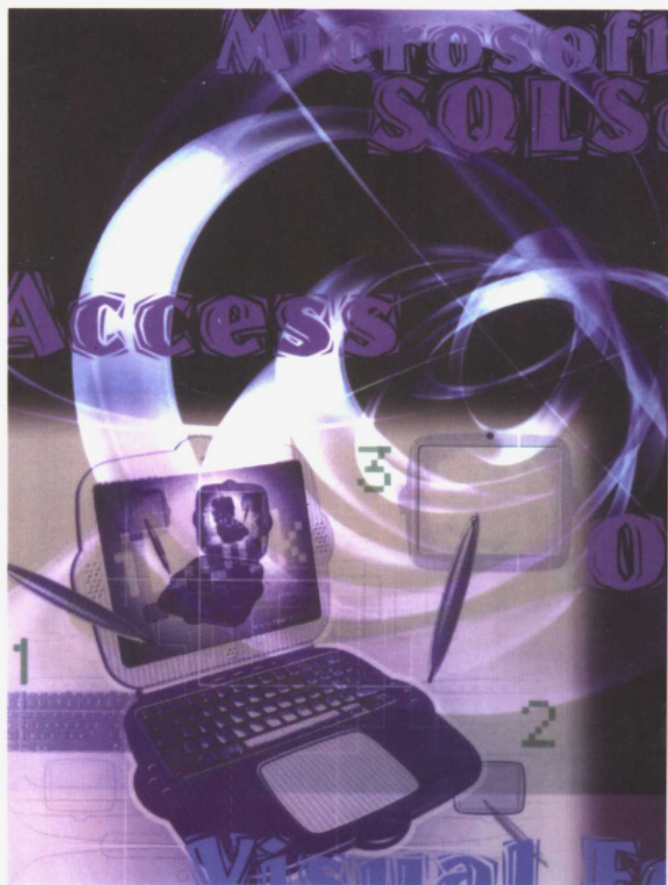


数据库系统 原理与应用教程

(第二版)

- ◆ 数据库系统基础
- ◆ 数据库建模——ER图、IDEF1x图和ODL方法
- ◆ 关系模型和关系模式
- ◆ 关系代数和Datalog语言
- ◆ SQL语言及其高级功能
- ◆ 分布式数据库技术
- ◆ 数据仓库技术
- ◆ 数据库系统应用工具和开发实例



闪四清 编著



清华大学出版社

数据库系统原理与应用教程（第二版）

本书是全面介绍数据库系统基本原理、技术实现和应用知识的最新教程。本书内容全面、概念清晰、语言流畅、图文并茂，把最新的数据库开发技术和应用实例与经典的数据库理论紧密结合，既可作为高等院校相关专业的数据库课程教材，也可作为初、中级数据库系统开发人员的参考书。

本书特色

● 真实案例

本书以一个贯穿全书始终的图书管理示例为主线，把各种数据库技术的知识要点串联成一个逻辑严密的整体。在这个示例不断深化、丰富和完善的过程中，读者也必将随之学习、理解和掌握数据库这门技术。

● 教法经典

本书按照读者学习数据库技术的自然过程将全书内容分为设计篇、编程篇、实现篇和提高篇四个部分，由浅入深，有利于读者循序渐进地掌握数据库技术。在每一章的开始，都有本章的内容简介和重点提示，在结束时给出该章的小结和对应的习题，既有利于老师教学，又有利于学生掌握重点内容和提高解决实际问题的能力。

● 思考和练习

本书在每一章的最后都安排了具有针对性的习题，包括问答题、思考题等。这些都是从教学与实践中提取出来的，对读者具有指导意义的习题。通过这些思考和练习，读者可以有效地巩固所学的技能 and 知识，并为以后的实际操作打下良好的基础。

高等院校计算机应用技术系列教材

数据库系统原理与应用教程

(第二版)

闪四清 编著

清华大学出版社

北 京

内 容 简 介

数据库技术是信息系统的一个核心技术。本书是一本全面介绍数据库技术基本原理、技术实现和应用等知识的教程。全书包括4部分共16章。第一部分讨论了数据库的建模方法,第二部分全面介绍了SQL语言及其理论基础,第三部分探讨了数据库技术深层次的实现方法,第四部分详细阐述了数据库技术的发展趋势和应用。

本书内容全面、概念清晰、语言流畅、图文并茂,可作为高等院校计算机相关专业数据库课程的教材,也可供从事于计算机软件工作的科技人员(如信息管理系统、电子商务网站开发人员等)参阅。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

