# 外卖行千锋商业运营平台

数仓

spark hive整合

spark hbase 整合

hbase的设计

任务调度

展示

SQL思路分析

讲一两个指标

项目的技术选型、项目描述、表设计

powerdesigner的使用

# 用户画像简介

什么是用户画像

用户画像的作用

如何构建用户画像

# 外卖行业用户画像的需求分析

1. 生成用户基本信息,业务数据提供用户基本信息和用户的扩展表

生成的用户基本信息包括：

Name,用户id,mid\_用户,登录名 ,用户性别,出生日期,年龄 ,星座 ,省份 ,城市 ,城市等级,邮箱 ,邮箱运营商 ,手机号 ,手机号段（前三位） ,手机运营商 ,注册时间 ,登录ip ,登录来源 ,邀请人 ,会员积分 ,已使用积分 ,会员等级名称,是否黑名单 ,是否结婚 ,学历 ,月收入 ,职业 ,性别模型值 ,是否孕妇 ,是否有小孩,是否有车 ,使用手机品牌 ,使用手机等级 ,使用手机种类数量,更换手机频率 ,是否马甲用户 ,马甲账户数量 ,用户忠诚度,用户购物类型 ,身材,身高。

1. 生成用户的访问模型

根据用户的app端点击日志和pc端的点击日志生成用户的访问模型，包括：

用户id,最近一次pc端访问日期,最近一次app端访问日期,最近一次pc端访问的session,最近一次pc端访问的cookies ,最近一次pc的pv量,最近一次pc端使用的浏览器,最近一次pc端的操作系统,最近一次访问的appname ,最近一次app端访问的操作系统,最近一次访问的ip,最近一次访问的城市,最近一次访问省份,第一次pc访问的日期 ,第一次app访问的日期,第一次pc访问的session,第一次访问的cookies ,第一次pc的pv ,pc第一次访问的浏览器名称,app第一次访问的浏览器 ,app第一次访问的名称,app第一次访问的os ,第一次访问的ip ,第一次访问的城市,第一次访问的省份 ,近7天访问app的次数 ,近15天访问app的次数,近30天访问app的次数,近60天访问app的次数,近90天访问的app的次数,近7天pc访问的次数 ,近15天pc访问的次数 ,近30天pc访问的次数 ,近60天访问的次数,近90天访问的次数 ,近30天pc访问的天数,近30天pc购买的次数,近30天pc端pv ,近30天pc端平均pv ,近30天pc访问的不同ip数,近30天pc使用最多的ip,近30天pc端不同的cookie ,近30天pc最常用的cookie,近30天pc端最常用的浏览器,近30天pc端最长用的os,近30天0到5点访问次数,近30天6到7点访问次数,近30天8到9点访问的次数,近30天10到12点访问的次数,近30天13到14点访问的次数,近30天15到17点访问的次数,近30天18到19点访问的次数,近30天20到21点访问的次数,近30天22到23点访问的次数。

1. 生成用户的消费模型

根据用户的订单，订单的收货地址，订单地址模型，订单商品表生成用户的消费模型

用户id,第一次下单时间,最近一次下单时间,首单距今时间,尾单距今时间,近30天订单数量（不含退拒）,近30天订单金额（不含退拒）,近60天订单数量（不含退拒）,近60天订单金额（不含退拒）,近90天订单数量（不含退拒）,近90天订单金额（不含退拒）,近30天订单数量（含退拒）,近30天订单金额（含退拒）,近60天购买次数（含退拒）,近60天购买金额（含退拒） ,近90天购买次数（含退拒）,近90天购买金额（含退拒） ,最大消费金额 ,最小消费金额,累计消费次数（不含退拒）,累计消费金额（不含退拒）,客单价（含退拒）,近90天的客单价（含退拒）,常用收货地址 ,常用支付方式 ,近30天购物车次数,近30天购物车件数,近30天购物车提交商品件数,近30天购物车成功率 ,近30天购物车放弃件数,退货商品数量 ,退货商品金额 ,拒收商品数量 ,拒收商品金额 ,最近一次退货时间,学校下单总数,公司下单总数 ,家里下单总数 ,上午下单总数,下午下单总数,晚上下单总数,凌晨下单总数。

