

国家信息产业部电子人才交流中心参与规划
“十一五”高等院校应用型规划教材

数据库

S h u J u K u

原理与应用

Y u a n L i J i Y i n g Y o n g

基于 SQL Server 2000
C# 与 ADO.NET 相结合

科海策划

常玉慧 主编
钱进 张俐 副主编



科学出版社

数据库原理与应用

主编 常玉慧

副主编 钱进 张俐

科学出版社

内 容 简 介

本书是在“数据库原理与应用”课程教学大纲的基础上,结合编者多年的教学经验,为培养计算机应用型人才而编写。

本书将数据库的基本原理和数据库应用系统与 SQL Server 2000 紧密地联系在一起。主要内容包括数据库系统概述、关系模型、标准查询语言、数据库规范化、数据库设计、数据库保护、数据库新技术和数据库应用系统。在理论讲解的过程中穿插大量教学实例,SQL 查询语言的所有例子均可在 SQL Server 2000 查询分析器中操作,数据库应用系统将 C# 与 ADO.NET 相结合。读者读完本书以后,可以参照学生管理信息系统来编制其他的数据库应用程序。

本书力求深入浅出,通俗易懂,可作为高等院校计算机或相关专业数据库课程的教材,也可作为参加全国各种软件水平考试用书,或作为广大数据库开发者的参考书。

图书在版编目(CIP)数据

数据库原理与应用/常玉慧,钱进,张俐编著.—北京:科学出版社,2006

ISBN 7-03-017719-3

I. 数… II. ①常…②钱…③张… III. 数据库系统 IV.TP311.13

中国版本图书馆 CIP 数据核字(2006)第 083168 号

责任编辑:潘秀燕 / 责任校对:刘雪莲
责任印刷:科海 / 封面设计:林陶

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>

北京市耀华印刷有限公司

科学出版社发行 各地新华书店经销

2006 年 8 月第一版

开本:16 开

2006 年 8 月第一次印刷

印张:17.75

印数:1-4000

字数:432 千字

定价:32.00 元

(如有印装质量问题,我社负责调换)

前 言

数据库技术是计算机科学技术中发展最快、应用最广的一个分支，是信息系统研究和开发的核心技术。全国数据库学术会议至今已经举办了 22 届了，各行各业的信息系统，包括在 Internet 上，数据库技术的应用比比皆是，这充分说明了数据库已成为了信息社会的重要基础，并已经成为了各个企业、学校不可缺少的一部分。

本书作为江苏技术师范学院的重点建设教材，在编写过程中将理论与实践紧密结合。在充分把握本课程教学体系结构的基础上，结合编者多年的教学经验，并汲取了其他同类教材的精华，又增加了实用性的数据库应用系统，力求使本教材体现“理论通俗易懂，实践跟上潮流”，真正满足培养计算机应用型人才的需要。

本书共分 8 章。

- 第 1 章主要介绍数据库系统的应用情况、发展历史和一些数据库的基本概念。
- 第 2 章主要介绍关系数据库理论模型。
- 第 3 章主要介绍关系数据库 SQL 标准查询语言，所举例子均能在 SQL Server 2000 查询分析器中操作。
- 第 4 章主要介绍关系模式的规范化理论，增加了大量的教学实例。
- 第 5 章主要介绍数据库设计的全过程，并增加了基于 UML 的数据库设计。
- 第 6 章主要讨论数据库的保护，包括数据库的安全性、完整性、并发控制和故障恢复。
- 第 7 章介绍了面向对象数据库系统、分布式数据库系统、并行数据库系统、数据仓库、数据挖掘等一些新技术，特别增加了比较典型数据挖掘技术，如关联规则、分类等。
- 第 8 章主要介绍了利用 ADO.NET 技术和高级编程语言 C#编写的一个实际的数据库应用系统案例，讨论了一个数据库系统的登录、数据处理、数据查询、统计报表等重要内容。

