

报名主题建模设计文档

1、建模设计

- 主要内容：按照需求划分事实表和维度表，事实数据表有哪些、维度数据表有哪些
- ==示例如下==
 - 事实表
 - customer_appeal：线索申诉信息表
 - customer_relationship：客户意向表
 - itcast_clazz：班级信息表
 - employee：员工信息表
 - scrm_department：员工部门表
 - 维度表
 - t_brand：品牌信息表
 - t_date：日期信息表
 -

2、分层设计

-  流程图
 - ODS（数据临时存储层）

属性	设计
功能	将各个数据源的 原始数据 几乎无处理地存放在数据仓库系统中，结构上与源系统基本保持一致，是数据仓库的数据准备区。
来源	MySQL，业务系统数据库
格式	Hive分区表，ORC格式

- DWD（明细数据层 Data Warehouse Detail）

属性	设计
功能	对来自ODS数据层做一些 数据清洗 和规范化的操作，提供更干净的数据，提供一定的 数据质量 保证，同时 划分事实表与维度表 。
来源	数据来自于ODS，一般保持和ODS层数据一样的粒度
格式	Hive分区表，orc格式

◦ **DWB（基础数据层 Data Warehouse Base）**

属性	设计
功能	基础数据层一般用作中间层。多张表数据关联在一起，降维操作，形成一张 大宽表 ，将后面需要用到的多张表的字段提前合并到一张表中
来源	数据来自于DWD
格式	Hive分区表，orc格式

◦ **DWS（服务数据层 Data Warehouse Service）**

属性	设计
功能	基于DWB上的基础数据，整合汇总成分析某一个主题域的服务数据，基于分析主题根据相关指标 维度进行提前统计聚合操作, 形成宽表==统计结果数据
来源	数据来自于DWB
格式	Hive分区表，orc格式

RPT（报表应用层 DA）

属性	设计
功能	对外提供数据报表的数据支持，提供每个需求对应的结果表
来源	数据来自于DWS
格式	Hive分区表，textFile格式

3、设计明细

- 主要内容：根据需求以及每一层的功能，描述每一层有哪些表以及每张表的信息

表名
注释
字段信息
列的分隔符
表的类型
文件类型
是否压缩
同步方式：仅ODS层及DWD有同步方式

• ==示例如下==

- 需求：统计每天每个品牌的订单个数、订单金额，统计每天每种品牌首字母的订单个数、订单金额

- 指标：订单个数、订单金额
- 维度：天、品牌、品牌首字母

◦ ODS层

■ 订单基础信息表

- 表名：ods_shop_order
- 注释：订单基础信息表
- 字段信息
image-20220426161358558
- 列的分隔符：\t
- 表的类型：分区表
- 文件类型：orc
- 是否压缩：snappy
- 同步方式：新增及更新同步

■ 品牌信息表

- 表名：ods_brand
- 注释：商品品牌信息表
- 字段信息

id	string,	'序列id'
store_id	string	'店铺id'
brand_pt_id	string	'平台品牌库品牌Id'
brand_name	string	'品牌名称'
brand_image	string	'品牌图片'
initial	string	'品牌首字母'
sort	INT	'排序'
is_use	INT	'0禁用1启用'
goods_state	INT	'商品品牌审核状态 1 审核中,2 通过,3 拒绝'
is_valid	INT	'0 : 失效, 1 : 开启'

- 列的分隔符：\t
- 表的类型：普通表
- 文件类型：textfile
- 是否压缩：否
- 同步方式：全量覆盖

○ DWD层

■ 订单基础信息表

- 表名: dwd_shop_order
- 注释: 订单基础信息表
- 字段信息

id	string	'根据一定规则生成的订单编号'
order_num	string	'订单序号',
buyer_id	string	'买家的userId'
store_id	string	'店铺的id'
order_from	int	'订单来源 1. 安卓; 2. ios; 3. 小程序H5; 4. PC'
order_state	int	'订单状态:1. 已下单; 2. 已付款, 3. 已确认; 4. 配送; 5. 已完成; 6. 退款; 7. 已取消'
create_date	string	'下单时间'
finnshed_time	string	'订单完成时间'
is_settlement	int	'是否结算; 0. 待结算订单; 1. 已结算订单;'
is_delete	int	'订单评价的状态: 0. 未删除; 1. 已删除; (默认0)'
evaluation_state	int	'订单评价的状态: 0. 未评价; 1. 已评价; (默认0)'
way	string	'取货方式: SELF自提; SHOP店铺负责配送'
is_stock_up	int	'是否需要备货 0: 不需要 1: 需要 2: 平台确认备货 3: 已完成备货 4: 平台已经将货物送至店铺'
is_valid	int	'是否有效 0: false; 1: true; 订单是否有效的标志'
dt	string	分区字段

