

Data Lake Compute Operation Guide Product Documentation





Copyright Notice

©2013-2025 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice

STencent Cloud

All trademarks associated with Tencent Cloud and its services are owned by the Tencent corporate group, including its parent, subsidiaries and affiliated companies, as the case may be. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Operation Guide

Console Operation Introduction

Data Development and Exploration

Data Exploration

SQL Editor

Data Query Task

SELECT Task

Querying Partition Table

Querying JSON Data

Querying Data from Other Sources

Using View

INSERT INTO

Querying Script Parameters

Obtaining Task Results

Query Script Analysis

Data Job

Overview

Configuring Data Access Policy

Creating Data Job

Managing Data Job

PySpark Dependency Package Management

Resource Management

Engine Management

Data Engine Introduction

SuperSQL Engine

SuperSQL Engine Overview

Purchasing Private Data Engine

Renewing SuperSQL Engine

Managing Private Data Engine

Engine-Level Parameter Settings

Disaster Recovery Cluster

Engine Kernel Version

Engine Network Configuration

Associating Tag with Private Engine Resource

Engine Local Cache

<u>S</u> Tencent Cloud

Custom Task Scheduling Pool Standard Engine Introduction of the Standard Engine System Standard Engine Introduction Standard Engine Kernel Versions Standard Engine Parameter Configuration **Engine Network Introduction** Gateway Introduction Standard Engine Startup and Stop Logs **Resource Group Resource Group Introduction Private Connection Private Connection Introduction Network Connection Configuration** Storage Configuration Managed Storage Configuration Binding a Metadata Acceleration Bucket Metadata Management Data Catalogs and DMC Data Table Management **Data View Management Function Management Partition Field Policy Ops Management Historical Task Instances** Historical task(Old version) Session Management **Insight Management** Task Insights System Management User and Permission Management **CAM Service** Permission Overview User and Work Group Sub-Account Permission Management Monitoring and Alarms Data Engine Monitoring Data Job Monitoring



Access Point Gateway Engine Monitoring Monitoring Alarm Configuration Audit Log

Operation Guide Console Operation Introduction Data Development and Exploration Data Exploration SQL Editor

Last updated : 2024-07-17 17:36:45

The SQL editor provided by Data Lake Compute (DLC) supports data querying using unified SQL statements, compatible with SparkSQL. You can complete data query tasks using standard SQL. You can access the SQL editor through data exploration, where you can perform simple data management, multi-session data queries, query record management, and download record management.

Data Management

Data management supports adding data sources, managing databases, and managing data tables.

Creating a data catalog

Currently, Data Lake Compute supports the management of COS and EMR Hive data catalogs. The directions are as follows:

1. Log in to the Data Lake Compute console and select the service region. You need to have the admin permission.

2. Select Data Explore on the left sidebar, hover over

on the **Database & table** tab, and click **Create catalog**.

Data Explore	hou *	SQL syntax reference 🖉 Data explo
Database Query	Ø + Query-2023-12-11 1 ● Query-2023-12-11 ● + ▼	🗘 Storage
Catalog DataLakeCatalog	Create a database Complet V 🖾 Save 🔾 Refresh 🛱 Format 蜿	🛢 demo 🔻 🔛 public-engine(SuperSQL-P 1.0-public
	Create catalog	
Select a target database		

For detailed directions, see Querying Data from Other Sources.

Managing a database

You can create, delete, and view the details of a database in the SQL editor.

Managing a data table

...

You can create, query, and view the details of a data table in the SQL editor.

Changing the default database

You can use the SQL editor to specify the default database for query tasks. If no database is specified in a query statement, the query will be executed in the default database.

- 1. Log in to the Data Lake Compute console and select the service region.
- 2. Select Data Explore on the left sidebar, hover over the target database name, click

, and click Set as default database to set the database as the default database.



3. You can also change the default database in the **Default database** selection box.

Query-2023-12-11 • + •	Default database	🌣 Storage cor
⊙ Running □ Save C Refresh □ Format	Select a default database '	Select a data engine
1	Q	
	iii i	
	10	

Data Query

Add Query Page

The SQL editor supports adding multiple pages for data querying, with each query page having independent configurations (default database, computation engine used, query records, etc.). This facilitates users in running and managing multiple tasks.

You can create a new query page by clicking on the

icon, and switch the editor interface by clicking on the tab bar.

Data Explo	ore 🔇 Guangzhou 🔻		SQL syntax reference IZ Data explore guide IZ
Database	Query Ø +	Query-2023-12-11	C Storage configuration
Catalog Datal	LakeCatalog v	O Running 🖾 Save 🖏 Refresh 🛱 Format 💂	😂 Select a default database 🔻 🔛 Select a data engine 🁻 🛛 🚥

For your convenience, you can save frequently used query pages by clicking the **Save** button. You can also quickly open your saved pages by clicking the

con.	
Cuery-cues-ic-line T	🗢 Seiect a default database 🔻 📄 Seiect a data engine 🄻 📔

For saved query page information, you can click the **Refresh** button to update and synchronize the saved information, ensuring the accuracy of the query statement.

\mathfrak{P} Storage configuration		+ +	Query-2023-12-11 (Query-2023-12-11 1 🌒
🛢 demo 🍸 📙 public-engine(SuperSQL-P 1.0-public) * 👘 ***	🛱 Format 🔤	🕄 Refresh	iplet 🔻 🗎 Save	• Running Com

The editor supports running multiple different SQL statements simultaneously. Clicking the **Run** button will execute all SQL statements within the editor, simultaneously dividing them into multiple SQL tasks.

If you need to run a portion of the statement, select the required statement and click **Partial run**.

Data	ibase C	Query Ø +	q	ery-2023-12-11 • + · ·	Storage configuration
Catalog	DataLakeCat	talog 👻		Partial run 🖾 Save 😳 Refresh 🛱 Format 😡	😫 Select a default database 🔻 📑 Select a data engine 🎽 👘
Select	a target databa	ise 🔻		1 SELECT * FROM 'DataLakeCatalog'.'a11'.'a' LIMIT 10; 2	

Engine Parameter Configuration

After selecting the data engine, you can configure parameters for the data engine. After selecting the data engine, click **Add** in Advanced Settings to configure.

Query-2023-12-11 • + •	坟 Storage configuration
O Running Complet ▼ 🖾 Save 🕃 Refresh 🚔 Format 📾	Select a default database 🔻 🔚 public-engine(SuperSQL-P 1.0-public) 🔻
1	E Data engine
	public-engine 🔻
	② Engine (kernel version) Different kernel versions support different SQL syntax rules. For details, see Kernel Versions.
	presto (SuperSQL-P 1.0-public)
	⊕ Create engine
	Advanced settings A Configuration description [2]
	1 dlc.query.execution.mode v async v -
	+ Add
Query result	

The currently supported configuration parameters are as follows:

Engine	Configuration name	Start Value	Configuration Notes
SparkSQL	spark.sql.files.maxRecordsPerFile	0	The maximum number of records that can be written to a single file. If this value is zero or negative, there are no restrictions.
	spark.sql.autoBroadcastJoinThreshold	10MB	Configure the maximum byte size of the table of all working nodes displayed when executing a connection. By setting this value to "-1", the display can be disabled.
	spark.sql.shuffle.partitions	200	Default Partition Count.
	spark.sql.sources.partitionOverwriteMode	static	When the value is set to static, all qualifying partitions will be deleted prior to executing the overwrite operation. For instance, in a partitioned table, there is a partition "2022-01". When using the INSERT OVERWRITE statement to write data to the "2022- 02" partition, the data in the "2021- 01" partition will also be overwritten. When the value is set to 'dynamic', partitions will not be deleted in advance, but will be overwritten during runtime for those partitions where data is written.



	spark.sql.files.maxPartitionBytes	128MB	The maximum number of bytes to be packaged into a single partition when reading a file.
Presto	use_mark_distinct	true	Determines whether the engine redistributes data when executing the distinct function. If the distinct function is called multiple times in a query, it is recommended to set this parameter to false.
	USEHIVEFUNCTION	true	Determines whether to use Hive functions when executing a query; if you need to use Presto native functions, please set the parameter to false.
	query_max_execution_time	-	This setting is used to establish a query timeout. If the execution time of a query exceeds the set time, the query will be terminated. The units supported are d-day, h-hour, m-minute, s-second, ms-millisecond (for example, 1d represents 1 day, 3m represents 3 minutes).
	dlc.query.execution.mode	async	The engine query execution mode is set to async mode by default. In this mode, the task will perform a complete query calculation, save the results to COS, and then return them to the user, allowing the user to download the query results after the query is completed. Users can also change this value to sync. In sync mode, queries may not necessarily perform full calculations. Once partial results are available, they will be directly returned to the user by the engine, without being saved to COS. Therefore, users can achieve lower query latency and duration, but the results are only saved in the system for 30 seconds. This mode is recommended for

	users who do not need to download the complete query results from
	COS, but expect lower query latency
	and duration, such as during the
	query exploration phase or BI result
	display.

Presto Execution Mode

When the user selects the Presto engine, Data Exploration supports the user to choose to run in "Fast Mode" or "Full Mode".

Quick Query: This offers faster speed, but the query results cannot be persistently saved. It is suitable for the exploration phase.

Full Mode: Execute a full query and save the data to object storage.



Search results

Through the SQL editor, you can directly view the query results. You can expand or collapse the display height of the query results by clicking the



Query-2023-12-11 +						
Partial run Co	omplet 🔻 🛅 Save 🖸 Refresh 🛱 Format 📾		Select a default database T public-engine(SuperSQL-P 1.0-public)	• ···		
1 SELECT * FR	COM 'DataLakeCatalog','demo'.'test_1' LIMIT 10;					
Query result Stati	istics		Run history Download histor	y 🔼 🗸		
Task ID SQL details Ex Query time 2.30s Scar 10 entries in total (up to	xport Suggestions 22 nned data volume 34.K8 Billable scanned volume 34.0 MB ① o 1.000 entries shown in the console)Copy ①					
id	pro_name	price	pro_date			
	r		1000	A		
	and the second se					
5	c					
1.1						
	re			- 1		
-		17.3				
				*		

You can configure the query result storage directory through the configuration button in the upper right corner, supporting configuration to the COS path or built-in storage.

ase	Query 🗘 🕂	Query-2024 Draft X	+ •					🌣 Storage confi
DataLakeC	Catalog 🔹		O C ™ E	Aa v	Default database	▼ 🗄 S	Gelect a data engine 🔻	⊱ SuperSQL Syntax ▼
target datal	base 🔻	1 CREATE DATABAS	E IF NOT EXISTS "Data	LakeCatalog"."demo2"	COMMENT 'for demo_	_test'		
_								
_database	Storage configuration	n				×		
ito_databa	Managed storage 🛈	Enable						
ito_databa	Managed storage type	General bucket						
ercion atabase1	Query result storage path	(i) Internal storage	User-defined storage					
atabase10		The SELECT query res	Its are stored in the internal	J storage of Data Lake Comp for 36 hours	oute, and the underlying			
atabase100		storage service is 000	The results will be retained i	or 50 fiburs.				
tabase11								
tabase12	Save Cance	ðl						
tabase14								

The console will return a maximum of 1000 results for a single task. If more results are needed, the API can be used. For instructions on API-related operations, refer to the API Documentation.

Query results can be downloaded locally when no COS storage path is specified. For detailed instructions, refer to Obtaining Task Results.

Querying statistical data

The query results under the Presto engine and SparkSQL engine support the display of optimized quantification with different characteristics.

The SparkSQL engine supports viewing:

- 1. Data Scanning Volume
- 2. Cache Acceleration
- 3. Adaptive Shuffle
- 4. Materialized View Acceleration

The Presto engine supports viewing:

- 1. Data Scanning Volume
- 2. Cache Acceleration
- 3. Materialized View Acceleration

Click on the **Statistics** column to review the statistical data and optimization suggestions for the query results.

٢	Data Explore	🛇 Guangzhou 🔻					SQL syntax reference 🖾 🛛 Data explor
	Database Query	φ +	Query-2023-12-12 • + •				🗘 Storage
Q	Catalog DataLakeCatalog	Ŧ	● Running Complet ▼ 🖺 Save 👯	Refresh 🛱 Format 💀		Select a default database 🔻 Ε	public-engine(SuperSQL-P 1.0-public)
	Select a target database	Ŧ	1				
٩	• 🛢 .						
	•						
	•						
) 🛢 (
) 🛢 d						
	* 🛢 d		Query result Statistics				Run history Download histor
	🔻 🕅 Table		Scanned data volume	Cache for acceleration ③		Materialized views for acceleration $\textcircled{0}$	
			2.5	Result cache hit No		No materialized view found	
			2.5кв	Fragment cache hit rate		Create a materialized view to further speed up queries	
	 BS View E E E E E E E E E E E E E E E E E E E			Autoxio cache hit fate ••			
	Contraction		Preprocess: 64.00ms		Execute: 4.62s		Get result: 82.00ms

Historical Queries

Each query page can save the running history of the past three months and supports viewing the query results of the past 24 hours. You can quickly find past task information through the running history. For detailed operations, refer to Task History Records.

Download History Management

Each query result's download task can be viewed in the **Download history**, where you can check the status of the download task and related parameter information.

Query result	Statistics			Run history Download history
Task ID SQL de Query time 4.6	tails Export Suggestions 🗹 25 Scanned data volume 2.5 KB Billable scanned volu	me 34.0 MB 🚯	Download history Task ID Source path	🗘 🗙 Execution st Task submission time Scanned da
10 entries in tot	al (up to 1,000 entries shown in the console)CopyI	price		No download record

Data Query Task SELECT Task

Last updated : 2024-07-17 16:04:41

You can query, analyze, and compute the data in a created database or data table with SQL statements.

Running a SELECT query task

1. Select the default database and compute resource.

You can select a default database. Then, when there is no database specified in a SQL statement, the statement will be executed in the default database.

You can select a public or private cluster as the compute resource.

2. Write a standard SQL statement and click **Running**.

Databa	ase Query	¢ +	Query-2023-12-12 • + •
Catalog	DataLakeCatalog	•	Running Complet ▼ □ Save □ Refresh □ Format □
Select a	target database	•	1 SELECT * FROM `DataLakeCatalog`.` ``` LIMIT 10;
• 81	-		
) 🛢 d			
) 🛢 c			
•₿∢			
• 8	7		
• 🛢 🌡			

In Data Lake Compute, a task can run for up to 30 minutes.

Data Lake Compute is serverless, so compute resources will be scheduled temporarily. It may take longer than usual to return the result of the first DML task.

3. The query result will be displayed in the console after the task is completed.

If you exit the console page, you cannot view the query result of a historical task there again. In this case, you can view the task result file in **Run history** or the query result COS bucket you configured.

Canceling a running query task

During task running, the **Run** button becomes **Terminated**, which you can click to cancel the task. Then, Data Lake Compute will not return the query result but will calculate the scanned data volume. If you use the public engine, the scanned data volume will incur fees. For billing details, see <u>Billing Overview</u>.



Querying Partition Table

Last updated : 2025-03-07 15:27:25

Storing data in partition catalogs can greatly reduce the scanned data volume of a computing task in Data Lake Compute and thereby significantly enhance the computing performance. The general practice of data partitioning is to store data in different catalogs by time. For example, data generated on the same day can be stored in the same catalog, and catalogs can be organized in a "year-month-day" structure. In Data Lake Compute, a table and its partitions must adopt the same data format.

Creating a Partition Table

To create a partition table, you need to specify the partition field in the table creation statement.

Adding Partitioned Data

Specifying a partition during data table creation is only to configure the partition field and doesn't allow running a query statement immediately to get data. You need to add partitioned data to a data table. If new partitioned data is added to the data catalog, you also need to add the partition information to the data table.

Manually adding a partition

Use the ALTER TABLE ADD PARTITION statement to add a specified partition catalog to a data table. If the partition catalog is compatible with the Hive partitioning rule (**partition column name=partition column value**), you don't need to specify the data path; otherwise, you need to refer SQL Syntax. Sample 1: Adding a single partition catalog

```
ALTER TABLE tabel_demo ADD
PARTITION (dt = '2021-01-01');
```

Sample 2: Adding multi-level nested partition catalogs

```
ALTER TABLE tabel_demo ADD
PARTITION (year = '2021', month='01', day='01');
```

Sample 3: Displaying the specified partition path

```
ALTER TABLE tabel_demo ADD
PARTITION (year = '2021', month='01', day='01') LOCATION 'cosn://tablea_demo';
```

Automatically adding a partition

Use the MSCK REPAIR TABLE statement to scan the data catalog specified during table creation. If there is a new partition catalog, the system will automatically add the partitions to the metadata of the data table. Details can be found in the SQL Syntax.Below is a sample:

MSCK REPAIR TABLE table_demo

System Restraints

MSCK REPAIR TABLE only adds partitions to the metadata of the data table but does not delete them. To delete an added partition, run the ALTER TABLE table-name DROP PARTITION statement.Details can be found in the SQL Syntax.

MSCK REPAIR TABLE is not recommended if the data volume is large, as the system will scan all the data, which may take a long time, cause the task to time out, and make the partition information of the data table incomplete. A partition catalog must be compatible with the Hive partitioning rule of **partition column name=partition column value**; otherwise, use ALTER TABLE ADD PARTITION to load a partition.Details can be found in the SQL Syntax.

Make sure that data of a table is stored in a separate folder. For example, if the cosn://tablea_a data in table A
and the s3://table_a/table_b data in table B are stored in COS and both tables are partitioned by string,
then MSCK REPAIR TABLE will add partitions of table B to table A. To avoid this, use separate folder structures,
such as cosn://tablea_a and cosn://tablea_b.

The statement may incur data read/write fees charged by COS. For more information, see Billing Overview.

Querying JSON Data

Last updated : 2024-07-17 16:18:53

Query steps

1. Create a data table and specify the JSON format for parsing.

```
CREATE EXTERNAL TABLE `order_demo`(
 `docid` string COMMENT 'from deserializer',
 `user` struct < id :int,
 username :string,
 name :string,
 shippingaddress :struct < address1 :string,
 address2 :string,
 city :string,
 state :string > COMMENT 'from deserializer',
 `children` array < string >
) ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe' LOCATION
 'cosn://dlc-bucket/order'
```

2. Run a query statement to query the JSON data. Data Lake Compute supports json_parse(),

json_extract_scalar() , and json_extract() parsing functions.

SELECT `user`.`shippingaddress`.`address1` FROM `order_demo` limit 10;

System restraints

The data must be in complete JSON format; otherwise, Data Lake Compute cannot parse it.

A data row cannot contain a line break, and the JSON format cannot be optimized visually; for example:

```
{"name":"Michael"}
{"name":"Andy", "age":30}
{"name":"Justin", "age":19}
```

Data Lake Compute will automatically recognize the first JSON level as the attribute column of a data table and recognize other nested structures as corresponding attribute values.

Querying Data from Other Sources

Last updated : 2025-01-03 15:40:27

Data Lake Compute allows you to query and analyze data in an external table. Currently, data from MySQL and EMR Hive can be connected to it. You can add and manage other data sources in the Data Lake Compute console.

Adding a data source

1. Log in to the Data Lake Compute console and select the service region. You need to have the permission to create data catalogs.

2. Select Data Explore on the left sidebar, hover over +, and click Create data catalog.



3. Select the data source type. Currently, MySQL and EMR Hive are supported. Before configuring MySQL, you need to add the Data Lake Compute subnet to the database's allowlist. Two configuration methods are supported: database instance and JDBC connection.

Catalog configuration Connection type * MySQL Connection name * It can contain up to 25 characters in letters, digits, and underscores (). Description Up to 50 characters Instance • Please select Data source VPC * Select a VPC Select a subne Connection	Create catalog		:
Connection type • MySQL • Connection name • It can contain up to 25 characters in letters, digits, and underscores (). Description Up to 50 characters Instance • Please select • Data source VPC • Select a VPC • Select a VPC • Select a subne • Enter a username Parsword •	1 Catalog configuration	> 2 Network configuration	
Connection name * It can contain up to 25 characters in letters, digits, and underscores (). Description Up to 50 characters Instance • Please select • Data source VPC * Select a VPC • Select a VPC • Select a subne • Username • Enter a username	Connection type *	MySQL	
Description Up to 50 characters Instance * Please select * Data source VPC * Select a VPC Select a subne *	Connection name *	It can contain up to 25 characters in letters, digits, and underscores ().	
Instance Please select Data source VPC • Select a VPC Select a subne Username • Enter a username	Description	Up to 50 characters	
Instance Please select Data source VPC * Select a VPC Select a subne Username * Enter a username			
Data source VPC Select a VPC Select a subne O IPs in total, 0 available Username Enter a username	Instance v *	Please select 🔻	
Username * Enter a username	Data source VPC *	Select a VPC 💌 Select a subne 🔻 🗘 0 IPs in total, 0 available	
Password * Enter a nassword	Username *	Enter a username	
Enter a passifiera	Password *	Enter a password	

Supported EMR Hive versions are 2.0.1, 2.1.0, 2.2.0, 2.2.1, 2.3.0, 2.4.0, 2.5.0, 2.5.1, and 2.6.0. The configuration is performed through the EMR access address.

4. Enter the data source information and click **Create connection**.

Note :

A data engine must be bound to the network configuration of the VPC where the data source resides. You can view the bound data engine during creation or create a network configuration and bind the data engine. For more information about network configuration, see Engine Network Configuration.

Managing Data

Currently, Data Lake Compute allows you to **view the database information of** and **preview data in** external tables.

Viewing database information

1. Log in to the Data Lake Compute console and select the service region. You need to have the permission to view data tables.

2. Select **Data Explore** on the left sidebar, hover over +, and click **Basic info**. You can view the basic information of a data table in the pop-up window.

Data Explo	ore 🔇 Gua	ngzhou 🔻
Database	Query	φ+
Catalog v	۲	÷ ¢
Select a target	database	•
• 🛢 i		•••
		Ba
▶ 🛢		Ad
•		

Previewing data in a data table

1. Log in to the Data Lake Compute console and select the service region. You need to have the permission to view data tables.

2. Select **Data Explore** > **Data table**, hover over ..., and click **Preview data**. Then, you can run a SQL statement to query and display data in the data table.

Database	Query	¢ +	Query-20	D24 Draft	×	+ •	7				
Catalog		•		6	۵.	Ċ	С	لا لا	1 2 3	Aa 🔻	
Select a target	database	•	1	SELECT * F	ROM	'a0"."	t1" L	.IMIT	10		
v											
▼ III Table											
⊞t1											
► BOView		Basic i	nfo								
► 🕅 Funct	tion	Preview	w data								
•		Delete	table								

Note:

Select the data engine bound to the network configuration of the VPC of the data source.

Using View

Last updated : 2025-01-03 15:27:27

In Data Lake Compute, a view is a logical table rather than a physical table. Whenever a view is referenced during a query, the query that defines the view will be executed. You can create a view through SELECT and reference it in future queries. Details can be found in the SQL Syntax.

System restraints

A view name is case-insensitive and can contain up to 128 letters and underscores.

Data Lake Compute doesn't support managing data access permissions through views.

INSERT INTO

Last updated : 2024-07-17 16:23:11

```
The INSERT INTO statement can insert a SELECT query result in the source table to the target table as a new row.
```

Querying Script Parameters

Last updated : 2024-07-17 16:23:47

Data Lake Compute allows you to configure date parameters to facilitate queries with scripts.

Data Lake Compute adopts the standard date format of yyyymmddhh24miss and uses the \${} command to set a date as a variable consisting of the date and time.

Date: It can be in any date format or a predefined system variable, such as yyyymmdd , yyyymm , yyyy-mmdd , yy , and dataDate .

 Time: It can be +/-N cycles and supports
 N/Nd
 Nm
 Nw
 Nh
 , and
 Nmi
 . It is compatible with various

 calculation formulas, such as
 7*N
 and
 N/24
 .
 .
 .
 .

Examples

+/- N Cycle	Method	Compatible Format	Example
N years later	\${yyyymmdd+Ny}	-	-
N years ago	\${yyyymmdd-Ny}	-	One year ago: \${yyyymmdd- 12m}: 20190920
N months later	-	\${yyyymmdd+Nm}	-
N months ago	\${yyyymmdd-Nm}	\$[add_months(yyyymmdd,- N)]	\${yyyymmdd-1m}: 20200820 \${yyyymm}: 202009 \${dataDate-1m}: 20200820
N weeks later	\${yyyymmdd+Nw}	\${yyyymmdd+7*N}	-
N weeks ago	\${yyyymmdd-Nw}	\${yyyymmdd-7*N}	-
N days later	\${yyyymmdd+N/Nd}	-	-
N days ago	\${yyyymmdd-N/Nd}	-	\${yyyymmdd-1}, \${dataDate-1}
N hours later	\${yyyymmddhh24+Nh}	\$[yyyymmddhh24+N/24]	-
N hours ago	\${yyyymmddhh24-Nh}	\$[yyyymmddhh24-N/24]	\${yyyymmddhh24-1h}: 2020092014

			\${dataDate-1h}: 2020092014
N minutes later	\${yyyymmddhh24mi+Nmi}	\$[yyyymmddhh24+N/24/60]	-
N minutes ago	\${yyyymmddhh24mi-Nmi}	\$[yyyymmddhh24-N/24/60]	\${yyyymmddhh24mi-10mi}, \${dataDate-10mi}

Note:

Make sure that the variable or the part before +/- in the variable is in line with the standard date format; otherwise, the system cannot recognize and use it.

Obtaining Task Results

Last updated : 2024-09-18 17:59:35

Using the Query Editor to Obtain Task Results

When you use the DLC console for task queries, the query results will be displayed in real-time below the editor.

y-2024 Draft x + ▼	☆ Storage configuration
	Default database 🔻 🔚 test 🍷 🕟 Standard - Presto Syntax 🍷 🛛 ***
SELECT * FROM `DataLakeCatalog`.`test`.`new_table_name2` LIMI	T 10;
ry result	Run history Download history A V
Task ID Export Query time 8s Scanned data volume 102 B Result Display 200 ▼ ① ① Total: 1 ② Copy Data I□ No. column_name1 \$ 1 1 1	column_name2 \$
r	y result Task ID Export Query time 8s Scanned data volume 102 B Result Display 200 ▼ ③ Total: 1 ③ Copy Data T No. column_name1 ‡ 1 1

A single SQL task in the console can display up to 1,000 rows of data. SQL tasks submitted via API and JDBC are not subject to this limitation.

You can view the query history for a single Session for up to 3 months by checking the running history. For more methods to query historical records, see History.

Output Format Configuration for Task Results

The results of data exploration are saved in CSV format by calling Spark's DataFrame.write. If the engine version is released later than April 2023, you can configure the output format of the exploration results.

1. Configure the format of the results output to CSV. The following parameters are supported:

Parameter	Default Value	Remark
livy.sql.result.format.option.sep livy.sql.result.format.option.delimiter	3	The separator bet the result is storec

		comma by default
livy.sql.result.format.option.encoding livy.sql.result.format.option.charset	UTF-8	String encoding fc For example: UTF 8859-1, UTF-16B UTF-16.
livy.sql.result.format.option.quote	\\"	Specifies whether quotation marks, v of escape charact
livy.sql.result.format.option.escape	////	Escape character of escape charact
livy.sql.result.format.option.charToEscapeQuoteEscaping		The characters th within quotation m
livy.sql.result.format.option.comment	\\u0000	Remark informatic
livy.sql.result.format.option.header	false	Specifies whether
livy.sql.result.format.option.inferSchema	false	Infers the data typ not inferred, all co strings.
livy.sql.result.format.option.ignoreLeadingWhiteSpace	true	Ignores leading er
livy.sql.result.format.option.ignoreTrailingWhiteSpace	true	Ignores trailing en
livy.sql.result.format.option.columnNameOfCorruptRecord	_corrupt_record	The name for the converted. This pa by spark.sql.column with table configur precedence.
livy.sql.result.format.option.nullValue		Specifies the stora values. The defau which case it can emptyValue types
livy.sql.result.format.option.nanValue	NaN	The storage forma values.
livy.sql.result.format.option.positiveInf	Inf	The storage forma
livy.sql.result.format.option.negativeInf	-Inf	The storage forma

livy.sql.result.format.option.compression or codec		The class name o algorithm. By defa applied. Short nar gzip, Iz4, and sna
livy.sql.result.format.option.timeZone	System default time zone	The default time z spark.sql.session. example, Asia/Sh configuration take
livy.sql.result.format.option.locale	en-US	Specifies the lang
livy.sql.result.format.option.dateFormat	yyyy-MM-dd	The default forma
livy.sql.result.format.option.timestampFormat	yyyy-MM- dd'T'HH:mm:ss.SSSXXX	The default forma LEGACY mode, it yyyy-MM-dd'T'HH
livy.sql.result.format.option.livy.sql.result.format.option.multiLine	false	Allows multiple lin
livy.sql.result.format.option.maxColumns	20480	The maximum nur
livy.sql.result.format.option.maxCharsPerColumn	-1	The maximum nur column1 means
livy.sql.result.format.option.escapeQuotes	true	Escapes quotation
livy.sql.result.format.option.quoteAll	quoteAll	Encloses the entir marks when writir
livy.sql.result.format.option.emptyValue	//"//"	The format used f empty values.
livy.sql.result.format.option.lineSep		The newline chara separation.

2. Configure the output format to a non-CSV format. Note that in this case, the console will not be able to display the results. However, you can read the result path using other methods. For details on where the result path is saved, see the next section.

The configuration option livy.sql.result.format supports saving in formats such as text, ORC, JSON, and Parquet.

Task Result Storage Location Configuration

Note: The Standard Engine - Presto is not supported. Full results can be obtained via JDBC.



DLC supports automatically saving query results to a COS path or DLC's managed storage through configuration. The configuration steps are as follows:

1. Log in to the DLC console, select the service region, and ensure that the login account has necessary COS-related permissions.

2. Go to the **Data Exploration Page**, click **Storage Configuration** in the upper right corner, and configure the settings for saving query results.

		Storage configuration
•	🗄 test 🔻	∑ Standard - Presto Syntax ▼ ● • • •

3. You can save the results to DLC's managed storage or COS. If you want to configure the path to COS, the operating account should have necessary COS-related permissions. Data storage fees will be based on COS pricing. The task results are stored in subfolders under the following COS path:

```
Data path for task results: COS directory
path/DLCQueryResults/yyyy/mm/dd/[QueryID]/data/XXXX.csv
Metadata path for task results: COS directory
path/DLCQueryResults/yyyy/mm/dd/[QueryID]/meta/result.meta.json
```

COS directory path: This is the COS directory path configured in the system settings.

/yyyy/mm/dd: The directory is organized based on the task execution date.

/data: This directory stores the query result data, with files in CSV format. DLC may generate multiple data files.

/meta: This directory stores the metadata for the queried data tables, with files in JSON format.

Note:

Storing SELECT query results in DLC's internal storage, with Cloud Object Storage as the underlying storage, and the results are retained for 36 hours.

When SELECT query results are stored in your COS bucket path, ensure that you have necessary COS-related permissions.

Downloading Task Results

Note: The Standard Engine - Presto is not supported. Full results can be obtained via JDBC.

DLC allows users to manually download query results to their local devices. If full result mode is not enabled, users can download the results of tasks with available query results to their local devices or manually save them to COS



(COS permissions are required).

The data downloaded or saved to COS correspond to the query results of the current SQL task, with a maximum of 500 results.

The maximum size for the local download is 50 MB.

If the results are configured to be saved to COS, they will be automatically stored in the COS path without the need for manual downloads.

Query Script Analysis

Last updated : 2024-08-07 17:08:48

To facilitate users in quickly handling repetitive query tasks, DLC provides script file analysis.

Note

The console allows saving up to 100 SQL scripts.

Creating a New Query Directory

- 1. Log in to DLC Console > Script Query Page.
- 2. On the query page, click Add Query Directory.



3. After filling in the directory configuration, you can save and complete the creation.

Add query	catalog	×
Basic info		
Catalog name		
Permission settings	An admin I	has all permissions by default and is not subject to the settings here
Available to	Work group	Add permissions for existing users in the work group and those to join later
	User	 Add permissions for individual users
		Confirm Cancel

Directory name: Supports Chinese characters, letters, and underscores (_), up to 25 characters.

Permission settings: You can set the visibility permissions for the script directory and the scripts within it based on the perspective of the workgroup or user.

Creating a New Query Script

1. Log in to DLC Console > Script Query Page, You can click the library

icon or directly add execution and save.

2. After the computation engine is selected, click Run to execute the script.

63	Data Evaluata		() nutty reference (2). Data and
3		gano	aug synox reference 📴 belo expr
88	Database Query	φ + Query-2023-12-19 ● + ×	tật Storag
Q	Enter a query name	Q. O Running B SAL C Refresh	🗟 - Select a default database 🎽 📘 Select a data engin
≣	▼ 酉 默认	1	2 Select the guery on
35	🗅 त	Create query	2. Select the query en
Ē	D	Modify 1. Create a query	
	D.	Delete catalog	
•	BF***		

Saving a Query Script

1. After the query is completed, click the Save button.

2. Queries created through the library will be saved under the directory of that library. Queries added through the tab bar can be saved directly in the root directory or an authorized library.

Data Explore 🕲 Guangzhou *				
Database Query ϕ +	test_db_save x Query-2023-12-12 • + •	⊅ Storage o		
Enter a query name Q	O Running Complet ▼ I Save C Refresh I Format Image: Complet Comp	🛢 demo 🔻 🔚 public-engine(SuperSQL-P 1.0-public)		
) Č	1			
) <u> </u>				
▶ Ĝ:				
È				
▶ 🔁 F				

3. Query table permissions can be customized according to the public scope of the library, and table usage permissions can be specified for the public scope.

Query name	Query-2023-12-12 14:15:56
Query catalog	Root directory 💌
	If you change the catalog, authorizations will be updated accordingly.
Permission settings	An admin has all permissions by default and is not subject to the setting: here
Available to Wor	k 🔹
grou	Add permissions for existing users in the work group and those to join later
Use	r 🗸
	Add permissions for individual users
Permissions Sele	All Read Edit Delete

Viewing script information

1. Hover the mouse pointer over the script name to view the script details.



Data Explo	ore (🛇 Guangzhou 🔻					SQL syntax reference 🛽 Data explore
Database	Query	φ +	test_db_save	x Query-2023-12-12 1	Query-2023-12-12 ●	+ •	🌣 Storage o
Enter a query na	ame	Q	Running	Complet 🔻 🛅 Save	🕄 Refresh 🔹 Forma	t Sal	😑 demo 🔻 🔛 public-engine(SuperSQL-P 1.0-public)
· õ		Query name	-				
🗋 ə' 👘	:890	Query catalog	00				
D I	^^5783	Update time	18:04:03				
D - 11	111198	68890					

2. Click the

...

icon next to the table you want to view, and select to open or query it.

Data Explore 🔇 Guan	gzhou v	SQL syntax reference 🖄 Data explore guide 🖾
Database Query		✿ Storage configuration
Enter a query name	Q, O Running Complet ▼ 🖾 Save 🗘 Refresh 🛱 Format 😡	😂 demo 🔻 🗄 public-engine(SuperSQL-P 1.0-public) 🔻 🛛 🚥
* 63WII.	1	
D		
	Open query	
0	Modify	
D	Generate schedule task	
	Delete	

Deleting a Query Script

Click the

icon next to the table you want to delete, and select to delete the script.

Data Explore	ngzhou *	SQL syntax reference [2] Data explore g
Database Query	Ø + Query-2023-12-12 1 ● Query-2023-12-12 1 ● Query-2023-12-12 ● + ▼	t¢: Storage co
Enter a query name	Q Complet ▼ B Save C Refresh B Format Sa	😑 demo 🍸 🔚 public-engine(SuperSQL-P 1.0-public) 🔻
- 6WI	1	
D		
D (Open query	
	Modify	
D	Generate schedule task	
D-1	Delete	

Note:

Deleted scripts cannot be restored. Operate with caution.
Data Job Overview

Last updated : 2024-07-17 16:36:54

Data Lake Compute provides Spark-based batch and flow computing capabilities for you to perform complex data processing and ETL operations through data jobs. Currently, data jobs support the following versions: Scala 2.12 Spark 3.1.2

Preparations

Before starting a data job, you need to create a data access policy to ensure data security as instructed in Configuring Data Access Policy.

Currently, only CKafka data source is supported for data job configuration, with more data sources to come in the future.

Billing mode

A data job is billed by the data engine usage. Currently, pay-as-you-go and monthly subscription billing modes are supported. For more information, see Data Engine Overview.

Pay-as-you-go: It is applicable to scenarios with a small number of data jobs or periodic usage. A data job is started after creation and automatically suspended after successful execution, after which no fees will be incurred. Monthly subscription: It is applicable to scenarios where a large number of data jobs are regularly executed. Resources are reserved in this mode, so you don't need to wait for data engine start.

Note:

As a data job differs from a SQL job in terms of the compute engine type, you need to purchase a separate data engine for Spark jobs; otherwise, you can't run data jobs on a SparkSQL data engine.

Job management

On the **Data job** management page, you can create, start, modify, and delete a data job.

1. Log in to the Data Lake Compute console and select Data job on the left sidebar.

2. Click Create job. For detailed directions, see Creating Data Job.

3. In the list, you can view the current task status of the data job. You can also manage the job as instructed in Managing Data Job.

