

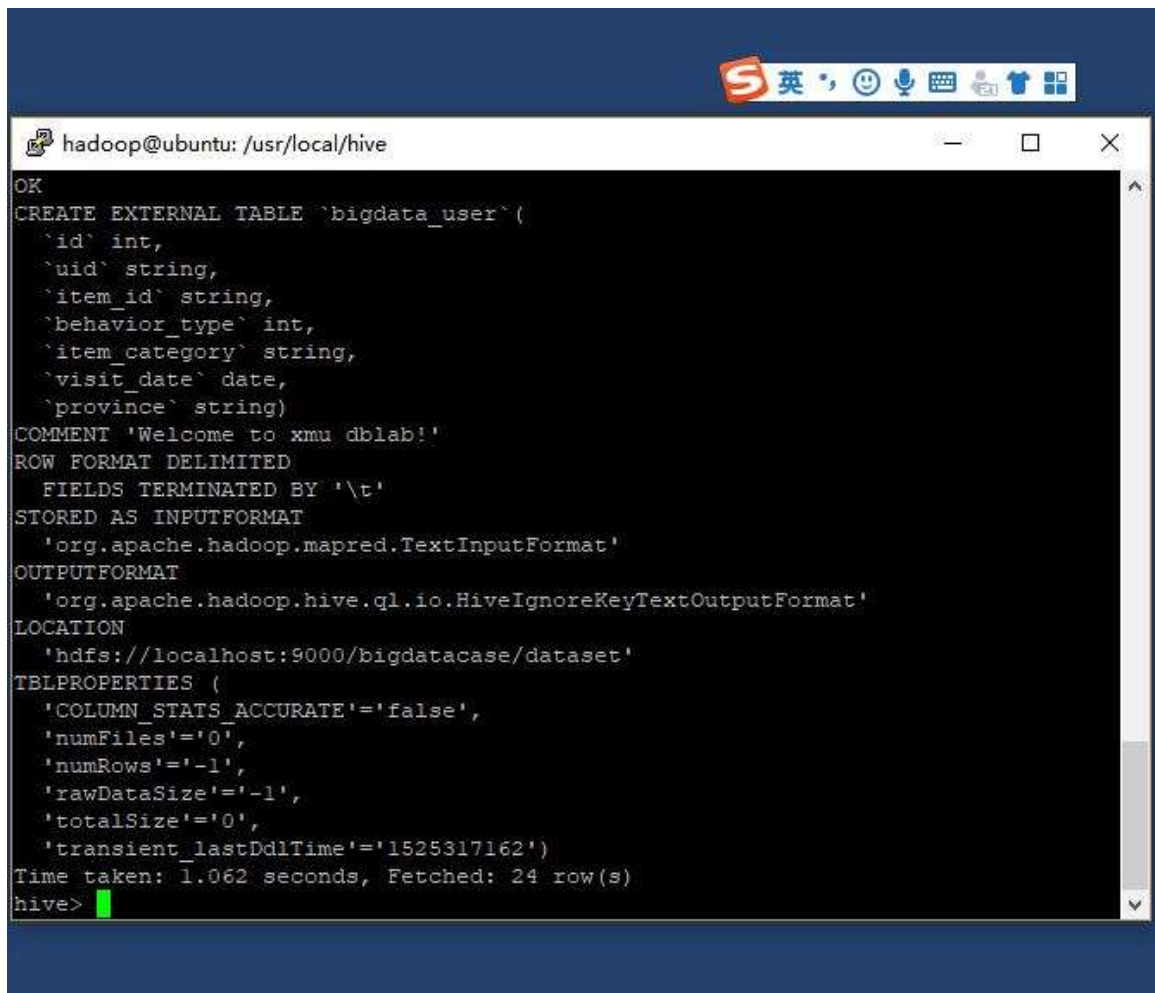
大数据案例-步骤二：Hive数据分析

一、任务清单

- 启动Hadoop和Hive
- 创建数据库和表
- 简单查询分析
- 查询条数统计分析
- 关键字条件查询分析
- 根据用户行为分析
- 用户实时查询分析

二、实验结果

- 启动Hadoop和Hive，创建数据库和表
 - 查看bigdata_user表的各种属性



```
hadoop@ubuntu: /usr/local/hive
OK
CREATE EXTERNAL TABLE `bigdata_user` (
  `id` int,
  `uid` string,
  `item_id` string,
  `behavior_type` int,
  `item_category` string,
  `visit_date` date,
  `province` string)
COMMENT 'Welcome to xmu dlab!'
ROW FORMAT DELIMITED
  FIELDS TERMINATED BY '\t'
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
  'hdfs://localhost:9000/bigdatacase/dataset'
TBLPROPERTIES (
  'COLUMN_STATS_ACCURATE'='false',
  'numFiles'='0',
  'numRows'='-1',
  'rawDataSize'='-1',
  'totalSize'='0',
  'transient_lastDdlTime'='1525317162')
Time taken: 1.062 seconds, Fetched: 24 row(s)
hive>
```

- 查看表的简单结构: `desc bigdata_user;`
- 简单查询分析
 - 查看前10位用户对商品的行为

- 查询前20位用户购买商品时的时间和商品的种类
- 利用列的别名查看用户对商品的行为和商品分类
- 查询条数统计分析
 - 用聚合函数count()计算出表内有多少条行数据
 - `select count(*) from bigdata_user;`
 - 查出uid不重复的数据有多少条
 - `select count(distinct uid) from bigdata_user;`
 - 查询不重复的数据有多少条(为了排除客户刷单情况)
 - `select count() from (select uid,itemid,behavior_type,itemcategory,visitdate,province from bigdatauser group by uid,itemid,behavior_type,itemcategory,visitdate,province having count()=1)a;`
- 关键字条件查询分析
 - 以关键字的存在区间为条件的查询
 - 查询2014年12月10日到2014年12月13日有多少人浏览了商品


```
select count(*) from bigdata_user where behavior_type='1' and visit_date<'2014-12-13' and visit_date>'2014-12-10';
```
 - 以月的第n天为统计单位，依次显示第n天网站卖出去的商品的个数


```
select count(distinct uid), day(visit_date) from bigdata_user where behavior_type='4' group by day(visit_date);
```
 - 关键字赋予给定值为条件，对其他数据进行分析
 - 取给定时间和给定地点，求当天发出到该地点的货物的数量


```
select count(*) from bigdatauser where province='江西' and visitdate='2014-12-12' and behavior_type='4';
```
- 根据用户行为分析
 - 查询一件商品在某天的购买比例或浏览比例
 - 查询有多少用户在2014-12-11购买了商品
 - 查询有多少用户在2014-12-11点击了该店
 - 查询某个用户在某一天点击网站占该天所有点击行为的比例

- 查询用户10001082在2014-12-12点击网站的次数
- 查询所有用户在这一天点击该网站的次数
- 查询某一天在该网站购买商品超过5次的用户id
- 用户实时查询分析
 - 某个地区的用户当天浏览网站的次数