

普通高校本科计算机专业特色教材精选·数据库

数据库管理系统概论

徐 述 习胜丰 杨轶芳 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

数据库管理系统概论

徐 述 习胜丰 杨轶芳 主 编
谭新良 何 骞 汪 彦 副主编



清华大学出版社
北京

内 容 简 介

本书系统阐述数据库技术的核心软件——数据库管理系统,详细讲解其基本功能、工作模式、系统结构和实现技术,并对新型数据库管理系统予以介绍和展望,为有兴趣的读者指明研读方向。全书共分为11章:第1章绪论;第2章数据库管理系统的数据组织与存储;第3章DBMS数据定义、操纵与完整性约束;第4章查询处理;第5章查询优化;第6章事务;第7章并发控制;第8章数据库安全;第9章数据库恢复;第10章数据库管理系统性能配置;第11章新型数据库管理系统。

本书可以作为高等学校计算机类专业、信息管理与信息系统等相关专业本科生和研究生“数据库”及相关课程的教材或教学参考书,也可供从事数据库管理系统研究、开发和应用的人员参考。

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

数据库管理系统概论/徐述等主编. —北京:清华大学出版社,2018

(普通高校本科计算机专业特色教材精选·数据库)

ISBN 978-7-302-50571-6

I. ①数… II. ①徐… III. ①数据库管理系统 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2018)第 141831 号

责任编辑:袁勤勇

封面设计:傅瑞学

责任校对:时翠兰

责任印制:李红英

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:三河市少明印务有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:15.5

字 数:350千字

版 次:2018年10月第1版

印 次:2018年10月第1次印刷

定 价:39.00元

产品编号:078766-01

前言

PREFACE

数据库管理系统是对数据进行存储、管理、处理与维护的系统软件，是数据库技术的核心部分。随着计算机硬件、软件技术的迅速发展，数据库技术在各行各业的广泛应用以及大数据时代的到来，数据库管理系统的理论与技术及其发展越来越成为计算机科学与技术教育中必不可少的部分。

本书详细讲述数据库管理系统的基本概念、工作模式、体系结构、功能模块组成以及主流实现技术；并介绍大数据时代下，数据库管理系统发展的新兴成果与方向。

全书分为 11 章，第 1 章对数据库管理系统及其工作模式、系统结构与主要功能进行综述；第 2 章讲述数据库管理系统的数据组织与存储，包括记录格式、文件格式、索引结构与实现技术；第 3 章讲述数据库管理系统的数据操纵与数据完整性约束；第 4 章讲述查询处理，包括查询处理的一般步骤、选择运算实现、排序与连接处理、表达式计算；第 5 章讲述查询优化，包括代数优化与物理优化；第 6 章讲述数据库管理系统运行与管理的基本单位——事务；第 7 章讲述并发控制，包括并发操作问题、并发事务调度的可串行化与可恢复性、并发控制主流技术，重点讨论封锁技术；第 8 章讲述数据库安全性控制，基于数据库系统安全模型与数据库管理系统安全控制模型讨论自主存取控制、审计、强制存取控制、数据加密等 DBMS 安全性措施；第 9 章讲述数据库恢复，包括故障类型、恢复原理、恢复算法、最佳恢复算法 ARIES 以及容灾备份机制等；第 10 章讲述数据库管理系统性能配置；第 11 章介绍新型数据库管理系统，包括面向对象数据库管理系统、对象关系数据库管理系统、XML 数据库管理系统、大数据管理系统。

本书针对数据库管理系统的内容详细丰富，弥补了数据库教学中对 DBMS 原理、实现技术与算法涉及不深的不足。读者可以根据需要阅读或者学习书中部分章节。

全书由徐述组织与执笔。湖南城市学院习胜丰教授、谭新良教授、何骞博士参与编撰和校阅，并提出了许多宝贵意见；杨轶芳老师、汪彦老师给予了英文检索与翻译支持。在此也感谢家人特别是女儿的默默付出与支持！

