

# 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 过程中如何将本地数据上传至 ODPS，学员可以根据本实验手册，去学习使用不同工具 Data IDE 或 Dship 将本地数据导入至 ODPS。

## 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 Dship 工具

### 2.1.1 安装新版客户端

step1：正确安装 Java 运行环境（JDK1.6 及以上）；

step2：官网产品导航中找到 dship 工具并下载解压；

[https://help.aliyun.com/document\\_detail/odps/download/download.html](https://help.aliyun.com/document_detail/odps/download/download.html)

**step3** : 修改配置/conf/odps\_config.ini 文件 ;

```
project_name=指定用户想进入的项目名称标识
access_id=*****
access_key=*****
end_point=http://service.odps.aliyun.com/api
tunnel_endpoint=http://dt.odps.aliyun.com
log_view_host=http://logview.odps.aliyun.com
https_check=true
```

**step4** : /bin/odpscmd 和/bin/odpscmd.bat 分别为 Linux 和 window 下的可执行文件 ;

**step5** : 上传命名 tunnel upload [options] <path> <[project.]table[/partition]>

---

**【备注】** 可通过 tunnel help upload 命令查看帮助。

### 2.1.2 使用 odpscmd 进入交互页面

根据自身系统环境选择双击打开/bin/odpscmd 或/bin/odpscmd.bat 进入交互页面。输入 help 查看帮助信息 :

```
Aliyun ODPS Command Line Tool
Version 0.20.0
@Copyright 2015 Alibaba Cloud Computing Co., Ltd. All rights
reserved.
odps@ project_name>help;
```

### 2.1.3 创建 ODPS 表

编写创建 ODPS 目标表的语句 , 如下所示 :

```
odps@ project_name>drop table if exists sale_detail;
```

```
odps@ project_name>create table sale_detail(shop_name
string,customer_id string,total_price double)partitioned
by(sale_date string,region string);
```

查看表的信息：

```
odps@ project_name>desc sale_detail;
```

### 2.1.4 添加分区

编写添加分区 DDL 语句：

```
odps@ project_name>alter table sale_detail add partition
(sale_date='201312', region='hangzhou');
```

查看分区信息：

```
odps@ project_name>show partitions sale_detail;
```

### 2.1.5 准备数据文件

新建 txt 文件，并输入数据如下。文件名为 data.txt：

```
shop9,97,100
shop10,10,200
shop11,11,150
```

### 2.1.6 上传数据

编写 tunnel 命令：

```
odps@ project_name>tunnel u d:\data.txt
```

```
sale_detail/sale_date=201312,region=hangzhou -s false;
```

### 2.1.7 查看数据

编写 select 查询语句：

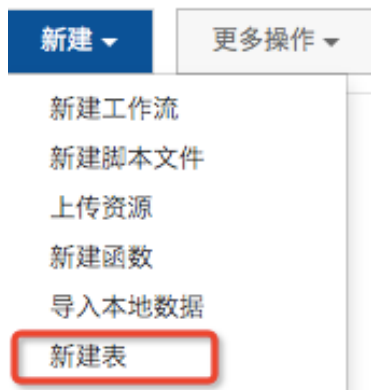
```
odps@ project_name>select * from sale_detail;
```

## 2.2 Data IDE

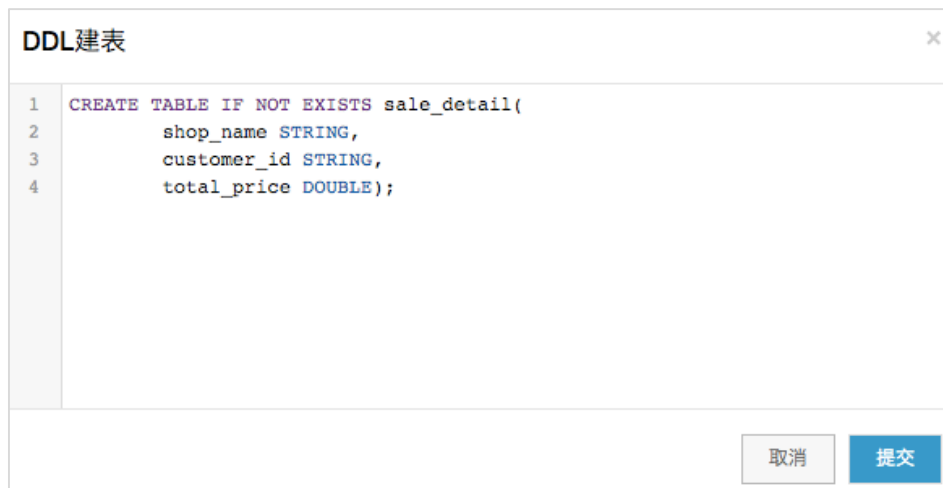
### 2.2.1 新建 ODPS 表

**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，点击新建表。



**step2：**跳转至新建表页面，点击 DDL 建表，编写 DDL 语句。



-----DDL 建表语句-----

```
CREATE TABLE IF NOT EXISTS sale_detail(  
    shop_name STRING,  
    customer_id STRING,  
    total_price DOUBLE);
```

step3：点击**提交**，选择数据表所属项目名、所属类目和生命周期。

基础信息

字段和分区信息

新建成功!

