



数据库原理 与应用

//

刘亚军 高莉莎 编著



清华大学出版社

21 世纪高等学校计算机应用技术规划教材

数据库原理与应用

刘亚军 高莉莎 编著

清华大学出版社
北京

内 容 简 介

本书主要介绍数据库原理与应用。全书分为两部分：一部分为数据库原理篇，内容包括概述、数据模型、关系数据库查询语言与存储过程、事务管理与查询优化、数据库的安全和完整性约束及其实现；另一部分为数据库应用篇，内容包括关系数据库设计理论、数据库设计的需求分析、数据库概念设计、数据库逻辑设计、数据库物理设计、数据库的实现、运行和维护、数据库应用示例以及大数据时代数据管理技术。附录给出了第1~11章习题的题解。

本书力求概念清晰、易于理解，并通过大量应用实例进一步讲述基本概念和方法，便于学生学习。本书的另一个特点是在数据库应用篇的数据库概念设计中，采用了UML E-R图进行语义建模，使学生能够掌握较新的建模方法。

本书适合作为大学本科计算机类专业数据库课程的教材，也可供考研学生和从事数据库应用的人士参考。

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

数据库原理与应用/刘亚军,高莉莎编著.--北京:清华大学出版社,2015

21世纪高等学校计算机应用技术规划教材

ISBN 978-7-302-39019-0

I. ①数… II. ①刘… ②高… 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(2015)第 017122 号

责任编辑:文 怡

封面设计:杨 兮

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 刷 者:北京富博印刷有限公司

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:22.5

字 数:562千字

版 次:2015年2月第1版

印 次:2015年2月第1次印刷

印 数:1~2000

定 价:39.00元

产品编号:061881-01

前言

本书是一本面向大学本科生的数据库课程教材。全书分为两个部分,共有 13 章。

第 1 篇是数据库原理篇,其中第 1 章介绍数据库技术的基本概念、数据管理技术的发展过程、数据模型和数据模式以及数据库系统体系结构;第 2 章介绍有代表性的数据模型,重点介绍关系数据模型的结构、约束、操作,另外还介绍了用于数据库设计的 E-R 模型;第 3 章介绍关系数据库查询语言 SQL 以及存储过程;第 4 章介绍事务的恢复、并发控制以及查询优化;第 5 章介绍数据库的安全性和完整性约束及其实现。

第 2 篇是数据库应用篇,其中第 6 章介绍关系数据库设计的理论知识;第 7 章将信息系统的业务流程分析和数据库设计相结合,介绍数据库设计的需求分析;第 8 章介绍数据库的概念设计并使用最新的 UML E-R 数据模型作为建模工具;第 9 章介绍数据库的逻辑设计;第 10 章介绍数据库的物理设计;第 11 章介绍数据库的实现、运行和维护;第 12 章介绍数据库的一些应用示例;第 13 章简单介绍了大数据时代的数据管理技术。第 1~11 章后面都安排了习题供学生练习,附录中给出了各章练习的解答。

本书着重数据库的基本概念、基本原理和基本应用,力图使读者能够对数据库原理和设计有较全面、深入的了解。通过掌握基本原理、概念和技术对今后的工作起到积极作用。

书中所有 SQL 例子都已在 SQL SERVER 2000 环境调试并运行。

本书作者长期从事数据库课程的教学工作并有多年的信息系统开发和数据库设计的经验,但由于水平有限,书中难免有疏漏和不足之处,恳请读者批评指正。

作者

2014 年 12 月于六朝古都

yjliu@seu.edu.cn

目 录

第 1 篇 数据库原理

第 1 章 概述	3
1.1 数据库技术的基本概念	3
1.1.1 数据	3
1.1.2 数据库	4
1.1.3 数据库管理系统	4
1.1.4 数据库系统	5
1.1.5 数据库管理员	5
1.1.6 数据库应用系统	5
1.1.7 数据目录	5
1.1.8 空值	6
1.2 数据管理技术的发展过程	6
1.2.1 人工管理阶段	6
1.2.2 文件管理阶段	7
1.2.3 数据库技术阶段	8
1.3 数据模型和数据模式	9
1.3.1 数据模型	9
1.3.2 数据模式	10
1.4 数据库系统体系结构	11
1.4.1 分布式数据库系统体系结构	11
1.4.2 客户机/服务器结构	11
1.4.3 浏览器/服务器结构	13
1.4.4 混合体系结构	14
本章习题	14
第 2 章 数据模型	15
2.1 层次数据模型	15
2.2 网状数据模型	17
2.3 关系数据模型	18
2.3.1 关系	18

