

数据湖探索

快速入门

文档版本 01
发布日期 2025-01-09



版权所有 © 华为技术有限公司 2025。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

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

如企业客户须获取漏洞信息，请参见如下网址：

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

目录

1 使用 DLI 提交 SQL 作业查询 OBS 数据.....	1
2 使用 DLI 提交 SQL 作业查询 RDS MySQL 数据.....	6
3 使用 DLI 提交 Flink OpenSource SQL 作业查询 RDS MySQL 数据.....	16
4 使用 DLI 提交 Flink Jar 作业.....	29
5 使用 DLI 提交 Spark Jar 作业.....	38
6 入门实践.....	44

1 使用 DLI 提交 SQL 作业查询 OBS 数据

场景描述

DLI可以查询存储在OBS中的数据，本节操作介绍使用DLI提交SQL作业查询OBS数据的操作步骤。

本例新建“sampledata.csv”文件上传OBS桶，并新建弹性资源池队列，使用DLI创建数据库和表，使用DLI提供的SQL编辑器查询表的1000条数据。

操作流程

使用DLI提交SQL作业查询OBS数据的基本流程如表1-1所示。

开始进行如下操作前，请务必参考[准备工作](#)完成必要操作。

表 1-1 使用 DLI 提交 SQL 作业查询 OBS 数据的操作步骤

操作步骤	说明
步骤1: 上传数据至OBS	使用DLI查询数据前，需要将数据文件上传至OBS中。
步骤2: 创建弹性资源池并添加队列	创建提交作业所需的计算资源。
步骤3: 创建数据库	DLI元数据是SQL作业开发的基础。在执行作业前您需要根据业务场景定义数据库和表。
步骤4: 创建表	数据库创建完成后，需要在数据库db1中基于OBS上的样本数据创建表。
步骤5: 查询数据	使用标准SQL语句进行数据的查询和分析。

准备工作

- 已注册华为账号并开通华为云，且在使用DLI前检查账号状态，账号不能处于欠费或冻结状态。
- 配置DLI委托访问授权
DLI使用过程中涉及到OBS、VPC、SMN等服务交互，首次使用DLI需要用户配置委托授权，允许访问这些依赖服务。

- a. 使用华为云账号登录DLI管理控制台，在左侧导航栏单击“全局配置 > 服务授权”。
- b. 在委托设置页面，勾选基础使用、跨源场景、运维场景的委托权限后，单击“更新委托权限”。
- c. 查看并了解更新委托的提示信息，单击“确定”。完成DLI委托权限的更新。

图 1-1 配置 DLI 委托访问授权



- d. 完成配置后，在IAM控制台的委托列表中，可查看到 dli_management_agency 的委托信息。

步骤 1：上传数据至 OBS

使用DLI查询数据前，需要将数据文件上传至OBS中。

1. 登录OBS管理控制台，进入OBS管理控制台页面。
2. 创建桶，本例桶名以“obs1”为例。
 - a. 单击页面右上角“创建桶”。
 - b. 进入“创建桶”页面，选择“区域”，输入“桶名称”。其他参数保持默认值或根据需要选择。

说明

创建OBS桶时，需要选择与DLI管理控制台相同的区域，不可跨区域执行操作。

- c. 单击“立即创建”。
3. 单击所建桶“obs1”，进入“对象”页面。
4. 勾选左侧列表中的“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”。

```
product_id,product_name
113,office_13
```

22,book_2
29,book_9

文件上传成功后，待分析的文件路径为“obs://obs1/sampleddata.csv”。
关于OBS管理控制台更多操作请参考《对象存储服务用户指南》。

步骤 2：创建弹性资源池并添加队列

本例创建弹性资源池“dli_resource_pool”、队列“dli_queue_01”。

1. 登录DLI管理控制台。
2. 在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。
3. 在弹性资源池管理界面，单击界面右上角的“购买弹性资源池”。
4. 在“购买弹性资源池”界面，填写具体的弹性资源池参数。

本例在华东-上海二区域购买按需计费的弹性资源池。相关参数说明如表1-2所示。

表 1-2 参数说明

参数名称	参数说明	配置样例
区域	选择弹性资源池所在区域。	华东-上海二
项目	每个区域默认对应一个项目，由系统预置。	系统默认项目
名称	弹性资源池名称。	dli_resource_pool
规格	选择弹性资源池规格。	标准版
CU范围	弹性资源池最大最小CU范围。	64-64
网段	规划弹性资源池所属的网段。如需使用DLI增强型跨源，弹性资源池网段与数据源网段不能重合。 弹性资源池网段设置后不支持更改。	172.16.0.0/19
企业项目	选择对应的企业项目。	default

5. 参数填写完成后，单击“立即购买”，在界面上确认当前配置是否正确。
6. 单击“提交”完成弹性资源池的创建。
7. 在弹性资源池的列表页，选择要操作的弹性资源池，单击操作列的“添加队列”。
8. 配置队列的基础配置，具体参数信息如下。

表 1-3 弹性资源池添加队列基础配置

参数名称	参数说明	配置样例
名称	弹性资源池添加的队列名称。	dli_queue_01

参数名称	参数说明	配置样例
类型	选择创建的队列类型。 <ul style="list-style-type: none"> • 执行SQL作业请选择SQL队列。 • 执行Flink或Spark作业请选择通用队列。 	-
执行引擎	SQL队列可以选择队列引擎为Spark或者HetuEngine。	-
企业项目	选择对应的企业项目。	default

9. 单击“下一步”，配置队列的扩缩容策略。
 单击“新增”，可以添加不同优先级、时间段、“最小CU”和“最大CU”扩缩容策略。
 本例配置的扩缩容策略如图1-2所示。

图 1-2 添加队列时配置扩缩容策略



表 1-4 扩缩容策略参数说明

参数名称	参数说明	配置样例
优先级	当前弹性资源池中的优先级数字越大表示优先级越高。本例设置一条扩缩容策略，默认优先级为1。	1
时间段	首条扩缩容策略是默认策略，不能删除和修改时间段配置。 即设置00-24点的扩缩容策略。	00-24
最小CU	设置扩缩容策略支持的最小CU数。	16
最大CU	当前扩缩容策略支持的最大CU数。	64

