

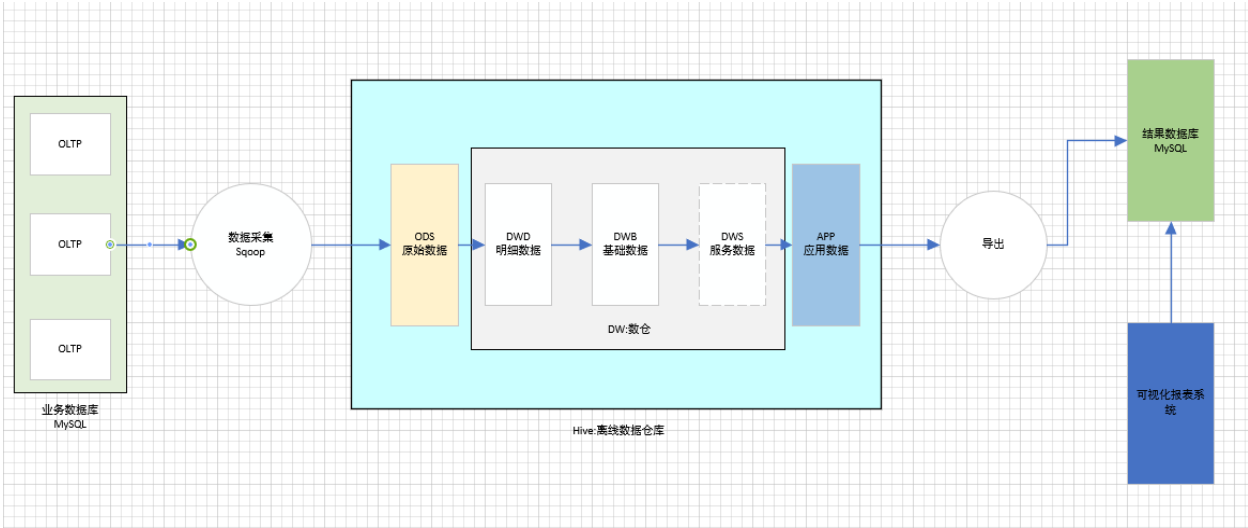
意向线索分析主题建模设计文档

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_dim
 - dim_customer_relationship: 客户意向表
 - dim_customer_clue: 客户线索表 新老学员
 - 维度表_fact
 - fact_customer: 客户静态信息表 地区
 - fact_customer_appeal: 线索申诉信息表 线索是否有效
 - fact_itcast_school: 校区信息表 校区
 - fact_itcast_subject: 学科信息表 学科
 - fact_scrm_department: 员工部门表 咨询中心
 - fact_employee: 员工信息表 关联得到部门维度

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(提前聚合【细粒度轻度聚合：天】), 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 (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DM
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1：每天/每月/每年线上线下以及新老学员的意向用户个数
 - 指标：意向用户个数
 - 维度：时间(每天/每月/每年)、线上线下、新老学员
- ODS层
 - 订单基础信息表
 - 表名：customer
 - 注释：记录了所有学员的所有信息，主要提供地区维度 id: 学员id area: 地区
 - 字段信息

```

-- 客户静态信息表
-- auto-generated definition
create temporary table customer
(
    id int,
    customer_relationship_id int comment '当前意向id',
    create_date_time string comment '创建时间',
    update_date_time string comment '最后更新时间',
    deleted string comment '是否被删除（禁用）',
    name string comment '姓名',
    idcard string comment '身份证号',
    birth_year int comment '出生年份',
    gender string comment '性别',
    phone string comment '手机号',
    wechat string comment '微信',
    qq string comment 'qq号',
    email string comment '邮箱',
    area string comment '所在区域',
    leave_school_date string comment '离校时间',
    graduation_date string comment '毕业时间',
    bxg_student_id string comment '博学谷学员ID，可能未关联到，不存在',
    creator int comment '创建人ID',
    origin_type string comment '数据来源',
    origin_channel string comment '来源渠道',
    tenant int,
    md_id int comment '中台id',
    dt string
);

alter table customer
    set tblproperties ('comment' = '客户静态信息表');

