面 向 2 1 世 纪 课 程 教 材

Textbook Series for 21st Century

数据库系统教程

(第 2 版)

施伯乐　丁宝康　汪　卫　编著

需要全本请访问 www.ertongbook.com

内 容 提 要

本书第1版列为教育部“面向21世纪课程教材”。这次再版时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、JDBC和XML等内容。

本书详细介绍数据库系统的基本原理、方法和发展趋势。内容包括:数据库系统的基本概念、关系模型的运算理论、SQL语言、规范化设计理论、数据库设计与ER模型、存储结构、系统实现技术、对象数据库、分布式数据库、中间件技术、数据库与WWW、XML技术。

本书第1版在2002年获教育部优秀教材二等奖。本书是作者长期从事数据库数学和科研活动所获得的结晶。全书内容丰富,叙述严谨清楚,每章后均有适量的习题,适于教和学。与本书配套,还有习题解答书等出版物。

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 2 版前言

《数据库系统教程》第 1 版自 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。

进入 21 世纪以后,无论是市场的需求还是技术条件的成熟,对象数据库技术和网络数据库技术的推广和普及已成定局,这些技术将成为新时代数据库技术应用的主流技术。为此,我们对第 1 版的结构作了如下的改动:

在第 2 章关系运算理论中增加了关系逻辑内容,在第 3 章的 SQL 语言中增加了递归查询内容,并把 QBE 内容缩写成对 SELECT 语句的图示解释。

把数据库设计从第 4 章中抽出来,单独成为新的第 5 章,并着重介绍 ER 模型,并列举了四个较大的 ER 模型实例。原来的第 5、6 章就顺延为新的第 6、7 章。在第 7 章中,对系统实现技术进行了改写,读起来更为通顺。

原来的第 8、9 章内容改写为新的第 8 章对象数据库系统,并增加了 ODMG2.0 标准内容,以及使用 UML 类图概念对象建模的内容。同时删去了原来的第 10 章。

对原来的第 7 章分布式数据库内容,改写成新的第 9 章。

增加了三章新的内容:第 10 章中间件,第 11 章数据库与 WWW,第 12 章 XML 技术。

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式,同时也适时地掌握对象数据库和网络数据库等技术。

全书共分 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,包括存储介质层次、文件结构、索引技术和散

列技术等内容。

第 7 章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,还介绍了概念对象建模的两种图示形式:对象联系图和 UML 的类图。

第 9 章介绍分布计算的两种形式:客户机/服务器系统和分布式数据库系统。

第 10 章介绍数据库技术的两种中间件技术:ODBC 技术和 JDBC 技术。

第 11 章介绍数据库与 WWW 的关系和结合,以及 JSP 和 PHP 等语言。

第 12 章介绍 XML 技术的基本概念和 XML 查询语言。

本书对某些较深的内容,在章节前加了符号"*";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) 讲授的重点应是第 1～5、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 章的存储结构内容,教师可作适当引导让学生自学,以拓宽知识面。

(3) 对于第 8 章的对象数据库内容,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。

(4) 对于第 9～12 章,教师可有针对性地选择某些内容向学生讲授。

本书是我们数据库教学小组 20 多年教学和科研成果的结晶,凝聚了几代人的心血。在此向给青年教师做出表率的罗文化、楼荣生等老教师致以崇高的敬意。本次修订工作第 5 章由许建军执笔,第 10 章由周华奇执笔,第 11、12 章由汪卫执笔,其余章节由丁宝康执笔。参与本书修订工作的还有陈长洪、张谧、吴爱华、庞引明、谢闽峰、吴清茂和曾宇昆等老师。最后由施伯乐、丁宝康和汪卫一起对全稿进行了修改和统稿。

