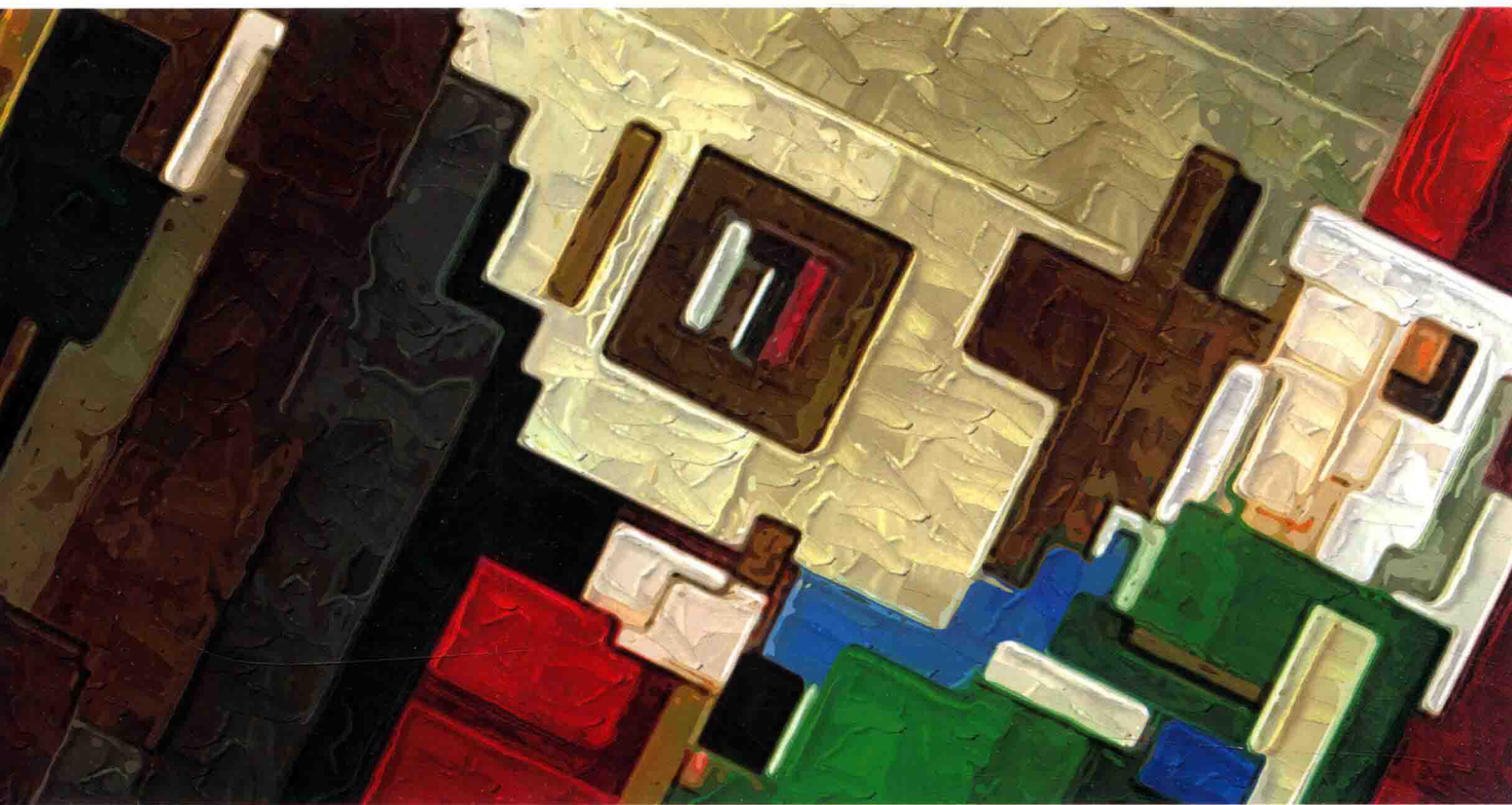


数据库原理

(第六版)

戴维·M. 克伦克 戴维·J. 奥尔 / 著
David M. Kroenke David J. Auer

张孝 / 译



Database Concepts
(Sixth Edition)



