



数据湖探索 (DLI)

用户操作手册

天翼云科技有限公司

目 录

1 数据湖探索简介	5
2 权限管理	6
2.1 权限管理概述	6
2.2 创建 IAM 用户并授权使用 DLI	9
2.3 DLI 自定义策略.....	11
2.4 DLI 资源.....	16
2.5 DLI 请求条件.....	16
2.6 常用操作与系统权限关系	17
3 入门操作指导	21
3.1 创建并提交 Spark SQL 作业.....	21
3.2 使用 TPC-H 样例模板开发并提交 Spark SQL 作业	23
3.3 创建并提交 Spark Jar 作业	24
3.4 创建并提交 Flink SQL 作业	26
4 DLI 控制台总览	34
5 SQL 编辑器	36
6 作业管理	42
6.1 SQL 作业管理.....	42
6.2 Flink 作业管理.....	46
6.2.1 Flink 作业管理概述.....	46
6.2.2 Flink 作业权限管理	49
6.2.3 准备 Flink 作业数据	51
6.2.4 创建 Flink SQL 作业	52
6.2.5 创建 Flink Jar 作业	56
6.2.6 调试 Flink 作业.....	60
6.2.7 操作 Flink 作业.....	61
6.2.8 Flink 作业详情.....	66
6.3 Spark 作业管理	71
6.3.1 Spark 作业管理概述	71
6.3.2 创建 Spark 作业.....	73

7 队列管理	77
7.1 队列管理概述	77
7.2 队列权限管理	80
7.3 创建队列	81
7.4 删除队列	83
7.5 修改队列网段	83
7.6 弹性扩缩容	83
7.7 弹性扩缩容定时任务	85
7.8 测试地址连通性	87
7.9 创建消息通知主题	87
8 数据管理	89
8.1 库表管理	89
8.1.1 库表管理概述	89
8.1.2 数据库权限管理	91
8.1.3 表权限管理	95
8.1.4 创建数据库和表	99
8.1.5 删除数据库和表	104
8.1.6 修改数据库和表所有者	104
8.1.7 导入数据	105
8.1.8 将 DLI 数据导出至 OBS	108
8.1.9 查看元数据	109
8.1.10 预览数据	110
8.2 程序包管理	110
8.2.1 程序包管理概述	110
8.2.2 程序包组和程序包权限管理	111
8.2.3 创建程序包	113
8.2.4 删除程序包	114
8.2.5 修改所有者	115
8.2.6 内置依赖包	115
9 作业模板	128
9.1 SQL 模板管理	128
9.2 Flink 模板管理	132
9.3 附录	136
9.3.1 SQL 模板下 TPC-H 样例数据说明	136
10 跨源连接	138
10.1 跨源连接和跨源分析概述	138
10.2 增强型跨源连接	142
10.2.1 增强型跨源连接概述	142

10.2.2 创建/查找/删除增强型跨源连接	143
10.2.3 绑定/解绑队列	145
10.2.4 修改主机信息	147
10.2.5 自定义路由信息	148
10.2.6 增强型跨源连接权限管理	148
10.3 跨源认证权限管理	149
10.4 创建及管理跨源认证	150
11 全局配置.....	155
11.1 全局变量	155
11.2 服务授权	156
12 DLI 使用 UDF 操作指导.....	158

1 数据湖探索简介

什么是 DLI

数据湖探索（Data Lake Insight，简称 DLI）是完全兼容 [Apache Spark](#)、[Apache Flink](#) 生态，提供一站式的流处理、批处理、交互式分析的 Serverless 融合处理分析服务。用户不需要管理任何服务器，即开即用。支持标准 SQL/Spark SQL/Flink SQL，支持多种接入方式，并兼容主流数据格式。数据无需复杂的抽取、转换、加载，使用 SQL 或程序就可以对云上 CloudTable、RDS、DWS、CSS、OBS、ECS 自建数据库以及线下数据库的异构数据进行探索。

功能优势

- 纯 SQL 操作
DLI 提供标准 SQL 接口，用户仅需使用 SQL 便可实现海量数据查询分析。
- 存算分离
DLI 的存储和计算解耦，降低成本的同时，提高了资源利用率。
- 免运维、高可用
用户无需感知底层运维、升级，跨 AZ 高可用，跨 AZ 双活。

DLI 核心引擎：Spark+Flink

- Spark 是用于大规模数据处理的统一分析引擎，聚焦于查询计算分析。DLI 在开源 Spark 基础上进行了大量的性能优化与服务化改造，不仅兼容 Apache Spark 生态和接口，性能较开源提升了 2.5 倍，在小时级即可实现 EB 级数据查询分析。
- Flink 是一款分布式的计算引擎，可以用来做批处理，即处理静态的数据集、历史的数据集；也可以用来做流处理，即实时地处理一些实时数据流，实时地产生数据的结果。DLI 在开源 Flink 基础上进行了特性增强和安全增强，提供了数据处理所必须的 Stream SQL 特性。

DLI 服务架构：Serverless

DLI 是无服务器化的大数据查询分析服务，其优势在于：

- 自动扩缩容：根据业务负载，对计算资源进行预估和自动扩缩容。

2 权限管理

2.1 权限管理概述

DLI 服务不仅在服务本身有一套完善的权限控制机制，同时还支持通过统一身份认证服务（Identity and Access Management，简称 IAM）细粒度鉴权，可以通过在 IAM 创建策略来管理 DLI 的权限控制。两种权限控制机制可以共同使用，没有冲突。

IAM 鉴权使用场景

企业用户在云上使用 DLI 服务时，需要对不同部门的员工使用 DLI 资源（队列）进行管理，包括资源的创建、删除、使用、隔离等。同时，也需要对不同部门的数据进行管理，包括数据的隔离、共享等。

DLI 使用 IAM 进行精细的企业级多租户管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全地控制云资源的访问。

通过 IAM，您可以在云帐号中给员工创建 IAM 用户，并使用策略来控制他们对云资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有 DLI 的使用权限，但是不希望他们拥有删除 DLI 等高危操作的权限，那么您可以使用 IAM 为开发人员创建用户，通过授予仅能使用 DLI，但是不允许删除 DLI 的权限策略，控制他们对 DLI 资源的使用范围。

说明

对于新建的用户，需要先登录一次 DLI，记录元数据，后续才可正常使用。

IAM 是云提供权限管理的基础服务，无需付费即可使用，您只需要为您帐号中的资源进行付费。

如果云帐号已经能满足您的需求，不需要创建独立的 IAM 用户进行权限管理，您可以跳过本章节，不影响您使用 DLI 服务的其他功能。

DLI 系统权限

如表 2-1 所示，包括了 DLI 的所有系统权限。

权限类别：根据授权精程度分为角色和策略。

- **角色**：IAM 最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于云各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。
- **策略**：IAM 最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。例如：针对 DLI 服务，管理员能够控制 IAM 用户仅能对某一类云服务器资源进行指定的管理操作。

表2-1 DLI 系统权限

系统角色/策略名称	描述	类别	授权方式
DLI FullAccess	数据湖探索所有权限。	系统策略	具体的授权方式请参考 2.2 创建 IAM 用户并授权使用 DLI。
DLI ReadOnlyAccess	数据湖探索只读权限。	系统策略	
Tenant Administrator	租户管理员。 <ul style="list-style-type: none"> • 操作权限：具有所有云服务的管理和使用权限。创建后，可通过 ACL 赋权给其他子用户使用。 • 作用范围：项目级服务。 	系统角色	
DLI Service Admin	DLI 服务管理员。 <ul style="list-style-type: none"> • 操作权限：具有数据湖探索服务队列、数据的管理和使用权限。创建后，可通过 ACL 赋权给其他子用户使用。 • 作用范围：项目级服务。 	系统角色	

DLI 权限分类

DLI 服务权限分类如表 2-2 所示，其可控制的资源请参考表 2-7。

表2-2 DLI 权限分类

权限大类	权限小类	控制台操作	SQL 语法	API 定义

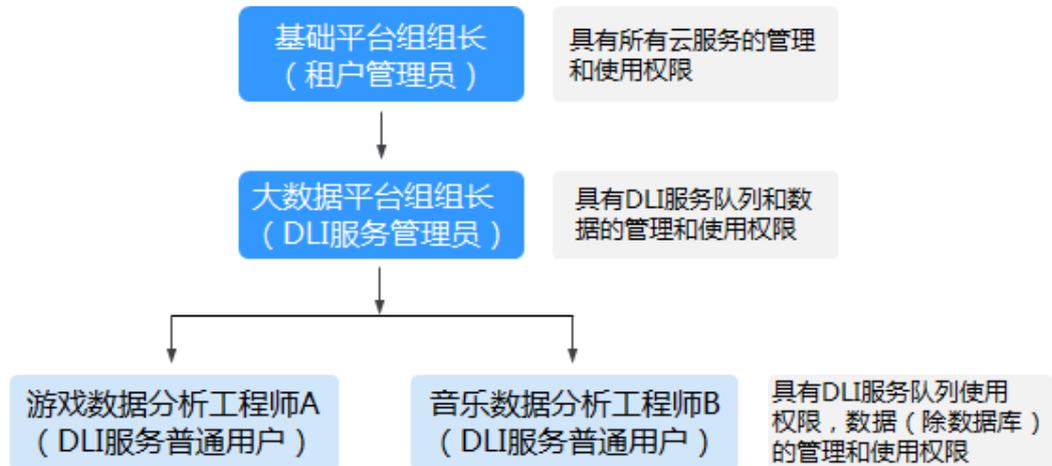
权限大类	权限小类	控制台操作	SQL 语法	API 定义
队列权限	队列管理权限	请参考 7.2 队列权限管理	无	请参考《数据湖探索 API 参考》中的“权限相关 API > 队列赋权”章节。
	队列使用权限			
数据权限	数据库权限	请参考 8.1.2 数据库权限管理和 8.1.3 表权限管理	请参考《数据湖探索 SQL 语法参考》中的“批作业 SQL 语法 > 数据权限管理 > 数据权限列表”章节。	请参考《数据湖探索 API 参考》中的“权限相关 API > 数据赋权（用户）”章节。
	表权限			
	列权限			
作业权限	Flink 作业	请参考 6.2.2 Flink 作业权限管理	无	请参考《数据湖探索 API 参考》中的“权限相关 API > 数据赋权（用户）”章节。
程序包权限	程序包组权限	请参考 8.2.2 程序包组和程序包权限管理	无	请参考《数据湖探索 API 参考》中的“权限相关 API > 数据赋权（用户）”章节。
	程序包权限			
跨源认证权限	跨源认证权限	请参考 10.3 跨源认证权限管理	无	请参考《数据湖探索 API 参考》中的“权限相关 API > 数据赋权（用户）”章节。

场景举例

某互联网公司，主要有游戏和音乐两大业务，使用 DLI 服务进行用户行为分析，辅助决策。

如图 2-1 所示，“基础平台组组长”在云上申请了一个“租户管理员”（Tenant Administrator）帐号，用于管理和使用云的各个服务。因为“大数据平台组”需要使用 DLI 进行数据分析，所有“基础平台组组长”增加了一个权限为“DLI 服务管理员”（DLI Service Admin）的子帐号用于管理和使用 DLI 服务。“基础平台组组长”按照公司两个业务对于数据分析的要求，创建了“队列 A”分配给“数据工程师 A”运行游戏数据分析业务，“队列 B”分配给“数据工程师 B”运行音乐数据分析业务，并分别赋予“DLI 普通用户”权限，具有队列使用权限，数据（除数据库）的管理和使用权限。

图2-1 权限分配



“数据工程师 A”创建了一个 gameTable 表用于存放游戏道具相关数据，userTable 表用于存放游戏用户相关数据。因为音乐业务是一个新业务，想在存量的游戏用户中挖掘一些潜在的音乐用户，所以“数据工程师 A”把 userTable 表的查询权限赋给了“数据工程师 B”。同时，“数据工程师 B”创建了一个 musicTable 用于存放音乐版权相关数据。

“数据工程师 A”和“数据工程师 B”对于队列和数据的使用权限如表 2-3 所示。

表2-3 使用权限说明

用户	数据工程师 A（游戏数据分析）	数据工程师 B（音乐数据分析）
队列	队列 A（队列使用权限）	队列 B（队列使用权限）
数据（表）	gameTable（表管理和使用权限）	musicTable（表管理和使用权限）
	userTable（表管理和使用权限）	userTable（表查询权限）

说明

队列的使用权限包括提交作业和终止作业两个权限。

2.2 创建 IAM 用户并授权使用 DLI

如果您需要对您所拥有的 DLI 资源进行精细的权限管理，您可以使用统一身份认证服务（Identity and Access Management，简称 IAM），具体 IAM 使用场景可以参考 2.1 权限管理概述。

如果云帐号已经能满足您的要求，不需要创建独立的 IAM 用户，您可以跳过本章节，不影响您使用 DLI 服务的其它功能。

本章节介绍创建 IAM 用户并授权使用 DLI 的方法，操作流程如图 2-2 所示。

前提条件

给用户组授权之前，请您先了解用户组可以添加的 DLI 权限，并结合实际需求进行选择。DLI 支持的系统权限，请参见：[DLI 系统权限](#)。

示例流程

图2-2 给用户授权 DLI 权限流程



1. 创建用户组并授权。

在 IAM 控制台创建用户组，并授予 DLI 服务普通用户权限“DLI ReadOnlyAccess”。

2. 创建用户并加入用户组。

在 IAM 控制台创建用户，并将其加入 1 中创建的用户组。

3. 用户登录并验证权限。

使用新创建的用户登录控制台，切换至授权区域，验证权限：

- 在“服务列表”中选择数据湖探索，进入 DLI 主界面。如果在“队列管理”页面可以查看队列列表，但是单击右上角“购买队列”，无法购买 DLI 队列（假设当前权限仅包含 DLI ReadOnlyAccess），表示“DLI ReadOnlyAccess”已生效。
- 在“服务列表”中选择除数据湖探索外（假设当前策略仅包含 DLI ReadOnlyAccess）的任一服务，若提示权限不足，表示“DLI ReadOnlyAccess”已生效。

2.3 DLI 自定义策略

如果系统预置的 DLI 权限，不满足您的授权要求，可以创建自定义策略。自定义策略中可以添加的授权项（Action）请参考《数据湖探索 API 参考》中的“权限策略和授权项”。

目前云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。
- JSON 视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写 JSON 格式的策略内容。

本章为您介绍常用的 DLI 自定义策略样例。

策略字段介绍

以授权用户拥有在所有区域中所有数据库的创建表权限为例进行说明：

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:create_table"
      ],
      "Resource": [
        "dli:*:*:database:*"
      ]
    }
  ]
}
```

- **Version**
版本信息，1.1：策略。IAM 最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。
- **Effect**
作用。包含两类：允许（Allow）和拒绝（Deny），既有 Allow 又有 Deny 的授权语句时，遵循 Deny 优先的原则。
- **Action**
授权项，指对资源的具体操作权限，不超过 100 个。

说明

- 格式为：服务名:资源类型:操作，例：dli:queue:submit_job。
- 服务名为产品名称，例如 dli、evs 和 vpc 等，服务名仅支持小写。资源类型和操作没有大小写，要求支持通配符号*，无需罗列全部授权项。
- 资源类型可以参考表 2-7 中的资源类型。
- 操作：操作以 IAM 服务中已经注册的 action 为准。
- **Condition**

限制条件：使策略生效的特定条件，包括条件键和运算符。

条件键表示策略语句的 **Condition** 元素中的键值，分为全局级条件键和服务级条件键。

- 全局级条件键（前缀为 **g:**）适用于所有操作。详细请参考策略语法中的条件键说明。
- 服务级条件键（前缀为服务缩写，如 **dli:**，目前 **DLI** 未提供服务级条件键）仅适用于对应服务的操作。

运算符与条件键一起使用，构成完整的条件判断语句。具体内容请参考表 2-4。

DLI 通过 **IAM** 预置了一组条件键，例如，您可以先使用 **hw:SourceIp** 条件键检查请求者的 IP 地址，然后再允许执行操作。下表显示了适用于 **DLI** 服务特定的条件键。

表2-4 DLI 请求条件

DLI 条件键	运算符	描述
g:CurrentTime	Date and time	接收到鉴权请求的时间。 说明 以“ISO 8601”格式表示，例如：2012-11-11T23:59:59Z。
g:MFAPresent	Boolean	用户登录时是否使用了多因素认证。
g:UserId	String	当前登录的用户 ID。
g:UserName	String	当前登录的用户名。
g:ProjectName	String	当前登录的 Project。
g:DomainName	String	当前登录的 Domain。

- **Resource**

格式为：服务名:region:domainId:资源类型:资源路径，通配符号*表示所有，资源类型和资源路径可以参考表 2-7。

示例：

"dli:*:*:queue:*": 表示所有的队列。

创建 DLI 自定义策略

用户可以根据场景设置不同级别的 **Action** 和 **Resource**。

1. 定义 Action

Action 由服务名:资源类型:操作三段组成，通配符为*。例如：

表2-5 Action

Action	说明
--------	----

Action	说明
dli:queue:submit_job	DLI 队列的提交操作
dli:queue:*	DLI 队列的全部操作
dli:*:*	DLI 所有资源类型的所有操作

更多操作与系统权限的关系请参考 2.6 常用操作与系统权限关系。

2. 定义 Resource

Resource 由<服务名:region:domainId:表 2-7>5 个字段组成，通配符号*表示所有资源。5 个字段可以灵活设置，资源路径可以按照场景需要，设置不同级别的权限控制。当需要设置该服务下的所有资源时，可以不指定该字段。Resource 定义请参考表 2-6。Resource 中的资源类型和资源路径请参考表 2-7。

表2-6 Resource

Resource	说明
DLI:*:*:table:databases.dbname.tables.*	DLI 服务，任意 region，任意帐号 ID 下，数据库名为 dbname 下的所有表资源。
DLI:*:*:database:databases.dbname	DLI 服务，任意 region，任意帐号 ID 下，数据库名为 dbname 的队列。
DLI:*:*:queue:queues.*	DLI 服务，任意 region，任意帐号 ID 下，任意队列资源。
DLI:*:*:jobs:jobs.flink.1	DLI 服务，任意 region，任意帐号 ID 下，作业 Id 为 1 的 flink 作业。

表2-7 DLI 的指定资源与对应路径

资源类型	资源名称	资源路径
queue	DLI 队列	queues.queueName
database	DLI 数据库	databases.dbname
table	DLI 表	databases.dbname.tables.tbname
column	DLI 列	databases.dbname.tables.tbname.columns.colname
jobs	DLI Flink 作业	jobs.flink.jobid
resource	DLI 程序包	resources.resourcename
group	DLI 程序包组	groups.groupname
datasourceauth	DLI 跨源认证信	datasourceauth.name

资源类型	资源名称	资源路径
h	息	

3. 将上述的所有字段拼接为一个 json 就是一个完整的策略了，其中 `action` 和 `resource` 均可以设置多个，当然也可以通过 IAM 提供的可视化界面进行创建，例如：
授权用户拥有 DLI 服务，任意 region，任意帐号 ID 下，任意数据库的创建删除权限，任意队列的提交作业权限，任意表的删除权限。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:create_database",
        "dli:database:drop_database",
        "dli:queue:submit_job",
        "dli:table:drop_table"
      ],
      "Resource": [
        "dli:*:*:database:*",
        "dli:*:*:queue:*",
        "dli:*:*:table:*"
      ]
    }
  ]
}
```

DLI 自定义策略样例

- 示例 1：允许
 - 授权用户拥有在所有区域中所有数据库的创建表权限

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:database:create table"
      ],
      "Resource": [
        "dli:*:*:database:*"
      ]
    }
  ]
}
```

- 授权用户拥有在数据库 db 中表 tb 中列 col 的查询权限

```
{
  "Version": "1.1",
  "Statement": [
    {
```

```
    "Effect": "Allow",
    "Action": [
        "dli:column:select"
    ],
    "Resource": [
        "dli:*:*:column:databases.db.tables.tb.columns.col"
    ]
}
]
```

- 示例 2: 拒绝

拒绝策略需要同时配合其他策略使用，否则没有实际作用。用户被授予的策略中，一个授权项的作用如果同时存在 **Allow** 和 **Deny**，则遵循 **Deny** 优先。

- 授权用户不能创建数据库，删除数据库，提交作业（default 队列除外），删除表

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "dli:database:create_database",
        "dli:database:drop_database",
        "dli:queue:submit_job",
        "dli:table:drop_table"
      ],
      "Resource": [
        "dli:*:*:database:*",
        "dli:*:*:queue:*",
        "dli:*:*:table:*"
      ]
    }
  ]
}
```

- 授权用户不能在队列为 **demo** 的队列上提交作业

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "dli:queue:submit_job"
      ],
      "Resource": [
        "dli:*:*:queue:queues.demo"
      ]
    }
  ]
}
```

2.4 DLI 资源

资源是服务中存在的对象。在 DLI 中，资源如下，您可以在创建自定义策略时，通过指定资源路径来选择特定资源。

表2-8 DLI 的指定资源与对应路径

资源类型	资源名称	资源路径
queue	DLI 队列	queues.queueName
database	DLI 数据库	databases.dbName
table	DLI 表	databases.dbName.tables.tbName
column	DLI 列	databases.dbName.tables.tbName.columns.colName
jobs	DLI Flink 作业	jobs.flink.jobId
resource	DLI 程序包	resources.resourceName
group	DLI 程序包组	groups.groupName
datasourceauth	DLI 跨源认证信息	datasourceauth.name

2.5 DLI 请求条件

您可以在创建自定义策略时，通过添加“请求条件”（Condition 元素）来控制策略何时生效。请求条件包括条件键和运算符，条件键表示策略语句的 Condition 元素，分为全局级条件键和服务级条件键。全局级条件键（前缀为 g:）适用于所有操作，服务级条件键（前缀为服务缩写，如 dli）仅适用于对应服务的操作。运算符与条件键一起使用，构成完整的条件判断语句。

DLI 通过 IAM 预置了一组条件键，例如，您可以先使用 hw:SourceIp 条件键检查请求者的 IP 地址，然后再允许执行操作。下表显示了适用于 DLI 服务特定的条件键。

表2-9 DLI 请求条件

DLI 条件键	运算符	描述
g:CurrentTime	Date and time	接收到鉴权请求的时间。 说明 以“ISO 8601”格式表示，例如：2012-11-11T23:59:59Z。
g:MFAPresent	Boolean	用户登录时是否使用了多因素认证。
g:UserId	String	当前登录的用户 ID。

DLI 条件键	运算符	描述
g:UserName	String	当前登录的用户名。
g:ProjectName	String	当前登录的 Project。
g:DomainName	String	当前登录的 Domain。

2.6 常用操作与系统权限关系

表 2-10 列出了 DLI SQL 常用操作与系统策略的授权关系，您可以参照该表选择合适的系统策略。更多 SQL 语法赋权请参考《数据湖探索 SQL 语法参考》中的“批作业 SQL 语法 > 数据权限管理 > 数据权限列表”章节。

表2-10 DLI 常用操作与系统权限的关系

资源	操作	说明	DLI FullAccess	DLI ReadOnlyAccess	T	D
队列	DROP_QUEUE	删除队列	√	×	√	√
	SUBMIT_JOB	提交作业	√	×	√	√
	CANCEL_JOB	终止作业	√	×	√	√
	RESTART	重启队列	√	×	√	√
	GRANT_PRIVILEGE	队列的赋权	√	×	√	√
	REVOKE_PRIVILEGE	队列权限的回收	√	×	√	√

资源	操作	说明	DLI FullAccess	DLI ReadOnlyAccess	T	D
	SHOW_PRIVILEGES	查看其他用户具备的队列权限	√	×	√	√
数据库	DROP_DATABASE	删除数据库	√	×	√	√
	CREATE_TABLE	创建表	√	×	√	√
	CREATE_VIEW	创建视图	√	×	√	√
	EXPLAIN	将 SQL 语句解释为执行计划	√	×	√	√
	CREATE_ROLE	创建角色	√	×	√	√
	DROP_ROLE	删除角色	√	×	√	√
	SHOW_ROLES	显示角色	√	×	√	√
	GRANT_ROLE	绑定角色	√	×	√	√
	REVOKE_ROLE	解除角色绑定	√	×	√	√
	SHOW_USERS	显示所有角色和用户的绑定关系	√	×	√	√
	GRANT_PRIVILEGE	数据库的赋权	√	×	√	√
	REVOKE_PRIVILEGE	数据库权限的回收	√	×	√	√
	SHOW_PRIVILEGES	查看其他用户具备的数据库权限	√	×	√	√

资源	操作	说明	DLI FullAccess	DLI ReadOnlyAccess	T	D
	DISPLAY_ALL_TABLES	显示数据库中的表信息	√	×	√	√
	DISPLAY_DATABASE	显示数据库信息	√	×	√	√
	CREATE_FUNCTION	创建函数	√	×	√	√
	DROP_FUNCTION	删除函数	√	×	√	√
	SHOW_FUNCTIONS	显示所有函数	√	×	√	√
	DESCRIBE_FUNCTION	显示函数详情	√	×	√	√
表	DROP_TABLE	删除表	√	×	√	√
	SELECT	查询表	√	×	√	√
	INSERT_INTO_TABLE	插入	√	×	√	√
	ALTER_TABLE_ADD_COLUMNS	添加列	√	×	√	√
	INSERT_OVERWRITE_TABLE	重写	√	×	√	√
	ALTER_TABLE_RENAME	重命名表	√	×	√	√
	ALTER_TABLE_	在分区表中添加分	√	×	√	√

资源	操作	说明	DLI FullAccess	DLI ReadOnlyAccess	T	D
	ADD_PARTITION	区				
	ALTER_TABLE_RENAME_PARTITION	重命名表分区	√	×	√	√
	ALTER_TABLE_DROP_PARTITION	删除分区表的分区	√	×	√	√
	SHOW_PARTITIONS	显示所有分区	√	×	√	√
	ALTER_TABLE_RECOVER_PARTITION	恢复表分区	√	×	√	√
	ALTER_TABLE_SET_LOCATION	设置分区路径	√	×	√	√
	GRANT_PRIVILEGE	表的赋权	√	×	√	√
	REVOKE_PRIVILEGE	表权限的回收	√	×	√	√
	SHOW_PRIVILEGES	查看其他用户具备的表权限	√	×	√	√
	DISPLAY_TABLE	显示表	√	√	√	√
	DESCRIBE_TABLE	显示表信息	√	×	√	√

3 入门操作指导

3.1 创建并提交 Spark SQL 作业

使用 DLI 提交 Spark SQL 作业查询数据。基本流程如下：

- 步骤 1: [登录云](#)
- 步骤 2: [上传数据至 OBS](#)
- 步骤 3: [登录 DLI 管理控制台](#)
- 步骤 4: [创建队列](#)
- 步骤 5: [创建数据库](#)
- 步骤 6: [创建表](#)
- 步骤 7: [查询数据](#)

如下操作以查询 OBS 的数据为例，DLI 的数据查询操作类同。

步骤 1：登录云

1. 打开产品首页。
2. 在登录页面输入“帐号名”和“密码”，单击“登录”。

步骤 2：上传数据至 OBS

DLI 可以查询存储在 OBS 中的数据，查询数据前，需要将数据文件上传至 OBS 中。

1. 在服务列表中，单击“存储”中的“对象存储服务 OBS”。
2. 进入 OBS 管理控制台页面。
3. 创建一个桶，桶名全局唯一，这里以桶名“obs1”为例。
 - a. 单击页面右上角“创建桶”。
 - b. 进入“创建桶”页面，输入“桶名称”。其他参数保持默认值或根据需要选择。
 - c. 单击“立即创建”。
4. 单击所建桶“obs1”，进入“概览”页面。

- 单击左侧列表中的“对象”，选择“上传对象”，将需要上传的文件“sampledata.csv”上传到指定目录，单击“上传”。

“sampledata.csv”样例文件可以通过新建“sampledata.txt”，复制如下英文逗号分隔的文本内容，再另存为“sampledata.csv”。

```
12,test
```

文件上传成功后，待分析的文件路径为“obs://obs1/sampledata.csv”。

关于 OBS 管理控制台更多操作请参考《对象存储服务控制台指南》。

OBS 上传文件指导，请参见《OBS 工具指南》。

说明

针对大文件场景，由于 OBS 管理控制台对文件大小和数量限制较多，所以推荐使用 OBS 工具上传大文件，如 OBS Browser+ 上传。

- OBS Browser+ 是一个比较常用的图形化工具，支持完善的桶管理和对象管理操作。推荐使用此工具创建桶或上传对象。

步骤 3：登录 DLI 管理控制台

- 在列表中，选择“数据湖探索 DLI”。
- 进入 DLI 管理控制台页面。第一次进入数据湖探索管理控制台需要进行授权，以获取访问 OBS 的权限。

说明：至少勾选“Tenant Administrator(全局服务)”权限。

步骤 4：创建队列

队列是使用 DLI 服务的基础，执行 SQL 作业前需要先创建队列。

- DLI 有预置的可用队列“default”。
- 用户也可根据需要自己创建队列。
 - 在 DLI 管理控制台，单击左侧导航栏中的“SQL 编辑器”，可进入 SQL 作业“SQL 编辑器”页面。

- 在左侧导航栏，选择队列页签，单击右侧的  创建队列。

创建队列详细介绍请参考《数据湖探索用户指南》>《创建队列》。

步骤 5：创建数据库

在进行数据查询之前还需要创建一个数据库，例如 db1。

说明

“default”为内置数据库，不能创建名为“default”的数据库。

- 在 DLI 管理控制台，单击左侧导航栏中的“SQL 编辑器”，可进入 SQL 作业“SQL 编辑器”页面。
- 在“SQL 编辑器”页面右侧的编辑窗口中，输入如下 SQL 语句，单击“执行”。阅读并同意隐私协议，单击“确定”。

```
create database db1;
```

数据库创建成功后，新建的数据库 db1 会在左侧“数据库”列表中出现。

📖 说明

在 DLI 管理控制台第一次单击“执行”操作时，需要阅读隐私协议，确认同意后才能执行作业，且后续“执行”操作将不会再提示阅读隐私协议。

步骤 6：创建表

数据库创建完成后，需要在数据库 db1 中基于 OBS 上的样本数据“obs://obs1/sampleddata.csv”创建一个表，例如 table1。

1. 在“SQL 编辑器”页面右侧的编辑窗口上方，选择队列“default”和数据库“db1”。
2. 在编辑窗口中，输入如下 SQL 语句，单击“执行”。

```
create table table1 (id int, name string) using csv options (path 'obs://obs1/sampleddata.csv');
```

表 table1 创建成功后，单击左侧“库表”页签，再单击 db1，新创建的表 table1 会在“表”区域下方显示。

步骤 7：查询数据

完成以上步骤后，就可以开始进行数据查询了。

1. 单击“SQL 编辑器”页面左侧的“数据库”页签，选择新创建的表 table1，双击表，在右侧编辑窗口中，自动输入 SQL 查询语句，例如查询 table1 表的 1000 条数据：