10. 单击“确定”完成添加队列配置。

步骤 3: 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。

1. 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“数据库”页面。
2. 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，左侧单击“”刷新数据库页面，新建的数据库db1会在“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 4：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 (product_id int, product_name string) using csv options (path 'obs://obs1');
```

创建表时只需指定数据文件所在的OBS存储路径，目录后不需要指定文件名。

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

步骤 5：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/失败后，可在SQL作业编辑窗口下方“查看结果”页签查看查询结果。

2 使用 DLI 提交 SQL 作业查询 RDS MySQL 数据

操作场景

DLI可以查询存储在RDS数据库中的数据，本节操作介绍使用DLI提交SQL作业查询RDS MySQL数据的操作步骤。

本例创建RDS实例并新建数据库和表，创建DLI弹性资源池队列，通过增强型跨源连接打通DLI弹性资源池和RDS实例的网络，并提交SQL作业跨源访问RDS的表数据。

操作流程

使用DLI提交SQL作业查询RDS MySQL数据操作流程如表2-1所示。

开始进行如下操作前，请务必参考[准备工作](#)完成必要操作。

表 2-1 使用 DLI 提交 SQL 作业查询 RDS MySQL 数据的操作流程

操作步骤	说明
步骤1: 创建RDS MySQL实例	样例场景需要创建RDS MySQL实例。
步骤2: 创建RDS数据库表	登录RDS实例创建数据库和表。
步骤3: 创建弹性资源池并添加队列	创建提交作业所需的计算资源。
步骤4: 创建增强型跨源连接	通过增强型跨源连接建立DLI弹性资源池与RDS实例的网络连通。
步骤5: 创建跨源认证	使用跨源认证保存DLI读写RDS数据的访问凭证。
步骤6: 提交SQL作业	使用标准SQL语句进行数据的查询和分析。

准备工作

- 已注册华为账号并开通华为云，且在使用DLI前检查账号状态，账号不能处于欠费或冻结状态。
- 配置DLI委托访问授权
DLI使用过程中涉及到OBS、VPC、SMN等服务交互，首次使用DLI需要用户配置委托授权，允许访问这些依赖服务。
 - a. 使用华为云账号登录DLI管理控制台，在左侧导航栏单击“全局配置 > 服务授权”。
 - b. 在委托设置页面，勾选基础使用、跨源场景、运维场景的委托权限后，单击“更新委托权限”。
 - c. 查看并了解更新委托的提示信息，单击“确定”。完成DLI委托权限的更新。

图 2-1 配置 DLI 委托访问授权



- d. 完成配置后，在IAM控制台的委托列表中，可查看到dli_management_agency的委托信息。

步骤 1: 创建 RDS MySQL 实例

本样例中，假设作业名称为“JobSample”，采用RDS服务作为数据源，创建RDS MySQL实例。

具体操作请参见[购买RDS for MySQL实例](#)。

1. 登录RDS管理控制台。
2. 在管理控制台左上角选择区域和项目。
3. 在“实例管理”页面，单击“购买数据库实例”。
4. 在购买数据库实例页面，选择计费模式，填写并选择实例相关信息后，单击“立即购买”。

根据业务规划配置参数信息，本例参数配置信息如所示。

更多RDS实例的参数配置信息请参考[购买RDS实例](#)。

表 2-2 RDS MySQL 实例参数配置信息

参数名称	参数说明	取值样例
计费模式	选择RDS实例的计费模式。	按需计费
区域	资源所在的区域。	华东-上海二
实例名称	实例名称。	rds-demo
数据库引擎	MySQL	MySQL
数据库版本	选用RDS for MySQL数据库时，请根据实际业务需求选择合适的数据库引擎版本。建议您选择当前可用的最高版本数据库，因其性能更稳定，安全性更高，使用更可靠。	8.0
实例类型	选择实例的主备类型。	单机
存储类型	实例的存储类型决定实例的读写速度。最大吞吐量越高，读写速度越快。	SSD云盘
可用区	对于单机实例，仅需选择单个可用区。	-
时区	由于世界各国与地区经度不同，地方时也有所不同，因此会划分为不同的时区。时区可在创建实例时选择，后期可修改。	默认
性能规格	实例的CPU和内存。不同性能规格对应不同连接数和最大IOPS。	2vCPUs 4GB
存储空间	如果存储类型为SSD云盘或极速型SSD，可设置存储空间自动扩容，当存储空间可用率过小时，会自动扩容存储空间。	40GB
磁盘加密	选择是否开启磁盘加密功能。	不加密
虚拟私有云	选择已有的虚拟私有云。如需重新创建VPC和子网，请参考。 说明 跨源场景数据源网段和弹性资源池的网段不能重合。	-
数据库端口	默认使用3306端口。	3306

参数名称	参数说明	取值样例
安全组	安全组限制实例的安全访问规则，加强云数据库RDS服务与其他服务间的安全访问。 数据源的安全组需放通DLI弹性资源池的网段。	-
设置密码	设置实例的登录密码。	-
管理员账号	root	root
管理员密码	设置管理员密码。	-
参数模板	数据库参数模板就像是数据库引擎配置值的容器，参数模板中的参数可应用于一个或多个相同类型的数据库实例。	系统默认模板：Default-MySQL-5.7
表名大小写	配置表名不区分大小写。	不区分大小写
企业项目	对于已成功关联企业项目的用户，仅需在“企业项目”下拉框中选择目标项目。	default
购买数量	实例购买数量	1

5. 单击“立即购买”，进入规格确认页面。
6. 单击“提交”，完成购买RDS MySQL实例。
7. 实例创建成功后，用户可以在“实例管理”页面对其进行查看和管理。

创建实例过程中，状态显示为“创建中”，创建完成的实例状态显示为“正常”。您可以通过“任务中心”查看详细进度和结果。

步骤 2：创建 RDS 数据库表

1. 登录RDS管理控制台。
2. 在管理控制台左上角选择区域和项目。
3. 在“实例管理页面”，找到您已经创建的RDS实例，记录并保存RDS实例的内网地址。

图 2-2 内网地址



4. 单击所创建RDS实例的“登录”，跳转至“数据管理服务-DAS”。输入相关账户信息，单击“测试连接”。显示连接成功后，单击“登录”，进入“实例登录”页面。

图 2-3 登录 RDS



图 2-4 实例登录



5. 登录RDS实例后，单击“新建数据库”，创建名称为“dli_demo”的数据库。
6. 单击“SQL查询”，执行如下SQL创建表。

```
CREATE TABLE `dli_demo`.`tabletest` (
  `id` VARCHAR(32) NOT NULL,
  `name` VARCHAR(32) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4;
```

步骤 3：创建弹性资源池并添加队列

执行跨源场景的SQL作业，不能使用系统已有的default队列，需要您创建SQL队列，本例创建弹性资源池“dli_resource_pool”、队列“dli_queue_01”。

1. 登录DLI管理控制台。
2. 在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。
3. 在弹性资源池管理界面，单击界面右上角的“购买弹性资源池”。
4. 在“购买弹性资源池”界面，填写具体的弹性资源池参数。

本例在华东-上海二区域购买按需计费的弹性资源池。相关参数说明如表2-3所示。

表 2-3 参数说明

参数名称	参数说明	配置样例
区域	选择弹性资源池所在区域。	华东-上海二

参数名称	参数说明	配置样例
项目	每个区域默认对应一个项目，由系统预置。	系统默认项目
名称	弹性资源池名称。	dli_resource_pool
规格	选择弹性资源池规格。	标准版
CU范围	弹性资源池最大最小CU范围。	64-64
网段	规划弹性资源池所属的网段。如需使用DLI增强型跨源，弹性资源池网段与数据源网段不能重合。 弹性资源池网段设置后不支持更改。	172.16.0.0/19
企业项目	选择对应的企业项目。	default

5. 参数填写完成后，单击“立即购买”，在界面上确认当前配置是否正确。
6. 单击“提交”完成弹性资源池的创建。
7. 在弹性资源池的列表页，选择要操作的弹性资源池，单击操作列的“添加队列”。
8. 配置队列的基础配置，具体参数信息如下。

表 2-4 弹性资源池添加队列基础配置

参数名称	参数说明	配置样例
名称	弹性资源池添加的队列名称。	dli_queue_01
类型	选择创建的队列类型。 <ul style="list-style-type: none"> • 执行SQL作业请选择SQL队列。 • 执行Flink或Spark作业请选择通用队列。 	-
执行引擎	SQL队列可以选择队列引擎为Spark或者HetuEngine。	-
企业项目	选择对应的企业项目。	default

9. 单击“下一步”，配置队列的扩缩容策略。
单击“新增”，可以添加不同优先级、时间段、“最小CU”和“最大CU”扩缩容策略。
本例配置的扩缩容策略如[图2-5](#)所示。

图 2-5 添加队列时配置扩缩容策略



表 2-5 扩缩容策略参数说明

参数名称	参数说明	配置样例
优先级	当前弹性资源池中的优先级数字越大表示优先级越高。本例设置一条扩缩容策略，默认优先级为1。	1
时间段	首条扩缩容策略是默认策略，不能删除和修改时间段配置。 即设置00-24点的扩缩容策略。	00-24
最小CU	设置扩缩容策略支持的最小CU数。	16
最大CU	当前扩缩容策略支持的最大CU数。	64

10. 单击“确定”完成添加队列配置。

步骤 4：创建增强型跨源连接

步骤1 在RDS实例所在的安全组上放通DLI队列网段

1. 在RDS管理控制台，在RDS实例列表页单击对应的RDS实例名称，进入到RDS实例的基本信息页面。
2. 单击“连接管理”配置安全组规则，在“入方向规则”中添加放通队列网段的规则。

例如，本示例队列网段为“172.16.0.0/19”，规则添加为：

- 优先级选择：1，策略选择：允许
- 类型：IPv4
- 协议端口：基本协议\自定义TCP，端口值不填
- 源地址为：选择IP地址，172.16.0.0/19

单击“确定”完成安全组规则添加。

步骤2 创建RDS与DLI的增强型跨源连接。

创建增强型跨源连接具体操作请参考[创建增强型跨源连接](#)。

说明

绑定跨源的弹性资源池的网段和数据源网段不能重合。

1. 在DLI管理控制台左侧导航栏中，选择“跨源管理”。
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：dlirds
 - 弹性资源池：选择**步骤3：创建弹性资源池并添加队列**中创建的弹性资源池。
 - 虚拟私有云：选择MySQL实例所在的VPC，即**步骤2：创建RDS数据库表**中选择的VPC。
 - 子网：选择MySQL实例所在的子网，即**步骤2：创建RDS数据库表**中选择的子网。

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”>“子网”，也可获取对应信息。
3. 单击“确定”，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：dlirds，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。
5. 测试队列与RDS实例连通性。
 - a. 单击“资源管理 > 队列管理”，选择**步骤3：创建弹性资源池并添加队列**添加的队列，单击“操作”列中的“更多 > 测试地址连通性”。

图 2-6 检测地址连通性



- b. 输入**步骤2：创建RDS数据库表**中记录的RDS MySQL实例内网地址，进行网络连通性测试。

说明

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”，即可获取对应信息。

- 测试结果显示可达，则DLI队列与MySQL实例的网络已经联通。

图 2-7 测试结果

测试地址连通性



- 如果测试结果不可达，请参考**创建跨源成功但测试网络连通性失败怎么办？**进行排查。

排查完成后，重新进行网络连通性测试。

---结束

步骤 5：创建跨源认证

跨源分析场景中，如果在作业中直接配置认证信息会触发密码泄露的风险，因此推荐您使用DLI提供的跨源认证方式来存储数据源的认证信息。

RDS数据源在Spark SQL作业场景可以创建Password类型的跨源认证。

1. 登录DLI管理控制台。
2. 选择“跨源管理 > 跨源认证”。
3. 单击“创建”。

填写认证信息，详细参数说明请参考[表2-6](#)。

表 2-6 跨源认证参数说明

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RDS实例用户名。
用户密码	RDS实例登录密码。

图 2-8 创建认证信息-Password

步骤 6: 提交 SQL 作业

本节示例以SQL作业为例说明如何跨源访问RDS表。

1. 在DLI管理控制台的左侧导航栏中，单击“SQL编辑器”，进入SQL作业编辑器页面。
2. 在“SQL编辑器”页面右侧的编辑窗口中，输入如下SQL语句创建数据库db1，单击“执行”。
`create database db1;`
3. 在编辑窗口中，选择队列“dli_queue_01”和数据库“db1”。输入如下SQL语句创建表，向RDS表插入数据，以及查询数据，单击“执行”。

根据界面显示的执行结果查询RDS表数据信息，整个操作完成。

