

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

# 数据库系统 原理与应用

Principles and Applications of Database System

孟彩霞 主编

张荣 乔平安 副主编

- 奠定基础理论知识
- 强化设计方法与技术
- 提高应用开发能力



高校系列

人民邮电出版社  
POSTS & TELECOM PRESS

21世纪高等学校计算机规划教材

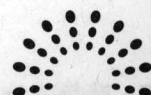
21st Century University Planned Textbooks of Computer Science

# 数据库系统 原理与应用

Principles and Applications of Database System

孟彩霞 主编

张荣 乔平安 副主编



高校系列

人民邮电出版社

北京

图书在版编目 (CIP) 数据

数据库系统原理与应用 / 孟彩霞主编. —北京: 人民邮电出版社, 2008.8  
21 世纪高等学校计算机规划教材  
ISBN 978-7-115-18070-4

I. 数… II. 孟… 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 (2008) 第 062602 号

# 数据库系统原理与应用

## 内 容 提 要

本书主要介绍数据库系统的基本概念和基本理论、应用技术和设计方法。内容包括数据模型和数据库系统结构、关系代数和关系演算、关系系统的查询优化、关系数据库标准语言 SQL、存储过程和触发器、关系数据库规范化理论、数据库系统设计、数据库的安全与保护、数据库应用系统的结构与开发、数据库技术的发展动态。附录中还介绍了目前流行的 SQL Server 2000, 并安排了 10 个实验。全书内容紧凑, 深入浅出, 通俗易懂, 注重实用。

本书可作为高等学校计算机、信息管理和信息系统等相关专业数据库课程的教材, 也可供从事计算机等相关专业的工程技术人员参考。

### 21 世纪高等学校计算机规划教材 数据库系统原理与应用

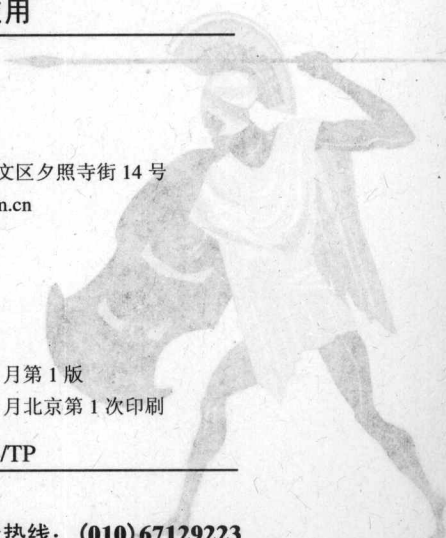
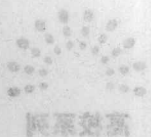
- ◆ 主 编 孟彩霞  
副 主 编 张 荣 乔平安  
责任编辑 张 鑫
  - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号  
邮编 100061 电子函件 315@ptpress.com.cn  
网址 <http://www.ptpress.com.cn>  
北京鑫正大印刷有限公司印刷
  - ◆ 开本: 787×1092 1/16  
印张: 16.5  
字数: 433 千字  
印数: 1—3 000 册
- 2008 年 8 月第 1 版  
2008 年 8 月北京第 1 次印刷

ISBN 978-7-115-18070-4/TP

定价: 26.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154





现今社会对人才的基本要求之一就是应用计算机的能力。在高等学校,培养学生应用计算机的能力,主要是通过计算机课程的体制改革,即计算机教学分层、分类规划与实施;密切联系实际,恰当体现与各专业其他课程配合;教学必须以市场需求为导向,目的是培养高素质创新型人才。

人民邮电出版社经过对教学改革新形势充分的调查研究,依据目前比较成熟的教学大纲,组织国内优秀的有丰富教学经验的教师编写一套体现教学改革最新形势的“高校系列计算机教材”。在本套教材的出版过程中,我社多次召开教材研讨会,广泛听取了一线教师的意见,也邀请众多专家对大纲和书稿做了认真的审读与研讨。本套教材具有以下特点。

### 1. 覆盖面广,突出教改特色

本套教材主要面向普通高等学校(包括计算机专业和非计算机专业),是在经过大量充分的调研基础上开发的计算机系列教材,涉及计算机教育领域中的所有课程(包括专业核心骨干课程与选修课程),适应了目前经济、社会对计算机教育的新要求、新动向,尤其适合于各专业计算机教学改革的特点特色。

### 2. 注重整体性、系统性

针对各专业的特点,同一门课程规划了组织结构与内容不同的几本教材,以适应不同教学需求,即分别满足不同层次计算机专业与非计算机专业(如工、理、管、文等)的课程安排。同时本套教材注重整体性的策划,在教材内容的选择上避免重叠与交叉,内容系统完善。学校可根据教学计划从中选择教材的各种组合,使其适合本校的教学特点。

### 3. 掌握基础知识,侧重培养应用能力

目前社会对人才的需要更侧重于其应用能力。培养应用能力,须具备计算机基础理论、良好的综合素质和实践能力。理论知识作为基础必须掌握,本套教材通过实践教学与实例教学培养解决实际问题的能力和知识综合运用能力。

### 4. 教学经验丰富的作者队伍

高等学校在计算机教学和教材改革上已经做了大量的工作,很多教师在计算机教育与科研方面积累了相当多的宝贵经验。本套教材均由有丰富教学经验的教师编写,并将这些宝贵经验渗透到教材中,使教材独具特色。

