

21 世纪高等院校教材

数据库原理 及其应用教程

黄德才 主编



3



科学出版社

www.sciencep.com

21 世纪高等院校教材

数据库原理及其应用教程

黄德才 主编

科学出版社

北京

内 容 简 介

本书详细介绍了数据库原理、方法及其应用开发技术。全书共分8章,分别介绍了数据模型,数据库系统的模式结构,关系模型与关系代数运算,关系数据库标准语言——SQL,关系模式的规范化设计理论,数据库的安全与保护,RAID技术,数据库设计的步骤和数据库实施,数据库应用系统的体系结构,Microsoft SQL Server 2000的安全管理、完整性策略、恢复技术、并发控制方法、触发器和存储过程,基于Delphi 6.0/Server 2000的C/S结构的简单数据库应用系统等内容。书中还介绍了数据库技术的新发展,如面向对象数据库技术、分布式数据库、数据仓库与数据挖掘技术等。书中配有较多的例子和适量的习题,随书光盘包括书中标准SQL例题和实验数据库应用系统,不仅有利教师进行多媒体教学,还便于学生自学。

本书既可作为普通高校有关专业“数据库原理及其应用”课程的教材,也可作为成人教育和自学考试同名课程的教材和参考书,还可供IT领域的科技人员参考。

图书在版编目(CIP)数据

数据库原理及其应用教程/黄德才主编. —北京:科学出版社,2002

(21世纪高等院校教材)

ISBN 7-03-010692-X

I. 数… II. 黄… 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(2002)第054336号

责任编辑:马长芳 段博原 王利

责任印制:刘秀平/封面设计:曹烨

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2002年8月第一版 开本: B5 (720×1000)

2003年3月第二次印刷 印张: 20 3/4

印数: 3 001—6 000 字数: 420 000

定价: 32.00元(含光盘)

(如有印装质量问题, 我社负责调换〈北燕〉)

前 言

数据库是计算机科学中十分重要而且发展迅速的重要分支, DBMS(数据库管理系统)同操作系统一样,已成为各种应用系统开发的基础平台。数据库原理及其应用课程的重要性也变得越来越突出,它不仅是普通高校计算机、信息管理等专业的主干课程,而且已成为许多非计算机专业学生的必修或选修课程;它还是政府上网工程中人员培训的必修科目。

对于绝大多数学生来说,学习数据库课程的主要目的不是去研究和开发商品化的 DBMS,而是为了应用现有 DBMS 和数据库应用系统开发工具,解决实际工作中的各类计算机应用问题。因此,高校计算机或信息管理类专业的大学生,必须兼顾数据库理论与实际应用知识的学习,不仅要掌握数据库的基本原理,还必须懂得如何将具体的 DBMS 与某种客户端应用开发工具结合,完成基于“C/S”或“B/S”结构的数据库应用系统的开发过程,才能真正理解和掌握数据库原理及应用的完整概念和知识,为开发数据库应用系统奠定坚实的基础。

然而,国内目前出版的各种计算机专业和信息管理专业的数据库教材,在内容安排和组织上都存在一些不尽人意的地方。这些教材基本上可以分为两类,一种是“数据库系统原理”类的教材,其特点是注重理论,主要内容放在数据库系统原理和数据库设计原理上,内容抽象且脱离实际,缺少数据库应用系统开发方面的知识,没能为学生架起从理论到实践的桥梁,使得学生在学习数据库理论知识后不知道如何应用。此外,这类教材中的不少书籍在介绍标准 SQL 语言时给出的例子,有些根本无法在当今最流行的 SQL Server 系统中运行(全国多数高校的计算机实验环境和学生自己的计算机环境都使用 SQL Server),使学习内容与实践环境严重脱节,给学生学习增加了困难。另一种是“数据库原理及其应用”类的教材,虽然增加了一些应用知识,但仅停留在数据库应用系统开发环境和一般开发技术的简单介绍层面上,这样反而给学生理解理论与进行实践设置了障碍。

本书是在作者多年从事数据库教学和进行数据库应用系统开发实践的基础上编写的一本体现当前数据库理论发展和应用技术水平,符合 21 世纪 IT 发展需要的新教材。本书内容丰富,层次分明,概念清楚,逻辑性强,理论叙述深入浅出,实际应用完整具体,本书不仅配有较多的例子和适量的习题,而且还有包括书中标准 SQL 例题和实验数据库应用系统的配套光盘。这不仅有利于教师进行多媒体教学,也将为学生自学提供极大的帮助。本书不仅介绍了数据库系统基本原理和标准 SQL(全部例子都在 SQL Server 2000 上运行通过),且首次介绍了将 Delphi 与 SQL Server 2000 结合开发的一个基于 C/S 结构的简单而完整的数据库应用系统,

