

Data Lake Insight

User Guide

Issue 01
Date 2024-08-19



Copyright © Huawei 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 Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI and other Huawei trademarks are trademarks 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 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.

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process*. For details about this process, visit the following web page:

<https://www.huawei.com/en/psirt/vul-response-process>

For vulnerability information, enterprise customers can visit the following web page:

<https://securitybulletin.huawei.com/enterprise/en/security-advisory>

Contents

1 DLI Console Overview.....	1
2 Creating an Elastic Resource Pool and Queues Within It.....	2
2.1 Overview of DLI Elastic Resource Pools and Queues.....	2
2.2 Creating an Elastic Resource Pool.....	4
2.3 Creating Queues in an Elastic Resource Pool.....	7
2.4 Creating a Standard Queue (Discarded and Not Recommended).....	10
2.5 Managing Elastic Resource Pools.....	13
2.5.1 Checking Basic Information.....	13
2.5.2 Managing Permissions.....	14
2.5.3 Binding a Queue.....	16
2.5.4 Setting CUs.....	16
2.5.5 Modifying Specifications.....	20
2.5.6 Managing Tags.....	22
2.5.7 Viewing Scaling History.....	23
2.5.8 Allocating to an Enterprise Project.....	24
2.6 Managing Queues.....	25
2.6.1 Queue Permission Management.....	26
2.6.2 Adjusting Scaling Policies for Queues in an Elastic Resource Pool.....	28
2.6.3 Allocating a Queue to an Enterprise Project.....	33
2.6.4 Creating an SMN Topic.....	34
2.6.5 Managing Queue Tags.....	36
2.6.6 Setting Queue Properties.....	38
2.6.7 Testing Address Connectivity.....	39
2.6.8 Modifying Queue Specifications.....	40
2.6.9 Deleting a Queue.....	41
2.6.10 Elastic Scaling of Standard Queues (Discarded and Not Recommended).....	42
2.6.11 Setting a Scheduled Elastic Scaling Task for a Standard Queue (Discarded and Not Recommended)	44
2.6.12 Changing the CIDR Block of a Standard Queue (Discarded and Not Recommended).....	48
2.7 Example Use Case: Creating an Elastic Resource Pool and Running Jobs.....	49
2.8 Example Use Case: Configuring Scaling Policies for Queues in an Elastic Resource Pool.....	55
3 Job Management.....	61
3.1 Overview.....	61

3.2 Creating and Managing SQL Jobs.....	62
3.2.1 Submitting a SQL Job Using the SQL Editor.....	62
3.2.2 Managing SQL Jobs.....	68
3.2.3 Exporting SQL Job Results.....	71
3.3 Creating and Managing Flink Jobs.....	75
3.3.1 Flink Job Management Overview.....	75
3.3.2 Managing Flink Job Permissions.....	77
3.3.3 Preparing Flink Job Data.....	80
3.3.4 Creating a Flink OpenSource SQL Job.....	81
3.3.5 Creating a Flink SQL Job.....	89
3.3.6 Creating a Flink Jar Job.....	95
3.3.7 Performing Operations on a Flink Job.....	101
3.3.8 Flink Job Details.....	108
3.3.9 Tag Management.....	115
3.3.10 Enabling Dynamic Scaling for Flink Jobs.....	117
3.4 Creating and Managing Spark Jobs.....	119
3.4.1 Spark Job Management.....	119
3.4.2 Creating a Spark Job.....	121
3.5 Setting the Priority for a Job.....	128
4 Data Management.....	132
4.1 Databases and Tables.....	132
4.1.1 Overview.....	132
4.1.2 Managing Database Permissions.....	136
4.1.3 Managing Table Permissions.....	141
4.1.4 Creating a Database or a Table.....	150
4.1.5 Deleting a Database or a Table.....	158
4.1.6 Modifying the Owners of Databases and Tables.....	159
4.1.7 Importing Data to the Table.....	159
4.1.8 Exporting Data from DLI to OBS.....	163
4.1.9 Viewing Metadata.....	166
4.1.10 Previewing Data.....	167
4.1.11 Managing Tags.....	167
4.2 Package Management.....	171
4.2.1 Overview.....	171
4.2.2 Managing Permissions on Packages and Package Groups.....	172
4.2.3 Creating a Package.....	176
4.2.4 Deleting a Package.....	178
4.2.5 Modifying the Owner.....	178
4.2.6 Built-in Dependencies.....	179
5 Enhanced Datasource Connections.....	208
5.1 Overview.....	208
5.2 Cross-Source Analysis Development Methods.....	210

5.3 Creating an Enhanced Datasource Connection.....	212
5.4 Deleting an Enhanced Datasource Connection.....	218
5.5 Modifying Host Information in an Elastic Resource Pool.....	218
5.6 Binding and Unbinding a Queue.....	220
5.7 Adding a Route.....	222
5.8 Deleting a Route.....	223
5.9 Enhanced Connection Permission Management.....	224
5.10 Enhanced Datasource Connection Tag Management.....	225
6 Datasource Authentication.....	228
6.1 Overview.....	228
6.2 Creating a CSS Datasource Authentication.....	230
6.3 Creating a Kerberos Datasource Authentication.....	232
6.4 Creating a Kafka_SSL Datasource Authentication.....	236
6.5 Creating a Password Datasource Authentication.....	239
6.6 Datasource Authentication Permission Management.....	242
7 Job Templates.....	244
7.1 Managing SQL Templates.....	244
7.2 Managing Flink Templates.....	249
7.3 Managing Spark SQL Templates.....	255
7.4 Appendix.....	256
7.4.1 TPC-H Sample Data in the SQL Template.....	256
8 DLI Agency Permissions.....	259
8.1 DLI Agency Overview.....	259
8.2 Updating the Permissions of a DLI Agency.....	262
8.3 Customizing DLI Agency Permissions.....	268
8.4 Agency Permission Policies in Common Scenarios.....	274
8.5 Guidelines for Configuring DLI Agency Permissions in Specific Scenarios.....	277
9 Creating a DLI Global Variable.....	278
10 Configuring a DLI Job Bucket.....	282
11 SQL Inspector.....	284
11.1 Overview.....	284
11.2 Creating a SQL Inspection Rule.....	285
11.3 System Rules.....	286
12 Creating a Custom Image.....	291
13 Permissions Management.....	296
13.1 Overview.....	296
13.2 Creating an IAM User and Granting Permissions.....	301
13.3 Creating a Custom Policy.....	303
13.4 DLI Resources.....	309

13.5 DLI Request Conditions.....	309
13.6 Common Operations Supported by DLI System Policy.....	310
14 Other Common Operations.....	315
14.1 Importing Data to a DLI Table.....	315
14.2 Viewing Monitoring Metrics.....	315
14.3 DLI Operations That Can Be Recorded by CTS.....	322
14.4 Quota Management.....	325
15 Permission Management for Global Variables.....	327

1 DLI Console Overview

The **Overview** page of the DLI console provides you with the DLI workflow and resource usage.

How to Use DLI

The process of using DLI is as follows:

1. **Create an elastic resource pool and create queues within it.**

Queues are DLI's compute resources. There are SQL queues and general-purpose queues. For a SQL queue, you can only submit Spark SQL jobs. For a general-purpose queue, you can submit Spark programs, Flink OpenSource SQL jobs, and Flink Jar jobs.

2. **Prepare data.**

Create databases and tables before you run a Spark SQL job. Upload a program package before you run a Spark job or a Flink Jar job.

3. **Edit and submit a job.**

After you set the job parameters, you can submit the job.

4. **View job status.**

Go to the **Job Management** page to view the job execution status.

2 Creating an Elastic Resource Pool and Queues Within It

2.1 Overview of DLI Elastic Resource Pools and Queues

DLI queues are classified into queues in an elastic resource pool, standard queues, and default queue.

- **Queues in an elastic resource pool:**

An elastic resource pool offers compute resources (CPU and memory) required for running DLI jobs, which can adapt to the changing demands of services.

You can create multiple queues within an elastic resource pool. These queues are associated with specific jobs and data processing tasks, and serve as the basic unit for resource allocation and usage within the pool. This means queues are specific compute resources required for executing jobs.

Queues within an elastic resource pool can be shared to execute jobs. This is achieved by correctly setting the queue allocation policy. This enhances queue utilization.

For how to buy an elastic resource pool and create queues within it, see [Creating an Elastic Resource Pool](#) and [Creating Queues in an Elastic Resource Pool](#).

- **Standard queue (deprecated and not recommended):** Previous-gen compute resources for running DLI jobs. Before buying such a queue, you need to estimate how many resources you will need.
- **default queue:**
 - The **default** queue is preset in DLI, which allocates resources as needed. If you are unsure about how much queue capacity you will require or you do not have a space for creating queues, you can use this **default** queue to run your jobs.
 - The **default** queue is typically used by users who are new to our service, but it can cause resource contention and prevent you from getting the resources you need for your jobs. It is recommended that you create your own queue to run your jobs.

Queues in an elastic resource pool are recommended, as they offer the flexibility to use resources with high utilization as needed.

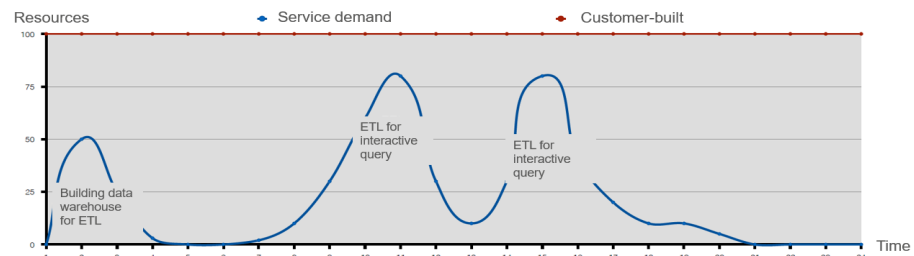
Use Cases of Elastic Resource Pools

Resources too fixed to meet a range of requirements.

The quantities of compute resources required for jobs change in different time of a day. If the resources cannot be scaled based on service requirements, they may be wasted or insufficient. **Figure 2-1** shows the resource usage during a day.

- After ETL jobs are complete, no other jobs are running during 04:00 to 07:00 in the early morning. The resources could be released at that time.
- From 09:00 to 12:00 a.m. and 02:00 to 04:00 p.m., a large number of ETL report and job queries are queuing for compute resources.

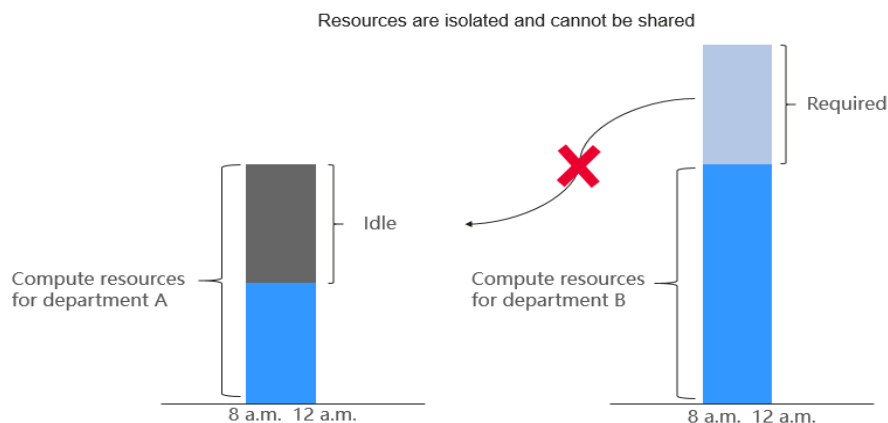
Figure 2-1 Fixed resources



Resources are isolated and cannot be shared.

A company has two departments, and each run their jobs on a DLI queue. Department A is idle from 08:00 to 12:00 a.m. and has remaining resources, while department B has a large number of service requests during this period and needs more resources to meet the requirements. Since the resources are isolated and cannot be shared between department A and B, the idle resources are wasted.

Figure 2-2 Resource waste due to resource isolation



Elastic resource pools can be accessed by different queues and automatically scaled to improve resource utilization and handle resource peaks.

You can use elastic resource pools to centrally manage and allocate resources. Multiple queues can be bound to an elastic resource pool to share the pooled resources.

2.2 Creating an Elastic Resource Pool

An elastic resource pool offers compute resources (CPU and memory) required for running DLI jobs, which can adapt to the changing demands of services.

You can create multiple queues within an elastic resource pool. These queues are associated with specific jobs and data processing tasks, and serve as the basic unit for resource allocation and usage within the pool. This means queues are specific compute resources required for executing jobs.

Queues within an elastic resource pool can be shared to execute jobs. This is achieved by correctly setting the queue allocation policy. This enhances queue utilization. This section describes how to create an elastic resource pool.

Notes and Constraints

- The billing mode of an elastic resource pool cannot be changed.
- The region of an elastic resource pool cannot be changed.
- For a pay-per-use resource pool, the dedicated resource mode is selected by default. The resource pool is billed by calendar hour since it is created.
- Jobs of Flink 1.10 or later can run in elastic resource pools.
- The CIDR block of an elastic resource pool cannot be changed once set.
- Associating an elastic resource pool with a queue:
 - Only pay-per-use queues (including dedicated queues) can be associated with elastic resource pools.
 - No resources are frozen.
- Currently, only yearly/monthly elastic resource pools can be scaled.
- You can only view the scaling history of resource pools in the last 30 days.
- Elastic resource pools cannot access the Internet.

For how to access the Internet from an elastic resource pool, see [Configuring the Connection Between a DLI Queue and a Data Source on the Internet](#).

- Changes to elastic resource pool CUs can occur when setting the CU, adding or deleting queues in an elastic resource pool, or modifying the scaling policies of queues in an elastic resource pool, or when the system automatically triggers elastic resource pool scaling. However, in some cases, the system cannot guarantee that the scaling will reach the target CUs as planned.
 - If there are not enough physical resources, an elastic resource pool may not be able to scale out to the desired target size.
 - The system does not guarantee that an elastic resource pool will be scaled in to the desired target size.

The system checks the resource usage before scaling in the elastic resource pool to determine if there is enough space for scaling in. If the

existing resources cannot be scaled in according to the minimum scaling step, the pool may not be scaled in successfully or only partially.

The scaling step may vary depending on the resource specifications, usually 16 CUs, 32 CUs, 48 CUs, 64 CUs, etc.

For example, if the elastic resource pool has a capacity of 192 CUs and the queues in the pool are using 68 CUs due to running jobs, the plan is to scale in to 64 CUs.

When executing a scaling in task, the system determines that there are 124 CUs remaining and scales in by the minimum step of 64 CUs.

However, the remaining 60 CUs cannot be scaled in any further.

Therefore, after the elastic resource pool executes the scaling in task, its capacity is reduced to 128 CUs.

Creating an Elastic Resource Pool

Step 1 In the navigation pane on the left, choose **Resources > Resource Pool**.

Step 2 On the displayed page, click **Buy Resource Pool** in the upper right corner.

Step 3 On the displayed page, set the following parameters:

Table 2-1 Parameters

Parameter	Description
Region	Select a region. Select a region near you to ensure the lowest latency possible.
Project	Each region corresponds to a project.
Name	Name of the elastic resource pool. <ul style="list-style-type: none">Only numbers, letters, and underscores (_) are allowed. The value cannot contain only numbers or start with an underscore (_) or number.The value can contain a maximum of 128 characters. NOTE The elastic resource pool name is case-insensitive. Uppercase letters will be automatically converted to lowercase letters.
Specifications	<ul style="list-style-type: none">Basic: 16 to 64 CUs<ul style="list-style-type: none">This edition is suitable for testing scenarios with low resource consumption and low requirements for resource reliability and availability.High reliability and availability are not supported.Queue properties and job priorities cannot be set.Notebook instances cannot be interconnected with.Standard: 64 CUs or higher This edition offers powerful computing capabilities, high availability, and flexible resource management. It is suitable for large-scale computing tasks and business scenarios with long-term resource planning needs.

Parameter	Description
CU Range	The maximum and minimum CUs allowed for the elastic resource pool.
Description	Description of the elastic resource pool
CIDR Block	<p>CIDR block the elastic resource pool belongs to. If you use an enhanced datasource connection, this CIDR block cannot overlap that of the data source. Once set, this CIDR block cannot be changed.</p> <p>Recommended CIDR block: 10.0.0.0~10.255.0.0/16~19 172.16.0.0~172.31.0.0/16~19 192.168.0.0~192.168.0.0/16~19</p>
Enterprise Project	<p>If the created elastic resource pool belongs to an enterprise project, select the enterprise project.</p> <p>Enterprise projects let you manage cloud resources and users by project.</p> <p>For details about how to set enterprise projects, see Enterprise Management User Guide.</p> <p>NOTE This parameter is displayed only for users who have enabled the Enterprise Management Service.</p>
Tags	<p>Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS).</p> <p>For details, see Tag Management Service User Guide.</p> <p>NOTE</p> <ul style="list-style-type: none">• A maximum of 20 tags can be added.• Only one tag value can be added to a tag key.• The key name in each resource must be unique. <p>• Tag key: Enter a tag key name in the text box.</p> <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters ([:=+-@) are allowed, but the value cannot start or end with a space or start with <code>_sys_</code>.</p> <p>• Tag value: Enter a tag value in the text box.</p> <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters ([:=+-@) are allowed. The value cannot start or end with a space.</p>

Step 4 Click **Buy** and confirm the configurations.

Step 5 Click **Pay**. Wait until the status of the elastic resource pool changes to **Available**. The elastic resource pool is successfully created.

Step 6 Refer to [Example Use Case: Creating an Elastic Resource Pool and Running Jobs](#) and [Example Use Case: Configuring Scaling Policies for Queues in an Elastic Resource Pool](#) to perform subsequent operations as needed.

----End

2.3 Creating Queues in an Elastic Resource Pool

You can add one or more queues to an elastic resource pool to run jobs. This section describes how to add a queue to an elastic resource pool.

NOTE

Queues added to an elastic resource pool are not billed separately, but billed as billing items of the elastic resource pool.

Notes and Constraints

- A queue named **default** is preset in DLI for you to experience. Resources are allocated on demand. You are billed based on the amount of data scanned in each job (unit: GB).
- Queue types:
 - For SQL: Spark SQL jobs can be submitted to SQL queues.
 - For general purpose: The queue is used to run Spark programs, Flink SQL jobs, and Flink Jar jobs.

The queue type cannot be changed. If you want to use another queue type, purchase a new queue.

- The billing mode of a queue cannot be changed.
- The region of a queue cannot be changed.
- Queues with 16 CUs do not support scale-out or scale-in.
- Queues with 64 CUs do not support scale-in.
- When creating a queue, you can only select cross-AZ active-active for yearly/monthly queues and pay-per-use dedicated queues. The price of a cross-AZ queue is twice that of a single-AZ queue.
- Newly created queues need to run jobs before they can be scaled in or out.
- DLI queues cannot access the Internet.

For how to access the Internet from an elastic resource pool, see [Configuring the Connection Between a DLI Queue and a Data Source on the Internet](#).

Adding a Queue

Step 1 In the navigation pane on the left, choose **Resources > Resource Pool**.

Step 2 Locate the target elastic resource pool and click **Add Queue** in the **Operation** column.

Step 3 On the **Add Queue** page, configure basic queue information. The following table describes the parameters.

Table 2-2 Queue information

Parameter	Description
Name	Name of the queue to add
Type	<ul style="list-style-type: none">For SQL: The queue is used to run SQL jobs.General general purpose: The queue is used to run Spark and Flink jobs.
Engine	If Type is For SQL , the queue engine can be spark or trino .
Enterprise Project	Select the enterprise project the queue belongs to. Queues under different enterprise projects can be added to an elastic resource pool. Enterprise projects let you manage cloud resources and users by project. For details about how to set enterprise projects, see Enterprise Management User Guide . NOTE This parameter is displayed only for users who have enabled the Enterprise Management Service.
Description	Description about the queue.
Tags	Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS). For details, see Tag Management Service User Guide . NOTE <ul style="list-style-type: none">A maximum of 20 tags can be added.Only one tag value can be added to a tag key.The key name in each resource must be unique. <ul style="list-style-type: none">Tag key: Enter a tag key name in the text box. NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with _sys_.Tag value: Enter a tag value in the text box. NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed. The value cannot start or end with a space.

Step 4 Click **Next**. On the displayed page, configure a scaling policy for the queue in the elastic resource pool.

Figure 2-3 Configuring a scaling policy when adding a queue

The screenshot shows a web interface for configuring scaling policies. At the top, there are two tabs: 'Basic Configuration' and 'Elastic Resources'. Below the tabs, there is a heading 'View scaling policies of all queues in pool_test'. A small text block explains that the priority ranges from 1 to 100 and that the period is in 'start time, end time' format. The form has four main sections: 'Priority' with a text input containing '1'; 'Period' with two dropdown menus showing '00' and '24'; 'Min CU' with a numeric input '16' and '+'/'-' buttons; and 'Max CU' with a numeric input '16' and '+'/'-' buttons. A 'Create' button is centered at the bottom of the form.

Click Create to add a scaling policy with specified priority, period, Minimum CUs, and Maximum CUs. The parameters of each scaling policy are as follows:

Table 2-3 Auto scaling policy

Parameter	Description
Priority	Priority of the scaling policy in the current elastic resource pool. A larger value indicates a higher priority. You can set a number ranging from 1 to 100.
Period	Time segment when the policy takes effects. It can be set only by hour. The start time is on the left, and the end time is on the right. <ul style="list-style-type: none"> The time range includes the start time but not the end time, that is, [start time, end time). For example, if the Period to 01 and 17, the scaling policy takes effect at 01:00 a.m. till 05:00 p.m. The periods of scaling policies with different priorities cannot overlap.
Min CU	Minimum number of CUs allowed by the scaling policy. <ul style="list-style-type: none"> In any time segment of a day, the total minimum CUs of all queues in an elastic resource pool cannot be more than the minimum CUs of the pool. If the minimum CUs of the queue is less than 16 CUs, both Max. Spark Driver Instances and Max. Prestart Spark Driver Instances set in the queue properties do not apply. Refer to Setting Queue Properties.
Max CU	Maximum number of CUs allowed by the scaling policy. In any time segment of a day, the maximum CUs of any queue in an elastic resource pool cannot be more than the maximum CUs of the pool.

NOTE

- The first scaling policy is the default policy, and its **Period** parameter configuration cannot be deleted or modified.
- Flink jobs cannot trigger automatic scaling of queues in an elastic resource pool.

- Step 5** Click **OK**. View all queues and scaling policies added to the elastic resource pool by referring to [Adjusting Scaling Policies for Queues in an Elastic Resource Pool](#).

----End

2.4 Creating a Standard Queue (Discarded and Not Recommended)

Standard queues are the previous-gen compute resources for running DLI jobs. Before buying such a queue, you need to estimate how many resources you will need.

Queues in an elastic resource pool are recommended, as they offer the flexibility to use resources with high utilization as needed. For how to buy an elastic resource pool and create queues within it, see [Creating an Elastic Resource Pool](#) and [Creating Queues in an Elastic Resource Pool](#).

NOTE

- If you use a sub-account to create a queue for the first time, log in to the DLI management console using the main account and keep records in the DLI database before creating a queue.
- It takes 6 to 10 minutes for a job running on a new queue for the first time.
- After a queue is created, if no job is run within one hour, the system releases the queue.
- Queues with 16 CUs do not support scale-out or scale-in.
- Queues with 64 CUs do not support scale-in.

Procedure


1. You can create a queue on the **Overview**, **SQL Editor**, or **Queue Management** page.
 - In the upper right corner of the **Overview** page, click Purchase Queue.
 - To create a queue on the **Queue Management** page:
 - i. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
 - ii. In the upper right corner of the **Queue Management** page, click **Buy Queue** to create a queue.
 - To create a queue on the **SQL Editor** page:
 - i. In the navigation pane of the DLI management console, click **SQL Editor**.
 - ii. Click **Queues**. On the tab page displayed, click  on the right to create a queue.
2. On the **Buy Queue** page displayed, set the parameters according to [Table 2-4](#).

Table 2-4 Parameters

Parameter	Description
Billing Mode	Pay-per-use Billing for CUH used = Number of CUs x Usage duration x Unit price. You are billed for used CUs on an hourly basis (rounded up to the nearest hour). The pay-per-use billing mode is adopted. You are advised to purchase the CUH package to enjoy preferential price.
Region	Select a region. Select a region near you to ensure the lowest latency possible.
Project	Each region corresponds to a project.
Name	Queue name <ul style="list-style-type: none">Only numbers, letters, and underscores (_) are allowed. The value cannot contain only numbers, start with an underscore (_), or be left unspecified.The value can contain a maximum of 128 characters. NOTE The queue name is case-insensitive. Uppercase letters will be automatically converted to lowercase letters.
Type	<ul style="list-style-type: none">For SQL: compute resources used for SQL jobs.For general purpose: compute resources used for Spark and Flink jobs. NOTE When a dedicated queue is not in use, its resources are still reserved and not released. This means that resources are constantly being held, regardless of whether the queue is being utilized. By using a dedicated queue, resources can be guaranteed to be available whenever jobs are submitted. You can create enhanced datasource connections for a dedicated queue. When buying a pay-per-use queue, you have the option to choose a dedicated queue. Dedicated queues are billed 24/7 regardless of whether they are used.
Specifications	The compute nodes' total number of CUs. One CU equals one vCPU and 4 GB of memory. DLI automatically assigns CPU and memory resources to each compute node, and the client does not need to know how many compute nodes are being used.
Enterprise Project	If the created queue belongs to an enterprise project, you can select the corresponding enterprise project. Enterprise projects let you manage cloud resources and users by project. For details about how to set an enterprise project, see Enterprise Management User Guide . NOTE This parameter is displayed only for users who have enabled the Enterprise Management Service.

Parameter	Description
Description	Description of the queue to be created. The value can contain a maximum of 128 characters.
Advanced Settings	<p>In the Queue Type area, select Dedicated Resource Mode and then click Advanced Settings.</p> <ul style="list-style-type: none">• Default: The system automatically configures the parameter.• Custom CIDR Block: You can specify the CIDR block. For details, see Changing the CIDR Block of a Standard Queue (Discarded and Not Recommended). If DLI enhanced datasource connection is used, the CIDR block of the DLI queue cannot overlap with that of the data source. Recommended CIDR block: 10.0.0.0~10.255.0.0/8~24 172.16.0.0~172.31.0.0/12~24 192.168.0.0~192.168.0.0/16~24 <p>Queue Type: When running an AI-related SQL job, select AI-enhanced. When running other jobs, select Basic.</p>
Tags	<p>Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS). For details, see Tag Management Service User Guide.</p> <p>NOTE</p> <ul style="list-style-type: none">• A maximum of 20 tags can be added.• Only one tag value can be added to a tag key.• The key name in each resource must be unique.• Tag key: Enter a tag key name in the text box. <p>NOTE</p> <p>A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with _sys_.</p> <ul style="list-style-type: none">• Tag value: Enter a tag value in the text box. <p>NOTE</p> <p>A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed. The value cannot start or end with a space.</p>

3. Click **Buy** to confirm the configuration.
4. Confirm the configuration and click **Submit**.

If the queue name already exists, the system displays a message indicating that the queue name already exists when you click **Submit**. In this case, click **Previous** to go back to modify the queue name.

- Once created, you can check and use it on the **Queue Management** page.

 **NOTE**

It takes 6 to 10 minutes for a job running on a new queue for the first time.


2.5 Managing Elastic Resource Pools

2.5.1 Checking Basic Information

After creating an elastic resource pool, you can check and manage it on the management console.

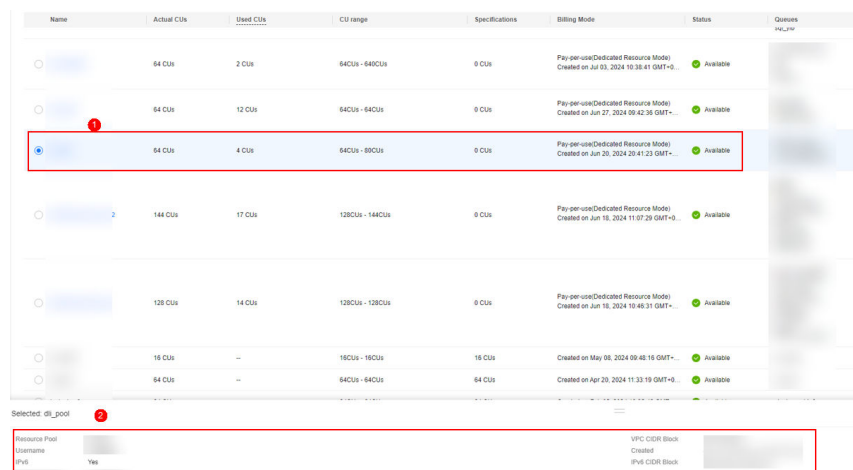
This section describes how to check basic information about an elastic resource pool on the management console, including the VPC CIDR block, IPv6 CIDR block, and creation time.

Procedure

- Log in to the DLI management console.
- In the navigation pane on the left, choose **Resources** > **Resource Pool**.
- On the displayed page, select the elastic resource pool you want to check.
 - In the upper right corner of the list page, click  to customize the columns to display and set the rules for displaying the table content and the **Operation** column.
 - In the search box above the list, you can filter the required elastic resource pool by name or tag.
- Click the name of the elastic resource pool you selected at the bottom of the page to check its details.

These details will include the name of the pool, the user who created it, the date it was created, the VPC CIDR block, and whether IPv6 is enabled. If IPv6 is enabled, the subnet's IPv6 CIDR block will also be displayed.

Figure 2-4 Basic information about an elastic resource pool



Name	Actual CUs	Used CUs	CU range	Specifications	Billing Mode	Status	Queues
[blurred]	64 CUs	2 CUs	64CUs - 640CUs	0 CUs	Pay-per-use(Dedicated Resource Mode) Created on Jul 03, 2024 10:38:41 GMT+08:00	Available	[blurred]
[blurred]	64 CUs	12 CUs	64CUs - 640CUs	0 CUs	Pay-per-use(Dedicated Resource Mode) Created on Jun 27, 2024 09:42:36 GMT+08:00	Available	[blurred]
[blurred]	64 CUs	4 CUs	64CUs - 640CUs	0 CUs	Pay-per-use(Dedicated Resource Mode) Created on Jun 20, 2024 20:41:23 GMT+08:00	Available	[blurred]
[blurred]	144 CUs	17 CUs	128CUs - 144CUs	0 CUs	Pay-per-use(Dedicated Resource Mode) Created on Jun 19, 2024 11:07:29 GMT+08:00	Available	[blurred]
[blurred]	128 CUs	14 CUs	128CUs - 128CUs	0 CUs	Pay-per-use(Dedicated Resource Mode) Created on Jun 18, 2024 10:46:31 GMT+08:00	Available	[blurred]
[blurred]	16 CUs	--	16CUs - 16CUs	16 CUs	Created on May 08, 2024 09:48:16 GMT+08:00	Available	[blurred]
[blurred]	64 CUs	--	64CUs - 640CUs	64 CUs	Created on Apr 20, 2024 11:33:19 GMT+08:00	Available	[blurred]

Selected: dl_pool	
Resource Pool	[blurred]
Username	[blurred]
IPv6	Yes
VPC CIDR Block	[blurred]
Created	[blurred]
IPv6 CIDR Block	[blurred]

2.5.2 Managing Permissions

Administrators can assign permissions of different operation scopes to users for each elastic resource pool.

Precautions

- The administrator and elastic resource pool owner have all permissions, which cannot be set or modified by other users.
- When you set resource pool permissions for a new user, ensure that the region of the user group to which the user belongs has the Tenant Guest permission. For details about the Tenant Guest permission and how to apply for the permission, see [Creating a User Group and Assigning Permissions](#) and [System Permissions](#).

Procedure

Step 1 In the navigation pane on the left of the DLI console, choose **Resources > Resource Pool**.

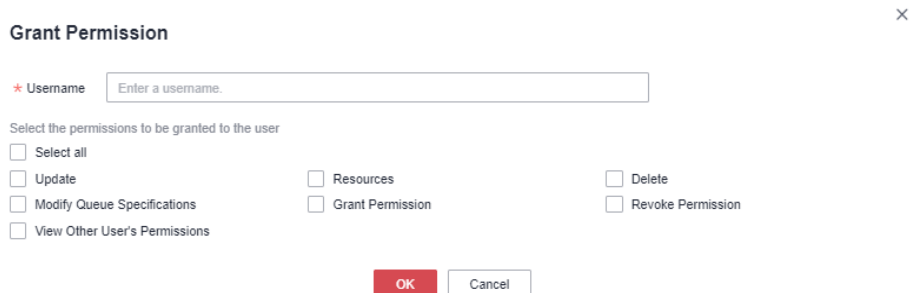
Step 2 Select the desired elastic resource pool and choose **More > Permissions** in the **Operation** column. The **User Permissions** area displays the list of users who have permissions of elastic resource pools.

You can assign permissions to new users, modify permissions for users who already have some permissions of elastic resource pools, and revoke all permissions of a user on a pool.

- Assign permissions to a new user.
A new user does not have permissions on the elastic resource pool.
 - a. Click **Set Permission** in the **Operations** column on **User Permissions** page. The **Set Permission** dialog box is displayed.
 - b. Set **Username** to the name of the desired IAM user, and select the required permissions for the user.
 - c. Click **OK** to.

[Table 2-5](#) describes the related parameters.

Figure 2-5 Managing permissions



Grant Permission ×

* Username

Select the permissions to be granted to the user

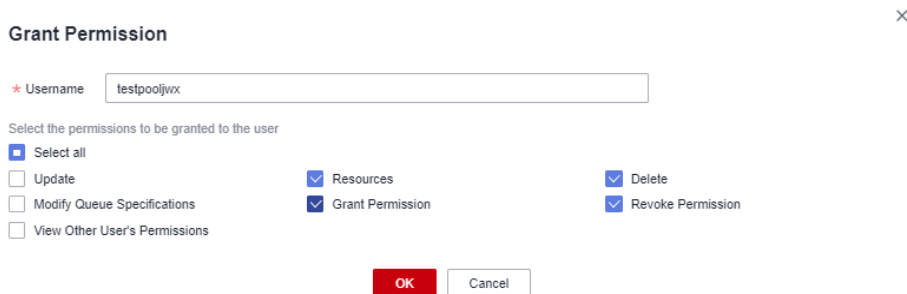
<input type="checkbox"/> Select all	<input type="checkbox"/> Resources	<input type="checkbox"/> Delete
<input type="checkbox"/> Update	<input type="checkbox"/> Grant Permission	<input type="checkbox"/> Revoke Permission
<input type="checkbox"/> Modify Queue Specifications		
<input type="checkbox"/> View Other User's Permissions		

Table 2-5 Parameters

Parameter	Description
Username	Name of the user you want to grant permissions to NOTE The username must be an existing IAM username and has logged in to the DLI management console.
Select the permissions to be granted to the user	<ul style="list-style-type: none"> ▪ Update: Update the description of an elastic resource pool. ▪ Resources: Add queues, delete queues, and configure scaling policies for queues in an elastic resource pool. ▪ Delete: Delete the elastic resource pool. ▪ Grant Permission: Grant the elastic resource pool permissions to other users. ▪ Revoke Permission: Revoke the permissions that other users have but you cannot revoke the owner's permissions. ▪ View Other User's Permissions: View the elastic resource pool permissions of other users.

- To assign or revoke permissions of a user who has some permissions on the elastic resource pool, perform the following steps:
 - a. In the list under **User Permissions**, select the user whose permissions need to be modified and click **Set Permission** in the **Operation** column.
 - b. In the displayed **Set Permission** dialog box, modify the permissions of the user. [Table 2-5](#) lists the detailed permission descriptions.
If **Set Permission** is gray, you are not allowed to change permissions on this elastic resource pool. You can apply to the administrator, elastic resource pool owner, or other authorized users for granting and revoking permissions.

Figure 2-6 Managing permissions



- c. Click **OK**.
- To revoke all permissions of a user on an elastic resource pool, perform the following steps:

In the list under **User Permissions**, select the desired user whose permissions and click **Set Permission** in the **Operation** column. Click **Yes** in the **Revoke Permission** dialog box.

----End

2.5.3 Binding a Queue

Scenario

If you want a queue to use resources in an elastic resource pool, bind the queue to the pool.

You can click **Associate Queue** on the **Resource Pool** page to bind a queue to an elastic resource pool, or bind a queue on the **Queue Management** page.

NOTE

Elastic resource pools support only Flink 1.10 or later. If jobs using Flink 1.7 run on a queue that is bound to an elastic resource pool, errors may occur due to incompatibility.

Prerequisites

- Both the elastic resource pool and queue are available.
- The queue you want to bind must be a dedicated queue in pay-per-use billing mode.
- No resources are frozen.
- Only queues under the same enterprise project can be bound to an elastic resource pool.

Associating a Queue

Step 1 In the navigation pane on the left, choose **Resources > Resource Pool**.

Step 2 Locate the row that contains the desired elastic resource pool, click **More** in the **Operation** column, and select **Associate Queue**.

Step 3 In the displayed dialog box, select the desired queue and click **OK**.

----End

Allocating a Queue to an Elastic Resource Pool

Step 1 In the navigation pane on the left, choose **Resources > Queue Management**.

Step 2 Locate the target queue and choose **More > Bind Resource Pool** in the **Operation** column.

Step 3 Select the desired elastic resource pool and click **OK**.

----End

2.5.4 Setting CUs

CU settings are used to control the maximum and minimum CU ranges for elastic resource pools to avoid unlimited resource scaling.

For example, an elastic resource pool has a maximum of 256 and two queues, and each queue must have at least 64 CUs. If you want to add another queue that needs at least 256 CUs to the elastic resource pool, the operation is not allowed due to the maximum CUs of the elastic resource pool.

Precautions

- In any time segment of a day, the total minimum CUs of all queues in an elastic resource pool cannot be more than the minimum CUs of the pool.
- In any time segment of a day, the maximum CUs of any queue in an elastic resource pool cannot be more than the maximum CUs of the pool.
- When you change the minimum CUs of a created elastic resource pool, ensure that the value is no more than the current CU value. Otherwise, the modification fails.

Setting CUs

- Step 1** In the navigation pane on the left, choose **Resources > Resource Pool**.
- Step 2** Locate the row that contains the desired elastic resource pool, click **More** in the **Operation** column, and select **Set CUs**.
- Step 3** In the **Set CUs** dialog box, set the minimum CUs on the left and the maximum CUs on the right. Click **OK**.

----End

What Are the Actual CUs, CU Range, and Specifications of an Elastic Resource Pool?

- **Actual CUs:** number of CUs that can be allocated in the elastic resource pool.
- **CU range:** CU settings are used to control the maximum and minimum CU ranges for elastic resource pools to avoid unlimited resource scaling.
 - The total minimum CUs of all queues in an elastic resource pool must be no more than the minimum CUs of the pool.
 - The maximum CUs of any queue in an elastic resource pool must be no more than the maximum CUs of the pool.
 - An elastic resource pool should at least ensure that all queues in it can run with the minimum CUs and should try to ensure that all queues in it can run with the maximum CUs.
- **Specifications:** The minimum CUs selected during elastic resource pool purchase are elastic resource pool specifications. Specifications are specific to yearly/monthly elastic resource pools. The specifications are billed on a yearly/monthly basis, while the extra CUs are billed in pay-per-use mode.

How Do I Increase the Minimum Value of the CU Range?

The minimum value of the CU range of an elastic resource pool is less than or equal to the actual CUs of the elastic resource pool. To adjust the minimum value of the CU range to be greater than the current CUs, increase the actual CUs first.

The following operations apply when the number of target CUs is less than or equal to the maximum value of the CU range. If the number of target CUs is

greater than the maximum value of the CU range, increase the maximum CUs of the elastic resource pool first.

- For a yearly/monthly elastic resource pool, you can increase its actual CUs by adjusting the maximum CUs of queues in or adding queues to the elastic resource pool. Then, you can change its specifications to be greater than or equal to the target CUs. In this case, the actual CUs are equal to the CUs provided by the specifications. Then, adjust the minimum value of the CU range to be equal to the target CUs.
- For a pay-per-use elastic resource pool, you can increase its actual CUs by adjusting the maximum CUs of queues in or adding queues to the elastic resource pool. Then, adjust the minimum value of the CU range to be equal to the target CUs.

Example

- Example 1: For a yearly/monthly elastic resource pool, the number of actual CUs is 64, the CU range is from 64 to 96, the number of CUs provided by the specifications is 64, and the target CU range is from 80 to 96.

Procedure

- a. You can increase the actual CUs of an elastic resource pool by adjusting the maximum CUs of queues in or adding queues to the elastic resource pool.

When the total number of maximum CUs of queues in the elastic resource pool is greater than its actual CUs, scale-out is triggered for the elastic resource pool. Actual CUs after scale-out = min (Total number of maximum CUs of queues, Maximum value of the CU range)

- b. Change the specifications of the elastic resource pool to 80 CUs. After the change, the minimum value of the CU range is automatically changed to the number of CUs provided by the specifications.

- Example 2: For a pay-per-use elastic resource pool, the number of actual CUs is 64, the CU range is from 64 to 96, and the target CU range is from 80 to 96.

Procedure

- a. You can increase the actual CUs of an elastic resource pool by adjusting the maximum CUs of queues in or adding queues to the elastic resource pool.

When the total number of maximum CUs of queues in the elastic resource pool is greater than its actual CUs, scale-out is triggered for the elastic resource pool. Actual CUs after scale-out = min (Total number of maximum CUs of queues, Maximum value of the CU range)

- b. Adjust the CU range to the target one (80 to 96).

NOTE

- The adjustment of the CU range, specifications change, and CU setting of an elastic resource pool take effect on the next hour.
- The adjustment of the actual CUs of an elastic resource pool by adding queues takes effect immediately.

How Do I Decrease the Maximum Value of the CU Range?

The minimum value of the CU range of an elastic resource pool is less than or equal to the actual CUs of the elastic resource pool. To adjust the maximum value of the CU range to be less than the current CUs, decrease the actual CUs first.

- For a yearly/monthly elastic resource pool, you can decrease its actual CUs by decreasing the maximum CUs of queues in or deleting queues from the elastic resource pool. Then, you can change its specifications to be less than or equal to the target CUs. In this case, the actual CUs are equal to the CUs provided by the specifications. Then, adjust the minimum value of the CU range to be equal to the target CUs.
- For a pay-per-use elastic resource pool, you can decrease its actual CUs by adjusting the maximum CUs of queues in or deleting queues from the elastic resource pool. Then, adjust the minimum value of the CU range to be equal to the target CUs.

Example

- Example 1: For a yearly/monthly elastic resource pool, the number of actual CUs is 96, the CU range is from 64 to 128, the number of CUs provided by the specifications is 96, and the target CU range is from 64 to 80.

Procedure

- a. You can decrease the actual CUs of an elastic resource pool by decreasing the maximum CUs of queues in or deleting queues from the elastic resource pool.

When the total number of maximum CUs of queues in the elastic resource pool is less than its actual CUs, scale-in is triggered for the elastic resource pool. Actual CUs after scale-in = min (Total number of maximum CUs of queues, Maximum value of the CU range)

- b. Change the specifications of the elastic resource pool to 80 CUs. After the change, the minimum value of the CU range is automatically changed to the number of CUs provided by the specifications.

- Example 2: For a pay-per-use elastic resource pool, the number of actual CUs is 96, the CU range is from 64 to 128, and the target CU range is from 64 to 80.

Procedure

- a. You can decrease the actual CUs of an elastic resource pool by decreasing the maximum CUs of queues in or deleting queues from the elastic resource pool.

When the total number of maximum CUs of queues in the elastic resource pool is less than its actual CUs, scale-in is triggered for the elastic resource pool. Actual CUs after scale-in = min (Total number of maximum CUs of queues, Maximum value of the CU range)

- b. Adjust the CU range to the target one (64 to 80).

NOTE

- The adjustment of the CU range, specifications change, and CU setting of an elastic resource pool take effect on the next hour.
- The adjustment of the actual CUs of an elastic resource pool by adding queues takes effect immediately.

2.5.5 Modifying Specifications

Scenario

If CUs of a yearly/monthly elastic resource pool cannot meet your service requirements, you can modify the CUs. In this case, you will be charged based on the number of CUs exceeding that of the yearly/monthly elastic resource pool.

For example, you have purchased an elastic resource pool with 64 CUs, and you find that most time data processing needs 128 CUs. You can add 64 CUs to the elastic resource pool and be billed based on a CU/hour basis. To save more, you can scale up your elastic resource pool to 128 CUs and be billed on a yearly/monthly basis for the 128-CU package.

If the current specifications of your elastic resource pool do not meet your service needs, you can modify them using the change specifications function.

Precautions

Currently, only yearly/monthly elastic resource pools can be scaled.

Scaling Out

1. In the navigation pane on the left of the console, choose **Resources > Resource Pool**.
2. Select the elastic resource pool you want and choose **More > Modify Specifications** in the **Operation** column.
3. In the **Modify Specifications** dialog page, set **Operation** to **Scale-out** and specify the number of CUs you want to add.

Figure 2-7 Scaling out

Name	erp_test
Billing Mode	Yearly/Monthly
Expired	Jun 29, 2023 23:59:59 GMT+08:00
Before Modification	64 CUs
Operation	<input checked="" type="radio"/> Scale-out <input type="radio"/> Scale-in
Amount (CUs)	<input type="text" value="16"/>
After Modification	80 CUs

4. Confirm the changes and click **OK**.
5. Choose **Job Management > SQL Jobs** to view the status of the SCALE_POOL SQL job.

If the job status is **Scaling**, the elastic resource pool is scaling out. Wait until the job status changes to **Finished**.

Scaling In

NOTE

By default, the minimum number of CUs is **16**. That is, when the specifications of an elastic resource pool are **16 CUs**, you cannot scale the pool down.

1. In the navigation pane on the left, choose **Resources > Resource Pool**.
2. Select the elastic resource pool you want and choose **More > Modify Specifications** in the **Operation** column.
3. In the **Modify Specifications** dialog page, set **Operation** to **Scale-in** and specify the number of CUs you want to add.

Figure 2-8 Scaling in

Name	<input type="text"/>
Billing Mode	Yearly/Monthly
Expired	<input type="text"/>
Before Modification	80 CUs
Operation	<input type="button" value="Scale-out"/> <input checked="" type="button" value="Scale-in"/>
Amount (CUs)	<input type="button" value="-"/> <input type="text" value="4"/> <input type="button" value="+"/>
After Modification	76 CUs

4. Confirm the changes and click **OK**.
5. Choose **Job Management > SQL Jobs** to view the status of the SCALE_POOL SQL job.

If the job status is **Scaling**, the elastic resource pool is scaling in. Wait until the job status changes to **Finished**.

2.5.6 Managing Tags

Tag Management

A tag is a key-value pair that you can customize to identify cloud resources. It helps you to classify and search for cloud resources. A tag consists of a tag key and a tag value.

If you use tags in other cloud services, you are advised to create the same tag (key-value pairs) for cloud resources used by the same business to keep consistency.

DLI supports the following two types of tags:

- Resource tags: non-global tags created on DLI
- Predefined tags: global tags created on Tag Management Service (TMS).
For more information about predefined tags, see [Tag Management Service User Guide](#).

DLI allows you to add, modify, or delete tags for queues.

Step 1 In the left navigation pane of the DLI console, choose **Resources > Resource Pool**.

Step 2 In the **Operation** column of the queue, choose **More > Tags**.

Step 3 The tag management page is displayed, showing the tag information about the current queue.

Step 4 Click **Add/Edit Tag**. The **Add/Edit Tag** dialog is displayed. Enter a tag and a value, and click **Add**.

Figure 2-9 Adding/Editing tags

Add/Edit Tag ×

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#)

To add a tag, enter a tag key and a tag value below.

10 tags available for addition.

Table 2-6 Tag parameters

Parameter	Description
Tag key	<p>You can specify the tag key in either of the following ways:</p> <ul style="list-style-type: none">Click the text box and select a predefined tag key from the drop-down list. To add a predefined tag, you need to create one on TMS and then select it from the Tag key drop-down list. You can click View predefined tags to go to the Predefined Tags page of the TMS console. Then, click Create Tag in the upper corner of the page to create a predefined tag. For details, see Creating Predefined Tags in <i>Tag Management Service User Guide</i>.Enter a tag key in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_ : = + - @) are allowed, but the value cannot start or end with a space or start with _sys_.</p>
Tag value	<p>You can specify the tag value in either of the following ways:</p> <ul style="list-style-type: none">Click the text box and select a predefined tag value from the drop-down list.Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_ : = + - @) are allowed. The value cannot start or end with a space.</p>

Step 5 Click **OK**.

Step 6 (Optional) To delete a tag, locate the row where the tag locates in the tag list and click **Delete** in the **Operation** column to delete the tag.

----End

2.5.7 Viewing Scaling History

Scenario

If you added a queue to or deleted one from an elastic resource pool, or you scaled an added queue, the CU quantity of the elastic resource pool may be changed. You can view historical CU changes of an elastic resource pool on the console.

⚠ CAUTION


When scaling in an elastic resource pool, Spark and SQL jobs may automatically retry. However, if the number of retries exceeds the limit, the jobs will fail and need to be manually executed again.

Prerequisites

Currently, you can only view the historical records generated within the last 30 days on the console.

Viewing Scaling History of an Elastic Resource Pool

1. In the navigation pane on the left, choose **Resources > Resource Pool**.
2. Select the desired elastic resource pool and choose **More > Expansion History** in the **Operation** column.
3. On the displayed page, select a duration to view the CU usage.
You can view the number of CUs before and after a scaling, and the target number of CUs.

The historical records can be displayed in charts or tables. Click  in the upper right corner to switch the display.

For example, the scaling is abnormal according to [Viewing Scaling History](#), and the [Figure 2-11](#) shows that the target number of CUs is 80, the original number of CUs is 64, and the scaled number of CUs is 64. The scaling fails.

Figure 2-10 Scaling history in a chart

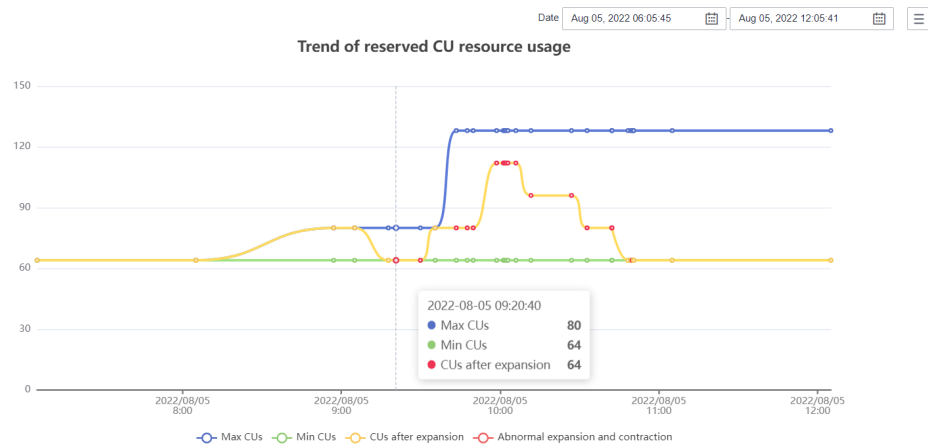


Figure 2-11 Scaling history in a table

Max CUs	Min CUs	Target CUs	CUs before expansion	CUs after expansion	Status	Operation time
128	64	128	80	80	fail	Aug 05, 2022 09:43:26 GMT+08:00
80	64	80	64	80	success	Aug 05, 2022 09:35:29 GMT+08:00
80	64	80	64	64	fail	Aug 05, 2022 09:29:53 GMT+08:00
80	64	80	64	64	fail	Aug 05, 2022 09:20:40 GMT+08:00
80	64	64	80	64	success	Aug 05, 2022 09:17:45 GMT+08:00

2.5.8 Allocating to an Enterprise Project

You can create enterprise projects matching the organizational structure of your enterprises to centrally manage cloud resources across regions by project. Then you can create user groups and users with different permissions and add them to enterprise projects.

DLI allows you to select an enterprise project when creating an elastic resource pool. This section describes how to bind an elastic resource pool to and modify an enterprise project.

NOTE

Modifying the enterprise project of an elastic resource pool will modify the enterprise projects of the queues in the elastic resource pool.

Only queues under the same enterprise project can be bound to an elastic resource pool.

Prerequisites

You have logged in to the Enterprise Project Management Service console and created an enterprise project by referring to [Creating an Enterprise Project](#).

Binding an Enterprise Project

When creating an elastic resource pool, you can select a created enterprise project for **Enterprise Project**.

Alternatively, you can click **Create Enterprise Project** to go to the Enterprise Project Management Service console to create an enterprise project and check existing ones.

For how to create a queue, see [Creating an Elastic Resource Pool](#) and [Creating Queues in an Elastic Resource Pool](#).

Modifying an Enterprise Project

You can modify the enterprise project bound to a created cluster as needed.

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Resources > Resource Pool**.
3. In the elastic resource pool list, locate the elastic resource pool for which you want to modify the enterprise project, click **More** in the **Operation** column, and select **Allocate to Enterprise Project**.
4. In the **Modify Enterprise Project** dialog box displayed, select an enterprise project.
Alternatively, you can click **Create Enterprise Project** to go to the Enterprise Project Management Service console to create an enterprise project and check existing ones.
5. After the modification, click **OK** to save the enterprise project information of the elastic resource pool.

Related Operations

For details about how to modify the enterprise project of a queue, see [Allocating a Queue to an Enterprise Project](#).

2.6 Managing Queues

2.6.1 Queue Permission Management

Scenario

- You can isolate queues allocated to different users by setting permissions to ensure data query performance.
- The administrator and queue owner have all permissions, which cannot be set or modified by other users.
- When setting queue permissions for a new user, ensure that the region of the user group to which the user belongs has the **Tenant Guest** permission. For details about the Tenant Guest permission and how to apply for the permission, see [Creating a User Group and Assigning Permissions](#) and [System Permissions](#).

Operations

Step 1 On the top menu bar of the DLI management console, click **Resources > Queue Management**.

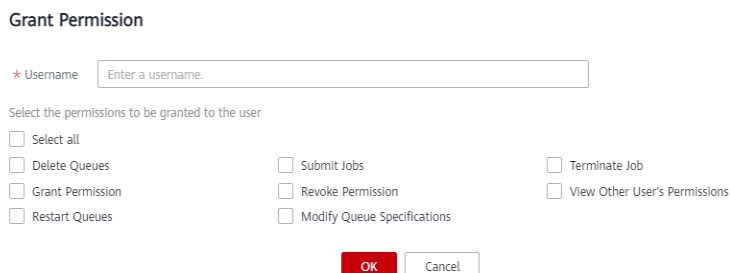
Step 2 Select the queue to be configured and choose **Manage Permissions** in the **Operation** column. The **User Permission Info** area displays the list of users who have permissions on the queue.

You can assign queue permissions to new users, modify permissions for users who have some permissions of a queue, and revoke all permissions of a user on a queue.

- Assign permissions to a new user.
A new user does not have permissions on the queue.
 - a. Click **Set Permission** on the right of **User Permissions** page. The **Set Permission** dialog box is displayed.
 - b. Specify **Username** and select corresponding permissions.
 - c. Click **OK**.

[Table 2-7](#) describes the related parameters.

Figure 2-12 Queue permission granting



Grant Permission

* Username

Select the permissions to be granted to the user

<input type="checkbox"/> Select all	<input type="checkbox"/> Submit Jobs	<input type="checkbox"/> Terminate Job
<input type="checkbox"/> Delete Queues	<input type="checkbox"/> Revoke Permission	<input type="checkbox"/> View Other User's Permissions
<input type="checkbox"/> Grant Permission	<input type="checkbox"/> Modify Queue Specifications	
<input type="checkbox"/> Restart Queues		

Table 2-7 Parameter description

Parameter	Description
Username	Name of the authorized user. NOTE The username is an existing IAM user name and has logged in to the DLI management console.
Permission Settings	<ul style="list-style-type: none">▪ Delete Queues: This permission allows you to delete the queue.▪ Submit Jobs: This permission allows you to submit jobs using this queue.▪ Terminate Jobs: This permission allows you to terminate jobs submitted using this queue.▪ Grant Permission: This permission allows you to grant queue permissions to other users.▪ Revoke Permission: This permission allows you to revoke the queue permissions that other users have but cannot revoke the queue owner's permissions.▪ View Other User's Permissions: This permission allows you to view the queue permissions of other users.▪ Restart Queues: This permission allows you to restart queues.▪ Modify Queue Specifications: This permission allows you to modify queue specifications.

- To assign or revoke permissions of a user who has some permissions on the queue, perform the following steps:
 - a. In the list under **User Permission Info** for a queue, select the user whose permissions need to be modified and click **Set Permission** in the **Operation** column.
 - b. In the displayed **Set Permission** dialog box, modify the permissions of the current user. [Table 2-7](#) lists the detailed permission descriptions.
If all options under **Set Permission** are gray, you are not allowed to change permissions on this queue. You can apply to the administrator, queue owner, or other authorized users for queue permission granting and revoking.

Figure 2-13 Setting queue permissions

The screenshot shows a 'Set Permission' dialog box. At the top, there is a text input field for 'Username'. Below it, the text 'Select the permissions to be granted to the user' is displayed. A list of permissions follows, each with a checkbox: 'Select all' (checked), 'Delete Queues', 'Grant Permission', 'Restart Queues', 'Submit Jobs' (checked), 'Revoke Permission', 'Modify Queue Specifications', 'Terminate Job', and 'View Other User's Permissions'. At the bottom, there are two buttons: 'OK' (highlighted in red) and 'Cancel'.

- c. Click **OK**.
- To revoke all permission of a user on the queue, perform the following steps:
In the user list under **Permission Info**, select the user whose permission needs to be revoked and click **Revoke Permission** under **Operation**. In the **Revoke Permission** dialog box, click **OK**. All permissions on this queue are revoked.

----End

2.6.2 Adjusting Scaling Policies for Queues in an Elastic Resource Pool

Multiple queues can be added to an elastic resource pool. For details about how to add a queue, see [Creating Queues in an Elastic Resource Pool](#). You can configure the number of CUs you want based on the compute resources used by DLI queues during peaks and troughs and set priorities for the scaling policies to ensure stable running of jobs.

Precautions

- In any time segment of a day, the total minimum CUs of all queues in an elastic resource pool cannot be more than the minimum CUs of the pool.
- In any time segment of a day, the maximum CUs of any queue in an elastic resource pool cannot be more than the maximum CUs of the pool.
- The periods of scaling policies cannot overlap.
- The period of a scaling policy can only be set by hour and specified by the start time and end time. For example, if you set the period to **00-09**, the time range when the policy takes effect is [00:00, 09:00). The period of the default scaling policy cannot be modified.
- In any period, compute resources are preferentially allocated to meet the minimum number of CUs of all queues. The remaining CUs (total CUs of the elastic resource pool – total minimum CUs of all queues) are allocated in accordance with the scaling policy priorities.
- After the queue is scaled out, the system starts billing you for the added CUs. So, if you do not have sufficient requirements, scale in your queue to release unnecessary CUs to save cost.

Table 2-8 CU allocation (without jobs)

Scenario	CUs
<p>An elastic resource pool has a maximum number of 256 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none">• Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64• Queue B: priority 10; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none">1. The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are 160 CUs remaining.2. The remaining CUs are allocated based on the priorities. Queue B is prior to queue A. Therefore, queue B gets 128 CUs, and queue A has 32 CUs.
<p>An elastic resource pool has a maximum number of 96 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none">• Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64• Queue B: priority 10; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none">1. The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are no remaining CUs.2. The allocation is complete.
<p>An elastic resource pool has a maximum number of 128 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none">• Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64• Queue B: priority 10; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none">1. The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are 32 CUs remaining.2. The remaining 32 CUs are preferentially allocated to queue B.
<p>An elastic resource pool has a maximum number of 128 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none">• Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64• Queue B: priority 5; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none">1. The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are 32 CUs remaining.2. The two queues have the same priority, the remaining 32 CUs are randomly allocated to the two queues.

Table 2-9 CU allocation (with jobs)

Scenario	Actual CUs of Elastic Resource Pool	CUs Allocated to Queue A	CUs Allocated to Queue B	Allocation Description
<p>Queues A and B are added to the elastic resource pool. The scaling policies are as follows:</p> <ul style="list-style-type: none"> Queue A: period: 00:00–09:00; minimum CU: 32; maximum CU: 64 Queue B: period: 00:00–09:00; minimum CU: 64; maximum CU: 128 	192 CUs	64 CUs	128 CUs	If the actual CUs of the elastic resource pool are greater than or equal to the sum of the maximum CUs of the two queues, the maximum CUs are allocated to both queues.
	96 CUs	32 CUs	64 CUs	The elastic resource pool preferentially meets the minimum CUs of the two queues. After the minimum CUs are allocated to the two queues, no CUs are allocatable.
	128 CUs	32 CUs to 64 CUs	64 CUs to 96 CUs	The elastic resource pool preferentially meets the minimum CUs of the two queues. That is, 32 CUs are allocated to queue A, 64 CUs are allocated to queue B, and the remaining 32 CUs are available. The remaining CUs are allocated based on the queue load and priority. The actual CUs of the queue change within the range listed.

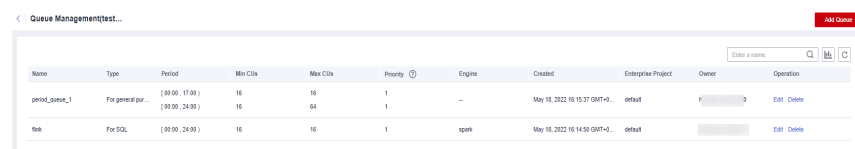
Managing Queues

- Step 1** In the navigation pane on the left, choose **Resources > Resource Pool**.
- Step 2** Locate the target elastic resource pool and click **Queue MGMT** in the **Operation** column. The **Queue Management** page is displayed.
- Step 3** View the queues added to the elastic resource pool.

Table 2-10 Queue parameters

Parameter	Description
Name	Name of the queue to add
Type	Queue type <ul style="list-style-type: none"> • For SQL • For general purpose
Period	The start and end time of the queue scaling policy. This time range includes the start time but not the end time, that is, [start time, end time).
Min CUs	Minimum number of CUs allowed by the scaling policy.
Max CUs	Maximum number of CUs allowed by the scaling policy.
Priority	Priority of the scaling policy for a queue in the elastic resource pool. The priority ranges from 1 to 100. A smaller value indicates a lower priority.
Engine	For a queue running SQL jobs, the engine is Spark. For a queue for general purpose, the engine can be Spark or Flink, but it is displayed by -- in this page.
Created	Time when a queue is added to the elastic resource pool
Enterprise Project	Enterprise project the queue belongs to. Queues under different enterprise projects can be added to an elastic resource pool.
Owner	User who added this queue
Operation	<ul style="list-style-type: none"> • Edit: Modify or add a scaling policy. • Delete: Delete the queue.

Figure 2-14 Managing queues



Step 4 Locate the target queue and click **Edit** in the **Operation** column.

Step 5 In the displayed **Queue Management** pane, perform the following operations as needed:

Figure 2-15 Editing scaling policies for a queue

The screenshot shows the 'Edit Queue' interface. At the top right, there is a close button (X). The main section is titled 'Elastic Resources' and contains a table of scaling policies. The table has four columns: 'Priority', 'Period', 'Min CU', and 'Max CU'. There are two rows of policies. The first row has a priority of 1, a period from 00 to 24, a minimum CU of 16, and a maximum CU of 64. The second row has a priority of 1, a period from 00 to 17, a minimum CU of 16, and a maximum CU of 16, with a 'Delete' button next to it. Below the table is a 'Create' button. Below the 'Elastic Resources' section is the 'Tags' section, which includes a text input field for a tag key and a tag value, and an 'Add' button. At the bottom right, there are 'OK' and 'Cancel' buttons.

Elastic Resources

The priority ranges from 1 to 100. If you do not set the priority for a specific period, the default value is 1. The new priority overrides the default priority, time segments, max and min CUs. You can only set the period to hours in [start time,end time) format.

Priority	Period	Min CU	Max CU
1	00 -- 24	16	64
1	00 -- 17	16	16

Tags

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#)

To add a tag, enter a tag key and a tag value below.

Enter a tag key:

Enter a tag value:

Add

10 tags available for addition.

OK Cancel

- **Add:** Click **Create** to add a scaling policy. Set **Priority**, **Period**, **Min CU**, and **Max CU**, and click **OK**.
- **Modify:** Modify parameters of an existing scaling policy and click **OK**.
- **Delete:** Locate the row that contains the scaling policy you want, click **Delete** and click **OK**.

NOTE

The **Priority** and **Period** parameters must meet the following requirements:

- **Priority:** The default value is 1. The value ranges from 1 to 100. A larger value indicates a higher priority.
- **Period:**
 - You can only set the period to hours in [start time,end time) format.
 - For example, if the **Period** to 01 and 17, the scaling policy takes effect at 01:00 a.m. till 05:00 p.m.
 - The periods of scaling policies with different priorities cannot overlap.
- **Max CUs and Min CUs:**
 - In any time segment of a day, the total minimum CUs of all queues in an elastic resource pool cannot be more than the minimum CUs of the pool.
 - In any time segment of a day, the maximum CUs of any queue in an elastic resource pool cannot be more than the maximum CUs of the pool.

Step 6 After you finish the settings, click statistics icon in the upper right corner of the queue list to view all scaling policies of all queue in the elastic resource pool.

