

新世纪高等学校计算机系列教材

# 数据库原理与设计

张龙祥 黄正瑞 龙 军 编著

邵佩英 审校

人民邮电出版社

图书在版编目(CIP)数据

数据库原理与设计/张龙祥, 黄正瑞, 龙军编著.北京: 人民邮电出版社, 2002.7

新世纪高等学校计算机系列教材

ISBN 7-115-10325-9

I. 数... II. ①张...②黄...③龙... 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(2002)第 048920 号

新世纪高等学校计算机系列教材

数据库原理与设计

◆ 编 著 张龙祥 黄正瑞 龙 军

审 校 邵佩英

责任编辑 邹文波

执行编辑 苗 颖

◆ 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>

读者热线: 010-67180876

北京汉魂图文设计有限公司制作

印刷厂印刷

新华书店总店北京发行所经销

◆ 开本: × 1/16

印张: 22.75

字数: 540 千字 2002 年 7 月第 1 版

印数: 1-0 000 册 2002 年 7 月北京第 1 次印刷

ISBN 7-115-10325-9/TP · 2890

定价: 28.00 元

本书如有印装质量问题, 请与本社联系 电话:(010) 67129223

# 新世纪高等学校计算机系列教材

## 编委会

主 任：袁开榜

执行副主任：张凤祥

副 主 任：杨学良 唐朔飞 朱庆生 王志英 张大方 刘乃琪  
吕晓春 李伟华 李茂青 张为群 邓亚平

编 委：程佩青 陈大正 陈道员 陈 旭 陈文伟 陈志国  
傅献祯 韩启明 蒋 伟 汲伟民 林 勇 石建斌  
王爱军 王 虹 王汝传 王晓丹 相万让 俞思伟  
余勇泉 邹桂章

## 序

新世纪伊始，以知识经济为主体的信息时代向人们揭示：成为时代宠儿的信息技术人才异常匮乏，正被世界各国相互争夺。据资料介绍，目前全球还缺少软件技术人才 40 余万人，美国每年需要增加 9.5 万名新的电脑专家。在我国，信息技术人才更是奇缺，每年约需补充 20 万名软件技术人才。国家在 35 所重点高校设置示范性软件学院，正好体现了我国奋起直追的战略决策。

为了推动我国计算机教育事业的发展，在美国 ACM/IEEE-CS 于 2001 年 12 月 15 日发布“CC2001 教程”之后，我国全国高等学校计算机教育研究会和中国计算机学会即将公布中国的“CCC2002 计划”之际，全国高等学校计算机教育研究会和全国计算机继续教育研究会，与人民邮电出版社友好协商，本着相互尊重，平等合作，各自发挥自身优势，共同发展的原则，组织出版了这套《新世纪高等学校计算机系列教材》。

这套教材是根据我国高等学校计算机专业(或非计算机专业)当前的课程设置、教学大纲的要求、学时安排及计算机技术的发展趋势而编写的。遴选的作者都是长期在教学第一线执教的教师，他们既有丰富的教学经验，又有出色的科研成果，能够把握教材的深度和实用性，还有较好的前瞻性。

这套教材虽然立足于当前教学的实际情况，但编委会要求作者尽可能参考美国“CC2001 教程”和我国“CCC2002 计划”提出的新思想，融入科技成果上反映的新技术，充分体现出教材的先进性、科学性和实用性，反映出自己的特色，使之成为内容新颖的精品，能和国际教材接轨。

为了保证这套教材的出版质量，全国高等学校计算机教育研究会、全国计算机继续教育研究会以及参加编审的同志们付出了辛勤的劳动，人民邮电出版社的有关领导及编辑们对这套教材的出版给予了有力的支持和鼓励，特对他们致以衷心的感谢。

希望这套教材的问世能为推动我国计算机教育事业的飞速发展，为培养我国面向 21 世纪科技精英发挥应有的作用。

全国高等学校计算机教育研究会理事长 袁开榜

全国计算机继续教育研究会理事长 张凤祥

2002 年 6 月 5 日

## 前 言

数据库技术是计算机科学技术中发展最快的领域之一。数据库系统已在当代的社会生活中获得了广泛的应用,渗透到了工农业生产、商业、行政管理、科学研究、教育、工程技术和国防军事等各行各业,而且已围绕数据库技术形成了一个巨大的软件产业,即数据库管理系统和各类工具软件的开发与经营。