```
select * from db1.table1 limit 1000;
```

2. 单击“执行”，系统开始查询。

SQL 语句执行成功后，可在 SQL 作业编辑窗口下方“查看结果”页签查看查询结果。

3.2 使用 TPC-H 样例模板开发并提交 Spark SQL 作业

为了便捷快速的执行 SQL 操作，DLI 支持定制模板或将正在使用的 SQL 语句保存为模板。保存模板后，不需编写 SQL 语句，可通过模板直接执行 SQL 操作。

当前系统提供了多条标准的 TPC-H 查询语句模板，可以根据当前需求选择使用。本样例演示通过一个 TPC-H 样例模板开发并提交 Spark SQL 作业的基本流程：

[步骤 1：登录云](#)

[步骤 2：登录 DLI 管理控制台](#)

[步骤 3：执行 TPC-H 样例模板并查看结果](#)

更多样例模板操作，请参考《数据湖探索用户指南》下的“用户指南 > 作业模板 > SQL 模板管理”。

步骤 1：登录云

1. 打开产品首页。
2. 在登录页面输入“帐号名”和“密码”，单击“登录”。

步骤 2：登录 DLI 管理控制台

1. 在列表中，选择“数据湖探索 DLI”。
2. 进入 DLI 管理控制台页面。

说明

第一次进入数据湖探索管理控制台需要进行授权，具体请参考《数据湖探索用户指南》>《全局配置》>《服务授权》。

步骤 3：执行 TPC-H 样例模板并查看结果

1. 在 DLI 管理控制台，选择“作业模板”>“SQL 模板”>“样例模板”，在“tpchQuery”下选择“Q1_价格摘要报告查询”样例模板，单击操作列的“执行”进入“SQL 编辑器”。
2. 在“SQL 编辑器”页面右侧的编辑窗口上方，“执行引擎”选择“spark”，“队列”选择“default”，“数据库”选择“default”，单击“执行”。
3. SQL 作业编辑窗口下方“查看结果”页签查看查询结果。

本示例使用系统预置的“default”队列和数据库进行演示，也可以在自建的队列和数据库下执行。

创建队列请参考《数据湖探索用户指南》>《创建队列》。创建数据库请参考《数据湖探索用户指南》下的“数据管理 > 库表管理 > 创建数据库和表”。

3.3 创建并提交 Spark Jar 作业

使用 DLI 提交 Spark 作业进行实时计算。基本流程如下：

[步骤 1：登录云](#)

[步骤 2：上传数据至 OBS](#)

[步骤 3：登录 DLI 管理控制台](#)

[步骤 4：创建队列](#)

[步骤 5：创建程序包](#)

[步骤 6：提交 Spark 作业](#)

步骤 1：登录云

使用 DLI 服务，首先要登录云。

1. 打开产品首页。
2. 在登录页面输入“帐号名”和“密码”，单击“登录”。

步骤 2：上传数据至 OBS

开发 Spark Jar 作业程序，编译并打包为“spark-examples.jar”。参考以下操作步骤上传该作业程序。

提交 Spark 作业之前，需要在 OBS 中上传数据文件。

1. 在服务列表中，单击“存储”中的“对象存储服务 OBS”。
2. 进入 OBS 管理控制台页面。
3. 创建一个桶，桶名全局唯一，这里以桶名“dli-test-obs01”为例。
 - a. 单击“创建桶”。
 - b. 进入“创建桶”页面，输入“桶名称”。其他参数保持默认值或根据需要选择。
 - c. 单击“立即创建”。
4. 单击所建桶“dli-test-obs01”，进入“概览”页面。
5. 单击左侧列表中的“对象”，选择“上传文件”，将需要上传的文件，例如“spark-examples.jar”上传到指定目录，单击“确定”。

例如，文件上传成功后，待分析的文件路径为“obs://dli-test-obs01/spark-examples.jar”。

关于 OBS 管理控制台更多操作请参考《对象存储服务控制台指南》。

OBS 上传文件指导，请参见《OBS 工具指南》。

说明

针对大文件场景，由于 OBS 管理控制台对文件大小和数量限制较多，所以推荐使用 OBS 工具上传大文件，如 OBS Browser+ 上传。

- OBS Browser+ 是一个比较常用的图形化工具，支持完善的桶管理和对象管理操作。推荐使用此工具创建桶或上传对象。

步骤 3：登录 DLI 管理控制台

使用 DLI 提交 Spark 作业，需要先进入 Spark 作业编辑页面。

1. 在列表中，选择“数据湖探索 DLI”。
2. 进入 DLI 管理控制台页面。

说明

第一次进入数据湖探索管理控制台需要进行授权，以获取访问 OBS 的权限。请参考《数据湖探索用户指南》>《全局配置》>《服务授权》。

步骤 4：创建队列

第一次提交 Spark 作业，需要先创建队列，例如创建名为“sparktest”的队列，队列类型选择为“通用队列”。

1. 在 DLI 管理控制台的左侧导航栏中，选择“队列管理”。
2. 单击“队列管理”页面右上角“创建队列”进行创建队列。

3. 创建名为“sparktest”的队列，队列类型选择为“通用队列”。创建队列详细介绍请参考《数据湖探索用户指南》>《创建队列》。
4. 单击“立即创建”，完成队列创建。

步骤 5：创建程序包

提交 Spark 作业之前需要创建程序包，例如“spark-examples.jar”。

1. 在管理控制台左侧，单击“数据管理”>“程序包管理”。
2. 在“程序包管理”页面，单击右上角“创建”可创建程序包。
3. 在“创建程序包”对话框，“包类型”选择“JAR”，“OBS 路径”选择[步骤 2：上传数据至 OBS](#)中“spark-examples.jar”的包路径，“分组设置”参数选择为“不分组”。
4. 单击“确定”，完成创建程序包。

程序包创建成功后，您可以在“程序包管理”页面查看和选择使用对应的包。

创建程序包详细介绍请参考《数据湖探索用户指南》>《创建程序包》。

步骤 6：提交 Spark 作业

1. 在 DLI 管理控制台，单击左侧导航栏中的“作业管理”>“Spark 作业”，单击“创建作业”，进入创建 Spark 作业页面。
2. 在 Spark 作业编辑页面中，“所属队列”选择[步骤 4：创建队列](#)中创建的队列，“应用程序”选择[步骤 5：创建程序包](#)创建的程序包。
其他参数请参考《数据湖探索用户指南》>《创建 Spark 作业》中关于 Spark 作业编辑页面的说明。
3. 单击 Spark 作业编辑页面右上方“执行”，阅读并同意隐私协议，单击“确定”。提交作业，页面显示“作业提交成功”。
4. （可选）可到“作业管理”>“Spark 作业”页面查看提交作业的状态及日志。

说明

在 DLI 管理控制台第一次单击“执行”操作时，需要阅读隐私协议，同意确定后，后续操作将不会再提示。

3.4 创建并提交 Flink SQL 作业

使用 DLI 提交 Flink SQL 作业进行实时计算。基本流程如下：

[步骤 1：登录云](#)

[步骤 2：准备数据源和数据输出通道](#)

[步骤 3：创建 OBS 桶保存输出数据](#)

[步骤 4：登录 DLI 管理控制台](#)

[步骤 5：创建队列](#)

步骤 6: [创建增强型跨源连接](#)

步骤 7: [创建跨源认证](#)

步骤 8: [配置安全组规则和测试地址连通性](#)

步骤 9: [创建 Flink SQL 作业](#)

样例场景需要创建一个 Flink SQL 作业，并且该作业有一个输入流和一个输出流。输入流用于从 DIS 读取数据，输出流用于将数据写入到 Kafka 中。

步骤 1: 登录云

使用 DLI 服务，首先要登录云。

1. 打开产品首页。
2. 在登录页面输入“帐号名”和“密码”，单击“登录”。

步骤 2: 准备数据源和数据输出通道

DLI Flink 作业支持其他服务作为数据源和数据输出通道，具体内容请参见《数据湖探索用户指南》>《Flink 作业管理》>《准备数据》。

本样例中，假设作业名称为“JobSample”，采用 DIS 服务作为数据源，开通数据接入服务（DIS），具体操作请参见《数据接入服务用户指南》中的“开通 DIS 通道”章节。采用分布式消息服务 Kafka 作为数据输出通道，创建 Kafka 专享版实例，具体操作请参见《分布式消息服务 Kafka 用户指南》中的“购买实例”章节。

- 创建用于作业输入流的 DIS 通道：
 - a. 登录 DIS 管理控制台。
 - b. 在管理控制台左上角选择区域和项目。
 - c. 单击“购买接入通道”配置相关参数。通道信息如下：
 - 区域：选择与 DLI 服务相同的区域
 - 通道名称：csinput
 - 通道类型：普通
 - 分区数量：1
 - 生命周期（小时）：24
 - 源数据类型：BLOB
 - 自动扩缩容：关闭
 - 企业项目：default
 - 高级配置：暂不配置
 - d. 单击“立即购买”，进入“规格确认”页面。
 - e. 单击“提交”，完成通道接入。
- 创建用于作业输出流的 Kafka 专享版实例：
 - a. 在创建 Kafka 实例前您需要提前准备相关依赖资源，包括 VPC、子网和安全组，并配置安全组。

- 创建 VPC 和子网的操作指导请参考《虚拟私有云用户指南》>创建虚拟私有云和子网，若需要在已有 VPC 上创建和使用新的子网，请参考《虚拟私有云用户指南》>为虚拟私有云创建新的子网。

📖 说明

- 创建的 VPC 与使用的 Kafka 服务应在相同的区域。
- 创建 VPC 和子网时，如无特殊需求，配置参数使用默认配置即可。
 - 创建安全组的操作指导请参考《虚拟私有云用户指南》>创建安全组，为安全组添加规则的操作指导请参考《虚拟私有云用户指南》>添加安全组规则。

更多信息请参考《分布式消息服务 Kafka 用户指南》中的“准备实例依赖资源”章节。

- b. 登录分布式消息服务 Kafka 管理控制台。
- c. 在管理控制台左上角选择区域。
- d. 在“Kafka 专享版”页面，单击右上角“购买 Kafka 实例”配置相关参数。实例信息如下：
 - 区域：选择与 DLI 服务相同的区域
 - 项目：默认
 - 可用区：默认
 - 实例名称：kafka-dliflink
 - 企业项目：default
 - 版本：默认
 - CPU 架构：默认
 - 规格：选择对应的规格
 - 代理个数：默认
 - 存储空间：默认
 - 容量阈值策略：默认
 - 虚拟私有云，子网：选择 a 中创建的虚拟私有云和子网。
 - 安全组：选择 a 中创建的安全组。
 - Manager 用户名：dliflink（用于登录实例管理页面）
 - 密码：****（请妥善保管密码，系统无法获取您设置的密码内容）
 - 确认密码：****
 - 更多配置：开启参数“Kafka SASL_SSL”，根据界面提示配置 SSL 认证的用户名和密码。其他参数可暂不配置。
- e. 单击“立即购买”，弹出“规格确认”页面。
- f. 单击“提交”，完成实例创建。
- g. 在分布式消息服务 Kafka 管理，单击“Kafka 专享版”，单击已创建的 Kafka 实例名称，例如 kafka-dliflink，进入实例详情页面。
- h. 在“基本信息 > 高级配置 > SSL 证书”所在行，单击下载按钮。下载压缩包到本地并解压，获取压缩包中的客户端证书文件：client.truststore.jks，给后续步骤做准备。

步骤 3：创建 OBS 桶保存输出数据

在本样例中，需要为作业“JobSample”开通对象存储服务（OBS），为 DLI Flink 作业提供 Checkpoint、保存作业日志和调试测试数据的存储功能。

具体操作请参见《对象存储服务控制台指南》中的“创建桶”章节。

1. 在 OBS 管理控制台左侧导航栏选择“对象存储”。
2. 在页面右上角单击“创建桶”，配置桶参数。
 - 区域：选择与 DLI 服务相同的区域
 - 桶名称：具体根据实际情况选择桶名，例如当前选择：smoke-test
 - 存储类别：标准存储
 - 桶策略：私有
 - 默认加密：关闭
 - 归档数据直读：关闭
 - 企业项目：default
 - 标签：不填写
3. 单击“立即创建”。

步骤 4：登录 DLI 管理控制台

1. 在列表中，选择“数据湖探索 DLI”。
2. 进入 DLI 管理控制台页面。第一次进入数据湖探索管理控制台需要进行授权，以获取访问 OBS 的权限。

步骤 5：创建队列

创建 DLI Flink SQL 作业，不能使用系统已有的 default 队列，需要您创建队列，例如创建名为“Flinktest”的队列。创建队列详细介绍请参考《数据湖探索用户指南》>《创建队列》。

1. 在 DLI 管理控制台总览页，单击右上角“购买队列”进入购买队列页面。
2. 配置参数。
 - 队列名称：Flinktest
 - 队列类型：通用队列。勾选“专属资源模式”。
 - 队列规格：16CUs
 - 企业项目：default
 - 描述：不填
 - 高级选项：自定义配置
 - 网段：配置的网段不能与 Kafka 的子网网段冲突
3. 单击“立即购买”，确认配置。
4. 配置确认无误，提交请求。

步骤 6：创建增强型跨源连接

创建 DLI Flink 作业，还需要创建增强型跨源连接。具体操作请参考《数据湖探索用户指南》>《跨源连接》>《增强型跨源连接》。

📖 说明

- 绑定跨源的 DLI 队列网段和数据源网段不能重合。
 - 系统 default 队列不支持创建跨源连接。
 - 访问跨源表需要使用已经创建跨源连接的队列。
1. 在 DLI 管理控制台左侧导航栏中，选择“跨源连接”。
 2. 选择“增强型跨源”页签，单击左上角的“创建”按钮。配置参数：
 - 连接名称：diskafka
 - 绑定队列：Flinktest
 - 虚拟私有云：vpc-dli
 - 子网：dli-subnet

📖 说明

创建跨源连接的虚拟私有云和子网需要和 Kafka 实例保持一致。

3. 单击“确定”，完成创建增强型跨源连接。
4. 在“增强型跨源”页签，单击创建的连接名称：diskafka，查看对等连接 ID 及连接状态，连接状态为“已激活”表示连接成功。

步骤 7：创建跨源认证

创建跨源认证的具体操作请参考《数据湖探索用户指南》>《跨源连接》>《跨源认证》。

1. 将**步骤 2：准备数据源和数据输出通道**中获取的 kafka 认证文件“client.truststore.jks”上传到**步骤 3：创建 OBS 桶保存输出数据**中的 OBS 桶“smoke-test”下。
2. 在 DLI 管理控制台选择“跨源连接”。
3. 在“跨源认证”页签，单击“创建”，创建认证信息。配置参数：
 - 认证信息名称：Flink
 - 类型：Kafka_SSL
 - Truststore 路径：obs://smoke-test/client.truststore.jks
 - Truststore 密码：dms@kafka其余参数可不用配置。
4. 单击“确定”，完成创建跨源认证。

步骤 8：配置安全组规则和测试地址连通性

1. DLI 管理控制台，单击“队列管理”，选择绑定的队列，点开队列左边的箭头，查看队列详情，获取队列的网段信息。

2. 登录分布式消息服务 Kafka 管理控制台，单击“Kafka 专享版”，单击已创建的 Kafka 实例名称，例如 kafka-dliflink，进入实例基本信息页面。
3. 在实例基本信息页面，在“连接地址”配置下的获取 Kafka 的连接地址和端口。
4. 在实例基本信息页面，在“网络”配置下的“安全组”，单击安全组名称，进入安全组配置页面。
5. 在 Kafka 实例对应的安全组配置页面，单击“入方向规则 > 添加规则”，协议选择“TCP”，端口选择“9093”，源地址填写 DLI 队列的网段。单击“确定”完成配置。
6. 登录 DLI 管理控制台，选择“队列管理”，在所在 Flink 队列行，单击“更多 > 测试地址连通性”，在“地址”参数中按照“IP:端口”的格式输入 Kafka 的连接地址和端口，单击“测试”，返回地址可达后进行后续操作步骤。注意多个地址要分开单独测试。

步骤 9：创建 Flink SQL 作业

准备好数据源和数据输出通道之后，就可以创建 Flink SQL 作业了。

1. 在 DLI 管理控制台的左侧导航栏中，单击“作业管理” > “Flink 作业”，进入“Flink 作业”页面。
2. 在“Flink 作业”页面右上角单击“创建作业”，弹出“创建作业”对话框。配置参数：
 - 类型：Flink SQL
 - 名称：DIS-Flink-Kafka
 - 描述：不填
 - 模板名称：不选择
3. 单击“确定”，进入作业“编辑”页面。
4. 编辑 SQL 作业。

在 SQL 语句编辑区域，输入详细的 SQL 语句。具体如下，注意以下加粗的参数值都需要根据注释提示修改。

```
CREATE SOURCE STREAM car_info (  
  a1 string,  
  a2 string,  
  a3 string,  
  a4 INT  
)  
WITH (  
  type = "dis",  
  region = "", //需要修改为当前 DLI 队列所在的 region  
  channel = "csinput",  
  encode = "csv",  
  FIELD_DELIMITER = ";"  
);  
  
CREATE SINK STREAM kafka_sink (  
  a1 string,  
  a2 string,  
  a3 string,  
  a4 INT
```

```
) // 输出字段

WITH (
  type="kafka",
  kafka_bootstrap_servers = "192.x.x.x:9093, 192.x.x.x:9093,
192.x.x.x:9093", //需要修改为 kafka 实例的连接地址
  kafka_topic = "testflink", // 要写入 kafka 的 topic,进入 kafka 控制台,单击已创建的
Kafka 实例名称,在 Topic 管理查看 Topic 名称
  encode = "csv", // 编码格式,支持 json/csv
  kafka_certificate_name = "Flink",
  kafka_properties_delimiter = ",",
  //kafka_properties 中的 username 和 password 的值 xxx 需要替换为步骤 2 中 kafka 创建 SSL
认证的用户名和密码
  kafka_properties =
"sasl.jaas.config=org.apache.kafka.common.security.plain.PlainLoginModule
required username=\"xxx\"
password=\"xxx\";sasl.mechanism=PLAIN,security.protocol=SASL_SSL"
);

INSERT INTO kafka_sink
SELECT * FROM car_info;

CREATE sink STREAM car_info1 (
  a1 string,
  a2 string,
  a3 string,
  a4 INT
)
WITH (
  type = "dis",
  region = "", //需要修改为当前 DLI 队列所在的 region
  channel = "csinput",
  encode = "csv",
  FIELD_DELIMITER = ";"
);

