实验 3-2 基于 Spark 构建用户画像

建议课时： 60 分钟

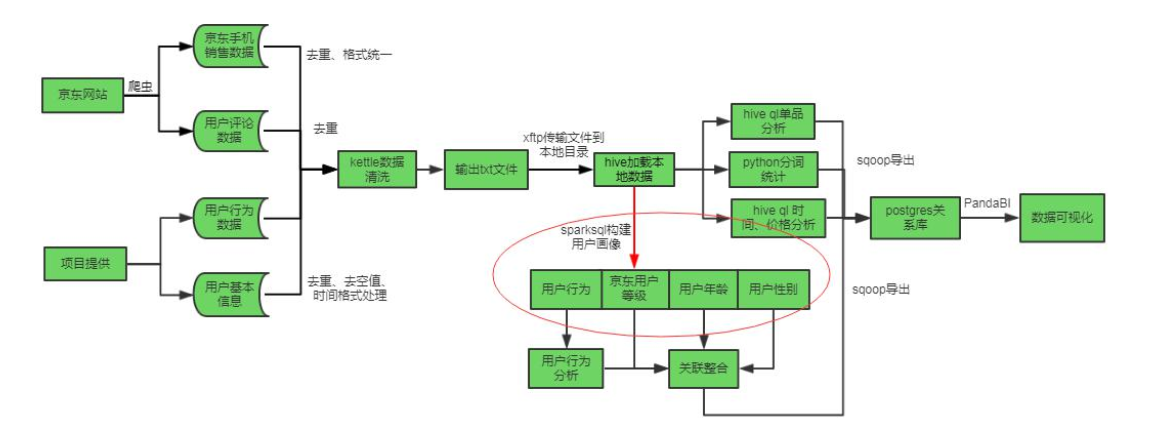
1. 实验目的

了解什么是用户画像；

熟悉用户画像的开发流程；  
了解电商用户画像如何构建；  
熟悉使用 Hive、 Spark SQL 进行数据开发；

1. 实验环境  
   Dsight 智慧实验室中的 hadoop 实验环境
2. 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 sparksql 整合 hive 实现用户画像的标签开发。

具体实验步骤如下：

（1） 开发用户性别标签表

create table person\_user\_tag\_gender(

user\_id string comment'用户编码',

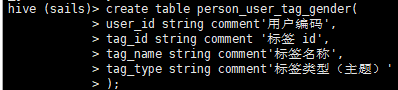
tag\_id string comment '标签 id',

tag\_name string comment'标签名称',

tag\_type string comment'标签类型（主题）'

);

comment '用户画像-用户行为标签表';



加载数据

insert into person\_user\_tag\_gender

select user\_id,

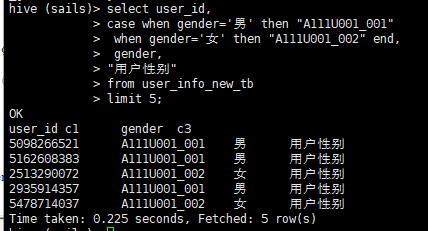
case when gender='男' then "A111U001\_001"

when gender='女' then "A111U001\_002" end,

gender,

"用户性别"

from user\_info\_new\_tb



（2）开发用户年龄标签表

建表语句：

create table person\_user\_tag\_user\_age\_region(

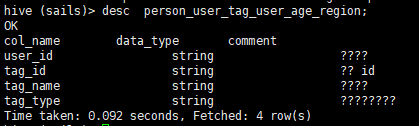
user\_id string comment'用户编码',

tag\_id string comment '标签 id',

tag\_name string comment'标签名称',

tag\_type string comment'标签类型（主题）'

);



加载数据

insert into person\_user\_tag\_user\_age\_region

select user\_id,

case when age\_region=1 then "A111U002\_001"

when age\_region=2 then "A111U002\_002"

when age\_region=3 then "A111U002\_003"

when age\_region=4 then "A111U002\_004"

when age\_region=5 then "A111U002\_005"

when age\_region=6 then "A111U002\_006"

