

国外经典教材

数据库系统原理

— Internet和Java应用指南

[美] Greg Riccardi 著 管永川 王 松 等 译

Principles of Database Systems with
Internet and Java Applications



清华大学出版社

数据库系统原理

——Internet 和 Java 应用指南

[美] Greg Riccardi 著

管永川 王 松等 译

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书围绕 Java、Internet 和虚拟的录像带出租连锁店的数据库例子, 展开了数据库建造和管理的讨论。全书从实际应用的层面揭示了核心 DBMS 理论, 探讨了信息管理、数据库系统、Java 编程以及通过 Web 与数据库进行交互等基本概念。

将数据库课程与 Internet 应用开发联系起来, 是本书的主要特色。

本书适合大学相关专业的本科生和研究生阅读, 可作为数据库相关课程的教材、参考书及研究资料, 也适合数据库应用自学者和 Internet 应用开发人员阅读和参考。

Principles of Database Systems with Internet and Java Applications

Greg Riccardi

Copyright © 2001 by Addison Wesley Publishing Company.

Original English language edition published by Addison Wesley Publishing Company

All right reserved.

No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher. For sale in the People's Republic of China Only.

本书中文简体版由 Addison Wesley Publishing Company 授权清华大学出版社出版发行, 未经出版者书面许可, 不得以任何方式复制或抄袭本书的任何部分。

北京市版权局著作权合同登记号: 图字 01-2002-4531 号

版权所有, 翻印必究。

本书封面贴有清华大学出版社激光防伪标签, 无标签者不得销售。

图书在版编目 (CIP) 数据

数据库系统原理: Internet 和 Java 应用指南/(美) 日卡尔迪著; 管永川, 王松等译.

—北京: 清华大学出版社, 2002

书名原文: Principles of Database Systems with Internet and Java Applications

ISBN 7-302-05895-4

I. 数... II. ①日... ②管... ③王... III. 数据库系统 IV. TP311.13

中国版本图书馆 CIP 数据核字 (2002) 第 072311 号

出 版 者: 清华大学出版社(北京清华大学学研大厦, 邮编 100084)

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

<http://www.tup.tsinghua.edu.cn>

责任编辑: 彭克里 李宝胜

印 刷 者: 北京昌平环球印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 787×1092 1/16 印张: 28 插页: 1 字数: 678 千字

版 次: 2002 年 11 月第 1 版 2002 年 11 月第 1 次印刷

书 号: ISBN 7-302-05895-4/TP·3500

印 数: 0001~4000

定 价: 55.00 元

前 言

本书对数据库方面的内容进行了简要介绍, 这些内容适用于以下领域: 本科生的数据库课程、面向应用的数据库/万维网交互课程以及专业开发。尤其值得一提的是, 本书对 Java 和 Internet 进行了更深层次的探讨, 从应用程序的角度来剖析核心 DBMS 理论。

本书覆盖了有关信息管理、数据库系统、Java 编程和在 WWW 网站上与数据库交互的基本内容。本书假设读者有编程背景, 它将帮助读者在设计和实现复杂信息系统方面提高他们的技巧。

信息管理是本书的中心内容, 它促进了数据模型和关系数据库系统信息表示的发展。读者既可以学习如何定义实体关系模型, 也可以学习如何在关系数据库中表示它的内容。通过这种方式, 他们可以熟练掌握 SQL——结构化查询语言, 以及关系数据库的一些高级特性, 也能学到建立高质量、大信息量应用程序的知识。另外, 他们还可了解如何让 Web 和 Java 协调工作, 帮助他们在更广的领域发布和收集信息。

本书涵盖了数据库和信息系统的內容, 包括分析信息需求、开发概念数据模型、概念性模型转换为关系模型、规范和改进关系模式、用 SQL 编写查询语句以及开发数据库应用。一些十分有意义的例程向学生展示了如何使用这些材料。另外, 本书还包括面向对象的模型和数据库、数据库性能和优化、约束和触发器、事务处理、备份与恢复、文件结构、索引和分布式对象技术等內容。