```

- 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
- 线索申诉信息表
 - 表名：customer_appeal
 - 注释：辅助判断线索是否有效 customer_relationship_first_id:意向id appeal_status:申诉状态
 - 字段信息

```

-- 线索申诉信息表
-- auto-generated definition
create temporary table customer_appeal
(
    id int comment '主键',
    customer_relationship_first_id int comment '第一条客户关系id',
    employee_id int comment '申诉人',
    employee_name string comment '申诉人姓名',
    employee_department_id int comment '申诉人部门',
    employee_tdepart_id int comment '申诉人所属部门',
    appeal_status int comment '申诉状态, 0:待稽核 1:无效 2:有效',
    audit_id int comment '稽核人id',
    audit_name string comment '稽核人姓名',
    audit_department_id int comment '稽核人所在部门',
    audit_department_name string comment '稽核人部门名称',
    audit_date_time string comment '稽核时间',
    create_date_time string comment '创建时间(申诉时间)',
    update_date_time string comment '更新时间',
    deleted string comment '删除标志位',
    tenant int,
    dt string
);

alter table customer_appeal
    set tblproperties ('comment' = '线索申诉信息表');

```

- 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: 仅新增
- 客户线索表
 - 表名: customer_clue
 - 注释: 线索分析: 记录了每个意向的信息的线索信息, 用于计算线索个数, 提供新老学员维度, 新线索就是新学员, 老线索就是老学员 id:线索id customer_relationship_id: 意向id clue_stat: VALID_NEW_CLUES是新线索
 - 字段信息

```

-- 客户线索表
-- auto-generated definition
create temporary table customer_clue
(
    id int,
    create_date_time string comment '创建时间',
    update_date_time string comment '最后更新时间',
    deleted string comment '是否被删除（禁用）',
    customer_id int comment '客户id',
    customer_relationship_id int comment '客户关系id',
    session_id string comment '七陌会话id',
    sid string comment '访客id',
    status string comment '状态（undeal待领取 deal 已领取 finish 已关闭 changePeer 已流转）',
    `user` string comment '所属坐席',
    create_time string comment '七陌创建时间',
    platform string comment '平台来源（pc-网站咨询|wap-wap咨询|sdk-app咨询|weixin-微信咨询）',
    s_name string comment '用户名称',
    seo_source string comment '搜索来源',
    seo_keywords string comment '关键字',
    ip string comment 'IP地址',
    referrer string comment '上级来源页面',
    from_url string comment '会话来源页面',
    landing_page_url string comment '访客着陆页面',
    url_title string comment '咨询页面title',
    to_peer string comment '所属技能组',
    manual_time string comment '人工开始时间',
    begin_time string comment '坐席领取时间',
    reply_msg_count int comment '客服回复消息数',
    total_msg_count int comment '消息总数',
    msg_count int comment '客户发送消息数',
    comment string comment '备注',
    finish_reason string comment '结束类型',

```

```

finish_user          string comment '结束坐席',
end_time             string comment '会话结束时间',
platform_description string comment '客户平台信息',
browser_name         string comment '浏览器名称',
os_info              string comment '系统名称',
area                 string comment '区域',
country              string comment '所在国家',
province             string comment '省',
city                 string comment '城市',
creator              int  comment '创建人',
name                 string comment '客户姓名',
idcard               string comment '身份证号',
phone                string comment '手机号',
itcast_school_id     int  comment '校区Id',
itcast_school        string comment '校区',
itcast_subject_id    int  comment '学科Id',
itcast_subject       string comment '学科',
wechat               string comment '微信',
qq                   string comment 'qq号',
email                string comment '邮箱',
gender               string comment '性别',
level                string comment '客户级别',
origin_type           string comment '数据来源渠道',
information_way       string comment '资讯方式',
working_years         string comment '开始工作时间',
technical_directions string comment '技术方向',
customer_state        string comment '当前客户状态',
valid                string comment '该线索是否是网资有效线索',
anticipat_signup_date string comment '预计报名时间',
clue_state            string comment '线索状态',
scrm_department_id   int  comment 'SCRM内部部门id',
superior_url          string comment '诸葛获取上级页面URL',
superior_source       string comment '诸葛获取上级页面URL标题',
landing_url           string comment '诸葛获取着陆页面URL',
landing_source        string comment '诸葛获取着陆页面URL来源',
info_url              string comment '诸葛获取留咨页URL',
info_source           string comment '诸葛获取留咨页URL标题'
);

info_source          string comment '诸葛获取留咨页URL标题',
origin_channel        string comment '投放渠道',
course_id             int,
course_name           string,
zhuge_session_id      string,
is_repeat             int  comment '是否重复线索(手机号维度) 0:正常 1:重复',
tenant                int  comment '租户id',
activity_id           string comment '活动id',
activity_name         string comment '活动名称',
follow_type           int  comment '分配类型, 0-自动分配, 1-手动分配, 2-自动转移, 3-手动单个转移, 4-手动批量转移, 5-公海领取',
shunt_mode_id         int  comment '匹配到的技能组id',
shunt_employee_group_id int comment '所属分派员工组',
dt                    string

);

alter table customer_clue
set tblproperties ('comment' = '客户线索表');

