



普通高等教育“十一五”国家级规划教材
计算机科学与技术系列教材 信息技术方向

数据库技术及应用

汤荷美 周立柱 冯建华 刘卫东 宋佳兴 编著



清华大学出版社

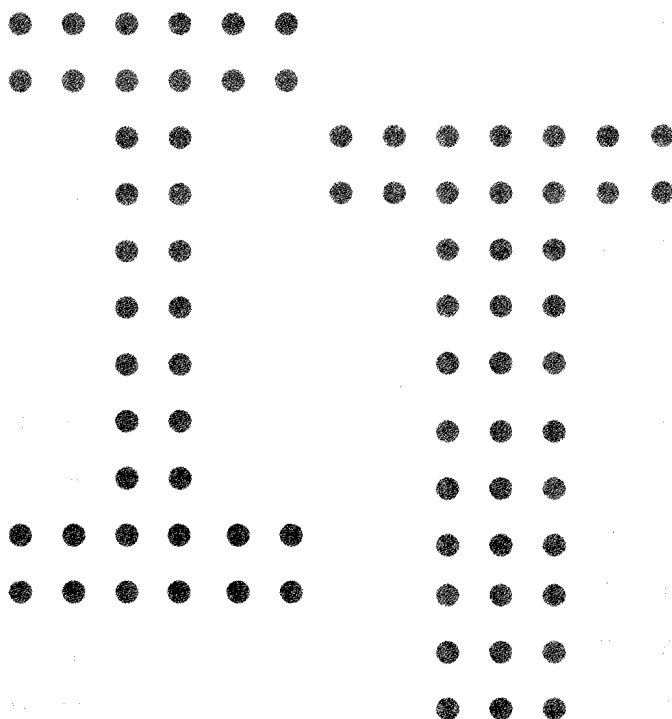
<http://www.tup.com.cn>



普通高等教育“十一五”国家级规划教材
计算机科学与技术系列教材 信息技术方向

数据库技术及应用

汤荷美 周立柱 冯建华 刘卫东 宋佳兴 编著



清华大学出版社
北京

内 容 简 介

本书紧密结合数据库技术应用需求,分为三个层次:数据库系统基本原理、数据操纵、应用开发,详细介绍数据库技术涉及的基本概念、原理、工具及方法。第一个层次数据库系统基本原理,内容包括了数据库系统的基本结构及组成、数据模型、数据存储技术、数据管理技术、分布式数据库技术等。第二个层次数据操纵,介绍 SQL 及 Oracle 的 PLSQL 关系数据库语言。第三个层次应用开发,从数据库工程角度介绍软件及数据库应用系统涉及的开发模型和开发技术,并结合 Java 和 .NET 面向对象开发环境,分别给出了一个简化的选课系统实现示例。

本教材参考 CC2005 课程体系和我国高等学校计算机科学与技术教学指导委员会编制的核心课程教学实施方案,内容覆盖了课程体系数据库中数据库技术知识点,贯穿强调基础、重视实践、内容实用的指导思想,以流行的 Oracle 大型分布式数据库系统作为实践教学对象,理论与实践结合,同时兼顾工程需求。

教材内容翔实,实用性强,可作为应用型计算机专业本科学生数据库课程的教材,也可供计算机专业工作人员及相关从业人员作为自学参考书。

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 13701121933

图书在版编目(CIP)数据

数据库技术及应用 / 汤荷美等编著. —北京:清华大学出版社, 2011.6

(计算机科学与技术系列教材 信息技术方向)

ISBN 978-7-302-25912-1

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

中国版本图书馆 CIP 数据核字(2011)第 110147 号

责任编辑:张瑞庆 薛 阳

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

地 址:北京清华大学学研大厦 A 座

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

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

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

印 装 者:北京嘉实印刷有限公司

经 销:全国新华书店

开 本:185×260

印 张:19.25

字 数:480 千字

版 次:2011 年 6 月第 1 版

印 次:2011 年 6 月第 1 次印刷

印 数:1~3000

定 价:29.50 元

产品编号:042635-01

计算机科学与技术系列教材 信息技术方向

编 委 会

主 任

陈道蓄

副 主 任

李晓明 陈 平

委 员

(按姓氏笔画为序)

