Formation权限。将 ALL 分配给 IAM_ALLOWED_PACYS 的值表示访问控制IAM权限。这被称为 “仅使用 IAM 访问控制” 的设置，是为了与Amazon Glue权限模型实施者IAM权限。

  唯一允许的值是空数组或包含向 IAM_ALLOWED_PACYS 授予 ALL 的单个 JSON 对象的数组。

  有关更多信息，请参阅 。更改数据湖的默认安全设置.
- `CreateTableDefaultPermissions` – PrincipalPermissions (p. 276) 对象的数组。

  指定是否由管理新创建的表的访问控制Lake Formation权限或完全由IAM权限。

  空值表示访问控制Lake Formation权限。将 ALL 分配给 IAM_ALLOWED_PACYS 的值表示访问控制IAM权限。这被称为 “仅使用 IAM 访问控制” 的设置，是为了与Amazon Glue权限模型实施者IAM权限。

  唯一允许的值是空数组或包含向 IAM_ALLOWED_PACYS 授予 ALL 的单个 JSON 对象的数组。

  有关更多信息，请参阅 。更改数据湖的默认安全设置.

- `TrustedResourceOwners` – UTF-8 字符串数组。

  资源拥有的账户 ID 列表，呼叫者的帐户可以用来共享其用户访问详细信息（用户 ARN）。用户 ARN 可以登录到资源所有者的CloudTrail日志。

  当您处于高信任度边界（例如同一团队或公司）时，您可能希望指定此属性。

操作

GetDataLakeSettings操作（Python：get_data_lake_settings）

检索的数据湖管理员列表Amazon Lake Formation-管理的数据湖。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，是账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Lake Formation环境。

响应

- `DataLakeSettings` – 一个 DataLake设置 (p. 281) 对象。

  表示列表的结构Lake Formation指定为数据湖管理员的承担者。

错误

- `InternalServiceException`
- `InvalidInputException`
- `EntityNotFoundException`

PutDataLakeSettings操作（Python：put_data_lake_settings）

设置对由管理的所有资源具有管理员权限的数据湖管理员列表Amazon Lake Formation. 有关管理员权限的更多信息，请参阅授权Lake FormationPermissions (权限).

此 API 将当前的数据湖管理员列表替换为通过的新列表。要添加管理员，请获取当前列表并将新管理员添加到该列表中，然后在此 API 中传递该列表。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

数据目录的标识符。默认情况下，是账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Lake Formation环境。

- `DataLakeSettings`–是必需的：一个 DataLake设置 (p. 281) 对象。

  表示列表的结构Lake Formation指定为数据湖管理员的承担者。

响应

- 无响应参数。

错误

- `InternalServiceException`
- `InvalidInputException`

凭据自动售货机 API

凭据自动售货机 API 描述了与使用Amazon Lake Formation服务来发送凭据以及注册和管理数据湖资源。

数据类型

- 筛选条件结构 (p. 283)
- ColumnNames (p. 283)
- Resource Info 结构 (p. 284)

筛选条件结构

此结构描述了基于筛选条件筛选表中的列的过滤。

字段

- `Field` – UTF-8 字符串（有效值：`RESOURCE_ARN` | `ROLE_ARN` | `LAST_MODIFIED`）。

  要在筛选条件下过滤的字段。
- `ComparisonOperator`— UTF-8 字符串（有效值：`EQ|NE|LE|LT|GE|GT|CONTAINS|NOT_CONTAINS|BEGINS_WITH|IN|BETWEEN`）。

  在筛选条件中使用的比较运算符。
- `StringValueList` – UTF-8 字符串数组。

  一个字符串，其中包含用于评估筛选条件的值。

ColumnNames

表中的列名的列表。

UTF-8 字符串数组。

表中的列名的列表。

Resource Info 结构

包含有关Amazon Lake Formation资源。

字段

- `ResourceArn` – UTF-8 字符串。

  资源的 Amazon 资源名称 (ARN)。
- `RoleArn` – UTF-8 字符串，与 Custom string pattern #5 (p. 321) 匹配。

  这些区域有：IAM注册资源的角色。
- `LastModified` – 时间戳。

  上次修改资源的日期和时间。

操作

- RegistRegerResource 操作（Python：register _resource）(p. 284)
- 取消注册资源操作（Python：deRegister_resource）(p. 285)
- ListResource 操作（Python：list_resource）(p. 285)

RegistRegerResource 操作（Python：register _resource）

将资源注册为由数据目录管理。

要添加或更新数据，Amazon Lake Formation需要对选定的读/写访问权限Amazon S3路径。选择您知道有权执行此操作的角色，或者选择 AWSServiceRoleForLakeFormationDataAccess 服务相关角色。当你注册第一个Amazon S3路径、服务相关角色和新的内联策略将代表您创建。Lake Formation将第一个路径添加到内联策略并将其附加到服务相关角色。当你注册后续路径时，Lake Formation将路径添加到现有策略。

以下请求注册了一个新的位置并给出Lake Formation允许使用服务相关角色访问该位置。

```
ResourceArn = arn:aws:s3:::my-bucket UseServiceLinkedRole = true
```

如果`UseServiceLinkedRole`未设为 true，您必须提供或设置`RoleArn`：

```
arn:aws:iam::12345:role/my-data-access-role
```

请求

- `ResourceArn`–：必需 UTF-8 字符串。

  您要注册的资源的 Amazon 资源名称 (ARN)。
- `UseServiceLinkedRole` – 布尔值。

  指定Amazon Identity and Access Management(IAM) 通过向数据目录注册此角色来实现与服务相关的角色。服务相关角色是一种与 Lake Formation 直接关联的独特类型的 IAM 角色。

  有关更多信息，请参阅为 Lake Formation 使用服务相关角色。
- `RoleArn` – UTF-8 字符串，与 Custom string pattern #5 (p. 321) 匹配。

注册资源的角色的标识符。

响应

- 无响应参数。

错误

- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AlreadyExistsException`
- `EntityNotFoundException`
- `ResourceNumberLimitExceededException`
- `AccessDeniedException`

取消注册资源操作（Python：deRegister_resource）

取消注册由数据目录管理的资源。

当你取消注册路径时，Lake Formation从附加到服务相关角色的内联策略中删除路径。

请求

- `ResourceArn–`：必需 UTF-8 字符串。

  您要取消注册的资源的 Amazon 资源名称 (ARN)。

响应

- 无响应参数。

错误

- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `EntityNotFoundException`

ListResource 操作（Python：list_resource）

列出注册由数据目录管理的资源。

请求

- `FilterConditionList`— 数组筛选条件 (p. 283)对象，不少于 1 个或不超过 20 个结构。

  资源的任何适用的行级和/或列级筛选条件。

- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  资源结果的最大数量。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索这些资源的第一个调用)。

响应

- `ResourceInfoList` – 资源信息 (p. 284) 对象的数组。

  数据湖资源的摘要。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索这些资源的第一个调用)。

错误

- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`

标记 API

标记 API 描述了与定义属性或键值对标签的权限模型的授权策略相关的数据类型和 API。

数据类型

- Tag 结构 (p. 286)
- LftagKey 资源结构 (p. 287)
- LFTAG 策略资源结构 (p. 287)
- TagedStable 结构 (p. 287)
- Tags 数据库结构 (p. 288)
- LfTag 结构 (p. 288)
- LftagPair 结构 (p. 288)
- LftagFtagCrror 结构 (p. 289)
- ColumlFTag 结构 (p. 289)

Tag 结构

LF-tag 键值对的结构。

字段

- `key` – UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节。

  LF 标签的钥匙。
- `value` – UTF-8 字符串，不超过 256 个字节。

LF-Tag 的值。

LftagKey 资源结构

包含 LF-tag 键和资源值的结构。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey`–：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  lf-tag 的键名。
- `TagValues`–：必需 — UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  属性可以获取的值列表。

LFTAG 策略资源结构

包含适用于资源的 LF-tag 策略的 LF-tag 条件列表的结构。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `ResourceType`–：必需 ：UTF-8 字符串（有效值：`DATABASE`|`TABLE`)。

  LF-tag 策略适用的资源类型。
- `Expression`–：必需 数组lfTag (p. 288)对象，不少于 1 个或不超过 5 个结构。

  适用于资源的 LF-tag 策略的 LF-tag 条件列表。

TagedStable 结构

描述带有 LF-Tags 的表资源的结构。

字段

- `Table` – 一个 TableResource (p. 272) 对象。

  附加了 LF 标签的表。
- `LFTagOnDatabase`— 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  附加到表所在数据库的 LF-Tags 列表。
- `LFTagsOnTable`— 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

附加到表格的 LF 标签列表。

- `LFTagsOnColumns` – ColumlFTag (p. 289) 对象的数组。

  附加到表中列的 LF 标签列表。

Tags 数据库结构

用 lf-Tags 描述数据库资源的结构。

字段

- `Database` – 一个 DatabaseResource (p. 272) 对象。

  附加了 LF 标签的数据库。
- `LFTags`— 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  附加到数据库的 LF 标签列表。

LfTag 结构

允许管理员在特定条件下授予用户权限的结构。例如，授予角色访问具有 LF-Tag 'Prod 的表中没有 LF-Tag “PII” 的所有列的权限。

字段

- `TagKey`–：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与匹配。Custom string pattern #9 (p. 321).

  lf-tag 的键名。
- `TagValues`–：必需 — UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  属性可以获取的值列表。

LftagPair 结构

包含 LF-tag 键值对的结构。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey`–：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与匹配。Custom string pattern #9 (p. 321).

  lf-tag 的键名。
- `TagValues`–：必需 — UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  属性可以获取的值列表。

LftagFtagCrror 结构

包含与`TagResource`要么`UnTagResource`operation.