为了满足大家对使用本书作为主教材的教和学的要求,我们还组织了习题解答书,随之一起出版。

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

2003 年 5 月于复旦大学

第1版前言

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。

需要全本请访问 www.ertongbook.com

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

目　录

第 1 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　数据库技术的基本术语 …… (4)
1.2　数据库的由来和发展 …… (4)
1.2.1　人工管理阶段 …… (5)
1.2.2　文件系统阶段 …… (5)
1.2.3　数据库阶段 …… (7)
1.2.4　高级数据库技术阶段 …… (9)
1.3　数据描述 …… (11)
1.3.1　概念设计中的数据描述 …… (11)
1.3.2　逻辑设计中的数据描述 …… (12)
1.3.3　物理设计中的数据描述 …… (13)
1.3.4　数据联系的描述 …… (13)
1.4　数据模型 …… (15)
1.4.1　数据模型的定义 …… (15)
1.4.2　实体联系模型 …… (16)
1.4.3　层次模型 …… (18)
1.4.4　网状模型 …… (19)
1.4.5　关系模型 …… (21)
1.4.6　对象模型 …… (22)
1.5　数据库的体系结构 …… (23)
1.5.1　数据库的三级体系结构 …… (23)
1.5.2　体系结构中的五个要素 …… (25)
1.5.3　两级数据独立性 …… (27)
1.5.4　用户与用户界面 …… (27)
1.5.5　本节小结 …… (27)
1.6　数据库管理系统 …… (28)
1.6.1　DBMS 的工作模式 …… (28)
1.6.2　DBMS 的主要功能 …… (29)
1.6.3　DBMS 的模块组成 …… (31)
1.7　数据库系统 …… (31)
1.7.1　DBS 的组成 …… (31)
1.7.2　DBS 的全局结构 …… (33)
1.7.3　DBS 结构的分类 …… (35)
1.7.4　DBS 的效益 …… (38)
小结 …… (39)
习题 1 …… (40)

第 2 章　关系模型和关系运算理论 …… (42)
2.1　关系模型的基本概念 …… (42)
2.1.1　基本术语 …… (42)
2.1.2　关系的定义和性质 …… (43)
2.1.3　关系模型的三类完整性规则 … (44)
2.1.4　ER 模型向关系模型的转换规则 …… (45)
2.1.5　关系模型的三级体系结构 …… (47)
2.1.6　关系模型的形式定义和优点 … (50)
2.1.7　关系查询语言和关系运算 …… (50)
2.2　关系代数 …… (51)
2.2.1　关系代数的五个基本操作 …… (51)
2.2.2　关系代数的四个组合操作 …… (53)
2.2.3　关系代数运算的应用实例 …… (56)
2.2.4　关系代数的七个扩充操作 …… (57)
*2.3　关系演算 …… (61)
2.3.1　元组关系演算 …… (61)

2.3.2 域关系演算 …… (64)
2.3.3 关系运算的安全约束和等价性 …… (66)
2.4 关系代数表达式的优化 …… (68)
2.4.1 关系代数表达式的优化问题 … (68)
2.4.2 关系代数表达式的等价变换规则 …… (69)
2.4.3 关系代数表达式的优化算法 … (71)
*2.5 关系逻辑 …… (76)
2.5.1 关系运算的成分 …… (76)
2.5.2 规则的安全性 …… (78)
2.5.3 从关系代数到关系逻辑的转换 …… (79)
2.5.4 递归过程 …… (82)
2.5.5 关系逻辑与关系代数的差异 … (84)
小结 …… (84)
习题 2 …… (85)

第 3 章 关系数据库语言 SQL …… (88)
3.1 SQL 的体系结构 …… (88)
3.1.1 SQL 的产生和发展 …… (88)
3.1.2 SQL 数据库的体系结构 …… (89)
3.1.3 SQL 的组成 …… (90)
3.2 SQL 的数据定义 …… (91)
3.2.1 SQL 模式的创建和撤销 …… (91)
3.2.2 基本数据类型 …… (92)
3.2.3 基本表的创建、修改和撤销 …… (92)
3.2.4 索引的创建和撤销 …… (95)
3.3 SQL 的数据查询 …… (96)
3.3.1 SELECT 查询语句 …… (96)
3.3.2 SELECT 语句完整的句法 …… (103)
3.3.3 SELECT 语句中的限定 …… (104)
3.3.4 基本表的连接操作 …… (110)
*3.3.5 SQL3 中的递归查询 …… (112)
3.4 SQL 的数据更新 …… (114)
3.4.1 数据插入 …… (114)
3.4.2 数据删除 …… (115)
3.4.3 数据修改 …… (116)
3.5 视图的定义和对视图的操作 …… (117)
3.5.1 视图的创建和撤销 …… (117)
3.5.2 对视图的更新操作 …… (118)
3.5.3 视图的优点 …… (119)
3.6 嵌入式 SQL …… (120)
3.6.1 SQL 的运行环境 …… (120)
3.6.2 嵌入式 SQL 的使用规定 …… (121)
3.6.3 嵌入式 SQL 的使用技术 …… (122)
3.6.4 动态 SQL 语句 …… (126)
小结 …… (127)
习题 3 …… (128)

