

可下载教学资料

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



高等学校教材
计算机科学与技术

数据库技术

——设计与应用实例

岳昆 编著 刘惟一 审校



清华大学出版社

高等学校教材
计算机科学与技术

数据库技术 ——设计与应用实例

岳昆 编著 刘惟一 审校

清华大学出版社
北京

内 容 简 介

本书深入浅出地讲述了数据库设计与应用的主要方法和实例,以基于 MS SQL Server 2000 数据库管理系统的实例介绍为重点。全书主要包括基础篇、提高篇和新技术篇 3 个部分。基础篇介绍数据库设计的一般方法以及数据库和表的基本操作;提高篇面向复杂应用介绍连接查询和嵌套查询、视图、存储过程和触发器,基于索引机制、基于关系模式反规范化和视图物化以及基于查询语句重写的查询优化,事务的操作和数据库的备份与还原;新技术篇介绍数据的 Web 发布和数据的挖掘与分析。

本书可单独作为数据库设计与应用相关课程的教材,也可作为数据库技术理论课程的配套实验指导书或作为数据库系统开发人员的参考资料。本书所讲述的内容适用于不同类型的读者。对于非计算机专业的数据库基础知识介绍或职业技术短期培训,可以基础篇作为教学的重点,有选择地介绍提高篇中的部分内容;对于计算机科学与技术专业的课程教学,可以基础篇和提高篇作为主要教学内容,并可根据具体情况选择新技术篇中的部分内容进行讲授。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

数据库技术——设计与应用实例/岳昆编著. —北京:清华大学出版社,2007.9

(高等学校教材·计算机科学与技术)

ISBN 978-7-302-15540-9

I. 数… II. 岳… III. 数据库系统 IV. TP311.13

中国版本图书馆 CIP 数据核字(2007)第 096235 号

责任编辑:梁颖 王冰飞

责任校对:时翠兰

责任印制:李红英

出版发行:清华大学出版社 地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn> 邮 编:100084

c-service@tup.tsinghua.edu.cn

社总机:010-62770175 邮购热线:010-62786544

投稿咨询:010-62772015 客户服务:010-62776969

印刷者:北京季蜂印刷有限公司

装订者:三河市漂源装订厂

经 销:全国新华书店

开 本:185×260 印 张:16.5 字 数:391 千字

版 次:2007 年 9 月第 1 版 印 次:2007 年 9 月第 1 次印刷

印 数:1~4000

定 价:23.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:022991-01

编审委员会成员

(按地区排序)

清华大学	周立柱	教授
	覃 征	教授
	王建民	教授
	刘 强	副教授
	冯建华	副教授
北京大学	杨冬青	教授
	陈 钟	教授
	陈立军	副教授
北京航空航天大学	马殿富	教授
	吴超英	副教授
	姚淑珍	教授
中国人民大学	王 珊	教授
	孟小峰	教授
	陈 红	教授
北京师范大学	周明全	教授
北京交通大学	阮秋琦	教授
北京信息工程学院	孟庆昌	教授
北京科技大学	杨炳儒	教授
石油大学	陈 明	教授
天津大学	艾德才	教授
复旦大学	吴立德	教授
	吴百锋	教授
	杨卫东	副教授
华东理工大学	邵志清	教授
华东师范大学	杨宗源	教授
	应吉康	教授
东华大学	乐嘉锦	教授
上海第二工业大学	蒋川群	教授
浙江大学	吴朝晖	教授
	李善平	教授
南京大学	骆 斌	教授
南京航空航天大学	秦小麟	教授
南京理工大学	张功萱	教授

南京邮电学院	朱秀昌	教授
苏州大学	龚声蓉	教授
江苏大学	宋余庆	教授
武汉大学	何炎祥	教授
华中科技大学	刘乐善	教授
中南财经政法大学	刘腾红	教授
华中师范大学	王林平	副教授
	魏开平	副教授
	叶俊民	教授
国防科技大学	赵克佳	教授
	肖 侬	副教授
中南大学	陈松乔	教授
	刘卫国	教授
湖南大学	林亚平	教授
	邹北骥	教授
西安交通大学	沈钧毅	教授
	齐 勇	教授
长安大学	巨永峰	教授
西安石油学院	方 明	教授
西安邮电学院	陈莉君	副教授
哈尔滨工业大学	郭茂祖	教授
吉林大学	徐一平	教授
	毕 强	教授
长春工程学院	沙胜贤	教授
山东大学	孟祥旭	教授
	郝兴伟	教授
山东科技大学	郑永果	教授
中山大学	潘小轰	教授
厦门大学	冯少荣	教授
福州大学	林世平	副教授
云南大学	刘惟一	教授
重庆邮电学院	王国胤	教授
西南交通大学	杨 燕	副教授

