

解决方案实践

基于 DLI 电商数据分析与处理

文档版本 1.0.0
发布日期 2023-05-30



版权所有 © 华为技术有限公司 2024。保留一切权利。

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

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

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

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

安全声明

漏洞处理流程

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

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

如企业客户须获取漏洞信息，请参见如下网址：

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

目录

1 方案概述	1
2 资源和成本规划	3
3 实施步骤	5
3.1 准备工作.....	5
3.2 快速部署.....	13
3.3 开始使用.....	21
3.4 快速卸载.....	41
4 附录	42
5 修订记录	43

1 方案概述

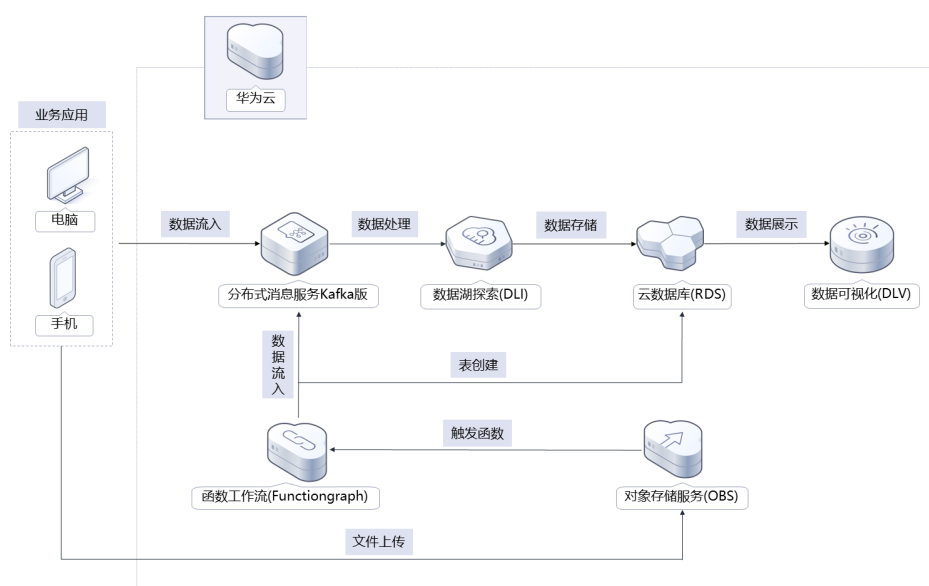
应用场景

当前线上购物无疑是最火热的购物方式，而电商平台则可以以多种方式接入，例如通过web方式访问、通过app的方式访问、通过微信小程序的方式访问等等。该解决方案可以帮助电商平台统计各平台的实时访问数据量、订单数、访问人数等等指标，从而能在显示大屏上实时展示相关数据。电商平台可以通过将每个商品的订单信息实时写入Kafka中或将业务数据表上传至对象存储 OBS桶中从而触发函数工作流自动将数据写入Kafka中，DLI根据当前可以获取到的业务数据，实时统计每种渠道的相关指标，输出存储到数据库中。方便电商平台及时了解数据变化，有针对性地调整营销策略。适用于通过web方式访问、通过app的方式访问、通过微信小程序的方式访问等等电商线上购物。

方案架构

该解决方案基于DLI Flink完成电商业务实时数据的分析处理，获取各个渠道的销售汇总数据。架构如下图所示：

图 1-1 方案架构



该解决方案会部署如下资源：

- 创建一台云数据库RDS for MySQL实例，用于存储统计的订单指标数据，从而展示在数据可视化DLV大屏。
- 创建分布式消息服务 kafka专享版实例，并创建 Topic，用于电商平台写入业务数据，为用户提供便捷高效的消息队列。
- 创建数据湖探索DLI增强型跨源队列，通过创建Flink作业，用于分析和处理电商平台的业务数据。
- 创建一个对象存储服务OBS桶，用于用户上传业务数据表csv文件。
- 使用函数工作流 FunctionGraph创建一个函数，通过OBS触发器，实现将用户上传至OBS桶中的业务数据写入Kafka，同时在RDS创建数据结果表模板。

用户可以购买并配置数据可视化 DLV大屏，快速定制数据大屏，将RDS中分析结果数据展示在大屏中。

方案优势

- 跨源分析
数据免搬迁，就可以关联分析存在OBS中的各个渠道的销售汇总数据。
- 纯SQL操作
DLI已对接多个数据源，直接通过SQL建表就可以完成数据源的映射。
- 一键部署
一键轻松部署，即可完成Kafka及DLI等资源发放，帮助用户轻松搭建电商数据分析处理平台。

约束与限制

- 部署该解决方案之前，您需注册华为账号并开通华为云，完成实名认证，且账号不能处于欠费或冻结状态。
- 第一次使用数据湖数据湖 DLI时需要登录DLI控制台，如果使用IAM子账户部署该解决方案，请参考部署指南确保该账号已经有DLI的相关授权，第一次使用DLI时候也需要用子账号登录一次控制台。具体请参考[数据湖探索 DLI帮助文档](#)。

2 资源和成本规划

该解决方案主要部署如下资源，不同产品的花费仅供参考，实际以收费账单为准，具体请参考华为云[官网价格](#)：

表 2-1 资源和成本规格

华为云服务	配置示例	每月预估花费
分布式消息服务Kafka版	<ul style="list-style-type: none">● 区域：华北-北京四● 按需计费：4.62元/小时● 规格：kafka.2u4g.cluster● 代理数：3个● 存储空间：超高IO 300GB● 购买时长：720小时● 购买量：1	3326.40 元
数据湖探索 DLI	<ul style="list-style-type: none">● 区域：华北-北京四● 按需计费：5.6元/小时● 类型：通用队列（专属资源模式）● CPU 架构：x86● AZ策略：单AZ● 规格：16 CUs● 购买时长：720小时● 购买数量：1	4032.00 元

华为云服务	配置示例	每月预估花费
云数据库 RDS for MySQL	<ul style="list-style-type: none"> ● 区域：华北-北京四 ● 按需计费：0.52元/小时 ● 数据库引擎：MySQL 5.7 ● 实例类型：单机 ● 性能规格：通用型 2 vCPUs 4 GB ● 存储空间：SSD云盘 100GB ● 购买时长：720小时 ● 购买量：1 	374.4 元
对象存储服务 OBS	<ul style="list-style-type: none"> ● 区域：华北-北京四 ● 存储空间：数据存储（多AZ存储） ● 默认存储类别：标准存储 ● 桶策略：私有 ● 请求费用：0.0100元/万次 ● 存储空间：0.0990元/GB/月 ● 流量费用： <ul style="list-style-type: none"> ● 公网流出流量 / 00:00-08:00（闲时）0.2500元/GB ● 公网流出流量 / 08:00-24:00（忙时）0.5000元/GB <p>费用包括存储空间、请求费用、流量费用两部分，详细请参考每月账单。</p> <p>费用包括存储空间、请求费用、流量费用两部分，具体请参考OBS计费详情。</p>	费用包括存储空间、请求费用、流量费用两部分，详细请参考每月账单。
函数工作流 FunctionGraph	<ul style="list-style-type: none"> ● 区域：华北-北京四 ● 产品：函数 ● 请求次数： <ul style="list-style-type: none"> 0-100万次：0元/100万次 100万次以上：1.33元/100万次 ● 计量时间： <ul style="list-style-type: none"> 0-400,000 GB/秒：0元/GB-秒 400,000 GB/秒以上：0.00011108元/GB-秒 	0 元
数据可视化 DLV	<ul style="list-style-type: none"> ● 区域：华北-北京四 ● 计费模式：包年包月 ● 版本：基础版 	5 元
合计		7737.8 元 + OBS 服务产生费用

3 实施步骤

- 3.1 准备工作
- 3.2 快速部署
- 3.3 开始使用
- 3.4 快速卸载

3.1 准备工作

须知

- DLI授权需要主账号或者用户组admin中的子账号进行操作，请[创建admin子账户](#)或给[IAM用户授权admin](#)。
- 用户首次使用子账号创建队列时，需要先使用主账号登录控制台，在DLI的数据库中保持记录，才能创建队列。

创建 rf_admin_trust 委托

步骤1 进入华为云官网，打开[控制台管理](#)界面，鼠标移动至个人账号处，打开“统一身份认证”菜单。

图 3-1 控制台管理界面



图 3-2 统一身份认证菜单



步骤2 进入“委托”菜单，搜索“rf_admin_trust”委托。

图 3-3 委托列表



- 如果委托存在，则不用执行接下来的创建委托的步骤
- 如果委托不存在时执行接下来的步骤创建委托

步骤3 单击步骤2界面中的“创建委托”按钮，在委托名称中输入“rf_admin_trust”，委托类型选择“云服务”，选择“RFS”，单击“下一步”。

图 3-4 创建委托

委托 / 创建委托

* 委托名称

* 委托类型 普通帐号
将帐号内资源的操作权限委托给其他华为云帐号。
 云服务
将帐号内资源的操作权限委托给华为云服务。

* 云服务

* 持续时间

描述

0/255

步骤4 在搜索框中输入“Tenant Administrator”权限，并勾选搜索结果。

图 3-5 选择策略

委托“rf_admin_trust”将拥有所选策略

策略搜索框: 从其他区域策略目录复制策略

名称	类型
<input checked="" type="checkbox"/> Tenant Administrator 全部云服务管理员 (IAM管理权限)	系统角色

步骤5 选择“所有资源”，并单击下一步完成配置。

图 3-6 设置授权范围

根据当前选择的策略，策略授予以下授权范围方案，更便于您最小化授权，可进行选择。了解如何根据应用场景选择最佳的授权范围方案

选择授权范围方案

所有资源
授权后，IAM用户可以按照权限使用帐号中所有资源，包括企业项目、区域项目和全局服务资源。

[展开其他方案](#)

步骤6 “委托”列表中出现“rf_admin_trust”委托则创建成功。

图 3-7 委托列表



----结束

创建 IAM Agency Management FullAccess 策略

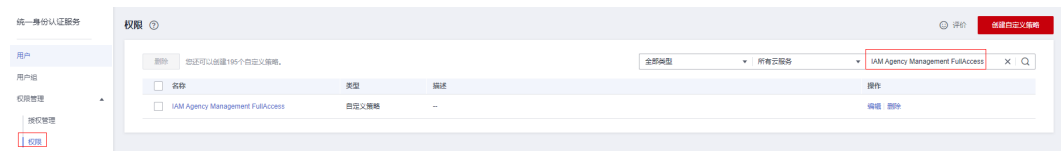
步骤1 打开“统一身份认证”菜单

图 3-8 统一身份认证菜单



步骤2 进入“权限管理”->“权限”菜单，在搜索框输入“IAM Agency Management FullAccess”当前账号是否存在IAM委托管理权限

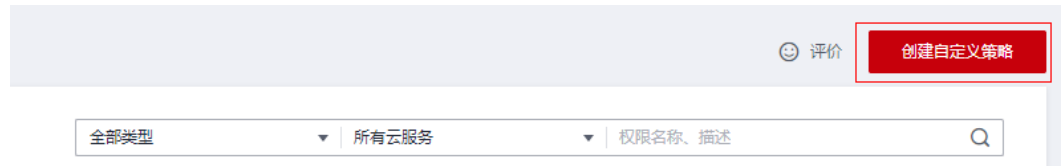
图 3-9 权限列表



- 如果搜索结果不为空，则当前账号已经存在IAM委托管理权限，不需要重复创建
- 如果过搜索结果为空，则继续创建“IAM Agency Management FullAccess”权限

步骤3 单击“创建自定义策略”按钮

图 3-10 创建自定义策略



步骤4 输入策略名称为“IAM Agency Management FullAccess”，选择“JSON视图”，在策略内容中输入如下JSON代码，单击确认按钮

图 3-11 创建自定义策略

* 策略名称

策略配置方式

* 策略内容

```
1 {  
2   "Version": "1.1",  
3   "Statement": [  
4     {  
5       "Effect": "Allow",  
6       "Action": [  
7         "iam:agencies:updateAgency",  
8         "iam:permissions:listRolesForAgencyOnDomain",  
9         "iam:permissions:revokeRoleFromAgencyOnDomain",  
10        "iam:permissions:listRolesForAgency",  
11        "iam:permissions:checkRoleForAgencyOnProject",  
12        "iam:roles:listRoles",  
13        "iam:agencies:deleteAgency",  
14        "iam:permissions:checkRoleForAgency",  
15        "iam:permissions:listRolesForAgencyOnProject",  
16        "iam:permissions:checkRoleForAgencyOnDomain",  
17        "iam:agencies:listAgencies",  
18        "iam:permissions:grantRoleToAgencyOnDomain",  
19        "iam:permissions:revokeRoleFromAgencyOnProject",  
20        "iam:agencies:getAgency",  
21        "iam:agencies:createAgency",  
22        "iam:permissions:grantRoleToAgency",  
23        "iam:permissions:grantRoleToAgencyOnProject",  
24        "iam:permissions:revokeRoleFromAgency"  
25      ]  
4     }  
3   ]  
1 }
```

