

软 件 工 程 系 列 规 划 教 材

# 数据库 系统及应用 (第2版)

◎ 魏祖宽 主编    ◎ 胡 旺 郑莉华 编著



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

TP311.13/354=2

软 件 工 程

2012

划 教 材

# 数据库 系统及应用

## (第2版)

TP311.13  
354=2

◎ 魏祖宽 主编 ◎ 胡 旺 郑莉华 编著



北方工业大学图书馆



C00273472

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书从实用性和先进性出发,通过一个完整的数据库应用实例,全面介绍数据库的基本理论及数据库的管理、设计与开发技术。全书共 14 章,主要内容包括:数据库系统概论、关系数据模型、关系数据库标准 SQL 语言、完整性约束、数据库安全、数据库恢复技术、并发控制、数据库设计方法、数据库高级开发技术、数据仓库技术、数据挖掘技术、地理信息系统和空间数据库、主流数据库产品介绍等。

本书可作为高等学校软件工程和计算机专业本科及研究生的教材,也可供相关领域的技术和管理人员学习、参考。

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

数据库系统及应用/魏祖宽主编;胡旺,郑莉华编著. —2 版. —北京:电子工业出版社,2012.4  
软件工程系列规划教材

ISBN 978-7-121-16653-2

I. ①数… II. ①魏… ②胡… ③郑… III. ①数据库系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2012)第 055773 号

策划编辑:王羽佳

责任编辑:王羽佳

印 刷:涿州市京南印刷厂  
装 订:

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 字数:621 千字

印 次:2012 年 4 月第 1 次印刷

印 数:4000 册 定价:39.90 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 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

# 出版说明

为适应我国经济结构战略性调整的要求和软件产业发展对人才的迫切需要,实现我国软件人才培养的跨越式发展,教育部和国家发展计划委员会于 2001 年联合批准在国内部分高等学校开办示范性软件学院,努力造就一批进入国际前沿、掌握关键技术、擅长顶层设计的技术带头人,培养一批具备不同专业背景且有市场观念的开发管理、工程管理和软件经营等复合型软件人才,形成一支有相当规模和质量、从事软件开发与应用的专业技术人员队伍。

经过多年的软件人才培养和教学实践,依据国内外企业对软件人才的知识和能力需求,以培养高层次、实用型软件人才为目标,我们组织长期从事软件工程硕士教学的专家教授,编写了一套软件工程专业学位系列教材。该系列教材主要包括《软件开发技术》、《数据库系统及应用》、《操作系统原理与 Linux 实例设计》、《面向对象技术与工具》及《计算机网络》。本次推出的软件工程硕士系列教材内容涵盖软件工程硕士需要掌握的主要知识和基本技术,具有领域宽、实用型强的特点,既可以作为软件学院工程硕士专业基础课教材,也可作为计算机专业高年级本科生和研究生的教材,还可供软件开发和管理人员作为参考书籍。我们还将陆续推出系列教材的习题解答和上机指导及教学用多媒体电子课件,便于教师备课和学生自学,请登录华信教育资源网 <http://www.hxedu.com.cn> 注册下载。

在本系列教材的编写过程中,得到了国内众多软件学院的任课教师和软件工程专家的大力支持和帮助,他们提出了许多中肯的意见和建议,对编写工作起到很大的指导作用,对此,编委会和作者表示衷心的感谢!

我们殷切希望本套教材的出版能对国内软件人才的培养起到推波助澜的作用。

尽管我们希望全力以赴编好这套教材,但由于水平和经验有限,难免存在不足和疏漏之处,恳请读者不吝指正。

软件工程系列规划教材编委会

# 前 言

计算机技术的发展不仅极大地促进了科学技术的发展,而且明显加快了经济信息化和社会信息化的进程。因此,计算机教育在各国都备受重视,具备计算机知识与使用能力已成为 21 世纪人才的基本素质之一。

数据库应用技术是其中的核心技术之一,以其为核心的各种数据库应用管理,无可争议地改变了政府部门和企事业单位的运营和管理方式。随着数据库应用广度和深度的扩展,不仅是计算机和信息技术行业,而且包括技术管理、工程管理甚至决策人员在内的众多行业都开始关心数据库技术。

教育部和国家发展计划委员会在全国设立了若干示范性软件学院,为了适应各软件学院正在开展的课程体系与教学内容的改革,及时反映相关研究成果,积极探索适应 21 世纪软件工程专业人才培养的教学模式,我们编写了这本数据库应用技术的教材。

本教材具有如下特色:

1. 根据读者的层次分类。将数据库技术分为数据库基础知识、数据库管理技术、数据库应用技术、现代数据库技术四个部分,且内容上保持连贯性。读者可以根据自身需求选择适当的内容阅读,不同层次的读者可以从不同的深度学习数据库知识。这样就使得本书更加符合软件工程应用的特点。

2. 面向软件工程理念,采用工程应用型学习方法,即“提出问题→解决问题→应用分析”的问题驱动方式,突出学生主动探究在整个教学中的作用。

3. 在内容描述上,我们换位思考,站在学生的角度阐述概念和理论,避免堆砌大量学生不常用的专业词汇,使得整个教材通俗易懂。

4. 在内容组织上,以一个典型的数据库应用系统(简化的医院管理数据库)为案例,在 MS SQL Server 平台上,以理论和实际相结合的方式,讲解数据库的概念和应用开发技术,以期达到高效的学习效果。

5. 本教材的内容分为 4 部分:① 数据库基础知识,讲述数据库的基本概念和理论知识,这部分是基础,面向所有读者;② 数据库管理技术,讲述数据库维护管理技术,面向关心数据库维护的读者;③ 数据库应用技术,讲述数据库的设计开发技术,面向关心行业应用系统开发的读者;④ 现代数据库技术及主流数据库产品介绍,介绍数据库技术的前沿热点及主要的数据库管理系统的商业产品,面向关心数据库发展动向及实际数据库产品的读者。

