

21世纪高等院校计算机教材

数据库原理及应用

孙梅 张寿华 孙浩军 编著

按照教学规律阐述基本理论并结合现在流行的数据库系统描述基本方法的应用



理实结合 内容详实

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

21 世纪高等院校计算机教材

数据库原理及应用

第2版 王珊 萨师煊 主编

清华大学出版社



清华大学出版社

清华大学出版社

21 世纪高等院校计算机教材

数据库原理及应用

孙梅 张寿华 孙浩军 编著

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

内 容 简 介

本书是关于数据库的基础教材,取材于数据库几十年发展的重要成果,并结合多年教学和实践经验精心编写而成。

注重理论与实践相结合是本书的特色,按照教学规律阐述基本理论并结合现在流行的数据库系统描述基本方法的应用。在介绍了数据模型、数据库的基本概念后,重点介绍了关系数据库基础知识、SQL 语言和关系数据库的设计模式,数据库的事务处理方法,并以 SQL Server 2000 为例阐述数据库的安全性和完整性,以及数据库的设计方法和数据库管理系统,最后,介绍了网络数据库和一些数据库的新技术。

本书可以作为高等院校计算机及信息专业的教材,也可作为中等专科学校计算机专业的教材,对从事计算机及信息专业领域的科研人员也是一本不错的参考书。

图书在版编目(CIP)数据

数据库原理及应用/孙梅,张寿华,孙浩军编著. —北京:中国铁道出版社,2007.1
21 世纪高等院校计算机教材
ISBN 978-7-113-07423-4

I. 数... II. ①孙... ②张... ③孙... III. 数据库系统—高等学校—教材 IV. TP311.13

中国版本图书馆 CIP 数据核字 (2007) 第 017998 号

书 名: 数据库原理及应用

作 者: 孙 梅 张寿华 孙浩军

出版发行: 中国铁道出版社 (100054, 北京市宣武区右安门西街 8 号)

策划编辑: 严晓舟 秦绪好

责任编辑: 苏 茜 翟玉峰 郑 双

封面制作: 白 雪

责任校对: 李 畅

印 刷: 河北省遵化市胶印厂

开 本: 787×1092 1/16 印张: 16.75 字数: 383 千

版 本: 2007 年 2 月第 1 版 2007 年 2 月第 1 次印刷

印 数: 1~5 000 册

书 号: ISBN 978-7-113-07423-4/TP·2043

定 价: 22.00 元

版权所有 侵权必究

本书封面贴有中国铁道出版社激光防伪标签,无标签者不得销售

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社计算机图书批销部调换。

21 世纪高等院校计算机教材

编写委员会

主 任：王凤先

副主任：马胜甫 刘振鹏

委 员：（按姓氏笔划排列）

于 锋	王学军	王卫华	王春红	白乙拉
刘立民	刘清波	李亚平	李继民	张晓莉
杨怀卿	孟玉芹	赵英杰	赵新生	赵福来
徐建民	韩宪忠			

前言

作为现代信息科学的主要支撑技术,数据库技术自 20 世纪 60 年代中期产生以来,已经发展成为计算机科学及应用学科的一个重要分支,它的出现、发展和广泛应用极大地促进了计算机科学及应用的发展。

在高等院校中,数据库原理及应用课程的重要性也越来越突出,它不仅是计算机、信息管理等专业的主干课程,而且还成为许多非计算机专业学生的必修或选修课程。由于数据库技术有完善的理论基础,同时又是应用性极强的课程,这就要求高校计算机或信息管理类专业的学生,既要数据库理论有较深入的理解,又要掌握必要的实际应用的能力。本书力求较好地理论与实践结合起来,一方面,将数据库系统理论的核心内容进一步精细提炼并较完整保留,在叙述理论内容时配备较丰富的例题;另一方面,以常用的 SQL Server 2000 作为实例,介绍数据库实现及应用技术,将相关内容贯穿至各个章节,既克服了对某一系统介绍所带来的多余篇幅,又保证了学生实际操作、练习与应用的需要。这样将数据库理论的各种知识有机地结合起来,为学生架起一座从理论到实践的桥梁。

