

数据库系统

设计、实现与管理（进阶篇）

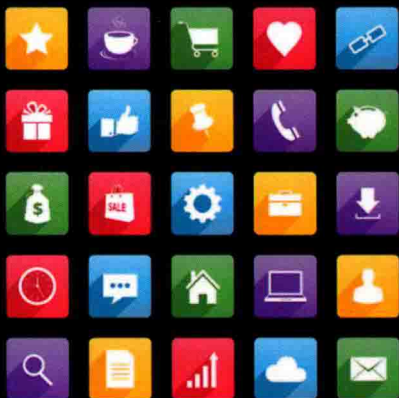
[英] 托马斯 M. 康诺利 (Thomas M. Connolly) 著 宁洪 李姗姗 王静 译
卡洛琳 E. 贝格 (Carolyn E. Begg)

Database Systems

A Practical Approach to Design, Implementation, and Management Sixth Edition

DATABASE SYSTEMS

A Practical Approach to
Design, Implementation, and Management
SIXTH EDITION



THOMAS CONNOLLY
CAROLYN BEGG



机械工业出版社
China Machine Press

计 算 机 科 学 丛 书

原书第6版

数据库系统

设计、实现与管理（进阶篇）

[英] 托马斯 M. 康诺利 (Thomas M. Connolly) 著 宁洪 李姗姗 王静 译
卡洛琳 E. 贝格 (Carolyn E. Begg)

Database Systems

A Practical Approach to Design, Implementation, and Management Sixth Edition

DATABASE SYSTEMS

A Practical Approach to
Design, Implementation, and Management
SIXTH EDITION



THOMAS CONNOLLY



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

数据库系统: 设计、实现与管理 (进阶篇)(原书第 6 版) / (英) 托马斯 M. 康诺利 (Thomas M. Connolly), (英) 卡洛琳 E. 贝格 (Carolyn E. Begg) 著; 宁洪, 李姗姗, 王静译.
—北京: 机械工业出版社, 2017.9

(计算机科学丛书)

书名原文: Database Systems: A Practical Approach to Design, Implementation, and Management, Sixth Edition

ISBN 978-7-111-58388-2

I. 数… II. ①托… ②卡… ③宁… ④李… ⑤王… III. 数据库系统—高等学校—教材—英文
IV. TP311.13

中国版本图书馆 CIP 数据核字 (2017) 第 267187 号

本书版权登记号: 图字 01-2014-5468

Authorized translation from the English language edition, entitled Database Systems: A Practical Approach to Design, Implementation, and Management, 6E, 978-0-13-294326-0 by Thomas M. Connolly, Carolyn E. Begg, published by Pearson Education, Inc., Copyright © 2015, 2010.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Chinese simplified language edition published by Pearson Education Asia Ltd., and China Machine Press Copyright © 2018.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

本书是数据库领域的经典著作, 内容系统、全面、实用, 被世界多所大学选为数据库相关课程的教材。中文版分为两册, 分别对应原书第一~五部分 (基础篇) 和第六~九部分 (进阶篇)。本书为进阶篇, 主要内容有: 分布式 DBMS 及复制服务器; 基于对象的 DBMS; 作为数据库应用平台的 Web 与 DBMS; 商务智能技术, 包括数据仓库、联机分析处理 (OLAP) 和数据挖掘。

本书既可作为数据库设计与管理相关课程的本科和研究生教材, 亦可作为数据库专业技术人员的参考书籍。

出版发行: 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: 张锡鹏

责任校对: 李秋荣

印刷: 三河市宏图印务有限公司

版次: 2018 年 1 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 29.75

书号: ISBN 978-7-111-58388-2

定价: 129.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

文艺复兴以来，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取得了垄断性的优势；也正是这样的优势，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，还揭示了学术的源变，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和发展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到积极的推动作用，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公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson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、研究及珍藏。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相助，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，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，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。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

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。随着计算机科学与技术专业学科建设的不断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将步入一个新的阶段，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。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：

华章网站：www.hzbook.com

电子邮件：hzjsj@hzbook.com

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



随着信息社会的到来，数据日益成为宝贵的资源，数据库作为数据管理的基本技术和工具正广泛应用于各行各业。然而，正如本书作者在前言中所述：“遗憾的是，正是由于数据库系统的简单性，许多用户有可能尚缺乏必要的知识，还不懂得如何开发正确且高效的系统，就开始创建数据库及其应用程序了。”本书的目标就是以读者易于接受和理解的方式介绍数据库设计、实现和管理的基本理论、方法与技术。本书既涵盖了数据库领域的经典内容，又反映了近期的研究成果，特别是为数据库概念设计、逻辑设计和物理设计提供了步骤完备的方法学，非常有实用价值。2004年我们曾翻译出版了本书的第3版，广受读者欢迎，后来的第4、5版扩充了数据库领域的新技术，特别是随着互联网、移动计算的发展衍生出来的许多应用技术，同时增强了可读性。第6版增加了云计算、时态数据库等内容，充实了第21章“数据管理中的职业、法律与道德问题”，与SQL相关的章节全面更新为2011年公布的新标准SQL: 2011。

本书的两位作者均具有丰富的数据库管理系统和数据库应用系统的设计经验，托马斯M. 康诺利(Thomas M. Connolly)曾参与设计世界上第一个商业可移动数据库管理系统RAPPORT和配置管理工具LIFESPAN，后者获英国设计奖；卡洛琳E. 贝格(Carolyn E. Begg)则是数据库技术应用于生物领域的专家。

