

# 多数据源联合分析查询操作指导

## 1 任务介绍

### 1.1 客户/订单信息管理系统

#### 1.1.1 功能概述

这是一个基于HBase组合rowkey的跨源场景案例分析。DLI提供信息的计算能力，CloudTable的Hbase数据库提供了快速点查能力。

**RowKey**是HBase中行的主键，HBase中的快速点查最佳方案是使用RowKey。当传统关系型数据库中需要使用多个字段进行联合过滤查询时，就需要使用到HBase的组合RowKey能力。因此RowKey的设计是至关重要的，关系到应用层的查询效率。而在DLI on CloudTable场景中，支持的组合RowKey能力，能够通过简单的SQL语句轻松实现。

#### 1.1.2 场景介绍

该场景中，我们将得到两份表单数据：客户信息和订单信息。

客户信息包含如下内容：

- 客户编号
- 客户名称
- 地址
- 电话号码

订单信息包含如下内容：

- 客户编号
- 订单编号
- 销售日期
- 货品信息（如物品装运位置、物品名称等，此处我们将简单设置成一个String类型的消息）

其中，客户编号将确定唯一的一条客户信息，客户编号 + 订单编号将确定唯一的一条订单信息。

### 1.1.3 设计思路

在该方案中，我们将客户信息作为内表存储在DLI中，将订单信息作为外表存储在CloudTable HBase中。其中，“客户编号 + 订单编号”将作为RowKey。

客户信息表（CUSTOMER\_INFO）的Schema信息如下：

标签	客户编号	客户名称	地址	电话号码
对应字段名	ID	NAME	LOCATION	PHONE

表1

订单信息表（ORDER\_INFO）的Schema信息如下：

标签	客户编号	订单编号	销售日期	货品信息
对应字段名	CUST_ID	ORDER_ID	SALE_DATE	ART_INFO

表2

订单信息表(ORDER\_INFO\_IN\_CT)在HBase中的逻辑视图如下：

标签	客户编号	订单编号	销售日期	货品信息
对应字段	RowKey		CF.DATE	CF.INFO

表3

## 2 任务执行

### 1. 前提条件

#### (1) 创建CloudTable集群

- 登录华为云，在华为云官网页面上方的导航栏，选择“产品”。
- 在“EI 企业智能”列表中，单击“表格存储服务”进入表格存储服务产品页面。

或直接点击链接进入表格存储服务产品页面：

<https://www.huaweicloud.com/product/cloudtable.html>

- 若导航栏中显示的区域不是“北京一”，点击切换区域到北京一。

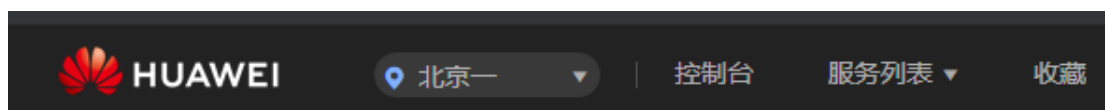


图 1 切换区域到北京一

- 在“表格存储服务”页面，单击“立即使用”进入总览页面，单击“购买集群”进入购买集群页面。若已有集群，可不购买，使用已有的即可。

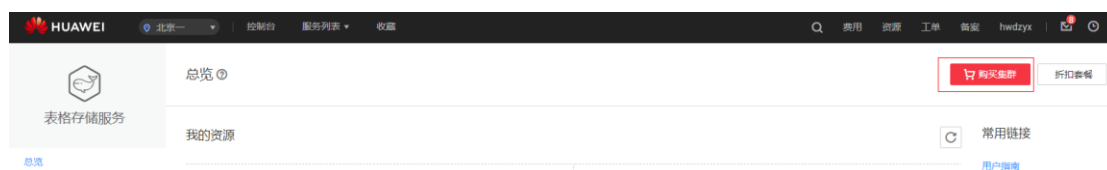


图 2 购买集群


在购买集群页面，参考图3进行集群设置。地域选择“北京一”，可用分区选择“可用区1”，集群配置、软件配置、计算单元使用默认配置。在网络配置中，若虚拟私有云显示为空，单击网络配置中的“创建虚拟私有云”链接，先创建一个虚拟私有云（参考图4），创建完成后再回到本页面点击虚拟私有云的刷新图标，将自动选择创建的虚拟私有云。若已有私有云，可不创建，选择使用默认显示的即可。配置完成后单击“立即购买”进入规格确认页面，单击“提交”，完成表格存储服务集群创建。