对关系数据库的讲解包括 Microsoft Access、SQL 数据库和面向对象设计工具等常见数据库工具的使用, 阐述了如何用 Microsoft Access 来描述在信息应用程序开发过程中用到的关系模型, 演示了 SQL 数据库与 ODBC 的交互以及 JDBC 与 Java 的交互。教师们可以在 Oracle 中使用这些程序项目, 或者随心所欲地将它们改写到任何关系数据库系统中。

本书讨论了如何用 Java 在面向对象的语言中表示信息的基础, 并将从 Java 如何与关系数据库交互开始介绍有关这方面编程语言。另外, 还在建立 Web 交互、表述文件结构和索引、建立分布式应用和设计面向对象的数据模型等方面介绍如何使用 Java。熟悉面向对象语言的学生可以很容易理解这些内容。而对于还不是 Java 程序员的学生, 也可以通过仔细阅读附录中关于 Java 的内容对 Java 有些了解, 同时可能需要考虑补充有关的知识。

我们讨论了应用程序和数据库的交互作用, 并用 Web 站点的实例对其进行了说明。Java 的 JDBC 类提供创建独立于数据库及平台的数据库应用方法。本书包括对这些类的详细讨论, 并提供了大量的例子。在学习了 HTML 和 CGI 编程后, 我们要求学生使用 Java 程序建立自己的 Web 站点, 并对 CGI 请求提供服务并产生 HTML 响应。另外, 我们还将讨论使用 Java 小服务程序代替 CGI 程序, 以及利用简单 Java 应用程序开发数据库用户

界面。

为了让读者理解文件结构的重要原理，我们从软件开发的角度的对它进行重点介绍，还向学生展示了在文件中表示信息的面向对象的类型。本书包括直接访问文件 I/O、buffer 打包和解包、索引和 B+树的类定义。最后，在文件结构部分结束时，我们讨论了有关实现关系数据库系统的方法。

数据库应用和实例

本书有许多独特之处，如它强调软件开发、对 Java 编程语言的使用以及将万维网作为与数据库交互工具的介绍等。介绍 Java 一般是从介绍面向对象的方法开始，接着使用这些方法建立数据库应用程序。它对数据库的支持是提供对 Internet 信息访问成功的关键。对 Java 和 Web 交互的详细讨论将阐明在建立有效应用时，哪些才是最需要的东西。总之，本书对那些想让自己的课程与 Internet 发展紧密相关的教师特别有吸引力。

复杂的信息应用程序在本书前几章中进行介绍，这些程序还将贯穿全书，用来说明各种概念和细节问题。BigHit Video Company——一个虚构的影视租售连锁店作为本书的主要实例。这些信息和应用足够用来建立从简单到复杂的应用程序。每一部分新内容都充分说明如何满足特殊信息应用程序的需求。

包括数据库、SQL 和本书中包含的例程等补充材料都可以在线获得。另外，有完整的 Microsoft Access 数据库例子可供教师们使用。我们细心地建立了这些数据库，是让学生们拓展他们的课程范围，并充实他们的课程内容。SQL 例子适合于 Access 或任何其他标准 SQL 数据库系统。我们努力保证本书中使用的代码和附加材料中用到的 SQL 代码来自通用的标准 SQL 语言的子集。

一本讲授数据库系统的书

计算机科学系的数据库系统高级课程是所有课程中非常重要的一门课程，因为它综合了用户、开发人员和软件系统维护员的需求。教师必须融合各种内容让学生懂得要满足这些需求应该做些什么。尽管学生在学习这门功课之前可能已经在计算机编程方面积累了相当多的经验，但是他们还是需要学习如何将分析、设计以及编程技巧结合起来设计出实用且有效率的软件。