本书内容丰富，脉络清晰，希望读者在掌握基本理论知识的同时，强化对应用系统的开发；同时也让读者了解一些数据库新技术，为适应数据库技术的飞速发展打下良好的基础。

本书可作为高等院校计算机或相关专业数据库课程的教材，也可作为参加全国各种软件水平考试用书，或作为广大数据库开发者的参考资料。

本书由常玉慧任主编，负责全书的统编、修改，钱进任副主编。其中，第1、2、4章由常玉慧编写；第3、5、7、8章由钱进编写；第6章由张俐编写。侯雪亚负责全书的编排工作。

本书承蒙徐亚平教授进行了认真的审阅，并提出了宝贵的意见，编者在此深表谢意。

本书编写时参考了相关的资料，谨向有关作者表示感谢。

为了便于教师使用本书进行教学，本书还为教师准备了电子课件。由于编写时间仓促，作者水平有限，书中难免存在缺点和错误，敬请读者批评指正。

如有任何问题，请发邮件与我们联系：

feedback@khp.com.cn

编 者

2006年7月

目 录

第 1 章 数据库系统概述..... 1

1.1 数据库的特点及相关概念	1
1.1.1 引言	1
1.1.2 数据、信息和数据处理	2
1.1.3 数据库系统	3
1.2 数据管理技术的发展	5
1.2.1 人工管理阶段	5
1.2.2 文件系统阶段	6
1.2.3 数据库系统阶段	7
1.3 数据抽象与数据库的三级模式	11
1.3.1 数据抽象	11
1.3.2 数据库的三级模式结构	12
1.3.3 数据库的二级映像功能与数据 独立性	13
1.4 数据模型	14
1.4.1 信息的三种世界及其描述	14
1.4.2 概念模型及其表示	15
1.4.3 数据模型	17
1.5 数据库技术的发展	21
1.6 小结	24
习题 1	24

第 2 章 关系数据库..... 25

2.1 关系模型及其定义	25
2.1.1 关系模型概述	25
2.1.2 关系模式	26
2.1.3 关系数据库	27
2.2 关系的数学定义	27

2.3 关系模型的完整性	30
2.3.1 实体完整性	30
2.3.2 参照完整性	30
2.3.3 用户定义完整性	31
2.4 关系代数	31
2.4.1 传统的集合运算	32
2.4.2 专门的关系运算	36
2.5 关系演算	41
2.5.1 元组关系演算	41
2.5.2 域关系演算	41
2.5.3 关系数据库查询语言	42
2.6 小结	48
习题 2	49

第 3 章 关系数据库标准语言 SQL 50

3.1 SQL 概述及特点	50
3.1.1 SQL 的发展历程	50
3.1.2 SQL 数据库的体系结构	51
3.1.3 SQL 的组成	52
3.2 SQL 的数据定义 (DDL)	52
3.2.1 SQL 语言基本知识	53
3.2.2 数据库的定义、修改和删除	54
3.2.3 基本表的创建、修改和删除	56
3.2.4 索引的创建和删除	59
3.3 SQL 的数据操纵	61
3.3.1 数据查询 SELECT 语句	61
3.3.2 单表查询	62

