

普通高校本科计算机专业特色教材精选·数据库

数据库原理

赵文栋 张少娴 徐正芹 主编



清华大学出版社

普通高校本科计算机专业特色教材精选·数据库

数据库原理

赵文栋	张少娴	徐正芹	主 编
彭来献	高 屹	李艾静	副主编
	徐任晖	王向东	参 编



清华大学出版社
北京

内 容 简 介

本书重点介绍了数据库的原理、设计与实现技术,对传统的数据库理论进行了精练,保留了核心与实用部分,并适当增加了实验内容,以突出实践能力的培养。全书由三部分共18章组成:第一部分介绍数据库基本原理,包括数据库的基础知识、数据模型、结构化查询语言 SQL 和数据的安全性等内容;第二部分介绍数据库的设计与开发,包括关系模式的规范化和数据库的设计与建模等内容;第三部分讨论数据库管理系统(DBMS)的设计与实现,以一个实际的 DBMS 为例,介绍 DBMS 的设计、存储管理、数据缓冲、数据操纵、索引技术、查询处理、事务管理和故障恢复等知识。

本书适合于高等院校计算机及相关专业的本科生和研究生教学使用,也可供相关领域技术人员参考。

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

数据库原理/赵文栋,张少娴,徐正芹主编. —北京:清华大学出版社,2019

(普通高校本科计算机专业特色教材精选·数据库)

ISBN 978-7-302-53594-2

I. ①数… II. ①赵… ②张… ③徐… 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(2019)第 174002 号

策划编辑:袁勤勇

责任编辑:袁勤勇

封面设计:傅瑞学

责任校对:梁毅

责任印制:丛怀宇

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 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

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市铭诚印务有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:22 字 数:519千字

版 次:2019年10月第1版 印 次:2019年10月第1次印刷

定 价:49.00元

产品编号:083850-01

前言

PREFACE

数据库是计算机专业公认比较枯燥难学的一门课程，但又是绝对重要和必须学好的基础理论。它的重要性不仅在于数据库技术本身不可替代的重要价值，还在于贯穿其中的计算机科学的理论精华和思维模式。对于广大读者来说，学习数据库相关知识是构筑自身计算机理论体系，建立计算机思维的必要训练过程。

真正掌握数据库知识需要大量的理论学习与实践。本书选取数据库理论的精要内容，并突出实验环节，追求的目标是在尽可能短的时间内，使读者清晰了解数据库的完整概念框架，理解其核心理论的本质，并通过实际动手操作真正掌握数据库原理的要义，在此基础上培养优秀数据库开发者应具备的全面素质，从而推开数据库开发殿堂的大门。

1. 目标读者群

本书适合于计算机及相关专业的本科生和研究生教学使用，也可作为软件开发人员(包括但不限于数据库开发人员)的技术参考书。希望本书的出版能够为计算机专业学生以及所有计算机爱好者提供一个快速学习数据库入门知识的渠道。

2. 学习理念与途径

数据库的应用方式有多种，学习的目的也有所不同。对数据库开发人员，主要有两类情形，一类是以管理大量结构化数据为目的，完成信息系统的开发。在网络化和 SOA 架构普及的今天，很少有应用或服务能够离开后台数据的支持，这是数据库应用的主要方式。另一类是对数据库技术的应用，典型的情形就是内嵌式数据库，这类应用的场合越来越多，很多系统内部都会内嵌一个精巧的数据库来实现数据的高效管理操作，或利用数据库管理系统的设计实现思想来开发其他专用系统。不论何种应用方式，能否获得成功越来越取决于开发者对数据库原理本身的掌

握程度。因此本书在学习理念上,更推崇在深入理解理论知识基础上的应用开发,使开发者不但知其然,还要知其所以然,从而能够以最合理的技术实现开发目标。

在本书的撰写过程中,特别研究借鉴了国外优秀数据库教材和课程的教学经验。对比以往国内在数据库教学方面的做法,明显感觉我们常常满足于平铺直叙地讲解理论,而忽视了通过实践让学生主动掌握知识这条重要的学习途径。很多学生感觉数据库理论十分抽象,主要原因一方面是对数据结构等基本理论的学习不够清晰,更重要的是忽视了数据库的操作实践,特别是对数据库实现原理的实验性剖析。当你要学习和研究一个系统时,最有效的办法是把自己置于原创者的位置,从设计的角度提出、思考和解决问题,这样会更敏锐和直接地接触所涉及知识的本质,从而专注于系统实现的原理,而非被动地接受和理解系统封装后的功能。本书一个重要的宗旨就是带读者进入到数据库内部去看看,将读者置于数据库设计者的位置进行研究性的学习。

