

“十一五”国家重点图书

计算机科学与技术学科前沿丛书

计算机科学与技术学科研究生系列教材（中文版）

数据库原理与实现

高屹 高岩 李雷 著



清华大学出版社



“十一五”国家重点图书 计算机科学与技术学科研究生系列教材（中文版）

数据库原理与实现

高屹 高岩 李雷 著



清华大学出版社

内 容 简 介

本书重点介绍了数据库的原理、设计与实现技术。本书对传统的数据库理论进行了精练,保留核心与实用部分,并适当增加实验内容,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,该部分以一个实际的 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内容。

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

数据库原理与实现/高屹,高岩,李雷著. —北京:清华大学出版社,2013.5

计算机科学与技术学科研究生系列教材(中文版)

ISBN 978-7-302-31286-4

I. ①数… II. ①高… ②高… ③李… III. ①数据库系统—研究生—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2013)第 008167 号

责任编辑:袁勤勇 顾 冰

封面设计:傅瑞学

责任校对:李建庄

责任印制:何 芊

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:三河市李旗庄少明印装厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:19.75

字 数:502 千字

版 次:2013 年 5 月第 1 版

印 次:2013 年 5 月第 1 次印刷

印 数:1~3000

定 价:29.00 元

产品编号:045519-01

前言

数据库是计算机专业公认比较枯燥难学的一门课程,但又是非常重要和必须学好的基础课。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培养。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精髓要义,并在此基础上建立一个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。希望本书的编写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手段。

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。对数据库开发人员来说,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 架构迅速普及的今天,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,成功的因素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,更推崇在理论知识深入理解基础上的应用开发,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,我们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读者主动掌握知识这条重要的学习途径。很多读者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,最深刻的办法是把自己置于原创者的角度,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角度来研究性地学习。

3. 本书特色

本书将数据库原理及设计实现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;并增加了实验内容,目的是为了让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3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系统的源代码,作为开放资源供读者研究实验。跟随这个实验系统的实现过程,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,比如,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,读者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。

本书主要作者在数据结构、数据库原理和C语言方面有着20年的教学经验,编写过多本教材,并长期从事信息系统与网络应用的开发工作,属实战型学者。在编写本书过程中,将作者在多年教学和开发中的积累进行了总结,将自行开发的“畅想数据库管理系统(CXDB)”的丰富资源作为剖析研究之用,真诚希望读者通过本书的阅读,特别是对CXDB的借鉴研究,不但在数据库原理的掌握上有满意的斩获,还能够感受到作者在数据库领域教学和科研实践上的那份感性与执著,正是这一点铸成了作者对数据库教学及开发工作的深刻认识和独特视角,也使得这本书在严苛的原理阐述之中被注入了鲜活的灵魂,让数据库的学习变得不再枯燥。因此请大家带着研究的冲动去剖析和分享CXDB带来的数据库设计思想,体会快乐学习。相信通过本书的研读,不论是初学者还是有一定数据库开发经验的人士,都能对数据库理论知识和设计思想产生新的认识,甚或重拾曾经被自己遗落的关键信息。

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。为了所有读者都能在这本书中学有所获并享受学习的乐趣,在本书的使用上有以下建议:

- 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1、2部分,研究生侧重第2、3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。其中部分带有星号的章节为选修内容,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

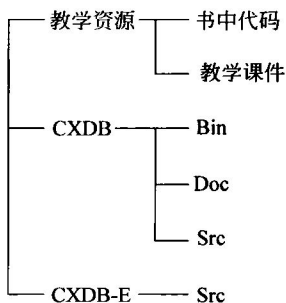
门学习。CXDB 的 SQL 语句语法与 Oracle 极其相似但有所简化,执行结果也与 Oracle 几乎完全相同。CXDB 的管理工具 CXDBManager 是绿色软件,免安装,用它来做演示和实验都很方便。即使对于有数据库使用和开发经验的读者,CXDB 也是比较新颖的内容,不会产生学习的枯燥感。当然读者也可以选用任何一个主流的 DBMS 产品作为实验环境,无论如何,强烈建议读者在实际的数据库环境中完成实验操作。不仅是通用的数据库原理实验,还包括作者精心设计的与 CXDB 有关的研究和实验。

- 本书所列参考资料,建议读者在系统学习本书的同时随时参阅,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说明。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,请读者批判接受。另外,特殊说明部分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。

