

数据库技术新进展

(第二版)

李昭原 主编

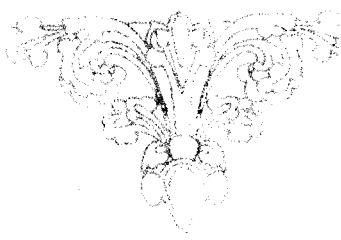
罗晓沛 主审



清华大学出版社

TP311.13/10=2

2007



数据库技术新进展

(第二版)

李昭原 主编 罗晓沛 主审

参加编写的作者(按章节顺序):

何新贵 院士

王 珊	马 垣	周龙骧	杜小勇	王翰虎	彭智勇	杨冬青
唐世渭	周傲英	马玉书	汤 庸	秦小麟	李战怀	刘连芳
刘云生	范 明	唐常杰	于 戈	李昭原	徐洁磐	马应章
陈 梅	王腾蛟	金培权	杨晓春	刘 瑞		
刘云峰	孔令波	陈世亮	徐 玮			

清华大学出版社
北京

内 容 简 介

本书主要介绍近年来数据库理论、技术和应用的最新发展。内容包括数据库技术回顾和展望、数据库理论、数据库管理系统、分布式数据库、面向对象数据库、Web 数据管理、流数据管理、智能数据库、特种数据库(主动数据库、模糊数据库、时态数据库、空间数据库、时空数据库、工程数据库、多媒体数据库、实时数据库、内存数据库)、数据仓库与数据挖掘、无线传感器网络数据管理、数据库应用开发方法与工具、数据库安全、数据库标准、数据库新产品等。内容新颖、丰富,具有先进性、实用性和可读性。既是一本专著,更是一本综合性的数据库教材。

本书可作为计算机专业、计算机应用专业以及各类信息技术、管理专业等学科的大学本科高年级学生和硕士、博士研究生的教材或参考书;也可作为数据库研究和应用工程开发的科技、管理、工程人员的工作参考书。

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 13501256678 13801310933

图书在版编目(CIP)数据

数据库技术新进展/李昭原主编;罗晓沛主审. —2版. —北京:清华大学出版社,2007.10
ISBN 978-7-302-15129-6

I. 数… II. ①李… ②罗… III. 数据库系统 IV. TP311.13

中国版本图书馆 CIP 数据核字(2007)第 059886 号

责任编辑:张瑞庆 赵晓宁

责任校对:李建庄

责任印制:李红英

出版发行:清华大学出版社

<http://www.tup.com.cn>

c-service@tup.tsinghua.edu.cn

社总机:010-62770175

投稿咨询:010-62772015

地 址:北京清华大学学研大厦 A 座

邮 编:100084

邮购热线:010-62786544

客户服务:010-62776969

印 刷 者:北京市清华园胶印厂

装 订 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:185×260 印 张:33 字 数:797 千字

版 次:2007 年 10 月第 2 版 印 次:2007 年 10 月第 1 次印刷

印 数:1~3000

定 价:49.00 元

本书如存在文字不清、漏印、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:010-62770177 转 3103 产品编号:021252-01



第二版 前言

在 1992 年召开的第三届全国数据库教学研讨会上,代表们一致认为:数据库技术的相关著作和教材,在国内已出版不少,但较全面和系统地反映 20 世纪 90 年代数据库技术新进展的内容、特点的书还甚为鲜见,大家一致倡议组织全国知名专家、教授集体编写一本较系统地介绍数据库技术新进展的教材,并委托北京航空航天大学李昭原教授负责筹划和组织。在 1994 年第五届全国数据库教学研讨会上,成立了编委会,并落实了编写大纲和作者。当时的编委会由罗晓沛教授、施伯乐教授、何新贵教授、王珊教授、唐世渭教授、李昭原教授、罗文化教授、徐洁磐教授、马应章教授组成,李昭原教授任主编、罗晓沛教授任主审。该书的内容包括数据库技术的进展、面向对象数据库、分布式数据库、并行数据库、知识库、主动数据库、模糊数据库、工程数据库、多媒体数据库管理、客户/服务器结构数据库、数据库应用快速开发工具、数据库标准化、数据库安全、当前流行的数据库 DBMS 等 14 章内容。编委会召开过两次审稿会议,于 1997 年 5 月由清华大学出版社出版和发行。全国各著名大学和一些地方大学都采用为计算机相关专业研究生的教材,一些企事业单位作为数据库技术研究、开发的主要参考书,收到了很好的社会反响。

