数据湖计算

操作指南

产品文档

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

文档目录

操作指南
- 数据探索
  - SQL 编辑器
  - 数据查询任务
    - SELECT 任务
    - 查询分区表
    - 查询 Json 数据
    - 查询其他数据源
    - 使用视图
    - INSERT INTO
    - 查询脚本参数
  - 查询脚本分析
- 数据管理
  - 数据目录及数据库管理
  - 数据表管理
  - 数据视图管理
  - 函数管理
  - 分区字段策略
- 数据作业
  - 数据作业概述
  - 配置数据访问策略
  - 创建数据作业
  - 管理数据作业
- 历史记录
- 引擎管理
  - SuperSQL 引擎
    - SuperSQL 引擎说明
    - 购买独享数据引擎
    - SuperSQL 引擎续费
    - 管理独享数据引擎
    - 容灾集群
    - 引擎内核版本
    - 引擎网络配置
    - 为独享引擎资源关联标签
    - 引擎本地缓存

版权所有：腾讯云计算（北京）有限责任公司

自定义任务调度池
运维管理
权限管理
访问管理 CAM 服务
DLC 权限概述
用户和工作组
子账号权限管理
存储配置
托管存储配置
元数据加速桶的绑定
审计日志
监控与告警
数据引擎监控
数据作业监控
接入点网关引擎监控
监控告警配置
查询脚本管理
系统约束
元数据信息
计算任务

操作指南
数据探索
SQL 编辑器

最近更新时间：2024-07-17 15:47:59

数据湖计算 DLC 提供的 SQL 编辑器支持使用统一的 SQL 语句进行数据查询，兼容 SparkSQL，您使用标准 SQL 即可完成数据查询任务。
您可以通过数据探索进入 SQL 编辑器，在编辑器内可完成简单的数据管理、多 Session 的数据查询、查询记录管理、下载记录管理。

数据管理

数据管理支持新增数据源、数据库管理及数据表管理。

新建数据目录

目前数据湖计算 DLC 支持管理 COS 及 EMR HIVE 的数据目录。操作步骤如下。
1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录角色需要管理员权限。
2. 进入数据探索，鼠标移入库表列表上方的

图标，单击**新建数据目录**即可进入新建流程。

详细操作指南可参见 查询其他数据源。

数据库管理

通过 SQL 编辑器，支持对数据库行进创建、删除、查看详情操作。

数据表管理

通过 SQL 编辑器，支持对数据表进行创建、查询、查看详情操作。

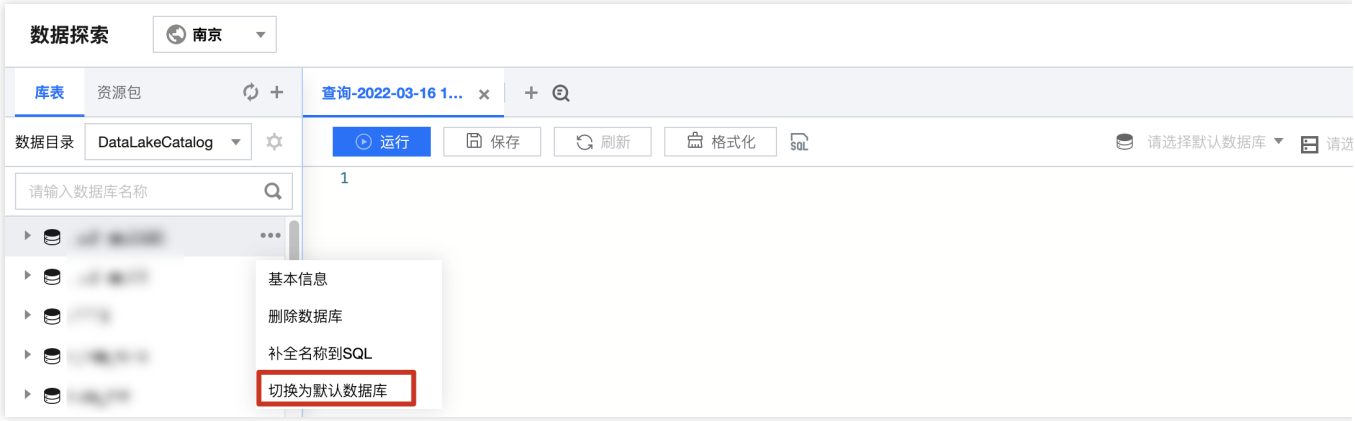
版权所有：腾讯云计算（北京）有限责任公司

默认数据库切换

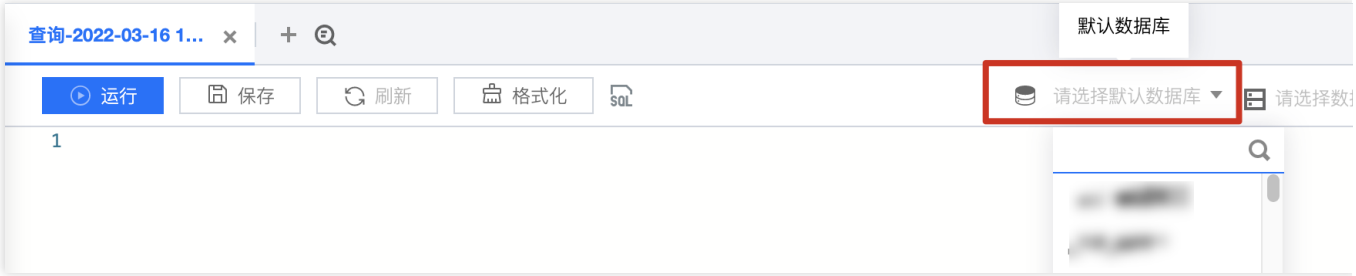
使用 SQL 编辑器时，可以指定查询任务的默认数据库，指定后若在查询语句中未申明数据库，则查询将在默认数据库下执行。

1. 登录 数据湖计算 DLC 控制台，选择服务地域。
2. 进入数据探索，鼠标悬停需指定的数据库名称，单击

图标，单击**切换为默认数据库**即可将该数据库指定为默认数据库。



3. 或可以直接在默认数据库选择框切换。



数据查询

Session 管理

SQL 编辑器支持多个 Session 进行数据查询，每个 Session 内的配置独立（默认数据库、使用的计算引擎、查询记录等），方便用户进行多个任务运行及管理。

您可以通过单击

图标创建 Session，单击 tab 栏进行编辑器界面的切换。

为了方便您的查询使用，常用的 Session 您可以点击保存按钮将 Session 进行保存，同时您可以通过点击

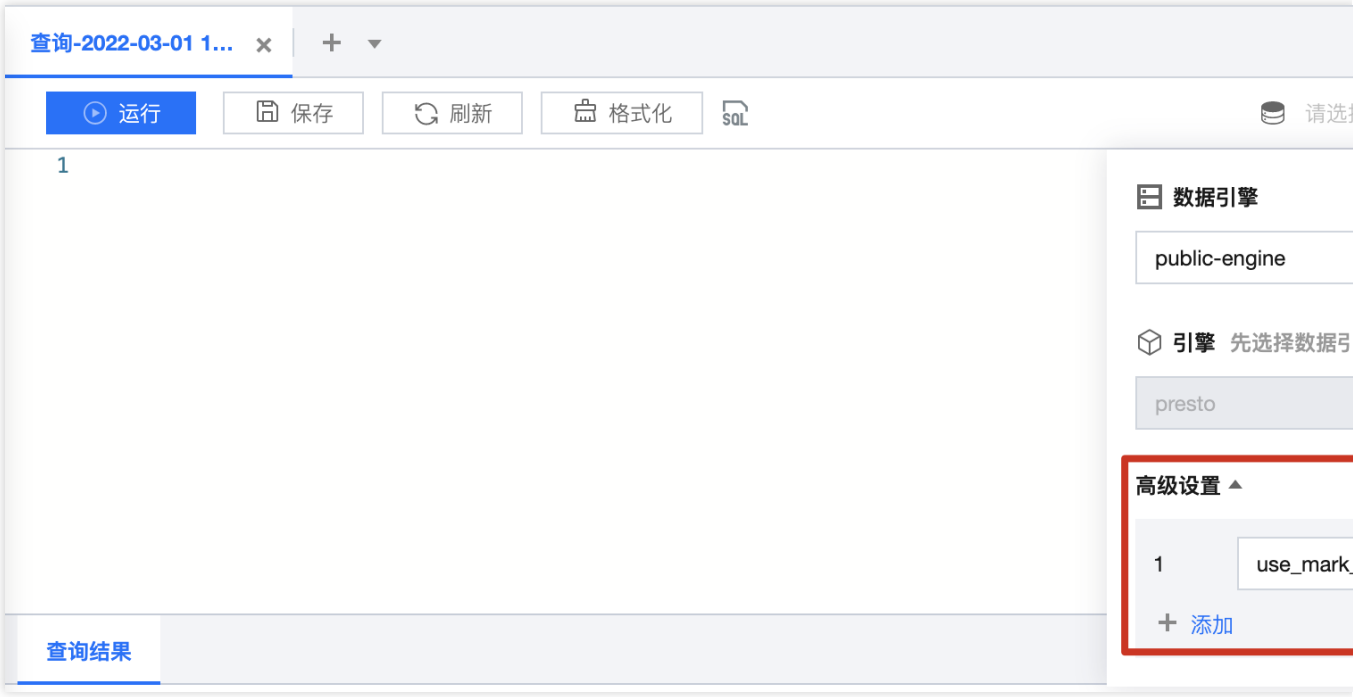
图标快速打开您已保存的 Session。

针对已保存的 Session，您可以单击**刷新**按钮来更新同步已保存的信息，保证查询语句的准确性。

编辑器支持同时运行多个不同的 SQL 语句，单击**运行**按钮将会把编辑器内所有的 SQL 语句进行运行，同时拆分为多个 SQL 任务。
如需运行部分语句，可选中需运行的语句后单击**部分运行**。

引擎参数配置

选择数据引擎后，支持对数据引擎进行参数配置，选择数据引擎后，在高级设置单击**添加**即可进行配置。



当前支持配置参数如下：

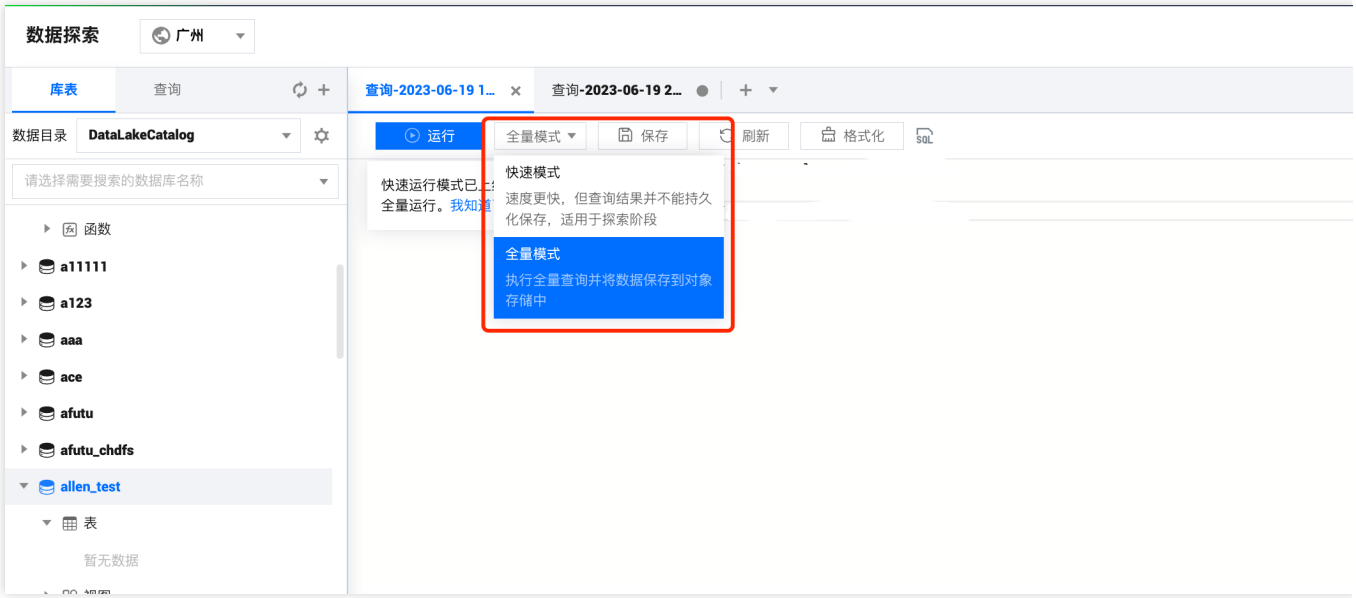
引擎	配置名称	初始值	配置说明
SparkSQL	spark.sql.files.maxRecordsPerFile	0	写入单个文件的最大记录数。 如果该值为零或为负，则没有限制。
	spark.sql.autoBroadcastJoinThreshold	10MB	配置执行连接时显示所有工作节点的表的最大字节大小。 通过将此值设置为“-1”，可以禁用显示。
	spark.sql.shuffle.partitions	200	默认分区数。
	spark.sql.sources.partitionOverwriteMode	static	该值为 static 时，在执行覆盖写操作之前，会删除所有符合条件的分区。 举例说明：分区表中有一个“2022-01”的分区，当使用 INSERT OVERWRITE 语句向表中写入“2022-02”这个分区的数据时，会把“2021-01”的分区数据也覆盖掉。 当该值为 dynamic 时，不会提前删除分区，而是在运行时覆盖那些有数据写入的分区。
	spark.sql.files.maxPartitionBytes	128MB	读取文件时要打包到单个分区中的最大字节数。
Presto	use_mark_distinct	true	决定引擎在执行 distinct 函数时是否进

			行数据重分布。 如果查询中多次调用 distinct 函数，推荐将该参数设置为 false。
	USEHIVEFUNCTION	true	执行查询时是否使用 Hive 函数；如需使用 presto 原生函数，请将参数设置为 false。
	query_max_execution_time	-	用于设置查询超时，在查询执行的时间超过设置的时间后，查询会被终止。单位支持 d-天，h-小时，m-分钟，s-秒，ms-毫秒（举例：1d 代表1天，3m 代表3分钟）。
	dlc.query.execution.mode	async	引擎查询执行模式，默认为 async 模式， 该模式任务会完成全量查询计算，并将结果保存到 COS，再返回给用户，允许用户在查询完成后下载查询结果。 用户也可以将该值改为 sync，在 sync 模式下，查询不一定会执行全量计算，部分结果可用后，会直接由引擎返回给用户，不再保存到 COS。因此用户可获得更低查询延迟和耗时， 但结果只在系统中保存30s。推荐不需要从 COS 下载完整查询结果，但期望更低查询延迟和耗时时使用该模式，例如查询探索阶段、BI 结果展示。

Presto 运行模式

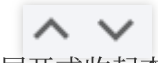
当用户选择的引擎为 Presto 引擎时，数据探索运行支持用户选择“快速模式”运行或“全量模式”运行

1. 快速查询：速度更快，但查询结果并不能持久化保存，适用于探索阶段。
2. 全量模式：执行全量查询并将数据保存到对象存储中。

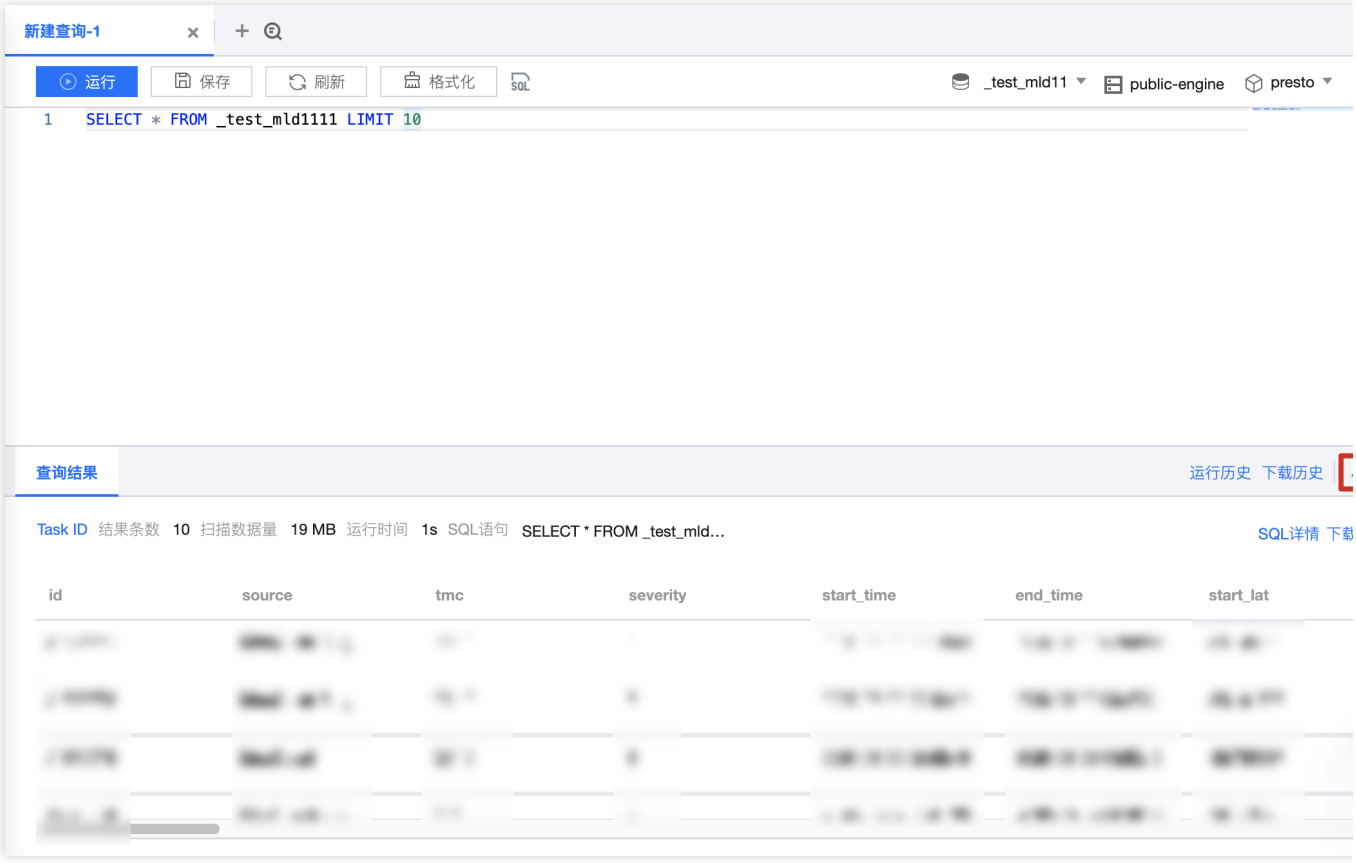


查询结果

通过 SQL 编辑器可直接查看查询结果，可以通过单击



图表展开或收起查询结果的展示高度。



控制台单个任务最多会返回1000条结果，如需更多结果可使用 API。
查询结果在未指定 COS 存储路径情况下支持下载到本地。

历史运行查询

每个 Session 可保存3个月内的运行历史，支持查看近24小时的查询结果。可通过运行历史快速查找过去执行的任务信息。

下载历史管理

每个 Session 的下载任务可在**下载历史**中查看，可查询下载任务状态及相关参数信息。

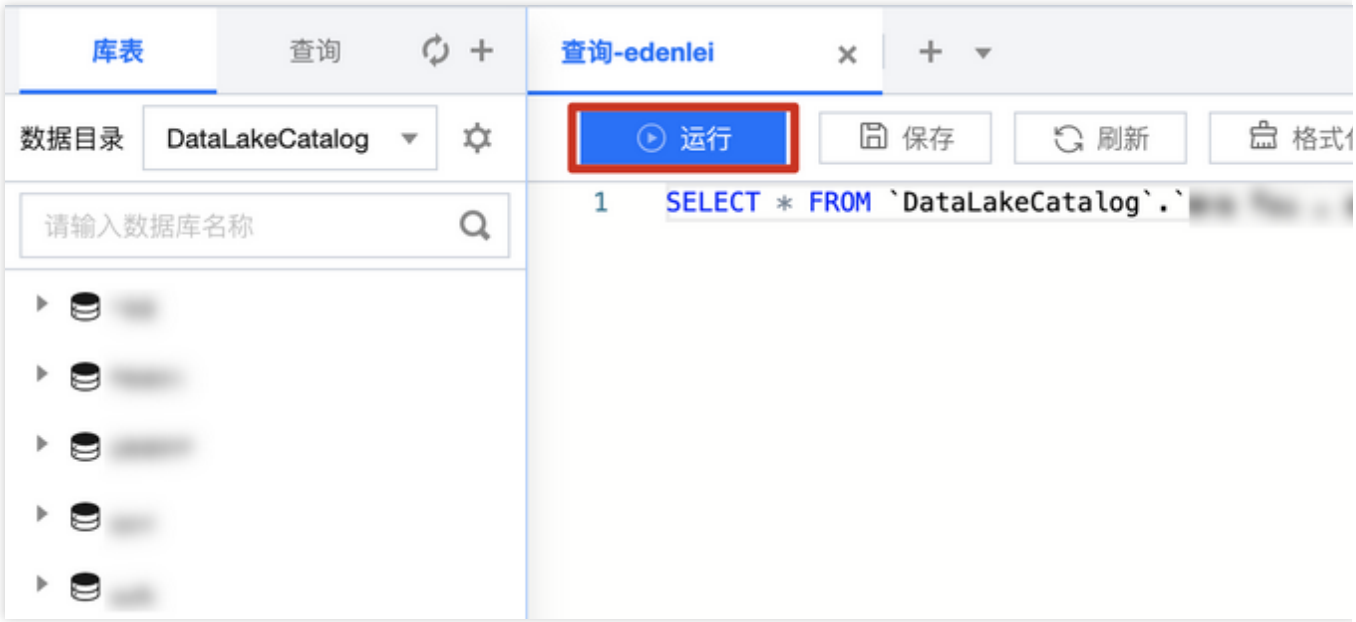
数据查询任务
SELECT 任务

最近更新时间：2024-07-17 16:16:17

用户可以通过 SQL 语句对创建的数据库和数据表中数据进行查询、分析和计算。

执行 SELECT 查询任务

1. 选择默认数据库及计算资源。
用户可以选择一个默认数据库，则在 SQL 语句中会将没有指定数据库的操作在默认数据库下执行。
计算资源可选共享集群或独享集群。
2. 运行 SQL 任务。写入标准SQL语句后，单击**运行**，即可开始运行 SQL 任务。

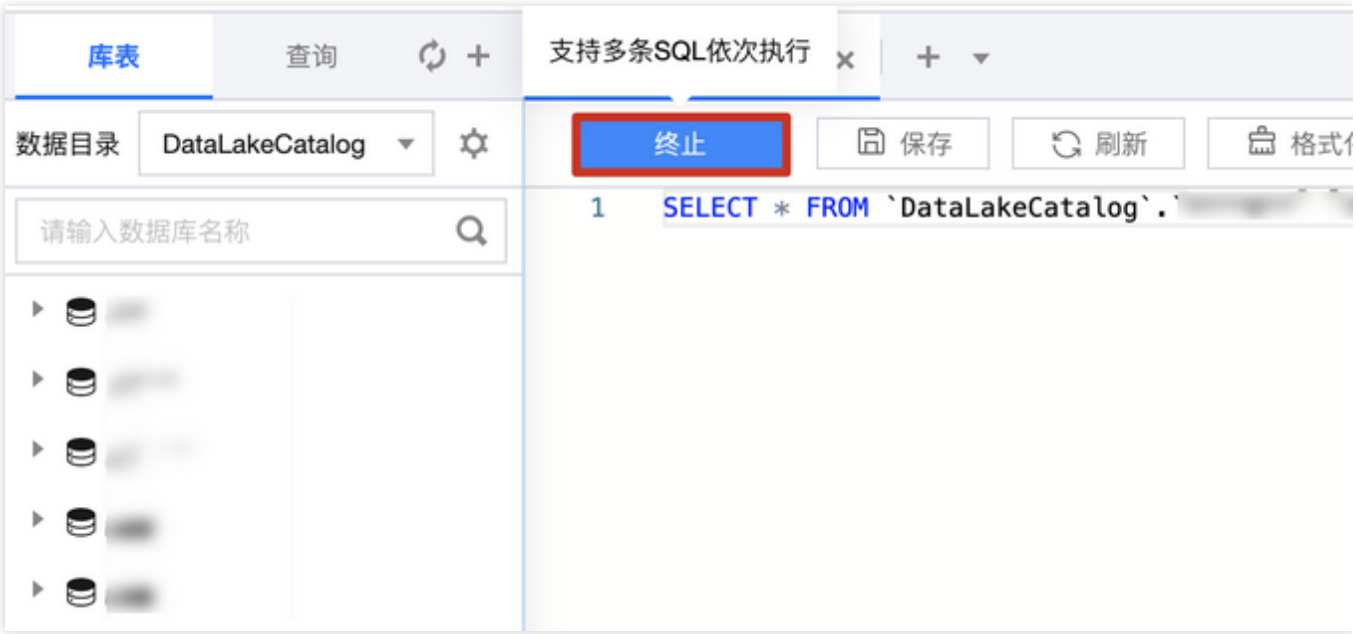


DLC 单个任务运行时间的上限为30分钟。
DLC 为 Serverless 架构，算力资源会临时被调度。在一段时间内，首次运行 DML 任务的结果返回时间可能会比正常任务稍慢。
注意：
通过控制台提交 SQL 任务时，单个 SQL 最大支持45万个字符，如超过该字符数，可通过 API 执行任务。
3. 任务执行完成，会在控制台展示查询结果。
若用户退出了控制台页面，则无法在控制台查看历史任务的查询结果，需要到历史运行或用户配置的查询结果 COS 桶中查看任务结果文件。

版权所有：腾讯云计算（北京）有限责任公司

取消正在执行的查询任务

在任务运行过程中，**运行**会切换为**终止**，可单击**终止**，取消本次查询。取消查询后，DLC 不会返回查询结果，但是 DLC 会统计已经执行的数据扫描量。若您使用的计算资源为public_engine，数据扫描量将会产生一定的费用，详细资费可参见 计费概述。



查询分区表

最近更新时间：2024-07-17 16:17:41

若用户将数据以分区目录的方式进行存储，可以减少 DLC 计算任务的数据扫描量，大幅度提升运算性能。对数据进行分区最常见的做法是将数据按照时间存储到不同目录下，例如，将同一天产生的数据放到一个目录下，可以按照“年-月-日”组织多层数据目录。DLC 中的同一个表及其分区必须使用相同的数据格式。

创建分区表

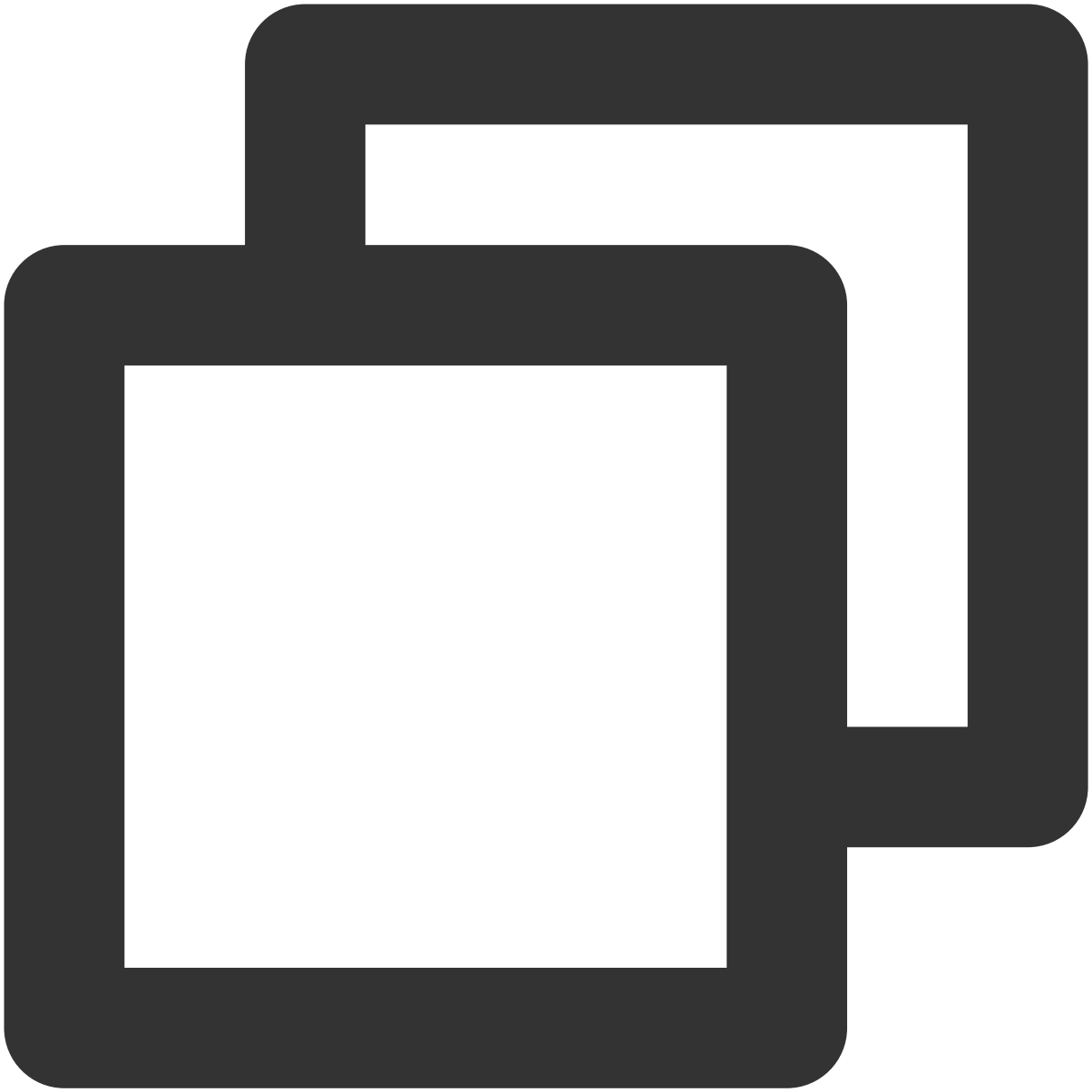
创建分区表需要在建表语句中指定分区字段。

添加分区数据

在创建数据表时指定分区仅是配置了分区字段，不能立刻执行查询语句获取数据。用户需要将分区数据添加至数据表。若有新的分区数据添加至数据目录，则需要添加分区信息到数据表。

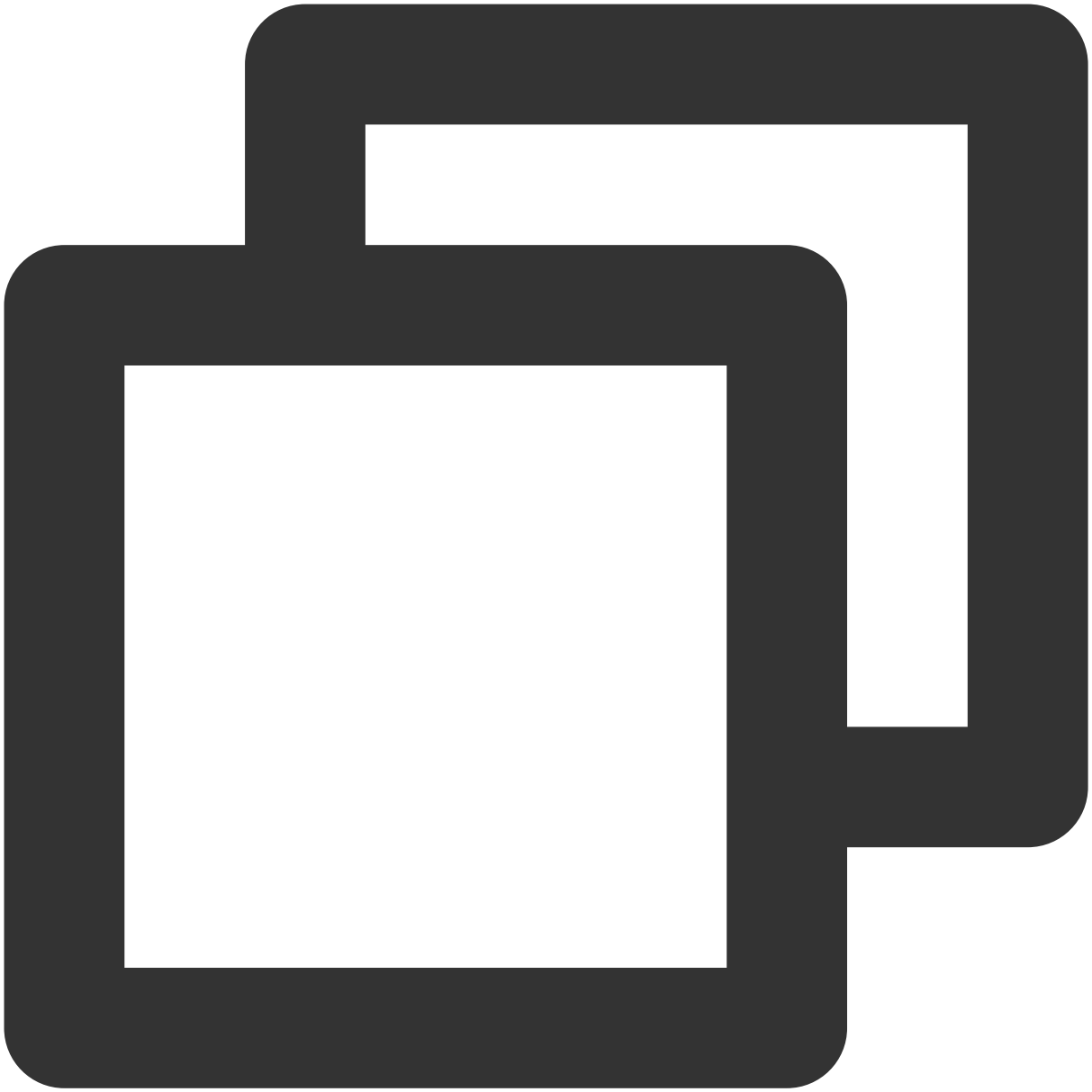
手动添加分区

使用 `ALTER TABLE ADD PARTITIO` 语句，会把指定的分区目录添加至数据表。若区目录兼容 Hive 的分区规则（**分区列名 = 分区列值**），则无需指定数据路径。否则，需要显示指定数据路径。
示例1：单个分区目录，分区目录符合



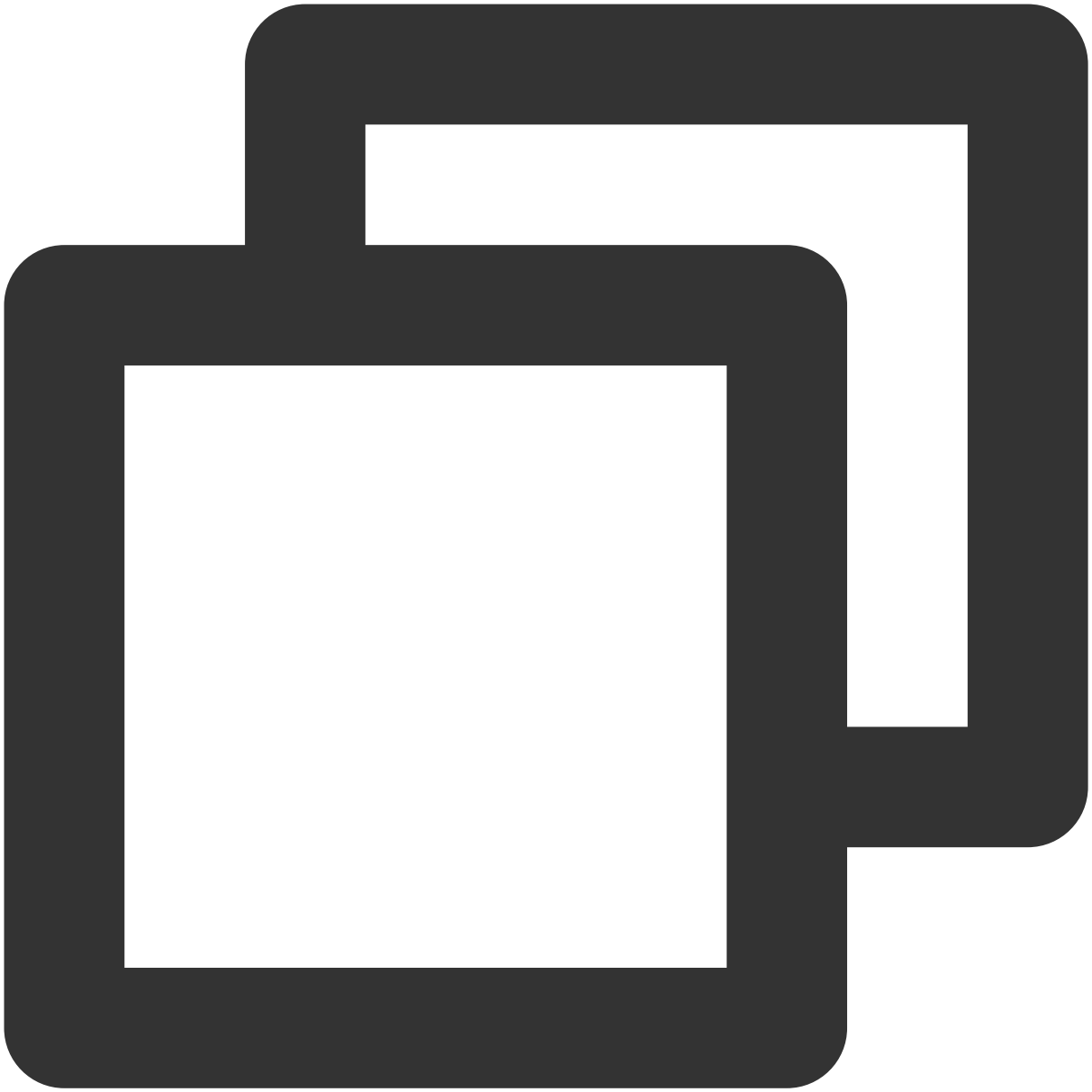
```
ALTER TABLE tabel_demo ADD
PARTITION (dt = '2021-01-01');
```