本书共分为 12 章。第 1 章数据库概论,简述了数据管理技术的发展、数据模型、数据库的基本构成和数据库管理系统等基本概念。第 2~4 章介绍了关系数据库基础知识、SQL 语言、数据库的安全性及完整性和关系数据库的设计模式。包括关系模式的概念及关系数据库的结构、关系运算和关系演算的基本知识;SQL 语言的发展、查询和安全控制的特点以及其他各种操作;介绍了数据库完整性和安全性的概念和控制的基本技术,并结合 SQL Server 2000 对这些技术加以讨论。第 5、6 章关系数据库模式设计及数据库设计,论述关系模式的设计问题、函数依赖理论及函数依赖的公理系统、关系模式的分解概念及关系模式的规范化理论和方法;介绍了数据库设计中各个阶段的划分和每个阶段的主要工作,讨论各个阶段的意义、原则和方法。第 7、8 章查询优化和事务处理,论述了查询优化的一般概念和常用技术;介绍事务处理的概念、数据库恢复的原理和实现技术、锁及封锁的概念及其处理技术、并发操作、调度和两段锁协议等内容。第 9 章数据库管理系统,进一步阐述 DBMS 的基本功能、系统结构以及进程组织方案,介绍语言翻译处理层、数据存取层以及缓冲区管理层的任务和功能。第 10、11 章数据库访问技术和网络环境下的数据库系统,介绍客户机/服务器系统的工作模式和技术特征、ODBC 的结构,分布式数据库体系结构、查询处理及事务处理技术。第 12 章介绍数据库的一些新技术与新应用。

本书第 1、2、4、11、12 章由孙梅编写,第 5~7 章由孙浩军编写,第 3、8、9、10 章由张寿华编写。由于作者水平有限,在本书中一定有不少问题,欢迎读者批评指正。

在本书的编写过程中,刘振鹏、袁方、陈俊芬、朱亮同志提出了许多有益的意见,硕士研究生刘志辉、孔令俊在习题的编写过程中做了很多工作,在此致以衷心的感谢。

编 者

2006 年 11 月

目 录

第 1 章 数据库概论	1
1-1 引言	1
1-1-1 数据与信息	1
1-1-2 数据处理与数据管理	1
1-1-3 数据库的基本概念	2
1-2 数据管理技术的发展	3
1-2-1 人工管理阶段	3
1-2-2 文件系统阶段	4
1-2-3 数据库系统阶段	5
1-3 数据模型	5
1-3-1 数据模型的组成要素	6
1-3-2 概念数据模型	7
1-3-3 层次模型	9
1-3-4 网状模型	9
1-3-5 关系模型	10
1-3-6 面向对象数据模型	11
1-4 数据库的体系结构	11
1-4-1 三级模式结构	11
1-4-2 数据独立性	12
1-5 数据库管理系统	13
1-5-1 数据库管理系统 (DBMS) 的主要功能	13
1-5-2 DBMS 的程序组成	14
1-5-3 用户访问数据的过程	15
1-6 数据库系统	16
1-6-1 DBS 的组成	16
1-6-2 DBS 的特点	17
习题	18
第 2 章 关系数据库基础知识	20
2-1 关系模型	20
2-1-1 关系模型的基本术语	20
2-1-2 关系模式和关系数据库	22
2-2 关系的完整性	23
2-2-1 关系模型的三要素	23
2-2-2 实体完整性	24

