高等院校信息管理与信息系统专业系列教材

数据库系统原理教程

王珊　　陈红编著

清华大学出版社

需要全本请访问 www.ertongbook.com

(京)新登字 158 号

内 容 提要

本书系统、完整地讲述了当前数据库技术的基本原理和应用实践。主要内容包括数据模型、数据库系统结构、关系数据库、SQL 语言、关系数据库设计理论、数据库保护、数据库设计、关系数据库管理系统实例、数据库技术新进展等。每章后均附有习题。

本书可作为高等院校信息管理专业数据库课程的教材，也可供从事计算机开发与应用的科研人员、工程技术人员以及其他有关人员参考。

版权所有,翻印必究。
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。

图书在版编目(CIP)数据

数据库系统原理教程／王珊,陈红编著. 一北京：清华大学出版社,1998.7
高等院校信息管理专业系列教材
ISBN 7-302-03009-X

Ⅰ.数… Ⅱ.陈… Ⅲ.数据库系统-理论-高等学校-教材 Ⅳ.TP311.13

中国版本图书馆 CIP 数据核字(98)第 14773 号

出 版 者： 清华大学出版社(北京清华大学学研楼,邮编 100084)
http://www.tup.tsinghua.edu.cn
责任编辑：范素珍
印 刷 者： 中国科学院印刷厂
发 行 者： 新华书店总店北京发行所
开 本 ： 787×1092 1/16 印张： 17.5 字数：434 千字
版 次 ： 1998 年 7 月第 1 版 1999 年 12 月第 5 次印刷
书 号 ： ISBN 7-302-03009-X/TP·1599
印 数 ： 30001～38000
定 价 ： 18.50 元

出版说明

20世纪三四十年代，长期搜索前进的古老的计算技术与刚走向成熟的电子技术结合。这一结合，不仅孕育了新一代计算工具——电子计算机,还产生了当时谁也没有料到的巨大效应：电子计算机——这种当初为计算而开发出来的工具，很快就超出计算的范畴，成为“信息处理机”的代名词；人类开始能够高效率地开发并利用信息；信息对人类社会的作用得以有效地发挥，并逐步超过材料和能源成为人类社会的重要支柱；信息产业急剧增长，信息经济高速发展，社会生产力达到了新的高度；人们的信息化意识不断加强，人类在信息资源方面开始更加激烈的竞争，社会发展走上信息化轨道。

科学技术是第一生产力，教育是基础。为了加速社会信息化的过程，以培养信息资源开发人才为目标的信息管理专业应运而生。

从与信息有关的学科纵向来看,信息管理处于信息学、信息技术、信息管理、信息经济、信息社会学这个层次的中间，它下以信息学和信息技术为基础，上与信息经济和信息社会学相关联。从其涉及的学科横向来看,它处在管理学、信息科学与技术、系统科学等有关学科领域的交叉点上。它对技术有极高的要求，又要求对组织的深刻理解和对行为的合理组织，反映了科学与人本融合的特点。这种交叉和融合正是信息管理的最重要特征，是别的学科或专业难以取代和涵盖的。

我国的信息管理专业创建于20世纪70年代末。在不到20年的时间里，已发展到150多个点，成为培养信息化人才的主要摇篮。其发展速度之快、影响之深远已令世人和学术界刮目相待。

然而，作为一个新的学科，这个专业的课程体系、教学内容以及教学方法都需要经历一个逐步完善、逐步成熟的过程。特别是教材的建设更需要经过长期的实践和探索。没有这样一个过程,具有专业特点、符合中国实际的教材是不可能产生的。近20年来大家一直在课程体系的完善和建设有自己专业特点的教材方面不断进行探讨。1991年全国10所财经类院校的经济信息管理专业负责人会聚在太原召开第一次教学研讨会。以后，1993年在大连、1995年在武汉、1997年在烟台,又有更多的院校参加了这一研讨。在讨论中,各校的同仁一致认为，教材建设是当务之急，它不仅直接体现和落实培养目标，同时也是学科建设的根本所在。目前一些课程缺乏专业特点,简单搬用其他专业教材的状况亟待改变。在武汉会议上，这一共识得到了与会的国家教委有关部门负责同志的赞许，清华大学出版社也对此表示了热情的支持。会议确定了首批计划编写八九本教材，由张基温教授主持实施，由清华大学出版社出版。在实施过程中，还聘请了魏晴宇、陈禹两位教授作为顾问。

