

数据库系统 设计、实现与管理

(第8版)

Peter Rob 著
Carlos Coronel

金名 张梅 等译

DATABASE SYSTEMS: DESIGN, IMPLEMENTATION
AND MANAGEMENT Eighth Edition

清华大学出版社



世界著名计算机教材精选

数据库系统设计、实现与管理

(第8版)

Peter Rob 著
Carlos Coronel

金名张梅等译

清华大学出版社
北京

北京市版权局著作权合同登记号 图字 01-2009-6827 号

Database Systems: Design, Implementation, and Management, 8E
Peter Rob, Carlos Coronel

Copyright © 2009 by Course Technology, of Cengage Learning.

Original edition published by Cengage Learning. All Rights reserved. 本书原版由圣智学习出版公司出版。版权所有,盗印必究。

Tsinghua University Press is authorized by Cengage Learning to publish and distribute exclusively this simplified Chinese edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

本书中文简体字翻译版由圣智学习出版公司授权清华大学出版社独家出版发行。此版本仅限在中华人民共和国境内(不包括中国香港、澳门特别行政区及中国台湾)销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可,不得以任何方式复制或发行本书的任何部分。

Cengage Learning Asia Pte. Ltd.
5 Shenton Way, # 01-01 UIC Building, Singapore 068808

本书封面贴有 Cengage Learning 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

数据库系统设计、实现与管理:第8版/(美)罗布(Rob, P.), (美)科尼尔(Coronel, C.)著;金名,张梅等译. —北京:清华大学出版社, 2012.7

书名原文: Database Systems: Design, Implementation, and Management

世界著名计算机教材精选

ISBN 978-7-302-29012-4

I. ①数… II. ①罗… ②科… ③金… ④张… III. ①数据库系统—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2012)第 124030 号

责任编辑:龙啟铭
封面设计:傅瑞学
责任校对:胡伟民
责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>
地 址:北京清华大学学研大厦 A 座 邮 编:100084
社 总 机:010-62770175 邮 购:010-62786544
投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn
质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:清华大学印刷厂

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:35.75

字 数:900千字

版 次:2012年7月第1版

印 次:2012年7月第1次印刷

印 数:1~3000

定 价:69.00元

产品编号:033747-01

译者序

正如本书的书名所指的那样，本书涵盖了数据库系统的设计、实现和管理 3 个方面的内容，全书分为数据库概念、数据库设计概念、高级数据库设计与实现、高级数据库概念、数据库与 Internet、数据库与管理 6 大部分共 15 章，首先从系统的角度介绍数据库及其设计的概念，接着全面阐述数据库设计，然后用示例讲解了数据库的具体实现，包括 Web 数据库的开发等，最后介绍了数据库系统的管理，如事务管理与并发控制、分布式数据库管理、系统数据库管理与安全、数据库性能调整与查询优化等内容。

本书是一本非常有特点的教材，其主要特点如下：

(1) 层次结构清晰。本书先讲解数据库的概念，接着介绍数据库的设计，然后介绍数据库的具体实现，使读者从抽象到具体，对数据库系统有着深刻的理解和掌握。

(2) 图文并茂。为了便于读者对数据库基本概念的理解，书中给出了大量形象的图形描述，此外，对于一些示例的操作过程和结果，都给出了相应的图形界面，便于读者理解和掌握。

(3) 示例丰富。为了配合 SQL 语言的讲解，书中给出了大量的示例，让读者了解每个语句的作用。最为突出的是，本书以一所大学的数据库管理系统为例，详尽地介绍了数据库的设计与实现的步骤。

总之，这是一本既系统又实用的数据库教材，不仅是高等学校本科生学习数据库的很好教材，对数据库设计、开发和管理人员，也非常有参考价值。

本书由金名和张梅主译，张长富、蔡建章、李匀、张建安、邓铁洪、徐君、李强、蒋恩俊、杨文保、李强、苏辛、周成兴、魏敬安、朱建波、徐志平、赵杰辉、傅祎、郭碧莲、郭洵、洪晓煜、黄宣达、江松波、柯渝、赖曲芳、廖阳、刘文红、李伟、郭涛、高磊、王振营、冯哲、韩毅、马以辉、李腾、邓卫、邓凡平、周云、董武、郑晓蕊、陈占军、倪泳智、黄虹、吕巧珍、裘蕾、金颖、陈河南、王嘉佳、吴建伟、宋雁、贺军等人也参与了本书的翻译工作，在此一并表示感谢！