2003 年第 20 届全国数据库学术会议期间,许多数据库专委会委员和与会代表提议,希望重新编写《数据库技术新进展》一书,希望全国研究生的数据库教学有一本全国性的教材或主要教学参考书。于是,我们成立了编审组,由罗晓沛教授、李昭原教授和马玉书教授组成,重新组织国内数据库技术知名专家、教授和学者重新编写《数据库技术新进展》一书,以每位数据库专家从事的数据库相关领域优势为基

基础,编写数据库研究与应用领域的理论、技术、应用和实现方法等,以反映近年来数据库理论、技术和应用的最新发展和最新研究成果。

2004—2005年期间,在主编、作者电子邮件沟通的基础上,召开了两次作者会议,通过面对面的讨论不断明确编写的目标、大纲、方法、要求。以达到“统一思想、统一结构、统一规范、统一写法”。2006年是交书稿、审议书稿、修改书稿的一年,许多书稿修改了若干遍,2006年底全书定稿。在内容上,力求“承上启下,突出新进展”,突出一个“新”字;在“结构、逻辑、内容、形式”等方面追求精练、紧凑、少而精,文字上力求“开门见山、直切主题、言简意赅、清晰通顺”,突出一个“精”字。通过3年的努力,新版的《数据库技术新进展》即将复印。其内容包括:数据库技术回顾和展望、数据库理论、数据库管理系统、分布式数据库、面向对象数据库、Web数据管理、流数据管理、智能数据库、特种数据库(主动数据库、模糊数据库、时态数据库、空间数据库、时空数据库、工程数据库、多媒体数据库、实时数据库、内存数据库)、数据仓库与数据挖掘、无线传感器网络数据管理、数据库应用开发方法与工具、数据库安全、数据库标准、数据库新产品等16章内容。

从1997年5月《数据库技术新进展》出版,至今已有整整十年。十年来,数据库技术作为信息技术的重要组成部分已得到长足的发展和进步。除了较传统的技术有新的内容引入外,比较两本书的目录就会发现,新书比旧版增加了不少新的,过去没有的内容,如Web数据管理、数据流管理、时空数据库、数据仓库与数据挖掘、传感器数据管理等,这些都是近年来数据库技术领域的热门研究内容,其成果应视为信息技术应用发展对数据库技术研究的推动和促进。

当前数据处理领域呈现的形势,一是应用需求的多样,促进了多种相关技术与数据库技术的有机结合,出现了很多数据处理新内容、新领域。二是数据管理对象的多样化,呈现出数据的内容和数量从有结构、半结构发展到无结构,从有限发展到无限(海量),从有序发展到无序,从低维发展到高维的趋势。总之,与传统、经典的数据库概念相比,当前的数据库研究对象与内容已经从单一到多样,从一枝独秀发展成五光十色的数据库大家族。

应用需求促进了数据库研究的领域和内容的极大丰富,研究成果反过来又促进了应用的深化,这些数据库技术产品及其衍生产品是将两者很好结合的媒介。可以认为,当前信息技术应用领域中的任何信息系统或应用系统都已离不开数据库技术及数据库管理系统的支持。本书虽力求完整,但也只能就技术发展的主要内容进行讨论。本书各章内容具有独立性,自成体系,全书力求保持系统性和连贯性。

