

1 实验目标

1.1 实验背景介绍

本实验为课程《ACA21107 深入理解阿里云产品-Data IDE》中的配套实验。
主要阐述在使用 Data IDE 过程中如何创建一次性调度 workflow 来适用于在两个项目空间中批量建表的场景。

1.2 实验环境架构

实验环境架构：阿里云开放数据处理服务 ODPS

1.3 云端服务环境

暂无

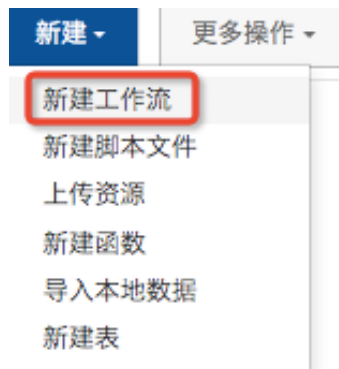
1.4 云端开发桌面环境

2 实验内容

2.1 新建 workflow

前提条件：用户已开通 Data IDE 并且加入某项目空间。

step1：进入 Data IDE，点击新建 workflow，勾选一次性调度。



新建工作流

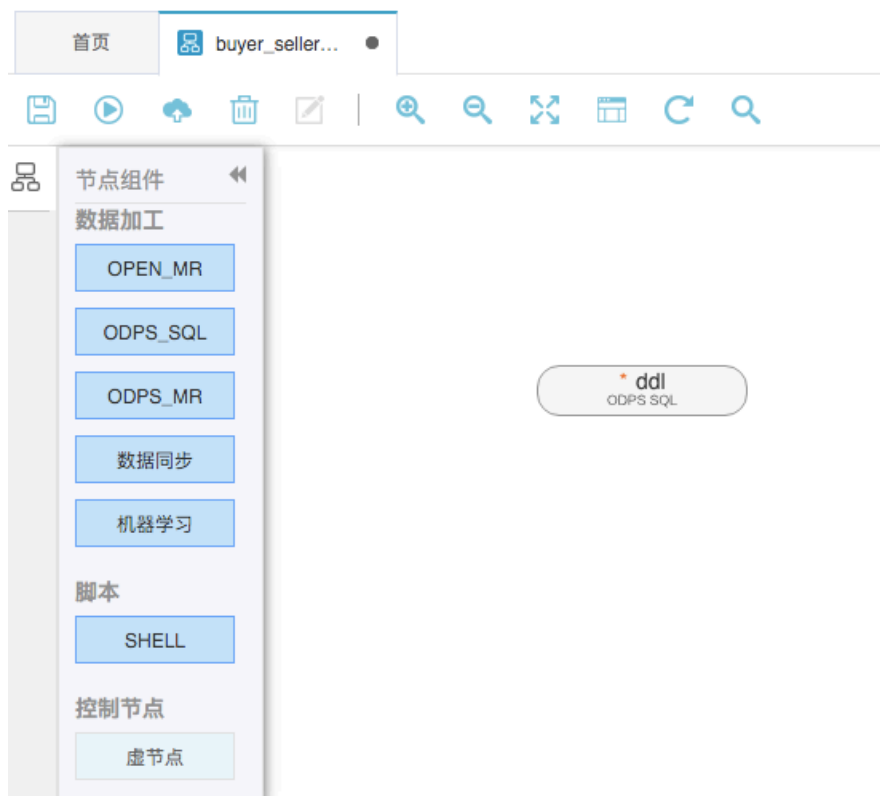
*工作流名称: buy_seller_interactions_feature_ddl