马殿富	王志坚	王志英	卢先和
张 钢	张彦铎	张瑞庆	杨 波
陈 峻	周立柱	孟祥旭	徐宝文
袁晓洁	高茂庭	董 东	蒋宗礼

序 言

随着高等教育规模的扩大以及信息化在社会经济各个领域的迅速普及,计算机类专业在校学生数量已在理工科各专业中遥遥领先。但是,计算机和信息化行业是一个高度多样化的行业,计算机从业人员从事的工作性质范围甚广。为了使得计算机专业能更好地适应社会发展的需求,从2004年开始,教育部高等学校计算机科学与技术教学指导委员会组织专家对国内计算机专业教育改革进行了深入的研究与探索,提出了以“培养规格分类”为核心思想的专业发展思路,将计算机科学与技术专业分成计算机科学(CS)、软件工程(SE)、计算机工程(CE)和信息技术(IT)四个方向,并且自2008年开始进入试点阶段。

以信息化技术的广泛应用为动力,实现信息化与工业化的融合,这是我们面临的重大战略任务。这一目标的实现依赖于培养出一支新一代劳动大军。除了计算机和网络等硬件、软件的研制开发生产人员外,必须要有更大量的专业人员从事信息化系统的建设并提供信息服务。

信息技术方向作为计算机科学与技术专业中分规格培养的一个方向,其目标就是培养在各类组织机构中承担信息化建设任务的专业人员。对他们的能力、素质与知识结构的要求尽管与计算机科学、软件工程、计算机工程等方向有交叉,但其特点也很清楚。信息技术方向培养能够熟练地应用各种软、硬件系统知识构建优化的信息系统,实施有效技术管理与维护。他们应该更了解各种计算机软、硬件系统的功能和性能,更善于系统的集成与配置,更有能力管理和维护复杂信息系统的运行。在信息技术应用广泛深入拓展的今天,这样的要求已远远超出了传统意义上人们对信息中心等机构技术人员组成和能力的理解。

信息技术在国外也是近年来才发展起来的新方向。其专业建设刚刚开始起步。本系列教材是国内第一套遵照教育部高等学校计算机科学与技术教学指导委员会编制的《高等学校计算机科学与技术专业发展战略研究报告暨专业规范(试行)》(以下简称专业规范),针对信息技术方向需要组织编写的教材,编委成员主要是教育部高等学校计算机科学与技术教学指导委员会制定专业规范信息技术方向研究组的核心成员。本系列教材的着重点是信息技术方向特色课程,即与计算机专业其他方向差别明显的课程的教材建设,力图通过这些教材,全面准确地体现专业规范的要求,为当前的试点工作以及今后信息技术方向更好的发展奠定良好的基础。

参与本系列教材编写的作者均为多年从事计算机教育的专家,其中多数人直接参与了计算机专业教育改革研究与专业规范的起草,对于以分规格培养为核心的改革理念有着深刻的理解。

当然,信息技术方向是全新的方向,这套教材的实用性还需要在教学实践中检验。本系列教材编委和作者按照信息技术方向的规范在这一新方向的教材建设方面做了很好的尝试,特别是把重点放在与其他方向不同的地方,为教材的编写提出了很高的要求,也有很大的难度,但对这一新方向的建设具有重要的意义。我希望通过本系列教材的出版,使得有更多的教育界的同仁参与到信息技术方向的建设中,更好地促进计算机教育为国家社会发展服务。

 2009.6.1

中国科学院院士
教育部高等学校计算机科学与技术教学指导委员会主任

前言

数据库技术诞生于 20 世纪 60 年代,它以方便计算机用户管理大量数据为目标,研究数据在计算机内的组织、存储、加工和处理的理论和方法。数据库技术以 E. F. Codd 提出的关系模型和关系代数为基础,建立起关系数据库理论,逐步发展出关系数据库产品,成为数据管理的有力工具。从 20 世纪 80 年代中期起,数据库应用开始在我国起步,从基于 dBASE 的财务管理系统开始,到各企业的综合信息管理系统,数据库技术进入了全面、深入的应用阶段。