```

- 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

■ 客户意向表

■ 表名: customer_relationship

- 注释: 意向分析和报名分析: 存储了所有意向和报名信息 id: 意向id, 唯一标记一条意向信息 create_date_time: 创建日期 deleted: 合法性判断的字段, 如果为1, 表示这一条意向信息已经被删除, 不用参加统计 customer_id: 用户id origin_type: 线上线下, NETSERVICE or PRESIGNUP 表示线上, 其他的值表示线下 itcast_school_id: 校区id itcast_subject_id: 学科id origin_channel: 来源渠道 creator: 负责这个意向信息的销售id payment_state: 支付状态 payment_time: 支付时间

■ 字段信息

```
-- 客户意向表
-- auto-generated definition
create temporary table customer_relationship
(
    id                                int,
    create_date_time                 string,
    update_date_time                 string comment '最后更新时间',
    deleted                           string comment '是否被删除(禁用)',
    customer_id                      int comment '所属客户id',
    first_id                         int comment '第一条客户关系id',
    belonger                         int comment '归属人',
    belonger_name                     string comment '归属人姓名',
    initial_belonger                 int comment '初始归属人',
    distribution_handler              int comment '分配处理人',
    business_scrm_department_id      int comment '归属部门',
    last_visit_time                   string comment '最后回访时间',
    next_visit_time                   string comment '下次回访时间',
    origin_type                       string comment '数据来源',
    itcast_school_id                 int comment '校区Id',
    itcast_subject_id                int comment '学科Id',
    intention_study_type              string comment '意向学习方式',
    anticipat_signup_date             string comment '预计报名时间',
    level                            string comment '客户级别',
    creator                           int comment '创建人',
    current_creator                   int comment '当前创建人: 初始==创建人, 当在公海拉回时为 拉回人',
    creator_name                     string comment '创建者姓名',
    origin_channel                    string comment '来源渠道',
    comment                           string,
    first_customer_clue_id            int comment '第一条线索id',
    last_customer_clue_id             int comment '最后一条线索id',
    process_state                     string comment '处理状态',
    process_time                      string comment '处理状态变动时间',
```



```

payment_state      string comment '支付状态',
payment_time       string comment '支付状态变动时间',
signup_state       string comment '报名状态',
signup_time        string comment '报名时间',
notice_state       string comment '通知状态',
notice_time        string comment '通知状态变动时间',
lock_state         string comment '锁定状态',
lock_time          string comment '锁定状态修改时间',
itcast_clazz_id    int comment '所属ems班级id',
itcast_clazz_time   string comment '报班时间',
payment_url        string comment '付款链接',
payment_url_time    string comment '支付链接生成时间',
ems_student_id     int comment 'ems的学生id',
delete_reason      string comment '删除原因',
deleter            int comment '删除人',
deleter_name       string comment '删除人姓名',
delete_time        string comment '删除时间',
course_id          int comment '课程ID',
course_name        string comment '课程名称',
delete_comment     string comment '删除原因说明',
close_state        string comment '关闭装填',
close_time         string comment '关闭状态变动时间',
appeal_id          int comment '申诉id',
tenant             int comment '租户',
total_fee          decimal(19) comment '报名费总金额',
belonged           int comment '小周期归属人',
belonged_time      string comment '归属时间',
belonger_time      string comment '归属时间',
transfer           int comment '转移人',
transfer_time      string comment '转移时间',
follow_type        int comment '分配类型, 0-自动分配, 1-手动分配, 2-自动转移, 3-手动单个转移, 4-手动批量转移, 5-公海领取',
transfer_bxg_oa_account string comment '转移到博学谷归属人OA账号',
transfer_bxg_belonger_name string comment '转移到博学谷归属人OA姓名',
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: 新增及更新同步
- 员工信息表
 - 表名: employee
 - 注释: 记录每个员工的信息: 为了关联得到部门维度 id: 员工id tdepart_id: 销售部门id
 - 字段信息

