

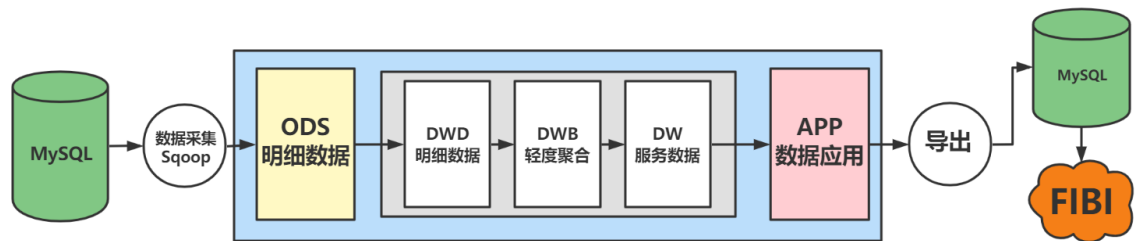
# 考勤分析主题建模设计文档

## 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
    - course\_table\_upload\_detail：班级课表
    - student\_leave\_apply：学生请假申请表(拉链表)
    - tbh\_class\_time\_table：班级作息时间表
    - tbh\_student\_signin\_record：学生打卡记录表
  - 维度表
    - class\_studying\_student\_count：班级在读学生人数表
    - calendar：日期表

## 2、分层设计

- 主要内容：描述整个项目的数仓分层的设计，分哪些层次，每一层的功能是什么等
- ==示例如下==



### ○ ODS（数据临时存储层）

属性	设计
功能	将MySQL数据源的 <b>原始数据</b> 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 <b>数据清洗</b> 和规范化的操作，提供更干净的数据，提供一定的 <b>数据质量保证</b> ，同时 <b>划分事实表与维度表</b> 。
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 <b>大宽表</b> ，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，基于考勤分析主题根据相关指标、维度进行提前统计聚合操作( <b>提前聚合【细粒度轻度聚合：天】</b> )，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

◦ **APP（数据应用层 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，orc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

- **需求1：每天每个班级各时间段的正常出勤人数、出勤率**

- 指标：出勤人数、出勤率
- 维度：天、班级
- 粒度：上午、下午、晚自习

- **需求2：每天各个班级各时间段的迟到人数、迟到率**

- 指标：迟到人数、迟到率
- 维度：天、班级
- 粒度：上午、下午、晚自习

- **需求3：每天各个班级各时间段的请假人数、请假率**

- 指标：请假人数、请假率
- 维度：天、班级
- 粒度：上午、下午、晚自习

- **需求4：每天各个班级各时间段的旷课人数、旷课率**

- 指标：旷课人数、旷课率
- 维度：天、班级
- 粒度：上午、下午、晚自习

- **ODS层**

- 日期表

- 表名：zxjy\_ods.calendar
- 注释：日期表
- 字段信息

```
id          int      comment '主键',  
datelist string comment '日期列'
```

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

- 学生请假申请表

- 表名：zxjy\_ods.student\_leave\_apply
- 注释：学生请假申请表
- 字段信息

```
id          int,
```

class_id	int	comment '班级id',
student_id	int	comment '学员id',
audit_state	int	comment '审核状态 0 待审核 1 通过 2 不通过',
audit_person	int	comment '审核人',
audit_time	string	comment '审核时间',
audit_remark	string	comment '审核备注',
leave_type	int	comment '请假类型 1 请假 2 销假',
leave_reason	int	comment '请假原因 1 事假 2 病假',
begin_time	string	comment '请假开始时间',
begin_time_type	int	comment '1: 上午 2: 下午',
end_time	string	comment '请假结束时间',
end_time_type	int	comment '1: 上午 2: 下午',
days	float	comment '请假/已休天数',
cancel_state	int	comment '撤销状态 0 未撤销 1 已撤销',
cancel_time	string	comment '撤销时间',
old_leave_id	int	comment '原请假id, 只有leave_type =2 销假的时候才有',
leave_remark	string	comment '请假/销假说明',
valid_state	int	comment '是否有效 (0: 无效 1: 有效)',
create_time	string	comment '创建时间',
dt	string	