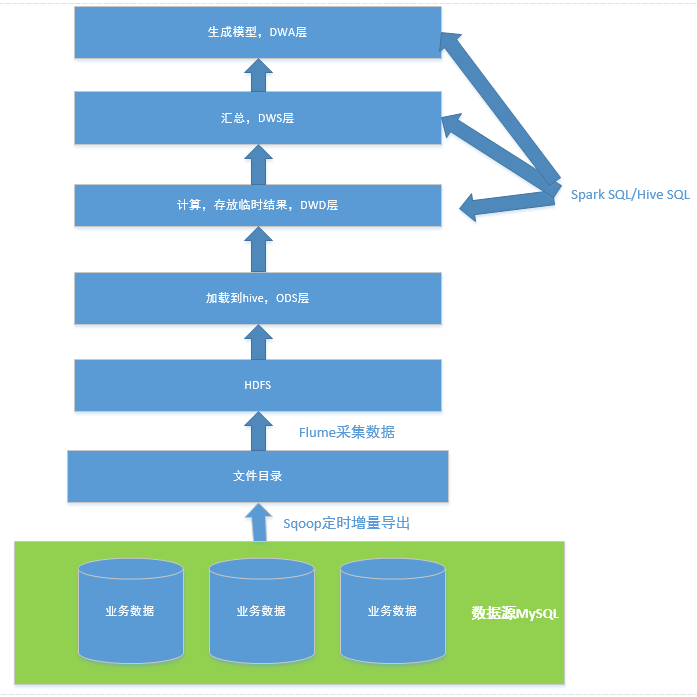
1. 生成用户的品类偏好

根据用户的购物车和订单的购买情况，生成用户的品类偏好，包括

user\_id ,一级类目,一级类目名称,二级类目 ,二级类目名称,三级类目 ,三级类目名称 ,订单30天购买数量,订单30天购买金额,订单90天购买数量,订单90天购买金额,订单180天购买数量,订单180天购买金额,订单总购买数量 ,订单总购买金额 ,购物车30天数量 ,购物车60天数量 ,购物车180天数量,购物车总数量 ,最近一次订单购买该类目的时间,最后一次购买该类目距今天数。

# 架构设计

源数据在mysql中，使用sqoop每天定时导出到文件夹中，使用flume采集文件目录，下沉到hive的ods层的hdfs中，dwd层定时从ods同步数据，并进行粗粒度的汇总，进行汇总计算写到dws中，最后形成用户模型存入到dm中



# 数仓建模

分为四层：ods，dwd，dws，dm

## ODS层：原始数据层

本层数据只保留一天

### 用户信息表：ods\_user

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Code | | DataType |
| 用户id | user\_id | | bigint |
| 登录名 | user\_name | | varchar(64) |
| 用户性别 | user\_gender | | tinyint |
| 出生日期 | user\_birthday | | datetime |
| 年龄 | user\_age | | int |
| 星座 | constellation | | varchar(8) |
| 省份 | province | | varchar(32) |
| 城市 | city | | varchar(32) |
| 城市等级 | | city\_level | tinyint |
| 邮箱 | | e\_mail | varchar(256) |
| 邮箱运营商 | | op\_mail | varchar(32) |
| 手机号 | | mobile | bigint |
| 手机号段 | | num\_seg\_mobile | tinyint |
| 手机运营商 | | op\_Mobile | varchar(64) |
| 注册时间 | | register\_time | datetime |
| 登录ip | | login\_ip | varchar(64) |
| 登录来源 | | login\_source | varchar(512) |
| 邀请人 | | request\_user | bigint |
| 会员积分 | | total\_score | decimal(18,2) |
| 已使用积分 | | used\_score | decimal(18,2) |
| 是否黑名单 | | is\_blacklist | tinyint |
| 是否结婚 | | is\_married | tinyint |
| 学历 | | education | varchar(128) |
| 月收入 | | monthly\_income | decimal(18,2) |
| 职业 | | profession | varchar(128) |

### 用户扩展信息表 ods\_user\_extend

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 性别 | user\_gender bigint | FALSE |
| 是否孕妇 | is\_pregnant\_woman | tinyint |
| 是否有小孩 | is\_have\_children | tinyint |
| 是否有车 | is\_have\_car | tinyint |
| 使用手机品牌 | phone\_brand | varchar(64) |
| 使用手机等级 | phone\_brand\_level | int |
| 使用手机种类数量 | phone\_cnt | int |
| 更换手机次数 | change\_phone\_cnt | int |
| 是否马甲用户 | is\_maja | tinyint |
| 马甲账户数量 | majie\_account\_cnt | int |
| 用户忠诚度 | loyal\_model | int |
| 用户购物类型 | shopping\_type\_model | int |
| 体重 | weight | int |
| 身高 | height | int |