Figure 2-16 Viewing statistic graphics

Name	Type	Period	Min Cls	Max Cls	Priority	Engine	Created	Enterprise Project	Owner	Operation
period_queue_1	For general pur...	(00:00, 17:00) (00:00, 24:00)	16	16	1	--	May 18, 2022 16:15:37 GMT+8...	default		Edit Delete
flink	For SQL	(00:00, 24:00)	16	16	1	spark	May 18, 2022 16:14:58 GMT+8...	default		Edit Delete

Figure 2-17 Statistic graphics of queues and scaling policies



Step 7 View the scaling task generated when the scaling starts. Go to **Job Management** > **SQL Jobs** and view the jobs of the **SCALE_QUEUE** type.

----End

2.6.3 Allocating a Queue to an Enterprise Project

You can create enterprise projects matching the organizational structure of your enterprises to centrally manage cloud resources across regions by project. Then you can create user groups and users with different permissions and add them to enterprise projects.

DLI allows you to select an enterprise project when creating a queue. This section describes how to bind a DLI queue to and modify an enterprise project.

NOTE

Currently, enterprise projects can be modified only for queues that have not been added to elastic resource pools.

Prerequisites

You have logged in to the Enterprise Project Management Service console and created an enterprise project by referring to [Creating an Enterprise Project](#).

Modifying an Enterprise Project

You can change the enterprise project associated with a queue that has been created.

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Resources** > **Queue Management**.

3. In the queue list, locate the queue for which you want to modify the enterprise project, click **More** in the **Operation** column, and select **Modify Enterprise Project**.
4. In the **Modify Enterprise Project** dialog box displayed, select an enterprise project.
Alternatively, you can click **Create Enterprise Project** to go to the Enterprise Project Management Service console to create an enterprise project and check existing ones.
5. After the modification, click **OK** to save the enterprise project information of the queue.

2.6.4 Creating an SMN Topic

Scenario

Once you have created an SMN topic, you can easily subscribe to it by going to the **Topic Management > Topics** page of the SMN console. You can choose to receive notifications via SMS or email. After the subscription is successful, if a job fails, the system automatically sends a message to the subscription endpoint you specified.

- If a job fails within 1 minute of submission, a message notification is not triggered.
- If a job fails after 1 minute of submission, the system automatically sends a message to the subscriber terminal you specified.

Procedure

1. On the **Resources > Queue Management** page, click **Create SMN Topic** on the upper left side. The **Create SMN Topic** dialog box is displayed.

Figure 2-18 Creating an SMN topic

Create SMN Topic

After creating a message notification topic, you can subscribe to it in different ways. After the subscription succeeds, any job failure will automatically be sent to your subscription endpoints.

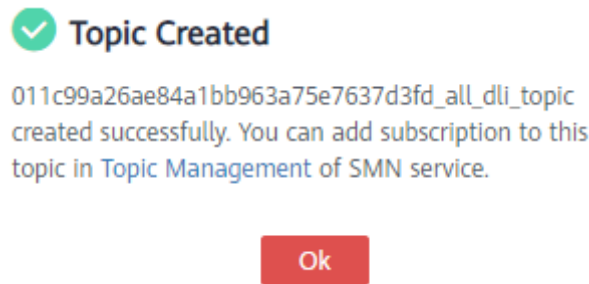
--Select-- ▼

2. Select a queue and click **OK**.

 **NOTE**

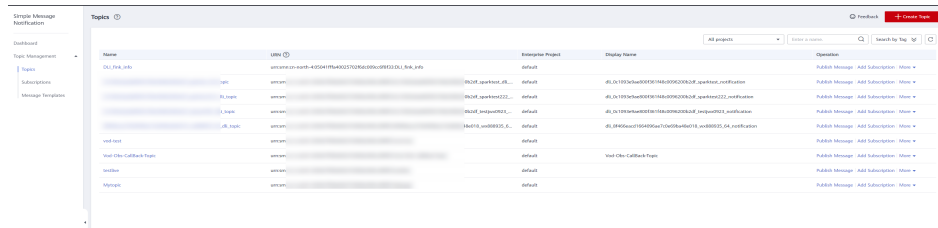
- You can select a single queue or all queues.
- If you create a topic for a queue and another topic for all queues, the SMN of all queues does not include the message of the single queue.
- After a message notification topic is created, you will receive a message notification only when a Spark job created on the subscription queue fails.

Figure 2-19 Successfully created a topic



3. Click **Topic Management** in to go to the **Topic Management** page of the SMN service.

Figure 2-20 Topic management page



4. In the **Operation** column of the topic, click **Add Subscription**. Select **Protocol** to determine the subscription mode.
 - If you select **SMS** for the subscription protocol, you need to enter the mobile number for receiving the confirmation SMS message in the **Endpoint** text box.
 - If you select **Email**, enter the email address for receiving the confirmation email in the **Endpoint** text box.

Figure 2-21 Adding a subscription

Add Subscription ×

Topic Name 0c1093e9ae800f361f48c0096200b2df_sparktest_dli_topic

* Protocol SMS

* Endpoint ? Endpoints Description

Endpoints	Description
<input type="text"/>	<input type="text"/>

+ Add Endpoint

OK Cancel

5. After you click the link in the SMS message or email, you will receive a message indicating that the subscription is successful.
6. Go to the **Subscriptions** page of SMN, and check that subscription status is **Confirmed**.

2.6.5 Managing Queue Tags

Tag Management

A tag is a key-value pair that you can customize to identify cloud resources. It helps you to classify and search for cloud resources. A tag consists of a tag key and a tag value.

If you use tags in other cloud services, you are advised to create the same tag (key-value pairs) for cloud resources used by the same business to keep consistency.

DLI supports the following two types of tags:

- Resource tags: non-global tags created on DLI
- Predefined tags: global tags created on Tag Management Service (TMS).
For more information about predefined tags, see [Tag Management Service User Guide](#).

DLI allows you to add, modify, or delete tags for queues.

- Step 1** In the navigation pane of the DLI management console, choose **Resources > Queue Management**.
- Step 2** In the **Operation** column of the queue, choose **More > Tags**.
- Step 3** The tag management page is displayed, showing the tag information about the current queue.
- Step 4** Click **Add/Edit Tag** to switch to the **Add/Edit Tag** dialog box. Enter a tag and a value, and click **Add**.

Figure 2-22 Adding/Editing tags

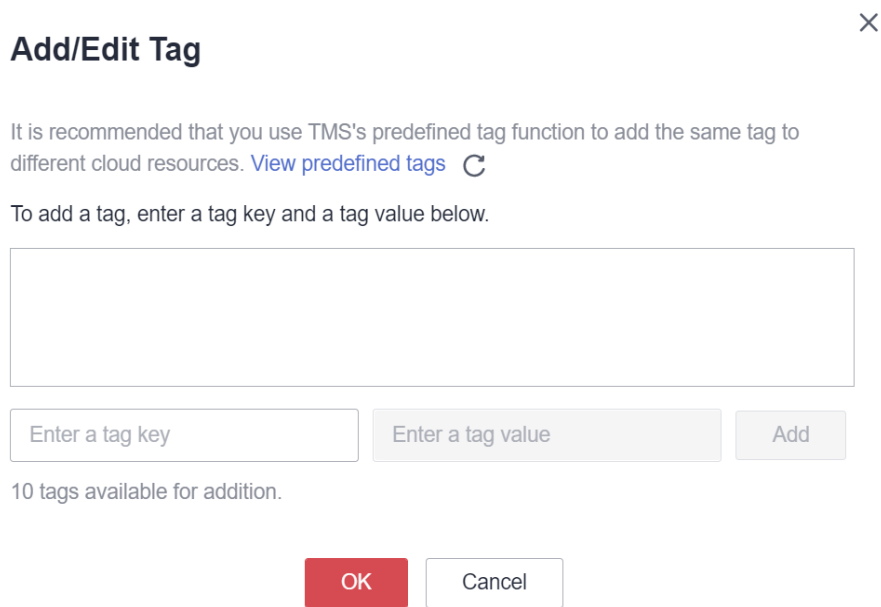


Table 2-11 Tag parameters

Parameter	Description
Tag key	<p>You can specify the tag key in either of the following ways:</p> <ul style="list-style-type: none"> Click the text box and select a predefined tag key from the drop-down list. To add a predefined tag, you need to create one on TMS and then select it from the Tag key drop-down list. You can click View predefined tags to go to the Predefined Tags page of the TMS console. Then, click Create Tag in the upper corner of the page to create a predefined tag. For details, see Creating Predefined Tags in <i>Tag Management Service User Guide</i>. Enter a tag key in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with _sys_.</p>
Tag value	<p>You can specify the tag value in either of the following ways:</p> <ul style="list-style-type: none"> Click the text box and select a predefined tag value from the drop-down list. Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed. The value cannot start or end with a space.</p>

 **NOTE**

- A maximum of 20 tags can be added.
- Only one tag value can be added to a tag key.
- The key name in each resource must be unique.

Step 5 Click **OK**.

Step 6 (Optional) To delete a tag, locate the row where the tag resides in the tag list and click **Delete** in the **Operation** column to delete the tag.

----End

2.6.6 Setting Queue Properties

Scenario

DLI allows you to set properties for queues.

You can set Spark driver parameters to improve the scheduling efficiency of queues.

This section describes how to set queue properties on the management console.

Notes and Constraints

- Only SQL queues of the Spark engine support configuring queue properties.
- Setting queue properties is only supported after the queue has been created.
- Currently, only queue properties related to the Spark driver can be set.
- Queue properties cannot be set in batches.
- For a queue in an elastic resource pool, if the minimum CUs of the queue is less than 16 CUs, both **Max. Spark Driver Instances** and **Max. Prestart Spark Driver Instances** set in the queue properties do not apply.

Procedure

1. In the navigation pane of the DLI management console, choose **Resources > Queue Management**.
2. In the **Operation** column of the queue, choose **More > Set Property**.
3. Go to the queue property setting page and set property parameters. For details about the property parameters, see [Table 2-12](#).

Table 2-12 Queue properties

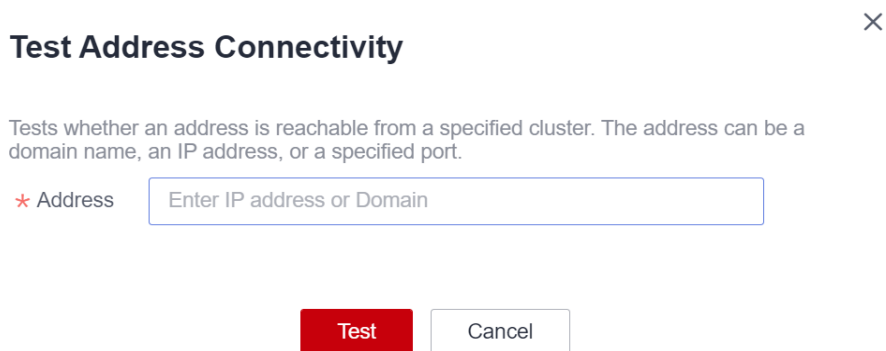
Property	Description	Value Range
Max. Spark Driver Instances	Maximum number of Spark drivers can be started on this queue, including the Spark driver that is prestarted and the Spark driver that runs jobs.	<ul style="list-style-type: none">For a 16-CU queue, the value is 2.For a queue that has more than 16 CUs, the value range is [2, queue CUs/16].If the minimum CUs of the queue is less than 16 CUs, this configuration item does not apply.
Max. Prestart Spark Driver Instances	Maximum number of Spark drivers can be prestarted on this queue. When the number of Spark drivers that run jobs exceeds the value of Max. Concurrency per Instance , the jobs are allocated to the Spark drivers that are prestarted.	<ul style="list-style-type: none">For a 16-CU queue, the value range is 0 to 1.For a queue that has more than 16 CUs, the value range is [2, queue CUs/16].If the minimum CUs of the queue is less than 16 CUs, this configuration item does not apply.
Max. Concurrency per Instance	Maximum number of jobs can be concurrently executed by a Spark driver. When the number of jobs exceeds the value of this parameter, the jobs are allocated to other Spark drivers.	1-32

4. Click **OK**.

2.6.7 Testing Address Connectivity

It can be used to test the connectivity between the DLI queue and the peer IP address specified by the user in common scenarios, or the connectivity between the DLI queue and the peer IP address bound to the datasource connection in datasource connection scenarios. The operations are as follows:

1. On the **Queue Management** page, locate the row containing the target queue, click **More** in the **Operation** column, and select **Test Address Connectivity**.
2. On the **Test Address Connectivity** page, enter the address to be tested. The domain name and IP address are supported, and the port number can be specified.

Figure 2-23 Testing address connectivity

Test Address Connectivity ×

Tests whether an address is reachable from a specified cluster. The address can be a domain name, an IP address, or a specified port.

* Address

Test Cancel

3. Click **Test**.

If the test address is reachable, a message is displayed on the page, indicating that the address is reachable.

If the test address is unreachable, the system displays a message indicating that the address is unreachable. Check the network configurations and retry. Network configurations include the VPC peering and the datasource connection. Check whether they have been activated.

2.6.8 Modifying Queue Specifications

Prerequisites

The specifications of a newly created yearly/monthly queue can be changed only when there were jobs running in this queue.

Precautions

- Currently, only the specifications of a yearly/monthly queue with more than 64 CUs can be changed.
- If **Status of queue xxx is assigning, which is not available** is displayed on the **Modify Queue Specifications** page, the specifications can be modified only after the queue resources are allocated.

Scaling Out

If the current queue specifications do not meet service requirements, you can add the number of CUs to scale out the queue.

 **NOTE**

Scale-out is time-consuming. After you perform scale-out on the **Modify Queue Specifications** page of DLI, wait for about 10 minutes. The duration is related to the CU amount to add. After a period of time, refresh the **Queue Management** page and check whether values of **Specifications** and **Actual CUs** are the same to determine whether the scale-out is successful. Alternatively, on the **Job Management** page, check the status of the **SCALE_QUEUE** SQL job. If the job status is **Scaling**, the queue is being scaled out.

The procedure is as follows:

1. On the left of the DLI management console, click **Resources > Queue Management**.
2. Select the queue to be scaled out and choose **More > Modify Queue Specifications** in the **Operation** column.
3. On the displayed page, select **Scale-out** for **Operation** and set the scale-out amount.
4. After confirming the fee, click **Submit**.

Scaling In

If the current queue specifications are too much for your computing service, you can reduce the number of CUs to scale in the queue.

NOTE

- Scale-in is time-consuming. After you perform scale-in on the **Modify Queue Specifications** page of DLI, wait for about 10 minutes. The duration is related to the CUs amount to reduce. After a period of time, refresh the **Queue Management** page and check whether values of **Specifications** and **Actual CUs** are the same to determine whether the scale-in is successful. Alternatively, on the **Job Management** page, check the status of the **SCALE_QUEUE** SQL job. If the job status is **Scaling**, the queue is being scaled in.
- The system may not fully scale in the queue to the target size. If the current queue is in use or the service volume of the queue is large, the scale-in may fail or only partial specifications may be reduced.
- By default, a queue runs with a minimum of 16 CUs. A queue with 16 CUs cannot be manually scaled in.

The procedure is as follows:

1. On the left of the DLI management console, click **Resources > Queue Management**.
2. Select the queue to be scaled in and choose **More > Modify Queue Specifications** in the **Operation** column.
3. On the displayed page, select **Scale-in** for **Operation** and set the scale-in amount.
4. After confirming the fee, click **Submit**.

2.6.9 Deleting a Queue

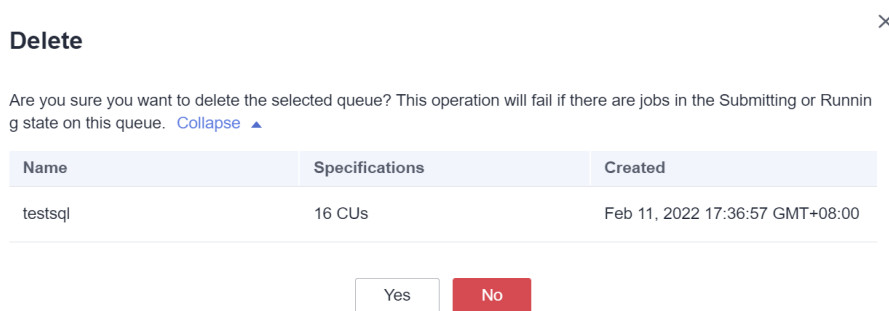
You can delete a queue based on actual conditions.

NOTE

- This operation will fail if there are jobs in the **Submitting** or **Running** state on this queue.
- Deleting a queue does not cause table data loss in your database.

Procedure

- Step 1** In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
- Step 2** Locate the row where the target queue locates and click **Delete** in the **Operation** column.

Figure 2-24 Deleting a Queue**NOTE**

If **Delete** in the **Operation** column is gray, the current user does not have the permission of deleting the queue. You can apply to the administrator for the permission.

Step 3 In the displayed dialog box, click **OK**.

----End

2.6.10 Elastic Scaling of Standard Queues (Discarded and Not Recommended)

Prerequisites

Newly created pay-per-use queues need to run jobs before they can be scaled in or out.

NOTE

The operations described in this section only apply to standard queues.

Notes and Constraints

- Queues with 16 CUs do not support scale-out or scale-in.
- Queues with 64 CUs do not support scale-in.
- Only queues whose billing mode is **Pay-per-use/By CUH** and **Pay-per-use/Dedicated resource mode** support elastic scaling.
- If **Status of queue xxx is assigning, which is not available** is displayed on the **Elastic Scaling** page, the queue can be scaled only after the queue resources are allocated.
- If there are not enough physical resources, a queue may not be able to scale out to the desired target size.
- The system does not guarantee that a queue will be scaled in to the desired target size. Typically, the system checks the resource usage before scaling in the queue to determine if there is enough space for scaling in. If the existing resources cannot be scaled in according to the minimum scaling step, the queue may not be scaled in successfully or only partially.

The scaling step may vary depending on the resource specifications, usually 16 CUs, 32 CUs, 48 CUs, 64 CUs, etc.

For example, if the queue size is 48 CUs and job execution uses 18 CUs, the remaining 30 CUs do not meet the requirement for scaling in by the minimum step of 32 CUs. If a scaling in task is executed, it will fail.

Scaling Out

If the current queue specifications do not meet service requirements, you can add the number of CUs to scale out the queue.

NOTE

Scale-out is time-consuming. After you perform scale-out on the **Elastic Scaling** page of DLI, wait for about 10 minutes. The duration is related to the CU amount to add. After a period of time, refresh the **Queue Management** page and check whether values of **Specifications** and **Actual CUs** are the same to determine whether the scale-out is successful. Alternatively, on the **Job Management** page, check the status of the **SCALE_QUEUE** SQL job. If the job status is **Scaling**, the queue is being scaled out.

The procedure is as follows:

1. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
2. Select the queue to be scaled out, click **More** in the **Operation** column, and select **Elastic Scaling**.
3. On the displayed page, select **Scale-out** for **Operation** and set the scale-out amount.

Figure 2-25 Scale-out

Name	sparktest
Billing Mode	Pay-per-use
Actual CUs	16 CUs
Opération	<input checked="" type="button" value="Scale-out"/> <input type="button" value="Scale-in"/>
Amount (CUs)	<input type="button" value="−"/> 16 <input type="button" value="+"/>
Final CU Count	32 CUs

4. Click OK.

Scaling In

If the current queue specifications are too much for your computing service, you can reduce the number of CUs to scale in the queue.

NOTE

- Scale-in is time-consuming. After you perform scale-in on the **Elastic Scaling** page of DLI, wait for about 10 minutes. The duration is related to the CU amount to reduce. After a period of time, refresh the **Queue Management** page and check whether values of **Specifications** and **Actual CUs** are the same to determine whether the scale-in is successful. Alternatively, on the **Job Management** page, check the status of the **SCALE_QUEUE** SQL job. If the job status is **Scaling**, the queue is being scaled in.
- By default, the minimum number of CUs is **16**. That is, when the queue specifications are **16 CUs**, you cannot scale in the queue.

The procedure is as follows:

1. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
2. Select the queue to be scaled in, click **More** in the **Operation** column, and select **Elastic Scaling**.
3. On the displayed page, select **Scale-in** for **Operation** and set the scale-in amount.

Figure 2-26 Manual scale-in

Name	createoutcuqueue
Billing Mode	Pay-per-use
Actual CUs	128 CUs
Opération	<input type="button" value="Scale-out"/> <input checked="" type="button" value="Scale-in"/>
Amount (CUs)	<input type="button" value="−"/> 16 <input type="button" value="+"/>
Final CU Count	112 CUs

4. Click OK.

2.6.11 Setting a Scheduled Elastic Scaling Task for a Standard Queue (Discarded and Not Recommended)

Scenario

When services are busy, you might need to use more compute resources to process services in a period. After this period, you do not require the same amount of resources. If the purchased queue specifications are small, resources may be insufficient during peak hours. If the queue specifications are large, resources may be wasted.

DLI provides scheduled tasks for elastic scale-in and -out in the preceding scenario. You can set different queue sizes (CUs) at different time or in different periods based on your service period or usage and the existing queue specifications to meet your service requirements and reduce costs.

 NOTE

The operations described in this section only apply to standard queues.

Precautions

- Newly created queues need to run jobs before they can be scaled in or out.
- Scheduled scaling tasks are available only for a queue with more than 64 CUs. That is, the minimum specifications of a queue are 64 CUs.
- A maximum of 12 scheduled tasks can be created for each queue.
- When each scheduled task starts, the actual start time of the specification change has a deviation of 5 minutes. It is recommended that the task start time be at least 20 minutes earlier than the time when the queue is actually used.
- The interval between two scheduled tasks must be at least 2 hours.
- Changing the specifications of a queue is time-consuming. The time required for changing the specifications depends on the difference between the target specifications and the current specifications. You can view the specifications of the current queue on the **Queue Management** page.
- If a job is running in the current queue, the queue may fail to be scaled in to the target CU amount value. Instead, it will be scaled in to a value between the current queue specifications and the target specifications. The system will try to scale in again 1 hour later until the next scheduled task starts.
- If a scheduled task does not scale out or scale in to the target CU amount value, the system triggers the scaling plan again 15 minutes later until the next scheduled task starts.

Creating Periodic Task

- If only scale-out or scale-in is required, you need to create only one task for changing specifications. Set the **Task Name**, **Final CU Count**, and **Executed** parameters. For details, see [Table 2-13](#).
- To set both scale-out and scale-in parameters, you need to create two periodic tasks, and set the **Task Name**, **Final CU Count**, and **Executed** parameters. For details, see [Table 2-13](#).

The procedure is as follows:

1. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
2. Locate the queue for which you want to schedule a periodic task for elastic scaling, and choose **More > Schedule CU Changes** in the **Operation** column.
3. On the displayed page, click **Create Periodic Task** in the upper right corner.
4. On the **Create Periodic Task** page, set the required parameters. Click **OK**.

Figure 2-27 Creating a periodic task

Create Periodic Task

* Task Name

Enable Task

Validity Period -

Total Number 32 CUs

* Final CU Count CUs

Repeat Select all
 Monday Tuesday Wednesday Thursday Friday Saturday
 Sunday

* Executed

Table 2-13 Parameters

Parameter	Description
Task Name	Enter the name of the periodic task. <ul style="list-style-type: none"> Only numbers, letters, and underscores (_) are allowed. The value cannot contain only numbers or start with an underscore (_) or be left unspecified. The value can contain a maximum of 128 characters.
Enable Task	Whether to enable periodic elastic scaling. The task is enabled by default. If disabled, the task will not be triggered on time.
Validity Period	Time segment for executing the periodic task. The options include Date and Time .
Actual CUs	Queue specifications before scale-in or scale-out.
Final CUs	Specifications after the queue is scaled in or out. <p>NOTE</p> <ul style="list-style-type: none"> By default, the maximum specifications of a queue are 512 CUs. The minimum queue specifications for scheduled scaling are 64 CUs. That is, only when Actual CUs are more than 64 CUs, the scheduled scaling can be performed. The value of Actual CUs must be a multiple of 16.

Parameter	Description
Repeat	<p>Time when scheduled scale-out or scale-in is repeat. Scheduled tasks can be scheduled by week in Repeat.</p> <ul style="list-style-type: none">• By default, this parameter is not configured, indicating that the task is executed only once at the time specified by Executed.• If you select all, the plan is executed every day.• If you select some options of Repeat, the plan is executed once a week at all specified days. <p>NOTE</p> <ul style="list-style-type: none">• You do not need to set this parameter if you only need to perform scale-in or scale-out once.• If you have set scaling, you can set Repeat as required. You can also set the repeat period together with the validity period.
Executed	<p>Time when scheduled scale-out or scale-in is performed</p> <ul style="list-style-type: none">• When each scheduled task starts, the actual start time of the specification change has a deviation of 5 minutes. It is recommended that the task start time be at least 20 minutes earlier than the time when the queue is actually used.• The interval between two scheduled tasks must be at least 2 hours.

After a periodic task is created, you can view the specification change of the current queue and the latest execution time on the page for scheduling CU changes.

Alternatively, on the **Queue Management** page, check whether the **Specifications** change to determine whether the scaling is successful.

You can also go to the **Job Management** page and check the status of the **SCALE_QUEUE** job. If the job status is **Scaling**, the queue is being scaled in or out.

Modifying a Scheduled Task

If a periodic task cannot meet service requirements anymore, you can modify it on the **Schedule CU Changes** page.

1. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
2. Locate the queue for which you want to schedule a periodic task for elastic scaling, and choose **More > Schedule CU Changes** in the **Operation** column.
3. On the displayed page, click **Modify** in the **Operation** column. In the displayed dialog box, modify the task parameters as needed.

Deleting a Scheduled Task

If you do not need the task anymore, delete the task on the **Schedule CU Changes** page.

1. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
2. Locate the queue for which you want to schedule a periodic task for elastic scaling, and choose **More > Schedule CU Changes** in the **Operation** column.
3. On the displayed page, click **Delete** in the **Operation** column. In the displayed dialog box, click **OK**.

2.6.12 Changing the CIDR Block of a Standard Queue (Discarded and Not Recommended)

If the CIDR block of the DLI queue conflicts with that of the user data source, you can change the CIDR block of the queue.

If the queue whose CIDR block is to be modified has jobs that are being submitted or running, or the queue has been bound to enhanced datasource connections, the CIDR block cannot be modified.

NOTE

The operations described in this section only apply to standard queues.

Procedure

NOTE

Currently, you can modify the CIDR block of for queues whose billing mode is **Yearly/Monthly** or **Pay-per-use/Dedicated resource mode**.

1. In the navigation pane on the left of the DLI management console, choose **Resources > Queue Management**.
2. Select the queue to be modified and click **Modify CIDR Block** in the **Operation** column.

Figure 2-28 Modifying a CIDR block

×

Modify CIDR Block

Name	testaaa
CIDR Block	<input type="text" value="172 . 16 . 0 . 0"/> / <input type="text" value="18"/>

If you need to use DLI enhanced datasource connections, the CIDR block entered here cannot be the same as that of the data source.
Recommended CIDR blocks:
10.0.0.0~10.255.0.0/8~24, 172.16.0.0~172.31.0.0/12~24, 192.168.0.0~192.168.0.0/16~24

OK Cancel

3. Enter the required CIDR block and click **OK**. After the CIDR block of the queue is successfully changed, wait for 5 to 10 minutes until the cluster to which the queue belongs is restarted and then run jobs on the queue.

Recommended CIDR block:
10.0.0.0~10.255.0.0/8~24
172.16.0.0~172.31.0.0/12~24
192.168.0.0~192.168.0.0/16~24

2.7 Example Use Case: Creating an Elastic Resource Pool and Running Jobs

This section walks you through the procedure of adding a queue to an elastic resource pool and binding an enhanced datasource connection to the elastic resource pool.

Figure 2-29 Process of creating an elastic resource pool

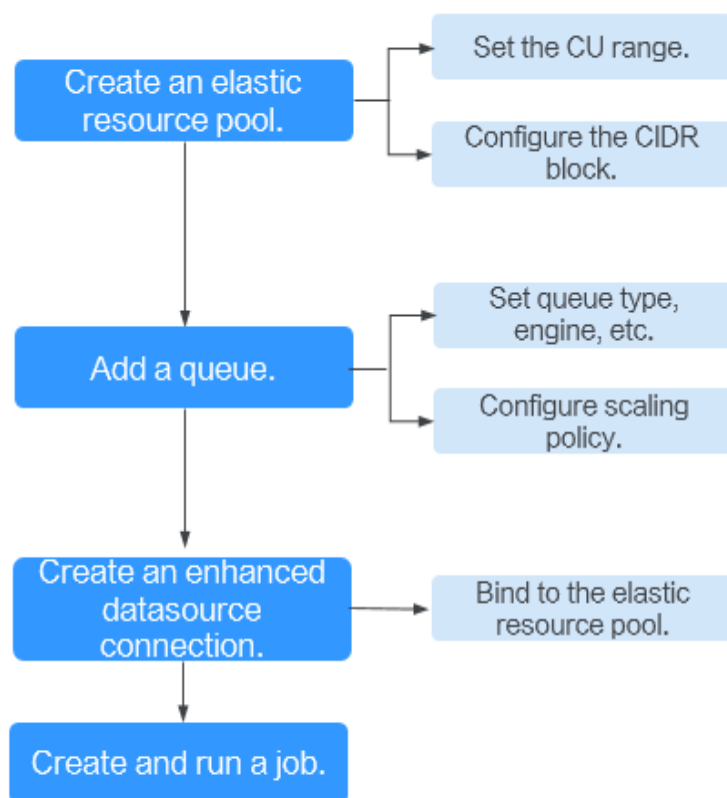


Table 2-14 Procedure

Step	Description	Reference
Create an elastic resource pool	Create an elastic resource pool and configure basic information, such as the billing mode, CU range, and CIDR block.	Creating an Elastic Resource Pool

Step	Description	Reference
Add a queue to the elastic resource pool	Add the queue where your jobs will run on to the elastic resource pool. The operations are as follows: <ol style="list-style-type: none">1. Set basic information about the queue, such as the name and type.2. Configure the scaling policy of the queue, including the priority, period, and the maximum and minimum CUs allowed for scaling.	Creating Queues in an Elastic Resource Pool Adjusting Scaling Policies for Queues in an Elastic Resource Pool
(Optional) Create an enhanced datasource connection.	If a job needs to access data from other data sources, for example, GaussDB(DWS) and RDS, you need to create a datasource connection. The created datasource connection must be bound to the elastic resource pool.	Creating an Enhanced Datasource Connection
Run a job.	Create and submit the job as you need.	Managing SQL Jobs Flink Job Management Overview Creating a Spark Job

Step 1: Create an Elastic Resource Pool

1. Log in to the DLI management console. In the navigation pane on the left, choose **Resources** > **Resource Pool**.
2. On the displayed page, click **Buy Resource Pool** in the upper right corner.
3. On the displayed page, set the following parameters:
 - **Name**: Enter the name of the elastic resource pool. For example, **pool_test**.
 - **CU range**: Minimum and maximum CUs of the elastic resource pool.
 - **CIDR Block**: Network segment of the elastic resource pool. For example, **172.16.0.0/18**.
 - Set other parameters as required.

Figure 2-30 Creating an elastic resource pool

The screenshot shows the 'Buy Resource Pool' configuration interface. At the top, there's a navigation bar with a back arrow and the title 'Buy Resource Pool'. Below this, there are several configuration sections:

- Billing Mode:** Two buttons, 'Yearly/Monthly' and 'Pay-per-use'. 'Pay-per-use' is selected.
- Region:** A dropdown menu. Below it, a note states: 'Resources are region-specific and cannot be used across regions through internal network connections. For low network latency and quick access, select the nearest region.'
- Project:** A dropdown menu.
- Name:** A text input field containing 'pool_test'.
- CU range:** Two input fields, both containing '64', with minus and plus buttons on either side. A help icon is to the right.
- Dedicated Resource Mode:** A checked checkbox with a help icon.
- Description:** A text area with a character count '0/256'.
- CIDR Block:** An input field containing '172.16.0.0' and a dropdown menu showing '/18'. Below this, a red note says 'An estimated 4096 CUs are available.' and a warning: 'This network segment cannot be changed after having been set. To use enhanced datasource connections, make sure that the network segment of the resource pool does not overlap with that of the datasources.' Below the warning, it lists 'Recommended CIDR blocks: 10.0.0.0-10.255.0.0/16-19,172.16.0.0-172.31.0.0/16-19,192.168.0.0-192.168.0.0/16-19'.
- Enterprise Project:** A dropdown menu showing 'default' and a 'Create Enterprise Project' link.

For details about how to create an elastic resource pool, see [Creating an Elastic Resource Pool](#).

4. Click **Buy**. Confirm the configuration and click **Pay**.
5. Go to the **Resource Pool** page to view the creation status. If the status is **Available**, the elastic resource pool is ready for use.

Step 2: Add a Queue to the Elastic Resource Pool

1. In the **Operation** column of the created elastic resource pool, click **Add Queue**.
2. Specify the basic information about the queue. The configuration parameters are as follows:
 - **Name:** Queue name
 - **Type:** Queue type In this example, select **For general purpose**.
 - For SQL:** The queue is used to run Spark SQL and Trino jobs.
 - For general purpose:** The queue is used to run Flink and Spark Jar jobs.
 - Set other parameters as required.

Figure 2-31 Creating a queue

The screenshot shows the 'Add Queue(pool_test)' configuration page. The 'Basic Configuration' step is active. The 'Name' field is 'general_test'. The 'Type' is 'For general purpose'. The 'Enterprise Project' is 'default'. The 'Description' field is empty. Below the description is a 'Tags' section with a text area and input fields for 'Enter a tag key' and 'Enter a tag value', and an 'Add' button. A note indicates '10 tags available for addition.'

3. Click **Next**. On the displayed page, set **Min CU** to **64** and **Max CU** to **64**.

Figure 2-32 Set scaling policy for the queue

The screenshot shows the 'Elastic Resources' step of the 'Add Queue(pool_test)' configuration. It displays a table for scaling policies. The table has columns for Priority, Period, Min CU, and Max CU. The values are set to 1, 00-24, 64, and 64 respectively. A 'Create' button is visible at the bottom.

Priority	Period	Min CU	Max CU
1	00 - 24	64	64

4. Click **OK**. The queue is added.

(Optional) Step 3: Create an Enhanced Datasource Connection

In this example, a datasource connection is required to connect to RDS. You need to create a datasource connection. If your job does not need to connect to an external data source, skip this step.

1. Log in to the RDS console and create an RDS DB instance. For details, see . Log in to the RDS instance, create a database and name it **test2**. For details, see [Buy a DB Instance](#).
2. Locate the row that contains the **test2** database, click **Query SQL Statements** in the **Operation** column. On the displayed page, enter the following statement to create table **tabletest2**. Click **Execute SQL**. The table creation statement is as follows:

```
CREATE TABLE `tabletest2` (  
  `id` int(11) unsigned,  
  `name` VARCHAR(32)  
) ENGINE = InnoDB DEFAULT CHARACTER SET = utf8mb4;
```
3. On the RDS console, choose **Instances** from the navigation pane. Click the name of a created RDS DB instance to view its basic information.

4. In the **Connection Information** pane, obtain the floating IP address, database port, VPC, and subnet.
5. Click the security group name. In the **Inbound Rules** tab, add a rule to allow access from the CIDR block of the elastic resource pool. For example, if the CIDR block of the elastic resource pool is **172.16.0.0/18** and the database port is **3306**, set the rule **Priority** to **1**, **Action** to **Allow**, **Protocol** to **TCP** and **Port** to **3306**, **Type** to **IPv4**, and **Source** to **172.16.0.0/18**.

Click **OK**. The security group rule is added.

6. Log in to the DLI management console. In the navigation pane on the left, choose **Datasource Connections**. On the displayed page, click **Create** in the **Enhanced** tab.
7. In the displayed dialog box, set the following parameters:
 - **Connection Name**: Name of the enhanced datasource connection
 - **Resource Pool**: Select the elastic resource pool created in [Step 1: Create an Elastic Resource Pool](#).

NOTE

If you cannot decide the elastic resource pool in this step, you can skip this parameter, go to the **Enhanced** tab, and click **More** > **Bind Resource Pool** in the **Operation** column of the row that contains this datasource connection after it is created.

- **VPC**: Select the VPC of the RDS DB instance obtained in [4](#).
- **Subnet**: Select the subnet of the RDS DB instance obtained in [4](#).
- Set other parameters as you need.

Click **OK**. Click the name of the created datasource connection to view its status. You can perform subsequent steps only after the connection status changes to **Active**.

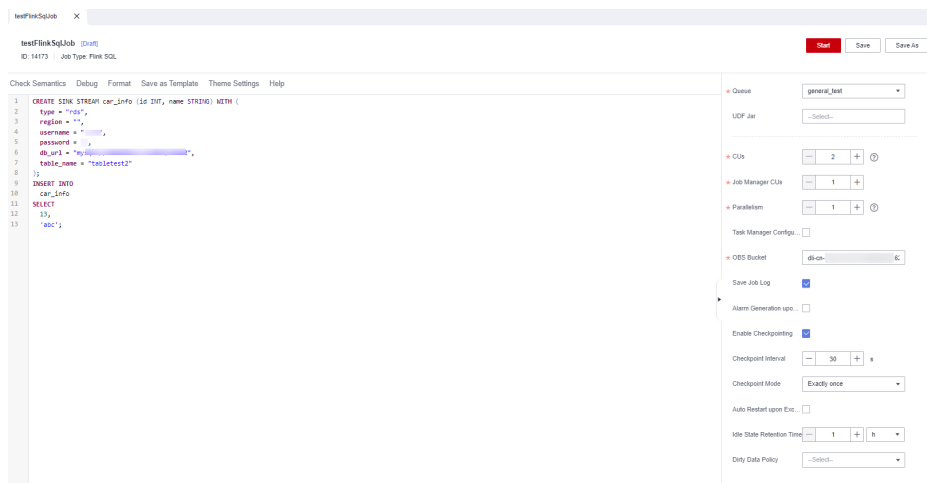
8. Click **Resources** > **Queue Management**, select the target queue, for example, **general_test**. In the **Operation** column, click **More** and select **Test Address Connectivity**.
9. In the displayed dialog box, enter *Floating IP address.Database port* of the RDS database in the **Address** box and click **Test** to check whether the database is reachable.

Step 4: Run a Job

Run a Flink SQL job on a queue in an elastic resource pool.

1. On the DLI management console, choose **Job Management** > **Flink Jobs**. On the **Flink Jobs** page, click **Create Job**.
2. In the **Create Job** dialog box, set **Type** to **Flink SQL** and **Name** to **testFlinkSqlJob**. Click **OK**.
3. On the job editing page, set the following parameters:

Figure 2-33 Creating a Flink SQL job

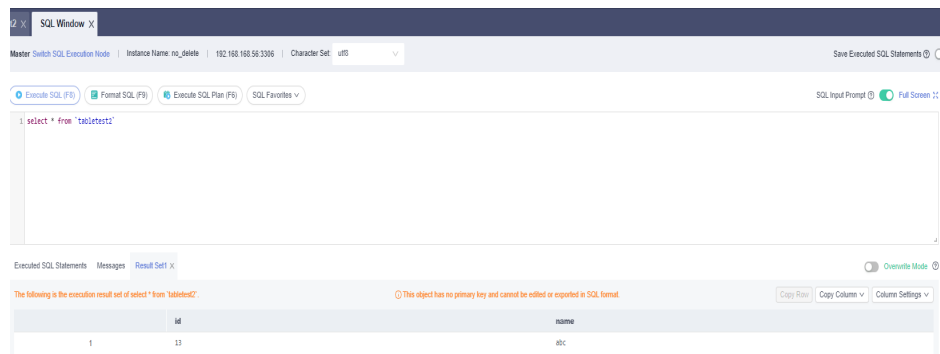


- **Queue:** Select the **general_test** queue added to the elastic resource pool in **Step 2: Add a Queue to the Elastic Resource Pool**.
- **Save Job Log:** Enable this function.
- **OBS Bucket:** Select an OBS bucket for storing job logs and grant access permissions of the OBS bucket as prompted.
- **Enable Checkpointing:** Enable this function.
- Enter the SQL statement in the editing pane. The following is an example. Modify the parameters in bold as you need.

```
CREATE SINK STREAM car_info (id INT, name STRING) WITH (  
  type = "rds",  
  region = "", /* Change the value to the current region ID. */  
  'pwd_auth_name'="xxxxx", // Name of the datasource authentication of the password type  
  created on DLI. If datasource authentication is used, you do not need to set the username and  
  password for the job.  
  db_url = "mysql://192.168.x.x:3306/test2", /* The format is mysql://floating IP address:port  
  number of the RDS database/database name. */  
  table_name = "tabletest2" /* Table name in RDS database */  
);  
INSERT INTO  
  car_info  
SELECT  
  13,  
  'abc';
```

4. Click **Check Semantic** and ensure that the SQL statement passes the check. Click **Save**. Click **Start**, confirm the job parameters, and click **Start Now** to execute the job.
5. Wait until the job is complete. The job status changes to **Completed**.
6. Log in to the RDS console, click the name of the RDS DB instance. On the displayed page, click the name of the created database, for example, **test2**, and click **Query SQL Statements** in the **Operation** column of the row that containing the **tabletest2** table.
7. On the displayed page, click **Execute SQL**. Check whether data has been written into the RDS table.

Figure 2-34 Query result



2.8 Example Use Case: Configuring Scaling Policies for Queues in an Elastic Resource Pool

Scenario

A company has multiple departments that perform data analysis in different periods during a day.

- Department A requires a large number of compute resources from 00:00 a.m. to 09:00 a.m. In other time segments, only small tasks are running.
- Department B requires a large number of compute resources from 10:00 a.m. to 10:00 p.m. Some periodical tasks are running in other time segments during a day.

In the preceding scenario, you can add two queues to an elastic resource pool: queue **test_a** for department A, and queue **test_b** for department B. You can add scaling policies for 00:00-09:00 and 10:00-23:00 respectively to the **test_a** and **test_b** queues. For jobs in other periods, you can modify the default scaling policy.

Table 2-15 Scaling policy

Queue	Period	Priority	CUs	Default Period	Default Priority	Default CUs	Remarks
test_a	[00:00, 09:00)	20	Minimum CU: 64 Maximum CU: 128	The time segments beyond [00:00, 09:00)	5	Minimum CU: 16 Maximum CU: 32	Jobs of department A

Queue	Period	Priority	CUs	Default Period	Default Priority	Default CUs	Remarks
test_b	[10:00, 23:00)	20	Minimum CU: 64 Maximum CU: 128	The time segments beyond [10:00, 23:00)	5	Minimum CU: 32 Maximum CU: 64	Jobs of department B

Precautions

- In any time segment of a day, the total minimum CUs of all queues in an elastic resource pool cannot be more than the minimum CUs of the pool.
- In any time segment of a day, the maximum CUs of any queue in an elastic resource pool cannot be more than the maximum CUs of the pool.
- The periods of scaling policies cannot overlap.
- The period of a scaling policy can only be set by hour and specified by the start time and end time. For example, if you set the period to **00-09**, the period when the policy takes effect is [00:00, 09:00). The period of the default scaling policy cannot be modified.
- In any period, compute resources are preferentially allocated to meet the minimum number of CUs of all queues. The remaining CUs (maximum CUs of the elastic resource pool – total minimum CUs of all queues) are allocated in accordance with the scaling policy priorities.
 - Scaling policies with smaller priority values are prior to over those with larger priority values.
 - If the scaling policies of two queues have the same priority, resources are randomly allocated to a queue. If there are remaining resources, they are randomly allocated until there is no more left.

Table 2-16 CU allocation

Scenario	CUs
<p>An elastic resource pool has a maximum number of 256 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none"> Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 128 Queue B: priority 10; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128 	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none"> The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are 160 CUs remaining. The remaining CUs are allocated based on the priorities. Queue B is prior to queue A. Therefore, queue B gets 64 CUs, and queue A has 96 CUs.
<p>An elastic resource pool has a maximum number of 96 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none"> Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64 Queue B: priority 10; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128 	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none"> The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are no remaining CUs. The allocation is complete.
<p>An elastic resource pool has a maximum number of 128 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none"> Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64 Queue B: priority 10; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128 	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none"> The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are 32 CUs remaining. The remaining 32 CUs are all preferentially allocated to queue B.
<p>An elastic resource pool has a maximum number of 128 CUs for queue A and queue B. The scaling policies are as follows:</p> <ul style="list-style-type: none"> Queue A: priority 5; period: 00:00–09:00; minimum CU: 32; maximum CU: 64 Queue B: priority 5; time period: 00:00–09:00; minimum CU: 64; maximum CU: 128 	<p>From 00:00 a.m. to 09:00 a.m.:</p> <ol style="list-style-type: none"> The minimum CUs are allocated to the two queues. Queue A has 32 CUs, and queue B has 64 CUs. There are 32 CUs remaining. The two queues have the same priority, the remaining 32 CUs are randomly allocated to the two queues.

Setting a Scaling Policy

- Step 1** Log in to the DLI management console and create an elastic resource pool. Set the minimum and maximum number of CUs of the pool to 128 and 256 respectively. For details, see [Creating an Elastic Resource Pool](#).
- Step 2** Choose **Resources > Resource Pool**. Locate the row that contains the created elastic resource pool, and click **Queue MGMT** in the **Operation** column.
- Step 3** Refer to [Creating Queues in an Elastic Resource Pool](#) to create the **test_a** queue and set the scaling policy.
- Set the priority of the default scaling policy to 5, **Min CU** to 16, and **Max CU** to 32.
 - Click create to add a scaling policy. Set the priority to 20, **Period** to 00--09, **Min CU** to 64, and **Max CU** to 128.

Figure 2-35 Adding a scaling policy for queue test_a

Priority	Period	Min CU	Max CU
5	00 -- 24	16	32
20	00 -- 09	64	128

- Step 4** View the scaling policy on the **Queue Management** page of the specific elastic resource pool.

Figure 2-36 Viewing the new scaling policy

Name	Type	Period	Min CUs	Max CUs	Priority	Engine	Created	Enterprise Project	Owner	Operation
test_a	For SQL	(00:00, 09:00) (00:00, 24:00)	64 16	128 32	20 5	spark	May 18, 2022 17:00:13 GMT+8	default		Edit Delete


Click  to view graphical statistics of priorities and CU settings for all time segments.

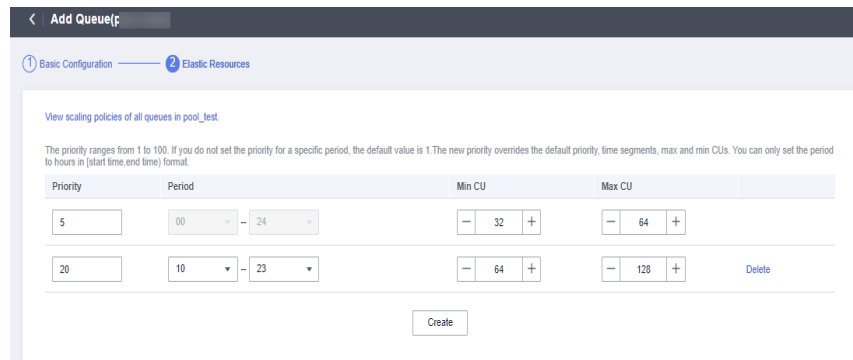
Figure 2-37 Scaling policies



Step 5 Refer to [Creating Queues in an Elastic Resource Pool](#) to create the **test_b** queue and set the scaling policy.

1. Set the priority of the default scaling policy to **5**, **Min CU** to **32**, and **Max CU** to **64**.
2. Click create to add a scaling policy. Set the priority to **20**, **Period** to **10--23**, **Min CU** to **64**, and **Max CU** to **128**.

Figure 2-38 Adding a scaling policy for queue test_b



Step 6 View the scaling policy on the **Queue Management** page of the specific elastic resource pool.

Figure 2-39 Viewing the new scaling policy

Name	Type	Period	Min CUs	Max CUs	Priority	Engine	Created	Enterprise Project	Owner	Operation
test_a	For SQL	[00:00, 08:00] [08:00, 24:00]	64	128	20	spark	May 18, 2022 17:47:40 GMT+08:00	default		Edit Delete
test_b	For SQL	[00:00, 24:00] [10:00, 23:00]	32	64	5	spark	May 18, 2022 17:46:59 GMT+08:00	default		Edit Delete


Click  to view graphical statistics on priorities and CU settings of the two queues for all time segments.

Figure 2-40 Scaling policies of both queues



----End

3 Job Management

3.1 Overview

DLI Job Type

DLI provides the following job types:

- **SQL job:** SQL jobs provide you with standard SQL statements and are compatible with Spark SQL and Presto SQL (based on Presto). You can query and analyze heterogeneous data sources on the cloud through visualized APIs, JDBC, ODBC, or Beeline. SQL jobs are compatible with mainstream data formats such as CSV, JSON, Parquet, Carbon, and ORC.
- **Flink job:** Flink jobs are real-time streaming big data analysis service jobs running on the public cloud. In full hosting mode, you only need to focus on Stream SQL services and execute jobs instantly without being aware of compute clusters. Flink jobs are fully compatible with Apache Flink APIs.
- **Spark job:** Spark jobs provide fully-managed Spark compute services. You can submit jobs through the GUI or RESTful APIs. Full-stack Spark jobs, such as Spark Core, DataSet, Streaming, MLLib, and GraphX jobs, are supported.

Constraints

- DLI supports the following types of jobs: Spark SQL, Spark Jar, Flink SQL, and Flink Jar.
- DLI supports the following Spark versions: Spark 3.3.1, Spark 3.1.1 (EOM), Spark 2.4.5 (EOM), and Spark 2.3 (EOS).
- DLI supports the following Flink versions: Flink Jar 1.15, Flink 1.12 (EOM), Flink 1.10 (EOS), and Flink 1.7 (EOS).
- SQL jobs support the Spark and Trino engines.
 - **Spark:** displays jobs whose execution engine is Spark.
 - **Trino:** displays jobs whose execution engine is Trino.
- SparkUI can only display the latest 100 jobs.
- A maximum of 1,000 job results can be displayed on the console. To view more or all jobs, export the job data to OBS.

- To export job run logs, you must have the permission to access OBS buckets. You need to configure a DLI job bucket on the **Global Configuration > Project** page in advance.
- The **View Log** button is not available for synchronization jobs and jobs running on the default queue.
- Only Spark jobs support custom images.
- An elastic resource pool supports a maximum of 32,000 CUs.
- Minimum CUs of a queue that can be created in an elastic resource pool:
 - General purpose queue: 4 CUs
 - SQL queue: Spark SQL queue: 8 CUs; Trino SQL queue: 16 CUs

3.2 Creating and Managing SQL Jobs

3.2.1 Submitting a SQL Job Using the SQL Editor

Introduction

You can edit and run SQL statements in the SQL job editor to execute data query.

The editor supports SQL:2003 and is compatible with Spark SQL. For details about the syntax, see [Data Lake Insight SQL Syntax Reference](#).

To access the SQL editor, choose **SQL Editor** in the left navigation pane of the DLI console, or click **Create Job** in the upper right corner of the **Job Management > SQL Jobs** page.

This topic describes the main functions of the SQL editor.

Notes

- If you access the SQL editor for the first time, the system prompts you to set a bucket for DLI jobs. The created bucket is used to store temporary data generated by DLI, such as job logs.

You cannot view job logs if you choose not to create the bucket. The bucket name will be set by the system.

On the OBS console, you can configure lifecycle rules for a bucket to periodically delete objects in it or change object storage classes.
- SQL statements can be executed in batches on the SQL editor page.
- Commonly used syntax in the job editing window is highlighted in different colors.
- Both single-line comment and multi-line comment are allowed. Use two consecutive hyphens (--) in each line to comment your statements.

Navigation pane

The navigation pane on the left consists of **Databases**, **Queues**, and **Templates** tabs.

Figure 3-1 SQL editor navigation pane

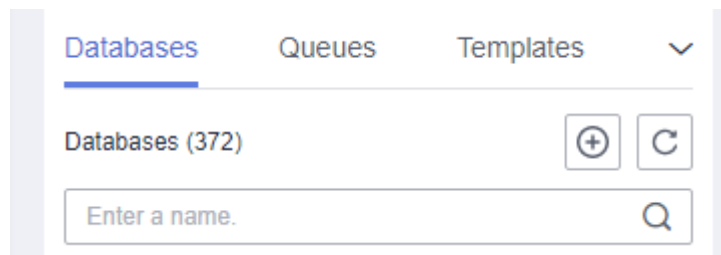


Table 3-1 Navigation pane buttons

No.	Name	Description
1	Databases	<p>Displays all the existing databases and tables in these databases.</p> <ul style="list-style-type: none"> Click a database name to view the tables in the database. Click a table name to view the metadata in the table. A maximum of 20 metadata records can be displayed. After you double-click a table name, a SQL query statement is automatically entered in the editing window.
2	Queues	Displays existing queues.
3	Templates	Click the drop-down button to view 22 built-in standard TPC-H query templates and custom templates.

SQL Editing Window

SQL job editing window is displayed in the upper right part of the page. For details about the parameters, see [Table 3-2](#).

The SQL statement editing area is below the operation bar. For details about keyboard shortcuts, see [Table 3-3](#).

Table 3-2 Components of the SQL job editing window

No.	Button & Drop-Down List	Description
1	Engine	<p>SQL jobs support the Spark and Trino engines.</p> <ul style="list-style-type: none"> Spark is suitable for offline analysis. Trino is suitable for interactive analysis.

N o.	Button & Drop- Down List	Description
2	Queues	Select a queue from the drop-down list box. If there is no queue available, the default queue is displayed. However, this queue is shared among all users for experience only. This may lead to resource contention, preventing you from obtaining the required resources for your jobs. So, you are advised to create your own queue for executing your jobs. For how to create a queue, see Creating an Elastic Resource Pool and Creating Queues in an Elastic Resource Pool . SQL jobs can only be executed on SQL queues.
3	Database	Select a database from the drop-down list box. If no database is available, the default database is displayed. For details about how to create a database, see Creating a Database or a Table . NOTE If you specify the database in the SQL statements, the database you choose from the drop-down list will not be used.
4	Execute	Click this button to run the SQL statements in the job editing window.
5	Format	Click this button to format the SQL statements.
6	Syntax Reference	Click this button to view the <i>Data Lake Insight SQL Syntax Reference</i> .
7	Settings	Add parameters and tags. Parameter Settings: Set parameters in key/value format for SQL jobs. Tags: Set tags in key/value format for SQL jobs.
8	More	The drop-down list includes the following options: <ul style="list-style-type: none">• Click Verify Syntax to check whether the SQL statements are correct.• Click Set as Template to set SQL statements as a template. For details, see Managing SQL Templates.• Click Change Theme to switch between dark and light modes.

Table 3-3 Keyboard shortcut description

Shortcut	Description
Ctrl+Enter	Execute SQL statements. You can run SQL statements by pressing Ctrl+R or Ctrl + Enter on the keyboard.

Shortcut	Description
Ctrl+F	Search for SQL statements. You can press Ctrl+F to search for a required SQL statement.
Shift+Alt+F	Format SQL statements. You can press Shift + Alt + F to format a SQL statement.
Ctrl+Q	Syntax verification. You can press Ctrl + Q to verify the syntax of SQL statements.
F11	Full screen. You can press F11 to display the SQL Job Editor window in full screen. Press F11 again to leave the full screen.

Executed Queries (Last Day) and View Result

After the SQL job is executed, you can view the execution history and result in the lower part of the editing area.

- **Executed Queries (Last Day)**

You can filter the execution history in the following ways:



- In the search box in the upper right corner of the **Executed Queries (Last Day)** pane, select a queue name or enter an execution statement in the search box.
- In the list, click the icon next to **Created** and choose **Ascending** or **Descending**.
- Select a job status from the **Status** list.

Table 3-4 Area description

Area	Description
<p>Executed Queries (Last Day)</p>	<p>The latest daily information about the submitted jobs, including the following items:</p> <ul style="list-style-type: none"> • Queues: Queue name • Username: User who executes the SQL statements • Type: Type of the SQL job • Status: Execution status of the SQL job • Query • Created • Operation <ul style="list-style-type: none"> - Edit: Edit the SQL statement. - SparkUI: Switch to the SparkUI page to view the SQL statement execution process. <p>NOTE</p> <p>When you execute a job on a created queue, the cluster is restarted. It takes about 10 minutes. If you click SparkUI before the cluster is created, an empty projectID will be cached. The SparkUI page cannot be displayed. You are advised to use a dedicated queue so that the cluster will not be released. Alternatively, wait for a while after the job is submitted (the cluster is created), and then check SparkUI.</p> <p>SparkUI can currently only display the latest 100 jobs.</p> <p>This function is not supported for synchronization jobs and jobs running on the default queue.</p> <ul style="list-style-type: none"> - More: The following operations vary depending on the SQL job types and running status. <ul style="list-style-type: none"> Cancel: Cancel a SQL job that is running or being submitted. Re-execute: Execute the SQL statement again. View Result: View the execution result of a QUERY job. Export Result: Export the execution results of a QUERY job to a specified OBS path. View Log: View the OBS path for storing SQL statement execution logs. <p>NOTE</p> <p>The View Log button is not available for synchronization jobs and jobs running on the default queue.</p>

- **View Result**

Table 3-5 Operations on the result tab

Operation	Description
Clear the result	Clear the displayed SQL statement query results.
View chart/ table	Click  to view the query result in a chart or table.
Export job results	Export the job execution results to the created OBS bucket.
Download result	Click  to download results of asynchronous DDL and QUERY statements to a local directory. For details, see Exporting Job Results to a Local Host .

SQL Query Procedure


1. Log in to the DLI management console. On the page displayed, choose **Job Management > SQL Jobs**. On the page displayed, click **Create Job**.



NOTE

On the SQL editor page, the system prompts you to create an OBS bucket to store temporary data generated by DLI jobs. In the **Set Job Bucket** dialog box, click **Setting**. On the page displayed, click the edit button in the upper right corner of the job bucket card. In the displayed **Set Job Bucket** dialog box, enter the job bucket path and click **OK**.

2. Select a queue from the queue list in the upper left corner of the SQL job editing window.
3. In the upper right corner of the SQL job editing window, select a database, for example, **qw**, from the **Databases** drop-down list.
4. Create a table, for example, **qw**. For details about how to create a database and table, see [Creating a Database or a Table](#).
5. In the SQL job editing window, enter the following SQL statement:

```
SELECT * FROM qw.qw LIMIT 10;
```

Alternatively, you can double-click the table name **qw**. The query statement is automatically entered in the SQL job editing window.
6. On top of the editing window, click **More > Verify Syntax** to check whether the SQL statement is correct.
 - a. If the verification fails, check the SQL statement syntax by referring to [Data Lake Insight SQL Syntax Reference](#).
 - b. If the syntax verification is successful, click **Execute**. Read and agree to the privacy agreement. Click **OK** to execute the SQL statement.
 - c. After the execution is complete, you can view the execution result in the area under the SQL job editing window.
7. (Optional) A maximum of 1000 records can be displayed in the query result on the current console. To view more or all data, click  to export the data to OBS.

- (Optional) In the **View Result** tab, click  to display the query result in a chart. Click  to switch back to the table view.

NOTE

- If no column of the numeric type is displayed in the execution result, the result cannot be represented in charts.
- You can view the data in a bar chart, line chart, or fan chart.
- In the bar chart and line chart, the X axis can be any column, while the Y axis can only be columns of the numeric type. The fan chart displays the corresponding legends and indicators.

Quickly Importing SQL Statements

- Double-click a table name in the navigation pane on the left to import the query statement of the selected table into the SQL statement editing window, and then click **Execute** to query.
- You can click **More** and choose **Save as Template** to save the SQL statement as a template for future use.

To use the SQL statement template, click **Templates** from the left pane of the SQL editor page. Double-click the required template in the template list, and modify it as required before executing the SQL statements.

3.2.2 Managing SQL Jobs

SQL jobs allow you to execute SQL statements in the SQL job editing window, import data, and export data.

SQL job management provides the following functions:


- Searching for Jobs:** Search for jobs that meet the search criteria.
- Viewing Job Details:** Display job details.
- Terminating a Job:** Stop a job in the **Submitting** or **Running** status.
- Exporting SQL Job Results:** Stores the data results of analyzed SQL jobs in a specified location in the desired format.

SQL Jobs Page

On the **Overview** page of the DLI console, click **SQL Jobs** to go to the SQL job management page. Alternatively, you can click **Job Management > SQL Jobs**. The job list displays all SQL jobs. If there are a large number of jobs, they will be displayed on multiple pages. You can switch to the specified page as needed. DLI allows you to view jobs in all statuses. By default, jobs in the job list are displayed in descending order of the job creation time.

Table 3-6 SQL Job management parameters

Parameter	Description
Queues	Name of the queue to which a job belongs

Parameter	Description
Engine	SQL jobs support the Spark and Trino engines. <ul style="list-style-type: none">● spark: displays jobs whose execution engine is spark.● trino: displays jobs whose execution engine is trino.
Username	Name of the user who executed the job.
Type	Job type. The following types are supported: <ul style="list-style-type: none">● IMPORT: A job that imports data to DLI● EXPORT: A job that exports data from DLI● DCL: Conventional DCLs and operations related to queue permissions● DDL: Conventional DDLs, including creating and deleting databases and tables● QUERY: A job that queries data by running SQL statements● INSERT: A job that inserts data by running SQL statements● UPDATE: A job that updates data.● DELETE: A job that deletes a SQL job.● DATA_MIGRATION: A job that migrates data.● RESTART_QUEUE: A job that restarts a queue.● SCALE_QUEUE: A job that changes queue specifications, including sale-out and scale-in.
Status	Job status. Possible values are as follows: <ul style="list-style-type: none">● Submitting● Running● Finished● Canceled● Failed● Scaling
Query	SQL statements for operations such as exporting and creating tables You can click  to copy the query statement.
Duration	Running duration of a job
Created	Time when a job is created. Jobs can be displayed in ascending or descending order of the job creation time.

Parameter	Description
Operation	<ul style="list-style-type: none"> ● Edit: Edit the job. ● Cancel <ul style="list-style-type: none"> - You can terminate a job only when the job is in Submitting or Running status. - A job whose status is Finished, Failed, or Canceled cannot be terminated. - If the Cancel button is gray, you are not allowed to perform this operation. ● Re-execute: Execute the job again. ● SparkUI: Display the Spark job execution page. <p>NOTE</p> <ul style="list-style-type: none"> ● When you execute a job on a created queue, the cluster is restarted. It takes about 10 minutes. If you click SparkUI before the cluster is created, an empty projectID will be cached. The SparkUI page cannot be displayed. You are advised to use a dedicated queue so that the cluster will not be released. Alternatively, wait for a while after the job is submitted (the cluster is created), and then check SparkUI. ● SparkUI can currently only display the latest 100 jobs. <ul style="list-style-type: none"> ● In addition to the preceding operations, the following operations are available for QUERY jobs and asynchronous DDL jobs. <ul style="list-style-type: none"> - Submit Download Request: Download the results of asynchronous DDL and QUERY statements to a local directory. For details, see Exporting Job Results to a Local Host. - View Result: View the job running result. - Export Result: Export the job running result to the created OBS bucket. ● In addition to the preceding operations, the EXPORT job also includes the following operations: <ul style="list-style-type: none"> - Download ● View Log: Save job logs to the temporary OBS bucket created by DLI. <p>NOTE</p> <p>The View Log button is not available for synchronization jobs and jobs running on the default queue.</p>


Searching for a Job

On the **SQL Jobs** page, you can search jobs with any of the following operations.

- Select a queue name.
- Select an engine.
- Set the date range.

- Enter a username, statement, tag, or job ID.
- Select the creation time in ascending or descending order.
- Select a job type.
- Select a job status.
- Select the job execution duration in ascending or descending order.

Viewing Job Details

On the **SQL Jobs** page, you can click  in front of a job record to view details about the job.

Job details vary with job types. The job details vary depending on the job types, status, and configuration options. The following describes how to load data, create a table, and select a job. For details about other job types, see the information on the management console.

- **Load data** (job type: IMPORT) include the following information: queue, job ID, username, type, status, execution statement, running duration, creation time, end time, parameter settings, label, number of results, scanned data, number of scanned data, number of error records, storage path, data format, database, table, table header, separator, reference character, escape character, date format, timestamp format, total CPU used, and output bytes.
- **Create table** (job type: DDL) include the following information: queue, job ID, username, type, status, execution statement, running duration, creation time, end time, parameter settings, tags, number of results, scanned data, and database.
- **Select** (job type: QUERY) include the following information: queue, job ID, username, type, status, execution statement, running duration, creation time, end time, parameter setting, label, number of results (results of successful executions can be exported), and scanned data, username, result status (results of successful tasks can be viewed. Failure causes of failed tasks are displayed), database, total CPU used, and output bytes.

NOTE

- **Total CPU Used (Core x ms)**: total CPU used during job execution.
- **Output Bytes**: number of output bytes after the job is executed.

Terminating a Job

On the **SQL Jobs** page, you can click **Terminate** in the **Operation** column to stop a submitting or running job.

3.2.3 Exporting SQL Job Results

Store the data results of analyzed SQL jobs in a specified location in the desired format.

By default, DLI stores SQL job results in its job bucket. You can also download job results to a local host or export job results to a specified OBS bucket.

Exporting Job Results to the DLI Job Bucket

DLI specifies a default OBS bucket for storing job results. You can configure the bucket information on the **Global Configuration > Project** page of the DLI management console. Once a job is complete, the system automatically stores its results to this bucket.