本书由北京航空航天大学李昭原教授任主编,负责对全书内容进行统编;中国科学院研究生院罗晓沛教授任主审,负责对全书内容进行统审,中国石油大学马玉书教授参加了第2、9、10、11章的编审工作和有关编审活动。参加编著的人员有:中国人民大学王珊教授(第1章);辽宁科技大学马垣教授(第2章);中国科学院数学所周龙骧教授、中国人民大学杜小勇教授(第3章);贵州大学王翰虎教授、陈梅副教授(第4章);武汉大学彭智勇教授(第5章);北京大学刘云峰博士、王腾蛟副教授、杨冬青教授、唐世渭教授、孙令波博士(第6章);复旦大学周傲英教授(第7章);中国石油大学马玉书教授(第8章);北京大学何新贵院士、中山大学汤庸教授、南京航空航天大学秦小麟教授、中国科技大学金培权副教授(第9章);西北工业大学李战怀教授、陈世亮讲师、广西大学刘连芳教授、华中科技大学刘云生教授(第10章);郑州大学范明教授、四川大学唐常杰教授(第11章);东北大学于戈教授、杨晓春副教授

(第 12 章);北京航空航天大学李昭原教授、徐玮博士、刘瑞副教授(第 13 章);南京大学徐洁磐教授(第 14 章);华北计算所马应章教授(第 15 章);北京航空航天大学李昭原教授(第 16 章)。

中国计算机学会数据库专业委员会主任王珊教授对本书的编写工作给予了积极的支持和关心,在此表示衷心感谢。

本书编写过程中得到 ORACLE 中国有限公司黄玮女士、徐向东博士,SYBASE 中国有限公司李岷先生、黄欣先生的积极支持和帮助,提供了大量的有关现代数据库产品的技术资料,特别是在第 16 章的编写中给予很大的支持,提出了很多宝贵意见。在此表示衷心感谢。

还要感谢清华大学出版社的领导对本书编写的积极支持和关心,特别是感谢责任编辑张瑞庆副教授的积极支持和工作。

北京航空航天大学研究生贾锦、王磊等同学作了大量的编辑排版等工作,对他们的辛勤劳动表示感谢。

本书采用统一大纲,分散编写,集中交流,集思广益与编审相结合的方式,力求全面、统一,但由于内容广泛,作者分散,并限于编审人水平,定有不完善、不妥当之处,请读者和专家批评指正。

编审者

2007 年 8 月



第一版 前言

数据库技术的相关著作和教材,国内已出版多种,但较全面和系统地反映 20 世纪 90 年代数据库技术新进展的内容、特点的书还甚为鲜见。在 1992 年召开的全国数据库教学研讨会上,很多从事多年数据库教学工作的同志一致倡议集体编写一本较系统地介绍数据库技术新进展的教材,并委托北京航空航天大学负责筹划和组织,1994 年第五届全国数据库教学研讨会成立了“数据库技术新进展”编委会,并组织编写了大纲和落实作者,经过两年的努力,本书已编写完成,不久即可与读者见面。

本书共分 14 章内容。第 1 章概述数据库技术的进展;第 2 章至第 9 章分别论述了新一代数据库:面向对象数据库、分布式数据库、并行数据库、知识库系统、主动数据库、模糊数据库、多媒体数据库、工程数据库等新型数据库的新概念和新技术原理;第 10 章介绍了客户/服务器结构的数据库系统;第 11 章介绍了新一代的数据库应用开发工具;第 12 章介绍了数据库标准化技术的发展;第 13 章介绍了数据库安全技术的发展;第 14 章介绍了当前流行的关系数据库系统 ORACLE、DB2、SYBASE、INFORMIX、INGRES 等的新技术与新产品。各章专题具有独立性,自成体系,而全书又注意系统性、连贯性,并形成一体。