经过两年多的工作,在全国许多高等院校的同仁共同努力下,其中7本已完成初稿。我们希望这批教材的问世，能够起到抛砖引玉的作用，对各校信息管理专业的建设与发展有所裨益。

近20年来的实践使我们对信息管理专业的重要性和困难有了切身的体会。一方面，席卷全球的信息化大潮把信息管理推到了时代发展的前沿，信息、信息管理、信息系统已经成

为全社会注视的热点。这为信息管理专业的建设创造了良好的外部条件，提供了难得的机遇。另一方面,信息技术的迅速发展与普及,多种社会经济因素的互相渗透和影响,前所未有的许多新问题、新情况的出现，又给这个专业的发展带来了很大的困难。我们深感责任之重大和任务之艰巨。在这套教材问世之时，我们再次表示这样一个心愿：希望与全国的同行共勉,为祖国信息化建设的宏伟事业多添一块砖,多加一块瓦,多出一份力,培养出更多的优秀人才。

由于如上种种原因，这套教材当然不会是完整的，也不会是完美的。它必然要不断补充、不断修改、不断完善。因此,对于它的任何修改意见,都是我们非常盼望的。希望能够在这套教材出版后，收到更多的意见和建议，使之逐步走向成熟。

全国高校计算机基础教育研究会
财经信息管理专业委员会
信息管理与信息系统专业教材编委会

1997.9

前　言

数据库技术产生于 20 世纪 60 年代末,发展至今已有近 30年的历史。数据库技术作为数据管理的最有效的手段，它的出现极大地促进了计算机应用的发展，目前基于数据库技术的计算机应用已成为计算机应用的主流。

近 30 年来，数据库技术本身也在不断地发展和完善。关系数据库已取代了早期的层次数据库与网状数据库，成为主流数据库，而新一代数据库也逐渐露出头角。本书以关系数据库为重点，比较全面系统地介绍了数据库的基本概念和基本技术。取材上力图反映当前数据库技术的发展水平和发展趋势。

本书共 8 章。第 1 章绪论,概述了数据管理的进展、数据模型、数据库管理系统和数据库工程的基本概念。

第 2 章至第 4 章讲解了关系数据库的数据模型、数据语言和数据理论，其中对关系数据库的标准语言 SQL 进行了深入介绍。

第 5 章详细讨论了数据库的安全性、完整性、并发控制和恢复等数据库保护技术，并以一个关系数据库产品为例，说明数据库保护技术在实际产品中是如何实现的。

第 6 章讲述了设计数据库应用系统的方法。重点放在设计关系数据库应用系统上。

第 7 章介绍关系数据库产品的发展过程和 5 个关系数据库产品实例。

第 8 章数据库技术的新进展介绍了数据库技术的发展过程和新一代数据库系统，包括分布式数据库、并行数据库、主动数据库、对象一关系数据库、数据仓库、工程数据库、统计数据库、空间数据库等。

为了方便读者学习，每章后面都附有一定量的习题。

在本书的编写过程中，山西财经大学的张基温教授提出了许多宝贵意见，在此表示诚挚的谢意。由于我们水平有限，书中难免存在许多不足之处，恳请读者批评指正。

王珊　　陈红

于中国人民大学信息学院数据与知识工程研究所

需要全本请访问 www.ertongbook.com

目　录

第 1 章　绪论 …… 1
1.1　引言 …… 1
1.1.1　数据、数据库、数据库系统、数据库管理系统 …… 1
1.1.2　数据库技术的产生与发展 …… 2
1.1.3　数据库技术的研究领域 …… 8
1.2　数据模型 …… 9
1.2.1　数据模型的要素 …… 9
1.2.2　概念模型 …… 10
1.2.3　数据模型 …… 13
1.3　数据库系统结构 …… 23
1.3.1　数据库系统的模式结构 …… 23
1.3.2　数据库系统的体系结构 …… 25
1.4　数据库管理系统 …… 27
1.4.1　数据库管理系统的功能与组成 …… 27
1.4.2　数据库管理系统的工作过程 …… 29
1.4.3　数据库管理系统的实现方法 …… 29
1.5　数据库工程与应用 …… 32
1.5.1　数据库设计的目标与特点 …… 32
1.5.2　数据库设计方法 …… 32
1.5.3　数据库设计步骤 …… 33
1.5.4　数据库应用 …… 34
习题 …… 36
第 2 章　关系数据库 …… 37
2.1　关系数据库概述 …… 37
2.2　关系数据结构 …… 38
2.3　关系的完整性 …… 42
2.4　关系代数 …… 45
2.4.1　传统的集合运算 …… 46
2.4.2　专门的关系运算 …… 46
2.5　关系演算 …… 52
2.5.1　元组关系演算语言 ALPHA …… 52
2.5.2　域关系演算语言 QBE …… 57
2.6　关系数据库管理系统 …… 62
习题 …… 64

