



普通高等院校计算机专业（本科）实用教程系列

数据库实用教程

丁宝康 董健全 编著 施伯乐 主审



清华大学出版社

<http://www.tup.tsinghua.edu.cn>



普通高等院校计算机专业(本科)实用教程系列

数据库实用教程

丁宝康 董健全 编著

施伯乐 主审

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书是为全国高等院校计算机专业及相关专业开设数据库课程而精心组织和编著的一本实用教材。

本书详细介绍了数据库基本原理、方法和应用技术。内容包括:数据库系统结构,关系运算,SQL 语言,规范化设计,实体联系模型,数据库设计全过程,存储技术,数据库管理机制,客户机/服务器系统,分布式数据库,ODBC 技术,对象关系数据库,面向对象数据库,PowerBuilder 和 SQL Server 应用简介。

本书内容丰富,概念阐述细致清楚。书中配有适量的例题和习题,便于学生学习。本书既可作为高等院校有关专业的数据库课程教材,也可作为从事信息领域工作的科技人员的参考书。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: 数据库实用教程

作 者: 丁宝康 董健全

出 版 者: 清华大学出版社(北京清华大学学研大厦,邮编: 100084)

<http://www.tup.tsinghua.edu.cn>

责任编辑: 闫红梅

印 刷 者: 清华大学印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 787×1092 1/16 印张: 24.75 字数: 598 千字

版 次: 2001 年 9 月第 1 版 2001 年 9 月第 1 次印刷

书 号: ISBN 7-302-04797-9/TP·2847

印 数: 0001 ~ 5000

定 价: 35.00 元

序 言

时光更迭、历史嬗递。中国经济带着她足以令世人惊叹的持续高速发展驶入了一个新的世纪,一个新的千年。世纪之初,以微电子、计算机、软件、通信技术为主导的信息技术革命给我们生存的社会所带来的变化令人目不暇接。软件是优化我国产业结构、加速传统产业改造和用信息化带动工业化的基础产业,是体现国家竞争力的战略性产业,是从事知识的提炼、总结、深化和应用的高智型产业;软件关系到国家的安全,是保证我国政治独立、文化不受侵蚀的重要因素;软件也是促进其他学科发展和提升的基础学科;软件作为 20 世纪人类文明进步的最伟大成果之一,代表了先进文化的前进方向。美国政府早在 1992 年“国家关键技术”一文中提出“美国在软件开发和应用上所处的传统领先地位是信息技术及其他重要领域竞争能力的一个关键因素”,“一个成熟的软件制造业的发展是满足商业与国防对复杂程序日益增长的要求所必需的”,“在很多国家关键技术中,软件是关键的起推动作用(或阻碍作用)的因素”。在 1999 年 1 月美国总统信息技术顾问委员会的报告“21 世纪的信息技术”中指出“从台式计算机、电话系统到股市,我们的经济与社会越来越依赖于软件”,“软件研究为基础研究方面最优先发展的领域”。而软件人才的缺乏和激烈竞争是当前国际的共性问题。各国、各企业都对培养、引进软件人才采取了特殊政策与措施。

为了满足社会对软件人才的需要,为了让更多的人可以更快地学到实用的软件理论、技术与方法。我们编著了《普通高等院校计算机专业(本科)实用教程系列丛书》。本套丛书面向普通高等院校学生,以培养面向 21 世纪计算机专业应用人才(以软件工程师为主)为目标,以简明实用、便于学习、反映计算机技术最新发展和应用为特色,具体归纳为以下几点:

1. 讲透基本理论、基本原理、方法和技术,在写法上力求叙述详细,算法具体,通俗易懂,便于自学。
2. 理论结合实际。计算机是一门实践性很强的科学,丛书贯彻从实践中来到实践中去的原则,许多技术理论结合实例讲,以便于学习和理解。
3. 本丛书形成完整的体系,每本教材既有相对独立性,又有相互衔接和呼应,为总的培养目标服务。
4. 每本教材都配以习题和实验,在各教学阶段安排课程设计或大作业,培养学生的实战能力与创新精神。习题和实验可以制作成光盘。