本书系统性强、结构清晰，特别注重理论联系实际，并以一个可运行的DreamHome(房屋租赁公司)案例贯穿全书，便于阅读和理解。本书既可作为计算机及相关专业本科生数据库管理或数据库设计的导论性教材(选取部分内容)，也可作为研究生或高年级本科生相关课程的教材，亦可作为IT专业人士(包括系统分析和设计人员、应用程序开发人员、系统程序员、数据库从业人员及独立的自学者)的参考书。

中文版分为两册，共九个部分。本书为进阶篇，包括原书的第六~九部分，主要讨论数据库的一些新技术。其中第六部分阐述当前数据库系统领域主要发展方向之一的分布式DBMS；第七部分讨论试图克服关系数据库系统种种缺陷的面向对象DBMS；第八部分涉及将DBMS集成到Web环境的问题、半结构化数据及其与XML的关系、XML查询语言以及XML到数据库的映射等内容；第九部分讨论商务智能，包括数据仓库、联机分析处理(OLAP)和数据挖掘。

宁洪教授全面负责本书的翻译，李姗姗提供了第八、九部分的部分初稿，王静提供了第六、七部分的部分初稿，全书由宁洪教授统稿并审校。

限于译者水平，译文中的疏漏和错误在所难免，欢迎批评指正。

本书是在2004年翻译的第3版的基础上完成的，借此机会向参与第3版翻译工作的其他译者致谢。

译者

国防科技大学计算机科学与技术系

背景

在过去的 30 年中，数据库的研究带来了巨大的生产力，使得数据库系统成为软件工程领域最重要的成果。目前，数据库作为信息系统的基本框架，已从根本上改变了许多公司的运作方式。特别是在最近几年里，随着这项技术本身的发展，产生了一些功能更强大、使用更方便的系统。这使得数据库系统变得越来越普及，用户类型也越来越广泛。遗憾的是，正是由于数据库系统的简单性，许多用户有可能尚缺乏必要的知识，还不懂得如何开发正确且高效的系统，就开始创建数据库及其应用程序了。这样很可能导致所谓的“软件危机”（software crisis，有时也称为“软件抑郁”（software depression））的延续。

编写本书的最初动因是我们在工业界的工作经历，当时我们为新软件系统中数据库的设计提供咨询，间或也解决遗留系统中存在的种种问题。进入学术界后，我们从另一类用户——学生那里发现了类似的问题。因此，本书的目标就是给出一本教程，尽可能清楚地介绍数据库的基础理论，并给出一套既能为专业技术人员亦能为非技术人员所用的数据库设计方法学。

本书针对当前主流的商用产品——关系数据库管理系统（DBMS）给出的设计方法学，已在学术界和工业界测试和使用了许多年。它包括三个主要阶段：数据库的概念设计、逻辑设计和物理设计。第一个阶段在不考虑任何物理因素的前提下设计概念数据模型，得到的数据模型在第二阶段被细化为逻辑数据模型，细化过程主要是去除在关系系统中无法表示的结构。在第三阶段，逻辑数据模型被转换成针对目标 DBMS 的物理设计，物理设计阶段主要考虑如何设计存储结构和访问方法，以便有效并安全地访问存储在辅存中的数据库。

该方法学按阶段被分为一系列步骤。对于缺少经验的设计者，最好按步骤进行设计，这里所提供的指南可帮助你完成整个过程。对于有经验的设计者，该方法学的指导作用显然会弱化，但经常可用于开发框架和检查列表。为了帮助读者学习使用上述方法学并理解其要点，整个方法学的描述中始终贯穿一个完整的 DreamHome 案例研究。附录 B 还给出了另外三个案例，供读者自行研究。

UML（统一建模语言）

越来越多的公司都在规范各自的数据建模方法，即选择一种特定的建模方法并在整个数据库开发项目中始终如一地使用它。一种在数据库概念设计和逻辑设计阶段较为通用的高级数据模型是 ER（实体 - 联系）模型，这也是本书采用的模型。由于当前还没有表示 ER 模型的标准方法，因此大部分书籍在描述关系 DBMS 的数据库设计时，常常使用下述两种表示方法之一：

- Chen 氏表示方法，即用矩形表示实体，用菱形表示联系，用线段连接矩形和菱形。
- Crow Feet（鸦爪）表示方法，仍用矩形表示实体，用实体间的连线表示联系，在一对多联系连线的多端有一个鸦爪标记。

当前,这两种表示方法都有计算机辅助软件工程(CASE)工具。然而,它们都难于使用和解释。本书的较早版本曾使用Chen氏表示方法,而在随后培生教育出版集团进行的一次问卷调查中,比较一致的意见是应该使用最新的称为UML(Unified Modeling Language,统一建模语言)的面向对象建模语言。UML表示方法结合了面向对象设计三大流派的成分:Rumbaugh的OMT建模语言,Booch的面向对象分析和面向对象设计,以及Jacobson的Objectory。

换用表示方法主要有以下三个原因:(1)UML正成为一种工业标准,例如,对象管理组(OMG)已经采纳UML作为对象方法的标准表示方法;(2)UML表达清楚并易于使用;(3)UML目前已被学术界用于面向对象分析与设计的教学,在数据库模块的教学中也使用UML将会更加一致。因此,在这个版本中,我们将采用UML的类图作为ER模型的表示方法。读者将会发现这种表示方法更加容易理解和使用。