```
CREATE TABLE IF NOT EXISTS rds_test USING JDBC OPTIONS (  
  'url' = 'jdbc:mysql://{ip}:{port}', //RDS内网地址及端口  
  'driver' = 'com.mysql.jdbc.Driver',  
  'dbtable' = 'dli_demo.tabletest', //在RDS创建的库名及表名  
  'passwdauth'='xxxxx' // DLI侧创建的Password类型的跨源认证名称。使用跨源认证则无需在作业中配置  
  账号和密码。  
)  
  
insert into rds_test VALUES ('123','abc');  
  
SELECT * from rds_test;
```

3 使用 DLI 提交 Flink OpenSource SQL 作业 查询 RDS MySQL 数据

操作场景

DLI Flink作业支持使用其他服务作为数据源和数据输出通道进行数据实时计算操作。

本例采用Kafka服务作为数据源通道，以RDS作为数据输出通道，介绍创建并提交Flink OpenSource SQL作业进行实时计算的操作步骤。

操作流程

样例场景需要创建一个Flink OpenSource SQL作业，且该作业包含一个输入流和一个输出流。输入流用于从Kafka读取数据，输出流用于将数据写入到RDS中。操作流程如[操作流程](#)所示。

开始进行如下操作前，请务必参考[准备工作](#)完成必要操作。

表 3-1 使用 DLI 提交 SQL 作业查询 RDS MySQL 数据的操作流程

操作步骤	说明
步骤1: 准备数据源通道	本样例场景需要创建Kafka实例作为数据源通道。
步骤2: 准备数据输出通道	本样例场景需要创建RDS实例作为数据输出通道。
步骤3: 创建OBS桶保存输出数据	创建OBS桶，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。
步骤4: 创建弹性资源池并添加队列	创建提交Flink作业所需的计算资源。
步骤5: 创建DLI连接Kafka的增强型跨源连接	通过增强型跨源连接建立DLI弹性资源池与Kafka实例的网络连通。
步骤6: 创建DLI连接RDS的增强型跨源连接	通过增强型跨源连接建立DLI弹性资源池与RDS实例的网络连通。

操作步骤	说明
步骤7: 使用DEW管理访问凭据, 并配置允许DLI访问DEW的委托	跨源分析场景中, 使用DEW管理数据源的访问凭证, 并创建允许DLI访问DEW的委托。
步骤8: 创建Flink OpenSource SQL作业	准备好数据源和数据输出通道后创建Flink OpenSource SQL作业分析数据。

准备工作

- 已注册华为账号并开通华为云, 且在使用DLI前检查账号状态, 账号不能处于欠费或冻结状态。
- 配置DLI委托访问授权
DLI使用过程中涉及到OBS、VPC、SMN等服务交互, 首次使用DLI需要用户配置委托授权, 允许访问这些依赖服务。
 - a. 使用华为云账号登录DLI管理控制台, 在左侧导航栏单击“全局配置 > 服务授权”。
 - b. 在委托设置页面, 勾选基础使用、跨源场景、运维场景的委托权限后, 单击“更新委托权限”。
 - c. 查看并了解更新委托的提示信息, 单击“确定”。完成DLI委托权限的更新。

图 3-1 配置 DLI 委托访问授权



- d. 完成配置后, 在IAM控制台的委托列表中, 可查看到 dli_management_agency 的委托信息。

步骤 1: 准备数据源通道

本例以Kafka数据作为数据源通道。

开通Kafka数据接入服务, 具体操作请参见[创建Kafka实例](#)。

1. 创建Kafka相关依赖资源

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[创建虚拟私有云和子网](#)，如果需要在已有VPC上创建和使用新的子网，请参考[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[创建安全组](#)，为安全组添加规则的操作指导请参考[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。

2. 创建用于作业输入流的Kafka专享版实例

- 登录分布式消息服务Kafka管理控制台。
- 在管理控制台左上角选择区域。
- 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：

更多Kafka专享版实例参数配置说明请参考[购买Kafka实例](#)。

表 3-2 购买 Kafka 专享版实例参数说明

参数名称	参数说明	配置样例
区域	不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	华东-上海二
可用区	可用区指在同一区域下，电力、网络隔离的物理区域，可用区之间内网互通，不同可用区之间物理隔离。	可用区1、可用区2、可用区3
版本	Kafka的版本号，实例创建后，版本号不支持修改。	3.x
部署架构	根据需求选择“单机”或“集群”。	集群
代理规格	请根据业务需求选择相应的代理规格。	kafka.2u4g.cluster.small
代理数量	选择代理数量。	3
单个代理存储空间	选择存储Kafka数据的磁盘类型和磁盘大小。Kafka实例创建后，磁盘类型不支持修改。	高I/O 100GB
虚拟私有云	选择已经创建好的或共享的虚拟私有云。	选择1中创建的虚拟私有云

参数名称	参数说明	配置样例
子网	选择已经创建好的或共享的子网。 子网在Kafka实例创建完成后，不支持修改。	选择1中创建的子网
安全组	选择已经创建好的安全组。	选择1中创建的安全组
接入方式	接入方式分为以下两种： <ul style="list-style-type: none"> 明文接入：表示客户端连接Kafka实例时，无需进行SASL认证。 密文接入：表示客户端连接Kafka实例时，需要进行SASL认证。 	明文接入
实例名称	按命名规则自定义实例名称。	kafka-dliflink
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，默认项目为default。	default

- d. 单击“立即购买”，弹出规格确认页面。
 - e. 确确实实例信息无误且阅读并同意《华为云用户协议》后，单击“提交”，完成实例创建,创建实例大约需要10~15分钟请耐心等待。
3. 创建Kafka的topic。
- a. 单击购买的Kafka实例名称，进入到Kafka实例的基本信息页面。
 - b. 单击“Topic管理 > 创建Topic”，创建一个Topic。Topic配置参数如下：
 - Topic名称。本示例输入为：testkafkatopic。
 - 分区数：1。
 - 副本数：1。
 其他参数保持默认即可。

步骤 2：准备数据输出通道

采用RDS作为数据输出通道，创建RDS MySQL实例。

1. 登录RDS管理控制台。
2. 在管理控制台左上角选择区域。
3. 单击“购买数据库实例”配置相关参数。主要参数的填下说明如下，其他参数保持默认值即可。

详细参数说明请参考[购买RDS for MySQL实例](#)。

表 3-3 RDS MySQL 实例参数配置信息

参数名称	参数说明	取值样例
计费模式	选择RDS实例的计费模式。	按需计费
区域	选择与DLI服务相同的区域。	华东-上海二
实例名称	实例名称。	rds-dliflink
数据库引擎	MySQL	MySQL
数据库版本	选用RDS for MySQL数据库时，请根据实际业务需求选择合适的数据库引擎版本。建议您选择当前可用的最高版本数据库，因其性能更稳定，安全性更高，使用更可靠。	8.0
实例类型	选择实例的主备类型。	主备
存储类型	实例的存储类型决定实例的读写速度。最大吞吐量越高，读写速度越快。	SSD云盘
可用区	对于单机实例，仅需选择单个可用区。	自定义
时区	由于世界各国与地区经度不同，地方时也有所不同，因此会划分为不同的时区。时区可在创建实例时选择，后期可修改。	默认
性能规格	实例的CPU和内存。不同性能规格对应不同连接数和最大IOPS。	2vCPUs 4GB
存储空间	如果存储类型为SSD云盘或极速型SSD，可设置存储空间自动扩容，当存储空间可用率过小时，会自动扩容存储空间。	40GB
磁盘加密	选择是否开启磁盘加密功能。	不加密
虚拟私有云、子网	选择已有的虚拟私有云和子网。如需重新创建VPC和子网，请参考。 说明 跨源场景数据源网段和弹性资源池的网段不能重合。	选择1中创建的虚拟私有云和子网。
数据库端口	默认使用3306端口。	3306
安全组	安全组限制实例的安全访问规则，加强云数据库RDS服务与其他服务间的安全访问。 数据源的安全组需放通DLI弹性资源池的网段。	选择1中创建的安全组。

参数名称	参数说明	取值样例
设置密码	设置实例的登录密码。	-
管理员账号	root	root
管理员密码	设置管理员密码。	-
参数模板	数据库参数模板就像是数据库引擎配置值的容器，参数模板中的参数可应用于一个或多个相同类型的数据库实例。	Default-MySQL-8.0
企业项目	对于已成功关联企业项目的用户，仅需在“企业项目”下拉框中选择目标项目。	default
购买数量	实例购买数量。	1

4. 单击“立即购买”，确认规格信息。
5. 单击“提交”，完成RDS实例的创建。
6. 登录MySQL，并使用下述命令在flink库下创建orders表。

登录MySQL，单击“SQL窗口”，在SQL查询页面输入以下创建表语句，创建RDS MySQL表。