The following conditions must be met if you want to read job results from the DLI job bucket:

- You have configured the job bucket on the **Global Configuration > Project** page of the DLI management console by referring to [Configuring a DLI Job Bucket](#).
- You have submitted a service ticket to request the whitelisting of the feature that allows writing job results to buckets.
- The user who executes jobs has been granted read and write permissions either on the job bucket or on the **jobs/result** path of the job bucket.

For details, see [Creating a Custom Bucket Policy](#).


For how to obtain job results from the DLI job bucket, see "Object Management" > [Downloading Objects](#) in *Object Storage Service User Guide*.

Exporting Job Results to a Specified Location in Another Bucket

In addition to storing job results in the default bucket, you can also export them to a specified location in another bucket, increasing the flexibility of job result management and making it easier to organize and manage them.

On the console, you can only view a maximum of 1,000 job results. To view additional results, you can export them to an OBS path. The procedure is as follows:

You can export job results on either the **SQL Jobs** or the **SQL Editor** page.

- **SQL Jobs** page: In the navigation pane on the left, choose **Job Management > SQL Jobs**. On the displayed page, locate the row containing a desired job, click **More** in the **Operation** column, and select **Export Result**.
- **SQL Editor** page: In the navigation pane on the left, choose **SQL Editor**. On the displayed page, once query statements are successfully executed, click  next to the **View Result** tab to export job results.

NOTE

- If there are no numerical columns in the query results, job results cannot be exported.
- Ensure that the user who exports job results has the read and write permissions on the OBS bucket.

Table 3-7 Parameters

Parameter	Mandatory	Description
Data Format	Yes	Choose a data format for the job results you want to export. The options include json and csv .
Queues	Yes	Select the queue where the job is executed. SQL jobs can only be executed on SQL queues.
Compression Format	No	Compression format of the data to be exported. The options are: <ul style="list-style-type: none"> • none • bzip2 • deflate • gzip
Storage Path	Yes	<p>Path in an OBS bucket where the job results are exported</p> <ul style="list-style-type: none"> • If Export Mode is set to New OBS directory, then You need to manually enter a directory name and ensure that the directory name does not exist. Otherwise, the system returns an error message and the export operation cannot be performed. <p>NOTE The folder name cannot contain special characters (\/:*?"<>) and cannot start or end with a period (.).</p> <p>For example, after selecting the storage path obs://bucket/src1/, you need to manually enter a directory name to change the path to obs://bucket/src1/src2/ and ensure that the src2 directory name does not exist under src1.</p> <p>The job result export path is obs://bucket/src1/src2/test.csv.</p> <ul style="list-style-type: none"> • If Export Mode is set to Existing OBS directory (Overwritten), then After selecting a bucket path, the job results are exported to that path. If there are files with the same name, they will be automatically overwritten. <p>For example, if you select obs://bucket/src1/ as the bucket path, then The job result export path is obs://bucket/src1/test.csv.</p>

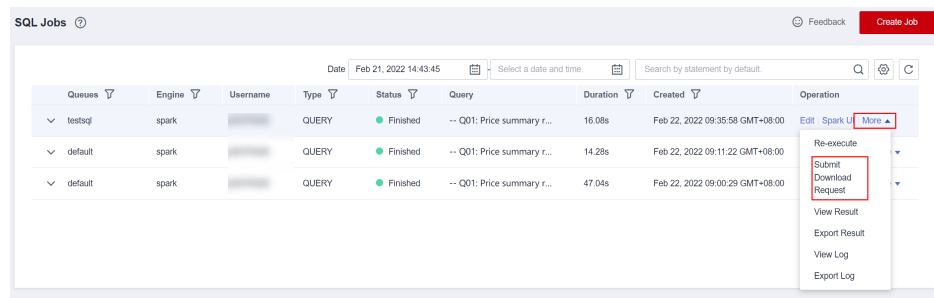
Parameter	Mandatory	Description
Export Mode	Yes	<ul style="list-style-type: none">• New OBS directory If you select this mode, a new folder path is created and the job results are saved to this path. This mode is used when you need to save exported results to a new location, making it easier to manage and track job results. If you select this option, you must manually enter an export directory in Storage Path and ensure that the directory must not exist. If the directory already exists, the system displays an error message and the export operation cannot be performed.• Existing OBS directory (Overwritten): When exporting job results, you can choose an existing file path as the output directory. If there is a file with the same name in that path, it will be automatically overwritten by the new exported job result file. This mode is used when you only need to save a single job result file in the same path, and you do not need to keep old job results.
Number of Results	No	Number of results to be exported If you do not specify or set it to 0 , all results will be exported.
Table Header	No	Whether the job results to be exported contain table headers

Exporting Job Results to a Local Host

You can download the results of asynchronous DDL and QUERY statements to a local directory. By default, you can download a maximum of 1,000 data records to a local host.

The procedure is as follows:

1. Locate the row containing a desired job whose asynchronous DDL or QUERY statement has been successfully executed, click **More** in the **Operation** column, and select **Submit Download Request**. In the displayed dialog box, click **OK**. After a few seconds, the **Submit Download Request** button would change to **Download**.

Figure 3-2 Selecting Submit Download Request

2. Click **Download** to download the results to your local host.

3.3 Creating and Managing Flink Jobs

3.3.1 Flink Job Management Overview

On the Job Management page of Flink jobs, you can submit a Flink job. Currently, the following job types are supported:

- **Flink OpenSource SQL** is compatible with the SQL syntax of the community Flink 1.10 version and can run on CCE queues only.
- **Flink SQL** uses SQL statements to define jobs and can be submitted to any general purpose queue.
- **Flink Jar**: custom Jar job based on Flink APIs.

Flink job management provides the following functions:

- [Managing Flink Job Permissions](#)
- [Creating a Flink SQL Job](#)
- [Creating a Flink Jar Job](#)
- [Editing a job](#)
- [Starting a Job](#)
- [Stopping a Job](#)
- [Deleting a Job](#)
- [Exporting a Job](#)
- [Importing a Job](#)
- [Modifying Name and Description](#)
- [Importing to a Savepoint](#)
- [Triggering a Savepoint](#)
- [Runtime Configuration](#)
- [Job Details](#)
- [Tag Management](#)

Flink Jobs Page

On the **Overview** page, click **Flink Jobs** to go to the Flink job management page. Alternatively, you can choose **Job Management > Flink Jobs** from the navigation

pane on the left. The page displays all Flink jobs. If there are a large number of jobs, they will be displayed on multiple pages. DLI allows you to view jobs in all statuses.

Table 3-8 Job management parameters

Parameter	Description
ID	ID of a submitted Flink job, which is generated by the system by default.
Name	Name of the submitted Flink job.
Type	Type of the submitted Flink job. Including: <ul style="list-style-type: none">• Flink SQL: Flink SQL jobs• Flink Jar: Flink Jar jobs• Flink OpenSource SQL: Flink OpenSource SQL jobs
Status	Job statuses, including: <ul style="list-style-type: none">• Draft• Submitting• Submission failed• Running: The billing starts. After the job is submitted, a normal result is returned.• Running exception: The billing stops. The job stops running due to an exception.• Downloading• Idle• Stopping• Stopped• Stopping failed• Creating the savepoint• Stopped due to arrears: The billing ends. The job is stopped because the user account is in arrears.• Restoring (recharged jobs): The account in arrears is recharged, and the job is being restored.• Completed
Description	Description of the submitted Flink job.
Username	Name of the user who submits a job.
Created	Time when a job is created.
Started	Time when a Flink job starts to run.
Duration	Time consumed by job running.

Parameter	Description
Operation	<ul style="list-style-type: none">● Edit: Edit a created job. For details, see Editing a Job.● Start: Start and run a job. For details, see Starting a Job.● More<ul style="list-style-type: none">– FlinkUI: After you click this button, the Flink job execution page is displayed. NOTE When you execute a job on a created queue, the cluster is restarted. It takes about 10 minutes. If you click FlinkUI before the cluster is created, an empty projectID will be cached. The FlinkUI page cannot be displayed. You are advised to use a dedicated queue so that the cluster will not be released. Alternatively, wait for a while after the job is submitted (the cluster is created), and then check FlinkUI.– Stop: Stop a Flink job. If this function is unavailable, jobs in the current status cannot be stopped.– Delete: Delete a job. NOTE A deleted job cannot be restored.– Modify Name and Description: You can modify the name and description of a job. For details, see Modifying Name and Description.– Import Savepoint: Import the data exported from the original CS job. For details, see Importing to a Savepoint.– Trigger Savepoint: You can click this button for jobs in the Running status to save the job status. For details, see Triggering a Savepoint.– Permissions: You can view the user permissions corresponding to the job and grant permissions to other users. For details, see Managing Flink Job Permissions.– Runtime Configuration: You can enable Alarm Generation upon Job Exception and Auto Restart upon Exception. For details, see Runtime Configuration.

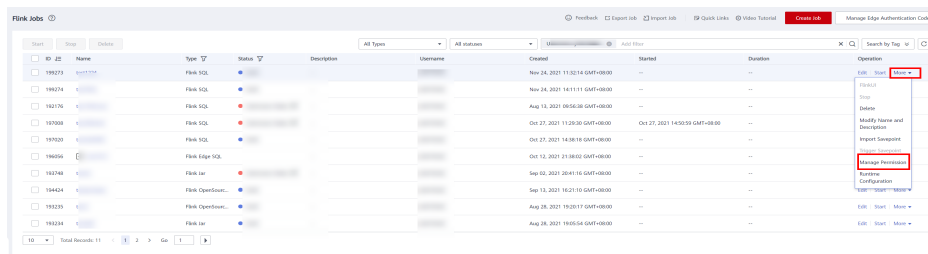
3.3.2 Managing Flink Job Permissions

Scenario

- You can isolate Flink jobs allocated to different users by setting permissions to ensure data query performance.
- The administrator and job creator have all permissions, which cannot be set or modified by other users.
- When setting job permissions for a new user, ensure that the region of the user group to which the user belongs has the **Tenant Guest** permission.

Flink Job Permission Operations

1. On the left of the DLI management console, choose **Job Management > Flink Jobs**.
2. Select the job to be configured and choose **More > Permissions** in the **Operation** column. The **User Permissions** area displays the list of users who have permissions on the job.



You can assign queue permissions to new users, modify permissions for users who have some permissions of a queue, and revoke all permissions of a user on a queue.

- Assign permissions to a new user.
 - A new user does not have permissions on the job.
 - i. Click **Grant Permission** on the right of **User Permissions** page. The **Grant Permission** dialog box is displayed.
 - ii. Specify **Username** and select corresponding permissions.
 - iii. Click **OK**.

Table 3-9 describes the related parameters.

Figure 3-3 Granting permissions

Grant Permission

* Username

Select the permissions to be granted to the user

<input type="checkbox"/> Select all	<input type="checkbox"/> Modify Jobs	<input type="checkbox"/> Delete Jobs
<input type="checkbox"/> Get Job Details	<input type="checkbox"/> Stop job	<input type="checkbox"/> Export Job
<input type="checkbox"/> Start job	<input type="checkbox"/> Revoke Permission	<input type="checkbox"/> View Other User's Permissions
<input type="checkbox"/> Grant Permission		

Table 3-9 Permission parameters

Parameter	Description
Username	Name of the user you want to grant permissions to. NOTE This username must be an existing IAM username. In addition, the user can perform authorization operations only after logging in to the Huawei Cloud platform.

Parameter	Description
Permissions to be granted to the user	<ul style="list-style-type: none"> ● Select all: All permissions are selected. ● View Job Details: This permission allows you to view the job details. ● Modify Job: This permission allows you to modify the job. ● Delete Job: This permission allows you to delete the job. ● Start Job: This permission allows you to start the job. ● Stop Job: This permission allows you to stop the job. ● Export Job: This permission allows you to export the job. ● Grant Permission: This permission allows you to grant job permissions to other users. ● Revoke Permission: This permission allows you to revoke the job permissions that other users have but cannot revoke the job creator's permissions. ● View Other User's Permissions: This permission allows you to view the job permissions of other users.

- To assign or revoke permissions of a user who has some permissions on the job, perform the following steps:
 - i. In the list under **User Permissions** for a job, select the user whose permissions need to be modified and click **Set Permission** in the **Operation** column.
 - ii. In the displayed **Set Permission** dialog box, modify the permissions of the current user. [Table 3-9](#) lists the detailed permission descriptions.
If all options under **Set Permission** are gray, you are not allowed to change permissions on this job. You can apply to the administrator, job creator, or other authorized users for job permission granting and revoking.
 - iii. Click **OK**.
- To revoke all permissions of a user on a job, perform the following steps:
In the list under **User Permissions** for a job, locate the user whose permissions need to be revoked, click **Revoke Permission** in the **Operation** column, and click **Yes**. After this operation, the user does not have any permission on the job.

Flink Job Permissions

- **View Job Details**
 - Tenants and the admin user can view and operate all jobs.

- Subusers and users with the read-only permission can only view their own jobs.

 NOTE

If another user grants any permission other than the job viewing permission to a subuser, the job is displayed in the job list, but the details cannot be viewed by the subuser.

- **Start Job**

You must have the permission to submit and start jobs.

- **Stop Job**

You must have the permission to stop queues and jobs.

- **Delete Job**

- If a job can be deleted, you can delete the job if you were granted this permission.
- If a job cannot be deleted, the system stops the job before you delete it. For details about how to stop a job, see [Stop Job](#). In addition, you must have the permission to delete the job.

- **Create Job**

- By default, sub-users cannot create jobs.
- To create a job, you must have this permission. Currently, only the admin user has the permission to create jobs. In addition, the user must have the permission of the related package group or package used by the job.

- **Modify Job**

When modifying a job, you need to have the permission to update the job and the permission to the package group or package used by the job belongs.

3.3.3 Preparing Flink Job Data

To create a Flink job, you need to enter the data source and data output channel, that is, source and sink. To use another service as the source or sink stream, you need to apply for the service first.

Flink jobs support the following data sources and output channels:

- DIS as the data input and output channel

To use DIS as the data source and output channel, you need to enable DIS first.

For details about how to create a DIS stream, see **Creating a DIS Stream** in the *Data Ingestion Service User Guide*.

After applying for a DIS stream, you can upload local data to DIS to provide data sources for Flink jobs in real time. For details, see **Sending Data to DIS** in the *Data Ingestion Service User Guide*.

An example is provided as follows:

```
1,lilei,bmw320i,28  
2,hanmeimei,audia4,27
```

- OBS as the data source

After you enable OBS, upload local files to OBS using the Internet. For detailed operations, see **Uploading a File** in the *Object Storage Service Console Operation Guide*.

- RDS as the output channel
To use RDS as the output channel, create an RDS instance. For details, see **Creating a DB Instance** in the *Relational Database Service User Guide*.
- SMN as the output channel
To use SMN as the output channel, create an SMN topic to obtain the URN resource ID and then add topic subscription. For detailed operations, see **Getting Started** in the *Simple Message Notification User Guide*.
- Kafka as the data input and output channel
If Kafka serves as both the source and sink streams, create an enhanced datasource connection between Flink jobs and Kafka. For details, see **Enhanced Datasource Connections**.
If the port of the Kafka server is listened on by the host name, you need to add the mapping between the host name and IP address of the Kafka Broker node to the datasource connection.
- CloudTable as the data input and output channel
To use CloudTable as the data input and output channel, create a cluster in CloudTable and obtain the cluster ID.
- CSS as the output channel
To use CSS as the data output channel, create a cluster in CSS and obtain the cluster's private network address. For details, see **Getting Started** in the *Cloud Search Service User Guide*.
- DCS as the output channel
To use DCS as the output channel, create a Redis cache instance in DCS and obtain the address used for Flink jobs to connect to the Redis instance. For details, see "Buying a DCS Redis Instance" in *Distributed Cache Service User Guide*.

3.3.4 Creating a Flink OpenSource SQL Job

This section describes how to create a Flink OpenSource SQL job.

DLI Flink OpenSource SQL jobs are fully compatible with the syntax of Flink provided by the community. In addition, Redis, GaussDB(DWS), and DIS data source types are added based on the community connector. For the syntax and constraints of Flink SQL DDL, DML, and functions, see **Table API & SQL**.

Prerequisites

- You have prepared the source and sink streams. For details, see **Preparing Flink Job Data**.
- Before creating a Flink OpenSource SQL job to access other external data sources, such as OpenTSDB, HBase, Kafka, GaussDB(DWS), RDS, CSS, CloudTable, DCS Redis, and DDS, you need to create a datasource connection to connect the job running queue to the external data source.
 - For details about the external data sources that can be accessed by Flink jobs, see **Cross-Source Analysis Development Methods**.
 - For details about how to create a datasource connection, see **Enhanced Datasource Connections**.

On the **Resources > Queue Management** page, locate the queue you have created, click **More** in the **Operation** column, and select **Test**

Address Connectivity to check if the network connection between the queue and the data source is normal. For details, see [Testing Address Connectivity](#).

Precautions

Before creating jobs and submitting tasks, you are advised to enable CTS to record DLI operations for queries, audits, and tracking. [DLI Operations That Can Be Recorded by CTS](#) lists DLI operations that can be recorded by CTS.

For details about how to enable CTS and view trace details, see [Cloud Trace Service Getting Started](#).

Creating a Flink OpenSource SQL Job

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 In the upper right corner of the **Flink Jobs** page, click **Create Job**.

Figure 3-4 Creating a Flink OpenSource SQL job

Create Job ×

Type

* Name

Description

Template Name

Tags

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#)

To add a tag, enter a tag key and a tag value below.

10 tags available for addition.

Step 3 Specify job parameters.

Table 3-10 Job configuration information

Parameter	Description
Type	Set Type to Flink OpenSource SQL . You will need to start jobs by compiling SQL statements.
Name	Job name. Enter 1 to 57 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed. NOTE The job name must be globally unique.
Description	Description of a job. It can contain a maximum of 512 characters.
Template Name	You can select a sample template or a custom job template. For details about templates, see Managing Flink Templates .
Tags	Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS). For details, see Tag Management Service User Guide . NOTE <ul style="list-style-type: none">A maximum of 20 tags can be added.Only one tag value can be added to a tag key.The key name in each resource must be unique. <ul style="list-style-type: none">Tag key: Enter a tag key name in the text box. NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_ :=+ - @) are allowed, but the value cannot start or end with a space or start with <code>_sys_</code>.Tag value: Enter a tag value in the text box. NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_ :=+ - @) are allowed. The value cannot start or end with a space.

Step 4 Click **OK** to enter the editing page.

Step 5 Edit an OpenSource SQL job.

Enter detailed SQL statements in the statement editing area. For details about SQL statements, see the *Data Lake Insight Flink OpenSource SQL Syntax Reference*.

Step 6 Click **Check Semantics**.

- You can **Start** a job only after the semantic verification is successful.
- If verification is successful, the message "The SQL semantic verification is complete. No error." will be displayed.

- If verification fails, a red "X" mark will be displayed in front of each SQL statement that produced an error. You can move the cursor to the "X" mark to view error details and change the SQL statement as prompted.

Step 7 Set job running parameters.

Figure 3-5 Setting running parameters for Flink OpenSource SQL

The screenshot displays the configuration panel for a Flink job. Key settings include:

- Queue:** Shared queue (Free for O...)
- CUs:** 2
- Job Manager CUs:** 1
- Parallelism:** 1
- OBS Bucket:** dl-
- Save Job Log:** Checked
- Enable Checkpointing:** Checked
- Checkpoint Interval:** 30 s
- Checkpoint Mode:** Exactly once
- Idle State Retention Time:** 1 h

Table 3-11 Running parameters

Parameter	Description
Queue	<p>Select a queue where you want to run your job and set the following parameter:</p> <p>UDF Jar: You can customize a UDF Jar file. Before selecting a JAR file to be inserted, upload the corresponding JAR file to the OBS bucket and choose Data Management > Package Management to create a package. For details, see Creating a Package.</p> <p>In SQL, you can call a user-defined function that is inserted into a JAR file.</p> <p>NOTE</p> <p>During job creation, a sub-user can only select a queue that has been allocated to the user.</p> <p>If the remaining capacity of the selected queue cannot meet the job requirements, the system automatically scales up the capacity and you will be billed based on the increased capacity. When a queue is idle, the system automatically scales in the queue.</p>

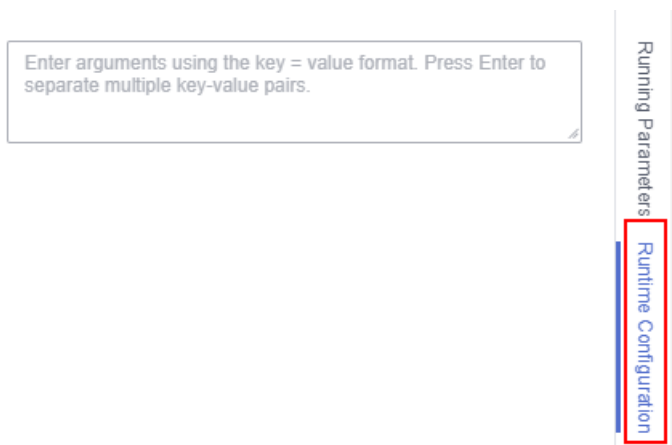
Parameter	Description
Agency	If you choose Flink 1.15 to execute your job, you can create a custom agency to allow DLI to access other services.
CUs	Sum of the number of compute units and job manager CUs of DLI. CU is also the billing unit of DLI. One CU equals 1 vCPU and 4 GB. The value is the number of CUs required for job running and cannot exceed the number of CUs in the bound queue.
Job Manager CUs	Number of CUs of the management unit.
Parallelism	Number of Flink OpenSource SQL jobs that run at the same time NOTE This value cannot be greater than four times the compute units (number of CUs minus the number of job manager CUs).
Task Manager Config	Whether Task Manager resource parameters are set If this option is selected, you need to set the following parameters: <ul style="list-style-type: none">● CU(s) per TM: Number of resources occupied by each Task Manager.● Slot(s) per TM: Number of slots contained in each Task Manager.
OBS Bucket	OBS bucket to store job logs and checkpoint information. If the OBS bucket you selected is unauthorized, click Authorize .
Save Job Log	Whether job running logs are saved to OBS. The logs are saved in the following path: <i>Bucket name/jobs/logs/Directory starting with the job ID</i> . CAUTION You are advised to configure this parameter. Otherwise, no run log is generated after the job is executed. If the job fails, the run log cannot be obtained for fault locating. If this option is selected, you need to set the following parameters: OBS Bucket: Select an OBS bucket to store job logs. If the OBS bucket you selected is unauthorized, click Authorize . NOTE If Enable Checkpointing and Save Job Log are both selected, you only need to authorize OBS once.

Parameter	Description
Alarm on Job Exception	<p>Whether to notify users of any job exceptions, such as running exceptions or arrears, via SMS or email.</p> <p>If this option is selected, you need to set the following parameters:</p> <p>SMN Topic</p> <p>Select a custom SMN topic. For details about how to customize SMN topics, see Creating a Topic in the <i>Simple Message Notification User Guide</i>.</p>
Enable Checkpointing	<p>Whether to enable job snapshots. If this function is enabled, jobs can be restored based on the checkpoints.</p> <p>If this option is selected, you need to set the following parameters:</p> <ul style="list-style-type: none">• Checkpoint Interval: interval for creating checkpoints, in seconds. The value ranges from 1 to 999999, and the default value is 30.• Checkpoint Mode can be set to either of the following values:<ul style="list-style-type: none">– At least once: Events are processed at least once.– Exactly once: Events are processed only once.• OBS Bucket: Select an OBS bucket to store your checkpoints. If the OBS bucket you selected is unauthorized, click Authorize. The checkpoint path is <i>Bucket name/jobs/checkpoint/Directory starting with the job ID</i>. <p>NOTE</p> <p>If Enable Checkpointing and Save Job Log are both selected, you only need to authorize OBS once.</p>
Auto Restart upon Exception	<p>Whether automatic restart is enabled. If enabled, jobs will be automatically restarted and restored when exceptions occur.</p> <p>If this option is selected, you need to set the following parameters:</p> <ul style="list-style-type: none">• Max. Retry Attempts: maximum number of retries upon an exception. The unit is times/hour.<ul style="list-style-type: none">– Unlimited: The number of retries is unlimited.– Limited: The number of retries is user-defined.• Restore Job from Checkpoint: This parameter is available only when Enable Checkpointing is selected.
Idle State Retention Time	<p>How long the state of a key is retained without being updated before it is removed in GroupBy or Window. The default value is 1 hour.</p>

Parameter	Description
Dirty Data Policy	Policy for processing dirty data. The following policies are supported: Ignore , Trigger a job exception , and Save . If you set this field to Save , the Dirty Data Dump Address must be set. Click the address box to select the OBS path for storing dirty data.

Step 8 (Optional) Set the runtime configuration as required.


Figure 3-6 Runtime configuration



Step 9 Click **Save**.

Step 10 Click **Start**. On the displayed **Start Flink Jobs** page, confirm the job specifications and the price, and click **Start Now** to start the job.

After the job is started, the system automatically switches to the **Flink Jobs** page, and the created job is displayed in the job list. You can view the job status in the **Status** column. Once a job is successfully submitted, its status changes from **Submitting** to **Running**. After the execution is complete, the status changes to **Completed**.

If the job status is **Submission failed** or **Running exception**, the job fails to submit or run. In this case, you can hover over the status icon in the **Status** column of the job list to view the error details. You can click  to copy these details. Rectify the fault based on the error information and resubmit the job.

 NOTE

Other buttons are as follows:

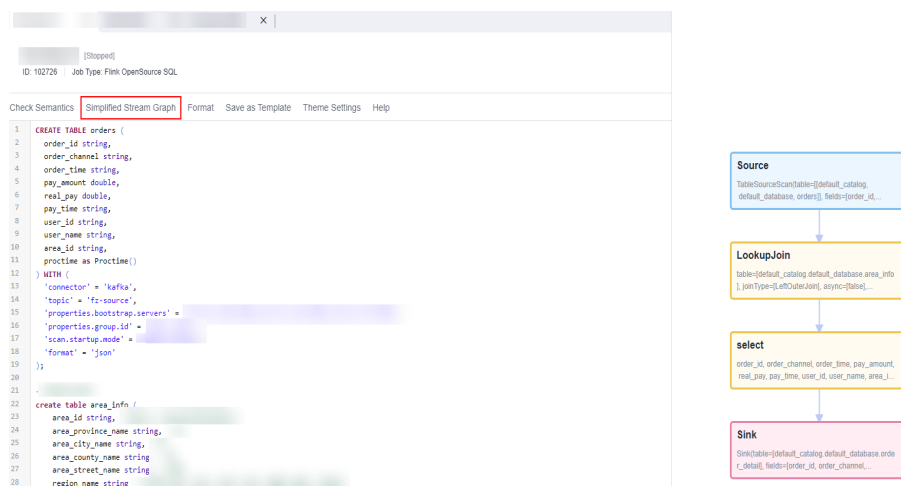
- **Save As:** Save the created job as a new job.
- **Static Stream Graph:** Provide the static concurrency estimation function and stream graph display function. See [Figure 3-8](#).
- **Simplified Stream Graph:** Display the data processing flow from the source to the sink. See [Figure 3-7](#).
- **Format:** Format the SQL statements in the editing box.
- **Set as Template:** Set the created SQL statements as a job template.
- **Theme Settings:** Set the theme related parameters, including **Font Size**, **Wrap**, and **Page Style**.
- **Help:** Redirect to the help center to provide you with the SQL syntax for stream jobs.

----End

Simplified Stream Graph

On the OpenSource SQL job editing page, click **Simplified Stream Graph**.

Figure 3-7 Simplified stream graph



The screenshot shows the Flink OpenSource SQL job editing page. The 'Simplified Stream Graph' tab is active, displaying a SQL script in the editor. The script includes a `CREATE TABLE orders` statement and a `WITH` clause defining a `LookupJoin` operation. To the right, a visual graph illustrates the data flow: a `Source` node (TableSourceScan) feeds into a `LookupJoin` node, which then feeds into a `select` node, and finally into a `Sink` node (Sink).

Static Stream Graph

On the OpenSource SQL job editing page, click **Static Stream Graph**.

 NOTE

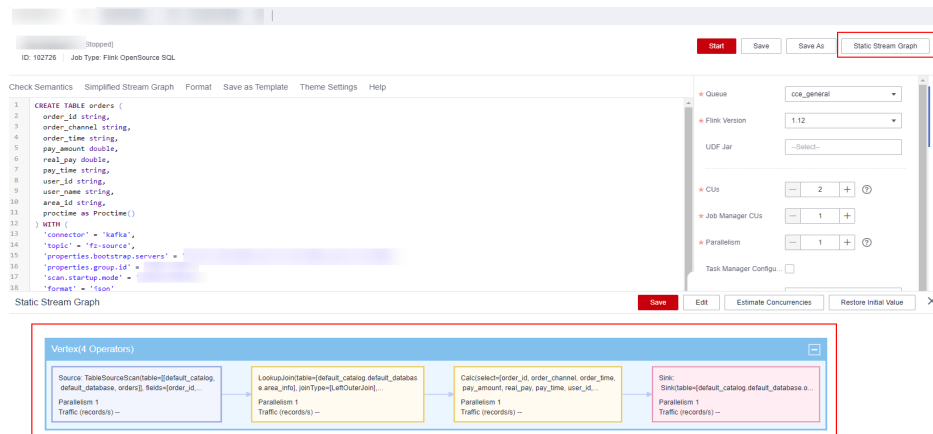
If you use a UDF in a Flink OpenSource SQL job, it is not possible to generate a static stream graph.

The **Static Stream Graph** page also allows you to:

- Estimate concurrencies. Click **Estimate Concurrencies** on the **Static Stream Graph** page to estimate concurrencies. Click **Restore Initial Value** to restore the initial value after concurrency estimation.
- Zoom in or out the page.
- Expand or merge operator chains.

- You can edit **Parallelism**, **Output rate**, and **Rate factor**.
 - **Parallelism**: indicates the number of concurrent tasks.
 - **Output rate**: indicates the data traffic of an operator. The unit is piece/s.
 - **Rate factor**: indicates the retention rate after data is processed by operators. Rate factor = Data output volume of an operator/Data input volume of the operator (Unit: %)

Figure 3-8 Static stream graph



3.3.5 Creating a Flink SQL Job

This section describes how to create a Flink SQL job. You can use Flink SQLs to develop jobs to meet your service requirements. Using SQL statements simplifies logic implementation. You can edit Flink SQL statements for your job in the DLI SQL editor. This section describes how to use the SQL editor to write Flink SQL statements.

DLI Flink OpenSource SQL jobs are fully compatible with the syntax of Flink 1.10 and 1.12 provided by the community. In addition, Redis, GaussDB(DWS), and DIS data source types are added based on the community connector.

Prerequisites

- You have prepared the source and sink streams. For details, see [Preparing Flink Job Data](#).
- When you use a Flink SQL job to access other external data sources, such as OpenTSDB, HBase, Kafka, GaussDB(DWS), RDS, CSS, CloudTable, DCS Redis, and DDS, you need to create a datasource connection to connect the job running queue to the external data source.
 - For details about the external data sources that can be accessed by Flink jobs, see [Cross-Source Analysis Development Methods](#).
 - For details about how to create a datasource connection, see [Enhanced Datasource Connections](#). After a datasource connection is created, you can choose **More > Test Address Connectivity** in the **Operation** column on the **Queue Management** page to check whether the network connection between the queue and the external data source is normal. For details, see [Testing Address Connectivity](#).

Precautions

Before creating jobs and submitting tasks, you are advised to enable CTS to record operations associated with DLI for later query, audit, and backtrack operations. DLI operations that can be recorded by CTS. For details, see [DLI Operations That Can Be Recorded by CTS](#).

For details about how to enable CTS and view trace details, see [Cloud Trace Service Getting Started](#).

Creating a Flink SQL Job

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 In the upper right corner of the **Flink Jobs** page, click **Create Job**.

Figure 3-9 Creating a Flink SQL job

Step 3 Specify job parameters.

Table 3-12 Job configuration information

Parameter	Description
Type	Set Type to Flink SQL . You will need to rewrite SQL statements to start the job.

Parameter	Description
Name	Name of a job. Enter 1 to 57 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed. NOTE The job name must be globally unique.
Description	Description of a job. It can contain a maximum of 512 characters.
Template Name	You can select a sample template or a custom job template. For details about templates, see Managing Flink Templates .
Tags	Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS). For details, see Tag Management Service User Guide . NOTE <ul style="list-style-type: none">A maximum of 20 tags can be added.Only one tag value can be added to a tag key.The key name in each resource must be unique. Tag key: Enter a tag key name in the text box. NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_:=-@) are allowed, but the value cannot start or end with a space or start with _sys_ . Tag value: Enter a tag value in the text box. NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_:=-@) are allowed. The value cannot start or end with a space.

Step 4 Click **OK** to enter the editing page.

Step 5 Edit a Flink SQL job.

Enter detailed SQL statements in the statement editing area. For details about SQL syntax, see the Data Lake Insight SQL Syntax Reference.

Step 6 Click **Check Semantics**.

- You can **Debug** or **Start** a job only after the semantic verification is successful.
- If verification is successful, the message "The SQL semantic verification is complete. No error." will be displayed.
- If verification fails, a red "X" mark will be displayed in front of each SQL statement that produced an error. You can move the cursor to the "X" mark to view error details and change the SQL statement as prompted.

Step 7 Set job running parameters.

Table 3-13 Running parameters

Parameter	Description
Queue	Select a custom queue as needed. NOTE <ul style="list-style-type: none">• During job creation, a sub-user can only select a queue that has been allocated to the user.• If the remaining capacity of the selected queue cannot meet the job requirements, the system automatically scales up the capacity and you will be billed based on the increased capacity. When a queue is idle, the system automatically scales in the queue.
UDF Jar	You can customize a UDF Jar file. Before you select a JAR file, upload the corresponding JAR package to the OBS bucket and choose Data Management > Package Management to create a package. For details, see Creating a Package . In SQL, you can call a user-defined function that is inserted into a JAR file.
CUs	Sum of the number of compute units and job manager CUs of DLI. CU is also the billing unit of DLI. One CU equals 1 vCPU and 4 GB. The configured number of CUs is the number of CUs required for job running and cannot exceed the number of CUs in the bound queue.
Job Manager CUs	Number of CUs of the management unit.
Parallelism	Number of Flink SQL jobs that run at the same time Properly increasing the number of parallel threads improves the overall computing capability of the job. However, the switchover overhead caused by the increase of threads must be considered. NOTE <ul style="list-style-type: none">• This value cannot be greater than four times the compute units (number of CUs minus the number of job manager CUs).• The priority of the number of parallel tasks on this page is lower than that set in the code.
Task Manager Config	Whether Task Manager resource parameters are set If this option is selected, you need to set the following parameters: <ul style="list-style-type: none">• CU(s) per TM: Number of resources occupied by each Task Manager.• Slot(s) per TM: Number of slots contained in each Task Manager.

Parameter	Description
OBS Bucket	<p>OBS bucket to store job logs and checkpoint information. If the OBS bucket you selected is unauthorized, click Authorize.</p> <p>NOTE If Enable Checkpointing and Save Job Log are both selected, you only need to authorize OBS once.</p>
Save Job Log	<p>Whether job running logs are saved to OBS. The logs are saved in the following path: <i>Bucket name/jobs/logs/Directory starting with the job ID</i>. To go to this path, go to the job list and click the job name. On the Run Log tab page, click the provided OBS link.</p> <p>CAUTION You are advised to select this parameter. Otherwise, no run log is generated after the job is executed. If the job is abnormal, the run log cannot be obtained for fault locating.</p> <p>If this option is selected, you need to set the following parameters:</p> <p>OBS Bucket: Select an OBS bucket to store job logs. If the OBS bucket you selected is unauthorized, click Authorize.</p> <p>NOTE If Enable Checkpointing and Save Job Log are both selected, you only need to authorize OBS once.</p>
Alarm on Job Exception	<p>Whether to notify users of any job exceptions, such as running exceptions or arrears, via SMS or email.</p> <p>If this option is selected, you need to set the following parameters:</p> <p>SMN Topic Select a custom SMN topic. For details about how to create a custom SMN topic, see "Creating a Topic" in <i>Simple Message Notification User Guide</i>.</p>


Parameter	Description
Enable Checkpointing	<p>Whether to enable job snapshots. If this function is enabled, jobs can be restored based on the checkpoints.</p> <p>If this option is selected, you need to set the following parameters:</p> <ul style="list-style-type: none">• Checkpoint Interval indicates the interval for creating checkpoints. The value ranges from 1 to 999999, and the default value is 30.• Checkpoint Mode can be set to either of the following values:<ul style="list-style-type: none">– At least once: Events are processed at least once.– Exactly once: Events are processed only once.• OBS Bucket: Select an OBS bucket to store your checkpoints. If the OBS bucket you selected is unauthorized, click Authorize. The checkpoint path is <i>Bucket name/jobs/checkpoint/Directory starting with the job ID</i>. <p>NOTE If Enable Checkpointing and Save Job Log are both selected, you only need to authorize OBS once.</p>
Auto Restart upon Exception	<p>Whether automatic restart is enabled. If enabled, jobs will be automatically restarted and restored when exceptions occur.</p> <p>If this option is selected, you need to set the following parameters:</p> <ul style="list-style-type: none">• Max. Retry Attempts: maximum number of retries upon an exception. The unit is times/hour.<ul style="list-style-type: none">– Unlimited: The number of retries is unlimited.– Limited: The number of retries is user-defined.• Restore Job from Checkpoint: This parameter is available only when Enable Checkpointing is selected.
Idle State Retention Time	<p>How long the state of a key is retained without being updated before it is removed in GroupBy or Window. The default value is 1 hour.</p>
Dirty Data Policy	<p>Select a policy for processing dirty data. The following policies are supported: Ignore, Trigger a job exception, and Save.</p> <p>NOTE Save indicates that the dirty data is stored to the OBS bucket selected above.</p>
Dirty Data Dump Address	<p>Set this parameter when Dirty Data Policy is set to Save. Click the address box to select the OBS path for storing dirty data.</p>

Step 8 (Optional) Set the runtime configuration as required.

Step 9 Click **Save**.

Step 10 Click **Start**. On the displayed **Start Flink Jobs** page, confirm the job specifications and the price, and click **Start Now** to start the job.

After the job is started, the system automatically switches to the **Flink Jobs** page, and the created job is displayed in the job list. You can view the job status in the **Status** column. Once a job is successfully submitted, its status changes from **Submitting** to **Running**. After the execution is complete, the status changes to **Completed**.

If the job status is **Submission failed** or **Running exception**, the job fails to submit or run. In this case, you can hover over the status icon in the **Status** column of the job list to view the error details. You can click  to copy these details. Rectify the fault based on the error information and resubmit the job.

NOTE

Other available buttons are as follows:

- **Save As:** Save the created job as a new job.
- **Format:** Format the SQL statements in the editing box.
- **Set as Template:** Set the created SQL statements as a job template.
- **Theme Settings:** Set the theme related parameters, including **Font Size**, **Wrap**, and **Page Style**.
- **Help:** Redirect to the help center to provide you with the SQL syntax for stream jobs.

----End

3.3.6 Creating a Flink Jar Job

A Flink Jar job involves developing a custom application Jar package based on Flink's capabilities and submitting it to a DLI queue for execution.

To create a Flink Jar job, you need to write and build your own application Jar package. This is suitable for users who require stream data processing and are proficient in Flink's secondary development capabilities.

This section describes how to create a Flink Jar job on the DLI management console.

Prerequisites

- When you use a Flink Jar job to access other external data sources, such as OpenTSDB, HBase, Kafka, GaussDB(DWS), RDS, CSS, CloudTable, DCS Redis, and DDS, you need to create a datasource connection to connect the job running queue to the external data source.
 - For details about the external data sources that can be accessed by Flink jobs, see [Cross-Source Analysis Development Methods](#).
 - For details about how to create a datasource connection, see [Enhanced Datasource Connections](#).

On the **Resources > Queue Management** page, locate the queue you have created, click **More** in the **Operation** column, and select **Test Address Connectivity** to check if the network connection between the queue and the data source is normal. For details, see [Testing Address Connectivity](#).

- When running a Flink Jar job, you need to build the secondary development application code into a Jar package and upload the JAR package to the created OBS bucket. Choose **Data Management > Package Management** to create a package. For details, see [Creating a Package](#).

 **NOTE**

- DLI does not support the download function. If you need to modify the uploaded data file, please edit the local file and upload it again.
- Flink dependencies have been built in the DLI server and security hardening has been performed based on the open-source community version. To avoid dependency package compatibility issues or log output and dump issues, be careful to exclude the following files when packaging:
 - Built-in dependencies (or set the package dependency scope to **provided** in Maven or SBT)
 - Log configuration files (example, **log4j.properties/logback.xml**)
 - JAR package for log output implementation (example, **log4j**).

Precautions

Before creating jobs and submitting tasks, you are advised to enable CTS to record operations associated with DLI for later query, audit, and backtrack operations. To view the DLI operations that can be recorded by CTS, see [DLI Operations That Can Be Recorded by CTS](#).

For details about how to enable CTS and view trace details, see the [Cloud Trace Service Getting Started](#).

Creating a Flink Jar Job

- Step 1** In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.
- Step 2** In the upper right corner of the **Flink Jobs** page, click **Create Job**.
- Step 3** Specify job parameters.

Table 3-14 Job configuration information

Parameter	Description
Type	Select Flink Jar .
Name	Job name. Enter 1 to 57 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed. NOTE The job name must be globally unique.
Description	Description of a job. It can be up to 512 characters long.

Parameter	Description
Tags	<p>Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS).</p> <p>For details, see Tag Management Service User Guide.</p> <p>NOTE</p> <ul style="list-style-type: none"> • A maximum of 20 tags can be added. • Only one tag value can be added to a tag key. • The key name in each resource must be unique. • Tag key: Enter a tag key name in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with <code>_sys_</code>.</p> <ul style="list-style-type: none"> • Tag value: Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed. The value cannot start or end with a space.</p>

Step 4 Click **OK** to enter the editing page.

Step 5 Select a queue.

Step 6 Configuring Flink Jar Job parameters

Table 3-15 Parameters

Parameter	Description
Queue	Select a queue where you want to run your job.
Application	Select the custom package that you have uploaded to the OBS bucket.
Main Class	<p>The name of the JAR package to be loaded, for example, KafkaMessageStreaming.</p> <ul style="list-style-type: none"> • Default: Specified based on the Manifest file in the JAR package. • Manually assign: You must enter the class name and confirm the class arguments (separated by spaces). <p>NOTE When a class belongs to a package, the main class path must contain the complete package path, for example, packageName.KafkaMessageStreaming.</p>

Parameter	Description
Class Arguments	<p>List of arguments of a specified class. The arguments are separated by spaces.</p> <p>Flink parameters support replacement of non-sensitive global variables. For example, if you add the global variable windowSize in Global Configuration > Global Variables, you can add the -windowsSize {{windowSize}} parameter for the Flink Jar job.</p>
JAR Package Dependencies	<p>Select a user-defined package dependency. The dependent program packages are stored in the classpath directory of the cluster.</p> <p>Before selecting a JAR file to be inserted, upload the corresponding JAR file to the OBS bucket and choose Data Management > Package Management to create a package. Select JAR as the package type. For details, see Creating a Package.</p> <p>For details about built-in dependency packages, see Built-in Dependencies.</p>
Other Dependencies	<p>User-defined dependency files. Other dependency files need to be referenced in the code.</p> <p>Before selecting a dependency file, upload the file to the OBS bucket and choose Data Management > Package Management to create a package. The package type is not limited. For details, see Creating a Package.</p> <p>You can add the following command to the application to access the corresponding dependency file. In the command, fileName indicates the name of the file to be accessed, and ClassName indicates the name of the class that needs to access the file.</p> <pre>ClassName.class.getClassLoader().getResource("userData/fileName")</pre>
Flink Version	<p>Before selecting a Flink version, you need to select the queue to which the Flink version belongs. Currently, versions 1.10, 1.11, and 1.12 are supported.</p> <p>NOTE</p> <p>You are not advised to use Spark/Flink engines of different versions for a long time.</p> <ul style="list-style-type: none"> • Doing so can lead to code incompatibility, which can negatively impact job execution efficiency. • Doing so may result in job execution failures due to conflicts in dependencies. Jobs rely on specific versions of libraries or components.
Agency	<p>If you choose Flink 1.15 to execute your job, you can create a custom agency to allow DLI to access other services.</p>

Parameter	Description
Runtime Configuration	<p>User-defined optimization parameters. The parameter format is key=value.</p> <p>Flink optimization parameters support replacement non-sensitive global variable. For example, if you create global variable phase in Global Configuration > Global Variables, optimization parameter table.optimizer.agg-phase.strategy={phase} can be added to the Flink Jar job.</p>

Step 7 Configure job parameters.

Table 3-16 Parameters


Parameter	Description
CUs	One CU consists of one vCPU and 4 GB of memory. The number of CUs ranges from 2 to 10000.
Job Manager CUs	Number of CUs allowed for the job manager. The value ranges from 1 to 4. The default value is 1.
Parallelism	<p>Maximum number of parallel operators in a job. The value ranges from 1 to 10000.</p> <p>NOTE</p> <ul style="list-style-type: none"> The value must be less than or equal to four times the number of compute units (CUs minus the number of job manager CUs). You are advised to set this parameter to a value greater than that configured in the code. Otherwise, job submission may fail.
Task Manager Config	<p>Whether Task Manager resource parameters are set</p> <p>If this option is selected, you need to set the following parameters:</p> <ul style="list-style-type: none"> CU(s) per TM: Number of resources occupied by each Task Manager. Slot(s) per TM: Number of slots contained in each Task Manager.
Save Job Log	<p>Whether to save the job running logs to the OBS bucket.</p> <p>CAUTION</p> <p>You are advised to select this parameter. Otherwise, no run log is generated after the job is executed. If the job is abnormal, the run log cannot be obtained for fault locating.</p> <p>If this option is selected, you need to set the following parameters:</p> <p>OBS Bucket: Select an OBS bucket to store job logs. If the selected OBS bucket is not authorized, click Authorize.</p>

Parameter	Description
Alarm on Job Exception	Whether to notify users of any job exceptions, such as running exceptions or arrears, via SMS or email. If this option is selected, you need to set the following parameters: SMN Topic Select a custom SMN topic. For details about how to customize SMN topics, see Creating a Topic in the <i>Simple Message Notification User Guide</i> .
Auto Restart upon Exception	Whether automatic restart is enabled. If enabled, jobs will be automatically restarted and restored when exceptions occur. If this option is selected, you need to set the following parameters: <ul style="list-style-type: none">● Max. Retry Attempts: maximum number of retries upon an exception. The unit is times/hour.<ul style="list-style-type: none">– Unlimited: The number of retries is unlimited.– Limited: The number of retries is user-defined.● Restore Job from Checkpoint: Restore the job from the saved checkpoint. If you select this parameter, you also need to set Checkpoint Path. Checkpoint Path: Select a path for storing checkpoints. This path must match that configured in the application package. Each job must have a unique checkpoint path, or, you will not be able to obtain the checkpoint.

Step 8 Click **Save** on the upper right of the page.

Step 9 Click **Start** on the upper right side of the page. On the displayed **Start Flink Job** page, confirm the job specifications and the price, and click **Start Now** to start the job.

After the job is started, the system automatically switches to the **Flink Jobs** page, and the created job is displayed in the job list. You can view the job status in the **Status** column. Once a job is successfully submitted, its status changes from **Submitting** to **Running**. After the execution is complete, the status changes to **Completed**.

If the job status is **Submission failed** or **Running exception**, the job fails to submit or run. In this case, you can hover over the status icon in the **Status** column of the job list to view the error details. You can click  to copy these details. Rectify the fault based on the error information and resubmit the job.

NOTE

Other available buttons are as follows:

Save As: Save the created job as a new job.

----End

3.3.7 Performing Operations on a Flink Job

After a job is created, you can perform operations on the job as required.

- [Editing a Job](#)
- [Starting a Job](#)
- [Stopping a Job](#)
- [Deleting a Job](#)
- [Exporting a Job](#)
- [Importing a Job](#)
- [Modifying Name and Description](#)
- [Importing to a Savepoint](#)
- [Triggering a Savepoint](#)
- [Runtime Configuration](#)

Editing a Job

You can edit a created job, for example, by modifying the SQL statement, job name, job description, or job configurations.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 In the row where the job you want to edit locates, click **Edit** in the **Operation** column to switch to the editing page.

Step 3 Edit the job as required.

For details about how to edit a Flink SQL job, see Step 5 to Step 7 in [Creating a Flink SQL Job](#).

For details about how to edit a user-defined Flink job, see Step 5 to Step 7 in [Creating a Flink Jar Job](#).

----End

Starting a Job

You can start a saved or stopped job.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Use either of the following methods to start jobs:

- Starting a single job
Select a job and click **Start** in the **Operation** column.
Alternatively, you can select the row where the job you want to start locates and click **Start** in the upper left of the job list.
- Batch starting jobs
Select the rows where the jobs you want to start locate and click **Start** in the upper left of the job list.

After you click **Start**, the **Start Flink Jobs** page is displayed.

Step 3 On the **Start Flink Jobs** page, confirm the job information and price. If they are correct, click **Start Now**.

After a job is started, you can view the job execution result in the **Status** column.

----End

Stopping a Job

You can stop a job in the **Running** or **Submitting** state.


Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Stop a job using either of the following methods:


- Stopping a job
Locate the row that contains the job to be stopped, click **More** in the **Operation** column, and select **Stop**.
Alternatively, you can select the row where the job you want to stop locates and click **Stop** in the upper left of the job list.
- Batch stopping jobs
Locate the rows containing the jobs you want to stop and click **Stop** in the upper left of the job list.

Step 3 In the displayed **Stop Job** dialog box, click **OK** to stop the job.

Figure 3-10 Stopping a job

 Are you sure you want to stop the following 1 jobs?

Name	Status	Description
	● Running	

Trigger Savepoint 

NOTE

- Before stopping a job, you can trigger a savepoint to save the job status information. When you start the job again, you can choose whether to restore the job from the savepoint.
- If you select **Trigger savepoint**, a savepoint is created. If **Trigger savepoint** is not selected, no savepoint is created. By default, the savepoint function is disabled.
- The lifecycle of a savepoint starts when the savepoint is triggered and stops the job, and ends when the job is restarted. The savepoint is automatically deleted after the job is restarted.

When a job is being stopped, the job status is displayed in the **Status** column of the job list. The details are as follows:

- **Stopping:** indicates that the job is being stopped.
- **Stopped:** indicates that the job is stopped successfully.
- **Stop failed:** indicates that the job failed to be stopped.

----End

Deleting a Job

If you do not need to use a job, perform the following operations to delete it. A deleted job cannot be restored. Therefore, exercise caution when deleting a job.

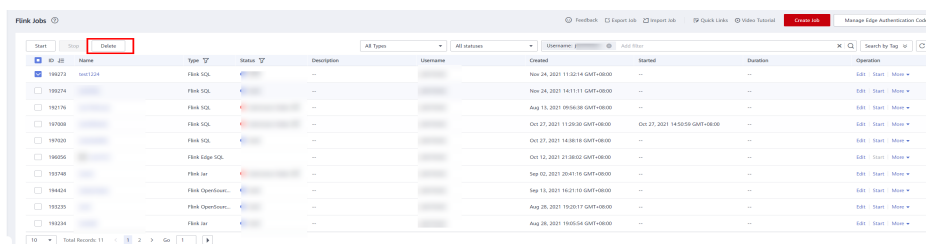
Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Perform either of the following methods to delete jobs:

- Deleting a single job

Locate the row containing the job you want to delete and click **More > Delete** in the **Operation** column.

Alternatively, you can select the row containing the job you want to delete and click **Delete** in the upper left of the job list.



- Deleting jobs in batches

Select the rows containing the jobs you want to delete and click **Delete** in the upper left of the job list.

Step 3 Click **Yes**.

----End

Exporting a Job

You can export the created Flink jobs to an OBS bucket.

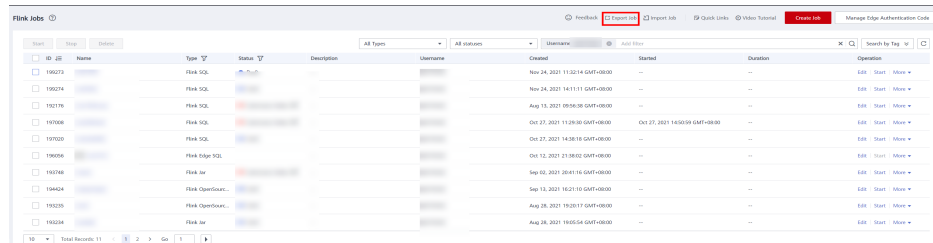
This mode is applicable to the scenario where a large number of jobs need to be created when you switch to another region, project, or user. In this case, you do not need to create a job. You only need to export the original job, log in to the system in a new region or project, or use a new user to import the job.

NOTE

When switching to another project or user, you need to grant permissions to the new project or user. For details, see [Managing Flink Job Permissions](#).

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click **Export Job** in the upper right corner. The **Export Job** dialog box is displayed.



Step 3 Select the OBS bucket where the job is stored. Click **Next**.

Step 4 Select job information you want to export.

By default, configurations of all jobs are exported. You can enable the **Custom Export** function to export configurations of the desired jobs.

Step 5 Click **Confirm** to export the job.

----End

Importing a Job

You can import the Flink job configuration file stored in the OBS bucket to the **Flink Jobs** page of DLI.

This mode is applicable to the scenario where a large number of jobs need to be created when you switch to another region, project, or user. In this case, you do not need to create a job. You only need to export the original job, log in to the system in a new region or project, or use a new user to import the job.

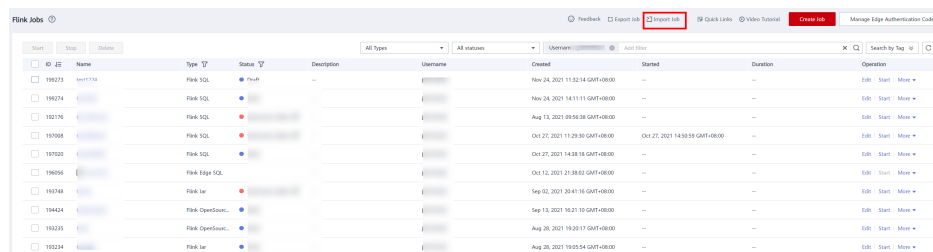
If you need to import a self-created job, you are advised to use the job creation function. For details, see [Creating a Flink SQL Job](#) and [Creating a Flink Jar Job](#).

NOTE

- When switching to another project or user, you need to grant permissions to the new project or user. For details, see [Managing Flink Job Permissions](#).
- Only jobs whose data format is the same as that of Flink jobs exported from DLI can be imported.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click **Import Job** in the upper right corner. The **Import Job** dialog box is displayed.



Step 3 Select the complete OBS path of the job configuration file to be imported. Click **Next**.

Step 4 Configure the same-name job policy and click next. Click **Next**.

- Select **Overwrite job of the same name**. If the name of the job to be imported already exists, the existing job configuration will be overwritten and the job status switches to **Draft**.
- If **Overwrite job of the same name** is not selected and the name of the job to be imported already exists, the job will not be imported.

Step 5 Ensure that **Config File** and **Overwrite Same-Name Job** are correctly configured. Click **Confirm** to import the job.

----End

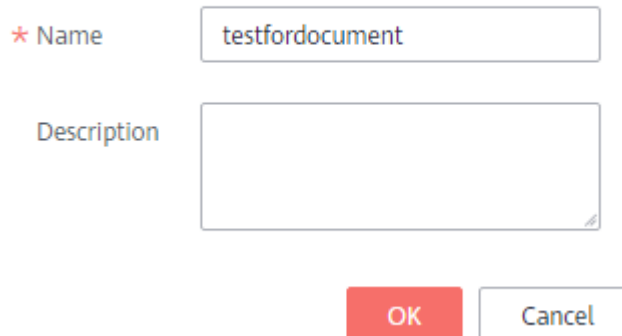
Modifying Name and Description

You can change the job name and description as required.

- Step 1** In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.
- Step 2** In the **Operation** column of the job whose name and description need to be modified, choose **More > Modify Name and Description**. The **Modify Name and Description** dialog box is displayed. Change the name or modify the description of a job.

Figure 3-11 Modifying name and description

Modify Name and Description



The dialog box titled "Modify Name and Description" contains two input fields. The first field, labeled "Name" with a red asterisk, contains the text "testfordocument". The second field, labeled "Description", is a larger text area that is currently empty. At the bottom of the dialog, there are two buttons: a red "OK" button and a white "Cancel" button with a grey border.

Step 3 Click **OK**.

----End

Triggering a Savepoint

When you need to stop a job, you can create a savepoint to save the job status information. In this case, when you restart the job, you can choose to restore the job from the latest savepoint.

 **NOTE**

- You can click **Trigger Savepoint** for jobs in the **Running** status to save the job status.
- The lifecycle of a savepoint starts when the savepoint is triggered and stops the job, and ends when the job is restarted. The savepoint is automatically deleted after the job is restarted.

Importing to a Savepoint

You can import a savepoint to restore the job status. For details about the savepoint, see [Checkpointing](#) at the official website of Flink.

You need to select the OBS path of the save point.

Runtime Configuration

You can select **Runtime Configuration** to configure job exception alarms and restart options.

 **NOTE**

Flink SQL jobs, Flink OpenSource SQL jobs, and Flink Jar jobs are supported.

1. In the **Operation** column of the Flink job, choose **More > Runtime Configuration**.
2. In the **Runtime Configuration** dialog box, set the following parameters:

Figure 3-12 Runtime configuration

✕

Runtime Configuration

Name testflinkjar

Alarm Generation upon...

* SMN Topic DLI_fink_info ▼
[Configure Topic](#)

Auto Restart upon Exc...

Max. Retry Attempts
 Unlimited
 Limited

Restore Job from Che...

* Checkpoint Path dli-test-obs01/

The checkpoint path must be the same as that you set in the application package. Note that the checkpoint path for each job must be unique. Otherwise, the checkpoint cannot be obtained.

Table 3-17 Running parameters

Parameter	Description
Name	Job name.
Alarm Generation upon Job Exception	Whether to report job exceptions, for example, abnormal job running or exceptions due to an insufficient balance, to users via SMS or email. If this option is selected, you need to set the following parameters: SMN Topic Select a user-defined SMN topic. For details about how to customize SMN topics, see Creating a Topic in the <i>Simple Message Notification User Guide</i> .

Parameter	Description
Auto Restart upon Exception	<p>Whether to enable automatic restart. If this function is enabled, any job that has become abnormal will be automatically restarted.</p> <p>If this option is selected, you need to set the following parameters:</p> <ul style="list-style-type: none">● Max. Retry Attempts: maximum number of retry times upon an exception. The unit is times/hour.<ul style="list-style-type: none">– Unlimited: The number of retries is unlimited.– Limited: The number of retries is user-defined.● Restore Job from Checkpoint: Restore the job from the saved checkpoint. <p>NOTE This parameter cannot be configured for Flink SQL jobs or Flink OpenSource SQL jobs.</p> <p>If this parameter is selected, you need to set Checkpoint Path for Flink Jar jobs.</p> <p>Checkpoint Path: Select the checkpoint saving path. The checkpoint path must be the same as that you set in the application package. Note that the checkpoint path for each job must be unique. Otherwise, the checkpoint cannot be obtained.</p>

3.3.8 Flink Job Details

After creating a job, you can view the job details to learn about the following information:

- [Viewing Job Details](#)
- [Checking Job Monitoring Information](#)
- [Viewing the Task List of a Job](#)
- [Viewing the Job Execution Plan](#)
- [Viewing Job Submission Logs](#)
- [Viewing Job Running Logs](#)
- [Viewing Job Tags](#)

Viewing Job Details

This section describes how to view job details. After you create and save a job, you can click the job name to view job details, including SQL statements and parameter settings. For a Jar job, you can only view its parameter settings.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click the name of the job to be viewed. The **Job Detail** tab is displayed.

In the **Job Details** tab, you can view SQL statements, configured parameters, and total cost for the job.

The following uses a Flink SQL job as an example.

Table 3-18 Description

Parameter	Description
Type	Job type, for example, Flink SQL
Name	Flink job name
Description	Description of a Flink job
Status	Running status of a job
Running Mode	The dedicated resource mode is used by default.
Flink Version	Version of Flink selected for the job.
Runtime Configuration	Displayed when a user-defined parameter is added to a job
CUs	Number of CUs configured for a job
Job Manager CUs	Number of job manager CUs configured for a job.
Parallelism	Number of jobs that can be concurrently executed by a Flink job
CU(s) per TM	Number of CUs occupied by each Task Manager configured for a job
Slot(s) per TM	Number of Task Manager slots configured for a job
OBS Bucket	OBS bucket name. After Enable Checkpointing and Save Job Log are enabled, checkpoints and job logs are saved in this bucket.
Save Job Log	Whether the job running logs are saved to OBS
Alarm on Job Exception	Whether job exceptions are reported
SMN Topic	Name of the SMN topic. This parameter is displayed when Alarm Generation upon Job Exception is enabled.
Auto Restart upon Exception	Whether automatic restart is enabled.
Max. Retry Attempts	Maximum number of retry times upon an exception. Unlimited means the number is not limited.
Savepoint	OBS path of the savepoint
Enable Checkpointing	Whether checkpointing is enabled

Parameter	Description
Checkpoint Interval	Interval between storing intermediate job running results to OBS. The unit is second.
Checkpoint Mode	Checkpoint mode. Available values are as follows: <ul style="list-style-type: none"> ● At least once: Events are processed at least once. ● Exactly once: Events are processed only once.
Idle State Retention Time	Defines for how long the state of a key is retained without being updated before it is removed in GroupBy or Window.
Dirty Data Policy	Policy for processing dirty data. The value is displayed only when there is a dirty data policy. Available values are as follows: <ul style="list-style-type: none"> Ignore Trigger a job exception Save
Dirty Data Dump Address	OBS path for storing dirty data when Dirty Data Policy is set to Save .
Created	Time when a job is created
Updated	Time when a job was last updated

----End

Checking Job Monitoring Information

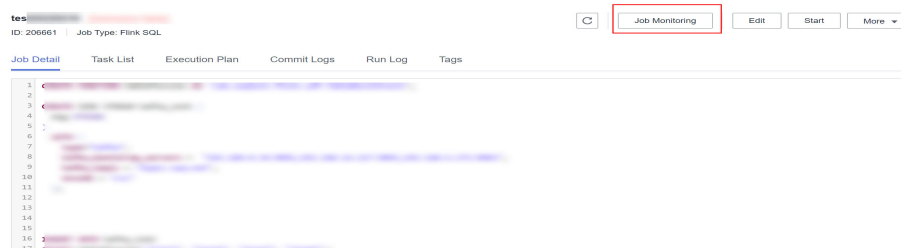
You can use Cloud Eye to view details about job data input and output.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click the name of the job you want. The job details are displayed.

Click **Job Monitoring** in the upper right corner of the page to switch to the Cloud Eye console.

Figure 3-13 Monitoring a Job



The following table describes monitoring metrics related to Flink jobs.

Table 3-19 Monitoring metrics related to Flink jobs

Name	Description
Flink Job Data Read Rate	Displays the data input rate of a Flink job for monitoring and debugging. Unit: record/s.
Flink Job Data Write Rate	Displays the data output rate of a Flink job for monitoring and debugging. Unit: record/s.
Flink Job Total Data Read	Displays the total number of data inputs of a Flink job for monitoring and debugging. Unit: records
Flink Job Total Data Write	Displays the total number of output data records of a Flink job for monitoring and debugging. Unit: records
Flink Job Byte Read Rate	Displays the number of input bytes per second of a Flink job. Unit: byte/s
Flink Job Byte Write Rate	Displays the number of output bytes per second of a Flink job. Unit: byte/s
Flink Job Total Read Byte	Displays the total number of input bytes of a Flink job. Unit: byte
Flink Job Total Write Byte	Displays the total number of output bytes of a Flink job. Unit: byte
Flink Job CPU Usage	Displays the CPU usage of Flink jobs. Unit: %
Flink Job Memory Usage	Displays the memory usage of Flink jobs. Unit: %
Flink Job Max Operator Latency	Displays the maximum operator delay of a Flink job. The unit is ms .
Flink Job Maximum Operator Backpressure	Displays the maximum operator backpressure value of a Flink job. A larger value indicates severer backpressure. 0 : OK 50 : low 100 : high

----End

Viewing the Task List of a Job

You can view details about each task running on a job, including the task start time, number of received and transmitted bytes, and running duration.

NOTE

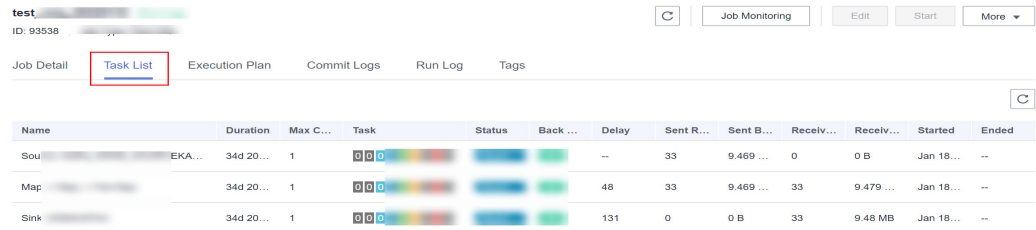
If the value is **0**, no data is received from the data source.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click the name of the job you want. The job details are displayed.

Step 3 On **Task List** and view the node information about the task.