改革开放以来,特别是党的十五大以来,我国教育事业取得了举世瞩目的辉煌成就,高等教育实现了历史性的跨越,已由精英教育阶段进入国际公认的大众化教育阶段。在质量不断提高的基础上,高等教育规模取得如此快速的发展,创造了世界教育发展史上的奇迹。当前,教育工作既面临着千载难逢的良好机遇,同时也面临着前所未有的严峻挑战。社会不断增长的高等教育需求同教育供给特别是优质教育供给不足的矛盾,是现阶段教育发展面临的基本矛盾。

教育部一直十分重视高等教育质量工作。2001年8月,教育部下发了《关于加强高等学校本科教学工作,提高教学质量的若干意见》,提出了十二条加强本科教学工作提高教学质量的措施和意见。2003年6月和2004年2月,教育部分别下发了《关于启动高等学校教学质量与教学改革工程精品课程建设工作的通知》和《教育部实施精品课程建设提高高校教学质量和人才培养质量》文件,指出“高等学校教学质量和教学改革工程”是教育部正在制定的《2003—2007年教育振兴行动计划》的重要组成部分,精品课程建设是“质量工程”的重要内容之一。教育部计划用五年时间(2003—2007年)建设1500门国家级精品课程,利用现代化的教育信息技术手段将精品课程的相关内容上网并免费开放,以实现优质教学资源共享,提高高等学校教学质量和人才培养质量。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作,提高教学质量的若干意见》精神,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上;精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展、顺应并符合新世纪教学发展的规律、代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻

性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。首批推出的特色精品教材包括:

(1) 高等学校教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。

(2) 高等学校教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 高等学校教材·电子信息——高等学校电子信息相关专业的教材。

(4) 高等学校教材·软件工程——高等学校软件工程相关专业的教材。

(5) 高等学校教材·信息管理与信息系统。

(6) 高等学校教材·财经管理与计算机应用。

清华大学出版社经过 20 多年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会

E-mail: dingl@tup.tsinghua.edu.cn

近年来数据库技术得到了迅速的发展,并且被广泛应用于现实生活中的多个领域。数据海量存储、数据挖掘与知识发现、Web 数据管理、异构数据集成等都是目前数据库领域理论研究和现实应用的热点,越来越多地受到人们的关注。数据库设计与应用是上述问题在理论上和工程化方面的基础,也是信息系统设计及实现的关键。

数据库设计与应用包含面向问题各个层面的一系列方法和技术,其中也体现了多方面的设计理念和策略。学习数据库设计与应用的相关知识,既要重视理论方法的理解,也要注重应用技能的训练;既要理解设计与应用的动机,也要掌握基本概念和常用策略。数据库设计与应用的特点要求理论方法与应用实例的学习紧密结合,学习理论方法后进行相应的上机练习,在真实的数据库环境中动手实践,这是熟练掌握数据库设计与应用方法,并达到学以致用目的的重要环节。理论方法与应用实例的学习相辅相成、互为补充,是学好数据库设计与应用的基本理论、主要方法和常用策略的有效手段。

本书以数据库设计与应用为主线,以具体实例为重点,以通过具体实例体现理论方法为目标,基于 MS SQL Server 2000 数据库管理系统主要介绍数据库设计与应用的实例,涉及从数据库概念建模到具体应用的各个环节,既讲原理,又讲应用。通过对本书内容的学习,使读者在真实的数据库管理系统(DBMS)环境中实施具体的数据库设计与应用,通过典型实例剖析问题实质,总结应用中的注意事项及一般原则,旨在对数据库技术的学习者起到抛砖引玉的作用,为读者进一步学习相关知识提供举一反三的范例,从而为读者提供一个学习数据库系统概论或数据库系统原理课程在实践环节的补充,为读者搭起一座同步学习数据库技术的理论方法与应用实例之间的桥梁。