策略描述

作用范围

```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Action": [  
        "iam:agencies:createAgency",  
        "iam:agencies:listAgencies",  
        ...  
      ]  
    }  
  ]  
}
```

```
"iam:agencies:getAgency",  
"iam:agencies:deleteAgency",  
"iam:agencies:updateAgency",  
"iam:permissions:revokeRoleFromAgencyOnProject",  
"iam:permissions:revokeRoleFromAgencyOnDomain",  
"iam:permissions:revokeRoleFromAgency",  
"iam:permissions:grantRoleToAgencyOnDomain",  
"iam:permissions:grantRoleToAgencyOnProject",  
"iam:permissions:grantRoleToAgency",  
"iam:permissions:listRolesForAgencyOnDomain",  
"iam:permissions:listRolesForAgencyOnProject",  
"iam:permissions:checkRoleForAgencyOnDomain",  
"iam:permissions:checkRoleForAgencyOnProject",  
"iam:permissions:listRolesForAgency",  
"iam:permissions:checkRoleForAgency",  
"iam:roles:listRoles"  
  ],  
  "Effect": "Allow"  
}  
]
```

步骤5 界面无报错，则成功创建IAM Agency Management FullAccess权限

----结束

给 rf_admin_trust 委托添加 IAM Agency Management FullAccess 策略

步骤1 打开“统一身份认证”菜单

图 3-12 统一身份认证菜单



步骤2 进入“委托”菜单，选择rf_admin_trust委托

图 3-13 委托列表



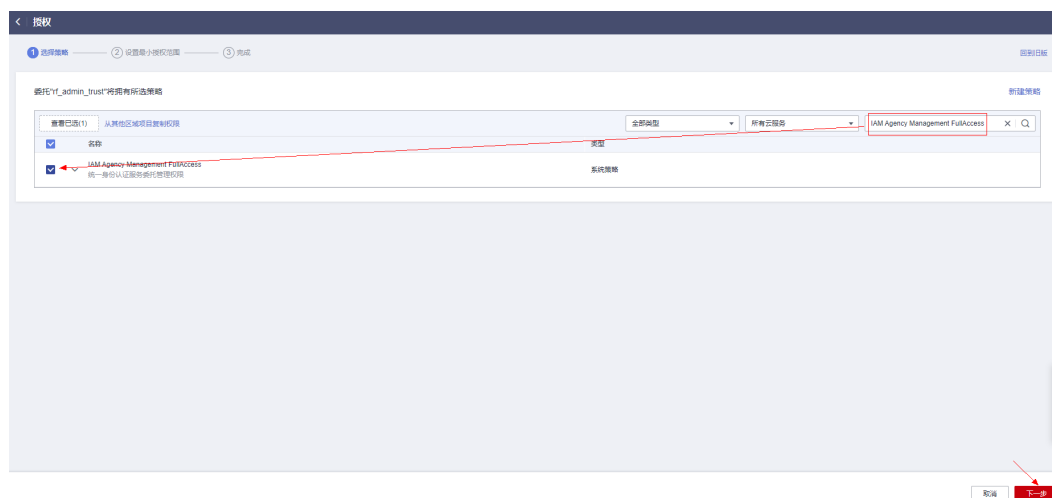
步骤3 进入“授权记录”菜单，单击“授权”按钮

图 3-14 授权记录



步骤4 在搜索框输入IAM Agency Management FullAccess，勾选过滤出来的记录，单击下一步，并确认完成权限的配置

图 3-15 配置 IAM Agency Management FullAccess 策略



步骤5 配置好后的情况：rf_admin_trust委托拥有Tenant Administrator和IAM Agency Management FullAccess权限

图 3-16 授权记录列表



----结束

获取子网 ID

本章节主要帮助用户在快速部署该解决方案之前，获取部分依赖的资源，以供一键部署时使用。

步骤1 登录[华为云官网控制台](#)，单击[虚拟私有云VPC](#)，选择该方案资源需要部署的VPC，单击该VPC下的[子网](#)，选择子网网段与数据湖探索 DLI所在网段（172.16.0.0/16）不重合的任一子网或，获取网络ID。

图 3-17 子网列表



名称	虚拟私有云	IPv4网段	IPv6网段	状态	可用区
subnet-cb0f	vpc-default	192.168.8.0/24	-- 启用IPv6	可用	可用区1
subnet-70b7	vpc-default	192.168.6.0/24	-- 启用IPv6	可用	可用区1
subnet-488b	vpc-default	192.168.3.0/24	-- 启用IPv6	可用	可用区1
subnet-87f9	vpc-default	192.168.1.0/24	-- 启用IPv6	可用	可用区1
subnet-default	vpc-default	192.168.0.0/24	-- 启用IPv6	可用	--

图 3-18 子网网络 ID



基本信息

名称	vpc-default	ID	dc2b-...ccce6
状态	可用	VPC网段	192.168.0.0/16 编辑网段
企业项目	default	描述	vpc-default

资源概览

弹性云服务器 添加	5	弹性云服务器 添加	0	弹性负载均衡 添加	0
弹性网卡 添加	8				

----结束

3.2 快速部署

本章节主要帮助用户快速部署该解决方案。

表 3-1 参数填写说明

参数名称	类型	是否必填	参数解释	默认值
subnet_id	String	必填	子网ID，该模板使用已有子网，该解决方案创建的数据湖探索 DLI队列所在网段默认 172.16.0.0/16，请确保所选子网网段不与DLI队列网段重合。该方案所涉及到的其他云服务资源将会部署在该子网下，查询子网ID请参考 获取子网ID 。	空
secgroup_name	String	必填	安全组名称，该模板新建安全组，安全组规则请参考 安全组规则修改（可选） 或者参考 Kafka帮助文档 修改安全组。取值范围：1-64个字符，支持数字、字母、中文、_(下划线)、-(中划线)、.(点)。	analyzing-e-commerce-data-with-dli-demo
obs_bucket_name	String	必填	对象存储服务OBS桶名称，命名格式为 {obs_bucket_name}-obs，用于存放应用数据，全局唯一。取值范围：长度 1~59个字符，以小写字母或数字开头，仅支持小写字母、数字、中划线(-)、英文句号(.)。	空
kafka_name	String	必填	分布式消息服务Kafka实例名称，取值范围：以字母开头，长度为4-64个字符，仅支持字母、数字、连字符(-)和下划线(_)。	analyzing-e-commerce-data-with-dli-demo

参数名称	类型	是否必填	参数解释	默认值
kafka_flavor_id	String	必填	Kafka实例规格 , 可选值 (除拉美-墨西哥城一以外区域): c6.2u4g.cluster、 c6.4u8g.cluster、 c6.8u16g.cluster、 c6.12u24g.cluster 、 c6.16u32g.cluster 。	c6.2u4g.cluster
kafka_storage_spec_type	String	必填	Kafka实例存储空间磁盘类型 , 创建完成后不可修改。请根据业务IO要求选择。可选值: dms.physical.storage.high.v2(高I/O磁盘类型)、 dms.physical.storage.ultra.v2 (超高I/O磁盘类型)。	dms.physical.storage.ultra.v2
kafka_storage_space	Number	必填	Kafka实例存储空间容量, 存储空间不支持扩容。存储空间主要用于存储消息 (包含副本, Kafka默认使用3副本), 除了存储消息外还需要预留部分空间用于存储日志和元数据。例如: 业务消息体积预估100GB, 则存储空间最少应为 $100\text{GB} * \text{副本数} + \text{预留磁盘大小}$ 100GB。取值范围: 300-900,000。	300

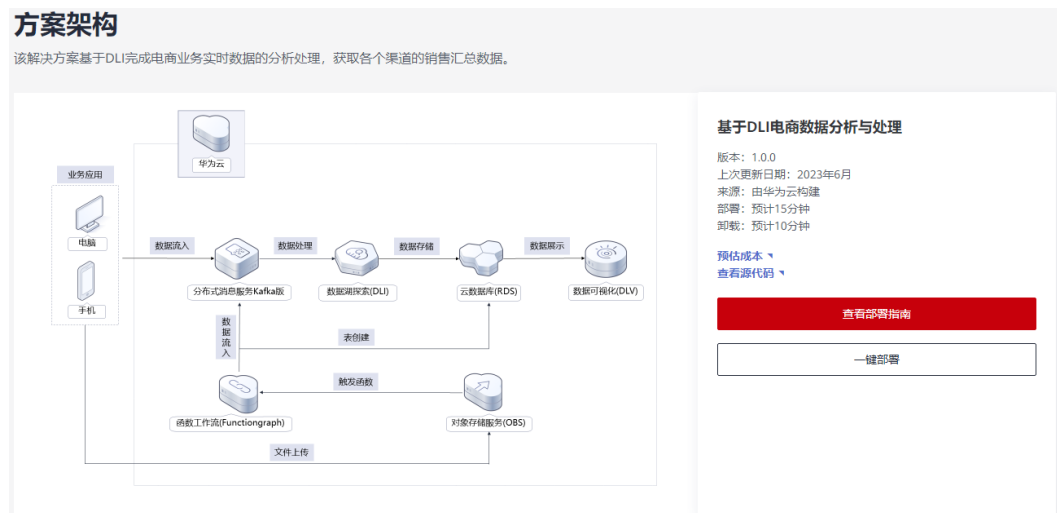
参数名称	类型	是否必填	参数解释	默认值
kafka_manager_password	String	必填	Kafka实例Manager用户密码，创建完成后，可参考部署指南登录Kafka控制台修改密码。取值范围：8-32个字符，密码至少必须包含大写字母、小写字母、数字和特殊字符`~!@#\$%^&*()-_+= [{]}:;\"'<.>?和空格中的三种，并且不能以-开头。不能与用户名或倒叙的用户名相同。	空
dli_queue_name	String	必填	数据湖探索DLI队列名称，第一次使用DLI时需要登录DLI控制台，如果使用IAM子账户部署该解决方案，请确保该账号已经有 DLI的相关授权 。取值范围：以字母开头，长度为1-48个字符，仅支持字母、数字和下划线（_），且不能以下划线或数字开头。	analyzing_e-commerce_data_with_dli_demo
dli_queue_specifications	Number	必填	DLI规格，是计算节点所有CU数的总和，1CU=1核4GB。DLI系统会自动分配各计算节点的内存和CPU大小，具体个数客户端不感知。可选值：16、64、256或512。	16

参数名称	类型	是否必填	参数解释	默认值
rds_name	String	必填	云数据库RDS for MySQL名称，不支持重名。取值范围：4-64个字符，以小写字母开头，由小写字母、数字、中划线(-)组成，且不能以中划线(-)结尾。	analyzing-e-commerce-data-with-dli-demo
rds_flavor_id	String	必填	云数据库RDS for MySQL规格 ，该方案默认创建单击版，请填写单击实例规格。	rds.mysql.n1.large.2 (通用型，单机实例 2vCPU4GB)
rds_volume_size	Number	必填	云数据库RDS实例磁盘大小，取值范围：40~4000，必须为10的整数倍。	100
rds_volume_type	String	必填	云数据库RDS实例磁盘类型，有效值：ULTRAHIGH (SSD)、LOCALSSD (表示本地SSD)、CLOUDSSD (SSD云盘，仅支持通用型和独享型规格实例)、ESSD (极速型SSD，仅支持独享型规格实例)。	CLOUDSSD
rds_password	String	必填	云数据库RDS for MySQL密码，创建完成后，可参考部署指南登录RDS控制台修改密码。取值范围：8-32个字符，密码至少必须包含大写字母、小写字母、数字和特殊字符(@#\$%^&_+=?,()&)中的三种。	空

参数名称	类型	是否必填	参数解释	默认值
functiongraph_name	String	必填	函数工作流 functiongraph 函数名称，取值范围：长度为2-57个字符，可包含字母、数字、_（下划线）和-（中划线），以字母开头，以字母或数字结尾。	analyzing-e-commerce-data-with-dli-demo