6. 本教材注重将计算机理论知识和现实中的工程应用相结合,适当引入数据库技术的最新发展,保持了教学内容的先进性和实用性。本书源于基础教育的教学实践及科研实践中的许多心得体会,凝聚了工作在教学和科研第一线教师多年的教学与科研成果。

通过学习本书，你可以了解：

- 数据库的概念、组成结构等基础知识；
- 关系数据库的核心——关系数据模型及数据库的操作语言 SQL；
- 数据库的完整性、安全性、恢复、并发控制等数据库管理技术；
- 数据库的应用设计方法和开发技术；
- 现代数据库技术的热点——数据挖掘、数据仓库及空间数据库；
- 目前的主流数据库管理系统产品——Oracle、SQLServer、DB2、MySQL 及 Sybase 的概况。

教学中，可以根据教学对象和学时等具体情况对书中的内容进行删减和组合，也可以进行适当扩展，参考学时为 32~64 学时。为适应教学模式和教学方法的改革，本教材每章配套安排了习题及参考答案、多媒体电子课件及相应的网络教学资源，请登录华信教育资源网 <http://www.hxedu.com.cn> 注册下载。

本书第 1、2、12、13、14 章由魏祖宽编写，第 3、4、5、6、7 章由郑莉华编写，第 8、9、10、11 章、案例数据库、习题及附录的习题答案由胡旺编写。全书由魏祖宽统稿并定稿。参加本书编写的还有电子科技大学的胡俊杰、张彪、张鑫、惠宇、周益民、朱小林、代林、胡红梅、陈佳、张乐信、刘小龙、张江泓，他们在收集基本素材、案例数据、相关技术资料及稿件校对等基础方面承担了大量工作。

本书在编写过程中参考了大量新近出版的相关资料和书籍，吸取了许多专家和同仁的宝贵经验，在此向他们深表谢意！

由于数据库应用技术发展迅速，作者学识有限，书中难免存在误漏之处，望广大读者批评指正。

作者

# 目 录

第 1 章 数据库系统概论	1	2.5 习题	41
1.1 数据库系统的应用及其研究领域	2	第 3 章 关系数据库标准语言 SQL	43
1.2 数据库系统概述	3	3.1 SQL 简介	44
1.2.1 数据库的基本概念和相关术语	3	3.1.1 SQL 的发展史	44
1.2.2 数据库技术的产生和发展	6	3.1.2 SQL 提供的功能	44
1.2.3 数据库系统的特点	8	3.1.3 SQL 的特点	45
1.3 数据模型	9	3.2 数据定义	46
1.3.1 概念模型	10	3.2.1 数据库的创建、修改与删除	46
1.3.2 数据模型	10	3.2.2 基本表的定义、修改与删除	49
1.4 数据库系统的结构与组成	11	3.3 数据查询	53
1.4.1 数据库系统的体系结构	11	3.3.1 格式	53
1.4.2 数据库系统的模式结构	12	3.3.2 简单查询	54
1.4.3 数据库系统的一般组成	14	3.3.3 连接查询	61
1.5 数据库语言	15	3.3.4 用 TOP 限制结果集	65
1.5.1 四大数据库语言	15	3.3.5 嵌套查询	65
1.5.2 结构化查询语言 SQL	16	3.3.6 集合查询	68
1.6 数据库系统的发展趋势及前沿技术	16	3.4 数据更新	69
1.7 习题	18	3.4.1 插入	69
第 2 章 关系数据模型	19	3.4.2 修改	70
2.1 数据模型	20	3.4.3 删除	71
2.1.1 概述	20	3.5 视图	71
2.1.2 概念模型	20	3.5.1 视图的基本概念	72
2.1.3 数据模型的基本要素	23	3.5.2 视图的定义与删除	72
2.1.4 数据模型的发展	24	3.5.3 视图的查询	74
2.2 关系数据模型	26	3.5.4 视图的更新	74
2.2.1 基本概念	27	3.5.5 视图的作用	75
2.2.2 关系数据模型的数据结构	27	3.6 索引	76
2.2.3 数据操作	29	3.6.1 索引的概念	76
2.2.4 数据约束	30	3.6.2 索引的分类	77
2.2.5 关系数据模型的优缺点	30	3.6.3 索引的创建与删除	78
2.3 关系	31	3.6.4 建立索引的原则	79
2.3.1 域、笛卡儿积和关系	31	3.7 习题	79
2.3.2 关系的性质	32	第 4 章 完整性约束	81
2.3.3 关系模式	32	4.1 数据库完整性概述	82
2.3.4 关系完整性	33	4.2 完整性约束的分类	82
2.4 关系代数和关系演算	34	4.3 完整性约束的定义	82
2.4.1 关系代数	34	4.3.1 Primary Keys 约束	82
2.4.2 关系演算	40	4.3.2 UNIQUE 约束	83
		4.3.3 NOT NULL 约束	84