示例2：多层分区目录嵌套



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

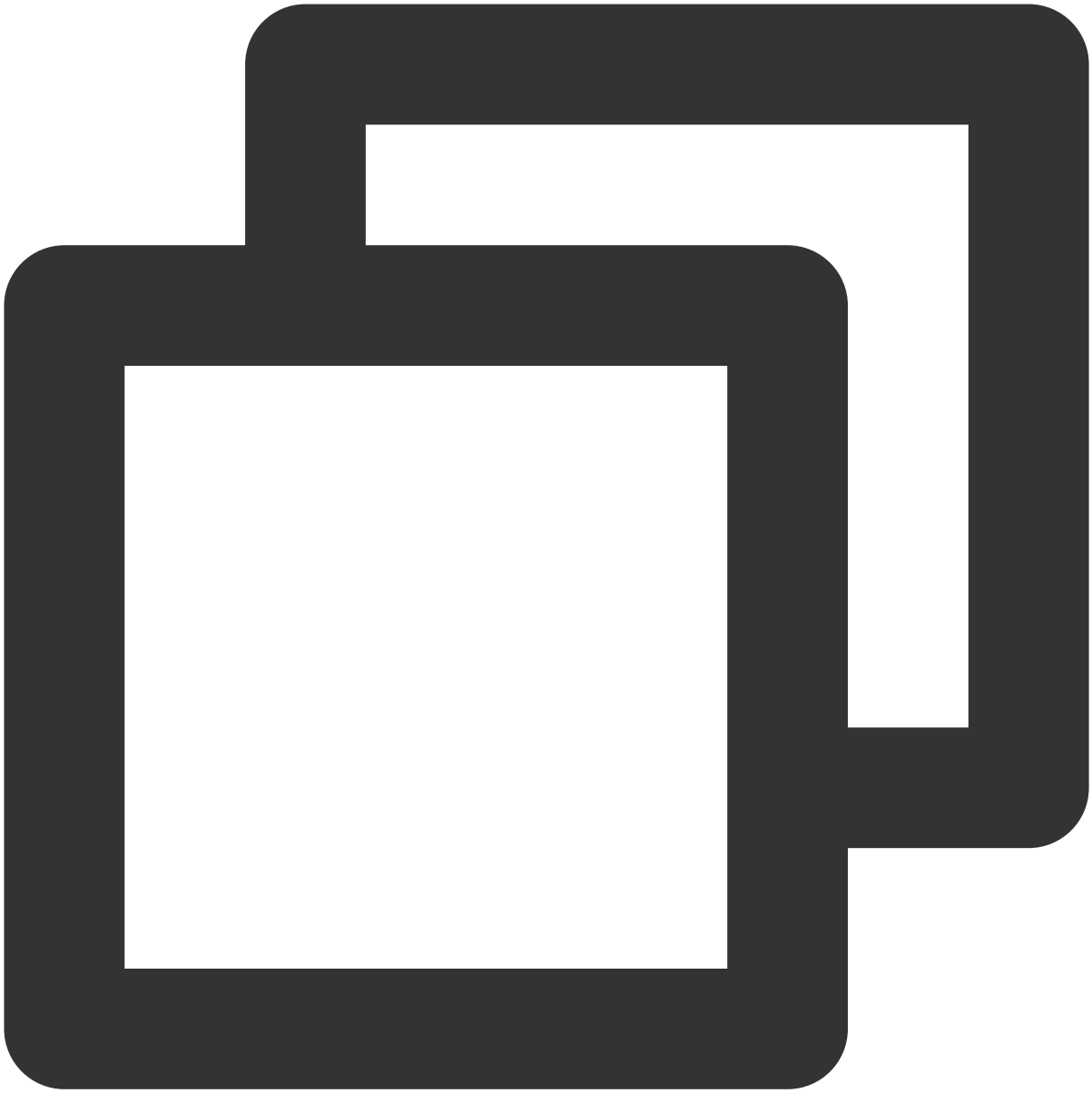
示例3：显示指定分区路径



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

自动添加分区

使用 `MSCK REPAIR TABLE` 语句，扫描在建表时指定的数据目录。若存在新的分区目录，则系统会自动将这些分区添加到数据表的元数据信息中。示例如下：



```
MSCK REPAIR TABLE table_demo
```

系统约束

`MSCK REPAIR TABLE` 仅向数据表元数据添加分区，不会删除它们。想要删除已经添加的分区，可执行语句 `ALTER TABLE table-name DROP PARTITION` 。

若数据量非常大，MSCK REPAIR TABLE不是推荐的方案。系统会扫描全量数据，并且消耗很长的时间，可能会导致任务超时，使得数据表的分区信息处于不完整状态。

分区目录必须兼容 Hive 的分区规则：**分区列名=分区列值**，若不一致，请使用 `ALTER TABLE ADD PARTITIO` 加载分区。

请确保在单独的文件夹层次结构中保留单独表的数据。例如，假设 COS 服务中有表 A 的数据 `cosn://tablea_a` 和表 B 的数据 `s3://table_a/table_b` 如果两个表都按字符串进行分区， `MSCK REPAIR TABLE` 会将表 B 的分区添加到表 A。为了避免这种情况，请使用单独的文件夹结构，如 `cosn://tablea_a` 和 `cosn://tablea_b` 。

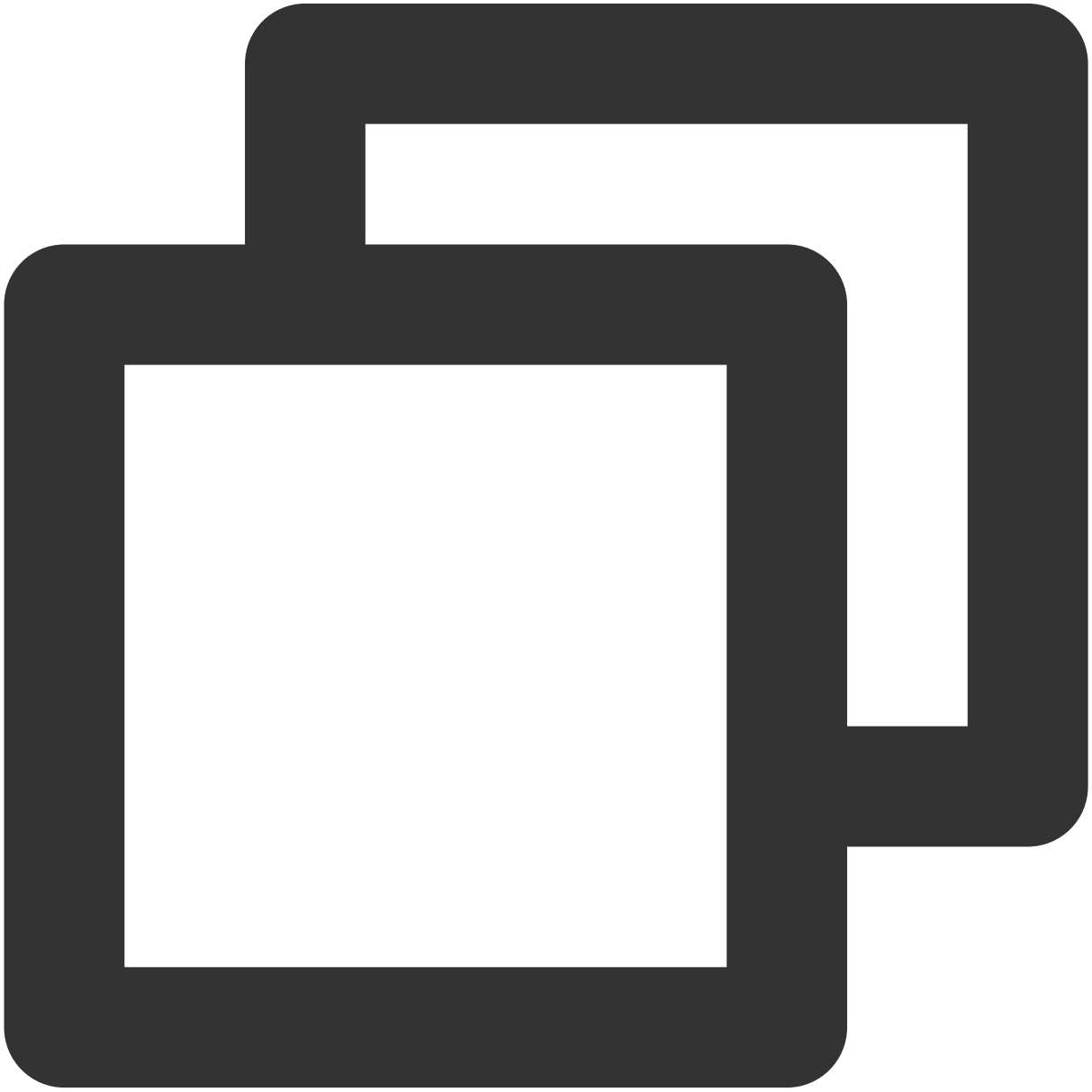
该语句可能会导致 COS 服务的数据读写开销，可参见 COS 服务的收费标准。

查询 Json 数据

最近更新时间：2024-07-17 16:18:24

查询步骤

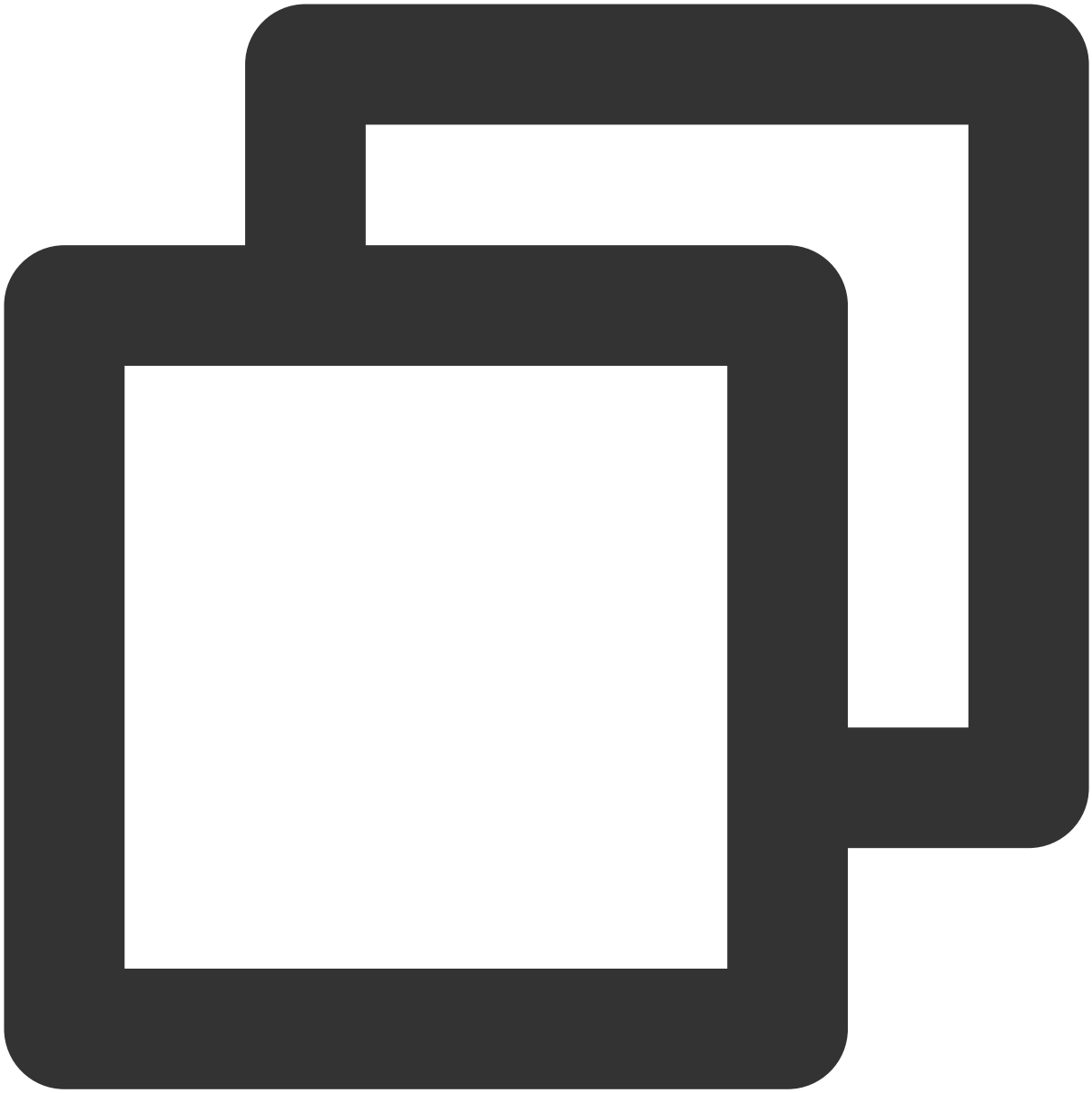
1. 创建数据表，指定 Json 解析格式。



```
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-b
```

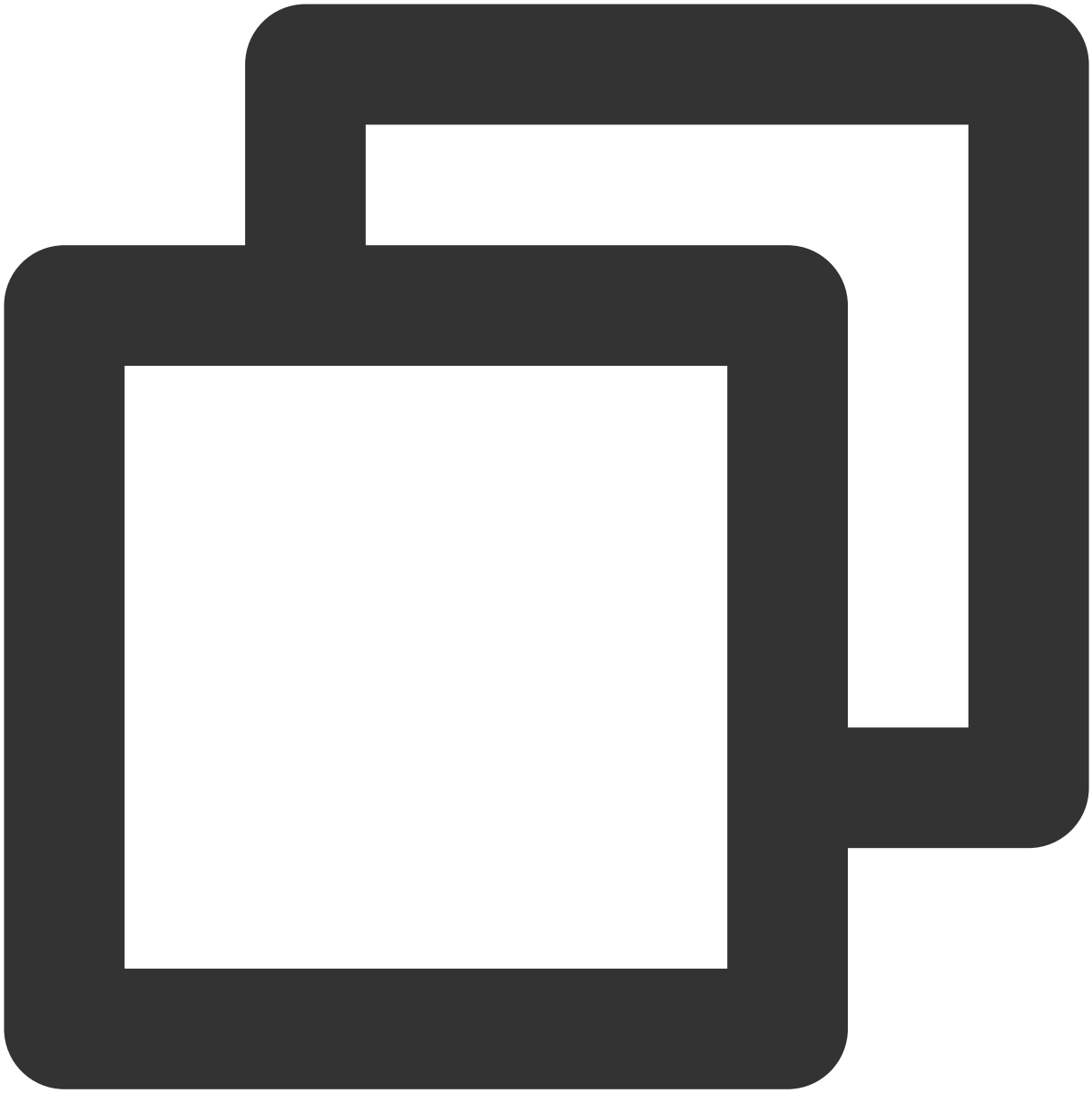
2. 执行查询语句查询 Json 数据。DLC 支持 Json 解析函数 `json_parse()`，`json_extract_scalar()`，`json_extract()` 等。



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

系统约束

必须是完整的 JSON 格式，否则 DLC 无法正常解析。
同一行数据不能有换行符，不能对 Json 进行可视化格式优化。例如：



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

DLC 会自动将 Json 的一层级识别为数据表的属性列，其余嵌套结构识别为对应的属性值。

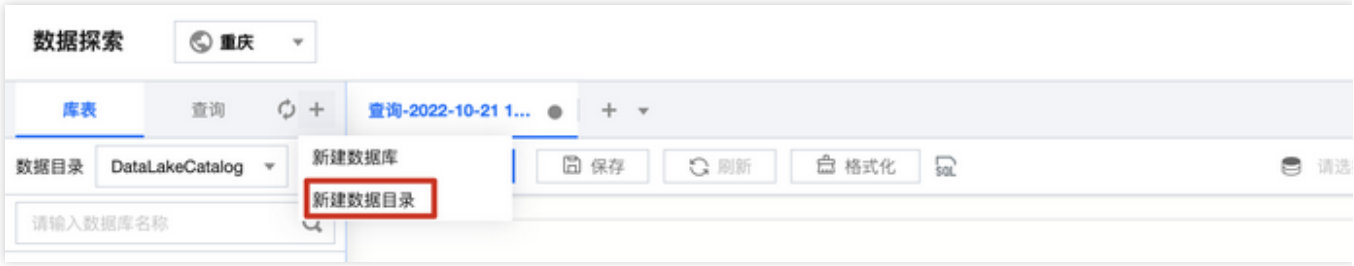
查询其他数据源

最近更新时间：2024-07-17 16:21:01

DLC 支持选择非托管的数据进行查询分析，目前支持 MySQL 及 EMR Hive 两种数据源的数据接入 DLC。用户可以通过 DLC 的控制台对其他数据源进行新增和管理操作。

新增数据源

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有创建数据目录的权限。
2. 进入**数据探索**页，鼠标移入“图标”出现菜单，单击**新建数据目录**。



3. 选择需要创建的数据源类型后，选择对应实例进行关联。

EMR Hive 目前支持2.0.1，2.1.0，2.2.0，2.2.1，2.3.0，2.4.0，2.5.0，2.5.1，2.6.0版本，通过 EMR 访问地址进行配置。

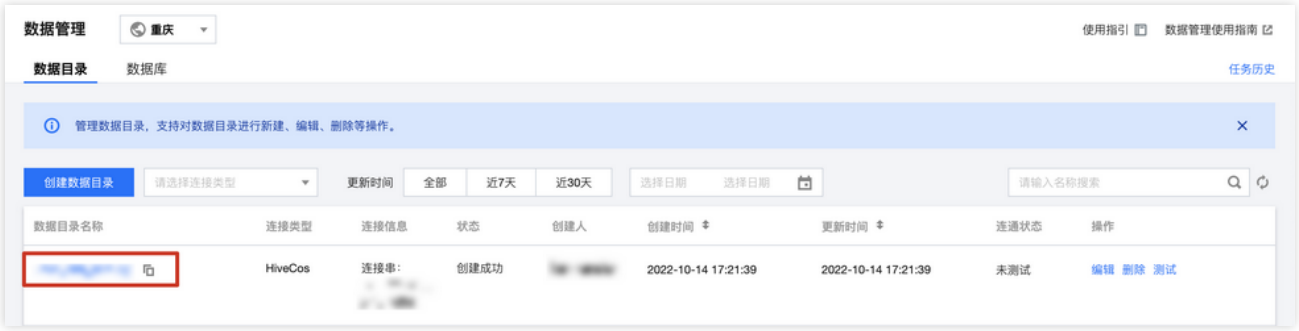
4. 填写数据源信息后单击**创建连接**，完成数据源的新增。

数据管理

目前 DLC 支持对非托管数据进行**数据库信息查看**及**数据表数据预览**功能。

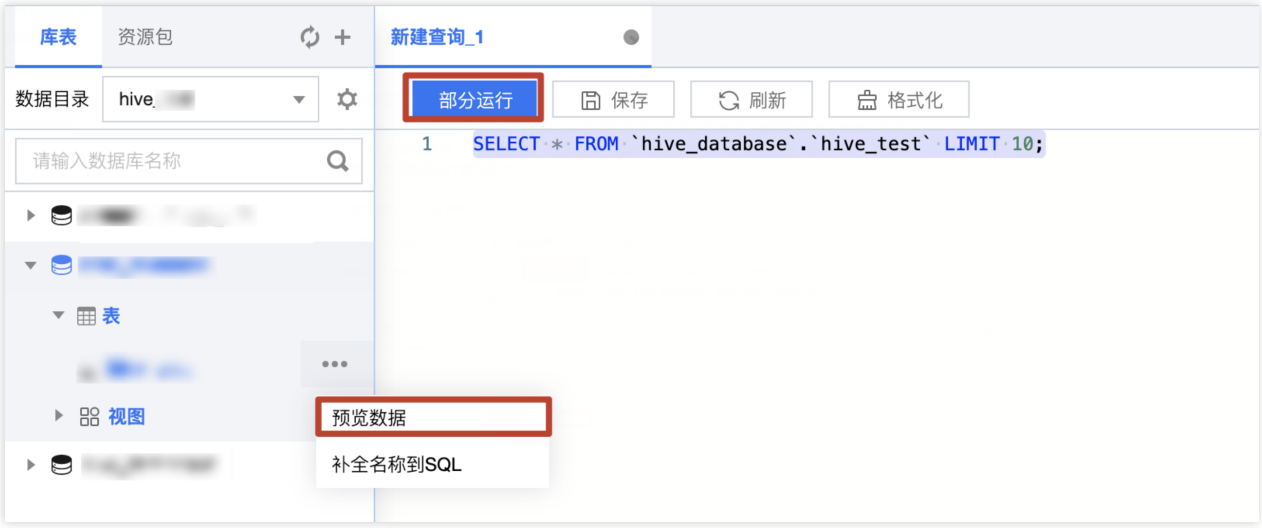
数据库信息查看

1. 登录 数据湖计算 DLC 控制台 ，选择服务所在区域，登录用户需要有数据表查看的权限。
2. 进入**数据探索**页，鼠标移入“图标”出现菜单，单击**基本信息**，弹窗内可查看数据表的基本信息。



数据表数据预览

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有数据表查看的权限。
2. 进入**数据探索**页，进入**数据表**，鼠标移入“图标”出现菜单，单击**预览数据**，执行 SQL 语句可将数据表的数据进行查询显示。



注意：

需选择与该数据源 VPC 所在网络配置绑定的数据引擎进行查询。

使用视图

最近更新时间：2024-07-17 16:22:27

DLC 中的视图是一个逻辑表，而非物理表。每次在查询中引用视图时，定义该视图的查询都会运行。可以从 `SELECT` 查询创建视图，然后在未来的查询中引用该视图。

系统约束

视图的名称不区分大小写，仅支持英文字符和下划线（_），最长为128个字符。

DLC 不支持使用视图管理数据的访问权限。

版权所有：腾讯云计算（北京）有限责任公司

INSERT INTO

最近更新时间：2024-07-17 16:23:25

INSERT INTO 语句可以将在源表上运行的 SELECT 查询结果作为新行插入到目标表中。

版权所有：腾讯云计算（北京）有限责任公司

查询脚本参数

最近更新时间：2024-07-17 16:24:02

为方便用户使用脚本参数进行查询，数据湖计算DLC支持配置日期参数进行查询。
数据湖计算 DLC 标准日期格式为：yyyymmddhh24miss，通过${}命令将日期设为变量，变量分为日期、时间两个部分，结果为两部分之和。
日期部分，可以是任意日期格式或者系统预定义变量，如 yyyymmdd、yyyymm、yyyy-mm-dd、yy、dataDate。
日期变化部分，可以+/-N个周期，支持N/Nd、Nm、Nw、Nh、Nmi。兼容支持计算公式，如7*N、N/24。
举例

获取加减周期	方法	兼容格式	举例
后 N 年	${yyyymmdd+Ny}	-	-
前 N 年	${yyyymmdd-Ny}	-	前一年：${yyyymmdd-12m}：20190920
后 N 月	-	${yyyymmdd+Nm}	-
前 N 月	${yyyymmdd-Nm}	$[add_months(yyyymmdd,-N)]	${yyyymmdd-1m}：20200820 ${yyyymm}：202009 ${dataDate-1m}：20200820
后 N 周	${yyyymmdd+Nw}	${yyyymmdd+7*N}	-
前 N 周	${yyyymmdd-Nw}	${yyyymmdd-7*N}	-
后 N 天	${yyyymmdd+N/Nd}	-	-
前 N 天	${yyyymmdd-N/Nd}	-	${yyyymmdd-1}、${dataDate-1}
后 N 小时	${yyyymmddhh24+Nh}	$[yyyymmddhh24+N/24]	-
前 N 小时	${yyyymmddhh24-Nh}	$[yyyymmddhh24-N/24]	${yyyymmddhh24-1h}：2020092014 ${dataDate-1h}： 2020092014
后 N 分钟	${yyyymmddhh24mi+Nmi}	$[yyyymmddhh24+N/24/60]	-
前 N 分钟	${yyyymmddhh24mi-Nmi}	$[yyyymmddhh24-N/24/60]	${yyyymmddhh24mi-10mi}、${dataDate-10mi}

注意：

使用时需保证变量或变量中+/-前的格式符合日期标准格式，否则将无法被系统识别使用。

查询脚本分析

最近更新时间：2024-08-07 17:09:00

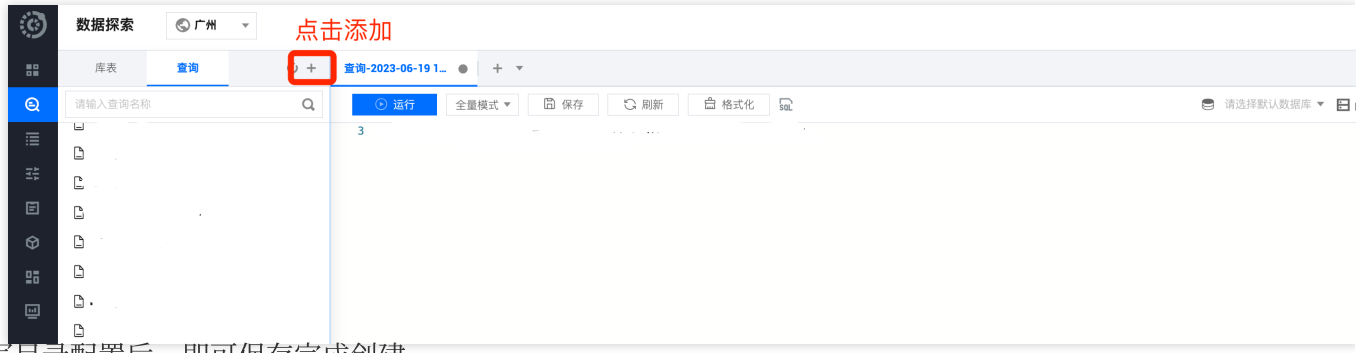
为方便用户快速处理重复的查询工作任务，数据湖计算 DLC 为用户提供了脚本文件分析。

说明

控制台中允许保存最多100个 SQL 脚本。

新建查询目录

1. 登录 数据湖计算 DLC 控制台 > 脚本查询页面。
2. 在查询页面中，单击添加查询目录。



3. 填写目录配置后，即可保存完成创建。

目录名称：支持中文、英文及“_”，最多25个字符。
权限设置：可根据工作组或用户视角，对该脚本目录及目录下的脚本，设置可见权限。

新建查询脚本

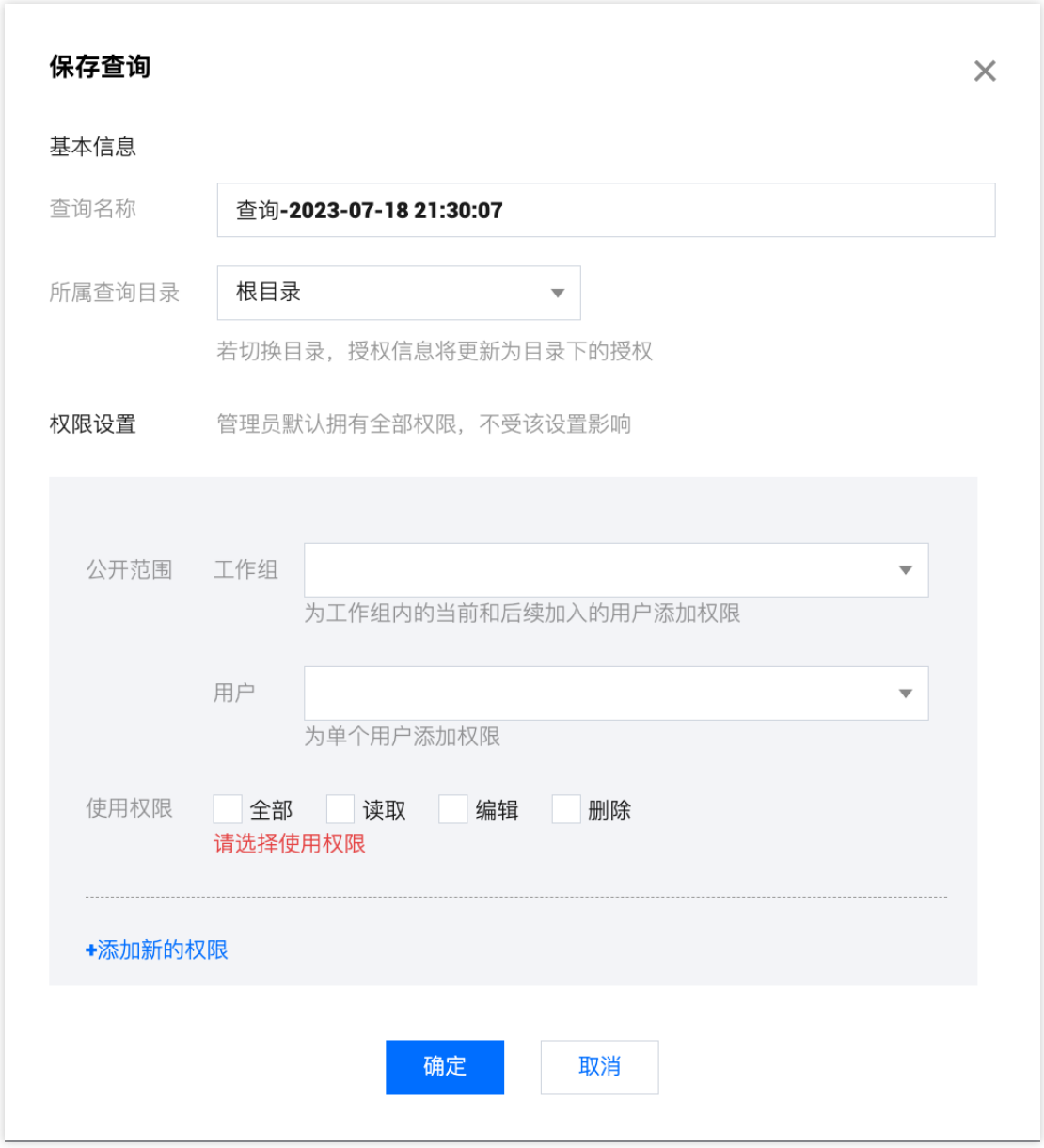
1. 登录 数据湖计算 DLC 控制台 > 脚本查询页面，可通过点击库

图标或直接新增执行与保存。
2. 选择计算引擎后，单击运行即可执行脚本。

保存查询脚本

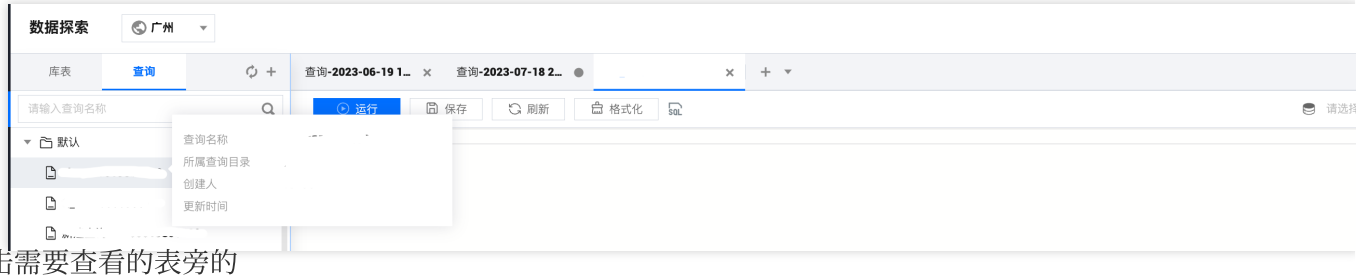
1. 查询完成后，单击“保存”按钮。
2. 通过库创建的查询，会保存在该库的目录下。通过 tab 栏添加查询的可直接存在根目录下或选择有权限的库保存。

3. 查询表的权限根据库的公开范围自定义选择，同时支持针对公开范围制定表的使用权限。



查看脚本信息

1. 鼠标悬停到需要查看的脚本名称，即可查看脚本详情信息。



2. 单击需要查看的表旁的

图标，即可选择打开或查询。

删除查询脚本

单击需要删除的表旁的

图标，即可选择删除脚本。

注意：

脚本删除后将无法恢复，请谨慎操作。

数据管理
数据目录及数据库管理

最近更新时间：2024-07-31 17:27:38

数据湖计算 DLC 的外部数据及托管存储数据可通过数据管理页、执行标准 SQL 语句及 API 进行管理。通过控制台数据管理页可创建、编辑数据目录，新建、查询、删除数据库表。

新建数据目录

说明：

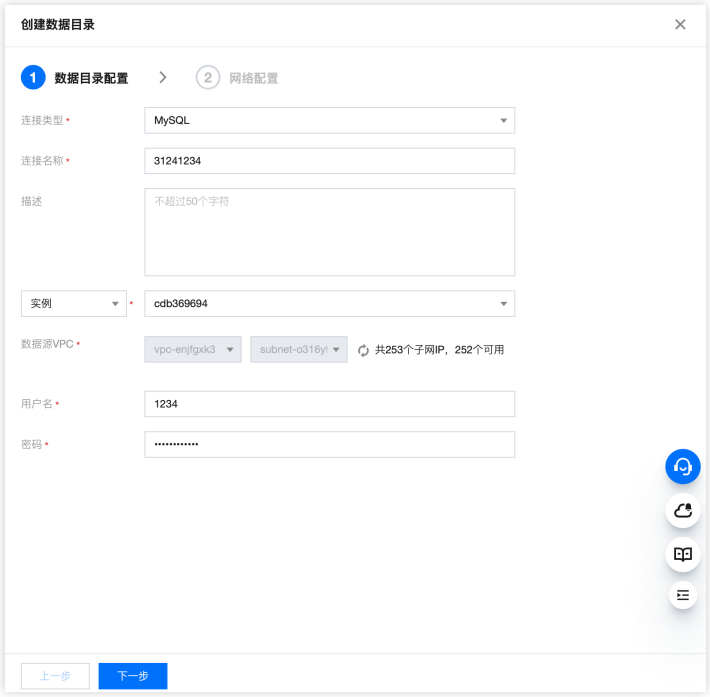
平台会默认帮您创建一个 DataLakeCatalog 用以湖上数据管理。

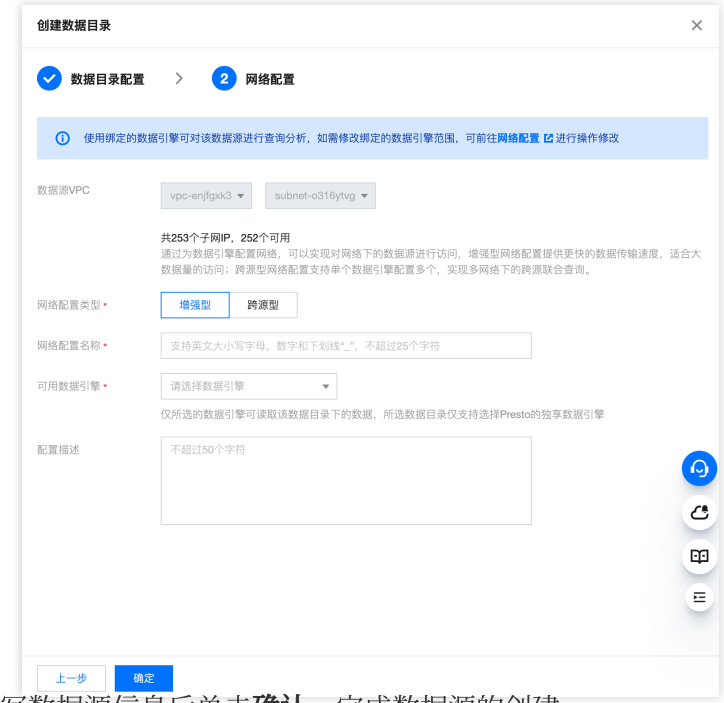
当您有外部数据源希望进行联邦分析时，可以参考以下流程创建外部数据源的数据目录。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录的账户须有创建目录权限，子账号权限开启可以参考子账号权限管理。

2. 进入数据管理，单击**创建数据目录**。

3. 进入数据源创建可视化界面。填写连接信息后，完成网络配置，打通引擎和外部数据源的网络。





4. 填写数据源信息后单击**确认**，完成数据源的创建。
5. 在数据目录列表查看连接信息、状态、创建人等信息。

编辑数据目录

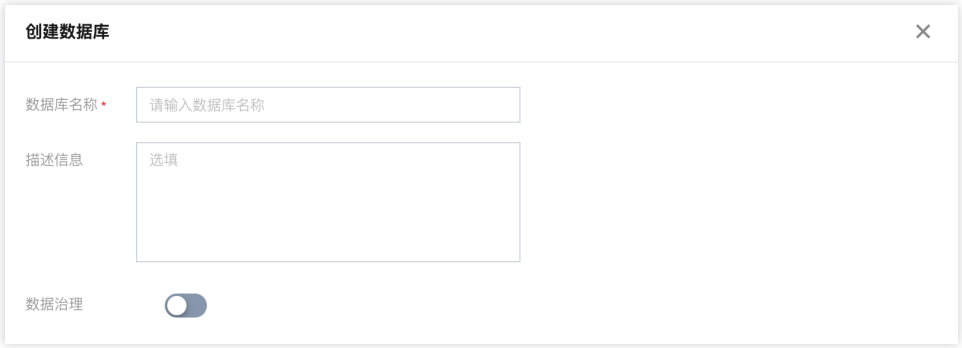
1. 单击**数据目录列表 > 操作 > 编辑**，支持修改数据目录的描述信息、网络配置信息、用户名、密码、运行集群等。

2. 修改后单击**创建**，数据目录重新创建。

新建数据库

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录的账户须有数据库新建权限。
2. 进入数据管理，单击数据目录下的目录名称，查看该目录下的数据库。
3. 单击**创建数据库**进入数据库创建可视化界面。

4. 填写数据库相关信息后保存，即可完成数据库新建。创建数据库时可以对整库开启数据优化。



数据库名称：全局唯一，支持英文大小写、数字、“_”，不允许数字开头，最多128个字符。
描述信息：支持中英文，最多2048个字符。
一个主账号最多可创建100个数据库。

查看数据库

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录的账户须有数据库查询权限。
2. 进入数据管理 > **数据库**，选择数据目录，单击**数据库名称**可进入数据库详情，可对数据库的数据表进行管理，详情操作指南可参见 数据表管理。

删除数据库

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录的账户须有数据库删除权限。
2. 进入 数据管理，单击**删除**，二次确认后即可将数据库删除。

数据表管理

最近更新时间：2024-07-31 17:28:16

用户可以使用数据湖计算 DLC 控制台或者 API，执行 DDL 语句创建数据库。

创建数据表

途径一：在数据探索进行创建

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有创建数据表的权限。
2. 进入 数据探索 模块，在左侧列表中，单击已有数据库，鼠标悬停表行，然后单击

