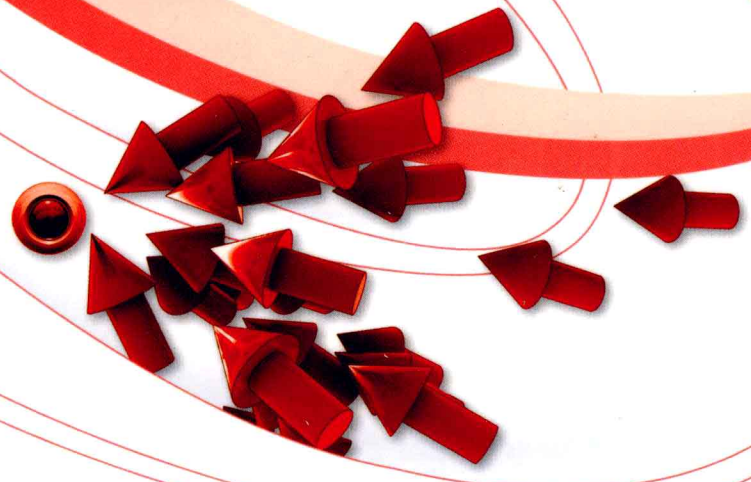

高等学校计算机规划教材

数据库原理与应用

高　凯　主编
张雪梅　倪素虹　张永强　白云飞　白红武　副主编

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http: //www.phei.com.cn

高等学校计算机规划教材

数据库原理与应用

高　凯　主编

张雪梅　倪素虹　张永强　白云飞　白红武　副主编

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

本书全面、系统地介绍了关系数据库基础理论及应用。在介绍关系数据库系统理论、SQL 语言、规范化理论、数据库设计、关系查询处理和查询优化、数据库完整性和安全性、数据库恢复和并发控制等理论的基础上，针对应用实际，介绍了以 VB、Delphi、VC++、C#等常用编程工具开发数据库应用程序的常用方法和技巧，介绍了基于数据库的 Web 应用程序开发、XML 技术等。本书配有 PPT、源代码等教学资源，方便教学和工程实践应用。

本书可作为普通高等学校计算机及相关专业“数据库原理与应用”课程的教材，也可作为从事数据库应用程序设计与开发的工程技术人员、希望了解数据库相关技术的爱好者的参考书。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容
版权所有 • 侵权必究

图书在版编目(CIP)数据

数据库原理与应用/高凯主编. —北京：电子工业出版社，2011.1
(高等学校计算机规划教材)
ISBN 978-7-121-12461-7

I. ①数… II. ①高… III. ①数据库系统－高等学校－教 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2010)第 236460 号

策划编辑：史鹏举
责任编辑：史鹏举　　文字编辑：侯丽平
印　　刷：北京市天竺颖华印刷厂
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×1092　1/16　印张：20.25　字数：585 千字
印　　次：2011 年 1 月第 1 次印刷
定　　价：37.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010) 88258888。

前　言

在社会信息化的今天，信息已经成为全社会宝贵的资源。就信息本身而言，大致可将其分为结构化信息、半结构化信息和非结构化信息三类。作为管理结构化信息的有效手段，数据库系统对于当今科研部门、政府机关、企事业单位等都是至关重要的；作为数据库系统中的核心，数据库管理系统DBMS——特别是关系型数据库管理系统(如 Microsoft SQL Server、Oracle、Sybase 等)——用于高效创建数据库和存储大量数据并对其进行有效管理和维护，其本身具有很大的工程实用价值。同时，数据库技术也是计算机科学与技术领域中发展飞快的一个分支，在其短短 40 年的发展历程中，已造就了包括 C. W. Bachman、E. F. Code、James Gray 等在内的多位图灵奖得主，发展成为以数据建模和 DBMS 为核心、具有很大工程实践价值的学科，数据库及其应用技术也因此成为计算机科学与技术及相关专业的重要专业课程。同时也要看到，虽然传统的关系型数据库系统在管理结构化信息方面具有很大优势，但在网络迅速普及的今天，半结构化信息和非结构化信息所占的比重正在逐步增大，已经逐渐成为重要的信息组织方式。因此，在本书前几章对传统的关系数据库系统理论、关系数据库标准语言 SQL、规范化理论、数据库设计、关系查询处理和查询优化、数据库完整性和安全性、数据库恢复和并发控制等基础理论进行综述的基础上，针对当前的应用实际，对基于数据库的 Web 应用程序开发与实现、XML 技术在数据库中的应用等进行概述，以使读者能对包括结构化、半结构化信息在内的数据管理技术有较全面的了解。

全书分为三个部分。

第一部分是原理篇，第 1～3 章讲述了关系数据库基础及 SQL，第 4 章讲述了关系规范化理论，第 5～8 章介绍了数据库设计理论、关系查询处理和查询优化、数据库完整性和安全性、数据库的恢复与并发控制理论等。实践出真知，对于有着广泛工程实践应用背景的数据库技术而言更是如此。在本书的编纂过程中，编者一直秉承着以能力培养为中心、以工程实践为手段、以工程案例驱动教学的讲练结合的编写原则，以期能帮助读者更好地学习和领会关系数据库管理系统——特别是广泛使用的 Microsoft SQL Server——在实践中的具体应用。

