

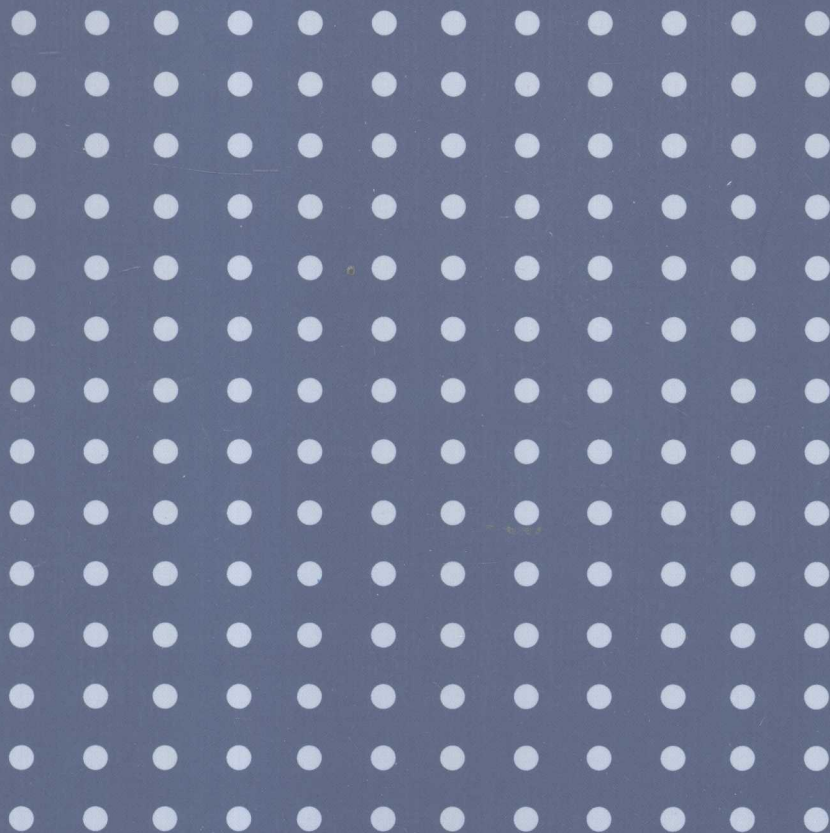


普通高等教育“十一五”国家级规划教材

重点大学计算机专业系列教材

数据库系统分析与实现

刘云生 编著



清华大学出版社



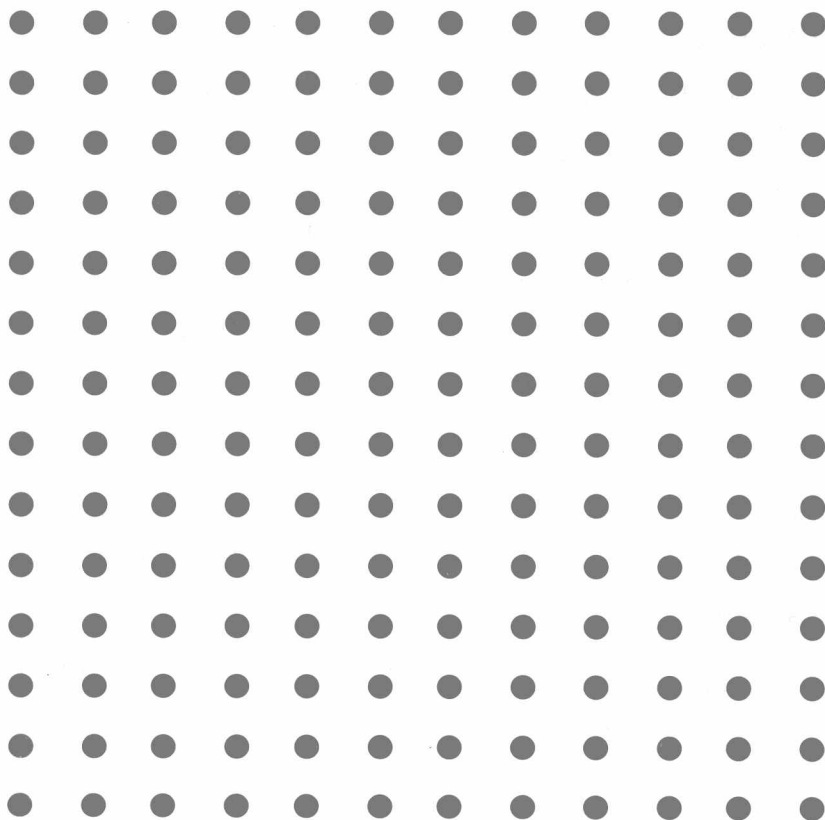


普通高等教育“十一五”国家级规划教材

重点大学计算机专业系列教材

数据库系统分析与实现

刘云生 编著



清华大学出版社

北京

内 容 简 介

本书是作者在30年来从事数据库的教学、研究、应用及DBMS开发工作的基础之上写成的。它将数据库系统理论、实现和应用紧密结合在一起,并以分析的观点、实现的视角、应用的立场来进行讨论,使读者不仅能“知其然”,还能“知其所以然”,而且还能懂得“如何应用”。它不仅包含了传统的数据库知识,还包含了支持非传统应用(如工程型、时间关键型)的现代数据库理论与技术。

本书内容非常丰富,共分5部分22章。系统基础部分为第1~4章,主要介绍数据库系统的基本概念与原理、各种典型的数据模型、SQL语言。系统结构部分为第5~7章,主要介绍数据库的系统结构、存储结构、DBMS结构。库系统实现部分为第8~13章,主要介绍查询处理与优化、事务处理、数据库安全与故障恢复等的实现技术。系统建立部分为第14~17章,主要介绍设计和建立一个数据库系统的过程、技术、工具与方法。系统发展部分为第18~22章,主要介绍实时、主动、内存、时态等现代(非传统)数据库的理论与技术。

本书适合作为高等学校计算机科学与技术、软件、电信、信息管理与信息系统及其他相关专业本科生和研究生的教材。也可供从事有关数据库应用开发、数据库系统管理、DBMS研发、数据库设计与建立、与非传统数据库技术相关等方面工作的科研和工程技术人员阅读与参考。

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 13701121933

图书在版编目(CIP)数据

数据库系统分析与实现/刘云生编著. —北京:清华大学出版社,2009.6

(重点大学计算机专业系列教材)

ISBN 978-7-302-19346-3

I. 数… II. 刘… 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(2009)第010777号

责任编辑:闫红梅

责任校对:李建庄

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