```
CREATE TABLE `flink`.`orders` (
  `order_id` VARCHAR(32) NOT NULL,
  `order_channel` VARCHAR(32) NULL,
  `order_time` VARCHAR(32) NULL,
  `pay_amount` DOUBLE UNSIGNED NOT NULL,
  `real_pay` DOUBLE UNSIGNED NULL,
  `pay_time` VARCHAR(32) NULL,
  `user_id` VARCHAR(32) NULL,
  `user_name` VARCHAR(32) NULL,
  `area_id` VARCHAR(32) NULL,
  PRIMARY KEY (`order_id`)
) ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_general_ci;
```

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。

表 3-4 OBS 桶参数说明

参数名称	参数说明	取值样例
区域	桶所属区域。请选择与DLI服务相同的区域。	华东-上海二

参数名称	参数说明	取值样例
桶名称	桶的名称。需全局唯一，不能与已有的任何桶名称重复，包括其他用户创建的桶。桶创建成功后，不支持修改名称，创建时，请设置合适的桶名。	obstest
默认存储类别	桶的存储类别。不同的存储类别可以满足客户业务对存储性能、成本的不同诉求。	标准存储
桶策略	桶的读写权限控制。 私有桶除桶ACL授权外的其他用户无桶的访问权限。	私有
企业项目	将桶加入到企业项目中统一管理。	default

- 单击“立即创建”。

步骤 4：创建弹性资源池并添加队列

创建DLI Flink OpenSource SQL作业，不能使用系统已有的default队列，需要您创建新队列，本例创建弹性资源池“dli_resource_pool”、队列“dli_queue_01”。

- 登录DLI管理控制台。
- 在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。
- 在弹性资源池管理界面，单击界面右上角的“购买弹性资源池”。
- 在“购买弹性资源池”界面，填写具体的弹性资源池参数。

本例在华东-上海二区域购买按需计费的弹性资源池。相关参数说明如表3-5所示。

表 3-5 参数说明

参数名称	参数说明	配置样例
区域	选择弹性资源池所在区域。	华东-上海二
项目	每个区域默认对应一个项目，由系统预置。	系统默认项目
名称	弹性资源池名称。	dli_resource_pool
规格	选择弹性资源池规格。	标准版
CU范围	弹性资源池最大最小CU范围。	64-64
网段	规划弹性资源池所属的网段。如需使用DLI增强型跨源，弹性资源池网段与数据源网段不能重合。 弹性资源池网段设置后不支持更改。	172.16.0.0/19

参数名称	参数说明	配置样例
企业项目	选择对应的企业项目。	default

- 参数填写完成后，单击“立即购买”，在界面上确认当前配置是否正确。
- 单击“提交”完成弹性资源池的创建。
- 在弹性资源池的列表页，选择要操作的弹性资源池，单击操作列的“添加队列”。
- 配置队列的基础配置，具体参数信息如下。

表 3-6 弹性资源池添加队列基础配置

参数名称	参数说明	配置样例
名称	弹性资源池添加的队列名称。	dli_queue_01
类型	选择创建的队列类型。 <ul style="list-style-type: none"> 执行SQL作业请选择SQL队列。 执行Flink或Spark作业请选择通用队列。 	-
执行引擎	SQL队列可以选择队列引擎为Spark或者HetuEngine。	-
企业项目	选择对应的企业项目。	default

- 单击“下一步”，配置队列的扩缩容策略。
单击“新增”，可以添加不同优先级、时间段、“最小CU”和“最大CU”扩缩容策略。
本例配置的扩缩容策略如图3-2所示。

图 3-2 添加队列时配置扩缩容策略



表 3-7 扩缩容策略参数说明

参数名称	参数说明	配置样例
优先级	当前弹性资源池中的优先级数字越大表示优先级越高。本例设置一条扩缩容策略，默认优先级为1。	1

参数名称	参数说明	配置样例
时间段	首条扩缩容策略是默认策略，不能删除和修改时间段配置。 即设置00-24点的扩缩容策略。	00-24
最小CU	设置扩缩容策略支持的最小CU数。	16
最大CU	当前扩缩容策略支持的最大CU数。	64

10. 单击“确定”完成添加队列配置。

步骤 5: 创建 DLI 连接 Kafka 的增强型跨源连接

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

- 增强型跨源仅支持按需专属队列。
- 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。

步骤1 在Kafka的安全组上放通DLI队列网段

1. 在Kafka管理控制台，选择“Kafka专享版”，单击对应的Kafka名称，进入到Kafka的基本信息页面。
2. 在“连接信息”中获取该Kafka的“内网连接地址”，在“基本信息”的“网络”中获取该实例的“虚拟私有云”和“子网”信息，方便后续操作步骤使用。
3. 单击“网络”中的安全组名称，在“入方向规则”中添加放通队列网段的规则。
例如，本示例队列网段为“10.0.0.0/16”，则规则添加为：优先级选择：1，策略选择：允许，协议选择：TCP，端口值不填，类型：IPv4，源地址为：10.0.0.0/16，单击“确定”完成安全组规则添加。

步骤2 创建DLI队列连接Kafka的增强型跨源连接。

1. 登录DLI管理控制台，在左侧导航栏单击“跨源管理”，在跨源管理界面，单击“增强型跨源”，单击“创建”。
2. 在增强型跨源创建界面，配置具体的跨源连接参数。具体参考如下。
 - 连接名称：设置具体的增强型跨源名称。本示例输入为：dli_kafka。
 - 弹性资源池：选择[步骤4: 创建弹性资源池并添加队列](#)中已经创建的资源名称。
 - 虚拟私有云：选择Kafka的虚拟私有云。
 - 子网：选择Kafka的子网。
 - 其他参数可以根据需要选择配置。

参数配置完成后，单击“确定”完成增强型跨源配置。单击创建的跨源连接名称，查看跨源连接的连接状态，等待连接状态为“已激活”后可以进行后续步骤。

3. 单击“资源管理 > 队列管理”，选择操作的队列，本示例为[步骤4: 创建弹性资源池并添加队列](#)中创建的队列，在操作列，单击“更多 > 测试地址连通性”。

- 在“测试连通性”界面，地址栏输入“Kafka内网地址:Kafka数据库端口”，单击“测试”测试DLI到Kafka网络是否可达。注意多个地址要分开单独测试。

----结束

步骤 6: 创建 DLI 连接 RDS 的增强型跨源连接

步骤1 在RDS的安全组上放通DLI队列网段。

如果RDS和Kafka在同一VPC下的同一安全组，则无需重复执行此步骤，在步骤**步骤1**中已在该安全组放通DLI的队列网段。

- 在RDS管理控制台，选择“实例管理”，单击对应的RDS实例名称，进入到RDS的基本信息页面。
- 在“基本信息”的“连接信息”中获取该实例的“内网地址”、“数据库端口”、“虚拟私有云”和“子网”信息，方便后续操作步骤使用。
- 单击“连接信息”中的安全组名称，在“入方向规则”中添加放通队列网段的规则。例如，本示例队列网段为“10.0.0.0/16”，则规则添加为：优先级选择：1，策略选择：允许，协议选择：TCP，端口值不填，类型：IPv4，源地址为：10.0.0.0/16，单击“确定”完成安全组规则添加。

步骤2 创建DLI队列连接RDS的增强型跨源连接。

如果RDS和Kafka在同一VPC和子网，则无需重复执行此步骤，在步骤**步骤2**中创建的增强型跨源连接已将网络打通。

如果Kafka和RDS实例分别在两个VPC和子网下，则要执行以下步骤创建DLI队列连接RDS的增强型跨源连接。

- 登录DLI管理控制台，在左侧导航栏单击“跨源管理”，在跨源管理界面，单击“增强型跨源”，单击“创建”。
- 在增强型跨源创建界面，配置具体的跨源连接参数。具体参考如下。
 - 连接名称：设置具体的增强型跨源名称。本示例输入为：dli_rds。
 - 弹性资源池：选择**步骤4: 创建弹性资源池并添加队列**中已经创建的队列名称。
 - 虚拟私有云：选择RDS的虚拟私有云。
 - 子网：选择RDS的子网。
 - 其他参数可以根据需要选择配置。

参数配置完成后，单击“确定”完成增强型跨源配置。单击创建的跨源连接名称，查看跨源连接的连接状态，等待连接状态为：“已激活”后可以进行后续步骤。

- 单击“资源管理 > 队列管理”，选择操作的队列，本示例为**步骤4: 创建弹性资源池并添加队列**中创建的队列，在操作列，单击“更多 > 测试地址连通性”。
- 在“测试连通性”界面，地址栏输入“RDS内网地址:RDS数据库端口”，单击“测试”测试DLI到RDS网络是否可达。

----结束

步骤 7: 使用 DEW 管理访问凭据，并配置允许 DLI 访问 DEW 的委托

跨源分析场景中，需要在connector中设置账号、密码等属性。但是账号密码等信息属于高度敏感数据，需要做加密处理，以保障用户的数据隐私安全。Flink1.15版本支持使用DEW管理凭据，在执行作业前需创建自定义委托并在作业中配置委托信息。

数据加密服务（Data Encryption Workshop, DEW）、云凭据管理服务（Cloud Secret Management Service, CSMS），提供一种安全、可靠、简单易用隐私数据加密解密方案。本例介绍Flink OpenSource SQL使用DEW管理RDS访问凭据的配置方法。

1. 创建DLI访问DEW的委托并完成委托授权。
2. 在DEW创建通用凭证。
3. 登录DEW管理控制台
4. 选择“凭据管理”，进入“凭据管理”页面。
5. 单击“创建凭据”，配置凭据基本信息
 - 凭据名称：待创建凭据的名称。本例名称为secretInfo。
 - 凭据值：配置RDS实例的用户名和密码。
 - 第一行凭据值的键为MySQLUsername，值为RDS实例的用户名。
 - 第二行凭据值的键为MySQLPassword，值为RDS实例的密码。

图 3-3 设置凭据值

设置凭据值

凭据键/值 明文

MySQLUsername 删除

MySQLPassword 删除

+ 添加

6. 按需完成其他参数的配置后，单击“确定”保存凭据。

步骤 8：创建 Flink OpenSource SQL 作业

准备好数据源和数据输出通道之后，就可以创建Flink OpenSource 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 OpenSource SQL
 - 名称：JobSample
 - 描述：不填
 - 模板名称：不选择
 - 标签：不填
3. 单击“确定”，进入作业“编辑”页面。
4. 设置作业运行参数。配置必选参数：
 - 所属队列：dli_queue_01
 - Flink版本：Flink 1.12。
 - 保存作业日志：勾选。
 - OBS桶：选择保存作业日志的OBS桶，根据提示进行OBS桶权限授权。

- 开启Checkpoint: 勾选。

其余参数可不用配置。


5. 单击“保存”，保存作业和相关参数。
6. 编辑Flink OpenSource SQL作业。

在Flink OpenSource SQL语句编辑区域，输入详细的Flink OpenSource 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 TABLE kafkaSource (  
  order_id string,  
  order_channel string,  
  order_time string,  
  pay_amount double,  
  real_pay double,  
  pay_time string,  
  user_id string,  
  user_name string,  
  area_id string  
) WITH (  
  'connector' = 'kafka',  
  'topic' = 'testkafkatopic', // 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  
  'properties.bootstrap.servers' = "192.168.0.237:9092,192.168.0.252:9092,192.168.0.137:9092", // 替  
  换为kafka的内网连接地址和端口,  
  'properties.group.id' = 'Group1d',  
  'scan.startup.mode' = 'latest-offset',  
  'format' = 'json'  
)  
);  
  
CREATE TABLE jdbcSink (  
  order_id string,  
  order_channel string,  
  order_time string,  
  pay_amount double,  
  real_pay double,  
  pay_time string,  
  user_id string,  
  user_name string,  
  area_id string  
) WITH (  
  'connector' = 'jdbc',  
  'url' = "jdbc:mysql://172.16.0.116:3306/rds-dliflink", // testrdsdb为创建的RDS的数据库名，IP和端口  
  替换为RDS MySQL的实例IP和端口  
  'table-name' = 'orders',  
  'pwd_auth_name' = 'xxxxx', // DLI侧创建的Password类型的跨源认证名称，使用跨源认证则无需在作业  
  中配置账号和密码  
  'sink.buffer-flush.max-rows' = '1'  
)  
);  
  
insert into jdbcSink select * from kafkaSource;
```

