LakeFormation

User Guide

Issue	01
Date	2024-07-22

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions

HUAWEI and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Contents

1 Preparations

1.1 Registering a Huawei Cloud Account

Register a Huawei Cloud account to access all services on Huawei Cloud and pay only for the services you use.

If you already have a Huawei Cloud account, skip this part.

Procedure

Step 1 Visit the **Huawei Cloud website**.

Step 2 Click **Register** and complete the registration as instructed. For details, see **Registering a Huawei ID and Enabling Huawei Cloud Services**.

The system redirects you to the personal information page.

Step 3 Complete real-name authentication for individual or enterprise accounts. For details, see **Real-Name Authentication**.

----End

1.2 Configuring Cloud Service Authorization

Before using LakeFormation, you need to authorize it to access related cloud services.

Procedure

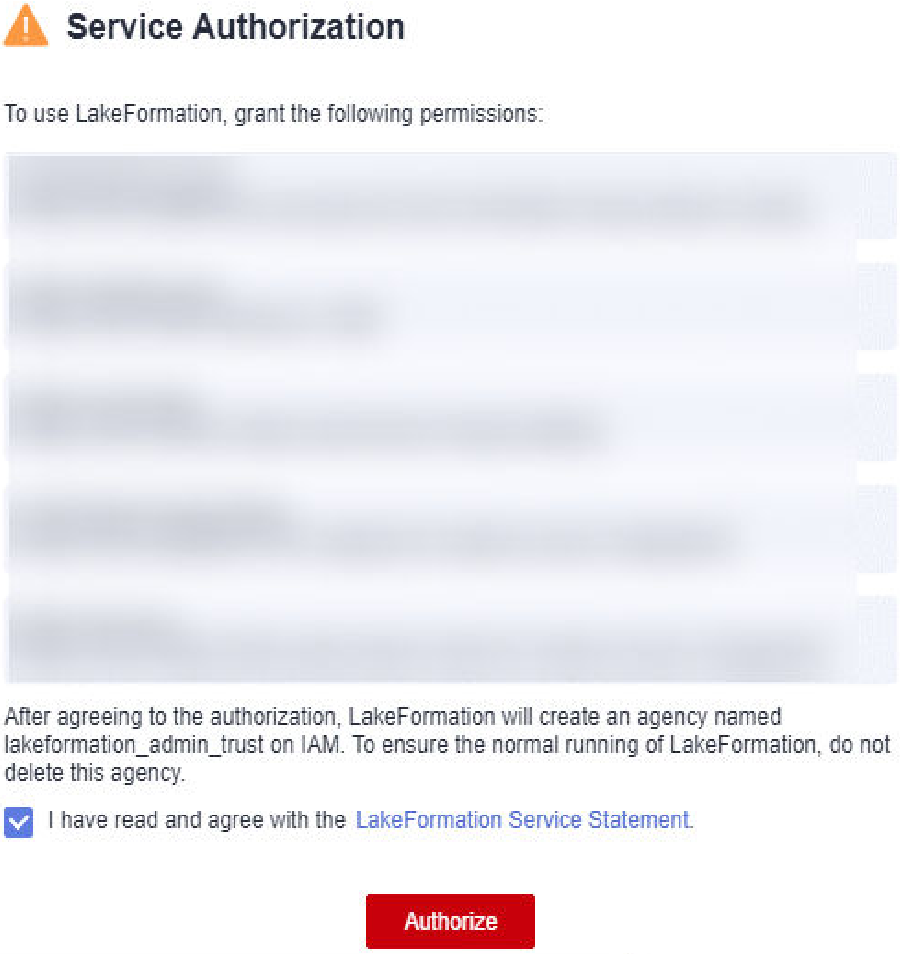
Step 1 Log in to the LakeFormation console.

Step 2 Click the service list and choose **Analytics** > **LakeFormation**. On the displayed page, click **Service Authorization** the left navigation pane.

- IAM ReadOnlyAccess: allows your instance to obtain user group and user information when running.
- OBS OperateAccess: provides the storage function for your instance.

- OBS AccessLabel: allows your instance to control permissions by using tagging.
- OBS Bucket Lifecycle: This permission is required to manage the lifecycle of instances.
- VPCEndpoint Administrator: allows you to operate on VPC endpoint for instance access management.
- DNS FullAccess: allows you to modify DNS private domain names for instance access management.

Figure 1-1 Authorization



Service Authorization

To use LakeFormation, grant the following permissions:

After agreeing to the authorization, LakeFormation will create an agency named lakeformation_admin_trust on IAM. To ensure the normal running of LakeFormation, do not delete this agency.

☑ I have read and agree with the LakeFormation Service Statement.

Step 3 Select "I have read and agree with the LakeFormation Service Statement." Click **Authorize**.

NOTE

After cloud service authorization, LakeFormation will create an agency named **lakeformation_admin_trust** in Identity and Access Management (IAM). Do not delete the agency when using LakeFormation.

If the agency fails to be automatically created, log in to the IAM console, delete the agency or contact the administrator to increase the quota, and check whether the current user has the permission to create an agency.

----End

1.3 Managing Permissions

1.3.1 Creating a User and Assigning Permissions

This chapter describes how to use **What Is IAM?** to implement fine-grained permissions control for your LakeFormation instances. With IAM, you can:

- Create IAM users for employees based on your enterprise's organizational structure. Each IAM user will have their own security credentials for accessing the resources.
- Assign users only the permissions required to perform a given task based on their job responsibilities.
- Entrust an account or a cloud service to perform professional and efficient O&M on your resources.

If your cloud account does not need individual IAM users, skip this section.

This section describes the procedure for granting permissions. See **Figure 1-2** for details.

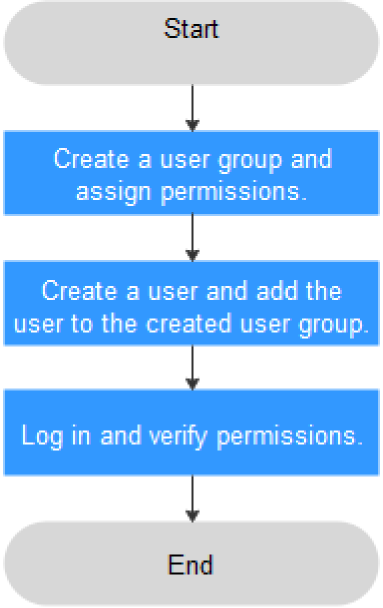
Prerequisites

Before assigning permissions to a user group, learn about LakeFormation permissions in **LakeFormation Permissions** and select them as needed.

If you need to assign permissions to other services, check all the policies supported by IAM in **System-defined Permissions**.

Procedure

Figure 1-2 Granting LakeFormation permissions


Start
Create a user group and assign permissions.
Create a user and add the user to the created user group.
Log in and verify permissions.
End

1. **Creating a User Group and Assigning Permissions**

   Create a user group on the IAM console, and assign LakeFormation permissions to the group.

2. **Create an IAM user**.

   Create a user on the IAM console and add the user to the group created in **1**.

3. **Log in** and verify permissions.

   Log in to the console as the user you created, and verify that the user has the assigned permissions.

   You can perform the following operation to verify a permission:

   Click the service list and choose LakeFormation. On the **Overview** page, click **Buy Instance** in the upper right corner. If the instance creation page is displayed, the **lakeformation:role:create permission** has already taken effect.

LakeFormation Permissions

By default, new IAM users lack permissions assigned. Add a user to one or more groups, and attach permissions policies or roles to these groups. Users inherit permissions from the groups to which they are added and can perform specified operations on cloud services.

You can grant permissions to a role or by creating a policy.

- Roles: A coarse-grained IAM authorization strategy to assign permissions based on user responsibilities. Only a limited number of service-level roles are available. Some roles depend other roles to take effect. When you assign such roles to users, remember to assign the roles they depend on. Roles are not an ideal choice for fine-grained authorization and secure access control.

  NOTE

  - Read-only permission authorization using IAM: To assign the read-only permission of LakeFormation in an IAM project to a sub-user, create a user group for this user as a tenant administrator and add the LakeFormation ReadOnlyAccess system-defined policy to this user group.
  - Enterprise project authorization: To assign all permission of LakeFormation in an enterprise project to a sub-user, create a user group for this user as a tenant administrator and add the LakeFormation CommonAccess permission to the user group and make this permission apply globally. Then, grant this permission to this enterprise project.

- Policies: A type of fine-grained authorization that defines permissions required to perform operations on specific cloud resources under certain conditions. This type of authorization is more flexible and ideal for secure access control. Most fine-grained policies split permissions by API. For details about how to customize IAM policies for LakeFormation, see **Creating a Custom Policy**.