本书由北京航空航天大学李昭原主编,中国科技大学研究生院罗晓沛主审,参加编写的人员有:中国人民大学王珊(第 1 章);复旦大学周傲英、楼荣生、罗文化(第 2 章);武汉大学郑振楣、郭敏(第 3 章);黑龙江大学李建中(第 4 章);石油大学马玉书、南京大学徐洁磐、郑州大学范明(第 5 章);国防科工委系统所何新贵(第 6,7 章);中国科技大学研究生院罗晓沛、邵佩英(第 8 章);北京航空航天大学李隆江(第 9

章);北京大学唐世渭、杨冬青(第10章);北京航空航天大学刘瑞、李昭原(第11章);电子部华北计算所马应章(第12章);中软总公司刘启原(第13章);北京航空航天大学李昭原、邓昉、莫炼等(第14章)。

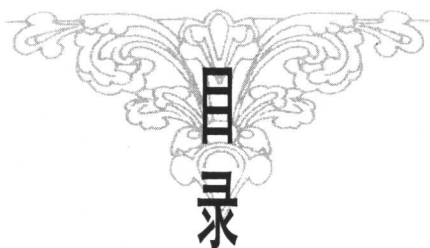
编委会名誉主编罗浇沛教授、施伯乐教授、教学研讨会负责人罗文化教授等都对本书的编写工作给予了积极支持和指导。编委会成员也都各尽其职,对本书出版作了积极贡献。在此表示衷心的感谢。在本书编写过程中也得到了各有关公司的积极支持。SYBASE 中国有限公司杨孝如先生、IBM 中国有限公司黄国威先生、ORACLE 中国有限公司黄玮女士、INFORMIX 中国有限公司王焰先生、华润资讯科技(北京)有限公司张一飞先生在本书第14章的编写过程中提供了大量材料,提出了许多宝贵意见。在此表示衷心的感谢。

北京航空航天大学研究生徐玮、莫炼等作了大量的画图、排版等工作,对他们辛勤的劳动表示感谢。

本书采用统一大纲,分散编写,集中交流,集思广益与统编相结合的编写方式,力求全面、系统,但由于时间短促,内容广泛,作者分散,并限于本人水平,可能有许多不完善不妥当之处,请同行专家和读者批评指正。

李昭原

1997年3月



第 1 章 数据库技术回顾和展望	1
1.1 概述	1
1.2 数据库发展的三个阶段	2
1.2.1 第一代数据库系统	3
1.2.2 第二代数据库系统	4
1.2.3 新一代数据库系统	5
1.3 数据库系统发展的特点	6
1.3.1 数据模型的发展	7
1.3.2 数据库技术与相关技术相结合	11
1.3.3 面向应用领域的数据库新技术	11
1.4 数据库技术发展趋势	12
1.4.1 数据来源和特性的变化	12
1.4.2 应用领域的变化	13
1.4.3 计算机硬件及相关技术的发展	14
1.4.4 当前若干研究热点	15
1.4.5 未来几年发展趋势预测	19
参考文献	23

第2章 数据库理论	24
2.1 数据库理论发展历程及新的挑战	24
2.2 值依赖及伪内涵	26
2.2.1 值依赖形式化定义	27
2.2.2 伪内涵	28
2.2.3 伪内涵计算方法	28
2.3 概率关系模型	29
2.3.1 联合分布与边缘分布	30
2.3.2 合并与约束	31
2.3.3 概率关系代数	32
2.4 序关系数据库理论	35
2.4.1 序关系模型	35
2.4.2 数据库的保序自同构	35
2.4.3 序代数及层次	36
2.4.4 非经典函数依赖	38
2.5 时态数据模型	40
2.5.1 时间模型	40
2.5.2 数据模型	41
2.5.3 时态函数依赖	42
参考文献	44
第3章 数据库管理系统	45
3.1 数据库系统、数据库管理系统和数据库	45
3.2 数据库管理系统的分层体系结构	45

