

DLF

用户指南

文档版本

7

发布日期

2020-08-12



版权所有 © 华为技术有限公司 2020。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默示的声明或保证。

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。

目录

1 环境准备	1
2 IAM 权限管理	2
2.1 创建用户并授权使用 DLF	2
3 数据管理	4
3.1 概述	4
3.2 数据连接	4
3.2.1 创建数据连接	4
3.2.2 编辑数据连接	10
3.2.3 删除数据连接	10
3.2.4 导出数据连接	11
3.2.5 导入数据连接	12
3.3 数据库	13
3.3.1 新建数据库	13
3.3.2 编辑数据库	15
3.3.3 删除数据库	15
3.4 命名空间	15
3.4.1 新建命名空间	15
3.4.2 删除命名空间	16
3.5 数据库模式	16
3.5.1 新建数据库模式	16
3.5.2 修改数据库模式	17
3.5.3 删除数据库模式	17
3.6 数据表	18
3.6.1 新建数据表（可视化模式）	18
3.6.2 新建数据表（DDL 模式）	23
3.6.3 查看表详情	24
3.6.4 删除数据表	25
3.7 列	25
4 数据集成	26
4.1 CDM 集群管理	26
4.2 DIS 通道管理	26
4.3 CS 作业管理	26

5 数据开发	27
5.1 脚本开发.....	27
5.1.1 新建脚本.....	27
5.1.2 开发 SQL 脚本.....	29
5.1.3 开发 Shell 脚本.....	32
5.1.4 重命名脚本.....	34
5.1.5 移动脚本.....	36
5.1.6 导出导入脚本.....	38
5.1.7 删除脚本.....	40
5.1.8 复制脚本.....	40
5.2 作业开发.....	41
5.2.1 新建作业.....	41
5.2.2 开发作业.....	43
5.2.3 重命名作业.....	49
5.2.4 移动作业.....	51
5.2.5 导出导入作业.....	52
5.2.6 删除作业.....	55
5.2.7 复制作业.....	56
6 解决方案	57
7 运维调度	59
7.1 运维概览.....	59
7.2 作业监控.....	59
7.2.1 批作业监控.....	59
7.2.2 实时作业监控.....	65
7.2.3 实时子作业监控.....	68
7.3 实例监控.....	70
7.4 补数据监控.....	71
7.5 通知管理.....	72
7.5.1 通知管理.....	72
7.5.2 通知周期概览.....	74
7.6 备份和还原资产.....	76
8 配置管理	79
8.1 主机连接管理.....	79
8.2 资源管理.....	81
9 配置	84
9.1 工作区.....	84
9.2 管理企业项目.....	85
9.3 环境变量.....	86
9.4 配置作业日志存储路径.....	88
9.5 配置委托.....	88

10 使用教程	96
10.1 开发一个 Spark 作业	96
10.2 开发一个 Hive SQL 脚本	98
11 参考	103
11.1 节点	103
11.1.1 节点概述	103
11.1.2 CDM Job	104
11.1.3 DIS Stream	107
11.1.4 DIS Dump	108
11.1.5 DIS Client	109
11.1.6 Rest Client	111
11.1.7 Import GES	116
11.1.8 MRS Kafka	117
11.1.9 Kafka Client	118
11.1.10 CS Job	119
11.1.11 DLI SQL	123
11.1.12 DLI Spark	126
11.1.13 DWS SQL	128
11.1.14 MRS SparkSQL	132
11.1.15 MRS Hive SQL	134
11.1.16 MRS Presto SQL	136
11.1.17 MRS Spark	137
11.1.18 MRS Spark Python	139
11.1.19 MRS Flink Job	140
11.1.20 MRS MapReduce	142
11.1.21 CSS	143
11.1.22 Shell	145
11.1.23 RDS SQL	146
11.1.24 ETL Job	148
11.1.25 OCR	151
11.1.26 Create OBS	153
11.1.27 Delete OBS	154
11.1.28 OBS Manager	155
11.1.29 Open/Close Resource	156
11.1.30 Data Quality Monitor	157
11.1.31 Sub Job	158
11.1.32 SMN	159
11.1.33 Dummy	162
11.1.34 For Each	163
11.2 表达式	164
11.2.1 表达式概述	164
11.2.2 基础操作符	165

11.2.3 日期和时间模式.....	166
11.2.4 Env 内嵌对象.....	167
11.2.5 Job 内嵌对象.....	167
11.2.6 StringUtil 内嵌对象.....	169
11.2.7 DateUtil 内嵌对象.....	169
11.2.8 JSONUtil 内嵌对象.....	170
11.2.9 Loop 内嵌对象.....	171
11.2.10 表达式使用示例.....	171
A 修订记录.....	174

1 环境准备

请执行以下步骤访问数据开发模块控制台：

步骤1 访问华为云管理控制台。

步骤2 单击管理控制台左上角的📍，选择区域和项目。

步骤3 在首页“所有服务”中，选择“EI企业智能 > 数据湖工厂”，进入数据开发模块服务的总览页面。

----结束

2 IAM 权限管理

2.1 创建用户并授权使用 DLF

如果您需要对您所拥有的DLF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），通过IAM，您可以：

- 根据企业的业务组织，在您的华为云账号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用DLF资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将DLF资源委托给更专业、高效的其他华为云账号或者云服务，这些账号或者云服务可以根据权限进行代运维。

如果华为云账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DLF服务的其它功能。

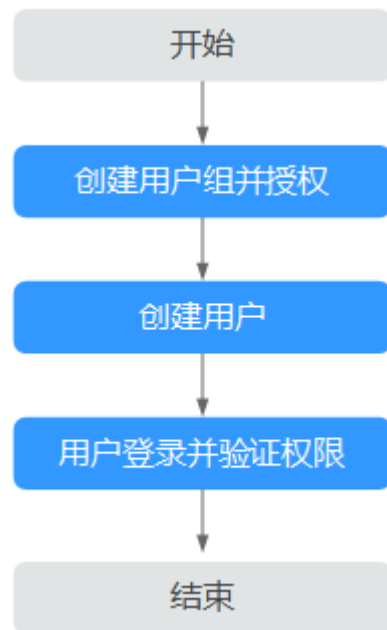
本章节为您介绍对用户授权的方法，操作流程如[图2-1](#)图1-1所示。

前提条件

给用户组授权之前，请您了解用户组可以添加的DLF权限，并结合实际需求进行选择，DLF支持的系统权限，请参见[DLF系统权限](#)。若您需要对除DLF之外的其它服务授权，IAM支持服务的所有策略请参见[权限策略](#)。

示例流程

图 2-1 给用户授权 DLF 权限流程



1. **创建用户组并授权**

在IAM控制台创建用户组，并授予数据湖工厂服务管理权限“DLF OperationAndMaintenanceAccess”。

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

新创建的用户登录控制台，验证数据湖工厂服务的管理权限。

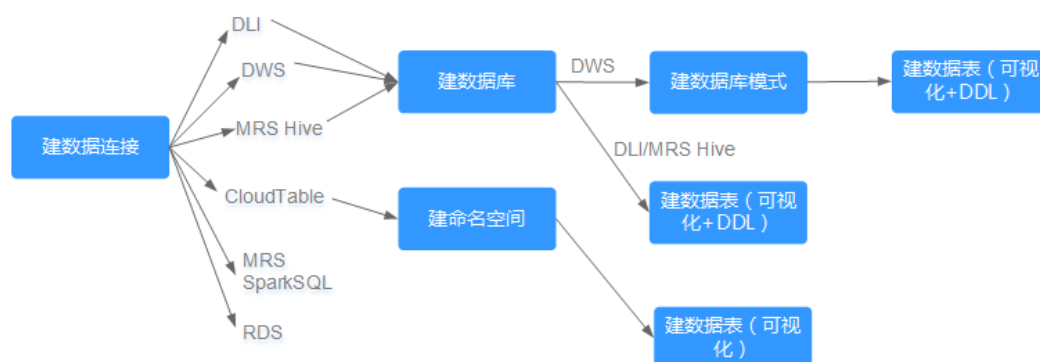
- a. 在“服务列表”中选择数据湖工厂服务，进入DLF主界面，单击右上角“购买数据湖工厂”，若未提示权限不足，表示“DLF OperationAndMaintenanceAccess”已生效。
- b. 在“服务列表”中选择除数据湖工厂服务外的任一服务，若提示权限不足，表示“DLF OperationAndMaintenanceAccess”已生效。

3 数据管理

3.1 概述

数据管理功能可以协助用户快速建立数据模型，为后续脚本和作业开发提供数据实体。数据管理的使用流程如下：

图 3-1 数据管理



1. 创建数据连接，连接华为云的相关服务。
2. 通过与相应服务建立连接，在数据开发模块中对相应服务进行更多数据操作，例如：管理数据库、管理命名空间、管理数据库模式、管理数据表。

3.2 数据连接

3.2.1 创建数据连接

数据连接对应于数据开发模块管理的数据实体，它保存了数据实体的连接信息。同一个数据连接下，可支持多个作业运行和多个脚本开发，当数据连接保存的信息发生变化时，您只需在连接管理中编辑修改该数据连接的信息。

目前支持创建如下类型的数据连接。

- DLI

- DWS
- MRS Hive
- MRS SparkSQL
- RDS

前提条件

- 已开通或创建相应的云服务。
例如，创建RDS类型的数据连接前，您需要先在RDS中创建数据库实例。
- 数据连接的最大配额为20，当前数据连接的数量未达到最大配额。

操作步骤

步骤1 创建数据连接的入口有两个，分别是“连接管理”页面和右侧区域。

- 通过“连接管理”页面
 - a. 在数据开发模块控制台的左侧导航栏，选择“数据管理 > 连接管理”。
 - b. 在页面的右上方，单击“新建数据连接”。
- 通过右侧区域
 - a. 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
 - b. 在右侧区域创建数据连接的方式有以下三种：
方式一：单击“新建数据连接”。

图 3-2 新建数据连接（方式一）




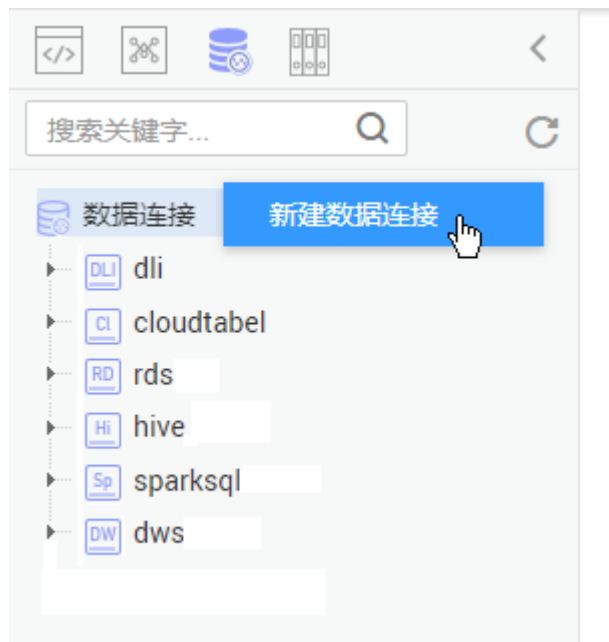
方式二：在左侧菜单选择 ，右键单击根目录“数据连接”，选择“新建数据连接”。

图 3-3 新建数据连接（方式二）




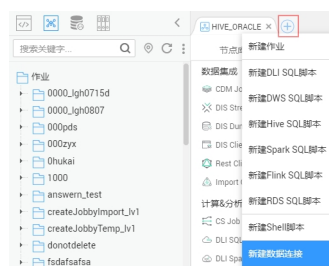
方式三：打开详细的脚本或者作业，单击 ，选择“新建数据连接”。

图 3-4 新建数据连接（方式三）



步骤2 在弹出的对话框中，选择“数据连接类型”，并参见表3-1配置相关参数。

表 3-1 数据连接

数据连接类型	参数说明	其他说明
DLI	请见表3-2	DLI类型的数据连接只能创建一个。
DWS	请见表3-3	-
MRS Hive	请见表3-4	-
MRS SparkSQL	请见表3-5	-
RDS	请见表3-6	-

步骤3 单击“测试”，测试数据连接的连通性。如果无法连通，数据连接将无法创建。

步骤4 单击“确定”，创建数据连接。

----结束

参数说明

表 3-2 DLI 数据连接

参数	是否必选	说明
数据连接名称	是	数据连接的名称，只能包含英文字母、数字、“_”，且长度为1~100个字符。

表 3-3 DWS 数据连接

参数	是否必选	说明
数据连接名称	是	数据连接的名称，只能包含英文字母、数字、“_”，且长度为1~100个字符。
集群名	否	选择DWS集群。如果不直接选择DWS集群，用户需要配置“访问地址”和“端口”。
访问地址	是/否	DWS集群的访问地址。 <ul style="list-style-type: none">当用户在“集群名”中选择了DWS集群时，系统会自动填写为对应DWS集群的访问地址。当不选择DWS集群时，用户需要手动填写为DWS集群的访问地址。
端口	是/否	DWS集群的访问端口。 <ul style="list-style-type: none">当用户在“集群名”中选择了DWS集群时，系统会自动填写为对应DWS集群的访问端口。当不选择DWS集群时，用户需要手动填写为DWS集群的访问端口。
用户名	是	DWS集群的管理员用户。
密码	是	DWS集群的管理员密码。
SSL连接	是/否	DWS支持SSL认证方式的连接，以加密DWS客户端与数据库之间传输的数据。SSL连接方式的安全性高于普通模式，从安全性考虑，建议用户开启SSL连接。
KMS密钥	是	通过KMS加密和解密用户密码和密钥对的密码，选择KMS中已创建的密钥。

参数	是否必选	说明
连接代理	是	DWS为非全托管服务，数据开发模块无法直接与非全托管服务进行连接。CDM集群提供了数据开发模块与非全托管服务通信的代理，所以创建DWS的数据连接时，请选择一个CDM集群。如果没有可用的CDM集群，请先创建一个CDM集群。

表 3-4 MRS Hive 数据连接

参数	是否必选	说明
数据连接名称	是	数据连接的名称，只能包含英文字母、数字、“_”，且长度为1~100个字符。
集群名	是	选择Hive所属的MRS集群。
连接方式	是	<p>选择DLF连接MRS的方式。</p> <p>通过代理连接</p> <p>使用CDM集群的通信代理功能，连接DLF和MRS，推荐使用该方式。</p> <p>选择该方式的用户还需配置如下参数：</p> <ul style="list-style-type: none"> ● 用户名（可选）：MRS Manager管理员用户。部分MRS集群不需要配置用户名，以具体集群为准。 ● 密码（可选）：MRS Manager管理员用户的密码。部分MRS集群不需要配置用户密码，以具体集群为准。 ● KMS密钥（可选）：通过KMS加密和解密用户密码和密钥对的密码，选择KMS中已创建的密钥。 ● 连接代理（必选）：选择可用的CDM集群。 <p>直接连接</p> <p>选择该方式的用户，将无法查看Hive数据表及字段内容；在线开发Hive SQL脚本时，执行结果只能在日志中查看。</p>

表 3-5 MRS SparkSQL 数据连接

参数	是否必选	说明
数据连接名称	是	数据连接的名称，只能包含英文字母、数字、“_”，且长度为1~100个字符。
集群名	是	选择SparkSQL所属的MRS集群。

参数	是否必选	说明
连接方式	是	<p>选择DLF连接MRS的方式。</p> <p>通过代理连接</p> <p>使用CDM集群的通信代理功能，连接DLF和MRS，推荐使用该方式。</p> <p>选择该方式的用户还需配置如下参数：</p> <ul style="list-style-type: none">● 用户名（可选）：MRS Manager管理员用户。部分MRS集群不需要配置用户名，以具体集群为准。● 密码（可选）：MRS Manager管理员用户的密码。部分MRS集群不需要配置用户密码，以具体集群为准。● KMS密钥（可选）：通过KMS加密和解密用户密码和密钥对的密码，选择KMS中已创建的密钥。● 连接代理（必选）：选择可用的CDM集群。 <p>直接连接</p> <p>选择该方式的用户，将无法查看Hive数据表及字段内容；在线开发SparkSQL脚本时，执行结果只能在日志中查看。</p>

表 3-6 RDS 数据连接

参数	是否必选	说明
数据连接名称	是	数据连接的名称，只能包含英文字母、数字、“_”，且长度为1~100个字符。
IP	是	RDS实例的连接地址。
端口	是	RDS实例的端口。
驱动程序名称	是	选择数据库驱动程序的名称： <ul style="list-style-type: none">● com.mysql.jdbc.Driver● org.postgresql.Driver
用户名	是	RDS实例的用户名，默认为root。
密码	是	RDS实例的密码。
KMS密钥	是	通过KMS加密和解密用户密码和密钥对的密码，选择KMS中已创建的密钥。


参数	是否必选	说明
驱动路径	是	JDBC驱动程序的路径。 请根据实际需求提前在MySQL和PostgreSQL官网下载JDBC驱动程序，并上传到OBS桶中。 <ul style="list-style-type: none">“驱动程序名称”选择的是“com.mysql.jdbc.Driver”时，请使用“mysql-connector-java-5.1.21.jar”驱动程序。“驱动程序名称”选择的是“org.postgresql.Driver”时，请使用“postgresql-42.2.2.jar”驱动程序。
连接代理	是	RDS为非全托管服务，数据开发模块无法直接与非全托管服务进行连接。CDM集群提供了数据开发模块与非全托管服务通信的代理，所以创建RDS的数据连接时，请选择一个CDM集群。如果没有可用的CDM集群，请先创建一个CDM集群。

3.2.2 编辑数据连接

数据连接创建完成后，用户可以根据需求修改数据连接的参数。

操作步骤

步骤1 编辑数据连接的入口有两个，分别是“连接管理”页面和右侧区域。

- 通过“连接管理”页面
 - 在数据开发模块控制台的左侧导航栏，选择“数据管理 > 连接管理”。
 - 在数据连接的“操作”列，单击“编辑”。
- 通过右侧区域
 - 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
 - 在左侧菜单选择，右键单击数据连接，选择“编辑”。

步骤2 在弹出的“编辑数据连接”页面，参考[参数说明](#)修改数据连接的参数。

步骤3 单击“测试”，测试数据连接的连通性。如果无法连通，数据连接将无法保存。

步骤4 单击“确定”，保存修改。

----结束

3.2.3 删除数据连接


当用户不需要使用某个数据连接时，可以参考如下操作删除该数据连接。

须知

如果某一个待删除的数据连接正在被脚本或作业关联，请确保强制删除数据连接后，不影响业务使用。如果希望脚本或作业能继续正常使用，请前往脚本或作业开发页面，重新关联可用的数据连接。

操作步骤

步骤1 删除数据连接的入口有两个，分别是“连接管理”页面和右侧区域。

- 通过“连接管理”页面
 - a. 在数据开发模块控制台的左侧导航栏，选择“数据管理 > 连接管理”。
 - b. 在数据连接的“操作”列，单击“删除”。
- 通过右侧区域
 - a. 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
 - b. 在左侧菜单选择 ，右键单击数据连接，选择“删除”。

步骤2 在弹出的“删除数据连接”页面，单击“确定”，删除数据连接。

----结束

3.2.4 导出数据连接

数据连接创建完成后，您可以在连接页面全量导出创建的数据连接。

导出连接，将存在的主机连接同步导出。

前提条件

已开通或创建相应的云服务，并[创建数据连接](#)。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。



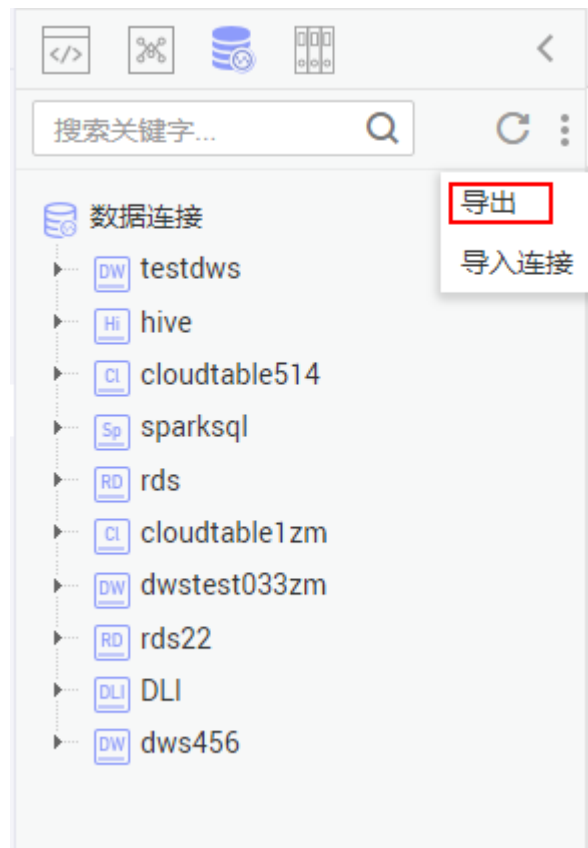
步骤3 在页面中选择 ，右键单击连接目录中的  > 导出。

图 3-5 导出数据连接



----结束

3.2.5 导入数据连接

通过导入数据连接功能，您可以快速的从本地导入数据到连接库中，增加新的数据连接。

前提条件

- 已获取连接数据源所需要的用户名、密码。
- 已开通对象存储服务，并在OBS中创建文件夹。
- 已将本地的数据连接上传至OBS文件夹。
- 数据连接的最大配额为20，当前数据连接的数量未达到最大配额。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。



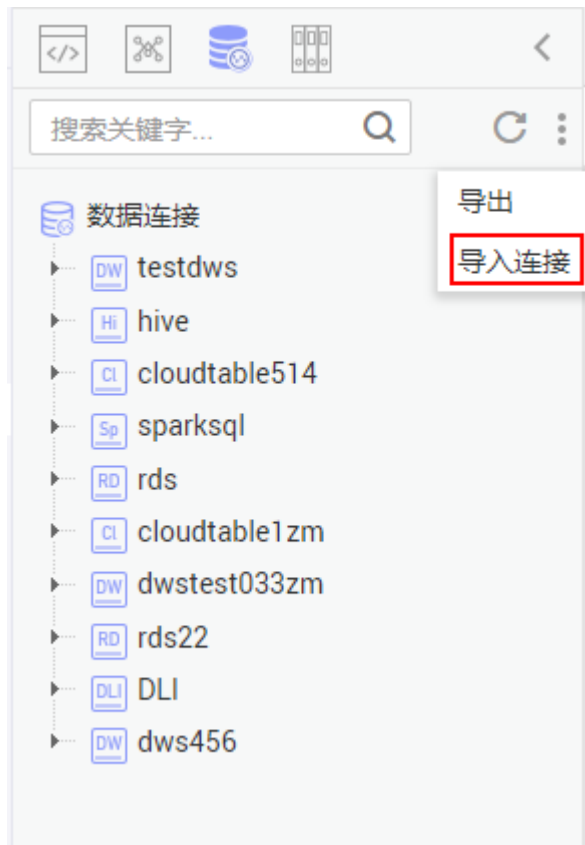
步骤3 在页面中选择 ，右键单击连接目录中的  > 导入连接。

图 3-6 导入数据连接



步骤4 在“导入连接”页面，选择已上传至OBS文件夹中的连接文件，以及重名处理策略。

图 3-7 导入连接



步骤5 单击“下一步”，根据提示导入数据连接。各数据连接的参数解释请参见[参数说明](#)。

----结束

3.3 数据库

3.3.1 新建数据库

数据连接创建完成后，用户可以在右侧区域中管理数据连接下的数据库。

目前支持新建如下数据库：

- DLI
- DWS
- MRS Hive


前提条件

已新建数据连接，请参见[创建数据连接](#)。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。

步骤3 在左侧菜单选择，右键单击数据连接名称，选择“新建数据库”，配置如[表3-7](#)所示的参数。

说明

DLI数据连接最多只能包含10个数据库，其他数据连接不限数据库的数量。

表 3-7 新建数据库

参数	是否必选	说明
数据库名称	是	数据库的名称，命名要求如下： <ul style="list-style-type: none">• DLI：只能包含英文字母、数字、“_”，只能以数字和字母开头，不能全部为数字，且长度为1~128个字符。• DWS：只能包含英文字母、数字、“_”、“\$”，只能以字母和下划线开头，不能全部为数字，且长度为1~63个字符。• MRS Hive：只能包含英文字母、数字、“_”，只能以数字和字母开头，不能全部为数字，且长度为1~128个字符。
描述	否	数据库的描述信息，填写要求如下： <ul style="list-style-type: none">• DLI：最大长度为256个字符。• DWS：最大长度为1024个字符。• MRS Hive：最大长度为1024个字符。

步骤4 单击“确定”，新建数据库。

----结束


3.3.2 编辑数据库

数据库创建完成后，用户可以根据需求修改DWS、MRS Hive数据库的描述信息。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。

步骤3 在左侧菜单选择，右键单击数据库名称，选择“修改”。

步骤4 在弹出的“修改数据库”页面，修改数据库的描述信息。

步骤5 单击“确定”，保存修改。

----结束

3.3.3 删除数据库

当用户不需要使用某个数据库时，可以参考如下操作删除该数据库。


前提条件

删除数据库时，请确保该数据库未被使用，且没有关联数据表。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。

步骤3 在左侧菜单选择，右键单击数据库名称，选择“删除”。

步骤4 在弹出“删除数据库”页面，单击“确定”，删除数据库。

----结束

3.4 命名空间

3.4.1 新建命名空间

CloudTable数据连接创建完成后，用户可以在右侧区域中管理CloudTable数据连接的命名空间。

前提条件

已新建CloudTable数据连接，请参见[创建数据连接](#)。

操作步骤


- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
- 步骤3** 在左侧菜单选择，右键单击CloudTable数据连接名称，选择“新建命名空间”，配置如表3-8所示的参数。

表 3-8 新建命名空间

参数	是否必选	说明
命名空间名称	是	命名空间的名称，只能包含英文字母、数字、“_”，且长度为1~200个字符。
描述	否	命名空间的描述信息，最大长度为1024个字符。

- 步骤4** 单击“确定”，新建命名空间。

----结束


3.4.2 删除命名空间

当用户不需要使用某个命名空间时，可以参考如下操作删除该命名空间。

前提条件

删除命名空间时，请确保该命名空间未被使用，且没有关联数据表。

操作步骤

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
- 步骤3** 在左侧菜单选择，右键单击命名空间名称，选择“删除”。
- 步骤4** 在弹出“删除命名空间”页面，单击“确定”，删除命名空间。

----结束

3.5 数据库模式

3.5.1 新建数据库模式

DWS数据连接创建完成后，用户可以在右侧区域中管理DWS数据连接的数据库模式。

前提条件

- 已新建DWS数据连接，请参见[创建数据连接](#)。
- 已新建DWS数据库，请参见[新建数据库](#)。

操作步骤


- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
- 步骤3** 在左侧菜单选择，单击DWS数据连接名称，选择需配置的数据库，右键单击数据库模式“schemas”，选择“新建模式”。
- 步骤4** 在弹出的“新建模式”页面，配置如[表3-9](#)所示的参数。

表 3-9 新建模式

参数	是否必选	说明
模式名称	是	数据库模式的名称。
描述	否	数据库模式的描述信息。


- 步骤5** 单击“确定”，新建数据库模式。

----结束

3.5.2 修改数据库模式

数据库模式创建完成后，用户可以根据需求修改数据库模式的描述信息。

操作步骤

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。
- 步骤3** 在左侧菜单选择，右键单击数据库模式名称，选择“修改”。
- 步骤4** 在弹出的“修改模式”页面，修改数据库模式的描述信息。
- 步骤5** 单击“确定”，保存修改。

----结束

3.5.3 删除数据库模式

当用户不需要使用某个数据库模式时，可以参考如下操作删除该数据库模式。


前提条件

默认的数据库模式不可删除。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”。

步骤3 在左侧菜单选择，右键单击数据库模式名称，选择“删除”。

步骤4 在弹出的“删除模式”页面，单击“确定”，删除数据库模式。

----结束

3.6 数据表

3.6.1 新建数据表（可视化模式）

用户可以通过可视化模式新建永久数据表。数据表新建完成后，可以提供给脚本和作业使用。

目前支持新建如下数据表：


- DLI
- DWS
- MRS Hive
- CloudTable

前提条件

- 已开通相应的云服务并在云服务中创建数据库。例如，新建DLI表前，请确保DLI服务已开通，并在DLI服务中创建好数据库。
- 已在数据开发模块中创建与数据表类型匹配的数据连接，请参见[创建数据连接](#)。

操作步骤

步骤1 通过“右侧区域”页面

1. 在DLF控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”，进入“右侧区域”页面。
2. 在左侧菜单选择，右键单击数据表“tables”，选择“新建数据表”。

步骤2 在弹出的对话框中，显示“配置基本属性”页面，选择“数据表连接类型”，并参见[表3-10](#)配置相关参数。

表 3-10 基本属性

数据连接类型	参数说明
DLI	请见表3-12的“基本属性”部分
DWS	请见表3-13的“基本属性”部分
MRS Hive	请见表3-14的“基本属性”部分
CloudTable	请见表3-15的“基本属性”部分