图标，单击 **创建原生表** 或者 **创建外部表。**

说明：

原生表是指DLC托管存储上的表，原生表您无需关注底层Iceberg存储格式，且具备数据优化等能力。使用原生表需要先开启托管存储，详情参考托管存储配置。

外部表的底层数据在您自己的 对象存储 上，创建外部表需要指定数据路径。

3. 单击**创建原生表/创建外部表**后，系统会自动在生成创建数据表的 SQL 模板，用户可以修改 SQL 模板创建数据表。单击 **运行** 后，执行创建数据表的 SQL 语句，完成创建。

途径二：在数据管理进行创建

数据管理模块支持管理托管存储于 DLC 的原生表及外部表。

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有创建数据表的权限。

2. 通过左侧菜单进入**数据管理**，进入**数据库**，单击数据表所在的数据库名称，进入数据库管理页面。

3. 单击**创建原生表**或**创建外部表**按钮，进入数据表配置页面。

原生表数据来源支持空表、本地上传、对象存储 COS 三种不同类型，选择不同数据来源对应不同的创建流程。原生表支持数据优化等能力，可以选择继承库治理规则或单独启闭。

3.1 创建空表：创建一个没有记录的空表。

数据表名称：非数字开头，支持大小写字母、数字及下划线—，最多128个字符。

支持填写数据表描述信息。

可手动添加和输入列名称和字段类型。支持 array/map/struct 三种复杂类型字段的配置。

3.2 本地上传：将本地表单文件上传至 DLC 形成数据表，支持100MB 以内的文件。

CSV：支持可视化配置解析 CSV 规则，包括压缩格式、列分割符号、字段域符。支持自动推断数据文件的 Schema 和将首行解析为列名。

Json：DLC 仅将 Json 的第一层级识别为列，支持自动推断 Json 文件的 Shema，系统会将 Json 第一层字段识别为列名。

支持 Parquet、ORC、AVRO 等常见大数据格式数据文件。

可手动添加和输入列名称和字段类型。

若选择了自动推断结构，DLC 会自动填充识别到的列、列名称、字段类型，若不正确请手动修改。

3.3 通过对象存储 COS 创建数据表。

通过读取当前账户下的COS数据桶进行数据表创建。

CSV：支持可视化配置解析 CSV 规则，包括压缩格式、列分割符号、字段域符。支持自动推断数据文件的 Schema 和将首行解析为列名。

Json：DLC 仅将 Json 的第一层级识别为列，支持自动推断 Json 文件的 Shema，系统会将 Json 第一层字段识别为列名。

支持 Parquet、ORC、AVRO 等常见大数据格式数据文件。

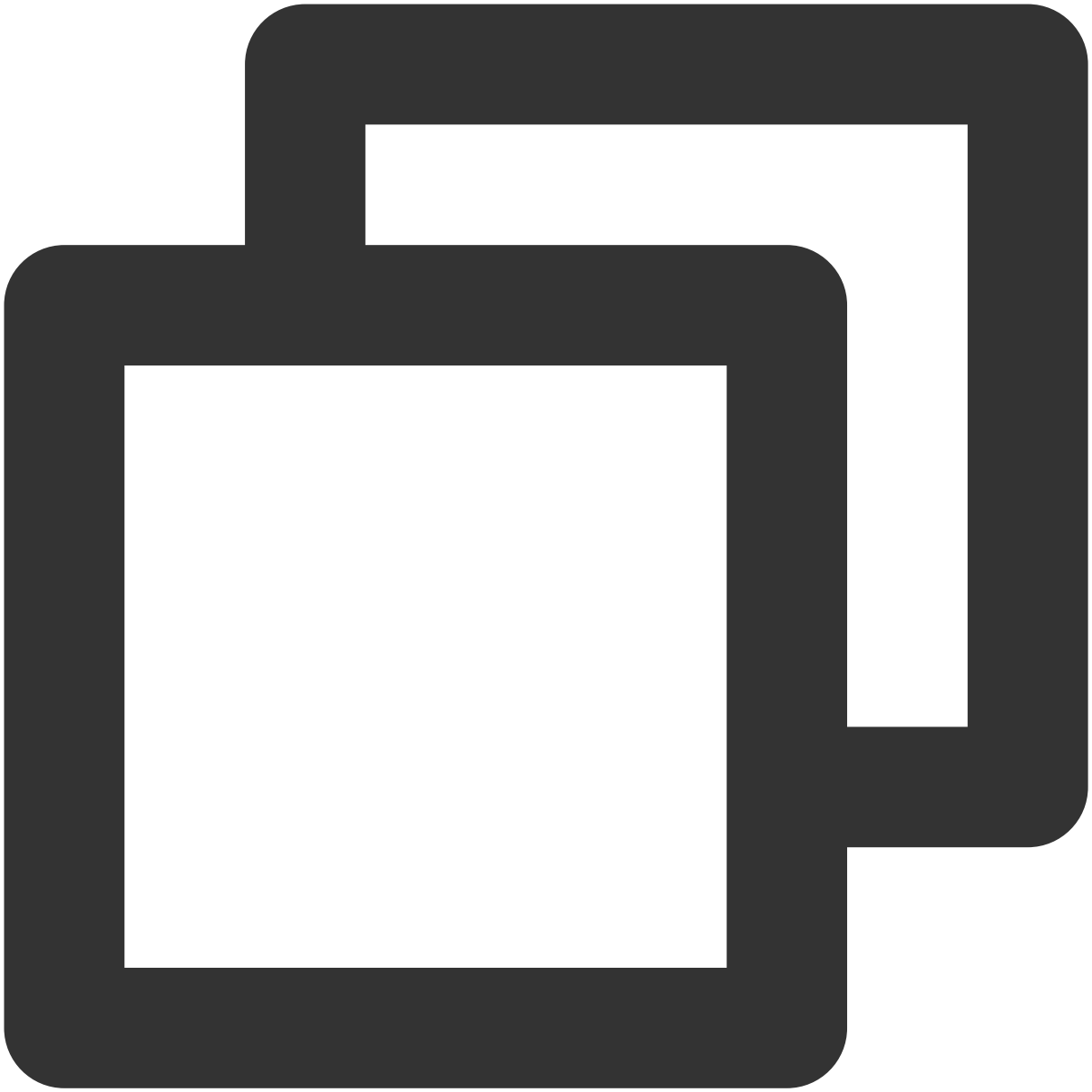
可手动添加和输入列名称和字段类型。

若选择了自动推断结构，DLC 会自动填充识别到的列、列名称、字段类型，若不正确请手动修改。

4. 数据分区通常为提高查询性能，会对大数据量的表进行分区。DLC 支持按照数据分区查询数据，用户需要在此步骤添加分区信息。通过分区您的数据，您可以限制每个查询扫描的数据量，从而提高查询性能并降低使用成本。DLC 遵从 Apache Hive 的分区规则。

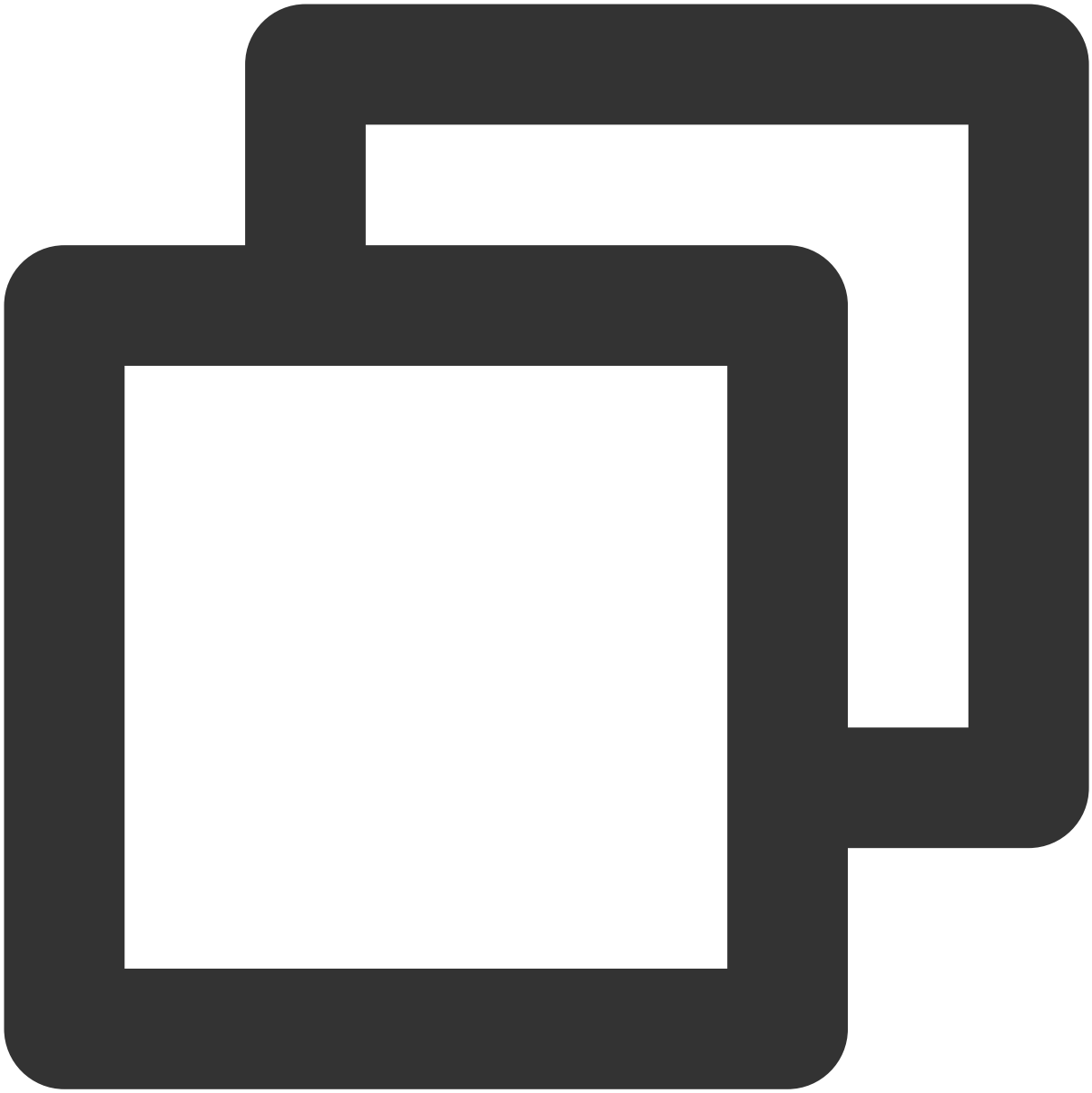
分区列对应表的 COS 路径下的一个子目录，目录的命名规则为**分区列名=分区列值**。

示例：



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

如果有多个分区列，则需要按照建表语句中指定的分区列顺序依次嵌套。



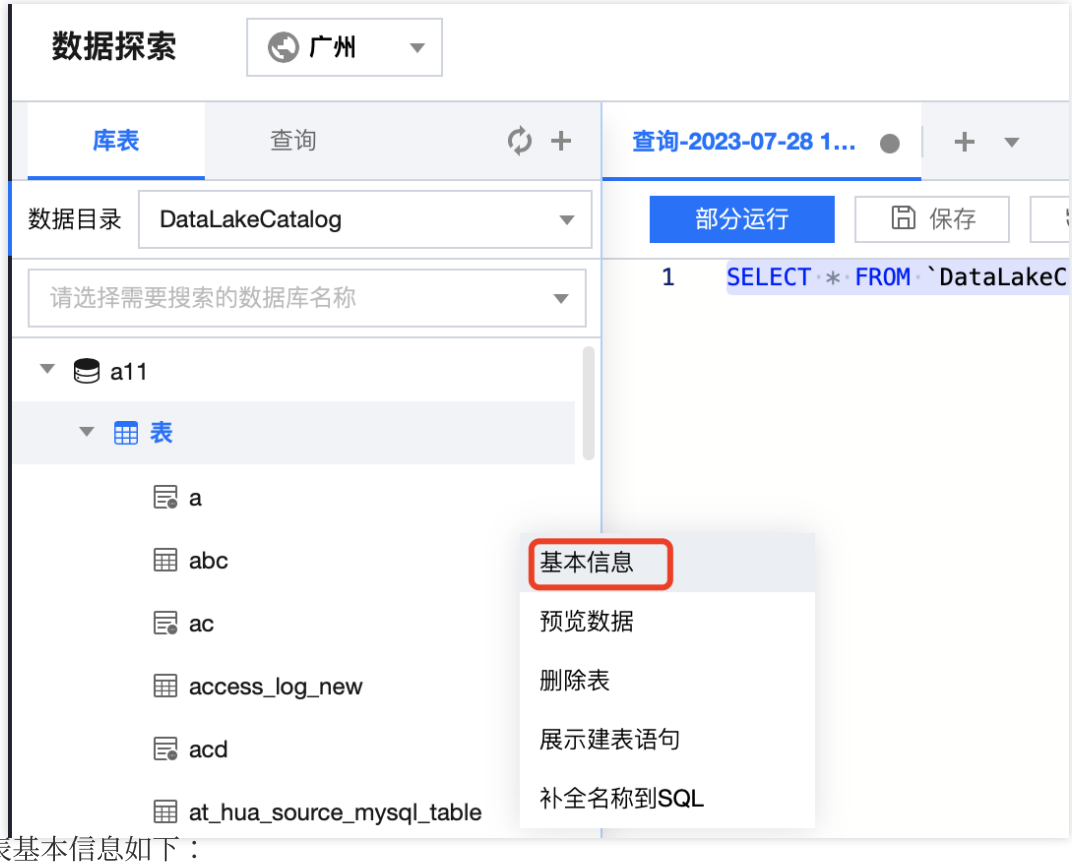
```
CREATE EXTERNAL TABLE IF NOT EXISTS `COSDataCatalog`.`dlc_demo`.`table_demo` (
    `_c0` string,
    `_c1` string,
    `_c2` string,
    `_c3` string
) PARTITIONED BY (`year` string, `month` string, `day` string)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES ('separatorChar' = ',', 'quoteChar' = '"')
STORED AS TEXTFILE
LOCATION 'cosn://bucket_name/folder_name/';
```

查询数据表基本信息

途径一：在数据探索中进行查询

在数据表项目中，鼠标悬停**数据表名称**行，然后单击

图标，在下拉菜单中单击**基本信息**，可以查看创建的数据表基本信息。



数据表基本信息如下：

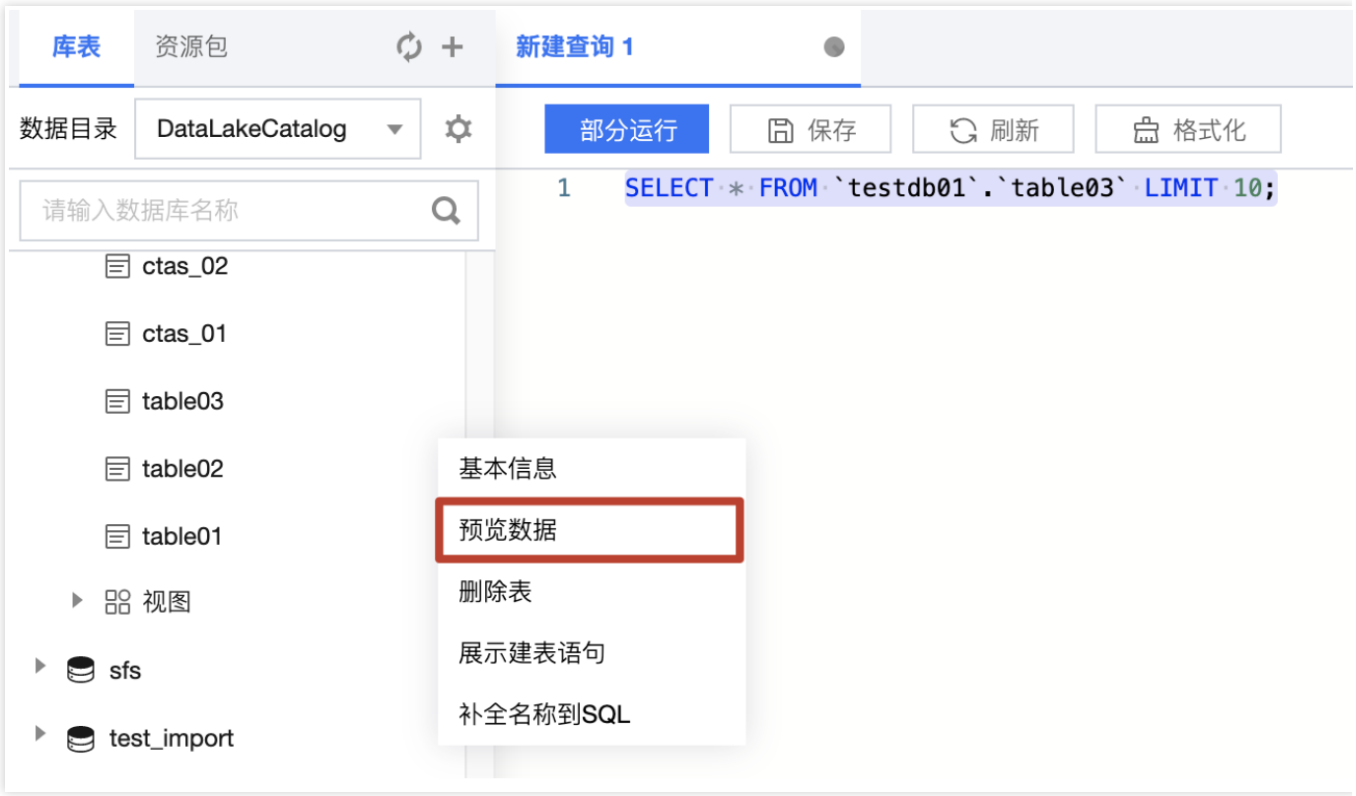
途径二：在数据管理进行查看

1. 登录数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有查看数据表的权限。
2. 通过左侧菜单进入**数据管理**页面，单击数据表所在的数据库名称，进入数据库管理页面，支持查询数据表的行数、存储空间、创建人、字段、分区等信息。

预览数据表数据

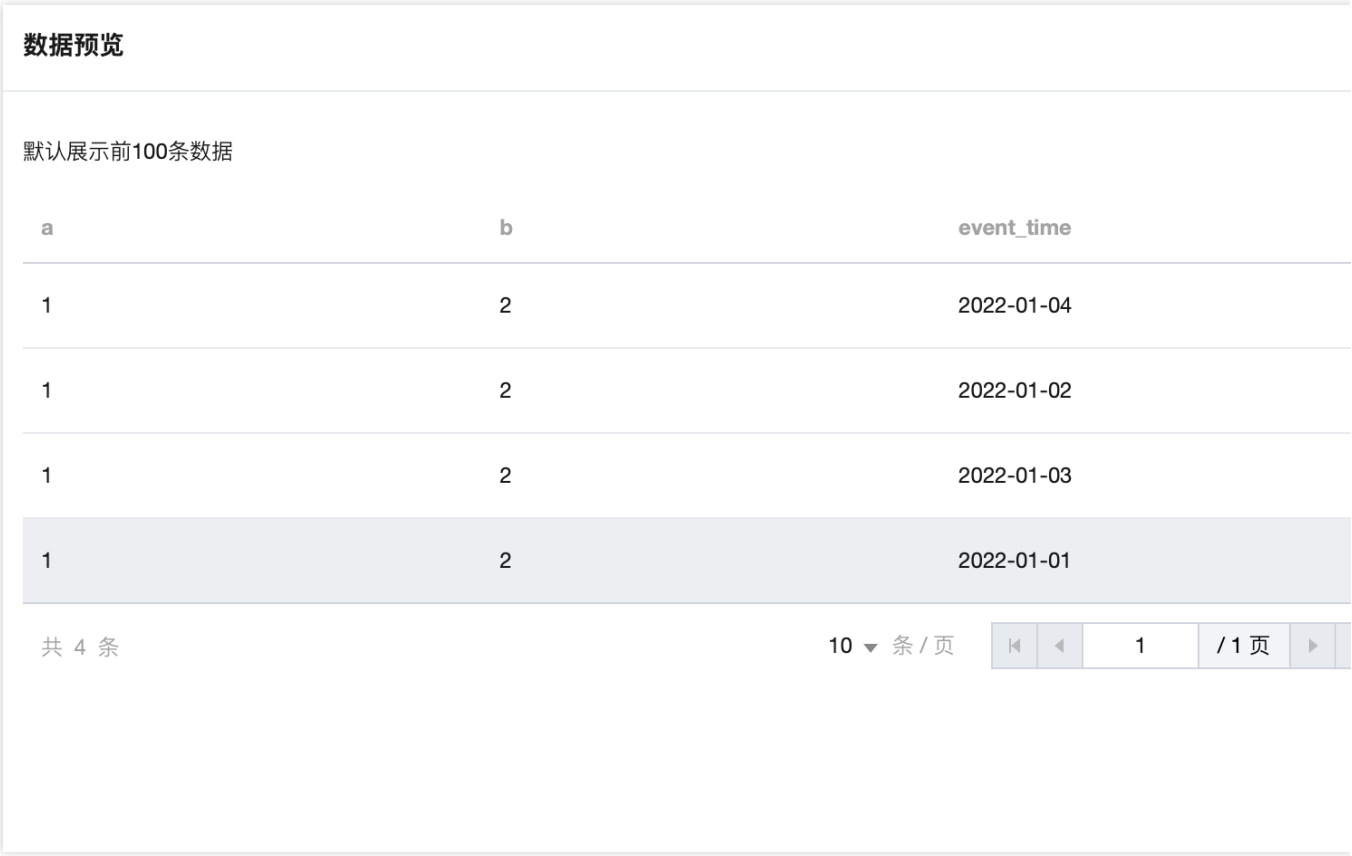
在数据表项目中，鼠标悬停**数据表名称**行，然后单击

图标，在下拉菜单中单击**预览数据**，DLC 会自动生成预览10条数据的 SQL 语句，执行 SQL 语句查询数据表前10条数据。



支持在**数据管理 > 数据库 > 数据表 > 数据表列表**中预览数据。

数据预览功能默认展示前100条数据。



编辑数据表信息

支持在数据管理模块编辑数据表的描述信息。

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有编辑数据表的权限。
2. 通过左侧菜单进入**数据管理 > 数据库**页面，单击数据表所在的数据库名称，进入数据库管理页面。
3. 找到需要编辑的数据，单击右侧的**编辑**按钮进行编辑。

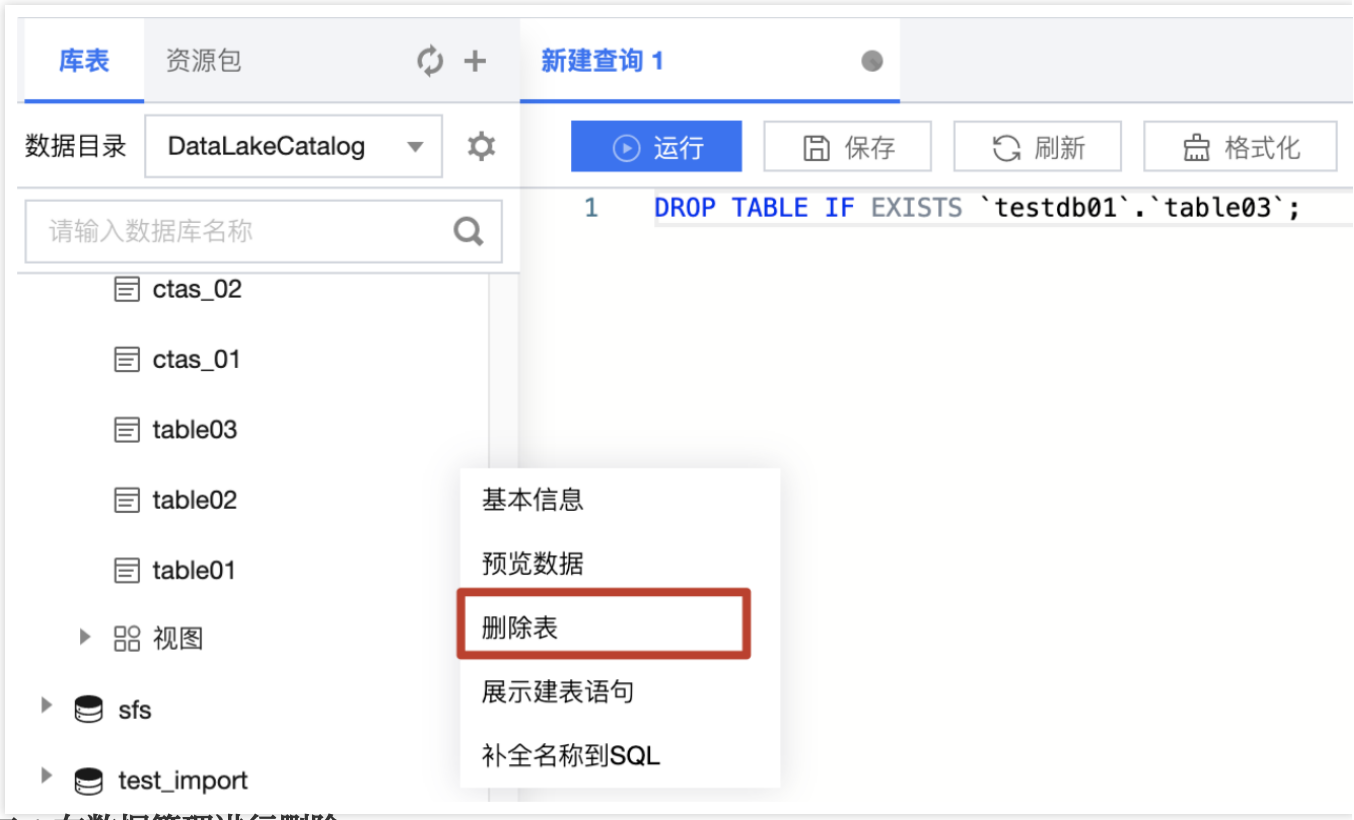
4. 修改后单击**确定**按钮即可完成编辑。

删除数据表

途径一：在数据探索进行删除

在数据表项目中，鼠标悬停**数据表名称**行，然后单击

图标，在下拉菜单中单击**删除表**，DLC 会自动生成删除数据表的 SQL 语句，执行 SQL 语句删除数据表。
删除外表，删除数据表仅是删除 DLC 中存储的元数据信息，不会影响数据源文件。
删除DataLakerCatalog目录下的数据表，将会清空所有该数据表数据，请谨慎操作。



途径二：在数据管理进行删除

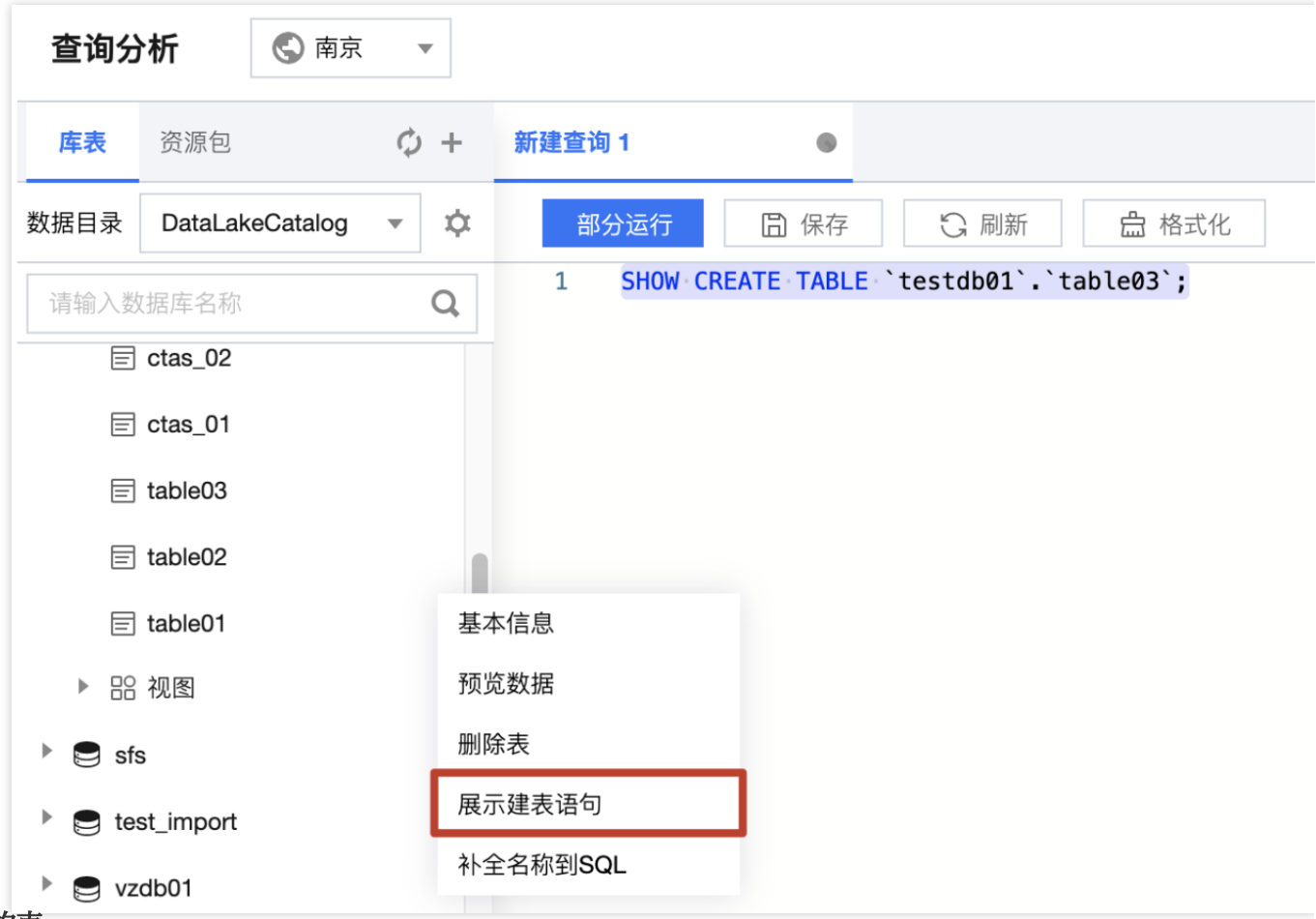
目前数据管理内仅支持管理托管存储于DLC的数据库表，外表请使用途径一进行删除。

1. 登录数据湖计算 DLC 控制台，选择服务所在区域，登录用户需要有删除数据表的权限。
2. 通过左侧菜单进入**数据管理 > 数据库**，单击数据表所在的数据库名称，进入数据库管理页面。
3. 单击需要删除的数据表后的**删除**按钮，二次确认后可删除对应的数据表，同时清空数据表数据。

展示建表语句

在数据表项目中，鼠标悬停**数据表名称**行，然后单击

图标，在下拉菜单中单击**展示建表语句**，DLC 会自动生成查看该数据表的建表语句，执行 SQL 语句查询建表语句。



系统约束

DLC 允许每个数据库下的数据表最大数量为4096个，每个数据表支持的最大分区数为100,000，每个数据表的属性列最大数量为4096个。

DLC 会把相同 COS 路径下的数据文件识别为同一张表的数据，请确保在单独的文件夹层次结构中保留单独表的数据。

DLC 不支持 COS 多版本数据，只能查询 COS 存储桶中最新版本的数据。

DLC 上创建的所有表是外表，在创建表的 SQL 语句中必须包含 EXTERNAL 关键字。

表名称在同一个数据库下必须唯一。

表名称不区分大小写，仅支持英文字符、数字和下划线（_），最长为128个字符。

若表为分区表时，需要手动执行 ADD PARTITION 语句或者 MSCK 语句，添加分区信息，才能查询到该分区数据。详情可参见 查询分区表。

使用 CSV 建表时，DLC 会默认把所有字段类型转化为 string，但是不影响原数据字段的计算和查询。

数据视图管理

最近更新时间：2024-07-31 17:28:47

数据湖计算 DLC 提供了数据视图查询能力，用户可通过管理数据视图进行便捷快速的数据查询及使用。

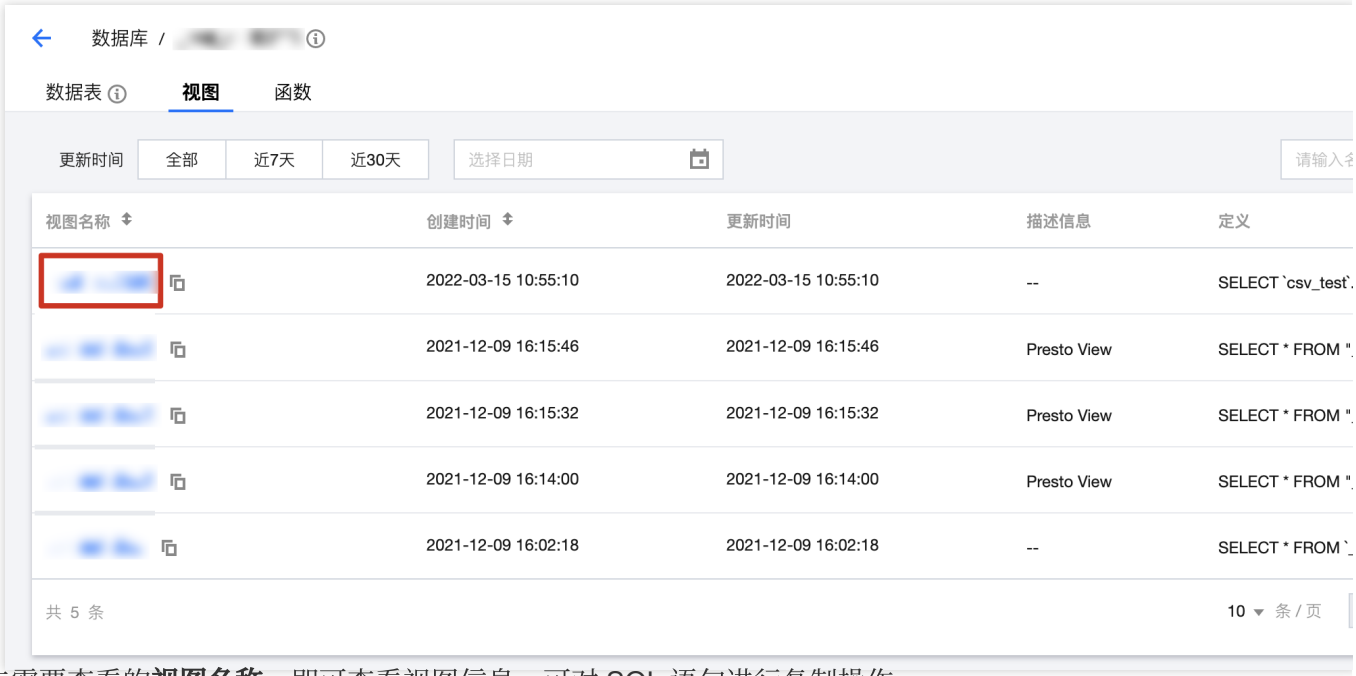
创建视图

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录用户须具有视图创建权限。
2. 进入**数据探索页**，通过 SQL 语句即可创建视图，语句详情参见 SQL 语法。
3. 选择计算资源，单击**运行**按钮之后即可完成视图创建。

查看视图

您可以通过数据探索使用 SQL 语句对视图进行查看，具体语法参见 SQL 语法。同时数据湖计算 DLC 也提供了可视化界面对视图进行管理，操作如下。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录用户须具有视图查询权限。
2. 进入数据管理页，单击视图所在的**数据库名称**进入数据库管理页。
3. 单击**视图**进入视图管理。

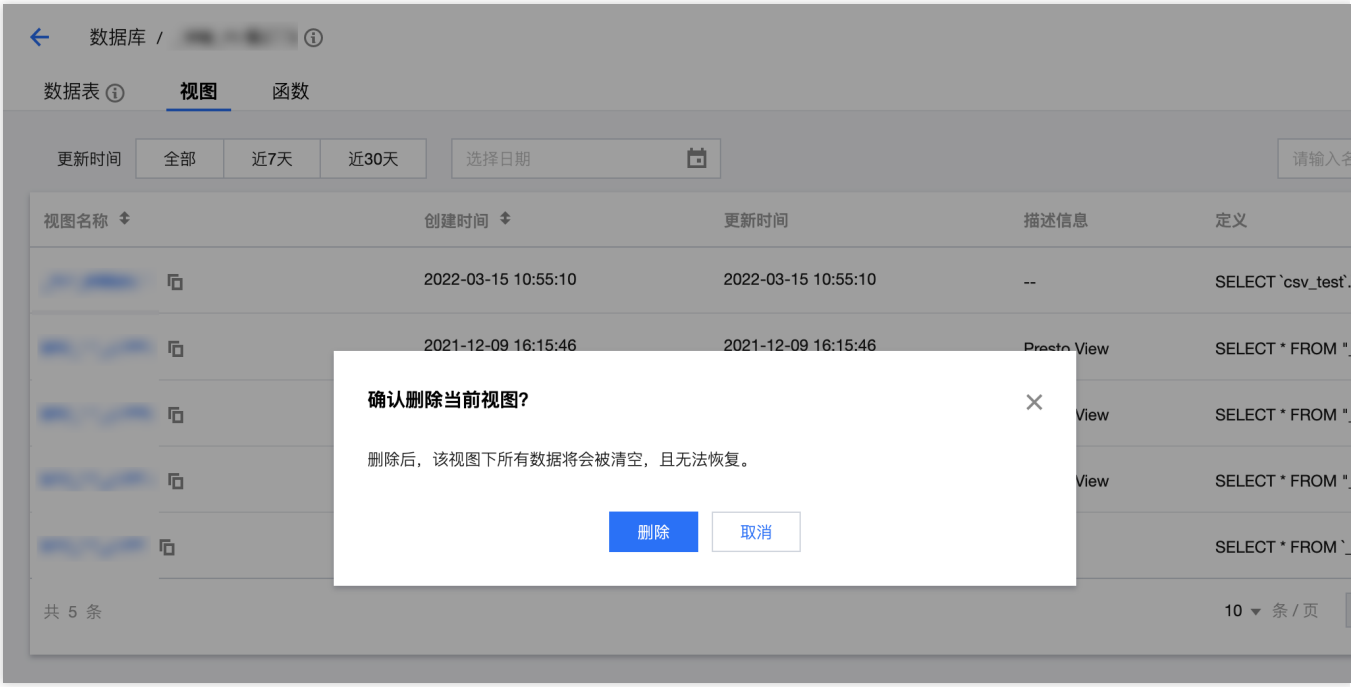


4. 单击需要查看的**视图名称**，即可查看视图信息，可对 SQL 语句进行复制操作。

删除视图

您可以通过数据探索使用 SQL 语句对视图进行查看，具体语法参见 SQL 语法。同时数据湖计算 DLC 也提供了可视化界面对视图进行管理，操作如下。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录用户须具有视图删除权限。
2. 进入数据管理页，单击视图所在的**数据库名称**进入数据库管理页。
3. 单击**视图**进入视图管理，单击**删除**按钮即可将视图删除。



