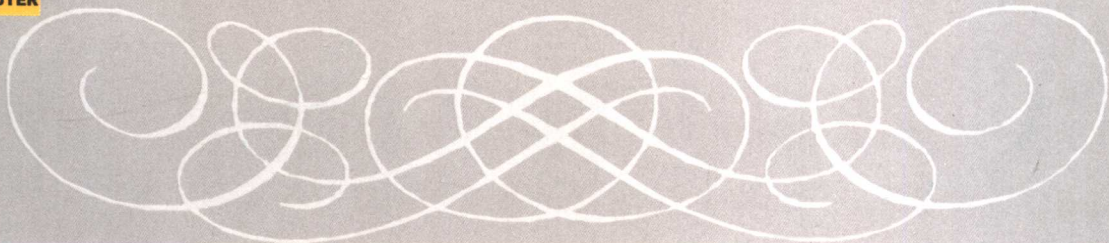


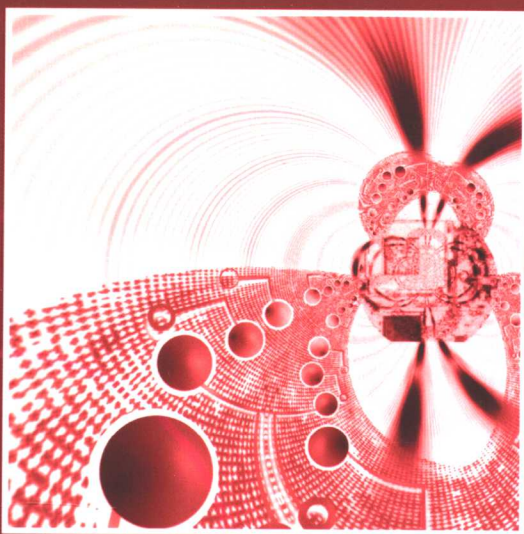


普通高等教育“十一五”计算机类规划教材



# 数据库 原理及应用

胡孔法 编著



机械工业出版社  
CHINA MACHINE PRESS



TP311.13/313

2008

普通高等教育“十一五”计算机类规划教材

# 数据库原理及应用

胡孔法 编著

机械工业出版社

本书将数据库基本原理、方法和应用技术相结合,以关系数据库系统为核心,在完整地论述数据库系统的基本概念、基本原理的基础上,对 Microsoft SQL Server 等主流的数据库管理系统、Visio 和 Power Designer、ASP.NET 和 ADO.NET 等数据库设计与软件开发工具进行讲解。主要内容包括数据模型、数据库系统的体系结构、关系数据库方法、关系数据库的结构化查询语言、关系模式的规范化理论、数据库设计、数据库保护、Microsoft SQL Server 2000、ASP.NET 和 ADO.NET 数据库开发技术、分布式数据库系统、数据仓库与数据挖掘、XML 数据库等。

本书可作为高等院校计算机专业、信息管理与信息系统以及相关专业的本科生的教材,还可作为广大软件设计与开发的工程技术人员等信息领域工作的科技人员的参考书。为方便教师教学,本书配有教学课件,欢迎选用本书作为教材的老师索取,索取邮箱:[llm7785@sina.com](mailto:llm7785@sina.com)。

### 图书在版编目(CIP)数据

数据库原理及应用/胡孔法编著. —北京:机械工业出版社,  
2008.1

普通高等教育“十一五”计算机类规划教材  
ISBN 978-7-111-22780-9

I. 数… II. 胡… III. 数据库系统—高等学校—教材  
IV. TP311.13

中国版本图书馆 CIP 数据核字(2007)第 174907 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

责任编辑:刘丽敏 责任校对:陈立辉

封面设计:张静 责任印制:洪汉军

北京铭成印刷有限公司印刷

2008 年 1 月第 1 版第 1 次印刷

184mm×260mm·22.25 印张·549 千字

标准书号:ISBN 978-7-111-22780-9

定价:32.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换  
销售服务热线电话:(010)68326294

购书热线电话:(010)88379639、88379641、88379643

编辑热线电话:(010)88379726

封面无防伪标均为盗版

# 前 言

在 20 世纪 60 年代末, 数据库技术是作为数据处理中的一门新技术发展起来的。目前, 数据库技术已是计算机软件领域的一个重要分支, 形成了较为完整的理论体系和实用技术。

随着数据库系统的推广, 计算机应用已深入人类社会的各个领域, 如当前的管理信息系统(MIS)、企业资源规划(ERP)、计算机集成制造系统(CIMS)、地理信息系统(GIS)、决策支持系统(DDS)等都是以数据库技术为基础。目前, 我国实施的国家信息化、“金”字工程、数字城市等都是以数据库为基础的大型计算机系统。数据库的建设规模和性能、数据库信息量的大小和使用水平已成为衡量一个国家信息化程度的重要标志。所以, 我国高等院校从 20 世纪 80 年代开始就把“数据库原理及应用”作为计算机专业的主要课程之一。目前, “数据库原理及应用”课程是高等院校计算机专业、信息管理与信息系统以及相关专业的一个重要专业基础课程。

本书将数据库基本原理、方法和应用技术相结合, 以关系数据库系统为核心, 在完整地论述数据库系统的基本概念、基本原理的基础上, 对 Microsoft SQL Server 等主流的数据库管理系统、Visio 和 Power Designer、ASP.NET 和 ADO.NET 等数据库设计与软件开发工具进行讲解, 并对分布式数据库、数据仓库与数据挖掘、XML 数据库等数据库新技术进行介绍, 使高等院校计算机专业、信息管理与信息系统以及相关专业的本科生以及广大软件设计与开发的工程技术人员, 能熟练地使用现有的数据库管理系统和软件设计与开发工具, 进行数据库结构的设计和数据库应用系统开发。

