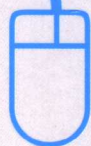


可下载教学资料

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



高等学校教材
计算机科学与技术

数据库系统教程

苑森淼 康 辉 编著



清华大学出版社

高等学校教材
计算机科学与技术

数据库系统教程

苑森森 康 辉 编著

清华大学出版社
北京

内 容 简 介

本书详细地介绍了数据库系统的原理。全书分为基础篇和系统篇两部分。基础篇完整地介绍了数据库建模、关系演算、数据库语言及其查询优化和编程环境；系统篇详细讲解了数据库系统的事务与并发控制、数据库的完整性与安全性、数据库设计的理论基础以及面向对象技术对关系型数据库的影响。

本书可作为高等院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。

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

数据库系统教程/苑森森,康辉编著. —北京:清华大学出版社,2008.1

(高等学校教材·计算机科学与技术)

ISBN 978-7-302-16015-1

I. 数… II. ①苑… ②康… 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(2007)第 133531 号

责任编辑:闫红梅

责任校对:梁 毅

责任印制:何 芊

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:清华大学印刷厂

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:20.75 字 数:502 千字

版 次:2008 年 1 月第 1 版 印 次:2008 年 1 月第 1 次印刷

印 数:1~4000

定 价:29.00 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:014364-01

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

数据库技术自 20 世纪 60 年代中期产生以来,无论是理论还是应用方面都已变得相当的重要和成熟,成为计算机科学的重要分支。数据库技术是计算机领域发展最快的学科之一,也是应用很广、实用性很强的一门技术。随着计算机技术飞速发展及其应用领域的扩大,特别是计算机网络和 Internet 的发展,数据库应用系统得到了突飞猛进的发展。目前,数据库技术已从第一代的层次、网状数据库系统,第二代的关系数据库系统,发展到以面向对象模型为主要特征的第三代数据库系统。

本书是编者结合多年的教学和科研经验,参阅了一些目前国内外比较优秀的同类教材,根据本学科的发展趋势及最新技术成果,注重使读者掌握理论基础,如数据建模、关系演算、数据库语言、并发控制、数据库设计理论等;又强调了数据库的体系结构及其原理,如查询优化、编程环境等;同时还突出了数据库的应用和新技术,如面向对象数据库和对象关系数据库等。

全书由吉林大学计算机科学与技术学院的苑森森和康辉共同编写完成。其中,苑森森编写了第 1~4 章,康辉编写了第 5~10 章。

由于编者水平所限,书中难免有错误与不足之处,恳请读者批评指正。

编 者

2007 年 6 月 27 日于长春

目 录

高等学校教材·计算机科学与技术

第 1 部分 基 础 篇

第 1 章 数据库系统导论	3
1.1 数据库系统	3
1.1.1 数据库系统的产生	3
1.1.2 数据库系统的定义	4
1.1.3 数据库与信息系统	5
1.2 客观世界的描述	7
1.2.1 信息的三个领域	7
1.2.2 如何实现对客观世界的描述	8
1.2.3 实体集之间的联系	9
1.2.4 对客观世界的描述举例	10
1.2.5 关键字	11
1.3 数据的关系模型	12
1.3.1 关系模型的基本概念	13
1.3.2 关系等价性问题	14
1.3.3 关系框架	14
1.3.4 基于集合论的关系定义	15
1.3.5 同类关系	16
1.3.6 关系的广义相等和广义包含	16
1.4 习题	17
第 2 章 数据库的概念设计	19
2.1 数据库设计导论	19
2.1.1 基本设计方法	19
2.1.2 数据库建模	20
2.1.3 数据库概念设计的表示方法	21
2.1.4 数据库概念设计方法	22

2.2	数据库概念设计的实体联系模型	22
2.2.1	E-R 模型的基本结构	22
2.2.2	数据库概念设计中的联系	24
2.2.3	联系的多向性和联系中的角色	24
2.2.4	联系的属性	25
2.2.5	把多向联系转换成二元联系	26
2.3	对象定义语言	27
2.3.1	ODL 中的属性	28
2.3.2	ODL 对联系的描述	30
2.3.3	三类联系方式的描述	31
2.3.4	关键字	32
2.4	关键字、约束和弱实体集	33
2.4.1	主关键字	33
2.4.2	数据库建模中主关键字的表示	34
2.4.3	关于约束	35
2.4.4	弱实体集	38
2.5	从模型设计到关系设计	40
2.5.1	数据库设计的一般方法	40
2.5.2	概念设计的基本方法	41
2.5.3	从 E-R 模型到关系的设计	41
2.5.4	从 ODL 设计到关系设计	44
2.5.5	E-R 和 ODL 模型向关系模型转换比较	51
2.6	数据定义语言	53
2.6.1	SQL 语言	53
2.6.2	SQL 支持的数据类型	54
2.6.3	基表的定义	55
2.6.4	基表模式修改语句	59
2.6.5	索引的建立和撤销	61
2.7	习题	62
第 3 章	关系运算	65
3.1	关系代数	65
3.1.1	基础知识	65
3.1.2	关系的集合运算	67
3.1.3	删除关系中部分数据的运算	69
3.1.4	连接关系的运算	71
3.1.5	关系代数运算表达式	76
3.1.6	改名运算	78
3.1.7	基本关系代数运算	79