步骤3 单击“下一步”，在“配置表结构”页面配置如表3-11所示的参数。

表 3-11 表结构

数据连接类型	参数说明
DLI	请见表3-12的“表结构”部分
DWS	请见表3-13的“表结构”部分
MRS Hive	请见表3-14的“表结构”部分
CloudTable	请见表3-15的“表结构”部分

步骤4 单击“确定”，新建数据表。

----结束

参数说明

表 3-12 DLI 数据表

参数	是否必选	说明
基本属性		
表名	是	数据表的名称。只能包含英文小写字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
别名	否	数据表的别名，只能包含中文字符、英文字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
数据连接	是	选择数据表所属的数据连接。
数据库	是	选择数据表所属的数据库。
数据位置	是	选择数据存储的位置： <ul style="list-style-type: none">● OBS● DLI


参数	是否必选	说明
数据格式	是	选择数据的格式。“数据位置”为“OBS”时，配置该参数。 <ul style="list-style-type: none"> parquet: 支持读取不压缩、snappy压缩、gzip压缩的parquet数据。 csv: 支持读取不压缩、gzip压缩的csv数据。 orc: 支持读取不压缩、snappy压缩的orc数据。 json: 支持读取不压缩、gzip压缩的json数据。
路径	是	选择数据存储的OBS路径。“数据位置”为“OBS”时，配置该参数。
表描述	否	数据表的描述信息。
表结构		
列名	是	填写列名，列名不能重复。
类型	是	选择数据类型，具体说明请参见《数据湖探索SQL语法参考》。
列描述	否	填写列的描述信息。
操作	否	单击  ，增加列。

表 3-13 DWS 数据表

参数	是否必选	说明
基本属性		
表名	是	数据表的名称。只能包含英文字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
别名	否	数据表的别名，只能包含中文字符、英文字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
数据连接	是	选择数据表所属的数据连接。
数据库	是	选择数据表所属的数据库。
模式	是	选择数据库的模式。
表描述	否	数据表的描述信息。

参数	是否必选	说明
高级选项	否	<p>提供以下高级选项：</p> <ul style="list-style-type: none"> ● 选择数据表的存储方式 <ul style="list-style-type: none"> - 行存模式 - 列存模式 ● 选择数据表的压缩级别 <ul style="list-style-type: none"> - 行存模式：压缩级别的有效值为 YES/NO。 - 列存模式：压缩级别的有效值为 YES/NO/LOW/MIDDLE/HIGH，还可以配置列存模式同一压缩级别下不同的压缩水平0-3（数值越大，表示同一压缩级别下压缩比越大）。
表结构		
列名	是	填写列名，列名不能重复。
数据分类	是	<p>选择数据类型的类别：</p> <ul style="list-style-type: none"> ● 数值类型 ● 货币类型 ● 布尔类型 ● 二进制类型 ● 字符类型 ● 时间类型 ● 几何类型 ● 网络地址类型 ● 位串类型 ● 文本搜索类型 ● UUID类型 ● JSON类型 ● 对象标识符类型
类型	是	选择数据类型，具体说明请参见《数据仓库服务开发指南》。
列描述	否	填写列的描述信息。
是否建ES索引	否	单击复选框时，表示需要建立ES索引。建立ES索引时，请同时在“CloudSearch集群名”中选择建立好的CSS集群。如何创建CSS集群，请参见《云搜索服务用户指南》

参数	是否必选	说明
ES索引数据类型	否	选择ES索引的数据类型： <ul style="list-style-type: none"> • text • keyword • date • long • integer • short • byte • double • boolean • binary
操作	否	单击  ，增加列。

表 3-14 MRS Hive 数据表

参数	是否必选	说明
基本属性		
表名	是	数据表的名称。只能包含英文小写字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
别名	否	数据表的别名，只能包含中文字符、英文字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
数据连接	是	选择数据表所属的数据连接。
数据库	是	选择数据表所属的数据库。
表描述	否	数据表的描述信息。
表结构		
列名	是	填写列名，列名不能重复。
数据分类	是	选择数据类型的类别： <ul style="list-style-type: none"> • 原始类型 • ARRAY • MAP • STRUCT • UNION

参数	是否必选	说明
类型	是	选择数据类型，具体说明请参见 LanguageManual DDL 。
列描述	否	填写列的描述信息。
操作	否	单击  ，增加列。

表 3-15 CloudTable 数据表

参数	是否必选	说明
基本属性		
表名	是	数据表的名称。只能包含英文字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
别名	否	数据表的别名，只能包含中文字符、英文字母、数字、“_”，不能为纯数字，不能以“_”开头，且长度为1~63个字符。
数据连接	是	选择数据表所属的数据连接。
命名空间	是	选择数据表所属的命名空间。
表描述	否	数据表的描述信息。
表结构		
列族名	是	填写列族名，列族名不能重复。
列族描述	否	填写列族的描述信息。
操作	否	单击  ，增加列。

3.6.2 新建数据表（DDL 模式）

用户可以通过DDL模式新建永久数据表和临时数据表。数据表新建完成后，可以提供给脚本和作业使用。

目前支持新建如下数据表：

- DLI
- DWS
- MRS Hive


前提条件

- 已开通相应的云服务并在云服务中创建数据库。例如，新建DLI表前，请确保DLI服务已开通，并在DLI服务中创建好数据库。

- 已在数据开发模块中创建与数据表类型匹配的数据连接，请参见[创建数据连接](#)。

操作步骤

步骤1 通过“右侧区域”页面

- 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”，进入“右侧区域”页面。
- 在左侧菜单选择，右键单击数据表“tables”，选择“新建数据表”。

步骤2 单击“DDL模式建表”，选择如[表3-16](#)所示的参数，并在下方的编辑器中输入SQL语句。

表 3-16 数据表参数

参数	说明
数据连接类型	选择数据表所属的数据连接类型。 <ul style="list-style-type: none">DLIDWSHIVE
数据连接	选择数据表所属的数据连接。
数据库	选择数据表所属的数据库。

步骤3 单击“确定”，新建数据表。


---结束

3.6.3 查看表详情

数据表创建完成后，用户可以查看数据表的基本信息、存储信息、字段信息，以及预览数据等。

操作步骤

步骤1 通过“右侧区域”页面

- 在DLF控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”，进入“右侧区域”页面。
- 在左侧菜单选择，右键单击数据表的名称，选择“查看表详情”。

步骤2 进入数据表详情页面，查看如[表3-17](#)所示的数据表信息。

表 3-17 表详情页面

页签名称	说明
表信息	显示数据表的基本信息和存储信息。

页签名称	说明
字段信息	显示数据表的字段信息。
数据预览	预览数据表的10条记录。
DDL	显示DLI/DWS数据表的DDL。


----结束

3.6.4 删除数据表

当用户不需要使用某个数据表时，可以参考如下操作删除该数据表。

操作步骤

步骤1 通过“右侧区域”页面

1. 在DLF控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”，进入“右侧区域”页面。
2. 在左侧菜单选择 ，右键单击数据表的名称，选择“删除”。

步骤2 在弹出的“删除数据表”页面，单击“确定”，删除数据表。

----结束


3.7 列

用户可以在“右侧区域”页面中查看数据表的列信息。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在DLF控制台的左侧导航栏，选择“数据开发 > 脚本开发” / “数据开发 > 作业开发”，进入“右侧区域”页面。

步骤3 在左侧菜单选择 ，展开数据连接目录，在数据表下查看对应的列信息。

----结束

4 数据集成

4.1 CDM 集群管理

为了协助用户快速迁移数据，数据开发模块集成了CDM集群管理功能。用户可通过“数据开发数据开发模块”控制台选择区的下拉选项，选择“数据集成”，进入数据集成管理界面。在左边的导航栏中选择CDM集群管理，进入CDM集群管理页面进行操作。同时，用户也可以直接访问CDM控制台进行相应操作。

CDM的使用方法请参见《云数据迁移用户指南》。

4.2 DIS 通道管理

为了方便用户进行数据接入，数据开发模块集成了DIS通道管理功能。用户可通过“数据开发数据开发模块”控制台选择区的下拉选项，选择“数据集成”，进入数据集成管理界面。在左边的导航栏中选择DIS通道管理，进入DIS通道管理页面进行操作。同时，用户也可以直接访问DIS控制台进行相应操作。

DIS的使用方法请参见《数据接入服务用户指南》。

4.3 CS 作业管理

为了协助用户快速进行流式大数据分析，数据开发模块集成了CS作业管理功能。用户可通过“数据开发数据开发模块”控制台选择区的下拉选项，选择“数据集成”，进入数据集成管理界面。在左边的导航栏中选择CS作业管理，进入CS作业管理页面。同时，用户也可以直接访问CS控制台进行相应操作。

CS的使用方法请参见《实时流计算服务用户指南》。

5 数据开发

5.1 脚本开发

5.1.1 新建脚本

数据开发模块的脚本开发功能支持在线编辑、调试和执行脚本，开发脚本前请先新建脚本。

新建目录（可选）

如果已存在可用的目录，可以不用新建目录。

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。
- 步骤3** 在脚本目录中，右键单击目录名称，选择“新建目录”。
- 步骤4** 在弹出的“新建目录”页面，配置如表5-1所示的参数。

表 5-1 脚本目录参数

参数	说明
目录名称	脚本目录的名称，只能包含英文字母、数字、中文字符、“_”、“-”，且长度为1~32个字符。
选择目录	选择该脚本目录的父级目录，父级目录默认为根目录。

- 步骤5** 单击“确定”，新建目录。

----结束

新建脚本

DLF目前支持新建以下几种脚本，用户可根据需要新建相应的脚本。

- DLI SQL脚本
- Hive SQL脚本
- DWS SQL脚本
- Spark SQL脚本
- Flink SQL脚本
- RDS SQL脚本
- PRESTO SQL脚本，仅亚太-新加坡区域支持。通过PRESTO SQL脚本执行select 查询语句后的查询内容将自动转储至OBS桶的s3a://dlf-log-{project_id}/temp路径下。
- Shell脚本

前提条件

脚本的最大配额是1000，请确保当前脚本的数量未达到最大配额。

操作步骤

步骤1 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。

步骤2 新建脚本的方式有如下两种：

方式一：在“右侧区域”，单击“新建SQL脚本” / “新建Shell脚本”。

图 5-1 新建 SQL 脚本（方式一）

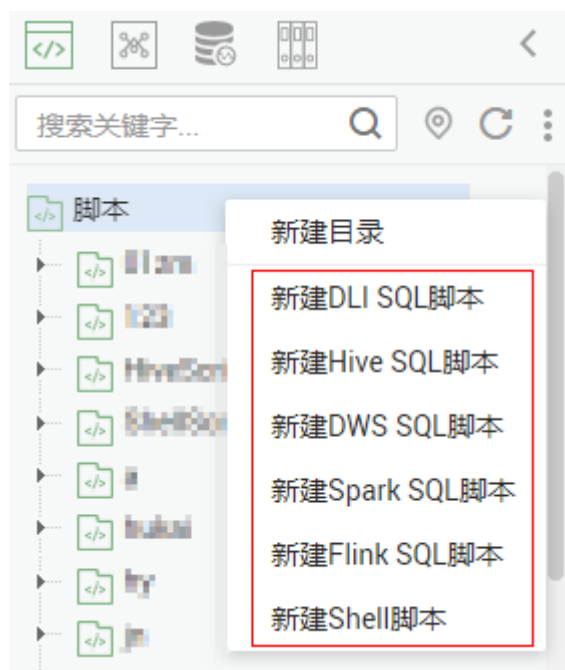


图 5-2 新建 Shell 脚本（方式一）



方式二：在脚本目录中，右键单击目录名称，选择新建相应的脚本。

图 5-3 新建脚本（方式二）



步骤3 进入脚本开发页面，具体操作请参见[开发SQL脚本](#)、[开发Shell脚本](#)。

----结束

5.1.2 开发 SQL 脚本

对SQL脚本进行在线开发、调试和执行，开发完成的脚本也可以在作业中执行（请参见[开发作业](#)）。

前提条件




- 已开通相应的云服务并在云服务中创建数据库。例如，开发DLI的脚本前，请确保DLI服务已开通，并在DLI服务中创建好数据库。Flink SQL脚本不涉及。
- 已在数据开发模块中创建与脚本的数据连接类型匹配的数据连接，请参见[创建数据连接](#)。Flink SQL脚本不涉及。
- 已新增SQL脚本，请参见[新建脚本](#)。

操作步骤

- 步骤1 登录DLF管理控制台。
- 步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。
- 步骤3 在脚本目录中，双击脚本名称，进入脚本开发页面。
- 步骤4 在编辑器上方，选择如[表5-2](#)所示的属性。Flink SQL脚本不涉及，请跳过该步骤。

表 5-2 SQL 脚本属性

属性	说明
数据连接	选择数据连接。

属性	说明
资源队列	<p>选择执行DLI作业的资源队列。当脚本为DLI SQL时，配置该参数。</p> <p>如需新建资源队列，请参考以下方法：</p> <ul style="list-style-type: none"> 单击 ，进入DLI的“队列管理”页面新建资源队列。 前往DLI管理控制台进行新建。 <p>如需以“key/value”的形式设置提交SQL作业的属性，请单击 。最多可设置10个属性，属性说明如下：</p> <ul style="list-style-type: none"> dli.sql.autoBroadcastJoinThreshold（自动使用BroadcastJoin的数据量阈值） dli.sql.shuffle.partitions（指定Shuffle过程中Partition的个数） dli.sql.cbo.enabled（是否打开CBO优化策略） dli.sql.cbo.joinReorder.enabled（开启CBO优化时，是否允许重新调整join的顺序） dli.sql.multiLevelDir.enabled（OBS表的指定目录或OBS表分区表的分区目录下有子目录时，是否查询子目录的内容；默认不查询） dli.sql.dynamicPartitionOverwrite.enabled（在动态分区模式时，只会重写查询中的数据涉及的分区，未涉及的分区不删除）
数据库	选择数据库。
表	显示数据库中已有的数据表，您也可以输入数据表名称，单击  进行查询。

步骤5 在编辑器中输入SQL语句（支持输入多条SQL语句），如需单独执行某部分SQL语句，请选中SQL语句再运行。为了方便脚本开发，数据开发模块提供了系统函数和脚本参数功能（Flink SQL和RDS SQL脚本除外）。

说明

SQL语句之间以“;”分隔。如果其它地方使用“;”，请通过“\”进行转义。例如：

```
select 1;
select * from a where b="dsfa\"; --example 1\;example 2.
```

- **系统函数**

单击编辑器右侧的“系统函数”，显示该数据连接类型支持的函数，您可以双击函数到编辑器中使用。

- **脚本参数**

在SQL语句中直接写入脚本参数，调试脚本时可以在脚本编辑器下方输入参数值。如果脚本被作业引用，在作业开发页面可以配置参数值，参数值支持使用EL表达式（参见[表达式概述](#)）。

脚本示例如下：

```
select ${str1} from data;
```

其中，str1是参数名称，只支持英文字母、数字、-、_、<和>，最大长度为16字符，且参数名称不允许重名。

步骤6 （可选）在编辑器上方，单击“格式化”，格式化SQL语句。Flink SQL脚本不涉及，请跳过该步骤。


步骤7 在编辑器上方，单击“运行”。SQL语句运行完成后，在编辑器下方可以查看脚本的执行历史、执行结果。Flink SQL脚本不涉及，请跳过该步骤。

用户可根据需求下载或转储执行结果，仅支持具有拥有DAYU Administrator或Tenant Administrator权限的用户下载和转储。参考如下：

- 下载结果：下载CSV格式的结果文件到本地。
- 转储结果：转储CSV格式的结果文件到OBS中，请参见[表5-3](#)。

表 5-3 转储结果

参数	是否必选	说明
数据格式	是	目前仅支持导出CSV格式的结果文件。
资源队列	否	选择执行导出操作的DLI队列。当脚本为DLI SQL时，配置该参数。
压缩格式	否	选择压缩格式。当脚本为DLI SQL时，配置该参数。 <ul style="list-style-type: none"> - none - bzip2 - deflate - gzip
存储路径	是	设置结果文件的OBS存储路径。选择OBS路径后，您需要在选择的路径后方自定义一个文件夹名称，系统将在OBS路径下创建文件夹，用于存放结果文件。
覆盖类型	是	如果“存储路径”中，您自定义的文件夹在OBS路径中已存在，选择覆盖类型。当脚本为DLI SQL时，配置该参数。 <ul style="list-style-type: none"> - 覆盖：删除OBS路径中已有的重名文件夹，重新创建自定义的文件夹。 - 存在即报错：系统返回错误信息，退出导出操作。

步骤8 在编辑器上方，单击，保存脚本。

如果脚本是新建且未保存过的，请配置如[表5-4](#)所示的参数。

表 5-4 保存脚本

参数	是否必选	说明
脚本名称	是	脚本的名称，只能包含五种字符：英文字母、数字、中文、中划线、下划线和点号，且长度小于等于128个字符。
描述	否	脚本的描述信息。
选择目录	是	选择脚本所属的目录，默认为根目录。

----结束

5.1.3 开发 Shell 脚本

对Shell脚本进行在线开发、调试和执行，开发完成的脚本也可以在作业中执行（请参见[开发作业](#)）。

前提条件

- 已新增Shell脚本，请参见[新建脚本](#)。
- 已新建主机连接，该主机用于执行Shell脚本，请参见[主机连接管理](#)。

操作步骤

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。
- 步骤3** 在脚本目录中，双击脚本名称，进入脚本开发页面。
- 步骤4** 在编辑器上方，配置如[表5-5](#)所示的属性。

表 5-5 Shell 脚本属性

参数	说明	示例
主机连接	选择执行Shell脚本的主机。	-

参数	说明	示例
参数	填写执行Shell脚本时，向脚本传递的参数，参数之间使用空格分隔，例如：a b c。此处的“参数”需要在Shell脚本中引用，否则配置无效。	例如输入如下shell交互式脚本，交互参数 1 2 3 对应 begin end exit，那么这里需要输入参数1 2 3 <pre>#!/bin/bash select ch in "begin" "end" "exit"; do case \$ch in "begin") echo "start something" ;; "end") echo "stop something" ;; "exit") echo "exit" break; ;; *) echo "Ignorant" ;; esac</pre>
交互式输入	填写交互式参数，即执行Shell脚本的过程中，需要用户输入的交互式信息（例如密码）。交互式参数之间以回车符分隔，Shell脚本根据交互情况按顺序读取参数值。	-

步骤5 在编辑器中编辑Shell语句。

为了方便脚本开发，DLF提供了脚本参数功能，使用方法如下：


在Shell语句中直接写入脚本参数名称和参数值。当Shell脚本被作业引用时，如果作业配置的参数名称与Shell脚本的参数名称相同，Shell脚本的参数值将被作业的参数值替换。

脚本示例如下：

```
a=1
echo ${a}
```

其中，a是参数名称，只支持英文字母、数字、-、_、<和>，最大长度为16字符，且参数名称不允许重名。

步骤6 在编辑器下方，单击“运行”。Shell语句运行完成后，在编辑器下方可以查看脚本的执行历史和执行结果。

步骤7 在编辑器上方，单击，保存脚本。

如果脚本是新建且未保存过的，请配置如表5-6所示的参数。

表 5-6 保存脚本

参数	是否必选	说明
脚本名称	是	脚本的名称，只能包含五种字符：英文字母、数字、中文、中划线、下划线和点号，且长度小于等于128个字符。
描述	否	脚本的描述信息。
选择目录	是	选择脚本所属的目录，默认为根目录。

----结束

5.1.4 重命名脚本

您可以通过重命名功能把脚本文件的当前名称修改为另一个名称。

本章节主要介绍如何重命名一份脚本文件。

前提条件

- 您已开发脚本，脚本目录中存在要重命名的脚本文件。
- 如何开发脚本，请参见[开发SQL脚本](#)和[开发Shell脚本](#)。

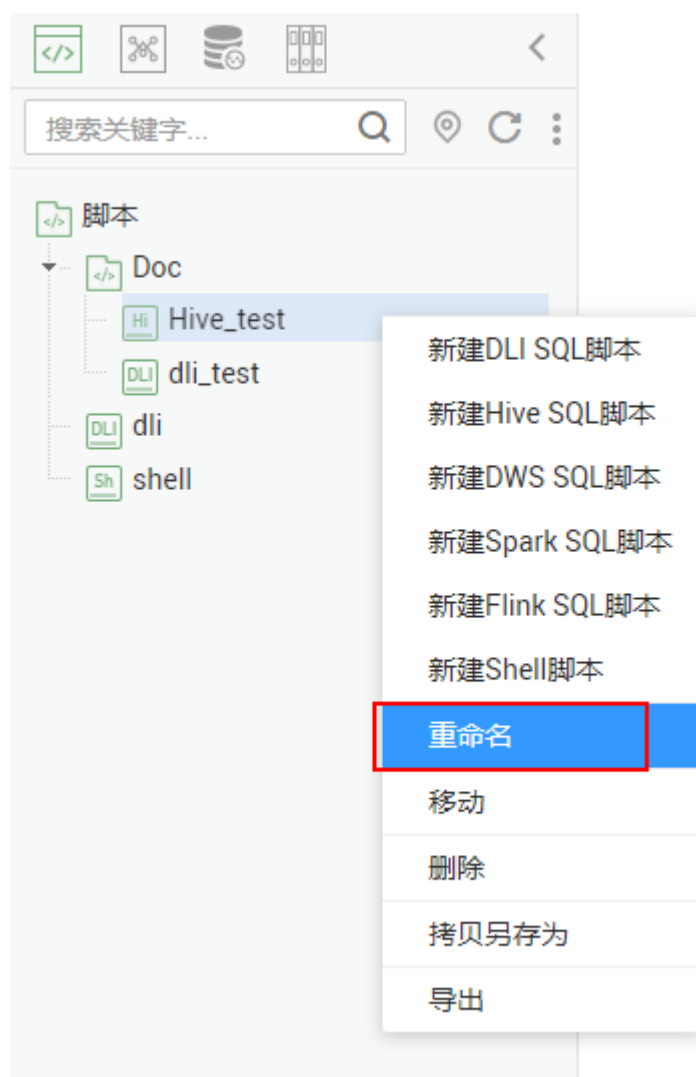
操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。

步骤3 在脚本目录中选择需要重命名的脚本，右键单击脚本名称，选择“重命名”。

图 5-4 重命名



说明

不支持重命名打开了的脚本文件。

步骤4 在弹出的“重命名脚本名称”页面，配置如表5-7所示的参数。

图 5-5 重命名脚本名称

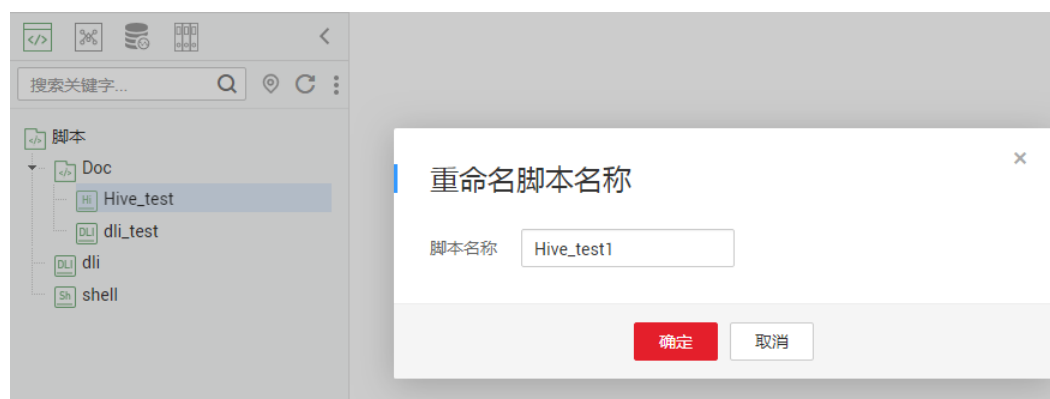


表 5-7 重命名脚本参数

参数	说明
脚本名称	脚本的名称，只能包含五种字符：英文字母、数字、中文、中划线、下划线和点号，且长度小于等于128个字符。

步骤5 单击“确定”，重命名脚本。

----结束

5.1.5 移动脚本

您可以通过移动功能把脚本文件从当前目录移动到另一个目录。

本章节主要介绍如何移动一份脚本文件。

前提条件

- 您已开发脚本，脚本目录中存在要移动的脚本文件。

如何开发脚本，请参见[开发SQL脚本](#)和[开发Shell脚本](#)。

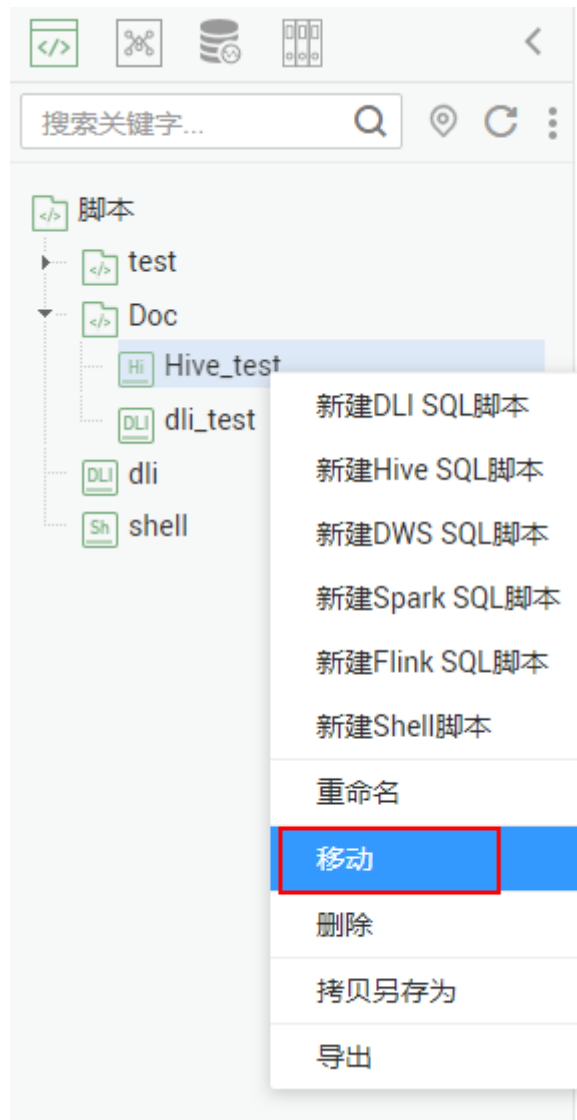
操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。

