21 世纪计算机应用技术系列规划教材

数据库原理及 CASE 技术教程

刘甫迎　党晋蓉　编著

人 民 邮 电 出 版 社

需要全本请访问 www.ertongbook.com

计算机应用基础编委会

主　任：聂克成

副主任：肖祥省　杨克复

委　员：曹克武　江俊辽　童　伟　冉兆春　杨秋黎
张金标　孙　雷

本书执行主编：肖祥省

副主编：杨秋黎　罗奕玥　张晓宇

编　者：肖祥省　杨秋黎　罗奕玥　张晓宇　郑利锋
王　峰　张金辉　邱天爽

图书在版编目（CIP）数据

数据库原理及 CASE 技术教程/刘甫迎，党晋蓉编著. —北京：人民邮电出版社，2005. 10
ISBN 7-115-14063-4

Ⅰ. 数… Ⅱ. ①刘… ②党… Ⅲ. 数据库系统 Ⅳ. TP311.13

中国版本图书馆 CIP 数据核字（2005）第111802号

内 容 提 要

本书共有 11 章和两个附录，详细介绍数据库的基础，数据库的由来和发展，数据模型，关系数据库，函数依赖、范式，PowerDesigner（PD 工具），概念数据模型（CDM），物理数据模型（PDM），面向对象模型（OOM），数据库的安全性与完整性、并发控制与恢复等数据库及 CASE 设计的基本概念、原理和技术。叙述后端大型数据库管理系统的工业标准——Oracle，并将其作为上述基本理论的具体例子贯穿全书。特别是书中还叙述了 Oracle、PL/SQL 以及 Oracle 的开发工具 Forms、Reports、企业管理器 Enterprise Manager 等内容，使本书同时又成了一本学习 Oracle 的基于 C/S 模式编程的教科书。

本书附有实例、实验指导书、习题和教学大纲，便于学习与教学，可作为高等学校及软件学院的教材也适于从事数据库软件开发和应用的人员参考。

21 世纪计算机应用技术系列规划教材

数据库原理及 CASE 技术教程

◆ 编　　著　刘甫迎　党晋蓉
责任编辑　潘春燕

◆ 人民邮电出版社出版发行　　北京市崇文区夕照寺街 14 号
邮编　100061　　电子函件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
北京艺辉印刷有限公司印刷
新华书店总店北京发行所经销

◆ 开本：787×1092　1/16
印张：21.5
字数：519 千字　　2005 年 10 月第 1 版
印数：1– 3 000 册　　2005 年 10 月北京第 1 次印刷

ISBN 7-115-14063-4/TP・5010

定价：28.00 元

读者服务热线：（010）67170985　印装质量热线：（010）67129223

编者的话



数据库是计算机管理信息的“仓库”，是当今信息时代广泛使用的不可或缺的重要工具。现在，数据库技术已经成为计算机领域中最重要的技术之一，它是软件学科中一个独立的分支，越来越多的人渴望学习数据库。

目前，国内高校该类课程的教学内容一般比较偏重于数据库理论或完全是数据库理论的教学，并且大多是传统的教学模式，即十分注重数据库原理公式的推导和介绍。这样，该类课程基本上是以数据库自身的体系为脉络展开教学的。例如，一般都是从网状、层次数据库再讲到关系数据库等，没有与具体数据库及其设计工具相结合。结果导致学生的数据库设计和应用能力不够。笔者认为数据库原理课程不单单是要学习数据库原理本身，更重要的是掌握数据库领域内的实践动手能力、分析问题解决问题的能力和掌握数据库管理员（DBA）岗位职业的技能。本教材就是在此改革的思路下编写的，其特点如下。

第一，打破传统的讲授该课程只讲纯理论的惯例，把目标瞄向大型数据库 Oracle，并将其作为数据库基本理论的具体例子贯穿整个课程教学，便于学生理解、应用。

第二，由于把课程教学目标纳入专业技能的培养目标中，因此目标具体明确，学生学习兴趣大，热情高。

第三，由于作者主持和参加过大量的与数据库有关的科研项目，积累了数据库系统开发的实践经验，并将科研经验引入本书，因此本书注重“产、学、研相结合”，并为后续课程学习乃至在毕业设计和就业实习阶段的项目科研活动提供数据库方面的支持。