为此，本书第二部分是应用篇，第 9～12 章分别介绍了以 VB、Delphi、VC++、C#等常用编程工具开发数据库应用程序的方法和技巧，并给出应用实例，围绕数据库应用系统的使用实际，选用适宜的开发工具完成任务。而对于每个功能的讲解，力求以明确的步骤和实例应用来指导操作，通过典型应用实例训练，使读者在掌握关系数据库理论的基础上，在实践中掌握相关的数据库编程技巧。第 13、14 章介绍了基于数据库的 Web 应用程序开发、XML 和数据库技术的应用等，以便读者能开发出基于数据库的 Web 相关程序，并对半结构化数据的处理有所了解。

第三部分是实验和辅助工具篇，包括数据库实验、SQL Server 系统存储过程介绍、Eclipse 的简单使用等。

全书理论联系实际，涉及面广，体系完整，内容新颖，条理清晰，组织合理，图例丰富，说明详细，适用于作为计算机科学与技术、软件工程、网络工程、通信工程、电子科学与技术等相关学科本科教学的教材，也可作为工程技术人员的技术参考书。

作为河北省电子信息类本科教学高地中的重点专业建设项目，为更好地服务于课程建设，**本书配有电子课件、完整的开发案例源代码等教学资源，**需要者可从华信教育资源网 http://www.hxedu.com.cn **免费注册下载。**

本书由高凯、张雪梅、倪素虹、张永强、白云飞、白红武、孟广学等人合作编写。高凯组稿并提出写作大纲，倪素虹撰写第 1、2 章，白云飞撰写第 3、4 章，张雪梅撰写第 5～8 章，张永强撰写第 9～12 章，高凯撰写第 13、14 章并完成书稿的最后审定和统稿工作。

在本书的写作过程中，得到了河北省教育厅科学研究计划(编号：2009435)、河北省科技支撑计划(编号：09243522D)和相关部门的资助。第 12 章中涉及的项目已经作为某高校教材征订系统中的一部分功能在使用；部分章节中的示例得到翰子昂教育培训机构的帮助(其提供的部分案例帮助我们设计了部分示例)。另外，在书稿的编纂过程中，王亚歌、王伟完成第 14 章部分编程和文字撰写工作并整理了参考文献。郭立炜、王俊社、周万珍、杨奎河、王晓红、吴会丛、王建霞、武卫东、许云峰、高国江、乔世权等均对本书的编写工作提出很多好的意见和建议，并完成了部分内容的编写和校对工作，而国内外众多的数据库应用技术的研究和相关网站亦为本书提供了良好的基础和参考，本书的顺利完成也得益于参阅了大量的相关工作及研究成果，在此谨向这些文献的作者和科研工作人员，特别是那些由于篇幅所限未能在参考文献中提及的作者，以及为本书提供帮助的老师、同仁和课题组成员致以诚挚的谢意和崇高的敬意。在本书写作过程中，也得到了电子工业出版社策划编辑史鹏举等的大力支持和帮助，在此一并表示衷心感谢。

由于我们的学识、水平均有限，书中不妥之处在所难免，恳请广大读者批评指正。

编　者

此为试读，需要完整PDF请访问：www.ertongbook.com

目　录

第一部分　原　理　篇

第 1 章　绪论 …… (2)
1.1　数据管理技术的产生和发展 …… (2)
1.1.1　人工管理阶段 …… (2)
1.1.2　文件系统阶段 …… (3)
1.1.3　数据库系统阶段 …… (3)
1.2　数据库基本概念 …… (5)
1.2.1　数据 …… (5)
1.2.2　数据库 …… (5)
1.2.3　数据库管理系统 …… (6)
1.2.4　数据库系统 …… (6)
1.3　数据模型 …… (7)
1.3.1　数据模型的组成要素 …… (7)
1.3.2　数据模型的分类 …… (8)
1.3.3　概念模型及表示方法 …… (8)
1.3.4　数据模型 …… (12)
1.4　数据库系统的结构 …… (14)
1.4.1　数据库系统的三级模式结构 …… (15)
1.4.2　数据库系统的二级映像 …… (15)
1.5　数据库管理系统 …… (16)
1.5.1　DBMS 的工作模式 …… (16)
1.5.2　DBMS 的功能 …… (17)
1.5.3　DBMS 的组成 …… (17)
1.6　小结 …… (18)
习题 1 …… (19)

第 2 章　关系数据库基础 …… (20)
2.1　关系模型的基本概念 …… (20)
2.1.1　域、笛卡儿积、关系、关系模式 …… (20)
2.1.2　码 …… (23)
2.2　关系模型的完整性约束 …… (24)
2.2.1　实体完整性规则 …… (24)
2.2.2　参照完整性规则 …… (25)
2.2.3　用户自定义完整性规则 …… (25)
2.3　关系操作 …… (26)
2.3.1　关系数据语言 …… (26)
2.3.2　关系代数 …… (26)
2.4　小结 …… (34)
习题 2 …… (34)