本书按照由浅入深的方式展开各部分内容的阐述,既注重各知识点的应用背景和基本原理的介绍,也注重各个实现细节的讨论;既基于应用背景讲原理,也基于典型范例教操作;既突出基本概念的阐述,也强调动手能力的训练。在各部分内容的讨论中,以“提示”、“说明”或“注意”的形式总结某一类实例的特征、具体问题的特殊性或同类问题不同方面的比较,解释实例的运行结果、蕴含的基本原理或涉及的背景知识,阐述设计、应用或开发中的常见问题、注意事项或实现技巧等。作为所学数据库设计与应用方法的练习和实践,本书既给出了每章的习题和课件,也给出了模拟实际应用的课程实习题,使得学生以由点到面的学习思路,先逐个掌握各个知识点,而后实现小型信息系统,从而达到融会贯通的学习效果。

本书介绍的内容围绕数据库设计与应用的主要方面,对于没有详细介绍的知识点也给出了学习的思路和文献阅读的提示。书中的内容并不是对 MS SQL Server 2000 所提供的工具和数据管理功能的逐一讲述,因此即使其他 DBMS(如 DB2、Oracle、Sybase 等)的用户,仍然可以较好地学习和理解书中所介绍方法与技术的实质。

本书以基础篇、提高篇和新技术篇为内容组织的框架,具体包括:

- 基础篇:从工程化角度出发的数据库概念建模和逻辑设计方法,是全书内容组织的基础;数据库及表的基本操作,主要包括基本 SQL 数据操纵、数据库的创建、表的创建和完整性约束的实现。

- 提高篇:面向复杂应用的数据操作,主要包括连接查询、嵌套查询的设计与实现,视图、存储过程和触发器的创建及应用方法;数据的查询优化,主要包括基于索引机制、基于关系模式反规范化和视图物化、基于查询语句重写的查询优化方法;数据库的备份与还原,主要包括事务的概念与操作,以及具体的数据库的备份与还原方法。

- 新技术篇:以数据的 Web 发布和数据的挖掘与分析为代表介绍数据库应用的新技术,前者包括数据的 HTML 和 XML 格式发布,后者包括关系数据和多维数据集的决策树和聚集分析技术。

此外,书中给出了一个贯穿全书的实际案例,相关章节讨论了所介绍方法在该案例中的具体应用。

本书的编写工作得到云南省自然科学基金项目(2005F0009Q)和“云南大学中青年骨干教师培养计划”的资助。在编写过程中,云南大学信息学院和清华大学出版社给予了极大的支持,云南大学信息学院计算机科学与工程系刘惟一教授对本书的选材和内容等方面提出了宝贵的意见和建议,对全书作了认真细致的审校。此外,复旦大学信息学院计算机科学与工程系周傲英教授、云南大学信息学院张学杰教授和周永恒教授、云南昆明科瑞迪电网科技开发有限公司王伯韬常务副总经理、云南电信有限公司昆明分公司曹琳工程师也给予了很多有益的帮助,在此一并表示衷心的感谢。由于作者水平有限,错误和疏漏之处难免,恳请各位专家和读者批评指正。

作者

2007 年 5 月

第一部分 基 础 篇

第 1 章 数据库设计	3
1.1 数据库设计过程	3
1.1.1 MIS 基本知识介绍	4
1.1.2 数据库设计及开发的过程	7
1.2 实体联系模型和关系模型	9
1.2.1 实体联系模型	9
1.2.2 关系模型	16
1.2.3 从 ER 模型到关系模型	22
1.2.4 规范化简介	27
1.3 案例研究	29
1.4 数据库产品简介及选择	31
1.4.1 几种常见的数据库系统	31
1.4.2 数据库管理系统的选择	37
习题	39
第 2 章 数据库及表的基本操作	41
2.1 基本 SQL 数据操纵	41
2.1.1 SQL 概述	41
2.1.2 数据的查询	43
2.1.3 数据的更新	49
2.2 数据库的创建	54
2.2.1 角色和用户	55
2.2.2 数据库的创建	61

2.3 表的创建及其完整性约束的实现	65
2.3.1 基本表的创建	65
2.3.2 实体完整性和参照完整性约束的有效性验证及实现	69
2.3.3 属性标识、CHECK 约束及默认值的创建	74
习题	78

