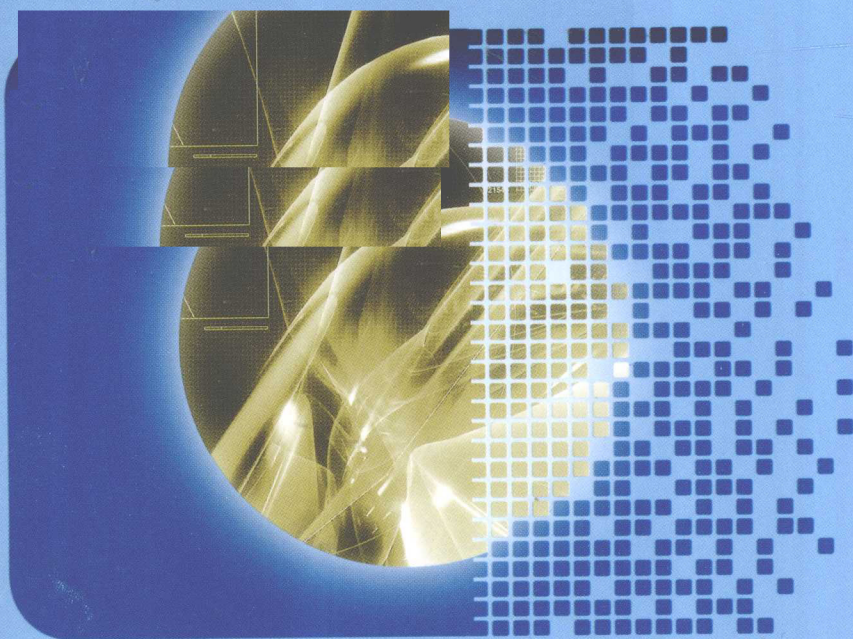




普通高等教育“十一五”国家级规划教材

数据库原理及其 应用教程 (第三版)

主 编 黄德才
副主编 许 芸 王文娟



科学出版社

www.sciencep.com

普通高等教育“十一五”国家级规划教材

数据库原理及其应用教程

(第三版)

主 编 黄德才

副主编 许 芸 王文娟

科 学 出 版 社

北 京

内 容 简 介

本书详细介绍了数据库原理、方法及其应用开发技术。全书共分8章,分别介绍了关系数据库标准语言——SQL,数据模型,数据库系统的模式结构,关系模型与关系代数运算,关系模式的规范化设计理论,数据库的安全与保护,RAID技术,数据库设计与实施,数据库应用系统的体系结构,Microsoft SQL Server 2008的安全管理、完整性策略、恢复技术、并发控制方法、触发器和存储过程,一个基于.NET框架的Microsoft Visual Studio 2008, C#/SQL Server 2008的C/S结构的简单数据库应用系统开发过程等内容。书中还介绍了数据库技术的新发展,如面向对象数据库技术、分布式数据库、XML数据库、数据仓库与数据挖掘技术等。

本书配有较多的例题和适量的习题,随书光盘包括书中所有SQL例题和分别用C#.NET(SQL Server 2008)、Delphi和Visual Basic(SQL Server 2000/2008)开发的数据库应用系统实例源程序和安装程序,并赠送PPT课件一套,不仅有利于教师进行多媒体教学,还便于学生自学。

本书既可作为普通高等院校相关专业“数据库原理及其应用”课程的教材,也可作为成人教育和自学考试同名课程的教材和参考书,还可作为IT领域科技人员的参考书。

图书在版编目(CIP)数据

数据库原理及其应用教程/黄德才主编.—3版.—北京:科学出版社,2010.6

(普通高等教育“十一五”国家级规划教材)

ISBN 978-7-03-027587-5

I. ①数… II. ①黄… III. ①数据库系统—高等学校—教材
IV. ①TP311.13

中国版本图书馆CIP数据核字(2010)第088716号

责任编辑:匡敏/责任校对:刘小梅
责任印制:张克忠/封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

丽泽印刷厂印刷

科学出版社发行 各地新华书店经销

*

2002年8月第一版 开本:787×1092 1/16

2006年8月第二版 印张:20 3/4

2010年6月第三版 字数:492 000

2010年6月第十次印刷 印数:26 001—31 000

定价:36.00元(含光盘)

(如有印装质量问题,我社负责调换)

