

大数据技术-第十章：购物网站用户行为分析综合案例

Hive数据分析



CONTENTS

01. 简单查询分析

02. 查询条数统计分析

03. 关键字条件查询分析

04. 根据用户行为分析



01

简单查询分析



简单查询分析

首先执行一条简单的指令：

```
hive> select behavior_type from bigdata_user limit 10; #查看  
前10位用户对商品的行为
```

如果要查出每位用户购买商品时的多种信息，输出语句格式如下：

select 列1, 列2, ..., 列n from 表名;

比如查询前20位用户购买商品时的时间和商品的种类，语句如下：

```
hive> select visit_date, item_category from bigdata_user limit 20;
```

有时在表中查询可以利用嵌套语句，如果列名太复杂可以设置该列的别名，以简化操作的难度，举例如下：

```
hive> select e.bh, e.it from (select behavior_type as bh,  
item_category as it from bigdata_user) as e limit 20;
```

02

查询条数统计分析



» 查询条数统计分析

1. 用聚合函数count()计算出表内有多少行数据

```
hive> select count(*) from bigdata_user;
```

2. 在函数内部加上distinct, 查出uid不重复的数据有多少条

```
hive> select count(distinct uid) from bigdata_user;
```

3. 查询不重复的数据有多少条(为了排除客户刷单情况)

```
hive> select count(*) from (select  
uid,item_id,behavior_type,item_category,visit_date,province from  
bigdata_user group by  
uid,item_id,behavior_type,item_category,visit_date,province  
having count(*)=1)a;
```

03

关键字条件查询分析



1.以关键字的存在区间为条件的查询

(1)查询2014年12月10日到2014年12月13日有多少人浏览了商品。

```
hive> select count(*) from bigdata_user where  
behavior_type='1' and visit_date<'2014-12-13' and  
visit_date>'2014-12-10';
```

(2)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。

```
hive> select count(distinct uid), day(visit_date) from  
bigdata_user where behavior_type='4' group by  
day(visit_date);
```

2.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。

```
hive> select count(*) from bigdata_user where province='江西'  
' and visit_date='2014-12-12' and behavior_type='4';
```


04

根据用户行为分析



根据用户行为分析

1. 查询一件商品在某天的购买比例或浏览比例

```
hive> select count(*) from bigdata_user where visit_date='2014-12-11' and behavior_type='4';#查询有多少用户在2014-12-11购买了商品
```

```
hive> select count(*) from bigdata_user where visit_date='2014-12-11';#查询有多少用户在2014-12-11点击了该店
```

2. 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（点击行为包括浏览、加入购物车、收藏、购买）

```
hive> select count(*) from bigdata_user where uid=10001082 and visit_date='2014-12-12';#查询用户10001082在2014-12-12点击网站的次数
```

```
hive> select count(*) from bigdata_user where visit_date='2014-12-12';#查询所有用户在这一天点击该网站的次数
```

➤ 根据用户行为分析

3. 给定购买商品的数量范围，查询某一天在该网站的购买该数量商品的用户id

```
hive> select uid from bigdata_user where behavior_type='4' and  
visit_date='2014-12-12' group by uid having  
count(behavior_type='4')>5;#查询某一天在该网站购买商品超过5次的用
```

户id

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG
DATA

智 / 能 / 科 / 技 放 / 眼 / 未 / 来