第二部分 提高篇

第3章 面向复杂应用的数据操作	83
3.1 连接查询	83
3.1.1 等值连接和自然连接	84
3.1.2 外连接	86
3.2 嵌套查询	90
3.2.1 几种基本的子查询	91
3.2.2 子查询的应用	100
3.3 集合查询	101
3.4 视图的创建及应用	102
3.4.1 视图的创建、查询及更新	103
3.4.2 视图机制下数据操作的执行效率	109
3.5 存储过程的创建及应用	110
3.5.1 存储过程的概念	111
3.5.2 存储过程的创建及执行	112
3.5.3 存储过程的应用	116
3.6 触发器的创建及应用	119
3.6.1 触发器的概念	120
3.6.2 触发器的创建和有效性验证	121
3.7 案例研究	123
习题	125
第4章 数据的查询优化	127
4.1 基于索引机制的查询优化	128
4.1.1 问题的提出	128
4.1.2 索引的概念	129
4.1.3 索引的创建	130
4.1.4 索引对数据查询操作性能的影响	141
4.1.5 维护索引的代价	148
4.1.6 索引选择和应用的一般原则	153

4.2 基于数据库模式反规范化和视图物化的查询优化	156
4.2.1 问题的提出	156
4.2.2 关系模式的反规范化	157
4.2.3 视图的物化	161
4.3 基于查询语句重写的查询优化	162
4.4 案例研究	166
习题	167
第 5 章 数据库的备份与还原	170
5.1 事务管理	170
5.1.1 事务的概念	171
5.1.2 数据库事务操作	173
5.2 数据库的备份与还原	179
5.2.1 基本概念	180
5.2.2 数据库的备份	182
5.2.3 数据库的还原	186
5.2.4 数据库的导出和导入	190
5.3 系统数据库的重建	193
习题	195

第三部分 新技术篇

第 6 章 数据的 Web 发布	199
6.1 数据的 HTML 格式发布	199
6.1.1 相关概念	199
6.1.2 将数据发布为 HTML 文件	200
6.2 数据的 XML 格式发布	206
6.2.1 XML 概述	207
6.2.2 检索和编写 XML 数据	210
6.2.3 将数据发布为 XML 文档	214
习题	218
第 7 章 数据挖掘与数据分析	220
7.1 背景知识	220
7.1.1 数据挖掘与 OLAP 概述	221
7.1.2 决策树和聚类分析简介	224
7.1.3 SQL Server 2000 Analysis Services 简介	225

7.2 多维数据集的基本操作	226
7.2.1 多维数据集的创建	226
7.2.2 多维数据集的钻取	231
7.3 基于决策树模型的挖掘与分析	233
7.3.1 用决策树创建关系挖掘模型	233
7.3.2 用决策树创建 OLAP 挖掘模型	238
7.4 基于聚集模型的挖掘与分析	240
习题	242
附录	243
参考文献	247

基础篇

如何构建最优的数据库模式及其上最有效的数据操作是数据库设计与应用的目标。这一部分中,作为数据库技术的基础,第1章介绍ER模型和关系模型,并从工程化的角度出发,介绍数据的概念建模和逻辑设计的方法;作为数据库的基本操作,第2章主要介绍基本SQL数据操纵、数据库的创建、表的创建和完整性约束的实现。

数据库设计

数据库技术是计算机科学技术中发展最快的领域之一,是一门以数据建模和数据管理为核心、内容丰富的学科,它研究的课题和使用的技术涉及其他广泛的领域,包括程序设计、数据结构、用户界面等。数据库技术也是应用最广泛的技术之一,被人们认为是用于信息资源开发、管理和服务的最有效的手段,从小型的单项事务处理系统到大型复杂的信息系统都采用数据库技术来保持数据的有效性、完整性和共享性。

数据库技术是信息系统开发和建设中的核心技术和重要基础,从数据库发展的早期开始,数据的存储和操作一直是主要的应用焦点。如何构建最优的数据库模式及其上最有效的数据操作,从而建立数据库及其应用系统,使之能有效地存储数据,满足各种用户的应用需求,即数据库设计,这是数据库应用领域的主要研究课题之一。

