



普通高等教育“十一五”国家级规划教材

数据库系统教程

(第3版)

施伯乐 丁宝康 汪卫 编著



高等教育出版社
Higher Education Press

主要内容

普通高等教育“十一五”国家级规划教材

数据库系统教程

(第3版)

施伯乐 丁宝康 汪卫 编著

高等教育出版社

图书在版编目(CIP)数据

数据库系统教程 / 施伯乐, 丁宝康, 汪卫编著. —3版.
—北京: 高等教育出版社, 2008.7
ISBN 978-7-04-024224-9

I. 数… II. ①施…②丁…③汪… 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 050346 号

策划编辑 刘 茜 责任编辑 关 旭 封面设计 于文燕 责任绘图 尹 莉
版式设计 陆瑞红 责任校对 胡晓琪 责任印制 韩 刚

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
总 机 010-58581000

经 销 蓝色畅想图书发行有限公司
印 刷 北京中科印刷有限公司

开 本 787×1092 1/16
印 张 26.75
字 数 600 000

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landracom.com>
<http://www.landracom.com.cn>
畅想教育 <http://www.widedu.com>

版 次 1999 年 12 月第 1 版
2008 年 7 月第 3 版
印 次 2008 年 7 月第 1 次印刷
定 价 33.20 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 24224-00

内容提要

本书第1版列为教育部“面向21世纪课程教材”。第2版作了较大的修改和补充,增加了关系逻辑、递归SQL、对象联系图、UML的类图、ODMG标准、ODBC和XML等内容。这次第3版被教育部列为普通高等教育“十一五”国家级规划教材,对内容又作了修订和充实。

本书详细介绍数据库系统的基本原理、方法和发展趋势。全书分为上、下两篇。上篇为基础篇,介绍传统的数据库技术,内容包括:数据库概论、关系模型和关系运算理论、关系数据库语言SQL、关系数据库的规范化设计、数据库设计与ER模型、数据库的存储结构和系统实现技术。下篇为发展篇,介绍数据库技术的发展,内容包括:对象数据库系统、分布式数据库系统、ODBC技术、XML技术和现代信息集成技术。

本书是作者长期从事数据库教学和科研活动所获成果的结晶。全书内容丰富,叙述清楚、严谨,每章后面均有适量的习题,适于教和学。另外,还有与本书配套的习题解答等出版物。

本书可作为高等学校计算机相关专业的数据库课程教材,也可供从事信息领域工作的科技人员和工程技术人员以及其他有关人员参阅。

第 3 版前言

《数据库系统教程》第 1 版由施伯乐、丁宝康、周傲英、田增平 4 人编著，被教育部列为“面向 21 世纪课程教材”。自 1999 年 12 月出版以来，在计算机界同仁和学生那里得到了很大的反响，并于 2002 年获教育部优秀教材二等奖。

该书第 2 版由施伯乐、丁宝康、汪卫 3 人编著，于 2003 年 8 月出版，对第 1 版作了较大的修改和补充。以第 2 版为基础的数据库“教材建设”和“教学方法研究”获得了成功，并于 2005 年获得国家级教学成果二等奖。

目前第 3 版又被教育部列为普通高等教育“十一五”国家级规划教材，为此我们对第 2 版的内容作了修订和充实。具体改动如下：

(1) 删去了原来的第 11 章“数据库与 WWW”，因为这部分内容放到“Web 数据库技术”教材中更合适。

(2) 为适应教学，对原来的第 12 章“XML 技术”进行了改写，现为第 11 章。

(3) 增加了新的第 12 章“现代信息集成技术”，介绍数据仓库、联机分析处理和数据挖掘等内容。

(4) 根据数据库技术的发展和教学的需要，对原来的第 1~10 章的内容进行了修改和补充，对例子进行了改写。

(5) 把全书分为上、下两篇。前 7 章为基础篇，介绍传统的数据库技术；后 5 章为发展篇，介绍数据库技术的发展内容。

本书的主要目的是使读者掌握数据库技术的基本原理、方法和应用技术，能有效地使用现有的数据库管理系统，掌握数据库结构的设计和数据库应用系统的开发方法，同时也适时地掌握数据库技术发展的动向。