第 3 章　关系数据库标准语言 SQL …… 66
3.1　SQL 概述 …… 66
3.1.1　SQL 的特点 …… 66
3.1.2　SQL 语言的基本概念 …… 68
3.2　数据定义 …… 68
3.2.1　定义、删除与修改基本表 …… 69
3.2.2　建立与删除索引 …… 71
3.3　查询 …… 72
3.3.1　单表查询 …… 73
3.3.2　连接查询 …… 82
3.3.3　嵌套查询 …… 87
3.3.4　集合查询 …… 95
3.3.5　小结 …… 97
3.4　数据更新 …… 98
3.4.1　插入数据 …… 98
3.4.2　修改数据 …… 99
3.4.3　删除数据 …… 100
3.5　视图 …… 101
3.5.1　定义视图 …… 102
3.5.2　查询视图 …… 105
3.5.3　更新视图 …… 107
3.5.4　视图的用途 …… 108
3.6　数据控制 …… 110
3.7　嵌入式 SQL …… 112
3.7.1　嵌入式 SQL 的一般形式 …… 113
3.7.2　嵌入式 SQL 语句与主语言之间的通信 …… 113
3.7.3　不用游标的 SQL 语句 …… 116
3.7.4　使用游标的 SQL 语句 …… 119
3.7.5　动态 SQL 简介 …… 125
习题 …… 126
第 4 章　关系数据库设计理论 …… 127
4.1　数据依赖 …… 127
4.1.1　关系模式中的数据依赖 …… 127
4.1.2　数据依赖对关系模式的影响 …… 128
4.1.3　有关概念 …… 129
4.2　范式 …… 130
4.2.1　第一范式 (1NF) …… 131
4.2.2　第二范式 (2NF) …… 132
4.2.3　第三范式 3NF) …… 133

4.2.4 BC 范式(BCNF) …… 134
4.2.5 多值依赖与第四范式(4NF) …… 136
4.3 关系模式的规范化 …… 139
4.3.1 关系模式规范化的步骤 …… 139
4.3.2 关系模式的分解 …… 140
习题 …… 143
第 5 章 数据库保护 …… 145
5.1 安全性 …… 145
5.1.1 安全性控制的一般方法 …… 145
5.1.2 ORACLE 数据库的安全性措施 …… 149
5.2 完整性 …… 153
5.2.1 完整性约束条件 …… 153
5.2.2 完整性控制 …… 155
5.2.3 ORACLE 的完整性 …… 158
5.3 并发控制 …… 161
5.3.1 并发控制概述 …… 161
5.3.2 并发操作的调度 …… 164
5.3.3 封锁 …… 165
5.3.4 死锁和活锁 …… 169
5.3.5 ORACLE 的并发控制 …… 172
5.4 恢复 …… 173
5.4.1 恢复的原理 …… 173
5.4.2 恢复的实现技术 …… 174
5.4.3 ORACLE 的恢复技术 …… 178
5.5 数据库复制与数据库镜象 …… 180
5.5.1 数据库复制 …… 180
5.5.2 数据库镜象 …… 182
习题 …… 183
第 6 章 数据库设计 …… 184
6.1 数据库设计的步骤 …… 184
6.2 需求分析 …… 185
6.2.1 需求分析的任务 …… 185
6.2.2 需求分析的方法 …… 185
6.2.3 数据字典 …… 190
6.3 概念结构设计 …… 192
6.3.1 概念结构设计的方法与步骤 …… 192
6.3.2 数据抽象与局部视图设计 …… 192
6.3.3 视图的集成 …… 196
6.4 逻辑结构设计 …… 200