Table 1-1 LakeFormation system permissions

Role/Policy Name	Description	Type	Dependency
LakeFormation FullAccess	Administrator permissions for LakeFormation. Users granted these permissions can use all LakeFormation functions.	System policy	N/A
LakeFormation ReadOnlyAccess	Read-only permissions for LakeFormation. Users granted these permissions can query LakeFormation data.	System policy	N/A
LakeFormation CommonAccess	Basic permissions for LakeFormation, including viewing, authorizing, and canceling the LakeFormation service agreement and basic permissions for dependent services such as OBS and TMS.	System policy	N/A

Table 1-2 LakeFormation fine-grained permissions

Operation Type	Item	Description
Read-only	lakeformation:instance:describe	Permission to query LakeFormation instances.
	lakeformation:catalog:describe	Permission to query the data catalogs of LakeFormation metadata.
	lakeformation:database:describe	Permission to query the databases of LakeFormation metadata.
	lakeformation:table:describe	Permission to query the data tables of LakeFormation metadata.
	lakeformation:function:describe	Permission to query the functions of LakeFormation metadata.
	lakeformation:policy:describe	Permission to query LakeFormation permission policies.
	lakeformation:policy:export	Permission to query LakeFormation permission policies in batches.

Operation Type	Item	Description
	lakeformation:agency:describe	Permission to query LakeFormation agencies.
	lakeformation:credential:describe	Permission to obtain the authentication information for accessing LakeFormation.
	lakeformation:group:describe	Permission to obtain the relationship between a LakeFormation user group and its associated roles.
	lakeformation:user:describe	Permission to obtain the relationship between a LakeFormation user and its associated roles.
	lakeformation:role:describe	Permission to query LakeFormation roles.
	lakeformation:configuration:describe	Permission to query user configurations.
	lakeformation:access:describe	Permission to query the client access permission.
	lakeformation:job:describe	Permission to query LakeFormation tasks.
Write	lakeformation:instance:create	Permission to create LakeFormation instances.
	lakeformation:role:create	Permission to create LakeFormation roles.
	lakeformation:policy:create	Permission to create LakeFormation permission policies.
	lakeformation:function:create	Permission to create the functions of LakeFormation metadata.
	lakeformation:catalog:create	Permission to create the data catalogs of LakeFormation metadata.
	lakeformation:database:create	Permission to create the databases of LakeFormation metadata.
	lakeformation:table:create	Permission to create the tables of LakeFormation metadata.

Operation Type	Item	Description
	lakeformation:access:create	Permission to create the client access permission.
	lakeformation:agency:create	Permission to create LakeFormation agencies.
	lakeformation:job:create	Permission to create LakeFormation tasks.
	lakeformation:instance:alter	Permission to modify LakeFormation instances.
	lakeformation:catalog:alter	Permission to modify the data catalogs of LakeFormation metadata.
	lakeformation:database:alter	Permission to modify the databases of LakeFormation metadata.
	lakeformation:table:alter	Permission to modify the tables of LakeFormation metadata.
	lakeformation:function:alter	Permission to modify the functions of LakeFormation metadata.
	lakeformation:role:alter	Permission to modify the relationship between a LakeFormation role and its associated user groups.
	lakeformation:group:alter	Permission to modify the relationship between a LakeFormation user group and its associated roles.
	lakeformation:user:alter	Permission to modify the relationship between a LakeFormation user and its associated roles.
	lakeformation:job:alter	Permission to modify LakeFormation tasks.
	lakeformation:instance:drop	Permission to delete LakeFormation instances.
	lakeformation:role:drop	Permission to delete LakeFormation roles.
	lakeformation:policy:drop	Permission to delete LakeFormation permission policies.

Operation Type	Item	Description
	lakeformation:function:drop	Permission to delete the functions of LakeFormation metadata.
	lakeformation:catalog:drop	Permission to delete the data catalogs of LakeFormation metadata.
	lakeformation:database:drop	Permission to delete the databases of LakeFormation metadata.
	lakeformation:table:drop	Permission to delete the tables of LakeFormation metadata.
	lakeformation:access:delete	Permission to delete the client access permission.
	lakeformation:agency:drop	Permission to delete LakeFormation agencies.
	lakeformation:job:drop	Permission to delete LakeFormation tasks.
	lakeformation:transaction:operate	Permission to operate LakeFormation transactions.
	lakeformation:instance:access	Permission to query a LakeFormation instance or apply for the access to it.
	lakeformation:job:exec	Permission to execute LakeFormation tasks.

1.3.2 Creating a Custom Policy

Custom policies can be created for LakeFormation to supplement system-defined policies.

To create a custom policy, choose either visual editor or JSON.

- Visual editor: Select cloud services, actions, resources, and request conditions. This does not require knowledge of policy syntax.
- JSON: Create a policy in the JSON format from scratch or based on an existing policy.

For details, see **Creating a Custom Policy**.

The following section contains examples of common LakeFormation custom policies.

Example Custom Policies

- Example 1: Grant the read-only permission on LakeFormation in batches to a user.

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

- Example 2: Deny data deletion.

  A policy with only "Deny" permissions must be used together with other policies. If the permissions granted to an IAM user contain both "Allow" and "Deny", the "Deny" permissions take precedence over the "Allow" permissions.

  The following method can be used if you need to assign permissions of the **Admin** policy to a user but you want to prevent the user from deleting LakeFormation catalogs, databases, and tables. Create a custom policy for denying deletion, and attach both policies to the group to which the user belongs. Then, the user can perform all operations on catalogs, databases, and tables except deleting them.

  Example policy denying MRS cluster deletion:

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

- Example 3: Create a custom policy that contains the actions of multiple services.

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

2 Instance Management

2.1 Creating an Instance

Before using LakeFormation, create an instance. Then, you can perform operations on the instance, such as managing metadata and setting metadata permissions.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Click **Buy Now** or **Buy Instance** in the upper right corner.

If a LakeFormation instance exists on the page, **Buy Instance** is displayed. Otherwise, **Buy Now** is displayed.

Step 4 Configure the parameters listed in the following table.

Table 2-1 Instance creation parameters

Parameter	Description
Type	Select the instance type. ● **Shared**: Resources are reused between shared instances to maximize the usage of resources such as clusters or GaussDB(for MySQL) instances. ● **Exclusive**: You are charged based on the upper limit of query per second (QPS) and metadata usage.
Region	Select a **region**. LakeFormation instances in different regions cannot communicate with each other over an intranet. For lower network latency and quick resource access, select the nearest region.

Parameter	Description
Billing Mode	Billing mode of instances. ● **Pay-per-use**: You are billed by the usage duration of LakeFormation instances.
Project	Select the project to which the instance belongs.
Name	User-defined LakeFormation instance name.
QPS	Maximum number of requests per second. You do not need to set this parameter when **Type** is set to **Shared**. You will be billed by LakeFormation based on your usage of metadata objects.
Enterprise Project	Enterprise project to which the cluster belongs. If no enterprise project is available, click **Create** to create one. An enterprise project facilitates project-level management and grouping of cloud resources and users.
Description	Description of the instance.
Label	Enter content in the boxes and click **Add** to add a tag. You are advised to click **View Predefined Tags** to use predefined tags and add the same tag to different cloud resources.

Step 5 Click **Buy Now**, confirm the configured information, and pay.

Step 6 Click **Back to Console**. You can view information about the newly created LakeFormation instance on the console.

NOTE

When creating an instance, pay attention to quota notification. If a resource quota is insufficient, increase the resource quota as prompted and create an instance.

Wait until the instance status changes to Running , indicating that the instance is created.

After an instance is created, you can view its basic information and data overview.

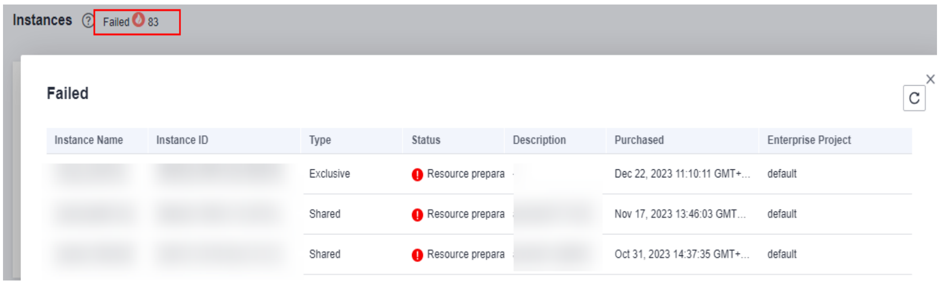
----End

Failed to Create an Instance

If an instance fails to be created, the **Failed** page is displayed.

Choose **Analytics** > **LakeFormation**. In the upper left corner of the **Overview** page, click **Failed** . In the displayed window, view information about the instance that fails to be created.