步骤3 在脚本目录中选择需要移动的脚本，右键单击脚本名称，选择“移动”。

图 5-6 移动



步骤4 在弹出的“移动脚本”页面，配置如表5-8所示的参数。

图 5-7 移动脚本

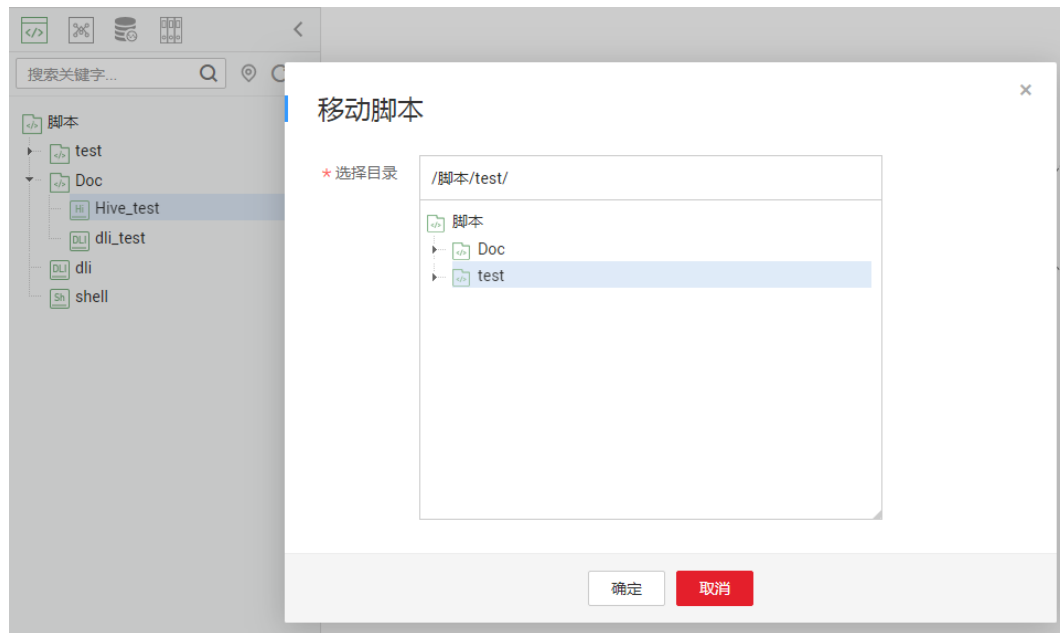


表 5-8 移动脚本参数

参数	说明
选择目录	选择脚本要移动到的目录，父级目录默认为根目录。

步骤5 单击“确定”，移动脚本。

----结束

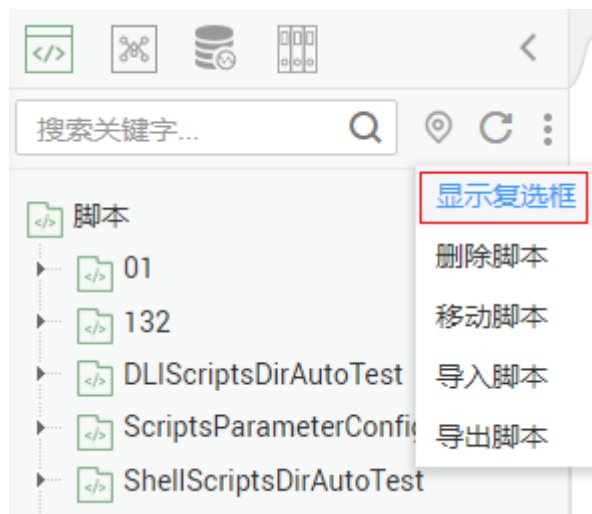
5.1.6 导出导入脚本

导出脚本

您可以在脚本目录中导出一个或多个脚本文件。

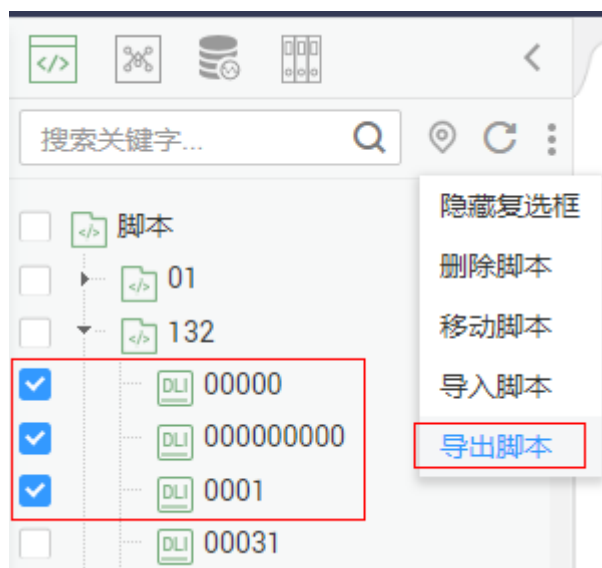
步骤1 单击脚本目录中的 ，选择“显示复选框”。

图 5-8 显示复选框



步骤2 勾选需要导出的脚本，单击  > 导出脚本。

图 5-9 选择并导出脚本



----结束

导入脚本

您可以在脚本目录中导入一个或多个脚本文件。


步骤1 单击作业目录中的  > 导入脚本，选择已上传至OBS中的脚本文件，以及重名处理策略。

图 5-10 导入脚本



步骤2 单击“下一步”，根据提示导入脚本。

----结束

5.1.7 删除脚本

当用户不需要使用某个脚本时，可以参考如下操作删除该脚本。

须知

如果某一个待删除的脚本正在被作业关联，请确保强制删除脚本后，不影响业务使用。如果希望作业能继续正常使用，请前往作业开发页面，重新关联可用的脚本。

普通删除

步骤1 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。


步骤2 在脚本目录中，右键单击脚本名称，选择“删除”。


步骤3 在弹出的“删除脚本”页面，单击“确定”，删除脚本。

----结束

批量删除

步骤1 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。

步骤2 在脚本目录顶部，单击 ，选择“显示复选框”，在脚本目录前出现复选框。

步骤3 选择需要删除的脚本，再次单击 ，选择“删除脚本”。

步骤4 在弹出的“删除脚本”页面，单击“确定”，批量删除脚本。

----结束

5.1.8 复制脚本

本章节主要介绍如何复制一个脚本。

前提条件

脚本目录中已存在要复制的脚本文件。

操作步骤

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在DLF控制台的左侧导航栏，选择“数据开发 > 脚本开发”。
- 步骤3** 在脚本目录中选择需要复制的脚本，右键单击脚本名称，选择“拷贝另存为”。
- 步骤4** 在弹出的“另存为”页面，配置如表5-9所示的参数。

表 5-9 脚本目录参数

参数	说明
脚本名称	脚本的名称，只能包含五种字符：英文字母、数字、中文、中划线、下划线和点号，且长度小于等于128个字符。 说明 复制后的脚本名称不能和原脚本名称相同。
选择目录	选择该脚本目录的父级目录，父级目录默认为根目录。

- 步骤5** 单击“确定”，复制脚本。

----结束

5.2 作业开发

5.2.1 新建作业

作业由一个或多个节点组成，共同执行以完成对数据的一系列操作。开发作业前请先新建作业。

新建目录（可选）

如果已存在可用的目录，可以不用新建目录。

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。
- 步骤3** 在作业目录中，右键单击目录名称，选择“新建目录”。
- 步骤4** 在弹出的“新建目录”页面，配置如表5-10所示的参数。

表 5-10 作业目录参数

参数	说明
目录名称	作业目录的名称，只能包含英文字母、数字、中文字符、“_”、“-”，且长度为1~32个字符。
选择目录	选择该作业目录的父级目录，父级目录默认为根目录。

步骤5 单击“确定”，新建目录。

----结束

新建作业

默认作业的最大配额是10000，请确保当前作业的数量未达到最大配额。

步骤1 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤2 新建作业的方式有如下两种：

方式一：在“右侧区域”页面的右侧，单击“新建作业”。

方式二：在脚本目录中，右键单击目录名称，选择“新建作业”。

步骤3 在弹出的“新建作业”页面，配置如表5-11所示的参数。

表 5-11 作业参数

参数	说明
作业名称	自定义作业的名称，只能包含英文字母、数字、中文、“-”、“_”、“.”，且长度为1~128个字符。
作业类型	选择作业的类型。 <ul style="list-style-type: none">批处理：按调度计划定期处理批量数据，主要用于实时性要求低的场景。实时处理：处理实时的连续数据，主要用于实时性要求高的场景。
创建方式	选择作业的创建方式。 <ul style="list-style-type: none">创建空作业：创建一个空的作业。基于模板创建：使用数据开发模块提供的模板来创建。
选择目录	选择作业所属的目录，默认为根目录。
作业责任人	填写该作业的责任人。
作业优先级	选择作业的优先级，提供高、中、低三个等级。

参数	说明
日志路径	选择作业日志的OBS存储路径。日志默认存储在以dlf-log-{Projectid}命名的桶中。 说明 若您想自定义存储路径，请参见 配置作业日志存储路径 选择您已在OBS服务侧创建的桶。

步骤4 单击“确定”，创建作业。

----结束

5.2.2 开发作业

对已有的作业进行编排。

前提条件

已[新建作业](#)。

编排作业节点

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤3 在作业目录中，双击作业名称，进入作业开发页面。

步骤4 拖动所需的节点至画布，鼠标移动到节点图标上，选中连线图标并拖动，连接到下一个节点上。

说明

每个作业最多可以包含200个节点。

----结束

配置作业基本信息

为作业配置责任人、优先级信息后，用户可根据责任人、优先级来检索相应的作业。操作方法如下：

单击画布的空白处，在右侧显示“作业基本信息”页签，单击此页签，展开配置页面，配置如[表5-12](#)所示的参数。

表 5-12 作业基本信息

参数	说明
作业责任人	自动匹配创建作业时配置的作业责任人，此处支持修改。

参数	说明
执行用户	执行作业的用户。如果输入了执行用户，则作业以执行用户身份执行；如果没有输入执行用户，则以提交作业启动的用户身份执行
作业优先级	自动匹配创建作业时配置的作业优先级，此处支持修改。
实例超时时间	配置作业实例的超时时间，设置为0或不配置时，该配置项不生效。如果您为作业设置了异常通知，当作业实例执行时间超过超时时间，将触发异常通知，发送消息给用户。
自定义字段	配置自定义字段的参数名称和参数值。



配置作业参数

作业参数为全局参数，可用于作业中的任意节点。操作方法如下：

单击画布的空白处，在右侧显示“作业参数配置”页签，单击此页签，展开配置页面，配置如表5-13所示的参数。

表 5-13 作业参数配置

功能	说明
参数	
新增	<p>单击“新增”，在文本框中填写作业参数的名称和参数值。</p> <ul style="list-style-type: none"> 参数名称 只支持英文字母、数字、“-”、“_”，最大长度为64字符，且参数名称不允许重名。 参数值 <ul style="list-style-type: none"> 函数类的参数以“\$”开头，例如： \$getCurrentTime(@@yyyyMMdd@@,0) 字符串类的参数直接填写字符串，例如：str1 字符串和函数结合使用时，字符串需要使用“@@”括起来，并使用“+”连接。例如： @@str1@@+ \$getCurrentTime(@@yyyyMMdd@@,0) 数值类的参数直接填写数值或运算表达式。 <p>参数配置完成后，在作业中的引用格式为：\${参数名称}</p>
修改	在参数名和参数值的文本框中直接修改，修改完成后，请保存。
保存	单击“保存”，保存作业参数的配置。

功能	说明
删除	在参数值文本框后方，单击  ，删除作业参数。
常量	
新增	<p>单击“新增”，在文本框中填写作业常量的名称和参数值。</p> <ul style="list-style-type: none"> 参数名称 只支持英文字母、数字、“-”、“_”，最大长度为64字符，且参数名称不允许重名。 参数值 <ul style="list-style-type: none"> 函数类的参数以“\$”开头，例如： \$getCurrentTime(@@yyyyMMdd@@,0) 字符串类的参数直接填写字符串，例如：str1 字符串和函数结合使用时，字符串需要使用“@@”括起来，并使用“+”连接。例如： @@str1@@+ \$getCurrentTime(@@yyyyMMdd@@,0) 数值类的参数直接填写数值或运算表达式。 <p>参数配置完成后，在作业中的引用格式为：\${参数名称}</p>
修改	在参数名和参数值的文本框中直接修改，修改完成后，请保存。
保存	单击“保存”，保存作业常量的配置。
删除	在参数值文本框后方，单击  ，删除作业常量。

配置作业调度任务

配置批处理作业的作业调度任务，支持单次调度、周期调度、事件驱动调度三种方式。操作方法如下：

单击画布的空白处，在右侧显示“调度配置”页签，单击此页签，展开配置页面，配置如表5-14所示的参数。

表 5-14 作业调度配置

参数	说明
调度方式	<p>选择作业的调度方式：</p> <ul style="list-style-type: none"> 单次调度：手动触发作业单次运行。 周期调度：周期性自动运行作业。 事件驱动调度：根据外部条件触发作业运行。
“周期调度”的参数	

参数	说明
生效时间	调度任务的生效时间段。
调度周期	选择调度任务的执行周期，并配置相关参数： <ul style="list-style-type: none"> 分钟 小时 天 周 月
依赖作业	选择依赖的作业，约束条件如下： <ul style="list-style-type: none"> 短周期的作业不能依赖长周期的作业。 调度周期为天的作业不能依赖调度周期为分钟的作业。 调度周期为周的作业不能依赖其他作业，也不能被其他作业依赖。 调度周期为月的作业只能依赖调度周期为天的作业。
依赖的作业失败后，当前作业处理策略	当依赖的作业执行失败后，选择当前作业的处理策略： <ul style="list-style-type: none"> 挂起 挂起当前作业，挂起的作业会阻塞后续作业的执行。您可以手动将依赖的作业强制成功，解决阻塞问题，具体请参见批作业监控。 继续执行 继续执行当前作业。 终止执行 终止执行当前作业，当前作业的状态为“取消”。
等待依赖作业的上一周期结束，才能运行	当作业依赖其他作业时，选择是否等待依赖作业的上一周期结束后才执行当前作业。
跨周期依赖	选择作业下实例之间的依赖关系。 <ul style="list-style-type: none"> 不依赖上一调度周期 自依赖（等待上一调度周期结束才能继续运行）
“事件驱动调度”的参数	
触发事件类型	选择触发作业运行的事件类型DIS。
DIS通道名称	选择DIS通道，当指定的DIS通道有新消息时，数据开发模块将新消息传递给作业，触发该作业运行。
事件处理并发数	选择作业并行处理的数量，最大并发数为128。
事件检测间隔	配置时间间隔，检测DIS通道是否有新的消息。时间间隔单位可以配置为秒或分钟。

参数	说明
读取策略	选择数据的读取位置： <ul style="list-style-type: none">● 从上次位置读取：首次启动时，从最新的位置读取数据。后续启动时，则从前一次记录的位置读取数据。● 从最新位置读取：每次启动都会从最新的位置读取数据。
失败策略	选择调度失败后的策略： <ul style="list-style-type: none">● 结束调度● 忽略失败，继续调度

配置节点属性

单击画布中的节点，在右侧显示“节点属性”页签，默认展开此配置页面，请参见[节点概述](#)配置具体节点的属性。

配置节点调度任务

配置实时处理作业的节点调度任务，支持单次调度、周期调度、事件驱动调度三种方式。操作方法如下：

单击画布中的节点，在右侧显示“调度配置”页签，单击此页签，展开配置页面，配置如[表5-15](#)所示的参数。

表 5-15 节点调度配置

参数	说明
调度方式	选择作业的调度方式： <ul style="list-style-type: none">● 单次调度：手动触发作业单次运行。● 周期调度：周期性自动运行作业。● 事件驱动调度：根据外部条件触发作业运行。
“周期调度”的参数	
生效时间	调度任务的生效时间段。
调度周期	选择调度任务的执行周期，并配置相关参数： <ul style="list-style-type: none">● 分钟● 小时● 天● 周● 月

参数	说明
跨周期依赖	选择作业下实例之间的依赖关系。 <ul style="list-style-type: none">● 不依赖上一调度周期● 自依赖（等待上一调度周期结束才能继续运行）
“事件驱动调度”的参数	
触发事件类型	选择触发作业运行的事件类型。
DIS通道名称	选择DIS通道，当指定的DIS通道有新消息时，数据开发模块将新消息传递给作业，触发该作业运行。
事件处理并发数	选择作业并行处理的数量，最大并发数为10。
事件检测间隔	配置时间间隔，检测DIS通道或OBS路径下是否有新的消息。时间间隔单位可以配置为秒或分钟。
失败策略	选择节点执行失败后的策略： <ul style="list-style-type: none">● 结束调度● 忽略失败，继续调度

更多节点功能

如需获取更多节点功能，请右键单击画布中的节点图标，根据实际需要选择如表5-16所示的功能。

表 5-16 更多节点功能

功能	说明
配置	进入该节点的“节点属性”页面。
删除	支持删除一个节点或同时删除多个节点。 <ul style="list-style-type: none">● 单节点删除：右键单击画布中的节点图标，选择删除或按快捷键Delete。● 多节点删除：按下键盘中的Ctrl，单击画布中需要删除的节点图标，在当前作业画布空白处单击右键，选择删除或按快捷键Delete。
复制	支持复制一个或多个节点至任意作业中： <ul style="list-style-type: none">● 单节点复制：右键单击画布中的节点图标，选择复制或按快捷键Ctrl+C，在作业画布空白处粘贴节点或按快捷键Ctrl+V，复制后的节点携带原节点的配置信息。● 多节点复制：按下键盘中的Ctrl，单击画布中需要复制的节点图标，在当前作业画布空白处单击右键选择复制或按快捷键Ctrl+C，在目标作业画布空白处粘贴或按快捷键Ctrl+V。复制后的节点携带原节点的配置信息，但不包含节点间的连接关系。

功能	说明
测试运行	测试运行该节点。
上传模拟数据	上传模拟数据至DIS Stream节点。提供以下模拟数据： <ul style="list-style-type: none">• 轨迹分析模拟数据• 锅炉异常监测模拟数据
添加便签	为该节点添加便签，每个节点可以有多个便签。
编辑脚本	跳转到脚本编辑页面，对关联的脚本进行编辑。仅关联了脚本的节点显示该选项。


保存和启动作业

作业编排和配置完成后，请执行以下操作：

批处理作业


步骤1 单击画布上方的测试运行按钮 ，测试作业。


步骤2 测试没问题后，单击画布上方的保存按钮 ，保存作业的配置信息。

步骤3 单击画布上方的执行调度按钮 ，启动作业。

----结束

实时处理作业

步骤1 单击画布上方的保存按钮 ，保存作业的配置信息。

步骤2 单击画布上方的提交按钮 ，提交并启动作业。

说明

DLF支持修改正在运行中的实时处理作业。作业修改完成后单击页面上方的“保存”按钮，在弹出的“保存作业”页面，再单击“确认”即可保存作业。

----结束

5.2.3 重命名作业

您可以通过重命名功能把作业文件的当前名称修改为另一个名称。

本章节主要介绍如何重命名一份作业文件。

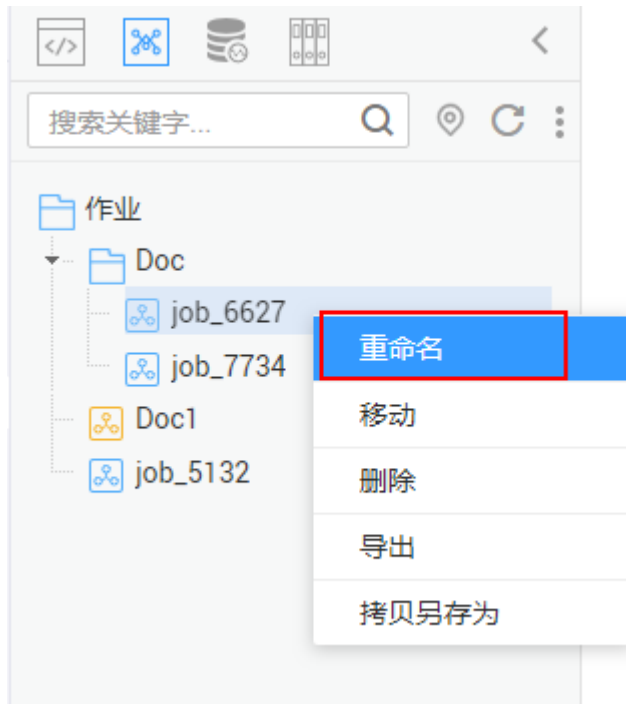
前提条件

您已[新建作业](#)，作业目录中存在要重命名的作业。

操作步骤

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。
- 步骤3** 在作业目录中选择需要重命名的作业，右键单击作业名称，选择“重命名”。

图 5-11 重命名



- 步骤4** 在弹出的“重命名作业名称”页面，配置如表5-17所示的参数。

图 5-12 重命名作业名称

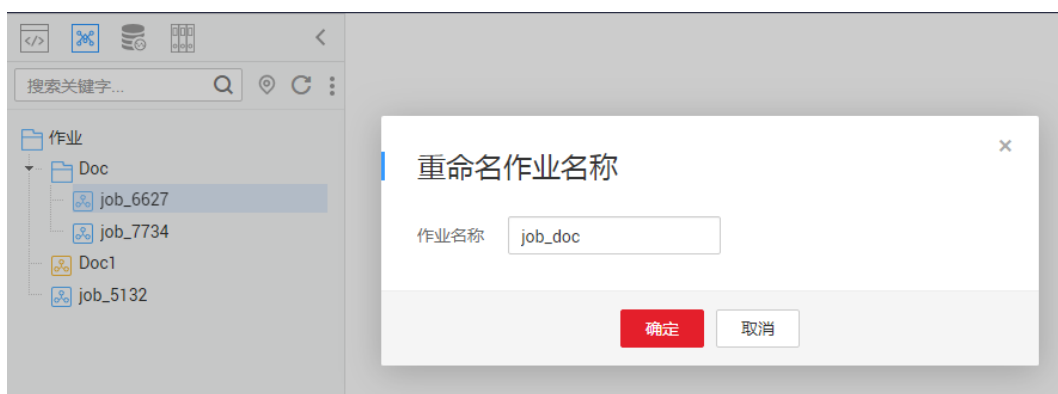


表 5-17 重命名作业参数

参数	说明
作业名称	自定义作业的名称，只能包含英文字母、数字、中文、“-”、“_”、“.”，且长度为1~128个字符。

步骤5 单击“确定”，重命名作业。

----结束

5.2.4 移动作业

您可以通过移动功能把作业文件从当前目录移动到另一个目录。

本章节主要介绍如何移动一份作业文件。

前提条件

您已[新建作业](#)，作业目录中存在要移动的作业。

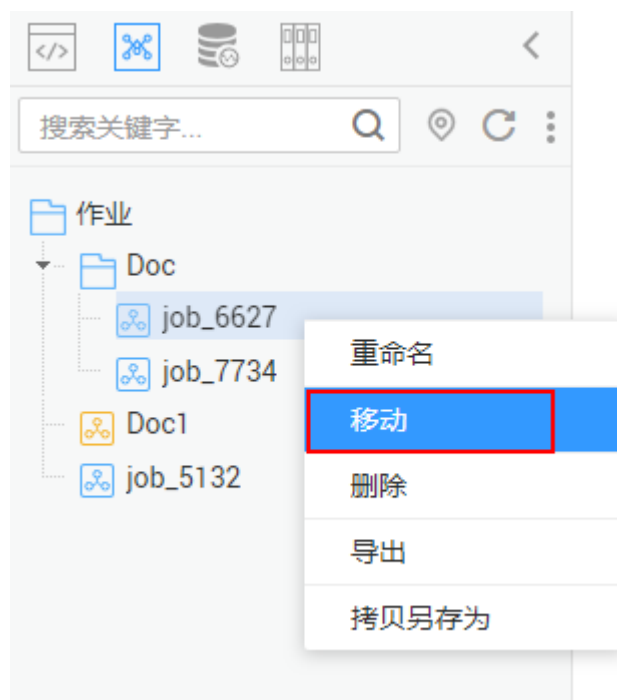
操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤3 在作业目录中选择需要移动的作业，右键单击作业名称，选择“移动”。

图 5-13 移动



步骤4 在弹出的“移动作业”页面，配置如[表5-18](#)所示的参数。

图 5-14 移动作业

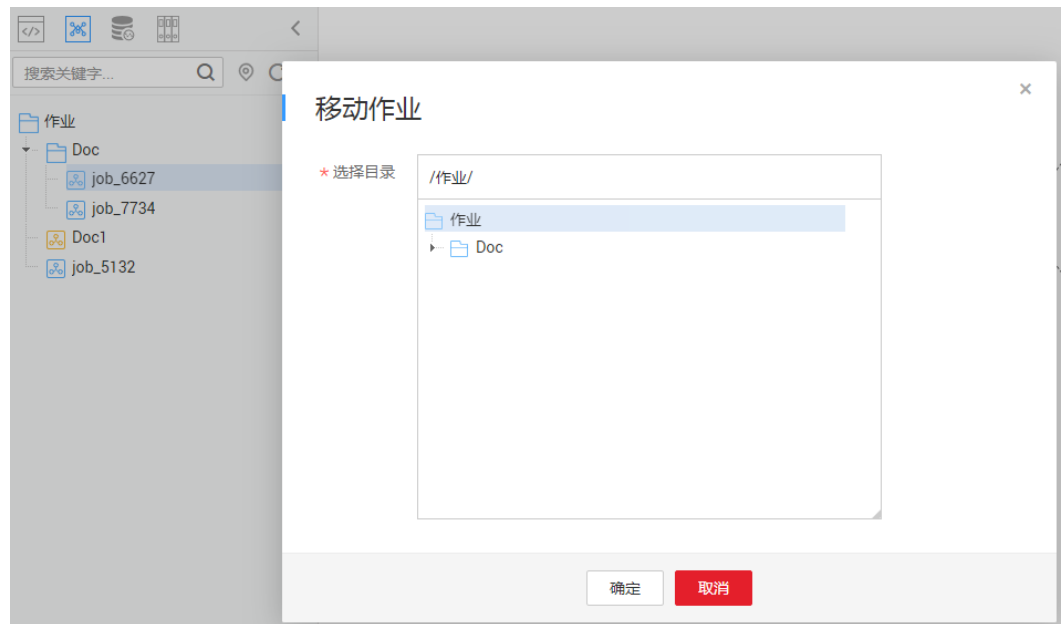


表 5-18 移动作业参数

参数	说明
选择目录	选择作业要移动到的目录，父级目录默认为根目录。


步骤5 单击“确定”，移动作业。

----结束

5.2.5 导出导入作业

导出作业

方式一：在作业开发页面导出某一个作业

步骤1 进入某一作业的开发页面，单击画布上方的导出按钮，选择导出作业的类型。

- 只导出作业：导出作业中节点的连接关系，以及各节点的属性配置，不包含密码等敏感信息。导出后，用户获取到zip格式的压缩包文件。
- 导出作业及其依赖脚本：导出作业中节点的连接关系、各节点的属性配置，以及作业的调度配置、参数配置、依赖的脚本，不包含密码等敏感信息。导出后，用户获取到zip格式的压缩包文件。

图 5-15 导出作业（方式一）



步骤2 单击“确定”，导出所需的作业文件。

----结束

方式二：在作业目录中导出一个或多个作业


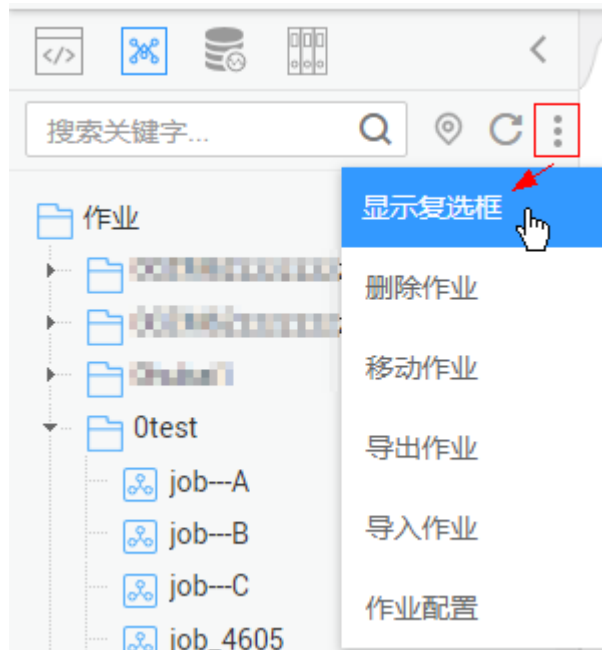
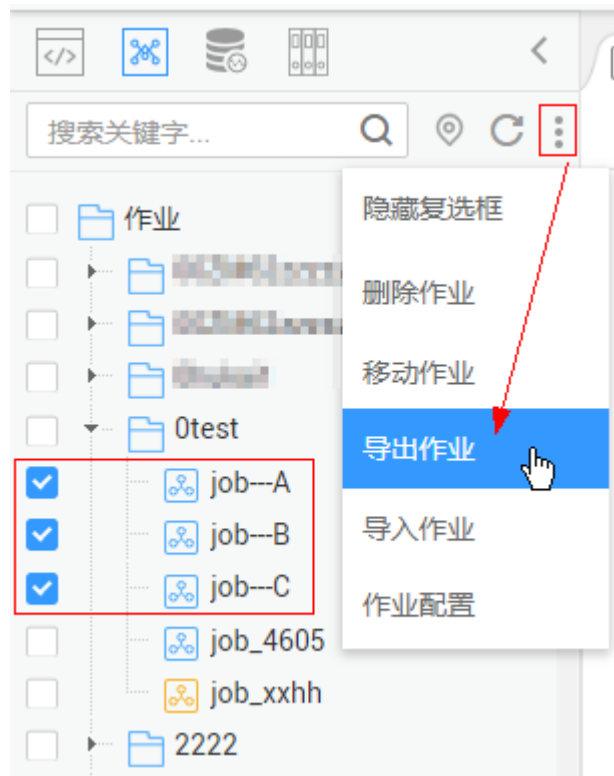
步骤1 单击作业目录中的 ，选择“显示复选框”。

图 5-16 显示复选框



步骤2 勾选需要导出的作业，单击  > 导出作业，导出作业及其依赖。

图 5-17 选择并导出作业



----结束

导入作业

在作业目录中导入一个或多个作业


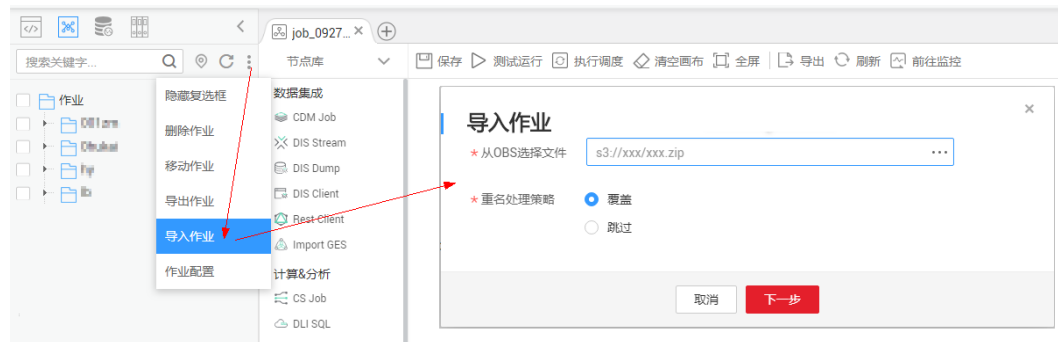
步骤1 单击作业目录中的  > 导入作业，选择已上传至OBS中的作业文件，以及重名处理策略。