Configuring Data Access Policy

Last updated : 2024-07-17 17:44:52

Data Access Policy (CAM role arn) Overview

A data access policy (CAM role arn) allows you to configure permissions in CAM for accessing data in data sources and COS during data job execution.

When configuring a data job in Data Lake Compute, you need to specify the data access policy to protect data security.

Directions

Step 1. Create a policy in CAM

1. Log in to the Tencent Cloud console and select **Cloud Access Management**. The logged-in account needs to have permissions to configure CAM; therefore, we recommend you use a root account or admin account.

2. Select **Policies** on the left sidebar to enter the policy management page. Click **Create Custom Policy** and select **Create by Policy Syntax**.

S Tencent Cloud	Overview Cloud Products ~		Searc	h for products C	Q 🖂 🖓 Orga	anization Account ~	Tools - Ticket -	Billing Center - English
Cloud Access Management	Policies							V Account ID:
E Dashboard								Account to.
은 Users 🗸 🗸	 Associate users or user groups with policies to 	grant permissions.						Account information
ዲ User Groups	Create Custom Policy			All Policies	Preset Policy	Custom Policies	Search by policy (C Security settings
Delicies								😔 Access management
🔄 Roles	Policy Name	Service Type T	Description			Last Modified		Project management
🛅 Identity Providers 👻	- Animal and a second s	-				. 2018-08-13 17:5	4:58	Security management
Federated * Account						2022-11-07 11:1	8:31	Preferences
(ছ) Access Key 👻						2021-08-09 10:4	2:42	Switch account
	-				,	2018-08-13 17:5	4:58	Associate User/User Grou
	a provide the	10000				2020-09-29 11:3	7:49	Associate User/User Grou
						2020-09-29 11:3	7:49	Associate User/User Grou

3. Search for COS in the policy template and select COS permission templates.

Create by Policy Syntax		
1 Select Policy Template > 2 Edit Policy		
Template Type: All Templates v cos	λ	
Select a template type		
All Templates (51 Total)Search "cos", to find 51 items-Back to Original List		
QcloudCOSBucketConfigRead Grant READ-only access on COS Bucket Configurations	QcloudCOSBucketConfigWrite Grant WRITE-only access on COS Bucket Configurations (exclude DELETE)	CcloudCOSDataFullControl Grant READ WRITE DELETE LIST access on COS
QcloudCOSDataReadOnly Grant READ-only access on COS (exclude LIST)	QcloudCOSDataWriteOnly Grant WRITE-only access on COS (exclude DELETE)	QcloudCOSFullAccess Full read-write access to Cloud Object Storage (COS)
QcloudCOSGetServiceAccess Access to the bucket list of Cloud Object Storage (COS)	O Grant LIST-only access on COS (List Buckets & Objects)	QcloudCOSMetaAccMgmt This preset policy is used to authorize sub-accounts to operate the CHDFS file system associated with COS bucket, which is enable Meta Accelerator configuration.
Nest		

The preset templates define read-only and read/write permission policies. If they don't meet your needs, create a custom policy template as instructed in Appendix.

4. Select the template, set a name for the policy, and click **Save**.

Step 2. Create a service role

1. Log in to the Tencent Cloud console and select **Cloud Access Management**. The logged-in account needs to have permissions to configure CAM; therefore, we recommend you use a root account or admin account.

2. Select **Role** on the left sidebar to enter the role management page. Click **Create Role** and select **Tencent Cloud Product Service**.



3. In the Role Entity service list, find and select Data Lake Compete and click Next.



Enter Role	Entity Info > (2) Configure Role Policy >	(3) Set Role Tag > (4) Review			
iduct Service	Cloud Security Integrated Platform(CSIP) (csip)	Tencent Cloud Advisor (advisor)	Aagis (aagis)	Shenbi Low-Code Application Platform as a Service (shenbi)	API Gateway (apign)
	Auto Scaling (as)	Application Service Workflow (asw)	TBaaS (thaas)	Billing (billing)	BlueKing (blueking)
	Cloud Physical Machine (bm)	BPasS (bpass)	Cloud Access Security Broker (casb)	Tencent Kubernetes Engine (tke)	Cloud Database (cdb)
	Cloud Data Coffer Service (cdcs)	TencentCloud Component Development Tools (cdevops)	CDN (cdn)	Cloud File Storage (cfs)	Cloud Firewall (cfw)
	Customer Growth Expert (oge)	Cloud Infinite (c)	Ciárica (cicaño)	Cloud Loader Balance (clb)	Cloud Audit (cloudaudit)
	Cloud Studie (cloudstudie)	Cloud Log Service (cls)	cloudWaf (waf)	Tencent Cloud Observability Platform (monitor)	Creativity Platform (cme)
	CODING DevOps (coding)	COS (cos)	Cloud Storage Gateway (csg)	Cloud Training Platform (ctp)	TencentDB for CTSDB (ctsdb)
	Cloud Virtual Machine (cvm)	Cloud Workload Protection (cxp)	Tencent Cloud Developer-TDP (devops)	Di (di)	Data Lake Compute (dic)
	Data Security Governance Center (dsgc)	Data Transfer Service (dts)	EventBridge (eb)	App Flow (els)	Elasticsearch MapReduce (emr)
	faceid (faceid)	Game Sever Bastic-scaling (gse)	Image AI (facerecognition)	DasS (idaas)	lotHub (lothub)
	IoT Suite (iotsuite)	Internet of Things Video (iotvideo)	Intelligent Viewdata Computing (iss)	Developer Laboratory (labs)	Cloud Streaming Services (live)
	CDB for MariaDB (TDSQL) (mariadb)	StreamLive (md)	StreamPackage (mdp)	Message Center (message)	Mobile Game Online Battle Engine (mgobe)
	Cloud MongoDB (mongodb)	Media Processing Service (mps)	Migration Service Platform (msp)	Media Transcoding Service (mts)	Network Assets Risk Monitor System (narms)
	Publicly Accessible Instance-RAI (pai)	Serverless Cloud Function (scf)	Stream Compute Service (scs)	Serverless Framework (sls)	Security Situation Awareness (ssa)
	Secrets Manager (ssm)	Tencent Cloud Base (tcb)	Tencent Cloud Display (tcd)	Tencent Cloud Mesh (tcm)	Tencent Container Registry (tcr)
	Tencent Container Security Service (tcss)	Tencent Database Middleware (tdm)	(a) LL (a)	TI Accelerator (tis)	Tencent Cloud Infrastructure as Code (tic)
	TencentCloud Ti Platform TI-ONE (tione)	Ti Self-Learning (tis)	Tencent Interactive Whiteboard (tiw)	Tencent Cloud Service Engine (tse)	Tencent Service Framework (tsf)
	Tencent Support System (tss)	Cloud Shield - Data Data Access Security Broker (dasb)	Video Moderation System (vm)	VOD (vod)	Vulnerability Scan Service (vss)
	WeData (wedata)	WeMall (wemall)	workorder (workorder)	YouMall (youmall)	Cloud Operations Console (zhiyun)
cases to choose	Data Lake Compute Allow Data Lake Compute to access your tencent cloud resources				
	Data Lake Compute - CheckDLCResource The current role is the DLC service role, which will access your other	service resources within the scope of the permissions of the associated p	seliq.		

- 4. In the policy configuration, find and select the policy created in Step 1 and click **Next**.
- 5. Set a name for the role and click **Save**.

Step 3. Get the role arn information

- 1. After creating the role in Step 2, return to the role list and find the created role.
- 2. Click **Role Name** to enter the role details page.

Delicies	Create Role					Search by role	ID/name/description (separ Q 🌣
🕃 Roles								
🗔 Identity Providers 👻	Role Name	Role ID	Role Entity	Description	Tag Information T	Max Session D	Creation Time	Operation
Federated ~ Account			Product Service - dlc		© 1	2 hours	2023-12-14 14:2	Delete
(양) Access Key 💙	-	14				12 hours	2023-12-06 16:2	Delete
						12 hours	2023-09-14 10:3	Delete

3. Find and copy the role arn information.

÷	•	
Ro	le Info	
Rol	e Name	
Rol	eArn	
Rol	e ID	
Des	scription	-/
Cre	ation Time	2023-12-14 14/21/29

Step 4. Configure the role arn in Data Lake Compute

1. Log in to the Data Lake Compute console with an admin account.



2. Select Data job on the left sidebar to enter the data job management page. Click Job configuration and select

CAM role arn.

3. Click Create role arn.

Data Lake Compute	Data job 🖏 Guangzhou 🔻				
E Overview	Spark job Job configuration	Session management			Job monitoring Task history Lo
) Data Explore	 The data job feature requires the 	access policy (CAM role) set by the Data Lake Compute admin to sp	ecify the scope of COS data access. For access	to data from other data sources, set the network via Network configuratio	n 🗹 first.
∃‡ Data Management					
🗄 Data Jobs	CAM role arn				Create role
😚 Data Engines 🗸 🔹	CAM role arn	Configuration description	Creator	Update time	Operation
Slobal * Configuration					
텔 Data Operations 👻			-0	No role available	
	Total items: 0				10 v / page H 4 1 / 1 page

4. Paste the role arn information obtained in Step 3 in the input box and click Save.

Appendix: Custom Policy Template

If the preset templates cannot meet your data management needs, you can configure a custom template in the following steps.

1. Log in to the Tencent Cloud console and select Cloud Access Management. The logged-in account needs to have permissions to configure CAM; therefore, we recommend you use a root account or admin account.

2. Select **Policies** on the left sidebar to enter the policy management page. Click **Create Custom Policy** and select **Create by Policy Generator**.



3. Select Allow as Effect and COS as Service. Select the resource scope as needed.

Cloud Access Management	← Create by Policy Generato)r	
B Dashboard			
옥 Users Y	1 Edit Policy >	Associate User/User Group/Role	
😫 User Groups		10011	
Delicies	Visual Policy Generator	JSON	
🖞 Roles	▼ COS(0 actions)		
Identity Providers 👻	Effect *	O Allow O Deny	
Federated Account	Service *	COS (cos)	
₽) Access Key Y	Action = Collapse	Select actions All actions (cos:*) Show More Add Custom Action Action Type Read Show More Urite Show More List Show More	
	Resource *	Select resource	
	Condition	Source IP (1) Add other conditions	
	+ Add Permissions		
	Next Characters: 114 (up t	o 6,144)	

If you need to manage specific resources, click Add a six-segment resource description to add resources. You

can use * to indicate all the resources. For more information, see Resource Description Method.

4. After completing the configuration, set a name for the policy and click **Save**. You can also select **Authorized Users** to authorize the policy to existing users.

Creating Data Job

Last updated : 2024-07-17 17:45:32

Preparations

Before creating a data job, you need to configure the CAM role arn to secure the data access from the data job. For detailed directions, see Configuring Data Access Policy.

Directions

- 1. Log in to the Data Lake Compute console and select Data job on the left sidebar.
- 2. Click Create job.

Create job	×
Basic info ▲ Job name *	Enter a job name It can contain up to 100 characters in Chinese characters, letters, digits, and underscores (_).
Job type *	Batch processing Stream processing SQL job
Data engine *	Select a Spark job engine The billing mode of the selected data engine prevails. For more info, see Data engine Z. For network configuration of the data engine, see Network configuration Z.
Program package *	COS Upload Select a data path Select a COS path COS permissions are required, and .jar/.py files are supported.
Main class *	This field is required if the program package is a .jar file
Program entry parameter	Enter program input parameters of up to 65,536 characters; separate two parameters by space
Job parameter (config)	Example: spark.network.timeout=120s
	-config info, the parameter info started with "spark.", one entry per line.
CAM role arn *	Select a CAM Role arn 🔻 🗘
Network configu	It determines the data access scope of a Spark job. For configurations, see Configure CAM role arn 🛿
Create job	Cancel

Configure parameters as follows:

Parameter	Description
Job name	It can contain up to 40 letters, digits, and underscores.
Job type	In batch: Batch data jobs based on Spark JAR In flow: Flow data jobs based on Spark Streaming
Data source connection	Data source for In batch data jobs. Currently, it can only be CKafka, which needs to be configured in advanced in Job configuration .
Data engine	It can be a Spark job data engine for which you have the permission. If you select Data source , you can only select a data engine connected to the data source.
Program package	The JAR format is supported.



	You can select a local file of up to 5 MB in size or a file in COS. If the local file exceeds 5 MB, upload it to COS for use. You can directly enter a COS path.
Dependency JAR resource	The JAR format is supported. You can select multiple resources. You can select a local file of up to 5 MB in size or a file in COS. If the local file exceeds 5 MB, upload it to COS for use. You can directly enter multiple COS paths and separate them by semicolon.
Dependency file resource	You can select a local file of up to 5 MB in size or a file in COS. If the local file exceeds 5 MB, upload it to COS for use. You can directly enter multiple COS paths and separate them by semicolon.
CAM role arn	The data access policy configured in Job configuration , which specifies the scope of data accessible to a data job. For more information, see Configuring Data Access Policy.
Main class	JAR package parameter in the main class. Separate multiple parameters by space.
Job parameter	-config information of the job, which starts with spark. in the format of $k=v$. Separate multiple parameters by line break. Example: spark.network.timeout=120s
Resource configuration	The engine resources that can be configured with the data job, the number of which cannot exceed the specifications of the selected data engine. Resource description: $1 \text{ CU} \approx 1$ -core 4 GB MEM Billable CUs = executor resource * executor quantity + driver resource Pay-as-you-go data engines are billed by the billable CUs.

3. After configuring the parameters, click **Save**.

Managing Data Job

Last updated : 2025-03-07 15:27:25

This document describes how to manage a data job. Edit a data job. Start and stop a data job task. View the data job and task details. Delete a data job.

Editing a data job

Note:

A running data job cannot be edited.

The type of a data job cannot be changed. To change it, create a new data job as instructed in Creating Data Job.

1. Log in to the Data Lake Compute console, select the service region, and select Data job on the left sidebar.

2. Find the target data job and click Edit.

Create job	Enter a job name or ID	Q, All	▼ All	¥			All Last 7	days Last 30 days	Select date Select date
Job name	Job ID	Job type	Job file	Current tasks	Task engine \$	Creator \$	Created at \$	Update time 💲	Operation
. q. 1					0		2023-10-17 21:20:34	2023-10-17 21:34:28	Monitor Edit Runn Delete

3. Edit the content and click Save.

Starting and stopping a data job task

You can start and stop a created data job to generate corresponding tasks. A data job can generate multiple task instances and be executed multiple times.

Data task statuses are as follows:

Status	Description
Not started	Initial status after creation.
Running	The data task is running, during which the data job cannot be edited or deleted.
Successful	The task is executed successfully.
Failed	Failed to run the task. You can query the error message through the log or SparkUI.



Canceled

The task is manually canceled.

You can start and stop a data job task in the following steps:

1. Log in to the Data Lake Compute console, select the service region, and select Data job on the left sidebar.

2. Find the target data job and click **Start** or **Stop** to change the task status.

Note:

Starting a task instance will use compute engine resources. If the usage exceeds the configured upper limit, the task will be put into a queue.

Create job	Enter a job name or ID	Q, All	▼ All	Ŧ			All Last 7	days Last 30 days	Select date Select date
Job name	Job ID	Job type	Job file	Current tasks	Task engine 💲	Creator 🕈	Created at \$	Update time 🗘	Operation
)	-	_			; j (j	2023-10-17 21:20:34	2023-10-17 21:34:28	Monitor Edit Runnin Delete

Viewing the Data Job and Task Details

Log in to the Data Lake Compute console, select the service region, and select Data job on the left sidebar.
 Click Job name to enter the data job details page.

Data job 🖏 Gui	angzhou 🔻					Spark job details
Spark job Job co	onfiguration Session man	agement				Job info Task history Monitoring and alerting
Create job Enter	a job name or ID 🛛 🔍 🔍	All	▼ All	Ŧ		Basic info *
Job name	Job ID	Job type	Job file	Current tasks	Task engine 🗘	Dab D
	a a	Batch processing			n (i)	Current task ID 🛛 🕶 🗗
	ā lieta ar	Batch processing			0	Current tasks Batch processing
5	n n	Batch processing	1		ı D	Data engine
р <u> </u>	. 6	Batch processing	/		0	Noine ·
	- TO	Batch processing			2	Program entry parameter
	To	Batch processing			()	Copy statement
	· 6	Batch processing				Job parameter
5		Batch processing			th (j)	Crestor
6	- To	Batch processing				Update time 2023-10-17 21:34:28
	6	Batch processing				Network configuration + Enhanced network

On the details page, you can view the basic information and task list of the data job. The task list contains the data



task information of the data job. You can view the task run log and SparkUI.

Job info	Task history	Monit	oring and a	lerting				
Select an exe	ecutic 🔻 🛛 Las	st 7 days	Last 30 da	ays 202	23-12-14 ~ 2023-12	2-20		Ref
Task ID	Executi	Task subr	nissi 🗘	Comput	Operation			
	Successful	2023-12-1	2 20:53:42	47.8s	Learn more Spa	ark UI		
Total items: 1					10 🔻 / page		1 /	1 page

Click **Learn more** or **Task ID** to view the task details, which include the basic information and run log of the task. Currently, the run log allows you to view the last 1,000 data entries.

Basic info	Run log					
Job nam∈ ∎ Up to latest 1,	■■■st Г 000 entries can be	Job II shown in the cor	nsole	ic F		
Last 7 day	/s Last 30 da	ys 2023-12	2-15 17:13:08 ~ 20	23-12-21 17:13:08 📋		
Desc by time	¢				Create download	task
_og name:				▼ Log level: S	elect 🔻	
				No data		
				No data		
				No data		
				No data		
			Ð	No data		
			-	No data		
			Ð	No data		

You can click Create download task to download the full log and click Log download to save the log locally.

l						×
Operated by	Job name	Job ID	Log name	Task ID	Status	Refresh Operation
20003763640 7	app_test_70z b4e	batch_6ee35 8a4-1	livy- 254afd95- 07	254afd95- 07ba-45	Completed	Save to local
			10 💌	/ page 🛛 🖌 🖣	1 / 1	I page 🕨 🕨
	Operated by 20003763640 7	Operated byJob name20003763640 7app_test_70z b4e	Operated byJob nameJob ID20003763640 7app_test_70z b4ebatch_6ee35 8a4-1	Operated by Job name Job ID Log name 20003763640 app_test_70z batch_6ee35 livy- 254afd95- 07 7 b4e aa4-1 254afd95- 07	Operated by Job name Job ID Log name Task ID 200003763640 app_test_70z batch_6ee35 livy- 254afd95- 07 254afd95- 07ba-45 07 10 ▼ / page ◄	Operated byJob nameJob IDLog nameTask IDStatus 20003763640 7app_test_70z b4ebatch_6ee35 8a4-1livy- 254afd95- 07254afd95- 07ba-45Completed 0 $10 \neq / page$ id1/ id



Note:

The download record will be saved for three days, after which you cannot save the log locally and need to create a new download task.

Deleting a data job

Note:

A data job with a running data task cannot be deleted.

1. Log in to the Data Lake Compute console, select the service region, and select Data job on the left sidebar.

2. Find the target data job, click **Delete** > **OK**.

Job name	Job ID	Job type	Job file	Current tasks	Task engine 🗘	Creator \$	Created at 🕈	Update time 🕈	Operation
· · · · · · ·	. 6	Batch processing			0	(j	2023-10-17 21:20:34	2023-10-17 21:34:28	Monitor Edit Run Delete

Note:

Note that deleting a data job will delete its data task information. Proceed with caution.

PySpark Dependency Package Management

Last updated : 2024-09-18 17:59:53

Currently, the basic running environment for DLC's PySpark uses Python 3.9.2.

Python dependencies for Spark jobs can be specified in the following two methods:

1. Use --py-files to specify dependency modules and files.

2. Use --archives to specify a virtual environment.

If your module or file is compiled by using pure Python to implement customized function, it is recommended to specify Python dependencies using the --py-files.

The <u>--archives</u> option allows you to package and use the entire development and test environment. This method supports compiled installations of C-related dependencies and is recommended when the environment is more complex.

Note:

The two methods mentioned above can be used simultaneously based on your needs.

Using --py-files to Specify Dependency Packages

This method is suitable for modules or files implemented in pure Python, without any C dependencies.

Step 1: Packaging Modules/Files

For external PyPI packages, use the pip command to install and package common dependencies in the local environment. The dependencies should be implemented in pure Python and should not be dependent on any C-related databases.

```
pip install -i https://mirrors.tencent.com/pypi/simple/ <packages...> -t dep
cd dep
zip -r ../dep.zip .
```

The single-file module (e.g., functions.py) and custom Python modules can be packaged by using the method mentioned above. It is important to ensure that custom Python modules are standardized according to Python's official requirements. For more details, see the official Python Packaging User Guide.

Step 2: Importing the Packaged Module

In the Data Lake DLC Console, create a job in the Data Job module. Use the -py-files parameter to import the packaged dep.zip file, which can be uploaded either through COS or directly from your local device.



Program package *	OCOS Upload	
	Select a data path	Select a COS path
	COS permissions are required, and .jar/.py files are supported.	

Using a Virtual Environment

A virtual environment can resolve issues with some Python dependency packages that are dependent on C databases. Users can compile and install dependency packages into the virtual environment as needed, and then upload the entire environment.

Since C-related dependencies involve compilation and installation, it is recommended to use an x86 architecture machine, Debian 11 (Bullseye) system, and Python 3.9.2 environment for packaging.

Step 1: Packaging the Virtual Environment

There are two methods to package a virtual environment: using Venv or Conda.

1. Packaging with Venv.

```
python3 -m venv pyvenv
source pyvenv/bin/activate
(pyvenv)> pip3 install -i [https://mirrors.tencent.com/pypi/simple/]
(https://mirrors.tencent.com/pypi/simple/) packages
(pyvenv)> deactivate
tar czvf pyvenv.tar.gz pyvenv/
```

2. Packaging with Conda.

```
conda create -y -n pyspark_env conda-pack <packages...> python=<3.9.x>
conda activate pyspark_env
conda pack -f -o pyspark_env.tar.gz
```

After packaging is completed, upload the packaged virtual environment file pyvenv.tar.gz to COS.

Note:

Use the tar command for packaging.

3. Use the provided packaging script.

To use the packaging script, you need to have docker installed. The script currently supports Linux and macOS environments.

```
bash pyspark_env_builder.sh -h
Usage:
    pyspark-env-builder.sh [-r] [-n] [-0] [-h]
    -r ARG, the requirements for python dependency.
    -n ARG, the name for the virtual environment.
    -o ARG, the output directory. [default:current directory]
    -h, print the help info.
```

Parameter	Description
-r	Specifies the location of the requirements.txt file.
-n	Specifies the name of the virtual environment (default: py3env).
-0	Specifies the local directory to save the virtual environment (default: the current directory).
-h	Prints help information.

```
# requirement.txt
requests
# Execute the following command.
bash pyspark_env_builder.sh -r requirement.txt -n py3env
```

After the script running is completed, you can obtain py3env.tar.gz in the current directory and then upload this file to COS.

Step 2: Specifying the Virtual Environment

In the Data Lake DLC console, create a job in the Data Operation Module following the instructions as shown in the screenshot below.

1. For the <u>--archives</u> parameter, enter the full path to the virtual environment. The name of the decompressed folder is After the #.

Note:

The # symbol is used to specify the decompression directory. The decompression directory will affect the configuration of the subsequent running environment parameters.

2. In the --config parameter, specify the running environment settings.

For the Venv packaging method, configure: spark.pyspark.python =

venv/pyspark_venv/bin/python3

For the Conda packaging method, configure: spark.pyspark.python = venv/bin/python3
For the script packaging method, configure: spark.pyspark.python = venv/bin/python3
Note:

Due to the differences in packaging methods between Venv and Conda, the directory structure will vary. You can decompress the .tar.gz file to check the relative path of the Python file.

Resource Management Engine Management Data Engine Introduction

Last updated : 2025-04-15 16:25:35

The DLC data engine is the foundation of DLC's data analysis and computation services. All calculations performed by users within DLC require the use of this data engine. Depending on the specific use case, users can select the appropriate engine type.

Engine Types

DLC offers two types of data engines for users to choose from: **Standard Engine** and **SuperSQL Engine**. The primary difference between these two engines lies in the SQL syntax they support. The Standard Engine uses native Spark and Presto syntax from the community, while the SuperSQL Engine supports DLC's independently developed unified syntax. This unified SuperSQL syntax can run on both Spark and Presto engines, effectively masking the syntax differences between them. This feature can significantly reduce usage costs in scenes where different analytics engines need to be used together. Below are the main characteristics of each engine and recommendations for selection:

Engine Types	Available Types	Main Features	Usage Requirements	Purchase Recommendations
Standard Engine	Spark Presto	Native syntax: Uses the native syntax from the Spark/Presto community, ensuring low learning and migration costs. Flexible usage: Supports both Hive JDBC and Presto JDBC. Integrated Spark: The standard Spark engine can execute SQL and Spark batch tasks.	Currently, a 2 CU specification free gateway is provided. If you need to upgrade the specification, upgrade the Gateway	 Require the use of native Spark/Presto syntax. Need to purchase a Spark engine for batch processing and offline SQL tasks. Prefer to use Hive JDBC and Presto JDBC.
SuperSQL Engine	SparkSQL	Unified syntax: A set of syntax applies to both	You need to learn the SuperSQL unified	1. Prefer to use a unified syntax for both Spark



Spark jobs\\nPrest	Spark and Presto engines. Supports federated queries.	syntax.\\nFor SQL/batch task scenes, it is recommended to purchase the corresponding engine type.	and Presto. 2. Need to perform federated queries.
-----------------------	--	---	---

For more detailed information, see the comparison table below or review the documentation for the Standard Engine and SuperSQL Engine Description.

Detailed Comparison of Standard Engine and SuperSQL Engine

Feature	Standard Engine	SuperSQL Engine	Description
Presto	1	\checkmark	Both engines support the Presto engine.
Spark	1	✓	The SuperSQL Engine is divided into SparkSQL and Spark job. The SparkSQL engine supports SQL jobs, while the Spark job engine supports Spark batch and streaming jobs as well as SQL jobs. The Standard Engine is an integrated Spark engine.
SQL Syntax	Native syntax	Unified syntax	The Standard Engine supports native Spark and Presto syntax. The SuperSQL Engine supports DLC's self-developed unified syntax.
Gateway	1		DLC, based on Apache Kyuubi, has developed its own Serverless gateway service, providing a more stable, secure, and high-performance task submission experience.
Resource Group	1		Resource groups are a unique feature of the Standard Spark Engine, allowing resources to be allocated as needed. SQL tasks can be submitted to a designated resource group for execution.
Shared Engine		✓	The SuperSQL Engine supports a shared mode, which is suitable for scenes with low analysis frequency and smaller data volumes.
Hive JDBC	1		The Standard Engine supports submitting tasks using Hive JDBC.



Presto JDBC	1		The Standard Engine supports submitting tasks using Presto JDBC.
DLC JDBC	✓	✓	Both types of engines support submitting tasks using DLC JDBC.
TencentCloud API Task Submission	1	✓	Both types of engines support submitting tasks using TencentCloud API or through the data exploration page in the console.
Federated Query		✓	The SuperSQL Engine provides federated query analysis capabilities. For instructions on adding a federated query data catalog, see Data Directory and DMC. The Standard Engine currently does not support federated queries.

If you have any questions about choosing between the Standard Engine or SuperSQL Engine, you can Submit a Ticket to contact us.

Engine Pricing

Data engines support both monthly subscription and pay-as-you-go subscription. For more information, see Billing Overview.

Limitations

The name of the data engine should be globally unique and cannot be changed.

The billing mode of the data engine cannot be switched.

The data engine does not support changing regions.

SuperSQL Engine SuperSQL Engine Overview

Last updated : 2025-03-07 15:27:25

Data engines empower the data analysis and computing service in Data Lake Compute. They are used in all computing operations and can be public or private based on your needs.

Public engine

The Data Lake Compute service comes with the shared public engine, which is applicable to low-frequency analysis use cases with small data volumes. With this highly flexible and available engine, you don't need to configure or manage resources. Fees are charged by the scanned data volume of running tasks. For billing details, see Billing Overview.

Since Data Lake Compute adopts serverless architecture, it needs to schedule the data engine for task execution for the first time over a period of time, which may take a longer time.

Private engine

A private engine is a dedicated data engine that you purchase on a pay-as-you-go basis. For billing details, see Billing Overview.

Pay-as-you-go: This billing mode is highly flexible and stable, where fees are charged by the CU usage. It is applicable to use cases where data is analyzed regularly, with compute resources elastically scaled based on the business load. Monthly subscription: This billing mode is applicable to use cases where large amounts of data require long-term and stable analysis, with compute resources elastically scaled based on the business load. It guarantees always available resources with no need to wait for resource startup. Fees are charged by month based on the cluster specification (elastic clusters are billed by CU usage though).

Compute engine types

A private engine can work with different compute engines in different use cases. SparkSQL: It is suitable for stable and efficient offline SQL tasks. Spark job: It is suitable for native Spark stream/batch data job processing. Presto: It is suitable for agile and fast interactive query and analysis. **Note:** The compute engine type does not affect the unit price of a private engine.

Engine scaling rules

The elastic scaling rules for the engine can be configured either in Create Engine or in the SuperSQL Engine.within the Console Data Engine.

Cluster count	- 1 +	
	Multiple clusters with fixed s	pecs can be configured in a data engine to increase task concurrency
Max task concurrency of a cluster	- 20 +	
	The max number of concurre longer compute time. When be queued up.	ent tasks that a cluster can process. A higher concurrency may result i the concurrency reaches the concurrency limit of a cluster, new tasks
Cluster scaling rules	Yes No Sca	ling rules 🖸
	Elastic clusters	- 1 +
		For elastic clusters, resources will be scaled based on the task concurrency and queue time and billed based on CU usage.
	Task queue-up time limit	- 1 + Minute
		The max task queue-up time. If it is set to 0, auto scaling will be triggered immediately after a task is queued. When the queue-up time exceeds this value, the cluster will be auto scaled after elastic resources are made available (the time varies by the number of resources required).

The number of clusters refers to the number of resident clusters in the engine. The sum of the total number of clusters and elastic clusters represents the maximum number of clusters the engine can scale to during elastic scaling. Basic rule: Engine scaling will only occur when the number of elastic clusters is greater than zero.

Scale-out rule: The system will scale out the data engine based on the configured rules when the number of queued tasks exceeds the available concurrent capacity, the task queue time surpasses the queue time limit, and no clusters are being initialized.

Scale-in rule: The system will scale in the data engine when the current number of clusters exceeds the number of resident clusters, the overall average load of the clusters is below 20%, and there are idle clusters. As shown in the figure below: During the purchase, the number of clusters is set to 2, the number of elastic clusters to 3, and the task queue time limit to 5 minutes. During high concurrency of cluster tasks, if the number of queued tasks exceeds 2 and the queue time exceeds 5 minutes, the system will scale out the data engine to alleviate the task queuing situation. After successful scale-out, if the task queuing situation is alleviated, clusters become idle, and the load is low, the system will scale in the data engine.



In the case of elastic scaling, the number of clusters in the data engine will not be less than the configured cluster count and will not exceed the sum of the configured cluster count and the elastic clusters.

For example, if the configured number of clusters is 2 and the number of elastic clusters is 3, after scaling out, the number of clusters will not exceed 5, and after scaling in, the number of clusters will not be fewer than 2.

Note:

The cluster count of a data engine cannot be smaller than the minimum cluster count. A pay-as-you-go cluster can be suspended if it is not needed.

Engine running status

A cluster may be in one of the following eight statuses: Starting, Running, Suspended, Suspending, Changing configuration, Isolated, Isolating, Recovering.

Starting: The cluster is being started. In this case, a pay-as-you-go private engine is not billed. A starting cluster cannot be selected for data computing.

Running: The cluster is running and can be selected for data computing.

Suspended: The cluster is suspended and cannot be selected for data computing.

Suspending: The cluster is being suspended and cannot be selected for data computing. This will affect running tasks. Changing configuration: The cluster is undergoing a configuration change and cannot be selected for data computing. Isolated: The cluster is isolated due to overdue payments and cannot be selected for data computing.

Isolating: The cluster is being isolated due to overdue payments and cannot be selected for data computing. This will affect running tasks.

Recovering: The cluster is being recovered from the **Isolated** status to the **Running** status after the account is topped up. It cannot be selected for data computing.

Purchasing Private Data Engine

Last updated : 2024-07-17 17:55:49

A private data engine in Data Lake Compute supports pay-as-you-go and monthly subscription billing modes. For billing details, see Billing Overview.

Private engine purchase

You can purchase on the Data Lake Compute purchase page or in the console as instructed below:

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin or financial collaborator permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Click **Create resource** in the top-left corner to enter the **Resource configuration** page. Configure the resource as needed and view the estimated price.

Data Lake Compute offers required; a private data eng the auto-suspension or sch	both public and priva gine can be billed on a neduled suspension p	te data engines. A p a pay-as-you-go bas olicy, with no fees cl	ublic data engine is ma sis or subscribed mont harged on it after susp	anaged by Data Lake Compu hly. For more billing info, see ension. For operations and n	te and billed by scanned Billing Overview 🖸 . A p otes, see Managing Priv	data volume, with no operation oay-as-you-go data engine can l ate Data Engines 🕻 .	or permission be configured with
Create resource Bill query L	2 Renewal manage	ment 🛂			Select a reso	urce tag or enter keyword(s) (se	parate two
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
ف DataEngine-iwxhwnud T	SparkSQL	Running	SuperSQL-S 3.5	Expire	No	Manual start, Manual suspension	Monitor Spec configura Parameter Configuration More ▼
DataEngine-p3d2xfq1 🗗	Presto	Starting (j)	SuperSQL-P 1.0			Auto-start, Manual suspensi	Monitor Spec configura Parameter Configuration More ▼
DataEngine-public-1313074 🗗	Presto	Running	SuperSQL-P 1.0- public	volume		Manual start, Manual suspension	Monitor Spec configura Parameter Configuration More ▼
Total items: 3						10 🔻 / page 🛛 🕅 🔌	1 / 1 page

4. Confirm the price and make the purchase.

Data Lake	Compute	Back					Documentation 🗳 Billing 🖄 Console
Engine edition	SuperSQL engine	Beta Standard engine	3				
Billing mode	Pay-as-you-go In this mode, a cluster is b loads and irregular task cy	Monthly subscriptio	n Detailed company	rison Ľ bended when no task is in	progress. A suspended	cluster incurs no cost. It is	suitable for data compute applications with certain task
Region	-Hong Kong/Macao/TaiWa	an (China Region)	Southeast Asia ——	Eastern U.S	Europe	Southeast Asia Pacific -	
	Hong Ko	ong	Singapore	Virginia	Frankfurt	Jakarta	
Cluster config	Cloud products in differen region nearest to your cus uration	It regions are not interco	nnected over private s latency.	networks and the region	cannot be changed afte	r you purchase the service.	. Please proceed with caution.We recommend you select thr
Basic configuration	on		_				
Compute engine type	e SparkSQL	Spark job	Presto				
	This is a memory engine fo SparkSQL engine.	or distributed SQL query	/. It supports real-time	e data write to SQL and re	eal-time result return in	Data Explore. It is suitable f	for applications with small loads. It runs faster than a
Kernel version	SuperSQL-P 1.0	developed Drecto_base	angina karnal for int	teractive query and analyti	tice. Suntay rulae sunno	rtad hu diffarant karnel vers	sions are clightly different. For more information on versions
	see Kernel Versions 🕻	developed Flesto-Daset	a engine kerner for filt	teraetive query and allaly	ασο, σγιταλ τατές σύμμο	nea by amerent kentel vers	avono are angituy different. For more mitor mation off Versions

Configuration parameter description:

Region: Cloud products in different regions are not interconnected over private networks and the region cannot be changed after you purchase the service. Proceed with caution.

Compute engine: Presto and Spark engines are supported. Note that the engine cannot be changed once purchased. Presto is suitable for faster interactive query and analysis and multi-source federated query, while Spark is suitable for more stable offline tasks with large data volumes.

Cluster spec: Cluster specification is measured in CU. 1 CU equals to 1 CPU core and 4 GB memory of compute resources. The specification determines the amount of compute resources during task execution and can be purchased as needed.

Note:

If you need more than 152 CUs, submit a ticket for assistance.

Min cluster count: Set the minimum number of clusters during cluster start or resident resources in a monthly subscribed cluster. Multiple clusters can deliver a higher concurrency.

Max cluster count: Set the maximum number of clusters for elastic scaling. If it is the same as the minimum cluster count, elastic scaling is not enabled for the cluster.

Auto-start: If it is enabled, a suspended data engine will be automatically started after receiving a task request.

Note:

As pay-as-you-go resources are not reserved, it is possible that they cannot be started right away. If you need resident and stable compute resources, purchase a monthly subscribed data engine instead.

Suspension policy: Configure the suspension method of a pay-as-you-go data engine. Automatic suspension and scheduled suspension are supported. A suspended pay-as-you-go data engine will not incur fees.

Auto-suspension: The data engine will be automatically switched to the **Suspended** status after it has been idle for a certain period of time.

Timing policy: You can configure scheduled start and suspension policies by week. The system will start or suspend clusters regularly as configured.