第三版前言

《数据库原理及其应用教程》(第二版)出版4年来,已被全国几十所高校选作“数据库原理及其应用”课程或同类课程的教材,并被评选为“普通高等教育‘十一五’国家级规划教材”。同时,广大读者对本书第二版提出了许多宝贵的意见和建议,为本书的改版工作做出了很大的贡献,在此谨向所有读者表示衷心的感谢。第三版除继承和发扬第二版的优点外,还着重考虑了以下两个现实问题:

(1) 知识更新问题。最近几年,数据库技术的发展十分迅速,应用领域更加广泛,不仅体现在 SQL 国际标准的修订和发布周期越来越短,而且体现在各大数据库厂商新版 DBMS 产品推出的速度越来越快,功能越来越丰富,更体现在业界对 XML 技术的普遍支持。因此,教材不仅要融入数据库最新技术发展和应用问题的讨论,还必须建立新的实验平台,包括更新 DBMS 实验环境以及应用系统的开发工具等。

(2) 人才需求问题。社会对计算机人才的需求发生了巨大的变化,以培养计算机精英人才为目标的学术研究型传统教育模式已不能完全满足社会 and 行业对计算机应用型人才的需求,计算机教育的定位由此发生了巨大变化,出现了以应用为基础的计算机应用型教育。基于此,中国计算机学会教育专业委员会于 2009 年 11 月在北京成功举办了全国首届计算机应用型人才培养论坛。

本书就是在这种时代背景和社会需求条件下,在浙江省精品课程建设项目支持和近几年教学改革与实践基础上进行修订和编写的。

本书以计算机应用型人才培养为目标,以项目需求为导向,核心实例贯穿主要知识点,并按照“实践-理论-再实践-再理论”的顺序精心组织教学内容和知识模块,即以数据应用系统操作演示,数据库标准语言 SQL 及其在 SQL Server 2008 中的应用开始,导出数据模型、关系代数、关系演算、关系模式规范化设计理论、数据库安全与保护、数据库设计与实施等数据库系统原理知识,再介绍数据库连接技术及用 C#. NET 开发工具完成的数据库应用系统实例,最后介绍数据库的新理论和新进展,符合学生学习新知识的思维习惯,可激发学生学习的趣味性和主动性,整体上提高教学效率和教学质量。

本书把第 1 章和第 7 章的所有内容和例题提升到 SQL Server 2008 平台,第 7 章增加了数据库连接技术,其应用系统实例改用 Microsoft Visual Studio 2008 平台的 C#. NET 开发语言来实现,同时还对第 2、3、4、5、8 章的内容进行了相应的补充和修改,并对有些章节的内容进行了全部重写,特别是在第 8 章补充了 XML 数据管理等数据库新技术知识。

本书配有教学光盘,其中有用课堂教学的新版 PPT 课件以及书中的 SQL 例题,还有用 C#. NET、Delphi 和 Visual Basic 开发的数据库应用系统实例源程序和安装程序,后面两个实例还配有详细的开发指南,既可以作为数据库系统实践教学的开发实例,也可以作为学生课余学习的案例。读者可以通过安装和使用这个应用系统来理解数据库管理系统与数据库应用系统之间的区别和联系,还可以在没有开发经验的情况下按照开发指南的步骤,自己开发并运行这两个应用系统实例(3~4 小时),使理论与实践相结合。

此外,本书还配套出版了《数据库原理及其应用教程——学习指导、例题分析、习题解答与标准试题库》学习指导书,若将其与本书配合使用,不仅可以帮助读者全面巩固所学知识,还有助于读者提高知识的灵活应用能力。

本书主要内容由黄德才、许芸、王文娟编写。黄德才任主编,负责草拟编写提纲,规划各章节基本内容,并完成第2、4、7、8章的编写和全书统稿工作;第1章和第3章主要由许芸编写;第5、6章和第1章第7节由王文娟编写。

本书编写工作得到浙江工业大学计算机科学与技术学院数据库教学团队同事和科学出版社编辑的大力支持,我的研究生王蕴同学完成了第1章所有实例在SQL Server 2008环境下的验证工作,吴灵芝同学用C#.NET语言开发完成了第7章的数据库应用系统实例,张毅刚同学完成了第8章第4节XML数据管理操作实例,在此一并表示衷心的感谢!本书有些章节还参考了参考文献所列国内外著作中的部分内容,谨此向各位作者致以衷心的感谢和深深的敬意!