图 5-18 导入作业定义及依赖



步骤2 单击“下一步”，根据提示导入作业。

说明

在导入作业过程中，若作业关联的数据连接、dis通道、dli 队列、ges图等在DLF系统中不存在时，系统会提示您重新选择。

----结束

操作示例

背景信息：


- 在DLF系统中创建一个DWS的数据连接“doctest”
- 在作业目录中创建实时作业“doc1”，作业中添加节点“DWS SQL”，配置节点的“数据连接”为“doctest”，配置“SQL脚本”和“数据库”。

步骤1 登录DLF管理控制台。

步骤2 在DLF控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤3 在作业搜索框中搜索作业“doc1”，导出作业到本地，并上传作业至OBS文件夹中。

步骤4 在DLF系统中删掉作业关联的dws数据连接“doctest”。

步骤5 单击作业目录中的  > 导入作业，选择上传至OBS文件夹中的作业，并设置重名处理策略。

步骤6 单击“下一步”，根据导入作业页面的提示重新选择数据连接。

图 5-19 导入作业



步骤7 单击“下一步”，再单击“关闭”。

----结束

5.2.6 删除作业

当用户不需要使用某个作业时，可以参考如下操作删除该作业，以减少作业的配额占用。

说明

作业删除后，将无法恢复，请确保删除作业后，不影响业务。

普通删除

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤3 在作业目录中，右键单击作业名称，选择“删除”。


步骤4 在弹出的“删除作业”页面，单击“确定”，删除作业。


----结束

批量删除

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤3 在作业目录顶部，单击，选择“显示复选框”，在作业目录前出现复选框。

步骤4 选择需要删除的作业，再次单击，选择“删除作业”。

步骤5 在弹出的“删除作业”页面，单击“确定”，批量删除作业。

----结束

5.2.7 复制作业

本章节主要介绍如何复制一份作业。

前提条件

作业目录中已存在要复制的作业文件。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。

步骤3 在作业目录中选择需要复制的作业，右键单击作业名称，选择“拷贝另存为”。

步骤4 在弹出的“另存为”页面，配置如表5-19所示的参数。

表 5-19 作业目录参数

参数	说明
作业名称	自定义作业的名称，只能包含英文字母、数字、中文、“-”、“_”、“.”，且长度为1~128个字符。
选择目录	选择该作业目录的父级目录，父级目录默认为根目录。

步骤5 单击“确定”，复制作业。

----结束

6 解决方案

背景信息

解决方案定位于为用户提供便捷的、系统的方式管理作业，更好地实现业务需求和目标。每个解决方案可以包含一个或多个业务相关的作业，一个作业可以被多个解决方案复用。

DLF目前支持处理以下几种方式的解决方案。

- [新建解决方案](#)
- [编辑解决方案](#)
- [导出解决方案](#)
- [导入解决方案](#)
- [升级解决方案](#)
- [删除解决方案](#)

新建解决方案


在数据开发模块的开发页面，新建一个解决方案，设置解决方案名称并选择业务相关的作业。

图 6-1 新建解决方案



步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”或“数据开发 > 作业开发”。

步骤3 在左侧目录上方，单击解决方案图标，显示解决方案目录。


步骤4 单击解决方案目录上方的，弹出“新建解决方案”页面，配置如表6-1所示的参数。

表 6-1 解决方案参数

参数	说明
名称	自定义解决方案的名称。
选择作业	选择解决方案包含的作业。

步骤5 单击“确定”，新建的解决方案将在左侧目录中显示。

---结束

编辑解决方案

在解决方案目录中，右键单击解决方案名称，选择“编辑”，修改如表6-1所示的参数。

导出解决方案

在解决方案目录中，右键单击解决方案名称，选择“导出”，导出zip格式的解决方案文件至本地。

导入解决方案

在解决方案目录中，右键单击根目录“解决方案”，选择“导入解决方案”，导入已上传到OBS中的解决方案文件。

升级解决方案

在解决方案目录中，右键单击解决方案名称，选择“升级”，导入已上传到OBS中的解决方案文件。升级解决方案时，会停止其中正在运行的作业，系统将依据用户配置的升级重启策略，判断是否在升级完成后重新启动作业。

删除解决方案

在解决方案目录中，右键单击解决方案名称，选择“删除”，删除解决方案。删除的解决方案不可恢复，请谨慎操作。

7 运维调度

7.1 运维概览

在“运维调度 > 运维概览”页面，用户可以通过图表的形式查看作业实例的统计数据，目前支持查看以下四种统计数据。

- 当天作业实例调度情况概览
- 近七天作业实例调度情况概览
- 近三十天执行时间最长的前十个作业实例
单击作业名称，跳转至“实例监控”页面，查看执行时间长的作业实例的详细运行记录。
- 近三十天运行异常次数最多的前十个作业实例
单击“运行异常次数”列的统计次数，跳转至“实例监控”页面，查看运行异常的作业实例的详细运行记录。

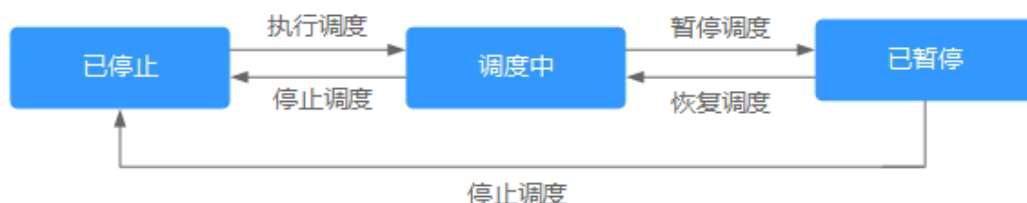
7.2 作业监控

7.2.1 批作业监控

批处理：调度作业

作业开发完成后，用户可以在“作业监控”页面中管理作业的调度任务，例如：执行调度、暂停调度、恢复调度、停止调度。

图 7-1 调度作业



- 步骤1 登录DLF管理控制台。
 - 步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。
 - 步骤3 单击“批作业监控”页签，进入批作业的监控页面。
 - 步骤4 在作业的“操作”列，单击“执行调度”/“暂停调度”/“恢复调度”/“停止”。
- 结束

批处理：调度依赖作业

当批处理作业在“调度配置”页签中配置有依赖作业，而依赖作业没有执行调度，您可以在“作业监控”页面对该批处理作业执行调度时设置是否启动依赖的作业。如何配置依赖作业，请参见[配置作业调度任务](#)。

- 步骤1 登录DLF管理控制台。
- 步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。
- 步骤3 单击“批作业监控”页签，选择配置好有依赖作业的作业。
- 步骤4 在作业的“操作”列，单击“执行调度”。

当依赖的作业没有执行调度，您在执行调度该作业时可以为只启动当前作业或同时启动依赖的作业。

图 7-2 启动作业



----结束

批处理：通知设置

用户可以通过该功能配置某一个作业的通知任务，当作业运行异常或成功时向相关人员发送通知。同时，用户也可以在“运维调度”中配置通知任务，请参见[通知管理](#)。

- 步骤1 登录DLF管理控制台。
 - 步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。
 - 步骤3 单击“批作业监控”页签，进入批作业的监控页面。
 - 步骤4 在作业的“操作”列，选择“更多 > 通知设置”，弹出“新建通知”页面，参考[表7-9](#)配置通知参数。
 - 步骤5 单击“确定”，为作业配置通知任务。
- 结束

批处理：实例监控

用户可以通过该功能查看某一个作业所有实例的运行记录。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 单击“批作业监控”页签，进入批作业的监控页面。

步骤4 在作业的“操作”列，选择“更多 > 实例监控”，查看该作业所有实例的运行记录。

- 实例的“操作”列，请参见[表7-2](#)。
- 节点的“操作”列，请参见[表7-3](#)。

----结束

批处理：调度配置

快速跳转至作业开发页面的“调度配置”。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 单击“批作业监控”页签，进入批作业的监控页面。

步骤4 在作业的“操作”列，选择“更多 > 调度配置”。

----结束

批处理：作业依赖关系视图

支持查看作业与其他作业的依赖关系。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 单击“批作业监控”页签，进入批作业的监控页面。

步骤4 单击作业名称，选择“作业依赖”页签。查看作业间的依赖关系。

图 7-3 作业依赖视图

作业监控 > Doc_batch

作业实例 作业详情 **作业依赖**



单击视图中的作业，将进入该作业的开发页面。

----结束

批处理：补数据

补数据是指作业执行一个调度任务，在过去某一段时间里生成一系列的实例。用户可以通过补数据，修正历史中出现数据错误的作业实例，或者构建更多的作业记录以便调试程序等。

只有配置了周期调度的作业，才支持使用该功能。如需查看补数据的执行情况，请参见[补数据监控](#)。

📖 说明

当作业正在补数据时，请勿修改作业配置，否则会影响补数据过程中生成的作业实例。

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。
- 步骤3** 单击“批作业监控”页签，进入批作业的监控页面。
- 步骤4** 在作业的“操作”列，选择“更多 > 补数据”。
- 步骤5** 弹出“补数据”对话框，配置如所示[表7-1](#)的参数。

图 7-4 补数据参数



表 7-1 参数说明

参数	说明
补数据名称	系统自动生成一个补数据的任务名称，允许修改。
作业名称	显示需要补数据的作业名称。
业务日期	选择需要补数据的时间段。
并行周期数	设置同时执行的实例数量，最多可同时执行5个实例。
需要补数据的下游作业	选择需要补数据的下游作业（指依赖于当前作业的作业），支持多选。

步骤6 单击“确定”，开始补数据，并进入“补数据监控”页面。

----结束

批处理：批量处理作业

DLF支持批量执行调度、停止作业、配置通知任务。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 单击“批作业监控”页签，进入批作业的监控页面。

步骤4 勾选作业名称前的复选框，单击“执行调度”/“停止”/“通知配置”，批量处理相应的作业。

----结束


批处理：查看最近的实例

用户可以通过该功能查看某一个作业正在运行的5条实例信息。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 单击“批作业监控”页签，进入批作业的监控页面。

步骤4 单击作业名称前方的 ，显示“最近的实例”页面，查看最近的实例包含的节点详细信息。

----结束

批处理：查看所有实例

用户可以通过“运行记录”页面查看某一个作业的所有运行记录，以及根据实际需求对实例/节点进行更多操作。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 单击“批作业监控”页签，进入批作业的监控页面。

步骤4 单击作业名称，进入“运行记录”页面。

支持批量停止、重跑、继续执行、强制成功多个实例，使用说明参见[表7-2](#)。

其中，批量重跑多个实例时，重跑的顺序如下：

- 如果作业不依赖上一调度周期，多个实例并行重跑。
- 如果作业自依赖，多个实例串行重跑，以上一调度周期中实例执行完成的先后顺序为准，先执行完成的先重跑。

图 7-5 批量操作




步骤5 在实例的“操作”列，提供如[表7-2](#)所示的操作。

表 7-2 操作（实例）

操作项	说明
停止	停止运行状态为“待运行”、“运行中”或“运行异常”的实例。
重跑	重新运行状态为“成功”或“取消”的实例。
查看等待作业实例	实例的状态为“等待运行”时，支持查看等待的作业实例。

操作项	说明
继续执行	实例的状态为“运行异常”时，支持继续运行实例中的后续节点。 说明 只有节点的“节点属性 > 高级 > 失败策略”设置为“挂起当前作业执行计划”时，才可以执行该操作。
更多 > 强制成功	强制将状态为“运行异常”、“取消”、“失败”的实例变更为“成功”状态。
更多 > 查看	跳转至作业开发页面，查看作业信息。

步骤6 单击实例前方的 ，显示该实例所有节点的运行记录。

步骤7 在节点的“操作”列，提供如表7-3所示的操作。

表 7-3 操作（节点）

操作项	说明
查看日志	查看节点的日志信息。
手工重试	节点的状态为“失败”时，支持重新运行节点。 说明 只有节点的“节点属性 > 高级 > 失败策略”设置为“挂起当前作业执行计划”时，才可以执行该操作。
强制成功	节点的状态为“失败”时，支持将该节点强制变更为“成功”状态。 说明 只有节点的“节点属性 > 高级 > 失败策略”设置为“挂起当前作业执行计划”时，才可以执行该操作。
更多 > 跳过	节点的状态为“待运行”或“已暂停节点”时，支持跳过该节点。
更多 > 暂停	节点的状态为“待运行”时，支持暂停运行该节点，该暂停节点的后续节点将会被阻塞。
更多 > 恢复	节点的状态为“已暂停”时，支持恢复运行该节点。

----结束

7.2.2 实时作业监控

用户可以在“实时作业监控”页面查看实时处理作业的运行状态、开始执行时间、结束执行时间等信息，以及启动和停止实时处理作业。

图 7-6 实时作业监控

批作业监控 **实时作业监控**

启动 停止 全部

<input type="checkbox"/>	名称	运行状态	开始时间	结束时间	创建人	操作
<input type="checkbox"/>	job_3091	已停止	-	-	df_00000000	启动 停止
<input type="checkbox"/>	job_ZXC	启动中	2019/01/04 15:34:29 GMT +08:00	-	df_00000000	启动 停止
<input type="checkbox"/>	job_8244	正常	2019/01/04 11:37:10 GMT +08:00	-	hmgul	启动 停止
<input type="checkbox"/>	job_9027	正常	2019/01/04 10:59:03 GMT +08:00	-	df_00000000	启动 停止
<input type="checkbox"/>	1111111	已停止	-	-	gandhuang	启动 停止

- 批量处理作业：勾选作业名称前的复选框，单击“启动”/“停止”
- 查看作业的节点信息：单击作业名称，在打开的页面中查看该作业的节点连接关系，以及节点的监控信息。
- 查看最近实例状态：单击作业名称前的下拉箭头，均可查看最近的实例状态。

实时作业：节点配置事件驱动调度

当您配置的实时作业中某个节点配置有事件驱动调度时，在打开的页面中右键单击配置有事件驱动调度的节点，选择“调度配置”，可以查看和修改节点的调度信息。

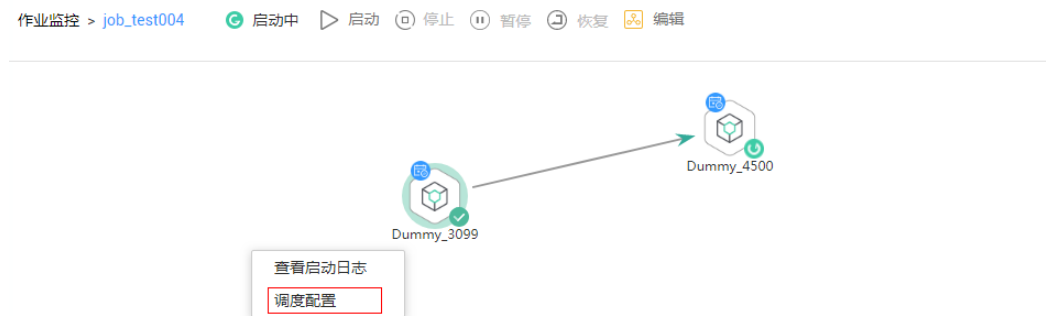
步骤1 登录DLF管理控制台。

步骤2 在DLF控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 选择“实时作业监控”页签，单击作业名称。

步骤4 单击配置有事件驱动调度的节点，选择“调度配置”。配置如表7-4所示的参数。
选择“查看启动日志”时，可以查看作业启动时的日志信息。

图 7-7 调度配置



- 当节点“事件驱动调度”的“触发事件类型”配置为“DIS”时：

图 7-8 DIS 调度策略配置

配置调度策略

* DIS通道名称

* 事件处理并发数

* 事件检测间隔

* 失败策略 挂起 忽略失败，读取下一个事件

表 7-4 调度配策略参数说明

参数	说明
DIS通道名称	选择DIS通道，当指定的DIS通道有新消息时，数据开发模块将新消息传递给作业，触发该作业运行。
事件处理并发数	选择作业并行处理的数量，最大并发数为10。
事件检测间隔	配置时间间隔，检测DIS通道是否有新的消息。时间间隔单位可以配置为秒或分钟。
失败策略	选择调度失败后的策略： <ul style="list-style-type: none">● 结束调度● 忽略失败，继续调度

----结束

实时作业暂停后重新运行

当您配置的实时作业中某个节点配置有事件调度时，您可以对这个节点配置“禁用”后恢复运行，恢复运行时可以重新选择运行位置。

步骤1 登录DLF管理控制台。

步骤2 在DLF控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 选择“实时作业监控”页签，单击作业名称。

步骤4 在“作业监控”页面，单击配置有事件调度的节点，选择“禁用”。

步骤5 设置禁用后，再右键单击选择“恢复”。弹出“恢复”对话框，配置如表7-5所示的参数。

图 7-9 恢复操作

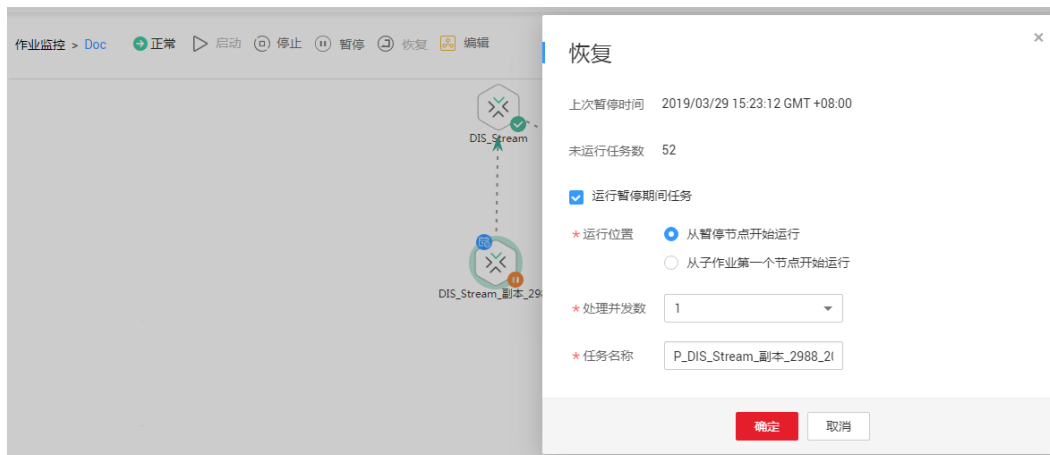


表 7-5 恢复参数说明

参数	说明
上次暂停时间	节点暂停运行的起始时间。
未运行任务数	节点暂停期间没有运行的任务数量。
“运行暂停期间任务”的参数	
运行位置	选择节点暂停运行后，恢复运行时的启动位置。 <ul style="list-style-type: none"> 从暂停节点开始运行 从子作业第一个节点开始运行
处理并发数	选择任务处理的数量。
任务名称	恢复的任务名称。

---结束

7.2.3 实时子作业监控

当用户配置的作业中某个节点配置有事件调度时，在单击此节点时可以查询子作业监控。在“子作业监控”页面可以对子作业设置停止、重跑、继续执行、强制成功、查看事件内容等操作。

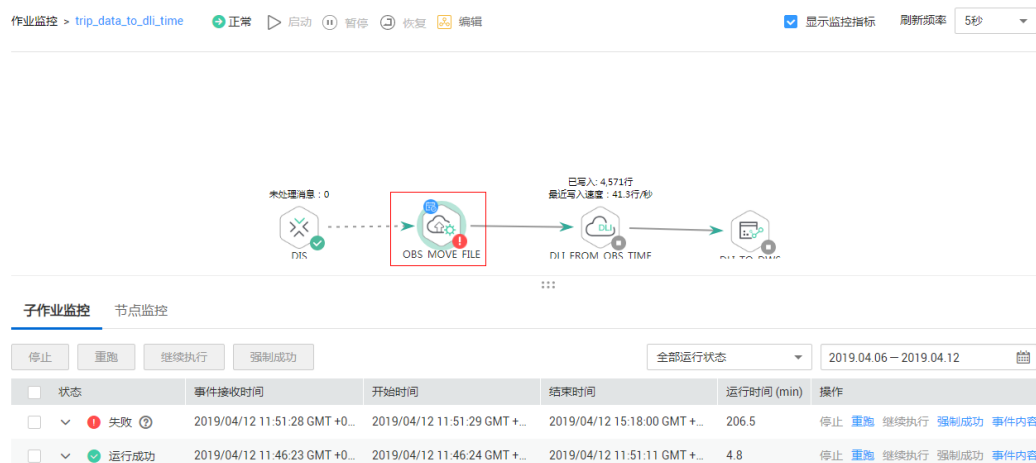
步骤1 登录DLF管理控制台。

步骤2 在DLF控制台的左侧导航栏，选择“运维调度 > 作业监控”。

步骤3 选择“实时作业监控”页签，单击作业名称。

步骤4 在“作业监控”页面，单击配置有事件调度的节点。如图7-10所示。

图 7-10 子作业监控



在“子作业监控”页面的“操作”列，提供如表7-6所示的操作。

表 7-6 操作

操作项	说明
停止	停止运行状态为“运行中”的子作业实例。
重跑	重新运行状态为“成功”或“失败”的子作业实例。
继续执行	子作业实例的状态为“运行异常”时，支持继续运行子作业实例中的后续节点。 说明 只有节点的“节点属性 > 高级 > 失败策略”设置为“挂起当前作业执行计划”时，才可以执行该操作。
强制成功	强制将状态为“失败”的子作业实例变更为“运行成功”状态。
事件内容	查看子作业的事件内容。

步骤5 单击“子作业监控”页面“状态”列下方的▼，显示该子作业节点的运行记录。

在节点的“操作”列，提供如表7-7所示的操作。

表 7-7 操作（节点）

操作项	说明
查看日志	查看节点的日志信息。
手工重试	节点的状态为“失败”时，支持重新运行节点。 说明 只有节点的“节点属性 > 高级 > 失败策略”设置为“挂起当前作业执行计划”时，才可以执行该操作。

操作项	说明
强制成功	节点的状态为“失败”时，支持将该节点强制变更为“成功”状态。 说明 只有节点的“节点属性 > 高级 > 失败策略”设置为“挂起当前作业执行计划”时，才可以执行该操作。
更多 > 跳过	节点的状态为“待运行”或“已暂停节点”时，支持跳过该节点。
更多 > 暂停	节点的状态为“待运行”时，支持暂停运行该节点，该暂停节点的后续节点将会被阻塞。
更多 > 恢复	节点的状态为“已暂停”时，支持恢复运行该节点。

----结束

7.3 实例监控

在数据开发模块控制台的左侧导航栏，选择“运维调度”，进入实例监控列表页面，用户可以在该页面中查看作业的实例信息，并根据需要对实例和节点进行更多操作，请参见[批处理：查看所有实例](#)。

当作业实例的运行状态为“失败”时，您可以通过“失败”旁边的提示信息查看作业实例失败的原因。如[图7-11](#)。

图 7-11 作业实例失败提示

实例监控

实例监控				
<input type="button" value="停止"/>	<input type="button" value="重跑"/>	<input type="button" value="继续执行"/>	<input type="button" value="强制成功"/>	
<input type="checkbox"/>	作业名称	状态	调度方式	计划时间
<input type="checkbox"/>	test222222222222...	✔ 运行成功	正常调度	2019/02/25 17:16:00 GMT...
<input type="checkbox"/>	job_9993	❗ 失败 ?	正常调度	2019/02/25 17:16:00 GMT...

重跑作业实例

您可以对运行成功或失败的作业实例设置重跑，配置重跑开始位置。

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在DLF控制台的左侧导航栏，选择“运维调度 > 实例监控”。
- 步骤3** 选择作业名称，在作业的“操作”列，单击“重跑”设置重跑作业实例；或单击作业名称左边的复选框，再选择“重跑”按钮设置作业重跑。

图 7-12 设置作业重跑



表 7-8 参数说明

参数	说明
重跑开始位置	<p>选择作业实例重跑的开始位置：</p> <ul style="list-style-type: none"> 从错误节点开始重跑：作业实例执行失败时，从实例执行失败的错误节点开始重跑。 从第一个节点开始重跑：从作业实例的第一个节点开始重跑。 从指定的节点开始重跑：从作业实例中指定的节点开始重跑。 <p>说明 以下两种情况，系统运行会从第一个节点开始重跑。</p> <ul style="list-style-type: none"> 如果作业中节点个数或者名称发生变化，从第一个节点开始重跑。 如果重跑成功状态的作业实例，从第一个节点开始重跑。

---结束

7.4 补数据监控

在数据开发模块控制台的左侧导航栏，选择“运维调度 > 补数据监控”，进入补数据的任务监控页面。

用户可以在图7-13的页面中，查看补数据的任务状态、业务日期、并行周期数、补数据作业名称，以及停止运行中的任务。

图 7-13 补数据监控主页

补数据名称	状态	业务日期	并行周期数	补数据作业名称	操作
P_job_p01_20181204_105826	运行中	2018/12/03 - 2018/12/03	2	job_p01.job_p02	停止
P_job_lbb_20181204_104121	成功	2018/12/03 - 2018/12/03	1		停止
P_job_lbb_20181204_103716	成功	2018/12/03 - 2018/12/03	1		停止
P_job_lb3_20181204_103330	成功	2018/11/29 - 2018/12/03	1	job_lb3	停止

在图7-13的页面中，单击补数据名称，进入图7-14的页面。在此页面，用户可以查看补数据的任务执行情况，以及手动干预实例和节点的执行（如需了解更多，请参见[批处理：查看所有实例](#)）。

图 7-14 补数据监控详情

补数据监控 > P_job_5433_20190409_172351

业务日期：2019/04/08 - 2019/04/08

任务	状态	计划时间	开始时间	结束时间	操作
2019/04/08	成功				
job_5433	运行成功	2019/04/08 23:58:00 GMT+08:00	2019/04/09 21:39:04 GMT+08:00	2019/04/09 21:39:56 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:56:00 GMT+08:00	2019/04/09 21:38:34 GMT+08:00	2019/04/09 21:39:10 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:54:00 GMT+08:00	2019/04/09 21:38:13 GMT+08:00	2019/04/09 21:39:01 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:52:00 GMT+08:00	2019/04/09 21:37:43 GMT+08:00	2019/04/09 21:38:31 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:50:00 GMT+08:00	2019/04/09 21:37:37 GMT+08:00	2019/04/09 21:38:10 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:48:00 GMT+08:00	2019/04/09 21:37:04 GMT+08:00	2019/04/09 21:37:40 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:46:00 GMT+08:00	2019/04/09 21:37:04 GMT+08:00	2019/04/09 21:37:35 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:44:00 GMT+08:00	2019/04/09 21:36:19 GMT+08:00	2019/04/09 21:37:01 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:42:00 GMT+08:00	2019/04/09 21:36:19 GMT+08:00	2019/04/09 21:37:01 GMT+08:00	停止 重抛 查看详情 继续执行 更多
job_5433	运行成功	2019/04/08 23:40:00 GMT+08:00	2019/04/09 21:35:19 GMT+08:00	2019/04/09 21:36:16 GMT+08:00	停止 重抛 查看详情 继续执行 更多

7.5 通知管理

7.5.1 通知管理

用户可以通过通知管理功能配置作业通知任务，当作业运行异常或成功时向相关人员发送通知。

配置通知

为作业配置通知前，已开通消息通知服务并配置主题；作业已提交，且不是“未启动”状态。

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 通知管理”。
- 步骤3** 在页面右侧的“通知管理”页签，单击“通知配置”，弹出“通知配置”页面，配置如表7-9所示的参数。

表 7-9 通知参数

参数	是否必选	说明
作业名称	是	选择作业。
通知方式	是	选择通知方式： <ul style="list-style-type: none"> ● 语音通知 ● 主题消息通知

参数	是否必选	说明
通知类型	是	选择通知类型： <ul style="list-style-type: none">运行异常/失败：作业的状态为“运行异常”或“失败”时，发送通知。运行成功：作业的状态为“成功”时，发送通知。未完成：该功能仅支持按天调度的作业配置。如果作业执行时间超过设置的未完成时间，则发送通知。 说明 实时作业只支持状态为 运行异常/失败 时发送通知，批处理作业在状态为运行成功和运行异常/失败时都能发送通知。
选择运行异常/失败的主题	是	选择作业运行异常/失败的消息主题。
选择运行成功主题	是	选择作业运行成功的消息主题。
开关	是	是否开启通知，默认开启。

步骤4 单击“确定”，为作业配置通知。

说明

数据开发模块的通知管理功能是通过消息通知服务来发送消息，消息通知服务的使用可能会产生费用，具体请咨询消息通知服务。

----结束

编辑通知

通知新建完成后，用户可以根据需求修改通知的参数。

步骤1 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 通知管理”。

步骤2 在页面右侧选择“通知管理”页签。

步骤3 在通知的“操作”列，单击“编辑”，弹出“编辑通知”页面，参考表7-9修改通知的参数。

步骤4 单击“确定”，保存修改。

----结束

关闭通知

用户可以在“编辑”中关闭通知任务，也可以在通知列表中关闭通知任务。

步骤1 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 通知管理”。

步骤2 在页面右侧选择“通知管理”页签。

步骤3 在通知的“开关”列，单击  ，切换到  时，通知为关闭状态。

----结束

查看通知记录

用户可以在通知记录中查看所有的通知信息。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 通知管理”。

步骤3 在页面右侧选择“通知记录”页签，进入通知记录页面。

----结束

删除通知

当用户不需要使用某个通知时，可以参考如下操作删除该通知。

步骤1 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 通知管理”。

步骤2 在页面右侧选择“通知管理”页签。

步骤3 在通知的“操作”列，单击“删除”，弹出“删除通知”页面。

步骤4 单击“确定”，删除通知。

----结束

7.5.2 通知周期概览

操作场景

用户可以按照天/周/月为调度周期配置通知任务，向相关人员发送通知。让相关人员可以定期跟踪作业的调度情况（作业调度成功数量，作业调度失败异常数量以及作业失败详情）。

前提条件

- 已开通消息通知服务并配置主题，为主题添加订阅。
- 已提交作业，且作业不是“未启动”状态。
- 已开通对象存储服务，并在OBS中创建文件夹。

配置通知

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“运维调度 > 通知管理”。

步骤3 在页面右侧的“周期概览”页签，单击“通知配置”，弹出“通知配置”页面，配置如表7-10所示的参数。

图 7-15 通知配置

The screenshot shows a '通知配置' (Notification Configuration) dialog box with the following fields and values:

- * 通知名称: notification
- * 调度周期: 周
- * 选择时间: 星期一, 星期二
- * 具体时间: 00 时 59 分
- * 选择概览通知的主题: doc (with a '查看主题' link)
- * 选择OBS桶: s3a://dlf-test/data/ (with an 'OBS' button)
- * 开关:

Buttons at the bottom: 确定 (OK), 取消 (Cancel).