第 3 章　关系数据库标准语言 SQL …… (37)
3.1　SQL 概述 …… (37)
3.1.1　SQL 的发展历史 …… (37)
3.1.2　SQL 的特点 …… (37)
3.1.3　SQL 数据库的体系结构 …… (38)
3.2　数据定义 …… (39)
3.2.1　模式的定义与删除 …… (40)
3.2.2　基本表的定义、删除和修改 …… (40)
3.2.3　索引的建立和删除 …… (43)
3.3　数据查询 …… (44)
3.3.1　单表无条件查询 …… (45)
3.3.2　单表带条件查询 …… (48)
3.3.3　分组查询和排序查询 …… (51)
3.3.4　连接查询 …… (53)
3.3.5　嵌套查询 …… (56)
3.3.6　集合查询 …… (60)
3.4　数据更新 …… (62)
3.4.1　插入数据 …… (62)
3.4.2　删除数据 …… (63)
3.4.3　修改数据 …… (63)
3.5　视图 …… (64)
3.5.1　建立视图 …… (64)
3.5.2　删除视图 …… (65)
3.5.3　查询视图 …… (66)
3.5.4　更新视图 …… (66)
3.5.5　视图的作用 …… (67)
3.6　小结 …… (68)
习题 3 …… (68)

第 4 章　规范化理论 …… (69)
4.1　问题的提出 …… (69)

4.1.1 关系模式中可能存在的问题 (69)
4.1.2 解决的方法 …… (70)
4.2 函数依赖 …… (70)
4.2.1 函数依赖的基本概念 …… (71)
4.2.2 函数依赖的推理规则 …… (72)
4.2.3 码的函数依赖表示 …… (75)
4.2.4 最小函数依赖集 …… (75)
4.3 规范化 …… (77)
4.3.1 范式 …… (77)
4.3.2 模式分解 …… (80)
4.3.3 关系模式规范化步骤 …… (86)
4.4 小结 …… (87)
习题 4 …… (87)

第 5 章 数据库设计 …… (89)
5.1 数据库设计概述 …… (89)
5.1.1 数据库设计的任务和特点 (89)
5.1.2 数据库设计方法 …… (90)
5.1.3 数据库设计步骤 …… (90)
5.1.4 数据库设计过程中的各级模式 …… (92)
5.2 需求分析 …… (92)
5.2.1 需求分析的任务 …… (92)
5.2.2 需求分析的方法和步骤 … (93)
5.2.3 需求分析的结果 …… (94)
5.3 概念结构设计 …… (95)
5.3.1 概念结构的特点 …… (95)
5.3.2 概念设计方法 …… (95)
5.3.3 概念设计的步骤 …… (95)
5.4 逻辑结构设计 …… (100)
5.4.1 E-R 模型到关系模式的转换 …… (100)
5.4.2 逻辑模式的优化 …… (103)
5.4.3 设计用户子模式 …… (103)
5.5 数据库的物理设计 …… (104)
5.5.1 数据库物理设计的内容和方法 …… (104)
5.5.2 确定数据库的物理结构 … (105)
5.5.3 评价物理结构 …… (106)
5.6 数据库的实施和维护 …… (106)
5.6.1 建立数据库和应用程序的调试 …… (106)
5.6.2 数据库的试运行 …… (106)
5.6.3 数据库的运行和维护 …… (107)
5.7 小结 …… (108)
习题 5 …… (108)

第 6 章 关系查询处理和查询优化 …… (109)
6.1 关系数据库系统的查询处理 … (109)
6.1.1 查询优化的必要性 …… (109)
6.1.2 查询处理的步骤 …… (111)
6.1.3 查询的执行代价 …… (111)
6.2 代数优化 …… (112)
6.2.1 关系代数表达式等价变换规则 …… (112)
6.2.2 查询优化一般策略 …… (113)
6.2.3 关系代数表达式的优化算法 …… (114)
6.3 物理优化 …… (116)
6.3.1 基于规则的优化方法 …… (116)
6.3.2 基于代价估算的优化方法 … (116)
6.4 SQL 查询语句的优化 …… (117)
6.4.1 合理使用索引 …… (117)
6.4.2 书写高效率的 SQL 查询语句 …… (118)
6.4.3 使用存储过程 …… (120)
6.5 小结 …… (123)
习题 6 …… (123)

第 7 章 数据库完整性和安全性 …… (124)
7.1 数据库完整性 …… (124)
7.1.1 实体完整性 …… (124)
7.1.2 参照完整性 …… (125)
7.1.3 用户定义的完整性 …… (127)
7.1.4 完整性约束命名子句 …… (128)
7.1.5 触发器 …… (129)
7.2 数据库安全性 …… (130)
7.2.1 数据库安全性概述 …… (130)
7.2.2 用户标识与鉴别 …… (130)
7.2.3 存取控制 …… (131)
7.2.4 其他安全控制方法 …… (135)
7.3 SQL Server 的完整性控制和安全控制 …… (137)
7.3.1 SQL Server 的完整性控制 … (137)
7.3.2 SQL Server 的安全控制 …… (139)
7.4 小结 …… (141)
习题 7 …… (142)

第 8 章　数据库恢复和并发控制 …… (143)
8.1　事务 …… (143)
8.1.1　事务的基本概念 …… (143)
8.1.2　事务的性质 …… (143)
8.2　数据库恢复技术 …… (144)
8.2.1　故障的种类 …… (144)
8.2.2　恢复的实现技术 …… (145)
8.2.3　故障恢复的策略 …… (146)
8.2.4　SQL Server 的数据库恢复技术 …… (149)
8.3　并发控制 …… (151)
8.3.1　并发操作带来的问题 …… (151)
8.3.2　封锁 …… (152)
8.3.3　并发调度的可串行性 …… (155)
8.3.4　两段锁协议 …… (156)
8.3.5　封锁的粒度 …… (157)
8.3.6　SQL Server 的封锁方式 …… (159)
8.4　小结 …… (159)
习题 8 …… (160)

