LakeFormation

用户指南

文档版本 06

发布日期 2025-01-10

华为云计算技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目 录

1 准备工作

1.1 注册华为云账号

在使用华为云服务之前您需要申请华为云账号。通过此账号，您可以使用所有华为云服务，并且只需为您所使用的服务付费。

如果您已有一个华为云账户，请跳到下一个任务。如果您还没有华为云账户，请参考以下步骤创建。

操作步骤

步骤1 打开**华为云网站**。

步骤2 单击“注册”，根据提示信息完成注册。详情请参考**注册华为账号并开通华为云**。

注册成功后，系统会自动跳转至您的个人信息界面。

步骤3 个人或企业账号实名认证请参考**实名认证**。

----结束

1.2 配置云服务授权

首次使用LakeFormation服务需要进行服务授权，授权相关云资源的权限。

云服务授权操作

步骤1 使用**注册华为云账号**创建的用户登录管理控制台。

步骤2 在服务列表中选择“大数据 > 湖仓构建 LakeFormation”，进入“服务授权”页面。

- IAM ReadOnlyAccess：实例运行的过程中需要有获取用户的用户组和用户信息的权限。
- OBS OperateAccess：实例的存储功能，需要获取访问对象存储等服务的权限。
- OBS AccessLabel：实例的权限控制功能，需要有打标签的能力。
- OBS Bucket Lifecycle：实例的生命周期管理功能，需要有操作生命周期的权限。

- VPCEndpoint Administrator：实例的接入管理功能，需要有操作VPC终端节点的权限。
- DNS FullAccess：实例的接入管理功能，需要有操作DNS内网域名的权限。

步骤3 勾选“同意LakeFormation服务声明”，并单击“同意授权”，完成服务授权。

说明

同意授权后，LakeFormation将在统一身份认证服务中创建名为lakeformation_admin_trust的委托，在使用LakeFormation服务期间，请不要删除该委托。

如果自动创建委托失败，则需要您登录到“统一身份认证服务”管理控制台，对委托进行删除或联系管理员增加限额并确认当前用户是否有委托创建权限。

----结束

1.3 权限管理

1.3.1 创建用户并授权使用 LakeFormation

如果需要对LakeFormation服务进行精细的权限管理，您可以使用**统一身份认证服务**（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的账号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用云服务资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将云服务资源委托给更专业、高效的其他账号或者云服务，这些账号或者云服务可以根据权限进行代运维。

如果账号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节。

本章节为您介绍对用户授权的方法，操作流程如**图1-1**所示。

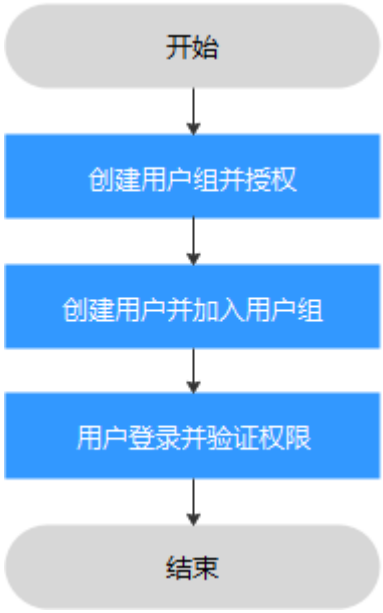
前提条件

给用户组授权之前，可参考**LakeFormation服务权限**了解用户组可以添加的LakeFormation权限，并结合实际需求进行选择。

如果您需要对除LakeFormation之外的其他服务授权，可参考**系统权限**查看IAM支持服务的所有策略。

操作流程

图 1-1 给用户授权 LakeFormation 权限流程



1. **创建用户组并授权**
   在IAM控制台创建用户组，并授予LakeFormation服务对应权限。
2. **创建用户并加入用户组**
   在IAM控制台创建用户，并将其加入**1.创建用户组并授权**中创建的用户组。
3. **用户登录**并验证权限
   以新创建的用户登录云服务控制台，切换至授权区域，验证权限是否生效。
   例如：
   在“服务列表”中选择LakeFormation服务，进入总览界面，单击右上角“购买实例”，实例创建界面正常展示，表示“lakeformation:role:create”权限已生效。

LakeFormation 服务权限

默认情况下，管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。

权限根据授权精细程度分为角色和策略。

- 角色：IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。

说明

- IAM项目只读授权指导：当租户管理员需要给某个子用户分配LakeFormation服务在某个IAM项目下的只读权限。可以给该用户创建一个用户组，同时在用户组将LakeFormation ReadOnlyAccess系统策略授权给指定IAM项目即可。
- 企业项目授权指导：当租户管理员需要给某个子用户分配LakeFormation服务在某个企业项目下的所有操作权限。可以给该用户创建一个用户组，同时在用户组中将LakeFormation CommonAccess授权给全局，将LakeFormation FullAccess授权给指定企业项目即可。

- 策略：IAM最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。多数细粒度策略以API接口为粒度进行权限拆分，LakeFormation的自定义IAM策略操作可参考**创建LakeFormation自定义IAM策略**。

表 1-1 LakeFormation 系统策略

系统角色/策略名称	描述	类别	依赖关系
LakeFormation FullAccess	LakeFormation管理员权限，拥有该权限的用户可以操作并使用所有LakeFormation服务功能。	系统策略	• IAM AgencyFullAccess • OBS OperateAccess • VPC FullAccess • VPCEndpoint FullAccess
LakeFormation ReadOnlyAccess	LakeFormation只读权限，拥有该权限的用户可以执行LakeFormation所有查询类功能。	系统策略	• IAM ReadOnlyAccess • OBS ReadOnlyAccess • VPC ReadOnlyAccess • VPCEndpoint ReadOnlyAccess
LakeFormation CommonOperations	LakeFormation基础权限，包含LakeFormation服务协议查看/授权/取消，以及OBS、TMS等周边依赖服务的基础权限集合。	系统策略	• IAM ReadOnlyAccess • OBS ReadOnlyAccess • VPC FullAccess • VPCEndpoint FullAccess

表 1-2 LakeFormation 的 IAM 权限列表

操作类型	操作项	描述
只读	lakeformation:access:describe	查询接入客户端。
	lakeformation:agency:describe	查询委托。
	lakeformation:catalog:describe	查询Catalog元数据。
	lakeformation:configuration:describe	查询配置。
	lakeformation:credential:describe	查询认证信息。
	lakeformation:database:describe	查询数据库元数据。
	lakeformation:file:describe	查询文件。
	lakeformation:function:describe	查询函数元数据。
	lakeformation:group:describe	查询用户组以及关联角色关系。
	lakeformation:instance:describe	查询实例。
	lakeformation:instance:listAuthorizedLocation	查询已授权给LakeFormation服务的OBS路径。
	lakeformation:instanceJob:describe	查询实例级任务。
	lakeformation:job:describe	查询任务。
	lakeformation:metadataEvent:describe	查询元数据事件。
	lakeformation:obs:describe	查询OBS桶列表。
	lakeformation:part:describe	查询分区。
	lakeformation:policy:describe	查询权限策略。
	lakeformation:policy:export	批量查询权限策略。
	lakeformation:role:describe	查询角色。
	lakeformation:table:describe	查询表元数据。
	lakeformation:tableFile:describe	查询文件。
	lakeformation:tableFileGroup:describe	查询表文件组元数据。

操作类型	操作项	描述
	lakeformation:tag:describe	查询资源标签。
	lakeformation:user:describe	查询用户以及关联角色关系。
写	lakeformation:access:create	创建接入客户端。
	lakeformation:access:delete	删除接入客户端。
	lakeformation:agency:create	创建委托。
	lakeformation:agency:drop	删除委托。
	lakeformation:catalog:alter	修改Catalog元数据。
	lakeformation:catalog:create	创建Catalog元数据。
	lakeformation:catalog:drop	删除Catalog元数据。
	lakeformation:database:alter	修改数据库元数据。
	lakeformation:database:create	创建数据库元数据。
	lakeformation:database:drop	删除数据库元数据。
	lakeformation:dataset:create	创建数据集元数据。
	lakeformation:file:create	创建文件。
	lakeformation:file:drop	删除文件。
	lakeformation:file:alter	修改文件。
	lakeformation:function:alter	修改函数元数据。
	lakeformation:function:create	创建函数元数据
	lakeformation:function:drop	删除函数元数据。
	lakeformation:group:alter	修改用户组以及关联角色关系。
	lakeformation:instance:access	申请接入服务。
	lakeformation:instance:alter	修改实例。
	lakeformation:instance:create	创建实例。
	lakeformation:instance:drop	删除实例。
	lakeformation:instanceJob:alter	修改任务。
	lakeformation:instanceJob:create	创建任务。
	lakeformation:instanceJob:drop	删除任务。
	lakeformation:instanceJob:exec	执行实例级任务。