Suspension after task end: After the specified time elapses, if a task is running, the system will automatically suspend the data engine within five minutes after the task ends.

Suspension after task pause: After the specified time elapses, if a task is running, the system will pause the task and suspend the data engine immediately.

Advanced configuration: If you need to use federated query, configure the IP range in the advanced configuration.

Tag: Set tags to categorize purchased resources and allocate costs. For more information, see Associating Tag with Private Engine Resource.

Bill query

You can query bills in the Data Lake Compute console in the following steps:

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin or financial collaborator permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Click **Bill query** to view the detailed bill and settlement information (the financial collaborator permission is required).

 Data Lake Compute offer required; a private data en the auto-suspension or so 	s both public and priv ngine can be billed or cheduled suspension	rate data engines. A p a pay-as-you-go bas policy, with no fees c	ublic data engine is m sis or subscribed mont harged on it after susp	anaged by Data Lake Comp hly. For more billing info, see ension. For operations and r	ute and billed by scanned Billing Overview 🖾 . A j notes, see Managing Priv	I data volume, with no operation pay-as-you-go data engine can I rate Data Engines ☑.	or permission
Create resource Bill query	Renewal manag	gement 🔀			Select a reso	urce tag or enter keyword(s) (se	oarate two
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
DataEngine-Iwxhwnucr g	SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Monitor Spec configuration Parameter Configuration More ▼
DataEngine-p3d2xtq1 In	Presto	Starting (i)	SuperSQL-P 1.0	Pay-as-you-go		Auto-start, Manual suspensi	Monitor Spec configuration Parameter Configuration More ▼
DataEngine-public-1313074 F	Presto	Running	SuperSQL-P 1.0- public	Pay by scanned data volume		Manual start, Manual suspension	Monitor Spec configuration Parameter Configuration More ▼
Total items: 3						10 ▼ / page 4 4	1 /1 page 🕨

Renewal management

For a monthly subscribed private data engine, you can perform renewal and other operations in the Data Lake Compute console > Renewal management > Resource management in the following steps:

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin or financial collaborator permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Click **Renewal management** to enter the resource list and renew resources (the financial collaborator permission is required).





Renewing SuperSQL Engine

Last updated : 2024-07-31 17:55:25

You can renew a monthly subscribed data engine that has not expired or is isolated in the Data Lake Compute console.

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin or financial collaborator permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Find the target data engine and click **More** > **Renew**. You can also renew resources that will expire soon (in seven days) by clicking **Renew** next to the expiration time.

Create resource Bill query 2	Renewal manager	nent 🛂			Select a resc	ource tag or enter keyword(s) (se	parate two C
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
Data Data	SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Monitor Spec configuration Parameter Configuration More
	Presto	Starting (i)	SuperSQL-P 1.0	Pay-as-you-go		Auto-start, Manual suspensi	Restart Terminate Renew

4. Check the renewal term and price and click **Confirm**. The renewal will be completed after the order is confirmed and paid.

Note:

The billing cycle of a data engine that is renewed from the isolated status will start from the expiration date of the previous cycle.

Managing Private Data Engine

Last updated : 2024-07-17 18:02:09

Note:

You don't need to manage the public engine, as it is managed by Data Lake Compute in a unified manner.

Modifying the private engine configuration

Note:

Fees may change as the private engine configuration changes. For more information, see Configuration Adjustment Fees Description.

Option 1. Data engine list

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin or financial collaborator permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Find the target private engine and click **Spec configuration** on the right to enter the configuration modification

page, where you can modify the cluster specification and elastic scaling policy.

4. After making changes, click **Save** to submit the order and make the payment.

Create resource Bill quer	ry 🗹 🛛 Renewal manag	gement 🛂			Select a reso	urce tag or enter keyword(s) (sej	parate two Q
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
1000	SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Monitor Spec configuration Parameter Configuration More ▼
	Presto	Starting (j)	SuperSQL-P 1.0	Pay-as-you-go		Auto-start, Manual suspens	Monitor Spec configuration Parameter Configuration More ▼
<u></u>	Presto	Running	SuperSQL-P 1.0- public	Pay by scanned data volume		Manual start, Manual suspension	Monitor Spec configuration Parameter Configuration More ▼
Total items: 3						10 🔻 / page 🛛 🗸	1 / 1 page 🕨

Option 2. Data engine details



1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin

or financial collaborator permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Find the target private engine and click the cluster name to enter the cluster details page, where you can modify the cluster specification and elastic scaling policy.

4. Adjust the parameters as needed and click **Save**.

← SuperSQL engine ■ ■ ■ ■ ■ ■						
Basic configuration Cluster monitoring	Alarm con					
Basic info	Configuration info Set start and stop policy Change spec configuration					
Engine name 🔳 📰 🖬 🖬 olī Resource ID DataEngine-p3d2xfq1 lī	Engine type Presto Kernel version SuperSQL-P 1.0 Engine Size 16 CU					
Description Region Hong Kong/Macao/TaiWan (China Region)-Hong Kong	Cluster count 1					
Engine Status Starting 🤣 Billing mode Pay-as-you-go	Auto-scaling Yes Elastic cluster count 4 Max task concurrency of a cluster 20 Task queue-up time limit 0 minute(s)					
Tag No tag Tags are used to categorize resources. To learn more, see Tag Documentation 12	Auto-start Yes Auto-suspension No Timing policy None					
	IP range of cluster 10.255.252.0/22 Network configuration					

Modifying the private engine information

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Find the target private engine and click the cluster name to enter the cluster details page, where you can modify

the cluster description, automatic start policy, and suspension policy.

4. Adjust the parameters as needed and click Save.

← SuperSQL engine ■ ■ ■ ■ ■	
Basic configuration Cluster monitoring	Alarm con
Basic info	Configuration info Set start and stop policy Change spec configuration
Engine name 🔳 📰 🖿 📄 o 🗗 Resource ID DataEngine-p3d2xfq1 🗗	Engine type Presto Kernel version SuperSQL-P 1.0 Engine Size 16 CU
Description	Cluster count 1
Region Hong Kong/Macao/TaiWan (China Region)-Hong Kong Engine Status Starting 🗘	Auto-scaling Yes Elastic cluster count 4 Max task concurrency of a cluster 20
Billing mode Pay-as-you-go	Task queue-up time limit 0 minute(s)
Tag No tag /> Tags are used to categorize resources. To learn more, see Tag Documentation I2	Auto-start Yes Auto-suspension No Timing policy None
	IP range of cluster 10.255.252.0/22
	Network configuration

Suspension policy: Configure the suspension method of a pay-as-you-go data engine. Automatic suspension and scheduled suspension are supported. A suspended pay-as-you-go data engine will not incur fees.

Auto-suspension: The data engine will be automatically switched to the **Suspended** status after it has been idle for 15 minutes.

Timing policy: You can configure scheduled start and suspension policies by week. The system will start or suspend clusters regularly as configured.

Suspension after task end: After the specified time elapses, if a task is running, the system will automatically suspend the data engine within five minutes after the task ends.

Suspension after task pause: After the specified time elapses, if a task is running, the system will pause the task and suspend the data engine immediately.

Enable suspension policy management

It supports the configuration of start & suspend policies for the exclusive data engine of billing by volume, which facilitates management and cost control.

Note:

If the pay-as-you-go data engine is not suspended, charges will be generated. If the data engine is not needed, suspend it in time.

Startup policy: Supports automatic start, manual start, and scheduled start of the data engine.

Automatic start: After the configuration, if the data engine is in the suspended state and a task is submitted to the data engine, the data engine will automatically start.

Manual start: After the configuration, if the data engine is in the suspended state, you need to manually start the data engine before processing data tasks.

Periodic startup: You can configure a weekly periodic startup policy. The system periodically starts the cluster based on the configuration rules.

Timing policy												
	Scheduled start	Mon	~	09:00		0						
	Scheduled	Mon	~	20:00		ß						
	Suspension											
	Suspension option	Suspension after task Suspe	nd after t	ask pause								
		The suspension rules that can be last task is ended. "Suspension af	set after th ter task pa	e scheduled suspensi use" means that resou	ion feature is enabl urces will be suspe	led. "Sus ended at	pension after the specified	task end" me suspension ti	eans that reso me, with ongo	urces will be su ving tasks pause	spended 5 mi ed by the syst	inutes afte tem.

Suspension policy: Supports the suspension mode of the data engine for charging by volume, including automatic suspension and scheduled suspension. Pay-as-you-go data engines do not incur any costs when suspended. Automatic suspension: After the configuration, the data engine automatically switches to the suspended state 10 minutes after there is no task, and the triggering time can be configured.

Auto-suspension	If this option is enabled, th	e data engine is automatically suspended after the set trigger time of no task; otherwise, the engine must be manually susper
	Auto-trigger time	- 10 + min
		valid range. F 666 milli, which will affect the time waiting for suspending the data engine.

Periodic policy - You can configure weekly periodic start and suspension policies. The system starts and suspends the cluster periodically based on the configuration rules.

Suspend after Completion: If a task is being executed by the data engine within the specified time, the data engine automatically suspends the task within 5 minutes after the task is completed.

Suspend after Automatic pause: If a task is being executed on the data engine within the specified time, the system suspends the task and immediately suspends the data engine.
Manually suspending/starting a private engine

Note:

Monthly subscribed resources are resident and cannot be suspended.

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin permission.

- 2. Click **Data engine** on the left sidebar to enter the data engine management page.
- 3. Find the target private engine, click More, and select Start or Suspend in the drop-down list.

Create resource Bill query	Z Renewal manager	ment 🛂			Select a resol	urce tag or enter keyword(s) (sep	parate two Q
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
100	SparkSQL	Suspend (i)	SuperSQL-S 1.0	Pay-as-you-go		Auto-start, Manual suspensi	Monitor Spec configuration Parameter Configuration More 💌
	SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Start Suspend Terminate

Terminating a private engine

You can terminate a data engine that is no longer needed. A monthly subscribed data engine will be returned automatically after termination. For more information, see Refund.

Note:

Note that a pay-as-you-go data engine cannot be recovered once terminated. Proceed with caution.

1. Log in to the Data Lake Compute console and select the service region. You need to have the Tencent Cloud admin permission.

2. Click **Data engine** on the left sidebar to enter the data engine management page.

3. Find the target private engine (only suspended clusters can be terminated), click **More**, and select **Terminate** in the drop-down list.

4. Confirm the termination.

Create resource Bill query	Renewal managen	nent 🖸			Select a reso	urce tag or enter keyword(s) (sep	parate two C
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
100	SparkSQL	Suspend (i)	SuperSQL-S 1.0	Pay-as-you-go		Auto-start, Manual suspensi	Monitor Spec configuration Parameter Configuration More ▼
	SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Start Suspend Terminate

Cluster running logs

Data Lake Compute provides running logs within 14 days for private engines to help you stay informed of the start,

suspension, and scaling of clusters. Cluster logs mainly include the following content:

Start time: The time when the cluster starts working.

Suspension time: The time when the cluster stops working.

Scale-out record: The time of the cluster scale-out and the number of added clusters.

Scale-in record: The time of the cluster scale-in and the number of removed clusters.

ta	rtup and stop logs	Kernel versior	nmanagement
Lo	og info		
	Time	Action	Details
		Cluster scali	Before expansion: number of clusters is 1, cluster size is 16CU, after expansion: number of clusters is 0, cluster size is 16CU
		Cluster susp	Cluster suspended
	1. S. M. S.	Scaling out	Before expansion: number of clusters is 0, cluster size is 16CU, after expansion: number of clusters is 1, cluster size is 16CU
		Scaling out	Before expansion: number of clusters is 0, cluster size is 16CU, after expansion: number of clusters is 1, cluster size is 16CU
-	Total items: 4		10 v / page 🕅 🖪 1 / 1 page 🕨 🕅

Engine-Level Parameter Settings

Last updated : 2024-09-04 11:22:53

Note:

Currently, only the SparkSQL Engine and Spark Job Engine are supported for engine configuration. Spark parameters are used to configure and optimize the settings of Apache Spark applications. In a self-built Spark environment, these parameters can be set via command-line options, configuration files, or programmatically. In DLC, you can specify Spark parameters within the SQL and code of the SparkSQL Engine and Spark Job Engine, or you can directly set engine-level parameters. The engine-level Spark parameter configuration is as follows.

Setting Engine-Level Parameters

1. Enter the SupersSQL Engine module, click **Parameter Configuration**, and the engine parameter side window will appear.



2. Under the Spark Job Engine, you can configure the default resource specifications and parameters for jobs. In the SparkSQL Engine, there's no need to adjust the default resource specifications for jobs.

S Tencent Cloud

Configuration	change
Default job resource spec	
Executor resource *	small(1CU) Select desired resources. 1 CU is approximately equivalent to 1-core CPU and 4 GB memory.
Executor count *	Dynamic Fixed - 1 + Resources to be used by each executor are those set in the above field
Driver resource *	small(1CU) Select desired resources. 1 CU is approximately equivalent to 1-core CPU and 4 GB memory.
Total resource size	2CU
Parameter Configuration	(
	+ Add

Using Engine-Level Parameters

Spark Job Engine Using Engine-Level Parameters

There are two entry points for submitting jobs in the Spark Job Engine: Data Job and Data Exploration. Both support the use of engine-level parameters.

When you create a data job, the engine-level parameters and resource configurations are inherited by default. You can override the engine-level parameters using job parameters (--config) and choose whether to inherit the engine-level

resource configurations. If the default configuration is selected, the engine-level resource configuration will be used.

Create job				×
Job parameter (config)	Example: spark.network.timeout=120	S		
	-config info, the parameter info starte	d with "spark.", one entry per line.		
CAM role arn *	Select a CAM Role arn		▼ Ø	
	It determines the data access scope of	of a Spark job. For configurations,	see Configure CAM role arn 🗹	
Spark image	▼			
	Built-in dependency packages vary by	/ image. For more details, see Spa	rk dependency package notes 🔀	
Network config Dependencies	uration ►			
Resource confi	guration 🔺			
O Default config	uration (i) Custom configuration			
Executor	small(1CU)			
resource Executor count	1			
Driver resource	small(1CLI)			
Shive resource				

When you use the Spark Job Engine to run SQL in Data Exploration, the engine-level parameters and resource configurations are inherited by default. You can override the engine-level parameters using the set command within the SQL, and choose whether to inherit the engine-level resource configurations.

🗄 Data engine		Refresh
stevensli_notebook		SuperSQL-Spark
The engine supportsS	SuperSQL Syntaxquery, Viewing Synta	ax Description 🛂 .
Syntax rules. For det	ersion) Different kernel versions su ails, see Kernel Versions.	pport different SQL
Spark job (Spark 3.5	5)	•
① Create engine		
Resource configurat	ion	
O Default configurat	ion 🕕 🔘 Custom configuration	
Executor resource *	small(1CU)	-
	Select desired resources. 1 CU is app to 1 vCore and 4 GB memory.	roximately equivalent
Executor count *	Oynamic O Fixed	
	- 1 +	
Driver resource *	small(1CU)	•
	Select desired resources. 1 CU is app to 1 vCore and 4 GB memory.	roximately equivalent
Total resource size	2CU	

SparkSQL Engine Using Engine-Level Parameters

The SparkSQL Engine does not have engine-level resource parameters, so tasks will use as much of the cluster's resources as possible. Currently, SQL needs to be submitted using the SparkSQL Engine within Data Exploration. When you run SQL in Data Exploration with the SparkSQL Engine, engine-level parameters are inherited by default. You can override these parameters using the set command within the SQL.



Disaster Recovery Cluster

Last updated : 2024-07-31 17:47:09

To ensure the stable operation of the compute engine under extreme scenarios, DLC provides an efficient and agile disaster recovery cluster capability. When you need a disaster recovery cluster, you can quickly switch to it to ensure normal service operation. The disaster recovery cluster is only charged during operation, for more details, please see Cost Description.

Operation step

- 1. Enter the DLC Console, click Data Engine to access the Data Engine Page.
- 2. Click on the Data Engine Resource Name to enter the Data Engine Detail Page.

Data Lake Compute	SuperSQL engine	Hong Kong 🔻						SuperSQL engin	e guide 🖄
 ■ Overview ● Data Explore □ Data Scheduling 	Data Lake Compute offers required; a private data en the auto-suspension or sc	both public and priva gine can be billed on a heduled suspension p	te data engines. A p a pay-as-you-go ba olicy, with no fees c	bublic data engine is m sis or subscribed mon harged on it after susp	nanaged by Data Lake Compute thly. For more billing info, see bension. For operations and n	te and billed by scanne Billing Overview ^[2] . A otes, see Managing Pri	d data volume, with no operation , pay-as-you-go data engine can vate Data Engines ≧ .	or permission be configured with	×
≟≑ Data Management	Create resource Bill query	Z Renewal manage	ment 🖸			Select a res	ource tag or enter keyword(s) (se	parate two	Q Ø
🖹 Data Job	Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation	
 ■ Task History Ø Insight Management BETA 	document_test r DataEngine-44ncfc/n r	SparkSQL	Suspend (i)	SuperSQL-S 1.0	Pay-as-you-go		Auto-start, Manual suspens	Monitor Spec configurat Parameter Configuration More ▼	tion
Engine Management SuperSQL Engine		SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Monitor Spec configurat Parameter Configuration More ▼	tion

3. Click Enable Disaster Recovery Cluster and wait for the disaster recovery cluster to initialize.



	Configuration into Set start and stop policy Change spec configuration
ocument_test a Resource ID DataEngine-44ncfc7n a lone ✓ Hong Kong/Macao/TaiWan (China Region)-Hong Kong Suspend ↓ Pay-as-you-go No tag ↓ Tags are used to categorize resources. To learn more, see Tag Documentation 12	Engine type SparkSQL Kernel version SuperSQL-S1.0 Engine Size 16 CU Cluster count 1 Auto-scaling Yes Elastic cluster count 1 Max task concurrency of a cluster 5 Task queue-up time limit 0 minute(s) Auto-start Yes Auto-suspension No Timing policy None IP range of cluster 10.2555.0.0/16 Network configuration
ter	
er	

4. After the disaster recovery cluster is enabled, in the disaster recovery cluster information, click **Switch to Disaster Recovery Cluster** to adjust the running cluster to the disaster recovery cluster. Subsequently, jobs directed to this data engine will be submitted to the disaster recovery cluster. The disaster recovery cluster serves as a transition during extreme failures of the data engine.

Failover cluster	Failover cluster configuration
Backup cluster name document_test_backlin Backup resource ID DataEngine-cof24oj5lin	Engine type SparkSQL Kernel version SuperSQL-S 1.0 Engine Size 16 CU Cluster count 1
Engine Status Starting C Switch to failover cluster	Auto-scaling Yes Elastic cluster count 1 Max task concurrency of a cluster 5
Billing mode Pay-as-you-go	Task queue-up time limit 0 minute(s)
	Auto-start Yes Auto-suspension No Timing policy None

5. Once the extreme failure is resolved, in the basic information of the data engine, click **Switch to Primary Cluster**, and the disaster recovery cluster will be suspended. Subsequently, jobs directed to this data engine will be submitted to the primary cluster.

Basic info
Engine name document_test Resource ID DataEngine-44ncfc7n
Description None
Region Hong Kong/Macao/TaiWan (China Region)-Hong Kong
Engine Status Suspend Switch to primary cluster
Billing mode Pay-as-you-go
Tag No tag 🤌
Tags are used to categorize resources. To learn more, see Tag Documentation 🛂

Disaster Recovery Cluster Specifications

The disaster recovery cluster always tries to match the specifications of the data engine itself to ensure that the original tasks can transition and run normally. When AS is enabled on the data engine itself, the AS rules of the disaster recovery cluster will be consistent with the data engine. At the same time, to save costs, the disaster recovery cluster always operates on a pay-as-you-go basis.

Note on Fees

There is no charge for enabling the disaster recovery cluster. When switching to the disaster recovery cluster and it is running, charges will be applied according to the pay-as-you-go rates for the same specifications as the data engine. Example:

1. When the data engine itself is a 16 CU SparkSQL engine with an annual and monthly subscription. After enabling the disaster recovery cluster, it becomes a 16 CU SparkSQL engine on a pay-as-you-go basis, and there is no charge while the disaster recovery cluster is suspended. When users switch to the disaster recovery cluster and it is running, additional charges for the disaster recovery cluster's use of CU duration will apply. For specific fees, please refer to Billing Overview.

2. When the data engine itself is a 16 CU SparkSQL engine on a pay-as-you-go basis. After enabling the disaster recovery cluster, it remains a 16 CU SparkSQL engine on a pay-as-you-go basis, and there is no charge while the disaster recovery cluster is suspended. When users switch to the disaster recovery cluster and it is running, with the primary cluster suspended, only the fees for the disaster recovery cluster's use of CU duration will be charged.

Engine Kernel Version

Last updated : 2024-07-31 17:47:29

DLC provides different kernel versions optimized for various use cases, with numerous features and performance enhancements. The available kernel versions are listed below.

If your scenario primarily involves interactive queries, it is recommended to use the Presto engine and SparkSQL engine with the latest kernel versions.

If your scenario primarily involves batch jobs, it is recommended to use the Spark job engine with the Spark 3.2 kernel version.

Engine Type	Kernel Version	Description
Presto	SuperSQL- P 1.0	Based on the native Presto 0.242 version, this implementation supports dynamic data source loading, enhanced Dynamic Filter, Iceberg V2 tables, INSERT OVERWRITE for non-partitioned tables, and execution of Hive UDFs.
SuperSQL- S 1.0		Based on the native Spark 3.2 version, this implementation supports Iceberg 1.1.0, Hudi 0.12.0, and Adaptive Shuffle Manager.
SparkSQL SuperSQL S 3.5	SuperSQL- S 3.5	Based on the native Spark3.5 version, this implementation supports Iceberg 1.5.0 and Adaptive Shuffle Manager. The current beta version is backward compatible with various SQL and data governance tasks of SuperSQL-S 1.0, providing a performance improvement of more than 33% over the S1.0 version.
SparkBatch	Spark 3.5	Based on the native Spark3.5 version, this implementation supports Iceberg 1.5.0, Python3 and Adaptive Shuffle Manager. The current beta version is backward compatible with various SQL, jar, pyspark and data governance tasks of Spark 3.2, with a performance improvement of more than 33% over Spark 3.2.
opanibaton	Spark 3.2	Based on the original Spark3.2 version, this implementation supports Iceberg 1.1.0, Hudi 0.12.0, Python3, and Adaptive Shuffle Manager.
	Spark 2.4	Based on the native Spark2.4 version, this implementation supports Iceberg 0.13.1, Python2, and Python3.

Engine Network Configuration

Last updated : 2024-07-31 17:47:50

DLC supports configuring the network (VPC) for the data engine, facilitating the management of data engine access to different data source networks.

Network Configuration Type

Based on different business scenarios, Data Lake Computing offers two types of network configurations.

Enhanced Network Configuration: Suitable for situations requiring high-speed, stable access to data within a single VPC.

Caution

Data engines of non-Spark job types can only be bound to one Enhanced Network Configuration. Cross-origin Network Configuration: Suitable for cross-origin federated data queries requiring access to multiple VPCs. A data engine can be bound to multiple Cross-origin Network Configurations.

Network Configuration Status

Initial: The network configuration is being initialized, and the network is not yet effective. Success: The network configuration is effective for the bound engine. Failure: Network configuration failed, it can be deleted and reconfigured.

Network Configuration Security Policies

If you have configured a Security Group Policy for the VPC, inbound rules need to be added for different types of network configurations.

Enhanced Network: In the Security Group, add inbound rules for the IP range of the VPC where the data source is located.

Cross-origin Network: In the Security Group, add inbound rules for the IP range where the network configuration's bound engine is located.

Create Network Configuration

1. log in to DLC console, select the service region.

- 2. Access Engine Management> Engine Network Configuration through the left navigation menu.
- 3. Click the **Create Network Configuration** button to enter the creation page.

Create network configur	ation				×
The enhanced type is suita network configuration can The cross-source type is s several sets of cross-source	ble for the scenario v be bound to a data e uitable for cross-sour e network configurat	where a fast and stable ingine. rce federated data que ions.	e VPC is required fo	or data access. Or VPCs. A data engir	ly a set of enhanced he can be bound with
Network configuration type *	Enhanced	Cross-source			
Configuration name *	Up to 35 character	rs in Chinese characte	rs, letters, and ι		
nstance source	Data Lake Comp	oute-hosted catalog	New network	configuration	
Catalog *	Please select		▼		
Data source VPC	Select a VPC 🔻	Select a subne 🔻	🗘 0 IPs in tota	l, 0 available	
	The data engine network of the data end of th	work will connect all su in the console.	ubnets in the VPC.	If existing network	s do not meet your needs, you
3ound data engines *	Select a data engir	ne 🔻			
Configuration description	Enter configuration	n info of up to 100 cha	racters		

Configure parameters as follows:

Configuration	Required	Filling Instructions
Network Configuration Type	Yes	Select based on use case: Enhanced Network Configuration: Suitable for scenarios requiring high-speed, stable access to data within a single VPC Cross-origin Network Configuration: Suitable for scenarios involving cross- origin federated query analysis requiring access to data across multiple VPCs
Configuration Name	Yes	Supports Chinese, English, and _, with a maximum of 35 characters
Instance Source	Yes	Supports two sources: DLC data directory: You can select the data directory that has been created under DLC's Data Management New Network Configuration: Choose a new data source to create a network connection. Currently, supported data sources include MySQL, Kafka, EMR HDFS (COS, HDFS, Chdfs), PostgreSQL, SQLServer, and ClickHouse. If the data source required for the network configuration is not yet supported, select Other and manually specify the VPC
Data directory	Yes	Based on the selected instance source, choose the corresponding data directory. The range of available data directories will be related to your account



		permissions
Bind data engine	Yes	Select the data engine associated with this network configuration. If the data engine is in an isolated or initializing status, it cannot be selected
Configuration description	No	No more than 100 characters

4. Fill out and save to create a network configuration.

Caution

After creation, the network will be in an initialization state, and its status can be viewed in the list afterward.

Delete network configuration

You can manage and delete network configurations that are no longer needed or have failed to configure by deleting them. The steps are as follows:

1. DLC Console, select the service region.

2. Access Engine Management> Engine Network Configuration through the left navigation menu.

3. Find the network configuration you wish to delete. You can filter search results, but be sure to select the correct Network Configuration Type.

4. Click the **Delete** button. After a secondary confirmation, the deletion will be complete.

Caution

After deletion, the data engine will not be able to use this network configuration. If access is required, it must be reconfigured. Please proceed with caution.

Modifying description information

You can modify the description of an existing network configuration by following these steps:

1. DLC Console, select the service region.

2. Access Engine Management> Engine Network Configuration through the left navigation menu.

3. Find the network configuration you wish to delete. You can filter search results, but be sure to select the correct Network Configuration Type.

4. Click the Modify description information button to edit and modify.

Associating Tag with Private Engine Resource

Last updated : 2025-01-03 15:27:27

Overview

A tag is used to categorize and manage resources. It consists of a tag key and a tag value. A tag key can correspond to multiple values. You can create tags and bind them to cloud resources for easier management. Data Lake Compute supports binding tags to private engines in the console or on the purchase page, thereby enabling multidimensional category management and bill breakdown for private engine resources.

Creating a Tag and Binding a Resource

Create a tag and bind it to a private engine for resource categorization and unified management.

Directions

1. Log in to the Tag console to create a tag as instructed in Creating Tags and Binding Resources.

2. Log in to the Data Lake Compute console.

3. Click SuperSQL Engine on the left sidebar to enter the Data engine list page.

4. Click a resource name to enter the resource details page. Click **Edit** to pop up the tag edit window and select a tag for binding.

Basic info		Configuration info	Set start and stop policy	Change spec configuratio
Engine name document_test i Resource ID DataEngine-44ncfc7ni		Engine type SparkSQL Kernel version	SuperSQL-S 1.0 Engine Size 16	CU Cluster count 1
Description None 🖍		Auto-scaling Yes Elastic cluster count	1 Max task concurrency of a clus	iter 5
Region Hong Kong/Macao/TaiWan (China Region)-Hong Kong Engine Status Starting		Task queue-up time limit 0 minute(s)		
Billing mode Pay-as-you-go Tag No tac ? Tags are used to categorize resources. To learn more, see	Edit tag	ion No X 6 bu can Labels [2	Timing policy None	
	Tag Key Tag	Value 💌 🗙		
Failover cluster	+ Add () Paste	ion		
Backup cluster name document_test_backling Backup resource ID D		el version	SuperSQL-S 1.0 Engine Size 16	CU Cluster count 1
Engine Status Suspend 🧔	Confirm	Cancel ster count	1 Max task concurrency of a clus	iter 5
Billing mode Pay-as-you-go		Table answer on these limits () animits(a)		

5. Click **Confirm** to bind the tag to the private engine. You can click **Edit** again to unbind or modify the tag.

Basic info	Configuration info Set start and stop policy Change spec configuration					
Engine name at_data_engine_prestol Resource ID DataEngine-p3d2xfq1	Engine type Presto Kernel version SuperSQL-P 1.0 Engine Size 16 CU					
Description autotest_presto_engine Region Hong Kong/Macao/TaiWan (China Region)-Hong Kong	Cluster count 1					
Engine Status Starting 🧔	Auto-scaling Yes Elastic cluster count 4 Max task concurrency of a cluster 20					
Billing mode Pay-as-you-go	Task queue-up time limit 0 minute(s)					
Tags are used to categorize resources. To learn more, see Tag Documentation	Auto-start Yes Auto-suspension No Timing policy None					
	IP range of cluster 10.255.252.0/22					
	Network configuration					

Binding a Tag on the Purchase Page

You can bind a tag when purchasing a private engine resource in both monthly subscription and pay-as-you-go billing modes.

Info configura	ation						
Resource name	Enter a name						
	It can contain up to 100 Ch	inese characters, lette	ers, digits, hyphens (-) an	nd underscores (_) o	only. A duplicate name	is not allowed.	
Description	Up to 250 characters						
	Optional, up to 250 charac	ters.					
Тад							
	Tag Key	•	Tag Value		7	Delete	
			+ Ad	d			
			() Pas	te			
	ОК	Cancel					
	Tags are used to categorize	e resources. To learn r	nore, see Tag Document a	ation 🛂			
Terms of agreement	I have read and agree	ee to the Service Le	vel Agreement for Dat	ta Lake Compute	andRefund Policy	ß	

Filtering Resources by Tag

You can filter resources by tag on the **SuperSQL Engine** page in the Data Lake Compute console.

Directions

- 1. Log in to the Data Lake Compute console and select **SuperSQL Engine**.
- 2. Select a tag in the tag search box. You can filter resources by tag key or tag key-value.

Create resource Bill query E	Renewal manager	ment 🖸	Tag	: Select a res	source tag or enter k	(eywo	rd(s) (separate two keywords with	a " ")	_		
Engine Name/ID	Engine type	Engine Status	Kernel	/ test	() Paste	•	123 💌	×	op policy	Cluster descripti	Operation
DataEngine-q6s3yxx+1	Presto	Suspend (j)	SuperS	OK	Cancel				Auto-suspension	Private engine	Monitor Presto UI Spec configuration Parameter Configuration More ▼

Create resource	Bill query 🖄 Renewal mar	nagement 🗹	Tag - tag key only	Select a resource tag or enter keyword(e) (ser	arate two keywords with a " ")		
Engine Name/ID	Engine type	Engine Status	Kernel version	Tag Key V X	Start and stop policy	Cluster descript	Operation
æ	Presto	Suspend (i)	SuperSQL-P 1.0	OK Cancel	Auto-start, Auto-suspension	Private engine	Monitor Presto UI Spec configuration Parameter Configuration More ▼

3. Click the search icon to get the list of engines with that tag.

Create resource Bill query	Renewal manager	ment 🛂	Tag - tag key or	nly: test 😒 Select a re	source tag or enter keyword	d(s) (separate two keywords with a '	")	8
Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Cluster description	Operation
at_data_engine_presto F DataEngine-p3d2xfq1 F	Presto	Starting (i)	SuperSQL-P 1.0	Pay-as-you-go		Auto-start, Manual suspension	Private engine	Monitor Spec configuration Parameter Configuration More ▼
Total items: 1						1	0 ▼ / page 🛛 🖣	1 /1 page

Allocating Costs by Tag

You can bind tags in the organization or business dimension for cost allocation by department, project team, region, etc.

Directions

1. Log in to the Tag console and create a tag.

2. Bind the tag to an engine resource in the tag console, on the **SuperSQL Engine** page in the Data Lake Compute console, or on the purchase page.

3. Go to the Billing Center to set a cost allocation tag. For more information, see Cost Allocation Tags.

4. Go to the Bill Overview page, select the aggregation by tag tab, and view the column chart and list of resources aggregated by tag key.

Engine Local Cache

Last updated : 2024-07-31 17:48:05

To ensure stable operation of Spark engine query analysis when network bandwidth is limited (e.g. during storage system throttling), the DLC Spark engine provides a local cache capability. When you need to cache table data, you can quickly enable caching by adding engine configuration.

Directions

1. Create a Spark Engine: For details, see Purchase Exclusive Data Engine.

2. Add Cache Configuration: Go to the DLC Console > Data Engine. Select the engine created in Step 1, click
 Parameter Configuration, and add the configuration items from Cache Configuration Item Explanation.
 Spark SQL Engine Configuration:

uperSQL引擎								
 数据引擎包括独享数据等 动挂起或定时挂起策略, 	引擎与共享数据引擎。p 挂起后将不会产生任(oublic-engine (共享引擎) 可费用,操作步骤及注意事	按扫描量计费,由新 项可参见 管理独享	系统进行管理,无需操作 数据引擎 ^[2]	() 修改引	擎参数配置将需要重启集群.		
创建资源 订单查询 🖸	续费管理 🖸				数据加密 🛈			
逐源名称/ID	引擎类型	内核版本 ()	运行状态	付费类型	参数配置	1 spark.hadoop.fs.cosn.impl	alluxio.hadoop.ShimFileSystem	-
biy 🔂	SparkSQL	SuperSQL-S 1.0	运行	按量计费		+添加		
qb901960 1	SparkSQL	SuperSQL-S 1.0	运行	按量计费				
_8km	Presto	SuperSQL-P 0.1	运行	按量计费				
	Spark作业		运行					
, , d	SparkSQL	-	运行	-				

Note:

After the configuration is added, the engine cluster will restart. It is recommended to enable the cache when no tasks are running to avoid affecting ongoing tasks.

3. To use the engine cache, go to Data Exploration, write the query SQL in the SQL interface, select the engine with the cache enabled, and execute the SQL. Once executed, the engine will cache the DLC external table data locally. When the SQL is executed again, the data will be fetched from the local cache, improving query efficiency. **Spark SQL Engine Query:**

	2-0 1.0)
运行历史	下载历史
	运行历史

Spark Batch Engine Query:

	$\leftarrow \rightarrow$ [] \models		Spark 3.2
1 set spark.ha 2 select test1 3 left join Da 4	doop.fs.cosn.impl=alluxio.hadoop.ShimFileSystem; .id,test1.name,test2.age from DataLakeCatalog.test_cry.h_test1 test1 taLakeCatalog.test_cry.h_test2 test2 on test1.id = test2.id		
查询结果 ————————————————————————————————————			下载历史
道道结果	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 Г		下载历史
直询结果	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 To 点击查看集群日志		下载历史
道询结果	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 盾 点击查看集群日志 ExecuteSQL: select test1.id,test1.name,test2.age from DataLakeCatalog.test_cry.h_test1 test1 left join DataLakeCatalog.test_cry.h_	_test2 test2 on test1.id = test2.id	下载历史
查询结果 2023-11-28 15:05:27	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 匠 点击查看集群日志 ExecuteSQL: select test1.id,test1.name,test2.age from DataLakeCatalog.test_cry.h_test1 test1 left join DataLakeCatalog.test_cry.h_ 当前任务状态: available 请等待	test2 test2 on test1.id = test2.id	下载历史
询结果 2023-11-28 15:05:27 2023-11-28 15:05:27	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 盾 点击查看集群日志 ExecuteSQL: select test1.id,test1.name,test2.age from DataLakeCatalog.test_cry.h_test1 test1 left join DataLakeCatalog.test_cry.h_ 当前任务状态: available 请等待 当前任务运行成功,点击查看运行结果	_test2 test2 on test1.id = test2.id	下载历史
询结果 2023-11-28 15:05:27 2023-11-28 15:05:27 2023-11-28 15:05:29	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 戶 点击查看集群日志 ExecuteSQL: select test1.id,test1.name,test2.age from DataLakeCatalog.test_cry.h_test1 test1 left join DataLakeCatalog.test_cry.h_ 当前任务状态: available 请等待 当前任务运行成功,点击查看运行结果 任务运行结束	test2 test2 on test1.id = test2.id	下载历史
询结果 2023-11-28 15:05:27 2023-11-28 15:05:27 2023-11-28 15:05:29 Task ID	TaskID: fdd1f66b-10f6-402a-b5a2-8e7af8c618c0 戶 点击查看集群日志 ExecuteSQL: select test1.id,test1.name,test2.age from DataLakeCatalog.test_cry.h_test1 test1 left join DataLakeCatalog.test_cry.h_ 当前任务状态: available 请等待 当前任务运行成功,点击查看运行结果 任务运行结束 SQL 开始时间 运行时长(_test2 test2 on test1.id = test2.id	下载历史 「 操作

Cache Description

Cache Configuration Items Description

Configuration Items	Configuration Values	Configuration Items Description

spark.hadoop.fs.cosn.impl alluxio.hadoop.ShimFileSystem Fixed v cache value to cache other th being a follow t If you r enablir item.	value; the configuration value is the implementation class. Configure this o enable the cache feature. If the feature is enabled, configuring a value han this will result in the engine not able to access COS data. Please the instructions carefully. need to disable the cache after ng it, please delete this configuration
--	--

Cache Usage Instructions

1. Engine Type Description

SparkSQL Engine: When the engine restarts, the cached data becomes invalid because it is a local cache.