### 用户地址标签表 ods\_user\_addr

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Code | Data Type | 备注 |
| 用户id | user\_id | bigint |  |
| 地址 | order\_addr | varchar(512) |  |
| 地址标签 | user\_order\_flag | tinyint | 1.学校，2单位，3家里 |

### 用户pc端访问记录 ods\_user\_pc\_click\_log

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 会话id | session\_id | varchar(128) |
| cookie的id | cookie\_id | varchar(128) |
| 访问时间 | visit\_time | datetime |
| 访问的URL | visit\_url | varchar(512) |
| 操作系统 | visit\_os | varchar(128) |
| 浏览器名称 | browser\_name | varchar(64) |
| 访问ip | visit\_ip | varchar(64) |
| 省份 | province | varchar(32) |
| 城市 | city | varchar(64) |
| 页面id | page\_id | int |
| 商品id | goods\_id | bigint |
| 商家id | shop\_id | bigint |

### 用户app端访问记录 ods\_user\_app\_click\_log

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 手机唯一标识 | imei | varchar(128) |
| 登录时间 | log\_time | datetime |
| 操作系统 | visit\_os | varchar(64) |
| 操作系统版本 | os\_version | varchar(32) |
| app的名称 | app\_name | varchar(256) |
| app的版本 | app\_version | varchar(32) |
| PUSH码 | device\_token | varchar(512) |
| 访问ip | visit\_ip | varchar(64) |
| 省份 | province | varchar(32) |
| 城市 | city | varchar(64) |

### 用户订单表 ods\_order

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| order\_id | order\_id | bigint |
| 订单编号 | order\_no | varchar(128) |
| 订单日期 | order\_date | datetime |
| 用户id | user\_id | bigint |
| 登录名 | user\_name | varchar(64) |
| 订单金额 | order\_money | decimal(18,2) |
| 订单类型 | order\_type | int |
| 订单状态 | order\_status | int |
| 支付状态 | pay\_status | int |
| 支付方式 | pay\_type | int |
| 订单来源 | order\_source | int |
| 订单更新时间 | update\_time | datetime |

### 用户订单商品表ods\_order\_item

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 订单id | order\_id | bigint |
| 订单名称 | order\_no | varchar(256) |
| 商品id | goods\_id | bigint |
| 商品编号 | goods\_no | varchar(64) |
| 商品名称 | goods\_name | varchar(256) |
| 商品数量 | goods\_amount | int |
| 商家id | shop\_id | bigint |
| 商家名称 | shop\_name | varchar(256) |
| 售卖价格 | curr\_price | decimal(18,2) |
| 市场价 | market\_price | decimal(18,2) |
| 折扣比例 | discount | decimal(18,2) |
| 成本价格 | cost\_price | decimal(18,2) |
| 一级类目 | first\_cart | bigint |
| 一级类目名称 | first\_cart\_name | varchar(256) |
| 二级类目 | second\_cart | bigint |
| 二级类目名称 | second\_cart\_name | varchar(256) |
| 三级类目 | third\_cart | bigint |
| 三级类目名称 | third\_cart\_name | varchar(256) |
| 备注 | desc | varchar(256) |

### 用户订单送货表ods\_order

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 订单id | order\_id | bigint |
| 订单号 | order\_no | varchar(64) 64 |
| 收货人姓名 | consignee | varchar(128) |
| 收货人地址id | area\_id | bigint |
| 地址id的地址名称 | area\_name | varchar(256) |
| 收货人地址 | address | varchar(512) |
| 收货人手机号 | mobile | bigint |
| 收货人电话 | telphone | varchar(32) |
| 使用代金券id | coupon\_id | bigint |
| 使用代金券金额 | coupon\_money | decimal(18,2) |
| 运费 | carriage\_money | decimal(18,2) |
| 创建时间 | create\_time | datetime |
| 更新时间 | update\_time | datetime |
| 入库时间 | dw\_date | datetime |