### 5. 配套资源完善

所有教材均配有 PPT 电子教案,部分教材配有实践教程、题库、教师手册、学习指南、习题解答、程序源代码、演示软件、素材、图书出版后要更新的内容等,以方便教与学。

我社致力于优秀教材的出版,恳请大家在使用的过程中,将发现的问题与提出的意见反馈给我们,以便再版时修改。



数据库技术是当前计算机领域中应用最广泛、发展最迅速的技术之一,已经成为计算机信息系统与应用系统的核心技术和重要基础。数据库系统在当代的社会生活中得到广泛的应用,并形成了一个巨大的产业。数据库课程已成为计算机及其相关专业的必修课程。

数据库技术始于 20 世纪 60 年代末,经过近 40 年的发展,已经形成以数据建模和 DBMS 核心技术为主,具有相当规模理论体系和实用技术的一门学科,成为计算机软件领域的一个重要分支。数据库技术的发展从早期的层次数据库系统与网状数据库系统,发展到当今流行的关系数据库系统,目前进一步发展的热点是新型的数据库系统。数据库技术的新原理、新技术不断出现,然而这些新型数据库系统大都建立在基本的数据库技术基础之上。

本书详细介绍了数据库系统的基本概念、原理、方法和应用技术,它是在作者多年的数据库课程教学和科研的基础上完成的。本书在编写中,把作者多年来从事有关数据库的教学体会和在科研实践中总结出来的数据库实用技术介绍给读者,使读者真正掌握数据库系统的基本理论和基本技术,提高数据库应用系统的开发能力。本书还介绍了数据库的最新研究发展状况,以了解其飞速发展的步伐。

全书共分 9 章。第 1 章主要介绍数据库系统的几个重要概念、数据库系统结构和数据模型;第 2~5 章介绍关系数据库,包括关系模型的基本概念,即关系模型的数据结构、数据操作和关系的完整性,关系数据库标准语言 SQL,关系数据库中存储过程和触发器的应用,关系数据库的规范化理论等;第 6 章介绍数据库应用系统设计的方法和技术;第 7 章介绍数据库的安全与保护,包括数据库的安全性、完整性、并发控制和数据库恢复;第 8 章介绍数据库应用系统的结构与开发,主要介绍 ODBC 数据访问模型和 ADO 数据访问模型;第 9 章介绍数据库技术的最新发展动态。每章后面均配有习题,以加强对数据库系统概念、方法的理解和掌握。全书内容紧凑,注重实用,深入浅出,通俗易懂。

数据库课程是一门实践性很强的课程,必须配合一定数量的上机实验。本书附录 A 介绍了“MS SQL Server 2000 数据库管理系统”;附录 B 安排了 10 个实验,可以作为课内上机实验题目。

本书讲授学时为 60~70 学时,其中实验学时为 20 学时。学习完本课程后,最好能够安排两周的“课程设计”,完成一个数据库应用系统的设计与开发。

本书由孟彩霞主编,负责全书内容的取材和组织,并编写了第 1、2、9 章以及附录 A 和附录 B,张荣编写了第 6、7 和 8 章,乔平安编写了第 3、4 和 5 章。最后由孟彩霞审校和统稿,张荣也参加了部分审校工作。尚小林提出了一些好的建议,在此表示感谢。