第四，本书内容新颖，涵盖了高版本对象关系数据库 Oracle9i、10g 等新技术。

第五，创新性地将 CASE（计算机辅助软件工程）设计技术与数据库原理结合起来进行教学，国内鲜见（第 5 章介绍了 CASE，第 7 章介绍用 PD 建立 CDM 和 PDM，第 9 章讲述 OOM 的用例图、时序图、类图）。

第六，选材上，遵循数据库原理的理论“以必需够用为度”，减轻学生负担。突出了数据库管理员（DBA）等岗位职业技能的内容。

第七，突出实践动手能力和实用性，突出案例（第 11 章为综合应用实例），配有实验指导书、习题、教学大纲等，力图使学生学习本书后基本可以编制和管理 Oracle 的程序和系统。

本书可作为高等院校及软件学院的教材，也适于从事数据库软件开发和应用人员参考。本书由刘甫迎教授、党晋蓉副教授编写。刘甫迎编写第 2 章、第 4 章、第 5 章、第 8 章、第

9 章和附录；党晋蓉编写第 1 章、第 3 章、第 6 章、第 7 章、第 10 章、第 11 章。全书由刘甫迎统稿。在编著和出版的过程中我们的学生李蕾蕾、楚丽娜、陈宇等协助做了不少工作，人民邮电出版社的编辑给予了很大的帮助，在此一并表示感谢！

由于水平有限，书中错误难免，请斧正。

作　者

2005 年 8 月

需要全本请访问 www.ertongbook.com

目　录

第 1 章　数据库基础 …… 1

1.1　数据库概念 …… 1
1.1.1　信息处理及数据处理 …… 1
1.1.2　数据模型和数据库技术 …… 2
1.2　数据库的由来和发展 …… 2
1.2.1　人工管理阶段（20 世纪 50 年代中期以前） …… 3
1.2.2　文件系统阶段（20 世纪 50 年代后期至 60 年代中后期） …… 4
1.2.3　数据库阶段（20 世纪 60 年代末开始） …… 4
1.2.4　高级数据库阶段（20 世纪 70 年代后期开始） …… 6
1.3　数据库的体系结构 …… 10
1.4　数据库系统和数据库管理员 …… 13
1.5　数据库管理系统 …… 16
习题 …… 19

第 2 章　关系模型及关系数据库 …… 20

2.1　关系模型和基本概念 …… 20
2.1.1　关系的定义 …… 20
2.1.2　关系模型 …… 21
2.1.3　关系数据语言概述 …… 22
2.2　关系代数 …… 23
2.2.1　传统的集合运算 …… 23
2.2.2　专门的关系运算 …… 24
2.3　关系演算 …… 27
2.3.1　元组关系演算 …… 28
2.3.2　域关系演算 …… 30
2.4　关系数据库标准语言——SQL …… 30

2.5 Oracle 对象关系数据库系统 …… 33
2.5.1 Oracle 系统的特点、产品结构及组成 …… 34
2.5.2 Oracle 的体系结构（实例的进程结构和内存结构） …… 35
2.5.3 Oracle 的配置方案 …… 42
2.5.4 Oracle 运行过程 …… 42
习题 …… 43

第 3 章 数据库的 SQL …… 45

3.1 Oracle 的 SQL 介绍 …… 45
3.2 定义、修改、删除表（Table） …… 47
3.2.1 创建表（Creat Table） …… 47
3.2.2 修改表结构 …… 53
3.2.3 索引的定义 …… 54
3.2.4 删除表、索引 …… 55
3.3 模式对象、直接量、函数和表达式 …… 56
3.3.1 模式对象 …… 56
3.3.2 对象及成分的命名 …… 56
3.3.3 直接量 …… 57
3.3.4 数据类型 …… 58
3.3.5 空值 …… 58
3.3.6 伪列 …… 58
3.3.7 注释 …… 59
3.3.8 算符 …… 59
3.3.9 函数 …… 62
3.3.10 表达式与条件 …… 64
3.4 数据操纵语言 …… 64
3.4.1 将新行插入（INSERT）表 …… 64
3.4.2 修改（UPDATE）表的行 …… 65
3.4.3 从表删除（DELETE）行 …… 66
3.5 视图 …… 66
3.5.1 视图定义 …… 66
3.5.2 视图的查询 …… 66
3.5.3 视图修改 …… 67
3.6 Oracle 的 PL/SQL …… 68
3.6.1 PL/SQL 概述 …… 68
3.6.2 PL/SQL 的结构 …… 69
3.6.3 控制结构 …… 77
3.6.4 游标（CURSOR） …… 79
习题 …… 83