计算机科学专业的数据库课程必须包括信息系统设计与实现、关系数据库系统和 SQL。由于学生们渴望成为软件专业人员，所以必须提高他们的综合技能并丰富他们的经验。最后，本书激励学生努力学习——通过学习一门非常有意思的课程可轻松实现目标。

本书是讲述数据库、Java 和万维网综合应用的一门课程，具有极大的挑战性。对学生非常有吸引力，因为它提供的是最新、最有趣的内容。

数据库课程最吸引人和最令人激动的部分应该是支持 Web 站点的数据库应用程序的

开发。在本书中，扩展了面向对象的编程内容，以展示应用程序如何运用 SQL 与数据库进行交互。本书涵盖了 HTML 基础知识、HTML 表单和 CGI 编程。它甚至包括了能替代 CGI 的 Java 小服务程序的使用。

在使用本书前，学生不必有丰富的 Java 使用经验。本书对 Java 的叙述从最基础的类开始，根据需要逐渐涉及更复杂的结构。了解任何面向对象语言的学生都能理解书中的编程实例和练习。

课程要点

本书至少讲述了数据库系统三方面的内容：传统数据库系统、面向对象的数据库应用和面向应用的数据库。每一方面的内容都包括相关的导论和一些章节来介绍数据模型、实体-关系模型及其扩展、关系模式、相关信息的处理和 SQL。我的经验是这些常用的内容约需要一个半学期的时间。这些内容将包含在以下章节中：

- 第 1 章：信息管理和数据库系统；
- 第 I 部分：第 2~5 章，信息模型和关系数据库；
- 第 II 部分：第 6~7 章，关系数据的处理。

在传统数据库系统课程中，剩余部分主要将重点放在关系数据库系统服务于信息系统的方法上。强调如何在文件结构和索引的基础上建立关系数据库系统 如何执行并优化查询，如何支持并发、事务、安全、备份与恢复。这些内容包括在以下的章节中

- 第 IV 部分：第 11~12 章，数据库的物理特性；
- 第 V 部分：第 13~15 章，在关系数据库系统中获得高性能和可靠性。

另外，在面向对象的数据库应用课程中，课程剩余部分的重点放在如何运用 Java 作为基本的编程语言，开发出具有丰富信息的应用程序。这部分的重点是在应用程序和关系数据库系统之间的交互，数据库应用程序对网站的支持，分布式信息系统 面向对象和对象-关系数据库系统。这些内容包含在以下章节中：

- 第 III 部分：第 8~10 章，数据库应用程序和万维网；
- 第 VI 部分：第 16~17 章，面向对象和分布式信息系统。

我更喜欢讲授面向应用的数据库系统，因为它将上面两门课程合并在一起。在这门课程的前半部分包括标准的内容，让学生使用 Microsoft Access 完成一个复杂的应用程序和用户界面。在后半部分中，我们讲述用 Java（第 8 章）编写关系数据库应用程序以及数据库与 Web 交互的内容（第 9 章）。在这一部分，学生使用 Java 和 SQL 数据库来建立一个网站。这门课程接下来对物理数据库（第 IV 部分）和查询过程（第 13 章）进行简要介绍，最后详细介绍了事务、安全、备份和恢复（第 14~15 章）。

使用 Java 开发数据库应用程序

在本书中，应用程序的开发主要依赖于 Java 编程语言。针对那些对 Java 不熟悉的读者

者,在本书附录中有 Java 入门的介绍。我在佛罗里达州立大学为本科生讲授了几年 Java 数据库课程。大多数学生能很好地使用 C++ 进行编程,但是却没有 Java 的编程经验。我总是告诉他们在建立数据库应用和网站时使用 Java 是非常简单的。他们开始时经常困惑,但最后总能成功。的确,他们很高兴地发现,在我们的编程项目中,不要求使用 Java 特性编程的项目是最困难的。特别是,没有要求用图形化用户界面进行编程的项目。