操作类型	操作项	描述
	lakeformation:instance:createSubscriber	创建元数据事件订阅者。
	lakeformation:instance:deleteSubscriber	删除元数据事件订阅者。
	lakeformation:job:alter	修改任务。
	lakeformation:job:create	创建任务。
	lakeformation:job:drop	删除任务。
	lakeformation:job:exec	执行任务。
	lakeformation:model:create	创建模型元数据。
	lakeformation:metadata:restore	恢复元数据。
	lakeformation:part:alter	修改分区。
	lakeformation:part:drop	删除分区。
	lakeformation:part:create	创建分区。
	lakeformation:policy:create	创建权限策略。
	lakeformation:policy:delegate	将权限策略委托给其他授权主体。
	lakeformation:policy:drop	删除权限策略。
	lakeformation:role:alter	修改角色以及关联用户组关系。
	lakeformation:role:create	创建角色。
	lakeformation:role:drop	删除角色。
	lakeformation:table:alter	修改表元数据。
	lakeformation:table:create	创建表元数据。
	lakeformation:table:drop	删除表元数据。
	lakeformation:tableFile:alter	修改表文件。
	lakeformation:tableFile:create	创建表文件。
	lakeformation:tableFile:drop	删除表文件。
	lakeformation:tableFileGroup:alter	修改表文件组元数据。
	lakeformation:tableFileGroup:create	创建表文件组元数据。
	lakeformation:tableFileGroup:drop	删除表文件组元数据。

操作类型	操作项	描述
	lakeformation:transaction:operate	操作事务。
	lakeformation:user:alter	修改用户以及关联角色关系。
权限管理	lakeformation:accessService:grant	授权接入服务。
	lakeformation:accessTenant:grant	授权接入租户。
	lakeformation:accessAgency:describe	查询接入委托信息。
	lakeformation:accessService:describe	查看接入服务。
	lakeformation:agreement:describe	查询服务协议授权。
	lakeformation:agreement:cancel	取消服务协议授权。
	lakeformation:agreement:grant	授权服务协议授权。
	lakeformation:instance:authorizeLocation	授权将OBS路径授权给LakeFormation服务。
	lakeformation:instance:cancelAuthorizeLocation	取消授权OBS路径。

1.3.2 创建 LakeFormation 自定义 IAM 策略

如果系统预置的LakeFormation权限，不满足您的授权要求，可以创建自定义策略。

目前华为云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。
- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见：**创建自定义策略**。

本章为您介绍常用的LakeFormation自定义策略样例。

LakeFormation 自定义策略样例

- 示例1：授权用户批量LakeFormation只读权限

```
{
    "Version": "1.1",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Action": [
                "lakeformation:instance:describe",
                "lakeformation:role:describe",
                "lakeformation:policy:export",
                "lakeformation:group:describe",
                "lakeformation:function:describe",
                "lakeformation:catalog:describe",
                "lakeformation:policy:describe",
                "lakeformation:table:describe",
                "lakeformation:database:describe"
            ]
        }
    ]
}
```

- 示例2：拒绝用户删除数据

  拒绝策略需要同时配合其他策略使用，否则没有实际作用。用户被授予的策略中，一个授权项的作用如果同时存在Allow和Deny，则遵循**Deny优先原则**。

  如果您给用户授予Admin的系统策略，但不希望用户拥有删除LakeFormation的Catalog、数据库、表的权限，您可以创建一条拒绝删除云服务的自定义策略，然后同时将Admin和拒绝策略授予用户，根据Deny优先原则，则用户可以对LakeFormation执行除了删除Catalog、数据库、表外的所有操作。

  拒绝策略示例如下：

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Deny",
            "Action": [
                "lakeformation:database:drop",
                "lakeformation:table:drop",
                "lakeformation:catalog:drop"
            ]
        }
    ]
}
```

- 示例3：多个云服务同时授权项策略

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "obs:bucket:CreateBucket"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:table:create",
                "lakeformation:database:create",
                "lakeformation:catalog:create"
            ]
        }
    ]
}
```

2 管理 LakeFormation 实例

2.1 创建 LakeFormation 实例

在使用LakeFormation时，您首先需要创建一个实例，后续的操作，如管理元数据、设置元数据权限等，都是基于您创建的实例进行的。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 单击页面右上角“立即购买”或“购买实例”，进入购买实例页面。

首次创建实例时界面显示“立即购买”，如果界面已有LakeFormation实例则显示为“购买实例”。

步骤4 按照需求配置以下参数。

表 2-1 购买 LakeFormation 实例

参数	参数说明	样例
类型	选择实例类型。 ● 共享：共享型实例之间，通过资源复用换取CCE集群或GaussDB(for MySQL)实例等资源的使用率最大化。 ● 独享：按照每秒查询率（QPS）上限和元数据使用量进行计费。	独享
计费模式	实例的计费模式。 ● 按需收费：按照LakeFormation实例实际使用时长计费。	按需收费
项目	选择实例所属的项目。	xxx

参数	参数说明	样例
名称	自定义LakeFormation实例名称。	lakeformation-test
QPS	每秒最大请求数。如果“实例类型”为“共享型”，则无需配置该参数。 LakeFormation将统计用户当前的元数据对象使用量，按量收费。	10000
企业项目	选择集群所属的企业项目。如果当前无可用企业项目，可以单击“新建企业项目”进行创建。 企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理。	xxx
描述	当前实例的描述信息。	-
标签	在标签键/值输入框输入内容后单击“添加”，即可添加标签。 如果您需要使用同一标签识别多种云资源，即所有服务均可在标签输入框下拉选择同一标签，可以单击“查看预定义标签”创建预定义标签。	-

步骤5 单击“立即购买”，确认配置的相关信息并支付。

步骤6 单击“返回控制台”，在控制台即可查看新创建的LakeFormation实例信息。

说明

创建实例时需要注意配额提醒。当资源配额不足时，建议按照界面提示申请足够的资源，再创建实例。

等待实例状态变为“运行中”表示实例已创建成功。

实例创建成功后，可以查看实例的基本信息及数据概况。

----结束

实例创建失败

如果实例创建失败，失败任务会自动转入“创建失败的实例”页面。

选择“大数据 > 湖仓构建 LakeFormation”，在“总览”页面单击左上方“创建失败的实例”，在打开的窗口中查看创建失败的实例信息。

2.2 配置 LakeFormation 实例

LakeFormation实例创建完成后，可以在实例“总览”页面对LakeFormation实例进行变更实例规格、设置默认实例等操作。

变更实例规格：对当前实例的QPS规格进行变更。仅独享型实例支持该操作。

设置默认实例：设置当前实例为默认实例。如果其他服务对接LakeFormation实例时，没有指定具体的实例ID，该操作将会修改服务对接的实例。

变更实例规格

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例。

步骤4 单击页面右上角“规格变更”，进入“规格变更”页面。如果当前实例为共享型，则不支持该操作。

步骤5 选择需要变更的QPS规格，单击“下一步”。

步骤6 确认当前实例的信息，及变更前后的实例规格后，单击“提交”。

----结束

设置默认实例

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例。

步骤4 查看实例“基本信息”中“是否为默认实例”的参数值。

- “true”表示当前实例为默认实例。
- “false”表示当前实例不为默认实例。

如果当前实例为默认实例，则无“设为默认实例”按钮。

步骤5 如果需要设置当前实例为默认实例，请单击页面右上角“设为默认实例”，勾选操作影响后单击“确定”，将当前实例设置为默认实例。

须知

如果其他服务对接LakeFormation实例时，没有指定具体的LakeFormation实例ID，系统将自动访问默认实例，变更默认实例后，可能对使用LakeFormation的周边服务产生影响，请谨慎操作。

----结束

2.3 删除 LakeFormation 实例

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例。

步骤4 单击页面右上角“删除当前实例”。

步骤5 在弹出的确认窗口中确认删除影响并勾选确认操作，单击“确定”后等待实例删除成功。

说明

- 删除后，实例会放在回收站，并继续计费直到从回收站删除。
  如果需要恢复已删除的实例，可在左侧导航栏单击“回收站”，单击“操作”列的“还原”，单击“确定”。
- 实例在回收站存放超过1天后，会自动删除，无法恢复。
- 为防止您的业务受到影响，实例在回收站中放置15分钟后，才能强制删除。

----结束

3 管理元数据

3.1 创建元数据存储路径

LakeFormation元数据映射的数据文件和目录存储在OBS并行文件系统中。在创建LakeFormation元数据之前，需要提前创建数据存储使用的OBS并行文件系统。