2.3.2	关系的键	19
2.3.3	关系的主属性和非主属性	20
2.3.4	关系的完整性约束	21
2.3.5	关系代数	22
2.3.6	关系演算	29
2.4	传统实体-联系数据模型	30
2.4.1	传统基本 E-R 模型	30
2.4.2	传统扩充 E-R 模型	33
2.5	UML E-R 数据模型	35
2.5.1	UML 基本 E-R 数据模型	35
2.5.2	UML 扩充 E-R 数据模型	37
	本章习题	38
第 3 章	关系数据库标准语言 SQL	40
3.1	SQL 的发展过程	40
3.2	SQL 的组成和基本结构	41
3.3	数据定义语言 DDL	41
3.3.1	基表模式的定义	42
3.3.2	基表模式的修改	44
3.3.3	索引的建立与撤销	46
3.4	查询语言 QL	46
3.4.1	基本 SQL 查询语句	47
3.4.2	较复杂的 SQL 查询语句	52
3.4.3	集合查询	59
3.5	数据操纵语言 DML	60
3.5.1	插入数据	60
3.5.2	修改数据	62
3.5.3	删除数据	63
3.6	视图	64
3.6.1	定义视图	64
3.6.2	查询视图	65
3.6.3	删除视图	65
3.6.4	更新视图	66
3.6.5	视图的作用	66
3.7	数据控制语言 DCL	66
3.7.1	授权	68
3.7.2	收回权限	69
3.7.3	完整性控制	70
3.8	嵌入式 SQL	70

3.8.1 嵌入式 SQL 介绍	70
3.8.2 嵌入式 SQL 的说明部分	71
3.8.3 嵌入式 SQL 的可执行语句	72
3.9 嵌入式 SQL 的实现	74
3.10 动态 SQL	75
3.11 SQL 的存储过程	77
本章习题	78
第 4 章 事务管理和查询优化	79
4.1 事务	79
4.2 系统恢复	81
4.2.1 系统恢复技术	81
4.2.2 故障的类型及恢复的对策	84
4.3 并发控制	85
4.3.1 并发的概念	85
4.3.2 加锁与锁的协议	87
4.3.3 死锁与活锁	89
4.3.4 可串行化调度与两段锁协议	92
4.3.5 多粒度锁	96
4.4 查询优化	98
4.4.1 查询优化的目的	98
4.4.2 代数优化	102
4.4.3 依赖于存取路径的规则优化	105
4.4.4 代价估算优化	108
本章习题	112
第 5 章 数据库的安全和完整性约束	114
5.1 数据库的安全	114
5.1.1 用户标识和鉴别	114
5.1.2 存取控制	115
5.1.3 视图定义和查询修改	116
5.1.4 数据加密	116
5.1.5 审计跟踪	116
5.2 数据库的完整性约束	117
5.2.1 域完整性约束	117
5.2.2 引用完整性约束	117
5.2.3 实体完整性约束	117
5.2.4 其他完整性约束	117
5.2.5 完整性约束的说明	118
本章习题	121

第2篇 数据库应用

第6章 关系数据库设计理论	125
6.1 关系模式设计中的一些语义问题	125
6.2 函数依赖	128
6.2.1 函数依赖的定义	128
6.2.2 函数依赖集的闭包	130
6.2.3 属性集的闭包	132
6.2.4 最小函数依赖	134
6.3 多值依赖	136
6.4 连接依赖	138
6.5 关系模式的分解	140
6.6 无损分解的验证	142
6.7 保持依赖的验证	146
6.8 关系模式的规范化	148
6.8.1 第一范式	148
6.8.2 第二范式	151
6.8.3 第三范式	151
6.8.4 BC 范式	152
6.8.5 无损连接和保持函数依赖分解成 3NF 模式集的算法	153
6.8.6 无损分解成 BCNF 模式的算法	157
6.8.7 第四范式	158
本章习题	159
第7章 数据库设计的需求分析	161
7.1 业务需求的确定	161
7.1.1 业务的流程	161
7.1.2 组成业务的数据	162
7.1.3 数据的处理	162
7.1.4 业务规则	163
7.2 数据需求的确定	163
7.3 处理需求的确定	166
7.4 数据的收集和分析	167
7.4.1 数据收集的方法	167
7.4.2 数据的分析	169
本章习题	170
第8章 数据库的概念设计	171
8.1 概念设计的基本方法	171