全书上、下两篇共分 12 章，具体内容如下。

第 1 章介绍数据库技术的基本概念，包括数据模型、数据库的体系结构和数据库系统的全局结构。

第 2 章介绍关系模型和关系运算理论，包括关系代数、关系演算和关系逻辑。

第 3 章介绍关系数据库标准语言 SQL 的数据定义、操纵、嵌入式等内容。

第 4 章介绍关系数据库的规范化设计理论，包括函数依赖、模式分解特性、范式和模式设计方法等问题。

第 5 章介绍数据库设计的全过程、ER 模型以及 ER 模型到关系模型的转换规则。

第 6 章介绍数据库的存储结构，包括文件结构、索引技术和散列技术等内容。

第 7 章介绍数据库系统实现的技术，包括事务概念及恢复、并发控制、完整性和安全性等

实现机制。

第8章先介绍面向对象概念建模的两种图示形式：对象联系图和UML类图。然后介绍对象数据库系统的两种形式（ORDB和OODB），介绍其定义语言和操纵语言。

第9章介绍分布式数据库系统的基本概念、数据存储、体系结构和查询处理等内容。

第10章介绍ODBC技术的体系结构、接口和SQL CLI等内容。

第11章介绍XML技术的由来、语法成分、文档类型定义、XML模式等内容。

第12章介绍现代信息集成技术中的DW、OLAP和DM等内容以及DSS的发展历程。

本书对某些较深的内容，在章节前加了“*”号；在每一道例题或定理的结尾处加了“□”符号，以示与正文的区别。

全书内容丰富，具有理论性、实用性和可操作性。每章后面均有适量的习题，以配合对知识点的掌握。教师讲授时应根据需要对内容作适当取舍。

本书可采取如下的教学方法：

(1) 讲授的重点应是基础篇中的第1~5章以及第7章。对于第2章和第4章中理论性较强的内容（目录中加“*”的节），可作适当压缩。第2章的重点是关系代数，第4章的重点是函数依赖和范式。

(2) 对于第6章（存储结构）的内容，教师可作适当引导让学生自学，以拓宽知识面。

(3) 对于发展篇中的内容，教师可作适当引导让学生自学，以提高学生的工作能力和研究水平。建议首选第8章对象数据库系统的内容。

本书是我们数据库教学小组30年来教学和科研成果的结晶，凝聚了几代人的心血。其中老一辈教师有罗文化、楼荣生、招兆铿、于玉、陈金海等，年轻一代的教师有顾宁、朱扬勇、张亮等。他们为复旦大学的数据库研究做出了贡献。

本书第3版由施伯乐、丁宝康、汪卫共同对第2版进行了修改和补充。参与修订工作的还有陈长洪、薛剑虹、朱顺虎、杨卫稼、张子鋈、曾宇昆、许建军、周华奇、张溢、吴爱华、庞引明、谢闽峰和吴清茂等。

上海交通大学尤晋元教授和东华大学乐嘉锦教授仔细审阅了第3版全稿，并提出许多宝贵的意见，在此向他们致以衷心的感谢。

为了满足大家对使用本书作为主教材教和学的需要，我们还组织编写了习题解答，随之一起出版。本书的教学课件即将传到网上，需要者可到中国高校计算机课程网 <http://computer.cncourse.com> 处下载。