如果已存在可用的OBS并行文件系统，可跳过该章节操作。

操作步骤

步骤1 登录管理控制台。

步骤2 在页面左上角单击，选择“存储 > 对象存储服务”，进入对象存储服务页面。

步骤3 选择“并行文件系统 > 创建并行文件系统”，进入创建页面，配置相关参数后单击“立即创建”。

- 文件系统名称：根据界面要求设置并行文件系统名称，例如“lakeformation-test”。
- 其他参数根据实际情况选择。

步骤4 在并行文件系统页面，单击已创建的文件系统名称，例如“lakeformation-test”。

步骤5 在左侧导航栏选择“文件”，单击“新建文件夹”，填写待创建的文件夹名称，单击“确定”。继续单击该文件夹名称，单击“新建文件夹”，可以创建其子文件夹。

参考该步骤，依次创建用于存放元数据的路径，例如：

- Catalog存储路径：lakeformation-test/catalog1
- 数据库存储路径：lakeformation-test/catalog1/database1
- 数据表存储路径：lakeformation-test/catalog1/database1/table1、lakeformation-test/catalog1/database1/table2
- 函数存储路径：lakeformation-test/catalog1/database1/udf1

----结束

3.2 管理 Catalog

数据目录（Catalog）是元数据管理对象，它可以包含多个数据库。

用户可在LakeFormation中创建并管理多个Catalog，用于不同外部集群的元数据隔离。

前提条件

- 已创建LakeFormation实例，且实例处于正常运行状态。
- Catalog数据存储到OBS中，当前用户需具有OBS相关操作权限。
- 已参考**创建元数据存储路径**提前创建了用于存储Catalog数据的OBS并行文件系统。

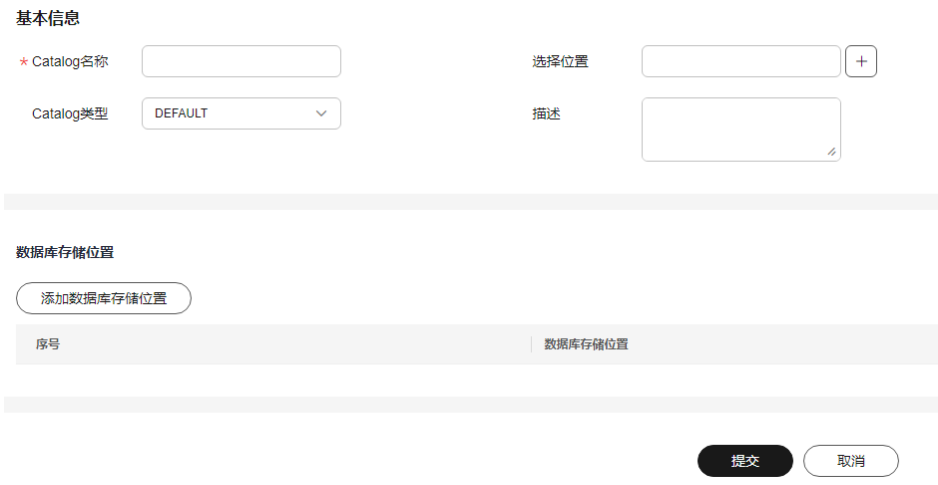
管理 Catalog

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“元数据 > Catalog”。

步骤4 单击“创建Catalog”，配置相关参数。



1. 在“基本信息”区域配置以下信息。

表 3-1 创建 Catalog

参数	参数说明
Catalog名称	填写待创建Catalog名称。 只能包含字母、数字和下划线，长度为1~256个字符。

参数	参数说明
Catalog类型	选择Catalog类型： – DEFAULT – CLICKHOUSE
选择位置	Catalog数据存储在OBS并行文件系统中的位置。可选参数。 单击“+”，选择位置后，单击“确定”。 – 如果配置该参数，则所选位置只能以“obs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1”。如果没有合适的并行文件系统，可以单击“前往OBS创建”进行创建。 – 建议选择未被其他Catalog选中的文件夹。
描述	所创建Catalog的描述信息。 长度为0~4000字节，1个中文字符对应3个字节。

2. （可选）单击“数据库存储位置”区域中的“添加数据库存储位置”。单击“+”可按照需求手动选择数据库存储位置，单击“确定”。支持添加多条。

   说明

   “数据库存储位置”为可选参数。如果配置了该参数，则该Catalog下的数据库位置必须选择为该Catalog“数据库存储位置”的子路径、或该Catalog“选择位置”的子路径。

3. 单击“提交”。

步骤5 创建完成后，即可在“Catalog”页面查看Catalog相关信息。

- 单击“操作”列的“编辑”可以修改Catalog配置信息。
- 单击“操作”列的“数据库”，可以查看当前Catalog下的数据库信息。
- 在“更多”中可以为当前Catalog进行授权、查看权限等操作。

步骤6 如果需要删除Catalog，可以选择“更多 > 删除”，确认操作影响后单击“确定”。

删除Catalog前需要提前删除该Catalog下的数据库。

说明

删除元数据时如果同步删除文件，数据将移入对应OBS桶的回收站（“lake-formation-trash-dir/table_id”OBS路径）目录下。

----结束

3.3 管理数据库

LakeFormation的一个Catalog下可以创建多个数据库，通过集中式的元数据管理，可以有效提升数据资产价值。

前提条件

- 已创建LakeFormation实例，且实例处于正常运行状态。
- 已创建待添加数据库的Catalog。
- 已参考**创建元数据存储路径**提前创建了用于存储数据库的OBS并行文件系统。

管理数据库

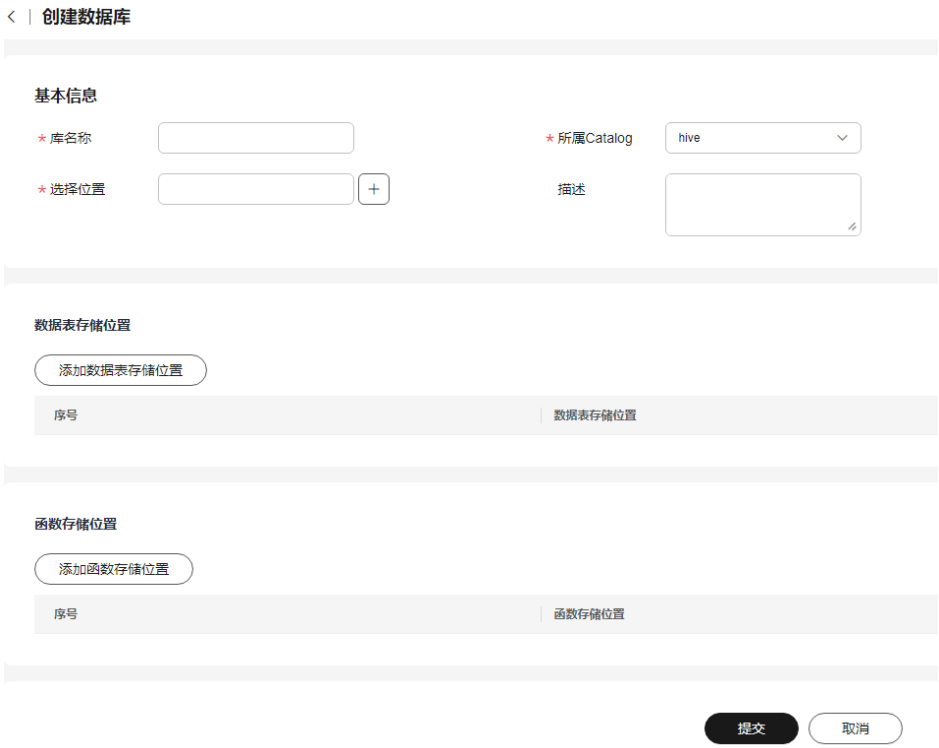
步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“元数据 > 数据库”。

步骤4 在右上角“Catalog”后的下拉框中选择待创建数据库所属的Catalog名称。可以查看当前Catalog中包含的数据库。

步骤5 单击“创建数据库”，配置相关参数。



1. 在“基本信息”区域配置以下信息。

表 3-2 创建数据库

参数	参数说明
库名称	填写待创建数据库名称。 只能包含中文、字母、数字、下划线，长度为1~128个字符。

参数	参数说明
所属Catalog	待创建数据库所属Catalog。
选择位置	数据库信息存储在OBS并行文件系统中的位置。 单击“+”，选择位置后，单击“确定”。 – 所选位置只能以“obs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1/database1”。如果没有合适的并行文件系统，可以单击“前往OBS创建”进行创建。 – 必须与所属的Catalog存储路径（即创建Catalog时配置的“选择位置”参数）不同。 – 如果所属Catalog配置了“数据库存储位置”参数，则此处该参数必须选择为所属Catalog“选择位置”的子路径、或“数据库存储位置”的子路径。
描述	所创建数据库的描述信息。 长度为0~4000字节，1个中文字符对应3个字节。