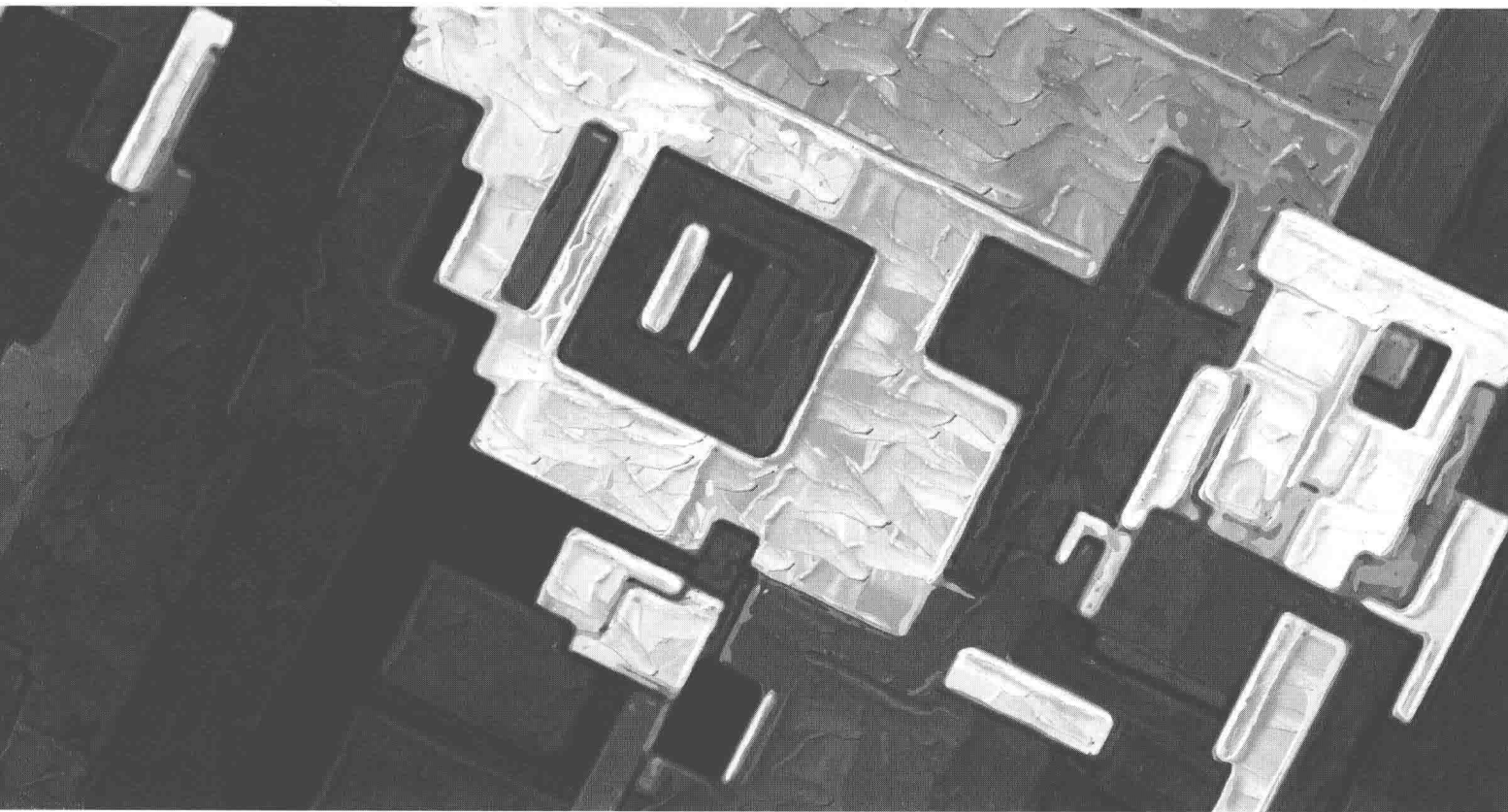
信息管理与信息系统
引进版教材系列

数据库原理

(第六版)

戴维·M. 克伦克 戴维·J. 奥尔 / 著
David M. Kroenke David J. Auer

张孝 / 译



Database Concepts
(Sixth Edition)

中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

数据库原理：第六版/戴维·M. 克伦克，戴维·J. 奥尔著；张孝译. —北京：中国人民大学出版社，2017. 1
信息管理与信息系统引进版教材系列
ISBN 978-7-300-23685-8

I. ①数… II. ①戴… ②戴… ③张… 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 (2016) 第 282723 号

信息管理与信息系统引进版教材系列
数据库原理 (第六版)

戴维·M. 克伦克 著

戴维·J. 奥尔

张孝 译

Shujuku Yuanli

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 涿州市星河印刷有限公司

规 格 215 mm×275 mm 16 开本

印 张 27.25 插页 1

字 数 822 000

邮政编码 100080

010-62511770 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2017 年 1 月第 1 版

印 次 2017 年 1 月第 1 次印刷

定 价 65.00 元

版权所有 侵权必究

印装差错 负责调换



科林·约翰逊是西雅图的一个小制造商的一名生产主管。几年前，科林想建立一个数据库来跟踪产品包中的组件。当时，他用电子表格方式来完成这项任务，但他无法从电子表格中得到他想要的报告。科林听说了 Microsoft Access，于是试图用它来解决他的问题。经过几天的失败与挫折，他买了几本流行的有关 Microsoft Access 的书，并试图从中学习如何使用 Access。但是他最终放弃了自己来尝试的想法而是请了一个顾问帮他构建有关的应用，这或多或少地满足了他的需求。

作为一个成功的商人，科林有强烈的愿望来实现自己的目标。作为一位经验丰富的 Windows 用户，他已经自学了如何使用 Microsoft Excel、Microsoft PowerPoint 以及其他一些面向生产的应用包。他很懊恼自己不能很好地利用 Microsoft Access 来解决自己的问题。“我相信我也能够用好 Access，但是不想投入太多的时间”，他想。这个故事有些特别值得注意的地方，因为这样的事情在过去的十年里或许已经在不同的人身上发生了数万次。

微软、甲骨文、IBM 和其他数据库管理系统 (DBMS) 供应商都意识到了这种情况并投入数百万美元来实现更好的图形界面、成百个多面板向导和许多示例应用程序。遗憾的是，这些努力都是治标不治本的。事实上，大多数用户都没有搞清楚这些向导正在代表他们干什么。一旦他们需要更改数据库结构或组件，比如表单和查询，他们将会淹没在这个复杂性的海洋中，措手不及。由于他们对基础知识知之甚少，这些用户会抓走任何一根似乎通向他们想要的方向的稻草。其后果是设计得比较糟糕的数据库和应用，以至于难以满足用户的实际需要。