字段

- `LFTag` – 一个 lftagPair (p. 288) 对象。

  lf-tag 的键名。
- `Error` – 一个 ErrorDetail (p. 320) 对象。

  LF-Tag 的附件或分离时发生的错误。

ColumlFTag 结构

包含列资源名称及其附加的 LF 标签的结构。

字段

- `Name` – 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 (p. 320) 匹配。

  列资源的名称。
- `LFTags`— 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  附加到列资源的 LF 标签。

操作

- Addlftag 到资源操作（Python：add_lf_tags_to_资源）(p. 289)
- 从资源操作中删除 FTags（Python：remove_lf_tags_resource）(p. 290)
- GetResourceFTags 操作（Python：get_resource_lf_tags）(p. 291)
- ListLfTags 操作（Python：list_lf_tags）(p. 291)
- CreateFTag 操作（Python：create_lf_tag）(p. 292)
- getLfTag 操作（Python：get_lf_tag）(p. 293)
- UpdateLfTag 操作（Python：update_lf_tag）(p. 294)
- DeleteFTag 操作（Python：delete_lf_tag）(p. 294)
- SearchTablebylfTags 操作（Python：search_table _by_lf_tags）(p. 295)
- SearchDatabesbylfTags 操作（Python：search_databases_by_lf_tags）(p. 296)

Addlftag 到资源操作（Python：add_lf_tags_to_资源）

将一个或多个 LF 标签附加到现有资源。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。

- `Resource`– ：必需 一个 资源 (p. 271) 对象。

  要附加 LF-Tag 的数据库、表或列资源。
- `LFTags`– ：必需 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  要附加到资源的 LF 标签。

响应

- `Failures` – lftag 错误 (p. 289) 对象的数组。

  标记资源的失败列表。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AccessDeniedException`
- `ConcurrentModificationException`

从资源操作中删除 FTags（Python：remove_lf_tags_resource）

从资源中删除 LF-Tag。只允许使用数据库、表或 tableWhwitColumn 资源。要标记列，请使用中的列包含列表`tableWithColumns`以指定列输入。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `Resource`– ：必需 一个 资源 (p. 271) 对象。

  要删除 LF-Tag 的数据库、表或列资源。
- `LFTags`– ：必需 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  要从资源中删除的 LF 标签。

响应

- `Failures` – lftag 错误 (p. 289) 对象的数组。

  取消标记资源的失败列表。

错误

- `EntityNotFoundException`

- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `GlueEncryptionException`
- `AccessDeniedException`
- `ConcurrentModificationException`

GetResourceFTags 操作（Python：get_resource_lf_tags）

返回应用于资源的 LF 标签。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `Resource`–：必需 一个 资源 (p. 271) 对象。

  要返回 LF-Tags 的数据库、表或列资源。
- `ShowAssignedLFTags` – 布尔值。

  指示是否显示分配的 LF 标签。

响应

- `LFTagOnDatabase`— 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  应用于数据库资源的 LF 标签列表。
- `LFTagsOnTable`— 数组lftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  应用于表资源的 LF 标签列表。
- `LFTagsOnColumns` – ColumlFTag (p. 289) 对象的数组。

  应用于列资源的 LF 标签列表。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `GlueEncryptionException`
- `AccessDeniedException`

ListLfTags 操作（Python：list_lf_tags）

列出请求者有权查看的 LF 标签。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `ResourceShareType` – UTF-8 字符串（有效值：`FOREIGN` | `ALL`）。

  如果资源共享类型为`ALL`，返回请求者有权查看的账户内 LF 标签和共享 LF 标签。如果资源共享类型为`FOREIGN`，返回请求者可以查看的所有共享 LF 标签。如果没有通过资源共享类型，请求者有权查看的给定目录 ID 中列出 LF-Tags。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  要返回的最大结果数量。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索此列表的第一个调用)。

响应

- `LFTags`— 数组IftagPair (p. 288)对象，不少于 1 个或不超过 50 个结构。

  请求有权查看的 LF 标签列表。
- `NextToken` – UTF-8 字符串。

  延续令牌 (如果当前列表片段不是最后一个，则呈现)。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AccessDeniedException`

CreateFTag 操作（Python：create_lf_tag）

使用指定的名称和值创建 LF-tag。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey`–：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与匹配。Custom string pattern #9 (p. 321).

  lf-tag 的键名。
- `TagValues`–：必需 — UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  属性可以获取的值列表。

响应

- 无响应参数。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `ResourceNumberLimitExceededException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AccessDeniedException`

getLfTag 操作（Python：get_lf_tag）

返回 lf-tag 定义。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey–`：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与匹配。Custom string pattern #9 (p. 321).

  lf-tag 的键名。

响应

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey` – UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与 Custom string pattern #9 (p. 321) 匹配。

  lf-tag 的键名。
- `TagValues`— UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  属性可以获取的值列表。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AccessDeniedException`

UpdateLfTag 操作（Python：update_lf_tag）

更新指定 LF-Tag 键的可能值列表。如果 LF-Tag 不存在，该操作将抛出 EntityNotFoundException。删除键值中的值将从可能值列表中删除。如果删除键值中的任何值附加到资源，则 API 会出错并出现 400 例外-“不允许更新”。在删除 lf-tag 键的值之前取消标记属性。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey`–：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与匹配。Custom string pattern #9 (p. 321).

  要为其添加或删除值的 LF-tag 的键名称。
- `TagValuesToDelete`— UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  要从 LF-Tag 中删除的 LF-Tag 值的列表。
- `TagValuesToAdd`— UTF-8 字符串数组，不少于 1 个或不超过 50 个字符串。

  要从 LF-tag 中添加的 lf-tag 值的列表。

响应

- 无响应参数。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `ConcurrentModificationException`
- `AccessDeniedException`

DeleteFTag 操作（Python：delete_lf_tag）

删除指定的 lf-tag 密钥名称。如果属性键不存在或 LF-Tag 不存在，则操作将不会执行任何操作。如果属性键存在，则操作将检查是否有资源标记为此属性键，如果是，API 会抛出 400 例外，并显示 “不允许删除” 消息，因为 LF-Tag 键仍附有资源。您可以考虑使用此 LF-tag 密钥取消标记资源。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `TagKey`–：必需 — UTF-8 字符串，长度不少于 1 个字节或超过 128 个字节，与匹配。Custom string pattern #9 (p. 321).