表 7-10 通知参数

参数	是否必选	说明
通知名称	是	设置发送的通知名称。
调度周期	是	选择通知发送的调度周期，可以设置为按“天”、“周”或“月”发送。 说明 按天发送，通知记录为以发送时间往前推24小时时间段的数据；按周发送，通知记录为往前推七天时间段的数据；按月发送，通知记录为往前推30天时间段的数据
选择时间	是	设置通知发送的具体日期。 <ul style="list-style-type: none"> 当调度周期为周时，可设置为一周中星期一至星期日的某一天或某几天。 当调度周期为月时，可设置为一月中每月1号至每月31号的某一天或某几天。
具体时间	是	设置通知发送的具体时间点，可以精确设置到小时和分钟。
选择概览通知的主题	是	单击下拉选项，设置通知发送的主题。
选择OBS桶	是	单击“OBS”设置通知记录数据存储的位置。

参数	是否必选	说明
开关	是	是否开启通知，默认开启。

步骤4 单击“确定”。

说明

数据开发模块的通知管理功能是通过消息通知服务来发送消息，消息通知服务的使用可能会产生费用，具体请咨询消息通知服务。

步骤5 通知配置完成后，您可以在通知的“操作”列进行如下操作。

- 单击“编辑”，打开“通知配置”页面，可以重新编辑通知。编辑完成后选择“确定”，保存修改。
- 单击“记录”，打开“查看记录”页面，可以查看作业的调度情况。
- 单击“删除”，打开“删除通知”页面，选择“确定”，删除通知。

----结束

7.6 备份和还原资产

通过备份功能，您可每日定时备份昨日系统中的所有作业、脚本、资源和环境变量。

通过还原功能，您可还原已备份的资产，包含作业、脚本、资源和环境变量。

前提条件

已开通对象存储服务，并在OBS中创建桶。

备份资产

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“备份管理”。

步骤3 单击“启动每日备份”，打开“OBS文件浏览”页面，选择OBS文件夹，设置备份数据的存储位置。

图 7-16 备份管理



说明

- 每日备份在每日0点开始备份昨日的所有作业、脚本、资源和环境变量，启动当日不会备份昨日的作业、脚本、资源和环境变量。
- 选择OBS存储路径时，若仅选择至桶名层级，则备份对象自动存储在以“备份日期”命名的文件夹内。环境变量，资源，脚本和作业分别存储在1_env,2_resources,3_scripts和4_jobs文件夹内。
- 备份成功后，在以“备份日期”命名的文件夹内，自动生成backup.json文件，该文件按照节点类型存储了作业信息，支持恢复作业前进行修改。
- 启动每日备份后，若想结束备份任务，您可以单击右边的“停止每日备份”。

----结束

还原资产

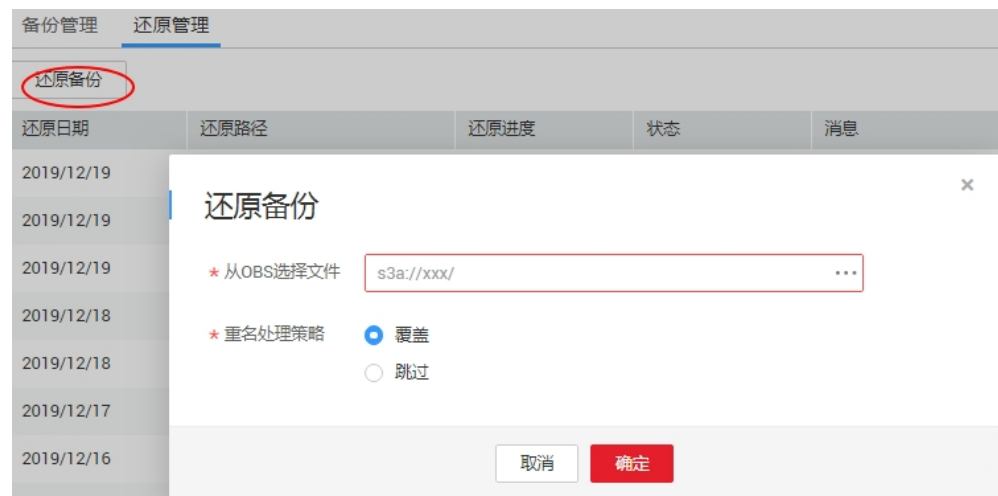
- 步骤1** 登录DLF管理控制台。
- 步骤2** 在数据开发模块控制台的左侧导航栏，选择“备份管理”。
- 步骤3** 选择“还原管理”页签，单击“还原备份”。

在还原备份对话框中，从OBS桶中选择待还原的资产存储路径，设置重名处理策略。

说明

- 待还原的资产存储路径请获取**备份资产**中生成的文件路径。
- 您可在还原资产前修改备份路径下的backup.json文件，支持修改连接名（connectionName）、数据库名（database）和集群名（clusterName）。

图 7-17 还原资产



步骤4 单击“确定”。

----结束

8 配置管理

8.1 主机连接管理

用户通过主机连接管理功能保存主机的连接信息，这些连接可以使用于脚本和作业。当主机的连接信息有变化时，只需在主机连接管理中编辑修改，而不需要到具体的脚本或作业中逐一修改。

新建主机连接

主机连接的最大配额为20，请确保当前主机连接的数量未达到最大配额。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 主机连接管理”。

步骤3 单击“创建主机连接”，弹出“创建主机连接”页面，配置如表8-1所示的参数。

表 8-1 参数说明

参数	是否必选	说明
主机连接名称	是	主机连接的名称，只能包含英文字母、数字、“_”，长度为1~100个字符，且名称不能重复。
主机地址	是	主机的地址。请参见《弹性云服务器用户指南》的“查看云服务器详细信息”页面获取。
端口	是	主机的端口号。
用户名	是	主机的用户名。
登录方式	是	选择主机的登录方式： <ul style="list-style-type: none">• 密钥对• 密码

参数	是否必选	说明
密钥对	是	主机的登录方式为密钥对时，用户获取并上传其私钥文件至OBS，在此处选择对应的OBS路径。
密钥对密码	否	如果密钥对未设置密码，则不需要填写该配置项。
密码	是	主机的登录方式为密码时，填写主机的登录密码。
绑定Agent	是	选择CDM集群，CDM集群提供Agent。
KMS密钥	是	通过KMS加密和解密用户密码和密钥对的密码，选择KMS中已创建的密钥。
主机连接描述	否	主机连接的描述信息。

步骤4 单击“测试”，测试主机的连通性。如果无法连通，主机连接将无法创建。

步骤5 单击“确定”，创建主机连接。

----结束

编辑主机连接

主机连接新建完成后，用户可以根据需求修改主机连接的参数。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 主机连接管理”。

步骤3 在主机连接的“操作”列，单击“编辑”，弹出“编辑主机连接”页面，参考表8-1修改主机连接的参数。

说明

为了确保主机连接的安全性，修改主机连接参数时，需要重新输入主机的密码。

步骤4 单击“测试”，测试主机的连通性。如果无法连通，主机连接将无法保存修改。

步骤5 单击“确定”，保存修改。

----结束

删除主机连接

当用户不需要使用某个主机连接时，可以参考如下操作删除该主机连接。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 主机连接管理”。

步骤3 在主机连接的“操作”列，单击“删除”，弹出“删除主机连接”页面。

步骤4 单击“确定”，删除主机连接。

须知

如果某一个待删除的主机连接正在被脚本或作业关联，请确保强制删除主机连接后，不影响业务使用。如果希望脚本或作业能继续正常使用，请前往脚本或作业开发页面，重新关联可用的主机连接。

----结束

8.2 资源管理

用户可以通过资源管理功能，上传自定义代码或文本文件作为资源，在节点运行时调用。可调用资源的节点包含DLI Spark，MRS Spark，MRS MapReduce和DLI Flink Job。

创建资源后，配置资源关联的文件。在作业中可以直接引用资源。当资源文件变更，只需要修改资源引用的位置即可，不需要修改作业配置。关于资源的使用样例请参见[开发一个Spark作业](#)。

新建目录（可选）

如果已存在可用的目录，可以不用新建目录。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。


步骤3 在资源目录中，单击，弹出“新建目录”页面，配置如表8-2所示的参数。

表 8-2 资源目录参数

参数	说明
目录名称	资源目录的名称，只能包含英文字母、数字、中文字符、“_”、“-”，且长度为1~32个字符。
选择目录	选择该资源目录的父级目录，父级目录默认为根目录。

步骤4 单击“确定”，新建目录。

----结束

新建资源

新建资源前，已开通OBS服务。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。

步骤3 单击“新建资源”，弹出“新建资源”页面，配置如表8-3所示的参数。

表 8-3 资源管理参数

参数	是否必选	说明
名称	是	资源的名称，只能包含英文字母、数字、中文字符、“_”、“-”，且长度为1~32个字符。
类型	是	选择资源的文件类型： <ul style="list-style-type: none">• jar• file• archive
资源位置	是	选择资源所在的位置： <ul style="list-style-type: none">• 本地• OBS
主Jar包	是	选择已上传到OBS中的主Jar包。“类型”为“jar”，且“资源位置”为“OBS”时，配置该参数。
依赖Jar包	否	选择已上传到OBS中的依赖Jar包。“类型”为“jar”，且“资源位置”为“OBS”时，配置该参数。
选择资源	是	选择具体的资源文件。
存储路径	是	选择资源的存储路径。“资源位置”为“本地”时，配置该参数。
描述	否	资源的描述信息。
选择目录	是	选择资源所属的目录，默认为根目录。

步骤4 单击“确定”，新建资源。

----结束

编辑资源

资源新建完成后，用户可以根据需求修改资源的参数。

步骤1 登录DLF管理控制台。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。

步骤3 在资源的“操作”列，单击“编辑”，弹出“编辑资源”页面，参考表8-3修改资源的参数。

步骤4 单击“确定”，保存修改。

----结束


删除资源

当用户不需要使用某个资源时，可以参考如下操作删除该资源。删除资源前，请确保该资源未被作业使用。

- 步骤1** 登录DLF管理控制台。
 - 步骤2** 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。。
 - 步骤3** 在资源的“操作”列，单击“删除”，弹出“删除资源”页面。
 - 步骤4** 单击“确定”，删除资源。
- 结束

导入资源

当用户想要导入某个资源时，可以参考如下操作导入该资源。

- 步骤1** 登录DLF管理控制台。
 - 步骤2** 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。。
 - 步骤3** 在资源目录中，右键单击 ，选择“导入资源”，弹出“导入资源”页面。
 - 步骤4** 导入已上传至OBS中的资源管理文件，单击“下一步”和“关闭”，导入资源。
- 结束

导出资源

当用户想要导出某个资源到本地时，可以参考如下操作导出该资源。

- 步骤1** 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。。
 - 步骤2** 在资源的“操作”列，单击“导出”，弹出“导出资源”页面。
 - 步骤3** 选择“只导出资源定义”或“导出资源定义和资源文件”，单击“确定”，导出资源。
- 结束

9 配置

用户可以在“配置”界面设置“环境变量”和“工作区”，配置日志存储路径。工作区的配置是项目级变量，环境变量的配置是工作区级变量。在工作区内设置好的环境变量，可以被该工作区的作业、脚本随意使用。

9.1 工作区

您可以自定义工作区，用以隔离资源。

操作步骤

步骤1 登录DLF管理控制台。

步骤2 在DLF控制台的左侧导航栏，选择“配置”。

步骤3 单击“工作区”，在“工作区配置”页面，单击新建，输入工作区名称并配置其所属的企业项目。

说明

- 已开通企业项目管理服务的用户才可以配置该参数。默认值为default。
- 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。
- 新建工作区您可以选择默认的企业项目“default”或其他已有的企业项目。如果要创建新的企业项目，请登录企业管理控制台进行创建，详细操作请参考《企业管理用户指南》。

图 9-1 新建工作区



步骤4 单击“确认”，建立工作区。

建立完一个工作区后，您还可以进行新增，编辑，删除等操作。

- 新增：单击“新建”，配置新的工作区。
- 编辑：单击“操作”列的“编辑”，修改工作区的名称。
- 删除：单击“操作”列的“删除”，删除工作区。

----结束

使用工作区

单击服务页上方“工作区”旁边的下拉三角，可以查看创建的“工作区”。

- 单击选择一个“工作区”，在该“工作区”创建“环境变量”，创建好的“环境变量”可以被该工作区的作业和脚本随意使用。如何使用环境变量，请参见[环境变量](#)。
- “工作区”内可以创建作业和脚本，如何创建请参见[脚本开发](#)和[作业开发](#)。
- “工作区”内可以配置日志存储路径，如何配置请参见[配置作业日志存储路径](#)。

9.2 管理企业项目

企业项目是一种云资源管理方式。企业管理（Enterprise Management）提供面向企业客户的云上资源管理、人员管理、权限管理、财务管理等综合管理服务。区别于管理控制台独立操控、配置云产品的方式，企业管理控制台以面向企业资源管理为出发点，帮助企业以公司、部门、项目等分级管理方式实现企业云上的人员、资源、权限、财务的管理。

已开通企业项目服务的用户，可以使用企业项目管理上的云服务资源。

绑定企业项目

用户可以在创建工作区时为工作区选择所属的企业项目，从而将DLF工作区与企业项目进行关联，详情请参见[工作区](#)。在选择“企业项目”的下拉列表中，将显示用户在企

业项目服务中已创建的项目。系统还内置了一个缺省的企业项目“default”，如果用户没有为工作区选择企业项目，将使用缺省项目“default”。

在工作区创建过程中，如果工作区与企业项目绑定成功，则工作区创建成功，如果绑定失败，系统会发送告警，工作区创建失败。

当删除DLF工作区时，DLF工作区与企业项目的关联关系就会被自动删除。

查看企业项目

工作区创建成功后，您可以在工作区配置页面查看工作区关联的企业项目。用户只能查询到有访问权限的项目下的工作区资源。

在工作区配置页面的列表中，查看工作区所属的企业项目。

图 9-2 查看企业项目

工作区配置

+新建

名称	ID	企业项目	操作
default	62099355b894428e8916573ae635f1f9	default	编辑 删除
testworkepsfsdagfd	8a2c21386c3c22f3016c3cdd43fa013e	testeps	编辑 删除

同时，在企业管理的控制台中，查询指定项目中的资源列表时，也可以查询到DLF服务的资源。

将通道迁入或迁出企业项目

一个DLF工作区只能关联一个企业项目。当工作区创建成功后，可以在企业管理的控制台中，执行迁出操作，将DLF工作区从当前所属的企业项目中迁出到另一个企业项目中；或者执行迁入操作，在指定的企业项目中迁入另一个企业项目中的DLF工作区。迁入迁出后，DLF工作区与新的企业项目进行关联，DLF工作区与原企业项目的关联关系将被自动解除。详细操作，请参考《企业管理用户指南》的“企业项目管理 > 如何管理资源”章节。

9.3 环境变量

本章节主要介绍环境变量的配置和使用。

操作步骤

- 步骤1** 登录DLF管理控制台。
- 步骤2** 在DLF控制台的左侧导航栏，选择“配置”。
- 步骤3** 单击“环境变量”，在“环境变量配置”页面，配置如表9-1所示的参数，单击“保存”。



图 9-3 环境变量配置

环境变量配置

表 9-1 环境变量参数配置

参数	是否必选	说明
参数名称	是	只支持英文字母、数字、“-”、“_”，最大长度为64字符，且参数名称不允许重名。
参数值	是	参数值当前支持常量和EL表达式，不支持系统函数。例如支持123, abc。 关于EL表达式的使用，请参见 表达式概述 。

配置完一个环境变量后，您还可以进行新增，修改，删除等操作。

- 新增：单击“新增”配置新的环境变量。
- 修改：参数值为常量时，直接在文本框中修改参数值；参数值为EL表达式时，可以单击文本框后方的  编辑EL表达式，修改参数值。修改完成后，请“保存”。
- 删除：在参数值文本框后方，单击  删除环境变量。

----结束

使用方法

当前配置好的环境变量支持如下两种使用方法：

1. `${环境变量名}`
2. `#{Env.get(“环境变量名”)}`

操作示例

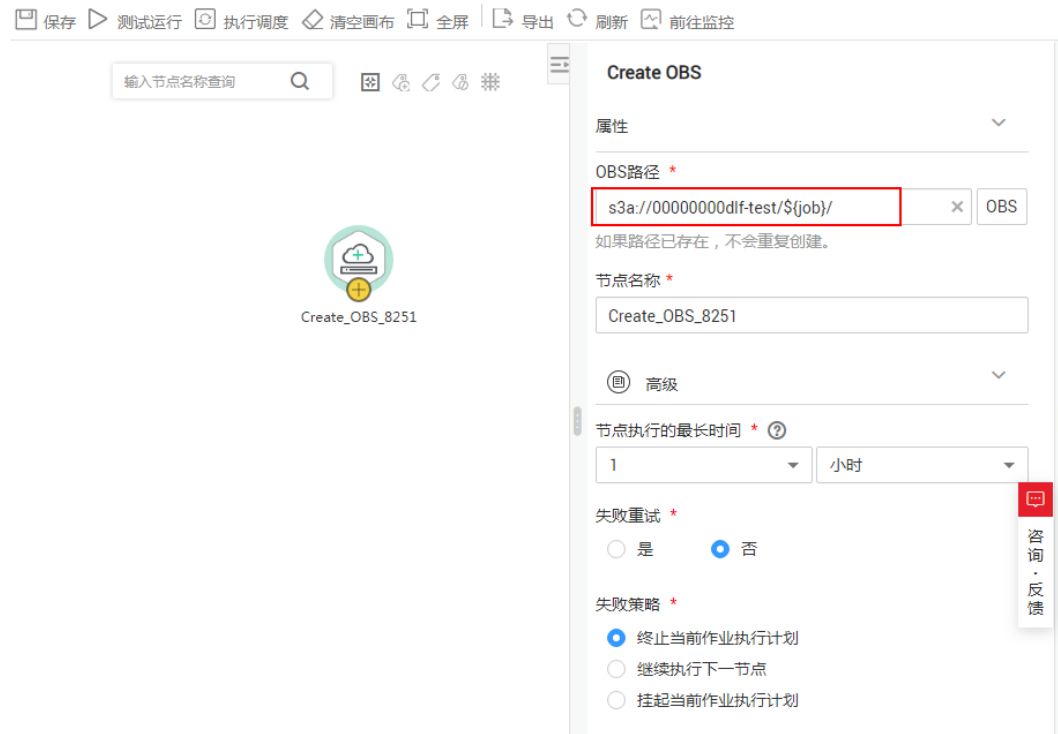
背景信息：

- 在DLF系统中已创建一个作业“test”
- 在环境变量中已新增一个变量，“参数名”为“job”，“参数值”为“123”。

步骤1 打开作业“test”，从左侧节点库中拖拽一个“Create OBS”节点。

步骤2 在节点属性页签中配置属性。

图 9-4 Create OBS



步骤3 单击“保存”后，选择“前往监控”页面监控作业的运行情况。

----结束

9.4 配置作业日志存储路径

作业日志和DLI脏数据默认存储在以dlf-log-{Project id}命名的OBS桶中，您也可以自定义日志存储路径，DLF支持您基于工作区全局配置OBS桶。

步骤1 登录DLF管理控制台。

步骤2 控制台上方选择对应的工作区。

步骤3 在DLF控制台的左侧导航栏，选择“配置”。

步骤4 单击“OBS桶”，进入OBS桶配置页面。

步骤5 单击“作业日志OBS桶”和“DLI脏数据OBS桶”后的OBS，选择OBS桶名称并确认。

----结束

9.5 配置委托

数据开发模块的作业执行中会遇到如下问题：

- 数据开发模块的作业执行机制是以启动作业的用户身份执行该作业。对于按照周期调度方式执行的作业，当启动该作业的IAM账号在调度周期内被删除后，系统无法获取用户身份认证信息，导致作业执行失败。

- 如果作业被低权限的用户启动，也会因为权限不足导致作业执行失败。

若需解决以上两个问题，则可配置委托。配置委托后，作业执行过程中，以委托的身份与其他服务交互，可以避免上述两种场景下作业执行失败。

委托的作用

由于华为云各服务之间存在业务交互关系，一些云服务需要与其他云服务协同工作，需要您创建云服务委托，将操作权限委托给该服务，让该服务以您的身份使用其他云服务，代替您进行一些资源运维工作。

委托的分类

委托分两类，工作空间委托和作业委托。

- 工作空间委托：工作空间级别的，全局委托。适用于该空间内的所有作业。
- 作业委托：适用于单个作业级别。

作业委托优先级高于工作空间委托，如果工作空间与作业级别的委托都没有配置，作业会以启动者的身份去执行。

约束限制

- 创建或修改委托需要用户具有Security Administrator权限。
- 配置工作空间级委托，需要用户具有DAYU Administrator或者Tenant Administrator权限。
- 配置作业级委托，具有查看列表委托权限的用户均可配置。

创建委托

1. 登录IAM服务控制台。
2. 选择“委托 > 创建委托”。
3. 设置“委托名称”。例如：DAYU_agency。
4. “委托类型”选择“云服务”，在“云服务”中选择“DAYU”，将操作权限委托给DAYU，让DAYU以您的身份使用其他云服务，代替您进行一些资源运维工作，如图9-5所示。
5. “持续时间”选择“永久”。

图 9-5 创建委托

* 委托名称

* 委托类型 普通账号 云服务

* 云服务 DAYU

* 持续时间

描述

0/255

6. 在“权限选择”区域中，单击“配置权限”。
7. 在弹出页面中搜索“Tenant Administrator”策略，勾选“Tenant Administrator”策略并单击“确定”，如图9-6所示。
 - 因Tenant Administrator策略具有除统一身份认证服务外，其他所有服务的所有执行权限。所以给委托服务DAYU配置Tenant Administrator，可访问周边所有服务。
 - 若您想达到对权限较小化的安全管控要求，Tenant Administrator可不配置，仅配置OBS OperateAccess权限（因作业执行过程中，需要往obs写执行日志信息，因此需要添加 OBS OperateAccess权限。）。然后再根据作业中的节点类型，配置不同的委托权限。例如某作业仅包含Import GES节点，可配置GES Administrator权限和OBS OperateAccess权限即可。详细方案请参考[配置权限](#)。

图 9-6 配置权限



8. 单击“确定”完成委托创建。

配置权限

将账号的操作权限委托给DAYU服务后，需要配置委托身份的权限，才可与其他服务进行交互。

为实现对权限较小化的安全管控要求，可根据作业中的节点类型，以服务为粒度，参见表9-2配置相应的服务Admin权限。

也可精确到具体服务的操作、资源以及请求条件等。根据作业中的节点类型，以对应服务API接口为粒度进行权限拆分，满足企业对权限最小化的安全管控要求。参见表9-3进行配置。例如包含Import GES节点的作业，您只需要创建自定义策略，并勾选ges:graph:getDetail（查看图详情），ges:jobs:getDetail（查询任务状态），ges:graph:access（使用图）这三个授权项即可。

须知

- MRS相关的节点（MRS Presto SQL、MRS Spark、MRS Spark Python、MRS Flink Job、MRS MapReduce），以及通过直连方式的（MRS Spark SQL、MRS Hive SQL）节点，由于部分MRS集群不支持委托方式提交作业，所以这类作业不能配置委托。
- 支持委托方式提交作业的MRS集群如下：
 - 非安全集群
 - 安全集群，集群版本大于 2.1.0，并且安装了MRS 2.1.0.1以上的补丁。
- 配置服务级Admin权限
因作业执行过程中，需要往obs写执行日志信息，因此粗粒度授权时，所有作业都需要添加 OBS OperateAccess权限。

表 9-2 配置相关节点的 admin 权限

节点名称	系统权限	权限描述
CDM Job、DIS Stream、DIS Dump、DIS Client	DAYU Administrator	智能数据湖运营平台服务的所有执行权限。
Import GES	GES Administrator	图引擎服务的所有执行权限。该角色有依赖，需要在同项目中勾选依赖的角色：Tenant Guest、Server Administrator。
<ul style="list-style-type: none"> MRS Presto SQL、MRS Spark、MRS Spark Python、MRS Flink Job、MRS MapReduce MRS Spark SQL、MRS Hive SQL（通过MRS API方式连接MRS集群的） 	MRS Administrator KMS Administrator	<p>MRS Administrator: MapReduce服务的所有执行权限。该角色有依赖，需要在同项目中勾选依赖的角色：Tenant Guest、Server Administrator。</p> <p>KMS Administrator: 数据加密服务加密密钥的管理员权限。</p>
MRS Spark SQL、MRS Hive SQL、MRS Kafka、Kafka Client（通过代理方式连接集群）	DAYU Administrator KMS Administrator	<p>DAYU Administrator: 智能数据湖运营平台服务的所有执行权限。</p> <p>KMS Administrator: 数据加密服务加密密钥的管理员权限。</p>
DLI Flink Job、DLI SQL、DLI Spark	DLI Service Admin	数据湖探索的所有执行权限。
DWS SQL、Shell、RDS SQL（通过代理方式连接数据源）	DAYU Administrator KMS Administrator	<p>DAYU Administrator: 智能数据湖运营平台服务的所有执行权限。</p> <p>KMS Administrator: 数据加密服务加密密钥的管理员权限。</p>
CSS	DAYU Administrator Elasticsearch Administrator	<p>DAYU Administrator: 智能数据湖运营平台服务的所有执行权限。</p> <p>Elasticsearch Administrator: 云搜索服务的所有执行权限。该角色有依赖，需要在同项目中勾选依赖的角色：Tenant Guest、Server Administrator。</p>
Create OBS、Delete OBS、OBS Manager	OBS OperateAccess	查看桶、上传对象、获取对象、删除对象、获取对象ACL等对象基本操作权限
SMN	SMN Administrator	消息通知服务的所有执行权限。

- 配置细粒度权限（根据各服务支持的授权项，创建自定义策略。）
创建自定义策略的详细操作请参见[创建自定义策略](#)。

说明

- 作业执行过程中，需要向OBS中写入执行日志。当采取精细化授权方式时，任何类型的作业均需要添加OBS的如下授权项：
 - obs:bucket:GetBucketLocation
 - obs:object:GetObject
 - obs:bucket>CreateBucket
 - obs:object:PutObject
 - obs:bucket:ListAllMyBuckets
 - obs:bucket:ListBucket
- CDM Job、DIS Stream、DIS Dump、DIS Client节点隶属于DAYU模块，DAYU不支持细粒度授权。因此包含这几类节点的作业，给服务配置权限仅支持DAYU Administrator。
- CSS不支持细粒度授权，且需要通过代理执行。因此包含这类节点的作业，需要配置DAYU Administrator和Elasticsearch Administrator权限。
- SMN不支持细粒度授权，因此包含这类节点的作业，需要配置SMN Administrator权限。