2-2-3 参照完整性	25
2-2-4 用户定义的完整性	25
2-3 关系代数	25
2-3-1 传统的集合运算	26
2-3-2 专门的关系运算	27
2-4* 关系演算	32
2-4-1 元组关系演算	32
2-4-2 域关系演算	33
习题	33
第3章 关系数据库 SQL 语言	35
3-1 SQL 语言概况	35
3-1-1 SQL 发展	35
3-1-2 SQL 语言的功能	36
3-1-3 SQL 语言特点	36
3-1-4 SQL 数据库的体系结构	37
3-2 SQL 的数据定义	38
3-2-1 定义基本表	38
3-2-2 基本表的修改与删除	41
3-2-3 索引的建立和删除	42
3-3 SQL 的数据查询	43
3-3-1 单表查询	45
3-3-2 连接查询	51
3-3-3 子查询	53
3-3-4 集合查询	58
3-4 SQL 的数据更新	59
3-4-1 SQL 数据插入	59
3-4-2 SQL 数据修改	60
3-4-3 SQL 数据删除	61
3-5 视图	61
3-5-1 视图的定义和删除	62
3-5-2 视图的查询	63
3-5-3 视图的更新	63
3-5-4 视图的优点	64
3-6 SQL 的数据控制功能	64
3-6-1 权限与角色	64
3-6-2 权限与角色授予和收回	65
3-7 嵌入式 SQL	67
3-7-1 嵌入式 SQL 概述	67
3-7-2 游标的使用	68

3-7-3 动态 SQL	69
3-8 存储过程	71
3-8-1 存储过程概述	71
3-8-2 存储过程的优点	72
3-8-3 存储过程的分类	73
3-8-4 存储过程的创建	74
3-8-5 删除存储过程	77
3-8-6 修改存储过程	77
习题	78
第 4 章 数据库的完整性与安全性	80
4-1 数据库的完整性	80
4-2 SQL Server 2000 的完整性	83
4-2-1 SQL Server 实现数据完整性的方法	83
4-2-2 SQL Server 2000 支持的约束	84
4-2-3 SQL Server 中的触发器	85
4-2-4 创建约束的其他选项	86
4-3 数据库的安全性	88
4-3-1 安全性分级	88
4-3-2 数据库安全控制的一般方法	89
4-4 SQL Server 的安全体系结构和安全认证模式	90
4-4-1 SQL Server 的安全体系结构	90
4-4-2 SQL Server 2000 的安全认证模式	91
4-5 SQL Server 2000 数据库安全性管理	92
4-5-1 数据库系统登录管理	92
4-5-2 数据库用户管理	92
4-5-3 数据库系统角色管理	93
4-5-4 SQL Server 权限管理	94
习题	96
第 5 章 关系数据库的模式设计	97
5-1 问题的提出	97
5-2 函数依赖	99
5-2-1 函数依赖的定义	99
5-2-2 码	100
5-2-3 函数依赖的推理规则	101
5-2-4 函数依赖集的等价与覆盖	102
5-3 模式分解的特性	103
5-3-1 模式分解的三个定义	103
5-3-2 分解的无损连接性和保持函数依赖性	105

5-4 关系模式的规范化	105
5-4-1 第一范式	105
5-4-2 第二范式	106
5-4-3 第三范式	107
5-4-4 BC 范式	107
5-4-5 多值依赖	108
5-4-6 第四范式	109
5-4-7 模式分解的算法	110
习题	112
第 6 章 数据库设计	114
6-1 数据库设计概述	114
6-1-1 数据库和信息系统	114
6-1-2 数据库设计的特征	115
6-1-3 数据库设计方法	115
6-1-4 数据库设计的步骤	117
6-2 需求分析	118
6-2-1 需求分析的任务	118
6-2-2 需求分析方法	119
6-2-3 数据字典 (DD)	120
6-3 概念结构设计	122
6-3-1 概念模型	122
6-3-2 数据抽象与局部 E-R 模型设计	123
6-3-3 视图的集成与优化	125
6-4 逻辑结构设计	126
6-4-1 E-R 图向关系模型的转换	127
6-4-2 数据模型的优化	128
6-4-3 设计用户外模式	129
6-5 物理结构设计	130
6-5-1 数据库的存取方法	130
6-5-2 确定数据库的存储结构	132
6-5-3* 物理结构性能评价	133
6-6 数据库的实现与维护	133
6-6-1 数据库系统的实现	133
6-6-2 数据库系统的试运行	134
6-6-3 数据库的运行和维护	134
习题	135
第 7 章 查询优化	137
7-1 查询优化概述	137