本书共分 13 章。第 1 章数据库系统概述, 讨论数据管理和数据库系统的发展过程。第 2 章数据模型, 着重介绍概念模型、层次模型、网状模型、关系模型、面向对象模型等数据库系统的数据模型的基本概念和设计方法。第 3 章数据库系统的体系结构, 主要介绍数据库系统的三级模式结构、DBS 的组成和全局结构、DBMS 工作模式、DBMS 系统结构等。第 4 章关系数据库方法, 介绍包括关系数据库基本概念、关系运算、关系演算、关系查询优化问题。第 5 章关系数据库的结构化查询语言, 主要介绍包括 SQL 概述、SQL 数据定义、SQL 数据查询、SQL 聚集函数、SQL 数据更新、SQL 中的视图。第 6 章主要介绍数据依赖、范式和模式设计方法等关系模式规范化理论。第 7 章主要介绍数据库设计的方法、步骤、辅助设计工具等。第 8 章介绍完整性控制、安全性控制、数据库恢复和并发控制等数据库保护措施。第 9 章介绍当前流行的关系数据库系统 Microsoft SQL Server 的基本知识, 以 Microsoft SQL Server 2000 为背景, 介绍数据库系统设计方法和 SQL Server 的高级应用技术。第 10 章主要介绍 ASP.NET 和 ADO.NET 基础知识、ASP.NET 连接数据库方法、ADO.NET 读取和操作数据库数据等基于 .NET 数据库开发技术。第 11 章主要介绍分布式数据库系统的基本概念、模式结构、数据分布、查询处理等。第 12 章主要介绍数据仓库和数据挖掘的基本概念、数据仓库设计方法与实现过程、数据挖掘的主要技术和过程。第 13 章介绍 XML 的基本知识, 以及 XML 文档、XML 模式、XPath 查询语言、XQuery 查询语言, 最后给出 XML 在数据管理方面的应用实例。

本书是作者根据十几年的软件开发与设计经验，结合在读硕士和博士期间在数据库及信息系统方面所从事的研究和多年的教学经验编写而成的。对于曾在这方面提供过帮助的陈峻教授、李斌教授、达庆利教授、董逸生教授以及纪北辉、刘海东、顾欣、张长海、陈竹西等研究生表示衷心的感谢。