## 区域

地域

北京一


温馨提示：页面左上角切换区域

\* 可用分区 


可用区1

可用区2

## 集群配置

\* 名称 

cloudtable-7053

\* 存储配额 


- 800 + GB

申请更多存储配额请单击[申请扩大配额](#)。CloudTable按照用户实际占用的存储空间进行收费。\* I/O类型 

普通I/O

超高I/O


## 网络配置

\* 虚拟私有云 

vpc-77f9

[创建虚拟私有云](#)\* 子网 

subnet-7802 (192.168....)

\* 安全组 

Sys-default

[创建安全组](#)

图 3 购买集群

创建虚拟私有云 <sup>①</sup> < 返回虚拟私有云列表

基本信息

区域

华北-北京一

不同区域的资源之间内网不互通。请选择靠近您客户的区域，可以降低网络时延、提高访问速度。

\* 名称

vpc-77f9

\* 网段

192 · 168 · 0 · 0 / 16

建议使用网段：10.0.0.0/8~24，172.16.0.0/12~24，192.168.0.0/16~24

标签

如果您需要使用同一标签标识多种云资源，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](#)

标签键 标签值

您还可以添加10个标签。

子网配置

默认子网

可用区 <sup>②</sup>

可用区2 可用区1 可用区3

\* 名称

subnet-7802

\* 子网网段

192 · 168 · 0 · 0 / 24

可用IP数:250 子网创建完成后，子网网段无法修改

高级配置

默认配置 自定义配置

添加子网

一次最多可以添加3个子网。

立即创建

图 4 创建虚拟私有云

- e) 单击表格存储服务总览页面的集群管理，可查看到集群正在创建，集群创建需要一段时间，请等待约10分钟，即可查看到创建完成的集群。单击创建的集群，可查看集群的相关信息。

集群管理 &gt; cloudtable-c916

## 集群信息

集群名称：	cloudtable-c916	集群ID：	066cd96c-f067-477b-ab16-b0d419965fd0
集群状态：	● 服务中	已使用存储容量 (GB)：	0.313
RS单元数量：	2	存储配额 (GB)：	800
TSD单元数量：	0	I/O类型：	普通I/O
计费模式：	按需	CloudTable版本：	v1.1.10
创建时间：	2018/11/01 16:03:33 GMT+08:00	OpenTSDB链接地址：	-
ZK链接地址：	cloudtable-c916-zk2-eNZUPBUV.cloudtable.co... 	IAM统一身份认证：	关闭

## 网络配置

地域：	华北-北京一	可用分区：	可用区1
虚拟私有云：	test-peer	子网：	subnet-b15d (192.168.1.0/24)
安全组：	dws-dws-demo-8000		

图 5 集群信息

(2) 购买弹性云服务器，用于通过HBase Shell访问表格存储服务集群。

- a) 登录华为云，在华为云官网页面上方的导航栏，选择“产品”。在“计算”列表中，单击“弹性云服务器 ECS”进入弹性云服务器产品页面，单击“立即购买”进入弹性云服务器购买配置页面，计费模式选择“按需计费”，区域选择“华北-北京一”，可用区选择“可用区 1”，规格使用默认规格“通用计算型 |s2.small.1|1vCPUs|1GB”，镜像选择“公共镜像|CentOS 7.5 64bit(40GB)”，虚拟私有云选择 (1) 中创建的虚拟私有云。密码和确认密码中配置服务器密码，单击“立即购买”到规格确认页面，单击“提交”，提交后返回云服务器列表，可查看到购买的弹性云服务器，单击“远程登录”，使用 root 用户名和配置的密码，可登录创建的云服务器后台，如图 6-9 所示，单击登录界面右上方的“Input Commands”，可粘贴命令在云服务器中执行。