注意

删除视图将清空视图下的所有数据，且无法恢复，请谨慎操作。

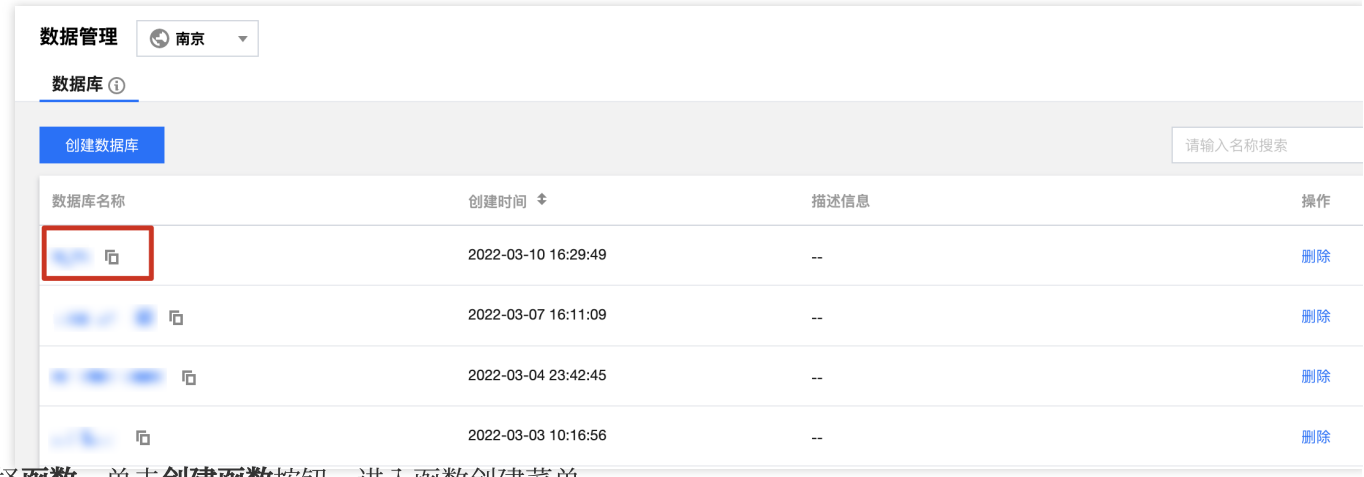
函数管理

最近更新时间：2024-07-31 17:29:05

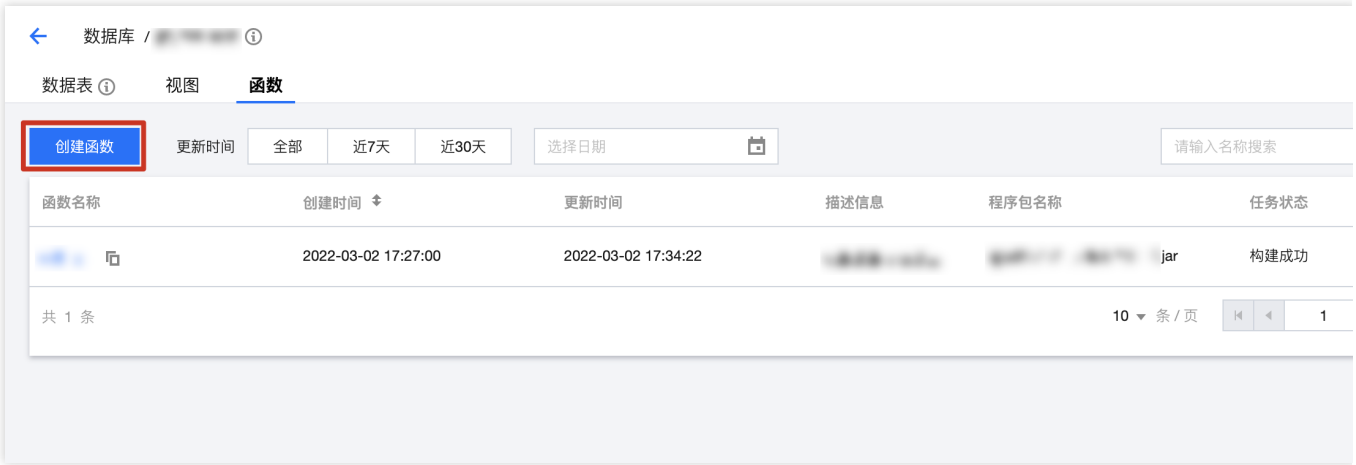
数据湖计算 DLC 中支持使用**用户自定义函数**对数据进行处理构建，并支持管理函数。

创建函数

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录账号需具备数据库操作权限。
2. 进入**数据管理页**，单击需要构建函数的**数据库名称**。



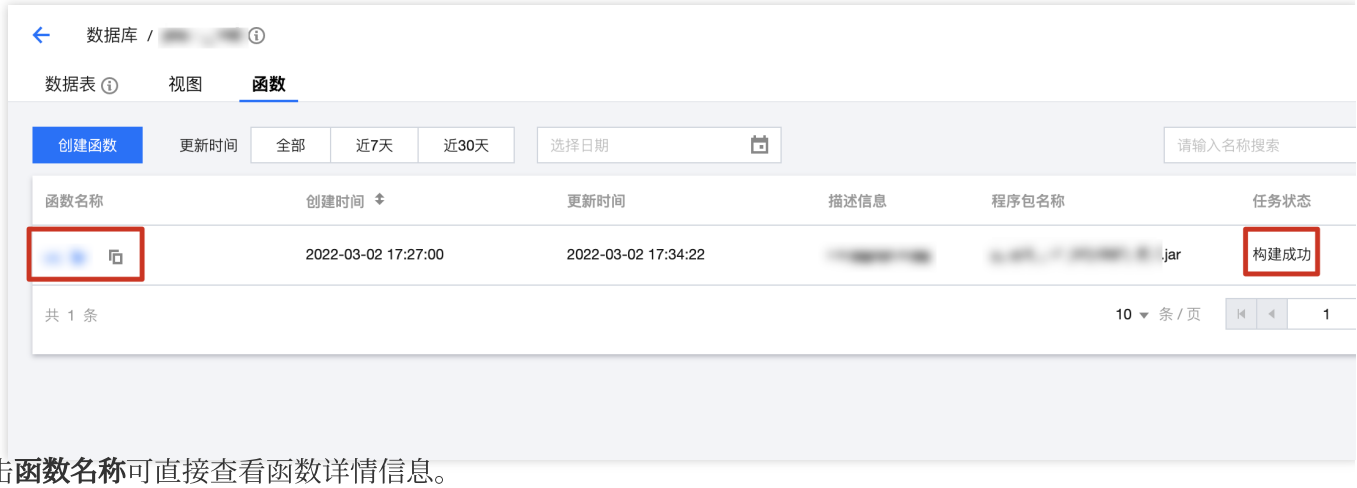
3. 选择**函数**，单击**创建函数**按钮，进入函数创建菜单。



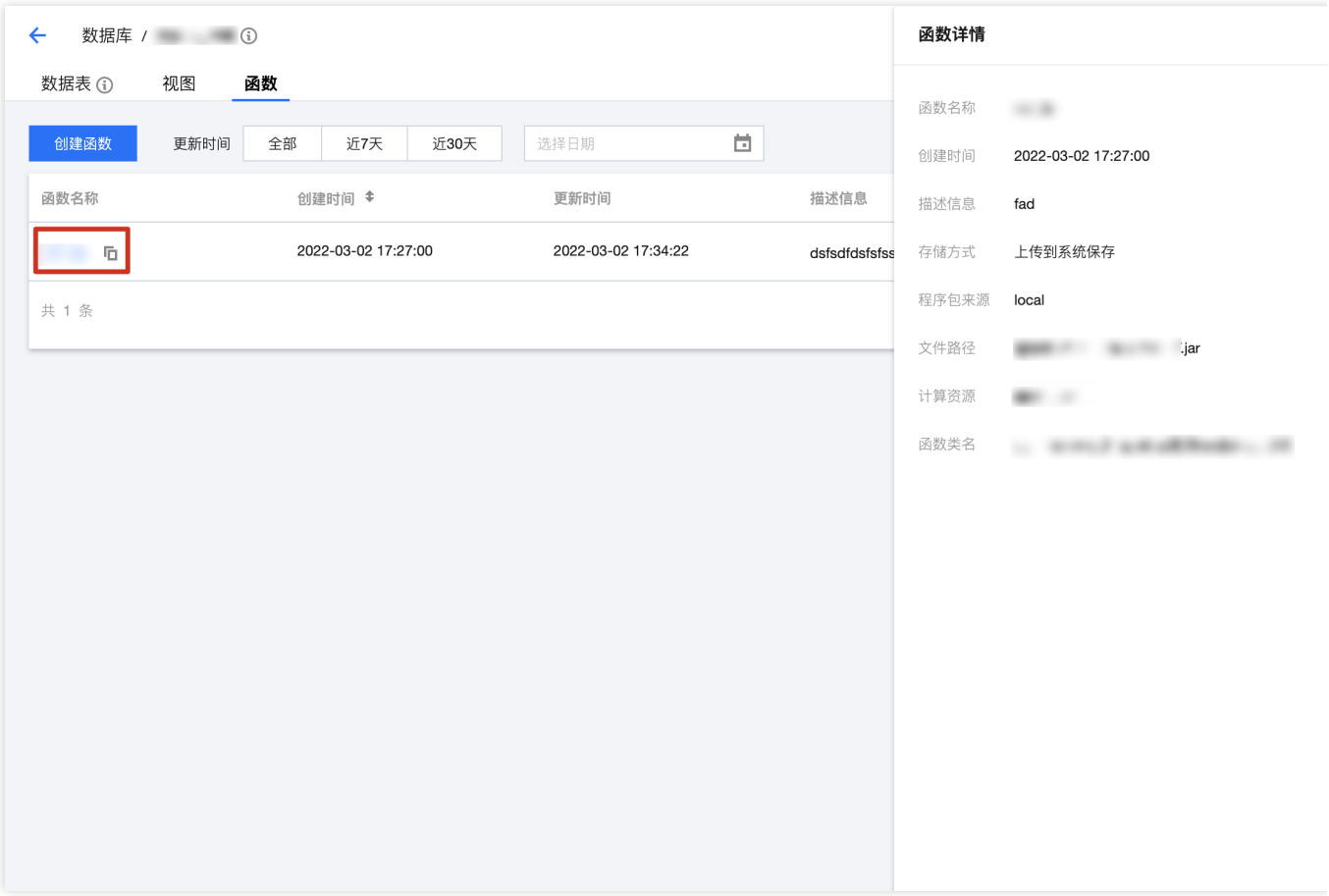
函数包支持本地上传或使用 COS 上已有的 JAR 包，本地上传仅支持 JAR 格式，最大5MB。
需选择 spark 集群进行运行，运行过程中将不会产生任何费用。
建议将函数包保存到系统，方便对函数包进行管理及使用。同时也支持挂载至指定 COS 路径。

查看函数信息

1. 登录 数据湖计算 DLC 控制台，登录账号需具备数据库操作权限。
2. 进入**数据管理页**，单击需要查看函数的**数据库名称**。
3. 选择函数，即可查看函数的构建状态，若构建失败可**编辑**后再次提交。

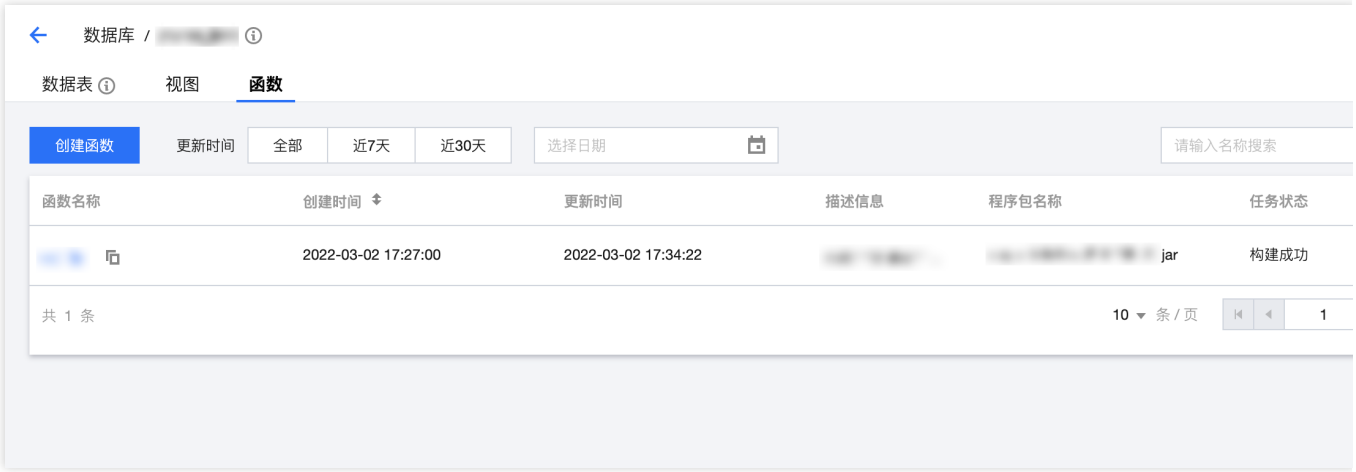


4. 单击**函数名称**可直接查看函数详情信息。



编辑函数信息

1. 登录 数据湖计算 DLC 控制台，选择服务地域，登录账号需具备数据库操作权限。
2. 进入**数据管理页**，单击需要查看函数的**数据库名称**。
3. 选择**函数**，单击**编辑**按钮进入函数信息编辑页面。



暂不可修改函数名称、存储方式以及上传方式，如需修改此类信息可重新创建函数。
修改函数信息后，将进行重新构建，请谨慎操作。

删除函数

针对无需继续使用管理的函数，可对函数进行删除操作。

1. 数据湖计算 DLC 控制台，选择服务地域，登录账号需具备数据库操作权限。
2. 进入**数据管理页**，单击需要查看函数的**数据库名称**。
3. 选择**函数**，单击**删除**按钮对无需继续使用的函数进行删除。

注意

删除后该函数下的数据将会被清空，且无法恢复，请谨慎操作。

分区字段策略

最近更新时间：2024-07-31 17:29:22

在 Hive 中，分区信息是以目录形式出现的。在 Iceberg 中，分区信息记录到最终底层数据文件中，这使得 Iceberg 的分区更加灵活，而且可以随着数据量的变化演变分区策略。在数据湖计算 DLC 中，您可以创建Iceberg 表来使用隐藏分区等特性。

说明：

原生表默认为 Iceberg 表，外部表根据文件格式可以选择 Hive 表或者 Iceberg 表，详细语法参考文档 CREATE TABLE。

隐藏式分区您在插入和查询数据时，不需要像 Hive 一样额外指定分区信息。

Iceberg 分区策略支持使用以下函数，不同字段及对应的分区转换策略如下表：

分区策略	字段类型	结果类型
identity	任何	源类型
bucket	int, long, decimal, date, time, timestamp, timestamptz, string, uuid, fixed, binary	int
truncate	int, long, decimal, string	源类型
year	date, timestamp, timestamptz	int
month	date, timestamp, timestamptz	int
day	date, timestamp, timestamptz	date
hour	timestamp, timestamptz	int

数据作业
数据作业概述

最近更新时间：2024-07-17 17:42:32

数据湖计算 DLC 提供了基于原生 Spark 的批处理、流计算能力，支持用户通过数据任务进行复杂的数据处理、ETL 等操作。
目前数据作业的 Spark 相关版本支持如下：
Scala 2.12.*版本。
Spark 3.1.2版本。

使用准备

在开始使用数据作业前，为了保证您的数据安全，您需要先创建数据访问策略，详细配置说明请参见 配置数据访问策略。
同时，我们支持使用 Ckafka 作为数据源进行数据作业配置。后续我们也将持续丰富数据源支持，为您提供更全面的数据作业使用场景。

计费模式

数据作业将根据使用数据引擎进行计费，目前支持按量计费和包年包月两种模式。详情可参见 数据引擎说明。
按量计费：适合数据作业量较少或周期性使用的场景，创建作业后拉起使用，作业运行完成后自动挂起不再产生费用。
包年包月：适合数据作业量较大同时很稳定使用的场景，按月付费预留资源，无需等待数据引擎拉起。
注意：
由于数据作业与 SQL 作业的计算引擎类型差异，需要单独购买 Spark 作业类型的数据引擎，无法使用 SparkSQL 类型的数据引擎运行数据作业。

作业管理

通过**数据作业**管理菜单，您可以对数据作业进行创建、启动、修改、删除。
1. 登录 数据湖计算控制台，单击左侧菜单**数据作业**进入数据作业管理页。
2. 单击**创建数据作业**按钮，即可创建新的数据作业，详细步骤请参见 创建数据作业。
3. 在列表内可以查看数据作业当前任务的状态，同时支持管理数据作业，详细步骤请参见 管理数据作业。

版权所有：腾讯云计算（北京）有限责任公司

配置数据访问策略

最近更新时间：2024-07-17 17:45:15

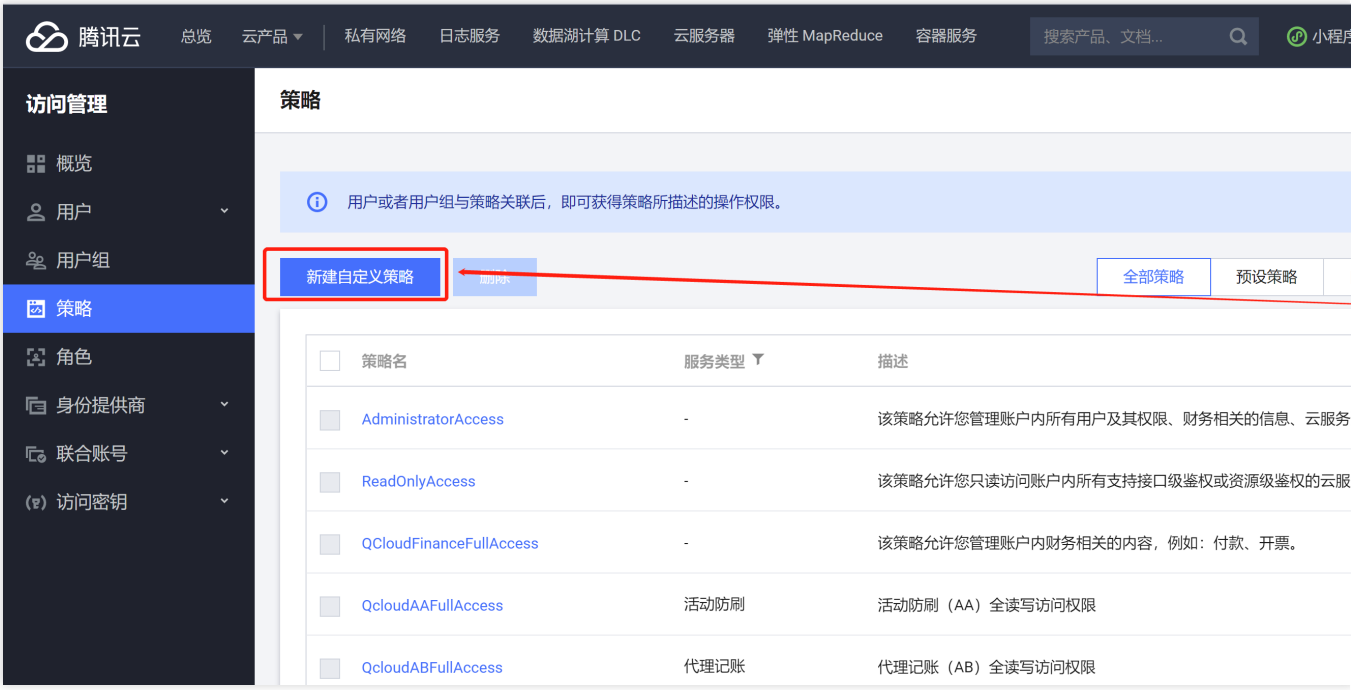
什么是数据访问策略（CAM role arn）

数据访问策略（CAM role arn）是为了保障数据作业运行过程中访问的数据源及对象存储 COS 上的数据安全，用户在访问管理（CAM）上对数据访问权限进行配置的策略。
在数据湖计算 DLC 中配置数据作业时，需指定对应的数据访问策略，以保证数据安全。

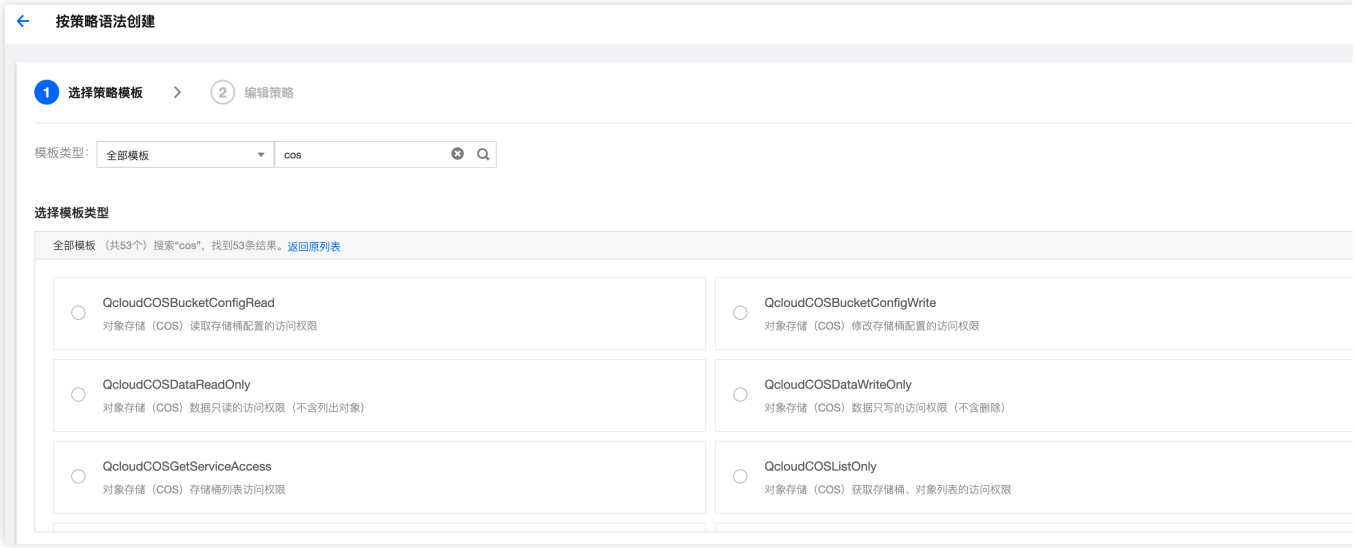
如何配置数据访问策略

步骤一：在访问管理 CAM 创建策略

1. 登录腾讯云账号，进入**访问管理**，登录账号需具备访问管理配置相关权限，建议使用主账号或管理员账号。
2. 单击左侧菜单的**策略**进入策略管理页，单击**新建自定义策略**，选择**按策略语法创建**，进入创建步骤。



3. 在策略模板中搜索 COS，选择**COS 相关权限模板**。

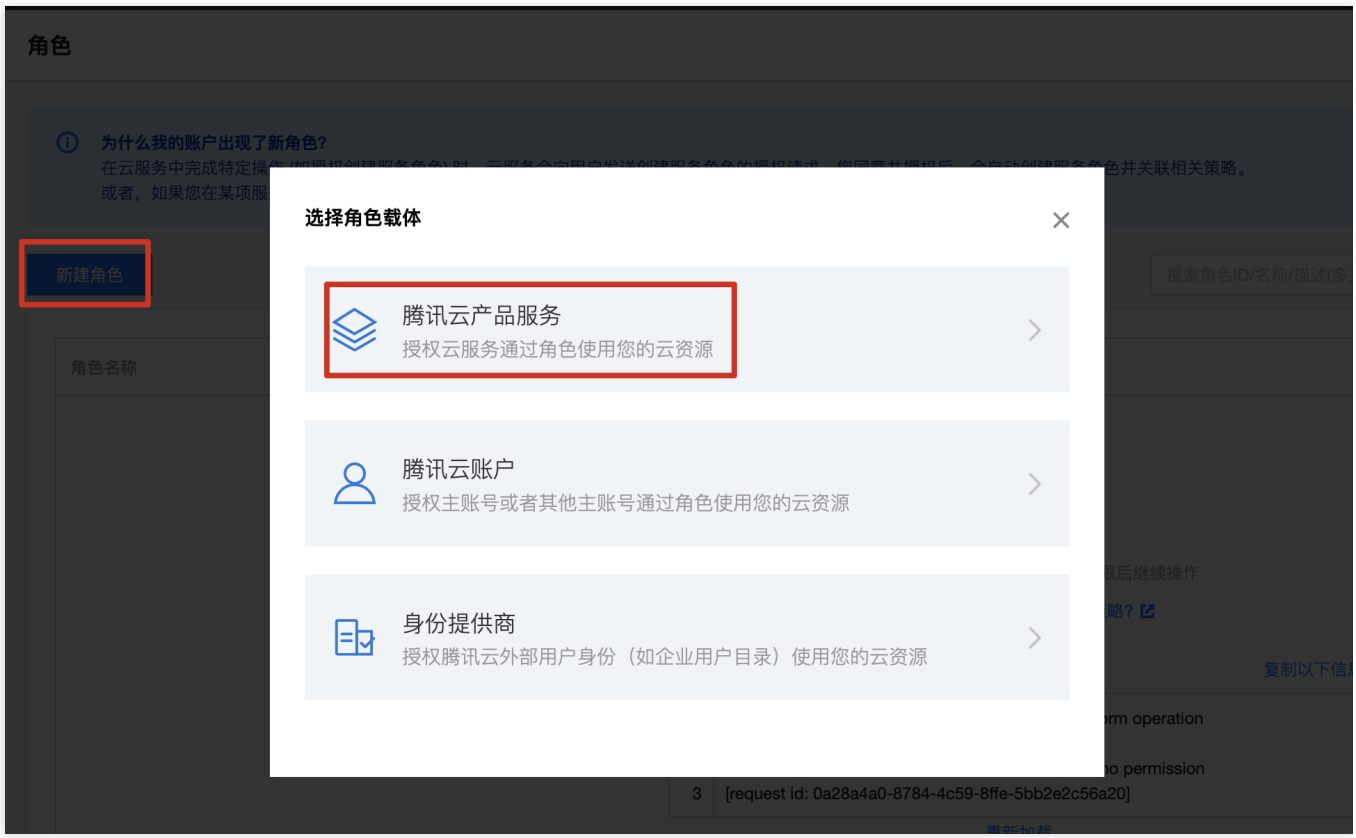


预置模板定义了只读、读写等权限策略场景，如不符合您的使用场景，可通过自定义策略模板进行创建，详细步骤参见 附录。

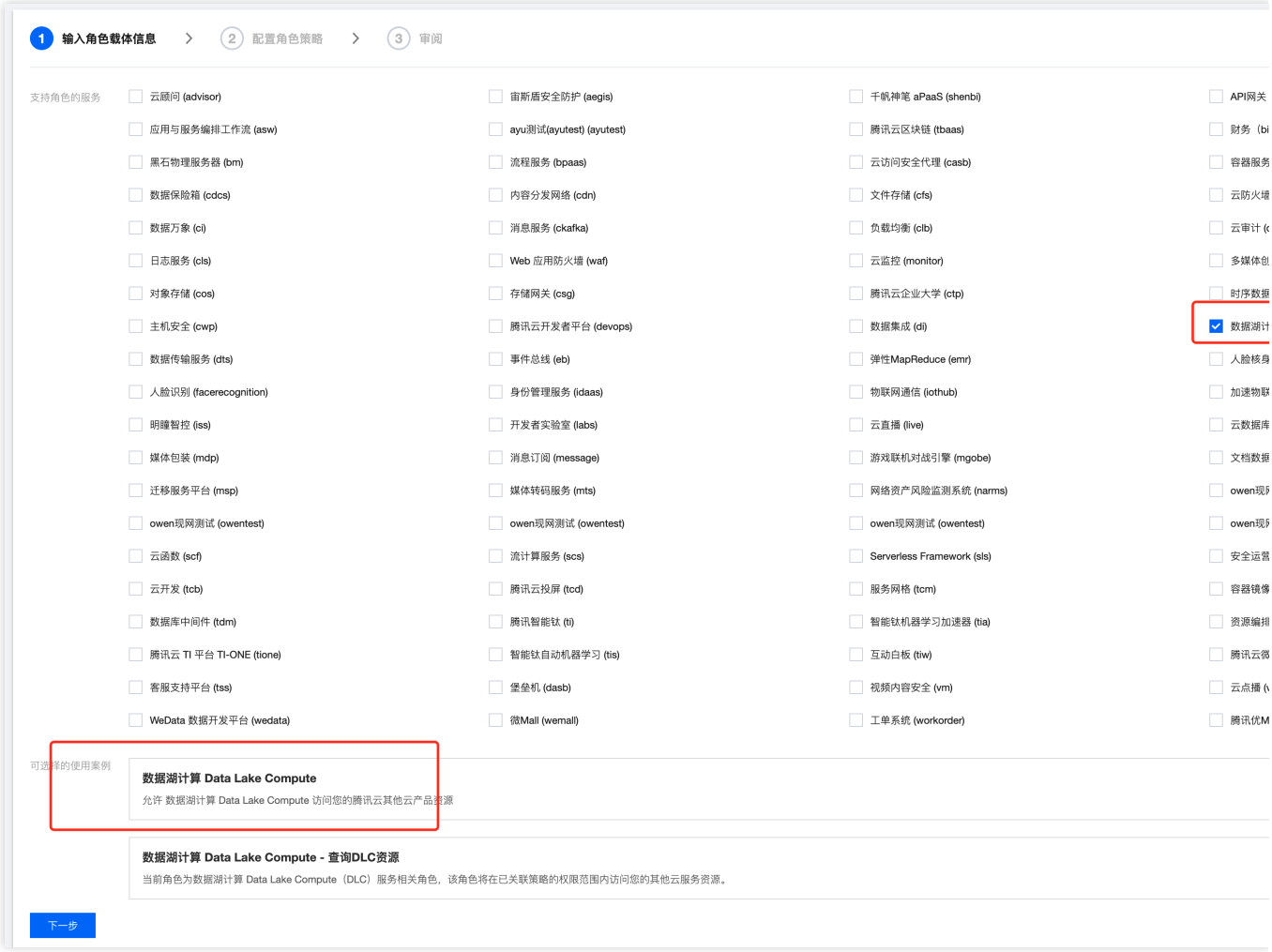
4. 选择模板后，命名策略后保存即可完成创建。

步骤二：创建服务角色

1. 登录腾讯云账号，进入**访问管理**，登录账号需具备访问管理配置相关权限，建议使用主账号或管理员账号。
2. 单击左侧菜单的**角色**进入角色管理页，单击**新建角色**，选择**腾讯云产品服务**后进入角色新建。



3. 在角色载体信息的服务列表中找到**数据湖计算 Data Lake Compete**并选择，然后单击下一步。

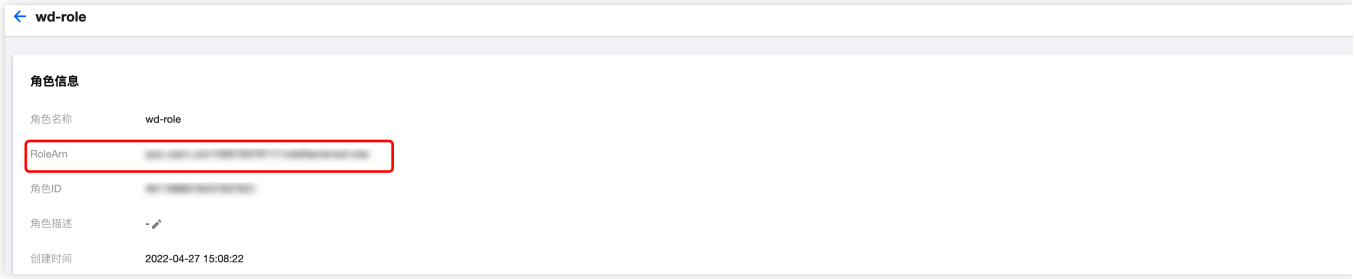


4. 在配置策略中，找到步骤一创建的策略并选择，然后单击下一步。
5. 对角色进行命名后保存，即可完成角色的创建。

步骤三：获取角色的 Role arn 信息

1. 在步骤二完成角色创建后，返回角色列表，找到刚才创建的角色。
2. 单击**角色名称**，进入角色详情页。

3. 找到 Role arn 信息，对其进行复制。



步骤四：将 Role arn 配置到数据湖计算 DLC

1. 登录 数据湖计算 DLC 控制台，登录账号需为 DLC 管理员。
2. 单击左侧菜单的**数据作业**进入数据作业管理页，单击**数据作业配置**并选择**CAM Role arn**。
3. 单击**创建 role arn**按钮，进入创建页面。

4. 将步骤三中获取到的role arn 信息粘贴到输入框内，保存即可完成配置。

附录：自定义策略模板

如预置的模板无法满足您对数据管理的诉求，您可配置自定义模板，配置步骤如下。

1. 登录 腾讯云账号，进入**访问管理**，登录账号需具备访问管理配置相关权限，建议使用主账号或管理员账号。
2. 单击左侧菜单的**策略**进入策略管理页，单击**新建自定义策略**，选择**按策略生成器创建**，进入创建步骤。

3. 效果选择**允许**，服务选择**对象存储（cos）**，资源范围根据需要进行选择。

如您需对特定资源进行管理，可通过单击**添加自定义资源六段式**来进行资源添加，资源支持使用“*” 来指代所有资源，详细说明可参见 资源描述方式 。

4. 配置完成后，对策略进行命名，单击保存即可完成创建。同时您可以选择**将此策略授权给用户**，即可将策略同时授权给已有用户。

创建数据作业

最近更新时间：2024-07-17 17:46:23

前期准备

在开始创建数据作业前，您需先完成数据访问策略配置，以保证数据作业能安全的访问您的数据。配置步骤请参见配置数据访问策略。

创建步骤

1. 登录 数据湖计算 DLC 控制台，单击左侧菜单**数据作业**进入数据作业管理页。
2. 单击**创建作业**按钮，进入创建页。

配置参数如下：

配置参数	说明
作业名称	支持中文、英文、数字和英文下划线，最多40个字符
作业类型	批处理：基于 Spark jar 的批处理数据作业 流处理：基于 Spark Streaming 的流式数据作业
数据源连接	批处理中需要访问的数据源连接选择，目前支持 ckafka，需在作业配置中进行提前配置
数据引擎	可选具有使用权限的 Spark 作业类型的数据引擎 若选择了数据源连接，则仅可选择与数据源连接的数据引擎
程序包	支持 jar 格式。

	支持选择对象存储 cos 内文件或本地上传，本地上传不可超过5MB，如程序包超过5MB可先将程序包上传到 cos 后使用。支持直接填写 cos 存储路径
依赖 jar 资源	支持 jar 格式，可选择多个 支持选择对象存储 cos 内文件或本地上传，本地上传不可超过5MB，如程序包超过5MB可先将程序包上传到 cos 后使用。支持直接填写 cos 存储路径，多个路径以“;”隔开
依赖 file 资源	支持选择对象存储 cos 内文件或本地上传，本地上传不可超过5MB，如程序包超过5MB可先将程序包上传到 cos 后使用。支持直接填写 cos 存储路径，多个路径以“;”隔开
CAM Role arn	在作业配置中配置的数据访问策略，指定了数据作业可访问的数据范围。详细配置说明参见 配置数据访问策略
Main Class	参数　　jar 包的主类参数，多个参数使用空格分割
作业参数	作业-config 信息，spark.开头的参数，按照 k=v 格式填写，多个参数换行填写。 示例：spark.network.timeout=120s
资源配置	可配置数据作业的引擎资源，可配置资源数不超过所选的数据引擎规格。资源说明：1CU≈1核4G 计费 CU 数 = Executor 资源 * Executor 数量 + Driver 资源 按量计费的数据引擎费用将按计算 CU 数的使用量收取

3. 完成配置填写后，保存即可完成创建。

管理数据作业

最近更新时间：2022-09-02 17:34:52

为了方便您更好的管理数据作业，我们提供了以下操作对数据作业进行管理。

- 编辑数据作业。
- 启动&停止数据作业任务。
- 查看数据作业&任务详情。
- 删除数据作业。

编辑数据作业

> 说明：
> 数据作业运行中不支持编辑。

不支持编辑数据作业的类型，如需变更类型可进行数据作业创建。详细步骤请参见 创建数据作业。

1. 登录 数据湖计算 DLC 控制台，选择服务区域，单击左侧菜单**数据作业**进入数据作业管理页。
2. 找到需编辑的数据作业，单击**编辑**按钮进入作业编辑页。

3. 修改需要编辑的内容后保存即可完成编辑。

启动&停止数据作业任务

通过已创建的数据作业，我们可以通过启动&停止操作来执行数据作业生成对应的任务。单个数据作业可以产生多个任务实例或被执行多次。
数据任务状态如下：

状态	说明
未启动	创建后的初始态
运行中	数据任务运行中，此时不可编辑、删除数据作业
成功	任务运行成功

状态	说明
失败	任务运行失败，可通过日志或 sparkui 进行错误信息查询
已取消	用户手动取消任务

启动&停止数据作业任务操作步骤如下：

1. 登录 数据湖计算 DLC 控制台，选择服务区域，单击左侧菜单**数据作业**进入数据作业管理页。
2. 找到需变更状态的数据作业，单击**启动**或**停止**按钮即可变更任务状态。

> 注意：
> 启动任务实例将消耗计算引擎资源，若资源超过了计算引擎配置的上限，任务将会产生排队。

查看数据作业&任务详情

1. 登录 数据湖计算 DLC 控制台，选择服务区域，单击左侧菜**数据作业**进入数据作业管理页。
2. 单击**作业名称**即可进入数据作业详情页。

详情页内可以查看数据作业的基本信息及任务列表。任务列表内包含了该数据作业对应的数据任务信息，支持查看任务的运行日志及 sparkui。

单击查看详情或任务 ID 即可进入任务详情，任务详情内包含了任务的基本信息及运行日志。运行日志目前仅支持查看最近的1000条数据。

单击**创建下载任务**即可对日志进行全量下载，可单击**日志下载**将日志保存到本地。

> 注意：
> 下载记录将会保存3天，超过3天将无法保存找本地，需重新创建下载任务。

删除数据作业

说明：
数据作业存在运行中的数据任务时，将无法被删除。

1. 登录 数据湖计算 DLC 控制台，，选择服务区域，单击左侧菜单**数据作业**进入数据作业管理页。
2. 找到需删除的数据作业，单击**删除**按钮，二次确认后即可完成删除。