Figure 2-1 Failed instances



Instances Failed 83

Failed

Instance Name	Instance ID	Type	Status	Description	Purchased	Enterprise Project
[redacted]	[redacted]	Exclusive	● Resource prepara	-	Dec 22, 2023 11:10:11 GMT+...	default
[redacted]	[redacted]	Shared	● Resource prepara	[redacted]	Nov 17, 2023 13:46:03 GMT...	default
[redacted]	[redacted]	Shared	● Resource prepara	[redacted]	Oct 31, 2023 14:37:35 GMT+...	default

2.2 Configuring an Instance

After a LakeFormation instance is created, you can perform operations on the LakeFormation instance on the **Overview** page, such as changing instance specifications and setting an instance as default.

Modifying Instance Specifications: Change the QPS value of the instance. Only exclusive instances support this operation.

Setting an Instance as Default: Set the instance as the default one. If no instance ID is specified when other services interconnect with the LakeFormation instance, this operation will modify the instance interconnected with the service.

Modifying Instance Specifications

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left.

Step 4 Click **Scale Instance** in the upper right corner of the page. The **Scale Instance** page is displayed. Shared instances do not support this operation.

Step 5 Select the target QPS valueand click **Next**.

Step 6 Confirm the instance information and its modified specification, and click **Submit**.

----End

Setting an Instance as Default

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left.

Step 4 Check the value of **Default Instance** in the **Basic Information** area.

- **true** indicates that the instance is the default instance.
- **false** indicates that the current instance is not the default instance.

If the current instance is the default instance, the **Set as Default** button is unavailable.

Step 5 To set the instance as the default one, click **Set as Default** in the upper right corner of the page, confirm the operation impact, and click **OK**.

NOTICE

If other services need to connect to LakeFormation and no LakeFormation instance ID is specified, the system automatically accesses the default instance. This default instance change may affect the services that use LakeFormation. Exercise caution when performing this operation.

----End

2.3 Deleting an Instance

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left.

Step 4 Click **Delete Instance** in the upper right corner.

Step 5 In the displayed dialog box, confirm the deletion impact, and click **OK** to delete the instance.

NOTE

- Deleted instances are moved to the recycle bin and continue to be billed until they are deleted from the recycle bin.
- Instances that have been stored in the recycle bin for more than one day will be automatically deleted and cannot be restored.
- Instances can be forcibly deleted only 15 minutes after they are moved to the recycle bin to prevent service interruption.

----End

3 Metadata Management

3.1 Creating a Metadata Storage Path

The data files and directories mapped to LakeFormation metadata are stored in the OBS parallel file system. Before creating LakeFormation metadata, you need to create an OBS parallel file system for data storage.

If an OBS parallel file system is available, skip this section.

Procedure

Step 1 Create an OBS parallel file system. For details, see Creating a Parallel File System. For example, the file system name is **lakeformation-test**.

Step 2 On the **Parallel File Systems** page, click the name of the created file system, that is **lakeformation-test**.

Step 3 Click **Files** in the navigation pane, click **Create Folder**, enter a folder name, and click **OK**. Click the folder name and click **Create Folder** to create a subfolder.

Repeat this step to create paths for storing metadata in sequence. The following paths are examples:

- Catalog storage path: **lakeformation-test/catalog1**
- Database storage path: **lakeformation-test/catalog1/database1**
- Table storage path: **lakeformation-test/catalog1/database1/table1** and **lakeformation-test/catalog1/database1/table2**
- Function storage path: **lakeformation-test/catalog1/database1/udf1**

----End

3.2 Managing Catalogs

A data catalog is a metadata management object that can contain multiple databases.

Multiple catalogs can be created and managed in LakeFormation to isolate metadata of different external clusters.

Prerequisites

- A LakeFormation instance has been created and is running properly.
- Catalog data is stored in OBS and the permission to perform operations on OBS is obtained.
- You have created an OBS parallel file system for storing catalog data by referring to **Creating a Metadata Storage Path**.

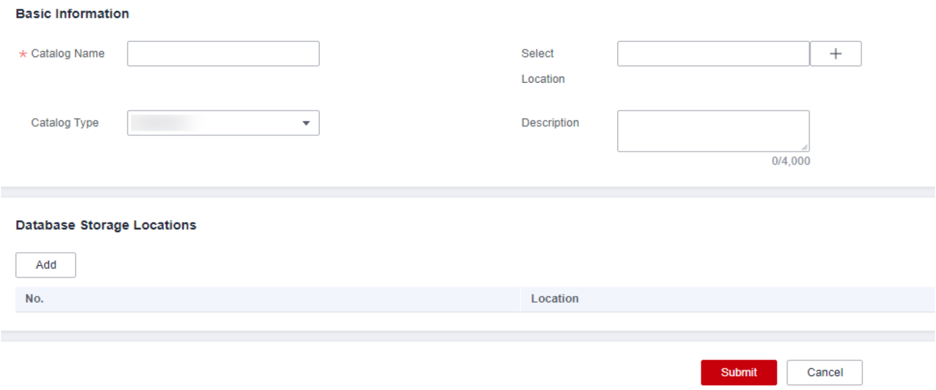
Creating a Catalog

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and choose **Metadata** > **Catalog** in the navigation pane.

Step 4 Click **Create** and set the parameters.



1. In the **Basic Information** area, set the related parameters.

Table 3-1 Parameters for creating a catalog

Parameter	Description
Catalog Name	Enter a catalog name. The value should contain 1 to 256 characters. Only letters, numbers, and underscores (_) are allowed.
Catalog Type	The options are as follows: – **DEFAULT** – **CLICKHOUSE**

Parameter	Description
Select Location	Select a location where catalog data is stored in the OBS parallel file system. Click +, select a location, and click **OK**. – The selected location must start with **obs://** and must contain one storage object. For example, select **obs://lakeformation-test/catalog1**. If no suitable parallel file system is available, click **go to OBS and create one**. – You are advised to select a folder that is not selected by other catalogs.
Description	Enter a description of the created catalog. The content length must be between 0 and 4000 bytes (3 bytes per Chinese character).

2. (Optional) Click **Add** under **Database Storage Locations**. On the **Add Database Storage Location** page, click + to manually select a database storage path as required. Click **OK**.

   Click ⊕ to add more storage paths. Click 🗑 to delete a storage path.

   NOTE

   Selecting database storage path is an optional operation. If a database storage path is added, the databases in this catalog must be stored in the subpath of the database storage path or that of the selected catalog location path.

3. Click **Submit**.

Step 5 After the catalog is created, you can view the catalog information on the **Catalog** page.

Click **Modify** in the **Operation** column to modify the configurations of a catalog.

Click **Database** in the **Operation** column to view the databases in a catalog.

Click **More** to authorize or delete a catalog, or view the permissions of a catalog.

NOTE

If files are also deleted when metadata is deleted, the metadata is moved to the recycle bin (OBS path **lake-formation-trash-dir/table_id**) of the corresponding OBS bucket.

----End

3.3 Managing Databases

Multiple databases can be created under a LakeFormation catalog. Centralized metadata management can maximize the value of data assets.

Prerequisites

- A LakeFormation instance has been created and is running properly.
- A catalog has been created.
- You have created an OBS parallel file system for storing databases by referring to **Creating a Metadata Storage Path**.

Create a Database

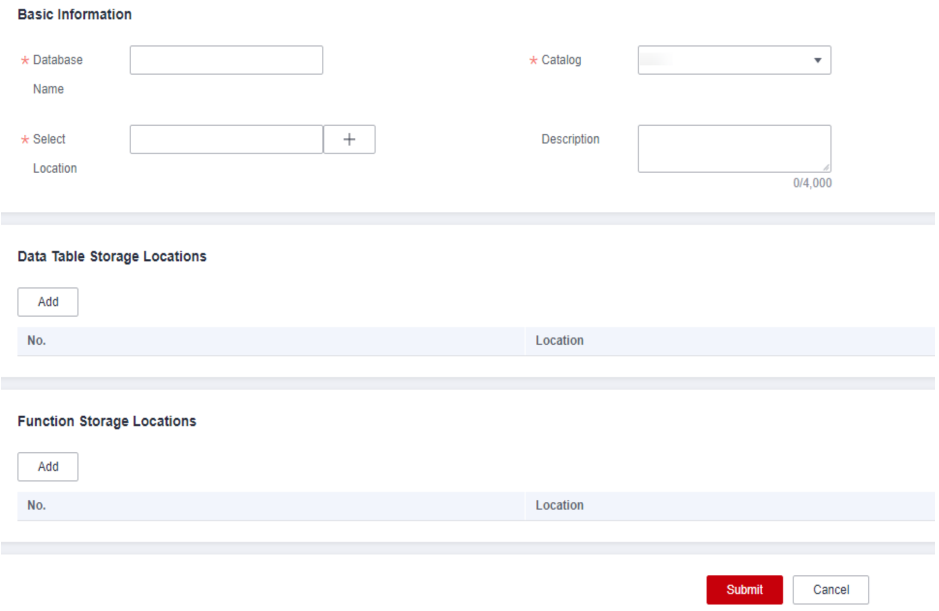
Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and choose **Metadata** > **Database** in the navigation pane.