本书可作为计算机等相关专业的数据库课程教材,讲授时可根据需要对内容进行适当取舍。本书也可供从事计算机等相关专业的工程技术人员学习和参考。

由于作者水平有限,编写时间仓促,书中难免存在错误和不足之处,恳请专家和读者批评指正。作者联系方式:mcxmcmcx@sina.com。

编 者

2008 年 4 月

# 目 录

## 第 1 章 数据库系统概述.....1

1.1 引言.....1

1.2 数据库基本概念.....2

1.2.1 信息、数据和数据处理.....2

1.2.2 数据管理技术的发展.....3

1.2.3 数据库系统.....6

1.3 数据库系统结构.....7

1.3.1 数据库系统的三级模式结构.....7

1.3.2 数据库系统的体系结构.....9

1.4 数据模型.....11

1.4.1 数据模型的概念.....11

1.4.2 数据模型的三要素.....12

1.4.3 概念模型.....13

1.4.4 常用的数据模型.....16

小结.....21

习题.....21

## 第 2 章 关系数据库.....23

2.1 关系数据结构.....23

2.1.1 关系的定义.....23

2.1.2 关系的性质.....26

2.1.3 关系模式.....27

2.1.4 关系数据库.....27

2.2 关系数据操作.....27

2.2.1 关系的基本操作.....28

2.2.2 关系数据语言.....28

2.3 关系的完整性.....29

2.3.1 完整性约束的分类.....29

2.3.2 实体完整性 (Entity Integrity).....29

2.3.3 参照完整性 (Referential Integrity).....29

2.3.4 用户定义的完整性.....31

2.4 关系代数.....31

2.4.1 传统的集合运算.....32

2.4.2 专门的关系运算.....33

2.4.3 关系代数运算的应用实例.....37

2.5 关系演算.....39

2.5.1 元组关系演算.....39

2.5.2 域关系演算.....41

2.5.3 关系运算的安全性和等价性.....41

2.6 关系系统的查询优化.....42

2.6.1 查询处理步骤.....43

2.6.2 一般问题.....44

2.6.3 查询优化的一般准则.....45

小结.....46

习题.....46

## 第 3 章 关系数据库标准语言 SQL.....49

3.1 SQL 概述.....49

3.1.1 SQL 的产生与发展.....49

3.1.2 SQL 的特点.....49

3.1.3 SQL 的基本概念.....51

3.2 SQL 的数据定义.....51

3.2.1 SQL 的数据定义功能.....51

3.2.2 基本表的定义、删除与修改.....51

3.2.3 索引的建立与删除.....53

3.3 SQL 的数据查询.....54

3.3.1 SELECT 语句的一般格式.....54

3.3.2 单表查询.....55

3.3.3 连接查询.....62

3.3.4 嵌套查询.....64

3.3.5 集合查询.....67

3.4 SQL 的数据更新.....68

3.4.1 插入数据.....68

3.4.2 修改数据.....69

3.4.3 删除数据.....69

3.5 视图.....70

3.5.1 视图的定义与删除.....70

3.5.2 查询视图.....71

3.5.3 更新视图.....72

3.5.4 视图的作用	73	5.3.7 关系模式的规范化步骤	111
3.6 SQL 提供的安全性与完整性	73	5.4 函数依赖的公理系统	112
3.6.1 安全性	74	5.4.1 Armstrong 公理系统	112
3.6.2 完整性约束	75	5.4.2 极小函数依赖集	115
3.7 嵌入式 SQL	78	5.5 关系模式的分解	116
3.7.1 嵌入式 SQL 的使用规定	78	5.5.1 关系模式分解的概念	117
3.7.2 嵌入式 SQL 的使用方法	80	5.5.2 具有无损连接性的关系模式分解	118
3.7.3 动态 SQL 简介	84	5.5.3 具有保持函数依赖的关系模式分解	120
小结	86	5.5.4 通过分解实现关系模式的规范化	121
习题	87	小结	123
<b>第 4 章 存储过程和触发器</b>	89	习题	124
4.1 存储过程	89	<b>第 6 章 数据库系统设计</b>	126
4.1.1 基本概念	89	6.1 数据库设计概述	126
4.1.2 创建和执行存储过程	90	6.1.1 数据库设计的基本任务	126
4.1.3 存储过程的修改和删除	94	6.1.2 数据库设计的基本步骤	127
4.2 触发器	95	6.1.3 数据库设计特点	129
4.2.1 基本概念	95	6.2 需求分析	130
4.2.2 定义触发器	96	6.2.1 需求分析的任务和方法	130
4.2.3 激活触发器	98	6.2.2 数据流图	131
4.2.4 删除触发器	99	6.2.3 数据字典	133
小结	99	6.3 概念结构设计	135
习题	100	6.3.1 概念结构的特点及设计方法	135
<b>第 5 章 关系数据库规范化理论</b>	101	6.3.2 概念结构设计的步骤	136
5.1 问题的提出	101	6.4 逻辑结构设计	141
5.1.1 关系模型可能存在的异常	102	6.4.1 E-R 模型向关系模型的转换	142
5.1.2 异常原因分析	103	6.4.2 关系模型的优化	143
5.2 函数依赖	103	6.4.3 外模式的设计	145
5.2.1 函数依赖的定义	104	6.5 物理结构设计	146
5.2.2 几种特殊的函数依赖	104	6.5.1 物理结构设计的内容和方法	146
5.2.3 键	105	6.5.2 确定物理结构	148
5.3 规范化	105	6.5.3 评价物理结构	148
5.3.1 范式及其类型	105	6.6 数据库的实施和维护	149
5.3.2 第一范式 (1NF)	106	6.6.1 数据的载入和应用程序的调试	149
5.3.3 第二范式 (2NF)	106	6.6.2 数据库的试运行	149
5.3.4 第三范式 (3NF)	107	6.6.3 数据库的运行和维护	150
5.3.5 BC 范式 (BCNF)	107	小结	151
5.3.6 多值依赖和第四范式 (4NF)	109	习题	151



<b>第 7 章 数据库的安全与保护</b> .....153	8.2.2 ODBC 的组成与结构.....180
7.1 安全与保护概述.....153	8.2.3 ODBC 的接口函数.....181
7.1.1 安全与保护定义.....153	8.2.4 ODBC 的基本工作流程.....187
7.1.2 安全标准简介.....154	8.3 ADO 数据访问模型.....191
7.2 数据库的安全性保护.....157	8.3.1 ADO 对象模型.....192
7.2.1 用户标识与鉴别 (Identification & Authentication).....157	8.3.2 ADO 的基本工作流程.....197
7.2.2 存取权限控制.....157	8.3.3 ADO 应用的简单实例.....198
7.2.3 视图机制.....159	小结.....200
7.2.4 审计.....159	习题.....201
7.2.5 数据加密.....160	<b>第 9 章 数据库技术的发展动态</b> .....202
7.3 数据库的完整性保护.....161	9.1 数据库技术的研究与发展.....202
7.3.1 完整性约束条件.....161	9.1.1 数据库技术研究的新特点.....202
7.3.2 完整性控制机制.....162	9.1.2 数据库技术的发展方向.....203
7.4 数据库的并发控制技术.....162	9.2 面向对象数据库系统.....204
7.4.1 事务及其特性.....162	9.2.1 面向对象数据模型.....205
7.4.2 数据库并发操作带来的问题.....163	9.2.2 对象关系数据库与对象数据库.....207
7.4.3 封锁.....165	9.3 并行数据库系统与分布式数据库系统.....209
7.4.4 活锁和死锁.....166	9.3.1 并行数据库系统.....209
7.4.5 并发调度的可串行性.....167	9.3.2 分布式数据库系统.....210
7.4.6 两段锁协议.....168	9.3.3 并行数据库系统与分布式数据库系统的区别.....214
7.4.7 封锁的粒度.....168	9.4 数据仓库与联机分析处理技术.....215
7.5 数据库的恢复技术.....169	9.4.1 数据仓库技术.....215
7.5.1 故障的种类.....169	9.4.2 联机分析处理技术.....219
7.5.2 恢复的实现技术.....170	9.4.3 数据挖掘技术.....220
7.5.3 恢复策略.....172	9.5 其他几种新型的数据库系统.....222
7.5.4 具有检查点的恢复技术.....173	9.5.1 多媒体数据库.....222
7.5.5 数据库镜像.....174	9.5.2 主动数据库.....223
小结.....175	9.5.3 工程数据库.....223
习题.....175	小结.....224
<b>第 8 章 数据库应用系统的结构与开发</b> .....177	习题.....225
8.1 数据库应用系统的结构.....177	<b>附录 A MS SQL Server 2000 数据库管理系统</b> .....226
8.1.1 集中式体系结构.....177	A.1 SQL Server 2000 概述.....226
8.1.2 客户机/服务器体系结构.....177	A.1.1 SQL Server 2000 的安装.....226
8.1.3 浏览器/服务器体系结构.....178	A.1.2 SQL Server 2000 的主要管理工具.....227
8.2 ODBC 数据访问模型.....179	A.1.3 SQL Server 2000 的配置.....229
8.2.1 ODBC 的提出.....179	