5. 关于附配资源

本书提供丰富的附配资源,既方便学习,也为更深入的研究提供帮助。附配资源的目录结构如下所示。



附配资源的详细内容在根目录的 ReadMe.txt 文件中说明。本书中需要用到附配资源时,用❖图标引导的文字指明在附配资源中的具体位置。

需要特别说明的是,在“教学资源”子目录下的文档“数据库原理与实现-参考资料.doc”,包含了本书的主要参考资料,包括附加的例题、对正文内容进一步(即更具体、更深入)的说明、实验结果和部分习题参考答案等。该文档采用与本书正文相同的目录,便于读者参阅。

在本书出版后,随着对示例系统的改进,根据读者的反馈,可能随时更新附配资源内容,请关注出版社网站(<http://www.tup.com.cn>)。也可到 <http://b1383571.xici.net> 下载,或直接与作者交流,作者的邮箱为:njgaoyi@yahoo.com.cn。

6. 致谢

首先要感谢清华大学出版社的广大员工,他们为本书的出版付出了卓有成效的劳动,坚

定了我们编写一部不一样的数据库教材的信心。

感谢陈彦德、蒋若江、张娟和冯径,他们允许在其主管的项目中首先采用了畅想数据库,使它经历了实践的检验。

感谢齐东元和田明,他们首先在教学中引入了畅想数据库,并对系统的改进和本书的编写提出了宝贵的意见。

还要感谢费康群,我会把对她的感谢永远珍藏心中。

作 者

2013 年 3 月

目 录

第 1 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	3
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	7
1.2.3 数据库的发展	8
1.2.4 数据库分类	10
1.3 DBMS 体系结构	11
习题 1	12
第 2 章 关系数据模型	13
2.1 关系模型基础	13
2.1.1 基本术语	13
2.1.2 关系的数学定义	15
2.2 关系代数	17
2.2.1 传统的集合运算	17
2.2.2 专门的关系运算	18
*2.2.3 关系演算	21
*2.3 基于包的关系代数	22
2.3.1 传统的集合运算	22
2.3.2 专门的关系运算	23
2.4 关系的三类完整性约束	24
2.4.1 实体完整性约束	24
2.4.2 参照完整性约束	25
2.4.3 用户自定义的完整性约束	26
实验 畅想数据库管理系统	26



习题 2	28
第 3 章 结构化查询语言 SQL	30
3.1 SQL 概述	30
3.1.1 SQL 的标准和特点	30
3.1.2 SQL 的术语	31
3.1.3 RDBMS 的体系结构	32
3.2 数据定义	32
3.2.1 数据类型	33
3.2.2 定义基表	34
3.2.3 修改与删除基表	38
3.3 基本数据查询	39
3.3.1 SELECT 语句的语法	39
3.3.2 单表查询	40
3.3.3 分组统计查询	43
3.4 连接查询	45
3.4.1 交叉连接	45
3.4.2 内连接	46
3.4.3 外连接	47
3.4.4 多表连接查询	49
3.5 数据更新	51
3.5.1 数据插入	51
3.5.2 数据修改	52
3.5.3 数据删除	52
*3.6 复杂的查询条件	53
3.6.1 查询条件表达式的语法	53
3.6.2 空值的判断	55
3.6.3 针对字符数据的查询	56
实验 SQL 实验	58
习题 3	60
第 4 章 数据库对象	62
4.1 视图	62
4.1.1 视图的概念	62
4.1.2 创建、删除视图	63
4.1.3 修改视图	64
4.1.4 使用视图数据	64
4.2 索引	64
4.2.1 索引的概念	64

4.2.2 使用索引	65
4.3 序列	66
实验 使用数据库对象	68
*4.4 存储过程	68
*4.5 触发器	69
习题 4	70
第 5 章 数据的安全性	72
5.1 事务和锁的概念	72
5.1.1 事务的概念	72
5.1.2 事务的 ACID 特性	73
5.1.3 事务生命周期	74
5.1.4 锁	75
5.2 使用事务	76
5.2.1 用户、连接和事务的关系	76
5.2.2 显式模式	77
5.2.3 隐性模式	77
5.2.4 自动提交模式	78
*5.3 特殊事务模型	79
5.3.1 嵌套事务	79
5.3.2 保存点	79
5.3.3 分布式事务	80
实验 事务的并发与隔离	80
5.4 SQL 中的授权机制	84
5.4.1 用户	84
5.4.2 权限	84
5.4.3 授权	85
5.4.4 解除授权	85
习题 5	86

第 2 部分 数据库设计与开发

第 6 章 关系模式的规范化	89
6.1 问题提出	89
6.2 函数依赖和模式分解	91
6.2.1 函数依赖	91
6.2.2 模式分解	92
6.3 关系范式	93
6.3.1 第一范式	95