### 类目码表 ods\_category\_code

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 一级类目id | first\_category\_id | int |
| 一级类目名称 | first\_category\_name | varchar(32) |
| 二级类目id | second\_category\_id | int |
| 二级类目名称 | second\_catery\_name | varchar(32) |
| 三级类目id | third\_category\_id | int |
| 三级类目名称 | third\_category\_name | varchar(32) |

### 购物车表 ods\_order\_cart

|  |  |  |  |
| --- | --- | --- | --- |
| Name | | Code | Data Type |
| id | | id | bigint |
| 会话id | | session\_id | varchar(64) |
| 用户id | | user\_id | bigint |
| 商品id  商品数量 | | goods\_id  goods\_num | bigint  int |
| 添加时间 | add\_time | | datetime |
| 取消时间 | cancle\_time | | datetime |
| 提交时间 | sumbit\_time | | datetime |

## DWD层：原始数据映射层

按天分区，本层基本上是ods层的直接映射，只对pc端的访问轨迹做了些汇聚

### 用户信息表：dwd\_user

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Code | | DataType |
| 用户id | user\_id | | bigint |
| 登录名 | user\_name | | varchar(64) |
| 用户性别 | user\_gender | | tinyint |
| 出生日期 | user\_birthday | | datetime |
| 年龄 | user\_age | | int |
| 星座 | constellation | | varchar(8) |
| 省份 | province | | varchar(32) |
| 城市 | city | | varchar(32) |
| 城市等级 | | city\_level | tinyint |
| 邮箱 | | e\_mail | varchar(256) |
| 邮箱运营商 | | op\_mail | varchar(32) |
| 手机号 | | mobile | bigint |
| 手机号段 | | num\_seg\_mobile | tinyint |
| 手机运营商 | | op\_Mobile | varchar(64) |
| 注册时间 | | register\_time | datetime |
| 登录ip | | login\_ip | varchar(64) |
| 登录来源 | | login\_source | varchar(512) |
| 邀请人 | | request\_user | bigint |
| 会员积分 | | total\_score | decimal(18,2) |
| 已使用积分 | | used\_score | decimal(18,2) |
| 是否黑名单 | | is\_blacklist | tinyint |
| 是否结婚 | | is\_married | tinyint |
| 学历 | | education | varchar(128) |
| 月收入 | | monthly\_income | decimal(18,2) |
| 职业 | | profession | varchar(128) |
| 入库时间 | | dw\_date | datetime |

### 用户扩展信息表 dwd\_user\_extend

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 性别 | user\_gender bigint | FALSE |
| 是否孕妇 | is\_pregnant\_woman | tinyint |
| 是否有小孩 | is\_have\_children | tinyint |
| 是否有车 | is\_have\_car | tinyint |
| 使用手机品牌 | phone\_brand | varchar(64) |
| 使用手机等级 | phone\_brand\_level | int |
| 使用手机种类数量 | phone\_cnt | int |
| 更换手机次数 | change\_phone\_cnt | int |
| 是否马甲用户 | is\_maja | tinyint |
| 马甲账户数量 | majie\_account\_cnt | int |
| 用户忠诚度 | loyal\_model | int |
| 用户购物类型 | shopping\_type\_model | int |
| 体重 | weight | int |
| 身高 | height | int |
| 入库时间 | dw\_date | datetime |

### 用户地址标签表 dwd\_user\_addr

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Code | Data Type | 备注 |
| 用户id | user\_id | bigint |  |
| 地址 | order\_addr | varchar(512) |  |
| 地址标签 | user\_order\_flag | tinyint | 1.学校，2单位，3家里 |
| 入库时间 | dw\_date | datetime | |

### 用户pc端访问记录 dws\_user\_pc\_click\_log

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Code | Data Type |  |
| 用户id | user\_id | bigint |  |
| 会话id | session\_id | varchar(128) |  |
| cookie的id | cookie\_id | varchar(128) |  |
| 访问时间 | in\_time | datetime | 每次会话的最小时间 |
| 跳出时间 | out\_time | datetime | 每次会话的最大访问时间 |
| 停留时长 | stay\_time | bigint | 单位是秒，如果最小访问是和最大访问时间相当，则默认按3秒计算，停留时长的计算方式：每次会话的最大时间-最小时间 |
| 操作系统 | visit\_os | varchar(128) |  |
| 浏览器名称 | browser\_name | varchar(64) |  |
| 访问ip | visit\_ip | varchar(64) |  |
| 省份 | province | varchar(32) |  |
| 城市 | city | varchar(64) |  |
| 入库时间 | dw\_date | datetime |  |
|  | | | |