新世纪曙光激人向上,催人奋进。江总书记在十五届五中全会上的讲话:“大力推进国民经济和社会信息化,是覆盖现代化建设全局的战略举措。以信息化带动工业化,发挥优势,实现社会生产力的跨越式发展。”指明了我国信息界前进的方向。21 世纪日趋开放的国策与更加迅速发展的科技会托起祖国更加辉煌灿烂的明天。

孙家广

2001 年 3 月

· I ·

普通高等院校计算机专业(本科)实用教程系列丛书

编 委 会

- 主 任 孙家广(清华大学教授,中国工程院院士)
成 员 (按姓氏笔划为序)
- 王玉龙(北方工业大学教授)
- 艾德才(天津大学教授)
- 刘 云(北方交通大学教授)
- 任爱华(北京航空航天大学副教授)
- 幸云辉(北京邮电大学教授)
- 张海藩(北京信息工程学院教授)
- 徐孝凯(中央广播电视大学副教授)
- 徐培忠(清华大学出版社编审)
- 樊孝忠(北京理工大学教授)
- 丛书策划 徐培忠 徐孝凯

2003/10

前 言

数据库是普通高校计算机专业和信息管理专业一门专业基础课。它的主要任务是研究如何存储、使用和管理数据。目的是使学生掌握数据库的基本原理、方法和应用技术,能有效使用现有的数据库管理系统和软件开发工具,掌握数据库结构的设计和数据库应用系统的开发方式。

数据库技术是计算机软件领域的一个重要分支,产生于20世纪60年代末。30多年来,数据库技术得到迅速发展,并形成较为完整的理论体系和一大批实用系统,同时造就了C.W. Bachman, E.F. Codd和J. Gray三位图灵奖得主。随着数据库系统的推广,计算机应用已深入到社会各个角落。当今的管理信息系统(MIS)、办公信息系统(OIS)、计算机辅助设计与制造(CAD/CAM)、计算机集成制造系统(CIMS)、地理信息系统(GIS)等都以数据库技术为基础。在20世纪90年代初,我国已在国民经济各领域装备了以数据库为基础的大型计算机系统。

在人类进入21世纪知识经济时代,信息已成为经济发展的战略资源,信息技术已成为社会生产力中重要的组成部分。人们充分认识到,数据库是信息化社会中信息资源管理与开发利用的基础。对于一个国家,数据库的建设规模、使用水平已成为衡量该国信息化程度的重要标志。因此,数据库是计算机专业的一门重要的课程。

本书详细介绍了数据库技术的基本原理、方法和应用技术。全书分7个部分:基础篇、运算篇、设计篇、管理篇、网络篇、发展篇和应用篇。具体内容如下:

- 第1章 介绍数据库技术的由来和发展过程。
- 第2章 介绍数据库系统的数据模型、体系结构和全局结构等内容。
- 第3章 介绍关系模型的运算理论:关系代数和关系演算。
- 第4章 介绍关系数据库标准语言SQL的全貌。
- 第5章 介绍关系数据库的模式设计理论,包括函数依赖、分解特性和范式等内容。
- 第6章 介绍实体联系模型的基本要素、设计过程,并给出许多实例。
- 第7章 介绍数据库应用系统设计的全过程,重点在概念设计和逻辑设计。
- 第8章 介绍数据库的存储结构,包括文件组织、索引和散列技术。
- 第9章 介绍数据库的管理机制,包括事务的概念及恢复、并发控制、完整性控制和安全性控制。
- 第10章 介绍当今热门的客户机/服务器系统的体系结构。
- 第11章 介绍分布式数据库系统的概念、数据存储、模式结构、查询处理。
- 第12章 介绍开放数据库互连(ODBC)技术的结构和接口。
- 第13章 介绍对象联系图、对象关系数据库的定义语言和查询语言。
- 第14章 介绍面向对象数据模型的基本概念和持久化C++系统。
- 第15章 介绍软件开发工具PowerBuilder 7.0的基本概念和数据库应用开发实例。