第二部分　应　用　篇

第 9 章　数据库应用技术概述 …… (162)
9.1　数据库访问接口 …… (162)
9.1.1　ODBC 简介 …… (162)
9.1.2　DAO 简介 …… (163)
9.1.3　OLE DB 简介 …… (164)
9.1.4　ADO 简介 …… (165)
9.1.5　ADO.NET 简介 …… (165)
9.1.6　内嵌式 SQL、RDO、DB 库简介 …… (166)
9.2　数据库应用体系结构 …… (167)
9.2.1　客户机/服务器工作模式 …… (168)
9.2.2　客户机/服务器模式的主要技术特征 …… (168)
9.2.3　客户机/服务器系统体系结构 …… (168)
9.2.4　分布式数据库系统 …… (171)
9.3　基于 B/S 架构的数据库应用程序 …… (172)
9.4　ODBC 数据访问接口及实例操作 …… (173)
9.4.1　ODBC 概述 …… (173)
9.4.2　MFC ODBC 数据库类 …… (180)
9.4.3　应用案例——VC++6.0 中使用 ODBC 开发个人通讯录系统 …… (185)
9.5　小结 …… (194)
习题 9 …… (194)

第 10 章　ADO 数据访问技术及操作 …… (195)
10.1　ADO 概述 …… (195)
10.2　ADO 数据库访问对象 …… (196)
10.2.1　Connection 对象 …… (196)
10.2.2　Command 对象 …… (197)
10.2,3　Recordset 对象 …… (198)
10.2.4　Parameter 对象 …… (201)
10.2.5　Field 对象 …… (201)
10.2.6　Record 对象 …… (202)
10.2.7　Stream 对象 …… (203)
10.2.8　Error 对象 …… (204)
10.2.9　Property 对象 …… (204)
10.3　高级语言操作 ADO 接口 …… (204)
10.3.1　在 VB 6.0 中使用 ADO …… (204)
10.3.2　在 VC 6.0 中使用 ADO …… (207)
10.3.3　在 Delphi 7 中使用 ADO …… (210)
10.4　应用案例——Delphi 中使用 ADO 开发个人通讯录系统 …… (213)
10.5　小结 …… (215)
习题 10 …… (215)

第 11 章　ADO.NET 数据访问技术及操作 …… (216)
11.1　.NET 平台及 VS.NET 2008 简介 …… (216)
11.1.1　.NET Framework 简介 …… (216)
11.1.2　VS.NET 2008 集成开发环境简介 …… (217)
11.1.3　C#开发语言简介 …… (218)
11.2　ADO.NET 概述 …… (218)
11.2.1　ADO.NET 的优点 …… (218)
11.2.2　ADO.NET 的结构 …… (219)
11.3　使用 ADO.NET 进行数据库访问 …… (220)

11.3.1 ADO.NET Connection 对象 …… (221)
11.3.2 ADO.NET Command 对象 …… (222)
11.3.3 ADO.NET DataReader 对象 …… (223)
11.3.4 ADO.NET DataAdapter 对象 …… (224)
11.3.5 ADO.NET DataSet 对象 …… (226)
11.4 应用案例——C#中使用 ADO.NET 开发个人通讯录系统 …… (227)
11.4.1 通过编写代码实现通讯录系统 …… (227)
11.4.2 使用 ADO.NET 控件实现通讯录系统开发 …… (229)
11.5 小结 …… (235)
习题 11 …… (235)

第 12 章 基于 C#的数据库系统应用案例——教材征订系统 …… (236)
12.1 系统设计 …… (236)
12.1.1 需求分析 …… (236)
12.1.2 概念结构设计 …… (237)
12.1.3 数据库设计 …… (238)
12.2 功能模块设计 …… (241)
12.2.1 系统功能模块 …… (241)
12.2.2 界面设计 …… (241)
12.3 功能实现步骤 …… (242)
12.3.1 数据库创建 …… (242)
12.3.2 公用模块设计 …… (242)
12.3.3 添加 app.config 文件 …… (247)
12.4 管理员端程序设计步骤 …… (247)
12.4.1 用户管理界面 …… (247)
12.4.2 用户管理信息编辑界面设计 …… (250)
12.4.3 登录窗体的设计 …… (254)
12.5 系统运行与调试 …… (256)
12.5.1 调试方法 …… (256)
12.5.2 调试过程 …… (256)
12.5.3 排错 …… (257)
12.6 小结 …… (258)
习题 12 …… (258)

第 13 章 基于数据库的 Web 应用程序开发 …… (259)
13.1 概述 …… (259)
13.2 常见的 Web 服务器简介 …… (260)
13.2.1 IIS …… (260)
13.2.2 Tomcat …… (262)
13.2.3 WebLogic …… (263)
13.3 服务器端的动态网页技术 …… (264)
13.3.1 JSP …… (264)
13.3.2 基于 ASP .NET 构建 Web 应用程序 …… (268)
13.4 客户端网页设计相关技术简介 …… (271)
13.4.1 基于 CSS+DIV 构建网页 …… (271)
13.4.2 客户端脚本语言 JavaScript 简介 …… (273)
13.5 小结 …… (276)
习题 13 …… (276)