限于作者水平,加之数据库技术的发展非常迅速,疏漏和不当之处在所难免,殷切希望广大师生和读者批评指正。作者的电子邮箱是hdc@zjut.edu.cn。

黄德才

2010年3月于杭州

第二版前言

数据库技术是计算机科学与技术中发展最快的领域之一,也是应用最为广泛的技术之一,它已成为信息时代的核心技术和重要基础。在《数据库原理及其应用教程》第一版于2002年8月出版发行至今的3年多时间里,一方面是数据库技术得到进一步发展,数据库原理及其应用技术方面的知识更加丰富;另一方面是各高校“数据库原理及其应用”课程的教学时数从前些年的70多学时逐年下降,有的学校甚至将该课程的教学时数降到48学时(含上机练习的学时)。这就为我们提出了如何解决教学内容增加,而教学学时下降的问题。本书第二版就是在这种教学需求背景下,在近几年教学实践基础上进行修订和编写的。本书首次在指导思想和体系结构上突破了传统教材从“理论—实践—再理论”的教学内容组织方法。

本书主要对第一版中第1、2、3章的全部内容进行重新组织和修改补充;对第7章增加了SQL Server常用工具和基于VB的数据库应用系统开发实例;此外,对第4章、第5章和第8章也进行适当修改,并补充了部分数据库技术的新知识,对一些限于篇幅而没能写进教材的数据库技术新进展,也指出了明确的参考文献,便于读者自学和扩大知识面。

全书从数据库标准语言SQL及其在SQL Server中的使用开始,导出数据模型、关系代数、关系演算、关系模式规范化设计理论、数据库安全与保护、数据库设计与实施等数据库系统原理与知识,再介绍数据库应用系统开发实例,最后介绍数据库的新理论和新发展,使整个教学过程贯穿“实践—理论—再实践—再理论”的逐渐发展过程,便于学生理解和掌握数据库的原理及其应用技术,使教学获得事半功倍的效果。

本书配有教学资料光盘,其中不仅有利于课堂教学的PPT教学课件以及教材中的SQL例题,还有用Delphi和Visual Basic开发的数据库应用系统实例源程序和安装程序。读者不仅可以通过安装和使用这个应用系统来理解数据库管理系统与数据库应用系统之间的区别和联系,还可以在没有开发经验的情况下按照开发步骤说明,自己开发完成并运行这个应用系统实例(3~4小时),真正帮助读者架起理论与实践的桥梁,这不仅能帮助读者切实理解客户端与服务器端的概念及其相互联系,增强实践能力,激发学习兴趣,更有利于提高教学质量和教学效果。

此外,本书还有配套的《数据库原理及其应用教程——学习指导、例题分析、习题解答与标准试题库》学习指导书。该书每章都包括“学习要求与重点”、“主要知识点”、“典型例题分析”、“教程习题答案”、“标准试题库(填空题、选择题、简答题和综合题)”和“标准试题库参考答案”6部分,最后的附录还给出一套模拟试题及参考答案。读者若将教材和指导书配合使用,不仅可以全面巩固所学知识,还有助于提高知识的灵活应用能力。

本书由黄德才任主编并完成全书的修订工作,许芸、王文娟和沈良忠参加了有关章节的校对工作。本书的编写得到了编者同事和科学出版社的大力支持,编者的研究生张良燕完成了用Visual Basic开发的数据库应用系统实例,王玲利和钟艳平还参加了第1章、第8章中部分内容的修改和增补工作,在此一并表示衷心的感谢!

限于编者水平所限,加之数据库技术的发展非常迅速,书中疏漏和不当之处在所难免,殷切希望广大师生和读者批评指正,以期在下次修订时予以纠正。

作者的电子邮件地址是 hdc@zjut.edu.cn。

黄德才

2006年2月于杭州

第一版前言

数据库是计算机科学中十分重要且发展迅速的重要分支, DBMS(数据库管理系统)同操作系统一样,已成为各种应用系统开发的基础平台。数据库原理及其应用课程的重要性越来越突出,它不仅是普通高校计算机、信息管理等专业的主干课程,而且已成为许多非计算机专业学生的必修或选修课程;它还是政府上网工程中人员培训的必修科目。