为什么像科林这些人可以学会使用文字处理器或电子表格产品，却学不会使用 DBMS 产品呢？原因可能在于：首先，对大多数人而言，那些基本的数据库概念是有些特别的。虽然大家都知道段落和边距是什么，但几乎没有人知道“关系是什么”。其次，似乎使用 DBMS 产品应该是比实际情况更容易。比如，有人问：“我想做的是跟踪某些东西。为什么这么难？”对没有关系模型知识的商务用户而言，在保存数据前就把一张销售发票拆分到五个单独的表中进行存储令他们感到很神秘。

这本书旨在帮助科林等人来理解、创建和使用某些 DBMS 产品的数据库。无论是在书店发现本书的个人，还是使用本书作为他们的教科书的学生，都可以这样获得提高。

新版本

学生和其他读者都将从本书这一版本的新内容和功能中受益。这些新内容或功能包括：

- 大数据是第 8 章中的新主题，其中包括了 NoSQL 运动的相关材料、非关系型的结构化数据存储（例如 Cassandra 和 HBase）的发展以及 Hadoop 分布式文件系统（HDFS）。



● 第7章 Web 数据库应用也支持数据输入 Web 表单页面。这样, Web 数据库应用就可以同时使用数据输入和数据读取网页来构建。

● Microsoft Access 2010 包括 Microsoft Access 切换面板 (switchboard) 表单 (参见附录 H “The Access Workbench—Section H—Microsoft Access 2010 Switchboards”), 它可以用来构建数据库应用的菜单。切换面板表单在数据库应用中是有帮助的; 它提供一个用户友好的主菜单, 简化了用户需要显示表单、打印报告以及执行查询的步骤。

● 每章都新设有一个独立的案例题集。案例题集中的问题一般不要求学生在前面的章节中已经完成同一个案例的工作, 但有一个有意安排的例外, 即关系数据建模和数据库设计要一起完成。

● 支持 Oracle 数据库 11g 第2版的速成 (Express) 版。新增加的附录 B——“Getting Started With Oracle Database 11g Release 2 Express Edition”, 将展示如何使用该产品以及 Oracle SQL 开发者图形用户界面 (GUI) 工具。

● 更新有关内容以反映新的 Microsoft SQL Server 2012 的速成版。虽然大多数主题向上兼容 Microsoft SQL Server 2008 R2 的速成版, 但本书现在使用的 SQL Server 2012 能够专门与 Office 2010 充分结合。

● 更新的内容已经反映 MySQL 5.5 和 MySQL 工作台的使用情况。

新增在线附录 J “Business Intelligence Systems”, 提供了与第8章中商务智能相关的详细材料。这些资料可供导师进一步深入这一主题。

必要的基本概念

面对今天的技术, 已经不可能成功使用 DBMS 而不预先学习一些基本概念。经过多年与商业用户开发数据库, 我们认为以下数据库概念是必不可少的:

- 关系模型的基础;
- 结构化查询语言 (SQL);
- 数据建模;
- 数据库设计;
- 数据库管理。

而且由于越来越多地使用互联网、万维网、常用的分析工具, 以及大数据和 NoSQL 运动的出现, 四个更本质的概念也需要考虑在内:

- Web 数据库处理;
- 数据仓库结构;
- 商务智能 (BI) 系统;
- 非关系型的结构化数据的存储和处理。

像科林这样的用户和将执行类似工作的学生都不必像未来从事信息系统的专业人员那样来学习这些主题。因此, 本教材只介绍那些本质概念, 这些概念是科林这种想要创建和使用小型数据库的用户所必需的。本书中的许多讨论都是重写和简化解释的主题, 其中很多主题已经在戴维·克伦克和戴维·奥尔的《数据库处理: 基础、设计和实现》一书中充分讨论了。^① 但是在撰写有关材料时, 我们一直在努力保证讨论仍然准确而不具误导性。如果学生已经学过一些高级的数据库课程, 也就没有不需要学习的东西了。

^① David M. Kroenke and David J. Auer, *Database Processing: Fundamentals, Design, and Implementation*, 12th edition (Upper Saddle River, NJ: Pearson/ Prentice Hall, 2012).



讲授与 DBMS 产品无关的概念

本书不假定学生将使用任何特定的 DBMS 产品。本书用 Microsoft Access 2010、Microsoft SQL Server 2012、Oracle 数据库 11g 第 2 版和 Oracle MySQL 5.5 来说明数据库的概念，以便学生可以将这些数据库产品作为工具并实际实验有关材料，但所有介绍的概念都是与 DBMS 无关的。当学生用这种方式学习有关材料时，他们会逐渐了解这些基本知识适用于任何数据库，无论是其中最小型的 Microsoft Access 还是最大型的 Microsoft SQL Server 或 Oracle 数据库。

这并不是说，这门课程不能使用某个 DBMS。恰恰相反，学生通过使用商业 DBMS 产品可以更好地掌握这些概念。该版包括了 Microsoft Access、SQL Server、Oracle 和 MySQL 足够的基本信息，这样读者就可以在课程中仅使用这些产品，而不需要第二本书或其他材料。

由于 Microsoft Access 广泛用于入门类数据库课程，所以每一章都随附一节“Access 工作台”，通过 Microsoft Access 来说明本章的概念和技术。“Access 工作台”从第 1 章中创建一个数据库和一个单表的主题开始，进而涉及各种不同主题，最后在第 7 章中介绍基于 Microsoft Access 数据库的 Web 数据库处理，以及在第 8 章使用 Microsoft Access 和 Microsoft Excel 来生成数据透视表（PivotTable）OLAP 报告。Microsoft Access 的材料涵盖了所有必要的基本内容以便学生能够有效地建立和使用 Microsoft Access 数据库，但不准备对 Microsoft Access 做全面的叙述。