限于水平，书中欠妥之处，敬请广大读者和专家批评指正。对本书的意见请按地址 dn@citiz.net 反馈给我们，谢谢。

施伯乐

2008年1月于复旦大学

II 目 录

2.4.2 关系代数表达式的等价变换规则	(60)	3.5 视图	(100)
2.4.3 关系代数表达式的启发式 优化算法	(62)	3.5.1 视图的创建和撤销	(100)
*2.5 关系逻辑	(65)	3.5.2 视图的更新	(101)
2.5.1 关系逻辑的成分	(65)	3.6 嵌入式 SQL	(102)
2.5.2 规则的安全性	(66)	3.6.1 嵌入式 SQL 的实现方式	(102)
2.5.3 从关系代数到关系逻辑的转换	(67)	3.6.2 嵌入式 SQL 的使用规定	(102)
2.5.4 递归过程	(70)	3.6.3 嵌入式 SQL 的使用技术	(104)
2.5.5 关系逻辑与关系代数的差异	(70)	3.6.4 动态 SQL 语句	(107)
2.6 小结	(70)	3.7 存储过程与 SQL/PSM	(108)
习题 2	(71)	3.7.1 数据库存储过程与函数	(108)
<hr/>		3.7.2 SQL/PSM	(110)
第 3 章 关系数据库语言 SQL	(74)	3.8 小结	(111)
3.1 SQL 简介	(74)	习题 3	(112)
3.1.1 SQL 的产生和发展	(74)	<hr/>	
3.1.2 SQL 数据库的体系结构	(75)	第 4 章 关系数据库的规范化设计	(116)
3.1.3 SQL 的组成	(76)	4.1 关系模式的设计问题	(116)
3.1.4 SQL 的特点	(76)	4.1.1 关系模式的外延和内涵	(116)
3.2 SQL 的数据定义	(76)	4.1.2 关系模式的冗余和异常问题	(116)
3.2.1 SQL 模式的创建和撤销	(76)	4.1.3 关系模式的非形式化设计准则	(118)
3.2.2 基本数据类型	(77)	4.1.4 本章的符号规定	(118)
3.2.3 基本表的创建、修改和撤销	(78)	4.2 函数依赖	(118)
3.2.4 索引的创建和撤销	(80)	4.2.1 函数依赖的定义	(119)
3.3 SQL 的数据查询	(81)	4.2.2 FD 的逻辑蕴涵	(120)
3.3.1 SELECT 查询语句的基本结构	(81)	4.2.3 FD 的推理规则	(120)
3.3.2 SELECT 语句的完整结构	(87)	4.2.4 FD 和关键码的联系	(122)
3.3.3 数据查询中的限制和规定	(89)	4.2.5 属性集的闭包	(122)
3.3.4 条件表达式中的比较操作	(90)	4.2.6 FD 推理规则的完备性	(123)
3.3.5 嵌套查询的改进写法	(94)	4.2.7 FD 集的最小依赖集	(123)
3.3.6 基本表的连接操作	(95)	4.3 关系模式的分解特性	(124)
*3.3.7 SQL3 中的递归查询	(96)	4.3.1 模式分解问题	(124)
3.4 SQL 的数据更新	(97)	4.3.2 无损分解	(125)
3.4.1 数据插入	(97)	4.3.3 模式分解的优缺点	(127)
3.4.2 数据删除	(98)	4.3.4 无损分解的测试方法	(127)
3.4.3 数据修改	(99)	4.3.5 保持函数依赖的分解	(129)
		4.3.6 模式分解与模式等价问题	(130)

4.4 关系模式的范式····· (130)	5.4 ER 模型实例分析····· (171)
4.4.1 第一范式····· (131)	5.4.1 库存管理信息系统的 ER 模型及转换····· (171)
4.4.2 第二范式····· (131)	5.4.2 公司车队信息系统的 ER 模型····· (173)
4.4.3 第三范式····· (132)	5.4.3 人事管理信息系统的 ER 模型····· (174)
4.4.4 BCNF····· (133)	5.4.4 旅游管理信息系统的 ER 模型····· (175)
4.4.5 分解成 BCNF 模式集的分解算法····· (134)	5.5 增强的 ER 模型····· (176)
4.4.6 分解成 3NF 模式集的合成算法····· (135)	5.5.1 弱实体与强实体····· (176)
4.4.7 模式设计方法小结····· (135)	5.5.2 子类实体与超类实体····· (177)
*4.5 模式的进一步规范化处理····· (135)	5.6 小结····· (178)
4.5.1 多值依赖的定义····· (136)	习题 5····· (179)
4.5.2 关于 FD 和 MVD 的推理规则集····· (137)	
4.5.3 第四范式····· (138)	
4.5.4 嵌入多值依赖····· (138)	
4.5.5 连接依赖和第五范式····· (139)	
4.6 小结····· (141)	
习题 4····· (142)	
<hr/>	
第 5 章 数据库设计与 ER 模型····· (147)	第 6 章 数据库的存储结构····· (182)
5.1 数据库设计的全过程····· (147)	6.1 文件组织····· (182)
5.1.1 规划阶段····· (148)	6.1.1 定长记录····· (182)
5.1.2 需求分析阶段····· (148)	6.1.2 变长记录····· (184)
5.1.3 概念设计阶段····· (149)	6.2 文件结构····· (187)
5.1.4 逻辑设计阶段····· (150)	6.2.1 4 种文件结构····· (187)
5.1.5 物理设计阶段····· (151)	6.2.2 顺序文件····· (188)
5.1.6 数据库的实现····· (151)	6.2.3 聚集文件····· (188)
5.1.7 数据库的运行与维护····· (153)	6.3 索引技术····· (189)
5.2 ER 模型····· (154)	6.3.1 索引机制····· (189)
5.2.1 ER 模型的基本元素····· (154)	6.3.2 有序索引的分类····· (190)
5.2.2 属性的分类····· (156)	6.3.3 主索引····· (190)
5.2.3 联系的设计····· (158)	6.3.4 辅助索引····· (193)
5.2.4 ER 模型的操作····· (161)	6.3.5 B ⁺ 树索引文件····· (194)
5.2.5 采用 ER 模型的数据库概念设计····· (162)	6.3.6 B 树索引文件····· (200)
5.3 ER 模型到关系模型的转换····· (168)	6.4 散列技术····· (201)
5.3.1 ER 图转换成关系模式集的算法····· (168)	6.4.1 散列机制····· (201)
5.3.2 采用 ER 模型的逻辑设计步骤····· (171)	6.4.2 散列索引····· (204)
	6.4.3 静态散列中的问题····· (205)
	6.4.4 可扩充散列结构····· (205)
	6.5 多键访问····· (210)
	6.5.1 单键查询的问题····· (210)
	6.5.2 网格文件····· (211)