insert into car_info1 select 'id','owner','brand',1;
insert into car_info1 select 'id','owner','brand',2;
insert into car_info1 select 'id','owner','brand',3;
insert into car_info1 select 'id','owner','brand',4;
insert into car_info1 select 'id','owner','brand',5;
insert into car_info1 select 'id','owner','brand',6;
insert into car_info1 select 'id','owner','brand',7;
insert into car_info1 select 'id','owner','brand',8;
insert into car_info1 select 'id','owner','brand',9;
insert into car_info1 select 'id','owner','brand',10;
```

5. 单击“语义校验”，确保语义校验成功。
6. 设置作业运行参数。配置必选参数：
 - 所属队列：Flinktest
 - CU 数量：2
 - 管理单元：1
 - 并行数：1

- 保存作业日志：勾选
- OBS 桶：选择作业日志保存的 OBS 桶，并进行授权。

其余参数可不用配置。

7. 单击“保存”，保存作业和相关参数。
8. 单击“启动”，进入“启动 Flink 作业”页面，确认作业规格和费用后，单击“立即启动”，启动作业。

启动作业后，系统将自动跳转到 Flink 作业管理页面，新创建的作业将显示在作业列表中，在“状态”列中可以查看作业状态。作业提交成功后，状态将由“提交中”变为“运行中”。

如果作业状态为“提交失败”或“运行异常”，表示作业提交或运行失败。用户可以在作业列表中的“状态”列中，将鼠标移动到状态图标上查看错误信息，单击可以复制错误信息。根据错误信息解决故障后，重新提交。

9. 作业运行完成后，可登录分布式消息服务 Kafka 管理控制台，查看对应的 Kafka 专享实例。单击实例名称，选择“消息查询”页签，选择 Flink SQL 作业中写入的 kafka 的 Topic 名称，单击“搜索”，在操作列单击“查看消息正文”查看写入的消息内容。

4 DLI 控制台总览

页面介绍

登录数据湖探索 DLI，进入“总览”页面。

按照从左到右，从上到下的顺序，说明如下。

表4-1 区域说明

序号	区域	描述
1	导航栏	数据湖探索管理控制台的导航栏，包括总览、SQL 编辑器、作业管理、队列管理、数据管理、作业模板、跨源管理和全局配置。
2	数据湖探索简介	数据湖探索（Data Lake Insight,DLI）是完全托管的批流结合大数据处理分析服务。用户不需要管理任何服务器，即开即用。支持标准 SQL/Spark SQL/Flink SQL，支持多种接入方式，并兼容主流数据格式。
3	创建队列	单击“创建队列”可直接创建队列。
4	SQL 作业简介	单击本区域右上角“创建作业”，可进入“SQL 编辑器”页面。详细介绍请参考 5 SQL 编辑器。
5	Flink 作业简介	单击本区域右上角“创建作业”，可进入 Flink 作业“作业管理”页面。详细介绍请参考 6.2 Flink 作业管理。
6	Spark 作业简介	单击本区域右上角“创建作业”，可进入 Spark 作业“作业编辑”页面。详细介绍请参考 6.3.2 创建 Spark 作业。
8	所有队列使用时长	可查看所有队列最近一天或一周或一个月的使用情况。
9	单队列使用时长	可查看对应单个队列最近一周或一个月的使用情况。
10	常用链接	包括：

序号	区域	描述
		<ul style="list-style-type: none">• 用户指南• API 参考• SDK 下载
11	创建作业操作步骤	<ol style="list-style-type: none">1. 新建队列2. 准备数据3. 编辑并提交作业4. 查看作业状态

5 SQL 编辑器

SQL 作业编辑器支持使用 SQL 语句执行数据查询操作。支持 SQL2003，兼容 SparkSQL。

在总览页面，单击左侧导航栏中的“SQL 编辑器”或 SQL 作业简介右上角的“创建作业”，可进入“SQL 编辑器”页面。

进入 SQL 编辑器，页面会提示系统将创建 DLI 临时数据桶。该桶用于存储使用 DLI 服务产生的临时数据，例如：作业日志、作业结果等。如果不创建该桶，将无法查看作业日志。可以，实现定时删除 OBS 桶中的对象或者定时转换对象的存储类别。桶名称为系统默认。

如果不需要创建 DLI 临时数据桶，并且希望不再收到该提示，可以勾选“下次不再提示”并单击“取消”。

“SQL 编辑器”页面中的区域和按键功能介绍如下：

导航栏

页面左侧的导航栏包括“数据库”列表和“队列”列表，说明如下。

表5-1 导航栏按键说明

序号	页签/按键	页签/按键名称	描述
1	数据库	数据库	显示已有的数据库及其下所有的表。 <ul style="list-style-type: none">单击数据库名，将显示该数据库中的表。单击表名，将在表名下显示该表中的元数据，最多可显示 20 个元数据。双击表名，将在作业编辑窗口自动输入 SQL 查询语句。
2	队列	队列	显示已有的队列。
3	模板	模板	自定义模板和内置的 SQL 样例模板（目前包含 22 条标准的 TPC-H 查询语句）。

序号	页签/按钮	页签/按钮名称	描述
4		创建	包括创建队列、数据库和表。具体操作请分别参考 7.3 创建队列和 8.1.4 创建数据库和表。
5		刷新	包括刷新已有的队列、数据库和表列表。
6		搜索	可输入关键字查找对应的数据库和表。

SQL 作业编辑窗口

页面右侧上方为 SQL 作业编辑窗口，按照从上至下，从左至右的顺序，说明如下。

表5-2 SQL 作业编辑窗口说明

序号	按钮	描述
2	队列	下拉选择需要使用的队列。如果没有可用队列，此处显示“请选择队列”，请先创建队列。SQL 作业只能在队列类型为“SQL 队列”下执行。具体队列创建可以参考 7.3 创建队列。
3	数据库	下拉选择需要使用的数据库。如果没有可用数据库，此处显示“请选择数据库”，请先创建数据库。 说明 如果 SQL 语句中指定了表所在的数据库，则此处选择的数据库无效。
4	执行	执行作业编辑窗口中的 SQL 语句。
5	格式化	格式化 SQL 语句。
6	语法参考	可跳转至《数据湖探索 SQL 语法参考》手册。
7	设置	包括设置“参数设置”和“标签”。 参数设置：以“key/value”的形式设置提交 SQL 作业的配置项。 标签：以“key/value”的形式设置 SQL 作业的标签。
8	更多	包括： <ul style="list-style-type: none"> 语法校验：判断 SQL 语句编写是否正确。 设为模板：将常用的 SQL 语句设为模板。具体操作请参见 9.1 SQL 模板管理。 切换主题：选择白底黑字或黑底白字。
9	SQL 语句编辑区	编辑 SQL 语句的区域。

序号	按键	描述
10	光标位置	标识光标在 SQL 语句编辑区中的位置（行，列）。
11	快捷键	快捷键介绍具体请参考表 5-3。

表5-3 快捷键说明

快捷键	描述
Ctrl+Enter	执行 SQL。通过按下键盘上的 Ctrl+R 或 Ctrl + Enter，您可以执行 SQL 语句。
Ctrl+F	搜索 SQL。通过按下键盘上的 Ctrl + F，您可以搜索需要的 SQL 语句。
Shift+Alt+F	格式化 SQL。通过按下键盘上的 Shift+Alt+F，您可以将 SQL 语句格式化。
Ctrl+Q	语法校验。通过按下键盘上的 Ctrl + Q，您可以对 SQL 语句进行语法校验。
F11	全屏。通过按下键盘上的 F11，您可将 SQL 作业编辑器窗口全屏。再次按下 F11，将从全屏复原。

“执行历史”和“查看结果”

页面右侧下方为 SQL 作业“执行历史（最近一天）”页签和“查看结果”页签，说明如下。

- 执行历史

表5-4 SQL 作业执行历史说明

区域	描述
执行历史（最近一天）	<p>执行历史（最近一天）：显示最近一天提交的作业的信息。包括：</p> <ul style="list-style-type: none"> • 队列：队列名称 • 用户名：执行 SQL 的用户 • 类型：SQL 作业的类型 • 状态：SQL 作业的执行状态 • 执行语句 • 创建时间 • 操作：

区域	描述
	<ul style="list-style-type: none"> - 编辑：重新编辑 SQL 语句 - SparkUI：跳转至 SparkUI 查看 SQL 语句执行进程 <p>说明</p> <p>新建队列，运行作业时会重新拉集群，大概需要 10 分钟左右才能拉好集群，在集群创建好之前单击 SparkUI 会导致缓存空的 projectID，从而导致无法查看 SparkUI。建议使用专属队列，集群不会被释放，就不会有该问题，或者提交作业后等一段时间再查看 SparkUI，确保集群已经拉好了，不要立即单击 SparkUI。</p> <p>目前 DLI 配置 SparkUI 只展示最新的 100 条作业信息。</p> <p>default 队列下运行的作业或者该作业为同步作业时不支持该操作。</p> <ul style="list-style-type: none"> - 更多：以下操作根据 SQL 类型和运行状态的不同显示会有差异，请以实际界面为准。 <p>终止：正在运行或提交中的 SQL 支持终止操作。</p> <p>重新执行：重新执行该 SQL 语句。</p> <p>查看结果：QUERY 作业支持查看作业执行结果。</p> <p>导出结果：QUERY 作业这次导出作业运行结果到指定 OBS 路径下。</p> <p>归档日志：SQL 语句运行日志存储的 OBS 路径。</p> <p>导出日志：导出 SQL 语句运行日志。</p> <p>说明</p> <p>导出日志用户需要具有创建 OBS 桶的权限。</p> <p>default 队列下运行的作业或者该作业为同步作业时不支持归档日志和导出日志操作。</p>

可以通过以下方式筛选执行历史：

- 在右上角选择队列名称或输入执行语句
- 在列表中选择创建时间顺序/倒序排列
- 在列表中选择作业状态

- 查看结果

表5-5 SQL 作业查看结果说明

序号	按键	描述
1	清空	清空显示的 SQL 语句执行结果。
2		以图形/表格的形式展示查询结果。

序号	按键	描述
3		<p>将查询结果导出到 OBS。具体操作介绍请参考导出查询结果。</p> <p>当前控制台界面查询结果最多显示 1000 条数据，如果需要查看更多或者全量数据，则可以通过“导出结果”功能将数据导出到 OBS 获取。</p>

SQL 查询操作步骤

1. 登录数据湖探索管理控制台，选择 SQL 作业，单击“创建作业”，进入“SQL 编辑器”页面。
2. 进入“SQL 编辑器”页面后，系统会提示“创建 DLI 临时数据桶”，用于存储使用 DLI 服务产生的临时数据，如：作业日志、作业结果。如果不创建该桶，您将无法查看作业日志。单击“确定”。系统将提示创建临时桶成功。
3. 在 SQL 作业编辑窗口右上方的“队列”列表中选择所使用的队列。创建队列操作步骤请参见 7.3 创建队列。
4. 在 SQL 作业编辑窗口右上方的“数据库”列表中选择所使用的数据库，例如选择数据库“qw”。
5. 创建表，例如表名为“qw”。创建数据库和表操作步骤请参见 8.1.4 创建数据库和表。
6. 在 SQL 作业编辑窗口输入表“qw”的 SQL 查询语句：


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

 或者双击左侧表名“qw”，上述查询语句会自动在作业编辑窗口中输入。
7. 单击“更多”中的“语法校验”，确认 SQL 语句书写是否正确。
 - a. 若语法校验失败，请参考《数据湖探索 SQL 语法参考》，检查 SQL 语句准确性。
 - b. 若语法校验通过，单击“执行”，阅读并同意隐私协议，单击“确定”后执行 SQL 语句。
 - c. SQL 语句执行成功后，在 SQL 作业编辑窗口下方会显示执行结果。
8. （可选）当前控制台界面查询结果最多显示 1000 条数据，如果需要查看更多或者全量数据，则可以通过  “导出结果”功能将数据导出到 OBS 获取。
9. （可选）在查看结果页签，单击右上侧  “结果图形化”按钮，查询结果将以图形形式呈现。再单击“表格展示”，查询结果可切换回表格形式。

说明

- 若执行结果中无数值列，则无法进行图形化。
- 图形类型包括柱状图、折线图、扇形图。
- 柱状图和折线图的 X 轴可为任意一列，Y 轴仅支持数值类型的列，扇形图对应图例和指标。

作业编辑窗口操作技巧

- 在 SQL 作业编辑窗口中导入 SQL 语句的简便方法
 - 可以双击左侧导航栏列表中的表名，即可将选定表的查询语句导入 SQL 语句编辑窗口中，单击“执行”，即可完成查询。
 - 可以通过单击“更多”，选择“设为模板”，将对应的 SQL 语句保存为模板，供将来执行使用。
需要使用时，通过单击“更多”，选择“选择模板”，在已有模板中双击所需的 SQL 语句，导入 SQL 作业编辑窗口中，单击“执行”，或根据需要进行修改后执行。
 - 可以批量执行 SQL 语句。
- 作业编辑窗口的特点
作业编辑窗口具有以下三个特点。
 - 颜色突出显示 - 常用语法采用不同颜色突出显示。
 - 注释支持 - 支持单行注释和多行注释。以“--”开头，后续内容即为注释。
 - 光标位置显示 - 可以用于确认位于作业编辑窗口内的鼠标光标的列号和行号。

6 作业管理

6.1 SQL 作业管理

SQL 作业包括在 5 SQL 编辑器窗口执行 SQL 语句，导入数据和导出数据等操作。

SQL 作业管理主要包括如下功能：

- **查找作业**：筛选符合对应条件的作业。
- **查看作业详情**：显示作业的详细信息。
- **终止作业**：终止“提交中”或“运行中”的作业。
- **导出查询结果**：当前控制台界面查询结果最多显示 1000 条数据，如果需要查看更多或者全量数据，则可以通过该功能将数据导出到 OBS 获取。

作业管理页面

在总览页面单击“SQL 作业”简介，或在左侧导航栏单击“作业管理”>“SQL 作业”，可进入 SQL 作业管理页面。SQL 作业管理页面显示所有 SQL 作业，作业数量较多时，系统分页显示，可根据需要跳转至指定页面。您可以查看任何状态下的作业。作业列表默认按创建时间降序排列，创建时间最近的作业显示在最前端。

表6-1 作业管理参数

参数	参数说明
队列	作业所属队列的名称。
用户名	执行该作业的用户名。
类型	作业的类型，包括如下。 <ul style="list-style-type: none">• IMPORT：导入数据到 DLI 的作业。• EXPORT：从 DLI 导出数据的作业。• DCL：包括传统 DCL，以及队列权限相关的操作。• DDL：与传统 DDL 操作一致，即创建和删除数据库，创建和删除表的作业。• QUERY：执行 SQL 查询数据的作业。

参数	参数说明
	<ul style="list-style-type: none"> • INSERT: 执行 SQL 插入数据的作业。 • UPDATE: 更新数据。 • DELETE: 删除 SQL 作业。 • DATA_MIGRATION: 数据迁移。 • RESTART_QUEUE: 重启队列。 • SCALE_QUEUE: 队列规格变更（扩容/缩容）。
状态	作业的状态信息，包括如下。 <ul style="list-style-type: none"> • 提交中 • 运行中 • 已成功 • 已取消 • 已失败 • 规格变更中
执行语句	作业的具体 SQL 语句以及导出、建表的操作，此处展示操作的描述。 单击  可复制对应的语句。
运行时长	作业的运行时长。
创建时间	每个作业的创建时间，可按创建时间顺序或倒序显示作业列表。
操作	<ul style="list-style-type: none"> • 编辑：重新编辑修改该作业。 • 终止： <ul style="list-style-type: none"> - 当作业状态在“提交中”和“运行中”时，“终止”按钮才生效。 - 当作业状态为“已成功”、“已失败”、“已取消”的作业不能终止。 - 当“终止”按钮为灰色时，表示无法执行终止操作。 • 重新执行：重新执行该作业。 • SparkUI：单击后，将跳转至 Spark 任务运行情况界面。 <p>说明</p> <ul style="list-style-type: none"> • 新建队列，运行作业时会重新拉集群，大概需要 10 分钟左右才能拉好集群，在集群创建好之前单击 SparkUI 会导致缓存空的 projectID，从而导致无法查看 SparkUI。建议使用专属队列，集群不会被释放，就不会有该问题，或者提交作业后等一段时间再查看 SparkUI，确保集群已经拉好了，不要立即单击 SparkUI。 • 目前 DLI 配置 SparkUI 只展示最新的 100 条作业信息。 • QUERY 作业和异步 DDL 作业除上述操作外，还包括： <ul style="list-style-type: none"> - 查看结果：查看作业运行结果。

参数	参数说明
	<ul style="list-style-type: none"> - 导出结果：将作业运行结果导出至用户创建的 OBS 桶中。具体操作请见导出查询结果。 • EXPORT 作业除上述操作外，还包括： <ul style="list-style-type: none"> - 立即下载 • 归档日志：将作业日志保存到系统创建的 DLI 临时 OBS 数据桶中。 • 导出日志：将日志导出至用户创建的 OBS 桶中进行查看。当作业状态在“运行中”时，不能导出日志。 <p>说明</p> <p>导出日志需要用户需要具有创建 OBS 桶的权限。</p> <p>default 队列下运行的作业或者该作业为同步作业时不支持归档日志和导出日志操作。</p>

查找作业

在“SQL 作业”页面，可以通过以下方式对作业进行过滤筛选，在页面中显示符合对应条件的作业。

- 选择队列名称
- 设置日期范围
- 输入用户名/执行语句/作业 ID
- 选择创建时间顺序/倒序排列
- 选择作业类型
- 选择作业状态
- 选择运行时长顺序/倒序排列

查看作业详情

在“SQL 作业”页面，选中一条作业，单击该作业对应的 ，可查看该条作业的详细信息。

不同类型的作业，显示的作业详情不同。以导入数据作业，建表作业和查询作业为例说明。作业详情根据作业类型、状态和配置选项不同显示可能存在差异，具体以实际界面显示为准。

- 导入数据（load data）作业（作业类型：IMPORT），包括以下信息：队列，作业 ID，用户名，类型，状态，执行语句，运行时长，创建时间，结束时间，参数设置，标签，结果条数，已扫描数据，扫描数据条数，错误记录条数，存储路径，数据格式，数据库，表，表头，分隔符，引用字符，转义字符，日期格式，时间戳格式。

- 建表 (create table) 作业 (作业类型: DDL), 包括以下信息: 队列, 作业 ID, 用户名, 类型, 状态, 执行语句, 运行时长, 创建时间, 结束时间, 参数设置, 标签, 结果条数, 已扫描数据, 数据库。
- 查询 (select) 作业 (作业类型: QUERY), 包括以下信息: 队列, 作业 ID, 用户名, 类型, 状态, 执行语句, 运行时长, 创建时间, 结束时间, 参数设置, 标签, 结果条数 (运行成功, 可导出结果), 已扫描数据, 执行用户, 结果状态 (运行成功, 可查看结果; 运行失败, 显示失败原因), 数据库。

终止作业

在“SQL 作业”页面, 可单击“操作”列的“终止”, 终止“提交中”或“运行中”的作业。

导出查询结果

当前控制台界面查询结果最多显示 1000 条数据, 如果需要查看更多或者全量数据, 则可以通过该功能将数据导出到 OBS 获取。具体操作步骤如下:

导出查询结果的操作入口有两个, 分别在“SQL 作业”和“SQL 编辑器”页面。

- 在“作业管理”>“SQL 作业”页面, 可单击对应作业“操作”列“更多”中的“导出结果”, 可导出执行查询后的结果。
- 在“SQL 编辑器”页面, 查询语句执行成功后, 在“查看结果”页签右侧, 单击  “导出结果”, 可导出执行查询后的结果。

说明

若查询结果中无数值列, 则无法导出查询结果。

表6-2 参数说明

参数名称	描述
数据源格式	导出查询结果数据的文件格式。当前只支持 json 格式。
队列	选择运行的队列。SQL 作业只能在队列类型为“SQL 队列”下执行。具体队列创建可以参考 7.3 创建队列。
压缩格式	导出查询结果数据的压缩方式, 选择如下压缩方式。 <ul style="list-style-type: none"> • none • bzip2 • deflate • gzip
存储路径	输入或选择 OBS 的路径。 说明 <ul style="list-style-type: none"> • 选择 OBS 桶后, 请在文本框中定义文件夹名称, 若该文件夹不存在, 则会在 OBS 中创建。

参数名称	描述
	<ul style="list-style-type: none">文件夹名称不能包含下列特殊字符：\/:*?"<> ，并且不能以“.”开头和结尾。
导出方式	导出查询结果数据的保存方式。 <ul style="list-style-type: none">随导出创建指定路径：指定的导出目录必须不存在，如果指定目录已经存在，系统将返回错误信息，无法执行导出操作。覆盖指定路径：在指定目录下新建文件，会删除已有文件。
结果条数	导出查询结果数据的数量。不填写数值或数值为“0”时，导出全部结果。
表头	设置导出查询结果数据是否含表头。

6.2 Flink 作业管理

6.2.1 Flink 作业管理概述

在 Flink 作业管理页面可提交 Flink 作业。目前有以下作业类型：

- Flink SQL 作业：使用 SQL 语句定义作业，可以提交到通用队列上。
- Flink Jar 作业：基于 Flink API 的自定义 Jar 包作业，可以运行在独享队列上。

Flink 作业管理主要包括如下功能：

- 6.2.2 Flink 作业权限管理
- 6.2.4 创建 Flink SQL 作业
- 6.2.5 创建 Flink Jar 作业
- 6.2.6 调试 Flink 作业
- [编辑作业](#)
- [启动作业](#)
- [停止作业](#)
- [删除作业](#)
- [导出作业](#)
- [导入作业](#)
- [名称和描述修改](#)
- [导入保存点](#)
- [触发保存点](#)
- [运行时配置](#)
- 6.2.8 Flink 作业详情

以及查看“使用指南”和“使用视频”。

委托权限设置

DLI 执行 Flink 作业需要进行委托授权，可在第一次登录管理控制台时进行设置，也可在“全局配置”>“11.2 服务授权”中进行修改。

具体权限如下：

- **Tenant Administrator(全局服务)**：DLI Flink 作业访问和使用 OBS 或者 DWS 数据源、日志转储（包括桶授权）、开启 checkpoint、作业导入导出等，需要获得访问和使用 OBS（对象存储服务）的 Tenant Administrator 权限。

📖 说明

由于云服务缓存需要时间，该权限 60 分钟左右才能生效。

- **DIS Administrator**：DLI Flink 作业访问和使用 DIS 数据源，需要获得访问和使用 DIS（数据接入服务）的 DIS Administrator 权限。

📖 说明

由于云服务缓存需要时间，该权限 30 分钟左右才能生效。

- **CloudTable Administrator**：DLI Flink 作业访问和使用 CloudTable 数据源，需要获得访问和使用 CloudTable（表格存储服务）的 CloudTable Administrator 权限。

📖 说明

由于云服务缓存需要时间，该权限 3 分钟左右才能生效。

- **Tenant Administrator(项目级)**：使用其他必须具有 Tenant Administrator 权限才能运行的服务也需要获得该权限。

📖 说明

由于云服务缓存需要时间，该权限 3 分钟左右才能生效。

作业管理页面

在总览页面单击“Flink 作业”简介，或在左侧导航栏单击“作业管理”>“Flink 作业”，可进入 Flink 作业管理页面。Flink 作业管理页面显示所有的 Flink 作业，作业数量较多时，系统分页显示，您可以查看任何状态下的作业。

表6-3 作业管理参数

参数	参数说明
ID	所提交 Flink 作业 ID，由系统默认生成。
名称	所提交 Flink 作业的名称。
类型	所提交 Flink 作业的类型。包括： <ul style="list-style-type: none">• Flink SQL：Flink SQL 作业• Flink Jar：Flink Jar 作业
状态	作业的状态信息，包括： <ul style="list-style-type: none">• 草稿

参数	参数说明
	<ul style="list-style-type: none"> 提交中 提交失败 运行中（提交作业后，返回正常结果） 运行异常（作业发生运行时异常，停止运行作业） 下载中 空闲 停止中 已停止 停止失败 保存点创建中 已完成
描述	所提交 Flink 作业的描述。
用户名	提交作业的用户名称。
创建时间	每个作业的创建时间。
开始时间	Flink 作业开始运行的时间。
运行时长	作业运行所消耗的时间。
操作	<ul style="list-style-type: none"> 编辑：编辑已经创建好的作业。具体请参见编辑作业。 启动：启动作业并运行。具体请参见启动作业。 更多 <ul style="list-style-type: none"> FlinkUI：单击后，将跳转至 Flink 任务运行情况界面。 <p>说明</p> <p>如果是新建队列，在该队列提交作业后，如果立即单击 FlinkUI，因为后台大约需要 10 分钟创建集群，会导致缓存空的 projectID，从而导致无法查看 FlinkUI。</p> <p>建议作业选择使用专属队列，后台集群不会被释放，避免上述问题产生。或者等待作业运行中时再查看 FlinkUI，确保集群已经拉好了，不要立即单击 FlinkUI。</p> <ul style="list-style-type: none"> 停止：停止 Flink 作业。如果该功能置灰，表示当前状态的作业不支持停止。 删除：删除作业。 <p>说明</p> <p>作业删除后不可恢复，请谨慎操作。</p> <ul style="list-style-type: none"> 名称和描述修改：修改作业名称和描述。具体请参考名称和描述修改。 导入保存点：导入原实时流计算服务作业导出的数据。具体请参考导入保存点。

参数	参数说明
	<ul style="list-style-type: none"> - 触发保存点：“运行中”的作业可以“触发保存点”，保存作业的状态信息。具体请参考触发保存点。 - 权限管理：查看作业对应的用户权限信息以及对其他用户授权。具体请参考 6.2.2 Flink 作业权限管理。 - 运行时配置：支持作业在运行时配置作业异常告警和异常自动重启。具体请参考运行时配置。

6.2.2 Flink 作业权限管理

Flink 作业权限操作场景

- 针对不同用户，可以通过权限设置分配不同的作业，不同用户之间的作业效率互不影响，保障作业性能。
- 管理员用户和作业的所有者拥有所有权限，不需要进行权限设置且其他用户无法修改其作业权限。

Flink 作业权限相关操作步骤

1. 在 DLI 管理控制台的左侧，选择“作业管理” > “Flink 作业”。
2. 选择待设置的作业，单击其“操作”列中的“更多” > “权限管理”。“用户权限信息”区域展示了当前具备此作业权限的用户列表。

权限设置有 3 种场景：为新用户赋予权限，为已有权限的用户修改权限，回收某用户具备的所有权限。

- 为新用户赋予权限

新用户指之前不具备此作业权限的用户。

- i. 单击“权限信息”右侧的“授权”，弹出“授权”对话框。
- ii. 填写“用户名”，并勾选对应权限。
- iii. 单击“确定”，完成新用户的添加。

待设置的参数说明如表 6-4 所示。

表6-4 Flink 作业授权参数说明

参数名称	描述
用户名	被授权用户的名称。 说明 该用户名称是已存在的 IAM 用户名称。并且该用户需要登录过云，才能进行授权操作。
权限设置	<ul style="list-style-type: none"> • 查看作业详情：查看此作业的作业详情。 • 更新作业：编辑修改此作业。 • 删除作业：删除此作业。

参数名称	描述
	<ul style="list-style-type: none">• 启动作业：启动该作业权限。• 停止作业：停止该作业。• 导出作业：导出该作业。• 赋权：当前用户可将作业的权限赋予其他用户。• 回收：当前用户可回收其他用户具备的该作业的权限，但不能回收该作业所有者的权限。• 查看其他用户具备的权限：当前用户可查看其他用户具备的该作业的权限。

- 为已有权限的用户赋予权限或回收权限。
 - i. 在对应作业“用户权限信息”区域的用户列表中，选择需要修改权限的用户，在“操作”列单击“权限设置”。
 - ii. 在作业“权限设置”对话框中，对当前用户具备的权限进行修改。详细权限描述如表 6-4 所示。

当“权限设置”中的选项为灰色时，表示您不具备修改此作业权限的权限。可以向管理员用户、作业所有者等具有赋权权限的用户申请“作业的赋权”和“作业权限的回收”权限。
 - iii. 单击“确定”完成权限设置。
- 回收某用户具备的所有权限。

在对应作业“权限信息”区域的用户列表中，选择需要删除权限的用户，在“操作”列单击“回收”。在“回收”对话框中单击“确定”后，此用户将不具备该作业的任意权限。

Flink 作业权限使用说明

- **查看作业**
 - 租户以及 admin 用户可以查看和操作所有作业。
 - 子用户以及拥有只读权限的用户只能查看自己的作业，以及他人赋权给该子用户任意权限的作业。
- **启动作业**
 - 使用独享队列时，用户需要同时拥有队列的提交作业权限以及作业的启动作业权限。
 - 使用共享队列时，用户只需要拥有作业的启动作业权限。
- **停止作业**
 - 使用独享队列时，用户需要同时拥有队列的停止作业权限以及作业的停止作业权限。
 - 使用共享队列时，用户只需要拥有作业的停止作业权限。
- **删除作业**
 - 如果作业在可删除状态，则用户拥有作业的删除权限即可。

- 如果作业在不可删除状态，用户删除作业时，系统会先停止作业，停止作业权限说明可以参考[•停止作业](#)，并且用户还需要拥有作业的删除权限。
- **创建作业**
 - 子用户默认不能创建作业。
 - 创建作业时，用户需要拥有创建作业的权限。目前只有 **admin** 用户创建作业的权限，同时用户还需要拥有该作业使用的相关程序包组权限或者程序包权限。
- **编辑作业**

编辑作业时，用户需要拥有更新作业的权限，同时用户还需要拥有该作业使用的相关程序包所属组权限或者程序包权限。

6.2.3 准备 Flink 作业数据

创建 Flink 作业需要输入数据源和数据输出通道，即常说的 **Source** 和 **Sink**。用户使用其他服务作为数据源或输出通道时，需要先开通相应服务。

Flink 作业支持以下数据源和输出通道：

- **DIS 数据源和输出通道**

如果用户作业需要 **DIS** 作为数据源和输出通道时，则要先开通数据接入服务 (**DIS**)。

用户如何开通 **DIS** 服务，具体操作请参见《数据接入服务用户指南》中的“开通 **DIS** 通道”章节。

申请 **DIS** 通道后，用户可以将本地数据通过 **DIS** 通道不断上传至 **DIS** 服务，实现向 **Flink** 作业提供实时流数据源，具体操作请参见《数据接入服务用户指南》中的“发送数据到 **DIS** 服务”章节。

样例数据如下所示：

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

- **OBS 数据源**

如果用户作业需要对象存储服务 (**OBS**) 作为数据源，则要先开通 **OBS** 服务，具体操作请参见《对象存储服务控制台指南》中的“开通 **OBS** 服务”章节。

开通 **OBS** 服务后，用户需要将本地文件通过 **Internet** 上传至 **OBS** 指定的位置，具体操作请参见《对象存储服务控制台指南》中的“上传文件”章节。

- **RDS 输出通道**

如果用户作业需要 **RDS** 作为输出通道，需要创建 **RDS** 实例，具体操作请参见《关系型数据库用户指南》中“创建实例”章节。

- **SMN 输出通道**

如果用户作业需要 **SMN** 作为输出通道，需要先在 **SMN** 中创建主题，获取 **URN** 资源标识，再添加订阅。具体操作请参见《消息通知服务用户指南》中“快速入门”章节。

- **Kafka 数据源和输出通道**

如果用户作业需要 **Kafka** 作为数据源和输出通道，则必须要通过创建增强型跨源连接与 **Kafka** 进行对接，具体操作请参见 10.2 增强型跨源连接。

如果 Kafka 服务端的端口监听在 hostname 上，则需要将 Kafka Broker 节点的 hostname 和 ip 的对应关系添加到跨源连接中。

- CloudTable 数据源和输出通道

如果用户作业需要 CloudTable 作为数据源和输出通道，需要先在 CloudTable 中创建集群，获取集群 ID。

- 云搜索服务输出通道

如果用户作业需要云搜索服务作为输出通道，需要先在云搜索服务中创建集群，获取集群内网访问地址。具体操作请参见《云搜索服务用户指南》中的“入门”章节。

- DCS 输出通道

如果用户作业需要 DCS 作为输出通道，需要先在 DCS 中创建 Redis 类型的缓存实例，获取 Redis 实例连接地址。具体操作请参见《分布式缓存服务用户指南》中的“快速入门”章节。

6.2.4 创建 Flink SQL 作业

本章节介绍如何新建 Flink SQL 作业。Flink SQL 是一种由用户根据自己的逻辑需求编写作业的方式。用 SQL 表达业务逻辑，可以简便快捷的实现业务。目前 Flink SQL 作业支持两种 Flink SQL 语句编辑方式：SQL 编辑器和可视化编辑器。本章节主要介绍使用 SQL 编辑器编写 Flink SQL 作业的方式。

前提条件

- 创建 Flink SQL 作业时，需要事先准备数据源以及数据输出通道，具体内容请参见 6.2.3 准备 Flink 作业数据。
- 创建 Flink SQL 作业，访问其他外部数据源时，如访问 OpenTSDB、HBase、Kafka、DWS、RDS、CSS、CloudTable、DCS Redis、DDS Mongo 等，需要先创建跨源连接，打通作业运行队列到外部数据源之间的网络。
 - 当前 Flink 作业支持访问的外部数据源详情请参考 10.1 跨源连接和跨源分析概述。
 - 创建跨源连接操作请参见 10.2 增强型跨源连接。创建完跨源连接后，可以通过“队列管理”页面，单击“操作”列“更多”中的“测试地址连通性”，验证队列到外部数据源之间的网络连通是否正常。详细操作可以参考 7.8 测试地址连通性。

创建 Flink SQL 作业

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入“Flink 作业”页面。

步骤 2 在“Flink 作业”页面右上角单击“创建作业”，弹出“创建作业”对话框。

步骤 3 配置作业信息。

表6-5 作业配置信息

参数	参数说明
----	------

参数	参数说明
类型	选择“Flink SQL”：用户通过编辑 SQL 语句来启动作业。
名称	作业名称，只能由字母、中文、数字、中划线和下划线组成，并且长度为 1~57 字节。 说明 作业名称必须是唯一的。
描述	作业的相关描述，长度为 0~512 字节。
模板名称	用户可以选择样例模板或自定义的作业模板。关于模板的详细信息，请参见 9.2 Flink 模板管理。

步骤 4 单击“确定”，进入作业“编辑”页面。

步骤 5 编辑 SQL 作业。

在 SQL 语句编辑区域，输入详细的 SQL 语句。相关 SQL 语句请参考《数据湖探索 SQL 语法参考》。

步骤 6 单击“语义校验”，确保语义校验成功。

- 只有语义校验成功后，才可以执行“调试”或“启动”作业的操作。
- 如果校验成功，提示“SQL 语义校验成功”。
- 如果校验失败，会在错误的 SQL 语句前面显示红色的“X”记号，鼠标移动到“X”号上可查看详细错误，请根据错误提示修改 SQL 语句。

步骤 7 设置作业运行参数。

表6-6 作业运行参数说明

参数	参数说明
所属队列	默认选择“共享队列”，用户也可以选择自定义的独享队列。 说明 <ul style="list-style-type: none"> • 当子用户在创建作业时，子用户只能选择已经被分配的队列。 • 当所选择队列的剩余容量不能满足作业需求时，系统会自动扩容。当队列空闲时，系统也会自动缩容。
UDF Jar	选择“独享队列”时可配置该参数。 用户自定义 UDF 文件，在选择 UDF Jar 之前需要将对应的 jar 包上传至 OBS 桶中，并在“数据管理 > 程序包管理”中创建程序包，具体操作请参考 8.2.3 创建程序包。 用户可以在 SQL 中调用插入 Jar 包中的自定义函数。
CU 数量	CU 数量为 DLI 的计算单元数量和管理单元数量总和，1CU=1 核 4G。 当前配置的 CU 数量为运行作业时所需的 CU 数，不能超过其绑

参数	参数说明
	定队列的 CU 数量。
管理单元	管理单元 CU 数量。
并行数	<p>并行数是指同时运行 Flink SQL 作业的最大任务数。适度增加并行数会提高作业整体算力，但也须考虑线程增多带来的切换开销。</p> <p>说明</p> <ul style="list-style-type: none"> • 最大并行数不能大于计算单元（CU 数量-管理单元）的 4 倍。 • 该界面并行数优先级低于代码中并行数设置。
TaskManager 配置	<p>用于设置 TaskManager 资源参数。</p> <p>勾选后需配置下列参数：</p> <ul style="list-style-type: none"> • “单 TM 所占 CU 数”：每个 TaskManager 占用的资源数量。 • “单 TM Slot”：每个 TaskManager 包含的 Slot 数量。
OBS 桶	<p>选择 OBS 桶用于保存用户作业日志信息、checkpoint 等信息。如果选择的 OBS 桶是未授权状态，需要单击“OBS 授权”。</p> <p>说明</p> <p>如果同时勾选了“开启 Checkpoint”和“保存作业日志”，OBS 授权一次即可。</p>
保存作业日志	<p>设置是否将作业运行时的日志信息保存到 OBS。日志信息的保存路径为：“桶名/jobs/logs/作业 id 开头的目录”。在作业列表中，单击对应的作业名称，在“运行日志”页签，可以单击页面提供的 OBS 链接跳转至对应的路径下。</p> <p>注意</p> <p>该参数建议勾选，否则作业运行完成后不会生成运行日志，后续如果作业运行异常则无法获取运行日志进行定位。</p> <p>勾选后需配置下列参数：</p> <p>“OBS 桶”：选择 OBS 桶用于保存用户作业日志信息。如果选择的 OBS 桶是未授权状态，需要单击“OBS 授权”。</p> <p>说明</p> <p>如果同时勾选了“开启 Checkpoint”和“保存作业日志”，OBS 授权一次即可。</p>
作业异常告警	<p>设置是否将作业异常告警信息，如作业出现运行异常或者欠费情况，以 SMN 的方式通知用户。</p> <p>勾选后需配置下列参数：</p> <p>“SMN 主题”：</p> <p>选择一个自定义的 SMN 主题。如何自定义 SMN 主题，请参见《消息通知服务用户指南》中“创建主题”章节。</p>
开启 Checkpoint	设置是否开启作业快照，开启后可基于 Checkpoint（一致性检查

参数	参数说明
	<p>点) 恢复作业。</p> <p>勾选后需配置下列参数：</p> <ul style="list-style-type: none"> • “Checkpoint 间隔”：Checkpoint 的时间间隔，单位为秒，输入范围 1~999999，默认值为 30s。 • “Checkpoint 模式”：支持如下两种模式： <ul style="list-style-type: none"> - At least once：事件至少被处理一次。 - Exactly once：事件仅被处理一次。 • “OBS 桶”：选择 OBS 桶用于保存用户 Checkpoint。如果选择的 OBS 桶是未授权状态，需要单击“OBS 授权”。 <p>Checkpoint 保存路径为：“桶名/jobs/checkpoint/作业 id 开头的目录”。</p> <p>说明</p> <p>如果同时勾选了“开启 Checkpoint”和“保存作业日志”，OBS 授权一次即可。</p>
异常自动重启	<p>设置是否启动异常自动重启功能，当作业异常时将自动重启并恢复作业。</p> <p>勾选后需配置下列参数：</p> <ul style="list-style-type: none"> • “异常重试最大次数”：配置异常重试最大次数。单位为“次/小时”。 <ul style="list-style-type: none"> - 无限：无限次重试。 - 有限：自定义重试次数。 • “从 Checkpoint 恢复”：需要同时勾选“开启 Checkpoint”才可配置该参数。
空闲状态保留时长	<p>用于清除 GroupBy 或 Window 经过最大保留时间后仍未更新的中间状态，默认设置为 1 小时。</p>
脏数据策略	<p>选择处理脏数据的策略。支持如下三种策略：“忽略”，“抛出异常”和“保存”。</p> <p>说明</p> <p>“保存”是指将脏数据保存到 OBS 桶中。</p>
脏数据转储地址	<p>“脏数据策略”选择“保存”时，配置该参数。单击地址框选择保存脏数据的 OBS 路径。</p>

步骤 8 (可选) 根据需要调试参数。作业调试功能只用于验证 SQL 逻辑，不会有数据写入操作。具体操作请参见 6.2.6 调试 Flink 作业。

步骤 9 (可选) 根据需要设置自定义配置。自定义配置为用户自定义。

步骤 10 单击“保存”，保存作业和相关参数。

步骤 11 单击“启动”，进入“启动 Flink 作业”页面，确认作业规格后，单击“立即启动”，启动作业。

启动作业后，系统将自动跳转到 Flink 作业管理页面，新创建的作业将显示在作业列表中，在“状态”列中可以查看作业状态。作业提交成功后，状态将由“提交中”变为“运行中”。运行完成后显示“已完成”。

如果作业状态为“提交失败”或“运行异常”，表示作业提交或运行失败。用户可以在作业列表中的“状态”列中，将鼠标移动到状态图标上查看错误信息，单击可以复制错误信息。根据错误信息解决故障后，重新提交。

说明

其他功能按钮说明如下：

- 另存为：将新建作业另存为一个新作业。
- 调试：对作业进行调试。具体操作请参见 6.2.6 调试 Flink 作业。
- 格式化：对 SQL 语句进行格式化。
- 设为模板：将新创建的作业设置为作业模板。
- 主题设置：设置页面主题，可以设置字体大小，自动换行和页面风格。

----结束

6.2.5 创建 Flink Jar 作业

本章节介绍如何新建 Flink Jar 作业。用户可以基于 Flink 的 API 进行二次开发，构建自己的应用 Jar 包并提交到 DLI 的队列运行，DLI 完全兼容开源社区接口。此功能需要用户自己编写并构建应用 Jar 包，适合对 Flink 二次开发有一定了解，并对流计算处理复杂度要求较高的用户。

前提条件

- 确保已创建独享队列。
- 创建 Flink Jar 作业，访问其他外部数据源时，如访问 OpenTSDB、HBase、Kafka、DWS、RDS、CSS、CloudTable、DCS Redis、DDS Mongo 等，需要先创建跨源连接，打通作业运行队列到外部数据源之间的网络。
 - 当前 Flink 作业支持访问的外部数据源详情请参考 10.1 跨源连接和跨源分析概述。
 - 创建跨源连接操作请参见 10.2 增强型跨源连接。

创建完跨源连接后，可以通过“队列管理”页面，单击“操作”列“更多”中的“测试地址连通性”，验证队列到外部数据源之间的网络连通是否正常。详细操作可以参考 7.8 测试地址连通性。
- 用户运行 Flink Jar 作业时，需要将二次开发的应用代码构建为 Jar 包，上传到已经创建的 OBS 桶中。并在 DLI “数据管理” > “程序包管理” 页面创建程序包，具体请参考 8.2.3 创建程序包。

说明

DLI 不支持下载功能，如果需要更新已上传的数据文件，可以将本地文件更新后重新上传。

- 由于 DLI 服务端已经内置了 Flink 的依赖包，并且基于开源社区版本做了安全加固。为了避免依赖包兼容性问题或日志输出及转储问题，打包时请注意排除以下文件：
 - 系统内置的依赖包，或者在 Maven 或者 Sbt 构建工具中将 scope 设为 provided
 - 日志配置文件（例如：“log4j.properties”或者“logback.xml”等）
 - 日志输出实现类 JAR 包（例如：log4j 等）

创建 Flink Jar 作业

- 步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入“Flink 作业”页面。
- 步骤 2 在“Flink 作业”页面右上角单击“新建作业”，弹出“新建作业”对话框。
- 步骤 3 配置作业信息。

表6-7 作业配置信息

参数	参数说明
类型	选择 Flink Jar。
名称	作业名称，只能由英文、中文、数字、中划线和下划线组成，并且长度为 1~57 字节。 说明 作业名称必须是唯一的。
描述	作业的相关描述，且长度为 0~512 字节。

- 步骤 4 单击“确定”，进入“编辑”页面。
- 步骤 5 选择队列。Flink Jar 作业只能运行在通用队列上。

📖 说明

- Flink Jar 作业只能运行在预先创建的独享队列上。
- 如果“所属队列”下拉框中无可用的独享队列，请先创建一个独享队列并将该队列绑定到当前用户。

- 步骤 6 配置 Flink Jar 作业参数。

表6-8 参数说明

名称	描述
应用程序	用户自定义的程序包。在选择程序包之前需要将对应的 Jar 包上传至 OBS 桶中，并在“数据管理>程序包管理”中创建程序包，具体操作请参考 8.2.3 创建程序包。内置依赖包请参考 Flink 1.7.2 依赖包 和 Flink 1.10 依赖包 。

名称	描述
主类	<p>指定加载的 Jar 包类名，如 <code>KafkaMessageStreaming</code>。</p> <ul style="list-style-type: none"> • 默认：根据 Jar 包文件的 Manifest 文件指定。 • 指定：必须输入“类名”并确定类参数列表（参数间用空格分隔）。 <p>说明</p> <p>当类属于某个包时，主类路径需要包含完整包路径，例如： <code>packagePath.KafkaMessageStreaming</code></p>
参数	<p>指定类的参数列表，参数之间使用空格分隔。</p> <p>Flink 参数支持全局变量替换。例如，在“全局配置”>“全局变量”中新增全局变量 <code>windowSize</code>，Flink Jar 作业就可以添加参数- <code>windowSize {{windowSize}}</code>。</p>
依赖 jar 包	<p>用户自定义的依赖程序包。依赖的相关程序包将会被放置到集群 <code>classpath</code> 下。</p> <p>在选择程序包之前需要将对应的 Jar 包上传至 OBS 桶中，并在“数据管理>程序包管理”中创建程序包，包类型选择“jar”。具体操作请参考 8.2.3 创建程序包。</p> <p>内置依赖包请参考 Flink 1.7.2 依赖包和 Flink 1.10 依赖包。</p>
其他依赖文件	<p>用户自定义的依赖文件。其他依赖文件需要自行在代码中引用。</p> <p>在选择依赖文件之前需要将对应的文件上传至 OBS 桶中，并在“数据管理>程序包管理”中创建程序包，包类型没有限制。具体操作请参考 8.2.3 创建程序包。</p> <p>通过在应用程序中添加以下内容可访问对应的依赖文件。其中，“<code>fileName</code>”为需要访问的文件名，“<code>ClassName</code>”为需要访问该文件的类名。</p> <pre>ClassName.class.getClassLoader().getResource("userData/fileName")</pre>
Flink 版本	选择 Flink 版本前，需要先选择所属的队列。当前支持“1.10”版本。
优化参数	<p>用户自定义的优化参数。参数格式为 <code>key=value</code>。</p> <p>Flink 优化参数支持全局变量替换。例如，在“全局配置”>“全局变量”中新增全局变量 <code>phase</code>，Flink Jar 作业就可以添加优化参数 <code>table.optimizer.agg-phase.strategy={{phase}}</code>。</p>

步骤 7 配置作业参数。

表6-9 参数说明

名称	描述
CU 数量	一个 CU 为 1 核 4G 的资源量。CU 数量范围为 2~400 个。
管理单元	设置管理单元的 CU 数，支持设置 1~4 个 CU，默认值为 1 个 CU。

名称	描述
并行数	<p>作业中每个算子的最大并行数。</p> <p>说明</p> <ul style="list-style-type: none"> 并行数不能大于计算单元（CU 数量-管理单元 CU 数量）的 4 倍。 并行数最好大于用户作业里设置的并发数，否则有可能提交失败。
TaskManager 配置	<p>用于设置 TaskManager 资源参数。</p> <p>勾选后需配置下列参数：</p> <ul style="list-style-type: none"> “单 TM 所占 CU 数”：每个 TaskManager 占用的资源数量。 “单 TM Slot”：每个 TaskManager 包含的 Slot 数量。
保存作业日志	<p>设置是否将作业运行时的日志信息保存到 OBS 桶。</p> <p>注意</p> <p>该参数建议勾选，否则作业运行完成后不会生成运行日志，后续如果作业运行异常则无法获取运行日志进行定位。</p> <p>勾选后需配置下列参数：</p> <p>“OBS 桶”：选择 OBS 桶用于保存用户作业日志信息。如果选择的 OBS 桶是未授权状态，需要单击“OBS 授权”。</p>
作业异常告警	<p>设置是否将作业异常告警信息，如作业出现运行异常或者欠费情况，以 SMN 的方式通知用户。</p> <p>勾选后需配置下列参数：</p> <p>“SMN 主题”：</p>
异常自动重启	<p>设置是否启动异常自动重启功能，当作业异常时将自动重启并恢复作业。</p> <p>勾选后需配置下列参数：</p> <ul style="list-style-type: none"> “异常重试最大次数”：配置异常重试最大次数。单位为“次/小时”。 <ul style="list-style-type: none"> 无限：无限次重试。 有限：自定义重试次数。 “从 Checkpoint 恢复”：从保存的 checkpoint 恢复作业。 <p>勾选该参数后，还需要选择“Checkpoint 路径”。</p> <p>“Checkpoint 路径”：选择 checkpoint 保存路径。必须和应用程序中配置的 Checkpoint 地址相对应。且不同作业的路径不可一致，否则无法获取准确的 Checkpoint。</p>

步骤 8 单击右上角“保存”，保存作业和相关参数。

步骤 9 单击右上角“启动”，进入“启动 Flink 作业”页面，确认作业规格，单击“立即启动”，启动作业。

启动作业后，系统将自动跳转到 Flink 作业管理页面，新创建的作业将显示在作业列表中，在“状态”列中可以查看作业状态。作业提交成功后，状态将由“提交中”变为“运行中”。运行完成后显示“已完成”。

如果作业状态为“提交失败”或“运行异常”，表示作业提交或运行失败。用户可以在作业列表中的“状态”列中，将鼠标移动到状态图标上查看错误信息，单击可以复制错误信息。根据错误信息解决故障后，重新提交。

说明

其他功能按钮说明如下：

另存为：将新建作业另存为一个新作业。

----结束

6.2.6 调试 Flink 作业

调试作业功能可以帮助用户在正式运行作业前，检测用户编写的 SQL 语句逻辑的正确性。

说明

- 目前只有 Flink SQL 作业支持本功能。
- 作业调试功能只用于验证 SQL 逻辑，不会有数据写入操作。

调试作业步骤

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入“Flink 作业”页面。

步骤 2 在已创建的 Flink SQL 作业“操作”列中，单击“编辑”，进入 Flink SQL 作业编辑页面。

如果是正在创建的新作业，可直接在“编辑”页面中，对作业进行调试。

步骤 3 单击“编辑”页面的“调试”，对已经编辑好的 SQL 语句进行解析。“编辑”页面右侧自动跳转到“调试参数”页签。

- “调试结果转储地址”：选择 OBS 桶用于保存调试日志，如果选择未授权的桶，单击“OBS 授权”进行授权。
- “数据输入方式”：可选择保存在 OBS 桶中的 CSV 数据，也可手动录入。
 - 选择“OBS (CSV)”方式
用户将 OBS 服务作为数据源，需要与 OBS 服务对接，在使用 DLI 服务前，需要先准备 OBS 数据，具体请参见 6.2.3 准备 Flink 作业数据。OBS 方式的数据输入格式为 CSV，以换行分隔多条记录，以逗号分隔单条记录中不同字段。同时，用户需要选择 OBS 中具体的对象作为输入源数据。
 - 选择“手动录入”方式
用户编写 SQL 语句作为数据源。手动录入方式需要填写单条记录的每一个字段值。

步骤 4 单击“开始调试”，调试完毕会弹出“调试结果”页面。

- 如果调试结果符合预期，说明作业可以正常运行。
- 如果调试结果不符合预期，说明作业业务逻辑可能存在问题，请修改 SQL 语句后再重新调试。

----结束

6.2.7 操作 Flink 作业

用户创建了新作业后，需要根据用户的实际需求对作业进行操作，具体操作有如下几种：

- [编辑作业](#)
- [启动作业](#)
- [停止作业](#)
- [删除作业](#)
- [导出作业](#)
- [导入作业](#)
- [名称和描述修改](#)
- [导入保存点](#)
- [触发保存点](#)
- [运行时配置](#)

编辑作业

用户可以对已经创建的作业进行编辑，如修改 SQL 语句、作业名称和描述、作业配置信息等。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 在需要编辑作业对应的“操作”列中，单击“编辑”，进入作业“编辑”页面。

步骤 3 根据实际需求编辑作业。

对于 Flink SQL 类型作业，编辑页面详情请参见 6.2.4 创建 Flink SQL 作业中的步骤 5~步骤 7。

对于 Flink 自定义类型作业，编辑页面详情请参见 6.2.5 创建 Flink Jar 作业中的步骤 5~步骤 7。

----结束

启动作业

用户可以启动已创建保存的作业或已经停止的作业。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 启动作业，有以下两种方式：

- 启动单个作业
选择一个作业，在对应的“操作”列中，单击“启动”。
也可以在作业列表中，勾选一个作业，单击作业列表左上方的“启动”。
- 批量启动作业
勾选多个作业，单击作业列表左上方的“启动”，可以启动多个作业。
单击“启动”后，跳转至“作业配置清单”页面。

步骤 3 在“作业配置清单”页面，确认作业信息，若无误，单击“立即启动”。

作业启动后，可在对应作业的“状态”列中查看运行成功或失败。

----结束

停止作业

当用户不需要运行某个作业时，用户可以将状态为“运行中”和“提交中”的作业停止。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 停止作业，有以下两种方式：

- 停止单个作业
选择需要停止的作业，在对应的“操作”列中，单击“更多 > 停止”。
也可以在作业列表中，勾选一个作业，单击作业列表左上方的“停止”。
- 批量停止作业
勾选多个需要停止作业，单击作业列表左上方的“停止”。可以停止多个作业。

步骤 3 在弹出的“停止作业”窗口中，单击“确认”，停止作业。

说明

- 在停止作业之前，用户可以触发保存点，保存作业的状态信息。当该作业再次启动时用户可以选择是否从保存点恢复。
- 勾选“触发保存点”表示创建保存点。不勾选“触发保存点”表示不创建保存点。默认不创建保存点。
- 保存点的生命周期从触发保存点并停止作业开始，重启作业后结束。保存点在重启作业后自动删除，不会一直保存。

停止作业过程中，在作业列表的“状态”列中将显示作业状态，说明如下：

- 如果在“状态”中显示“停止中”，表示正在停止作业。
- 如果在“状态”中显示“已停止”，表示停止作业成功。

- 如果在“状态”中显示“停止失败”，表示停止作业失败。

----结束

删除作业

当用户不再需要使用某个作业时，可以参考如下操作删除该作业。作业删除后，将不可恢复，请谨慎操作。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 删除作业，有以下两种方式：

- 删除单个作业

在需要删除作业对应的“操作”列中，单击“更多 > 删除”，弹出“删除作业”页面。

也可以在作业列表中，勾选一个作业，单击作业列表左上方的“删除”，弹出“删除作业”页面。

- 批量删除作业

勾选多个需要删除作业，单击作业列表左上方的“删除”，弹出“删除作业”页面，可以删除多个作业。

步骤 3 单击“确定”，完成作业的删除。

----结束

导出作业

用户可以将所创建的 Flink 作业导出至 OBS 桶中。

适用于当用户切换区域、项目或用户时，需要创建相同的作业，而作业比较多的情况。此时，不需要重新创建作业，只需要将原有的作业导出，再在新的区域、项目或者使用新的用户登录后，导入作业即可。

说明

切换项目或用户时，需要对新项目或用户授权，具体请参考 6.2.2 Flink 作业权限管理。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击右上角“导出作业”，打开“导出作业”对话框。

步骤 3 选择保存作业的 OBS 桶。单击“下一步”。

步骤 4 选择待导出的作业。

默认导出所有作业，也可以勾选“自定义导出”选择需要导出的作业。

步骤 5 单击“确认导出”，完成导出作业。

----结束

导入作业

用户可以将保存在 OBS 桶中的 Flink 作业配置文件导入至 DLI 的 Flink 作业管理中。

适用于当用户切换区域、项目或用户时，需要创建相同的作业，而作业较多的情况。此时，不需要重新创建作业，只需要将原有的作业导出，再在新的区域、项目或者使用新的用户登录后，导入作业即可。

如果需要导入自建的作业，建议使用创建作业的功能。具体请参考 6.2.4 创建 Flink SQL 作业，6.2.5 创建 Flink Jar 作业。

说明

- 切换项目或用户时，需要对新项目或用户授权，具体请参考 6.2.2 Flink 作业权限管理。
- 仅支持导入与从 DLI 导出的 Flink 作业相同数据格式的作业。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击右上角“导入作业”，打开“导入作业”对话框。

步骤 3 选择需导入的作业配置文件的完整 OBS 路径。单击“下一步”。

步骤 4 配置同名作业策略。单击“下一步”。

- 勾选“配置同名替换”，若待导入的作业名已存在，则覆盖已存在的作业配置，并且作业状态重置为草稿。
- 不勾选“配置同名替换”，若待导入的作业名已存在，则不导入同名作业的配置。

步骤 5 确认“配置文件”和“同名作业策略”配置无误。单击“确认导入”，完成导入作业。

----结束

名称和描述修改

用户可以根据需要修改作业名称和描述。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 在需要修改名称和描述的作业对应的“操作”列中，单击“更多 > 名称和描述修改”，弹出“属性修改”页面。修改作业名称和描述。

步骤 3 单击“确定”完成修改。

----结束

触发保存点

状态为“运行中”的作业可以“触发保存点”，保存作业的状态信息。

导入保存点

导入保存点用于恢复作业状态。关于保存点，可参见 flink 官方文档 [Checkpointing 说明](#)。

具体操作为：选择保存点的 OBS 路径，选择完成后，单击“确认”完成导入。

运行时配置

用户可通过选择“运行时配置”配置作业异常告警和重启选项。

📖 说明

支持 Flink SQL 作业，Flink jar 作业。

1. 在对应 Flink 作业“操作”列>“更多”中单击“运行时配置”。
2. 在“运行时配置”页面配置以下参数。

表6-10 作业运行参数说明

参数	参数说明
作业异常告警	<p>设置是否将作业异常告警信息，如作业出现运行异常或者欠费情况，以 SMN 的方式通知用户。</p> <p>勾选后需配置下列参数：</p> <p>“SMN 主题”：</p> <p>选择一个自定义的 SMN 主题。如何自定义 SMN 主题，请参见《消息通知服务用户指南》中“创建主题”章节。</p>
异常自动重启	<p>设置是否启动异常自动重启功能，当作业异常时将自动重启并恢复作业。</p> <p>勾选后需配置下列参数：</p> <ul style="list-style-type: none">• “异常重试最大次数”：配置异常重试最大次数。单位为“次/小时”。<ul style="list-style-type: none">- 无限：无限次重试。- 有限：自定义重试次数。• “从 Checkpoint 恢复”：从已保存的 checkpoint 恢复作业。 <p>说明</p> <p>Flink SQL 作业需要在编辑页面勾选“开启 Checkpoint”才可配置该参数。</p> <p>勾选该参数后，Flink jar 作业还需要选择“Checkpoint 路径”。</p> <p>“Checkpoint 路径”：选择 checkpoint 保存路径。必须和应用程序中配置的 Checkpoint 地址相对应。且不同作业的路径不可一致，否则无法获取准确的 Checkpoint。</p>

6.2.8 Flink 作业详情

创建作业后，用户可以通过查看作业详情，了解如下信息：

- [查看作业详情](#)
- [查看作业监控](#)
- [查看作业任务列表](#)
- [查看作业执行计划](#)
- [查看提交作业日志](#)
- [查看作业运行日志](#)

查看作业详情

用户作业创建完成并保存后，用户可以单击作业名查看作业的详细信息，包括作业的 SQL 语句和参数设置信息，如果是 jar 作业只可以看到参数设置信息。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击需要查看的作业名称，进入“作业详情”页面。

在“作业详情”页签，用户可以查看作业的 SQL 语句、参数设置信息。

以某个 Flink SQL 作业为例进行说明。

表6-11 参数说明

参数名称	参数说明
类型	作业类型，如 Flink SQL 作业。
名称	Flink 作业名称。
描述	Flink 作业描述信息。
状态	Flink 作业运行状态。
运行模式	如果作业所属队列是共享队列，则作业的运行模式是共享模式。 如果作业所属队列是自定义的独享队列，则作业的运行模式是独享模式。
所属队列	如果作业所属队列是共享队列，则显示共享队列。 如果作业所属队列是自定义的独享队列，则显示具体队列名称。
UDF Jar	作业选择非共享队列，并且配置 UDF Jar 参数时显示。
优化参数	作业添加了自定义参数配置时显示该参数。
CU 数量	作业配置的 CU 数量。

参数名称	参数说明
管理单元	作业配置的管理单元 CU 数量。
并行数	作业配置的同时运行 Flink 作业的任务数。
单 TM 所占 CU 数	作业配置的每个 TaskManager 所占 CU 数量。
单 TM Slot 数	作业配置的每个 TaskManager Slot 数量。
OBS 桶	作业配置了 OBS 桶显示。当勾选“开启 Checkpoint”和“保存作业日志”参数后，用于保存 Checkpoint 和作业运行日志。
保存作业日志	开启或关闭。
作业异常告警	开启或关闭。
SMN 主题	作业配置的 SMN 主题名称。当作业勾选了“作业异常告警”参数时显示。
异常自动重启	开启或关闭。
异常重试最大次数	异常重试最大次数为无限时显示为-1，其余为设置的具体值。
从 Checkpoint 恢复	开启或关闭。
ID	作业 ID。
保存点路径	保存点的 OBS 存储路径。
开启 Checkpoint	开启或关闭。
Checkpoint 间隔	将作业运行的中间结果保存到 OBS 的间隔时间，单位为秒。
Checkpoint 模式	Checkpoint 模式： <ul style="list-style-type: none"> at_least_once: 事件至少被处理一次 exactly_once: 事件仅被处理一次
空闲状态保留时长	用于清除 GroupBy 或 Window 经过最大保留时间后仍未更新的中间状态。
脏数据策略	作业出现脏数据时的处理策略。作业配置脏数据策略时显示该参数。 忽略。 抛出异常。 保存。
脏数据转储地址	“脏数据策略”选择“保存”时，保存脏数据的 OBS 路径。
创建时间	作业创建的具体时间。
更新时间	作业最近一次更新的时间。

----结束

查看作业监控

用户可以通过云监控服务（CES）查看作业数据输入输出的详细信息。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击需要查看的作业名称，进入“作业详情”页面。

单击页面右上角的“作业监控”，将跳转至云监控服务（CES）。

Flink 作业包含如下监控指标。

表6-12 Flink 作业监控指标

指标名称	说明
Flink 作业数据输入速率	展示用户 Flink 作业的数据输入速率，供监控和调试使用。 单位：条/秒。
Flink 作业数据输出速率	展示用户 Flink 作业的数据输出速率，供监控和调试使用。 单位：条/秒。
Flink 作业数据输入总数	展示用户 Flink 作业的数据输入总数，供监控和调试使用。 单位：条。
Flink 作业数据输出总数	展示用户 Flink 作业的数据输出总数，供监控和调试使用。 单位：条。
Flink 作业字节输入速率	展示用户 Flink 作业每秒输入的字节数。单位：字节/秒。
Flink 作业字节输出速率	展示用户 Flink 作业每秒输出的字节数。单位：字节/秒。
Flink 作业字节输入总数	展示用户 Flink 作业字节的输入总数。单位：字节。
Flink 作业字节输出总数	展示用户 Flink 作业字节的输出总数。单位：字节。
Flink 作业 CPU 使用率	展示用户 Flink 作业的 CPU 使用率。单位：%。
Flink 作业内存使用率	展示用户 Flink 作业的内存使用率。单位：%。
Flink 作业最大算子延迟	展示用户 Flink 作业的最大算子延迟时间，单位 ms。
Flink 作业最大算子反压	展示用户 Flink 作业的最大算子反压值，数值越大，反压越严重。 0 ：表示 OK

指标名称	说明
	50: 表示 Low 100: 表示 High

----结束

查看作业任务列表

用户可以查看作业运行时每个任务的详细信息，例如任务的开始时间、收发字节数和运行时长等。

说明

如果数据为零，表示没有从数据源接收到数据。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击需要查看的作业名称，进入“作业详情”页面。

步骤 3 在“任务列表”页签，可以查看任务的节点信息。

查看算子任务列表，具体参见下表：

表6-13 算子任务列表参数

参数	说明
名称	算子名称。
持续时间	算子运行的持续时间。
最大并行数	算子中并行的 Task 的个数。
任务	算子的任务有以下几种： <ul style="list-style-type: none"> • 红色数字表示已失败的 Task 个数。 • 浅灰色数字表示已取消的 Task 个数。 • 黄色数字表示取消中的 Task 个数。 • 绿色数字表示已完成的 Task 个数。 • 蓝色数字表示运行中的 Task 个数。 • 天蓝色数字表示部署中的 Task 个数。 • 深灰色数字表示排队中的 Task 个数。
状态	算子任务对应的状态。
反压状态	算子的工作负荷状态。包含如下几种状态： <ul style="list-style-type: none"> • OK：表示工作负荷正常。 • LOW：表示工作负荷略高。DLI 处理数据的速度比较快。

参数	说明
	<ul style="list-style-type: none">HIGH: 表示工作负荷高。源端输入数据的速度比较慢。
时延	指事件从源端算子到达本算子的过程中消耗的时间，单位为毫秒（ms）。
发送的记录数	算子发送数据的记录。
发送的字节数	算子发送的字节数。
接受的字节数	算子接收的字节数。
接受的记录数	算子收到数据的记录。
开始时间	算子运行开始时间。
结束时间	算子运行结束时间。

----结束

查看作业执行计划

用户通过查看执行计划了解到运行中的作业的算子流向。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击需要查看的作业名称，进入“作业详情”页面。

步骤 3 单击“执行计划”，在“执行计划”页签可以查看作业的算子流向。

单击对应的节点，在页面右侧显示对应的信息。

- 滚动鼠标滚轮可对流图进行缩放查看。
- 流图展示当前运行作业的实时算子流图信息。

----结束

查看提交作业日志

用户可以通过查看提交日志排查提交作业异常的故障。

步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。

步骤 2 单击需要查看的作业名称，进入“作业详情”页面。

步骤 3 在“提交日志”页签，可以查看提交作业的过程信息。

----结束

查看作业运行日志

用户可以通过查看运行日志排查作业运行异常的故障。

