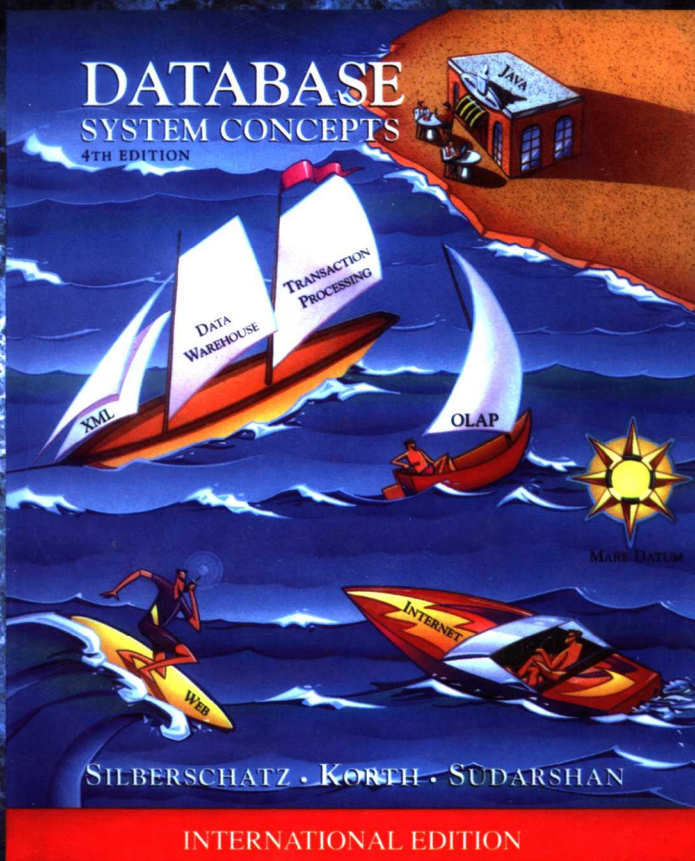


计 算 机 科 学 丛 书

原书第4版

数据库系统概念

Abraham Silberschatz Henry F. Korth S. Sudarshan 著 杨冬青 唐世渭 等译
贝尔实验室 贝尔实验室 印度理工学院 北京大学



Database System Concepts

Fourth Edition



机械工业出版社
China Machine Press



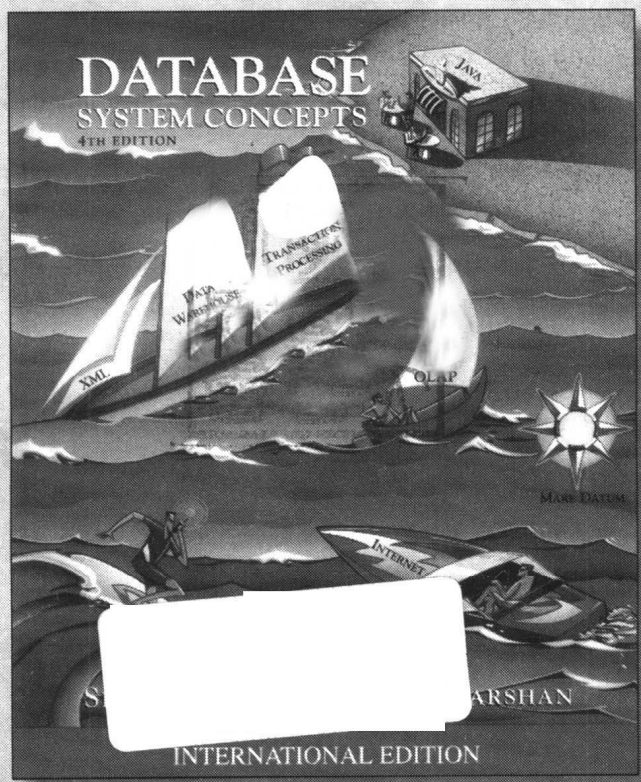
计 算 机 科 学 丛 书

原书第4版

TP311.13
2X136

数据库系统概念

Abraham Silberschatz 贝尔实验室 Henry F. Korth 贝尔实验室 S. Sudarshan 印度理工学院 著 杨冬青 唐世渭 等译
北京 大学



Database System Concepts Fourth Edition

机械工业出版社
China Machine Press

本书是经典的数据库系统教科书《Database System Concepts》的最新修订版，它被国际上许多著名大学采用，包括斯坦福大学、德克萨斯大学、印度理工学院等。而且，这本书前一版的中文版被国内的许多大学采用为本科生和研究生的数据库课程的教材或主要教学参考书，收到了良好的效果。