# 第 1 章

## 数据库系统概述

数据库是数据管理的最新技术，是计算机科学技术中发展最快的技术之一，也是应用最为广泛的技术之一，它早已成为计算机科学的重要分支。数据库技术的应用已渗透到工农业生产、商业、行政、科学研究、工程技术和国防军事等领域的每一个部门，并随着 Internet 的出现遍布社会的每一个角落。

今天，人类社会已处于信息化时代，信息已经成为各个部门的重要财富和资源，而数据库是信息化社会中资源管理与开发利用的基础，数据库的建设规模和应用水平已成为衡量一个国家信息化程度的重要标志。因此，数据库课程早已成为计算机科学与技术相关专业的重要课程，尤其是信息专业的核心课程，也是许多非计算机专业的选修课程。

本章主要是对数据库系统做一概括性介绍。首先介绍数据库系统的一些基本概念，然后介绍数据库系统结构和数据模型，以使读者对数据库技术的基本概念和基本知识有所了解。

### 1.1 引言

数据库技术诞生于 20 世纪 60 年代末，至今已走过了近 40 年的历程，特别是近 20 年，数据库技术及其应用得到了迅猛的发展。数据库系统从早期的层次数据库和网状数据库，发展到目前占主流地位的关系数据库，已形成了较为完整的理论体系。

1968 年，IBM 公司推出层次模型的 IMS (Information Management System) 数据库系统；1969 年，美国数据系统语言协会 (Conference On Data System Language, CODASYL) 的数据库任务小组 (DataBase Task Group, DBTG) 发表了一系列的报告中提出了网状模型；1970 年，IBM 研究中心的 E.F.Codd 博士发表了关于关系模型的著名论文。这 3 件工作奠定了现代数据库技术的基础。由于 C.M.Bachman 在网状模型和 DBTG 报告中的贡献，他在 1973 年荣获美国计算机学会 (ACM) 授予的图灵奖。E.F.Codd 在关系模型上作出了杰出的开拓性贡献，在 1981 年获得了图灵奖。

20 世纪 70 年代和 80 年代是数据库蓬勃发展期，不仅推出了一些网状系统和层次系统，还围绕关系数据库进行了大量的研究和开发工作，使得关系数据库的理论和系统日趋完善。随着计算机硬件性能的改善，关系系统逐步代替网状系统和层次系统，而占领了市场。迄今为止，在数据库产品市场，出现了许多关系数据库产品，如 Oracle、DB2、Sybase、Informix、SQL Server 等。

数据库系统的出现使信息系统从以加工数据的程序为中心转向围绕共享的数据为中心的新阶段。这样既便于数据的集中管理，又有利于应用程序的研制和维护，提高了数据的利用率和相容性。20 世纪 80 年代后不仅在大型机上，在多数微机上也配置了数据库管理系统，使数据库技术



得到更加广泛的应用和普及。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和制造(CAD/CAM)、办公信息系统、地理信息系统等都应用了数据库技术。数据库的建设规模、数据库中信息量的大小及使用程度已经成为衡量一个企业信息化程度的重要标志。

在国内,20世纪70年代开始使用数据库的单位主要集中在国家部委、国防、气象和石油等一些特殊行业和部门。而数据库技术真正得到广泛的应用,是从20世纪80年代的dBASE II开始,尽管dBASE II和随后的xBASE系列都不能称为一个完备的数据库管理系统,但是它们都支持基本的关系数据模型,使用起来非常方便,加之该系统能在微机上实现,一般也能满足中、小规模的信息管理系统的应用,为数据库技术的普及和广泛应用奠定了基础。

数据库技术是在操作系统的文件系统基础上发展起来的,而且数据库管理系统本身要在操作系统的支持下才能工作。数据库与数据结构之间的联系十分密切,数据库技术不仅涉及数据结构中的链表、树、图等知识,而且还丰富了数据结构的内容。应用程序是使用数据库系统最基本的方式,因为系统中大量的应用程序都是用高级语言加上数据库的操作语言联合编制的。离散数学中的集合论、数理逻辑是关系数据库的理论基础,很多概念、术语、思想都直接用到关系数据库中。因此,数据库技术课程是一门综合性较强的课程。