Step 4 Select the name of the desired catalog from the **Catalog** drop-down list box in the upper right corner. You can view the databases contained in this catalog.

Step 5 Click **Create**.



1. In the **Basic Information** area, set the related parameters.

Table 3-2 Parameters for creating a database

Parameter	Description
Database Name	Enter a database name. The value should contain 1 to 128 characters. Only letters, numbers, and underscores (_) are allowed.
Catalog	Select the name of the catalog to which the database to be created belongs.

Parameter	Description
Select Location	Select a location where the database information is stored in the OBS parallel file system. Click +, select a location, and click **OK**. – The selected location must start with **obs://** and must contain one storage object. For example, select **obs://lakeformation-test/catalog1/database1**. If no suitable parallel file system is available, click **go to OBS and create one**. – The path must be different from the storage path of its catalog (the one specified by **Select Location** on the catalog creation page). – If a database storage path is added for the catalog to which the database belongs, set this parameter to the subpath of the database storage path or that of the path specified by **Select Location** during catalog creation.
Description	Description of the created database. The content length must be between 0 and 4000 bytes (3 bytes per Chinese character).

2. (Optional) Click **Add** under **Data Table Storage Locations**. On the **Add Table Storage Location** page, click + to manually select a database storage path as required. Click **OK**.

   Click ⊕ to add more storage paths. Click 🗑 to delete a storage path.

   NOTE

   – Selecting data table storage path is an optional operation.

   – The data table storage path can be set to the catalog path and its subpaths, or database storage location path and its subpaths.

   – If a data table storage path is added, the tables in this database must be the subpath of the data table storage path or that of the selected database location path.

3. (Optional) Click **Add** under **Function Storage Locations**. On the **Add Function Storage Location** page, click + to manually select a function storage path as required. Click **OK**.

   Click ⊕ to add more storage paths. Click 🗑 to delete a storage path.

   NOTE

   – Selecting function storage path is an optional operation.

   – The function storage path can be set to the catalog path and its subpaths, or database storage location path and its subpaths.

   – If a function storage path is added, the functions in this database must be the subpath of the function storage path or that of the selected database location path.

4. Click **Submit**.

Step 6 After the database is created, you can view the database name/ID, catalog, owner, and storage location on the **Database** page.

Click **Modify** in the **Operation** column to modify the configurations of a database.

Click **Table** in the **Operation** column to view the tables in a database.

Click **More** to authorize or delete a database, or view the permissions of a database.

NOTE

If files are also deleted when metadata is deleted, the metadata is moved to the recycle bin (OBS path **lake-formation-trash-dir/table_id**) of the corresponding OBS bucket.

----End

3.4 Managing Tables

You can manage metadata databases and metadata tables in the data catalog and create data tables based on the service plan.

Prerequisites

- A LakeFormation instance has been created and is running properly.
- A catalog and a database have been created.
- You have created an OBS parallel file system for storing tables by referring to **Creating a Metadata Storage Path**.

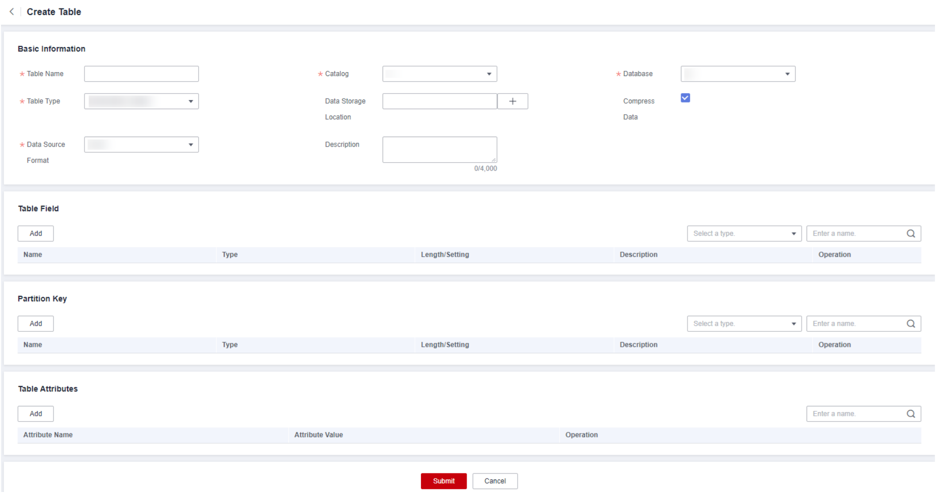
Creating a Table

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select a LakeFormation instance from the drop-down list box on the left, choose **Metadata** > **Table**, and select names of the target catalog and database from the **Catalog** and **Database** drop-down lists in the upper right corner. You can view the tables contained in the selected database.

Step 4 Click **Create** and set related parameters.



1. In the **Basic Information** area, set the related parameters.

Table 3-3 Parameters for creating a table

Parameter	Description
Table Name	Enter a metadata table name. The value should contain 1 to 256 characters. Only letters, numbers, and underscores (_) are allowed.
Catalog	Select the catalog to which the table to be created belongs.
Database	Select the database to which the table to be created belongs.
Table Type	Type of the table to be created. Currently, the following types are supported: – **MANAGED_TABLE**: management table. If a management table or partition is deleted, the data and metadata associated with the table or partition are deleted. – **EXTERNAL_TABLE**: external table. An external table is used when the file already exists or is stored in a remote location. – **VIRTUAL_VIEW**: virtual view. It does not store actual data or occupy physical space. – **MATERIALIZED_VIEW**: materialized view. It stores actual data and occupies physical space.

Parameter	Description
Data Storage Location	File directory of the OBS parallel file system to which the table is mapped. Click +, select a location for storing the table in the OBS parallel file system, and click **OK**. – The selected location must start with **obs://** and must contain one storage object. For example, select **obs://lakeformation-test/catalog1/database1/table1**. If no suitable parallel file system is available, click **go to OBS and create one**. – The path must be different from the storage path of the catalog and database to which the belongs. – If a table storage path is added for the database to which the table belongs, set this parameter to the subpath of the table storage path or that of the path specified by **Select Location** during database creation.
Compress Data	Whether to compress a data table. By using table compression, you can store data in a compressed format and improve performance and storage space.
Data Source Format	Data source format of the table to be created. Currently, the following types are supported: – **Avro** – **Json** – **Xml** – **Parquet** – **Csv** – **Orc** – **Text** – **Rc** – **Sequence** – **Custom** Parameters **Input Format**, **Output Format**, **Serde name**, and **SerializationLib** are displayed if **Data Source Format** is set to **Custom**. Set these parameters based on the site requirements.

Parameter	Description
Separator	This parameter is displayed if **Data Source Format** is set to **Csv**. The values include: – **Comma(,)** – **Vertical bar(|)** – **Semicolon(;)** – **Tab(\u0009)** – **Ctrl-A(\u0001)**
Description	Description of the created table. The content length must be between 0 and 4000 bytes (3 bytes per Chinese character).

2. (Optional) Click **Add** in the **Table Field** area. Manually add metadata table fields as required. Click **OK**.

   Click ⊕ to add more table fields. Click 🗑 to delete a table field.

   A table field is an independent piece of information that forms a record in a table.
3. (Optional) Click **Add** in the **Partition Key** area. Manually add the partition key of the metadata as required. Click **OK**.

   Click ⊕ to add more partition keys. Click 🗑 to delete a partition key.

   A partition key is an ordered set of one or more table columns. The values in the table partition keys are used to determine the data partition that a row belongs to.
4. (Optional) Click **Add** in the **Table Attributes** area. Add metadata table attributes as required. Click **OK**.

   Click ⊕ to add more table attributes. Click 🗑 to delete a table attribute.

   A table attribute enables you to tag table definitions with your own metadata key-value pairs.
5. Click **Submit**.

Step 5 After the table is created, you can view the table name/ID, catalog, database, type, and storage location on the **Table** page.

Click **Modify** in the **Operation** column to modify the configurations of a table.

Click **More** and then click **Authorize** in the **Operation** column to authorize a table.

Click **More** to authorize or delete a datable, or view the permissions of a table.

NOTE

If files are also deleted when metadata is deleted, the metadata is moved to the recycle bin (OBS path **lake-formation-trash-dir/table_id**) of the corresponding OBS bucket.

Step 6 Click a table name to view its detailed metadata information.

- The format and serialization information includes the storage format, input format, and output format.