本书包含了数据模型、关系模型、基于对象的系统和XML、数据存储和检索、事务处理、数据库系统体系结构等方面的内容，在第八部分“实例研究”中还讨论了特定的商品化数据库系统的细节。同时，本书还提供了关于网状模型和层次模型的联机附录。文献注解中指出了首先发表各个研究结果的文章，并列出了进一步阅读的参考文献。本书中的概念都以直观的方式加以描述，其中的许多概念基于银行运行示例，另外，书中省略了形式化证明而采用图表和示例来描述概念及预期结果，增强了本书的易读性。本书会使读者在数据库设计、数据库语言、数据库系统实现等方面得到明显提高。

本书既可作为本科生三、四年级的数据库课程教材，也可作为研究生一年级的教科书。同时，本书还提供了高级内容供数据库领域的技术人员阅读。

Abraham Silberschatz, Henry F. Korth, S. Sudarshan: Database System Concepts, Fourth Edition (ISBN 0-07-112268-0).

Copyright © 2002, 1999, 1991, 1986 by The McGraw-Hill Companies, Inc.

Original language published by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Simplified Chinese translation edition jointly published by McGraw-Hill Education (Asia) Co. and China Machine Press.

本书中文简体字翻译版由机械工业出版社和美国麦格劳-希尔教育(亚洲)出版公司合作出版。未经出版者预先书面许可，不得以任何方式复制或抄袭本书的任何部分。

本书封面贴有McGraw-Hill公司防伪标签，无标签者不得销售。

版权所有，侵权必究。

本书版权登记号：图字：01-2001-3945

图书在版编目(CIP)数据

数据库系统概念(原书第4版)/西尔伯沙茨(Silberschatz, A.)等著;杨冬青等译.-北京:机械工业出版社,2003.3

(计算机科学丛书)

书名原文: Database System Concepts, Fourth Edition

ISBN 7-111-11625-9

I. 数… II. ①西… ②杨… III. 数据库系统-教材 IV. TP311.13

中国版本图书馆CIP数据核字(2003)第009751号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:杨海玲

北京忠信诚胶印厂印刷·新华书店北京发行所发行

2003年3月第1版第1次印刷

787mm×1092mm 1/16·50印张

印数:0 001-5 000册

定价:69.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域中取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭开了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短、从业人员较少的现状下，美国等发达国家在其计算机科学发展的几十年间积淀的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章图文信息有限公司较早意识到“出版要为教育服务”。自1998年开始，华章公司就将工作重点放在了遴选、移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall, Addison-Wesley, McGraw-Hill, Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum, Stroustrup, Kernighan, Jim Gray等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及收藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专诚为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都步入一个新的阶段。为此，华章公司将加大引进教材的力度，在“华章教育”的总规划之下出版三个系列的计算机教材：除“计算机科学丛书”之外，对影印版的教材，则单独开辟出“经典原版书库”；同时，引进全美通行的教学辅导书“Schaum's Outlines”系列组成“全美经典学习指导系列”。为了保证这三套丛书的权威性，同时也为了更好地为学校和老师服务，华章公司聘请了中国科学院、北京大学、清华大学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成“专家指导委员会”，为我们提供选题意见和出版监督。

这三套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业

的教学度身订造的。其中许多教材均已为M. I. T., Stanford, U.C. Berkeley, C. M. U. 等世界名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色——有的出自语言设计者之手、有的历经三十年而不衰、有的已被全世界的几百所高校采用。在这些圆熟通博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证，但我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。教材的出版只是我们的后续服务的起点。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

电子邮件：hzedu@hzbook.com

联系电话：(010) 68995264

联系地址：北京市西城区百万庄南街1号

邮政编码：100037

专家指导委员会

(按姓氏笔画顺序)

尤晋元	王 珊	冯博琴	史忠植	史美林
石教英	吕 建	孙玉芳	吴世忠	吴时霖
张立昂	李伟琴	李师贤	李建中	杨冬青
邵维忠	陆丽娜	陆鑫达	陈向群	周伯生
周克定	周傲英	孟小峰	岳丽华	范 明
郑国梁	施伯乐	钟玉琢	唐世渭	袁崇义
高传善	梅 宏	程 旭	程时端	谢希仁
裘宗燕	戴 葵			

译者序

数据库系统是对数据进行存储、管理、处理和维护的软件系统，是现代计算环境中的一个核心成分。随着计算机硬件、软件技术的飞速发展和计算机系统在各行各业的广泛应用，数据库技术的发展尤其迅速，引人注目。有关数据库系统的理论和技术是计算机科学技术教育中必不可少的部分。《数据库系统概念》是一本经典的、备受赞扬的数据库系统教科书。其内容由浅入深，既包含数据库系统基本概念，又反映数据库技术新进展。它被国际上许多著名大学所采用，包括斯坦福大学、德克萨斯大学、印度理工学院（孟买）等。

我们于2000年初将《数据库系统概念》一书的第3版译成中文，由机械工业出版社出版，国内许多大学采用其作为本科生和研究生数据库课程的教材或主要教学参考书，收到了良好的效果。现在我们又翻译了2002年出版的该书第4版。第4版保持了前三版的总体风格，同时重点强调了数据库技术的新进展。第4版的内容大体上可以分为四个部分。

