



面向 21 世 纪 课 程 教 材
Textbook Series for 21st Century

数据 库 系 统 教 程



清华大学出版社
TERTIADIAO HUAPU

面向 21 世 纪 课 程 教 材
Textbook Series for 21st Century

数据库系统教程

施伯乐 丁宝康 周傲英 田增平 编著



高 等 教 育 出 版 社
HIGHER EDUCATION PRESS

(京)112号

内容提要

本书是教育部“高等教育面向 21 世纪教学内容和课程体系改革计划”的研究成果，是面向 21 世纪课程教材和教育部高等学校计算机科学与技术学科“九五”规划教材。本书详细介绍了数据库系统的原理、方法和发展趋势。内容包括：数据库系统的基本概念、关系模型的运算及理论、关系数据库语言 SQL 和 QBE、关系数据库的设计、数据库的存储结构、系统实现技术、分布式数据库、面向对象数据库、对象关系数据库及数据库技术的回顾与展望。

本书可作为高等院校计算机有关专业的数据库课程教材，也可供从事计算机软件工作的科技人员和工程技术人员以及其他有关人员参阅。

图书在版编目(CIP)数据

数据库系统教程 / 施伯乐等编著 . —北京 : 高等教育出版社 , 1999

面向 21 世纪课程教材

ISBN 7-04-007496-6

I . 数 … II . 施 … 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 (1999) 第 64438 号

数据库系统教程

施伯乐 丁宝康 周傲英 田增平 编著

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

邮政编码 100009

电 话 010-64054588

传 真 010-64014048

网 址 <http://www.hep.edu.cn>

经 销 新华书店北京发行所

印 刷 中国科学院印刷厂

纸张供应 山东高唐纸业集团总公司

开 本 787×960 1/16

版 次 1999 年 12 月第 1 版

印 张 23.75

印 次 1999 年 12 月第 1 次印刷

字 数 440 000

定 价 24.90 元

凡购买高等教育出版社图书，如有缺页、倒页、脱页等

质量问题，请在所购图书销售部门联系调换。

版权所有 侵权必究

前　　言

数据库系统产生于 20 世纪 60 年代末。30 年来, 数据库技术得到迅速发展, 已形成较为完整的理论体系和一大批实用系统, 现已成为计算机软件领域的一个重要分支。

在 20 世纪 50 年代, 数据管理还处于人工管理阶段, 而计算机应用主要是用于科学计算。60 年代, 出现了操作系统, 从而使数据管理进入文件系统阶段。随着计算机应用的普及, 计算机逐步从科学计算转向企业管理。为了克服文件系统的数据冗余弱点, 方便用户操作和提高程序开发的生产率, 在 60 年代末产生了数据库系统。数据库系统产生之后, 显示了其强大的生命力。70 年代, 层次、网状数据库系统迅速研制成功, 并在商业上得到广泛应用。当时, 关系数据库的研究还集中在理论和实验系统的开发上, 直至 80 年代初才形成产品。由于关系数据库有较好的理论基础, 并具有操作方便等优点, 因此关系数据库的商用系统迅速占领市场, 并逐步取代了层次和网状数据库系统。1987 年 ISO 组织研究并颁布关系数据库语言 SQL 标准。90 年代数据库技术进一步发展, 推出许多新型数据库系统, 以适应用户提出的新需求, 并进而渗透到多媒体、人工智能、网络等领域。

随着数据库系统的推广使用, 计算机应用已深入到工农业生产、商业、金融、行政管理、科学的研究和工程技术的各个领域; 当今的管理信息系统(MIS)、办公自动化(OA)、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)和知识库系统等, 也都以数据库技术为基础。90 年代初, 我国已在邮电、银行、电力、铁路、气象、民航、情报、公安、军事、航天、财税等行业装备了以数据库为基础的大型计算机系统。

在人类迈向 21 世纪知识经济时代的时候, 信息变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。对于一个国家来说, 数据库的建设规模和使用水平已成为衡量该国信息化程度的重要标志。因此, 数据库课程是计算机领域中一门重要课程。