- 步骤 1 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入 Flink 作业管理页面。
- 步骤 2 单击需要查看的作业名称，进入“作业详情”页面。
- 步骤 3 在“运行日志”页签，可以查看运行中作业的 JobManager 和 TaskManager 信息。

JobManager 和 TaskManager 信息每分钟刷新一次，默认展示最近一分钟的运行日志。

如果作业配置了保存作业日志的 OBS 桶，更多历史日志信息可以到保存日志的 OBS 桶中下载查看。

如果作业没有运行，则无法查看 TaskManager 信息。

----结束

6.3 Spark 作业管理

6.3.1 Spark 作业管理概述

DLI 在开源 Spark 基础上进行了大量的性能优化与服务化改造，兼容 Apache Spark 生态和接口，执行批处理任务。

DLI 还支持使用 Spark 作业访问 DLI 元数据。

Spark 作业管理主要包括如下功能：

- 6.3.2 创建 Spark 作业
- [重新执行作业](#)
- [查找作业](#)
- [终止作业](#)

以及查看“使用指南”和“使用视频”。

作业管理页面

在总览页面单击“Spark 作业”简介，或在左侧导航栏单击“作业管理”>“Spark 作业”，可进入 Spark 作业管理页面。Spark 作业管理页面显示所有的 Spark 作业，作业数量较多时，系统分页显示，您可以查看任何状态下的作业。

表6-14 作业管理参数

参数	参数说明
作业 ID	所提交 Spark 作业的 ID，由系统默认生成。
名称	所提交 Spark 作业的名称。

参数	参数说明
队列	所提交 Spark 作业所在的队列。
用户名	执行 Spark 作业的用户名称。
状态	作业的状态信息，包括如下。 <ul style="list-style-type: none"> • 启动中：正在启动 • 运行中：正在执行任务 • 已失败：session 已退出 • 已成功：session 运行成功 • 恢复中：正在恢复任务
创建时间	每个作业的创建时间，可按创建时间顺序或倒序显示作业列表。
最后修改时间	作业运行完成的时间。
操作	<ul style="list-style-type: none"> • 编辑：可修改当前作业配置，重新执行作业。 • SparkUI：单击后，将跳转至 Spark 任务运行情况界面。 <p>说明</p> <ul style="list-style-type: none"> • 状态为“启动中”的作业不能查看 SparkUI 界面。 • 目前 DLI 配置 SparkUI 只展示最新的 100 条作业信息。 • 终止作业：终止启动中和运行中的作业。 • 重新执行：重新运行该作业。 • 归档日志：将作业日志保存到系统创建的 DLI 临时数据桶中。 • 导出日志：将日志导出至用户创建的 OBS 桶中进行查看。 <p>说明</p> <ul style="list-style-type: none"> • 用户需要具有创建 OBS 桶的权限。 • 当作业状态在“运行中”时，不能导出日志。 • 提交日志：查看提交作业的日志。 • Driver 日志：查看运行作业的日志。

重新执行作业

在“Spark 作业”页面，单击对应作业“操作”列中的“编辑”，跳转至“Spark 作业编辑”页面，可根据需要修改参数，执行作业。

查找作业

在“Spark 作业”页面，选择“状态”或“队列”。系统将根据设置的过滤条件，在作业列表显示符合对应条件的作业。

终止作业

在“Spark 作业”页面，单击对应作业“操作”列中的“更多”>“终止作业”，可停止启动中和运行中的作业。

导出日志

在“Spark 作业”页面，单击对应作业“操作”列中的“更多”>“导出日志”，在弹窗中输入已创建的 OBS 桶地址，单击“确定”。

6.3.2 创建 Spark 作业

Spark 作业编辑页面支持执行 Spark 作业，为用户提供全托管式的 Spark 计算服务。

在总览页面，单击 Spark 作业右上角的“创建作业”，或在 Spark 作业管理页面，单击右上角的“创建作业”，均可进入 Spark 作业编辑页面。

进入 Spark 作业编辑页面，页面会提示系统将创建 DLI 临时数据桶。该桶用于存储使用 DLI 服务产生的临时数据，例如：作业日志、作业结果等。如果不创建该桶，将无法查看作业日志。桶名称为系统默认。

如果不需要创建 DLI 临时数据桶，并且希望不再收到该提示，可以勾选“下次不再提示”并单击“取消”。

前提条件

- 请先将所要依赖的程序包通过“数据管理>程序包管理”页面上传至对应的 OBS 桶中。具体操作请参考 8.2.3 创建程序包。
- 创建 Spark 作业，访问其他外部数据源时，如访问 OpenTSDB、HBase、Kafka、DWS、RDS、CSS、CloudTable、DCS Redis、DDS Mongo 等，需要先创建跨源连接，打通作业运行队列到外部数据源之间的网络。
 - 当前 Spark 作业支持访问的外部数据源详情请参考 10.1 跨源连接和跨源分析概述。
 - 创建跨源连接操作请参见 10.2 增强型跨源连接。创建完跨源连接后，可以通过“队列管理”页面，单击“操作”列“更多”中的“测试地址连通性”，验证队列到外部数据源之间的网络连通是否正常。详细操作可以参考 7.8 测试地址连通性。

界面说明

- 左侧导航栏
在创建 Spark 作业页面，左侧导航栏包括“队列”页签和“程序包”页签。

表6-15 左侧导航栏说明

序号	页签/按键	页签/按键名称	描述
1	队列	队列	显示已有的队列。

序号	页签/按键	页签/按键名称	描述
2		程序包	显示已有的程序包。
3		模板	Spark 作业模板。
4		创建	创建队列/程序包。
5		刷新	包括刷新已有的队列和程序包列表。
6		搜索	在程序包页签，可以输入程序包名称进行搜索。

- 作业编辑窗口

在作业编辑窗口，可以选择使用“表单模式”或者“API 模式”进行参数设置。

以下以“表单模式”页面进行说明，“API 模式”即采用 API 接口模式设置参数及参数值，具体请参考《数据湖探索 API 参考》。

- 选择运行队列：具体参数请参考表 6-16。

表6-16 运行队列参数说明

参数名称	参数描述
队列	下拉选择要使用的队列。

- 作业配置：具体参数请参考表 6-17。

表6-17 作业配置参数说明

参数名称	参数描述
作业名称	设置作业名称。
应用程序	选择需要执行的程序包。包括“.jar”和“.py”两种类型。
主类	输入主类名称。当应用程序类型为“.jar”时，主类名称不能为空。
应用程序参数	<p>用户自定义参数，多个参数请以 Enter 键分隔。</p> <p>应用程序参数支持全局变量替换。例如，在“全局配置”>“全局变量”中新增全局变量 key 为 batch_num，可以使用{{batch_num}}，在提交作业之后进行变量替换。</p>
Spark 参数	<p>以“key=value”的形式设置提交 Spark 作业的属性，多个参数以 Enter 键分隔。</p> <p>Spark 参数 value 支持全局变量替换。例如，在“全局配置”>“全局变量”中新增全局变量 key 为 custom_class，可以使用 "spark.sql.catalog"={{custom_class}}，在提交作业之后进行变量替</p>

参数名称	参数描述
	换。 说明 Spark 作业不支持自定义设置 jvm 垃圾回收算法。
依赖 jar 包	运行 spark 作业依赖的 jars。可以输入 jar 包名称，也可以输入对应 jar 包文件的的 OBS 路径，格式为：obs://桶名/文件夹路径名/包名。
依赖 python 文件	运行 spark 作业依赖的 py-files。可以输入 Python 文件名称，也可以输入 Python 文件对应的 OBS 路径，格式为：obs://桶名/文件夹路径名/文件名。
其他依赖文件	运行 spark 作业依赖的其他 files。可以输入依赖文件名称，也可以输入对应的 OBS 路径，格式为：obs://桶名/文件夹路径名/文件名。
依赖分组	在创建程序包时，如果选择了分组，在此处选择对应的分组，则可以同时选中该分组中的所有程序包和文件。创建程序包操作请参考 8.2.3 创建程序包。
访问元数据	是否通过 Spark 作业访问元数据。。
是否重试	作业失败后是否进行重试。 选择“是”需要配置以下参数： “最大重试次数”：设置作业失败重试次数，最大值为“100”。
高级配置	<ul style="list-style-type: none"> • 暂不配置 • 现在配置：包括以下两项参数 <ul style="list-style-type: none"> - 选择依赖资源：具体参数请参考表 6-18。 - 计算资源规格：具体参数请参考表 6-19。

表6-18 选择依赖资源参数说明

参数名称	参数描述
Module 名称	DLI 系统提供的用于执行跨源作业的依赖模块访问各个不同的服务，选择不同的模块： <ul style="list-style-type: none"> • MRS HBase: sys.datasource.hbase • DDS: sys.datasource.mongo • MRS OpenTSDB: sys.datasource.opentsdb • DWS: sys.datasource.dws • RDS MySQL: sys.datasource.rds • RDS PostGre: sys.datasource.rds • DCS: sys.datasource.redis • CSS: sys.datasource.css
资源包	运行 spark 作业依赖的 jar 包。

表6-19 计算资源规格参数说明

参数名称	参数描述
资源规格	下拉选择所需的资源规格。系统提供 3 种资源规格供您选择。资源规格中如下配置项支持修改： <ul style="list-style-type: none">• Executor 内存• Executor CPU 核数• Executor 个数• driver CPU 核数• driver 内存 最终配置结果以修改后数据为准。
Executor 内存	在所选资源规格基础上自定义 Executor 内存规格。
Executor CPU 核数	在所选资源规格基础上自定义 Executor CPU 核数。
Executor 个数	在所选资源规格基础上自定义 Executor 个数。
driver CPU 核数	在所选资源规格基础上自定义 Driver CPU 核数。
driver 内存	在所选资源规格基础上自定义 Driver 内存规格。

说明

Spark 作业参数计算：

- $CU \text{ 数} = \text{driver CPU 核数} + \text{Executor 个数} * \text{Executor CPU 核数}$

注意：因为本身 DLI 集群管理面和 driver 都会占用一部分 CU 资源，“Executor 个数 * Executor CPU 核数”实际要小于队列的计算资源 CU 数。

- $\text{内存数} = \text{driver 内存} + (\text{Executor 个数} * \text{Executor 内存})$

创建 Spark 作业步骤

1. 在 Spark 作业编辑页面中，输入相关参数，具体请参考关于 Spark 作业编辑窗口的说明。
2. 单击 Spark 作业编辑页面右上方“执行”，提交作业，页面显示“批处理作业提交成功”。
3. （可选）可在“Spark 作业”管理页面查看提交作业的状态及日志。

7 队列管理

7.1 队列管理概述

队列

DLI 服务中的队列即为计算资源，计算资源是使用 DLI 服务的基础，用户执行的一切作业都需要使用计算资源。

当前 DLI 服务包括“SQL 队列”和“通用队列”两种队列类型。其中，SQL 队列用于运行 SQL 作业，通用队列兼容老版本的 Spark 队列，用于运行 Spark 作业和 Flink 作业。

说明

- SQL 队列每隔 12 小时会自动重启一次。
- 不支持不同类型队列直接转换，如需更换队列，请重新购买新的队列。
- default 队列是系统预置的默认队列，不需要单独删除。

计算资源和存储资源的区别

表7-1 计算资源和存储资源的区别

资源名称	获取方式	主要作用
计算资源	在 DLI 管理控制台创建队列。	执行查询。
存储资源	DLI 服务自带 5G 配额。	存储数据库和 DLI 表。

说明

- 存储资源是 DLI 服务内部的存储资源，用于存储数据库和 DLI 表，体现用户存储在 DLI 中的数据量。
- 存储资源无需创建，DLI 默认提供 5G 配额。

- DLI 服务预置了名为“default”的队列。用户在不明确所需队列容量或没有可创建队列空间的情况下，可以使用该队列执行作业。
- “default”队列只用于用户体验，所有人共用，用户间可能会出现抢占资源的情况，不能保证每次都可以得到资源执行相关操作。建议使用自建队列执行作业。

专属队列

专属队列是指队列对应的资源为专属资源，空闲时不释放，即无论是否使用均保留资源的队列类型。专属队列可以保证提交作业时资源一定存在。

队列弹性扩缩容

DLI 提供了队列弹性扩缩容的功能。用户在创建指定规格队列后，可根据需要进行弹性扩缩容。

根据业务情况，手动更改队列规格。具体操作请参考 7.6 弹性扩缩容。

说明

新创建的队列需要运行作业后才可进行扩缩容。

队列弹性扩缩容定时任务

DLI 提供了队列弹性扩缩容定时任务的功能。用户在创建队列后，可根据需要进行弹性扩缩容定时任务。

根据业务情况，设置队列自动扩缩容的时间，由系统定时触发队列扩缩容。具体操作请参考 7.7 弹性扩缩容定时任务。

说明

新创建的队列需要运行作业后才可进行扩缩容。

队列自动扩缩容

Flink 作业使用队列，DLI 可根据作业大小自动触发扩缩容，用户无需进行操作。

说明

新创建的队列需要运行作业后才可进行扩缩容。

队列管理页面

队列管理主要包括如下功能：

- 7.2 队列权限管理
- 7.3 创建队列
- 7.4 删除队列
- 7.5 修改队列网段
- 7.6 弹性扩缩容

- 7.7 弹性扩缩容定时任务
- 7.8 测试地址连通性
- 7.9 创建消息通知主题

📖 说明

DLI 作业执行失败需要通过 SMN 发送通知消息，因此需要获得访问和使用 SMN（消息通知服务）的 SMN Administrator 权限。

队列管理页面显示用户创建所有的队列和服务预置的 default 队列。队列列表默认按创建时间排列，创建时间最近的队列显示在最前端。

表7-2 队列管理参数

参数	参数说明
名称	队列的名称。
类型	队列的类型。 <ul style="list-style-type: none"> • SQL 队列 • 通用队列 • Spark 队列（兼容老版本）
规格	队列大小，单位：CUs。 CUs 是队列的计价单位。1CUs=1Core 4GMem。不同规格的队列对应的计算能力不一样，规格越高计算能力越好。
实际 CUs	当前队列实际大小值。
弹性扩缩容	定时扩缩容的目标 CU 值，或当前规格 CU 值的最大值和最小值。
用户名	队列所有者。
描述	创建队列时，对队列的描述。若无描述，则显示“--”。
操作	<ul style="list-style-type: none"> • 删除：删除所选队列。如果队列中有正在提交或者正在运行的作业，将不支持删除操作。 • 权限管理：查看队列对应的用户权限信息以及对其他用户授权。 • 更多 <ul style="list-style-type: none"> - 重启：强制重启队列。 <p>说明</p> <p>只有 SQL 队列有“重启”操作。</p> <ul style="list-style-type: none"> - 弹性扩缩容：可以根据需要选择“扩容”或“缩容”，目标值大小必须为 16CU 的整数倍。 - 弹性扩缩容定时任务：可以根据业务周期或使用情况，在不同的时间或周期内设置不同的队列大小，系统将定时自动进行“扩容”或“缩容”。目标值大小必须为 16CU 的整数倍。 - 修改网段：使用 DLI 增强型跨源时，DLI 队列网段与数据源网段不能重合，可根据需要进行修改。

参数	参数说明
	<ul style="list-style-type: none"> 测试地址连通性：测试队列到指定地址是否可达，支持域名和 ip，可指定端口。

7.2 队列权限管理

队列权限管理操作场景

- 针对不同用户，可以通过权限设置分配不同的队列，不同用户之间的作业效率互不影响，保障作业性能。
- 管理员用户和队列的所有者拥有所有权限，不需要进行权限设置且其他用户无法修改其队列权限。

队列权限相关操作步骤

步骤 1 在 DLI 管理控制台的左侧，选择“队列管理”。

步骤 2 选择待设置的队列，单击其“操作”列中的“权限管理”。“用户权限信息”区域展示了当前具备此队列权限的用户列表。

权限设置有 3 种场景：为新用户赋予权限，为已有权限的用户修改权限，回收某用户具备的所有权限。

- 为新用户赋予权限
新用户指之前不具备此队列权限的用户。
 - a. 单击“权限信息”右侧的“授权”，弹出“授权”对话框。
 - b. 填写“用户名”，并勾选对应权限。
 - c. 单击“确定”，完成新用户的添加。
 待设置的参数说明如表 7-3 所示。

表7-3 参数说明

参数名称	描述
用户名	被授权的用户名称。 说明 该用户名称是已存在的 IAM 用户名称且该用户登录过 DLI 管理控制台。
权限设置	<ul style="list-style-type: none"> • 删除队列：删除此队列。 • 提交作业：向此队列提交作业。 • 终止作业：终止提交到此队列的作业。 • 赋权：当前用户可将队列的权限赋予其他用户。 • 回收：当前用户可回收其他用户具备的该队列的权限，但不能

参数名称	描述
	<p>回收该队列所有者的权限。</p> <ul style="list-style-type: none">• 查看其他用户具备的权限：当前用户可查看其他用户具备的该队列的权限。• 重启队列权限：重启此队列的权限。• 规格变更：修改队列规格的权限。

- 为已有权限的用户赋予权限或回收权限。
 - a. 在对应队列“权限信息”区域的用户列表中，选择需要修改权限的用户，在“操作”列单击“权限设置”。
 - b. 在队列“权限设置”对话框中，对当前用户具备的权限进行修改。详细权限描述如表 7-3 所示。

当“权限设置”中的选项为灰色时，表示您不具备修改此队列权限的权限。可以向管理员用户、队列所有者等具有赋权权限的用户申请“队列的赋权”和“队列权限的回收”权限。
 - c. 单击“确定”完成权限设置。
- 回收某用户具备的所有权限。

在对应队列“权限信息”区域的用户列表中，选择需要删除权限的用户，在“操作”列单击“回收用户权限”。在“回收用户权限”对话框中单击“确定”后，此用户将不具备该队列的任意权限。

----结束

7.3 创建队列

执行作业前需要先创建队列。

说明

- 用户首次使用子帐号创建队列时，需要先使用主帐号登录控制台，在 DLI 的数据库中保持记录，才能创建队列。
- 新队列第一次运行作业时，需要一定的时间，通常为 6~10 分钟。
- 队列创建完成后，如果在 1 小时内未运行作业，系统将进行释放。

创建队列步骤

1. 创建队列的操作入口有三个，分别在“总览”页面、“SQL 编辑器”页面和“队列管理”页面。
 - 单击总览页面右上角“创建队列”进行创建队列。
 - 在“队列管理”页面创建队列。
 - i. 在 DLI 管理控制台的左侧导航栏中，选择“队列管理”。
 - ii. 单击“队列管理”页面右上角“创建队列”进行创建队列。

- 在“SQL 编辑器”页面创建队列。
 - i. 在 DLI 管理控制台的左侧导航栏中，选择“SQL 编辑器”。
 - ii. 在左侧导航栏的队列页签，单击右侧的  创建队列。

表7-4 参数说明

参数名称	描述
名称	队列的名称。 <ul style="list-style-type: none"> • 只能包含数字、英文字母和下划线，但不能是纯数字，不能以下划线开头，且不能为空。 • 输入长度不能超过 128 个字符。 说明 队列名称不区分大小写，系统会自动转换为小写。
类型	<ul style="list-style-type: none"> • SQL 队列：SQL 作业的计算资源。 • 通用队列：Spark 作业、Flink 作业的计算资源。 说明 可勾选“专属资源模式”，该模式下可创建增强型跨源。
规格	按需选择队列规格。1CU=1 核 4GB。队列规格指的是计算节点所有 CU 数的总和，DLI 系统会自动分配各计算节点的内存和 CPU 大小，具体计算节点个数客户端不感知。 <ul style="list-style-type: none"> • “固定规格”包括“16CU”、“64CU”、“256CU”、“512CU”。
描述	所创建队列的相应描述。输入长度不能超过 256 个字符。
高级选项	在“队列类型”中，勾选了“专属资源模式”后，需要选择“高级选项”。 <ul style="list-style-type: none"> • 默认配置：由系统自动配置。 • 自定义配置： <ul style="list-style-type: none"> “网段”：指定使用的网段范围，请参考 7.5 修改队列网段。如需使用 DLI 增强型跨源，DLI 队列网段与数据源网段不能重合。 “队列特性”：运行 AI 相关 SQL 作业时选择“AI 增强型”队列，运行其他作业时选择“基础型”队列。

2. 单击“立即创建”，完成队列创建。
 队列创建成功后，您可以在“队列管理”页面查看和选择使用对应的队列。

说明

新队列第一次运行作业时，需要一定的时间，通常为 6~10 分钟。

7.4 删除队列

根据实际使用情况，您可以通过删除操作释放队列。

📖 说明

- 如果待删除的队列中有正在提交或正在运行的作业，将不支持删除操作。
- 删除队列不会导致您数据库中的表数据丢失。

删除队列步骤

步骤 1 在 DLI 管理控制台左侧，选择“队列管理”。

步骤 2 选择待删除的队列，单击“操作”列的“删除”删除。

📖 说明

如果“操作”列的“删除”为灰色，表示当前用户没有删除队列的权限。您可以向管理员申请删除队列的权限。

步骤 3 在弹出的确认对话框中，单击“是”。

----结束

7.5 修改队列网段

使用增强型跨源时，如果 DLI 队列的网段和用户数据源的网段发生冲突，您可以通过修改网段操作更改队列的网段。

📖 说明

如果待修改网段的队列中有正在提交或正在运行的作业，或者改队列已经绑定了增强型跨源，将不支持修改网段操作。

修改队列网段步骤

1. 在 DLI 管理控制台左侧，选择“队列管理”。
2. 选择待修改的队列，单击“操作”列“更多”中的“修改网段”。
3. 填写需要的网段后，单击“确定”。

7.6 弹性扩缩容

前提条件

新创建的队列需要运行作业后才可进行弹性扩缩容。

注意事项

- 如果在“弹性扩缩容”页面提示“Status of queue xxx is assigning, which is not available”，表示需要等待队列资源分配完毕才可进行扩缩容。

弹性扩容

当前队列规格不满足业务需要时，可以通过手动变更队列规格来扩容当前队列。

📖 说明

扩容属于耗时操作，在 DLI “弹性扩缩容” 页面执行扩容操作后，需要等待大约 10 分钟，具体时长和扩容的 CU 值有关，等待一段时间后，可以通过刷新“队列管理”页面，对比“规格”和“实际 CUs”大小是否一致来判断是否扩容成功。或者在“作业管理”页面，查看“SCALE_QUEUE”类型 SQL 作业的状态，如果作业状态为“弹性扩缩容中”，表示队列正在扩容中。

操作步骤如下：

1. 在 DLI 管理控制台左侧，选择“队列管理”。
2. 选择需要扩容的队列，单击“操作”列“更多”中的“弹性扩缩容”。
3. 在“弹性扩缩容”页面，“变更方式”选择“扩容”，设置扩容的 CU 值。
4. 确认无误后，单击“确定”。

弹性缩容

当计算业务较小，不需要那么大的队列规格时，可以通过手动变更队列规格来缩容当前队列。

📖 说明

- 缩容属于耗时操作，在 DLI “弹性扩缩容” 页面执行缩容操作后，需要等待大约 10 分钟，具体时长和缩容的 CU 值有关，等待一段时间后，可以通过刷新“队列管理”页面，对比“规格”和“实际 CUs”大小是否一致来判断是否缩容成功。或者在“作业管理”页面，查看“SCALE_QUEUE”类型 SQL 作业的状态，如果作业状态为“弹性扩缩容中”，表示队列正在缩容中。
- 系统不保证完全缩容到设定的目标大小。如果当前队列正在使用或者队列业务量比较大，会出现缩容不成功，或者缩容一部分规格的情况。
- 系统默认最小 CU 值为 16CU，即当队列规格为 16CUs 时，不能进行手动缩容。

操作步骤如下：

1. 在 DLI 管理控制台左侧，选择“队列管理”。
2. 选择需要扩容的队列，单击“操作”列“更多”中的“弹性扩缩容”。
3. 在“弹性扩缩容”页面，“变更方式”选择“缩容”，设置缩容的 CU 值。
4. 确认无误后，单击“确定”。

7.7 弹性扩缩容定时任务

弹性扩缩容定时任务使用场景

通常，用户业务繁忙的场景是有周期性的，在某个周期内，用户需要更多的计算资源来处理业务，过了这个周期，则不需要那么多资源。如果用户购买的队列规格比较小，在业务繁忙时会存在资源不足的情况；而如果购买的队列规格比较大，又可能会存在资源浪费的情况。

基于以上场景，DLI 提供了队列弹性扩缩容定时任务功能。用户可以根据自己的业务周期或者使用情况，基于现有队列规格，在不同的时间或者周期内设置不同的队列大小，以满足自己的业务需求，节约成本。

使用弹性扩缩容定时任务注意事项

- 新创建的队列需要运行作业后才可进行扩缩容。
- 目前只支持规格为 64CUs 以上的队列进行定时弹性扩缩容任务，即队列最小规格为 64CUs。
- 对于每个队列，最多支持创建 12 个定时任务。
- 每个定时任务开始时，弹性扩缩容的实际开始的时间有 5 分钟误差。建议扩容时间定时至少比实际使用队列的时间提前 20 分钟。
- 每个定时任务之间需要至少有 2 小时的间隔。
- 队列的定时弹性扩缩容属于耗时操作，变更所消耗的时间取决于扩缩容目标规格与当前规格的差值大小，用户在“队列管理”页面中可以查看当前队列的规格。
- 如果当前队列有作业正在运行时，可能无法缩容到目标 CU 值，而是缩容到当前队列规格和目标规格中间的某个值，系统将在 1 小时后继续尝试进行缩容，直至下一个定时任务开始。
- 当一个定时任务没有扩容或者缩容到目标 CU 值时，系统会在约 15 分钟后再次触发扩缩计划，直到下一个定时任务开始。
- 定时任务的“任务名称”、“最终 CUs”和“执行时间”是创建和修改定时任务时的必选项，用户可根据业务需要进行填写。
- 定时任务的“重复规律”支持按周为周期进行调度。
 - 默认不选，表示“不重复”，即只在“执行时间”执行一次；
 - 如果全选，表示该计划每天都会执行；
 - 如果选择部分，则选择规律的计划在对应的时间每周都会被执行一次。

创建弹性扩缩容定时任务

- 如果只设定扩容或者缩容，只需创建一个弹性扩缩容定时任务。设定“任务名称”、“最终 CUs”和“执行时间”即可，具体请参考表 7-5。
- 如果需要同时设定扩容和缩容，则需要创建两个弹性扩缩容定时任务，分别设定扩容和缩容的“任务名称”、“最终 CUs”和“执行时间”，具体请参考表 7-5。

操作步骤如下：

1. 在 DLI 管理控制台左侧，选择“队列管理”。

2. 选择需要设置弹性扩缩容定时任务的队列，单击“操作”列“更多”中的“弹性扩缩容定时任务”。
3. 在“弹性扩缩容定时任务”页面，单击右上角的“创建定时任务”。
4. 在“创建定时任务”页面，设置参数。单击“确定”。

表7-5 参数说明

参数名称	描述
任务名称	<p>输入定时任务的名称。</p> <ul style="list-style-type: none"> • 只能包含数字、英文字母和下划线，但不能是纯数字，不能以下划线开头，且不能为空。 • 输入长度不能超过 128 个字符。
激活任务	<p>激活队列扩缩容定时任务。默认开启。如果关闭，则系统不会触发执行当前设置的定时规格变更任务。</p>
有效期	<p>设置执行定时任务的时间段。包括“日期”和“时间”。如果没有时间段的限制，该参数值为空即可，表示永久生效，如果用户需要指定当前任务的生效时间段，可根据业务需要进行选择。</p> <p>说明</p> <ul style="list-style-type: none"> • “有效期”中的“开始时间”需要晚于当前的系统时间。 • 如果只设置了扩容，在“有效期”结束之后，系统不会自动缩容，需要手动修改或设置缩容定时任务。反之亦然。即为单次执行定时扩缩容。 • 如果同时设置了扩容和缩容，在有效期内会按照设定扩缩容，在“有效期”结束之后，将保持最后一次设定的队列规格。
实际 CUs	<p>队列扩容或缩容前的规格。</p>
最终 CUs	<p>队列扩容或缩容后的规格。</p> <p>说明</p> <ul style="list-style-type: none"> • 系统默认队列最大规格为 512CUs。 • 进行定时扩缩容操作的队列最小规格为 64CUs，即当“实际 CUs”小于 64CUs 时，不能进行定时扩缩容。 • 最终规格只能为 16 的倍数。
重复规律	<p>选择执行定时扩缩容的周期。</p> <p>说明</p> <ul style="list-style-type: none"> • 如果只是单次执行扩容或者缩容，无需选择“执行周期”。 • 如果同时设置了扩缩容，可根据需要选择“执行周期”，还可与“有效期”进行配合使用。
执行时间	<p>执行定时扩容或者缩容的时间。</p>

定时任务创建后，可以在“弹性扩缩容定时任务”页面查看当前队列的规格变化情况，以及计划最近一次的执行时间。

或者在“队列管理”页面，查看“规格”大小是否改变来判断是否扩缩容成功。

或者在“作业管理”页面，查看“SCALE_QUEUE”类型作业的状态，如果作业状态为“规格变更中”，表示队列正在扩缩容中。

修改弹性扩缩容定时任务

如果设定的定时任务不再满足业务需求，可以在“弹性扩缩容定时任务”页面，单击对应定时任务“操作”列中的“修改”，进入“创建定时任务”页面进行修改。

删除弹性扩缩容定时任务

如果不再需要定时修改队列规格，可以在“弹性扩缩容定时任务”页面，单击对应定时任务“操作”列中的“删除”，进入“删除定时任务”页面，单击“是”，进行删除。

7.8 测试地址连通性

可用于测试普通场景中 DLI 队列与用户指定对端地址之间的连通性，或者跨源连接场景中 DLI 队列与跨源连接绑定的对端地址之间连通性。具体操作如下：

1. 在“队列管理”页面，选择需要测试地址连通性的队列，单击“操作”列“更多”中的“测试地址连通性”。
2. 在“测试地址连通性”页面填写需要测试的地址。支持域名和 ip，可指定端口。
3. 单击“测试”。

如果测试地址可连通，页面上将提示地址可达。

如果测试地址不可连通，页面上将提示地址不可达，请检查网络配置后重试。检查网络配置即检查所测试的 VPC 对等连接或跨源连接是否处于已激活状态。

7.9 创建消息通知主题

确定创建消息通知主题后，您可在消息通知服务的“主题管理”页面中，对对应的主题“添加订阅”，选择不同方式（例如短信或者邮件等）进行订阅；订阅成功后，若作业失败，则系统将会自动发送消息到您指定的订阅终端。具体操作如下：

1. 在队列管理页面，单击左上角“创建消息通知主题”，弹出对话框。
2. 选择队列，单击“确定”。弹出对话框。

说明

- 选择队列时，可以选择单个队列，也可以选择所有队列。如果单个队列和所有队列的终端不一致，那么，即选择了单个队列，又选择了所有队列进行订阅时，所有队列的消息通知中将不包含那个单队列的消息。

- 创建消息通知主题后，只有在创建 session 或者创建 batch 失败时才会收到消息通知。
3. 单击“主题管理”，跳转至消息通知服务“主题管理”页面。
 4. 在对应主题的“操作”栏中，单击“添加订阅”，选择“协议”，确定订阅方式。
 - 如果选择“邮件”协议，需要在“订阅终端”中填写接收确认邮件的邮箱地址。
 5. 通过单击邮件中的链接确认后，将收到“订阅成功”的信息。
 6. 返回“主题管理”所在页面，单击左侧“订阅”，对应的订阅状态为“已确认”，表示订阅成功。

8.1 库表管理

8.1.1 库表管理概述

库表管理主要包括如下功能：

- 8.1.2 数据库权限管理
- 8.1.3 表权限管理
- 8.1.4 创建数据库和表
- 8.1.5 删除数据库和表
- 8.1.6 修改数据库和表所有者
- 8.1.7 导入数据
- 8.1.8 将 DLI 数据导出至 OBS
- 8.1.9 查看元数据
- 8.1.10 预览数据

DLI 表和 OBS 表的区别

- DLI 表的数据存储在 DLI 服务内部，查询性能更好，适用于对时延敏感类的业务，如交互类的查询等。
- OBS 表的数据存储在 OBS 上，适用于对时延不敏感的业务，如历史数据统计分析等。

库表管理页面

库表管理页面显示用户创建所有的数据库，您可以查看所有者，表数量等信息。

表8-1 库表管理参数

参数	参数说明
数据库名称	<ul style="list-style-type: none">• 数据库名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。

参数	参数说明
	<ul style="list-style-type: none"> 数据库名称大小写不敏感且不能为空。 输入长度不能超过 128 个字符。
用户名	数据库所有者。
表数量	对应数据库中表的数量。
描述	创建数据库时，对数据库的描述。如果没有描述，则显示“--”。
企业项目	显示所属的企业项目。企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理。
操作	<ul style="list-style-type: none"> 权限管理：可查看“权限信息”，进行“授权”、“权限设置”和“回收用户权限”。 表管理：可查看对应数据库下的表。具体描述请参见表管理页面。 创建表：在对应数据库中创建表。 删除数据库：删除所选数据库。 修改数据库：修改数据库的所有者。必须是相同账户下已存在的用户名。

表管理页面

在“数据管理”页面中，单击对应数据库名称或“操作”列中的“表管理”，可进入其“表管理”页面。

“表管理”页面显示用户在当前数据库中创建所有的表，您可以查看表类型，数据存储位置等信息。表列表默认按创建时间排列，创建时间最近的表显示在最前端。

表8-2 表管理参数

参数	参数说明
表名称	<ul style="list-style-type: none"> 表名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 表名称大小写不敏感且不能为空。 表名称支持包含“\$”符号。例如：\$test。 输入长度不能超过 128 个字符。
表类型	表的类型。包括以下三种类型： <ul style="list-style-type: none"> Managed：数据存储位置为 DLI 的表。 External：数据存储位置为 OBS 的表。 View：视图。视图只能通过 SQL 语句创建。 说明

参数	参数说明
	视图中包含的表或视图信息不可被更改，如有更改可能会造成查询失败。
所有者	创建表的用户。
存储位置	包括 DLI、OBS、View、CloudTable、CSS 等数据存储位置。
表大小	表中数据的大小。只有 Managed 类型表会显示具体数值，其他类型表显示 "--"。
数据源路径	<ul style="list-style-type: none"> • 数据存储位置为 OBS 时，显示对应的 OBS 路径。 • 数据存储位置为 DLI 和 View 时，显示 "--"。 • 数据存储位置为 CloudTable、CSS 等跨源服务时，显示其对应的 URL。
创建时间	创建表的时间。
最后访问时间	最后对表进行操作的时间。
操作	<ul style="list-style-type: none"> • 权限管理：可查看“权限信息”，进行“授权”、“权限设置”和“回收用户权限”。 • 更多： <ul style="list-style-type: none"> - 删除：在对应数据库中删除表。 - 修改所有者：修改表的所有者。必须是相同账户下已存在的用户名。 - 导入：将存储在 OBS 桶中的数据导入至 DLI 表或 OBS 表。 - 导出：仅支持将 DLI 表的数据导出至 OBS 桶中。 - 表属性：可查看“元数据”和“预览”表中的数据。

8.1.2 数据库权限管理

数据库权限操作场景

- 针对不同用户，可以通过权限设置分配不同的数据库，不同用户之间的作业效率互不影响，保障作业性能。
- 管理员用户和数据库的所有者拥有所有权限，不需要进行权限设置且其他用户无法修改其数据库权限。

注意事项

- 数据库和表赋权对象具有层级关系，用户赋予上一层级的权限会自动继承到下一层级对象上，层级关系为：数据库>表>列。
- 数据库所有者、表所有者、被赋予“赋权权限”的用户都可以对数据库和表赋权。
- 列只能继承查询权限。“可继承权限”详细信息请参见 8.1.2 数据库权限管理。

- 回收权限时，只能在初始赋权的层级上回收。在哪一层赋权的，在哪一层进行权限回收。赋予权限和回收权限需要在同一层级操作。例如：在数据库上给用户赋予插入权限，那么在数据库下面的表就有了插入权限，回收这个插入权限，只能在数据库上回收，不能在表上回收。
- 如果数据库被删除后，再重新创建同名的数据库，数据库权限不会继承，需要对操作该数据库的用户或项目重新进行赋权。

例如，testdb 数据库给用户 A 赋予了删除数据库的权限，后续执行了删除 testdb 数据库，并重新创建了 testdb 数据库。如果希望 A 用户继续保留删除 testdb 数据库的权限，则需要重新对 A 用户赋予该权限。

查看数据库权限

- 在管理控制台左侧，单击“数据管理”>“库表管理”。
- 单击所选数据库“操作”栏中的“权限管理”，将显示该数据库对应的权限信息。

权限设置有 3 种场景：为新用户或项目赋予权限、为已有权限的用户或项目修改权限、回收某用户或项目具备的所有权限。

为新用户或项目赋予权限

为新用户或新项目赋予权限，新用户或新项目指之前不具备此数据库权限的用户或项目。

- 在数据库权限管理页面右上角单击“授权”。
- 在“授权”弹出框中，选择“用户授权”或“项目授权”，填写需要授权的用户名或选择需要授权的项目，选择相应的权限。具体权限说明请参考表 8-3。

表8-3 参数说明

参数	描述
授权对象	选择“用户授权”或“项目授权”。
用户名/项目名	<ul style="list-style-type: none"> 选择“用户授权”时，输入数据库新增用户对应的 IAM 用户名称。 <p>说明</p> <p>该用户名称是已存在的 IAM 用户名称且该用户登录过 DLI 管理控制台。</p> <ul style="list-style-type: none"> 选择“项目授权”时，选择当前区域下需要授权的项目。 <p>说明</p> <p>选择“项目授权”时：</p> <ul style="list-style-type: none"> 如果赋权选择不可继承权限，则在该项目中无法查看对应数据库中表的信息。 如果赋权选择可继承权限，则在该项目中可查看该数据库内所有表的信息。
非继承权限	选中权限即对用户或项目进行赋权，取消勾选即对用户权限或项目权限进行回收。

参数	描述
	<p>非继承权限只作用于当前数据库。</p> <ul style="list-style-type: none"> • 同时适用于“用户授权”和“项目授权”的权限包括： <ul style="list-style-type: none"> - 删除数据库：删除当前数据库。 - 创建表：在当前数据库创建表。 - 创建视图：在当前数据库创建视图。 - 将 SQL 语句解释为执行计划：执行 explain 语句。 - 创建角色：在当前数据库创建角色。 - 删除角色：删除当前数据库中的角色。 - 显示角色：显示当前用户的角色。 - 绑定角色：在当前数据库绑定角色。 - 解除角色绑定：在当前数据库解除角色绑定。 - 显示所有角色和用户的绑定关系：显示所有角色和用户的绑定关系。 - 创建函数：在当前数据库创建函数。 - 删除函数：删除当前数据库中的函数。 - 显示所有函数：显示当前数据库中的所有函数。 - 显示函数详情：显示当前函数详情。 • 只适用于“用户授权”的权限包括： <ul style="list-style-type: none"> - 显示所有表：显示当前数据库下的所有表。 - 显示数据库：显示当前数据库的信息。
继承权限	<p>选中权限即对用户进行赋权，取消勾选即对用户权限进行回收。</p> <p>继承权限可作用到当前数据库及其所有的表上，但是表中的列只能继承其中的查询权限。</p> <p>以下权限同时适用于“用户授权”和“项目授权”。</p> <ul style="list-style-type: none"> • 删除表：删除数据库下的表。 • 查询表：在当前表内查询。 • 显示表信息：显示当前表的信息。 • 插入：在当前表内插入数据。 • 添加列：在当前表中增加列。 • 重写：在当前表内插入覆盖数据。 • 赋权：用户可将数据库的权限赋予其他用户或项目。 • 回收：用户可回收其他用户或项目具备的此数据库的权限，但是不能回收数据库所有者的权限。 • 添加分区：在分区表中添加新的分区。 • 删除分区：删除分区表中已有的分区。 • 设置分区的路径：将分区表中的某个分区路径设置为用户指定的 OBS 路径。

参数	描述
	<ul style="list-style-type: none">• 重命名表分区：对分区表中的分区重新命名。• 重命名表：对表重新命名。• 恢复表分区：从文件系统中导出分区信息保存到元数据中。• 显示所有分区：显示分区表中的所有分区。• 查看其他用户具备的权限：查看其他用户或项目具备的当前数据库的权限。

3. 单击“确定”，完成授权。

为已有权限的用户或项目修改权限

某用户或项目已具备此数据库的一些权限时，可为此用户或项目赋予或取消权限。

说明

当“权限设置”中的选项为灰色时，表示对应帐号不具备修改此数据库的权限。可以向管理员用户、数据库所有者等具有赋权权限的用户申请数据库的“赋权”和数据库权限的“回收”权限。

1. 在“用户权限信息”列表中找到需要设置权限的用户：
 - 若用户为子用户，可进行“权限设置”。
 - 若用户为管理员用户，只能查看“权限信息”。

在“项目权限信息”列表中找到需要设置权限的项目，进行“权限设置”。

2. 在子用户或项目的“操作”栏中单击“权限设置”，可弹出数据库“权限设置”对话框。
数据库用户或项目详细的权限描述请参考表 8-3。
3. 单击“确定”，完成权限设置。

回收某用户或项目具备的所有权限

回收某用户具备的所有权限，或回收某项目具备的所有权限。

- 在“用户权限信息”区域的用户列表中，选择需要回收权限的子用户，在“操作”栏中单击“回收”，在“回收用户权限”对话框中单击“确定”后，此用户将不具备数据库的任意权限。

说明

用户为管理员用户时，“回收”为灰色，表示不可回收该用户的权限。

- 在“项目权限信息”区域的项目列表中，选择需要回收权限的项目，在“操作”栏中单击“回收”，在“回收项目权限”对话框中单击“确定”后，此项目将不具备数据库的任意权限。

8.1.3 表权限管理

表权限操作场景

- 针对不同用户，可以通过权限设置分配不同的数据库，不同用户之间的作业效率互不影响，保障作业性能。
- 管理员用户和数据库的所有者拥有所有权限，不需要进行权限设置且其他用户无法修改其数据库权限。
- 给新用户设置数据库权限时，该用户所在用户组需具有 Tenant Guest 权限。

注意事项

- 如果表被删除后，再重新创建同名的表，表权限不会继承，需要对操作该表的用户和项目重新进行权限赋予。
例如，testTable 表给用户 A 赋予了删除表的权限，后续执行了删除 testTable 表，并重新创建了 testTable 表。如果希望 A 用户继续保留删除 testTable 表的权限，则需要重新对 A 用户赋予该权限。

查看表权限

1. 在管理控制台左侧，单击“数据管理”>“库表管理”。
2. 单击需要设置权限的表所在的数据库名，进入该数据库的“表管理”页面。
3. 单击所选表“操作”栏中的“权限管理”，将显示该表对应的权限信息。
表权限设置有 3 种场景：为新用户或项目赋予权限，为已有权限的用户或项目修改权限，回收某用户或项目具备的所有权限。

为新用户或项目赋予权限

为新用户或项目赋予权限，新用户或项目指之前不具备此表任何权限的用户或项目。

1. 单击表权限管理页面右上角的“授权”按钮。
2. 在弹出的“授权”对话框中选择相应的权限。
 - DLI 表具体权限说明请参考表 8-4。

表8-4 参数配置

参数	描述
授权对象	选择“用户授权”或“项目授权”。
用户名/项目	<ul style="list-style-type: none">• 选择“用户授权”时，输入表新增用户对应 IAM 用户名称。 说明 该用户名称是已存在的 IAM 用户名称且该用户登录过 DLI 管理控制台。• 选择“项目授权”时，选择当前区域下需要授权的项目。 说明 选择“项目授权”时，只能查看被授权的表及其所在数据库的信息。

参数	描述
非继承权限	<p>选中权限即对用户或项目进行赋权，取消勾选即对用户权限或项目权限进行收回。</p> <ul style="list-style-type: none"> 同时适用于“用户授权”和“项目授权”的权限包括： <ul style="list-style-type: none"> 查询表：在当前表内查询数据。 显示表信息：显示当前表的信息。 显示创建表语句：显示当前表的创建语句。 删除表：删除当前表。 重命名表：对当前表重新命名。 插入：在当前表内插入数据。 重写：在当前表内插入覆盖数据。 添加列：在当前表中增加列。 赋权：当前用户可将表的权限赋予其他用户。 回收：当前用户可回收其他用户具备的此表的权限，并且不能回收表所有者的权限。 查看其他用户具备的权限：查看其他用户具备的当前表的权限。 分区表还具有以下权限： <ul style="list-style-type: none"> 删除分区：删除分区表中的分区。 显示所有分区：显示分区表中的所有分区。 只适用于“用户授权”的权限包括： <ul style="list-style-type: none"> 显示表：显示当前表。

- OBS 表具体权限说明请参考表 8-5。

表8-5 参数配置

参数	描述
授权对象	选择“用户授权”或“项目授权”。
用户名/项目	<ul style="list-style-type: none"> 选择“用户授权”时，输入表新增用户对应 IAM 用户名称。 说明 该用户名称是已存在的 IAM 用户名称且该用户登录过 DLI 管理控制台。 选择“项目授权”时，选择当前区域下需要授权的项目。 说明 选择“项目授权”时，只能查看被授权的表及其所在数据库的信息。
非继承权限	<p>选中权限即对用户或项目进行赋权，取消勾选即对用户权限或项目权限进行收回。</p> <ul style="list-style-type: none"> 同时适用于“用户授权”和“项目授权”的权限包括：

参数	描述
	<ul style="list-style-type: none"> - 显示创建表语句：显示当前表的创建语句。 - 显示表信息：显示当前表的信息。 - 查询表：在当前表内查询数据。 - 删除表：删除当前表。 - 重命名表：对当前表重新命名。 - 插入：在当前表内插入数据。 - 重写：在当前表内插入覆盖数据。 - 添加列：在当前表中增加列。 - 赋权：当前用户可将表的权限赋予其他用户或项目。 - 回收：当前用户或项目可回收其他用户或项目具备的此表的权限，并且不能回收表所有者的权限。 - 查看其他用户具备的权限：查看其他用户具备的当前表的权限。 <p>分区表还具有以下权限：</p> <ul style="list-style-type: none"> - 添加分区：在分区表中添加新的分区。 - 删除分区：删除分区表中的任意分区。 - 设置分区的路径：将分区表中的某个分区路径设置为用户指定的 OBS 路径。 - 重命名表分区：对分区表中的分区重新命名。 - 恢复表分区：从文件系统中导出分区信息保存到元数据中。 - 显示所有分区：显示分区表中的所有分区。 <ul style="list-style-type: none"> • 只适用于“用户授权”的权限包括： <ul style="list-style-type: none"> - 显示表：显示当前表。

- View 具体权限说明请参考表 8-6。

📖 说明

View 只能通过 SQL 语句进行创建，不能通过“创建表”页面进行创建。

表8-6 参数配置

参数	描述
授权选择	选择“用户授权”或“项目授权”。
用户名/项目	<ul style="list-style-type: none"> • 选择“用户授权”时，输入表新增用户对应 IAM 用户名称。 说明 该用户名称是已存在的 IAM 用户名称且该用户登录过 DLI 管理控制台。 • 选择“项目授权”时，选择当前区域下需要授权的项目。 说明

参数	描述
	选择“项目授权”时，只能查看被授权的表及其所在数据库的信息。
非继承权限	<p>选中权限即对用户或项目进行赋权，取消勾选即对用户权限或项目权限进行收回。</p> <ul style="list-style-type: none"> • 同时适用于“用户授权”和“项目授权”的权限包括： <ul style="list-style-type: none"> - 显示表信息：显示当前表的信息。 - 显示创建表语句：显示当前表的创建语句。 - 删除表：删除当前表。 - 查询表：在当前表内查询数据。 - 重命名表：对当前表重新进行命名。 - 赋权：当前用户或项目可将表的权限赋予其他用户或项目。 - 回收：当前用户或项目可回收其他用户或项目具备的此表的权限，并且不能回收表所有者的权限。 - 查看其他用户具备的权限：查看其他用户具备的当前表的权限。 • 只适用于 <ul style="list-style-type: none"> - 显示表：显示当前表。

3. 单击“确定”，完成表权限设置。

为已有权限的用户或项目修改权限

某用户或项目已具备此表的一些权限时，可为此用户或项目赋予或回收权限。

📖 说明

当“权限设置”中的选项为灰色时，表示您不具备修改此表的权限。可以向管理员用户、表所有者等具有赋权权限的用户申请表的“赋权”和表权限的“回收”权限。

1. 在“用户权限信息”列表中找到需要设置权限的用户：
 - 若用户为子用户且不是表的所有者，可进行“权限设置”。
 - 若用户为管理员用户或表的所有者，只能查看“权限信息”。

在“项目权限信息”列表中找到需要设置权限的项目，进行“权限设置”。
2. 在子用户或项目的“操作”栏中单击“权限设置”，可弹出表“权限设置”对话框。
 - DLI 表用户或项目权限说明请参考表 8-4。
 - OBS 表用户或项目权限说明请参考表 8-5。
 - View 用户或项目权限说明请参考表 8-6。
3. 单击“确定”，完成表权限设置。

回收某用户或项目具备的所有权限

回收某用户具备的所有权限，或回收某项目具备的所有权限。

- 在“用户权限信息”区域的用户列表中，选择需要回收权限的子用户，在“操作”栏中单击“回收”，确定后，此用户将不具备表的任意权限。

📖 说明

以下情况中，“回收”为灰色，表示不可回收该用户的权限。

- 用户为管理员用户
- 子用户是表的所有者
- 子用户只有可继承权限
- 在“项目权限信息”区域的项目列表中，选择需要回收权限的项目，在“操作”栏中单击“回收”，确定后，此项目将不具备表的任意权限。

📖 说明

当项目只有可继承权限时，“回收”为灰色，表示不可回收该项目的权限。

8.1.4 创建数据库和表

数据库和表说明

数据库是按照数据结构来组织、存储和管理数据的建立在计算机存储设备上的仓库。

表是数据库最重要的组成部分之一。表是由行与列组合成的。每一列被当作是一个字段。每个字段中的值代表一种类型的数据。

数据库是一个框架，表是其实质内容。一个数据库包含一个或者多个表。

DLI 中数据库的概念、基本用法与 Oracle 数据库基本相同，是 DLI 管理权限的基础单元，赋权以数据库为单位。

用户可通过管理控制台或 SQL 语句创建数据库和表。本章节介绍在管理控制台创建数据库和表的操作步骤。

📖 说明

View 只能通过 SQL 语句进行创建，不能通过“创建表”页面进行创建。

注意事项

- 当 OBS 的目录下有同名文件夹和文件时，创建 OBS 表指向该路径会优先指向文件而非文件夹。

创建数据库

步骤 1 创建数据库的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面创建数据库。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。

- b. 在库表管理页面右上角，单击“创建数据库”可创建数据库。
- 在“SQL 编辑器”页面创建数据库。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。
 - b. 在左侧导航栏单击“数据库”页签右侧  可创建数据库。

步骤 2 在“创建数据库”页面，参见表 8-7 输入数据库名称和描述信息。

表8-7 参数说明

参数名称	描述	示例
数据库名称	<ul style="list-style-type: none"> • 数据库名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 • 数据库名称大小写不敏感且不能为空。 • 输入长度不能超过 128 个字符。 	DB01
描述	该数据库的描述。	-

说明

“default”为内置数据库，不能创建名为“default”的数据库。

步骤 3 单击“确定”，完成数据库创建。

数据库创建成功后，您可以在“库表管理”页面或者“SQL 编辑器”页面查看和选择使用对应的数据库。

----结束

创建表

创建表前，请确保已创建数据库。

步骤 1 创建表的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

说明

此处创建表的方式不支持创建 View, HBase (MRS) 表、OpenTSDB (MRS) 表、DWS 表、RDS 表和 CSS 表等跨源连接表。可通过 SQL 方式创建 View 和跨源连接表，具体请参考《数据湖探索 SQL 语法参考》中关于“创建视图”和“创建跨源连接表”的章节。

- 在“数据管理”页面创建表。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。
 - b. 在库表管理页面中，选择需要建表的数据库。在其“操作”栏中，单击“更多”>“创建表”，可创建当前数据库下的表。
- 在“SQL 编辑器”页面创建表。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。

- b. 在“SQL 编辑器”页面的左侧导航栏单击“数据库”页签。有两种方式创建表。
- 鼠标左键单击数据库名，进入“表”区域，单击右侧 ，创建当前数据库下的表。
 - 鼠标左键单击对应数据库右侧的 ，在列表菜单中选择“创建表”，创建当前数据库下的表。

步骤 2 在“创建表”页面，填写参数。

- 当数据位置为 DLI 时，请参见表 8-8 填写相关参数；
- 当数据位置为 OBS 时，请参见表 8-8 和表 8-9 填写相关参数。

表8-8 通用参数说明

参数名称	描述	示例
表名称	<ul style="list-style-type: none"> ● 表名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 ● 表名称大小写不敏感且不能为空。 ● 表名称支持包含“\$”符号。例如：\$test。 ● 输入长度不能超过 128 个字符。 	table01
数据位置	数据存储位置，当前支持 DLI 和 OBS。	OBS
描述	该表的描述。	-
列类型	选择为“普通列”或“分区列”。	普通列
列名称	表的列名。列名应至少包含一个字母，并允许下划线（_），但不支持纯数字。 可选择“普通列”或“分区列”。“分区列”是分区表专用的，对用户数据进行分区，可提高查询效率。 说明 列名不区分大小写，不能相同。	name
数据类型	与“列名”对应，表示该列的数据类型。 <ul style="list-style-type: none"> ● 字符串（string）：字符串类型。 ● 有符号整数（int）：存储空间为 4 字节。 ● 日期类型（date）：所表示日期的范围为 0000-01-01 to 9999-12-31。 ● 双精度浮点型（double）：存储空间为 8 字节。 ● 布尔类型（boolean）：存储空间为 1 字节。 ● 固定有效位数和小数位数的数据类型（decimal）：有效位数为 1~38 之间的正整数，包含 1 和 38；小数位数为小于 10 的整数。 	string

参数名称	描述	示例
	<ul style="list-style-type: none"> 有符号整数（smallint/short）：存储空间为 2 字节。 有符号整数（bigint/long）：存储空间为 8 字节。 时间戳（timestamp）：表示日期和时间，可达到小数点后 6 位。 单精度浮点型（float）：存储空间为 4 字节。 有符号整数（tinyint）：存储空间为 1 字节。仅 OBS 表支持。 	
列描述	该列的描述。	-
操作	<ul style="list-style-type: none"> 增加列 删除列 <p>说明</p> <p>当列数较多时，建议您使用 SQL 语句创建表，或直接从本地 Excel 导入列信息。</p>	-

表8-9 数据位置为 OBS 的参数说明

参数名称	描述	示例
数据格式	<p>支持以下数据格式。</p> <ul style="list-style-type: none"> Parquet: DLI 支持读取不压缩、snappy 压缩、gzip 压缩的 parquet 数据。 CSV: DLI 支持读取不压缩、gzip 压缩的 csv 数据。 ORC: DLI 支持读取不压缩、snappy 压缩的 orc 数据。 JSON: DLI 支持读取不压缩、gzip 压缩的 json 数据。 Carbon: DLI 支持读取不压缩的 carbon 数据。 Avro: DLI 支持读取不压缩的 avro 数据。 	CSV
存储路径	<p>输入或选择 OBS 路径。路径可以是文件夹，也可以是路径。</p> <p>说明</p> <p>如果需要将 OBS 上的数据导入 OBS 表，则建表路径必须是文件夹。建表路径是文件时，会导致数据导入失败。</p>	obs://obs1/sampled_ata.csv
表头:无/有	<p>当“数据格式”为“CSV”时，该参数有效。设置导入数据源是否含表头。</p> <p>选中“高级选项”，勾选“表头:无”前的方框，“表头:无”显示为“表头:有”，表示有表头；去勾</p>	-

参数名称	描述	示例
	选即为“表头:无”，表示无表头。	
自定义分隔符	<p>当“数据格式”为“CSV”，并在自定义分隔符前的方框打勾时，该参数有效。</p> <p>选中高级选项，支持选择如下分隔符。</p> <ul style="list-style-type: none"> • 逗号(,) • 竖线() • 制表符(\t) • 其他：输入自定义分隔符 	逗号(,)
自定义引用字符	<p>当“数据格式”为“CSV”，并在自定义引用字符前的方框打勾时，该参数有效。</p> <p>选中高级选项，支持选择如下引用字符。</p> <ul style="list-style-type: none"> • 单引号(') • 双引号(") • 其他：输入自定义引用字符 	单引号(')
自定义转义字符	<p>当“数据格式”为“CSV”，并在自定义转义字符前的方框打勾时，该参数有效。</p> <p>选中高级选项，支持选择如下转义字符。</p> <ul style="list-style-type: none"> • 反斜杠(\) • 其他：输入自定义转义字符 	反斜杠(\)
日期格式	<p>当“数据格式”为“CSV”和“JSON”时此参数有效。</p> <p>选中“高级选项”，该参数表示表中日期的格式，默认格式为“yyyy-MM-dd”。</p>	2000-01-01
时间戳格式	<p>当“数据格式”为“CSV”和“JSON”时此参数有效。</p> <p>选中“高级选项”，该参数表示表中时间戳的格式，默认格式为“yyyy-MM-dd HH:mm:ss”。</p>	2000-01-01 09:00:00

步骤 3 单击“确定”，完成表创建。

表创建成功后，您可以在“表管理”页面或者“SQL 编辑器”页面查看和选择使用对应的表。

步骤 4 (可选) DLI 表创建完成后，您可以选择是否直接向该表导入数据。

----结束

8.1.5 删除数据库和表

根据实际使用情况，您可以通过删除数据库和表。

注意事项

- 具有正在运行中的作业的数据库或者表不能删除。
- 管理员用户、数据库的所有者和具有删除数据库权限的用户可以删除数据库。管理员用户、表的所有者和具有删除表权限的用户可以删除表。

说明

数据库和表删除后，将不可恢复，请谨慎操作。

删除表

删除表的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面删除表。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。
 - b. 单击需删除表的数据库名，进入该数据库的“表管理”页面。
 - c. 选中目标表，单击“操作”栏中的“更多 > 删除”。
 - d. 在弹出的确认对话框中，单击“是”。
- 在“SQL 编辑器”页面删除表。
 - a. 在 SQL 作业管理控制台的顶部菜单栏中，选择“SQL 编辑器”。
 - b. 在左侧导航栏选择“数据库”页签，鼠标左键单击需要删除表的数据库名，进入“表”区域。
 - c. 鼠标左键单击对应表右侧的 ，在列表菜单中选择“删除”。
 - d. 在弹出的确认对话框中，单击“确定”。

删除数据库

1. 在管理控制台左侧，单击“数据管理”>“库表管理”。
2. 单击需要删除的数据库“操作”栏中的“更多 > 删除数据库”。

说明

需要删除的数据库中含有表时，不能执行删除操作。需要先删除其中的表。

3. 在弹出的确认对话框中，单击“是”。

8.1.6 修改数据库和表所有者

在实际使用过程中，开发人员创建了数据库和表，交给测试人员进行测试，测试人员测试完成后，再交给运维人员进行体验，在这种情况下，可以通过修改数据库和表的所有者，将数据转移给其他所有者。

修改数据库所有者

修改数据库所有者的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面修改数据库所有者。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。
 - b. 在“库表管理”页面选中需要修改的数据库，单击“操作”栏中的“更多 > 修改数据库”。
 - c. 在弹出的对话框中，输入新的所有者用户名（已存在的用户名），单击“确定”。
- 在“SQL 编辑器”页面修改数据库所有者。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。
 - b. 在左侧导航栏单击选择“数据库”页签，鼠标左键单击对应数据库右侧的 ，在列表菜单中选择“修改数据库”。
 - c. 在弹出的确认对话框中，输入新的所有者用户名（已存在的用户名），单击“确定”。

修改表所有者

1. 在管理控制台左侧，单击“数据管理”>“库表管理”。
2. 单击需要修改的表对应数据库名，进入该数据库的“表管理”页面。
3. 单击目标表“操作”栏中的“更多”>“修改所有者”。
4. 在弹出的对话框中，输入新的所有者用户名（已存在的用户名），单击“确定”。

8.1.7 导入数据

支持将 OBS 上的数据导入到 DLI 中创建的表中。

注意事项

- 创建 OBS 表时指定的路径必须是文件夹，若建表路径是文件将导致导入数据失败。
- 导入数据时只能指定一个路径，路径中不能包含逗号。
- 当 OBS 的目录下有同名文件夹和文件时，数据导入指向该路径会优先指向文件而非文件夹。
- 若将 CSV 格式数据导入分区表，需在数据源中将分区列放在最后一列。
- 不建议对同一张表并发导入数据，因为有一定概率发生并发冲突，导致导入失败。
- 导入文件支持 CSV, Parquet, ORC, JSON 和 Avro 格式，且文本格式仅支持 UTF-8。

前提条件

待导入的数据已存储到 OBS 上。

导入数据步骤

步骤 1 导入数据的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面导入数据。
 - a. 在管理控制台的左侧，选择“数据管理”>“库表管理”。
 - b. 单击需导入数据的表对应的数据库名称，进入该数据库的“表管理”页面。
 - c. 在目标表“操作”栏中选择“更多”中的“导入”，弹出“导入数据”页面。
- 在“SQL 编辑器”页面导入数据。
 - a. 在管理控制台的左侧，单击“SQL 编辑器”。
 - b. 在“SQL 编辑器”页面左侧导航栏选择“数据库”页签，鼠标左键单击需要导入数据的表对应的数据库名，进入“表”区域。
 - c. 鼠标左键单击对应表右侧的 ，在列表菜单中选择“导入”，弹出“导入数据”页面。

步骤 2 在“导入数据”页面，参见表 8-10 填写相关信息。

表8-10 参数说明

参数名称	描述	示例
数据库	当前表所在的数据库。	-
表名称	当前表名称。	-
队列	选择队列。	-
文件格式	导入数据源的文件格式。导入支持 CSV, Parquet, ORC, JSON, Avro 格式。	CSV
数据源路径	直接输入路径或单击  选择 OBS 的路径，若没有合适的桶可直接跳转 OBS 创建。 说明 路径同时支持文件和文件夹。	obs://DLI/sampleddata.csv
表头:无/有	当“文件格式”为“CSV”时该参数有效。设置导入数据源是否含表头。 选中“高级选项”，勾选“表头:无”前的方框，“表头:无”显示为“表头:有”，表示有表头；去勾选即为“表头:无”，表示无表头。	-
自定义分隔符	当“文件格式”为“CSV”，勾选自定义分隔符前的方框时，该参数有效。 支持选择如下分隔符。 <ul style="list-style-type: none"> • 逗号(,) • 竖线() 	默认值: 逗号(,)

参数名称	描述	示例
	<ul style="list-style-type: none"> 制表符(\t) 其他：输入自定义分隔符 	
自定义引用字符	当“文件格式”为“CSV”，勾选自定义引用字符前的方框时，该参数有效。 支持选择如下引用字符。 <ul style="list-style-type: none"> 单引号(') 双引号(") 其他：输入自定义引用字符 	默认值： 单引号(')
自定义转义字符	当“文件格式”为“CSV”，并在自定义转义字符前的方框打勾时，该参数有效。 选中高级选项，支持选择如下转义字符。 <ul style="list-style-type: none"> 反斜杠(\) 其他：输入自定义转义字符 	默认值： 反斜杠(\)
日期格式	当“文件格式”为“CSV”和“JSON”时此参数有效。 选中“高级选项”，该参数表示表中日期的格式，默认格式为“yyyy-MM-dd”。	2000-01-01
时间戳格式	当“文件格式”为“CSV”和“JSON”时此参数有效。 选中“高级选项”，该参数表示表中时间戳的格式，默认格式为“yyyy-MM-dd HH:mm:ss”。	2000-01-01 09:00:00
错误数据存储路径	当“文件格式”为“CSV”和“JSON”时此参数有效。 选中“高级选项”，该参数表示可将错误数据保存到对应的 OBS 路径中。	obs://DLI/

步骤 3 单击“确定”，系统开始导入数据。

步骤 4 有两种方式可查看导入的数据。

说明

目前预览只显示导入的前十条数据。

- 在“数据管理”>“库表管理”页面，单击数据库名，在表管理界面对应表的“操作”栏选择“更多”中的“表属性”，在弹框的“预览”页签中，可查看导入的数据
- 在“SQL 编辑器”的“数据库”页签中，单击数据库名称，进入对应的表列表，鼠标左键单击对应表右侧的 ，在列表菜单中选择“表属性”，在弹框的“预览”页签中，可查看导入的数据。

步骤 5（可选）可以在“作业管理 > SQL 作业”页面，查看该导入作业的状态以及执行结果。

----结束

8.1.8 将 DLI 数据导出至 OBS

支持将数据从 DLI 表中导出到 OBS 服务中，导出操作将在 OBS 服务新建文件夹，或覆盖已有文件夹中的内容。

注意事项

- 支持导出 json 格式的文件，且文本格式仅支持 UTF-8。
- 只支持将 DLI 表（表类型为“Managed”）中的数据导出到 OBS 桶中，且导出的路径必须指定到文件夹级别。
- 支持跨帐号导出数据，即，如果 B 帐号对 A 帐号授权后，A 帐号拥有 B 帐号 OBS 桶的元数据信息和权限信息的读取权限，以及路径的读写权限，则 A 帐号可将数据导出至 B 帐号的 OBS 路径中。

导出数据步骤

步骤 1 导出数据的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面导出数据。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。
 - b. 单击需导出数据的表对应的数据库，进入该数据的“表管理”页面。
 - c. 在对应表（DLI 表）的“操作”栏中选择“更多”中的“导出”，弹出“导出数据”页面。
- 在“SQL 编辑器”页面导出数据。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。
 - b. 在左侧导航栏选择“数据库”页签，鼠标左键单击需要导出数据的表对应的数据库名，进入“表”区域。
 - c. 鼠标左键单击需要导出数据的表（Managed 表，即 DLI 表）右侧的 ，在列表菜单中选择“导出”，选择弹出“导出数据”页面。

步骤 2 在“导出数据”对话框，参考表 8-11 填写导出数据相关信息。

表8-11 参数说明

参数名称	描述
数据库	当前表所在的数据库。
表名称	当前表名称。
数据格式	导出数据的文件格式。当前只支持 json 格式。
队列	选择队列。

参数名称	描述
压缩格式	导出数据的压缩方式，选择如下压缩方式。 <ul style="list-style-type: none"> • none • bzip2 • deflate • gzip
存储路径	<ul style="list-style-type: none"> • 输入或选择 OBS 路径。 • 导出路径必须为 OBS 桶中不存在的文件夹，即用户需在 OBS 目标路径后创建一个新文件夹。 • 文件夹名称不能包含下列特殊字符：\/:*?"<> ，并且不能以“.”开头和结尾。
导出方式	导出数据的保存方式。 <ul style="list-style-type: none"> • 随导出创建指定路径：指定的导出目录必须不存在，如果指定目录已经存在，系统将返回错误信息，无法执行导出操作。 • 覆盖指定路径：在指定目录下新建文件，会删除已有文件。
表头:无/有	设置导出数据是否含表头。

步骤 3 单击“确定”即可导出数据。