7-2	查询优化的一般策略.....	139
7-3	基于关系代数表达式的优化算法.....	139
7-3-1	关系代数表达式的等价变换规则.....	139
7-3-2	关系代数表达式的优化步骤.....	140
7-4	分解查询的优化方法.....	144
7-4-1	分解处理.....	144
7-4-2	结局处理.....	147
7-5	连接运算的优化.....	148
7-6	代价估算优化.....	150
7-6-1	查询执行代价的组成和代价模型.....	151
7-6-2	选择操作的代价估算.....	152
7-6-3	连接操作的代价估算.....	153
	习题.....	155
第 8 章	事务处理.....	157
8-1	事务.....	157
8-1-1	事务的概念.....	157
8-1-2	事务的性质.....	158
8-1-3	事务的状态.....	159
8-2	恢复技术.....	160
8-2-1	故障种类.....	160
8-2-2	数据库恢复的原理.....	161
8-3	数据库恢复的技术.....	163
8-3-1	事务故障的恢复.....	163
8-3-2	系统故障的恢复.....	163
8-3-3	介质故障的恢复.....	164
8-3-4	具有检查点的恢复技术.....	164
8-3-5	数据库镜像技术.....	166
8-4	SQL Server 2000 的恢复技术.....	167
8-4-1	SQL Server 的备份形式和操作方式.....	167
8-4-2	SQL Server 的数据备份和恢复策略.....	167
8-5	并发控制引论.....	168
8-5-1	并发的目的.....	168
8-5-2	并发操作带来的问题.....	169
8-6	封锁.....	170
8-6-1	封锁协议.....	170
8-6-2	活锁和死锁.....	171
8-7	并发调度的可串行性.....	174
8-8	协议.....	174
8-8-1	两段锁协议.....	174

8-8-2 基于时间戳的协议	175
8-8-3 乐观协议	177
8-9 封锁的粒度	178
8-9-1 多粒度封锁	178
8-9-2 意向锁	178
8-10 SQL Server 2000 的并发控制技术	179
8-10-1 SQL Server 的事务类型	179
8-10-2 SQL Server 的空间管理及锁的级别	179
8-10-3 SQL Server 锁的类型	180
习题	181
第 9 章 数据库管理系统	182
9-1 DBMS 概述	182
9-1-1 DBMS 的基本功能	182
9-1-2 DBMS 和操作系统	183
9-1-3 DBMS 的进程结构	185
9-1-4 DBMS 的系统结构	187
9-2 语言处理层	188
9-2-1 语言处理层的任务	188
9-2-2 解释方法	190
9-2-3 预编译方法	190
9-3 数据存取层	191
9-3-1 数据存取层的系统结构	192
9-3-2 存取层的功能子系统	193
9-3-3 缓冲区管理	195
习题	195
第 10 章 数据库访问技术	196
10-1 ODBC	196
10-1-1 ODBC 概述	196
10-1-2 配置 ODBC 数据源	200
10-1-3 ODBC API 使用基础	203
10-2 JDBC	206
10-2-1 JDBC 介绍	206
10-2-2 JDBC 的结构	208
10-2-3 JDBC 接口概述	210
10-3 ADO 简介	211
10-3-1 OLE DB 概述	211
10-3-2 ADO 概述	212
习题	218