许多人都认为 Java 是用来建立生动的网页的一种语言。的确,它是建立运行在 Web 浏览器上的用户界面的一种有效工具。但是 Java 还是开发本书中所能见到的各种数据库应用程序的最好的编程工具。

Java 真正的优点在于它工作在 Web 服务器后台和各应用程序之间的能力。它可以在任何平台上与任何数据库服务器进行交互,这对动态网站的开发质量来说是至关重要的,且其分布式对象能力是令人叹服的。

本书中的代码例子为开发 Java 应用程序和成为一名有经验的 Java 数据库程序员指引了方向。对于大多数程序,学生可以采用一个或多个包含 Java 类来形成自己的解决方案。每一处只做很少的变化就可产生一个完全不同的应用程序。我的学生发现他们可以很容易地理解 Java 代码,因为他们已经熟悉了 C++。当他们完成了一项又一项的作业之后,他们开始逐渐了解这门语言,能力也在不断提高。在毕业之后,他们经常发现 Java 和数据库编程是他们最挣钱的技巧。

理解本书的代码实例需要仔细阅读并访问本书的网站上的完整实现过程。书中的这些代码实例经常是支离破碎的,且注释有限。而网站上的代码则比较完整,并且有更详细的文档。

补充材料

教师和学生可以在网上找到大量的补充材料。读者可以从 <ftp://ftp.aw.com/cseng/authors/riccardi/database> 目录下找到所有的 Java 代码、Microsoft Access 样本数据库和 SQL 样本数据库。尽管大多数材料都可以免费使用,但是教师手册、测试样本、习题答案和更广泛的应用只能通过与 Addison-Wesley 出版社的当地销售代表联系来得到。

所有数据库模式、SQL 语句和本书中出现的应用程序都可以在网上得到。这些材料包括 Microsoft Access 数据库、建立和组建表所需要的 SQL 代码、编制应用程序和建设网站所需要的 Java 代码,以及在 Unix 和 Windows 系统中实现应用程序与 Web 网的集成所需要的代码。

在网站上提供了支持 Web 浏览器、Web 服务器和数据库应用程序之间交互的两个版本的 Java 包。一个版本支持在 Microsoft Windows 平台使用 Access、ODBC 和微软发布的可自由使用的个人网络服务器 PWS。在这个版本中,以一种教师和学生容易在任何 Windows 平台上安装的方式实现网站实例。另一个版本支持在 Unix 平台上使用 Apache Web 服务器和 JDBC 支持的 SQL 数据库。所有的例子都使用免费的 MySQL 数据库系统和 Oracle 8 数据库系统进行了测试。Java 数据库应用程序既可以在 Unix 系统上运行,也可以在 Windows 系统上运行。

在线教师手册包括课程演讲记录、练习答案和测试例题。这些材料只提供给讲授这门课程的教师。我们会不时地在网站上添加一些最新练习和大量的项目来保持本书是最新的。它可以通过联系 Addison-Wesley 出版社在当地的销售代表而得到。

请随时通过 e-mail (Greg.Riccardi@acm.org) 来与我联系, 提出或询问书中的错误, 并得到最新的补充材料。

致 谢

我要感谢 Addison-Wesley 出版社的编辑和出版团队的支持与鼓励。组稿编辑 Maite Suarez-Rivas 坚持认为我可以写出一本很精彩的书, 并且她确实让我做到了。项目编辑 Katherine Harutunian 现正与我在两本书中进行合作, 我实在找不出比他更有帮助和更容易相处的编辑了。在本书的出版过程中, Pat Mahtani 和 Diane Freed 给予了极大的帮助。