6.3.2 第二范式	96
6.3.3 第三范式	97
6.3.4 非规范化	99
*6.4 Boyce-Codd 范式(BCNF)	100
实验 规范化设计案例	101
习题 6	102
第 7 章 数据库设计与建模	103
7.1 数据库设计的工作	103
7.2 数据库系统的设计方法及步骤	106
7.3 概念模型设计与 ER 模型	107
7.3.1 基本 ER 元素	108
7.3.2 高级 ER 元素	112
7.3.3 建立 ER 模型	114
7.4 逻辑模式设计	117
7.4.1 概念模型向关系模式的转化	117
7.4.2 继续转化为逻辑模式	122
*7.5 选择索引	123
7.5.1 单个索引的收益	124
7.5.2 主键上的索引	125
7.5.3 索引的选择	125
实验 数据库建模实验	125
习题 7	126
第 8 章 数据库应用程序开发	127
8.1 数据库系统的体系结构	127
8.1.1 数据库的访问	127
8.1.2 几种常见的体系结构	128
8.1.3 三层体系结构	129
8.2 应用程序接口	130
8.2.1 ODBC	130
8.2.2 JDBC	130
实验 使用 ODBC 连接数据库	132
习题 8	133

第 3 部分 DBMS 的实现

第 9 章 DBMS 的设计	137
9.1 需求分析	137

9.1.1	关于数据量	138
9.1.2	关于并发度	139
9.1.3	CXDB 教学版简介	139
9.2	接口设计	140
9.2.1	数据接口设计	141
9.2.2	语言接口设计	141
9.2.3	编程接口设计	142
9.3	总体设计	143
9.3.1	总体设计的任务	143
9.3.2	DBMS 的一般体系结构	144
9.3.3	CXDB-E 的体系结构	145
实验	基于 CXDB-E 的编程	147
习题 9		148
第 10 章	存储管理	149
10.1	数据存储技术概述	149
10.1.1	存储器的层次	149
10.1.2	磁盘的读写	150
10.1.3	磁盘调度	151
10.1.4	磁盘缓存(Disk Cache)	151
10.1.5	提高磁盘 I/O 速度的方法	152
10.1.6	RAID	152
10.2	数据在文件中的组织	153
10.2.1	CXDB-E 文件格式	154
10.2.2	CXDB-E 的文件头格式	154
10.3	页的寻址	155
实验	CXDB-E 的存储管理	156
习题 10		157
第 11 章	数据缓冲	158
11.1	数据缓冲层概述	158
11.1.1	缓冲区的概念	158
11.1.2	散列表	159
11.2	缓冲区管理	160
11.2.1	文件页和缓冲区页	160
11.2.2	缓冲区管理的功能	162
11.2.3	缓冲区的大小	163
11.3	缓冲区替换策略	163
实验	CXDB-E 的数据缓冲	165



习题 11	165
第 12 章 数据操纵	166
12.1 记录的存储技术.....	166
12.1.1 记录在页中的存放.....	166
12.1.2 记录的寻址.....	167
12.1.3 溢出页的使用.....	168
12.2 CXDB-E 的数据页格式.....	168
12.2.1 数据记录格式.....	169
12.2.2 数据页格式.....	170
* 12.2.3 溢出页格式.....	171
* 12.2.4 空闲页格式.....	171
12.3 游标.....	172
12.3.1 游标种类.....	172
12.3.2 游标操作.....	173
12.3.3 游标结构.....	174
实验 CXDB-E 的记录操纵.....	174
习题 12	175
第 13 章 索引技术	176
13.1 索引技术概述.....	176
13.2 B 树索引.....	178
13.2.1 B 树的概念.....	178
13.2.2 B 树的基本算法.....	179
* 13.2.3 B 树基本算法的讨论.....	182
13.2.4 B+ 树.....	182
13.3 CXDB-E 的索引页格式.....	183
* 13.4 用于索引的 B 树算法.....	184
13.4.1 需要考虑的问题.....	185
13.4.2 算法.....	186
实验 CXDB-E 的索引操纵.....	192
13.5 其他索引技术.....	193
13.5.1 散列表索引.....	193
13.5.2 位图索引.....	193
13.5.3 多维索引.....	194
习题 13	194
第 14 章 查询处理	195
14.1 查询处理的一般过程.....	195