如果你需要比本书目前内容更深入地介绍 Microsoft Access 或其他 DBMS 产品，你可能需要在本书基础上用针对某个特定的 DBMS 的教材或其他材料进行补充。

关键术语、复习题、练习、案例和项目

因为让学生应用他们所学的概念非常重要，所以每章都包括了关键术语、复习题、练习（其中包括“Access 工作台”相关的练习）、案例问题以及贯穿全书的三个项目。学生在阅读并理解本章内容之后应该知道每个关键术语的含义并能够回答复习题。每个练习要求学生在一个小问题或任务上运用所学的概念。

本书所介绍的三个项目丽园、詹姆斯河珠宝以及安妮女王古玩店都是正在进行的项目并贯穿本书所有章节。在每个实例中，要求学生申请的项目能够适用本章项目概念。教师能够在《教师手册》中找到有关这些项目使用的更多信息，并能从本书网站（www.pearsonhighered.com/kroenke）的密码保护的教师部分获得这些数据库和数据。

本书使用的软件

就像我们以与 DBMS 无关的方式展开讨论一样，只要有可能，我们会选择尽可能独立于操作系统的软件。令人吃惊的是很多优秀软件都可以从网上找到。



因此尽管本书中的例子是用微软操作系统、SQL Server 2012 的速成版、Microsoft Access 2010、Microsoft Excel 2010 和 IIS Web 服务器创建的,但其中大部分可以很容易地通过使用 Linux、MySQL 5.5 的社区服务器版、OpenOffice.org 的 Base 和 Calc,以及 Apache Web 服务器来完成。另外,书中使用的其他一些软件产品,如 PHP 和 Eclipse 等也都可用于多个操作系统。

新版的变化

该版本中最显著的变化是涉及了目前迅速发展的有关大数据的使用情况和相关的 NoSQL 运动。能够存储和处理极大的数据集的需要正在改变数据库世界。虽然这些发展使得本书中介绍的基础知识能够保持不变,但是需要读者对本书核心的关系数据库与非关系型的结构化存储的理解,并为读者提供在大数据环境中的使用。因此,现在的第 8 章围绕大数据来组织,同时数据仓库、数据库服务器集群、分布式数据库以及商务智能 (BI) 系统概述等主题也都在该章中找到了合适的位置。为了找到与前一版的《数据库原理》(*Database Concepts*) 中关于 BI 的相同介绍,我们已经将 BI 的有关资料从第 8 章移到了附录 J。

另一个显著变化是本书增加了对 Oracle 数据库 11g 第 2 版的支持。虽然《数据库原理》一书一直专注于概念而不是特定的 DBMS 产品,但是我们也提供了对 Microsoft Access 2010、Microsoft SQL Server 2012 和 MySQL 5.5 的充分介绍以便这些概念能够被付诸实践。我们已经扩大了介绍范围,其中包括了同级别的教材对 Oracle 数据库的引用和说明,并添加了一个新的附录介绍 Oracle 数据库 11g 第 2 版的速成版和 Oracle SQL Developer GUI 实用程序(见附录 B “Getting Started with Oracle Database 11g Release 2 Express Edition”)。

最后,我们还新增了与各章独立的案例题集。虽然每章的项目紧密绑在一起,但案例问题不要求学生已经完成前面章节中同一案例的工作。但有一个贯穿第 4 章和第 5 章的特别例外将关系数据建模和数据库设计捆绑在一起,要一起完成,但每个章节还各自包括一个独立的案例。虽然在某些情况下不同章节中可能使用相同的基本的指定案例,但每个实例仍然完全独立于任何其他实例,同时我们在网站 www.pearsonhighered.com/kroenke 提供了所需的 Microsoft Access 2010 数据库和 SQL 脚本文本。

当然,我们还更新了本书中的所有其他产品的信息。具体而言,涵盖了 MySQL 5.5 和新发布的 Microsoft SQL Server 2012。

我们保留和改进了本书之前版本中的几个特点:

- 在每章使用小节“Access 工作台”来介绍 Microsoft Access 基本知识,现在包括 Microsoft Access 切换面板 (Switchboards) (附录 H “The Access Workbench—Section H—Microsoft Access 2010 Switchboards”,也可在线获得)。
- Microsoft SQL Server 2012 速成版 (附录 A “Getting Started with Microsoft SQL Server 2012 Express Edition”,在线获得) 和 MySQL 社区服务器版 5.5 (附录 C “Getting Started with MySQL 5.5 Community Server Edition”,在线获得) 使用导引。
- 使用贯穿全书的三个示例数据库的充分开发的数据集:韦奇伍德太平洋公司、希瑟·斯威尼设计以及 Wallingford 汽车公司。
- 关于 Web 数据库处理的各主题所使用的 PHP 脚本语言和 Eclipse IDE 中包括了网页代码输入表单。
- 重组后的第 8 章仍然保留了多维数据库模型和 OLAP 的介绍。

为了给这些新资料提供空间,我们不得不将以前一些有价值的资料移到本书的在线附录中。其中就包括了詹姆斯河珠宝 (James River Jewelry) 项目问题集,它现在是在线附录 D “James River Jewelry Project