3.3.3 多表连接查询.....	64	4.3.3 第三范式 (3NF)	102
3.3.4 嵌套查询	66	4.3.4 改进的 3NF——BCNF	102
3.3.5 集合查询	71	4.3.5 多值依赖与 4NF	104
3.3.6 SQL 数据更新	72	4.3.6 规范化小结	106
3.4 视图的创建和删除	75	4.3.7 关系模式的规范化及 反规范化	106
3.4.1 视图的作用	75	4.4 关系模式的分解算法	107
3.4.2 创建视图	76	4.4.1 关系模式分解的算法基础	107
3.4.3 删除视图	77	4.4.2 关系模式的分解	111
3.4.4 查询视图	78	4.4.3 分解的无损连接性	111
3.4.5 更新视图	78	4.4.4 分解的函数依赖保持性	114
3.5 数据控制语句	79	4.4.5 分解算法	115
3.5.1 授予权限	79	4.5 小结	118
3.5.2 收回权限	82	习题 4	119
3.6 存储过程和触发器	82	第 5 章 数据库设计	120
3.6.1 存储过程概述	82	5.1 数据库设计概述	120
3.6.2 存储过程的相关操作	83	5.1.1 数据库设计的任务、内容 和特点	121
3.6.3 触发器概述	87	5.1.2 数据库设计方法	122
3.6.4 触发器的相关操作	87	5.1.3 数据库设计的基本步骤	124
3.7 嵌入式 SQL	89	5.2 需求分析	126
3.7.1 嵌入式 SQL 概述	89	5.2.1 需求分析的任务、内容和方法	127
3.7.2 SQL 与主语言之间的通信	90	5.2.2 需求分析的具体步骤	128
3.7.3 嵌入式 SQL 的使用技术	92	5.2.3 数据流图和数据字典	129
3.7.4 动态 SQL 简介	94	5.3 概念结构设计	134
3.8 小结	95	5.3.1 概念结构概述	134
习题 3	95	5.3.2 数据抽象与局部视图设计	137
第 4 章 关系数据库规范化理论. 96		5.3.3 全局概念结构设计	138
4.1 问题的提出	96	5.4 逻辑结构设计	140
4.2 函数依赖	97	5.4.1 E-R 图向关系模型的转换	141
4.2.1 属性间关系	98	5.4.2 用关系规范化理论优化关系 数据模型	142
4.2.2 函数依赖	98	5.4.3 设计用户子模式	143
4.3 关系的范式	100		
4.3.1 第一范式 (1NF)	101		
4.3.2 第二范式 (2NF)	101		

5.5 数据库的物理设计	144	6.5 数据库的并发控制	189
5.6 数据库实施	146	6.5.1 并发操作带来的问题	189
5.7 数据库运行与维护	148	6.5.2 封锁技术	190
5.8 基于 UML 的关系数据库设计	149	6.5.3 封锁协议	192
5.8.1 UML 概述	149	6.5.4 封锁带来的问题	193
5.8.2 基于 UML 的数据库设计	150	6.5.5 并发调度的可串行性	195
5.8.3 对象模型的建立	151	6.5.6 SQL Server 2000 锁机制	196
5.8.4 UML 类图中的类映射成表 的策略	153	6.6 小结	201
5.9 综合举例	155	习题 6	201
5.10 小结	157		
习题 5	158	第 7 章 数据库新技术	203
第 6 章 数据库保护	159	7.1 数据库新技术概述	203
6.1 事务	159	7.2 面向对象数据库系统	205
6.1.1 事务的定义	159	7.2.1 面向对象数据库系统概述	205
6.1.2 事务的 ACID 性质	160	7.2.2 面向对象的基本概念	206
6.2 数据库的完整性	161	7.2.3 面向对象设计方法	208
6.2.1 完整性子系统和完整性规则	161	7.2.4 面向对象数据库	208
6.2.2 完整性约束的分类	162	7.2.5 面向对象数据库语言	210
6.2.3 触发器	163	7.2.6 面向对象数据库的模式演进	210
6.2.4 SQL Server 2000 的数据库 完整性及实现方法	164	7.2.7 面向对象数据库的事务管理	212
6.3 数据库的安全性	169	7.3 分布式数据库系统	213
6.3.1 安全性	169	7.3.1 分布式数据库概述	213
6.3.2 权限	169	7.3.2 分布式数据库系统的体系结构	215
6.3.3 常用的安全性措施	171	7.3.3 分布式数据库管理系统	218
6.3.4 SQL Server 2000 安全模式	173	7.3.4 分布式数据库查询处理和优化	219
6.4 数据库恢复	181	7.3.5 分布事务管理	221
6.4.1 备份	181	7.4 并行数据库系统	223
6.4.2 恢复	181	7.4.1 并行数据库系统体系结构	223
6.4.3 故障的种类和恢复方法	182	7.4.2 并行数据库系统并行处理 技术简介	225
6.4.4 SQL Server 2000 的数据备份和 恢复机制	182	7.5 数据仓库与数据挖掘	229
		7.5.1 数据仓库概念与结构	229
		7.5.2 数据仓库设计和实现	234
		7.5.3 联机分析处理 (OLAP)	235