4.3.4	CHECK 约束	84	6.5.4	RAID 的应用	129
4.3.5	Foreign-Key 约束	86	6.6	习题	129
4.3.6	域约束	89	<b>第7章</b>	<b>并发控制</b>	<b>130</b>
4.3.7	断言	89	7.1	事务的并发执行	131
4.4	完整性约束的修改	90	7.2	并发执行可能引起的问题	131
4.4.1	约束的命名	90	7.3	可串行化	132
4.4.2	约束的修改	90	7.3.1	串行调度	132
4.5	完整性约束的验证	91	7.3.2	可串行化调度	133
4.6	数据库产品对完整性的支持	93	7.3.3	可恢复性	136
4.6.1	Oracle 的完整性约束	94	7.4	基于锁的并发控制协议	137
4.6.2	DB2 的完整性约束	95	7.4.1	封锁	137
4.6.3	SQL Server 2000 的完整性约束	96	7.4.2	两段锁协议(2PL)	139
4.7	习题	97	7.4.3	锁的升级及更新锁	140
<b>第5章</b>	<b>数据库安全</b>	<b>98</b>	7.5	活锁与死锁	141
5.1	计算机安全概述	99	7.5.1	活锁	141
5.2	数据库安全性控制	100	7.5.2	死锁的形成	141
5.3	自主访问控制(DAC)	101	7.5.3	死锁的预防	142
5.3.1	权限	101	7.5.4	死锁的检测及处理	143
5.3.2	授权	102	7.6	多粒度封锁	144
5.3.3	授权图	103	7.7	习题	145
5.3.4	授权和视图	105	<b>第8章</b>	<b>关系数据库设计理论</b>	<b>146</b>
5.3.5	角色	106	8.1	关系模式设计中的问题	147
5.3.6	收回权限	106	8.2	函数依赖	148
5.4	强制访问控制(MAC)	107	8.2.1	函数依赖的定义	149
5.5	跟踪审计	108	8.2.2	FD 的逻辑蕴涵	150
5.6	数据库安全性的其他技术	109	8.2.3	FD 的 Armstrong 公理	150
5.6.1	数据加密	109	8.2.4	函数依赖与码的联系	151
5.6.2	鉴定	109	8.2.5	属性集的闭包	152
5.7	习题	111	8.2.6	FD 推理规则的完备性	152
<b>第6章</b>	<b>数据库恢复技术</b>	<b>112</b>	8.2.7	FD 集的最小依赖集	153
6.1	事务概念	113	8.3	模式分解	154
6.2	数据库恢复概述	115	8.3.1	模式分解问题	154
6.3	恢复的实现技术	115	8.3.2	无损分解	154
6.3.1	日志	115	8.3.3	保持函数依赖的分解	157
6.3.2	更新事务的执行与恢复	116	8.3.4	模式分解与模式等价问题	159
6.3.3	检查点	120	8.4	规范化	160
6.3.4	数据转储	122	8.4.1	第一范式(1NF)	160
6.4	故障的种类及恢复策略	123	8.4.2	第二范式(2NF)	161
6.4.1	故障的分类	123	8.4.3	第三范式(3NF)	161
6.4.2	不同故障的恢复策略	123	8.4.4	BC 范式(BCNF)	162
6.5	RAID	126	8.4.5	模式设计的原则	163
6.5.1	RAID 的优点	126	8.5	多值函数依赖与 4NF	164
6.5.2	RAID 的分级	126	8.5.1	多值函数依赖	164
6.5.3	RAID 级别的选择	128	8.5.2	FD 与 MVD 的推理规则集	165



8.5.3	4NF	165	10.4	触发器	210
8.6	连接依赖和 5NF	165	10.4.1	触发器的概念	210
8.7	本章小结	166	10.4.2	触发器的作用	211
8.8	习题	166	10.4.3	触发器的工作过程	211
			10.4.4	触发器的操作	212
<b>第 9 章</b>	<b>数据库设计方法</b>	<b>168</b>	10.5	嵌入式 SQL	215
9.1	数据库设计概述	169	10.5.1	ESQL 的处理过程	215
9.1.1	数据库的设计方法	169	10.5.2	ESQL 与主语言之间的通信	215
9.1.2	数据库开发生命周期方法	171	10.5.3	ESQL 的使用规定	215
9.1.3	数据库设计的基本过程	172	10.5.4	ESQL 的使用技术	216
9.2	数据库需求分析	175	10.6	数据库的连接访问	217
9.2.1	需求描述与分析	175	10.6.1	ODBC	217
9.2.2	需求分析的步骤	176	10.6.2	JDBC	219
9.2.3	数据字典	179	10.7	XML 数据库	219
9.3	概念数据库设计	180	10.7.1	XML 数据模型	220
9.3.1	概念数据库设计的必要性	180	10.7.2	XML 数据库	220
9.3.2	概念设计的方法和步骤	181	10.8	本章小结	222
9.3.3	概念设计工具: E-R 图	182	10.9	习题	222
9.4	数据库逻辑结构设计及优化	190	<b>第 11 章</b>	<b>数据仓库技术</b>	<b>223</b>
9.4.1	逻辑设计环境	190	11.1	商业智能	224
9.4.2	E-R 模式向关系模型的转换	191	11.1.1	商业智能是什么	224
9.4.3	用关系规范化理论对关系数据模型进行优化	191	11.1.2	商业智能过程	224
9.5	数据库的物理设计	192	11.1.3	商业智能的体系结构	225
9.5.1	数据库的物理结构确定	192	11.1.4	商业智能应用	226
9.5.2	物理结构的评价	193	11.1.5	商业智能工具	227
9.6	数据库的実施和运行维护	194	11.2	数据仓库的概念与特征	228
9.6.1	实际数据库结构的建立	194	11.2.1	数据仓库的产生背景	228
9.6.2	装入数据	194	11.2.2	数据仓库的概念	229
9.6.3	应用程序开发和调试	194	11.2.3	数据仓库与传统数据库的比较	230
9.6.4	数据库试运行	195	11.3	数据仓库的体系结构	231
9.6.5	数据库运行和维护	195	11.3.1	数据仓库系统的总体结构	231
9.7	习题	196	11.3.2	数据仓库系统的分层结构	233
			11.3.3	带有 ODS 的数据仓库体系结构	234
<b>第 10 章</b>	<b>数据库高级开发技术</b>	<b>197</b>	11.4	多维数据模型	235
10.1	游标	198	11.4.1	数据立方体	235
10.1.1	游标种类	198	11.4.2	星型模式	236
10.1.2	游标操作	198	11.4.3	雪花模式	237
10.2	存储过程	201	11.4.4	事实星座模式	237
10.2.1	存储过程的概念	201	11.5	数据仓库的 ETL 过程	238
10.2.2	存储过程的种类	202	11.6	联机分析处理	238
10.2.3	存储过程的操作	203	11.6.1	OLAP 的基本概念	238
10.3	函数	205	11.6.2	OLAP 的基本操作	238
10.3.1	SQL 中函数的概念	205	11.6.3	OLAP 的特点	239
10.3.2	函数类型	206			
10.3.3	函数的操作	206			



