## 达摩盘分析

1. 项目概述

达摩盘（DMP，Data Management Platform）是阿里巴巴旗下的一款数据管理平台，主要用于精准营销和用户画像分析，帮助商家在阿里生态（如淘宝、天猫、支付宝等）中更精准地触达目标用户。、基础特征包括年龄、性别、年代、身高、体重、星座 6 类标签。

1. 需求范围

|  |  |  |  |
| --- | --- | --- | --- |
| **标签类别** | **子标签** | **数据来源** | **更新频率** |
| 年龄 | 6种  （18-50岁以上） | 淘宝平台用户行为日志数据 | 每日更新 |
| 性别 | 男性、女性和家庭用户 | 用户行为数据、商品一级品类列表 | 每日更新+全量更新：每月对历史用户进行一次全量重新计算 |
| 身高 | 数值（cm） | 实名资料、订单尺码、智能硬件 | 实时/T+1更新 |

|  |  |  |  |
| --- | --- | --- | --- |
| 体重 | 数值（kg）或NULL | 订单信息、会员资料、智能设备 | T+1日/实时更新 |
| 星座 | 12星座或未知 | 用户生日字段 | 实时/离线批量更新 |

1. 用户年龄

* 主数据：用户出生年龄
* 辅助数据：用户行为日志（类目浏览、品牌偏好等）

1. 用户性别

* 主数据：用户出生性别加上一个家庭用户。
* 辅助数据：用户行为日志、商品一级品类列表。

用户来源是从用户行为、商品一级品类列表来获取。获取性别高级做法，从用户行为来判断，以购买行为、加购/收藏、浏览行为，依次降低权重，来判断用户性别。

1. 用户体重

用户体重的单位是千克，数据需要经过清洗和转换操作来让单位统一，统一单位为千克。如果用户的体重信息填写的过高、过低，则要剔除这些很有可能为无用信息的信息。

1. 用户身高

用户身高标的单位是千克，

1. ADS层表设计
2. 表结构

表（ads\_user\_profile\_basic\_tags）结构如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 类型 | 是否主键 | 说明 |
| user\_id | BIGINT | 是 | 用户唯一ID |
| age\_group | VARCHAR | 否 | 年龄段标签（如25-29岁） |
| gender | VARCHAR | 否 | 性别标签（男/女/家庭） |
| height\_cm | VARCHAR | 否 | 身高（厘米，1位小数） |
| weight\_kg | VARCHAR | 否 | 体重（千克，1位小数） |
| constellation | VARCHAR | 否 | 星座（如“白羊座”） |
| dt | DATE | 分区键 | 数据日期（T+1更新） |

1. 分区与索引

* 分区存储：按日期（dt）进行分区，保留近一年的数据。
* 索引：user\_id是主键索引，age\_group和gender是二级索引。