Figure 3-14 Task list



View the operator task list. The following table describes the task parameters.

Table 3-20 Parameter description

Parameter	Description
Name	Name of an operator.
Duration	Running duration of an operator.
Max Concurrent Jobs	Number of parallel tasks in an operator.
Task	Operator tasks are categorized as follows: <ul style="list-style-type: none"> • The digit in red indicates the number of failed tasks. • The digit in light gray indicates the number of canceled tasks. • The digit in yellow indicates the number of tasks that are being canceled. • The digit in green indicates the number of finished tasks. • The digit in blue indicates the number of running tasks. • The digit in sky blue indicates the number of tasks that are being deployed. • The digit in dark gray indicates the number of tasks in a queue.
Status	Status of an operator task.
Back Pressure Status	Working load status of an operator. Available options are as follows: <ul style="list-style-type: none"> • OK: indicates that the operator is in normal working load. • LOW: indicates that the operator is in slightly high working load. DLI processes data quickly. • HIGH: indicates that the operator is in high working load. The data input speed at the source end is slow.

Parameter	Description
Delay	Duration from the time when source data starts being processed to the time when data reaches the current operator. The unit is millisecond.
Sent Records	Number of data records sent by an operator.
Sent Bytes	Number of bytes sent by an operator.
Received Bytes	Number of bytes received by an operator.
Received Records	Number of data records received by an operator.
Started	Time when an operator starts running.
Ended	Time when an operator stops running.

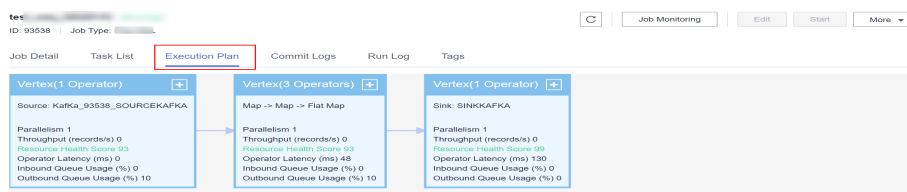
----End

Viewing the Job Execution Plan

You can view the execution plan to understand the operator stream information about the running job.

- Step 1** In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.
- Step 2** Click the name of the job you want. The job details are displayed.
- Step 3** Click the **Execution Plan** tab to view the operator flow direction.

Figure 3-15 Execution plan



Click a node. The corresponding information is displayed on the right of the page.

- Scroll the mouse wheel to zoom in or out.
- The stream diagram displays the operator stream information about the running job in real time.

----End

Viewing Job Submission Logs

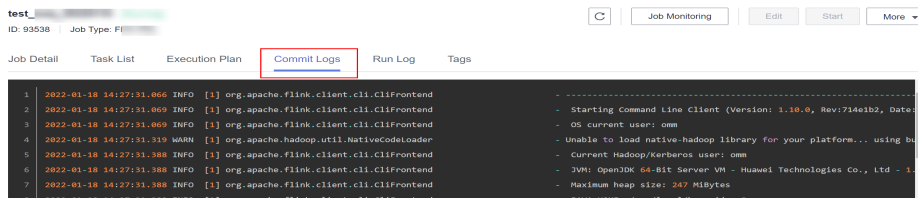
You can view the submission logs to locate the fault.

- Step 1** In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click the name of the job you want. The job details are displayed.

Step 3 In the **Commit Logs** tab, view the information about the job submission process.

Figure 3-16 Commit logs



----End

Viewing Job Running Logs

You can view the run logs to locate the faults occurring during job running.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click the name of the job you want. The job details are displayed.

Step 3 On the **Run Log** tab page, you can view the **Job Manager** and **Task Manager** information of the running job.

Figure 3-17 Run logs



Information about JobManager and TaskManager is updated every minute. Only run logs of the last minute are displayed by default.

If you select an OBS bucket for saving job logs during the job configuration, you can switch to the OBS bucket and download log files to view more historical logs.

If the job is not running, information on the **Task Manager** page cannot be viewed.

----End

Viewing Job Tags

You can view, add, modify, and delete job tags.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 Click the name of the job to be viewed. The **Job Details** page is displayed.

Step 3 On the **Tags** tab page, view the tag information about the current job.



For more information about job tags, see [Tag Management](#).

----End

3.3.9 Tag Management

A tag is a key-value pair customized by users and used to identify cloud resources. It helps users to classify and search for cloud resources. A tag consists of a tag key and a tag value.

DLI allows you to add tags to Flink jobs. You can add tags to Flink jobs to identify information such as the project name, service type, and background. If you use tags in other cloud services, you are advised to create the same tag key-value pairs for cloud resources used by the same business to keep consistency.

DLI supports the following two types of tags:

- Resource tags: indicate non-global tags created on DLI.
- Predefined tags: global tags created on Tag Management Service (TMS).
For more information about predefined tags, see [Tag Management Service User Guide](#).

This section includes the following content:

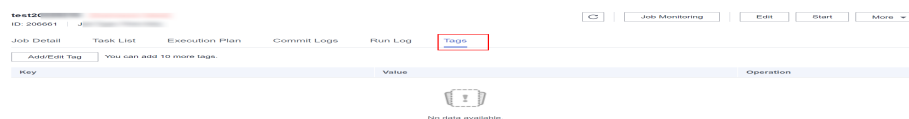
- [Managing a Job Tag](#)
- [Searching for a Job by Tag](#)

Managing a Job Tag

DLI allows you to add, modify, or delete tags for jobs.

- Step 1** In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.
- Step 2** Click the name of the job to be viewed. The **Job Details** page is displayed.
- Step 3** Click **Tags** to display the tag information about the current job.

Figure 3-18 Managing a job tag



- Step 4** Click **Add/Edit Tag** to open to the **Add/Edit Tag** dialog box.
- Step 5** Configure the tag parameters in the **Add/Edit Tag** dialog box.

Figure 3-19 Adding a tag

✕

Add/Edit Tag

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#) ↻

To add a tag, enter a tag key and a tag value below.

Enter a tag key

Enter a tag value

Add

10 tags available for addition.

OK

Cancel

Table 3-21 Tag parameters

Parameter	Description
Tag key	<p>You can perform the following operations:</p> <ul style="list-style-type: none"> Click the text box and select a predefined tag key from the drop-down list. To add a predefined tag, you need to create one on TMS and then select it from the Tag key drop-down list. You can click View predefined tags to go to the Predefined Tags page of the TMS console. Then, click Create Tag in the upper corner of the page to create a predefined tag. For details, see Creating Predefined Tags in <i>Tag Management Service User Guide</i>. Enter a tag key in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=@) are allowed, but the value cannot start or end with a space or start with _sys_.</p>
Tag value	<p>You can perform the following operations:</p> <ul style="list-style-type: none"> Click the text box and select a predefined tag value from the drop-down list. Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=@) are allowed. The value cannot start or end with a space.</p>

NOTE

- A maximum of 20 tags can be added.
- Only one tag value can be added to a tag key.
- The key name in each resource must be unique.

Step 6 Click **OK**.

Step 7 (Optional) In the tag list, locate the row where the tag you want to delete resides, click **Delete** in the **Operation** column to delete the tag.

----End

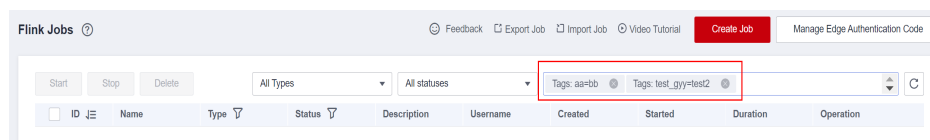
Searching for a Job by Tag

If tags have been added to a job, you can search for the job by setting tag filtering conditions to quickly find it.

Step 1 In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.

Step 2 In the upper right corner of the page, click the search box and select **Tags**.

Figure 3-20 Searching for jobs by tag



Step 3 Choose a tag key and value as prompted. If no tag key or value is available, create a tag for the job. For details, see [Managing a Job Tag](#).

Step 4 Choose other tags to generate a tag combination for job search. You can add a maximum of 10 tags to search for jobs. If you specify more than one tag, jobs containing all the specified tags will be displayed.

Step 5 Click search icon. The target job will be displayed in the job list.

----End

3.3.10 Enabling Dynamic Scaling for Flink Jobs

Scenario

In actual job operations, the compute resources required by a job vary depending on the data volume. As a result, compute resources are wasted when the volume is small and are insufficient when the volume is large.

DLI provides dynamic scaling to dynamically adjust the compute resources used by a job based on the job load, such as the data input and output volume, data input and output rate, and backpressure, to improve resource utilization.

NOTE

Currently, dynamic scaling can be enabled only for jobs of Flink 1.12 or later.

Notes

- During dynamic scaling of a Flink job, if queue resources are occupied and the remaining resources are insufficient for starting the job, the job may fail to be restored.
- When the resources that can be used by a Flink job are dynamically scaled in or out, the background job needs to be stopped and then restored from the savepoint. So, the job cannot process data before the restoration is successful.
- Savepoints need to be triggered during scaling. So, you must configure an OBS bucket, save logs, and enable checkpointing.
- Do not set the scaling detection period to a small value to avoid frequent job start and stop.
- The restoration duration of a scaling job is affected by the savepoint size. If the savepoint size is large, the restoration may take a long time.
- To adjust the configuration items of dynamic scaling, you need to stop the job, edit the job, and submit the job for the modification to take effect.

Procedure

Dynamic scaling applies to Flink OpenSource SQL and Flink Jar jobs.

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Job Management > Flink Jobs**.
3. Select the job for which you want to enable dynamic scaling and click **Edit** in the **Operation** column.
 - For a Flink OpenSource SQL job, click **Runtime Configuration** on the right to configure dynamic scaling parameters.
 - For a Flink Jar job, click the **Runtime Configuration** box to configure dynamic scaling parameters.

Table 3-22 Dynamic scaling parameters

Parameter	Default Value	Description
flink.dli.job.scale.enable	false	Whether to enable dynamic scaling, that is, whether to allow DLI to adjust the resources used by jobs based on job loads and job priorities. If this parameter is set to false , the function is disabled. If this parameter is set to true , the function is enabled. The default value is false .

Parameter	Default Value	Description
flink.dli.job.scale.interval	30	Interval for checking whether to scale the resources for the current job, in minutes. The default value is 30 . For example, 30 indicates that the job is checked every 30 minutes to determine whether to scale in or out the resources used by the job. Note: This configuration is effective only when dynamic scaling is enabled.
flink.dli.job.cu.max	Initial CU value	Maximum number of CUs that can be used by the current job during dynamic scaling. If this parameter is not set, the default value is the initial total number of CUs of the job. Note: The value of this parameter cannot be smaller than the total number of CUs configured by the user. In addition, this parameter is effective only when dynamic scaling is enabled.
flink.dli.job.cu.min	2	Minimum number of CUs that can be used by the current job during dynamic scaling. The default value is 2 . Note: The value of this parameter cannot be greater than the total number of CUs configured by the user. In addition, this parameter is effective only when dynamic scaling is enabled.

3.4 Creating and Managing Spark Jobs

3.4.1 Spark Job Management

Based on the open-source Spark, DLI optimizes performance and reconstructs services to be compatible with the Apache Spark ecosystem and interfaces, and executes batch processing tasks.

DLI also allows you to use Spark jobs to access DLI metadata.

Spark job management provides the following functions:

- [Creating a Spark Job](#)
- [Re-executing a Job](#)
- [Searching for a Job](#)
- [Terminating a Job](#)

In addition, you can click **Quick Links** to switch to the details on *User Guide*.

Spark Jobs Page

On the **Overview** page, click **Spark Jobs** to go to the SQL job management page. Alternatively, you can click **Job Management > Spark Jobs**. The page displays all Spark jobs. If there are a large number of jobs, they will be displayed on multiple pages. DLI allows you to view jobs in all statuses.

Table 3-23 Job management parameters

Parameter	Description
Job ID	ID of a submitted Spark job, which is generated by the system by default.
Name	Name of a submitted Spark job.
Queues	Queue where the submitted Spark job runs
Username	Name of the user who executed the Spark job
Status	Job status. The following values are available: <ul style="list-style-type: none">● Starting: The job is being started.● Running: The job is being executed.● Failed: The session has exited.● Finished: The session is successfully executed.● Restoring: The job is being restored.
Created	Time when a job is created. Jobs can be displayed in ascending or descending order of the job creation time.
Last Modified	Time when a job is completed.
Operation	<ul style="list-style-type: none">● Edit: You can modify the current job configuration and re-execute the job.● SparkUI: After you click this button, the Spark job execution page is displayed. NOTE<ul style="list-style-type: none">● The SparkUI page cannot be viewed for jobs in the Starting state.● Currently, only the latest 100 job information records are displayed on the SparkUI of DLI.● Terminate Job: Cancel a job that is being started or running.● Re-execute: Run the job again.● Archive Log: Save job logs to the temporary bucket created by DLI.● Commit Log: View the logs of submitted jobs.● Driver Log: View the logs of running jobs.

Re-executing a Job

On the **Spark Jobs** page, click **Edit** in the **Operation** column of the job. On the Spark job creation page that is displayed, modify parameters as required and execute the job.

Searching for a Job

On the **Spark Jobs** page, select **Status** or **Queues**. The system displays the jobs that meet the filter condition in the job list.

Terminating a Job

On the **Spark Jobs** page, choose **More > Terminate Job** in the **Operation** column of the job that you want to stop.

3.4.2 Creating a Spark Job

DLI provides fully-managed Spark computing services by allowing you to execute Spark jobs.

On the **Overview** page, click **Create Job** in the upper right corner of the **Spark Jobs** tab or click **Create Job** in the upper right corner of the **Spark Jobs** page. The Spark job editing page is displayed.

On the Spark job editing page, a message is displayed, indicating that a temporary DLI data bucket will be created. The created bucket is used to store temporary data generated by DLI, such as job logs and job results. You cannot view job logs if you choose not to create the bucket. The bucket will be created and the default bucket name is used.

If you do not need to create a DLI temporary data bucket and do not want to receive this message, select **Do not show again** and click **Cancel**.

Prerequisites

- You have uploaded the dependencies to the corresponding OBS bucket on the **Data Management > Package Management** page. For details, see [Creating a Package](#).
- Before creating a Spark job to access other external data sources, such as OpenTSDB, HBase, Kafka, GaussDB(DWS), RDS, CSS, CloudTable, DCS Redis, and DDS, you need to create a datasource connection to enable the network between the job running queue and external data sources.
 - For details about the external data sources that can be accessed by Spark jobs, see [Cross-Source Analysis Development Methods](#).
 - For details about how to create a datasource connection, see [Enhanced Datasource Connections](#).

On the **Resources > Queue Management** page, locate the queue you have created, click **More** in the **Operation** column, and select **Test Address Connectivity** to check if the network connection between the queue and the data source is normal. For details, see [Testing Address Connectivity](#).

Procedure

1. In the left navigation pane of the DLI management console, choose **Job Management > Spark Jobs**. The **Spark Jobs** page is displayed.
Click **Create Job** in the upper right corner. In the job editing window, you can set parameters in **Fill Form** mode or **Write API** mode.
The following uses the **Fill Form** as an example. In **Write API** mode, refer to the *Data Lake Insight API Reference* for parameter settings.
2. Select a queue.
Select the queue you want to use from the drop-down list box.
3. Select a Spark version. Select a supported Spark version from the drop-down list. The latest version is recommended.

NOTE

You are not advised to use Spark/Flink engines of different versions for a long time.

- Doing so can lead to code incompatibility, which can negatively impact job execution efficiency.
 - Doing so may result in job execution failures due to conflicts in dependencies. Jobs rely on specific versions of libraries or components.
4. Configure the job.
Configure job parameters by referring to [Table 3-24](#).

Figure 3-21 Spark job configuration

The screenshot shows a configuration form for a Spark job. The fields and their values are as follows:

- Job Name (--name):** test-catoonfile
- Application:** obs://ai-test/wwxTest/catConfigFile.py
- Main Class (--class):** This parameter cannot be left blank if the Application is a JAR package.
- Application Parameters:** Specify each parameter on a separate line.
- Spark Arguments (--conf):** Enter arguments using the key = value format. Press Enter to separate multiple key-value pairs.
- Job Type:** Basic (selected), AI-enhanced, Image
- JAR Package Dependencies (--jars):** Use the DLI program package for production services
- Python File Dependencies (--py-files):** Use the DLI program package for production services
- Other Dependencies (--files):** Use the DLI program package for production services
- Group Name:** Use the DLI program package for production services
- Access Metadata:** Yes (selected), No
- Retry upon Failure:** Yes (selected), No
- Advanced Settings:** Skip (selected), Configure

Table 3-24 Job configuration parameters

Parameter	Description
Job Name (--name)	Set a job name.
Application	Select the package to be executed. The value can be .jar or .py . You can select the name of a JAR or pyFile package that has been uploaded to the DLI resource management system. You can also specify an OBS path, for example, obs://Bucket name/Package name . Spark 3.3.x or later supports only packages in OBS paths.
Main Class (--class)	Enter the name of the main class. When the application type is .jar , the main class name cannot be empty.

Parameter	Description
Application Parameters	<p>User-defined parameters. Separate multiple parameters by Enter.</p> <p>These parameters can be replaced with global variables. For example, if you create a global variable batch_num on the Global Configuration > Global Variables page, you can use {{batch_num}} to replace a parameter with this variable after the job is submitted.</p>
Spark Arguments (--conf)	<p>Enter a parameter in the format of key=value. Press Enter to separate multiple key-value pairs. For details about the parameters, see Spark Configuration.</p> <p>These parameters can be replaced with global variables. For example, if you create a global variable custom_class on the Global Configuration > Global Variables page, you can use "spark.sql.catalog"={{custom_class}} to replace a parameter with this variable after the job is submitted.</p> <p>NOTE</p> <ul style="list-style-type: none">• The JVM garbage collection algorithm cannot be customized for Spark jobs.• If the Spark version is 3.1.1, configure Spark parameters (--conf) to select a dependent module. For details about the example configuration, see Table 3-25.
Job Type	<p>Set this parameter when you select a CCE queue. Type of the Spark image used by a job. The options are as follows:</p> <ul style="list-style-type: none">• Basic: Basic images provided by DLI. Select this option for non-AI jobs.• Image: Custom Spark images. Select an existing image name and version on SWR.
JAR Package Dependencies (--jars)	<p>JAR file on which the Spark job depends. You can enter the JAR package name or the corresponding OBS path. The format is as follows: obs://Bucket name/Folder name/Package name.</p>
Python File Dependencies (--py-files)	<p>py-files on which the Spark job depends. You can enter the Python file name or the corresponding OBS path of the Python file. The format is as follows: obs://Bucket name/Folder name/File name.</p>
Other Dependencies (--files)	<p>Other files on which the Spark job depends. You can enter the name of the dependency file or the corresponding OBS path of the dependency file. The format is as follows: obs://Bucket name/Folder name/File name.</p>
Group Name	<p>If you select a group when creating a package, you can select all the packages and files in the group. For details about how to create a package, see Creating a Package.</p>
Access Metadata	<p>Whether to access metadata through Spark jobs.</p>

Parameter	Description
Retry upon Failure	Indicates whether to retry a failed job. If you select Yes , you need to set the following parameters: Maximum Retries: Maximum number of retry times. The maximum value is 100 .
Advanced Settings	<ul style="list-style-type: none">• Skip• Configure<ul style="list-style-type: none">– Select Dependency Resources: For details about the parameters, see Table 3-26.– Configure Resources: For details about the parameters, see Table 3-27.

Table 3-25 Spark Parameter (--conf) configuration

Datasource	Example Value
CSS	spark.driver.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/css/* spark.executor.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/css/*
GaussDB(DWS)	spark.driver.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/dws/* spark.executor.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/dws/*
HBase	spark.driver.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/hbase/* spark.executor.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/hbase/*
OpenTSDB	spark.driver.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/opentsdb/* spark.executor.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/opentsdb/*
RDS	spark.driver.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/rds/* spark.executor.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/rds/*
Redis	spark.driver.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/redis/* spark.executor.extraClassPath=/usr/share/extension/dli/ spark-jar/datasource/redis/*

Datasource	Example Value
Mongo	spark.driver.extraClassPath=/usr/share/extension/dli/spark-jar/datasource/mongo/* spark.executor.extraClassPath=/usr/share/extension/dli/spark-jar/datasource/mongo/*

Figure 3-22 Creating a Spark job - advanced settings

Select Dependency Resources

Module Name

Resource Package

Configure Resources

Resource Specifications

Start by selecting one of the resource sets, but then you can modify the executors, executor memory, the number of driver cores, and the driver memory as needed.

Executor Memory

Executor Cores

Executors

Driver Cores

Driver Memory

- Set the following parameters in advanced settings:
 - Select Dependency Resources:** For details about the parameters, see [Table 3-26](#).
 - Configure Resources:** For details about the parameters, see [Table 3-27](#).

NOTE

The parallelism degree of Spark resources is jointly determined by the number of Executors and the number of Executor CPU cores.

Maximum number of tasks that can be concurrently executed = Number of Executors x Number of Executor CPU cores

You can properly plan compute resource specifications based on the compute CUs of the queue you have purchased.

Note that Spark tasks need to be jointly executed by multiple roles, such as **driver** and **executor**. So, the number of executors multiplied by the number of executor CPU cores must be less than the number of compute CUs of the queue to prevent other roles from failing to start Spark tasks. For more information about roles for Spark tasks, see [Apache Spark](#).

Calculation formula for Spark job parameters:

- CUs = Driver Cores + Executors x Executor Cores
- Memory = Driver Memory + (Executors x Executor Memory)

Table 3-26 Parameters for selecting dependency resources

Parameter	Description
modules	<p>If the Spark version is 3.1.1, you do not need to select a module. Configure Spark parameters (--conf).</p> <p>Dependency modules provided by DLI for executing datasource connection jobs. To access different services, you need to select different modules.</p> <ul style="list-style-type: none"> • CloudTable/MRS HBase: sys.datasource.hbase • DDS: sys.datasource.mongo • CloudTable/MRS OpenTSDB: sys.datasource.opentsdb • DWS: sys.datasource.dws • RDS MySQL: sys.datasource.rds • RDS PostGre: sys.datasource.rds • DCS: sys.datasource.redis • CSS: sys.datasource.css <p>DLI internal modules include:</p> <ul style="list-style-type: none"> • sys.res.dli-v2 • sys.res.dli • sys.datasource.dli-inner-table
Resource Package	JAR package on which the Spark job depends.

Table 3-27 Resource specification parameters

Parameter	Description
Resource Specifications	<p>Select a resource specification from the drop-down list box. The system provides three resource specification options for you to choose from.</p> <p>Resource specifications involve the following parameters:</p> <ul style="list-style-type: none"> • Executor Memory • Executor Cores • Executors • Driver Cores • Driver Memory <p>If modified, your modified settings of the items are used.</p>

Parameter	Description
Executor Memory	Customize the configuration item based on the selected resource specifications. Memory of each Executor. It is recommended that the ratio of Executor CPU cores to Executor memory be 1:4.
Executor Cores	Number of CPU cores of each Executor applied for by Spark jobs, which determines the capability of each Executor to execute tasks concurrently.
Executors	Number of Executors applied for by a Spark job
Driver Cores	Number of CPU cores of the driver
Driver Memory	Driver memory size. It is recommended that the ratio of the number of driver CPU cores to the driver memory be 1:4.

- Click **Execute** in the upper right corner of the Spark job editing page.
After the message "Batch processing job submitted successfully" is displayed, you can view the status and logs of the submitted job on the **Spark Jobs** page.

3.5 Setting the Priority for a Job

Scenario

In actual job operations, jobs have different importance and urgency levels. So, the compute resources required for normal operations of important and urgent jobs need to be guaranteed.

DLI allows you to set the priority of each Spark job, Spark SQL job, and Flink job. When compute resources are insufficient, the resources are preferentially used for jobs with higher priorities.

NOTE

- Priorities can only be set for jobs running in elastic resource pools.
- SQL jobs in elastic resource pools support priorities.
- Priorities can be set for jobs of Spark 2.4.5 or later.
- Priorities can be set for jobs of Flink 1.12 or later.

Notes

- You can set the priority for each job. The value ranges from 1 to 10. A larger value indicates a higher priority. Compute resources are preferentially allocated to high-priority jobs. That is, if compute resources required for high-priority jobs are insufficient, compute resources for low-priority jobs are reduced.
- The default priority of Flink jobs running on a general-purpose queue is 5.

- The default priority of Spark jobs running on a general-purpose queue is 3.
- The default priority of jobs running on a SQL queue is 3.
- The change to the job priority takes effect only after the job is stopped, edited, and submitted.
- For Flink jobs, enable dynamic scaling by setting `flink.dli.job.scale.enable` to **true** by referring to [Enabling Dynamic Scaling for Flink Jobs](#), and then set the job priority.
- The change to the priority for a Flink job takes effect only after the job is stopped, edited, and submitted.

Procedure for Setting the Priorities for Flink OpenSource SQL Jobs

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Job Management > Flink Jobs**.
3. Locate the row containing the job for which you want to set the priority and click **Edit** in the **Operation** column.
4. On the right of the page displayed, click **Runtime Configuration**.
5. In the text box, enter the following statements to enable dynamic scaling and then set the job priority:

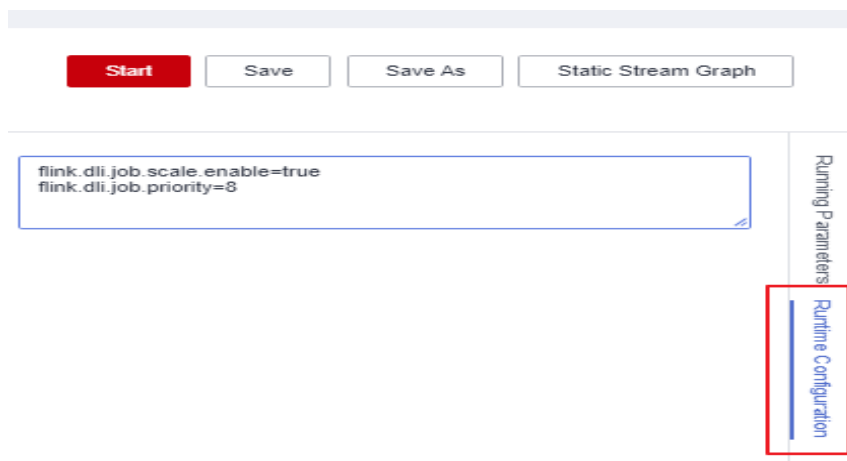
NOTE

For Flink jobs, you must set `flink.dli.job.scale.enable` to **true** to enable dynamic scaling, and then set the job priority.

For more parameter settings for enabling dynamic scaling, see [Enabling Dynamic Scaling for Flink Jobs](#).

```
flink.dli.job.scale.enable=true  
flink.dli.job.priority=x
```

Figure 3-23 Example configuration for a Flink OpenSource SQL job



Procedure for Setting the Priorities for Flink Jar Jobs

In the **Runtime Configuration** text box, enter the following statement. *x* indicates the priority value.

```
flink.dli.job.priority=x
```

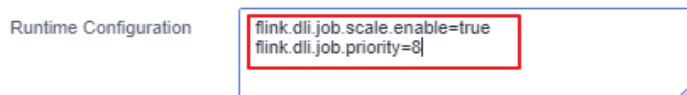

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Job Management > Flink Jobs**.
3. Locate the row containing the job for which you want to set the priority and click **Edit** in the **Operation** column.
4. In the **Runtime Configuration** text box, enter the following statements to enable dynamic scaling and set the job priority:

NOTE

For Flink jobs, you must set **flink.dli.job.scale.enable** to **true** to enable dynamic scaling, and then set the job priority.

For more parameter settings for enabling dynamic scaling, see [Enabling Dynamic Scaling for Flink Jobs](#).

```
flink.dli.job.scale.enable=true  
flink.dli.job.priority=x
```

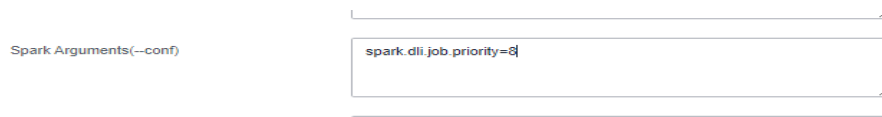
Figure 3-24 Example configuration for a Flink Jar job

Procedure for Setting the Priorities for Spark Jobs

In the **Spark Arguments(--conf)** text box, enter the following statement. *x* indicates the priority value.

```
spark.dli.job.priority=x
```

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Job Management > Spark Jobs**.
3. Locate the row containing the job for which you want to set the priority and click **Edit** in the **Operation** column.
4. In the **Spark Arguments(--conf)** text box, configure the **spark.dli.job.priority** parameter.

Figure 3-25 Example configuration for a Spark job

Procedure for Setting the Priorities for Spark SQL Jobs

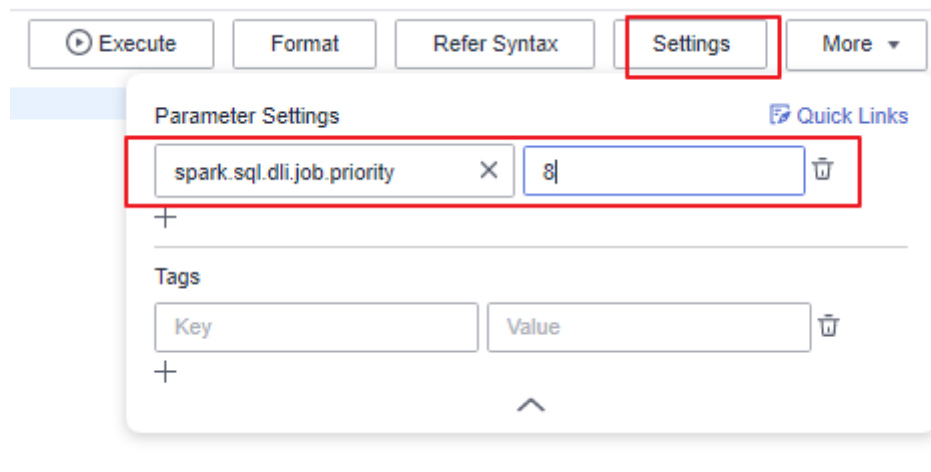
Click **Settings**. In the **Parameter Settings** area, configure the following parameter. *x* indicates the priority value.

```
spark.sql.dli.job.priority=x
```

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Job Management > SQL Jobs**.
3. Locate the row containing the job for which you want to set the priority and click **Edit** in the **Operation** column.

4. Click **Settings**. In the **Parameter Settings** area, configure the **spark.sql.dli.job.priority** parameter.

Figure 3-26 Example configuration for a Spark SQL job



4 Data Management

4.1 Databases and Tables

4.1.1 Overview

DLI database and table management provide the following functions:

- [Database Permission Management](#)
- [Table Permission Management](#)
- [Creating a Database or a Table](#)
- [Deleting a Database or a Table](#)
- [Changing the Owners of Databases and Tables](#)
- [Importing Data](#)
- [Exporting Data](#)
- [Viewing Metadata](#)
- [Previewing Data](#)

Difference Between DLI Tables and OBS Tables

- Data stored in DLI tables is applicable to delay-sensitive services, such as interactive queries.
- Data stored in OBS tables is applicable to delay-insensitive services, such as historical data statistics and analysis.

Notes and Constraints

- **Database**
 - **default** is the database built in DLI. You cannot create a database named **default**.
 - DLI supports a maximum of 50 databases.
- **Table**
 - DLI supports a maximum of 5,000 tables.

- DLI supports the following table types:
 - **MANAGED**: Data is stored in a DLI table.
 - **EXTERNAL**: Data is stored in an OBS table.
 - **View**: A view can only be created using SQL statements.
 - **Datasource table**: The table type is also **EXTERNAL**.
- You cannot specify a storage path when creating a DLI table.
- **Data import**
 - Only OBS data can be imported to DLI or OBS.
 - You can import data in CSV, Parquet, ORC, JSON, or Avro format from OBS to tables created on DLI.
 - To import data in CSV format to a partitioned table, place the partition column in the last column of the data source.
 - The encoding format of imported data can only be UTF-8.
- **Data export**
 - Data in DLI tables (whose table type is **MANAGED**) can only be exported to OBS buckets, and the export path must contain a folder.
 - The exported file is in JSON format, and the text format can only be UTF-8.
 - Data can be exported across accounts. That is, after account B authorizes account A, account A has the permission to read the metadata and permission information of account B's OBS bucket as well as the read and write permissions on the path. Account A can export data to the OBS path of account B.

Databases and Tables Page

The **Databases and Tables** page displays all created databases. You can view the database information, such as the owner and the number of tables.

Table 4-1 Database and table management parameters

Parameter	Description
Database Name	<ul style="list-style-type: none">● The database name can contain only digits, letters, and underscores (_), but cannot contain only digits or start with an underscore (_).● The database name is case insensitive and cannot be left unspecified.● It cannot exceed 128 characters.
Username	Database owner.
Tables	Number of tables in the database.
Description	Description of the database specified during database creation. If no description is provided, -- is displayed.

Parameter	Description
Enterprise Project	Enterprise project to which the database belongs. An enterprise project facilitates project-level management and grouping of cloud resources and users. For details about how to set enterprise projects, see Enterprise Management User Guide .
Operation	<ul style="list-style-type: none">● Permissions: View the permission information and perform user authorization, permission settings, and user permission revocation.● Tables: View the tables in the corresponding database. For details, see Table Management Page.● Create Table: This permission allows you to create a table in the corresponding database.● Modify Database. This permission allows you to change the owner of the database. The username must exist under the same account.● Drop Database: This permission allows you to delete the selected database.● Modify Enterprise Project: Add the enterprise project.● Tags: Add or edit tags.

Table Management Page

From the **Data Management** page, click the database name or **Tables** in the **Operation** column to switch to the table management page.

The displayed page lists all tables created in the current database. You can view the table type, data storage location, and other information. Tables are listed in chronological order by default, with the most recently created tables displayed at the top.

Table 4-2 Table management parameters

Parameter	Description
Table Name	<ul style="list-style-type: none">● The table name can contain only digits, letters, and underscores (_), but cannot contain only digits or start with an underscore (_).● The table name is case insensitive and cannot be left unspecified.● The table name can contain the dollar sign (\$). An example value is \$test.● It cannot exceed 128 characters.

Parameter	Description
Table Type	<p>Table type. Available options are as follows:</p> <ul style="list-style-type: none"> ● Managed: Indicates that data is stored in a DLI table. ● External: Indicates that data is stored in an OBS table. ● View: Indicates the view type. You can only create views using SQL statements. <p>NOTE The table or view information contained in the view cannot be modified. If the table or view information is modified, the query may fail.</p>
Owner	User who creates the table.
Storage Location	DLI, OBS, View, CloudTable, and CSS data location
Size	Size of the data in the table. The value is displayed only for tables of the Managed type. For tables of other types, -- is displayed.
Data Source Path	<ul style="list-style-type: none"> ● If Data Location is OBS, the corresponding OBS path is displayed. ● If Data Location is DLI and View, -- is displayed. ● When the data storage location is a datasource connection service such as CloudTable and CSS, the corresponding URL is displayed.
Created	Time when the table is created.
Last Accessed	<p>Last time when an operation was performed on the table.</p> <p>The last access time of a table refers only to the last time it was updated, not the time it was read (SELECT operation).</p>
Operation	<ul style="list-style-type: none"> ● Manage Permissions: This operation allows you to view the permission information and perform user authorization, permission settings, and user permission revocation. ● More: <ul style="list-style-type: none"> – Delete: Delete a table from the corresponding database. – Modify Owner: Change the owner of a table. The username must exist under the same account. – Import: Import data stored in an OBS bucket to a DLI or OBS table. – Properties: View data in Metadata and Preview tabs.

4.1.2 Managing Database Permissions

Scenario

- By setting permissions, you can assign varying database permissions to different users.
- The administrator and database owner have all permissions, which cannot be set or modified by other users.
- When setting database permissions for a new user, ensure that the region of the user group to which the user belongs has the **Tenant Guest** permission. For details about the Tenant Guest permission and how to apply for the permission, see [Creating a User Group and Assigning Permissions](#) and [System Permissions](#).

Precautions

- Lower-level objects automatically inherit permissions granted to upper-level objects. The hierarchical relationship is database > table > column.
- The database owner, table owner, and **authorized** users can assign permissions on the database and tables.
- Columns can only inherit the query permission. For details about **Inheritable Permissions**, see [Managing Database Permissions](#).
- The permissions can be revoked only at the initial level to which the permissions are granted. You need to grant and revoke permissions at the same level. You need to grant and revoke permissions at the same level. For example, after you are granted the insertion permission on a database, you can obtain the insertion permission on the tables in the database. Your insertion permission can be revoked only at the database level.
- If you create a database with the same name as a deleted database, the database permissions will not be inherited. In this case, you need to grant the database permissions to users or projects.

For example, user A is granted with the permission to delete the **testdb** database. Delete the database and create another one with the same name. You need to grant user A the deletion permission of the **testdb** database again.

Viewing Database Permissions

1. On the left of the management console, choose **Data Management > Databases and Tables**.
2. Locate the row where the target database resides and click **Manage Permissions** in the **Operation** column.

Permissions can be granted to new users or projects, modified for users or projects with existing permissions, or revoked from a user or project.

Granting Permissions to a New User or Project

Here, the new user or project refers to a user or a project that does not have permissions on the database.

1. Click a database you need. In the displayed **Database Permission Management** page, click **Grant Permission** in the upper right corner.

- In the displayed dialog box, select **User** or **Project**, enter the username or select the project to be authorized, and select the required permissions. For details about the permissions, see [Table 4-3](#).

Table 4-3 Parameters

Parameter	Description
Authorization Object	Select User or Project .
Username/ Project Name	<ul style="list-style-type: none">If you select User, enter the IAM username when adding a user to the database. NOTE The username is an existing IAM user name and has logged in to the DLI management console.If you select Project, select the project to be authorized in the current region. NOTE When Project is selected:<ul style="list-style-type: none">If you set Non-inheritable Permissions, you cannot view tables in the corresponding database within the project.If you set Inheritable Permissions, you can view all tables in the database within the project.

Parameter	Description
Non-Inherited Permissions	<p>Select a permission to grant it to the user, or deselect a permission to revoke it.</p> <p>Non-inherited permissions apply only to the current database.</p> <ul style="list-style-type: none"> • The following permissions are applicable to both user and project authorization: <ul style="list-style-type: none"> - Drop Database: This permission allows you to delete the current database. - Create Table: This permission allows you to create tables in the current database. - Create View: This permission allows you to create views in the current database. - Execute SQL EXPLAIN: This permission allows you to execute an EXPLAIN statement and view information about how this database executes a query. - Create Role: This permission allows you to create roles in the current database. - Delete Role: This permission allows you to delete roles of the current database. - View Role: This permission allows you to view the role of the current user. - Bind Role: This permission allows you to bind roles to the current database. - Unbind Role: This permission allows you to bind roles from the current database. - View All Binding Relationships: This permission allows you to view the binding relationships between all roles and users. - Create Function: This permission allows you to create a function in the current database. - Delete Function: This permission allows you to delete functions from the current database. - View All Functions: This permission allows you to view all functions in the current database. - View Function Details: This permission allows you to view details about the current function. • The following permissions can only be granted to users: <ul style="list-style-type: none"> - View All Tables: This permission allows you to view all tables in the current database. <p>NOTE If this permission of a specific database is not granted, all tables in the database will not be displayed.</p> <ul style="list-style-type: none"> - View Database: This permission allows you to view the information about the current database.

Parameter	Description
	NOTE If this permission is not granted, the database will not be displayed.

Parameter	Description
Inherited Permissions	<p>Select a permission to grant it to the user, or deselect a permission to revoke it.</p> <p>Inherited permissions are applicable to the current database and all its tables. However, only the query permission is applicable to table columns.</p> <p>The following permissions can be granted to both user and project.</p> <ul style="list-style-type: none">● Drop Table: This permission allows you to delete tables in a database.● Select Table: This permission allows you to query data of the current table.● View Table Information: This permission allows you to view information about the current table.● Insert: This permission allows you to insert data into the current table.● Add Column: This permission allows you to add columns to the current table.● Overwrite: This permission allows you to insert data to overwrite the data in the current table.● Grant Permission: This permission allows you to grant database permissions to other users or projects.● Revoke Permission: This permission allows you to revoke the permissions of the database that other users have but cannot revoke the database owner's permissions.● Add Partition to Partition Table: This permission allows you to add a partition to a partition table.● Delete Partition from Partition Table: This permission allows you to delete existing partitions from a partition table.● Configure Path for Partition: This permission allows you to set the path of a partition in a partition table to a specified OBS path.● Rename Table Partition: This permission allows you to rename partitions in a partition table.● Rename Table: This permission allows you to rename tables.● Restore Table Partition: This permission allows you to export partition information from the file system and save the information to metadata.● View All Partitions: This permission allows you to view all partitions in a partition table.● View Other Users' Permissions: This permission allows you to query other users' permission on the current database.

3. Click **OK**.

Modifying Permissions for an Existing User or Project

For a user or project that has some permissions on the database, you can revoke the existing permissions or grant new ones.

NOTE

If the options in **Set Permission** are gray, the corresponding account does not have the permission to modify the database. You can apply to the administrator, database owner, or other authorized users for granting and revoking permissions of databases.

1. In the **User Permission Info** list, find the user whose permission needs to be set.
 - If the user is a sub-user, you can set permissions for it.
 - If the user is already an administrator, you can only view the permissions information.

In the **Project Permission Info** list, locate the project for which you want to set permissions and click **Set Permission**.

2. In the **Operation** column of the sub-user or project, click **Set Permission**. The **Set Permission** dialog box is displayed.

For details about the permissions of database users or projects, see [Table 4-3](#).

3. Click **OK**.

Revoking All Permissions of a User or Project

Revoke all permissions of a user or a project.

- In the user list under **User Permission Info**, locate the row where the target sub-user resides and click **Revoke Permission** in the **Operation** column. In the displayed dialog box, click **OK**. In this case, the user has no permissions on the database.

NOTE

If a user is an administrator, **Revoke Permission** is gray, indicating that the user's permission cannot be revoked.

- In the **Project Permission Info** area, select the project whose permissions need to be revoked and click **Revoke Permission** in the **Operation** column. After you click **OK**, the project does not have any permissions on the database.

4.1.3 Managing Table Permissions

Operation Scenario

- By setting permissions, you can assign varying table permissions to different users.
- The administrator and table owner have all permissions, which cannot be set or modified by other users.

- When setting table permissions for a new user, ensure that the user group the user belongs to has the **Tenant Guest** permission.

Precautions

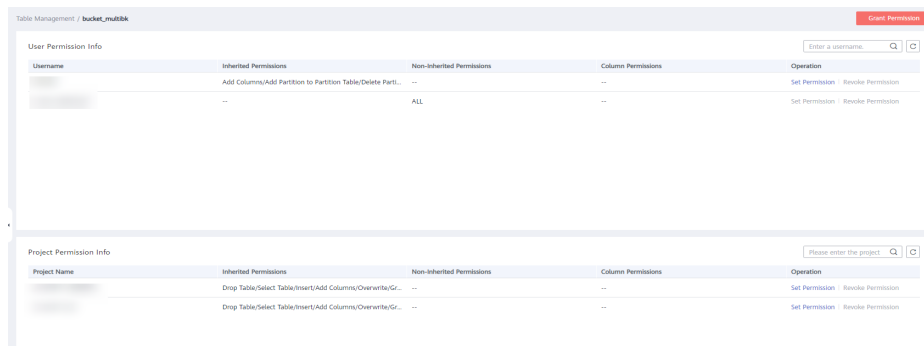
- By the rules in [Common Operations Supported by DLI System Policy](#), you need to authorize a user to view tables in a database of the owner account.
- If you create a table with the same name as a deleted table, the table permissions will not be inherited. In this case, you need to grant the table permissions to users or projects.

For example, user A is granted with the permission to delete the **testTable** table. Delete the table and create another one with the same name. You need to grant user A the deletion permission of the **testTable** table again.

Viewing Table Permissions

1. On the left of the management console, choose **Data Management > Databases and Tables**.
2. Click the database name in the table whose authority is to be set. The **Table Management** page of the database is displayed.
3. Locate the row where the target table resides and click **Manage Permissions** in the **Operation** column.

Figure 4-1 Table permission management



Permissions can be granted to new users or projects, modified for users or projects with existing permissions, or revoked from a user or project.

Granting Permissions to a New User or a Project

Here, the new user or project refers to a user or a project that does not have permissions on the database.

1. Click the table you need. In the displayed table permissions page, click **Grant Permission** in the upper right corner.
2. In the displayed **Grant Permission** dialog box, select the required permissions.
 - For details about the DLI table permissions, see [Table 4-4](#).

Figure 4-2 Granting DLI table permissions to a user

Grant Permission ×

* Authorization Object: User Project

* Username:

Select the permissions to be granted to the user

Non-inheritable Permissions

<input type="checkbox"/> Select all	<input type="checkbox"/> Select Table	<input type="checkbox"/> View Table Info
<input type="checkbox"/> View Table	<input type="checkbox"/> Drop Table	<input type="checkbox"/> Rename Table
<input type="checkbox"/> View Table Creation Statement	<input type="checkbox"/> Overwrite Table	<input type="checkbox"/> Add Column
<input type="checkbox"/> Insert	<input type="checkbox"/> View Other User's Permissions	<input type="checkbox"/> Grant Permission
<input type="checkbox"/> Access Metadata		
<input type="checkbox"/> Revoke Permission		

OK Cancel

Figure 4-3 Granting DLI table permissions to a project

Grant Permission ×

* Authorization Object: User Project

* Project:

Select the permissions to be granted to the user

Non-inheritable Permissions

<input type="checkbox"/> Select all	<input type="checkbox"/> View Table Info	<input type="checkbox"/> View Table Creation Statement
<input type="checkbox"/> Select Table	<input type="checkbox"/> Rename Table	<input type="checkbox"/> Insert
<input type="checkbox"/> Drop Table	<input type="checkbox"/> Add Column	<input type="checkbox"/> Access Metadata
<input type="checkbox"/> Overwrite Table	<input type="checkbox"/> Grant Permission	<input type="checkbox"/> Revoke Permission
<input type="checkbox"/> View Other User's Permissions		

OK Cancel

Table 4-4 Parameter description

Parameter	Description
Authorization Object	Select User or Project .
Username/Project	<ul style="list-style-type: none"> If you select User, enter the IAM username when granting table permissions to the user. <p>NOTE The username is an existing IAM user name and has logged in to the DLI management console.</p> <ul style="list-style-type: none"> If you select Project, select the project to be authorized in the current region. <p>NOTE If you select Project, you can only view information about the authorized tables and their databases.</p>

Parameter	Description
Non-inheritable Permissions	<p>Select a permission to grant it to the user, or deselect a permission to revoke it.</p> <ul style="list-style-type: none">• The following permissions are applicable to both user and project authorization:<ul style="list-style-type: none">– Select Table: This permission allows you to query data of the current table.– View Table Information: This permission allows you to view information about the current table.– View Table Creation Statement: This permission allows you to view the statement for creating the current table.– Drop Table: This permission allows you to delete the current table.– Rename Table: Rename the current table.– Insert: This permission allows you to insert data into the current table.– Overwrite: This permission allows you to insert data to overwrite the data in the current table.– Add Column: Add columns to the current table.– Grant Permission: The current user can grant table permissions to other users.– Revoke Permission: The current user can revoke the table's permissions that other users have but cannot revoke the table owner's permissions.– View Other Users' Permissions: This permission allows you to query other users' permission on the current table. <p>The partition table also has the following permissions:</p> <ul style="list-style-type: none">– Delete Partition: This permission allows you to delete existing partitions from a partition table.– View All Partitions: This permission allows you to view all partitions in a partition table. <ul style="list-style-type: none">• The following permissions can only be granted to users:<ul style="list-style-type: none">– View Table: This permission allows you to display the current table.

- For details about the OBS table permissions, see [Table 4-5](#).

Figure 4-4 Granting OBS table permissions to a user

Figure 4-5 Granting OBS table permissions to a project

Table 4-5 Parameter description

Parameter	Description
Authorization Object	Select User or Project .
Username /Project	<ul style="list-style-type: none"> If you select User, enter the IAM username when granting table permissions to the user. NOTE The username is an existing IAM user name and has logged in to the DLI management console. If you select Project, select the project to be authorized in the current region. NOTE If you select Project, you can only view information about the authorized tables and their databases.

Parameter	Description
Non-inheritable Permissions	<p>Select a permission to grant it to the user, or deselect a permission to revoke it.</p> <ul style="list-style-type: none">• The following permissions are applicable to both user and project authorization:<ul style="list-style-type: none">– View Table Creation Statement: This permission allows you to view the statement for creating the current table.– View Table Information: This permission allows you to view information about the current table.– Select Table: This permission allows you to query data of the current table.– Drop Table: This permission allows you to delete the current table.– Rename Table: Rename the current table.– Insert: This permission allows you to insert data into the current table.– Overwrite: This permission allows you to insert data to overwrite the data in the current table.– Add Column: This permission allows you to add columns to the current table.– Grant Permission: This permission allows you to grant table permissions to other users or projects.– Revoke Permission: This permission allows you to revoke the table's permissions that other users or projects have but cannot revoke the table owner's permissions.– View Other Users' Permissions: This permission allows you to query other users' permission on the current table. <p>The partition table also has the following permissions:</p> <ul style="list-style-type: none">– Add Partition: This permission allows you to add a partition to a partition table.– Delete Partition: This permission allows you to delete existing partitions from a partition table.– Configure Path for Partition: This permission allows you to set the path of a partition in a partition table to a specified OBS path.– Rename Table Partition: This permission allows you to rename partitions in a partition table.– Restore Table Partition: This permission allows you to export partition information from the file system and save the information to metadata.

Parameter	Description
	<ul style="list-style-type: none"> - View All Partitions: This permission allows you to view all partitions in a partition table. • The following permissions can only be granted to users: <ul style="list-style-type: none"> - View Table: This permission allows you to view the current table.

- For details about the view permissions, see [Table 4-6](#).

 **NOTE**

A view can be created only by using SQL statements. You cannot create a view on the **Create Table** page.

Figure 4-6 Granting view permissions to a user

Grant Permission

* Authorization Object User Project

* Username

Select the permissions to be granted to the user

Non-Inherited Permissions

Select all

View Table View Table Info View Table Creation Statement

Drop Table Select Table Rename Table

View Other User's Permissions Grant Permission Revoke Permission

OK Cancel

Figure 4-7 Granting view permissions to a project

Grant Permission

* Authorization Object User Project

* Enterprise Project

Select the permissions to be granted to the user

Non-Inherited Permissions

Select all

View Table Info View Table Creation Statement Drop Table

Select Table Rename Table View Other User's Permissions

Grant Permission Revoke Permission

OK Cancel

Table 4-6 Parameter description

Parameter	Description
Authorization Object	Select User or Project .
Username/ Project	<ul style="list-style-type: none">If you select User, enter the IAM username when adding a user to the database. NOTE The username is an existing IAM user name and has logged in to the DLI management console.If you select Project, select the project to be authorized in the current region. NOTE If you select Project, you can only view information about the authorized tables and their databases.
Non-inheritable Permissions	<p>Select a permission to grant it to the user, or deselect a permission to revoke it.</p> <ul style="list-style-type: none">The following permissions are applicable to both user and project authorization:<ul style="list-style-type: none">View Table Information: This permission allows you to view information about the current table.View Table Creation Statement: This permission allows you to view the statement for creating the current table.Drop Table: This permission allows you to delete the current table.Select Table: This permission allows you to query data of the current table.Rename Table: Rename the current table.Grant Permission: The current user or project can grant table permissions to other users or projects.Revoke Permission: The current user or project can revoke the table's permissions that other users or projects have but cannot revoke the table owner's permissions.View Other Users' Permissions: This permission allows you to query other users' permission on the current table.Only applicable to<ul style="list-style-type: none">View Table: This permission allows you to view the current table.

3. Click **OK**.

Modifying Permissions for an Existing User or Project

For a user or project that has some permissions on the database, you can revoke the existing permissions or grant new ones.

NOTE

If all options under **Set Permission** are gray, you are not allowed to change permissions on this table. You can apply to the administrator, table owner, or other authorized users for granting and revoking table permissions.

1. In the **User Permission Info** list, find the user whose permission needs to be set.
 - If the user is a sub-user and is not the owner of the table, you can set permissions.
 - If the user is an administrator or table owner, you can only view permissions.

In the **Project Permission Info** list, locate the project for which you want to set permissions and click **Set Permission**.

2. In the **Operation** column of the sub-user or project, click **Set Permission**. The **Set Permission** dialog box is displayed.
 - For details about DLI table user or project permissions, see [Table 4-4](#).
 - For details about OBS table user or project permissions, see [Table 4-5](#).
 - For details about View table user or project permissions, see [Table 4-6](#).
3. Click **OK**.

Revoking All Permissions of a User or Project

Revoke all permissions of a user or a project.

- In the user list under **User Permission Info**, locate the row where the target sub-user resides and click **Revoke Permission** in the **Operation** column. In the displayed dialog box, click **OK**. In this case, the user has no permissions on the table.

NOTE

In the following cases, **Revoke Permission** is gray, indicating that the permission of the user cannot be revoked.

- The user is an administrator.
- The sub-user is the owner of the table.
- The sub-user has only inheritable permissions.
- In the **Project Permission Info** area, select the project whose permissions need to be revoked and click **Revoke Permission** in the **Operation** column. After you click **OK**, the project does not have any permissions on the table.

NOTE

If a project has only inheritable permissions, **Revoke Permission** is gray, indicating that the permissions of the project cannot be revoked.

4.1.4 Creating a Database or a Table

Definition of Database and Table in DLI

A database, built on the computer storage device, is a data warehouse where data is organized, stored, and managed based on its structure.

The table is an important part of the database. It consists of rows and columns. Each column functions as a field. Each value in a field (column) represents a type of data.

The database is a framework and the table contains data content. A database has one or more tables.

You can create databases and tables on the management console or using SQL statements. This section describes how to create a database and a table on the management console.

For details about using SQL statements, see [Creating a Database](#), [Creating an OBS Table](#), and [Creating a DLI Table](#).

NOTE


A view can be created only by using SQL statements. You cannot create a view on the **Create Table** page.

Precautions

- If a folder and a file have the same name in the OBS directory, the file path is preferred as the path of the OBS table to be created.
- When creating a database or table, you need to grant permissions to other users so that they can view the created database or table. For details, see [Common Operations Supported by DLI System Policy](#).

Creating a Database

Step 1 You can create a database on either the **Data Management** page or the **SQL Editor** page.

- To create a database on the **Data Management** page:
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. In the upper right corner of the **Databases and Tables** page, click **Create Database** to create a database.
- To create a database on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation pane on the left, click  next to **Databases**.

Step 2 In the displayed **Create Database** dialog box, specify **Name** and **Description** by referring to [Table 4-7](#).

Figure 4-8 Creating a database

Create Database

×

You can create 1 more databases. [Increase quota.](#)

★ Databases

Description

0/128

★ Enterprise Project Enterprise Project ↻ ⓘ Create Enterprise Project

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#) ↻

To add a tag, enter a tag key and a tag value below.

Tags

10 tags available for addition.

Table 4-7 Description

Parameter	Description
Database Name	<ul style="list-style-type: none"> Only numbers, letters, and underscores (_) are allowed. The value cannot contain only numbers or start with an underscore (_). The database name is case insensitive and cannot be left blank. The value can contain a maximum of 128 characters. <p>NOTE The default database is a built-in database. You cannot create the default database.</p>
Enterprise Project	<p>If the created queue belongs to an enterprise project, you can select the corresponding enterprise project.</p> <p>Enterprise projects let you manage cloud resources and users by project.</p> <p>For details about how to set enterprise projects, see Enterprise Management User Guide.</p> <p>NOTE This parameter is displayed only for users who have enabled the Enterprise Management Service.</p>
Description	Description of a database.

Parameter	Description
Tags	<p>Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS).</p> <p>For details, see Tag Management Service User Guide.</p> <p>NOTE</p> <ul style="list-style-type: none">• A maximum of 20 tags can be added.• Only one tag value can be added to a tag key.• The key name in each resource must be unique.• Tag key: Enter a tag key name in the text box. <p>NOTE</p> <p>A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:=-@) are allowed, but the value cannot start or end with a space or start with <code>_sys_</code>.</p> <ul style="list-style-type: none">• Tag value: Enter a tag value in the text box. <p>NOTE</p> <p>A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:=-@) are allowed. The value cannot start or end with a space.</p>

Step 3 Click **OK**.

After a database is created, you can view and select the database for use on the **Databases and Tables** page or **SQL Editor** page.

----End

Creating a Table



Before creating a table, ensure that a database has been created.

Step 1 You can create a table on either the **Databases and Tables** page or the **SQL Editor** page.

NOTE

Datasource connection tables, such as View tables, HBase (CloudTable/MRS) tables, OpenTSDB (CloudTable/MRS) tables, GaussDB(DWS) tables, RDS tables, and CSS tables, cannot be created. You can use SQL to create views and datasource connection tables. For details, see sections **Creating a View** and **Creating a Datasource Connection Table** in the *Data Lake Insight SQL Syntax Reference*.

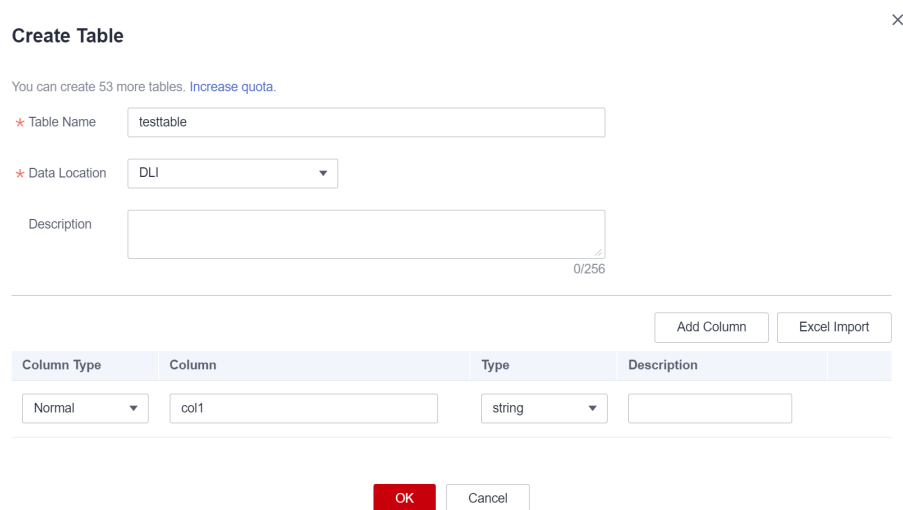
- To create a table on the **Data Management** page:
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. On the **Databases and Tables** page, select the database for which you want to create a table. In the **Operation** column, click **More > Create Table** to create a table in the current database.

- To create a table on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation pane of the displayed **SQL Editor** page, click **Databases**. You can create a table in either of the following ways:
 - Click a database name. In the **Tables** area, click  on the right to create a table in the current database.
 - Click  on the right of the database and choose **Create Table** from the shortcut menu to create a table in the current database.

Step 2 In the displayed **Create Table** dialog box, set parameters as required.

- If you set **Data Location** to **DLI**, set related parameters by referring to [Table 4-8](#).

Figure 4-9 Creating a DLI table



Create Table ×

You can create 53 more tables. [Increase quota](#).

* Table Name

* Data Location

Description 0/256

Column Type	Column	Type	Description
Normal	col1	string	

- If you set **Data Location** to **OBS**, set related parameters by referring to [Table 4-8](#) and [Table 4-9](#).

Figure 4-10 Creating an OBS table

Create Table ×

You can create 53 more tables. [Increase quota.](#)

* Table Name

* Data Location

* Data Format

* Storage Path

Description 0/256

Column Type	Column	Type	Description
<input type="text" value="Normal"/>	<input type="text" value="col1"/>	<input type="text" value="string"/>	<input type="text"/>

Advanced Settings

Table 4-8 Common parameters

Parameter	Description	Example
Table Name	<ul style="list-style-type: none"> - Only numbers, letters, and underscores (_) are allowed. The value cannot contain only numbers or start with an underscore (_). - The table name is case insensitive and cannot be left unspecified. - The table name can contain the dollar sign (\$). An example value is \$test. - The value can contain a maximum of 128 characters. 	table01
Data Location	Data storage location. Currently, DLI and OBS are supported.	DLI
Description	Description of the table.	-
Column Type	Available values: Normal or Partition	Normal

Parameter	Description	Example
Column	<p>Name of a column in a table. The column name must contain at least one letter and can contain underscores (_). It cannot contain only numbers.</p> <p>You can select Normal or Partition. Partition columns are dedicated to partition tables. User data is partitioned to improve query efficiency.</p> <p>NOTE The column name is case-insensitive and must be unique.</p>	name
Type	<p>Data type of a column. This parameter corresponds to Column Name.</p> <ul style="list-style-type: none"> - string: The data is of the string type. - int: Each integer is stored on four bytes. - date: The value ranges from 0000-01-01 to 9999-12-31. - double: Each number is stored on eight bytes. - boolean: Each value is stored on one byte. - decimal: The valid bits are positive integers between 1 to 38, including 1 and 38. The decimal digits are integers less than 10. - smallint/short: The number is stored on two bytes. - bigint/long: The number is stored on eight bytes. - timestamp: The data indicates a date and time. The value can be accurate to six decimal points. - float: Each number is stored on four bytes. - tinyint: Each number is stored on one byte. Only OBS tables support this data type. 	string
Column Description	Description of a column.	-
Operation	<ul style="list-style-type: none"> - Add Column - Delete <p>NOTE If the table to be created includes a great number of columns, you are advised to use SQL statements to create the table or import column information from the local EXCEL file.</p>	-

Table 4-9 Parameter description when **Data Location** is set to **OBS**

Parameter	Description	Example
Data Format	<p>DLI supports the following data formats:</p> <ul style="list-style-type: none"> - Parquet: DLI can read non-compressed data or data that is compressed using Snappy and gzip. - CSV: DLI can read non-compressed data or data that is compressed using gzip. - ORC: DLI can read non-compressed data or data that is compressed using Snappy. - JSON: DLI can read non-compressed data or data that is compressed using gzip. - Avro: DLI can read uncompressed Avro data. 	CSV
Storage Path	<p>Enter or select an OBS path. The path can be a folder or a path.</p> <p>NOTE If you need to import data stored in OBS to the OBS table, set this parameter to the path of a folder. If the table creation path is a file, data fails to be imported.</p>	obs://obs1/ sampledata.csv
Table Header: No/Yes	<p>This parameter is valid only when Data Format is set to CSV. Whether the data source to be imported contains the table header.</p> <p>Click Advanced Settings and select the checkbox next to Table Header: No. If the checkbox is selected, the table header is displayed. If the checkbox is deselected, no table header is displayed.</p>	-
User-defined Delimiter	<p>This parameter is valid only when Data Format is set to CSV and you select User-defined Delimiter.</p> <p>The following delimiters are supported:</p> <ul style="list-style-type: none"> - Comma (,) - Vertical bar () - Tab character (\t) - Others: Enter a user-defined delimiter. 	Comma (,)

Parameter	Description	Example
User-defined Quotation Character	This parameter is valid only when Data Format is set to CSV and you select User-defined Quotation Character . The following quotation characters are supported: <ul style="list-style-type: none">- Single quotation mark (')- Double quotation marks (")- Others: Enter a user-defined quotation character.	Single quotation mark (')
User-defined Escape Character	This parameter is valid only when Data Format is set to CSV and you select User-defined Escape Character . The following escape characters are supported: <ul style="list-style-type: none">- Backslash (\)- Others: Enter a user-defined escape character.	Backslash (\)
Date Format	This parameter is valid only when Data Format is set to CSV or JSON . This parameter specifies the format of the date in the table and is valid only Advanced Settings is selected. The default value is yyyy-MM-dd . For definition of characters involved in the date pattern, see Table 3 in Importing Data to the Table .	2000-01-01
Timestamp Format	This parameter is valid only when Data Format is set to CSV or JSON . This parameter specifies the format of the timestamp in the table and is valid only Advanced Settings is selected. The default value is yyyy-MM-dd HH:mm:ss . For definition of characters involved in the time pattern, see Table 3 in Importing Data to the Table .	2000-01-01 09:00:00

Step 3 Click **OK**.

After a table is created, you can view and select the table for use on the **Data Management** page or **SQL Editor** page.

Step 4 (Optional) After a DLI table is created, you can decide whether to directly import data to the table.

----End

4.1.5 Deleting a Database or a Table

You can delete unnecessary databases and tables based on actual conditions.

Precautions


- You are not allowed to delete databases or tables that are being used for running jobs.
- The administrator, database owner, and users with the database deletion permission can delete the database. The administrator, database owner, and users with the table deletion permission can delete the table.

NOTE

If a database or table is deleted, it cannot be recovered. Exercise caution when performing this operation.

Deleting a Table

You can delete a table on either the **Data Management** page or the **SQL Editor** page.

- Delete the table on the **Data Management** page.
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. Locate the row where the database whose tables you want to delete, click the database name to switch to the **Table Management** page.
 - c. Locate the row where the target table locates and click **More > Delete** in the **Operation** column.
 - d. In the displayed dialog box, click **Yes**.
- Delete a table on the **SQL Editor** page.
 - a. On the top menu bar of the DLI management console, click **SQL Editor**.
 - b. In the navigation tree on the left, click **Databases**. Click the name of a database where the table you want belongs. The tables of the selected database are displayed.
 - c. Click  on the right of the table and choose **Delete** from the shortcut menu.
 - d. In the dialog box that is displayed, click **OK**.