由于水平有限，如有不妥之处，恳请读者指正。

目 录

第 1 部分 数据库概念

第 1 章 数据库系统 3	第 2 章 数据模型 25
1.1 数据与信息..... 3	2.1 数据建模及数据模型..... 25
1.2 数据库和数据库管理系统简介..... 5	2.2 数据模型的重要性..... 26
1.2.1 DBMS 的作用和优势..... 5	2.3 数据模型基本组成..... 27
1.2.2 数据库类型..... 7	2.4 业务规则..... 28
1.3 数据库设计为什么很重要..... 8	2.4.1 发现业务规则..... 29
1.4 历史根源：文件和文件系统..... 9	2.4.2 将业务规则转化成数据模型组件..... 29
1.5 文件系统数据管理存在的问题..... 12	2.5 数据模型的发展..... 30
1.5.1 结构依赖和数据依赖..... 13	2.5.1 层次模型..... 30
1.5.2 字段定义和命名规范..... 13	2.5.2 网状模型..... 31
1.5.3 数据冗余..... 15	2.5.3 关系模型..... 32
1.6 数据库系统..... 16	2.5.4 实体联系模型..... 34
1.6.1 数据库系统环境..... 17	2.5.5 面向对象模型..... 35
1.6.2 DBMS 功能..... 19	2.5.6 数据模型合并..... 37
1.6.3 管理数据库系统：焦点转移..... 21	2.5.7 数据库模型和互联网..... 38
本章小结..... 22	2.5.8 数据模型：总结..... 38
思考题..... 22	2.6 数据抽象程度..... 41
习题..... 23	2.6.1 外模型..... 41
	2.6.2 概念模型..... 43
	2.6.3 内模型..... 43
	2.6.4 物理模型..... 44
	本章总结..... 45
	思考题..... 45
	习题..... 46

第 2 部分 数据库设计概念

第 3 章 关系数据库模型 53	3.1.1 表及其特性..... 54
3.1 数据的逻辑视图..... 53	3.2 码..... 56

3.3 完整性规则	60	5.3.3 转化为第三范式	125
3.4 关系集合操作	62	5.4 改进数据库设计	126
3.5 数据字典和系统目录	67	5.5 代理码的考虑	130
3.6 关系数据库中的联系	68	5.6 高阶范式	131
3.6.1 1:M 联系	68	5.6.1 鲍依斯-科得范式	
3.6.2 1:1 联系	69	(BCNF)	131
3.6.3 M:N 联系	71	5.6.2 第四范式(4NF)	133
3.7 再论数据冗余	74	5.7 规范化和数据库设计	135
3.8 索引	76	5.8 反规范化	138
3.9 Codd 的关系数据库规则	77	本章总结	140
本章总结	78	思考题	142
思考题	78	习题	143
习题	79	第6章 高级数据建模	150
第4章 实体联系(ER)模型	87	6.1 扩展实体关系模型	150
4.1 实体联系模型(ERM)	87	6.1.1 实体超类和实体	
4.1.1 实体	88	子类	150
4.1.2 属性	88	6.1.2 特殊化层次结构	151
4.1.3 联系	92	6.1.3 继承	152
4.1.4 互连和基数	93	6.1.4 子类辨识器	153
4.1.5 存在依赖性	93	6.1.5 不相交/重叠约束	153
4.1.6 联系强度	94	6.1.6 完备性约束	154
4.1.7 弱实体	96	6.1.7 特殊化与一般化	155
4.1.8 联系的参与	97	6.2 实体簇	155
4.1.9 联系的度	99	6.3 实体完整性与主码选择	156
4.1.10 递归联系	100	6.3.1 自然码与主码	157
4.1.11 联合(复合)		6.3.2 主码选择原则	157
实体	102	6.6.3 什么时候使用	
4.2 开发 ER 图	104	复合主码	158
4.3 数据库设计挑战:		6.3.4 什么时候使用	
冲突目标	110	代理主码	159
本章总结	112	6.4 设计实例:学习灵活的	
思考题	112	数据库设计	160
习题	114	6.4.1 设计示例#1:实现	
第5章 数据库表的规范化	117	1:1 型联系	160
5.1 数据库表和规范化	117	6.4.2 设计示例#2:维护	
5.2 规范化的需求	118	时变数据的历史	161
5.3 规范化处理	120	6.4.3 设计示例#3:扇形	
5.3.1 转化为第一范式	121	陷阱	163
5.3.2 转化为第二范式	124		