## 1.2 数据库基本概念

### 1.2.1 信息、数据和数据处理

人类的社会活动和生产活动,离不开对信息的收集、保存、利用和处理,特别是当今社会生产力发展突飞猛进,新技术层出不穷,信息量迅速剧增,人类社会进入了信息化的阶段。那么究竟什么是信息呢?信息是人们用以对客观世界直接进行描述的,可在人们之间进行传递的一些知识。信息需要被加工和处理,需要被交流和使用。随着计算机技术的迅速发展,计算机具有的高速处理能力和存储容量巨大的特点,使得人们有可能对大量的信息进行保存和加工处理。为了记载信息,人们使用各种各样的符号和它们的组合来表示信息,这些符号及其组合就是数据。数据是信息的具体表示形式,信息是数据的有意义的表现。由此可见,信息和数据有一定的区别,但在有些场合信息和数据难以区分,信息本身就是数据化了的,数据本身是一种信息。因此在很多场合不对它们进行区分,信息处理与数据处理往往指同一个概念,计算机之间交换数据也可以说成是交换信息等。

数据(Data)是数据库中存储的基本对象。数据在大多数人头脑中的第一个反映就是数字,如26、2 000、53.4、-12.5、¥5 400、\$ 750等。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。数据这个概念已经不局限于普通意义上的数字,数据的种类很多,包括文字、图形、图像、声音、学生的档案记录、货物的运输情况等,这些都是数据。

数据可以定义为描述事物的符号记录。描述事物的符号可以是数字,也可以是文字、图形、图像、声音、语言等,数据有多种表现形式,它们都可以经过符号化后存入计算机。早期的计算机主要用于科学计算,处理的数据是数值型数据,现在计算机存储和处理的对象十分广泛,表示这些对象的数据也越来越复杂了。

数据的表现形式还不能完全表达其内容,需要经过解释,数据和关于数据的解释是不可分的。

例如, 95 是一个数据, 它可以是一个学生某门课的成绩, 也可以是某个人的体重, 还可以是一个实验室中计算机的台数。数据的解释是指对数据含义的说明, 数据的含义称为数据的语义, 数据与其语义是不可分的。

在日常生活中, 人们直接用自然语言描述事物。例如, 这样描述一名学生, 李林同学, 男, 1986 年出生, 陕西省西安市人, 2004 年入学, 就读于计算机科学系。而在计算机中常常这样来描述, (李林, 男, 1986, 陕西省西安市, 计算机科学系, 2004), 即把学生的姓名、性别、出生年份、籍贯、所在系别、入学时间等组织在一起, 组成一个记录。这里的学生记录就是描述学生的数据。这样的数据是有结构的, 记录是计算机中表示和存储数据的一种格式或者一种方法。

有了数据就产生了数据处理的问题, 人们收集到的各种数据需要经过加工处理。数据处理包括对数据的收集、记载、分类、排序、存储和计算等工作, 其目的是使有效的信息资源得到合理和充分地利用, 从而促进社会生产力的发展。

数据处理经过了手工处理、机械处理和电子数据处理 3 个阶段。今天, 用计算机进行数据处理方法的研究已成为计算机技术中的主要课题之一, 数据库技术已成为社会信息化时代不可缺少的方法和工具。

## 1.2.2 数据管理技术的发展

数据处理的核⼼问题是数据管理, 数据管理是指对数据进行分类、组织、编码、存储、检索、维护等。在应用需求的推动下, 随着计算机硬件和软件的发展, 数据管理技术先后经历了 3 个发展阶段, 即人工管理阶段、文件系统管理阶段和数据库系统管理阶段。

### 1. 人工管理阶段

在 20 世纪 50 年代中期以前, 计算机主要用于科学计算。当时的硬件状况是, 外存只有纸带、卡片和磁带, 没有磁盘等直接存取的存储设备; 软件状况是, 没有操作系统和管理数据的软件; 数据处理方式是批处理。当时对数据的管理是由程序员个人考虑和安排的, 一个程序对应于一组数据, 进行程序设计时, 往往也要对数据的结构、存储方式和输入/输出方式等进行设计。严格地说, 这种管理只是一种技巧, 这是数据自由管理的方式, 因此, 这一阶段又称为自由管理阶段。其特点包括以下几方面。

(1) 数据不能长期保存。当时计算机主要用于科学计算, 当用户需要计算某一课题时, 就临时将有关数据输入, 计算完毕后输出运算结果。随着计算任务的完成, 数据空间与程序空间一起被释放, 计算机在处理过程中不长期保存数据。

(2) 没有专门的软件对数据进行管理。由于没有专门的数据管理软件负责数据的管理工作, 数据管理是由应用程序自己管理。程序员不仅要规定数据的逻辑结构, 而且还要在编制程序时设计物理结构, 即要设计数据的存储结构、存取方法和输入/输出方式等。因此程序员负担很重。

(3) 数据不共享。由于数据是面向应用程序的, 一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时, 由于必须各自定义, 无法互相利用, 因此存在大量的冗余数据。

(4) 数据不具有独立性。由于一组数据只能对应一个程序, 即程序依赖于数据, 如果数据的类型、格式或输入/输出方式等逻辑结构或物理结构发生变化, 必须对应用程序做出相应的修改, 这就加重了程序员的负担。