Questions”;关于 SQL 视图的资料（上一版第 3A 章），现在是在线附录 E “SQL Views”；以及有关报表系统和数据挖掘的商务智能系统的材料，现在是在线附录 J “Business Intelligence Systems”。

本书概述

本教材由 8 章和 10 个在线附录组成（全部附录可从 www.pearsonhighered.com/kroenke 随时获得）。第 1 章解释了为什么使用数据库、它们的组件是什么，以及它们是如何开发的。学生将了解使用数据库的目的及其应用，以及数据库与电子表格存在哪些不同和改进。第 2 章介绍了关系模型并定义了基本的关系术语。同时还介绍了有关规范化和规范化过程的基本思想。

第 3 章给出了基本的 SQL 语句。介绍了用于数据定义的基本 SQL 语句、SQL SELECT 和数据修改语句。本章只有必需的语句，而没有介绍高级 SQL 语句。在线附录 E 中增加了对 SQL 视图的介绍。

接下来的两个章节涉及数据库的设计。第 4 章讨论使用实体—联系（E—R）模型来进行数据建模。这一章描述了数据建模需求，介绍 E—R 的基本术语和概念并给出了一个短小的 E—R 建模的案例应用（希瑟·斯威尼设计）（Heather Sweeney Designs）。第 5 章描述了数据库设计并解释了规范化的本质。第 4 章中案例的数据模型在第 5 章中转换成关系型设计。

我们在这一版本中继续使用在较早版本中规范化的更有效的讨论。我们提出了一个描述性过程，能够通过四个步骤完成关系的规范化。这种方法不仅使规范化任务更容易，也使得规范化原则更容易理解。因此，本书保留了这种做法。如果教师想了解更详细的范式细节，可以参阅第 5 章，其中包括了大多数范式的简短定义。

最后三章考虑数据库管理和应用中的数据库使用问题。第 6 章为数据库管理提供了一个概述，综述了并发控制、安全、备份和恢复技术。数据库管理是一个重要的话题，因为它适用于所有的数据库，包括个人和单用户数据库。

第 7 章介绍了基于 Web 的数据库处理的应用，包括对开放式数据库连接（ODBC）的讨论和 PHP 脚本语言的使用，还讨论了可扩展标记语言（XML）的出现和基本概念。

第 8 章讨论新兴的大数据和 NoSQL 运动。讨论了商务智能系统（BI）和数据仓库的架构，但 BI 系统的许多细节已被移到在线附录 J。第 8 章还讨论了多维数据库。本章给出了如何构建希瑟·斯威尼设计的一个多维数据库，然后用它生成一个数据透视表 PivotTable 的联机分析处理（OLAP）的报告。

附录 A 提供了一个 SQL Server 2012 速成版的简要介绍，附录 B 简要介绍了 Oracle 数据库 11g 第 2 版速成版，附录 C 则为 MySQL 5.5 提供了类似的介绍。Microsoft Access 则出现在每章都包括的“Access 工作台”。附录 D 现在包含了詹姆斯河珠宝项目问题。附录 E 介绍 SQL 视图语句。附录 F 介绍了系统分析和设计，这给第 4 章（数据建模）和第 5 章（数据库设计）提供了技术环境。附录 G 简要介绍了 Microsoft Visio 2010，Visio 2010 可以用于数据建模（第 4 章）和数据库设计（第 5 章）。另一个有用的数据库设计工具是 MySQL 工作台。在附录 C 中讨论 MySQL 工作台。附录 H 扩充了第 5 章的“Access 工作台”部分，增加了对 Microsoft Access 2010 切换面板的介绍。附录 I 用来补充第 7 章的内容，详细说明了启动和运行微软 IIS Web 服务器、PHP 和 Eclipse PHP 开发工具（PDT）的细节。最后，附录 J 为商务智能（BI）系统提供补充材料，并为第 8 章的报告系统和数据挖掘提供了更多的讨论。

为了使《数据库原理》各版本之间的更新保持一致，我们会根据需要将更新发布在网站 www.pearsonhighered.com/kroenke。教师资源和学生资料也可在网站上获得，所以最好不时检查一下该网站。



致 谢

在过去的 30 多年中,数据库和数据库应用一直是一个愉快而有回报的活动。我们相信,数据库应用及其支撑数据库的重要性在未来将不断提升,就像目前在谷歌 (Google)、Facebook 和其他产品中不断发展的大数据结构所展示的那样。而这些是本书第一版出版时所没有的。我们希望在这本书中所提到的概念、知识和技术能够帮助学生成功地参与到这个新兴的数据库世界中。

我们要衷心感谢以下审稿人所给出的富有洞察力和有益的意见:

David Chou, 东密歇根大学

Geoffrey Decker, 北伊利诺伊大学

Deena Engel, 纽约大学

Jean Hendrix, 阿肯色大学蒙蒂塞洛分校

Malini Krishnamurthi, 美国加利福尼亚州立大学富勒尔分校

Rashmi Malhotra, 圣约瑟夫大学

Gabriel Petersen, 北卡罗来纳中央大学

Eliot Rich, 纽约州立大学奥尔巴尼分校

Bond Wetherbe, 美国得克萨斯州理工大学

Diana Wolfe, 俄克拉何马城俄克拉何马州立大学

要感谢我们的编辑 Bob Horan; 我们的编辑项目经理 Kelly Loftus; 生产项目经理 Jane Bonnell; 项目经理 Jennifer Welsch。感谢他们在项目推进中所表现的敬业精神、洞察力、支持和帮助。我们还要感谢 Marcia Williams 为终稿所提出的详细的意见。最后,戴维·克伦克想感谢他的妻子 Lynda; 戴维·奥尔想感谢他的妻子 Donna。感谢她们在完成这个项目的过程中所表现的爱、鼓励和耐心。