6.4.1 E-R 图向数据模型的转换 …… 200
6.4.2 数据模型的优化 …… 203
6.4.3 设计用户子模式 …… 204
6.5 数据库物理设计 …… 205
6.6 数据库实施 …… 207
6.7 数据库运行与维护 …… 210
习题 …… 213
第 7 章 关系数据库管理系统实例 …… 214
7.1 关系数据库管理系统产品概述 …… 214
7.2 ORACLE …… 216
7.3 SYBASE …… 221
7.4 INFORMIX …… 226
7.5 DB2 …… 231
7.6 INGRES …… 236
习题 …… 240
第 8 章 数据库技术新进展 …… 241
8.1 数据库技术发展概述 …… 241
8.2 数据模型及数据库系统的发展 …… 242
8.2.1 第一代数据库系统 …… 242
8.2.2 第二代数据库系统 …… 243
8.2.3 新一代数据库技术的研究和发展 …… 244
8.3 数据库技术与其它相关技术相结合 …… 248
8.3.1 分布式数据库 …… 248
8.3.2 并行数据库 …… 253
8.3.3 多媒体数据库 …… 257
8.3.4 主动数据库 …… 258
8.3.5 对象-关系数据库 …… 259
8.4 面向应用领域的数据库新技术 …… 260
8.4.1 数据仓库 …… 260
8.4.2 工程数据库 …… 265
8.4.3 统计数据库 …… 266
8.4.4 空间数据库 …… 266
习题 …… 267
参考文献 …… 269

第1章 绪 论

数据库技术产生于 20 世纪 60 年代中期,是数据管理的最新技术,是计算机科学的重要分支，它的出现极大地促进了计算机应用向各行各业的渗透。本章将介绍数据库的有关概念以及为什么要发展数据库技术，从中不难看出数据库技术的重要性所在。

1.1 引 言

1.1.1 数据、数据库、数据库系统、数据库管理系统

数据、数据库、数据库系统和数据库管理系统是与数据库技术密切相关的 4 个基本概念。

1. 数据(data)

说起数据，人们首先想到的是数字。其实数字只是最简单的一种数据。数据的种类很多，在日常生活中数据无处不在：文字、图形、图象、声音、学生的档案记录、货物的运输情况……,这些都是数据。

为了认识世界，交流信息，人们需要描述事物。数据实际上是描述事物的符号记录。在日常生活中人们直接用自然语言（如汉语）描述事物。在计算机中，为了存储和处理这些事物,就要抽出对这些事物感兴趣的特征组成一个记录来描述。例如,在学生档案中,如果人们最感兴趣的是学生的姓名、性别、出生年月、籍贯、所在系别、入学时间,那么可以这样描述：

(李明,男,1972,江苏,计算机系,1990)

数据与其语义是不可分的。对于上面一条学生记录，了解其语义的人会得到如下信息：李明是个大学生,1972 年出生,江苏人,1990 年考入计算机系；而不了解其语义的人则无法理解其含义。可见，数据的形式本身并不能完全表达其内容，需要经过语义解释。

2. 数据库(database,简称 DB)

收集并抽取出一个应用所需要的大量数据之后，应将其保存起来以供进一步加工处理和抽取有用信息。保存方法有很多种:人工保存、存放在文件里、存放在数据库里,其中数据库是存放数据的最佳场所，其原因将在 1.1.2 节中介绍。

所谓数据库就是长期储存在计算机内、有组织的、可共享的数据集合。数据库中的数据按一定的数据模型组织、描述和储存,具有较小的冗余度,较高的数据独立性和易扩展性,并可为各种用户共享。

3. 数据库管理系统(database management system ,简称 DBMS)

收集并抽取出一个应用所需要的大量数据之后，如何科学地组织这些数据并将其存储

在数据库中，又如何高效地处理这些数据呢？完成这个任务的是一个软件系统——数据库管理系统。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件。

数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制。数据库管理系统使用户能方便地定义数据和操纵数据，并能够保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

4. 数据库系统(database system,简称 DBS)

数据库系统是指在计算机系统中引入数据库后的系统构成，一般由数据库、数据库管理系统(及其开发工具)、应用系统、数据库管理员和用户构成。应当指出的是,数据库的建立、使用和维护等工作只靠一个 DBMS 远远不够，还要有专门的人员来完成，这些人称为数据库管理员(database administrator,简称 DBA)。