图 1.1 人工管理阶段程序与数据的关系

## 2. 文件系统管理阶段

20世纪50年代后期到60年代中期,计算机软硬件都得到了发展,计算机应用领域拓宽,不仅用于科学计算,还大量用于数据管理。这时硬件方面已出现了磁盘、磁鼓等直接存取的存储设备;软件方面,操作系统中已经有了专门的数据管理软件,即文件系统;处理方式上不仅有了批处理,而且能够联机实时处理。和人工管理阶段相比,该阶段的数据管理具有如下优点。

(1) 数据可以长期保存。

(2) 由文件系统管理数据。在文件系统阶段,由专门的软件即文件系统进行数据管理。文件系统把数据组织成相互独立的数据文件,利用“按文件名进行访问,按记录进行存取”的管理技术,可以对文件进行修改、插入和删除的操作。文件系统实现了记录内的结构性,但整体无结构。程序和数据之间由文件系统提供存取方法进行转换,使应用程序与数据之间有了一定的独立性,程序员可以不必过多地考虑物理细节,将精力集中于算法。而且数据在存储上的改变不一定反映在程序上,大大节省了维护程序的工作量。

但是,文件系统仍存在以下缺点。

(1) 数据共享性差,冗余度大。文件系统中,一个数据文件基本上对应于一个应用程序,即数据仍然是面向应用的。当不同的应用程序具有部分相同的数据时,也必须建立各自的数据文件,而不能共享相同的数据,因此数据的冗余度大,浪费存储空间。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困难。

(2) 数据独立性差。文件系统中的文件是为某一特定应用服务的,文件的逻辑结构对该应用程序来说是优化的,因此要想对现有的数据再增加一些新的应用会很困难,系统不容易扩充。一旦数据的逻辑结构改变,必须修改应用程序;应用程序的改变,也将引起文件的数据结构改变。因此数据与程序之间仍缺乏独立性。可见,文件系统仍然是一个不具有弹性的无结构的数据集合,即数据文件之间是孤立的,不能反映现实世界事物之间的内在联系。

在文件系统阶段,程序与数据之间的关系如图1.2所示。

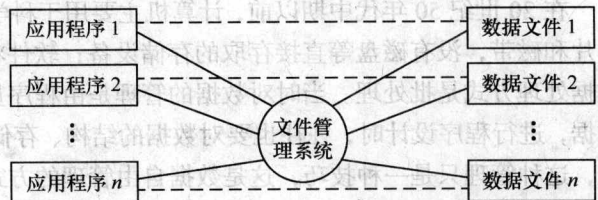


图 1.2 文件系统阶段程序与数据的关系

## 3. 数据库系统管理阶段

20世纪60年代后期以来,计算机软硬件技术得到了飞速发展,同时,计算机用于管理的规模越来越大,应用也越来越广泛,数据量急剧增长,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。这时硬件已有大容量的磁盘,硬件价格下降;软件则价格上升,为编制和维护系统软件及应用程序所需的成本相对增加;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。

为了实现多用户、多应用共享数据,使数据为尽可能多的应用服务,以文件系统作为数据管理手段已经不能满足应用的需求,于是数据库技术便应运而生,出现了统一管理数据的专门软件系统,即数据库管理系统(DataBase Management System, DBMS)。

从文件系统到数据库系统,标志着数据管理技术的飞跃。与文件系统管理阶段相比,数据库系统管理阶段具有以下优点。

(1) 数据结构化。数据结构化是数据库主要特征之一,是数据库系统与文件系统的根本区别。

在文件系统中,相互独立的文件的记录内部是有结构的,传统文件的最简单形式是等长同格式的记录集合,但记录之间是没有联系的,并且文件是面向某一应用的。而实际系统往往涉及许多应用,在数据库系统中不仅要考虑某个应用的数据结构,还要考虑整个组织的数据结构。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。

在数据库系统中，数据不再针对某一应用，而是面向全组织，具有整体的结构化。另外存取数据的方式也很灵活，可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。而在文件系统中，数据的最小存取单位是记录。

(2) 数据的共享性高、冗余度低、易扩充。由于数据库系统中的数据不再面向某个应用而是面向整个系统，因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余，节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。

由于数据库系统中的数据是面向整个系统，是有结构的数据，因此不仅可以被多个应用共享使用，而且容易增加新的应用，可以适应各种应用需求。当应用需求改变或增加时，只要重新选取整体数据的不同子集，便可以满足新的要求，这就使得数据库系统具有弹性大、易扩充的特点。

(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，包括数据的物理独立性和数据的逻辑独立性。

数据的物理独立性是指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的。也就是说，数据在磁盘上的数据库中怎样存储是由 DBMS 管理的，用户程序不需要了解，应用程序要处理的只是数据的逻辑结构，这样当数据的物理存储改变了，应用程序不用改变。

数据的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，用户程序也可以不变。

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 DBMS 负责，从而简化了应用程序的编制，大大减少了应用程序的维护和修改工作。

数据独立性是由 DBMS 的二级映像功能来保证的，稍后再介绍。

(4) DBMS 对数据进行统一的管理和控制。数据库对系统中的用户来说是共享资源，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。为此，DBMS 必须提供以下几方面的数据控制和保护功能。

① 数据的安全性保护。数据的安全性是指保护数据以防止被不合法的使用所造成的数据泄密和破坏，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，学生对于课程的成绩只能进行查询，不能修改。

② 数据的完整性控制。数据的完整性是指数据的正确性、有效性和相容性。完整性检查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。例如，月份是 1~12 之间的正整数，学生学号必须唯一，学生所在的院系必须是有效存在的院系等。