戴维·M. 克伦克, 华盛顿州西雅图

戴维·J. 奥尔, 华盛顿州贝林汉姆



作者简介

戴维·M. 克伦克，1967年作为兰德（RAND）公司的暑期实习生进入电脑行业。从那时起，他的职业生涯已经跨越了教育、工业、咨询和出版。

他曾任教于华盛顿大学、科罗拉多州立大学和西雅图大学。多年来，他为大学教授主办了几十场教学研讨会。1991年国际信息系统协会授予他“年度电脑教育家”的称号。

在工业领域，克伦克曾在美国空军和波音公司的电脑服务部门工作，他是三家创业公司的法人。他也是 Microrim 公司产品营销和开发部的副总裁和数据库部门的首席技术专家，同时也是语义对象数据模型之父。克伦克的咨询客户包括 IBM 公司、微软、计算机科学公司，以及许多其他公司和组织。

他写的教科书包括：《数据库处理：基础、设计和实施》（*Database Processing: Fundamentals, Design, and Implementation*），首次出版于1977年，现在是第十二版。他于2003年撰写了《数据库的概念》（*Database Concepts*，现在你正在阅读第六版）。克伦克还出版了许多其他教科书，包括经典的《商业计算机系统》（*Business Computer Systems*）（1981）。最近，他创作了三本书：《MIS 经验》（*Experiencing MIS*）（第三版）、《MIS 必备》（*MIS Essentials*）（第二版）以及《使用 MIS》（*Using MIS*）（第五版）。

作为一个狂热的水手，克伦克还编写了：*Know Your Boat: The Guide to Everything That Makes Your Boat Work*。克伦克住在华盛顿西雅图，已婚，有两个孩子和三个孙子。

戴维·J. 奥尔，1994年以来，一直在西华盛顿大学经济及工商管理学院（CBE）担任信息系统和技术服务主任，并担任 CBE 决策科学系的讲师。自1981年以来，他曾任教 CBE 课程的定量方法、生产和经营管理、统计、金融、管理信息系统。除了管理 CBE 的计算机、网络和其他技术资源，他还教授管理信息系统课程。他曾任教管理信息系统和业务数据库开发课程并负责开发 CBE 的网络基础设施的课程，包括计算机硬件和操作系统、电信和网络管理。他与人合著了几本有关 MIS 的教科书。

奥尔拥有华盛顿大学英语文学学士学位，西华盛顿大学数学和经济学学士学位、经济学硕士学位，以及辅导心理学硕士学位。他曾担任美国空军军官，还曾作为员工援助计划（EAP）的组织发展专家和治疗师。

奥尔和他的妻子 Donna 住在华盛顿州贝林汉姆。他有两个孩子和五个孙子。



第一部分 数据库基础

第1章 入 门	3
为什么要使用数据库?	3
什么是数据库系统?	12
第2章 关系模型	52
关 系	52
键的类型	55
空值问题	64
函数依赖和规范化	64
第3章 结构化查询语言	98
示例数据库	99
数据定义类 SQL	103
插入关系数据类 SQL	113
关系查询类 SQL	117
关系数据修改和删除类 SQL	143
表和约束修改及删除类 SQL	146
SQL 视图	149

第二部分 数据库设计

第4章 数据建模和实体—联系模型	197
需求分析	198
实体—联系数据模型	198
E—R 图	203
开发 E—R 图的例子	210



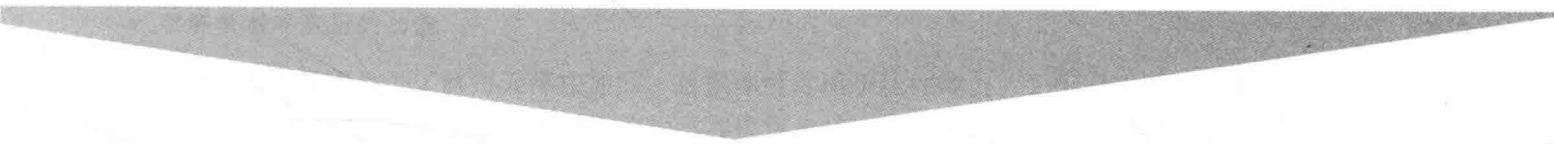
第 5 章 数据库设计	233
将数据模型转化成数据库设计	233
用关系模型表示实体	234
表示联系	243
希瑟·斯威尼设计中的数据库设计	255

第三部分 数据库管理

第 6 章 数据库管理	275
实施希瑟·斯威尼设计数据库	276
控制、安全和可靠性需求	276
并发控制	278
游标类型	287
数据库安全	288
数据库备份和恢复	294
DBA 的附加职责	297
第 7 章 数据库处理类应用	324
数据库处理环境	324
Web 数据库处理	328
数据库处理和 XML	352
第 8 章 大数据、数据仓库及商务智能系统	371
商务智能系统	373
业务系统和 BI 系统的联系	373
报表系统和数据挖掘应用	374
数据仓库和数据集市	375
联机分析处理 (OLAP)	383
分布式数据库处理	387
对象—关系数据库	389
大数据和 NoSQL 运动	389
在线附录	411
词汇表	412