SparkBatch Engine: The SparkBatch engine runs tasks at the session level. Once the task execution is complete, the cached data becomes invalid.

2. Table Type Description

Currently, only DLC external tables are cached.

Custom Task Scheduling Pool

Last updated : 2024-07-31 17:48:18

Application scenario

Applicable Engine: Spark SQL Engine.

When you submit multiple tasks to the engine, for example, submitting multiple SQL tasks to the Spark SQL cluster simultaneously, the tasks submitted by the business may have dependencies, so the engine will default to scheduling these tasks in a FIFO manner when scheduling and executing.

However, in some special cases, you may need to define the priorities of certain tasks yourself, for example in the following scenario:

The submitted task has a high priority and needs to be executed with the highest priority, not wanting it to queue for cluster resources.

The submitted task has a low priority, hoping that it will not preempt resources from other tasks as much as possible. It will be executed when resources are available, and it will queue when resources are not.

Customize Scheduling Rules

In the Spark SQL Engine, each executed SQL task Job is split into a collection of multiple tasks, TaskSet, and our scheduling is based on TaskSet. Whenever the cluster has idle resources, it takes a Task from all Job's TaskSet according to the scheduling algorithm for dispatch execution.

Our scheduling algorithm is to define multiple scheduling pools, placing Job/TaskSet in the corresponding scheduling pool, and obtaining the Task that needs to be dispatched for execution according to the scheduling pool.

Scheduling Pool and Its Attributes

You can define multiple scheduling pools, each with four attributes:

name: The name of the scheduling pool, which you can name yourself. It can be named default, indicating the default scheduling pool.

schedulingMode: The scheduling rule, supporting two modes: FIFO and FAIR. The scheduling algorithm when there are multiple TaskSets within a scheduling pool.

FIFO: Tasks are dispatched in the order that TaskSets are submitted.

FAIR: Tasks from multiple TaskSets are dispatched fairly. The specific dispatch rules are related to the minShare and weight attributes of the scheduling pool.

minShare: The minimum number of cores required, must be greater than 0, that is, the minimum number of Tasks that can run. During scheduling, priority is given to the number of Tasks running in the scheduling pool reaching minShare.

weight: The weight. Scheduling pools with a higher weight will have their Tasks prioritized. Weight comparison will only occur after minShare is met.

The scheduling configuration requires you to write an xml file, in the following formats:

```
<?xml version="1.0"?>
<allocations>
<pool name="production">
<schedulingMode>FAIR</schedulingMode>
<weight>1</weight>
<minShare>2</minShare>
</pool>
<pool name="test">
<schedulingMode>FIFO</schedulingMode>
<weight>2</weight>
<minShare>3</minShare>
</pool>
</allocations>
```

Scheduling Configuration Reference Example

You can refer to the settings for three scheduling pools:

Default Scheduling Pool default:schedulingMode = FIFO, weight = 1, minShare = (Cluster Cores - Driver Cores). This scheduling pool is the default submission pool for tasks, with ordinary priority. Execution is in sequential order, and it can utilize all of the cluster's computing resources.

Slow Task Scheduling Pool straggler:schedulingMode = FAIR, weight = 1, minShare = 1. This scheduling pool is dedicated to slow task submissions, with ordinary priority. Since minShare = 1, it does not preempt resources from tasks submitted to the default pool. Tasks in the straggler scheduling pool are executed when the cluster has more available resources.

High Priority Scheduling Pool special:schedulingMode = FIFO, weight = 1000, minShare = (Cluster Cores - Driver Cores). This scheduling pool is for tasks that need priority execution in special circumstances. However, due to the presence of minShare, this pool does not monopolize all cluster resources. Tasks in both the default and special pools continue to be executed, typically dispatching an equal number of Tasks from each pool.

Taking a 16CU cluster (with the driver being 4CU) as an example, the configuration for this reference example is as follows:

```
<?xml version="1.0"?>
<allocations>
  <pool name="default">
    <schedulingMode>FIFO</schedulingMode>
    <weight>1</weight>
    <minShare>12</minShare>
  </pool>
  <pool name="straggler">
```



```
<schedulingMode>FAIR</schedulingMode>
<weight>1</weight>
<minShare>1</minShare>
</pool>
<pool name="special">
<schedulingMode>FIFO</schedulingMode>
<weight>1000</weight>
<minShare>12</minShare>
</pool>
</allocations>
```

Operation method

1. After preparing the xml file for the scheduling pool, place it in a path on cos, for example cosn://bucketappid/fairscheduler.xml.

2. Add the following configuration in the engine settings.

Data Lake Compute	SuperSQL engine	🛇 Hong Kong 🔻						SuperSQL eng
E Overview	Data Lake Compute offer required; a private data	ers both public and priva engine can be billed on a	te data engines. A p a pay-as-you-go bas	ublic data engine is m sis or subscribed mont	anaged by Data Lake Computitive State Stat	te and billed by scanned Billing Overview 🗹 . A	l data volume, with no operati bay-as-you-go data engine ca	on or permission an be configured with
📃 Data Scheduling	the auto-suspension or	scheduled suspension p	blicy, with no tees c	narged on it after susp	ension. For operations and no	ntes, see managing Priv	ate Data Engines 🗠 .	
∃≟ Data Management	Create resource Bill quer	y Renewal manage	ment			Select a resource	ce tag or enter keyword(s) (se	parate two
🗐 Data Job	Engine Name/ID	Engine type	Engine Status	Kernel version	Billing mode	Auto-renewal	Start and stop policy	Operation
 Task History Insight Management 	1 ☐ DataEngine-ksyfgcnl 1 ☐	Spark job	Starting (i)	Spark 3.2	Pay-as-you-go	-	Manual start, Manual suspension	Monitor Spec configurati Parameter Configuration More ▼
Engine Management SuperSQL Engine Standard Engine BETA	1717	Presto	Suspend (1)	SuperSQL-P 1.0	Pay-as-you-go	-	Auto-start, Auto-suspens	Monitor Presto UI Spec configuration Parameter Configuration More ▼
Ops Management		Spark job	Running	Spark 3.2	Pay-as-you-go		Auto-start, Manual suspe	Monitor Spec configurati Parameter Configuration More ▼
Management Storage Configuration	90%	SparkSQL	Suspend (j)	SuperSQL-S 1.0	Pay-as-you-go		Auto-start, Auto-suspens	Monitor Spec configurati Parameter Configuration More ▼
 Audit Log Monitoring & Alerting Id 		SparkSQL	Running	SuperSQL-S 3.5	Monthly subscription 2024-08-02 11:37:06 Expire	No	Manual start, Manual suspension	Monitor Spec configurati Parameter Configuration More ▼

Parameter configuration spark.scheduler.allocation.file, set to the path of your scheduling pool xml file cosn://bucketappid/fairscheduler.xml.



Configuration change	
If engine paramet	er configurations are changed, you must restart the cluster to apply the new configurations
Data encryption (j)	
Parameter Configuration	1 sqark.scheduler.allocation.file cc
	+ Add

This operation requires restarting the cluster.

3. When submitting a task, specify the following parameters as task parameters: spark.scheduler.pool = the name of the scheduling pool to submit to. If it is the default scheduling pool, it does not need to be specified.

	🗄 🛯 🔹 🖬 🖬	n étan yaka	an Jacob State Ba
1 r* >=+ 'stin 's. AC-12* 2	🗄 Data engine	Refresh	und gesch kannen juli andere kannen som
3 Martin Contractor Contractor	n filologia	Shared <	
	The engine supportsSuperSQL Syntaxquery.		
	Sensine (kernel version) Different kernel versions sur rules. For details, see Kernel Versions.	oport different SQL syntax	
	1979 - 2088 A.C. 1970 - 197	▼	
	Create engine		
	Execution mode Complete m 🔻		
Categoria (Advanced settings Co	nfiguration description 🛽	ownload history
р С блан бори. Х болуна, па балаанын коле	1 Select configuration. More V	-	

Notes

Scheduling occurs at the time node when: the cluster has idle resources and there is a task that needs scheduling. Therefore, if the cluster is already fully occupied by a task, for example, a slow task, it must wait for one Task of that task to be completed before beginning to schedule other tasks with higher priority. Therefore, it is important to note



that the time consumption of a single Task of a slow task should be relatively reasonable; otherwise, it might still lead to long periods of occupying cluster resources.

Standard Engine Introduction of the Standard Engine System

Last updated : 2024-09-04 11:12:04

The Standard Engine system is composed of several key components: Engine Network, Gateway, Resource Group, Endpoint, and Executor. Before using the DLC Standard Engine, you should understand these concepts:

Public netwo	ork access(DLC JDBC、TencentCloue	dAPI) onnection rate limit 20QPS learn more 🗳	Engine network: name 10.255.0.0/16	3		Delete engine
	User VPC	Private connection (i) 1 Details	E Gateway Running	Details	Standard engine	
			2 cu		2	
			Engine network: name 10.255.0.0/16	3		Delete engine
Public network		Private connection (i) 0 Create	🗎 Gateway	Details	Standard engine	
access	Submission machine		None		0	
			Engine network: name 10.255.0.0/16	3		Delete engine
		Private connection () 3 Details	📋 Gateway	Details	Standard engine	
			8 cu		6	

Concepts

The table below provides a brief introduction to several key concepts within the Standard Engine system. For more detailed information, you can click the relevant links.

Concept	Description
Engine Network	The Engine Network is a managed private connection that deploys the gateway and the Standard Engine within a logically isolated network environment. Users can customize the IP address range and subnet of the Engine Network according to their business needs.
Gateway	The gateway, implemented based on the Kyuubi big data component, serves as the access point for the Standard Engine services, providing users with a more efficient and stable task submission experience.
Standard Engine	The Standard Engine is a type of computing resource provided by DLC that helps users quickly launch compute clusters of a certain scale. It offers comprehensive support for native syntax and behavior, allowing users familiar with the big data ecosystem to get started more quickly and use the system with ease.
Resource Group	The Standard Spark Engine supports further on-demand division of engine resources through the use of resource groups. A resource group is a collection of a portion of the Standard Spark Engine's computing resources and corresponding configurations. SQL tasks can be submitted to a designated resource group for execution.



Private Link	Through a private connection, users can establish a link between their account's VPC and the Standard Engine's network, allowing tasks to be submitted via servers within that VPC.
Executor	After an endpoint is created, any server within the user account's VPC associated with that endpoint can serve as an executor for task submissions.

Task Submission Methods

Users can submit tasks in various ways:

- 1. Through JDBC on the executor, as shown in the diagram.
- 2. Submit SQL tasks via the Data Exploration page in the DLC console.
- 3. Submit Spark batch and streaming jobs via the Data Jobs page in the DLC console.
- 4. Submit tasks through the TencentCloud API.

Quick Purchase and Configuration of the Standard Engine

1. If you are purchasing the Standard Engine for the first time, DLC recommends following the Standard Engine Configuration Guide in the documentation to quickly set up the Standard Engine.

2. Once the purchase is completed, you can submit tasks via the Data Exploration page or the executor.

Standard Engine Introduction

Last updated : 2024-09-04 11:13:49

The Standard Engine is a type of computing resource provided by DLC that helps users quickly launch compute clusters of a certain scale. It offers comprehensive support for native syntax and behavior, enabling users who are familiar with the big data ecosystem to get started quickly and use it with ease.

Types of Standard Engine

Users can choose different Standard Engine kernels based on their needs to address various use cases. The Standard Engine is divided into the following types:

Spark: Suitable for stable and efficient offline SQL tasks, as well as native Spark streaming/batch data processing jobs.

Presto: Suitable for agile and rapid interactive query analysis.

Gateway: The Gateway is a special type of Standard Engine implemented based on native Kyuubi. The Gateway is used to connect users to the Spark/Presto computing engines and submit tasks, serving as a prerequisite for using other computing engines.

Note:

Different types of engines do not affect the unit price of engine billing. For detailed pricing information, see the Billing Overview.

Engine Elasticity

Currently, only the annual subscription Spark Standard Engine supports the configuration of pay-as-you-go for resource elasticity.

Elastic Cluster Specifications	
	Elastic resource rate - 16 + Cu
	Elastic cluster specifications indicate that the cluster specifications are dynamically scaled based on the task load as long as they do not exceed the resident cluster specifications, and the billing mode pav-as-you-go. For example, if the resident cluster specifications are 256 CUs, the maximum elastic cluster specifications are 256 CUs.

As shown in the diagram, tasks and resource groups will prioritize using the resources from the monthly or annual subscription. If a user's submitted task exhausts the resources from this subscription, any subsequent tasks will automatically use the configured pay-as-you-go elastic resources. In the diagram, after Task 03 depletes the subscription resources, it continues to use the pay-as-you-go resources.





Note:

1. Pay-as-you-go elastic resources are charged based on the actual computing resources used.

2. If a task or resource group is scheduled to use pay-as-you-go resources, it will continue to use those resources even if the monthly or annual subscription resources are later freed up. The resource group will only be rescheduled to use the subscription resources after it has been restarted.

3. A single Spark Standard Engine cannot set elastic resources exceeding the amount of resources in the annual or monthly subscription. For example, a 128 CU annual or monthly subscription engine can set up to 128 CU of elastic resources. If you need to configure more elastic resources, contact us through a ticket.

Standard Engine Terminology

Terminology	Description
Cluster Type	When purchasing a Standard Spark Engine, you can choose the cluster type. The standard type is 1 CU \approx 1 core with 4 GB memory, and the memory type is 1 CU \approx 1 core with 8 GB memory. Different types have different unit prices. For more details, see the Billing Overview.

Elastic Cluster Specifications	The monthly or annual subscription Spark Engine allows users to configure elastic specifications. Once the resources from the subscription package are exhausted, the system will automatically allocate pay-as-you-go resources based on user configuration.
Gateway Name	The name of the gateway must be globally unique. It cannot share the same name as any other gateway or compute engine.
Engine Name	The name of the engine must be globally unique. It cannot share the same name as any other gateway or compute engine.
Engine Type	The Standard Engine types are categorized into Presto Engine and Spark Engine. The gateway is also a special type of Standard Engine.
Engine Status	The status of the Standard Engine varies based on the current operation of the cluster. The statuses include: Starting, Running, Ready, Paused, Pausing, Modifying, Isolated, Isolating, and Recovering. Starting: The cluster resources are being initiated. Pay-as-you-go for the engine does not occur during this time. Clusters in the starting status cannot be selected for data computation tasks. Running: The cluster is running and can be selected for data computation tasks. Ready: Similar to the running status, this status indicates that the engine is available for use. Paused: The cluster is paused and cannot be selected for data computation tasks. Pausing: The cluster is in the process of switching to the paused status. This transition may affect any running tasks, and the cluster cannot be selected for data computation during this time. Modifying: The cluster is undergoing configuration changes. During this period, it cannot be selected for data computation tasks. Isolated: The cluster has been isolated due to account arrears and cannot be selected for data computation tasks. Isolated: The cluster is in the process of being isolated due to account arrears. This transition may affect any running tasks, and the cluster cannot be selected for data computation during this time. Recovering: The cluster is in the process of being isolated due to account arrears. This transition may affect any running tasks, and the cluster cannot be selected for data computation during this time.
Resource Group Count	The current number of resource groups under the Standard Spark Engine.
Used Resources / Total Resources	The quantity of resources currently used by the engine and the total available resources of the engine. The total resource count includes both the persistent resources and the elastic resources. Used resources include those occupied by the DLC deployment service system. There may be some delay in the reported data.
Payment Type	Payment types include annual/monthly subscription and pay-as-you-go. The gateway only supports the annual/monthly subscription model.

	The Standard Spark and Presto engines support both annual/monthly subscription and pay- as-you-go.
Auto- Renewal	Indicates whether the monthly or annual subscription engine will automatically renew as it approaches expiration.
Engine Size	 The total available resources of the engine, measured in CUs. For monthly or annual subscription engines, the size includes both the engine's persistent capacity and the elastic capacity billed on a pay-as-you-go basis. Note: For monthly or annual subscription engines, a one-time payment is required at the time of purchase. The engine's status does not affect billing costs. For pay-as-you-go engines, charges are based on the user's usage: The Standard Presto Engine incurs charges while running, but not when suspended. Some costs may be incurred during the engine's startup phase. The Standard Spark Engine does not incur charges while in a ready status. Costs are only incurred when tasks are submitted or when a resource group is started and running.

Standard Engine Kernel Versions

Last updated : 2024-09-04 11:14:22

The kernel versions used by the DLC Standard Engine are described as follows:

Engine Type	Kernel Version	Description
Spark	Standard-S 1.0	Standard-S 1.0 is a self-developed engine kernel based on Spark 3.2, compatible with native Spark syntax and behavior, and suitable for offline SQL tasks. It also supports Iceberg 1.1.0, Hudi 0.12.0, and Python 3, and includes support for Adaptive Shuffle Manager.
Presto	Standard-P 1.0	Standard-P 1.0 is a self-developed engine kernel based on Presto 0.242, compatible with native Presto syntax and behavior, and suitable for interactive query analysis. It also supports dynamic data source loading, enhanced Dynamic Filtering, Iceberg V2 tables, INSERT OVERWRITE for non-partitioned tables, and the execution of Hive UDFs.

Standard Engine Parameter Configuration

Last updated : 2025-03-12 18:03:39

Spark parameters are used to configure and optimize settings for Apache Spark applications.

In a self-built Spark, these parameters can be set through command line options, configuration files, or programmatically.

In the DLC standard engine, you can set Spark parameters on the engine, which will take effect when users submit Spark jobs or submit interactive SQL using custom configurations.

Note:

- 1. The standard engine dimension configuration only takes effect for Spark jobs and Batch SQL tasks.
- 2. Only after the engine dimension configuration is added will the new tasks take effect.

Setting Standard Spark Engine Parameters

- 1. Enter the standard engine feature.
- 2. Select the engine that needs to be configured on the list page.
- 3. Click Parameter Configuration , and the engine parameter side window pops up.
- 4. In "Parameter Configuration", click Add , add the target configuration and then click Confirm .

Data Lake Compute	Standard engine	S Beijing Finance 🗸						
Overview Council Data Explore	Overview	Pi	rivate Link 2	Gateway 2cu S	Suspend	Standard	engine 3 Cre	ate engine network 🗘 U
Data Scheduling Data Data Management	A standard pay-as-yo	u-go Spark engine is charg	ed based on running task	s and resource groups. When no tasks or	resource groups are runnir	g, no fee is charged. <mark>Billing</mark>	Overview 🗹	
E Data Job	Create resource	Bill query Renew	al management				Enter a keyword	C
Task History	Engine Name/ID	Engine type T	Engine Status T	Engine Network Name/ID T	Resource Groups	Used Resources/T	Access link	Operation
 Insight Management Engine Management 	_service DataEngine-rggb1ltt	Standard Spark	Ready	farley_t271 DataEngine-Network-Odzxigkb	2	2/32	HiveJDBC jdbc:hive2://192.168.100.8:10009/?spark.engine= jdbc:hive2://172.16.0.6:10009/?spark.engine=sL_	ESI_5 Manage Resource Grou Serv Spec configuration
SuperSQL Engine							DLCJDBC jdbc:dlc:dlc.tencentcloudapi.com?task_tvp	Parameter Configuratio

Resource Group Dimension Parameters

Parameters of Resource Group for SQL Analysis Only Scenario

Adding Parameters When a Resource Group Is Created

When a resource group is created, select SQL analysis only and add parameters in the Parameter Management at the bottom.

Note:

1. Static parameters can only take effect after the resource group restarts, while dynamic parameters do not require a restart of the resource group to take effect.

2. For details on dynamic parameters and static parameters, see the official website of Spark.

3. The configuration of resource group for SQL analysis only scenario takes effect only when SQL tasks are run using that resource group.



Modifying Resource Group Parameters

1. In Standard Engine List Page , select the engine to be modified and click Enter .

2. On the resource group management page, select a resource group for SQL analysis only scenario and click the

Details button.

Data Lake Compute	← Standard engine / t	272_spark							
B Overview	Basic configuration	Cluster monitoring	Resource Group Manage	ement					Alarm con
Data Explore	• The pay-as-you-go engin	e is charged based on runr	ning resource groups. No fee is	charged when the resource	groups are suspended. For det	ails, see Billing Overview 🛙	1		
📃 Data Scheduling	A resource group is a gen groups. A resource group	eral term for a group of co can remain resident after i	mputing resources and corresp it is started to reduce the cold	oonding configurations. You o startup time.	can classify resources of the Sp	oark engine into different res	ource groups as required, a	and SQL tasks can run in	specified resource
그는 Data Management	Create Resource Group	Batch Restart	Batch Start Batch Susp	bend	Execution stat	tus Select	▼ Resource Gr	oup Name En K the ke	nyword of the re
🗉 Data Job									
Task History	Resource Group Nam	Resource group statu	IS Scenario T	Engine Name/Status	Gateway Name/Status	Resource Specification	Resource Group Usag.	Creator	Operation
Insight Management	only_sql rg-4xrhjdgvbh	Running	SQL analysis only	t272_spark Ready	default-gateway-ka755 Running	2CU Custom configuration	CU	100010537383	Details Start Rest Suspend Terminat

3. On the details page, click **Edit** in the parameter management panel to add parameters or modify and delete added parameters. Similarly, static parameters can only take effect after the resource group is restarted, while dynamic



parameters do not require a restart of the resource group to take effect.

Overview						
Data Explore	Basic info		Edit	Enabling/Disablir	ng Policy 🚯	
Data Scheduling	Resource Group Name	only_sql		Auto-start	Enable	
Data	Binding Engine	t272_spark		Auto-suspension	Not enabled	
Management	Scenario	SQL analysis only		Concurrent Tasks	5	
Data Job		It supports resource allocation for tasks in SQL query analysis scenarios.				
Task History	Resource configuratio	n		Network configu	ration	
Insight Management	Configuration Method	Custom configuration		Cross-source netwo	rk configuration	All cross-source network configurations bound to the data engine will take and do not need to be selected.
ine Management	Executor resource	small(1CU)		Enhanced network c	onfiguration	
SuperSQL Engine	Executor count	1				
Standard Engine	Driver resource	small(1CU)				
Engine Network						
Configuration	Parameter Manageme	nt				ſ
Allanagament						

4. After the modification is completed, click **Save**. Then you can choose Restart Now, or you can choose Not Restart and Save Only and then restart the resource group at an appropriate time later to make the configuration take effect.

😚 Standard Engine				
 Engine Network Configuration 	Parameter Management			
Ops Management	Modified static parameters take effect only a	fter the cluster is restarted.		
ở Permission Management	The configuration is updated and takes effect after the cluster is restarted.			
Storage Configuration	To make the changes take effect immediately, restart the cluster now. Alternatively, you can manually restart the cluster later to make the changes take effect.		1	
🗎 Audit Log	Restart Now Not Restart and Save Only			
Monitoring & Alerting	Save Cancel			

Resource Group Parameters for AI (Machine Learning) Scenario

Note:

1. Currently, only the Spark MLlib-type AI resource groups support adding configurations.

2. Currently, only static configurations can be added, which only take effect on new notebook sessions and do not take effect on existing sessions.

3. The AI Resource Group feature is a whitelist feature. To ensure that it meets your usage scenarios, please submit

a ticket contact us for assessment and enablement.

4. This resource group only supports Standard Spark engine Standard-S version 1.1.

Adding Parameters When an AI Resource Group Is Created

As shown in the figure below, select the Spark MLlib type when the AI resource group is created, and choose to add parameters in the Parameter Management panel at the bottom.
Data Lake Compute	← Standard engine / t2	272_spark				Create Resource Group	
E Overview	Basic configuration C	Cluster monitoring	esource Group Manag	ement		Framework Type	ML open-source framework Python Spark MLlib
Data Explore	• The pay-as-you-go engine	is charged based on running	resource groups. No fee is	s charged when the resource group	ps a		Execute tasks using Pyspark.
Data Scheduling	A resource group is a gene groups. A resource group of	ral term for a group of compu can remain resident after it is	uting resources and corresp started to reduce the cold	ponding configurations. You can cl startup time.	lass	Built-in image	O Built-in image Custom image
∃≟⊨ Data							Please select
Management	Create Resource Group	Batch Restart Bate	ch Start Batch Sus	pend			Select an image as the default image, so that this image will be used by default durin resource group execution. You car also change the images during execution.
🗐 Data Job							
Task History	Resource Group Nam	Resource group status	Scenario T	Engine Name/Status	Gat	Resource configuration	
🛷 Insight	only_sql	Running	SQL analysis only	t272_spark	defa	Resource Group Usage Limit •	- 16 + CU
Management	ig-anijugvon			Ready	Run		1 CU is about 1-core CPU and 4 GB of memory.
Engine Management	pythonti	Ready	Machine learning	t272_spark	defa	Configuration Method	Quick configuration Custom configuration
SuperSQL Engine	19-11440002297			Ready	Run	Resource Group Specifications	
Standard Engine	sparkml2	Initialize	SQL analysis only	t272_spark	defa	Repeated croup opcontentions	
Engine Network	1g-121Wyoreina			Reauy	Run		The resource group specification specified currently may differ from the actual specified currently may differ from the
Configuration	sparkml	Ready	Machine learning	t272_spark	defa		For configuration details, see Resource Group Specifications
Ops Management	Ig-zusikk/ivs			Ready	Run	Network configuration	
2 Permission	default-rg-kvgl1m04zp	Suspend	SQL analysis only	t272_spark	defa		
Management	rg-kvgtim04zp			Ready	Run	Cross-source network configuration	All cross-source network configurations bound to the data engine will take effect and need to be selected.
Storage	Total items: 5					Enhanced network configuration	bind_DataEngine-a4wvak8f_eg_default
Configuration	Berennenenenen						Select up to one enhanced network configuration if needed.
🗄 Audit Log							
Monitoring &						Parameter Management	
Alerting						Static Parameters 🛈	Add

Modifying AI Resource Group Parameters

1. In Standard Engine List Page , select the engine to be modified and click Enter.

2. On the resource group management page, select a Spark MLlib resource group and click **Details**.

3. On the details page, click **Edit** in the parameter management panel to add parameters or modify and delete the added parameters. Note that the modified parameters only take effect on the notebook session pulled after modification, and do not take effect on the existing sessions.

Overview	Basic info	Edit	t	Network configuration	
Data Explore	Resource Group Name	sparkml		Cross-source network configuration	All cross-source network configurations bound to the data engine will take ef
Data Scheduling	Binding Engine	t272_spark			and do not need to be selected.
Data Management	Scenario	Machine learning		Enhanced network configuration	
Data Job		Allocating resources to the tasks using Python, ML frameworks, or Pyspark to train the Al models .			
Task History	Framework Type	Spark MLlib			
Insight Management	Built-in image	Built-in image			
ine Management		Select an image as the default image, so that this image will be used by default during resource group execution. You can also change the images during execution.			
SuperSQL Engine					
Standard Engine Engine Network Configuration	Resource Configuration	16CU 1 CU is about 1-core CPU and 4 GB of memory.			
Managament	Configuration Method	Custom configuration			
Permission	Executor resource	small(1CU)			
Management	Executor count	1			
Storage Configuration	Driver resource	small(1CU)			
Audit Log					
Monitoring &	Parameter Management				

Data Exploration Parameters

Note:

1. Currently, only the resource group for SQL analysis only scenario supports adding parameters on the Data Explore page.

2. Note that only dynamic Spark configurations take effect in the subsequent executions against SQL, and static parameters cannot take effect.

3. The parameter configuration at the data exploration level is of higher priority than that at the engine level and resource group level.

As shown in the figure below, on the Data Explore page, select the Standard-Spark engine for Data engine, select the option Select Resource Group for Resource configuration, and click Advanced settings on the page to add configurations.



As shown in the figure below, you can select a built-in configuration or enter the configuration manually.

sl_service			Standar	d-Spark 🔻
The engine supportsS	tandard-Spark Sy	ntaxquery	, Viewing Syntax D	escription 🗹
Engine (kernel ve syntax rules. For det	ersion) Different ails, see Kernel V	kernel ve /ersions.	rsions support dif	erent SQL
Standard Spark (Sta	indard-S 1.1)			-
🕀 Create engine				
Advanced settings 🔺	s.		Configuration d	escription 🛽
1 kyuubi.operati	on.result.saveToFi	le.er 🔻	Value	_
+ Select configuration	on. More 🔻			
Resource configurat	+ Manua	ally Enter Co	onfiguration	
O Select Resource G	roup 🚯 🗌 Cus	stom config	guration	
default-rg-8ka70u9	gef ID: rg-8ka70	u9gef	Suspend	•
Executor resource	large(4CU)			
Executor count	2 large(4CU)			
Driver resource				
Driver resource Status	Suspend			

Spark Job Parameters

Note:

1. Modifications to job parameters only take effect in the jobs that are launched subsequently and will not take effect in the running jobs.

2. The priority of job parameters is higher than that of engine-level parameters.

Adding Parameters When a Job Is Created

Enter Data Job, click **Create job**, and add parameters in Job parameter.

ta Lake Compute	Data job 🕓 Beiji	ng Finance 🗸				Create job					
Overview	Spark job Job co	onfiguration Sess	sion management			Basic info +					
Data Fundana	Create job Enter a		Q All	▼ All		Job name •	Enter a job name				
Data Explore							It can contain up to 100	characters in Chinese cl	haracters, letter	s, digits, an	d underscores (_).
Data Scheduling	Job name	Job ID	Job type	Job file	Current ta	Job type •	Batch processing	Stream processing	SQL job		
Data Management	test_for_example	batch_2f2411	SQL job	spark-sql-executor-1.0.1.jar	0	Data engine •	Select a Spark job en	igine.		(*	
							The billing mode of the	selected data engine pre	vails. For more	info, see Su	perSQL engine 🗹 .
Task History Insight	stevensli_post_test	batch_3dfb43	SQL job	spark-sql-executor-1.0.1.jar	0	Program package •	Configuration of the da	ta engine, see Network c	onfiguration 🗹 .		
Management							Select a data path				Select a COS path
ie Management	abobby_py_job	batch_8d1fa0	Batch processing	lakefs_client_test.py	0		COS permissions are re	equired, and .jar/.py files a	are supported.		
SuperSQL Engine	Total items: 3					Main class •	This field is required if	the program package is a	.jar file		
Standard Engine Engine Network Configuration						Program entry parameter	Enter program input p parameters by space	arameters of up to 65,536	characters; sepa	rate two	
Permission Management						Job parameter (config)	Example: spark.netwo	rk.timeout=120s			
Storage Configuration							-coolio info, the	ator info started with "	ark 1 and art-	norlina	J
Audit Log							-coning info, the param	eter into started with "sp	ark, one entry	per inte.	
						CAM role arn •	Select a CAM Role an	'n		*	φ

Editing Parameters of an Existing Job

1. Click Data Job, select an existing job and click Edit.

Data Lake Compute	Data job	🔇 Beijing Finance 🗸								
Overview	Spark job	Job configuration	Session management						Job monitoring	Task history Lo
Data Explore	Create job	Enter a job name or ID	Q, All	• All		¥	All	Last 7 days Last	30 days Select date	Select date
📃 Data Scheduling	Job name	Job ID	Job type	Job file	Current tasks	Task engine 💲	Creator \$	Created at \$	Update time \$	Operation
^{⊒≗} Data Management	test_for_examp	ple batch_2f241	1 SQL job	spark-sql-executor-1.0.1.jar	0	rickyspark	tonywwang 🚯	2025-01-15 23:08:02	2025-01-15 23:08:02	Monitor Edit
🖃 Data Job										
Task History	stevensli_post	_test batch_3dfb4	3 SQL job	spark-sql-executor-1.0.1.jar	0	rickyspark	100006124200 🛞	2025-01-09 10:58:17	2025-01-09 10:58:17	Monitor Edit Running Mo
Management	abobby_py_job	batch_8d1fa	0 Batch processing	lakefs_client_test.py	0	t272_spark	abobbywang 🚯	2024-12-31 14:30:37	2024-12-31 15:27:21	Monitor Edit
SuperSQL Engine	Total items: 3							10 - /	page 🕅 🛋 1	/1 page

2. On the Edit job page, modify the job parameters and click **Save** after the modification.

Edit job				×				
Basic info 🔺								
Job name *	test_for_example							
	It can contain up to 100 cha	racters in Chinese ch	aracters, letters, digits, and uno	derscores (_).				
Job type *	Batch processing	Stream processing	SQL job					
Data engine *	rickyspark		Standard-Spark 🔻					
SQL script	Configuration of the data en Configuration of the data en Control Enter SQL Select I	ected data engine prev ngine, see Network co Data Lake Compute q	alls. For more into, see Standar figuration 겉. Jery file	ra engine 🕻 . For network				
	select 1;			,				
Job parameter (config)	spark.network.timeout = 12	20s						

Engine Network Introduction

Last updated : 2025-03-12 18:03:39

Concept

The engine network is built on a Virtual Private Cloud (VPC) and assigns computing engines (such as the standard Spark engine and the standard Presto engine) with fixed network addresses, for example, 10.255.0.0/16. Each engine network is provided with a gateway for external access to standard engines within the network. This allows computing engines to be accessed via JDBC from either a private network (VPC) or a public network.

Note:

If you need to access resources in different VPCs, such as using a DLC engine to access EMR HDFS data, it is recommended to select an IP range with sufficient available addresses that do not conflict with those used by other products. You can purchase multiple computing engines under the same engine network and manage them centrally through the gateway.



Use Limits

Note:

The IP range should be consistent with the VPC IP range settings and created manually. Once created, it cannot be modified.

1. Use any of the following private IP ranges:10.0.0.0 - 10.255.255.255 (mask range: 12-28)

172.16.0.0 - 172.31.255.255 (mask range: 12-28)

192.168.0.0 - 192.168.255.255 (mask range: 16-28)

2. Make sure that a subnet with sufficient IP addresses is allocated to the engine network to prevent IP address exhaustion, which could hinder Pod creation in large-scale workloads. If the required scale is uncertain, it is

recommended to use the default configuration.

3. When federated queries is used, ensure that the engine IP range does not overlap with the data source IP range.

4. Engine network configuration: Custom network settings can be configured during the initial purchase. To make changes later, submit a ticket to apply for that.

Network Segmentation

Standard engines under each engine network are managed by a gateway. Proper segmentation of engine networks helps balance the gateway load efficiently and mitigates the risk of single point of failure. We recommend segmenting networks based on business departments or task types.

Segmentation by Business

We recommend segmenting engine networks based on business departments. For example, each business department should have at least one engine network.

Segmentation by Task

We recommend segmenting engine networks based on task types. For example, you can create separate engine networks for different tasks such as BI analysis, data governance, and data analysis.

Note:

The above engine network segmentation recommendations are provided based on our experience for reference. You can also adjust the segmentation based on your actual needs, such as creating a dedicated engine network for handling of large-scale tasks according to the task scale.

Private Network Access

Creating a private link allows you to establish a secure and stable connection between your VPC and the gateway, enabling access to standard engines. On the Cloud Access Management page, you can create a private link, select the source VPC and subnet to be accessed, and obtain an access link upon completion. Any machine within the source VPC can then be connected to standard engines in the engine network.



Data Lake Compute	← Cloud Access Management				
B Overview					
Data Explore	Overview	Private Link 0	Gateway 2cu Running	Standard engine 2	¢
≣≣ Data Management					
E Data Job	Private Link				
Task History	Used to connect users' VPCs (submission serve	r) to the engine. You can use Hive JDBC 🗾 or	Presto JDBC 🙋 .		×
Insight Management	Create a Private Unk				
Engine Management	Private Link Name	VPC	Access link	Operation	
SuperSQL Engine	test		jdbc:hive2://	ine=(DataEngineName);spark.resourcegroup=(Reso Ii roperties=presto.engine:(DataEngineName);region:a Ii Details II Deteite	
Engine Network Configuration	Total items: 1			10 ¥ / page 14 < 1 / 1	I page 🕨 H
Ops Management					
o [≠] Permission Management	Gateway				
Storage Configuration	The gateway is a gateway service that helps u	sers build connections between the local data	and the DLC standard engine.		×
🖬 Audit Log	Through the gateway, you can use the console In the test period, the gateway of 2 CUs is free	 DBC, or other methods to submit SQL queri of charge. If you have any questions, submit a 	ies, analyses, and other tasks to the standard engine.Learn more 🗹 a ticket.		
△ Monitoring & Alerting	Spec configuration Start Susp	end Monitor Ø			
	Gateway default-s				
Ξ	Resource ID DataEng				

Public Network Access

Standard engines in the engine network can also be accessed via the public network. For example, certain BI tools deployed on the public network may require a public network connection to the engine.