第 4 章 关系数据库的规范化设计 …… (132)
4.1 关系模式的设计问题 …… (132)
4.1.1 关系模式的外延和内涵 …… (132)
4.1.2 关系模式的冗余和异常问题 …… (133)
4.1.3 本章的符号规定 …… (134)
4.2 函数依赖 …… (135)
4.2.1 函数依赖的定义 …… (135)
4.2.2 FD 的逻辑蕴涵 …… (136)
4.2.3 FD 的推理规则 …… (136)
4.2.4 FD 和关键码的联系 …… (138)
4.2.5 属性集的闭包 …… (138)
4.2.6 FD 推理规则的完备性 …… (139)
4.2.7 FD 集的最小依赖集 …… (140)
4.3 关系模式的分解特性 …… (141)
4.3.1 模式分解问题 …… (141)
4.3.2 无损分解 …… (142)
4.3.3 无损分解的测试方法 …… (144)
4.3.4 保持函数依赖的分解 …… (146)
4.3.5 模式分解与模式等价问题 …… (147)
4.4 关系模式的范式 …… (148)
4.4.1 第一范式(1NF) …… (149)

4.4.2　第二范式(2NF) …… (149)
4.4.3　第三范式(3NF) …… (150)
4.4.4　BCNF(Boyce - Codd NF) …… (152)
4.4.5　分解成 BCNF 模式集的算法 … (153)
4.4.6　分解成 3NF 模式集的算法 …… (154)
4.4.7　模式设计方法的原则 …… (154)
*4.5　模式的进一步规范化处理 …… (155)
4.5.1　多值依赖的定义 …… (155)
4.5.2　关于 FD 和 MVD 的推理规则集 …… (156)
4.5.3　第四范式(4NF) …… (157)
4.5.4　嵌入多值依赖 …… (158)
4.5.5　连接依赖和第五范式 …… (158)
小结 …… (160)
习题 4 …… (161)

第 5 章　数据库设计与 ER 模型 …… (166)
5.1　数据库设计概述 …… (166)
5.1.1　软件工程与软件生存期 …… (166)
5.1.2　数据库工程与数据库生存期 …… (167)
5.1.3　数据库设计的输入、输出 …… (168)
5.2　数据库设计的全过程 …… (169)
5.2.1　规划阶段 …… (169)
5.2.2　需求分析阶段 …… (170)
5.2.3　概念设计阶段 …… (171)
5.2.4　逻辑设计阶段 …… (172)
5.2.5　物理设计阶段 …… (174)
5.2.6　数据库的实现 …… (175)
5.2.7　数据库的运行与维护 …… (177)
5.3　ER 模型 …… (178)
5.3.1　ER 模型的基本元素 …… (178)
5.3.2　属性的分类 …… (180)
5.3.3　联系的设计 …… (182)
5.3.4　ER 模型的操作 …… (185)
5.3.5　采用 ER 方法的数据库概念设计 …… (186)
5.4　ER 模型到关系模型的转换 …… (193)
5.4.1　ER 图转换成关系模式集的规则 …… (193)
5.4.2　采用 ER 方法的逻辑设计步骤 …… (194)
5.5　ER 模型实例分析 …… (195)
5.5.1　库存销售信息管理系统的 ER 模型及转换 …… (195)
5.5.2　公司车队信息系统的 ER 模型 …… (197)
5.5.3　人事管理信息系统的 ER 模型 …… (198)
5.5.4　旅游管理信息系统的 ER 模型 …… (199)
5.6　ER 模型的扩充 …… (200)
5.6.1　弱实体 …… (200)
5.6.2　子类实体与超类实体 …… (201)
小结 …… (202)
习题 5 …… (202)

第 6 章　数据库的存储结构 …… (205)
6.1　物理存储介质 …… (205)
6.1.1　存储介质层次 …… (205)
6.1.2　磁盘 …… (207)
6.1.3　第三级存储器 …… (210)
6.2　文件组织 …… (211)
6.2.1　定长记录 …… (211)
6.2.2　变长记录 …… (213)
6.3　文件结构 …… (216)
6.3.1　四种文件结构 …… (216)
6.3.2　顺序文件 …… (217)
6.3.3　聚集文件 …… (218)
6.4　索引技术 …… (219)
6.4.1　索引技术的分类 …… (219)
6.4.2　有序索引的分类 …… (219)