第1~10章讲述数据库系统的基本概念，包括对数据库系统的性质和目标的综述、关系数据模型和几种关系语言、数据库的完整性约束及其在数据库设计中的应用。这一部分中还介绍了新型的数据库系统——面向对象数据库和对对象-关系数据库，以及XML数据表示和数据查询。

第11~20章讨论数据库系统实现技术，包括数据存储结构、数据存取技术、查询优化方法、事务处理系统的基本概念和并发控制、故障恢复技术。这一部分中还介绍了数据库系统体系结构，以及并行数据库系统和分布式数据库系统中所采用的一些主要策略和技术。

第21~24章讨论数据库研究中的新进展和新的数据库应用。主要内容包括数据库的Web接口、数据库性能调整 and 性能测试标准、电子商务中的数据库问题，联机分析处理（OLAP）、数据挖掘、数据仓库工程、用于查询正文数据的信息检索技术等高级查询问题，时态数据、空间和地理数据、多媒体数据、移动式和个人数据库管理等高级数据类型和新应用，以及事务处理监控器、高性能事务系统、实时事务系统和事务工作流等高级事务处理问题。

第25~27章对Oracle、IBM DB2和Microsoft SQL Server三种商品化数据库系统进行实例研究，结合这几种具体产品来讨论前面各部分描述的各种实现技术是如何应用到实际系统中的。

这本书的第一部分以及第二和第三部分的部分内容可以作为本科数据库概论课程的教材或主要参考资料，第二和第三部分的其余内容可以用于研究生的数据库课程教学，第四部分可以作为帮助学生了解实际系统的补充材料。

杨冬青、唐世渭组织并参加了本书的翻译和审校工作，参加翻译的还有余剑纯、张俊虎、郭迎春、章悦、王涓、陈佳、郭炜、应莺，张立新为本书的翻译做了许多资料整理工作。

限于译者水平，译文中疏漏和错误难免，欢迎批评指正。

译者

2002年于北京大学

译者简介



杨冬青 北京大学信息科学技术学院教授，博士生导师，网络与信息系统研究所副所长，数据库与信息系统研究室主任。中国计算机学会数据库专委会委员，中国计算机学会普及工作委员会副主任。1969年毕业于北京大学数学力学系数学专业，从事数据库与信息系统领域研究、开发与教学20余年，曾获国家科技进步二等奖等多项奖励。



唐世渭 北京大学信息科学技术学院教授，博士生导师。北京大学视觉与听觉信息处理国家重点实验室主任。中国计算机学会数据库专委会副主任。1964年毕业于北京大学数学力学系计算数学专业，从事数据库与信息系统领域研究、开发与教学20余年，曾获国家科技进步二等奖等多项奖励。

前 言

数据库管理已经从一种专门的计算机应用发展为现代计算环境的一个核心成分。因此，有关数据库系统的知识已成为计算机科学教育中的一个必不可少的部分。在本书中，我们讲述数据库管理的基本概念，这些概念包括数据库设计、数据库语言及数据库系统实现等多个方面。

本书可作为本科生三年级或四年级数据库入门课程的教科书，也可作为研究生一年级的教科书。除了可用于入门课程的基本内容外，本书还包括了作为课程补充或作为高级课程介绍性材料的高级内容。

我们仅要求读者熟悉基本的数据结构、计算机组织和一种高级程序设计语言，例如Java、C或Pascal。书中的概念都以直观的方式加以描述，其中的许多概念基于我们的银行运行例子。本书中包括了一些重要的理论结果，但省略了形式化证明。对于各个研究结果，文献注解中指出了首先发表并证明该结果的文章，文献注解中还列出了进一步阅读的参考文献。在应该给出证明的地方，我们使用图表和例子来说明为什么结论是正确的。

本书中所包括的基本概念和算法通常基于商品化的或试验性的数据库系统中所采用的概念和算法。然而我们以一种一般化的形式来描述这些概念和算法，而不是和一个特定的数据库系统联系在一起。我们在第八部分“实例研究”中讨论特定的商品化数据库系统的细节。

在《数据库系统概念》的第4版中，我们保持了前三版的总体风格，同时重点强调了数据库管理的进展。新增加了几章来讨论新的技术。对每一章都进行了编辑，并对大多数的章进行了修改扩充。下面我们对各章做一个简短描述。