随着信息技术的不断发展和进步,各种信息系统逐步成为支撑各行业甚至整个社会正常运行的重要基础。数据库技术也面临新的挑战,主要体现在以下几个方面:首先,计算机网络尤其是互联网技术和应用的飞速发展,使数据库应用从单机走向网络,数据管理从集中式管理转向分布式管理;第二,传统数据库以关系模型为理论基础,而随着应用的不断延伸,各种半结构化数据、非结构化数据的管理也成为必然,对数据模式提出了更多的要求;第三,数据规模的扩大,不同类别的应用,如 OLTP、OLAP、数据仓库到数据中心等使 TB 规模的数据库成为常事。为应对这些挑战,近些年来,数据库领域中涌现出许多新的研究成果,也开发出许多新的数据库产品,数据库应用也在不断深化,延伸到越来越多的应用领域。

数据库领域研究成果的广泛应用,使得数据库技术成为高等学校计算机专业的核心课程。ACM 的计算机专业课程体系 CC2005 中,把计算机专业划分为计算机科学(CS)、计算机工程(CE)、软件工程(SE)、信息技术(IT)、信息系统(IS)5 个方向,而在其中每个方向,数据库都占有重要的地位。本教材参考 CC2005 课程体系和我国高等学校计算机教学指导委员会编制的核心课程教学实施方案,覆盖了课程体系中数据库技术的知识点,并结合数据库应用的实例,力图让读者在掌握数据库基本理论的同时,能对开发数据库应用系统有一个比较全面的认识,并在教学过程中进行基本的实践。

本书紧密结合数据库应用的现实需求,介绍数据库管理系统 DBMS 的基本概念,分析数据存储和管理的基本策略、数据索引的建立和查找算法,并介绍了事务的概念和实现机制,对当前数据库前沿技术如 XML 数据库等也有所涉及,最后,综述了数据库应用系统的基本开发方法和过程,并分别用 Java 和 .NET 开发环境给出了一个简化的选课系统的实现示例。全书以 Oracle 数据库为教

学对象,理论和实践相结合,力图为读者全面展示数据库系统的基本概念和实现技术,以及数据库管理方法和应用系统开发的基本过程。本书可作为应用型计算机专业本科学生数据库课程的教材,也可用作计算机专业工作人员在从事数据库应用时的参考书。

本书各章节主要内容如下所示。

第1章:数据库系统概述

本章简要介绍数据管理技术所经历的人工管理、文件管理和数据库管理的三个发展阶段,重点描述了数据库管理系统 DBMS 的组成和结构,说明了数据库应用系统的计算模式,最后,对本书所用的 Oracle 数据库做了总体说明。

第2章:数据模型

数据库管理系统的目标是通过数据的管理,支持对现实世界中各种应用的管理、分析和预测。现实世界的各种应用是纷繁复杂的,数据模型就是要对复杂的现实世界中的数据进行抽象,即仅保留与管理目标直接相关的数据,在计算机中建立一个能反映现实世界数据的使用情况的模型。本章给出了数据模型的概念和常用的几种模型,并对其中的实体-关系模型(ER 模型)和关系模型进行了详细介绍。

第3章:关系数据库语言 SQL 与 PLSQL

SQL 是当前使用关系数据库的标准语言,它既能对其中的数据进行操作(查询、增加、修改和删除),也能对关系数据库本身进行定义和维护(创建关系表、维护索引等),是学习和使用关系数据库的主要手段之一。本章对 SQL 的特点、各种查询功能进行了详细描述,并给出了具体实例。PLSQL 是 Oracle 对 SQL 进行的扩充,把 SQL 的数据操纵能力和过程语言的数据处理能力相结合,弥补 SQL 本身对数据结果集处理能力的不足,并可实现与其他高级语言的嵌套使用。

第4章:数据库设计

数据库设计的任务是根据应用系统的功能需求,为应用系统建立一个结构合理、使用方便、运行高效、与需求吻合的数据库。其核心是要在对应用系统运行规律准确刻画和描述的基础上,完成数据库设计的需求分析、概念设计、逻辑设计、物理设计、实现与维护等5个阶段的工作,本章具体描述了每个阶段的主要任务和一般设计方法。

第5章:数据库存储技术

数据管理的基础是数据的存储,数据存储技术直接影响数据库功能的实现和性能的提高。本章讨论了多种数据记录存储结构,以及数据文件中记录的不同组织方式,包括堆文件组织、顺序文件组织、散列文件组织、簇集(clustering)文件组织、B⁺树文件组织等。建立数据索引是提高数据库性能的必要手段,本章还讨论了不同的索引对数据访问性能的影响以及各自的特点。