第 11 章 网络环境下的数据库系统	219
11-1 客户机/服务器系统	219
11-1-1 客户机/服务器数据库系统	220
11-1-2 客户机/服务器数据库系统的实现	220
11-1-3 客户机/服务器数据库系统的特点	222
11-2 分布式数据库	223
11-2-1 分布式数据库的概念	223
11-2-2 分布式数据库系统的特点	224
11-2-3 分布式数据库系统的体系结构	225
11-2-4 分布式查询处理	227
11-2-5 分布式事务处理	228
11-2-6 Web 与数据库的连接	231
习题	234
第 12 章 数据库新技术与新应用	235
12-1 面向对象数据库	235
12-1-1 面向对象数据库的主要概念	235
12-1-2 对象类之间的联系	237
12-1-3 面向对象数据库系统 (OODBS)	238
12-2 多媒体数据库	239
12-2-1 多媒体数据管理的基本要求和功能	239
12-2-2 多媒体数据库的结构	240
12-2-3 多媒体数据库的实现	242
12-3 主动数据库	243
12-3-1 主动数据库的产生及描述	243
12-3-2 主动数据库管理系统的实现	244
12-3-3 主动数据库的研究进展	246
12-3-4 主动数据库的应用	246
12-4 数据仓库与数据挖掘	246
12-4-1 数据仓库的概念及特点	247
12-4-2 数据仓库的工具	248
12-4-3 数据挖掘	249
习题	251
参考文献	252

第 1 章 数据库概论

本章目标

- 数据处理与数据管理的基本概念
- 了解数据管理技术的发展
- 数据模型的组成要素及几个主要的数据模型
- 数据库的体系结构
- 数据库管理系统及主要功能和组成

20 世纪中期以来,现代数据库技术从无到有,得到迅速发展,今天,数据库系统已遍布政府机关、社会团体和各个行业,存储着它们赖以正常运转的数据资源,成为数据处理的公用支撑技术。数据库的广泛应用显著地提高了工作效率和质量,产生出极大的社会、经济效益。数据库技术的发展一方面受到应用需求的牵引和驱动,另一方面又促进和拓宽了计算机的应用领域。同时也改变了人们运用信息的方式,有力地推动了社会信息化的进程。本章简要介绍数据库技术的发展历史及相关的基本概念。

1-1 引 言

数据库技术的主要目的是有效地管理和存取大量的数据资源。在计算机应用中,数据处理和以数据处理为基础的信息系统所占的比重最大。人类的一切活动都离不开数据,离不开信息。

1-1-1 数据与信息

数据 (Data): 是指存储在某一媒体上可加以鉴别的符号资料,如银行的账目数据、学校的管理数据、企业的生产管理和产品销售数据以及电影图片资料等。数据的概念包括两个方面,其一,数据内容是事物特性的反映或描述;其二,数据是存储在某一媒体上符号的集合。这里的符号,不仅指数字、字母、文字和其他特殊字符,而且还包括图形、图像、动画、影像、声音等多媒体数据。存储不仅是在纸上,还包括记录在磁介质、光介质上、半导体存储器里。

信息 (Information): 是关于现实世界事物的存在方式或运动形态的综合反映,是人们进行各种活动所需要的知识,它反映了客观事物的物理状态。信息是可存储、加工、传递和再生的。人们用大脑存储信息,叫做记忆。计算机存储器、录音、录像等技术的发展,进一步扩大了信息存储的范围。借助计算机,还可以对收集到的信息进行整理。

数据是载荷信息的物理符号或称为载体。信息是人们消化理解了的数据,是对客观世界的认识,即知识。如某一企业的计算机年产量是 10 万台,这个 10 万台是一个数据,它表示了此企业生产计算机的能力这一信息。从这一例子可以看出,信息是由数据来表示的,二者是不可分离又有一定区别的两个概念。可以说,信息是被消化了的数据,它依赖于数据而存在;反过来,数据是信息的具体表现形式,它使信息更容易理解和应用。

1-1-2 数据处理与数据管理

数据处理 (Data Processing): 是指将数据转换成信息的过程。广义地讲,处理包括对数

据的收集、存储、加工、分类、检索、传播等一系列活动。狭义地讲，处理是指对所输入的数据进行加工处理。可用以下公式简单表示：

$$\text{信息} = \text{数据} + \text{处理}$$

数据是原料，是输入，而信息是产出，是输出结果。“信息处理”的确切含义应该是为了得到信息而对数据进行的处理。

数据管理 (Data Management)：是指对数据进行组织、分类、编码、存储、检索和维护等。

现代数据管理和数据处理都是现代计算机系统的基本支撑技术和重要分支。可以预测，随着社会信息化水平的提高，人们对数据管理的要求也将越来越高，而数据库技术是现阶段数据管理的主要形式。数据库技术的发展将对今后社会的发展发挥重要作用。