### 用户app端访问记录 dwd\_user\_app\_click\_log

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 手机唯一标识 | imei | varchar(128) |
| 登录时间 | log\_time | datetime |
| 操作系统 | visit\_os | varchar(64) |
| 操作系统版本 | os\_version | varchar(32) |
| app的名称 | app\_name | varchar(256) |
| app的版本 | app\_version | varchar(32) |
| PUSH码 | device\_token | varchar(512) |
| 访问ip | visit\_ip | varchar(64) |
| 省份 | province | varchar(32) |
| 城市 | city | varchar(64) |
| 入库时间 | dw\_date | datetime |

### 用户订单表 dwd\_order

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| order\_id | order\_id | bigint |
| 订单编号 | order\_no | varchar(128) |
| 订单日期 | order\_date | datetime |
| 用户id | user\_id | bigint |
| 登录名 | user\_name | varchar(64) |
| 订单金额 | order\_money | decimal(18,2) |
| 订单类型 | order\_type | int |
| 订单状态 | order\_status | int |
| 支付状态 | pay\_status | int |
| 支付方式 | pay\_type | int |
| 订单来源 | order\_source | int |
| 订单更新时间 | update\_time | datetime |
| 入库时间 | dw\_date | datetime |

### 用户订单商品表dwd\_order\_item

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 用户id | user\_id | bigint |
| 订单id | order\_id | bigint |
| 订单名称 | order\_no | varchar(256) |
| 商品id | goods\_id | bigint |
| 商品编号 | goods\_no | varchar(64) |
| 商品名称 | goods\_name | varchar(256) |
| 商品数量 | goods\_amount | int |
| 商家id | shop\_id | bigint |
| 商家名称 | shop\_name | varchar(256) |
| 售卖价格 | curr\_price | decimal(18,2) |
| 市场价 | market\_price | decimal(18,2) |
| 折扣比例 | discount | decimal(18,2) |
| 成本价格 | cost\_price | decimal(18,2) |
| 一级类目 | first\_cart | bigint |
| 一级类目名称 | first\_cart\_name | varchar(256) |
| 二级类目 | second\_cart | bigint |
| 二级类目名称 | second\_cart\_name | varchar(256) |
| 三级类目 | third\_cart | bigint |
| 三级类目名称 | third\_cart\_name | varchar(256) |
| 备注 | desc | varchar(256) |
| 入库时间 | dw\_date | datetime |

### 用户订单送货表dwd\_order

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 订单id | order\_id | bigint |
| 订单号 | order\_no | varchar(64) 64 |
| 收货人姓名 | consignee | varchar(128) |
| 收货人地址id | area\_id | bigint |
| 地址id的地址名称 | area\_name | varchar(256) |
| 收货人地址 | address | varchar(512) |
| 收货人手机号 | mobile | bigint |
| 收货人电话 | telphone | varchar(32) |
| 使用代金券id | coupon\_id | bigint |
| 使用代金券金额 | coupon\_money | decimal(18,2) |
| 运费 | carriage\_money | decimal(18,2) |
| 创建时间 | create\_time | datetime |
| 更新时间 | update\_time | datetime |
| 入库时间 | dw\_date | datetime |
| 入库时间 | dw\_date | datetime |

### 类目码表 dwd\_category\_code

|  |  |  |
| --- | --- | --- |
| Name | Code | Data Type |
| 一级类目id | first\_category\_id | int |
| 一级类目名称 | first\_category\_name | varchar(32) |
| 二级类目id | second\_category\_id | int |
| 二级类目名称 | second\_catery\_name | varchar(32) |
| 三级类目id | third\_category\_id | int |
| 三级类目名称 | third\_category\_name | varchar(32) |
| 入库时间 | dw\_date | datetime |

### 购物车表 dwd\_order\_cart

|  |  |  |  |
| --- | --- | --- | --- |
| Name | | Code | Data Type |
| id | | id | bigint |
| 会话id | | session\_id | varchar(64) |
| 用户id | | user\_id | bigint |
| 商品id  商品数量 | | goods\_id  goods\_num | bigint  int |
| 添加时间 | add\_time | | datetime |
| 取消时间 | cancle\_time | | datetime |
| 提交时间 | sumbit\_time | | datetime |
| 入库时间 | dw\_date | | datetime |