学生无需系统开发经验即可在3~4小时内建成并运行该系统;这真正为学生架起了理论与实践的桥梁。实践表明,这不仅能帮助学生真正理解客户端与服务端的概念及其相互关系,增强学生的实践能力,还有利于提高教学质量和教学效果,培养学生的系统开发能力以及创新能力。

全书共分8章。第1章为绪论,主要内容有数据库的常用术语、计算机数据管理技术的产生和发展、数据模型、数据库系统的模式结构、DBMS的功能等;第2章为关系数据库,主要介绍关系模型的基本概念、关系代数运算、关系演算及三种关系查询语言——ISBL、ALPHA、QBE等;第3章为关系数据库标准语言——SQL,主要有SQL的特点、数据定义、数据查询、数据更新、视图定义、数据控制、嵌入式SQL以及查询优化等有关知识;第4章为关系模式的规范化设计理论,内容有关系模式可能存在的异常、关系模式的函数依赖、关系模式的范式(1NF~5NF)、关系模式的分解特性等;第5章为数据库的安全与保护,包括安全性、完整性、并发控制、事务处理与故障恢复技术、RAID技术介绍等;第6章为数据库设计,介绍数据库设计概念、数据库设计的步骤(数据库规划、需求分析、概念结构设计、逻辑结构设计、物理结构设计)和数据库实施等内容;第7章为数据库应用系统开发,内容包括数据库应用系统的体系结构,Microsoft SQL Server 2000的主要特点、安装、基本工具、安全管理、完整性策略、恢复技术和并发控制及其触发器和存储过程,一个基于Delphi 6.0/Server 2000的C/S结构的简单数据库应用系统;第8章为数据库技术新发展,其内容有数据库家族概述、面向对象数据库技术、分布式数据库、数据仓库与数据挖掘等。讲授完本书全部内容大约需要64学时,对于专科生或教学学时数较少的本科生可以不讲授标有“*”的内容。书中标有“**”的内容主要安排给学生课后自学,如果时间充裕,也可选择其中部分内容用于课堂讲授。

本书可作为普通高校计算机专业、信息管理与信息系统专业和数学专业“数据库原理及其应用”专业课程的教材,也可作为成人教育和自学考试同名课程的教材和参考书,还可供IT领域科技人员参考。

本书主要由黄德才、许芸、王文娟编写,黄德才任主编并草拟提纲、规划各章节基本内容。第1、4、7、8章由黄德才负责编写,第2、3章由许芸编写,第5、6章由王文娟编写。黄德才完成全书的修改和统稿工作,并对第2、3、5、6章进行了补充。本书的编写得到了同事们和科学出版社的大力支持,我的研究生郭海东、沈良忠还参加了第7章部分内容的编写工作,在此一并表示衷心的感谢!

限于编者水平,加之数据库技术的发展非常迅速,疏漏和不当之处在所难免。真切希望广大师生和读者来信批评指正,以期在修订时予以完善。

作者的电子邮件地址是:huangdc@mail.hz.zj.cn。

黄德才

2002年7月于杭州

目 录

前言

第 1 章 绪论	1
1.1 数据库的常用术语	1
1.1.1 数据与数据处理	1
1.1.2 信息与数据的关系	2
1.1.3 数据独立性	3
1.1.4 数据库系统	3
1.2 计算机数据管理技术的产生和发展	4
1.2.1 人工管理阶段	4
1.2.2 文件系统阶段	5
1.2.3 数据库系统阶段	7
1.2.4 数据库技术的发展	10
1.2.5 数据库技术的主要研究领域	11
1.3 数据模型	12
1.3.1 数据模型的构成	12
1.3.2 数据模型的分类	13
1.3.3 实体-联系 (E-R) 模型	15
1.3.4 常用的结构数据模型	22
1.4 数据库系统的模式结构	28
1.4.1 数据库系统模式的概念	29
1.4.2 数据库的三级模式	29
1.4.3 数据库的二级映像功能与数据独立性	31
1.5 DBMS 的功能	33
1.6 数据库系统的组成	34
习题 1	36
第 2 章 关系数据库	37
2.1 关系模型的基本概念	37
2.1.1 关系模型概述	37
2.1.2 关系数据结构	38
2.1.3 完整性规则	41
2.2 关系代数	42