要删除的 LF-Tag 的键名。

响应

- 无响应参数。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AccessDeniedException`

SearchTablebylfTags 操作（Python：search_table _by_lf_tags）

此操作允许搜索`TABLE`资源`LFTag`。这将由想要授予用户对某些 LF-Tags 的权限的管理员使用。在获得补助之前，管理员可以使用`SearchTablesByLFTags`找到给定的所有资源`LFTags` 对于验证返回的资源是否可以共享是否有效。

请求

- `NextToken` – UTF-8 字符串。

  延续标记 (如果这不是检索此列表的第一个调用)。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  要返回的最大结果数量。
- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- `Expression`–：必需 数组lfTag (p. 288)对象，不少于 1 个或不超过 5 个结构。

  条件的列表 (`LFTag`结构）以在表资源中搜索。

响应

- `NextToken` – UTF-8 字符串。

  延续令牌 (如果当前列表片段不是最后一个，则呈现)。
- `TableList` – 标记表 (p. 287) 对象的数组。

  满足 LF-Tag 条件的表的列表。

错误

- `EntityNotFoundException`
- `InternalServiceException`

- InvalidInputException
- OperationTimeoutException
- GlueEncryptionException
- AccessDeniedException

SearchDatabesbylfTags 操作 (Python : search_databases_by_lf_tags)

此操作允许搜索DATABASE资源TagCondition. 想要授予用户对某些特定权限的管理员使用此操作TagConditions. 在获得补助之前，管理员可以使用SearchDatabasesByTags找到给定的所有资源TagConditions用于验证返回的资源是否可以共享。

请求

- NextToken – UTF-8 字符串。

  延续标记 (如果这不是检索此列表的第一个调用)。
- MaxResults – 数字 (整数) ，不小于 1 或大于 1000。

  要返回的最大结果数量。
- CatalogId – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  数据目录的标识符。默认情况下，账户 ID。数据目录是持久性元数据存储。它包含数据库定义、表定义以及其他用于管理您的控制信息。Amazon Lake Formation一个环境。
- Expression– : 必需 数组lfTag (p. 288)对象，不少于 1 个或不超过 5 个结构。

  条件的列表 (LFTag结构) 以在数据库资源中搜索。

响应

- NextToken – UTF-8 字符串。

  延续令牌 (如果当前列表片段不是最后一个，则呈现)。
- DatabaseList – 标记数据库 (p. 288) 对象的数组。

  满足 LF-Tag 条件的数据库列表。

错误

- EntityNotFoundException
- InternalServiceException
- InvalidInputException
- OperationTimeoutException
- GlueEncryptionException
- AccessDeniedException

事务 API

事务 API 描述了以事务方式更新中表格内容的操作Amazon Lake Formation.

数据类型

事务描述结构

一个包含事务相关信息的结构。

字段

- `TransactionId` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  事务的 ID。
- `TransactionStatus` – UTF-8 字符串（有效值：`ACTIVE` | `COMMITTED` | `ABORTED` | `COMMIT_IN_PROGRESS`）。

  状态为 “活动”、“已提交” 或 “已中止”。
- `TransactionStartTime` – 时间戳。

  事务开始的时间。
- `TransactionEndTime` – 时间戳。

  提交或中止事务的时间（如果当前处于活动状态）。

Virtual Object 结构

一个定义Amazon S3如果交易取消，则要删除的对象，前提是`VirtualPut`是在写对象之前调用的。

字段

- `Uri`–必填项：统一资源标识符 (uri)，长度不少于 1 个字节或超过 1024 个字节，与匹配。URI address multi-line string pattern (p. 320).

  的路径Amazon S3对象。必须以 s3://开头
- `ETag` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  的 eTagAmazon S3对象。

操作

StartTranstransaction 操作（Python：start_transaction）

启动新交易并返回其交易 ID。事务 ID 是不透明的对象，您可以用它来识别交易。

请求

- `TransactionType` – UTF-8 字符串（有效值：`READ_AND_WRITE` | `READ_ONLY`）。

  指示此事务应为只读还是读写。使用只读事务 ID 进行的写入将被拒绝。只读事务不需要提交。

响应

- `TransactionId` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  事务的不透明标识符。

错误

- `InternalServiceException`
- `OperationTimeoutException`

事务操作（Python：commit _transaction）

尝试提交指定的事务。如果事务以前已中止，则返回异常。如果对同一事务多次调用此 API 操作，则是幂等的。

请求

- `TransactionId`–必填项：— UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Custom string pattern #11 (p. 321).

  要提交的交易。

响应

- `TransactionStatus` – UTF-8 字符串（有效值：`ACTIVE` | `COMMITTED` | `ABORTED` | `COMMIT_IN_PROGRESS`）。

  事务的状态。

错误

- `InvalidInputException`
- `EntityNotFoundException`
- `InternalServiceException`
- `OperationTimeoutException`
- `TransactionCanceledException`
- `ConcurrentModificationException`

Canceltransaction 操作（ Python ：ancel_transaction ）

尝试取消指定的事务。如果事务以前已提交，则返回异常。

请求

- `TransactionId`–必填项：— UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Custom string pattern #11 (p. 321).

  要取消的事务。

响应

- 无响应参数。

错误

- `InvalidInputException`
- `EntityNotFoundException`
- `InternalServiceException`
- `OperationTimeoutException`
- `TransactionCommittedException`
- `TransactionCommitInProgressException`
- `ConcurrentModificationException`

ExendDtransaction 操作（ Python ：extend_transaction ）

向服务表明指定的事务仍处于活动状态，不应将其视为空闲和中止。

除非明确延长，否则长时间处于空闲状态的写入事务将自动中止。

请求

- `TransactionId` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  要延长的交易。

响应

- 无响应参数。

错误

- `InvalidInputException`
- `EntityNotFoundException`
- `InternalServiceException`
- `OperationTimeoutException`
- `TransactionCommittedException`

- `TransactionCanceledException`
- `TransactionCommitInProgressException`

事务操作（Python：describpe_transaction）

返回单个事务的详细信息。

请求

- `TransactionId`–必填项：— UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Custom string pattern #11 (p. 321).

  要返回状态的交易。

响应

- `TransactionDescription` – 一个 交易说明 (p. 297) 对象。

  返回`TransactionDescription`包含事务相关信息的对象。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `InternalServiceException`
- `OperationTimeoutException`

Listtransaction 操作（Python：list_transaction）

返回有关交易及其状态的元数据。为了防止响应无限期增长，只返回未提交的事务和可用于时间旅行查询的事务。

此操作可以帮助您识别未提交的事务或获取有关交易的信息。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  要列出交易记录的目录。默认值为调用者的账户 ID。
- `StatusFilter` – UTF-8 字符串（有效值：`ALL` | `COMPLETED` | `ACTIVE` | `COMMITTED` | `ABORTED`）。

  指示要返回的交易状态的过滤器。选项全部 | 已完成 | 已提交 | 已中止 | 活动。默认为 `ALL`。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  要在单个调用中返回的最大交易数。
- `NextToken`— UTF-8 字符串，不超过 4096 个字节。

  延续令牌（如果这不是检索事务的第一个调用）。

响应

- `Transactions` – 交易说明 (p. 297) 对象的数组。

交易清单。每笔交易的记录都是TransactionDescription对象。

- NextToken— UTF-8 字符串，不超过 4096 个字节。

  一个延续令牌，指示是否有其他数据可用。

错误

- InvalidInputException
- InternalServiceException
- OperationTimeoutException

DeleteObjects 操作（Python：delete_objects_on_ance_transaction）

对于特定的受管理表，请提供Amazon S3将在当前事务期间写入的对象，如果事务被取消，则可以自动删除该对象。没有这个电话，不Amazon S3当事务取消时，对象将自动删除。

这些区域有：Amazon GlueETL 库函数write_dynamic_frame.from_catalog()包括自动呼叫的选项DeleteObjectsOnCancel在写之前。有关更多信息，请参阅 。回滚Amazon S3写入.

请求

- CatalogId – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  这些区域有：Amazon Glue包含受管理表的数据目录。默认为当前账户 ID。
- DatabaseName–必填项：— UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  包含受管理表的数据库。
- TableName–必填项：— UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  受管理表的名称。
- TransactionId–必填项：— UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Custom string pattern #11 (p. 321).

  写入进行的事务的 ID。
- Objects–必填项：数组虚拟对象 (p. 297)对象，不少于 1 个或不超过 100 个结构。

  VirtualObject 结构的列表，指示Amazon S3如果事务取消，则要删除的对象。

响应

- 无响应参数。

错误

- InternalServiceException
- InvalidInputException
- OperationTimeoutException

- `EntityNotFoundException`
- `TransactionCommittedException`
- `TransactionCanceledException`
- `ResourceNotReadyException`
- `ConcurrentModificationException`

异常

Progrestransaction 异常结构

包含与正在进行的事务提交相关的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

事务中止异常结构

包含有关指定交易已中止且无法用于的错误的详细信息`UpdateTableObjects`.

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

事务委员会异常结构

包含有关指定事务已经提交且无法用于的错误的详细信息`UpdateTableObjects`.

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

TransactionCanceledException 结构

包含与已取消的交易相关的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

事务争议异常结构

包含有关可重试错误的详细信息，该错误表明由于争用或冲突而导致提交未成功。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

ResourceNotReadyException 结构

包含与尚未准备好进行交易的资源相关的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

对象 API

Object API 在中描述了受管理的表对象Amazon Lake Formation.