IV 目 录

6.5.3 分区散列技术	(212)	7.3.1 并发操作带来的 3 个问题	(224)
6.6 小结	(212)	7.3.2 封锁技术	(226)
习题 6	(213)	7.3.3 封锁带来的问题	(229)
<hr/>			
第 7 章 系统实现技术	(216)	7.3.4 并发操作的调度	(231)
7.1 事务	(216)	7.3.5 SQL 对事务并发处理的支持	(232)
7.1.1 事务的定义	(216)	*7.3.6 基于时标的并发控制	(232)
7.1.2 事务的 ACID 性质	(217)	7.4 数据库的完整性	(234)
7.1.3 事务的状态变迁图	(218)	7.4.1 完整性子系统	(234)
7.2 数据库的恢复	(219)	7.4.2 SQL 中的完整性约束	(235)
7.2.1 存储器结构	(219)	*7.4.3 SQL3 中的触发器	(239)
7.2.2 恢复的基本原则和实现方法	(221)	7.5 数据库的安全性	(242)
7.2.3 故障类型和恢复方法	(222)	7.5.1 安全性问题	(242)
7.2.4 检查点技术	(223)	7.5.2 SQL 中的安全性机制	(243)
7.2.5 SQL 对事务的支持	(224)	7.5.3 常用的安全性措施	(246)
7.3 数据库的并发控制	(224)	7.6 小结	(249)
		习题 7	(250)
<hr/>			
下 篇 发 展 篇			
<hr/>			
第 8 章 对象数据库系统	(255)	8.4.3 引用类型的定义	(270)
8.1 面向对象的数据类型系统	(256)	8.5 ORDB 的查询语言	(272)
8.2 对象联系图	(257)	8.5.1 对 SELECT 语句的新规定	(273)
8.2.1 对象联系图的成分	(257)	8.5.2 嵌套与解除嵌套	(274)
8.2.2 数据的概化/特化	(259)	8.5.3 复合值的创建和查询	(275)
8.3 使用 UML 类图对概念对象建模	(260)	8.5.4 Oracle 中查询的两种技术	(276)
8.3.1 统一建模语言概述	(260)	8.5.5 函数和过程	(277)
8.3.2 用类图表达类和关联	(260)	8.6 OODBS 的基本概念	(281)
8.3.3 用类图表达关联类	(263)	8.6.1 ODMG 标准	(282)
8.3.4 用类图表达概化/特化	(264)	8.6.2 OODBS 的定义	(283)
8.3.5 用类图表达聚合	(265)	8.6.3 OODB 的基本概念	(283)
8.4 ORDB 的定义语言	(266)	8.7 ODMG 对象模型	(285)
8.4.1 从关系模型到对象关系模型的演变	(266)	8.7.1 对象和文字	(286)
8.4.2 继承性的定义	(268)	8.7.2 接口、类和继承	(287)
		8.7.3 类外延、关键码和工厂对象	(288)
		8.7.4 ODMG OD L	(289)