步骤 4（可选）您可以在“作业管理”>“SQL 作业”页面查看导出作业的“状态”、“执行语句”等信息。

1. 在“作业类型”中选择“EXPORT”，输入导出数据的时间段，即可查询出对应条件下的作业列表。
2. 单击导出作业名称前的 ，可查看导出作业的详细信息。

----结束

8.1.9 查看元数据

元数据说明

- 元数据（Metadata）是用来定义数据类型的数据。主要是描述数据自身信息，包含源、大小、格式或其它数据特征。数据库字段中，元数据用于诠释数据仓库的内容。
- 创建表时，会定义元数据，由列名、类型、列描述三列组成。
- “元数据”页面将显示目标表的列名、列类型、类型和描述。

查看元数据步骤

查看元数据的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面查看元数据。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。
 - b. 单击需导出数据对应数据库名称，进入该数据库“表管理”页面。
 - c. 单击目标表“操作”栏中的“更多”，选择“表属性”，即可在“元数据”页签查看该表的元数据信息。
- 在“SQL 编辑器”页面查看元数据。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。
 - b. 在“SQL 编辑器”页面的左侧导航栏中，选择“数据库”页签。
 - c. 单击对应数据库名，进入该数据库的表列表。
 - d. 鼠标左键单击对应表右侧的☰，在列表菜单中选择“表属性”，即可在“元数据”页签查看该表的元数据信息。

8.1.10 预览数据

“预览页面”将显示对应表的前 10 条数据。

预览数据步骤

预览数据的入口有两个，分别在“数据管理”和“SQL 编辑器”页面。

- 在“数据管理”页面预览数据。
 - a. 在管理控制台左侧，单击“数据管理”>“库表管理”。
 - b. 单击需导出数据对应数据库名称，进入该数据库“表管理”页面。
 - c. 单击目标表“操作”栏中的“更多”，选择“表属性”。
 - d. 单击“预览”页签，即可预览该表数据。
- 在“SQL 编辑器”页面预览数据。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。
 - b. 在“SQL 编辑器”页面的左侧导航栏中，选择“数据库”页签。
 - c. 鼠标左键单击对应数据库名，进入该数据库的表列表。
 - d. 鼠标左键单击对应表右侧的☰，在列表菜单中选择“表属性”，单击“预览”页签，即可预览该表数据。

8.2 程序包管理

8.2.1 程序包管理概述

Spark 程序包管理主要包括如下功能：

- 8.2.2 程序包组和程序包权限管理
- 8.2.3 创建程序包
- 8.2.4 删除程序包

说明

支持批量删除程序包。

- 8.2.5 修改所有者

程序包管理页面

表8-12 程序包管理参数

参数	参数说明
分组名称	程序包所属分组的名称。若不分组，则显示“--”。
名称	程序包名称。
所有者	上传程序包的用户名称。
包类型	程序包的类型。支持的包类型如下： <ul style="list-style-type: none">• JAR：用户 jar 文件。• PyFile：用户 Python 文件。• File：用户文件。
状态	创建程序包的状态。 <ul style="list-style-type: none">• 上传中（Uploading）：表示程序包正在上传。• 已成功（Finished）：表示程序包已经上传。• 已失败（Failed）：表示程序包上传失败。
创建时间	创建程序包的时间。
更新时间	更新程序包的时间。
操作	权限管理：对程序包用户进行权限管理。 删除：删除程序包。 更多： <ul style="list-style-type: none">• 修改所有者：修改程序包用户。

8.2.2 程序包组和程序包权限管理

程序包组和程序包权限操作场景

- 针对不同用户，可以通过权限设置分配不同的程序包组或程序包，不同用户之间的作业效率互不影响，保障作业性能。
- 管理员用户和程序包组或程序包的所有者拥有所有权限，不需要进行权限设置且其他用户无法修改其程序包组或程序包权限。
- 给新用户设置程序包组或程序包权限时，该用户所在用户组需具有 Tenant Guest 权限。

在“程序包管理”页面，单击程序包“操作”列中的“权限管理”，进入“用户权限信息”页面，可以对用户进行程序包组或程序包的授权、设置权限和回收权限。

说明

- 如果创建程序包时选择了分组，则权限管理为对应程序包组的权限管理。
- 如果创建程序包时选择了不分组，则权限管理为对应程序包的权限管理。

程序包组/程序包用户授权

单击页面右上角“授权”可对用户进行程序包组/程序包授权。

- 程序包组授权

表8-13 程序包组授权参数说明

参数名称	描述
用户名	被授权的 IAM 用户的名称。 说明 该用户名称是已存在的 IAM 用户名称。
权限设置	<ul style="list-style-type: none"> • 使用组：使用该组的程序包。 • 更新组：更新该组内程序包，包含创建组内程序包。 • 查询组：查询组内程序包详情。 • 删除组：删除该组的程序包。 • 赋权：当前用户可将组的权限赋予其他用户。 • 回收：当前用户可回收其他用户具备的该组的权限，但不能回收该组所有者的权限。 • 查看其他用户具备的权限：当前用户可查看其他用户具备的该组的权限。

- 程序包授权

表8-14 程序包授权参数说明

参数名称	描述
用户名	被授权的 IAM 用户的名称。 说明 该用户名称是已存在的 IAM 用户名称。
权限设置	<ul style="list-style-type: none"> • 使用程序包：使用该程序包。 • 更新程序包：更新该程序包。 • 查询程序包：查询该程序包。 • 删除程序包：删除该程序包。

参数名称	描述
	<ul style="list-style-type: none">• 赋权：当前用户可将程序包的权限赋予其他用户。• 回收：当前用户可回收其他用户具备的该程序包的权限，但不能回收该程序包所有者的权限。• 查看其他用户具备的权限：当前用户可查看其他用户具备的该程序包的权限。

设置程序包组和程序包权限

单击对应子用户“操作”列中的“权限设置”可修改该用户的权限。详细权限描述如表 8-13 和表 8-14 所示。

当“权限设置”中的选项为灰色时，表示您不具备修改此程序包组或程序包权限的权限。可以向管理员用户、组所有者等具有赋权权限的用户申请“程序包组或程序包的赋权”和“程序包组或程序包权限的回收”权限。

回收程序包组和程序包权限

单击对应子用户“操作”列中的“回收”将删除该用户的权限。该子用户将不具备该程序包组或程序包的任意权限。

程序包组和程序包权限使用说明

- 程序包组
查询权限，用户可以查看该用户创建的程序包组以及该组下的所有程序包，也可以查看被赋权任意权限的程序包组。
程序包组作为一个单元，用于管理行为一致的程序包，所以只能赋权给用户程序包组相关权限。
- 程序包
查询权限，用户可以查看该用户创建的程序包，也可以查看被赋权任意权限的程序包。

8.2.3 创建程序包

DLI 支持用户通过批处理方式将程序包提交至通用队列中运行。

说明

如果用户需要更新程序包，可以使用相同的程序包或文件上传至 DLI 的同一个位置（同一个分组），直接覆盖原有的程序包或文件。

前提条件

所使用的程序包需提前上传至 OBS 服务中保存。

创建程序包步骤

1. 在管理控制台左侧，单击“数据管理”>“程序包管理”。
2. 在“程序包管理”页面，单击右上角“创建”可创建程序包。
3. 在“创建程序包”对话框，参见表 8-15 设置相关参数。

表8-15 参数说明

参数名称	描述
包类型	支持的包类型如下： <ul style="list-style-type: none">• JAR: 用户 jar 文件• PyFile: 用户 Python 文件• File: 用户文件
OBS 路径	选择对应程序包的 OBS 路径。 说明 <ul style="list-style-type: none">• 程序包需提前上传至 OBS 服务中保存。• 只支持选择文件。
分组设置	可选择“已有分组”，“创建新分组”或“不分组”。
分组名称	<ul style="list-style-type: none">• 选择“已有分组”：可选择已有的分组。• 选择“创建新分组”：可输入自定义的组名称。• 选择“不分组”：不需要选择或输入组名称。 说明 <ul style="list-style-type: none">• 如果选择分组，则对应的权限管理为对应程序包组的权限管理。• 如果选择不分组，则对应的权限管理为对应程序包的权限管理。 程序包组和程序包权限管理请参考 8.2.2 程序包组和程序包权限管理。

4. 单击“确定”，完成创建程序包。
程序包创建成功后，您可以在“程序包管理”页面查看和选择使用对应的包。

8.2.4 删除程序包

根据实际使用情况，您可以通过删除操作释放程序包。

删除程序包步骤

1. 在管理控制台左侧，单击“数据管理”>“程序包管理”。
2. 单击待删除的程序包“操作”列中的“删除”。
3. 在弹出的确认对话框中，单击“是”。

8.2.5 修改所有者

在“程序包管理”页面，单击程序包“操作”列中的“更多 > 修改所有者”，可以修改对应程序包的所有者。

- 如果该程序包进行过分组设置，则可以参考下图，选择“组”或者“程序包”进行修改。
- 如果该程序包没有进行过分组设置，则可以参考下图，直接修改该程序包的所有者。

表8-16 参数说明

参数名称	描述
分组名称	<ul style="list-style-type: none">• 如果创建程序包时选择了分组，显示所在的分组名称。• 如果创建程序包时没有选择分组，则不显示该参数。
名称	程序包名称。
选择类型	<ul style="list-style-type: none">• 如果创建程序包时选择了分组，可选择修改“组”的所有者或者“程序包”的所有者。• 如果创建程序包时没有选择分组，则不显示该参数。
用户名	程序包所有者的名称。 说明 该用户名称是已存在的 IAM 用户名称。

8.2.6 内置依赖包

DLI 内置依赖包是平台默认提供的依赖包，用户打包 Spark 或 Flink jar 作业 jar 包时，不需要额外上传这些依赖包，以免与平台内置依赖包冲突。

Spark 2.3.2 依赖包

- accessors-smart-1.2.jar
- activation-1.1.1.jar
- aircompressor-0.8.jar
- alluxio-2.3.1-luxor-SNAPSHOT-client.jar
- antlr-2.7.7.jar
- antlr4-runtime-4.8-1.jar
- antlr-runtime-3.4.jar
- aopalliance-1.0.jar
- aopalliance-repackaged-2.4.0-b34.jar
- apache-log4j-extras-1.2.17.jar
- arpack_combined_all-0.1.jar
- arrow-format-0.8.0.jar

- arrow-memory-0.8.0.jar
- arrow-vector-0.8.0.jar
- asm-5.0.4.jar
- audience-annotations-0.5.0.jar
- automaton-1.11-8.jar
- avro-1.7.7.jar
- avro-ipc-1.7.7.jar
- avro-ipc-1.7.7-tests.jar
- avro-mapred-1.7.7-hadoop2.jar
- aws-java-sdk-bundle-1.11.271.jar
- base64-2.3.8.jar
- bcpkix-jdk15on-1.66.jar
- bcprov-jdk15on-1.66.jar
- bonecp-0.8.0.RELEASE.jar
- breeze_2.11-0.13.2.jar
- breeze-macros_2.11-0.13.2.jar
- calcite-avatica-1.2.0-incubating.jar
- calcite-core-1.2.0-incubating.jar
- calcite-linq4j-1.2.0-incubating.jar
- carbondata-bloom-1.6.1.1032-2.0.0.dli-20201012.110925-6.jar
- carbondata-cli-1.6.1.1032-2.0.0.dli-20201012.110947-6.jar
- carbondata-common-1.6.1.1032-2.0.0.dli-20201012.110840-6.jar
- carbondata-common-plugin-1.6.1.1032-2.0.0.dli-20201012.110945-6.jar
- carbondata-core-1.6.1.1032-2.0.0.dli-20201012.110843-6.jar
- carbondata-format-1.6.1.1032-2.0.0.dli-20201012.110953-12.jar
- carbondata-hadoop-1.6.1.1032-2.0.0.dli-20201012.110849-6.jar
- carbondata-hive-1.6.1.1032-2.0.0.dli-20201012.110911-6.jar
- carbondata-mv-core-1.6.1.1032-2.0.0.dli-20201012.110936-6.jar
- carbondata-mv-plan-1.6.1.1032-2.0.0.dli-20201012.110933-6.jar
- carbondata-processing-1.6.1.1032-2.0.0.dli-20201012.110846-6.jar
- carbondata-secondary_index-1.6.1.1032-2.0.0.dli-20201012.110942-6.jar
- carbondata-spark2-1.6.1.1032-2.0.0.dli-20201012.110857-6.jar
- carbondata-spark-common-1.6.1.1032-2.0.0.dli-20201012.110903-6.jar
- carbondata-spark-datasource-1.6.1.1032-2.0.0.dli-20201012.110900-6.jar
- carbondata-store-sdk-1.6.1.1032-2.0.0.dli-20201012.110852-6.jar
- carbondata-streaming-1.6.1.1032-2.0.0.dli-20201012.110922-6.jar
- carbondata-vector-1.6.1.1032-2.0.0.dli-20201012.110956-6.jar
- checker-qual-2.11.1.jar
- chill_2.11-0.8.4.jar
- chill-java-0.8.4.jar
- commons-beanutils-1.9.4.jar
- commons-cli-1.2.jar

- commons-codec-2.0-20130428.202122-59.jar
- commons-collections-3.2.2.jar
- commons-collections4-4.2.jar
- commons-compiler-3.0.8.jar
- commons-compress-1.4.1.jar
- commons-configuration2-2.1.1.jar
- commons-crypto-1.0.0-20191105.jar
- commons-daemon-1.0.13.jar
- commons-dbc-1.4.jar
- commons-dbc2-2.7.0.jar
- commons-httpclient-3.1.jar
- commons-io-2.5.jar
- commons-lang-2.6.jar
- commons-lang3-3.5.jar
- commons-logging-1.2.jar
- commons-math3-3.4.1.jar
- commons-net-2.2.jar
- commons-pool-1.5.4.jar
- commons-pool2-2.8.0.jar
- commons-text-1.3.jar
- compress-lzf-1.0.3.jar
- core-1.1.2.jar
- curator-client-4.2.0.jar
- curator-framework-4.2.0.jar
- curator-recipes-2.7.1.jar
- datanucleus-api-jdo-3.2.6.jar
- datanucleus-core-3.2.10.jar
- datanucleus-rdbms-3.2.9.jar
- derby-10.12.1.1.jar
- dnsjava-2.1.7.jar
- ehcache-3.3.1.jar
- eigenbase-properties-1.1.5.jar
- error_prone_annotations-2.3.4.jar
- failureaccess-1.0.1.jar
- fastutil-8.2.3.jar
- ffmpeg-4.3.1-1.5.4.jar
- ffmpeg-4.3.1-1.5.4-linux-x86_64.jar
- flatbuffers-1.2.0-3f79e055.jar
- generex-1.0.2.jar
- geronimo-jcache_1.0_spec-1.0-alpha-1.jar
- gson-2.2.4.jar
- guava-29.0-jre.jar

- guice-4.0.jar
- guice-servlet-4.0.jar
- hadoop-annotations-3.1.1-ei-302002.jar
- hadoop-auth-3.1.1-ei-302002.jar
- hadoop-aws-3.1.1-ei-302002.jar
- hadoop-client-3.1.1-ei-302002.jar
- hadoop-common-3.1.1-ei-302002.jar
- hadoop-hdfs-3.1.1-ei-302002.jar
- hadoop-hdfs-client-3.1.1-ei-302002.jar
- hadoop-mapreduce-client-common-3.1.1-ei-302002.jar
- hadoop-mapreduce-client-core-3.1.1-ei-302002.jar
- hadoop-mapreduce-client-jobclient-3.1.1-ei-302002.jar
- hadoop-minikdc-3.1.1-hw-ei-302002.jar
- hadoop-yarn-api-3.1.1-ei-302002.jar
- hadoop-yarn-client-3.1.1-ei-302002.jar
- hadoop-yarn-common-3.1.1-ei-302002.jar
- hadoop-yarn-registry-3.1.1-ei-302002.jar
- hadoop-yarn-server-common-3.1.1-ei-302002.jar
- hadoop-yarn-server-web-proxy-3.1.1-ei-302002.jar
- hamcrest-core-1.3.jar
- HikariCP-java7-2.4.12.jar
- hive-common-1.2.1-2.1.0.dli-20201111.064115-91.jar
- hive-exec-1.2.1-2.1.0.dli-20201111.064444-91.jar
- hive-metastore-1.2.1-2.1.0.dli-20201111.064230-91.jar
- hk2-api-2.4.0-b34.jar
- hk2-locator-2.4.0-b34.jar
- hk2-utils-2.4.0-b34.jar
- hppc-0.7.2.jar
- htrace-core4-4.2.0-incubating-hw-1.0.0.jar
- httpclient-4.5.4.jar
- httpcore-4.4.7.jar
- ivy-2.4.0.jar
- j2objc-annotations-1.3.jar
- jackson-annotations-2.10.0.jar
- jackson-core-2.10.0.jar
- jackson-core-asl-1.9.13-atlassian-4.jar
- jackson-databind-2.10.0.jar
- jackson-dataformat-yaml-2.10.0.jar
- jackson-datatype-jsr310-2.10.3.jar
- jackson-jaxrs-base-2.10.3.jar
- jackson-jaxrs-json-provider-2.10.3.jar
- jackson-mapper-asl-1.9.13-atlassian-4.jar

- jackson-module-jaxb-annotations-2.10.3.jar
- jackson-module-paranamer-2.10.0.jar
- jackson-module-scala_2.11-2.10.0.jar
- jakarta.activation-api-1.2.1.jar
- jakarta.xml.bind-api-2.3.2.jar
- janino-3.0.8.jar
- javacpp-1.5.4.jar
- javacpp-1.5.4-linux-x86_64.jar
- javacv-1.5.4.jar
- JavaEWAH-1.1.7.jar
- javassist-3.18.1-GA.jar
- javax.annotation-api-1.2.jar
- javax.inject-1.jar
- javax.inject-2.4.0-b34.jar
- javax.servlet-api-3.1.0.jar
- javax.ws.rs-api-2.0.1.jar
- java-xmlbuilder-1.1.jar
- javolution-5.3.1.jar
- jaxb-api-2.2.11.jar
- jcip-annotations-1.0-1.jar
- jcl-over-slf4j-1.7.26.jar
- jdo-api-3.0.1.jar
- jersey-client-2.23.1.jar
- jersey-common-2.23.1.jar
- jersey-container-servlet-2.23.1.jar
- jersey-container-servlet-core-2.23.1.jar
- jersey-guava-2.23.1.jar
- jersey-media-jaxb-2.23.1.jar
- jersey-server-2.23.1.jar
- jets3t-0.9.4.jar
- jetty-http-9.4.31.v20200723.jar
- jetty-io-9.4.31.v20200723.jar
- jetty-security-9.4.31.v20200723.jar
- jetty-server-9.4.31.v20200723.jar
- jetty-servlet-9.4.31.v20200723.jar
- jetty-util-9.4.31.v20200723.jar
- jetty-util-ajax-9.4.31.v20200723.jar
- jetty-webapp-9.4.31.v20200723.jar
- jetty-xml-9.4.31.v20200723.jar
- joda-time-2.9.3.jar
- jodd-core-4.2.0.jar
- json-20200518.jar

- json4s-ast_2.11-3.2.11.jar
- json4s-core_2.11-3.2.11.jar
- json4s-jackson_2.11-3.2.11.jar
- json-sanitizer-1.2.1.jar
- json-smart-2.3.jar
- jsp-api-2.1.jar
- jsr305-3.0.2.jar
- jta-1.1.jar
- jtransforms-2.4.0.jar
- jul-to-slf4j-1.7.26.jar
- junit-4.11.jar
- kerb-admin-1.0.1.jar
- kerb-client-1.0.1.jar
- kerb-common-1.0.1.jar
- kerb-core-1.0.1.jar
- kerb-crypto-1.0.1.jar
- kerb-identity-1.0.1.jar
- kerb-server-1.0.1.jar
- kerb-simplekdc-1.0.1.jar
- kerb-util-1.0.1.jar
- kerby-asn1-1.0.1.jar
- kerby-config-1.0.1.jar
- kerby-pkix-1.0.1.jar
- kerby-util-1.0.1.jar
- kerby-xdr-1.0.1.jar
- kryo-shaded-3.0.3.jar
- kubernetes-client-4.9.2-20200804.jar
- kubernetes-model-4.9.2-20200804.jar
- kubernetes-model-common-4.9.2-20200804.jar
- leveledbjni-all-1.8-20191105.jar
- libfb303-0.9.3.jar
- libthrift-0.12.0.jar
- listenablefuture-9999.0-empty-to-avoid-conflict-with-guava.jar
- log4j-1.2.17-cloudera1.jar
- log4j-rolling-appender-20131024-2017.jar
- logging-interceptor-3.14.4.jar
- luxor-encrypt-2.1.0-20201106.065437-53.jar
- luxor-fs3-2.1.0-20201106.065612-53.jar
- luxor-obs-fs3-2.1.0-20201106.065616-53.jar
- luxor-rpc_2.11-2.1.0-20201106.065541-53.jar
- luxor-rpc-protobuf2-2.1.0-20201106.065551-53.jar
- lz4-java-1.7.1.jar

- machinist_2.11-0.6.1.jar
- macro-compat_2.11-1.1.1.jar
- metrics-core-3.1.5.jar
- metrics-graphite-3.1.5.jar
- metrics-jmx-4.1.12.1.jar
- metrics-json-3.1.5.jar
- metrics-jvm-3.1.5.jar
- minlog-1.3.0.jar
- mssql-jdbc-6.2.1.jre7.jar
- netty-3.10.6.Final.jar
- netty-all-4.1.51.Final.jar
- nimbus-jose-jwt-8.19.jar
- objenesis-2.1.jar
- okhttp-3.14.4.jar
- okio-1.17.2.jar
- opencsv-2.3.jar
- opencsv-4.6.jar
- opencv-4.3.0-2.jar
- orc-core-1.4.4-nohive.jar
- orc-mapreduce-1.4.4-nohive.jar
- oro-2.0.8.jar
- osgi-resource-locator-1.0.1.jar
- paranamer-2.8.jar
- parquet-column-1.8.3.jar
- parquet-common-1.8.3.jar
- parquet-encoding-1.8.3.jar
- parquet-format-2.3.1.jar
- parquet-hadoop-1.8.3.jar
- parquet-hadoop-bundle-1.6.0.jar
- parquet-jackson-1.8.3.jar
- parquet-format-2.3.1.jar
- parquet-hadoop-1.8.3.jar
- parquet-hadoop-bundle-1.6.0.jar
- parquet-jackson-1.8.3.jar
- postgresql-42.2.14.jar
- protobuf-java-2.5.0.jar
- py4j-0.10.7.jar
- pyrolite-4.13.jar
- re2j-1.1.jar
- RoaringBitmap-0.5.11.jar
- scala-compiler-2.11.12.jar
- scala-library-2.11.12.jar

- scalap-2.11.0.jar
- scala-parser-combinators_2.11-1.1.0.jar
- scala-reflect-2.11.12.jar
- scala-xml_2.11-1.0.5.jar
- secComponentApi-1.0.5c.jar
- shapeless_2.11-2.3.2.jar
- slf4j-api-1.7.30.jar
- slf4j-log4j12-1.7.30.jar
- snakeyaml-1.24.jar
- snappy-java-1.1.7.5.jar
- spark-catalyst_2.11-2.3.2.0101-2.1.0.dli-20201111.073826-143.jar
- spark-core_2.11-2.3.2.0101-2.1.0.dli-20201111.073836-134.jar
- spark-graphx_2.11-2.3.2.0101-2.1.0.dli-20201111.073847-129.jar
- spark-hive_2.11-2.3.2.0101-2.1.0.dli-20201111.073854-132.jar
- spark-kubernetes_2.11-2.3.2.0101-2.1.0.dli-20201111.073916-85.jar
- spark-kvstore_2.11-2.3.2.0101-2.1.0.dli-20201111.073933-127.jar
- spark-launcher_2.11-2.3.2.0101-2.1.0.dli-20201111.073940-127.jar
- spark-mllib_2.11-2.3.2.0101-2.1.0.dli-20201111.073946-127.jar
- spark-mllib-local_2.11-2.3.2.0101-2.1.0.dli-20201111.073953-127.jar
- spark-network-common_2.11-2.3.2.0101-2.1.0.dli-20201111.073959-127.jar
- spark-network-shuffle_2.11-2.3.2.0101-2.1.0.dli-20201111.074007-127.jar
- spark-om_2.11-2.3.2.0101-2.1.0.dli-20201111.074019-125.jar
- spark-repl_2.11-2.3.2.0101-2.1.0.dli-20201111.074028-125.jar
- spark-sketch_2.11-2.3.2.01012.1.0.dli-20201111.074035-125.jar
- spark-sql_2.11-2.3.2.0101-2.1.0.dli-20201111.074041-126.jar
- spark-streaming_2.11-2.3.2.0101-2.1.0.dli-20201111.074100-123.jar
- spark-tags_2.11-2.3.2.0101-2.1.0.dli-20201111.074136-123.jar
- spark-tags_2.11-2.3.2.0101-2.1.0.dli-20201111.074141-124-tests.jar
- spark-unsafe_2.11-2.3.2.0101-2.1.0.dli-20201111.074144-123.jar
- spark-uquery_2.11-2.3.2.0101-2.1.0.dli-20201111.074906-210.jar
- spark-yarn_2.11-2.3.2.0101-2.1.0.dli-20201111.074151-123.jar
- spire_2.11-0.13.0.jar
- spire-macros_2.11-0.13.0.jar
- ST4-4.3.1.jar
- stax2-api-3.1.4.jar
- stax-api-1.0-2.jar
- stream-2.7.0.jar
- stringtemplate-3.2.1.jar
- token-provider-1.0.1.jar
- univocity-parsers-2.5.9.jar
- validation-api-1.1.0.Final.jar
- woodstox-core-5.0.3.jar

- xbean-asm5-shaded-4.4.jar
- xercesImpl-2.12.0.jar
- xml-apis-1.4.01.jar
- xz-1.0.jar
- zjsonpatch-0.3.0.jar
- zookeeper-3.5.6-ei-302002.jar
- zookeeper-jute-3.5.6-ei-302002.jar
- zstd-jni-1.4.4-11.jar

Flink 1.7.2 依赖包

- bcpkix-jdk15on-1.60.jar
- bcprov-jdk15on-1.60.jar
- commons-codec-1.9.jar
- commons-configuration-1.7.jar
- deeplearning4j-core-0.9.1.jar
- deeplearning4j-nlp-0.9.1.jar
- deeplearning4j-nn-0.9.1.jar
- ejml-cdense-0.33.jar
- ejml-core-0.33.jar
- ejml-ddense-0.33.jar
- ejml-dsparse-0.33.jar
- ejml-experimental-0.33.jar
- ejml-fdense-0.33.jar
- ejml-simple-0.33.jar
- ejml-zdense-0.33.jar
- elsa-3.0.0-M7.jar
- esdk-obs-java-3.1.3.jar
- flink-cep_2.11-1.7.0.jar
- flink-cep-scala_2.11-1.7.0.jar
- flink-dist_2.11-1.7.0.jar
- flink-gelly_2.11-1.7.0.jar
- flink-gelly-scala_2.11-1.7.0.jar
- flink-ml_2.11-1.7.0.jar
- flink-python_2.11-1.7.0.jar
- flink-queryable-state-runtime_2.11-1.7.0.jar
- flink-shaded-curator-1.7.0.jar
- flink-shaded-hadoop2-uber-1.7.0.jar
- flink-table_2.11-1.7.0.jar
- guava-26.0-jre.jar
- hadoop-cloud-3.1.1-41-20201014.085840-4.jar
- httpasyncclient-4.1.2.jar
- httpclient-4.5.12.jar

- httpcore-4.4.4.jar
- httpcore-nio-4.4.4.jar
- java-xmlbuilder-1.1.jar
- jna-4.1.0.jar
- libtensorflow-1.12.0.jar
- log4j-api-2.8.2.jar
- log4j-core-2.8.2.jar
- log4j-over-slf4j-1.7.21.jar
- logback-classic-1.2.3.jar
- logback-core-1.2.3.jar
- nd4j-api-0.9.1.jar
- nd4j-native-0.9.1.jar
- nd4j-native-api-0.9.1.jar
- nd4j-native-platform-0.9.1.jar
- okhttp-3.14.8.jar
- okio-1.14.0.jar
- slf4j-api-1.7.21.jar
- tensorflow-1.12.0.jar

Flink 1.10 依赖包

2020 年 12 月之后创建的新队列才能使用 Flink 1.10 依赖包。

- bcpkix-jdk15on-1.60.jar
- bcprov-jdk15on-1.60.jar
- commons-codec-1.9.jar
- commons-configuration-1.7.jar
- deeplearning4j-core-0.9.1.jar
- deeplearning4j-nlp-0.9.1.jar
- deeplearning4j-nn-0.9.1.jar
- ejml-cdense-0.33.jar
- ejml-core-0.33.jar
- ejml-ddense-0.33.jar
- ejml-dsparse-0.33.jar
- ejml-experimental-0.33.jar
- ejml-fdense-0.33.jar
- ejml-simple-0.33.jar
- ejml-zdense-0.33.jar
- elsa-3.0.0-M7.jar
- esdk-obs-java-3.20.6.1.jar
- flink-cep_2.11-1.10.0.jar
- flink-cep-scala_2.11-1.10.0.jar
- flink-dist_2.11-1.10.0.jar

- flink-python_2.11-1.10.0.jar
- flink-queryable-state-runtime_2.11-1.10.0.jar
- flink-sql-client_2.11-1.10.0.jar
- flink-state-processor-api_2.11-1.10.0.jar
- flink-table_2.11-1.10.0.jar
- flink-table-blink_2.11-1.10.0.jar
- guava-26.0-jre.jar
- hadoop-3.1.1-41.jar
- httpasyncclient-4.1.2.jar
- httpclient-4.5.3.jar
- httpcore-4.4.4.jar
- httpcore-nio-4.4.4.jar
- java-xmlbuilder-1.1.jar
- jna-4.1.0.jar
- libtensorflow-1.12.0.jar
- log4j-over-slf4j-1.7.26.jar
- logback-classic-1.2.3.jar
- logback-core-1.2.3.jar
- nd4j-api-0.9.1.jar
- nd4j-native-0.9.1.jar
- nd4j-native-api-0.9.1.jar
- nd4j-native-platform-0.9.1.jar
- okhttp-3.14.8.jar
- okio-1.14.0.jar
- secComponentApi-1.0.5.jar
- slf4j-api-1.7.26.jar
- tensorflow-1.12.0.jar

Python 3.7.5 依赖包

- absl-py-0.8.1
- astor-0.8.0
- backports.functools_lru_cache-1.5
- backports.weakref-1.0.post1
- blaze-0.10.1
- Click-7.0
- contextlib2-0.6.0.post1
- cycler-0.10.0
- dataclasses-0.6
- datashape-0.5.2
- decorator-4.4.0
- Flask-1.1.1
- Flask_Cors-3.0.8

- Flask-1.16.0
- funcsigs-1.0.2
- gast-0.2.2
- google_pasta-0.1.7
- grpcio-1.24.1
- h5py-2.10.0
- itsdangerous-1.1.0
- Jinja2-2.10.3
- joblib-0.14.0
- Keras-2.3.1
- Keras_Applications-1.0.8
- Keras_Preprocessing-1.1.0
- kiwisolver-1.1.0
- Markdown-3.1.1
- MarkupSafe-1.1.1
- matplotlib-3.1.1
- mock-3.0.5
- multipledispatch-0.6.0
- multipledispatch2-0.4.0
- networkx-2.2
- numpy-1.16.1
- odo-0.5.0
- opt_einsum-3.1.0
- pandas-0.24.1
- pip-19.2.3
- protobuf-3.10.0
- psutil-5.6.3
- pyarrow-0.11.1
- pyparsing-2.4.2
- python-dateutil-2.8.0
- pytz-2019.3
- PyYAML-5.1.2
- scikit-learn-0.21.3
- scipy-1.1.0
- setuptools-41.2.0
- six-1.12.0
- SQLAlchemy-1.3.10
- subprocess32-3.5.4
- tensorboard-1.13.1
- tensorflow-1.13.1
- tensorflow-estimator-1.13.0
- termcolor-1.1.0



- toolz-0.10.0
- Werkzeug-0.16.0
- wheel-0.33.6
- wrapt-1.11.2

9 作业模板

9.1 SQL 模板管理

为了便捷快速的执行 SQL 操作，DLI 支持定制模板或将正在使用的 SQL 语句保存为模板。保存模板后，不需编写 SQL 语句，可通过模板直接执行 SQL 操作。

SQL 模板包括样例模板和自定义模板。当前系统默认的样例模板包括 22 条标准的 TPC-H 查询语句，可以满足用户大部分的 TPC-H 需求场景测试，TPC-H 样例说明请参考 9.3.1 SQL 模板下 TPC-H 样例数据说明。

SQL 模板管理主要包括如下功能：

- 样例模板
- 自定义模板
- 创建模板
- 执行模板
- 查找模板
- 修改模板
- 删除模板

表格设置

在“SQL 模板”页面右上角，单击“设置”可以选择是否按照分组展示模板。

如果选择“按分组展示”，有以下三种展示方式：

- 展开第一个分组
- 全部展开
- 全部收起

样例模板

当前样例模板包括 22 条标准的 TPC-H 查询语句，您可以查看模板名称、描述、语句等信息。TPC-H 样例说明请参考 9.3.1 SQL 模板下 TPC-H 样例数据说明。

表9-1 模板管理参数

参数	参数说明
名称	模板名称。 <ul style="list-style-type: none">模板名称只能包含数字、英文字母和下划线，但不能是纯数字，不能以下划线开头，且不能为空。输入长度不能超过 50 个字符。
描述	所创建模板的相应描述。
语句	创建为模板的 SQL 语句。
操作	执行：单击“执行”将跳转至 SQL 编辑器页面，可根据需要进行修改或直接执行操作。具体请参考 执行模板 。

当前已有的样例模板包括如下场景：

- 价格摘要报告查询
- 最小代价供应商分析
- 运送优先级分析
- 订单优先级检查分析
- 当地供应商数量分析
- 预测收入变化分析
- 货运量分析
- 国家市场份额分析
- 产品类型利润估量分析
- 返修零件分析
- 重要库存标志分析
- 货运模式和命令优先分析
- 消费者分配分析
- 促销效果分析
- 最高贡献供应商分析
- 零件/供应商关系分析
- 小批量订单收入分析
- 大订单顾客分析
- 折扣收入分析
- 潜在零件改进分析
- 不能按时交货供应商分析
- 全球销售机会分析

自定义模板

自定义模板页面显示用户创建的所有模板，您可以查看模板名称、描述、语句等信息。

表9-2 模板管理参数

参数	参数说明
名称	模板名称。 <ul style="list-style-type: none">模板名称只能包含数字、英文字母和下划线，但不能是纯数字，不能以下划线开头，且不能为空。输入长度不能超过 50 个字符。
描述	所创建模板的相应描述。
语句	创建为模板的 SQL 语句。
操作	<ul style="list-style-type: none">执行：单击“执行”将跳转至 SQL 编辑器页面，可根据需要进行修改或直接执行操作。具体请参考执行模板。修改：单击“修改”可在弹出的“修改模板”对话框中，根据需要修改模板的信息。具体请参考修改模板。

创建模板

创建模板的操作入口有两个，分别在“作业模板”和“SQL 编辑器”页面。

- 在“作业模板”页面创建模板。
 - 在管理控制台左侧，单击“作业模板”>“SQL 模板”。
 - 在“SQL 模板”页面，单击右上角“创建模板”。
输入模板名称、语句和描述信息，详细参数介绍请参见表 9-3。

表9-3 参数说明

参数名称	描述
名称	模板名称。 <ul style="list-style-type: none">模板名称只能包含数字、英文字母和下划线，但不能是纯数字，不能以下划线开头，且不能为空。输入长度不能超过 50 个字符。
语句	需要保存为模板的 SQL 语句。
描述	该模板的相应描述。
分组设置	<ul style="list-style-type: none">已有分组创建新分组

参数名称	描述
	<ul style="list-style-type: none">• 不分组
分组名称	“分组设置”选择“已有分组”或者“创建新分组”时，需要填写分组名称。

- c. 单击“确定”，完成模板创建。
- 在“SQL 编辑器”页面创建模板。
 - a. 在管理控制台左侧，单击“SQL 编辑器”。
 - b. 单击 SQL 作业编辑窗口右上方的“更多”，选择“设为模板”，可将编辑窗口中的 SQL 语句设置为模板。
输入模板名称、语句和描述信息，详细介绍请参见表 9-3。
 - c. 单击“确定”，完成模板创建。

执行模板

执行模板操作步骤如下：

1. 在管理控制台左侧，单击“作业模板”>“SQL 模板”。
2. 在“SQL 模板”页面，勾选相应的模板，单击“操作”列的“执行”，将跳转至“SQL 编辑器”页面，并在 SQL 作业编辑窗口中自动输入对应的 SQL 语句。
3. 在 SQL 作业编辑窗口右上方，单击“执行”运行 SQL 语句，执行结束后，可以在 SQL 作业编辑窗口下方区域中查看执行结果。

查找模板

在“SQL 模板”页面，可在右上方搜索框中输入模板名称关键字，查找与之匹配的模板。

修改模板

修改模板仅支持对自定义模板进行操作，具体步骤如下：

- 步骤 1** 在“SQL 模板”页面，单击“自定义模板”，选中需修改的模板，单击“操作”列的“修改”。
- 步骤 2** 在弹出的“修改模板”对话框中，根据需要修改模板的名称、语句和描述。
- 步骤 3** 单击“确定”，保存修改结果。

----结束

删除模板

在“SQL 模板”页面，单击“自定义模板”，勾选一个或多个待删除的模板，单击“删除”，可删除选中的模板。

9.2 Flink 模板管理

Flink 模板包括样例模板和自定义模板。用户可以在已有的样例模板中进行修改，来实现实际的作业逻辑需求，节约编辑 SQL 语句的时间。也可以根据自己的习惯和方法自定义作业模板，方便后续可以直接调用或修改。

Flink 模板管理主要包括如下功能：

- [Flink SQL 样例模板](#)
- [自定义模板](#)
- [新建模板](#)
- [基于模板新建作业](#)
- [修改模板](#)
- [删除模板](#)

Flink SQL 样例模板

Flink SQL 样例模板列表显示已有的 Flink SQL 样例作业模板，Flink SQL 样例模板列表参数说明如表 9-5 所示。

表9-4 Flink SQL 样例模板列表参数

参数	参数说明
名称	模板名称，只能由英文、中文、数字、中划线和下划线组成，并且长度为 1~64 个字符。
描述	模板的相关描述，且长度为 0~512 个字符。
操作	“创建作业”：直接在该模板下创建作业，创建完后，系统跳转到“作业管理”下的作业编辑页面。

当前已有的样例模板包括如下场景：

- NGINX 访问日志实时 ETL 入库
- 套牌车辆检测
- 电子围栏告警
- 车辆偏航告警
- 车辆超速告警
- 流式随机森林异常检测
- CloudTable-DLI Flink-CloudTable
- DIS-DLI Flink-CSS(Elasticsearch)
- DIS-DLI Flink-CloudTable(OpenTSDB)
- DIS-DLI Flink-CloudTable
- DIS-DLI Flink-DCS
- DIS-DLI Flink-DDS(MongoDB)

- DIS-DLI Flink-DIS
- DIS-DLI Flink-OBS-DWS
- DIS-DLI Flink-SMN
- Kafka-DLI Flink-Kafka
- OBS-DLI Flink-RDS
- Stream-Join-Table(DCS)
- Stream-Join-Table(RDS)

自定义模板

自定义模板列表显示所有的 jar 作业模板，自定义模板列表参数说明如表 9-5 所示。

表9-5 自定义模板列表参数

参数	参数说明
名称	模板名称，只能由英文、中文、数字、中划线和下划线组成，并且长度为 1~64 个字符。
描述	模板的相关描述，且长度为 0~512 个字符。
创建时间	创建模板的时间。
更新时间	最后修改模板的时间。
操作	<ul style="list-style-type: none"> • “编辑”：对已经创建好的模板进行修改。 • “创建作业”：直接在该模板下创建作业，创建完后，系统跳转到“作业管理”下的作业编辑页面。 • 更多： <ul style="list-style-type: none"> - “删除”：将已经创建的模板删除。

新建模板

创建作业模板，有以下四种方法。

- 进入“作业模板”页面新建模板。
 - a. 在 DLI 管理控制台的左侧导航栏中，单击“作业模板”>“Flink 模板”。
 - b. 单击页面右上角“创建模板”，弹出“创建模板”页面。
 - c. 输入“名称”和“描述”。

表9-6 模板配置信息

参数	参数说明
名称	模板名称，只能由字母、中文、数字、中划线和下划线组成，并且长度为 1~64 个字符。
	说明

参数	参数说明
	模板名称必须是唯一的。
描述	模板的相关描述，且长度为 0~512 字符。

- d. 单击“确定”，进入“编辑”页面。

界面介绍

功能	描述
名称	可以修改模板名称。
描述	可以修改模板描述。
保存方式	<ul style="list-style-type: none"> 修改：将修改保存至当前的模板中。 新增：将修改另存为新的模板。
SQL 语句编辑区域	输入详细的 SQL 语句，实现业务逻辑功能。SQL 语句的编写请参考《数据湖探索 SQL 语法参考》。
保存	保存修改。
创建作业	使用当前模板创建作业。
格式化	对 SQL 语句进行格式化，将 SQL 语句格式化后，需要重新编辑 SQL 语句。
主题设置	更改字体大小、自动换行、页面风格（黑色底或白色底）等配置。

- e. 在 SQL 语句编辑区域，输入 SQL 语句，实现业务逻辑功能。SQL 语句的编写请参考《数据湖探索 SQL 语法参考》。
- f. SQL 编辑完成后，单击右上角的“保存”，完成创建模板。
- g. （可选）如果不需要进行修改，也可以单击右上角的“创建作业”基于当前模板创建作业。创建作业请参考 6.2.4 创建 Flink SQL 作业、6.2.5 创建 Flink Jar 作业。
- 基于现有作业模板新建模板
 - 在 DLI 管理控制台的左侧导航栏中，单击“作业模板”>“Flink 模板”，单击“自定义模板”页签。
 - 在自定义模板列表中，单击所需作业模板“操作”列中的“编辑”，进入“模板编辑”页面。
 - 修改完成后，“保存方式”选择“新增”。
 - 单击右上角“保存”，完成另存一个新模板。
 - 基于新建作业新建模板
 - 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入“Flink 作业”页面。

- b. 单击右上角“创建作业”，弹出“创建作业”页面。
 - c. 配置作业信息，输入“名称”和“描述”，选择“模板”。
 - d. 单击“确认”，进入“作业编辑”页面。
 - e. SQL 编辑完成后，单击“设为模板”，弹出“设为模板”窗口。
 - f. 输入“名称”和“描述”，单击“确认”，完成另存一个新模板。
- 基于现有作业新建模板
 - a. 在 DLI 管理控制台的左侧导航栏中，单击“作业管理”>“Flink 作业”，进入“Flink 作业”页面。
 - b. 在作业列表中，选择一个需要设置为模板的作业，在“操作”列单击“编辑”，进入“作业编辑”页面。
 - c. SQL 编辑完成后，单击“设为模板”，弹出“设为模板”窗口。
 - d. 输入“名称”和“描述”，单击“确认”，完成另存一个新模板。

基于模板新建作业

用户可以基于样例模板或者自定义模板新建作业。

1. 在 DLI 管理控制台的左侧导航栏中，单击“作业模板”>“Flink 模板”。
2. 在样例模板列表中，单击对应模板“操作”列中的“创建作业”。创建作业请参考 6.2.4 创建 Flink SQL 作业和 6.2.5 创建 Flink Jar 作业。

修改模板

用户创建完自定义模板后，可以根据实际需求修改自定义模板。样例模板不支持修改，但是可以查看。

1. 在 DLI 管理控制台的左侧导航栏中，单击“作业模板”>“Flink 模板”，单击“自定义模板”页签。
2. 在自定义模板列表中，选择一个需要修改的模板，单击模板名称或该模板“操作”列中的“编辑”，进入“编辑”页面。
3. 在 SQL 语句编辑区，根据需要修改 SQL 语句。
4. “保存方式”选择“修改”。
5. 单击右上角“保存”，保存当前模板修改的内容。

删除模板

用户可以根据需求删除不需要的自定义模板，不支持删除样例模板。模板删除后无法恢复，请谨慎操作。

1. 在 DLI 管理控制台的左侧导航栏中，单击“作业模板”>“Flink 模板”，单击“自定义模板”页签。
2. 在自定义模板列表中，勾选需要删除的模板，支持多选，单击自定义模板列表左上方的“删除”。

用户也可以在自定义模板列表中，勾选需要删除的模板，单击“操作”栏中“更多 > 删除”，删除对应的模板。

3. 在弹出的删除确认窗口中，单击“是”。

9.3 附录

9.3.1 SQL 模板下 TPC-H 样例数据说明

TPC-H 样例数据简介

TPC-H（商业智能计算测试）是美国交易处理效能委员会（TPC,Transaction Processing Performance Council）组织制定的用来模拟决策支持类应用的一个测试集。目前，在学术界和工业界普遍用来评价决策支持技术方面应用的性能。这种商业测试可以全方位评测系统的整体商业计算综合能力，对厂商的要求更高，同时也具有普遍的商业实用意义，目前在银行信贷分析和信用卡分析、电信运营分析、税收分析、烟草行业决策分析中都有广泛的应用。

TPC-H 基准测试是由 TPC-D（由 TPC 组织于 1994 年指定的标准，用于决策支持系统方面的测试基准）发展而来的。TPC-H 用 3NF 实现了一个数据仓库，共包含 8 个基本关系，其数据量可以设定从 1G~3T 不等。TPC-H 基准测试包括 22 个查询 (Q1~Q22)，其主要评价指标是各个查询的响应时间，即从提交查询到结果返回所需时间。TPC-H 基准测试的度量单位是每小时执行的查询数(QphH@size)，其中“H”表示每小时系统执行复杂查询的平均次数，“size”表示数据库规模的大小，能够反映出系统在处理查询时的能力。TPC-H 是根据真实的生产运行环境来建模的,这使得它可以评估一些其他测试所不能评估的关键性能参数。总而言之，TPC 组织颁布的 TPC-H 标准满足了数据仓库领域的测试需求,并且促使各个厂商以及研究机构将该项技术推向极限。

本示例将演示 DLI 直接对存储在 OBS 中的 TPC-H 数据集进行查询的操作，DLI 已经预先生成了 100M 的 TPC-H-2.18 的标准数据集，已将数据集上传到了 OBS 的 tpch 文件夹中，并且赋予了只读访问权限，方便用户进行查询操作。

TPC-H 的测试和度量指标

TPC-H 测试分解为 3 个子测试：数据装载测试、Power 测试和 Throughput 测试。建立测试数据库的过程被称为装载数据，装载测试是为测试 DBMS 装载数据的能力。装载测试是第一项测试，测试装载数据的时间，这项操作非常耗时。Power 测试是在数据装载测试完成后，数据库处于初始状态，未进行其它任何操作，特别是缓冲区还没有被测试数据库的数据，被称为 raw 查询。Power 测试要求 22 个查询顺序执行 1 遍，同时执行一对 RF1 和 RF2 操作。最后进行 Throughput 测试，也是最核心和最复杂的测试，更接近于实际应用环境，与 Power 测试比对 SUT 系统的压力有非常大的增加，有多个查询语句组，同时有一对 RF1 和 RF2 更新流。

测试中测量的基础数据都与执行时间有关，这些时间又可分为：装载数据的每一步操作时间、每个查询执行时间和每个更新操作执行时间，由这些时间可计算出：数据装载时间、Power@Size、Throughput@Size、QphH@Size 和\$/QphH@Size。

Power@Size 是 Power 测试的结果，被定义为查询时间和更改时间的几何平均值的倒数，公式如下：

$$\text{TPC-H Power@Size} = \frac{3600 * SF}{\sqrt[24]{\prod_{i=1}^{i=22} QI(i,0) * \prod_{j=1}^{j=2} RI(j,0)}}$$

其中：Size 为数据规模；SF 为数据规模的比例因子；QI (i, 0) 为第 i 个查询的时间，以秒为单位；R (Ij, 0) 为 RFj 更新的时间，以秒为单位。

Throughput@Size 是 Throughput 测试的结果，被定义为所有查询执行时间平均值的倒数，公式如下：

$$\text{QphH@Size} = \sqrt{\text{Power @ Size} * \text{Throughput @ Size}}$$

业务场景

用户可以通过 DLI 内置的 TPC-H 测试套件进行简单高效的交互式查询，无需用户上传数据，即可以体验 DLI 的核心功能。

DLI 内置 TPC-H 的优势

- 用户只需要登录 DLI，完成授予权限，即可操作 SQL 语句，无需用户自己创建表和导入数据。
- 预置 22 条 TPC-H SQL 查询模板，功能丰富，可满足大部分的商业场景，无需用户自行下载 TPC-H 的查询语句，省时省力。
- 用最小的时间代价体验 serverless 化的 DLI 产品，领略数据湖带给我们的全新体验。

注意事项

子账号使用 TPC-H 测试套件时，需要主账号为子账号赋权 OBS 访问权限和查看主账号表的权限；如果主账号未登录过 DLI 服务，子账号除上述权限外，还需要创建数据库和创建表的权限。

操作说明

具体操作指导详见 9.1 SQL 模板管理。

10 跨源连接

10.1 跨源连接和跨源分析概述

DLI 支持原生 Spark 的 DataSource 能力，并在其基础上进行了扩展，能够通过 SQL 语句、Spark 作业或者 Flink 作业访问其他数据存储服务并导入、查询、分析处理其中的数据。

跨源连接

在使用 DLI 进行跨源分析前，需要先建立跨源连接，打通数据源之间的网络。

DLI 增强型跨源连接底层采用对等连接，直接打通 DLI 队列与目的数据源的 VPC 网络，通过点对点的方式实现数据互通。

📖 说明

- 系统 default 队列不支持创建跨源连接。
- 跨源连接需要使用 VPC、子网、路由、对等连接功能，因此需要获得 VPC（虚拟私有云）的 VPC Administrator 权限。可在 11.2 服务授权中进行设置。

跨源分析

增强型跨源支持 DLI 服务已实现的所有跨源业务，并且通过可以 UDF、Spark 作业和 Flink 作业等方式实现与自建数据源之间的访问。

目前 DLI 支持跨源访问的数据源包括：CloudTable HBase，CloudTable OpenTSDB，CSS，DCS Redis，DDS Mongo，DIS，DMS Kafka，DWS，MRS HBase，MRS Kafka，MRS OpenTSDB，OBS，RDS MySQL，RDS PostGre，SMN。

📖 说明

- 访问跨源表需要使用已经创建跨源连接的队列。
- 跨源表不支持 Preview 预览功能。

跨源分析流程

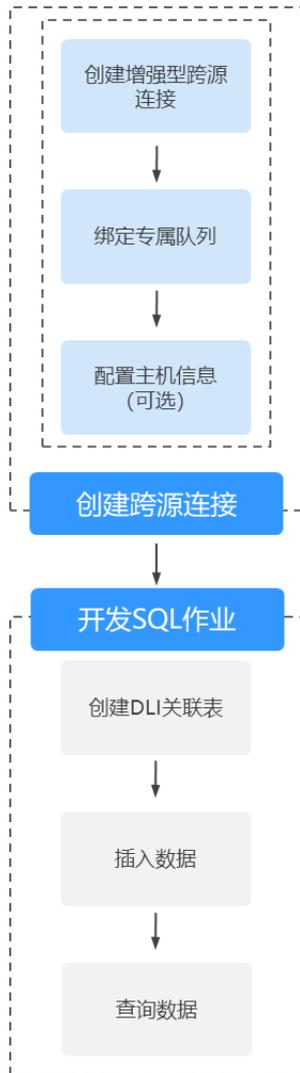
使用 DLI 进行跨源分析，需要先建立跨源连接，再开发不同的作业访问数据源。操作步骤如下：

1. 创建跨源连接。包括以下两种创建方式：
 - 通过管理控制台创建跨源连接。
 - 通过 API 接口创建跨源连接。
2. 开发 DLI 作业访问数据源。包括以下三种访问方式：
 - 开发 SQL 作业访问数据源。
 - 开发 Spark 作业访问数据源。
 - 开发 Flink 作业访问数据源。

以下分别为开发 SQL 作业、Spark 作业和 Flink 作业进行跨源连接的基本流程。

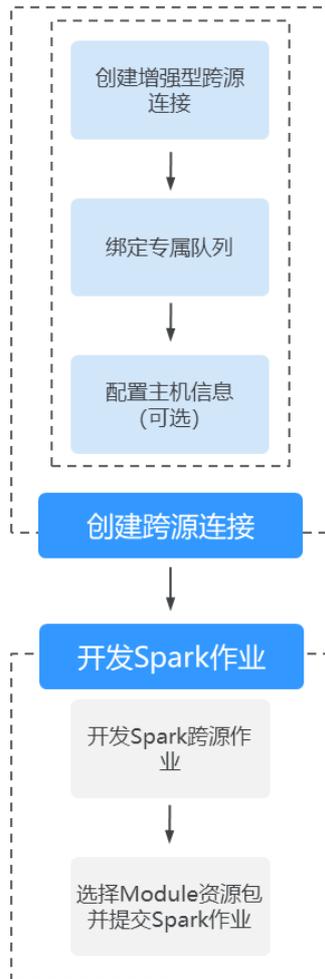
- SQL 作业

图10-1 SQL 跨源分析流程



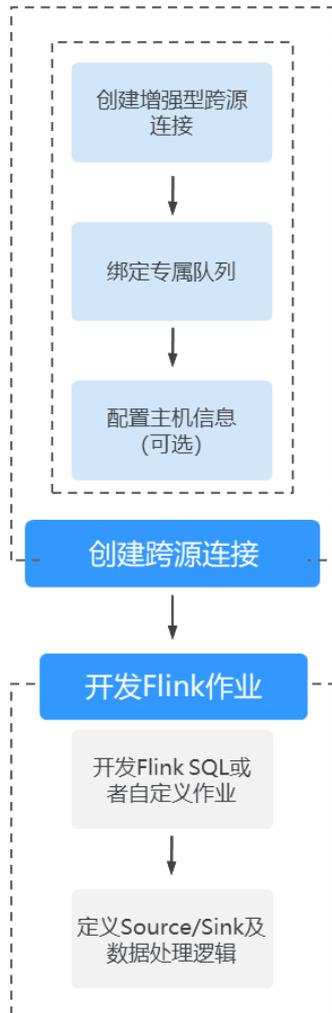
- Spark 作业

图10-2 Spark 跨源分析流程



- Flink 作业

图10-3 Flink 跨源分析流程



10.2 增强型跨源连接

10.2.1 增强型跨源连接概述

DLI 增强型跨源连接底层采用对等连接，直接打通 DLI 集群与目的数据源的 vpc 网络，通过点对点的方式实现数据互通。增强型跨源支持所有 DLI 服务已实现的跨源业务，包括 CloudTable 的 Hbase 和 OpenTSDB，MRS 的 OpenTSDB，DWS，RDS，CSS，DCS，DDS 等数据源。并且通过 UDF、Spark 作业和 Flink 作业方式能够实现与自建数据源之间的访问。

📖 说明

- 绑定跨源的 DLI 队列网段和数据源网段不能重合。
- 系统 default 队列不支持创建跨源连接。
- 访问跨源表需要使用已经创建跨源连接的队列。

- 跨源表不支持 Preview 预览功能。

增强型跨源主要包括如下功能：

- 10.2.2 创建/查找/删除增强型跨源连接
- 10.2.3 绑定/解绑队列
- [修改主机信息](#)
- 10.2.5 自定义路由信息
- 10.2.6 增强型跨源连接权限管理

增强型跨源页面

显示所有的增强型跨源连接，连接数量较多时，系统分页显示。

表10-1 跨源连接列表参数

参数	参数说明
连接名称	所创建的跨源连接名称。
连接状态	跨源连接的状态信息，目前控制台仅显示“已激活”状态的连接。
虚拟私有云	目的数据源所使用的虚拟私有云。
子网	目的数据源所使用的子网。
创建时间	每个连接的创建时间，可按创建时间顺序或倒序显示连接列表。
操作	<ul style="list-style-type: none">• 删除连接：用于删除已创建的增强型跨源连接。• 路由信息：查看增强型跨源连接自定义的路由信息。• 更多：<ul style="list-style-type: none">- 修改主机信息：用户自定义配置主机/域名对应的 IP 信息。- 绑定队列：用于为增强型跨源连接绑定队列。- 解绑队列：用于解除增强型跨源连接与队列之间的绑定关系。- 添加路由：添加增强型跨源连接的自定义路由。- 删除路由：删除增强型跨源连接的自定义路由。- 权限管理：对其他项目进行授权或者回收权限。

10.2.2 创建/查找/删除增强型跨源连接

创建增强型跨源连接

以创建 MRS 服务 HBase 跨源连接为例进行说明。

📖 说明

只支持通过增强型跨源访问 MRS HBase。

1. 在 MRS 服务中申请集群。
如果已有可用集群，可不用重新申请。
2. 在 DLI 管理控制台左侧导航栏中，选择“跨源管理”。
3. 选择“增强型跨源”页签，单击左上角的“创建”按钮。
输入连接名称，选择绑定队列（可选），虚拟私有云，子网，输入主机信息（可选），详细参数介绍请参见表 10-2。

表10-2 参数说明

参数	参数说明
连接名称	所创建的跨源连接名称。 <ul style="list-style-type: none"> • 名称只能包含数字、英文字母、下划线。不能为空。 • 输入长度不能超过 64 个字符。
绑定队列	可选参数，用于绑定需要使用跨源的队列。 说明 使用增强型跨源连接之前必须绑定队列且对等连接的状态是“active”。
虚拟私有云	目的数据源所使用的虚拟私有云。
子网	目的数据源所使用的子网。
主机信息	可选参数，访问 MRS 的 HBase 集群时需要配置 Zookeeper 实例的主机名与对应的 IP 地址。每行填写一条记录，填写格式为：“IP 主机名/域名”。 获取 MRS 集群主机名与 IP 地址，以 MRS3.x 集群为例，步骤如下： <ol style="list-style-type: none"> 1. 登录 MRS 管理控制台页面。 2. 单击“集群列表 > 现有集群”，在集群列表中单击指定的集群名称，进入集群信息页面。 3. 选择“组件管理”页签； 4. 单击进入“Zookeeper”服务； 5. 选择“实例”页签，可以查看对应业务 IP，可选择任意一个业务 IP。 说明 如果 MapReduce 服务集群存在多个 IP，创建跨源连接时填写其中任意一个业务 IP 即可。

4. 单击“确定”，完成连接创建。
5. 创建完跨源连接后，连接 Kafka、DWS、RDS 等实例还需要在实例的安全组下添加 DLI 网段的安全组规则。以连接 RDS 为例：

- a. 在 DLI 管理控制台，单击“队列管理”，选择您所绑定的队列，单击队列名称旁的  按钮，获取队列的网段信息。
 - b. 在 RDS 控制台“实例管理”页面，单击对应实例名称，查看“连接信息”>“数据库端口”，获取 RDS 数据库实例端口。
 - c. 单击“连接信息”>“安全组”对应的安全组名称，跳转到 RDS 实例安全组管理界面。单击“入方向规则 > 添加规则”，协议选择“TCP”，端口选择 RDS 数据库实例端口，源地址填写 DLI 队列的网段。单击“确定”完成配置。
6. 测试 DLI 队列与连接实例的连通性。以下通过测试 DLI 与 RDS 实例连通性举例说明。
- a. 在 RDS 控制台“实例管理”页面，单击对应实例名称，查看“连接信息”>“内网地址”，即可获取 RDS 内网地址。查看“连接信息”>“数据库端口”，获取 RDS 数据库实例端口。
 - b. 在 DLI 管理控制台，单击“队列管理”，选择您所绑定的队列，单击“操作”列中的“更多”>“测试地址连通性”。
 - c. 按照：“IP:端口”格式输入 RDS 实例连接地址和端口，进行网络连通性测试。

查找增强型跨源连接

在“增强型跨源”页面，可在搜索框中输入连接名称关键字，查找与之匹配的连接。

查看增强型跨源连接详情

在“增强型跨源”页面，选中一条连接，单击该连接对应的  ，可查看该条连接的详细信息。包括：连接 ID 和主机信息。

删除增强型跨源连接

在“增强型跨源”页面，可单击“操作”列的“删除连接”，删除不需要的连接。

说明

当“连接状态”为“创建中”时，连接不可删除。

10.2.3 绑定/解绑队列

约束限制

- 绑定跨源的 DLI 队列网段和数据源网段不能重合。
- 不支持绑定系统预置的 default 队列。

绑定队列

使用增强型跨源连接之前必须绑定队列且对等连接的状态是“active”。

- 方式一

在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“更多 > 绑定队列”，在弹出的对话框中，选择要绑定的队列（支持多选），单击“确定”。

- 方式二

单击选择的连接名称，进入该连接的“详情”页面。单击左上角的“创建”，在弹出的对话框中，选择要绑定的队列（支持多选），单击“确定”。

查看绑定队列详情

在“增强型跨源”页面，选中一条连接，单击选中的连接名称，可以查看绑定队列相关信息。

表10-3 跨源连接队列详情列表参数

参数	参数说明
对等连接 ID	增强型跨源在该队列所在集群中创建的对等连接 ID。 说明 每一个增强型跨源对每一个绑定的队列都会创建一个对等连接。该对等连接用于实现跨 VPC 通信，请确保数据源使用的安全组开放了 DLI 队列网段的访问，并且在使用跨源过程中不要删除该对等连接。
名称	已绑定的队列名称。
连接状态	跨源连接的状态信息，包括以下三种状态： <ul style="list-style-type: none"> • 创建中 • 已激活 • 已失败 说明 当连接状态显示为“已失败”时，单击左边对应的  ，可查看详细错误信息。
更新时间	每个连接的更新时间，可按更新时间顺序或倒序显示连接列表。
操作	解绑队列：用于解除跨源连接与队列之间的绑定关系。

解绑队列

不需要使用增强型跨源连接时，可将所绑定的队列进行解绑，释放资源。

- 方式一

在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“更多” > “解绑队列”，在弹出的对话框中，选择要解绑的队列，单击“确定”。

- 方式二

单击选择的连接名称，进入该连接的“详情”页面。选择要解绑的队列，单击该连接“操作”列中的“解绑队列”，在弹出的对话框中，选择要解绑的队列，单击“是”。

10.2.4 修改主机信息

修改主机信息

方法一：通过 MRS 节点的 “/etc/hosts” 信息填写

步骤 1 以 root 用户登录 MRS 的任意一个主机节点。

步骤 2 执行以下命令获取 MRS 对应主机节点的 hosts 信息，复制保存。

cat /etc/hosts