6.4.4 设计示例#4: 冗余	本章总结	166
联系	思考题	167
6.5 数据建模一览表	习题	167

第 3 部分 高级数据库设计与实现

第 7 章 结构化查询语言 (SQL)

简介	7.5.3 增加字段	210
7.1 SQL 概述	7.5.4 删除字段	210
7.2 数据定义命令	7.5.5 高级数据更新	211
7.2.1 数据库模型	7.5.6 复制部分表	212
7.2.2 创建数据库	7.5.7 增加主码和外码	214
7.2.3 数据库模式	7.5.8 删除数据库中的表	215
7.2.4 数据类型	7.6 高级 SELECT 查询	215
7.2.5 创建表结构	7.6.1 查询结果排序	215
7.2.6 SQL 约束	7.6.2 显示唯一值	217
7.2.7 SQL 索引	7.6.3 聚集函数	217
7.3 数据操纵命令	7.6.4 数据分组	221
7.3.1 添加表记录	7.7 虚拟表: 创建视图	223
7.3.2 保存表的修改	7.8 数据库表连接	224
7.3.3 显示表记录	7.8.1 带别名的表连接	227
7.3.4 更新表记录	7.8.2 递归连接	227
7.3.5 恢复表的内容	7.8.3 外连接	228
7.3.6 删除表记录	本章总结	229
7.3.7 用 SELECT 子查询	思考题	230
插入表记录	习题	233
7.4 SELECT 查询	第 8 章 高级 SQL	241
7.4.1 条件约束查询	8.1 关系集合运算符	241
7.4.2 算术运算符: 优先级	8.1.1 UNION	242
规则	8.1.2 UNION ALL	243
7.4.3 逻辑运算符: AND、	8.1.3 INTERSECT	244
OR 和 NOT	8.1.4 MINUS	245
7.4.4 特殊运算符	8.1.5 语法替换	246
7.5 高级数据定义命令	8.2 SQL 连接运算符	248
7.5.1 修改字段的数据	8.2.1 交叉连接	249
类型	8.2.2 自然连接	250
7.5.2 修改字段的数据	8.2.3 用子句连接	251
特征	8.2.4 在子句中连接	251
	8.2.5 外连接	252

8.3	子查询和关联查询	255	思考题	301
8.3.1	WHERE 子查询	256	习题	302
8.3.2	IN 子查询	257	第9章 数据库设计	306
8.3.3	HAVING 子查询	258	9.1 信息系统	306
8.3.4	多行子查询运算符: ANY 和 ALL	258	9.2 信息系统开发生命周期	308
8.3.5	FROM 子查询	260	9.2.1 计划	308
8.3.6	属性列表子查询	261	9.2.2 分析	309
8.3.7	关联查询	263	9.2.3 详细的系统设计	310
8.4	SQL 函数	265	9.2.4 实现	310
8.4.1	日期和时间函数	265	9.2.5 维护	311
8.4.2	数字函数	268	9.3 数据库生命周期	311
8.4.3	字符串函数	269	9.3.1 数据库初步研究	311
8.4.4	转换函数	270	9.3.2 数据库设计	315
8.5	Oracle 子查询	272	9.3.3 实现和装载	328
8.6	可更新的视图	275	9.3.4 测试和评价	331
8.7	过程 SQL	277	9.3.5 运行	332
8.7.1	触发器	281	9.3.6 维护和演化	332
8.7.2	存储过程	289	9.4 数据库设计策略	333
8.7.3	带游标的 PL/SQL 处理	293	9.5 集中式和分散式设计的 对比	335
8.7.4	PL/SQL 存储函数	295	本章总结	336
8.8	嵌入式 SQL	295	思考题	336
	本章总结	299	习题	336

第4部分 高级数据库概念

第10章 事务管理与并发控制	341	10.2.4 调度器	351
10.1 什么是事务	341	10.3 用锁方法进行并发控制	352
10.1.1 演化事务结果	343	10.3.1 锁粒度	353
10.1.2 事务的性质	345	10.3.2 锁类型	355
10.1.3 用 SQL 的事务 管理	346	10.3.3 两阶段加锁	357
10.1.4 事务日志	346	10.3.4 死锁	357
10.2 并发控制	348	10.4 用时间戳方法进行并发 控制	359
10.2.1 更新丢失	348	10.4.1 WAIT/DIE 和 WOUND/WAIT 调度方案	359
10.2.2 未提交数据	349		
10.2.3 不一致检索	350		