购买弹性云服务器

计费模式

包年/包月

按需计费

区域

华北-北京一

可用区

可用区2

可用区1

可用区3

规格

最新系列

vCPUs 全部

内存 全部

请输入规格名称

通用计算型

通用计算增强型

内存优化型

高性能计算型

磁盘增强型

超高性能型

GPU加速型

FPGA加速型

通用入门型

了解如何选择弹性云服务器类型

规格名称	vCPUs/内存	基准/最大带宽	内网收发包
<input checked="" type="radio"/> s2.small.1	1vCPUs   1GB	0.1/0.5 Gbit/s	50,000 pps
<input type="radio"/> s2.medium.2	1vCPUs   2GB	0.1/0.5 Gbit/s	50,000 pps
<input type="radio"/> s2.medium.4	1vCPUs   4GB	0.1/0.5 Gbit/s	50,000 pps
<input type="radio"/> s2.large.2	2vCPUs   4GB	0.2/0.8 Gbit/s	100,000 pps
<input type="radio"/> s2.large.4	2vCPUs   8GB	0.2/0.8 Gbit/s	100,000 pps

当前配置

计费模式

按需计费

地域

北京一

可用区

可用区1

云服务器名称

ecs-7582

规格

通用计算型 | s2.small.1 | 1vCPUs | 1GB

镜像

CentOS 7.5 64bit

系统盘

高IO, 40GB

备份策略

defaultPolicy | 启用 | 02:45 | 每天 | 90天

虚拟私有云

test-peer

安全组

Sys-default

网卡

subnet-b15d(192.168.1.0/24)

弹性公网IP

规格: 静态BGP  
计费方式: 按带宽计费 带宽: 5 Mbit/s

购买数量

1

配置费用

¥0.35/小时

云服务器备份套餐包超出部分按需费用

¥0.268/GB/月

以上是参考价格, 具体扣费请以账单为准。

了解计费详情

立即购买

图 6 购买弹性云服务器

虚拟私有云

test-peer

新建虚拟私有云

网卡

主网卡

subnet-b15d(192.168.1.0/24)

自动分配IP地址

查看已使用IP地址

增加一块网卡

您还可以增加 11 块网卡

安全组

如何配置安全组?

Sys-default (入方向: TCP/8000, 3389, 22 | 出方向: ...)

新建安全组

弹性公网IP

如有互联网访问需求, 请先规划您的弹性公网IP资源。单击[这里](#)查看弹性公网IP。

现在购买

使用已有

暂不购买

规格

静态BGP

全动态BGP

带宽类型

独享带宽

共享带宽

计费方式

按带宽计费

按流量计费

带宽

5 Mbit/s

1 100 200 300 500 1,000 2,000 5 Mbit/s

当前配置

计费模式

按需计费

地域

北京一

可用区

可用区1

云服务器名称

ecs-7582

规格

通用计算型 | s2.small.1 | 1vCPUs | 1GB

镜像

CentOS 7.5 64bit

系统盘

高IO, 40GB

备份策略

defaultPolicy | 启用 | 02:45 | 每天 | 90天

虚拟私有云

test-peer

安全组

Sys-default

网卡

subnet-b15d(192.168.1.0/24)

弹性公网IP

规格: 静态BGP  
计费方式: 按带宽计费 带宽: 5 Mbit/s

购买数量

1

配置费用

¥0.35/小时

云服务器备份套餐包超出部分按需费用

¥0.268/GB/月

以上是参考价格, 具体扣费请以账单为准。

了解计费详情

立即购买

图 7 购买弹性云服务器

登录方式

密码

密钥对

用户名

root

密码

请设置密码, 密码长度至少8位, 且包含大写字母、小写字母、数字和特殊字符。

确认密码

密码

密码

高级配置

暂不配置

现在配置

云服务器名称

ecs-7582

允许重名

购买数量

1

镜像

CentOS 7.5 64bit

系统盘

高IO, 40GB

备份策略

defaultPolicy | 启用 | 02:45 | 每天 | 90天

虚拟私有云

test-peer

安全组

Sys-default

网卡

subnet-b15d(192.168.1.0/24)