本书以当前流行的关系数据库技术为中心,以数据库语言 SQL 为重点,着重介绍基于关系数据模型的集中式数据库和网络数据库的原理、设计和应用,同时适当介绍分布式数据库系统、面向对象数据库系统、对象-关系数据库系统和数据仓库的基本概念与技术。

在计算机网络环境下使用数据库系统已经是当前的数据库应用的主流方式,为此本书专列一章(第9章)讨论网络数据库系统,包括客户机/服务器系统和浏览器/服务器系统。

本书介绍的 SQL 的内容以最新的国际标准 SQL3 为准,着重介绍 SQL3 的核心部分和重要的扩充内容,尤其是对面向对象概念的支持。这些新扩充内容目前尚未被 DBMS 产品所完全支持。

本书的内容编排向未来的 2002 计算机教学大纲靠拢,以奠定数据库理论基础、培养数据库开发能力为目标,强调理论与实践紧密结合,数据库原理与设计应用并重,既有数据库基本原理的严谨叙述,又详细说明数据库设计的方法和过程。

本书的参考学时数为 54~74 学时。在材料的组织上本书逻辑隐含三大部分:

第一部分由第 1 章~第 4 章构成,介绍数据库系统的基础知识与应用;

第二部分由第 5 章~第 9 章构成,介绍数据库系统的设计与网络应用;

第三部分由第 10 章~第 12 章构成,介绍分布式数据库系统、面向对象数据库系统和数据仓库。

第一部分和第二部分可作为必修内容(其中个别内容也可作选修),第三部分为选修内容,在实际教学中可以根据具体情况选用。

本书由张龙祥、黄正瑞、龙军合作完成。黄正瑞编写了第 4 章、第 5 章、第 7 章,龙军编写了第 9 章、第 12 章,张龙祥制定

了编写大纲，编写了其余各章，并最后统编全书稿。

在本书的编写过程中得到了全国高等学校计算机教育研究会和全国计算机继续教育研究会的指导与支持，特在此表示衷心的感谢。

数据库技术的发展十分迅速，作者的水平有限，书中难免存在缺点和错误，殷切希望同行专家和读者批评指正。

编 者

# CONTENTS 目录

第 1 章 基础知识.....	1
1.1 数据库系统.....	2
1.1.1 信息世界.....	2
1.1.2 数据管理.....	3
1.1.3 数据库系统的组成.....	5
1.1.4 数据库系统的作用与优越性.....	6
1.2 数据模型.....	7
1.2.1 信息的 3 个世界.....	7
1.2.2 数据模型的特性及分类.....	8
1.2.3 层次模型.....	10
1.2.4 网状模型.....	11
1.2.5 关系模型.....	12
1.2.6 数据模型的相互转换和比较.....	13
1.3 数据库系统的构造.....	14
1.3.1 数据库系统结构.....	14
1.3.2 数据库概念模式.....	15
1.3.3 数据库外模式.....	16
1.3.4 数据库内模式.....	17
1.3.5 数据库管理员.....	17
1.3.6 数据库系统的体系结构.....	18
1.4 数据库管理系统.....	19
1.4.1 什么是数据库管理系统.....	19
1.4.2 数据描述语言.....	20
1.4.3 数据操作语言.....	21
1.4.4 数据库管理例行程序.....	22
1.4.5 数据库管理系统的工作过程.....	23
小结.....	25