第 16 章 介绍 SQL Server 7.0 的基本组成。

全书每章后均配有适量的习题,以加强对数据库基本原理、方法的理解和掌握。

本书内容丰富,讲授时应根据需要对内容做适当取舍。根据我们多年授课的经验,可把全书分成若干板块,建议如下:

(1) 讲授。重点讲授第 1~7 章及第 9 章的内容。其中对于第 3 章中关系演算和第 5 章中理论性较强的内容,可根据情况适当压缩。

(2) 介绍。对于第 10~12 章网络篇的内容,教师可有针对性地选择某些内容,向学生讲授。

(3) 自学。对于第 8 章存储结构内容及第 13、14 章面向对象数据库内容,教师可作适当引导让学生自学,作为课外作业,以提高学生的工作能力和研究水平,拓宽知识面。

(4) 实习。第 15、16 章的内容可根据具体实习环境酌情采用。

本书第 1、2、5、10~14 章由丁宝康撰写,第 3、4、6~9、15、16 章由董健全撰写。复旦大学施伯乐教授审阅了全稿,并提出了许多宝贵的意见。

本书组稿时,我们曾和西安交通大学顾学春教授、中国人民大学信息学院何军教授就全书的结构、取材进行了三次探讨。本书的出版还得到了复旦大学、上海大学、上海(国际)数据库研究中心的支持。姜连生、夏根女、陈长洪、杨卫稼、徐庆莲、徐美娟、王晓雯和丁宁等老师为本书的出版做了大量工作。在此谨向他们表示衷心的感谢。

限于水平,书中欠妥之处,敬请广大读者和专家批评指正。

丁宝康 复旦大学信息与工程学院
董健全 上海大学计算机学院
jqdong@yc.shu.edu.cn

2001 年 7 月

目 录

第1部分 基础篇

第1章 数据库发展史	1
1.1 数据管理技术的发展	1
1.1.1 人工管理阶段	2
1.1.2 文件系统阶段	2
1.1.3 文件系统的缺陷	3
1.1.4 倒排文件系统阶段	3
1.2 数据库技术的产生和发展	4
1.2.1 数据库技术的产生	4
1.2.2 数据库阶段的数据管理特点	5
1.2.3 数据库技术的术语	7
1.2.4 数据库技术的发展	8
小结	9
习题1	9
第2章 数据库系统结构	11
2.1 数据描述	11
2.1.1 概念设计中的数据描述	11
2.1.2 逻辑设计中的数据描述	11
2.1.3 存储介质层次及数据描述	12
2.1.4 数据联系的描述	13
2.2 数据模型	15
2.2.1 数据模型的定义	15
2.2.2 实体联系模型	16
2.2.3 层次模型	17
2.2.4 网状模型	19
2.2.5 关系模型	20
2.2.6 面向对象模型	21
2.3 数据库的体系结构	22
2.3.1 三级模式结构	22
2.3.2 三级结构和两级映象	24
2.3.3 两级数据独立性	25

2.3.4 用户及用户界面	25
2.4 数据库管理系统	26
2.4.1 DBMS(数据库管理系统)的工作模式	26
2.4.2 DBMS的主要功能	27
2.4.3 DBMS的模块组成	28
2.5 数据库系统	28
2.5.1 DBS的组成	28
2.5.2 DBS的全局结构	30
2.5.3 DBS结构的分类	32
2.5.4 DBS的效益	34
小结	34
习题2	35

第2部分 运算篇

第3章 关系运算	36
3.1 关系数据模型	36
3.1.1 关系数据模型的定义	36
3.1.2 关键码和表之间的联系	38
3.1.3 关系模式、关系子模式和存储模式	38
3.1.4 关系模型的完整性规则	41
3.1.5 关系模型的形式定义	42
3.2 关系代数	43
3.2.1 关系查询语言和关系运算	43
3.2.2 关系代数的五个基本操作	43
3.2.3 关系代数的组合操作	46
3.2.4 关系代数表达式及其应用实例	48
3.2.5 扩充的关系代数操作	50
3.3 关系演算	51
3.3.1 元组关系演算	52
3.3.2 域关系演算	53
3.3.3 关系运算的安全性和等价性	54
3.4 查询优化	55
3.4.1 关系代数表达式的优化问题	55
3.4.2 关系代数表达式的等价变换规则	56
3.4.3 优化的一般策略	58
3.4.4 优化算法	59
小结	63
习题3	63