7.5.4 数据挖掘	237	8.2.1 ADO.NET 对象及其编程	248
7.6 小结	243	8.2.2 利用 ADO.NET 控件编写 数据库应用程序	254
习题 7	243	8.3 综合应用数据库系统	258
第 8 章 使用 ADO.NET 开发应用 系统	244	8.3.1 学生管理信息系统主要功能	258
8.1 ADO.NET 概述	244	8.3.2 学生管理信息系统数据库设计	259
8.1.1 ADO.NET 的新特点	244	8.3.3 程序开发	260
8.1.2 ADO.NET 对象模型	245	8.4 小结	274
8.1.3 访问数据库方式	247	习题 8	274
8.2 使用 ADO.NET 开发数据库 应用程序	248	参考文献	275

第 1 章

数据库系统概述

教学目标

- 了解数据和信息的基本含义及它们之间的相互关系
- 了解数据管理技术的发展历程
- 理解数据库系统的组成和功能
- 理解数据库、数据库管理系统和数据库系统的概念及它们之间的联系
- 掌握数据库系统的三级模式结构及它们之间的映像
- 了解现实世界、信息世界和数据世界的含义及它们之间的相互关系
- 理解概念模型和数据模型的相关概念
- 掌握常见的四种数据模型
- 了解数据库技术新进展

1.1 数据库的特点及相关概念

1.1.1 引言

信息化是继工业化之后社会生产力发展的一个新阶段，将对社会经济的发展乃至整个人类文明产生巨大的影响。在信息化的社会里，信息资源对人类社会生活的重要性不断提高，信息资源的占有与利用水平已成为衡量一个国家或企业综合实力和竞争力的重要标志。进入 20 世纪 90 年代，随着信息技术及产业的迅猛发展，信息化浪潮在蓬勃兴起。我国对推进信息化建设给予了高度重视。从 1993 年开始，国家信息化重大工程项目如金桥、金卡、金关、金税等“金”系列开始启动，信息资源开发与利用迅速发展，领域信息化和社会信息化取得了明显进展。特别是远程教育、远程医疗、网上娱乐、数字图书馆等新业务得到不断应用，社会和经济效益明显。

在计算机的三大主要应用领域（科学计算、数据处理和过程控制）中，数据处理是计算机应用的主要方面。数据库技术是作为数据处理中的一门技术而发展起来的，它所研究的问题是如何科学地、有效地获取数据、组织数据、存储和处理数据。数据库技术已越来越成为现代信息技术的重要组成部分，是现代计算机信息系统和诸多其他计算机应用系统

的基础和核心，是一个国家实现信息化的重要基础。一个国家的数据库建设规模、数据库信息量的大小和使用频度已经成为衡量这个国家信息化程度的重要标志之一了。

1.1.2 数据、信息和数据处理

1. 数据

数据是记录在一定介质上并可鉴别的符号，可以是字符、数字、文本、声音、图形、图像、图表、图片等，可以被计算机所识别、存储和处理。

数据具有四个基本特征：

(1) 数据有“类型”和“值”之分。例如，学生姓名其类型为字符型，其值为“张三”等等。

(2) 数据受数据类型和取值范围的约束。例如，学生年龄为整数类型，取值范围在18~35之间。

(3) 数据有定性数据和定量数据之分。例如，学生身高可以用“很高”来表示，也可以用1.75米来表示。

(4) 数据具有多种表示形式。例如，学生性别可以用“男”或“女”来表示，也可以用0或1来表示。

用数据描述的对象可以是实实在在的事物，如一个学生的学号、姓名、性别、年龄、班级等，也可以是抽象的事物，如用文本描述的一个想法，用图画描述的一个画面等等。这些都是数据，都可以输入到计算机中，通过计算机对其进行适当的管理和操作。用数据描述客观事物，即用数据的形式表示事物的特性或特征。反之，事物的特性或特征也即数据的语义。

例如，一个学生的数据如学号、姓名、性别、年龄和班级等，描述了一个学生的有关特征和特性。反过来，这些特征也就是相应数据的语义。数据是需要解释的，也就是要对数据的含义或语义进行解释才有意义，因为数据的形式即符号本身并不能完全表达其内涵。

2. 信息

信息是经过加工后具有新的事实知识的数据。具体来说，信息是对数据进行加工、处理的结果，这种数据形式对接受者来说是有意义的，而且对当前和将来的决策具有明显的或实际的价值。