北卡罗莱纳州州立大学的 Munindar P. Singh 教授、罗杰斯特技术学院的 Henry A. Etlinger、德克萨斯 A&M 大学的 Salih Yurttas、佐治亚州州立大学的 Arijit Sengupta 和俄克拉荷马大学的 Le Gruenwald 都对本书进行了全面和深入的审校。亚利桑那州州立大学的 Suzanne Dietrich 教授、Wayne 大学的 Bill Groshy 和 Nova 东南大学的 Junping Sun 也帮助审定了部分章节。

讲授数据库系统和与佛罗里达州州立大学的教师和学生一同工作的经验对我有很大的帮助。我还要特别感谢 Lawrence Dennis、Bryon Ehlmann、Charles W. Ford Jr.、Lois Wright Hawkes、Samuel J. Eaves II、Dmitriy Blaginin、Troy Cochran 和 Shanmaguraja Ramaswamy。佛罗里达州州立大学的 COP4710 数据库班的学生同意使用本书的初版作为他们的教材。我对所有这些学生表示感谢, 特别是 Shang-Wen Cheng、Ahmed Moussa、Gabrielle Reed、Tilak Madan、Andres Naranjo、Huey Ling Toh、Ronie Robinson、William Maher、Ron Steedman 和 Solomon Williams。

我欠我的妻子 Ann 和女儿 Mary、Christina 和 Elizabeth 太多。由于我致力于本书的写作, 她们多次容忍了我在精神上和肉体上对她们的冷漠。我把所有的成功归功于她们的支持和鼓励。

目 录

前言

第 1 章 信息管理和数据库系统	1
1.1 数据库在经济领域的重要性	2
1.2 数据库如何表述信息	3
1.3 数据库系统的相关人员	4
1.4 信息管理	5
1.5 数据库和万维网	7
1.6 数据库的概念和体系	8
1.7 信息系统的生命周期	11
1.8 数据库系统简史	12
本章总结	13
关键术语	14
练习	15
推荐阅读	16

第 I 部分 信息模型和关系数据库

第 2 章 用数据模型表示信息	19
2.1 调查并规范需求	19
2.2 组织信息	20
2.3 实体-关系模型	21
2.4 实体-关系图解	29
2.5 针对 BigHit Video 的 ER 模型	32
本章总结	38
关键术语	39
练习	40
推荐阅读	43
第 3 章 高级数据模型	44
3.1 增强型 ER 建模	44
3.2 面向对象的数据建模	48
3.3 BigHit Video 影带出租店的 OO 模型	51

本章总结	52
关键术语	53
练习	53
推荐阅读	55
第 4 章 关系数据模型	56
4.1 关系模型入门	56
4.2 关系模式	57
4.3 将 ER 模型转换成关系模式	67
4.4 用 Microsoft Access 构建数据库	70
本章总结	72
关键术语	72
练习	73
推荐阅读	74
第 5 章 改进数据库设计质量	75
5.1 属性之间的功能依赖性	76
5.2 范式	80
5.3 规范化示例	86
5.4 引用完整性和其他约束	91
本章总结	92
关键术语	92
练习	93
推荐阅读	94
 第 II 部分 处理关系数据	
第 6 章 关系数据的处理技术	97
6.1 使用关系代数来处理信息	98
6.2 使用关系演算来描述查询	110
6.3 用 QBE 和 Microsoft Access 来定义查询	111
本章总结	114
关键术语	114
练习	115
推荐阅读	119
第 7 章 SQL——结构化查询语言	120
7.1 用 SQL 的 Select 语句进行查询	121
7.2 用 SQL 修改数据库内容	133