数据类型

Table 对象结构

指定受管理表的详细信息。

字段

- `Uri` - 统一资源标识符 (uri)，不少于 1 个字节或超过 1024 个字节，与 URI address multi-line string pattern (p. 320) 匹配。

  这些区域有：Amazon S3对象的位置。
- `ETag` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

这些区域有：Amazon S3Tag 对象的 ETag。返回方`GetTableObjects`用于验证并用于标识基础数据的更改。

- `Size` – 数字（长型）。

  大小的大小Amazon S3对象以字节为单位。

Transform 对象结构

包含分区值和表对象列表的结构。

字段

- `PartitionValues`— UTF-8 字符串数组，不少于 1 个或不超过 100 个字符串。

  分区值的列表。
- `Objects` – 表对象 (p. 303) 对象的数组。

  表对象的列表

addoBjectInput 结构

要添加到受管理表的新对象。

字段

- `Uri`–：必需 统一资源标识符 (uri)，长度不少于 1 个字节或超过 1024 个字节，与匹配。URI address multi-line string pattern (p. 320).

  这些区域有：Amazon S3对象的位置。
- `ETag`–：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Custom string pattern #11 (p. 321).

  这些区域有：Amazon S3Tag 对象的 ETag。返回方`GetTableObjects`用于验证并用于标识基础数据的更改。
- `Size`–：必需 数字 (long)。

  大小的大小Amazon S3对象以字节为单位。
- `PartitionValues`— UTF-8 字符串数组，不少于 1 个或不超过 100 个字符串。

  对象的分区值的列表。必须为与表关联的每个分区键指定一个值。

  支持的数据类型包括整数、长型、日期 (yyyy-MM-dd)、时间戳 (yyy-mm-dd hh: mm: SSXXX 或 yyy-mm-dd hh: mm: SS")、字符串和小数点数。

DeceObjectInput 结构

要从受管理的表中删除的对象。

字段

- `Uri`–：必需 统一资源标识符 (uri)，长度不少于 1 个字节或超过 1024 个字节，与匹配。URI address multi-line string pattern (p. 320).

这些区域有：Amazon S3要删除的对象的位置。

- `ETag` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  这些区域有：Amazon S3Tag 对象的 ETag。返回方`GetTableObjects`用于验证并用于标识基础数据的更改。
- `PartitionValues`— UTF-8 字符串数组，不少于 1 个或不超过 100 个字符串。

  对象的分区值的列表。必须为与受管理表关联的每个分区键指定一个值。

WriteOpform 结构

定义要在受管理表中添加或删除的对象。

字段

- `AddObject` – 一个 addobject输入 (p. 304) 对象。

  要添加到受管理表的新对象。
- `DeleteObject` – 一个 删除对象输入 (p. 304) 对象。

  要从受管理的表中删除的对象。

操作

- GetTable对象操作（Python：get_table_object）(p. 305)
- UpdateTable 对象操作（Python：update_table_object）(p. 306)

GetTable对象操作（Python：get_table_object）

返回一组Amazon S3对象组成指定的受管理表。可以为时间旅行查询指定交易 ID 或时间戳。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  包含受管理表的目录。默认为来电者的帐户。
- `DatabaseName`–：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  包含受管理表的数据库。
- `TableName`–：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  要检索对象的受管理表。
- `TransactionId` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  在该 ID 处读取受管理的表内容的事务 ID。如果此事务已中止，将返回错误。如果未设置，则默认为最近提交的事务。无法与 `QueryAsOfTime` 一起指定。

- `QueryAsOfTime` – 时间戳。

  截至读取受管理的表内容的时间。如果未设置，将使用最近的事务提交时间。无法与 `TransactionId` 一起指定。
- `PartitionPredicate` – 谓词字符串，不超过 2048 个字节，与 URI address multi-line string pattern (p. 320) 匹配。

  根据受管理表中定义的分区键筛选返回的对象的谓词。
  - 支持的比较运算符有：=、>、<、>=、<=
  - 支持的逻辑运算符有：AND
  - 支持的数据类型包括整数、长型、日期 (yyyy-MM-dd)、时间戳 (yyy-mm-dd hh: mm: SSXXX 或 yyy-mm-dd hh: mm: SS")、字符串和小数点数。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  指定在页面中返回多少值。
- `NextToken`— UTF-8 字符串，长度不超过 4096 个字节。

  延续标记（如果这不是检索这些对象的第一次调用）。

响应

- `Objects` – 分区对象 (p. 304) 对象的数组。

  按分区键组织的对象的列表。
- `NextToken`— UTF-8 字符串，长度不超过 4096 个字节。

  一个延续令牌，指示是否有其他数据可用。

错误

- `EntityNotFoundException`
- `InternalServiceException`
- `InvalidInputException`
- `OperationTimeoutException`
- `TransactionCommittedException`
- `TransactionCanceledException`
- `ResourceNotReadyException`

UpdateTable 对象操作（Python：update_table_object）

更新清单Amazon S3对象组成指定的受管理表。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  包含要更新的受管理表的目录。默认为来电者的帐户 ID。
- `DatabaseName`–：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

包含要更新的受管理表的数据库。

- `TableName`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  要更新的受管理表。

- `TransactionId`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Custom string pattern #11 (p. 321).

  要执行写入的交易。

- `WriteOperations`– ：必需 数组写操作 (p. 305)对象，不少于 1 个或不超过 100 个结构。

  列表`WriteOperation`定义要添加到受管理表的清单中或从清单中删除的对象的对象。

响应

- 无响应参数。

错误

- `InternalServiceException`
- `InvalidInputException`
- `OperationTimeoutException`
- `EntityNotFoundException`
- `TransactionCommittedException`
- `TransactionCanceledException`
- `TransactionCommitInProgressException`
- `ResourceNotReadyException`
- `ConcurrentModificationException`

数据筛选数据 API

数据筛选器 API 介绍了如何在中管理数据单元格筛选器Amazon Lake Formation.

数据类型

- DataCellsFilter 结构 (p. 307)
- TrowFilter 结构 (p. 308)

DataCellsFilter 结构

描述某些行上的某些列的结构。

字段

- `TableCatalogId`– ：必需 Matalog id 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  表所属目录的 ID。

- `DatabaseName`– : 必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  中的数据库Amazon Glue数据目录。
- `TableName`– : 必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  数据库中的表。
- `Name`– : 必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配Single-line string pattern (p. 320).

  用户给数据筛选器单元格指定的名称。
- `RowFilter` – 一个 ROW 过滤器 (p. 308) 对象。

  PartiQL 谓词。
- `ColumnNames` – UTF-8 字符串数组。

  列名称的列表。
- `ColumnWildcard` – 一个 ColumnWildcard (p. 275) 对象。

  包含排除项的通配符。

  您必须指定`ColumnNames`列表或`ColumnWildCard`.

TrowFilter 结构

PartiQL 谓词。

字段

- `FilterExpression` – 谓词字符串，不超过 2048 个字节，与 URI address multi-line string pattern (p. 320) 匹配。

  筛选条件表达式。
- `AllRowsWildcard`— 一个名为的空结构`AllRowsWildcard`.

  所有行的通配符。

操作

- CellateCellsFilter 操作（Python：create_data_Cells_filter）(p. 308)
- DeleteCellsCellsFilter 操作（Python：delete_data_Cells_filter）(p. 309)
- 列表数据 CellsCellsFilter 操作（Python：list_data_cells_filter）(p. 310)

CellateCellsFilter 操作（Python：create_data_Cells_filter）

创建数据单元格筛选器以允许用户授予对某些行上某些列的访问权限。

请求

- `TableData`– : 必需 一个 DataCellsFilter (p. 307) 对象。

一个DataCellsFilter包含有关数据单元格筛选器信息的结构。

响应

- 无响应参数。

错误

- AlreadyExistsException
- InvalidInputException
- EntityNotFoundException
- ResourceNumberLimitExceededException
- InternalServiceException
- OperationTimeoutException
- AccessDeniedException

DeleteCellsCellsFilter 操作（Python：delete_data_Cells_filter）

删除数据单元格筛选器。

请求

- TableData – 一个 DataCellsFilter (p. 307) 对象。

  一个DataCellsFilter包含有关数据单元格筛选器信息的结构。
- TableCatalogId – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  表所属目录的 ID。
- DatabaseName – 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 (p. 320) 匹配。

  中的数据库Amazon Glue数据目录。
- TableName – 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 (p. 320) 匹配。

  数据库中的表。
- Name – 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 (p. 320) 匹配。

  用户给数据筛选器单元格指定的名称。

响应

- 无响应参数。

错误

- InvalidInputException

- `EntityNotFoundException`
- `InternalServiceException`
- `OperationTimeoutException`
- `AccessDeniedException`