- The field information includes the table field name, type, and description, and the field name, type, and description of the partition key.
- The table attribute information includes the name and value of each attribute in the table.

Locate the row that contains the target table and click **Edit** to modify the fields in it.

----End

3.5 Managing Functions

You can manage metadata in the data catalog and create functions based on the service plan.

Prerequisites

- A LakeFormation instance has been created and is running properly.
- A catalog and a database have been created.
- You have created an OBS parallel file system for storing functions by referring to **Creating a Metadata Storage Path** if a function location is added.

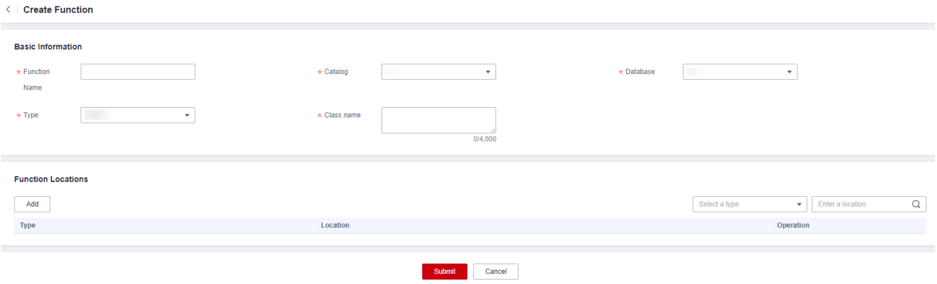
Creating a Function

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left and choose **Metadata** > **Function**. In the upper right corner, select the names of the target catalog and database from the **Catalog** and **Database** drop-down lists. You can view the functions contained in the selected database.

Step 4 Click **Create** and set related parameters.



1. In the **Basic Information** area, set the related parameters.

Table 3-4 Parameter for creating a function

Parameter	Description
Function Name	Enter a metadata function name. The value should contain 1 to 256 characters. Only letters, numbers, and underscores (_) are allowed.
Catalog	Catalog to which the function to be created belongs.
Database	Database to which the function to be created belongs.
Type	Type of the function to be created. Currently, **JAVA** is supported.
Class Name	Enter a function class name.

2. (Optional) Click **Add** under **Function Locations** to add the function package type and location as required. Click **OK**.

   Click ⊕ to add more function package types and locations. Click 🗑 to delete a function package type and location.

   NOTE

   - Selecting function package type and location is an optional operation.
   - If a function storage path is added for the database to which the function belongs, the storage location must be the path specified by **Select Location** during database creation or its subpaths, or the function storage path or its subpaths.

3. Click **Submit**.

Step 5 After the function is created, you can view the function name/ID, catalog, database, type, and class name on the **Function** page.

Click **Modify** in the **Operation** column to modify the configurations of a function.

Click **More** and then click **Authorize** in the **Operation** column to authorize a function.

Click **More** to authorize or delete a function, or view the permissions of a function.

NOTE

If files are also deleted when metadata is deleted, the metadata is moved to the recycle bin (OBS path **lake-formation-trash-dir/table_id**) of the corresponding OBS bucket.

----End

4 Data Permission Management

4.1 Data Permission Overview

Data lake permissions can be configured in three dimensions: database, table, and function.

Cloud service administrators can configure permissions of different user groups for different managed objects to centrally manage data lake resources.

NOTE

You can centrally manage permissions on resources in the data lake on the LakeFormation console. IAM users and user groups can also be associated with fine-grained permission policies of LakeFormation for authorization. For details, see **Creating a Custom Policy**. If there are a large number of data resources in the data lake, you are advised to use the LakeFormation console to centrally manage permissions on resources in the data lake.

The following table lists the main elements in LakeFormation permission configuration.

Table 4-1 Permission configuration elements

Element	Description
Authorization Entity	You can specify any user, user group, or role to be the authorization entity. The name of an authorization entity cannot contain hyphens (-). Otherwise, the operation may fail.
Granted To	● **Resources** – Data catalogs – Databases – Tables – Columns – Functions ● **Paths**

Element	Description
Operation Type	Access permission on the authorization object that the authorization entity has. Different authorization objects support different operations. For details, see **Table 4-2**.
Grant Authorization Permission	Whether to grant the authorization permission. After the authorization permission is granted, the authorization entity can grant the permission to other authorization entities.

Table 4-2 Operation rights of different authorization objects

Object	Operation Type	Description
Catalog	ALL	Perform all operations on catalogs.
	ALTER	Modify catalogs.
	CREATE_DATABASE	Create databases.
	DROP	Delete catalogs.
	DESCRIBE	Check the metadata of catalogs or switch catalogs.
	LIST_DATABASE	View the resource list in a catalog.
Database	ALL	Perform all operations on databases.
	ALTER	Modify databases.
	DROP	Delete databases.
	DESCRIBE	Check the metadata of databases or switch databases.
	LIST_TABLE	View the resource list in a database.
	LIST_FUNC	View functions in a database.
	CREATE_TABLE	Create a table in a database.
	CREATE_FUNC	Create a function in a database.
Table	ALL	Perform all operations on tables.
	ALTER	Modify tables.
	DROP	Delete tables.
	DESCRIBE	Check the metadata of tables.
	UPDATE	Update table data.

Object	Operation Type	Description
	INSERT	Insert table data.
	SELECT	Query data in a table.
	DELETE	Delete data from a table.
Column	SELECT	Query column data in a table.
Function	ALL	Perform all operations on functions.
	ALTER	Modify functions.
	DROP	Delete functions.
	DESCRIBE	Check the metadata of functions.
	EXEC	Execute functions.
Path	READ	Read files stored in a path.
	WRITE	Write data into the files stored in a path.

NOTE

Permission administrators are categorized into two types: system permission administrators and service permission administrators, each with distinct IAM permission requirements and scopes of permission management.

- System permission administrator
  - System permission administrators need the following IAM operation permissions: **lakeformation:policy:describe**, **lakeformation:policy:create**, and **lakeformation:policy:drop**.
  - System permission administrators can assign metadata permissions to other authorization entities and revoke the assigned permissions as well.
- Service permission administrator
  - Service permission administrators need the following IAM operation permissions: **lakeformation:policy:describe** and **lakeformation:policy:delegate**.
  - Service permission administrators are authorized to assign and revoke metadata permissions only after they have received the requisite permissions from the service permission administrator.

Assume that there are system permission administrator (**User A**), service permission administrator (**User B**), and common user (**User C**). After User A has granted User B the **ALL** permissions for **catalog1**, along with authorization rights, User B is then able to grant permissions such as **DESC** for **catalog1** to User C. However, User B does not have the authority to grant permissions for any other catalogs to User C.

4.2 Granting permissions

You can centrally manage permissions on resources in the data lake and grant permissions to different authorization entities on the LakeFormation console.

Before authorization, ensure that the entity to be authorized exists. For example, the IAM user group has been created.

NOTE

You can centrally manage permissions on resources in the data lake on the LakeFormation console. IAM users and user groups can also be associated with fine-grained permission policies of LakeFormation for authorization. For details, see **Creating a Custom Policy**. If there are a large number of data resources in the data lake, you are advised to use the LakeFormation console to centrally manage permissions on resources in the data lake.

Adding an Authorization Policy

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left and choose **Data Permissions** > **Data Authorization**.

Step 4 Click **Authorize**. In the displayed dialog box, set parameters by referring to the following table and click **OK**.

NOTE

The name of an authorization entity (including users, user groups, and roles) cannot contain hyphens (-). Otherwise, the operation may fail.

Table 4-3 Data authorization parameters

Parameter		Description
Entity Type	User group	**Select User Group(s)**: Select the user group to be authorized, for example, IAM user group. You can create one on the IAM console in advance.
	Role	**Role**: Select the role to be authorized. You can create roles in advance by following the instructions provided in **Creating a Role and Binding a User with It**.
	User	**Select User(s)**: Select the user to be authorized.
Granted To		• **Resources**: authorizes the resources in LakeFormation instances. • **Paths**: authorizes the paths in the OBS service. This authorization type is used to grant permissions to foreign tables or functions.
Resource Type	Catalog	This parameter is displayed when **Granted To** is set to **Resources**. Select the type of the resource to be authorized. **NOTE** When granting the **SELECT** permission to a table, you need to select columns at the same time. For example, set Column to * to select all columns.
	Database	
	Table	
	Column	
	Function	

Parameter	Description
Row Filter Criterion	Whether to set row filtering criteria for the permission policy. For example, set the row filtering condition to **department = "financial"**. This parameter is displayed when **Resource Type** is set to **Table** or **Column**. After the row filter criteria are set, the operation type can only be **SELECT**.
Paths	This parameter is displayed when **Granted To** is set to **Paths**. Click + to select a path in the authorized OBS file system. You can select a maximum of 10 paths.
Operation Type	Select the operation type to be authorized. Different authorization types have different operation types. For details, see **Table 4-2**.
Grant Authorization Permission	Whether to grant the authorization permission. After authorization permission is granted, an authorization entity has the permission to authorize an object to other authorization entities.