11.6.4	OLAP 与 OLTP 的比较	239	13.1.2	GIS 的组成	267
11.7	数据仓库结构设计	240	13.1.3	GIS 的主要用途	270
11.7.1	数据仓库的工作过程	240	13.2	GIS 的主要功能和行业应用	271
11.7.2	数据仓库系统开发过程	241	13.2.1	GIS 的主要功能	271
11.7.3	数据仓库的元数据设计	243	13.2.2	GIS 的应用领域	274
11.7.4	数据仓库性能	243	13.2.3	GIS 行业应用案例——配电网 可视化管理系统	275
11.8	本章小结	244	13.3	主流 GIS 和空间数据库产品	279
11.9	习题	244	13.3.1	GIS 的发展和国内外的 GIS 产品	279
<b>第 12 章</b>	<b>数据挖掘技术</b>	<b>245</b>	13.3.2	ArcGIS	280
12.1	数据挖掘概述	246	13.3.3	SuperMap GIS	281
12.1.1	数据挖掘的起源和发展	246	13.3.4	MapGIS	283
12.1.2	数据挖掘的定义和任务	246	13.4	习题	284
12.1.3	一个典型的数据挖掘过程	247	<b>第 14 章</b>	<b>主流数据库产品介绍</b>	<b>285</b>
12.2	数据挖掘的传统技术	247	14.1	Oracle 数据库	286
12.2.1	数据挖掘的数据对象	247	14.1.1	Oracle 简介	286
12.2.2	数据挖掘发现的知识模式	248	14.1.2	Oracle 发展简史	286
12.2.3	数据挖掘的主要技术	249	14.1.3	Oracle 特点	287
12.3	数据挖掘工具及其评价标准	250	14.1.4	Oracle 11g 新特性	287
12.3.1	数据挖掘工具的分类	251	14.1.5	Oracle 体系结构	289
12.3.2	几种常用的数据挖掘工具	251	14.1.6	Oracle 开发套件	290
12.3.3	数据挖掘工具的评价标准	256	14.2	MS SQL Server 数据库	292
12.3.4	数据挖掘工具的选择原则	256	14.2.1	SQL Server 简介	292
12.4	数据挖掘技术的新发展	257	14.2.2	SQL Server 发展简史	292
12.4.1	文本数据挖掘	257	14.2.3	SQL Server 2008 新特性	292
12.4.2	Web 数据挖掘	258	14.2.4	SQL Server 系统结构	294
12.4.3	可视化数据挖掘	258	14.2.5	SQL Server 的工具和组件	294
12.4.4	空间数据挖掘	259	14.3	DB2 数据库	295
12.4.5	分布式数据挖掘	260	14.3.1	DB2 数据库简介	295
12.5	数据挖掘及其他系统的关系	260	14.3.2	DB2 发展简史	296
12.5.1	与数据仓库集成的数据挖掘 体系	260	14.3.3	DB2 核心数据库特点	296
12.5.2	数据仓库与数据挖掘的关系	261	14.3.4	DB2 工具	296
12.5.3	数据仓库与专家系统、OLAP 的关系	262	14.3.5	WebSphere 应用服务器	297
12.6	数据挖掘的应用现状及前景	262	14.3.6	开发工具	298
12.6.1	商业零售行业	263	14.4	MySQL 数据库	298
12.6.2	金融和保险服务行业	263	14.4.1	MySQL 数据库简介	298
12.6.3	科学研究领域	263	14.4.2	MySQL 发展简史	298
12.6.4	电信网络管理	263	14.4.3	MySQL 特点	299
12.6.5	其他主要应用领域	264	14.4.4	MySQL 工具	299
12.7	习题	264	14.5	数据库产品的简略比较	300
<b>第 13 章</b>	<b>地理信息系统和空间数据库</b>	<b>265</b>	<b>附录 A</b>	<b>HIS 案例</b>	<b>303</b>
13.1	GIS 和空间数据库概念	266	<b>参考文献</b>		<b>309</b>
13.1.1	什么是 GIS	266			

# 第1章 数据库系统概论

数据库是数据管理的最新技术，是计算机学科的重要分支。十余年来，数据库管理系统已从专业的应用程序包发展成为通用的系统软件。由于数据库系统具有数据结构化、最低冗余度、较高的程序与数据独立性等优点，较大的信息管理系统都是以数据库作为基础的。数据管理技术的发展经历了3个阶段：