列表数据 CellsCellsFilter 操作（Python：list_data_cells_filter）

列出表中的所有数据单元格筛选器。

请求

- `Table` – 一个 TableResource (p. 272) 对象。

  中的一张表Amazon Glue数据目录。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这是延续调用)。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  响应的最大大小。

响应

- `DataCellsFilters` – DataCellsFilter (p. 307) 对象的数组。

  列表`DataCellFilter`结构。
- `NextToken` – UTF-8 字符串。

  延续令牌（如果尚未返回所有请求的数据单元格筛选条件）。

错误

- `InvalidInputException`
- `OperationTimeoutException`
- `InternalServiceException`
- `AccessDeniedException`

查询 API

查询 API 允许您共享中管理的数据湖的事务一致性数据Amazon S3、Amazon Redshift和其他Amazon服务。

数据类型

- WorkUnit 范围结构 (p. 311)
- GetWorkUnits 响应结构 (p. 311)
- GetQueryState 响应结构 (p. 311)
- GetWorkUnit 结果 (p. 312)

WorkUnit 范围结构

定义用于查询执行服务的工作单元 ID 的有效范围。

字段

- `WorkUnitIdMax`– : 必需 数字 (long)。

  定义范围内的最大工作单元 ID。最大值包括在内。
- `WorkUnitIdMin`– : 必需 数字 (long)。

  定义范围内的最小工作单元 ID。
- `WorkUnitToken`– : 必需 UTF-8 字符串。

  用于查询执行服务的工作令牌。

GetWorkUnits 响应结构

输出的结构。

字段

- `NextToken` – UTF-8 字符串。

  对返回的标记列表进行分页的延续令牌 (如果列表的当前片段不是最后一个，则返回)。
- `QueryId`– : 必需 UTF-8 字符串。

  计划查询操作的 ID。
- `WorkUnitRanges`– : 必需 WorkUnit 范围 (p. 311) 对象数组。

  一个`WorkUnitRangeList`对象，该对象指定用于查询执行服务的工作单元 ID 的有效范围。

GetQueryState 响应结构

输出的结构。

字段

- `Error` – UTF-8 字符串。

  在操作失败时显示错误消息。
- `State`– : 必需 UTF-8 字符串 (有效值 : `PENDING=""|WORKUNITS_AVAILABLE=""|ERROR=""|FINISHED=""|EXPIRED=""`)。

  之前提交的查询的状态。可能的状态包括 :
  - 待定 : 查询处于待处理状态。
  - WORKUNITS_SALABLE : 一些工作单元已准备好进行检索和执行。
  - 已完成 : 查询计划已成功完成，所有工作单元都已准备好进行检索和执行。

- 错误：查询发生错误，例如无效的查询 ID 或后端错误。

GetWorkUnit 结果

输出的结构。

字段

- `ResultStream` – Blob。

  从返回的行`GetWorkUnitResults`作为 Apache Arrow v1.0 消息的流进行操作。

查询计划上下文结构

包含有关查询计划信息的结构。

字段

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  所涉分区所在的数据目录的 ID。如果没有提供，则默认情况下使用 Amazon 账户 ID。
- `DatabaseName`– ：必需 匹配 UTF-8 字符串，至少 1 个字节，与匹配Single-line string pattern (p. 320).

  包含该表的数据库。
- `QueryAsOfTime` – 时间戳。

  截至读取表内容的时间。如果未设置，将使用最近的事务提交时间。无法与 `TransactionId` 一起指定。
- `QueryParameters` – 键值对的映射数组。

  每个键是一个 UTF-8 字符串。

  每个值是一个 UTF-8 字符串。

  由键值对组成的映射。
- `TransactionId` – UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与 Custom string pattern #11 (p. 321) 匹配。

  在该 ID 处读取表内容的事务 ID。如果未提交此事务，则读取将被视为该事务的一部分，并将看到其写入内容。如果此事务已中止，将返回错误。如果未设置，则默认为最近提交的事务。无法与 `QueryAsOfTime` 一起指定。

执行/统计结构

与处理查询语句相关的统计信息。

字段

- `AverageExecutionTimeMillis` – 数字（长型）。

  执行请求的平均时间。
- `DataScannedBytes` – 数字（长型）。

  扫描的数据量（以字节为单位）。

- `WorkUnitsExecutedCount` – 数字（长型）。

  已执行的工作单位的数量。

规划/统计结构

与处理查询语句相关的统计信息。

字段

- `EstimatedDataToScanBytes` – 数字（长型）。

  以字节为单位的扫描数据的估计值。
- `PlanningTimeMillis` – 数字（长型）。

  处理请求所花费的时间。
- `QueueTimeMillis` – 数字（长型）。

  请求排队等待处理的时间。
- `WorkUnitsGeneratedCount` – 数字（长型）。

  生成的工作单元的数量。

操作

- StartQueryPlanning 操作（Python：start_query_plan）(p. 313)
- GetQueryState 操作（Python：get_query_state）(p. 314)
- GetWorkUnits 操作（Python：get_Work_Units）(p. 314)
- GetWorkUnit 结果操作（Python：get_work_unit_Results）(p. 315)
- GetQueryStartyStartyStartyInstistics 操作（Python：get_query_统计）(p. 316)

StartQueryPlanning 操作（Python：start_query_plan）

提交处理查询语句的请求。

此操作生成的工作单元可以使用`GetWorkUnits`只要查询状态为 WORKUNITS_SALABLE 或已完成，就会立即操作。

请求

- `QueryPlanningContext`–：必需 一个 查询规划上下文 (p. 312) 对象。

  包含有关查询计划信息的结构。
- `QueryString`–：必需 UTF-8 字符串，至少为 1 个字节。

  用作计划程序服务输入的 PartiQL 查询语句。

响应

输出的结构。

- `QueryId`– ：必需 UTF-8 字符串。

  计划查询操作的 ID 可用于获取作为操作结果生成的实际工作单元描述符。ID 还用于获取查询状态并作为输入`Execute`operation.

错误

- `InternalServiceException`
- `InvalidInputException`
- `AccessDeniedException`
- `ThrottledException`

GetQueryState 操作（Python：get_query_state）

返回之前提交的查询的状态。客户应该进行民意调查`GetQueryState`以便在检索工作单元之前监控计划的当前状态。查询状态只对进行初始调用的委托人可见`StartQueryPlanning`.

请求

- `QueryId`– ：必需 UTF-8 字符串，长度不少于 36 个字节或超过 36 个字节。

  计划查询操作的 ID。

响应

输出的结构。

- `Error` – UTF-8 字符串。

  在操作失败时显示错误消息。
- `State`– ：必需 UTF-8 字符串（有效值：`PENDING=""|WORKUNITS_AVAILABLE=""|ERROR=""|FINISHED=""|EXPIRED=""`）。

  之前提交的查询的状态。可能的状态包括：
  - 待定：查询处于待处理状态。
  - WORKUNITS_SALABLE：一些工作单元已准备好进行检索和执行。
  - 已完成：查询计划已成功完成，所有工作单元都已准备好进行检索和执行。
  - 错误：查询发生错误，例如无效的查询 ID 或后端错误。

错误

- `InternalServiceException`
- `InvalidInputException`
- `AccessDeniedException`

GetWorkUnits 操作（Python：get_Work_Units）

检索由`StartQueryPlanning`operation.

请求

- `NextToken` – UTF-8 字符串。

延续标记 (如果这是延续调用)。

- `PageSize` – 数字 (integer)。

  进入的每个页面的大小Amazon服务电话。这不会影响命令的输出中返回的项目数。设置较小的页面大小会导致对Amazon服务，每次调用检索的项目数较少。这有助于防止Amazon超时服务调用。
- `QueryId`–：必需 UTF-8 字符串，长度不少于 36 个字节或超过 36 个字节。

  计划查询操作的 ID。

响应

输出的结构。

- `NextToken` – UTF-8 字符串。

  对返回的标记列表进行分页的延续令牌 (如果列表的当前片段不是最后一个，则返回)。
- `QueryId`–：必需 UTF-8 字符串。

  计划查询操作的 ID。
- `WorkUnitRanges`–：必需 WorkUnit 范围 (p. 311) 对象数组。

  一个`WorkUnitRangeList`对象，该对象指定用于查询执行服务的工作单元 ID 的有效范围。

错误

- `WorkUnitsNotReadyYetException`
- `InternalServiceException`
- `InvalidInputException`
- `AccessDeniedException`
- `ExpiredException`

GetWorkUnit 结果操作（Python：get_work_unit_Results）

返回查询生成的工作单元。工作单位可以按任意顺序并行执行。

请求

- `QueryId`–：必需 UTF-8 字符串，长度不少于 36 个字节或超过 36 个字节。

  要获取结果的计划查询操作的 ID。
- `WorkUnitId`–：必需 数字 (long)，至多为 “无”。

  要获得结果的工作单元 ID。枚举生成的值`WorkUnitIdMin`到`WorkUnitIdMax`（含）`WorkUnitRange`在的输出中`GetWorkUnits`.
- `WorkUnitToken`–：必需 UTF-8 字符串，至少为 1 个字节。

  用于查询执行服务的工作令牌。令牌输出来自`GetWorkUnits`.