信息具有四个基本特征：

(1) 信息的内容是有关现实世界的事实和知识。

(2) 信息是有用的，是人们活动的必需知识。

(3) 信息能够在空间和时间上被传递，在空间上传递信息称为信息通信，在时间上传递信息称为信息存储。

(4) 信息需要一定的表示形式，信息与其表现符号不可分离。

在现实生活中，数据与信息两个术语常常被混淆，但是数据与信息在概念上还是有区别的。不是所有数据都能成为信息，只有抽取加工之后，具有新的事实知识的数据才能成

为信息。数据经过加工处理之后形成的信息，仍然以数据的形式表现，此时数据是信息的载体，也是人们能动地认识和改造世界的媒体。例如，某经销部门看到的某个地区或商店的日销售额或月销售额，是经过累加一天或一个月的销售记录而得到的。这个数据，对于销售总管来说是一个原始数据，他需要根据各地区各商店的销售额来规划产品的市场销售计划和策略，从而提高其产品的市场占有率和利润。同样是这个数据，对于本地或本商店的销售人员也可以看作是信息，可用来规划下一步的促销行动。由此可以得出，数据和信息之间的关系是原料和成品之间的关系，同时又具有相对性。

信息还具有以下几点作用：

- (1) 它可以提高人们对事物的认识，减少人们活动的盲目性。
- (2) 信息是社会机体进行活动的纽带，社会的各个组织通过信息网相互了解并协同工作，使整个社会协调发展。
- (3) 信息又是管理活动的核心。

3. 数据处理

数据处理是指将数据转换成信息的过程，即对数据进行收集、组织、整理、加工、存储和传播等一系列活动过程。数据处理的过程主要包括数据管理——收集信息、表示数据、组织和保存数据；数据加工——对数据进行变换、抽取和运算；数据传播——在空间或时间上以各种形式传播信息（不改变数据的结构、性质和内容）。通过数据管理获得数据；通过数据加工得到更有用的数据（以指导或控制人的决策行为或事物的变化趋势）；通过数据传播，使更多的人共享信息。

1.1.3 数据库系统

数据库（Data Base, DB）、数据库管理系统（Data Base Management System, DBMS）和数据库系统（Data Base System, DBS）是数据库技术中最基本、最常用的三个既有区别又有联系的术语。

1. 数据库

数据库是长期储存在计算机内的、有组织的、可共享的数据集合，也是现实世界中相互关联的大量数据及数据间关系的集合。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。例如，汽车修理商店的数据库包含能够识别顾客（包括顾客姓名、地址、单位电话、家庭电话及信用卡号）及其汽车（包括制造商、型号和年份）以及每辆车的维修史（包括维修日期、维修人员姓名、维修类型和维修费用）的数据。

数据库具有如下特性：

- (1) 数据库是具有逻辑关系和确定意义的数据集合。逻辑上无关的数据集合不能称为数据库。
- (2) 数据库是针对明确的应用目标而设计、建立和加载的。每个数据库都具有一组用户，并为这些用户的应用程序服务。

(3) 一个数据库表示了现实世界的某些方面(称为小世界)。一个数据库应随其所表示的小世界的改变而改变。

2. 数据库管理系统

数据库管理系统是一个通用的软件系统,是位于用户和操作系统之间的数据管理软件,由一组计算机程序构成。数据库管理系统能够对数据库进行有效的管理,包括存储管理、安全性管理、完整性管理等。数据库管理系统又能提供一个软件环境,使用户能方便快速地建立、维护、检索、存取和处理数据库中的信息。在操作系统的支持下,对内负责对数据库中的数据进行管理,对外负责对用户提供操作数据库的界面。

成熟的 DBMS 产品有大家熟悉的 Access, Foxpro 等小型 DBMS 软件,也有 IBM 的 DB2、Oracle、微软的 SQL Server 系列产品、Informix 等大型的 DBMS 软件。

数据库管理系统是数据库系统的一个重要组成部分。数据库管理系统主要包括以下几个方面的功能:

(1) 数据定义功能

DBMS 提供数据定义语言(Data Definition Language, DDL),通过它可以方便地定义数据库中数据对象的逻辑结构。

(2) 数据操纵功能

DBMS 提供数据操纵语言,通过它可以操纵数据库中的数据。如对数据库中的数据进行查询、插入、删除和修改等操作。

数据操纵语言可以分为两大类。一类是嵌入主语言中的,如 C、COBOL 等高级语言中,这类数据操纵语言本身不能独立使用,因此称为宿主型数据操纵语言。另一类是交互式命令语言,这类语言语法简单,可以独立使用,所以称为自主型数据操纵语言。

(3) 数据库的运行管理

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,以保证数据的安全性、完整性、并发控制和数据库的恢复。

(4) 数据库的建立和维护功能

包括数据库初始数据的输入、转换功能,数据库的转储、恢复功能,数据库的重组功能和性能监视、分析功能等,这些功能是由一些实用程序来完成的。

3. 数据库系统

数据库系统是一个采用数据库技术有机结合的人机系统。广义地讲,一个数据库系统应该包括数据库应用系统、数据库应用开发接口、数据库管理系统、操作系统、数据库和用户,如图 1-1 所示。它就是一个采用数据库技术的计算机存储系统,该系统的目标是存储信息并支持用户检索和更新所需要的信息。

数据库系统用户划分为以下五个类别:

(1) 数据库设计者:定义数据库,按用户需求建立和合并数据视图。

(2) 系统分析员和应用程序员:定义数据处理事务及编程实现。

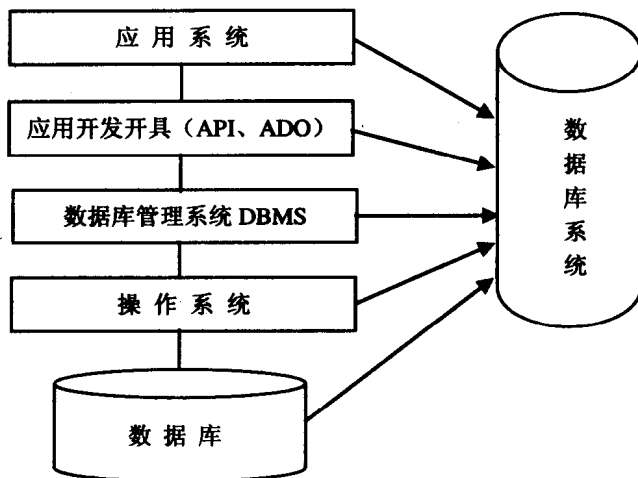


图 1-1 数据库系统

(3) 数据库管理员 (DBA)：监督管理共享资源，即数据库与 DBMS。数据库是为多用户共享的，因此 DBA 负责进行规划、设计、协调、维护和管理。

(4) 最终用户 (EU)：访问数据库的用户。最终用户可以划分为以下三类：

- 简单用户：使用数据库应用系统进行事务处理的用户，如操作员。
- 偶然用户：偶然使用查询语言访问数据库进行简单统计和分析的用户，如管理层人员。
- 复杂用户：通过 DBMS 访问数据库，并对数据进行分析、决策的用户，如决策层人员。

(5) 其他人员：DBMS 及有关实用程序的开发人员、系统维护人员等。

1.2 数据管理技术的发展

数据管理具体就是指人们对数据进行收集、分类、组织、存储、检索和维护的一系列活动。伴随着计算机软硬件技术的发展，数据管理技术经历了人工管理、文件管理、数据库管理三个阶段。每一阶段的发展以数据存储冗余不断减小、数据独立性不断增强、数据操作更加方便和简单为标志，各有特点。

1.2.1 人工管理阶段

在计算机出现之前，人们利用纸张记录数据，利用计算工具（算盘、计算尺）进行计算数据，并主要使用人的大脑来管理和利用这些数据。而早期的计算机主要用于数值计算，也无管理数据的软件。因此，从计算机记录的数据来看，数据量小，数据无结构，用户直接管理，数据仅依赖特定的应用，而且数据间缺乏逻辑组织，缺乏独立性。

