电商商品关联分析数仓流程分析文档

一、数仓分层架构设计

（一）ODS 层（操作数据存储层）

1. 表结构与作用

ods\_goods\_info：存储商品基础信息，包含 goods\_id（商品唯一标识）、goods\_name（商品名称）、shop\_id（所属店铺 ID） ，通过 dt 分区管理不同时间数据，作为商品信息的原始数据源，为下游提供商品基本属性关联依据。

ods\_user\_behavior：记录用户对商品的交互行为，字段 behavior\_id（行为记录 ID）、user\_id（用户 ID）、shop\_id（店铺 ID）、goods\_id（商品 ID）、behavior\_type（行为类型：浏览、加购、支付等）、behavior\_time（行为时间） ，借助 dt 分区留存用户行为轨迹，是分析商品关联关系的核心行为数据源。

ods\_orders：存储订单交易数据，涵盖 order\_id（订单 ID）、shop\_id（店铺 ID）、goods\_id（商品 ID）、user\_id（下单用户 ID）、order\_time（下单时间） ，以 dt 分区维护订单数据，用于后续分析商品购买关联。

2. 数据插入逻辑

通过模拟数据插入语句，按照一定规则（如随机生成商品、用户、店铺关联关系，随机行为类型、时间等 ）向 ODS 层表中插入数据，为下游分层提供基础 “原材料”，保证数据覆盖商品、用户行为、订单等业务场景。

（二）DWD 层（数据仓库明细层）

1. 表结构与作用

dwd\_goods\_relation：基于 ODS 层用户行为数据，清洗并提炼商品关联行为关系，字段 shop\_id（店铺 ID）、main\_goods\_id（主商品 ID，用户先交互商品 ）、related\_goods\_id（关联商品 ID，用户后交互商品 ）、user\_id（用户 ID）、behavior\_type（行为类型）、behavior\_time（行为时间） ，以 dt 分区存储，用于清晰呈现用户与商品交互过程中商品间的关联链路，是商品关联分析的关键明细层数据。

2. 数据加载逻辑

利用窗口函数 LAG ，按 user\_id、shop\_id 分区并依 behavior\_time 排序，识别出用户在同一店铺、同一用户维度下先交互的 “主商品” 和后交互的 “关联商品”，同时对主商品为空等异常情况做处理（如标记为 无主商品 ），从 ODS 层 ods\_user\_behavior 清洗过滤出有效行为数据并加载到该层，为后续聚合分析提供关联关系基础。

（三）DWS 层（数据仓库汇总层）

1. 表结构与作用

dws\_goods\_drain\_hot：对 DWD 层数据进一步聚合分析，字段 shop\_id（店铺 ID）、goods\_id（商品 ID）、is\_drain（是否为引流品，1 是 0 否 ）、is\_hot（是否为热销品，1 是 0 否 ） ，以 dt 分区存储，用于标记商品在店铺维度下的引流、热销属性，为业务提供商品价值判断的中间分析结果。

2. 数据加载逻辑

按 shop\_id、goods\_id 分组，基于 DWD 层 dwd\_goods\_relation 数据，通过 COUNT 函数结合 CASE WHEN 语句，统计商品浏览、支付等行为次数，依据次数阈值（如浏览次数 > 100 标记为引流品，支付次数 > 50 标记为热销品 ）判断并标记商品是否为引流品或热销品，将聚合标记后的数据加载到该层，提炼出业务分析的关键属性。

（四）ADS 层（应用数据层）

1. 表结构与作用

ads\_goods\_linkage：关联多维度数据，输出直接支撑业务决策的数据，字段 shop\_id（店铺 ID）、drain\_goods\_id（引流品 ID）、drain\_goods\_name（引流品名称）、hot\_goods\_id（热销品 ID）、hot\_goods\_name（热销品名称）、co\_buy\_cnt（共同购买次数，关联强度 ） ，以 dt 分区存储，用于直观呈现引流品与热销品的关联关系及关联购买强度，直接服务于运营商品推荐、连带销售等策略制定。

2. 数据加载逻辑

关联 DWS 层 dws\_goods\_drain\_hot（获取商品引流、热销标记 ）、ODS 层 ods\_goods\_info（获取商品名称 ）、ods\_orders（获取订单关联购买信息 ）数据，通过多表 JOIN 操作（关联店铺、商品 ID 等条件 ），按 shop\_id、drain\_goods\_id 等分组，用 COUNT(\*) 统计共同购买次数，将关联整合后的数据加载到该层，形成直接可用的业务分析结果。