本书较为详细地介绍了数据库系统的基本概念、原理、方法及应用技术。全

书共分 10 章。

第一章介绍数据库系统的基本概念,包括数据模型、数据库的体系结构、数据库系统的全局结构。

第二章介绍关系模型的运算理论:关系代数和关系演算。

第三章介绍关系数据库标准语言 SQL 和另一个对数据库发展有重大作用的 QBE 语言。

第四章介绍关系数据库的规范化设计理论以及数据库设计的全过程。

第五章介绍数据库的存储结构,包括存储介质、文件组织、索引和散列技术。

第六章介绍数据库系统的实现技术,包括事务的概念及并发控制、恢复、完整性和安全性实现机制。

第七章介绍分布式数据库的概念、结构和方法。

第八章介绍基于面向对象程序设计思想的面向对象数据模型的概念以及持久化 C++ 系统。

第九章介绍基于关系模型的对象关系模型的概念和语言。

第十章回顾数据库技术的发展历史、综述现状、介绍新的应用、分析面临的挑战并提出新的研究方向。

全书每章后面均配有适量的习题,以加强对数据库系统原理、方法的理解和掌握。

本书是计算机有关专业的教材。讲授时应根据需要对内容作适当取舍。建议如下:

① 重点讲授第一章至第四章及第六章内容。其中对于第二章中关系演算和第四章中理论性较强的内容可根据情况适当压缩。

② 对于第五章的存储结构内容,教师可作适当引导并让学生自学,以提高学生的工作能力和研究水平,拓宽知识面。

③ 对于第七章以及后面涉及到数据库系统的进一步发展的内容,教师可有针对性地选择某些内容,向学生传授。

本书由施伯乐负责内容的取材、组织和审定。第一章至第六章由丁宝康执笔,第七章由田增平执笔,第八章至第十章由周傲英执笔。

复旦大学楼荣生教授仔细校阅了全稿,在此向他表示感谢。

上海交通大学尤晋元教授审阅了全稿,并提出许多宝贵的意见,在此向他致以衷心的感谢。

限于水平,书中难免有欠妥之处,欢迎广大读者和专家批评指正。

作 者

1999 年 4 月于复旦大学

责任编辑 鲍 涌
封面设计 张 楠
责任绘图 李维平
版式设计 史新薇
责任校对 王效珍
责任印制 宋克学

目 录

第一章 数据库概论	(1)
1.1 引言	(1)
1.1.1 数据与信息	(2)
1.1.2 数据处理与数据管理	(3)
1.1.3 数据库技术的基本概念	(3)
1.2 数据库的由来和发展	(4)
1.2.1 人工管理阶段	(4)
1.2.2 文件系统阶段	(5)
1.2.3 数据库阶段	(6)
1.2.4 高级数据库技术阶段	(8)
1.3 数据描述	(10)
1.3.1 现实世界	(10)
1.3.2 信息世界	(10)
1.3.3 机器世界	(10)
1.3.4 类型和值	(11)
1.3.5 物理存储中的数据描述	(11)
1.3.6 逻辑数据和物理数据	(12)
1.3.7 数据之间的联系	(13)
1.3.8 实体间的联系	(13)
1.4 数据模型	(14)
1.4.1 数据模型的定义	(14)
1.4.2 实体联系模型	(15)
1.4.3 层次模型	(17)
1.4.4 网状模型	(18)
1.4.5 关系模型	(19)
1.4.6 面向对象模型	(21)