Step 5 Click **Cancel Authorization** in the **Operation** column and click **OK** to cancel the authorization of a user group or role.

NOTICE

Revoked authorizations cannot be recovered.

----End

Adding Permissions for a Specified Resource

You can add permissions for a specified resource (database or table).

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left.

Step 4 Access the page for granting permissions to a specified resource.

- Catalog: In the navigation pane, choose **Metadata** > **Catalog**. In the **Operation** column of the catalog to be authorized, choose **More** > **Authorize**.
- Database: In the navigation pane on the left, choose **Metadata** > **Database**. In the upper right corner, select the name of the catalog to which the target

database belongs from the **Catalog** drop-down list, and choose **More** > **Authorize** in the **Operation** column of the database.

- Data Table: In the navigation pane, choose **Metadata** > **Table**. In the upper right corner, select the catalog to which the target data table belongs and the database name from the **Catalog** and **Database** drop-down lists. Click **Authorize** in the **Operation** column of the data table.
- Function: In the navigation pane on the left, choose **Metadata** > **Function**. In the upper right corner, select the catalog to which the target function belongs and the name of the database from the **Catalog** and **Database** drop-down lists. Click **Authorize** in the **Operation** column of the function.

Step 5 Configure related information by referring to **Table 4-3** and click **OK**.

Step 6 Choose **Data Permissions** > **Data Authorization** to view the authorization information.

After the authorization is complete, users in the selected user group or users or user groups with the selected role can perform operations on the current database.

----End

4.3 Canceling Authorization

This section describes how to cancel an authorized permission.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left and choose **Data Permissions** > **Data Authorization**.

Step 4 Search for the authorization policy to be canceled and click **Operation** next to the authorization policy and click **Revoke Authorization**.

Step 5 Click **OK** to revoke the authorization.

NOTICE

Revoked authorizations cannot be recovered.

----End

4.4 Querying Authorization

This section describes how to query an authorized permission.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left and choose **Data Permissions** > **Data Authorization**.

You can select the target permission information using **Authorization Entity**, **Entity Type**, and **Entity Source** in the upper right corner as filters.

Step 4 View the data authorization information in the displayed list.

The following table lists the information items:

Table 4-4 Authorization information

Item	Description
Policy Type	The values include: ● **DEFAULT**: default permission policy. ● **ROW_FILTER**: row filtering permission policy, including row filtering criteria.
Authorization Entity	Name of the authorized entity.
Entity Type	Type of the authorized entity. **GROUP** indicates the user group, **ROLE** indicates role, and **USER** indicates user.
Authorization Object	Name or path of the authorized resource. If the authorization type is set to **Resources**, the format is *Catalog*.**[***Database***]**.**[***Table***]**.
Resource Type	The values include: ● **CATALOG**: catalog ● **DATABASE**: database ● **TABLE**: table ● **COLUMN**: column ● **FUNC**: function ● **URI**: path
Permission	Name of the authorized permission. For details about the permission description, see **Table 4-2**.
Authorized Permission	Authorized permission.

----End

4.5 Performing Role-based Authorization

If a role has some permissions on resources (such as databases), users or user groups with this role also have the corresponding resource operation permissions.

NOTE

If the service interconnected with the LakeFormation instance requires role authorization, the agency for interconnecting with LakeFormation must contain the permissions of the role.

For example, if the query permission of a role is required after LakeFormation is interconnected with an MRS cluster, select **lakeformation:role:describe** when creating a LakeFormation agency.

Creating a Role and Binding a User with It

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the target LakeFormation instance from the drop-down list box on the left and choose **Data Permission** > **Role**.

Step 4 Click **Create**, set **Role Name** and **Description**, and click **OK**.

Step 5 On the **Roles** page, click **Add** in the **Operation** column, select the target role and user, and click **OK**.

NOTE

- You can also choose **Data Permissions** > **Users** in the navigation pane, locate the row that contains the user to be bound with the role, click **Add** in the **Operation** column, select the target role, and click **OK**.
- After the role is authorized, the users bound with the role have its permissions.

----End

Granting Permissions to the Created Role

Grant permissions to the created role. For details, see **Granting permissions**.

5 Data Migration Management

5.1 Granting the Job Management Permission

Scenario

LakeFormation supports full or incremental migration of metadata and permissions from external services to the current LakeFormation instance for unified management.

Before managing jobs, you need to delegate LakeFormation access permissions to the current user so that related data can be written during metadata and permission migration.

Prerequisites

You have created a user by following the instructions provided in **Creating a User and Assigning Permissions** and and added the user to the **admin** user group.

Procedure

Step 1 Log in to the LakeFormation console as a user in the admin user group.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and choose **Tasks** > **Job Authorization** in the navigation pane.

Step 4 Click **Authorized** to grant the LakeFormation job management permission to the current user.

To cancel the permission, click **Cancel Authorization**.

NOTE

After the authorization is approved, LakeFormation automatically creates an agency named **lakeformation_job_trust**. Do not delete the agency during job running.

----End

5.2 Migrating Metadata

Scenario

Migrate external metadata to LakeFormation and store the data in OBS for unified management.

Prerequisites

- A catalog for storing migration metadata has been created for the current instance.
- The target user has the permission to perform operations on OBS and the catalog for storing migration metadata.
- You have created an OBS parallel file system for storing migrated data.
- The name of the table owner can contain 1 to 49 characters, including only letters, digits, and underscores (_). The value cannot contain other characters such as hyphens (-).
- If metadata in multiple MRS clusters needs to be migrated to the same LakeFormation instance, the database names of the MRS clusters must be unique.
- If multiple migrations are necessary, the table column updates must adhere to compatibility requirements, ensuring both column order and column type consistency.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and choose **Tasks** > **Metadata Migration** in the navigation pane.

Step 4 Click **Create Migration Task**, set related parameters, and click **Submit**.

Table 5-1 Creating a metadata migration task

Parameter	Description
Task Name	Name of the metadata migration task.
Description	Description of the created migration task.

Parameter	Description
Data Source	Type of the data to be migrated. The options are as follows: ● **DLF** ● **MRS RDS for MySQL** ● **OpenSource HiveMetastore for MySQL** ● **MRS RDS FOR PostgreSQL** ● **MRS LOCAL GaussDB**
JDBC URL	JDBC URL of the metadata to be migrated. Set this parameter when **Data Source Type** is not set to **DLF**. **NOTE** Some examples are as follows: ● JDBC URL of the MySQL data source type: **jdbc:mysql://** *IP address*:*Port number*/*Database name* **?useSSL=false&permitMysqlScheme** ● JDBC URL of the PostgreSQL data source type: **jdbc:postgresql://** *IP address*:*Port number*/*Database name* **?socketTimeout=600** **socketTimeout** indicates the socket timeout interval for the connection between the migration client and the database. ● When configuring the network, the URL will contain the IP address associated with the EIP linked to the data source. In addition, you need to set the following parameters: ● Username: username for accessing the data source. ● Password: password for accessing the data source. If the user has a password, this parameter is mandatory. Otherwise, leave this parameter blank.
Access Point	Access point of the metadata service to be migrated. This parameter is displayed when **Data Source** is set to **DLF**. In addition, you need to set the following parameters: ● **Access Key**: Obtain the AK from DLF O&M personnel. ● **Secret Key**: Obtain the SK from DLF O&M personnel.
Source Catalog	Name of the catalog to which the metadata to be migrated belongs.
Target Catalog	Name of the catalog to which metadata is migrated in LakeFormation.
Conflict Resolution	Policy for resolving conflicts during migration. Currently, only **Update old metadata** is supported.

Parameter	Description
Default Owner	Default owner of metadata after migration. This parameter is displayed when **Data Source** is set to **DLF**. ● If the configured default owner does not have the corresponding metadata operation permission, the migrated metadata cannot be added, deleted, modified, or queried. In this case, you can grant permissions to the owner or migrate permissions. ● If all metadata can be used properly before the migration, you do not need to set this parameter.
Log Path	Storage location of logs generated during migration. Click + to select a path. The path must exist in OBS. If the path is customized, the migration task will fail.
Force Table Creation	Selecting this option will bypass OBS path restrictions when creating an internal table.
Metadata Objects to Migrate	Select the metadata objects to be migrated. The available options are: ● **All**: databases, functions, data tables, and partitions. ● **Database**: databases. ● **Function**: functions. ● **Table**: tables. ● **Partition**: partitions. **NOTE** – Select **All** to migrate all metadata for the first migration task. – Ensure that the upper-level directory of the selected metadata exists if **All** is not selected. For example, you need to ensure that the target catalog contains the database (for example, **DB_1**) where the tables are located if you plan to set this parameter to **Table**. Otherwise, the table migration will fail. – Ensure that the function class name exists if you plan to set this parameter to **Function** to guarantee a successful function migration task.