这一阶段的数据管理有下列特点：

(1) 数据不能保存

当时计算机的硬件存储设备昂贵, 计算机主要用于计算, 计算完后, 数据也不需要长期保存。在需要计算时, 利用卡片、纸带等将数据输入, 经过运算后输出结果, 数据处理的过程就结束了, 数据空间随程序空间一起被释放。

(2) 数据不能独立

数据是输入程序的重要组成部分, 数据和程序作为一个不可分割的整体同时提供给计算机计算使用。对数据进行管理, 就像现在的操作系统可以以目录、文件的形式管理数据。程序员不仅要知道数据的逻辑结构, 也要规定数据的物理结构, 程序员对存储结构, 存取方法及输入输出的格式有绝对的控制权, 要修改数据必须修改程序。要对 100 组数据进行同样的运算, 就要给计算机输入 100 个独立的程序, 因为数据无法独立存在。

(3) 没有文件的概念

数据的组织完全由程序员自行设计和安排。即使人们发现了这样做的弊病, 也无可奈何。因为此时计算机的外存能力是很弱的。

(4) 数据是面向程序的

一组数据对应一个程序。不同程序的数据之间是相互独立、彼此无关的, 即使两个不同程序涉及到相同的数据, 也必须各自定义, 无法互相利用, 互相参照。数据不但高度冗余, 而且不能共享。

所以有人也称这一数据管理阶段为无管理阶段。这一阶段应用程序与数据之间的关系如图 1-2 所示。

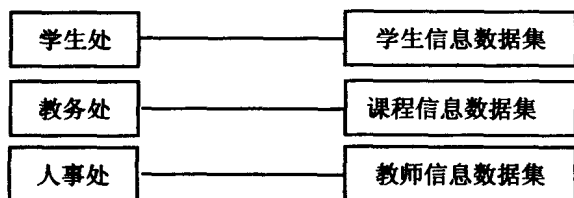


图 1-2 人工管理阶段程序与数据的关系

1.2.2 文件系统阶段

这一时期的数据管理是把计算机中的数据组织成相互独立的数据文件, 并可按文件的名字进行访问, 实现对数据的存取。数据可以长期保存在计算机外存上, 可以对数据进行反复处理, 并支持文件的查询、修改、插入和删除等操作, 这就是文件系统。

这一时期的特点是:

(1) 数据长期保留

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处理, 即经常有查询、修改和删除等操作。

(2) 数据的独立性

由于有了操作系统, 利用文件系统进行专门的数据管理, 使得程序员可以集中精力进行算法设计, 而不必过多地考虑细节。比如要保存数据时, 只需编写保存指令, 而不必所

有的程序员都还要精心设计一套程序，控制计算机如何物理地保存数据。在读取数据时，只要给出文件名，而不必知道文件的具体存放地址。文件的逻辑结构和物理存储结构由系统进行转换，程序与数据有了一定的独立性，数据的改变不一定要引起程序的改变。数据文件中有 100 条记录时，使用某一个查询程序。当文件中有 1000 条记录时，仍可使用同一个查询程序。

(3) 实时处理

由于有了直接存取设备，也有了索引文件、链接存取文件、直接存取文件等存储机制，所以既可以采用顺序批处理，也可以采用实时处理方式。数据的存取以记录为基本单位。

文件系统数据管理如图 1-3 所示。文件系统实现了记录内的结构化，但从文件的整体来看却是无结构的，其数据面向特定的应用程序。因此，数据共享性、独立性差，且冗余度大，管理和维护的代价也很大。

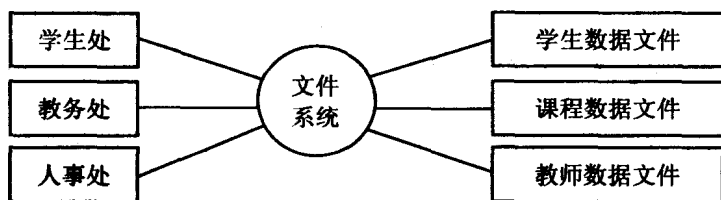


图 1-3 文件管理阶段程序与数据的关系

1.2.3 数据库系统阶段

1. 数据库系统的发展阶段

从 20 世纪 60 年代后期开始，数据管理进入数据库系统阶段。这一时期，计算机管理的规模日益庞大，应用越来越广泛，数据量急剧增长，要求共享数据的呼声越来越强。此时的计算机有了大容量磁盘，硬件价格下降，编制软件和维护软件的费用相对在增加。联机实时处理的要求更多，并开始提出和考虑并行处理。在这样的背景下，数据管理技术进入数据库系统阶段。