```

create temporary table employee
(
    id                int,
    email             string comment '公司邮箱，OA登录账号',
    real_name         string comment '员工的真实姓名',
    phone             string comment '手机号，目前还没有使用；隐私问题OA接口没有提供这个属性，',
    department_id     string comment 'OA中的部门编号，有负值',
    department_name    string comment 'OA中的部门名',
    remote_login      string comment '员工是否可以远程登录',
    job_number        string comment '员工工号',
    cross_school      string comment '是否有跨校区权限',
    last_login_date   string comment '最后登录日期',
    creator           int comment '创建人',
    create_date_time  string comment '创建时间',
    update_date_time  string comment '最后更新时间',
    deleted           string comment '是否被删除（禁用）',
    scrm_department_id int comment 'SCRM内部部门id',
    leave_office      string comment '离职状态',
    leave_office_time string comment '离职时间',
    reinstated_time   string comment '复职时间',
    superior_leaders_id int comment '上级领导ID',
    tdepart_id        int comment '直属部门',
    tenant            int,
    ems_user_name     string,
    dt                string
);

alter table employee
    set tblproperties ('comment' = '员工信息表');

```

- 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
- 班级信息表
 - 表名：itcast_clazz
 - 注释：用于报名分析：线索分析和意向分析用不到，提供学科和校区维度
 - id：班级id
 - itcast_school_id：校区id
 - itcast_school_name：校区名称
 - itcast_subject_id：学科id
 - itcast_subject_name：学科名称
 - 字段信息

Data Source: zx_hive

Schema: zx_ods

Table: itcast_clazz

-- 班级信息表

-- auto-generated definition

```
create temporary table itcast_clazz
(
    id                int comment 'ems课程id(非自增)',
    create_date_time  string comment '创建时间',
    update_date_time  string comment '最后更新时间',
    deleted           string comment '是否被删除（禁用）',
    itcast_school_id  string comment 'ems校区ID',
    itcast_school_name string comment 'ems校区名称',
    itcast_subject_id string comment 'ems学科ID',
    itcast_subject_name string comment 'ems学科名称',
    itcast_brand       string comment 'ems品牌',
    clazz_type_state   string comment '班级类型状态',
    clazz_type_name     string comment '班级类型名称',
    teaching_mode       string comment '授课模式',
    start_time         string comment '开班时间',
    end_time           string comment '毕业时间',
    comment            string comment '备注',
    detail             string comment '详情(比如: 27期)',
    uncertain          string comment '待定班(0:否,1:是)',
    tenant             int,
    dt                string
);

alter table itcast_clazz
    set tblproperties ('comment' = '班级信息表');
```

- 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
- 校区信息表
 - 表名: itcast_school
 - 注释: 意向分析 id: 校区id name: 校区名称
 - 字段信息

Data Source: zx_hive

Schema: zx_ods

Table: itcast_school

```
-- 校区信息表
-- auto-generated definition
create temporary table itcast_school
(
    id                int,
    create_date_time  string comment '创建时间',
    update_date_time  string comment '最后更新时间',
    deleted           string comment '是否被删除（禁用）',
    name              string comment '校区名称',
    code              string,
    tenant            int,
    dt                string
);

alter table itcast_school
    set tblproperties ('comment' = '校区信息表');
```

- 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
- 学科信息表
 - 表名: itcast_school
 - 注释: 意向分析 id: 学科id name: 学科名称
 - 字段信息

```

-- 学科信息表
-- auto-generated definition
create temporary table itcast_subject
(
    id                int,
    create_date_time  string comment '创建时间',
    update_date_time  string comment '最后更新时间',
    deleted           string comment '是否被删除（禁用）',
    name              string comment '学科名称',
    code              string,
    tenant            int,
    dt                string
);

alter table itcast_subject
    set tblproperties ('comment' = '学科信息表');

```

- 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
- 员工部门表
 - 表名：scrm_department
 - 注释：意向分析、报名分析 记录了整个公司所有销售部门信息 id：部门id name：部门名称
 - 字段信息

```

-- 员工部门表
-- auto-generated definition
create temporary table scrm_department
(
    id                int comment '部门id',
    name              string comment '部门名称',
    parent_id         int comment '父部门id',
    create_date_time  string comment '创建时间',
    update_date_time  string comment '更新时间',
    deleted           string comment '删除标志',
    id_path           string comment '编码全路径',
    tdepart_code       int comment '直属部门',
    creator           string comment '创建者',
    depart_level       int comment '部门层级',
    depart_sign        int comment '部门标志，暂时默认1',
    depart_line        int comment '业务线，存储业务线编码',
    depart_sort        int comment '排序字段',
    disable_flag       int comment '禁用标志',
    tenant            int,
    dt                string
);

alter table scrm_department
    set tblproperties ('comment' = '员工部门表');

```

- 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

• 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