# 详细设计

## 用户访问模型

将用户表和用户扩展表合并为一个表dm\_user\_basic

## ODS=>DWD ods层的数据抽取到mid层

有些表映射吴变化，只是增加dw\_date字段，dw字段为入数据仓库的时间，取值为: from\_unixtime(unix\_timestamp())，包括

mid\_order

mid\_order\_cart

mid\_order\_desc

mid\_order\_goods

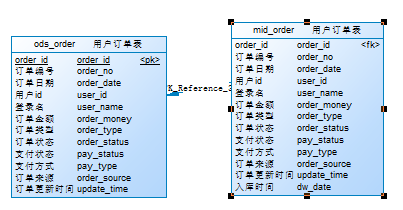
mid\_user\_wide

mid\_user\_order\_addr\_model

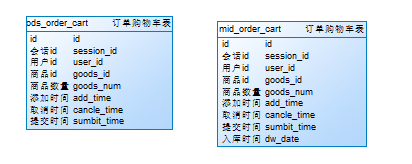
mid\_category\_code

字段一一对应进行抽取

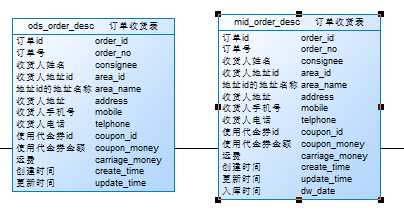
mid\_order



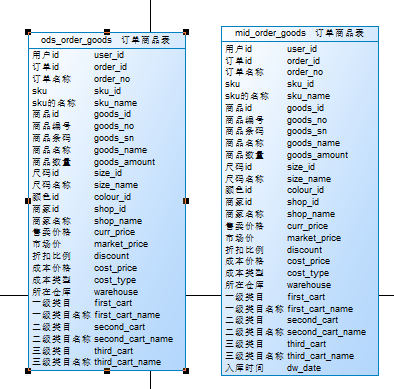
mid\_order\_cart



mid\_order\_desc



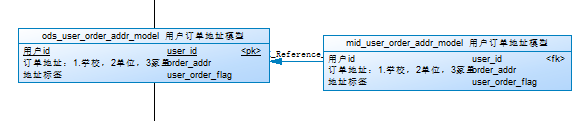
mid\_order\_goods



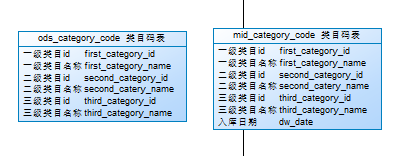
mid\_user\_wide



mid\_user\_order\_addr\_model



mid\_category\_code



## ODS=>DWD统计用户PC端的访问轨迹

按照session\_id,

cookie\_id,

visit\_os,

browser\_name,

visit\_ip,

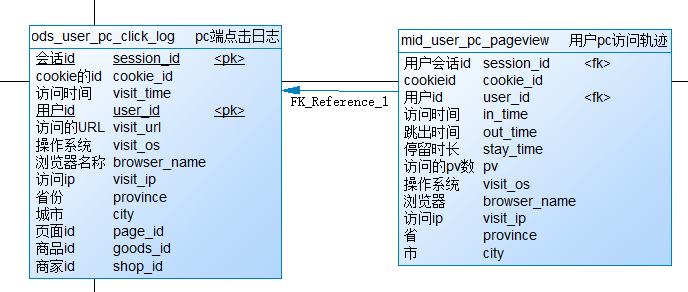
province,

city分

组统计访问间visit\_time,跳出时间out\_time,停留时间 stay\_time,访问的pv数pv,

其中访问时间为最小访问时间，跳出时间为最大访问时间，停留时间为最大访问时间-最小访问时间段的秒数，如果最大访问时间和最小访问时间相等，则停留时间按5秒算。

表映射关系如下

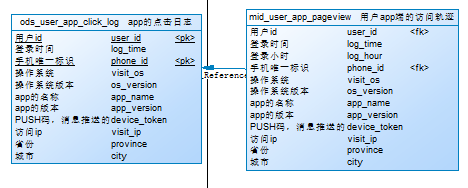


语句：

|  |
| --- |
| INSERT overwrite TABLE mid.mid\_user\_pc\_pageview  pc\_pageview PARTITION (dt = '2018-12-20')  SELECT  t.session\_id,  t.cookie\_id,  t.user\_id,  MIN(visit\_time) in\_time,  MAX(visit\_time) out\_time,  (  case  WHEN MIN(visit\_time) = MAX(visit\_time)  then 5  else unix\_timestamp(MAX(visit\_time)) - unix\_timestamp(MIN(visit\_time))  end  ) stay\_time,  COUNT(1) pv,  t.visit\_os,  t.browser\_name,  t.visit\_ip,  t.province,  t.city  FROM  bdm.bdm\_user\_pc\_click\_log t  WHERE dt = '2018-12-20'  GROUP BY t.session\_id,  t.cookie\_id,  t.user\_id,  t.visit\_os,  t.browser\_name,  t.visit\_ip,  t.province,  t.city ; |

## ODS=>DWD统计app端用户访问轨迹

原样抽取数据即可，增加一个登录的小时，使用hour(log\_time)即可获得



语句