- 人工管理阶段
- 文件系统阶段
- 数据库系统阶段

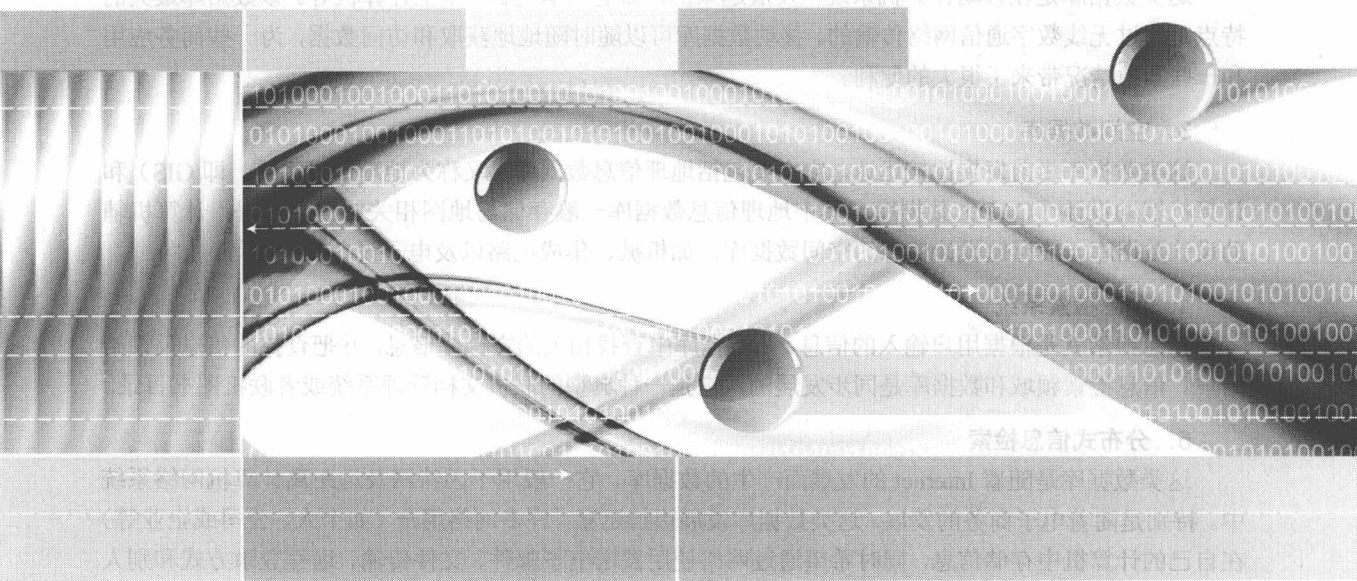
数据库作为数据管理技术发展的最新阶段，其体系结构由3级模式构成：

- 模式
- 外模式
- 内模式

在这一章里，我们将学习数据库系统的基本概念和术语，了解数据管理技术的发展历史，学习数据库系统的体系结构，包括三级模式与两级映像，了解数据库的语言，为后续的学习打下一个良好的基础。

学习目标：

- 掌握数据库的基本概念和相关术语
- 了解数据库技术的产生与发展
- 掌握数据库管理技术发展的三个阶段
- 理解数据库中的各种数据模型
- 了解数据库系统的体系结构和一般组成
- 理解数据库系统的模式结构
- 了解数据库的各种语言





数据库和数据库系统已经成为现代社会日常生活中的重要组成部分,在日常的工作和生活中,大多数人都会或多或少地与数据库打交道。比如,到银行存钱或取钱;预定机票或宾馆房间;在图书馆的计算机管理系统中查找图书条目;或者从网上商店购买商品,所有这些活动都会涉及访问数据库系统。

在传统的数据库应用中,大多数信息都是以文本或数字形式来存储和访问的。随着技术的发展,不断出现许多新的数据库系统应用。多媒体数据库可以存储图片、视屏片段以及语音消息。地理信息系统可以存储和分析地图、气象数据和卫星图像。许多公司和政府机构使用数据仓库和联机分析处理来提取、分析大型数据库中的有用信息以辅助决策。实时数据库和主动数据库技术则用于控制工业和制造业的生产过程。数据库搜索技术还用到万维网上,改善信息搜索,满足互联网用户查找信息的要求。在本章,我们将介绍数据库应用的基础知识,为后续的学习打下基础。



### 1.1 数据库系统的应用及其研究领域

数据库的应用领域非常广泛,不管是学校、公司或大型企业,还是政府部门,都需要使用数据库来存储数据信息。例如,学校要存储学生的基本信息、课程信息和成绩等;医院可以用数据库来存储医生、病人、药品的基本信息,以及病人的就诊信息等;银行需要用数据库来存储客户的信息、账户以及银行的交易记录;电信业需要数据库来存储通话记录,产生每月清单,维护预付电话卡的余额和存储通信网络的信息;金融业需要数据库来存储股票、债券等金融票据的持有、出售和买入的信息……数据库已经渗入到社会日常生活中的各个方面。

传统数据库中的很大一部分用于商务领域,如证券行业、银行、销售部门、医院、公司或企业单位,以及国家政府部门、国防军工领域、科技发展领域等。随着信息时代的发展,数据库也相应产生了一些新的应用领域,主要表现在下面6个方面。

#### 1. 多媒体数据库

这类数据库主要存储与多媒体相关的数据,如声音、图像和视频等数据。多媒体数据最大的特点是数据连续,而且数据量比较大,存储需要的空间较大。

#### 2. 移动数据库

这类数据库是在移动计算机系统上发展起来的,如笔记本电脑、掌上计算机等。该数据库最大的特点是通过无线数字通信网络传输的。移动数据库可以随时随地地获取和访问数据,为一些商务应用和一些紧急情况带来了很大的便利。

#### 3. 空间数据库

这类数据库目前发展比较迅速。它主要包括地理信息数据库(又称为地理信息系统,即GIS)和计算机辅助设计(CAD)数据库。其中地理信息数据库一般存储与地图相关的信息数据;计算机辅助设计数据库一般存储设计信息的空间数据库,如机械、集成电路以及电子设备设计图等。