第 14 章 XML 与数据库技术 …… (277)
14.1 基于 XML 的数据库信息管理 …… (277)
14.1.1 XML 及其与数据库的互操作 …… (278)
14.1.2 XML 的基本规范和简单结构 …… (279)
14.1.3 XML 数据的显示 …… (279)
14.1.4 数据库信息转换为 XML …… (281)
14.1.5 XML 与数据库信息的互转换 …… (283)
14.2 基于 XML 的网络异构数据的采集与处理 …… (285)
14.2.1 异构资源采集 …… (286)
14.2.2 利用 HttpClient 检索数据 …… (286)
14.2.3 构造 XML 文档 …… (288)
14.2.4 在浏览器端显示 XML 文档 …… (291)
14.2.5 提供统一检索接口 …… (293)
14.3 小结 …… (294)
习题 14 …… (294)

第三部分　实验和辅助工具篇

实验 1　熟悉 SQL Server 2000 环境及数据库管理 …… (296)
实验 2　数据定义 …… (297)
实验 3　数据操纵 …… (298)
实验 4　简单查询 …… (300)
实验 5　连接查询和嵌套查询 …… (300)
实验 6　数据库的完整性控制 …… (301)
实验 7　数据库的安全性 …… (302)

第一部分 原 理 篇

第 1 章 绪论
第 2 章 关系数据库基础
第 3 章 关系数据库标准语言 SQL
第 4 章 规范化理论
第 5 章 数据库设计
第 6 章 关系查询处理和查询优化
第 7 章 数据库完整性和安全性
第 8 章 数据库恢复和并发控制

此为试读，需要完整PDF请访问：www.ertongbook.com

第1章 绪　论

近年来，随着计算机技术与网络技术的飞速发展，信息技术在各个领域的深度融合推动了数据库技术在各个领域的广泛应用。从小型事务处理到大型信息系统，从联机事务处理到联机分析处理，从一般企业管理到计算机辅助设计与制造、电子政务、电子商务、地理信息系统等，数据库系统已经渗透到人们工作和生活的方方面面。因此，数据库技术已成为当今计算机信息系统的核心技术，是计算机技术和应用发展的基础。经过40多年的发展，它已形成了较为完整的理论体系和实用技术。

本章首先回顾了数据管理技术的发展过程，然后介绍数据库的基本概念和数据模型，最后阐述数据库系统的体系结构。

1.1 数据管理技术的产生和发展

研制计算机的初衷是进行复杂的科学计算，随着计算机技术的发展和数据处理的需求，需要对大量的数据进行处理来得到有价值的信息，但由于数据量大，数据结构复杂，数据之间有复杂的逻辑联系，所以需要一个通用、高效而又使用方便的管理软件，以便把数据有效地管理起来，这样，数据管理技术也就相应产生了。

随着计算机硬件和软件的发展，数据管理技术大致经历了三个阶段：人工管理阶段、文件系统阶段和数据库系统阶段。

1.1.1 人工管理阶段

在20世纪50年代中期以前，计算机技术相对落后，计算机主要用于科学计算。硬件方面，计算机的外存只有磁带、卡片、纸带，没有磁盘等直接存取的存储设备，存储量非常小。软件方面，只有汇编语言，没有操作系统和管理数据的软件。数据处理的方式是批处理，也即机器一次处理一批数据，直到运算完成为止，然后才能进行另外一批数据的处理，中间不能被打断，原因是当时的外存如磁带、卡片等只能顺序输入。

这个阶段数据管理的特点如下：

(1) 数据不保存。在需要计算时，利用卡片、纸带等将数据输入，经过运算得到运算结果，数据处理的过程就结束了。

(2) 没有专用的软件对数据进行管理。数据需要由应用程序自己管理，每个应用程序不仅要规定数据的逻辑结构，而且要设计物理结构，包括存储结构、存取方法、输入方式等。因此，程序员负担很重。

(3) 数据不共享。数据是面向应用程序的，一组数据只能对应一个应用程序。多个应用程序若涉及某些相同的数据，也必须各自定义，无法相互利用，互相参照。数据不但高度冗余，而且不能共享。

(4) 数据不具有独立性。数据是程序的组成部分，即程序和数据是一个不可分割的整体，数据和程序同时提供给计算机运算使用。对数据进行管理，就像现在的操作系统以目录、文件的形式管理数据。程序员不仅要知道数据的逻辑结构，也要规定数据的物理结构，程序员对存储结构、存取方法及输入/输出的格式有绝对的控制权，要修改数据必须修改程序。例如，要对100组数据进行同样的运算，就要给计算机输入100个独立的程序，因为数据无法独立存在。

在人工管理阶段，程序与数据之间的关系如图1.1所示。

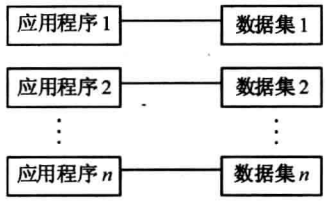

图 1.1　人工管理阶段程序与数据之间的关系

1.1.2　文件系统阶段

从 20 世纪 50 年代后期到 60 年代中期，数据管理发展到文件系统阶段。此时的计算机不仅用于科学计算，还大量应用于数据管理。在硬件方面，有了磁盘、磁鼓等直接存储设备；在软件方面，出现了高级语言和操作系统，且操作系统中有了专门管理数据的软件，一般称之为文件系统。从处理方式上讲，不仅有了文件批处理，而且能够联机实时处理，联机实时处理是指在需要的时候随时从存储设备中查询、修改或更新，因为操作系统的文件管理功能提供了这种可能。