第 6 章：事务管理与并发控制

数据库事务指构成单一逻辑工作单元的一系列操作，这些操作必须全部成功完成，才可以更新数据库中的数据。事务是数据库管理系统保证数据库的语义完整性，简化数据库错误恢复，提高应用系统可靠性的重要支撑技术。本章讨论了事务的 ACID 特性，以及这些特性的具体实现方法，同时讨论了支持多用户同时访问数据库的并发控制方法、死锁处理等技术。

第 7 章：数据库管理与维护

正确的设计和高效的实现是数据库应用系统正常运行的前提，而数据库管理和维护是应用系统能发挥作用的基础和保障，其任务是通过数据库管理员对数据库系统的日常维护管理，保证数据的安全性、正确性、一致性和可靠性。另外，还需要对数据库运行状态进行监控和优化，动态规划存储空间的使用，以使数据库系统能适应数据量的增长和性能的提高。本章以 Oracle 数据库为例，详细讨论了数据管理和维护的主要工作。

第 8 章：分布式数据库

计算机网络技术的发展，使数据库技术从传统的集中处理方式走向了网络时代，分布式数据库也应运而生。它以数据的物理分布性和全局数据逻辑统一性为特点，为数据的异地处理及相关的应用问题提供了很好的解决方案，广泛应用于跨地域、企业级的数据处理中。本章介绍了分布式数据库的主要特点和一般结构，并讨论了 Oracle 分布式数据库的实现及其支持的操作，说明了它的透明性。

第 9 章：XML 数据库基础

XML 是一种专门为 Internet 设计的标记语言，它为 Internet 上不同应用程序之间进行数据交换提供了标准。由于其在对半结构化数据的表示、存储、处理和传输方面的独特优势，XML 数据在互联网上得到了广泛的应用。XML 数据库主要研究对 XML 文档数据的管理技术，是当前数据库技术的发展前沿。本章介绍了 XML 数据的特点，以及 XML 数据管理的基本要求与实现。

第 10 章：数据库应用系统分析与设计

数据库应用系统的开发是一项复杂的工程，也是数据库技术应用的关键。应用系统的开发与运行应遵循软件工程的一般原则。本章介绍软件工程中系统分析设计的典型方法，分析了数据库应用系统的结构，叙述了数据库应用系统的开发过程，并详细说明了各开发阶段的主要工作。

第 11 章：Java 语言数据库编程

网络技术的飞速发展使数据库应用系统逐步进入网络环境中，Java 是当前流行的网络应用开发语言和平台。本章首先简单介绍 Java 语言和 Java 环境下数据库应用系统开发的

一般过程,然后,以一个简化的选课系统为实例,从需求分析开始,经过系统逻辑设计、物理设计等过程,最后,给出了该系统的主要功能模块的具体实现。

第 12 章: .NET 平台数据库编程

.Net 平台是微软公司推出的网络应用开发环境。本章对其做了简单的介绍,同样以第 11 章的选课系统为例,给出了 .Net 环境下各功能模块的具体实现。

本书全面介绍了数据库技术的核心内容,并结合实例给出了设计开发数据库应用系统的一般过程,内容比较丰富,教学时可根据课时的多少进行适当的取舍,如课时比较紧张,可将第 8 章、第 9 章内容作为课外阅读材料,根据具体的实验环境选择第 11 章或 12 章其中一章进行讲授。数据库系统是一门实验性较强的课程,在教学过程中安排一定的教学实验是必需的,实验可以用书中的实例作为基础,适当扩充一些如用户管理等的功能来完成。