1. See Private Network Access to create a private link. For example: private network access JBDC link string.

jdbc:hive2://172.22.0.202:10009/?spark.engine={DataEngineName};spark.resourcegrou

2. Go to the Cloud Load Balancer console, create a public network access instance, and select Configure listener.

Cloud Load Balancer	Instance management	Singapore	1 *					(i) Ya	u're using the New version	of monitor dashboard.	Switch to Old v	ersion H	Help of C	LB 🗹
E Overview	Tencent Cloud Load Ba	lance provides ser	rice via domain names	with dynamic VIPs. T	The SLA increases from	99.95% to 99.99%.	For more details, see	Release Notes a	and Announcements 🗹 .				• >	×
Cloud Load Balancer														
🗄 Instance List	Create		t Edit tage	More x					Droject All projecte			0	C 1	-
Certificate	Create	Assign to projec	Eun tags	more +					Project. All projects			4	9 -	*
Management	D/Name \$ Mor	Status	Domain	VIP/EIP	Availability zone	Network T	Network	Instance s	T Health status	Billing mode T	Tags 7	Operation		
E Custom Configuration					Primary:		umo-dkiu/213r			Pay-as-you-go -		[
🛱 Log Management 👻	test_clb	Normal		43.163.	Singapore Zone 4 Secondary:	Public network	Default-VPC (172.22.0.0/16)	Shared	Normal	Creation: Feb 24, 2025 19:01:00	-	More *	Sterier	
Idle Instances					Singapore Zone 1					(UTC+08:00)				
Gateway Cloud Load Balancer	Total items: 1									20 🔻 / page	R 4 1	/ 1 pag	je 🔸	н
田 Instance List														
l目 Target Groups														

3. Go to the Create Listener page, create a listener and select TCP for Listening Protocol. The port should match the private link port by default: 10009 (for accessing the standard Spark engine) or 10999 (for accessing the standard Presto engine).

Cloud Load Balancer	← lb-10c73yvc (test_clb)
E Overview	Basic information Listener management Redirection configurations Monitoring Security groups
Cloud Load Balancer	We support one-click activation of free WAF :
Certificate Management	Note: When dustom redirection policies are o
르늘 Custom Configuration	Basic configuration 2 Health check 3 Session persistence The set of t
🛅 Log Management 🛛 👻	Create Name test_tcp
Idle Instances	Vp to 60 charactersBLANK
	Listener Port 10009
l Target Groups	TCP/UDP/TCP SSL/QUIC listener(Configured) Port range: 1 - 65535 WRR * WRR scheduling is based on the number of new connections. The real server with higher weight stands more chances to be polled.
	You've not created any lister

4. Bind the backend service to the created listener. Select the IP type and enter the private link IP address created earlier, such as 172.22.0.202. Use port 10009 (for accessing the standard Spark engine) or port 10999 (for accessing the standard Presto engine).

Cloud Load Balancer	HTTP/HTTPS listener(Configured0	
B Overview	Create	
Cloud Load Balancer	You've not created any listeners. Create now	Click the left node to view details
🗄 Instance List		
Certificate Management	TCP/UDP/TCP SSL/QUIC listener(Configured1	
E Custom Configuration	Create	
🛅 Log Management 👻	test_tcp(TCP:10009) 🖉 📋	Listener detailsExpand -
Idle Instances	test_tcp(TCP:10009)	Backend service bound
Gateway Cloud Load Balancer		Bind Kodify port Modify weight Unbind Search by private IPs; separate Q Ø
E Instance List		ID/Name Port health status() IP address Port Weight Ope
I Target Groups		
		Listener created. PleaseBind with backend service

5. Use the public network VIP provided by CLB along with port 10009 or 10999 to access engine resources. This converts the access link into a public network connection.

jdbc:hive2://{public network VIP}:10009/?spark.engine={DataEngineName};spark.reso

Accessing the Public Network in the Engine

By default, standard engines do not support public network access. If you need to access the public network, such as for installing Python packages in the notebook using magic %pip, submit a ticket to apply.

Gateway Introduction

Last updated : 2025-03-12 18:03:39

The DLC gateway is a Serverless unified access gateway service deeply optimized based on Apache Kyuubi. Through the gateway, you can achieve stable and secure access to DLC data and standard computing engines based on Hive JDBC/Presto JDBC/DLC JDBC/TencentCloud API standard interfaces, reducing the complexity of managing access to large-scale computing engines. For example, you can submit SQL tasks and ETL jobs to specified standard computing engines through the gateway.

Public netwo	ork access(DLC JDBC、TencentCloudA	API) onnection rate limit 20QPS learn more 🗹	Engine network: name 10.255.0.0/16			Delete engine
	User VPC	Private connection (i) 1 Details	E Gateway Running	Details	Standard engine	
			2 cu		2	
			Engine network: name 10.255.0.0/16			Delete engine
Public network		Private connection (i) 0 Create	🖨 Gateway	Details	Standard engine	
access	Submission machine		None		0	
			Engine network: name 10.255.0.0/16			Delete engine
		Private connection () 3 Details	📋 Gateway	Details	Standard engine	
			8 cu		6	

DLC Gateway

The gateway is a unique service of the DLC standard engine, offering users strengths such as reduced query latency, security and high availability, and flexible integration:

Reduced query latency: The DLC gateway can significantly reduce the time taken on the query link, and improve performance of data interactive analysis, especially for small data volumes.

Support for more access methods: The gateway supports Hive JDBC/Presto JDBC connects to the DLC standard engine, catering to various query scenarios.

Enterprise-level security: Identity authentication and sub-user engine permission control are performed through CAM authentication parameters (AK/SK).

High availability: The gateway provides higher availability and load balancing and supports scaling out for extremely high query concurrency.

Architecture

🔗 Tencent Cloud

As shown in the figure below, only one gateway can be created under an engine network. This gateway can simultaneously manage all standard Spark engines and Presto engines created under the engine network. By default, a user can only have one engine network and can only create one gateway. If the business scenario is complex and there are high requirements for concurrency and other performances, or if some more important businesses require environment isolation, it is recommended that users create multiple engine networks and multiple gateways to physically isolate different tasks.

Note:

1. Creating multiple engine networks and gateways requires the backend to enable the allowlist. Contact DLC development personnel to conduct the operations.

2. Different engine networks and gateways are physically isolated and cannot communicate with each other or access each other's engines.



Creating an Engine Network and Gateway

When the allowlist is not enabled, users have one engine network by default and cannot create another engine network, as shown in the figure below. Users do not need to manually create gateways. When users create the first engine or submit the first task under that engine network, DLC will create a free gateway with specifications of 2 CUs by default under that engine network.

Data Lake Compute St	andard engine	🕽 Virginia 🗸					
 Overview Data Explore Data Scheduling 	Overview Public Network Access	(DLC JDBC, Cloud AP)Connection Frequency Limit: 20 QPSLearn more	Engine Network:default-network-0 (10.	1.255.0.0/16)		
🗄 Data		User VPC	Private Link (1) 0 Go to Create	🗎 Gateway	Details	Standard engine	Purchase
Management	Public Network Access	Submission Machine		After you purchase the first standard engine, a free gateway with 2 CUs is automatically created.		0	
Task History							
P Insight	A standard pay-as-you-go S	ipark engine is charged ba	ased on running tasks and resource groups. When no tasks	or resource groups are running, no fee is ch	harged.Billing Overview 🖸		;

After the allowlist is enabled, users can create multiple engine networks, as shown below. Users can create an engine network by clicking **Create engine network**. The created engine network does not have a gateway initially. Similarly, when users create the first engine or submit the first task under that engine network, DLC will create a free gateway with the specifications of 2 CUs by default under that engine network.



Users can see which engine network the current engine belongs to through the Engine Network Name/ID column on the engine list page.

Data Lake Compute	Standard engine	🛇 Nanjing ~					
 Overview Data Explore 	Overview	Private Link 0	Gateway -	cu	Standard eng	ine 0 Create	engine network
 Data Scheduling Data Management 	A standard pay-as-you	-go Spark engine is charged based on running	tasks and resource groups. When no tasks or	r resource groups are runnin	g, no fee is charged.Billing	Overview 🗠	×
Data JobTask History	Engine Name/ID	Engine type T Engine Status 1	T Engine Network Name/ID T	Resource Groups	Used Resources/T	Access link	Operation
Insight Management Engine Management SuperSQL Engine	vz-test DataEngine-Ocpzrnrm	Standard Spark Ready	test DataEngine-Network-b0z30xkp	2	16/16	DLCJDBC jdbc:dlc:dlc.tencentcloudapi.com?task_type=Spark	Cloud Access Management Monitor Manage Resource Group S Spec configuration Parameter Configuration

Click the Unfold button on the upper right corner to view the engine network list information as shown in the figure below. Click the Details button to view the detailed information of the current engine network, including the number of standard engines under the current engine network, the number of user VPCs connected with the engine network, and the specifications of the gateway.

To avoid wrong cancellation, the system does not allow users to directly delete the engine network. Only when the number of standard engines under the current engine network is 0 can users click Delete Engine Network to delete the engine network.

Data Lake Compute	Standard engine	💲 Nanjing 🗸					
OverviewData Explore	Overview					Create	engine network
Data Scheduling Data	Public Network Access		Private Link ① 0 Go to Create	Engine Network:test (10.255.0.0/16)	Details	Standard engine	Delete Engine Network
Management	PUDIIC Network Access	Machine		2 cu		1	
Task History Insight			Private Link ① 0 Go to Create	Engine Network:sparkLoad (10.255.0.0/16)			Delete Engine Network
Management				Cateway Running	Dotaile	Standard engine	Purchase



Data Lake Compute	← Cloud Access Management
Overview	
Data Explore	
Data Scheduling	If you want to access the engine and data with JDBC, service access nodes first
E Data Management	
🗐 Data Job	
Task History	Total items: 0
Insight Management	Gateway
Engine Management	
59 SuperSQL Engine	 The gateway is a gateway service that helps users build connections between the local data and the DLC standard engine. Through the gateway, you can use the console JDBC, or other methods to submit SQL gueries, analyses, and other tasks to the standard engine.
😚 Standard Engine	 In the test period, the gateway of 2 CUs is free of charge. If you have any questions, submit a ticket.
 Engine Network Configuration 	Spec configuration Start Suspend Monitor Ø
Ops Management	Gateway default-gateway-37trqb8x
o" Permission	Name
Management	Resource ID DataEngine-9sd77nzw
Storage Configuration	Spec 2CU
🖬 Audit Log	Status Running
Monitoring & Alerting C	Tag C Loading Tags are used to categorize resources. To learn more, see Tag Documentation ⊠
=	

Gateway Specifications

The DLC will automatically create a gateway with the specifications of 2 CUs for each engine network, and this gateway will not incur any fees. However, the gateway of 2 CUs is only suitable for the testing environment. It is recommended that users scale out the gateway for the production environment.

The DLC offers various gateway specifications. It is recommended to select the gateway specifications based on the number of engines to be managed, the maximum query concurrency QPS of the business scenario, and others. See the following table for details.

Gateway Specifications	Whether the Gateway Supports HA	Number of Managed Spark Resource Groups	Number of Managed Presto Engines	Number of Spark SQL/Presto SQL Concurrent Queries	Number of Concurrent Spark MLlib Notebook Sessions Created Transiently/Max Recommended	Number of Concurrent Spark Batch Tasks Submitted Transiently/Number of Spark Batch Tasks Running Simultaneously
2 CU	No	50	4	100	10/20	30/50

16 CU	Yes	150	12	200	20/80	80/150
32 CU	Yes	400	35	600	100/200	220/400
64 CU	Yes	700	70	1000	200/300	400/600

Upgrading Specifications

Data Lake Compute (DLC) provides 2 CU specifications for users by default. When the business scenario cannot be met and it is necessary to upgrade the specifications, purchase is required to obtain them.

Note:

1. Gateway configuration adjustment will lead to interruption and failure of all currently running tasks. Proceed with caution.

2. The entire change process is expected to take 10 to 15 minutes. If the gateway status does not return to running for a long time, submit a ticket for resolution.

If users need to upgrade the configuration of the gateway, they can follow the steps below.

1. Click on the left side of the sidebar. Standard engine to enter the engine list page.

2. Click Standard Engine on the left to enter the engine list page. At the top of the page, find the to-be-operated engine network and click **Gateway> Details** to enter the engine network details page.

Data Lake Compute	Standard engine	🛇 Nanjing 🗸										
Overview Data Explore	Overview						Create en	jine network 🗘 Hid				
Data Scheduling			Engine Networl		Engine Network:test (10.255.0.0/16)	1	Delete Engine Network					
E Data Management	Public Network Access	Submission Machine	Private Link (i) 0 Go to Create	Gateway Running	Details	Standard engine	Purchase					
Data Job					2 CU		1					
Task History	,											
🖉 Insight					Private Link @ 0	Contra Consta	Engine Network:sparkLoad (10.255.0.0/16)		1	Delete Engine Network		
Management			Private Link () 0	Go to Create	Catoway Dunning	Datalla	Ctandard angina	Durahasa				
ngine Management												
SuperSQL Engine	A standard pay-as-you-go	Spark engine is charged	based on running tasks and resource	e groups. When no t	tasks or resource groups are running, no fee is charged.Bil	ling Overview 🗹		×				
Standard Engine	Create resource Bill	query Renewal	management			Enter a keyword		Q Ø				

3. Scroll down to the bottom of the details page and click the Spec configuration button of the gateway.

Data Lake Compute	← Cloud Access Management
Overview	
E Data Explore	
Data Scheduling	If you want to access the service
급 Data Management	
🗐 Data Job	
Task History	Total items: 0
Insight Management	Gateway
Engine Management	
5 SuperSQL Engine	The gateway is a gateway service that helps users build connections between the local data and the DLC standard Through the gateway you can use the console. IDBC, or other methods to submit SQL queries, analyses, and oth
Standard Engine	 In the test period the gateway of 2 CUs is free of charge. If you have any questions, submit a ticket.
Engine Network Configuration	Spec configuration Start Suspend Monitor Ø
Ops Management	Gateway default-gateway-37trob8x
ở Permission	Name
Management	Resource ID DataEngine-9sd77nzw
Configuration	Spec 2CU
📩 Audit Log	Status Running
Monitoring & Alerting	Tag No tag 🖉 Tags are used to categorize resources. To learn more, see Tag Documentation 🗳

4. In the pop-up Configuration change page, select the specifications to change to and click Confirm.

FAQs

How to solve the API timeout error when tasks are submitted via JDBC?

First, check the gateway status through the console to see if it is normal and running. If the status of the gateway is Suspend, you can click the Start button to start the gateway and try again. Enter the engine network details page, go to the gateway details at the bottom, and click theStart button.



Gateway	
• The gate	way is a gateway service that help users build connections between the local data and the DLC standard eng
• Through	the gateway, you can use the physical sole, JDBC, or other methods to submit SQL queries, analyses, and other ta
• In the tes	t period, the gateway of 2003 is free of charge. If you have any questions, submit a ticket.
Spec config	uration Start Suspend Monitor Ø
Gateway Name	default-gateway-a7dppz3p
Resource ID	DataEngine-00gv3kh1
Spec	2CU
Status	Suspend
Tag	No tag 🖉
	Tags are used to categorize resources. To learn more, see Tag Documentation 🛂

How to determine whether the current gateway load is normal?

The DLC provides basic monitoring of the gateway, and the health status of the gateway can be judged through the monitoring information. Enter the engine network details page, go to the gateway details at the bottom, and click the **Monitor** button to enter the gateway monitoring page.

Gateway	
The gateThroughIn the te	eway is a gateway service that helps users build connections between the local data and the DLC stan the gateway, you can use the console, JDBC, or other methods to submit SQL queries, analyses, and st period, the gateway of 2 CUs is free of charge. If you have any questions, submit a ticket.
Spec confi	guration Start Suspend Monitor
Gateway Name	default-gateway-37trqb8x
Resource ID	DataEngine-9sd77nzw
Spec	2CU
Status	Running
Тад	No tag 🖉 Tags are used to categorize resources. To learn more, see Tag Documentation 🗳

As shown in the figure below, you can see the monitoring information of the gateway's CPU, memory, task threads and other aspects. If the CPU or memory load exceeds 70%, you need to consider whether the gateway load is high and scale out for the gateway.



Observability Platform	Back to List dataengine-00gv3kh1 Monitoring		
Monitor Overview			
Dashboard ~	12 hours 📋 🕓 Time granularity: 1 min	✓ Ø Disable ✓ ··· ✓ Show legends	
E Instance Group	cpu_usage_rate(%) (i)	cpu_usage_core(Core) ()	cpu_usage_seconds(ms) ①
	20 11:33 18.10	1.33 0.36	1133 0.36
Alarm 🗸	15	0.27	0.27
Management	10	0.18	0.18
	0 08:48 09:57 11:06 12:15 13:24 14:33 15:42 16:51 18:00 19:09 20:18	08:48 10:00 11:12 12:24 13:36 14:48 16:00 17:12 18:24 19:36 20:	08:48 10:00 11:12 12:24 13:36 14:48 16:00 17:12 18:24 19:36 20:
Managed Service for Prometheus	dataengine-00gv3kh1 Max: 18.10 Min: 0.70 Avg: 1.04	dataengine-00gv3kh1 Max: 0.36 Min: 0.01 Avg: 0.02	dataengine-00gv3kh1 Max: 0.36 Min: 0.01 Avg: 0.02

Meanwhile, users can configure alarms in Tencent Cloud Observability Platform (TCOP). When the CPU utilization and the memory usage of the gateway exceed certain limits, the alarms can reach customers in the first place, enabling them to carry out operations such as scale-out of the gateway in advance.

The configuration process is as follows:

1. Enter the TCOP console, select Alarm Configuration, and click Create Policy.

Observability Platform	Alarm configuration	/				Scan QR code to join our communit	on WeChat or WeCom B
Monitor Overview	Cloud Product Monitoring	Alarm Policy Convergence rule	Trigger Condition Templat	e Notification Template	Notification Content Temp	late Scheduling Management	
Dashboard ^	• APM	Create Policy Delete More	~	Adv	vanced Filter Search by Tag, P	olicy Name/ID	Q C ֎ ∓
Dashboard	RUM Cloud Broke	Policy Name Monitoring Ty	Policy Type Alarm Rule	Project T	Associated In Notificatio	□ Last M ↓ □ Alarm On	☑ Operation
Instance Group	Cloud Probe Monitor Terminal	policy-o15a6g89 services	Public Load drop_total_conns Balancer About in_drop_pkts > 10 drop/usage out_drop_pkts > 1 monitor	> 10count Default Project Count/s, 0Count/s	· •	2025/02/10 11:23:03	Copy Delete Alarm Records Set to Default Policy
Alarm Management	Performance Monitoring	fault Tencent Cloud policy-gwju51ig services	ckafka- ib_produce_band instance-broker ib_consume_banc disk_usage_perce	width_per Default Project width_per ntage > 8	3	100018379117 2025/02/10 00:17:29	Copy Delete Alarm Records Set to Default Policy
Alarm Configuration Alarm		cos-amoro- optimizer- 1305424723- dofault-alarm	COS Internet Traffic >	5000MB, Default Project	1	100034446136 2025/01/22 17:27:32	Copy Delete Alarm Records

2. Policy: Any policy

Policy Type: datalake/gateway (dim)

Filters (AND): Select the region where the gateway resides and select the gateway that requires alarm enabled. Multiple filters are allowed.

Trigger Condition: Manually configure the trigger conditions. As shown in the figure below, it It is configured that if either the CPU load or the memory usage exceeds 70%, an alarm will be triggered. Users can configure other alarm trigger conditions according to their needs.

3. Click **Next step:Configure Alarm Notification**. As shown in the figure below, if <u>If there is an alarm notification</u> template, you can reuse the existing template. If there is not, you can create a template and select the users to be notified after the alarm is triggered or select the WeChat group that the alarms are to be distributed to.

	a Lake Compute Cloud Object Storage Cloud Virtual Machine Q. Supports searching f Shortcolf/ Organization Tools Support Cost • EN Q 🗔 93140060380-
(0) ← Alarm configuration / CreateAla	rm Policy
Cloud Product Monitoring APM Basic Int	htigure Alarm > (2) Configure Alarm Net Notification
Cloud Probe Cloud Probe Monitor Remarks Terminal Performance	me Up to 80 characters R can contain up to 100 characters
	rr Alarm Rule c could Product Monthors c decide Product Monthors c for for v to get every (few v to get
Previo	us step Next step: Configure Alarm Notification

4. After the notification template is configured, click **Complete**.

(0)	← Alarm configuration / CreateAlarm Policy	Create Notific	fication Template	×
	Cloud Product	Basic Info		
	Configure Alarm > 2 Confi Policy Notifi	Template Name	Up to 60 characters	
	RUM Configure Alarm Notification	Notification Type	J 🗹 Alarm Trigger 🛛 V Alarm Recovery	
	Cloud Probe Monitor To add an alarm recipient (group), you need to select a no You cannot receive notifications about the alarms to	Notification Language	English 🗸	
	Terminal Performance Notification Select Template Create Tem	Tag	Ting Key Ting Value	
	Monitoring Template You have selected 0 notification template Name		+ Ad () Paste	
		Notifications	IS (Fill in at least one item)	
		User	You can add a user only for receiving messages.Guide for Adding Recipient [2	
	Advanced Configuration(N/A, only metric alarm conditi		Recipient User 👻 📿 Add User	Delete
	Previous step Complete		Notification 🔽 Mon 🔽 Tue 💟 Wed 💟 Thu 💟 Fri 💟 Sat 🗹 Sun Cycle	
			Notification 00.00:00 ~ 23:59:59 () ()	
			Receiving Zemail ZSMS WeChat WeCom Call(Enable Now)	
			Add User Notification	
		API Callback	API Caliback Ltdi Ltdi	Delete
0		L	Configure API Caliback. CM vill send alarm notifications to the URL or corresponding group. View Usage Guides 12	

Standard Engine Startup and Stop Logs

Last updated : 2025-03-21 12:29:26

The log feature of Standard Engine Startup and Stop Logs records the startup and suspension events of each engine, making it easy to monitor engine status, troubleshoot, and optimize resource management.

Operation Steps

Log in to Data Lake Compute (DLC) Console > Resource Management > Standard Engine, choose service region.
 Startup and stop logs of different operation objects:

Gateway: Unfold the overview, click **Details**, and view the startup and stop logs of the gateway on the details page. Presto engine: Select the engine instance you want to view in the engine list, click **engine name**, and enter the basic configuration page to view the startup and stop logs of the computing engine.

Spark engine resource group: In the engine list, select the engine instance you want to view, click **resource group management**, select the resource group you want to view, and click **resource group name** to enter the resource group details page to view the startup and stop logs of the resource group.

Startup and Stop Log List

Note:

Support for Spark engine resource group startup and shutdown logs requires a gateway restart operation after March 20, 2025. Specific operation steps: Click on **Engine Network > Gateway > Details** on the overview card to enter the engine network details page, click **Suspend**, and then click **Start**.

Field Name	Description
Traceld	Traceld is a unique identifier for a start-stop process. It can associate the logs of different actions within the same process, helping users identify which logs belong to the same operation or request.
Time	Starting an action corresponds to the operation start time, and completing an action corresponds to the operation completion time.
Action	The actions include CLUSTER_SCALE_IN、CLUSTER_SUSPEND、CLUSTER_SCALE_UP, etc.
Details	CU adjustment of objects before and after operation.

Resource Group Resource Group Introduction

Last updated : 2025-01-23 17:05:12

The resource group is a secondary queue division of the computing resources within a Standard Spark Engine. Resource groups belong to a parent Standard Engine, and resource groups under the same engine share resources with each other. The computing units (CUs) of the DLC Standard Spark Engine can be allocated to multiple resource groups as needed. You can configure each resource group's minimum and maximum CU limits, start and stop policies, concurrency, and dynamic/static parameters to efficiently manage resource isolation and workload in complex scenes such as multi-tenancy and multi-tasking.

For example, you can create separate resource groups within a Standard Spark Engine, such as a Report Resource Group, a Data Warehouse Resource Group, and a Historical Backfill Resource Group. You can set the upper and lower limits of computing units (CUs) for each resource group and assign relevant SQL tasks or jobs, such as reports and data warehouse tasks, to the appropriate resource group, ensuring resource isolation between different types of tasks and preventing individual large queries from monopolizing resources for extended periods.

Features

Resource Group Isolation

Resource groups enable resource isolation within the Standard Spark Engine. You can assign specific resource groups to different users or queries, effectively isolating resources and preventing a single user or large query from monopolizing most of the computing engine's resources.



Resource Group Elasticity

By configuring the number of Executors in a resource group for dynamic allocation, the resource group can adjust the resources used by SQL tasks or jobs based on the workload, effectively improving resource utilization. The dynamic allocation configuration is shown in the diagram below:



Configuration of	shange
Default job resource spec	
Executor	small(1CU)
resource *	Select desired resources. 1 CU is approximately equivalent to 1-core CPU and 4 GB memo
Executor count *	O Dynamic Fixed
	Minimum – 2 + Maximum – 5 +
	Resources to be used by each executor are those set in the above field
Driver resource *	small(1CU)
	Select desired resources. 1 CU is approximately equivalent to 1-core CPU and 4 GB memor
Total resource size	3CU ~ 6CU

Both Task 01 and Task 02 are set to dynamic allocation, each using 8 CUs at Time A. By Time B, Task 01 only requires 4 CUs, releasing 4 CUs of idle resources for Task 02 to use, thereby improving overall resource utilization. This process is illustrated in the diagram below:



Usage Limitations

The resource group name should be globally unique. It is recommended to use an all-English name.

Terminology

Description	Illustration	Default Resource Groups
(System created by default) Exist upon engine creation, and	The default resource group starts in a suspended status, with settings for automatic start and automatic suspension. The default resource group supports modification of resource configurations.	default-rg-i0wg28z6 F rg-i0wg28z6p4 F

named as default-rg-xxx.	The default resource group supports configuring start/stop policies, setting concurrency limits, and adjusting dynamic/static parameters. The default resource group cannot be deleted.
(User manually created) The custom resource group supports the modification of resource configurations.	The custom resource group supports configuring start/stop policies, setting concurrency limits, and adjusting dynamic/static parameters. The custom resource group can be deleted.

Private Connection Private Connection Introduction

Last updated : 2024-09-04 11:15:28

Endpoints are built on Private Link. If you need to access engines and data through JDBC or other methods, you can create an endpoint to establish a secure and stable private connection between your VPC and the access point.

rk access(DLC JDBC、TencentCloud	IAPI) onnection rate limit 20QPS learn more 🗹	Engine network: name 10.255.0.0/16			Delete engine
User VPC	Private connection ① 1 Details	Gateway Running	Details	Standard engine	
		2 cu		2	
		Engine network: name 10.255.0.0/16			Delete engine
	Private connection ① 0 Create	📋 Gateway	Details	Standard engine	
Submission machine		None		0	
		Engine network: name 10.255.0.0/16			Delete engine
	Private connection ① 3 Details	📋 Gateway	Details	Standard engine	
		8 cu		6	
	rk access(DLC JDBC, TencentCloud	rk access(DLC JDBC, TencentCloudAPI) onnection rate limit 20QPS learn more [2] User VPC Private connection ① 1 Details Private connection ① 0 Create Private connection ① 3 Details	rk access(DLC JDBC, TencentCloudAPI) onnection rate limit 200PS learn more [2] Engine network: name i 10.255.0.0/16 User VPC Private connection ① 1 Details Private connection ① 0 Create Private connection ① 0 Create Private connection ① 3 Details Engine network: name i 10.255.0.0/16 Engine network: name i	rk access(DLC JDBC, TencentCloudAPI) onnection rate limit 20QPS learn more [2] Engine network: name I 10.255.0.0/16 User VPC Private connection ① 1 Details Private connection ① 0 Create Private connection ① 0 Create Private connection ① 3 Details Private connection ① 3 Details Bout	rk access[DLC JDBC, TencentCloudAPI) onnection rate limit 200PS learn more [2] Engine network: name 110.255.0.0/16 User VPC Private connection ① 1 Details 2 cu Private connection ① 0 Create Engine network: name 110.255.0.0/16 Submission machine Private connection ① 0 Create Engine network: name 110.255.0.0/16 Private connection ① 0 Create Engine network: name 110.255.0.0/16 Private connection ① 0 Create Engine network: name 110.255.0.0/16 Private connection ① 1 Details Engine network: name 110.255.0.0/16 Private connection ① 1 Details Engine network: name 110.255.0.0/16 Private connection ① 1 Create None Private connection ① 3 Details Engine network: name 110.255.0.0/16 Engine network: name 110.255.0.0/16 Engine network: name 110.255.0.0/16 Private connection ① 3 Details Engine network: name 110.255.0.0/16

Usage Limitations

- 1. A maximum of 4 endpoints can be created.
- 2. For private connection billing, see Private Link Billing.

Network Connection Configuration

Last updated : 2025-04-09 20:49:29

Data Lake Compute (DLC) supports configuring network (VPC) for data engine, facilitating management of engine access to different data source networks.

Network Configuration Type

According to different business scenarios, DLC provides two network configuration types.

Enhanced network configuration: suitable for accessing the data under one VPC with high speed and stability. **Note**:

1. A data engine of a non-Spark job type can only be bound to one enhanced network configuration.

2. If you use an enhanced network, the subnet IP address under your VPC will be used. Please ensure sufficient subnet IP addresses.

Cross-origin network configuration: suitable for cross-origin federated data query that needs to access multiple VPCs. A data engine can support binding multiple cross-origin network configurations.

Network Configuration Status

Initializing: The network configuration is being initialized. At this point, the network is not active. Success: The network configuration takes effect on the bound engine. Failure: The network configuration fails and can be deleted and reconfigured.

Network Configuration Security Policy

If you have configured a security group policy for the VPC, you need to add inbound rules for different network configuration types.

Enhanced network: Add inbound rules for the IP range of the VPC where the data source is located to the security group.

Cross-source network: Add inbound rules for the IP range of the engine bound to the network configuration to the security group.

Create Network Configuration

- 1. Log in to the DLC console and choose service region.
- 2. Enter **Resource Management > Network Connection Configuration** through the left sidebar.
- 3. Click the **Create Network Configuration** button to enter the Create Configuration page.

The configuration parameters are as follows:

Configuration Content	Required or Not	Filling Instructions
Network Configuration Type	Yes	Select according to the use case Enhanced network configuration: suitable for data scenarios that require high- speed and stable access to a VPC. Cross-origin network configuration: suitable for cross-origin federated query analysis scenarios that need to access data under multiple VPCs.
Configuration Name	Yes	Supports Chinese, English, and _, with a number of characters not more than 35.
Instance source	Yes	Two sources are supported: Data catalog of DLC: Option the data catalog that has created a connection in the data management of DLC currently New network configuration: Select a new data source to create a network connection. Currently, the data source supports MySQL, Kafka, EMR HDFS (COS, HDFS, Chdfs), Postgresql, SqlServer, Clickhouse. If the data source associated with the network configuration to be created is not yet supported, you can select another option and manually specify a VPC.
Catalog	Yes	Select the corresponding data catalog according to the source of the selected instance. The range of selectable data catalogs will be related to your account permission.
Data source VPC	No	The data engine network will connect all subnets in the VPC.
Bound data engine	Yes	Select the data engine associated with this network configuration. If the data engine is in isolated or initializing status, it will be unable to select.
Configuration Description	No	Not more than 100 characters.

4. Fill in, complete the settings and save. Then you can create a network configuration.

Note:

Once created, the network is in the initialization state. Subsequently, you can view the status in the list.

Delete Network Configuration

You can perform a deletion operation to manage the deletion of network configurations that are no longer needed or have failed to configure. Directions:

1. DLC console, choose service region.

2. Enter **Engine Management > Engine Network Configuration** through the left sidebar.

3. Find the network configuration that needs to be deleted. Support filtering search. Note the selection of network configuration type.

4. Click the **Delete** button. Just complete the deletion after secondary confirmation.

Note:

After deletion, this data engine will not be able to use this network configuration. If you need access, reconfiguration is required. Proceed with caution.

Modify Description Information

You can modify the description information of the configured network configuration by modifying the description information. Directions:

1. DLC console, choose service region.

2. Enter Engine Management > Engine Network Configuration through the left sidebar.

3. Find the network configuration that needs to be deleted. Support filtering search. Note the selection of network configuration type.

4. Click the **Modify Description Information** button to edit.

Storage Configuration Managed Storage Configuration

Last updated : 2024-07-31 17:30:11

Managed storage refers to the storage space hosted on the Data Lake product, with COS as the underlying storage. Managed storage contains data such as native tables, user program packages, and query results. Therefore, to utilize the capabilities of native tables and data optimization, it is necessary to enable managed storage first. The native tables on managed storage are by default in the Iceberg format, so you don't need to manage the underlying file contents. For details on managed storage billing, please refer to Billing Overview. This document introduces how to enable and configure managed storage.

Enable Managed Storage

Step 1: Enter Managed Storage Configuration

You can enter the managed storage configuration in the Data Exploration module or the Global Configuration > Storage Configuration module.



Step 2: Open Managed Storage

1. Check to enable managed storage and save.

Here, you can specify the managed storage type as either a Metadata Acceleration Bucket or an Ordinary Bucket. The billing for both is consistent, but it is necessary to separately configure engine access permissions for the Metadata Acceleration Bucket. For details, please refer to Binding of Metadata Acceleration Bucket.

2. The query result path is used to temporarily store SQL query results, Spark Job Shuffle data, etc. You need to specify a path to ensure the normal operation of jobs and tasks. If you have enabled managed storage, it is recommended to configure the query result path as **Managed Storage**. You can also configure the query result path to your own account's COS bucket path.

Storage configuration		
Managed storage	Enable	
Managed storage type	General bucket	
Query result storage path 🛈	Managed storage	User-defined storage
Save Cancel		

View managed bucket

After enabling managed storage, a bucket will be created, and you can view the buckets and data on managed storage in the Data Management module.

0	Data management	🔇 Guangzhou 👻						Data Manag	jem
	Catalog Database	Bucket list							
© ≣	O You can view the files	you have stored in the built-in buckets in	the bucket list to efficiently manage your	business. You can view and download billing details in the Billin	ig Center, and view billing standards in Billing Overv	view 🖸 .			
Ξŧ	Bucket name		Usage		Metadata acceleration bucket		Created at	Operation	
E		1230 T <u>a</u>	JGB		20		2021-12-16 18:58:21	View	
Ŷ	Total items: 1							10 ≠ / page H < 1 /1 pi	age

Destroy Managed Storage

Destroying data is a high-risk action; only after all database table data has been deleted, can you proceed to destroy managed storage. Destroying managed storage requires administrator privileges.

Step one: Delete database table data

To destroy managed storage, you must first delete all database table data on the managed storage.

You can refer to the Data Catalog and DMC and Data Table Management documents to delete the database table data, or you can run the DROP Syntax in the Data Exploration module to delete the database table data.

Step two: Destroy Managed Storage

After deleting the database table data, you can destroy managed storage on the managed storage configuration tab under the Storage Configuration module.

Destroying managed storage will delete all DLC managed buckets, so please proceed with caution.

Binding a Metadata Acceleration Bucket

Last updated : 2024-07-31 17:30:27

DLC supports the binding of Fusion Bucket to accelerate Query Analysis Performance. To use this feature, you need to create a Metadata Acceleration Bucket. DLC Managed Storage provides Metadata Acceleration Bucket. Use COS Bucket under the user's account. For details, please see COS>Metadata Acceleration.

When accessing the DLC Metadata Acceleration Bucket, binding of permissions is necessary. The Permission Binding Process is as follows.

Bind Data Engine and Metadata Acceleration Bucket

1. log in to Data Lake Computing Console, enter Common Management > Storage Configuration.

2. Enter the **Metadata Acceleration Bucket Configuration Page**, select the bucket you want to bind, and click **Configure**.

Note:

Only Metadata Acceleration Buckets are displayed on the Metadata Acceleration Bucket page; ordinary buckets (buckets without the metadata acceleration feature enabled) will not be shown.