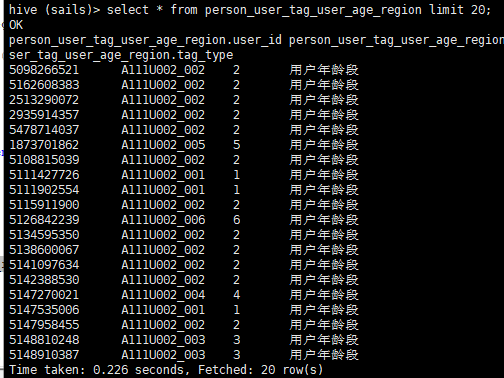
when age\_region=7 then "A111U002\_007" end,

age\_region,

"用户年龄段"

from user\_info\_new\_tb

查询：



（3）开发用户等级标签表

建表

create table person\_user\_tag\_grade(

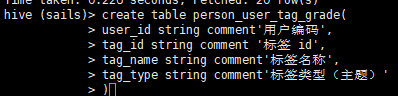
user\_id string comment'用户编码',

tag\_id string comment '标签 id',

tag\_name string comment'标签名称',

tag\_type string comment'标签类型（主题）'

)



插入数据

insert into person\_user\_tag\_grade

select user\_id,

case when user\_grade='注册会员' then "A111U003\_001"

when user\_grade='铜牌会员' then "A111U003\_002"

when user\_grade='银牌会员' then "A111U003\_003"

when user\_grade='金牌会员' then "A111U003\_004"

when user\_grade='钻石会员' then "A111U003\_005"

when user\_grade='企业会员' then "A111U003\_006"

when user\_grade='PLUS会员[试用]' then "A111U003\_007"

when user\_grade='PLUS会员' then "A111U004\_008"

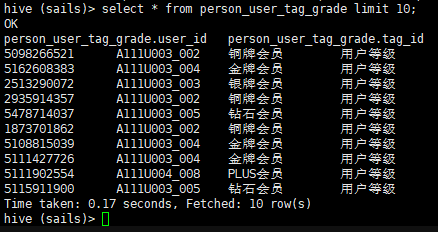
else "A111U003\_000"

end,

user\_grade,

"用户等级"

from user\_info\_new\_tb



（4）开发用户行为标签表

用户行为标签： 是根据用户在产品上的访问行为、 下单行为提取用户标签，用于定位用 户在产品上的访问情况， 进而根据用户的浏览习惯、 消费偏好做推荐和营销。

主要数据来源： 用户行为表

注： 在项目工程实践中， 数据主要来源于业务类数据表、 日志数据表和埋点数据表， 本次案例真实数据无法获取， 简单模拟了用户的基本行为数据包括（点击、 加入购物 车、 购买、 关注商品）  
标签开发的技术工具： 和用户属性标签开发相同

建表语句：

create table person\_user\_tag\_action(

user\_id string comment'用户编码',

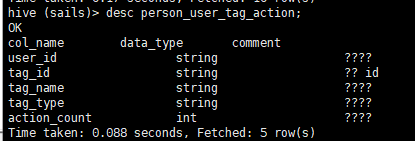
tag\_id string comment '标签 id',

tag\_name string comment'标签名称',

tag\_type string comment'标签类型',

action\_count int comment'行为次数'

);



导入数据

insert into person\_user\_tag\_action

select user\_id,

case

when user\_action=0 then "B211U001\_001"

when user\_action=1 then "B211U001\_002"

when user\_action=2 then "B211U001\_003"

when user\_action=3 then "B211U001\_004"

end ,

case

when user\_action=0 then "点击"

when user\_action=1 then "加入购物车"

when user\_action=2 then "关注"

when user\_action=3 then "购买"

end,

"用户行为",

cast(count(\*)/2 as int)

from user\_action\_tb

group by user\_id,user\_action

limit 10