虽然本书的作者都具有数据库方面的研究、开发经历,对数据库技术有一定的掌握和见解,但限于时间和水平有限,书中肯定还存在一些不足之处,敬请读者不吝指正。

作 者

2011 年 4 月于清华园

目 录

第 1 章 数据库系统概述	1
1.1 基本概念	1
1.1.1 数据与信息	1
1.1.2 数据处理	2
1.1.3 计算机信息系统	2
1.1.4 数据管理	2
1.2 文件系统与数据库系统	2
1.3 数据库系统的组成	4
1.3.1 数据库	5
1.3.2 数据库管理系统	5
1.3.3 应用开发工具与应用程序	6
1.3.4 数据库管理员及相关人员	6
1.4 数据库系统结构	6
1.4.1 三级模式结构	6
1.4.2 Oracle 数据库的三级模式结构	8
1.5 数据库应用系统计算模式	8
1.5.1 主机-终端计算模式	8
1.5.2 客户机/服务器计算模式(C/S)	9
1.5.3 网络计算模式(B/S)	10
1.5.4 分布式数据库计算模式	11
1.6 Oracle 数据库系统	11
1.6.1 Oracle 数据库系统的产品结构	11
1.6.2 Oracle 系统体系结构	12
1.6.3 Oracle 数据库的存储结构	14
1.6.4 Oracle 的数据字典	16
小结	16
习题	17
第 2 章 数据模型	18
2.1 模型与数据模型	18
2.1.1 概念模型	18
2.1.2 基本数据模型	19
2.2 实体-关系(ER)模型	22



2.2.1	ER 模型元素	22
2.2.2	ER 模型图形表示	23
2.3	关系模型与关系数据库	24
2.3.1	关系模型	24
2.3.2	关系的定义	27
2.3.3	关系代数与操作	28
2.3.4	关系代数与 SQL	31
2.4	关系的完整性	31
2.4.1	主码约束	31
2.4.2	外来码约束	31
2.4.3	域(Domain)约束	32
	小结	32
	习题	33
第 3 章	关系数据库语言 SQL 与 PLSQL	34
3.1	SQL 概述	34
3.1.1	SQL 的特点	34
3.1.2	SQL 的基本成分	35
3.1.3	实例	38
3.2	数据定义语言	39
3.2.1	关系表的创建与维护	39
3.2.2	视图的定义与维护	40
3.3	数据更新	41
3.3.1	INSERT 语句	41
3.3.2	UPDATE 语句	42
3.3.3	DELETE 语句	42
3.4	数据查询	43
3.4.1	SELECT 及其子句的用法	43
3.4.2	集合操作——UNION、INTERSECT、MINUS 子句	48
3.4.3	连接查询——JOIN 操作	48
3.4.4	嵌套查询(子查询)	50
3.5	过程语言	53
3.5.1	PLSQL 的特点	53
3.5.2	PLSQL 的基本结构	54
3.5.3	PLSQL 基础	55
3.5.4	存储过程	65

3.5.5 存储函数	66
3.5.6 触发器	67
3.6 Oracle 数据库操作环境简介	69
3.6.1 注册及退出 Oracle	69
3.6.2 SQL 命令的编辑及执行	70
3.6.3 Oracle 数据库的安装	71
3.6.4 Oracle 操作环境的设置	72
小结	72
习题	72
第 4 章 数据库设计	74
4.1 需求分析	74
4.1.1 需求分析任务	74
4.1.2 获取需求	74
4.1.3 分析及描述需求	75
4.1.4 需求审核与确认	76
4.1.5 功能建模方法	76
4.2 数据库概念设计	79
4.2.1 概念设计任务	79
4.2.2 概念设计方法与步骤	80
4.3 数据库逻辑设计	81
4.3.1 初始模式设计	82
4.3.2 模式优化	83
4.3.3 完整性设计	85
4.3.4 安全模式设计	85
4.3.5 外模式设计	85
4.4 数据库物理设计	86
4.4.1 确定数据的存储结构	86
4.4.2 确定分布策略	86
4.4.3 定义及维护索引	86
4.4.4 定义及维护聚集	87
4.5 实现与维护	88
4.6 关系数据库设计理论	89
4.6.1 基本概念	89
4.6.2 规范化设计方法	93
小结	97

习题	97
第5章 数据库存储技术	99
5.1 物理存储介质	99
5.1.1 三级存储体系	99
5.1.2 磁盘	99
5.1.3 RAID	100
5.1.4 第三级存储	101
5.2 文件组织	102
5.2.1 定长记录	102
5.2.2 变长记录	105
5.3 文件中记录的组织	108
5.3.1 堆文件组织	108
5.3.2 顺序文件组织	108
5.3.3 散列文件组织	109
5.3.4 簇集文件组织	109
5.4 数据字典的存储	111
5.4.1 关系的元数据	111
5.4.2 用户的元数据	111
5.4.3 统计数据和描述数据	111
5.4.4 索引的元数据	112
5.4.5 系统表	112
5.5 数据库中的索引	112
5.5.1 基本的索引结构	112
5.5.2 评价索引的标准	113
5.6 顺序索引	113
5.6.1 索引顺序文件	113
5.6.2 多级索引	115
5.6.3 索引的更新	116
5.6.4 辅助索引	117
5.7 B ⁺ 树索引文件	118
5.7.1 B ⁺ 树索引结构	118
5.7.2 B ⁺ 树索引的缺点	120
5.7.3 B ⁺ 树上的查询	120
5.7.4 B ⁺ 树的更新	120
5.7.5 B ⁺ 树文件组织	121