第 4 章 结构化查询语言 SQL	65
4.1 SQL 概述	65
4.1.1 SQL 的发展历程	65
4.1.2 SQL 数据库的体系结构	65
4.1.3 SQL 的组成	66
4.2 SQL 的数据定义	67
4.2.1 SQL 模式的创建和撤消	67
4.2.2 SQL 提供的基本数据类型	68
4.2.3 基本表的创建、修改和撤消	68
4.2.4 视图的创建和撤消	71
4.2.5 索引的创建和撤消	71
4.3 SQL 的数据查询	72
4.3.1 SELECT 语句格式	73
4.3.2 单表查询	73
4.3.3 多表查询	77
4.3.4 聚合函数	80
4.3.5 数据分组	81
4.3.6 集合操作	82
4.4 SQL 的数据更新	86
4.4.1 数据插入	86
4.4.2 数据删除	87
4.4.3 数据修改	87
4.4.4 对视图的更新操作	88
4.5 嵌入式 SQL	89
4.5.1 SQL 语言的运行环境	89
4.5.2 嵌入式 SQL 的使用规定	90
4.5.3 嵌入式 SQL 的使用技术	91
4.5.4 动态 SQL 语句	94
小结	95
习题 4	96

第 3 部分 设计篇

第 5 章 规范化设计	99
5.1 关系模式的设计问题	99
5.1.1 关系模型的外延和内涵	99
5.1.2 泛关系模式与数据库模式	99
5.1.3 关系模式的冗余和异常问题	100
5.1.4 本章的符号规定	101

5.2	函数依赖	101
5.2.1	函数依赖的定义	102
5.2.2	FD 的逻辑蕴涵	102
5.2.3	FD 的推理规则	103
5.2.4	FD 和关键码的联系	104
5.2.5	属性集的闭包	104
5.2.6	FD 集的最小依赖集	105
5.3	关系模式的分解特性	105
5.3.1	模式分解问题	105
5.3.2	无损分解	106
5.3.3	无损分解的测试方法	108
5.3.4	保持函数依赖的分解	109
5.3.5	模式分解与模式等价问题	110
5.4	关系模式的范式	111
5.4.1	第一范式(1NF)	111
5.4.2	第二范式(2NF)	111
5.4.3	第三范式(3NF)	112
5.4.4	BCNF(Boyce-Codd NF)	113
5.4.5	分解成 BCNF 模式集的方法	114
5.4.6	分解成 3NF 模式集的方法	114
5.4.7	模式设计方法的原则	114
5.5	模式的进一步规范化	115
5.5.1	多值依赖的定义	115
5.5.2	关于 FD 和 MVD 的推理规则集	116
5.5.3	第四范式(4NF)	117
5.5.4	联接依赖	117
5.5.5	第五范式(5NF)	118
	小结	119
	习题 5	119
第 6 章	实体联系模型	122
6.1	ER 模型的基本元素	122
6.2	属性的分类	123
6.2.1	基本属性和复合属性	123
6.2.2	单值属性和多值属性	124
6.2.3	导出属性	125
6.2.4	空值属性	126
6.3	联系的设计	126