1.5 数据库的体系结构	(22)
1.5.1 数据库的三级体系结构	(22)
1.5.2 概念模式	(24)
1.5.3 外模式	(24)
1.5.4 内模式	(24)
1.5.5 模式/内模式映象	(25)
1.5.6 外模式/模式映象	(25)
1.5.7 用户	(25)
1.5.8 用户界面	(25)
1.5.9 本节小结	(26)
1.6 数据库管理系统	(26)
1.6.1 DBMS 的主要功能	(27)
1.6.2 DBMS 的组成	(28)
1.6.3 用户访问数据的过程	(28)
1.7 数据库系统	(29)
1.7.1 DBS 的组成	(29)
1.7.2 DBS 的全局结构	(31)
1.7.3 DBS 结构的分类	(33)
1.7.4 DBS 的效益	(36)
习题	(37)
第二章 关系模型	(38)
2.1 关系模型的基本概念	(38)
2.1.1 基本术语	(38)
2.1.2 关系的定义	(39)
2.1.3 关系模式	(40)
2.1.4 子模式	(41)
2.1.5 存储技术	(42)
2.1.6 关系模型的三类完整性规则	(43)
2.1.7 关系模型的形式定义	(44)
2.1.8 关系模型的优点	(44)
2.1.9 ER 模型向关系模型的转换	(44)
2.2 关系代数	(45)
2.2.1 关系查询语言和关系运算	(45)
2.2.2 关系代数的五种基本操作	(46)
2.2.3 关系代数的其它操作	(48)
2.2.4 关系代数运算的应用实例	(50)
2.2.5 扩充的关系代数操作	(51)
2.3 关系演算	(53)

2.3.1 元组关系演算	(54)
2.3.2 域关系演算	(57)
2.3.3 关系运算的安全约束	(58)
2.3.4 三种关系运算形式的等价性	(59)
2.4 查询优化	(60)
2.4.1 关系代数表达式的优化问题	(60)
2.4.2 关系代数表达式的等价变换规则	(61)
2.4.3 优化的一般策略	(63)
2.4.4 关系代数表达式的优化算法	(64)
习题	(67)
第三章 关系数据库语言 SQL 和 QBE	(71)
3.1 SQL 的体系结构	(71)
3.1.1 SQL 的产生和发展	(71)
3.1.2 SQL 数据库的体系结构	(72)
3.1.3 SQL 的组成	(73)
3.2 SQL 的数据查询	(73)
3.2.1 SELECT 查询语句	(73)
3.2.2 查询例子	(74)
3.2.3 聚合函数	(78)
3.2.4 SELECT 语句完整的句法	(78)
3.2.5 SELECT 子句中的规定	(79)
3.2.6 条件表达式中的算术比较操作	(80)
3.2.7 属性和关系的别名操作	(81)
3.2.8 字符串的匹配操作	(81)
3.2.9 集合的并、交、差操作	(81)
3.2.10 空值的处理	(82)
3.2.11 集合的比较操作	(83)
3.2.12 导出关系的使用	(85)
3.2.13 关系的联接	(86)
3.3 SQL 的数据更新	(88)
3.3.1 插入操作	(88)
3.3.2 删除操作	(89)
3.3.3 修改操作	(90)
3.4 SQL 的数据定义	(91)
3.4.1 模式的定义和撤消	(91)
3.4.2 SQL2 提供的基本数据类型	(92)
3.4.3 关系(基本表)的定义、修改和撤消	(92)
3.4.4 索引的定义和撤消	(95)

3.5 视图的定义和对视图的操作	(96)
3.5.1 视图的定义	(96)
3.5.2 视图的撤消	(97)
3.5.3 对视图的更新操作	(97)
3.5.4 视图的优点	(98)
3.6 嵌入式 SQL	(99)
3.6.1 SQL 语言的运行环境	(99)
3.6.2 SQL 和宿主语言的接口	(99)
3.6.3 嵌入式 SQL 的使用规定	(100)
3.6.4 嵌入式 SQL 的使用技术	(101)
3.6.5 动态 SQL 语句	(104)
3.7 QBE 的数据查询	(105)
3.7.1 QBE 概述	(105)
3.7.2 QBE 语言的命令结构	(106)
3.7.3 在单个关系上的数据查询	(107)
3.7.4 在多个关系上的数据查询	(109)
3.7.5 查询结果关系的形成	(110)
3.7.6 查询结果的排序	(110)
3.7.7 聚合操作	(111)
3.8 QBE 的数据更新	(112)
3.8.1 删除操作	(112)
3.8.2 插入操作	(113)
3.8.3 修改操作	(114)
3.8.4 QBE 的完备性	(114)
习题	(115)
第四章 关系数据库的规范化设计	(120)
4.1 关系模式的设计问题	(120)
4.1.1 关系模型的外延和内涵	(120)
4.1.2 关系数据库设计理论的主要内容	(120)
4.1.3 泛关系模式与数据库模式	(121)
4.1.4 关系模式的存储异常问题	(121)
4.1.5 冗余和数据依赖	(123)
4.1.6 本章的符号规定	(123)
4.2 函数依赖	(123)
4.2.1 函数依赖的定义	(123)
4.2.2 FD 的逻辑蕴涵	(124)
4.2.3 FD 的推理规则	(125)
4.2.4 FD 和键的联系	(126)