第 4 章　关系数据库设计理论 …… 87

4.1　引言 …… 87
4.2　函数依赖 …… 88
4.3　范式 …… 90
4.3.1　第一范式（1NF） …… 90
4.3.2　第二范式（2NF） …… 91
4.3.3　第三范式（3NF） …… 91
4.3.4　BCNF …… 92
4.4　多值依赖和 4NF …… 93
习题 …… 95

第 5 章　PowerDesigner CASE 技术 …… 96

5.1　CASE 工具概述 …… 96
5.2　PD 分析设计过程及若干级建模技术 …… 99
5.2.1　PD 的分析设计过程 …… 99
5.2.2　PD 的若干级建模功能 …… 100
5.3　PD 的功能 …… 101
5.3.1　PD 的一般功能 …… 101
5.3.2　PD10 的主要特征 …… 102
5.4　PD 的分析设计环境 …… 103
5.4.1　PD 的软硬件要求 …… 103
5.4.2　PD 主界面 …… 105
5.4.3　PD 的分析设计环境 …… 107
5.4.4　PD 的公共资源 …… 108
习题 …… 109

第 6 章　E-R 模型与概念数据模型（CDM） …… 110

6.1　实体—联系模型（E-R） …… 110
6.1.1　实体和实体集合 …… 110
6.1.2　联系和联系集合 …… 111
6.1.3　属性、映射限制和关键字 …… 112
6.1.4　实体联系 E-R 图解和将其归纳为表 …… 115
6.1.5　概括和聚集 …… 119
6.1.6　E-R 数据库模式文件设计 …… 120
6.2　用 PD 创建 CDM …… 122
6.2.1　概念数据模型（CDM）基础 …… 122
6.2.2　用 PD 建立 CDM …… 123
6.2.3　业务规则 …… 126

6.2.4 定义 CDM 中的域 …… 128
6.2.5 定义数据项 …… 132
6.2.6 定义实体 …… 135
6.2.7 定义联系 …… 140
6.2.8 定义继承 …… 144
6.3 由 CDM 生成 PDM …… 148
习题 …… 150

第 7 章 物理数据模型（PDM）及建库 …… 151

7.1 用 PD 创建 PDM …… 151
7.1.1 PDM 及其与 DBMS 的关系 …… 151
7.1.2 建立物理数据库模型（PDM） …… 152
7.1.3 定义索引 …… 155
7.1.4 定义视图 …… 157
7.1.5 生成数据库 SQL 脚本 …… 161
7.1.6 PDM 中的用户管理 …… 162
7.1.7 检查 PDM 中的对象 …… 163
7.1.8 逆向工程 …… 166
7.2 数据库的实现与维护 …… 170
7.2.1 数据库的实现 …… 170
7.2.2 其他设计工作 …… 171
7.2.3 运行与维护 …… 172
7.3 用 Oracle 建立数据库 …… 172
7.3.1 数据库结构和空间管理 …… 172
7.3.2 用企业管理器工具建库表 …… 179
习题 …… 209

第 8 章 数据库保护 …… 211

8.1 数据库的安全性 …… 211
8.1.1 数据库的存取控制和用户的建立 …… 212
8.1.2 特权和角色 …… 215
8.1.3 审计 …… 221
8.2 数据完整性（数据库触发器） …… 221
8.2.1 完整性约束 …… 221
8.2.2 数据库触发器 …… 222
8.3 并发控制 …… 225
8.3.1 数据库不一致的类型 …… 225
8.3.2 封锁 …… 226
8.3.3 Oracle 多种一致性模型 …… 226

