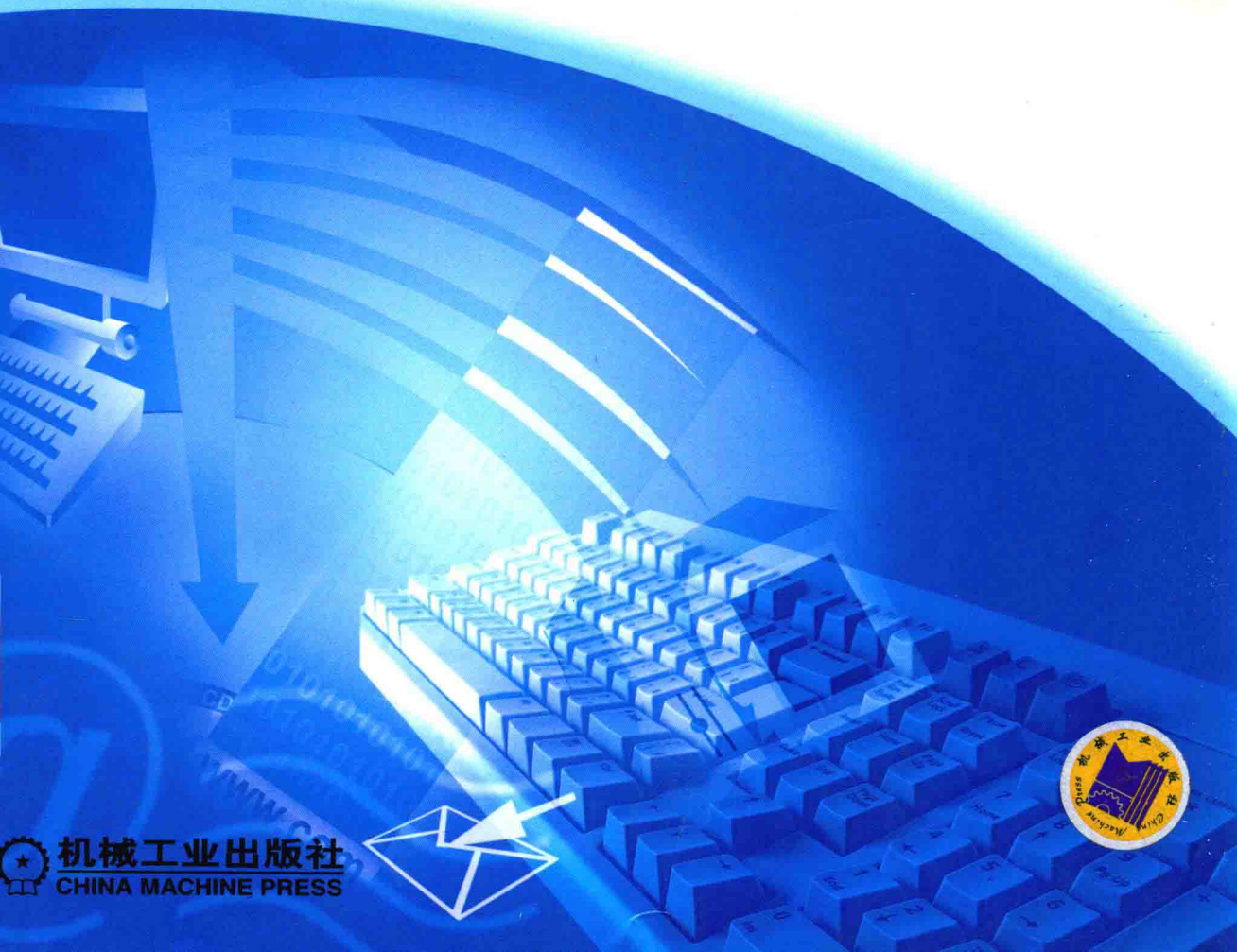




普通高等教育规划教材

# 数据库原理及应用

靳敏 主编



机械工业出版社  
CHINA MACHINE PRESS



7/11.13

105

普通高等教育应用型人才培养规划教材

普通高等教育规划教材

数据库原理及应用

# 数据库原理及应用

主编 靳 敏

副主编 夏素霞 张益星

参 编 周 屹

主 审 李 斌

陈志强 华北航天工业学院

陈 峻 扬州大学

苏 群 黑龙江科技学院

曹 群 哈尔滨工程大学

曹 群 哈尔滨工程大学

曹 群 哈尔滨工程大学

江苏工业学院图书馆  
藏书章

中国标准书号 (2003) 第 134144 号

机械工业出版社 (北京) 出版

责任编辑: 王小求  
封面设计: 王 燕

北京中印书局印刷

2004年5月第1版

787mm×1092mm

25.00元



机械工业出版社

01.11.1975  
201

本书系统地阐述了数据库系统的基本概念、基本原理，并结合 PowerBuilder 数据库开发工具和具体的应用实例，详细讲述了数据库系统的开发过程。全书共分 8 章。第 1 章主要介绍数据库的基本概念、基本原理；第 2 章~第 4 章介绍当前最流行、应用最广泛、理论研究最多的关系数据库，对其基本概念、关系代数理论、查询语言、规范化理论、查询优化等作了充分介绍；第 5 章介绍数据库保护技术；第 6 章和第 7 章主要介绍 PB 和 SQL Server 数据库应用系统开发与设计；第 8 章介绍数据库系统的研究与新发展。

本书理论与应用并重，概念清楚，系统性强，内容丰富，力图反映数据库领域的最新研究成果。

本书可作为普通高等教育应用型本科院校计算机各专业数据库的教材和教学参考书，也可供从事数据库系统开发的人员及其他有关人员参考。

沁 風 齋 參  
藏 本 庫 主

### 图书在版编目 (CIP) 数据

数据库原理及应用/靳敏主编. —北京: 机械工业出版社, 2004.1

普通高等教育规划教材

ISBN 7-111-13764-7

I. 数… II. 靳… 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 (2003) 第 124144 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: 王小东 版式设计: 霍永明 责任校对: 姚培新  
王玉鑫

