

## 实验 4-4 Hive ql 实现价格、时间段销量分析

建议课时：60 分钟

### 一、实验目的

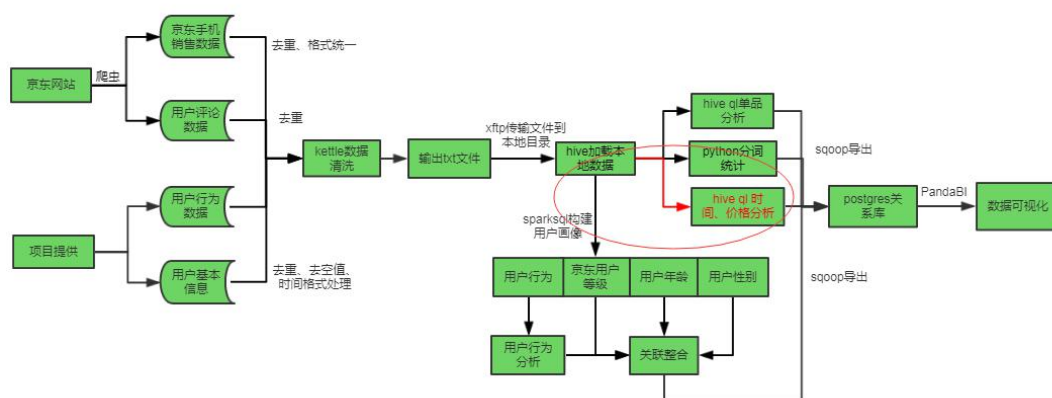
- 掌握 Hql 语句的编写；
- 了解时间段和价格两个指标的分析方法；

### 二、实验环境

Dsight 智慧实验室的 hadoop 环境

### 三、实验步骤

本节实验所做内容如下红色标注：



本节实验主要是通过 Hive ql 语句实现各时间段、各价格区间手机销量的统计。

具体实验步骤如下：

#### 1. 划分时间段统计手机销量

##### (1) 新建时间段手机销量表

表结构设计：

```
drop table if exists date_range_sail_count;

create table date_range_sail_count
(
    date_range string comment '交易时间段',
    sail_count int comment '手机销量'
)

comment '各时间段手机销量表';
```

## (2) 编辑 Hql 语句关联查询数据并插入新表

关联手机销售信息表和用户行为表

各时间段销量表字段、数据：

```
Time taken: 0.139 seconds, Fetched: 185 row(s)
hive> desc period_sail_count_tb;
OK
_c0                bigint
deal_time          string
Time taken: 0.04 seconds, Fetched: 2 row(s)
hive> select * from period_sail_count_tb limit 5;
OK
201176  10-1
60613   10-10
20896   10-11
36366   10-12
28850   10-13
Time taken: 0.107 seconds, Fetched: 5 row(s)
hive> 
```

## 2. 划分价格区间统计手机销量

### (1) 新建价格区间中间表，存储手机销量、价格区间

<1000 1000 元以下

[1000,2000] 1000-2000 元

[2001,3000] 2000-3000 元

[3001-5000] 3000-5000 元

[5001-8000] 5000-8000 元

[8001-10000] 8000-10000 元

>10000 10000 元以上

### (2) 在中间表基础上计算销量并存储到价格区间手机销量结果表

表结构设计：

```
drop table if exists price_range_sail_count;

create table price_range_sail_count
(
```

```
price_range string comment '价格区间',  
sail_count int comment '手机销量'  
)  
comment '各价格区间手机销量表';
```

(3) 编辑 Hql 语句计算并查询数据插入结果表

```
hive> desc price_range_sail_count_result;  
OK  
price_range      string  
_c1              bigint  
Time taken: 0.041 seconds, Fetched: 2 row(s)  
hive> select * from price_range_sail_count_result limit 5;  
OK  
1000-2000元      19943943  
10000元及以上    375013  
1000元以下       19167842  
2000-3000元      8674382  
3000-5000元      10500136  
Time taken: 0.143 seconds, Fetched: 5 row(s)  
hive> █
```

## 四、实验成果

本次实验完成后，需要得到以下结果：

- 统计不同时间段手机销量；
- 统计不同价格区间手机销量；