2. （可选）单击“数据表存储位置”区域中的“添加数据表存储位置”。单击“+”按照需求手动选择数据表存储位置，单击“确定”。支持添加多条。

   说明

   – “数据表存储位置”为可选参数。
   – “数据表存储位置”可选择为所属Catalog路径及其子路径、或“数据库存储位置”路径及其子路径。
   – 如果配置了该参数，则该数据库下的数据表位置必须是该数据库“数据表存储位置”的子路径、或数据库“选择位置”的子路径。

3. （可选）单击“函数存储位置”区域中的“添加函数存储位置”。单击“+”按照需求手动选择函数存储位置，单击“确定”。支持添加多条。

   说明

   – “函数存储位置”为可选参数。
   – “函数存储位置”可选择为所属Catalog路径及其子路径、或“数据库存储位置”路径及其子路径。
   – 如果配置了该参数，则该数据库下的函数位置必须选择为该数据库“函数存储位置”或数据库“选择位置”的子路径。

4. 单击“提交”。

步骤6 创建完成后，即可在“数据库”页面查看库名称/ID、所属Catalog、数据库拥有者、存储位置等信息。

单击“操作”列的“编辑”可以修改数据库配置信息。

单击“操作”列的“数据表”，可以查看当前数据库下的数据表信息。

在“更多”中可以为当前数据库进行授权、查看权限等操作。

步骤7 如果需要删除数据库，可以选择“更多 > 删除”，确认操作影响，并根据界面提示确认是否删除其他数据后，单击“确定”。

- 同时删除数据库下的表：如果当前数据库下存在未删除的数据表或函数，则必须勾选此选项，否则会报错。**删除后的数据无法恢复，请谨慎操作！**
- 同时删除存储在OBS的数据：可选配置，删除后数据将会放入回收站目录下，可以在过期删除前恢复。

说明

删除元数据时如果同步删除文件，数据将移入对应OBS桶的回收站（“lake-formation-trash-dir/table_id”OBS路径）目录下。

----结束

3.4 管理数据表

用户在数据目录（Catalog）中可对元数据库和元数据表进行管理，按照业务规划创建对应数据表。

前提条件

- 已创建LakeFormation实例，且实例处于正常运行状态。
- 已创建待创建数据表的数据库及其所属Catalog。
- 已参考**创建元数据存储路径**提前创建了用于存储数据表的OBS并行文件系统。

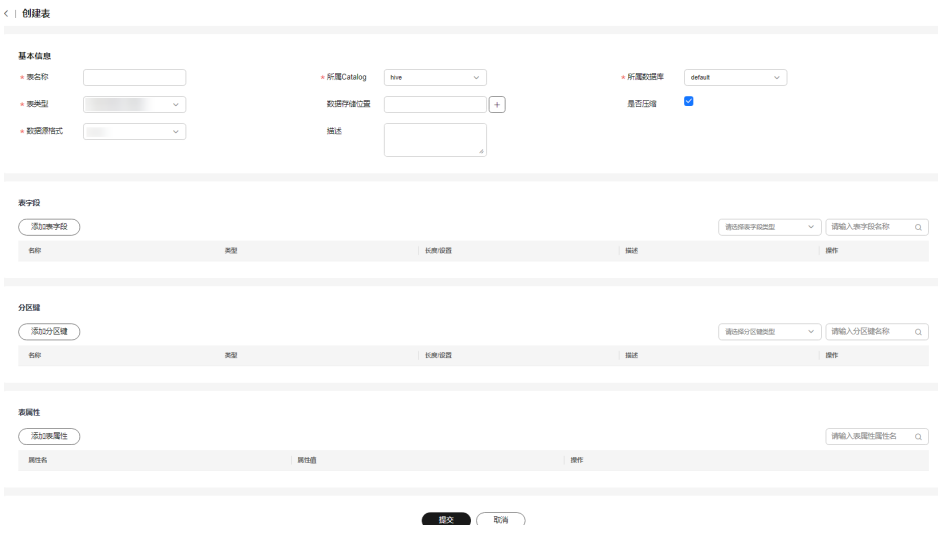
管理数据表

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，选择“元数据 > 表”，在右上角“Catalog”和“数据库”后的下拉框中分别选择待创建表的Catalog、数据库的名称。可以查看当前数据库中包含的数据表。

步骤4 单击“创建表”，配置相关参数。



1. 在“基本信息”区域配置以下信息。

表 3-3 基本信息配置参数

参数	参数说明
表名称	填写待创建的元数据表名称。 只能包含中文、字母、数字、下划线，长度为1~256个字符。
所属Catalog	待创建表所属的Catalog。
所属数据库	待创建表所属的数据库。
表类型	待创建表的类型，目前支持以下类型： – MANAGED_TABLE：管理表。如果删除管理表或分区，则与该表或分区关联的数据和元数据将删除。 – EXTERNAL_TABLE：外部表。当文件已经存在或位于远程位置时，使用外部表。 – VIRTUAL_VIEW：虚拟视图。不存储实际的数据，不占用物理空间。 – MATERIALIZED_VIEW：物化视图。存储实际的数据，占用物理空间。
数据存储位置	表所映射的OBS并行文件系统的文件目录。 单击“+”，选择表存储在OBS并行文件系统中的位置，单击“确定”。 – 可选参数，如果不配置，则数据表存储路径为“*上层数据库存储路径/表名*” – 如果配置该参数，则所选位置只能以“obs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1/database1/table1”。如果没有合适的并行文件系统，可以单击“前往OBS创建”进行创建。 – 必须与所属的Catalog、数据库的存储路径不同。 – 如果所属数据库配置了“数据表存储位置”参数，则此处存储位置必须选择为所属数据库“选择位置”的子路径、或“数据表存储位置”的子路径。
是否压缩	数据表是否压缩。 压缩表能够使表中的数据以压缩格式存储，表压缩能提升性能，减少存储空间。

参数	参数说明
数据源格式	待创建表的数据源格式，目前支持以下类型： – Avro – Json – Xml – Parquet – Csv – Orc – Text – Rc – Sequence – 自定义 如果选择为“自定义”需要根据实际需求配置“输入格式”、“输出格式”、“Serde name”、“SerializationLib”参数。
分隔符	数据源格式为“Csv”时，需设置字段分隔符。目前支持以下类型： – 逗号（,） – 竖线（|） – 分号（;） – Tab（\u0009） – Ctrl-A（\u0001）
描述	所创建表的描述信息。 长度为0~4000字节，1个中文字符对应3个字节。

2. （可选）单击“表字段”区域中的“添加表字段”。按照需求手动添加元数据的表字段，单击“确定”。支持添加多条。

   表字段：表字段是表中组成记录的一条条独立的信息。

3. （可选）单击“分区键”区域中的“添加分区键”。按照需求手动添加元数据的分区键，单击“确定”。支持添加多条。

   分区键：分区键是一个或多个表列的有序集合。表分区键列中的值用来确定每个表行所属的数据分区。

4. （可选）单击“表属性”区域中的“添加表属性”。按照需求添加元数据的表属性，单击“确定”。支持添加多条。

   表属性：使您能够使用自己的元数据键/值对来标记表定义。

5. 单击“提交”。

步骤5 创建完成后，即可在数据表页面查看表名称/ID、所属Catalog、所属数据库、类型、存储位置等信息。

- 单击“操作”列的“编辑”可以修改数据表配置信息。
- 在“更多”中可以为当前数据表进行授权、查看权限等操作。
- 单击数据表名称，可以查看当前数据表的详细元数据信息。

- 其中格式与序列化信息包含存储格式、输入格式、输出格式等。
- 其中字段信息包含表字段名称、类型、描述，以及分区键的字段名称、类型、描述。
- 其中表属性信息包含Table的各个属性的属性名、属性值。

单击“编辑”按钮，可对数据表相关字段进行修改。

步骤6 如果需要删除数据表，可以选择“更多 > 删除”，确认操作影响，并确认是否“同时删除存储在OBS的数据”，单击“确定”。

说明

删除元数据时如果同步删除文件，数据将移入对应OBS桶的回收站（“lake-formation-trash-dir/table_id”OBS路径）目录下。

----结束

3.5 管理函数

用户在数据目录（Catalog）中可对元数据进行管理，按照业务规划创建对应函数。

前提条件

- 已创建LakeFormation实例，且实例处于正常运行状态。
- 已创建待添加函数的数据库及其所属Catalog。
- 如果配置“函数位置”参数，需已参考**创建元数据存储路径**提前创建了用于存储函数的OBS并行文件系统。

创建函数

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，选择“元数据 > 函数”。在右上角“Catalog”和“数据库”后的下拉框中分别选择待创建函数的Catalog、数据库的名称。可以查看当前数据库中包含的函数。