本书的组织

本书组织成八个主要部分和三个附录：

- **综述**（第1章）。第1章对数据库系统的性质和目标进行一般性综述。我们解释了数据库系统的概念是如何发展起来的，各种数据库系统的共同特性是什么，数据库系统能为用户做什么，以及数据库系统如何与操作系统接口。我们还引入了一个数据库应用系统的例子：包括多个支行的一个银行企业。这个例子是贯穿全书的运行实例。这一章讲动机、历史和性质。
- **数据模型**（第2章和第3章）。第2章介绍实体-联系模型，该模型为数据库设计问题，以及我们在数据模型的约束下捕获现实应用的语义时所遇到的问题，提供了一个高层视图。第3章介绍关系数据模型，包括与之相关的关系代数和关系演算。
- **关系模型**（第4~7章）。第4章介绍最有影响的面向用户的关系语言——SQL。第5章介绍另外两种关系语言——QBE和Datalog。这两章对数据操纵——查询、修改、插入、删除进行了描述。算法和设计问题在以后的章节中描述。因此，这两章适合于入门课程和某些只想了解数据库系统是什么，而不想深入掌握其内部算法和结构的人员阅读。

第6章从数据库完整性和安全性的角度介绍了约束；第7章说明在关系数据库的设计

中如何使用约束。第6章介绍了参照完整性、触发器和断言等完整性维护机制以及授权机制。这一章的要点是保护数据库，免遭意外的和有意的破坏。

第7章介绍关系数据库设计理论，包括数据依赖和规范化等，重点强调了提出各种范式的原因以及它们的直观含义。这一章还详细描述了数据库设计的整个过程。

- **基于对象的系统和XML**（第8~10章）。第8章介绍面向对象数据库，包括面向对象程序设计的概念以及如何用这些概念去形成一个数据模型的基础。在这儿不需要读者事先具有面向对象语言的知识。第9章介绍对象-关系数据库，说明SQL:1999标准如何对关系数据模型进行扩展，以包括面向对象的特性，例如继承性、复杂类型和对象标识。

第10章介绍数据表示的XML标准，它正日益广泛地应用于数据通信和复杂数据类型存储。这一章还描述了XML查询语言。

- **数据存储和查询**（第11~14章）。第11章讨论磁盘、文件和文件系统结构，以及关系和对象数据到文件系统的映像。在第12章中介绍了多种数据存取技术，包括散列、B+ 树索引和网格文件索引。第13章和14章的重点是查询计算算法，以及基于保持等价性的查询转换的查询优化。

这几章旨在使读者理解数据库存储和检索的内部机制。

- **事务管理**（第15~17章）。第15章介绍事务处理系统的基本概念，包括事务的原子性、一致性、隔离性和持久性以及可串行性概念。

第16章的重点是并发控制，我们介绍了保证可串行性的几种技术，包括封锁、时间戳和乐观（有效性检查）技术。在这一章中还讨论了死锁问题。第17章讨论了在系统崩溃和磁盘故障情况下保证事务正确执行的主要技术。这些技术包括日志、影子页面、检查点和数据库转储。

- **数据库系统体系结构**（第18~20章）。第18章介绍计算机系统体系结构，并描述了作为基础的计算机对于数据库系统的影响。在这一章中我们讨论了集中式系统、客户-服务器系统、并行体系结构和分布式体系结构，以及网络类型。第19章讨论分布式数据库系统，在分布式数据库系统的环境下再一次讨论数据库设计、事务管理、查询计算和优化问题。这一章还包括了故障时的系统可用性问题，并介绍了LDAP目录系统。

在第20章关于并行数据库的讨论中，我们探讨了各种并行技术，包括I/O并行、查询间并行和查询内并行，以及操作间并行和操作内并行。这一章中还讨论了并行系统设计。

- **其他主题**（第21~24章）。第21章讨论数据库应用开发和管理，包括数据库接口（特别是Web接口）、性能调整、性能测试标准、标准化和电子商务中的数据库问题。第22章讨论查询技术，包括决策支持系统和信息检索。决策支持领域中讨论的话题包括联机分析处理（OLAP）技术、SQL:1999对OLAP的支持、数据挖掘和数据仓库工程。这一章还描述了用于查询正文数据的信息检索技术，包括Web搜索引擎中使用的基于超级链接的技术。

第23章讨论高级数据类型和新应用，包括时态数据、空间数据和地理数据、多媒体数据以及移动式和个人数据库管理中的问题。最后，第24章讨论高级事务处理，包括事务处理监控器、高性能事务系统、实时事务系统以及事务工作流。

- **实例研究**（第25~27章）。在这一部分我们对三个领先的商品化数据库系统进行实例研究，包括Oracle、IBM DB2和Microsoft SQL Server。这几章中列举了上述每一种产品

的独有特性，描述了它们的内部结构，提供了关于各个产品的丰富的有用信息，帮助你了解前面各部分描述的各种实现技术是如何在实际系统中使用的。这几章中还包括了实际系统设计中的几个有趣的方面。