用文件系统管理数据的特点如下：

(1) 数据可长期保存。数据可以长期保留在外存上反复处理，即可以经常对数据文件进行查询、修改、插入和删除等操作。

(2) 由文件系统对数据的存取进行管理。文件系统把数据组织成相互独立的数据文件，利用“按名访问，按记录存取”的管理技术，对文件进行修改、插入和删除等操作。因此，程序员只与文件名打交道，不必明确数据的物理存储结构，大大减轻了程序员的负担。

(3) 文件组织多样化。文件有顺序文件、链接文件、索引文件等，因而对文件的记录可顺序访问，也可随机访问，更便于存储和查找数据。但文件之间相互独立、缺乏联系。数据之间的联系要通过程序去构造。

(4) 程序与数据之间有一定独立性。由文件系统进行数据管理，程序和数据之间由软件提供的存取方法进行转换，数据存储发生变化不一定影响程序的运行，既可大大节省维护的工作量，又可减轻程序员的负担。

与人工管理阶段相比，文件系统阶段对数据的管理有了很大的进步，但一些根本性问题仍没有彻底解决，主要表现在以下三方面：

(1) 数据冗余度大，一致性差。在文件系统中，一个文件基本上对应于一个应用程序。如果多个应用程序的数据完全相同，可以以文件为单位实现共享。当不同的应用程序具有部分相同的数据时，仍需建立各自的独立数据文件，而不能共享相同的数据。因此，数据的冗余度大，空间浪费严重。同时由于相同数据的重复存储、各自管理，容易造成数据的不一致性，给数据的修改和维护带来了困难。

(2) 数据联系弱。不同的数据文件之间相互独立，缺乏联系。虽然某些数据之间存在着紧密的联系，但是由于实现的复杂性，很少在系统中提供数据之间的紧密联系。

(3) 数据独立性差。数据和程序相互依赖，一旦改变数据的逻辑结构，必须修改相应的应用程序。而应用程序发生变化，如改用另一种程序设计语言来编写程序，有时也需修改数据结构。因此，数据和程序之间缺乏独立性。

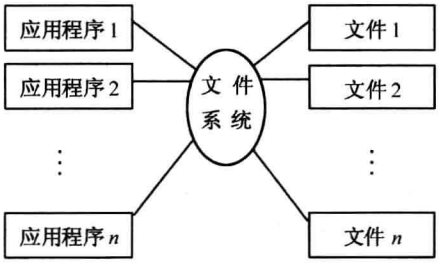

图 1.2　文件系统阶段程序与数据之间的关系

在文件系统阶段，程序与数据之间的关系如图 1.2 所示。

1.1.3　数据库系统阶段

20 世纪 60 年代后期，计算机用于管理的规模更为庞大，应用越来越广泛，数据量急剧增长，同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。硬件已有了大容量的磁盘，硬件

价格下降。在处理方式上，联机实时处理要求更多，并开始提出和考虑分布处理。在这样的背景下，以文件系统作为数据管理手段已经不能满足应用的需求，为解决多用户、多应用共享数据的需求，出现了统一管理数据的专门软件系统——数据库系统。

从文件系统到数据库系统，是数据管理技术的一个飞跃。数据库系统管理数据的特点如下：

(1) 数据结构化

数据结构化是数据库系统与文件系统的根本区别。在文件系统阶段，只考虑了同一文件记录内部数据项之间的联系，而不同文件的记录之间是没有联系的，也就是说，从整体上看数据是无结构的。在数据库中，实现了整体数据的结构化，把文件系统中的简单的记录结构变成了记录和记录之间的联系所构成的结构化数据。在描述数据的时候，不仅要描述数据本身，还要描述数据之间的联系。

(2) 数据共享性高、冗余小，易扩充

数据库系统从全局角度看待和描述数据，数据不再面向某个应用程序而是面向整个系统，因此数据可以被多个用户、多个应用共享使用。这样便减小了不必要的数据冗余，节约存储空间，同时也避免了数据之间的不相容性与不一致性。

由于数据面向整个系统，是有结构的数据，不仅可被多个应用共享使用，而且容易增加新的应用，这就使得数据库系统弹性大，易于扩充，可以适应各种用户的要求。

(3) 数据独立性高

数据独立性是数据库领域中一个常用术语，是指数据库中的数据与应用程序之间依赖性弱，独立性强，包括数据的物理独立性和数据的逻辑独立性。

物理独立性是指用户的应用程序与存储在磁盘数据库中的数据是相互独立的。也就是说，数据在数据库中怎样存储是由数据库系统管理的，应用程序不需要了解，应用程序要处理的只是数据的逻辑结构，即使数据的物理存储改变了，应用程序也不用改变。

逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。也就是说，数据的逻辑结构改变了，应用程序也可以不变。

由于数据与程序的独立，把数据的定义从程序中分离出去，加上数据的存取又由数据库系统负责，从而简化了应用程序，大大减少了应用程序的维护和修改的工作量。

(4) 有统一的数据控制功能

数据库为多个用户和应用程序所共享，对数据的存取往往是并发的，即多个用户可以同时存取数据库中的数据，甚至可以同时存取数据库中的同一个数据，为确保数据库数据的正确有效和数据库系统的有效运行，数据库系统提供下述四方面的数据控制功能。