```
[root@node-master1nqtl ~]# cat /etc/hosts
:1 localhost localhost.localdomain localhost6 localhost6.localdomain6
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
10.10.10.10 hadoop.hadoop.com
10.10.10.10 manager
192.168.0.44 node-master1nqtl.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-master1nqtl.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.190 node-master2V8cb.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-master2V8cb.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.188 node-ana-corezyNL.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-ana-corezyNL.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.136 node-ana-corezyKB.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-ana-corezyKB.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
192.168.0.174 node-ana-coreVPY0.b27fd346-a6fb-42ef-bcc0-d964639baafb.com node-ana-coreVPY0.b27fd346-a6fb-42ef-bcc0-d964639baafb.com.
[root@node-master1nqtl ~]#
[root@node-master1nqtl ~]#
[root@node-master1nqtl ~]#
```

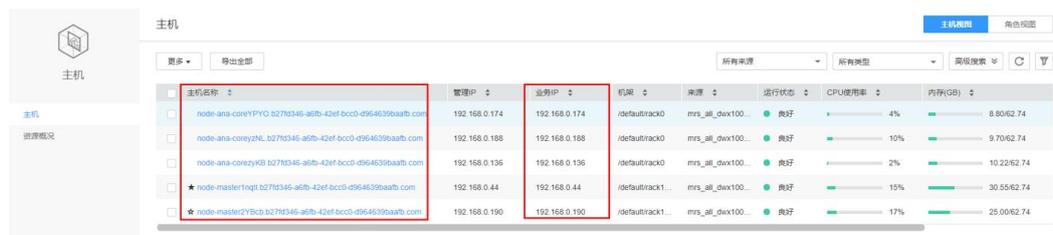
步骤 3 在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“修改主机信息”，在弹出的对话框中，将步骤 2 中复制的 MRS 主机信息粘贴进去。单击“确定”完成。

----结束

方法二：登录 MRS 的 FusionInsight Manager 获取主机信息填写

步骤 1 登录 MRS 的 FusionInsight Manager 界面。

步骤 2 在 FusionInsight Manager 界面，单击“主机”。在主机页面，分别获取 MRS 的“主机名称”和“业务 IP”。



主机名称	管理IP	业务IP	机架	来源	运行状态	CPU使用率	内存(GB)
node-ana-coreVPY0.b27fd346-a6fb-42ef-bcc0-d964639baafb.com	192.168.0.174	192.168.0.174	/defaulttrack0	mrs_all_oux100...	良好	4%	8.80/62.74
node-ana-corezyNL.b27fd346-a6fb-42ef-bcc0-d964639baafb.com	192.168.0.188	192.168.0.188	/defaulttrack0	mrs_all_oux100...	良好	10%	9.70/62.74
node-ana-corezyKB.b27fd346-a6fb-42ef-bcc0-d964639baafb.com	192.168.0.136	192.168.0.136	/defaulttrack0	mrs_all_oux100...	良好	2%	10.22/62.74
node-master1nqtl.b27fd346-a6fb-42ef-bcc0-d964639baafb.com	192.168.0.44	192.168.0.44	/defaulttrack1...	mrs_all_oux100...	良好	15%	30.55/62.74
node-master2V8cb.b27fd346-a6fb-42ef-bcc0-d964639baafb.com	192.168.0.190	192.168.0.190	/defaulttrack1...	mrs_all_oux100...	良好	17%	25.00/62.74

步骤 3 在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“修改主机信息”，在弹出的对话框中，填写主机信息。单击“确定”完成。

主机信息格式为：“步骤 2 中获取的业务 IP 步骤 2 中获取的主机名称”，多条信息之间以换行分隔。

例如：

192.168.0.22 node-masterxxx1.com

192.168.0.23 node-masterxxx2.com

----结束

说明

- 修改主机信息每次均为全量覆盖。
- 主机名/域名限制为长度 128，数字字母下划线("_")横杠("-")句点(".")组成，字母开头。
- 填写的主机信息格式为：“IP 主机名/域名”，多条信息之间以换行分隔。

10.2.5 自定义路由信息

自定义路由信息

创建增强型跨源连接并绑定队列后，系统会自动配置路由。同时，您也可以为所绑定的队列添加自定义路由。

- 查看路由信息
在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“路由信息”，可以查看该跨源自定义的路由信息。
- 添加路由
在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“更多”>“添加路由”，或者在“路由信息”页面单击“添加路由”，可以添加该跨源的自定义路由。在弹出的对话框中，填写路由名称与路由网段即可，详细参数请参考表 10-4。

表10-4 自定义路由详情列表参数

参数	参数说明
路由名称	自定义路由的名称，在同一个增强型跨源中唯一。名称规则为：长度 1~64 字节，数字、字母、下划线("_")、中划线("-")组成。
路由 IP	自定义路由网段，允许不同路由的网段之间有交集，但不允许完全相同。

- 删除路由
在“增强型跨源”页面，选择一条连接，单击该连接“操作”列中的“更多”>“删除路由”，或者在“路由信息”页面，选中一条自定义路由信息，单击该路由对应的“删除路由”，在弹出的对话框中选择要删除的路由信息，单击“是”。

10.2.6 增强型跨源连接权限管理

权限管理

对其他项目进行对应增强型跨源连接的授权或者回收权限。

- 授权：
 - a. 登录 DLI 管理控制台，单击“跨源管理 > 增强型跨源”，选择具体的跨源连接，在“操作”列，单击“更多 > 权限管理”。在权限管理对话框中选择“赋权”并输入需要授权的项目 ID，单击“确定”完成授权。

- b. 对项目进行授权后，使用被授权的项目对应的用户进行登录或切换到对应的项目下，在其对应的“增强型跨源”页面中可显示授权的跨源连接，可以将所创建的队列与该跨源连接进行绑定，实现跨项目创建跨源连接和创建路由的功能。

📖 说明

- 如果被授权的项目属于相同区域（region）的不同用户，则需使用被授权项目所属的用户帐号进行登录。
- 如果被授权的项目属于相同区域（region）的同一用户，则需使用当前帐号切换到对应的项目下。

例如：项目 B 需要访问项目 A 上的数据源，对应操作如下。

- 对于项目 A：
 - i. 使用项目 A 对应的帐号登录 DLI 服务。
 - ii. 通过对应数据源的 VPC 信息在 DLI 服务中创建增强型跨源连接“ds”。
 - iii. 将增强型跨源连接“ds”授权给项目 B。
- 对于项目 B：
 - i. 使用项目 B 对应的帐号登录 DLI 服务。
 - ii. 对增强型跨源连接“ds”进行绑定队列操作。
 - iii. （可选）设置主机信息，创建路由。

通过上述操作项目 A 的增强型跨源连接与项目 B 的队列创建了对等连接和路由，即可在项目 B 的队列上创建作业访问项目 A 的数据源。

- 回收：在“权限管理”对话框中选择“回收”并选择需要回收权限的项目 ID。

10.3 跨源认证权限管理

跨源认证权限管理操作场景

- 针对不同用户，可以通过权限设置分配不同的跨源认证，不同用户之间的作业效率互不影响，保障作业性能。
- 管理员用户和跨源认证的所有者拥有所有权限，不需要进行权限设置且其他用户无法修改其跨源认证权限。

在“跨源认证”页签，单击需要修改的认证信息“操作”列中的“权限管理”，进入“用户权限信息”页面，对用户进行跨源认证授权、设置权限和回收权限。

跨源认证用户授权

登录 DLI 管理控制台，单击“跨源管理 > 跨源认证”，选择要进行授权的跨源认证，在“操作”列，单击“权限管理”。单击页面右上角“授权”可对用户进行跨源认证授权。

表10-5 用户授权参数说明

参数名称	描述
用户名	被授权的 IAM 用户的名称。 说明 该用户名称是已存在的 IAM 用户名称。
权限设置	<ul style="list-style-type: none">• 使用：使用该跨源认证。• 更新：更新该跨源认证。• 删除：删除该跨源认证。• 赋权：当前用户可将跨源认证的权限赋予其他用户。• 回收：当前用户可回收其他用户具备的该跨源认证的权限，但不能回收该跨源认证所有者的权限。• 查看其他用户具备的权限：当前用户可查看其他用户具备的该跨源认证的权限。

设置跨源认证权限

单击对应子用户“操作”列中的“权限设置”可修改该用户的权限。详细权限描述如表 10-5 所示。

当“权限设置”中的选项为灰色时，表示您不具备修改此跨源认证权限的权限。可以向管理员用户、组所有者等具有赋权权限的用户申请“跨源认证的赋权”和“跨源认证权限的回收”权限。

回收跨源认证权限

单击对应子用户“操作”列中的“回收”将删除该用户的权限。该子用户将不具备该跨源认证的任意权限。

10.4 创建及管理跨源认证

跨源认证使用场景

跨源认证用于管理访问 CSS 和 MRS 安全环境的认证信息，以及加密存储访问 DWS、RDS、DDS、DCS 数据源的密码。

- 创建 CSS 安全集群时，需要指定用户名、密码、CSS 安全集群，会有一个证书可供下载。访问 CSS 安全集群，需要这三个信息，为了便于从 DLI 连接 CSS 安全集群，将这三个信息存储到 DLI 服务中，即为连接 CSS 安全集群的“跨源认证信息”。
- 创建 MRS 安全集群时，需要开启 Kerberos 认证。下载认证凭证，认证凭证中包含“krb5.conf”和“user.keytab”文件。在从 DLI 连接 MRS 安全集群时，可将这两个文件存储到 DLI 服务中，即为连接 MRS 安全集群的“跨源认证信息”。

- 创建 Kafka 集群，需要开启 SSL 访问配置。可在 MRS 服务，Kafka 组件的“服务配置”中进行配置，也可以通过搭建开源 Kafka 集群，修改配置文件进行开启。
- 通过跨源密码认证，加密存储用于访问 DWS、RDS、DDS、DCS 数据源的密码。

创建跨源认证信息

1. 创建需要访问的数据源。

说明

如果已有可用集群，可不用重新申请。

- 创建 CSS 安全集群：在 CSS 服务中创建集群，“集群版本”选择“6.5.4”或“6.5.4”以上版本，并开启“安全模式”。
 - 创建 MRS 安全集群：在 MRS 服务中申请集群，选择“自定义购买”，“集群版本”选择“MRS 2.1.0”或“MRS 2.1.0”以上版本，并开启“Kerberos 认证”。
 - 创建 Kafka 集群：可在 MRS 服务，Kafka 组件的“服务配置”中进行配置。在 MRS 服务配置 SSL 具体操作可参考《MapReduce 服务用户指南》中《安全配置》章节。
 - 创建 DWS 集群：在 DWS 服务中申请数据仓库集群，具体请参考《数据仓库服务管理指南》。
 - 创建 RDS 数据源：在 RDS 服务中申请数据库实例，具体请参考《云数据库快速入门》。
 - 创建 DCS 数据源：在 DCS 服务中申请缓存实例，具体请参考《分布式缓存指南》。
 - 创建 DDS 数据源：在 DDS 服务中申请数据库实例，具体请参考《文档数据库服务快速入门》。
2. 下载认证凭证。访问 DWS、RDS、DCS、DDS 数据源请跳过该步骤。
 - CSS 安全集群：在“集群管理”页面中，单击对应的集群名称，进入“基本信息”页面，找到“安全模式”，下载 CSS 安全集群的证书。
 - MRS 安全集群：在“集群列表”页面中，单击对应的集群名称，进入集群信息页面，可“下载认证凭证”。下载认证凭证后，需进行解压，得到“krb5.conf”和“user.keytab”文件。
 3. 上传“认证凭证”。访问 DWS、RDS、DCS、DDS 数据源请跳过该步骤。
将获取到的认证凭证文件上传到自定义的 OBS 桶中。
 4. 在 DLI 管理控制台选择“跨源管理”。
 5. 在“跨源认证”页签，单击“创建”，创建认证信息。
 - CSS

表10-6 参数说明

参数	参数说明
类型	选择 CSS。
认证信息名	所创建的跨源认证信息名称。

参数	参数说明
称	<ul style="list-style-type: none"> 名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 输入长度不能超过 128 个字符。 建议名称中包含 CSS 安全集群的名称，便于区分不同集群的安全认证信息。
用户名	安全集群的登录用户名。
用户密码	安全集群的登录密码。
Certificate 路径	上传“安全证书”的 OBS 路径。

- MRS

表10-7 参数说明

参数	参数说明
类型	选择 kerberos。
认证信息名称	所创建的跨源认证信息名称。 <ul style="list-style-type: none"> 名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 输入长度不能超过 128 个字符。 建议名称中包含 MRS 安全集群的名称，便于区分不同集群的安全认证信息。
用户名	安全集群的登录用户名。
krb5_conf 路径	上传“krb5.conf”文件的 OBS 路径。 说明 “krb5.conf”中需移除[libdefaults]下的“renew_lifetime”配置项，否则可能会遇到“Message stream modified (41)”问题。
keytab 路径	上传“user.keytab”文件的 OBS 路径。

- Kafka

表10-8 参数说明

参数	参数说明
类型	选择 Kafka_SSL。
认证信息名称	所创建的跨源认证信息名称。

参数	参数说明
	<ul style="list-style-type: none"> 名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 输入长度不能超过 128 个字符。
Truststore 路径	上传 SSL TRUSTSTORE 文件的 OBS 路径。
Truststore 密码	truststore 密码，默认为 dms@kafka
Keystore 路径	上传 SSL KEYSTORE(密钥和证书)文件的 OBS 路径。
Keystore 密码	keystore(密钥和证书)密码。
Key 密码	keystore 中的私钥密码。

- password (跨源密码认证)

访问 DWS、RDS、DCS 和 DDS 数据源通过该方式创建跨源认证。

说明

目前只支持 Spark SQL。

表10-9 参数说明

参数	参数说明
类型	选择 Password。
认证信息名称	所创建的跨源认证信息名称。 <ul style="list-style-type: none"> 名称只能包含数字、英文字母和下划线，但不能是纯数字，且不能以下划线开头。 输入长度不能超过 128 个字符。
用户名称	访问数据源的用户名。
用户密码	访问数据源的密码。

查找认证信息

在“跨源认证”页签，可在搜索框中输入认证信息名称，查找与之匹配的认证信息。为了用户信息的安全，不会返回密码字段。

更新认证信息

在“跨源认证”页签，单击需要修改的认证信息“操作”列中的“更新”，更改认证信息。目前仅支持更新改用户名称和用户密码。若需要更新证书，请删除该认证信息，再重新创建。

📖 说明

用户名和密码非必填，若不填，则表示不修改该字段。

删除认证信息

在“跨源认证”页签，单击需要修改的认证信息“操作”列中的“删除”，删除不需要的认证信息。

11 全局配置

11.1 全局变量

全局变量可用于简化复杂参数。例如，可替换长难复杂变量，提升 SQL 语句可读性。

创建变量

1. 在 DLI 控制台左侧导航栏中单击“全局配置 > 全局变量”。
2. 在“全局变量”页面，单击右上角“创建变量”，可创建新的全局变量。

表11-1 创建变量参数说明

参数名称	描述
变量名称	所创建的全局变量名称。
敏感变量	如果是涉及到密码或者其他敏感信息，可设置为：是，不涉及到敏感信息选择为：否。
变量值	全局变量的值。

3. 创建全局变量之后，在 SQL 语法中使用“{{xxxx}}”代替设置为全局变量的参数值即可，其中“xxxx”为变量名称。例如，在建表语句中，设置表名为全局变量 abc，即可用{{abc}}代替实际的表名。