本书得到国家自然科学基金项目(60773103、60673060)、江苏省“青蓝工程”优秀青年骨干教师人才计划、扬州大学出版基金资助等项目的支持。

本书可作为计算机及相关的本科生教材，也可供研究生、广大软件设计和开发人员参考。书中错漏之处，敬请指正。

胡孔法

# 目 录

## 前言

## 第 1 章 数据库系统概述 ..... 1

### 1.1 数据库技术的产生与发展 ..... 1

#### 1.1.1 数据管理的发展 ..... 1

#### 1.1.2 数据和数据管理技术 ..... 2

#### 1.1.3 数据库系统的三个发展阶段 ..... 4

### 1.2 数据库系统基本概念 ..... 5

#### 1.2.1 数据库 ..... 5

#### 1.2.2 数据库系统 ..... 5

### 1.3 数据库系统的特点 ..... 6

## 第 2 章 数据模型 ..... 9

### 2.1 数据模型概述 ..... 9

#### 2.1.1 数据模型的基本组成 ..... 9

#### 2.1.2 数据模型的发展 ..... 10

### 2.2 E-R 数据模型 ..... 10

#### 2.2.1 基本概念 ..... 10

#### 2.2.2 E-R 图 ..... 12

#### 2.2.3 扩充 E-R 数据模型 ..... 13

### 2.3 层次数据模型 ..... 15

#### 2.3.1 基本概念和结构 ..... 16

#### 2.3.2 数据操作 ..... 17

#### 2.3.3 数据约束 ..... 18

#### 2.3.4 层次数据模型的优缺点 ..... 20

### 2.4 网状数据模型 ..... 20

#### 2.4.1 基本概念和结构 ..... 20

#### 2.4.2 数据操作 ..... 22

#### 2.4.3 数据约束 ..... 23

#### 2.4.4 网状数据模型的优缺点 ..... 23

### 2.5 关系数据模型 ..... 24

#### 2.5.1 基本概念 ..... 24

#### 2.5.2 关系数据模型的数据结构 ..... 24

#### 2.5.3 数据操作 ..... 26

#### 2.5.4 数据约束 ..... 26

#### 2.5.5 关系数据模型的优缺点 ..... 27

### 2.6 面向对象数据模型 ..... 27

#### 2.6.1 对象和对象标识符 ..... 28

#### 2.6.2 属性和方法 ..... 28

#### 2.6.3 封装和消息传递 ..... 28

#### 2.6.4 类和实例 ..... 29

#### 2.6.5 类层次结构和继承 ..... 29

#### 2.6.6 持久性和版本 ..... 30

#### 2.6.7 多态、重载、重定义 与动态联编 ..... 30

#### 2.6.8 面向对象数据模型与关系数据 模型比较 ..... 31

## 第 3 章 数据库系统的体系结构 ..... 32

### 3.1 数据库的体系结构 ..... 32

#### 3.1.1 三级模式结构 ..... 32

#### 3.1.2 两级映像和两级数据独立性 ..... 34

### 3.2 数据库系统 ..... 35

#### 3.2.1 DBS 的组成 ..... 35

#### 3.2.2 DBS 的全局结构 ..... 37

#### 3.2.3 DBS 结构的分类 ..... 38

### 3.3 数据库管理系统 ..... 40

#### 3.3.1 DBMS 的工作模式 ..... 40

#### 3.3.2 DBMS 的主要功能 ..... 41

### 3.4 DBMS 的系统结构 ..... 42

#### 3.4.1 DBMS 的进程结构和 多线索结构 ..... 43

#### 3.4.2 DBMS 的组成 ..... 43

#### 3.4.3 DBMS 的层次结构 ..... 45

### 3.5 几种典型结构的 DBMS ..... 46

#### 3.5.1 Client/Server 结构的 DBMS ..... 46

#### 3.5.2 基于 Web 的数据库系统 ..... 48

## 第 4 章 关系数据库方法 ..... 53

### 4.1 关系数据库的基本概念 ..... 53

#### 4.1.1 关系的形式化定义 ..... 53

#### 4.1.2 关系模式、关系子模式和



存储模式 .....	55	5.6.2 视图查询 .....	113
4.1.3 关系模型的完整性规则 .....	57	5.6.3 视图更新 .....	113
4.1.4 关系数据库模式 .....	59	5.6.4 视图删除 .....	114
4.2 关系代数 .....	60	5.6.5 视图的作用 .....	115
4.2.1 关系代数的五个基本操作 .....	60	5.7 嵌入式 SQL .....	115
4.2.2 关系代数的组合操作 .....	63	5.7.1 程序设计环境下的 SQL .....	115
4.2.3 关系代数表达式应用举例 .....	68	5.7.2 单行选择语句 .....	118
4.3 关系演算 .....	69	5.7.3 游标 .....	119
4.3.1 元组关系演算 .....	69	<b>第6章 关系模式的规范化理论</b> .....	122
4.3.2 域关系演算 .....	70	6.1 关系模式设计中的问题 .....	122
4.4 关系查询优化 .....	71	6.2 函数依赖 .....	123
4.4.1 查询优化的一般策略 .....	72	6.2.1 函数依赖定义 .....	123
4.4.2 关系代数优化 .....	72	6.2.2 关键字和超关键字 .....	124
4.4.3 基于存取路径的规则优化 .....	77	6.3 数据依赖的公理系统 .....	125
<b>第5章 关系数据库的结构化</b>		6.3.1 函数依赖的逻辑蕴涵 .....	125
<b>查询语言</b> .....	85	6.3.2 Armstrong 公理系统 .....	126
5.1 SQL 概述 .....	85	6.3.3 函数依赖集的等价和覆盖 .....	129
5.2 SQL 的数据定义语言 .....	87	6.4 关系模式的分解及其问题 .....	131
5.2.1 数据类型 .....	87	6.4.1 分解的无损连接性 .....	131
5.2.2 数据库模式的定义 .....	87	6.4.2 保持函数依赖性 .....	137
5.2.3 基本表的定义 .....	89	6.5 关系模式的规范化 .....	139
5.2.4 基本表的修改和删除 .....	92	6.5.1 范式 .....	140
5.2.5 域 .....	93	6.5.2 模式分解的算法 .....	142
5.2.6 索引的建立和删除 .....	93	6.6 多值函数依赖与4NF .....	145
5.3 SQL 数据查询 .....	94	6.6.1 BCNF 关系模式存在的问题 .....	146
5.3.1 ALPHA 关系演算语言 .....	94	6.6.2 多值函数依赖 .....	146
5.3.2 SQL 查询语句格式 .....	96	6.6.3 4NF .....	148
5.3.3 简单查询 .....	97	6.7 连接依赖和5NF .....	149
5.3.4 复杂查询 .....	99	<b>第7章 数据库设计</b> .....	150
5.4 SQL 聚集函数 .....	106	7.1 数据库设计概述 .....	150
5.4.1 聚集函数的运算符 .....	106	7.1.1 数据库系统生存期 .....	150
5.4.2 数据分组 .....	107	7.1.2 数据库设计方法 .....	151
5.4.3 数据排序 .....	108	7.1.3 数据库设计的基本过程 .....	153
5.5 SQL 数据更新 .....	108	7.2 规划 .....	154
5.5.1 元组插入 .....	108	7.3 需求分析 .....	155
5.5.2 元组删除 .....	110	7.3.1 需求描述与分析 .....	155
5.5.3 元组修改 .....	110	7.3.2 需求分析阶段的输入和输出 .....	155
5.6 SQL 中的视图 .....	111	7.3.3 需求分析的步骤 .....	156
5.6.1 视图定义 .....	112	7.3.4 数据字典 .....	160