8.8 ODMG OQL.....	(290)	9.5.3 基于连接的优化方法.....	(321)
8.8.1 OQL 中的 SELECT 语句.....	(291)	9.6 分布式数据库中的并发控制和恢复技术.....	(322)
8.8.2 OQL 表达式的附加格式.....	(293)	9.6.1 DDB 中的问题.....	(322)
8.8.3 OQL 中对象的赋值和建立.....	(295)	9.6.2 基于数据项识别副本的分布式并发控制.....	(322)
8.8.4 C++语言的绑定.....	(296)	9.6.3 基于投票方法的分布式并发控制.....	(323)
8.9 OODB 与 RDB、ORDB 的比较.....	(297)	9.6.4 分布式恢复.....	(324)
8.9.1 OODB与RDB在概念设计上的区别.....	(297)	9.7 小结.....	(324)
8.9.2 OODB 与 ORDB 的比较.....	(298)	习题 9.....	(325)
8.10 小结.....	(299)		
习题 8.....	(300)		
<hr/>		<hr/>	
第 9 章 分布式数据库系统.....	(303)	第 10 章 ODBC 技术.....	(327)
9.1 DDBS 的定义和特点.....	(303)	10.1 中间件.....	(327)
9.1.1 从集中式、分散式到分布式.....	(303)	10.1.1 中间件的定义.....	(327)
9.1.2 DDBS 的定义.....	(304)	10.1.2 中间件的作用.....	(328)
9.1.3 DDBS 的特点.....	(305)	10.2 ODBC 简介.....	(329)
9.1.4 DDBS 的优缺点.....	(306)	10.2.1 ODBC 的概念.....	(329)
9.1.5 DDBS 的分类.....	(308)	10.2.2 ODBC 的体系结构.....	(330)
9.2 分布式数据存储.....	(308)	10.2.3 ODBC 的特性.....	(333)
9.2.1 数据分片.....	(308)	10.3 ODBC 接口.....	(334)
9.2.2 数据分配.....	(309)	10.3.1 ODBC 应用程序的基本流程.....	(334)
9.3 DDB 的体系结构.....	(310)	10.3.2 ODBC 句柄.....	(335)
9.3.1 体系结构.....	(311)	10.3.3 数据源的连接与断开.....	(337)
9.3.2 分布透明性.....	(313)	10.3.4 SQL 语句的执行.....	(337)
9.4 DDBMS.....	(314)	10.3.5 查询结果的获取.....	(339)
9.4.1 DDBMS 的组成.....	(314)	10.4 ODBC 的符合性级别.....	(340)
9.4.2 DDBMS 的功能.....	(315)	10.4.1 API 符合性的 3 个级别.....	(340)
9.4.3 DDBMS 的组成.....	(315)	10.4.2 SQL 符合性的 3 个级别.....	(341)
9.4.4 DDBMS 的同构性程度和局部自治性程度.....	(316)	10.4.3 ODBC API与SQL CLI之间的协调.....	(342)
9.4.5 FDBS 的异构性.....	(317)	10.4.4 SQL CLI 与嵌入式 SQL 的比较.....	(343)
9.4.6 FDBS 的 5 层模式结构.....	(318)	10.4.5 SQL CLI 的应用.....	(343)
9.5 分布式查询处理.....	(318)	10.4.6 典型的数据库应用系统开发工具.....	(347)
9.5.1 查询代价的估算方法.....	(318)		
9.5.2 基于半连接的优化策略.....	(319)		