弹性公网IP

规格: 静态BGP  
计费方式: 按带宽计费 带宽: 5 Mbit/s

购买数量

1

配置费用

¥0.35/小时

云服务器备份套餐包超出部分按需费用

¥0.268/GB/月

以上是参考价格, 具体扣费请以账单为准。

了解计费详情

立即购买

图 8 购买弹性云服务器



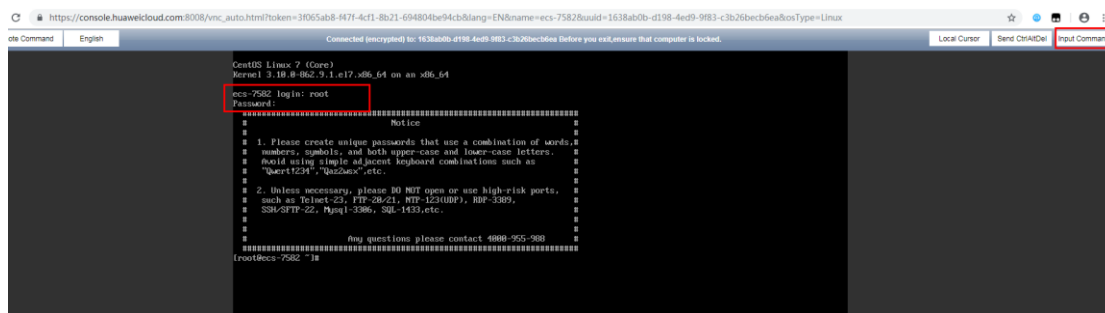


图 9 弹性云服务器后台

(3) 通过弹性云服务器访问表格存储服务集群。

访问[https://support.huaweicloud.com/usermanual-cloudtable/cloudtable\\_01\\_0097.html](https://support.huaweicloud.com/usermanual-cloudtable/cloudtable_01_0097.html)，根据指导，使用一键部署客户端的方法使用HBase Shell访问集群，注意bin/hbase shell命令要在hbase目录中执行，如图10-12。

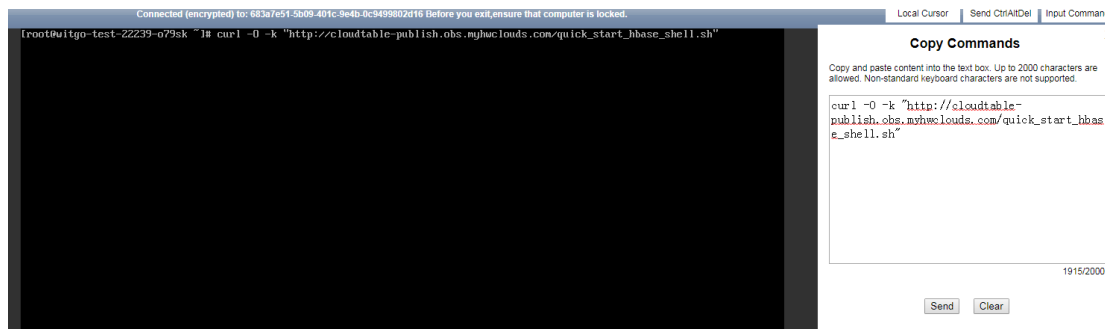


图 10 下载一键式部署工具

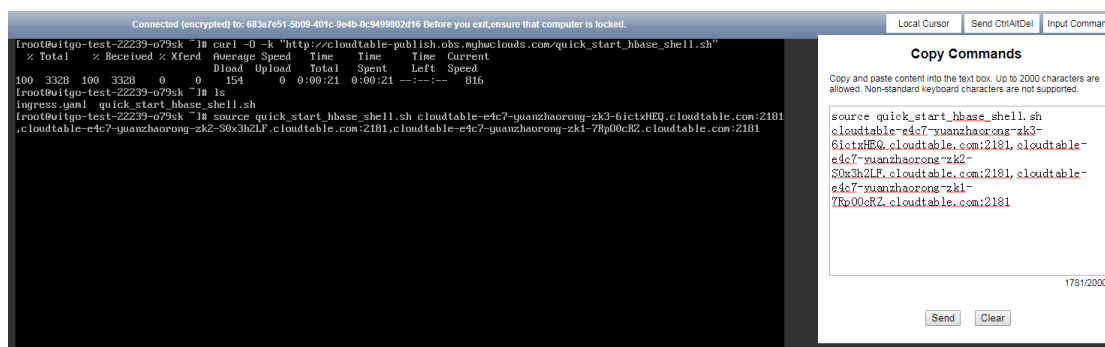


图 11 使用工具部署客户端

```
[root@ecs-7582 ~]# bin/hbase shell
-bash: bin/hbase: No such file or directory
[root@ecs-7582 ~]# ls
anaconda-ks.cfg          hbase                    jdk-8u151-linux-x64.tar.gz
geomesa-hbase_2.11-1.3.4 hbase-1.3.1-bin.tar.gz  quick_start_hbase_shell.sh
[root@ecs-7582 ~]# cd hbase
[root@ecs-7582 hbase]# bin/hbase shell
```