7.4 概念设计 .....	162	<b>第9章 Microsoft SQL Server</b>	
7.4.1 概念设计的必要性 .....	162	<b>2000</b> .....	207
7.4.2 概念设计的主要步骤 .....	163	9.1 SQL Server 简介 .....	207
7.4.3 采用 E-R 模型方法的数据库		9.1.1 SQL Server 2000 的特点 .....	207
概念设计 .....	164	9.1.2 SQL Server 2000 环境介绍 .....	208
7.5 数据库逻辑结构设计及优化 .....	173	9.1.3 SQL Server 2000 的工具介绍 .....	209
7.5.1 逻辑设计环境 .....	173	9.2 数据库的创建、修改和删除 .....	212
7.5.2 E-R 模型向关系模型的转换 .....	174	9.2.1 创建 SQL Server 数据库 .....	212
7.5.3 用关系规范化理论对关系数据		9.2.2 修改数据库 .....	215
模型进行优化 .....	180	9.2.3 删除数据库 .....	215
7.6 数据库的物理设计 .....	182	9.3 表和主键的创建 .....	217
7.7 数据库的实现 .....	182	9.3.1 表的创建 .....	217
7.8 数据库的运行与维护 .....	183	9.3.2 主键的创建 .....	217
7.9 Power Designer 辅助设计工具 .....	183	9.3.3 用 SQL 查询分析器的命令	
7.9.1 绘制 E-R 图 .....	183	方式来创建表 .....	220
7.9.2 后台生成 SQL 数据库及数据表 .....	186	9.4 数据的插入、修改、删除和查询 .....	220
<b>第8章 数据库保护</b> .....	190	9.4.1 数据的插入 .....	220
8.1 事务 .....	190	9.4.2 数据的修改 .....	222
8.1.1 事务的定义 .....	190	9.4.3 数据的删除 .....	222
8.1.2 事务的 ACID 准则 .....	190	9.4.4 数据的查询 .....	223
8.2 数据库完整性 .....	191	9.5 数据库的备份和恢复 .....	223
8.2.1 完整性子系统和完整性规则 .....	191	9.5.1 数据库的备份 .....	223
8.2.2 SQL 中的完整性约束 .....	192	9.5.2 数据库的恢复 .....	226
8.2.3 SQL3 的触发器 .....	193	9.6 高级应用技术 .....	228
8.3 数据库安全性 .....	194	9.6.1 存储过程 .....	228
8.3.1 数据库安全性级别 .....	194	9.6.2 触发器 .....	232
8.3.2 数据访问权限 .....	194	<b>第10章 ASP.NET 和 ADO.NET</b>	
8.3.3 SQL 中的安全性机制 .....	195	<b>数据库开发技术</b> .....	238
8.4 数据库恢复技术 .....	196	10.1 ASP.NET 基础知识 .....	238
8.4.1 恢复的定义、原则和方法 .....	196	10.1.1 Visual Studio.NET 的特色 .....	238
8.4.2 故障恢复方法 .....	197	10.1.2 .NET Framework .....	239
8.4.3 具有检查点的恢复技术 .....	197	10.1.3 ASP.NET 开发环境介绍 .....	240
8.5 并发控制 .....	198	10.2 ADO.NET 介绍 .....	240
8.5.1 并发所引起的问题 .....	198	10.2.1 ADO.NET 与数据管理简介 .....	240
8.5.2 封锁 .....	200	10.2.2 ADO.NET 命名空间 .....	241
8.5.3 活锁和死锁 .....	202	10.3 连接数据库 .....	241
8.5.4 死锁的防止、检测和处理 .....	203	10.3.1 连接 Microsoft SQL Server	
8.5.5 并发调度的可串行化 .....	204	数据库 .....	241
8.5.6 两段封锁协议 .....	206	10.3.2 连接 Microsoft Access	



数据库 .....	245	<b>第 12 章 数据仓库与数据挖掘</b> .....	302
10.4 读取和操作数据 .....	246	12.1 概述 .....	302
10.5 数据集 .....	250	12.2 数据仓库概念 .....	303
10.5.1 使用 DataAdapter 类 .....	251	12.2.1 数据仓库的定义 .....	304
10.5.2 使用 DataTables、DataColumns 和 DataRows .....	251	12.2.2 数据仓库的结构 .....	305
10.6 DataReader 类 .....	253	12.3 数据仓库的设计与实现 .....	311
10.7 DataGrid 和 GridView 控件 .....	256	12.3.1 数据仓库的设计原则 .....	311
10.8 DataList 控件 .....	259	12.3.2 数据仓库的三级数据模型 .....	312
<b>第 11 章 分布式数据库系统</b> .....	272	12.3.3 数据仓库的设计步骤 .....	313
11.1 概述 .....	272	12.3.4 数据仓库的实现结构 .....	316
11.1.1 分布式数据库的发展需求 .....	272	12.4 数据挖掘 .....	316
11.1.2 分布式数据库系统的定义 .....	274	12.4.1 数据挖掘的定义 .....	316
11.1.3 分布式数据库系统的特点 .....	275	12.4.2 数据挖掘技术的应用研究 现状 .....	317
11.1.4 分布式数据库系统的环境 .....	278	12.4.3 数据挖掘主要技术 .....	319
11.1.5 分布式数据库系统的分类 .....	279	12.4.4 数据挖掘的过程 .....	321
11.2 分布式数据库系统体系结构 .....	282	12.4.5 DM 与 OLAP .....	322
11.2.1 分布式数据库系统的模式 结构 .....	282	<b>第 13 章 XML 开发技术</b> .....	324
11.2.2 分布式数据库管理系统 .....	285	13.1 引言 .....	324
11.2.3 分布式数据库系统结构 .....	287	13.2 XML 文档 .....	325
11.3 数据分布 .....	289	13.3 XML 文档类型定义 .....	328
11.3.1 数据分布的目的 .....	289	13.3.1 元素定义 .....	329
11.3.2 数据分布的概念 .....	291	13.3.2 属性定义 .....	330
11.3.3 数据分布的方式 .....	292	13.4 XML 模式 .....	331
11.3.4 数据分片方式与原则 .....	293	13.5 XPath 查询语言 .....	336
11.3.5 数据分片操作 .....	294	13.6 XQuery 查询语言 .....	337
11.3.6 数据分布带来的问题 .....	297	13.7 XML 的应用 .....	338
11.4 查询优化 .....	298	13.7.1 基于 XML 的数据交换与异构 数据集成 .....	338
11.4.1 查询树的变换 .....	298	13.7.2 XML 索引与查询处理 .....	339
11.4.2 副本选择 .....	299	13.7.3 XML 文档聚类 .....	342
11.4.3 查询分解 .....	300	13.7.4 XML 流处理 .....	344
11.4.4 半连接在分布连接中的应用 .....	301	<b>参考文献</b> .....	346

