

## # 离线数据仓库建设

### ## 业务数据介绍

现有业务系统数据如下

#### - 订单信息表--->order

字段名	类型	描述
order_id	string	订单唯一id
order_date	string	下单日期
order_amount	int	订单数量
product_ID	int	产品ID
product_type	string	产品品类
product_name	string	产品名称
customer_ID	string	客户ID
customer_name	string	顾客名字
customer_sex	int	顾客性别
order_type	int	交易类型
region_ID	int	销售区域ID
sale_region_big	string	大区
sale_region_nation	string	国家
sale_region_small	string	城市
price	int	单价
sales_amount	int	销售金额

#### - 产品成本信息表---->cost

字段名	类型	描述
order_id	string	订单编号
product_cost	string	产品成本

#### - 销售区域信息表--->region

字段名	类型	描述
sale_region_id	int	销售区域ID
sale_region_big	string	销售大区
sale_region_nation	string	国家
sale_region_small	string	区域

#### - 产品信息表--->product

字段名	类型	描述
product_ID	string	产品ID
product_price	string	产品单价

#### - 下单时间信息表--->date

字段名	类型	描述
order_date	string	下单日期

#### - 客户信息表--->customers

字段名	类型	描述
customer_id	string	客户ID
customer_name	string	客户姓名
customer_birth	string	客户生日
customer_tel	string	客户的电话
customer_sex	string	客户性别
customer_height	string	客户身高
customer_weight	string	客户体重

## ## 项目要求

### ### 使用星型模型建模

- 使用星型模型进行建模
  - 事实表为清洗后的order表
  - 维度表
    - 订单日期信息表
    - 客户信息表
    - 产品信息表
    - 销售区域信息表
    - 产品成本信息表
- 最终效果如下:

#### - 订单信息表--->order

字段名	字段类型	描述
order_id	string	订单编号
order_date	date	下单日期
order_amount	int	订单数量
product_ID	int	产品ID
customer_ID	string	客户ID
order_type	int	交易类型
region_ID	int	销售区域ID
sales_amount	int	销售金额

#### - 产品成本信息表---> cost

字段名	类型	描述
order_id	string	订单编号
product_costs	int	产品成本费用,单位:元

#### - 客户信息表--->customer

字段名	类型	描述
customer_id	string	客户ID
customer_name	string	客户姓名
customer_birth	string	客户生日
customer_tel	string	客户电话
customer_sex	int	客户性别
customer_height	string	客户身高
customer_weight	string	客户体重

#### - 订单日期信息--->dates

字段名	类型	描述
softball_date	date	订单生成日期
softball_year	string	订单生成日期的年
softball_month	string	订单生成日期的月
softball_day	string	订单生成日期的日
softball_quarter	string	订单生成日期的季度
softball_year_month	string	订单生成日期的年月:YYYY-mm
softball_year_quarter_month	string	订单生成日期的年季度月:YYYY-第q季度-MM
softball_week	string	订单生成日期的第几周

#### - 产品信息表--->products

字段名	类型	描述
product_ID	string	产品ID
product_name	string	产品名称
product_type	string	产品类别
product_price	string	产品单价

#### - 销售区域信息表--->regions

字段名	类型	描述
sale_region_id	int	销售区域ID
sale_region_big	string	销售大区
sale_region_nation	string	国家
sale_region_small	string	区域

### ### 数据仓库分层

- 创建DW层, 用来提供服务给业务人员
- 根据 年, 月分区

字段名	类型	描述
order_id	string	订单编号
order_date	date	订单生成日期
order_amount	int	订单数量
product_ID	int	产品id
customer_ID	string	顾客id
order_type	int	订单交易类型
region_ID	int	订单销售区域id
sales_amount	int	订单销售金额