4.2.5 属性集的闭包	(127)
4.2.6 FD 推理规则的完备性	(127)
4.2.7 函数依赖集的规范覆盖	(128)
4.3 关系模式的分解特性	(130)
4.3.1 模式分解问题	(130)
4.3.2 无损联接	(130)
4.3.3 无损联接的测试算法	(133)
4.3.4 保持函数依赖的分解	(134)
4.3.5 本节小结	(135)
4.4 关系模式的范式	(135)
4.4.1 第一范式(1NF)	(136)
4.4.2 第二范式(2NF)	(136)
4.4.3 第三范式(3NF)	(136)
4.4.4 BCNF(Boyce-Codd NF)	(137)
4.4.5 分解成 BCNF 模式集的算法	(138)
4.4.6 分解成 3NF 模式集的算法	(140)
4.4.7 模式设计方法的原则	(141)
4.5 数据依赖和模式的规范化处理	(142)
4.5.1 多值依赖的定义	(142)
4.5.2 关于 FD 和 MVD 的推理规则集	(143)
4.5.3 属性集的依赖基	(144)
4.5.4 MVD 与无损联接的联系	(144)
4.5.5 第四范式(4NF)	(145)
4.5.6 分解成 4NF 模式集的算法	(146)
4.5.7 MVD 与 FD 的联系与区别	(146)
4.5.8 嵌入多值依赖(EMVD)	(147)
4.5.9 联接依赖(JD)	(147)
4.5.10 投影联接范式(PJNF)	(148)
4.6 数据库设计的全过程	(148)
4.6.1 数据库设计方法	(149)
4.6.2 规划阶段	(151)
4.6.3 需求分析阶段	(152)
4.6.4 概念设计阶段	(154)
4.6.5 逻辑设计阶段	(159)
4.6.6 物理设计阶段	(161)
4.6.7 数据库的实现	(162)
4.6.8 数据库的运行和维护	(162)
习题	(163)

第五章 数据库的存储结构	(166)
5.1 物理存储介质	(166)
5.1.1 存储介质层次	(166)
5.1.2 磁盘	(168)
5.1.3 第三级存储器	(171)
5.2 文件组织	(171)
5.2.1 定长记录	(172)
5.2.2 变长记录	(174)
5.3 文件中记录的组织	(177)
5.3.1 文件中记录的组织方式	(177)
5.3.2 顺序文件组织	(177)
5.3.3 聚类文件组织	(179)
5.4 索引技术	(180)
5.4.1 索引机制的分类	(180)
5.4.2 有序索引的分类	(180)
5.4.3 主索引	(181)
5.4.4 辅助索引	(184)
5.4.5 B ⁺ 树索引文件	(185)
5.4.6 B 树索引文件	(192)
5.5 散列技术	(193)
5.5.1 散列文件组织	(193)
5.5.2 散列索引	(196)
5.5.3 静态散列中的问题	(197)
5.5.4 可扩充散列结构	(197)
5.6 多键访问	(202)
5.6.1 单键查询的问题	(202)
5.6.2 网格文件	(203)
5.6.3 分区散列技术	(205)
习题	(205)
第六章 系统实现技术	(208)
6.1 系统目录	(208)
6.1.1 数据字典与系统目录	(208)
6.1.2 关系 DBMS 的目录	(209)
6.1.3 系统目录的扩充 ER 图	(211)
6.1.4 系统目录和 DBMS 各子系统的联系	(211)
6.2 事务	(212)
6.2.1 事务的基本概念	(212)