2.2.1	传统的集合运算	43
2.2.2	专门的关系运算	44
2.2.3	关系运算举例	48
2.3	关系演算	49
2.3.1	元组关系演算	49
2.3.2	域关系演算	51
2.3.3	关系运算的安全限制	52
2.4	三种关系查询语言*	53
2.4.1	基于关系代数的语言——ISBL	54
2.4.2	元组关系演算语言——ALPHA	56
2.4.3	域关系演算语言——QBE	58
	习题 2	62
第 3 章	关系数据库标准语言——SQL	64
3.1	SQL 概述	64
3.1.1	SQL 的发展	64
3.1.2	SQL 的特点	64
3.2	SQL 的数据定义	67
3.2.1	基本表的定义、删除与修改	68
3.2.2	建立与删除索引	72
3.3	SQL 的数据查询	73
3.3.1	简单的选择与投影查询	73
3.3.2	连接查询	80
3.3.3	嵌套查询	83
3.3.4	集合查询	89
3.4	SQL 的数据更新	89
3.4.1	插入数据	90
3.4.2	修改数据	90
3.4.3	删除数据	91
3.5	SQL 的视图	92
3.5.1	定义视图	92
3.5.2	查询视图	94
3.5.3	更新视图	94
3.6	SQL 的数据控制	95
3.6.1	授权	96
3.6.2	收回权限	98
3.7	嵌入式 SQL	98

3.7.1	嵌入式 SQL 的介绍	99
3.7.2	嵌入式 SQL 语句与主语言之间的通信	100
3.7.3	PowerBuilder 中的嵌入式 SQL 语句	101
3.7.4	在 Delphi 中使用 SQL 语句	105
3.8	查询优化	107
3.8.1	查询优化概述	107
3.8.2	查询实例分析	109
3.8.3	查询优化的一般策略	111
3.8.4	关系代数的等价公式	111
3.8.5	查询优化的一般步骤	113
	习题 3	114
第 4 章	关系模式的规范化设计理论	116
4.1	问题的提出	116
4.1.1	关系模式可能存在的异常	116
4.1.2	异常原因分析	117
4.1.3	异常问题的解决	118
4.2	关系模式的函数依赖	119
4.2.1	再论关系与关系模式	120
4.2.2	函数依赖的一般概念	120
4.2.3	候选键与主键	122
4.2.4	函数依赖的推理规则	123
4.3	关系模式的规范化	132
4.3.1	范式及其类型	133
4.3.2	第一范式 (1NF)	133
4.3.3	第二范式 (2NF)	134
4.3.4	第三范式 (3NF)	136
4.3.5	BC 范式 (BCNF)	137
4.3.6	多值依赖	139
4.3.7	第四范式 (4NF)	142
4.3.8	关系模式规范化步骤	143
4.4	关系模式的分解特性	145
4.4.1	模式分解中存在的问题	145
4.4.2	无损连接	146
4.4.3	无损连接的测试	147
4.4.4	保持函数依赖的分解	151
4.4.5	分解成 3NF 的模式集	153

4.4.6 关系模式设计原则	154
习题 4	154
第 5 章 数据库的安全与保护	157
5.1 安全与保护概述	157
5.2 数据库的安全性保护	157
5.2.1 用户鉴别	158
5.2.2 存取权限控制	158
5.2.3 视图机制	159
5.2.4 跟踪审查	160
5.2.5 数据加密存储	160
5.3 数据库的完整性保护	161
5.3.1 完整性约束的分类	161
5.3.2 完整性控制	163
5.3.3 触发器	166
5.4 数据库的并发控制技术	167
5.4.1 事务及特性	167
5.4.2 数据库的并发控制	168
5.4.3 并发的目的	169
5.4.4 并发所引起的问题	169
5.4.5 并发控制方法	170
5.4.6 并发调度的可串行性	173
5.4.7 时标技术与两段锁协议	174
5.5 数据库的恢复技术	177
5.5.1 故障的种类	177
5.5.2 恢复技术	179
5.6 数据库的复制与相关技术	183
5.6.1 数据库的复制	183
5.6.2 数据库的镜像	184
5.6.3 RAID 技术简介	184
习题 5	186
第 6 章 数据库设计与实施	187
6.1 数据库设计概述	187
6.1.1 数据库设计概念	187
6.1.2 数据库设计步骤	188
6.1.3 数据库设计的特点	190
6.2 数据库规划	190