#### 4. 信息检索系统

信息检索就是根据用户输入的信息,从数据库中查找相关的文档或信息,并把查找的信息反馈给用户。信息检索领域和数据库是同步发展的,它是一种典型的联机文档管理系统或者联机图书目录。

#### 5. 分布式信息检索

这类数据库是随着Internet的发展而产生的数据库。它一般用于因特网及远距离计算机网络系统中。特别是随着电子商务的发展,这类数据库发展更加迅猛。许多网络用户(如个人、公司或企业等)在自己的计算机中存储信息,同时希望通过网络使用发送电子邮件、文件传输、远程登录方式和别人共享这些信息。分布式信息检索满足了这一要求。

## 6. 专家决策系统

专家决策系统也是数据库应用的一部分。由于越来越多的数据可以联机获取，特别是企业通过这些数据可以对企业的发展做出更好的决策，以使企业更好地运行。由于人工智能的发展，使得专家决策系统的应用更加广泛。



## 1.2 数据库系统概述



### 1.2.1 数据库的基本概念和相关术语

在系统地介绍数据库的基本概念之前，这里首先介绍一些数据库最常用的术语和基本概念。

#### 1. 数据、信息与数据处理

在计算机应用中，数据处理和以数据处理为基础的信息系统占据着很大的比重。

人类的一切活动都离不开数据，离不开信息。在不同的领域里，信息的含义有所不同。一般认为信息是数据、消息中包含的意义。数据和信息有时可以混用，例如，数据处理也称为信息处理；有时必须分清，例如，不能把信息系统称为数据系统。

##### (1) 数据 (Data)

数据是一种符号序列，它的内容是事物特性的反映。数据是对现实世界的事物采用计算机能够识别、存储和处理的方式进行的描述，或者说就是计算机化的信息。数据的概念在数据处理领域中得到不断的发展，目前数据不仅包括数字、字母、文字和其他特殊字符，而且还包括图形、图像、声音等多媒体数据。

在计算机中，为了存储和处理现实世界中的事物，就要抽出对这些事物感兴趣的特征组成一个记录来描述。例如，在员工档案中，如果人们最感兴趣的是员工的姓名、性别、年龄、籍贯、所在部门、进入公司的时间和薪水，那么可以这样描述：

(余颖，女，20，四川，销售部，2008，3000)

这里的员工记录就是数据。对于上面这条员工记录，了解其含义的人会得到如下信息：余颖是公司职员，女，20岁，四川人，2008年进入公司，月薪3000元。而不了解其语意的人则无法理解其含义。

##### (2) 信息

信息是经过加工处理的数据，是人们消理解了的数据，是数据的具体含义。数据与信息既有联系又有区别。数据是信息的符号表示或载体，而信息则是数据的具体含义，对数据的语义解释。而且同一数据也可能有不同的解释。数据一般都可以表示成某种信息，但并非任何数据都能包含对人们来说有用的信息。信息是抽象的，不随数据设备所决定的数据形式而变化；而数据的表现形式却具有可选择性。

信息是反映客观现实世界的知识，用不同的数据形式可以表示同样的信息。例如，同样的新闻可以通过报纸、电台和电视来报道，它的表现形式不同，但其信息的内容可以相同。

##### (3) 数据处理

数据处理是指将数据转换成信息的过程。广义地讲，它包括对数据的收集、存储、加工、分类、检索、传播等一系列活动。狭义地讲，它是指对所输入的数据进行加工整理。基本目的是从大量、已知数据出发，根据事物之间的固有联系和规律，通过分析归纳、演绎推导等手段，提出对人们有价值、有意义的信息，作为决策的依据。数据的加工可以简单，也可以相当复杂。简单加工包括组织、编码、分类、排序等；复杂加工可以使用统计学方法、数学模型等对数据进行深层次的加工。

因此，可得：信息 = 数据 + 数据处理。

## 2. 数据库的基本概念

数据库是什么呢?顾名思义,即存放数据的仓库。只不过这个仓库是在计算机存储设备上,而且数据是按一定的格式存放的。实质上,数据库就是一些存在了很长时间——常常是许多年的信息的聚集。通常意义下,“数据库”这个术语是指由数据库管理系统管理的数据集合。

### (1) 数据库

数据库(DB, Database)是指长期存储在计算机内的、有组织的、可共享的大量数据的集合。数据库中的数据按照一定的数据模型组织、描述和存储,具有较小的冗余度、较高的数据独立性和易扩展性,并为各种用户共享。

### (2) 数据库管理系统

了解了数据和数据库的概念,下一个问题就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。完成这个任务的即是数据库管理系统(DBMS)。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件,主要目标是使数据成为方便各种用户使用的资源,并提高数据的安全性、完整性和可用性。

DBMS一般具有如下功能:

① 数据定义功能,提供数据定义语言DDL(Data Definition Language),对各级数据模式进行精确定义。

② 数据操纵功能,提供数据操纵语言DML(Data Manipulation Language),可以对数据库中的数据进行追加、插入、修改、删除、检索等操作。

③ 数据库运行控制功能,提供数据控制语言DCL(Data Control Language),可以对数据库中的数据进行并发控制、数据的安全性控制、数据的完整性控制。

④ 数据组织、存储和管理功能,DBMS分类组织、存储和管理各种数据,包括数据字典、用户数据、数据的存取路径。确定以何种文件结构和存取方式在存储级上组织这些数据,实现数据之间的联系。

⑤ 数据库的建立和维护功能,主要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,数据库的重组功能和性能监视、分析功能等。这些功能通常是由一些实用程序或管理工具完成的。