注意：
删除数据作业将同时删除对应的数据任务信息，请谨慎操作。

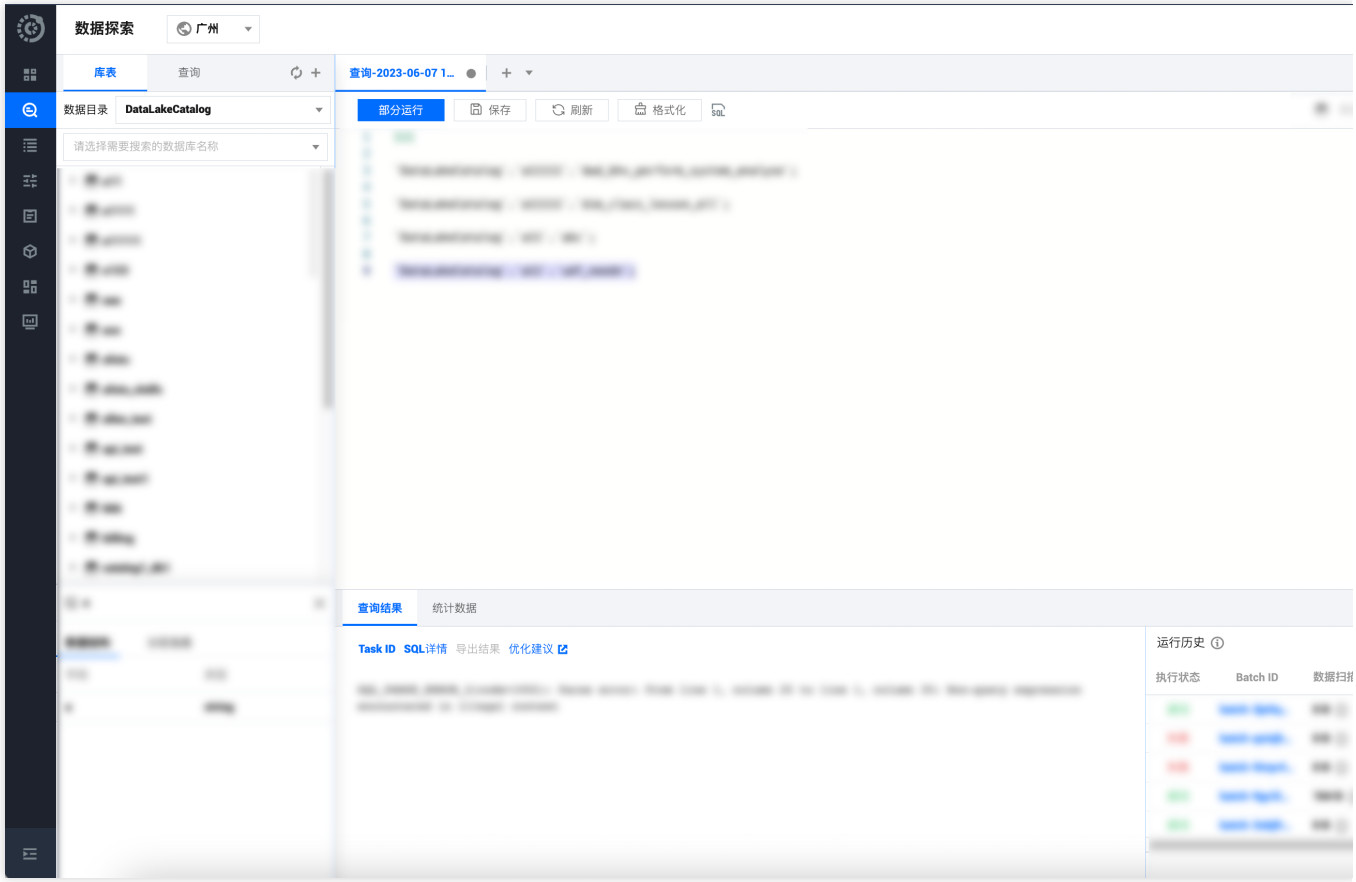
历史记录

最近更新时间：2024-07-31 17:29:42

为了方便用户查询历史任务记录，数据湖计算 DLC 提供了3种途径对历史任务进行查询处理。

查询编辑器中查看运行历史任务

1. 登录 数据湖计算 DLC 控制台，选择服务地域。
2. 进入**数据探索页面**，在单个 Session 内单击**运行历史**即可查看该 Session 下的任务运行历史记录。
3. 单击历史记录 **Batch ID** 即可在左侧查看对应的运行结果



每个 Session 的运行历史独立，最多保存45天的运行历史记录。
历史任务结果数据保存24小时，如需查看超过24小时的任务结果，需重新运行任务。

数据管理功能中查看数据导入历史

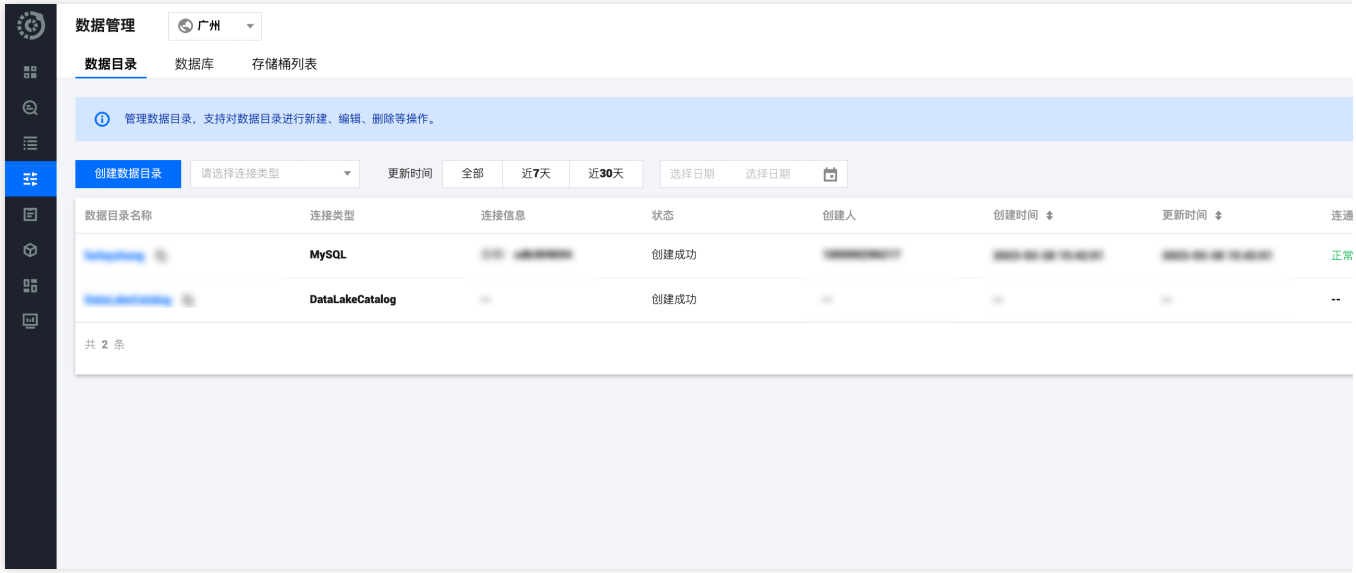
1. 登录 数据湖计算 DLC 控制台 > 数据管理，选择服务地域。

注意：

登录账号需要有数据库相关权限。

2. 单击右上角的**任务历史**即可查询数据导入历史任务。

3. 支持查看近45天的历史任务

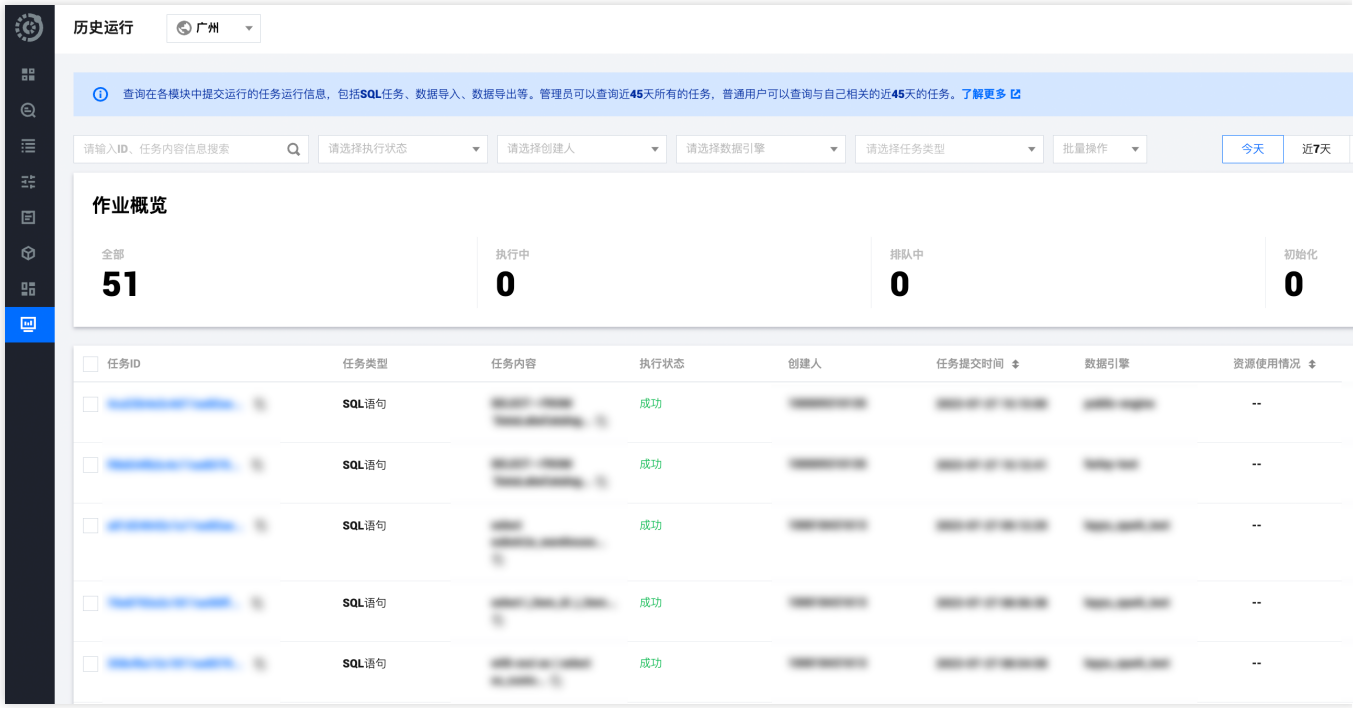


历史运行功能中查看历史任务

1. 登录 数据湖计算 DLC 控制台 > 历史运行，选择服务地域。

2. 进入历史运行页面，管理员可查看近45天所有的历史运行任务，普通用户可查询与自己相关的近45天的任务。

3. 支持对任务类型、执行状态、创建人、数据类型等方式的筛选查看。



4. 单击**查看详情**，可查看任务运行详情及任务运行结果。

历史任务结果数据保存24小时，如需查看超过24小时的任务结果，需重新运行任务，可直接**复制语句**至数据探索进行任务运行。

可直接单击**任务 ID** 快速切换查看任务运行详情。

运行中的任务，可对任务进行**取消**操作。

引擎管理
SuperSQL 引擎
SuperSQL 引擎说明

最近更新时间：2024-07-31 17:51:53

DLC 的数据引擎是 DLC 的数据分析计算服务的基础，用户在 DLC 进行的所有计算都需要使用数据引擎。根据用户的使用场景不同，可选择共享引擎或独享引擎。

共享引擎

共享引擎（public-engine）是 DLC 服务开通后自带的数据引擎，适合分析频率低、计算数据量较小的场景使用。用户无需配置、管理资源，按任务扫描量计费（具体资费参见 计费概述 ），不运行则不计费，具有高灵活、高可用的特点。
DLC 为 Serverless 架构，在一段时间内首次执行任务需要调度数据引擎，等待时间可能稍长。

独享引擎

独享引擎是用户付费购买的专属于自己的数据引擎，资源使用按量计费，具体资费参见 计费概述。
按量计费：适合分析数据具有周期性，需根据业务峰谷进行弹性伸缩的用户，具有高灵活、高稳定对的特点，按 CU 使用量付费。
包年包月：适合长期大量稳定的数据分析需求，可根据业务峰谷进行弹性伸缩，无需等待资源拉起，随时可用，按集群规格按月付费，弹性扩容集群按 CU 使用量计费。

计算引擎类型

根据不同的使用场景，独享引擎可以选择不同的计算引擎来应对不同场景。
SparkSQL：适用于稳定高效的离线 SQL 任务。
Spark 作业：适用于 Spark 原生的流式/批式数据作业处理。
Presto：适用于敏捷、快速的交互式查询分析。
注意：
不同计算引擎类型不影响独享引擎计费单价。

引擎扩缩容规则

基本规则：当配置数据引擎的最大集群数＞最小集群数时，该数据引擎才会产生扩缩容。
扩容规则：当数据引擎目前存在的排队任务＞空闲的并发容量，且没有正在初始化的集群时，系统将会按照配置规则对数据引擎进行扩容。
缩容规则：当数据引擎目前的集群数＞最小集群数，系统将会把处于闲置状态一定时间的集群进行销毁处理。
注意：
数据引擎的集群数量不会少于配置的最小集群数，按量付费集群若无需使用，可对集群进行挂起操作。

引擎运行状态

集群根据当前运行情况，分为启动中、运行、暂停、暂停中、变配中、隔离、隔离中、恢复中八个状态。
启动中：该集群资源正在被拉起，按量计费的独享引擎此时不计费。启动中的集群无法被数据计算选中使用。
运行：该集群正在运行，可被数据计算选中使用。
暂停：该集群暂停使用，无法被数据计算选中使用。
暂停中：该集群正在切换为暂停状态，会影响正在运行的任务，无法被数据计算选中使用。
变配中：该集群正在进行配置变更，配置变更期间将无法被数据计算使用。
隔离：由于账号欠费导致的集群被隔离，无法被数据计算选中使用。
隔离中：由于账号欠费导致，集群正在切换为隔离状态，会影响正在运行的任务，无法被数据计算选中使用。
恢复中：账号通过充值不再欠费后，集群由隔离状态恢复到运行状态的过程，无法被数据计算选中使用。

版权所有：腾讯云计算（北京）有限责任公司

购买独享数据引擎

最近更新时间：2024-07-17 17:58:46

数据湖计算 DLC 的独享数据引擎目前支持按量计费及包年包月两种付费模式，详细价格参见：计费概述。

购买独享引擎

您可通过官网购买页搜索数据湖计算 DLC 或在控制台进行购买，控制台购买流程如下：

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户需具有腾讯云管理员或财务权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 单击左上角**创建资源**按钮，进入资源配置购买页面，根据需求完成配置后即可查看预估价格。

4. 确认价格后即可进行购买。

购买配置参数说明：

地域：处于不同地域的云产品内网不互通，购买后不能更换，请您谨慎选择。建议选择最靠近您客户的地域，可降低访问时延。

计算引擎：支持 Presto、Spark 两种引擎，购买后不能更换，请谨慎选择。Presto 引擎适用于更快的交互查询分析，且支持多频联合查询。Spark 引擎适用于大数据量的离线任务，任务运行更加稳定。

集群规格：集群规格以 CU 为单位，1CU 约等同于1核4GB 内存的计算资源。规格大小决定了任务执行时的算力资源数量，可根据您的实际情况进行购买。

注意：

如您需购买超过152CUs 的资源，请通过工单联系我们提前备货，以免购买时存货不足影响您的使用。

最小集群数：集群启动时的最小集群个数，购买包年包月集群的常驻资源数，多个集群可提供更高的并发能力。

最大集群数：弹性扩容的最大集群个数，若设置等于最小集群数，则代表该集群不进行弹性扩容。

自动启动：若开启，当数据引擎挂起时，收到一次任务请求后将自动启动。

注意：

由于按量计费资源非预留资源，可能存在无法拉起的情况。若您需要常驻稳定运行的计算资源可选择购买包年包月的数据引擎。

挂起策略：支持配置按量计费的数据引擎的挂起方式，支持自动挂起和定时挂起两种策略。按量计费的数据引擎挂起后将不会产生任何费用。

自动挂起-将在数据引擎无任务后一段时间自动将数据引擎切换为暂停状态。

定时策略-可配置按周的定时启动、挂起策略，系统将根据配置规则对集群定期定时进行启动及挂起。

结束任务后挂起：到达指定时间后，若数据引擎有正在执行的任务，数据引擎将会在任务执行完毕之后5分钟内自动挂起。

自动暂停后挂起：到达指定时间后，若数据引擎有正在执行的任务，系统会将任务暂停然后立即挂起数据引擎。

高级配置：若您需要使用联邦查询，可在高级配置进行网段信息配置。

版权所有：腾讯云计算（北京）有限责任公司

标签：支持对购买资源进行标签分类及分账，详细可参见：为独享引擎资源关联标签 文档。

查询账单

可通过数据湖计算 DLC 控制台对您在产品内的花费账单进行查询，操作步骤如下：

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户需具有腾讯云管理员或财务权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 单击**订单查询**按钮即可查看详细订单及扣费情况（需财务权限）。

续费管理

针对包年包月的独享数据引擎，您可通过数据湖计算 DLC 控制台的续费管理入口进入您的资源管理页面进行续费等相关操作，操作步骤如下：

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户需具有腾讯云管理员或财务权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 单击**续费管理**按钮即可进入资源列表进行续费管理操作（需财务权限）。

版权所有：腾讯云计算（北京）有限责任公司

数据引擎
广州
数据引擎包括独享数据引擎与共享数据引擎。共享数据引擎public_engine由系统管理，按扫描量计费；独享引擎支持自动弹性伸缩、自动启动与挂起，按CU时计费。更多权限操作指引，请参考 数据引擎操作指引，引擎计费信息 请参考 计费指引
创建资源
订单查询
续费管理
选择资源标签进行筛选
请输入关键字搜索
资源名称/ID
数据引擎
运行状态
付费类型
集群规模
创建时间
创建人
描述
操作
presto
运行
按扫描量计费
--
2022-05-07 16:52:30
共享集群
配置 更多
共 1 条
10 条 / 页
1
/ 1 页

SuperSQL 引擎续费

最近更新时间：2024-07-31 17:55:02

您可在数据湖计算 DLC 控制台对未到期或隔离中的包年包月数据引擎进行续费操作。

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有腾讯云管理员或财务权限。
2. 通过左侧导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 找到需要续费的数据引擎，单击**更多**然后选择“续费”，临近到期时间（7天内）的资源也可直接点击到期时间旁的**续费**按钮进行续费。

4. 确认续费时长信息及价格后，单击**确认续费**即可提交续费订单，确认订单并扣费成功即可完成续费。

注意：

隔离中的数据引擎被续费恢复使用后，续费周期的起始时间为上一个周期的到期日。

管理独享数据引擎

最近更新时间：2024-07-17 18:05:00

说明：

共享数据引擎有 DLC 统一管理，无需用户进行管理。

变更独享引擎配置

注意：

独享引擎配置变更可能导致资费变更，详细资费变化参见 调整配置费用说明。

途径一：数据引擎列表

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有腾讯云管理员或财务权限。
2. 导航菜单栏单击**数据引擎**，进入数据引擎管理页面。
3. 在数据引擎中找到需要修改的独享引擎，单击右侧**配置**按钮进入配置修改页面，可修改集群规格、弹性策略。
4. 根据需要调整参数后，单击保存提交订单支付后即可发起配置变更。

途径二：数据引擎详情

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有腾讯云管理员或财务权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 在数据引擎中找到需要修改的独享引擎，单击集群名称进入集群详情页面，可修改集群规格、弹性策略。
4. 根据需要，调整参数后，单击保存即可完成配置修改。

修改独享引擎信息

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有腾讯云管理员权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。

版权所有：腾讯云计算（北京）有限责任公司

3. 在数据引擎中找到需要修改的独享引擎，单击集群名称进入集群详情页，**可修改集群描述、自动启动、挂起策略**。
4. 根据需要，调整参数后，单击保存即可完成信息修改。

挂起策略：支持配置按量计费的数据引擎的挂起方式，支持自动挂起和定时挂起两种策略。按量计费的数据引擎挂起后将不会产生任何费用。
自动挂起：将在数据引擎无任务后15分钟自动将数据引擎切换为暂停状态。
定时策略：可配置按周的定时启动、挂起策略，系统将根据配置规则对集群定期定时进行启动及挂起。
结束任务后挂起：到达指定时间后，若数据引擎有正在执行的任务，数据引擎将会在任务执行完毕之后5分钟内自动挂起。
自动暂停后挂起：到达指定时间后，若数据引擎有正在执行的任务，系统会将任务暂停然后立即挂起数据引擎。

手动挂起&启动独享引擎

注意：
包年包月资源为常驻资源，不支持进行挂起操作。
1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有管理员权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 在数据引擎中找到需要暂停/启动的独享引擎，鼠标悬停至**更多**，在下拉菜单中单击启动/挂起按钮。

独享引擎销毁

如您无需再使用数据引擎，可对数据引擎进行销毁操作。包年包月的数据引擎销毁将产生自助退还操作，相关说明可参见 退费说明。

注意：

按量付费的数据引擎销毁后将无法恢复，请谨慎操作。

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有腾讯云管理员权限。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 在数据引擎中找到需要销毁的独享引擎（仅状态为“暂停”的集群可被销毁），鼠标悬停至**更多**，在下拉菜单中单击销毁按钮。
4. 二次确认后，独享引销毁成功。

集群运行日志

用户购买的独享引擎，DLC 提供14天内的集群运行日志，帮助用户了解集群的启动、挂起、扩缩容情况。集群日志主要包括一下内容：

启动时间：集群开始工作的时间。

挂起时间：集群停止工作的时间。

扩容记录：集群扩容时间及扩容数量。

缩容记录：集群缩容时间及缩容数量。

查看集群运行日志

1. 登录 数据湖计算 DLC 控制台，选择服务所在区域，登录的用户须具有数据引擎监控权限（详情请参见 权限文档）。
2. 导航菜单单击**数据引擎**，进入数据引擎管理页面。
3. 在数据引擎中找到需要查看日志的独享引擎，单击**集群名称**进入详情页查看。

容灾集群

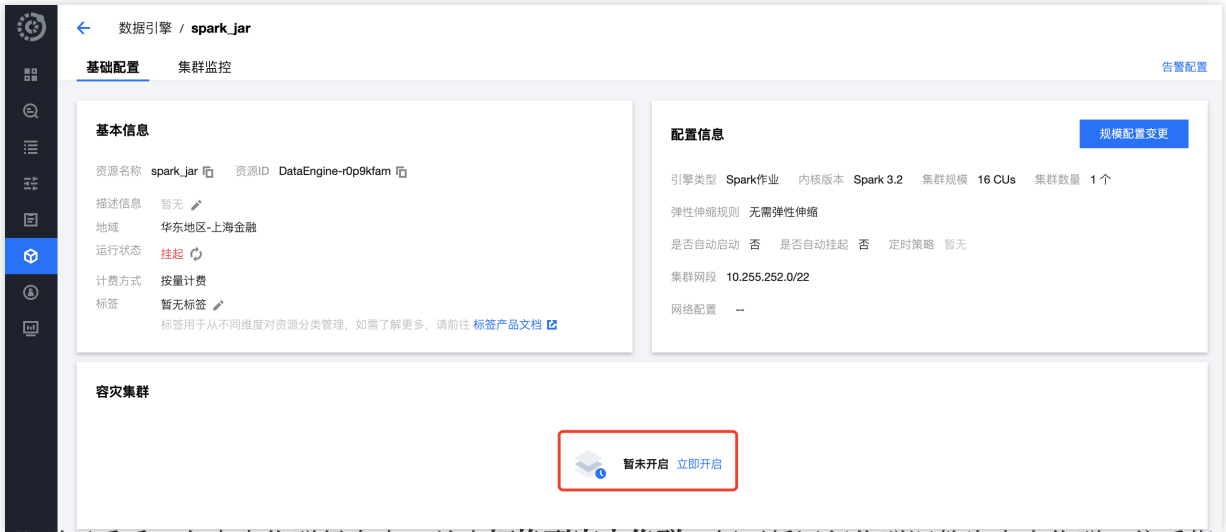
最近更新时间：2024-07-31 17:47:16

为保证极端场景下计算引擎稳定运行，DLC 提供高效敏捷的容灾集群能力。当您需要容灾集群时，可以快速切换到容灾集群，保障服务正常运行。容灾集群只在运行时收费，详情请见费用说明。

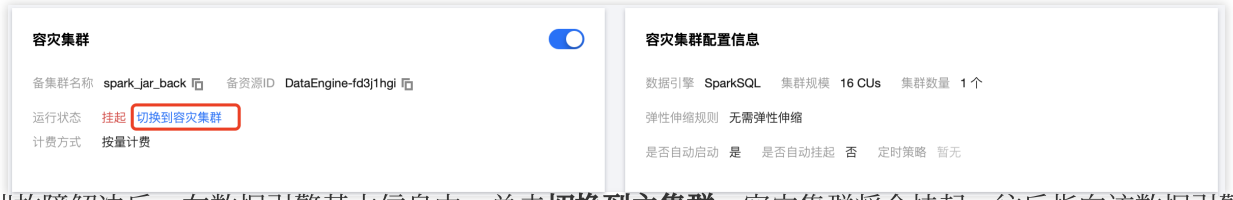
操作步骤

1. 进入 DLC 控制台，单击数据引擎进入数据引擎页面。
2. 单击数据引擎资源名称，进入数据引擎详情页面。

3. 单击**开启容灾集群**，等待容灾集群初始化。



4. 容灾集群开启后，在容灾集群信息中，单击**切换到容灾集群**，便可将运行集群调整为容灾集群，往后指向该数据引擎的作业都会提交到容灾集群。容灾集群用以数据引擎极端故障时过渡。



5. 极端故障解决后，在数据引擎基本信息中，单击**切换到主集群**，容灾集群将会挂起，往后指向该数据引擎的作业都会提交到主集群。

容灾集群规格

容灾集群总是尽可能地与数据引擎本身的规格保持一致，以保证原有任务可以正常过渡运行。当数据引擎本身开启了弹性伸缩，容灾集群的弹性伸缩规则会与数据引擎保持一致。同时为了节约费用，容灾集群总是按量计费模式的。

费用说明

容灾集群开启不收费。当切换到容灾集群时且容灾集群运行时，会按照数据引擎同规格下按量计费的标准进行收费。

示例：

1. 当数据引擎本身为包年包月的16 CU SparkSQL引擎时。开启容灾集群后，容灾集群为按量计费的16 CU SparkSQL引擎，此时容灾集群挂起不收费。当用户切换到容灾集群后，容灾集群运行，将额外收取容灾集群使用 CU 时长的费用。具体费用可参考计费概述。
2. 当数据引擎本身为按量计费的16 CU SparkSQL 引擎时。开启容灾集群后，容灾集群为按量计费的16 CU SparkSQL 引擎，此时容灾集群挂起不收费。当用户切换到容灾集群后，容灾集群运行，主集群挂起，只收取容灾集群使用 CU 时长的费用。

版权所有：腾讯云计算（北京）有限责任公司

引擎内核版本

最近更新时间：2024-07-31 17:47:43

DLC 根据不同应用场景提供了不同的内核版本，并针对内核做了大量的特性和性能优化，内核版本一览如下所示。
如果您的场景主要为交互式查询，推荐使用 Presto 引擎和 SparkSQL 引擎和最新的内核版本。
如果您的场景主要使用批作业，推荐使用 Spark 作业引擎和 Spark3.2内核版本。

引擎类型	内核版本	说明
Presto	SuperSQL-P 1.0	基于原生 Presto 0.242 版本实现，在此基础上支持动态加载数据源、Dymanic Filter 增强、 Iceberg V2 表、非分区表的 INSERT OVERWRITE、支持执行 Hive UDF。
SparkSQL	SuperSQL-S 1.0	基于原生 Spark3.2版本实现，在此基础上支持 Iceberg 1.1.0、Hudi 0.12.0、自适应 Shuffle Manager。
Spark 作业	Spark 3.2	基于原生 Spark3.2版本实现，在此基础上支持 Iceberg 1.1.0、Hudi 0.12.0、Python3、支持自适应 Shuffle Manager。
	Spark 2.4	基于原生 Spark2.4版本实现，在此基础上支持 Iceberg 0.13.1、Python2、Python3。

版权所有：腾讯云计算（北京）有限责任公司

引擎网络配置

最近更新时间：2024-07-31 17:47:56

数据湖计算 DLC 支持对数据引擎配置网络（VPC），方便管理数据引擎访问不同的数据源网络。

网络配置类型

根据不同的业务场景，数据湖计算提供了两种网络配置类型。
增强型网络配置：适用于需高速稳定访问单个 VPC 下的数据。
注意
非 Spark 作业类型的数据引擎仅可绑定一个增强型网络配置。
跨源型网络配置：适用于需访问多个VPC的数据跨源联邦查询，一个数据引擎支持绑定多个跨源型网络配置。

网络配置状态

初始中：该网络配置正在初始化，此时网络未生效。
成功：网络配置对绑定的引擎生效。
失败：网络配置失败，可删除后重新配置。

网络配置安全策略

若您为 VPC 配置了安全组策略，则需对不同的网络配置类型添加入站规则。
增强型网络：需要在安全组添加数据源所在 VPC 的网段入站规则。
跨源型网络：需要在安全组添加网络配置绑定的引擎所在网段入站规则。

创建网络配置

1. 登录 数据湖计算 DLC 控制台，选择服务地域。
2. 通过左侧导航菜单进入**数据引擎 > 网络配置**。
3. 单击**创建网络配置**按钮，进入创建配置页。

配置参数如下：

配置内容	是否必填	填写说明
网络配置类型	是	根据应用场景进行选择： 增强型网络配置：适用于需高速稳定访问一个 VPC 下的数据场景 跨源型网络配置：适用于需访问多个 VPC 下数据的跨源联邦查询分析场景
配置名称	是	支持中文、英文和_，字符数不超过35个
实例来源	是	支持两种来源： DLC 的数据目录：可选当前已在数据湖计算 DLC 的数据管理内创建连接的数据目录 新的网络配置：选择一个新的数据源来创建网络连接，目前数据源支持 MySQL、Kafka、EMR HDFS（COS、HDFS、Chdfs）、Postgresql、SqlServer、

		Clickhouse。若需要创建的网络配置关联的数据源暂未支持，可选其他，手动指定VPC
数据目录	是	根据所选的实例来源，选择对应的数据目录。可选数据目录范围将与您的账号权限有关
绑定数据引擎	是	选择该网络配置关联的数据引擎，若数据引擎处于隔离或启动中状态时，将无法选择
配置描述	否	不超过100个字符

4. 填写完成保存即可创建网络配置。

注意

创建后网络为初始化状态，后续可在列表查看状态。

删除网络配置

您可通过删除操作对无需使用或配置失败的网络配置进行管理删除，操作步骤如下：

1. 数据湖计算 DLC 控制台，选择服务地域。
2. 通过左侧导航菜单进入**数据引擎 > 网络配置**。
3. 找到需要删除的网络配置，支持进行筛选查找，注意选择网络配置类型。
4. 单击**删除**按钮，二次确认后即可完成删除操作。

注意

删除后该数据引擎将无法使用该网络配置，如需访问需重新配置，请谨慎操作。

修改描述信息

您可通过修改描述信息对已配置的网络配置进行描述信息修改，操作步骤如下：

1. 数据湖计算 DLC 控制台，选择服务地域。
2. 通过左侧导航菜单进入**数据引擎 > 网络配置**。
3. 找到需要删除的网络配置，支持进行筛选查找，注意选择网络配置类型。
4. 单击**修改描述信息**按钮，即可进行编辑修改。

为独享引擎资源关联标签

最近更新时间：2022-08-16 09:42:00

概述

标签用于对资源进行分类整理。标签由标签键和标签值组成，一个标签键可以对应多个值。用户可以为云资源创建和绑定标签对云资源进行管理。DLC 支持在产品控制台和新购页为独享引擎绑定标签，通过标签对独享引擎资源进行多维度分类管理以及账单细分。

创建标签并绑定资源

通过创建标签并绑定独享引擎来完成资源的分类和统一管理。

操作步骤

1. 登录 标签控制台 创建标签，具体操作参考 创建标签并绑定资源。
2. 登录 DLC 控制台。
3. 在左侧导航栏中，单击**数据引擎**模块，进入数据引擎列表页。
4. 单击资源名称进入资源详情页，单击**编辑**按钮，弹出标签编辑窗口，选择标签进行绑定。

5. 单击**确认**，成功为引擎绑定标签。再次单击编辑，可进行解绑、修改操作。

新购页标签绑定

用户可以在购买独享引擎资源时绑定标签，包年包月、按量付费模式下均支持标签的绑定。

支持按标签进行过滤

支持在 DLC 控制台数据引擎模块按照标签对资源进行过滤。

版权所有：腾讯云计算（北京）有限责任公司

操作步骤

1. 登录 DLC 控制台，进入数据引擎模块。
2. 在标签搜索框内进行选择标签，支持按照标签键和标签键-值两种方式进行过滤，可以根据多个标签值进行筛选。

3. 选择标签后单击搜索按钮，筛选出包含该标签的引擎列表。

按标签分账

您可以基于组织或业务维度为资源规划标签（例如：部门、项目组、地区等）实现成本的分摊管理。

操作步骤

1. 创建标签。登录 标签控制台 创建标签。
2. 绑定资源。在标签控制台、DLC 产品控制台-数据引擎模块或购买页为引擎资源绑定标签。
3. 设置分账标签。前往 费用中心 设置设置分账标签，详见 分账标签。
4. 前往 账单概览 页面，选择按标签汇总页签，查看根据标签键汇总的相关资源的柱状图和列表。

引擎本地缓存

最近更新时间：2024-07-31 17:48:11

为保证在网络带宽受限的情况下（例如触发存储系统的限流），Spark 引擎查询分析能稳定运行，DLC Spark 引擎提供本地缓存（Local Cache）能力。当您需要缓存表数据时，可以通过添加引擎配置快速开启缓存功能。

操作步骤

1. 创建 Spark 引擎，详情请参见 购买独享数据引擎。
2. 添加缓存配置，进入 DLC 控制台 > 数据引擎，选择步骤1创建的引擎，单击**参数配置**，将 缓存配置项说明 中的配置项添加到引擎配置中。

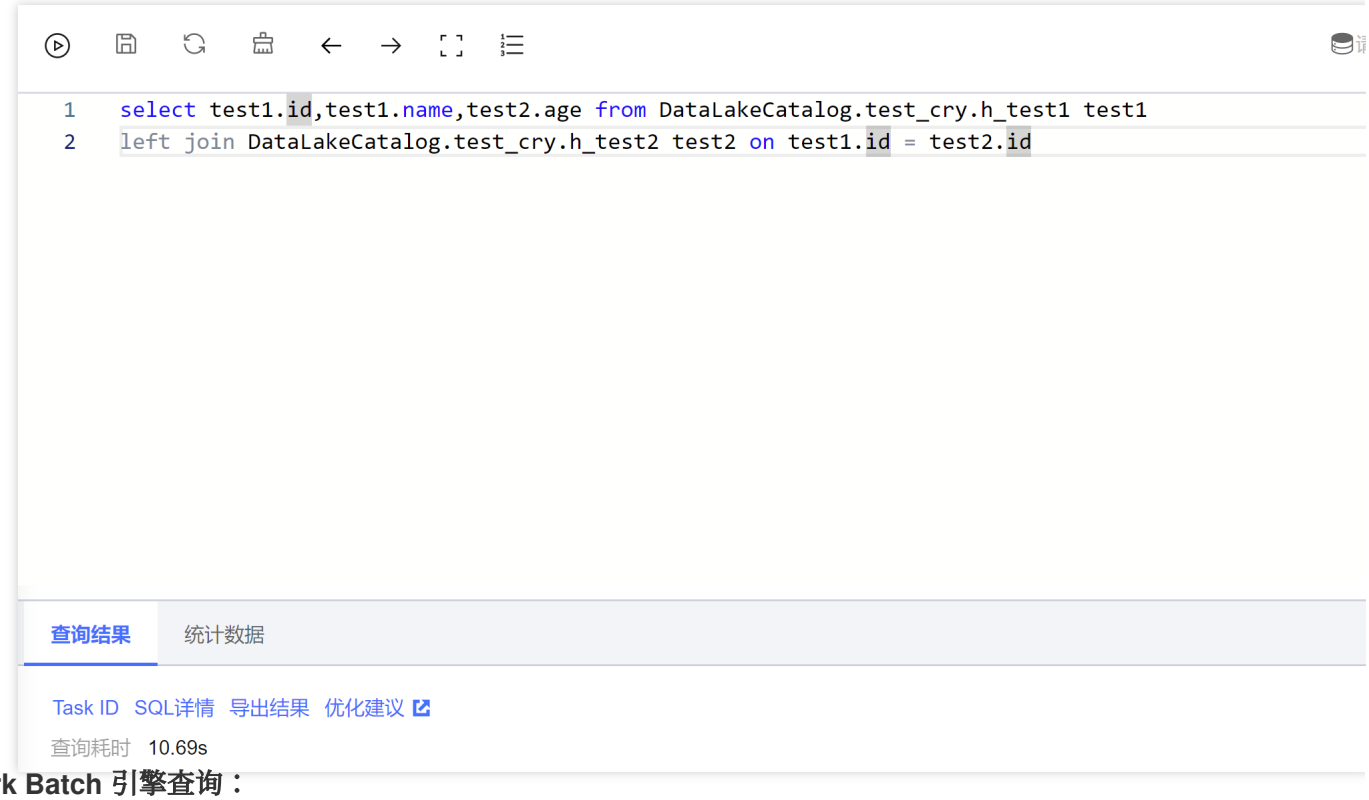
Spark SQL 引擎配置：

说明：

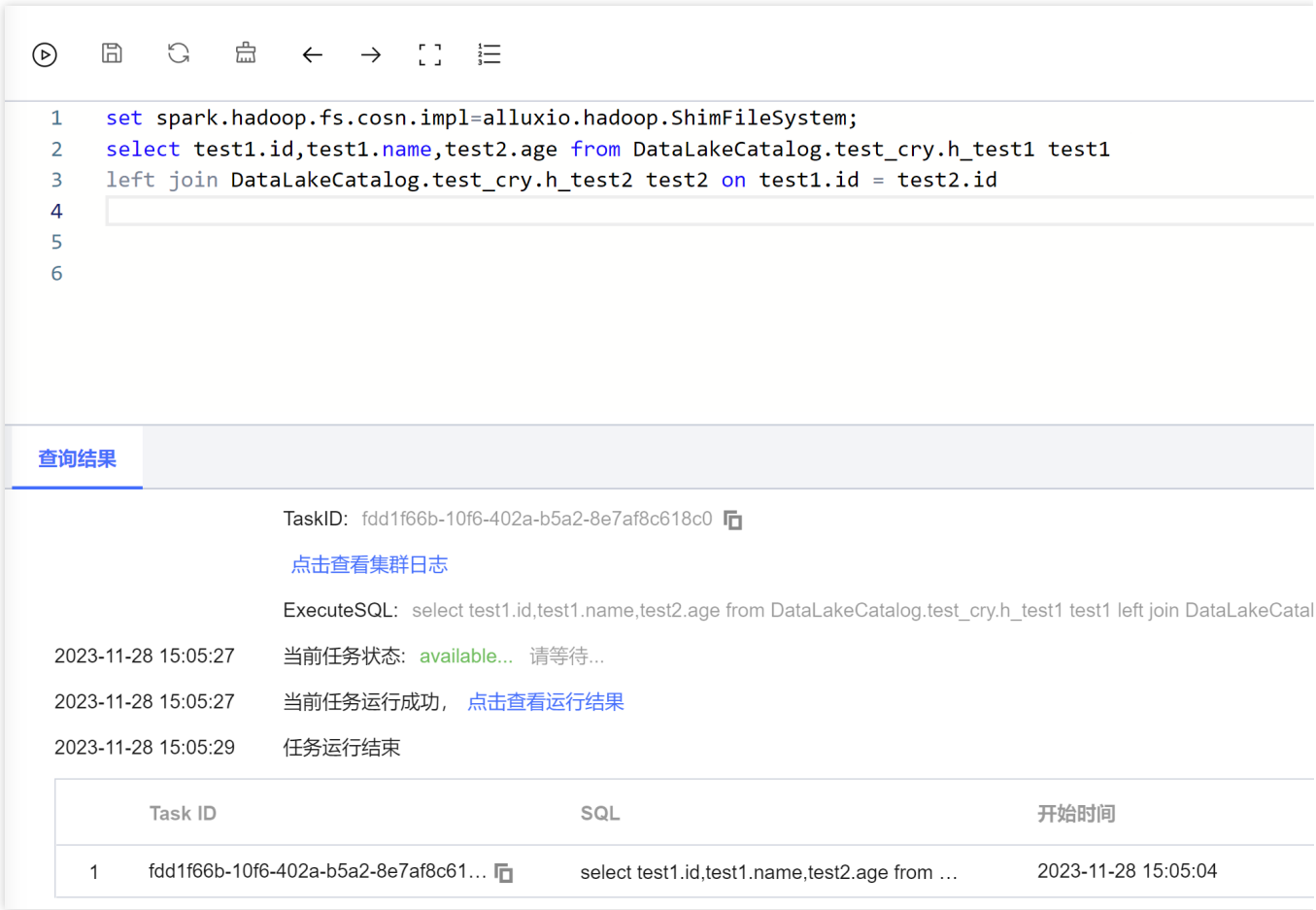
添加配置后，引擎集群会重启，建议在没有跑任务的情况下开启缓存，避免影响正在运行的任务。