限于水平，书中难免存在疏漏、欠妥之处，欢迎读者批评指正。

徐 述

2018年3月

目 录

CONTENTS

第 1 章 绪论	1
1.1 数据库管理系统	1
1.1.1 数据库管理系统概述	1
1.1.2 数据库管理系统工作模式	2
1.2 数据库管理系统结构	3
1.2.1 应用层	4
1.2.2 语言处理层	4
1.2.3 存储管理层	4
1.3 语言处理层	5
1.4 存储管理层	7
1.4.1 数据存取	7
1.4.2 缓冲区管理	7
1.4.3 数据存储的物理组织	8
1.5 数据库管理系统基本功能	10
1.6 小结	11
思考题	11
第 2 章 数据库管理系统的数据组织与存储	13
2.1 数据库系统存储结构	13
2.1.1 数据库磁盘存储器中的数据结构	13
2.1.2 数据库系统存储介质	16
2.1.3 存储介质层次结构	18
2.2 数据文件的记录格式	19
2.2.1 定长记录格式	19
2.2.2 变长记录格式	22

2.3 数据文件格式	23
2.3.1 文件格式	23
2.3.2 顺序文件	24
2.3.3 聚集文件	25
2.4 索引技术	26
2.4.1 索引基本概念	26
2.4.2 顺序索引	27
2.4.3 辅助索引	30
2.4.4 索引的更新	31
2.4.5 索引的自动生成	32
2.5 B ⁺ 树索引文件	32
2.5.1 B ⁺ 树结构	33
2.5.2 B ⁺ 树的查询	34
2.5.3 B ⁺ 树的更新	35
2.5.4 B ⁺ 树文件组织	36
2.5.5 B树索引文件	37
2.6 散列索引文件	38
2.6.1 散列技术	38
2.6.2 静态散列索引	40
2.6.3 可扩充散列结构	42
2.7 小结	45
思考题	46
第3章 DBMS 数据定义、操纵与完整性约束	47
3.1 SQL 概述	47
3.1.1 数据定义语言	48
3.1.2 数据操纵语言	48
3.1.3 数据完整性控制	48
3.1.4 数据控制语言	48
3.1.5 事务管理	48
3.1.6 嵌入式 SQL 和动态 SQL	48
3.2 项目工程公司数据库	49
3.3 DBMS 数据定义	52
3.3.1 模式的定义与删除	53
3.3.2 基本表的定义、修改与删除	54
3.3.3 视图建立与删除简介	57
3.3.4 索引的建立、修改与删除	57
3.4 DBMS 数据操纵	58

3.4.1 数据查询	58
3.4.2 数据更新	67
3.4.3 视图	69
3.5 DBMS 完整性约束	74
3.5.1 完整性概述	74
3.5.2 实体完整性	75
3.5.3 参照完整性	77
3.5.4 非空约束	79
3.5.5 唯一约束	79
3.5.6 CHECK 约束	80
3.5.7 完整性约束命名	80
3.5.8 触发器	81
3.6 小结	84
思考题	84
第 4 章 查询处理	87
4.1 概述	87
4.2 查询的选择运算实现	89
4.2.1 使用单文件扫描和索引的选择	89
4.2.2 涉及比较的选择	91
4.2.3 复合条件选择	92
4.3 查询的排序处理	93
4.3.1 外部归并排序算法	93
4.3.2 外部归并排序的代价分析	94
4.4 查询的连接处理	95
4.4.1 嵌套循环算法	95
4.4.2 索引嵌套循环连接	96
4.4.3 归并连接算法	96
4.4.4 散列连接算法	98
4.5 表达式计算	101
4.5.1 物化	101
4.5.2 流水线	101
4.6 小结	102
思考题	103
第 5 章 查询优化	105
5.1 概述	105
5.2 代数优化	106