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京季蜂印刷有限公司

装 订 者:三河市兴旺装订有限公司

经 销:全国新华书店

开 本:185×260 印 张:35.75 字 数:865千字

版 次:2009年6月第1版 印 次:2009年6月第1次印刷

印 数:1~3000

定 价:49.00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:018074-01

出版说明

随着国家信息化步伐的加快和高等教育规模的扩大,社会对计算机专业人才的需求不仅体现在数量的增加上,而且体现在质量要求的提高上,培养具有研究和实践能力的高层次的计算机专业人才已成为许多重点大学计算机专业教育的主要目标。目前,我国共有 16 个国家重点学科、20 个博士点一级学科、28 个博士点二级学科集中在教育部部属重点大学,这些高校在计算机教学和科研方面具有一定优势,并且大多以国际著名大学计算机教育为参照系,具有系统完善的教学课程体系、教学实验体系、教学质量保证体系和人才培养评估体系等综合体系,形成了培养一流人才的教学和科研环境。

重点大学计算机学科的教学与科研氛围是培养一流计算机人才的基础,其中专业教材的使用和建设则是这种氛围的重要组成部分,一批具有学科方向特色优势的计算机专业教材作为各重点大学的重点建设项目成果得到肯定。为了展示和发扬各重点大学在计算机专业教育上的优势,特别是专业教材建设上的优势,同时配合各重点大学的计算机学科建设和专业课程教学需要,在教育部相关教学指导委员会专家的建议和各重点大学的大力支持下,清华大学出版社规划并出版本系列教材。本系列教材的建设旨在“汇聚学科精英、引领学科建设、培育专业英才”,同时以教材示范各重点大学的优秀教学理念、教学方法、教学手段和教学内容等。

本系列教材在规划过程中体现了如下一些基本组织原则和特点。

1. 面向学科发展的前沿,适应当前社会对计算机专业高级人才的培养需求。教材内容以基本理论为基础,反映基本理论和原理的综合应用,重视实践和应用环节。

2. 反映教学需要,促进教学发展。教材要能适应多样化的教学需要,正确把握教学内容和课程体系的改革方向。在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养,为学生知识、能力、素质协调发展创造条件。

3. 实施精品战略,突出重点,保证质量。规划教材建设的重点依然是专业基础课和专业主干课;特别注意选择并安排了一部分原来基础比较好的优秀教材或讲义修订再版,逐步形成精品教材;提倡并鼓励编写体现重点大学