6.4.3 主索引 …… (220)
6.4.4 辅助索引 …… (223)
6.4.5 B^+ 树索引文件 …… (224)
6.4.6 B 树索引文件 …… (231)
6.5 散列技术 …… (232)
6.5.1 散列机制 …… (232)
6.5.2 散列索引 …… (235)
6.5.3 静态散列中的问题 …… (236)
6.5.4 可扩充散列结构 …… (236)
6.6 多键访问 …… (241)
6.6.1 单键查询的问题 …… (241)
6.6.2 网格文件 …… (242)
6.6.3 分区散列技术 …… (244)
小结 …… (245)
习题 6 …… (246)

第 7 章 系统实现技术 …… (248)
7.1 系统目录 …… (248)
7.1.1 关系 DBMS 的目录 …… (248)
7.1.2 系统目录的扩充 ER 图 …… (250)
7.1.3 系统目录和 DBMS 各子系统的联系 …… (250)
7.2 事务 …… (252)
7.2.1 事务的定义 …… (252)
7.2.2 事务的 ACID 性质 …… (253)
7.2.3 事务的状态变迁图 …… (255)
7.3 数据库的恢复 …… (256)
7.3.1 存储器结构 …… (257)
7.3.2 恢复的基本原则和实现方法 …… (259)
7.3.3 故障类型和恢复方法 …… (260)
7.3.4 检查点技术 …… (261)
*7.3.5 阴影页技术 …… (263)
7.3.6 SQL 对事务的支持 …… (265)
7.4 数据库的并发控制 …… (265)
7.4.1 并发操作带来的四个问题 …… (265)
7.4.2 封锁技术 …… (268)
7.4.3 封锁带来的问题 …… (272)
7.4.4 并发操作的调度 …… (274)
7.4.5 SQL 中事务的存取模式和隔离级别 …… (275)
*7.4.6 基于时标的并发控制 …… (276)
*7.5 并发事务的可串行化和可恢复性 …… (278)
7.5.1 并发事务的可串行化 …… (279)
7.5.2 并发事务的可恢复性 …… (284)
7.6 数据库的完整性 …… (286)
7.6.1 完整性子系统和完整性规则 …… (286)
7.6.2 SQL 中的完整性约束 …… (286)
*7.6.3 SQL3 中的触发器 …… (291)
7.7 数据库的安全性 …… (295)
7.7.1 安全性问题 …… (295)
7.7.2 权限问题 …… (296)
7.7.3 SQL 中的安全性机制 …… (297)
7.7.4 常用的安全性措施 …… (299)
小结 …… (303)
习题 7 …… (303)

第 8 章 对象数据库系统 …… (306)
8.1 对象联系图 …… (307)
8.1.1 从关系到嵌套关系、复合对象 …… (307)
8.1.2 引用类型 …… (309)
8.1.3 对象联系图的成分 …… (310)
8.1.4 数据的概化/特化 …… (312)
8.2 面向对象的数据类型系统 …… (313)
8.3 ORDB 的定义语言 …… (314)
8.3.1 ORDBS 的定义 …… (315)
8.3.2 数据类型的定义 …… (315)
8.3.3 继承性的定义 …… (316)
8.3.4 引用类型的定义 …… (318)

8.3.5　SQL3 中的定义语言 …… (319)
8.4　ORDB 的查询语言 …… (320)
8.4.1　以关系为值的属性 …… (320)
8.4.2　路径表达式 …… (321)
8.4.3　Oracle 中查询的两种技术 …… (322)
8.4.4　嵌套与解除嵌套 …… (324)
8.4.5　函数的定义和使用 …… (325)
8.4.6　复合值的创建和查询 …… (325)
8.5　OODBS 概述 …… (326)
8.5.1　ODMG 标准 …… (326)
8.5.2　OODBS 的定义 …… (327)
8.5.3　面向对象数据模型的概念 …… (328)
8.6　ODMG1.0 标准 …… (332)
8.6.1　持久化程序设计语言 …… (332)
8.6.2　ODMG C++ 对象定义语言 (C++ ODL) …… (334)
8.6.3　ODMG C++ 对象操纵语言 (C++ OML) …… (336)
8.7　ODMG2.0 标准 …… (337)
8.7.1　ODMG 对象模型 …… (337)
8.7.2　ODMG 对象定义语言(ODL) … (339)
8.7.3　ODMG 对象查询语言(OQL) … (340)
8.8　OODB 与 ORDB 的比较 …… (346)
*8.9　使用 UML 类图来概念对象建模 … (347)
8.9.1　统一建模语言(UML)概述 …… (347)
8.9.2　用类图表达类和关联 …… (348)
8.9.3　用类图表达关联类 …… (351)
8.9.4　用类图表达概化/特化 …… (352)
8.9.5　用类图表达聚合 …… (354)
小结 …… (355)
习题 8 …… (356)