6.2.2 事务的 ACID 性质	(213)
6.2.3 事务的状态变迁	(215)
6.2.4 事务的并发执行	(216)
6.2.5 并发事务的可串行化	(218)
6.2.6 并发事务的可恢复性	(222)
6.2.7 SQL 中的事务定义	(223)
6.3 数据库的并发控制	(224)
6.3.1 数据库的并发操作带来的问题	(224)
6.3.2 排它型封锁(X 封锁)和 PX 协议	(226)
6.3.3 活锁和死锁	(227)
6.3.4 由事务的 ROLLBACK 引起的丢失更新问题	(229)
6.3.5 共享型封锁(S 封锁)	(230)
6.3.6 两段封锁法	(232)
6.3.7 基于时标的并发控制	(235)
6.4 数据库的恢复	(238)
6.4.1 数据库系统遭遇的故障	(238)
6.4.2 存储器结构	(238)
6.4.3 恢复和原子性的联系	(240)
6.4.4 基于日志的恢复技术	(241)
6.4.5 阴影页技术	(246)
6.4.6 并发事务的恢复	(247)
6.5 数据库的完整性	(249)
6.5.1 完整性子系统	(249)
6.5.2 完整性规则	(250)
6.5.3 SQL2 中的完整性约束	(250)
6.5.4 SQL3 中的触发器	(256)
6.6 数据库的安全性	(258)
6.6.1 安全性	(258)
6.6.2 权限	(259)
6.6.3 权限的转授与回收	(260)
6.6.4 SQL 中的安全性机制	(261)
6.6.5 数据加密法	(263)
6.6.6 统计数据库的安全性	(264)
6.6.7 自然环境的安全性	(265)
习题	(266)
第七章 分布式数据库	(270)
7.1 分布式数据库系统	(270)
7.1.1 从集中式数据库系统到分布式数据库系统	(270)

7.1.2 分布式数据库系统的定义	(271)
7.1.3 分布式数据库系统的透明性	(272)
7.1.4 分布式数据库系统的优点	(273)
7.1.5 分布式数据库系统的弱点	(274)
7.1.6 分布式数据库系统的分类	(274)
7.1.7 分布式数据库系统与客户/服务器体系结构	(275)
7.1.8 分布式数据库系统中存在的问题	(275)
7.2 分布式数据库设计及系统的体系结构	(276)
7.2.1 分布式数据库设计的任务	(276)
7.2.2 数据分片(data fragmentation)	(277)
7.2.3 数据分配(data distribution)	(278)
7.2.4 分布式数据库系统的体系结构	(280)
7.2.5 分布透明性(distribution transparency)	(282)
7.2.6 分布式数据库管理系统(DDBMS)的功能及组成	(284)
7.3 查询处理及优化	(285)
7.3.1 分布式查询处理的任务	(285)
7.3.2 一个分布式查询实例	(285)
7.3.3 全局查询到片段查询的变换	(287)
7.3.4 限定关系的代数	(289)
7.3.5 片段查询的简化	(289)
7.3.6 查询优化的目标	(292)
7.3.7 基于半联接的优化策略	(293)
7.3.8 基于联接的优化策略	(294)
7.4 分布式事务管理及恢复	(295)
7.4.1 分布式事务管理及相关问题	(295)
7.4.2 事务管理的目标	(296)
7.4.3 分布式事务的结构	(296)
7.4.4 分布式事务的计算结构	(298)
7.4.5 分布式事务恢复	(298)
7.4.6 两段提交协议(two-phase commitment protocol)	(300)
7.5 分布式并发控制	(301)
7.5.1 封锁	(302)
7.5.2 基于封锁的并发控制方法	(302)
7.5.3 全局死锁	(304)
7.5.4 基于时标的并发控制方法	(305)
7.5.5 分布式并发控制的乐观方法	(309)
7.6 事务处理管理器	(310)
7.6.1 集中式数据库、分布式数据库及事务处理(TP)管理器	(310)