3.3 数据库管理系统的用户接口层	47
3.4 数据库管理系统的数据库系统层	49
3.5 数据库管理系统的存取系统层	52
3.6 数据库管理系统的存储系统层	54
3.7 数据库管理系统的存储结构	56
3.8 数据库系统的目录结构	58
3.9 数据库管理系统并发控制的实现	60
3.9.1 基于封锁的调度	61
3.9.2 基于时间戳的调度方法	63
3.9.3 适当修改写规则可以减少回退率	64
3.9.4 基于有效性检验的调度方法	65
3.9.5 多版本并发控制机制	66
3.10 数据库管理系统的恢复技术	66
3.11 数据库管理系统的查询优化	70
3.12 数据库管理系统的索引技术	73
参考文献	75
 第4章 分布式数据库	 76
4.1 第一代分布式数据库系统	76
4.1.1 概述	76
4.1.2 分布式数据库系统四层结构	76
4.1.3 数据分片和数据复制	77
4.1.4 分布式查询处理及其优化	78
4.1.5 分布式并发控制和可靠性协议	79
4.2 新一代分布式数据库技术	80
4.2.1 分布式数据库系统结构	80

4.2.2 动态数据复制和缓存技术	82
4.2.3 分布式查询优化	83
4.2.4 高级事务模型	84
4.2.5 分布式内容管理	85
4.3 多数据库系统	86
4.3.1 多数据库系统体系结构	87
4.3.2 模式集成	88
4.3.3 多数据库系统的查询处理及优化	89
4.4 虚拟数据库	90
4.4.1 虚拟数据库管理系统体系结构	92
4.4.2 数据集成	93
4.5 P2P 数据管理技术	94
4.5.1 P2P 数据管理体系结构	95
4.5.2 P2P 资源管理	96
4.5.3 P2P 分布式查询及优化	96
4.5.4 P2P 数据管理系统实例	100
参考文献	102
第5章 面向对象数据库	103
5.1 面向对象数据库概述	103
5.1.1 面向对象数据模型	103
5.1.2 面向对象数据库语言	106
5.1.3 面向对象数据库系统	108
5.1.4 面向对象数据库应用	109
5.2 对象关系数据库	112

5.2.1 对象关系数据模型	113
5.2.2 对象关系数据库语言	114
5.2.3 对象关系数据库系统	118
5.2.4 对象关系数据库应用	120
5.3 对象代理数据库	122
5.3.1 对象代理数据模型	123
5.3.2 对象代理数据库语言	124
5.3.3 对象代理数据库系统	128
5.3.4 对象代理数据库应用	132
参考文献	134
第 6 章 Web 数据管理	135
6.1 Web 数据的提取与集成	135
6.1.1 半结构化数据模型	136
6.1.2 Web 数据提取技术	137
6.1.3 Web 数据集成与共享	142
6.2 XML 数据管理	145
6.2.1 XML 数据模型	145
6.2.2 XML 数据的查询	151
6.2.3 XML 数据的索引	157
6.3 Web 环境下基于 XML 的关系数据发布	163
6.3.1 基于 XML 的关系数据发布的概述	163
6.3.2 基于 XML 的数据发布系统的体系结构	165
6.3.3 基于 XML 的关系数据发布方法	166
参考文献	170

第 7 章 流数据管理	173
7.1 流数据管理的应用背景和特点	173
7.1.1 应用背景	173
7.1.2 流数据应用的特点	174
7.2 流数据管理系统 (DSMS) 的体系结构	176
7.2.1 流数据管理系统的结构	176
7.2.2 流数据管理原型系统	177
7.3 流数据上的概要数据结构维护	186
7.3.1 概要数据结构及其维护的理论基础	186
7.3.2 哈希方法	187
7.3.3 基于直方图的方法	188
7.3.4 基于抽样的方法	190
7.3.5 基于小波的方法	191
7.3.6 滑动窗口上的概要数据结构维护	192
7.4 连续查询处理	194
7.4.1 流数据上的查询语言	194
7.4.2 查询执行	195
7.4.3 查询优化与调度	197
7.4.4 降载	198
7.5 流数据上的数据挖掘	199
7.5.1 流数据挖掘与传统数据的不同	199
7.5.2 流数据上的频繁项和频繁项集维护	200
7.5.3 流数据上的聚类分析	203
7.5.4 流数据上的分类器构造	205

