

# 1、ODS

## 1.1 刷卡主数据

1、在hive中创建表

```
create EXTERNAL TABLE IF NOT EXISTS ods.ods_szt_data(
    car_no String comment '车牌号',
    card_no String comment '卡号',
    close_date String comment '结算日期',
    company_name String comment '公司名称',
    conn_mark String comment '联程标记',
    deal_date String comment '交易日期时间',
    deal_money String comment '交易金额',
    deal_type String comment '交易类型',
    deal_value String comment '交易值',
    equ_no String comment '设备编码',
    station String comment '线路站点'
)
comment '深圳地铁刷卡信息表'
PARTITIONED BY (
    day string comment '天分区'
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS INPUTFORMAT 'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
location '/daas/szt/ods/ods_szt_data';
```

2、增加分区

```
alter table ods.ods_szt_data add if not exists partition(day='2018-09-01');
```

3、将数据上传到表的目录下

```
hadoop dfs -put data.txt /daas/szt/ods/ods_szt_data/day=2018-09-01
```

# 2、DWD

## 2.1 地铁出入站表

1、在hive中创建表

```
create EXTERNAL TABLE IF NOT EXISTS dwd.dwd_fact_szt_in_out_detail(
    car_no String comment '车牌号',
    card_no String comment '卡号',
    close_date String comment '结算日期',
    company_name String comment '公司名称',
    conn_mark String comment '联程标记',
    deal_date String comment '交易日期时间',
```

```

    deal_money String comment '交易金额',
    deal_type String comment '交易类型',
    deal_value String comment '交易值',
    equ_no String comment '设备编码',
    station String comment '线路站点'
)

comment '地铁出入站'
PARTITIONED BY (
    day string comment '天分区'
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS INPUTFORMAT 'org.apache.hadoop.mapred.TextInputFormat'
    OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
location '/daas/szt/dwd/dwd_fact_szt_in_out_detail';

```

## 2、将脚本上传到服务器

dwd\_fact\_szt\_in\_out\_detail.sh

## 3、在dolphinscheduler中创建一个shell 节点

节点名称: dwd\_fact\_szt\_in\_out\_detail

运行标志:  正常  禁止执行

描述: dwd\_fact\_szt\_in\_out\_detail.sh

任务优先级: MEDIUM

Worker分组: default

失败重试次数: 0 (次)

失败重试间隔: 1 (分)

超时告警:

脚本:

```

1 sh /root/szt/dwd_fact_szt_in_out_detail.sh 2018-09-01
2

```

资源: 请选择资源

自定义参数:

## 4、启动任务

# 2.2 地铁进站表

## 1、在hive中创建表

```

create EXTERNAL TABLE dwd.dwd_fact_szt_in_detail(
    car_no String comment '车牌号',
    card_no String comment '卡号',
    close_date String comment '结算日期',
    company_name String comment '公司名称',
    conn_mark String comment '联程标记',
    deal_date string comment '交易日期时间',
    deal_money String comment '交易金额',
    deal_type String comment '交易类型',
    deal_value String comment '交易值',
    equ_no String comment '设备编码',
    station String comment '线路站点'
)
comment '地铁入站'
PARTITIONED BY (
    day string comment '天分区'
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS INPUTFORMAT 'org.apache.hadoop.mapred.TextInputFormat'
    OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
location '/daas/szt/dwd/dwd_fact_szt_in_detail';

```

## 2、将项目打包上传到dolphinscheduler

The screenshot shows the 'File Management > szt' page in the DolphinScheduler web interface. The page has a header with navigation links: 首页 (Home), 项目管理 (Project Management), 资源中心 (Resource Center), 数据源中心 (Data Source Center), 监控中心 (Monitoring Center), and 安全中心 (Security Center). On the right, there are language and user selection dropdowns. The main content area displays a table of files. The table has columns: 编号 (ID), 名称 (Name), 是否文件夹 (Is Directory), 文件名称 (File Name), 描述 (Description), 大小 (Size), 更新时间 (Last Updated), and 操作 (Operations). A single row is shown: ID 1, Name 'szt-dwd-1.0.jar', Is Directory '否' (No), File Name 'szt-dwd-1.0.jar', Description 'dwd\_fact\_szt\_in\_detail', Size '5 KB', Last Updated '2021-08-09 16:54:25', and Operations with icons for download, edit, delete, and refresh. Below the table are pagination controls: '<' (Previous), '1' (Current page), '>' (Next), '10条/页' (10 items per page), and '跳转至' (Jump to) with a page number input field.

编号	名称	是否文件夹	文件名称	描述	大小	更新时间	操作
1	szt-dwd-1.0.jar	否	szt-dwd-1.0.jar	dwd_fact_szt_in_detail	5 KB	2021-08-09 16:54:25	

## 3、在工作流增加一个spark 节点

节点名称

运行标志  正常  禁止执行

描述

任务优先级  Worker分组

失败重试次数  (次) 失败重试间隔  (分)

超时告警

程序类型

Spark版本

主函数的Class

主程序包

部署方式  cluster  client  local

Driver核心数  Driver内存数

Executor核心数  Executor内存数

## 2.3 地铁出站表

1、在hive中创建表

```

create EXTERNAL TABLE IF NOT EXISTS dwd.dwd_fact_szt_out_detail(
    car_no String comment '车牌号',
    card_no String comment '卡号',
    close_date String comment '结算日期',
    company_name String comment '公司名称',
    conn_mark string comment '联程标记',
    deal_date String comment '交易时间',
    deal_money String comment '交易金额',
    deal_type String comment '交易类型',
    deal_value String comment '交易值',
    equ_no String comment '设备编码',
    station String comment '线路站点'
)
comment '地铁出站'
PARTITIONED BY (
    day string comment '天分区'
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS INPUTFORMAT 'org.apache.hadoop.mapred.TextInputFormat'

```

```
    OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'  
    location '/daas/szt/dwd/dwd_fact_szt_out_detail';
```

2、将脚本上传到服务器

dwd\_fact\_szt\_out\_detail.sh

3、在dolphinscheduler增加一个shell节点