封面设计: 饶 薇 责任印制: 闫 焱

北京京丰印刷厂印刷·新华书店北京发行所发行

2004 年 2 月第 1 版·第 1 次印刷

787mm×1092mm<sup>1</sup>/<sub>16</sub>·17.75 印张·437 千字

定价: 25.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换  
本社购书热线电话 (010) 68993821、88379646

封面无防伪标均为盗版





### 2. 突出特色，强化应用

围绕培养目标，以工程应用为背景，通过理论与工程实际相结合，构建工程应用型本科教育系列教材特色。本套系列教材的内容、结构遵循如下九字方针：知识新、结构新、重应用。教材内容的要求概括为：“精”、“新”、“广”、“用”。“精”指在融会贯通教学内容的基础上，挑选出最基本的内容、方法及典型应用；“新”指在将本学科前沿的新进展和有关的技术进步新成果、新应用等纳入教学内容，以适应科学技术发展的需要。妥善处理好传统内容的继承与现代内容的引进。用现代的思想、观点和方法重新认识基础内容和引入现代科技的新内容，并将这些按新的教学系统重新组织；“广”指在保持本学科基本体系下，处理好与相邻以及交叉学科的关系；“用”指注重理论与实际融会贯通，特别是注入工程意识，包括经济、质量、环境等诸多因素对工程的影响。

### 3. 抓住重点，合理配套

工程应用型本科教育系列教材的重点是专业课（专业基础课、专业课）教材的建设，并做好与理论课教材建设同步的实践教材的建设，力争做好与之配套的电子教材的建设。

### 4. 精选编者，确保质量

遴选一批既具有丰富的工程实践经验，又具有丰富的教学实践经验的教师担任编写任务，以确保教材质量。

我们相信，本套系列教材的出版，对我国工程应用型人才培养质量的提高，必将产生积极作用，会为我国经济建设和社会发展作出一定的贡献。

机械工业出版社颇具魄力和眼光，高瞻远瞩，及时提出并组织编写这套系列教材，他们为编好这套系列教材做了认真细致的工作，并为该套系列教材的出版提供了许多有利的条件，在此深表衷心感谢！

编委会主任 刘国荣教授  
湖南工程学院院长

# 前 言

数据库技术自 20 世纪 60 年代后期出现以后,一直活跃在数据处理领域。二十多年来,数据库技术得到了迅速发展,已经成为计算机科学的一个重要分支。由于数据库技术已经成为开发各种信息系统的重要工具,它的推广使用,使得计算机应用深入到工农业生产、商业管理、办公自动化、科学研究和工程技术的各个领域。随着社会各个领域的迫切要求和广泛应用,人们对数据库技术重要性的认识越来越明确,对它的应用和学习要求也越来越迫切。

本书是编者在多年从事数据库系统教学和科研的基础上编写而成的。在编写过程中,本着由易到难、循序渐进、理论与实践并重、突出特色的原则,力求做到概念清楚、深入浅出、理论联系实际和面向应用。全书结构合理,选材精练,重点突出应用最广泛的关系数据库的设计,对它的概念、原理进行了较为深入的讨论。为了突出实际应用,结合 PowerBuilder 数据库开发工具和具体的应用实例,详细讲述了数据库系统的开发过程,以使学生在学习过程中能更好地把理论与实际结合起来,从而达到更好的学习效果。另外,在内容选材上,我们还注意尽量反映这一领域的新方法、新技术,以使学生能对数据库领域的前沿动态能有初步的了解。

全书共分 8 章。第 1 章主要介绍数据库的基本概念、基本原理;第 2 章至第 4 章介绍当前最流行、应用最广泛、理论研究最多的关系数据库,对其基本概念、关系代数理论、查询语言、规范化理论、查询优化等作了充分介绍;第 5 章介绍数据库保护技术;第 6 章和第 7 章介绍 PB 和 SQL Server 数据库应用系统开发与设计;第 8 章介绍数据库系统的研究与新发展,力图反映数据库领域的最新研究成果。

本书由靳敏任主编,由夏素霞、张益星任副主编。第 1、2、7 章由靳敏编写,第 3、6 章由夏素霞编写,第 4、5、8 章由张益星编写,周屹参与编写了第 7 章部分内容。全书由靳敏统稿和定稿。