计算机专业教学内容和课程体系改革成果的教材。

4. 主张一纲多本,合理配套。专业基础课和专业主干课教材要配套,同一门课程可以有多本具有不同内容特点的教材。处理好教材统一性与多样化的关系;基本教材与辅助教材以及教学参考书的关系;文字教材与软件教材的关系,实现教材系列资源配套。

5. 依靠专家,择优落实。在制订教材规划时要依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主编。书稿完成后要认真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平的以老带新的教材编写队伍才能保证教材的编写质量,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

教材编委会

前言

数据库是当前发展最快、最受人关注、应用最广泛的科学技术之一。数据库系统已经成为现代信息系统不可或缺的核心组成部分。在当今社会,不仅是传统的商业、管理和行政事务型应用已离不开数据库,那些实时、过程 and 控制的工程型应用领域也要求并且已经开始使用现代(非传统)数据库。因此,人们越来越普遍地要求全面学习和掌握数据库的理论知识、系统技术、应用方法及其新发展。

本书旨在将数据库的基础理论、系统实现技术和应用开发方法“紧密整合”(seamless integration),以提供一个统一分层的数据库知识体系;将有关数据库理论、技术与方法的叙述、分析、实现和应用有机地结合,使之形成一个较完整的数据库方法论体系;对现代(非传统)数据库的新发展,如实时、主动、内存、时态等数据库给予系统介绍,从而充实和丰富传统的数据库领域。

本书是作者 30 年来从事数据库教学、理论与实现技术研究、数据库管理系统(DBMS)尤其是现代(非传统)DBMS 的开发、数据库设计与应用的学习心得和工作总结,且汇入了作者在国外一些著名大学课堂实地听课和阅读国内外大量相关著作与论文的体会。它以分析的观点、从实现的角度、站在开发与应用的立场来进行讨论,希望不仅说明数据库系统“是什么”,还进一步分析“为什么”,且讨论“如何做”,使读者不仅能“知其然”,还能“知其所以然”,懂得“如何应用”。它不仅包含了关于管理、事务、决策分析型应用的传统数据库知识,还包含了作为作者多年研究成果的支持非传统应用(即工程实现、过程控制、实时处理等领域应用)的现代数据库理论与技术。

其内容分为五部分:

- 第一部分 数据库系统基础(第 1~4 章): 主要介绍数据库系统的基本概念与原理、数据的基本表示技术和各种典型的数据模型、当前最流行的数据库语言 SQL 及其使用。它侧重应用,适合于数据库应用的人们。
- 第二部分 数据库系统结构(第 5~7 章): 主要介绍数据库系统的体系结构、数据库存储结构、DBMS 实现结构。它侧重系统,适合于数据库系统开发与管理的人们。

- 第三部分 数据库系统实现(第 8~13 章): 主要介绍查询处理与优化、事务的概念及处理、数据库安全与故障恢复等方面的实现技术。它侧重实现, 适合于数据库管理系统实现与研发的人们。
- 第四部分 数据库系统建立(第 14~17 章): 主要介绍设计和建立一个数据库系统的过程、技术、工具与方法。它侧重设计, 适合于数据库设计与应用系统开发的人们。
- 第五部分 现代数据库技术新发展(第 18~22 章): 主要介绍实时、主动、内存、时态数据库, 以及工程、多媒体、Web 和空间数据库等现代(非传统)数据库的理论与技术。它侧重发展, 适合于需要数据库新技术的人们。

本书的手稿已在计算机学院、软件学院对本科生和研究生讲授了多次, 他们有的阅读了原手稿, 并提出过意见。尤其是软件学院本科 2006 级“数据库研讨组”的同学们录入了本书大部分的电子版。借此向他们致谢。

对于书中的许多内容特别是“现代数据库技术新发展”部分, 以及在成书的过程中, 作者的多届博士生、硕士生曾从各个不同的方面、以不同的形式做了许多工作。在此, 一并向他们表示诚挚的谢意。

诚如前面所说, 书中的许多方面是作者的学习与实践体会, 有的内容尤其是“现代数据库技术新发展”部分中的许多方面是作者的研究心得, 再加之作者才学疏浅, 水平与能力有限, 因此书中见仁见智之说、不妥或不足之处, 恐在所难免。切盼学术界同仁与各方读者不吝赐教。