响应

输出的结构。

- `ResultStream` – Blob。

  从返回的行`GetWorkUnitResults`作为 Apache Arrow v1.0 消息的流进行操作。

错误

- `InternalServiceException`
- `InvalidInputException`
- `AccessDeniedException`
- `ExpiredException`
- `ThrottledException`

GetQueryStartyStartyStartyInstistics 操作（Python：get_query_统计）

检索有关计划和执行查询的统计信息。

请求

- `QueryId`–：必需 UTF-8 字符串，长度不少于 36 个字节或超过 36 个字节。

  计划查询操作的 ID。

响应

- `ExecutionStatistics` – 一个 执行统计 (p. 312) 对象。

  网络 ACL 和安全组都允许 (因此可到达您的实例) 的发起 ping 的`ExecutionStatistics`结构包含执行统计信息。
- `PlanningStatistics` – 一个 规划统计/统计 (p. 313) 对象。

  一个`PlanningStatistics`结构包含查询规划统计信息。
- `QuerySubmissionTime` – UTF-8 字符串。

  提交查询的时间。

错误

- `StatisticsNotReadyYetException`
- `InternalServiceException`
- `InvalidInputException`
- `AccessDeniedException`
- `ExpiredException`
- `ThrottledException`

异常

- IntoReadyetCreadyet 异常结构 (p. 317)
- WorkUnteReadyet 异常结构 (p. 317)

IntoReadyetCreadyet 异常结构

包含有关统计信息未准备就绪的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

WorkUnteReadyet 异常结构

包含与工作单元尚未准备就绪相关的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

ExcireExeption 结构

包含有关查询请求过期的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

CLOTTLEException 结构

包含有关限制查询请求的错误的详细信息。

字段

- `Message` – UTF-8 字符串。

  描述错误的消息。

存储 API

存储 API 允许您管理中受管理表的存储优化Amazon Lake Formation.

数据类型

StorageOptimizer 结构

描述存储优化程序的配置和详细信息的结构。

字段

- `StorageOptimizerType` – UTF-8 字符串（有效值：`compaction="COMPACTION" | garbage_collection="GARBAGE_COLLECTION" | index="INDEX" | copy_on_write="COPY_ON_WRITE" | all="GENERIC"`）。

  特定类型的存储优化器。支持的值为 `compaction`。
- `Config` – 键值对的映射数组。

  每个键是一个 UTF-8 字符串。

  每个值是一个 UTF-8 字符串。

  存储优化程序配置的映射。目前只包含一个键值对：`is_enabled`表示加速度为真或假。
- `ErrorMessage` – UTF-8 字符串。

  包含有关任何错误（如果存在）的信息的消息。

  当加速结果为已启用状态时，错误消息为空。

  当加速结果处于禁用状态时，该消息将描述错误或仅表示 “被用户禁用”。
- `Warnings` – UTF-8 字符串。

  包含有关任何警告（如果存在）的信息的消息。
- `LastRunDetails` – UTF-8 字符串。

  当加速结果为已启用状态时，将包含上次作业运行的详细信息。

操作

- ListTable StorageOptimizer 操作（Python：list_table_storage_Optimizer）(p. 318)
- UpdateTableStorageOptimizer 操作（Python：update_table_storage_Optimizer）(p. 319)

ListTable StorageOptimizer 操作（Python：list_table_storage_Optimizer）

返回与指定表关联的所有存储优化程序的配置。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  表的目录 ID。
- `DatabaseName`–：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  表所在数据库的名称。

- `TableName`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  表名称.
- `StorageOptimizerType` – UTF-8 字符串（有效值：`compaction="COMPACTION" | garbage_collection="GARBAGE_COLLECTION" | index="INDEX" | copy_on_write="COPY_ON_WRITE" | all="GENERIC"`）。

  要列出的特定类型的存储优化器。支持的值为 `compaction`。
- `MaxResults` – 数字（整数），不小于 1 或大于 1000。

  每次调用时要返回的存储优化器的数量。
- `NextToken` – UTF-8 字符串。

  延续标记 (如果这是延续调用)。

响应

- `StorageOptimizerList` – 存储优化器 (p. 318) 对象的数组。

  与表关联的存储优化程序的列表。
- `NextToken` – UTF-8 字符串。

  对返回的标记列表进行分页的延续令牌 (如果列表的当前片段不是最后一个，则返回)。

错误

- `EntityNotFoundException`
- `InvalidInputException`
- `AccessDeniedException`
- `InternalServiceException`

UpdateTableStorageOptimizer 操作（Python：update_table_storage_Optimizer）

更新表的存储优化程序的配置。

请求

- `CatalogId` – 目录 id 字符串，长度不少于 1 个字节或超过 255 个字节，与 Single-line string pattern (p. 320) 匹配。

  表的目录 ID。
- `DatabaseName`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  表所在数据库的名称。
- `TableName`– ：必需 UTF-8 字符串，长度不少于 1 个字节或超过 255 个字节，与匹配。Single-line string pattern (p. 320).

  要启用存储优化程序的表的名称。
- `StorageOptimizerConfig`– ：必需 键值对的映射数组。

每个键是一个 UTF-8 字符串（有效
值：compaction="COMPACTION"|garbage_collection="GARBAGE_COLLECTION"|index="INDEX"|copy_on_

每个值都是键值对的映射数组。

每个键是一个 UTF-8 字符串。

每个值是一个 UTF-8 字符串。

要启用存储优化程序的表的名称。

响应

- Result – UTF-8 字符串。

  操作失败的响应。

错误

- EntityNotFoundException
- InvalidInputException
- AccessDeniedException
- InternalServiceException

常见数据类型

常见数据类型介绍 Amazon 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 (p. 320) 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 (p. 320) 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}]*"

Lake Formation 角色和 IAM 权限参考

本章列出了一些建议Amazon Lake Formation角色和他们的建议Amazon 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” (p. 167).

Amazon Lake Formation角色

下表列出了建议的Amazon Lake Formation角色。

Lake Formation 角色

角色	描述
IAM 管理员（超级用户）	（必需）可以创建 IAM 用户和角色的用户。Has the`AdministratorAccess` Amazon托管策略。拥有所有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。
数据湖管理员	（必需）具有以下权限的用户：注册 Amazon S3 位置、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Amazon CloudTrail日志。IAM 权限少于 IAM 管理员，但足以管理数据湖。无法添加其他数据湖管理员。
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 Lake Formation 对爬网程序和工作流创建的 Data Catalog 表的权限的用户。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 (p. 113)。
数据分析人员	（可选）可以对数据湖运行查询的用户，例如Amazon Athena. 只有足够的权限运行查询。
工作流角色	（必需）代表用户运行工作流的角色。您可在从蓝图创建工作流时指定此角色。

角色 A 建议的权限

以下是针对每个角色的建议权限。不包括 IAM 管理员，因为该用户拥有所有资源的所有权限。

主题

数据湖管理员权限

Important

在以下策略中，将*<account-id>*具有有效的Amazon账号，并替换*<workflow_role>*使用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 (p. 326).

策略类型	策略
Amazon 托管策略	• AWSLakeFormationDataAdmin • AWSGlueConsoleFullAccess（可选） • CloudWatchLogsReadOnlyAccess（可选） • AWSLakeFormationCrossAccountManager（可选） • AmazonAthenaFullAccess（可选） 有关可选的Amazon托管策略，请参阅the section called “创建数据湖管理员” (p. 12).