10.5 用优化方法进行并发控制.....	360	12.5 DDBMS 的组成.....	403
10.6 数据库恢复管理.....	360	12.6 数据层和分布式处理.....	404
10.6.1 事务恢复.....	361	12.6.1 单点处理与单点 数据.....	404
本章总结.....	364	12.6.2 多点处理与单点 数据.....	405
思考题.....	365	12.6.3 多点处理与多点 数据.....	406
习题.....	365	12.7 分布式数据库的透明性.....	407
第 11 章 数据库性能调整与 查询优化器.....	368	12.8 分布透明性.....	408
11.1 数据库性能调整的概念.....	368	12.9 事务处理透明性.....	410
11.1.1 性能调整: 客户和 服务器.....	369	12.9.1 分布式请求和 分布式事务.....	410
11.1.2 DBMS 架构.....	370	12.9.2 分布式并发控制.....	413
11.1.3 数据库统计.....	371	12.9.3 两阶段提交协议.....	414
11.2 查询处理.....	373	12.10 性能透明性和查询优化.....	415
11.2.1 SQL 分析阶段.....	373	12.11 分布式数据库设计.....	416
11.2.2 SQL 执行阶段.....	375	12.11.1 数据分割.....	416
11.2.3 SQL 提取阶段.....	375	12.11.2 数据复制.....	419
11.2.4 查询处理瓶颈.....	375	12.11.3 数据放置.....	421
11.3 索引和查询优化.....	376	12.12 客户/服务器与 DDBMS 的比较.....	421
11.4 优化选择.....	378	12.13 C. J. Date 关于分布式数 据库的 12 条告诫.....	422
11.4.1 使用提示影响优化 选择.....	379	本章总结.....	423
11.5 SQL 性能调整.....	380	思考题.....	424
11.5.1 索引选择性.....	380	习题.....	424
11.5.2 条件表达式.....	381	第 13 章 业务智能和数据仓库.....	427
11.6 查询制定.....	383	13.1 数据分析需求.....	427
11.7 DBMS 性能调整.....	384	13.2 业务智能.....	428
11.8 查询优化示例.....	386	13.3 业务智能架构.....	429
本章总结.....	391	13.4 决策支持数据.....	433
思考题.....	392	13.4.1 运营数据对决策 支持数据.....	433
习题.....	393	13.4.2 决策支持数据库 需求.....	435
第 12 章 分布式数据库管理系统.....	397	13.5 数据仓库.....	437
12.1 分布式数据库管理系统 的发展.....	397	13.5.1 12 条定义数据 仓库的规则.....	440
12.2 DDBMS 的优点和缺点.....	399		
12.3 分布式处理和分布式 数据库.....	400		
12.4 分布式数据库管理系统 的特征.....	402		

13.5.2	决策支持架构 风格	441	13.7.6	用于星形模式的 性能提高技术	458
13.6	联机分析处理	442	13.8	实现数据仓库	461
13.6.1	多维数据分析 技术	442	13.8.1	数据仓库作为 一个积极的决策 支持框架	461
13.6.2	高级数据库支持	443	13.8.2	需要用户参与的 全公司范围的 努力	461
13.6.3	易于使用的终端 用户界面	444	13.8.3	满意三部曲: 数据、分析和 用户	461
13.6.4	客户端/服务器端 架构	444	13.8.4	应用数据库 设计过程	462
13.6.5	OLAP 架构	445	13.9	数据挖掘	463
13.6.6	关系联机分析 处理	448	13.10	OLAP 的 SQL 扩展	465
13.6.7	多维联机分析 处理	450	13.10.1	ROLLUP 扩展	466
13.6.8	ROLAP 对 MOLAP	451	13.10.2	CUBE 扩展	467
13.7	星形模式	452	13.10.3	物化视图	468
13.7.1	事实	452	本章小结		471
13.7.2	维度	453	思考题		472
13.7.3	属性	453	习题		473
13.7.4	属性层次结构	455			
13.7.5	星形模式表示	456			

第 5 部分 数据库与 Internet

第 14 章	数据库互连和 Web 技术	479	14.2.2	Web 服务器接口	491
14.1	数据库连接	479	14.2.3	Web 浏览器	492
14.1.1	本机 SQL 连接	480	14.2.4	客户端扩展	494
14.1.2	ODBC、DAO 和 RDO	481	14.2.5	Web 应用服务器	495
14.1.3	OLE-DB	483	14.3	可扩展标记语言 (XML) ...	495
14.1.4	ADO.NET	485	14.3.1	文档类型定义 (DTD) 与 XML 模式	497
14.1.5	Java 数据库连接 (JDBC)	488	14.3.2	XML 表示	500
14.2	互联网数据库	489	14.3.3	XML 应用程序	501
14.2.1	Web 到数据库 中间件: 服务器 端的扩展	490	本章小结		503
			思考题		504
			习题		504