8.3.4　封锁机制 …… 227
8.3.5　手工数据封锁 …… 228
8.4　数据库后备和恢复 …… 229
8.4.1　数据库恢复所使用的结构 …… 229
8.4.2　在线日志 …… 229
8.4.3　归档日志 …… 231
8.4.4　数据库后备 …… 231
8.4.5　数据库恢复 …… 232
习题 …… 233

第 9 章　PowerDesigner 的面向对象模型（OOM） …… 235

9.1　UML 概述 …… 235
9.1.1　UML——统一建模语言 …… 235
9.1.2　PD 与 UML …… 236
9.1.3　本章示例说明 …… 241
9.2　用 PD 建立 OOM …… 242
9.2.1　OOM 建立概述 …… 242
9.2.2　设计用例图 …… 245
9.2.3　设计时序图 …… 250
9.2.4　设计类图 …… 254
9.3　从 OOM 生成源程序 …… 263
9.3.1　从 OOM 生成 Java 源文件 …… 263
9.3.2　编译和运行 Java 应用程序 …… 266
习题 …… 267

第 10 章　分布式数据库系统 …… 269

10.1　分布式数据库系统（客户/服务器结构、服务器-服务器结构） …… 269
10.1.1　分布式数据库的特征 …… 271
10.1.2　全功能分布式数据库的规则和目标 …… 271
10.2　分布式数据库的连接 …… 272
10.2.1　分布式数据库全局名和数据库链 …… 273
10.2.2　连接客户和服务器 …… 275
10.2.3　连接服务器到其他服务器 …… 279
10.3　分布式查询处理及其他 …… 280
10.3.1　建立分布式查询 …… 280
10.3.2　在分布式事务中的其他语句 …… 280
10.4　事务管理 …… 281
10.4.1　事务 …… 281
10.4.2　Oracle 的事务管理 …… 281

10.5 数据库应用程序开发工具——Oracle Developer …… 283
10.5.1 Forms Builder …… 283
10.5.2 Reports Builder …… 305
习题 …… 311

第 11 章 综合实例——出版公司信息管理系统 …… 313

11.1 运行环境 …… 313
11.2 具体操作 …… 313
11.3 源代码 …… 321

附录 A 《数据库原理及 CASE 技术教程》教学大纲 …… 326

附录 B 实验指导书 …… 328

主要参考文献 …… 333

需要全本请访问 www.ertongbook.com

第 1 章
数据库基础



本章介绍数据库的概念、发展史、体系结构、数据库系统以及数据库管理系统等，以便对数据库有一个基本的了解。

1.1 数据库概念

1.1.1 信息处理及数据处理

诞生于 20 世纪中叶的计算机科学，较之于其他现代科学技术的发展更为迅猛。21 世纪到来之际，它几乎可以称为“知识爆炸”了。21 世纪是信息和知识的社会，如何组织和利用这些庞大的信息和知识，已成为衡量一个国家科学技术水平高低的重要标志。

早在上世纪 60 年代，数据库技术作为现代信息系统基础的一门软件学科便已诞生。现在，数据库技术已成为计算机领域中最重要的技术之一，它是软件学科中一个独立的分支。它的出现使得计算机应用渗透到工农业生产、商业、行政、教育、科学研究、工程技术和国防军事的各个部门。管理信息系统（MIS）、办公自动化系统（OA）、决策支持系统等都是使用了数据库管理系统或数据库技术的计算机应用系统。

数据库（Database，DB）是存储在一起的相关数据的集合，是存储数据的“仓库”。因此，要理解数据库就需要先了解在数据处理领域中常遇到的两个基本概念：“信息”（information）和“数据”（data）。

“信息”是关于现实世界事物存在方式和运动状态的一种组合反映。例如，上课用的黑板，它的颜色是黑的，矩形样，尺寸是长 3.2m、高 1.4m，由木料制作而成。这些都是关于黑板的信息，都是关于黑板存在状态的反映，它们从不同角度“反映”或“刻画”了黑板这个事物。信息传递需要物质载体，信息的获取和传递要消耗能量。信息可以感知，可以存储，并且可以加工、传递和再生。电子计算机是信息处理领域中最先进的工具之一，可以对收集到的信息进行取舍、加工和整理。

数据，通常是指用符号记录下来的可加以鉴别的信息。例如，为了描述黑板的信息，可以用“黑色、矩形、3.2m×1.4m”这样一组数据来表示。由于“黑色”、“矩形”、“3”、“m”……这些符号已经被人们赋予了特定的语义，所以它们就具有了传递信息的功能。