对于绝大多数学生来说,学习数据库课程的主要目的不是去研究和开发商品化的 DBMS,而是为了应用现有的 DBMS 和数据库应用系统开发工具,解决实际工作中的各类计算机应用问题。因此,高校计算机或信息管理类专业的学生,必须兼顾数据库理论与实际应用知识的学习,不仅要掌握数据库的基本原理,还必须懂得如何将具体的 DBMS 与某种客户端应用开发工具结合,完成基于 C/S 或 B/S 结构的数据库应用系统的开发过程,才能真正理解和掌握数据库原理及应用的完整概念和知识,为开发数据库应用系统奠定坚实的基础。

然而,国内目前出版的各种计算机专业和信息管理专业的数据库教材,在内容安排和组织上都存在一些不尽如人意的地方。这些教材基本上可以分为两类,一种是“数据库系统原理”类的教材,其特点是注重理论,主要内容放在数据库系统原理和数据库设计原理上,内容抽象且脱离实际,缺少对数据库应用系统开发方面的知识,没能为学生架起从理论到实践的桥梁,使得学生在学习数据库理论知识后却不知道如何应用。此外,这类教材中的不少书籍在介绍标准 SQL 语言时给出的例子,有些根本无法在当今最流行的 SQL Server 系统中运行(全国多数高校的计算机实验环境和学生自己的计算机环境都使用 SQL Server),使学习内容与实践环境严重脱节,给学生学习增加了困难。另一种是“数据库原理及其应用”类的教材,虽然增加了一些应用知识,但仅停留在数据库应用系统开发环境和一般开发技术的简单介绍层面上,这样反而给学生理解理论与实践之间的联系设置了障碍。

本书是在作者多年数据库教学经验和数据库应用系统开发实践的基础上编写的一本体现当前数据库理论发展和应用技术水平,符合 21 世纪 IT 发展需要的新教材。本书内容丰富,层次分明,概念清楚,逻辑性强,理论叙述深入浅出,实际应用完整具体,本书不仅配有较多的例子和适量的习题,而且还有包括书中标准 SQL 例题和实验数据库应用系统的配套光盘。这不仅有利于教师进行多媒体教学,也将为学生自学提供极大的帮助。本书不仅介绍了数据库系统基本原理和标准 SQL(全部例子都在 SQL Server 2000 上运行通过),且首次介绍了将 Delphi 与 SQL Server 2000 结合开发的一个基于 C/S 结构的简单而完整的数据库应用系统,学生不需要系统开发经验即可在 3~4 小时内建立完成并运行该系统,这真正为学生架起了理论与实践的桥梁。实践表明,这不仅能帮助学生真正理解客户端与服务器端的概念及其相互关系,增强学生的实践能力,还有利于提高教学质量和教学效果,培养学生的系统开发能力和创新能力。

全书共分 8 章,第 1 章绪论,主要内容有数据库的常用术语、计算机数据管理技术的

产生和发展、数据模型、数据库系统的模式结构、DBMS 的功能等;第 2 章关系数据库,主要介绍关系模型的基本概念、关系代数运算、关系演算及其三种关系查询语言——ISBL, ALPHA, QBE 等;第 3 章关系数据库标准语言——SQL,主要有 SQL 的特点、数据定义、数据查询、数据更新、视图定义、数据控制、嵌入式 SQL 以及查询优化等有关知识;第 4 章关系模式的规范化设计理论,内容有关系模式可能存在的异常、关系模式的函数依赖、关系模式的范式(1NF~5NF)、关系模式的分解特性等;第 5 章数据库的安全与保护,包括安全性、完整性、并发控制、事务处理与故障恢复技术、RAID 技术介绍等;第 6 章数据库设计,介绍数据库设计概念、数据库设计的步骤(数据库规划、需求分析、概念结构设计、逻辑结构设计、物理结构设计)和数据库实施等内容;第 7 章数据库应用系统开发,内容包括数据库应用系统的体系结构,Microsoft SQL Server 2000 的主要特点、安装、基本工具、安全管理、完整性策略、恢复技术和并发控制及其触发器和存储过程,一个基于 Delphi 6.0/Server 2000 的 C/S 结构的简单数据库应用系统;第 8 章数据库技术新发展,其内容有数据库家族概述、面向对象数据库技术、分布式数据库、数据仓库与数据挖掘等。讲授完本书全部内容大约需要 64 学时,对于专科生或教学学时数较少的本科生可以不讲授标有“*”的内容。书中标有“*”的内容主要安排给学生课后自学,如果时间充裕,也可选择其中部分内容用于课堂讲授。