5.2.1	关系代数表达式等价变换规则	106
5.2.2	基于启发式规则的代数优化	108
5.2.3	代数优化实例	109
5.3	物理优化	112
5.3.1	基于启发式规则的物理优化	112
5.3.2	基于代价估算的物理优化	113
5.4	基于语义的查询优化	113
5.5	小结	114
	思考题	114
第6章	事务	115
6.1	事务的概念	115
6.2	事务的 ACID 性质	116
6.3	一个简单的事务实例	116
6.4	事务抽象模型与状态变迁	118
6.5	SQL 中事务的存取模式和隔离级别	120
6.6	小结	121
	思考题	121
第7章	并发控制	123
7.1	事务的并发执行	123
7.1.1	事务并发执行的必要性	123
7.1.2	事务并发执行趋势	124
7.1.3	并发操作带来的问题	124
7.1.4	并发事务调度可串行化与可恢复性	126
7.1.5	并发控制技术	133
7.2	封锁技术	134
7.2.1	封锁类型	134
7.2.2	封锁协议	135
7.2.3	两段锁协议	138
7.2.4	封锁的实现	141
7.3	封锁带来的问题	142
7.3.1	活锁	142
7.3.2	死锁	142
7.4	多粒度封锁	145
7.4.1	多粒度树	145
7.4.2	意向锁	146
7.4.3	多粒度封锁协议	148

7.5	时间戳技术	148
7.5.1	时间戳	148
7.5.2	时间戳排序协议	149
7.5.3	改进的时间戳协议——Thomas 写规则	150
7.6	多版本机制与快照隔离	151
7.6.1	多版本并发控制	151
7.6.2	多版本两段锁协议	152
7.6.3	快照隔离	153
7.7	幻行现象	155
7.8	小结	157
	思考题	158
第 8 章	数据库安全	159
8.1	数据库安全概述	159
8.1.1	威胁数据库的安全因素	159
8.1.2	数据库安全标准简介	160
8.2	数据库系统安全控制	163
8.2.1	数据库系统安全模型	163
8.2.2	数据库管理系统安全性控制模型	163
8.2.3	用户身份标识与鉴别	164
8.3	存取控制概述	165
8.3.1	自主存取控制	166
8.3.2	强制存取控制	172
8.4	审计	173
8.4.1	审计事件	173
8.4.2	审计的作用	174
8.4.3	AUDIT 语句和 NOAUDIT 语句	174
8.4.4	ORACLE 的审计技术	174
8.5	数据加密	175
8.5.1	加密技术	175
8.5.2	数据库中的加密支持	176
8.6	更高安全性保护	177
8.6.1	推理控制	177
8.6.2	隐蔽信道	178
8.6.3	数据隐私	179
8.7	小结	179
	思考题	180

第9章 数据库恢复	181
9.1 故障类型	181
9.1.1 事务故障	181
9.1.2 系统故障	182
9.1.3 介质故障	182
9.2 恢复机制下的存储器与数据访问	182
9.2.1 存储器种类	182
9.2.2 稳定存储器的实现	183
9.2.3 事务数据访问机制	183
9.3 恢复的基本原理与实现方法	184
9.3.1 恢复与事务原子性	184
9.3.2 日志恢复的基本原则与实现方法	185
9.3.3 影子复制恢复的基本原理	185
9.4 日志恢复技术	186
9.4.1 数据转储	186
9.4.2 日志文件格式	187
9.4.3 日志登记原则	188
9.4.4 使用日志重做和撤销事务	189
9.4.5 检查点	191
9.5 缓冲区管理	192
9.5.1 日志记录缓冲	192
9.5.2 数据库缓冲	193
9.5.3 模糊检查点	194
9.6 恢复算法	194
9.6.1 事务故障恢复——事务回滚	195
9.6.2 系统故障恢复	195
9.6.3 介质故障后的恢复	197
9.7 ARIES 恢复技术	197
9.7.1 ARIES 特点	198
9.7.2 ARIES 数据结构	198
9.7.3 ARIES 恢复算法	200
9.7.4 ARIES 恢复算法特征	202
9.8 容灾备份系统	203
9.9 小结	205
思考题	206