习题.....	25
第 2 章 关系数据模型.....	26
2.1 关系数据模型的基本概念与术语.....	26
2.1.1 关系、元组、属性和域.....	27
2.1.2 关键字.....	28
2.1.3 关系模型的数据操作.....	29
2.1.4 数据操作语言.....	29
2.1.5 完整性约束.....	29
2.2 关系代数.....	30
2.2.1 一般的集合运算.....	31
2.2.2 专门的关系运算.....	33
2.2.3 5 种基本操作的实现.....	35
2.3 关系演算.....	36
2.3.1 元组关系演算.....	36
2.3.2 域关系演算.....	38
2.4 关系运算的安全限制.....	38
2.5 关系代数表达式的优化.....	41
2.5.1 优化的一般策略.....	41
2.5.2 关系代数表达式的等价代换规则.....	42
2.5.3 关系代数表达式的优化算法.....	43
小结.....	46
习题.....	47
第 3 章 关系数据库语言 SQL.....	49
3.1 SQL 语言组成.....	50
3.1.1 关系与表.....	50
3.1.2 SQL 的数据类型.....	51
3.1.3 SQL 的语句类型.....	53
3.2 SQL 的数据定义.....	54
3.2.1 表的建立和删除.....	54
3.2.2 表的扩充与修改.....	56
3.2.3 视图 ( View ).....	58
3.2.4 索引.....	59
3.2.5 数据库模式的创建与删除.....	59
3.3 SQL 的数据查询.....	60
3.3.1 单表查询.....	61
3.3.2 多表查询.....	63
3.3.3 表达式与函数的使用.....	66

3.3.4	相关子查询.....	71
3.3.5	关系代数运算.....	73
3.4	SQL 的数据操纵.....	76
3.4.1	插入数据.....	76
3.4.2	修改数据.....	77
3.4.3	删除数据.....	78
3.5	SQL 的数据控制.....	78
3.6	嵌入式 SQL.....	79
3.6.1	概述.....	79
3.6.2	内嵌 SQL 语句的 C 程序组成.....	80
3.6.3	无游标的操作.....	82
3.6.4	带游标的查询操作.....	83
3.7	构造数据类型、域定义与动态 SQL.....	86
3.7.1	构造数据类型.....	86
3.7.2	域定义.....	87
3.7.3	动态 SQL.....	89
	小结.....	90
	习题.....	91
第 4 章	数据库管理系统.....	93
4.1	数据库管理系统和操作系统的相互关系.....	93
4.1.1	三种可能的配置方案.....	93
4.1.2	DBMS 对操作系统的要求.....	94
4.2	数据库管理系统的层次结构.....	95
4.3	数据的物理存储与管理.....	99
4.3.1	外存上的页面编址.....	99
4.3.2	系统缓冲区作为存储接口.....	101
4.3.3	页面替换策略.....	102
4.4	数据存取管理.....	102
4.4.1	记录在数据页面上的存储.....	102
4.4.2	记录编址.....	103
4.4.3	数据库文件结构.....	105
4.4.4	主关键字的存取路径结构.....	108
4.4.5	数据记录集合的存取路径结构.....	110
4.5	事务处理.....	114
4.5.1	事务.....	114
4.5.2	事务的状态.....	115
4.5.3	更新事务的执行与恢复.....	116
4.6	并发控制.....	117

4.6.1	并发控制的概念 .....	117
4.6.2	可串行化的判断 .....	120
4.6.3	封锁机制与两段式调度协议 .....	122
4.7	关系运算的实现算法 .....	127
4.8	关系数据库管理系统 SQL Server .....	130
	小结 .....	134
	习题 .....	135
第 5 章 关系数据库设计理论基础 .....		137
5.1	关系模式及其评价 .....	137
5.1.1	关系模式 .....	137
5.1.2	关系模式的评价 .....	138
5.2	函数依赖 .....	139
5.2.1	函数依赖的定义 .....	139
5.2.2	函数依赖的逻辑蕴涵 .....	141
5.2.3	关键字 .....	141
5.3	函数依赖公理体系 .....	142
5.3.1	Armstrong 公理 .....	142
5.3.2	Armstrong 公理的完备性 .....	143
5.3.3	闭包计算 .....	145
5.3.4	函数依赖集的等价和最小集 .....	146
5.4	关系模式的分解 .....	147
5.4.1	分解的无损连接性 .....	148
5.4.2	分解的函数依赖保持性 .....	152
5.5	关系模式的规范化 .....	153
5.5.1	第一范式 (1NF) .....	153
5.5.2	第二范式 (2NF) .....	153
5.5.3	第三范式 (3NF) .....	154
5.5.4	BCNF .....	155
5.5.5	分解算法 .....	157
5.6	多值依赖和第四范式 .....	159
5.6.1	多值依赖的定义 .....	160
5.6.2	多值依赖公理 .....	161
5.6.3	第四范式 (4NF) .....	162
5.7	连接依赖和第五范式 .....	164
5.7.1	连接依赖 .....	164
5.7.2	第五范式 (5NF) .....	165
	小结 .....	166
	习题 .....	166