- 列的分隔符: \t
 - 表的类型: 分区表
 - 文件类型: orc
 - 是否压缩: snappy
 - 同步方式: 新增及更新同步
- ### ■ 品牌信息表
- 表名: dwd_brand
 - 注释: 商品品牌信息表
 - 字段信息

id	string,	'序列id'
store_id	string	'店铺id'
brand_pt_id	string	'平台品牌库品牌Id'
brand_name	string	'品牌名称'
brand_image	string	'品牌图片'
initial	string	'品牌首字母'
sort	INT	'排序'
is_use	INT	'0禁用1启用'
goods_state	INT	'商品品牌审核状态 1 审核中,2 通过,3 拒绝'
is_valid	INT	'0 : 失效, 1 : 开启'

- 列的分隔符: \t
- 表的类型: 普通表
- 文件类型: textfile
- 是否压缩: 否
- 同步方式: 全量覆盖

○ DWB层

- 订单信息宽表表
 - 表名: dwb_shop_order
 - 注释: 订单基础信息表
 - 字段信息

```
-- 原来的字段
id                string      '根据一定规则生成的订单编号'
order_num         string      '订单序号',
buyer_id          string      '买家的userId'
store_id          string      '店铺id'
order_from        int         '订单来源 1.安卓; 2.ios; 3.小程序H5; 4.PC'
order_state       int         '订单状态:1.已下单; 2.已付款, 3. 已确认 ;4.配送; 5.已完成; 6.退款;7.已取消'
create_date       string      '下单时间'
finnshed_time     string      '订单完成时间'
is_settlement     int         '是否结算;0.待结算订单; 1.已结算订单;'
is_delete         int         '订单评价的状态:0.未删除; 1.已删除;(默认0)'
evaluation_state  int         '订单评价的状态:0.未评价; 1.已评价;(默认0)'
way               string      '取货方式:SELF自提;SHOP店铺负责配送'
is_stock_up       int         '是否需要备货 0: 不需要 1: 需要 2: 平台确认备货 3:已完成备货 4平台已经将货物送至店铺 '
is_valid          int         '是否有效 0: false; 1: true; 订单是否有效的标志'

-- 从支付表关联得到的字段
pay_id            string      '支付id'
pay_amt           double      '支付金额'
pay_time          string      '支付时间'

-- 从退款表关联得到的字段
```

refund_id	string	'退款id'
refund_amt	double	'退款金额'
refund_time	string	'退款时间'
-- 从订单商品明细表关联得到字段		
pid	string	'商品id'
brand_id	string	'商品品牌id'
-- 拉链表字段		
endTime	string	'拉链表状态终止时间'
.....		
dt	string	分区字段

- 列的分隔符: \t
- 表的类型: 分区表
- 文件类型: orc
- 是否压缩: snappy

◦ DIM层

■ 品牌信息表

- 表名: dim_brand
- 注释: 商品品牌信息表
- 字段信息

id	string,	'序列id'
brand_pt_id	string	'平台品牌库品牌Id'
brand_name	string	'品牌名称'
initial	string	'品牌首字母'
is_use	INT	'0禁用 1启用'
goods_state	INT	'商品品牌审核状态 1 审核中,2 通过,3 拒绝'
is_valid	INT	'0 : 失效, 1 : 开启'

- 列的分隔符: \t
- 表的类型: 普通表
- 文件类型: textfile
- 是否压缩: 否
- 同步方式: 全量覆盖

◦ DWS层

■ 订单主题事实表

- 表名: dws_order
- 注释: 订单主题事实表
- 字段

brand_id	string	'品牌id'
brand_initial	string	'品牌首字母'
order_cnt	int	'订单个数'
order_amt	dicimal(8,2)	'订单金额'
group_type	string	'0-所有品牌维度, 1-所有品牌首字母维度'
dt	string	'分区字段'

- 列的分隔符: \t
- 表的类型: 分区表
- 文件类型: orc
- 是否压缩: snappy

○ ADS层

- 昨日不同品牌订单个数占比图: 饼图
- 不同品牌订单个数占比信息表
 - 表名: ads_brand_order_cnt_rate
 - 注释: 不同品牌订单个数占比信息表
 - 字段

brand_initial	string	'品牌首字母'
order_cnt	int	'订单个数'
order_amt	dicimal(8,2)	'订单金额'
dt	string	'分区字段'

- 列的分隔符: \t
- 表的类型: 分区表
- 文件类型: textfile
- 是否压缩: 否