劉雲生

2008 年秋于武昌喻家山

目录

第一部分 数据库系统基础

第 1 章 绪论	3
1.1 数据库管理技术	3
1.1.1 文件管理系统	4
1.1.2 数据库系统	5
1.2 数据库系统的发展	7
1.3 数据库的研究	9
1.3.1 数据库理论研究	9
1.3.2 数据库管理系统研究	10
1.3.3 数据库应用研究	10
1.4 数据库应用	11
1.5 数据库系统评估	12
1.5.1 数据特征	13
1.5.2 应用特征	13
1.5.3 DBMS 支持	13
1.6 现代数据库技术	14
习题 1	14
第 2 章 数据库系统概念	15
2.1 数据及其表示	15
2.1.1 数据与信息	15
2.1.2 数据表示的两个领域	16
2.1.3 数据项间的联系	17
2.1.4 记录间的联系	18
2.2 数据的视图	19
2.2.1 数据抽象	19

2.2.2	数据库的分层视图	19
2.2.3	数据独立性	21
2.2.4	模式与实例	22
2.3	数据模型	23
2.3.1	模型的概念	23
2.3.2	数据模型的概念	24
2.3.3	数据模型的分析	24
2.4	数据库语言	25
2.4.1	数据定义语言	25
2.4.2	数据操纵语言	26
2.4.3	宿主语言	26
2.5	数据库组织	27
2.6	查询	28
2.7	事务管理	28
2.7.1	接纳	28
2.7.2	资源管理	29
2.7.3	数据库一致性保护	29
2.7.4	故障恢复	29
2.7.5	执行管理	29
2.8	存储数据管理	29
2.9	数据库用户	30
2.9.1	数据处理人员	30
2.9.2	非数据处理人员(最终用户)	31
2.10	数据库系统概述	31
2.10.1	数据库系统组成	31
2.10.2	关系数据库系统	33
2.10.3	网状数据库系统	36
2.10.4	层次数据库系统	40
习题 2		43
第 3 章	数据库数据模型	44
3.1	数据模型的概念	44
3.2	E-R 模型	45
3.2.1	实体与实体集	45
3.2.2	属性、关键字及值域	46
3.2.3	联系与联系集	47
3.2.4	E-R 图	49
3.2.5	标识	50
3.2.6	联系的成员籍	51

3.2.7 弱实体	52
3.3 关系模型	53
3.3.1 概述	53
3.3.2 基本概念	54
3.3.3 关系的一般特征	56
3.3.4 关系数据完整性	56
3.3.5 关系代数	57
3.3.6 关系演算	63
3.4 记录式模型	66
3.4.1 网状模型	66
3.4.2 层次模型	68
3.4.3 数据完整性约束	70
3.5 语义模型	70
3.5.1 语义模型概述	70
3.5.2 扩展的关系模型	72
3.5.3 角色模型	74
3.5.4 函数模型	77
3.6 对象数据模型	78
3.6.1 面向对象的基本概念	78
3.6.2 对象数据结构	81
3.6.3 类结构	82
3.6.4 方法与消息	83
习题 3	84
第 4 章 SQL 数据库语言	86
4.1 SQL 的组成	86
4.1.1 主要部件	86
4.1.2 基本查询结构	87
4.2 数据定义及其维护	87
4.2.1 基本关系的定义及变更	87
4.2.2 视图的定义与撤销	89
4.2.3 索引的建立	89
4.3 基本查询	89
4.3.1 无条件简单查询	90
4.3.2 条件查询	90
4.3.3 空值查询	92
4.3.4 查询结果排序	93
4.4 使用聚集函数的查询	93
4.4.1 聚集函数	93

4.4.2	消去重复	94
4.4.3	查询结果分组	94
4.4.4	分组筛选	94
4.5	复杂查询	95
4.5.1	关系的连接	95
4.5.2	简单连接查询	96
4.5.3	元组变量	97
4.5.4	查询的并、交、差	97
4.5.5	嵌套的子查询	98
4.6	数据库的变更	101
4.6.1	数据插入	101
4.6.2	数据删除	101
4.6.3	数据修改	102
4.7	数据控制	102
4.7.1	安全性控制	102
4.7.2	完整性控制	104
4.7.3	事务执行控制	107
4.7.4	事务并发控制	107
4.8	嵌入式 SQL	109
4.8.1	阻抗不匹配问题	109
4.8.2	SQL 与主语言的接口	109
4.8.3	游标	110
4.9	动态 SQL	113
4.9.1	嵌入式动态 SQL 语句	113
4.9.2	参数与变元	114
4.9.3	SELECT 语句的处理	115
4.9.4	动态 SQL 游标的使用	116
4.10	SQL 的触发器	116
习题 4		117