5.8	散列文件组织	121
5.8.1	散列文件的操作	121
5.8.2	散列函数	121
5.8.3	桶溢出控制	122
5.9	散列索引	123
5.10	顺序索引和散列的比较	124
5.11	多码访问	124
	小结	125
	习题	125
第6章	事务管理与并发控制	127
6.1	事务的概念	127
6.1.1	背景知识	127
6.1.2	事务的特性	128
6.2	事务的状态	129
6.2.1	基本术语	129
6.2.2	抽象事务模型	130
6.3	原子性和持久性的实现	130
6.4	事务的并发执行	131
6.4.1	为什么要并发执行	131
6.4.2	调度	132
6.5	封锁协议	132
6.5.1	锁	132
6.5.2	基本的封锁协议	133
6.5.3	基本封锁协议的问题	134
6.5.4	两阶段封锁协议	135
6.5.5	加强的两阶段封锁协议	136
6.5.6	商用 DBMS 中封锁协议的实现	137
6.6	时间戳排序协议	138
6.6.1	时间戳	138
6.6.2	时间戳排序协议	139
6.7	死锁处理	140
6.7.1	死锁问题	140
6.7.2	死锁预防	141
6.7.3	死锁检测与恢复	142
	小结	143

习题	143
第7章 数据库管理与维护	144
7.1 安全性	144
7.1.1 制定安全策略	144
7.1.2 用户管理	145
7.1.3 特权与角色管理	146
7.1.4 启用审计	149
7.2 完整性	150
7.3 可靠性	151
7.3.1 数据库转储(备份)	151
7.3.2 数据库恢复	152
7.4 日常管理	153
7.4.1 启动及关闭数据库	153
7.4.2 存储空间管理	155
7.5 性能优化	156
7.5.1 Oracle 的优化器	156
7.5.2 应用优化	157
7.5.3 数据库及实例的调整	158
小结	160
习题	161
第8章 分布式数据库	162
8.1 概述	162
8.1.1 实现模型	162
8.1.2 实现目标	163
8.1.3 分布式数据库参考模式结构	164
8.2 分布式 Oracle 数据库系统	165
8.2.1 概述	165
8.2.2 Oracle 分布式数据库系统结构	166
8.2.3 分布式 Oracle 系统支持的操作	167
8.2.4 分布式 Oracle 数据库的透明性	172
小结	174
习题	174

第 9 章 XML 基础	175
9.1 基本知识	175
9.1.1 Web 与 HTML	175
9.1.2 什么是 XML	175
9.1.3 XML 的优点	176
9.1.4 基本概念	177
9.2 XML 查询语言	178
9.2.1 数据模型和形式语义	179
9.2.2 XPath 1.0 与 XPath 2.0	179
9.2.3 示范用例	179
9.3 XML 数据更新	182
小结	183
习题	184
 第 10 章 数据库应用系统分析与设计	 185
10.1 软件开发模型	185
10.1.1 瀑布模型	185
10.1.2 原型模型	186
10.1.3 螺旋模型	187
10.2 数据库应用系统周期模型	188
10.2.1 数据库应用系统的组成结构	188
10.2.2 数据库应用系统的分类	189
10.2.3 数据库应用系统开发模型	190
10.3 项目规划	192
10.3.1 确定目标	192
10.3.2 可行性分析	193
10.3.3 开发计划	193
10.4 需求分析	194
10.4.1 任务和内容	194
10.4.2 需求分析步骤	194
10.5 系统设计	196
10.5.1 数据库设计	196
10.5.2 应用程序设计	197
10.5.3 系统总体设计	199
10.6 实现与部署	199
10.6.1 数据库实现	200