(1) 数据的安全性(security)控制。数据的安全性是指保护数据以防止不合法使用数据造成数据的泄露和破坏，保证数据的安全和机密，使每个用户只能按规定对某些数据以某些方式进行使用和处理。例如，系统提供口令检查或其他手段来验证用户身份，防止非法用户使用系统；也可以对数据的存取权限进行限制，只有通过权限检查后才能执行相应的操作。

(2) 数据的完整性(integrity)控制。数据的完整性是指系统通过设置一些完整性规则以确保数据的正确性、有效性和相容性。完整性控制将数据控制在有效的范围内，或保证数据之间满足一定的关系。其中，有效性是指数据是否在其定义的有效范围内，如月份只能用 1～12 之间的正整数表示。正确性是指数据的合法性，如年龄属于数值型数据，只能含 0, 1, … ,9，不能含字母或特殊符号。相容性是指表示同一事实的两个数据应相同，否则就不相容，如一个人不能有两个性别。

(3) 并发(concurrency)控制。多用户同时存取或修改数据时，可能会发生相互干扰而提供给用户不正确的数据，并使数据库的完整性受到破坏，因此必须对多用户的并发操作加以控制和协调。

(4) 数据恢复(recovery)。计算机系统出现各种故障是很正常的，数据库中的数据被破坏、丢失也是可能的。当数据库被破坏或数据不可靠时，系统应有能力将数据库从错误状态恢复到最近某一时刻的正确状态。

在数据库系统阶段，程序与数据之间的关系如图1.3所示。

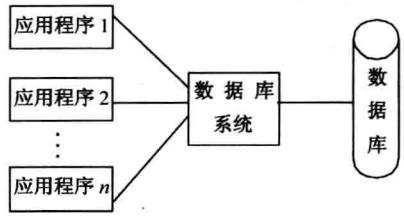

图 1.3　数据库系统阶段程序与数据之间的关系

从文件系统发展到数据库系统是信息处理领域的一个重大变化。在文件系统阶段，人们对信息处理方式关心的中心问题是应用系统功能的设计，程序设计处于主导地位，数据只起着服从程序设计需要的作用。在数据库系统管理阶段，人们关心的问题是数据结构的设计，它是整个数据库应用的核心，而数据库应用的设计则处于以既定的数据结构为基础的外围地位。

从数据库技术的发展过程和演变趋势来看，数据库系统本身也在不断发展，从最初的层次数据库系统、网状数据库系统，向关系型数据库系统、关系对象数据库系统、对象数据库系统发展。本书重点介绍关系数据库系统。

1.2　数据库基本概念

下面介绍一些数据库常用的术语和基本概念。

1.2.1　数据

数据是数据库中存储的基本对象，它有多种表现形式，在大多数人的头脑中，数据就是数字，其实数字只是最简单的一种数据，是数据的一种传统和狭义的理解。广义地理解，数据的种类很多，文字、图形、图像、声频、视频等，这些都可以作为数据，经过数字化后存入计算机。可以对数据做如下定义：数据是指描述事物的符号记录，这些符号可以是文字、图形、图像、声频、视频等。

数据的含义称为数据的语义，数据与其语义是不可分的。例如，学生档案表中有一个记录的描述如下：

(081637121，黎明，男，1987-5-11，计算机系)

这里的学生记录就是数据。对于这条记录，了解其含义后将得到如下信息：学号 081637121，姓名黎明，男，出生于 1987 年 5 月 11 日，在计算机系读书；而不了解含义的人则无法得到有用信息。可见，数据的形式本身还不能完全表达其内容，需要经过语义解释，数据与其语义是不可分的。

1.2.2　数据库

顾名思义，数据库是存放数据的仓库。只不过这个仓库是在计算机存储设备上，而且数据是按照一定的格式存放的。

在科学技术飞速发展的今天，人们的视野越来越广，数据量急剧增加。过去人们把数据存放在文件柜里，现在人们借助计算机和数据库技术科学地保存和管理大量复杂的数据，以便能方便而充分地利用这些宝贵的信息资源。

数据库是一个长期存储在计算机内、有组织的、可共享的、统一管理的数据集合。数据库中的数据按一定的数据模型组织、描述和储存，具有较小的冗余度、较高的数据独立性和易扩展性，可以为各种用户共享。

数据的长期存储、有组织和可共享是数据库的三个基本特点。

1.2.3 数据库管理系统

数据库管理系统(DBMS)是操纵和管理数据库的软件系统，它由一组计算机程序构成，管理并控制数据资源的使用。在计算机软件系统的体系结构中，数据库管理系统位于用户和操作系统之间。

数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制。数据库管理系统使用户能方便地定义数据和操纵数据，并能够保证数据的安全性、完整性，以及多用户对数据的并发使用及发生故障后的系统恢复。

1.2.4 数据库系统

数据库系统是指在计算机系统中引入数据库后的系统。数据库系统不但能够实现有组织地、动态地存储大量相关的数据，而且为数据处理和信息资源共享提供了便利条件。数据库系统主要由四部分组成：数据库、软件系统、硬件系统和用户。

(1) 数据库

在一个数据库系统中，可以根据需要创建多个数据库，并且数据库中的数据通常可以被多个用户所共享。

(2) 软件系统

软件系统主要包括数据库管理系统及其开发工具、操作系统(OS)和应用系统等。数据库系统的核心软件是数据库管理系统 DBMS 和数据库应用系统 DBAS(Database Application System)。