本书可作为普通高校计算机专业、信息管理与信息系统专业和信息计算与数学专业学生“数据库原理及其应用”专业课程的教材,也可作为成人教育和自学考试同名课程的教材和参考书,还可作为 IT 领域科技人员的参考书。

本书主要由黄德才、许芸、王文娟编写,黄德才任主编并草拟提纲、规划各章节基本内容。第 2、4、7、8 章由黄德才负责编写,第 1、3 章由许芸编写,第 5、6 章由王文娟编写。黄德才完成全书的修改和统稿工作,并对第 1、3、5、6 章进行了补充。本书的编写得到了编者同事和科学出版社的大力支持,编者的研究生郭海东、沈良忠还参加了第 7 章内容的编写工作,在此表示衷心的感谢!

限于编者水平所限,加之数据库技术的发展非常迅速,书中疏漏和不当之处在所难免,真切希望广大师生和读者来信批评指正,以期在修订时予以完善。

作者的电子邮件地址是 huangdc@mail.hz.zj.cn。

黄德才
2002 年 7 月于杭州

目 录

第三版前言

第二版前言

第一版前言

第 1 章 关系数据库标准语言——SQL	(1)
1.1 SQL 概述	(1)
1.1.1 SQL 的发展	(1)
1.1.2 SQL 的特点	(2)
1.1.3 基本概念	(3)
1.2 SQL 的数据定义	(4)
1.2.1 数据库的建立与删除	(5)
1.2.2 基本表的定义、删除与修改	(7)
1.2.3 建立与删除索引	(14)
1.3 SQL 的数据查询	(15)
1.3.1 简单的选择与投影查询	(15)
1.3.2 连接查询	(21)
1.3.3 嵌套查询	(25)
1.3.4 集合查询	(29)
1.4 SQL 的数据更新	(30)
1.4.1 插入数据	(30)
1.4.2 修改数据	(30)
1.4.3 删除数据	(31)
1.5 SQL 的视图	(32)
1.5.1 定义视图	(33)
1.5.2 查询视图	(34)
1.5.3 更新视图	(35)
1.6 SQL 的数据控制	(35)
1.6.1 创建用户	(36)
1.6.2 授权	(39)
1.6.3 收回权限	(42)
1.7 嵌入式 SQL	(42)
1.7.1 嵌入式 SQL 的标志	(43)
1.7.2 嵌入式 SQL 语句与主语言之间的通信	(44)
1.7.3 PowerBuilder 中的嵌入式 SQL 语句	(46)
1.7.4 在 C#中使用 SQL 语句	(49)

习题	(51)
第2章 数据管理与数据库	(53)
2.1 数据库的常用术语	(53)
2.1.1 数据与信息	(53)
2.1.2 数据独立性	(55)
2.1.3 数据库系统	(55)
2.2 计算机数据管理技术的产生和发展	(56)
2.2.1 人工管理阶段	(56)
2.2.2 文件系统阶段	(57)
2.2.3 数据库系统阶段	(58)
2.2.4 数据库技术的发展	(61)
2.2.5 数据库技术的主要研究领域	(62)
2.3 数据模型	(63)
2.3.1 数据模型的构成	(63)
2.3.2 数据模型的分类	(64)
2.3.3 实体-联系(E-R)模型	(66)
2.3.4 常用的结构数据模型	(71)
2.4 数据库系统的模式结构	(77)
2.4.1 数据库系统模式的概念	(77)
2.4.2 数据库的三级模式	(78)
2.4.3 数据库的二级映像功能与数据独立性	(79)
2.5 DBMS 的功能	(81)
2.6 数据库系统的组成	(82)
习题	(84)
第3章 关系数据库模型	(86)
3.1 关系模型的基本概念	(86)
3.1.1 关系模型概述	(86)
3.1.2 关系数据结构	(87)
3.1.3 完整性规则	(90)
3.2 关系代数	(91)
3.2.1 传统的集合运算	(92)
3.2.2 专门的关系运算	(93)
3.2.3 关系运算举例	(96)
3.3 关系演算	(97)
3.3.1 元组关系演算	(98)
3.3.2 域关系演算	(99)
3.3.3 关系运算的安全限制	(100)
3.4 查询优化	(101)
3.4.1 查询实例分析	(101)

