

1、ODS

1.1 刷卡主数据

1、再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



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