Storage configuration 🔇 Guangzhou *			
Managed storage configuration Metadata acceleration-ena	abled bucket configuration		
① A metadata acceleration-enabled bucket can be used only after the p	roduct permission and VPC access permission are configured.		
		Enter a	a bucket name Q. All
Bucket name	Bucket type	Associated Data Lake Compute data engines	Operation
: 3 10			Configuration
3 10	1070700		Configuration

3. Click **Bind** to bind the data engine that needs to access this bucket to the Metadata Acceleration Bucket.



un access to metadata acc		a bucket				
letadata acceleration-enabled bu	ucket name					
letadata acceleration-enabled bu	ucket type U					
ind data engine						
Enter an engine name	Q					Ģ
Data engine name				Operation		
				Bind		
ri				Unbind		
Total items: 2			10 🔻 / page	 1	/ 1 page	► ►
ssociate Tencent Cloud pro	oducts	Resource			Operatio	'n
ssociate Tencent Cloud pro	oducts	Resource No data j	yet		Operatio	'n
ssociate Tencent Cloud pro	oducts	Resource No data y Add produ	vet		Operatio	'n
et HDFS user Edit	oducts	Resource No data y Add produ	vet uct		Operatio	n
et HDFS user Edit	oducts	Resource No data t Add produ	yet uct		Operatio	n
et HDFS user Edit uperuser Edit ote: This section enables you to	oducts manage the tenant in	Resource No data : Add produ	yet uct		Operatio	n
et HDFS user Edit uperuser Edit ote: This section enables you to et access to HDFS metadat	oducts manage the tenant in ta	Resource No data t Add produ	yet uct		Operatio	л
et HDFS user Edit uperuser Edit ote: This section enables you to et access to HDFS metadat VPC name/ID	oducts manage the tenant in ta	Resource No data t Add produ	yet 		Operatio	n

Bind computing resources of SCS

If you use SCS to stream data into the lake, and the storage written to is a Metadata Acceleration Bucket, then you need to configure access permissions for the Metadata Acceleration Bucket under Storage Configuration. Under the Tencent Cloud Product Binding section, create a new product, select Stream Computing Oceanus and the corresponding resources, then click save.



dit access to metadata acce	leration-enabled l	DUCKEL					
etadata acceleration-enabled buc	ket name						
etadata acceleration-enabled buc	ket type 🛛 U						
ind data engine							
Enter an engine name	Q						¢
Data engine name					Operation		
					Bind		
ri -					Unbind		
						14	h hi
Total items: 2	ducts		10 🔻 / page		1	/ I page	
Total items: 2 ssociate Tencent Cloud proc Product	ducts	Resource	10 🔻 / page	4	1	Operatio	n
Total items: 2	v.	Resource	10 ¥ / page	v	1	Operatio Save Car	n
Total items: 2 ssociate Tencent Cloud prod Product Stream Compute Service t HDFS user Edit	ducts T	Resource Select	10 • / page	v .	1	Operatio Save Car	n
Total items: 2	ducts T	Resource Select	10 v / page	v V	1	Operatio	n
Total items: 2	ducts anage the tenant info	Resource Select	10 V / page	¥ 4	1	Operatio	n
Total items: 2	anage the tenant info	Resource Select	10 ▼ / page	▼ ▼	1	Operatio	n

Bind computing resources of non-DLC data engines

Sometimes, the computing resources you need to access the Metadata Acceleration Bucket are not from a DLC data engine. In this case, you can configure access permissions for the Metadata Acceleration Bucket under Storage Configuration.

HDFS User Configuration is used to configure the super user of your computing resources accessing DLC, usually root/hadoop/presto/flink.

HDFS Metadata Permissions Configuration is used to configure the VPC Network Environment you allow to access DLC, usually the VPC where the computing resources of the above mentioned non-DLC data engines are located.

e: This section enables you to manag	e the tenant information of compute nodes.		
access to HDFS metadata			
PC name/ID	Node IP	Operation	
		Edit Delete	
		Edit Delete	
-			
		Edit Delete	

Metadata Management Data Catalogs and DMC

Last updated : 2024-07-31 17:27:26

External data and managed storage data in DLC can be managed through the Data Management Page by executing standard SQL statements and APIs. Through the Console Data Management Page, you can create, edit data catalogs, and create, query, delete databases and tables.

Creating a data catalog

Note:

The platform will automatically create a DataLakeCatalog for you for data management on the lake.

When you have external data sources and wish to perform federated analysis, you can follow the process below to create a data catalog for external data sources.

1. Log in to DLC console, select the service region. The account used to log in must have the permission to create a catalog. For enabling sub-account permissions, refer to Sub-account Permission Management.

2. Enter Data Management, click Create Catalog.

Data management	Suangzhou ▼							Data Management	
Catalog Database	Bucket list							Та	
Perform creation, edit, deletion, or other management operations on catalogs.									
Create catalog Sele	ct a connection type 💌 Update time	All Last 7 days I	Last 30 days Select date Select date	te 🖬			Enter a	name	
Catalog name	Connection type	Connection info	Status	Creator	Created at 🕈	Update time 💲	Connectivity status	Operation	
G	MySQL	Co	Created successfully		2023-10-27 16:17:43	2023-10-27 16:17:43	Normal	Edit Delete Test	

3. Enter the data source creation visual interface. After filling in the connection information, complete the network configuration to connect the engine with the external data source.


Create catalog		×
1 Catalog configuration	> (2) Network configuration	
Connection type *	MySQL •	
Connection name *	te]
Description	Up to 50 characters	
Instance v	cd "5 *	
Data source VPC *	vpc-i 🔿 x 💌 subnet-c 3fr1j 💌 🖒 253 IPs in total, 245 available	
Username •	чи: .	
Password •		

Create catalog	
Catalog configuration	> 2 Network configuration
Use the bound data e engine via Network o	engine to query and analyze data from this data source. You can change the scope of the bound data configuration 🖉.
Data source VPC	vpc-73vy8arx ▼ subnet-c93fr1js ▼
	253 IPs in total, 245 available You can configure a network for a data engine to access data sources over it. Enhanced network configurat offers faster data transmission and thus is suitable for accessing a large volume of data. Cross-source netw configuration allow you to set several networks for one data engine for cross-source federated data query across several networks.
Network configuration type *	Enhanced Cross-source
Network configuration name •	It can contain up to 25 characters in letters, digits, and underscores ().
Available data engines *	Select a data engine 🔻
	Only the selected data engine can read data under this catalog. Only Presto private data engines are availa for this selected catalog.
Configuration description	Up to 50 characters

4. After filling in the data source information, click **Confirm** to complete the creation of the data source.

5. In the Data Catalog List, view connection information, status, creator, and other information.

Edit Data Catalog



1. Click Data Catalog List > Operations > Edit to modify the Data Catalog's description information, network

configuration information, username, password, and running cluster, etc.

Edit catalog		>
Connection type *	MySQL	
Connection name *		
Description	Up to 50 characters	
JDBC 🔻	jdbcmy Example: jdbcmysol://jo:port: database name is not required.	
Data source VPC *	vpc-73 Carx 💌 subnet -00-11 💌 🆒 253 IPs in total, 245 available	
Username *		
Password *		
Use the bound dat via Network confi Data engines boun	a engine to query and analyze data from this data source. You can change the scope of the bound data engine guration [2] . d to the data source VPC:	

2. After modifications, click **Create** to reconstruct the Data Catalog.

New database

1. Log in to DLC Console, select the service region. The account used to log in must have database creation permissions.

2. Enter Data Management, click on the directory name under the Data Catalog to view the databases within that directory.

3. Click Create Data Catalog to enter the Database Creation Visual Interface.



Data management	🖏 Guangzhou 👻			Use guide 🧭 🛛 Data Managemer
Catalog Database	Bucket list			
 This module allows you 	u to manage databases under different catalogs. You can	click the name of a database to manage its tables, views, and other data objects, or t	o manually import data to tables. Learn more	2. Data operations require relevant data permissions. For more permission guide, see here 💋.
Create database Data	aLakeCatalog 🔹			Enter a name
Database name 🗘	Created at \$	Description	Creator	Operation
E	2023-10-26 17:51:31			Edit Delete
G	2023-10-26 17:51:30		-	Edit Delete

4. After filling in the relevant database information and saving, the database creation is complete. When creating a database, you can enable data optimization for the entire database.

		>
Database name *	Enter a database name	
Description	Optional	
L Data governance		

Database Name: Globally unique, supports English case-sensitive letters, numbers, "_", cannot start with a number, up to 128 characters.

Description: Supports both Chinese and English, up to 2,048 characters.

A root account can create up to 100 databases.

View Database

1. Log in to DLC Console, select the service region. The account used to log in must have database query permissions.

2. Enter Data Management > Database, select the data directory, click Database Name to access the database details, manage the database's tables. For a detailed operation guide, refer to Data Table Management.



Data management	🖏 Guangzhou 🔻			Use guide Ø	Data Managemen
Catalog Database	Bucket list				т
① This module allows ye	ou to manage databases under different catalogs. You can i	lick the name of a database to manage its tables, views, and other data objects, or	to manually import data to tables. Learn more	2. Data operations require relevant data permissions. For more permission guide, see here	2.
Create database Dat	taLakeCatalog 💌			Enter a name	
Database name 🕈	Created at \$	Description	Creator	Operation	
	2023-10-26 17:51:31			Edit Delete	

Dropping a Database

1. Log in to DLC Console, select the service region. The account used to log in must have database deletion permissions.

2. Enter Data Management, click **Delete**. After confirming a second time, the database can be deleted.

Delete this database			×
Delete all data in a database	before deleting it.	A deleted database	e cannot be recovered.
	Delete	Cancel	

Data Table Management

Last updated : 2024-07-31 17:27:51

Users can use the DLC console or API to execute DDL statements to create a database.

Creating Table

Approach one: Create in Data Exploration

Log in to the DLC console, select the service region, log in to users need to have the permission to create tables.
 Enter the Data Exploration module, in the left list, click on an existing database, hover over the table row, then click the

...

icon, click Create Native Table or Create External Table.

Note:

A native table refers to a table on the DLC managed storage. With a native table, you don't need to worry about the underlying Iceberg storage format, and it has capabilities like data optimization. To use a native table, you need to enable managed storage first, see Managed Storage Configuration for details.

The underlying data of the external table resides on your own COS. Creating an external table requires specifying the data path.

Data Explo	ore 🕓 Guan	gzhou 🔻	
Database	Query	¢ +	Query-2023-12-13 ● + ▼
Catalog DataL	akeCatalog	•	💿 Running 🖾 Save 😋 Refresh 🛱 Format 🔊
Select a target	database	•	1
* 🛢 ;			
🕨 🎹 Table	e	•••	
▶ 🔡 View		Create	native table
▶ 📧 Funct	tion	Create	e external table

3. After clicking **Create Native Table/Create External Table**, the system will automatically generate an SQL template for creating a data table. Users can modify the SQL template to create a data table. After clicking **Run**, the SQL statement to create the data table is executed, completing the creation.

Data Explore S Guangzhou	Ŧ	
Database Query Q	¢ +	Query-2023-12-13 • + •
Catalog DataLakeCatalog	•	Partial run 🖾 Save 🗘 Refresh 🛱 Format 🔊
Select a target database	•	1 CREATE TABLE IF NOT EXISTS 2 ·`db_name`.`new_table_name`(3 `column_name1`.column_tyne1
		4 `column_name2` column_type2
Table		<pre>5) TBLPROPERTIES ('format-version' = '1', 'write.upsert.enabled' = 'false');</pre>
▶ BB View		
▶ 🔊 Function		

Approach two: Create in Data Management

The Data Management module supports managing native tables and external tables stored in DLC.

1. Log in to the DLC console, select the service region, log in to users need to have the permission to create tables.

2. Through the left menu, enter **Data Management**, enter **Database**, click the name of the database where the data table is located, enter the DMC page.

Data management Catalog Database	S Guangzhou *			Use guide \$	ð Data Managemer
 This module allows you 	u to manage databases under different catalogs. You can	click the name of a database to manage its tables, views, and other data objects, or to	manually import data to tables. Learn more	2. Data operations require relevant data permissions. For more permission guide, see her	e 12.
Create database Data	aLakeCatalog 🔻			Enter a name	
Database name 💲	Created at \$	Description	Creator	Operation	
	2023-10-26 17:51:31			Edit Delete	
G	2023-10-26 17:51:30			Edit Delete	

3. Click **Create Native Table** or **Create External Table** button to enter the data table configuration page.

Database / c Data table View Function				Task history Storage o
Data tables under the database. You can manage basic info, fields; and .	other info in the native and external tables, and import da	ta from local system or COS and export data to COS asynchronous)	y. You can view the task running details in the task history.	For the billing mode of the native table, see Billing Overview 🗹
Create native table Create external table Select a table type Data table name \$ Table type	Update time All Last 7 days Rows Table size	Last 30 days Select date Select date Governance status Created at \$	Batch delete Update time \$ Creator	Enter a name Description Operation
		No data		

Native table data sources support three different types: empty table, local upload, and COS COS. Choosing different data sources corresponds to different creation processes. Native tables support capabilities such as data optimization and can choose to inherit database governance rules or individually turn them on/off.

3.1 Create Empty Table: Create an empty table with no records.



Data Table Name: Cannot start with a number, supports uppercase and lowercase letters, numbers, and underscores

-, with a maximum of 128 characters.

Support for entering data table description information.

Manually add and enter column names and field types. Supports the configuration of three complex type fields: array/map/struct.

Create native ta	ble	×
Data table source	Blank table 🔹	
	Create a table for specific data and import data, or directly create a blank table.	
Data table name	Enter a data table name	
Data table version	Select 💌	
	Iceberg table version. v1: Analytic data tables; v2: Supports row-level updates and deletes.	
Description	Optional	
Field info	Field name Field type Field configuration Description	Operation
	No data	
	Add	
Partitioning		
Inherit database	• Yes No	
governance rules	The current data table inherits the governance rules of the database as follows:	
Data governance		
Attributes 🕨		
Confirm	Cancel	Show SQL

3.2 Local Upload: Upload local form files to DLC to create data tables, supports files up to 100MB.

CSV: Supports visual configuration of CSV parsing rules, including Compression Format, Column Splitting Symbol, Field Domain Symbol. Supports automatic inference of the data file's Schema and parsing the first row as Column Names.

Json: DLC only recognizes the first level of Json as columns, supports automatic inference of the Json file's Schema. The system will recognize the first level fields of Json as Column Names.

Supports common Big Data Format files like Parquet, ORC, AVRO, etc.

Manually add and enter Column Names and Field Types.

If the Automatic Structure Inference is selected, DLC will automatically fill in the detected columns, Column Names, and Field Types. If incorrect, please manually modify.



Create native ta	ble	
Data table source	Upload 🔻	
	Create a table for specific data and import data, or directly	create a blank table.
Data path *	Select file	
	You can upload a file of up to 100 MB. For files larger than 1 or other tools.	100 MB, please use the COS mode or import them with API
Data format	Select a data format	
Data table name	Enter a data table name	
Data table version	Select 💌	
Description	Optional	
Field info	Infer structure Automatically infer the data structure based on the select modify the data structure.	ted file. Please confirm the data structure info, or manually
	Field name Field type Field	d configuration Description Operation
	No	data
	Add	
Partitioning		
Inherit	• Yes No	
ualaOdSe		Show SC

3.3 Create a data table through COS COS.

Create a data table by reading the COS data buckets under the current account.

CSV: Supports visual configuration of CSV parsing rules, including Compression Format, Column Splitting Symbol,

Field Domain Symbol. Supports automatic inference of the data file's Schema and parsing the first row as Column Names.

Json: DLC only recognizes the first level of Json as columns, supports automatic inference of the Json file's Schema. The system will recognize the first level fields of Json as Column Names.

Supports common Big Data Format files like Parquet, ORC, AVRO, etc.

Manually add and enter Column Names and Field Types.

If the Automatic Structure Inference is selected, DLC will automatically fill in the detected columns, Column Names, and Field Types. If incorrect, please manually modify.

🔗 Tencent Clou	Jd
----------------	----

Create native ta	ble ×
Data table source	COS 💌
	Create a table for specific data and import data, or directly create a blank table.
Data path *	Select a data path Select a COS path
	You can upload a file of up to 100 MB. For files larger than 100 MB, please use the COS mode or import them with API or other tools.
Data format	Select a data format 🔻
Data table name	Enter a data table name
Data table version	Select 💌
	Iceberg table version. v1: Analytic data tables; v2: Supports row-level updates and deletes.
Description	Optional
Field info	Infer structure
	Automatically infer the data structure based on the selected file. Please confirm the data structure info, or manually modify the data structure.
	Field name Field type Field configuration Description Operation
	No data
	Add
Partitioning	
Inherit	• Yes No
Confirm	Cancel Show SQL

4. Data Partitioning is often used to enhance Query Performance and is applied to large volume tables. DLC supports data querying by Data Partitioning. Users need to add partition information at this step. By partitioning your data, you can limit the amount of data scanned with each query, thereby improving Query Performance and reducing usage costs. DLC adheres to Apache Hive's partitioning rules.

The partition column corresponds to a subdirectory under the COS path of the table, with the directory naming convention being **Partition Column Name=Partition Column Value**. Example:

```
cosn://nanjin-bucket/CSV/year=2021/month=10/day=10/demo1.csv
cosn://nanjin-bucket/CSV/year=2021/month=10/day=11/demo2.csv
```

If there are multiple partition columns, they need to be nested in the order specified in the create table statement.

```
CREATE EXTERNAL TABLE IF NOT EXISTS `COSDataCatalog`.`dlc_demo`.`table_demo` (
   `_c0` string,
   `_c1` string,
   `_c2` string,
   `_c3` string
) PARTITIONED BY (`year` string, `month` string, `day` string)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES ('separatorChar' = ',', 'quoteChar' = '"')
STORED AS TEXTFILE
LOCATION 'cosn://bucket_name/folder_name/';
```

Query basic information of the data table

Approach one: Query in Data Exploration

In the Data Table Item, mouse hover over the Data Table Name row, then click the

...

icon, in the Dropdown Menu click **Basic info** to view the basic information of the created data table.

Data Exp	olore) Guangzhou 🔻				
Database	Query	φ+	Query-2023-12-13 1.	Query-	2023-12-13 🌒	+ •
Catalog Dat	taLakeCatalog	•	● Running	Complet 🔻	🛱 Save	Ġ Refre
Select a targ	et database	v	1			
• 8						
🔻 🆽 Tab	ble					
E		•••				
Ē		Basic	info			
Ē	i	Previ	ew data			
E	3	Delet	te table			
E		Show	v table creation statement	ts		
E	5 101155	e Add	name to SQL			

The basic information of the data table is as follows:

Basic info of	data table	×
Table name		
Table type		
Database		
Mapping view		
Created at	The second se	
Data path	· · · ·	
Data format	- and the second se	
Description		

Approach two: View in Data Management

1. Log in to the DLC Console, select the service region. Users need to have the permission to view data tables.

2. Through the left menu, enter the **Data Management** page, click the name of the database where the data table is located, enter the DMC page. It supports querying information such as the number of rows, storage space, creator, fields, partitions, etc.

ata management	S Guangzhou -							Us	e guide 🧭 🛛 Data Manage
	bucket list								
(i) This module allows you to	to manage databases under different catalogs. Y	ou can click the name of a dat	abase to manage its tables, vi	ews, and other data objects, or to many	ually import data to tables. Lea	arn more 🖾 . Data operations re	equire relevant data per	missions. For more permission guide	, see here 🗹 .
Create database DataLa	akeCatalog 👻							Enter a	name
Database name 💲	Created at \$		Descript	ion	Creator			Operation	
• •	2023-10-26 17:51:31							Edit Delete	
· 6	2023-10-26 17:51:30							Edit Delete	
6	2023-10-26 17:51:30							Edit Delete	
. 6	2023-10-25 20:45:24							Edit Delete	
	2023-10-08 12:08:07							Edit Delete	
- Database / Data table View	Function								Task history Stora
Database / Data table View Data tables under the data	Function	other info in the native and ext	ernal tables, and import data	from local system or COS and export d	ata to COS asynchronously. Yo	u can view the task running det	alls in the task history. F	For the billing mode of the native tab	Task history Storag
Database / Data table View Data tables under the dat Create nutive table Create nutive table Cneate nutive table	Function Itabase You can manage basic info, fields, and o reate external table Select a table type	other info in the native and ext	ernal tables, and import data	from local system or COS and export di Last 30 days Select date	ata to COS asynchronously. Yo Select date	u can view the task running det Batch delete	ails in the task history. F	For the billing mode of the native tab	Task history Storag He, see Billing Overview L2 name
Database / Data table View Data tables under the dat Create native table Create native table Cn	Function Natabase. You can manage basic info, fields, and of reste external table Select a table type Table type	Therinfo in the native and ext Update time Rows \$	All Last 7 days	from local system or COS and export d Last 30 days Select date Governance status	ata to COS asynchronously. Yo Select date 🗂 Created at ‡	u can view the task running det Batch dekete Update time ‡	alls in the task history. F Creator	For the billing mode of the native tab Enter a Description	Task history Store ike, see Billing Overview [2 name Operation
- Database / Data table View ① Data tables under the dat Create native table Create native table Create native table ① Data table name \$ ③ Data table name \$ ③ Data table name \$ ③ Data table name \$ ④ Data table name \$ ● D	Function tabase. You can manage basic info, fields, and or reste external table Table type	other info in the native and ext Update time Rows \$	All Last 7 days Table size *	from local system or COS and export d Last 30 days Select date Governance status 	ata to COS asynchronously. Yo Select date Created at ‡ 2023-08-09 17:01:22	u can view the task running det Batch delete Update time # 2023-10-09 143338	ails in the task history. F Creator	For the billing mode of the native tab Enter a Description 	Task history Store Ne, see Billing Overview I2 name Operation Basic info Edit Import data Es Delete
- Database / Data table View	Function Function Function Fablese You can manage basic info, fields, and of sele external table Select a table type Table type Table type	other into in the native and ext Update time Rows #	ernal tables, and import data All Last 7 days Table size ‡	from local system or COS and export date Last 30 days Select date Governance status 	ata to COS asynchronously. Yo Select date Created at 2023-08-09 17/01:22 2023-08-09 17/00:33	u can view the task running det Batch ddete Update time * 2023-10-09 1433:38 2023-10-09 1433:27	alls in the task history, F Creator 	For the billing mode of the native tab Enter a Description 	Task history Storag ke, see Billing Overview Ed rame Operation Basic info Edit Import Edit Import Edit Import Edit Delete

Preview Data Table Data

In the Data Table Item, hover the mouse over the Data Table Name row, then click the

icon, in the Dropdown Menu click **Preview Data**. DLC will automatically generate a SQL Statements to preview the first 10 rows of data, executing the SQL Statements to query the top 10 rows of the data table.

Data preview			×
Select a data engine	public-engine •	Execute Show first 100 entrie	s by default
id	pro_name	price	pro_date
	-		2.30 -
			2
		10	
	, /	a	
1			
	_		
Total items: 17		10 💌 / page	I 1 / 2 pages ▶ ▶

Support for previewing data in **Data Management > Database > Data Table > Data Table List**. The Data Preview Function by default displays the first 100 rows of data.

Editing Data Table Information

Support editing the Description information of the data table in the Data Management module.

1. Log in to the DLC Console, select the Service Region. Users need to have the permission to edit data tables.

2. Through the left menu, enter the **Data Management** > **Database** page, click the name of the database where the data table is located, enter the DMC page.

3. Find the data you need to edit, click the **Edit** button on the right to edit.

← Database /										
Data table View	Function								Task history Storage configu	ration
 Data tables under the . 	he database. You can manage basic info, fields, and c	other info in the native a	nd external tables, and import data fr	om local system or COS and export o	data to COS asynchronously. Ye	ou can view the task running de	tails in the task history.	For the billing mode of the native table, see	Billing Overview 🗹 🛛 🗙	¢
Create native table	Create external table Select a table type		e time All Last 7 days	Last 30 days Select date	Select date 📋	Batch delete		Enter a name	Q	φ
Data table name 🗘	Table type	Rows \$	Table size 💲	Governance status	Created at 💲	Update time 🗘	Creator	Description	Operation	
- - 6			100		2023-07-12 14:46:58	2023-07-12 14:47:24		-	Basic info Edit Preview Import data Export data Delete	
				**	2023-07-12 14:36:32	2023-07-12 14:45:59			Basic info Edit Preview Import data Export data Delete	
Total items: 2								10 🔻 / page 🛛 H 🖂	1 /1 page →	H

4. After modification, click the **Confirm** button to complete the editing.

Edit data table		×
Data table name	Terrar Control of Cont	
Data table version	VI VI	
	Iceberg table version, v1: Analytic data tables; v2: Supports row-level updates and deletes.	
Upsert		
Created at	2023-07-12 14:46:58	
Update time	2023-07-12 14:47:24	
Description		
Inherit	• Yes No	
governance rules	The current data table inherits the governance rules of the database as follows:	
Data		
governance		
Confirm	Cancel	

Dropping a Table

Approach one: Delete in Data Exploration

In the Data Table Items, hover the mouse over the Data Table Name row, then click the

...

icon, in the dropdown menu click **Delete**. DLC will automatically generate the SQL statement to drop the data table, execute the SQL statement to drop the table.

Dropping an external table, dropping a data table only removes the metadata stored in DLC, it does not affect the data source file.

Deleting tables under the DataLakerCatalog directory will clear all data of that table, proceed with caution.

← Database /									
Data table View	Function								Task history Storage co
 Data tables under the . 	e database. You can manage basic info, fields, and o	ther info in the native and exte	rnal tables, and import data	from local system or COS and export da	ta to COS asynchronously. Y	ou can view the task running det	ails in the task histor	y. For the billing mode of the native	able, see Billing Overview 🕻
Create native table	Create external table Select a table type	▼ Update time	All Last 7 days	Last 30 days Select date	Select date 📑	Batch delete			er a name
Data table name 🗘	Table type	Rows \$	Table size 💲	Governance status	Created at \$	Update time 🗘	Creator	Description	Operation
					2023-07-12 14:46:58	2023-07-12 14:47:24		**	Basic info Edit Prei Import data Export (Delete
- 🥅 n	-				2023-07-12 14:36:32	2023-07-12 14:45:59	1		Basic info Edit Prei Import data Export (Delete
Total items: 2								10 💌 / page	H < 1 /1 page

Approach two: Delete in Data Management

Currently, Data Management only supports the management of databases and tables hosted in DLC. For external tables, please use approach one for deletion.

1. log in to the DLC Console, select the service region, users need to have the permission to delete data tables.

2. Through the left menu, enter **Data Management** > **Database**, click the name of the database where the data table is located, to enter the DMC page.

3. Click the **Delete** button after the data table you wish to delete, after confirmation, the corresponding data table can be deleted and its data cleared.

Data Explore 🖏 Guangzhou	Ŧ
Database Query	φ + Query-2023-12-13 ● + ▼
Catalog DataLakeCatalog	▼ 🕑 Running 🖾 Save 😋 Refresh 🛱 Format 🚮
Select a target database	1 SHOW CREATE TABLE ` . ` `;
▼ ₿ ;	
🔻 🇮 Table	
Ē	
	Basic info
E :	Preview data
	Delete table
E	Show table creation statements
	Add name to SQL
Ē	

Show create table statement

In the Data Table Item, hover the mouse over the Data Table Name row, then click the

icon, in the dropdown menu click **Show table creation statements**. DLC will automatically generate the SQL statement to view the create table statement for that data table, execute the SQL statement to query the create table statement.

Data Explore 🔇 Guangzhou	~
Database Query	Ø + Query-2023-12-13 ● + ▼
Catalog DataLakeCatalog	▼ 🕞 Running 🖾 Save 😋 Refresh 🛱 Format 😡
Select a target database	1 SHOW CREATE TABLE `.`.`;
▼ 🛢 i	
Table	
Ē	•••
	Basic info
	Preview data
	Delete table
E	Show table creation statements
	Add name to SQL
Ē	

System constraints

DLC allows up to 4096 data tables under each database, supports a maximum of 100,000 partitions per data table, and the maximum number of attribute columns per data table is 4096.

DLC will recognize data files under the same COS path as data from the same table, please ensure data for separate tables is kept in separate folder hierarchies.

DLC does not support multi-version data in COS; it can only query the latest version of data in a COS bucket.

All tables created on DLC are external tables, and the SQL statement to create the table must include the EXTERNAL keyword.

Table names must be unique within the same database.

Table names are case-insensitive and only support letters, numbers, and underscores (_), with a maximum length of 128 characters.

If the table is a partitioned table, you must manually execute the ADD PARTITION statement or the MSCK statement to add partition information before you can query the partition data. For more details, see Query partitioned table. When creating a table with CSV, DLC will by default convert all field types to string, but this does not affect the computation and querying of raw data fields.

Data View Management

Last updated : 2024-07-31 17:28:41

DLC provides data view query capabilities, allowing users to quickly and easily perform data queries and use through the management of data views.

Create View

1. log in to DLC console, select the service region, log in users must have the permission to create views.

2. Enter the **Data Exploration page**, you can create views using SQL statements. For details of the statement, see SQL Syntax.

3. Select the computing resource, click the **Running** button to complete view creation.

Database Query Ø +	Query-2023-12-14 • + •	🗘 Storage c
Catalog DataLakeCatalog v	O Running Complet ▼ Ib Save C Refresh Ib Format Ib	😑 Select a default database 🔻 🔚 public-engine(SuperSQL-P 1.0-public)
Select a target database 💌	<pre>1 create or replace view db1.v1 as select x,y from tb1; 2 create view test_view (id comment 'test c1', name length comment 'test name c2') as select id, length(name) from test;</pre>	
• S :		
•		

View Views

You can view the view using SQL statements through Data Exploration, see SQL Syntax for specific syntax.

Meanwhile, DLC also offers a Visual Interface for managing views, with the following operations.

1. Log in to the DLC console, select the service region, log in users must have the permission to query views.

- 2. Enter the Data Management page, click on the Database Name where the view is located to enter the DMC page.
- 3. Click View to enter View Management.

ta table View	Function						Task history
Views under the data	atabase. To create a view, run view creation SQL statement	ts in Data Explore. Learn more 🗹					
Jodate time All	Last 7 days Last 30 days Select date	Select date					
w name 🗘	View type	Created at ‡	Update time	Description	Definition	Creator	Operation
		2023-04-10 21:11:19	2023-04-10 21:11:19		SELEC		Delete

4. Click the View Name you want to inspect to view its information. You can copy the SQL statement.

View info		
View name	t	
Created at	2023-04-10 21:11:19	
Definition	1 SELECT 2 `` 3 `` 4 `` 5 FROM 6 `datalakecatalog`.` 7 LIMIT 8 10 Copy statement	

Delete View

You can view the view using SQL statements through Data Exploration, see SQL Syntax for specific syntax.

Meanwhile, DLC also offers a Visual Interface for managing views, with the following operations.

1. Log in to DLC Console, select the service region, users must have view deletion permissions.

- 2. Enter the Data Management page, click on the Database Name where the view is located to enter the DMC page.
- 3. Click View to enter View Management, then click the Delete button to delete the view.

ata table	View	Function									Task history Sto
 Views und 	der the di	atabase. To create -	a view, run view creati	on SQL statements in Data Explo	re. Learn more	2					
Update time	All	Last 7 days	Last 30 days		t						
ew name 🏼 🗘			View type		Created at ‡		Update time	Description	Definition	Creator	Operation
- ro					2023-04-10 21	:11:19	2023-04-10 21:11:19		SELECT 'c	1	Delete
al items: 1						Delete ti	nis view		×		10 ¥ / page H 4 1 /
						After the v	iew is deleted, all of its data wi	I be permanently cleared.			
							Delet	Cancel			

Caution

Deleting a view will clear all data under the view and cannot be recovered. Please proceed with caution.

Function Management

Last updated : 2025-03-07 15:27:24

Data Lake Compute (DLC) supports using **user-defined functions** to process and build data, as well as managing functions.

Creating a Function

1. Log in to the DLC Console and select the service region. Ensure the logged-in account has database operation permissions.

2. Go to the Data Management Page and click the database name where you want to create the function.

Data management	© Guangzhou ▼ Bucket list			Use guide 💋 🛛	Data Management Ta
 This module allows 	you to manage databases under different catalogs. You can cl	ick the name of a database to manage its tables, views, and other data objects, or to ma	nually import data to tables. Learn more 🗹 . Data o	perations require relevant data permissions. For more permission guide, see here 🗹 .	
Create database	DataLakeCatalog 🔹			al	۵
Database name 🗘	Created at \$	Description	Creator	Operation	
	2022-10-09 19:25:58	-		Edit Delete	
<u>-1</u> 6	2022-09-05 16:29:27	-		Edit Delete	
	2022-06-30 00:24:32			Edit Delete	
	2022-06-08 11:56:03	1111		Edit Delete	
Total items: 4				10 💌 / page 🛛 H 🔄 1	/ 1 page

3. Select the **function**, then click the **Create Function** button to enter the function creation menu.

← Database / ε					
Data table View Func	tion				Task history Storage cor
 Custom Spark functions under 	the database. The .jar package	es can be uploaded to Data Lake Compute or mounted to your COS bu	ckets. Uploading to Data Lake Compute is recommended for central manag	gement of program packages. Learn more	2
Create function Update time	e All Last 7 days	Last 30 days Select date Select date			Enter a name
Function name	Created at \$	Update time Description	Program package name	Task status	Operation
	2022-12-12 20:01:42	2022-12-12 20:01:42	cosn	Created successfully	Edit Delete
Total items: 1					10 🔻 / page 🛛 🖌 🚽 1 👘 / 1 page

Description Optional Storage mode O Save on system O Mount on a specified COS path The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Program package source O Upload O COS File path * Select file Only a .jar package of up to 5 MB is supported Function class name * Enter a function class name	Function name *	Enter a function name
Storage mode Save on system Mount on a specified COS path The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Program package source Upload COS File path * Select file Only a .jar package of up to 5 MB is supported Function class name * Enter a function class name	Description	Optional
Program package source Upload COS File path * Select file Only a .jar package of up to 5 MB is supported Function class name * Enter a function class name		Save on system
File path * Only a .jar package of up to 5 MB is supported Function class name * Enter a function class name	Storage mode	The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path.
Only a .jar package of up to 5 MB is supported	Storage mode Program package source	 Development of a specified COS path The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Upload OCOS
Function class name * Enter a function class name	Storage mode Program package source File path *	The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Upload COS Select file
	Storage mode Program package source File path *	The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Upload COS Select file Only a .jar package of up to 5 MB is supported
	Storage mode Program package source File path * Function class name *	The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Upload COS Select file Only a .jar package of up to 5 MB is supported Enter a function class name
	Storage mode Program package source File path * Function class name *	 Gave on system Conduction a specified COS path The storage mode of the function package. You can upload and save the function package to the system (recommended), or directly save it at a specified COS path. Upload COS Select file Only a .jar package of up to 5 MB is supported Enter a function class name

The function packet supports local uploads or the use of existing JAR files in COS. Local uploads only support JAR format, with a maximum size of 5 MB.

Select the Spark cluster to run the function. There will be no fees incurred during the execution.

It is recommended to save the function package to the system for easy management and use. You can also mount it to a specified COS path.

Viewing Function Information

- 1. Log in to the DLC Console and ensure the account has database operation permissions.
- 2. Go to the **Data Management Page** and click the **database name** where the function is located.
- 3. Select the function to view its build status. If the build fails, you can edit and resubmit it.



← Database /						
Data table View Functio	'n					Task history Storage con
 Custom Spark functions under the 	e database. The .jar packages o	can be uploaded to Data Lake (Compute or mounted to your COS buckets. Uploa	ding to Data Lake Compute is recommended for central i	management of program packages. Learn more 🗳	
Create function Update time	All Last 7 days	Last 30 days Select	date Select date 🛅			Enter a name
Function name	Created at ‡	Update time	Description	Program package name	Task status	Operation
	2022-12-12 20:01:42	2022-12-12 20:01:42		cosn/	Created successfully	Edit Delete
Total items: 1						10 ▼ / page 🛛 🖂 1 / 1 page

4. Click the Function Name to view detailed information about the function.

Database /							Function details	
Data table View Function	on						Eunction name	
Custom Spark functions under the	e database. The jar packages o	can be uploaded to Data Lak	e Compute or mounted to your COS I	buckets. Uploading to Data	Lake Compute is recommended for central managem	ent of program packages. Learn m	Created at	2022-12-12 20:01:42
J					,		Description	
Create function Update time	All Last 7 days	Last 30 days Sele	ect date 🛛 Select date 🛅				Storage mode	Mount on a specified COS path
Function name	Created at \$	Update time	Description		Program package name	Task status	File path	cosn://
To and the second se	2022-12-12 20:01:42	2022-12-12 20:01:42			cosn://c	Created successfully		
Total items: 1							Function class name	q

Editing Function Information

1. Log in to the DLC Console and select the service region, and ensure the logged-in account has database operation permissions.

- 2. Go to the Data Management Page and click the database name where the function is located.
- 3. Select the **function**, then click the **Edit** button to enter the function information editing page.

← Database / 、										
Data table View Functio	n								Task hist	ory Storage configu
 Custom Spark functions under the 	e database. The .jar packages	can be uploaded to Data La	ke Compute or mounted to your	r COS buckets. Uploading to Da	ta Lake Compute is recommended for centr	al management c	of program packages. Learn mor	e 🖸		>
Create function Update time	All Last 7 days	Last 30 days Se	lect date Select date 🛅						Enter a name	Q
Function name	Created at \$	Update time	Description		Program package name		Task status		Operation	
· · · · · · · · · · · · · · · · · · ·	2022-12-12 20:01:42	2022-12-12 20:01:42			cosn://		Created successfully		Edit Delete	
Total items: 1									10 🔻 / page 🛛 H 🔄 1	/1 page →

Edit function				×
Function name *				
Created at	2022-12-12 20:01:42			
Description	0 [°] ,			
Storage mode	Save on system O Mount on a specified COS path The storage mode of the function package. You can upload a (recommended), or directly save it at a specified COS path.	and save the fu	nction package to the system	1
File path *	cosn://		Select a COS path	
Function class name *	Only a .jar package of up to 100 MB is supported.]		
Confirm	Cancel			

Currently, you cannot modify the function name, storage method, or upload method. If you need to change this information, you must recreate the function.

