

计算机科学本科核心课程教材

Databases

A Beginner's Guide

Third Edition



数据库基础教程

(第3版)

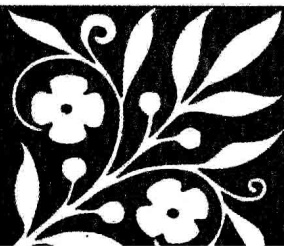
清华大学出版社

计算机科学本科核心课程教材

Databases

A Beginner's Guide

Third Edition



数据库基础教程

(第3版)

清华大学出版社
北京

Andy Oppel

Databases: A Beginner's Guide, Third Edition

EISBN: 978-0-07-160846-6

Copyright © 2009 The McGraw-Hill Companies, Inc.

All Rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including without limitation photocopying, recording, taping, or any database, information or retrieval system, without the prior written permission of the publisher.

This authorized Chinese translation edition is jointly published by McGraw-Hill Education (Asia) and Tsinghua University Press. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong, Macao SAR and Taiwan.

Copyright © 2010 by McGraw-Hill Education (Asia), a division of the Singapore Branch of The McGraw-Hill Companies, Inc. and Tsinghua University Press.

版权所有。未经出版人事先书面许可,对本出版物的任何部分不得以任何方式或途径复制或传播,包括但不限于复印、录制、录音,或通过任何数据库、信息或可检索的系统。

本授权中文简体字翻译版由麦格劳-希尔(亚洲)教育出版公司和清华大学出版社合作出版。此版本经授权仅限在中华人民共和国境内(不包括香港特别行政区、澳门特别行政区和台湾)销售。

版权© 2010 由麦格劳-希尔(亚洲)教育出版公司与清华大学出版社所有。

北京市版权局著作权合同登记号 图字:01-2009-3868

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。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

数据库基础教程:第3版/(美)奥佩尔(Oppel, A.)著;远红亮等译. —北京:清华大学出版社, 2010.11

书名原文: Databases: A Beginner's Guide, Third Edition

ISBN 978-7-302-22950-6

I. ①数… II. ①奥… ②远… III. ①数据库系统—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2010)第 105534 号

责任编辑:龙啟铭

责任校对:李建庄

责任印制:王秀菊

出版发行:清华大学出版社

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

社总机:010-62770175

投稿与读者服务:010-62795954, jsjic@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

地址:北京清华大学学研大厦 A 座

邮编:100084

邮购:010-62786544

印装者:北京鑫海金澳胶印有限公司

经销:全国新华书店

开本:185×230 印张:21.75

字数:471 千字

版次:2010 年 11 月第 1 版

印次:2010 年 11 月第 1 次印刷

印数:1~3000

定价:39.50 元

产品编号:032798-01

译者序

数据库技术是计算机科学中非常重要的一项技术,几乎所有的应用系统开发都会用到数据库技术;数据库的应用更是当今人们日常生活中不可或缺的部分,差不多每个人每天都要或多或少、直接或间接地使用到数据库。因此无论是计算机专业领域的工作者或是对 IT 技术感兴趣的人们,学习和掌握数据库知识是非常有益的。本书作为一本讲授数据库技术的入门教程,非常好地完成了指南的作用,它非常全面地介绍了数据库技术,涉及方方面面内容,全面而不失重点,对于数据库中至关重要的主题 SQL、范式化、数据建模等都给予了充分的讲解,除此之外本书的一大特色还在于:它使用非常流行的 Access 和 Oracle 作为样例数据库系统对读者进行实践操作练习指导,不至于使初学者在学习了一大堆理论之后对数据库技术还是一头雾水,另外本书后面几章对数据库的高级主题进行了介绍,对于想更进一步学习数据库技术的读者给予了方向性的指导。

本书总共分为 4 个部分,包含 13 章。第 1 部分介绍数据库基本概念,由 4 章组成,全面地介绍了数据库的基本知识,另外单独一章介绍 SQL 这个重要知识点;第 2 部分介绍数据库开发,也由 4 章组成,介绍了数据库的生命周期,从数据建模到物理数据库设计进行了全面介绍,另外单独一章对数据库最为重要主题的范式化进行了全面介绍。第 3 部分为数据库实现,介绍了数据库部署、数据库安全性、数据库与外部系统互连以及数据仓库和 SQL/XML 等高级主题,能够使读者非常全面地俯瞰到数据库技术的绝大多数内容;第 4 部分为附录部分,包含了各个章节的自我测试题答案以帮助读者进行自我检验,以及动手实践练习题解决方案。