数据库系统原理与应用教程/闪四清 编著. —2版 — 北京:清华大学出版社,2004

(高等院校计算机应用技术系列教材)

ISBN 7-302-08156-5

I. 数… II. 闪… 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(2004)第013346号

出 版 者: 清华大学出版社 地 址: 北京清华大学学研大厦

<http://www.tup.com.cn> 邮 编: 100084

社 总 机: 010-62770175 客户服务: 010-62776969

组稿编辑: 曹 康

文稿编辑: 胡辰浩

封面设计: 孔祥丰

版式设计: 康 博

印 刷 者: 北京市清华园胶印厂

装 订 者: 三河市金元装订厂

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 24.25 字数: 575千字

版 次: 2004年3月第1版 2004年3月第1次印刷

书 号: ISBN 7-302-08156-5/TP·5894

印 数: 1~6000

定 价: 35.00元

本书如存在文字不清、漏印以及缺页、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)62770175-3103 或 (010)62795704

第一版前言

当今是一个信息爆炸的时代,信息已经成为社会和经济发展的主要支柱之一。大量信息的产生、处理、存储、传播和使用推动了社会的进步和经济的发展。信息系统是一种以加工、处理信息为主的计算机系统。数据库技术作为一种存储和使用信息的信息系统核心技术正在发挥着越来越重要的作用。例如,现在的银行无法再依赖手工记账来处理复杂的储蓄、结算、信贷和信用卡等业务,现在的邮政、电信也无法离开信息系统开展各种邮政和电信业务,现在的航空运输也无法离开信息系统来管理庞大的分布式售票网点,现在的加工企业无法离开信息系统来管理庞大的零部件仓库、计划、生产、销售和财务等,现在的图书出版业无法依赖手工的方式管理大量图书的出版、库存和销售等业务。

本书对数据库技术进行了全面的阐述和研究。在结合大量实例和本人教学体会的基础上,对数据库技术的各个领域进行了深入浅出的剖析,对数据库技术的重点和难点进行了详细的描述,力求做到概念清晰、生动活泼。

本书包括了4部分共16章内容。第一部分是设计篇,重点介绍了ER图、ODL方法、IDEFlx图和关系模型等数据库建模技术。第二部分是编程篇,对数据库技术中的关系代数、Datalog语言等理论以及标准SQL语言进行了讲述。第三部分是实现篇,讨论了有关索引、查询、事务和并发控制等数据库技术内核的实现问题。第四部分是提高篇,对典型的桌面型数据库产品、分布式数据库产品以及信息系统的开发方法进行了介绍和分析。

在每章的开始,对该章将要涉及的内容和作用进行了简单的介绍,然后指出了学完本章后应该掌握的重要内容。每章正文内容结束之后,有一个对本章所讲述的内容进行总结和评价的小结。另外,每章后面都附有思考和练习。思考题侧重于加深读者对本章涉及概念的理解,练习题侧重于培养学生应用本章学到的知识来解决实际问题的能力。

本教程还有一个显著的特点,就是以贯穿全书的图书管理示例为主线,把各种数据库技术的知识要点,串联成一个逻辑严密的整体。这个示例不断深化、丰富和完善的过程,实际上就是读者不断学习、理解和掌握数据库技术的过程。

限于本人水平有限,书中难免有不妥之处,欢迎广大读者批评指正。

作 者

2001年1月

前 言

本书于 2001 年出版后, 由于内容新颖、结构合理、概念清晰、示例真实连贯、讲述生动活泼, 受到了广大读者的欢迎, 连续印刷了多次。

两年多来, 数据库技术又有了许多新的发展, 读者迫切希望全面掌握新技术、新知识。同时本人在教学实践中又积累了更多的经验。本书的再版及时反映了数据库技术的新进展并全面总结了本人的教学经验。

本次修订过程中, 对整个教程的结构进行了较大的改动, 以便更好地满足读者的需要。第二版重写了本书的第 4 部分, 增加了分布式数据库技术和数据仓库技术的内容, 删除了原来的第 14 章和第 15 章, 将删除的部分内容融入到了其他章节中。在第 9 章和第 10 章中, 增加了“动手练习”一节, 目的是培养学生的实际动手能力。此外, 对第一版中的不妥之处和错误之处进行了认真的修订。