第一部分 数据库基础



第一部分将介绍关系数据管理的基本概念和技术。第1章介绍数据库技术，讨论为什么使用数据库，并描述一个数据库系统的组成部分。第2章介绍关系模型和定义关系数据库中的关键术语。同时还给出了关系数据库设计的基本原则。第3章介绍结构化查询语言——创建和处理关系数据库的一个国际标准。

在学会这些基本的数据库概念之后，第二部分将集中于数据库建模、设计和实施。最后，本书第三部分将讨论数据库管理、Web数据库应用、大数据和商务智能（BI）系统。

第 1 章

入 门

本章目标

- 确定本书的目的和范围
- 了解使用列表的潜在问题
- 理解使用数据库的原因
- 了解如何使用相关表来避免使用列表时的问题
- 知道一个数据库系统的组成部分
- 学习数据库中的元素
- 了解一个数据库管理系统 (DBMS) 的用途
- 理解数据库应用的功能

对数据库技术重要性的了解每天都在加深。数据库可以用于任何地方：它们是电子商务和其他 Web 应用的关键组件。数据库位于整个组织范围级的运营和决策支持应用的“心脏”。数以千计的工作组和成百万的个人也在使用数据库。据估计，当今世界有超过 1 000 万的活跃数据库。

本书的目的是教给读者必要的数据库概念、技术和技巧。这些都是数据库开发人员开始自己的职业生涯所需要的。本书不会讲授数据库技术中所有的重要技术，但它会提供足够的背景知识使你能够创建自己的个人数据库，并能够作为一名团队成员参与更大、更复杂的数据库的开发。你也能够独自提出恰当的问题并自学得更多。

在第 1 章中，我们来调研使用数据库的原因。我们首先描述使用列表时可能出现的问题。通过一系列的例子来说明如何使用相关的表集以帮助你避免这些问题。接下来，我们将描述数据库系统的组件并解释数据库中的各个要素、数据库管理系统 (DBMS) 的目的以及数据库应用的功能。

为什么要使用数据库？

数据库能够帮助人们跟踪一些事情。可能令你感到疑惑的是：当使用一个简单的列表就可以达到这样的目的时，我们为什么还需要一个专门的学期（和课程）来掌握这样的技术呢？很多人使用列表来跟踪事情，而且这样的列表确实在某些时候很有价值。然而在另外一些情况下，简单的列表会导致数据不一致和其他问题。



在本节中，我们考察几个不同的列表，并展示其中可能存在的一些问题。正如你将看到的，我们可以通过将列表分割为数据表来解决问题。这些表是数据库的关键组成部分。本书的大部分内容涉及数据表的设计和表所包含的数据的有关技术。

列表存在的问题

图 1—1 显示了一个存储在电子表单中的简单的学生名单列表 Student。^① 学生名单是一个非常简单的列表，并且对这样的清单列表，电子表单很好用。即使名单很长，也可以按字母顺序对姓氏、名字或电子邮件地址排序，找到想要的任何条目。你可以更改数据值，添加一个新学生的数据，或删除学生数据。使用类似图 1—1 中的 Student 列表，这些操作都不是问题，并且没有必要使用数据库。使该列表以电子表单形式存储就很好。

	A	B	C
1	LastName	FirstName	Email
2	Andrews	Matthew	Matthew.Andrews@ourcampus.edu
3	Brisbon	Lisa	Lisa.Brisbon@ourcampus.edu
4	Fischer	Douglas	Douglas.Fischer@ourcampus.edu
5	Hwang	Terry	Terry.Hwang@ourcampus.edu
6	Lai	Tzu	Tzu.Lai@ourcampus.edu
7	Marino	Chip	Chip.Marino@ourcampus.edu
8	Thompson	James	James.Thompson@ourcampus.edu

图 1—1 电子表单中的学生列表

但是，假如我们改变学生列表，在其中加入导师数据（如图 1—2 所示）。你仍然可以用多种不同方法对新的包含导师的学生列表进行排序并从中查找一个表项。但如果对这个名单进行更改，则可能导致修改问题（modification problem）。例如，假设要删除的数据是学生 Chip Marino。如图 1—3 所示，如果恰好要删除第 7 行，那么这时不仅移除了 Chip Marino 的数据，同时也删除的事实是：一位名为 Tran 的导师以及 Tran 教授的电子邮件地址 Ken. Tran @ ourcampus. edu。

	A	B	C	D	E
1	LastName	FirstName	Email	AdviserLastName	AdviserEmail
2	Andrews	Matthew	Matthew.Andrews@ourcampus.edu	Baker	Linda.Baker@ourcampus.edu
3	Brisbon	Lisa	Lisa.Brisbon@ourcampus.edu	Valdez	Richard.Valdez@ourcampus.edu
4	Fischer	Douglas	Douglas.Fischer@ourcampus.edu	Baker	Linda.Baker@ourcampus.edu
5	Hwang	Terry	Terry.Hwang@ourcampus.edu	Taing	Susan.Taing@ourcampus.edu
6	Lai	Tzu	Tzu.Lai@ourcampus.edu	Valdez	Richard.Valdez@ourcampus.edu
7	Marino	Chip	Chip.Marino@ourcampus.edu	Tran	Ken.Tran@ourcampus.edu
8	Thompson	James	James.Thompson@ourcampus.edu	Taing	Susan.Taing@ourcampus.edu