第 6 章 非关系数据模型 .....	169
6.1 实体 - 联系模型 .....	170
6.1.1 基本概念 .....	170
6.1.2 ER 图 .....	171
6.1.3 实体间的依赖 .....	175
6.1.4 组合关键字 .....	177
6.1.5 扩充 ER 模型 .....	178
6.2 实体 - 联系模型转换为关系模式 .....	179
6.3 面向对象数据模型 .....	182
6.3.1 对象 .....	182
6.3.2 类 .....	183
6.3.3 继承 .....	184
6.3.4 对象标识 .....	185
6.4 基于逻辑的数据模型 .....	185
6.4.1 基于一阶谓词的数据模型 .....	185
6.4.2 逻辑数据库 .....	186
小结 .....	188
习题 .....	188
第 7 章 关系数据库设计 .....	190
7.1 信息系统及其与数据库的关系 .....	190
7.2 数据库设计与数据库生命周期 .....	191
7.2.1 数据库设计的含义 .....	191
7.2.2 数据库设计的目标 .....	192
7.2.3 数据库设计的一般过程 .....	193
7.3 数据库设计方法 .....	194
7.3.1 数据库设计方法概述 .....	194
7.3.2 基于 3NF 的数据库设计方法 .....	195
7.3.3 LRA 方法 .....	196
7.3.4 DBMS 的选择和全关系 DBMS 的评价准则 .....	199
7.4 关系数据库的设计方法 .....	202
7.4.1 数据需求分析 .....	202
7.4.2 概念模型设计 .....	205
7.4.3 逻辑设计 .....	212
7.4.4 物理设计 .....	215
7.5 数据库设计评价 .....	219
7.5.1 数据库设计评价准则 .....	219
7.5.2 一个通用的分析方法 .....	220

7.5.3 设计策略 .....	221
7.6 用 UML 设计关系数据库模式 .....	223
小结 .....	224
习题 .....	225
<b>第 8 章 数据库的完整性与安全性 .....</b>	<b>226</b>
8.1 概述 .....	226
8.2 数据的完整性约束 .....	227
8.2.1 完整性的语义约束和检查 .....	228
8.2.2 SQL 中的完整性约束 .....	229
8.3 数据库的安全性 .....	232
8.3.1 DBMS 安全模型 .....	232
8.3.2 视图与安全性 .....	233
8.3.3 访问控制 .....	234
8.3.4 数据密码 .....	236
8.3.5 跟踪审计 .....	237
8.3.6 统计数据库的安全性 .....	238
8.4 数据库的故障与恢复 .....	239
8.4.1 数据库的故障类型 .....	239
8.4.2 恢复机制 .....	240
8.4.3 事务级故障的恢复 .....	243
8.4.4 系统与介质故障的恢复 .....	244
8.4.5 SQL 对恢复操作的支持 .....	245
小结 .....	247
习题 .....	247
<b>第 9 章 网络数据库系统 .....</b>	<b>249</b>
9.1 基于客户机/服务器模式的数据库系统 .....	250
9.1.1 客户机/服务器系统的组成 .....	250
9.1.2 三层结构的客户机/服务器系统 .....	251
9.2 服务器数据库的建立 .....	253
9.2.1 数据库的建立 .....	253
9.2.2 用户注册与授权 .....	257
9.2.3 数据源的建立 .....	261
9.2.4 数据库设计 .....	261
9.3 服务器数据操纵 .....	262
9.3.1 流程控制语言 .....	262
9.3.2 触发器 .....	265
9.3.3 存储过程 .....	267