1-1-3 数据库的基本概念

数据库 (DataBase)：是指长期存储在计算机存储介质上，按照一定格式存放的可共享的数据集合。数据库中的数据按一定数据模型组织、描述和存储，具有较小的冗余度、较高的数据独立性和可扩充性，可为多个不同用户所共享。

科学地组织和存储数据，高效地提取和维护数据，是数据管理的主要任务。数据库管理系统就是为完成这一任务而设计的一套计算机系统软件。

数据库管理系统 (DataBase Management System, DBMS)：是指用于数据库建立、使用和维护的管理软件。它建立在操作系统的基础上，按照用户的要求对数据库进行统一的管理、维护和控制，它是用户和计算机应用系统之间的桥梁。

数据库系统 (DataBase System, DBS)：是指计算机系统引入数据库后的系统。一般来说，它是由数据库、数据库管理系统、应用系统、数据库管理员和用户组成，如图 1-1 所示。在不引起混淆的情况下通常把数据库系统简称为数据库。

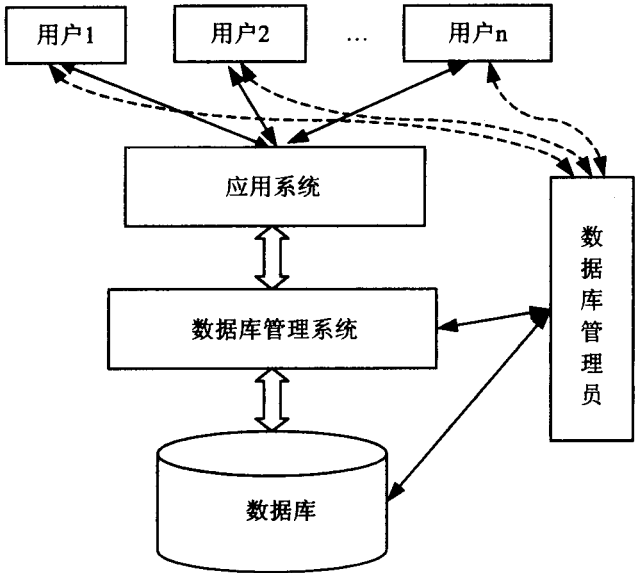


图 1-1 数据库系统

1-2 数据管理技术的发展

随着现代科学技术的飞速发展，人们的视野迅速扩大，社会信息量和数据量急剧增加，数据管理也经历了从人工管理、文件系统管理到数据库系统管理的发展过程。

1-2-1 人工管理阶段

第一阶段是数据的手工处理阶段。这是自远古时代到 20 世纪 50 年代一直使用的数据处理方法，由于当时的社会生产力低，科学技术不够发达，因此，数据处理的方法保持着简单、低级的手工操作。人们只能借助于如算盘、计算尺和低等计算机处理数据。这种方法的效率低，可处理的数据量小且可靠性差。计算机主要用于科学计算，硬件中的外存只有卡片、纸带和磁带，没有磁盘等直接存取设备。软件只有汇编语言，没有操作系统和管理数据的软件，数据处理的方式基本上是批处理。此阶段的主要特点有：