希望广大读者从本书中学习到自己需要的知识, 真正地掌握数据库系统的理论和应用。这正是本次修订的目的和本人的心愿。

作 者

2003 年 12 月

目 录

第 I 部分 设计篇

第 1 章 步入数据库系统世界	3
1.1 概述	3
1.1.1 数据库系统应用示例	4
1.1.2 数据库系统的作用	6
1.1.3 常见的数据库系统	8
1.1.4 数据库应用程序的开发过程	11
1.2 数据库技术的演变	12
1.2.1 数据库技术的环境	12
1.2.2 传统的文件管理阶段	13
1.2.3 现代的数据库管理系统阶段	14
1.3 数据库系统架构	16
1.3.1 数据存储架构	16
1.3.2 数据视图管理架构	18
1.3.3 数据库应用架构	19
1.4 数据库管理系统组件	23
1.4.1 数据和元数据	23
1.4.2 输入	24
1.4.3 查询处理器	25
1.4.4 存储管理器	26
1.4.5 事务管理器	26
1.5 数据库技术的发展趋势	27
1.5.1 数据库技术面临的挑战	27
1.5.2 数据库技术的未来	27
1.6 小结	29
1.7 思考和练习	29
第 2 章 数据库建模——ER 图	31
2.1 ER 图的基本概念	31

2.1.1 实体集和实体集中的属性	32
2.1.2 联系集和联系集中的属性	35
2.1.3 绘制 ER 图	37
2.2 ER 图设计原则	39
2.2.1 真实性原则	40
2.2.2 避免冗余原则	40
2.2.3 简单性原则	40
2.2.4 恰当性原则: 用实体集还是用属性	41
2.2.5 恰当性原则: 用实体集还是用联系集	41
2.3 对约束建模	42
2.3.1 映射约束	42
2.3.2 键码约束	45
2.3.3 参照完整性约束	47
2.3.4 其他类型的约束	48
2.4 弱实体集	48
2.4.1 弱实体集产生的原因	49
2.4.2 对弱实体集的要求	50
2.4.3 弱实体集表示方法	50
2.5 ER 图的高级技术	51
2.5.1 特殊化技术	51
2.5.2 概括技术	52
2.5.3 属性继承技术	53
2.5.4 概括约束设计	53
2.5.5 聚集技术	54
2.6 小结	56
2.7 思考和练习	56
第 3 章 数据库建模——IDEF1x 图	58
3.1 概述	58
3.1.1 IDEF 方法族简介	58
3.1.2 IDEF1x 图的概念和特点	60
3.2 IDEF1x 图的语法和语义	61
3.2.1 实体	61
3.2.2 连接联系	62
3.2.3 分类联系	66
3.2.4 非确定联系	68

3.2.5 属性	69
3.2.6 主键码和次键码	71
3.2.7 外键码	73
3.3 使用 IDEF1x 图建立数据库模型的过程	75
3.3.1 准备阶段	75
3.3.2 定义实体阶段	78
3.3.3 定义联系阶段	79
3.3.4 定义键码阶段	81
3.3.5 定义属性阶段	85
3.3.6 文件编制阶段	87
3.4 小结	88
3.5 思考和练习	88
第 4 章 数据库建模——ODL 方法	90
4.1 面向对象概述	90
4.1.1 面向对象模型和关系模型比较	90
4.1.2 面向对象基本概念	91
4.2 ODL 的基本结构	93
4.2.1 ODL 概念	93
4.2.2 ODL 接口说明	94
4.2.3 ODL 中的属性	95
4.2.4 ODL 中的联系	96
4.2.5 反向联系的概念	97
4.2.6 联系的多重性	99
4.2.7 ODL 中的类型	101
4.2.8 ODL 中方法署名的说明	104
4.3 子类和约束	106
4.3.1 ODL 中的子类	106
4.3.2 ODL 中的约束	110
4.4 小结	115
4.5 思考和练习	115
第 5 章 关系模型	117
5.1 基本概念	117
5.2 从 ER 图到关系模型设计的转换	122
5.2.1 ER 图和关系模式的比较	123
5.2.2 实体集到关系模式的转换	123