第6版的更新之处

- 扩展了第3章“数据库的结构与Web”,增加了云计算。
- 修改了第21章“数据管理中的职业、法律与道德问题”。
- 增加了“数据仓库与时态数据库”(31.5节)。
- 每章后增加了新的思考题和习题。
- 修改了与SQL相关的章节,全面反映2011年公布的新标准SQL:2011。
- 修订了第26章“复制与移动数据库”。
- 修改了关于Web-DBMS集成和XML的章节。
- 与Oracle相关的内容一律修改为针对Oracle 11g。

读者对象

本书可作为本科生数据库管理或数据库设计的导论性教材,也可作为研究生或高年级本科生相关课程的教材,学时可分为一到两个学期。通常信息系统、商业IT或计算机科学等专业都包含这类课程。

本书还可以作为一些IT专业人士的参考书,如系统分析和设计人员、应用程序开发人员、系统程序员、数据库从业人员及独立的自学者。随着当今数据库系统的广泛使用,这些专业人士可能来自于需要数据库的任何类型的公司。

读者在学习关于物理数据库设计的第18章和关于查询处理的第23章之前,如果对附录F中介绍的文件组织和数据结构相关概念有清楚的了解,那么将会有所帮助。理想的情况是这些背景知识已从前导课程中获得。如果不具备这个条件,则可以在开始数据库课程后,学完第1章立即学习附录F。

如果读者已经掌握了一门高级编程语言,比如C,那么在学习附录I的嵌入式与动态SQL和28.3节的ObjectStore时会更有成效。

突出特点

(1)为数据库逻辑设计和概念设计提供了易用、逐步指导的方法学,该方法学基于广泛采用的实体-联系模型并将规范化作为验证技术。此外,通过一个完整的案例研究来说明如何使用这套方法学。

(2)为数据库物理设计提供了易用、逐步指导的方法学,包括:逻辑设计到物理实现的

映射，文件组织方法的选择，适合应用程序的索引结构，以及何时引入可控冗余。此外，通过一个完整的案例研究来说明如何使用这套方法学。

(3) 用独立的章节来讲解以下三个主题：数据库设计阶段在整个系统开发生命周期中的位置与作用；如何使用实况发现技术来获取系统需求；如何将 UML 用于整个方法学。

(4) 每章都采用清晰且易于理解的表述方法，如突出显示定义，明确给出各章学习目标，在各章最后进行小结。通篇使用了大量示例和图表来说明概念。来自现实生活的 DreamHome 案例研究贯穿全书，另外还给出若干案例供学生选作课程实践题目。

(5) 扩充了下列最新的正式标准及事实标准：结构化查询语言 (SQL)，举例查询 (QBE)，面向对象数据库的对象数据管理组 (ODMG) 标准。

(6) 利用三章的篇幅，以教程式风格介绍 SQL 标准，包含交互式 and 嵌入式 SQL。

(7) 专设一章讨论 IT 和数据库中的职业、法律与道德问题。

(8) 全面讨论了与分布式 DBMS 和复制服务器相关的概念和问题。

(9) 全面介绍了基于对象的 DBMS 中的一些概念和问题。回顾了 ODMG 标准，介绍了在最新公布的 SQL 版本 SQL:2011 中出现的各种对象管理机制。

(10) 扩展了作为数据库应用平台的 Web 部分的内容，并给出多个 Web 数据库访问的代码示例。具体包括容器管理持久性 (CMP)、Java 数据对象 (JDO)、Java 持久性 API (JPA)、JDBC、SQLJ、ActiveX 数据对象 (ADO)、ADO.NET 和 Oracle PL/SQL Pages (PSP)。

(11) 介绍了半结构化数据及其与 XML 的关系，扩展了 XML 的内容和相关术语，包括 XML Schema、XQuery、XQuery 数据模型和形式语义。还讨论了在数据库中集成 XML，以及为发布 XML 而在 SQL:2008 和 SQL:2011 中所做的扩展。

(12) 全面介绍了数据仓库、联机分析处理 (OLAP) 和数据挖掘。

(13) 全面介绍了用于数据仓库数据库设计的维度建模技术，并且通过一个完整的案例来演示如何使用该方法进行数据仓库数据库设计。

(14) 介绍了 DBMS 系统实现的有关概念，包括并发技术和恢复控制、安全以及查询处理和查询优化。

教学方法

在开始撰写本书之前，我们的目标之一就是写一本让读者容易接受和理解的教材，而不管读者具备怎样的背景知识和经验。根据我们使用教材的经验以及从很多同事、客户和学生中吸收的意见，实际上存在若干读者喜爱和不喜爱的设计特性。考虑到这些因素，本书决定采用如下的风格和结构：

- 在每章的开头明确说明该章的学习目标。
- 清楚定义每一个重要的概念，并用特殊格式突出显示。
- 通篇大量使用图表来支持和阐明概念。
- 面向实际应用：为了做到这点，每章都包含了许多实际有效的示例以说明所描述的概念。
- 每章最后配有小结，涉及该章所有主要的概念。
- 每章最后配有思考题，问题的答案都可以在书中找到。
- 每章最后配有习题，教师可用其测试学生对章节内容的理解，自学者也可进行自测。全部习题的答案可以在原书配套的教辅资源“教师答案手册”中找到。

教辅资源[⊖]

适用于本教材的教辅资源包括：

- 课程 PPT。
- 教师答案手册，包括所有课后思考题和习题的答案示例。
- 其他资源的配套网站：www.pearsonhighered.com/connolly-begg。

上述资源仅提供给在 www.pearsonhighered.com/irc 上注册过的教师。请与当地的销售代表联系。

本书结构[⊖]