After the function information is modified, it will be rebuilt. Please operate with caution.

Deleting a Function

For functions that are no longer needed, you can delete them.

1. Log in to the DLC Console and select the service region. Ensure the logged-in account has database operation permissions.

- 2. Go to the Data Management Page and click the database name where the function is located.
- 3. Select the **function**, then click the **delete** button to remove the function that is no longer needed.

← Database /							
Data table View Function	n					Task history	itorage configuration
 Custom Spark functions under the 	database. The .jar packages ca	n be uploaded to Data La	ke Compute or mounted to your CO!	5 buckets. Uploading to Data Lake Compute is recommended for cent	ral management of program packages	. Learn more 🖪	×
Create function Update time	All Last 7 days	Last 30 days	lect date Select date 🛅				QØ
Function name	Created at ‡	Update time	Description	Program package name	Task status	Operation	
	2022-12-12 20:01:42	2022-12-12 20:01:42		cosn://	Created successfully	Edit Delete	
Total items: 1						10 💌 / page 🛛 🖂 1 🛛 /	1 page 🗼 🕅
		Delete	this function	×			
		After th	e function is deleted, all of its data w	ill be permanently cleared.			
			Delete	Cancel			

Note

After deletion, the data under this function will be cleared and cannot be recovered. Please operate with caution.

Partition Field Policy

Last updated : 2024-07-31 17:29:14

In Hive, partition information appears in the form of directories. In Iceberg, partition information is recorded in the underlying data files, making Iceberg's partitions more flexible and allowing the partitioning strategy to evolve with changes in data volume. In DLC, you can create Iceberg tables to utilize features such as hidden partitions. **Note:**

By default, native tables are Iceberg tables. External tables, depending on the file format, can choose between Hive or Iceberg tables. For detailed syntax, refer to the document CREATE TABLE.

With hidden partitions, when inserting and querying data, you do not need to specify partition information additionally as required in Hive.

Iceberg partition strategy supports the use of the following functions, with different fields and corresponding partition transformation strategies as shown in the table:

Partitioning Strategy	Field Type	Result Type
identity	Any	Source Type
bucket	int, long, decimal, date, time, timestamp, timestamptz, string, uuid, fixed, binary	int
truncate	int, long, decimal, string	Source Type
year	date, timestamp, timestamptz	int
month	date, timestamp, timestamptz	int
day	date, timestamp, timestamptz	date
hour	timestamp, timestamptz	int

Ops Management Historical Task Instances

Last updated : 2025-06-12 12:01:53

Historical Task Instances focus on recording and managing various types of tasks performed by users in DLC for subsequent tracking, review, and optimization. Through the Historical Task Instances feature, users can quickly view the execution status of tasks, including start and end times, execution status (such as successful or failed), input and output details, and generated logs or error information. It provides users with the convenience of auditing and retrieval, helping users identify task health status, potential issues, and optimize resource configuration, etc.

Operation Steps

1. Log in to Data Lake Compute (DLC) Console > Ops Management > Historical Task Instances and choose service region.

2. Enter the historical task instances page. Administrators can view all historical operation tasks in the past 45 days, and general users can query tasks related to themselves in the past 45 days.

3. Support filtering and viewing by task type, task status, creator, task time range, task name, ID, content, subchannel, and other methods.

4. Click the task ID/name. Support view task details, including modules such as basic information, running result, task insights, and task logs.

5. Support user click to modify task configuration, quickly enter job details to adjust configuration for optimization.

Historical Task Instances List

Note:

The *field supports after enabling the insight feature. For enablement method, please see How to Enable Insight

Feature.

Field Name	Description
Task ID	Unique identifier of the task.
Task name	Prefix_yyyymmddhhmmss_eight-digit uuid, where yyyymmddhhmmss is the task execution time. Prefix rule: 1. The job task submitted by the console is prefixed with the job name. For example, if the user-created job is customer_segmentation_job and it is executed at 21:25:10 on November



	 26, 2024, the task id will be customer_segmentation_job_20241126212510_f2a65wk1. According to the current data format restriction, the job name should be <= 100 characters. 2. SQL type submitted on the data exploration page, prefixed with sql_query. Example: sql_query_20241126212510_f2a65wk1. 3. Data optimization tasks, according to the prefixes of different sub-types of optimization tasks, among them: 3.1 The prefix of the optimizer is only optimizer. 3.2 The SQL type of the optimized instance is optimizer_sql. 3.3 The batch type of the optimized instance is optimizer_batch. 3.4 Configuration task created when configuring the data optimization policy: optimizer_config. 4. Import data task, prefixed with import, for example: import_20241126212510_f2a65wk1. 5. Export data task, prefixed with export, for example: export_20241126212510_f2a65wk1. 6. Wedata submission, prefixed with customized, for example: customized_20241126212510_f2a65wk1. 8. Tasks created for metadata operations on the metadata management page, prefixed with metadata, for example: metadata_20241126212510_f2a65wk1.
Task status	Starting Executing Queuing up Successful Failed Canceled Expired Task run timeout
Task content	Detailed content of the task. For job type tasks, it is a hyperlink to job details; for SQL type tasks, it is the complete sql statement.
Task type	Be divided into Job type, SQL type.
Task source	The origin of this task. Support data exploration tasks, data job tasks, data optimization tasks, import tasks, export tasks, metadata management, Wedata tasks, and API submission tasks.
Sub-channel	Users can customize sub-channels when submitting tasks via the API.
Compute resource	The computing engine/resource group used to run the task.
Consumed CU*H	During task execution, CU*H consumption occurs. Please note that the final CU consumption is subject to the bill, and the final result may vary. In the Spark scenario, it is approximately equal to the sum of Spark task execution durations divided by 3600.
Compute time	1. If the task supports insight feature, it is the execution time within the engine.

	 If the task does not support insight feature: 1 For a Spark SQL task, it is the platform scheduling time + consumed queuing time within the engine + execution time within the engine. 2 For a Spark job task, it is the platform scheduling time + engine startup duration + queuing time within the engine + execution time within the engine. The execution time within the engine is the duration from the start execution of the first task of a Spark task to the task completion.
Scanned data volume	The physical data volume read from storage by this task is approximately equal to the sum of Stage Input Size in Spark UI in the Spark scenario.
*Scanned data records	The number of physical data entries read from storage by this task is, in the Spark scenario, approximately equal to the sum of Stage Input Records in Spark UI.
Creator	If it is a job type task, it refers to the creator of the job.
Executor	The user running the task.
Submitted at	The time when the user submits tasks.
*Engine execution time	The time when the first preemption of the CPU starts execution of the task, the start execution time of the first task within the Spark engine.
*Number of output files	The collection of this metric requires upgrading the Spark engine kernel to a version later than 2024.11.16. Total number of files written by tasks through statements such as Insert. Case-insensitive to task type.
*Output small- sized files	The collection of this metric requires upgrading the Spark engine kernel to a version later than 2024.11.16. Small File Definition: An individual file size of the output that is less than 4 MB is defined as a small file (controlled by the parameter spark.dlc.monitorFileSizeThreshold, with a default value of 4 MB, which can be configured globally or at the task level for the engine). This metric definition: Total number of small files written by tasks through statements such as insert. Case-insensitive to task type.
*Total output lines	The number of records output after this task processes data is, in the Spark scenario, approximately equal to the sum of Stage Output Records in Spark UI.
*Total output size	The Size of the record output after this task processes data is, in the Spark scenario, approximately equal to the sum of Stage Output Size in Spark UI.
*Data shuffle lines	Approximately equal to the sum of Stage Shuffle Read Records in Spark UI in the Spark scenario.
*Data shuffle size	Approximately equal to the sum of Stage Shuffle Read Size in Spark UI in the Spark scenario.



*Health status

Analyze the task to judge the health status of the task and determine whether optimization is required. Please see Task Insight for details.

Historical Task Instances Details

Basic Info

1. Users can view specific task content in **execution content**. For SQL tasks, view the complete SQL statement; for job tasks, view job details and job parameters.

2. Users can view relevant content about task resources in **resource consumption**, including consumed CU*H, computational overhead, scanned data volume, compute resource, kernel version, Driver resource, Executor resource, and count of Executors.

3. Users can view basic information of tasks in **basic info**, including task name, task ID, task type, task source, creator, executor, submission time, and engine execution time.

4. For tasks running on the SuperSQL SparkSQL or SuperSQL Presto engine, users can view the task running progress bar in **query statistics**, which includes the time taken for stages such as creating tasks, scheduling tasks, executing tasks, and obtaining results.

Running Result

After task completion, users can query the task result on the execution result page. There are two types of task results:

1. Write file information: For file writing tasks running on SuperSQL, standard engine, or Spark kernel engine, support user viewing of write file information.

Average file size

minimum file size

maximum file size

Total file size

2. Execution result: SQL task query statement, which can display the query result of the current task and support users to download query results.

Task Insight

After task completion, users can view task insight results on the task insight page. It supports analyzing the aggregate metrics that each task has executed and insights into optimizable issues. Based on the actual execution situation of the current task, DLC task insight will combine data analysis and algorithm rules to provide corresponding optimization suggestions. For details, please see Task Insight.

Task Log

Users can view the logs of the current task on the task log page.

Note:

Only the job type supports task log viewing.

1. Support switching logs of nodes in different clusters through Pod Name, including Driver, Executor, etc.

2. Support three log level filters: All, Error, Warning.

3. This page only displays the last 1000 logs. If you need to view all log entries, you can export logs.

4. Support viewing log export records and the status of export tasks. In log export records, users can save log files locally.

Historical task(Old version)

Last updated : 2025-03-21 12:22:27

To facilitate users in querying historical task records, DLC provides three methods to search and process historical tasks.

View historical tasks run in the Query Editor

1. Log in to DLC console, select the service region.

2. Enter the **Data Exploration Page**, click on **Run History** within a single Session to view the task run history for that Session.

3. Click on the history record Batch ID to view the corresponding execution results on the left

٢	Data Explo	re 🔇 Hon	g Kong 🔻											Data explor
	Database	Query	¢ +	•	SessionNam Draft x	SessionName4fcbb ×	SessionName81af2 ×	SessionNamecdd7 ×	SessionNamef1f2d ×	SessionName60646 ×	SessionNamedcc7 ×	SessionName09a8	b x → +	▼ \$torage
Q	Catalog DataLa	keCatalog	Ŧ	Ø		C 🖏 🗄 🛤 🔻						≎⊷ -	Si suka ng ali	d sandaras
≣	Select a target o	atabase	Ψ.	•	15-16-16/16-									
	1. 5 4													
E	< 36 C													
₿	 Personal 	ብ ምርሳትሌ <u>ም</u> ን												
d	 Pjejár svoj 	Non page												
_	n State and	8.8 až s												
59	n di sedar T													
 	t Gestare. 													
~	i (Tra Sovi) I Marcana													
	. Carro	:												
_	- 5a-40													
ď	· Serier													
6	المراجع ال	-												
Ö	- Çesnav	,												
₽	L Russer	8		Que	ery result								Run history Do	wnload history
	 Zersines 	1		Re										
	t Saletsan	5		sult	Query time 338ms Billat	ble scanned volume 34.0 MB	0							
	ト ね-Vz-ost				$\operatorname{Conv} \operatorname{She}_{2}(\mathbb{R} \otimes \mathbb{R})$	h Geben ⊈, weglew	R							
	 Q² (s² a) set 			Que	Dig Najawan di					ave Viti 1				
	 All articles 			ry										
	i ayan yeye			St	in the second									
IJ	L Prais and	-		atistic	2 af 40darae	SCIENCENTRI AMPLEX	s Hogic (Jan 1973), Hogi							

Each Session's run history is independent, and a maximum of 45 days of run history is kept.

Historical task result data is saved for 24 hours. To view task results beyond 24 hours, the task must be rerun.

View data import history in the Data Management feature

1. log in to DLC Console > Data Management, select the service region.

Note:

Log in to the account requires database-related permissions.

- 2. Click on **Task History** in the top right corner to query data import history tasks.
- 3. Supports viewing historical tasks from the past 45 days

٢	Data job	🔇 Guangzhou 🔻									_
	Spark job	Job configuration Sess	ion management								Job monitoring Task history Lo
©	Create job	in the second se	O, All	* All	Ŧ				All	Last 7 days Last 30 days	Select date Select date
:= 72	Job name	Job ID	Job ty	type	Job file	Current tasks	Task engine \$	Creator \$	Created at \$	Update time \$	Operation
	(11) B		6				h (j)		2023+10+08 20:10:22	2023-10-11 19:01:43	Monitor Edit Running I
Ø	Total items: 1									10 💌 / page	I ≤ 1 /1 page
٩											

View historical tasks in the Historical Operation feature

- 1. log in to DLC Console > Historical Operation, select the service region.
- 2. Enter the Historical Operation page, where administrators can view all historical operation tasks from the past 45 days, and ordinary users can query tasks related to themselves from the past 45 days.
- 3. Supports filtering by task type, execution status, creator, data type, etc.

٢	Run history 🕲 Guangzho	u v								Histori
## ©	This module displays the statu	s of tasks submitted in other modules, in	cluding SQL tasks and data import/exp	ort tasks. An admin can query all	tasks in the last 45 days, while a ge	neral user can query tasks related to them i	n the last 45 days.Learn more g	2		
≣	9 Q	Q Select an execution status	Select a job or task creator 💌	Select a data engine	Select a task type	▼ Batch operat ▼		Today Last 7	lays Last 30 days	2023-12-18 ~ 2023-12-18
	Job overview									
ଦ (କ	^{лн}		Executing			Queuing up		Initialize		
	Task ID	Task fyne	Task content	Execution status	Creator	Task submission time \$	Data engine	Resource usage \$	Kernel version	Operation
		S	tiog To	Successful	u (i)	2023-12-18 17:33:28		-		Learn more
	Total items: 1								10 👻 / page	н н 1 /1 раде

4. click **Run Details** to see the task execution details and results.

	Running result Query statistics
k ID	21 FD
ator	
k type	SQL statement
nel sion	
ta engine	All set in the set of
nning time	2023-12-18 17:33:28to2023-12-18 17:33:31
k time	1.5s
inned data ume	
a entries	U
ery tement	
SELEC FROM Da	taLakeCatalog`.`
LIMIT 10	

Historical task result data is saved for 24 hours. To view task results beyond 24 hours, the task must be rerun. You can directly **Copy Statement** to Data Exploration to execute the task.

You can directly click **Task ID** to quickly switch and view the task execution details.

For tasks that are running, you can **Cancel** them.

Session Management

Last updated : 2025-03-21 12:22:27

The session management feature is used to record and trace notebook interactive sessions submitted to the DLC engine through the API or Wedata. Users can perform operations such as SQL queries, data processing, and model training through sessions.

Prerequisites

Environment preparation for Data Lake Compute (DLC). Enable Tencent Cloud DLC engine service. Creating a session requires purchasing a job type engine. SuperSQL job engine. Standard engine: Spark engine or machine learning resource group.

Operation Steps

- 1. Log in to DLC Console > Ops Management > Session Management and choose service region.
- 2. Enter the session management page, and users can view all the historical session records.
- 3. Support filtering and viewing by engine type, status, Kind, engine name, Session ID, and Session Name.
- 4. Click Session Name/ID. View session details is supported.
- 5. Support users to click kill to close the session on the console.
- 6. Support user viewing of the Spark UI of the session.

Session List

Field Name	Description
Session Name/ID	Unique identifier for the session. Sessions created by the SuperSQL job engine only have a Session ID. Session ID rule: livy-session-uuid. Sessions created by the standard engine or Spark engine User-submitted Notebook, prefixed with session_test User-submitted batch SQL, prefixed with temporary-rg
Status	State of the current session, can be divided into



	 not_started: The session has not been started. This status indicates that the session request has been accepted, but the session has not yet started for some reason (for example, insufficient resources or configuration problems). Users need to check related configurations or resource status to start the session. Starting: The session is starting. This status means Livy is allocating resources and initializing the environment for a new Spark session. idle: The session has started successfully and is in idle state. At this point, you can submit Spark jobs. The Livy session is ready to process requests. busy: The session is processing one or more jobs. This status indicates that the session is executing tasks and cannot accept new job requests until the current job is completed. shutting down: The session is deactivating. This status means the user has requested to stop the session and Livy is performing clearing and resource release operations. The session may stay in this status for a period of time until all running jobs are completed and resources are released. error: The session is unable to function normally, possibly due to insufficient resources, configuration errors, or other problems. dead: The session has died and cannot be recovered. killed: The session has died and cannot be recovered. success: The session has been successfully completed. This status usually indicates that all jobs in the session have been successfully executed and completed. The session can be considered successful y executed and completed. The session can be considered successful in this status, and users can view the results or output.
Engine	Computing engine.
Kind	Session type Spark Pyspark SQL Machine Learning Python MLlib
Creator	The user who creates a session.
Validity period	The running time of the session.

Insight Management Task Insights

Last updated : 2025-04-17 15:22:36

Task insights are made from the task perspective, helping you quickly identify the completed tasks for analysis and providing optimization suggestions.

Prerequisites

- 1. SuperSQL SparkSQL and Spark job engines:
- 1. For engines purchased after July 18, 2024, task insights are enabled by default.
- 2. For Spark kernel versions prior to July 18, 2024, the engine kernel should be upgraded to enable task insights. For details on upgrading, see How to Enable Insights.
- 3. Standard Spark engine:
- 1. For engines purchased after December 20, 2024, task insights are supported by default.
- 2. For engines purchased before December 20, 2024, manual activation of task insights is not supported. Submit a ticket to contact after-sales service for activation.

Other types of engines do not support task insights currently.

Directions

Log in to the DLC Console, select the Insight Management feature, and switch to the task insights page.

Insights Overview

Daily-level statistics offer insights into the distribution and trend of tasks requiring optimization, providing a more intuitive understanding of daily tasks.

Task Insights


The task insights feature supports analyzing the summary metrics of each executed task and identifying the possible optimization issues.

After a task is completed, users only need to select the task to be analyzed and click **Task Insights** in the operation column to view the details.

Based on the actual execution of the current task, DLC task insights leverage data analysis and algorithmic rules to provide the corresponding optimization recommendations.

How to Enable the Insights Feature?

Upgrading Kernel Image for Existing SuperSQL Engines

Note:

For engines purchased after July 18, 2024, or existing engines upgraded to kernel versions after July 18, 2024, Insights are automatically enabled. You can skip this step.

Directions

1. Go to the SuperSQL Engine list page and select the engine for which you want to enable the insights feature.

2. On the engine details page, click **Kernel version management > Version upgrade** (default upgrade to the latest kernel version).

Overview of Key Insight Metrics

Metric Name	Metric Definition
Engine execution time	Reflects the time the first task was executed on the Spark engine (the time when the task first preempted the CPU for execution).
Execution time within the engine	Reflects the time actually required for computing, namely, the time taken from the start of the first task execution in a Spark task to the completion of the Spark task. More specifically, it is the sum of the duration from the start of the first task to the completion of the last task for each Spark stage. This sum does not include the queuing time of the task before it starts (that is, excluding other time such as the time required for scheduling between task submission and the start of execution of the Spark task), nor include the time spent waiting for task execution due to insufficient executor resources between multiple Spark stages during the task execution process.
Queuing time (time spent	Specifies the time taken from task submission to the start execution of the first



waiting for execution)	Spark task. The time taken may include the cold startup duration of the first execution of the engine, the queuing time caused by the concurrent limit of the configuration task, the time waiting for executor resources due to full resources within the engine, and the time taken to generate and optimize the Spark execution plan.
Consumed CU*H	Specifies the sum of the CPU execution duration of each core of the Spark Executor used in computing, per hour (not equivalent to the duration of starting machines in the cluster, because the machines may not participate in task computing after they start. Eventually, the cluster's CU fee is subject to the bill). In the Spark scenario, it approximately equals to the sum of the execution durations of the Spark task (seconds) / 3600 (per hour).
Data scan size	The amount of physical data read from storage by this task. In the Spark scenario, it approximately equals to the sum of the Stage Input Size in Spark UI.
Total output size	The size of the records output after this task processes the data. In the Spark scenario, it approximately equals to the sum of the Stage Output Size in Spark UI.
Data shuffle size	In the Spark scenario, it approximately equals to the sum of the Stage Shuffle Read Records in Spark UI.
Number of output files	(This metric requires the Spark engine kernel to be upgraded to a version after November 16, 2024)The total number of files written by tasks through statements such as insert.
Number of output small files	(This metric requires the Spark engine kernel to be upgraded to a version after November 16, 2024)Small files are defined as output files with a size less than 4 MB (controlled by the parameter spark.dlc.monitorFileSizeThreshold, default 4 MB, configurable at the engine or task level). This metric represents the total number of small files written by tasks through statements such as insert.
Parallel task	Displays the parallel execution of tasks, making it easier to analyze affected tasks (up to 200 entries).

Overview of Insight Algorithms

Insight Type	Algorithm Description (Continuously Improving and Adding New Algorithms)
Resource preemption	SQL execution task delay time is greater than 1 minute after stage submission, or delay exceeds 20% of the total runtime (the threshold formula dynamically adjusts based on task runtime and data volume).



Shuffle exception	Stage execution encounters shuffle-related error stack information.
Slow task	Task duration in a stage is greater than twice the average duration of other tasks in the same stage (the threshold formula dynamically adjusts based on task runtime and data volume).
Data skew	Task shuffle data is greater than twice the average shuffle data size of other tasks (the threshold formula dynamically adjusts based on task runtime and data volume).
Disk or memory insufficiency	Error stack information during stage execution includes OOM, insufficient disk space, or COS bandwidth limitation errors related to disk or memory insufficiency.
Excessive small file output	 (This insights type requires the Spark engine kernel to be upgraded to a version after November 16, 2024)See the metric number of output small files in the list, and the presence of excessive small file output is determined if any of the following conditions are met: 1. Partitioned tables: The number of small files written out by a partition exceeds 200. 2. Non-partitioned tables: The total number of output small files exceeds 1000. 3. If partitioned or non-partitioned tables output more than 3,000 files with an average file size less than 4 MB.

System Management User and Permission Management CAM Service

Last updated : 2025-01-03 15:27:28

Data Lake Compute has a complete data access control mechanism and divides permissions into operation permissions and data permissions. The former is managed by CAM, while the latter is managed by the permission module of Data Lake Compute.

A root account has all the operation and data permissions of Data Lake Compute by default.

If a sub-user is granted the operation permissions of Data Lake Compute, the sub-user can grant the data permissions to other sub-users and can be regarded as an "admin" of this type of sub-users.

If a sub-user is granted the data read/write permissions, the sub-user can query data as permitted. The data permissions are granted by an "admin".

The data permissions of all sub-users other than root accounts are granted by an "admin". They cannot query data which they don't have permissions on.

A root account has all the operation permissions of Data Lake Compute by default and can grant sub-users the access permissions of Data Lake Compute through CAM, so that the sub-users can have corresponding operation permissions of Data Lake Compute.

Directions

1. Create and authorize a sub-user.

In the CAM console, create a sub-user and grant permissions as instructed in Sub-user authorization.

Preset policy QcloudDLCFullAccess : All the operation permissions in Data Lake Compute.

Custom policy: Specified operation permissions of Data Lake Compute.

2. Log in to the Data Lake Compute console with a sub-user account and verify the permissions.

If the operation succeeds, the authorization has taken effect.

Operation permission category

Data Lake Compute operation permissions are categorized by API as follows.

Permission Type	Description
Metadata	Manipulate the metadata information of databases and data tables managed in Data

management	Lake Compute.
Task management	Submit and view tasks in Data Lake Compute.
Permission management	Manage users' data access permissions.
System configuration	Perform basic configurations of the Data Lake Compute service.

Sub-user authorization

If you access Data Lake Compute as a root account, skip this step.

1. Create a sub-account as instructed in Creating and Authorizing Sub-account.

2. Create a custom policy.

On the Policies page in the CAM console, click Create Custom Policy.

In the pop-up window, click Create by Policy Syntax.

On the Create by Policy Syntax page, select Blank Template and click Next.

In the template, enter the **Policy Name** (e.g., DLCDataAccess) and **Description**, copy the following policy, paste it into **Policy Content**, and click **Complete**. A sub-user bound to the custom policy can log in to the Data Lake Compute console to run SQL tasks but cannot manage data permissions. For more information, see Sub-Account Permission Management.

```
{
  "version": "2.0",
  "statement": [
      {
          "effect": "allow",
          "action": [
              "dlc:DescribeStoreLocation",
              "dlc:DescribeTable",
              "dlc:DescribeViews",
              "dlc:CancelTask",
              "dlc:CreateDatabase",
              "dlc:CreateScript",
              "dlc:CreateTable",
              "dlc:CreateTask",
              "dlc:DeleteScript",
              "dlc:DescribeDatabases",
              "dlc:DescribeScripts",
              "dlc:DescribeTables",
              "dlc:DescribeTasks",
              "dlc:DescribeQueue"
          ],
```



```
"resource": [
"*"
]
}
]
```

Select Pol	licy Template > 2 Edit Policy
Policy Name *	DLCDataAccess
Aft	ter the policy is created, its name cannot be modified.
Description	
Policy Content	Use Legacy Version
1	
2 "ve	ersion": "2.0",
3 "st	tatement": [
4	{
5	"effect": "allow".
6	"action": [
7	"dlc:DescribeStoreLocation".
8	"dlc:DescribeSchielderon",
0	"dlc:DescribeViews"
10	"dlc:CancelTask"
11	"dlc:CneateDatabase"
12	"dlc:CreateScrint"
12	"dlc:CreateTable"
14	"dle:CreateTack"
14	"dle:DeleteSeriet"
15	alc:Deletescript ,
16	"dlc:DescribeDatabases",
17	"dlc:DescribeScripts",
	"dlc:DescribeTables",
18	"dlc:DescribeTasks",
18 19	"dlc:DescribeQueue",
18 19 20	
18 19 20 21	"dlc:DescribeTaskResult"
18 19 20 21 Policy Syntay Descr	<pre>"dlc:DescribeTaskResult" intion P2 CAM-enabled Services P2</pre>

3. Bind the preset or custom policy to a sub-account, and the sub-account can log in to and access Data Lake Compute. For more information, see Setting Sub-user Permissions.



Preset policy: QcloudDLCFullAccess .

Custom policy: The policy customized in the above steps for accessing Data Lake Compute.

Permission Overview

Last updated : 2024-07-17 15:42:58

Data Lake Compute permissions include data permissions and data engine permissions. If you have the admin permission, you can log in to the Data Lake Compute console or use an API to grant a sub-user data and data engine permissions. Sub-users cannot use, modify, or delete data or data engines before they are authorized.

User and work group

Data Lake Compute provides the user mode and work group mode for personnel permission management.

User: You can select users in CAM, including sub-accounts and collaborator accounts.

Work group: It is a group of users with the same permissions managed in the product.

Note:

If users are granted different permissions from those granted in their work groups, all the granted permissions will take effect.

A work group allows you to quickly grant permissions to a batch of users, so it is recommended for batch user authorization. For detailed directions, see User and User Group.

User type

In Data Lake Compute, User type can be Admin or General user.

Admin: An admin have all the data, engine, and task permissions and can add, authorize, and remove users and work groups in Data Lake Compute.

General user: A general user is added by an admin, has no Data Lake Compute permissions by default, and needs to be authorized. Only data and engine permissions that can be **regranted** can be granted to general users.

Permission and Operation	Admin	General User
Data permissions	All	None by default (to be authorized by an admin)
Data engine permissions	All	None by default (to be authorized by an admin)
User management	Yes	No
Work group management	Yes	No
Authorization scope	All	Permissions that can be regranted



Note:

The above permissions only include those defined in Data Lake Compute. To perform purchase, configuration adjustment, and refund operations that involve billing, log in to the CAM console and get the financial collaborator permission <code>QCloudFinanceFullAccess</code> (for detailed directions, see Creating and Authorizing Sub-account).

Data permissions

Data Lake Compute data permissions allow operations on data catalogs, databases, and data tables. To facilitate your management and configuration, permissions can be granted in the standard or advanced mode. In standard mode, you can grant roles while ignoring the specific permission configuration (for more information on roles and permissions, see Sub-Account Permission Management). The authorization granularity can be data catalog, database, or data table. This mode is suitable for quick authorization with no complex permission management involved.

In advanced mode, you can grant permissions at the database, data table, view, or function level. It is suitable for refined permission management.

CREATE DATABASEIIIICataglogALTER DATABASEIIIIIDatabaseDROP DATABASEIIIIIDatabaseDROP DATABASEIIIIIDatabaseIIIIIIIDatabaseDROP TABLEIIIIIDatabaseIIIIIIIDatabaseDROP TABLEIIIIIIDatabase/TableII <td< th=""><th>Action</th><th>CREATE</th><th>ALTER</th><th>DROP</th><th>SELECT</th><th>INSERT</th><th>DELETE</th><th>Target</th></td<>	Action	CREATE	ALTER	DROP	SELECT	INSERT	DELETE	Target
ALTER DATABASE·····DatabaseDROP DATABASE·······DatabaseCREATE ABLE AS SELECT·······DatabaseCREATE ABLE AS SELECT·······DatabaseDROP TABLE········Database/TableDROP TABLE·········Database/TableDROP TABLE··········Database/TableDROP TABLE··········Database/TableDROP TABLE··········Database/TableDROP TABLE··········Database/TableDROP TABLE···············Database/TableDROP TABLE···	CREATE DATABASE	1	-	-	-	-	-	Cataglog
DROP DATABASE·····DatabaseCREATE TABLE AS SELECT······Database/TableDROP TABLE········Database/TableDROP TABLE········TableALTER TABLE········TableALTER TABLE··········ALTER TABLE············ALTER TABLE··· </td <td>ALTER DATABASE</br></td> <td>-</td> <td>1</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>Database</td>	ALTER 	-	1	-	-	-	-	Database
CREATE ABLE AS SELECT·····DatabaseCREATE TABLE AS 	DROP DATABASE	-	-	✓	-	-	-	Database
CREATE TABLE AS SELECTImage: SelectImage: Sele	CREATE TABLE	1	-	-	-	-	-	Database
DROP TABLETableALTER TABLE LOCATIONTableALTER PARTITION LOCATIONALTER 	CREATE TABLE AS SELECT	✓	-	-	✓	✓	-	Database/Table
ALTER TABLE LOCATIONTableALTER PARTITION LOCATIONTable	DROP TABLE	-	-	1	-	-	-	Table
ALTER PARTITION LOCATION	ALTER TABLE	-	1	-	-	-	-	Table
	ALTER PARTITION LOCATION	-	✓	-	-	-	-	Table

SQL statements for permission operations are as follows:

🔗 Tencent Cloud

ALTER TABLE ADD PARTITION	-	1	-	-	-	-	Table
ALTER TABLE DROP PARTITION	-	1	-	-	-	-	Table
ALTER TABLE	-	1	-	-	-	-	Table
CREATE VIEW	1	-	-	-	-	-	Database
ALTER VIEW PROPERTIES	-	1	-	-	-	-	View
ALTER VIEW RENAME	-	✓	-	-	-	-	View
DROP VIEW PROPERTIES	-	✓	1	-	-	-	View
DROP VIEW	-	-	1	-	-	-	View
SELECT TABLE	-	-	-	1	-	-	Table
INSERT	-	-	-	-	1	-	Table
INSERT OVERWRITE	-	-	-	-	\checkmark	1	Table
CREATE FUNCTION	<i>√</i>	-	-	-	-	-	Database
DROP FUNCTION	-	-	1	-	-	-	Function
SELECT VIEW	-	-	-	1	-	-	View
SELECT FUNCTION	-	-	-	1	-	-	Function

Data engine permissions

Data Lake Compute data engine permissions allow using, modifying, manipulating, monitoring, and deleting data engines as detailed below:

Use: The permission to use engines to perform tasks.

Modify: The permission to modify the basic information and configuration information of engines (modifying the

configuration information requires the CAM financial collaborator permission).

Manipulate: The permission to suspend and restart engines.

Monitor: The permission to view the running tasks and monitoring information of engines.

Delete: The permission to return engines.

Permission granting

A single user can be granted multiple permissions. For detailed directions, see Sub-Account Permission Management.

User and Work Group

Last updated : 2024-07-17 15:44:57

Data Lake Compute provides the user mode and work group mode for personnel permission management. For more information on permissions, see Permission Overview.

Description

User: You can select users in CAM, including sub-accounts and collaborator accounts.

Work group: It is a group of users with the same permissions managed in the product.

Note:

If users are granted different permissions from those granted in their work groups, all the granted permissions will take effect.

A work group allows you to quickly grant permissions to a batch of users, so it is recommended for batch user authorization.

User Management

User management requires Data Lake Compute operation permissions. For more information, see CAM Service.

Adding a user

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

2. Click Add user to add an account with a specified user ID to Data Lake Compute for management.

Permission management	🔇 Guangzhou 🔻					Use guide Ø Permission manageme
User Work group						
Both sub-account and coordi	inator users need to be granted with permissions to use	edata, engines, and other resources. A user may be ass	sociated with one or more work groups to inherit all of	their permissions. An admin user has all resource permi	ssions. For more permission guidelines, see here 🖄 .	
Add user Batch delete						Enter a user ID or name
User ID	Username	User type 🕥	Description	Added by	Add time \$	Operation
1				-	2023-12-06 17:35:24	Edit Authorize 🔻 Delete
					2023-03-07 16:56:41	Edit Authorize 👻 Delete
Total items: 2						10 v / page 🛛 K 🔍 1 / 1 page



3. After entering the **User ID**, bind the user to a work group (which requires the admin permission). If binding is not needed, directly click **Complete**.

← Add u	iser
1 Basi	c info > (2) Blind work group
User ID	
Username	Enter a username 0
User type	Seted •
Description	An admit has all permissions for all resources (including data and engines), and can manage other admits eacept the root account use; A general user needs to be granted with neevest permissions or associated with a work group to access corresponding resources.
4 Back	Net 1

Viewing user information

A Data Lake Compute admin can modify the basic information and permissions of a user.

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

2. Search for the target **User ID** and click the **Username** to view the user information and permissions.

Permission management	🖏 Guangzhou 👻				View user
User Work group					User ID
Both sub-account and coordin	nator users need to be granted with permissions to use da	ata, engines, and other resources. A user may be assi	ociated with one or more work groups to inherit all of	their permissions. An admin user ha	Username
Add user Batch delete					User type C
User ID	Username	User type ①	Description	Added by	Description
1				1 4	Tork group Data permasion angine permasion
1	4		i i i i i i i i i i i i i i i i i i i	-	Work group name Description Add time * Added by
Total items: 2					
					No data
					Total items: 0 10 💌 / page 🔣 4 1 / 1 page >>

Editing user information

You can edit the description and work group of a user. For detailed directions, see Sub-Account Data Authorization. 1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page. 2. Search for the target user account ID and click Edit in the Operation column to enter the edit page.

Removing a user

If you don't want a user to use Data Lake Compute any more, you can use an admin account to remove the user.

Then, the Data Lake Compute permission granted to the user will be revoked.

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

2. Search for and select one or multiple target user account IDs and click **Batch remove** to remove them from Data Lake Compute.

Permission management	🖏 Guangzhou 🔻					Use guide Ø Permission manageme
User Work group						
Both sub-account and coordi	nator users need to be granted with permissions to use d	sta, engines, and other resources. A user may be ass	ociated with one or more work groups to inherit all of t	heir permissions. An admin user has all resource permiss	ions. For more permission guidelines, see here 🗳 .	
Add user Batch delete]					Enter a user ID or name
Jser ID	Username	User type (i)	Description	Added by	Add time \$	Operation
		-	-		2023-12-05 16:49:02	Edit Authorize 🔻 Delete
		1000			2023-11-22 14:53:11	Edit Authorize 🔻 Delete
		-		1000	2023-10-23 11:46:29	Edit Authorize 🔻 Delete

Work Group Management

Work group management requires Data Lake Compute operation permissions. For more information, see CAM Service.

Adding a work group

You can manage permissions that need to be repeatedly granted to users through a work group. The following describes how to add a work group.

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

- 2. Click **Work group** to enter the work group management page.
- 3. Click Add work group, enter relevant information, and click Confirm.

Permission manager	ment 🔇 Guangzhou 🔻	r				Use guide 🧭 🛛 Permission manageme
User Work grou	p					
Batch add users to	o a work group to batch grant ther	n the permissions of	data, engines, and other resou	rces of this work group. There is no need to add an admin to a work group. For more permission g	guides, see here 😰 .	
Add work group	Batch remove					Enter a work group name
Work group ID	Work group name	User count	Description	Added by	Add time \$	Operation
					2023-08-22 14:30:42	Edit Authorize 🔻 Remove
🗆 😡				Distance in the second s	2021-11-10 13:32:11	Edit Authorize 🔻 Remove
Total items: 2						10 🕶 / page 🛛 K 🔍 1 👘 / 1 page

Viewing work group information

You can view the information of a work group in the following steps:

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

2. Click **Work group** to enter the work group management page.

3. Search for the target work group and click **Work group ID** or **Work group name** to view the work group information.