- **联机附录。**虽然大多数新的数据库应用系统使用关系模型或面向对象模型，但网状数据模型和层次数据模型也仍然在使用中。为了满足希望了解这些数据模型的读者的需要，我们给出了描述网状数据模型和层次数据模型的附录，分别为附录A和附录B。这两个附录只是在本书的Web页面上联机提供（<http://www.bell-labs.com/topic/books/db-book>）。

第4版

对本书第4版的产生起指导作用的是我们收到的关于前面几版的许多意见和建议，我们在孟买的印度理工学院讲授本课程的体会，以及我们对于数据库技术发展方向的分析。

我们的基本过程是重写每一章，将较旧的内容修改为最新的内容，添加关于数据库技术当前进展的讨论，并改进学生难于理解的那些话题的描述。现在每一章都有一个术语回顾列表，它可以帮助读者复习该章讨论的关键主题。我们还在大多数的章后面增加了一个“工具”节，提供与该章内容有关的软件工具的信息。我们还增加了新的习题，更新了参考文献。

我们增加了关于XML的新的一章，以及三个实例研究的章，描述领先的商品化数据库系统Oracle、IBM DB2和Microsoft SQL Server。

我们将所有的章组织为几个部分，并重新组织了几个章的内容。为方便那些熟悉本书第3版的读者，我们对于主要的变化说明如下：

- **实体-联系模型。**我们将实体-联系（E-R）模型的讨论已经进行了改善。增加了一些例子，修改了一些例子，目的在于给读者更加直观的引导。增加了关于各种不同的E-R表示法的总结，以及关于UML的新的一节。
- **关系数据库。**第4章关于SQL的讨论现在参考的是SQL:1999标准，该标准是本书第3版出版之后发布的。我们大大扩展了关于SQL的讨论，将with子句、扩充了的嵌入式SQL以及最近几年中应用日益广泛的ODBC和JDBC包括在我们的讨论范围中。我们去掉了第5章中关于Quel的介绍，因为对它的使用已经不多了。我们修改了关于QBE的描述，以去掉某些有歧义的地方，并增加了关于Microsoft Access数据库中使用的QBE的介绍。

第6章现在包括完整性约束和安全性。关于安全性的讨论从第3版的第19章移到了现在的第6章，第6章还介绍了触发器。第7章包括数据库设计和范式。关于函数依赖的讨论从第3版的第6章移到了现在的第7章。我们对第7章进行了许多改写，对于函数依赖的处理提供了几个简捷的算法，并且扩充了关于整个数据库设计过程的讨论。我们将有关多值依赖推理的公理、PJNF和DKNF移到了附录中。

- **基于对象的数据库。**我们改进了第8章中对于面向对象数据库的介绍，修改了关于ODMG的讨论。第9章中关于对象-关系数据库的讨论也进行了修改，特别地，用SQL:1999标准代替了第3章中使用的扩展SQL。
- **XML。**第10章是第4版中新增的一章，关于XML。
- **存储、索引和查询处理。**第11章讨论存储和文件结构，它是第3版中的第10章。我们对其内容进行了修改。在过去的几年中，磁盘驱动器和其他存储机制的许多特性发生了

很大的变化，因此我们的讨论进行了相应的修改。我们修改了关于RAID的介绍，以反映技术发展的趋势，我们还扩充了关于数据字典（目录）的介绍。

第12章是第3版中的第11章，关于索引的介绍，现在包括了位图索引。我们对B⁺树插入算法进行了简化，并提供了搜索算法的伪代码，删除了分段散列，因为其使用已经不多见了。

我们对查询处理的讨论进行了重新组织，将原来的一章（第3版中的第12章）分为两章，其中一章讨论查询处理（第13章），另一章讨论查询优化（第14章）。关于代价估算和查询优化的所有细节都移到了第14章，而将第13章集中在查询处理算法上。我们删掉了为不同的运算计算精确的I/O操作数的几个详细（且冗长）的公式。第14章现在包括优化算法的伪代码，以及关于嵌套子查询和物化视图的优化的新的几节。

- **事务处理。**第15章对事务进行介绍，在第3版中是第13章，我们对它进行了修改，删除了视图可串行化的测试。

第16章讨论并发控制，它包括了新的一节，关于锁管理器的实现。较弱的一致性级别的讨论原来在第3版的第20章中，现在包括在本章中。我们扩展了关于索引结构的并发控制的讨论，给出了关于蟹行协议（crabbing protocol）的细节，该协议是B-link协议的简化。我们还给出了旨在避免幻象的下一码（next-key）封锁方法。第17章讨论恢复，它现在包括了关于ARIES恢复算法的描述。这一章还包括了为在故障情况下提供高可用性的远程备份系统，这是“24×7”应用所需要的一个重要特性。