# 第 1 章 数据库系统概述

在 20 世纪 60 年代末，数据库技术是作为数据处理中的一门新技术发展起来的。目前，数据库技术已是计算机软件领域的一个重要分支，形成了较为完整的理论体系和实用技术。本章首先介绍数据库技术的发展过程，然后介绍数据库中的基本概念，以使读者对数据库的概貌有所了解。

## 1.1 数据库技术的产生与发展

### 1.1.1 数据管理的发展

数据库系统的萌芽出现于 20 世纪 60 年代。计算机开始广泛地应用于数据管理，对数据的共享提出了越来越高的要求。传统的文件系统已经不能满足人们的需要，能够统一管理和共享数据的数据库管理系统 (DBMS) 应运而生。数据模型是数据库系统的核心和基础，各种 DBMS 软件都是基于某种数据模型的。所以通常也按照数据模型的特点将传统数据库系统分成网状数据库、层次数据库和关系数据库三类。

世界上第一个 DBMS 应当是美国通用电气公司 Bachman 等人在 1964 年开发成功的 IDS (Integrated Data Store)。IDS 奠定了网状数据库的基础，并得到了广泛的应用。IDS 具有数据模式和日志的特征，但它只能在 GE 主机上运行，并且数据库只有一个文件，数据库所有的表必须通过手工编码来生成。

最著名、最典型的层次数据库系统是 IBM 公司在 20 世纪 60 年末开发的 IMS (Information Management System)，IMS 是一种适合其主机的层次数据库，这是 IBM 公司研制的最早的大型数据库系统程序产品。

网状数据库和层次数据库已经很好地解决了数据的集中和共享问题，但是在数据独立性和抽象级别上仍有很大欠缺。用户在对这两种数据库进行存取时，仍然需要明确数据的存储结构，指出存取路径，而后来出现的关系数据库则较好地解决了这些问题。

1970 年，IBM 的研究员 E. F. Codd 博士在刊物《Communication of the ACM》上发表了一篇名为“A Relational Model of Data for Large Shared Data Banks”的论文，提出了关系模型的概念，奠定了关系模型的理论基础。尽管在这之前的 1968 年 Childs 已经提出了面向集合的模型，然而这篇论文被普遍认为是数据库系统历史上具有划时代意义的里程碑。后来 Codd 又陆续发表多篇文章，论述了范式理论和衡量关系系统的 12 条标准，用数学理论奠定了关系数据库的基础。关系模型有严格的数学基础，抽象级别比较高，而且简单清晰，便于理解和使用。但是当时也有人认为关系模型是理想化的数据模型，用来实现 DBMS 是不现实的，尤其担心关系数据库的性能难以接受，更有人视其为当时正在进行的网状数据库规范化工作的严重威胁。为了促进对问题的理解，1974 年 ACM 牵头组织了一次研讨会，会上开展了一场分别以 Codd 和 Bachman 为首的支持和反对关系数据库两派



之间的辩论。这次著名的辩论推动了关系数据库的发展，不少工业和学术单位投入到关系 DBMS 原型的研制，使其最终成为现代数据库产品的主流。在众多的关系 DBMS 原型中，功能最全面、技术上最有代表性的是 IBM 公司的 System R 和加州大学伯克利分校的 INGRES。这两个原型系统全面地提供了比较成熟的关系 DBMS 技术，为研制商品化的关系 DBMS 提供了技术上的准备。IBM 公司在 System R 的基础上先后推出了 SQL/DS 和 DB2 两个商品化的关系 DBMS。关系型数据库系统以关系代数为坚实的理论基础，经过几十年的发展和实际应用，技术越来越成熟和完善。其代表产品有 Oracle、Sybase、Informix、MS SQL Server 等。

20 世纪 80 年代是关系数据库的全盛时代。随着计算机的广泛应用，新的应用又提出新的要求。人们开始发现关系数据库的许多限制和不足，这又推动了数据库技术的新一轮的研究，一是改造和扩充关系数据库，以适应新的应用需求；二是改用新的数据模型，例如面向对象数据模型等，研制新型的数据库。目前，这两方面都取得了很大进展。

数据管理与数据处理一样，都是计算机系统的最基本的支撑技术。尽管计算机科学技术经历了飞速的发展，但数据管理的这一地位没有变化。可以预期，数据管理将作为计算机科学技术的一个重要分支一直发展下去，社会愈信息化，对数据管理的要求也愈高。推动这门学科发展的动力是计算机应用，发展这门学科的基础是新的硬、软件技术。数据库是现阶段数据管理的主要形式。

### 1.1.2 数据和数据管理技术

#### 1. 数据

为了了解世界、研究世界和交流情况，人们需要描述事物。人们常常抽象那些感兴趣事物的特征或属性来对事物进行描述。

例如：一个大学生可用如下的记录来描述：<李定,021231142,男,1984,江苏,计算机系,2002>。表示李定是位大学生，学号是 021231142，性别是男性，1984 年出生，江苏人，2002 年考入计算机系。这种对事物描述的符号记录称为数据。

(1) 数据的含义 数据是载荷信息的媒体，是对现实世界中客观事物的符号表示，它可以是数值数据，也可以是非数值数据，如声音、图像等，它能输入计算机，并能为其处理的符号序列。