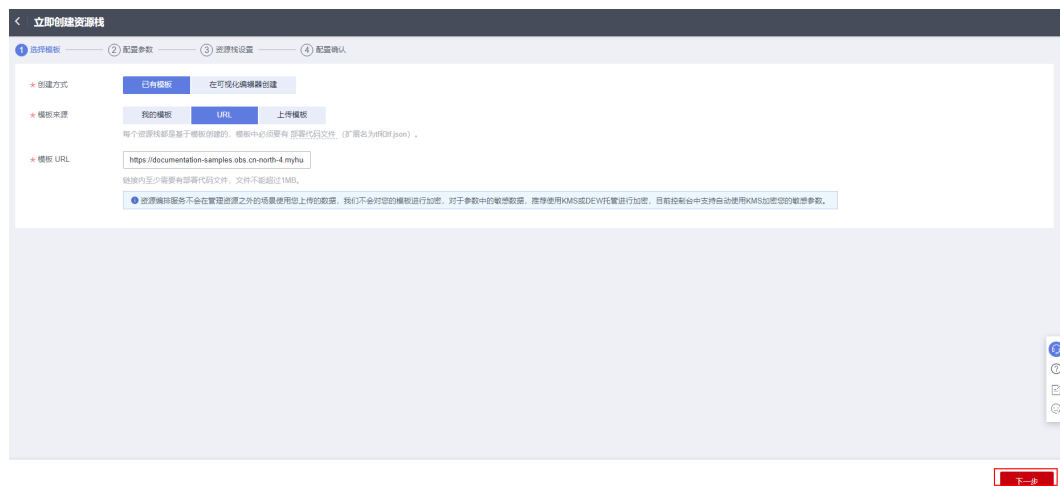
步骤1 登录[华为云解决方案实践](#)，选择“基于DLI电商数据分析与处理”解决方案。单击“一键部署”，跳转至解决方案创建堆栈界面。

图 3-19 解决方案实施库



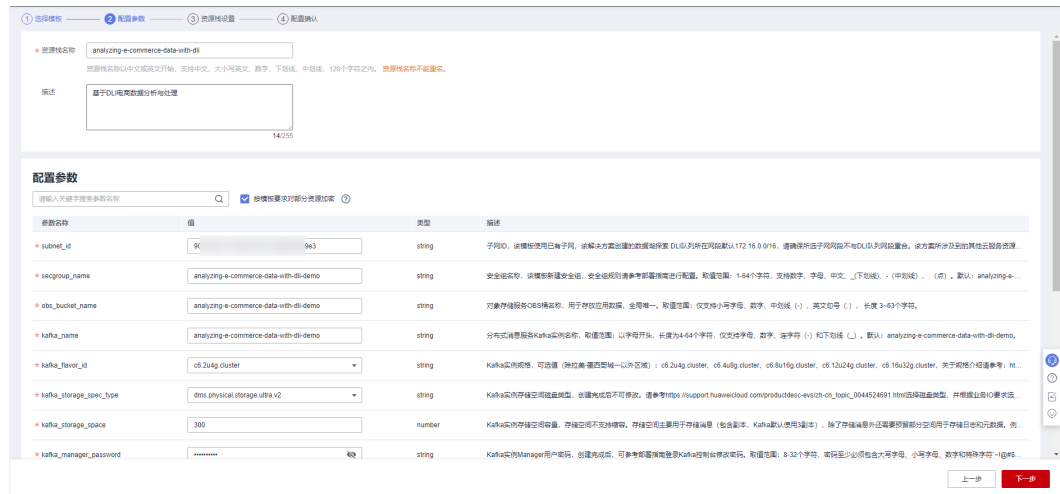
步骤2 在选择模板界面中，单击“下一步”。

图 3-20 选择模板



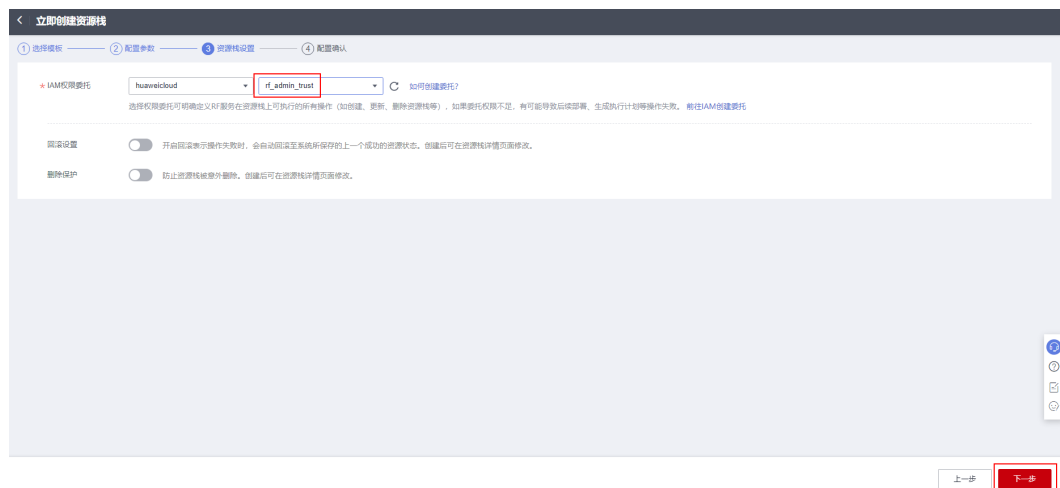
步骤3 在配置参数界面中，参考表3-1完成自定义参数填写，单击“下一步”。

图 3-21 配置参数



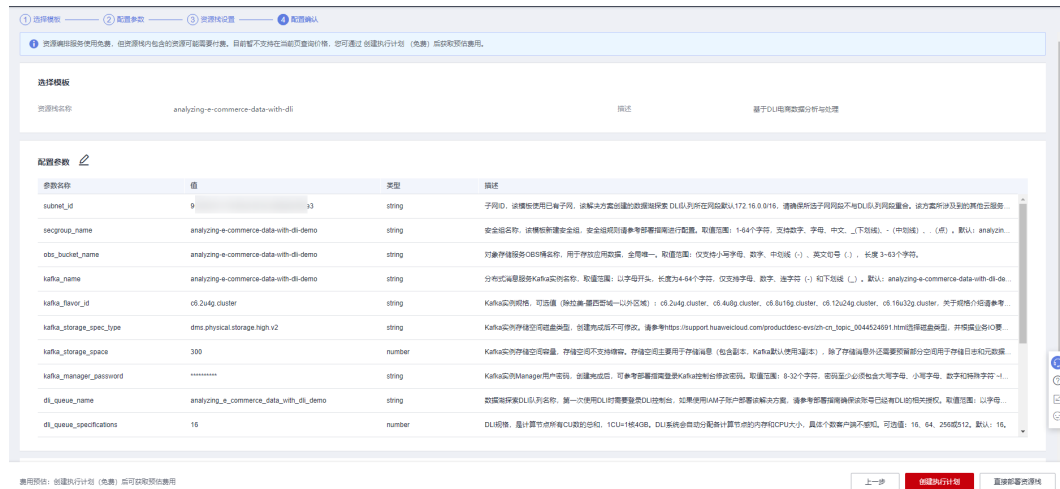
步骤4 在资源栈设置页面中，权限委托选择“rf_admin_trust”，单击“下一步”。

图 3-22 资源栈设置



步骤5 在配置确认页面中，单击“创建执行计划”。

图 3-23 配置确认



步骤6 在弹出的创建执行计划框中，自定义填写执行计划名称，单击“确定”。

图 3-24 创建执行计划



步骤7 待执行计划状态为“创建成功，待部署”后，单击“部署”，并且在弹出的执行计划确认框中单击“执行”。

图 3-25 执行计划

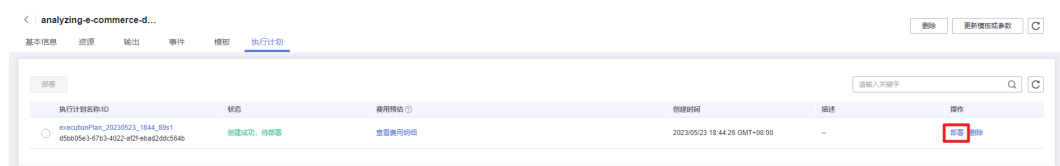
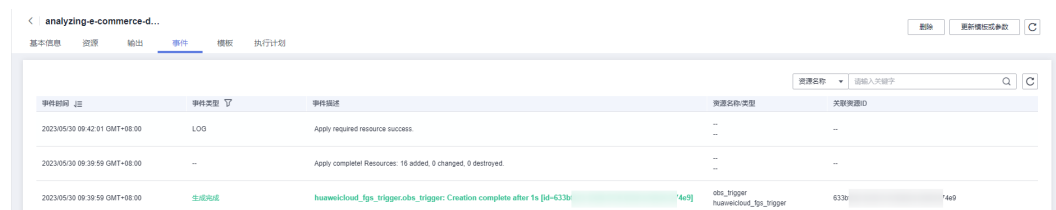


图 3-26 执行计划确认



步骤8 等待解决方案自动部署。部署成功后，单击“事件”，回显结果如下：

图 3-27 资源创建成功



----结束

3.3 开始使用

安全组规则修改（可选）

须知

- Kafka默认使用9092端口，该方案默认针对子网网段及DLI网段开放，请参考[修改安全组规则](#)，配置IP地址白名单，以便能正常访问服务。
- 云数据库RDS 默认使用3306端口，该方案默认针对子网网段及DLI网段开放，请参考[修改安全组规则](#)，配置IP地址白名单，以便能正常访问服务。

安全组实际是网络流量访问策略，包括网络流量入方向规则和出方向规则，通过这些规则为安全组内具有相同保护需求并且相互信任的云服务器、云容器、云数据库等实例提供安全保护。

如果您的实例关联的安全组策略无法满足使用需求，比如需要添加、修改、删除某个TCP端口，请参考以下内容进行修改。

- 添加安全组规则：根据业务使用需求需要开放某个TCP端口，请参考[添加安全组规则](#)添加入方向规则，打开指定的TCP端口。
- 修改安全组规则：安全组规则设置不当会造成严重的安全隐患。您可以参考[修改安全组规则](#)，来修改安全组中不合理的规则，保证云服务器等实例的网络安全。
- 删除安全组规则：当安全组规则入方向、出方向源地址/目的地址有变化时，或者不需要开放某个端口时，您可以参考[删除安全组规则](#)进行安全组规则删除。

步骤1 查看OBS桶。在控制台单击“服务列表”，选择“[对象存储服务 OBS](#)”，单击进入OBS页面。该桶将作为用户上传电商数据文件的存储桶。

图 3-28 OBS 桶



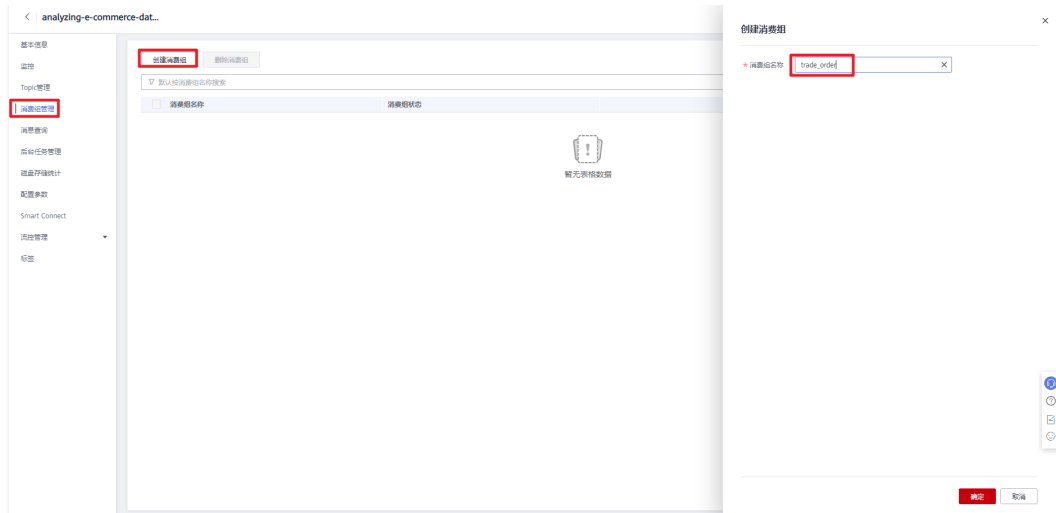
步骤2 查看Kafka。在控制台单击“服务列表”，选择“[分布式消息服务Kafka版](#)”，在“Kafka专享版”页面找到该方案所创建的Kafka实例。