与对第3版的处理一样，讲课教师可以选择只介绍事务处理概念（只讲第15章）或详细讲解所有内容（讲第15~17章）。

- **数据库系统体系结构。**第18章概述数据库系统体系结构。我们对本章进行了修改，以反映当前的技术，这一章是第3版的第16章。我们交换了并行数据库和分布式数据库这两章的顺序。第19章讨论分布式数据库，这是比较基本的内容，任何一个与数据库打交道的人都应该熟悉该内容。第20章（第3版中是第16章）讨论并行数据库查询处理技术，对此内容感兴趣的主要是那些希望学习数据库内部技术的读者。

我们对第19章关于分布式数据库的讨论做了重要的改写，减少了关于命名和透明性的描述，增加了对于故障时的操作的描述，包括为提供高可用性而采用的并发控制技术。我们缩减了关于三阶段提交协议的描述和关于全局死锁的分布式检测的描述，因为这两者在实践中都用得不多。我们去掉了第3版的第20章中关于异构数据库查询处理问题的描述，增加了关于目录系统，特别是LDAP的新一节，因为这些技术在分布式环境中作为保证信息可用性的机制被广泛使用。

- **其他主题。**虽然我们对整本书都进行了修订和更新，但我们把与正在进展的数据库研究和新的数据库应用有关的描述集中在新增加的四章中，即从第21章到第24章。

第21章是第4版中新的一章，讨论应用开发和管理。这一章描述了如何建立数据库的Web接口，包括servlet和服务端脚本的其他机制的描述，这些都是新的内容。关于性能调整的一节（原先在第19章中）包括了关于5分钟规则和1分钟规则的新内容以及一些新的例子。关于物化视图选择的一节也是新的。有关性能测试和标准的内容进行了更新。我们增加了关于电子商务的新一节，重点在电子商务中的数据库问题。还增加了新的一节，关于遗留系统。

第22章是关于高级查询和信息检索的，包括关于OLAP，特别是SQL:1999有关数据分析的扩展的新内容。有关数据仓库工程和数据挖掘的描述也大大扩充了。有关信息检索，特别是Web搜索的描述做了重要扩充。在第3版中相关的内容在第21章中。

第23章讨论高级数据类型和新应用。这一章的内容包括时态数据、空间数据、多媒体数据和移动数据库。这些内容是第3版第21章内容的更新的版本。第24章讨论高级事务处理，这一章的内容包括TP监控器、工作流系统、主存和实时数据库、长事务以及多媒体数据库中的事务管理。这些内容是第3版第20章内容的更新版本。

- **实例研究。**关于Oracle、IBM DB2和Microsoft SQL Server的实例研究是第4版新增的内容。这几章概述了上述每一种产品的特性，并描述了它们的内部结构。

教师注意

本书包括基本内容和高级内容，在一个学期内不可能讲授所有内容。我们用符号“***”把某些节标记为高级内容，如果愿意的话可以省略这些节，而且仍能保持内容的连续性。

教师可以将课程安排成包括本书各章的不同子集。下面我们列出一些可能的安排：

- 如果学生在本课程中不使用QBE或Datalog，可以不讲第5章。
- 如果面向对象的内容在另一门高级课程中讲述，那么可以不讲第8章、第9章和11.9节。或者，它们可以构成对象数据库高级课程的基础。
- 对于入门课程来说，可以不讲第10章（XML）和第14章（查询优化）。
- 我们对事务处理的讨论（第15章至第17章）和对数据库系统体系结构的讨论（第18章至第20章）都包括一章综述（分别为第15章和第18章）和后续的两章详细讨论。如果你计划把详细讨论的几章推迟到高级课程中去讲授，那么可以只讲第15章和第18章。
- 尽管21.1节可以包括在数据库入门课程中，但第21~24章更适合于高级课程或学生自学。可以在本书的Web主页（见下一节）中找到基于本书内容的教学大纲样板。

Web主页和教学补充材料

本书的Web主页的网址是

<http://www.bell-labs.com/topic/books/db-book>

该主页包括以下内容：

- 介绍本书所有章节的幻灯片。
- 某些习题的答案。
- 三个附录。
- 最新勘误表。
- 本书用户贡献的补充材料。

我们将写出完整的解答手册，仅提供给教师。关于如何得到解答手册的进一步信息，请发电子邮件到customer.service@mcgraw-hill.com。在美国可以打电话800-338-3987。McGraw-Hill关于本书的网页是

<http://www.mhhe.com/silberschatz>

与我们和其他读者联系

我们提供了一个邮件列表，使本书读者之间可以互相通信，也可以与我们通信。如果你愿意加入这一列表，请发一条消息到db-book@research.bell-labs.com，消息中包括你的姓名、所在单位、头衔和电子邮件地址。

我们已尽最大努力避免在本书中出现排版错误、内容失误等。然而，与新发布的软件相类似，错误在所难免。在本书的主页中有一个最新勘误表。如果你能指出尚未包含在最新勘误表中的本书的疏漏之处，我们将十分感激。