本书由李斌主审。在本书的编审过程中,得到了许多同行的大力支持,提出了许多宝贵意见,在此表示深深的谢意。

由于时间仓促及作者水平有限,错误和不当之处在所难免,敬请广大读者批评指正。

编 者

2003 年 10 月

# 目 录

序	
前言	
<b>第 1 章 数据库系统导论</b> .....	1
1.1 引言 .....	1
1.1.1 数据与数据处理 .....	1
1.1.2 数据管理技术的发展过程 .....	2
1.1.3 数据库的概念 .....	3
1.1.4 数据库系统的主要特征 .....	6
1.2 现实世界的描述 .....	9
1.2.1 现实世界 .....	9
1.2.2 信息世界 .....	10
1.2.3 计算机世界 .....	10
1.2.4 常用术语的说明 .....	10
1.3 数据模型 .....	11
1.3.1 层次模型 .....	11
1.3.2 网状模型 .....	12
1.3.3 关系模型 .....	13
1.3.4 对传统数据模型的评价 .....	14
1.3.5 E-R 数据模型 .....	17
1.4 数据库系统结构 .....	20
1.4.1 数据库系统结构的一些概念 .....	20
1.4.2 数据库系统的组成 .....	22
1.4.3 数据库管理系统 .....	24
1.4.4 数据库语言 .....	26
1.4.5 数据库管理员 .....	29
1.4.6 数据字典 .....	30
1.4.7 数据库系统的工作流程 .....	32
习题 .....	33
<b>第 2 章 关系数据库及查询优化</b> .....	34
2.1 关系模型的基本概念 .....	34
2.1.1 笛卡儿积与关系定义 .....	34
2.1.2 关系的一些基本概念 .....	35
2.2 关系模型及其描述 .....	36
2.2.1 关系模型概述 .....	36
2.2.2 关系模式 .....	37
2.2.3 关系模型的完整性规则 .....	39
2.3 关系代数 .....	39
2.3.1 传统的集合运算 .....	39
2.3.2 专门的关系运算 .....	40
2.4 关系演算语言 .....	44
2.4.1 元组关系演算语言 ALPHA .....	45
2.4.2 域关系演算语言 QBE .....	48
2.5 查询优化概述及一般策略 .....	50
2.5.1 查询优化概述 .....	50
2.5.2 查询优化的一般策略 .....	51
2.6 SQL Server 数据库 .....	52
2.6.1 SQL Server 特点 .....	52
2.6.2 SQL Server 的数据库管理 .....	52
2.6.3 数据完整性 .....	61
习题 .....	63
<b>第 3 章 关系数据库标准语言——SQL</b> .....	64
3.1 SQL 语言概述 .....	64
3.1.1 SQL 的特点 .....	64
3.1.2 基本概念 .....	65
3.1.3 SQL 语言的主要组成 .....	65
3.2 Transact—SQL 的附加语言元素 .....	66
3.2.1 Transact—SQL 基本元素简介 .....	66
3.2.2 Transact—SQL 的几个常用语句 .....	70
3.2.3 Transact—SQL 的执行环境 .....	71
3.3 Transcat—SQL 语言的数据定义功能 .....	73
3.3.1 数据库的创建 .....	73
3.3.2 基本表的创建 .....	78
3.3.3 索引的建立和撤消 .....	84
3.4 SQL 语言数据操纵功能 .....	89
3.4.1 SQL 数据查询 .....	90
3.4.2 SQL 数据插入、删除和修改 .....	98

3.5 视图 .....	102	4.3.2 模式分解 .....	140
3.5.1 理解视图 .....	102	4.3.3 保持函数依赖性的分解 .....	141
3.5.2 视图的定义和删除 .....	103	4.3.4 无损连接分解 .....	141
3.5.3 视图的查询与更新 .....	105	4.4 数据库设计 .....	142
3.6 存储过程技术 .....	106	4.4.1 数据库设计的任务、内容和特点 .....	142
3.6.1 存储过程的概念 .....	106	4.4.2 数据库设计的步骤 .....	143
3.6.2 存储过程的分类 .....	107	4.4.3 需求分析 .....	144
3.6.3 存储过程的创建 .....	108	4.4.4 概念结构设计 .....	148
3.6.4 删除存储过程 .....	110	4.4.5 逻辑结构设计 .....	156
3.7 触发器 .....	111	4.4.6 数据库物理设计 .....	157
3.7.1 触发器的概念 .....	111	4.4.7 数据库实施 .....	160
3.7.2 创建和删除触发器 .....	111	4.4.8 数据库运行和维护 .....	162
3.7.3 使用触发器的一点建议 .....	115	习题 .....	163
3.8 数据库的完整性 .....	115	<b>第5章 数据库的并发控制与恢复</b> .....	164
3.8.1 完整性的概念 .....	115	5.1 数据库的并发控制 .....	164
3.8.2 完整性控制的实现方式 .....	116	5.1.1 事务及并发控制的基本概念 .....	164
3.9 数据库的安全性 .....	120	5.1.2 并发操作与数据的不一致性 .....	165
3.9.1 安全性的概念 .....	120	5.1.3 封锁及封锁协议 .....	167
3.9.2 数据库的安全性措施 .....	120	5.1.4 死锁和活锁 .....	170
3.10 嵌入式 SQL .....	122	5.1.5 并发调度的可串行性 .....	172
3.10.1 嵌入式 SQL 语句与主语言之间的通信 .....	123	5.1.6 两段锁协议 .....	173
3.10.2 不用游标的 SQL 语句 .....	123	5.1.7 基于时间标志的并发控制 .....	173
3.10.3 使用游标的 SQL 语句 .....	125	5.2 SQL Server 的并发控制策略 .....	175
3.10.4 动态 SQL 语句 .....	127	5.3 数据库的恢复 .....	177
习题 .....	129	5.3.1 故障的种类 .....	178
<b>第4章 关系数据库理论</b> .....	131	5.3.2 数据恢复的实现技术 .....	178
4.1 基本概念 .....	131	5.3.3 数据库恢复策略 .....	180
4.1.1 函数依赖 .....	131	5.3.4 具有检查点的数据恢复技术 .....	182
4.1.2 平凡函数依赖和非平凡函数依赖 .....	132	5.4 SQL Server 的恢复策略 .....	183
4.1.3 完全函数依赖与部分函数依赖 .....	132	5.4.1 SQL Server 的备份类型 .....	183
4.1.4 传递函数依赖 .....	132	5.4.2 SQL Server 备份的操作方式 .....	184
4.1.5 码 .....	132	5.4.3 SQL Server 的数据备份或恢复策略 .....	184
4.2 关系模式的规范化 .....	133	习题 .....	185
4.2.1 规范化理论的主要内容 .....	133	<b>第6章 Power Builder 8.0 基础</b> .....	186
4.2.2 范式 .....	133	6.1 Power Script 语言 .....	186
4.2.3 多值依赖与第四范式(4NF) .....	137	6.1.1 Power Script 语言基础 .....	186
4.2.4 关系模式的规范化 .....	139	6.1.2 数据类型 .....	187
4.3 模式分解 .....	140	6.1.3 变量、常量和表达式 .....	188
4.3.1 函数依赖集的投影 .....	140	6.1.4 过程控制语句 .....	189