7. 单击“语义校验”，确保语义校验成功。
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. 连接Kafka集群，向Kafka相应的topic中发送如下测试数据：
Kafka生产和发送数据的方法请参考：[连接实例生产消费信息](#)。

```
{"order_id":"202103241000000001", "order_channel":"webShop", "order_time":"2021-03-24 10:00:00",  
"pay_amount":"100.00", "real_pay":"100.00", "pay_time":"2021-03-24 10:02:03", "user_id":"0001",  
"user_name":"Alice", "area_id":"330106"}
```

```
{"order_id":"202103241606060001", "order_channel":"appShop", "order_time":"2021-03-24 16:06:06",  
"pay_amount":"200.00", "real_pay":"180.00", "pay_time":"2021-03-24 16:10:06", "user_id":"0001",  
"user_name":"Alice", "area_id":"330106"}
```

10. 查看表中数据，在MySQL中执行sql查询语句。

```
select * from orders;
```

其结果参考如下（以下数据为从MySQL中复制的结果）。

```
202103241000000001,webShop,2021-03-24 10:00:00,100.0,100.0,2021-03-24  
10:02:03,0001,Alice,330106  
202103241606060001,appShop,2021-03-24 16:06:06,200.0,180.0,2021-03-24  
16:10:06,0001,Alice,330106
```


4 使用 DLI 提交 Flink Jar 作业

操作场景

Flink Jar作业适用于需要自定义流处理逻辑、复杂的状态管理或特定库集成的数据分析场景。需要用户自行编写并构建Jar作业程序包，在提交Flink Jar作业前，将Jar作业程序包上传至OBS，并将程序包与数据和作业参数一起提交以运行作业。