```
create table {{table_name}} (String1 String, int4 int, varchar1 varchar(10))  
partitioned by (int1 int,int2 int,int3 int)
```

📖 说明

- 只有创建全局变量的用户才可以使用对应的变量。
- 不推荐在建表语句的 OPTIONS 关键字中使用全局变量。

修改变量

在“全局变量”页面，单击变量“操作”列中的“修改”，可修改对应的变量值。

说明

只有创建全局变量的用户才可以修改对应的变量。

删除变量

在“全局变量”页面，单击变量“操作”列中的“删除”，可删除对应的变量。

说明

- 只有创建全局变量的用户才可以删除对应的变量。
- 变量删除后，SQL 中将无法使用该变量。

11.2 服务授权

前提条件

服务授权需要主帐号或者用户组 `admin` 中的子帐号进行操作。

服务授权操作

进入数据湖探索管理控制台后，为保证正常使用数据湖探索功能，建议先进行委托权限设置。

如果需要对委托权限进行调整，可在“服务授权”页面中进行修改。目前 DLI 需要设置的委托权限参考表 11-2。

1. 根据需要勾选对应委托权限，单击“更新委托授权”。授权需要主帐号或者用户组 `admin` 中的子帐号进行操作。界面会提示“委托权限更新成功”，表示修改成功。
2. 同意授权后，DLI 将在统一身份认证服务 IAM 为您创建名为 `dli_admin_agency` 的委托，授权成功后，可以进入服务委托列表查看。**注意：委托 `dli_admin_agency` 创建成功后，请勿删除。**

表11-2 DLI 委托权限列表

权限名	详细信息	备注
Tenant Administrator (全局服务)	DLI Flink 作业访问和使用 OBS 或者 DWS 数据源、日志转储（包括桶授权）、开启 checkpoint、作业导入导出等，需要获得访问和使用 OBS（对象存储服务）的 Tenant Administrator 权限。	由于云服务缓存需要时间，该权限 60 分钟左右才能生效。
DIS Administrator	DLI Flink 作业访问和使用 DIS 数据源，需要获得访问和使用 DIS（数据接入服务）的 DIS Administrator 权限。	由于云服务缓存需要时间，该权限 30 分钟左右才能生效。

权限名	详细信息	备注
VPC Administrator	DLI 跨源连接需要使用 VPC、子网、路由、对等连接功能，因此需要获得使用 VPC（虚拟私有云）的 VPC Administrator 权限。	由于云服务缓存需要时间，该权限 3 分钟左右才能生效。
SMN Administrator	DLI 作业执行失败需要通过 SMN 发送通知消息，因此需要获得访问和使用 SMN（消息通知服务）的 SMN Administrator 权限。	由于云服务缓存需要时间，该权限 3 分钟左右才能生效。
Tenant Administrator(项目级)	使用其他必须具有 Tenant Administrator 权限才能运行的服务也需要获得该权限。	由于云服务缓存需要时间，该权限 3 分钟左右才能生效。

12 DLI 使用 UDF 操作指导

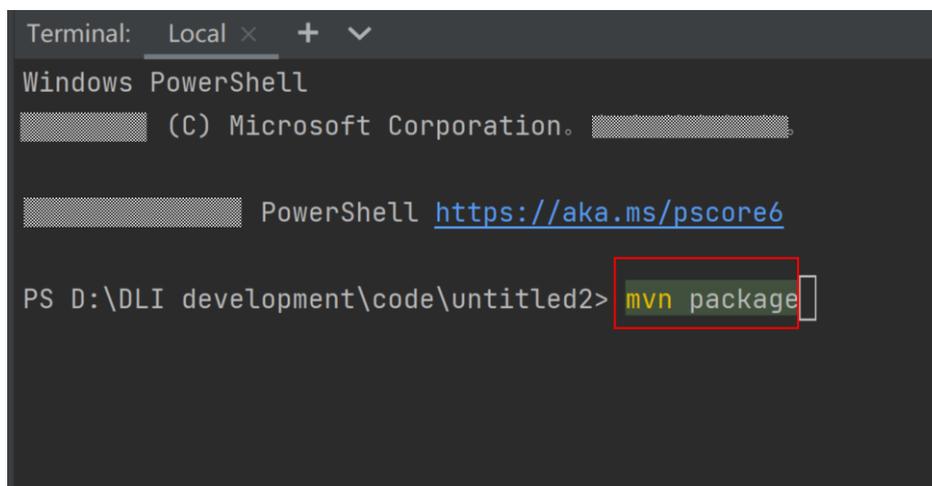
DLI 支持用户使用 Hive UDF（User Defined Function，用户定义函数）进行数据查询等操作。

使用 UDF 操作步骤

1. 编写 UDF。
 - a. 在 pom 文件中添加以下配置信息。