6.2 Power Builder 8.0 应用程序设计入门 ..... 190

6.3 菜单和 MDI 的应用程序的设计 ..... 191

6.4 常用控件的使用 ..... 195

6.5 数据库的建立与应用 ..... 199

6.6 游标的使用 ..... 203

6.7 用嵌入式 SQL 操作数据库 ..... 204

6.8 多表更新 ..... 205

6.9 数据更新及自动配置 ODBC 的方法 ..... 207

6.10 通用查询程序 ..... 209

习题 ..... 214

**第 7 章 PB 数据库应用系统设计实例——超市仓库管理系统 ..... 215**

7.1 用户需求 ..... 215

7.1.1 超市仓库管理系统及 POS 机的简介 ..... 215

7.1.2 超市仓库管理系统及 POS 机的用户需求 ..... 215

7.2 超市仓库管理系统及 POS 机的分析和设计 ..... 216

7.2.1 超市仓库管理系统及 POS 机的分析和概要设计 ..... 216

7.2.2 超市仓库管理系统及 POS 机数据库设计 ..... 216

7.3 超市仓库管理系统及 POS 机的具体实现 ..... 220

7.3.1 后台数据库的设计 ..... 220

7.3.2 系统模块的具体实现 ..... 220

7.3.3 收银机的设计 ..... 232

7.3.4 系统的实现 ..... 234

7.4 主要原码清单 ..... 235

7.4.1 仓库管理系统 ..... 235

7.4.2 收银机 ..... 246

**第 8 章 数据库系统的研究与新进展 ..... 254**

8.1 分布式数据库系统 ..... 254

8.1.1 分布式数据库系统概述 ..... 254

8.1.2 分布式数据库系统的体系结构 ..... 255

8.1.3 分布透明性 ..... 257

8.1.4 分布式数据库系统 (DDBS) 的组成 ..... 258

8.1.5 分布式数据库的应用 ..... 258

8.2 面向对象的数据库系统 ..... 258

8.2.1 面向对象的数据模型 ..... 259

8.2.2 类和类层次 ..... 261

8.2.3 面向对象的数据库语言 ..... 263

8.2.4 对象关系数据库管理系统 ..... 263

8.3 其他新型数据库系统 ..... 264

8.3.1 演绎数据库系统 ..... 264

8.3.2 模糊数据库系统 ..... 266

8.3.3 并行数据库 ..... 267

8.3.4 主动数据库 ..... 269

8.3.5 多媒体数据库 ..... 270

8.3.6 工程数据库 ..... 270

习题 ..... 271

**参考文献 ..... 272**

# 第 1 章 数据库系统导论

数据库是计算机科学的一个重要研究领域，是专门研究数据管理的技术，从 20 世纪 60 年代末开始，经过 30 多年的发展，已成为计算机软件学科的一个重要分支。本章从数据管理技术出发，介绍数据库系统及相关名词术语。

## 1.1 引言

### 1.1.1 数据与数据处理

计算机的出现，将数据处理带入了一个新的时代。数据处理的基本问题是数据的组织、存储、检索、维护及加工利用，这些正是数据库系统所要研究解决的问题。

#### 1. 信息与数据

数据是数据库系统研究和处理的对象。数据又离不开信息，它们既有联系又有区别。

(1) 信息 信息是现实世界中各种事物（包括有生命的和无生命的、有形的和无形的）的存在方式、运动形态以及它们之间的相互联系等诸要素在人脑中的反映，通过人脑的抽象后形成的概念。这些概念不仅被人们认识和理解，而且人们可以对它进行推理、加工和传播。

(2) 数据 数据一般是指信息的一种符号化表示方法，就是说用一定的符号表示信息，而采用什么符号，完全是人为规定。例如，为了便于用计算机处理信息，就得把信息转换为计算机能够识别的符号，即采用 0 和 1 两个符号编码来表示各种各样的信息。所以数据的概念包括两个方面的含义：一是数据的内容是信息，二是数据的表现形式是符号。

数据在数据处理领域中涵盖的内容非常广泛，这里的“符号”不仅仅指数字、字母、文字等常见符号，它还包括图形、图像、声音等多媒体数据。

(3) 信息与数据的关系 信息与数据的关系是既有联系又有区别。数据是承载信息的物理符号或称之为载体，而信息是数据的内涵。二者的区别是：数据可以表示信息，但不是任何数据都能表示信息，同一数据也可以有不同的解释。正如人们常说的“如果给计算机输入的是垃圾，输出的也将是垃圾”。信息是抽象的，同一信息可以有不同的数据表示方式。例如，在足球世界杯期间，同一场比赛的新闻，可以在报纸上以文字形式、在电台以声音形式、在电视上以图像形式来表现。

#### 2. 数据处理

数据处理是指将数据转换成信息的过程，这一过程主要是指对所输入的数据进行加工整理，包括对数据的收集、存储、加工、分类、检索和传播等一系列活动，其根本目的就是大量的、已知的数据出发，根据事物之间的固有联系和运动规律，采用分析、推理、归纳等手段，提取出对人们有价值、有意义的信息，作为某种决策的依据。

我们可以用如图 1-1 所示的过程简单地表示出信息与数据之间的关系。

图 1-1 中数据是输入，而信息是输出结果。人们有时说的“信息处理”，其真正含义应

该是为了产生信息而处理数据。例如，学生的“出生日期”是有生以来不可改变的基本特征之一，属于原始数据，而“年龄”是当年与出生日期相减而得到的数字，具有相对性，可视为二次数据。同样道理，“参加工作时间”、产品的“购置日期”是职工和产品的原始数据，工龄、产品的报废日期则是经过简单计算所得到的结果。

在数据处理活动中，计算过程相对比较简单，很少涉及复杂的数学模型，但是却有数据量大，且数据之间有着复杂的逻辑联系的特点。因此数据

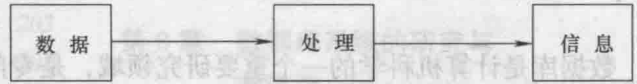


图 1-1 信息与数据之间的关系

处理任务的矛盾焦点不是计算，而是把数据管理好。数据管理是指数据的收集、整理、组织、存储、查询、维护和传送等各种操作，是数据处理的基本环节，是任何数据处理任务必有的共性部分。因此，对数据管理应当加以突出，集中精力开发出通用而又方便实用的软件，把数据有效地管理起来，以便最大限度地减轻计算机软件用户的负担。数据库技术正是瞄准这一目标而逐渐完善起来的一门计算机软件技术。