步骤4 单击“创建函数”，配置相关参数。

1. 在“基本信息”区域配置以下信息。

表 3-4 基本信息配置参数

参数	参数说明
函数名称	填写待创建的元数据函数名称。 只能包含字母、数字、下划线，长度为1~256个字符。
所属Catalog	待创建函数的所属Catalog。
所属数据库	待创建函数的所属数据库。
类型	待创建函数的类型，目前支持“JAVA”类型。
函数类名	填写函数类名。

2. （可选）单击“函数位置”区域中的“添加函数位置”，按照需求手动添加函数包类型和函数位置，单击“确定”。支持添加多条。

   说明

   - “函数位置”为可选参数。
   - 如果函数所属数据库配置了“函数存储位置”参数，则此处存储位置必须选择为所属数据库“选择位置”的路径及其子路径、或“函数存储位置”的路径及其子路径。

3. 单击“提交”。

步骤5 创建完成后，即可在“函数”页面查看函数名称/ID、所属Catalog、所属数据库、类型、函数类名等信息。

单击“操作”列的“编辑”可以修改函数配置信息。

单击“操作”列的“授权”，可以为当前函数进行授权操作。

在“更多”中可以为当前函数进行查看权限等操作。

步骤6 如果需要删除函数，可以选择“更多 > 删除”，确认操作影响后，单击“确定”。

说明

删除元数据时如果同步删除文件，数据将移入对应OBS桶的回收站（“lake-formation-trash-dir/table_id”OBS路径）目录下。

----结束

4 管理数据权限

4.1 数据权限概述

数据湖权限支持配置数据库、数据表、函数等维度的权限。

云服务管理员可针对不同的管理对象配置不同用户组的权限，统一对数据湖资源进行管理。

说明

用户可通过LakeFormation管理控制台对数据湖内的资源进行统一权限管理，对于IAM用户/用户组，也可以通过关联LakeFormation服务的细粒度权限策略进行特性权限场景的授权，参见**创建LakeFormation自定义IAM策略**。当湖内数据资源较多时，建议通过LakeFormation管理控制台对数据湖内的资源进行统一权限管理。

LakeFormation配置权限时需要包含如下要素：

表 4-1 LakeFormation 权限要素

权限要素	描述
授权主体	被授予权限的对象，使其具备针对某数据资源的指定访问权限的用户组、角色、IAM用户、委托用户等身份，如某一用户组、某一角色等。 授权主体（用户组、角色、IAM用户、委托用户）名称中不能包含中划线（-），否则可能造成操作失败。
授权类型	• 数据湖中管理的资源： – 数据目录（Catalog） – 数据库（Database） – 数据表（Table） – 数据列（Column） – 函数（Function） • OBS路径。
操作类型	主体对授权类型的访问权限，不同授权类型支持的操作类型各不相同，可参见**表4-2**。

权限要素	描述
赋予授权权限	是否赋予授权权限，赋予授权权限后，授权主体便可以将拥有将权限授权给其他授权主体。

表 4-2 不同授权类型的操作权限

授权类型	操作类型	权限说明
Catalog	ALL	Catalog的所有操作权限。
	ALTER	修改Catalog。
	CREATE_DATABASE	创建数据库。
	DROP	删除Catalog。
	DESCRIBE	查看Catalog的元数据信息或切换Catalog。
	LIST_DATABASE	查看Catalog下资源列表。
数据库	ALL	数据库的所有操作权限。
	ALTER	修改数据库。
	DROP	删除数据库。
	DESCRIBE	查看数据库的元数据信息或切换数据库。
	LIST_TABLE	查看数据库下资源列表。
	LIST_FUNC	查看某一数据库下的函数。
	CREATE_TABLE	在数据库中创建表。
	CREATE_FUNC	在数据库中创建函数。
表	ALL	表的所有操作权限。
	ALTER	修改表。
	DROP	删除表。
	DESCRIBE	查看表的元数据信息。
	UPDATE	更新表数据。
	INSERT	插入表数据。
	SELECT	查询表内数据。
	DELETE	删除表的数据。
列	SELECT	查询表内的列数据。
函数	ALL	函数的所有操作权限。

授权类型	操作类型	权限说明
	ALTER	修改函数。
	DROP	删除函数。
	DESCRIBE	查看函数的元数据信息。
	EXEC	执行函数。
路径	READ	路径下文件的读权限。
	WRITE	路径下文件的写权限。

说明

权限管理员通常分为系统权限管理员与业务权限管理员，需要具备的IAM权限与权限管理范围不同

- 系统权限管理员
  - 需要拥有以下IAM操作权限：lakeformation:policy:describe、lakeformation:policy:create、lakeformation:policy:drop。
  - 权限管理范围：可将任意元数据权限授予给其他授权主体，可撤销任意元数据权限。
- 业务权限管理员
  - 需要拥有以下IAM操作权限： lakeformation:policy:describe、lakeformation:policy:delegate。
  - 权限管理范围：需要被业务管理员赋予授权权限，才能将元数据权限授予其他授权主体，或撤销元数据权限。

例如：系统中存在系统权限管理员User A、业务权限管理员User B、普通用户User C。User A将Catalog1的ALL权限授予User B并同时赋予授权权限后，User B可将Catalog1的DESC等权限授予User C。但User B无法将其他Catalog的权限授予User C。

4.2 新增授权

用户可通过LakeFormation管理控制台对数据湖内的资源进行统一权限管理，针对不同的授权主体进行授权。

在进行授权前，需确认待授权主体已存在，例如IAM用户组已提前创建。

说明

用户可通过LakeFormation管理控制台对数据湖内的资源进行统一权限管理，对于IAM用户/用户组，也可以通过关联LakeFormation服务的细粒度权限策略进行特性权限场景的授权，参见**创建LakeFormation自定义IAM策略**。当湖内数据资源较多时，建议通过LakeFormation管理控制台对数据湖内的资源进行统一权限管理。

统一添加授权策略

步骤1 登录管理控制台。

步骤2 在左上角单击“ ☰ ”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，选择“数据权限 > 数据授权”。

步骤4 单击“授权”，在弹出的窗口中参考下表配置参数后，单击“确定”。

说明

- 主体类型（用户组、角色、IAM用户、委托用户）名称中不能包含中划线（-），否则可能造成操作失败。
- 对单表授权权限策略数量不超过256个。

表 4-3 元数据权限授权

参数	参数说明
主体类型	• 用户组：选择待授权的用户组，例如IAM用户组，可提前到IAM服务管理控制台创建。 • 角色：选择待授权的角色。可提前参考**创建角色并绑定用户**章节创建角色。 • IAM用户：选择待授权的IAM用户。 • 委托用户：选择待授权的委托用户。
授权类型	• 资源：表示对LakeFormation实例中的资源进行授权。 • 路径：表示对OBS文件系统中的路径进行授权。该授权类型用于给外表或函数授权。
资源类型	选择待授权资源类型。“授权类型”选择“资源”时配置该参数。 并需要根据实际需求选择待授权的“Catalog”、“数据库”、“表”、“列”、“函数”。 **说明** 为表授予“SELECT”权限时，需要同时选择列，例如设置“列”为“*”。
行过滤条件	为权限策略设置行过滤条件。资源类型为“表”、“列”时显示该参数。 • 设置格式为：*列名 操作符 列值* 支持使用=、<=、<、>、>=、like等格式。 例如行过滤条件设置为：department = "financial"，表示选择表中“department”列中值为“financial”的行内容。 • 设置行过滤条件后，操作类型仅可以选择“SELECT”。

参数	参数说明
列脱敏类型	选择列脱敏类型，对列数据进行脱敏处理。资源类型为“列”时显示该参数。 ● PARTIAL_MASK：遮掩部分数据。 ● REDACT：修订原有列的数据。 ● HASH：使用哈希算法加密。 ● NULLIFY：用NULL值替换原值。 ● UNMASKED：原样显示。 ● DATA_ONLY_SHOW_YEAR：仅显示日期字符串的年份部分。 ● CUSTOM：用户自定义脱敏规则。 **说明** ● 设置列脱敏规则后，操作类型仅可以选择“SELECT”。 ● LakeFormation仅提供动态脱敏策略管理与获取功能。LakeFormation不处理动态脱敏策略冲突。
列脱敏参数	列脱敏类型对应的选项。可以参考表4-4进行配置。资源类型为“列”时显示该参数。
路径	“授权类型”选择“路径”时配置该参数。 单击 + 选择授权的OBS文件系统中的路径。可以选择多个路径，但不能超过10个。
操作类型	选择授权的操作类型。不同的授权类型对应的操作类型不同，详见表4-2。
赋予授权权限	是否授予授权权限。 赋予授权权限后，授权主体便拥有将对象授权其他授权主体的权限。