6.3	需求分析	191
6.3.1	需求分析的任务	191
6.3.2	需求分析的步骤	192
6.3.3	数据字典	195
6.4	概念结构设计	197
6.4.1	基本方法	198
6.4.2	主要设计步骤	199
6.5	逻辑结构设计	203
6.5.1	E-R 模式到关系模式的转换	203
6.5.2	关系模式的优化	206
6.6	物理结构设计	208
6.6.1	聚簇设计	208
6.6.2	索引设计	209
6.6.3	分区设计	209
6.7	数据库的实施和维护	210
6.7.1	数据库的建立与调整	210
6.7.2	数据库系统的试运行	211
6.7.3	数据库系统的运行和维护	211
	习题 6	212
第 7 章	数据库应用系统开发	214
7.1	数据库应用系统的结构	215
7.1.1	集中式结构	215
7.1.2	二层客户机/服务器结构	216
7.1.3	三层客户机/服务器结构	217
7.1.4	多层结构	219
7.2	Microsoft SQL Server 2000	220
7.2.1	SQL Server 的主要特点	220
7.2.2	SQL Server 的安装**	222
7.2.3	SQL Server 的基本工具	227
7.2.4	SQL Server 的安全性管理	233
7.2.5	SQL Server 的完整性策略	234
7.2.6	SQL Server 的恢复技术	236
7.2.7	SQL Server 的并发控制	238
7.2.8	数据库编程	241
7.3	一个简单的数据库应用系统**	245
7.3.1	图形化客户端开发工具——Delphi	245

7.3.2 一个数据库应用系统的实例	249
习题 7	274
第 8 章 数据库技术新发展	275
8.1 数据库家族概述	275
8.2 面向对象数据库技术	276
8.2.1 面向对象方法学简介	276
8.2.2 面向对象数据模型	278
8.2.3 面向对象数据库管理系统	282
8.3 分布式数据库	286
8.3.1 分布式数据库系统的产生	286
8.3.2 分布式数据库系统的定义	288
8.3.3 分布式数据库系统的主要特点	289
8.3.4 分布式数据库系统的模式结构	291
8.3.5 分布式数据库管理系统及其分类	293
8.3.6 分布式数据库系统的优缺点	295
8.4 数据仓库与数据挖掘	296
8.4.1 数据的事务处理与分析处理	296
8.4.2 什么是数据仓库	298
8.4.3 数据仓库数据的四个基本特征	299
8.4.4 数据仓库系统的组成	301
8.4.5 数据仓库中数据的组织	302
8.4.6 数据仓库工具	303
8.4.7 数据挖掘工具	307
习题 8	319
参考文献	321

第 1 章 绪 论

数据库技术是计算机数据管理的最新技术，是计算机科学的重要分支。当计算机的主要应用领域从科学计算转变到数据及事务处理时，数据库技术便应运而生并成为计算机科学中的重要领域。今天，数据库技术不仅在企业管理信息系统 (MIS)、计算机集成制造系统 (CIMS)、办公信息系统 (OIS)、地理信息系统 (GIS)、Internet 技术等方面得到广泛应用，而且越来越多的新应用领域都采用数据库来存储和处理它们的信息资源。对于一个国家来说，数据库的建设规模、数据库信息量的大小和使用频度、数据库的安全性和可靠性已成为衡量一个国家信息化程度的重要标志。因此，数据库原理及其应用课程已成为计算机科学与技术专业、信息管理专业的重要专业课程之一。

本章将从计算机数据管理的发展过程出发，介绍数据库、数据库系统及其相关的基本概念，使读者从宏观上了解数据库技术的总体发展概貌，为后面章节的学习打下基础。

1.1 数据库的常用术语

在系统地介绍数据库原理及相关知识之前，这里先介绍一些数据库中最常用的术语和基本概念。

1.1.1 数据与数据处理

提起数据，大多数人想到的是能够进行加、减、乘、除运算以及各种统计计算的数值，如 3、6、9 等。其实，数值只是一种最简单的数据，是一种传统和狭义的数据。从广义上讲，数据 (data) 是现实世界的抽象表示，是描述客观事物特征或性质的某种符号表示。描述事物的符号可以是数值，也可以是文字、图形、图像、声音、语言等多种形式，它们都可以经过数字化处理后存入计算机。因此，从计算机数据管理的角度，可更一般地把数据定义为：凡是能够经过数字化处理存入计算机的符号都称为数据。

