**理解仓库建模的两种模型：星型模型和雪花模型**

1. 星型模型和雪花模型的概念。
2. 星型模式 vs 雪花模型建模特点。
3. 什么情况下应用什么样的模型？
4. 建模其他事项。

## 星型模型和雪花模型的概念？

直接上图：







很明显，上图左侧为星型模式，右侧为雪花模型。详细概念如下：

1. **星型模式**：是一种多维的数据关系，它由一个事实表和一组维表组成。每个维表都有一个维作为主键，所有这些维的主键组合成事实表的主键。强调的是对维度进行预处理，将多个维度集合到一个事实表，形成一个宽表，
2. **雪花模型：**当有一个或多个维表没有直接连接到事实表上，而是通过其他维表连接到事实表上时，其图解就像多个雪花连接在一起，故称雪花模型。雪花模型是对星型模型的扩展。它对星型模型的维表进一步层次化，原有的各维表可能被扩展为小的事实表，形成一些局部的 "层次 " 区域，这些被分解的表都连接到主维度表而不是事实表。

## 星型模式 vs 雪花模型建模特点。

|  |  |  |
| --- | --- | --- |
| **比较项** | **星型模型** | **雪花模型** |
| **行数** | 多 | 少 |
| **可读性** | 容易 | 难 |
| **表个数** | 少 | 多 |
| **搜索维的时间** | 快 | 慢 |
| **冗余度** | 高 | 低 |
| **对事实表的影响** | 增加事实表的宽度 | 事实表字段较少，降低数据库存储负担。 |

**关于表格中冗余度和对事实表的影响说明：**星型模型会将所有的分析维度都作为事实表的一个直接维度，数据冗余较大的。但是，在一张表中表达出所有的类别属性，对数据库的存储空间较大的。在此基础上，对星型架构的基础上扩展雪花架构，一定程度上降低了分析查询的性能，提升了数据仓库的存储容量。

## 什么情况下应用什么样的模型？

星型模式架构主要有两点优势：

1. 提高查询的效率。由于数据的组织已经过预处理，主要数据都在庞大的事实表中，所以只要扫描事实表就可以进行查询，而不必进行多次连接。查询访问效率较高。同时，由于维表一般都很小，可以放在高速缓存中，与事实表作连接时其速度较快。便于用户理解。
2. 设计维护较为简单。

雪花模式架构主要有如下优势：

1. 正规化，数据冗余少。
2. 效率较低，有些数据需要连接才能获取。
3. 规范化操作较复杂，导致设计及后期维护复杂。

实际应用中，可以采取上述两种模型的混合体：如：中间层使用雪花结构以降低数据冗余度，数据集市部分采用星型以方便数据提取及和分析。

有时候规范化及存储空间考虑和效率会矛盾。一般会牺牲空间（规范化）来换取好的性能，把尽可能多的维度信息存在一张“大表”里面是最快的。通常会视情况而定，采取折中的策略。

另外，星型模型，需要更新维度成员时，必须连事实表同时更新。雪花型，有时只需要更新雪花维度中的一层即可，无需更改庞大的事实表。

例如：时间维度，年，季等属性一般星型模型。而像产品和产品的分类，可以考虑建立雪花模型。可以建立分类的查找表，对分析信息进行再分析。

## 建模其他事项。

工具：ERWIN 等建模工具。

范式：符合 3NF 的关系型数据模型。有时候数据仓库考虑性能等，会不规范。参看星型模型和雪花模型的概念。