7.3 用 SQL 创建和处理模式	136
本章总结	139
关键术语	140
练习	140
推荐阅读	142

第Ⅲ部分 数据库应用和万维网

第 8 章 面向关系型数据库的应用程序编程	145
8.1 数据库应用程序编程概论	145
8.2 JDBC 包和数据库连接	151
8.3 用 JDBC 连接数据库	153
8.4 用 DBConnect 类来连接数据库	156
8.5 执行 Select 语句	158
8.6 一个简单的 Java SQL 应用	160
8.7 执行 Insert 和 Update 语句	163
8.8 执行其他语句	164
8.9 用 Java 写一个简单的 BigHit Video 应用	165
8.10 Microsoft Access 用户界面	166
本章总结	168
关键术语	170
练习	171
推荐阅读	173
第 9 章 支持万维网上的数据库交互	174
9.1 HTML 入门	174
9.2 HTML 的活动 URL 和表单	178
9.3 用 Java 生成 HTML	183
9.4 用 Java 进行 CGI 编程	192
9.5 Java 小服务程序	194
9.6 为 BigHit Video 影带出租店设计的小服务程序	196
本章总结	197
关键术语	198
练习	199
推荐阅读	200
第 10 章 用 JDBC 改进面向对象的应用	201
10.1 JDBC 中的事务处理	201
10.2 预处理语句和过程调用语句	203

10.3 JDBC 的高级特性	204
10.4 JDBC 2.0 的结果集	204
10.5 结果集和 HTML 表单的集成	211
本章总结	214
关键术语	215
练习	215
推荐阅读	216

第IV部分 数据库的物理特征

第 11 章 以文件形式管理信息	219
11.1 文件系统的硬件特征	220
11.2 在 Java 中处理流和文件	225
11.3 在流中表示值的方法	230
11.4 通过数据流读写二进制值	232
11.5 把关系表表示成文件	238
11.6 把文件组织成记录	239
11.7 按地址直接访问记录	243
11.8 更新和删除记录	244
本章总结	245
关键术语	246
练习	247
推荐阅读	250
第 12 章 改进按值访问信息的技术	251
12.1 以内容为基础用索引来访问记录	251
12.2 二级索引	255
12.3 多级索引和 B+树	258
12.4 用散列表表示索引	266
12.5 指定物理数据库特征	270
本章总结	273
关键术语	274
练习	275
推荐阅读	277

第V部分 改进关系型数据库系统的性能和可靠性

第 13 章 查询处理和查询优化	281
13.1 处理选择查询	282

13.2 处理投影和避免重复	290
13.3 处理联合查询	292
13.4 针对复杂关系表达式的查询计划和查询优化	296
13.5 Oracle8 中的查询优化和数据库分析	300
本章总结	301
关键术语	302
练习	303
推荐阅读	306
第 14 章 事务管理	307
14.1 基本事务管理	307
14.2 并发事务处理	313
14.3 可恢复的事务时间表	321
14.4 串行化的事务时间表	322
本章总结	328
关键术语	329
练习	331
推荐阅读	333
第 15 章 数据库服务器的可靠性和安全性	334
15.1 备份和故障恢复	334
15.2 关系型数据库系统的安全问题	339
15.3 存储过程和函数	343
15.4 分布式数据库	344
本章总结	348
关键术语	349
练习	350
推荐阅读	351
 第 VI 部分 面向对象和分布式的信息系统	
第 16 章 开发面向对象的分布式应用	355
16.1 创建分布式应用	355
16.2 企业 JavaBeans(EJB)	364
16.3 Java 中的事务处理	368
本章总结	370
关键术语	371
练习	372
推荐阅读	373

第 17 章 用面向对象的数据模型来表示信息	374
17.1 用 Java 数据模型表示信息和操作	374
17.2 在 Java 中实现概念模型继承	381
17.3 对象的持久性	385
17.4 从 ER 模型转变为 Java 对象模型	390
17.5 案例分析: BigHit Video 系统中使用 Java 技术	392
17.6 对象数据库	393
17.7 对象-关系数据库	402
本章总结	403
关键术语	404
练习	405
推荐阅读	406
附录 A Java 技术概论	407
A.1 Java 和 C++ 的异同点	407
A.2 编译和执行	413
A.3 包和名	414
A.4 类和虚拟方法分级结构	416
A.5 异常和异常控制	419
A.6 应用程序编程接口 (API)	422
推荐阅读	424
参考文献	425