1 基本信息设置

DDL建表

\* 表名:

sale\_detail

别名:

请输入中文名

\* 项目名:

odps.dataplus\_private\_test\_4\_dev

所属类目:

测试一级类目

测试二级

描述:

请输入描述

2 存储生命周期设置

\* 生命周期:

永久

取消

下一步

step4：点击**下一步**，进入**字段和分区信息**页面并点击提交。

基础信息

字段和分区信息

新建成功!

3 字段信息设置

字段英文名	中文名	字段类型	描述	设置权限	操作
shop_name		STRING			上移 下移 编辑 删除
customer_id		STRING			上移 下移 编辑 删除
total_price		DOUBLE			上移 下移 编辑 删除

+ 新增字段

4 是否设置分区: ☒ 否 ☐ 是

取消

上一步

提交

step5：新建表成功。

数据表管理

刷新

新建表

我拥有的表

我收藏的表

我管理的表

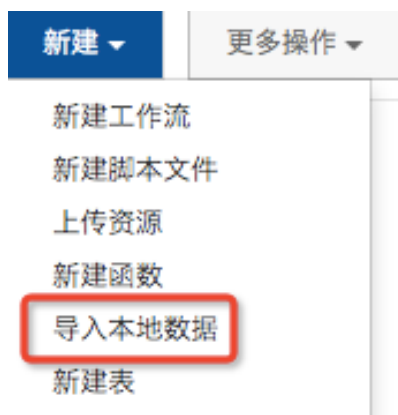
请输入表名/项目名

搜索

表名	ODPS项目	项目名	创建时间	物理存储	生命周期	人气	最近30天浏览人数	操作
sale_detail	odps.dataplus_p...	D+测试项目4(DEV)	2016-02-19 10:46:14	-	永久	0	0	更多

## 2.2.2 上传本地数据

step1：点击上传本地数据。



step2：选择本地文件 data.txt，并取消勾选首行为标题选项。

本地数据导入

已选文件: data.txt 只支持.txt和.csv文件类型

分隔符号: 逗号

自定义分隔符: ☐ 分隔符

原始字符集: GBK

导入起始行: 1

首行为标题: ☒ 是

col1	col2	col3
shop9	97	100
shop10	10	200
shop11	11	150

下一步

取消

step3：点击**下一步**，输入表名 sale\_detail 进行模糊查找。

本地数据导入

导入至表: sale\_detail

字段匹配: ☒ 按位置匹配 ☐ 按名称匹配

目标字段	源字段
shop_name	空字段
customer_id	空字段
total_price	空字段

上一步

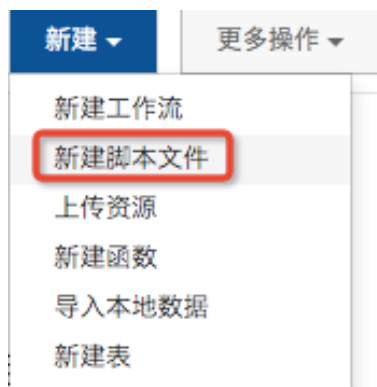
导入

取消

step4：点击**导入**。

### 2.2.3 查看数据

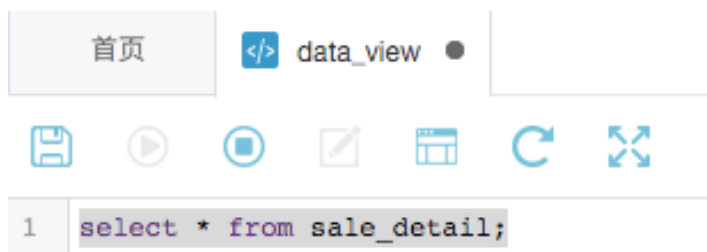
step1：点击**新建脚本文件**。



step2：填写文件名称、选择脚本文件类型为 ODPS SQL 并输入描述。

A screenshot of a '新建脚本文件' (New Script File) dialog box. It has three input fields: '\*文件名称:' (File Name) with the value 'data\_view', '\*类型:' (Type) with a dropdown menu showing 'ODPS SQL', and '\*描述:' (Description) with the value '数据查看'. At the bottom right, there are two buttons: '取消' (Cancel) and '提交' (Submit).

step3：编写数据查看 select 语句。



```
select * from sale_detail;
```