Deleting a Database

1. On the left of the management console, choose **Data Management > Databases and Tables**.
2. Locate the row where the target database locates and click **More > Drop Database** in the **Operation** column.

NOTE

You cannot delete databases that contain tables. To delete a database containing tables, delete the tables first.


3. In the displayed dialog box, click **Yes**.

4.1.6 Modifying the Owners of Databases and Tables

During actual use, developers create databases and tables and submit them to test personnel for testing. After the test is complete, the databases and tables are transferred to O&M personnel for user experience. In this case, you can change the owner of the databases and tables to transfer data to other owners.

Modifying the Database Owner

You can change the owner of a database on either the **Data Management** page or the **SQL Editor** page.

- On the **Data Management** page, change the database owner.
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. On the **Databases and Tables** page, locate the database you want and click **More > Modify Database** in the **Operation** column.
 - c. In the displayed dialog box, enter a new owner name (an existing username) and click **OK**.
- Change the database owner on the **SQL Editor** page.
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation tree on the left, click **Databases**, click  on the right of the database you want to modify, and choose **Modify Database** from the shortcut menu.
 - c. In the displayed dialog box, enter a new owner name (an existing username) and click **OK**.

Modifying the Table Owner

1. On the left of the management console, choose **Data Management > Databases and Tables**.
2. Click the name of the database corresponding to the table to be modified. The **Manage Tables** page of the database is displayed.
3. In the **Operation** column of the target table, choose **More > Modify Owner**.
4. In the displayed dialog box, enter a new owner name (an existing username) and click **OK**.

4.1.7 Importing Data to the Table

You can import data from OBS to a table created in DLI.

Precautions

- Only one path can be specified during data import. The path cannot contain commas (,).
- To import data in CSV format to a partitioned table, place the column to be partitioned in the last column of the data source.
- You are advised not to concurrently import data in to a table. If you concurrently import data into a table, there is a possibility that conflicts occur, leading to failed data import.

- The imported file can be in CSV, Parquet, ORC, JSON, and Avro format. The encoding format must be UTF-8.

Prerequisites

The data to be imported has been stored on OBS.

Procedure

Step 1 You can import data on either the **Data Management** page or the **SQL Editor** page.

- To import data on the **Data Management** page:
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. Click the name of the database corresponding to the table where data is to be imported to switch to the table management page.
 - c. Locate the row where the target table resides and choose **More > Import** in the **Operation** column. The **Import** dialog box is displayed.

Figure 4-11 Importing data

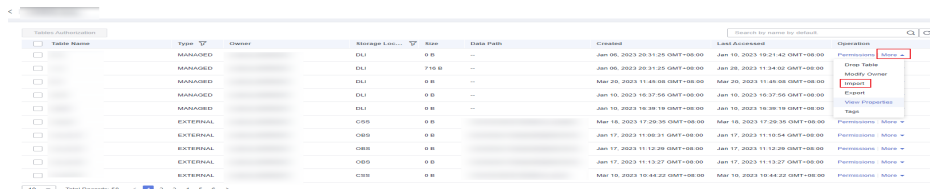


Table Name	Type	Owner	Message Length	Size	State	Created	Last Updated	Operation
	MANAGED		DLI	0 B	---	Jan 06, 2023 20:31:25 GMT+08:00	Jan 10, 2023 10:21:42 GMT+08:00	Permissions More
	MANAGED		DLI	710 B	---	Jan 06, 2023 20:31:25 GMT+08:00	Jan 28, 2023 11:24:02 GMT+08:00	Create Table
	MANAGED		DLI	0 B	---	Mar 20, 2023 11:48:08 GMT+08:00	Mar 20, 2023 11:28:08 GMT+08:00	Modify Owner
	MANAGED		DLI	0 B	---	Jan 10, 2023 10:37:55 GMT+08:00	Jan 10, 2023 10:37:55 GMT+08:00	Export
	MANAGED		DLI	0 B	---	Jan 10, 2023 10:39:18 GMT+08:00	Jan 10, 2023 10:39:18 GMT+08:00	Open Inspector
	MANAGED		DLI	0 B	---	Jan 10, 2023 10:39:18 GMT+08:00	Jan 10, 2023 10:39:18 GMT+08:00	Tags
	EXTERNAL		ORC	0 B	---	Mar 16, 2023 17:20:26 GMT+08:00	Mar 16, 2023 17:20:26 GMT+08:00	Permissions More
	EXTERNAL		ORC	0 B	---	Jan 17, 2023 11:08:21 GMT+08:00	Jan 17, 2023 11:10:24 GMT+08:00	Permissions More
	EXTERNAL		ORC	0 B	---	Jan 17, 2023 11:12:28 GMT+08:00	Jan 17, 2023 11:12:28 GMT+08:00	Permissions More
	EXTERNAL		ORC	0 B	---	Jan 17, 2023 11:13:27 GMT+08:00	Jan 17, 2023 11:13:27 GMT+08:00	Permissions More
	EXTERNAL		ORC	0 B	---	Mar 10, 2023 10:44:22 GMT+08:00	Mar 10, 2023 10:44:22 GMT+08:00	Permissions More


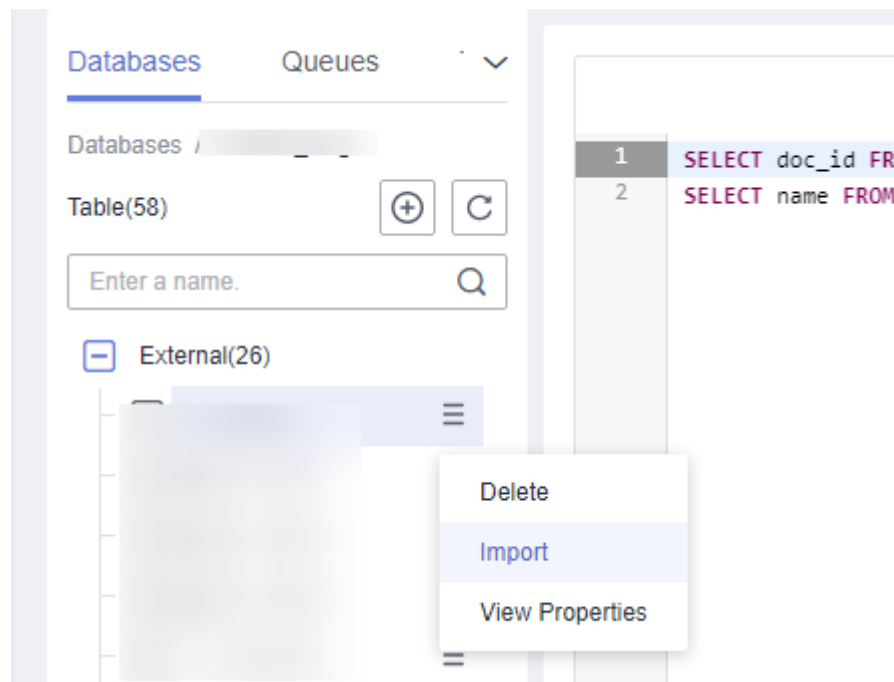

- To import data on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation tree on the left of **SQL Editor**, click **Databases** to see all databases. Click the database where the target table belongs. The table list is displayed.
 - c. Click  on the right of the table and choose **Import** from the shortcut menu. The **Import** page is displayed.

Figure 4-12 SQL editor - importing data



Step 2 In the **Import** dialog box, set the parameters based on [Table 4-10](#).

Table 4-10 Description

Parameter	Description	Example
Databases	Database where the current table is located.	-
Table Name	Name of the current table.	-
Queues	Queue where the imported data will be used	-
File Format	Format of the data source file to be imported. The CSV, Parquet, ORC, JSON, and Avro formats are supported. Encoding format. Only UTF-8 is supported.	CSV
Path	<p>You can directly enter a path or click  and select an OBS path. If no bucket is available, you can directly switch to the OBS management console and create an OBS bucket.</p> <ul style="list-style-type: none"> • When creating an OBS table, you must specify a folder as the directory. If a file is specified, data import may be failed. • If a folder and a file have the same name in the OBS directory, the file path is preferred as the path of the data to be imported. <p>NOTE The path can be a file or folder.</p>	obs://DLI/ sampledat a.csv

Parameter	Description	Example
Table Header: No/Yes	<p>This parameter is valid only when File Format is set to CSV. Whether the data source to be imported contains the table header.</p> <p>Click Advanced Settings and select the checkbox next to Table Header: No. If the checkbox is selected, the table header is displayed. If the checkbox is deselected, no table header is displayed.</p>	-
User-defined Delimiter	<p>This parameter is valid only when File Format is set to CSV and you select User-defined Delimiter.</p> <p>The following delimiters are supported:</p> <ul style="list-style-type: none"> • Comma (,) • Vertical bar () • Tab character (\t) • Others: Enter a user-defined delimiter. 	Default value: comma (,)
User-defined Quotation Character	<p>This parameter is valid only when File Format is set to CSV and User-defined Quotation Character is selected.</p> <p>The following quotation characters are supported:</p> <ul style="list-style-type: none"> • Single quotation mark (') • Double quotation marks (") • Others: Enter a user-defined quotation character. 	Default value: double quotation marks (")
User-defined Escape Character	<p>This parameter is valid only when File Format is set to CSV and you select User-defined Escape Character.</p> <p>The following escape characters are supported:</p> <ul style="list-style-type: none"> • Backslash (\) • Others: Enter a user-defined escape character. 	Default value: backslash (\)
Date Format	<p>This parameter is valid only when File Format is set to CSV or JSON.</p> <p>This parameter specifies the format of the date in the table and is valid only Advanced Settings is selected. The default value is yyyy-MM-dd. For definition of characters involved in the date pattern, see Table 3 in Importing Data to the Table.</p>	2000-01-01


Parameter	Description	Example
Timestamp Format	This parameter is valid only when File Format is set to CSV or JSON . This parameter specifies the format of the timestamp in the table and is valid only Advanced Settings is selected. The default value is yyyy-MM-dd HH:mm:ss . For definition of characters involved in the time pattern, see Table 3 in Importing Data to the Table .	2000-01-01 09:00:00
Error Records Path	This parameter is valid only when File Format is set to CSV or JSON . The parameter specifies the error data is stored in the corresponding OBS path and is valid only Advanced Settings is selected.	obs://DLI/

Step 3 Click **OK**.

Step 4 You can view the imported data in either of the following ways:

 **NOTE**

Currently, only the first 10 records are displayed.

- Choose **Data Management > Databases and Tables** in the navigation pane of the console. Locate the row that contains the database where the target table belongs and click **More > View Properties** in the **Operation** column. In the displayed dialog box, click the **Preview** tab to view the imported data.
- On the **Databases** tab of the **SQL Editor**, click the database name to go to the table list. Click  on the right of a table name and choose **View Properties** from the shortcut menu. In the displayed dialog box, click **Preview** to view the imported data.

Step 5 (Optional) View the status and execution result of the importing job on the **Job Management > SQL Jobs** page.

----End

4.1.8 Exporting Data from DLI to OBS

You can export data from a DLI table to OBS. During the export, a folder is created in OBS or the content in the existing folder is overwritten.

Precautions

- The exported file can be in JSON format, and the text format can only be UTF-8.
- Only the data in the DLI table (the table type is **Managed**) can be exported to the OBS bucket, and the export path must be specified to the folder level.
- Data can be exported across accounts. That is, after account B authorizes account A, account A can export data to the OBS path of account B if account A has the permission to read the metadata and permission information about the OBS bucket of account B and read and write the path.

Procedure


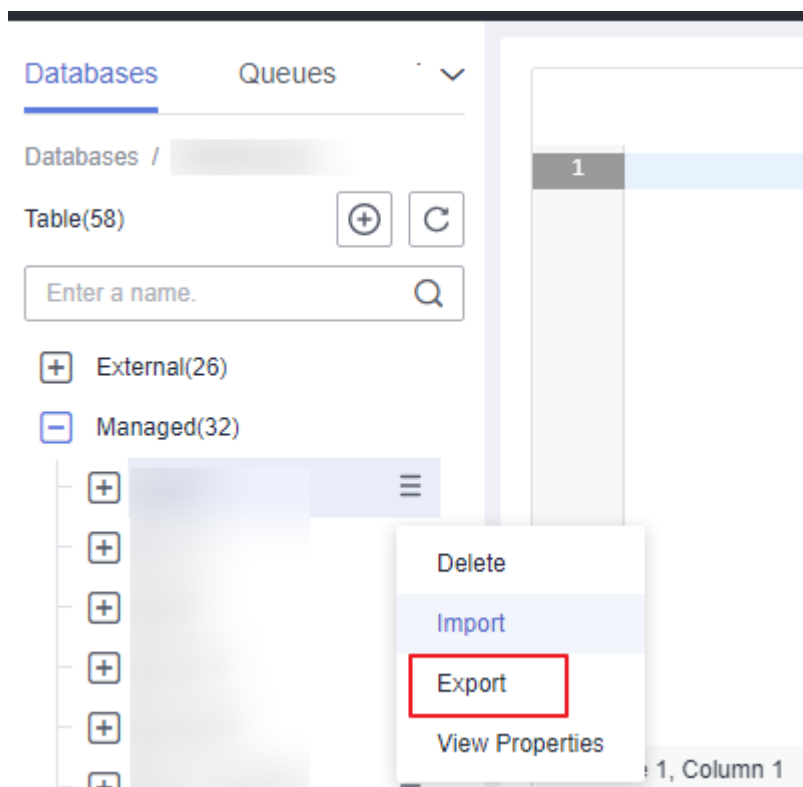
- Step 1** You can export data on either the **Data Management** page or the **SQL Editor** page.
- To export data on the **Data Management** page:
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. Click the name of the database corresponding to the table where data is to be exported to switch to the **Manage Tables** page.
 - c. Select the corresponding table (DLI table) and choose **More > Export** in the **Operation** column. The **Export Data** page is displayed.
 - To export data on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation tree on the left, click **Databases** to see all databases. Click the database name corresponding to the table to which data is to be exported. The tables are displayed.
 - c. Click  on the right of the managed table (DLI table) whose data is to be exported, and choose **Export** from the shortcut menu.

Figure 4-13 Exporting the table



- Step 2** In the displayed **Export Data** dialog box, specify parameters by referring to [Table 4-11](#).

Figure 4-14 Exporting data

Export Result
×

★ Databases db1

★ Table Name src

★ Data Format json

Queues

Compression Format

★ Storage Path

★ Export Mode
 New OBS directory
 Existing OBS directory (Overwritten)

Table Header
 No
 Yes


Table 4-11 Parameter description

Parameter	Description
Databases	Database where the current table is located.
Table Name	Name of the current table.
Data Format	Format of the file storing data to be exported. Formats other than JSON will be supported in later versions.
Queue	Select a queue.
Compression Format	Compression format of the data to be exported. The following compression formats are supported: <ul style="list-style-type: none"> ● none ● bzip2 ● deflate ● gzip
Storage Path	<ul style="list-style-type: none"> ● Enter or select an OBS path. ● The export path must be a folder that does not exist in the OBS bucket. Specifically, you need to create a folder in the target OBS directory. ● The folder name cannot contain the special characters of \/:*?"<> , and cannot start or end with a dot (.).

Parameter	Description
Export Mode	Storage mode of the data to be exported. <ul style="list-style-type: none">• New OBS directory: If the specified export directory exists, an error is reported and the export operation cannot be performed.• Existing OBS directory (Overwritten): If you create a file in the specified directory, the existing file will be overwritten.
Table Header: No/Yes	Whether the data to be exported contains the table header.

Step 3 Click **OK**.

Step 4 (Optional) You can view the job status (indicated by **Status**), statements (indicated by **Statement**), and other information about exporting jobs on the **SQL Jobs** page.

1. Select **EXPORT** from the **Job Type** drop-down list box and specify the time range for exporting data. The jobs meeting the requirements are displayed in the job list.
2. Click  to view details about an exporting job.

----End

4.1.9 Viewing Metadata


Metadata Description

- Metadata is used to define data types. It describes information about the data, including the source, size, format, and other data features. In database fields, metadata interprets data content in the data warehouse.
- When you create a table, metadata is defined, consisting of the column name, type, and description.
- The **Metadata** page displays information about the target table, including **Column Name**, **Column Type**, **Data Type**, and **Description**.

Procedure

You can view metadata on either the **Data Management** page or the **SQL Editor** page.

- To view metadata on the **Data Management** page:
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. On the displayed **Data Management** page, click the name of the database where the target table whose data you want to export resides to switch to the **Manage Tables** page.
 - c. Click **More** in the **Operation** column of the target table and select **View Properties**. In the **Metadata** tab, view the metadata of the table.


- To view metadata on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation pane of the displayed **SQL Editor** page, click **Databases**.
 - c. Click the corresponding database name to view the tables in the database.
 - d. Click  on the right of the table and choose **View Properties** from the shortcut menu. On the **Metadata** tab page, view the metadata of the table.

4.1.10 Previewing Data

The **Preview** page displays the first 10 records in the table.

Procedure

You can preview data on either the **Data Management** page or the **SQL Editor** page.

- To preview data on the **Data Management** page:
 - a. On the left of the management console, choose **Data Management > Databases and Tables**.
 - b. On the displayed **Data Management** page, click the name of the database where the target table whose data you want to export resides to switch to the **Manage Tables** page.
 - c. Click **More** in the **Operation** column of the target table and select **View Properties**.
 - d. Click the **Preview** tab to preview the table data.
- To preview data on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the navigation pane of the displayed **SQL Editor** page, click **Databases**.
 - c. Click the corresponding database name to view the tables in the database.
 - d. Click  on the right of the corresponding table, choose **View Properties** from the list menu, and click the **Preview** tab to preview the data of the table.

4.1.11 Managing Tags

Tag Management

A tag is a key-value pair that you can customize to identify cloud resources. It helps you to classify and search for cloud resources. A tag consists of a tag key and a tag value. If you use tags in other cloud services, you are advised to create the same tag (key-value pairs) for cloud resources used by the same business to keep consistency.

DLI supports the following two types of tags:

- Resource tags: non-global tags created on DLI
- Predefined tags: global tags created on Tag Management Service (TMS).
For more information about predefined tags, see [Tag Management Service User Guide](#).

This section describes how to add, modify, and delete tags for databases and tables.

Database Tags

- Step 1** In the navigation pane on the left, choose **Data Management > Databases and Tables**.
- Step 2** Locate the row that contains the target database, and click **More > Tags** in the **Operation** column.
- Step 3** The tag management page is displayed, and the tags (if there are) are displayed.
- Step 4** On the displayed page, click **Add/Edit Tag**. The **Add/Edit Tag** dialog box is displayed.

Enter a tag key and a tag value in the text boxes and click **Add**.

Figure 4-15 Adding/Editing tags

Add/Edit Tag ×

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#) ↻

To add a tag, enter a tag key and a tag value below.

test = test ×

dli 001 Add

19 tags available for addition.

OK Cancel

Table 4-12 Tag parameters

Parameter	Description
Tag key	<p>You can specify the tag key in either of the following ways:</p> <ul style="list-style-type: none">Click the text box for tag key and select a predefined tag key from the drop-down list. To add a predefined tag, you need to create one on TMS and then select it from the Tag key drop-down list. You can click View predefined tags to go to the Predefined Tags page of the TMS console. Then, click Create Tag in the upper corner of the page to create a predefined tag. For details, see Creating Predefined Tags in <i>Tag Management Service User Guide</i>.Enter a tag key in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with _sys_.</p>
Tag value	<p>You can specify the tag value in either of the following ways:</p> <ul style="list-style-type: none">Click the tag value text box and select a predefined tag value from the drop-down list.Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed. The value cannot start or end with a space.</p>

NOTE

- A maximum of 20 tags can be added.
- Only one tag value can be added to a tag key.
- The key name in each resource must be unique.

Step 5 Click **OK**. The database tag is added.

To delete a tag, click **Delete** in the **Operation** column of the target tag.

----End

Table Tags

Step 1 In the navigation pane on the left, choose **Data Management > Databases and Tables**.

Step 2 Click a database name to view the tables in the database.

Step 3 Locate the row that contains the target table and click **More > Tag** in the **Operation** column.

Step 4 The tag management page is displayed, and the tags (if there are) are displayed.

Step 5 On the displayed page, click **Add/Edit Tag**. The **Add/Edit Tag** dialog box is displayed.

Enter a tag key and a tag value in the text boxes and click **Add**

Figure 4-16 Adding/Editing tags

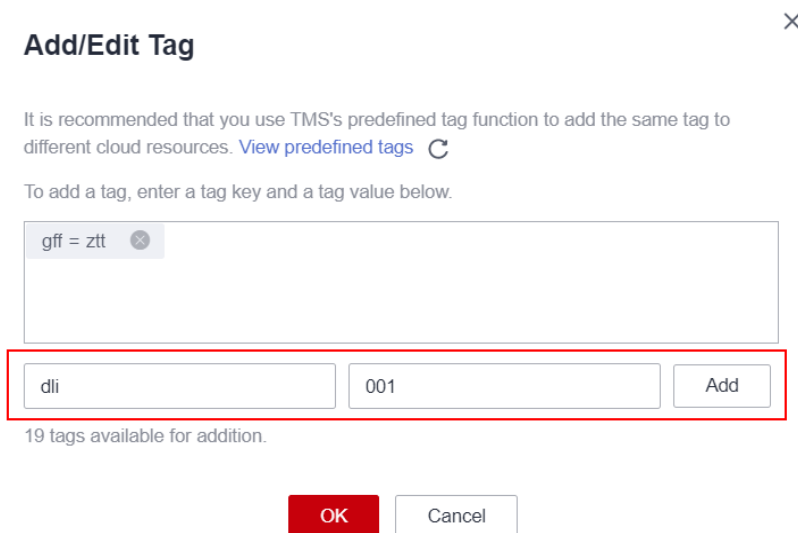


Table 4-13 Tag parameters

Parameter	Description
Tag key	<p>You can specify the tag key in either of the following ways:</p> <ul style="list-style-type: none"> Click the text box for tag key and select a predefined tag key from the drop-down list. To add a predefined tag, you need to create one on TMS and then select it from the Tag key drop-down list. You can click View predefined tags to go to the Predefined Tags page of the TMS console. Then, click Create Tag in the upper corner of the page to create a predefined tag. For details, see Creating Predefined Tags in <i>Tag Management Service User Guide</i>. Enter a tag key in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with _sys_.</p>

Parameter	Description
Tag value	<p>You can specify the tag value in either of the following ways:</p> <ul style="list-style-type: none">• Click the tag value text box and select a predefined tag value from the drop-down list.• Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:=-@) are allowed. The value cannot start or end with a space.</p>

 **NOTE**

- A maximum of 20 tags can be added.
- Only one tag value can be added to a tag key.
- The key name in each resource must be unique.

Step 6 Click **OK**. The table tag is added.

To delete a tag, click **Delete** in the **Operation** column of the target tag.

----End

4.2 Package Management

4.2.1 Overview

Package management provides the following functions:

- [Managing Package Permissions](#)
- [Creating a Package](#)
- [Deleting a Package](#)

 **NOTE**

You can delete program packages in batches.

- [Modifying the Owner](#)

Notes and Constraints

- A package can be deleted, but a package group cannot be deleted.
- The following types of packages can be uploaded:
 - **JAR**: JAR file
 - **PyFile**: User Python file
 - **File**: User file
 - **ModelFile**: User AI model file

Package Management Page

Table 4-14 Package management parameters

Parameter	Description
Group Name	Name of the group to which the package belongs. If the package is not grouped, -- is displayed.
Package Name	Name of a package.
Owner	Name of the user who uploads the package.
Type	Type of a package. The following package types are supported: <ul style="list-style-type: none">• JAR: JAR file• PyFile: User Python file• File: User file• ModelFile: User AI model file
Status	Status of the package to be created. <ul style="list-style-type: none">• Uploading: The file is being uploaded.• Finished: The resource package has been uploaded.• Failed: The resource package upload failed.
Created	Time when a package is created.
Updated	Time when the package is updated.
Operation	Manage Permissions: Manage user permissions for a package. Delete: Delete the package. More: <ul style="list-style-type: none">• Modify Owner: Modify the owner of the package.• Tags: Add or edit package tags.

4.2.2 Managing Permissions on Packages and Package Groups

Scenario

- You can isolate package groups or packages allocated to different users by setting permissions to ensure data query performance.
- The administrator and the owner of a package group or package have all permissions. You do not need to set permissions and the permissions cannot be modified by other users.
- When you set permissions on a package group or a package to a new user, the region of the user group the user belong to must have the Tenant Guest permission. For details about the Tenant Guest permission and how to apply for the permission, see [Creating a User Group and Assigning Permissions](#) and [System Permissions](#).

On the **Package Management** page, click **Manage Permissions** in the **Operation** column of the target package. On the displayed **User Permission Info** page, you can grant permissions for the package group or package, set and revoke user permissions.

 **NOTE**

- If you select a group when creating a package, you can manage permissions of the corresponding program package group.
- If you select **No grouping** when creating a package, you can manage permissions of the corresponding package.

Granting Permissions on Package Groups/Packages

Click **Grant Permission** in the upper right corner of the page.

- Granting permissions on package groups

Figure 4-17 Granting permissions on package groups

Grant Permission

* Username

Select the permissions to be granted to the user

Select all
 Use Group
 Update Group
 Get Group

Delete Group
 Grant Permission
 Revoke Permission

View Other User's Permissions

Table 4-15 Permission parameters

Parameter	Description
Username	Name of the authorized IAM user. NOTE The username is the name of an existing IAM user.

Parameter	Description
Select the permissions to be granted to the user	<ul style="list-style-type: none"> • Use Group: This permission allows you to use the package of this group. • Update Group: This permission allows you to update the packages in the group, including creating a package in the group. • Query Group: This permission allows you to query the details of a package in a group. • Delete Group: This permission allows you to delete the package of the group. • Grant Permission: This permission allows you to grant group permissions to other users. • Revoke Permission: This permission allows you to revoke the group permissions that other users have but cannot revoke the group owner's permissions. • View Other User's Permissions: This permission allows you to view the group permissions of other users.

- Granting permissions on packages

Figure 4-18 Granting permissions on package groups

Grant Permission

* Username

Select the permissions to be granted to the user

Select all
 Use Resource
 Update Resource
 Get Resource

Delete Resource
 Grant Permission
 Revoke Permission

View Other User's Permissions

Table 4-16 Permission parameters

Parameter	Description
Username	Name of the authorized IAM user. NOTE The username is the name of an existing IAM user.

Parameter	Description
Select the permissions to be granted to the user	<ul style="list-style-type: none">● Use Package: This permission allows you to use the package.● Update Package: This permission allows you to update the package.● Query Package: This permission allows you to query the package.● Delete Package: This permission allows you to delete the package.● Grant Permission: This permission allows you to grant package permissions to other users.● Revoke Permission: This permission allows you to revoke the package permissions that other users have but cannot revoke the package owner's permissions.● View Other User's Permissions: This permission allows you to view the package permissions of other users.

Setting Permissions on Package Groups and Packages

Click **Set Permission** in the **Operation** column of the sub-user to modify the permission of the user. [Table 4-15](#) and [Table 4-16](#) list the detailed permission descriptions.

If the **Set Permission** button is gray, you do not have the permission to modify the package group or package. You can apply to the administrator, group owner, or other users who have the permissions on granting and revoking permissions of package groups or packages.

Revoking Permissions on Package Groups and Packages

Click **Revoke Permission** in the **Operation** column of a sub-user to revoke the user's permissions. After the operation, the sub-user does not have any permission on the package group or package.

Permissions Description

- Package group permissions
Querying permissions. A group owner can view the created package group and all packages in the group, and can also view package groups on which they have all permissions.
A package group is a unit. If you select a group when creating a package, you can grant only the permissions of the package group to other users.
- Package permissions
Querying permissions. A package owner can view the created packages, and can also view packages on which they have all permissions.

4.2.3 Creating a Package

DLI allows you to submit program packages in batches to the general-use queue for running.

NOTE

If you need to update a package, you can use the same package or file to upload it to the same location (in the same group) on DLI to overwrite the original package or file.

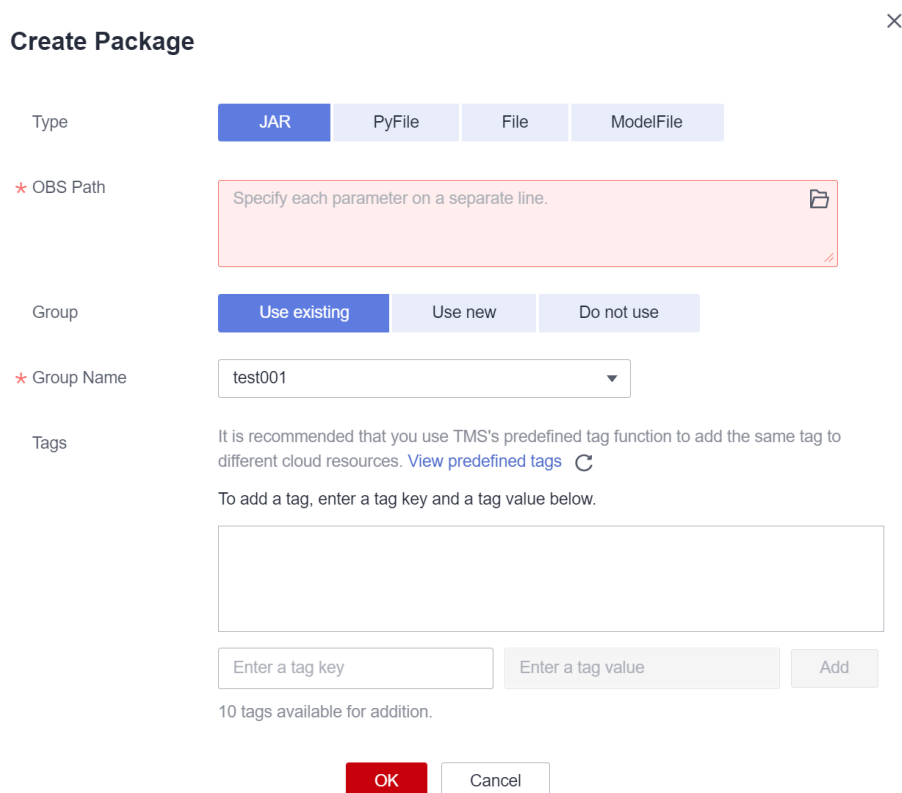
Prerequisites

All software packages must be uploaded to OBS for storage in advance.

Procedure

1. On the left of the management console, choose **Data Management > Package Management**.
2. On the **Package Management** page, click **Create** in the upper right corner to create a package.
3. In the displayed **Create Package** dialog box, set related parameters by referring to [Table 4-17](#).

Figure 4-19 Creating a package




Create Package ×

Type JAR PyFile File ModelFile

* OBS Path

Group Use existing Use new Do not use

* Group Name

Tags
It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#) 
To add a tag, enter a tag key and a tag value below.

10 tags available for addition.

Table 4-17 Parameter description

Parameter	Description
Package Type	The following package types are supported: <ul style="list-style-type: none">● JAR: JAR file● PyFile: User Python file● File: User file● ModelFile: User AI model file
Package File Path	Select the OBS path of the corresponding packages. NOTE <ul style="list-style-type: none">● The packages must be uploaded to OBS for storage in advance.● Only files can be selected.
Group Policy	You can select Use existing group , Use new group , or No grouping .
Group Name	<ul style="list-style-type: none">● Use existing group: Select an existing group.● Use new group: Enter a custom group name.● No grouping: No need to select or enter a group name. NOTE <ul style="list-style-type: none">● If you select a group, the permission management refers to the permissions of the corresponding package group.● If no group is selected, the permission management refers to the permissions of the corresponding package. For details about how to manage permissions on package groups and packages, see Managing Permissions on Packages and Package Groups .
Tag	Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS). For details, see Tag Management Service User Guide . NOTE <ul style="list-style-type: none">● A maximum of 20 tags can be added.● Only one tag value can be added to a tag key.● The key name in each resource must be unique. <ul style="list-style-type: none">● Tag key: Enter a tag key name in the text box. NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_ : = + @) are allowed, but the value cannot start or end with a space or start with _sys_.● Tag value: Enter a tag value in the text box. NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_ : = + @) are allowed. The value cannot start or end with a space.

4. Click **OK**.

After a package is created, you can view and select the package for use on the **Package Management** page.

4.2.4 Deleting a Package

You can delete a package based on actual conditions.

Procedure

1. On the left of the management console, choose **Data Management > Package Management**.
2. Click **Delete** in the **Operation** column of the package to be deleted.
3. In the dialog box that is displayed, click **Yes**.

4.2.5 Modifying the Owner

To change the owner of a package, click **More > Modify Owner** in the **Operation** column of a package on the **Package Management** page.

- If the package has been grouped, you can modify the owner of the **Group** or **Resource** of it.

Figure 4-20 Modifying the package owner

Modify Owner

Group Name

Name

Select Type Group Resource

Username

- If the package has not been grouped, change its owner directly.

Figure 4-21 Modifying the owner of a package

Modify Owner
×

Name

★ Username

OK
Cancel

Table 4-18 Description

Parameter	Description
Group Name	<ul style="list-style-type: none"> If you select a group when creating a package, the name of the group is displayed. If no group is selected when creating a package, this parameter is not displayed.
Name	Name of a package.
Select Type	<ul style="list-style-type: none"> If you select a group when creating a package, you can change the owner of the group or package. If no group is selected when creating a package, this parameter is not displayed.
Username	Name of the package owner. NOTE The username is the name of an existing IAM user.

4.2.6 Built-in Dependencies

DLI built-in dependencies are provided by the platform by default. In case of conflicts, you do not need to upload them when packaging JAR packages of Spark or Flink Jar jobs.

Spark 3.1.1 Dependencies

Table 4-19 Spark 3.1.1 dependencies

Dependency		
accessors-smart-1.2.jar	hive-shims-scheduler-3.1.0-h0.cbu.mrs.321.r10.jar	metrics-graphite-4.1.1.jar

Dependency		
activation-1.1.1.jar	hive-spark-client-3.1.0-h0.cbu.mrs.321.r10.jar	metrics-jmx-4.1.1.jar
aggdesigner-algorithm-6.0.jar	hive-standalone-metastore-3.1.0-h0.cbu.mrs.321.r10.jar	metrics-json-4.1.1.jar
aircompressor-0.16.jar	hive-storage-api-2.7.2.jar	metrics-jvm-4.1.1.jar
algebra_2.12-2.0.0-M2.jar	hive-vector-codegen-3.1.0-h0.cbu.mrs.321.r10.jar	minlog-1.3.0.jar
annotations-17.0.0.jar	hk2-api-2.6.1.jar	netty-3.10.6.Final.jar
ant-1.10.9.jar	hk2-locator-2.6.1.jar	netty-all-4.1.86.Final.jar
ant-launcher-1.10.9.jar	hk2-utils-2.6.1.jar	netty-buffer-4.1.86.Final.jar
antlr4-runtime-4.8-1.jar	hppc-0.7.2.jar	netty-codec-4.1.86.Final.jar
antlr-runtime-3.5.2.jar	httpclient-4.5.6.jar	netty-codec-dns-4.1.86.Final.jar
aopalliance-1.0.jar	httpcore-4.4.10.jar	netty-codec-haproxy-4.1.86.Final.jar
aopalliance-repackaged-2.6.1.jar	istack-commons-runtime-3.0.8.jar	netty-codec-http2-4.1.86.Final.jar
apiguardian-api-1.1.0.jar	ivy-2.5.0.jar	netty-codec-http-4.1.86.Final.jar
arpack_combined_all-0.1.jar	jackson-annotations-2.13.2.jar	netty-codec-memcache-4.1.86.Final.jar
arrow-format-2.0.0.jar	jackson-core-2.13.2.jar	netty-codec-mqtt-4.1.86.Final.jar
arrow-memory-core-2.0.0.jar	jackson-core-asl-1.9.13-atlassian-4.jar	netty-codec-redis-4.1.86.Final.jar
arrow-memory-netty-2.0.0.jar	jackson-databind-2.13.2.jar	netty-codec-smtp-4.1.86.Final.jar
arrow-vector-2.0.0.jar	jackson-dataformat-yaml-2.13.2.jar	netty-codec-socks-4.1.86.Final.jar
asm-5.0.4.jar	jackson-datatype-jsr310-2.11.2.jar	netty-codec-stomp-4.1.86.Final.jar
audience-annotations-0.5.0.jar	jackson-mapper-asl-1.9.13-atlassian-4.jar	netty-codec-xml-4.1.86.Final.jar

Dependency		
automaton-1.11-8.jar	jackson-module-jaxb-annotations-2.13.2.jar	netty-common-4.1.86.Final.jar
avatica-1.22.0.jar	jackson-module-scala_2.12-2.13.2.jar	netty-handler-4.1.86.Final.jar
avatica-core-1.16.0.jar	jaeger-client-1.6.0.jar	netty-handler-proxy-4.1.86.Final.jar
avatica-metrics-1.16.0.jar	jaeger-core-1.6.0.jar	netty-handler-ssl-ocsp-4.1.86.Final.jar
avatica-server-1.16.0.jar	jaeger-thrift-1.6.0.jar	netty-resolver-4.1.86.Final.jar
avro-1.8.2.jar	jaeger-tracerresolver-1.6.0.jar	netty-resolver-dns-4.1.86.Final.jar
avro-ipc-1.8.2.jar	jakarta.activation-api-1.2.1.jar	netty-resolver-dns-classes-macos-4.1.86.Final.jar
avro-mapred-1.8.2.jar	jakarta.annotation-api-1.3.5.jar	netty-resolver-dns-native-macos-4.1.86.Final-osx-aarch_64.jar
java-sdk-bundle-1.11.856.jar	jakarta.el-3.0.3.jar	netty-resolver-dns-native-macos-4.1.86.Final-osx-x86_64.jar
base64-2.3.8.jar	jakarta.el-api-3.0.3.jar	netty-transport-4.1.86.Final.jar
bcpkix-jdk15on-1.69.jar	jakarta.inject-2.6.1.jar	netty-transport-classes-epoll-4.1.86.Final.jar
bcprov-jdk15on-1.69.jar	jakarta.servlet-api-4.0.3.jar	netty-transport-classes-kqueue-4.1.86.Final.jar
bcutil-jdk15on-1.69.jar	jakarta.validation-api-2.0.2.jar	netty-transport-native-epoll-4.1.86.Final-linux-aarch_64.jar
bonecp-0.8.0.RELEASE.jar	jakarta.ws.rs-api-2.1.6.jar	netty-transport-native-epoll-4.1.86.Final-linux-x86_64.jar
breeze_2.12-1.0.jar	jakarta.xml.bind-api-2.3.2.jar	netty-transport-native-kqueue-4.1.86.Final-osx-aarch_64.jar

Dependency		
breeze-macros_2.12-1.0.jar	jamon-runtime-2.4.1.jar	netty-transport-native-kqueue-4.1.86.Final-osx-x86_64.jar
caffeine-2.8.1.jar	janino-3.0.16.jar	netty-transport-native-unix-common-4.1.86.Final.jar
calcite-core-1.22.0.jar	JavaEWAH-0.3.2.jar	netty-transport-rxtx-4.1.86.Final.jar
calcite-druid-1.19.0.jar	java-sdk-core-3.0.12.jar	netty-transport-sctp-4.1.86.Final.jar
calcite-linq4j-1.22.0.jar	javassist-3.25.0-GA.jar	netty-transport-udt-4.1.86.Final.jar
cats-kernel_2.12-2.0.0-M4.jar	javax.activation-api-1.2.0.jar	nimbus-jose-jwt-8.19.jar
checker-qual-3.5.0.jar	javax.annotation-api-1.3.2.jar	objenesis-2.5.1.jar
chill_2.12-0.9.5.jar	javax.inject-1.jar	okhttp-3.14.9.jar
chill-java-0.9.5.jar	javax.jdo-3.2.0-m3.jar	okio-1.17.2.jar
classmate-1.5.1.jar	java-xmlbuilder-1.1.jar	opencsv-2.3.jar
commons-beanutils-1.9.4.jar	javax.servlet-api-3.1.0.jar	opentelemetry-api-1.16.0.jar
commons-cli-1.2.jar	javax.transaction-api-1.3.jar	opentelemetry-context-1.16.0.jar
commons-codec-1.15.jar	javax.ws.rs-api-2.1.1.jar	opentelemetry-semconv-1.16.0-alpha.jar
commons-collections-3.2.2.jar	javalution-5.5.1.jar	opentracing-api-0.33.0.jar
commons-compiler-3.0.16.jar	jaxb-api-2.2.11.jar	opentracing-noop-0.33.0.jar
commons-compress-1.21.jar	jaxb-runtime-2.3.2.jar	opentracing-tracerresolver-0.1.8.jar
commons-configuration2-2.1.1.jar	jboss-logging-3.4.1.Final.jar	opentracing-util-0.33.0.jar
commons-crypto-1.0.0-20191105.jar	jboss-threads-2.3.3.Final.jar	orc-core-1.6.8.jar
commons-daemon-1.0.13.jar	jcip-annotations-1.0-1.jar	orc-mapreduce-1.6.8.jar

Dependency		
commons-dbcp-1.4.jar	jcl-over-slf4j-1.7.36.jar	orc-shims-1.6.8.jar
commons-dbcp2-2.6.0.jar	jcodings-1.0.57.jar	orc-tools-1.6.7-h0.cbu.mrs.321.r10.jar
commons-digester-2.1.jar	jdo-api-3.2.jar	oro-2.0.8.jar
commons-httpclient-3.1.jar	jersey-client-2.34.jar	osgi-resource-locator-1.0.3.jar
commons-io-2.8.0.jar	jersey-common-2.34.jar	paranamer-2.8.jar
commons-lang-2.4.jar	jersey-container-servlet-2.34.jar	parquet-column-1.12.2.jar
commons-lang-2.6.jar	jersey-container-servlet-core-2.34.jar	parquet-common-1.12.2.jar
commons-lang3-3.10.jar	jersey-hk2-2.34.jar	parquet-encoding-1.12.2.jar
commons-logging-1.2.jar	jersey-server-2.34.jar	parquet-format-structures-1.12.2.jar
commons-math3-3.4.1.jar	jets3t-0.9.4-1.0.0.jar	parquet-hadoop-1.12.2.jar
commons-net-3.1.jar	jettison-1.1.jar	parquet-hadoop-bundle-1.12.0-ei-2.0.jar
commons-pool2-2.6.1.jar	jetty-http-9.4.41.v20210516.jar	parquet-jackson-1.12.2.jar
commons-text-1.10.0.jar	jetty-io-9.4.41.v20210516.jar	postgresql-42.3.5.jar
commons-validator-1.7.jar	jetty-rewrite-9.4.43.v20210629.jar	protobuf-java-2.5.0.jar
compress-lzf-1.0.3.jar	jetty-security-9.4.41.v20210516.jar	py4j-0.10.9.jar
core-1.1.2.jar	jetty-server-9.4.41.v20210516.jar	pyrolite-4.30.jar
curator-client-2.13.0.jar	jetty-servlet-9.4.41.v20210516.jar	re2j-1.1.jar
curator-framework-2.13.0.jar	jetty-util-9.4.41.v20210516.jar	RoaringBitmap-0.9.0.jar

Dependency		
curator-recipes-2.13.0.jar	jetty-util-ajax-9.4.41.v20210516.jar	scala-collection-compat_2.12-2.1.1.jar
datanucleus-api-jdo-4.2.4.jar	jetty-webapp-9.4.41.v20210516.jar	scala-compiler-2.12.16.jar
datanucleus-core-4.1.17.jar	jetty-xml-9.4.41.v20210516.jar	scala-library-2.12.16.jar
datanucleus-rdbms-fi-4.1.19-302022.jar	JLargeArrays-1.5.jar	scala-parser-combinators_2.12-1.1.2.jar
derby-10.14.2.0.jar	jline-3.21.0.jar	scala-reflect-2.12.16.jar
disruptor-3.4.2.jar	joda-time-2.10.5.jar	scala-xml_2.12-1.2.0.jar
dli-catalog-client-2.3.7-20240108.090504-101.jar	jodd-core-3.5.2.jar	secComponentApi-1.1.8.jar
dli-catalog-hive3-client-2.3.7-20240108.090513-100.jar	jodd-util-6.0.0.jar	serializer-2.7.2.jar
dli-catalog-hive-extension-2.3.7-20240108.090517-100.jar	joni-2.1.43.jar	shapeless_2.12-2.3.3.jar
dnsjava-2.1.7.jar	jpam-1.1.jar	shims-0.9.0.jar
dropwizard-metrics-hadoop-metrics2-reporter-0.1.2.jar	jsch-0.1.72.jar	sketches-core-0.9.0.jar
error_prone_annotations-2.18.0.jar	json-20210307.jar	slf4j-api-1.7.30.jar
esdk-obs-java-optimised-3.22.10.2.jar	json4s-ast_2.12-3.7.0-M5.jar	slf4j-log4j12-1.7.25.jar
esri-geometry-api-2.2.0.jar	json4s-core_2.12-3.7.0-M5.jar	snakeyaml-1.30.jar
fastutil-6.5.6.jar	json4s-jackson_2.12-3.7.0-M5.jar	snappy-java-1.1.8.2.jar
flatbuffers-java-1.9.0.jar	json4s-scalap_2.12-3.7.0-M5.jar	spark-avro_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
generex-1.0.2.jar	json-path-2.4.0.jar	spark-catalyst_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
glassfish-corba-omgapi-4.2.2.jar	json-smart-2.3.jar	spark-core_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar

Dependency		
gson-2.8.9.jar	jsr305-3.0.0.jar	spark-graphx_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
gson-fire-1.8.5.jar	JTransforms-3.1.jar	spark-hive_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
guava-14.0.1.jar	jul-to-slf4j-1.7.36.jar	spark-kubernetes_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
guice-3.0.jar	kafka-clients-2.8.0.jar	spark-kvstore_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
guice-assistedinject-3.0.jar	kerb-admin-2.0.2.jar	spark-launcher_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
guice-servlet-4.0.jar	kerb-client-2.0.2.jar	spark-mllib_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-annotations-3.1.1-h0.cbu.mrs.313.r9.jar	kerb-common-2.0.2.jar	spark-mllib-local_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-archives-3.3.1-h0.cbu.mrs.321.r10.jar	kerb-core-2.0.2.jar	spark-network-common_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-auth-3.3.1-h0.cbu.mrs.321.r16.jar	kerb-crypto-2.0.2.jar	spark-network-shuffle_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-3.3.1-h0.cbu.mrs.321.r16.jar	kerb-identity-2.0.2.jar	spark-quota-manager_2.12-3.1.1-2.3.7.dli-SNAPSHOT.jar
hadoop-client-3.1.1-h0.cbu.mrs.313.r9.jar	kerb-server-2.0.2.jar	spark-repl_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-common-3.3.1-h0.cbu.mrs.321.r10.jar	kerb-simplekdc-2.0.2.jar	spark-sketch_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-distcp-3.3.1-h0.cbu.mrs.321.r10.jar	kerb-util-2.0.2.jar	spark-sql_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-hdfs-3.3.1-h0.cbu.mrs.321.r16.jar	kerby-asn1-2.0.2.jar	spark-streaming_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-hdfs-client-3.3.1-h0.cbu.mrs.321.r10.jar	kerby-config-2.0.2.jar	spark-tags_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar
hadoop-3.1.1-52.1.jar	kerby-pkix-2.0.2.jar	spark-unsafe_2.12-3.1.1-h1.cbu.dli.20230607.r1.jar

Dependency		
hadoop-mapreduce-client-common-3.1.1-h0.cbu.mrs.313.r9.jar	kerby-util-2.0.2.jar	spark-query_2.12-3.1.1-2.3.7.dli-SNAPSHOT.jar
hadoop-mapreduce-client-core-3.1.1-h0.cbu.mrs.313.r9.jar	kerby-xdr-2.0.2.jar	spire_2.12-0.17.0-M1.jar
hadoop-mapreduce-client-jobclient-3.1.1-h0.cbu.mrs.313.r9.jar	kotlin-stdlib-1.4.21.jar	spire-macros_2.12-0.17.0-M1.jar
hadoop-mapreduce-client-nativetask-3.3.1-h0.cbu.mrs.321.r10.jar	kotlin-stdlib-common-1.4.21.jar	spire-platform_2.12-0.17.0-M1.jar
hadoop-registry-3.3.1-h0.cbu.mrs.321.r10.jar	kryo-shaded-4.0.2.jar	spire-util_2.12-0.17.0-M1.jar
hadoop-shaded-guava-1.1.1.jar	kubernetes-client-5.4.1-20211025.jar	sqlline-1.3.0.jar
hadoop-shaded-protobuf_3_7-1.1.1.jar	kubernetes-model-admissionregistration-5.4.1-20211025.jar	ST4-4.0.4.jar
hadoop-yarn-api-3.1.1-h0.cbu.mrs.313.r9.jar	kubernetes-model-apiextensions-5.4.1-20211025.jar	stax2-api-4.2.1.jar
hadoop-yarn-client-3.1.1-h0.cbu.mrs.313.r9.jar	kubernetes-model-apps-5.4.1-20211025.jar	stax-api-1.0.1.jar
hadoop-yarn-registry-3.3.1-h0.cbu.mrs.321.r10.jar	kubernetes-model-autoscaling-5.4.1-20211025.jar	stream-2.9.6.jar
hbase-asyncfs-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-batch-5.4.1-20211025.jar	streamingClient
hbase-client-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-certificates-5.4.1-20211025.jar	streamingClient010
hbase-common-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-common-5.4.1-20211025.jar	swagger-annotations-2.2.8.jar
hbase-hadoop2-compat-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-coordination-5.4.1-20211025.jar	tephra-api-0.6.0.jar

Dependency		
hbase-hadoop-compat-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-core-5.4.1-20211025.jar	tephra-core-0.6.0.jar
hbase-http-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-discovery-5.4.1-20211025.jar	tephra-hbase-compat-1.0-0.6.0.jar
hbase-logging-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-events-5.4.1-20211025.jar	threetenbp-1.3.5.jar
hbase-metrics-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-extensions-5.4.1-20211025.jar	threeten-extra-1.5.0.jar
hbase-metrics-api-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-flowcontrol-5.4.1-20211025.jar	tink-1.6.0.jar
hbase-procedure-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-metrics-5.4.1-20211025.jar	token-provider-2.0.2.jar
hbase-protocol-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-networking-5.4.1-20211025.jar	tomcat-servlet-api-8.5.61.jar
hbase-protocol-shaded-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-node-5.4.1-20211025.jar	transaction-api-1.1.jar
hbase-replication-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-policy-5.4.1-20211025.jar	twill-api-0.6.0-incubating.jar
hbase-server-2.4.14-h0.cbu.mrs.321.r10.jar	kubernetes-model-rbac-5.4.1-20211025.jar	twill-common-0.6.0-incubating.jar
hbase-shaded-gson-4.1.4.jar	kubernetes-model-scheduling-5.4.1-20211025.jar	twill-core-0.6.0-incubating.jar
hbase-shaded-jersey-4.1.4.jar	kubernetes-model-storageclass-5.4.1-20211025.jar	twill-discovery-api-0.6.0-incubating.jar
hbase-shaded-jetty-4.1.4.jar	leveldbjni-all-1.8-20191105.jar	twill-discovery-core-0.6.0-incubating.jar
hbase-shaded-miscellaneous-4.1.4.jar	libfb303-0.9.3.jar	twill-zookeeper-0.6.0-incubating.jar
hbase-shaded-netty-4.1.4.jar	libthrift-0.14.1-ei-311001.jar	univocity-parsers-2.9.1.jar

Dependency		
hbase-shaded-protobuf-4.1.4.jar	log4j-1.2.17-cloudera1.jar	us-common-1.0.66.jar
hbase-unsafe-4.1.4.jar	log4j-api-2.17.1.jar	velocity-1.7.jar
hbase-zookeeper-2.4.14-h0.cbu.mrs.321.r10.jar	log4j-rolling-appender-20131024-2017.jar	velocity-engine-core-2.3.jar
hibernate-validator-6.2.5.Final.jar	logging-interceptor-3.14.9.jar	wildfly-client-config-1.0.1.Final.jar
HikariCP-2.6.1.jar	luxor-cluster-quota-manager-transport_2.12-2.3.7-20231226.034700-559.jar	wildfly-common-1.5.2.Final.jar
hive-classification-3.1.0-h0.cbu.mrs.321.r10.jar	luxor-encrypt-2.3.7-20231226.034423-1046.jar	woodstox-core-5.4.0.jar
hive-common-3.1.0-h0.cbu.mrs.321.r10.jar	luxor-fs3-2.3.7-20231226.034438-1039.jar	xalan-2.7.2.jar
hive-exec-3.1.0-h0.cbu.mrs.321.r10-core.jar	luxor-obs-fs3-2.3.7-20231226.034443-1038.jar	xbean-asm7-shaded-4.15.jar
hive-llap-client-2.3.3-ei-12-20210120.005053-2.jar	luxor-rpc_2.12-2.3.7-20231226.034653-560.jar	xercesImpl-2.12.2.jar
hive-llap-common-3.1.0-h0.cbu.mrs.321.r10.jar	luxor-scc-adapter-2.3.7-20231226.034418-1045.jar	xml-apis-1.4.01.jar
hive-llap-tez-3.1.0-h0.cbu.mrs.321.r10.jar	luxor-transport-2.3.7-20231226.034433-1038.jar	xnio-api-3.8.4.Final.jar
hive-metastore-3.1.0-h0.cbu.mrs.321.r10.jar	lz4-java-1.7.1.jar	xz-1.5.jar
hive-serde-3.1.0-h0.cbu.mrs.321.r10.jar	machinist_2.12-0.6.8.jar	zjsonpatch-0.3.0.jar
hive-service-rpc-3.1.0-h0.cbu.mrs.321.r10.jar	macro-compat_2.12-1.1.1.jar	zookeeper-3.5.6-ei-302002.jar
hive-shims-0.23-3.1.0-h0.cbu.mrs.321.r10.jar	memarts-ccsdk-1.0.jar	zookeeper-jute-3.5.6-ei-302002.jar
hive-shims-3.1.0-h0.cbu.mrs.321.r10.jar	memory-0.9.0.jar	zstd-jni-1.4.9-1.jar

Dependency		
hive-shims-common-3.1.0-h0.cbu.mrs.321.r10.jar	metrics-core-4.1.1.jar	-

Spark 2.4.5 Dependencies

Table 4-20 Spark 2.4.5 dependencies

Dependency		
JavaEWAH-1.1.7.jar	httpClient-4.5.6.jar	lucene-queryparser-7.7.2.jar
RoaringBitmap-0.7.45.jar	httpcore-4.4.10.jar	lucene-sandbox-7.7.2.jar
ST4-4.3.1.jar	ivy-2.4.0.jar	luxor-encrypt-2.0.0-20220623.010726-213.jar
accessors-smart-1.2.jar	jackson-annotations-2.11.4.jar	luxor-fs3-2.0.0-20220623.010750-209.jar
activation-1.1.1.jar	jackson-core-2.11.4.jar	luxor-obs-fs3-2.0.0-20220623.010756-209.jar
aircompressor-0.16.jar	jackson-core-asl-1.9.13-atlassian-4.jar	luxor-rpc_2.11-2.0.0-20220623.010737-182.jar
alluxio-2.3.1-luxor-SNAPSHOT-client.jar	jackson-databind-2.11.4.jar	luxor-transport-2.0.0-20220623.010744-71.jar
annotations-17.0.0.jar	jackson-dataformat-yaml-2.11.4.jar	lz4-java-1.7.1.jar
antlr-2.7.7.jar	jackson-datatype-jsr310-2.11.2.jar	machinist_2.11-0.6.1.jar
antlr-runtime-3.4.jar	jackson-jaxrs-base-2.10.3.jar	macro-compat_2.11-1.1.1.jar
antlr4-runtime-4.8-1.jar	jackson-jaxrs-json-provider-2.10.3.jar	metrics-core-3.1.5.jar
aopalliance-1.0.jar	jackson-mapper-asl-1.9.13-atlassian-4.jar	metrics-graphite-3.1.5.jar

Dependency		
aopalliance-repackaged-2.4.0-b34.jar	jackson-module-jaxb-annotations-2.10.3.jar	metrics-jmx-4.1.12.1.jar
apache-log4j-extras-1.2.17.jar	jackson-module-paranamer-2.11.4.jar	metrics-json-3.1.5.jar
arpack_combined_all-0.1.jar	jackson-module-scala_2.11-2.11.4.jar	metrics-jvm-3.1.5.jar
arrow-format-0.12.0.jar	jakarta.activation-api-1.2.1.jar	minlog-1.3.0.jar
arrow-memory-0.12.0.jar	jakarta.xml.bind-api-2.3.2.jar	mssql-jdbc-6.2.1.jre7.jar
arrow-vector-0.12.0.jar	janino-3.0.9.jar	netty-all-4.1.51.Final.jar
asm-5.0.4.jar	java-util-1.9.0.jar	nimbus-jose-jwt-8.19.jar
audience-annotations-0.5.0.jar	java-xmlbuilder-1.1.jar	objenesis-2.5.1.jar
automaton-1.11-8.jar	javassist-3.18.1-GA.jar	okhttp-3.14.9.jar
avro-1.8.2.jar	javax.annotation-api-1.2.jar	okio-1.17.2.jar
avro-ipc-1.8.2.jar	javax.inject-1.jar	opencsv-2.3.jar
avro-mapred-1.8.2.jar	javax.inject-2.4.0-b34.jar	opencsv-4.6.jar
java-sdk-bundle-1.11.856.jar	javax.servlet-api-3.1.0.jar	opencv-4.3.0-2.jar
base64-2.3.8.jar	javax.ws.rs-api-2.0.1.jar	orc-core-1.6.8-nohive.jar
bcpkix-jdk15on-1.66.jar	javalution-5.3.1.jar	orc-mapreduce-1.6.8-nohive.jar
bcprov-jdk15on-1.67.jar	jaxb-api-2.2.11.jar	orc-shims-1.6.8.jar
bonecp-0.8.0.RELEASE.jar	jcip-annotations-1.0-1.jar	oro-2.0.8.jar
breeze-macros_2.11-0.13.2.jar	jcl-over-slf4j-1.7.30.jar	osgi-resource-locator-1.0.1.jar
breeze_2.11-0.13.2.jar	jdo-api-3.0.1.jar	paranamer-2.8.jar
calcite-avatica-1.2.0-incubating.jar	jersey-client-2.23.1.jar	parquet-column-1.12.2.jar
chill-java-0.9.3.jar	jersey-common-2.23.1.jar	parquet-common-1.12.2.jar

Dependency		
chill_2.11-0.9.3.jar	jersey-container-servlet-2.23.1.jar	parquet-encoding-1.12.2.jar
commons-beanutils-1.9.4.jar	jersey-container-servlet-core-2.23.1.jar	parquet-format-structures-1.12.2.jar
commons-cli-1.2.jar	jersey-guava-2.23.1.jar	parquet-hadoop-1.12.2.jar
commons-codec-1.15.jar	jersey-media-jaxb-2.23.1.jar	parquet-hadoop-bundle-1.6.0.jar
commons-collections-3.2.2.jar	jersey-server-2.23.1.jar	parquet-jackson-1.12.2.jar
commons-collections4-4.2.jar	jets3t-0.9.4.jar	postgresql-42.2.14.jar
commons-compiler-3.0.9.jar	jettison-1.1.jar	protobuf-java-2.5.0.jar
commons-compress-1.4.1.jar	jetty-http-9.4.34.v20201102.jar	py4j-0.10.7.jar
commons-configuration2-2.1.1.jar	jetty-io-9.4.34.v20201102.jar	pyrolite-4.13.jar
commons-crypto-1.0.0-20191105.jar	jetty-security-9.4.34.v20201102.jar	re2j-1.1.jar
commons-daemon-1.0.13.jar	jetty-server-9.4.34.v20201102.jar	scala-compiler-2.11.12.jar
commons-dbcp2-2.7.0.jar	jetty-servlet-9.4.34.v20201102.jar	scala-library-2.11.12.jar
commons-httpclient-3.1.jar	jetty-util-9.4.34.v20201102.jar	scala-parser-combinators_2.11-1.1.2.jar
commons-io-2.5.jar	jetty-util-ajax-9.4.34.v20201102.jar	scala-reflect-2.11.12.jar
commons-lang-2.6.jar	jetty-webapp-9.4.34.v20201102.jar	scala-xml_2.11-1.0.5.jar
commons-lang3-3.5.jar	jetty-xml-9.4.34.v20201102.jar	secComponentApi-1.0.6.jar
commons-logging-1.2.jar	joda-time-2.9.3.jar	shapeless_2.11-2.3.2.jar
commons-math3-3.4.1.jar	jodd-core-3.5.2.jar	shims-0.7.45.jar

Dependency		
commons-net-3.1.jar	json-20200518.jar	slf4j-api-1.7.30.jar
commons-pool2-2.8.0.jar	json-io-2.5.1.jar	slf4j-log4j12-1.7.30.jar
commons-text-1.3.jar	json-sanitizer-1.2.1.jar	snakeyaml-1.26.jar
compress-lzf-1.0.3.jar	json-smart-2.3.jar	snappy-java-1.1.8.2.jar
core-1.1.2.jar	json4s-ast_2.11-3.5.3.jar	solr-core-7.7.2.jar
crypter-0.0.6.jar	json4s-core_2.11-3.5.3.jar	solr-solrj-7.7.2.jar
curator-client-4.2.0.jar	json4s-jackson_2.11-3.5.3.jar	spark-avro_2.11-2.4.5.0100-2.0.0.dli-20220617.085536-9.jar
curator-framework-4.2.0.jar	json4s-scalap_2.11-3.5.3.jar	spark-avro_2.11-4.0.0.jar
curator-recipes-2.7.1.jar	jsp-api-2.1.jar	spark-catalyst_2.11-2.4.5.0100-2.0.0.dli-20220617.085405-16.jar
datanucleus-api-jdo-3.2.6.jar	jsr305-1.3.9.jar	spark-core_2.11-2.4.5.0100-2.0.0.dli-20220617.085327-16.jar
datanucleus-core-3.2.10.jar	jta-1.1.jar	spark-graphx_2.11-2.4.5.0100-.0.dli-20220617.085336-16.jar
datanucleus-rdbms-3.2.9.jar	jtransforms-2.4.0.jar	spark-hive_2.11-2.4.5.0100-2.0.0.dli-20220617.085423-16.jar
derby-10.14.2.0.jar	jts-core-1.16.1.jar	spark-kubernetes_2.11-2.4.5.0100-2.0.0.dli-20220617.085519-16.jar
dnsjava-2.1.7.jar	jul-to-slf4j-1.7.30.jar	spark-kvstore_2.11-2.4.5.0100-2.0.0.dli-20220617.085249-16.jar
ecj-3.21.0.jar	junit-4.11.jar	spark-launcher_2.11-2.4.5.0100-2.0.0.dli-20220617.085435-16.jar

Dependency		
ehcache-3.3.1.jar	kerb-admin-1.0.1.jar	spark-mllib-local_2.11-2.4.5.0100-2.0.0.dli-20220617.085349-16.jar
expiringmap-0.5.9.jar	kerb-client-1.0.1.jar	spark-mllib_2.11-2.4.5.0100-2.0.0.dli-20220617.085342-16.jar
fastutil-8.2.3.jar	kerb-common-1.0.1.jar	spark-network-common_2.11-2.4.5.0100-2.0.0.dli-20220617.085254-16.jar
flatbuffers-java-1.9.0.jar	kerb-core-1.0.1.jar	spark-network-shuffle_2.11-2.4.5.0100-2.0.0.dli-20220617.085300-16.jar
fst-2.50.jar	kerb-crypto-1.0.1.jar	spark-om_2.11-2.4.5.0100-2.0.0.dli-20220617.085316-16.jar
generex-1.0.2.jar	kerb-identity-1.0.1.jar	spark-repl_2.11-2.4.5.0100-2.0.0.dli-20220617.085430-16.jar
geronimo-jcache_1.0_spec-1.0-alpha-1.jar	kerb-server-1.0.1.jar	spark-sketch_2.11-2.4.5.0100-2.0.0.dli-20220617.085243-16.jar
gson-2.2.4.jar	kerb-simplekdc-1.0.1.jar	spark-sql_2.11-2.4.5.0100-2.0.0.dli-20220617.085414-16.jar
guava-14.0.1.jar	kerb-util-1.0.1.jar	spark-streaming_2.11-2.4.5.0100-2.0.0.dli-20220617.085359-16.jar
guice-4.0.jar	kerby-asn1-1.0.1.jar	spark-tags_2.11-2.4.5.0100-2.0.0.dli-20220617.085322-16.jar