3. 本书特色

本书将数据库原理及与设计实现关系最为密切的理论知识精心抽取出来,重点讲解并辅以实验代码,以尽可能降低学习难度。而增加实验内容,目的是为了读者更快形成对数据库总体架构的直观认识,并能专注于核心概念的理解。注重数据库原理与数据库实现技术的结合,在本书的第三部分专门安排了与数据库管理系统(DBMS)设计相关的内容,特别是精心准备了一个典型的精简型 DBMS 的源代码,作为开放资源供读者研究和实验。跟随这个实验性系统的实现过程,读者将能够真正了解数据库系统内部的实现机理。这一收获对于数据库应用开发者也是有益的,带来的直接好处是能够洞察数据库应用系统开发中不同实现方法在技术本质上的区别,更容易设计出高效、稳固的系统。例如,在探究了数据存储管理、索引、查询优化等一系列技术实现细节之后,您将有能力在对海量数据进行复杂查询时,找到快速获得结果的最佳方案。

4. 使用本书的建议

本书内容兼顾数据库初学者和有一定开发经验的读者。为了让所有读者都能在这本书中学有所获并享受学习的乐趣,对于本书的使用有以下建议:

- 对于课程教学,建议从最基本的数据库理论讲起,包括数据模型和数据库体系结构等,这部分内容篇幅不大,即使以前学过,再复习一遍也不会感到乏味并能有所收获。具体安排上,本科生侧重第一、二部分,研究生侧重第二、三部分。程序设计爱好者可从头开始研读,并可在任何一章停止,都会有所收获。其中带有星号的部分章节为选修内容,供读者拓展阅读。
- 本书编写力求严谨,每个术语的使用都经过认真思考,希望读者在进行理论学习时也能秉承严谨作风,从细节入手,深入研究。
- 本书的实验用 DBMS 选用我们自行开发的畅想数据库管理系统(CXDB)。CXDB 是一个小型、内嵌式的数据库管理系统,小巧易用,免安装维护,特别适合数据库的入门学习。CXDB 的 SQL 语句的语法与 Oracle 极其相似但有所简化,执行结果也与 Oracle 几乎完全相同。CXDB 的管理工具 CXDBManager 是绿色软件,免安

装，用它来做演示和实验都很方便。即使对于有数据库使用和开发经验的读者，CXDB 也是比较新颖的内容，不会产生学习的枯燥感。当然读者也可以选用任何一个主流的 DBMS 产品作为实验环境。无论如何，强烈建议读者在实际的数据库环境中完成实验操作，不仅包括通用的数据库原理实验，还包括作者精心设计的、与 CXDB 有关的研究和实验。

- 对于本书所列参考资料，建议读者在系统学习本书的同时随时查阅，对有精力和兴趣的读者，建议有选择地对所列书目进行深入阅读。
- 特殊说明部分

书中灰色背景的部分为特殊说明，其内容一般是解释性的。数据库相关的理论比较抽象，作者试图用解释性的语言进一步说明问题，以方便读者理解。其中有的类比不一定贴切，有些内容仅代表作者本人的观点，供读者参考。另外，特殊说明部分中的有些内容是综合论述，不一定限于所在章节，涉及的某些概念可能会在后续章节中深入讨论。

作者

2019年2月

图书资源支持

感谢您一直以来对清华版图书的支持和爱护。为了配合本书的使用,本书提供配套的资源,有需求的读者请扫描下方的“书圈”微信公众号二维码,在图书专区下载,也可以拨打电话或发送电子邮件咨询。

如果您在使用本书的过程中遇到了什么问题,或者有相关图书出版计划,也请您发邮件告诉我们,以便我们更好地为您服务。

我们的联系方式:

地 址: 北京市海淀区双清路学研大厦 A 座 701

邮 编: 100084

电 话: 010-62770175-4608

资源下载: <http://www.tup.com.cn>

客服邮箱: tupjsj@vip.163.com

QQ: 2301891038 (请写明您的单位和姓名)

