实验 4-5　Sqoop 实现分析数据的导出

建议课时：60 分钟

一、实验目的

- 掌握 sqoop 工具导出数据的使用方法；
- 熟练编写 sqoop export 命令；
- 熟练使用 Danastudio 平台建 postgres 表；

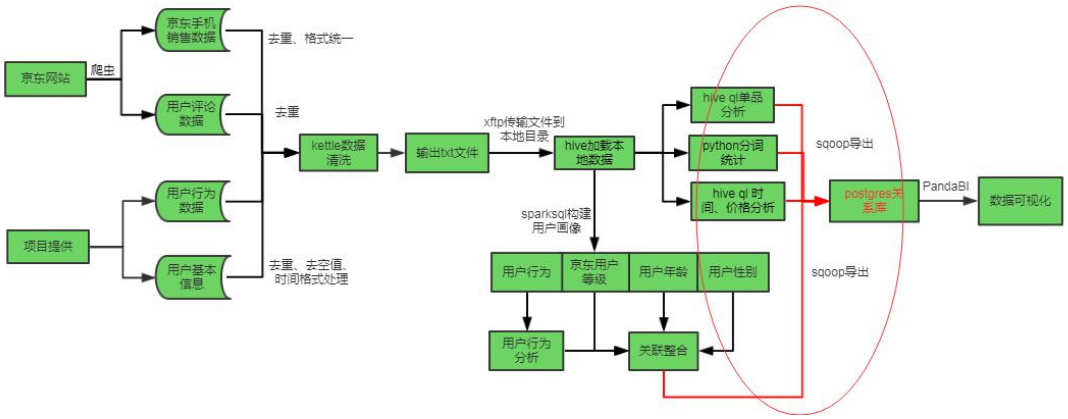
二、实验环境

Dsight 智慧实验室的 hadoop 环境

Danastudio 平台

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 Sqoop 工具将分析结果数据导出到 postgres 关系库中，方便后续大屏可视化。

具体实验步骤如下：

1. 关系库 postgres 中目标表的创建

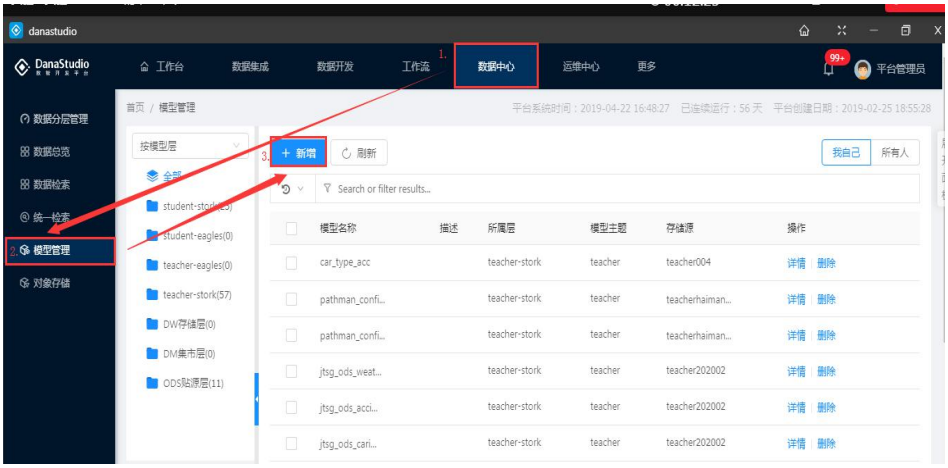
新建手机品牌热销 Top10 表、华为手机单品销量 Top20 表、苹果手机单品销量 Top20 表、用户评论热词统计表、各地区手机销量表、各时间段手机销量表、各价格区间手机销量表、用户标签宽表