在日常生活中，人们直接用自然语言，如汉语、英语等表示数据。在计算机中，为了方便存储和处理这些数据，通常将描述事物特征的若干数据组成一个数据记录 (record)。

例 1.1 在学生档案管理系统中，如果人们最感兴趣的是学生的姓名、性别、出生日期、籍贯、所在系别、入学日期等数据，那么就先把它们组成如下形

式：

学生（姓名，性别，出生日期，籍贯，所在系别，入学日期）

并将其称为记录型（record type），也称为记录的逻辑结构。它是对学生特征的一个抽象描述。这里“学生”称为记录名，其中的姓名、性别、出生日期等称为字段（field）。记录型中每个字段取确定的值时，例如，

（李明，男，1972/08/21，江苏，计算机系，1990/09/01）

就成为一个记录，这个学生记录也是数据。若干记录的集合就构成一个数据文件。

对于上面这条学生记录，了解其含义的人会得到如下信息：李明是个大学生，男，1972年8月21日出生，江苏人，1990年9月1日进入计算机系学习。对于不了解其语义的人则无法理解其含义。可见，数据的组织形式还不能完全表达其内容，需要经过数据的语义解释。数据的语义就是数据的含义，数据的解释是指对数据含义的说明。因此，数据与其语义是不可分的，即数据和关于数据的解释是不可分的。

数据处理（data processing）是指对各种数据进行收集、存储、检索、分类、加工和传输等一系列活动过程的总和。

1.1.2 信息与数据的关系

自20世纪40年代信息论和控制论被提出以来，科学界一直在对信息的定义进行积极的探索，不同领域的专家从不同的角度都做了不同的解释。据1986年统计，公开发表的有关信息的定义有85种，估计现在已有百余种，但尚无一个令大家都接受的定义。多数学者认为，信息是对客观世界中各种事物的变化和特征的反映，是客观事物之间相互作用和联系的表征，是客观事物经过感知或认识后的再现。下面，我们介绍一个在计算机数据管理及其应用的有关课程中常用的信息定义。

信息（information）是对客观世界中各种事物的运动状态和特征的抽象描述，是数据处理的结果，载荷信息的物理符号是数据。

因此，信息与数据的关系可以简单地表示为

$$\text{信息} = \text{数据} + \text{数据处理}$$

具体地说，信息与数据有如下的区别和联系：

- 1) 数据是描述客观事物特征或性质的某种符号；
- 2) 数据是载荷信息的物理符号并被称为载体；
- 3) 数据能表示信息，但并非任何数据都是信息；
- 4) 信息是数据处理的结果，是人们能理解的、有用的数据。

1.1.3 数据独立性

数据独立性 (data independency) 是数据库领域中一个常用术语, 它是指数据与用户的应用程序之间的独立性, 包括数据的物理独立性和数据的逻辑独立性。

1) 数据的物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的。也就是说, 数据在磁盘上的数据库中怎样存储是数据库管理系统负责管理的, 用户程序不需要了解, 应用程序要处理的只是数据的逻辑结构, 这样当数据的物理存储改变时, 应用程序也不用改变。

2) 数据的逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。也就是说, 数据的逻辑结构改变了, 用户程序也可以不变。

例 1.2 设在数据库中有一个存储学习成绩的记录型 (学号, 课程名, 成绩), 若将其修改为 (学号, 课程名, 成绩, 课程编号), 则原先关于成绩记录型 (学号, 课程名, 成绩) 的有关应用程序不需作任何修改仍可使用。

1.1.4 数据库系统

数据库系统与数据库、数据库管理系统是三个既有联系又有区别的重要基本概念。

1. 数据库

通俗地讲, 数据库就是存放数据的仓库, 只不过这个仓库是放在计算机的存储设备上的, 而且数据是按一定的格式存放的。按数据库理论的定义, 数据库 (database, DB) 是长期存储在计算机内的、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, 具有最小的冗余度、较高的数据独立性和易扩展性, 并可为各种用户共享等特点。数据库本身不是独立存在的, 它是数据库系统的组成部分。