6.3.1	联系的元数	126
6.3.2	联系的连通词	126
6.3.3	联系的基数	130
6.4	ER 模型的扩充	130
6.4.1	依赖联系与弱实体	130
6.4.2	超类和子类	131
6.4.3	演绎和归纳	132
6.5	ER 模型实例分析	133
	小结	139
	习题 6	139
第 7 章	数据库设计	141
7.1	数据库设计概述	141
7.1.1	软件生存期	141
7.1.2	数据库系统生存期	142
7.1.3	数据库设计的具体步骤	143
7.2	规划	144
7.3	需求分析	145
7.3.1	需求描述与分析	145
7.3.2	需求分析阶段的输入和输出	146
7.3.3	需求分析的步骤	146
7.4	概念设计	150
7.4.1	概念设计的必要性	150
7.4.2	概念模型	151
7.4.3	概念设计的主要步骤	152
7.4.4	数据抽象	152
7.4.5	ER 模型的操作	154
7.4.6	采用 ER 方法的数据库概念设计	156
7.5	逻辑设计	163
7.5.1	逻辑设计环境	163
7.5.2	逻辑设计的步骤	164
7.5.3	ER 模型向关系模型的转换	165
7.5.4	关系数据库的逻辑设计	168
7.6	物理设计	170
7.7	数据库的实现	171
7.8	数据库的运行与维护	171
	小结	171
	习题 7	172

第4部分 管理篇

第8章 存储技术	173
8.1 文件组织	173
8.1.1 定长记录	173
8.1.2 变长记录	175
8.2 文件结构	178
8.2.1 四种文件结构	178
8.2.2 顺序文件	178
8.2.3 聚集文件	180
8.3 索引技术	181
8.3.1 索引机制	181
8.3.2 有序索引的分类	181
8.3.3 主索引	181
8.3.4 辅助索引	184
8.3.5 B ⁺ 树索引文件	185
8.3.6 B树索引文件	191
8.4 散列技术	192
8.4.1 散列机制	192
8.4.2 散列索引	195
8.4.3 静态散列中的问题	196
8.4.4 可扩充散列结构	196
8.5 多键访问	200
8.5.1 单键查询的问题	200
8.5.2 网格文件	201
8.5.3 分区散列技术	202
小结	203
习题8	203
第9章 数据库的管理	206
9.1 事务的概念	206
9.1.1 事务的定义	206
9.1.2 事务的ACID性质	206
9.2 数据库的恢复	208
9.2.1 恢复的定义、原则和方法	208
9.2.2 故障类型和恢复方法	209
9.2.3 检查点机制	209
9.2.4 运行记录优先原则	211

9.2.5	SQL对事务的支持	211
9.3	数据库的并发控制	211
9.3.1	并发操作带来的三个问题	211
9.3.2	封锁机制	214
9.3.3	活锁和死锁	216
9.3.4	并发调度的可串行化	217
9.3.5	SQL中事务的存取模式和隔离级别	218
9.4	数据库的完整性	219
9.4.1	完整性子系统和完整性规则	219
9.4.2	SQL中的完整性约束	219
9.4.3	SQL3中的触发器	223
9.5	数据库的安全性	225
9.5.1	安全性级别	225
9.5.2	权限	225
9.5.3	SQL中的安全性机制	226
9.5.4	数据加密	227
9.5.5	自然环境的安全性	228
小结	228
习题9	229

第5部分 网络篇

第10章	客户机/服务器系统	231
10.1	“分布计算”概念	231
10.2	客户机/服务器计算模式	233
10.2.1	C/S结构的工作模式	233
10.2.2	C/S环境下应用成分的分布	233
10.2.3	C/S主要技术特征	234
10.3	C/S系统的体系结构	235
10.3.1	C/S模式的体系结构	235
10.3.2	C/S结构的演变	236
10.3.3	网络服务器的类型	238
10.4	C/S系统的管理问题	239
10.4.1	C/S系统的并行处理与并发控制	239
10.4.2	C/S系统的完整性	240
10.4.3	两类DBMS的出现	240
小结	241
习题10	241