3. 使用引擎缓存，进入 数据探索，SQL 界面编写查询 SQL，选择开启缓存的引擎，执行 SQL，执行完毕。引擎会将 SQL 中涉及的 DLC 外表缓存到本地，当再次执行 SQL 时，数据会从本地缓存获取，提升查询效率。

Spark SQL 引擎查询：



Spark Batch 引擎查询：



缓存说明

缓存配置项说明

配置项	配置值	配置项说明
spark.hadoop.fs.cosn.impl	alluxio.hadoop.ShimFileSystem	固定值；配置值为缓存功能实现类。配置该值为开启缓存功能；开启缓存功能时，若配置非该值的其他值，会导致引擎访问不到 COS 数据，请谨遵指引配置。 若开启缓存后需关闭缓存，请删除该配置项。

缓存使用说明

1. 引擎类型说明

   SparkSQL 引擎：当引擎重新启动后，因缓存为本地缓存，故原缓存的数据会失效。

   SparkBatch 引擎：SparkBatch 引擎运行任务为 Session 级别，任务执行完，缓存数据失效。

2. 表类型说明

   目前只缓存 DLC 外表。

版权所有：腾讯云计算（北京）有限责任公司

自定义任务调度池

最近更新时间：2024-07-31 17:48:28

使用场景

适用引擎：Spark SQL 引擎。
当您向引擎提交多个任务时，例如同时提交多个 SQL 任务到 Spark SQL 集群，由于业务提交的任务可能存在前后依赖，所以引擎在调度执行这些任务时，会默认按照 FIFO 的方式来调度这些任务。
但是某些特殊情况，可能您需要自己定义某些任务的优先级，例如以下场景：
提交的任务优先级很高，需要最高优先级执行这个任务，不希望它排队等待集群资源。
提交的任务优先级很低，希望这个任务尽量不抢占其他任务的资源，有资源时就执行，没资源时就排队等待。

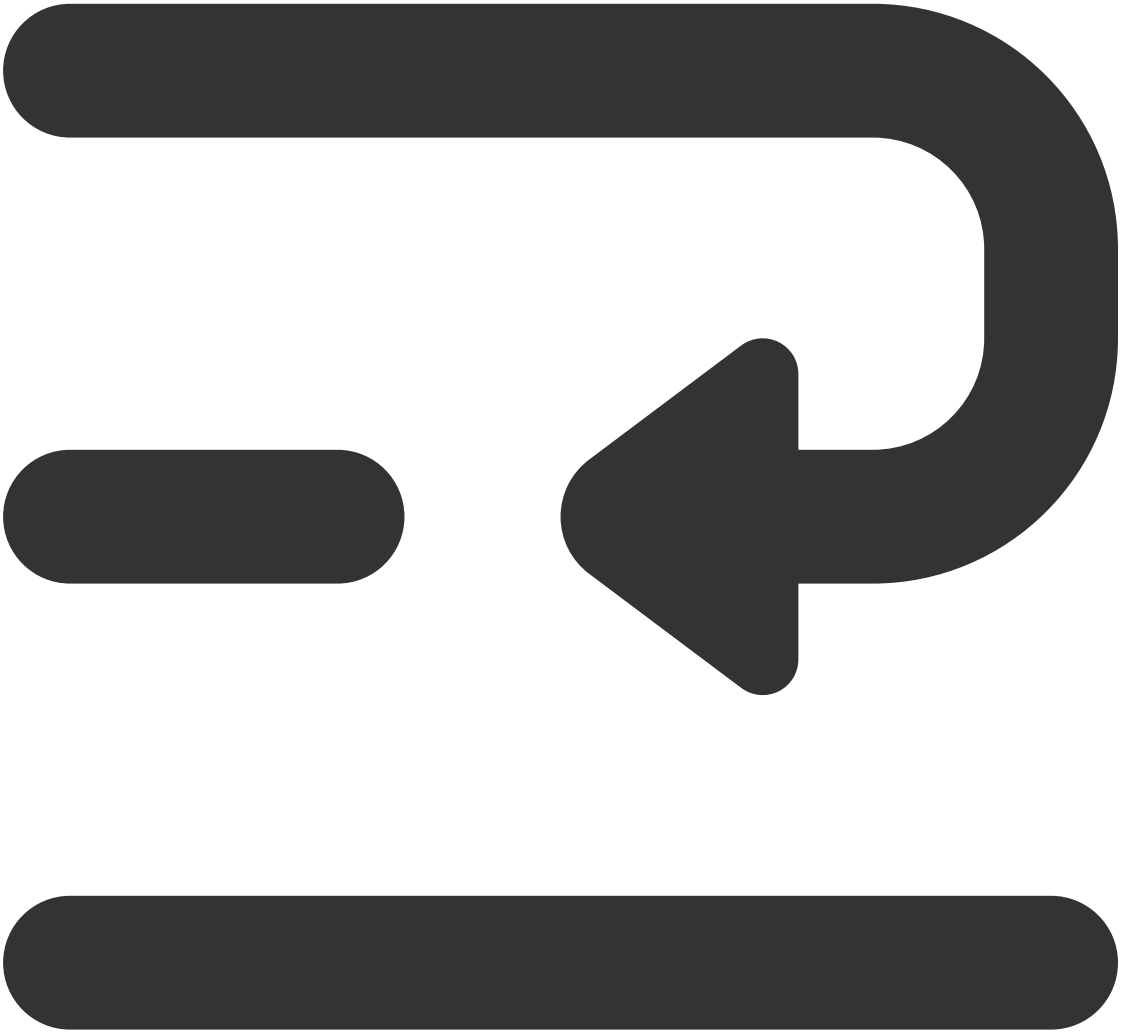
自定义调度规则

在 Spark SQL 引擎中，每个执行的 SQL 任务 Job 都会被切分成多个 Task 的集合 TaskSet，我们的调度是基于 TaskSet 来调度的。每当集群有空闲资源时，就从所有 Job 的 TaskSet 中按调度算法取一个 Task 下发执行。
我们的调度算法就是定义多个调度池，将 Job/TaskSet 放在对应的调度池中，按调度池来获取需要下发执行的 Task。

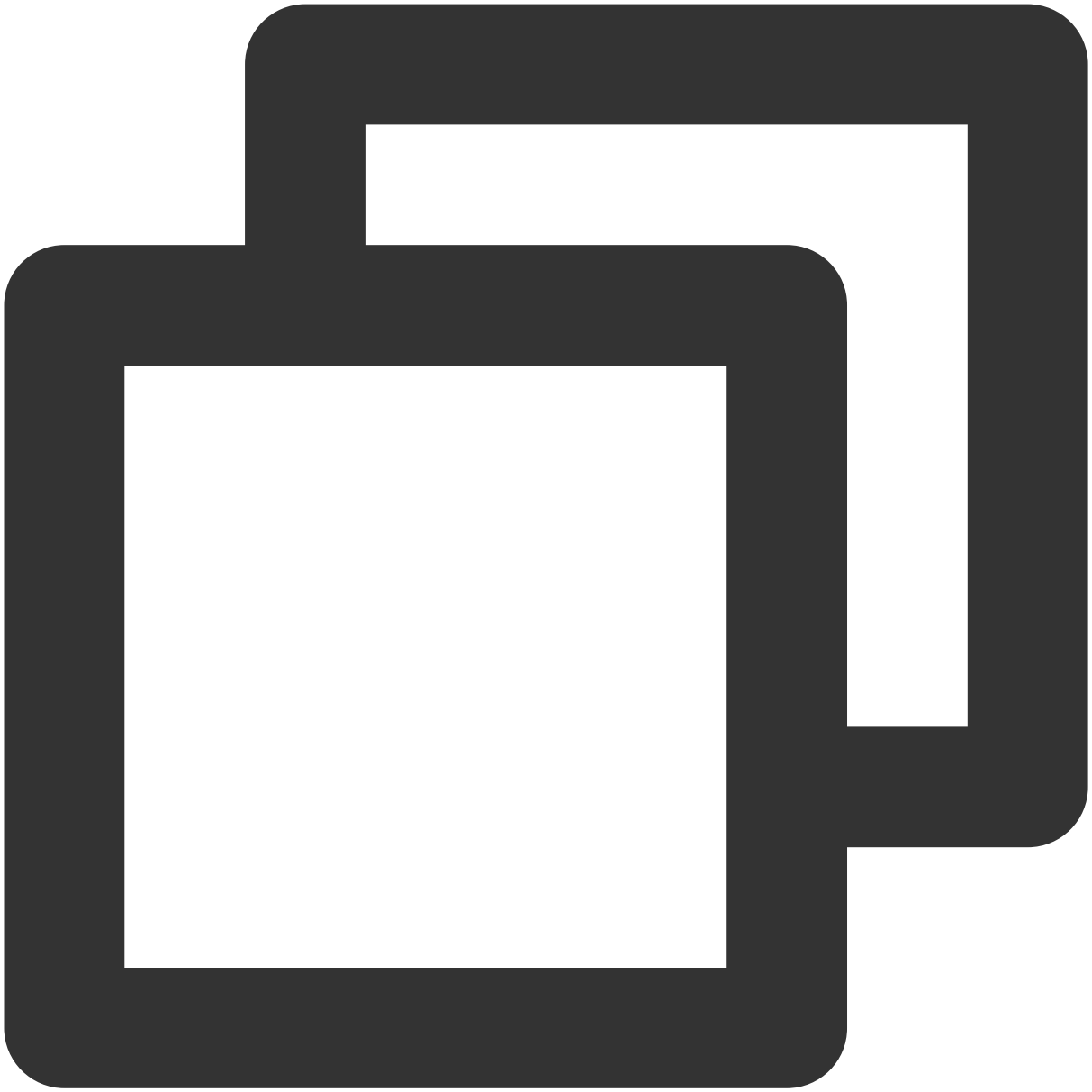
调度池及其属性

您可以定义多个调度池，每个调度池有四个属性：
name：调度池的名称，您可以自己命名。可以命名为 default，表示默认的调度池。
schedulingMode：调度规则，支持 FIFO 和 FAIR 两种模式。一个调度池内有多个 TaskSet 时的调度算法。
FIFO：按照 TaskSet 提交的先后顺序依次下发。
FAIR：多个 TaskSet 的任务都公平地被下发。具体下发规则与调度池的 minShare 和 weight 属性有关。
minShare：最小满足的核数，必须大于0，也就是最小可以运行的 Task 数。在调度时会优先满足该调度池内运行的 Task 数目达到 minShare。
weight：权重。权重高的调度池，Task 会优先被调度。在满足 minShare 后才会对比权重。
调度的配置需要您撰写一个 xml 文件，格式如下：

版权所有：腾讯云计算（北京）有限责任公司



版权所有：腾讯云计算（北京）有限责任公司



```
<?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>
```

调度配置参考示例

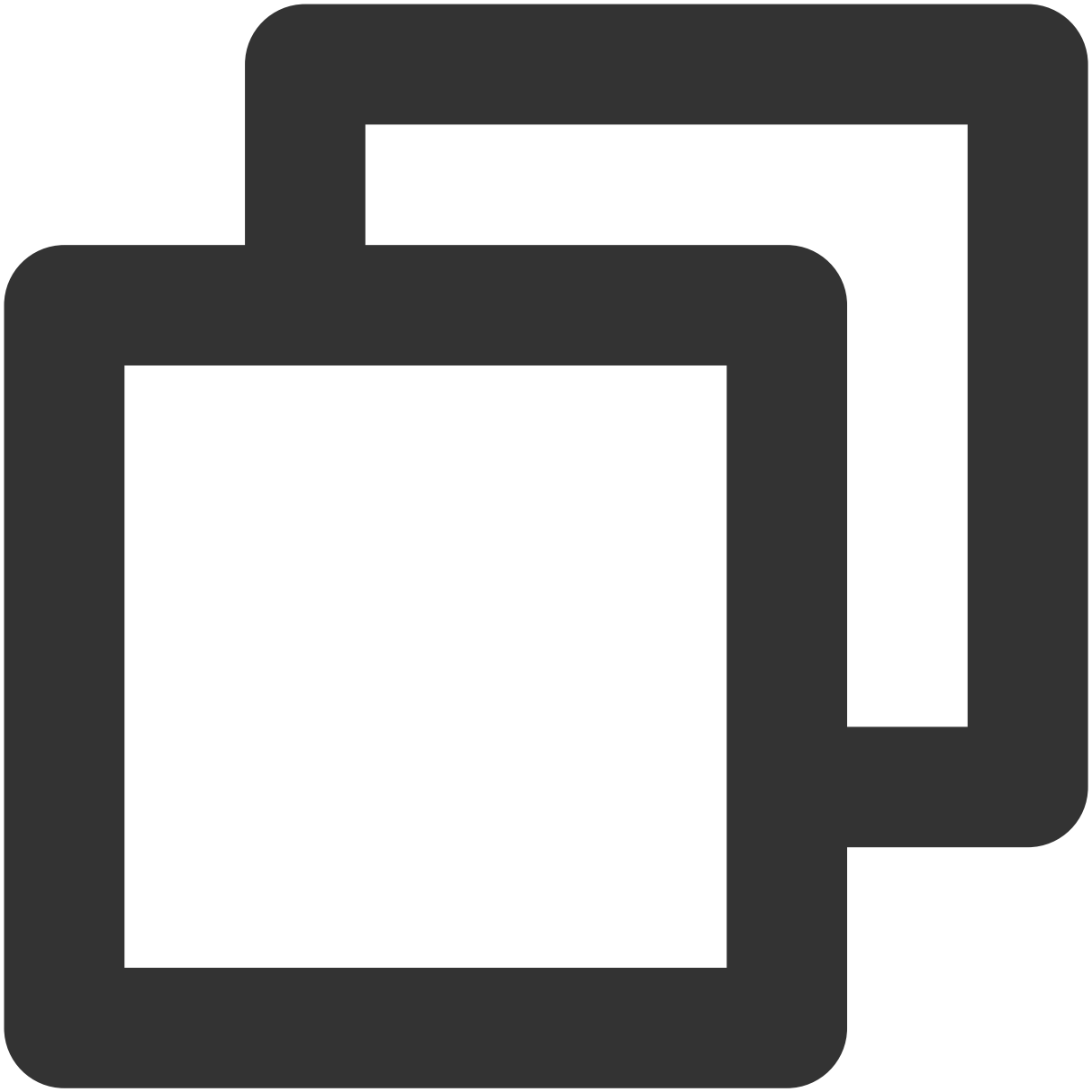
您可以参考设置三个调度池：

默认调度池 default：schedulingMode = FIFO，weight = 1，minShare =（集群核数 - driver 核数）。该调度池是任务默认提交的调度池，优先级普通，按照先后顺序执行，可以使用集群全部的计算资源。

慢任务调度池 straggler：schedulingMode = FAIR，weight = 1，minShare = 1。该调度池专用于慢任务提交，优先级普通。由于minShare = 1，并不会抢占其他提交到 default 调度池的任务的资源。当集群有更多空闲资源时才会执行 straggler 调度池中的 Task。

高优调度池 special：schedulingMode = FIFO，weight = 1000，minShare =（集群核数 - driver 核数）。该调度池用于特殊情况需要优先执行的任务。但是，由于 minShare 的存在，该调度池并不会抢占集群全部资源，default 调度池中的任务仍然也会继续执行，一般情况下 default 池和 special 池各发下同等数量的 Task。

以16CU的集群（driver 为4CU）为例，这种参考示例的配置如下：

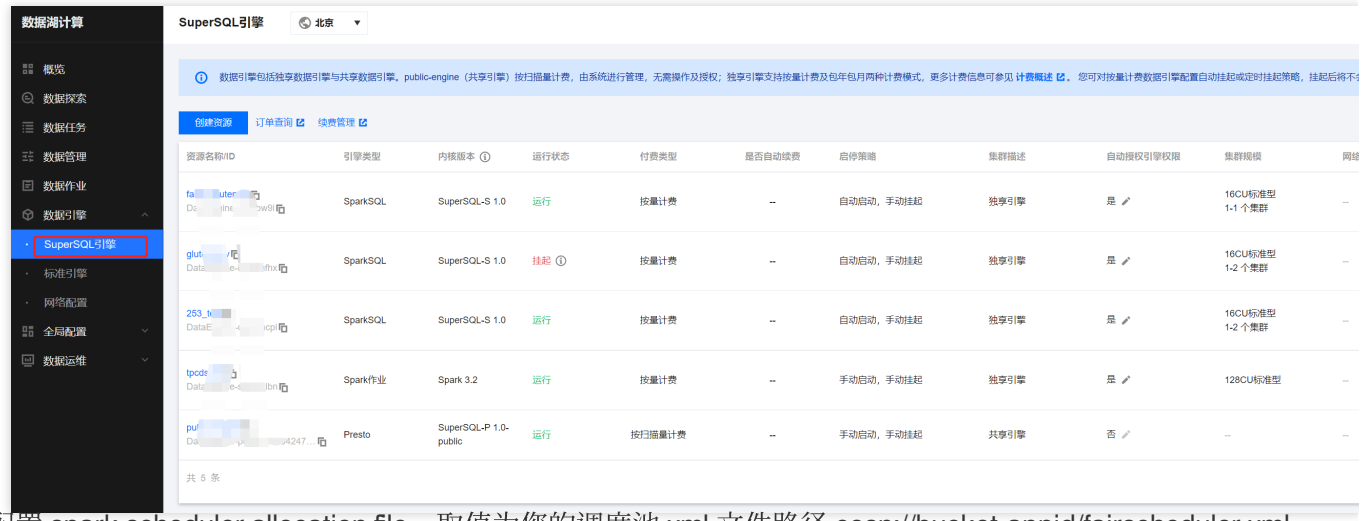


```
<?xml version="1.0"?>
<allocations>
  <pool name="default">
    <schedulingMode>FIFO</schedulingMode>
    <weight>1</weight>
    <minShare>12</minShare>
  </pool>
  <pool name="straggler">
    <schedulingMode>FAIR</schedulingMode>
    <weight>1</weight>
    <minShare>1</minShare>
```

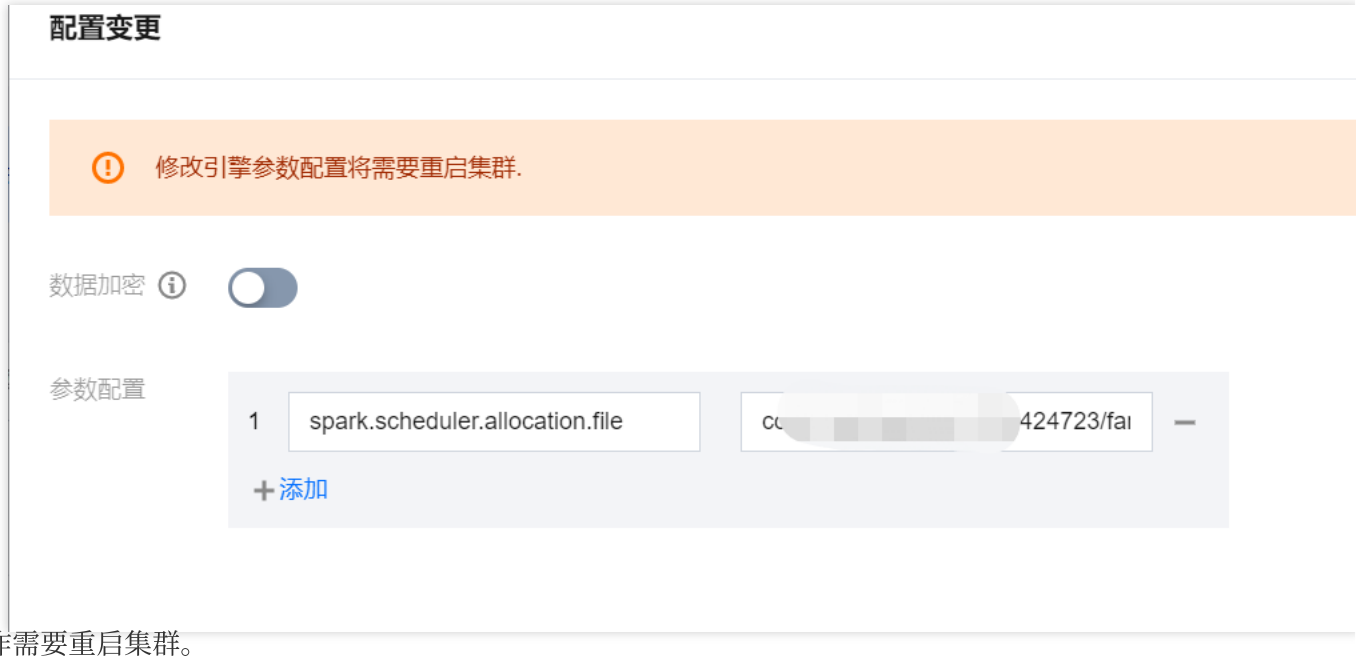
```
    </pool>
    <pool name="special">
      <schedulingMode>FIFO</schedulingMode>
      <weight>1000</weight>
      <minShare>12</minShare>
    </pool>
  </allocations>
```

操作方式

1. 在准备好调度池的 xml 文件后，放置在 cos 的某个路径，例如 cosn://bucket-appid/fairscheduler.xml。
2. 在引擎配置中新增以下配置。

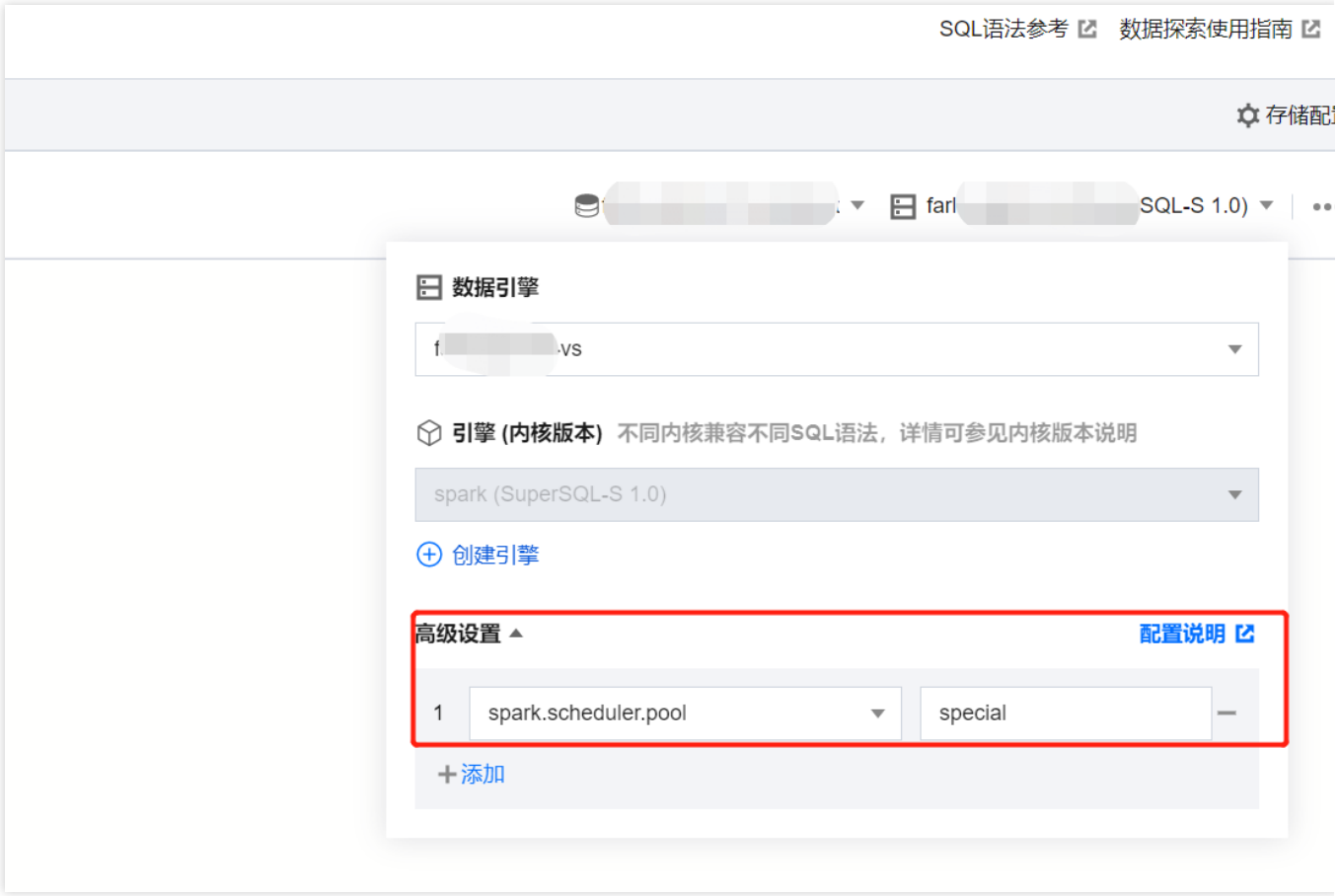


参数配置 spark.scheduler.allocation.file，取值为您的调度池 xml 文件路径 cosn://bucket-appid/fairscheduler.xml。



该操作需要重启集群。

3. 在任务提交时，指定以下参数作为任务参数。spark.scheduler.pool = 要提交到的调度池 name。如果是 default 调度池，可以不指定。



注意事项

调度发生的时间节点在：集群有空闲资源，而有任务需要调度时。因此，如果集群已经被某个任务，例如某个慢任务，全部占满后，必须等待该任务有一个 Task 执行完成，才会开始调度其他优先级更高的任务。所以需要注意，慢任务的单个 Task 耗时最好要相对合理，否则仍然可能会导致长时间占据集群资源。

运维管理
权限管理
访问管理 CAM 服务

最近更新时间：2024-07-17 15:40:54

数据湖计算 DLC 具备完善的数据访问权限机制，DLC 的权限分为操作权限和数据权限。操作权限由访问管理 CAM 服务进行管理，数据权限由 DLC 权限模块进行管理。
主账号默认拥有 DLC 全部操作权限和数据权限。
若子用户被授予 DLC 数据权限管理的操作权限，则该子用户可以将数据权限授予给其他子用户，可以将该类用户视作“管理员”。
若子用户被授予数据读写权限，则该子用户可以运行有权限数据的查询任务，其数据权限是被“管理员”进行分配。
除主账号外的所有子用户的数据权限都是靠“管理员”进行分配。用户不能查询没有权限的数据。
主账户默认拥有 DLC 全部操作权限。主账号通过访问管理 CAM 将 DLC 的访问权限授予给子用户，使子用户拥有对应的 DLC 操作权限。

操作步骤

1. 创建子用户并授权
 在 CAM 控制台创建子用户，并授予对应的权限。具体可参见 子账户授权。
预设策略 QcloudDLCFullAccess：DLC 的全部操作权限。
自定义策略：DLC 指定的操作权限。
2. 子用户登录数据湖计算 DLC 控制台并验证权限。
若子用户登录控制台并执行授权操作成功，则授权生效。

操作权限分类

如下表所示，按照 DLC 接口，将 DLC 操作分类如下。

权限类型	说明
元数据管理	操作 DLC 管理的数据库和表的元数据信息
任务管理	提交和查看 DLC 任务
权限管理	管理用户的数据访问权限
系统配置	DLC 服务基础配置

子账户授权

如果您使用主账号访问数据湖计算 DLC，可以跳过该步骤。

1. 创建子账户的操作步骤，可参见 创建子账号并授权。

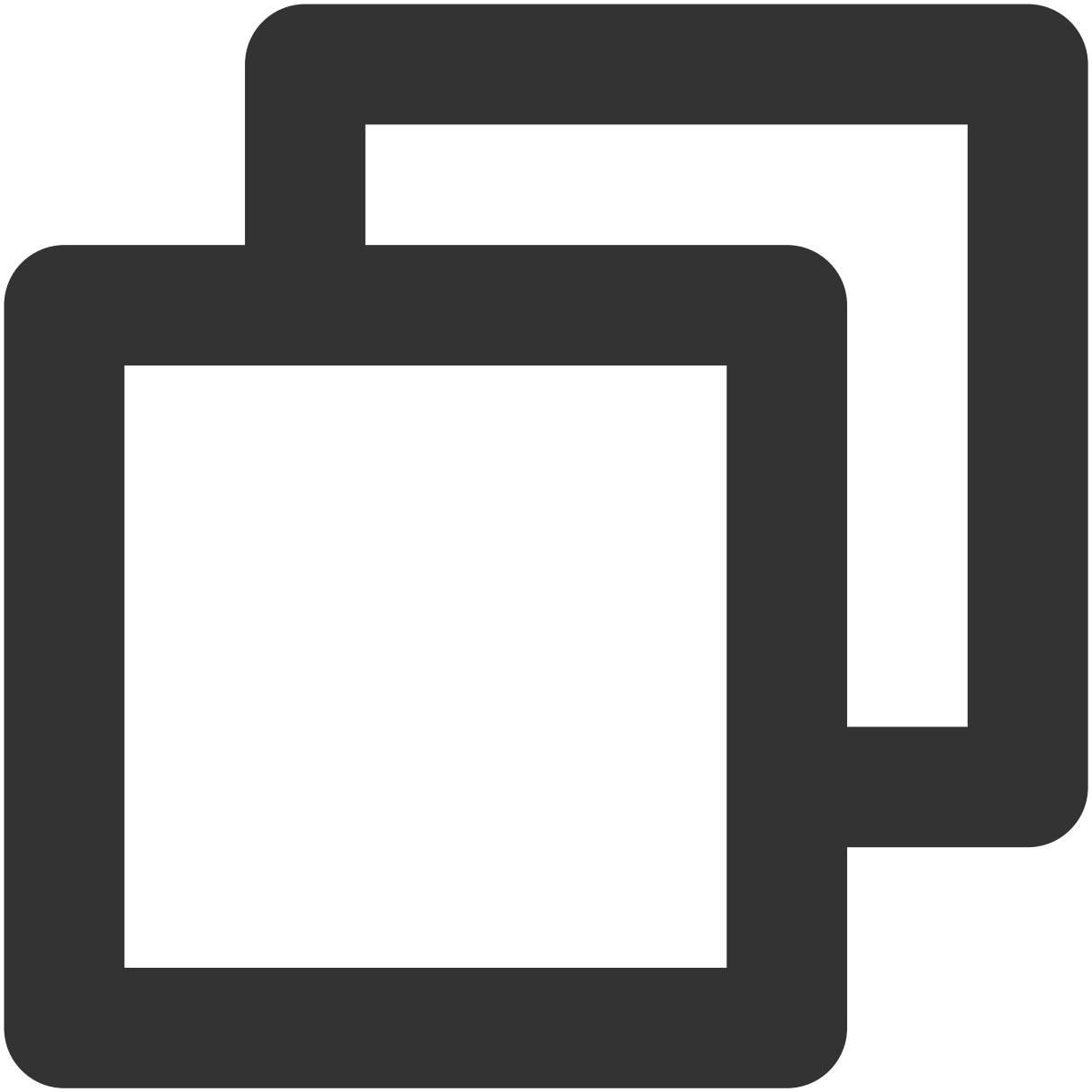
2. 创建自定义策略。

在访问管理控制台的 策略 页面，单击**新建自定义策略**新建策略。

在弹出的选择策略创建方式中，单击**按策略语法创建**，进入编辑策略页面。

在按策略语法创建页面中，选择**空白模板**，单击**下一步**。

在模板中，输入策略名称和描述（建议策略名称为 DLCDataAccess），将如下复制策略粘贴至**策略内容**。填写完成后，单击**完成**，即可成功创建自定义策略。拥有该自定义策略权限的子用户，可以登录数据湖计算 DLC 控制台执行 SQL 任务。但无法操作数据权限管理。



```
{
  "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": [
                "*"
            ]
        }
    ]
}
```

3. 将预设策略或者自定义策略绑定给访问数据湖计算 DLC 的子账户后，该子账户可以登录和访问数据湖计算 DLC，具体可参见 子用户权限设置。

预设策略：QcloudDLCFullAccess。

自定义策略：依据上述步骤自定义创建的数据湖计算 DLC 访问策略。

DLC 权限概述

最近更新时间：2024-07-17 15:42:05

DLC 权限包括数据权限和数据引擎权限。拥有管理权限的用户可以登录 DLC 控制台或者使用 API， 将数据和数据引擎权限授权给子用户。任何子用户的权限都需要被授予，否则无法执行数据和引擎的使用、修改、删除等操作。

用户与工作组

数据湖计算 DLC 为客户提供了两种人员管理模式：用户模式与工作组模式，两种模式皆可对 DLC 的使用者进行权限管理。
用户：CAM 中的用户，包括子账号、协作者账号。
工作组：产品内部管理的包括一批用户的组，组内用户权限相同。
说明：
当用户被赋予的权限与所在工作组权限不同时，两者权限取并集。
通过工作组，您可以快速对人员进行权限赋予，无需重复对每个用户进行授权，企业用户推荐使用工作组模式进行授权。详细操作步骤参见 用户与工作组 。

用户类型

数据湖计算 DLC 的用户类型分为**DLC 管理员**和**普通用户**。
DLC 管理员：负责管理使用 DLC 的用户，拥有所有的数据、引擎、任务等权限，可对用户及工作组进行管理操作（新增、授权、移除）。
普通用户：DLC 的普通使用者，由 DLC 管理员添加，默认无任何 DLC 权限，需要授权以获取相应权限。仅可对拥有**再授权**权限的数据、引擎权限进行授权。

权限&操作	DLC 管理员	普通用户
数据权限	所有权限	默认无，由 DLC 管理员授权
数据引擎权限	所有权限	默认无，由 DLC 管理员授权
用户管理	可进行	不可操作
工作组管理	可进行	不可操作
授权范围	所有权限	权限中**可授权**的权限

说明：

上述权限仅包括数据湖计算 DLC 内定义的权限，购买/变更配置/退费等关于费用的操作需额外前往访问管理 CAM 授予财务权限 QCloudFinanceFullAccess（授权操作可参见：创建子账号并授权），不包含在上述权限中。

数据权限

DLC 的数据权限包括对数据目录、数据库和数据表的操作权限。为方便您的管理及配置，我们提供了两种权限授予模式：常规权限设置、高级权限设置。

常规权限设置模式下，您可以直接给用户授予角色（角色与权限对应参见 子账号权限管理），而无需关注具体的权限配置。授权粒度覆盖数据目录、数据库、数据表。适合快速授权、无需复杂权限管理场景。

高级权限设置模式下，您可以对单个数据库及数据库内的数据表、视图、函数进行全面授权，适合需要精细权限管理场景。

操作权限对应的 SQL 操作如下：

Action	CREATE	ALTER	DROP	SELECT	INSERT	DELETE	Target
CREATE DATABASE	✓	-	-	-	-	-	Cataglog
ALTER DATABASE	-	✓	-	-	-	-	Database
DROP DATABASE	-	-	✓	-	-	-	Database
CREATE TABLE	✓	-	-	-	-	-	Database
CREATE TABLE AS SELECT	✓	-	-	✓	✓	-	Database/Table
DROP TABLE	-	-	✓	-	-	-	Table
ALTER TABLE LOCATION	-	✓	-	-	-	-	Table
ALTER PARTITION LOCATION	-	✓	-	-	-	-	Table
ALTER TABLE ADD PARTITION	-	✓	-	-	-	-	Table
ALTER TABLE	-	✓	-	-	-	-	Table

DROP PARTITION							
ALTER TABLE	-	✓	-	-	-	-	Table
CREATE VIEW	✓	-	-	-	-	-	Database
ALTER VIEW PROPERTIES	-	✓	-	-	-	-	View
ALTER VIEW RENAME	-	✓	-	-	-	-	View
DROP VIEW PROPERTIES	-	✓	✓	-	-	-	View
DROP VIEW	-	-	✓	-	-	-	View
SELECT TABLE	-	-	-	✓	-	-	Table
INSERT	-	-	-	-	✓	-	Table
INSERT OVERWRITE	-	-	-	-	✓	✓	Table
CREATE FUNCTION	✓	-	-	-	-	-	Database
DROP FUNCTION	-	-	✓	-	-	-	Function
SELECT VIEW	-	-	-	✓	-	-	View
SELECT FUNCTION	-	-	-	✓	-	-	Function

数据引擎权限

DLC 的数据引擎操作权限包括对数据引擎的使用、修改、操作、监控、删除权限。具体权限如下：

使用：选择使用该引擎进行任务的权限。

修改：可以修改引擎的基础信息及配置信息（配置信息修改需同时具备 CAM 财务权限）。

操作：对引擎进行挂起、重启操作的权限。

监控：查看引擎的运行任务与监控信息的权限。

删除：对引擎进行退费处理权限。

版权所有：腾讯云计算（北京）有限责任公司

权限授予

单个用户可被授予多组权限，授权详细操作参见 子账号权限管理。

用户和工作组

最近更新时间：2024-07-17 15:43:57

数据湖计算 DLC 为客户提供了两种人员管理模式：用户模式与工作组模式，两种模式皆可对 DLC 的使用者进行权限管理，DLC 相关权限可参见 DLC 权限概述。

基本说明

用户：CAM 中的用户，包括子账号、协作者账号。
工作组：产品内部管理的包括一批用户的组，组内用户权限相同。
说明：
当用户被赋予的权限与所在工作组权限不同时，两者权限取并集。
通过工作组，您可以快速对人员进行权限赋予，无需重复对每个用户进行授权，企业用户推荐使用工作组模式进行授权。

用户管理

用户管理需具备数据湖计算 DLC 的相关操作权限，详情参见 访问管理 CAM 服务。

添加用户

1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 单击**添加用户**，通过用户 ID 将该账号添加进入 DLC 内进行管理。
3. 输入用户 ID 后，可直接为用户绑定工作组（需管理员权限），无需绑定可单击**完成**即可将用户添加入数据湖计算 DLC 中。

查看用户信息

数据湖计算 DLC 的管理员可以修改用户基本信息及权限信息。
1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 找到需要查看的用户**账号 ID**，可通过搜索快速查找。单击该**用户名称**即可查看该用户相关的信息及权限。

编辑用户信息

您可以通过编辑功能修改用户的描述信息及工作组，权限相关变更操作可参见 子账号数据权限授权。
1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 找到需要编辑的用户账号ID，可通过搜索快速查找。单击条目操作栏的**编辑**按钮即可进入编辑页。

移除用户

如果您不希望该用户继续使用数据湖计算 DLC，可以使用管理员账号对用户进行移除，移除后该用户将被回收数据湖计算 DLC 的使用权限，无法进行使用。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 找到需要移除的用户账号 ID，可通过搜索快速查找。单击该用户条目前的选择框，支持批量进行移除，选择完毕后，单击**批量移除**即可将选中的用户从数据湖计算 DLC 中移除。
3.