(2) 数据与信息的联系与区别 数据是数据库系统研究和处理的对象。数据与信息是分不开的，它们既有联系又有区别。

所谓数据，通常指用符号记录下来的、可以识别的信息。

信息与数据之间存在着固有的联系，数据是信息的符号表示或称为载体；信息则是数据的内涵，是对数据语义的解释。

#### 2. 数据管理技术

数据管理是指数据的收集、整理、分类、组织、存储、查询、检索、维护等操作，这部分操作是数据处理业务的基本环节，而且是任何数据处理业务中必不可少的共有部分。

数据处理是指从某些已知的数据出发，推导加工出一些新的数据，这些新的数据又表示了新的信息。

数据处理是与数据管理相联系的，数据管理技术的优劣，将直接影响数据处理的效率。

### 3. 数据管理技术的发展

计算机数据处理的速度和规模是人工方式或机械方式所无可比拟的，随着数据处理量的增长，产生了数据管理技术。数据管理技术的发展，与计算机硬件、系统软件及计算机应用的范围有着密切的联系。数据管理技术的发展经历了人工管理阶段、文件管理阶段和数据库系统阶段三个发展阶段。

(1) 人工管理阶段 20世纪50年代中期以前，计算机主要用于科学计算，数据管理处于人工管理阶段，数据处理的方式基本上是批处理。

当时计算机主要用于科学计算，数据量小、结构简单，如高阶方程、曲线拟和等。外存主要为顺序存取设备，如磁带、卡片、纸带，没有磁盘等直接存取设备。没有操作系统，没有数据管理软件，用户用机器指令编码，通过纸带机输入程序和数据，程序运行完毕后，由用户取走纸带和运算结果，再让下一用户上机操作。人工管理阶段主要特点是没有专用的软件对数据进行管理，由应用程序管理数据；数据面向应用程序，即一组数据对应一个应用程序，数据不能共享；应用程序完全依赖于数据，数据与程序没有独立性；数据不保存在计算机内，数据一般也不需要长期保存。

(2) 文件管理阶段 20世纪50年代后期至60年代中期，数据管理进入文件系统阶段。这个阶段是将数据组织成若干个相互独立的文件，用户通过操作系统对文件进行打开、读写、关闭等操作。

这时计算机不但用于科学计算，还用于信息管理。外部存储器已有磁盘、磁鼓等直接存取设备。有了专门管理数据的软件，即文件系统，用于文件存储空间的管理、目录管理、文件读写管理、文件保护、向用户提供操作接口等。

文件管理阶段的主要特点是数据以“文件”形式可长期保存在外部存储器的磁盘上；数据的逻辑结构与物理结构有了区别，但比较简单；文件组织已多样化，有索引文件、链接文件和直接存取文件等；数据不再属于某个特定的程序，可以重复使用，即数据面向应用；对数据的操作以记录为单位。

随着数据管理规模的扩大，数据量急剧增加，文件系统显露出数据存在大量冗余(redundancy)，造成了数据不一致(inconsistency)，数据之间的联系也比较弱(poor data relationship)等缺陷。

(3) 数据库系统阶段 20世纪60年代后期以来，为了克服文件系统的弊端，适应日益增长的数据处理的需求，人们开始探索新的数据管理方法和工具。在这一时期，存储技术取得了重要进展，大容量和快速存取的磁盘相继投入市场，为新型数据管理技术的研制提供了良好的物质基础。

由于计算机管理的数据量大，关系复杂，共享性要求强，外存出现了大容量磁盘、光盘。软件价格上升，硬件价格下降，编制和维护软件及应用程序成本相对增加，其中维护的成本更高，力求降低，这些也推动了数据库技术的发展。

数据管理技术进入数据库阶段的标志是20世纪60年代末的三个里程碑：1968年美国IBM公司推出层次模型的IMS系统；1969年美国数据系统语言委员会CODASYL组织发布了DBTG报告，总结了当时各式各样的数据库，提出网状模型；1970年美国IBM公司的E. F. Codd连续发表论文，提出关系模型，奠定了关系数据库的理论基础。



数据库系统阶段的主要特点是采用数据模型表示复杂的数据结构；有较高的数据独立性；数据库系统为用户提供了方便的用户接口；数据库系统提供数据库的并发控制、数据库的恢复、数据的完整性和数据安全性四方面的数据控制功能；增加了系统的灵活性和安全性。

### 1.1.3 数据库系统的三个发展阶段

#### 1. 第一代数据库系统

20世纪70年代，以层次型数据库和网状型数据库为代表的第一代数据库系统得到广泛应用。它们基本实现了数据管理中的“集中控制与数据共享”这一目标。

#### 2. 第二代数据库系统

20世纪80年代出现了以关系型数据库为代表的第二代数据库系统。

由于关系数据库具有坚实的理论基础、结构简单、操作方便，且在查询效率和数据库性能等方面取得了突破性的进展，因此在这一时期关系数据库得到了迅猛发展。关系数据库已逐渐成为数据库发展的主流，一批性能不断改善、版本不断更新的商品化关系数据库软件相继投入使用，如 Oracle、Sybase、Informix、Ingres 等关系数据库系统已广泛用于大型信息管理系统。关系数据库系统得到了以下方面的发展：

1) 关系数据库系统的发展促进了数据库系统的小型化。20世纪80年代推出了一批适于微机环境下运行的关系数据库系统，美国 Ashton-Tate 公司先后在1981年推出微机关系数据库软件 dBASE II，1984年升级到 dBASE III，1986年又推出 dBASE III Plus。1987年美国 FOX SOFTWARE 公司推出了 FoxBase + (V2.0)，之后又推出 Foxpro 等软件。Oracle 公司在1986年也研制出 Professional Oracle 版微机数据库软件。