10.5 小结	(347)	11.6 XML 数据的存储	(373)
习题 10	(348)	11.6.1 关系数据库	(373)
<hr/>			
第 11 章 XML 技术	(349)	11.6.2 非关系的数据存储	(374)
11.1 XML 的由来	(349)	11.7 小结	(375)
11.1.1 从 SGML、HTML 到 XML	(349)	习题 11	(375)
11.1.2 XML 和 HTML 的区别	(350)	<hr/>	
11.1.3 XML 在数据交换中的作用	(352)	第 12 章 现代信息集成技术	(377)
11.2 XML 文档的语法成分	(352)	12.1 数据仓库	(377)
11.2.1 XML 声明	(352)	12.1.1 DW 概述	(377)
11.2.2 元素	(353)	12.1.2 DW 的组织结构	(380)
11.2.3 属性	(355)	12.1.3 DW 存储的多维数据模型	(381)
11.2.4 引用	(355)	12.1.4 数据处理	(383)
11.2.5 注释	(356)	12.1.5 DW 的设计和发展阶段	(387)
11.2.6 名字空间	(356)	12.2 联机分析处理技术	(391)
11.3 文档类型定义	(357)	12.2.1 OLAP 概述	(392)
11.3.1 元素类型声明	(357)	12.2.2 OLAP 的数据组织	(396)
11.3.2 属性声明	(358)	12.2.3 OLAP 应用开发实例	(398)
11.3.3 DTD 的局限性	(361)	12.3 数据挖掘	(399)
11.4 XML 模式	(361)	12.3.1 DM 概述	(400)
11.4.1 XML 模式的示例	(361)	12.3.2 DM 与 DW、OLAP 的联系与区别	(401)
11.4.2 XML 模式的基本成分	(363)	12.3.3 DM 应用过程	(402)
11.4.3 XML 模式的优点	(366)	12.3.4 DM 的分析方法和用到的技术	(404)
11.5 XML 查询和转换	(366)	12.3.5 DM 的应用领域	(405)
11.5.1 XPath	(367)	12.4 新决策支持系统概述	(406)
11.5.2 XSLT	(369)	12.4.1 DSS 的发展历程	(406)
11.5.3 XQuery	(371)	12.4.2 综合 DSS 的结构图	(408)
11.5.4 XML API	(372)	12.5 小结	(410)
参考文献	(413)	习题 12	(411)

上 篇

基 础 篇

在 20 世纪 60 年代末, 随着软件工程的诞生, 数据库技术也作为软件学科的一个分支应运而生了。经过 40 年的发展, 已形成较为完整的数据库理论体系和实用技术。

本篇介绍传统的数据库技术, 分为 7 章: 第 1 章数据库概论, 第 2 章关系模型和关系运算理论, 第 3 章关系数据库语言 SQL, 第 4 章关系数据库的规范化设计, 第 5 章数据库设计与 ER 模型, 第 6 章数据库的存储结构, 第 7 章系统实现技术。

本篇第 2 章中的关系演算和关系逻辑、第 3 章中的递归查询、第 4 章中的多值依赖和连接依赖、第 7 章中的时标技术和触发器等内容, 属于较深的内容, 初学者或非计算机专业的学生可以跳过它, 不作为课程内容。

第 2 章中的关系代数、第 3 章中的 SELECT 语句、第 4 章中的函数依赖和范式、第 5 章中的 ER 模型、第 7 章中的 SQL 实现技术等内容, 是学生必须掌握的内容, 也是考核的重点, 这些内容也是传统数据库技术的精髓所在。

第 6 章“数据库的存储结构”实际上是“数据结构”内容的继续和延伸。由于课程时间的限制, 建议教师作适当引导, 让学生自学并撰写读书笔记, 以拓宽知识面。

第 1 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逐渐扩展到企业、行政部门, 数据处理已成为计算机的主要应用。在 60 年代末, 数据库技术是作为数据处理中的一门新技术发展起来的。经过近 40 年的发展, 数据库技术已是计算机软件领域的一个重要分支, 形成了较为完整的理论体系和实用技术。

本章介绍数据库系统的基本概念, 使读者对数据库概貌有个基本的了解。1.1 节为引言, 1.2 节介绍数据库的由来和发展, 1.3 节介绍数据库设计中各阶段的数据描述, 1.4 节介绍数据库设计中各阶段的数据抽象(即数据模型), 1.5 节介绍数据库管理系统的工作模式和主要功能, 1.6 节介绍数据库系统的组成和全局结构。

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。如何组织和利用如此庞杂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。

今天, “数据库”这个名词已是家喻户晓, 成为人们日常生活中不可缺少的一部分。例如, 人们可能会去银行取款, 可能需要预订飞机票, 可能要在网上购物等, 数据库的应用领域已遍及现代社会的每一个角落。根据调查, 数据库(Database)这个名词起源于 20 世纪 50 年代初, 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该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, 但基本上都是文件系统的扩充或倒排文件系统。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, 数据库系统语言协会) 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 发表论文提出关系模型。这 3 件事奠定了现代数据库技术的基础。

20 世纪 70 年代是数据库蓬勃发展的年代, 网状系统和层次系统占据了整个数据库商用市场, 而关系系统仅处于实验阶段。许多计算机厂商研制了各种数据库管理系统, 许多学者发表了大量研究性论文, 使数据库技术在实践中和理论上得到飞速发展。数据库技术也日趋成熟。