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

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

（可选）内联策略（如果您的账户正在授予或接收跨账户 Lake Formation 权限）。此政策用于接受或拒绝Amazon RAM资源共享邀请，以及允许向组织授予跨账户权	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Effect": "Allow",</pre>

策略类型	策略
限。ram:EnableSharingWithAwsOrganization只有数据湖管理员才需要Amazon Organizations管理账户。	"Action": ["ram:AcceptResourceShareInvitation", "ram:RejectResourceShareInvitation", "ec2:DescribeAvailabilityZones", "ram:EnableSharingWithAwsOrganization"], "Resource": "*" }] }

数据工程师权限

Important

在以下策略中，将*<account-id>*具有有效的Amazon账号，并替换*<workflow_role>*具有工作流角色的名称。

策略类型	策略
Amazon 托管式策略	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:RemoveLFTagsFromResource", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags", "lakeformation:GetWorkUnits", "lakeformation:GetWorkUnitResults", "lakeformation:StartQueryPlanning", "lakeformation:GetQueryState", "lakeformation:GetQueryStatistics"], "Resource": "*" }] }
内联策略（用于对受管控表的操作，包括事务内的操作）	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow",

策略类型	策略

	<pre> "Action": [
	                "lakeformation:StartTransaction",
	                "lakeformation:CommitTransaction",
	                "lakeformation:CancelTransaction",
	                "lakeformation:ExtendTransaction",
	                "lakeformation:DescribeTransaction",
	                "lakeformation:ListTransactions",
	                "lakeformation:GetTableObjects",
	                "lakeformation:UpdateTableObjects",
	                "lakeformation:DeleteObjectsOnCancel"
	            ],
	            "Resource": "*"
	        }
	    ]
	}</pre>

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

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

数据分析员权限

策略类型	策略
Amazon 托管式策略	`AmazonAthenaFullAccess`

策略类型	策略

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，包括事务内的操作）	{
	"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": "*"
	}
	]
	}

工作流角色权限

此角色具有运行工作流所需的权限。您可在创建工作流时指定具有这些权限的角色。

Important

在以下策略中，将*<region>*具有有效的Amazon区域标识符（例如us-east-1），*<account-id>*具有有效的Amazon账号，*<workflow_role>*工作流的名称，以及*<your-s3-cloudtrail-bucket>*针对您的 Amazon S3 路径Amazon CloudTrail日志。

策略类型	策略
Amazon 托管式策略	AWSGlueServiceRole

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

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

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

Lake Formation

如果您在使用时遇到问题AmazonLake Formation，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。

Error: 的Lake Formation 权限不足 <Amazon S3 location>

有人试图在资源所指向的 Amazon S3 位置上创建或更改没有数据位置权限的数据目录资源。

如果数据目录数据库或表指向 Amazon S3 位置，则当您授予Lake Formation 权限时`CREATE_TABLE`要么`ALTER`，还必须授予的`DATA_LOCATION_ACCESS`该位置的权限。如果要向外部账户或组织授予这些权限，则必须包括授予选项。

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账户中的委托人（用户或角色）授予权限。在授予`DATA_LOCATION_ACCESS`从另一个账户获得的权限，则必须指定目录 ID (Amazon账户 ID）的所有者账户。所有者账户是注册该地点的账户。

有关更多信息，请参阅 底层数据访问控制 (p. 238) 和 授予数据位置权限 (p. 141)。

Error: “Glue API 的加密密钥权限不足”

有人试图在没有的情况下授予 Lake Formation 权限Amazon Identity and Access Management(IAM) 权限Amazon KMS加密数据目录的加密密钥。

我的Amazon Athena或者使用清单的 Amazon Redshift 查询失败

Lake Formation 不支持使用清单的查询。

Error: “Lake Formation 权限不足：必需在目录上创建标记”

用户/角色必须是数据湖管理员。

跨账户访问问题排除

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，但收件人看不到资源

- 收件人账户中的用户是否为数据湖管理员？ 共享时，只有数据湖管理员才能看到该资源。
- 您是否使用命名资源方法与组织外部的账户共享？ 如果是，则收件人账户的数据湖管理员必须接受资源共享邀请Amazon Resource Access Manager(Amazon RAM)。

  有关更多信息，请参阅 the section called “接受Amazon RAM资源共享邀请” (p. 124)。
- 您是否在中使用账户级（数据目录）资源策略Amazon Glue? 如果是，则如果您使用 named resources 方法，则必须在策略中包含一个特殊语句，以授权Amazon RAM代表您共享保单。

  有关更多信息，请参阅 the section called “使用这两者来管理跨账户权限Amazon GlueLake Format” (p. 250)。
- 您是否已经Amazon Identity and Access Management(IAM) 权限是否需要授予跨账户访问权限？

  有关更多信息，请参阅 the section called “跨账户访问权限前提条件” (p. 243)。
- 你授予权限的资源不得具有任何 Lake Formation 权限授予`IAMAllowedPrincipals`组中)。
- 有没有`deny`账户级策略中关于资源的声明？

收件人账户中的委托人可以查看数据目录资源，但无法访问基础数据

收款人账户中的委托人必须具有必需的Amazon 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 (p. 246)。