9.4	数据库访问接口.....	269
9.4.1	固有调用.....	269
9.4.2	ODBC.....	269
9.4.3	JDBC.....	271
9.5	SQL 会话期与调用级接口 CLI.....	273
9.5.1	SQL 的会话期.....	273
9.5.2	调用级接口 SQL/CLI.....	273
9.6	客户端应用系统.....	274
9.7	基于浏览器/服务器模式的数据库系统.....	275
9.7.1	浏览器/服务器模式.....	275
9.7.2	浏览器/服务器的工作原理和过程.....	276
9.7.3	浏览器/服务器系统的工作方式.....	277
9.7.4	浏览器/服务器系统的实施方案.....	278
9.8	ASP 的数据库访问技术.....	279
9.8.1	ASP 文件.....	280
9.8.2	服务器组件.....	281
9.8.3	访问服务器数据库.....	282
9.9	JSP 的数据库访问技术.....	284
9.10	Internet/Intranet 信息系统多层体系结构.....	286
9.10.1	多层应用软件的体系结构.....	286
9.10.2	多层应用软件体系的实现技术.....	287
	小结.....	289
	习题.....	289
第 10 章	分布式数据库系统.....	291
10.1	概述.....	291
10.1.1	什么是分布式数据库系统.....	291
10.1.2	分布式数据库系统的目标与优点.....	293
10.2	分布式数据库的模式结构.....	294
10.2.1	数据分布.....	294
10.2.2	分布式数据库的模式结构.....	295
10.3	分布式数据库管理系统.....	298
10.3.1	分布式数据库管理系统的功能与组成.....	298
10.3.2	分布式数据库管理系统的主要技术问题.....	299
10.4	分布式查询处理.....	300
10.4.1	分布式查询的处理过程与特点.....	300
10.4.2	基于半连接的分布式查询处理.....	301
10.5	分布式事务模型.....	302
10.6	并发控制.....	303

小结.....	304
习题.....	305
第 11 章 面向对象数据库系统 .....	306
11.1 概述 .....	306
11.1.1 面向对象数据库系统的功能要求 .....	307
11.1.2 面向对象数据库系统的实现途径 .....	307
11.2 面向对象数据模型.....	308
11.2.1 对象结构 .....	308
11.2.2 对象间的联系 .....	309
11.2.3 对象标识 .....	310
11.2.4 对象参照完整性约束 .....	310
11.3 面向对象数据库语言 .....	310
11.3.1 对象的持久性 .....	311
11.3.2 ODMG-93.....	312
11.4 面向对象数据库管理系统.....	315
11.4.1 面向对象数据库管理系统的基本构造.....	315
11.4.2 存储结构 .....	315
11.4.3 方法的实现 .....	316
11.4.4 长事务 .....	316
11.4.5 版本管理 .....	317
11.5 对象-关系数据库系统.....	318
11.5.1 对象-关系数据库系统的特征 .....	318
11.5.2 SQL 3.....	320
小结.....	324
习题.....	324
第 12 章 数据仓库 .....	325
12.1 概述 .....	325
12.1.1 数据仓库的产生 .....	325
12.1.2 数据仓库定义 .....	327
12.1.3 数据仓库系统的主要特征 .....	327
12.2 数据仓库构造.....	327
12.2.1 数据仓库系统结构 .....	327
12.2.2 数据仓库模型 .....	329
12.2.3 元数据 .....	331
12.3 联机分析处理 OLAP .....	332
12.3.1 OLAP 数据模型.....	332
12.3.2 OLAP 基本操作.....	334

# CONTENTS

12.3.3	OLAP 实现技术.....	334
12.4	数据仓库的开发.....	337
12.4.1	数据仓库开发的流程.....	337
12.4.2	构造数据仓库的过程.....	338
12.4.3	执行信息系统.....	339
12.5	数据挖掘.....	340
12.5.1	什么是数据挖掘.....	340
12.5.2	数据挖掘的主要目标与特点.....	340
12.5.3	数据挖掘模型.....	341
	小结.....	343

# 第 1 章 Database

## 基础知识

随着科学技术的发展，人们掌握和处理的信息越来越多，然而要想充分地开发与利用这些信息资源，就必须对大量的信息进行识别、存储、处理与传递。众所周知，人脑在信息的识别、信息的分析与综合、推理和联想等方面有很强的能力，但在记忆信息、快速处理信息方面的能力较弱，因此人脑要从大量而又经常变化的信息中提取一些关键信息并不是件很容易的事。以电子计算机为基础的数据库技术，由于具有信息存储量大、处理和传输速度快、逻辑推理严密、重复性高且不会疲劳、能够有效合理地存储各种信息、能为有关应用准确快速地提供有用信息等特点，正好弥补了人脑加工处理信息等方面的能力的不足，使其很快成为了信息处理的强有力工具。