随着计算机、网络和通信等信息技术的高速发展, 电子商务、电子政务等得到迅速的发展, 并由此产生出大量的数据。人们把这些数据保存在数据库中, 其目的是为了更方便而充分地利用这些数据资源, 并将其加工处理, 抽取有用的决策信息, 为社会经济发展服务。

2. 数据库管理系统

数据库管理系统 (database management system, DBMS) 是位于用户与操作系统 (OS) 之间的, 使人们能对数据库中的数据进行科学的组织、高效地存取和维护管理的一种数据管理软件 (图 1.1)。它主要为用户提供数据定义、数据操纵、数据库控制、数据库的建立和维护等功能。DBMS 是数据库系统的核心组成部分, 其详细内容将在 1.5 节介绍。

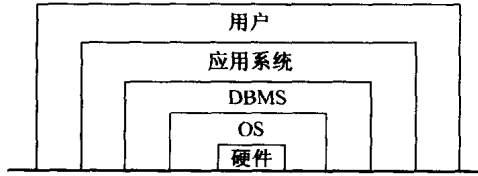


图 1.1 DBMS 在数据库系统中的地位

3. 数据库系统

数据库系统 (database system, DBS) 是指在计算机系统中引入 DB、DBMS 及其应用软件之后的系统, 因此, DBS 一般由如下几个部分组成: 硬件、软件 (包括开发工具)、数据库、数据库管理员和用户构成。本章 1.6 节将介绍数据库系统各个部分的作用。应当指出的是, 数据库的建立、使用和维护等工作仅依靠 DBMS 还远远不够, 还需要专门的人员来完成, 这些人员被称为数据库管理员 (database administrator, DBA)。

在不会引起混淆的情况下, 人们又常常把数据库系统简称为数据库。

1.2 计算机数据管理技术的产生和发展

所谓数据管理是指对数据进行分类、组织、编码、存储、检索和维护的一系列操作, 它是数据处理的前提和中心问题。

自 1946 年第一台电子计算机问世以来, 随着计算机硬件技术和软件技术的发展, 计算机的应用领域已从单纯的科学计算逐步渗透到数据管理领域。现在, 用于数据管理的计算机数量已远远超过用于科学计算的计算机。在应用需求的推动和计算机硬件、软件发展的基础上, 计算机数据管理技术不断更替和完善, 主要经历了如下三个阶段:

- 1) 人工管理阶段 (20 世纪 50 年代中期以前);
- 2) 文件系统阶段 (从 20 世纪 50 年代后期到 60 年代中期);
- 3) 数据库系统阶段 (从 20 世纪 60 年代后期至今)。

1.2.1 人工管理阶段

在这一阶段中, 计算机的主要应用领域是科学计算。从硬件上看, 计算机内存空间小, 计算速度低; 外存只有磁带、卡片和纸带, 没有像磁盘这样快速的直接存取的存储设备。从软件上看, 计算机没有操作系统, 更没有数据管理软件, 数据处理是以批处理方式进行的。

人工管理阶段具有如下特点:

1) 数据不保存。由于当时计算机主要用于科学计算，侧重于提高计算的速度和精度，相对而言数据量较少，一般不需要将数据长期保存。此外，限于内存和外存的空间和速度，其数据也不便于长期保存。因此，只是在计算某一课题时将数据输入计算机，数据用完后就从内存中撤走。不仅对用户数据如此处置，对系统软件涉及的数据有时也是这样。

2) 程序员管理数据。由于当时没有相应的软件系统负责数据的管理工作，应用程序中涉及的数据需要由程序员自己管理，即程序员在程序中不仅要编写规定数据逻辑结构的程序代码，而且还要编写设计数据物理结构的代码，包括数据存储结构、存取方法和输入方式的代码等。

3) 数据不共享。数据共享 (data sharing) 是指多个用户、多种语言、多个应用程序相互覆盖地使用一些共同的数据集合。在人工管理阶段，由于数据是在程序中定义的，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，由于必须在各自的程序中定义，无法或很难互相利用、互相参照，因此程序与程序之间的相同数据无法共享。

4) 数据不具有独立性。由于数据的逻辑结构和物理结构均在程序编码中定义，如果因为某种原因，需要改变数据的逻辑结构或物理结构时，必须对应用程序代码做相应的修改，即数据与程序之间不具有独立性，这不仅加重了程序员的负担，且使用起来也很不方便、费时费力。