第 9 章　分布式数据库系统 …… (359)
9.1　集中计算与分布计算 …… (359)
9.2　客户机/服务器系统 …… (361)
9.2.1　C/S 系统的计算模式 …… (361)
9.2.2　C/S 系统的体系结构 …… (363)
9.2.3　C/S 结构的演变 …… (365)
9.2.4　网络服务器的类型 …… (367)
9.3　分布式数据库系统概述 …… (369)
9.3.1　从集中式、分散式到分布式 … (369)
9.3.2　DDBS 的定义 …… (369)
9.3.3　DDBS 的特点与优缺点 …… (371)
9.3.4　DDBS 的分类 …… (374)
9.4　分布式数据存储 …… (375)
9.4.1　数据分片(Data Fragmentation) … (375)
9.4.2　数据分配(Data Allocation) …… (376)
9.5　分布式数据库的体系结构 …… (377)
9.5.1　体系结构 …… (378)
9.5.2　分布透明性 …… (380)
9.5.3　分布式数据库管理系统 …… (382)
9.6　分布式查询处理 …… (384)
9.6.1　查询例子分析 …… (384)
9.6.2　查询代价的估算方法 …… (386)
9.6.3　基于半连接的优化策略 …… (386)
9.6.4　基于连接的优化方法 …… (389)
9.7　分布式数据库中的并发控制和恢复技术 …… (390)
9.7.1　DDB 中的问题 …… (390)
9.7.2　基于数据项识别拷贝的分布式并发控制 …… (390)
9.7.3　基于投票方法的分布式并发控制 …… (392)
9.7.4　分布式恢复 …… (392)
小结 …… (393)
习题 9 …… (394)

第 10 章　中间件技术 …… (396)
10.1　中间件 …… (396)
10.1.1　中间件的定义 …… (396)
10.1.2　中间件的作用 …… (397)
10.2　ODBC 结构 …… (398)

需要全本请访问 www.ertongbook.com

10.2.1 ODBC 概念 …… (398)
10.2.2 ODBC 的体系结构 …… (399)
10.2.3 ODBC 的特性 …… (404)
10.3 ODBC 接口 …… (404)
10.3.1 ODBC 应用程序的基本流程 …… (404)
10.3.2 ODBC 句柄 …… (405)
10.3.3 数据源的连接与断开 …… (408)
10.3.4 SQL 语句的执行 …… (408)
10.3.5 查询结果的获取 …… (410)
10.4 ODBC 的符合性级别 …… (411)
10.4.1 API 符合性的三个级别 …… (411)
10.4.2 SQL 符合性的三个级别 …… (413)
10.4.3 ODBC API 与 SQL CLI 之间的协调 …… (414)
10.4.4 典型的 DB 应用系统开发工具 …… (414)
10.5 Java 和 JDBC …… (415)
10.5.1 Java 简介 …… (415)
10.5.2 JDBC 的提出 …… (417)
10.5.3 JDBC 的基本功能 …… (418)
10.5.4 JDBC 与 CGI 的比较 …… (418)
10.6 JDBC 的结构 …… (420)
10.6.1 JDBC 数据库设计方法 …… (420)
10.6.2 保持一致性的措施 …… (421)
10.6.3 JDBC 驱动程序 …… (422)
10.7 JDBC API …… (423)
10.7.1 JDBC API 的目标 …… (423)
10.7.2 JDBC API 接口概貌 …… (424)
10.7.3 JDBC 的接口和类 …… (426)
10.7.4 JDBC 数据库应用程序的编程 …… (427)
小结 …… (430)
习题 10 …… (431)

