
《数据管理及数仓建模》

-段方

某世界 100 强企业大数据/AI 总设计

师 教授 北京大学博士后

1 概述

1.1 企业的数字化转型

1.2 数字化转型的核心——大数据

1.3 数据仓库基础

1.4 数据湖概念和架构

1.5 大数据的核心——数据建模

1.6 数据建模概览

1.7 如何表述数据之间的关系？

1.8 数据管理的相关规范

1.8.1 数据标准

1.8.2 业务标准

1.8.3 管理标准

1.9 【例】附件——某企业数据治理案例介绍

2 数据建模的技术、流程和注意点

- 2.1 什么是数据模型？
- 2.2 数据仓库的数据模型
- 2.3 大数据系统不用数据模型吗？
- 2.4 数据建模的基础——第三范式
- 2.5 数据建模的流程
- 2.6 数据模型是数据治理的基础
- 2.7 数据接口的标准是数据模型的基础
- 2.8 数据字典概念基础
- 2.9 数据建模的关键点
- 2.10 【例】附件——某企业数据字典举例

3 数据整合、处理和展示

- 3.1 数据内容
 - 3.1.1 客户数据
 - 3.1.2 产品数据
 - 3.1.3 营销数据
 - 3.1.4 员工数据
 - 3.1.4 服务数据
 - 3.1.5

3.2 数据标准的颗粒度

3.2.1 详细数据

3.2.2 汇总数据

3.2.3 衍生数据

3.3 数据如何收集

3.3.1 从内部各个 IT 系统获取

3.3.2 外部数据获取方法

3.3.3 批量获取/实时获取

3.4 数据清洗过程

3.4.1 数据抽取

3.4.2 数据转换

3.4.3 数据加载

3.5 跨平台数据接口标准

3.5.1 数据接口的定义

3.5.2 数据接口的管理职责

3.5.3 接口设计的原则

3.6 逻辑模型设计及实例

3.6.1 原因

3.6.2 逻辑模型设计的原则

3.6.3 设计实例

3.7 物理模型设计及实例

3.7.1 原因

3.7.2 物理模型设计原则

3.7.3 设计实例

3.8 数据字典及实例

3.8.1 概念

3.8.2 范围和内容

3.8.3 设计实例

3.9 【案例】附件-某企业的数据资产治理案例

4 数据仓库和数据湖技术及案例

4.1 数据仓库的引出

4.2 数据仓库的流派之争

4.3 非结构化数据带来的挑战

4.4 从数据仓库到数据湖

4.5 数据仓库底层技术的演进

4.6 数据湖的底层技术变化

4.7 某企业的数据建模举例

4.8 【案例】附件-某企业数据仓库/数据湖案例

5 建模方法论

5.1 数据模型基本方法

- 5.2 数据模型与行业应用的关系
- 5.3 第三范式
- 5.4 实体-关系模型
- 5.5 建模的三阶段
 - 5.5.1 概念模型
 - 5.5.2 逻辑模型
 - 5.5.3 物理模型
- 5.6 建立模型的关键问题
 - 5.6.1 逻辑模型与业务的关系
 - 5.6.2 逻辑模型与需求的关系
 - 5.6.3 逻辑模型与业务系统的关系
- 5.7 【例】附件——某企业数据仓库建模举例

6 建模工具

- 6.1 Erwin 建模工具介绍
- 6.2 主要特点
- 6.3 主要功能
- 6.4 Erwin 支持的各种接口
- 6.5 【例】附件——Erwin 工具介绍

7 数据仓库逻辑模型（示例）

- 7.1 数据仓库模型概述
- 7.2 某企业数据模型概述及举例
- 7.3 参与人主题域
- 7.4 服务主题域
- 7.5 资源主题域
- 7.6 营销主题域
- 7.7 事件主题域
- 7.8 账务主题域
- 7.9 财务主题域
- 7.10 【例】附件——某企业的数据仓库建模举例

8 汇总层数据模型

- 8.1 什么是汇总层数据模型？
- 8.2 某企业汇总层数据模型及案例
- 8.3 详单汇总
- 8.4 账务汇总
- 8.5 个人用户汇总
- 8.6 集团客户汇总
- 8.7 产品信息汇总等
- 8.8 【例】附件——某企业的数据质量管理介绍

9 数据接口设计

9.1 数据接口设计要求

9.2 数据接口分类

9.3 数据接口实现机制

9.4 接口双方职责

9.5 数据转换规则

9.6 消息级接口规则

9.7 【例】附件——某企业的接口规范举例

10 大数据系统数据模型

10.1 HDOOP 系统中不需要数据模型了吗？

10.2 Hadoop 的生态及组件

10.3 对内服务和对外服务的关系

10.4 大数据系统的物理模型

10.5 参照表的设计

10.6 实体命名规则

10.7 实体定义

10.8 数据湖的数据建模

10.9 【例】附件——某企业大数据系统建模举例

11 总结