### 1.1.2 数据管理技术的发展过程

数据以及数据之间的关系可以从逻辑和物理两方面进行描述和组织。数据的逻辑描述是指按用户要求的数据间逻辑关系来组织和表达数据的，而物理描述则将涉及数据在计算机内的存储方式，是以存储数据的时空概念来描述数据间的关系的，前者可称为数据的逻辑结构，后者则称为数据的物理结构。

如何用计算机对数据进行组织、编目、定位、存储、检索和维护是数据管理的任务。数据管理是数据处理的中心问题。数据管理技术与数据处理方式有密切的关系，并且直接影响到数据处理的效率。在数据处理的手工处理阶段和机械处理阶段，数据管理技术也是手工的。但是，在计算机进入数据处理领域之后，原来的那套手工管理方式就不能适应计算机自动处理数据的需要了。而且，在跨入信息社会的今天，需要管理的数据越来越多，如果仍然沿用原来的那套方式，不仅需要耗费大量的人力，而且很难使这些数据发挥应有的作用。为此，许多计算机专家，特别是软件工作者，对数据管理技术做了大量的研究工作，使数据管理技术经历了一个由简单到复杂的不断完善的发展过程。

#### 1. 人工管理阶段（20 世纪 50 年代中期以前）

在这一阶段中，计算机除了硬件设备之外没有任何软件可供使用，放在用户面前的只有硬件的裸机。用户在这种裸机上进行数据处理，除了编制自己的课题程序外，还必须考虑数据的逻辑定义和组织，以及数据在计算机存储设备内的物理存储方式和地址。这时，数据的引用是按物理地址进行的，表示处理流程的程序和作为处理对象的数据，相互结合成一个整体。数据的管理、基本上是手工的、分散的，计算机还没有在数据管理中发挥应有的作用，因此严重地影响了计算机的使用效率。这一时期数据与程序之间的关系如图 1-2 所示。



图 1-2 人工管理阶段数据和程序之间的关系

#### 2. 文件系统阶段（20 世纪 50 年代后期至 20 世纪 60 年代中期）

把数据组织成文件的形式，是计

计算机数据管理的重大进展，也是计算机在数据管理中直接发挥作用的开始。众所周知，无组织的一大堆数据是毫无意义的，也是没有实用价值的。软件将数据按照一定的规则组织起来，成为一个有效的数据组合体，赋予它一个名字，称为文件名或文件标识，用户通过文件名来访问该文件。这时用户就不必过多地考虑物理细节了，而是由文件系统充当应用程序和数据文件之间的接口，这一时期应用程序和数据的关系如图 1-3 所示。

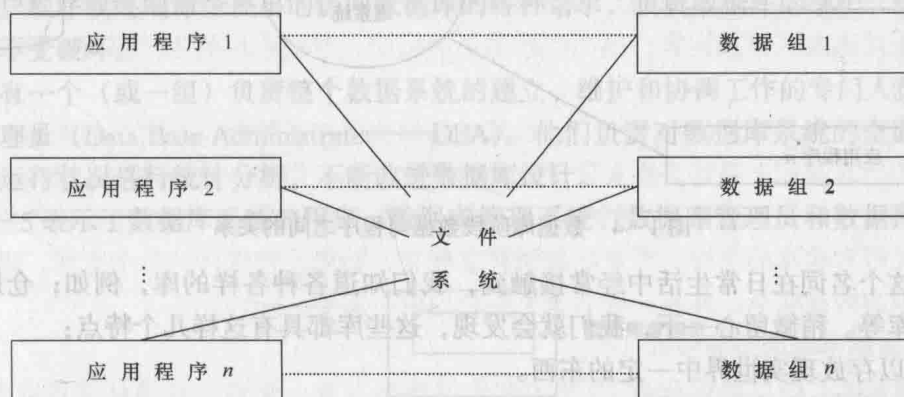


图 1-3 文件系统阶段数据和程序之间的关系

虽然这一阶段较人工管理阶段有了很大的改进，但仍存在明显的弱点，即文件本身基本上还是对应于一个或几个应用程序，或者说数据还是面向应用的。从逻辑结构上讲，还与前一阶段没有什么变化，尽管程序不必要直接与文件打交道而有软件做为接口，但它仍然是一个不具有弹性的无结构的信息集合，存在着冗余度大、空间浪费、文件不易扩充等缺点，从而反映不了现实世界事物之间的广泛的内在联系。

### 3. 数据库系统阶段 (20 世纪 60 年代后期开始)

在这一时期，磁盘技术已经成熟，并作为主要外存而广泛使用。计算机硬件的价格大幅度下降，可靠性增强，为数据管理技术的发展奠定了物质基础。另外，计算机用于管理，其规模更加庞大，从而使数据量急剧增加。对数据进行集中控制，充分提供数据共享的要求日益迫切。

在这样的背景下产生了一种新的数据管理技术即数据库技术。数据库技术克服了以前所有管理方式的缺点，试图提供一种完善的、更高级的数据管理方式。它的基本思想是解决多用户数据共享的问题，实现对数据的集中统一管理，具有较高的数据独立性，并为数据提供各种保护措施。这一时期，数据库管理软件作为用户与数据的接口，程序和数据的关系如图 1-4 所示。

#### 1.1.3 数据库的概念

数据库系统是计算机软件的一个重要分支，是近 20 年来发展起来的一门新兴学科，它和计算机网络、人工智能被称为当今计算机技术界的三大热门技术。目前，虽有人对数据库的设计原则和方法进行总结和探讨，使之通用化、标准化和理论化。但总的说来，它的概念、原理和方法还处于从工程实践向理论过渡的阶段，它的概念、原理和方法还在继续变化和发展。另外，数据库是一个很复杂的系统，涉及面很广，难以用简练的语言准确地概括其全部特征。因此，给它下一个确切的定义是困难的，所以我们先从简单分析入手，逐步认识什么是数据库。