人们所获得的信息正爆炸性地增长,为了从巨大而复杂的数据集中获得最大效益,用户必须借助工具以简化数据管理工作,数据库管理系统(Database Management System, DBMS)是辅助用户管理和利用大规模数据集的软件。DBMS 管理数据具有数据独立性、有效的数据存取、数据完整性和安全性、并发存储和故障恢复、减少应用程序开发时间等优点。本书以 SQL Server 2000 数据库管理系统作为所介绍方法和实例的平台,通过有代表性的实例,深入介绍如何根据用户需求设计结构良好的数据库,并对数据库中的数据进行有效的处理。

本章主要以工程化的方法从总体上介绍数据建模的方法和数据库设计的过程,是后续章节的基础。本章第 1.1 节简单介绍数据库设计的过程及相应的设计方法。第 1.2 节介绍实体联系模型和关系模型,以及从 ER 模型到关系模型的转换。第 1.3 节给出数据库设计的典型案例。第 1.4 节简单介绍几种主要的数据库产品,并讨论数据库产品的选择。

1.1 数据库设计过程

对于一个数据密集型应用来说,数据库是它的一个核心部分,也是大型软件系统的重要组成部分。数据库把企业应用中涉及到的大量数据按一定的模型组织起来,提供存储、检索和维护数据的功能,使得信息系统可以方便地满足用户对数据存储、检索和维护的应用需求。数据库是信息系统中各个部分紧密结合的关键,数据的存储结构和优化策略直接决定了数据的有效性、完整性,以及查询的效率。因此只有对数据库进行合理的逻辑设计和有效

的物理设计才能开发出完善而且高效的信息系统。

大型数据库的设计和开发是一个涉及多学科的综合性和工程,它需要结合计算机程序设计、软件工程原理、应用领域知识等,因此这里把软件工程的原理和方法应用到数据库的建设中,讨论数据库设计的过程及其中相关的主要方法。

作为以数据库为核心的信息系统中有代表性的一类,也作为本书所介绍内容的一个应用背景,本节首先简单介绍管理信息系统的基本知识,进而介绍数据库设计的过程。

1.1.1 MIS 基本知识介绍

管理信息系统(Management Information Systems, MIS)在现代社会中已深入到各行各业,由于计算机技术的迅速发展和普及, MIS 事实上已成为计算机 MIS。目前,由于开发高质量 MIS 的能力大大落后于计算机硬件的发展,加上社会对 MIS 发展和完善需求的增加,以及对 MIS 开发过程中出现的错误认识和行为将导致 MIS 开发的失败,这些情况直接影响了以数据库为核心的信息管理及决策支持等应用。因此对 MIS 有关的内容进行深入研究,提高工作效率,提高 MIS 开发成功率已变得十分重要。

1. MIS 的概念及解释

MIS 是一个不断发展的新型学科, MIS 的定义随着计算机技术和通信技术的进步也在不断更新,在现阶段普遍认为 MIS 是由人和计算机设备或其他信息处理手段组成并用于管理信息的系统。

(1) MIS 的对象就是信息。

信息是经过加工的数据,是对决策者有价值的数据。信息的主要特征是来源分散且数量庞大。信息来源于生产第一线,来源于社会环境,来源于市场,来源于行政管理等部门。信息具有时间性,信息的加工具有多种形式。企业从信息管理的角度可划分为物流和信息流。生产过程是一个物流投入产出的不可逆过程;管理过程是信息流的过程,且具有信息反馈的特征。

(2) 系统是由相互联系、相互作用的若干要素按一定法则组成并具有一定功能的整体。

系统有两个以上要素,各要素和整体之间、整体和环境之间存在一定的有机联系。系统由输入、处理、输出、反馈、控制五个基本要素组成。信息系统是对输入的数据进行处理,输出信息的系统。

(3) 管理信息由信息的采集、信息的传递、信息的存储、信息的加工、信息的维护和信息的使用六个方面组成。

任何地方只要有管理就必然有信息,如果形成系统就形成了 MIS。计算机设备使 MIS 更有效,尤其是现代社会中, MIS 已和计算机设备不可分离,因此一般来说 MIS 就是计算机 MIS。MIS 包括计算机、网络通信设备等硬件成分,包括操作系统、应用软件包等软件成分,并随着计算机技术和通信技术的迅速发展还会出现更多的内容,计算机设备并不是 MIS 的必要条件。

2. MIS 的结构、特征及相关学科

MIS 由信源、信宿、信息处理、信息用户和信息管理者五个部分组成。

完善的 MIS 具有以下四个标准:确定的信息需求,信息的可采集与可加工,可以通过