第二部分 数据库系统结构

第 5 章	数据库系统体系结构	121
5.1	数据库系统的一般功能模型	121
5.2	数据库系统的层次结构模型	122
5.3	集中式数据库系统结构	125
5.4	分布式数据库系统结构	126
5.5	客户机/服务器数据库系统结构	127
5.5.1	一般 C/S 系统	127

5.5.2	事务服务器系统	128
5.5.3	数据服务器系统	129
5.6	并行数据库系统结构	130
5.6.1	无共享结构	130
5.6.2	全共享结构	131
5.6.3	共享磁盘结构	131
5.6.4	层次结构	131
5.7	移动数据库系统结构	131
习题 5		132
第 6 章	数据库存储结构	133
6.1	数据库存储空间组织	133
6.1.1	存储层次结构	133
6.1.2	磁盘结构	135
6.1.3	RAID(磁盘阵列)技术	139
6.1.4	磁盘空间管理	142
6.2	缓冲区管理	143
6.2.1	缓冲区管理器	143
6.2.2	缓冲区替换策略	144
6.3	物理数据库模型	145
6.3.1	逻辑存储结构到物理存储结构的映射	145
6.3.2	存储数据的安置	146
6.3.3	存储文件结构	147
6.3.4	页结构	149
6.3.5	存储记录格式	151
6.3.6	特殊数据项的存储处理	153
6.3.7	记录的存储处理	157
6.4	文件的组织形式	159
6.5	顺序结构	160
6.6	索引结构	161
6.6.1	KTA 变换	161
6.6.2	索引的组织	162
6.6.3	索引的特性	162
6.6.4	索引存取	164
6.7	Hash 结构	165
6.7.1	一般 Hash 组织模型	165
6.7.2	桶式 Hash	169
6.7.3	撒播 Hash	169
6.7.4	索引 Hash	170

6.7.5	可扩展的 Hash	170
6.7.6	线性扩展 Hash	172
6.7.7	多目录 Hash	174
6.8	树索引结构	175
6.8.1	B-树结构	175
6.8.2	B ⁺ -树结构	179
6.8.3	B-树的性能分析	180
6.8.4	索引顺序结构	181
6.9	链表结构	183
6.9.1	整体链表结构	183
6.9.2	多分链表结构	184
6.9.3	链表的查询	185
6.9.4	链表结构的数据维护	186
6.10	倒排表结构	186
习题 6		187
第 7 章	数据库管理系统体系结构	190
7.1	数据的集成管理	190
7.1.1	数据独立	190
7.1.2	数据共享	193
7.1.3	数据的集中控制	194
7.1.4	数据可用性	194
7.2	DBMS 的功能	195
7.3	DBMS 的组成与结构	195
7.4	数据字典系统	197
7.4.1	数据字典结构	197
7.4.2	数据字典的管理	197
7.4.3	数据字典系统的实现	198
7.5	DBMS 的进程结构	199
7.5.1	N:0 式结构	199
7.5.2	N:N 式结构	200
7.5.3	N:1 式结构	200
7.5.4	N:M 式结构	201
习题 7		202

第三部分 数据库系统实现

第 8 章	数据库使用	205
8.1	数据库运行	205

8.1.1 运行环境的构成	205
8.1.2 运行环境的类型	206
8.1.3 处理方式	207
8.2 数据库使用方式	207
8.2.1 编程使用方式	207
8.2.2 交互使用方式	209
8.3 用户请求的实现过程	209
8.4 数据库的存取	211
8.4.1 逻辑数据存取	212
8.4.2 物理数据存取	212
8.4.3 数据库存取实现过程	212
习题 8	214
第 9 章 查询处理	215
9.1 查询处理概述	215
9.2 查询处理步骤	217
9.2.1 语法分析	217
9.2.2 语法树	218
9.2.3 语句翻译	219
9.2.4 逻辑计划生成	220
9.2.5 逻辑查询计划优化	221
9.2.6 物理查询计划生成	221
9.2.7 物理计划优化	223
9.2.8 查询计划执行	224
9.3 查询优化技术	224
9.3.1 关系代数的等价性定律	224
9.3.2 逻辑查询优化的一般策略	227
9.3.3 物理操作实现的算法	227
9.4 查询的代价估算	230
9.4.1 操作代价模型	230
9.4.2 一趟算法的代价	231
9.4.3 基于排序两趟算法的代价	231
9.4.4 基于 Hash 两趟算法的代价	232
9.4.5 基于索引算法的代价	233
9.5 查询计划的选择方法	233
9.5.1 相关技术	233
9.5.2 逻辑查询计划的启发式选优	234
9.5.3 物理查询计划的选择方法	234
习题 9	236