表 9-3 自定义策略

节点名称	授权项
Import GES	<ul style="list-style-type: none"> ges:graph:access ges:graph:getDetail ges:jobs:getDetail
<ul style="list-style-type: none"> MRS Presto SQL、MRS Spark、MRS Spark Python、MRS Flink Job、MRS MapReduce MRS Spark SQL、MRS Hive SQL（通过MRS API方式连接MRS集群的） 	<ul style="list-style-type: none"> mrs:job:delete mrs:job:stop mrs:job:submit mrs:cluster:get mrs:cluster:list mrs:job:get mrs:job:list kms:dek:crypto kms:cmk:get
MRS Spark SQL、MRS Hive SQL、MRS Kafka、Kafka Client（通过代理方式连接集群）	<ul style="list-style-type: none"> kms:dek:crypto kms:cmk:get DAYU Administrator(角色)

节点名称	授权项
DLI Flink Job、DLI SQL、DLI Spark	<ul style="list-style-type: none">• dli:jobs:get• dli:jobs:update• dli:jobs:create• dli:queue:submit_job• dli:jobs:list• dli:jobs:list_all
DWS SQL、Shell、RDS SQL（通过代理方式连接数据源）	<ul style="list-style-type: none">• kms:dek:crypto• kms:cmk:get• DAYU Administrator(角色)
Create OBS、Delete OBS、OBS Manager	<ul style="list-style-type: none">• obs:bucket:GetBucketLocation• obs:bucket:ListBucketVersions• obs:object:GetObject• obs:bucket:CreateBucket• obs:bucket>DeleteBucket• obs:object>DeleteObject• obs:object:PutObject• obs:bucket:ListAllMyBuckets• obs:bucket:ListBucket

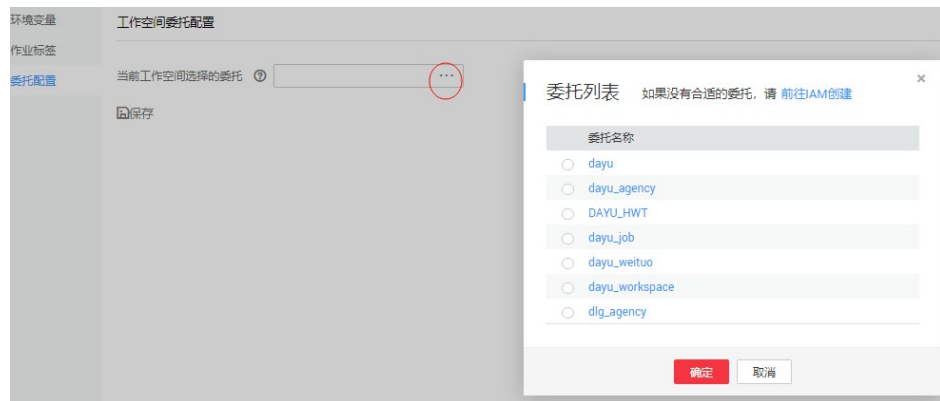
配置工作空间级委托


注意

工作空间级别的委托影响所有的作业，请慎重配置。特别是部分作业中包含 MRS 相关的节点。

1. 在数据开发主界面的左侧导航栏，选择“配置”。
2. 单击“委托配置”，在工作空间委托配置页面配置委托。
3. 在委托列表中选择合适的委托，也可重新创建委托。创建委托和配置权限，请参见[创建委托](#)。

图 9-7 配置工作空间级委托



4. 单击“确定”，回到工作空间委托配置页面，再单击，创建工作空间级委托成功。

配置作业级委托

说明

支持新建作业时，配置作业级委托。也支持修改已有作业的委托。

新建作业时配置委托

1. 在数据开发主界面的左侧导航栏，选择“数据开发 > 作业开发”。
2. 在作业目录处，单击右键，选择“新建作业”。系统弹出新建作业对话框，若已配置过工作空间级委托，则该作业默认使用工作空间级委托。您也可从委托列表中，选择其他已创建的委托。

图 9-8 配置作业委托

修改已有作业的委托

1. 在数据开发主界面的左侧导航栏，选择“数据开发 > 作业开发”。
2. 在作业目录处，双击选中已有作业。在节点编排页面右侧，选择“作业基本信息”。系统弹出作业信息基本配置对话框，若已配置过工作空间级委托，则该作业默认使用工作空间级委托。您也可从委托列表中，选择其他已创建的委托。

10 使用教程

10.1 开发一个 Spark 作业

在本章节用户可以学习到数据开发模块资源管理、作业编辑等功能。

场景说明

用户在使用DLI服务时，大部分时间会使用SQL对数据进行分析处理，有时候处理的逻辑特别复杂，无法通过SQL处理，那么可以通过Spark作业进行分析处理。本章节通过一个例子演示如何在数据开发模块中提交一个Spark作业。

操作流程如下：

1. 创建DLI集群，通过DLI集群的物理资源来运行Spark作业。
2. 获取Spark作业的演示JAR包，并在数据开发模块中关联到此JAR包。
3. 创建数据开发模块作业，通过DLI Spark节点提交Spark作业。

环境准备

- 已开通对象存储服务OBS，并创建桶，例如“s3a://dlfexample”，用于存放Spark作业的JAR包。
- 已开通数据湖探索服务DLI，并创建Spark集群“spark_cluster”，为Spark作业提供运行所需的物理资源。

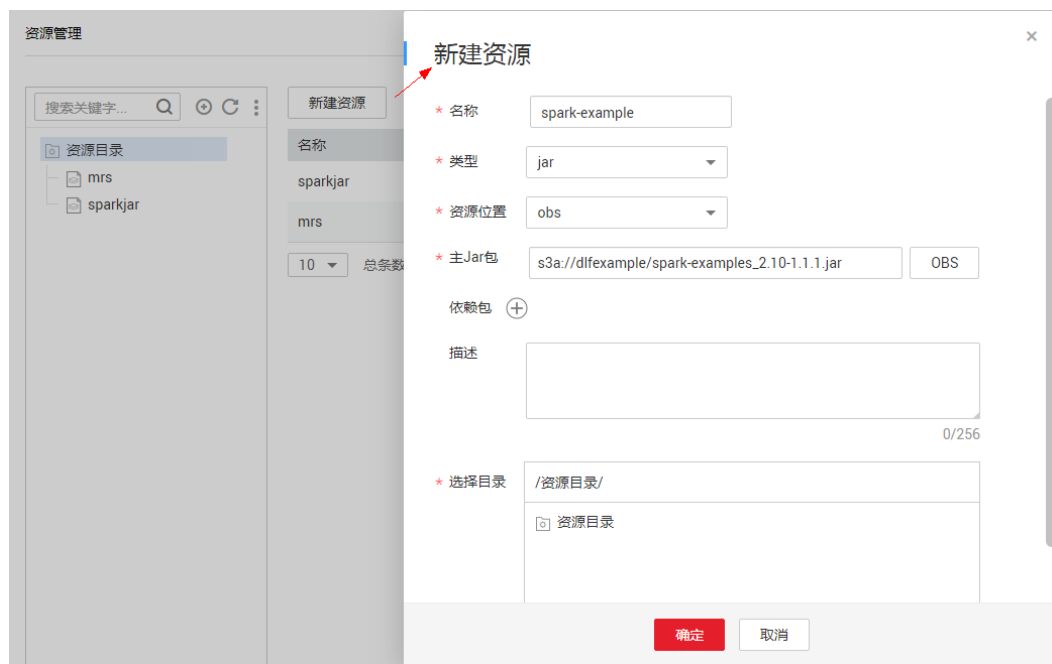
获取 Spark 作业代码

本示例使用的Spark作业代码来自maven库（下载地址：[Spark作业代码](#)），下载spark-examples_2.10-1.1.1.jar，此Spark作业是计算 π 的近似值。

步骤1 获取Spark作业代码JAR包后，将JAR包上传到OBS桶中，存储路径为“s3a://dlfexample/spark-examples_2.10-1.1.1.jar”。

步骤2 在数据开发模块控制台的左侧导航栏，选择“配置管理 > 资源管理”。在数据开发模块中创建一个资源关联到**步骤1**的JAR包，资源名称为“spark-example”。

图 10-1 创建资源



----结束

提交 Spark 作业

用户需要在数据开发模块中创建一个作业，通过作业的DLI Spark节点提交Spark作业。

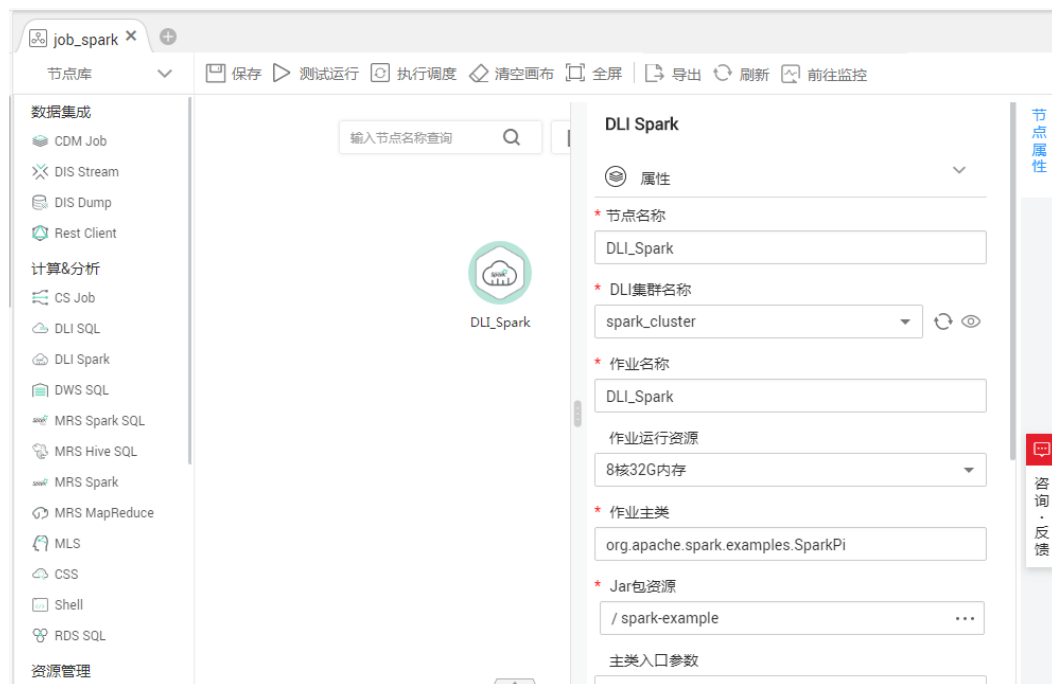
步骤1 创建一个DLF空作业，作业名称为“job_spark”。

图 10-2 创建作业



步骤2 然后进入作业开发页面，拖动DLI Spark节点到画布并单击，配置节点的属性。

图 10-3 配置节点属性



关键属性说明：

- DLI集群名称：已开通数据湖探索服务DLI，并创建Spark集...中创建的Spark集群。
- 作业运行资源：DLI Spark节点运行时，限制最大可以使用的CPU、内存资源。
- 作业主类：DLI Spark节点的主类，本例的主类是“org.apache.spark.examples.SparkPi”。
- Jar包资源：步骤2中创建的资源。


步骤3 作业编排完成后，单击 ，测试运行作业。

图 10-4 作业日志（仅参考）



步骤4 如果日志运行正常，单击  保存作业。

----结束

10.2 开发一个 Hive SQL 脚本

本章节介绍如何在数据开发模块上进行Hive SQL开发。

场景说明

数据开发模块作为一站式大数据开发平台，支持多种大数据工具的开发。Hive是基于Hadoop的一个数据仓库工具，可以将结构化的数据文件映射为一张数据库表，并提供简单的SQL查询功能；可以将SQL语句转换为MapReduce任务进行运行。

环境准备

- 已开通MapReduce服务MRS，并创建MRS集群“MRS_1009”，为Hive SQL提供运行环境。

MRS集群创建时，需要注意：

- 集群关闭Kerberos认证。
- 组件要包含Hive。

- 已开通云数据迁移服务CDM，并创建CDM集群“cdm-7357”，为数据开发模块提供数据开发模块与MRS通信的代理。

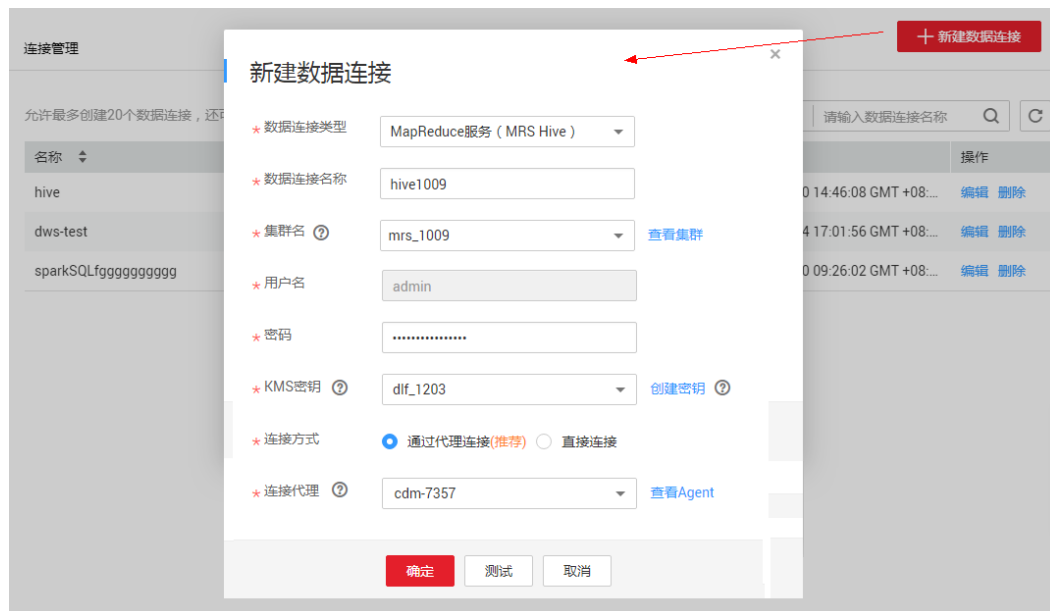
CDM集群创建时，需要注意：

- 虚拟私有云、子网、安全组与MRS集群“MRS_1009”可以网络互通。

建立 Hive 的数据连接

开发Hive SQL前，我们需要在数据开发模块中建立一个到MRS Hive的连接，数据连接名称为“hive1009”。

图 10-5 创建数据连接



关键参数说明：

- 集群名：**已开通MapReduce服务MRS，并创建MR...**中创建的MRS集群。
- 绑定Agent：**已开通数据迁移服务CDM，并创建CDM集群“c...**中创建的CDM集群。


开发 Hive SQL 脚本

在数据开发模块中创建一个Hive SQL脚本，脚本名称为“hive_sql”。在编辑器中输入SQL语句，通过SQL语句来实现业务需求。

图 10-6 开发脚本



关键说明：

- 图10-6中的脚本开发区为临时调试区，关闭脚本页签后，开发区的内容将丢失。请单击，将脚本保存至指定的目录中。
- 数据连接：[建立Hive的数据连接](#)创建的连接。

开发 Hive SQL 作业

Hive SQL脚本开发完成后，我们为Hive SQL脚本构建一个周期执行的作业，使得该脚本能定期执行。

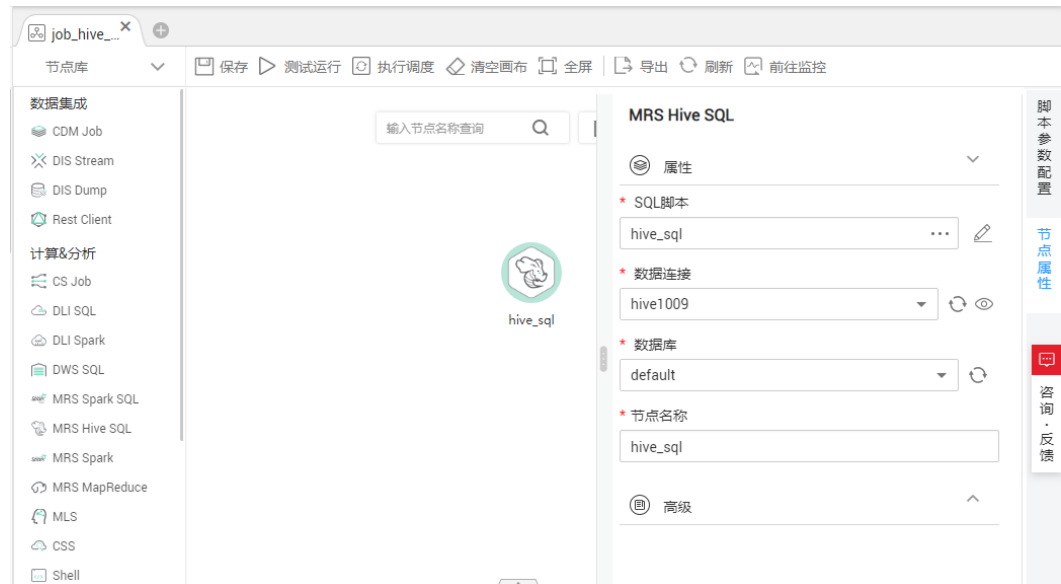
步骤1 创建一个数据开发模块空作业，作业名称为“job_hive_sql”。

图 10-7 创建 job_hive_sql 作业




步骤2 然后进入到作业开发页面，拖动MRS Hive SQL节点到画布中并单击，配置节点的属性。

图 10-8 配置 MRS Hive SQL 节点属性



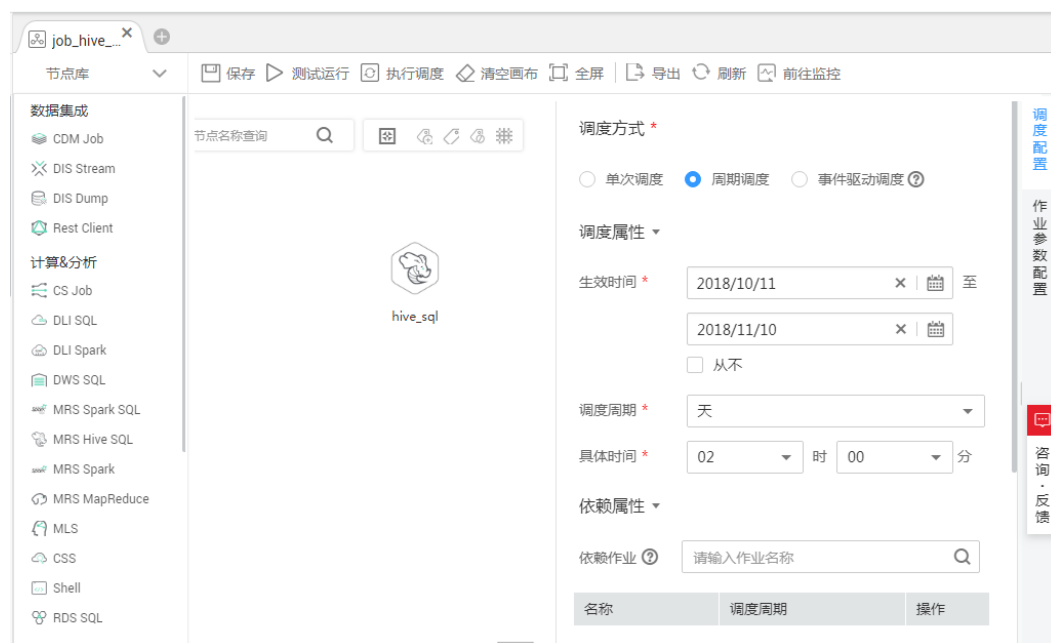
关键属性说明：

- SQL脚本：关联[开发Hive SQL脚本](#)中开发完成的Hive SQL脚本“hive_sql”。
- 数据连接：默认选择SQL脚本“hive_sql”中设置的数据连接，支持修改。
- 数据库：默认选择SQL脚本“hive_sql”中设置的数据库，支持修改。
- 节点名称：默认显示为SQL脚本“hive_sql”的名称，支持修改。

步骤3 作业编排完成后，单击 ，测试运行作业。



步骤4 如果日志运行正常，单击画布空白处，在右侧的“调度配置”页面，配置作业的调度策略。

图 10-9 配置调度方式



说明：

- 2018/10/11至2018/11/10，每天2点执行一次作业。

步骤5 最后我们需要保存作业（单击 ），并执行调度作业（单击 ），实现作业每天自动运行。

----结束

11 参考

11.1 节点

11.1.1 节点概述

节点定义对数据执行的操作。数据开发模块提供数据集成、计算&分析、数据库操作、资源管理等类型的节点，用户根据业务模型选择所需的节点。

- 节点参数支持使用EL表达式，EL表达式的使用方法详见[表达式概述](#)。
- 节点间的连接方式支持串行和并行。
串行连接：按顺序逐个执行节点，当A节点执行完成后，再执行B节点。
并行连接：A节点和B节点同时执行。

图 11-1 连接示意图



11.1.2 CDM Job

功能

通过CDM Job节点执行一个预先定义的CDM作业，实现数据迁移功能。

参数

用户可参考表11-1，表11-2和表11-3配置CDM Job节点的参数。配置血缘关系用以标识数据流向，在数据资产模块中可以查看。

表 11-1 属性参数

参数	是否必选	说明
CDM集群名称	是	选择待执行的CDM作业所属的CDM集群。
CDM作业名称	是	选择待执行的CDM作业。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。







表 11-2 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>

参数	是否必选	说明
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> • 终止当前作业执行计划 • 继续执行下一作业 • 挂起当前作业执行计划 • 终止当前以及后续节点执行计划

表 11-3 血缘关系

参数	说明
输入	
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS和DLI类型。</p> <ul style="list-style-type: none"> • DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 • OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 • CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 • DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输入功能的参数配置。
取消	单击“取消”，取消节点输入功能的参数配置。

参数	说明
编辑	单击  ，修改节点输入功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输入功能的参数配置。
查看表详情	单击  ，查看节点输入血缘关系创建数据表的详细信息。
输出	
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS，DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输出功能的参数配置。
取消	单击“取消”，取消节点输出功能的参数配置。
编辑	单击  ，修改节点输出功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输出功能的参数配置。
查看表详情	单击  ，查看节点输出血缘关系创建数据表的详细信息。

11.1.3 DIS Stream

功能

通过DIS Stream节点查询DIS通道的状态，如果DIS通道运行正常，继续执行后续的节点；如果DIS通道运行异常，DIS Stream将报错并退出，此时如果需要继续执行后续的节点，请配置“失败策略”为“继续执行下一节点”，请参见表11-5。

参数

用户可参考表11-4和表11-5配置DIS Stream节点的参数。

表 11-4 属性参数


参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
通道名称	是	选择或输入待查询的DIS通道，输入通道名称时支持引用作业参数和使用EL表达式（参见 表达式概述 ）。 如需新建DIS通道，请参考以下方法： <ul style="list-style-type: none">单击，前往数据开发模块的“数据集成 > DIS通道管理”页面新建DIS通道。前往DIS管理控制台进行新建。

表 11-5 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>

参数	是否必选	说明
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">● 终止当前作业执行计划● 继续执行下一节点● 挂起当前作业执行计划● 终止后续节点执行计划

11.1.4 DIS Dump


功能

通过DIS Dump节点配置DIS的数据转储任务。

参数

用户可参考[表11-6](#)和[表11-7](#)配置DIS Dump节点参数。

表 11-6 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
通道名称	是	选择或输入待配置的DIS通道，输入通道名称时支持引用作业参数和使用EL表达式（参见 表达式概述 ）。 如需新建DIS通道，请参考以下方法： <ul style="list-style-type: none">● 单击，前往数据开发模块的“数据集成 > DIS通道管理”页面新建DIS通道。● 前往DIS管理控制台进行新建。
转储任务重名策略	是	选择重名策略。当“转储服务类型”配置的转储任务名称出现重名时，DIS Dump将根据重名策略进行下一步操作。 <ul style="list-style-type: none">● 忽略：不添加转储任务，并退出DIS Dump，DIS Dump的状态为“成功”。● 覆盖：继续添加转储任务，覆盖已存在的重名转储任务。


参数	是否必选	说明
转储服务类型	是	1. 选择转储服务类型，目前支持转储至： <ul style="list-style-type: none"> ● OBS：通道里的流式数据存储在DIS中，并周期性导入对象存储服务OBS；通道里的实时文件数据传输完成后，导入OBS。 2. 单击  ，在弹出的对话框中配置转储任务的参数（参数说明请见《数据接入服务用户指南》）。

表 11-7 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> ● 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> - 最大重试次数 - 重试间隔时间（秒） ● 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> ● 终止当前作业执行计划 ● 继续执行下一节点 ● 挂起当前作业执行计划 ● 终止后续节点执行计划

11.1.5 DIS Client

功能

通过DIS Client节点可以给DIS通道发送消息。

参数

用户可参考[表11-8](#)配置DIS Client节点参数。

表 11-8 属性参数


参数	是否必选	说明
是否使用DIS数据通道连接	否	若使用数据通道连接, 可以向其他账号的DIS通道发送消息; 若不使用, 仅能给本账号下所有region的通道发送消息。
数据通道连接名称	否	选择“管理中心”中已创建的DIS连接名称。
通道所属Region	否	使用DIS Client节点发送消息至目标DIS通道时, 目标通道所在的Region。
节点名称	是	节点名称, 可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符, 长度为1~128个字符。
通道名称	是	需要发送消息的DIS通道。可以直接输入DIS通道地址或单击...选择DIS通道。
发送数据	是	发送到DIS通道的文本内容。可以直接输入文本或单击  使用EL表达式编辑。
相关作业	否	单击...选择相关作业, 您可以选择批作业, 也可以实时作业。 说明 您最多只能选择10个作业。 选择完相关作业, 单击“前往监控”在“作业监控”页面选择DIS Client节点时, 单击页面下方的“查看相关作业”按钮可以查看相关作业。 在“相关作业”页面, 单击“查看”能跳转到对应的作业。

表 11-9 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间, 如果节点配置了重试, 在超时时间内未执行完成, 该节点将不会再重试, 直接置为失败状态。

参数	是否必选	说明
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一节点 挂起当前作业执行计划 终止后续节点执行计划

11.1.6 Rest Client

功能

通过Rest Client节点执行一个华为云内的RESTful请求，目前只支持IAM Token认证鉴权方式的RESTful请求。

参数

用户可参考[表11-10](#)，[表11-11](#)和[表11-12](#)配置Rest Client节点的参数。

表 11-10 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
URL地址	是	填写请求主机的IP或域名地址，以及端口号。例如： https://192.160.10.10:8080
HTTP方法	是	<p>选择请求的类型：</p> <ul style="list-style-type: none"> GET POST PUT DELETE

参数	是否必选	说明
请求头	否	<p>单击 +，添加请求消息头，参数说明如下：</p> <ul style="list-style-type: none"> 参数名称 选择参数的名称，选项为“Content-Type”、“Accept-Language”、“X-Auth-Token”。 参数值 填写参数的值，只支持英文字母、数字、“-”、“_”、“/”、“;”，最大长度为64字符。
URL参数	否	<p>填写URL参数，格式为“参数=值”形式的字符串，字符串间以换行符分隔。当“HTTP方法”为“GET”时，显示该配置项。参数说明如下：</p> <ul style="list-style-type: none"> 参数 只支持英文字母、数字、“-”、“_”，最大长度为32字符。 值 只支持英文字母、数字、“-”、“_”、“\$”、“{”和“}”，最大长度为64字符。
请求消息体	是	填写Json格式的请求消息体。当“HTTP方法”为“POST”、“PUT”时，显示该配置项。
是否需要判断返回值	否	<p>设置是否判断返回消息的值和预期的一致。当“HTTP方法”为“GET”时，显示该配置项。</p> <ul style="list-style-type: none"> YES：检查返回消息中的值是否和预期的一致。 NO：不检查，请求返回200响应码（表示节点执行成功）。
返回值字段路径	是	<p>填写Json响应消息中某个属性的路径（下称：Json属性路径），每个Rest Client节点都只能配置一个属性的路径。当“是否需要判断返回值”为“YES”时，显示该配置项。</p> <p>例如，返回结果为：</p> <pre> { "param1": "aaaa", "inner": { "inner": { "param4": 2014247437 }, "param3": "cccc" }, "status": 200, "param2": "bbbb" } </pre> <p>其中“param4”属性的路径为“inner.inner.param4”。</p>