第 11 章 数据库与 WWW …… (433)
11.1 网络基本概念 …… (433)
11.1.1 Internet …… (433)
11.1.2 WWW …… (435)
11.1.3 网络协议模型 …… (435)
11.2 WWW 与数据库 …… (438)
11.2.1 WWW 与数据库的结合 …… (438)
11.2.2 WWW 与数据库交互的方法 …… (439)
11.2.3 CGI、JDBC 和 API 的比较 …… (443)
11.3 公共网关接口(CGI) …… (444)
11.3.1 CGI 程序 …… (444)
11.3.2 CGI 特点 …… (446)
11.4 ASP 技术 …… (449)
11.4.1 ASP 的特点 …… (449)
11.4.2 ASP 服务器端程序的原理 …… (450)
11.4.3 ASP 的对象存取 Mdb 数据库方法 …… (450)
11.5 JSP 技术 …… (452)
11.5.1 JSP 脚本语言 …… (452)
11.5.2 JSP 中的指令 …… (453)
11.5.3 JSP 中的内置对象(Implicit Objects) …… (455)
11.5.4 JSP 应用例子 …… (457)
11.6 PHP 技术 …… (460)
11.6.1 PHP 语言简介 …… (460)
11.6.2 PHP 应用例子 …… (460)
11.6.3 PHP 连接数据库的方法 …… (462)
小结 …… (466)
习题 11 …… (466)

* 第 12 章 XML 技术 …… (468)
12.1 XML 概述 …… (468)
12.1.1 从 SGML、HTML 到 XML …… (468)
12.1.2 XML、DTD 与 XML Schema …… (469)
12.1.3 XML 数据库 …… (471)
12.2 常用的 XML 查询语言:XQuery …… (472)

12.2.1　XQuery 的基本功能 …… (472)
12.2.2　基本概念 …… (474)
12.2.3　简单查询 …… (475)
12.2.4　各种类型的查询 …… (478)
12.2.5　复杂查询 …… (484)
12.3　基于关系数据库的 XML 处理 …… (489)
12.3.1　关系数据的 XML 发布 …… (489)
12.3.2　XML 存储和查询系统的体系结构 …… (491)
12.3.3　XML 数据到关系数据库的存储映射 …… (492)
12.3.4　产生映射方法的步骤 …… (494)
12.3.5　将 XML 文档中的数据装入关系表 …… (500)
12.3.6　基于关系数据库的 XML 查询 …… (501)
小结 …… (503)
习题 12 …… (504)

参考文献 …… (506)

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逐渐扩展到企业、行政等社会各领域,数据处理已成为计算机的主要应用。在 60 年代末,数据库技术就是作为数据处理中的一门新技术发展起来的。它是计算机软件领域的一个重要分支,经过 30 多年的发展,形成了较为完整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,然后介绍数据库中的基本概念,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,数据库技术作为软件学科的一个分支应运而生了。

现在,“数据库”这个名词已是家喻户晓。据查,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。这三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,随着计算机硬件性能的改善,关系系统由于使用简便,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网络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,他在 1973 年荣获美国计算机学会(ACM)图灵奖。E.F.Codd 在关系模型上作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 1998 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、网络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,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。

本书第 2～7 章介绍传统的数据库技术,内容包括关系运算理论、SQL 语言、规范化设计理论、ER 模型、存储结构和系统实现技术。第 8 章介绍对象数据库技术的两种形式:对象关系数据库和面向对象数据库。第 9 章介绍分布式数据库系统。第 10 章介绍中间件技术。第 11、12 章介绍数据库与 WWW 及 XML 技术。

1.1.1　数据和信息

数据,是数据库系统研究和处理的对象。数据与信息十分不开的,它们既有联系又有区别。

信息是许多学科广泛使用的概念。在不同的领域中,其含义有所不同。但人们普遍认为信息与能源、材料科学并列,构成现代社会的三大支柱。

在数据处理领域,一般把信息理解为关于现实世界事物存在方式或运动状态

的反映。例如,我们上课用的黑板,颜色是黑的,形状是矩形,尺寸是长 3 m、高 1.2 m,材料是木材,这些都是关于黑板的信息,是黑板存在状态的反映。

信息有许多重要的特征:信息来源于物质和能量;信息是可以感知的;信息是可以存储的;信息是可以加工、传递和再生的。这些特点,构成了信息的最重要的自然属性。作为信息的社会属性,信息已经成为社会上各行各业不可缺少的重要资源之一。人类获取、积累并利用信息是认识和改造客观世界的必要过程。借助信息,人类才能获得知识,才能有效地组织社会各种活动。因此,信息是人类维持正常活动不可缺少的资源。