第 1 章 信息管理和数据库系统

学习目标

在本章中，您将学习：

- 信息管理系统的用途和优点
- 为信息系统管理提供支持的主要理论
- 数据库管理系统的基本体系结构
- 数据库系统在商业和 Web 上的应用
- 与数据存储、数据表示和应用的无关性设计有哪些好处
- 数据库系统历史中的主要事件

计算机是一种信息处理机，而信息管理是软件开发中最具挑战性的一面。大型应用程序软件开发包括确定应用程序的信息内容、定义信息模型、用特定的计算机数据结构表示那些模型，并且提供能够创建、维护和保护信息内容的软件机制。

数据库处于计算的中心。今天的数据库管理系统（DBMS）是数十年研究和开发的结果。一个商业的 DBMS 应用能建立起正确、高效、可靠和安全的系统。的确，也正是那些运用数据库的专业人员的功劳，才使得计算机系统在企业界的应用繁荣昌盛。在今天计算机经济中，数据库系统和数据库应用在不断增加，并在世界经济中占据越来越重要的地位。商业应用依赖数据库来保存重要商业信息，保证商业信息的精确性，以及让它更加易于使用。

数据库系统构成了表述人们和计算机所需信息的主要手段。因此，对数据库系统性能的彻底理解，对计算机软件开发人员进行个人的和专业的开发是至关重要的。

正是由于有了这些数据库系统和数据库专家才使得许多公司和企业如此成功。数据库方面的改进对很多机构来说都具有重大意义。因此，那些希望与企业共创未来的人们必须时刻紧跟这些改进。另外，了解如何设计和实现数据库系统和数据库应用程序的人才对这些机构有非常重要的价值，并且对这些人才的需求量非常大。

本书的重点是介绍在计算和数据库管理技术方面的信息管理。特别是指 Internet 应用程序中的信息以及支持万维网站的信息。网站使用这些信息来提供内容、搜索、电子商务和用户管理。大多数网站的软件开发都涉及到在 Web 服务器和数据库之间的交互。本书中，我们强调了 Java 编程语言的用法，这种语言支持网站将信息保存在数据库中。

1.1 数据库在经济领域的重要性

我们可以通过两个来自零售业的例子来理解数据库对商务的重要性：Wal-Mart（沃尔玛）公司——世界上最大的零售商，L. L. Bean——一个重要的邮购零售商。

在这两个公司中，数据库系统被广泛用来记录详细的销售记录。这些信息帮助管理者做出运作和战略决策。尽管两个公司在数据库的使用中有一些共同之处，但在销售目标以及从系统得到的益处方面还是有明显的差别。

像许多零售商业一样，Wal-Mart 使用电子货币注册器连接到数据库系统中，来记录在每个商店中购买的每一种商品。每次当客户购买时，都产生一条记录，包括商品名称、数量、单价、购买总额、支付方式等。通过分析这些信息可以确定某个商店或整个企业的信息。每次购买记录的详细内容允许进行如下类型的分析：

- 分类销售
 - 商品的日销售量与商品库存的比较
 - 特定商品和类似商品销售的季节性变化
 - 不同特性的类似商品的相关销售
- Market-basket 收集（一次选购所有商品）
 - 总购买量的平均值和变化
 - 所购份数和商品价格平均值和变化
 - 一次购买中所售各类商品相关性
- 客户分析
 - 多数客户的行为
 - 个别客户的喜好

对商品销售的分析可以使公司及时确定需要购进哪些商品来对客户喜欢购买的货物保持一定的库存。分析后得出结论：浅色游泳衣在盛夏比春季卖得好，以及人们在周四会比周一买更多的木炭。使用这些信息，零售店可以更好地利用库存和资金。