第 10 章 数据库管理系统性能配置	207
10.1 性能配置	207
10.1.1 瓶颈位置	208
10.1.2 硬件调整	208
10.1.3 数据库系统参数调整	210
10.1.4 模式与事务调整	210
10.2 性能基准程序	211
10.2.1 任务集	211
10.2.2 数据库应用类型	212
10.2.3 TPC 基准测试	213
10.3 数据库标准	217
10.3.1 SQL 标准	217
10.3.2 数据库连接标准	218
10.3.3 对象数据库标准	218
10.3.4 XML 标准	218
10.4 小结	219
思考题	219
第 11 章 新型数据库管理系统	221
11.1 数据库管理系统发展的三个阶段	222
11.1.1 第一代数据库管理系统——基于格式化模型 DBMS	222
11.1.2 第二代数据库管理系统——关系 DBMS	222
11.1.3 第三代数据库管理系统——新一代 DBMS	222
11.2 基于新型数据模型的数据库管理系统	223
11.2.1 面向对象数据库管理系统	223
11.2.2 关系对象数据库管理系统	224
11.2.3 XML 数据库管理系统	224
11.3 大数据管理系统	225
11.3.1 大数据	225
11.3.2 大数据建模——基于分析的用户建模	227
11.3.3 大数据管理系统	228
11.4 小结	230
思考题	231
参考文献	232

第 1 章

CHAPTER 1

绪 论

数据库技术是计算机科学技术的重要分支,是迄今为止数据管理的有效技术。设计数据库的目的是方便管理大量信息。作为数据库技术的核心软件——数据库管理系统(Database Management System,DBMS),其主要目的是提供一种方便、高效地存取数据库信息的途径。数据库管理系统对数据的管理涉及信息存储结构、信息操作及信息安全等方面。

在信息资源已经成为重要资源和财富的今天,作为信息系统核心和基础的数据技术已经普及各行各业。从信息管理系统到联机事务处理,从计算机辅助设计与制造到电子政务、电子商务,从地理信息系统到大数据兴起,信息在社会各个方面已经离不开数据库技术与数据库管理系统的支撑。数据库管理系统如何有效地管理与维护数据,主流的数据库管理系统技术有哪些都是本书所关注的重点。

本章介绍数据库管理系统的概念、系统结构、工作原理与基本功能。目前,数据库管理系统大部分为关系型数据库管理系统(Relational Database Management System,RDBMS),如无特殊说明,本书讨论的数据库管理系统均为关系型数据库管理系统。

1.1 数据库管理系统

数据库管理系统是数据库系统的核心,是介于操作系统与用户之间管理数据库的系统软件。

1.1.1 数据库管理系统概述

数据库管理系统是一种操纵和管理数据库的大型系统软件,用于建立、使用和维护数据库。它对数据库进行统一的管理和控制,以保证数据库的安全性和完整性。用户通过数据库管理系统访问数据库中的数据,数据库管理员也通过数据库管理系统进行数据库的维护工作。数据库管理系统可供多个应用程序或用户通过不同的方法同时或分时去建立、修改和

查询数据库。

DBMS 提供数据定义语言(Data Definition Language, DDL)和数据操作语言(Data Manipulation Language, DML),供用户定义数据库的模式结构与权限约束,实现对数据的插入、删除、修改以及检索等操作。

数据库管理系统就是实现把用户层面上抽象的逻辑数据处理转换为计算机中具体的物理数据处理的软件。有了数据库管理系统,用户就可以在抽象意义下处理数据,而不必考虑这些数据在计算机中的布局和物理位置。