在不引起混淆的情况下人们常常把数据库系统简称为数据库。

数据库系统可以用图 1-1 表示。

数据库系统在整个计算机系统中的地位如图 1-2 所示。

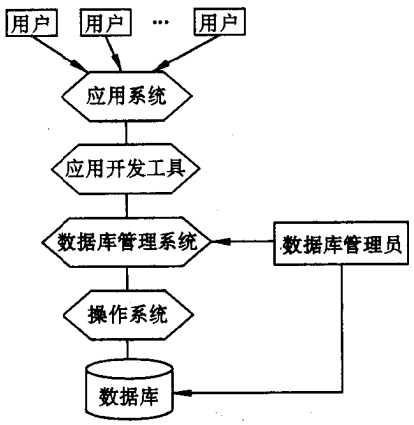


图 1-1 数据库系统

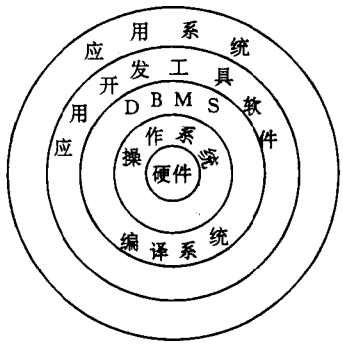


图 1-2 数据库在计算机系统中的地位

1.1.2 数据库技术的产生与发展

数据库技术是应数据管理任务的需要而产生的。

数据管理是指如何对数据进行分类、组织、编码、储存、检索和维护,它是数据处理的中心问题。随着计算机硬件和软件的发展，数据管理经历了人工管理、文件系统和数据库系统三个发展阶段。这三个阶段的比较如表 1-1 所示。

1. 人工管理阶段

在 20 世纪 50 年代中期以前，计算机主要用于科学计算。当时的硬件状况是，外存只有纸带、卡片、磁带,没有磁盘等直接存取的存储设备 软件状况是,没有操作系统,没有管理数据的软件；数据处理方式是批处理。

需要全本请访问 www.ertongbook.com

表 1-1　数据管理三个阶段的比较

		人工管理阶段	文件系统阶段	数据库系统阶段
背景	应用背景	科学计算	科学计算、管理	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理 批处理	联机实时处理 分布处理 批处理
特点	数据的管理者	人	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用程序	现实世界
	数据的共享程度	无共享 冗余度极大	共享性差 冗余度大	共享性高 冗余度小
	数据的独立性	不独立，完全依赖于	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构 整体无结构	整体结构化，用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

人工管理数据具有如下特点：

(1) 数据不保存。由于当时计算机主要用于科学计算，一般不需要将数据长期保存，只是在计算某一课题时将数据输入，用完就撤走。不仅对用户数据如此处置，对系统软件有时也是这样。

(2) 数据需要由应用程序自己管理，没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构，而且要设计物理结构，包括存储结构、存取方法、输入方式等。因此程序员负担很重。

(3) 数据不共享。数据是面向应用的，一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时，由于必须各自定义，无法互相利用、互相参照，因此程序与程序之间有大量的冗余数据。

(4) 数据不具有独立性，数据的逻辑结构或物理结构发生变化后，必须对应用程序做相应的修改，这就进一步加重了程序员的负担。

人工管理阶段应用程序与数据之间的对应关系可用图 1-3 表示。

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期，计算机的应用范围逐渐扩大，计算机不仅用于科学计算，而且还大量用于管理。这时硬件上已有了磁盘、磁鼓等直接存取存储设备；软件方

面，操作系统中已经有了专门的数据管理软件，一般称为文件系统；处理方式上不仅有了文件批处理，而且能够联机实时处理。

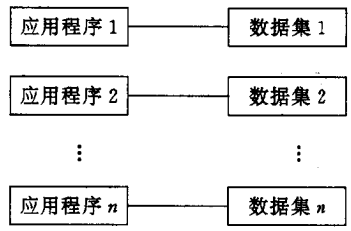


图 1-3　人工管理阶段应用程序与数据之间的对应关系

用文件系统管理数据具有如下特点：

(1) 数据可以长期保存。由于计算机大量用于数据处理，数据需要长期保留在外存上，反复进行查询、修改、插入和删除等操作。

(2) 由专门的软件即文件系统进行数据管理，程序和数据之间由软件提供的存取方法进行转换，使应用程序与数据之间有了一定的独立性，程序员可以不必过多地考虑物理细节，将精力集中于算法。而且数据在存储上的改变不一定反映在程序上，大大节省了维护程序的工作量。