(1) 数据不保存

因为当时计算机主要用于科学计算，对于数据保存的需求尚不迫切。在计算某一课题时将数据输入主存，运算处理后将结果输出，计算机在处理数据过程中不保存程序和数据。

(2) 数据依靠应用程序管理

每个应用程序都要规定数据的逻辑存储结构，设计物理结构、存取方法和输入方式等，程序员编写程序的工作量很大。

(3) 数据不共享

数据是面向应用的，一组数据只能用于一个应用程序。当多个应用程序使用相同数据时，也需要各自定义它们的结构，造成程序之间大量的冗余数据。

(4) 数据不具有独立性

由于这一阶段程序是依赖于数据的，当数据的类型、格式等逻辑结构或物理结构发生变化时，必须对应用程序做出相应的修改，从而加大了程序的应用成本。

人工管理阶段应用程序与数据之间的对应关系如图 1-2 所示。

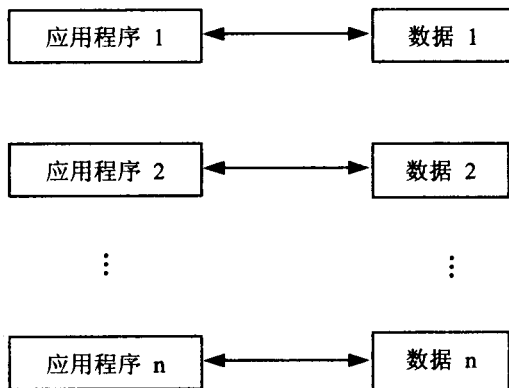


图 1-2 人工管理阶段应用程序与数据之间的对应关系

1-2-2 文件系统阶段

第二阶段是文件系统阶段，在 20 世纪 50 年代后期至 20 世纪 60 年代中期。随着计算机技术的发展，这一阶段计算机不仅用于科学计算，还大量用于信息管理。此时在计算机硬件方面，磁盘、磁鼓等直接存取设备的出现为高效数据管理提供了基础；在软件方面出现了高级语言和具有数据管理功能的操作系统；文件批处理功能和联机实时处理增强了数据处理能力。

与人工管理相比文件系统管理数据具有以下优点：

(1) 数据以文件形式长期保存

数据以文件形式，长期保存在计算机的外存储设备上，应用程序可以随时对数据进行查询、修改及增删等处理。

(2) 文件系统可对数据的存取进行管理

程序员只与存储在外存储设备上的文件的文件名和记录打交道。不必关心数据的物理存储，大大减轻了程序员的负担。文件系统有了一定的结构，如顺序文件、倒排文件和索引文件等，从而实现了不仅可对文件记录的顺序访问，也可以随机访问；应用程序与数据间有了一定的独立性；数据由专门的软件即文件系统进行管理，由软件提供的存取方法进行转换。数据发生变化不一定影响程序的运行，从而大大节省了程序的运行成本。

但是，文件系统仍存在以下缺陷：

(1) 数据共享性较差

数据文件之间没有建立有机的联系，一个数据文件基本上对应一个应用程序，当不同的应用程序需要相同的数据时，也必须建立各自的数据文件系统，数据不能共享，造成数据的不一致性，给数据维护带来不必要的困难。

(2) 数据独立性差

由于文件是为某一特定应用服务的，因而应用程序和数据有相互依赖的关系，若数据的逻辑结构发生变化，必须修改相应的应用程序；反之，若应用程序发生变化，如应用程序需要改用不同的高级语言重写，也将导致数据文件系统的结构改变。

文件系统阶段应用程序与数据之间的对应关系如图 1-3 所示。

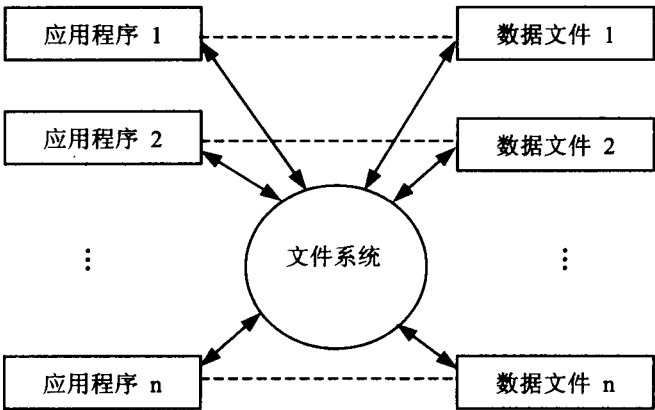


图1-3 文件系统阶段应用程序与数据之间的对应关系