第一部分 背景

本书的第一部分介绍数据库系统和数据库设计。

第 1 章引入数据库管理的概念。主要阐述了数据库前身，即基于文件的系统之不足及数据库方法所具备的优势。

第 2 章总览数据库环境。主要讨论了三层 ANSI-SPARC 体系结构的优点，介绍了目前最通用的数据模型，列出了多用户 DBMS 应提供的各种功能。

第 3 章考察各种多用户 DBMS 结构，讨论了数据库领域不同类型的中间件。分析 Web 服务，它能为用户和 SOA（面向服务的结构）提供新型的业务服务。该章简要描述分布式 DBMS 和数据仓库的结构，后面还将详细讨论。该章还给出一个抽象 DBMS 的内部结构以及 Oracle DBMS 的逻辑结构和物理结构，这一部分内容在数据库管理初级课程中可以略去。

第二部分 关系模型与语言

本书的第二部分介绍关系模型和关系语言，即关系代数和关系演算、QBE（举例查询）和 SQL（结构化查询语言）。这部分还介绍了两种非常流行的商用系统：Microsoft Access 和 Oracle。

第 4 章介绍当前最流行的数据模型——关系模型背后的概念，这是最常被选作商用标准的模型。具体安排是首先介绍术语并说明其与数学上的关系的联系，然后讨论关系完整性规则，包括实体完整性和引用完整性。这一章最后概述视图，第 7 章还将进一步讨论视图。

第 5 章介绍关系代数和关系演算，并用示例加以说明。这部分内容在数据库管理初级课程中可以略去。然而，在第 23 章学习查询处理和第 24 章学习分布式 DBMS 的分段时需要用到关系代数的知识。此外，虽然不是绝对有必要，但是了解过程式的代数与非过程式的演算之间的区别将有利于学习第 6 章和第 7 章介绍的 SQL 语言。

第 6 章介绍 SQL 的数据操作语句 SELECT、INSERT、UPDATE 和 DELETE。该章通过一系列有效的示例，以教程式的风格说明了这些语句的主要概念。

第 7 章讨论 SQL 标准中主要的数据定义机制。该章仍采用教程式风格，介绍 SQL 的数据类型、数据定义语句、完整性增强特性（IEF）和数据定义语句中一些更高级的特性，包

⊖ 关于本书教辅资源，只有使用本书作为教材的教师才可以申请，需要的教师请联系机械工业出版社华章公司，电话：13601156823，邮箱：wanguang@hzbook.com。——编辑注

⊖ 中文版分为两册，分别对应原书第一～五部分（基础篇）和第六～九部分（进阶篇）。本书为进阶篇。——编辑注

括访问控制语句 GRANT 和 REVOKE。此外，将再次讨论视图以及用 SQL 如何创建视图。

第 8 章涉及 SQL 的一些高级特性，包括 SQL 的编程语言 (SQL/PSM)、触发器和存储过程。

第 9 章介绍对象关系 DBMS，并详细描述了 SQL 新标准 SQL:2011 中的各种对象管理特性。该章还讨论了如何扩展查询处理和查询优化机制，以高效处理扩展的各种数据类型。该章最后将讨论 Oracle 中的对象关系特性。

第三部分 数据库分析与设计

本书的第三部分讨论数据库分析和设计的主要技术，以及这些技术的实际运用方法。

第 10 章总览数据库系统开发生命周期的各个主要阶段。特别强调了数据库设计的重要性，并说明这个过程如何被分为概念、逻辑和物理数据库设计三个阶段。此外，还描述了应用程序的设计（功能方面）对数据库设计（数据方面）的影响。数据库系统开发生命周期的关键阶段是选择合适的 DBMS。这一章讨论了对 DBMS 的选择过程，提供了一系列方针和建议。

第 11 章讨论数据库开发者应于何时使用实况发现技术，以及捕获何种类型的实况。这一章描述了最常用的实况发现技术及其优缺点。通过 DreamHome 案例研究说明在数据库系统生命周期的早期阶段如何应用这些技术。

第 12 章和第 13 章介绍了实体 - 联系模型和扩展的实体 - 联系 (EER) 模型，在 EER 模型中，允许使用更高级的数据建模技术，如子类、超类和分类。EER 模型是一种流行的高级概念数据模型，也是这里讨论的数据库设计方法学的一种基本技术。这两章还为读者介绍了如何使用 UML 来表示 ER 图。

第 14 章和第 15 章阐述了规范化背后的一系列概念，它是逻辑数据库设计方法学中的另一项重要技术。通过从一个完整的案例中抽取的几个有效部分，说明如何从一种范式转换到另一种范式，以及将数据库逻辑设计转换为某一更高范式（直至第五范式）的好处。

第四部分 方法学

本书的第四部分介绍了一种数据库设计方法学。该方法学分为三个阶段，分别是概念数据库设计、逻辑数据库设计和物理数据库设计。每个部分都使用 DreamHome 案例研究加以阐述。

第 16 章为概念数据库设计提供逐步指导的方法学。该章说明了如何将设计分解成多个基于各自视图的易于管理的部分，还给出了标识实体、属性、联系和关键字的方法。

第 17 章为关系模型的逻辑数据库设计提供逐步指导的方法学。该章阐述了如何将概念数据模型映射到逻辑数据模型，以及如何针对所需的事务使用规范化技术来验证逻辑数据模型。对于有多个用户视图的数据库系统，这一章还介绍了如何将得到的多个数据模型合并为一个能表示所有视图的全局数据模型。