3.2	关系代数运算实现方法讨论	80
3.2.1	关系求并、求差和求交运算的实现方法	80
3.2.2	选择和投影运算的实现方法	82
3.2.3	关系的乘积和连接运算的实现	83
3.3	关系演算	83
3.3.1	元组关系演算	83
3.3.2	元组演算和关系代数运算的等价性	85
3.3.3	元组演算举例	86
3.3.4	域关系演算	87
3.3.5	域演算举例	88
3.4	逻辑查询语言	89
3.4.1	Datalog 中的谓词和原子	90
3.4.2	Datalog 规则	91
3.4.3	用 Datalog 规则表示五种基本关系代数运算	92
3.4.4	关于 Datalog 规则的讨论	95
3.4.5	用 Datalog 规则表示的运算举例	97
3.5	Datalog 对递归过程的描述	99
3.5.1	一个递归处理的例子	100
3.5.2	使用 Datalog 描述递归运算	101
3.5.3	Datalog 递归规则中的求反问题	104
3.6	习题	106
第 4 章	数据库语言	110
4.1	SQL 的基本查询结构	110
4.1.1	SQL 的运算符和影射块	111
4.1.2	投影运算	113
4.1.3	选择运算	114
4.1.4	字符串和日期比较	116
4.1.5	字符串模式匹配作为选择条件	117
4.1.6	ORDER BY 子句	118
4.2	连接及其他二目运算的 SQL 表示方法	118
4.2.1	用影射块描述连接运算	119
4.2.2	关于复合运算表示方法的讨论	120
4.2.3	关于多关系查询的实现	121
4.2.4	其他二目运算	124
4.2.5	使用集合运算符构造选择条件表达式	126
4.2.6	影射块嵌套	127
4.2.7	关系的别名	129
4.3	基于包的关系运算和查询	130

4.3.1	为什么使用包运算	130
4.3.2	基于包的求并、求交和求差	131
4.3.3	基于包的其他关系运算	133
4.3.4	SQL 中的一些运算规定	134
4.4	聚合函数和分组	136
4.4.1	聚合函数	136
4.4.2	GROUP BY 子句	137
4.4.3	HAVING 子句	139
4.4.4	关于 SQL 查询语言小结	141
4.5	SQL 的数据操作语句	143
4.5.1	插入元组的语句	143
4.5.2	删除元组的语句	146
4.5.3	修改元组分量值的语句	147
4.6	数据库逻辑结构和视图的定义	147
4.6.1	数据库的逻辑结构	148
4.6.2	视图的定义	149
4.6.3	面向视图的查询处理	152
4.6.4	视图的更新问题	155
4.6.5	从模式到物理存储的映射	157
4.7	递归查询	158
4.7.1	SQL3 中的递归描述	158
4.7.2	WITH 语句中定义视图	160
4.7.3	求反问题	162
4.8	习题	164
第 5 章	查询优化及实现	170
5.1	查询优化导论	170
5.1.1	问题的提出	170
5.1.2	优化的一般方法及策略	170
5.2	代数优化基础	171
5.2.1	关系代数等价变换规则	171
5.2.2	广义包含与广义相等的运算规则	174
5.3	查询处理	175
5.3.1	查询的实现过程概述	175
5.3.2	语法树及其优化算法	177
5.4	主要查询操作实现及优化	178
5.4.1	选择操作实现方法及优化	179
5.4.2	连接操作的实现方法和优化	182
5.4.3	投影运算的实现	188