2) 随着数据库技术的发展和计算机网络的广泛应用，分布式数据库系统在20世纪80年代得到了很大的发展。

如1986年 Oracle 公司和 Ingres 公司先后推出 SQL \* Star 和 Ingres/Star 的开放型分布式关系数据库系统。

3) 随着计算机的广泛应用，特别是一些新的应用领域不断提出新的应用要求(如图形图像处理、GIS 空间数据管理、XML 数据管理)，关系型数据库、层次型数据库、网状型数据库等传统数据库都表现出不同程度的局限性。因此人们在20世纪80年代后期又提出研制新一代数据库的设想。

#### 3. 第三代数据库系统

20世纪80年代末90年代初，新一代数据库技术的研究和开发已成为数据库领域学术界和工业界的研究热点。

例如，人们要求对图形、图像、声音等多媒体数据、时态数据、空间数据、知识信息以及各种复杂对象等非常规数据进行有效处理，为了适应应用需求，人们提出了许多新概念、新思想和新方法，以及一些新的数据模型和新数据库管理系统的体系结构。如多媒体数据库、时态数据库、空间数据库、面向对象数据库、分布式数据库、并行数据库系统、数据仓库、移动数据库、XML 数据管理技术等。

## 1.2 数据库系统基本概念

### 1.2.1 数据库

#### 1. 数据库

数据库(Database, 简记为DB)是长期存储在计算机内、有组织的、统一管理的相关数据的集合。DB能为各种用户共享,具有较小冗余度、数据间联系紧密而又有较高的数据独立性等特点。

#### 2. 数据库管理系统

数据库管理系统(Database Management System, 简记为DBMS)是位于用户与操作系统(OS)之间的一层数据管理软件,它为用户或应用程序提供访问DB的方法,包括DB的建立、查询、更新及各种数据控制。DBMS总是基于某种数据模型,可以分为层次型、网状型、关系型和面向对象型等。

### 1.2.2 数据库系统

数据库系统的简单结构如图1.1所示。

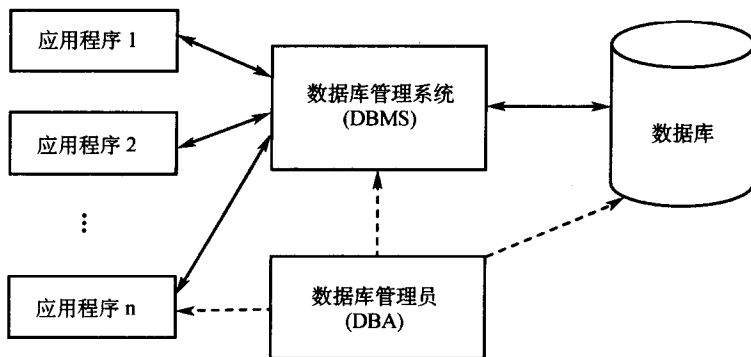


图 1.1 数据库系统的简单结构

图1.1中的数据库是数据的汇集,它们经一定的组织形式存储在磁盘介质中。数据管理系统(DBMS)是管理数据库的软件,它实现数据库系统的各种功能。应用程序是指以数据库为基础的各种应用程序,它必须通过DBMS访问数据库。数据库管理员是负责进行数据库规划、设计、协调、维护和管理等工作的人员。由这些应用程序、数据库管理系统、数据库和数据库管理员构成了数据库系统。

从图1.1可以看出,DBMS是数据库系统的核心,DBMS一般是由厂家提供的通用软件,如Oracle公司提供Oracle 9i、IBM公司的DB2等。DBMS的功能因产品而异,但现代DBMS一般必须具备以下基本功能:

#### 1. 提供高级的用户接口

用户通过DBMS看到的是与数据物理存储细节无关的数据逻辑形式,通过SQL等非过程数据库语言来对数据库进行操纵。为了适应不同用户的需要,DBMS通常提供各种接口。



## 2. 查询处理和优化

这里查询是泛指用户对数据库所提出的访问请求，包括数据修改、数据定义等。由于用户用非过程数据库语言来查询数据库，查询处理过程的拟定和优化都由 DBMS 来完成。查询处理和优化是 DBMS 的基本任务，它将对 DBMS 的性能影响颇大。

## 3. 数据目录管理

由于数据库中保存了持久和共享的数据，数据的定义也长期保存在数据库中，这就构成了数据目录。数据目录不但包含数据的逻辑属性，还包含数据的存储结构定义、数据访问和管理所必须的信息。DBMS 在完成各种功能时都离不开数据目录，数据目录管理是 DBMS 的基本功能。

## 4. 并发控制

现代 DBMS 一般允许多用户并发访问数据库，这就不可避免产生冲突，例如对同一数据，一个用户要读，另一个用户要写；如果是并发执行，则可能产生不正确的结果。为了解决冲突，DBMS 中需要有并发控制机制。