*描述: 买家与卖家交互行为信息原始特征集相关建表任务

*调度类型: ☒ 一次性调度 ☐ 周期调度

取消 创建

step2: 点击创建工作流, 并向工作流中拖入 ODPS SQL 节点。



step3 : 双击 ODPS SQL 进入 ODPS 代码编辑页面。输入多个建表语句。

-----DDL 批量建表语句-----

```
CREATE TABLE odps_buyer_seller_interactions (  
  
    thedate bigint COMMENT '日期',  
  
    buyer_id string COMMENT '买家 id',  
  
    seller_id string COMMENT '卖家 id',  
  
    shop_id string COMMENT '店铺 id',  
  
    cat_id bigint COMMENT '二级类目 id',  
  
    pv bigint COMMENT '浏览次数',  
  
    add_cart_num bigint COMMENT '添加购物车次数',  
  
    auction_collect_num bigint COMMENT '收藏次数',  
  
    alipay_trade_num bigint COMMENT '购买次数'  
  
)  
  
COMMENT '买家和卖家交互行为信息'  
  
PARTITIONED BY(  
  
    dt string  
  
)  
  
LIFECYCLE 7;  
  
  
CREATE TABLE odps_buyer_seller_interactions_feature (  
  
    masked_buyer_id string COMMENT '模糊用户 id',  
  
    masked_seller_id string COMMENT '模糊卖家 id',
```

```
masked_shop_id string COMMENT '模糊店铺 id',

cat_id bigint COMMENT '类目 id',

pv double COMMENT '流量',

add_cart_num double COMMENT '添加购物车次数',

auction_collect_num double COMMENT '商品收藏数',

alipay_trade_num double COMMENT '购买次数',

target bigint COMMENT '模型预测目标'

)

COMMENT '最佳实践特征模型表'

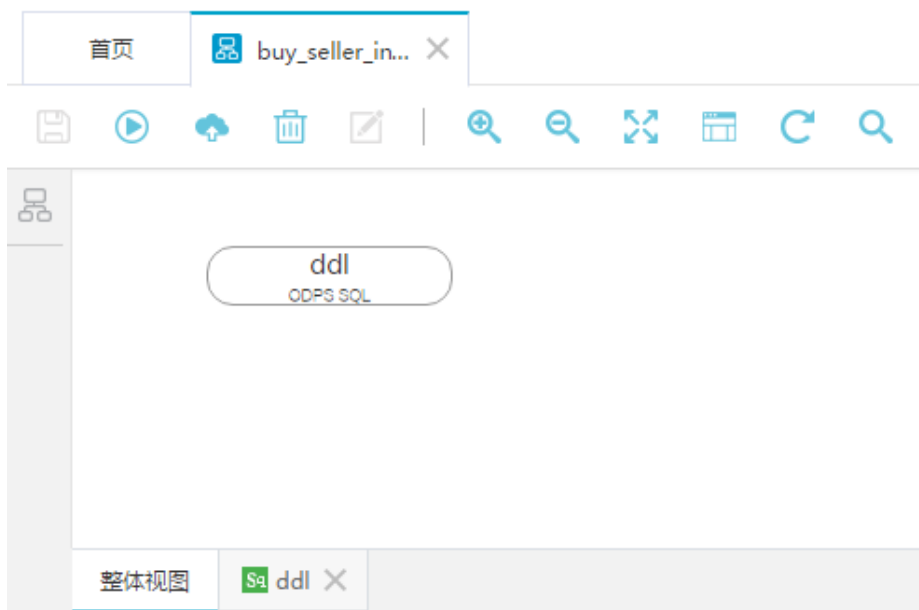
PARTITIONED BY (

    dt string

)

LIFECYCLE 10;
```

step3：点击保存并提交。



step4：点击**运行并创建**冒烟测试工作流，跳转至运维中心测试页面：

冒烟工作流

实例名称: buyer_seller_interactions_feature_ddl_2016_01_20

*业务日期: 2016-01-19

[取消](#) [创建](#)

定义 运维 测试 补数据

workflows buyer_seller_interactions_featu 查询

运行日期: 2016-01-20 业务日期: YYYY-MM-DD

☐ 我的工作流 ☐ 我的工作流(今天测的) 更多

✓ buyer_seller_interactions_featu
1-20 10:45:46~10:45:55(dur 9s)

✗ buyer_seller_interactions_featu
1-20 10:44:52~10:44:56(dur 4s)

workflows buyer_seller_inter... 工作流

工作流名称: buyer_seller_interactions_featu

当前状态: 运行成功

状态描述: 关联的工作流实例运行成功

应用名称: 日志_开发


作业类型: FLOW

开始时间: 2016-01-20 10:45:46

结束时间: 2016-01-20 10:45:55

责任人: yangyl

2.2 发布 workflow

step1：进入数据开发>发布管理，点击创建发布包。将新建的工作流添加至待发布列表。点击待发布列表 1。



step2：填写发布描述信息，并点击执行发布（项目管理员、运维和部署角色才能够直接进行发布，开发角色只能创建发布包）。

2.3 目标项目中执行 workflow

step1：点击切换至目标项目空间中。



step2：导航至运维中心>任务管理>定义，找到发布的工作流，在视图中右键选择测试 workflow。

定义

运维

测试

补数据

工作流

名称:

查询

☐ 我的工作流

☐ 我的工作流(今天修改的)

更多

工作流	修改日期
buy_seller_interactio_ddl	2016-01-14 13:19:02
testffzz	2016-01-12 17:42:54
无惟3	2015-12-30 20:10:14
f	2015-12-30 17:38:36
test_sync_copy_9	2015-12-30 16:27:31
无惟4	2015-12-30 16:20:21
test_sync_copy_8	2015-12-30 13:25:37

buy_seller_inter...

工作流

- 展开父工作流
- 展开子工作流
- 查看工作流操作日志
- 查看工作流属性
- 编辑工作流
- 测试工作流
- 补数据工作流