1.1.2 数据库管理系统工作模式

作为数据库系统的核心组成部分,数据库系统中管理数据的最重要软件,DBMS 负责对数据库的一切操作,包括定义、查询、更新及各种控制。

数据库管理系统的工作模式如图 1-1 所示。

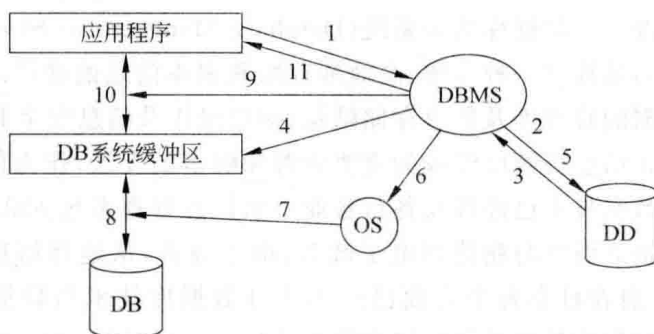


图 1-1 数据库管理系统的工作模式

DBMS 工作模式描述如下。

(1) 数据库管理系统接受用户应用程序提出的数据处理请求。例如,查询某关系的某一行数据。

(2) 数据库管理系统首先对数据请求进行语法检查、语义检查以及用户存取权限检查。具体做法:DBMS 读取数据字典,检查数据请求涉及的关系及相应的字段是否存在,该用户是否有权限操作数据等,确认语义正确与权限合法之后,DBMS 开始执行数据请求命令;否则,拒绝执行,返回错误信息。

(3) 数据库管理系统进行查询优化。优化器根据数据字典中的相关信息进行优化,将数据请求转换成一串单存取操作序列。DBMS 重复以下各步骤逐个执行单存取操作,直到序列结束。

(4) 数据库管理系统在系统缓冲区中查找单存取操作涉及的记录,若找到该记录则转到(9),否则转到(5)。

(5) 若单存取操作涉及的记录不在系统缓冲区中,则数据库管理系统查找内模式,决定到哪个数据文件用什么方式读取相应的物理记录。

(6) 数据库管理系统根据(5)的结果,向操作系统发出读取物理记录的命令。

(7) 操作系统执行读取记录的相关操作。

(8) 操作系统将记录从数据库的存储区送至系统缓冲区。

(9) 数据库管理系统根据结构化查询语言(Structured Query Language, SQL)命令和数据字典的外模式定义,导出用户要读取的记录格式。

(10) 数据库管理系统将转换格式后的记录从系统缓冲区传送到应用程序用户工作区。

(11) 数据库管理系统将单存取操作的执行状态信息返回给应用程序。执行状态信息指读取成功或者不成功的系统提示、例外状态信息等。

数据库管理系统的各层模块紧密配合,相互依赖,共同完成对数据库的操作。其他操作,如插入、删除、修改,与查询操作类似。

1.2 数据库管理系统结构

从系统结构角度来讲,数据库管理系统与操作系统一样,其体系结构是分层的。分层的DBMS体系结构有助于数据库管理系统的设计与维护,也有利于用户更清楚地认识数据库管理系统。

数据库管理系统体系结构从高到低分为应用层、语言处理层和存储管理层,如图1-2所示。

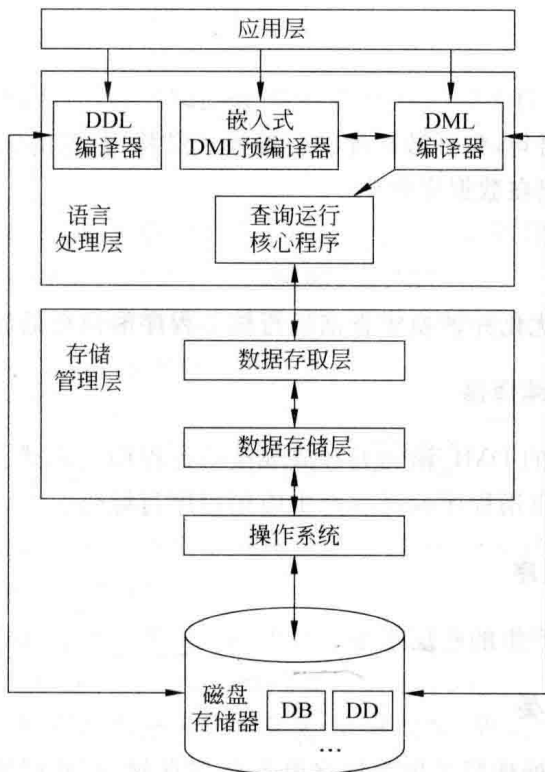


图 1-2 DBMS 体系结构

数据库管理系统之下是操作系统,操作系统为数据库管理系统提供最基本的磁盘读写服务,处理对象是数据文件的物理块。操作系统提供存取原语和基本存取方法作为接

口供数据库管理系统存储管理层调用,保证数据库管理系统对数据逻辑上的读写真实地映射到物理数据文件上。本书讨论 DBMS 的功能与结构,略去了操作系统。

1.2.1 应用层

最上层是应用层,位于关系数据库管理系统核心之外。应用层处理的对象是各式各样的数据库应用,如利用开发工具开发的应用程序、存储过程、嵌入式 SQL、终端用户发出的事务请求或者交互式 SQL 等。应用层是关系数据库管理系统与用户或应用程序的界面层。