第 18 章和第 19 章为关系系统的物理数据库设计提供逐步指导的方法学。该章阐述了如何将逻辑数据库设计阶段开发的全局数据模型转换成某关系系统的物理设计。方法学中还说明了如何通过选择文件组织方式和存储结构，以及何时引入可控冗余来改善实现的性能。

第五部分 可选的数据库专题

第五部分阐述了我们认为对于现代数据库管理课程必要的四个专题。

第 20 章讨论数据库的安全和管理问题。安全不仅要考虑 DBMS，还包括整个环境。该

章将讨论 Microsoft Office Access 和 Oracle 提供的一些安全保障，并专门阐述了在 Web 环境下的一些安全问题，并给出了解决这些问题的方法。最后讨论数据管理和数据库管理的任务。

第 21 章考虑有关 IT 和数据管理与治理的职业、法律与道德问题。主要内容包括区分数据和数据库管理员面对的问题和场景中哪些属法律范畴、哪些属道德范畴；各项新的规章给数据和数据库管理员提出了哪些新的要求和职责；萨班斯 - 奥克斯利法案和巴塞尔 II 协议等法规对数据和数据库管理功能有何影响，等等。

第 22 章集中讨论了数据库管理系统应该提供的三种功能，即事务管理、并发控制及故障恢复。这些技术用于确保当多个用户访问数据库或出现硬件 / 软件部件错误时数据库是可靠且一致的。该章还讨论了一些更适合于长寿事务的高级事务模型，最后分析了 Oracle 中的事务管理。

第 23 章阐述查询处理和查询优化。该章讨论查询优化的两种主要技术：一种是使用启发式规则安排查询中操作的顺序，另一种是通过比较不同策略的相对代价选择资源耗费最少的策略。最后分析了 Oracle 中的查询处理。

第六部分 分布式 DBMS 与复制

第六部分阐述分布式 DBMS。分布式 DBMS 技术是当前数据库系统领域一个主要的发展方向。本书前面各章主要介绍集中数据库系统，即由单个 DBMS 控制的位于单个节点的单一逻辑数据库。

第 24 章讨论分布式 DBMS 的概念与问题。使用分布式 DBMS 时，用户既可以访问自己结点上的数据库，也可以访问存储在远程结点上的数据。

第 25 章阐述与分布式 DBMS 相关的各个高级概念。具体地说，重点阐述与分布式事务管理、并发控制、死锁管理以及数据库恢复相关的协议。此外，还讨论了 X/Open 分布式事务处理 (DTP) 协议。最后分析了 Oracle 中的数据分布机制。

第 26 章讨论利用复制服务器替代分布式 DBMS 的方案以及与移动数据库相关的问题。该章也分析了 Oracle 中的复制机制。

第七部分 对象 DBMS

本书前面各章都在关注关系模型和关系系统，其原因是这类系统在当前传统业务数据库应用中占据主导地位。不过，关系系统并不是万能的，在数据库领域发展面向对象 DBMS 就是试图克服关系系统的一些缺陷。第 27 章和第 28 章就专门叙述这一方面的发展细节。

第 27 章首先引入基于对象的 DBMS (object-based DBMS) 的概念，查看业已出现的各类新兴的数据库应用，说明关系数据模型因其种种弱点而对这些新兴的应用无能为力。然后讨论面向对象 DBMS (object-oriented DBMS) 的概念，从介绍面向对象数据模型及持久性编程语言开始。接下来，分析通常 DBMS 所用的两层存储模型与面向对象 DBMS 所用的单层存储模型的区别及对数据访问的影响。此外，还讨论了提供编程语言持久性的不同方法、指针混写的不同技术、版本控制、模式进化和面向对象 DBMS 体系结构等问题。该章也简要介绍了如何将本书第四部分介绍的方法学推广到面向对象 DBMS 中。

第 28 章介绍面向对象管理组 (Object Data Management Group, ODMG) 推荐的新的对象模型，这一模型已成为面向对象 DBMS 的事实标准。该章还介绍了一个商用的面向对象

数据库——ObjectStore。

第八部分 Web 与 DBMS

本书的第八部分涉及将 DBMS 集成到 Web 环境的问题，以及半结构化数据及其与 XML 的关系、XML 查询语言和 XML 到数据库的映射。

第 29 章阐述将 DBMS 集成到 Web 环境的问题。首先简单介绍 Internet 和 Web 技术，然后说明 Web 作为数据库应用平台的适宜性，并讨论这种方法的优缺点。随后讨论若干种将 DBMS 集成到 Web 环境的方法，包括脚本语言、CGI、服务器扩展、Java、ADO 和 ADO.NET，以及 Oracle Internet Platform。

第 30 章阐述半结构化数据，然后讨论 XML 及 XML 如何成为 Web 上数据表示和交换的流行标准。该章讨论 XML 相关技术，如名字空间、XSL、XPath、XPointer、XLink、SOAP、WSDL 和 UDDI，等等。该章还阐述怎样用 XML 模式定义 XML 文档的内容模型，以及怎样用资源描述框架（RDF）为元数据交换提供框架。此外，还讨论了 XML 的查询语言，具体集中在由 W3C 提出的 XQuery。该章也讨论了为支持 XML 发布，或更广义地说为在数据库中映射和存储 XML 而对 SQL:2011 的扩展。

第九部分 商务智能

本书的最后一部分考虑与商务智能有关的主要技术，包括数据仓库、联机分析处理（OLAP）和数据挖掘。

第 31 章讨论数据仓库，包括它的定义、进化过程及潜在优缺点。该章阐述数据仓库的体系结构、主要组成部分和相关工具与技术，讨论数据集市及其开发和管理的有关问题。此外也讨论了数据仓库中与时态数据管理关联的概念及实践。最后分析了 Oracle 中的数据仓库机制。