- 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: 增量同步 (新增和更新)

○ 班级在读学生人数表

- 表名: zxjy\_ods.class\_studying\_student\_count
- 注释: 班级在读学生人数表
- 字段信息

id	int	comment '主键',
school_id	string	comment '校区id',
subject_id	string	comment '学科id',
class_id	string	comment '班级id',
studying_student_count	string	comment '在读班级人数',
studying_date	string	comment '在读日期'
dt	string	

- 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: 增量同步 (仅新增)

○ 班级课表

- 表名: zxjy\_ods.course\_table\_upload\_detail
- 注释: 班级课表
- 字段信息

id	int	comment 'id',
base_id	int	comment '课程主表id',
class_id	int	comment '班级id',
class_date	string	comment '上课日期',
content	string	comment '课程内容',
teacher_id	int	comment '老师id',
teacher_name	string	comment '老师名字',
job_number	string	comment '工号',
classroom_id	int	comment '教室id',
classroom_name	string	comment '教室名称',
is_outline	int	comment '是否大纲 0 否 1 是',
class_mode	int	comment '上课模式 0 传统全天 1 AB上午 2 AB下午 3 线上直播',
is_stage_exam	int	comment '是否阶段考试 (0: 否 1: 是)',
is_pay	int	comment '代课费 (0: 无 1: 有)',
tutor_teacher_id	int	comment '晚自习辅导老师id',
tutor_teacher_name	string	comment '辅导老师姓名',
tutor_job_number	string	comment '晚自习辅导老师工号',
is_subsidy	int	comment '晚自习补贴 (0: 无 1: 有)',
answer_teacher_id	int	comment '答疑老师id',
answer_teacher_name	string	comment '答疑老师姓名',
answer_job_number	string	comment '答疑老师工号',
remark	string	comment '备注',
create_time	string	comment '创建时间'
dt	string	

- 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: 增量同步 (仅新增)
- 班级作息时间表
  - 表名: zxjy\_ods.tbh\_class\_time\_table
  - 注释: 日期表
  - 字段信息

id	int	comment '主键id',
class_id	int	comment '班级id',
morning_template_id	int	comment '上午出勤模板id',
morning_begin_time	string	comment '上午开始时间',
morning_end_time	string	comment '上午结束时间',
afternoon_template_id	int	comment '下午出勤模板id',
afternoon_begin_time	string	comment '下午开始时间',
afternoon_end_time	string	comment '下午结束时间',
evening_template_id	int	comment '晚上出勤模板id',
evening_begin_time	string	comment '晚上开始时间',
evening_end_time	string	comment '晚上结束时间',
use_begin_date	string	comment '使用开始日期',
use_end_date	string	comment '使用结束日期',
create_time	string	comment '创建时间',
create_person	int	comment '创建人',
remark	string	comment '备注'
dt	string	

- 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: 增量同步 (新增和更新)
- 学生打卡记录表
  - 表名: zxjy\_ods.tbh\_student\_signin\_record
  - 注释: 学生打卡记录表
  - 字段信息

```

id                int                comment '主键id',
normal_class_flag int                comment '是否正课 1 正课
time_table_id     int                comment '作息时间表id 关联
class_id          int                comment '班级id',
student_id        int                comment '学员id',
signin_time       string             comment '签到时间',
signin_date       string             comment '签到日期',
inner_flag        int                comment '内外网标志 0
signin_type       int                comment '签到类型 1 心跳
share_state       int                comment '共享屏幕状态 0
inner_ip          string             comment '内网ip地址'
dt                string

```

- 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: 增量同步 (仅新增)

## 分割线--后面的层做到再继续完善

### • 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

- 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

- 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