资源下载、样书申请



书圈



扫一扫, 获取最新目录

用微信扫一扫右边的二维码,即可关注清华大学出版社公众号“书圈”。

目 录

CONTENTS

第一部分 数据库原理

第 1 章 数据库基础知识	3
1.1 数据管理技术的发展	3
1.1.1 数据和数据处理	3
1.1.2 数据管理的三个阶段	4
1.2 数据库的基本概念	5
1.2.1 数据库系统	5
1.2.2 数据模型	8
1.2.3 数据库的发展	9
1.2.4 数据库分类	11
1.3 DBMS 体系结构	12
习题 1	13
第 2 章 关系数据模型	15
2.1 关系模型基础	15
2.1.1 基本术语	15
2.1.2 关系的数学定义	17
2.2 关系代数	19
2.2.1 传统的集合运算	20
2.2.2 专门的关系运算	22
2.2.3* 关系演算	24
2.3* 基于包的关系代数	25
2.3.1 传统的集合运算	25
2.3.2 专门的关系运算	27

2.4	关系的三类完整性约束	27
2.4.1	实体完整性约束	28
2.4.2	参照完整性约束	28
2.4.3	用户自定义的完整性约束	29
实验	畅想数据库管理系统	29
习题 2	32
第 3 章	结构化查询语言(SQL)	35
3.1	SQL 概述	35
3.1.1	SQL 的标准和特点	35
3.1.2	SQL 的术语	36
3.1.3	RDBMS 的体系结构	37
3.2	数据定义	38
3.2.1	数据类型	38
3.2.2	定义基表	40
3.2.3	修改与删除基表	44
3.3	基本数据查询	45
3.3.1	SELECT 语句的语法	45
3.3.2	单表查询	45
3.3.3	分组统计查询	49
3.4	连接查询	51
3.4.1	交叉连接	51
3.4.2	内连接	52
3.4.3	外连接	53
3.4.4	多表连接查询	56
3.5	数据更新	57
3.5.1	数据插入	57
3.5.2	数据修改	58
3.5.3	数据删除	59
3.6*	复杂的查询条件	60
3.6.1	查询条件表达式的语法	60
3.6.2	空值的判断	62
3.6.3	针对字符数据的查询	62
实验	SQL 实验	64
习题 3	67
第 4 章	数据库对象	69
4.1	视图	69

4.1.1	视图的概念	69
4.1.2	创建和删除视图	70
4.1.3	修改视图	71
4.1.4	使用视图数据	71
4.2	索引	72
4.2.1	索引的概念	72
4.2.2	使用索引	73
4.3	序列	73
实验	使用数据库对象	75
4.4*	存储过程	75
4.5*	触发器	77
习题 4	78
第 5 章	数据的安全性	79
5.1	事务和锁的概念	79
5.1.1	事务的概念	79
5.1.2	事务的 ACID 特性	80
5.1.3	事务生命周期	82
5.1.4	锁	83
5.2	使用事务	83
5.2.1	用户、连接和事务的关系	83
5.2.2	显式模式	84
5.2.3	隐性模式	85
5.2.4	自动提交模式	86
5.3*	特殊事务模型	86
5.3.1	嵌套事务	86
5.3.2	保存点	87
5.3.3	分布式事务	87
实验	事务的并发与隔离	88
5.4	SQL 中的授权机制	92
5.4.1	用户	92
5.4.2	权限	92
5.4.3	授权	93
5.4.4	解除授权	93
习题 5	94

第二部分 数据库设计与开发

第 6 章 关系模式的规范化	97
6.1 问题提出	97
6.2 函数依赖和模式分解	99
6.2.1 函数依赖	99
6.2.2 模式分解	101
6.3 关系范式	102
6.3.1 第一范式	104
6.3.2 第二范式	105
6.3.3 第三范式	107
6.3.4 非规范化	108
6.4* Boyce-Codd 范式(BCNF)	109
实验 规范化设计案例	111
习题 6	111
第 7 章 数据库设计与建模	113
7.1 数据库设计的工作	113
7.2 数据库系统的设计方法及步骤	116
7.3 概念模型设计与 ER 模型	118
7.3.1 基本 ER 元素	118
7.3.2 高级 ER 元素	123
7.3.3 建立 ER 模型	125
7.4 逻辑模式设计	128
7.4.1 概念模型向关系模式的转换	129
7.4.2 继续转换为逻辑模式	134
7.5* 选择索引	135
7.5.1 单个索引的收益	135
7.5.2 主键上的索引	136
7.5.3 索引的选择	136
实验 数据库建模实验	137
习题 7	137
第 8 章 数据库应用程序开发	139
8.1 数据库系统的体系结构	139
8.1.1 数据库的访问	139