第 32 章提供了设计用于决策支持的数据仓库和数据集市数据库的方法。该章描述了维度建模技术的基本概念并将其与传统的实体 - 联系建模技术进行比较；给出了数据仓库设计方法学指南，并通过扩展的 DreamHome 案例研究说明如何实际使用该方法学。该章最后说明如何用 Oracle Warehouse Builder 设计数据仓库。

第 33 章考虑联机分析处理（OLAP）。首先讨论了何谓 OLAP 以及 OLAP 应用的主要特性，然后讨论了多维数据的表示及主要的 OLAP 工具，最后讨论了 SQL 标准针对 OLAP 的扩展以及 Oracle 对 OLAP 的支持。

第 34 章考虑数据挖掘（DM）。首先讨论了何谓 DM 以及 DM 应用的主要特性，然后讨论了数据挖掘操作的主要特性和相关技术，最后描述了 DM 过程和 DM 工具的主要特性，以及 Oracle 对 DM 的支持。

附录

附录 A 给出 DreamHome 案例研究的说明，它将在全书通篇使用。

附录 B 给出另外三个案例研究，供学生课程设计时选用。

附录 C 给出有别于 UML 的另外两种建模表示法，即 Chen 氏表示方法和 Crow Feet 表示方法。

附录 D 总结了第 16 ~ 19 章讨论的概念、逻辑和物理数据库设计方法学。

附录 E 简单介绍用 C# 实现的一个称为 Pyrrho 的轻量级 RDBMS，它能说明本书讨论的许多概念，还能下载使用。

在线附录[⊖]

附录 F 介绍文件组织和存储结构的相关概念，它们对理解第 18 章讨论的物理数据库设计和第 23 章讨论的查询处理是必要的。

附录 G 给出 Codd 的关于关系 DBMS 的 12 条规则，它是鉴别关系 DBMS 的标准。

附录 H 介绍了两种最常用的商用关系 DBMS：Microsoft Office Access 和 Oracle。在本书的许多章节中，都涵盖这两种 DBMS 如何实现各种机制的内容，例如安全和查询处理等。

附录 I 借助 C 语言示例程序说明嵌入式和动态 SQL，还介绍了开放数据库互连 (ODBC)，这一标准现在已经成为访问异构 SQL 数据库的业界标准。

附录 J 讨论如何估计 Oracle 数据库的磁盘空间需求。

附录 K 概述面向对象的主要概念。

附录 L 提供若干 Web 脚本示例，补充第 29 章关于 Web 和 DBMS 的讨论。

附录 M 介绍交互式查询语言 QBE (举例查询)，对于非专业用户来说，它是访问数据库时最易使用的语言之一。附录中将使用 Microsoft Office Access 来说明 QBE 的用法。

附录 N 给出第三代 DBMS 宣言。

附录 O 介绍 Postgres，它是一个早期的对象关系 DBMS。

本书的逻辑组织及建议的阅读路线见图 0-1。

纠错和建议

如此大部头的一本教材难免出现错误、分歧、遗漏和混乱，恳请各位读者为未来的再版和编辑留下你的意见。任何建议、纠错和建设性意见都可发邮件告诉我：thomas.connolly@uws.ac.uk。

致谢

这本书是我们在工业界、研究机构和学术界工作多年的结晶。要想列出在此过程中帮助过我们的所有人是很难的。我们在此对任何不巧被遗漏的人表示歉意。首先，我们把最特别的感谢和道歉送给我们的家人，这些年我们完全埋头工作，怠慢甚至忽略了他们。

我们想要感谢审阅本书早期版本的那些人：得克萨斯技术大学的 William H. Gwinn，位于莱斯特的德蒙福特大学的 Adrian Larnar，斯克莱德大学的 Andrew McGettrick，南加州大学的 Dennis McLeod，加州大学的 Josephine DeGuzman Mendoza，俄克拉何马大学的 Jeff Naughton，诺瓦东南大学的 Junping Sun，佐治亚理工大学的 Donovan Young，布拉德富大学的 Barry Eaglestone，IBM 的 John Wade，米兰理工大学的 Stephano Ceri，位于厄斯特松德的瑞中大学的 Lars Gillberg，位于哈利法克斯的圣玛丽大学的 Dawn Jutla，伦敦城市大学的 Julie McCann，北卡罗来纳州立大学的 Munindar Singh，英国赫斯利的 Hugh Darwen，法国巴黎大学的 Claude Delobel，英国雷丁大学的 Dennis Murray，格拉斯哥大学的 Richard Cooper，厄勒布鲁大学的 Emma Eliason，斯德哥尔摩大学和皇家技术学院的

⊖ 在线附录为付费内容，需要的读者可向培生教育出版集团北京代表处申请购买。——编辑注

Sari Hakkarainen, 芝加哥洛约拉大学的 Nenad Jukic, 安特卫普大学的 Jan Paredaens, 丹尼尔·韦伯斯特学院的 Stephen Priest 以及来自我们系的 John Kawala 和 Peter Knaggs, 还有许多人, 谢谢你们花了那么多时间看我们的书稿。我们也想感谢 Anne Strachan 对第 1 版的贡献。