8.2	概念设计的数据模型	172
8.3	局部视图的设计	172
8.3.1	确定局部视图的设计范围	172
8.3.2	确定实体及实体的主键	174
8.3.3	定义实体间的联系	175
8.3.4	给实体及联系加上描述属性	189
8.4	高级建模技术	191
8.4.1	特殊化和普遍化	191
8.4.2	超类/子类关系的约束	193
8.5	视图集成概述	193
8.5.1	视图集成的策略	195
8.5.2	视图集成的步骤	196
8.6	实体的集成	197
8.7	联系的集成	202
8.7.1	相同元数、相同角色的联系的集成	202
8.7.2	相同元数、不同角色的联系的集成	204
8.7.3	不同元的联系集成	205
8.8	新老数据模式的集成	208
8.9	实例	209
	本章习题	211
第 9 章	数据库的逻辑设计	213
9.1	E-R 图到关系模式的映射	213
9.1.1	实体到关系的映射	213
9.1.2	基本 E-R 图映射为关系的方法	215
9.1.3	扩充 E-R 数据模式的映射	220
9.2	关系模式的优化	222
9.3	关系模式的调整	227
9.3.1	改善数据库性能	227
9.3.2	节省存储空间的一些考虑	227
9.4	外模式的设计	228
	本章习题	230
第 10 章	数据库的物理设计	232
10.1	确定记录的存储结构	232
10.1.1	数据项的存储技术	232
10.1.2	记录在物理块上的分配	233
10.1.3	物理块在磁盘上的分配	234
10.1.4	数据压缩方法	235
10.2	确定数据库的存储结构	236

10.2.1	确定数据的存放位置	236
10.2.2	选择文件的组织方式	236
10.2.3	确定系统配置	239
10.3	簇集设计	239
10.4	索引的选择	240
10.5	评价物理结构	247
	本章习题	248
第 11 章	数据库的实现、运行与维护	249
11.1	数据库的实现	249
11.2	数据库的运行	251
11.3	数据库的维护	251
	本章习题	255
第 12 章	数据库设计示例	256
12.1	客户订购登记管理	256
12.1.1	需求分析	256
12.1.2	概念设计	257
12.1.3	逻辑设计	260
12.2	学生住宿管理	266
12.2.1	需求分析	266
12.2.2	概念设计	267
12.2.3	逻辑设计	269
12.3	工资管理	272
12.3.1	需求分析	273
12.3.2	概念设计	273
12.3.3	逻辑设计	276
12.4	人力资源管理	279
12.4.1	需求分析	279
12.4.2	概念设计	280
12.4.3	逻辑设计	282
12.5	发票处理	284
12.5.1	需求分析	284
12.5.2	概念设计	285
12.5.3	逻辑设计	286
12.6	保险业务管理	287
12.6.1	需求分析	288
12.6.2	概念设计	289
12.6.3	逻辑设计	291
12.7	车辆租赁管理	293

12.7.1	需求分析	293
12.7.2	概念设计	294
12.7.3	逻辑设计	296
12.8	飞机订票系统	298
12.8.1	需求分析	298
12.8.2	概念设计	299
12.8.3	逻辑设计	300
12.9	酒店客房预订系统	301
12.9.1	需求分析	301
12.9.2	概念设计	302
12.9.3	逻辑设计	304
12.10	学生工作管理	305
12.10.1	需求分析	305
12.10.2	概念设计	306
12.10.3	逻辑设计	308
第 13 章	大数据时代的数据管理技术	310
13.1	关系数据库时代的数据管理技术	310
13.2	大数据时代的数据管理技术	311
13.2.1	传统关系数据库的瓶颈	311
13.2.2	大数据的特点	312
13.2.3	NoSQL 数据库	313
13.2.4	MapReduce	314
13.2.5	Hadoop	315
各章习题参考答案		317
第 1 章		317
第 2 章		318
第 3 章		321
第 4 章		327
第 5 章		330
第 6 章		332
第 7 章		335
第 8 章		336
第 9 章		341
第 10 章		345
第 11 章		347
参考文献		348

第1篇 数据库原理

第 1 章

概 述

近年来,计算机科学技术发展迅速,而数据库技术是计算机科学技术发展最快的领域之一,同时也是应用最广泛的技术之一。在信息管理自动化程度日益提高的今天,数据库技术越来越多地渗透到了人们工作和生活的各个方面。

数据库的基础知识是从事信息产业工作人员和相关专业工作人员的必备知识与技能,同时也是进一步深入研究数据库原理及其应用的出发点。本章主要介绍数据库的基本概念、数据管理技术的发展过程、数据库以及数据库系统体系结构。