本书主要由远红亮翻译，张艳艳对翻译文档进行了编辑整理工作，其中部分章节内容由张浩、张静文、林杰、蔡玲玲、李克新、刘捷、王辉、陈洁、张凯、王颖、王伟、刘海燕、李建伟、刘静、王鑫鑫、郑浩、孙峰、吴莎莎等翻译，在此对大家表示衷心感谢。

由于译者知识有限，书中的翻译错误和不妥之处在所难免，读者若发现翻译处理不当之处，欢迎批评指正。

关于作者

Andrew J. (Andy) Oppel 不仅是马里兰州男童拉丁学校的优秀毕业生，同时也是特兰西瓦尼亚大学的优秀毕业生，他于 1974 年在该大学获取计算机科学专业的学士学位。从那时起，他一直投身于信息技术的各个领域，包括程序员、程序员/分析师、系统架构师、项目经理、高级数据库管理员、数据库组经理和顾问、数据库设计师、数据建模员、数据架构师。另外，他还在加利福尼亚大学伯克利分校做了 20 多年兼职教员，并于 2000 年获得最受尊敬教员奖。在加利福尼亚大学伯克利分校，他开发了三门课程，“数据库管理系统概念”、“关系数据库管理系统导论”和“数据建模与数据库设计”。同时，他于 2003 年获得 Oracle 9i 数据库认证。现在他受雇于 Blue Shield of California 公司，做高级数据建模师。除了计算机系统领域的工作之外，Andy 喜欢音乐（吉他和主唱）、业余无线电（亚太部副主任、美国无线电中继联盟）和英式足球（裁判讲师、美式足球）。

Andy 在广泛的应用领域中进行过数以百计的数据库设计与实现工作，这些领域包括医学研究、银行、保险、服装制作、电信、无线通信和人力资源。他是 *Databases Demystified* 一书的作者（McGraw-Hill, 2004）和 *SQL Demystified* 的作者（McGraw-Hill, 2005），同时，他还是 *SQL: A Beginner's Guide*（McGraw-Hill, 2009）的合著者。他在数据库产品方面的经验包括 IMS、DB2、Sybase ASE、Microsoft SQL Server、Microsoft Access、MySQL 和 Oracle（版本 7、8、8i、9i 和 10g）。

如果有任何方面的问题，请通过 andy@andyoppel.com 联系 Andy。

关于本书的技术编辑

Todd Meister 使用微软技术做过 10 多年的开发工作。他作为技术编辑参与过 50 多本书，这些书的主题从 SQL Server 包含到 .NET 框架都有。除

了做技术编辑工作之外，他在印第安那州波尔州立大学的计算服务部做助理主任。他同妻子 Kimberly 和 4 个可爱的孩子住在印第安那州中部。Todd 的联系地址为 tmeister@sycamore resolutions.com。

致谢

非常多的人参与到本书的出版过程中——这其中大部分人的名字我还不清楚。第一，McGraw-Hill 的编辑和职员为本书花费了大量的时间和精力。这里我特别感谢编辑部主任 Wendy Rinaldi，他在我所有书的出版过程中给予了我最多的建议和精神支持。事实上，正是 Wendy 开启了我的 McGraw-Hill 作者之路。另外我想感谢 Lisa Theobald，她做了出色的复制编辑工作，以及所有的编辑们、校对员、编索引员、设计员、插图员和其他参与人员。我要特别感谢技术编辑 Todd Meister，他在本书细节方面的用心以及整书的编辑过程中所做的有帮助的工作。最后我要感谢我的家庭给予我的支持和理解，因为我在已经非常繁忙的生活当中插入了编写工作。

前言

FOREWORD

在 35 年前，数据库只有在特殊的研究实验室才能见得到，那时计算机科学家们想尽各种办法让数据库更加高效有用，并把他们的研究成果发表在无数的研究论文中。而今天数据库是信息技术 (IT) 产业中随处可见的东西。我们每天直接或间接地使用着数据库——银行事务、旅行预订、雇佣关系、网站搜索、在线或离线交易，以及其他大多数记录进数据库或有数据库服务的事务。