第 10 章 事务	237
10.1 事务的概念	237
10.2 事务的 ACID 特性	238
10.2.1 原子性	238
10.2.2 一致性	239
10.2.3 隔离性	239
10.2.4 永久性	240
10.3 事务的分类	240
10.3.1 按构造分类	240
10.3.2 按执行时间分类	241
10.3.3 按使用数据的方式分类	241
10.3.4 按执行的方式分类	241
10.4 事务的状态	241
10.5 事务的正确性	243
10.6 复杂事务的模型	243
10.6.1 复杂事务的一般模型	243
10.6.2 典型复杂事务类型	244
10.6.3 复杂事务的特性	245
10.6.4 嵌套事务	246
习题 10	247
第 11 章 事务处理	248
11.1 事务处理的任务	248
11.2 事务经历模型	249
11.2.1 事务管理原语操作	250
11.2.2 数据库原语操作	250
11.2.3 事务经历	251
11.2.4 事务处理过程	252
11.3 事务调度	253
11.4 可串行性	255
11.4.1 效果可串行性	255
11.4.2 视图可串行性	255
11.4.3 冲突可串行性	256
11.4.4 可串行性的测试	258
11.5 并发调度的原则	262
11.5.1 事务相互干扰的问题	262
11.5.2 可串行性的程度	264
11.5.3 可恢复性	264

11.5.4	隔离的级别	266
11.6	封锁式并发控制	267
11.6.1	锁	267
11.6.2	加锁与解锁	268
11.6.3	封锁的主要问题	269
11.6.4	两阶段封锁	270
11.6.5	2PL 的变种	271
11.6.6	锁的升级与更新锁	273
11.6.7	增量锁	274
11.6.8	封锁管理及其实现	275
11.7	多层结构数据的封锁	277
11.7.1	多粒度锁	277
11.7.2	树结构的封锁	278
11.8	幻象处理与索引的封锁	280
11.9	死锁处理	282
11.9.1	死锁的预防	282
11.9.2	死锁检测	284
11.9.3	死锁解除	285
11.10	无锁的并发控制	285
11.10.1	基于时间戳的方法	285
11.10.2	乐观的方法	289
11.10.3	多版本的方法	291
习题 11		293
第 12 章	故障恢复	295
12.1	数据库故障恢复概述	295
12.2	故障分类	297
12.2.1	事务失败	297
12.2.2	系统崩溃	297
12.2.3	介质故障	297
12.3	日志	298
12.3.1	日志记录类型	298
12.3.2	日志记录格式	299
12.3.3	日志管理	300
12.4	检验点	301
12.4.1	清晰检验点	302
12.4.2	含混检验点	303
12.4.3	模糊检验点	303
12.5	备份	304

12.5.1	清晰备份	305
12.5.2	模糊备份	305
12.6	基于日志的恢复	306
12.6.1	事务失败恢复	306
12.6.2	系统崩溃的恢复	307
12.6.3	介质故障恢复	311
12.7	其他恢复策略	312
12.7.1	ARIES	312
12.7.2	影子法	313
12.7.3	逻辑日志法	314
习题 12		314
第 13 章	数据库安全	315
13.1	数据库安全概述	315
13.2	用户识别	316
13.2.1	用户的个人特征识别	317
13.2.2	用户的特有东西识别	317
13.2.3	用户的专门知识识别	317
13.3	授权	317
13.3.1	授权方案	318
13.3.2	权限转授	318
13.3.3	视图与程序授权	319
13.4	存取控制模型	319
13.4.1	存取控制的层次	320
13.4.2	存取控制矩阵	321
13.4.3	逻辑存取控制模型	321
13.5	存取控制策略	322
13.5.1	自主存取控制	322
13.5.2	强制存取控制	322
13.5.3	多级安全与多实例性	323
13.5.4	隐通道	324
13.5.5	统计数据库的安全性	324
13.6	数据加密	325
13.6.1	基本概念	325
13.6.2	基本的传统加密方法	326
13.6.3	改进的传统加密方法	327
13.6.4	公钥加密法	329
13.7	数据库安全的其他问题	330
13.7.1	安全级别	330