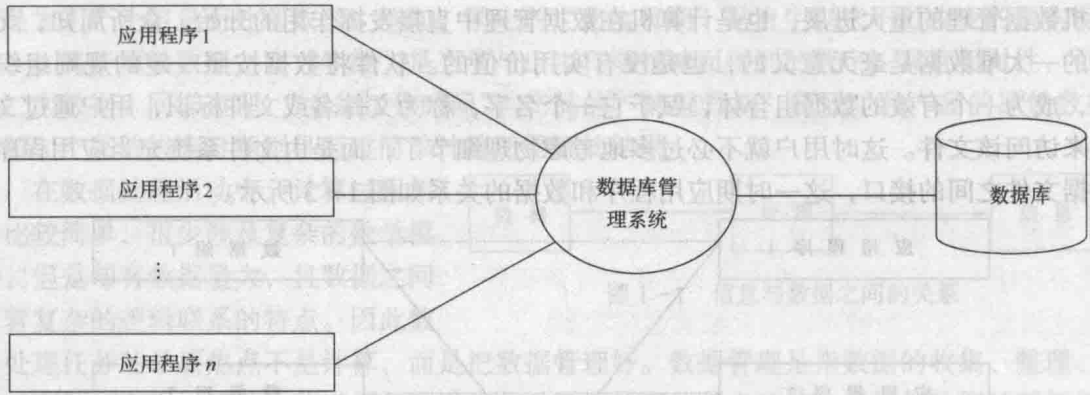


图 1-4 数据库阶段数据与程序之间的关系

库，这个名词在日常生活中经常接触到，我们知道各种各样的库，例如；仓库、书库、金库、血库等。稍微留心一下，我们会发现，这些库都具有这样几个特点：

- 1) 可以存放现实世界中一定的东西。
- 2) 这些东西是有条理有组织地存放在库中的。
- 3) 库中这些东西由专人来进行管理。

例如，书库中的书就是有次序地摆在书架上的，并由图书管理员对它们进行管理。显然，库中的东西如果不是按照一定的次序存放，并由专人管理，那么就不能有效地利用库中的空间，也不可能方便地找出所需要的东西。

顾名思义，数据库就是存储数据的“仓库”。但它和我们前面讲的库是有所不同的。首先，数据不是存放在容器或空间中，而是存放在计算机的外存储器上（如磁盘），并且是有组织地存放的。数据的管理和利用是通过计算机的数据管理软件——数据库管理系统来完成的。因此，我们讲的数据库，不单是指存有数据的计算机外存，而是指存放在外存上的数据集合以及管理它们的计算机软件的总和，通常称为数据库系统。

目前，对于什么是数据库还没有一个统一的、公认的定义。比较流行的有关数据库的文献中常见的数据库定义有如下三种：

(1) DBTG 的定义 数据库是由一个特定的模式控制的所有记录、系和域组成的。如果有多个数据库，则每一个数据库必须有自己的模式。并假定不同数据库的内容是彼此无关的。

(2) C.J.Date 的定义 它从三个方面描述数据库：①存储在磁鼓、磁盘或其他外存介质上的数据集合——这是指数据库自身；②存在以这种数据为背景而运行的若干个应用程序，可进行检索、修改、插入和删除等操作，另外可能有一些联机用户，利用远程终端与数据库相互作用；③数据库是集成的，即包括许多用户的数据；每个用户只享用其中的一小部分，且不同用户使用的部分以多种方式重叠，也就是单独的数据片能够被许多不同用户所共享。

(3) J.Martin 的定义 数据库是存储在一起的相关数据的集合，这些数据去掉了有害的或不必要的冗余，为多种应用服务，数据的存储独立于使用它的程序；对数据库插入新数据，修改和检索原有数据均能按一种公用的和可控制方式进行；数据被结构化，为今后的应用研究提供基础。当某个系统中存在结构上完全分开的若干数据库时，则说该系统包含一个数据库集合。

综合上述三种定义，通常一个数据库系统包含下列内容：

(1) 有一个结构化的相关数据的集合。在这个数据集中没有有害的或不必要的冗余，能够为多种应用服务，它独立于应用程序而存在。这种结构化的数据集合就是数据库本身，是数据库系统的核心和管理对象。

(2) 有一个负责数据库管理和维护的软件系统，称为数据库管理系统 (Data Base Management System——DBMS)。它对数据库中数据的各种操作，提供一种公用的方法。它接受并完成用户程序或终端命令提出的访问数据库的各种请求，负责数据库的维护，保护数据库中的数据不受破坏。

(3) 有一个 (或一组) 负责整个数据系统的建立、维护和协调工作的专门人员，这就是数据库管理员 (Data Base Administrator——DBA)。他们负责对数据库系统的全面管理与监督，并对运行状况进行统计分析，不断改进数据库设计。