第 11 章 分布式数据库系统	242
11.1 DDBS 概述	242
11.1.1 集中式系统与分布式系统	242
11.1.2 DDBS 的定义	243
11.1.3 DDBS 的特点	244
11.1.4 DDBS 的优缺点	245
11.1.5 DDBS 的分类	246
11.2 分布式数据存储	247
11.2.1 数据分片	247
11.2.2 数据分配	248
11.3 DDB 的模式结构	249
11.3.1 六层模式结构	250
11.3.2 六层结构的特征	251
11.3.3 分布透明性	251
11.4 DDBMS 的功能及组成	253
11.4.1 DDBS 的组成	253
11.4.2 DDBMS 的功能	254
11.4.3 DDBMS 的组成	254
11.5 分布式查询处理	255
11.5.1 查询代价的估算方法	255
11.5.2 基于半联接的优化策略	256
11.5.3 基于联接的优化策略	257
小结	258
习题 11	259
第 12 章 ODBC 技术	260
12.1 中间件	260
12.1.1 中间件的定义	260
12.1.2 中间件的作用	260
12.2 ODBC 概念和作用	261
12.3 ODBC 的结构	262
12.3.1 ODBC 的体系结构	262
12.3.2 应用程序	263
12.3.3 驱动程序管理器	263
12.3.4 DBMS 驱动程序	263
12.3.5 ODBC 的数据源管理	266
12.4 ODBC 接口	266
12.4.1 ODBC 应用程序的基本流程	266
12.4.2 ODBC 句柄	267

12.4.3	ODBC 连接	269
12.4.4	SQL 语句的执行	269
12.4.5	光标函数的使用	271
12.5	ODBC 的符合性级别	272
12.5.1	API 符合性的三个级别	272
12.5.2	SQL 符合性的三个级别	273
12.6	ODBC API 与 SQL CLI	274
12.7	典型的数据库应用系统开发工具	274
12.7.1	PowerBuilder	275
12.7.2	Delphi	276
12.7.3	Visual Basic	276
12.7.4	Developer/2000	277
小结	278
习题 12	278

第 6 部分 发展篇

第 13 章	对象关系数据库	280
13.1	对象联系图	280
13.1.1	从关系到嵌套关系、复合关系	280
13.1.2	引用类型	282
13.1.3	对象联系图的成分	282
13.1.4	数据的泛化/细化	284
13.2	ORDB 的定义语言	284
13.2.1	ORDB 的定义	285
13.2.2	数据类型的定义	285
13.2.3	继承性的定义	286
13.2.4	引用类型的定义	288
13.3	ORDB 的查询语言	289
13.3.1	以关系为值的属性	289
13.3.2	路径表达式	289
13.3.3	嵌套与解除嵌套	290
小结	291
习题 13	292
第 14 章	面向对象数据库	294
14.1	面向对象数据库系统概述	294
14.1.1	ODMG 组织和标准	294
14.1.2	OODBS 的概念	295

14.1.3	OODBS 存在的问题和局限性	295
14.2	面向对象数据模型的基本概念	296
14.2.1	对象	296
14.2.2	类	296
14.2.3	继承性	297
14.2.4	对象标识	298
14.2.5	对象包含	299
14.3	持久化程序设计语言	299
14.3.1	持久化语言与嵌入式语言的区别	299
14.3.2	持久化语言的基本概念	300
14.4	持久化C++ 系统	301
14.4.1	ODMG C++ 对象定义语言	301
14.4.2	ODMG C++ 对象操纵语言	303
14.5	OODB 与 ORDB 的比较	303
	小结	304
	习题 14	305

第 7 部分 应用篇

第 15 章	PowerBuilder 7.0 数据库应用开发简介	306
15.1	PowerBuilder 7.0 集成开发环境	306
15.1.1	开发工具和集成开发环境	306
15.1.2	使用画板	309
15.2	一个完整的实例	309
15.2.1	规划	310
15.2.2	创建数据库	312
15.2.3	连接数据库	316
15.2.4	创建表和数据操作	319
15.2.5	建立应用	323
15.2.6	建立数据窗口	324
15.2.7	建立窗口	331
15.2.8	运行应用程序	338
15.2.9	创建另外两个窗口	339
15.2.10	添加菜单	341
15.3	PowerScript 编程语言	345
15.3.1	基本语法规则	345
15.3.2	运算符和数据类型	346
15.3.3	变量及其作用域	347
15.3.4	实例变量(Instance)的访问权限	347