第 6 部分 数据库与管理

第 15 章 数据库管理和安全	509	15.9 DBA 工作: Oracle 数据库管理	536
15.1 数据是公司资产	509	15.9.1 Oracle 数据库管理工具	537
15.2 机构中数据库的角色需求	510	15.9.2 默认登录	537
15.3 数据库引入: 特定考虑	511	15.9.3 保证 RDBMS 自动开始	538
15.4 数据库管理职能开发	512	15.9.4 创建表空间和数据文件	539
15.5 数据库环境中人的构成	515	15.9.5 管理数据库对象: 表、视图、触发器和存储过程	541
15.5.1 DBA 的管理角色	517	15.9.6 管理用户和构建安全性	542
15.5.2 DBA 的技术角色	522	15.9.7 自定义数据库初始化参数	543
15.6 安全性	527	15.9.8 创建新数据库	544
15.6.1 安全策略	528	本章小结	549
15.6.2 安全漏洞	528	思考题	550
15.6.3 数据库安全	529	词汇表	553
15.7 数据库管理工具	530		
15.7.1 数据字典	531		
15.7.2 CASE 工具	533		
15.8 开发数据管理策略	535		

第 1 部分

数据库概念

第 1 章 数据库系统

第 2 章 数据模型

相关变革

今天，我们理所当然地享受着关系型数据库给我们带来的好处：能够快速且方便地在廉价的计算机上存取和改变数据。但是，直到 20 世纪 70 年代末，用于存储大量数据的层次型数据库还是比较难于操纵和更改的。在设计层次型数据库之前，程序员需要知道客户希望利用这些数据做什么。对层次数据进行增加或改变是一个非常耗时且代价高昂的过程。由此导致的结果是，需要在一大堆卡片目录中查找一本图书、使用未显示去年变化的地图、不得不买来报纸查找股票信息等。

然而，1970 年，IBM 数学家 E. F. Codd 的一篇论文改变了这一切。当时并没有人意识到 Codd 的晦涩难懂的理论将会引起一场类似于个人计算机和因特网发展的技术革命。SQL（当今数据库系统使用最流行的计算机语言）发明者之一 Don Chamberlin 说：“Codd 使用一些奇怪的数学符号，但当时没人把它当回事。”后来，Codd 组织了一次学术报告会。当 Chamberlin 听到 Codd 将长达 5 页的复杂程序减少为 1 行时发出惊叹：“这太神奇了！”。

这次学术报告会促使了 IBM 筹资支持 System R 项目的研制。该项目旨在开发一个关系型数据库原型系统，并最终促使了 SQL 和 DB2 的诞生。然而，IBM 在极为关键的一段时间内搁置了 System R，因其当时在 IMS（1968 年推出的一个可靠的高端数据库系统）上获得了大量收益。由于 IBM 没有意识到 System R 研究的市场潜力，因此允许其员工公开发表相关论文。

当时，这些论文的读者之一 Larry Ellison 刚刚成立了一家小公司。他招募了一些来自于 System R 项目和 California 大学的程序员，于 1979 年向市场推出了第一个基于 SQL 的关系型数据库，并完全领先于 IBM。到了 1983 年，这家公司发布了该数据库的一个可移植版本，年收入已超 500 万美元，并将其更名为 Oracle。在激烈竞争的刺激下，IBM 终于在 1980 年发布了它的第一个关系型数据库 SQL/DS。

2007 年，全球数据库管理系统的总销量达到 150 亿美元，其中 Oracle 占据大约一半的市场份额，IBM 尾随其后，约占四分之一。Microsoft 的 SQL Server 市场份额增长快于竞争对手，已攀升至 14%。

第 1 章 数据库系统

本章介绍以下内容：

- 数据与信息区别。
- 什么是数据库？有哪些类型的数据库？为什么数据库能为决策提供支持？
- 数据库设计的重要性。
- 现代数据库怎样从文件系统演变而来？
- 文件系统管理数据存在的缺陷。
- 数据库系统由哪些部分组成？数据库系统与文件系统的区别是什么？
- 数据库管理系统（DBMS）的主要功能。

正确决策的做出，需要从未经加工的事实中得到有用的信息。我们将这些未经加工的事实称为数据。当数据存储于数据库中时，管理效率可能最高。通过本章学习，我们可以了解到什么是数据库？它的作用是什么？与其他数据管理方法相比，为什么数据库能得到更好的效果？我们也可了解到各种类型的数据库以及数据库设计的重要性。