14.1.1	编译	195
14.1.2	优化	197
14.1.3	执行	198
14.2	物理查询计划操作符	198
14.2.1	关系的扫描	198
14.2.2	物理操作符的代价计算模型	199
14.2.3	衡量代价的参数	199
14.2.4	两阶段多路归并排序	200
14.2.5	物理操作符迭代器	201
14.2.6	算法分类	202
14.3	选择操作	203
14.4	连接操作	205
14.4.1	一趟算法	205
14.4.2	嵌套循环连接	205
14.4.3	基于排序的两趟算法	206
14.4.4	基于散列的两趟算法	208
14.4.5	基于索引的算法	209
14.5	聚合操作	210
14.5.1	一趟算法	210
14.5.2	基于排序的两趟算法	212
14.5.3	基于散列的两趟算法	212
14.5.4	基于索引的算法	213
*14.6	关系代数的等价性定律	213
14.6.1	交换律和结合律	213
14.6.2	与选择有关的定律	213
14.6.3	与投影有关的定律	214
14.6.4	与连接和积有关的定律	215
14.6.5	与分组和聚合有关的定律	215
14.7	选择逻辑查询计划	215
14.7.1	将分析树转换为关系代数表达式树	215
14.7.2	逻辑查询计划的优化	217
14.8	选择物理查询计划	217
习题 14		219
* 第 15 章	CXDB-E 的 SQL 层	221
15.1	SQL 语句的编译	221
15.1.1	编译的一般步骤	221
15.1.2	lex 与 yacc 介绍	222
15.1.3	SQL 语句的语法描述	223

15.1.4	SQL 语句的内存结构	225
15.2	CXDB-E 的查询处理	226
15.2.1	语义分析	227
15.2.2	选择查询计划	228
15.2.3	查询执行	232
15.2.4	条件表达式向索引的转化	234
实验	SQL 层的实现	235
习题 15		235
第 16 章	事务管理	236
16.1	事务调度	236
16.1.1	串行调度和并行调度	237
16.1.2	可串行化	238
16.1.3	冲突可串行化	239
16.2	事务隔离级别	240
16.2.1	并行调度造成的一致现象	240
16.2.2	事务隔离级别	242
16.3	并发控制	243
16.3.1	并发控制的分类	243
16.3.2	基于锁的并发控制	244
16.4	锁的类型	244
16.4.1	共享锁与排他锁	245
16.4.2	锁的相容矩阵	245
16.4.3	锁升级	246
16.4.4	更新锁	246
16.4.5	意向锁	247
16.5	封锁粒度和封锁协议	248
16.5.1	锁的粒度	248
16.5.2	两段锁协议	249
16.6	加锁产生的问题	251
16.6.1	阻塞	252
16.6.2	死锁	252
16.6.3	死锁的预防	252
16.6.4	死锁检测与解除	253
16.6.5	活锁	254
*16.7	CXDB 加锁的实现	254
16.7.1	CXDB 加锁的粒度	254
16.7.2	CXDB 表级锁的类型	255
16.7.3	CXDB 的死锁检测与解决	257



实验 死锁.....	257
习题 16	259
第 17 章 故障恢复	261
17.1 故障恢复概述.....	261
17.1.1 故障类型.....	261
17.1.2 故障恢复.....	262
17.2 Undo 日志	263
17.2.1 日志的概念.....	264
17.2.2 日志记录.....	264
17.2.3 使用 Undo 日志的恢复	265
17.2.4 检查点.....	267
17.2.5 模糊检查点.....	268
17.3 Redo 日志	268
17.3.1 使用 Redo 日志的恢复	269
17.3.2 Redo 日志的检查点	270
17.3.3 使用带检查点的 Redo 日志的恢复	270
17.4 Undo/Redo 日志	270
17.4.1 使用 Undo/Redo 日志的恢复	271
17.4.2 Undo/Redo 日志的检查点	271
17.5 防备介质故障.....	272
17.5.1 备份类型.....	272
17.5.2 使用备份和日志的恢复.....	273
习题 17	274
第 18 章 数据库技术的发展	275
18.1 数据仓库.....	276
18.1.1 数据仓库的发展与定义.....	276
18.1.2 数据仓库的特点.....	277
18.1.3 数据仓库的组成.....	278
18.1.4 新一代数据仓库.....	279
18.2 XML 和 XQuery	280
18.2.1 半结构化数据.....	280
18.2.2 XML	281
18.2.3 XPath	282
18.2.4 XQuery	284
附录 A 人员数据库设计	287
A.1 表定义和表数据	287



- A.2 创建语句 288
- 附录 B 图书借阅数据库设计 291
 - B.1 规章制度 291
 - B.2 与藏书有关的需求 292
 - B.3 数据库设计 293
- 参考文献..... 300