同大多数快速发展的技术一样，工业标准已经滞后于数据库技术的发展了，结果导致了大量的数据库产品，而每一个产品都体现着特定软件厂商的数据库视角。另外，出现了大量不同种类的数据库模型；其中关系模型是最为流行的。本书将介绍所有主要的数据库模型，包括层次模型、网络模型、关系模型、面向对象模型和对象关系模型。因为关系模型和对象关系模型是 IT 工业中的主流模型而且在可预见的未来它们都将是这样的，因此本书主要关注于关系模型和对象关系模型。

实现数据库最重要的工作在于正确地设计数据库结构。如果没有对数据库将要解决的问题有比较透彻的理解，没有关于数据组织的最佳实践知识，那么实现的数据库很可能变成可怕的工程，在项目完成后仍然要花费大量时间和精力。本书关注于需求到工作数据模型之间的转换工作，该过程称为范式化，该过程已经被证明是设计关系数据库非常有效的一项技术。事实上，范式化过程可以成功应用到其他数据库模型中。另外，基于一种思想，没有开过车的人不可能去设计汽车，本书将介绍结构化查询语言 (Structure Query Language, SQL)，这样可以使读者在开始设计数据库之前先“驾驶”数据库。我利用作为数据库设计者、管理员和教员方面的经验为读者提供自我帮助方面的指导，使读者能够顺利进入数据库技术这个迷人且复杂的世界。本书中包含的例子使用到了微软 Access 和 Oracle。本书中用到了这些厂商提

供的开放示例数据库（微软 Access 北风贸易公司和 Oracle 人力资源数据库模式）可以使读者在自己的计算机上随意地进行示例练习。每章的末尾有一个自我测试练习题可以帮助读者强化本章所学到的知识。

读者对象

本书适合于任何希望在数据库设计和管理上获取基础知识的人们，无论是个人还是专业人士。本书被专门设计适合于数据库技术的新手和相对新手；不过，希望进修范式化、数据库设计和管理知识的读者也会发现本书非常有用。无论你是有经验的开发人员，还是有一些开发经历，或者你是数据库管理员，或是编程和数据库方面的新手，本书为大家提供了非常强大的基础指导，对于希望学习更多关于数据库技术的读者来说是非常有用的。事实上，在试图理解和使用数据库的时候，下面这些人们会发现本书是非常有帮助的：

- 数据库设计和 SQL 编程的新手。
- 想要更好地理解如何设计、实现和访问数据库的分析师或经理。
- 想要学习更多关于数据库设计的数据库管理员。
- 必须完成 SQL 数据库特设查询的技术支持人员或测试/QA 工程师。
- 编写应用程序的 Web 程序员。
- 在应用程序源代码中嵌入 SQL 的第三代语言（3GL）的程序员。
- 任何想学习数据库设计并编写 SQL 代码来创建和访问数据库的人们。

无论你属于上面的哪一类，需要注意的是，本书专门为那些想要学习标准数据库设计技术的人们而设计，其内容并不专门针对任何具体的开发商产品。这使得读者能够将本书学到的知识应用于现实世界，而不必局限于任何的产品标准。当然，读者同时也要清楚自己所使用的数据库的具体实现情况，特别是具体的 SQL 语言，但是本书为读者提供的基础知识能够使读者轻松地从一个 RDBMS 转到下一个，帮助读者建立一个牢固的数据库设计理论。最终，读者将发现本书对于数据库新手来说是一个非常有用的工具，特别对于关系数据库新手来说，而不管他使用的哪个开发商的产品。读者可以很容易地将本书学到的知识应用于特定的 RDBMS。

本书内容

本书由 3 部分组成。第 1 部分介绍数据库基本概念以及如何在数据库中使用 SQL 语句创建和访问对象。第 2 部分介绍数据库开发的基础，包括数据库生命周期、使用范式化过程进行数据库逻辑设计、转换数据库逻辑设计为物理数据库、进行数据和过程建模。