5.2.3	联系集到关系模式的转换	124
5.2.4	弱实体集到关系模式的转换	126
5.2.5	ER 图中其他元素的转换	127
5.2.6	键码的转换	129
5.3	从 IDEF1x 图到关系模型设计的转换	130
5.4	从 ODL 设计到关系模型设计的转换	131
5.4.1	ODL 设计和关系模式设计的比较	131
5.4.2	从 ODL 属性到关系模式的属性	132
5.4.3	ODL 设计中非原子属性的转换	132
5.4.4	ODL 设计中联系的转换	136
5.4.5	ODL 设计中联系和反向联系问题	137
5.4.6	ODL 设计中子类结构的转换	137
5.5	小结	139
5.6	思考和练习	139
第 6 章	关系模式的规范化设计	141
6.1	概述	141
6.1.1	异常问题	141
6.1.2	泛关系模式和数据库模式	142
6.2	函数依赖	143
6.2.1	函数依赖的定义	143
6.2.2	函数依赖和键码	144
6.2.3	逻辑蕴含	145
6.2.4	函数依赖的推理规则	146
6.2.5	属性集的闭包	147
6.2.6	正则覆盖	149
6.3	关系模式的分解	150
6.4	关系模式的范式	151
6.4.1	第一范式	151
6.4.2	BCNF 范式	152
6.4.3	分解成 BCNF 模式的算法	152
6.4.4	函数依赖的投影	154
6.4.5	第三范式	155
6.4.6	第二范式	157
6.5	多值依赖	157
6.5.1	多值依赖的概念	158

6.5.2 多值依赖的推理规则	159
6.5.3 第四范式	159
6.6 范式之间的关系	160
6.7 数据库模式的实例	161
6.7.1 图书管理数据库模式	161
6.7.2 计算机产品信息管理数据库模式	163
6.8 小结	164
6.9 思考和练习	165

第 II 部分 编 程 篇

第 7 章 关系代数基本理论	169
7.1 概述	169
7.2 关系代数的基本运算	170
7.2.1 集合运算	170
7.2.2 选择运算	172
7.2.3 投影运算	174
7.2.4 笛卡儿积运算	175
7.2.5 改名运算	176
7.3 关系代数的附加运算	177
7.3.1 自然连接运算	177
7.3.2 θ 连接运算	178
7.3.3 除法运算	179
7.3.4 赋值运算	180
7.4 扩充的关系代数运算	181
7.4.1 广义投影运算	181
7.4.2 外连接运算	182
7.4.3 聚集运算	183
7.5 关系演算	183
7.5.1 元组关系演算	184
7.5.2 域关系演算	184
7.6 关系代数的修改运算	185
7.6.1 插入运算	185
7.6.2 删除运算	186
7.6.3 修改运算	186

7.7	关系代数表达式的优化策略	187
7.8	小结	188
7.9	思考和练习	188
第 8 章	Datalog 语言	190
8.1	基本概念	190
8.1.1	基本结构	190
8.1.2	一般规则	192
8.1.3	安全规则	193
8.1.4	外延谓词和内涵谓词	195
8.2	关系代数向 Datalog 规则的转换	196
8.2.1	从集合运算到 Datalog 规则	196
8.2.2	从投影运算到 Datalog 规则	197
8.2.3	从笛卡儿乘积到 Datalog 规则	197
8.2.4	从选择运算到 Datalog 规则	198
8.2.5	从连接运算到 Datalog 规则	200
8.2.6	从多重运算到 Datalog 规则	200
8.3	递归原理	201
8.3.1	关系代数存在的问题	201
8.3.2	计算最小固定点	203
8.3.3	使用 Datalog 规则表示固定点公式	204
8.4	包的运算	205
8.4.1	包的意义	205
8.4.2	包的关系运算	206
8.4.3	包的逻辑运算	209
8.5	小结	209
8.6	思考和练习	210
第 9 章	SQL 语言初步	212
9.1	概述	212
9.2	简单查询语句	213
9.2.1	最简单的查询语句	214
9.2.2	投影	215
9.2.3	选择	217
9.2.4	字符串比较	220
9.2.5	日期和时间的比较	221
9.2.6	排序结果	221

9.2.7 副本	223
9.3 高级查询语句	224
9.3.1 连接查询	224
9.3.2 集合查询	230
9.3.3 聚集查询	232
9.3.4 子查询	234
9.4 数据库更新语句	236
9.4.1 插入数据	237
9.4.2 删除数据	238
9.4.3 修改数据	238
9.5 数据定义语句	239
9.5.1 数据类型和域	240
9.5.2 管理表	241
9.5.3 默认值	243
9.5.4 索引定义	243
9.5.5 管理视图	244
9.6 递归查询	246
9.7 动手练习	247
9.8 小结	248
9.9 思考和练习	249
第 10 章 SQL 语言高级功能	251
10.1 约束	251
10.1.1 键码约束	251
10.1.2 外键码约束	253
10.1.3 非空约束	254
10.1.4 CHECK 约束	255
10.1.5 断言约束	256
10.2 触发器	258
10.2.1 触发器的结构	258
10.2.2 触发器示例	260
10.3 游标	261
10.3.1 定义游标	261
10.3.2 游标的选项	262
10.4 嵌入式 SQL 语言	263
10.4.1 接口和声明	263