图 3-29 Kafka 实例



步骤3 创建Kafka消费组。单击“消费组管理>创建消费组”，输入消费组名字“trade_order” (使用demo样例时请填写该名称)，单击“确定”保存。

图 3-30 消费组



步骤4 查看RDS。在控制台单击“服务列表”，选择“云数据库RDS”，单击进入RDS页面。在“实例管理页面”，找到该解决方案已经创建的RDS实例。

图 3-31 云数据库 RDS 实例



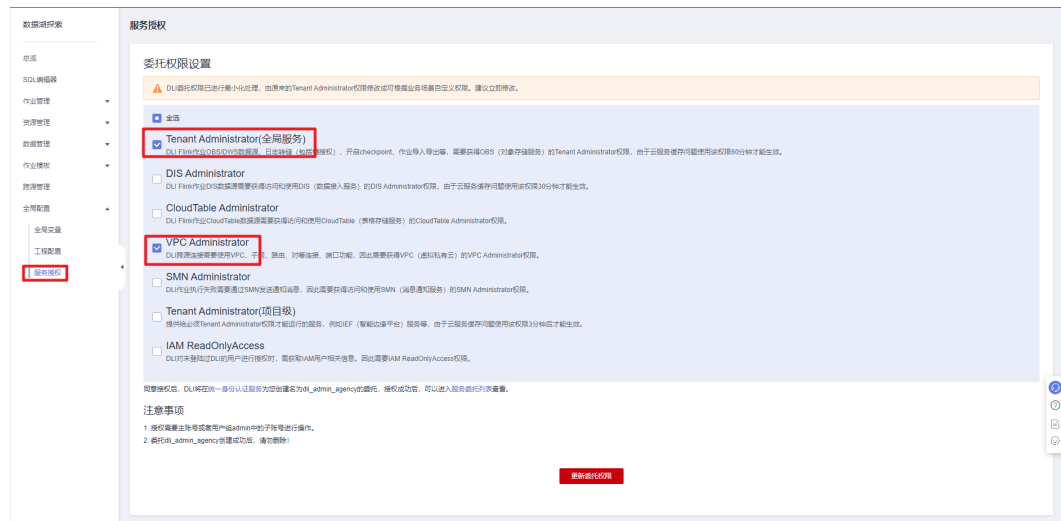
步骤5 查看DLI。在控制台单击“服务列表”，选择“数据湖探索 DLI”，单击进入DLI服务页面。单击“资源管理 > 队列管理”，查询创建的DLI队列。

图 3-32 DLI 队列



步骤6 配置DLI服务授权。单击“全局变量>服务授权”，选定以下两项委托授权，单击“更新委托权限”。

图 3-33 DLI 服务授权

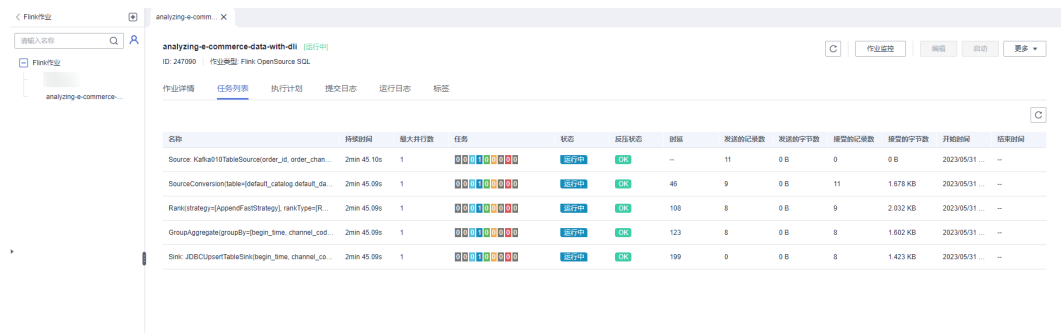


步骤7 参考[数据湖探索 DLI使用指南](#)创建Flink作业，启动成功后作业状态为“运行中”。

步骤8 体验Demo样例。获取[Demo数据](#)样例文件，上传至[3.3开始使用步骤1](#)中的OBS桶中，即可触发函数 workflows 去主动将数据写入Kafka，并自动在MySQL中创建结果表。手动创建请参考[云数据库RDS for MySQL使用指南](#)创建结果存储表，待Flink作业创建成功后参考[分布式消息服务 Kafka使用指南](#)连接Kafka手动写入数据（或参考[Kafka开发指南-python文档](#)及[函数工作流FunctionGraph定制化开发](#)修改代码并重新部署，上传业务数据csv文件至OBS桶即可自动写入数据）。

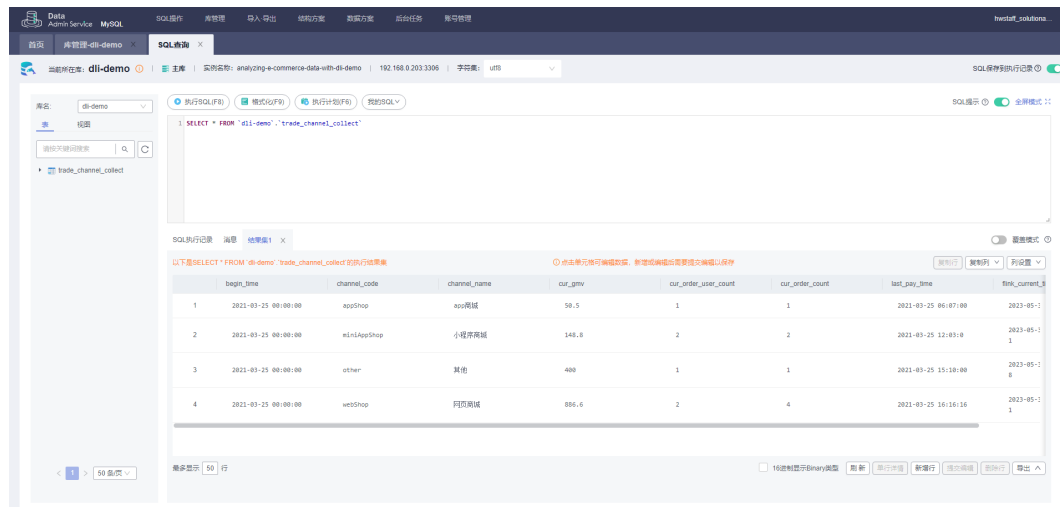
步骤9 进入Flink作业，单击“任务列表”，查看任务状态。

图 3-34 任务列表



步骤10 登录数据库，“SQL操作”>“SQL查询”，执行如下SQL语句，即可查询到经过Flink作业处理后的结果数据。

图 3-35 结果数据表

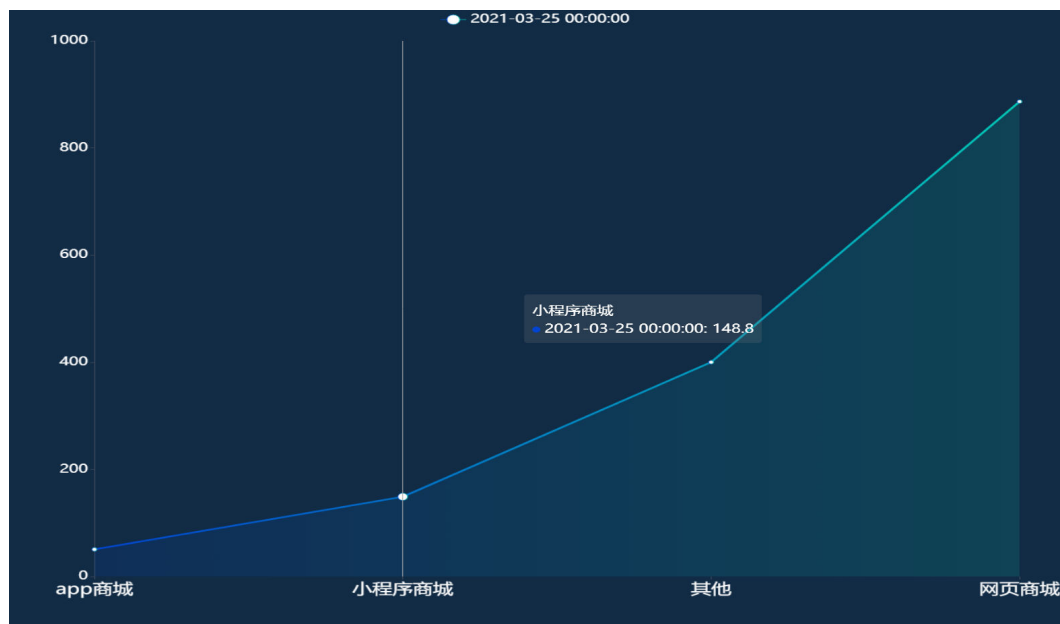


The screenshot shows a SQL query execution interface with a table of results. The table has the following columns: begin_time, channel_code, channel_name, cur_gmv, cur_order_user_count, cur_order_count, last_pay_time, and time_current. The data rows are as follows:

begin_time	channel_code	channel_name	cur_gmv	cur_order_user_count	cur_order_count	last_pay_time	time_current
2021-03-25 00:00:00	appShop	app商城	50.5	1	1	2021-03-25 06:07:00	2023-05-11
2021-03-25 00:00:00	miniAppShop	小程序商城	148.8	2	2	2021-03-25 12:03:00	2023-05-11
2021-03-25 00:00:00	other	其他	400	1	1	2021-03-25 15:18:00	2023-05-11
2021-03-25 00:00:00	webShop	网页商城	886.4	2	4	2021-03-25 16:16:16	2023-05-11

步骤11 如需在可视化数据 DLV大屏展示数据，请参考配置数据可视化 DLV大屏进行配置。配置效果如下：

图 3-36 DLV 大屏展示



----结束

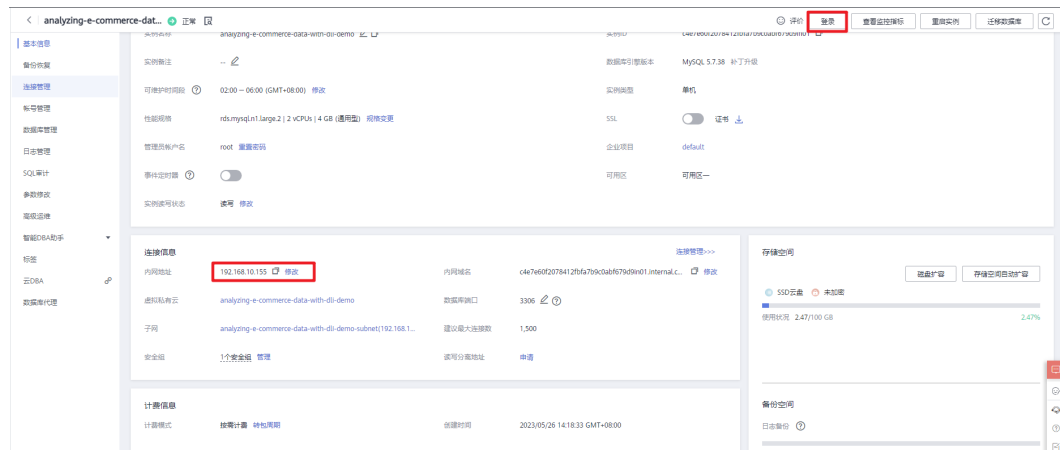
云数据库 RDS for MySQL 使用指南

步骤1 进入RDS页面，在“实例管理页面”，找到该解决方案已经创建的RDS实例，获取其内网地址。

图 3-37 云数据库 RDS 实例



图 3-38 RDS 内网地址



步骤2 单击所创建RDS实例的“登录”，跳转至“数据管理服务-DAS”。输入相关账户信息，单击“测试连接”。显示连接成功后，单击“登录”，进入“实例登录”页面。登录RDS实例后，单击“新建数据库”，创建名称为“dli-demo”的数据库。