第3部分着重介绍数据库实现中的安全问题,以及数据库高级主题——在线分析过程(OLAP)和集成对象与XML文档到数据库的问题,使你在第1、2部分的基础上更加深入地理解数据库。另外,附录中还包括自我检测题的答案和贯穿本书的体验练习题的解决方案。

内容概要

以下以章节任务为重点概要介绍本书的内容。

第1部分 数据库概念

第1部分介绍数据库基本概念以及如何在数据库中使用SQL语句创建和访问对象。

第1章 数据库基础 本章介绍数据库的基础概念和定义,包括数据库的共同属性、流行的数据库模型、数据库发展简史以及为何着重关注关系型数据库。

第2章 关系数据库组件之探讨 本章介绍构成关系模型的概念、逻辑和物理组件。概念数据库设计以技术独立的方式对数据进行研究和建模。逻辑数据库设计是概念数据库到逻辑数据库的一个转换和映射的过程,是选择数据库模型的过程(关系型、面向对象型、对象关系型等),最后一步设计是物理数据库设计,将逻辑数据库设计映射成为针对特定数据库管理系统的物理设计。

第3章 基于表单的数据库查询 本章通过微软Access中基于表单的查询工具概要介绍构造和执行数据库查询的概念,为后续章节介绍数据库设计理论提供基础。

第4章 SQL介绍 本章介绍的SQL语言,已成为普遍的关系数据库查询语言,几乎每一个现代的数据库管理系统都支持它。其原因是在SQL语言功能和标准的开发中投入了大量时间和精力,使SQL语言在关系型数据库管理系统产品中具有高度的可移植性。

第2部分 数据库开发

第2部分介绍数据库开发的基础,包括数据库生命周期、使用范式化过程进行数据库逻辑设计、转换数据库逻辑设计为物理数据库、数据和过程建模。

第5章 数据库生命周期 本章介绍数据库设计过程的整体框架,以更好地理解数据库设计的具体步骤。数据库生命周期(计算机系统生命周期)是一个描述所有事件发生的一个术语,是从了解数据库的需求,到数据库的开发和部署,最后到数据库停止服务的那一天为止的一个时间周期。

第6章 使用范式进行数据库设计 本章介绍使用范式进行逻辑数据库设计的过程。就关系型数据库而言,范式是本书最重要的一个主题,因为范式告诉我们怎样最好地设计数据表。

第7章 数据与过程建模 本章详细介绍实体-关系图和数据建模。本章第2部分还包括高级别调查的过程设计概念和图表技术。

第8章 物理数据库设计 本章介绍数据库设计人员的物理数据库设计工作，将逻辑数据库设计转换成一个或多个物理数据库设计。

第3部分 数据库实现

第3部分着重介绍数据库实现中的安全问题，以及数据库高级主题——在线分析过程（OLAP）和对象与XML文档与数据库整合问题，使你在第1、2部分的基础上更加深入地理解数据库。另外，附录中还包括自我检测题的答案和贯穿本书的体验练习题的解决方案。

第9章 数据库与外部互连 本章介绍数据库部署模式的发展，数据库逐渐地开始同数据库使用者以及企业计算集成连接（一种内部组织结构，组织企业的所有计算资源，包括数据库、应用程序、计算机硬件和计算机网络）中的其他计算机系统相互连接应用。本章介绍数据库同其他应用连接的方法——使用浏览器作为主要的用户接口，该方法已经成为大部分现代应用系统的方法。同时介绍当前数据库同应用程序连接的方法，包括ODBC连接（适用于大部分编程语言）和其他使用Java进行连接的方法（适用于面向对象编程语言）。

第10章 数据库安全性 本章介绍安全性需求，考虑数据库服务器和客户端的部署安全性、数据库访问安全性，以及安全监视和安全审计问题的探讨。

第11章 部署数据库 本章介绍一些关于数据库应用系统的开发问题，包括指针处理、事务管理、性能优化和变更控制。

第12章 在线分析处理数据库 本章介绍数据库分析过程的概念，包括数据仓库和数据集市、数据挖掘和其他数据分析技术，以及针对这些特殊数据库的数据库设计变化。

第13章 集成XML文档和对象入数据库 本章介绍一系列方法将XML和对象内容集成到数据库。

第4部分 附录

附录中包括自我检测题的答案和贯穿本书的体验练习题的解决方案。

附录A 自我测试题答案 本附录包括每章后面自我测试题的答案。

附录B 体验练习题解决方案 本附录包括出现在每一章中的体验练习题的解决方案，包括图和一些SQL代码。

章阐释

从上面概要部分可以看出，本书以章为单位进行组织。每章关注一些重要的技能和概念以及一些你需要理解这些概念和应用这些技能的背景信息。每章还额外包含一些帮助你理解这些信息的材料。

