

LakeFormation

Quick Start

Issue 01
Date 2024-01-31



Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of 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 How to Use LakeFormation.....	1
2 Creating a LakeFormation Instance and a Client Connection.....	4
3 Constructing a Data Lake.....	7
4 Interconnecting MRS Clusters with LakeFormation.....	10

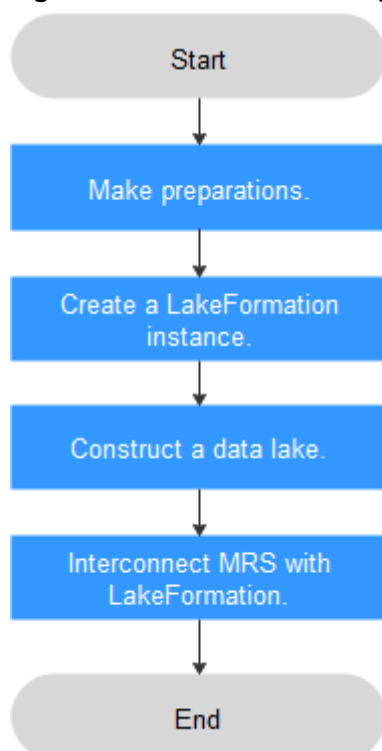
1 How to Use LakeFormation

This LakeFormation tutorial describes how to create a LakeFormation instance and interconnect it with MRS clusters to implement unified data lake metadata and permission management.

General Use Procedure

The following figure shows the procedure of interconnecting MRS with LakeFormation.

Figure 1-1 Procedure of using LakeFormation



Restrictions and Constraints

The restrictions and constraints of interconnecting LakeFormation with MRS clusters are as follows.

- **Before interconnecting LakeFormation with MRS clusters, pay attention to the following restrictions:**
 - MRS clusters and LakeFormation instances must belong to the same cloud account and region.
 - The VPC where the access client created by LakeFormation resides must be in the same VPC as MRS clusters.
 - MRS clusters can only interconnect with Hive Catalog of LakeFormation instances.
 - For existing MRS clusters, you need to migrate the metadata database and permission policies to the LakeFormation instance, and then configure the interconnection.
- **After MRS is interconnected with LakeFormation, MRS components are subject to the following constraints:**
 - Hive does not support temporary tables.
 - Hive does not support cross-cluster column encryption.
 - Hive WebHCat cannot interconnect with LakeFormation.
 - If the table directory is not empty when Hive creates an internal table, the table cannot be created.
 - Before creating a Hudi table, you need to add the path authorization of the Hudi table directory on LakeFormation to grant OBS read and write permissions.
 - Fields in a Hudi table cannot be edited on the LakeFormation console. You can only add, delete, or modify table fields on the Hudi client.
 - When Flink reads and writes Hive tables, only **hive_sync.mode=jdbc** can be used to synchronize Hive tables. HMS is not supported.
 - When Spark tries to access the **default** library on the client, it needs permission for the OBS path. If Spark uses a user account that lacks this permission, the client shows a message saying so. However, this does not affect the database creation, which still succeeds.
- **After MRS is interconnected with LakeFormation, the permission policy restrictions are as follows:**
 - In LakeFormation authorization, only LakeFormation roles can be used as authorization entities. Users or user groups cannot be used as authorization entities.
 - The PolicySync process does not modify the default policy of the RangerAdmin Hive module in the cluster. The default policy still takes effect.
 - After the PolicySync process is started, it compares the permissions with those of LakeFormation instances and deletes the non-default policies that do not exist in LakeFormation. You are advised to migrate the permission policies to LakeFormation instances first.
 - For the Hive module on the RangerAdmin web UI, do not add or delete non-default policies. Grant permissions on the data permission page of LakeFormation instances.
 - After the interconnection between the MRS cluster and LakeFormation is canceled, the non-default policies of RangerAdmin will not be cleared. You need to manually clear them.

- Hive does not support SQL statements for granting permissions. You need to grant permissions on the **Data Permissions** page.

2 Creating a LakeFormation Instance and a Client Connection

Preparing a User


If you use LakeFormation for the first time, create a user and grant permissions to the user by referring to [Preparations](#).

This section demonstrates the entire process of using LakeFormation. You need to complete the following preparations in advance:

- An IAM service user and user group for authorization have been created in IAM, and the user has been associated with the user group.
- A parallel file system has been created in OBS, and a folder for mapping data tables has been created in the file system.
- You have created an agency required for interconnecting with LakeFormation. For example, if you plan to interconnect MRS clusters with LakeFormation, create an agency by referring to [Preparations](#).