工作组管理

工作组管理需具备数据湖计算 DLC 的相关操作权限，详情参见 访问管理 CAM 服务。

添加工作组

您可以将需要重复赋予的权限以工作组的形式管理，添加工作组操作流程如下。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 单击**工作组**，进入工作组管理页面。
3. 单击**添加工作组**，填写相关信息后确认，即可完成工作组的添加。
4.

查看工作组信息

编辑工作组支持修改工作组描述信息及包括的用户。编辑工作组操作流程如下。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 单击**工作组**，进入工作组管理页面。
3. 找到需要查看的工作组，可通过搜索快速定位，单击**工作组 ID** 或**工作组名称**即可查看工作组信息。

编辑工作组信息

编辑工作组支持修改工作组描述信息及包括的用户。编辑工作组操作流程如下。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 单击**工作组**，进入工作组管理页面。
3. 找到需要编辑的**工作组名称**，单击操作栏的**编辑**按钮即可进入编辑界面。

如需编辑描述信息，可单击

进行编辑。

可通过绑定用户按钮将数据湖计算 DLC 内的用户添加至工作组。

勾选用户后点击批量移除，或点击用户操作栏的移除按钮，都可将用户从工作组移除。移除后用户将不再拥有该工作组授予的权限，但不影响其他已授予的权限。

删除工作组

数据湖计算 DLC 管理员可以将工作组进行移除。

注意：

移除后该工作组下的所有授权用户的该工作组相关权限将会被回收。移除后将不可恢复，请谨慎操作。

1. 登录 数据湖计算 DLC 控制台，选择服务地域，进入**权限管理页**。
2. 单击**工作组**，进入工作组管理页面。
3. 找到需要移除的工作组名称，单击表头的选项框后批量移除，或单击操作栏的移除按钮，都可以将对应工作组移除。

子账号权限管理

最近更新时间：2024-07-17 15:46:45

用户权限

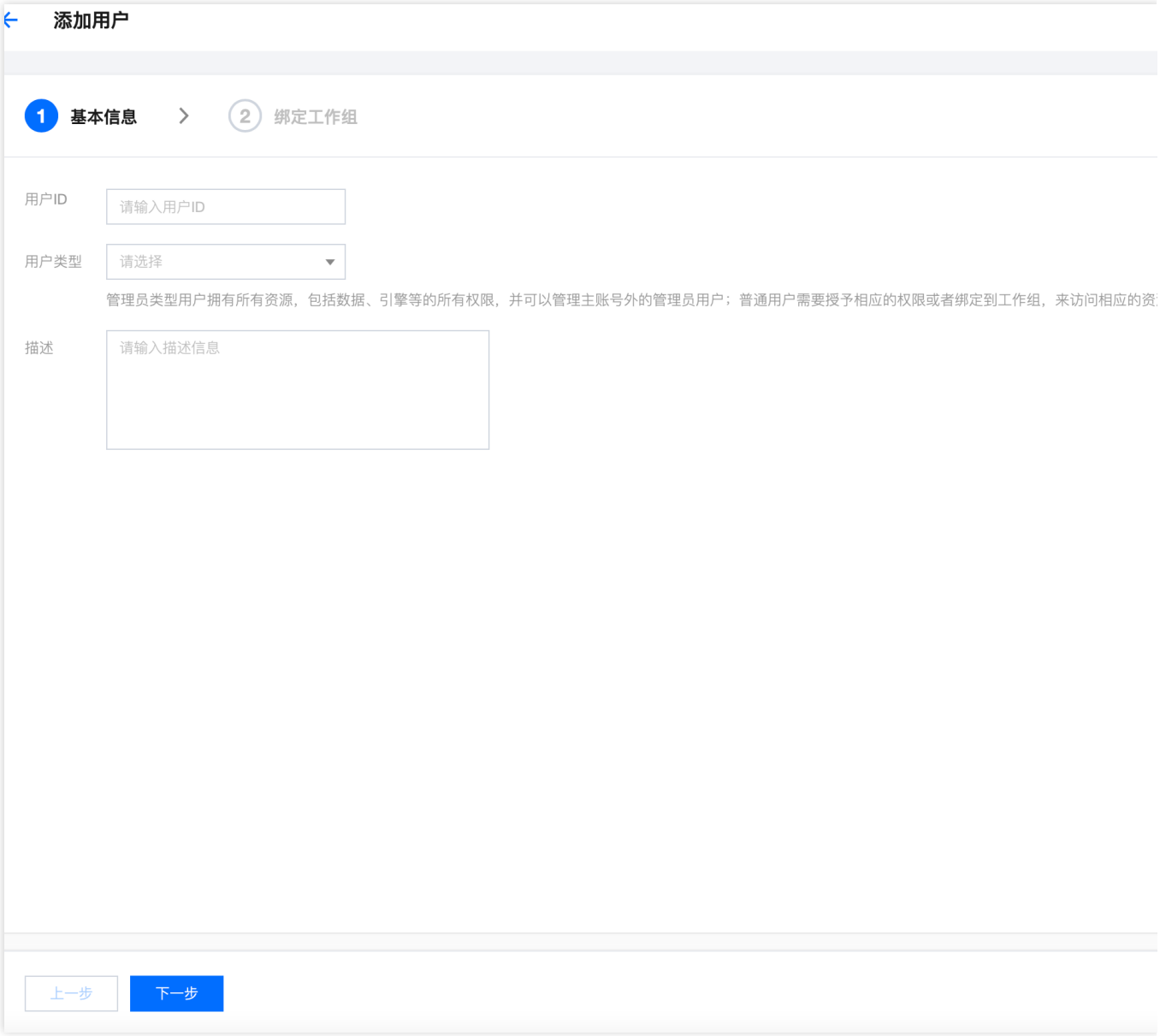
用户权限包括数据权限和引擎权限（详细权限说明可参见 DLC 权限概述），需要有相应的数据权限，才能访问 DLC 中的数据。DLC 提供了库表级的权限管理，以及列级别的细粒度权限管理，方便您在不同场景下对用户或者组进行数据授权，精细化管理数据权限。另外，可以通过相应的引擎权限来管理资源。

用户与工作组

可以单独为用户授予权限，也可以创建包括一批用户的工作组，进行权限授权。用户与工作组管理操作可参见 用户和工作组。
用户：CAM 中的用户，包括子账号、协作者账号。
工作组：产品内部管理的包括一批用户的组，组内用户权限相同。
说明：
当用户被赋予的权限与所在工作组权限不同时，两者权限取并集。
通过工作组，您可以快速对人员进行权限赋予，无需重复对每个用户进行授权，企业用户推荐使用工作组模式进行授权。

添加用户权限

为指定用户授予权限。
1. 设置用户为管理员/普通用户。管理员类型用户不需要绑定工作组即拥有所有资源，包括数据、引擎等的所有权限，并可以管理主账号外的管理员用户。**该权限需要谨慎设置**。

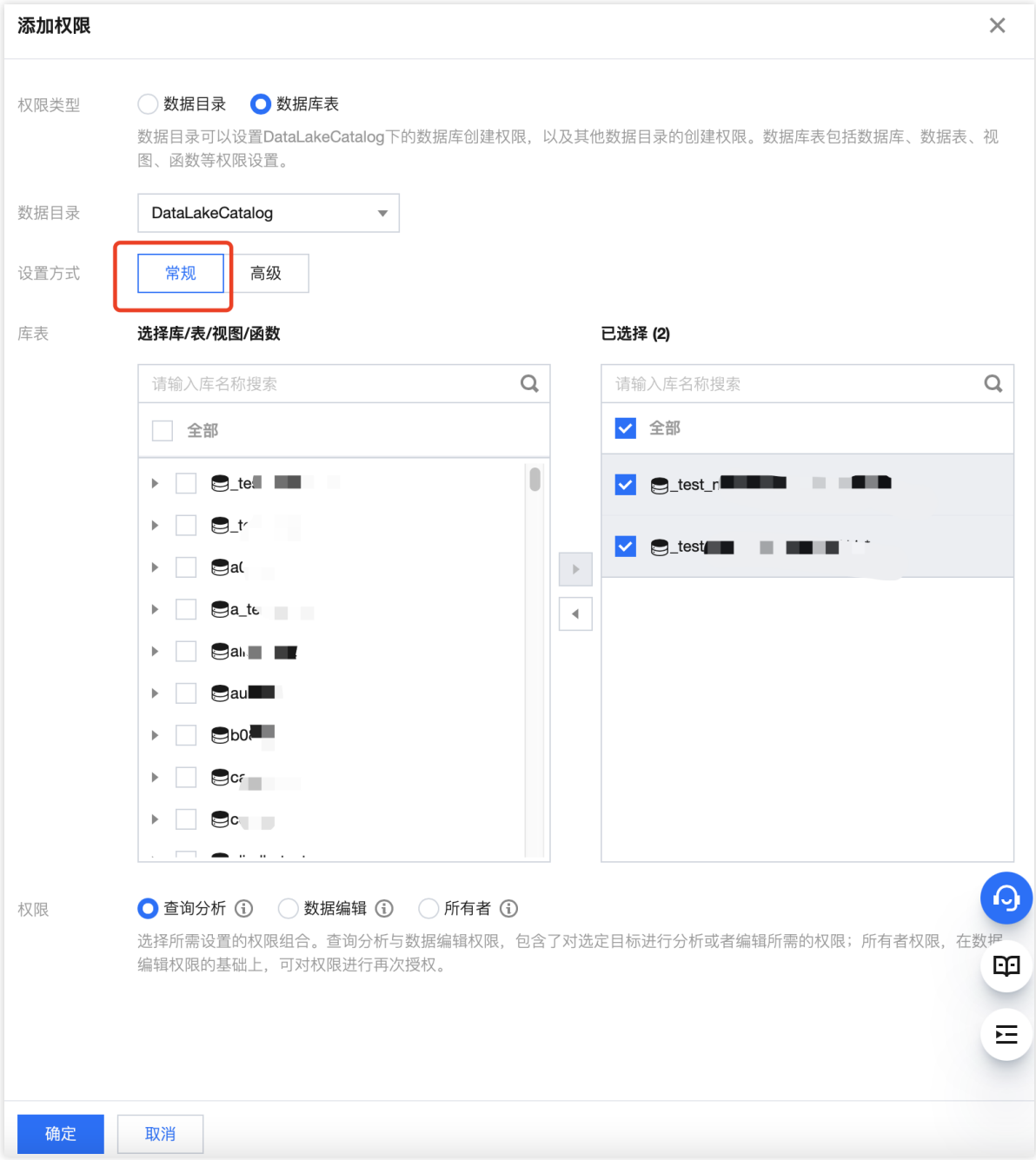


2. 绑定工作组：对于普通用户，需要授予相应的权限或者绑定到工作组，来访问相应的资源。

3. 添加数据权限：在用户列表中选择授权操作，选择数据权限，可以选择授予用户数据目录或库表的权限。

添加数据目录权限。数据目录可以设置 DataLakeCatalog 下的数据库创建权限，以及其他数据目录的创建权限。

添加数据库表权限。可以通过常规设置和高级设置两种模式来添加权限。常规权限设置：可以添加指定目录下的数据库表权限，设置查询分析、数据编辑、所有者权限。



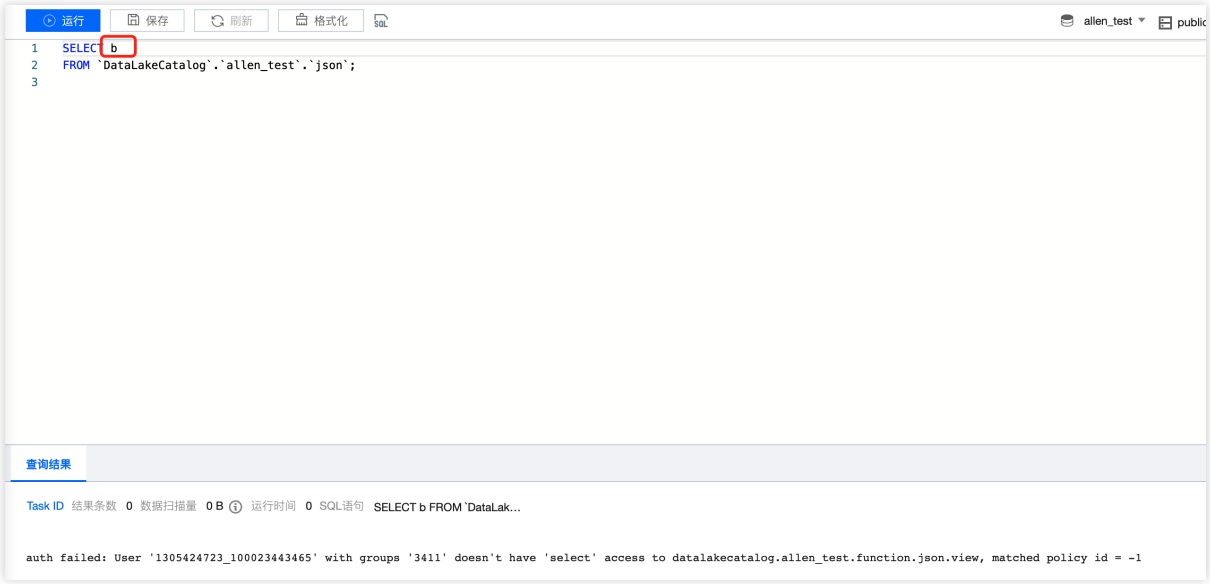
具体权限如下：

权限类型	数据库	数据表	视图与函数
查询分析权限	对数据库中所有表、试图、函数的查询权限。创建数据表的权限。	查询	查询
数据编辑权限	库的修改、删除、建表权限。 所有表、视图、函数的所有权限。	数据的查询、插入、更新、删除。• 表的修改、删除。	查询、创建、修改、删除

所有者权限（在数据编辑权限的基础上，可对权限进行再次授权。）	库的修改、删除、建表。 所有表、视图、函数的所有权限。	数据的查询、插入、更新、删除。 表的修改、删除。	查询、创建、修改、删除

高级权限设置：选择单个数据库，可继续设置库下表、视图、函数的查询、插入、更新、删除，选择多个数据库时，将只设置数据库的权限。
高级模式下支持在列级别进行权限设置。选择单个数据表，可添加列的查询权限。支持选择一个/多个列或选择全部列进行授权。

单击**确定**添加完成后，在数据探索模块执行查询，输入以下sql语句预览“**col1**”的信息，运行后可查看该列的预览结果。
对于该数据表中未授予列权限的数据列 **b**，输入 sql 语句查看数据列 **b** 的信息，该查询因为权限问题无法查询。

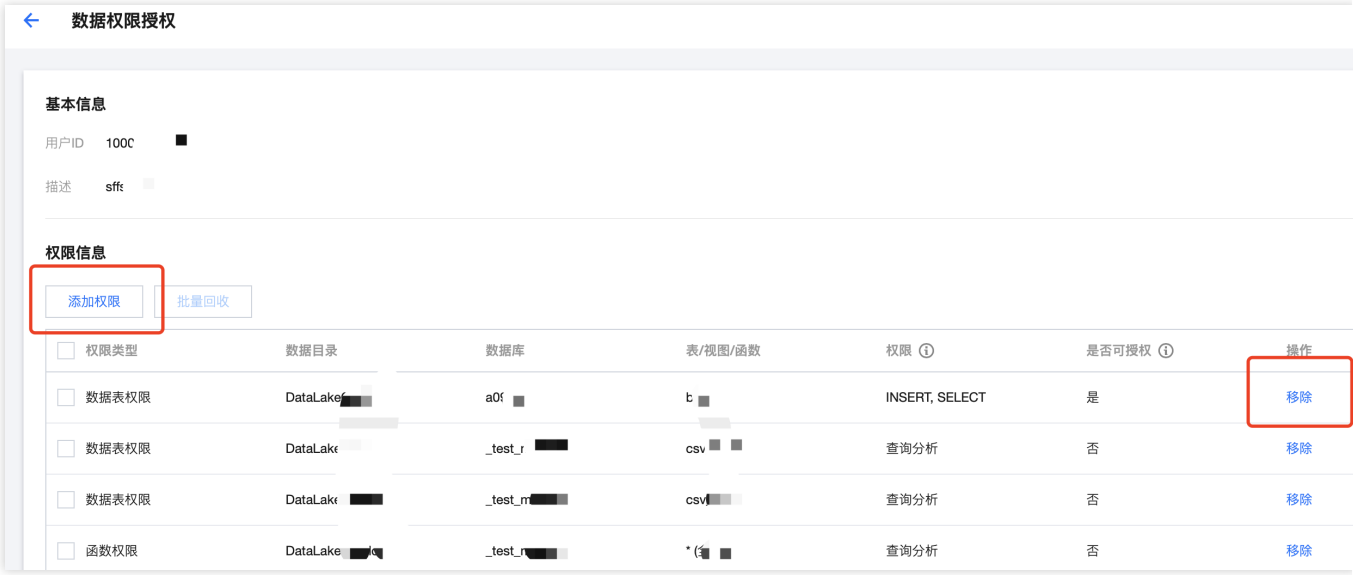


4. 添加引擎权限：在用户列表中选择授权操作，选择引擎权限，为指定资源授予使用、修改、操作、监控、删除权限。

修改用户权限

1. 在用户列表中选择授权，选择数据权限或引擎权限。

进入授权界面（以数据权限为例），通过添加权限或移除权限，进行权限的修改（引擎权限的修改执行同样的操作）。



2. 修改工作组或用户类型，可单击**操作 > 编辑**，进入用户编辑用户的界面，支持修改用户名称、用户类型、描述。普通用户可添加或移除工作组。

单击**编辑**按钮，可更改用户类型。

查看用户权限

1. 单击用户列表中的用户 ID，进入查看用户详情页。

2. 查看用户所在工作组、数据权限、引擎权限等信息。

移除用户权限

对于需要回收的权限，可在用户的权限列表中，将该权限移除。移除用户需要管理员操作。

添加与移除组的权限

添加与移除工作组需要管理员操作，与用户的数据权限操作相似。组中的用户拥有该组所拥有的所有权限。可以通过将一批用户绑定到工作组，并授予该组数据、引擎等资源权限，来批量管理用户权限。管理员不需要绑定到工作组。

存储配置
托管存储配置

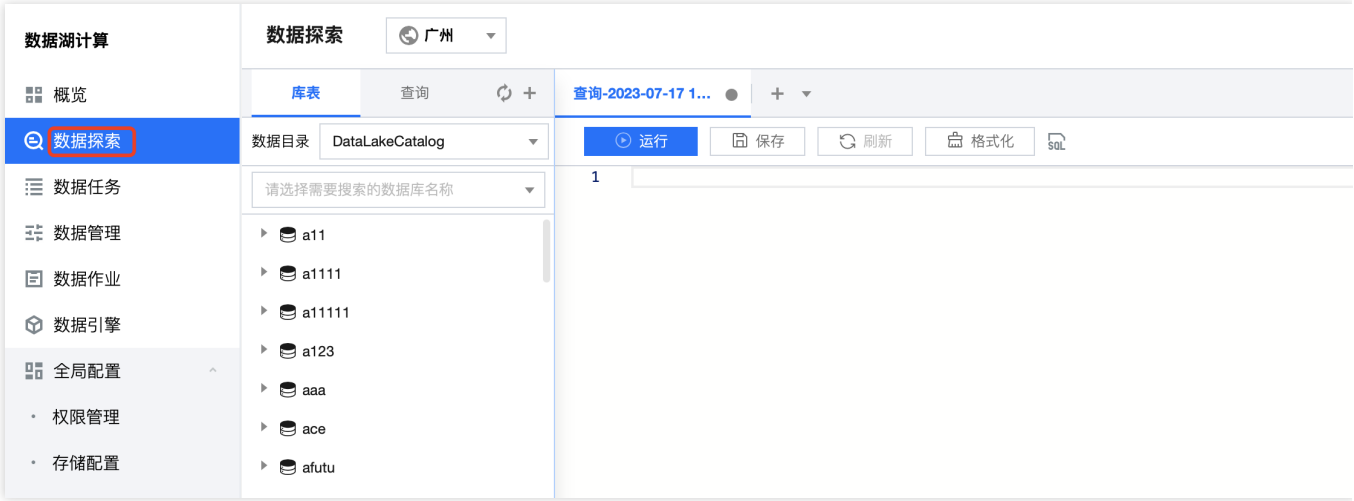
最近更新时间：2024-07-31 17:30:19

托管存储是指用户在数据湖产品上托管的存储空间，存储底层为对象存储。托管存储上存放着原生表、用户程序包、查询结果等数据，因此如果需要使用原生表及其数据优化等能力，需要先开启托管存储。托管存储上的原生表默认为 Iceberg 格式表 ，您无须管理底层的文件内容。托管存储计费详情请参见 计费概述。
本篇文档介绍托管存储的开启和配置步骤。

开启托管存储

步骤一：进入托管存储配置

您可以在 数据探索 模块或者 全局配置 > 存储配置 模块中进入托管存储配置。



步骤二：打开托管存储

1. 勾选启用托管存储并保存。
此处您可以指定托管存储类型为元数据加速桶或者普通桶，元数据加速桶和普通桶在计费上一致，但是需要单独配置引擎访问元数据加速桶的权限，详情请参见 元数据加速桶的绑定。
2. 查询结果路径用以临时存储 SQL 查询的结果数据、Spark 作业 Shuffle 数据等等，您需要指定一个路径以保证作业和任务的正常运行。如果您开启了托管存储，建议查询结果路径配置为**托管存储**。您也可以把查询结果路径配置到自己账号下的 COS 存储桶 路径。

版权所有：腾讯云计算（北京）有限责任公司

查看托管存储桶

开启托管存储后会创建一个存储桶，可以在 数据管理 模块查看托管存储上的桶以及数据。

销毁托管存储

销毁数据是高危行为，只有当库表数据都删除后，您才可以销毁托管存储。销毁托管存储需要管理员权限。

步骤一：删除库表数据

销毁托管存储需要先删除托管存储上的所有库表数据。
您可以参考 数据目录及数据库管理和数据表管理这两篇文档删除掉库表数据，也可以在 数据探索 模块 运行 DROP 语法删除掉库表数据。

步骤二：销毁托管存储

删除库表数据后，您可以在存储配置模块下的托管存储配置页签，销毁托管存储。
销毁托管存储会把所有的 DLC 托管存储桶都删除，请谨慎操作。

元数据加速桶的绑定

最近更新时间：2024-07-31 17:30:34

数据湖计算 DLC 支持绑定融合桶加速查询分析性能，使用该功能您需要创建元数据加速桶，DLC 托管存储提供元数据加速桶，使用用户账号下的对象存储桶 详情请参见 对象存储 > 元数据加速。
引擎访问 DLC 元数据加速桶时，需要绑定权限，权限绑定流程如下。

绑定数据引擎和元数据加速桶

1. 登录 数据湖计算控制台，进入公共管理 > 存储配置。
2. 进入**元数据加速桶配置页**，选择需要绑定的桶，单击**配置**。

说明：
元数据加速桶页面下只展示元数据加速桶，此处不会展示普通桶（未开启元数据加速功能的桶）。

3. 单击**绑定**，将需要访问该桶的数据引擎绑定到该元数据加速桶上。

绑定流计算 Oceanus 的计算资源

如果您使用 流计算 Oceanus 来流式写入数据湖，且写入的存储为元数据加速桶，则需要在存储配置下配置访问元数据加速桶的权限。在 腾讯云产品绑定 这一栏下，新建产品，选择流计算Oceanus和对应的资源后点击保存即可。

绑定非 DLC 数据引擎的计算资源

有时候您需要访问元数据加速桶的计算资源并非 DLC 的计算引擎。这时您可以在存储配置下配置访问元数据加速桶的权限。

HDFS 用户配置用以配置您访问 DLC 的计算资源的超级用户，一般为 root/hadoop/presto/flink。

HDFS 元数据权限配置用以配置您允许访问 DLC 的 VPC 网络环境，一般为上述非 DLC 数据引擎的计算资源所在的 VPC。

HDFS 用户配置 编辑

超级用户 root
hadoop
presto

说明：HDFS 用户配置用于管理计算节点的租户信息

HDFS 元数据权限配置

VPC网络名称/ID	节点IP地址	操作
vpc-emr (10.0.0.0/20)	10.0.0.0/20	编辑 删除
	新增权限配置	

审计日志

最近更新时间：2024-07-31 17:31:00

数据湖计算 DLC 提供了基于腾讯云操作审计服务的操作日志审计服务，保障您能实时了解系统操作记录，核对操作信息。

使用须知

使用数据湖计算 DLC 的审计日志服务前，您需要开通腾讯云的**操作审计服务**。如暂未开通该服务，可使用主账号开通该服务。

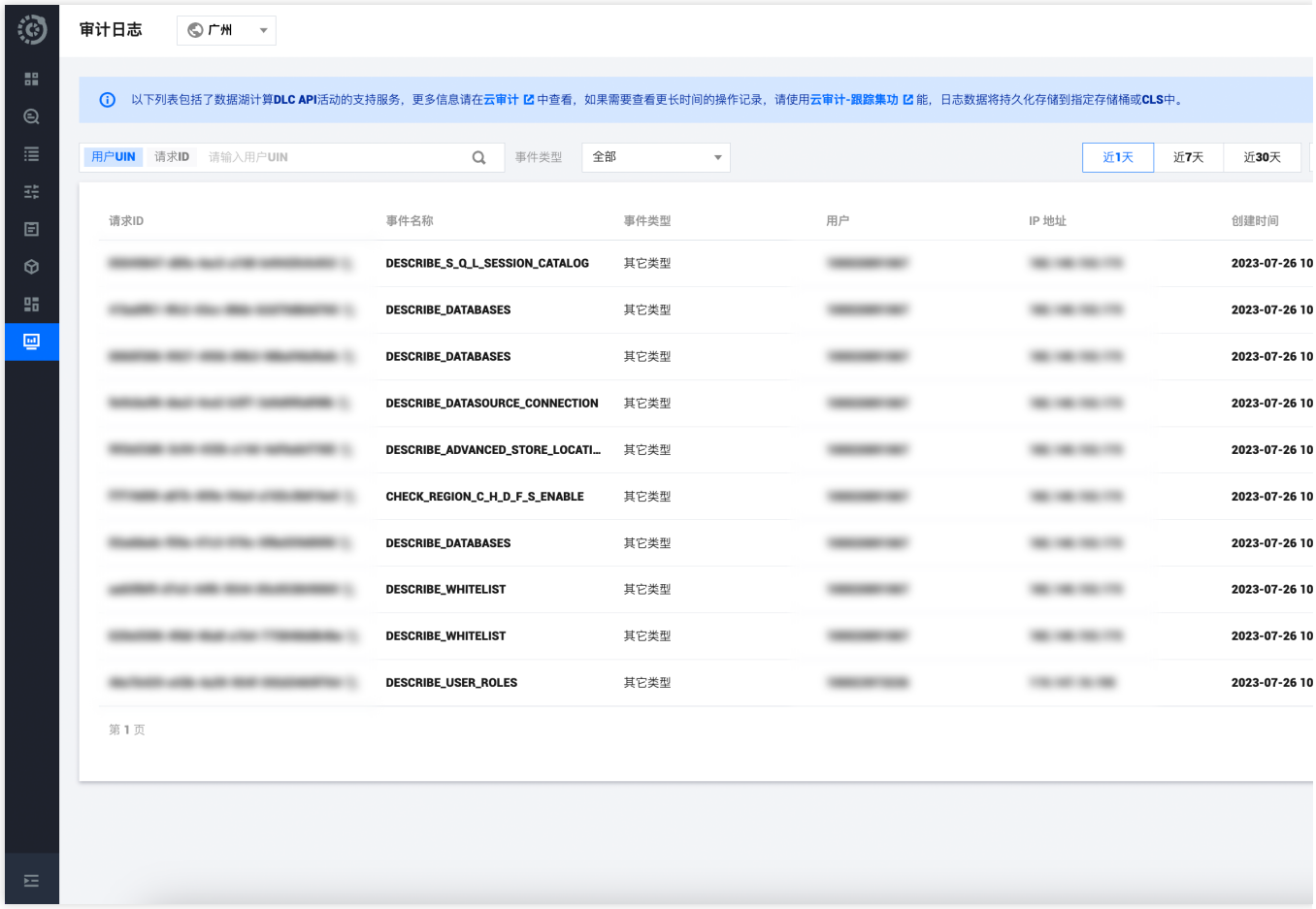
使用说明

数据湖计算控制台目前最多展示近3个月的日志信息，如需查看更早的日志信息，可至操作审计查看。
审计日志包含了控制台操作及 API 调用操作。目前支持查看引擎管理、任务管理、数据源管理、工作组管理、用户管理、调度任务实例管理、调度任务管理、调度计划管理日志信息。

操作指南

1. 登录 数据湖计算控制台，选择**服务地域**。
2. 通过左侧菜单**数据运维**，选择审计日志功能。
3. 支持基于用户 UIN 或请求 ID 进行日志查询。
4. 详细日志信息可单击**查询详情**查看。

版权所有：腾讯云计算（北京）有限责任公司


审计日志
广州
以下列表包括了数据湖计算DLC API活动的支持服务，更多信息请在云审计 中查看，如果需要查看更长时间的操作记录，请使用云审计-跟踪集功 能，日志数据将持久化存储到指定存储桶或CLS中。
用户UIN
请求ID
请输入用户UIN
事件类型
全部
近1天
近7天
近30天
请求ID
事件名称
事件类型
用户
IP 地址
创建时间
DESCRIBE_S_Q_L_SESSION_CATALOG 其它类型 2023-07-26 10
DESCRIBE_DATABASES 其它类型 2023-07-26 10
DESCRIBE_DATABASES 其它类型 2023-07-26 10
DESCRIBE_DATASOURCE_CONNECTION 其它类型 2023-07-26 10
DESCRIBE_ADVANCED_STORE_LOCATI... 其它类型 2023-07-26 10
CHECK_REGION_C_H_D_F_S_ENABLE 其它类型 2023-07-26 10
DESCRIBE_DATABASES 其它类型 2023-07-26 10
DESCRIBE_WHITELIST 其它类型 2023-07-26 10
DESCRIBE_WHITELIST 其它类型 2023-07-26 10
DESCRIBE_USER_ROLES 其它类型 2023-07-26 10
第 1 页

监控与告警
数据引擎监控

最近更新时间：2024-07-31 17:31:25

数据湖计算 DLC 提供了基于腾讯云可观测平台服务的数据引擎的监控服务，保障您能实时了解数据引擎的运行情况，配置数据告警。告警配置方式可参见 监控告警配置。

使用须知

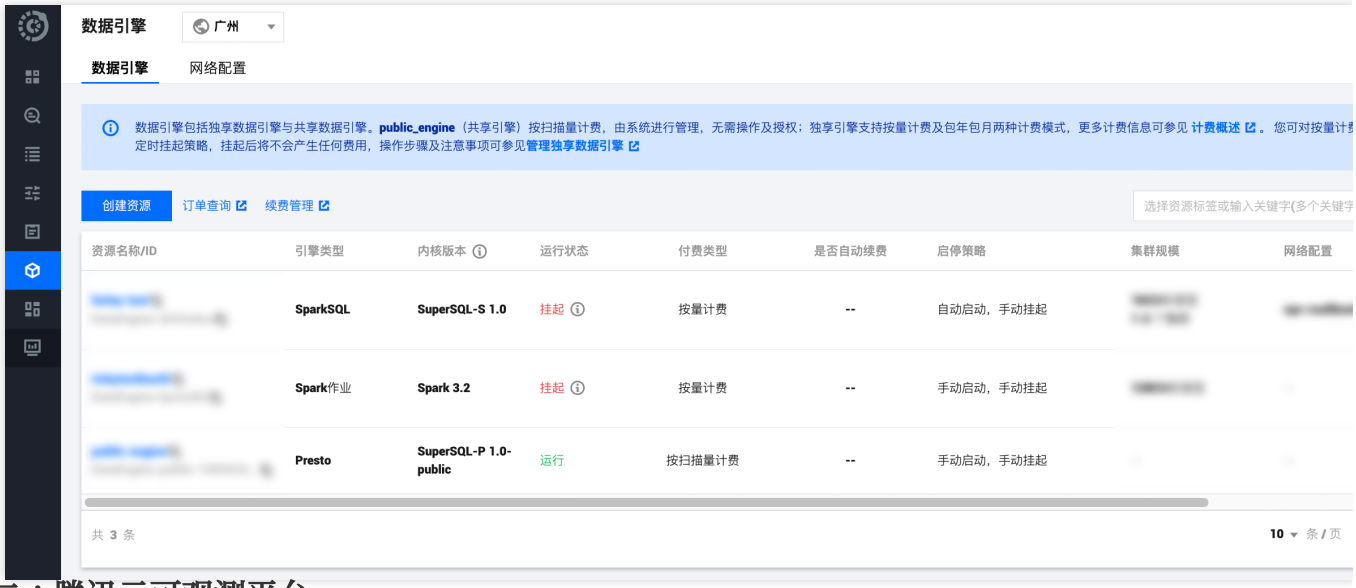
使用数据湖计算 DLC 的监控服务前，您需要开通腾讯云可观测平台服务。如暂未开通该服务，可使用主账号开通该服务。
腾讯云可观测平台服务使用过程中可能产生相关资费，详细资费信息参见 计费概述。

监控入口

入口一：数据湖计算 DLC 控制台

注意：
账号需有数据引擎的监控权限。
1. 登录 数据湖计算 DLC 控制台，选择服务地域。
2. 左侧菜单栏进入**数据引擎**页。
3. 查看方式支持：
方式1：选择引擎类型，进入匹配引擎监控列表。
方式2：选择引擎列表中目标引擎，单击**监控**查看目标引擎监控。

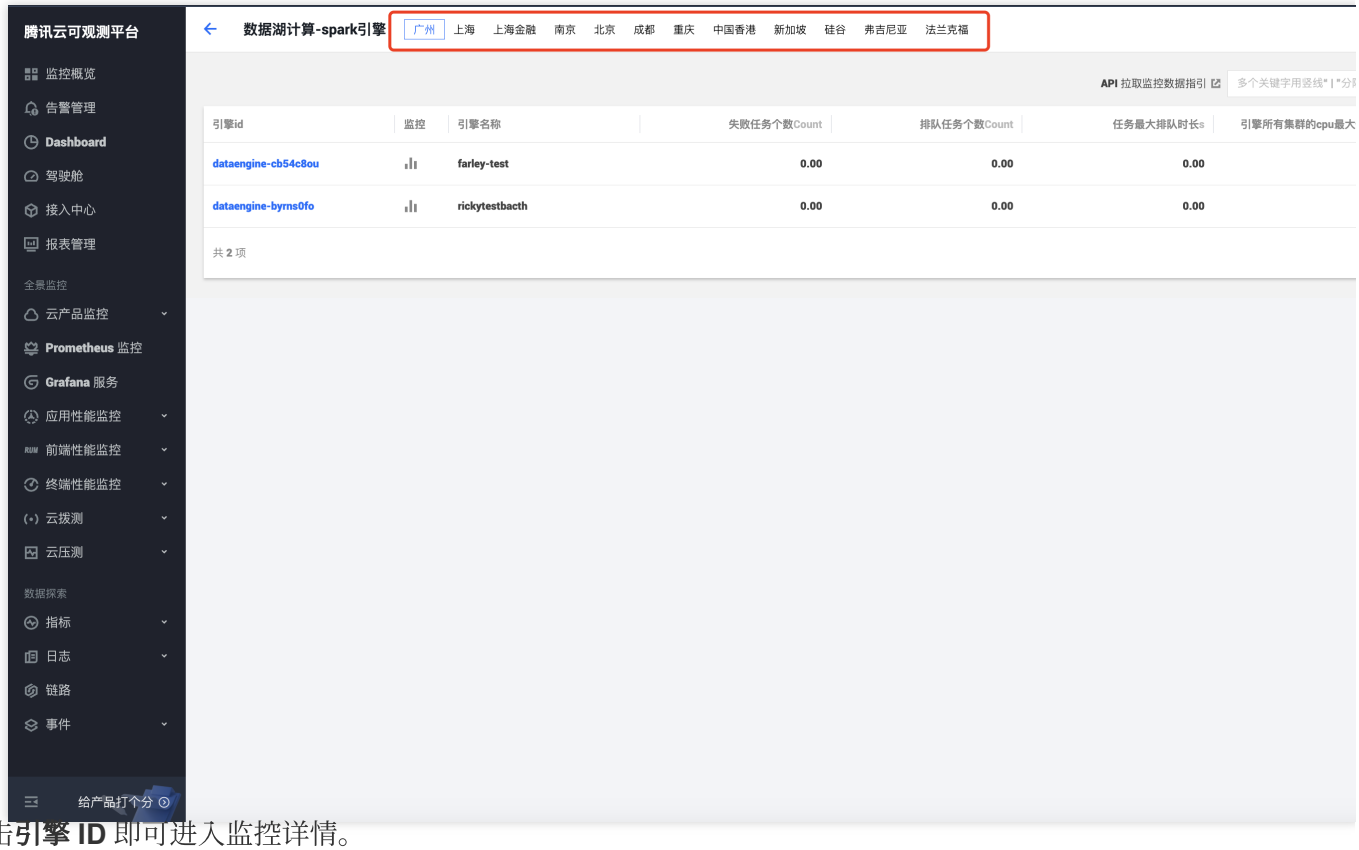
版权所有：腾讯云计算（北京）有限责任公司



入口二：腾讯云可观测平台

1. 登录 腾讯云可观测平台，登录账号需具备相关权限。
2. 左侧菜单选择**云产品监控**，找到数据湖计算 DLC，并选择需要查看监控的类型。

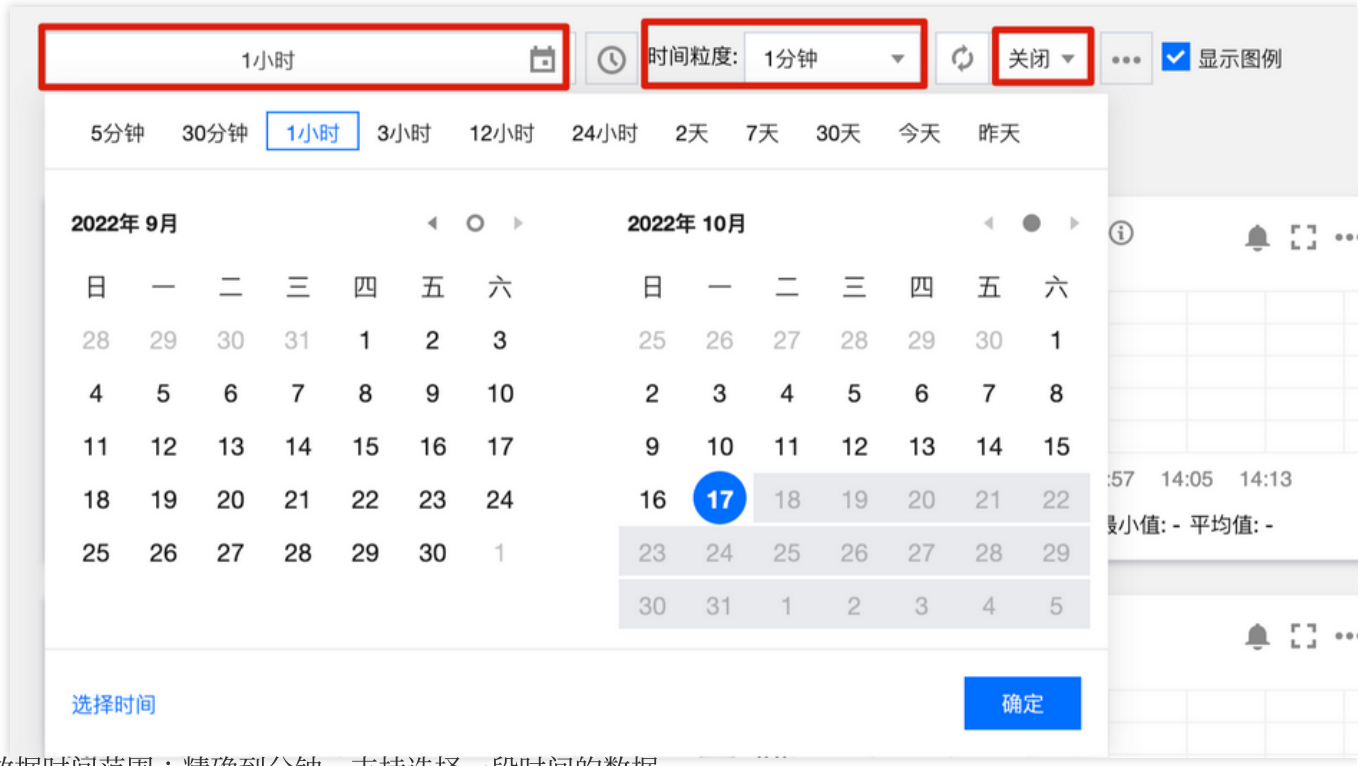
3. 选择监控类型后进入监控页，选择对应地域即可查看该地域下的监控资源信息。



4. 单击**引擎 ID** 即可进入监控详情。

监控粒度配置

支持通过监控顶部配置监控数据时间范围、时间粒度、自动更新时间范围。



监控数据时间范围：精确到分钟，支持选择一段时间的数据。

时间粒度：监测点间隔时间，支持配置1分钟或5分钟。

自动更新数据：页面数据自动刷新方式配置，支持配置关闭、30s、5min、30min、1h。

监控数据对比

支持选择一段时间的数据进行监控对比。单击选择对比时间范围后即可在下方数据罗盘查看对比数据。

监控指标

监控类型	监控指标

CPU	所有 Driver 节点 CPU 最大使用率
	所有 Executor 节点 CPU 最大使用率
	所有 Driver 节点 CPU 平均使用率
	所有 Executor 节点 CPU 平均使用率
	所有集群 CPU 最大使用率
	所有集群 CPU 平均使用率
内存	所有 Driver 节点内存最大使用率
	所有 Executor 节点内存最大使用率
	所有 Driver 节点内存平均使用率
	所有 Executor 节点内存平均使用率
	所有集群内存最大使用率
	所有集群内存平均使用率
任务	取消任务个数
	失败任务个数
	初始化任务个数
	任务平均初始化时长
	任务最大初始化时长
	排队任务个数
	任务平均排队时长
	任务最大排队时长
	运行中任务个数
	成功任务个数
网络	所有 Driver 节点网络最大入带宽
	所有 Executor 节点网络最大入带宽
	所有 Driver 节点网络平均入带宽