我们很高兴能收到你对改进本书的建议，也很欢迎你对本书主页做出对其他读者有用的贡献，如程序设计练习、课程实习建议、联机实验室和个别指导以及讲课要点。

请将电子邮件发到db-book@research.bell-labs.com。来信请寄到Avi Silberschatz, Bell Laboratories, Room 2T-310, 600 Mountain Avenue, Murray Hill, NJ 07974, USA。

致谢

使用本书第3版的许多学生提供了很多有用的意见，对于我们形成第4版帮助很大。此外，许多人口头地或书面地与我们关于此书交换过意见，提出了建议。在此我们不可能一一列出，我们谨对下列人员致以特别的感谢：

- 佛罗里达理工学院的Phil Bernhard、俄亥俄州立大学的Eitan M.Gurari、Old Dominion大学的Irwin Levinstein、佐治亚理工学院的Ling Liu、乔治·梅森大学的Ami Motro、Case Western Reserve大学的Bhagirath Narahari和Meral Ozsoyoglu、伦敦皇家学院的Odinaldo Rodriguez，他们是本书的审阅者。他们的意见对于我们形成第4版帮助很大。
- Soumen Chakrabarti、Sharad Mehrotra、Krithi Ramamtitham、Mike Reiter、Sunita Sarawagi、N.L.Sarda、Dilys Thomas，感谢他们对本书中的几章内容给出了广泛和宝贵的反馈意见。
- Phil Bohannon，他编写了第10章的第一稿，描述XML。
- Oracle的Hakan Jakobsson、IBM的Sriram Padmanabhan、Microsoft的César Galindo-Legaria、Goetz Graefe、José A. Blakeley、Kalen Delaney、Michael Rys、Michael Zwilling、Sameet Agarwal、Thomas Casey，他们编写了几个附录，分别对Oracle、IBM DB2和Microsoft SQL Server数据库系统进行描述。
- Yuri Breitbart，他帮助我们编写了分布式数据库这一章。Mike Reiter，他帮助我们编写了安全性的几节。Jim Melton，他帮助我们阐明SQL:1999。
- Marilyn Turnamian和Nandprasad Joshi，她们做了很好的秘书工作，对于第4版的完成起了重要作用。

最后，Sudarshan感谢他的妻子Sita的爱和支持，感谢他两岁的儿子Madhur的爱，并感谢他的母亲Indira的支持。Hank感谢他的妻子Joan和孩子Abby、Joe的爱和理解。Avi感谢他的妻子Haya和他的儿子Aaron在他修改本书时的耐心和支持。

A.S.

H.F.K.

S.S.

目 录

出版者的话	
专家指导委员会	
译者序	
前言	
第1章 引言	1
1.1 数据库系统的应用	1
1.2 数据库系统与文件系统	2
1.3 数据视图	3
1.3.1 数据抽象	4
1.3.2 实例和模式	5
1.4 数据模型	5
1.4.1 实体-联系模型	5
1.4.2 关系模型	6
1.4.3 其他数据模型	7
1.5 数据库语言	7
1.5.1 数据定义语言	7
1.5.2 数据操纵语言	8
1.5.3 来自应用程序的数据库访问	9
1.6 数据库用户和管理员	9
1.6.1 数据库用户和用户界面	9
1.6.2 数据库管理员	10
1.7 事务管理	10
1.8 数据库系统结构	11
1.8.1 存储管理器	11
1.8.2 查询处理器	12
1.9 应用程序体系结构	12
1.10 数据库系统的历史	12
1.11 小结	15
术语回顾	15
习题	16
文献注解	16
工具	16

第一部分 数据模型

第2章 实体-联系模型	18
2.1 基本概念	18
2.1.1 实体集	18
2.1.2 联系集	20
2.2 约束	22
2.2.1 映射基数	22
2.2.2 参与约束	23
2.3 码	23
2.3.1 实体集	23
2.3.2 联系集	24
2.4 设计问题	25
2.4.1 用实体集还是用属性	25
2.4.2 用实体集还是用联系集	25
2.4.3 二元联系集与n元联系集	26
2.4.4 联系属性的布局	27
2.5 实体-联系图	27
2.6 弱实体集	32
2.7 扩展的E-R特性	33
2.7.1 特殊化	33
2.7.2 一般化	34
2.7.3 属性继承	35
2.7.4 一般化约束	35
2.7.5 聚集	36
2.7.6 可选择的E-R符号	37
2.8 数据库E-R模式的设计	39
2.8.1 设计阶段	39
2.8.2 银行企业的数据库设计	40
2.9 将E-R模式转换为表	41
2.9.1 用表表示强实体集	42
2.9.2 用表表示弱实体集	43
2.9.3 用表表示联系集	43