Parameter	Description
Add Location Rule	- If the prefix of the metadata storage path is not **obs://**, click **Add Location Rule** to replace the prefix with **obs://** and ensure that the corresponding OBS storage path exists. For example, if the current metadata storage path is **file:/a/b**, set **Original Path** to **file:/** and **New Path** to **obs://**. Ensure that the **obs://a/b** path exists in the OBS parallel file system, the new metadata storage path is **obs://a/b**. - You can create multiple rules at the same time. If a rule conflict occurs, the rule on the top of the page prevails.
Network Connection	Select a network connection scheme. You are advised to select **EIP**. In **EIP** is selected, you need to also select **Security Group ID**, which corresponds to the security group ID of the VPC associated with the data source.
Event notification policy (Currently, this function is in the OBT phase.)	(Optional) Once this option is configured, a notification (via SMS or email) will be sent when a specific event (such as task success or failure) occurs. - Event Notification: If this function is enabled, event notifications will be activated. - Event Notification Topic: Select the topic to be notified. You can configure the topic using Simple Message Notification (SMN) on the management console. - Event: Specifies the status of the topic to be notified. The value can be either **Task succeeded** or **Task failed**.

Step 5 Click **Start** in the **Operation** column to run the migration task.

NOTE

- Before running a migration task, you need to authorize the task by referring to **Granting the Job Management Permission**.
- After the migration task starts, if new metadata is added to the source database, the new metadata will not be migrated. You need to run the migration task again to migrate the new metadata. You can also use the metadata discovery function to migrate new metadata. For details, see **Using the Metadata Discovery Function**.
- If the task fails to be executed, you can click **Start** in the **Operation** column to retry after rectifying the fault.

You can click **Metadata** on the navigation pane and click the name of target metadata object to view the metadata object after the migration. For example, choose **Metadata** > **Database** to view the migrated database.

Click **Edit** or **Delete** in the **Operation** column to modify or delete a task.

Step 6 Click **View Log** in the **Operation** column to view the logs generated during task running.

By default, the latest 50 lines of logs are displayed.

You can click the hyperlink at the bottom of the log to view the complete log. For details about the configuration, see section "Downloading an Object" in *Object Storage Service 3.0 (OBS) 3.24.3h&s User Guide (for Huawei Cloud Stack 8.3.1) in Object Storage Service 3.0 (OBS) 3.24.3h&s Usage Guide (for Huawei Cloud Stack 8.3.1)*.

The following table lists some error messages in logs and their causes.

Error Message	Cause
field 'storageDescriptor.location' must match '^(obs|har)://.+/.+$'	Incorrect location rule is configured. (The metadata storage path should start with **obs://**.)
Invalid input parameter	The input parameter of the metadata is invalid or LakeFormation does not support such metadata.
Incorrect type of column *xxx*.	The column type is invalid or LakeFormation is incompatible with the column type.
No permission to perform this operation on resources.	The default owner is incorrectly configured or the owner does not have the metadata operation permission.

Error Message	Cause
Error creating transactional connection factory	The LakeFormation server is disconnected from the data source. The solution is as follows: 1. Check whether the username, password, AK, and SK of the data source are correct. 2. Check whether the database entered in JDBC URL is correct. 3. Check whether the IP address entered in JDBC URL is correct. If the data source type is MRS local metadata, an active/standby DBServer switchover may occur. In this case, you need to bind the EIP to the active node again. 4. Check whether the security group of the database connection port is enabled. – For tasks using an EIP connection mode, 0.0.0.0/0 needs to be allowed in the data source's security group rules before execution. – When opting for the VPC peer connection mode, the data source's security group must permit access from the VPC peering connection's peer IP address.
The entered VPC network segment conflicts with the LakeFormation network segment.	When the VPC peer connection mode is chosen, a conflict occurs between the VPC network segment of the data source and that of the LakeFormation server. In this case, you can choose to use EIP for migration.
The log is not found.	Check whether the log path exists. ● If the log path already exists, contact LakeFormation O&M personnel for assistance. ● If the log path does not exist, modify the log path in the task configuration to ensure that the log path exists in OBS.
The path should be a sub path of the catalog storage location or database location list	The path must be a subpath of the catalog storage location or database storage location list.

Error Message	Cause
Incorrect Partition Value	The entered partition value is incorrect. Check whether the number and type of the entered partition key list of the table match those of the entered partition value list.
Database does not exist	The database does not exist. Verify whether the database is present.
Location doesn't exist in the OBS Parallel File Systems	The path does not exist in the OBS parallel file system.

----End

5.3 Migrating Permissions

Scenario

After metadata migration is complete, you can migrate metadata permissions to LakeFormation. After the migration is successful, the default owner bound to the metadata will have the metadata operation permissions.

Prerequisites

- Metadata has been migrated. (For how to migrate metadata, see **Migrating Metadata**.)
- You have obtained the permission to perform operations on OBS and have created an OBS parallel file system for storing data.
- The permission policy files to be migrated have been exported and uploaded to the OBS parallel file system. For details about how to export permissions, contact the corresponding service support personnel.
- Authorization entities (except roles) in the permission policy have been created in advance and their names are consistent. Metadata objects contained in the permission policy already exist and their names are consistent.

  If the migration type is set to **DLF**, the mapping and migration policies are as follows:

  - RAM user: IAM user (If the corresponding IAM user does not exist, the permission policy will not be migrated.)
  - RAM role: IAM user group (If the corresponding IAM user group does not exist, the permission policy will not be migrated.)
  - DLF role: LakeFormation role (If this role does not exist, it will be automatically created.)
- If **Policy Type** is Ranger, only the **allow** permission of Ranger can be migrated. The **deny** permission cannot be migrated.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and choose **Tasks** > **Permission Migration** in the navigation pane on the left.

Step 4 Click **Create Migration Task**, set related parameters, and click **Submit**.

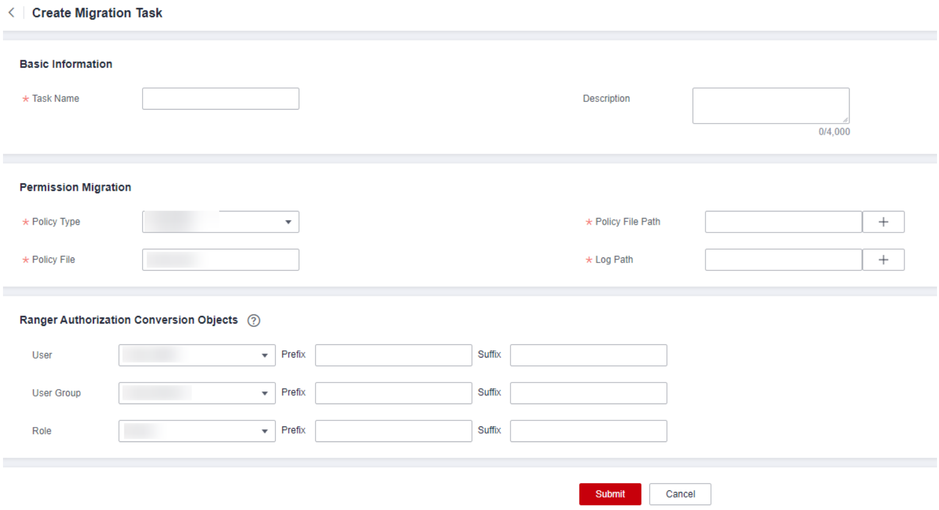


Table 5-2 Creating a permission migration task

Parameter	Description
Task Name	Name of the permission migration task to be created.
Description	Description of the created migration task.
Policy Type	Type of the permission policy to be migrated. ● **DLF** ● **RANGER**
Policy File Path	Storage location of the permission policy file to be migrated in OBS.
Policy File	Name of the file whose permission policies are to be migrated.
Log Path	Storage location of logs generated during migration.
Catalog ID	Catalog name of the permission source. This parameter needs to be specified when **Policy Type** is set to **DLF**.

Parameter	Description
Ranger Authorization Conversion Objects	Parameter for specifying the conversion relationship between authorization objects of the Ranger permission policy. The prefix and suffix are added to the name of the authorization object. This parameter is displayed when **Policy Type** is set to **RANGER**. You need to configure **User**, **User Group**, **Role**, **Prefix**, and **Suffix** as well. You are advised to convert non-IAM users and non-IAM user groups in Ranger to roles.

Step 5 Click **Run** in the **Operation** column to run the migration task.

Click **View Log** to view the logs generated during task running.

