# Hive数据分析

**【实验名称】**

Hive数据分析

**【实验目的】**

掌握数据仓库Hive概念及其基本原理、SQL语句、数据库查询分析

**【实验要求】**

 数据仓库Hive基本操作、创建数据库和表、使用SQL语句进行查询分析

**【实验设备】**

基于综合实验1在hadoop1主节点上进行操作

**【实验步骤】**

## 1.创建Hive数据库

### 1.1启动MySQL数据库

[root@hadoop1 ~]# service mysqld restart



### 1.2进入Hive

[root@hadoop1 ~]#hive



### 1.3在hive中创建dblab数据库

hive> create database dblab;

hive> use dblab;





### 1.4创建外部表

hive> CREATE EXTERNAL TABLE dblab.bigdata\_user(id INT,uid STRING,item\_id STRING,behavior\_type INT,item\_category STRING,visit\_date DATE,province STRING) COMMENT 'Welcome to xmudblab!' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE LOCATION '/bigdatacase/dataset';



## 2.Hive查询数据

### 2.1查看dblab数据库中的表

hive> use dblab;

hive> show tables;



### 2.2查看bigdata\_user表的各种属性

hive> show create table bigdata\_user;



### 2.3查看表的简单结构

hive> desc bigdata\_user;



### 2.4查询相关数据

hive> select \* from bigdata\_user limit 10;



hive> select behavior\_type from bigdata\_user limit 10;



## 3.Hive数据分析

### 3.1简单查询分析

查看前10位用户对商品的行为

hive> use dblab;

hive> select behavior\_type from bigdata\_user limit 10;



### 3.2查询前20位用户购买商品的时间和商品的种类

hive> select visit\_date, item\_category from bigdata\_user limit 20;



### 3.3查询条数统计分析

用聚合函数 count()计算出表内有多少行数据

hive> select count(\*) from bigdata\_user;



### 3.4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;



### 3.5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

hive> select count(\*) from bigdata\_user where behavior\_type='1' and visit\_date<'2014-12-13' and visit\_date>'2014-12-10';



### 3.6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个数

hive> select count(distinct uid), day(visit\_date) from bigdata\_user where behavior\_type='4' group by day(visit\_date);



### 3.7关键字赋予定值为条件,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';



### 3.8查询一件商品在某天的购买比例或者浏览比例

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'and behavior\_type='4';



### 3.9查询有多少用户在2014-12-11点击了该店

hive> select count(\*) from bigdata\_user where visit\_date ='2014-12-11';



### 3.10查询用户10001082在2014-12-12点击网站的次数

hive> select count(\*) from bigdata\_user where uid=10001082 and visit\_date='2014-12-12';



### 3.11查询所有用户在这一天点击该网站的次数

hive> select count(\*) from bigdata\_user where visit\_date='2014-12-12';



### 3.12查询某一天在该网站购买商品超过5次的用户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;