2.9.4 复合属性	44
2.9.5 多值属性	45
2.9.6 用表表示一般化	45
2.9.7 用表表示聚集	45
2.10 统一建模语言UML**	45
2.11 小结	46
术语回顾	48
习题	48
文献注解	51
工具	52
第3章 关系模型	53
3.1 关系数据库的结构	53
3.1.1 基本结构	53
3.1.2 数据库模式	55
3.1.3 码	57
3.1.4 模式图	58
3.1.5 查询语言	59
3.2 关系代数	59
3.2.1 基本运算	59
3.2.2 关系代数的形式化定义	65
3.2.3 附加运算	65
3.3 扩展的关系代数运算	69
3.3.1 广义投影	69
3.3.2 聚集函数	69
3.3.3 外连接	71
3.3.4 空值**	73
3.4 数据库的修改	74
3.4.1 删除	74
3.4.2 插入	75
3.4.3 更新	75
3.5 视图	76
3.5.1 视图定义	76
3.5.2 通过视图进行更新与空值	77
3.5.3 用视图定义视图	77
3.6 元组关系演算	78
3.6.1 查询的例子	79
3.6.2 形式化定义	81
3.6.3 表达式的安全性	81
3.6.4 语言的表达能力	82

3.7 域关系演算**	82
3.7.1 形式化定义	82
3.7.2 查询的例子	82
3.7.3 表达式的安全性	83
3.7.4 语言的表达能力	84
3.8 小结	84
术语回顾	85
习题	85
文献注解	88

第二部分 关系数据库

第4章 SQL	90
4.1 背景	90
4.2 基本结构	91
4.2.1 select子句	91
4.2.2 where子句	92
4.2.3 from子句	93
4.2.4 更名运算	93
4.2.5 元组变量	94
4.2.6 字符串操作	94
4.2.7 排列元组的显示次序	95
4.2.8 重复	95
4.3 集合运算	96
4.3.1 并运算	96
4.3.2 交运算	97
4.3.3 差运算	97
4.4 聚集函数	97
4.5 空值	99
4.6 嵌套子查询	100
4.6.1 集合成员资格	100
4.6.2 集合的比较	101
4.6.3 测试是否为空关系	102
4.6.4 测试是否存在重复元组	103
4.7 视图	104
4.8 复杂查询	104
4.8.1 派生关系	104
4.8.2 with子句	105
4.9 数据库的修改	106

4.9.1 删除	106	5.2.7 递归的能力	145
4.9.2 插入	107	5.2.8 其他语言的递归	146
4.9.3 更新	108	5.3 用户界面和工具	147
4.9.4 视图的更新	109	5.3.1 表单和图形用户界面	147
4.9.5 事务	110	5.3.2 报表生成器	148
4.10 关系的连接**	110	5.4 小结	149
4.10.1 举例	110	术语回顾	149
4.10.2 连接类型和条件	112	习题	149
4.11 数据定义语言DDL	113	文献注解	151
4.11.1 SQL中的域类型	114	工具	152
4.11.2 SQL的模式定义	115	第6章 完整性和安全性	153
4.12 嵌入式SQL	117	6.1 域约束	153
4.13 动态SQL	119	6.2 参照完整性	154
4.13.1 ODBC**	119	6.2.1 基本概念	154
4.13.2 JDBC**	121	6.2.2 E-R模型中的参照完整性	155
4.14 其他SQL特性	123	6.2.3 数据库的修改	156
4.14.1 模式、目录、环境	123	6.2.4 SQL中的参照完整性	156
4.14.2 过程扩展与存储过程	123	6.3 断言	158
4.15 小结	124	6.4 触发器	158
术语回顾	124	6.4.1 触发器的需求	159
习题	125	6.4.2 SQL中的触发器	159
文献注解	127	6.4.3 何时不用触发器	161
第5章 其他关系语言	128	6.5 安全性和授权	162
5.1 Query-by-Example	128	6.5.1 安全性违例	162
5.1.1 在一个关系上的查询	128	6.5.2 授权	163
5.1.2 在多个关系上的查询	130	6.5.3 授权与视图	164
5.1.3 条件框	131	6.5.4 权限的授予	164
5.1.4 结果关系	132	6.5.5 角色的概念	165
5.1.5 元组的显示次序	133	6.5.6 审计追踪	166
5.1.6 聚集操作	133	6.6 SQL中授权	166
5.1.7 数据库的修改	134	6.6.1 SQL中的权限	166
5.1.8 在Microsoft Access中的QBE	136	6.6.2 角色	167
5.2 Datalog	137	6.6.3 授权的权限	167
5.2.1 基本结构	137	6.6.4 其他特性	168
5.2.2 Datalog规则的语法	139	6.6.5 SQL授权的局限	168
5.2.3 非递归Datalog的语义	140	6.7 加密和鉴定	168
5.2.4 安全性	142	6.7.1 加密技术	169
5.2.5 Datalog中的关系运算	143	6.7.2 鉴定	170
5.2.6 Datalog中的递归	143	6.8 小结	170