数据库是从文件系统演变而来的。虽然利用文件系统进行数据管理已经过时，但是理解文件系统的特性却十分重要，因为它是严重制约数据管理效率的源头。在本章中，我们还可以了解到数据库系统如何克服文件系统数据管理的缺陷。

1.1 数据与信息

为弄清数据库设计的驱动原因，必须理解数据和信息的差别。数据（data）是未经加工的事实。所谓“未经加工”，表明它未被处理，其包含的内在涵义也未被揭示。例如，假设我们想了解一个计算机实验室的用户对实验室服务有什么样的看法时，通常会调查用户访问该实验室的情况。图 1.1（a）是给实验室用户设计的 Web 调查问卷表。当用户填写完该表后，表中未经加工的数据被保存到图 1.1（b）所示的数据库中。虽然我们已获得这些事实，但是这种格式的数据并没有太大用处，因为阅读一页又一页的 0 和 1 并不能提供多少有意义的信息。然而，我们若将这些未经加工的数据转化成如图 1.1（c）所示的数据汇总时，就能够快速回答诸如“实验室用户由哪些人员组成”之类的问题。在图 1.1（c）中，可以迅速得出该实验室的大部分用户是三年级学生（24.59%）和四年级学生（53.01%）。由于图形能够增强人们快速从数据中提取有意义信息的能力，图 1.1（d）给出了该实验室调查问卷数据汇总的柱状图。

与数据相反，信息（Information）是对原始数据进行加工处理后的结果，并揭示了其包含的内在含义。数据处理既可以是简单地从有组织的数据中发现模式，也可以是使用复

Business Computer Lab Satisfaction Survey

College of Business
COMPUTER LAB

This survey is designed to obtain student feedback regarding the services provided by the Business Computer Lab and identify areas in which we need to improve. Please answer each question as accurately as possible.

Using the Lab

What is your academic classification?
 Freshman Sophomore Junior Senior Graduate Student Other

Do you own a computer?
 Yes No

How often do you use the Business Computer Lab?
 Five or more times / week
 Three or four times / week
 Once or twice / week
 Once / month or less

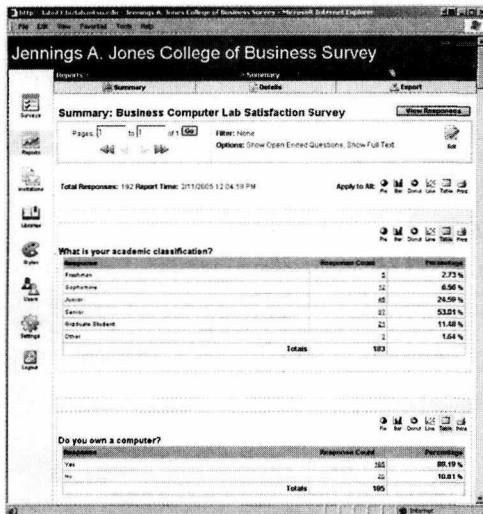
What do you **primarily** use the Business Computer Lab for?
 You may check **more than one**
 Internet (e.g. Web Browsing, Chat) Presentations (e.g. MS PowerPoint)
 Email Access Academic Programs (e.g., Moodle, Canvas, MS Visio, etc.)
 Word Processing (e.g. MS Word) Games
 Spreadsheets (e.g. MS Excel) Other: _____

What do you like **MOST** about the Business Computer Lab?
 You may check **more than one**
 Email Up to date software
 24 Hour access Laptop printers
 Availability of computer Scanners