表 4-4 列脱敏参数说明

列脱敏类型	列脱敏参数（示例）	脱敏描述
PARTIAL_MASK	show last 4	只显示最后四个字符数据，例如“*******1234”。
	show first 4	只显示最前四个字符数据，例如“1234*******”。
REDACT	#	用#修订原来列的数据，例如“###########”。
	*	用*修订原来列的数据，例如“**********”。
HASH	hash256	使用hash256算法加密列中的数据。

列脱敏类型	列脱敏参数（示例）	脱敏描述
NULLIFY	无	用NULL值替换原值。
UNMASKED	无	原样显示。
DATA_ONLY_SHOW_YEAR	无	仅显示日期字符串的年份部分。
CUSTOM	自定义	用户自定义脱敏规则。

步骤5 如果需要取消授权，单击“操作”列中的“取消授权”，单击“确认”。

须知

取消授权后则无法恢复，请谨慎操作。

如果需要修改行过滤条件，单击“操作”列中的“更多”按钮，单击“编辑行过滤条件”。仅配置了行过滤条件时支持操作。

如果需要修改列脱敏配置，单击“操作”列中的“更多”按钮，单击“编辑列脱敏参数”。仅配置了列脱敏时支持操作。

----结束

为指定资源添加授权

用户可基于数据湖资源视角，为指定的某个资源（数据库、表）添加授权。

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例。

步骤4 进入指定资源的授权界面。

- Catalog：在左侧导航栏选择“元数据 > Catalog”，选择待授权Catalog“操作”列的“更多 > 授权”。
- 数据库：在左侧导航栏选择“元数据 > 数据库”，在右上角“Catalog”后的下拉框中选择待授权数据库所属的Catalog名称，选择待授权数据库“操作”列的“更多 > 授权”。
- 数据表：在左侧导航栏选择“元数据 > 表”，在右上角“Catalog”和“数据库”后的下拉框中分别选择待授权数据表所属的Catalog、数据库的名称，单击待授权数据表“操作”列的“授权”。
- 函数：在左侧导航栏选择“元数据 > 函数”，在右上角“Catalog”和“数据库”后的下拉框中分别选择待授权函数所属的Catalog、数据库的名称，单击待操作函数“操作”列的“授权”。

步骤5 参考表4-3配置相关信息后，单击“确定”。

步骤6 在“数据权限 > 数据授权”界面即可查看已授权的相关信息。

授权完成后，所选用户组中的用户、或拥有所选角色的用户或者用户组即可对当前数据库进行相关操作。

----结束

4.3 取消授权

本章节主要为您说明如何取消已有授权权限。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“  ”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，选择“数据权限 > 数据授权”。

步骤4 搜索想要取消的授权策略，单击要取消的授权信息后的“取消授权”按钮。

步骤5 单击“确认”，完成取消授权操作。

须知

取消授权后则无法恢复，请谨慎操作。

----结束

4.4 查询授权

本章节主要为您说明如何进行查询已有的数据授权信息。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“  ”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，选择“数据权限 > 数据授权”。

可以在右上方“OBS授权路径”、“授权主体”、“主体类型”、“主体来源”中搜索待查看权限的对应信息。

步骤4 在列表中查看数据授权信息。

相关字段说明如下：

表 4-5 数据授权参数

参数	说明
权限策略类型	包含以下类型： • DEFAULT：默认权限策略。 • ROW_FILTER：行过滤权限策略，包含行过滤条件。
授权主体	被授权的主体名称。
主体类型	被授权的主体类型。“GROUP”为用户组，“ROLE”为角色、“USER”为用户。
主体来源	被授权的主体来源。“IAM”表示来自IAM（用户、用户组），“LOCAL”表示来自LakeFormation，“AGENTTENANT”表示来自IAM委托。
授权对象	授权的资源名称或路径。 如果授权类型为资源，则格式为*Catalog*.[*Database*].[*Table*]。
资源类型	包括以下类型： • CATALOG：数据目录 • DATABASE：数据库 • TABLE：数据表 • COLUMN：数据列 • FUNC：函数 • URI：路径
权限	授权的权限名称，关于权限描述可参考**表4-2**。
授权权限	所授权的权限。

----结束

4.5 管理角色

某个角色拥有资源（比如数据库）的某些权限，则拥有这个角色的用户或者用户组也拥有了对应的资源操作权限。

说明

如果与LakeFormation实例对接的服务需要使用角色授权，则在创建对接LakeFormation权限的委托时必须包含角色的相关权限。

例如，LakeFormation与MRS集群对接后，需要使用角色的查询权限，则在**准备工作**创建LakeFormation委托时需要勾选“lakeformation:role:describe”。

创建角色并绑定用户

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，选择“数据权限 > 角色”。

步骤4 单击“创建角色”，在弹出的窗口中填写“角色名称”和“描述”后，单击“确定”。

步骤5 在“角色”页面，选择“操作”列的“添加IAM用户”或“添加委托用户”，选择绑定的角色名称勾选需要绑定的用户，单击“确定”。

说明

- 也可以在左侧导航栏选择“数据权限 > 用户”，在待绑定角色的用户所在行“操作”列单击“加入角色”，选择绑定的角色名称，单击“确定”。
- 为该角色授权后，绑定的用户将同时拥有对应的权限。
- 您也可以在LakeFormation与MRS集群对接后，在Ranger WebUI界面为MRS集群内的用户或用户组绑定该角色，具体操作请参考**通过Ranger为MRS集群内用户绑定LakeFormation角色**。

步骤6 如果需要为已创建的角色授权，可参考**新增授权**章节进行操作。

----结束

5 管理数据迁移

5.1 任务授权

操作场景

LakeFormation支持将外部服务的元数据及其权限全量或增量迁移至当前LakeFormation实例中，对元数据及权限进行统一管理。

进行任务管理操作前，需要为当前用户委托访问LakeFormation的相关权限，用于元数据、权限迁移时写入相关数据。

前提条件

已提前参考**创建用户并授权使用LakeFormation**章节创建用户，并加入admin用户组。

操作步骤

步骤1 使用加入admin用户组的用户登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“任务管理 > 任务授权”。

步骤4 单击“同意授权”，对当前用户授予LakeFormation任务管理权限。

如果需要取消用户的LakeFormation任务管理权限，请单击“取消授权”。

说明

同意授权后，LakeFormation将自动创建名为lakeformation_job_trust的委托，在任务运行期间，请勿删除该委托。

----结束

5.2 元数据迁移

操作场景

用户可以参考该章节将外部的元数据迁移至LakeFormation并将数据存储在OBS中进行统一管理。

说明

在迁移hive元数据时，为避免迁移元数据时发生路径冲突，建议hive Catalog路径与default数据库路径保持一致。

前提条件

- 当前实例已创建存储迁移元数据的Catalog。
- 待操作用户具有OBS相关操作权限、具有已创建存储迁移元数据的Catalog的操作权限。
- 已创建了用于存储迁移数据的OBS并行文件系统。
- 表的Owner只能包含字母、数字和下划线(_)，且长度为1~49个字符。不能包含中划线(-)等其他字符。
- 如果需要迁移多个MRS集群中的元数据到同一个LakeFormation实例，MRS集群之间的Database名称不能重复。
- 如果需要进行多次迁移，表的列更新需要满足列排序和列类型一致的兼容性要求。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“ ”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“任务管理 > 元数据迁移”。

步骤4 单击“创建迁移任务”，配置相关参数后，单击“提交”。

表 5-1 创建元数据迁移任务

参数	参数说明
任务名称	填写待创建元数据迁移任务名称。
描述	所创建迁移任务的描述信息。

参数	参数说明
数据源类型	选择待迁移数据的数据源类型。 ● DLF：第三方数据湖构建（Data Lake Formation，DLF） ● MRS云数据库RDS(for MySQL) ● 开源HiveMetastore(for MySQL) ● MRS云数据库RDS(for PostgreSQL) ● MRS本地数据库
JDBC URL	待迁移元数据JDBC链接的URL。“数据源类型”不为“DLF”时配置该参数。 **说明** 例如： ● MySQL数据源类型JDBC URL为：jdbc:mysql://*IP地址:端口/数据库名称*?useSSL=false&permitMysqlScheme ● PostgreSQL数据源类型JDBC URL为：jdbc:postgresql://*IP地址:端口/数据库名称*?socketTimeout=600 socketTimeout为迁移客户端和数据库连接的socket超时时长。 ● 迁移除MRS本地数据库以外的数据源类型时，URL中的IP为数据源绑定的弹性公网IP。 此外还需配置以下参数： ● 用户名：访问数据源所使用的用户。 ● 密码：访问数据源所使用的用户密码。 如果所使用用户存在密码，则必须填写；如果用户无密码，则为空即可。
服务接入点	配置待迁移元数据服务接入点。 “数据源类型”为“DLF”时配置该参数。此外还需配置以下参数： ● Access Key：联系DLF服务运维人员获取AK信息。 ● Secret Key：联系DLF服务运维人员获取SK信息。
源Catalog	待迁移元数据所属Catalog名称。
迁移至Catalog	元数据迁移至LakeFormation中Catalog的名称。
冲突解决策略	迁移过程中发生冲突的解决策略。 当前仅支持“创建并更新元数据”。
默认缺省Owner	迁移后元数据的默认Owner。“数据源类型”为“DLF”时配置该参数。 ● 如果配置的默认Owner没有对应的元数据操作权限，迁移后的元数据将无法进行增删改查等操作，此时可以手动给Owner授权或者进行权限迁移。 ● 如果迁移前所有元数据都能正常使用，则不需要配置该参数。