1.1 启动 Danastudio

在 Dsight 实验室打开 danastudio 环境，进入 danastudio

1.2 新增数据源和目标表

点击数据中心 ---> 模型管理 ---> 新增



选择存储源（student+自己学号或 teacher+工号,存储源名称即目标数据库名） ---> 数据层选择 student-stork 或 teacher-stork ---> 类型、模式选择默认 ---> 模型名（即目标表名）---> 按导出字段添加字段 ---> 点击下边的完成按钮即可。

注：导出的字段和目标表的字段要一致

2. Sqoop export 导出命令的编写

➢ 编写导出各年龄段手机销售数据的 sqoop 命令：

bin/sqoop export \

--connect jdbc:postgresql://192.168.50.78:14103/teacher123 \

--username stork \

--password stork \

--table age_region_sail_info \

--export-dir /data/hive/warehouse/sail.db/age_region_sail_info \

--input-fields-terminated-by ','

注：

192.168.50.78：Postgres 主机地址（Danastudio IP 地址）

14103：Postgres 固定 IP

teacher123：postgres 数据库名称（danastudio 中存储源的名称，student+学号或 teacher+工号）

--export-dir /data/hive/warehouse/sail.db/age_range_sail_count 数据存储路径，可以通过以下命令查询：

show create table age_range_sail_count;

- 按照以上方式编写 sqoop 脚本导出其他几张表的数据

3. 执行 Sqoop 脚本导出数据

（1）进入 sqoop 目录下：cd /opt/sqoop

（2）执行导出命令

bin/sqoop export \

--connect jdbc:postgresql://192.168.50.78:14103/teacher123 \

--username stork \

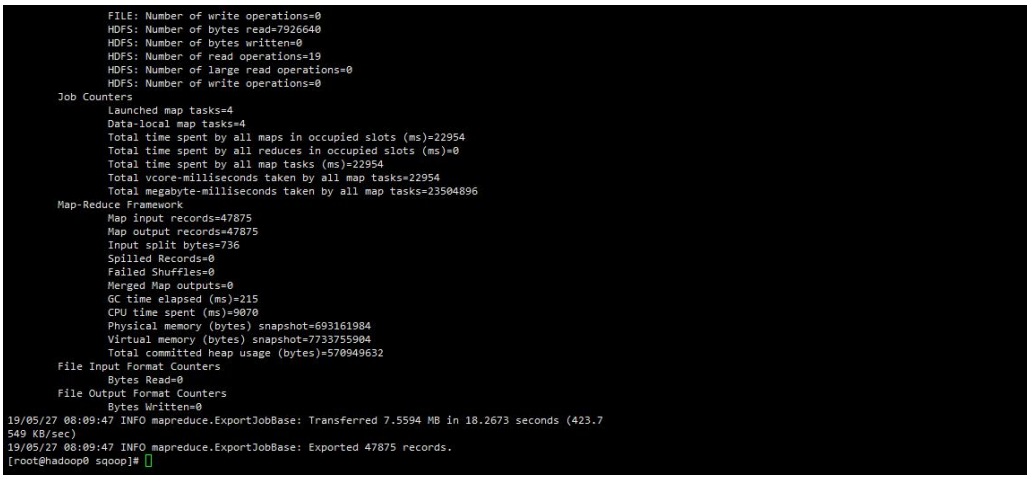
--password stork \

--table age_region_sail_info\

--export-dir /data/hive/warehouse/sail.db/age_region_sail_info\

--input-fields-terminated-by ','

出现以下信息表示导出成功：



四、实验成果

本次实验完成后，需要得到以下结果：

- 导出用户各年龄段销量表数据到 postgres；
- 导出各地区手机单品销量数据到 postgres；
- 导出手机销量 Top10 表中数据到 postgres；
- 导出华为手机单品销量 Top20 数据到 postgres；
- 导出苹果手机单品销量 Top20 数据到 postgres；
- 导出各时间段手机销量数据到 postgres；
- 导出各价格区间手机销量数据到 postgres；
- 导出用户评论词频统计 Top200 数据到 postgres；
- 导出用户标签宽表数据到 postgres；