本例介绍通过DLI控制台提交Flink Jar作业程序包的基本流程。由于不同的业务需求，Jar包的具体编写会有所差异。建议您参考DLI提供的示例代码，并根据实际业务场景进行相应的编辑和定制。获取[DLI样例代码](#)。

操作流程

使用DLI提交Flink Jar作业的操作流程请参考[表4-1](#)。

开始进行如下操作前，请务必参考[准备工作](#)完成必要操作。

表 4-1 使用 DLI 提交 Flink Jar 作业的操作流程

操作步骤	说明
步骤1: 开发Jar包并上传数据至OBS	准备Flink Jar作业程序包并上传至OBS。
步骤2: 购买弹性资源池并添加队列	创建提交Flink作业所需的计算资源。
步骤3: 使用DEW管理访问凭据	跨源分析场景中，使用DEW管理数据源的访问凭证。
步骤4: 创建自定义委托允许DLI访问DEW读取凭证	创建允许DLI访问DEW的委托。
步骤5: 创建Flink Jar作业并配置作业信息	创建Flink Jar作业分析数据。

准备工作

- 已注册华为账号并开通华为云，且在使用DLI前检查账号状态，账号不能处于欠费或冻结状态。
- 配置DLI委托访问授权
DLI使用过程中涉及到OBS、VPC、SMN等服务交互，首次使用DLI需要用户配置委托授权，允许访问这些依赖服务。
 - a. 使用华为云账号登录DLI管理控制台，在左侧导航栏单击“全局配置 > 服务授权”。
 - b. 在委托设置页面，勾选基础使用、跨源场景、运维场景的委托权限后，单击“更新委托权限”。
 - c. 查看并了解更新委托的提示信息，单击“确定”。完成DLI委托权限的更新。

图 4-1 配置 DLI 委托访问授权



- d. 完成配置后，在IAM控制台的委托列表中，可查看到dli_management_agency的委托信息。

步骤 1：开发 Jar 包并上传数据至 OBS

DLI控制台不提供Jar包的开发能力，您需要在线下完成Jar包的开发。Jar包的开发样例请参考Flink Jar开发基础样例。

参考[Flink作业样例代码](#)开发Flink Jar作业程序，编译并打包为“flink-examples.jar”。参考以下操作步骤上传该作业程序。

提交Flink作业之前，需要在OBS中上传数据文件。

1. 登录管理控制台
2. 在服务列表中，单击“存储”中的“对象存储服务OBS”，进入OBS管理控制台页面。
3. 创建桶，这里以桶名“dli-test-obs01”为例。

- a. 单击“创建桶”。
- b. 进入“创建桶”页面，选择“区域”，输入“桶名称”。其他参数保持默认值或根据需要选择。

📖 说明

创建OBS桶时，需要选择与DLI管理控制台相同的区域，不可跨区域执行操作。

- c. 单击“立即创建”。
4. 单击所建桶“dli-test-obs01”，进入“对象”页面。
5. 选择左侧列表中的“对象”，选择“上传对象”，将需要上传的文件，例如“flink-examples.jar”上传到指定目录，单击“确定”。

本例文件上传成功后，路径为“obs://dli-test-obs01/flink-examples.jar”。

关于OBS管理控制台更多操作请参考《对象存储服务用户指南》。

步骤 2：购买弹性资源池并添加队列

执行跨源场景的SQL作业，不能使用系统已有的default队列，需要您创建SQL队列，本例创建弹性资源池“dli_resource_pool”、队列“dli_queue_01”。

1. 登录DLI管理控制台。
2. 在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。
3. 在弹性资源池管理界面，单击界面右上角的“购买弹性资源池”。
4. 在“购买弹性资源池”界面，填写具体的弹性资源池参数。

本例在华东-上海二区域购买按需计费的弹性资源池。相关参数说明如[表4-2](#)所示。

表 4-2 参数说明

参数名称	参数说明	配置样例
区域	选择弹性资源池所在区域。	华东-上海二
项目	每个区域默认对应一个项目，由系统预置。	系统默认项目
名称	弹性资源池名称。	dli_resource_pool
规格	选择弹性资源池规格。	标准版
CU范围	弹性资源池最大最小CU范围。	64-64
网段	规划弹性资源池所属的网段。如需使用DLI增强型跨源，弹性资源池网段与数据源网段不能重合。 弹性资源池网段设置后不支持更改。	172.16.0.0/19
企业项目	选择对应的企业项目。	default

5. 参数填写完成后，单击“立即购买”，在界面上确认当前配置是否正确。
6. 单击“提交”完成弹性资源池的创建。
7. 在弹性资源池的列表页，选择要操作的弹性资源池，单击操作列的“添加队列”。

8. 配置队列的基础配置，具体参数信息如下。

表 4-3 弹性资源池添加队列基础配置

参数名称	参数说明	配置样例
名称	弹性资源池添加的队列名称。	dli_queue_01
类型	选择创建的队列类型。 <ul style="list-style-type: none"> 执行SQL作业请选择SQL队列。 执行Flink或Spark作业请选择通用队列。 	-
执行引擎	SQL队列可以选择队列引擎为Spark或者HetuEngine。	-
企业项目	选择对应的企业项目。	default

9. 单击“下一步”，配置队列的扩缩容策略。
 单击“新增”，可以添加不同优先级、时间段、“最小CU”和“最大CU”扩缩容策略。
 本例配置的扩缩容策略如图4-2所示。

图 4-2 添加队列时配置扩缩容策略



表 4-4 扩缩容策略参数说明

参数名称	参数说明	配置样例
优先级	当前弹性资源池中的优先级数字越大表示优先级越高。本例设置一条扩缩容策略，默认优先级为1。	1
时间段	首条扩缩容策略是默认策略，不能删除和修改时间段配置。 即设置00-24点的扩缩容策略。	00-24
最小CU	设置扩缩容策略支持的最小CU数。	16
最大CU	当前扩缩容策略支持的最大CU数。	64

10. 单击“确定”完成添加队列配置。

步骤 3: 使用 DEW 管理访问凭据

跨源分析场景中，需要在connector中设置账号、密码等属性。但是账号密码等信息属于高度敏感数据，需要做加密处理，以保障用户的数据隐私安全。

数据加密服务（Data Encryption Workshop，DEW）安全、可靠、简单易用隐私数据加解密方案。

本例介绍在DEW创建通用凭证的操作步骤。

1. 登录DEW管理控制台
2. 选择“凭据管理”，进入“凭据管理”页面。
3. 单击“创建凭据”，配置凭据基本信息。
 - 本例中第一行key对应的value为用户的AK（Access Key Id）
 - 本例中第二行key对应的value为用户的SK（Secret Access Key）

图 4-3 在 DEW 中配置访问凭据



4. 后续在DLI Flink jar作业编辑界面设置访问凭据的参数示例。


```
flink.hadoop.fs.obs.bucket.USER_BUCKET_NAME.dew.access.key=USER_AK_CSMS_KEY_obstest1
flink.hadoop.fs.obs.bucket.USER_BUCKET_NAME.dew.secret.key=USER_SK_CSMS_KEY_obstest1
flink.hadoop.fs.obs.security.provider=com.dli.provider.UserObsBasicCredentialProvider
flink.hadoop.fs.dew.csms.secretName=obsAksKflink.hadoop.fs.dew.endpoint=kmsendpoint
flink.hadoop.fs.dew.csms.version=v6flink.hadoop.fs.dew.csms.cache.time.second=3600flink.dli.job.agency.name=agencyname
```

步骤 4: 创建自定义委托允许 DLI 访问 DEW 读取凭证

1. 登录管理控制台。
2. 单击右上方登录的用户名，在下拉列表中选择“统一身份认证”。
3. 在左侧导航栏中，单击“委托”。
4. 在“委托”页面，单击“创建委托”。
5. 在“创建委托”页面，设置如下参数：
 - 委托名称：按需填写，例如“dli_dew_agency_access”。
 - 委托类型：选择“云服务”。
 - 云服务：（“委托类型”选择“云服务”时出现此参数项。）在下拉列表中选择“DLI”。
 - 持续时间：选择“永久”。
 - 描述：非必选，可以填写“拥有OBS OperateAccess权限的委托”。
6. 配置完委托的基本信息后，单击“下一步”。
7. 授予当前委托所需的权限策略，单击“新建策略”。