图 3-39 数据库登录

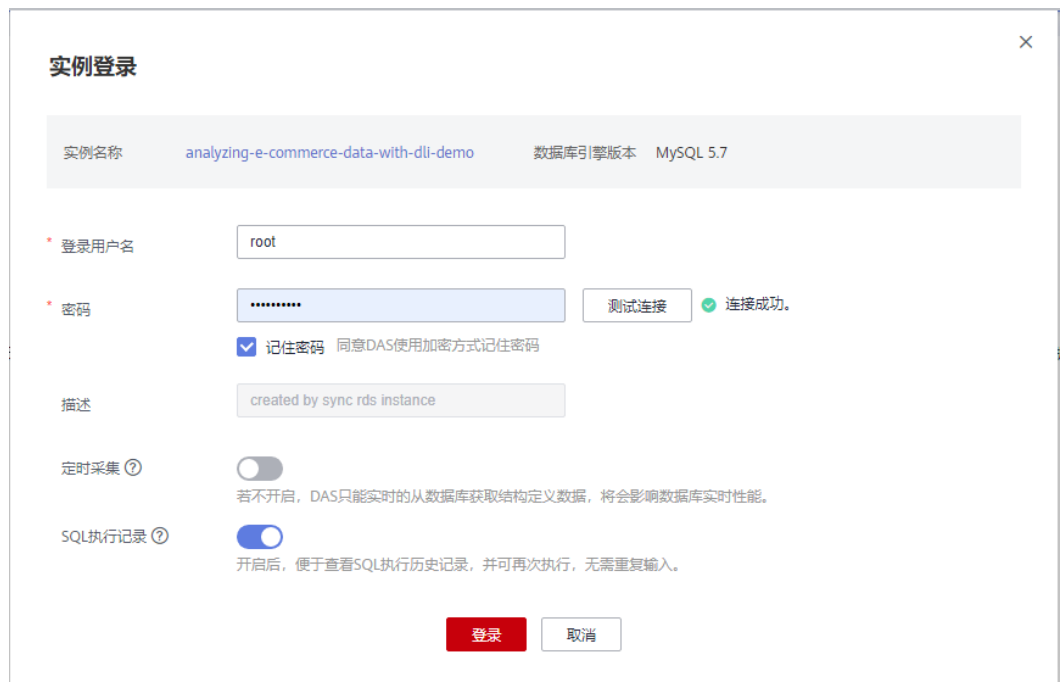


图 3-40 新建数据库



图 3-41 数据库参数填写

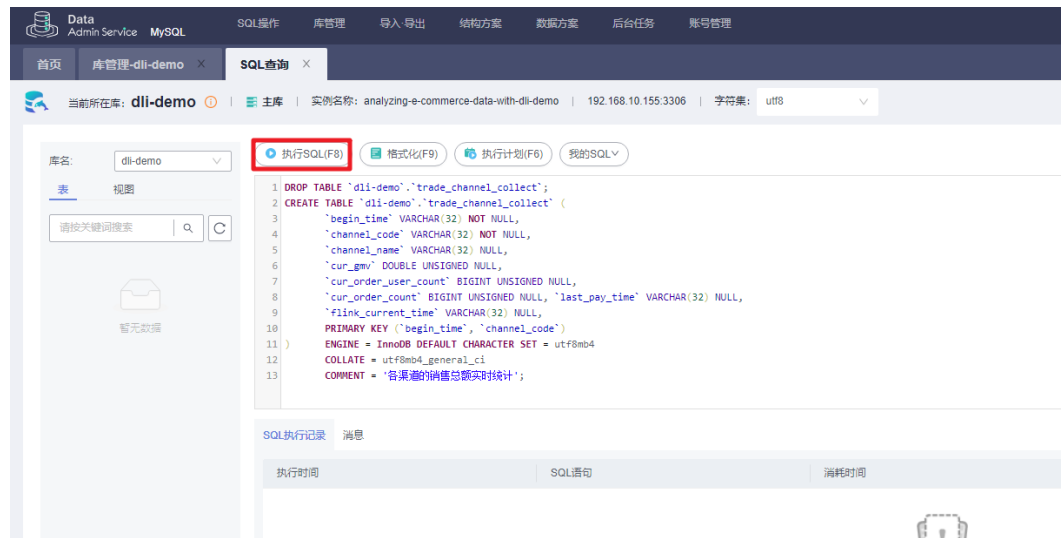


步骤3 单击“SQL窗口”，根据实际业务场景创建参考如下SQL代码创建表，表相关字段含义在表1 数据源表和表2 结果表中有详细介绍。

图 3-42 切换 SQL 视图



图 3-43 执行 SQL 表创建



以下SQL代码仅供参考，请根据实际业务场景修改表信息：

```
DROP TABLE `dli-demo`.`trade_channel_collect`;
CREATE TABLE `dli-demo`.`trade_channel_collect` (
  `begin_time` VARCHAR(32) NOT NULL,
  `channel_code` VARCHAR(32) NOT NULL,
  `channel_name` VARCHAR(32) NULL,
  `cur_gmv` DOUBLE UNSIGNED NULL,
  `cur_order_user_count` BIGINT UNSIGNED NULL,
  `cur_order_count` BIGINT UNSIGNED NULL, `last_pay_time` VARCHAR(32) NULL,
  `flink_current_time` VARCHAR(32) NULL,
  PRIMARY KEY (`begin_time`, `channel_code`)
)
ENGINE = InnoDB DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_general_ci
COMMENT = '各渠道的销售总额实时统计';
```

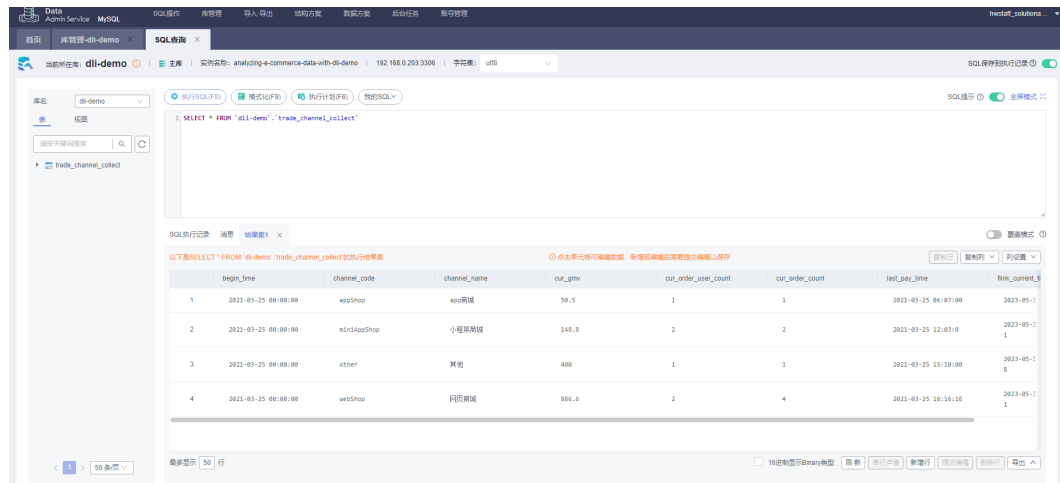
步骤4 查询表数据。登录MySQL实例，单击“SQL窗口”执行如下SQL语句，即可查询到经过Flink作业处理后的结果数据。

```
SELECT * FROM `dli-demo`.`trade_channel_collect`;
```

图 3-44 切换 SQL 窗口



图 3-45 查询结果表

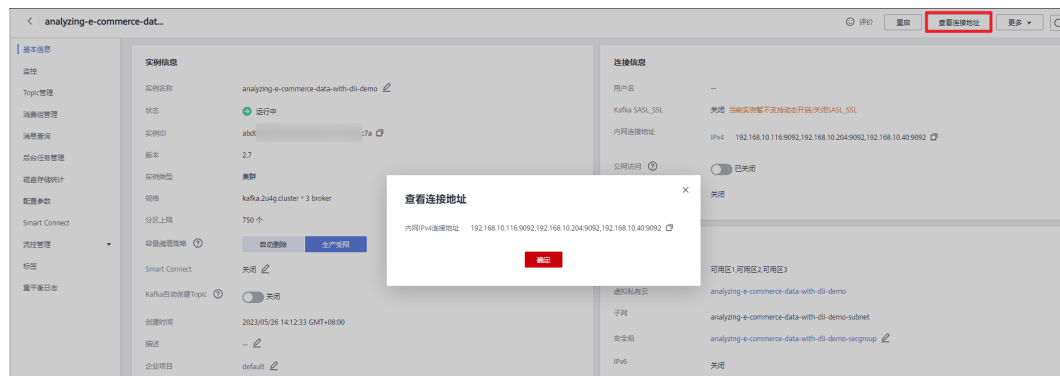


----结束

分布式消息服务 Kafka 使用指南

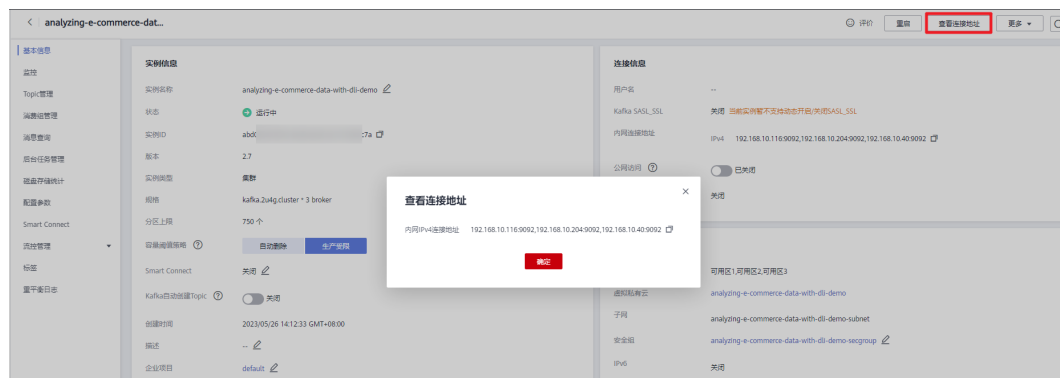
步骤1 在控制台单击“服务列表”，选择“[分布式消息服务Kafka版](#)”，在“Kafka专享版”页面找到该方案所创建的Kafka实例。更多介绍请参考[Kafka介绍](#)。