从上面的例子中，可以看到信息和数据之间的固有联系：数据是信息的符号表示或称为载体，信息则是数据的内涵，是对数据的语义解释。另一方面，某一具体信息与表示它的数据的这种对应关系又因环境而异。同一信息可能有不同的符号表示，同一数据在不同场合可能有不同的解释。数据处理领域中的数据概念，较之科学计算领域中的数据概念已经大大地拓宽了。定义中所说的“符号”，不仅包含数字符号，而且包含文字，图像和其他符号；所谓“记录下来”，也不仅是指用笔写在纸上，它还包括磁记录、光刻等各种记录形式。

由于信息是现实世界中事物的存在方式和运动状态的反映，而现实世界的事物常常是相互关联的，这就使得人们在了解、掌握事物之间固有联系和运动规律的基础上，可以从一些已知的信息出发，经过演绎推理，导出新的信息，为人类社会生活的各种需要服务，这就是常说的“信息处理”。例如，根据上述黑板的尺寸数据和木工定额标准，可以推算出制作黑板所需的木材数量和制作费用。

一般，人们称原始信息为源数据。对源数据进行综合推导加工后，可以得出新的数据。这些结果数据表示了新的信息，可以作为某种决策的依据（或用于新的推导加工），这整个过程通常称为“数据处理”。

电子计算机使大规模的数据处理成为可能，它和通信、网络技术融合在一起，进一步推动了信息处理的进程，极大地增强了人类社会信息处理的能力。

1.1.2 数据模型和数据库技术

数据库系统指的是实现有组织地、动态地存储大量相关数据，方便用户访问的计算机软、硬件资源组成的系统。数据库技术则是研究数据库的结构、存储、设计和使用的一门软件学科。因此，数据库技术主要是研究如何存储、使用和管理数据。在计算机应用中，数据处理占的比重最大，而数据库系统是数据处理的核心机构，所以它的效能往往决定了整个计算机应用的经济效益。

数据模型是对现实世界客观事物及其联系的描述，它反映数据项之间和记录之间的联系。在数据库技术中，总是使用模型的概念来描述数据库的结构与语义。因此，数据库离不开数据模型。在数据库中，常用的三种数据模型是：层次模型（hierarchical mode）、网状模型（network model）和关系模型 relational model）。此外，还有面向对象模型（object_oriented model）等。

数据库这门学科与其他基础软件、系统软件、应用软件有着密切的联系，例如：数据库与操作系统联系紧密，数据库技术是在操作系统的文件系统基础上发展起来的，并且数据库系统必须在操作系统支持下才能工作。数据库与网络技术与多媒体技术的关系很密切，例如：分布式数据库要用上网络，数据库甚至可在 Internet 上交换海外的数据、声音、图像、图片等多媒体信息。数据库技术与数据结构的关系也是不可分的，数据库技术不仅要用到数据结构的知识，而且丰富了数据结构的内容。程序设计是使用数据库系统的最基本方式，数据库中大量的应用程序过程大都是用高级语言加上数据库的操纵语言编写的。此外，集合论、数理逻辑是关系数据库的理论基础，其很多概念、术语、思想都直接用到关系数据库中。

1.2 数据库的由来和发展

数据库这个名词起源于 20 世纪 50 年代，当时美国为了战争的需要，把各种情报集中在

一起，存入计算机，称为 Information Base 或 Database。1963 年美国 Honeywell 公司的 IDS（Integrated Data Store）系统投入运行，揭开了数据库技术的序幕。1965 年美国利用数据库设计了阿波罗登月火箭，推动了数据库技术的产生。当时社会上产生了许多行行色色的 Database 或 Databank，但基本上都是文件系统的扩充。1968 年美国 IBM 公司推出了层次模型的 IMS 数据库系统，并于 1969 年形成产品；1969 年，推出 COBOL 语言的美国数据系统语言协会（Conference on Date System Language，CODASYL），组织了数据库任务组（DBTG），并发表了网状数据库系统的标准文本（1971 年正式通过）；1970 年初，IBM 公司的高级研究员 E.F.Codd 发表论文，提出了数据库的关系模型，从而奠定了关系数据库的理论基础。