参数	是否必选	说明
请求成功标志位	是	填写请求成功标志位，如果响应消息的返回值与请求成功标志位中的某一个匹配，表示节点执行成功。当“是否需要判断返回值”为“YES”时，显示该配置项。 请求成功标志位只支持英文字母、数字、“-”、“_”、“\$”、“{”、“}”，多个值使用“;”分隔。
请求失败标志位	否	填写请求失败标志位，如果响应消息的返回值与请求失败标志位中的某一个匹配，表示节点执行失败。当“是否需要判断返回值”为“YES”时，显示该配置项。 请求失败标志位只支持英文字母、数字、“-”、“_”、“\$”、“{”、“}”，多个值使用“;”分隔。
请求间隔时间（秒）	是	如果响应消息的返回值与请求成功标志位不匹配，将每隔一段时间查询一次，直到响应消息的返回值与请求成功标志位一致。节点执行的超时时间默认为1小时，如果1小时内查询的结果始终为不匹配，那么节点的状态将置为失败。当“是否需要判断返回值”为“YES”时，显示该配置项。
响应消息体解析为传递参数定义	否	设置作业变量与Json属性路径的对应关系，参数间以换行符分隔。 例如： <code>var4=inner.inner.param4</code> 其中，“var4”为作业变量，作业变量只支持英文字母、数字，最大长度为16字符；“inner.inner.param4”为Json属性路径。 仅该节点的后续节点引用该参数才会生效，引用该参数时，格式为： <code>\${var4}</code> 。







表 11-11 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> ● 终止当前作业执行计划 ● 继续执行下一节点 ● 挂起当前作业执行计划 ● 终止后续节点执行计划

表 11-12 血缘关系

参数	说明
输入	
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS和DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输入功能的参数配置。
取消	单击“取消”，取消节点输入功能的参数配置。

参数	说明
编辑	单击  ，修改节点输入功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输入功能的参数配置。
查看表详情	单击  ，查看节点输入血缘关系创建数据表的详细信息。
输出	
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS，DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输出功能的参数配置。
取消	单击“取消”，取消节点输出功能的参数配置。
编辑	单击  ，修改节点输出功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输出功能的参数配置。
查看表详情	单击  ，查看节点输出血缘关系创建数据表的详细信息。

11.1.7 Import GES

功能

通过Import GES节点可以将OBS桶中的文件导入到GES的图中。

参数

用户可参考[表11-13](#)和[表11-14](#)配置Import GES节点的参数。

表 11-13 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
图名称	是	选择需要导入的图。 如需新建GES图，请前往GES管理控制台进行新建。
元数据	是	选择对应的元数据。
边数据集	是	选择对应的边数据集。
点数据集	否	选择对应的点数据集。若不选择，则以边数据集中的点作为点数据集来源。
边处理	是	边处理支持如下几种方式： <ul style="list-style-type: none">• 允许重复边• 不允许重复，忽略之后的重复边• 不允许重复，覆盖之前的重复边
离线导入	否	是否离线导入，取值为true或者false，默认取false。 <ul style="list-style-type: none">• 是：表示离线导入，导入速度较快，但导入过程中图处于锁定状态，不可读不可写。• 否：表示在线导入，相对离线导入，在线导入速度略慢，但导入过程中图并未锁定，可读不可写。
重复边忽略Label	否	重复边的定义，是否忽略Label。取值为true或者false，默认取true。 <ul style="list-style-type: none">• true 表示重复边定义不包含Label，即用<源点，终点>标记一条边，不包含Label。• false 表示重复边定义包含Label，即用<源点，终点，Label>标记一条边。
日志存储路径	否	用于存储导入图过程中不符合元数据定义的点、边数据集和详细日志。

表 11-14 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.8 MRS Kafka

功能

MRS Kafka主要是查询Topic未消费的消息数。

参数

用户可参考[表11-15](#)和[表11-16](#)配置MRS Kafka的参数。

表 11-15 属性参数

参数	是否必选	说明
数据连接	是	选择管理中心中已创建的MRS Kafka连接。
Topic名称	是	选择MRS Kafka中已创建的Topic，使用SDK或者命令行创建。

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。

表 11-16 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">终止当前作业执行计划继续执行下一节点挂起当前作业执行计划终止后续节点执行计划

11.1.9 Kafka Client

功能

通过Kafka Client向Kafka的Topic中发送数据。

参数

用户可参考[表11-17](#)配置Kafka Client节点的参数。

表 11-17 属性参数


参数	是否必选	说明
数据连接	是	选择管理中心中已创建的MRS Kafka连接。
Topic名称	是	选择需要上传数据的Topic，如果有多个partition，默认发送到partition 0。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
发送数据	是	发送到Kafka的文本内容。可以直接输入文本或单击  使用EL表达式编辑。

表 11-18 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一节点 挂起当前作业执行计划 终止后续节点执行计划

11.1.10 CS Job

功能

通过CS Job节点执行一个预先定义的CS作业，实现实时流式大数据分析。


背景信息

该节点方便用户启动或者查询CS作业是否正在运行。当选择作业类型不是‘选择已存在的实时作业’时，DLF会根据在节点中配置的作业情况，进行创建和启动作业。方便用户自定义作业以及使用DLF作业参数。

参数

用户可参考[表11-19](#)和[表11-20](#)配置CS Job节点的参数。

表 11-19 属性参数

参数	是否必选	说明
作业类型	是	选择CS的作业类型： <ul style="list-style-type: none"> 选择已存在的实时流作业 Flink SQL作业 Flink自定义作业 Spark自定义作业
选择已存在的实时流作业		
实时流作业名称	是	选择需要执行的CS作业。 如需新建CS作业，请参考以下方法： <ul style="list-style-type: none"> 单击，前往数据开发模块的“数据集成”页面新建CS作业。 前往CS管理控制台进行新建。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
Flink SQL作业		
SQL脚本	是	选择需要执行的脚本。如果脚本未创建，请参考 新建脚本 和 开发SQL脚本 先创建和开发脚本。
脚本参数	是	关联的SQL脚本如果使用了参数，此处显示参数名称，请在参数名称后的输入框配置参数值。参数值支持使用 EL表达式 。
CloudStream集群	是	选择CS集群。如需创建新的CS集群，请前往CS控制台操作。
SPUs	是	选择流处理单元，1SPU=1核4G的资源配置。
并行数	是	选择同时运行CS作业的任务数，建议设置为SPU的1~2倍。
UDF Jar	否	SQL中可调用插入Jar包中的自定义函数。用户需先将Jar包上传至OBS桶中。

参数	是否必选	说明
异常自动启动	否	设置是否启动异常自动重启功能，当CS作业异常时自动重启并恢复CS作业。
实时流作业名称	是	Flink SQL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。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
Flink自定义作业		
jar包路径	是	用户需先上传自定义的Jar包至OBS桶中，在此处才能选择对应Jar包的OBS路径。
入口类	否	指定加载的Jar包入口类名（如：KafkaMessageStreaming）。若不指定，则根据Jar包文件的Manifest文件确定。
入口参数	否	指定入口类的参数列表，参数之间使用空格分隔（如:test tmp/result.txt）。
CloudStream集群	是	选择CS集群。如需创建新的CS集群，请前往CS控制台操作。
SPUs	是	选择流处理单元，1SPU=1核4G的资源配置。
管理单元	是	设置Driver节点使用的SPU数量。
并行数	是	选择同时运行CS作业的任务数，建议设置为SPU的1~2倍。
异常自动启动	否	设置是否启动异常自动重启功能，当CS作业异常时自动重启并恢复CS作业。
实时流作业名称	是	Flink自定义作业名称，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。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
Spark自定义作业		
jar包路径	是	用户需先上传自定义的Jar包至OBS桶中，在此处才能选择对应Jar包的OBS路径。
入口类	否	指定加载的Jar包入口类名（如：KafkaMessageStreaming）。若不指定，则根据Jar包文件的Manifest文件确定。
入口参数	否	指定入口类的参数列表，参数之间使用空格分隔（如:test tmp/result.txt）。

参数	是否必选	说明
CloudStream集群	是	选择CS集群。如需创建新的CS集群，请前往CS控制台操作。
SPUs	是	选择流处理单元，1SPU=1核4G的资源配置。
管理单元	是	设置Driver节点使用的SPU数量。
Executor个数	是	设置Executor节点的数量。
Executor SPUs	是	设置每个Executor节点使用的SPU数量。
异常自动启动	否	设置是否启动异常自动重启功能，当CS作业异常时自动重启并恢复CS作业。
实时流作业名称	是	Spark自定义作业名称，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。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。

表 11-20 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.11 DLI SQL

功能

通过DLI SQL节点传递SQL语句到DLI中执行，实现多数据源分析探索。



原理

该节点方便用户在DLF的周期与实时调度中执行DLI相关语句，可以使用参数变量为用户的数仓进行增量导入，分区处理等动作。

参数

用户可参考[表11-21](#)，[表11-22](#)和[表11-23](#)配置DLI SQL节点的参数。

表 11-21 属性参数

参数	是否必选	说明
SQL或脚本	是	<p>可以选择SQL语句或SQL脚本。</p> <ul style="list-style-type: none"> SQL语句 单击“SQL语句”参数下的文本框，在“SQL语句”页面输入需要执行的SQL语句。 SQL脚本 在“SQL脚本”参数后选择需要执行的脚本。如果脚本未创建，请参考新建脚本和开发SQL脚本先创建和开发脚本。 <p>说明 若选择SQL语句方式，DLF将无法解析您输入SQL语句中携带的参数。</p>
数据库名称	是	默认选择SQL脚本中设置的数据库，支持修改。
队列名称	是	<p>默认选择SQL脚本中设置的DLI队列，支持修改。 如需新建资源队列，请参考以下方法：</p> <ul style="list-style-type: none"> 单击，进入DLI的“队列管理”页面新建资源队列。 前往DLI管理控制台进行新建。
脚本参数	否	<p>关联的SQL脚本如果使用了参数，此处显示参数名称，请在参数名称后的输入框配置参数值。参数值支持使用EL表达式。</p> <p>若关联的SQL脚本，脚本参数发生变化，可单击刷新按钮同步。</p>
节点名称	是	<p>默认显示为SQL脚本的名称，支持修改。规则如下： 节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。</p>




参数	是否必选	说明
是否记录脏数据	是	单击 <input type="radio"/> 选择节点是否记录脏数据。 <ul style="list-style-type: none"> 是：记录脏数据 否：不记录脏数据




表 11-22 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

表 11-23 血缘关系

参数	说明
输入	

参数	说明
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS和DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输入功能的参数配置。
取消	单击“取消”，取消节点输入功能的参数配置。
编辑	单击  ，修改节点输入功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输入功能的参数配置。
查看表详情	单击  ，查看节点输入血缘关系创建数据表的详细信息。
输出	

参数	说明
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS，DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输出功能的参数配置。
取消	单击“取消”，取消节点输出功能的参数配置。
编辑	单击  ，修改节点输出功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输出功能的参数配置。
查看表详情	单击  ，查看节点输出血缘关系创建数据表的详细信息。

11.1.12 DLI Spark

功能

通过DLI Spark节点执行一个预先定义的Spark作业。

参数

用户可参考[表11-24](#)和[表11-25](#)配置DLI Spark节点的参数。

表 11-24 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
DLI队列	是	下拉选择需要使用的队列。
作业名称	是	填写DLI Spark作业的名称，只能包含英文字母、数字、“_”，且长度为1~64个字符。默认与节点的名称一致。
作业运行资源	否	选择作业运行的资源规格： <ul style="list-style-type: none">● 8核32G内存● 16核64G内存● 32核128G内存
作业主类	是	填写Spark作业的主类，即Jar包的主类。
Spark程序资源包	是	用户自定义开发的Spark应用程序的Jar包。在选择资源包之前，您需要先将Jar包及其依赖包上传至OBS桶中，并在“资源管理”页面中新建资源，具体操作请参考 新建资源 。
主类入口参数	否	填写程序的入口参数，参数之间使用Enter键分隔。
Spark作业运行参数	否	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。具体参数请参见 Spark Configuration 。
Module名称	否	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">● CloudTable/MRS HBase: sys.datasource.hbase● CloudTable/MRS OpenTSDB: sys.datasource.opentsdb● RDS MySQL: sys.datasource.rds● RDS PostGre: sys.datasource.rds● DWS: sys.datasource.dws● CSS: sys.datasource.css

表 11-25 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">● 是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">- 最大重试次数- 重试间隔时间（秒）● 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">● 终止当前作业执行计划● 继续执行下一作业● 挂起当前作业执行计划● 终止当前以及后续节点执行计划

11.1.13 DWS SQL

功能

通过DWS SQL节点传递SQL语句到DWS中执行。

背景信息

该节点方便用户在DLF的批处理作业和实时处理作业中执行DWS相关语句，可以使用参数变量为用户的数据仓库进行增量导入，分区处理等操作。

参数

用户可参考[表11-26](#)，[表11-27](#)和[表11-28](#)配置DWS SQL节点的参数。

表 11-26 属性参数


参数	是否必选	说明
SQL或脚本	是	<p>可以选择SQL语句或SQL脚本。</p> <ul style="list-style-type: none"> SQL语句 单击“SQL语句”参数下的文本框，在“SQL语句”页面输入需要执行的SQL语句。 SQL脚本 在“SQL脚本”参数后选择需要执行的脚本。如果脚本未创建，请参考新建脚本和开发SQL脚本先创建和开发脚本。 <p>说明 若选择SQL语句方式，DLF将无法解析您输入SQL语句中携带的参数。</p>
数据连接	是	默认选择SQL脚本中设置的数据连接，支持修改。
数据库	是	默认选择SQL脚本中设置的数据库，支持修改。
脚本参数	否	<p>关联的SQL脚本如果使用了参数，此处显示参数名称，请在参数名称后的输入框配置参数值。参数值支持使用EL表达式。</p> <p>若关联的SQL脚本，脚本参数发生变化，可单击刷新按钮同步。</p>
脏数据表	否	填写SQL脚本中定义的脏数据表名称。
节点名称	是	<p>默认显示为SQL脚本的名称，支持修改。规则如下： 节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。</p>




表 11-27 高级参数




参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> ● 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> - 最大重试次数 - 重试间隔时间（秒） ● 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> ● 终止当前作业执行计划 ● 继续执行下一作业 ● 挂起当前作业执行计划 ● 终止当前以及后续节点执行计划

表 11-28 血缘关系

参数	说明
输入	

参数	说明
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS和DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输入功能的参数配置。
取消	单击“取消”，取消节点输入功能的参数配置。
编辑	单击  ，修改节点输入功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输入功能的参数配置。
查看表详情	单击  ，查看节点输入血缘关系创建数据表的详细信息。
输出	

参数	说明
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS，DLI类型。</p> <ul style="list-style-type: none"> • DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 • OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 • CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 • DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输出功能的参数配置。
取消	单击“取消”，取消节点输出功能的参数配置。
编辑	单击  ，修改节点输出功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输出功能的参数配置。
查看表详情	单击  ，查看节点输出血缘关系创建数据表的详细信息。

11.1.14 MRS SparkSQL

功能

通过MRS Spark SQL节点实现在MRS中执行预先定义的SparkSQL语句。

参数

用户可参考[表11-29](#)和[表11-30](#)配置MRS Spark SQL节点的参数。

表 11-29 属性参数


参数	是否必选	说明
SQL脚本	是	选择需要执行的脚本。如果脚本未创建，请参考 新建脚本 和 开发SQL脚本 先创建和开发脚本。
数据连接	是	默认选择SQL脚本中设置的数据连接，支持修改。
数据库	是	默认选择SQL脚本中设置的数据库，支持修改。
脚本参数	否	关联的SQL脚本如果使用了参数，此处显示参数名称，请在参数名称后的输入框配置参数值。参数值支持使用 EL表达式 。 若关联的SQL脚本，脚本参数发生变化，可单击刷新按钮  同步。
运行程序参数	否	为本次执行的作业配置相关优化参数（例如线程、内存、CPU核数等），用于优化资源使用效率，提升作业的执行性能。 说明 若集群为MRS 1.8.7版本或MRS 2.0.1之后版本，需要配置此参数。 MRS SparkSQL作业的运行程序参数，请参见《MapReduce用户指南》中的“ 运行SparkSql作业 > 表2 运行程序参数”。
节点名称	是	默认显示为SQL脚本的名称，支持修改。规则如下： 节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。

表 11-30 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.15 MRS Hive SQL


功能

通过MRS Hive SQL节点执行数据开发模块中预先定义的Hive SQL脚本。

参数

用户可参考[表11-31](#)和[表11-32](#)配置MRS Hive SQL节点的参数。

表 11-31 属性参数

参数	是否必选	说明
SQL脚本	是	选择需要执行的脚本。如果脚本未创建，请参考 新建脚本 和 开发SQL脚本 先创建和开发脚本。
数据连接	是	默认选择SQL脚本中设置的数据连接，支持修改。
数据库	是	默认选择SQL脚本中设置的数据库，支持修改。
脚本参数	否	<p>关联的SQL脚本如果使用了参数，此处显示参数名称，请在参数名称后的输入框配置参数值。参数值支持使用EL表达式。</p> <p>若关联的SQL脚本，脚本参数发生变化，可单击刷新按钮同步。</p>

参数	是否必选	说明
运行程序参数	否	<p>为本次执行的作业配置相关优化参数（例如线程、内存、CPU核数等），用于优化资源使用效率，提升作业的执行性能。</p> <p>说明 若集群为MRS 1.8.7版本或MRS 2.0.1之后版本，需要配置此参数。</p> <p>MRS Hive SQL作业的运行程序参数，请参见《MapReduce用户指南》中的“运行HiveSql作业 > 表2运行程序参数”。</p>
节点名称	是	<p>默认显示为SQL脚本的名称，支持修改。规则如下： 节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。</p>

表 11-32 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.16 MRS Presto SQL

功能

通过MRS Presto SQL执行数据开发模块中预先定义的Presto SQL脚本。

参数

用户可参考[表11-33](#)和[表11-34](#)配置MRS Presto SQL的参数。

表 11-33 属性参数


参数	是否必选	说明
SQL脚本	是	选择需要执行的脚本。如果脚本未创建，请参考 新建脚本 和 开发SQL脚本 先创建和开发脚本。
数据连接	是	默认选择SQL脚本中设置的数据连接，支持修改。
模式	是	默认选择SQL脚本中设置的数据库，支持修改。
脚本参数	否	关联的SQL脚本如果使用了参数，此处显示参数名称，请在参数名称后的输入框配置参数值。参数值支持使用 EL表达式 。 若关联的SQL脚本，脚本参数发生变化，可单击刷新按钮  同步。
节点名称	是	默认显示为SQL脚本的名称，支持修改。规则如下： 节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。

表 11-34 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.17 MRS Spark


功能

通过MRS Spark节点实现在MRS中执行预先定义的Spark作业。

参数

用户可参考[表11-35](#)和[表11-36](#)配置MRS Spark节点的参数。

表 11-35 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
MRS集群名	是	<p>选择MRS集群。</p> <p>如需新建集群，请参考以下方法：</p> <ul style="list-style-type: none"> 单击，进入“集群列表”页面新建MRS集群。 前往MRS管理控制台进行新建。
Spark作业名称	是	MRS作业名称，只能包含英文字母、数字、“_”，且长度为1~64个字符。

参数	是否必选	说明
Jar包资源	是	选择Jar包。在选择Jar包之前，您需要先将Jar包上传至OBS桶中，并在“资源管理”页面中新建资源将Jar包添加到资源管理列表中，具体操作请参考 新建资源 。
Jar包参数	否	Jar包的参数。
运行程序参数	否	为本次执行的作业配置相关优化参数（例如线程、内存、CPU核数等），用于优化资源使用效率，提升作业的执行性能。 说明 若集群为MRS 1.8.7版本或MRS 2.0.1之后版本，需要配置此参数。 MRS Spark作业的运行程序参数，请参见《MapReduce用户指南》中的 运行Spark作业 。
输入数据路径	否	选择输入数据所在的路径。
输出数据路径	否	选择输出数据存储的路径。

表 11-36 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.18 MRS Spark Python

功能

通过MRS Spark Python节点实现在MRS中执行预先定义的Spark Python作业。

参数

用户可参考[表11-37](#)和[表11-38](#)配置MRS Spark Python节点的参数。

表 11-37 属性参数


参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
MRS集群名	是	选择支持spark python的mrs集群。MRS只有特定版本支持spark python的集群，请先测试运行，保证集群支持。 如需新建集群，请参考以下方法： <ul style="list-style-type: none">• 单击，进入“集群列表”页面新建MRS集群。• 前往MRS管理控制台进行新建。 如何新建集群，请参见《MapReduce服务用户指南》的 创建集群 。
作业名称	是	输入MRS Spark Python作业名称，只能包含英文字母、数字、“_”，且长度为1~64个字符。
参数	是	输入MRS的执行程序参数，多个参数间使用Enter键分隔。
属性	否	输入key=value格式的的参数，多个参数间使用Enter键分割。

表 11-38 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一节点 挂起当前作业执行计划 终止后续节点执行计划

11.1.19 MRS Flink Job


功能

通过MRS Flink节点实现在MRS中执行预先定义的Flink作业。

参数

用户可参考[表11-39](#)和[表11-40](#)配置MRS Flink节点的参数。

表 11-39 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
MRS集群名	是	<p>选择MRS集群。</p> <p>如需新建集群，请参考以下方法：</p> <ul style="list-style-type: none"> 单击，进入“集群列表”页面新建MRS集群。 前往MRS管理控制台进行新建。
Flink作业名称	是	MRS作业名称，只能包含英文字母、数字、“_”，且长度为1~64个字符。

参数	是否必选	说明
Flink作业资源包	是	选择Jar包。在选择Jar包之前，您需要先将Jar包上传至OBS桶中，并在“资源管理”页面中新建资源将Jar包添加到资源管理列表中，具体操作请参考 新建资源 。
Flink作业执行参数	否	Flink作业执行的程序关键参数，该参数由用户程序内的函数指定。多个参数间使用空格隔开。
运行程序参数	否	为本次执行的作业配置相关优化参数（例如线程、内存、CPU核数等），用于优化资源使用效率，提升作业的执行性能。 说明 若集群为MRS 1.8.7版本或MRS 2.0.1之后版本，需要配置此参数。 MRS Spark作业的运行程序参数，请参见《MapReduce用户指南》中的 运行Spark作业 。
输入数据路径	否	选择输入数据所在的路径。
输出数据路径	否	选择输出数据存储的路径。

表 11-40 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.20 MRS MapReduce

功能

通过MRS MapReduce节点实现在MRS中执行预先定义的MapReduce程序。

参数

用户可参考[表11-41](#)和[表11-42](#)配置MRS MapReduce节点参数。

表 11-41 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
MRS集群名	是	选择MRS集群。 如需新建集群，请参考以下方法： <ul style="list-style-type: none">单击，进入“集群列表”页面新建MRS集群。前往MRS管理控制台进行新建。
MapReduce作业名称	是	MRS作业名称，只能包含英文字母、数字、“_”，且长度为1~64个字符。
Jar包资源	是	选择Jar包。在选择Jar包之前，您需要先将Jar包上传至OBS桶中，并在“资源管理”页面中新建资源将Jar包添加到资源管理列表中，具体操作请参考 新建资源 。
Jar包参数	否	Jar包的参数。
输入数据路径	否	选择输入数据所在的路径。
输出数据路径	否	选择输出数据存储的路径。

表 11-42 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.21 CSS

功能

通过CSS节点执行云搜索请求，实现在线分布式搜索功能。

参数

用户可参考[表11-43](#)和[表11-44](#)配置CSS节点的参数。

表 11-43 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
CloudSearch集群	是	选择CloudSearch集群，该集群已在CloudSearch服务中创建好。目前仅支持使用5.5.1版本的集群。
CDM集群名称	是	<p>选择CDM集群，CDM集群提供的Agent为CSS的运行环境。</p> <p>如果下拉框中未提供CDM集群，请访问CDM管理控制台创建集群。</p>

参数	是否必选	说明
请求类型	是	支持以下请求类型： <ul style="list-style-type: none"> • GET • POST • PUT • HEAD • DELETE
请求参数	否	请求参数。 假设用户需要查询dlf_search索引中dlfdata映射类型的信息，请求参数可填写为： /dlf_search/dlfdata/_search
请求体	否	Json格式的请求消息体。
CloudSearch输出路径	否	选择输出数据的存储路径。

表 11-44 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> • 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> - 最大重试次数 - 重试间隔时间（秒） • 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> • 终止当前作业执行计划 • 继续执行下一作业 • 挂起当前作业执行计划 • 终止当前以及后续节点执行计划

11.1.22 Shell

功能

通过Shell节点执行用户指定的Shell脚本。

说明

Shell节点的后续节点可以通过EL表达式`#{Job.getNodeOutput()}`，获取Shell脚本最后4000字符的标准输出。

使用示例：

获取某个Shell脚本（脚本名称为shell_job1）输出值包含“<name>jack<name1>”的内容，EL表达式如下所示：

```
#{StringUtil.substringBetween(Job.getNodeOutput("shell_job1"),"<name>","<name1>")}
```

参数

用户可以参考[表11-45](#)和[表11-46](#)配置Shell节点的参数。

表 11-45 属性参数

参数	是否必选	说明
SQL或脚本	是	可以选择SQL语句或SQL脚本。 <ul style="list-style-type: none"> SQL语句 单击“SQL语句”参数下的文本框，在“SQL语句”页面输入需要执行的SQL语句。 SQL脚本 在“脚本路径”参数后选择需要执行的脚本。如果脚本未创建，请参考新建脚本和开发Shell脚本先创建和开发脚本。 说明 若选择SQL语句方式，DLF将无法解析您输入SQL语句中携带的参数。
主机连接	是	选择执行Shell脚本的主机。
参数	否	填写执行Shell脚本时，向脚本传递的参数，参数之间使用空格分隔，例如：a b c。此处的“参数”需要在Shell脚本中引用，否则配置无效。
交互式输入	否	填写交互式参数，即执行Shell脚本的过程中，需要用户输入的交互式信息（例如密码）。交互式参数之间以回车符分隔，Shell脚本根据交互情况按顺序读取参数值。
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。

表 11-46 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.23 RDS SQL

功能

通过RDS SQL节点传递SQL语句到RDS中执行。

参数

用户可参考[表11-47](#)和[表11-48](#)配置RDS SQL节点的参数。

表 11-47 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
数据连接	是	选择数据连接。

参数	是否必选	说明
数据库	是	填写数据库名称，该数据库已创建好，建议不要使用默认数据库。
SQL或脚本	是	<p>可以选择SQL语句或SQL脚本。</p> <ul style="list-style-type: none"> SQL语句 单击“SQL语句”参数下的文本框，在“SQL语句”页面输入需要执行的SQL语句。 SQL脚本 在“脚本路径”参数后选择需要执行的脚本。如果脚本未创建，请参考新建脚本和开发SQL脚本先创建和开发脚本。 <p>说明 若选择SQL语句方式，DLF将无法解析您输入SQL语句中携带的参数。</p>

表 11-48 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.24 ETL Job

功能

通过ETL Job节点可以从指定数据源中抽取数据，经过数据准备对数据预处理后，导入到目标数据源。

参数

用户可参考[表11-49](#)和[表11-50](#)配置ETL Job节点的参数。

表 11-49 属性参数


参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
ETL配置	是	<p>单击  配置需要转换的源端数据和目的端数据。</p> <p>当前支持的源端数据为DLI类型、OBS类型和MySQL类型。</p> <ul style="list-style-type: none">当源端数据为DLI类型时，支持的目的端数据类型为DWS、GES、CSS、OBS、DLI。当源端数据为MySQL类型时，支持的目的端数据类型为MySQL。当源端数据为OBS类型时，支持的目的端数据类型为DLI、DWS。 <p>须知</p> <ul style="list-style-type: none">DLI到DWS端的数据转换： 因为DLF调用DWS的集群时，需要走网络代理。所以导入数据到DWS时，需要提前先在DLF中创建DWS的数据连接。 DLI导入数据到DWS时，DWS的表需要先创建好。DLI到CSS端的数据转换： DLI导入数据到CSS集群时，需要在DLI侧提前创建好关联对应CSS集群的跨源连接，请参见《数据湖探索用户指南》。
SQL模板	否	单击“配置”按钮获取SQL模板。




表 11-50 高级参数




参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">● 是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">- 最大重试次数- 重试间隔时间（秒）● 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">● 终止当前作业执行计划● 继续执行下一节点● 挂起当前作业执行计划● 终止后续节点执行计划

表 11-51 血缘关系

参数	说明
输入	

参数	说明
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS和DLI类型。</p> <ul style="list-style-type: none"> ● DWS <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。 - schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。 ● OBS <ul style="list-style-type: none"> - 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。 ● CSS <ul style="list-style-type: none"> - 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。 - 索引名称（必选）：输入CSS类型的索引名称。 ● DLI <ul style="list-style-type: none"> - 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。 - 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。 - 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输入功能的参数配置。
取消	单击“取消”，取消节点输入功能的参数配置。
编辑	单击  ，修改节点输入功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输入功能的参数配置。
查看表详情	单击  ，查看节点输入血缘关系创建数据表的详细信息。
输出	