Dependency		
guice-servlet-4.0.jar	kerby-config-1.0.1.jar	spark-unsafe_2.11-2.4.5.0100-2.0.0.dli-20220617.085311-16.jar
hadoop-annotations-3.1.1-ei-302002.jar	kerby-pkix-1.0.1.jar	spark-uquery_2.11-2.4.5.0100-2.0.0.dli-SNAPSHOT.jar
hadoop-auth-3.1.1-ei-302002.jar	kerby-util-1.0.1.jar	spark-yarn_2.11-2.4.5.0100-2.0.0.dli-20220617.085531-16.jar
hadoop-3.1.1-ei-302002.jar	kerby-xdr-1.0.1.jar	spire-macros_2.11-0.13.0.jar
hadoop-client-3.1.1-ei-302002.jar	kryo-shaded-4.0.2.jar	spire_2.11-0.13.0.jar
hadoop-common-3.1.1-ei-302002.jar	kubernetes-client-5.4.1-20211025.jar	stax-api-1.0-2.jar
hadoop-hdfs-3.1.1-ei-302002.jar	kubernetes-model-admissionregistration-5.4.1-20211025.jar	stax2-api-3.1.4.jar
hadoop-hdfs-client-3.1.1-ei-302002.jar	kubernetes-model-apiextensions-5.4.1-20211025.jar	stream-2.7.0.jar
hadoop-3.1.1-46.jar	kubernetes-model-apps-5.4.1-20211025.jar	stringtemplate-3.2.1.jar
hadoop-mapreduce-client-common-3.1.1-ei-302002.jar	kubernetes-model-autoscaling-5.4.1-20211025.jar	threeten-extra-1.5.0.jar
hadoop-mapreduce-client-core-3.1.1-ei-302002.jar	kubernetes-model-batch-5.4.1-20211025.jar	tink-1.6.0.jar
hadoop-mapreduce-client-jobclient-3.1.1-ei-302002.jar	kubernetes-model-certificates-5.4.1-20211025.jar	token-provider-1.0.1.jar
hadoop-minikdc-3.1.1-ei-302002.jar	kubernetes-model-common-5.4.1-20211025.jar	tomcat-api-9.0.39.jar
hadoop-yarn-api-3.1.1-ei-302002.jar	kubernetes-model-coordination-5.4.1-20211025.jar	zookeeper-jute-3.5.6-ei-302002.jar

Dependency		
hadoop-yarn-client-3.1.1-ei-302002.jar	kubernetes-model-core-5.4.1-20211025.jar	tomcat-el-api-9.0.39.jar
hadoop-yarn-common-3.1.1-ei-302002.jar	kubernetes-model-discovery-5.4.1-20211025.jar	tomcat-jasper-9.0.39.jar
hadoop-yarn-registry-3.1.1-ei-302002.jar	kubernetes-model-events-5.4.1-20211025.jar	tomcat-jasper-el-9.0.39.jar
hadoop-yarn-server-applicationhistoryservice-3.1.1-ei-302002.jar	kubernetes-model-extensions-5.4.1-20211025.jar	tomcat-jsp-api-9.0.39.jar
hadoop-yarn-server-common-3.1.1-ei-302002.jar	kubernetes-model-flowcontrol-5.4.1-20211025.jar	tomcat-juli-9.0.39.jar
hadoop-yarn-server-resourcemanager-3.1.1-ei-302002.jar	kubernetes-model-metrics-5.4.1-20211025.jar	tomcat-servlet-api-9.0.39.jar
hadoop-yarn-server-web-proxy-3.1.1-ei-302002.jar	kubernetes-model-networking-5.4.1-20211025.jar	tomcat-util-9.0.39.jar
hamcrest-core-1.3.jar	kubernetes-model-node-5.4.1-20211025.jar	tomcat-util-scan-9.0.39.jar
hive-common-1.2.1-2.0.0.dli-20220528.090500-402.jar	kubernetes-model-policy-5.4.1-20211025.jar	univocity-parsers-2.7.3.jar
hive-exec-1.2.1-2.0.0.dli-20220528.090521-401.jar	kubernetes-model-rbac-5.4.1-20211025.jar	zstd-jni-1.4.9-1.jar
hive-metastore-1.2.1-2.0.0.dli-20220528.090509-402.jar	kubernetes-model-scheduling-5.4.1-20211025.jar	validation-api-1.1.0.Final.jar
hive-shims-0.23-1.2.1-2.0.0.dli-20220528.090445-403.jar	kubernetes-model-storageclass-5.4.1-20211025.jar	velocity-1.7.jar
hive-shims-1.2.1-2.0.0.dli-20220528.090455-403.jar	leveldbjni-all-1.8-20191105.jar	woodstox-core-5.0.3.jar

Dependency		
hive-shims-common-1.2.1-2.0.0.dli-20220528.090441-404.jar	libfb303-0.9.3.jar	xbean-asm6-shaded-4.8.jar
hive-shims-scheduler-1.2.1-2.0.0.dli-20220528.090450-403.jar	libthrift-0.12.0.jar	xercesImpl-2.12.0.jar
hk2-api-2.4.0-b34.jar	log4j-1.2.17-cloudera1.jar	xml-apis-1.4.01.jar
hk2-locator-2.4.0-b34.jar	log4j-rolling-appender-20131024-2017.jar	xz-1.0.jar
hk2-utils-2.4.0-b34.jar	logging-interceptor-3.14.9.jar	zjsonpatch-0.3.0.jar
hppc-0.7.2.jar	lucene-analyzers-common-7.7.2.jar	zookeeper-3.5.6-ei-302002.jar
htrace-core4-4.2.0-incubating-1.0.0.jar	lucene-core-7.7.2.jar	-

Spark 2.3.2 Dependencies

Table 4-21 Spark 2.3.2 dependencies

Dependency		
accessors-smart-1.2.jar	HikariCP-java7-2.4.12.jar	logging-interceptor-3.14.4.jar
activation-1.1.1.jar	hive-common-1.2.1-2.1.0.dli-20201111.064115-91.jar	luxor-encrypt-2.1.0-20201106.065437-53.jar
aircompressor-0.8.jar	hive-exec-1.2.1-2.1.0.dli-20201111.064444-91.jar	luxor-fs3-2.1.0-20201106.065612-53.jar
alluxio-2.3.1-luxor-SNAPSHOT-client.jar	hive-metastore-1.2.1-2.1.0.dli-20201111.064230-91.jar	luxor-obs-fs3-2.1.0-20201106.065616-53.jar
antlr-2.7.7.jar	hk2-api-2.4.0-b34.jar	luxor-rpc_2.11-2.1.0-20201106.065541-53.jar

Dependency		
antlr4-runtime-4.8-1.jar	hk2-locator-2.4.0-b34.jar	luxor-rpc-protobuf2-2.1.0-20201106.065551-53.jar
antlr-runtime-3.4.jar	hk2-utils-2.4.0-b34.jar	lz4-java-1.7.1.jar
aopalliance-1.0.jar	hppc-0.7.2.jar	machinist_2.11-0.6.1.jar
aopalliance-repackaged-2.4.0-b34.jar	htrace-core4-4.2.0-incubating-1.0.0.jar	macro-compat_2.11-1.1.1.jar
apache-log4j-extras-1.2.17.jar	httpClient-4.5.4.jar	metrics-core-3.1.5.jar
arpack_combined_all-0.1.jar	httpcore-4.4.7.jar	metrics-graphite-3.1.5.jar
arrow-format-0.8.0.jar	ivy-2.4.0.jar	metrics-jmx-4.1.12.1.jar
arrow-memory-0.8.0.jar	j2objc-annotations-1.3.jar	metrics-json-3.1.5.jar
arrow-vector-0.8.0.jar	jackson-annotations-2.10.0.jar	metrics-jvm-3.1.5.jar
asm-5.0.4.jar	jackson-core-2.10.0.jar	minlog-1.3.0.jar
audience-annotations-0.5.0.jar	jackson-core-asl-1.9.13-atlassian-4.jar	mssql-jdbc-6.2.1.jre7.jar
automaton-1.11-8.jar	jackson-databind-2.10.0.jar	netty-3.10.6.Final.jar
avro-1.7.7.jar	jackson-dataformat-yaml-2.10.0.jar	netty-all-4.1.51.Final.jar
avro-ipc-1.7.7.jar	jackson-datatype-jsr310-2.10.3.jar	nimbus-jose-jwt-8.19.jar
avro-ipc-1.7.7-tests.jar	jackson-jaxrs-base-2.10.3.jar	objenesis-2.1.jar
avro-mapred-1.7.7-hadoop2.jar	jackson-jaxrs-json-provider-2.10.3.jar	okhttp-3.14.4.jar
java-sdk-bundle-1.11.271.jar	jackson-mapper-asl-1.9.13-atlassian-4.jar	okio-1.17.2.jar
base64-2.3.8.jar	jackson-module-jaxb-annotations-2.10.3.jar	opencsv-2.3.jar
bcpxix-jdk15on-1.66.jar	jackson-module-paranamer-2.10.0.jar	opencsv-4.6.jar
bcprov-jdk15on-1.66.jar	jackson-module-scala_2.11-2.10.0.jar	opencv-4.3.0-2.jar

Dependency		
bonecp-0.8.0.RELEASE.jar	jakarta.activation-api-1.2.1.jar	orc-core-1.4.4-nohive.jar
breeze_2.11-0.13.2.jar	jakarta.xml.bind-api-2.3.2.jar	orc-mapreduce-1.4.4-nohive.jar
breeze-macros_2.11-0.13.2.jar	janino-3.0.8.jar	oro-2.0.8.jar
calcite-avatica-1.2.0-incubating.jar	javacpp-1.5.4.jar	osgi-resource-locator-1.0.1.jar
calcite-core-1.2.0-incubating.jar	javacpp-1.5.4-linux-x86_64.jar	paranamer-2.8.jar
calcite-linq4j-1.2.0-incubating.jar	javacv-1.5.4.jar	parquet-column-1.8.3.jar
checker-qual-2.11.1.jar	JavaEWAH-1.1.7.jar	parquet-common-1.8.3.jar
chill_2.11-0.8.4.jar	javassist-3.18.1-GA.jar	parquet-encoding-1.8.3.jar
chill-java-0.8.4.jar	javax.annotation-api-1.2.jar	parquet-format-2.3.1.jar
commons-beanutils-1.9.4.jar	javax.inject-1.jar	parquet-hadoop-1.8.3.jar
commons-cli-1.2.jar	javax.inject-2.4.0-b34.jar	parquet-hadoop-bundle-1.6.0.jar
commons-codec-2.0-20130428.202122-59.jar	javax.servlet-api-3.1.0.jar	parquet-jackson-1.8.3.jar
commons-collections-3.2.2.jar	javax.ws.rs-api-2.0.1.jar	parquet-format-2.3.1.jar
commons-collections4-4.2.jar	java-xmlbuilder-1.1.jar	parquet-hadoop-1.8.3.jar
commons-compiler-3.0.8.jar	javalution-5.3.1.jar	parquet-hadoop-bundle-1.6.0.jar
commons-compress-1.4.1.jar	jaxb-api-2.2.11.jar	parquet-jackson-1.8.3.jar
commons-configuration2-2.1.1.jar	jcip-annotations-1.0-1.jar	postgresql-42.2.14.jar
commons-crypto-1.0.0-20191105.jar	jcl-over-slf4j-1.7.26.jar	protobuf-java-2.5.0.jar

Dependency		
commons-daemon-1.0.13.jar	jdo-api-3.0.1.jar	py4j-0.10.7.jar
commons-dbcp-1.4.jar	jersey-client-2.23.1.jar	pyrolite-4.13.jar
commons-dbcp2-2.7.0.jar	jersey-common-2.23.1.jar	re2j-1.1.jar
commons-httpclient-3.1.jar	jersey-container-servlet-2.23.1.jar	RoaringBitmap-0.5.11.jar
commons-io-2.5.jar	jersey-container-servlet-core-2.23.1.jar	scala-compiler-2.11.12.jar
commons-lang-2.6.jar	jersey-guava-2.23.1.jar	scala-library-2.11.12.jar
commons-lang3-3.5.jar	jersey-media-jaxb-2.23.1.jar	scalap-2.11.0.jar
commons-logging-1.2.jar	jersey-server-2.23.1.jar	scala-parser-combinators_2.11-1.1.0.jar
commons-math3-3.4.1.jar	jets3t-0.9.4.jar	scala-reflect-2.11.12.jar
commons-net-2.2.jar	jetty-http-9.4.31.v20200723.jar	scala-xml_2.11-1.0.5.jar
commons-pool-1.5.4.jar	jetty-io-9.4.31.v20200723.jar	secComponentApi-1.0.5c.jar
commons-pool2-2.8.0.jar	jetty-security-9.4.31.v20200723.jar	shapeless_2.11-2.3.2.jar
commons-text-1.3.jar	jetty-server-9.4.31.v20200723.jar	slf4j-api-1.7.30.jar
compress-lzf-1.0.3.jar	jetty-servlet-9.4.31.v20200723.jar	slf4j-log4j12-1.7.30.jar
core-1.1.2.jar	jetty-util-9.4.31.v20200723.jar	snakeyaml-1.24.jar
curator-client-4.2.0.jar	jetty-util-ajax-9.4.31.v20200723.jar	snappy-java-1.1.7.5.jar
curator-framework-4.2.0.jar	jetty-webapp-9.4.31.v20200723.jar	spark-catalyst_2.11-2.3.2.0101-2.1.0.dli-20201111.073826-143.jar

Dependency		
curator-recipes-2.7.1.jar	jetty-xml-9.4.31.v20200723.jar	spark-core_2.11-2.3.2.0101-.0.dli-20201111.073836-134.jar
datanucleus-api-jdo-3.2.6.jar	joda-time-2.9.3.jar	spark-graphx_2.11-2.3.2.0101-2.1.0.dli-20201111.073847-129.jar
datanucleus-core-3.2.10.jar	jodd-core-4.2.0.jar	spark-hive_2.11-2.3.2.0101-.0.dli-20201111.073854-132.jar
datanucleus-rdbms-3.2.9.jar	json-20200518.jar	spark-kubernetes_2.11-2.3.2.0101-2.1.0.dli-20201111.073916-85.jar
derby-10.12.1.1.jar	json4s-ast_2.11-3.2.11.jar	spark-kvstore_2.11-2.3.2.0101-2.1.0.dli-20201111.073933-127.jar
dnsjava-2.1.7.jar	json4s-core_2.11-3.2.11.jar	spark-launcher_2.11-2.3.2.0101-2.1.0.dli-20201111.073940-127.jar
ehcache-3.3.1.jar	json4s-jackson_2.11-3.2.11.jar	spark-mllib_2.11-2.3.2.0101-2.1.0.dli-20201111.073946-127.jar
eigenbase-properties-1.1.5.jar	json-sanitizer-1.2.1.jar	spark-mllib-local_2.11-2.3.2.0101-2.1.0.dli-20201111.073953-127.jar
error_prone_annotations-2.3.4.jar	json-smart-2.3.jar	spark-network-common_2.11-2.3.2.0101-2.1.0.dli-20201111.073959-127.jar
failureaccess-1.0.1.jar	jsp-api-2.1.jar	spark-network-shuffle_2.11-2.3.2.0101-2.1.0.dli-20201111.074007-127.jar

Dependency		
fastutil-8.2.3.jar	jsr305-3.0.2.jar	spark-om_2.11-2.3.2.0101-.0.dli-20201111.074019-125.jar
ffmpeg-4.3.1-1.5.4.jar	jta-1.1.jar	spark-repl_2.11-2.3.2.0101-2.1.0.dli-20201111.074028-125.jar
ffmpeg-4.3.1-1.5.4-linux-x86_64.jar	jtransforms-2.4.0.jar	spark-sketch_2.11-2.3.2.0101-2.1.0.dli-20201111.074035-125.jar
flatbuffers-1.2.0-3f79e055.jar	jul-to-slf4j-1.7.26.jar	spark-sql_2.11-2.3.2.0101-2.1.0.dli-20201111.074041-126.jar
generex-1.0.2.jar	junit-4.11.jar	spark-streaming_2.11-2.3.2.0101-2.1.0.dli-20201111.074100-123.jar
geronimo-jcache_1.0_spec-1.0-alpha-1.jar	kerb-admin-1.0.1.jar	spark-tags_2.11-2.3.2.0101-2.1.0.dli-20201111.074136-123.jar
gson-2.2.4.jar	kerb-client-1.0.1.jar	spark-tags_2.11-2.3.2.0101-2.1.0.dli-20201111.074141-124-tests.jar
guava-29.0-jre.jar	kerb-common-1.0.1.jar	spark-unsafe_2.11-2.3.2.0101-2.1.0.dli-20201111.074144-123.jar
guice-4.0.jar	kerb-core-1.0.1.jar	spark-uquery_2.11-2.3.2.0101-2.1.0.dli-20201111.074906-210.jar
guice-servlet-4.0.jar	kerb-crypto-1.0.1.jar	spark-yarn_2.11-2.3.2.0101-2.1.0.dli-20201111.074151-123.jar
hadoop-annotations-3.1.1-ei-302002.jar	kerb-identity-1.0.1.jar	spire_2.11-0.13.0.jar

Dependency		
hadoop-auth-3.1.1-ei-302002.jar	kerb-server-1.0.1.jar	spire-macros_2.11-0.13.0.jar
hadoop-3.1.1-ei-302002.jar	kerb-simplekdc-1.0.1.jar	ST4-4.3.1.jar
hadoop-client-3.1.1-ei-302002.jar	kerb-util-1.0.1.jar	stax2-api-3.1.4.jar
hadoop-common-3.1.1-ei-302002.jar	kerby-asn1-1.0.1.jar	stax-api-1.0-2.jar
hadoop-hdfs-3.1.1-ei-302002.jar	kerby-config-1.0.1.jar	stream-2.7.0.jar
hadoop-hdfs-client-3.1.1-ei-302002.jar	kerby-pkix-1.0.1.jar	stringtemplate-3.2.1.jar
hadoop-3.1.1-41.jar	kerby-util-1.0.1.jar	token-provider-1.0.1.jar
hadoop-mapreduce-client-common-3.1.1-ei-302002.jar	kerby-xdr-1.0.1.jar	univocity-parsers-2.5.9.jar
hadoop-mapreduce-client-core-3.1.1-ei-302002.jar	kryo-shaded-3.0.3.jar	validation-api-1.1.0.Final.jar
hadoop-mapreduce-client-jobclient-3.1.1-ei-302002.jar	kubernetes-client-4.9.2-20200804.jar	woodstox-core-5.0.3.jar
hadoop-minikdc-3.1.1-ei-302002.jar	kubernetes-model-4.9.2-20200804.jar	xbean-asm5-shaded-4.4.jar
hadoop-yarn-api-3.1.1-ei-302002.jar	kubernetes-model-common-4.9.2-20200804.jar	xercesImpl-2.12.0.jar
hadoop-yarn-client-3.1.1-ei-302002.jar	leveldbjni-all-1.8-20191105.jar	xml-apis-1.4.01.jar
hadoop-yarn-common-3.1.1-ei-302002.jar	libfb303-0.9.3.jar	xz-1.0.jar
hadoop-yarn-registry-3.1.1-ei-302002.jar	libthrift-0.12.0.jar	zjsonpatch-0.3.0.jar
hadoop-yarn-server-common-3.1.1-ei-302002.jar	listenablefuture-9999.0-empty-to-avoid-conflict-with-guava.jar	zookeeper-3.5.6-ei-302002.jar

Dependency		
hadoop-yarn-server-web-proxy-3.1.1-ei-302002.jar	log4j-1.2.17-cloudera1.jar	zookeeper-jute-3.5.6-ei-302002.jar
hamcrest-core-1.3.jar	log4j-rolling-appender-20131024-2017.jar	zstd-jni-1.4.4-11.jar

Flink 1.15 Dependencies

Obtain information about the Flink 1.15 dependencies from the logs of a Flink job.

1. Check the logs of a Flink job.
 - a. Log in to the DLI console. In the navigation pane on the left, choose **Job Management > Flink Jobs**.
 - b. Click the name of the desired job. On the displayed page, click the **Run Log** tab.
 - c. Check the latest run logs. For more logs, check the OBS bucket where the job logs are stored.
2. Search for dependency information in the logs.
Search for **Classpath**: in the logs to check the dependencies.

Flink 1.12 Dependencies

Table 4-22 Flink 1.12 dependencies

Dependency		
bcpkix-jdk15on-1.60.jar	flink-json-1.12.2-ei-313001-dli-2022011002.jar	libtensorflow-1.12.0.jar
bcprov-jdk15on-1.60.jar	flink-kubernetes_2.11-1.12.2-ei-313001-dli-2022011002.jar	log4j-1.2-api-2.17.1.jar
clickhouse-jdbc-0.3.1-ei-313001-SNAPSHOT.jar	flink-metrics-prometheus_2.11-1.12.2-ei-313001-dli-2022011002.jar	log4j-api-2.17.1.jar
commons-codec-1.9.jar	flink-obs-hadoop-fs-2.0.0-20220226.034421-73.jar	log4j-core-2.17.1.jar
commons-configuration-1.7.jar	flink-s3-fs-hadoop-1.12.2.jar	log4j-slf4j-impl-2.17.1.jar

Dependency		
dataflow-fs-obs-2.0.0-20220226.034402-190.jar	flink-shaded-zookeeper-3.6.3-ei-313001-SNAPSHOT.jar	luxor-encrypt-2.0.0-20220405.072004-199.jar
deeplearning4j-core-0.9.1.jar	flink-sql-avro-1.12.2-ei-313001-dli-2022011002.jar	luxor-fs3-2.0.0-20220405.072025-195.jar
deeplearning4j-nlp-0.9.1.jar	flink-sql-avro-confluent-registry-1.12.2-ei-313001-dli-2022011002.jar	luxor-obs-fs3-2.0.0-20220405.072030-195.jar
deeplearning4j-nn-0.9.1.jar	flink-table_2.11-1.12.2-ei-313001-dli-2022011002.jar	manager-hadoop-security-crypter-8.1.3-313001-SNAPSHOT.jar
ejml-cdense-0.33.jar	flink-table-blink_2.11-1.12.2-ei-313001-dli-2022011002.jar	manager-wc2frm-8.1.3-313001-SNAPSHOT.jar
ejml-core-0.33.jar	guava-18.0.jar	mrs-obs-provider-3.1.1.49.jar
ejml-ddense-0.33.jar	guava-26.0-jre.jar	nd4j-api-0.9.1.jar
ejml-dsparse-0.33.jar	hadoop-hdfs-client-3.1.1-ei-302002.jar	nd4j-native-0.9.1.jar
ejml-experimental-0.33.jar	hadoop-3.1.1-46.jar	nd4j-native-api-0.9.1.jar
ejml-fdense-0.33.jar	hadoop-plugins-8.1.3-313001-SNAPSHOT.jar	nd4j-native-platform-0.9.1.jar
ejml-simple-0.33.jar	httpasyncclient-4.1.2.jar	okhttp-3.14.8.jar
ejml-zdense-0.33.jar	httpClient-4.5.3.jar	okio-1.14.0.jar
elsa-3.0.0-M7.jar	httpcore-4.4.4.jar	ranger-obs-client-0.1.1.jar
flink-changelog-json-1.12.2-ei-313001-dli-2022011002.jar	httpcore-nio-4.4.4.jar	secComponentApi-1.0.5.jar
flink-csv-1.12.2-ei-313001-dli-2022011002.jar	java-xmlbuilder-1.1.jar	slf4j-api-1.7.26.jar

Dependency		
flink-dist_2.11-1.12.2-ei-313001-dli-2022011002.jar	jna-4.1.0.jar	tensorflow-1.12.0.jar

Flink 1.10 Dependencies

For details about the sample code for a Flink 1.10 program, see [Writing Data to OBS Using Flink Jar](#).

Only queues created after December 2020 can use the Flink 1.10 dependencies.

Table 4-23 Flink 1.10 dependencies

Dependency		
bcpkix-jdk15on-1.60.jar	esdk-obs-java-3.20.6.1.jar	java-xmlbuilder-1.1.jar
bcprov-jdk15on-1.60.jar	flink-cep_2.11-1.10.0.jar	jna-4.1.0.jar
commons-codec-1.9.jar	flink-cep-scala_2.11-1.10.0.jar	libtensorflow-1.12.0.jar
commons-configuration-1.7.jar	flink-dist_2.11-1.10.0.jar	log4j-over-slf4j-1.7.26.jar
deeplearning4j-core-0.9.1.jar	flink-python_2.11-1.10.0.jar	logback-classic-1.2.3.jar
deeplearning4j-nlp-0.9.1.jar	flink-queryable-state-runtime_2.11-1.10.0.jar	logback-core-1.2.3.jar
deeplearning4j-nn-0.9.1.jar	flink-sql-client_2.11-1.10.0.jar	nd4j-api-0.9.1.jar
ejml-cdense-0.33.jar	flink-state-processor-api_2.11-1.10.0.jar	nd4j-native-0.9.1.jar
ejml-core-0.33.jar	flink-table_2.11-1.10.0.jar	nd4j-native-api-0.9.1.jar
ejml-ddense-0.33.jar	flink-table-blink_2.11-1.10.0.jar	nd4j-native-platform-0.9.1.jar
ejml-dsparse-0.33.jar	guava-26.0-jre.jar	okhttp-3.14.8.jar
ejml-experimental-0.33.jar	hadoop-3.1.1-41.jar	okio-1.14.0.jar
ejml-fdense-0.33.jar	httpasyncclient-4.1.2.jar	secComponentApi-1.0.5.jar
ejml-simple-0.33.jar	httpClient-4.5.3.jar	slf4j-api-1.7.26.jar
ejml-zdense-0.33.jar	httpcore-4.4.4.jar	tensorflow-1.12.0.jar

Dependency		
elsa-3.0.0-M7.jar	httpcore-nio-4.4.4.jar	-

Flink 1.7.2 Dependencies

For details about the sample code of a Flink 1.7.2 program, see **luxor-demo\dliflink-demo** in [DLI examples](#).

Table 4-24 Flink 1.7.2 dependencies

Dependency		
bcpkix-jdk15on-1.60.jar	esdk-obs-java-3.1.3.jar	httpcore-4.4.4.jar
bcprov-jdk15on-1.60.jar	flink-cep_2.11-1.7.0.jar	httpcore-nio-4.4.4.jar
commons-codec-1.9.jar	flink-cep-scala_2.11-1.7.0.jar	java-xmlbuilder-1.1.jar
commons-configuration-1.7.jar	flink-dist_2.11-1.7.0.jar	jna-4.1.0.jar
deeplearning4j-core-0.9.1.jar	flink-gelly_2.11-1.7.0.jar	libtensorflow-1.12.0.jar
deeplearning4j-nlp-0.9.1.jar	flink-gelly-scala_2.11-1.7.0.jar	log4j-over-slf4j-1.7.21.jar
deeplearning4j-nn-0.9.1.jar	flink-ml_2.11-1.7.0.jar	logback-classic-1.2.3.jar
ejml-cdense-0.33.jar	flink-python_2.11-1.7.0.jar	logback-core-1.2.3.jar
ejml-core-0.33.jar	flink-queryable-state-runtime_2.11-1.7.0.jar	nd4j-api-0.9.1.jar
ejml-ddense-0.33.jar	flink-shaded-curator-1.7.0.jar	nd4j-native-0.9.1.jar
ejml-dsparse-0.33.jar	flink-shaded-hadoop2-uber-1.7.0.jar	nd4j-native-api-0.9.1.jar
ejml-experimental-0.33.jar	flink-table_2.11-1.7.0.jar	nd4j-native-platform-0.9.1.jar
ejml-fdense-0.33.jar	guava-26.0-jre.jar	okhttp-3.14.8.jar
ejml-simple-0.33.jar	hadoop-3.1.1-41-20201014.085840-4.jar	okio-1.14.0.jar
ejml-zdense-0.33.jar	httpasyncclient-4.1.2.jar	slf4j-api-1.7.21.jar
elsa-3.0.0-M7.jar	httpClient-4.5.12.jar	tensorflow-1.12.0.jar

Dependency		
log4j-api-2.16.0.jar	log4j-core-2.16.0.jar	log4j-api-2.8.2.jar
log4j-core-2.8.2.jar	-	-

5 Enhanced Datasource Connections

5.1 Overview

What Is Enhanced Datasource Connection?

In cross-source data analysis scenarios, DLI needs to connect to external data sources. However, due to the different VPCs between the data source and DLI, the network cannot be connected, which results in DLI being unable to read data from the data source. DLI's enhanced datasource connection feature enables network connectivity between DLI and the data source.

This section will introduce a solution for cross-VPC data source network connectivity:

- Creating an enhanced datasource connection: Establish a VPC peering connection to connect DLI and the data source's VPC network.
- Testing network connectivity: Verify the connectivity between the queue and the data source's network.

For details about the data sources that support cross-source access, see [Cross-Source Analysis Development Methods](#).

CAUTION

In cross-source development scenarios, there is a risk of password leakage if datasource authentication information is directly configured. You are advised to use the datasource authentication provided by DLI. For details, see [Overview](#).

Notes and Constraints

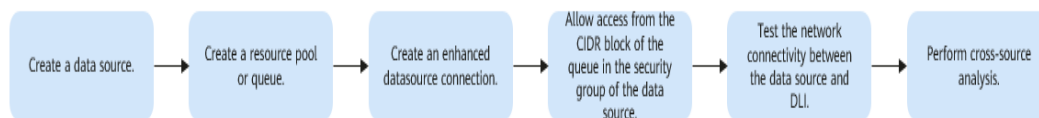
- Datasource connections cannot be created for the **default** queue.
- Flink jobs can directly access DIS, OBS, and SMN data sources without using datasource connections.
- Enhanced connections can only be created for yearly/monthly and pay-per-use queues.

- **VPC Administrator** permissions are required for enhanced connections to use VPCs, subnets, routes, VPC peering connections.
You can set these permissions by referring to [Service Authorization](#).
- If you use an enhanced datasource connection, the CIDR block of the elastic resource pool or queue cannot overlap with that of the data source.
- Only queues bound with datasource connections can access datasource tables.
- Datasource tables do not support the preview function.
- When checking the connectivity of datasource connections, the notes and constraints on IP addresses are:
 - The IP address must be valid, which consists of four decimal numbers separated by periods (.). The value ranges from 0 to 255.
 - During the test, you can add a port after the IP address and separate them with colons (:). The port can contain a maximum of five digits. The value ranges from 0 to 65535.
For example, **192.168.xx.xx** or **192.168.xx.xx:8181**.
- When checking the connectivity of datasource connections, the notes and constraints on domain names are:
 - The domain name can contain 1 to 255 characters. Only letters, numbers, underscores (_), and hyphens (-) are allowed.
 - The top-level domain name must contain at least two letters, for example, **.com**, **.net**, and **.cn**.
 - During the test, you can add a port after the domain name and separate them with colons (:). The port can contain a maximum of five digits. The value ranges from 0 to 65535.
For example, **example.com:8080**.

Cross-Source Analysis Process

To use DLI for cross-source analysis, you need to create a datasource connection to connect DLI to the data source, and then develop jobs to access the data source.

Figure 5-1 Cross-source analysis flowchart



Helpful Links

Creation Method	Enhanced Datasource Connection
Console	Creating an Enhanced Datasource Connection
API	Creating an Enhanced Datasource Connection

5.2 Cross-Source Analysis Development Methods

Cross-Source Analysis

If DLI needs to access external data sources, you need to establish enhanced datasource connections to enable the network between DLI and the data sources, and then develop different types of jobs to access the data sources. This is the process of DLI cross-source analysis.

This section describes how to develop data sources supported by DLI for cross-source analysis.

Notes

- Flink jobs can directly access DIS, OBS, and SMN data sources without using datasource connections.
- You are advised to use enhanced datasource connections to connect DLI to data sources.

Cross-Source Analysis Development Methods

[Table 5-1](#) lists the data sources supported by DLI and the corresponding development methods.

Table 5-1 Syntax reference for cross-source analysis

Service	Spark SQL Job	Spark Jar Job	Flink OpenSource SQL Job	Flink Jar Job
CloudTable HBase	<ul style="list-style-type: none"> • Create an HBase association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	<ul style="list-style-type: none"> • HBase source table • HBase result table • HBase dimension table 	-
CloudTable OpenTSDB	<ul style="list-style-type: none"> • Create an OpenTSDB association table • Insert data • Insert data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	-	-

Service	Spark SQL Job	Spark Jar Job	Flink OpenSource SQL Job	Flink Jar Job
CSS	<ul style="list-style-type: none"> • Create a CSS association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	<ul style="list-style-type: none"> • Elasticsearch result table 	-
DCS Redis	<ul style="list-style-type: none"> • Create a DCS association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code • Java sample code 	<ul style="list-style-type: none"> • Redis source table • Redis result table • Redis dimension table 	Flink job sample
DDS	<ul style="list-style-type: none"> • Create a DDS association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code • Java sample code 	-	-
DMS	-	-	<ul style="list-style-type: none"> • Kafka source table • Kafka result table 	-
GaussDB(DWS)	<ul style="list-style-type: none"> • Create a GaussDB(DWS) association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	<ul style="list-style-type: none"> • GaussDB(DWS) source table • GaussDB(DWS) result table • GaussDB(DWS) dimension table 	Flink job sample

Service	Spark SQL Job	Spark Jar Job	Flink OpenSource SQL Job	Flink Jar Job
MRS HBase	<ul style="list-style-type: none"> • Create an HBase association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	<ul style="list-style-type: none"> • HBase source table • HBase result table • HBase dimension table 	Flink job sample
MRS Kafka	-	-	<ul style="list-style-type: none"> • Kafka source table • Kafka result table 	Flink job sample
MRS OpenTSDB	<ul style="list-style-type: none"> • Create an OpenTSDB association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	-	-
RDS for MySQL	<ul style="list-style-type: none"> • Create an RDS association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	<ul style="list-style-type: none"> • MySQL CDC source table 	-
RDS PostGre	<ul style="list-style-type: none"> • Create an RDS association table • Insert data • Query data 	<ul style="list-style-type: none"> • Scala sample code • PySpark sample code 	<ul style="list-style-type: none"> • Postgres CDC source table 	-

5.3 Creating an Enhanced Datasource Connection

Scenario

Create an enhanced datasource connection for DLI to access, import, query, and analyze data of other data sources.

For example, to connect DLI to the MRS, RDS, CSS, Kafka, or GaussDB(DWS) data source, you need to enable the network between DLI and the VPC of the data source.

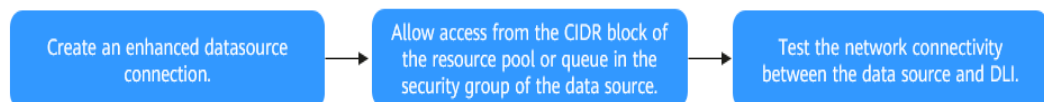
Create an enhanced datasource connection on the console.

Notes and Constraints

- Datasource connections cannot be created for the **default** queue.
- Flink jobs can directly access DIS, OBS, and SMN data sources without using datasource connections.
- Enhanced connections can only be created for yearly/monthly and pay-per-use queues.
- **VPC Administrator** permissions are required for enhanced connections to use VPCs, subnets, routes, VPC peering connections.
You can set these permissions by referring to [Service Authorization](#).
- If you use an enhanced datasource connection, the CIDR block of the elastic resource pool or queue cannot overlap with that of the data source.
- Only queues bound with datasource connections can access datasource tables.
- Datasource tables do not support the preview function.
- When checking the connectivity of datasource connections, the notes and constraints on IP addresses are:
 - The IP address must be valid, which consists of four decimal numbers separated by periods (.). The value ranges from 0 to 255.
 - During the test, you can add a port after the IP address and separate them with colons (:). The port can contain a maximum of five digits. The value ranges from 0 to 65535.
For example, **192.168.xx.xx** or **192.168.xx.xx:8181**.
- When checking the connectivity of datasource connections, the notes and constraints on domain names are:
 - The domain name can contain 1 to 255 characters. Only letters, numbers, underscores (_), and hyphens (-) are allowed.
 - The top-level domain name must contain at least two letters, for example, **.com**, **.net**, and **.cn**.
 - During the test, you can add a port after the domain name and separate them with colons (:). The port can contain a maximum of five digits. The value ranges from 0 to 65535.
For example, **example.com:8080**.

Process

Figure 5-2 Enhanced datasource connection creation flowchart



Prerequisites

- An elastic resource pool or queue has been created.
- You have obtained the VPC, subnet, private IP address, port, and security group information of the external data source.
- The security group of the external data source has allowed access from the CIDR block of the elastic resource pool or queue.

Procedure

Step 1 Create an Enhanced Datasource Connection

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Datasource Connections**.
3. On the displayed **Enhanced** tab, click **Create**.

Set parameters based on [Table 5-2](#).

Table 5-2 Parameters

Parameter	Description
Connection Name	Name of the created datasource connection. <ul style="list-style-type: none">- Only letters, numbers, and underscores (_) are allowed. The parameter must be specified.- A maximum of 64 characters are allowed.
Resource Pool	It binds an elastic resource pool or queue that uses a datasource connection. This parameter is optional. Only dedicated queues charged in yearly/monthly or pay-per-use billing mode can be bound to elastic resource pools. In regions where this function is available, an elastic resource pool with the same name is created by default for the queue created in "Creating a Queue." NOTE Before using an enhanced datasource connection, you must bind a queue and ensure that the VPC peering connection is in the Active state.
VPC	VPC used by the data source.
Subnet	Subnet used by the data source.
Route Table	Route table of the subnet. NOTE <ul style="list-style-type: none">- The route table is associated with the subnet used by the destination data source, which is not the table containing the route you add by Manage Route in the Operation column. The route you add on the Manage Route page is contained in the route table associated with the subnet used by the queue to be bound.- The subnet used by the destination data source must be different from that used by the queue to be bound. Otherwise, a segment conflict occurs.

Parameter	Description
Host Information	<p>In this text field, you can configure the mapping between host IP addresses and domain names so that jobs can only use the configured domain names to access corresponding hosts. This parameter is optional.</p> <p>For example, when accessing the HBase cluster of MRS, you need to configure the host name (domain name) and IP address of the ZooKeeper instance. Enter one record in each line in the format of <i>IP address Host name/Domain name</i>.</p> <p>Example:</p> <pre>192.168.0.22 node-masterxxx1.com 192.168.0.23 node-masterxxx2.com</pre> <p>For details about how to obtain host information, see How Do I Obtain MRS Host Information?.</p>
Tags	<p>Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS).</p> <p>For details, see Tag Management Service User Guide.</p> <p>NOTE</p> <ul style="list-style-type: none">- A maximum of 20 tags can be added.- Only one tag value can be added to a tag key.- The key name in each resource must be unique. <p>- Tag key: Enter a tag key name in the text box.</p> <p>NOTE</p> <p>A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (<code>_</code>, <code>:=+@</code>) are allowed, but the value cannot start or end with a space or start with <code>_sys_</code>.</p> <p>- Tag value: Enter a tag value in the text box.</p> <p>NOTE</p> <p>A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (<code>_</code>, <code>:=+@</code>) are allowed. The value cannot start or end with a space.</p>

4. Click **OK**.

After the creation is complete, the enhanced datasource connection is in the **Active** state, indicating that the connection is successfully created.

Step 2 Security Group Where the Data Source Belongs Allows Access from the CIDR Block of the Elastic Resource Pool

1. On the DLI management console, obtain the network segment of the elastic resource pool or queue.

- Choose **Resources** > **Queue Management** from the left navigation pane. On the page displayed, locate the queue on which jobs are running, and click the button next to the queue name to obtain the CIDR block of the queue.
2. Log in to the VPC console and find the VPC the data source belongs to.
 3. On the network console, choose **Virtual Private Cloud** > **Network Interfaces**. On the **Network Interfaces** tab page displayed, search for the security group name, click **More** in the **Operation** column, and select **Change Security Group**.
 4. In the navigation pane on the left, choose **Access Control** > **Security Groups**.
 5. Click the name of the security group to which the external data source belongs.
 6. Click the **Inbound Rules** tab and add a rule to allow access from the CIDR block of the queue. See [Figure 5-3](#).

Configure the inbound rule parameters according to [Table 5-3](#).

Figure 5-3 Adding an inbound rule

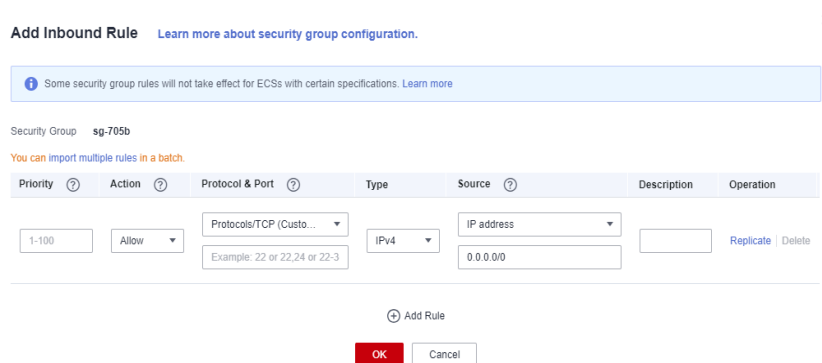


Table 5-3 Inbound rule parameters

Parameter	Description	Example Value
Priority	Priority of a security group rule. The priority value ranges from 1 to 100. The default value is 1, indicating the highest priority. A smaller value indicates a higher priority of a security group rule.	1
Action	Action of the security group rule.	Allow

Parameter	Description	Example Value
Protocol & Port	<ul style="list-style-type: none">– Network protocol. The value can be All, TCP, UDP, ICMP, or GRE.– Port: Port or port range over which the traffic can reach your instance. The port ranges from 1 to 65535.	In this example, select TCP . Leave the port blank or set it to the data source port.
Type	Type of IP addresses.	IPv4
Source	Allows access from IP addresses or instances in another security group.	In this example, enter the obtained queue CIDR block.
Description	Supplementary information about the security group rule. This parameter is optional.	–

Step 3 Test the Connectivity Between the DLI Queue and the Data Source

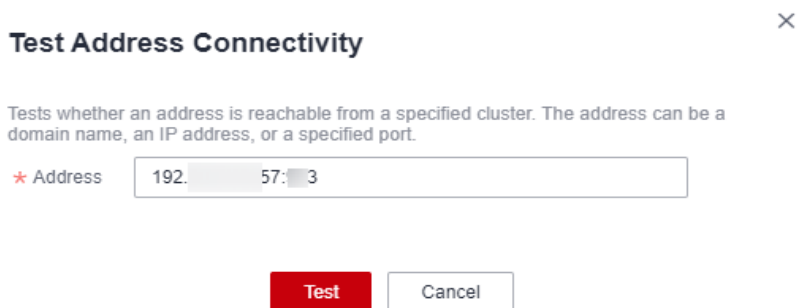
1. Obtain the private IP address and port number of the data source.
Take the RDS data source as an example. On the **Instances** page, click the target DB instance. On the page displayed, locate the **Connection Information** pane and view the private IP address. In the **Connection Information** pane, locate the **Database Port** to view the port number of the RDS DB instance.
2. In the navigation pane of the DLI management console, choose **Resources > Queue Management**.
3. Locate the queue bound with the enhanced datasource connection, click **More** in the **Operation** column, and select **Test Address Connectivity**.
4. Enter the data source connection address and port number to test the network connectivity.

Format: *IP address:Port number*

CAUTION

Before testing the connection, ensure that the security group of the external data source has allowed access from the CIDR block of the queue.

Figure 5-4 Testing the network connectivity between the queue and the data source



----End

Related Operations

- [Why Is a Datasource Connection Successfully Created But the Network Connectivity Test Fails?](#)
- [Configuring the Connection Between a DLI Queue and a Data Source in a Private Network](#)
- [Configuring the Connection Between a DLI Queue and a Data Source in the Internet](#)

5.4 Deleting an Enhanced Datasource Connection

Scenario

Delete an enhanced datasource connection that is no longer used on the console.

Procedure

1. Log in to the DLI management console.
2. In the left navigation pane, choose **Datasource Connections**.
3. On the **Enhanced** tab page displayed, locate the enhanced datasource connection to be deleted and click **Delete** in the **Operation** column.
4. In the dialog box displayed, click **Yes**.

5.5 Modifying Host Information in an Elastic Resource Pool

Scenario

Host information is the mapping between host IP addresses and domain names. After you configure host information, jobs can only use the configured domain names to access corresponding hosts. After a datasource connection is created, you can modify the host information.

When accessing the HBase cluster of MRS, you need to configure the host name (domain name) and IP address of the instance.

Constraints

You have obtained the MRS host information by referring to [How Do I Obtain MRS Host Information?](#)

Modifying Host Information

1. Log in to the DLI management console.
2. In the left navigation pane, choose **Datasource Connections**.
3. On the **Enhanced** tab page displayed, locate the enhanced datasource connection to be modified, click **More** in the **Operation** column, and select **Modify Host**.
4. In the **Modify Host** dialog box displayed, enter the obtained host information. Enter host information in the format of *Host IP address Host name*. Information about multiple hosts is separated by line breaks.

Example:

```
192.168.0.22 node-masterxxx1.com
```

```
192.168.0.23 node-masterxxx2.com
```

Obtain the MRS host information by referring to [How Do I Obtain MRS Host Information?](#)

5. Click **OK**.

How Do I Obtain MRS Host Information?

- **Method 1: View MRS host information on the management console.**

To obtain the host name and IP address of an MRS cluster, for example, MRS 3.x, perform the following operations:

- a. Log in to the MRS management console.
- b. On the **Active Clusters** page displayed, click your desired cluster to access its details page.
- c. Click the **Components** tab.
- d. Click **ZooKeeper**.
- e. Click the **Instance** tab to view the corresponding service IP addresses. You can select any service IP address.
- f. Modify host information by referring to [Modifying Host Information](#).

NOTE

If the MRS cluster has multiple IP addresses, enter any service IP address when creating a datasource connection.

- **Method 2: Obtain MRS host information from the /etc/hosts file on an MRS node.**
 - a. Log in to any MRS node as user **root**.
 - b. Run the following command to obtain MRS hosts information. Copy and save the information.

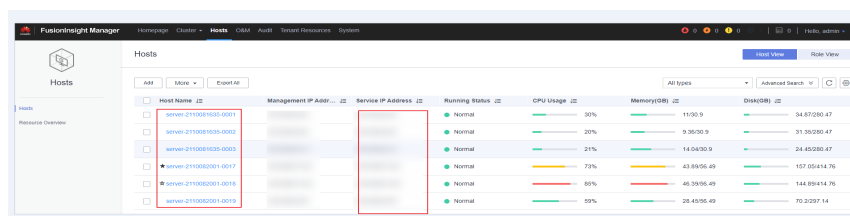
cat /etc/hosts

Figure 5-5 Obtaining hosts information

```
[root@node-master1nqt1 ~]# cat /etc/hosts
::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
10.10.10.10 hadoop.hadoop.com
10.10.10.10 manager
192.168.0.44 node-master1nqt1.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-master1nqt1.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.198 node-master2Y8cb.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-master2Y8cb.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.188 node-ana-corezyNL.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-ana-corezyNL.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.136 node-ana-corezyKB.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-ana-corezyKB.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.174 node-ana-coreVPY0.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-ana-coreVPY0.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
[root@node-master1nqt1 ~]#
[root@node-master1nqt1 ~]#
[root@node-master1nqt1 ~]#
```

- c. Modify host information by referring to [Modifying Host Information](#).
- **Method 3: Log in to FusionInsight Manager to obtain host information.**
 - a. Log in to FusionInsight Manager.
 - b. On FusionInsight Manager, click **Hosts**. On the **Hosts** page, obtain the host names and service IP addresses of the MRS hosts.

Figure 5-6 FusionInsight Manager



- c. Modify host information by referring to [Modifying Host Information](#).

5.6 Binding and Unbinding a Queue

Constraints

- The CIDR block of the DLI queue that is bound with a datasource connection cannot overlap with that of the data source.
- The default queue cannot be bound with a connection.

Binding a Queue


Before using an enhanced datasource connection, you must bind a queue and ensure that the VPC peering connection is in the **Active** state.

- **Method 1**
In the **Enhanced** tab, select a connection and click **Bind Queue** in the **Operation** column. In the dialog box that is displayed, select the queues to be bound and click **OK**.
- **Method 2**
Click the name of the selected connection. The **Details** page is displayed. Click **Create** in the upper left corner. In the dialog box that is displayed, select the queues to be bound and click **OK**.

Viewing Details about a Bound Queue

On the **Enhanced** tab page, select a connection and click the connection name to view the information about the bound queue.

Table 5-4 Parameters in the details list of datasource connection queues

Parameter	Description
VPC Peering ID	ID of the VPC peering connection created in the cluster to which the queue belongs. NOTE A VPC peering connection is created for each queue bound to an enhanced datasource connection. The VPC peering connection is used for cross-VPC communication. Ensure that the security group used by the data source allows access from the DLI queue CIDR block, and do not delete the VPC peering connection during the datasource connection.
Name	Name of a bound queue.
Connection Status	Datasource connection status. The following three statuses are available: <ul style="list-style-type: none">• Creating• Active• Failed NOTE If the connection status is Failed , click  on the left to view the detailed error information.
Updated	Time when a connection is updated. The connections in the connection list can be displayed according to the update time in ascending or descending order.
Operation	Unbind Queue: This operation is used to unbind a datasource connection from a queue.

Unbinding a Queue

If you do not need to use an enhanced datasource connection, you can unbind the queue from it to release resources.

- Method 1

On the **Enhanced** tab page, select a connection and click **More > Unbind Queue** in the **Operation** column. In the dialog box that is displayed, select the queues to be unbound and click **OK**.

- Method 2

Click the name of the target connection. The **Details** page is displayed. Select the queue to be unbound and click **Unbind Queue** in the **Operation** column. In the displayed dialog box, confirm the queue to be unbound and click **Yes**.

5.7 Adding a Route

Scenario

A route is configured with the destination, next hop type, and next hop to determine where the network traffic is directed. Routes are classified into system routes and custom routes.

After an enhanced connection is created, the subnet is automatically associated with the default route. You can add custom routes as needed to forward traffic destined for the destination to the specified next hop.

NOTE

- When an enhanced connection is created, the associated route table is the one associated with the subnet of the data source.
- The route to be added in the **Add Route** dialog box must be one in the route table associated with the subnet of the resource pool.
- The subnet of the data source must be different from that used by the resource pool. Otherwise, a network segment conflict occurs.

Procedure

1. Log in to the DLI management console.
2. In the left navigation pane, choose **Datasource Connections**.
3. On the **Enhanced** tab page displayed, locate the row containing the enhanced connection to which a route needs to be added, and add the route.
 - Method 1:
 - i. On the **Enhanced** tab page displayed, locate the enhanced datasource connection to which a route needs to be added and click **Manage Route** in the **Operation** column.
 - ii. Click **Add Route**.
 - iii. In the **Add Route** dialog box, enter the route information. For details about the parameters, see [Table 5-5](#).
 - iv. Click **OK**.
 - Method 2:
 - i. On the **Enhanced** tab page displayed, locate the enhanced datasource connection to which a route needs to be added, click **More** in the **Operation** column, and select **Add Route**.
 - ii. In the **Add Route** dialog box, enter the route information. For details about the parameters, see [Table 5-5](#).
 - iii. Click **OK**.

Table 5-5 Parameters for adding a custom route

Parameter	Description
Route Name	Name of a custom route, which is unique in the same enhanced datasource scenario. The name can contain 1 to 64 characters. Only digits, letters, underscores (_), and hyphens (-) are allowed.
IP Address	Custom route CIDR block. The CIDR block of different routes can overlap but cannot be the same. Do not add the CIDR blocks 100.125.xx.xx and 100.64.xx.xx to prevent conflicts with the internal CIDR blocks of services such as SWR. This can lead to failure of the enhanced datasource connection.

4. After adding a route, you can view the route information on the route details page.

5.8 Deleting a Route

Scenario

Delete a route that is no longer used.

Constraints

A custom route table cannot be deleted if it is associated with a subnet.

Procedure

1. Log in to the DLI management console.
2. In the left navigation pane, choose **Datasource Connections**.
3. On the **Enhanced** tab page displayed, locate the row containing the enhanced connection from which the route needs to be deleted, and delete the route.
 - Method 1:
 - i. On the **Enhanced** tab page displayed, locate the enhanced connection from which the route needs to be deleted and click **Manage Route** in the **Operation** column.
 - ii. Locate the route to be deleted and click **Delete** in the **Operation** column.
 - iii. In the dialog box displayed, click **OK**.
 - Method 2:
 - i. On the **Enhanced** tab page displayed, locate the enhanced connection from which the route needs to be deleted, click **More** in the **Operation** column, and select **Delete Route**.
 - ii. In the **Delete Route** dialog box displayed, confirm the route information.

- iii. Click **Yes**.

5.9 Enhanced Connection Permission Management

Scenario

Enhanced connections support user authorization by project. After authorization, users in the project have the permission to perform operations on the enhanced connection, including viewing the enhanced connection, binding a created resource pool to the enhanced connection, and creating custom routes. In this way, the enhanced connection can be used across projects. Grant and revoke permissions to and from a user for an enhanced connection.

NOTE

- If the authorized projects belong to different users in the same region, you can use the user account of the authorized projects to log in.
- If the authorized projects belong to the same user in the same region, you can use the current account to switch to the corresponding project.

Use Cases

Project B needs to access the data source of project A. The operations are as follows:

- For Project A:
 - a. Log in to DLI using the account of project A.
 - b. Create an enhanced datasource connection **ds** in DLI based on the VPC information of the corresponding data source.
 - c. Grant project B the permission to access the enhanced datasource connection **ds**.
- For Project B:
 - a. Log in to DLI using the account of project B.
 - b. Bind the enhanced datasource connection **ds** to a queue.
 - c. (Optional) Set host information and create a route.

After creating a VPC peering connection and route between the enhanced datasource connection of project A and the queue of project B, you can create a job in the queue of project B to access the data source of project A.

Procedure

1. Log in to the DLI management console.
2. In the left navigation pane, choose **Datasource Connections**.
3. On the **Enhanced** tab page displayed, locate the desired enhanced connection, click **More** in the **Operation** column, and select **Manage Permission**.
 - **Granting permission**
 - i. In the **Permissions** dialog box displayed, select **Grant Permission for Set Permission**.

- ii. Enter the project ID.
 - iii. Click **OK** to grant the resource pool operation permission to the project.
- **Revoking permission**
- i. In the **Permissions** dialog box displayed, select **Revoke Permission** for **Set Permission**.
 - ii. Select a project ID.
 - iii. Click **OK** to revoke the resource pool operation permission from the specified project.

5.10 Enhanced Datasource Connection Tag Management

Scenario

A tag is a key-value pair customized by users and used to identify cloud resources. It helps users to classify and search for cloud resources. A tag consists of a tag key and a tag value.

If you use tags in other cloud services, you are advised to create the same tag key-value pairs for cloud resources used by the same business to keep consistency.

DLI supports the following two types of tags:

- Resource tags: non-global tags created on DLI.
- Predefined tags: global tags created on Tag Management Service (TMS).
For more information about predefined tags, see [Tag Management Service User Guide](#).

DLI allows you to add, modify, or delete tags for datasource connections.

Procedure

1. In the left navigation pane of the DLI management console, choose **Datasource Connections**.
2. In the **Operation** column of the link, choose **More > Tags**.
3. The tag management page is displayed, showing the tag information about the current connection.
4. Click **Add/Edit Tag**. The **Add/Edit Tag** dialog is displayed. Add or edit tag keys and values and click **OK**.

Figure 5-7 Adding/Editing a tag

×

Add/Edit Tag

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#) ↻

To add a tag, enter a tag key and a tag value below.

Enter a tag key

Enter a tag value

Add

10 tags available for addition.

OK

Cancel

Table 5-6 Tag parameters

Parameter	Description
Tag key	<p>You can perform the following operations:</p> <ul style="list-style-type: none"> Click the text box and select a predefined tag key from the drop-down list. <p>To add a predefined tag, you need to create one on TMS and then select it from the Tag key drop-down list. You can click View predefined tags to go to the Predefined Tags page of the TMS console. Then, click Create Tag in the upper corner of the page to create a predefined tag.</p> <p>For details, see Creating Predefined Tags in <i>Tag Management Service User Guide</i>.</p> <ul style="list-style-type: none"> Enter a tag key in the text box. <p>NOTE A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=@) are allowed, but the value cannot start or end with a space or start with _sys_.</p>
Tag value	<p>You can perform the following operations:</p> <ul style="list-style-type: none"> Click the text box and select a predefined tag value from the drop-down list. Enter a tag value in the text box. <p>NOTE A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=@) are allowed. The value cannot start or end with a space.</p>

5. Click **OK**.
6. (Optional) To delete a tag, locate the row where the tag locates in the tag list and click **Delete** in the **Operation** column to delete the tag.

6 Datasource Authentication

6.1 Overview

What Is Datasource Authentication?

When analyzing across multiple sources, it is not recommended to configure authentication information directly in a job as it can lead to password leakage. Instead, you are advised to use either Data Encryption Workshop (DEW) or datasource authentication provided by DLI to securely store data source authentication information.

- DEW is a comprehensive cloud-based encryption service that addresses data security, key security, and complex key management issues. You are advised to use DEW to store authentication information for data sources. For details, see [Data Encryption Workshop \(DEW\)](#).
- Datasource authentication is used to manage authentication information for accessing specified data sources. After datasource authentication is configured, you do not need to repeatedly configure data source authentication information in jobs, improving data source authentication security while enabling DLI to securely access data sources.

This section describes how to use datasource authentication provided by DLI.

Notes and Constraints

- Only Spark SQL and Flink OpenSource SQL 1.12 jobs support datasource authentication.
- Flink jobs can use datasource authentication only on queues created after May 1, 2023.
- DLI supports four types of datasource authentication. Select an authentication type specific to each data source.
 - CSS: applies to 6.5.4 or later CSS clusters with the security mode enabled.
 - Kerberos: applies to MRS security clusters with Kerberos authentication enabled.
 - Kafka_SSL: applies to Kafka with SSL enabled.

- Password: applies to GaussDB(DWS), RDS, DDS, and DCS.

Datasource Authentication Types

DLI supports four types of datasource authentication. Select an authentication type specific to each data source.

- **CSS:** applies to 6.5.4 or later CSS clusters with the security mode enabled. During the configuration, you need to specify the username, password, and authentication certificate of the cluster and store the information in DLI through datasource authentication so that DLI can securely access CSS data sources. For details, see [Creating a CSS Datasource Authentication](#).
- **Kerberos:** applies to MRS security clusters with Kerberos authentication enabled. During the configuration, you need to specify MRS cluster authentication credentials, including the `krb5.conf` and `user.keytab` files. For details, see [Creating a Kerberos Datasource Authentication](#).
- **Kafka_SSL:** applies to Kafka with SSL enabled. During the configuration, you need to specify the KafkaTruststore path and password. For details, see [Creating a Kafka_SSL Datasource Authentication](#).
- **Password:** applies to GaussDB(DWS), RDS, DDS, and DCS data sources. During the configuration, you need to store the passwords of the data sources in DLI. For details, see [Creating a Password Datasource Authentication](#).

Jobs That Can Connect to Data Sources Through Datasource Authentication

Different types of jobs can connect to data sources through different types of datasource authentication.

- For details about the data sources that Spark SQL jobs can connect to through datasource authentication and their constraints, see [Table 6-1](#).
- For details about the data sources that Flink OpenSource SQL 1.12 jobs can connect to through datasource authentication and their constraints, see [Table 6-2](#).

Table 6-1 Data sources that Spark SQL jobs can connect to through datasource authentication

Datasource Authentication Type	Data Source	Notes and Constraints
CSS	CSS	The CSS cluster version must be 6.5.4 or later. The security mode has been enabled for the CSS cluster.
Password	GaussDB(DWS), RDS, DDS, and Redis	-

Table 6-2 Data sources that Flink OpenSource SQL 1.12 jobs can connect to through datasource authentication

Table Type	Datasource Authentication Type	Data Source	Notes and Constraints
Source table	Kerberos	HBase	Kerberos authentication has been enabled for the MRS cluster.
		Kafka	Kerberos authentication has been enabled for MRS Kafka.
	Kafka_SSL	Kafka	SASL_SSL authentication has been enabled for DMS Kafka. SASL authentication has been enabled for MRS Kafka. SSL authentication has been enabled for MRS Kafka.
	Password	GaussDB(DWS), RDS, and Redis	-
Result table	Kerberos	HBase	Kerberos authentication has been enabled for the MRS cluster.
		Kafka	Kerberos authentication has been enabled for MRS Kafka.
	Kafka_SSL	Kafka	SASL_SSL authentication has been enabled for DMS Kafka. SASL authentication has been enabled for MRS Kafka. SSL authentication has been enabled for MRS Kafka.
	Password	GaussDB(DWS), RDS, CSS, and Redis	-
Dimension table	Kerberos	HBase	Kerberos authentication has been enabled for the MRS cluster.
	Password	GaussDB(DWS), RDS, and Redis	-

6.2 Creating a CSS Datasource Authentication

Scenario

Create a CSS datasource authentication on the DLI console to store the authentication information of the CSS security cluster to DLI. This will allow you to

access to the CSS security cluster without having to configure a username and password in SQL jobs.

Create a datasource authentication for a CSS security cluster on the DLI console.

Notes

A CSS security cluster has been created and has met the following conditions:

- The cluster version is 6.5.4 or later.
- The security mode has been enabled for the cluster.

Procedure

1. Download the authentication credential of the CSS security cluster.
 - a. Log in to the CSS management console and choose **Clusters > Elasticsearch**.
 - b. On the **Clusters** page displayed, click the cluster name.
 - c. On the **Cluster Information** page displayed, find the security mode and download the certificate of the CSS security cluster.
2. Upload the authentication credential to the OBS bucket.
3. Create a datasource authentication.
 - a. Log in to the DLI management console.
 - b. Choose **Datasource Connections**. On the page displayed, click **Datasource Authentication**.
 - c. Click **Create**.

Configure CSS authentication parameters according to [Table 6-3](#).

Table 6-3 Parameters

Parameter	Description
Authentication Certificate	Name of the datasource authentication information to be created. <ul style="list-style-type: none">• The name can contain only digits, letters, and underscores (_), but cannot contain only digits or start with an underscore (_).• The length of the database name cannot exceed 128 characters.• It is recommended that the name contain the CSS security cluster name to distinguish security authentication information of different clusters.
Type	Select CSS .
Username	Username for logging in to the security cluster.
Password	The password of the security cluster

Parameter	Description
Certificate Path	Enter the OBS path to which the security certificate is uploaded, that is, the OBS bucket address in 2 .

Figure 6-1 Creating a datasource authentication-CSS

Create Authentication ×

Type

* Authentication Certificate

* Username ?

* Password ?

* Certificate Path ?

4. Create a table to access the CSS cluster.

When creating a table, associate the table with the created datasource authentication to access the CSS cluster.

For example, when using Spark SQL to create a table for accessing the CSS cluster, configure **es.certificate.name** to set the datasource authentication name and then connect to the CSS security cluster.

Use Spark SQL to create a table for accessing the CSS cluster by referring to [Creating a DLI Table and Associating It with CSS](#).

6.3 Creating a Kerberos Datasource Authentication

Scenario

Create a Kerberos datasource authentication on the DLI console to store the authentication information of the data source to DLI. This will allow you to access to the data source without having to configure a username and password in SQL jobs.

 NOTE

- When Kerberos authentication is enabled for MRS Kafka but SSL authentication is disabled, create a Kerberos authentication. When creating a table, configure **krb_auth_name** to associate the datasource authentication.
- If Kerberos authentication and SSL authentication are both enabled for MRS Kafka, you need to create Kerberos and Kafka_SSL authentications. When creating a table, configure **krb_auth_name** and **ssl_auth_name** to associate the datasource authentications.
- Datasource authentication is not required when Kerberos authentication is disabled but SASL authentication is enabled for MRS Kafka (for example, when a username and a password are used for PlainLoginModule authentication).
- When Kerberos authentication is disabled but SSL authentication is enabled for MRS Kafka, you need to create a Kafka_SSL authentication. When creating a table, configure **ssl_auth_name** to associate the datasource authentication.
- When Kerberos authentication is disabled but SASL authentication and SSL authentication are enabled for MRS Kafka, you need to create a Kafka_SSL authentication. When creating a table, configure **ssl_auth_name** to associate the datasource authentication.

Data Sources Supported by Kerberos Datasource Authentication

[Table 6-4](#) lists the data sources supported by Kerberos datasource authentication.

Table 6-4 Data sources supported by Kerberos datasource authentication

Job Type	Table Type	Data Source	Notes and Constraints
Flink OpenSource SQL	Source table	HBase	Kerberos authentication has been enabled for the MRS cluster.
		Kafka	Kerberos authentication has been enabled for MRS Kafka.
	Result table	HBase	Kerberos authentication has been enabled for the MRS cluster.
		Kafka	Kerberos authentication has been enabled for MRS Kafka.
	Dimension table	HBase	Kerberos authentication has been enabled for the MRS cluster.

Procedure

1. Download the authentication credential of the data source.
 - a. Log in to MRS Manager.
 - b. Choose **System > Permission > User**.
 - c. Click **More**, select **Download Authentication Credential**, save the file, and decompress it to obtain the **keytab** and **krb5.conf** files.

2. Upload the authentication credential to the OBS bucket.
3. Create a datasource authentication.
 - a. Log in to the DLI management console.
 - b. Choose **Datasource Connections**. On the page displayed, click **Datasource Authentication**.
 - c. Click **Create**.

Configure Kerberos authentication parameters according to [Table 6-5](#).

Table 6-5 Parameters

Parameter	Description
Type	Select Kerberos .
Authentication Certificate	Name of the datasource authentication to be created. <ul style="list-style-type: none">• The name can contain only digits, letters, and underscores (_), but cannot contain only digits or start with an underscore (_).• The name can contain a maximum of 128 characters.• It is recommended that the name contain the MRS security cluster name to distinguish security authentication information of different clusters.
Username	Username for logging in to the security cluster.
krb5_conf Path	OBS path to which the krb5.conf file is uploaded. NOTE The renew_lifetime configuration item under [libdefaults] must be removed from krb5.conf . Otherwise, the "Message stream modified (41)" error may occur.
keytab Path	OBS path to which the user.keytab file is uploaded.