(3) 数据共享性差。在文件系统中，一个文件基本上对应于一个应用程序，即文件仍然是面向应用的。当不同的应用程序具有部分相同的数据时，也必须建立各自的文件，而不能共享相同的数据，因此数据的冗余度大，浪费存储空间。同时由于相同数据的重复存储、各自管理，给数据的修改和维护带来了困难，容易造成数据的不一致性。

(4) 数据独立性低。文件系统中的文件是为某一特定应用服务的，文件的逻辑结构对该应用程序来说是优化的，因此要想对现有的数据再增加一些新的应用会很困难，系统不容易扩充。一旦数据的逻辑结构改变，必须修改应用程序，修改文件结构的定义。而应用程序的改变，例如，应用程序改用不同的高级语言等，也将引起文件的数据结构的改变。因此数据与程序之间仍缺乏独立性。可见，文件系统仍然是一个不具有弹性的无结构的数据集合，即文件之间是孤立的，不能反映现实世界事物之间的内在联系。

文件系统阶段应用程序与数据之间的关系如图 1-4 所示。

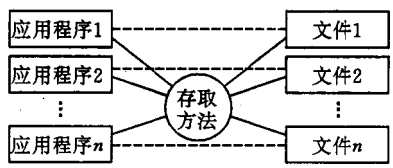


图 1-4　文件系统阶段应用程序与数据之间的对应关系

3. 数据库系统阶段

20 世纪 60 年代后期以来，计算机用于管理的规模更为庞大，应用越来越广泛，数据量急剧增长，同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。这时硬件已有大容量磁盘，硬件价格下降，软件价格上升，为编制和维护系统软件及应用程序所需的

成本相对增加 在处理方式上,联机实时处理要求更多,并开始提出和考虑分布处理。在这种背景下，以文件系统作为数据管理手段已经不能满足应用的需求，于是为解决多用户、多应用共享数据的需求，使数据为尽可能多的应用服务，就出现了数据库技术，出现了统一管理数据的专门软件系统——数据库管理系统。

用数据库系统来管理数据具有如下特点：

(1) 数据结构化

数据结构化是数据库与文件系统的根本区别。

在文件系统中，相互独立的文件的记录内部是有结构的。传统文件的最简单形式是等长同格式的记录集合。例如，一个学生人事记录文件，每个记录都有如图 1-5 的记录格式。

学生人事记录

学号	姓名	性别	系别	年龄	政治面貌	家庭出身	籍贯	家庭成员	奖惩情况

图 1-5

其中前 8 项数据是任何学生必须具有的而且基本上是等长的，而各个学生的后两项数据其信息量大小变化较大。如果采用等长记录形式存储学生数据，为了建立完整的学生档案文件，每个学生记录的长度必须等于信息量最多的记录的长度，因而会浪费大量的存储空间。所以最好是采用变长记录或主记录与详细记录相结合的形式建立文件。即将学生人事记录的前 8 项作为主记录,后两项作为详细记录,则每个记录有如图 1-6 记录格式,某个学生记录如图 1-7 所示。

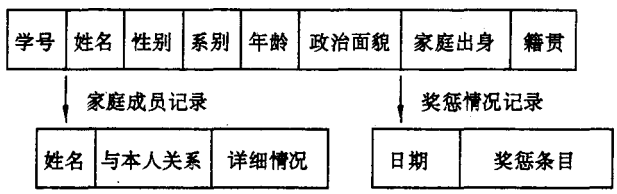


图 1-6

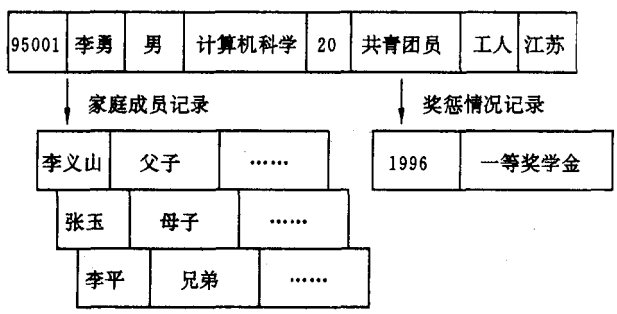


图 1-7

这样可以节省许多存储空间，灵活性也有相对提高。

但这样建立的文件仍有局限性，因为这种灵活性只对一个应用而言。一个学校或一个组

织涉及许多应用，在数据库系统中不仅要考虑某个应用的数据结构，还要考虑整个组织的数据结构。例如，一个学校的管理信息系统中不仅要考虑学生的人事管理，还要考虑学籍管理、选课管理等,可按图 1-8 方式为该校的信息管理系统组织学生数据。