3.4.2	查询处理与优化技术	(103)
3.4.3	关系代数等价公式	(105)
3.4.4	查询优化策略	(107)
3.4.5	查询优化计算步骤	(108)
习题		(109)
第4章	关系模式的规范化设计理论	(111)
4.1	问题的提出	(111)
4.1.1	关系模式可能存在的异常	(111)
4.1.2	异常原因的分析	(112)
4.1.3	异常问题的解决	(113)
4.2	关系模式的函数依赖	(114)
4.2.1	再论关系与关系模式	(114)
4.2.2	函数依赖的一般概念	(114)
4.2.3	候选键与主键	(116)
4.2.4	函数依赖的推理规则	(117)
4.3	关系模式的规范化	(125)
4.3.1	范式及其类型	(126)
4.3.2	第一范式(1NF)	(126)
4.3.3	第二范式(2NF)	(127)
4.3.4	第三范式(3NF)	(128)
4.3.5	BC范式(BCNF)	(129)
4.3.6	多值依赖	(131)
4.3.7	第四范式(4NF)	(134)
4.3.8	关系模式规范化步骤	(135)
4.4	关系模式的分解特性	(136)
4.4.1	模式分解中存在的问题	(136)
4.4.2	无损连接	(137)
4.4.3	无损连接的测试	(138)
4.4.4	保持函数依赖的分解	(141)
4.4.5	分解成3NF的模式集	(143)
4.4.6	关系模式设计原则	(143)
习题		(144)
第5章	数据库的安全与保护	(146)
5.1	安全与保护概述	(146)
5.2	数据库的安全性保护	(147)
5.2.1	用户鉴别	(147)
5.2.2	存取权限控制	(148)
5.2.3	视图机制	(149)
5.2.4	跟踪审查	(149)

5.2.5 数据加密存储	(150)
5.3 数据库的完整性保护	(150)
5.3.1 完整性约束的分类	(150)
5.3.2 完整性控制	(152)
5.3.3 触发器	(155)
5.4 数据库的并发控制技术	(156)
5.4.1 事务及特性	(156)
5.4.2 数据库的并发控制	(157)
5.4.3 并发的目的	(158)
5.4.4 并发所引起的问题	(158)
5.4.5 并发控制方法	(160)
5.4.6 并发调度的可串行性	(163)
5.4.7 时标技术	(166)
5.5 数据库的恢复技术	(167)
5.5.1 故障的种类	(167)
5.5.2 恢复技术	(169)
5.5.3 检查点机制	(172)
5.6 数据库的复制与相关技术	(173)
5.6.1 数据库的复制	(173)
5.6.2 数据库的镜像	(174)
5.6.3 RAID 技术简介	(174)
习题	(176)
第6章 数据库设计与实施	(177)
6.1 数据库设计概述	(177)
6.1.1 数据库设计的概念	(177)
6.1.2 数据库设计的步骤	(178)
6.1.3 数据库设计的特点	(179)
6.2 数据库规划	(180)
6.3 需求分析	(180)
6.3.1 需求分析的任务	(181)
6.3.2 需求分析的步骤	(181)
6.3.3 数据字典	(184)
6.4 概念结构设计	(186)
6.4.1 基本方法	(186)
6.4.2 主要设计步骤	(187)
6.5 逻辑结构设计	(191)
6.5.1 E-R 模式到关系模式的转换	(191)
6.5.2 关系模式的优化	(193)
6.6 物理结构设计	(195)