1.2.2 语言处理层

语言处理层的对象是数据库语言,以 SQL 为主;向上提供应用层需要的数据,处理对象是关系或者视图,即元组的集合。

语言处理层的功能是对数据库语言的各类语句进行语法分析、视图转换、完整性检查、安全性检查、查询优化等;通过对下层基本模块的调用,生成可执行代码,代码的执行即可完成数据库语句的功能要求。

查询处理器有四个主要模块:DDL 编译器,DML 编译器,嵌入式 DML 预编译器及查询运行核心程序。

1. DDL 编译器

编译或解释 DDL 语句,包括数据库的三级模式结构、数据的完整性约束、保密与权限限制等约束,并将其登记在数据字典中。

2. DML 编译器

对 DML 语句进行优化并转换成查询运行核心程序能执行的底层指令。

3. 嵌入式 DML 预编译器

将嵌入在主语言中的 DML 语句处理成规范的过程调用形式。应用程序主语言编译程序和 DML 编译器对应用程序编译后产生应用程序目标码。

4. 查询运行核心程序

执行 DML 编译器产生的底层指令。

1.2.3 存储管理层

存储管理层为语言处理层提供数据存取与数据存储,是底层数据和应用程序之间的接口。

数据存取处理的对象是元组,应用层处理的关系或视图在数据存取角度被转换为单记录操作。DBMS 执行表扫描、排序,元组的查询、更新(增、删、改)、封锁等基本操作,以完成数据记录的存取、存取路径维护、事务管理、并发控制与恢复等工作。

数据存储处理的对象是数据页和系统缓冲区,DBMS 通过执行文件的逻辑打开/关闭、读页、写页、缓冲区读写、页面淘汰等操作,完成系统缓冲区的管理、内外存交换以及外存数据管理等功能。

有些书籍将存储管理层分为数据存取与数据存储上下两层。本书将其合称为存储管理层。

存储管理器可以分成四个功能模块:权限和完整性管理器,事务管理器,文件管理器及缓冲区管理器。

1. 权限和完整性管理器

权限和完整性管理器检查用户访问数据的合法性,测试应用程序是否满足完整性约束条件。

2. 事务管理器

数据库系统的逻辑工作单元称为事务,事务由数据库的操作序列组成。数据库系统是以事务为基本逻辑单位来运行的。事务管理器负责保证并发事务操作的正确执行,也负责保证数据库的一致性。

3. 文件管理器

文件管理器负责数据库磁盘空间的合理分配,管理物理文件的存储结构与存取方式。

4. 缓冲区管理器

缓冲区管理器为应用程序开辟数据库系统缓冲区,负责将数据从磁盘读出并送入内存的缓冲区,决定哪些数据进入高速缓冲存储器(Cache)。

原理上数据库管理系统体系结构按照图 1-2 划分,具体系统在划分细节上是多样的,根据 DBMS 的实现环境以及系统规模灵活处理。

1.3 语言处理层

数据库管理系统向应用程序提供多种形式的语言,如交互式 SQL、嵌入式 ESQL、存储过程等,这些语言都由语言处理层支持。

语言处理层的任务就是将各种用户应用提交给数据库管理系统的数据库操作,并将其转换成对 DBMS 内层可执行的基本存取模块的调用序列。

用户应用涉及的数据库操作分为数据定义、数据操纵、数据控制三类,对应 DBMS 数据定义语言、数据操纵语言以及数据控制语言。SQL 集三种语言于一体。

对于数据定义,语言处理层完成语法分析后,将其翻译成内部表示,然后存储在数据字典中。

对于数据控制的定义部分,如安全保密定义、存取权限定义、完整性定义等的处理与数据定义相同。