5.4.4 集合运算的实现	189
5.4.5 组合操作问题	191
5.5 习题	192
第 6 章 数据库系统及 SQL 编程环境	193
6.1 数据库系统的逻辑结构	193
6.1.1 数据库系统的三级模式结构	193
6.1.2 数据库系统的层次转换与映像	195
6.2 DBMS 的系统结构	197
6.2.1 DBMS 与环境的集成	197
6.2.2 DBMS 进程结构	199
6.2.3 DBMS 的体系结构	200
6.3 编程环境中的 SQL	207
6.3.1 嵌入式 SQL 综述	207
6.3.2 嵌入式 SQL 与主语言的通信	212
6.3.3 带有游标的查询	214
6.3.4 防止并发更新的保护措施	217
6.4 动态 SQL	218
6.4.1 直接执行的动态 SQL	219
6.4.2 带动态参数的动态 SQL	220
6.4.3 查询类动态 SQL	220
6.5 存储过程	223
6.5.1 什么是存储过程	223
6.5.2 创建存储过程	225
6.5.3 调用存储过程	225
6.6 习题	226

第 2 部分 系 统 篇

第 7 章 数据库的事务管理和并发访问控制	231
7.1 事务	231
7.1.1 问题的提出	231
7.1.2 事务的基本概念	232
7.2 数据库恢复系统	234
7.2.1 恢复机制	234
7.2.2 恢复处理的实现方法	238
7.2.3 故障类型及其恢复对策	244
7.3 事务执行的并发和控制	246
7.3.1 并发控制概述	246

7.3.2	并发控制的正确性准则	248
7.4	并发控制机制	250
7.4.1	基于锁的协议	250
7.4.2	多粒度封锁	255
7.5	死锁的发现和處理	258
7.5.1	死锁的處理	258
7.5.2	死锁預防	259
7.5.3	死锁检测与恢复	259
7.6	习题	261
第8章	数据库的约束、触发程序和安全	262
8.1	SQL 中的主键和外键	262
8.1.1	Create Table 语句中的完整性约束	262
8.1.2	主键、外键和参照完整性	263
8.2	对属性值的约束	265
8.2.1	CHECK 约束	265
8.2.2	域约束	266
8.2.3	断言	266
8.3	约束更新	267
8.3.1	数据库的更新	267
8.3.2	SQL 的约束更新	268
8.3.3	SQL 触发程序	268
8.4	SQL2 的安全和授权	270
8.4.1	权限	271
8.4.2	SQL 中的安全性机制	272
8.5	习题	273
第9章	数据库设计的理论基础	275
9.1	关系框架的设计问题	275
9.1.1	关系框架的评价	275
9.1.2	关系框架的分解	276
9.2	函数依赖	277
9.2.1	函数依赖的基本概念	277
9.2.2	键	278
9.3	函数依赖规则	279
9.3.1	函数依赖的逻辑蕴含	279
9.3.2	阿姆斯特朗公理	279
9.3.3	闭包、覆盖和最小覆盖	280
9.4	关系数据库模式设计	285

9.4.1	2NF	285
9.4.2	BCNF	286
9.4.3	3NF	288
9.4.4	多值依赖与 4NF	290
9.5	习题	294
第 10 章	对象和对象-关系数据库系统	295
10.1	面向对象数据库系统	295
10.1.1	ODMG 标准	295
10.1.2	OODBS 的定义	296
10.1.3	ODMG2.0 标准	296
10.2	ODMG 对象查询语言	300
10.2.1	OQL 中的 SELECT 语句	300
10.2.2	OQL 表达式的附加格式	301
10.2.3	OQL 中对象的赋值和建立	303
10.3	SQL3 概述	304
10.3.1	SQL3 中的元组对象	305
10.3.2	SQL3 的抽象数据类型	306
10.4	ODL/OQL 和 SQL3 方法的比较	309
10.5	习题	310
参考文献		311

基 础 篇

第1章 数据库系统导论

第2章 数据库的概念设计

第3章 关系运算

第4章 数据库语言

第5章 查询优化及实现

第6章 数据库系统及 SQL 编程环境

第1章

数据库系统导论

1.1 数据库系统

任何计算机应用都离不开对数据的处理,所以在计算机应用的历史中,数据管理系统一直是人们研究和开发的重要课题。随着计算机应用的发展,处理的数据量越来越大,出现了一类所谓的数据密集型应用。这类应用的特点是:系统运行涉及的数据量很大,须存储于外存;数据不随程序运行结束而消失,长期保存于计算机系统(现阶段主要是硬磁盘)中,即系统使用的是所谓的持久数据(persistent data);数据应支持多种应用共享。

数据密集型应用,是计算机应用中涉及领域最广的一类应用,包括各种信息管理系统,如办公、银行、保险,各种民航、交通的售票系统,情报和文献检索系统等。这类系统所面临的一个共同的问题是如何对它们要处理的数据进行安全有效的利用。