Figure 6-2 Creating a datasource authentication – Kerberos

4. Create a table to access the MRS cluster.
When creating a data source, associate the data source with the created datasource authentication to access the data source.

Table 6-6 lists the fields used to associate with the datasource authentication during table creation.

Table 6-6 Fields that are used to associate with Kerberos datasource authentication during table creation

Job Type	Data Source	Parameter	Mandatory	Data Type	Description
Flink OpenSource SQL	HBa se	krb_au th_n ame	No	String	This field is used to associate datasource authentications when source, result, and dimension tables are created.

Job Type	Data Source	Parameter	Mandatory	Data Type	Description
	Kafka	krb_auth_name	No	String	<p>This field is used to associate datasource authentications when source and result tables are created.</p> <p>Name of the created Kerberos datasource authentication.</p> <p>If SASL_PLAINTEXT and Kerberos authentication are both used, you need to configure the following parameters:</p> <ul style="list-style-type: none">• 'properties.sasl.mechanism' = 'GSSAPI'• 'properties.security.protocol' = 'SASL_PLAINTEXT'

For details about how to create a table, see *Data Lake Insight Syntax Reference*.

- Flink OpenSource SQL Syntax Reference: [Creating an HBase Source Table](#)

6.4 Creating a Kafka_SSL Datasource Authentication

Scenario

Create a Kafka_SSL datasource authentication on the DLI console to store the Kafka authentication information to DLI. This will allow you to access to Kafka instances without having to configure a username and password in SQL jobs.

 NOTE

- When Kerberos authentication is enabled for MRS Kafka but SSL authentication is disabled, create a Kerberos authentication. When creating a table, configure **krb_auth_name** to associate the datasource authentication.
- If Kerberos authentication and SSL authentication are both enabled for MRS Kafka, you need to create Kerberos and Kafka_SSL authentications. When creating a table, configure **krb_auth_name** and **ssl_auth_name** to associate the datasource authentications.
- Datasource authentication is not required when Kerberos authentication is disabled but SASL authentication is enabled for MRS Kafka (for example, when a username and a password are used for PlainLoginModule authentication).
- When Kerberos authentication is disabled but SSL authentication is enabled for MRS Kafka, you need to create a Kafka_SSL authentication. When creating a table, configure **ssl_auth_name** to associate the datasource authentication.
- When Kerberos authentication is disabled but SASL authentication and SSL authentication are enabled for MRS Kafka, you need to create a Kafka_SSL authentication. When creating a table, configure **ssl_auth_name** to associate the datasource authentication.

Data Sources Supported by Kafka_SSL Datasource Authentication

[Table 6-7](#) lists the data sources supported by Kafka_SSL datasource authentication.

Table 6-7 Data sources supported by Kafka_SSL datasource authentication

Job Type	Table Type	Data Source	Notes and Constraints
Flink OpenSource SQL	Source table and result table	Kafka	SASL_SSL authentication has been enabled for DMS Kafka. SASL authentication has been enabled for MRS Kafka. SSL authentication has been enabled for MRS Kafka.

Procedure

1. Download the authentication credential.
 - **DMS Kafka**
 - i. Log in to the DMS (for Kafka) console and click a Kafka instance to access its details page.
 - ii. In the connection information, find the SSL certificate and click **Download**.
Decompress the downloaded **kafka-certs** package to obtain the **client.jks** and **phy_ca.crt** files.
 - **MRS Kafka**
 - i. Log in to MRS Manager.

- ii. Choose **System > Permission > User**.
 - iii. Click **More**, select **Download Authentication Credential**, save the file, and decompress it to obtain the truststore file.
 2. Upload the authentication credential to the OBS bucket.
 3. Create a datasource authentication.
 - a. Log in to the DLI management console.
 - b. Choose **Datasource Connections**. On the page displayed, click **Datasource Authentication**.
 - c. Click **Create**.
Configure Kafka authentication parameters according to [Table 6-8](#).

Table 6-8 Parameters

Parameter	Description
Type	Select Kafka_SSL .
Authentication Certificate	Name of the datasource authentication to be created. <ul style="list-style-type: none">• The name can contain only digits, letters, and underscores (<code>_</code>), but cannot contain only digits or start with an underscore (<code>_</code>).• The name can contain a maximum of 128 characters.
Truststore Path	OBS path to which the SSL truststore file is uploaded. <ul style="list-style-type: none">• For MRS Kafka, enter the OBS path of the Truststore.jks file.• For DMS Kafka, enter the OBS path of the client.jks file.
Truststore Password	Truststore password.
Keystore Path	OBS path to which the SSL keystore file (key and certificate) is uploaded.
Keystore Password	Keystore (key and certificate) password.
Key Password	Password of the private key in the keystore file.

4. Access Kafka with SASL_SSL authentication enabled.
When creating a data source, associate the data source with the created datasource authentication to access the data source.
[Table 6-9](#) lists the fields used to associate with the datasource authentication during table creation.

Table 6-9 Fields that are used to associate with Kafka_SSL datasource authentication during table creation

Parameter	Mandatory	Data Type	Description
ssl_auth_name	No	String	<p>This field is used to associate datasource authentications when source, result, and dimension tables are created.</p> <p>Name of the created Kafka_SSL datasource authentication. This configuration is used when SSL is configured for Kafka.</p> <ul style="list-style-type: none">• If only SSL is used, configure the following parameter: 'properties.security.protocol' = 'SSL';• If SASL_SSL is used, configure the following parameters:<ul style="list-style-type: none">- 'properties.security.protocol' = 'SASL_SSL',- 'properties.sasl.mechanism' ='GSSAPI or PLAIN'- 'properties.sasl.jaas.config' = 'org.apache.kafka.common.security.plain.PlainLoginModule required username=\"xxx\" password=\"xxx\";'

For details about how to create a table, see *Data Lake Insight Syntax Reference*.

- Flink OpenSource SQL Syntax Reference: [Creating a Kafka Source Table](#)

6.5 Creating a Password Datasource Authentication

Scenario

Create a password datasource authentication on the DLI console to store passwords of the GaussDB(DWS), RDS, DCS, and DDS data sources to DLI. This will allow you to access to the data sources without having to configure a username and password in SQL jobs.

Data Sources Supported by Password Datasource Authentication

[Table 6-10](#) lists the data sources supported by password datasource authentication.

Table 6-10 Data sources supported by password datasource authentication

Job Type	Table Type	Data Source
Spark SQL	-	GaussDB(DWS), RDS, DDS, and Redis
Flink OpenSource SQL	Source table	GaussDB(DWS), RDS, and Redis
	Result table	GaussDB(DWS), RDS, CSS, and Redis
	Dimension table	GaussDB(DWS), RDS, and Redis

Procedure

1. Create a datasource authentication.
 - a. Log in to the DLI management console.
 - b. Choose **Datasource Connections**. On the page displayed, click **Datasource Authentication**.
 - c. Click **Create**.
Configure authentication parameters according to [Table 6-11](#).

Table 6-11 Parameters

Parameter	Description
Type	Select Password .
Authentication Certificate	Name of the datasource authentication to be created. <ul style="list-style-type: none">• The name can contain only digits, letters, and underscores (_), but cannot contain only digits or start with an underscore (_).• The name can contain a maximum of 128 characters.
Username	Username for accessing the data source.
Password	Password for accessing the data source.

Figure 6-3 Creating a datasource authentication – Password

✕

Create Authentication

Type

✱ Authentication Certificate

Username

?

✱ Password

👁 ?

OK
Cancel

2. Access the data source.

When creating a data source, associate the data source with the created datasource authentication to access the data source.

Table 6-12 lists the fields used to associate with the datasource authentication during table creation.

Table 6-12 Fields that are used to associate with password datasource authentication during table creation

Job Type	Parameter	Mandatory	Data Type	Description
Spark SQL	passwd auth	No	String	Name of datasource authentication. It is applicable to GaussDB(DWS), RDS, DDS, and Redis data sources.
Flink OpenSource SQL	pwd_auth_name	No	String	This field is used to associate datasource authentications when source, result, and dimension tables are created. Set pwd_auth_name to the name of the password datasource authentication. If this parameter is set, you do not need to configure a username and a password of the data source in SQL jobs.

For details about how to create a table, see *Data Lake Insight Syntax Reference*.

- Flink OpenSource SQL Syntax Reference: [Creating a GaussDB\(DWS\) Source Table](#)

6.6 Datasource Authentication Permission Management

Scenario

Grant permissions on a datasource authentication to users so multiple user jobs can use the datasource authentication without affecting each other.

Notes

- The administrator and the datasource authentication owner have all permissions. You do not need to set permissions for them, and their datasource authentication permissions cannot be modified by other users.
- When setting datasource authentication permissions for a new user, ensure that the user group to which the user belongs has the **Tenant Guest** permission.

For details about the **Tenant Guest** permission and how to apply for the permission, see [System Permissions](#) and [Creating a User Group](#) in *Identity and Access Management User Guide*.

Granting Permissions on Datasource Connections

1. Log in to the DLI management console.
2. Choose **Datasource Connections**. On the page displayed, click **Datasource Authentication**.
3. Locate the row containing the datasource authentication to be authorized and click **Manage Permission** in the **Operation** column. The **User Permissions** page is displayed.
4. Click **Grant Permission** in the upper right corner of the page. On the **Grant Permission** dialog box displayed, grant permissions on this datasource authentication to other users.

Figure 6-4 Granting permissions on datasource connections

Grant Permission

* Username

Select the permissions to be granted to the user

<input type="checkbox"/> Select all	<input type="checkbox"/> Update	<input type="checkbox"/> Delete
<input type="checkbox"/> Access	<input type="checkbox"/> Revoke Permission	<input type="checkbox"/> View Other User's Permissions
<input type="checkbox"/> Grant Permission		

Table 6-13 Permission granting parameters

Parameter	Description
Username	Name of the IAM user to whom permissions on the datasource connection are to be granted. NOTE The username is the name of an existing IAM user.
Select the permissions to be granted to the user	<ul style="list-style-type: none">• Access: This permission allows you to access the datasource connection.• Update: This permission allows you to update the datasource connection.• Delete: This permission allows you to delete the datasource connection.• Grant Permission: This permission allows you to grant the datasource connection permission to other users.• Grant Permission: This permission allows you to revoke the datasource connection permission to other users. However, you cannot revoke the permissions of the datasource connection owner.• View Other User's Permissions: This permission allows you to view the datasource connection permissions of other users.

Modifying the Permissions of Current User

1. Log in to the DLI management console.
2. Choose **Datasource Connections**. On the page displayed, click **Datasource Authentication**.
3. Locate the row containing the datasource authentication to be authorized and click **Manage Permission** in the **Operation** column. The **User Permissions** page is displayed.
4. Click **Set Permission** in the **Operation** column to modify the permissions of the current user. [Table 6-13](#) lists the detailed permission descriptions.

NOTE

- If all options under **Set Permission** are gray, you are not allowed to change permissions on this datasource connection. You can apply to the administrator, group owner, or other users who have the permission to grant permissions for the permissions to grant and revoke the datasource authentication permissions.
- To revoke all permissions of the current user, click **Revoke Permission** in the **Operation** column. The IAM user will no longer have any permission on the datasource authentication.

7 Job Templates

7.1 Managing SQL Templates

To facilitate SQL operation execution, DLI allows you to customize query templates or save the SQL statements in use as templates. After templates are saved, you do not need to compile SQL statements. Instead, you can directly perform the SQL operations using the templates.

SQL templates include sample templates and custom templates. The default sample template contains 22 standard TPC-H query statements, which can meet most TPC-H test requirements. For details, see [TPC-H Sample Data in the SQL Template](#).

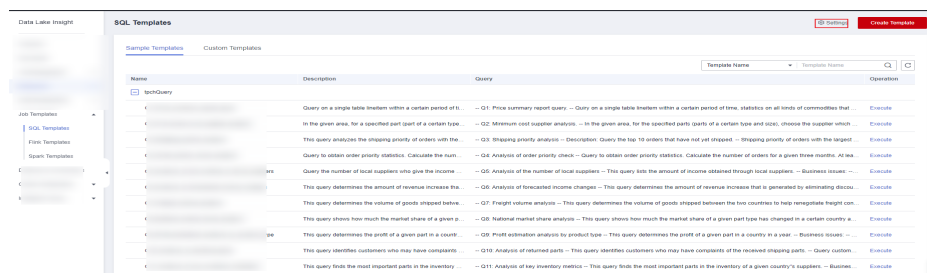
SQL template management provides the following functions:

- [Sample Templates](#)
- [Custom Templates](#)
- [Creating a Template](#)
- [Executing the Template](#)
- [Searching for a Template](#)
- [Modifying a Template](#)
- [Deleting a Template](#)

Table Settings

In the upper right corner of the **SQL Template** page, click **Set Property** to determine whether to display templates by group.

Figure 7-1 Template settings



If you select **Display by Group**, the following display modes are available:

- Expand the first group
- Expand all
- Collapse All

Sample Templates

The current sample template contains 22 standard TPC-H query statements. You can view the template name, description, and statements. For details about TPC-H examples, see [TPC-H Sample Data in the SQL Template](#).

Table 7-1 Template management parameters

Parameter	Description
Name	<p>Indicates the template name.</p> <ul style="list-style-type: none"> ● A template name can contain only digits, letters, and underscores (_), but cannot start with an underscore (_) or contain only digits. It cannot be left empty. ● The template name can contain a maximum of 50 characters.
Description	Description of the template you create.
Statement	SQL statement created as the template.
Operation	Execute: After you click this button, the system switches to the SQL Editor page, where you can modify or directly perform the statement as required. For details, see Executing the Template .

The existing sample templates apply to the following scenarios:

- Price summary report query
- Minimum cost supplier analysis

- Shipping priority analysis
- Analysis of order priority check
- Analysis of the number of local suppliers
- Analysis of forecasted income changes
- Freight volume analysis
- National market share analysis
- Profit estimation analysis by product type
- Analysis of returned parts
- Analysis of key inventory indicators
- Freight mode and command priority analysis
- Consumer allocation analysis
- Promotion effect analysis
- Analysis of the supplier with the largest contribution
- Analysis of the relationship between parts and suppliers
- Revenue analysis of small-lot orders
- Customer analysis for large orders
- Discounted revenue analysis
- Analysis of potential component improvements
- Analysis of suppliers who fail to deliver goods on time
- Global sales opportunity analysis

Custom Templates

The custom template list displays all templates you have created. You can view the template name, description, statements, and more.

Table 7-2 Template management parameters

Parameter	Description
Name	Indicates the template name. <ul style="list-style-type: none">• A template name can contain only digits, letters, and underscores (_), but cannot start with an underscore (_) or contain only digits. It cannot be left empty.• The template name can contain a maximum of 50 characters.
Description	Description of the template you create.
Statement	SQL statement created as the template.

Parameter	Description
Operation	<ul style="list-style-type: none"> • Execute: After you click this button, the system switches to the SQL Editor page, where you can modify or directly perform the statement as required. For details, see Executing the Template. • Modify: Click Modify. In the displayed Modify Template dialog box, modify the template information as required. For details, see Modifying a Template.

Creating a Template

You can create a template on either the **Job Templates** or the **SQL Editor** page.

- To create a template on the **Job Templates** page:
 - a. On the left of the management console, choose **Job Templates > SQL Templates**.
 - b. On the **SQL Templates** page, click **Create Template** to create a template.

Enter the template name, SQL statement, and description information. For details, see [Table 7-3](#).

Figure 7-2 Creating a template

Create Template

You can create 89 more templates. [Increase quota](#).

* Name

* Statement
0/10,000

Description
0/256

Group Use Existing Use new Do not use

* Group Name

OK Cancel

Table 7-3 Parameter description

Parameter	Description
Name	Indicates the template name. <ul style="list-style-type: none">• A template name can contain only digits, letters, and underscores (_), but cannot start with an underscore (_) or contain only digits. It cannot be left empty.• The template name can contain a maximum of 50 characters.
Statement	SQL statement to be saved as a template.
Description	Description of the template you create.
Group	<ul style="list-style-type: none">• Use existing• Use new• Do not use
Group Name	If you select Use existing or Use new , you need to enter the group name.

- c. Click **OK**.
- To create a template on the **SQL Editor** page:
 - a. On the left of the management console, click **SQL Editor**.
 - b. In the SQL job editing area of the displayed **SQL Editor** page, click **More** in the upper right corner and choose **Save as Template**.
Enter the template name, SQL statement, and description information.
For details, see [Table 7-3](#).
 - c. Click **OK**.

Executing the Template

Perform the template operation as follows:

1. On the left of the management console, choose **Job Templates > SQL Templates**.
2. On the **SQL Templates** page, select a template and click **Execute** in the **Operation** column. The **SQL Editor** page is displayed, and the corresponding SQL statement is automatically entered in the SQL job editing window.
3. In the upper right corner of the SQL job editing window, Click **Execute** to run the SQL statement. After the execution is complete, you can view the execution result below the current SQL job editing window.

Searching for a Template

On the **SQL Templates** page, you can enter the template name keyword in the search box on the upper right corner to search for the desired template.

Modifying a Template

Only custom templates can be modified. To modify a template, perform the following steps:

- Step 1** On the **SQL Templates** page, locate the target template and click **Modify** in the **Operation** column.
- Step 2** In the displayed **Modify Template** dialog box, modify the template name, statement, and description as required.
- Step 3** Click **OK**.

----End

Deleting a Template

On the **SQL Templates** page, select one or more templates to be deleted and click **Delete** to delete the selected templates.

7.2 Managing Flink Templates

Flink templates include sample templates and custom templates. You can modify an existing sample template to meet the actual job logic requirements and save time for editing SQL statements. You can also customize a job template based on your habits and methods so that you can directly invoke or modify the template in later jobs.

Flink template management provides the following functions:

- [Flink SQL Sample Template](#)
- [Flink OpenSource SQL Sample Template](#)
- [Custom Templates](#)
- [Creating a Template](#)
- [Creating a Job Based on a Template](#)
- [Modifying a Template](#)
- [Deleting a Template](#)

Flink SQL Sample Template

The template list displays existing sample templates for Flink SQL jobs. [Table 1](#) describes the parameters in the template list.

The scenarios of sample templates can be different, which are subject to the console.

Table 7-4 Parameters in the Flink SQL sample template list

Parameter	Description
Name	Template name. Enter 1 to 64 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed.
Description	Description of a template. It contains 0 to 512 characters.
Operation	Create Job: Create a job directly by using the template. After a job is created, the system switches to the Edit page under Job Management .

Flink OpenSource SQL Sample Template

The template list displays existing sample templates for Flink SQL OpenSource jobs. [Table 1](#) describes the parameters in the template list.

Table 7-5 Parameters in the sample template list for Flink OpenSource SQL jobs

Parameter	Description
Name	Template name. Enter 1 to 64 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed.
Description	Description of a template. It contains 0 to 512 characters.
Operation	Create Job: Create a job directly by using the template. After a job is created, the system switches to the Edit page under Job Management .

The existing sample templates apply to the following scenarios:

- Create a wide table of order information from the dimension table of address information.
- Generate statistics on indicators such as the daily transaction amount, number of orders, and number of paid users in real time.
- Generate statistics on offerings with the highest real-time click-through rate.

Custom Templates

The custom template list displays all Jar job templates. [Table 1](#) describes parameters in the custom template list.

Table 7-6 Parameters in the custom template list

Parameter	Description
Name	Template name. Enter 1 to 64 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed.
Description	Description of a template. It contains 0 to 512 characters.
Created	Time when a template is created.
Updated	Latest time when a template is modified.
Operation	<ul style="list-style-type: none">● Edit: Modify a template that has been created.● Create Job: Create a job directly by using the template. After a job is created, the system switches to the Edit page under Job Management.● More:<ul style="list-style-type: none">- Delete: Delete a created template.- Tags: View or add tags.

Creating a Template

You can create a template using any of the following methods:

- Creating a template on the **Template Management** page
 - a. In the left navigation pane of the DLI management console, choose **Job Templates > Flink Templates**.
 - b. Click **Create Template** in the upper right corner of the page. The **Create Template** dialog box is displayed.
 - c. Specify **Name** and **Description**.

Figure 7-3 Creating a Flink template

Create Template
×

Type Flink SQL ▼

* Name Enter a name.

Description Description

Tags

It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources. [View predefined tags](#) ↻

To add a tag, enter a tag key and a tag value below.

10 tags available for addition.

Table 7-7 Template parameters

Parameter	Description
Type	Template type <ul style="list-style-type: none"> Flink SQL job template Flink OpenSource SQL job template
Name	Template name. Enter 1 to 64 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed. <p>NOTE The template name must be unique.</p>
Description	Description of a template. It contains 0 to 512 characters.

Parameter	Description
Tags	<p>Tags used to identify cloud resources. A tag includes the tag key and tag value. If you want to use the same tag to identify multiple cloud resources, that is, to select the same tag from the drop-down list box for all services, you are advised to create predefined tags on the Tag Management Service (TMS).</p> <p>For details, see Tag Management Service User Guide.</p> <p>NOTE</p> <ul style="list-style-type: none"> • A maximum of 20 tags can be added. • Only one tag value can be added to a tag key. • The key name in each resource must be unique. • Tag key: Enter a tag key name in the text box. <p>NOTE</p> <p>A tag key can contain a maximum of 128 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed, but the value cannot start or end with a space or start with <code>_sys_</code>.</p> <ul style="list-style-type: none"> • Tag value: Enter a tag value in the text box. <p>NOTE</p> <p>A tag value can contain a maximum of 255 characters. Only letters, numbers, spaces, and special characters (_.:+=-@) are allowed. The value cannot start or end with a space.</p>

d. Click **OK** to enter the editing page.

The [Table 7-8](#) describes the parameters on the template editing page.

Table 7-8 Template parameters

Parameter	Description
Name	You can modify the template name.
Description	You can modify the template description.
Saving Mode	<ul style="list-style-type: none"> • Save Here: Save the modification to the current template. • Save as New: Save the modification as a new template.
SQL statement editing area	In the area, you can enter detailed SQL statements to implement business logic. For details about how to compile SQL statements, see Data Lake Insight SQL Syntax Reference .
Save	Save the modifications.
Create Job	Use the current template to create a job.
Format	Format SQL statements. After SQL statements are formatted, you need to compile SQL statements again.

Parameter	Description
Theme Settings	Change the font size, word wrap, and page style (black or white background).

- e. In the SQL statement editing area, enter SQL statements to implement service logic. For details about how to compile SQL statements, see [Data Lake Insight SQL Syntax Reference](#).
 - f. After the SQL statement is edited, click **Save** in the upper right corner to complete the template creation.
 - g. (Optional) If you do not need to modify the template, click **Create Job** in the upper right corner to create a job based on the current template. For details about how to create a job, see [Creating a Flink SQL Job](#), and [Creating a Flink Jar Job](#).
- Creating a template based on an existing job template
 - a. In the left navigation pane of the DLI management console, choose **Job Templates > Flink Templates**. Click the **Custom Templates** tab.
 - b. In the row where the desired template is located in the custom template list, click **Edit** under **Operation** to enter the **Edit** page.
 - c. After the modification is complete, set **Saving Mode** to **Save as New**.
 - d. Click **Save** in the upper right corner to save the template as a new one.
 - Creating a template using a created job
 - a. In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.
 - b. Click **Create Job** in the upper right corner. The **Create Job** page is displayed.
 - c. Specify parameters as required.
 - d. Click **OK** to enter the editing page.
 - e. After the SQL statement is compiled, click **Set as Template**.
 - f. In the **Set as Template** dialog box that is displayed, specify **Name** and **Description** and click **OK**.
 - Creating a template based on the existing job
 - a. In the left navigation pane of the DLI management console, choose **Job Management > Flink Jobs**. The **Flink Jobs** page is displayed.
 - b. In the job list, locate the row where the job that you want to set as a template resides, and click **Edit** in the **Operation** column.
 - c. After the SQL statement is compiled, click **Set as Template**.
 - d. In the **Set as Template** dialog box that is displayed, specify **Name** and **Description** and click **OK**.

Creating a Job Based on a Template

You can create jobs based on sample templates or custom templates.

1. In the left navigation pane of the DLI management console, choose **Job Templates > Flink Templates**.

2. In the sample template list, click **Create Job** in the **Operation** column of the target template. For details about how to create a job, see [Creating a Flink SQL Job](#) and [Creating a Flink Jar Job](#).

Modifying a Template

After creating a custom template, you can modify it as required. The sample template cannot be modified, but you can view the template details.

1. In the left navigation pane of the DLI management console, choose **Job Templates > Flink Templates**. Click the **Custom Templates** tab.
2. In the row where the template you want to modify is located in the custom template list, click **Edit** in the **Operation** column to enter the **Edit** page.
3. In the SQL statement editing area, modify the SQL statements as required.
4. Set **Saving Mode** to **Save Here**.
5. Click **Save** in the upper right corner to save the modification.

Deleting a Template

You can delete a custom template as required. The sample templates cannot be deleted. Deleted templates cannot be restored. Exercise caution when performing this operation.

1. In the left navigation pane of the DLI management console, choose **Job Templates > Flink Templates**. Click the **Custom Templates** tab.
2. In the custom template list, select the templates you want to delete and click **Delete** in the upper left of the custom template list.

Alternatively, you can delete a template by performing the following operations: In the custom template list, locate the row where the template you want to delete resides, and click **More > Delete** in the **Operation** column.

3. In the displayed dialog box, click **Yes**.

7.3 Managing Spark SQL Templates

Scenario

You can modify a sample template to meet the Spark job requirements, saving time for editing SQL statements.

Currently, the cloud platform does not provide preset Spark templates. You can customize Spark job templates. This section describes how to create a Spark job template.

Creating a Management Job Template

To create a Spark job template, you can save a Spark job information as a template.

1. In the left navigation pane of the DLI management console, choose **Job Templates > Spark Templates**. The **Spark Templates** page is displayed.

2. Configure job parameters by referring to [Creating a Spark Job](#).
3. After you finish editing the job, click **Save as Template**.
4. Enter the template name and other information as you need.
5. Set a template group for future management.
6. Click **OK**.

7.4 Appendix

7.4.1 TPC-H Sample Data in the SQL Template

TPC-H Sample Data

TPC-H is a test set developed by the Transaction Processing Performance Council (TPC) to simulate decision-making support applications. It is widely used in academia and industry to evaluate the performance of decision-making support technology. This business test has higher requirements on vendors, because it can comprehensively evaluate the overall business computing capability. With universal business significance, is widely used in analysis of bank credit, credit card, telecom operation, tax, as well as tobacco industry decision-making analysis.

The TPC-H benchmark test is developed from TPC-D (a standard specified by TPC in 1994 and used as the test benchmark for decision-making support systems). TPC-H implements a 3NF data warehouse that contains eight basic relationships, with a data volume range from 1 GB to 3 TB. The TPC-H benchmark test includes 22 queries (Q1 to Q22). The main evaluation indicator is the response time of each query (from submission to result return). The unit of the TPC-H benchmark test is the query number per hour (QphH@size). **H** indicates the average number of complex queries per hour. **size** indicates the size of database, which reflects the query processing capability of the system. TPC-H can evaluate key performance parameters that other tests cannot evaluate, because it is modeled based on the actual production and operation environment. In a word, the TPC-H standard by TPC meets the test requirements of data warehouse and motivate vendors and research institutes to stretch the limit of this technology.

In this example, DLI directly queries the TPC-H dataset on OBS. DLI has generated a standard TPC-H-2.18 dataset of 100 MB which is uploaded to the tpch folder on OBS. The read-only permission is granted to you to facilitate query operations.

TPC-H Test and Metrics

TPC-H test is divided into three sub-tests: data loading test, Power test, and Throughput test. Data loading indicates the process of setting up a test database, and the loading test is to test the data loading ability of DBMS. The first test is data loading test that tests data loading time, which is time-consuming. The second test is Power test, also called raw query. After data loading test is complete, the database is in the initial state without any other operation, especially the data in the buffer is not tested. Power test requires that the 22 queries be executed once in sequence and a pair of RF1 and RF2 operations be executed at the same time. The third test is Throughput test, the core and most complex test, more similar to the actual application environment. With multiple

query statement groups and a pair of RF1 and RF2 update flows, Throughput test pose greater pressure on the SUT system than Power test does.

The basic data in the test is related to the execution time (the time of each data loading step, each query execution, and each update execution), based on which you can calculate the data loading time, Power@Size, Throughput@Size, qphH@Size and \$/QphH@Size.

Power@Size is the result of the Power test, which is defined as the reciprocal of the geometric average value of the query time and change time. The formula is as follows:

$$\text{TPC-H Power@Size} = \frac{3600 * SF}{\sqrt[24]{\prod_{i=1}^{i=22} QI(i,0) * \prod_{j=1}^{j=2} RI(j,0)}}$$

Size indicates the data size. SF is the scaling factor of data scale. QI (i, 0) indicates the time of the ith query, in seconds. R (I j, 0) is the update time of RFj, in seconds.

Throughput@Size is the Throughput test result, which is defined as the reciprocal of the average value of all query execution time. The formula is as follows:

$$\text{QphH@Size} = \sqrt{\text{Power @ Size} * \text{Throughput @ Size}}$$

Service Scenario

You can use the built-in TPC-H test suite of DLI to perform interactive query without uploading data.

Advantages of DLI Built-in TPC-H

- You can log in to DLI and get permission to run SQL statements without creating tables or import data.
- The 22 preset TPC-H SQL query templates with rich functions meet the requirements of most business scenarios. You do not need to download TPC-H query statements, which saves your time and energy.
- Data Lake gives you brand-new experience of serverless DLI product within the minimum time.

Precautions

When a sub-account uses the TPC-H test suite, the main account needs to grant the sub-account the OBS access permission and the permission to view the main account table. If the master account has not logged in to DLI, the sub-account needs to have the permissions to create databases and tables in addition to the preceding permissions.

Developing and Submitting a Spark SQL Job Using the TPC-H Sample Template

DLI allows you to create custom templates or save currently used SQL statements as templates for quick and convenient SQL operations. Once a template is saved, you can execute SQL operations directly through the template without the need to write SQL statements.

1. Log in to the DLI management console.
2. On the DLI management console, choose **Job Templates > SQL Templates** in the navigation pane on the left. On the displayed **Sample Templates** tab, click the plus sign next to **tpchQuery** to find the **Q1_Price_summary_report_query** template. Then, click **Execute** in the **Operation** column.
3. In the upper part of the editing window, set **Engine** to **Spark**, **Queues** to **default**, and **Databases** to **default**, and click **Execute**.
4. Check the query result on the **View Result** tab below the editing window.

This example uses the **default** queue and database preset in the system as an example. You can also run the command in a self-created queue and database.

8 DLI Agency Permissions

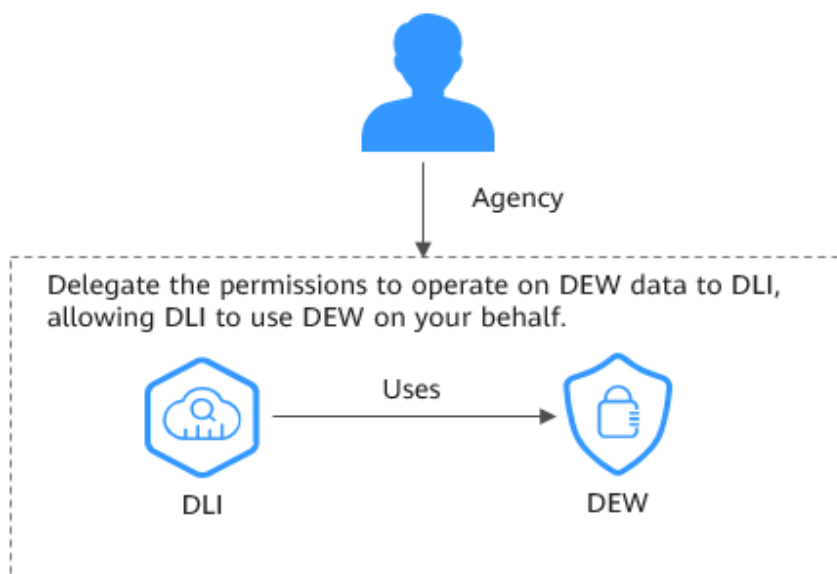
8.1 DLI Agency Overview

What Is an Agency?

Huawei Cloud services often interact with each other, with some of which dependent on other services. You can create an agency to delegate DLI to use other cloud services and perform resource O&M on your behalf.

For example, the AK/SK required by DLI Flink jobs is stored in DEW. To allow DLI to access DEW data during job execution, you need to provide an IAM agency to delegate the permissions to operate on DEW data to DLI.

Figure 8-1 DLI service agency



DLI Agencies

Before using DLI, you are advised to set up DLI agency permissions to ensure the proper functioning of DLI.

- By default, DLI provides the following agencies: **dli_admin_agency**, **dli_management_agency**, and **dli_data_clean_agency**. The names of these agencies are fixed, but the permissions contained in them can be customized. In other scenarios, you need to create custom agencies. For details about the agencies, see [Table 8-1](#).
- DLI upgrades **dli_admin_agency** to **dli_management_agency** to meet the demand for fine-grained agency permissions management. The new agency has the necessary permissions for datasource operations, notifications, and user authorization operations. For details, see [Updating the Permissions of a DLI Agency](#).
- To use Flink 1.15, Spark 3.3.1 (Spark general queue scenario), or a later version to execute jobs, perform the following operations:
Create an agency on the IAM console and add the agency information to the job configuration. For details, see [Customizing DLI Agency Permissions](#).
 - Common scenarios for creating an agency: DLI is allowed to read and write data from and to OBS, dump logs, and read and write Flink checkpoints. DLI is allowed to access DEW to obtain data access credentials and access catalogs to obtain metadata.
 - You cannot use the default agency names **dli_admin_agency**, **dli_management_agency**, or **dli_data_clean_agency**. It must be unique.
- If the engine version is earlier than Flink 1.15, **dli_admin_agency** is used by default during job execution. If the engine version is earlier than Spark 3.3.1, user authentication information (AK/SK and security token) is used during job execution.
This means that jobs whose engine versions are earlier than Flink 1.15 or Spark 3.3.1 are not affected by the update of agency permissions and do not require custom agencies.
- To maintain compatibility with existing job agency permission requirements, **dli_admin_agency** will still be listed in the IAM agency list even after the update.

 NOTE

- Only the tenant account or a member account of user group **admin** can authorize the service.
- Do not delete the agency created by the system by default.

Table 8-1 DLI agencies

Permission	Type	Description
dli_admin_agency	Default agency	This agency has been discarded and is not recommended. Upgrade the agency to dli_management_agency as soon as possible. For details about how to update an agency, see Updating the Permissions of a DLI Agency .

Permission	Type	Description
dli_management_agency	Default agency	DLI system agency, which is used to delegate operation permissions to DLI so that DLI can use other cloud services and perform resource O&M operations on your behalf. This agency grants permissions for datasource operations, message notifications, and user authorization operations. For details about the permissions of an agency, see Table 8-2 .
dli_data_clean_agency	Default agency, which needs to be authorized by users	Data cleanup agency, which is used to clean up data according to the lifecycle of a table and clean up lakehouse table data. You need to create a DLI agency named dli_data_clean_agency on IAM and grant permissions to it. You need to create an agency and customize permissions for it. However, the agency name is fixed to dli_data_clean_agency . For details about the permission policies of an agency, see Agency Permission Policies in Common Scenarios .
Other custom agencies	Custom agency	When using Flink 1.15, Spark 3.3, or a later version to execute jobs, create an agency on the IAM console and add new agency information to the job configuration. For details, see Customizing DLI Agency Permissions . Common scenarios for creating an agency: DLI is allowed to read and write data from and to OBS to transfer logs. DLI is allowed to access DEW to obtain data access credentials and access catalogs to obtain metadata. You cannot use the default agency names dli_admin_agency , dli_management_agency , or dli_data_clean_agency . It must be unique. For details about the permission policies of an agency, see Agency Permission Policies in Common Scenarios .

Table 8-2 Permissions contained in the `dli_management_agency` agency

Policy	Description
IAM ReadOnlyAccess	To authorize IAM users who have not logged in to DLI, you need to obtain their information. So, the permissions contained in the IAM ReadOnlyAccess policy are required.
DLI Datasource Connections Agency Access	Permissions to access and use VPCs, subnets, routes, and VPC peering connections
DLI Notification Agency Access	Permissions to send notifications through SMN when a job fails to be executed

8.2 Updating the Permissions of a DLI Agency

By updating the permissions of a DLI agency, you can minimize the number of permissions granted. This approach aligns with the current service requirements and mitigates the risks associated with granting excessive permissions.

To resolve this issue and implement fine-grained agency permission control, DLI upgrades the system agency from **dli_admin_agency** to **dli_management_agency**. The new agency contains the permissions required for obtaining IAM user information, performing datasource operations, and sending message notifications. This effectively prevents uncontrolled permission issues related to the services associated with DLI. After the upgrade, the DLI agency is more flexible and more suitable for scenario-based agency customization for medium- and large-sized enterprises.

This section describes how to update the permissions of a DLI agency.

Notes and Constraints

- Only the tenant account or a member account of user group **admin** can authorize the service.
- DLI authorization needs to be conducted by project. The permissions of required agencies must be updated separately in each project. This means you need to switch to the corresponding project and then update the agency by following the instructions provided in this section.
- Among the permissions contained in **dli_management_agency**:
 - The authorization scope of the **IAM ReadOnlyAccess** policy covers all global service resources in all regions.
 - If you select this policy when updating a DLI agency in any region, this policy's permissions apply to the projects in all regions.
 - If you do not select this policy when updating an agency in any project, this policy's permissions will be revoked from all regions. This means that all projects cannot obtain IAM user information.

CAUTION

When updating agency permissions, you will need to select the **IAM ReadOnlyAccess** policy, as its authorization scope covers global service resources. If you deselect it and update the agency, all its permissions will become invalid in all regions and projects.

- The authorization scope of the **DLI Datasource Connections Agency Access** and **DLI Notification Agency Access** policies covers the project resources in specified regions.

These policies' permissions only apply to projects for which these policies are selected and the DLI agency permissions are updated. Projects for which these policies are not selected do not have the permissions required in datasource scenarios and the permission to send notifications using SMN.

Example 1 and **Example 2** demonstrate the agency permission differences caused by updating a DLI agency for different projects in a region.

Updating DLI Agency Permissions (dli_management_agency)

This part describes how to update **dli_management_agency**. If you still use **dli_admin_agency**, refer to the following:

1. In the navigation pane of the DLI console, choose **Global Configuration > Service Authorization**.
2. On the displayed page, select permissions for scenarios.

Click  on a permission card to view its detailed permission policies.

Table 8-3 describes these agencies.

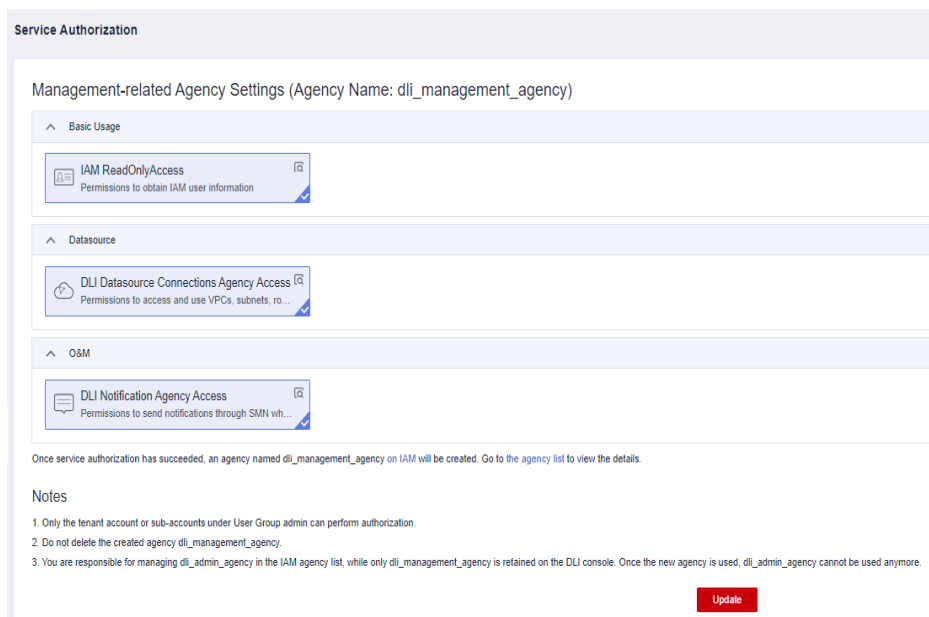
Table 8-3 Permissions contained in the dli_management_agency agency

Use Case	Agency	Description
Basic usage	IAM ReadOnlyAccess	To authorize IAM users who have not logged in to DLI, you need to obtain their information. So, the permissions contained in the IAM ReadOnlyAccess policy are required. IAM ReadOnlyAccess is a global policy. Make sure you select this policy. If you do not select it, all its permissions will become invalid in all regions, and the system cannot obtain IAM user information.
Datasource	DLI Datasource Connections Agency Access	Permissions to access and use VPCs, subnets, routes, and VPC peering connections

Use Case	Agency	Description
O&M	DLI Notification Agency Access	Permissions to send notifications through SMN when a job fails to be executed

3. Select the policies to be included in **dli_management_agency** and click **Update**.

Figure 8-2 Updating agency permissions



4. View and understand the notes for updating the agency, and click **OK**. The DLI agency permissions are updated.
 - Once the agency permissions are updated, your **dli_admin_agency** will be upgraded to **dli_management_agency**. This new agency will have the necessary permissions for datasource operations, notifications, and user authorization operations.
 - To use Flink 1.15, Spark 3.3.1 (Spark general queue scenario), or a later version to execute jobs, perform the following operations:
Create an agency on the IAM console and add the agency information to the job configuration. For details, see [Customizing DLI Agency Permissions](#).
 - Common scenarios for creating an agency: DLI is allowed to read and write data from and to OBS, dump logs, and read and write Flink checkpoints. DLI is allowed to access DEW to obtain data access credentials and access catalogs to obtain metadata.
 - You cannot use the default agency names **dli_admin_agency**, **dli_management_agency**, or **dli_data_clean_agency**. It must be unique.
 - If the engine version is earlier than Flink 1.15, **dli_admin_agency** is used by default during job execution. If the engine version is earlier than Spark

3.3.1, user authentication information (AK/SK and security token) is used during job execution.

- To maintain compatibility with existing job agency permission requirements, **qli_admin_agency** will still be listed in the IAM agency list even after the update.
- Do not delete the agency created by the system by default.

Follow-Up Operations

In addition to the permissions provided by **qli_management_agency**, you need to create an agency on the IAM console and add information about the new agency to the job configuration for scenarios like allowing DLI to read and write data from and to OBS to transfer logs, or allowing DLI to access DEW to obtain data access credentials. For details, see [Customizing DLI Agency Permissions](#) and [Agency Permission Policies in Common Scenarios](#).

- When Flink 1.15, Spark 3.3.1 (Spark general queue scenario), or a later version is used to execute jobs, you need to create an agency on the IAM console.
- If the engine version is earlier than Flink 1.15, **qli_admin_agency** is used by default during job execution. If the engine version is earlier than Spark 3.3.1, user authentication information (AK/SK and security token) is used during job execution.

This means that jobs whose engine versions are earlier than Flink 1.15 or Spark 3.3.1 are not affected by the update of agency permissions and do not require custom agencies.

Common service scenarios where you need to create an agency:

- Data cleanup agency required for clearing data according to the lifecycle of a table and clearing lakehouse table data. You need to create a DLI agency named **qli_data_clean_agency** on IAM and grant permissions to it. You need to create an agency and customize permissions for it. However, the agency name is fixed to **qli_data_clean_agency**.
- **Tenant Administrator** permissions are required to access data from OBS to execute Flink jobs on DLI, for example, obtaining OBS data sources, log dump (including bucket authorization), checkpointing enabling, and job import and export.
- The AK/SK required by DLI Flink jobs is stored in DEW. To allow DLI to access DEW data during job execution, you need to create an agency to delegate the permissions to operate on DEW data to DLI.
- To allow DLI to access DLI catalogs to retrieve metadata when executing jobs, you need to create a new agency that grants DLI catalog data operation permissions to DLI. This will enable DLI to access DLI catalogs on your behalf.
- Cloud data required by DLI Flink jobs is stored in LakeFormation. To allow DLI to access catalogs to retrieve metadata during job execution, you need to create an agency to delegate the permissions to operate on catalog data to DLI.

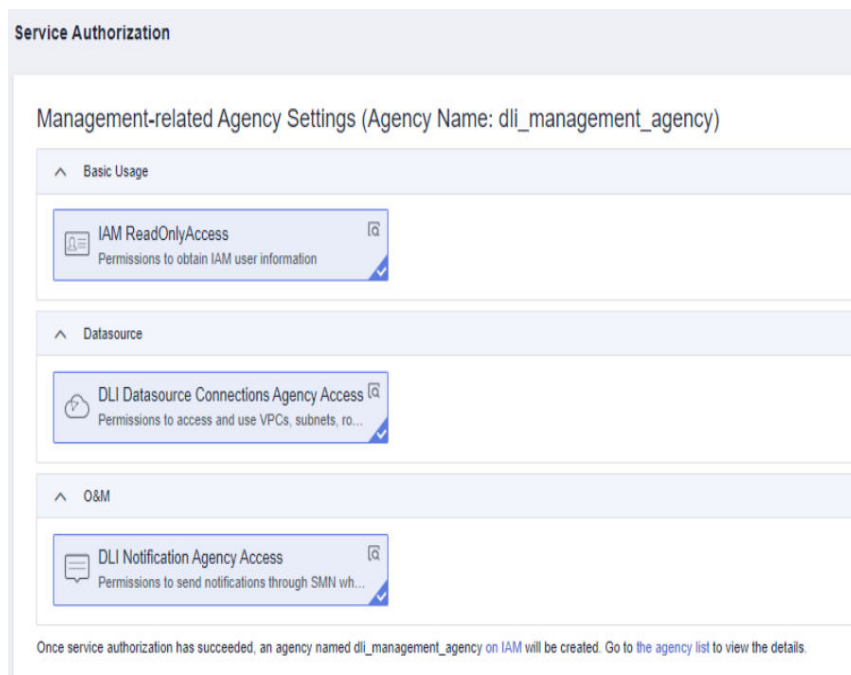
When creating an agency, you cannot use the default agency names **qli_admin_agency**, **qli_management_agency**, or **qli_data_clean_agency**. It must be unique.

For more information about custom agency operations, see [Customizing DLI Agency Permissions](#) and [Agency Permission Policies in Common Scenarios](#).

Example 1

- **Operation instruction:** A DLI user upgrades `dli_admin_agency` to `dli_management_agency` for project A in **CN North-Beijing4**.
 - a. On the DLI management console, switch to project A in the **CN North-Beijing4** region and choose **Global Configuration > Service Authorization**.
 - b. Select the policies under **Basic Usage**, **Datasource**, and **O&M**.

Figure 8-3 Updating a DLI agency for project A in CN North-Beijing4



- c. Click **Update**.
- **Permission description:** The agency permissions are updated for project A in **CN North-Beijing4**.
 - The **IAM ReadOnlyAccess** policy's permissions are granted to global service resources, meaning that all regions and projects have these permissions.
 - The **DLI Datasource Connections Agency Access** and **DLI Notification Agency Access** policies contain only regional permissions, meaning that their permissions only apply to project A in **CN North-Beijing4**.

Example of Agency Permissions for Project A in CN North-Beijing4	Example of Agency Permissions for Project B in CN North-Beijing4
<p>dli_management_agency</p> <p>The new agency contains the following policy permissions:</p> <ul style="list-style-type: none"> • IAM ReadOnlyAccess • DLI Datasource Connections Agency Access • DLI Notification Agency Access 	<p>dli_management_agency</p> <p>The new agency contains the following policy permissions:</p> <ul style="list-style-type: none"> • IAM ReadOnlyAccess

Example 2

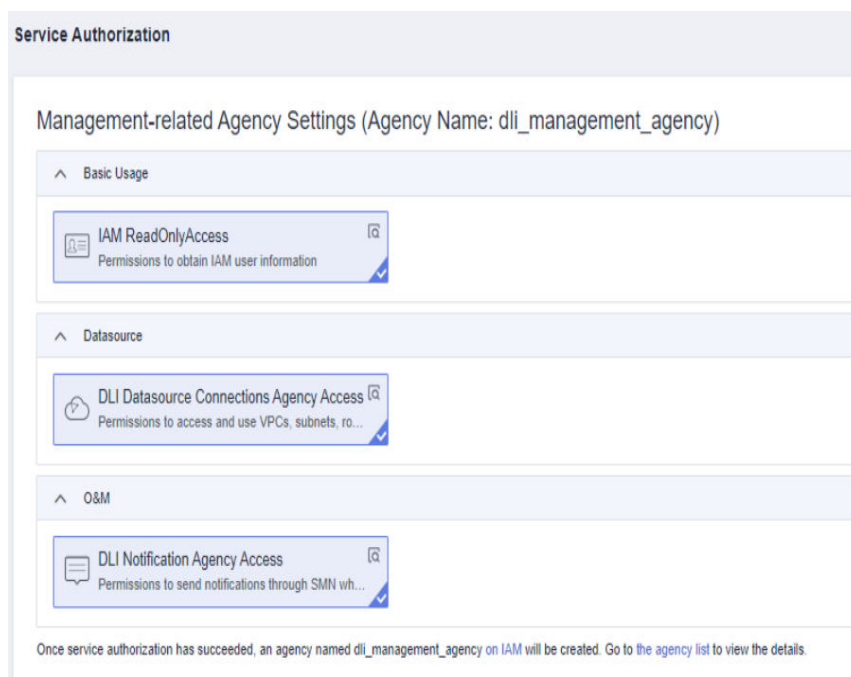
Operation instruction: To assign the permissions of the **DLI Datasource Connections Agency Access** and **DLI Notification Agency Access** policies to project B in the **CN North-Beijing4** region, perform the following operations to update the agency permissions of project B in **CN North-Beijing4**:

1. On the DLI management console, switch to project B in the **CN North-Beijing4** region and choose **Global Configuration > Service Authorization**.
2. Select the policies under **Basic Usage**, **Datasource**, and **O&M**.

 **CAUTION**

When updating the agency permissions for project B, you will need to select the **IAM ReadOnlyAccess** policy, as its authorization scope covers global service resources. If you deselect it and update the agency, all its permissions will become invalid in all regions and projects.

Figure 8-4 Updating a DLI agency for project B in CN North-Beijing4



3. Click **Update**.

Permission description:

The authorization scope of the **DLI Datasource Connections Agency Access** and **DLI Notification Agency Access** policies covers the project resources in specified regions. Following updates to agency permissions, project B in **CN North-Beijing4** has the permission to obtain IAM user information, perform datasource operations, and send message notifications.

Example of Agency Permissions in Region A	Example of Agency Permissions in Region B
<p>dli_management_agency</p> <p>The new agency contains the following policy permissions:</p> <ul style="list-style-type: none"> • IAM ReadOnlyAccess • DLI Datasource Connections Agency Access • DLI Notification Agency Access 	<p>dli_management_agency</p> <p>The new agency contains the following policy permissions:</p> <ul style="list-style-type: none"> • IAM ReadOnlyAccess • DLI Datasource Connections Agency Access • DLI Notification Agency Access

8.3 Customizing DLI Agency Permissions

When Flink 1.15, Spark 3.3, or a later version is used to execute jobs and the required agency is not included in the DLI system agency **dli_management_agency**, you need to create an agency on the IAM console and add information about the new agency to the job configuration. **dli_management_agency** contains the permissions required for datasource operations, message notifications, and user authorization operations. For other

agency permission requirements, you need to create custom DLI agencies. For details about **dli_management_agency**, see [DLI Agency Overview](#).

This section walks you through on how to create a custom agency, complete service authorization, and add information about the new agency to the job configuration.

DLI Custom Agency Scenarios

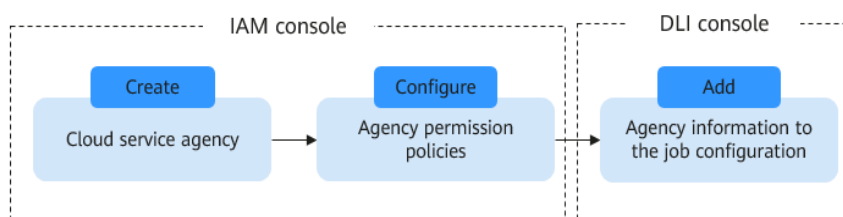
Table 8-4 DLI custom agency scenarios

Scenario	Agency Name	Description	Permission Policy
Allowing DLI to clear data according to the lifecycle of a table	dli_data_clean_agency	Data cleanup agency, which is used to clean up data according to the lifecycle of a table and clean up lakehouse table data. You need to create an agency and customize permissions for it. However, the agency name is fixed to dli_data_clean_agency .	Data Cleanup Agency Permission Configuration
Allowing DLI to read and write data from and to OBS to transfer logs	Custom	For DLI Flink jobs, the permissions include downloading OBS objects, obtaining OBS/GaussDB(DWS) data sources (foreign tables), transferring logs, using savepoints, and enabling checkpointing. For DLI Spark jobs, the permissions allow downloading OBS objects and reading/writing OBS foreign tables.	Permission Policies for Accessing and Using OBS
Allowing DLI to obtain data access credentials by accessing DEW	Custom	DLI jobs use DEW-CSMS' secret management.	Permission to Use DEW's Encryption Function
Allowing DLI to access DLI catalogs to retrieve metadata	Custom	DLI accesses catalogs to retrieve metadata.	Permission to Access DLI Catalog Metadata

Scenario	Agency Name	Description	Permission Policy
Allowing DLI to access LakeFormation catalogs to retrieve metadata	Custom	DLI accesses LakeFormation catalogs to retrieve metadata.	Permission to Access LakeFormation Catalog Metadata

Procedure

Figure 8-5 Process of creating a custom agency



Notes and Constraints

- The custom agency name cannot be the default agency names **qli_admin_agency**, **qli_management_agency**, or **qli_data_clean_agency**. It must be unique.
- The name of the agency that allows DLI to clear data according to the lifecycle of a table must be **qli_data_clean_agency**.
- Custom agencies can be configured only for jobs executed by engines like Flink 1.15, Spark 3.3.1 (Spark general queue scenario), or later versions.
- Once the agency permissions are updated, your **qli_admin_agency** will be upgraded to **qli_management_agency**. This new agency will have the necessary permissions for datasource operations, notifications, and user authorization operations. If you have other agency permission needs, you will need to create custom agencies. For details about **qli_management_agency**, see [DLI Agency Overview](#).
- Common scenarios for creating an agency: DLI is allowed to read and write data from and to OBS, dump logs, and read and write Flink checkpoints. DLI is allowed to access DEW to obtain data access credentials and access catalogs to obtain metadata. For details about agency permissions in these scenarios, see [Agency Permission Policies in Common Scenarios](#).

Step 1: Create a Cloud Service Agency on the IAM Console and Grant Permissions

1. Log in to the management console.
2. In the upper right corner of the page, hover over the username and select **Identity and Access Management**.
3. In the navigation pane of the IAM console, choose **Agencies**.

4. On the displayed page, click **Create Agency**.
5. On the **Create Agency** page, set the following parameters:
 - **Agency Name:** Enter an agency name, for example, **dli_obs_agency_access**.
 - **Agency Type:** Select **Cloud service**.
 - **Cloud Service:** This parameter is available only when you select **Cloud service** for **Agency Type**. Select **Data Lake Insight (DLI)** from the drop-down list.
 - **Validity Period:** Select **Unlimited**.
 - **Description:** You can enter **Agency with OBS OperateAccess permissions**. This parameter is optional.
6. Click **Next**.
7. Click the agency name. On the displayed page, click the **Permissions** tab. Click **Authorize**. On the displayed page, click **Create Policy**.
8. Configure policy information.
 - a. Enter a policy name, for example, **dli-obs-agency**.
 - b. Select **JSON**.
 - c. In the **Policy Content** area, paste a custom policy.

In this example, the permissions allow access and usage of OBS in various scenarios. For DLI Flink jobs, this includes downloading OBS objects, obtaining OBS/GaussDB(DWS) data sources (foreign tables), transferring logs, using savepoints, and enabling checkpointing. For DLI Spark jobs, the permissions allow downloading OBS objects and reading/writing OBS foreign tables.

For details about how to configure common agency permissions for Flink jobs, see [Agency Permission Policies in Common Scenarios](#).

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "obs:bucket:GetBucketPolicy",
        "obs:bucket:GetLifecycleConfiguration",
        "obs:bucket:GetBucketLocation",
        "obs:bucket:ListBucketMultipartUploads",
        "obs:bucket:GetBucketLogging",
        "obs:object:GetObjectVersion",
        "obs:bucket:GetBucketStorage",
        "obs:bucket:GetBucketVersioning",
        "obs:object:GetObject",
        "obs:object:GetObjectVersionAcl",
        "obs:object:DeleteObject",
        "obs:object:ListMultipartUploadParts",
        "obs:bucket:HeadBucket",
        "obs:bucket:GetBucketAcl",
        "obs:bucket:GetBucketStoragePolicy",
        "obs:object:AbortMultipartUpload",
        "obs:object:DeleteObjectVersion",
        "obs:object:GetObjectAcl",
        "obs:bucket:ListBucketVersions",
        "obs:bucket:ListBucket",
        "obs:object:PutObject"
      ],
      "Resource": [
        "OBS:*:bucket:bucketName", // Replace bucketName with the actual bucket name.
      ]
    }
  ]
}
```

```
    "OBS:*:*:object:*"  
  ],  
},  
{  
  "Effect": "Allow",  
  "Action": [  
    "obs:bucket:ListAllMyBuckets"  
  ]  
}  
]
```

- d. Enter a policy description as required.
9. Click **Next**.
10. On the **Select Policy/Role** page, select **Custom policy** from the first drop-down list and select the custom policy created in **8**.
11. Click **Next**. On the **Select Scope** page, set the authorization scope.

For details about authorization operations, see [Creating a User Group and Assigning Permissions](#).

- **All resources:** IAM users will be able to use all resources, including those in enterprise projects, region-specific projects, and global services under your account based on assigned permissions.
- **Global services:** Global services are deployed in all regions, so you can access them seamlessly without having to switch between regions. However, global services cannot be authorized based on regional projects. They include services like Object Storage Service (OBS) and Content Delivery Network (CDN). After authorization, you can use the selected services based on your permissions.
- **Region-specific projects:** IAM users will be able to use resources in the selected region-specific projects based on assigned permissions.
- **Enterprise projects:** IAM users will be able to use resources in the selected enterprise projects based on assigned permissions. For example, an enterprise project may contain resources that are deployed in different regions. After you associate the enterprise project with the IAM users, they can access the resources in this enterprise project based on the assigned permissions.

In this example, the custom policy is an OBS agency. So, select **Global services**. If a DLI agency is used, you are advised to select **Region-specific projects**.

12. Click **OK**.

It takes 15 to 30 minutes for the authorization to be in effect.

Step 2: Set Agency Permissions for a Job

When Flink 1.15, Spark 3.3, or a later version is used to execute jobs, you need to add information about the new agency to the job configuration.

Otherwise, If you do not specify an agency for Spark 3.3.1 jobs, the jobs cannot use OBS. If you do not specify an agency for a Flink 1.15 job, checkpointing cannot be enabled, savepoints cannot be used, logs cannot be transferred, and data sources such as OBS and GaussDB(DWS) cannot be used.

 **CAUTION**

- You can only specify an agency for Flink 1.15 and Spark 3.3.1 jobs running on queues in an elastic resource pool.
- After specifying an agency for a job, be careful when modifying the permissions granted to the agency. Any changes made may impact the job's normal operation.

- **Specifying an agency for a Flink Jar job**

- a. Log in to the DLI console. In the navigation pane, choose **Job Management > Flink Jobs**.
- b. Select a desired job and click **Edit** in the **Operation** column.
- c. In the job configuration area on the right, configure agency information.
 - **Flink Version:** Select **1.15**.
 - **Runtime Configuration:** Configure the key-value information of the new agency. The key is fixed to **flink.dli.job.agency.name**, and the value is the custom agency name.

In this example, set the key-value information to **flink.dli.job.agency.name=dli_obs_agency_access**.

- **Specifying an agency for a Flink OpenSource SQL job**

- a. Log in to the DLI console. In the navigation pane, choose **Job Management > Flink Jobs**.
- b. Select a desired job and click **Edit** in the **Operation** column.
- c. In the job configuration area on the right, configure agency information.
 - On the **Running Parameters** tab, ensure that the selected Flink version is **1.15**.
 - Click **Runtime Configuration**. Configure the key-value information of the new agency. The key is fixed to **flink.dli.job.agency.name**, and the value is the custom agency name.

In this example, set the key-value information to **flink.dli.job.agency.name=dli_obs_agency_access**.

- **Specifying an agency for a Spark job**

- a. Log in to the DLI console. In the navigation pane, choose **Job Management > Spark Jobs**.
- b. Select the target job and click **Edit** in the **Operation** column.
- c. In the job configuration area on the right, configure agency information.
 - Ensure that the selected Spark version is **3.3.1**.
 - In the **Spark Arguments(--conf)** area, configure the key-value information of the new agency. The key is fixed to **spark.dli.job.agency.name**, and the value is the custom agency name.

In this example, set the key-value information to **spark.dli.job.agency.name=dli_obs_agency_access**.

8.4 Agency Permission Policies in Common Scenarios

This section provides agency permission policies for common scenarios, which can be used to configure agency permission policies when you customize your permissions. The "Resource" in the agency policy should be replaced according to specific needs.

Data Cleanup Agency Permission Configuration

Application scenario: Data cleanup agency, which is used to clean up data according to the lifecycle of a table and clean up lakehouse table data. You need to create an agency and customize permissions for it. However, the agency name is fixed to **qli_data_clean_agency**.

NOTE

Set the authorization scope of an agency as follows:

- For an OBS agency, select **Global services**.
- For a DLI agency, select **Region-specific projects**.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "obs:object:GetObject",
        "obs:object:DeleteObject",
        "obs:bucket:HeadBucket",
        "obs:bucket:ListBucket",
        "obs:object:PutObject"
      ]
    }
  ]
}
```

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:table:showPartitions",
        "dli:table:select",
        "dli:table:dropTable",
        "dli:table:alterTableDropPartition"
      ]
    }
  ]
}
```

Permission Policies for Accessing and Using OBS

Application scenario: For DLI Flink jobs, the permissions include downloading OBS objects, obtaining OBS/GaussDB(DWS) data sources (foreign tables), transferring logs, using savepoints, and enabling checkpointing. For DLI Spark jobs, the permissions allow downloading OBS objects and reading/writing OBS foreign tables.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "obs:bucket:GetBucketPolicy",
        "obs:bucket:GetLifecycleConfiguration",
        "obs:bucket:GetBucketLocation",
        "obs:bucket:ListBucketMultipartUploads",
        "obs:bucket:GetBucketLogging",
        "obs:object:GetObjectVersion",
        "obs:bucket:GetBucketStorage",
        "obs:bucket:GetBucketVersioning",
        "obs:object:GetObject",
        "obs:object:GetObjectVersionAcl",
        "obs:object:DeleteObject",
        "obs:object:ListMultipartUploadParts",
        "obs:bucket:HeadBucket",
        "obs:bucket:GetBucketAcl",
        "obs:bucket:GetBucketStoragePolicy",
        "obs:object:AbortMultipartUpload",
        "obs:object:DeleteObjectVersion",
        "obs:object:GetObjectAcl",
        "obs:bucket:ListBucketVersions",
        "obs:bucket:ListBucket",
        "obs:object:PutObject"
      ],
      "Resource": [
        "OBS:*:bucket:bucketName ,// Replace bucketName with the actual bucket name.",
        "OBS:*:object:*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "obs:bucket:ListAllMyBuckets"
      ]
    }
  ]
}
```

Permission to Use DEW's Encryption Function

Application scenario: DLI Flink and Spark jobs use DEW-CSMS' secret management.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "csms:secretVersion:get",
        "csms:secretVersion:list",
        "kms:dek:decrypt"
      ]
    }
  ]
}
```

Permission to Access DLI Catalog Metadata

Application scenario: DLI Flink and Spark jobs are authorized to access DLI metadata.

```
{
  "Version": "1.1",
```

```
"Statement": [  
  {  
    "Effect": "Allow",  
    "Action": [  
      "dli:table:showPartitions",  
      "dli:table:alterTableAddPartition",  
      "dli:table:alterTableAddColumns",  
      "dli:table:alterTableRenamePartition",  
      "dli:table:delete",  
      "dli:column:select",  
      "dli:database:dropFunction",  
      "dli:table:insertOverwriteTable",  
      "dli:table:describeTable",  
      "dli:database:explain",  
      "dli:table:insertIntoTable",  
      "dli:database:createDatabase",  
      "dli:table:alterView",  
      "dli:table:showCreateTable",  
      "dli:table:alterTableRename",  
      "dli:table:compaction",  
      "dli:database:displayAllDatabases",  
      "dli:database:dropDatabase",  
      "dli:table:truncateTable",  
      "dli:table:select",  
      "dli:table:alterTableDropColumns",  
      "dli:table:alterTableSetProperties",  
      "dli:database:displayAllTables",  
      "dli:database:createFunction",  
      "dli:table:alterTableChangeColumn",  
      "dli:database:describeFunction",  
      "dli:table:showSegments",  
      "dli:database:createView",  
      "dli:database:createTable",  
      "dli:table:showTableProperties",  
      "dli:database:showFunctions",  
      "dli:database:displayDatabase",  
      "dli:table:alterTableRecoverPartition",  
      "dli:table:dropTable",  
      "dli:table:update",  
      "dli:table:alterTableDropPartition"  
    ]  
  }  
]
```

Permission to Access LakeFormation Catalog Metadata

Application scenario: DLI Spark jobs are authorized to access LakeFormation metadata.