图 1—2 包含导师信息的学生列表

类似地，更新此列表中的值也可能产生意想不到的后果。例如，如果改变第 8 行的 AdviserEmail，将产生不一致的数据。本次变动后，第 5 行表示 Taing 教授的一个电子邮件地址，而第 8 行则给同名教授列出了不同的电子邮件地址。或者他们是同一位教授吗？单从这份列表中，我们不能断定到底是只有一个 Taing 教授，而他有两个不一致的电子邮件地址，还是有一位 Taing 教授，他们有不同的电子邮件地址。由于执行该更新，我们使得列表中增加了混乱和不确定性。

最后，当我们要添加一个还没有学生的教授的数据时应该怎么做？例如，George Green 教授还没有学生，但我们仍然要记录他的电子邮件。如图 1—3 所示，我们必须插入一个值不完整的行。在数据库字段，

^① 为了方便识别和引用正在讨论的列表，我们会将本章中列表名的每个单词的第一个字母大写。同样地，我们将与列表相关联的数据库表的名称大写。

这些不完整的值称为空值（null value）。

	A	B	C	D	E
1	LastName	FirstName	Email	AdviserLastName	AdviserEmail
2	Andrews	Matthew	Matthew.Andrews@ourcampus.edu	Baker	Linda.Baker@ourcampus.edu
3	Brisbon	Lisa	Lisa.Brisbon@ourcampus.edu	Valdez	Richard.Valdez@ourcampus.edu
4	Fischer	Douglas	Douglas.Fischer@ourcampus.edu	Baker	Linda.Baker@ourcampus.edu
5	Hwang	Terry	Terry.Hwang@ourcampus.edu	Taing	Susan.Taing@ourcampus.edu
6	Lai	Tzu	Tzu.Lai@ourcampus.edu	Valdez	Richard.Valdez@ourcampus.edu
7	Marino	Chip	Chip.Marino@ourcampus.edu	Tran	Ken.Tran@ourcampus.edu
8	Thompson	James	James.Thompson@ourcampus.edu	Taing	Sue.Taing@ourcampus.edu
9	???	???	???	Green	George.Green@ourcampus.edu

图 1—3 学生列表（包含导师信息）的更新问题

在这种情况下，术语“空值”表示一种缺失值，但用于数据库时该术语还有其他含义。我们将在下一章中详细讨论有关空值的问题，我们将看到空值总是有问题的，而且我们要尽可能避免使用空值。

现在，在上面两个例子中究竟发生了什么？我们有一个三列的简单列表，又增加了两列，于是产生了几个问题。问题不仅仅是列表有五列，而非三列。考虑另外一个有五列的不同列表：包含宿舍（Residence）的学生列表，如图 1—4 所示。这个列表中有五列，但它没有遇到图 1—3 所示的包含导师信息的学生列表的任何问题。

在图 1—4 的包含宿舍信息的学生列表中，我们可以删除学生 Chip Marino 的数据，而且仅仅删除了该学生的数据。没有出现意想不到的后果。同样，我们可以改变学生 Tzu Lai 的宿舍值而不会导致任何不一致。最后，我们可以添加学生 Garret Ingram 的数据而且没有任何空值。

图 1—3 包含导师信息的学生列表和图 1—4 包含宿舍信息的学生列表之间存在本质上的区别。从这两个图中，你能确定什么区别吗？本质区别在于：图 1—4 包含宿舍信息的学生列表只关注一件事情：在该列表中的所有数据只和学生有关。相反，图 1—3 包含导师信息的学生列表则涉及两件事情：一些数据是关于学生的，而另外一些数据是关于导师的。一般说来，如果一个列表中涉及两个或两个以上的不同事物，就会产生数据修改问题。

	A	B	C	D	E
1	LastName	FirstName	Email	Phone	Residence
2	Andrews	Matthew	Matthew.Andrews@ourcampus.edu	301-555-2225	123 15th St Apt 21
3	Brisbon	Lisa	Lisa.Brisbon@ourcampus.edu	301-555-2241	Dorsett Room 201
4	Fischer	Douglas	Douglas.Fischer@ourcampus.edu	301-555-2257	McKinley Room 109
5	Hwang	Terry	Terry.Hwang@ourcampus.edu	301-555-2229	McKinley Room 208
6	Ingram	Garrett	Garrett.Ingram@ourcampus.edu	301-555-2223	Dorsett Room 218
7	Lai	Tzu	Tzu.Lai@ourcampus.edu	301-555-2231	McKinley Room 115
8	Marino	Chip	Chip.Marino@ourcampus.edu	301-555-2243	234 16th St Apt 32
9	Thompson	James	James.Thompson@ourcampus.edu	301-555-2245	345 17th St Apt 43

图 1—4 包含宿舍的学生列表

为了强调这个观点，我们考察图 1—5 的包含导师和学部信息的学生列表。这个列表中的数据关于三个不同的事物：学生、导师和学部。正如从图中可见，插入、更新和删除数据所引发的问题变得更糟。例如，要改变 AdviserLastName 的值，可能只需要修改 AdviserEmail，或者也可能需要同时改变 AdviserEmail、Department 和 AdminLastName 的值。可以想象一下，如果这个列表很长，比如有几千行，并且如果有多个人处理它，列表就会在很短的时间内变得一团糟。

使用关系数据库的表

使用列表的问题最初在 20 世纪 60 年代被确定，自此以后人们开发了许多不同的技术来解决这些问题。随着时间的推移，关系模型（relational model）作为一种方法论，成为其中的领导性解决方案，而且现在几