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

数据库系统是由数据库及其管理软件组成的系统,一般由数据库、数据库管理系统、数据库管理员(DBA)、用户和应用程序组成。一个数据库系统应该具有如下特征:

① 允许用户使用一种叫作数据定义语言的专用语言,建立新的数据库和指定它们的模式(schema)。

② 使用户能够用适当的语言查询数据和更新数据。

③ 支持存储大量的数据——G( $10^9$ )字节以上,经过很长一段时间以后仍保证其安全,同时允许对数据库合法的查询和更新。

④ 控制多用户的同时访问,使得一个用户的访问不影响其他用户,保证同时访问不会损害数据。

上面介绍了与数据库相关的一些基本的、抽象的概念,接下来介绍一些具体概念。

### (4) 实体与表

在数据库系统中,一个实体就是一个人、一个地方、一个事件,或者一个我们将要为其收集数据的物体。客观存在并可相互区别的事物称为实体。比如,在学校中,学生、教师员工、课程等都是实体。学生可以组成一个实体集。一个实体集就是实体的一个特定集合。每个实体都有某些称为属性的

特征。如学生实体可能包含以下属性：学生学号、姓名、性别、入学时间、专业方向等。每个属性必须恰当地命名，以便让用户能够知道它的内容，比如学生实体，属性姓名可以存储为 STU\_NAME，性别可以存储为 STU\_SEX。

表是二维结构，它包括行和列。一个表包括一组相关的实体——实体集。因此，术语实体集和表经常互换使用。表 1.1 列出了一个表的特征。

表 1.1 表的特征

1	表是二维结构，它包括行和列
2	每个表行描述实体集中的一个实体
3	每个表列描述一个属性，每一列有一个明确的名字
4	每行与列的交叉描述一个数据值
5	每个表必须有一个属性或者一个属性复合体来唯一标识每一行
6	一列中的所有值必须是同一数据格式
7	每列有一个明确的数值范围
8	行和列的排列顺序对 DBMS 并不重要

### (5) 关系列表和关系数据库

数据库是一个自描述的数据元素的集合以及这些元素之间的联系，数据库中表之间的关系是数据元素间的重要联系。

#### ① 关系列表

关系数据库使用表来组织数据元素，每一个表对应于一个应用实体，而每行则代表实体的一个示例。例如，医院管理系统中医生实体对应数据库中的表 Doctor，该表中的每一行则代表不同的医生。医生数据表的属性构成见表 1-2。

表 1.2 医生数据表的属性构成

医生编号	医生姓名	医生性别	医生年龄	所属部门	技术等级	工资
------	------	------	------	------	------	----

联系通过将来自于一个表的行标识符（医生编号）出现在一个表示诊断的行中，从而建立了该诊断与这位医生之间的联系。诊断数据表的属性构成见表 1.3。

表 1.3 诊断数据表的属性构成

诊断编号	患者编号	医生编号	症状描述	诊断描述	就诊时间
------	------	------	------	------	------

这种表之间通过属性进行联系，构成了关系列表，是关系数据库的一个基础。

#### ② 关系数据库

1970 年，IBM 研究室的 Ted Codd 发表了一篇具有很大影响的关于关系数据库模型的论文，数据库系统发生了显著的变化。Codd 提出数据库系统应为用户提供这样一种观点，即数据库系统是用一种称为“关系”的表来组织数据的。而在背后，可能有一个很复杂的数据结构，以保证对各种查询的快速响应。但与以前的数据库系统的用户不同，关系数据库系统的用户并不关心数据的存储结构，而是使查询能用很高级的语言来实现，从而大大提高了数据库开发人员的效率。

下面先对关系做一下简单介绍，增进读者对关系模型的了解，同时，我们给出一个 SQL 的例子，以便读者了解关系模型如何支持高级语言的查询。

**【例 1-1】**关系就是表。表的各列以属性开始，属性是列的入口。表 1.4 是一个名为 Doctor（医生）的关系，记录的是医生的信息。

表 1.4 医生数据表

医生编号	医生姓名	医生性别	医生年龄	所属部门	技术等级	工资
001	李红	女	30	101	主任医师	3000
002	张兰	女	35	104	主治医师	1800

假设要查询李红的工资，SQL 的查询语句如下：

```
SELECT 工资
FROM   Doctor
WHERE  医生姓名='李红';
```

## 1.2.2 数据库技术的产生和发展

数据库技术是应数据库管理任务的需要而产生的。数据管理是指对各种数据进行分类、组织、编码、存储、检索和维护,它是数据处理的核心问题。人们借助计算机进行数据处理是从20世纪60年代开始的。

在应用需求的推动下,在计算机硬件、软件发展的基础上,数据管理技术经历了人工管理、文件系统、数据库系统三个阶段。

### 1. 人工管理阶段

在这一阶段(20世纪50年代中期以前),计算机主要用于科学计算。外部存储器只有磁带、卡片和纸带等,还没有磁盘等直接存取存储设备。软件只有汇编语言,尚无数据管理方面的软件,数据处理方式基本是批处理。这个阶段有如下特点:

① 计算机系统不提供对用户数据的管理功能。用户编制程序时,必须全面考虑好相关的数据,包括数据的定义、存储结构以及存取方法等。程序和数据是一个不可分割的整体。数据脱离了程序就无任何存在的价值,数据无独立性。

② 数据不能共享。不同的程序均有各自的数据,这些数据对不同的程序通常是不相同的,不可共享。即使不同的程序使用了相同的一组数据,这些数据也不能共享,程序中仍然需要各自加入这组数据,谁也不能省略。基于这种数据的不可共享性,必然导致程序与程序之间存在大量的重复数据,增加了数据管理的复杂性,浪费了存储空间。