```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": [  
        "lakeformation:table:drop",  
        "lakeformation:table:create",  
        "lakeformation:policy:create",  
        "lakeformation:database:create",  
        "lakeformation:database:drop",  
        "lakeformation:database:describe",  
        "lakeformation:catalog:alter",  
        "lakeformation:table:alter",  
        "lakeformation:database:alter",  
        "lakeformation:catalog:create",  
        "lakeformation:function:describe",  
        "lakeformation:catalog:describe",  
        "lakeformation:function:create",  
      ]  
    }  
  ]  
}
```

```

        "lakeformation:table:describe",
        "lakeformation:function:drop",
        "lakeformation:transaction:operate"
    ]
}
]
}

```

8.5 Guidelines for Configuring DLI Agency Permissions in Specific Scenarios

Table 8-5 Guidelines for configuring DLI agency permissions in specific scenarios

Type	Description
Flink job	Guideline for using DEW to manage and access credentials for Flink OpenSource SQL jobs. When writing the output data of Flink jobs to MySQL or GaussDB(DWS), set attributes such as the username and password in the connector.
	Guideline for Flink Jar jobs to acquire an AK/SK to read and write data from and to OBS.
	DLI provides a common interface to obtain temporary credentials for Flink job agencies set by users during job launch. The interface encapsulates the obtained temporary credentials for the job agency in the com.huaweicloud.sdk.core.auth.BasicCredentials class. Guideline for obtaining a temporary credential for a Flink job agency.
Spark job	Guideline for Spark Jar jobs to acquire an AK/SK to read and write data from and to OBS.
	Guideline for obtaining a temporary credential for a Spark Jar job agency.

9 Creating a DLI Global Variable

What Is a Global Variable?

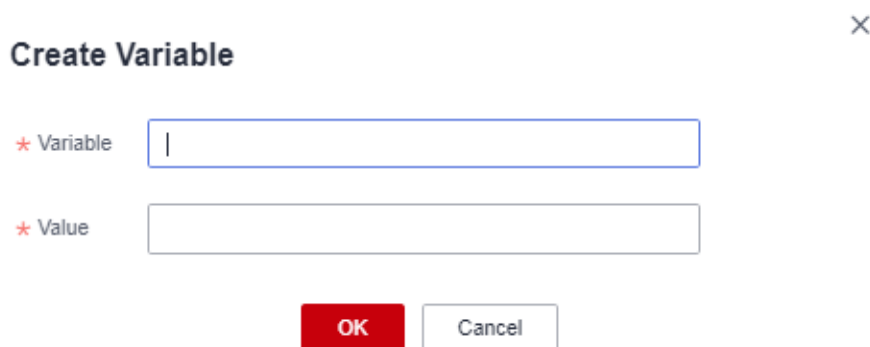
DLI allows you to set variables that are frequently used during job development as global variables on the DLI management console. This avoids repeated definitions during job editing and reduces development and maintenance costs. Global variables can be used to replace long and difficult variables, simplifying complex parameters and improving the readability of SQL statements.

This section describes how to create a global variable.

Creating a Global Variable

1. In the navigation pane of the DLI console, choose **Global Configuration > Global Variables**.
2. On the **Global Variables** page, click **Create** in the upper right corner to create a global variable.

Figure 9-1 Creating a global variable



Create Variable ×

* Variable

* Value

OK

Table 9-1 Global variable parameters

Parameter	Description
Variable	Name of the created global variable.

Parameter	Description
Value	Global variable value.

- After creating a global variable, use `{{xxxx}}` in the SQL statement to replace the parameter value set as the global variable. `xxxx` indicates the variable name. For example, if you set global variable `abc` to represent the table name, replace the actual table name with `{{abc}}` in the table creation statement.

```
create table {{table_name}} (String1 String, int4 int, varchar1 varchar(10))  
partitioned by (int1 int,int2 int,int3 int)
```

NOTE

Do not use global variables in **OPTIONS** of the table creation statements.

Related operations:

- **Modifying a global variable**

On the **Global Variables** page, locate a desired variable and click **Modify** in the **Operation** column.

NOTE

If there are multiple global variables with the same name in the same project under an account, delete the redundant global variables to ensure that the global variables are unique in the same project. In this case, all users who have the permission to modify the global variables can change the variable values.

- **Deleting a global variable**

On the **Global Variables** page, click **Delete** in the **Operation** column of a variable to delete the variable value.

NOTE

- If there are multiple global variables with the same name in the same project under an account, delete the global variables created by the user first. If there are only unique global variables, all users who have the delete permission can delete the global variables.
- After a variable is deleted, the variable cannot be used in SQL statements.

Permissions Management for Global Variables

You can assign different users different global variables through permission settings. The administrator and owners of global variables have all permissions. You do not need to set permissions for them, and their global variable permissions cannot be modified by other users.

When setting global variables for a new user, the region hosting the user's group must have the **Tenant Guest** permission. For details about the **Tenant Guest** permission and how to apply for the permission, see [System-defined Permissions](#) and [Creating a User Group and Assigning Permissions](#).

• **Granting permissions on a global variable to a user**

- In the navigation pane on the left of the DLI console, choose **Global Configuration > Global Variables**. On the displayed page, locate a desired global variable and click **Set Permission** in the **Operation**

column. On the displayed **User Permissions** page, you can grant, set, and revoke permissions on the global variable to users.

- b. Click **Grant Permission** in the upper right corner.

Figure 9-2 Granting permissions of a global variable to a user

Table 9-2 Global variable parameters

Parameter	Description
Username	Name of the IAM user who is granted permissions NOTE This username must be an existing IAM username.
Permissions	<ul style="list-style-type: none"> ● Update: This permission allows you to update the global variable. ● Delete: This permission allows you to delete the global variable. ● Grant Permission: This permission allows you to grant permissions of the global variable to other users. ● Revoke Permission: This permission allows you to revoke the global variable permissions that other users have but cannot revoke the global variable owner's permissions. ● View Other User's Permissions: This permission allows you to view the global variable permissions of other users.

- **Granting permissions on a global variable to a user**

On the **User Permissions** page, locate a desired IAM user and click **Set Permission** in the **Operation** column. [Table 9-2](#) lists the permission parameters.

If all permission options are grayed out, it means you do not have the authority to modify the permissions on this global variable. You can request the modification permission from users who have authorization, such as the administrator or group owners.

- **Revoking permissions on a global variable from a user**

On the **User Permissions** page, locate a desired IAM user and click **Revoke Permission** in the **Operation** column. Once the revoke operation is complete, the IAM user will no longer have any permissions on the global variable.

10 Configuring a DLI Job Bucket

Scenario

When you use a DLI job, you need an OBS bucket to store temporary data, such as job logs and results generated during job running.

Configure a DLI job bucket on the **Global Configuration > Project** page of the DLI management console.

Before the configuration, create an OBS bucket or parallel file system (PFS). In big data scenarios, you are advised to create a PFS. PFS is a high-performance file system provided by OBS, with access latency in milliseconds. PFS can achieve a bandwidth performance of up to TB/s and millions of IOPS, which makes it ideal for processing high-performance computing (HPC) workloads.

For details about PFS, see [Parallel File System Feature Guide](#).

Notes

- The OBS bucket must be used to store temporary data generated by DLI, such as job logs and job results.
- Do not use the OBS bucket for other purposes.
- The OBS bucket must be set and modified by the main account. Member users do not have the permission.
- If the bucket is not configured, you will not be able to view job logs.
- You can create lifecycle rules to automatically delete objects or change storage classes for objects that meet specified conditions.
- Inappropriate modifications of the job bucket may lead to loss of historical data.

Procedure


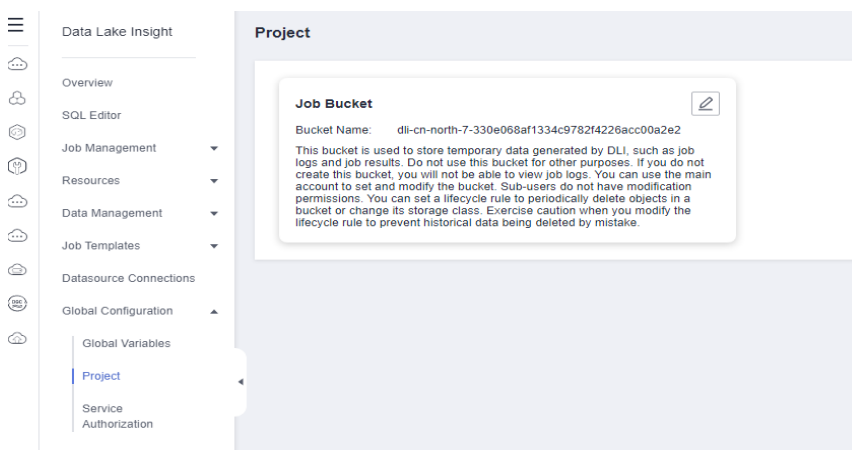
1. In the navigation pane of the DLI console, choose **Global Configuration > Project**.
2. On the **Project** page, click  next to **Job Bucket** to configure bucket information.

Figure 10-1 Project




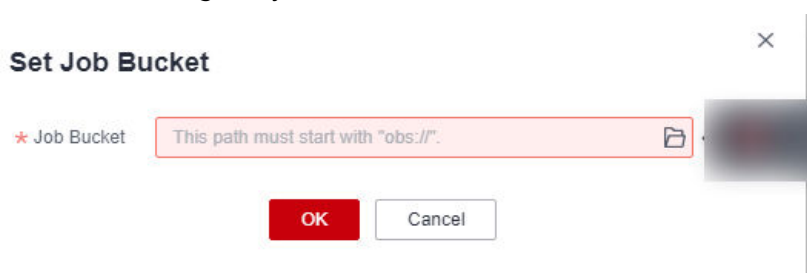
3. Click  to view available buckets.
4. Select the bucket for storing the temporary data of the DLI job and click **OK**. Temporary data generated during DLI job running will be stored in the OBS bucket.

Figure 10-2 Setting the job bucket



11 SQL Inspector

11.1 Overview

What Is SQL Inspection?

There are numerous SQL engines in the current big data field, which bring diversity to the solutions but also expose some issues such as varying quality of SQL input statements, difficult SQL problem localization, and excessive resource consumption by large SQL statements.

Poor quality SQL can have unforeseeable impacts on data analysis platforms, affecting system performance or platform stability.

 **NOTE**

Only Spark 3.3 or later supports this function.

Functions

DLI offers this feature to allow you to create inspection rules for the Spark SQL engine. This helps prevent common issues like large or low-quality SQL statements by providing pre-event information, blocking, and in-event circuit breaker. You do not need to change your SQL submission method or syntax, so it is easy to implement without affecting service operations.

- You can configure SQL inspection rules in a visualized manner and also have the ability to query and modify these rules.
- During query execution and response, each SQL engine proactively inspects SQL statements based on the rules.
- Administrators can choose to display hints on, intercept, or block SQL statements. The system logs SQL inspection events in real time for SQL audit. O&M engineers can analyze the logs, assess the quality of SQL statements on the live network, identify potential risks, and take preventive measures.

SQL inspection rules contain the following types of actions:

- **Info:** The system provides hints or blocks SQL statements based on SQL syntax rules when a SQL request meets certain conditions.

- **Block:** The system blocks SQL statements based on rules of data table statistics and metadata information. The system will cease processing a SQL request when certain conditions are met.
- **Circuit Breaker:** The system performs circuit breaking on SQL statements based on system states (such as CPU, memory, and I/O) during the runtime of the SQL statements. The system will perform circuit breaking on a SQL task when certain conditions are met.

Notes and Constraints of DLI SQL Inspection Rules

- SQL inspection is only supported by Spark 3.3 or later.
- Only one inspection rule can be created for an action or a queue.
- Each rule can be associated with a maximum of 50 SQL queues.
- A maximum of 1,000 rules can be created for each project.

11.2 Creating a SQL Inspection Rule

Scenario

You can create SQL inspection rules for specified SQL queues on the **SQL Inspector** page. The system will prompt, block, or perform circuit breaking on SQL requests that trigger the rules.

NOTE

When creating or modifying a SQL inspection rule, evaluate the appropriateness of enabling the rules and setting the threshold based on the service scenario. This will help avoid the negative impact of unreasonable inspection rules blocking or performing circuit breaking on relevant SQL requests on the services.

Notes and Constraints

- SQL inspection is only supported by Spark 3.3 or later.
For SQL queues, check the Spark engine version on the SparkUI page.
- Only one inspection rule can be created for an action or a queue.
- Each rule can be associated with 50 SQL queues.
- A maximum of 1,000 rules can be created for a project.

Procedure

1. Log in to the DLI management console.
2. In the navigation pane on the left, choose **Global Configuration > SQL Inspector**.
3. On the displayed **SQL Inspector** page, click **Create Rule** in the upper right corner. In the **Create Rule** dialog box, set parameters based on the table below.

Table 11-1 Parameters for creating a SQL inspection rule

Parameter	Description
Rule Name	Name of a SQL inspection rule
System Rules	Select an inspection rule. For details about the system inspection rules supported by DLI, see Creating a SQL Inspection Rule .
Queues	Select the queues the rules are bound to.
Description	Enter a rule description.
Rule Action	Actions that the current SQL inspection rule supports. SQL rules support the following types of actions: <ul style="list-style-type: none">• Info: Record logs and provide a hint for handling the SQL request. If the rule has parameters, you need to configure the threshold.• Block: Intercept the SQL request that meets the rule. If the rule has parameters, you need to configure the threshold.• Circuit Breaker: Perform circuit breaking on the SQL request that meets the inspection rules. If the rule has parameters, you need to configure the threshold.

4. Click **OK**.

View the added inspection rule on the **SQL Inspector** page. The rule takes effect dynamically.

To modify a rule, click **Modify** in its **Operation** column.

11.3 System Rules

[Table 11-2](#) lists the system inspection rules supported by DLI.

Default system rules are SQL inspection rules automatically created by DLI when a queue is created. These rules are bound to the queue and cannot be deleted.

Default system rules include **Scan files number**, **Scan partitions number**, **Count(distinct) occurrences**, and **Not in<Subquery>**.

Only one inspection rule can be created for an action or a queue.

For every supported action, a default system rule is created. For example, when a queue is created, a **Scan files number** rule is automatically created for both the **Info** and **Block** actions.

Table 11-2 System inspection rules supported by DLI

Rule ID	Rule Name	Description	Type	Applicable Engine	Action	Value	Default System Rule	Example SQL Statement
dynamic_0001	Scan files number	Maximum number of files to be scanned	Dynamic	Spark	Info Block	Value range: 1-200000 Default value: 200000	Yes	N/A
dynamic_0002	Scan partitions number	Maximum number of partitions involved in the operations (select, delete, update, and alter) that can be performed on a table	Dynamic	Spark	Info Block	Value range: 1-500000 Default value: 5000	Yes	DELETE FROM table_name WHERE column_name = value
running_0002	Memory used(MB)	Peak memory usage of the SQL statement	Running	Spark	Circuit Breaker	Value range: 1-8388608	No	N/A
running_0003	Run time(S)	Maximum running duration of the SQL statement	Running	Spark	Circuit Breaker	Unit: second Value range: 1-43200	No	N/A

Rule ID	Rule Name	Description	Type	Applicable Engine	Action	Value	Default System Rule	Example SQL Statement
running_0004	Scan data(GB)	Maximum amount of data to be scanned	Running	Spark	Circuit Breaker	Unit: GB Value range: 1-10240	No	N/A
static_0001	Count(distinct) occurrences	Maximum number of occurrences of count(distinct) in the SQL statement	Static	Spark	Info Block	Value range: 1-100 Default value: 10	Yes	SELECT COUNT(DISTINCT deviceId), COUNT(DISTINCT collDeviceId) FROM table GROUP BY deviceName, collDeviceName, collCurrentVersion;
static_0002	Not in<Subquery>	Check whether not in <subquery> is used in the SQL statement	Static	Spark	Info Block	Value range: Yes, No Default value: Yes	Yes	SELECT * FROM Orders o WHERE Orders.Order_ID not in (Select Order_ID FROM HeldOrders h where h.order_id = o.order_id);

Rule ID	Rule Name	Description	Type	Applicable Engine	Action	Value	Default System Rule	Example SQL Statement
static_0003	Join occurrences	Maximum number of joins in the SQL statement	Static	Spark	Info Block	Value range: 1-50	No	SELECT name, text FROM table_1 JOIN table_2 ON table_1.Id = table_2.Id
static_0004	Union occurrences	Maximum number of union all times in the SQL statement	Static	Spark	Info Block	Value range: 1-100	No	select * from tables t1 union all select * from tables t2 union all select * from tables t3
static_0005	Subquery nesting layers	Maximum number of subquery nesting layers	Static	Spark	Info Block	Value range: 1-20	No	select * from (with temp1 as (select * from tables) select * from temp1);
static_0006	Sql size(KB)	Maximum size of a SQL file	Static	Spark	Info Block	Unit: KB Value range: 1-1024	No	N/A

Rule ID	Rule Name	Description	Type	Applicable Engine	Action	Value	Default System Rule	Example SQL Statement
static_0007	Cartesian product	Limitation of Cartesian products when multiple tables are being associated	Static	Spark	Info Block	Value range: 0-1	No	select * from A,B;

12 Creating a Custom Image

Use Cases

To enhance the functions and performance of Spark and Flink jobs, you can create custom images by downloading the basic image provided by DLI and adding dependencies (files, JAR files, or software) and private capabilities required for job execution. This changes the container runtime environment for the jobs.

For example, you can add a Python package or C library related to machine learning to a custom image to help you extend functions.

NOTE

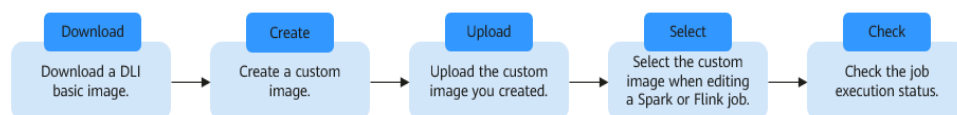
To use the custom image function, you need to have basic knowledge of [Docker](#).

Notes and Constraints

- The basic images provided by DLI must be used to create custom images.
- You cannot modify the DLI components and directories in the basic images.
- Only Spark and Flink Jar jobs are supported.

Use Process

Figure 12-1 Process of using a custom image



1. Obtain DLI basic images.
2. Use Dockerfile to pack dependencies (files, JAR files, or software) required for job execution into the basic image to create a custom image.
3. Publish the custom image to SoftWare Repository for Container (SWR).
4. On the DLI job editing page, select the created image and run the job.
5. Check the job execution status.

Obtaining DLI Basic Images

Table 12-1 Obtaining DLI basic images

Image Type	Architecture
General image	x86
General image	Arm
Notebook image	x86
Notebook image	Arm

Creating a Custom Image

The following describes how to package TensorFlow into an image to generate a custom image with TensorFlow installed. Then, you can use the image to run jobs in DLI.

Step 1 Prepare the container environment.

For details, see "Step 1: Install the Container Engine" in [Uploading an Image Through a Container Engine Client](#).

Step 2 Log in to the prepared container environment as user **root** and run a command to obtain the basic image.

In this example, the Spark basic image is used and downloaded to the container image environment in [Step 1](#) by running the following command:


```
docker pull Address for downloading the basic image
```

For details about the address, see [Creating a Custom Image](#).

For example, to download the Spark basic image, run the following command:

```
docker pull swr.xxx/dli-public/spark_general-x86_64:3.3.1-2.3.7.1720240419835647952528832.202404250955
```

Step 3 Access SWR.

1. Log in to the SWR management console.
2. In the navigation pane on the left, choose **Dashboard** and click **Generate Login Command** in the upper right corner. On the displayed page, click  to copy the login command.
3. Run the login command on the VM where the container engine is installed.

Step 4 Create an organization. If an organization has been created, skip this step.

1. Log in to the SWR management console.
2. In the navigation pane on the left, choose **Organization Management**. On the displayed page, click **Create Organization** in the upper right corner.
3. Enter the organization name and click **OK**.

Step 5 Write a Dockerfile.

vi Dockerfile

Pack TensorFlow into the image as follows:

```
ARG BASE_IMG=swr.xxx/dli-public/spark_general-  
x86_64:3.3.1-2.3.7.1720240419835647952528832.202404250955//Replace xxx with the URL of the basic  
image.  
  
FROM ${BASE_IMG} as builder  
USER omm //Run this command as user omm.  
RUN set -ex && \  
    mkdir -p /home/omm/.pip && \  
    pip3 install tensorflow==1.13.1 \  
    --user --no-cache-dir --trusted-host pypi.cloudartifact.dgg.dragon.tools.huawei.com \  
    -i https://pypi.cloudartifact.dgg.dragon.tools.huawei.com/artifactory/api/pypi/cbu-pypi-public/simple  
Copy the content to the basic image.  
USER omm
```

The following steps are included:

1. Set the available repository address of pip.
2. Use pip3 to install the TensorFlow algorithm package.
3. Copy the content in the temporary image builder where the algorithm package is installed to the basic image (this step is to reduce the image size) to generate the final custom image.

Step 6 Use Dockerfile to generate a custom image.

Format of the image packaging command:

```
docker build -t [Custom organization name]/[Custom image name]: [Image version] --build-arg  
BASE_IMG= [DLI basic image path] -f Dockerfile .
```

The DLI basic image path is the image path in [Table 12-1](#).

The following is an example:

```
docker build -t mydli/spark:2.4 --build-arg BASE_IMG=swr.xxx/dli-public/spark_general-  
x86_64:3.3.1-2.3.7.1720240419835647952528832.202404250955 -f Dockerfile .
```

Step 7 Add a tag to the custom image.

docker tag [Organization name]/[Image name]:[Image version][Image repository address]/[Organization name]/[Image name:version] in [Step 6](#)

The following is an example:

```
docker tag mydli/spark:2.4 swr.xxx/testdli0617/spark:2.4.5.tensorflow
```

Step 8 Upload the custom image.

docker push [Image repository address]/[Organization name]/[Image name:Version]

Set [Image repository address]/[Organization name]/[Image name:Version] the same as those in [Step 7](#).

The following is an example:

```
docker push swr.xxx/testdli0617/spark:2.4.5.tensorflow
```

Step 9 When submitting a Spark or Flink JAR job in DLI, select a custom image.

- Open the Spark job or Flink job editing page on the management console, select the uploaded and shared image from the custom image list, and run the job.

If you select a non- shared image, the system displays a message indicating that the image is not authorized. You can use the image only after it is authorized. Click **Authorize** as prompted as shown in [Figure 12-3](#). Configure other job execution parameters and execute the job.

Figure 12-2 Selecting a custom image on the DLI Spark job editing page

The screenshot shows the 'DLI Spark job editing page' with a sidebar on the left containing navigation options: 'Fill Form' (selected), 'Write API', 'Execute', and 'View History'. The main content area is titled 'Select a Queue' and contains several configuration sections: 'Queue' (dropdown menu), 'Job Configurations' (Job Name, Application, Main Class, Application Parameters, Spark Arguments), 'Job Type' (Basic, AI-enhanced, Image - selected), 'JAR Package Dependencies', 'Python File Dependencies', 'Other Dependencies', 'Group Name', 'Access Metadata', 'Retry upon Failure', and 'Advanced Settings' (Skip, Configure). The 'Job Type' section is highlighted with a red box, showing the 'Image' tab selected and a dropdown menu with 'Configure Image' and 'Use Custom Image' options.

Figure 12-3 Authorizing a Spark job image

The screenshot shows the 'DLI Spark job editing page' with the 'Job Type' section set to 'Image'. The 'Application' field is filled with 'sparkPi.jar'. The 'Main Class' field is empty. The 'Application Parameters' field contains '5'. The 'Spark Arguments' field contains 'spark.kubernetes.executor.deleteOnTermination=false'. The 'Job Type' section is highlighted with a red box, showing the 'Image' tab selected and a dropdown menu with 'Configure Image' and 'Use Custom Image' options. Below the dropdown menu, a red error message box is displayed: 'The el_dlics not authorized.' with an 'Authorize' button highlighted in red.

Figure 12-4 Selecting a custom image on the DLI Flink Jar job editing page

test_s [redacted]

ID: 93229 | Job Type: Flink Jar

* Queue

* Application [View Built-in Dependencies](#)

Main Class

Default main class is specified by the Manifest file of the application.

Class Arguments

JAR Package Dependenc... [View Built-in Dependencies](#)

Other Dependencies

Job Type

[Configure ImageUse](#) [Custom Image](#)

Flink Version

Runtime Configuration

- Specify the image parameter in job parameters on API to use a custom image to run a job.

For details about Spark jobs, see [Creating a Batch Processing Job](#). For details about Flink Jar jobs, see [Creating a Flink Jar Job](#).

----End

13 Permissions Management

13.1 Overview

DLI has a comprehensive permission control mechanism and supports fine-grained authentication through Identity and Access Management (IAM). You can create policies in IAM to manage DLI permissions. You can use both the DLI's permission control mechanism and the IAM service for permission management.

Application Scenarios of IAM Authentication

When using DLI on the public cloud, enterprise users need to manage DLI resources (queues) used by employees in different departments, including creating, deleting, using, and isolating resources. In addition, data of different departments needs to be managed, including data isolation and sharing.

DLI uses IAM for refined enterprise-level multi-tenant management. IAM provides identity authentication, permissions management, and access control, helping you securely access to your public cloud resources.

With IAM, you can use your Huawei Cloud account to create IAM users for your employees, and assign permissions to the users to control their access to specific resource types. For example, some software developers in your enterprise need to use DLI resources but must not delete them or perform any high-risk operations. To achieve this result, you can create IAM users for the software developers and grant them only the permissions required for using DLI resources.

NOTE

For a new user, you need to log in for the system to record the metadata before using DLI.

IAM is free to use, and you only need to pay for the resources in your account.

If your Huawei Cloud account does not need individual IAM users for permissions management, skip over this section.

DLI System Permissions

[Table 13-1](#) lists all the system-defined roles and policies supported by DLI.

Type: There are roles and policies.

- Roles: A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. Only a limited number of service-level roles are available. When using roles to grant permissions, you also need to assign other roles on which the permissions depend. However, roles are not an ideal choice for fine-grained authorization and secure access control.
- Policies: A type of fine-grained authorization mechanism that defines permissions required to perform operations on specific cloud resources under certain conditions. This mechanism allows for more flexible policy-based authorization, meeting requirements for secure access control. For example, you can grant DLI users only the permissions for managing a certain type of ECSs.

For details about the system policies you need to perform common SQL operations, see [Common Operations Supported by DLI System Policy](#).

Table 13-1 DLI system permissions

Role/Policy Name	Description	Category	Dependency
DLI FullAccess	Full permissions for DLI.	System-defined policy	This role depends on other roles in the same project. <ul style="list-style-type: none">• Creating a datasource connection: VPC ReadOnlyAccess• Creating yearly/monthly resources: BSS Administrator• Creating a tag: TMS FullAccess and EPS FullAccess• Using OBS for storage: OBS OperateAccess• Creating an agency: Security Administrator

Role/Policy Name	Description	Category	Dependency
DLI ReadOnlyAccess	<p>Read-only permissions for DLI.</p> <p>With read-only permissions, you can use DLI resources and perform operations that do not require fine-grained permissions. For example, create global variables, create packages and package groups, submit jobs to the default queue, create tables in the default database, create datasource connections, and delete datasource connections.</p>	System-defined policy	None
Tenant Administrator	<p>Tenant administrator</p> <ul style="list-style-type: none"> • Job execution permissions for DLI resources. After a database or a queue is created, the user can use the ACL to assign rights to other users. • Scope: project-level service 	System-defined role	None
DLI Service Administrator	<p>DLI administrator.</p> <ul style="list-style-type: none"> • Job execution permissions for DLI resources. After a database or a queue is created, the user can use the ACL to assign rights to other users. • Scope: project-level service 	System-defined role	None

For details, see [Creating an IAM User and Granting Permissions, How Do I Create an IAM user?](#) and [How Do I Modify a User Policy?](#)

DLI Permission Types

Table 13-2 lists the DLI service permissions. For details about the resources that can be controlled by DLI, see **Table 13-7**.

Table 13-2 DLI permission types

Permission Type	Subtype	Console Operations	SQL Syntax	API Definition
Queue Permissions	Queue management permissions	For details, see Queue Permission Management .	None	For details, see "Granting Users with the Queue Usage Permission" in the <i>Data Lake Insight API Reference</i> .
	Queue usage permission			
Data Permissions	Database permissions	For details, see Managing Database Permissions and Managing Table Permissions .	For details, see SQL Syntax of Batch Jobs > Data Permissions Management > Data Permissions List in the <i>Data Lake Insight SQL Syntax Reference</i> .	For details, see Permission-related APIs > Granting Users with the Data Usage Permission in the <i>Data Lake Insight API Reference</i> .
	Table permissions			
	Column permissions			
Job Permissions	Flink job permissions	For details, see Managing Flink Job Permissions .	None	For details, see Permission-related APIs > Granting Users with the Data Usage Permission in the <i>Data Lake Insight API Reference</i> .

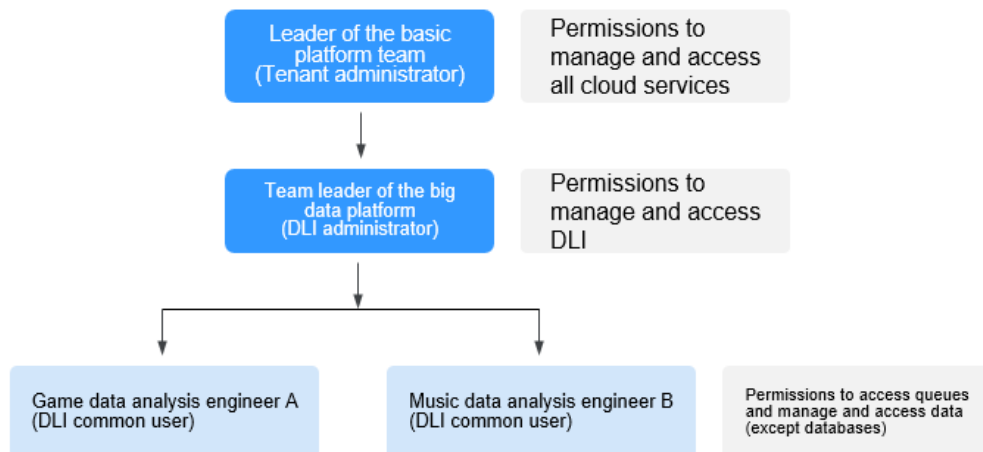
Permission Type	Subtype	Console Operations	SQL Syntax	API Definition
Package Permissions	Package group permissions	For details, see Managing Permissions on Packages and Package Groups .	None	For details, see Permission-related APIs > Granting Users with the Data Usage Permission in the <i>Data Lake Insight API Reference</i> .
	Package permissions			
Datasource Connection Permissions	Datasource connection permissions	For details, see Datasource Authentication Permission Management .	None	For details, see Permission-related APIs > Granting Users with the Data Usage Permission in the <i>Data Lake Insight API Reference</i> .

Examples

An Internet company mainly provides game and music services. DLI is used to analyze user behaviors and assist decision making.

As shown in [Figure 13-1](#), the **Leader of the Basic Platform Team** has applied for a **Tenant Administrator** account to manage and use public cloud services. The **Leader of the Basic Platform Team** creates a subaccount with the **DLI Service Administrator** permission to manage and use DLI, as the **Big Data Platform Team** requires DLI for data analysis. The **Leader of the Basic Platform Team** creates a **Queue A** and assigns it to **Data Engineer A** to analyze the gaming data. A **Queue B** is also assigned to **Data Engineer B** to analyze the music data. Besides granting the queue usage permission, the **Leader of the Basic Platform Team** grants data (except the database) management and usage permissions to the two engineers.

Figure 13-1 Granting permissions



The **Data Engineer A** creates a table named **gameTable** for storing game prop data and a table named **userTable** for storing game user data. The music service is a new service. To explore potential music users among existing game users, the **Data Engineer A** assigns the query permission on the **userTable** to the **Data Engineer B**. In addition, **Data Engineer B** creates a table named **musicTable** for storing music copyrights information.

Table 13-3 describes the queue and data permissions of **Data Engineer A** and **Data Engineer B**.

Table 13-3 Permission description

User	Data Engineer A (game data analysis)	Data Engineer B (music data analysis)
Queues	Queue A (queue usage permission)	Queue B (queue usage permission)
Data (Table)	gameTable (table management and usage permission)	musicTable (table management and usage permission)
	userTable (table management and usage permission)	userTable (table query permission)

NOTE

The queue usage permission includes job submitting and terminating permissions.

13.2 Creating an IAM User and Granting Permissions

You can use Identity and Access Management (IAM) to implement fine-grained permissions control on DLI resources. For details, see [Overview](#).

If your cloud account does not need individual IAM users, then you may skip over this section.

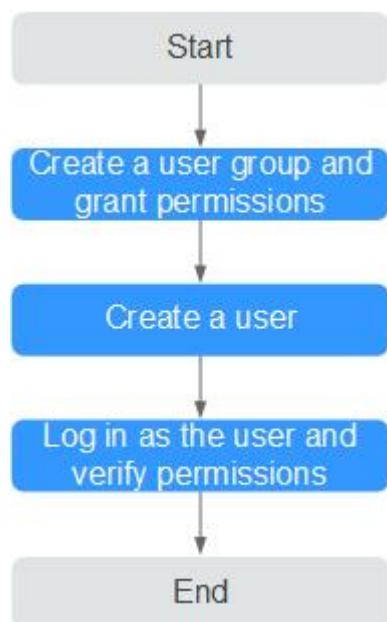
This section describes how to create an IAM user and grant DLI permissions to the user. [Figure 13-2](#) shows the procedure.

Prerequisites

Before assigning permissions to user groups, you should learn about system policies and select the policies based on service requirements. For details about system permissions supported by DLI, see [DLI System Permissions](#).

Process Flow

Figure 13-2 Process for granting DLI permissions



1. Create a user group and grant the permission to it.
Create a user group on the IAM console, and assign the **DLI ReadOnlyAccess** permission to the group.
2. Create a user and add the user to the user group.
Create a user on the IAM console and add the user to the group created in [1](#).
3. Log in and verify the permission.
Log in to the management console using the newly created user, and verify that the user's permissions.
 - Choose **Service List > Data Lake Insight**. The DLI management console is displayed. If you can view the queue list on the **Queue Management** page but cannot buy DLI queues by clicking **Buy Queue** in the upper right corner (assume that the current permission contains only **DLI ReadOnlyAccess**), the **DLI ReadOnlyAccess** permission has taken effect.
 - Choose any other service in **Service List**. If a message appears indicating that you have insufficient permissions to access the service, the **DLI ReadOnlyAccess** permission has already taken effect.

13.3 Creating a Custom Policy

Custom policies can be created as a supplement to the system policies of DLI. You can add actions to custom policies. For the actions supported for custom policies, see "Permissions Policies and Supported Actions" in the *Elastic Volume Service API Reference*.

You can create custom policies in either of the following two ways:

- Visual editor: Select cloud services, actions, resources, and request conditions without the need to know policy syntax.
- JSON: Create a policy in the JSON format from scratch or based on an existing policy.

. This section describes common DLI custom policies.

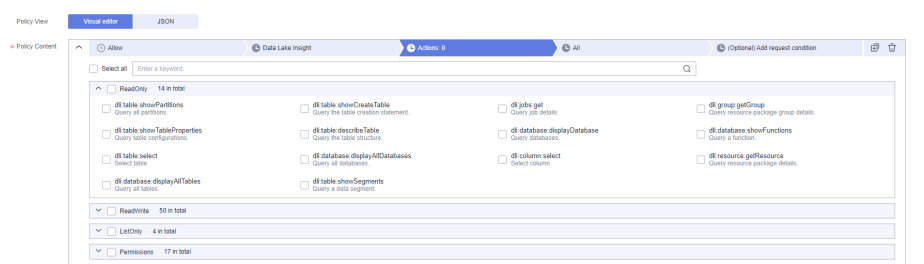
Policy Field Description

The following example assumes that the authorized user has the permission to create tables in all databases in all regions:

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:createTable"
      ],
      "Resource": [
        "dli:*:*:database:*"
      ]
    }
  ]
}
```

- Version
1.1 indicates a fine-grained permission policy that defines permissions required to perform operations on specific cloud resources under certain conditions.
- Effect
 The value can be **Allow** and **Deny**. If both **Allow** and **Deny** are found in statements, the **Deny** overrides the **Allow**.
- Action
 Specific operation on a resource. A maximum of 100 actions are allowed, as shown in [Figure 13-3](#).

Figure 13-3 DLI actions



 NOTE

- The format is *Service name.Resource type.Action*, for example, **dli:queue:submit_job**.
 - *Service name*: product name, such as **dli**, **evs**, and **vpc**. Only lowercase letters are allowed. Resource types and operations are not case-sensitive. You can use an asterisk (*) to represent all operations.
 - *Resource type*: For details, see [Table 13-7](#).
 - *Action*: action registered in IAM.
- Condition

Conditions determine when a policy takes effect. A condition consists of a condition key and operator.

A condition key is a key in the **Condition** element of a statement. There are global and service-level condition keys.

- Global condition keys (prefixed with **g:**) apply to all actions. For details, see condition key description in Policy Syntax.
- Service-level condition keys apply only to operations of the specific service.

An operator is used together with a condition key to form a complete condition statement. For details, see [Table 13-4](#).

IAM provides a set of DLI predefined condition keys. The following table lists the predefined condition keys of DLI.

Table 13-4 DLI request conditions

Condition Key	Type	Operator	Description
g:CurrentTime	Global	Date and time	Time when an authentication request is received NOTE The time is expressed in the format defined by ISO 8601 , for example, 2012-11-11T23:59:59Z .
g:MFAPresent	Global	Boolean	Whether multi-factor authentication is used during user login
g:UserId	Global	String	ID of the current login user
g:UserName	Global	String	Current login user
g:ProjectName	Global	String	Project that you have logged in to
g:DomainName	Global	String	Domain that you have logged in to
g:ResourceTag	Global	StringEquals	Resource tag value.

- Resource

The format is *Service name:Region:Domain ID:Resource type:Resource path*. The wildcard (*) indicates all options. For details about the resource types and path, see [Table 13-7](#).

Example:

dli:*:*:queue:* indicates all queues.

Creating a Custom Policy

You can set actions and resources of different levels based on scenarios.

1. Define an action.

The format is *Service name:Resource type:Action*. The wildcard is *. Example:

Table 13-5 Action

Action	Description
dli:queue:submit_job	Submission operations on a DLI queue
dli:queue:*	All operations on a DLI queue
dli:*:*	All operations on all DLI resource types

For more information about the relationship between operations and system permissions, see [Common Operations Supported by DLI System Policy](#).

2. Define a resource.

The format is *Service name:Region:Domain ID:Resource type:Resource path*. The wildcard (*) indicates all resources. The five fields can be flexibly set. Different levels of permission control can be set for resource paths based on scenario requirements. If you need to set all resources of the service, you do not need to specify this field. For details about the definition of Resource, see [Table 13-6](#). For details about the resource types and paths in Resource, see [Table 13-7](#).

Table 13-6 Resource

Resource	Description
DLI:*:*:table:databases.dbname.tables.*	DLI, any region, any account ID, all table resources of database dbname
DLI:*:*:database:databases.dbname	DLI, any region, any account ID, resource of database dbname
DLI:*:*:queue:queues.*	DLI, any region, any account ID, any queue resource

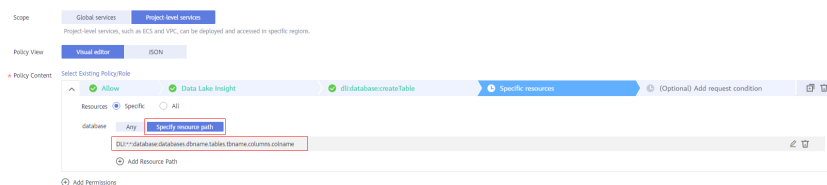
Resource	Description
DLI:*:*:jobs:jobs.flink.1	DLI, any region, any account ID, Flink job whose ID is 1

Table 13-7 DLI resources and their paths

Resource Type	Resource Names	Path
queue	DLI queue	queues.queueName
database	DLI database	databases.dbName
table	DLI table	databases.dbName.tables.tbName
column	DLI column	databases.dbName.tables.tbName.columns.columnName
jobs	DLI Flink job	jobs.flink.jobId
resource	DLI package	resources.resourceName
group	DLI package group	groups.groupName
datasource auth	DLI cross-source authentication information	datasourceauth.name
edsconnections	Enhanced datasource connection	edsconnections.connection ID

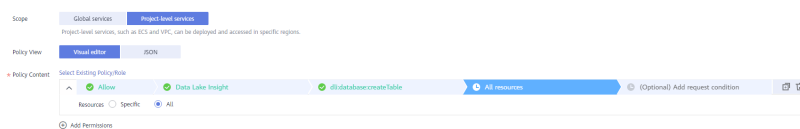
- Specific resources:

Figure 13-4 Specific resources



- All resources: all resources of the service

Figure 13-5 All resources



- Combine all the preceding fields into a JSON file to form a complete policy. You can set multiple actions and resources. You can also create a policy on the visualized page provided by IAM. For example:

Create a policy that grants users the permission to create and delete databases, submit jobs for any queue, and delete tables under any account ID in any region of DLI.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:createDatabase",
        "dli:database:dropDatabase",
        "dli:queue:submitJob",
        "dli:table:dropTable"
      ],
      "Resource": [
        "dli:*:database:*",
        "dli:*:queue:*",
        "dli:*:table:*"
      ]
    }
  ]
}
```

Example Custom Policies

- Example 1: Allow policies

- Allow users to create tables across all databases in all regions:

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:createTable"
      ],
      "Resource": [
        "dli:*:database:*"
      ]
    }
  ]
}
```

- Allow users to query column **col** in the table **tb** of the database **db** in the region where the user is located:

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:column:select"
      ],
      "Resource": [
        "dli:xx:*:column:databases.db.tables.tb.columns.col"
      ]
    }
  ]
}
```

- Example 2: Deny policies

A deny policy must be used together with other policies. That is, a user can set a deny policy only after being assigned some operation permissions. Otherwise, the deny policy does not take effect.

If the permissions assigned to a user contain both Allow and Deny actions, the Deny actions take precedence over the Allow actions.

- Deny users to create or delete databases, submit jobs (except the default queue), or delete tables.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "dli:database:createDatabase",
        "dli:database:dropDatabase",
        "dli:queue:submitJob",
        "dli:table:dropTable"
      ],
      "Resource": [
        "dli:*:database:*",
        "dli:*:queue:*",
        "dli:*:table:*"
      ]
    }
  ]
}
```

- Deny users to submit jobs in the demo queue.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "dli:queue:submitJob"
      ],
      "Resource": [
        "dli:*:queue:queues.demo"
      ]
    }
  ]
}
```

- Example 3: Tag authentication. You need to specify an action and bind it to a condition, and specify the key and value of **g:ResourceTag**.

Condition g: ResourceTag indicates a resource with the *key=value* tag. Only the operation on this resource contained in the policy action list is allowed.

The key is case insensitive. Fuzzy match for the value is not supported.

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:dropDatabase",
        "dli:table:select",
        "dli:database:createTable",
        "dli:table:dropTable"
      ],
      "Condition": {
        "StringEquals": {
          "g:ResourceTag/key": [
            "value"
          ]
        }
      }
    }
  ]
}
```

```
}  
  }  
] }  
}
```

13.4 DLI Resources

A resource is an object that exists within a service. You can select DLI resources by specifying their paths.

Table 13-8 DLI resources and their paths

Resource Type	Resource Names	Path
queue	DLI queue	queues.queueName
database	DLI database	databases.dbName
table	DLI table	databases.dbName.tables.tbName
column	DLI column	databases.dbName.tables.tbName.columns.colName
jobs	DLI Flink job	jobs.flink.jobId
resource	DLI package	resources.resourceName
group	DLI package group	groups.groupName
datasourceauth	DLI cross-source authentication information	datasourceauth.name
edsconnections	Enhanced datasource connection	edsconnections. <i>connection ID</i>

13.5 DLI Request Conditions

Request conditions are useful in determining when a custom policy takes effect. A request condition consists of a condition key and operator. Condition keys are either global or service-level and are used in the Condition element of a policy statement. Global condition keys (starting with **g:**) are available for operations of all services, while service-level condition keys (starting with a service name such as **dli**) are available only for operations of a specific service. An operator is used together with a condition key to form a complete condition statement.

IAM provides a set of DLI predefined condition keys. The following table lists the predefined condition keys of *DLI*.

Table 13-9 DLI request conditions

Condition Key	Type	Operator	Description
g:CurrentTime	Global	Date and time	Time when an authentication request is received NOTE The time is expressed in the format defined by ISO 8601 , for example, 2012-11-11T23:59:59Z .
g:MFAPresent	Global	Boolean	Whether multi-factor authentication is used during user login
g:UserId	Global	String	ID of the current login user
g:UserName	Global	String	Current login user
g:ProjectName	Global	String	Project that you have logged in to
g:DomainName	Global	String	Domain that you have logged in to
g:ResourceTag	Global	StringEquals	Resource tag value.

13.6 Common Operations Supported by DLI System Policy

Table 13-10 lists the common operations supported by each system policy of DLI. Choose proper system policies according to this table. For details about the SQL statement permission matrix in DLI in terms of permissions on databases, tables, and roles, see **SQL Syntax of Batch Jobs > Data Permissions Management > Data Permissions List** in the *Data Lake Insight SQL Syntax Reference*.

Table 13-10 Common operations supported by each system permission

Resource	Operation	Description	DLI FullAccess	DLI ReadOnlyAccess	Tenant Administrator	DLI Service Administrator
Queue	DROP_QUEUE	Deleting a Queue	√	×	√	√
	SUBMIT_JOB	Submitting a job	√	×	√	√

Resource	Operation	Description	DLI FullAccess	DLI ReadOnlyAccess	Tenant Administrator	DLI Service Administrator
	CANCEL_JOB	Terminating a Job	√	×	√	√
	RESTART	Restarting a queue	√	×	√	√
	GRANT_PRIVILEGE	Granting permissions to a queue	√	×	√	√
	REVOKE_PRIVILEGE	Revoking permissions to a queue	√	×	√	√
	SHOW_PRIVILEGES	Viewing the queue permissions of other users	√	×	√	√
Database	DROP_DATABASE	Deleting a database	√	×	√	√
	CREATE_TABLE	Creating a table	√	×	√	√
	CREATE_VIEW	Creating a view	√	×	√	√
	EXPLAIN	Explaining the SQL statement as an execution plan	√	×	√	√
	CREATE_ROLE	Creating a role	√	×	√	√
	DROP_ROLE	Deleting a role	√	×	√	√
	SHOW_ROLES	Displaying a role	√	×	√	√
	GRANT_ROLE	Binding a role	√	×	√	√
	REVOKE_ROLE	Unbinding a role	√	×	√	√

Resource	Operation	Description	DLI FullAccess	DLI ReadOnlyAccess	Tenant Administrator	DLI Service Administrator
	SHOW_USERS	Displaying the binding relationships between all roles and users	√	×	√	√
	GRANT_PRIVILEGE	Granting permissions to the database	√	×	√	√
	REVOKE_PRIVILEGE	Revoking permissions to the database	√	×	√	√
	SHOW_PRIVILEGES	Viewing database permissions of other users	√	×	√	√
	DISPLAY_ALL_TABLES	Displaying tables in a database	√	√	√	√
	DISPLAY_DATABASE	Displaying databases	√	√	√	√
	CREATE_FUNCTION	Creating a function	√	×	√	√
	DROP_FUNCTION	Deleting a function	√	×	√	√
	SHOW_FUNCTIONS	Displaying all functions	√	×	√	√
	DESCRIBE_FUNCTION	Displaying function details	√	×	√	√
Table	DROP_TABLE	Deleting tables	√	×	√	√
	SELECT	Querying tables	√	×	√	√
	INSERT INTO TABLE	Inserting table data	√	×	√	√

Resource	Operation	Description	DLI FullAccess	DLI ReadOnlyAccess	Tenant Administrator	DLI Service Administrator
	ALTER_TABLE_ADD_COLUMNS	Adding a column	√	×	√	√
	INSERT_OVERWRITE_TABLE	Overwriting a table	√	×	√	√
	ALTER_TABLE_RENAME	Renaming a table	√	×	√	√
	ALTER_TABLE_ADD_PARTITION	Adding partitions to the partition table	√	×	√	√
	ALTER_TABLE_RENAME_PARTITION	Renaming a table partition	√	×	√	√
	ALTER_TABLE_DROP_PARTITION	Deleting partitions from a partition table	√	×	√	√
	SHOW_PARTITIONS	Displaying all partitions	√	×	√	√
	ALTER_TABLE_RECOVER_PARTITION	Restoring table partitions	√	×	√	√
	ALTER_TABLE_SET_LOCATION	Setting the partition path	√	×	√	√
	GRANT_PRIVILEGE	Granting permissions to the table	√	×	√	√
	REVOKE_PRIVILEGE	Revoking permissions to the table	√	×	√	√
	SHOW_PRIVILEGES	Viewing table permissions of other users	√	×	√	√

Resource	Operation	Description	DLI FullAccess	DLI ReadOnlyAccess	Tenant Administrator	DLI Service Administrator
	DISPLAY_TABLE	Displaying a table	√	√	√	√
	DESCRIBE_TABLE	Displaying table information	√	×	√	√
Enhanced datasource connection	BIND_QUEUE	Binding an enhanced datasource connection to a queue It is only used to grant permissions across projects.	×	×	×	×

14 Other Common Operations

14.1 Importing Data to a DLI Table

Importing Data Using OBS

On the DLI management console, you can import data stored in OBS into DLI tables.

To import OBS data to a DLI table, either choose **Data Management > Databases and Tables** in the navigation pane on the left, locate a desired database, and click **Tables** in the **Operation** column, or choose **SQL Editor** in the navigation pane on the left.

For details, see [Importing Data to the Table](#).

14.2 Viewing Monitoring Metrics

Description

This section describes metrics reported by DLI to Cloud Eye as well as their namespaces and dimensions. You can use the management console or APIs provided by Cloud Eye to query the metrics of the monitored object and alarms generated for DLI.

Namespace

SYS.DLI

Metric

Table 14-1 DLI metrics

Metric ID	Name	Description	Value Range	Monitored Object	Monitoring Period (Raw Data)
queue_cu_num	Queue CU Usage	Displays the number of CUs applied by the user queue	≥ 0	Queues	5 minutes
queue_job_launching_num	Number of Jobs Being Submitted	Displays the number of jobs in the Submitting state in the user queue.	≥ 0	Queues	5 minutes
queue_job_running_num	Number of Running Jobs	Displays the number of running jobs in the user queue.	≥ 0	Queues	5 minutes
queue_job_succeed_num	Number of Finished Jobs	Displays the number of completed jobs in the user queue.	≥ 0	Queues	5 minutes
queue_job_failed_num	Failed Jobs	Displays the number of failed jobs in the user queue.	≥ 0	Queues	5 minutes

Metric ID	Name	Description	Value Range	Monitored Object	Monitoring Period (Raw Data)
queue_job_cancelled_num	Number of Canceled Jobs	Displays the number of canceled jobs in the user queue.	≥ 0	Queues	5 minutes
queue_alloc_cu_num	Allocated CUs (queue)	Displays the CU allocation for user queues.	≥ 0	Queues	5 minutes
queue_min_cu_num	Minimum CUs for Queue	Displays the minimum number of CUs for a user queue.	≥ 0	Queues	5 minutes
queue_max_cu_num	Maximum CUs for Queue	Displays the maximum number of CUs for a user queue.	≥ 0	Queues	5 minutes
queue_priority	Queue Priority	Displays the priority of a user queue.	1-100	Queues	5 minutes
queue_cpu_usage	Queue CPU Usage	Displays the CPU usage of user queues.	0-100	Queues	5 minutes
queue_disk_usage	Queue Disk Usage	Displays the disk usage of user queues.	0-100	Queues	5 minutes

Metric ID	Name	Description	Value Range	Monitored Object	Monitoring Period (Raw Data)
queue_disk_used	Max Disk Usage	Displays the maximum disk usage of user queues.	0-100	Queues	5 minutes
queue_mem_usage	Queue Memory Usage	Displays the memory usage of user queues.	0-100	Queues	5 minutes
queue_mem_used	Used Memory	Displays the memory usage rate of the user queues.	≥ 0	Queues	5 minutes
flink_read_records_per_second	Flink Job Data Read Rate	Displays the data input rate of a Flink job for monitoring and debugging.	≥ 0	Flink jobs	10 seconds
flink_write_records_per_second	Flink Job Data Write Rate	Displays the data output rate of a Flink job for monitoring and debugging.	≥ 0	Flink jobs	10 seconds
flink_read_records_total	Flink Job Total Data Read	Displays the total number of data inputs of a Flink job for monitoring and debugging.	≥ 0	Flink jobs	10 seconds

Metric ID	Name	Description	Value Range	Monitored Object	Monitoring Period (Raw Data)
flink_write_records_total	Flink Job Total Data Write	Displays the total number of output data records of a Flink job for monitoring and debugging.	≥ 0	Flink jobs	10 seconds
flink_read_bytes_per_second	Flink Job Byte Read Rate	Displays the number of input bytes per second of a Flink job.	≥ 0	Flink jobs	10 seconds
flink_write_bytes_per_second	Flink Job Byte Write Rate	Displays the number of output bytes per second of a Flink job.	≥ 0	Flink jobs	10 seconds
flink_read_bytes_total	Flink Job Total Read Byte	Displays the total number of input bytes of a Flink job.	≥ 0	Flink jobs	10 seconds
flink_write_bytes_total	Flink Job Total Write Byte	Displays the total number of output bytes of a Flink job.	≥ 0	Flink jobs	10 seconds
flink_cpu_usage	Flink Job CPU Usage	Displays the CPU usage of Flink jobs.	0-100	Flink jobs	10 seconds

Metric ID	Name	Description	Value Range	Monitored Object	Monitoring Period (Raw Data)
flink_mem_usage	Flink Job Memory Usage	Displays the memory usage of Flink jobs.	0-100	Flink jobs	10 seconds
flink_max_op_latency	Flink Job Max Operator Latency	Displays the maximum operator delay of a Flink job. The unit is ms .	≥ 0	Flink jobs	10 seconds
flink_max_op_backpressure_level	Flink Job Maximum Operator Backpressure	Displays the maximum operator backpressure value of a Flink job. A larger value indicates severer backpressure. 0 : OK 50 : low 100 : high	0-100	Flink jobs	10 seconds
elastic_resource_pool_cpu_usage	CPU Usage of Elastic Resource Pool	Displays the CPU usage of elastic resource pools.	0-100	Elastic resource pools	5 minutes
elastic_resource_pool_mem_usage	Memory Usage of Elastic Resource Pool	Displays the memory usage of elastic resource pools.	0-100	Elastic resource pools	5 minutes

Metric ID	Name	Description	Value Range	Monitored Object	Monitoring Period (Raw Data)
elastic_resource_pool_disk_usage	Disk Usage of Elastic Resource Pool	Displays the disk usage of elastic resource pools.	0-100	Elastic resource pools	5 minutes
elastic_resource_pool_disk_max_usage	Maximum Disk Usage of Elastic Resource Pool	Displays the maximum disk usage of elastic resource pools.	0-100	Elastic resource pools	5 minutes
elastic_resource_pool_cu_num	CU Usage of Elastic Resource Pool	Displays the CU usage of elastic resource pools.	≥ 0	Elastic resource pools	5 minutes
elastic_resource_pool_alloc_cu_num	Allocated CUs of Elastic Resource Pool	Displays the CU allocation of elastic resource pools.	≥ 0	Elastic resource pools	5 minutes
elastic_resource_pool_min_cu_num	Minimum CUs of Elastic Resource Pool	Displays the minimum number of CUs of elastic resource pools.	≥ 0	Elastic resource pools	5 minutes
elastic_resource_pool_max_cu_num	Maximum CUs of Elastic Resource Pool	Displays the maximum number of CUs of elastic resource pools.	≥ 0	Elastic resource pools	5 minutes

Dimension

Table 14-2 Dimension

Key	Value
queue_id	Queue
flink_job_id	Flink job

Viewing DLI Monitoring Metrics on Cloud Eye

1. Search for Cloud Eye on the management console.
2. In the navigation pane on the left of the Cloud Eye console, click **Cloud Service Monitoring > Data Lake Insight**.
3. Select a queue to view its information.

14.3 DLI Operations That Can Be Recorded by CTS

With CTS, you can record operations associated with DLI for later query, audit, and backtrack operations.

Table 14-3 DLI operations that can be recorded by CTS

Operation	Resource Type	Trace Name
Creating a database	database	createDatabase
Deleting a database	database	deleteDatabase
Changing the database owner	database	alterDatabaseOwner
Creating a table	table	createTable
Deleting tables	table	deleteTable
Exporting table data	table	exportData
Importing table data	table	importData
Modifying the owner of a table	table	alterTableOwner
Creating a queue	queue	createQueue
Deleting a queue	queue	dropQueue
Granting permissions to a queue	queue	shareQueue
Modifying the CIDR block of a queue	queue	replaceQueue

Operation	Resource Type	Trace Name
Restarting a queue	queue	queueActions
Scaling out/in a queue	queue	queueActions
Submitting a job (SQL)	queue	submitJob
Canceling a job (SQL)	queue	cancelJob
Granting DLI the permission to access OBS buckets	obs	obsAuthorize
Checking the SQL syntax	job	checkSQL
Creating a job	job	createJob
Updating a job	job	updateJob
Deleting a job	job	deleteJob
Creating a Flink OpenSource SQL job	jobs	createStreamSqlJob
Updating a Flink OpenSource SQL job	jobs	updateStreamSqlJob
Deleting a Flink job	jobs	deleteStreamJobs
Stopping a Flink job	jobs	stopStreamJobs
Submitting a Flink job	jobs	submitStreamJobs
Creating a Flink Jar job	jobs	createStreamJarJob
Updating a Flink Jar job	jobs	updateStreamJarJob
Checking Flink jobs	jobs	checkStreamJob
Importing a savepoint	jobs	dealSavepoint
Purchasing CUH packages	order	orderPackage
Freezing resources	resource	freezeResource
Unfreezing resources	resource	unfreezeResource
Terminating resources	resource	deleteResource
Clearing resources	resource	cleanResource
Granting data permissions	data	dataAuthorize
Granting permissions on other projects	data	authorizeProjectData
Exporting query results	data	storeJobResult
Saving a SQL template	sqlTemplate	saveSQLTemplate

Operation	Resource Type	Trace Name
Updating a SQL template	sqlTemplate	updateSQLTemplate
Deleting a SQL template	sqlTemplate	deleteSQLTemplate
Creating a Flink template	template	createStreamTemplate
Updating a Flink template	template	updateStreamTemplate
Checking Flink templates	template	checkStreamTemplate
Deleting a Flink template	flinkTemplate	deleteStreamTemplate
Creating a data upload task	uploader	createUploadJob
Obtaining the authentication to perform a data upload task	uploader	getUploadAuthInfo
Submitting a data upload task	uploader	commitUploadJob
Creating a datasource authentication and uploading a certificate	authInfo	uploadAuthInfo
Updating a datasource authentication	authInfo	updateAuthInfo
Deleting a datasource authentication	authInfo	deleteAuthInfo
Updating the quota	quota	updateQuota
Uploading a resource package	pkgResource	uploadResources
Deleting a resource package	pkgResource	deleteResource
Creating a basic datasource connection	datasource	createDatasourceConn
Deleting a basic datasource connection	datasource	deleteDatasourceConn
Reactivating a basic datasource connection	datasource	reactivateDSConnection
Creating an enhanced datasource connection	datasource	createConnection
Deleting an enhanced datasource connection	datasource	getConnection
Binding a queue	datasource	associateQueueToDataSourceConn

Operation	Resource Type	Trace Name
Unbinding a queue	datasource	disassociateQueueToDataSourceConn
Modifying the host information	datasource	updateHostInfo
Adding a route	datasource	addRoute
Deleting a route	datasource	deleteRoute
Creating a topic	smn	createTopic
Creating an agency	agency	createAgencyV2
Creating a batch processing job	batch	createBatch
Canceling a batch processing job	batch	cancelBatch
Creating a session	session	createSession
Deleting a session	session	deleteSession
Creating a statement	statement	createStatement
Canceling execution of a statement	statement	cancelStatement
Creating a global variable	globalVar	createGlobalVariable
Deleting a global variable	globalVar	deleteGlobalVariable
Modifying a global variable	globalVar	updateGlobalVariable

For details about how to enable CTS and view trace details, see the [Cloud Trace Service Getting Started](#).

For details about key fields in the CTS trace structure, see [Trace StructureTrace Structure](#) in the *Cloud Trace Service User Guide*.

14.4 Quota Management

What Is a Quota?

A quota limits the quantity of a resource available to users, thereby preventing spikes in the usage of the resource.

You can also request for an increased quota if your existing quota cannot meet your service requirements.

How Do I View My Quotas?


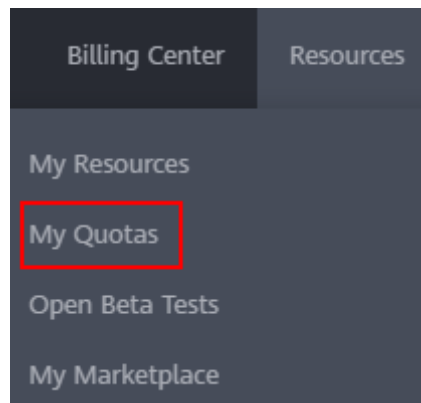
1. Log in to the management console.
2. Click  in the upper left corner and select a region and a project.
3. In the upper right corner of the page, choose **Resources > My Quotas**.
The **Service Quota** page is displayed.

Figure 14-1 My quotas

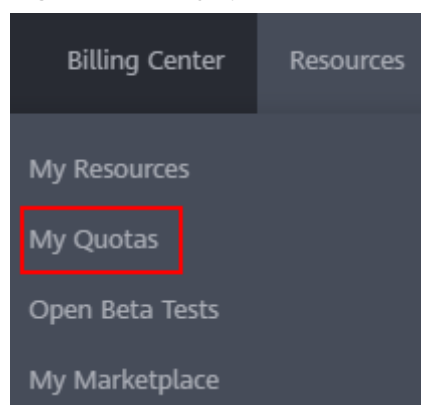


4. View the used and total quota of each type of resources on the displayed page.
If a quota cannot meet service requirements, increase a quota.

How Do I Apply for a Higher Quota?

1. Log in to the management console.
2. In the upper right corner of the page, choose **Resources > My Quotas**.
The **Service Quota** page is displayed.

Figure 14-2 My quotas



3. Click **Increase Quota**.
4. On the **Create Service Ticket** page, configure parameters as required.
In the **Problem Description** area, fill in the content and reason for adjustment.
5. Select the agreement and click **Submit**.

15 Permission Management for Global Variables

Scenario

- You can grant permissions on a global variable to users.
- The administrator and the global variable owner have all permissions. You do not need to set permissions for them, and their global variable permissions cannot be modified by other users.
- When setting global variables for a new user, ensure that the region of the user group the user belongs to has the **Tenant Guest** permission. For details about the Tenant Guest permission and how to apply for the permission, see [Creating a User Group and Assigning Permissions](#) and [System Permissions](#).

Granting Permissions on a Global Variable to a User

1. Choose **Global Configuration > Global Variables**, locate the row containing the desired global variable, and click **Set Permission** in the **Operation** column. On the **User Permissions** page displayed, you can grant permissions for the global variable, set and revoke user permissions.
2. Click **Grant Permission** in the upper right corner of the page.

Figure 15-1 Granting permissions of a global variable to a user

Grant Permission

★ Username

Select the permissions to be granted to the user

<input type="checkbox"/> Select all	<input type="checkbox"/> Delete	<input type="checkbox"/> Grant Permission
<input type="checkbox"/> Update	<input type="checkbox"/> View Other User's Permissions	
<input type="checkbox"/> Revoke Permission		

Table 15-1 Global variable parameters

Parameter	Description
Username	Name of the IAM user to whom permissions on the datasource connection are to be granted. NOTE The username is the name of an existing IAM user.
Permissions to be granted to the user	<ul style="list-style-type: none">• Update: This permission allows you to update the global variable.• Delete: This permission allows you to delete the global variable.• Grant Permission: This permission allows you to grant permissions of the global variable to other users.• Revoke Permission: This permission allows you to revoke the global variable permissions that other users have but cannot revoke the global variable owner's permissions.• View Other User's Permissions: This permission allows you to view the global variable permissions of other users.

Granting Global Variable Permissions

Click **Set Permission** in the **Operation** column of the sub-user to modify their permissions. [Table 15-1](#) lists the permission parameters.

If all permission options are grayed out, you are not allowed to change permissions on this global variable. You can apply to the administrator, group owner, or other authorized users for required permissions on the global variable.

Revoking Global Variable Permissions

Click **Revoke Permission** in the **Operation** column of a sub-user to revoke their permissions. After this operation, the sub-user does not have any permission on the global variable.