```
<dependency>  
  <groupId>org.apache.hive</groupId>  
  <artifactId>hive-exec</artifactId>  
  <version>1.2.1</version>  
</dependency>
```
 - b. 代码中需要添加导入：`import org.apache.hadoop.hive.ql.exec.UDF;`
 - c. 根据业务需要新建或修改样例代码中 `SumUdfDemo.java` 的内容。
2. 生成 jar 包。使用“`mvn package`”命令编译打包程序，生成所需的 Jar 包。
例如，使用 IDEA 工具编写好程序代码后，单击工具界面最下方的“Terminal”，在命令行窗口输入“`mvn package`”编译打包程序。

图12-1 编译打包程序



根据编译结果提示，在对应的路径下获取生成的 jar 包文件，例如 TestUDF.jar

3. 上传 TestUDF.jar 至 OBS 上。上传数据至 OBS 可参考 3.1 创建并提交 Spark SQL 作业中“步骤 2: 上传数据至 OBS”。
4. 将 TestUDF.jar 上传到 DLI 的程序包管理中，方便后续统一管理。
 - a. 登录 DLI 管理控制台，单击“数据管理 > 程序包管理”。
 - b. 在“程序包管理”页面，单击右上角的“创建”创建程序包。
 - c. 在“创建程序包”对话框，配置以下参数。
 - i. 包类型：选择“JAR”。
 - ii. OBS 路径：3 中程序包所在的 OBS 路径。
 - iii. 分组设置和组名称根据情况选择设置，方便后续识别和管理程序包。
 - d. 单击“确定”，完成创建程序包。
5. 创建 Function。

在 DLI Console 上输入下列命令创建 Function:

```
CREATE FUNCTION fun1 AS 'com.demo.SumUdfDemo' using jar 'obs://udf/TestUDF.jar';
```

6. 重启原有 SQL 队列，使得创建的 Function 生效。
 - a. 登录数据湖探索管理控制台，选择“队列管理”，在对应“SQL 队列”类型作业的“操作”列，单击“重启”。
 - b. 在“重启队列”界面，选择“确定”完成队列重启。

7. 使用 Function。

在查询语句中使用 5 中创建的 Function:

```
select fun1(ip) from ip_tables;
```

8. 删除 Function。

如果不再使用该 Function，可执行以下语句删除 Function:

```
Drop FUNCTION fun1;
```

样例代码

SumUdfDemo.java 中的样例代码如下所示:

```
package com.demo;
import org.apache.hadoop.hive.ql.exec.UDF;
public class SumUdfDemo extends UDF {
    public Int evaluate(Int a, Int b) {
        return a + b;
    }
}
```