(3) 硬件系统

硬件系统是指运行数据库系统所需要的硬件设备。任何一个计算机系统都需要有中央处理器、存储器和输入/输出设备等硬件。一个数据库系统需要有足够快的处理器来处理数据，以便快速响应用户的数据处理和数据检索请求，同时还需要有足够容量的内存与外存来存储大量的数据。对于分布式数据库系统，还需要有网络通信设备的支持。

(4) 用户

用户是指使用和管理数据库的人，他们可以对数据库进行存储、维护和检索等操作。数据库系统中用户可分为三类：终端用户、应用程序员和数据库管理员。

终端用户主要是使用数据库的各级管理人员、工程技术人员、科研人员，一般为非计算机专业人员。应用程序员负责为终端用户设计和编制应用程序，以便终端用户对数据库进行存/取操作。

数据库管理员(Database Administrator，DBA)是指对数据库进行设计、维护和管理的专门人员。DBA 是数据库所属单位的代表。一个单位决定开发一个数据库系统时，首先就应确定 DBA 的人选。DBA 不仅应当熟悉系统软件，还应熟悉本单位的业务工作。DBA 应自始至终参加整个数据库系统的研制开发工作，开发成功后，DBA 将全面负责数据库系统的管理、维护和正常使用。其主要职责如下：

(1) 参与数据库设计的全过程，决定数据库的结构和内容。

(2) 定义数据的安全性和完整性，负责分配用户对数据库的使用权限和口令管理。

(3) 监督控制数据库的使用和运行，改进和重新构造数据库系统。当数据库受到破坏时，应负责恢复数据库；当数据库的结构需要改变时，完成对数据结构的修改。

因此，DBA 不仅要有较高的技术水平，还应具有了解和阐明管理要求的能力。特别对于大型数据库系统，DBA 极为重要。对于常见的微机数据库系统，通常只有一个用户，常常不设 DBA，DBA 的职责由应用程序员或终端用户代替。

在不引起混淆的情况下，数据库系统可以简称为数据库。

数据库系统的组成如图1.4所示。

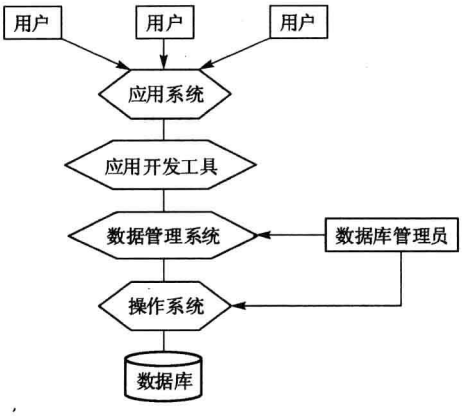

图 1.4　数据库系统的组成

1.3　数据模型

数据库是某个企业、组织或部门所涉及的数据的综合，它不仅要反映数据本身的内容，而且要反映数据之间的联系。由于计算机不可能直接处理现实世界中的具体事物，所以人们必须事先把具体事物转换成计算机能够处理的数据。也就是首先要数字化，把现实世界中具体的人、物、活动、概念用数据模型这个工具来抽象、表示和处理。通俗地讲，数据模型是现实世界的模拟。现有的数据库系统都是基于某种数据模型的。

数据模型应满足三个方面的要求：一是能比较真实地模拟现实世界；二是容易为人所理解；三是便于在计算机上实现。一种数据模型要很好地满足这三方面的要求，在目前尚很困难。在数据库系统中应针对不同的使用对象和应用目的，采用不同的数据模型。数据模型是数据库系统的核心和基础。

1.3.1　数据模型的组成要素

一般来讲，任何一种数据模型都是严格定义的概念的集合。这些概念必须能精确描述系统的静态特性、动态特性和完整性约束条件。因此，数据模型通常都是由数据结构、数据操作和完整性约束三个要素组成的。

1．数据结构

数据结构是所研究的对象类型的集合。这些对象是数据库的组成成分，包括两类：一类是与数据类型、内容、性质有关的对象，如网状模型中的数据项、记录，关系模型中的域、属性、关系等；另一类是与数据之间联系有关的对象，如网状模型中的系型(Set Type)。

数据结构是刻画一个数据模型性质最重要的方面。因此在数据库系统中，人们通常按照其数据结构的类型来命名数据模型。例如，层次结构、网状结构和关系结构的数据模型分别命名为层次模型、网状模型和关系模型。数据结构是对系统静态特性的描述。

2．数据操作

数据操作是指一组用于指定数据结构的任何有效的操作。数据库中的主要操作有查询和更新两大类。数据模型要给出这些操作确切的含义、操作规则和实现操作的语言，因此，数据操作规定了数据模型的动态特性。

3．完整性约束

数据的约束条件是一组完整性规则的集合，它定义了给定数据模型中数据及其联系所具有的制约和依存规则，用来限定相容的数据库状态的集合和可容许的状态改变，以保证数据库中数据的正确性、有效性和相容性。每种数据模型都规定有通用的和特殊的完整性约束条件。

(1) 通用的完整性约束条件。通常把具有普遍性的问题归纳成一组通用的约束规则，只有在满足给定约束规则的条件下才允许对数据库进行更新操作。例如，关系模型中通用的约束规则是实体完整性和参照完整性。

此为试读，需要完整PDF请访问：www.ertongbook.com