1.1.1 数据库系统的产生

早期的数据管理采用的是文件系统(file system),如 COBOL 语言、DOS 操作系统,便是典型的支持文件操作的语言和操作系统。文件管理系统的优点是:根据访问的方便性把数据组成若干命名文件,文件中的数据元素可以是简单类型的,也可以是结构(如记录)类型的,用户直接使用操作系统命令完成对指定文件的打开和读、写等操作。

使用文件系统管理数据,对大规模的数据管理任务,实际上是无法支持的。文件系统的主要缺点如下:

(1) 系统提供的数据库管理能力贫乏,只有打开、关闭、读、写等少数几个初级的文件操作命令。对文件的查询、修改、添加数据等操作,均需根据具体文件组织方式,在应用程序中编写,从而使数据的组织渗透到应用程序之中。当应用程序需要处理的数据关系复杂时,这对应用程序编写人员的要求是很高的,他们必须了解各文件的具体结构,并且还要知道它们之间的关联关系。

(2) 文件结构的修改会导致应用程序的大量修改。由于数据所描述的客观事物是不断发展变化的,所以这种文件结构的修改是客观需要的。这就使应用系统变得很脆弱,可用性差,维修困难。

(3) 文件系统只靠几个简单的操作系统命令操作文件,难以支持多用户访问同一文件。

因为它不提供对文件并发访问的控制与管理,严重妨碍了数据的共享和系统运行效率。例如在网络环境下,数据服务器不支持多用户并发访问,对系统运行效率的影响是显而易见的。

(4) 文件系统对数据的安全和保密难以保证。

针对文件系统存在的问题,人们研究、开发了数据库系统(database system),对各种应用程序共享的大规模数据实施统一管理。这种对数据库中数据进行统一管理的软件系统,叫数据库管理系统(database management system,DBMS)。DBMS的关键作用是对应用程序和数据物理存储之间,实现了相当程度的隔离。这主要是 DBMS 对用户透明地建立、保持、维护了一个称为数据字典(data dictionary)的元数据库,其中保存了数据库中所有数据文件的逻辑结构和物理存储信息,使应用程序对数据库中数据的操作,如查找、更新、添加等,在一定程度上实现了非过程化。应用程序对数据库操作主要说明做什么,至于如何做,则主要由 DBMS 自己决定。这在很大程度上避免了文件系统的所谓“文件结构渗透于应用程序中”带来的一系列的缺点,从而使数据库系统获得了广泛应用,已经成为大规模数据处理系统的基础软件。

1.1.2 数据库系统的定义

数据库系统是实现动态存储,管理大量相关数据,方便用户和应用程序访问的计算机软件、硬件资源组成的数据存储和管理系统,所以数据库系统主要包括数据库和数据库管理系统(DBMS)两部分。DBMS 是决定数据库系统性能、支持数据库运行的基础软件系统,是近几十年发展起来的一个计算机软件分支。从总体来看,它已基本完成了从实践向理论的过渡,但其理论、概念和方法还在随应用的发展而变化。由于数据库是一个应用广泛、结构复杂的系统,很难用简单的语言概括其全部含义。尽管如此,在数据库发展的初期,人们就给出了数据库各种各样的定义。其中 J Martin 的定义比较全面:

数据库是存储在一起的相关的数据集合,该数据集合可为多种应用服务而无有害或不必要的冗余;数据的存储独立于使用它的程序;对数据库的操作,如插入新数据、删除原有数据及查询、修改等,均以公用的和可控制的方式进行;数据被结构化。

定义由四层含义组成。

第一层是数据库的基本概念。数据库是存储在一起的相关的数据集合,它为多种应用服务,而无有害或不必要的冗余。这里有几个概念应予以说明:

(1) 存储在一起的数据集合。是指存储在同一个 CPU 控制之下,或存储在多个可以互相通信的 CPU 控制之下的存储介质上,但在同一个 DBMS 管理下的数据。

(2) 相关的数据集合。这里的“相关”并不是指数据库中数据之间必须存在什么函数关系。由于数据库是客观世界的数据描述,也就是为满足用户需要,数据库必须描述一个最小的客观事物(实体)集合及它们之间的联系。所以这里所说的“相关”,实际上主要是由于数据库描述的客观事物之间存在联系,使描述这些事物的数据之间也关联起来。如一个大学的数据库中关于各系的数据与教务、学生、人事等部门的数据,因它们之间存在的工作上的联系,而使它们之间的数据“相关”起来。

(3) 数据库的数据应能为多种应用服务,而无有害或不必要的冗余。这句话说明的是