Permission managem	ent 🕲 Guangzhou	r				View wo	ork gro	up					
User Work group	-					Work grou	ip name						
Batch add users to a	work group to batch grant ther	n the permissions o	f data, engines, and other resources	of this work group. There is no need to add an admin to a work grou	up. For more permission guides, see here 😰 .	Descriptio	n						
Add work group	Batch remove					User	Dat	ita permission	Engine pe	rmission			
Work group ID	Work group name	User count	Description	Added by	1	Catalog,	Datab	Select	*	Permission source			
			1000	10.00	2	Select	¥						
				100000	2	Permiss	ion	Catalog	Database	Table/View/Fun.	Column	Permission (Permis
Total items: 2						Databas	e per	DataLakeCatalog					
						Total Iter	ms: 1				10 🔻 / page	: H 4 1	/ 1 page
						Row-lev	el pern	missions					
						Permission	n source	Select 🔻				Enter a database or tab	le name
						Permiss	ion	Catalog	Database	Data table	Row filte	er expression	
										No d	əta		
										·			
						Total ite	ms: 0				10 🔻 / page		/ 1 page

Editing work group information

You can modify the description and users of a work group in the following steps:

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

- 2. Click **Work group** to enter the work group management page.
- 3. Find the target Work group name and click Edit in the Operation column.



÷	Edit work group						
	Work group name 1						
	Bind user Batch remove	An associated user will obtain	all permission of this work group				
	User ID	Username	User type	Description	Add time \$	Added by	Operation
					2023-08-22 14:57:06		Remove
					2023-08-10 20:57:13		Remove
	Total items: 2						10 v / page K 4 1 / 1 page >

To edit the description, click



You can click **Bind user** to add Data Lake Compute users to the work group.

Select multiple target users and click **Batch remove**, or click **Remove** in the **Operation** column of a specific target user. Removed users will no longer have the permissions of the work group, which does not affect other permissions granted to them though.

Deleting a work group

A Data Lake Compute admin can remove work groups.

Note:

After a work group is removed, all its permissions granted to users in it will be revoked. Note that a removed work group cannot be recovered. Proceed with caution.

1. Log in to the Data Lake Compute console, select the service region, and go to the **Permission management** page.

2. Click **Work group** to enter the work group management page.

3. Select multiple target work groups and click **Batch remove**, or click **Remove** in the **Operation** column of a specific target work group.

Permission manageme	ent 🔇 Guangzhou 🔻						Use guide 💋	Permission manageme
User Work group								
Batch add users to a	work group to batch grant them	the permissions of	f data, engines, and other resources of this work group. There is no need to	add an admin to a work group. For more permission guides, see here 😫 .				
Add work group	Batch remove						Enter a work g	group name
Work group ID	Work group name	User count	Description	Added by	Add time \$	Operation		
•					2023-08-22 14:30:42	Edit Authorize 🕈	Remove	
<u> </u>					2021-11-10 13:32:11	Edit Authorize 🕈	Remove	
Total items: 2						10 v /	page H ┥	1 / 1 page

Sub-Account Permission Management

Last updated : 2024-07-17 15:46:12

User permission

User permissions include data permissions and engine permissions (for more information on permissions, see Permission Overview). The former is required to access data in Data Lake Compute, while the latter is used for resource management. Data Lake Compute enables permission management at the database, table, and column levels, so that you can authorize a user or work group for refined data permission management in different use cases.

User and work group

You can authorize a user or create and authorize a work group of users. For detailed directions, see User and Work Group.

User: You can select users in CAM, including sub-accounts and collaborator accounts.

Work group: It is a group of users with the same permissions managed in the product.

Note:

If users are granted different permissions from those granted in their work groups, all the granted permissions will take effect.

A work group allows you to quickly grant permissions to a batch of users, so it is recommended for batch user authorization.

Granting a user a permission

Grant permissions to the specified user.

1. Set a user to **Admin** or **General user**. Admins have the permissions of all the data and engines by default with no need to be bound to a work group. They can also manage admin users other than the root account. **Set an admin with caution.**



← Add u	·
1 Basic	fo > (2) Bind work group
User ID	
Username	Enter a username D
User type	Select 👻
Description	Enter a description
4	
Back	Ned

2. Bind a work group: General users need to be granted permissions or bound to a work group before they can access resources.

to a work group grants the user all permit	ssions on the group.		
me Description	Add time \$	Added by	Operation
	-		
		10 v / page H 4 1 / 1	page 🕨 🕨
	to a work group grants the user all permis	to a work group grants the user all permissions on the group.	to a work group grants the user all permissions on the group. Imme Description Add time Added by

3. Add a data permission: In the User list, click Authorize in the Operation column and select Data permission to grant permissions at the data catalog or database/table level.

🔶 Add user					
Basic info > 2 Bind work gr	oup				
Bind work group Batch remove Binding	g a user to a work group grants the user all permissions on the group.				
Vork group ID	Work group name	Description	Add time \$	Added by	Operation
	10	-	2023-12-07 15:37:30		Remove
Total items: 1				10 ¥ / page	H ← 1 /1 page →

Add a data catalog permission. You can grant permissions to create databases under DataLakeCatalog and create other data catalogs.

Permission type O Catalog O Database & table The catalog option covers permissions to create databases under DataLakeCatalog and other catalogs, while th table option covers permissions of databases, data tables, views, and functions.	
The catalog option covers permissions to create databases under DataLakeCatalog and other catalogs, while th table option covers permissions of databases, data tables, views, and functions.	
	e database anc
Permission Create database under DataLakeCatalog Create catalog	
Authorizable Yes	

Add a database/table permission: You can grant permissions in **Standard** or **Advanced** mode. In standard mode, you can grant database/table permissions in the specified catalog and set **Query & analytics**, **Data edit**, and **Owner** permissions.

Add permis	sion			×
Permission typ	 Catalog Otabase & table The catalog option covers permissions to create databas table option covers permissions of databases, data table 	es under l s, views, a	DataLakeCatalog and other catalogs, while nd functions.	the database an
Catalog	DataLakeCatalog 💌			
etting mode	Standard Advanced			
Database	Select a database/view/function		Selected (0)	
	Enter a database name	2	Enter a database name	Q
	All		All	
		•		
ermission	Query analysis () Edit data () Owned	by 🚯		
	Select a target permission set. "Query & analytics" and "I targets; "Owner" grants the permission to re-authorize p	ermission	cover the permissions required to analyze s in addition to data edit permissions.	or edit selected

Specific permissions are as follows:

Permission Type	Database	Data Table	View and Function
Query & analytics	• Query all the tables, views, and functions in databases.• Create data tables.	Query	Query
Data edit	 Modify and delete databases and create tables. Permissions of all the tables, views, and functions. 	• Query, insert, update, and delete data.• Modify and delete tables.	Query, create, modify, and delete.
Owner (grants the permission to re-authorize permissions in addition to data edit permissions)	 Modify and delete databases and create tables. Permissions of all the tables, views, and functions. 	• Query, insert, update, and delete data.• Modify and delete tables.	Query, create, modify,



	and delete.
--	-------------

Advanced permission settings: When selecting a single database, you can further set the permissions to query, insert, update, and delete tables, views, and functions; when selecting multiple databases, you can only set permissions at the database level.

In advanced mode, you can set permissions at the column level. When selecting a single data table, you can add the permission to query columns. You can select one or more columns or all of them for authorization.

emission type Catalog Catalog Catalog catalog Catalog Ethic point cater catalogs and table option covers permissions of databases, data tables, views, and functions. atalog Catalog ething mode statose Catalog Catalog statose Catalog Catalog Statose and table option covers permissions of tables, views, and functions. atalog Catalog Catalog Catalog Statose and catalog	Add permission	
atalog DataLakeCatalog etting mode Standard Advanced atabase Image: Imag	Permission type	Catalog Database & table The catalog option covers permissions to create databases under DataLakeCatalog and other catalogs, while the database and table option covers permissions of databases, data tables, views, and functions.
etting mode Standard Advanced batabase Image: Imag	Catalog	DataLakeCatalog 👻
batabase Wen selecting a single database, you can continue to set permissions for tables, views, functions, and columns; but wen selecting more than one databases, you can only set permissions at the database level. ame Data table column column SELECT () Ves	etting mode	Standard Advanced
When selecting a single database, you can continue to set permissions for tables, views, functions, and columns; but when selecting more than one databases, you can only set permissions at the database level. Iame Data table Iolumn Iolumn Iolumn permission SELECT () Idea table Yes	latabase	s st 🛇
ame Data table In		When selecting a single database, you can continue to set permissions for tables, views, functions, and columns; but when selecting more than one databases, you can only set permissions at the database level.
olumn permission SLECT () uthorizable 'Yes	ame	Data table 🔻 🔅 🗇 n 🕲
olumn permission SELECT () uthorizable Yes	olumn	col1 🕲
uthorizable Yes	olumn permission	SELECT ①
	uthorizable	Yes

Click **Confirm** and perform queries in the **Data Explore** module. Enter the following SQL statement to preview the information of **col1** and run the statement to view the preview result of the column.



The permission is not granted for data column **b** in the data table. If you enter the SQL statement to view the information of **b**, the query cannot be performed due to lack of permission.

4. Add an engine permission: In the **User list**, click **Authorize** in the **Operation** column and select **Engine permission** to grant permissions to use, modify, manipulate, monitor, and delete specified resources.

Permission management	🛇 Guangzhou 🔻					Use guide Ø Permission managem
User Work group						
Both sub-account and coordinator a	users need to be granted with permissions to u	use data, engines, and other resources. A user may be associ	ated with one or more work groups to inherit all of	their permissions. An admin user has all resource permi	ssions. For more permission guidelines, see here 🖪 .	
Add user Batch delete						Enter a user ID or name
User ID	Username	User type ①	Description	Added by	Add time \$	Operation
11111111			-		2023-12-05 16:49:02	Edit Authorize 🔻 Delete
		w	-		2023-11-22 14:53:11	Edit Engine permission

Modifying a user permission



1. In the User list, click Authorize and select Data permission or Engine permission.

Permission management	🛇 Guangzhou 🔻					Use guide 🧭 Permission manageme
User Work group						
Both sub-account and coordin	ator users need to be granted with permissions to us	e data, engines, and other resources. A user may be as:	sociated with one or more work groups to inherit all of th	eir permissions. An admin user has all resource per	missions. For more permission guidelines, see here 🗳 .	
Add user Batch delete						Enter a user ID or name
User ID	Username	User type 🚯	Description	Added by	Add time \$	Operation
					2023-12-05 16:49:02	Edit Authorize 💌 Delete

The following takes data permission as an example. On the **Data permission authorization** page, click **Add permission** or **Remove** to modify a permission. The steps for engine permission modification are similar.

← Grant data permissions								
Basic info								
User ID								
Username								
User type								
Description								
Catalog/Database/Table								
Add permission Batch reposse	An admin has all catalog/databa	se/table permissions. Adding or dele	ting this permission does not apply to an a	admin.				
Permission type	Catalog	Database	Table/View/Function	Column	Permission ①	Permission source	Authorizable 🕥	Operation
Database permission					1000	User	No	Remove
Total items: 1							10 💌 / page	

2. Modify **Work group** or **User type**. Click **Operation** > **Edit** to enter the **Edit user** page, where you can modify the **Username**, **User type**, and **Description**. You can also add/remove general users to/from a work group.

Permission management	🔇 Guangzhou 🔻					Use guide Ø Permission managem
User Work group						
Both sub-account and coordin	ator users need to be granted with permissions to use	e data, engines, and other resources. A user may be asso	ociated with one or more work groups to inherit all of th	eir permissions. An admin user has all resource permiss	ions. For more permission guidelines, see here [2].	
Add user Batch delete						Enter a user ID or name
User ID	Username	User type 🕥	Description	Added by	Add time \$	Operation
			-	10000	2023-12-05 16/49/02	Edit Authorize 🔻 Delete
		-			2023-11-22 14:53:11	Edit Authorize 🔻 Delete
					2023-10-23 11:46:29	Edit Authorize 🔻 Delete

Click Edit to modify User type.



← Edit user				
User ID				
Username				
User type Admin Admin	nins excent the root account user. A general u	ser needs to be granted with relevant nermiss	Since or associated with a work rows to access corresponding requires	
An ophin has an permissions for an resources (including laws and engines), and can manage other dam	ning except the root occount user A general o	ser reces to be granted with relevant perma	ions of associated many work group to access corresponding resources.	
Description 🖉				
Bind work group Blick remove Binding a user to a work group grants the user all permissions on the	he group.			
Work group ID Work group name	Description	Add time \$	Added by	Operation
		No data		
Total Items: 0				10 v / page H H H 1 / 1 page +

Viewing a user's permissions

1. Click a user ID in the user list to enter the user details page.

Permission management	🔇 Guangzhou 🔻					Use guide 🧭 🛛 Permission manageme
User Work group						
Both sub-account and coordin	ator users need to be granted with permissions to use	data, engines, and other resources. A user may be as	sociated with one or more work groups to inherit all of th	heir permissions. An admin user has all resource per	missions. For more permission guidelines, see here 💈 .	
Add user Batch delete						Enter a user ID or name
User ID	Username	User type 🚯	Description	Added by	Add time \$	Operation
			-		2023-12-05 16:49:02	Edit Authorize 🔻 Delete

2. View the user's work group, data permission, and engine permission information

View user						
User ID						
Username st						
User type 🛛 🖌						
Description						
Work group	Data permis	ision Engi	ne permission			
Catalog/Datab	ase/Table					
Include the user's	data permissions an	d those inherited f	from a work group			
Permission type	Select	Ŧ	Permission source		Enter a database or ta	ble name 🛛 🔍 🔍
Select 💌						¢
Permission	Catalog	Database	Table/View/Fun	Colum	n Permission (j)	Permissi
Function per						r ,
Function per						9
Admin permi				1		-
•						
Total items: 3			10	🔻 / page	₩ ◀ 1	/ 1 page 🕨 🕨

Revoking a user's permissions

Remove permissions to be revoked from the permission list of a user. This operation requires the admin permission.



Permission management	🛇 Guangzhou 🔻					Use guide Ø Permission manageme
User Work group						
Both sub-account and coordinato	r users need to be granted with permissions to u	se data, engines, and other resources. A user may be associ	ated with one or more work groups to inherit all of	their permissions. An admin user has all resource perm	ssions. For more permission guidelines, see here 🖄	
Add user Batch delete						Enter a user ID or name
User ID	Username	User type 🕥	Description	Added by	Add time \$	Operation
				-	2023-12-05 16:49:02	Edit Authorize 🔻 Delete
	-	in a second second		-	2023-11-22 14:53:11	Edit Authorize 🔻 Delete

Adding and removing a work group permission

Only admins can add or remove work group permissions in a similar way to manipulate data permissions. Users in a work group have all the permissions of the group, so you can bind users to a work group to grant them the data and engine permissions of the work group. Admins don't need to be bound to a work group.

Permission manage	ement 🔇 Guangzhou	,					Use guide Ø	Permission manageme
User Work gro	up							
Batch add users	to a work group to batch grant ther	n the permissions of	data, engines, and other resources of this work group. There is no need to	add an admin to a work group. For more permission guides, see here 🕻 .				
Add work group	Batch remove						Enter a work (group name
Work group ID	Work group name	User count	Description	Added by	Add time \$	Operation		
					2023-08-22 14:30:42	Edit Authorize 🔻	Remove	
			-		2021-11-10 13:32:11	Edit Authorize 🔻	Remove	
Total items: 2						10 🔻 /	page H 4	1 / 1 page

Monitoring and Alarms Data Engine Monitoring

Last updated : 2024-07-31 17:31:18

Data Lake Compute (DLC) provides monitoring services for data engines based on the Tencent Cloud Observability Platform (TCOP), ensuring you can understand the real-time status of data engines and configure data alarms. For alarm configuration methods, see Monitoring Alarm Configuration.

Usage Notice

Before using the Data Lake Compute (DLC) monitoring service, you need to activate the TCOP service. If this service is not yet activated, you can use the root account to activate it.

The use of the TCOP service may incur related charges. For detailed pricing information, see Billing Overview.

Monitoring Access

Access Point I: Data Lake Compute (DLC) Console

Note:

The account must have monitoring permissions for the data engine.

- 1. Log in to the DLC console and select the service region.
- 2. Navigate to the SuperSQL engine page from the left menu.
- 3. Viewing methods supported:

Method 1: Select the engine type to enter the matching engine monitoring list.

Method 2: Select the target engine from the engine list and click **Monitoring** to view the target engine monitoring.

Data engine	🖇 Guanga	thou *												Data engl
Data engine N	etwork co	nfiguration												Clust
 Data Lake Comp engine can be complexed 	ute offers t infigured w	oth public and priva ith the auto-suspen	ste data engines. A publi sion or scheduled susper	c data engine is mar nsion policy, with no	aged by Data Lake Compute fees charged on it after susp	and billed by scanned o ension. For operations a	data volume, with no operation or and notes, see Managing Private	permission required; a private Data Engines 🙆 .	data engine can be billed	on a pay-as-you-go basis or sub	scribed monthly. For mo	re billing info, see Billing Ov	arview 🛃 . A pa	Spark engin y-as-you-go < Presto engir
Create resource	ill query 🛙	Renewal manage	ement 🗹									Select a resource tag or ente	r keyword(s) (se	parate two
Resource name/ID		Engine type	Kernel version	Running sta	Billing mode	Auto-renewal	Start and stop policy	Cluster description	Cluster spec	Network configuration	Created at \$	Creator	Descripti	o Operation
th To Do K		-	-	Suspend (1)	Pay-as-you-go		Manual start, Manual suspension	Private engine	64CU (standard)		2023-05-15 20:26:07	0		Monitor Spec configuration Parameter Configuration More ▼
e D	6	F	Sur-	Running	Pay by scanned data volume		Manual start, Manual suspension	Public engine			2022-03-17 21:00:18		共享集群	Monitor Spec configura More 🔻
Total items: 2												10 💌 /	page H	< 1 / 1 page

Access Point Two: TCOP

1. Log in to the TCOP with an account that has the necessary permissions.

2. Select **Cloud Product Monitoring** from the left menu, find Data Lake Compute DLC, and choose the type of monitoring you need to view.

Observability Platform	Cloud Virtual Machine Gaingthou Shanghai Nanjing Beijing Chengdu Chongoing Hong Kong China Singapore Bangkok Jakarta Mumbai Seoul Tokyo Silkon Valley Virginia Toronto Frankfurt Saopaulo
G Managed Service for Grafana	Heath status in last 24 hours-klarm(0) Aret(0)
Monitoring Platform	Last 7 days Select Time 🗇 Guide to Cetting Monitoring Data Va AP 12 Use ") to split more than one keywords, and press Enter to "Q
Event Bridge *	Drifest Name Monitor. Network Tree IP4 Addresses IP46 Addresses Status Project T CPUIRIENTION. # Destriktion
Data Usage Monitoring	No results found
Cloud Product Monitoring	
Ocloud Virtual Machine	0 items; in total
Cloud Block Storage	
Cloud Load Balance	
Cloud Database *	
🔒 Private Network 👻	
 Cloud Object Storage 	
Cloud File Storage	
⊚ DLC ^	
• datalake- sparkengine	
 datalake-datatask 	
 datalake- prestoengine 	
Application Performance Management	
Application	
_	

3. After selecting the monitoring type, you will enter the monitoring page. Select the corresponding region to view the monitoring resource information for that region.

	_								
Observability Platform	datalake-sparkengine	Guangzhou Shangha	Shanghai Finance	Nanjing Beijing Beijing	Finance Chengdu Chongqing H	long Kong, China Singapore Silicon Valley	Virginia Frankfurt		
G Managed Service							Guide to Getting Mo	nitoring Data via API 😰 Use ' ' to split more than one	keywords, and press Enter to :
Monitoring Platform	dataEngineld	Monitor dat	aEngineName		task_failed_numCount	task_queue_numCount	task_queue_time_ma	cluster_cpu_usage_max%	cluster_mem_u
🗊 Event Bridge 👻		di g				100			
Data Usage Monitoring	1 item(s) in total							Lines pa	er page 20 💌 🖂 🕺 1,
Cloud Product Monitoring									
Cloud Virtual Machine									
Cloud Block Storage									
Cloud Load Balance									
🗧 Cloud Database 🗳									
🔒 Private Network 👻									
Cloud Object Storage									
Cloud File Storage									
⊚ DLC ^									
 datalake- sparkengine 									
 datalake-datatask 									
 datalake- prestoengine 									
Application Performance Management									
Application									
Э									

4. Click the **Engine ID** to enter the detailed monitoring page.

Monitoring Granularity Configuration



You can configure the monitoring data time range, time granularity, and auto-update interval at the top of the monitoring page.

		11	nour	_		Ċ	1	<u>с</u> т	me grar	nularity	1 mi	'n	¥	φ	Disable 🔻 🕶 Y Show legends
5 m 30 c	inutes Jays	30 n Today	ninutes Yest	1 ho erday	bur	3 hours	12 h	ours	24 hou	rs 2	days !	7 day	s		verage(%) 🛈
Nov 2	2023				4	0 >		Dec 2	023				4	• •	
Su	Mo	Tu	We	Th	Fr	Sa		Su	Mo	Tu	We	Th	Fr	Sa	
29			1	2	3	4		26	27		29		1	2	59 10:06 10:13 10:20 10:27
5	6	7	8	9	10	11		3	4	5	6	7	8	9)fo Max: - Min: - Avg: -
12	13	14	15	16	17	18		10	11	12	13	14	15	16	
19	20	21	22	23	24	25		17	18	19		21			a max(%)
26	27	28	29	30	1	2		24	25		27	28	29		
									1	2		4	5		
Select	time													ОК	

Monitoring data time range: Accurate to the minute, supports selecting data for a specific time period.

Time granularity: Interval between monitoring points, configurable to 1 minute or 5 minutes.

Auto-update data: Configures the automatic refresh interval for page data, with options to set it to off, 30 seconds, 5 minutes, 30 minutes, or 1 hour.

Monitoring Data Comparison

You can select a time period for data comparison. After selecting the comparison time range through one click, you can view the comparison data in the data compass below.

Time granularity: 1 min 👻 🗘 Disable 💌 🚥 🗹 Show legends
Week-over-Week (Last Week) Day-over-Day (Yesterday)
0.6

Monitoring Metrics

Monitoring Metrics						
Maximum CPU utilization of all Driver nodes						
Maximum CPU utilization of all Executor nodes						
Average CPU utilization of all Driver nodes						



	Average CPU utilization of all Executor nodes
	Maximum CPU utilization of all clusters
	Average CPU utilization of all clusters
	Maximum memory utilization of all Driver nodes
	Maximum memory utilization of all Executor nodes
Momony	Average memory utilization of all Driver nodes
Memory	Average memory utilization of all Executor nodes
	Maximum memory utilization of all clusters
	Average memory utilization of all clusters
	Number of canceled tasks
	Number of failed tasks
	Number of initialized tasks
	Average task initialization time
Taaka	Maximum task initialization time
14585	Number of queued tasks
	Average task queue time
	Maximum task queue time
	Number of running tasks
	Number of successful tasks
Network	Maximum inbound bandwidth of all Driver nodes network
	Maximum inbound bandwidth of all Executor nodes network
	Average inbound bandwidth of all Driver nodes network
	Average inbound bandwidth of all Executor nodes network
	Maximum outbound bandwidth of all Driver nodes network
	Maximum outbound bandwidth of all Executor nodes network



	Average outbound bandwidth of all Driver nodes network						
	Average outbound bandwidth of all Executor nodes network						
	Maximum cloud disk utilization of all Driver nodes						
Claud Diak	Maximum cloud disk utilization of all Executor nodes						
Cloud Disk	Average cloud disk utilization of all Driver nodes						
	Average cloud disk utilization of all Executor nodes						
	Job Engine CU Count						
00	CU Utilization						

Data Job Monitoring

Last updated : 2024-07-31 17:31:39

DLC provides monitoring services for data jobs based on TCOP service, ensuring that you can understand the operation of data jobs in real time and configure data alarms.

Notes

Before using the monitoring service of DLC, you need to activate the TCOP service (for usage details, refer to TCOP Documentation). If the service has not been activated, it can be done using the root account. Fees may be incurred during the use of TCOP service; for detailed fee information, refer to TCOP Billing Overview.

Monitoring Entrance

Entrance one: DLC Console

1. Log in to DLC Console > Data Job, and select the service region.

2. Or enter the Data Job page from the left sidebar.

3. In the top right corner, click **Job Monitoring** to go to the monitoring page. Or click the **Monitoring** feature of the target job to enter its monitoring page.

1	Data job	🔇 Guangzhou 🤋									
	Spark job	Job configuration	on Session management								Job monitoring Task history Lo
ର =	Create job		🕲 Q, Ali	* All	×				All	Last 7 days Last 30 days	Select date Select date
=	Job name		Job ID	Job type	Job file	Current tasks	Task engine \$	Creator \$	Created at \$	Update time \$	Operation
	-	ō	c 6				h (i)	5 (I)	2023-10-17 21:20:34	2023-10-17 21:34:28	Monitor Edit Running I
Ø	Total items: 1									10 💌 / page	H ≺ 1 /1 page
۵											

Entrance two: TCOP

1. Log in to TCOP Console. Account must have the required permissions.

2. In the left menu, select Cloud Product Monitoring, find DLC, and choose the type of monitoring you wish to view.

Observability Platform	Cloud Virtual Machine Guargetou Shanghai Nanjing Beijing Chengdu Chongoing Hong Kong China Singapore Bangkok Jakarta Mumbai Seoul Tokyo Silicon Valley Virginia Toronto Frankfurt Saogaulo
Managed Service for Grafana	Health status in last 24 hoursolverm(i) Alert(ii)
Monitoring Platform	Last hour Last 24 hours Last 7 days Select Time 🔝 Guide to Getting Monitoring Data via API 😰 Use '' to split more than one keywords, and press Enter to R 🗘
Event Bridge *	D/Host Name Monifor_ Network Type IPv4 Addresses IPv6 Addresses Status Project T CPUUBIRZono_ # MemoryUBIZ_ # DiakUBIZzdon_ # PublicBandwi. # HestBh Subts ① Number of
Data Usage Monitoring	No results found
Cloud Product Monitoring	
Cloud Virtual Machine	
Cloud Block Storage	
Cloud Load Balance	
Cloud Database 👻	
🔒 Private Network 👻	
Cloud Object Storage	
Cloud File Storage	
◎ DLC ^	
 datalake- sparkengine 	
 datalake-datatask 	
 datalake- prestoengine 	
Application Performance Management	
Application	

3. After selecting the monitoring type, enter the monitoring page and select the respective region to view the monitoring job information for that region.

 datalake-datatask 	Guangzhou Shanghai	Shanghai Finance	Nanjing B	leijing Beijing Finance	Chengdu (Chongqing	Hong Kong, China	Singapore	Silicon Valley	Virginia]				
												Guide to Getting Monitoring Data via API 😰	(Use 'j' to split r	more than $ \mathbf{Q} $
datataskname				Monitor	datataskid							workflowname			
				di											
1 item(s) in total														Lines per page 20 💌	H 4 1/1

4. Click **Job ID** to enter the monitoring details.

Monitoring Granularity Configuration

Supports configuring the monitoring data time period, time granularity, and automatic update time range through the monitoring settings at the top.

		11	nour			Ċ	1	() Ti	me grar	nularity	1 mi	n	v	¢	Disable 🔻 🕶 🗸 Show legends
5 mi 30 d	inutes Jays	30 n Today	ninutes Yest	1 ho erday	our	3 hours	12 ho	ours	24 hou	rs 2	days	7 day	s		verage(%) 🕕
Nov 2	2023				4	0 ⊦		Dec 2	023				4	• •	
Su	Мо	Tu	We	Th	Fr	Sa		Su	Mo	Tu	We	Th	Fr	Sa	
29	30	31	1	2	3	4		26	27	28	29	30	1	2	50 10:05 10:13 10:20 10:27
5	б	7	8	9	10	11		3	4	5	б	7	8	9)fo Max: - Min: - Avg: -
12	13	14	15	16	17	18		10	11	12	13	14	15	16	
19	20	21	22	23	24	25		17	18	19	20	21	22	23	* max(%)
26	27	28	29	30	1	2		24	25	26	27	28	29		
									1	2		4	5	6	
Select	time													ОК	

Monitoring Data Time Range: Precise to minutes, supports selecting data for a specific period.

Time Granularity: Monitoring point interval time, supports configuring for 1 minute or 5 minutes.

Automatic Data Update: Page data auto-refresh configuration, supports configuring off, 30s, 5min, 30min, 1h.

Monitoring Data Comparison

Supports selecting data for a specific period to compare monitoring data. After clicking to select the comparison time range, you can view the comparison data in the data compass below.

Monitoring		
1 hour	Time granularity: 1 min	✓ Disable ▼ ··· ✓ Show legends
driver_cpu_usage_average(%) ③	Week-over-Week (Last Week) Day-over-Day (Yesterday)	ı_usage_average(%) 🛈
0.8	Custom Date	
0.6	0.6	

Monitoring Metric

Monitoring Type	Monitoring Metric
Job	Job error Log Count
	Job warn Log Count


Access Point Gateway Engine Monitoring

Last updated : 2024-07-31 17:31:54

DLC provides monitoring services for the access point gateway engine based on TCOP service, ensuring you can understand the gateway status in real time.

Notes

Before using DLC's monitoring service, you need to activate the TCOP service (for usage details, see TCOP Documentation). If the service has not been activated yet, it can be activated using the root account. TCOP service usage may incur related tariffs, for detailed tariff information, see TCOP Billing Overview.

Monitoring Entrance

Entrance one: DLC Console

1. Log in to the <1>Standard Engine> page, and select the Service Region.

2. Select the Standard Engine, and click on **Monitoring** at the access point to enter the monitoring data display interface.

Configuration Entrance: TCOP

1. Log in to the TCOP Console, the account must have the relevant permissions.

2. From the left menu, select Cloud Product Monitoring, enter the Policy Management page under Alarm Management, select Data Lake Computing, and choose the corresponding Access Point Gateway Engine.

Access Point Gateway Engine Monitoring Configuration Type

Creating alarm policy

1. DLC Access Point Gateway supports alarm capabilities. Log in to TCOP, click **Alarm Management**, and select the Policy Management page.

2. Click **New Policy**, for policy type choose "Data Lake Computing". Access Point Gateway supports alarms for three dimensions, including:

"Gateway" alarm dimension is: appid/gatewayid.

"Gateway (Multi-dimensional)" alarm dimension is: appid/gatewayid/instanceid.

"Gateway Engine (Multi-dimensional)" alarm dimension is: appid/gatewayid/engineid/processid.

Name	Supported Dimensions	Advantages and Use Cases			
Gateway (Multi- dimensional)	Supports: CPU, Memory, Disk, Network Fine-grained Alerting. For example, to configure an alert for the CPU utilization of an Access Point Gateway, you can choose to configure one, several instances under a specific Access Point Gateway, or any instance node triggering the threshold to alert.	Alert supports more dimensions, and the alert method is more flexible. Basic Metrics are recommended to use this approach.			
API Gateway	Mainly aimed at monitoring the overall load situation of the current gateway, aggregating basic metrics according to Access Point Gateway Nodes, and supporting Service-level Metric Alerts. For example: execute_statement_num (number of statements executed), opened_operation_num (number of operations opened), launch_engine_num (number of engines started), engine_process_thread_num (number of threads started by the engine).	Supports Dashboard. Suitable for Single-node access point gateway or service metric alert.			
Gateway Engine (Multidimensional)	The Gateway Engine refers to the monitoring and alarm of the process of starting the DLC engine by the Access Point Gateway. For example: engine_process_thread_num (number of threads started by the engine), mainly aimed at monitoring the process information of the engine started by the current Access Point Gateway	Supports fine-grained alerting, for example: commonly configure any engine's process count under a specific Access Point Gateway ID to reach the threshold to trigger an alert. Suitable for alerting on process metrics started by the Access Point Gateway.			

Monitoring Alarm Configuration

Last updated : 2024-07-31 17:32:15

Configuring New Alarm Policy

Supports configuring monitoring alarms for specific metrics. You can go to Creating Alarm Policy to configure the content of the alarm.

Observability Platform	← Create Ala	rm Policy								
Monitor Overview										
Dashboard *	Source Addition Configure Addition									
田 Instance Group	- Basic Info									
Alarm Management	The New York State Stat									
Alarma List	Policy Name									
	Remarks									
Alarm ^ Configuration										
Alarm Policy										
Silence Alarm	Configure Alar	Configure Alarm Rule								
 (0) Trigger Condition Template 	Monitoring Type	Cloud Product Manitaring APM RUM Cloud Probe Manitar								
	Policy Type	Cloud Virtual Machine *								
Template	Project 🛈	Debut Project I textus. The close of the c								
Cloud Native Monitor										
😫 Managed Service	lag	lág Key lág Valué X								
for Prometheus										
G Managed Service										
tor Gratana	Alarm Object	Instance ID v Select object v								
Monitoring Platform		CNU - Back Monter supports alarm policy configuration by tag now, allowing newsy purchased instances to be automatically associated with alarm policies Vew Details 2								
🗉 Event Bridge 🛛 👻	Trigger Condition	🕐 Seeket Tempste 🕜 Configue manually 💆 Apply preset fogger conditions 🛈 (Currently, event alarm notifications cannot be configured through the fogger condition tempste)								
Data Usage Monitoring		Metric Alarm Event Alarm								
Cloud Product Monitoring										
Cloud Virtual		When meeting any v of the following metric conditions, the metric will trigger an alam. Enable alam live feature.								
Machine										
Cloud Block		If CPULISitization * (statistical perior * > * 95 55 at 5 consecutive * then Atam every 2 hours * ① ①								
Ξ										

Or click the monitoring content for which you need to configure an alarm to enter the configuration page, where you can configure the content of the alarm.



1 hour		Time granularity: 1 mir	т Ф	Disable 💌 🚥 🔽	Show legends	
	D	A		(j)		::
1			1			
0.8			0.8			
0.6			0.6			
0.4			0.4			
0.2			0.2			
0			0			

Managing an alarm policy

To manage configured alarm policies, you can perform configuration management through the Policy Management page.

Observability Platform	Alarm Management
Monitor Overview	Alarm Dashboard Alarm Records Policy Management Basic Configuration
Dashboard *	0 If you have any questions or supportions, scare QR code to join our community on WeChast or WeCom.
日 Instance Group	
Alarm Management	Create Paircy Delete More * Poly Name/10 th
Alarm List	Palley Name Monitoring Tree Policy Type Alarm Bule Project T Associated Instances Notification Template Y Last Modified 1 Alarm On-Off T Operation
🛱 Alarm	
Configuration	Liteut tours Clearing Constant
Alarm Policy	Cocy Deter md 202311/14 205614 Cocy Deter
 Slience Alarm 	A second s
 ([©]) Trigger Condition Template 	Total items 1
A Notification	
Template	
Cloud Native Monitor	
Managed Service for Prometheus	
G Managed Service	
for Grafana	
Monitoring Platform	
Event Bridge *	
Data Usage Monitoring	
Cloud Product Monitoring	
Cloud Virtual	
Machine	
Cloud Block	
Ξ	

Configuration Instructions

Configuration Item	Configuration Instructions
Policy name	Name of the alarm policy, up to 60 characters
Remarks	Remarks for the alarm policy, up to 100 characters
Monitoring Type	Please select Cloud Product Monitoring
Policy Type	Please select DLC
Policy Tag	Support for managing policy content via Tag requires relevant permissions to operate
Alarm Object	You can configure alarms for Instance ID (supports multiple selections), grouped instances, and all instances
Alert Configuration Template	You can choose a template or configure manually. Administrators need to create the template in advance, and it supports configuring multiple alert rules
Notification Template	Supports creating or selecting existing notification templates, with support for configuring up to 3 templates

Audit Log

Last updated : 2024-07-31 17:30:53

DLC provides an operation log audit service based on Tencent Cloud's CloudAudit service, ensuring you can understand the system operation records in real time and check the operation information.

Notes

Before using the audit CLS of DLC, you need to activate Tencent Cloud's CloudAudit service. If the service is not yet activated, you can activate it with the primary account.

Use Instructions

The Data Lake Computing Console currently displays up to 3 months of log information. To view older log information, you can go to CloudAudit.

The audit logs contain console operations and API call operations. Currently, it supports viewing log information for engine management, task management, data source management, workgroup management, user management, scheduled task instance management, scheduled task management, and scheduling plan management.

Operation Guide

- 1. log in to Data Lake Computing Console, select Service Region.
- 2. Through the left menu **Data Operation and Maintenance**, select the Audit Log feature.
- 3. Supports log queries based on user UIN or request ID.
- 4. Detailed log information can be viewed by clicking **Query Details**.

٢	Run history	🔇 Guangzhou 🔻										Histori
# ©	This module displays the status of tasks submitted in other modules, including SQL tasks and data import/export tasks. An admin can query all tasks in the last 45 days, while a general user can query tasks related to them in the last 45 days. Learn more 12											
≣	1 Select a task type v Select								2023-12-18 ~ 2023-12-18			
H H	Job overview											
Ŷ	All Executing				Queuing up			Initialize				
٢	1			0			0			C		
				1								
	Task ID		Task type	Task content	Execution status	Creator	Task submission time \$	Data engine	Resource usage	\$ Kern	el version	Operation
		0 15	S	tiog Th	Successful	<u>()</u> د	2023-12-18 17:33:28		-			Learn more
	Total items: 1										10 💌 / page	