NOTE

- Before running a migration task, you need to authorize the task by referring to **Granting the Job Management Permission**.
- If the task fails to be executed, you can click **Start** in the **Operation** column to retry after rectifying the fault.

Click **Edit** or **Delete** in the **Operation** column to modify or delete a task.

After the migration task is complete, you can click **Data Permissions** and **Data Authorization** to view the migrated LakeFormation permission policies.

----End

5.4 Using the Metadata Discovery Function

Scenario

If data is stored in OBS parallel file systems but is not associated with metadata in LakeFormation, you can use the metadata discovery function to construct metadata corresponding to the data to support the computing and analysis of SQL engines or user applications.

NOTICE

The metadata discovery feature is currently in OBT and is free of charge. However, once it is officially launched, fees will be charged based on the resources consumed by metadata discovery tasks.

Prerequisites

- Authorization has been enabled by referring to **Granting the Job Management Permission**.
- The data to be discovered has been uploaded to the OBS parallel file system. That is, the data has been uploaded from S3 or HDFS to the planned path of

the OBS parallel file system in the region where the LakeFormation instance is located.

- The catalog and database for metadata discovery have been prepared and created.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and choose **Tasks** > **Metadata Discovery** in the navigation pane.

Step 4 Click **Create Discovery Task**, set related parameters, and click **Submit**.

Table 5-3 Creating a discovery task

Parameter	Description
Task Name	Name of the metadata discovery task.
Description	Description of the created metadata discovery task.
Data Storage Location	Location where the discovered metadata is stored in the OBS parallel file system. Click +, select a location, and click **OK**.
Discovery File Type	Type of the discovered file. The options include: ● **Automatic discovery** (including Parquet, ORC, JSON, Avro, and CSV) ● **Parquet** ● **ORC** ● **JSON** ● **CSV** (If you select this type, you also need to configure parameters such as **Delimiter**, **Escape Character**, .**Quotation Character**, and **Use first row as column name**.) ● Avro **NOTE** ● If the data storage location contains file name extensions of the same type, it is recommended to choose the matching discovery file type. ● Should there be a variety of file name extensions present, selecting **Automatic discovery** is advisable. ● In the absence of a suffix for the file, opt for the appropriate type. Note that **Automatic discovery** defaults to identifying Parquet files, and may not recognize files of other formats.

Parameter	Description
Log Path	Storage location of logs generated when a metadata discovery task is executed. Click + to select a path. The path must exist in OBS. If the path is customized, the discovery task will fail.
Target Catalog	Name of the catalog to which the metadata to be discovered belongs.
Target Database	Name of the database to which the metadata to be discovered belongs.
Conflict Resolution	Method used to resolve the issue of duplicate metadata names during metadata discovery. ● **Create and update metadata** ● **Create metadata only**
Default Owner	Default owner of metadata after a metadata discovery task is executed. To avoid authorization failure, ensure that the selected entity's name does not contain hyphens (-).
File Sampling Rate	(Optional) File sampling frequency. When the sampling rate is **0**, all partitions after the current partition table are skipped if an empty file is found. This method reduces the operation time, but reduces the accuracy.
Rediscovery Method	Execute the discovery policy for metadata discovery again. ● Full discovery: When you perform the discovery operation again, all files in the data storage location are discovered. ● Incremental discovery: When you perform the discovery operation again, the system discovers the files added to the data storage location after the last task (successfully executed) starts.
Entity Type	(Optional) By default, selecting an entity assigns it read permission on the data storage location. ● You can choose user groups, roles, or users to be the authorization entity. To avoid authorization failure, ensure that the selected entity's name does not contain hyphens (-). ● If you want to grant the write permission as well, select **Write Permission**.

Step 5 Click **Run** in the **Operation** column to run the migration task.

- Click **Stop** to stop a running task.

- Click **View Log** to view the logs generated during task running.

  By default, the latest 50 lines of logs are displayed.

  You can click the hyperlink at the bottom of the log to view the complete log. For details about the configuration, see section "Downloading an Object" in *Object Storage Service 3.0 (OBS) 3.24.3h&s User Guide (for Huawei Cloud Stack 8.3.1) in Object Storage Service 3.0 (OBS) 3.24.3h&s Usage Guide (for Huawei Cloud Stack 8.3.1)*.
- Click **Edit** or **Delete** in the **Operation** column to modify or delete a task.

Step 6 After the migration task is complete, choose **Metadata** > **Table**. In the upper right corner, select the target catalog and database from the **Catalog** and **Database** drop-down lists to view the discovered tables.

----End

6 Managing Clients

On the access management page, you can create and manage clients, view client IP addresses, and use the IP addresses to connect multiple services to LakeFormation instances.

Procedure

Step 1 Log in to the LakeFormation console.

Step 2 In the upper left corner, click ☰ and choose **Analytics** > **LakeFormation** to access the LakeFormation console.

Step 3 Select the LakeFormation instance to be operated from the drop-down list on the left and click **Clients** in the navigation pane.

Step 4 Click **Create**. In the displayed dialog box, set the following parameters and click **OK**.

If no suitable VPC or subnet is available, click **create one** to access the VPC console to create one.

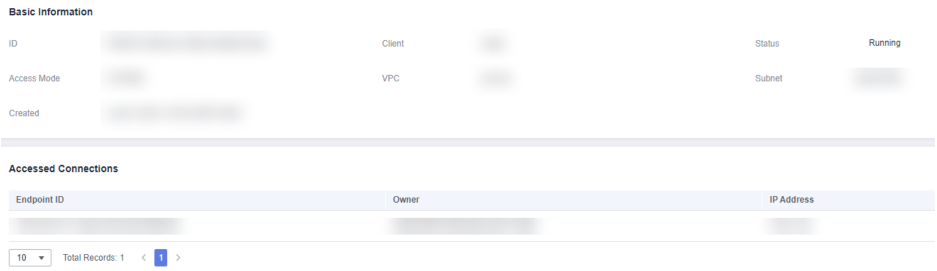
Table 6-1 Parameters for creating an access management client

Parameter	Description
Client	Enter a client name.
VPC	VPC where the service to be connected is located.
Subnet	Subnet where the service to be connected is located.

Step 5 Click **View Details** in the **Operation** column to open the details page.

In the **Basic Information** area, you can view the ID, client name, status, access mode, VPC, and subnet.

In the **Accessed Connections** area, you can view the VPC endpoint ID and access IP address.



Basic Information

ID	[redacted]	Client	[redacted]	Status	Running
Access Mode	[redacted]	VPC	[redacted]	Subnet	[redacted]
Created	[redacted]				

Accessed Connections

Endpoint ID	Owner	IP Address
[redacted]	[redacted]	[redacted]

Step 6 Interconnect other cloud services with LakeFormation based on the obtained information, such as the access IP address.

For details, see the operation guide for interconnecting the corresponding cloud service with LakeFormation. For how to interconnect LakeFormation with MRS, see **Configuring a LakeFormation Data Connection**

----End

7 Viewing Audit Logs

Cloud Trace Service (CTS) records operations on LakeFormation instances for query, audit, and backtrack.

Operations Recorded by CTS

Table 7-1 Operations recorded by CTS

Operation	Resource Type	Trace Name
Creating a catalog	Catalog	createCatalog
Deleting a catalog	Catalog	dropCatalog
Modifying a catalog	Catalog	alterCatalog
Creating a database	Database	createDatabase
Deleting a database	Database	dropDatabase
Modifying a database	Database	alterDatabase
Creating a data table	Table	createTable
Deleting a data table.	Table	dropTable
Modifying a data table.	Table	alterTable
Clearing the data in a table	Table	truncateTable
Creating a function	Function	createFunction
Modifying the attributes of a function	Function	alterFunction
Deleting a function	Function	dropFunction
Creating an instance	instance	createInstance
Modifying an instance	instance	updateInstance

Operation	Resource Type	Trace Name
Deleting an instance	instance	deleteInstance
Granting permissions	Access	grantAccess
Revoking authorization	Access	revokeAccess
Updating statistics about a specified column in a table	TableColumnStatistic	setTableColumnStatistics
Deleting statistics about a specified column in a table	TableColumnStatistic	deleteTableColumnStatistics
Adding column restrictions for creating tables in batches	TableConstraint	addConstraints
Deleting a column restriction	TableConstraint	deleteConstraints
Adding partition information in batches	Partition	addPartitions
Modifying partition information in batches	Partition	alterPartitions
Deleting partition information in batches	Partition	dropPartitions
Clearing list information in batches	Partition	truncatePartitions
Setting partition column statistics in batches	PartitionColumnStatistic	setPartitionColumnStatistics
Deleting partition column statistics	PartitionColumnStatistic	deletePartitionColumnStatistics

Viewing an Audit Log

Query LakeFormation traces on the CTS console.

For details, visit **Querying Real-Time Traces**.