

关系数据库的理论是大数据领域建模的基础,有兴趣的读者可以参考相关的经典数据库理论书籍,如《数据库系统概念》。

8.3 从 OLTP 和 OLAP 系统的区别看模型方法论的选择

OLTP 系统通常面向的主要数据操作是随机读写,主要采用满足 3NF 的实体关系模型存储数据,从而在事务处理中解决数据的冗余和一致性问题;而 OLAP 系统面向的主要数据操作是批量读写,事务处理中的一致性不是 OLAP 所关注的,其主要关注数据的整合,以及在一次性的复杂大数据查询和处理中的性能,因此它需要采用一些不同的数据建模方法。

8.4 典型的数据仓库建模方法论

8.4.1 ER 模型

数据仓库之父 Bill Inmon 提出的建模方法是从全企业的高度设计一个 3NF 模型,用实体关系 (Entity Relationship, ER) 模型描述企业业务,在范式理论上符合 3NF。数据仓库中的 3NF 与 OLTP 系统中的 3NF 的区别在于,它是站在企业角度面向主题的抽象,而不是针对某个具体业务流程的实体对象关系的抽象。其具有以下几个特点:

- 需要全面了解企业业务和数据。
- 实施周期非常长。
- 对建模人员的能力要求非常高。

采用 ER 模型建设数据仓库模型的出发点是整合数据,将各个系统