③ 数据库恢复。计算机系统的软硬件故障、操作员的失误以及恶意的破坏都会影响到数据库中数据的正确性，甚至造成数据库部分或全部数据的丢失。因此 DBMS 必须具有将数据库从错误状态恢复到某一已知的正确状态（也称为完整状态或一致状态）的功能。

④ 并发控制。当多个用户的并发进程同时存取、修改数据库时，可能会发生相互干扰而得到错误的结果或使得数据库的完整性遭到破坏，因此必须对多用户的并发操作加以控制和协调。

数据库技术发展到今天已经是一门成熟的技术，无论是从数据库的技术水平，还是从数据库的应用水平，今天与过去不可同日而语，但数据库的最基本特征并没有变化。在数据库系统中，程序与数据之间的关系如图 1.3 所示。

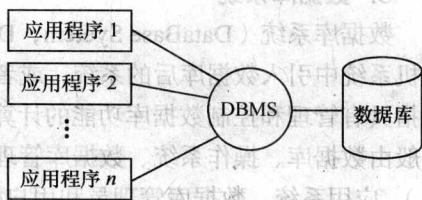


图 1.3 数据库系统阶段程序与数据的关系

## 1.2.3 数据库系统

### 1. 数据库

顾名思义，数据库 (DataBase, DB) 是存放数据的仓库。仓库是保存和管理物资的，并能根据其服务对象的要求随时提供所需物资。数据库是存储和管理数据，并负责向用户提供所需数据的“机构”。只不过这个仓库是在计算机存储设备上，而且数据是按一定的格式存放的。

数据库是指长期储存在计算机内的、有组织的、可共享的数据集合。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，并可为各种用户共享。

例如，一个部门可能同时有职工文件（职工号、姓名、地址、部门、工资等）和业务档案文件（职工号、姓名、部门、完成项目、评价等），在这两个文件中存在着一定的冗余（重复）数据（职工号、姓名、部门）。在构造数据库时，就可以消除这 3 项数据的冗余，只存储一套数据，因为数据库中的数据可为用户共享。

概括地说，数据库数据具有永久存储、有组织和可共享 3 个基本特点。

### 2. 数据库管理系统

介绍了数据和数据库的概念，接下来的问题是如何科学地组织和存储数据，如何高效地获取和维护数据，完成这个任务的是一个系统软件——数据库管理系统。

数据库管理系统 (DataBase Management System, DBMS) 是位于用户与操作系统之间的一层数据管理软件，是一个帮助用户建立、使用和管理数据库的软件系统，是数据库与用户之间的接口。它的基本功能应包括以下几个方面。

(1) 数据定义功能。DBMS 提供数据定义语言 (Data Definition Language, DDL)，用户通过 DDL 可以方便地对数据库中的数据对象进行定义。

(2) 数据操纵功能。DBMS 提供数据操纵语言 (Data Manipulation Language, DML)，用户通过使用 DML 操纵数据，实现对数据库的基本操作，如查询、插入、删除和修改。

(3) 数据库的事务管理和运行管理。数据库在建立、运行和维护时由 DBMS 统一管理、统一控制，以保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

(4) 数据库的建立和维护功能。它包括数据库初始数据的输入、转换功能，数据库的转储、恢复功能，数据库的重组功能，以及性能监视、分析功能等。

(5) 其他功能。它包括 DBMS 与网络中其他软件系统的通信功能；一个 DBMS 与另一个 DBMS 或文件系统的数据库转换功能；异构数据库之间的互访和互操作功能等。

数据库管理系统是数据库系统的一个重要组成部分。

### 3. 数据库系统

数据库系统 (DataBase System, DBS) 是指在计算机系统中引入数据库后的系统，或者说数据库系统是指具有管理和控制数据库功能的计算机系统。DBS 一般由数据库、操作系统、数据库管理系统 (及其工具)、应用系统、数据库管理员和用户构成，如图 1.4 所示。

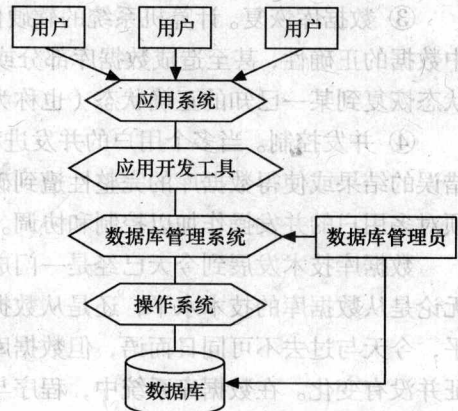


图 1.4 数据库系统示意图

应当指出的是,数据库的建立、管理、使用和维护等工作只靠一个 DBMS 远远不够,在数据库系统中,还需要有专门的管理机构来监督和管理数据库系统,数据库管理员(DataBase Adiministrator, DBA)则是这个机构的一个(组)人员。DBA 是负责全面管理和控制数据库系统正常运行的人员,他承担着创建、监控和维护整个数据库结构的责任。DBA 应该对程序语言和系统软件(如 OS、DBMS 等)都比较熟悉,还要了解各应用部门的所有业务工作。DBA 的素质在一定程度上决定了数据库应用的水平,所以他们是数据库系统中最重要的一类人员。

由于数据库系统数据量通常都很大,加之 DBMS 丰富的功能使得自身的规模也很大,因此数据库系统对硬件提出了较高的要求,这些要求包括要有足够大的内存;要有足够大的磁盘存放数据库,足够的磁带做数据备份;要求系统有较高的通信能力,以提高数据传输率。