6.6.1 聚簇设计	(196)
6.6.2 索引设计	(196)
6.6.3 分区设计	(196)
6.7 数据库的实施和维护	(197)
6.7.1 数据库的建立与调整	(197)
6.7.2 数据库系统的试运行	(198)
6.7.3 数据库系统的运行和维护	(198)
习题	(199)
第7章 数据库应用系统开发	(201)
7.1 数据库应用系统的结构	(201)
7.1.1 单用户结构	(202)
7.1.2 集中式结构	(202)
7.1.3 二层客户机/服务器结构	(203)
7.1.4 三层客户机/服务器结构	(204)
7.1.5 多层结构	(205)
7.2 Microsoft SQL Server 2008	(207)
7.2.1 SQL Server 发展简史	(207)
7.2.2 SQL Server 2008 的版本	(209)
7.2.3 SQL Server 2008 的基本操作	(210)
7.2.4 SQL Server 2008 的安全性管理	(214)
7.2.5 SQL Server 的完整性策略	(216)
7.2.6 SQL Server 的恢复技术	(218)
7.2.7 SQL Server 数据库的备份和恢复	(220)
7.2.8 SQL Server 的并发控制	(220)
7.2.9 数据库编程	(223)
7.3 数据库应用系统实例	(225)
7.3.1 常见的数据库访问技术	(226)
7.3.2 Microsoft Visual Studio 2008	(228)
7.3.3 一个数据库应用系统实例	(231)
习题	(259)
第8章 数据库技术新发展	(260)
8.1 数据库家族概述	(260)
8.2 面向对象数据库技术	(261)
8.2.1 面向对象方法学简介	(261)
8.2.2 面向对象数据模型	(262)
8.2.3 面向对象数据库管理系统	(266)
8.3 分布式数据库	(273)
8.3.1 分布式数据库系统的产生	(273)
8.3.2 分布式数据库系统的定义	(274)

8.3.3	分布式数据库系统的主要特点	(275)
8.3.4	分布式数据库系统的模式结构	(278)
8.3.5	分布式数据库管理系统及其分类	(279)
8.3.6	分布式数据库系统的优缺点	(281)
8.4	XML 数据库技术	(282)
8.4.1	XML 概述	(282)
8.4.2	XML 基础知识	(287)
8.4.3	XML 数据库	(293)
8.5	数据仓库与数据挖掘	(295)
8.5.1	数据的事务处理与分析处理	(295)
8.5.2	数据仓库的概念	(296)
8.5.3	数据仓库数据的基本特征	(297)
8.5.4	数据仓库系统的组成	(298)
8.5.5	数据仓库数据的组织	(299)
8.5.6	数据仓库工具	(300)
8.5.7	数据挖掘工具	(305)
	习题	(315)
	参考文献	(317)

第 1 章 关系数据库标准语言——SQL

数据库技术是计算机数据管理的最新技术,是计算机科学的重要分支。当计算机的主要应用领域从科学计算转变到数据及事务处理时,数据库技术应运而生并成为计算机科学的重要领域。今天,数据库技术不仅在企业管理信息系统(MIS)、计算机集成制造系统(CIMS)、办公信息系统(OIS)、地理信息系统(GIS)、Internet 技术等许多方面得到广泛应用,且越来越多的新应用领域都采用数据库来存储和处理它们的信息资源。对于一个国家来说,数据库的建设规模、数据库信息量的大小和使用频度、数据库的安全性和可靠性已成为衡量一个国家信息化程度的重要标志。因此,数据库原理及其应用课程成为计算机科学与技术、软件工程、信息管理等专业的重要专业课程之一。

SQL 是 Structured Query Language 的缩写,其读音为[si:kwə]或直接读字母“S-Q-L”,中文翻译为“结构化查询语言”。它是关系数据库中使用的标准数据查询语言,被 IBM 公司最早使用在其开发的数据库系统中,并于 1987 年得到国际标准化组织的支持成为国际标准。目前,SQL 已被广泛地应用在各种商品化关系数据库管理系统(RDBMS)之中,且是在相当一段时间内无法被取代的重要数据查询语言。可以毫不夸张地说,如果不能掌握和熟练使用 SQL 语言,就很难成为数据库方面的顶级高手。

1.1 SQL 概述

1.1.1 SQL 的发展

SQL 语言由 Boyce 和 Chamberlin 于 1974 年提出。在 1975 ~ 1979 年,IBM 公司 San Jose Research Laboratory 研制了著名的关系数据库管理实验系统 System R 并实现了这种语言。SQL 语言功能丰富,语言简洁,所以备受用户及计算机工业界欢迎,被众多计算机公司和软件公司所采用。经各公司的不断修改、扩充和完善,SQL 语言最终发展成为关系数据库的标准语言。

1986 年 10 月美国国家标准局(American National Standard Institute, ANSI)的数据库委员会 X3H2 批准了 SQL 作为关系数据库语言的美国标准,同年公布了 SQL 标准文本(简称 SQL-86)。国际标准化组织(International Organization for Standardization, ISO) 1987 年也通过了 ANSI 的 SQL-86 标准,使其成为国际标准。