参数	参数说明
日志存储位置	运行迁移任务时，产生的日志存储位置。可单击+进行选择。 **该路径必须已在OBS中存在**，如果为自定义路径将会导致迁移任务失败。
是否强制建表	勾选此项将会跳过建立内表时对OBS路径的限制。
元数据过滤策略	迁移过程中元数据的过滤策略。 ● 按元数据类型 ● 按自定义规则
过滤策略存储位置	迁移的自定义元数据过滤策略文件在OBS并行文件系统中的存储位置。 “元数据过滤策略”选择“按自定义规则”时配置该参数。
过滤策略文件名	迁移的自定义元数据过滤策略文件名。 “元数据过滤策略”选择“按自定义规则”时配置该参数。
迁移元数据对象	勾选待迁移的元数据对象。“元数据过滤策略”选择“按元数据类型”时配置该参数。 ● 全选：迁移数据库、函数、数据表、分区 ● Database：数据库 ● Function：函数 ● Table：数据表 ● Partition：分区 **说明** – 如果为首次迁移，建议选择“全部”，迁移全部元数据。 – 如果仅勾选函数、数据表、分区的一种或几种时，需确保勾选元数据的上层目录存在。例如，只勾选了Table，则需要确保迁移目标Catalog中已包含该Table所在的数据库（例如DB_1）。否则会导致Table迁移成功的数量将为零个。 – 如果勾选函数，需要确认函数类名存在，否则会导致函数迁移失败。
添加location规则	● 如果迁移的数据来源中，元数据的存储路径前缀不为“obs://”，则需要单击“添加location规则”配置规则将前缀替换为“obs://”，并且确保存在对应的OBS存储路径。 例如，当前元数据的存储路径为“file:/a/b”，则“路径”填写“file:/”，“替换成”填写“obs://”，并确保OBS并行文件系统中存在“obs://a/b”路径，则生成元数据时新的元数据存储路径为“obs://a/b”。 ● 可以同时创建多条规则，当规则发生冲突时，以排在界面最上方的规则为准。

参数	参数说明
网络连接	选择网络连接方案。 推荐选择“EIP”，使用EIP方式连接网络。 同时需要选择“安全组ID”，即数据源所在VPC的安全组ID，用于打通网络。
事件通知策略 （当前该功能为公测阶段）	（可选）配置该选项后，发生特定事件（例如任务成功、任务失败等）后会发送通知（短信、邮件等）。 • 事件通知开关：开启后表示启用事件通知。 • 事件通知主题：选择需要通知的主题，可以在管理控制台选择“消息通知服务 SMN”进行配置。 • 事件：需要通知的主题状态，可选择“任务成功”、“任务失败”。

步骤5 创建完成后，单击“操作”列的“运行”即可运行当前迁移任务。

说明

- 在运行迁移任务前，需要已对用户进行任务授权，详情请参考**任务授权**。
- 迁移任务开始运行后，源数据库如果有新增的元数据，则新增的元数据将不会被迁移，需要再次运行迁移任务。也可以使用元数据发现功能，迁移新增的元数据，具体请参考**元数据发现**。
- 如果任务运行失败，在修复故障后可再次单击“操作”列的“运行”进行重试。

迁移任务完成后，可以在对应的元数据界面进行查看。例如进入“元数据 > 数据库”页面查看迁移完成的数据库。

单击操作列的“编辑”或“删除”，可以修改或者删除当前任务。

步骤6 单击“操作”列“查看日志”可以查看任务运行产生的日志。

默认显示最近的50行日志。

可单击日志最下方超链接查看完整日志，具体配置请参考**下载对象**章节。

日志中常见报错信息及对应原因如下：

日志报错信息	报错原因
field 'storageDescriptor.location' must match '^(obs|har)://.+/.+$'	请配置正确的location规则，确保元数据路径以“obs://”开头。
Invalid input parameter	元数据的入参非法，或者LakeFormation暂时不支持此类元数据。
Incorrect type of column *xxx*.	列类型非法，或者LakeFormation不兼容此列类型。
No permission to perform this operation on resources.	请检查默认缺省Owner是否配置正确，以及是否有元数据操作权限。

日志报错信息	报错原因
Error creating transactional connection factory	LakeFormation服务端与数据源连接不通。解决思路如下： 1. 数据源的用户名/密码或AK/SK是否正常。 2. JDBC URL中填写的数据库是否准确。 3. JDBC URL中填写的IP是否准确。如果数据源类型为MRS本地元数据，DBServer可能发生主备倒换，需要重新绑定弹性IP到主节点。 4. 数据库连接端口安全组是否放开。 – 如果选择EIP连接方式，任务运行前，数据源的安全组规则需要把0.0.0.0/0全部放开。 – 如果选择VPC对等连接方式，任务运行前，数据源的安全组规则需要把对等连接中对端访问IP放开。
输入的vpc网段与lakeformtion网段冲突	选择VPC对等连接时，数据源所在的VPC网段和LakeFormation服务端所在网段冲突。此时可以选择EIP方式进行迁移。
无日志内容	请确认日志路径是否存在。 ● 日志路径已存在，请联系LakeFormation运维人员协助处理。 ● 日志路径不存在，请修改任务配置中的日志路径，确保日志路径在OBS中存在。
The path should be a sub path of the catalog storage location or database location list	路径应为Catalog存储位置或者数据库存储位置列表的子路径。
Incorrect Partition Value	输入的分区值错误， 请检查输入的表的分区键列表与输入的分区值列表数量和类型是否匹配。
Database does not exist	数据库不存在，请检查数据库是否存在。
Location doesn't exist in the OBS Parallel File Systems	路径在OBS并行文件系统中不存在。
Folder obs://*xxxx/yyyy/* is not empty in the OBS	建表时OBS目录不能非空，迁移时需要勾选强制建表选项跳过该建表限制。

----结束

5.3 权限迁移

操作场景

在完成元数据迁移后，可以将对应元数据的权限迁移至LakeFormation，迁移成功后为元数据绑定的默认Owner将会拥有元数据的操作权限。

前提条件

- 已参考**元数据迁移**完成元数据迁移。
- 当前用户具有OBS相关操作权限，且已创建用于存储数据的OBS并行文件系统。
- 需将待迁移的权限策略文件导出，并上传至OBS并行文件系统中。权限导出操作可联系对应服务支持人员。
- 权限策略中授权主体（除角色外）需要提前创建，且名称需保持一致；权限策略中包含的元数据已存在，且名称一致。

  如果迁移类型为DLF，其对应关系及迁移策略如下：
  - RAM用户：IAM用户（如果对应的IAM用户不存在，该权限策略不进行迁移）
  - RAM角色：IAM用户组（如果对应的IAM用户组不存在，该权限策略不进行迁移）
  - DLF角色：LakeFormation角色（不存在会自动创建）
- 如果迁移类型为Ranger，则仅支持Ranger的allow权限迁移，不支持deny权限迁移。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“任务管理 > 权限迁移”。

步骤4 单击“创建迁移任务”，配置相关参数后，单击“提交”。

表 5-2 创建权限迁移任务

参数	参数说明
任务名称	填写待创建权限迁移任务名称。
描述	所创建迁移任务的描述信息。
权限策略类型	选择待迁移权限策略类型。 • DLF：第三方数据湖构建（Data Lake Formation，DLF）权限策略 • RANGER：MRS集群中Ranger权限策略

参数	参数说明
日志存储位置	运行迁移任务时，产生的日志存储位置。
权限策略文件存储位置	待迁移的权限策略文件在OBS并行文件系统中的存储位置。
权限策略文件名	待迁移权限策略的文件名称。
Catalog ID	填写权限来源的Catalog名称。 “权限策略类型”选择“DLF”时配置该参数。
授权主体转换关系	手动指定对应的权限策略的授权主体的转换关系，前、后缀值会为最终授权主体名称添加对应的前缀、后缀。 需要分别配置“用户转换对象”、“用户组转换对象”、“角色转换对象”，及其“前缀”和“后缀”。建议非IAM用户、非IAM用户组授权主体转换为角色。 “权限策略类型”为“DLF”时无需配置该参数。
事件通知策略 （当前该功能为公测阶段）	（可选）配置该选项后，发生特定事件（例如任务成功、任务失败等）后会发送通知（短信、邮件等）。 • 事件通知开关：开启后表示启用事件通知。 • 事件通知主题：选择需要通知的主题，可以在管理控制台选择“消息通知服务 SMN”进行配置。 • 事件：需要通知的主题状态，可选择“任务成功”、“任务失败”。