图 12 访问hbase集群

(4) 创建订单信息表。

在CloudTable集群中创建相关的HBase表ORDER\_INFO\_IN\_CT。

HBase相关建表语句：

```
create 'ORDER_INFO_IN_CT','CF'
```

```
hbase(main):014:0> create 'ORDER_INFO_IN_CT','CF'
2018-11-01 22:41:33,279 INFO [main] client.HBaseAdmin: Created ORDER_INFO_IN_CT
0 row(s) in 1.2200 seconds
=> Hbase::Table - ORDER_INFO_IN_CT
hbase(main):015:0>
```

图 13 HBase中创建表

## 2. 创建队列

(1) 进入数据湖探索 (DLI) 控制台。

a) 在华为云官网页面上方的导航栏，选择“EI 企业智能”。

b) 在“EI 大数据”列表中，单击“数据湖探索”。

或直接点击链接：



<https://www.huaweicloud.com/product/dli.html>

c) 在“数据湖探索 DLI”页面，单击“进入控制台”。

d) 在 DLI 控制台，单击 SQL 作业的“创建作业”超链接（如图 14），进入 SQL 作业的“作业编辑器”界面。



图 14 DLI控制台

在左侧导航栏中，单击  选择队列，单击 （如图 15），参考图 16 创建容量为 4CU 的队列 test。

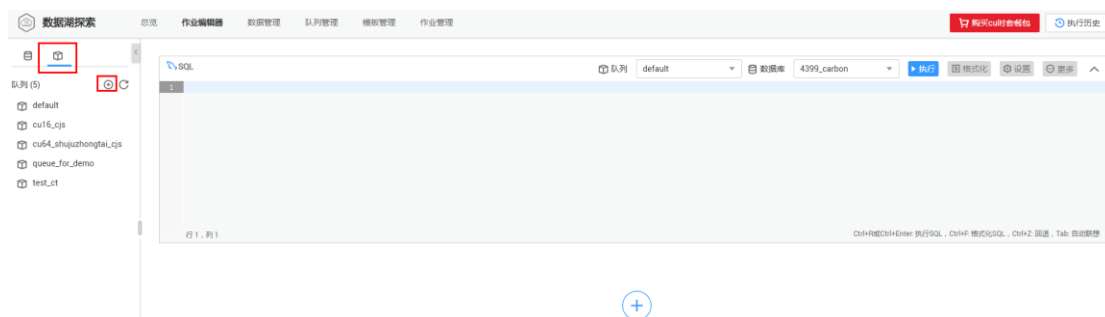


图 15 创建队列操作

## 创建队列

提示：剩余CU配额为95，1CU=4核16GB。 [了解更多...](#)