图 3-46 Kafka 实例链接地址



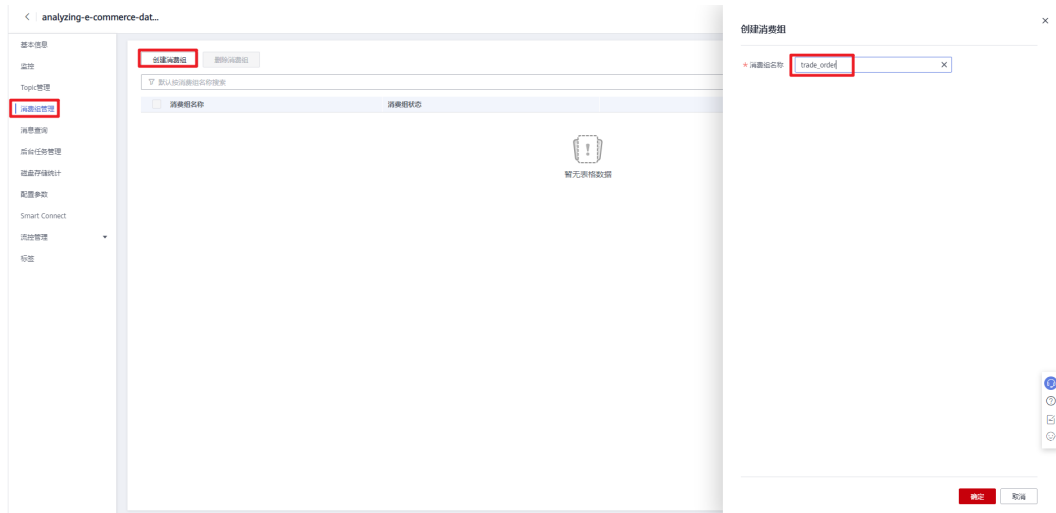
步骤2 进入实例详情页面。单击“基本信息”，获取“连接地址”。

图 3-47 Kafka 实例链接地址



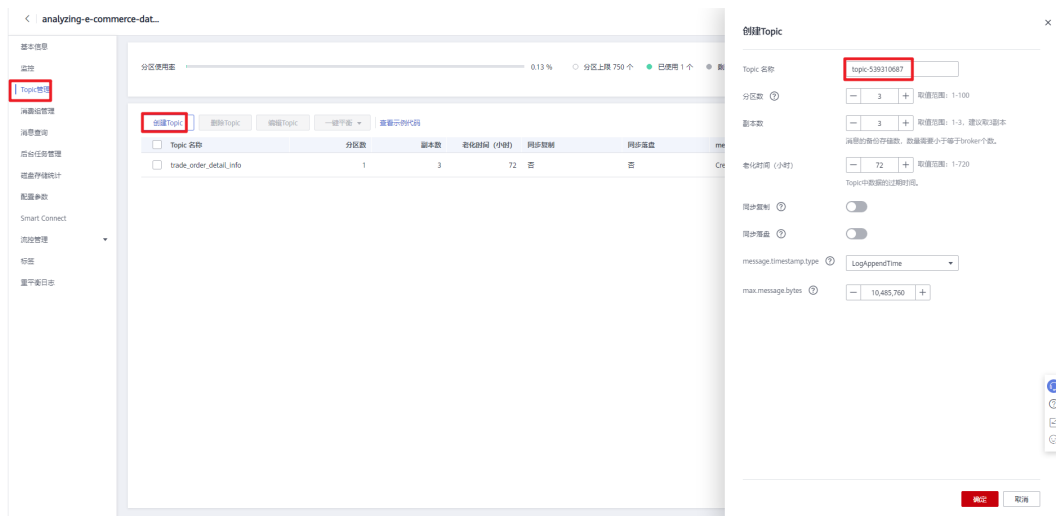
步骤3 单击“消费组管理>创建消费组”，输入消费组名字，单击“确定”保存。

图 3-48 消费组



步骤4 单击“Topic管理”，创建一个Topic。

图 3-49 创建 Topic



步骤5 使用Kafka客户端向指定topic发送数据，模拟实时数据流。具体方法请参考[DMS-连接实例生产消费信息](#)。

图 3-50 模拟数据写入 Kafka

```
(ldi@kafka-client bin)$ ./kafka-console-producer.sh --broker-list 192.168.0.3:9092,192.168.0.147:9092,192.168.0.192:9092 --topic trade_order_detail info
>{"order_id":"202103241000000001", "order_channel":"webShop", "order_time":"2021-03-24 10:00:00", "pay_amount":"100.00", "real_pay":
"100.00", "pay_time":"2021-03-24 10:02:03", "user_id":"0001", "user_name":"Alice", "area_id":"330106"} {"order_id":"202103241606060001", "order_channel":"appShop", "order_time":"2021-03-24 16:06:06", "pay_amount":"200.00", "real_pay":
"180.00", "pay_time":"2021-03-24 16:10:06", "user_id":"0001", "user_name":"Alice", "area_id":"330106"} {"order_id":"202103251202020001", "order_channel":"miniAppShop", "order_time":"2021-03-25 12:02:02", "pay_amount":"60.00", "rea
l_pay":"60.00", "pay_time":"2021-03-25 12:03:00", "user_id":"0002", "user_name":"Bob", "area_id":"330110"} {"order_id":"202103251505050001", "order_channel":"qqShop", "order_time":"2021-03-25 15:05:05", "pay_amount":"500.00", "real_pa
y":"400.00", "pay_time":"2021-03-25 15:10:00", "user_id":"0003", "user_name":"Cindy", "area_id":"330108"} {"order_id":"202103252020200001", "order_channel":"webShop", "order_time":"2021-03-24 20:20:20", "pay_amount":"600.00", "real_p
ay":"480.00", "pay_time":"2021-03-25 00:00:00", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"} {"order_id":"202103260808080001", "order_channel":"webShop", "order_time":"2021-03-25 08:08:08", "pay_amount":"300.00", "real_pay":"240.00", "pay_time":"2021-03-25 08:10:00", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"} {"order_id":"202103261313130001", "order_channel":"webShop", "order_time":"2021-03-25 13:13:13", "pay_amount":"100.00", "real_pay":
"100.00", "pay_time":"2021-03-25 16:16:16", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"} {"order_id":"202103270606060001", "order_channel":"appShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"50.50", "real_pa
y":"50.50", "pay_time":"2021-03-25 06:07:00", "user_id":"0001", "user_name":"Alice", "area_id":"330106"} {"order_id":"202103270606060002", "order_channel":"webShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"66.60", "real_pay":"66.60", "pay_time":"2021-03-25 06:07:00", "user_id":"0002", "user_name":"Bob", "area_id":"330110"} {"order_id":"202103270606060003", "order_channel":"miniAppShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"88.80", "real_pay":"88.80", "pay_time":"2021-03-25 06:07:00", "user_id":"0003", "user_name":"Cindy", "area_id":"330108"} {"order_id":"202103270606060004", "order_channel":"webShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"99.90", "real_pay":"99.90", "pay_time":"2021-03-25 06:07:00", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"}
>
```

步骤6 登录Kafka实例，发送命令如下：

进入Kafka客户端文件的“/bin”目录下，执行如下命令进行生产消息：

```
./kafka-console-producer.sh --broker-list Kafka连接地址 --topic Topic名称
```

示例数据如下：

```
{"order_id":"202103241000000001", "order_channel":"webShop", "order_time":"2021-03-24 10:00:00", "pay_amount":"100.00", "real_pay":"100.00", "pay_time":"2021-03-24 10:02:03", "user_id":"0001", "user_name":"Alice", "area_id":"330106"} {"order_id":"202103241606060001", "order_channel":"appShop", "order_time":"2021-03-24 16:06:06", "pay_amount":"200.00", "real_pay":"180.00", "pay_time":"2021-03-24 16:10:06", "user_id":"0001", "user_name":"Alice", "area_id":"330106"} {"order_id":"202103251202020001", "order_channel":"miniAppShop", "order_time":"2021-03-25 12:02:02", "pay_amount":"60.00", "real_pay":"60.00", "pay_time":"2021-03-25 12:03:00", "user_id":"0002", "user_name":"Bob", "area_id":"330110"} {"order_id":"202103251505050001", "order_channel":"qqShop", "order_time":"2021-03-25 15:05:05", "pay_amount":"500.00", "real_pay":"400.00", "pay_time":"2021-03-25 15:10:00", "user_id":"0003", "user_name":"Cindy", "area_id":"330108"} {"order_id":"202103252020200001", "order_channel":"webShop", "order_time":"2021-03-24 20:20:20", "pay_amount":"600.00", "real_pay":"480.00", "pay_time":"2021-03-25 00:00:00", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"} {"order_id":"202103260808080001", "order_channel":"webShop", "order_time":"2021-03-25 08:08:08", "pay_amount":"300.00", "real_pay":"240.00", "pay_time":"2021-03-25 08:10:00", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"} {"order_id":"202103261313130001", "order_channel":"webShop", "order_time":"2021-03-25 13:13:13", "pay_amount":"100.00", "real_pay":"100.00", "pay_time":"2021-03-25 16:16:16", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"} {"order_id":"202103270606060001", "order_channel":"appShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"50.50", "real_pay":"50.50", "pay_time":"2021-03-25 06:07:00", "user_id":"0001", "user_name":"Alice", "area_id":"330106"} {"order_id":"202103270606060002", "order_channel":"webShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"66.60", "real_pay":"66.60", "pay_time":"2021-03-25 06:07:00", "user_id":"0002", "user_name":"Bob", "area_id":"330110"} {"order_id":"202103270606060003", "order_channel":"miniAppShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"88.80", "real_pay":"88.80", "pay_time":"2021-03-25 06:07:00", "user_id":"0003", "user_name":"Cindy", "area_id":"330108"} {"order_id":"202103270606060004", "order_channel":"webShop", "order_time":"2021-03-25 06:06:06", "pay_amount":"99.90", "real_pay":"99.90", "pay_time":"2021-03-25 06:07:00", "user_id":"0004", "user_name":"Daisy", "area_id":"330102"}
```

步骤7 登录Kafka控制台，单击“消息查询”可以查看消息信息。

图 3-51 消息列表

Topic 名称	分区	偏移量	消息大小 (KB)	创建时间	操作
trade_order_detail_info	0	153	240	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	152	244	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	151	238	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	150	240	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	149	242	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	148	242	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	147	242	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	146	241	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	145	241	2023/05/31 13:22:06 GMT+08:00	查看详情正文
trade_order_detail_info	0	144	242	2023/05/31 13:22:06 GMT+08:00	查看详情正文

----结束

数据湖探索 DLI 使用指南

- 步骤1** 在控制台单击“服务列表”，选择“**数据湖探索 DLI**”，单击进入DLI服务页面。更多介绍请参考**DLI介绍**。
- 步骤2** 单击DLI控制台左侧“作业管理 > Flink作业 > 创建作业”。

图 3-52 创建作业

创建作业

类型: Flink OpenSource SQL

* 名称: 请输入名称

描述: 请输入描述

模板名称: -请选择模板名称-

标签: 如果您需要使用同一标签识别多种云资源，即所有服务均可在标签输入框下拉选择同一标签，建议在TMS中创建预定义标签。查看预定义标签

在下方键/值输入框输入内容后单击添加，即可将标签加入此处

请输入标签键 请输入标签值 添加

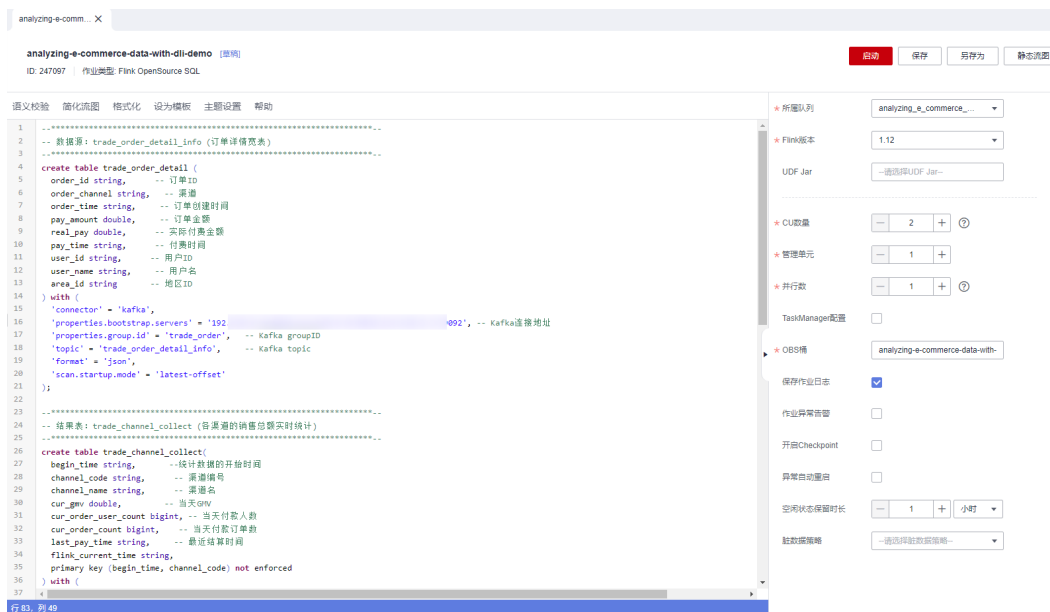
您还可以添加20个标签。

确定 取消

- 类型：选择作业类型为：Flink OpenSource SQL。
- 名称：自定义

步骤3 单击“确定”，进入作业编辑作业页面，具体SQL示例如下，部分参数值需要根据RDS和DMS对应的信息进行修改，更多SQL语法请参考[Flink OpenSource SQL 1.12语法概览](#)（注意Flink作业日志存储OBS桶建议避免选用本方案创建的OBS桶，否则会导致函数 workflows 重复调用，该动作不影响方案正常使用）。

图 3-53 Flink 作业示例



```
-----  
-- 数据源：trade_order_detail_info (订单详情宽表)  
-----  
create table trade_order_detail (  
  order_id string, -- 订单ID  
  order_channel string, -- 渠道  
  order_time string, -- 订单创建时间  
  pay_amount double, -- 订单金额  
  real_pay double, -- 实际付费金额  
  pay_time string, -- 付费时间  
  user_id string, -- 用户ID  
  user_name string, -- 用户名  
  area_id string -- 地区ID  
) with (  
  "connector" = "kafka",  
  "properties.bootstrap.servers" = "xxxx:9092,xxxx:9092,xxxx:9092", -- Kafka连接地址  
  "properties.group.id" = "xxxx", -- Kafka groupID  
  "topic" = "xxxx", -- Kafka topic  
  "format" = "json",  
  "scan.startup.mode" = "latest-offset"  
)  
;  
-----  
-- 结果表：trade_channel_collect (各渠道的销售总额实时统计)  
-----  
create table trade_channel_collect(  
  begin_time string, --统计数据的开始时间  
  channel_code string, -- 渠道编号  
  channel_name string, -- 渠道名  
  cur_gmv double, -- 当天GMV  
  cur_order_user_count bigint, -- 当天付款人数  
  cur_order_count bigint, -- 当天付款订单数  
);
```

```

last_pay_time string, -- 最近结算时间
flink_current_time string,
primary key (begin_time, channel_code) not enforced
) with (
"connector" = "jdbc",
"url" = "jdbc:mysql://xxx:3306/xxx", -- mysql连接地址及mysql表所在数据库名, jdbc格式
"table-name" = "xxx", -- mysql表名
"driver" = "com.mysql.jdbc.Driver",
"username" = "xxx", -- mysql用户名
"password" = "xxx", -- mysql密码
"sink.buffer-flush.max-rows" = "1000",
"sink.buffer-flush.interval" = "1s"
);
--*****_
-- 临时中间表
--*****_
create view tmp_order_detail
as
select *
, case when t.order_channel not in ("webShop", "appShop", "miniAppShop") then "other"
else t.order_channel end as channel_code --重新定义统计渠道 只有四个枚举值[webShop、appShop、
miniAppShop、other]
, case when t.order_channel = "webShop" then _UTF16"网页商城"
when t.order_channel = "appShop" then _UTF16"app商城"
when t.order_channel = "miniAppShop" then _UTF16"小程序商城"
else _UTF16"其他" end as channel_name --渠道名称
from (
select *
, row_number() over(partition by order_id order by order_time desc ) as rn --去除重复订单数据
, concat(substr("2021-03-25 12:03:00", 1, 10), " 00:00:00") as begin_time
, concat(substr("2021-03-25 12:03:00", 1, 10), " 23:59:59") as end_time
from trade_order_detail
where pay_time >= concat(substr("2021-03-25 12:03:00", 1, 10), " 00:00:00") --取今天数据, 为了方便运行, 这里使用"2021-03-25 12:03:00"替代cast(LOCALTIMESTAMP as string)
and real_pay is not null
) t
where t.rn = 1;
-- 按渠道统计各个指标
insert into trade_channel_collect
select
begin_time --统计数据的开始时间
, channel_code
, channel_name
, cast(COALESCE(sum(real_pay), 0) as double) as cur_gmv --当天GMV
, count(distinct user_id) as cur_order_user_count --当天付款人数
, count(1) as cur_order_count --当天付款订单数
, max(pay_time) as last_pay_time --最近结算时间
, cast(LOCALTIMESTAMP as string) as flink_current_time --flink任务中的当前时间
from tmp_order_detail
where pay_time >= concat(substr("2021-03-25 12:03:00", 1, 10), " 00:00:00")
group by begin_time, channel_code, channel_name;