7.6.2 TP管理器的概念	(311)
7.6.3 TP管理器的体系结构	(312)
习题	(313)
第八章 面向对象数据库	(315)
8.1 新型数据库应用	(315)
8.2 面向对象数据模型	(316)
8.2.1 对象结构	(316)
8.2.2 对象类	(317)
8.2.3 继承	(318)
8.2.4 多重继承	(321)
8.2.5 对象标识	(323)
8.2.6 对象包含	(324)
8.3 面向对象语言	(325)
8.4 持久化程序设计语言	(325)
8.4.1 对象的持久性	(326)
8.4.2 对象标识和指针	(327)
8.4.3 持久对象的存储和访问	(328)
8.5 持久化 C++ 系统	(329)
8.5.1 ODMG C++ 对象定义语言	(329)
8.5.2 ODMG C++ 对象操纵语言	(331)
8.6 小结	(333)
习题	(333)
第九章 对象关系数据库	(335)
9.1 嵌套关系	(335)
9.2 复杂类型和面向对象	(338)
9.2.1 结构类型和聚集类型	(338)
9.2.2 继承	(339)
9.2.3 引用类型	(342)
9.3 复杂类型查询	(343)
9.3.1 关系值属性	(343)
9.3.2 路径表达式	(344)
9.3.3 嵌套化和非嵌套化	(345)
9.3.4 函数	(346)
9.3.5 复杂值和对象的生成	(347)
9.4 面向对象和对象关系数据库的比较	(348)
9.5 小结	(349)
习题	(349)

第十章 数据库技术的回顾与展望	(351)
10.1 数据库系统的发展回顾	(351)
10.2 数据库技术的现状	(353)
10.2.1 数据建模	(353)
10.2.2 查询语言	(353)
10.2.3 查询优化和计值算法	(353)
10.2.4 数据管理	(354)
10.2.5 事务管理和并发处理	(354)
10.2.6 分布式系统	(354)
10.3 数据库的新应用	(354)
10.3.1 多媒体数据库(multimedia database)	(355)
10.3.2 主动数据库(active database)	(355)
10.3.3 并行数据库(parallel database)	(356)
10.3.4 工作流数据库(workflow database)	(357)
10.4 数据库技术展望	(359)
10.4.1 数据库技术面临的挑战	(359)
10.4.2 新的研究方向	(359)
参考文献	(362)

第一章 数据库概论

从 20 世纪 50 年代中期开始,计算机的应用由科学研究部门扩展到企业、行政部门,数据处理迅速上升为计算机应用的主要方面。在计算机的三大应用(科学计算、数据处理与过程控制)中,数据处理所占比重约为 70% 左右。在 60 年代末,数据库技术就是作为数据处理中的一门技术发展起来的。数据库技术是计算机软件领域的一个重要分支,已形成相当规模的理论体系和实用技术。本章先回顾数据管理技术的发展过程,然后介绍数据库中主要的概念,以使读者对数据库的概貌有所了解。

1.1 引言

当前人类社会正处于信息社会,人类知识以惊人的速度增长。如何组织和利用这样庞大的知识成为急需解决的问题。在 20 世纪 60 年代的“软件危机”中,数据库技术作为一门软件学科应运而生了。

据查,数据库(DataBase,记作 DB)这个名词起源于 20 世纪 50 年代,当时美国为了战争的需要,把各种情报集中在一起,存储在计算机里,称为 Information Base 或 DataBase。

1963 年,美国 Honeywell 公司的 IDS(Integrated Data Store)系统投入运行,揭开了数据库技术的序幕。1965 年,美国一家火箭公司利用数据存储系统帮助设计了阿波罗登月火箭,推动了数据库技术的产生。当时美国社会上产生了许多形形色色的 DataBase 或 Databank,但是,它们基本上都是文件系统的扩充或是倒排文件系统。1968 年,美国 IBM 公司推出层次模型的 IMS 数据库系统(1969 年形成产品);1969 年,美国 CODASYL(Conference On Data System Language,数据系统语言协会)组织的数据库任务组(DBTG)发表关于网状模型的 DBTG 报告(1971 年正式通过);1970 年,IBM 公司的 E. F. Codd 发表论文提