* [公共维度汇总层（DIM）维表规范](https://help.aliyun.com/knowledge_detail/114454.html?spm=a2c4g.11186631.2.2.249f501abZoFzk" \l "title-03s-vq3-3gj)
* [建表示例](https://help.aliyun.com/knowledge_detail/114454.html?spm=a2c4g.11186631.2.2.249f501abZoFzk#title-wlx-yks-tib)

公共维度汇总层（DIM）基于维度建模理念，建立整个企业的一致性维度。

公共维度汇总层（DIM）主要由维度表（维表）构成。维度是逻辑概念，是衡量和观察业务的角度。维表是根据维度及其属性将数据平台上构建的物理化的表，采用宽表设计的原则。因此，公共维度汇总层（DIM）首先需要定义维度。

**定义维度**

在划分数据域、构建总线矩阵时，需要结合对业务过程的分析定义维度。本教程以A电商公司的营销业务板块为例，在交易数据域中，我们重点考察确认收货（交易成功）的业务过程。

在确认收货的业务过程中，主要有商品和收货地点（本教程中，假设收货和购买是同一个地点）两个维度所依赖的业务角度。从商品角度可以定义出以下维度：

* 商品ID
* 商品名称
* 商品价格
* 商品新旧程度

0表示全新，1表示闲置，2表示二手。

* 商品类目ID
* 商品类目名称
* 品类ID
* 品类名称
* 买家ID
* 商品状态

0表示正常，1表示用户删除，2表示下架，3表示从未上架。

* 商品所在城市
* 商品所在省份

从地域角度，可以定义出以下维度：

* 买家ID
* 城市code
* 城市名称
* 省份code
* 省份名称

作为维度建模的核心，在企业级数据仓库中必须保证维度的唯一性。以A公司的商品维度为例，有且只允许有一种维度定义。例如，省份code这个维度，对于任何业务过程所传达的信息都是一致的。

**设计维表**

完成维度定义后，您可以对维度进行补充，进而生成维表。维表的设计需要注意：

* 建议维表单表信息不超过1000万条。
* 维表与其他表进行Join时，建议您使用Map Join。
* 避免过于频繁的更新维表的数据。

在设计维表时，您需要从下列方面进行考虑：

* 维表中数据的稳定性。

例如，A公司电商会员通常不会出现消亡，但会员数据可能在任何时候更新，此时要考虑创建单个分区存储全量数据。如果存在不会更新的记录，您可能需要分别创建历史表与日常表。日常表用于存放当前有效的记录，保持表的数据量不会膨胀；历史表根据消亡时间插入对应分区，使用单个分区存放分区对应时间的消亡记录。

* 维表是否需要垂直拆分。

如果一个维表存在大量属性不被使用，或由于承载过多属性字段导致查询变慢，则需要考虑对字段进行拆分，创建多个维表。

* 维表是否需要水平拆分。

如果记录之间有明显的界限，可以考虑拆成多个表或设计成多级分区。

* 核心维表的产出时间。通常有严格的要求。

设计维表的主要步骤如下：

1. 初步定义维度。

保证维度的一致性。

1. 确定主维表（中心事实表，本教程中采用星型模型）。

此处的主维表通常是数据引入层（ODS）表，直接与业务系统同步。例如，s\_auction是与前台商品中心系统同步的商品表，此表即是主维表。

1. 确定相关维表。

数据仓库是业务源系统的数据整合，不同业务系统或者同一业务系统中的表之间存在关联性。根据对业务的梳理，确定哪些表和主维表存在关联关系，并选择其中的某些表用于生成维度属性。以商品维度为例，根据对业务逻辑的梳理，可以得到商品与类目、卖家和店铺等维度存在关联关系。

1. 确定维度属性。

主要包括两个阶段。第一个阶段是从主维表中选择维度属性或生成新的维度属性；第二个阶段是从相关维表中选择维度属性或生成新的维度属性。以商品维度为例，从主维表（s\_auction）、类目、卖家和店铺等相关维表中选择维度属性或生成新的维度属性。维度属性的设计需要注意：

* + 尽可能生成丰富的维度属性。
  + 尽可能多地给出富有意义的文字性描述。
  + 区分数值型属性和事实。
  + 尽量沉淀出通用的维度属性。

**公共维度汇总层（DIM）维表规范**

公共维度汇总层（DIM）维表命名规范：dim\_{业务板块名称/pub}\_{维度定义}[\_{自定义命名标签}]，pub是与具体业务板块无关或各个业务板块都可公用的维度。例如，时间维度，举例如下：

* 公共区域维表**dim\_pub\_area**
* A公司电商板块的商品全量表**dim\_asale\_itm**

**建表示例**

本例中，最终的维表建表语句如下所示。

**CREATE** **TABLE** **IF** **NOT** **EXISTS** dim\_asale\_itm

(

item\_id **BIGINT** **COMMENT** '商品ID',

item\_title **STRING** **COMMENT** '商品名称',

item\_price **DOUBLE** **COMMENT** '商品成交价格\_元',

item\_stuff\_status **BIGINT** **COMMENT** '商品新旧程度\_0全新1闲置2二手',

cate\_id **BIGINT** **COMMENT** '商品类目ID',

cate\_name **STRING** **COMMENT** '商品类目名称',

commodity\_id **BIGINT** **COMMENT** '品类ID',

commodity\_name **STRING** **COMMENT** '品类名称',

umid **STRING** **COMMENT** '买家ID',

item\_status **BIGINT** **COMMENT** '商品状态\_0正常1用户删除2下架3未上架',

city **STRING** **COMMENT** '商品所在城市',

prov **STRING** **COMMENT** '商品所在省份'

)

**COMMENT** '商品全量表'

PARTITIONED **BY** (ds **STRING** **COMMENT** '日期,yyyymmdd');

**CREATE** **TABLE** **IF** **NOT** **EXISTS** dim\_pub\_area

(

buyer\_id **STRING** **COMMENT** '买家ID',

city\_code **STRING** **COMMENT** '城市code',

city\_name **STRING** **COMMENT** '城市名称',

prov\_code **STRING** **COMMENT** '省份code',

prov\_name **STRING** **COMMENT** '省份名称'

)

**COMMENT** '公共区域维表'

PARTITIONED **BY** (ds **STRING** **COMMENT** '日期分区,格式yyyymmdd')

LIFECYCLE 3600;