实验 3-1 Hive 加载电商源数据

建议课时：60 分钟

一、 实验目的

 掌握hive 建库、建表语句；

 掌握hive 加载数据的几种方式；

 熟练hive 表数据的常用查询语句；

二、 实验环境

Dsight 智慧实验室中的 hadoop 环境

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Hive 加载本地源数据到 hdfs 分布式文件系统。

创建用户数据库

create database sails；

使用数据库

use sails；

创建销售表

create table goods\_sail\_tb(

goods\_id string,

goods\_name string,

goods\_property string,

store\_name string,

stroe\_id string,

goods\_url string,

goods\_price float,

keyword string,

sail\_count int,

good\_rate int,

brand string,

model string,

color string,

time\_to\_market string,

operate\_system string)

row format delimited fields terminated by ','

l

打入本地iPhone.txt数据

load data local inpath '/home/hadoop/data/data1/iphone.txt' into table goods\_sail\_tb;

创建user\_action表

create table user\_action(

user\_id string,

goods\_id string,

user\_action int,

month string,

day string)

row format delimited fields terminated by ','

lines terminated by '\n';

导入本地action.txt数据

load data local inpath '/home/hadoop/data/data1/action.txt' into table user\_action;

创建user\_info\_tb表

create table user\_info\_tb(

user\_id string,

user\_name string,

addr string ,

gender string,

birthday string)

row format delimited fields terminated by ','

lines terminated by '\n';

导入数据

load data local inpath '/home/hadoop/data/data1/userinfo.txt' into table user\_info\_tb;

创建user\_info\_new\_tb表

create table user\_info\_new\_tb(

user\_id string,

user\_name string,

addr string ,

gender string,

age\_region int,

age\_region\_alias string,

user\_grade string)

row format delimited fields terminated by ','

lines terminated by '\n';

创建评论表

create table user\_comment\_tb(

goods\_id string,

comment string,

releasetime string ,

user\_id string,

user\_name string,

user\_level string,

goods\_color string,

answercount string,

e\_score string,

resource string)

row format delimited fields terminated by ','

lines terminated by '\n';

导入数据

load data local inpath '/home/hadoop/data/data1/comment.txt' into table user\_comment\_tb;

创建中间表

create table userinfo\_mid as

select i.user\_id,i.user\_name,i.addr,i.gender,

floor(datediff(concat('',from\_unixtime(unix\_timestamp()),''),

concat('',birthday,'')) / 365) age,

c.user\_level

from user\_info\_tb i

join user\_comment\_tb c on i.user\_id=c.user\_id;

分段

insert into user\_info\_new\_tb select user\_id,user\_name,addr,gender,

case when age>=0 and age<=17 then 1

when age >= 18 and age <=24 then 2

when age >= 25 and age <= 29 then 3

when age>= 30 and age <= 34 then 4

when age>= 35 and age <= 39 then 5

when age>=40 and age<=49 then 6

when age>=50 then 7 end as age\_region,

case when age>=0 and age<=17 then "18岁以下"

when age>=18 and age<=24 then "[18,24]"

when age>=25 and age<=29 then "[25,29]"

when age>=30 and age<=34 then "[30,34]"

when age>=35 and age<=39 then "[35,39]"

when age>=40 and age<=49 then "[40,49]"

when age>=50 then "50岁以上" end as age\_region\_alias

,user\_level from userinfo\_mid ;