8.1.2 几种常见的体系结构	140
8.1.3 三层体系结构	142
8.2 应用程序接口	142
8.2.1 ODBC	143
8.2.2 JDBC	143
实验 使用 ODBC 连接数据库	145
习题 8	146

第三部分 DBMS 的实现

第 9 章 DBMS 的设计	149
9.1 需求分析	149
9.1.1 关于数据量	150
9.1.2 关于并发度	151
9.1.3 CXDB 教学版简介	152
9.2 接口设计	153
9.2.1 数据接口设计	153
9.2.2 语言接口设计	153
9.2.3 编程接口设计	154
9.3 总体设计	155
9.3.1 总体设计的任务	155
9.3.2 DBMS 的一般体系结构	156
9.3.3 CXDB-E 的体系结构	158
实验 基于 CXDB-E 的编程	160
习题 9	161
第 10 章 存储管理	163
10.1 数据存储技术概述	163
10.1.1 存储器的层次	163
10.1.2 磁盘的读写	164
10.1.3 磁盘调度	165
10.1.4 磁盘缓存	166
10.1.5 提高磁盘 I/O 速度的方法	166
10.1.6 RAID	167
10.2 数据在文件中的组织	168
10.2.1 CXDB-E 文件格式	168
10.2.2 CXDB-E 的文件头格式	169

10.3 页的寻址	170
实验 CXDB-E 的存储管理	171
习题 10	171
第 11 章 数据缓冲	173
11.1 数据缓冲层概述	173
11.1.1 缓冲区的概念	173
11.1.2 散列表	174
11.2 缓冲区管理	175
11.2.1 文件页和缓冲区页	175
11.2.2 缓冲区管理的功能	177
11.2.3 缓冲区的大小	178
11.3 缓冲区替换策略	179
实验 CXDB-E 的数据缓冲	181
习题 11	181
第 12 章 数据操纵	183
12.1 记录的存储技术	183
12.1.1 记录在页中的存放	183
12.1.2 记录的寻址	185
12.1.3 溢出页的使用	185
12.2 CXDB-E 的数据页格式	186
12.2.1 数据记录格式	186
12.2.2 数据页格式	187
12.2.3* 溢出页格式	188
12.2.4* 空闲页格式	189
12.3 游标	190
12.3.1 游标种类	190
12.3.2 游标操作	190
12.3.3 游标结构	191
实验 CXDB-E 的记录操纵	192
习题 12	193
第 13 章 索引技术	195
13.1 索引技术概述	195
13.2 B 树索引	197
13.2.1 B 树的概念	197
13.2.2 B 树的基本算法	198

13.2.3*	B 树基本算法的讨论	202
13.2.4	B+ 树	203
13.3	CXDB-E 的索引页格式	204
13.4*	用于索引的 B 树算法	205
13.4.1	需要考虑的问题	205
13.4.2	算法	206
实验	CXDB-E 的索引操纵	213
13.5	其他索引技术	214
13.5.1	散列索引	214
13.5.2	位图索引	215
13.5.3	多维索引	215
习题 13	216
第 14 章	查询处理	217
14.1	查询处理的一般过程	217
14.1.1	编译	218
14.1.2	优化	219
14.1.3	执行	220
14.2	物理查询计划操作符	220
14.2.1	关系的扫描	221
14.2.2	物理操作符的代价计算模型	221
14.2.3	衡量代价的参数	221
14.2.4	两阶段多路归并排序	222
14.2.5	物理操作符迭代器	224
14.2.6	算法分类	225
14.3	选择操作	226
14.4	连接操作	228
14.4.1	一趟算法	228
14.4.2	嵌套循环连接	228
14.4.3	基于排序的两趟算法	229
14.4.4	基于散列的两趟算法	231
14.4.5	基于索引的算法	232
14.5	聚合操作	233
14.5.1	一趟算法	234
14.5.2	基于排序的两趟算法	235
14.5.3	基于散列的两趟算法	236
14.5.4	基于索引的算法	236
14.6*	关系代数的等价性定律	236