自 SQL 成为国际标准语言以后,各个数据库厂家纷纷推出支持标准 SQL 的数据库管理系统或接口软件,使不同数据库系统之间的相互操作有了共同的基础,使数据库技术得到革命性的发展。此外,SQL 的影响力已远远超出了数据库领域本身,对数据库以外的领域也产生了很大影响,有不少软件产品将 SQL 语言的数据查询功能与图形功能、软件工程工具、软件开发工具和人工智能程序结合起来。因此,SQL 已成为数据库领域中使用最为广泛的主流语言之一。

自 1987 年开始,ISO 与 ANSI 积极合作并不断地修改和完善 SQL 标准,于 1989 年公

布了 SQL—89 标准,1992 年公布了 SQL—92 标准,也称 SQL2;1999 年公布了标准 SQL—99,也称 SQL3,且各大数据库厂商纷纷宣布在自己的新版本中支持这个标准;2003 年 ISO 在以前所有标准的基础上增加了一些新的功能,形成了标准 SQL—2003。该版本对 SQL3 存在的一些问题进行了改进,还支持 XML,支持 Window 函数和 Merge 语句等;2006 年又发布了 SQL—2006,继续增强 XML 方面的特性;2008 年又发布了 SQL—2008,据媒体报道,此次标准的制定已经有中国专家参与其中,只不过最近两个版本发布后在业界没有引起多大的反响,也很少看到有该技术领域的人士讨论此事。但从标准 SQL—99 到标准 SQL—2008 的发布可以看出,标准修订的周期越来越短,反映了数据库对技术需求的变化速度。

因此,所有这些标准并不仅仅是对 SQL 语言本身的完善,而是随着 IT 技术的发展和各个领域实际应用的需求,对基于 SQL 语言的整个数据库管理系统应该支持的功能进行适时的补充和完善。例如,随着面向对象技术发展和多媒体数据应用的需求,SQL 标准增加了基本对象支持、增强对象支持和多媒体支持等功能;为了实现不同应用程序系统之间交换数据和应用程序跨平台性,SQL 标准增加了对 XML 和 Java 的支持;为了帮助企业高层分析决策,增加了数据仓库和联机分析处理(OLAP)的支持功能等。目前,世界上几乎所有著名的数据库管理系统,如 SQL Server、Oracle、Sybase 和 DB2 等,都支持并实现了 SQL 语言,而且还有自己独特的改进和扩展。读者若想了解更多 SQL 标准的相关知识,可阅读本书后的参考文献。

1.1.2 SQL 的特点

SQL 语言之所以能够为用户和业界所接受,并成为国际标准,是因为它是一种综合的、功能极强同时又简便易学的语言。SQL 语言集数据查询(data query)、数据操纵(data manipulation)、数据定义(data definition)和数据控制(data control)功能于一体,主要特点包括:

1) 综合统一。数据库系统的主要功能是通过数据库支持的数据语言来实现的,SQL 语言的综合统一性体现在语言风格统一和操作过程统一两方面。

① 语言风格统一。SQL 语言将数据定义语言 DDL、数据操纵语言 DML、数据控制语言 DCL 的功能集于一体,语言风格统一,可以独立完成数据库生命周期中的全部活动,包括定义关系模式、插入数据、建立数据库、查询、更新、维护、数据库重构、数据库安全性控制等一系列操作要求,这就为数据库应用系统的开发提供了良好的环境。用户在数据库系统投入运行后,还可根据需要随时地、逐步地修改模式,且并不影响数据库的运行,从而使系统具有良好的可扩展性。

② 操作过程统一。在关系模型中实体和实体间的联系均用关系表示,这种单一的数据结构使数据的查找、插入、删除、修改等每一种操作都只需一种操作符,从而克服了非关系系统由于信息表示方式的多样性带来的操作复杂性。例如,在 DBTG 中,需要两种插入操作符:STORE 用来把记录存入数据库,CONNECT 用来把记录插入系值以建立数据之间的联系。

而非关系模型(层次模型、网状模型)的语言一般都分为模式数据定义语言(schema data definition language,模式 DDL)、外模式数据定义语言(subschema data definition lan-