在人工管理阶段，应用程序与数据之间的一一对应关系可用图 1.2 表示。其中，左边表示数据从程序外部通过键盘等设备进入程序的情形，右边表示数据集集成在程序中的情形。

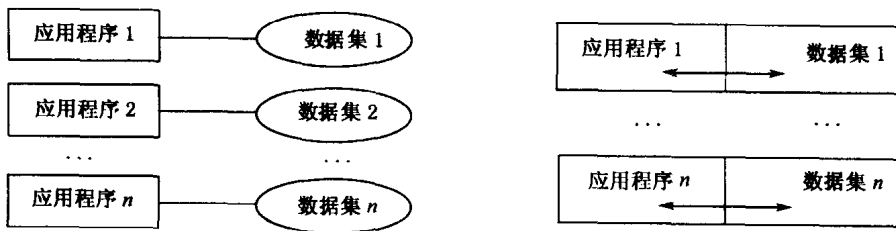


图 1.2 人工管理阶段应用程序与数据之间的对应关系

1.2.2 文件系统阶段

在这一阶段中，计算机不仅仅用于科学计算，而且还开始大量地用于数据处理。从硬件上看，计算机内存空间增大，计算速度得到很大的提高，外存有了磁盘、磁鼓等快速的直接存取存储设备；从软件上看，计算机配备了操作系统，且操作系统中已经有了专门的数据管理软件——文件系统。从数据处理方式上看，

不仅有批处理，而且能够联机实时处理。

文件系统阶段有如下特点：

1) 数据可以长期保存。由于计算机大量用于数据处理，数据需要长期保存以便反复使用。把一批数据以一个文件的形式保存在磁盘等外存储器上，为数据的长期保存和反复使用（查询、修改、插入和删除等）提供了保证。

2) 文件的多样化和结构化。由于有了直接存取设备，也就出现了顺序文件、索引文件、随机文件等多种文件组织形式。数据文件的存取基本上以记录为单位组织，而记录则是由一些字段按特定的结构组成，每个记录的长度均相等，如1.1.1节介绍的学生记录。因此，文件是记录的集合，文件系统实现了记录内的结构性，但在整体上数据是无结构的。

3) 文件系统管理数据。操作系统提供的文件管理系统是应用程序与数据文件的专门接口。程序和数据文件之间的存取操作由文件系统根据“按文件名访问，按记录进行存取”的管理技术自动完成。文件的逻辑结构到物理结构的转换由文件系统来自动完成，而不必由程序员设计考虑，从而使得程序和数据分开，数据与程序之间有了一定的独立性。这样，程序员可以集中精力于数据的逻辑结构，而不必或者很少考虑其物理结构。数据在物理结构上的改变，通常不会反映在程序上，从而可以大大地节省维护程序的工作量。

文件系统的上述特点比人工管理阶段有了很大的改进，但仍存在如下缺点：

1) 数据冗余度大。数据冗余（data redundancy）是指同一组数据在多个文件中同时出现所引起的数据重复现象。在文件系统中，一个文件基本上对应于一个应用程序，即文件系统中文件仍由应用程序来定义。当不同的应用程序具有部分相同的数据时，也必须建立各自的数据文件，而不能共享相同的数据，例如，在人事文件中包含了企业所有雇员的信息，而销售文件中又包含了属于销售人员的雇员信息。此外由于相同的数据在不同的文件中重复存储、并由各自的程序管理，容易造成数据的不一致性，给数据的修改和维护带来困难。

2) 数据独立性较差。由于文件的逻辑结构是在应用程序中定义的，文件系统中的文件是为某一特定应用服务的，因此，一旦数据的逻辑结构改变，必须修改应用程序中关于文件结构的定义。反过来，应用程序的改变，例如应用程序改用不同的高级语言编写等，也将引起数据文件结构的改变。因此数据与程序之间独立性较差。

3) 数据联系弱。数据联系（data relationship）是指不同文件中的数据之间的联系。在文件系统阶段，不同文件中的数据相互之间是独立的，缺乏相互关联的方法，因此数据联系弱。虽然文件之间的某些数据存在着紧密的逻辑联系，但是由于实现的复杂性，很少在文件管理系统中提供这些数据之间的联系方法，因此，各个数据文件之间是孤立的，不能反映现实世界中客观事物之间的内在联系。