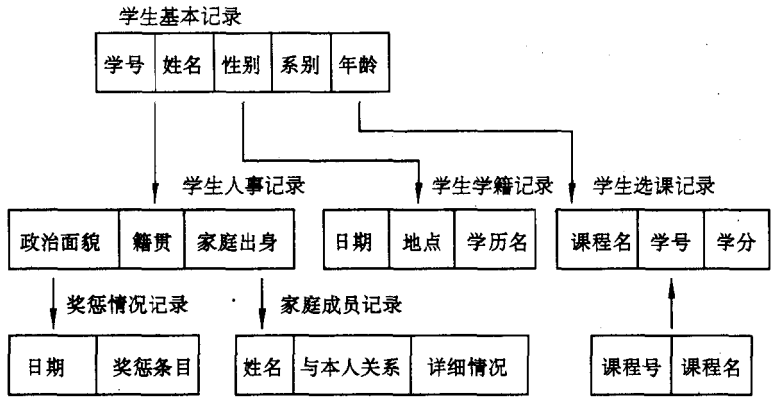


图 1-8

这种数据组织方式为各个管理提供必要的记录，使学校的学生数据结构化了。这就要求在描述数据时不仅要描述数据本身，还要描述数据之间的联系。文件系统尽管其记录内部已有了某些结构，但记录之间没有联系。数据库系统实现整体数据的结构化，这是数据库的主要特征之一，也是数据库系统与文件系统的本质区别。

在数据库系统中，不仅数据是结构化的，而且存取数据的方式也很灵活，可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。而在文件系统中,数据的最小存取单位是记录，粒度不能细到数据项。

(2) 数据的共享性好,冗余度低

数据的共享程度直接关系到数据的冗余度。数据库系统从整体角度看待和描述数据，数据不再面向某个应用而是面向整个系统。上例中的学生基本记录就可以被多个应用共享使用。这样既可以大大减少数据冗余，节约存储空间，又能够避免数据之间的不相容性与不一致性。所谓数据的不一致性是指同一数据不同拷贝的值不一样。采用人工管理或文件系统管理时，由于数据被重复存储，当不同的应用和修改不同的拷贝时就易造成数据的不一致。

(3) 数据独立性高

数据库系统提供了两方面的映象功能，从而使数据既具有物理独立性，又有逻辑独立性。

数据库系统的一个映象功能是数据的总体逻辑结构与某类应用所涉及的局部逻辑结构之间的映象或转换功能。这一映象功能保证了当数据的总体逻辑结构改变时，通过对映象的相应改变可以保持数据的局部逻辑结构不变，由于应用程序是依据数据的局部逻辑结构编写的，所以应用程序不必修改。这就是数据与程序的逻辑独立性，简称数据的逻辑独立性。

数据库系统的另一个映象功能是数据的存储结构与逻辑结构之间的映象或转换功能。这一映象功能保证了当数据的存储结构（或物理结构）改变时，通过对映象的相应改变可以

保持数据的逻辑结构不变，从而应用程序也不必改变。这就是数据与程序的物理独立性，简称数据的物理独立性。

数据与程序之间的独立性，使得可以把数据的定义和描述从应用程序中分离出去。另外,由于数据的存取由 DBMS 管理，用户不必考虑存取路径等细节，从而简化了应用程序的编制，大大减少了应用程序的维护和修改。

(4) 数据由 DBMS 统一管理和控制

由于对数据实行了统一管理，而且所管理的是有结构的数据，因此在使用数据时可以有很灵活的方式，可以取整体数据的各种合理子集用于不同的应用系统，而且当应用需求改变或增加时，只要重新选取不同子集或者加上一小部分数据，便可以有更多的用途，满足新的要求。因此使数据库系统弹性大，易于扩充。

除了管理功能以外，为了适应数据共享的环境,DBMS 还必须提供以下几方面的数据控制功能。

数据的安全性(security)