图 1-5 表示了数据库系统的用户、数据库管理系统、数据库管理员和数据库之间的相互关系。

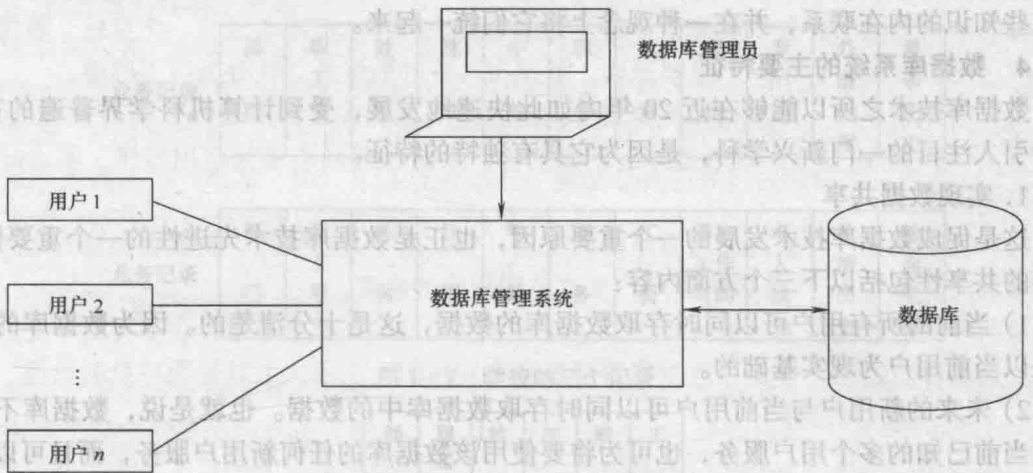


图 1-5 数据库系统

数据库保存的是属于企业和事业部门、团体和个人的有关数据，例如，银行账目数据、学校的教学管理数据、政府部门的管理、统计和计划数据等。特别应该强调的是，设计数据库保存这些数据的目的是，不仅仅是为了扩展人们的记忆，而主要是帮助人们去控制与之相关的事物。因此，数据库往往不是孤立存在的，通常是一个更大的信息系统的一部分，两者的关系如图 1-6 所示。该图表示，人们从观测客观事物中得到大量信息，对这些信息进行记录、整理和归类 (总称规范)，然后将规范信息数据化，送入数据库中保存起来，其中一部分信息可能直接送入控制决策机构。另一方面，控制决策机构 (它既可由一

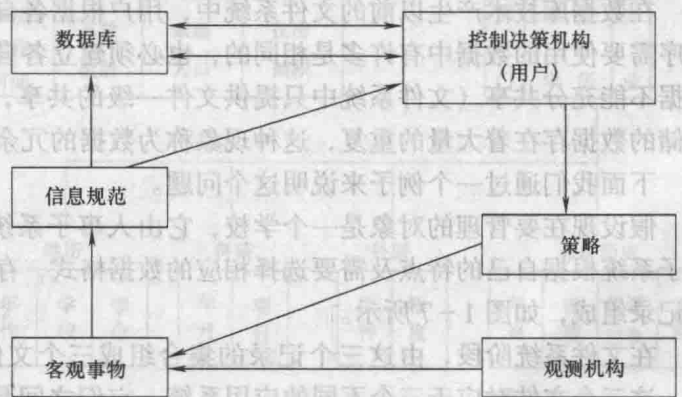


图 1-6 信息控制系统

些人组成,又可是一个自动控制系统)向数据库发出询问,并利用数据库响应后提供的信息(以及其他有关信息)做出决策,再控制客观事物。例如,就一个企业的产品销售信息系统而言,客观事物可以是产品的销售情况和用户对该产品的反映情况等,控制决策机构指企业领导机关,策略是指提高产品质量扩大产品销售的措施。

当数据库的管理人员把观测客观事物得到的信息规范化、数据化送入数据库后,企业领导机关可以通过询问数据库得到产品的销售情况和用户对产品的反映,根据这些情况,企业领导就可以提出改进产品销售,提高产品质量的措施,并在今后的生产经营中付诸实行(控制客观事物)。

数据库系统作为软件的一个分支,与其他基础软件和系统软件有密切的关系,它几乎涉及到软件的所有知识,是许多重要软件技术的综合应用。例如,数据结构、操作系统、编译技术、程序设计等知识在数据库都将用到。所以数据库系统是一门综合性的软件技术,是一门很有意义很有趣味的学科。要更好地研究掌握它,必须了解软件的各个方面,以便加深理解这些知识的内在联系,并在一种观念上将它们统一起来。

#### 1.1.4 数据库系统的主要特征

数据库技术之所以能够在近 20 年内如此快速地发展,受到计算机科学界普遍的重视,成为引人注目的一门新兴学科,是因为它具有独特的特征。

##### 1. 实现数据共享

这是促成数据库技术发展的一个重要原因,也正是数据库技术先进性的一个重要体现。数据的共享性包括以下三个方面内容:

1) 当前的所有用户可以同时存取数据库的数据,这是十分清楚的。因为数据库的研制正是以当前用户为现实基础的。

2) 未来的新用户与当前用户可以同时存取数据库中的数据。也就是说,数据库不仅可以为当前已知的多个用户服务,也可为将要使用该数据库的任何新用户服务,而且可以同时为多个新用户服务。

3) 用户可以通过多种程序设计语言或命令语言与数据库接口。如 COBOL、PL/1、FORTRAN 以及终端命令和数据库管理系统提供的数据操纵语言等,从而可以用多种语言使用数据库。

##### 2. 减少数据的冗余度

在数据库技术产生以前的文件系统中,用户根据各自的应用建立文件,即使不同的应用程序需要使用的数据中有许多是相同的,也必须建立各自的文件,其原因就在于文件系统中数据不能充分共享(文件系统中只提供文件一级的共享,不提供记录级的共享)。因此使得存储的数据存在着大量的重复,这种现象称为数据的冗余。

下面我们通过一个例子来说明这个问题。

假设现在要管理的对象是一个学校,它由人事子系统、业务子系统、总务子系统组成,各子系统根据自己的特点及需要选择相应的数据格式,存放自己感兴趣的数据。它由这样三种记录组成,如图 1-7 所示。

在文件系统阶段,由这三个记录的集合组成三个文件:人事文件、业务文件、总务文件。这三个文件对应于三个不同的应用系统,它们之间是互不联系的。而在这三个文件中,有许多项目如部门、职工号、姓名、性别、年龄、职务、工资是相同的,但是在文件系统阶

段中，数据与应用系统是一一对应的，所以必须分别存储，这样就势必出现很大的数据冗余，浪费存储空间。而这种冗余还会带来其他问题，如在更新操作时，必须要进行多次更新操作，以更新某些冗余的副本，增加不必要的机器时间。再者由于多个副本的存在，系统容易出现不一致的信息。而在数据库中，是从整体观点来组织数据的，其数据不是面向个别的应用，而是为多种应用所共享的，从而避免了不必要的冗余。例如上面的例子在数据库中三个文件就不再是互不相干的了，而是将这三个文件视为统一的整体，也就是说，把这三个文件进行分解，重新组合，去掉重复部分，构成一个整体。图 1-8 给出了经过重新组合形成的数据结构。