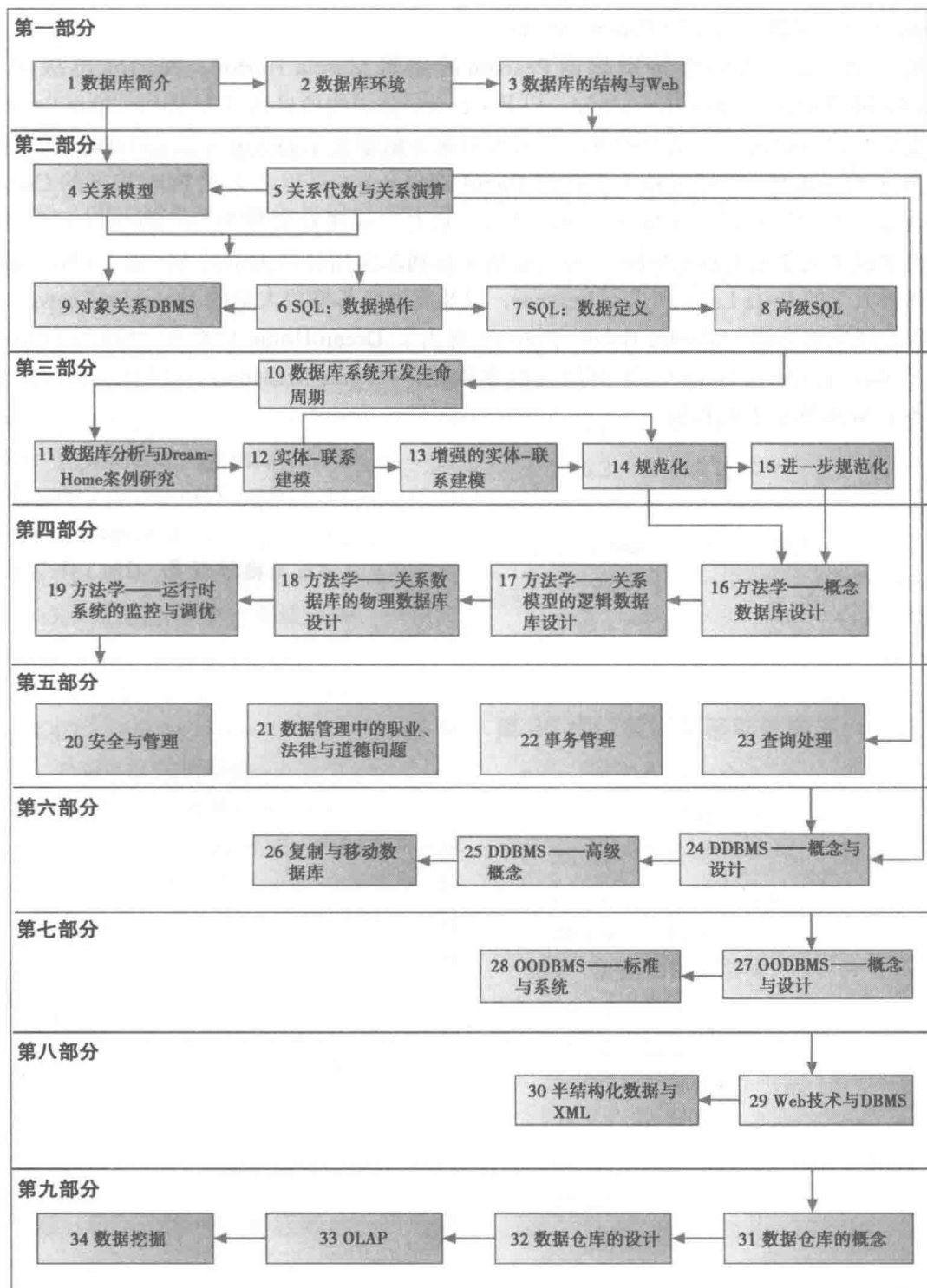


图 0-1 本书的逻辑组织及建议的阅读路线

我们还要感谢拉马尔大学的 Kakoli Bandyopadhyay, 北得克萨斯大学的 Jiangping Chen, 苟地比肯学院的 Robert Donnelly, 多明尼克大学的 Cyrus Grant, 华盛顿大学的 David G. Hendry, 斯特林大学的 Amir Hussain, 俄克拉何马州立大学的 Marilyn G. Kletke, 伦敦城市大学 CCTM 系知识管理研究组的 Farhi Marir, 印第安纳大学伯明顿主校区的 Javed Mostafa, 曼彻斯特大学的 Goran Nenadic, 旧金山州立大学的 Robert C. Nickerson, 丹佛大学的 Amy Phillips 和劳伦斯技术大学的 Pamela Smith。

关于第 6 版, 我们想特别感谢 Pearson 的编辑 Marcia Horton, Pearson 出版团队的 Kayla Smith-Tarbox 和 Marilyn Lloyd, 以及 Cenveo 公司的项目负责人 Vasundhara Sawhney。我们也要感谢下面这些人做出的贡献: 瑞典斯德哥尔摩大学的 Nikos Dimitrakas, 卡迪夫大学的 Tom Carnduff, 匹斯堡技术学院的 David Kingston, 马里兰大学 Park 校区的 Catherine Anderson, 罗切斯特理工学院的 Xumin Liu, 澳大利亚国立大学的 Dr. Mohammed Yamin, 伊兹密尔经济大学的 Cahit Aybet, 芬兰哈格 - 赫利尔应用科学大学的 Martti Laiho, 德国罗伊特林根大学的 Fritz Laux 和 Tim Lessner, 以及英国西苏格兰大学的 Malcolm Crowe。