数据的安全性是指保护数据，防止不合法使用数据造成数据的泄密和破坏，使每个用户只能按规定对某些数据以某些方式进行访问和处理。

数据的完整性(integrity)

数据的完整性指数据的正确性、有效性和相容性。即将数据控制在有效的范围内，或要求数据之间满足一定的关系。

- 并发(concurrency)控制

当多个用户的并发进程同时存取、修改数据库时，可能会发生相互干扰而得到错误的结果，并使得数据库的完整性遭到破坏，因此必须对多用户的并发操作加以控制和协调。

数据库恢复(recovery)

计算机系统的硬件故障、软件故障、操作员的失误以及故意的破坏也会影响数据库中数据的正确性，甚至造成数据库部分或全部数据的丢失。 DBMS 必须具有将数据库从错误状态恢复到某一已知的正确状态（也称为完整状态或一致状态）的功能，这就是数据库的恢复功能。

数据库管理阶段应用程序与数据之间的对应关系可用图 1-9 表示。

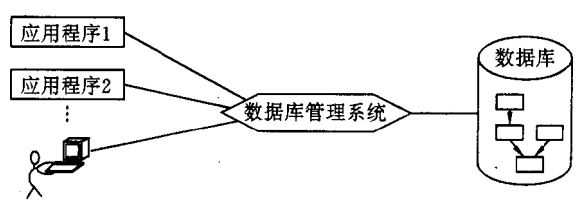


图 1-9 程序与数据的对应关系

综上所述，数据库是长期存储在计算机内有组织的大量的共享的数据集合。它可以供各种用户共享，具有最小冗余度和较高的数据独立性。DBMS 在数据库建立、运用和维护时对数据库进行统一控制，以保证数据的完整性、安全性，并在多用户同时使用数据库时进行并发控制，在发生故障后对系统进行恢复。

数据库系统的出现使信息系统的研制从以加工数据的程序为中心转向围绕共享的数据库来进行。这样既便于数据的集中管理,又有利于应用程序的研制和维护,提高了数据的利用率和相容性,提高了决策的可靠性。

数据库技术从20世纪60年代中期产生到今天仅仅30年的历史，但其发展速度之快，使用范围之广是其他技术所不及的。60年代末出现了第一代数据库——网状数据库、层次数据库，70年代出现了第二代数据库——关系数据库。目前关系数据库系统已逐渐淘汰了网状数据库和层次数据库，成为当今最为流行的商用数据库系统。而80年代出现的以面向对象模型为主要特征的数据库系统又在向关系数据库系统挑战。数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透、互相结合,成为当前数据库技术发展的主要特征。我们将在第8章中比较详细地介绍数据库技术的最新发展。

1.1.3 数据库技术的研究领域

目前虽然已有了一些比较成熟的数据库技术，但随着计算机硬件的发展和应用范围的扩大，数据库技术也需要不断向前发展，概括地讲，当前数据库学科主要的研究范围有以下三个领域。

1. 数据库管理系统软件的研制

数据库管理系统DBMS是数据库系统的基础。DBMS的研制包括研制DBMS本身以及以DBMS为核心的一组相互联系的软件系统。研制的目标是扩大功能、提高性能和提高用户的生产率。

随着数据库应用领域的不断扩大，许多新的应用领域如自动控制、计算机辅助设计等要求数据库能够处理与传统数据类型不同的新的数据类型，例如，声音、图象等非格式化数据，面向对象的数据库系统、扩展的数据库系统、多媒体数据库系统等的兴起就是应这些新的需求和应用背景而产生的。

2. 数据库设计

数据库设计的主要任务是在DBMS的支持下，按照应用的要求，为某一部门或组织设计一个结构合理、使用方便、效率较高的数据库及其应用系统。其中主要的研究方向是数据库设计方法学和设计工具，包括数据库设计方法、设计工具和设计理论的研究，数据模型和数据建模的研究，计算机辅助数据库设计方法及其软件系统的研究，数据库设计规范和标准的研究等。

3. 数据库理论

数据库理论的研究主要集中于关系的规范化理论、关系数据理论等。近年来，随着人工智能与数据库理论的结合以及并行计算机的发展，数据库逻辑演绎和知识推理、并行算法等理论研究，以及演绎数据库系统、知识库系统和数据仓库的研制都已成为新的研究方向。

需要全本请访问 www.ertongbook.com