1.1 数据库技术的基本概念

计算机在发展的初期只用于复杂的科学计算,后来随着软硬件技术的发展以及字符串处理能力的引入,计算机开始具有了数据处理能力。数据库技术是数据管理的最新技术,也是计算机科学的一个重要分支。数据库是信息系统的核心和基础,数据库技术的出现促进了计算机应用向各行各业的渗透。

1.1.1 数据

数据(data)是数据库中存储和管理的基本对象,是描述事物属性的一种符号记录。数据可分为两大类:一类是能够参与数值运算的数值型数据,如学生成绩、职工工资等;另一类是不能参与数值运算的非数值型数据,如文字、图形、图像、声音等。数据有多种形式,如学生的档案记录、学生的选课情况、学生的照片等,它们都可以经过数字化后存入计算机。

在计算机中,为了存储和处理这些事物,要选择能够描述事物特征的一组数据组成一个记录。

例如:在学生的档案中,如果人们最感兴趣的是学生的姓名、性别、出生年月、籍贯、所在系别、入学时间,那么可以这样描述某个学生的档案记录:

(李明,男,1978-10-25,江苏,计算机系,1990)

该记录表示李明是个大学生,1978年10月25日出生,男,江苏人,1990年考入计算机系。但数据的表现形式并不能完全表达其内容,不了解其语义的人无法理解其含义,必须经过语义解释才能被人理解。语义解释是指对数据含义的说明,数据的含义称为数据的语义。人们通过解释、推论、分析、综合等方法,从数据所获得的有意义的内容称为信息。因此,数据与其语义信息是密不可分的。数据是信息存在的一种形式,只有通过解释或处理才能成为有用的信息。

1.1.2 数据库

数据库(database,DB)是长期存储在计算机内部的逻辑上相关、可共享的数据集合。所谓“逻辑上相关”是指数据库中存储的是数据和数据之间的逻辑关系。数据库中的数据通常按一定的数据模型组织、描述和存储,具有较小的冗余度、较高的数据独立性和易扩展性,并可为各种用户共享。

1.1.3 数据库管理系统

数据库管理系统(database management system,DBMS)是位于用户与操作系统之间的一层数据管理软件,其主要功能包括以下几个方面:

- ① 数据定义功能:用数据描述语言定义模式、外模式和内模式。
- ② 数据操纵功能:用数据操纵语言实现对数据的操作。包括数据的查询、插入、删除和修改。
- ③ 数据库的运行管理功能:对数据库的安全性、完整性、故障恢复和并发操作等方面的管理功能。
- ④ 数据库的建立和维护功能:对数据库数据的初始装载、数据库转储、数据库重组和记录日志文件。

因此,DBMS 是数据库系统的一个重要组成部分。

DBMS 的一般工作原理为:

① 用户编写的应用程序经过接口软件处理后,抽出其中数据库语言语句,转换成一种最基本的数据库语言,交词法和语法器分析,产生相应的语法树。然后进行授权检查,检查用户是否有权访问语法树中所涉及的数据对象。如果授权检查通过,则继续执行;否则返回适当消息,拒绝执行。

② 通过授权检查以后,就可对语法树进行语义分析和处理。对数据定义语句、查询语句、数据操纵语句和数据控制语句分别做不同的处理。其中的查询语句是最复杂和最基本的,这部分功能常统称为查询处理。在查询处理时,还存在多种存取路径的选择问题,这就是所谓查询优化。

③ 经过语义分析和查询处理,就形成了语句的执行计划,并用 DBMS 内部定义的存取原语表示。存取原语是一些基本操作命令,例如打开文件、关闭文件、取一记录、建立索引等。存取原语由存取机制执行。在执行过程中,还须有并发控制,以防止多用户并发访问数据库时引起的数据不一致。数据是重要的资源,任何破坏都会导致严重的后果。但是,再好的系统也会发生故障。在发生故障时,恢复机制能够使数据库恢复到最近的一致状态或先前的某个一致状态。

DBMS 是建立在操作系统之上的软件系统,是操作系统的用户。计算机系统的硬件和各种资源由操作系统统一管理。DBMS 若有分配内存、创建或撤销进程、访问磁盘等要求,必须通过系统调用请求操作系统为其服务。

DBMS 须按查询处理所确定的执行计划对数据进行各种处理,以获得所需的查询结果,并通过接口以一定的格式提供给应用程序或用户。

1.1.4 数据库系统

数据库系统(database system, DBS)是指在计算机系统中引入数据库后的系统构成,一般由数据库、数据库管理系统、应用程序、数据库管理员和用户构成,如图 1.1 所示。