8. 配置策略信息。
 - a. 输入策略名称，本例：dli-dew-agency。
 - b. 选择“JSON视图”。
 - c. 在策略内容中粘贴自定义策略。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "csms:secretVersion:get",
        "csms:secretVersion:list",
        "kms:dek:decrypt"
      ]
    }
  ]
}
```
 - d. 按需输入策略描述。
9. 新建策略完成后，单击“下一步”，返回委托授权页面。
10. 选择步骤8新建的自定义策略。
11. 单击“下一步”，选择委托的授权范围。本例选择“所有资源”。
了解更多授权操作说明请参考[创建用户组并授权](#)。
12. 单击“确定”，完成授权。
授权后需等待15-30分钟才可生效。

步骤 5：创建 Flink Jar 作业并配置作业信息

步骤1 创建Flink Jar作业。

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作业列表页面。
2. 单击“创建作业”，弹出“创建作业”对话框。
本例选择Flink Jar作业类型，作业名称为Flink_Jar_for_test。

图 4-4 新建 Flink Jar 作业

The screenshot shows a 'Create Job' (创建作业) form. At the top right is a close button (X). The form has the following sections:

- 类型 (Type):** A dropdown menu with 'Flink Jar' selected.
- * 名称 (Name):** A required text input field with the placeholder '请输入名称'.
- 描述 (Description):** A text input field with the placeholder '请输入描述'.
- 标签 (Tags):** A section with explanatory text: '如果您需要使用同一标签识别多种云资源, 即所有服务均可在标签输入框下拉选择同一标签, 建议在TMS中创建预定义标签。查看预定义标签 C'. Below this is a large text area with the instruction '在下方键/值输入框输入内容后单击添加, 即可将标签加入此处'. At the bottom of this section are three input fields: '请输入标签键', '请输入标签值', and a '添加' (Add) button. Below these fields is the text '您还可以添加20个标签。'

At the bottom of the form are two buttons: a red '确定' (Confirm) button and a white '取消' (Cancel) button.

3. 单击“确定”进入作业配置页面。

步骤2 配置作业的基本信息。

参考表4-5配置作业的基本信息。

表 4-5 参数说明

名称	是否必选	描述
所属队列	是	选择作业运行时使用的队列资源。
应用程序	是	选择 步骤1: 开发Jar包并上传数据至OBS 中自定义的程序包。
主类	是	<p>指定加载的Jar包类名。</p> <p>“主类”参数用于指定Flink作业的入口, 即包含main方法的类。这是Flink作业启动时首先执行的类。</p> <p>当应用程序类型为.jar时, 必须提供主类名称。</p> <p>主类名称区分大小写, 必须准确无误。</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 当类属于某个包时, 需携带包路径, 例如: packagePath.KafkaMessageStreaming</p>
Flink版本	是	<p>选择作业运行时所使用的Flink的版本。</p> <p>本例选择使用Flink1.15版本时请在作业中配置允许DLI访问的云服务的委托信息。</p>

名称	是否必选	描述
委托	否	选择Flink1.15版本时需要用户自行配置委托，否则可能影响作业运行。

步骤3 Flink Jar作业的高级配置。

参考表4-6配置Flink Jar作业。

表 4-6 Flink Jar 作业的高级配置

名称	是否必选	描述
CU数量	是	一个CU为1核4G的资源量。CU数量范围为2~400个。
管理单元	是	设置管理单元的CU数，支持设置1~4个CU，默认值为1个CU。
并行数	是	作业中每个算子的最大并行数。 说明 <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	否	用于设置TaskManager资源参数。 勾选后需配置下列参数： <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	否	设置是否将作业运行时的日志信息保存到OBS。 勾选后需配置下列参数： “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”。
作业异常告警	否	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。 勾选后需配置下列参数： “SMN主题”： 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。

名称	是否必选	描述
异常自动重启	否	<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”：从最新保存的checkpoint恢复作业。勾选该参数后，还需要选择“Checkpoint路径”。 <ul style="list-style-type: none"> “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。


步骤4 单击右上角“保存”，保存作业和相关参数。

步骤5 单击右上角“启动”，进入“启动Flink作业”页面。

步骤6 在“启动Flink作业”页面确认作业规格和费用，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，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。

----结束

5 使用 DLI 提交 Spark Jar 作业

操作场景

DLI允许用户提交编译为Jar包的Spark作业，Jar包中包含了Jar作业执行所需的代码和依赖信息，用于在数据查询、数据分析、机器学习等特定的数据处理任务中使用。在提交Spark Jar作业前，将程序包上传至OBS，并将程序包与数据和作业参数一起提交以运行作业。

本例介绍通过DLI控制台提交Jar作业程序包的基本流程。由于不同的业务需求，Jar包的具体编写会有所差异。建议您参考DLI提供的示例代码，并根据实际业务场景进行相应的编辑和定制。获取[DLI样例代码](#)。

操作流程

使用DLI提交Spark Jar作业的操作流程如表5-1所示。

开始进行如下操作前，请务必参考[准备工作](#)完成必要操作。

表 5-1 使用 DLI 提交 Spark Jar 作业的操作流程

操作步骤	说明
步骤1: 上传数据至OBS	准备Spark Jar作业程序包并上传至OBS。
步骤2: 创建弹性资源池并添加队列	创建提交Spark Jar作业所需的计算资源。
步骤3: 使用DEW管理访问凭证	跨源分析场景中，使用DEW管理数据源的访问凭证，并创建允许DLI访问DEW的委托。
步骤4: 创建自定义委托允许DLI访问DEW读取凭证	创建允许DLI访问DEW的委托。
步骤5: 提交Spark作业	创建Spark Jar作业分析数据。

准备工作

- 已注册华为账号并开通华为云，且在使用DLI前检查账号状态，账号不能处于欠费或冻结状态。

- 配置DLI委托访问授权

DLI使用过程中涉及到OBS、VPC、SMN等服务交互，首次使用DLI需要用户配置委托授权，允许访问这些依赖服务。

 - a. 使用华为云账号登录DLI管理控制台，在左侧导航栏单击“全局配置 > 服务授权”。
 - b. 在委托设置页面，勾选基础使用、跨源场景、运维场景的委托权限后，单击“更新委托权限”。
 - c. 查看并了解更新委托的提示信息，单击“确定”。完成DLI委托权限的更新。

图 5-1 配置 DLI 委托访问授权



- d. 完成配置后，在IAM控制台的委托列表中，可查看到dli_management_agency的委托信息。

步骤 1：上传数据至 OBS

参考[Spark作业样例代码](#)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. 登录管理控制台
2. 在服务列表中，单击“存储”中的“对象存储服务OBS”，进入OBS管理控制台页面。
3. 创建桶，这里以桶名“dli-test-obs01”为例。
 - a. 单击“创建桶”。
 - b. 进入“创建桶”页面，选择“区域”，输入“桶名称”。其他参数保持默认值或根据需要选择。

📖 说明

创建OBS桶时，需要选择与DLI管理控制台相同的区域，不可跨区域执行操作。

- c. 单击“立即创建”。
4. 单击所建桶“dli-test-obs01”，进入“对象”页面。
5. 选择左侧列表中的“对象”，选择“上传对象”，将需要上传的文件，例如“spark-examples.jar”上传到指定目录，单击“确定”。

本例文件上传成功后，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《对象存储服务用户指南》。

步骤 2：创建弹性资源池并添加队列

本例创建弹性资源池“dli_resource_pool”、队列“dli_queue_01”。

1. 登录DLI管理控制台。
2. 在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。
3. 在弹性资源池管理界面，单击界面右上角的“购买弹性资源池”。
4. 在“购买弹性资源池”界面，填写具体的弹性资源池参数。

本例在华东-上海二区域购买按需计费的弹性资源池。相关参数说明如表5-2所示。

表 5-2 参数说明

参数名称	参数说明	配置样例
区域	选择弹性资源池所在区域。	华东-上海二
项目	每个区域默认对应一个项目，由系统预置。	系统默认项目
名称	弹性资源池名称。	dli_resource_pool
规格	选择弹性资源池规格。	标准版
CU范围	弹性资源池最大最小CU范围。	64-64
网段	规划弹性资源池所属的网段。如需使用DLI增强型跨源，弹性资源池网段与数据源网段不能重合。 弹性资源池网段设置后不支持更改。	172.16.0.0/19
企业项目	选择对应的企业项目。	default

5. 参数填写完成后，单击“立即购买”，在界面上确认当前配置是否正确。
6. 单击“提交”完成弹性资源池的创建。
7. 在弹性资源池的列表页，选择要操作的弹性资源池，单击操作列的“添加队列”。
8. 配置队列的基础配置，具体参数信息如下。

表 5-3 弹性资源池添加队列基础配置

参数名称	参数说明	配置样例
名称	弹性资源池添加的队列名称。	dli_queue_01

参数名称	参数说明	配置样例
类型	选择创建的队列类型。 <ul style="list-style-type: none"> • 执行SQL作业请选择SQL队列。 • 执行Flink或Spark作业请选择通用队列。 	-
执行引擎	SQL队列可以选择队列引擎为Spark或者HetuEngine。	-
企业项目	选择对应的企业项目。	default

9. 单击“下一步”，配置队列的扩缩容策略。
 单击“新增”，可以添加不同优先级、时间段、“最小CU”和“最大CU”扩缩容策略。
 本例配置的扩缩容策略如图5-2所示。

图 5-2 添加队列时配置扩缩容策略



表 5-4 扩缩容策略参数说明

参数名称	参数说明	配置样例
优先级	当前弹性资源池中的优先级数字越大表示优先级越高。本例设置一条扩缩容策略，默认优先级为1。	1
时间段	首条扩缩容策略是默认策略，不能删除和修改时间段配置。 即设置00-24点的扩缩容策略。	00-24
最小CU	设置扩缩容策略支持的最小CU数。	16
最大CU	当前扩缩容策略支持的最大CU数。	64

10. 单击“确定”完成添加队列配置。

步骤 3: 使用 DEW 管理访问凭证

Spark Jar作业的输出数据写入到OBS时，需要配置AKSK访问OBS，为了确保AKSK数据安全，您可以通过数据加密服务（Data Encryption Workshop，DEW）、云凭据管

理服务（Cloud Secret Management Service, CSMS），对AKSK统一管理，有效避免程序硬编码或明文配置等问题导致的敏感信息泄露以及权限失控带来的业务风险。

本节操作介绍在DEW创建通用凭证。

1. 登录DEW管理控制台
2. 选择“凭据管理”，进入“凭据管理”页面。
3. 单击“创建凭据”。配置凭据基本信息。
分别配置AK和SK对应的凭据键值
 - 本例中第一行key对应的value为用户的AK（Access Key Id）
 - 本例中第二行key对应的value为用户的SK（Secret Access Key）

图 5-3 在 DEW 中配置访问凭据



4. 后续在DLI Spark jar作业编辑界面设置访问凭据的参数示例。