20 世纪 60 年代末出现了数据库技术，当时主要用于商用数据处理。20 世纪 70 年代曾经被称为数据库的年代，数据库技术得到了蓬勃的发展和迅速地推广应用，渗透到了工农业生产、商业、行政管理、科学研究、教育、工程技术和国防军事等各个部门。数据库成了管理信息系统（MIS）、办公室自动化系统（OA）、决策支持系统（DSS）等各类应用系统的核心部分。可以说哪里有信息处理，哪里就有数据库技术的应用。

20 世纪 80 年代以来，计算机网络技术和面向对象的程序设计方法的迅速发展，新的计算机环境对数据库系统给出了新的支持，而新的应用对数据库系统提出了新的要求，因而在传统的层次数据库、网状数据库、关系数据库的基础上产生了多种新型数据库系统，如：按其数据对象的类型分有图形图像数据库系统、工程数据库系统、统计

数据库系统、多媒体数据库系统；按其体系结构和服务方式有分布式数据库系统、并行数据库系统、主动数据库系统；按其数据模型有面向对象数据库系统、对象-关系数据库系统、演绎数据库系统；按其智能化程度有智能数据库系统等。当代数据库技术的应用也已从常规的商用数据处理扩展到许多新的应用领域，例如地球观测系统、地理信息系统、协同设计、数字出版、电子商务、多媒体应用等。

人们把早期的层次数据库系统与网状数据库系统称为第一代数据库系统，把当今流行的关系数据库系统称为第二代数据库系统。当前正在发展的热点是新型的第三代乃至第四代数据库系统。数据库技术的发展方兴未艾，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的基础上。

本章是对数据库系统的一个概括叙述，主要介绍数据库、数据模型、数据库系统、数据库管理系统等有关的基本概念和基础知识。

## 1.1 数据库系统

### 1.1.1 信息世界

人类社会错综复杂，存在着各种各样的事物，各种事物有各自的属性，事物之间存在着各种联系，人们对各种事物有相应的管理活动，所有的这些都跟信息有关。农民从事农业生产需要土质、产量、肥料价格、气候情况等信息；工厂制定生产计划需要市场需求、材料来源、生产能力等有关信息；在日常生活中需要天气、物价、亲友的通信地址及电话号码、日程安排等信息。总之，人们的一切活动时刻都和大量的信息打交道，所有的这些信息就构成了信息世界。人类社会的一切活动都离不开对信息的处理，由于信息世界中的绝大部分信息都可以用数据来表示，因此，信息处理实际上也可以说是数据处理。从数据处理技术发展的历史来看，数据处理技术大致经历了手工处理、机械处理和电子处理 3 个阶段。

从原始社会到 19 世纪末，由于社会生产力和科学技术的限制，数据处理处于低级的手工处理阶段。这一阶段主要使用的计算工具有算盘、基于齿轮结构的六位加法器、计算尺和微分机等。该阶段的特点是计算工具低级、精确度差、处理能力低，并且离不开手工操作。

1890 年，美国中央统计局的 H.Hollerith 为了编制人口统计表的需要，发明了卡片制表机，使数据处理跨入了机械处理阶段。该机能以半自动方式进行卡片的穿孔、校验、分类、整理和制表等工作。与手工阶段相比，一部分手工操作由机械所代替，因此，无论在数据处理的能力和效率上都有很大的提高，但其效率受到机械设备性能的限制。

1946 年第一台电子计算机 ENIAC 的诞生，标志着一个崭新的数据处理阶段——电子处理阶段的开始。电子计算机以它自动、快速的处理，大容量的数据存储，灵活的输入/输出，彻底改变了以往数据处理效率低，手工操作多，可靠性差，与当时社会生产力的发展不相适应的落后状态。随着计算机硬件和软件的发展，特别是高速中央处理器和大容量磁盘存储器的应用，使计算机不仅能进行各类数值计算，同时还能对文字和图像等各种数据进行处理，使数据处理技术获得突飞猛进的发展。本书所讨论的数据处理均指电子数据处理。