上世纪 70 年代是数据库蓬勃发展的年代。网状系统和层次系统占领着市场，关系系统则处于实验阶段，IBM 公司研制出了原型关系语言 System R。1979 年关系软件（Relational Software）公司推出了第一个基于 SQL 的商用关系数据库产品 Oracle。

上世纪 80 年代起，关系型数据库产品已相当成熟，取代了网状系统和层次系统的市场。与此同时关系数据库理论日臻完善，走向更高级的阶段，有了分布式数据库系统（Ditributed database systems）等。后来，从不同的计算机应用领域提出了许多数据库的非传统应用课题，诸如多媒体数据、空间数据、时序数据、科学数据、复杂对象、知识、超文本管理等。为了适应这类应用的需要，提出了不少新的概念、新的数据模型和系统结构。经过几年的研究和实践，逐步形成了面向对象数据库系统（Object_Oriented Database Systems）、主动数据库系统（Active Database Systems）、大型知识库系统（Large Knowledge Base Systems）、数据库中的知识发现（Knowledge Discovery in Database）以及科学数据库（Science Database）等热点。可以预言，在 21 世纪，数据库技术必将获得更加长足的发展。

综观数据管理技术的发展可知：它与硬件（主要是外部存储器）、软件以及计算机应用的范围有密切关系。数据管理技术的发展，大致经过以下四个阶段：人工管理阶段、文件系统阶段、数据库阶段以及高级数据库阶段。

1.2.1　人工管理阶段（20 世纪 50 年代中期以前）

这一时期的计算机主要用于科学计算。硬件中的外存只有卡片、纸带等；软件只有汇编语言，没有数据管理方面的软件，数据处理的方式基本上是批处理。这个时期的数据管理特点如下：

1. 数据不保存

进行某一课题计算时将原始数据随程序一起输入主存，运算处理后将结果数据输出。任务完成后，数据空间同程序空间一起释放。

2. 没有专用软件对数据进行管理

每个应用程序要包括存储结构、存取方法、输入输出方式等，数据结构与程序不具有独立性，一旦存储结构改变，就必须由应用程序员修改程序。由于程序直接面向存储结构，因此不存在逻辑结构与物理结构的区别。

3. 只有程序（PROGRAM）的概念、没有文件（FILE）的概念

即使有文件，也大多是顺序文件，其他组织方式必须由程序员自行设计与安排。

4. 数据面向应用

即一组数据对应于一个程序。由于各应用程序处理的数据可能会有这样那样的联系，因

此应用程序的数据处理之间会存在很多的重复劳动。

5. 对数据的存取以记录为单位

其灵活性差。

1.2.2 文件系统阶段（20 世纪 50 年代后期至 60 年代中后期）

这一时期的计算机不仅用于科学计算，还大量用于信息管理。外存已有磁盘、磁鼓等直接存取设备。软件方面出现了高级语言和操作系统。操作系统中的文件系统（有时也称为“信息处理模块”）是专门处理外存的数据管理软件。处理数据方式有批处理，也有联机实时处理。这一阶段数据管理情况的特点和存在的问题如下。

1. 特点

（1）数据可长期保存在外存的磁盘上。用户经常随时通过程序对文件进行查询、修改以及删除等处理。由于计算转向管理，数据处理的工作量增大。

（2）数据的物理结构与逻辑结构有了区别，但较简单。程序与设备之间有设备独立性（程序只需用文件名与数据打交道，不必关心数据的物理位置），由操作系统的文件系统提供存取方法（读/写）。由存取方法实现数据的逻辑结构与物理结构之间的转换。

（3）文件的形式已多样化，有索引文件，链接文件和直接存取等，因而对文件的记录可顺序访问，也可随机访问。但文件之间是独立的，联系要通过程序去构造，文件的共享性差。

（4）由于有了存储文件，数据不再仅仅属于某个特定的程序，可以重复使用。但文件结构的设计仍然是基于特定的用途，程序仍然是基于特定的物理结构和存取方法编制的。因此，数据结构与程序之间的依赖关系并未根本改变。