我们还应该感谢 Malcolm Bronte-Stewart 提出了 DreamHome 的想法, Moira O'Donnell 保证了 Wellmeadows Hospital 案例研究的准确性, Alistair McMonnies 和 Richard Beeby 帮助准备了 Web 网站上的材料。

Thomas M. Connolly
Carolyn E. Begg
格拉斯哥, 2013 年 2 月

出版者的话
译者序
前言

第六部分 分布式 DBMS 与复制

第 24 章 DDBMS——概念与设计 2

24.1 引言	3
24.1.1 概念	3
24.1.2 DDBMS 的优缺点	6
24.1.3 同构 DDBMS 和异构 DDBMS	8
24.2 网络概述	10
24.3 DDBMS 的功能和体系结构	13
24.3.1 DDBMS 的功能	13
24.3.2 DDBMS 的参考体系结构	13
24.3.3 联邦 MDBS 的参考体系结构	14
24.3.4 DDBMS 的组成结构	15
24.4 分布式关系数据库设计	16
24.4.1 数据分配	17
24.4.2 分段	18
24.5 DDBMS 的透明性	24
24.5.1 分布透明性	24
24.5.2 事务透明性	26
24.5.3 性能透明性	29
24.5.4 DBMS 透明性	31
24.5.5 DDBMS 的透明性小结	31
24.6 Date 关于 DDBMS 的 12 条 规则	31
本章小结	32
思考题	33
习题	34

第 25 章 DDBMS——高级概念 35

25.1 分布式事务管理	36
25.2 分布式并发控制	36
25.2.1 目标	36

25.2.2 分布串行性	37
25.2.3 锁协议	37
25.3 分布式死锁管理	39
25.4 分布式数据库恢复	42
25.4.1 分布式环境中的故障	42
25.4.2 故障对恢复的影响	43
25.4.3 两段式提交	44
25.4.4 三段式提交	48
25.4.5 网络分区	51
25.5 X/Open 分布式事务处理模型	53
25.6 分布式查询优化	55
25.6.1 数据定位	56
25.6.2 分布式连接	58
25.6.3 全局优化	59
25.7 Oracle 中的分布特性	62
本章小结	67
思考题	67
习题	68

第 26 章 复制与移动数据库 69

26.1 数据库复制简介	70
26.1.1 复制的应用	71
26.1.2 复制模型	71
26.1.3 复制协议的功能模型	71
26.1.4 一致性	72
26.2 复制的体系结构	73
26.2.1 基于内核的复制	73
26.2.2 基于中间件的复制	73
26.2.3 更新处理	74
26.2.4 更新传播	75
26.2.5 更新场所 (数据所有权)	76
26.2.6 终止协议	78
26.3 复制模式	79
26.3.1 积极主备份	79
26.3.2 懒惰主备份	83
26.3.3 积极随处更新	86

26.3.4	懒惰随处更新	86
26.3.5	使用统一全序广播的随处更新	89
26.3.6	SI 与统一全序广播复制	92
26.4	移动数据库简介	97
26.4.1	移动 DBMS	98
26.4.2	与移动 DBMS 相关的问题	98
26.5	Oracle 中的复制	108
本章小结		113
思考题		114
习题		114
第七部分 对象 DBMS		
第 27 章	OODBMS——概念与设计	118
27.1	下一代数据库系统	119
27.2	OODBMS 简介	120
27.2.1	面向对象 DBMS 的定义	120
27.2.2	函数数据模型	122
27.2.3	持久型编程语言	125
27.2.4	开发 OODBMS 的可选策略	126
27.3	OODBMS 的持久性	127
27.3.1	指针切换技术	128
27.3.2	访问对象	131
27.3.3	持久性模式	132
27.3.4	正交持久性	133
27.4	OODBMS 中的问题	135
27.4.1	事务	135
27.4.2	版本	135
27.4.3	模式演化	136
27.4.4	体系结构	139
27.4.5	基准测试	141
27.5	OODBMS 的优点和缺点	142
27.5.1	优点	143
27.5.2	缺点	144
27.6	ORDBMS 与 OODBMS 的比较	145
27.7	面向对象数据库设计	146
27.7.1	面向对象数据建模与概念 数据建模的比较	146
27.7.2	联系和引用完整性	147
27.7.3	行为设计	149

27.8	采用 UML 的面向对象分析与 设计	150
27.8.1	UML 图	151
27.8.2	UML 在数据库设计方法学 中的应用	155
本章小结		156
思考题		157
习题		158

第 28 章 OODBMS——标准与系统 159

28.1	对象管理组	159
28.1.1	背景	160
28.1.2	公共对象请求代理架构	162
28.1.3	其他 OMG 规范	166
28.1.4	模型驱动的架构	167
28.2	对象数据标准 ODMG 3.0	170
28.2.1	对象数据管理组	170
28.2.2	对象模型	171
28.2.3	对象定义语言	177
28.2.4	对象查询语言	178
28.2.5	ODMG 标准的其他部分	183
28.2.6	将概念设计映射为逻辑 (面向对象的)设计	186
28.3	ObjectStore	187
28.3.1	总体结构	187
28.3.2	构建 ObjectStore 应用程序	189
28.3.3	ObjectStore 中的数据定义	190
28.3.4	ObjectStore 中的数据操作	193
本章小结		196
思考题		197
习题		197

第八部分 Web 与 DBMS

第 29 章 Web 技术与 DBMS 200

29.1	Internet 和 Web 简介	201
29.1.1	企业内联网与外联网	202
29.1.2	电子贸易和电子商务	203
29.2	Web	203
29.2.1	超文本传输协议	204