14.6.1	交换律和结合律	236
14.6.2	与选择有关的定律	237
14.6.3	与投影有关的定律	238
14.6.4	与连接和积有关的定律	238
14.6.5	与分组和聚合有关的定律	238
14.7	选择逻辑查询计划	238
14.7.1	将分析树转换为关系代数表达式树	239
14.7.2	逻辑查询计划的优化	240
14.8	选择物理查询计划	241
习题 14		242
第 15 章 CXDB-E 的 SQL 层		245
15.1	SQL 语句的编译	245
15.1.1	编译的一般步骤	246
15.1.2	lex 与 yacc 介绍	246
15.1.3	SQL 语句的语法描述	248
15.1.4	SQL 语句的内存结构	249
15.2	CXDB-E 的查询处理	250
15.2.1	语义分析	251
15.2.2	选择查询计划	252
15.2.3	查询执行	256
15.2.4	条件表达式向索引的转换	259
实验	SQL 层的实现	259
习题 15		260
第 16 章 事务管理		261
16.1	事务调度	261
16.1.1	串行调度和并行调度	262
16.1.2	可串行化	265
16.1.3	冲突可串行化	266
16.2	事务隔离级别	267
16.2.1	并行调度造成的不一致现象	267
16.2.2	事务隔离级别	269
16.3	并发控制	270
16.3.1	并发控制的分类	270
16.3.2	基于锁的并发控制	270
16.4	锁的类型	271
16.4.1	共享锁与排他锁	271

- 16.4.2 锁的相容矩阵..... 271
- 16.4.3 锁升级..... 272
- 16.4.4 更新锁..... 273
- 16.4.5 意向锁..... 274
- 16.5 封锁粒度和封锁协议..... 275
 - 16.5.1 锁的粒度..... 275
 - 16.5.2 两段锁协议..... 276
- 16.6 加锁产生的问题..... 279
 - 16.6.1 阻塞..... 279
 - 16.6.2 死锁..... 280
 - 16.6.3 死锁的预防..... 280
 - 16.6.4 死锁检测与解除..... 281
 - 16.6.5 活锁..... 282
- 16.7* CXDB 加锁的实现 282
 - 16.7.1 CXDB 加锁的粒度..... 282
 - 16.7.2 CXDB 表级锁的类型..... 283
 - 16.7.3 CXDB 的死锁检测与解决..... 285
- 实验 死锁..... 285
- 习题 16 287

第 17 章 故障恢复 289

- 17.1 故障恢复概述..... 289
 - 17.1.1 故障类型..... 289
 - 17.1.2 故障恢复基础..... 290
- 17.2 Undo 日志 292
 - 17.2.1 日志的概念..... 292
 - 17.2.2 日志记录..... 293
 - 17.2.3 使用 Undo 日志的恢复..... 294
 - 17.2.4 检查点..... 295
 - 17.2.5 模糊检查点..... 296
- 17.3 Redo 日志 297
 - 17.3.1 使用 Redo 日志的恢复 297
 - 17.3.2 Redo 日志的检查点 298
 - 17.3.3 使用带检查点的 Redo 日志的恢复 299
- 17.4 Undo/Redo 日志 299
 - 17.4.1 使用 Undo/Redo 日志的恢复 299
 - 17.4.2 Undo/Redo 日志的检查点 300
- 17.5 防备介质故障..... 301

17.5.1	备份类型	301
17.5.2	使用备份和日志的恢复	302
习题 17		302
第 18 章	数据库技术的发展	305
18.1	数据仓库	306
18.1.1	数据仓库的发展与定义	306
18.1.2	数据仓库的特点	307
18.1.3	数据仓库的组成	308
18.1.4	新一代数据仓库	309
18.2	XML 和 XQuery	310
18.2.1	半结构化数据	311
18.2.2	XML	311
18.2.3	XPath	313
18.2.4	XQuery	314
附录 A	人员数据库设计	319
A.1	表定义和表数据	319
A.2	创建语句	321
附录 B	图书借阅数据库设计	323
B.1	规章制度	323
B.2	与藏书有关的需求	324
B.3	数据库设计	325
参考文献		333