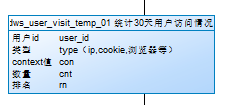
|  |
| --- |
| INSERT overwrite TABLE mid.mid\_user\_app\_pageview PARTITION (dt = '2018-12-20')  SELECT  t.user\_id,  t.log\_time,  HOUR(t.log\_time) log\_hour,  t.phone\_id,  t.visit\_os,  t.os\_version,  t.app\_name,  t.app\_version,  t.device\_token,  t.visit\_ip,  t.province,  t.city  FROM  ods.ods\_user\_app\_click\_log t  WHERE dt = '2018-12-20' ; |

## DWD=>DWS 用户宽表

将mid\_user\_wide和mid\_user\_extend表进行合并，装入dws\_user\_basic表中



## DWD=>DWS统计近30天用户访问汇总信息

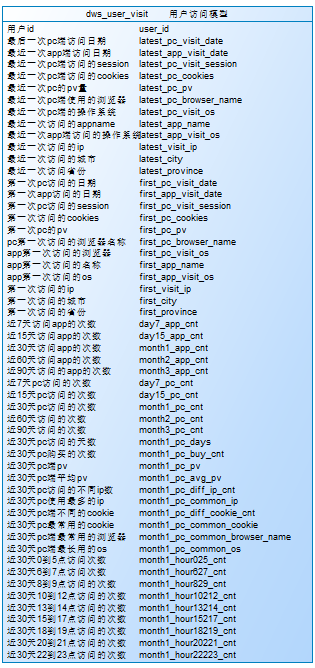


计算近30天的“visit\_ip”，‘cookie\_id’，‘browser\_name’，‘visit\_os’，的各项统计值，并将结果集进行合并

|  |
| --- |
| create table dws.dws\_user\_visit\_temp\_01 as  select  t.user\_id,  t.type,  t.con,  t.cnt,  t.rn  from  (select  b.user\_id,  b.con,  b.type,  b.cnt,  row\_number () over (  distribute by b.user\_id,  b.type sort by b.cnt desc  ) rn  from  (select  a.user\_id,  a.visit\_ip con,  'visit\_ip' type, --近30天常用的IP地址  count(1) cnt  from  mid.mid\_user\_pc\_pageview a  where dt >= date\_add('2018-12-20', -29)  group by  a.user\_id,  a.visit\_ip  union  all  select  a.user\_id,  a.cookie\_id con,  'cookie\_id' type, --近30天常用的cookie  count(1) cnt  from  mid.mid\_user\_pc\_pageview a  where dt >= date\_add('2018-12-20', -29)  group by  a.user\_id,  a.cookie\_id  union  all  select  a.user\_id,  a.browser\_name con,  'browser\_name' type, --近30天常用的游览器  count(1) cnt  from  mid.mid\_user\_pc\_pageview a  where dt >= date\_add('2018-12-20', -29)  group by  a.user\_id,  a.browser\_name  union  all  select  a.user\_id,  a.visit\_os con,  'visit\_os' type, --近30天常用的系统  count(1) cnt  from  mid.mid\_user\_pc\_pageview a  where dt >= date\_add('2018-12-20', -29)  group by a.user\_id,  a.visit\_os) b) t ; |

’

## DWD=>DWS统计各项指标，形成用户访问模型



# 环境搭建

# 项目可视化