专家询问

每章包括一到两个专家询问的部分，提供一些关于本章内容可能会被询问到的一些问题。

自我测试

每章结尾有一个自我测试，有一系列题目帮助测试你对本章内容和技能的学习情况，自我测试答案见附录 A。

体验练习

大部分章节包含一到两个体验练习题，解决这些题目需要用本章学习到的知识。每个练习题被拆分成几个步骤，通过这几个步骤实现一个特定任务。这些练习包含的相关文件可以从 www.mhprofessional.com 下载。单击 **computing**，再单击页面左边的 **Downloads** 进入到下载页面，找到本书的列表，然后选择你需要的文件进行下载。这些文件包括 SQL 语句和一些图。

要完成这些体验练习，需要能够访问 RDBMS 同时拥有访问权限并运行 SQL 语句。如果你正在访问网络 RDBMS 数据库，你需要同数据库管理员联系确认你是否有创建数据库的权限，是否需要一些特别的权限来创建对象。同时，确认你是否需要一些特别的参数来创建数据库（例如，日志文件大小），是否限制你能够使用的命名和其他应用。在使用任何数据库产品前，请先查阅相关的文档。

目 录

CONTENTS

第 1 部分 数据库概念

第 1 章 数据库基础	3
1.1 数据库特性	3
1.1.1 数据库管理系统	4
1.1.2 数据抽象层次	5
1.1.3 物理数据独立性	7
1.1.4 逻辑数据独立性	8
1.2 流行的数据库模型	8
1.2.1 纯文本文件	8
1.2.2 层次模型	11
1.2.3 网状模型	12
1.2.4 关系模型	14
1.2.5 面向对象模型	15
1.2.6 对象关系模型	17
1.3 数据库简史	17
1.4 为什么关注于关系模型	20
第 1 章自我测试题	20
第 2 章 关系数据库组件之探讨	23
2.1 概念数据库设计组件	23
2.1.1 实体	24
2.1.2 属性	25
2.1.3 关系	25
2.1.4 业务规则	29
2.2 体验练习 2-1: 北风贸易公司数据库之探讨	30

2.3	逻辑/物理数据库设计组件	32
2.3.1	表	32
2.3.2	列和数据类型	34
2.3.3	约束	35
2.3.4	完整性约束	40
2.3.5	视图	42
第 2 章自我测试题		44
第 3 章	基于表单的数据库查询	47
3.1	QBE: 基于表单查询的根源	47
3.2	微软 Access 入门	48
3.3	微软 Access Relationships 面板	54
3.4	微软 Access 表设计视图	56
3.5	使用微软 Access 创建查询	57
3.6	体验练习 3-1: 罗列出所有的客户	60
3.7	体验练习 3-2: 有选择地显示数据列	61
3.8	体验练习 3-3: 查询结果排序	62
3.9	体验练习 3-4: 高级排序	63
3.10	体验练习 3-5: 有选择地显示数据行	65
3.11	体验练习 3-6: 复合行的选取	67
3.12	体验练习 3-7: 使用不等于运算符	68
3.13	体验练习 3-8: 表连接查询	69
3.14	体验练习 3-9: 限制连接查询结果	71
3.15	体验练习 3-10: 外连接查询	72
3.16	体验练习 3-11: 微软 Access SQL	74
3.17	体验练习 3-12: 多表连接查询与计算列	76
3.18	体验练习 3-13: 聚合函数	78
3.19	体验练习 3-14: 自连接查询	80
第 3 章自我测试题		83
第 4 章	SQL 介绍	86
4.1	SQL 简史	88
4.2	Oracle SQL 入门	89
4.3	体验练习 4-1: 解锁 HR 账户与 HR 身份登录	90
4.4	数据在哪里	93
4.4.1	使用目录视图查找数据库对象	93