## 5. 恢复功能

这要求数据库在发生故障时，甚至遭到破坏时，也要能够恢复到数据一致状态。

## 6. 完整性约束功能

由于数据库中的数据是持久和共享的，保证其正确性是至关重要的。在 DBMS 中不但要对数据进行语法检查，还要对数据进行语义检查。数据在语义上的约束称为完整性约束。

例如，每个学生的学号应该是唯一的，年龄不能是负数或大于 200 等，都是完整性约束。

## 7. 访问控制

DBMS 应有控制用户访问数据库中数据的访问权限控制。例如对于学生成绩这一数据，允许所有教务人员读，但只允许成绩登录人员写，只允许学生查看自己的课程成绩等。

# 1.3 数据库系统的特点

## 1. 实现数据的集中化控制

以往文件系统中文件是按用户进行分散管理的，而数据库系统则能集中控制和管理数据。

(1) 数据库中的数据是集成式的 通常将一个部门所涉及的全部数据都组织在一个数据库中。例如我们通常利用数据库系统将一个学校的人事档案管理、学生学籍管理、教学管理等各种相关的数据集中在一个数据库中进行统一维护和管理，学校各职能部门可以通过访问权限从数据库中提取相应的所需数据。

(2) 数据库管理员 (DBA) 通过数据库管理员对本单位的数据库进行管理和维护。

## 2. 数据的冗余度小，易扩充

通过数据库系统可以将相同的数据在数据库中只存储一次，并为不同的应用共享，这比文件系统的数据库冗余要小得多。

## 3. 采用一定的数据模型实现数据结构化

数据库中的数据是按照一定的数据模型来组织、描述和存储的，称为数据结构化。在数

数据库系统中常用的数据模型有层次模型、网状模型和关系模型等。通常每一个数据库系统只选用一种数据模型，根据选用的数据模型不同，数据库分为层次型数据库、网状型数据库和关系型数据库。

#### 4. 避免了数据的不一致性

数据的不一致性是指数据的不相容性和矛盾性，即数据违反了数据完整性约束条件，使同一数据在数据库内重复出现不同的值。例如，职工工资在工资单上和人事档案中有不同的值，这就是数据的不一致性，主要是由于存在数据冗余所造成的。

在数据库系统中，由于冗余度的大量减少，避免了大量数据的不一致产生。同时数据库系统提供了对数据进行控制和检查功能，从而使得数据在更新时能同时更新数据的所有副本，以保证数据的一致性。

#### 5. 实现数据共享

在数据库系统中，一组数据集合为多种应用程序和多个用户共同使用。主要体现在当前所有用户(批处理用户、终端用户)可以同时使用数据库；数据库既能满足用户的需求，又能适应新的应用需求，即增加新应用需求不会影响原有的应用；具有多种用户接口，可以提供与 PL/SQL、C/C++、COBOL、VB 等数据库接口。

#### 6. 提供数据库保护

数据库系统提供了安全性控制、完整性控制、并发控制、故障的检测与恢复等一系列数据库控制功能，以确保数据库中的数据是安全可靠的。

#### 7. 数据独立性

数据独立性是指数据库中的数据独立于应用程序，即数据的逻辑结构、存储结构与存取方式的改变不影响应用程序。数据独立性一般分为数据的逻辑独立性和数据的物理独立性。其示意图如图 1.2 所示。

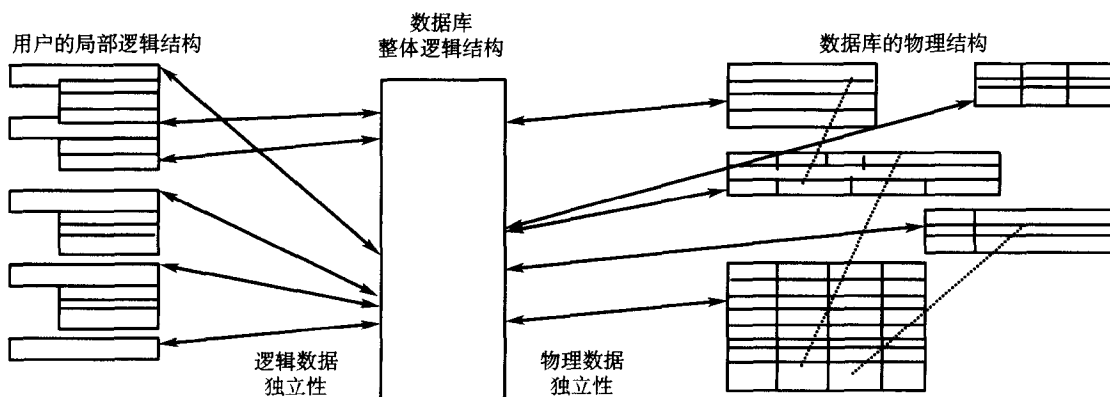


图 1.2 数据独立性示意图

(1) 数据逻辑独立性 数据逻辑独立性是指数据库总体逻辑结构的改变(如修改数据定义、增加新的数据类型、改变数据间的联系等)不需要修改应用程序。

(2) 数据物理独立性 数据物理独立性是指数据的物理结构(存储结构、存取方式等)的改变，如存储设备的更换、物理存储格式和存取方式的改变等不影响数据库的逻辑结构，因而不会引起应用程序的变化。





## 8. 数据由 DBMS 统一管理和控制

数据库中的多用户并发共享数据须通过 DBMS 进行统一管理和控制。

(1) 数据的安全性保护 数据的安全性保护是指保护数据以防止不合法的使用而造成数据的泄密和破坏。

(2) 完整性保护 完整性保护是指将数据控制在有效范围内(如年龄在 1~200)或使数据之间满足一定的关系(外键值为引用实体的关键字值或为空值),以保证数据的正确性、有效性和相容性。

(3) 并发控制 当多个用户进程同时对数据库进行并发存取时,必须对多用户的并发操作加以控制和协调。

(4) 数据库恢复 当数据库系统发生故障时, DBMS 必须具有将数据库从错误状态恢复到某个正确状态的功能。