7.5.5 流数据上的突变检测与趋势预测	208
7.5.6 多条流数据上的监视与挖掘	209
参考文献	211
第 8 章 智能数据库	217
8.1 智能数据库及其发展	217
8.1.1 智能数据库的定义与发展	217
8.1.2 演绎数据库与基于一阶谓词逻辑的知识库系统	218
8.2 智能数据库系统的研究方向与发展趋势	221
8.2.1 人类知识系统存在的主要问题及其解决途径	221
8.2.2 扩大知识的表示形式与处理范围	222
8.2.3 推理方法与搜索策略的改进	223
8.2.4 知识库语言的改进与扩充	229
8.2.5 知识库的查询优化、智能检索和知识库校验算法研究	230
8.2.6 知识获取(机器学习)与知识发现	234
8.3 智能数据库和智能信息系统的最新实现技术	234
8.3.1 基于计算智能的智能数据库与智能信息系统	234
8.3.2 网络计算与分布式智能数据库系统	242
参考文献	250
第 9 章 特种数据库 (1)	251
9.1 主动数据库	251
9.1.1 主动数据库技术的发展和展望	251
9.1.2 主动数据库的应用领域	253
9.1.3 主动数据库的一般模型	254

9.1.4 主动数据库管理系统	255
9.1.5 基于主动数据库的应用系统	257
9.2 模糊数据库	257
9.2.1 模糊数据的表示方法	258
9.2.2 各种模糊数间的语义距离	258
9.2.3 模糊数据库的特征	260
9.2.4 模糊数据模型	261
9.2.5 模糊数据库管理系统	263
9.2.6 模糊数据库的发展小史和展望	264
9.3 时态数据库	265
9.3.1 时态数据库的发展	265
9.3.2 时态数据库的基本概念	266
9.3.3 时态关系模型及语言	268
9.3.4 时态应用模式及热点课题	272
9.4 空间数据库	273
9.4.1 空间数据模型	274
9.4.2 空间拓扑分析操作	276
9.4.3 空间数据结构与索引	278
9.4.4 空间数据库管理系统实现技术	280
9.4.5 三维空间数据库	282
9.5 时空数据库	283
9.5.1 时空数据与时空数据管理	283
9.5.2 时空语义	284
9.5.3 时空数据模型	284
9.5.4 时空数据库语言	288

9.5.5 时空数据的存储结构和索引	289
9.5.6 时空查询处理	290
9.5.7 时空数据库未来研究的若干方向	291
参考文献	291
第 10 章 特种数据库 (2)	293
10.1 工程数据库	293
10.1.1 工程数据库的基本概念	293
10.1.2 工程数据模型	296
10.1.3 工程数据库语言	300
10.1.4 工程数据库管理系统	300
10.1.5 工程数据库中的版本管理方法	302
10.1.6 工程数据库应用实例	305
10.2 多媒体数据库	310
10.2.1 多媒体技术	310
10.2.2 多媒体数据的管理与查询	316
10.2.3 多媒体数据的查询与检索	320
10.3 实时数据库	323
10.3.1 实时数据库的发展	323
10.3.2 实时数据库的特征	324
10.3.3 实时数据库管理系统	325
10.3.4 实时数据库事务	326
10.3.5 实时事务处理	328
10.3.6 实时数据库的数据管理	330
10.3.7 实时数据库的 I/O 调度	332