几乎和信息同样广泛使用的另一个概念是"数据"。所谓数据,通常指用符号记录下来的、可以识别的信息。例如描述黑板的信息,用一组数据"黑色、矩形、3 m、1.2 m"表示。由于这些符号已被人们赋予了特定的语义,因此它们就具有传递信息的功能。

可以看出,信息与数据之间存在着固有的联系:数据是信息的符号表示或称为载体;信息则是数据的内涵,是对数据语义的解释。数据表示了信息,而信息只有通过数据形式表示出来才能被人们理解和接受。尽管两者在概念上不尽相同,但通常人们并不严格去区分它们。

1.1.2 数据处理与数据管理

数据处理也可称为信息处理,因为当把客观事物表示成数据后,这些数据便被人们赋予了特定的含义,从而为人们提供了不必直接观察和度量事物就可以获得有关信息的手段。数据处理是指从某些已知的数据出发,推导加工出一些新的数据,这些新的数据又表示了新的信息。在具体操作中,涉及到数据库收集、管理、加工利用乃至信息输出的演变与推导全过程。

在数据处理中,通常计算比较简单,而数据的管理比较复杂。数据管理是指数据的收集、整理、组织、存储、维护、检索、传送等操作,这部分操作是数据处理业务的基本环节,而且是任何数据处理业务中必不可少的共有部分。对数据管理部分,应研制出一个通用、高效而又使用方便的管理软件,把数据有效地管理起来,以便最大限度地减轻程序员的负担;至于处理业务中的加工计算,因不同业务各不相同,要靠程序员根据业务情况编写应用程序加以解决。所以,数据处理是与数据管理相联系的,数据管理技术的优劣,将直接影响数据处理的效率。数据库技术正是瞄准这一目标研究、发展并完善起来的专门技术。

1.1.3　数据库技术的基本术语

在数据库应用中,常用到 DB、DBMS、DBS 等术语,形式定义如下。

定义 1.1　数据库(DataBase,简记为 DB)

DB 是长期存储在计算机内、有组织的、统一管理的相关数据的集合。DB 能为各种用户共享,具有较小冗余度、数据间联系紧密而又有较高的数据独立性等特点。

定义 1.2　数据库管理系统(DataBase Management System,简记为 DBMS)

DBMS 是位于用户与操作系统(OS)之间的一层数据管理软件,如图 1.1 所示,它为用户或应用程序提供访问 DB 的方法,包括 DB 的建立、查询、更新及各种数据控制。

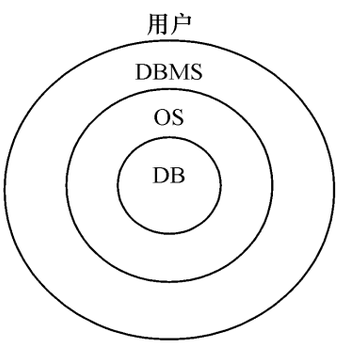


图 1.1　系统层次图

DBMS 总是基于某种数据模型,可以分为层次型、网状型、关系型和面向对象型等。

定义 1.3　数据库系统(DataBase System,简记为 DBS)

DBS 是实现有组织地、动态地存储大量关联数据、方便多用户访问的计算机硬件、软件和数据资源组成的系统,即它是采用数据库技术的计算机系统。

定义 1.4　数据库技术

数据库技术是研究数据库的结构、存储、设计、管理和使用的一门软件课程。

数据库技术是在操作系统的文件系统基础上发展起来的,而且 DBMS 本身要在操作系统支持下才能工作。数据库与数据结构之间的联系也很密切,数据库技术不仅要用到数据结构中链表、树、图等知识,而且还丰富了数据结构的内容。应用程序是使用数据库系统最基本的方式,因为系统中大量的应用程序都是用高级语言(例如 COBOL、C 等)加上数据库的操纵语言联合编制的。集合论、数理逻辑是关系数据库的理论基础,很多概念、术语、思想都直接用到关系数据库中。因此,数据库技术是一门综合性较强的课程。

1.2　数据库的由来和发展

计算机的数据处理应用,首先要把大量的信息以数据形式存放在存储器中。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 年生产的第一台

需要全本请访问 www.ertongbook.com