```

表 3-2 数据源表：电商业务订单详情宽表

字段名	字段类型	说明
order_id	string	订单ID
order_channel	string	订单生成的渠道(即web方式、app方式等)
order_time	string	订单时间
pay_amount	double	订单金额

字段名	字段类型	说明
real_pay	double	实际支付金额
pay_time	string	支付时间
user_id	string	用户ID
user_name	string	用户姓名
area_id	string	订单地区ID

表 3-3 结果表：各渠道的销售总额实时统计表。

字段名	字段类型	说明
begin_time	varchar(32)	开始统计指标的时间
channel_code	varchar(32)	渠道编号
channel_name	varchar(32)	渠道名
cur_gmv	double	当天GMV
cur_order_user_count	bigint	当天付款人数
cur_order_count	bigint	当天付款订单数
last_pay_time	varchar(32)	最近结算时间
flink_current_time	varchar(32)	Flink数据处理时间

📖 说明

作业逻辑说明如下：

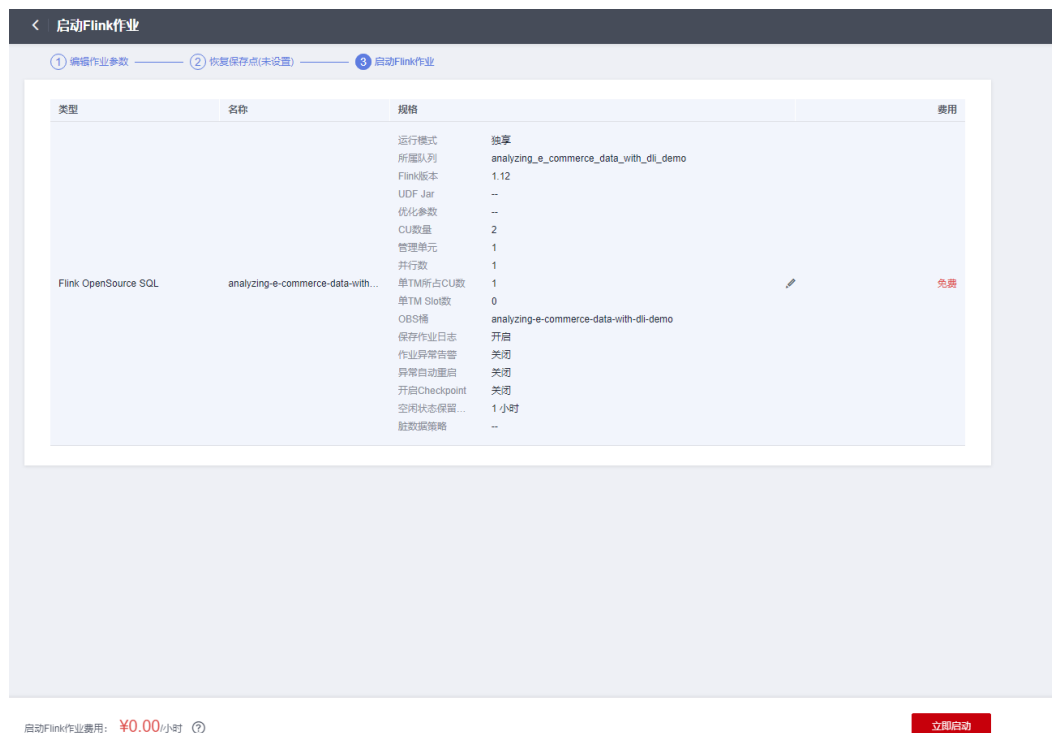
1. 创建一个Kafka源表，用来从Kafka指定Topic中读取消费数据；
2. 创建一个结果表，用来通过JDBC向MySQL中写入结果数据。
3. 实现相应的处理逻辑，以实现各个指标的统计。

为了简化最终的处理逻辑，使用创建视图进行数据预处理。

1. 利用over窗口条件和过滤条件结合以去除重复数据（该方式是利用了top N的方法），同时利用相应的内置函数concat和substr将当天的00:00:00作为统计的开始时间，当天的23:59:59作为统计结束时间，并筛选出支付时间在当天凌晨00:00:00后的订单数据进行统计（为了方便模拟数据的构造，这里使用"2021-03-25 12:03:00"替代cast(LOCALTIMESTAMP as string)）。
2. 根据这些数据的订单渠道利用内置的条件函数设置channel_code和channel_name的值，从而获取了源表中的字段信息，以及begin_time、end_time和channel_code、channel_name的值。
4. 根据需要对相应指标进行统计和筛选，并将结果写入到结果表中。

步骤4 选择所创建的DLI通用队列提交作业。

图 3-54 提交作业



步骤5 等待作业状态会变为“运行中”，单击作业名称，可以查看作业详细运行情况。

图 3-55 作业运行状态



----结束

配置数据可视化 DLV 大屏

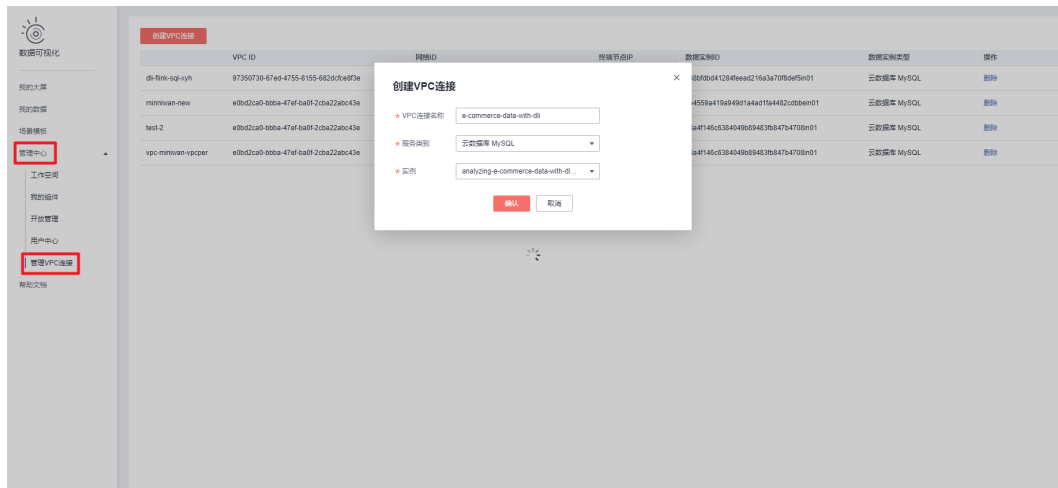
步骤1 配置DLV大屏。在控制台单击“服务列表”，选择“数据可视化 DLV”，单击进入DLV页面。更多介绍请参考[DLV介绍](#)。

图 3-56 DLV 主页



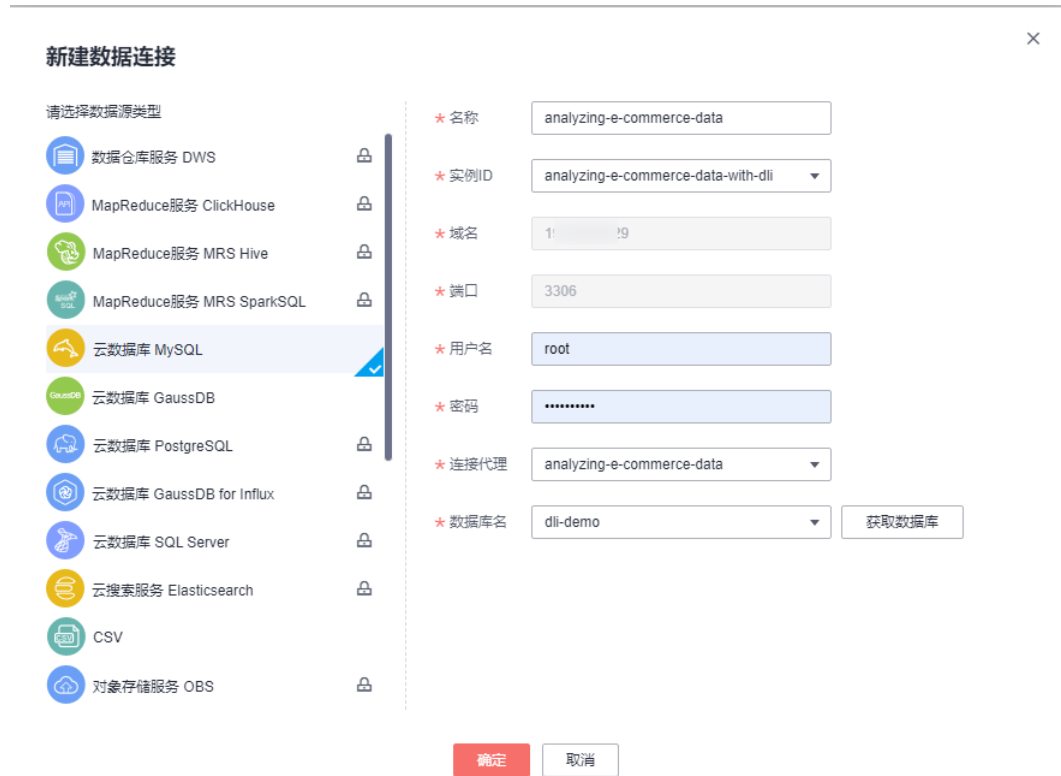
步骤2 创建VPC连接。单击“管理中心”>“管理VPC连接”，单击“创建VPC连接”。

图 3-57 VPC 链接



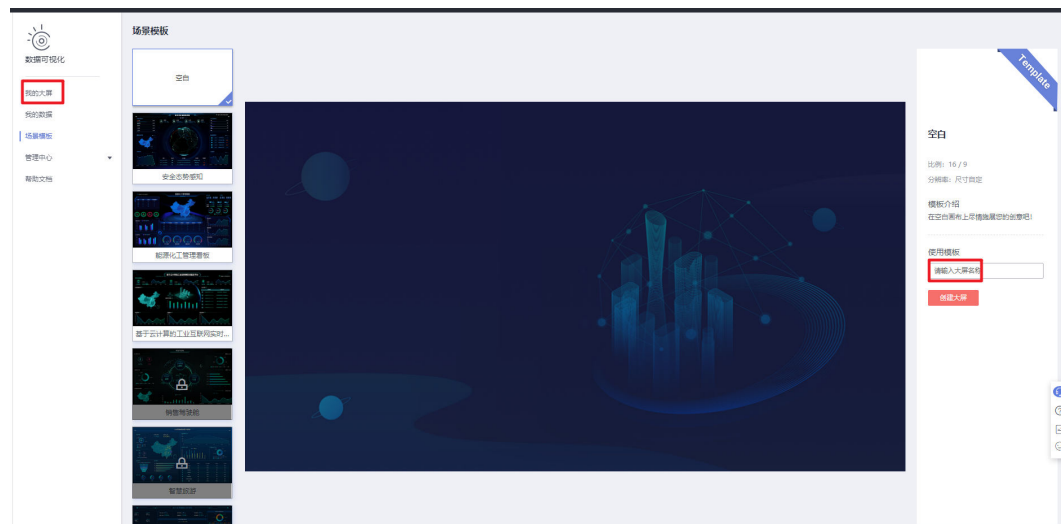
步骤3 单击“我的数据”>“新建数据连接”，填写相关值，单击“确定”（名称支持数字、字母、中文、_（下划线）、-（中划线），长度1-32位）。

图 3-58 新建数据连接



步骤4 具体配置方法可参考[DLV开发大屏](#)。单击“我的大屏”>“新建大屏”，选择空白模板，填写大屏名称并创建大屏。

图 3-59 新建数据连接



步骤5 选择大屏，单击编辑图标，进入大屏开发配置，例如现状图，单击“数据”导入 MySQL 数据，执行 SQL 查询 RDS MySQL，并设置 x、y、s 坐标轴具体，样式设置参考[组件指南](#)，预览大屏实时展示。