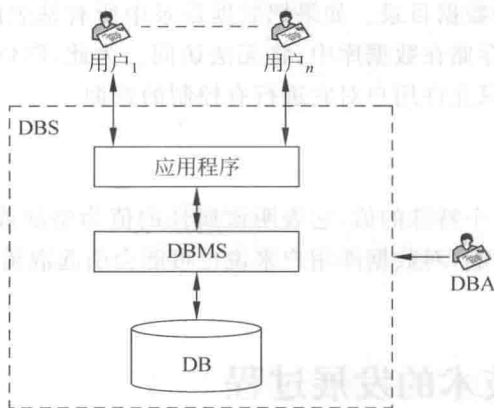


图 1.1 数据库系统的组成

图 1.1 中,用户是指与数据库系统的设计、创建、使用和维护等工作相关的人员;数据库是指应用中实际存储数据的一组关系表;DBMS 是指管理、控制数据库系统和执行数据库操作的系统软件,是 DBS 的重要组成部分;DBA 是指数据库管理员,其职责是负责数据库的规划、设计、协调、维护和管理,保证 DBS 正常运行。

1.1.5 数据库管理员

数据库是一种共享的资源,需要有人进行数据库的规划、设计、协调、维护和管理等工作,负责这些工作的人员或集体称为数据库管理员(database administrator, DBA)。DBA 的具体职责是:

- ① 决定数据库中的信息内容和结构。
- ② 决定数据库的存储结构和存取策略。
- ③ 定义数据的安全性要求和完整性约束条件。
- ④ 监控数据库的使用和运行。

1.1.6 数据库应用系统

数据库应用系统(database application system)是指系统开发人员利用数据库系统资源开发出来的,面向某一类实际应用的应用软件系统。例如,应用于教务部门的学生选课管理系统、应用于人事部门的人事管理系统以及应用于财务部门的财务管理系统等。

1.1.7 数据目录

数据目录(catalog)是一组关于数据的数据,也叫元数据。在高级程序设计语言中,程序所用到的数据由程序中的说明语句定义,程序运行结束了,这些说明也就失效了。DBMS 的任务是管理大量的、共享的、持久的数据。有关这些数据的定义和描述须长期保存在系统

中,一般就把这些元数据组成若干表,称之为数据目录,由系统管理和使用。

数据目录的内容包括基表、视图的定义以及存取路径(索引、散列等)、访问权限和用于查询优化的统计数据等的描述。数据目录只能由系统定义并为系统所有,在初始化时由系统自动生成。数据目录是被频繁访问的数据,同时又是十分重要的数据,几乎 DBMS 的每一部分在运行时都要用到数据目录。如果把数据目录中所有基表的定义全部删去,则数据库中的所有数据,尽管还存储在数据库中,将无法访问。为此,DBMS 一般不允许用户对数据目录进行更新操作,而只允许用户对它进行有控制的查询。

1.1.8 空值

空值在数据库中是一个特殊的值,它表明该属性的值为空缺或未知。空值的存在会对关系代数运算产生一些影响,对数据库用户来说也可能会引起混淆,所以应当尽量避免使用空值。

1.2 数据管理技术的发展过程

数据库技术是应管理任务的需要而产生的。反过来,数据库系统的广泛应用又促进了数据库技术的进一步发展和创新。

在应用需求的推动下,在计算机硬件、软件发展的基础上,数据管理技术经历了人工管理、文件系统、数据库系统三个阶段。

1.2.1 人工管理阶段

20 世纪 50 年代中期以前,计算机主要用于科学计算。当时的硬件状况是外存只有纸带、卡片、磁带,没有磁盘等直接存取的存储设备;软件状况是没有操作系统,没有管理数据的软件;数据处理方式是批处理方式。

这个阶段的人工管理数据具有如下特点:

1. 数据不保存

由于当时计算机主要用于科学计算,一般不需要将数据长期保存,只是在计算某一课题时将数据输入,用完就撤走。不仅对用户数据如此处置,对系统软件有时也是这样。

2. 应用程序管理数据

数据需要由应用程序自己管理,没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构,而且要设计物理结构,包括存储结构、存取方法、输入方式等。因此程序员负担很重。

3. 数据不共享

数据是面向应用的,一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时,由于必须各自定义,无法互相利用、互相参照,程序与程序之间有大量的冗余数据。