\* 队列名称

test

\* 队列容量

4CU

16CU

64CU

\* 计费方式

按CU时

按SQL计算量

描述



0/256

确定

取消

图 16 创建队列界面

### 3. 创建数据库

在“作业编辑器”左侧导航栏中，单击  选择数据库，单击 （如图 17）创建数据库 DB1。

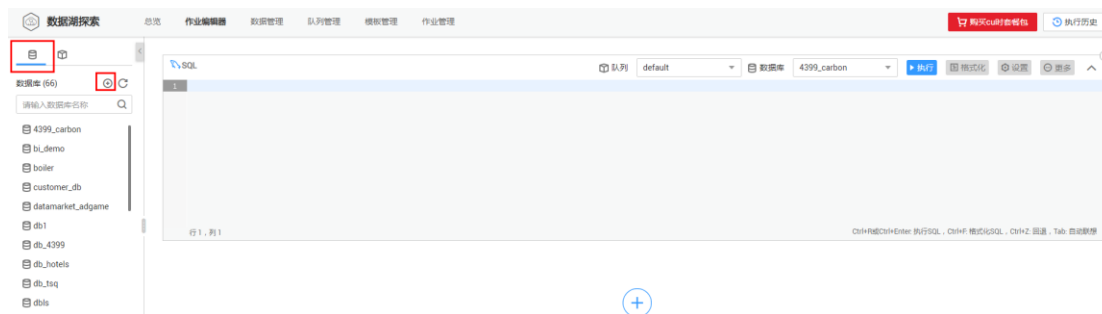


图 17 创建数据库操作

创建数据库

提示：您还可以创建27个数据库。

\* 数据库名称

DB1

描述

test

4/100

确定

取消

图 18 创建数据库

数据库创建成功后，创建的数据库 DB1 会显示在左侧导航栏“数据库”列表中。

### 4. 创建客户信息表（内表）

在作业编辑器界面，选择创建的队列和数据库，在SQL编辑器中输入如下

内容并执行：

```
CREATE TABLE IF NOT EXISTS CUSTOMER_INFO (  
ID STRING,  
NAME STRING,  
LOCATION STRING,  
PHONE STRING)
```

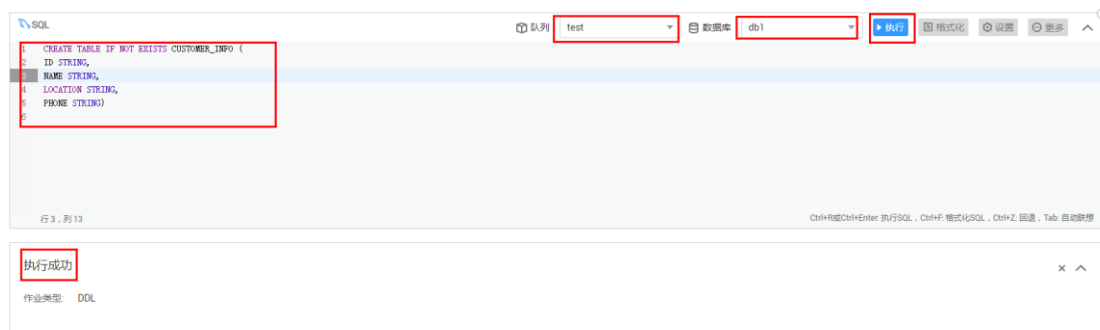


图 19 创建客户信息表

## 5. 创建订单信息表

在SQL编辑器中输入如下内容（注意红色部分为创建的表格存储服务集群

ID，见图5所示）并执行，在HBase中创建ORDER\_INFO\_IN\_CT：

```
CREATE TABLE IF NOT EXISTS ORDER_INFO (  
CUST_ID STRING,  
ORDER_ID STRING,  
SALE_DATE STRING,  
ART_INFO STRING) USING CLOUDTABLE OPTIONS (  
'ClusterId'='066cd96c-f067-477b-ab16-b0d419965fd0',  
'TableName'='ORDER_INFO_IN_CT',
```

```
'RowKey'='CUST_ID:1, ORDER_ID:2',  
'Cols'='SALE_DATE:CF.DATE,ART_INFO:CF.INFO')
```

## 6. 插入客户信息数据

假设有一条用户信息如下：

**ID: 1, NAME: 华为, LOCATION: 杭州市滨江区江虹路410号, PHONE: 0571xxxxxxxx**

在SQL编辑器中输入SQL语句并执行：

```
INSERT INTO CUSTOMER_INFO VALUES("1","华为","杭州市滨江区江  
虹路410号","0571xxxxxxxx")
```