（5）对数据的存取基本上还是以记录为单位。

2. 缺陷

这种文件系统，未能彻底体现用户观点下的数据逻辑结构独立于数据在外存的物理结构的要求。因此，数据的物理结构修改时，仍然需要修改用户的应用程序。

文件系统有三大缺陷：

（1）数据冗余性（redundancy）。由于文件之间缺乏联系，造成每个应用程序都有对应的文件，有可能同样的数据在多个文件中重复存储。

（2）不一致性（inconsistency）。由于数据冗余造成数据的不一致性，在进行更新操作时，稍不小心同样的数据在不同的文件中就可能不一样。

（3）数据联系弱（poor data relationship），这是文件之间独立、缺乏联系造成的。

由于这些原因，促使人们开始研究一种新的数据管理技术，从而导致在上世纪 60 年代末产生了数据库技术，进入数据管理技术的第三阶段。

1.2.3 数据库阶段（20 世纪 60 年代末开始）

20 世纪 60 年代末，磁盘技术取得了重大进展，大容量（数百兆字节以上）和快速存取的磁盘陆续进入市场，成本有了很大的下降，这为数据库技术的实现提供了物质条件。

上世纪 60 年代中期，出现的大多数系统（database 或 databank）还不能称为真正的数据库系统。数据管理技术进入数据库阶段的标志是前面讲到的 60 年代后期的三大事件：即 1969 年 IBM 公司推出的 IMS 产品（层次数据库系统）和 CODASYL 研究和建议的 DBTG 系统（网

状数据库系统），以及 1970 年起，IBM 公司 E.F.Codd 连续发表的一系列论文，奠定了关系数据库理论的基础。

上世纪 70 年代以来，数据库技术得到迅速发展，并投入了实际的应用。根据数据库系统支持的数据模型（data model），其发展已经历了三代：从上面讲到的层次网状代，到关系模型代以及后面高级数据库阶段涉及的面向对象代。关系模型代从上世纪 70 年代初 E.F.Codd 奠定了关系数据库理论基础后，70 年代末推出了一些试验系统，80 年代初出现了一批商品化的关系数据库系统，如 Oracle、SQL/DS、DB2、INGRES、INFORMIX、UNIFY 以及 dBASE、FoxBASE 等。SQL 语言已在 1986 年被美国 ANSI 和国际标准化组织（ISO）采纳，作为关系数据库语言的国际标准。

与文件系统相比，数据库系统克服了文件系统的缺点，提供了对数据更深层次和更有效的管理。概括起来，数据库技术的管理方式具有以下特点。

1. 采用复杂的数据模型（结构）

数据模型不仅要描述数据本身的特点，还要描述数据之间的联系。这种联系是通过存取路径来实现的。通过一切存取路径来表示自然的数据联系是数据库与传统文件的根本区别。这样数据不再面向特定的某个或多个应用，而是面向整个应用系统。数据冗余明显减少，实现了数据共享。

2. 有较高的数据独立性

在数据库系统中，系统提供映象的功能，确保应用程序对数据结构和存取方法有较高的独立性。数据的物理结构和逻辑结构之间的差别可以很大。用户以简单的逻辑结构操作数据，根本不去考虑数据的物理结构。在数据库系统中，数据库结构分成用户的逻辑结构、整体逻辑结构和物理结构，如图 1-1 所示。用户（应用程序或终端用户）的数据和外存中的数据之间的转换由数据库管理系统实现。为提高效率、减少冗余或增加新的数据，常需改变数据结构。在改变物理结构时，不影响整体逻辑结构、用户的逻辑结构以及应用程序，这样就认为数据库达到了物理数据的独立性。在改变整体逻辑时，不影响用户的逻辑结构以及应用程序，这样就认为数据库达到了逻辑数据的独立性。

