# Hadoop综合练习

**计科2班(国际班)3119009450邹佳豪**

## 1.数据说明：

搜狗五百万数据，是经过处理后的搜狗搜索引擎生产数据，具有真实性，大数据性，能够较好的满足分布式计算应用开发课程设计的数据要求。

## 2. 数据的格式：

访问时间 ts string

用户ID uid string

查询词 keyword string

该URL在返回结果中的排名 rank int

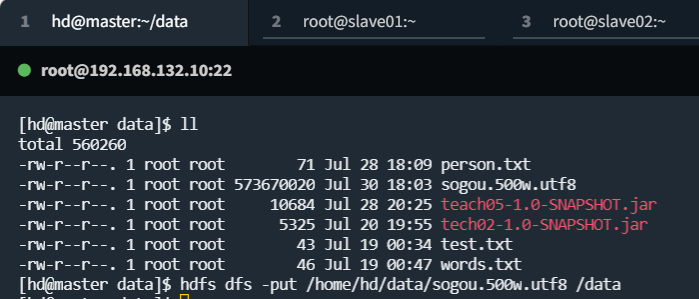
用户点击的顺序号 sorder int

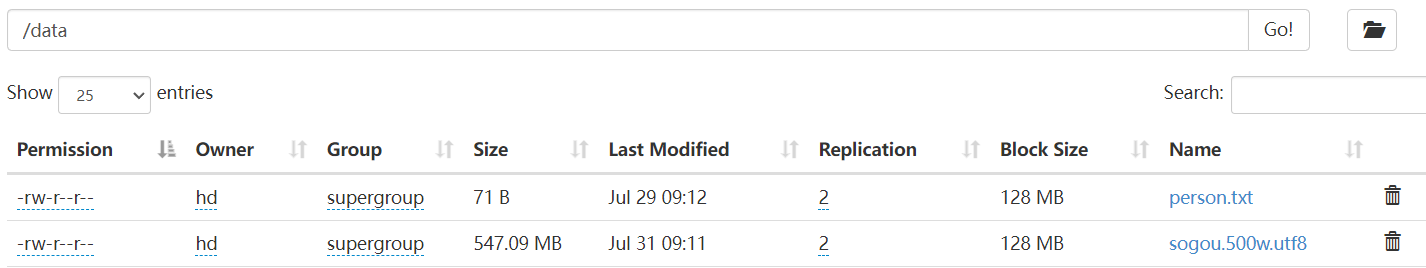
用户点击的URL url string

## 操作要求：

1. 将原始数据加载到HDFS。

截图





1. 将数据导入到hive仓库

|  |
| --- |
| 创建数据库hive> create database sogou;  使用数据库Hive> use sogou;  创建表：hive>  CREATE TABLE sogou\_data  (ts string,uid string,keyword string,rank int,sorder int,url string)  ROW FORMAT DELIMITED  FIELDS TERMINATED BY '\t';  Hive表加载数据 |

1. 查看hive sogou表前100条数据信息

|  |
| --- |
|  |

1. 通过idea编写代码,用sparkSQL 查询[查询频度排名（频度最高的前50词）

],保存到MariaDB

再通过flask显示视图到页面

|  |
| --- |
| SQL:select keyword,count(\*) query\_count from sogou group by keyword order by query\_count desc limit 50;  改为  SQL:select keyword,count(\*) as count from t\_sogou group by keyword order by count desc limit 50;  代码截图    效果截图 |

1. 通过idea编写代码,用sparkSQL 查询[查询次数大于2次的用户总数],保存到MariaDB

再通过flask显示视图到页面

|  |
| --- |
| SQL:select count(\*) from (select count(\*) as query\_count from sogou group by uid having query\_count > 2) a;  代码截图  select count(\*) from (select count(\*) as query\_count from t\_sogou group by uid having query\_count > 2);  效果截图    一列数据不好呈表 |

1. 通过idea编写代码,用sparkSQL 查询[查询搜索过”仙剑奇侠传“的uid，并且次数大于3],保存到MariaDB

再通过flask显示视图到页面

|  |
| --- |
| SQL:select uid,count(\*) as cnt from sogo where keyword='仙剑奇侠传'group by uid having cnt > 3;  改为:select uid,count(\*) as cnt from t\_sogou where keyword='仙剑奇侠传'group by uid having cnt > 3;  代码截图      效果截图 |

1. 通过idea编写代码,用sparkSQL 查询[关键词包含有‘小米’，‘百度’，‘仙剑奇侠传’的数据查询出来，并且统计查询次数],保存到MariaDB

再通过flask显示视图到页面

|  |
| --- |
| SQL:SELECT keyword , count(keyword) as countkeyword from sougou where (keyword LIKE '%百度%' \nor keyword LIKE '%小米%' \nor keyword LIKE '%仙剑奇侠传%' \n) GROUP by keyword  代码截图  SELECT keyword , count(keyword) as countkeyword from t\_sogou where (keyword LIKE '%百度%' or keyword LIKE '%小米%' or keyword LIKE '%仙剑奇侠传%' ) GROUP by keyword;  效果截图 |