公共汇总粒度事实层（DWS）

更新时间：2020-06-17 20:01:02

[**本页目录**](javascript:void(0))

* [公共汇总事实表设计原则](https://help.aliyun.com/knowledge_detail/114458.html?spm=a2c4g.11186631.2.4.e4e2501a6yEpjn#title-8om-e1e-4k0)
* [公共汇总事实表规范](https://help.aliyun.com/knowledge_detail/114458.html?spm=a2c4g.11186631.2.4.e4e2501a6yEpjn#title-v7z-bn4-r5b)
* [建表示例](https://help.aliyun.com/knowledge_detail/114458.html?spm=a2c4g.11186631.2.4.e4e2501a6yEpjn#title-emt-brj-mn1)

公共汇总粒度事实层以分析的主题对象作为建模驱动，基于上层的应用和产品的指标需求构建公共粒度的汇总指标事实表。公共汇总层的一个表通常会对应一个派生指标。

**公共汇总事实表设计原则**

聚集是指针对原始明细粒度的数据进行汇总。DWS公共汇总层是面向分析对象的主题聚集建模。在本教程中，最终的分析目标为：最近一天某个类目（例如：厨具）商品在各省的销售总额、该类目Top10销售额商品名称、各省用户购买力分布。因此，我们可以以最终交易成功的商品、类目、买家等角度对最近一天的数据进行汇总。

**注意**

* 聚集是不跨越事实的。聚集是针对原始星形模型进行的汇总。为获取和查询与原始模型一致的结果，聚集的维度和度量必须与原始模型保持一致，因此聚集是不跨越事实的。
* 聚集会带来查询性能的提升，但聚集也会增加ETL维护的难度。当子类目对应的一级类目发生变更时，先前存在的、已经被汇总到聚集表中的数据需要被重新调整。

此外，进行DWS层设计时还需遵循以下原则：

* 数据公用性：需考虑汇总的聚集是否可以提供给第三方使用。您可以判断，基于某个维度的聚集是否经常用于数据分析中。如果答案是肯定的，就有必要把明细数据经过汇总沉淀到聚集表中。
* 不跨数据域。数据域是在较高层次上对数据进行分类聚集的抽象。数据域通常以业务过程进行分类，例如交易统一划到交易域下， 商品的新增、修改放到商品域下。
* 区分统计周期。在表的命名上要能说明数据的统计周期，例如\_1d表示最近1天，td表示截至当天，nd表示最近N天。

**公共汇总事实表规范**

公共汇总事实表命名规范：dws\_{业务板块缩写/pub}\_{数据域缩写}\_{数据粒度缩写}[\_{自定义表命名标签缩写}]\_{统计时间周期范围缩写}。

* 关于统计实际周期范围缩写，缺省情况下，离线计算应该包括最近一天（\_1d），最近N天（\_nd）和历史截至当天（\_td）三个表。如果出现\_nd的表字段过多需要拆分时，只允许以一个统计周期单元作为原子拆分。即一个统计周期拆分一个表，例如最近7天（\_1w）拆分一个表。不允许拆分出来的一个表存储多个统计周期。
* 对于小时表（无论是天刷新还是小时刷新），都用\_hh 来表示。
* 对于分钟表（无论是天刷新还是小时刷新），都用\_mm来表示。

举例如下：

* dws\_asale\_trd\_byr\_subpay\_1d（A电商公司买家粒度交易分阶段付款一日汇总事实表）
* dws\_asale\_trd\_byr\_subpay\_td（A电商公司买家粒度分阶段付款截至当日汇总表）
* dws\_asale\_trd\_byr\_cod\_nd（A电商公司买家粒度货到付款交易汇总事实表）
* dws\_asale\_itm\_slr\_td（A电商公司卖家粒度商品截至当日存量汇总表）
* dws\_asale\_itm\_slr\_hh（A电商公司卖家粒度商品小时汇总表）---维度为小时
* dws\_asale\_itm\_slr\_mm（A电商公司卖家粒度商品分钟汇总表）---维度为分钟

DWS层数据存储及生命周期管理规范请参见[CDM汇总层设计规范](https://help.aliyun.com/document_detail/117438.html#concept-226672)。