4.4.2 使用对象浏览器查看数据库对象	95
4.5 体验练习 4-2: 使用 Application Express Object Browser	96
4.6 数据查询语言 (DQL): SELECT 语句	97
4.6.1 列出所有的行与列	97
4.6.2 限制显示的列	98
4.6.3 查询结果排序	99
4.6.4 选取显示的行	100
4.6.5 表连接查询	107
4.6.6 聚合函数	113
4.7 数据操纵语言 (DML)	116
4.7.1 事务支持 (COMMIT 与 ROLLBACK)	116
4.7.2 INSERT 语句	117
4.7.3 UPDATE 语句	119
4.7.4 DELETE 语句	120
4.8 数据定义语言 (DDL) 语句	120
4.8.1 CREATE TABLE 语句	121
4.8.2 ALTER TABLE 语句	122
4.8.3 CREATE VIEW 语句	123
4.8.4 CREATE INDEX 语句	124
4.8.5 DROP 语句	124
4.9 数据控制语言 (DCL) 语句	125
4.9.1 GRANT 语句	125
4.9.2 REVOKE 语句	126
第 4 章自我测试题	126

第 2 部分 数据库开发

第 5 章 数据库生命周期	131
5.1 传统生命周期	132
5.1.1 计划阶段	133
5.1.2 需求收集	134
5.1.3 概念层设计	136
5.1.4 逻辑层设计	136
5.1.5 物理层设计	137
5.1.6 构建阶段	137

5.1.7 实现与展现	138
5.1.8 后续支持	138
5.2 非传统生命周期	139
5.2.1 原型法	140
5.2.2 快速应用系统开发	140
5.3 工程三角形	141
5.4 体验练习 5-1: 工程数据库管理任务	141
第5章自我测试题	143
第6章 使用范式进行数据库设计	146
6.1 范式化的需求	148
6.1.1 插入异常	148
6.1.2 删除异常	149
6.1.3 更新异常	149
6.2 应用范式化过程	149
6.2.1 选择一个主键码	151
6.2.2 第一范式: 去除重复数据	153
6.2.3 第二范式: 去除部分依赖关系	155
6.2.4 第三范式: 去除传递依赖性	157
6.2.5 高于第三范式的范式	159
6.3 非范式化	163
6.4 实践问题	163
6.5 体验练习 6-1: UTLA 教学信息跟踪	164
6.6 体验练习 6-2: 计算机图书公司	166
用户视图	167
第6章自我测试题	169
第7章 数据与过程建模	172
7.1 实体关系建模	172
7.1.1 ERD 格式	173
7.1.2 超类与子类	179
7.1.3 绘画 ERD 图的指导方针	183
7.2 过程模型	184
7.2.1 流程图	184
7.2.2 功能层次图	186
7.2.3 泳道图	187

7.2.4 数据流图	188
7.3 关联实体与过程	190
7.4 体验练习 7-1: 使用信息工程格式绘制 ERD 图	191
第 7 章自我测试题	192
第 8 章 物理数据库设计	196
8.1 表设计	196
8.1.1 实现超类型与子类型	200
8.1.2 命名约定	203
8.2 集成业务规则与数据完整性	206
8.2.1 NOT NULL 约束	207
8.2.2 主键码约束	208
8.2.3 参照性 (外键码) 约束	208
8.2.4 唯一性约束	209
8.2.5 核对性约束	209
8.2.6 数据类型、精确度和范围	210
8.2.7 触发器	210
8.3 设计视图	211
8.4 增加索引来提高性能	212
8.5 体验练习 8-1: 逻辑模型到物理数据库模型的映射	213
第 8 章自我测试题	215

第 3 部分 数据库实现

第 9 章 数据库与外部互连	221
9.1 部署模型	221
9.1.1 中央式模型	222
9.1.2 分布式模型	223
9.1.3 客户机/服务器模型	224
9.2 数据库同 Web 互连	228
9.2.1 Internet 与 Web 介绍	229
9.2.2 Web “技术栈”中相应的组件	230
9.2.3 从 Web 页面调用事务处理	230
9.3 数据库同应用系统互连	232
9.3.1 使用 ODBC 进行数据库连接	232
9.3.2 使用 OLE DB 进行数据库连接	233