

数仓建模有哪些方式？

可回答：1) 星型，雪花，星座模型；2) 建模常用理论；3) 数据建模的几种方式了解吗

问过的一些公司：阿里(2022.11)，阿里云(2022.10)，小米(2022.09)x2，众安保险(2022.07)，字节(2022.06)，快手(2021.09)，唯品会(2021.09)，海康威视(2021.09)，58同城(2021.09)，美团(2021.09)(2021.08)(2020.09)，字节社招(2021.03)(2020.09)，58同城(2020.09)，猿辅导(2020.08)，转转(2020.09)

参考答案：

一般回答前两种就行，详细部分参考“数仓建模常用模型吗？区别、优缺点？”

1、维度建模

维度建模按数据组织类型划分可分为星型模型、雪花模型、星座模型。

1) 星型模型

星型模型主要是维表和事实表，以事实表为中心，所有维度直接关联在事实表上，呈星型分布。

2) 雪花模型

雪花模型，在星型模型的基础上，维度表上又关联了其他维度表。这种模型维护成本高，性能方面也较差，所以一般不建议使用。尤其是基于hadoop体系构建数仓，减少join就是减少shuffle，性能差距会很大。

星型模型可以理解为，一个事实表关联多个维度表，雪花模型可以理解为一个事实表关联多个维度表，维度表再关联维度表。

3) 星座模型

数据仓库由多个主题构成，包含多个事实表，而维表是公共的，可以共享（例如两张事实表共用一些维度表时，就叫做星型模型），这种模式可以看做星型模式的汇集，因而称作星系模式或者事实星座模式。

2、范式建模（ER建模）

从全企业的高度设计一个3NF模型，用实体加关系描述的数据模型描述企业业务架构，在范式理论上符合3NF。此建模方法，对建模人员的能力要求非常高。

特点：设计思路自上而下，适合上游基础数据存储，同一份数据只存储一份，没有数据冗余，方便解耦，易维护，缺点是开发周期一般比较长，维护成本高。

3、Data Vault模型

DataVault由Hub（关键核心业务实体）、Link（关系）、Satellite（实体属性）三部分组成，是Dan Linstedt发起创建的一种模型方法论，它是在ER关系模型上的衍生，同时设计的出发点也是为了实现数据的整合，并非为数据决策分析直接使用。

4、Anchor模型

Anchor模型是对Data Vault模型做了进一步规范化处理，它是一个高度可扩展的模型，所有的扩展只是添加而不是修改，因此它将模型规范到6NF，基本变成了K-V结构模型。企业很少使用。



蓦然 送你一张星球优惠券

「旧时光大数据」

立减

¥ 30

新人立减券

2024/12/31 12:00 后失效

知识星球

长按扫码领取优惠



公众号：旧时光大数据