```
spark.hadoop.fs.obs.bucket.USER_BUCKET_NAME.dew.access.key= USER_AK_CSMS_KEY_obstest1
spark.hadoop.fs.obs.bucket.USER_BUCKET_NAME.dew.secret.key= USER_SK_CSMS_KEY_obstest1
spark.hadoop.fs.obs.security.provider=com.dli.provider.UserObsBasicCredentialProvider
spark.hadoop.fs.dew.csms.secretName=obsAkSkspark.hadoop.fs.dew.endpoint=kmsendpoint
spark.hadoop.fs.dew.csms.version=v3spark.dli.job.agency.name=agencyname
```

步骤 4：创建自定义委托允许 DLI 访问 DEW 读取凭证

1. 登录管理控制台。
2. 单击右上方登录的用户名，在下拉列表中选择“统一身份认证”。
3. 在左侧导航栏中，单击“委托”。
4. 在“委托”页面，单击“创建委托”。
5. 在“创建委托”页面，设置如下参数：
 - 委托名称：按需填写，例如“dli_dew_agency_access”。
 - 委托类型：选择“云服务”。
 - 云服务：（“委托类型”选择“云服务”时出现此参数项。）在下拉列表中选择“DLI”。
 - 持续时间：选择“永久”。
 - 描述：非必选，可以填写“拥有OBS OperateAccess权限的委托”。
6. 配置完委托的基本信息后，单击“下一步”。
7. 授予当前委托所需的权限策略，单击“新建策略”。
8. 配置策略信息。
 - a. 输入策略名称，本例：dli-dew-agency。
 - b. 选择“JSON视图”。

- c. 在策略内容中粘贴自定义策略。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "csms:secretVersion:get",
        "csms:secretVersion:list",
        "kms:dek:decrypt"
      ]
    }
  ]
}
```

- d. 按需输入策略描述。

9. 新建策略完成后，单击“下一步”，返回委托授权页面。
10. 选择步骤8新建的自定义策略。
11. 单击“下一步”，选择委托的授权范围。本例选择“所有资源”。
了解更多授权操作说明请参考[创建用户组并授权](#)。
12. 单击“确定”，完成授权。
授权后需等待15-30分钟才可生效。

步骤 5：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作业编辑页面配置作业参数信息：
 - 队列：选择[步骤2：创建弹性资源池并添加队列](#)中创建的队列。
 - Spark版本：选择Spark引擎版本。本例选择3.3.1版本。
 - 应用程序：选择[步骤1：上传数据至OBS](#)创建的程序包。
 - 委托：请选择[步骤4：创建自定义委托允许DLI访问DEW读取凭证](#)创建的委托，用于访问DEW中存储的凭证。

其他参数请参考[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。

6 入门实践

我们整理了从队列网络连通、不同类型的作业分析、数据迁移场景的常用开发指南和最佳实践内容，帮助您更好的使用DLI进行大数据分析和处理。

表 6-1 DLI 常用开发指南与最佳实践

场景	操作指导	描述
队列网络连通	配置DLI队列与内网数据源的网络连通	DLI在创建运行作业需要连接外部其他数据源，本节操作介绍队列与外部数据源之间的网络连通方法，如：DLI连接MRS、RDS、CSS、Kafka、DWS时，需要打通DLI和外部数据源之间的网络。
	配置DLI 队列与公网网络连通	介绍DLI队列在公网访问场景下网络连通的方法。通过配置SNAT规则，添加到公网的路由信息，可以实现队列到和公网的网络打通。
Spark SQL 作业开发	使用Spark SQL作业分析OBS数据	介绍使用Spark SQL作业创建OBS表、导入OBS表数据、插入和查询OBS表数据等操作。
Flink OpenSource SQL作业开发	从Kafka读取数据写入到RDS	介绍使用Flink OpenSource SQL作业从Kafka读取数据写入到RDS的开发示例。
	从Kafka读取数据写入到DWS	介绍使用Flink OpenSource SQL作业从Kafka读取数据写入到DWS的开发示例。
	从Kafka读取数据写入到Elasticsearch	介绍使用Flink OpenSource SQL作业从Kafka读取数据写入到Elasticsearch的开发示例。
	从MySQL CDC源表读取数据写入到DWS	介绍使用Flink OpenSource SQL作业从MySQL CDC源表读取数据写入到DWS的开发示例。
	从PostgreSQL CDC源表读取数据写入到DWS	介绍使用Flink OpenSource SQL作业从PostgreSQL CDC源表读取数据写入到DWS的开发示例。

场景	操作指导	描述
Flink Jar作业开发	Flink Jar作业开发基础样例	通过自定义作业与MRS进行交互为例进行介绍Flink Jar作业开发样例。
	使用Flink Jar写入数据到OBS开发指南	介绍将kafka数据处理后写入到OBS的操作样例。
	使用Flink Jar连接开启SASL_SSL认证的Kafka	介绍使用Flink OpenSource SQL连接开启SASL_SSL认证的Kafka。
	使用Flink Jar读写DIS开发指南	介绍Flink Jar作业读写DIS数据的操作方法。
Spark Jar作业开发	使用Spark Jar作业读取和查询OBS数据	介绍从编写Spark程序代码读取和查询OBS数据、编译打包到提交Spark Jar作业的开发样例。
数据迁移	迁移Hive数据至DLI	介绍如何通过CDM数据同步功能，迁移MRS Hive数据至DLI。
	迁移Kafka数据至DLI	介绍如何通过CDM数据同步功能，迁移MRS Kafka数据至DLI。
	迁移Elasticsearch数据至DLI	介绍如何通过CDM数据同步功能，迁移Elasticsearch类型的CSS集群数据至DLI。
	迁移RDS数据至DLI	介绍如何通过CDM数据同步功能，迁移关系型数据库RDS数据至DLI。
	迁移DWS数据至DLI	介绍如何通过CDM数据同步功能，迁移数据仓库服务DWS数据至DLI。