	所有 Executor 节点网络平均入带宽
	所有 Driver 节点网络最大出带宽
	所有 Executor 节点网络最大出带宽
	所有 Driver 节点网络平均出带宽
	所有 Executor 节点网络平均出带宽
云盘	所有 Driver 节点云盘最大使用率
	所有 Executor 节点云盘最大使用率
	所有 Driver 节点云盘平均使用率
	所有 Executor 节点云盘平均使用率
CU	作业引擎 CU 数
	CU 使用率

数据作业监控

最近更新时间：2024-07-31 17:31:46

数据湖计算 DLC 提供了基于腾讯云可观测平台服务的数据作业的监控服务，保障您能实时了解数据作业的运行情况，配置数据告警。

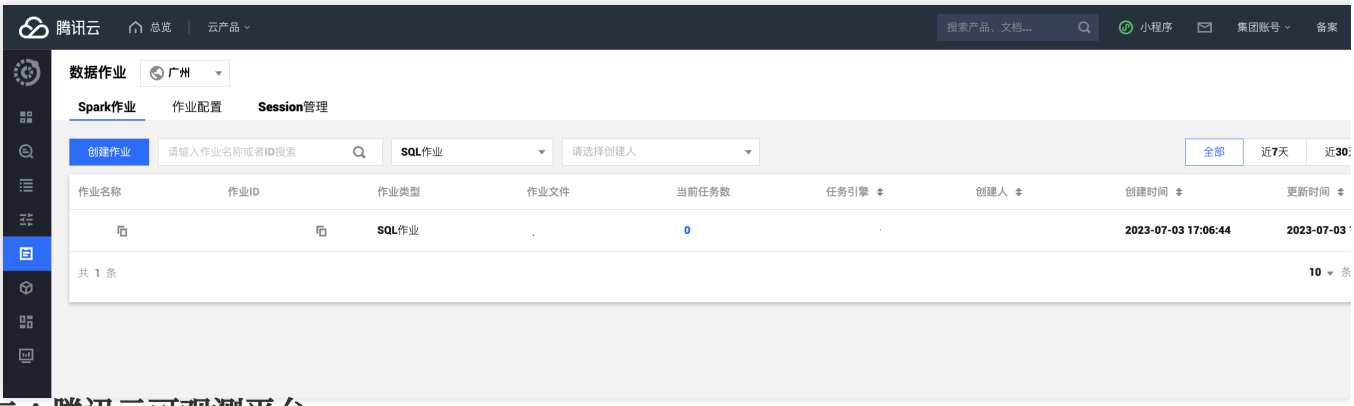
使用须知

使用数据湖计算 DLC 的监控服务前，您需要开通腾讯云可观测平台服务（使用详情可参见 腾讯云可观测平台文档）。如暂未开通该服务，可使用主账号开通该服务。
腾讯云可观测平台服务使用过程中可能产生相关资费，详细资费信息参见 腾讯云可观测平台计费概述。

监控入口

入口一：数据湖计算 DLC 控制台

1. 登录 数据湖计算 DLC控制台 > 数据作业，选择服务地域。
2. 或在左侧菜单栏进入数据作业页。
3. 在右上角单击**作业监控**，将进入监控页。或单击目标作业的**监控**功能，进入目标作业监控页。



入口二：腾讯云可观测平台

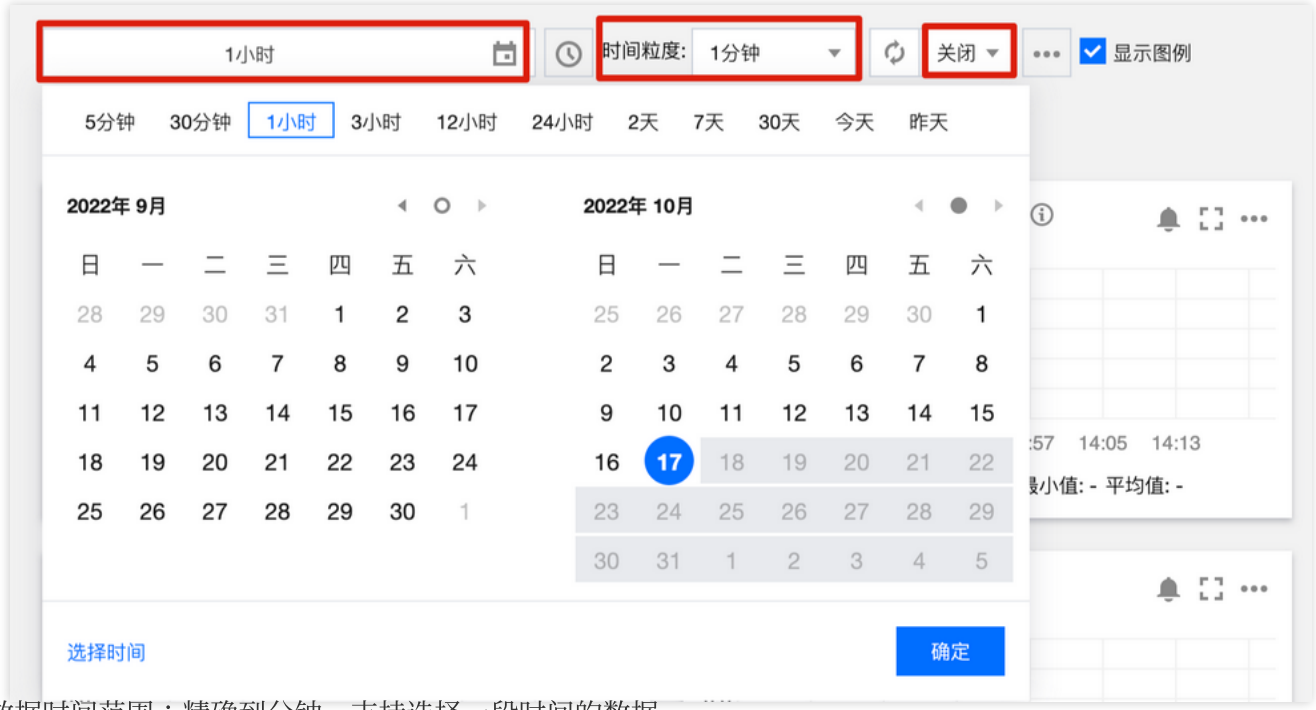
1. 登录 腾讯云可观测平台控制台，登录账号需具备相关权限。
2. 左侧菜单选择云产品监控，找到数据湖计算 DLC，并选择需要查看监控的类型。

3. 选择监控类型后进入监控页，选择对应地域即可查看该地域下的监控作业信息。

4. 单击**作业 ID** 即可进入监控详情。

监控粒度配置

支持通过监控顶部配置监控数据时间范围、时间粒度、自动更新时间范围。



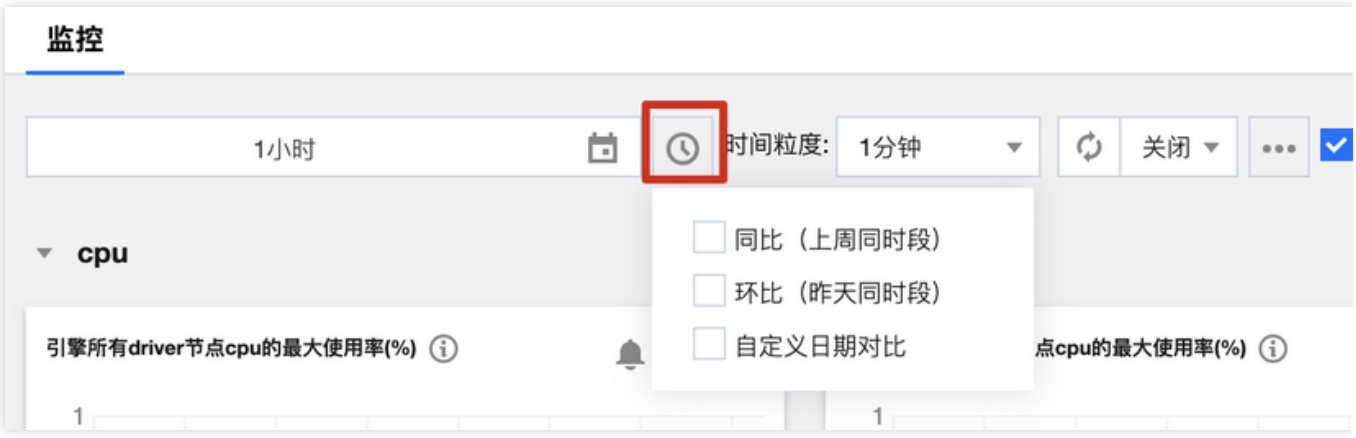
监控数据时间范围：精确到分钟，支持选择一段时间的数据。

时间粒度：监测点间隔时间，支持配置1分钟或5分钟。

自动更新数据：页面数据自动刷新方式配置，支持配置关闭、30s、5min、30min、1h。

监控数据对比

支持选择一段时间的数据进行监控对比。点击选择对比时间范围后即可在下方数据罗盘查看对比数据。



监控指标

监控类型	监控指标
作业	作业 error 日志数

	作业 warn 日志数

接入点网关引擎监控

最近更新时间：2024-07-31 17:32:00

数据湖计算 DLC 提供了基于腾讯云可观测平台服务的接入点网关引擎的监控服务，保障您能实时了解网关情况。

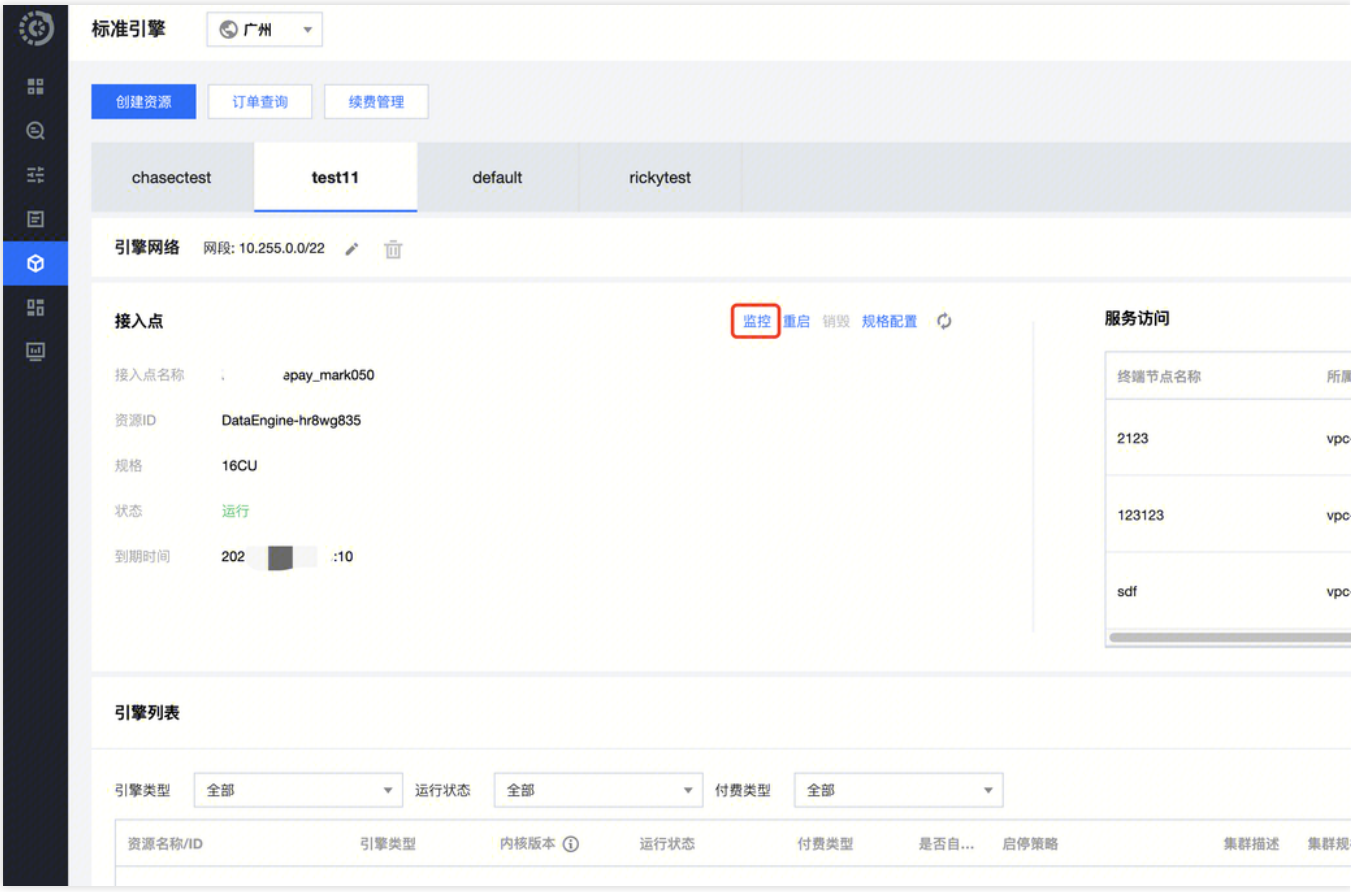
使用须知

使用数据湖计算 DLC 的监控服务前，您需要开通腾讯云可观测平台服务（使用详情可参见 腾讯云可观测平台文档）。如暂未开通该服务，可使用主账号开通该服务。
腾讯云可观测平台服务使用过程中可能产生相关资费，详细资费信息参见 腾讯云可观测平台计费概述。

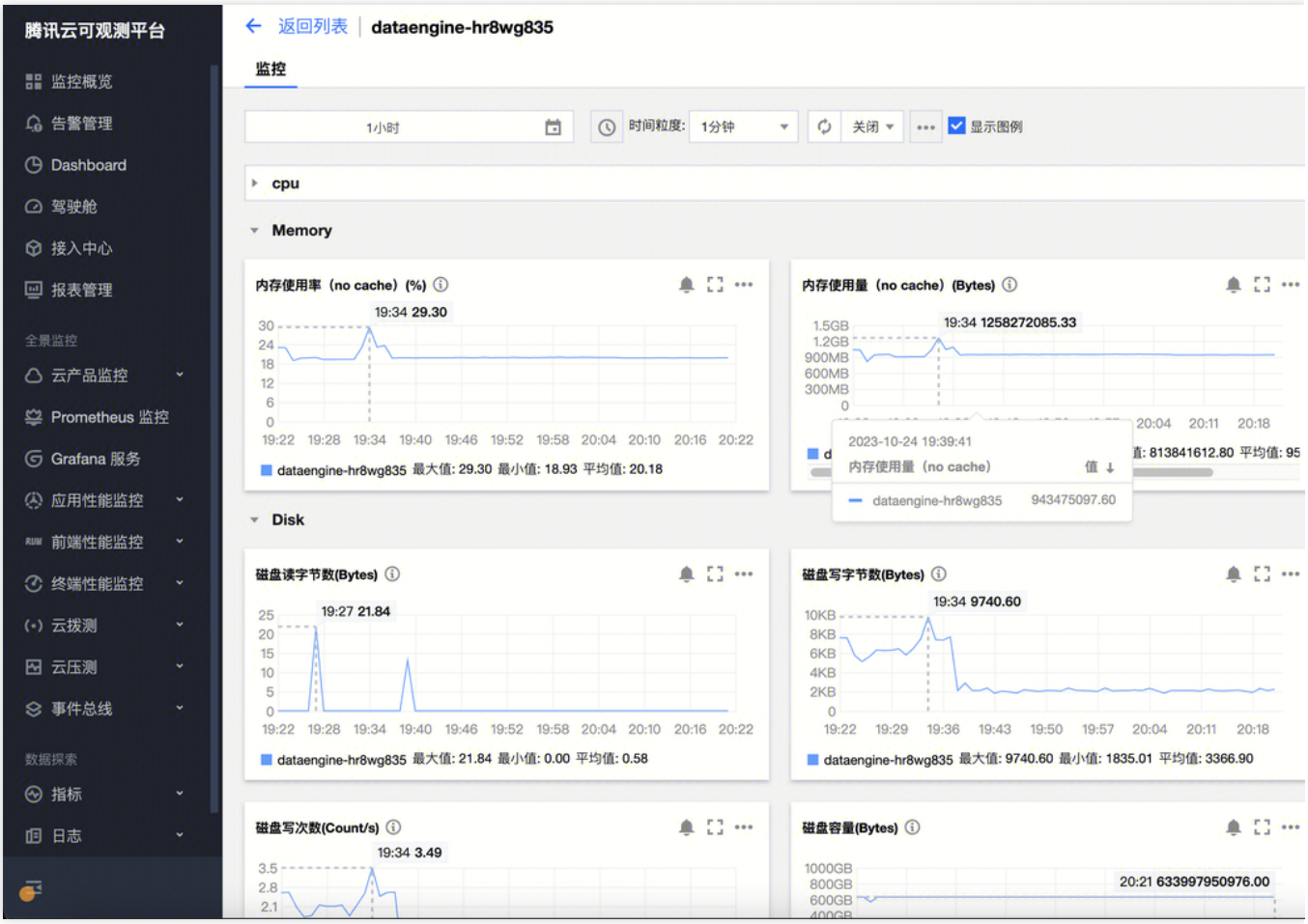
监控入口

入口一：数据湖计算 DLC 控制台

1. 登录数据引擎 > 标准引擎 页面，择服务地域。
2. 选择标准引擎，在接入点处单击**监控**，即可进入监控数据展示界面。


标准引擎
广州
创建资源
订单查询
续费管理
chasectest
test11
default
rickytest
引擎网络
网段: 10.255.0.0/22
接入点
监控
重启
销毁
规格配置
服务访问
接入点名称
apay_mark050
资源ID
DataEngine-hr8wg835
规格
16CU
状态
运行
到期时间
202
:10
终端节点名称
2123
123123
sdf
vpc
引擎列表
引擎类型
全部
运行状态
全部
付费类型
全部
资源名称/ID
引擎类型
内核版本
运行状态
付费类型
是否自...
启停策略
集群描述

监控数据展示界面如下：



配置入口：腾讯云可观测平台

1. 登录 腾讯云可观测平台控制台，登录账号需具备相关权限。
2. 左侧菜单选择云产品监控，进入告警管理的策略管理页面，选择数据湖计算选择对应的接入点网关引擎。

接入点网关引擎监控配置类型

新建告警策略

1. 数据湖计算 DLC 接入点网关支持告警能力，进入腾讯云可观测平台，单击**告警管理**，选择策略管理页。
2. 单击**新建策略**，策略类型选择“数据湖计算”。接入点网关支持三种维度的告警，其中：

“网关”告警维度为：appid/gatewayid。

“网关（多维）”告警维度为：appid/gatewayid/instanceid。

“网关引擎（多维）”告警维度为：appid/gatewayid/engineid/processid。

名称	支持维度	优势及适用场景
网关（多维）	支持：CPU、内存、磁盘、网络细粒度告警。 例如：对接入点网关的 CPU 利用率进行告警配置，可以选择配置某个接入点网关下的一个、多个实例或者任意一个实例节点触发阈值就告警。	告警支持的维度更多，告警方式更加灵活。 基础指标推荐使用此方式。
网关	主要针对当前网关整体负载情况进行监控，对基础指标按照接入点网关节点进行聚合，同时支持服务级别的指标告警。	支持 Dashboard。适用于单节点接入点网关或者需要对服务指标告警。

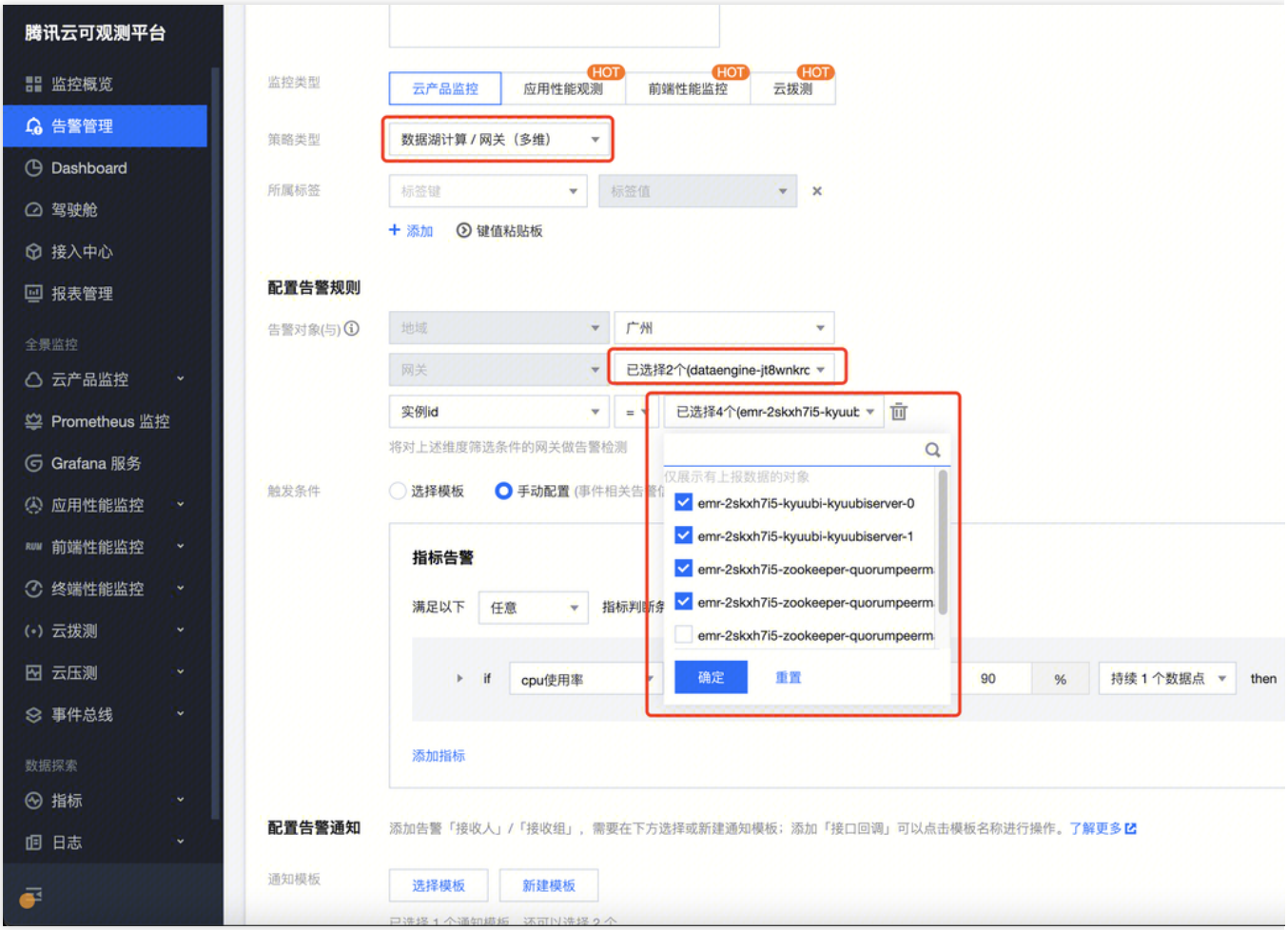
	例如：execute_statement_num（执行 statement 数量），opened_operation_num（打开的 operation 数量），launch_engine_num（启动的引擎数量），engine_process_thread_num（engine 启动的线程数）。	
网关引擎（多维）	网关引擎指接入点网关对启动 DLC 引擎的进程进行监控告警。 例如： engine_process_thread_num（engine 启动的线程数），主要针对当前接入点网关启动的引擎的进程信息进行监控	支持细粒度告警，例如：常配置某个接入点网关 ID 下任意一个引擎的进程数达到阈值就告警。适用于对接入点网关启动的进程指标告警。

配置监控项示例

网关（多维）配置

示例1：配置某个接入点网关接入点id的任意一个实例的 CPU 使用率达到90%或者内存使用率达到80%告警，配置如下图所示：

示例2：要配置某个接入点网关id的其中一个或者多个实例的 CPU 使用率达到90%告警（这种配置方式只能对选中的实例进行告警，集群重建后需要重新配置），配置如下图所示：



网关配置

示例1：配置某个接入点网关 ID 中的执行的 statement 数量大于100告警，配置如下图所示：

网关引擎（多维）

示例1：配置某个接入点网关 ID 中启动的任意引擎进程中线程数量达到500告警，配置如下图所示：

示例2：配置某个接入点网关 ID，某个引擎中任意一个进程的线程数达到500告警，配置如下图所示：

腾讯云可观测平台
监控概览
告警管理
Dashboard
驾驶舱
接入中心
报表管理
全景监控
云产品监控
Prometheus 监控
Grafana 服务
应用性能监控
前端性能监控
终端性能监控
云拨测
云压测
事件总线
数据探索
指标
日志
监控类型
云产品监控
应用性能观测
前端性能监控
云拨测
HOT
策略类型
数据湖计算 / 网关引擎（多维）
所属标签
标签键
标签值
添加
键值粘贴板
配置告警规则
告警对象(与)
地域
广州
网关
已选择1个(dataengine-jt8wnkrc
引擎id
=
c669e4bc_42da_4e09_94a1_bi
将对上述维度筛选条件的网关引擎做告警检测
触发条件
选择模板
手动配置 (事件相关告警信息暂不支持通过触发条件模板配置)
指标告警
满足以下
任意
指标判断条件时，触发告警
启用告警分级功能
if
engine启动的线...
统计粒度1分钟
>
500
Count
持续 1 个数据点
添加指标
配置告警通知
添加告警「接收人」/「接收组」，需要在下方选择或新建通知模板；添加「接口回调」可以点击模板名称进行操作。了解更多
通知模板
选择模板
新建模板

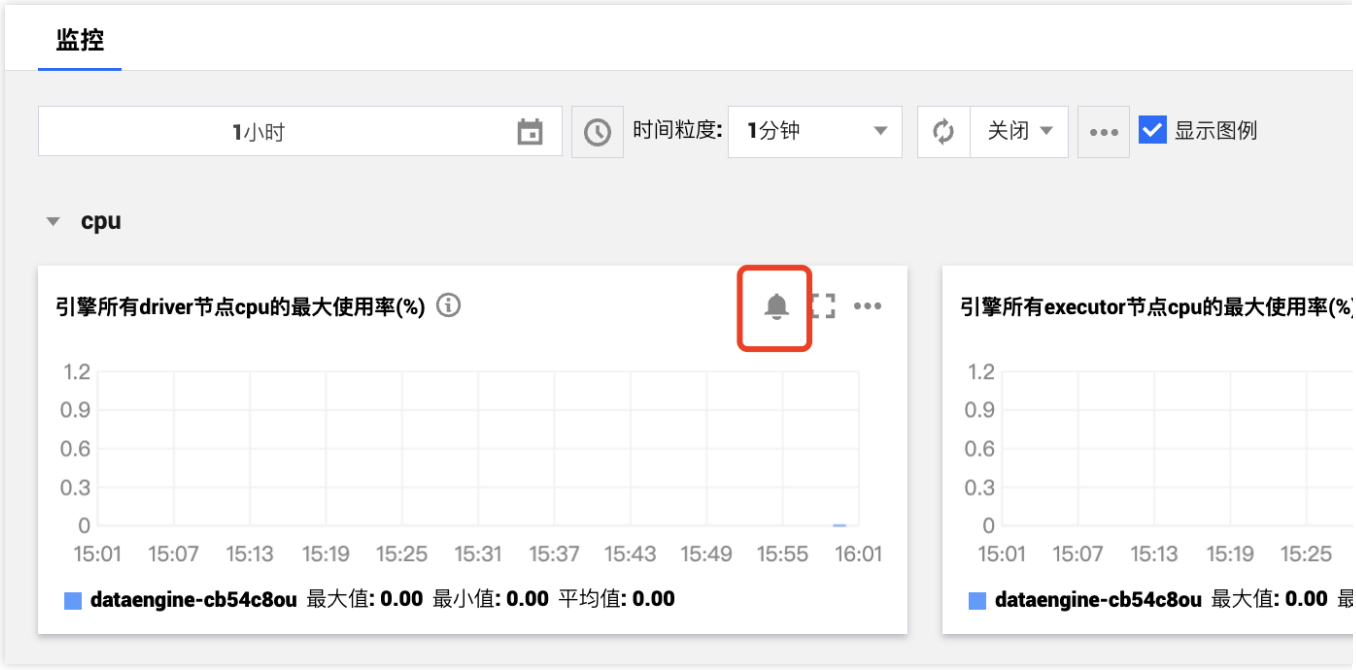
监控告警配置

最近更新时间：2024-07-31 17:32:23

配置新告警策略

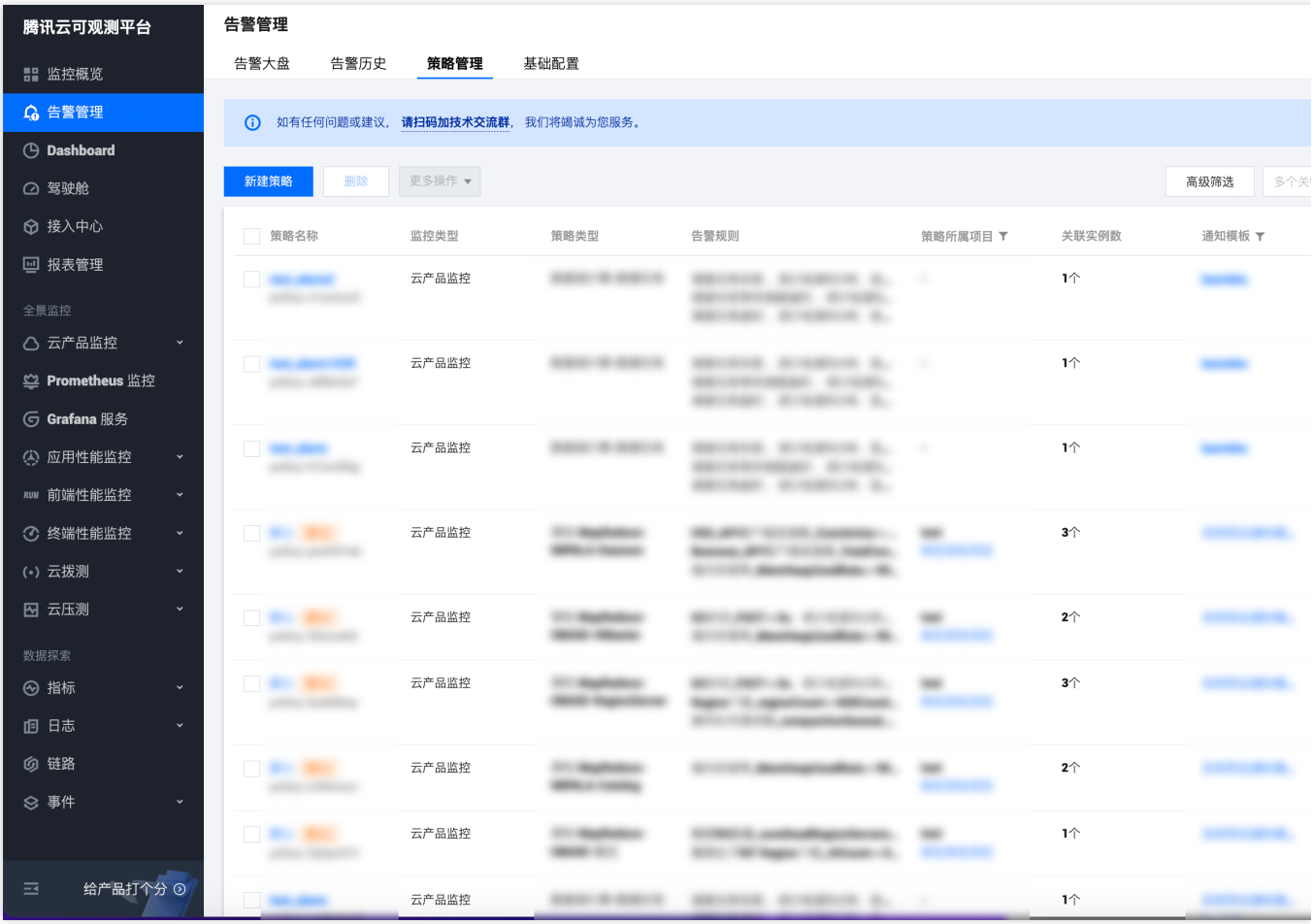
支持对特定的指标配置监控告警，可前往新建告警策略配置告警内容。

或单击需要配置告警的监控内容进入配置页面，即可配置告警内容。



管理告警策略

如需管理已配置的告警策略，可通过策略管理页面进行配置管理。



配置说明

配置项	配置说明
策略名称	告警策略的名称，最多60个字符
备注	告警策略的备注信息，最多100个字符
监控类型	请选择云产品监控
策略类型	请选择数据湖计算 DLC
策略标签	支持通过标签配置管理策略内容，需具备相关权限才可进行操作
告警对象	可针对实例 ID（支持多选）、分组实例和全部实例进行告警配置
告警配置模板	支持选择模板或手动配置，模板需管理员提前创建后使用，支持配置多个告警规则
通知模板	支持创建或选择已有的通知模板，支持配置3个模板

版权所有：腾讯云计算（北京）有限责任公司

版权所有：腾讯云计算（北京）有限责任公司

查询脚本管理

最近更新时间：2022-08-16 09:42:00

为方便用户快速处理重复的查询工作任务，数据湖计算 DLC 为用户提供了脚本文件管理。

> 说明：
> 控制台中允许保存最多100个 SQL 脚本。

新建脚本

1. 登录 数据湖计算 DLC 控制台，进入查询分析页面，单击**资源包菜单**。

2. 在资源包中，鼠标悬停脚本文件行，然后单击　　图标，单击**新建脚本**进入新建流程。

3. 填写脚本内容后，即可保存完成新建。

- 脚本名称：支持中文、英文及“_”，最多25个字符。
- 脚本描述：不超过2048个字符。
- SQL 语句：支持标准 SQL 语句，SQL语句不超过1000个字符，大小不超过2MB。

执行脚本

1. 登录 数据湖计算 DLC 控制台，进入查询分析页面，单击**资源包菜单**。

2. 在资源包中，鼠标悬停到需要执行的脚本名称，单击　　图标，单击**复制脚本到 SQL** 按钮。
3. 选择计算引擎后，单击**运行**即可执行脚本。

查看脚本信息

1. 登录 数据湖计算 DLC 控制台，进入查询分析页面，单击**资源包菜单**。

2. 在资源包中，鼠标悬停到需要查看的脚本名称，点击　　图标，单击**查看详情**按钮，即可查看脚本详情信息。

删除脚本

1. 登录 数据湖计算 DLC 控制台，进入查询分析页面，单击**资源包菜单**。

2. 在资源包中，鼠标悬停到需要查看的脚本名称，点击　　图标，单击**删除**按钮，即可删除脚本详情信息。

注意：
脚本删除后将无法恢复，请谨慎操作。

系统约束
元数据信息

最近更新时间：2024-07-17 18:17:37

数据库、数据表、属性列、分区的个数。

项目	最大数量
每个账户的数据库数量	1,000
每个账户的数据表数量	10,000
每个数据库的表数	4,096
每个数据表的列数	4,096
每个表的分区数	10,000
每个主账户的分区数	1,000,000
每张表字段数量上限	4096
每个账户的自定义函数数量	100
可创建 catalog 数量	20

数据库

名称：127个字符，同一个数据链接下，不允许有相同的数据库名称。
描述：2048 个字符。
外表的数据地址（COS 地址）：888个字符（COS 路径长度限制）。
参数：`Map<string:string>` 的形式，每个参数的长度限制为127个字符，总长度限制为3000个字符。

数据表/视图

名称：127个字符，同一个数据库下，不允许有相同的数据表名称。
描述：1000个字符。
外表的数据地址（COS 地址）：888个字符（COS 路径长度限制）。
参数：`Map<string:string>` 的形式，每个参数的长度限制为127个字符，总长度限制为512000个字符。

属性列

名称：127个字符，同一个数据表下，不允许有相同的属性列名称。
描述：256个字符。

版权所有：腾讯云计算（北京）有限责任公司

字段对应的值：131072个字符。若超过将无法创建。

分区

分区字段名称：127个字符。

计算任务

最近更新时间：2022-08-16 09:42:00

- 单条 SQL 语句大小限制为2MB。