图 3-60 新建数据连接

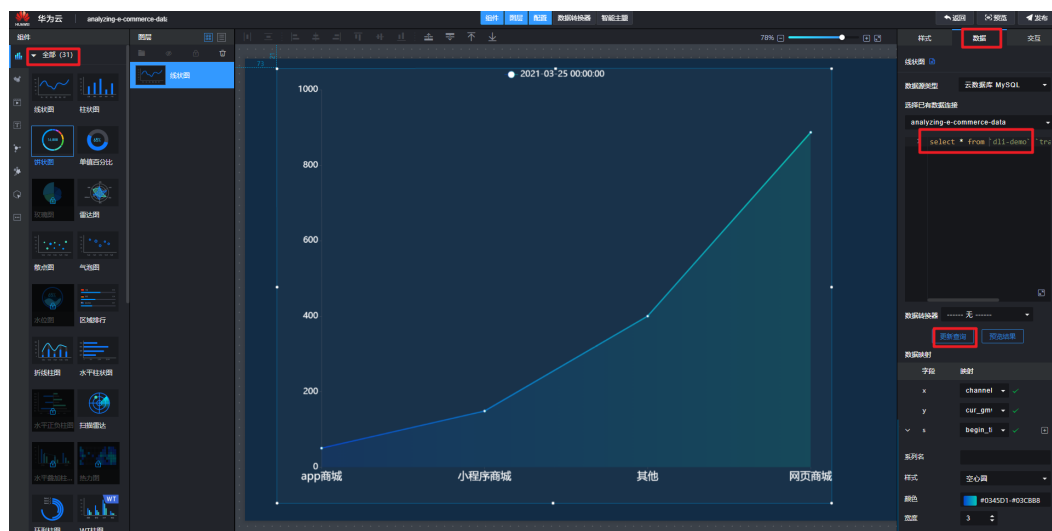
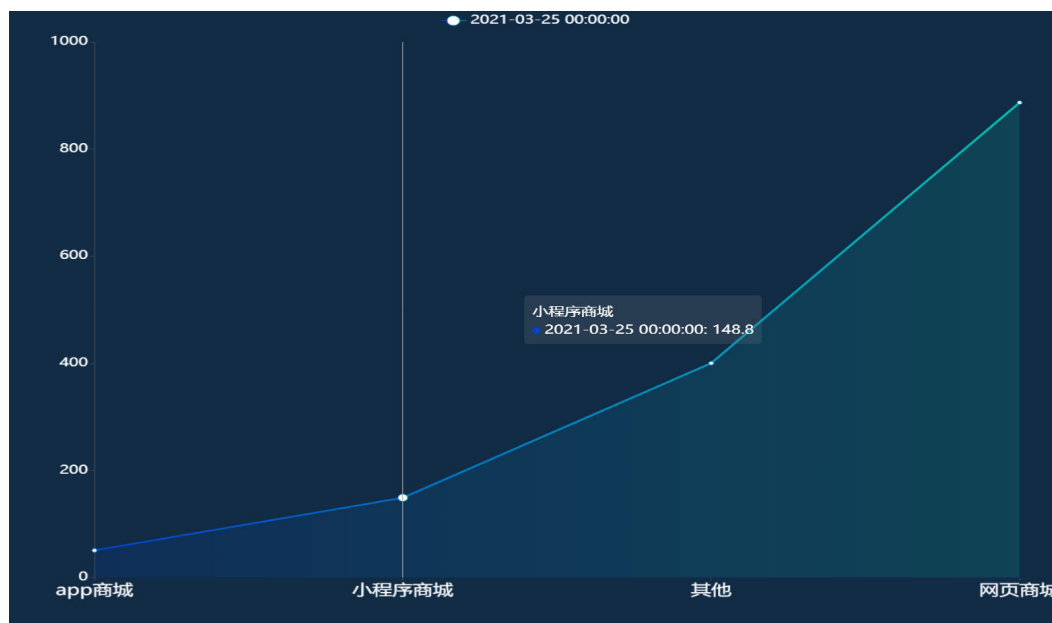


图 3-61 预览效果



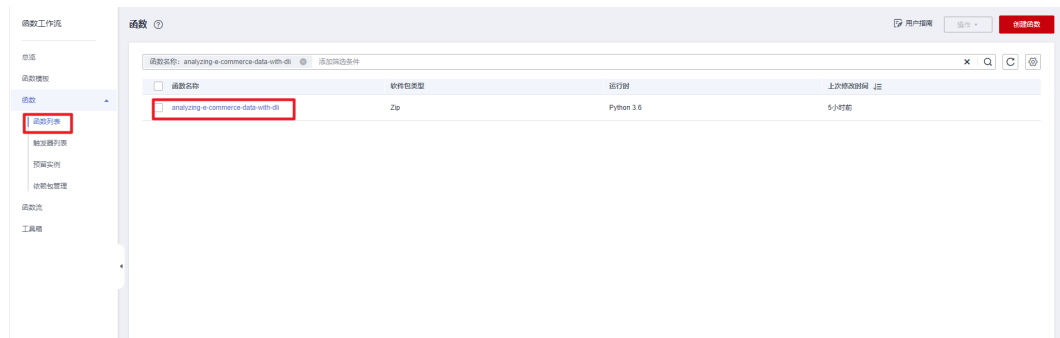
----结束

函数 workflow FunctionGraph 定制化开发

在实际使用过程中，用户可以根据自己的业务场景参考[函数 workflow FunctionGraph](#)对代码进行定制化开发。

- 步骤1** 在控制台单击“服务列表”，选择“[函数 workflow FunctionGraph](#)”，单击进入函数主页面，单击“函数”>“函数列表”，打开该解决方案创建的函数。

图 3-62 函数



步骤2 函数代码相关配置。用户可以根据实际情况，进行二次定制化代码开发及参数配置。

图 3-63 函数代码

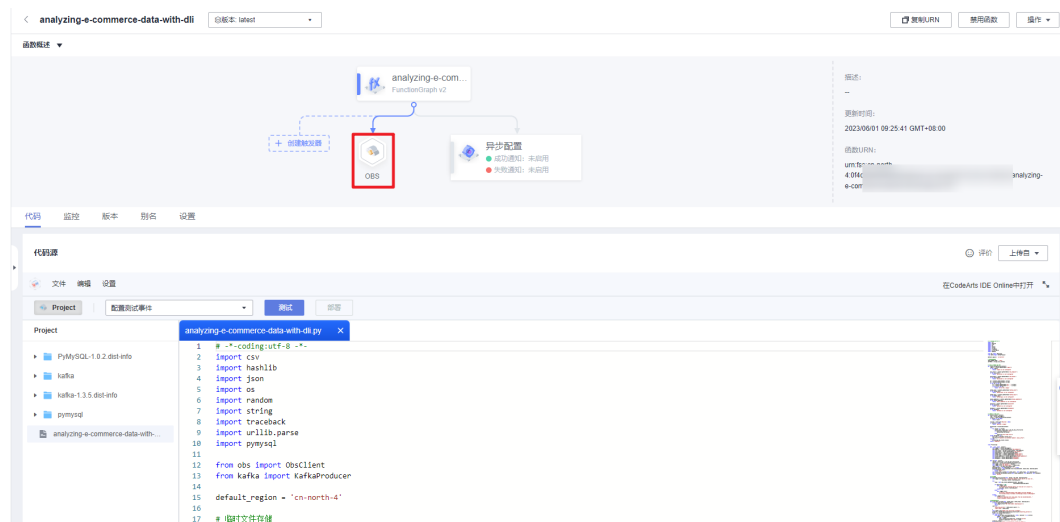
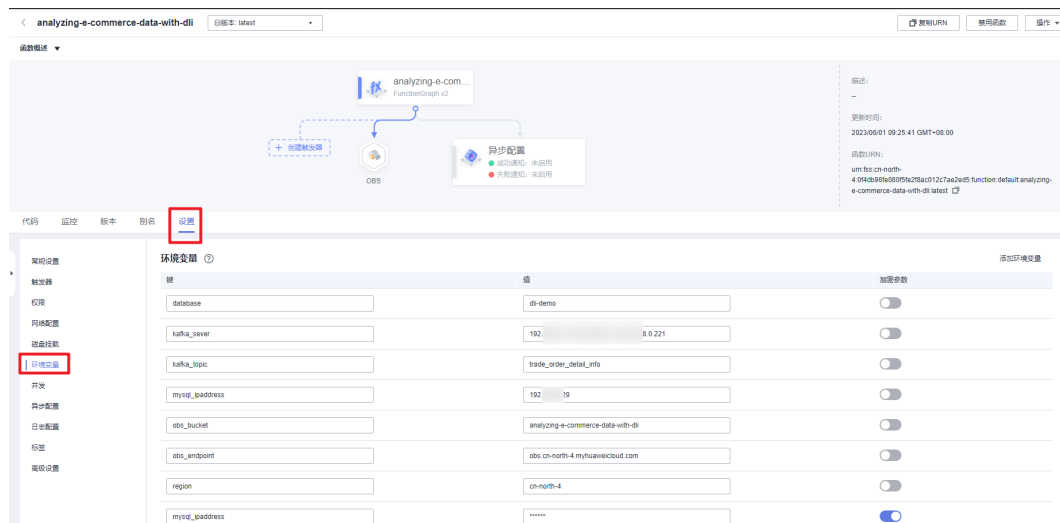


图 3-64 环境变量



----结束

3.4 快速卸载

说明

- 该解决方案涉及到对象存储服务 OBS桶，如果OBS桶中有数据的话会导致资源栈删除失败。请确保数据以及迁移备份后清空OBS桶中的数据，再卸载该解决方案。
- 如果在DLI下创建了Flink作业，请先删除该作业，再卸载该解决方案。

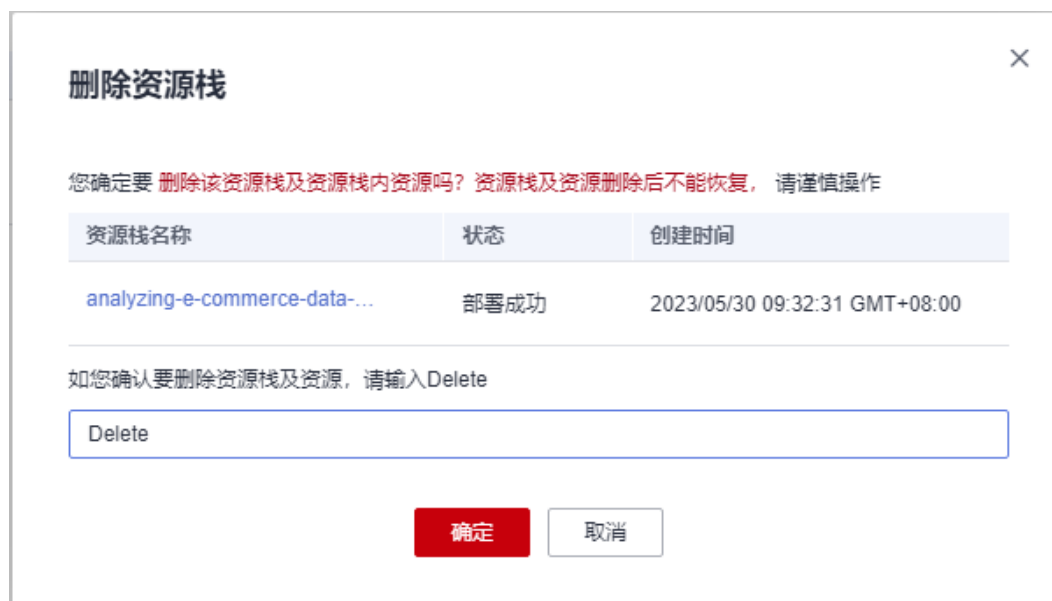
步骤1 解决方案部署成功后，单击该方案堆栈后的“删除”。

图 3-65 一键卸载



步骤2 在弹出的删除堆栈确认框中，输入Delete，单击“确定”，即可卸载解决方案。

图 3-66 删除堆栈确认



----结束

4 附录

名词解释

基本概念、云服务简介、专有名词解释

- 对象存储服务 OBS: 对象存储服务 (Object Storage Service, OBS) 是一个基于对象的存储服务, 为客户提供海量、安全、高可靠、低成本的数据存储能力, 使用时无需考虑容量限制, 并且提供多种存储类型供选择, 满足客户各类业务场景诉求。
- 数据湖探索 DLI: 数据湖探索 (Data Lake Insight, 简称DLI) 是完全兼容Apache Spark、Apache Flink、openLookeng (基于Presto) 生态, 提供一站式的流处理、批处理、交互式分析的Serverless融合处理分析服务, 支持数据入湖、数据仓库、BI、AI融合等能力。
- 云数据库 RDS for MySQL: 云数据库 RDS for MySQL拥有即开即用、稳定可靠、安全运行、弹性伸缩、轻松管理、经济实用等特点, 让您更加专注业务发展。
- 分布式消息服务Kafka版: 分布式消息服务Kafka版是一个高吞吐、高可用的消息中间件服务, 适用于构建实时数据管道、流式数据处理、第三方解耦、流量削峰去谷等场景, 具有大规模、高可靠、高并发访问、可扩展且完全托管的特点, 是分布式应用上云必不可少的重要组件。
- 数据可视化 DLV: 数据可视化服务 (Data Lake Visualization) 是一站式数据可视化平台, 适配云上云下多种数据源, 提供丰富多样的2D、3D可视化组件, 采用拖拽式自由布局, 旨在帮助您快速定制和应用属于您自己的数据大屏。
- 函数工作流 FunctionGraph: 函数工作流 (FunctionGraph) 是一项基于事件驱动的函数托管计算服务。通过函数工作流, 只需编写业务函数代码并设置运行的条件, 无需配置和管理服务器等基础设施, 函数以弹性、免运维、高可靠的方式运行。此外, 按函数实际执行资源计费, 不执行不产生费用。

5 修订记录

表 5-1 修订记录

发布日期	修订记录
2023-05-30	第一次正式发布。