通过 Market-basket 分析可以发现那些经常被同时购买的货物组合。这些信息可以用来有效地组织商品摆放方式。例如，如果确定购买牛奶和购买面包有一定联系，就可以将面包放在前面，而将牛奶放在后面，这样就促使客人走到商店的后面去。他们可能看见并挑选面包，然后去寻找牛奶。

要让 Wal-Mart 所关心的所有客户的个人行为都被记录是不太可能的，因为这个公司有好几百万的客户。然而，当客户通过支票或者信用卡进行购买时，信息被记录下来并可用来确认购买者。使用同一张卡进行多次购买被认为是来自同一个人或同一个家庭。

L. L. Bean 在记录客户总数（少）和为每一位客户存储的信息总数（多）方面的信息与 Wal-Mart 不同。L. L. Bean 记录了每一位客户的姓名和地址信息，可以直接将这些信息与购买的商品联系在一起。这些搭配可以让公司确定购买商品类型。例如，购买野营设备的人们将收到适合户外用的商品价目。系统还可以让公司根据客户喜欢购买的商品来与客户进行联系。“您喜欢的那种球衣现在正在销售”可能是销售提醒设备告诉一个近

几年来购买了一些类似球衣的电话购物客户的信息。L. L. Bean 通过分析客户的个人资料为他们提供超值的服

由于受数据库容量的限制,因此 Wal-Mart 可从销售信息中提取的信息量也受到了限制。一年中销售数据达到数千吉字节。管理这些数据,保证这些信息的准确性,保护这些信息不被竞争对手访问,开发应用程序来分析它以及分发这些结果都是件非常困难的事情,并且费用昂贵。

总之,数据库系统领域的应用在显著增长。微软每月销售 100 万套 Access 数据库。Oracle 公司通过专门开发数据库系统而成为世界第二大软件公司。每天在各主要媒体或 Web 上的新站点,我们都可以看到有关数据库影响我们生活的例子。

1.2 数据库如何表述信息

数据库专家给我们描绘了数据与信息的区别。数据是经过组织的比特的集合,而信息是有特定释义或意义的数据。计算机中某个文件以特定字节为单位保存的值就是数据。与信息内容相关联的值需要它的类型(有 integer、string、float 等)、名称和名称所表示的内容。如果我们知道一个文件的第 1003~1022 字节有一个特殊的十六进制值,那么就知道了字节的数据内容。如果知道字节中有一个 20 个字符的字符串表示一次交易中的客户姓氏,我们就知道了客户的信息内容。

建立大信息量应用程序重要的第一步,是提出一个信息内容的规范。数据库系统提供了将这种规范转换为包含信息内容的数据库的方法。

数据库管理系统(DBMS)是软件和数据的结合体:

- 实物数据库:包含数据内容的文件的集合;
- 模式(Schema):实物数据库的信息内容的规范;
- 数据库引擎:支持访问和修改数据库内容的软件;
- 数据定义和处理语言:支持模式定义和数据库访问的编程语言。

关系数据库管理系统(RDBMS)是与关系模型相关联的 DBMS,所有的数据都保存在表中。表中的一行表示一个数据库对象,表中的一列表示各对象某种属性的值或特征。

例如,图 1.1 包括一个样本表,它可能是 BigHit Video 公司数据库中的一部分。这个公司是一家假想的影带出租连锁店,是本书的主要数据库例子。它还展示了对表结构或者模式的逻辑定义以及在 DBMS 种用来创建表的建表语句。表模式包括一个名称(custom)和一系列名称。建表语句还可以做更多的事情,指定每列的类型信息。表本身包括列和一组行的定义。表中每一行表示一个单一客户的信息,并且对应于每一列都有一个值。

数据库系统的模式存储在系统中。也就是说,数据库系统是自描述的。当图 1.1 的建表语句在数据库中执行时,表模式就存储在了数据库中。数据库的内容可以根据这个模式信息进行存取,而不是根据任何实物特征来访问。