1. **大数据相关**

**数据仓库建模理论**

-数据建模的目标

|  |  |
| --- | --- |
| 访问性能 | 能够快速查询所需的数据，减少数据I/O |
| 数据成本 | 减少不必要的数据冗余，实现计算结果数据复用，降低大数据系统中的存储成本和计算成本 |
| 使用效率 | 改善用户应用体验，提高使用数据的效率 |
| 数据质量 | 改善数据统计口径的不一致性，减少数据计算错误的可能性，提供高质量的、一致的数据访问平台 |

因此，数仓建模需要通过更好的组织、存储数据，以便在性能、成本、效率和数据质量之间找到最佳平衡点。

-关系模式范式

关系型数据库设计时，遵照一定的规范要求，目的在于降低数据的冗余性和数据的一致性，目前业界范式有：

|  |  |
| --- | --- |
| **第一范式（1NF）** | 域都是原子性的，即数据库表的每一列都是不可分割的原子数据项 |
| **第二范式（2NF）** | 在1NF的基础上，实体的属性完全依赖于主关键字，不能存在仅依赖主关键字一部分的属性，也就是不存在局部依赖 |
| **第三范式（3NF）** | 在2NF的基础上，任何非主属性不依赖于其它非主属性，也就是不存在传递依赖 |

-四种建模方式

|  |  |
| --- | --- |
| ER实体模型 |  |
| 维度建模 |  |

**数据仓库指标体系**

1. **Java相关**