用数据库系统来管理数据比文件系统具有明显的优点，从文件系统到数据库系统，标志着数据管理技术的飞跃。下面来详细地讨论数据库系统的特点及其带来的优点。

2. 数据库系统阶段数据管理技术的特点

数据库的特点是数据不再只针对某一个特定的应用，而是面向全组织，具有整体的结构性，共享性高，冗余度减小，具有一定的程序与数据之间的独立性，并且对数据进行统一的控制，保证数据的完整性和安全性，提高了数据管理效率。数据库也是以文件方式存储数据的，但它是数据的一种高级组织形式。在应用程序和数据库之间，由数据库管理软件 DBMS 把所有应用程序中使用的相关数据汇集起来，按统一的数据模型，以记录为单位存储在数据库中，为各个应用程序提供方便、快捷的查询及使用。这一阶段管理数据的方法如图 1-4 所示。

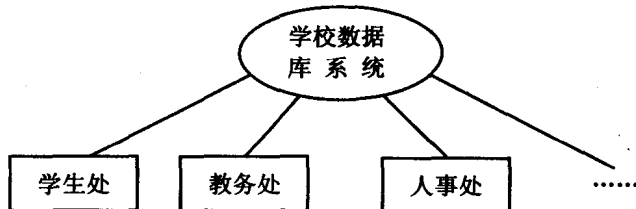


图 1-4 数据库系统阶段程序与数据的关系

这一阶段使用的数据管理方法有以下特点：

(1) 信息完整，功能通用

数据库系统的通用性是指抽象出所有文件的元数据并统一存储管理。所谓元数据是指数据库的说明信息，包括数据库每个文件的结构、每个数据项的存储格式和数据类型、数据的完整性约束等。元数据存储称为数据字典的特殊文件中。应用程序如果要访问数据库，首先通过 DBMS 从数据字典中取得有关的元数据，然后根据这些信息获取数据库的数据。

在文件系统中，文件的元数据分散在不同的应用程序中，所以文件与能够阅读这些文件的应用程序形成了特定的关系，其他的应用程序由于不了解元数据而无法访问文件。数据库系统的通用性则是抽象出所有文件的元数据，统一了存储和管理。

(2) 程序与数据独立

由于数据库系统把所有文件的元数据存放在数据字典中作统一存储和管理，从而实现了应用程序与数据的独立性，即当文件结构改变时，应用程序不必改变。这是因为通过 DBMS 完成对数据字典的相应修改。

相比之下，在文件系统中，文件的元数据嵌套在应用程序中，文件结构的任何改变将引起所有访问该文件的应用程序的改变。

例如，应用程序原来设计为可以访问任课教师文件（教师号，姓名，电话，部门），它包含对文件结构描述的元数据。但当文件新增数据项“学位”后，应用程序必须修改才能继续使用。

(3) 数据抽象

数据库系统提供了数据的抽象概念表示，使得用户不必了解数据库文件的存储结构、存储位置、存取方法等细节就能访问数据库。数据模型是提供数据抽象概念的有利工具，使用逻辑概念表示数据。不同类型的数据库系统使用不同的数据模型实现数据的抽象概念表示。数据模型屏蔽了数据存放的物理细节，这种方法表达的数据更容易被用户理解。在数据库系统中，每个文件的存储结构、存取方法等细节存放在数据字典中，访问数据库时，用户只需引用数据的抽象概念表示，DBMS 负责从数据字典中提取文件的存储结构、存取方法等细节，把用户引用的抽象概念转换为物理表示，完成用户的访问要求。

(4) 支持数据的不同视图

视图是关系数据库系统提供给用户以多种角度观察数据库中数据的重要机制。一个视图可以是一个数据库子集，也可以是多个数据库的子集按某种方式构成的虚拟数据库（不是实际存储的数据库）。