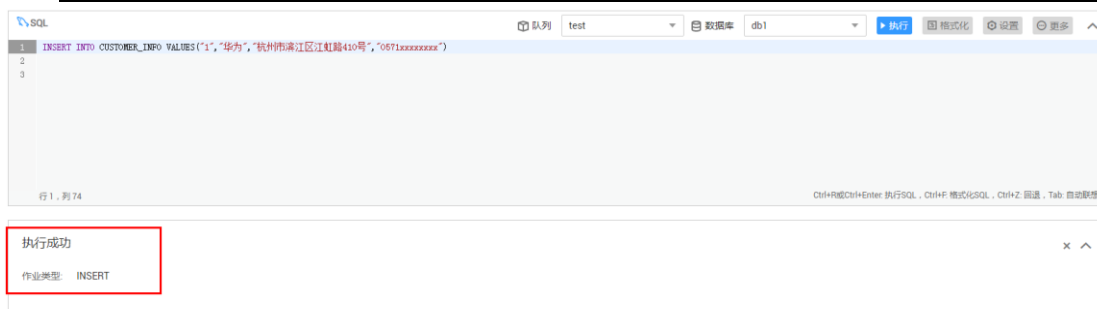


图 20 插入客户信息

## 7. 插入订单数据

假设有一条订单信息如下：

**CUST\_ID: 1, ORDER\_ID: 01, SALE\_DATE: 2018-01-01, ART\_INFO: XXX**

在SQL作业编辑器中输入SQL并执行：

```
INSERT INTO ORDER_INFO VALUES("1", "01", "2018-01-01",  
"XXX")
```

## 8. 生成订单派送信息

订单派送信息将包含客户信息与订单信息，需要使用CT表与DLI内表的关联在SQL作业编辑器中输入SQL并执行：

```
SELECT * FROM CUSTOMER_INFO JOIN ORDER_INFO WHERE  
  
ORDER_INFO.CUST_ID='1' AND  
  
ORDER_INFO.ORDER_ID='01' AND  
  
CUSTOMER_INFO.ID=ORDER_INFO.CUST_ID
```

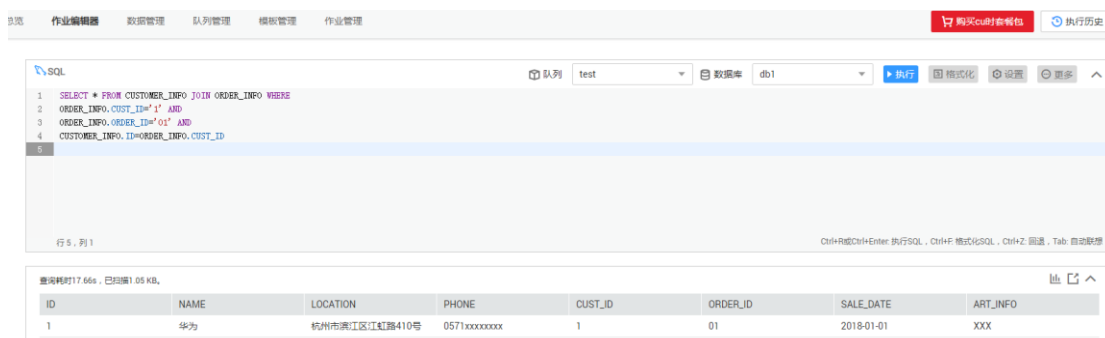


图 21 获取订单派送信息

## 3 打卡任务

**任务描述：**在上述案例中，插入一条新的订单数据，生成新订单的派送信息。

订单数据如下：

CUST\_ID: 1, ORDER\_ID: 02, SALE\_DATE: 2018-01-02, ART\_INFO: YYY

**打卡要求：**对包含查询结果的界面进行截图，截图中需包含学员的华为云用户名信息。参考下图：



控制台

服务列表

收藏

常用

资源

工单

告警

11

执行历史

作业编辑器 数据管理 队列管理 模板管理 作业管理

购买GaussDB

SQL

队列 test

数据库 db1

执行

格式化

设置

更多

1 SELECT \* FROM CUSTOMER\_INFO JOIN ORDER\_INFO WHERE

2 ORDER\_INFO.CUST\_ID='1' AND

3 ORDER\_INFO.ORDER\_ID='02' AND

4 CUSTOMER\_INFO.ID=ORDER\_INFO.CUST\_ID

5

6

行 3, 列 30

Ctrl+R或Ctrl+Enter 执行SQL, Ctrl+F 格式化SQL, Ctrl+Z 后退, Tab 自动联想

查询耗时11.29s, 已扫描1.05 KB.

ID	NAME	LOCATION	PHONE	CUST_ID	ORDER_ID	SALE_DATE	ART_INFO
1	华为	杭州市滨江区江虹路410号	0571xxxxxxxx	1	02	2018-01-02	YYY