20 世纪 80 年代, 关系系统由于使用简便以及硬件性能的改善, 逐步代替网状系统和层次系统占领了市场。同时关系数据库的理论日趋完善, 走向更高级的阶段。数据库的理论以及新系统的研究继续进行, 不断深化。此时关系数据库语言 SQL 成了国际标准语言。

20 世纪 90 年代, 关系数据库已成为数据库技术的主流。由于受到计算机应用领域以及其他分支学科的影响, 数据库技术与面向对象技术、网络技术相互渗透, 人们在对象数据库技术和 Web 数据库技术方面进行了深入的研究。

由于 C.W.Bachman 在网状模型和 DBTG 报告中的贡献, 使得他在 1973 年荣获美国计算机学会 (ACM) 图灵奖。E.F.Codd 在关系模型上做出了杰出的开拓性贡献, 在 1981 年获得了图灵奖。J.Gray 在数据库技术的实用化和事务管理方面发挥了关键的作用, 因而在 1999 年也获得了图灵奖。

进入 21 世纪以后, 无论是市场的需求还是技术条件的成熟, 对象数据库技术、Web 数据库技术的推广和普及已成定局。数据库技术使得计算机应用迅速渗透到工农业生产、商业、行政、科学研究、工程技术和国防军事各个部门, 渗透到社会的每一角落, 并改变着人们的工作方式和生活方式。管理信息系统 (MIS)、办公信息系统 (OIS)、计算机集成制造系统 (CIMS)、地理信息系统 (GIS)、因特网 (Internet) 技术等都是使用了数据库技术的计算机应用系统。

1.2 数据库的由来和发展

要进行数据处理, 首先要把大量的数据存放在存储器中。存储器的容量、存储速率直接影响到数据管理技术的发展。1956 年生产的第一张磁盘, 其容量仅为 5 MB, 而现在已经突破 1 TB (见表 1.1)。目前光盘已经被广泛应用, 容量已达数个 GB。存储器的发展, 为数据库技术提供了良好的物质基础。

表 1.1 磁盘容量的发展

时间 (年)	1956	1965	1971	1978	1981	1985	1995	2003	2007
容量/轴	5 MB	30 MB	100 MB	600 MB	1.2 GB	5 GB	10 GB	180 GB	1 TB

注: 1 TB=1 024 GB, 1 GB=1 024 MB

使用计算机以后, 数据处理的速度和规模无论相对于手工方式还是机械方式都是无可比拟的, 随着数据处理量的增长, 产生了数据管理技术。数据管理技术的发展, 与计算机硬件 (主要是外部存储器)、系统软件及计算机应用的范围有着密切的联系。数据管理技术的发展经历了人工管理、文件系统、数据库阶段和高级数据库阶段。下面分别介绍。

1.2.1 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,其他工作还没有展开。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。数据处理的方式基本上是批处理。这个时期的数据管理有下列特点:

(1) 数据不保存在计算机内。计算机主要用于计算,一般不需要长期保存数据。在进行某一课题计算时,将原始数据随程序一起输入内存,运算处理后将结果数据输出。随着计算任务的完成,用户作业退出计算机系统,数据空间随着程序空间一起被释放。

(2) 没有专用的软件对数据进行管理。每个应用程序都要包括存储结构、存取方法、输入/输出方式等内容。程序中的存取子程序随着存储结构的改变而改变,因而数据与程序不具有独立性。存储结构改变时,应用程序必须改变。此时,由于程序直接面向存储结构,因此数据的逻辑结构与物理结构没有区别。

(3) 只有程序(Program)的概念,没有文件(File)的概念。数据的组织方式必须由程序员自行设计与安排。

(4) 数据面向程序,即一组数据对应一个程序。

1.2.2 文件系统阶段

在这一阶段(20世纪50年代后期至60年代后期),计算机不仅用于科学计算,还用于信息管理。随着数据量的增加,解决数据的存储、检索和维护问题成为紧迫的需要,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、磁鼓等直接存取存储设备。1956年生产的第一张磁盘,其容量仅为5MB,20世纪60年代已达到数十个MB。软件领域出现了高级语言和操作系统。操作系统中的文件系统是专门管理外存的数据管理软件。数据处理的方式有批处理,也有联机实时处理。

这一阶段的数据管理有以下特点:

(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(读/写)。

(3) 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。

(4) 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。但是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。

□ 在文件系统阶段,由于具有设备独立性,因此当改变存储设备时,不必改变应用程序。但