Error: “关联失败，因为呼叫者未获得授权”Amazon RAM资源共享邀请

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
        }
    ]
}
```

出现此错误是因为`glue:PutResourcePolicy`被调用Amazon Glue当接收账户接受资源共享邀请时。要解决该问题，请允许`glue:PutResourcePolicy`由接收账户使用的代入角色执行的操作。

Error: “未获得授予资源权限的授权”

试图授予对另一个账户拥有的数据库或表的跨账户权限。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。

Error: “访问被拒绝，无法检索Amazon组织信息”

您的账户是AmazonOrganizations 管理账户，并且您没有检索组织信息（例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 (p. 244)。

Error: “<organization-ID>未找到组织”

尝试与组织共享资源，但未启用与组织共享。启用与组织共享资源。

有关更多信息，请参阅 。启用与 共享AmazonOrganizations中的Amazon RAM用户指南.

Error: “Lake Formation 权限不足 非法组合”

用户共享了数据目录资源，而 Lake Formation 权限被授予`IAMAllowedPrincipals`资源的组。用户必须撤销所有 Lake Formation 权限`IAMAllowedPrincipals`在共享资源之前。

ConcurrentModificationException 对外部账户的授予/撤销请求时

当用户在 LF-tag 策略上为委托人发出多个并发授予和/或撤销权限请求时，Lake Formation 会抛出 ConcurrentModificationException. 用户需要catch 异常并重试失败的授予/撤销请求。使用批处理版本的`GrantPermissions`/`RevokePermissions`API 操作-the section called “批量格兰特权限（batch_grant_权限）” (p. 278)和the section called “批量撤销权限（batch_revoke_权限）” (p. 279)通过减少并发授予/撤销请求的数量在一定程度上缓解了这个问题。

蓝图和工作流程疑难解答

使用此处的信息可帮助您诊断和修复蓝图和工作流程问题。

主题

我的蓝图失败，“User: <user-ARN>is not authorized to performPassRole 在资源上:<role-ARN>”

没有足够权限传递所选角色的用户尝试创建蓝图。

更新用户的 IAM 策略以使其能够传递角色，或者要求用户选择具有所需密码角色权限的其他角色。

有关更多信息，请参阅 Lake Formation 角色和 IAM 权限参考 (p. 322)。

“User: <user-ARN>is not authorized to perform:PassRole 在资源上:<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 “为工作流创建 IAM 角色” (p. 11)。

我的工作流程中的爬虫失败，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。为该角色授予`CREATE_TABLE`数据库的权限。

我的工作流程中的爬虫失败，并显示 “发生了错误 (AccessDeniedException) 在调用 CreateTable operation.”

一个可能的原因是工作流角色在目标存储位置上没有数据位置权限。给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`” (p. 173)。

的已知问题Amazon Lake Formation

查看 Amazon Lake Formation 的以下已知问题。

主题

对表元数据筛选的限制

Amazon 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类似的 API 检索有关表的元数据时`glue:GetTable`，表对象中的列列表仅包含他们有权访问的字段。务必要了解这种元数据筛选的局限性。

尽管 Lake Formation 提供了有关集成服务列权限的元数据，但实际筛选查询响应中的列是集成服务的责任。支持列级筛选的 Lake Formation 客户端，包括Amazon Athena、Amazon Redshift Spectrum 和 Amazon EMR，根据在 Lake Formation 注册的列权限筛选数据。用户将无法读取他们无权访问的任何数据。目前，Amazon GlueETL 不支持列筛选。

> Note
>
> EMR 集群不完全由Amazon. 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）存储在`Parameters`map as table 属性。这些属性未经修改返回，任何用户都可以通过以下方式访问这些属性`SELECT`任何列的权限。

例如，AvroSerDe将表架构的 JSON 表示 JSON 存储在名为的表属性中`avro.schema.literal`，有权访问表的所有用户可以使用该表。我们建议您避免在表属性中存储敏感信息，并注意用户可以学习 Avro 格式表的完整架构。此限制特定于表的元数据。

Amazon Lake Formation移除任何以开头的表属性`spark.sql.sources.schema`CH when respon`glue:GetTable`或类似的请求（如果来电者没有）`SELECT`表中所有列的权限。这可以防止用户获得有关使用 Apache Spark 创建的表的其他元数据的访问权限。在 Amazon EMR 上运行时，Apache Spark 应用程序仍然可以读取这些表，但可能无法应用某些优化，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 会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

如果您使用列级权限排除某列，然后重命名该列，则该列不会再从查询中排除，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，然后从架构中删除一列，则查询可能会返回错误的数据，并且可能不遵守列级权限。

解决办法：请改为创建新表。

表分区必须在通用路径下添加

Lake Formation 希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。当你使用搜寻器向目录添加分区时，这可以无缝运行。但是，如果您手动添加分区，并且这些分区不在父表中设置的位置下，则无法访问数据。

在工作流程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程时，如果目标数据库不存在，则可以创建目标数据库。当您这样做时，登录的 IAM 用户将获得`CREATE_TABLE`对创建的数据库的权限。但是，工作流程生成的搜寻器在尝试创建表时会扮演工作流的角色。这失败了，因为角色没有`CREATE_TABLE`对数据库的权限。

解决办法：如果您在工作流程设置期间通过控制台创建数据库，则在运行工作流程之前，必须为与工作流程关联的角色指定为`CREATE_TABLE`对您刚创建的数据库的权限。

删除然后重新创建用户时出现问题

以下场景会导致错误的 Lake Formation 权限返回`lakeformation:ListPermissions`：

1. 创建用户并授予 Lake Formation 权限。
2. 请删除用户。
3. 使用同一名称重新创建用户。

`ListPermissions`返回两个条目，一个条目用于旧用户，另一个条目用于新用户。如果您尝试撤消授予旧用户的权限，则新用户的权限将被撤销。

`GetTables`和`SearchTables`API 不会更新其值`IsRegisteredWithLakeFormation`参数

数据目录 API 操作存在一个已知的限制，例如`GetTables`和`SearchTables`请勿更新其值`IsRegisteredWithLakeFormation parameter`，并返回默认值，即 false。建议使用`GetTable`用于查看正确值的 API`IsRegisteredWithLakeFormation parameter`.

数据目录 API 操作不会更新`IsRegisteredWithLakeFormation`参数

数据目录 API 操作存在一个已知的限制，例如`GetTables`和`SearchTables`请勿更新其值`IsRegisteredWithLakeFormation`参数，并返回默认值，即 false。建议使用`GetTable`用于查看正确值的 API`IsRegisteredWithLakeFormation`参数。

Lake Formation 操作不支持Amazon Glue架构注册表

Lake Formation 操作不支持Amazon Glue包含 a 的表`SchemaReference`中的`StorageDescriptor`将用于架构注册表.

Amazon Lake Formation 的文档历史记录

下表介绍对 Amazon Lake Formation 文档的一些重要更改。

update-history-change	update-history-description	update-history-date
跨账户资源共享的更新 (p. 335)	添加了如何操作的说明跨账户资源份额在 Lake Formation 工作。记录了对AWSLakeFormationCrossAccountManager政策。	2022 年 5 月 6 日
新教程 (p. 335)	添加了有关创建 goverend 表、保护数据湖和共享数据湖的新教程。有关更多详细信息，请参阅试用部分。	2022 年 4 月 20 日
新 Lake Formation 登陆页面 (p. 335)	更新了Lake Formation登录页面，包括教程的链接，这些链接提供了有关如何使用 Lake Formation 构建数据湖、采集数据、共享和保护数据湖的分步说明。	2022 年 4 月 20 日
Support 凭据自动售货 (p. 335)	添加了有关凭据自动售货机的信息，该信息支持 Lake Formation，允许第三方服务通过使用凭据自动售货机 API 操作与 Lake Formation 集成 有关更多信息，请参阅 。Lake Formation 中的凭证自动售货机如何.	2022 年 2 月 28 日
Support 受管理的表和高级数据筛选 (p. 335)	添加了有关受管表的信息，这些表支持 ACID 事务、自动数据压缩和时间旅行查询。添加了有关创建数据筛选器以支持列级安全性、行级安全性和单元格级安全性的信息。有关更多信息，请参阅 。Lake Formation 中的受管理表和Lake Formation 中的数据过滤和细胞级安全.	2021 年 11 月 30 日
对 VPC 接口终端节点的 Support (p. 335)	添加了有关为 Lake Formation 创建虚拟私有云 (VPC) 接口终端节点的信息，以便您的 VPC 与 Lake Formation 之间的通信完全在Amazon网络。有关更多信息，请参阅 。将 Lake Formation 与 VPC 端点结合使用.	2021 年 10 月 11 日
支持 VPC 终端节点策略 (p. 335)	添加了有关对 Lake Formation 中的 Virtual Private Cloud (VPC) 端点策略的支持的信息。有关更多信	2021 年 10 月 11 日

	息，请参阅 。将 Lake Formation 与 VPC 端点结合使用.	
Support 基于标签的访问控制 (p. 335)	基于 Lake Formation 标签的访问控制提供了一种新的、更具可扩展性的方式，通过使用 LF-Tags 来管理对数据目录资源和底层数据 有关更多信息，请参阅 。Lake Formation 标签的访问控制.	2021 年 5 月 7 日
在亚马逊 EMR 上筛选数据的新选择加入要求。 (p. 335)	添加了有关选择加入以允许亚马逊 EMR 筛选由 Lake Formation 管理的数据的要求的信息。有关更多信息，请参阅 。允许在亚马逊 EMR 上过滤数据.	2020 年 10 月 9 日
Support 授予对数据目录数据库的完全跨账户权限 (p. 335)	添加了有关授予对跨数据目录数据库的完整 Lake Formation 权限Amazon账户，包括`CREATE_TABLE`. 有关更多信息，请参阅 。共享数据目录数据库.	2020 年 10 月 1 日
对该项的支持Amazon Athena 用户通过 SAML 进行身份验证。 (p. 335)	添加了有关对通过 JDBC 或 ODBC 驱动程序进行连接并通过 SAML 身份提供商（如 Okta 和 Microsoft Active Directory 联合服务（AD FS）进行身份验证的 Athena 用户的支持信息。有关更多信息，请参阅 。Amazon与 Lake Formation 的服务集成.	2020 年 9 月 30 日
使用加密的数据目录 Support 跨账户访问 (p. 335)	添加了有关在加密数据目录时授予跨账户权限的信息。有关更多信息，请参阅 。跨账户访问前提条件.	2020 年 7 月 30 日
对数据湖的跨账户访问的 Support (p. 335)	添加了有关授予的信Amazon Lake Formation数据目录数据库和表对外部的权限Amazon帐户和组织，以及访问从外部帐户共享的数据目录对象。有关更多信息，请参阅 。跨账户访问.	2020 年 7 月 7 日
与亚马逊集成 QuickSight (p. 335)	添加了有关如何向 Amazon 授予 Lake Formation 权限的信息 QuickSight 企业版用户，以便他们可以访问驻留在注册 Amazon S3 位置的数据集。有关更多信息，请参阅 。授予数据目录权限.	2020 年 6 月 29 日
设置和入门章节的更新 (p. 335)	重组并改进了设置和入门章节。更新了推荐的Amazon Identity and Access Management数据湖管理员的 (IAM) 权限。	2020 年 2 月 27 日

对该项的支持Amazon Key Management Service (p. 335)	添加了有关 Lake Formation 的支持的信息Amazon Key Management Service(Amazon KMS) 简化了集成服务的设置，以便在注册的 Amazon Simple Storage Service (Amazon S3) 位置读写加密数据。添加了有关如何注册加密的 Amazon S3 位置的信息。Amazon KMS keys. 有关更多信息，请参阅 将 Amazon S3 位置添加到您的数据湖 (p. 102)。	2020 年 2 月 27 日
蓝图和数据湖管理员 IAM 策略的更新 (p. 335)	澄清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
安全章节重写和升级章节修订 (p. 335)	改进了安全性和升级章节。	2019 年 10 月 29 日
超级权限替换所有权限 (p. 335)	更新了 “安全” 和 “升级” 章节以反映权限的替换情况`All`和`Super`.	2019 年 10 月 10 日
补充、更正和澄清 (p. 335)	根据反馈进行了补充、更正和澄清。修改了安全章节。更新了安全性和升级章节以反映组的替换情况`Everyone`和`IAMAllowedPrincipals`.	2019 年 9 月 11 日
新指南 (p. 335)	这是 Amazon Lake Formation 开发人员指南的初始版本。	2019 年 8 月 8 日

Amazon词汇表

有关最新Amazon术语，请参阅《Amazon一般参考》中的Amazon术语表。

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。