通常,在一台能够满足数据库应用系统开发需求的计算机上先安装一个具体的数据库管理系统,而它必须安装在一个具体的操作系统之上,然后开发人员利用应用开发工具并根据用户需求开发一个具体的应用系统,从而形成一个完整的数据库系统。DBA 的任务就是管理和维护这个数据库系统正常运行。

在具备了硬件环境、操作系统等其他系统软件和某个具体的数据库管理系统的情况下,对数据库应用开发人员来说,就是如何使用这个环境来表达用户的要求,并转换成有效的数据库结构,构造较优的数据库模式等,这就涉及数据库设计问题。在后续章节中将对数据库设计问题进行介绍。

在不引起混淆的情况下常常把数据库系统简称为数据库。

## 1.3 数据库系统结构

从不同的角度考察,可以得出不同的数据库系统结构。从数据库管理系统角度来看,数据库系统通常采用三级模式结构,这是数据库管理系统内部的结构;从数据库最终用户的角度来看,数据库系统的结构分为单用户结构、主从式结构、分布式结构、客户机/服务器结构、浏览器/服务器结构等,这是数据库系统外部的结构。

### 1.3.1 数据库系统的三级模式结构

虽然实际的数据库管理系统软件产品种类繁多,它们支持不同的数据模型,使用不同的数据库语言,建立在不同的操作系统之上,数据的存储结构也各不相同,但它们在结构上通常都具有相同的特征。即从 DBMS 的角度看,数据库系统结构采用三级模式结构,并提供两级映像功能。

在数据库系统中,用户看到的数据和计算机中存放的数据是有区别的,它们之间通过两次映像变换相互联系起来。

#### 1. 三级模式

数据库系统的三级模式结构是指数据库系统是由外模式、模式、内模式构成,如图 1.5 所示。

(1) 模式 (Schema),又称为逻辑模式,是数据库中全体数据的逻辑结构和特征的描述,是所有用户的公共数据视图。它是数据库系统模式结构的中间层,即不涉及数据的物理存储细节和硬件环境,也与具体的应用程序、所使用的应用开发工具及高级程序设计语言无关。



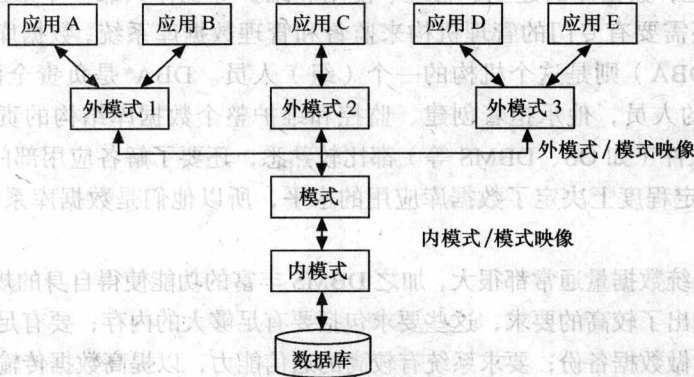


图 1.5 数据库系统的三级模式结构

模式实际上是数据库数据在逻辑级上的视图。一个数据库只有一个模式。数据库模式以某一种数据模型为基础，统一综合地考虑了所有用户的需求，并将这些需求有机地结合成一个逻辑整体。定义模式时不仅要定义数据的逻辑结构，如数据记录由哪些数据项构成，数据项的名称、类型、取值范围等，而且要定义这些数据之间的联系，定义与数据有关的安全性、完整性要求。

在数据库的三级模式结构中，数据库模式即全局逻辑结构是数据库的中心与关键，它独立于数据库的其他层次，因此设计数据库模式结构时应首先确定数据库的逻辑模式。

DBMS 提供模式描述语言（模式 DDL）来严格地定义模式。

(2) 外模式 (External Schema)，又称子模式 (Subschema) 或用户模式，它是数据库用户（包括应用程序员和最终用户）能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示。

外模式通常是模式的子集。一个数据库可以有多个外模式。由于它是各个用户的数据视图，如果不同的用户在应用需求、看待数据的方式、对数据保密的要求等方面存在差异，则其外模式描述就是不同的。即使对模式中同一数据，在外模式中的结构、类型、长度、保密级别等都可以不同。另一方面，同一外模式也可以为某一用户的多个应用系统所使用，但一个应用程序只能使用一个外模式。

外模式是保证数据库安全性的一个有力措施。每个用户只能看见和访问所对应的外模式中的数据，数据库中的其余数据是不可见的。也就是说，用户对数据库的操作，只能与外模式发生联系，按照外模式的结构存储和操纵数据，不必关心模式。

数据库的外模式面向具体的应用程序，它定义在逻辑模式之上，但独立于存储模式和存储设备。当应用需求发生较大变化，相应外模式不能满足其视图要求时，该外模式就需要做相应改动，所以设计数据库外模式时应充分考虑到应用的扩充性。

DBMS 提供子模式描述语言（子模式 DDL）来严格地定义子模式。

(3) 内模式 (Internal Schema)，又称存储模式 (Storage Schema)，一个数据库只有一个内模式。它是数据物理结构和存储方式的描述，是数据在数据库内部的表示方式。例如，记录的存储方式是顺序存储，还是按照某个（些）属性值的升（降）序存储，还是按照属性值聚簇 (Cluster) 存储；索引按照什么方式组织，是 B+ 树索引，还是 Hash 索引；数据是否压缩存储，是否加密；数据的存储记录结构有何规定，如定长结构或变长结构等。

数据库的内模式依赖于它的全局逻辑结构，但独立于数据库的用户视图，即外模式，也独立