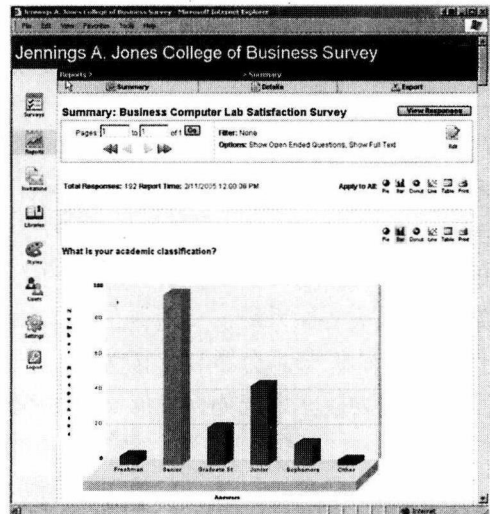
(a) 初始屏幕

Acad. Clas.	Computer	How Often	Primary	Most	Frequency	Frequency	Frequency
2	Gr	N	5W	0	1	1	1
3	Sen	Y	5W	1	0	0	1
4	Sen	Y	1W	0	0	0	0
5	Sen	Y	1W	1	0	0	0
6	Sen	Y	3W	0	0	0	0
7	Gr	N	5W	0	0	0	1
8	Sen	Y	1W	1	0	0	1
9	Sen	Y	3W	1	0	1	0
10	Sen	Y	1W	0	0	0	1
11	Sen	Y	5W	0	1	0	0
12	Jun	Y	1W	1	0	0	1
13	Sen	N	5W	1	0	0	1
14	Jun	Y	1W	0	0	0	0
15	Sen	Y	5W	0	0	0	0
16	Jun	Y	1W	0	1	0	0
17	Sen	Y	1W	0	0	1	0
18	Sen	Y	1W	0	0	1	0
19	Gr	N	5W	1	0	0	1
20	Gr	Y	1M	0	1	0	0
21	Gr	Y	5W	0	0	1	1
22	Sen	N	3W	0	0	1	1
23	Jun	Y	1W	1	0	0	0
24	Sen	Y	3W	1	1	0	1
25	Jun	Y	1W	1	0	0	1
26	Jun	Y	1W	1	0	0	0
27	Sen	Y	1M	1	0	0	1
28	Sen	Y	5W	0	0	0	1
29	Gr	Y	1M	0	0	1	0
30	Gr	Y	5W	1	1	0	1
31	Jun	N	5W	1	0	1	1
32	Jun	Y	3W	1	0	0	0
33	Gr	Y	5W	0	1	0	1

(b) 原始数据



(c) 汇总信息



(d) 图形格式的信息

图 1.1 将未经加工的数据转化为信息

杂的统计模型进行预测或推断。为揭示数据所包涵的具体意义，往往需要知道信息的背景信息。例如，平均温度是 105 度并没有太多含义，除非你知道它的背景知识，比如，是华氏温度还是摄氏温度？是机器温度、体温还是室外气温？信息可作为制订决策的基础。例如，从调查问卷每个问题的数据汇总中，可以知道该实验室的优点和缺点，从而有助于做出合理的决定，以更好地满足用户需求。

需要记住的是，未经加工的数据必须以适当地格式进行存储、处理和表示。例如，在图 1.1 (c) 中，学生分类是基于新生、二年级、三年级、四年级和研究生进行组织和显示的。被调查人回答的 Yes/No 需要转化成 Y/N 的格式进行存储。在处理诸如声音、视频和图像等复杂数据类型时，则需采用更加复杂的格式。

在当今信息时代，产生准确、相关以及及时的信息是做出正确决策的关键，而做出正确决策又是企业在全球市场生存的关键。我们正进入所谓的“知识时代”，而数据是信息和知识的基础。换言之，数据是信息的躯体和关于特定主题的事实。知识 (Knowledge) 意味着当信息应用到具体环境时对信息的认知、通晓和理解。知识的一个关键特征是“新”

知识可从“旧”知识推导而来。

对上述内容小结如下：

- 数据是信息的基本组成。
- 信息是对数据进行加工处理而产生的。
- 信息是用于揭示数据内在的含义。
- 准确、相关以及及时的信息是做出正确决策的关键。
- 做出正确决策是企业在全球环境下生存的关键。

如果希望获取及时和有用的信息，则必须获得准确的数据。准确的数据必须以正确的方式产生，并以易于存取和处理的格式进行存储。像任何基本资源一样，必须小心管理数据所处的环境。数据管理（data management）是一门研究数据是如何正确产生、存储和检索的学科。对于任何商业单位、政府部门、服务组织或慈善机构，如果能理解数据在其日常活动中所扮演的重要角色，则对以数据管理作为它们的核心活动就不会感到惊奇。

1.2 数据库和数据库管理系统简介

高效的数据管理通常需要数据库支持。数据库（database）是一个共享和集成的计算机结构，用于存储下列数据的集合：

- 用户数据，即终端用户感兴趣的未经处理的事实。
- 元数据（Metadata），即描述数据的数据，用于集成和管理终端用户数据。

元数据提供了对数据特征的描述，以及数据库中数据之间的联系集。例如，元数据可存储数据元素的名称、元素值的数据类型（如数值、日期和文本等）、能否为空值等信息。因此，元数据提供了关于数据值及如何使用这些数据的补充和扩展信息。简言之，元数据呈现了数据库中数据的完整描述。由于元数据具有这些特征，因此数据库被称为“自描述数据的集合”。