Creating a LakeFormation 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 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 Parameters for creating a LakeFormation instance


Parameter	Description	Example Value
Type	Select the instance type. <ul style="list-style-type: none">• Shared• Exclusive	Exclusive
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.	xxx
Billing Mode	Billing mode of instances. <ul style="list-style-type: none">• Pay-per-use	Pay-per-use
Project	Select the project to which the instance belongs.	xxx
Name	User-defined LakeFormation instance name.	lakeformation-test
QPS	Maximum number of requests per second. You do not need to set this parameter when Type is set to Shared .	10000
Enterprise Project	Enterprise project to which the cluster belongs. If no enterprise project is available, click Create to create one.	xxx
Description	Description of the instance.	-
Label	Enter content in the boxes and click Add to add a tag.	-

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.

----End

Creating an Access Client

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 created LakeFormation instance (for example, **lakeformation-test**). For how to create an instance, see the operations described in [Creating a LakeFormation Instance](#). 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 2-2 Parameters for creating an access management client

Parameter	Description
Client	Enter a client name.
VPC	Select the VPC to which the client to be connected belongs. The value must be the same as that of the service to be connected.
Subnet	Select the subnet to which the client to be connected belongs. The value must be the same as that of the service to be connected.

- Step 5** Click **View Details** in the **Operation** column to view the VPC endpoint ID, access domain name, and access IP address.

----End

3 Constructing a Data Lake

MRS can only connect to the **hive** catalog and the **default** database in LakeFormation.

After a LakeFormation instance is created, you can create catalogs and metadata such as databases and tables as required.

For example, perform the following steps to create a catalog named **hive** (fixed name, which cannot be customized) and create a database named **default** (fixed name, which cannot be customized) in the catalog. Assign permissions for the two data tables **table_A** and **table_B** contained in the database.

Creating an OBS Path for Storing Metadata


- 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/table**
- Function storage path: **lakeformation-test/catalog1/database1/udf1**

----End

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 created LakeFormation instance (for example, **lakeformation-test**). For how to create an instance, see the operations described in [Creating a LakeFormation Instance and a Client Connection](#). Choose **Metadata > Catalog** in the navigation pane.

You only need one **hive** catalog per instance. If you already have one in the current instance, you can skip creating another one.

Step 4 Click **Create** in the upper right corner. On the displayed page, set the following parameters, and click **Submit**.


- **Catalog Name: hive**
The name is fixed and cannot be customized.
- **Select Location:** Click **+**, select a storage location, for example, **obs://lakeformation-test/catalog1**, and click **OK**.
- **Catalog Type: DEFAULT**
- Retain the default values for other parameters.

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

----End

Creating 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 created LakeFormation instance (for example, **lakeformation-test**). For how to create an instance, see the operations described in [Creating a LakeFormation Instance and a Client Connection](#). Choose **Metadata > Database** in the navigation pane.

Step 4 Select **hive** from the **Catalog** drop-down list box in the upper right corner.

You only need one **default** database per instance. If you already have one in the current instance, you can skip creating another one.


Step 5 Click **Create**, set related parameters, and click **Submit**.

- **Database Name: default**
The name is fixed and cannot be customized.
- **Catalog: hive**
- **Select Location:** Click **+**, select a storage location, for example, **obs://lakeformation-test/catalog1/database1**, and click **OK**.
- Retain the default values for other parameters.


Step 6 After the catalog is created, you can view the database information on the **Database** page.

----End

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 the created LakeFormation instance (for example, **lakeformation-test**) from the drop-down list box on the left. For how to create an instance, see [Creating a LakeFormation Instance and a Client Connection](#). Choose **Metadata > Table** in the navigation pane and select **hive** and **default** from the **Catalog** and **Database** drop-down lists in the upper right corner.
- Step 4** Click **Create**, set related parameters, and click **Submit**.
- **Table Name:** Enter a table name, for example, **table_A**.
 - **Data Storage Location:** Click **+**, select a location for storing the table in the OBS parallel file system, for example, **obs://lakeformation-test/catalog1/database1/table1**, and click **OK**.
 - Set other parameters based on the site requirements.
- Step 5** Create another table named **table_B** and set the storage path to **lakeformation-test/catalog1/database1/table2** by referring to [Step 4](#).
- Step 6** After these two tables are created, you can check these tables on the **Table** page.
- End

Authorizing Metadata

- 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**, authorize the **hive** catalog and the **default** database as required, and click **OK**.
- For details about authorization configurations and requirements, see [Granting permissions](#).
- End

4 Interconnecting MRS Clusters with LakeFormation

Preparations

For how to prepare for data connection, see [Preparations](#). Create an agency and configure a LakeFormation data connection by following the instructions.

Connecting MRS to LakeFormation

Only MRS 3.3.0-LTS clusters can interconnect with LakeFormation instances.

For how to interconnect MRS with LakeFormation instances, see [Interconnecting MRS 3.3.0-LTS Clusters with LakeFormation](#).