③ 不单独保存数据。基于数据与程序是一个整体,数据只为本程序所使用,数据只有与相应的程序一起保存才有价值,否则就毫无用处。所以,所有程序的数据均不单独保存。

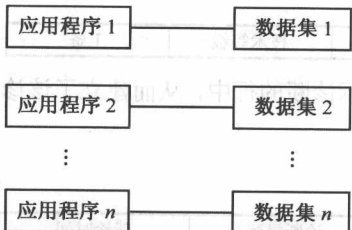


图 1.1 人工管理阶段应用程序与数据之间的对应关系

在人工管理阶段,程序与数据之间的一一对应关系可用图 1.1 表示。

### 2. 文件系统阶段

在这一阶段(20世纪50年代后期至60年代中期),计算机不仅用于科学计算,还用在信息管理方面。随着数据量的增加,数据的存储、检索和维护问题的解决成为紧迫的需要,数据结构和数据管理技术迅速发展起来。此时,外部存储器已有磁盘、磁鼓等直接存取的存储设备。软件领域出现了操作系统和高级软件。操作系统中的文件系统是专门管理外存的数据管理软件,文件是操作系统管理的重要资源之一。数据处理方式有批处理,也有联机实时处理。这个阶段有如下特点:

① 数据以“文件”形式可长期保存在外部存储器的磁盘上。由于计算机的应用转向信息管理,因此对文件要进行大量的查询、修改和插入等操作。

② 数据的逻辑结构与物理结构有了区别,但比较简单。程序与数据之间具有“设备独立性”,即程序只需用文件名就可与数据打交道,不必关心数据的物理位置。由操作系统的文件系统提供存取方法。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,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。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。文件结构的设计仍然是基于特定的用途。程序基于特定的物理结构和存取方法,因此程序与数据结构之间的依赖关系并未根本改变,对数据的操作以记录为单位。这是由于文件中只存储数据,不存储文件记录的结构描述信息。文件的建立、存取、查询、插入、删除、修改等所有操作,都要用程序来实现。



随着数据管理规模的扩大,数据量急剧增加,文件系统显露出一些缺陷:

① 数据冗余。由于文件之间缺乏联系,造成每个应用程序都有对应的文件,有可能同样的数据在多个文件中重复存储。

② 不一致性。这往往是由数据冗余造成的,在进行更新操作时,稍不谨慎,就可能使同样的数据在不同的文件中不一样。

③ 数据联系弱。这是由于文件之间相互独立、缺乏联系造成的。

文件系统阶段是数据管理技术发展中的一个重要阶段。在这一阶段中,得到充分发展的数据结构和算法丰富了计算机科学,为数据管理技术的进一步发展打下了基础,现在仍是计算机软件科学的重要基础。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,修改文件结构的定义。应用程序的改变,例如程序改变不同的高级语言等,也将引起文件的数据结构的改变。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备弹性的无结构数据集合,即文件之间是孤立的,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。

### 3. 数据库系统阶段

现实世界是复杂的,反映现实世界的各类数据之间必然存在错综复杂的联系。为反映这种复杂的数据结构,让数据资源能为多种应用需要服务,并为多个用户所共享,同时为了让用户能更方便地使用这些数据资源,在计算机科学中,逐渐形成了数据管理这一独立分支。计算机中的数据及数据的管理统一由数据库系统来完成。为此,数据管理技术进入数据库系统阶段。

数据库系统克服了文件系统的缺陷,提供了对数据更高级、更有效的管理。这个阶段的程序和数据的联系通过数据库管理系统(DBMS)来实现,如图 1.3 所示。

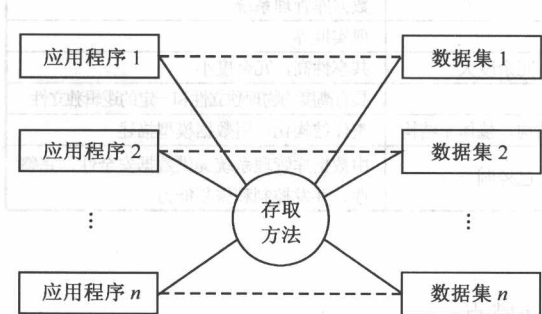


图 1.2 文件系统阶段应用程序与数据之间的对应关系

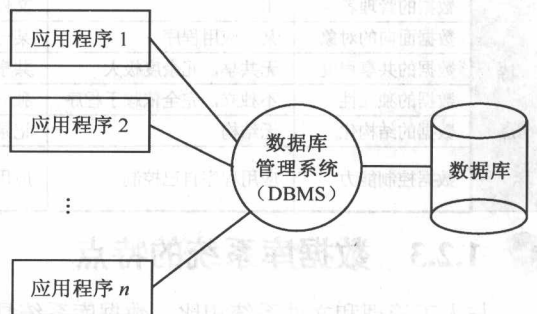


图 1.3 数据库系统阶段应用程序与数据之间的对应关系

概括起来,数据库系统阶段的数据管理具有以下特点:

#### (1) 采用数据模型表示复杂的数据结构

数据模型不仅描述数据本身的特征,还要描述数据之间的联系,这种联系通过存取路径实现。通过所有存取路径表示自然的数据联系是数据库与传统文件的根本区别。这样,数据不再面向特定的某个或多个应用,而是面向整个应用系统。数据冗余明显减少,实现了数据共享。

#### (2) 有较高的数据独立性

数据的逻辑结构与物理结构之间的差别可以很大。用户以简单的逻辑结构操作数据而无须考虑数据的物理结构。数据库的结构分成用户的局部逻辑结构、数据库的整体逻辑结构和物理结构三级。用户的数据和外存中的数据之间转换由数据库管理系统实现。