一个数据库管理系统（Database Management System, DBMS）是一组程序的集合，它负责管理数据库的结构，并控制对数据库中数据的访问。在某种意义上，数据库可以看做一个组织得很好的电子档案柜，而其中的内容则由功能强大的软件（即数据库管理系统）帮助管理。

1.2.1 DBMS 的作用和优势

DBMS 是用户和数据库之间的中介。数据库结构本身是作为一个文件的集合进行存储，而访问这些文件中数据的唯一途径是 DBMS。图 1.2 表示了 DBMS 呈现给终端用户（或应用程序）的单一和集成的数据视图。DBMS 负责接受所有应用程序的请求，并将它们翻译成用于执行这些请求的复杂数据库操作，从而对应用程序和用户隐藏了数据库的内部复杂性。应用程序既可以是使用某种程序语言（如 Visual Basic.NET、Java 或 C++）来编写，也可能是通过具体 DBMS 的实用程序来创建。

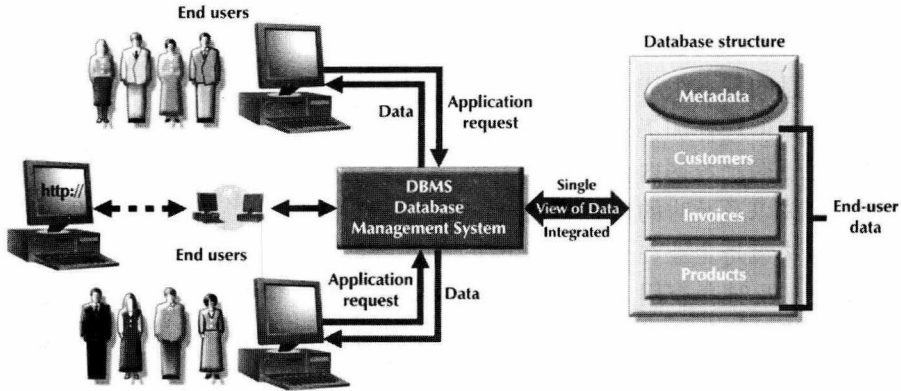


图 1.2 DBMS 管理终端用户与数据库之间的交互

在应用程序和数据库之间提供 DBMS 具有许多优点。首先，DBMS 能使数据库中的数据被多个应用程序或用户共享。其次，DBMS 可将许多不同用户的数据视图集成为一个单一且包罗万象的数据仓储。

由于数据是用于推导信息的关键原始材料，因此，应设计出好的方法对它们进行管理。通过本书的学习，将会发现 DBMS 能够帮助人们高效地管理数据。特别地，DBMS 可带来以下好处：

- 提高数据共享度。DBMS 能帮助用户创建一个便于存取大量且有组织数据的环境，使得用户能对环境的变化做出快速反应。
- 提高数据安全性。存取数据的用户数越多，数据的安全风险就越大。许多公司都要投入大量的时间、精力和资金来确保数据的安全性。DBMS 则为加强数据隐私和安全方面提供了一个较好的框架。
- 良好的数据集成能力。对管理有序的数据进行更大范围存取有助于得到企业操作的集成视图和全局数据视图，这样使得观察公司某个部门的活动对其他部门所产生影响变得更加容易。
- 减少数据不一致。数据不一致性（Data Inconsistency）是指相同数据在不同位置出现不同的版本。例如，一个公司的销售部门存储了一个名为 Bill Brown 的销售代理人信息，但人事部门却将他的名字存为 William G. Brown；公司某地区的销售部门显示某产品的价格为 45.95 美元，而公司总部的销售部门却显示该产品的价格为 43.95 美元。这些都是数据不一致。一个设计良好的数据库能大大减低产生数据不一致的可能性。
- 提高数据存取能力。DBMS 能快速回答即席查询。从数据库观点来看，一个查询（Query）就是发送给 DBMS 的一个特定的数据操作请求，即读取或更新数据。简言之，一个查询就是一个提问，一个即席查询（Ad Hoc Query）就是一个即兴提问。DBMS 将查询结果集（Query Result Set）作为答案返回给应用程序。例如，当终端用户处理大量销售数据时，可能希望 DBMS 能快速回答下列问题（即席查询）：
 - 过去 6 个月产品销售额是多少？
 - 过去 3 个月每位销售人员的奖金是多少？