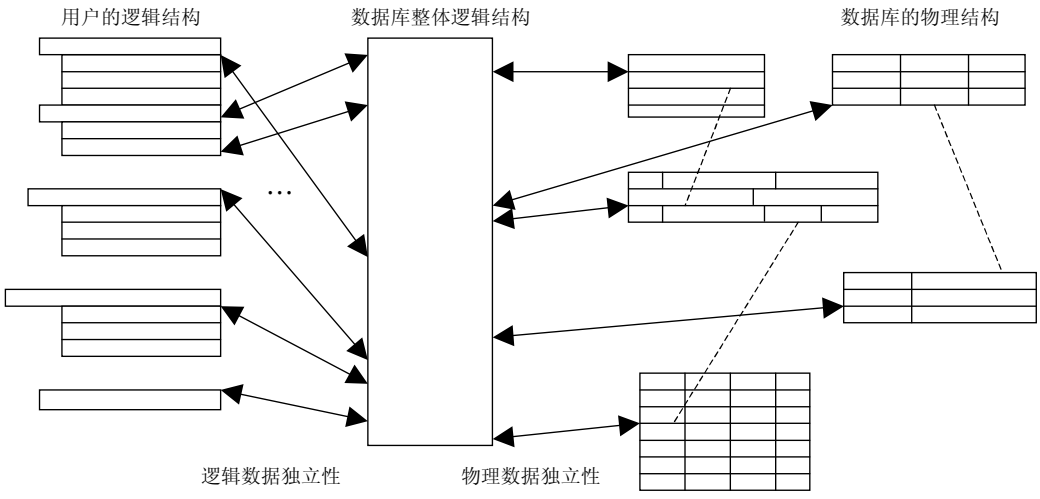


图 1-1　数据库系统的结构

3. 数据库系统为用户提供了方便的用户接口

用户可使用查询语言或简单的终端命令来操作数据库，也可以通过程序方式（如高级语言 C、FORTRAN 等，以及数据库操纵语言编制的程序）操作数据库。

4. 提供下述四方面的数据控制功能

（1）数据完整性：保证数据库始终包含正确的数据。用户可设计一些完整性规则以确保数据值的正确性。例如可把数据值限制在某个范围内，并对数据值之间的联系进行各种检验。

（2）数据安全性：保证数据的安全和机密，防止数据丢失或被窃取。

（3）数据库的并发控制：避免并发程序之间的相互干扰，防止数据库数据被破坏，杜绝提供给用户不正确的数据。

（4）数据的恢复：在数据库被破坏时或数据不可靠时，系统有能力把数据库恢复到最近某个时刻的正确状态。

5. 以数据项作为操作单位

在数据库中，对数据的操作除了以记录为单位外，还可以数据项为单位。

数据库阶段的程序和数据的联系可用图 1-2 表示。

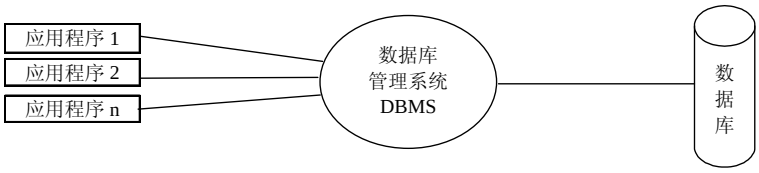


图 1-2　程序和数据的联系

综上所述，可以把数据库定义为：为某个特定组织多种应用服务的、具有尽可能小冗余度的互相关联的存储数据的集合。该集合的数据结构独立于使用数据的程序，对数据的增添、删除、修改及检索，由系统统一控制，采用的数据模型有利于将来应用的发展。

从文件系统发展到数据库技术，是信息处理领域的一个重大变化。在文件系统阶段程序设计处于主导地位，数据只起着服从程序设计需要的作用；而在数据库方式下，数据开始占据了中心位置，数据的结构设计成为信息系统首先关心的问题，而利用这些数据的应用程序设计则退居到以既定的数据结构为基础的外围地位。

目前，国内外数据库应用已相当普及，各行业都建立了以数据库技术为基础的大型计算机网络系统，并在国际互联网（Internet）的基础上建立了国际性联机检索系统，其应用深入到了人类社会生活的各个领域，甚至家庭。

1.2.4　高级数据库阶段（20 世纪 70 年代后期开始）

高级数据库阶段的主要标志，是分布式数据库系统、面向对象数据库、智能数据库系统的出现。

1. 分布式数据库系统（distributed database systems）

集中式数据库把数据集中在一个数据库中进行集中管理。这种管理方式减少了数据冗余和不一致性，数据间的联系比文件系统强得多。但集中式系统也有弱点，如系统庞大，通信

需要全本请访问 www.ertongbook.com