步骤5 创建完成后，单击“操作”列的“运行”即可运行当前迁移任务。

单击“查看日志”可以查看任务运行产生的日志。

说明

- 在运行迁移任务前，需要已对用户进行任务授权，详情请参考**任务授权**。
- 如果任务运行失败，在修复故障后可再次单击“操作”列的“运行”进行重试。

单击操作列的“编辑”或“删除”，可以修改或者删除当前任务。

迁移任务成功以后，可以在“数据权限 > 数据授权”页面查看成功迁移的LakeFormation权限策略。

----结束

5.4 元数据发现

操作场景

当数据存储在OBS并行文件系统中，而在LakeFormation还未与对应的元数据关联时，可以通过元数据发现，来构造这些数据对应的元数据信息，从而支撑SQL引擎或者用户的应用程序的计算与分析。

须知

当前元数据发现特性属于公测阶段，公测期产品完全免费，商业化后会根据元数据发现消耗资源收取资源费用。

元数据发现当前仅支持Spark on Hudi。

前提条件

- 已参考**任务授权**开启授权。
- 已上传待检测的数据至OBS并行文件系统，即已从S3或HDFS将数据上传复制到LakeFormation实例所在Region的OBS并行文件系统的规划路径下。
- 元数据发现的目标Catalog、目标Database已规划和创建。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“任务管理 > 元数据发现”。

步骤4 单击“创建发现任务”，配置相关参数后，单击“提交”。

表 5-3 创建发现任务

参数	参数说明
任务名称	填写待创建元数据发现任务名称。
描述	所创建元数据发现任务的描述信息。
数据存储位置	发现的数据表存储在OBS并行文件系统中的位置。 单击“+”，选择位置后，单击“确定”。

参数	参数说明
发现文件类型	选择所发现文件的类型。目前支持以下类型： ● 自动发现（包含Parquet、ORC、Json、Csv、Avro类型） ● Parquet ● ORC ● Json ● Csv（选择该类型，还需配置“分隔符”、“逃逸字符”、“引号字符”、“是否把第一行当做列名”等参数） ● Avro **说明** ● 如果数据存储位置下的文件后缀为同一类型，建议选择对应发现类型。 ● 如果数据存储位置下的文件后缀为多种类型，建议选择“自动发现”。 ● 如果数据存储位置下的文件不带后缀，建议选择对应类型。如果选择“自动发现”，则系统默认以Parquet类型文件进行发现，其他类型文件将会发现失败。
日志存储位置	运行元数据发现任务时，产生的日志存储位置。单击+选择路径。 该路径必须已在OBS中存在，如果为自定义路径将会导致发现任务失败。
目标Catalog	待发现元数据所属Catalog名称。
目标Database	待发现元数据所属数据库名称。
冲突解决策略	元数据发现过程中，存在同名元数据时的解决策略。 ● 创建并更新元数据 ● 仅创建元数据
默认缺省Owner	元数据发现任务执行后元数据的默认Owner。 如果选择的授权主体名称中带有中划线，此功能可能有失败风险。
文件采样率	（可选）文件采样频率。 采样率为0时，遇到空文件会跳过当前分区表之后的所有分区。该方法减少操作时间，但是准确性会降低。
重新发现策略	再次执行元数据发现时的发现策略。 ● 全量发现：再次执行发现操作时，发现数据存储位置下的所有文件。 ● 增量发现：再次执行发现操作时，发现上次任务（运行成功的）开始运行后，数据存储位置下新增的文件。

参数	参数说明
执行策略	选择当前迁移任务的执行策略。 ● 手动执行：手动触发执行迁移任务。 选择该方式后，需要在任务创建完成后，单击“操作”列的“运行”运行当前迁移任务。 ● 调度执行：周期性自动执行迁移任务。 选择该方式后，可根据实际需要选择调度执行的周期（“每月”、“每周”、“每日”、“每小时”）并配置对应参数。
主体类型	（可选）选择了主体后将默认为主体赋予数据存储位置的读权限。 ● 可选择为“用户组”、“角色”、“IAM用户”、“委托用户”，并选择具体授权的主体。 如果选择的授权主体名称中带有中划线，此功能可能有失败风险。 ● 如果需要对主体授予写权限，可勾选“赋予写权限”。
事件通知策略	（可选）配置该选项后，发生特定事件（例如任务成功、任务失败等）后会发送通知（短信、邮件等）。 ● 事件通知开关：开启后表示启用事件通知。 ● 事件通知主题：选择需要通知的主题，可以在管理控制台选择“消息通知服务 SMN”进行配置。 ● 事件：需要通知的主题状态，可选择“任务成功”、“任务失败”。

步骤5 创建完成后，单击“操作”列的“运行”即可运行当前迁移任务。

- 单击“停止”即可停止正在运行的任务。
- 单击“查看日志”可以查看任务运行产生的日志。

  默认显示最近的50行日志。

  可单击日志最下方超链接查看完整日志，具体配置请参考**下载对象**章节。
- 单击操作列的“编辑”或“删除”，可以修改或者删除当前任务。

步骤6 迁移任务运行完成后，可以进入“元数据 > 表”页面，在右上角“Catalog”和“数据库”后的下拉框中分别选择目标Catalog、目标Database的名称，查看已发现的数据表信息。

----结束

6 管理接入客户端

用户可以通过接入管理页面，简单快速地创建并管理接入客户端，可以在客户端详情中获取接入IP等信息，用于多种服务接入LakeFormation实例。

操作步骤

步骤1 登录管理控制台。

步骤2 在左上角单击“ ”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择待操作的LakeFormation实例，在左侧导航栏选择“接入管理”。

步骤4 单击“创建客户端”，在弹出的窗口中填写以下参数后，单击“确定”。

如果没有合适的虚拟私有云或子网，可以单击“前往VPC创建”。

表 6-1 创建接入管理客户端参数

参数	说明
客户端名称	自定义接入客户端的名称。
虚拟私有云	待接入服务所在的虚拟私有云。
所属子网	待接入服务所在的子网。

步骤5 单击“操作”列“查看详情”，打开详细信息页面。

“基本信息”区域可以查看ID、客户端名称、状态、接入模式、虚拟私有云、所属子网等信息。

“接入连接列表”区域可以查看终端节点ID、接入IP等信息。

步骤6 根据获取的接入IP等信息，完成其他云服务与LakeFormation的对接。

详细操作请参考对应云服务对接LakeFormation的操作指导。例如与MRS对接，则参考**配置LakeFormation数据连接**章节进行操作。

----结束

7 查看审计日志

云审计服务是安全解决方案中专业的日志审计服务，记录了LakeFormation的相关操作事件，方便您日后的查询、审计和回溯。

支持审计日志的操作

表 7-1 云审计服务支持的 LakeFormation 服务操作

操作名称	资源类型	事件名称
创建Catalog	Catalog	createCatalog
删除Catalog	Catalog	dropCatalog
修改Catalog	Catalog	alterCatalog
创建数据库	Database	createDatabase
删除数据库	Database	dropDatabase
修改数据库	Database	alterDatabase
创建数据表	Table	createTable
删除数据表	Table	dropTable
修改数据表	Table	alterTable
清空表的数据	Table	truncateTable
创建函数	Function	createFunction
修改函数属性	Function	alterFunction
删除函数	Function	dropFunction
创建实例	instance	createInstance
修改实例	instance	updateInstance
删除实例	instance	deleteInstance
授予权限	Access	grantAccess

操作名称	资源类型	事件名称
取消授权	Access	revokeAccess
更新表的指定列统计信息	TableColumnStatistic	setTableColumnStatistics
删除表的指定列统计信息	TableColumnStatistic	deleteTableColumnStatistics
批量创建表的列限制条件	TableConstraint	addConstraints
删除列限制条件	TableConstraint	deleteConstraints
批量添加分区信息	Partition	addPartitions
批量修改分区信息	Partition	alterPartitions
批量删除分区信息	Partition	dropPartitions
批量清空列表信息	Partition	truncatePartitions
批量设置分区的统计信息	PartitionColumnStatistic	setPartitionColumnStatistics
删除分区列的统计信息	PartitionColumnStatistic	deletePartitionColumnStatistics

查看审计日志

用户需要在云审计服务CTS的管理控制台查询LakeFormation服务的事件列表。

详情请参考：**查看审计日志（审计事件）**。