参数	说明
新建	<p>单击“新建”，在“类型”的下拉选项中选择要新建的类型。可以选择DWS，OBS，CSS，DLI类型。</p> <ul style="list-style-type: none">• DWS<ul style="list-style-type: none">- 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DWS的数据连接。- 数据库（必选）：单击...，在弹出的“数据库”窗口选择DWS的数据库。- schema（必选）：单击...，在弹出的“schema”窗口选择DWS的数据库模式。- 表名（必选）：单击...，在弹出的“表名”窗口选择DWS的数据表。• OBS<ul style="list-style-type: none">- 路径（必选）：单击...，在弹出的“OBS文件浏览”窗口选择OBS路径。• CSS<ul style="list-style-type: none">- 集群名称（必选）：单击...，在弹出的“CloudSearch集群”窗口选择CloudSearch集群。- 索引名称（必选）：输入CSS类型的索引名称。• DLI<ul style="list-style-type: none">- 连接名称（必选）：单击...，在弹出的“连接名称”窗口选择DLI的数据连接。- 数据库（必选）：单击...，在弹出的“数据库”窗口选择DLI的数据库。- 表名（必选）：单击...，在弹出的“表名”窗口选择DLI的数据表。
确认	单击“确认”，保存节点输出功能的参数配置。
取消	单击“取消”，取消节点输出功能的参数配置。
编辑	单击  ，修改节点输出功能的参数配置，修改完成后，请保存。
删除	单击  ，删除节点输出功能的参数配置。
查看表详情	单击  ，查看节点输出血缘关系创建数据表的详细信息。

11.1.25 OCR

功能

OCR节点支持识别图片上的文字信息，识别结果将存储到用户指定的OBS目录下。

 说明

OCR节点仅支持在华北-北京四使用。

参数

用户可参考[表11-52](#)和[表11-53](#)配置OCR节点的参数。

表 11-52 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
图片路径	是	需要识别的图片在OBS上的路径。
识别类型	是	可识别的图片类型，当前仅支持通用文字识别。可支持扫描文件、电子文档、书籍、票据和表单等多种场景的文字识别。
识别结果存储目录	是	指定识别结果存储的OBS目录。

表 11-53 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">终止当前作业执行计划继续执行下一节点挂起当前作业执行计划终止后续节点执行计划

11.1.26 Create OBS

功能

通过Create OBS节点在OBS服务中创建桶和目录。

参数

用户可参考[表11-54](#)和[表11-55](#)配置Create OBS节点的参数。

表 11-54 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
OBS路径	是	创建OBS桶或目录的路径。 <ul style="list-style-type: none">创建桶：在“//”后输入OBS桶名称，OBS桶名称不允许重名。创建OBS目录：选择需要创建目录的路径，在路径后输入“/目录名”，目录名不允许重名。

表 11-55 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">终止当前作业执行计划继续执行下一节点挂起当前作业执行计划终止后续节点执行计划

11.1.27 Delete OBS

功能

通过Delete OBS节点在OBS服务中删除桶和目录。

参数

用户可参考[表11-56](#)和[表11-57](#)配置Delete OBS节点的参数。

表 11-56 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
OBS路径	是	删除OBS桶或目录的路径。 说明 删除的文件将无法恢复，如需保留文件，请在删除前备份该桶下的数据。

表 11-57 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">终止当前作业执行计划继续执行下一节点挂起当前作业执行计划终止后续节点执行计划

11.1.28 OBS Manager

功能

通过OBS Manager节点可以将OBS文件移动或复制到指定目录下。

参数

用户可参考[表11-58](#)和[表11-59](#)配置OBS Manager节点的参数。

表 11-58 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
操作类型	是	通过节点可以执行的操作： <ul style="list-style-type: none"> • 移动文件 • 复制文件
源文件	是	OBS桶中需要被移动或复制的OBS文件。
目的目录	是	存放待移动或复制OBS文件的新目录
文件过滤器	否	输入文件过滤的通配符，满足该过滤条件的文件才会被移动或复制。当不指定该参数时，默认移动或复制所有源文件。例如：匹配文件名以.csv结尾的文件，输入通配符*.csv。

表 11-59 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> • 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> - 最大重试次数 - 重试间隔时间（秒） • 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>

参数	是否必选	说明
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">• 终止当前作业执行计划• 继续执行下一节点• 挂起当前作业执行计划• 终止后续节点执行计划

11.1.29 Open/Close Resource

功能

通过Open/Close Resource节点按需开启或关闭华为云服务。

参数

用户可参考[表11-60](#)和[表11-61](#)配置Open/Close Resource节点的参数。

表 11-60 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
服务	是	选择需要开机/关机的服务： <ul style="list-style-type: none">• ECS• CDM• CS
开关机设置	是	选择开关机类型： <ul style="list-style-type: none">• 开• 关
开关机对象	是	选择需要开机/关机的具体对象，例如开启某个CDM集群。

表 11-61 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.30 Data Quality Monitor

功能

通过Data Quality Monitor节点可以对运行的数据进行质量监控。

参数

用户可参考[表11-62](#)和[表11-63](#)配置Data Quality Monitor节点的参数。

表 11-62 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
数据质量规则名称	是	选择对运行的数据进行质量监控的规则名称。如何创建规则，请参见《智能数据湖运营平台用户指南》。

表 11-63 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一节点 挂起当前作业执行计划 终止后续节点执行计划

11.1.31 Sub Job

功能

通过Sub Job节点可以调用另外一个批处理作业。

参数

用户可参考[表11-64](#)和[表11-65](#)配置Sub Job节点的参数。

表 11-64 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
子作业名称	是	选择需要调用的子作业名称。 说明 您只能选择已存在的批处理作业名称，此批处理作业不能为作业本身，并且该批处理作业为不包含Sub Job节点的作业。

参数	是否必选	说明
子作业参数名称	是/否	<ul style="list-style-type: none"> 当节点属性中子作业参数配置为空时，子作业使用自身参数变量执行。父作业的“子作业参数名称”不显现。 当节点属性中子作业参数配置了数据时，子作业将使用配置参数变量执行。此时父作业的“子作业参数名称”显现，并且节点属性中子作业参数配置的数据或者EL表达式，将根据父作业的环境变量读取替换。

表 11-65 高级参数

参数	是否必选	说明
节点状态轮询时间（秒）	是	设置轮询时间（1~60秒），每隔x秒查询一次节点是否执行完成。
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> 最大重试次数 重试间隔时间（秒） 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	<p>节点执行失败后的操作：</p> <ul style="list-style-type: none"> 终止当前作业执行计划 继续执行下一作业 挂起当前作业执行计划 终止当前以及后续节点执行计划

11.1.32 SMN

功能

通过SMN节点向用户发送通知消息。

参数

用户可参考[表11-66](#)和[表11-67](#)配置SMN节点的参数。

表 11-66 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
主题名称	是	选择消息的主题，该主题已在SMN服务中创建好。
消息标题	否	自定义消息的标题，长度必须少于512个字符。
消息类型	是	选择消息的发送格式。 <ul style="list-style-type: none">● 文本消息：按文本格式发送的消息。● JSON消息：按JSON格式发送的消息，用户可对不同的订阅者类型发送不同的消息。<ul style="list-style-type: none">- 手动输入JSON格式的消息：在“消息内容”直接输入。- 通过工具自动生成JSON格式的消息：单击“生成JSON消息”，在弹出的对话框中填写“消息”和选择“协议”。● 模板消息：按模板格式发送的消息，即固定格式的消息，可以通过tag的方式来处理变量的部分。<ul style="list-style-type: none">- 手动输入模板格式的消息：在“消息内容”直接输入。- 通过工具自动生成模板格式的消息：单击“生成模板消息”，在弹出的对话框中，选择“模板名称”，并设置{tag}的值。

参数	是否必选	说明
消息内容	是	<p>填写消息的内容，不同消息类型的填写要求如下：</p> <ul style="list-style-type: none"> • 文本消息：大小不超过10KB。 • JSON消息：JSON消息中必须有Default协议，大小不超过10KB。 示例如下： <pre>{ "default": "Dear Sir or Madam, this is a default message.", "email": "Dear Sir or Madam, this is an email message.", "http": "{message:'Dear Sir or Madam, this is an HTTP message.'}", "https": "{message:'Dear Sir or Madam, this is an HTTPS message.'}", "sms": "This is an SMS message." }</pre> <ul style="list-style-type: none"> • 模板消息：大小不超过10KB。 示例如下： <pre>"message_template_name":"confirm_message", "tags":{" "topic_urn":"urn:smn:regionId:xxxx:SMN_01" }</pre> <p>其中，“message_template_name”为模板名称，“tags”为模板中所有的tag标签。</p> <p>如需了解更多SMN的配置说明，请参见《消息通知服务用户指南》。</p>

表 11-67 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。
失败重试	是	<p>节点执行失败后，是否重新执行节点。</p> <ul style="list-style-type: none"> • 是：重新执行节点，请配置以下参数。 <ul style="list-style-type: none"> - 最大重试次数 - 重试间隔时间（秒） • 否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>

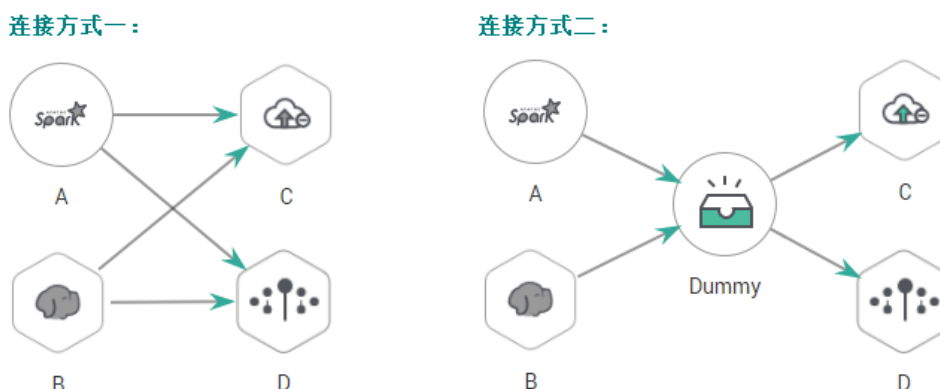
参数	是否必选	说明
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none"> • 终止当前作业执行计划 • 继续执行下一节点 • 挂起当前作业执行计划 • 终止后续节点执行计划

11.1.33 Dummy

功能

Dummy节点是一个空的节点，不执行任何操作。用于简化节点的连接视图，便于用户理解复杂节点流的连接关系。

图 11-2 连接方式对比



参数

用户可参考[表11-68](#)配置Dummy节点的参数。

表 11-68 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。

11.1.34 For Each

功能

该节点可以指定一个子作业循环执行，并支持用一个数据集对子作业中的变量进行循环替换。

参数

用户可参考[表11-69](#)配置For Each节点的参数。

表 11-69 属性参数

参数	是否必选	说明
节点名称	是	节点名称，可以包含中文、英文字母、数字、“_”、“-”、“/”、“<”、“>”等各类特殊字符，长度为1~128个字符。
循环执行的子作业	是	选择需要循环执行的子作业。
数据集	是	For循环算子需要定义一个数据集，这个数据集用来循环替换子作业中的变量。数据集的来源： <ul style="list-style-type: none">来自于上游节点的输出，比如DLI SQL的select语句，或者Shell节点的echo等。当前只支持Hive SQL、Spark SQL 和DLI SQL的select语句来自于一个CSV格式的文件。
子作业并发数	是	循环产生的子作业可以并发执行，您可设置并发数。
子作业后缀	否	For循环生成的子任务名称：For循环节点名称 + 下划线 + 后缀。 后缀可配置，如果不配置，则按照数字顺序依次递增。
作业运行参数	否	<ul style="list-style-type: none">节点属性中子作业参数配置为空时，子作业使用自身参数变量执行。节点属性中子作业参数配置后，将使用配置参数变量执行。节点属性中子作业参数配置的方法或者EL表达式，将根据父作业的环境变量读取替换。

表 11-70 高级参数

参数	是否必选	说明
节点执行的最长时间	是	设置节点执行的超时时间，如果节点配置了重试，在超时时间内未执行完成，该节点将不会再重试，直接置为失败状态。

参数	是否必选	说明
失败重试	是	节点执行失败后，是否重新执行节点。 <ul style="list-style-type: none">是：重新执行节点，请配置以下参数。<ul style="list-style-type: none">最大重试次数重试间隔时间（秒）否：默认值，不重新执行节点。 <p>说明 如果作业节点配置了重试，并且配置了超时时间，该节点执行超时后将不会再重试，直接置为失败状态。</p>
失败策略	是	节点执行失败后的操作： <ul style="list-style-type: none">终止当前作业执行计划继续执行下一节点挂起当前作业执行计划终止后续节点执行计划

11.2 表达式

11.2.1 表达式概述

数据开发模块作业中的节点参数可以使用表达式语言（Expression Language，简称 EL），根据运行环境动态生成参数值。数据开发模块 EL 表达式使用简单的算术和逻辑计算，引用内嵌对象，包括作业对象和一些工具类对象。

作业对象：提供了获取作业中上一个节点的输出消息、作业调度计划时间、作业执行时间等属性和方法。

工具类对象：提供了一系列字符串、时间、JSON 操作方法，例如从一个字符串中截取一个子字符串、时间格式化等。

语法

表达式的语法：

```
#{expr}
```

其中，“expr”指的是表达式。“#”和“{}”是数据开发模块 EL 中通用的操作符，这两个操作符允许您通过数据开发模块内嵌对象访问作业属性。

举例

在 Rest Client 节点的参数“URL 参数”中使用 EL 表达式

“tableName=#{JSONUtil.path(Job.getNodeOutput("get_cluster"),"tables[0].table_name")}”，如图 11-3 所示。

表达式说明如下：

1. 获取作业中“get_cluster”节点的执行结果（“Job.getNodeOutput("get_cluster)"），执行结果是一个JSON字符串。
2. 通过JSON路径（“tables[0].table_name”），获取JSON字符串中字段的值。

图 11-3 表达式示例



11.2.2 基础操作符

EL表达式支持大部分Java提供的算术和逻辑操作符。

操作符列表

表 11-71 基础操作符

操作符	描述
.	访问一个Bean属性或者一个映射条目
[]	访问一个数组或者链表的元素
()	组织一个子表达式以改变优先级
+	加
-	减或负
*	乘
/ 或 div	除
% 或 mod	取模
== 或 eq	测试是否相等
!= 或 ne	测试是否不等
< 或 lt	测试是否小于

操作符	描述
> 或 gt	测试是否大于
<= 或 le	测试是否小于等于
>= 或 ge	测试是否大于等于
&& 或 and	测试逻辑与
或 or	测试逻辑或
! 或 not	测试取反
empty	测试是否空值
?:	类似if else表示式。如果?前面的语句为true, 返回?和:之间的表达式的值; 否则返回:后面的值。

举例

如果变量a为空, 返回default, 否则返回a本身。EL表达式如下:

```
#{empty a?"default":a}
```

11.2.3 日期和时间模式

EL表达式中的日期和时间可以按用户指定的格式进行显示, 日期和时间格式由日期和时间模式字符串指定。日期和时间模式字符串由A到Z、a到z的非引号字母组成, 字母的含义如表11-72所示。

表 11-72 字母含义

字母	描述	示例
G	纪元标记	AD
y	年	2001
M	年中的月份	July 或 07
d	月份中的日期	10
h	12小时制 (1~12) 的小时	12
H	24小时制 (0~23) 的小时	22
m	分钟数	30
s	秒数	55
S	毫秒数	234
E	星期几	Tuesday
D	年中的日期	360

字母	描述	示例
F	月份中第几周周几	2(second Wed. in July)
w	年中的第几周	40
W	月份中的第几周	1
a	A.M./P.M.标记	PM
k	24小时制（1~24）的小时	24
K	12小时制（0~11）的小时	10
z	时区	Eastern Standard Time
'	文字定界符	无示例
"	单引号	无示例

11.2.4 Env 内嵌对象

Env内嵌对象提供了获取环境变量值的方法。

方法

表 11-73 方法说明

方法	描述
String get(String name)	获取指定名称环境变量值。

举例

获取环境变量名称为**test**的参数值，EL表达式如下：

```
#{Env.get("test")}
```

11.2.5 Job 内嵌对象

Job为作业对象，提供了获取作业中上一节点的输出消息、作业调度计划时间、作业执行时间等属性和方法。

属性和方法

表 11-74 属性说明

属性	类型	描述
name	String	作业名称。

属性	类型	描述
planTime	java.util.Date	作业调度计划时间，即周期调度配置的时间，例如每天凌晨1:01调度作业。
startTime	java.util.Date	作业执行时间，有可能与planTime同一个时间，也有可能晚于planTime（由于作业引擎繁忙等）。
eventData	String	当作业使用事件驱动调度时，从DIS通道获取的消息。
projectId	String	当前DLF所处项目ID。

表 11-75 方法说明

方法	描述
String getNodeStatus(String nodeName)	获取指定节点运行状态，成功状态返回success，失败状态返回fail。 例如，判断节点是否运行成功，可以使用如下判断条件： <code>#{(Job.getNodeStatus("test")) == "success" }</code>
String getNodeOutput(String nodeName)	获取指定节点的输出。此方法只能获取前面依赖节点的输出。目前只支持获取Rest Client节点的输出，即REST请求的返回消息。
String getParam(String key)	获取作业参数。
String getPlanTime(String pattern)	获取指定pattern的计划时间字符串，pattern为日期、时间模式，请参考 日期和时间模式 。
String getYesterday(String pattern)	获取执行pattern的计划时间前一天的时间字符串，pattern为日期、时间模式，请参考 日期和时间模式 。
String getLastHour(String pattern)	获取执行pattern的计划时间前一小时的时间字符串，pattern为日期、时间模式，请参考 日期和时间模式 。
String getRunningData(String nodeName)	获取指定节点运行中记录的数据。此方法只能获取前面依赖节点的输出。当前只支持获取DLI SQL节点运行中记录的DLI作业id，例如，想要获取DLI节点第3条语句的job ID（DLI节点名为DLI_INSERT_DATA），可以这样使用： <code>#{JSONUtil.path(Job.getRunningData("DLI_INSERT_DATA"),"jobIds[2]")}</code> 。
String getInsertJobId(String nodeName)	返回指定DLI SQL或Transform Load节点第一个DLI Insert SQL语句的作业ID，不指定参数nodeName时，获取前面一个节点第一个DLI Insert SQL语句的作业ID，如果无法获取到作业ID，返回null值。

举例

获取作业中节点名称为**test**的输出，EL表达式如下：

```
#{Job.getNodeOutput("test")}
```

11.2.6 StringUtil 内嵌对象

StringUtil内嵌对象提供了一系列字符串操作方法，例如从一个字符串中截取一个子字符串。

StringUtil内部是由org.apache.commons.lang3.StringUtils实现的，具体使用方法请参考[apache commons文档](#)。

举例

假设变量a为字符串No.0010，返回.后面的子字符串，EL表达式如下：

```
#{StringUtil.substringAfter(a,".")}
```

11.2.7 DateUtil 内嵌对象

DateUtil内嵌对象提供了一系列时间格式化、时间计算方法。

方法

表 11-76 方法说明

方法	描述
String format(Date date, String pattern)	将Date类型时间按指定pattern格式为字符串。
Date addMonths(Date date, int amount)	给date添加指定月数后，返回新Date对象，amount可以是负数。
Date addDays(Date date, int amount)	给date添加指定天数后，返回新Date对象，amount可以是负数。
Date addHours(Date date, int amount)	给date添加指定小时数后，返回新Date对象，amount可以是负数。
Date addMinutes(Date date, int amount)	给date添加指定分钟数后，返回新Date对象，amount可以是负数。
int getDay(Date date)	从date获取天，例如：date为2018-09-14，则返回14。
int getMonth(Date date)	从date获取月，例如：date为2018-09-14，则返回9。
int getYear(Date date)	从date获取年，例如：date为2018-09-14，则返回2018。
Date now()	返回当前时间。
long getTime(Date date)	将Date类型时间转换为long类型。

方法	描述
Date parseDate(String str, String pattern)	字符串按pattern转换为Date类型，pattern为日期、时间模式，请参考 日期和时间模式 。

举例

以作业调度计划时间的前一天时间作为子目录名称，生成一个OBS路径，EL表达式如下：

```
#{"obs://test/"+DateUtil.format(DateUtil.addDays(Job.planTime,-1),"yyyy-MM-dd")}
```

11.2.8 JSONUtil 内嵌对象

JSONUtil内嵌对象提供了JSON对象方法。

方法

表 11-77 方法说明

方法	描述
Object parse(String jsonStr)	将json字符串转换为对象。
String toString(Object jsonObject)	将对象转换为json字符串。
Object path(String jsonStr,String jsonPath)	返回json字符串指定路径下的字段值。类似于XPath，path方法可以通过路径检索或设置JSON，其路径中可以使用.或[]等访问成员、数值，例如：tables[0].table_name。

举例

字符串变量str的内容如下：

```
{
  "cities": [{
    "name": "Shenzhen",
    "areaCode": "0755"
  },
  {
    "name": "city2",
    "areaCode": "010"
  },
  {
    "name": "Shanghai",
    "areaCode": "021"
  }
]}
```

获取深圳的电话区号，EL表达式如下：

```
#[JSONUtil.path(str,"cities[0].areaCode")]
```


11.2.9 Loop 内嵌对象

使用Loop内嵌对象可获取for each数据集中的数据。

属性

表 11-78 属性说明

属性	类型	描述
dataArray	String	For循环算子输入的数据集，是一个二维数组。
current	String	For循环算子当前遍历到的数据行,是一个一维数组。
offset	Int	For循环当前的偏移量，从0开始。 Loop.dataArray[Loop.offset] = Loop.current。

11.2.10 表达式使用示例

通过本示例，用户可以了解数据开发模块 EL表达式的如下应用：

- 如何在数据开发模块的SQL脚本中使用变量？
- 作业如何传递参数给SQL脚本变量？
- 在参数中如何使用EL表达式？

背景信息

使用数据开发模块的作业编排和作业调度功能，每日通过统计交易明细表，生成日交易统计报表。

本示例涉及的数据表如下所示：

- trade_log：记录每一笔交易数据。
- trade_report：根据trade_log统计产生，记录每日交易汇总。

前提条件

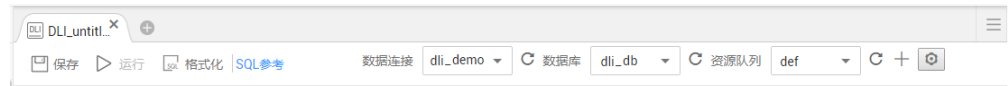
- 已建立DLI的数据连接，以“dli_demo”数据连接为例。
如未建立，请参考[创建数据连接](#)进行操作。
- 已在DLI中创建数据库，以“dli_db”数据库为例。
如未创建，请参考[新建数据库](#)进行操作。
- 已在“dli_db”数据库中创建数据表trade_log和trade_report。
如未创建，请参考[新建数据表（可视化模式）](#)或[新建数据表（DDL模式）](#)进行操作。

操作步骤

步骤1 新建和开发SQL脚本。


1. 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 脚本开发”。
2. 进入右侧区域页面，选择“新建SQL脚本 > DLI”。
3. 进入SQL脚本开发页面，在脚本属性栏选择“数据连接”、“数据库”、“资源队列”。

图 11-4 脚本属性



4. 在脚本编辑器中输入以下SQL语句。

```
INSERT OVERWRITE TABLE trade_report
SELECT
  sum(trade_count),
  '${yesterday}'
FROM
  trade_log
where
  date_format(trade_time, 'yyyy-MM-dd') = '${yesterday}'
```

5. 单击 ，将脚本的名称设置为“generate_trade_report”。

步骤2 新建和开发作业。

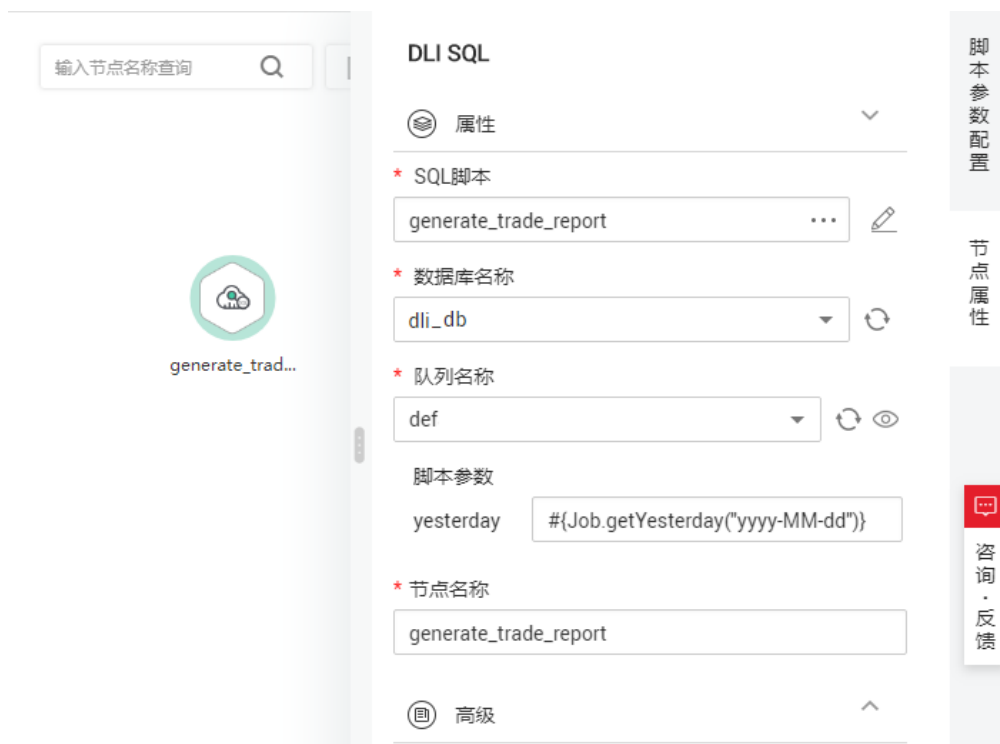
1. 在数据开发模块控制台的左侧导航栏，选择“数据开发 > 作业开发”。
2. 进入右侧区域页面，单击“新建作业”，新建一个名称为“job”的空作业。

图 11-5 新建作业 02



3. 进入作业开发页面，将DLI SQL节点拖至画布中，单击其图标并配置“节点属性”。

图 11-6 节点属性



关键属性说明：



- SQL脚本：关联**步骤1**中开发完成的SQL脚本“generate_trade_report”。
- 数据库名称：自动填写SQL脚本“generate_trade_report”中选择的数据数据库。
- 队列名称：自动填写SQL脚本“generate_trade_report”中选择的资源队列。
- 脚本参数：显示SQL脚本“generate_trade_report”中的参数“yesterday”，输入以下EL表达式作为其参数值。

```
#{Job.getYesterday("yyyy-MM-dd")}
```

EL表达式说明：Job为作业对象，通过getYesterday方法获取作业计划执行时间前一天的时间，时间格式为yyyy-MM-dd。

假设作业计划执行时间为2018/9/26 01:00:00，这个表达式计算结果是2018-09-25，该计算结果将替换**SQL脚本**中的\${yesterday}参数。替换后的SQL内容如下：

```
INSERT OVERWRITE TABLE trade_report
SELECT
  sum(trade_count),
  '2018-09-25'
FROM
  trade_log
where
  date_format(trade_time, 'yyyy-MM-dd') = '2018-09-25'
```

4. 单击 ，测试运行作业。
5. 作业测试无问题后，单击 ，保存作业配置。

----结束

A 修订记录

发布日期	修改说明
2020-08-12	第七次正式发布。 新增章节： <ul style="list-style-type: none">● 配置委托，支持创建委托，作业执行过程中，以委托的身份与其他服务交互。● MRS Flink Job，新增此节点。● 屏蔽 函数描述。
2019-09-30	第六次发布。 新增PRESTO SQL脚本，修改 新建脚本 。
2019-07-08	第五次发布。 <ul style="list-style-type: none">● 支持MRS Python Spark作业，新增MRS Spark Python● 血缘关系增强，修改如下节点：<ul style="list-style-type: none">- CDM Job- Rest Client- DLI SQL- DWS SQL- ETL Job
2019-06-21	第四次正式发布。 <ul style="list-style-type: none">● 支持修改保存运行中的实时作业，修改开发作业

发布日期	修改说明
2019-05-28	第三次正式发布。 新增如下章节： <ul style="list-style-type: none"> ● 重命名脚本 ● 重命名作业 ● 配置 ● Env内嵌对象 修改如下章节： <ul style="list-style-type: none"> ● DLI SQL ● DWS SQL ● Rest Client ● MRS SparkSQL ● MRS Hive SQL ● Shell ● ETL Job ● Job内嵌对象
2019-05-24	第二次正式发布。 新增如下章节： <ul style="list-style-type: none"> ● 导出数据连接 ● 导入数据连接 ● 通知周期概览 ● 备份和还原资产 修改如下章节： <ul style="list-style-type: none"> ● 导出导入作业 ● ETL Job ● DIS Client
2019-04-28	第一次正式发布。