图 1-7 学校的三个记录

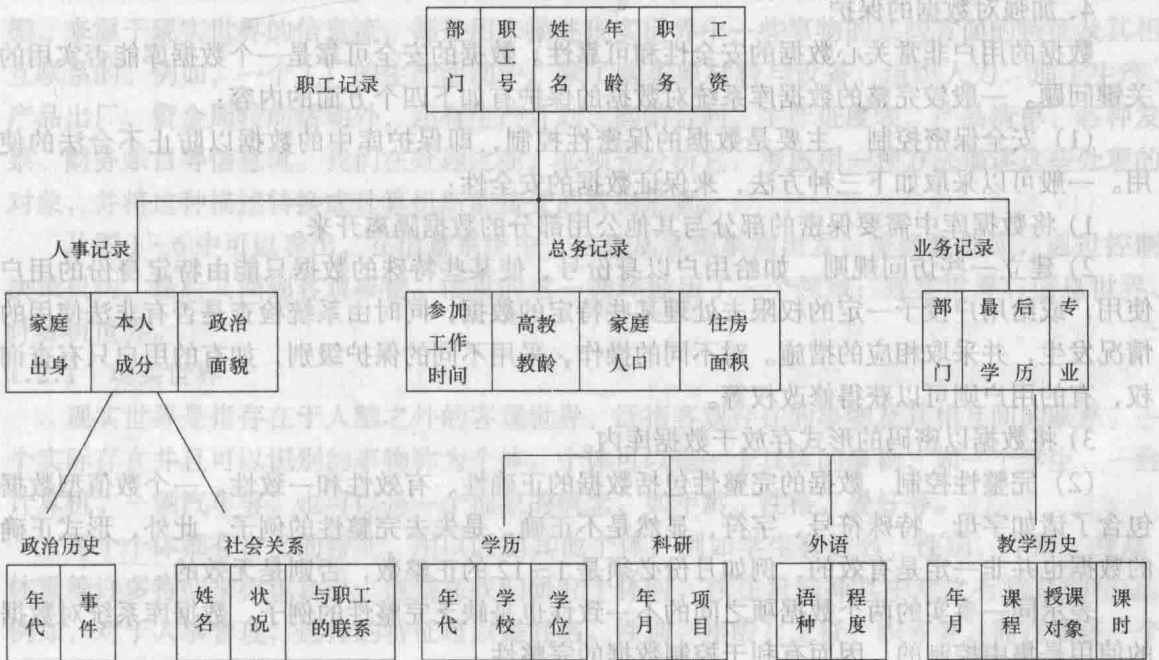


图 1-8 组合后的数据结构



从图 1-8 中可以看出, 我们设计了各部门都要用的职工记录, 并带上了人事记录、业务记录和总务记录等, 这样就把整个组织的数据结构化了。这就要求在描述数据时, 不仅要描述数据本身, 还要描述数据之间的联系, 这种联系往往是通过存取路径来表现的。在这种情况下就大大地减少了数据的冗余度, 节省空间, 提高存取速度, 避免不相容。此外, 由于从整体的观点看数据, 所以数据不是面向应用的, 而是面向系统的。这样它的弹性大, 可以扩充, 应用数据时可以有灵活的方式。即可以取整体模型的各种合理的子集, 用于不同的应用系统。

从理论上说数据库中是没有冗余的, 但实际上, 许多数据库为了改善访问时间, 或者由于比较简单的存取方法, 数据库中还存在着某种程度的冗余性。但是它去掉了有害的或不必要的冗余。

### 3. 维护数据的独立性

所谓数据的独立性, 就是应用程序不必因为数据的存储结构的变化而修改, 即应用程序与数据的结构之间不存在依赖关系, 这是数据库系统所努力追求实现的一个目标。数据库系统结构之所以复杂, 这是一个重要原因。数据库系统的数据独立性分为以下两级:

(1) 物理独立性 即数据库的物理结构的变化(如物理设备的更换、物理位置的变化、存取方法的改变等), 不影响数据库的逻辑结构, 从而也就影响不到应用程序, 不会导致应用程序的修改。

(2) 逻辑独立性 数据库逻辑结构的变化(如数据定义的修改, 新数据类型的增加, 数据间联系的变更等), 不致影响到用户原有应用程序的修改。

这两种独立性统称为数据独立性。数据独立性的目的, 就是应用程序尽可能不受数据的影响。这两种数据独立性是靠数据库管理系统来实现的, 它可大大减轻程序员的负担。

### 4. 加强对数据的保护

数据的用户非常关心数据的安全性和可靠性。数据的安全可靠是一个数据库能否实用的关键问题。一般较完整的数据库系统对数据的保护有如下四个方面的内容:

(1) 安全保密控制 主要是数据的保密性控制, 即保护库中的数据以防止不合法的使用。一般可以采取如下三种方法, 来保证数据的安全性:

1) 将数据库中需要保密的部分与其他公用部分的数据隔离开来。

2) 建立一些访问规则 如给用户以身份号, 使某些特殊的数据只能由特定身份的用户使用; 或给用户授予一定的权限去处理某些特定的数据, 同时由系统检查是否有非法使用的情况发生, 并采取相应的措施。对不同的操作, 采用不同的保护级别, 如有的用户只有查询权, 有的用户则可以获得修改权等。

3) 将数据以密码的形式存放于数据库内。

(2) 完整性控制 数据的完整性包括数据的正确性、有效性和一致性。一个数值型数据包含了诸如字母、特殊符号、字符, 显然是不正确, 是失去完整性的例子。此外, 形式正确的数据也并非一定是有效的, 例如月份必须是 1~12 的正整数, 否则是无效的。

表示同一事实的两个数据项之间的不一致性也是缺乏完整性的例子。数据库系统对数据的使用是集中控制的, 因而有利于控制数据的完整性。

(3) 并发控制 数据的不相容性主要是由于数据的共享而引起的。不同用户同时使用数据库, 可能引起对数据的干扰。例如, 当某个用户需要修改某一数据时, 他必须先读出这个