10.4.2	动态的 SQL 语句	264
10.5	安全控制和权限管理	265
10.5.1	权限的类型	265
10.5.2	授权和收权	266
10.6	动手练习	267
10.7	小结	268
10.8	思考和练习	268

第Ⅲ部分 实 现 篇

第 11 章	索引和散列技术	273
11.1	概述	273
11.1.1	文件组织方式	273
11.1.2	顺序文件组织	274
11.1.3	聚集文件组织	275
11.2	索引技术	276
11.2.1	基本概念	276
11.2.2	顺序索引	276
11.2.3	B ⁺ 树索引文件	280
11.3	散列技术	280
11.3.1	基本概念	281
11.3.2	散列索引	281
11.4	Microsoft SQL Server 系统中的索引	282
11.4.1	索引类型	282
11.4.2	创建索引的方法	284
11.5	小结	285
11.6	思考和练习	285
第 12 章	查询处理技术	286
12.1	概述	286
12.2	查询处理的代价模型	288
12.3	单个关系运算的代价估计	289
12.3.1	选择运算的代价估计	289
12.3.2	连接运算的代价估计	290
12.3.3	集合运算的代价估计	291
12.4	表达式运算的代价估计	291

12.4.1 实体化方法	291
12.4.2 流水线方法	292
12.5 Microsoft SQL Server 系统的查询处理器	293
12.5.1 特性概述	293
12.5.2 交互式操作	294
12.5.3 执行计划	295
12.5.4 优化索引	296
12.5.5 优化隐藏	297
12.6 小结	298
12.7 思考和练习	299
第 13 章 事务和并发控制	300
13.1 事务	300
13.1.1 事务的基本概念	300
13.1.2 事务的状态变迁图	301
13.1.3 事务的并发执行	302
13.1.4 事务的一致性级别	304
13.2 并发控制	305
13.2.1 并发操作带来的问题	305
13.2.2 锁	305
13.2.3 死锁	306
13.3 Microsoft SQL Server 系统的事务和锁	307
13.3.1 事务的类型	307
13.3.2 事务的工作原理	308
13.3.3 锁的粒度	310
13.3.4 锁的类型	313
13.3.5 死锁和死锁的解决方案	315
13.4 小结	316
13.5 思考和练习	316

第IV部分 提 高 篇

第 14 章 分布式数据库技术	319
14.1 概述	319
14.2 分布式数据库的概念和特点	321
14.2.1 分布式数据库的概念	322

14.2.2	分布式数据库的特点	322
14.3	分布式数据存储技术	324
14.3.1	数据复制	324
14.3.2	数据分片	326
14.3.3	数据混合存储	329
14.4	分布式数据查询技术	330
14.4.1	分布式数据查询示例	331
14.4.2	基于半连接的优化策略	332
14.4.3	基于连接的优化策略	333
14.5	Microsoft SQL Server 的分布式计算特点	333
14.5.1	Microsoft Distributed Transaction Coordinator	333
14.5.2	链接服务器	334
14.5.3	数据复制	334
14.6	小结	338
14.7	思考和练习	338
第 15 章	数据仓库技术	339
15.1	概述	339
15.2	数据挖掘和数据中心库	341
15.2.1	数据挖掘技术	342
15.2.2	数据中心库	344
15.3	数据仓库的支持工具	345
15.3.1	OLAP 服务	345
15.3.2	数据转换服务	346
15.3.3	PivotTable 服务	347
15.4	数据仓库的实现步骤	347
15.4.1	确定用户需求	348
15.4.2	设计和建立数据库	348
15.4.3	提取和加载数据	351
15.5	数据的粒度、分割和元数据	352
15.6	小结	353
15.7	思考和练习	353
第 16 章	信息系统的开发过程	354
16.1	信息系统概述	354
16.2	开发信息系统的方法论	356
16.2.1	方法类型	356

16.2.2 开发策略.....	357
16.3 信息系统开发阶段.....	358
16.3.1 调查研究阶段.....	358
16.3.2 系统分析阶段.....	360
16.3.3 系统设计阶段.....	362
16.3.4 系统实施阶段.....	365
16.3.5 系统评价阶段.....	365
16.4 影响信息系统开发成功和失败的因素.....	366
16.5 小结.....	367
16.6 思考和练习.....	368
参考文献.....	369