

高等院校计算机专业教育改革推荐教材

数据库应用 技术基础

黄志球 李 清 等编著



机械工业出版社
CHINA MACHINE PRESS



高等院校计算机专业教育改革推荐教材

数据库应用技术基础

黄志球 李 清 等编著

机 械 工 业 出 版 社

本书将数据库的基础理论、数据库管理系统和数据库应用技术融为一体,以关系数据库为重点,系统地介绍了相关的理论、方法和技术。第1篇介绍了数据库系统的基本概念、数据模型、关系数据库理论和标准查询语言SQL;第2篇描述了数据库管理系统的组成、功能和实现技术;第3篇面向应用,介绍了数据库应用系统和数据库应用的新技术。各章均附有习题,附录还配备了上机实验。

本书内容丰富、注重理论和技术基础,反映了当前数据库的新技术和新发展,可作为计算机专业或相关专业教材,也可供广大从事信息领域的技术人员参考。

图书在版编目(CIP)数据

数据库应用技术基础/黄志球等编著. —北京:机械工业出版社,2003.10
高等院校计算机专业教育改革推荐教材
ISBN 7-111-12898-2

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(2003)第 071141 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策 划:胡毓坚

责任编辑:田 梅

责任印制:闫 焱

北京京丰印刷厂印刷·新华书店北京发行所发行

2003 年 10 月第 1 版·第 1 次印刷

787mm×1092mm $1/16$ ·19.5 印张·480 千字

0 001—5 000 册

定价:27.00 元

凡购本图书,如有缺页、倒页、脱页,由本社发行部调换

本社购书热线电话(010) 68993821、88379646

封面无防伪标均为盗版

高等院校计算机专业教育改革推荐教材

编委会成员名单

主 编 刘大有

副主编 王元元

编 委 (按姓氏笔画排序)

刘晓明 李师贤 张桂芸 徐汀荣

耿亦兵 顾军华 黄国兴 薛永生

编者的话

计算机科学技术日新月异的飞速发展和计算机科学技术专业教育的相对滞后,已是不争的事实。

有两个发人深省的现象:一是,由于非计算机专业的学生既具有一门非计算机专业的专业知识,又具有越来越高的计算机应用技术水平,从而使计算机专业的学生感受到一种强烈的冲击和压力;二是,创建软件学院的工作已有近两年的历史,但软件学院的计算机专业教育的定位仍在探讨之中。

我们认为计算机科学与技术专业(以下简称计算机专业)教育的改革势在必行,正确认识和划分计算机专业教育的层次,对该专业的教育改革无疑是一个非常重要的问题。我国的计算机专业教育主要分三个层次。一般说来,这三个层次通常分布在以下三类高等院校:

第一层次主要以具有计算机一级学科博士学位授予权的教育部属重点高等院校为代表(包括具有两个博士点的大学)。这一类大学本科着重培养理论基础比较坚实、技术掌握熟练、有一定研究和开发能力的计算机专业学科型人才,其中部分学生(约本科生的 10 %)可攻读博士学位。

第二层次主要以具有一个计算机二级学科专业博士点的教育部属高等院校为代表。这一类高等院校本科着重培养有一定的理论基础、技术掌握比较熟练、有一定的研究或开发能力的计算机专业人才,其中一部分培养成学科型人才,另一部分培养成应用型人才,一小部分学生(约本科生的 5 %)可攻读博士学位。

第三层次主要以具有计算机二级学科专业硕士点的省属高等院校为代表。这一类高等院校本科面向企业应用,侧重培养对计算机技术或部分计算机技术掌握比较熟练,有一定的开发、应用能力的计算机专业应用型人才,其中很小一部分学生(约本科生的 2.5 %)可攻读博士学位。

国家教育部、计委批准的或省教育厅批准的示范性软件学院,就其培养目标和办学特色而言,分别与第二层次中应用型人才培养部分以及第三层次比较相近,但在如下方面有所不同:将软件工程课程作为专业教学重点;更加强调英语教学,更加重视实践能力培养,并对两者有更高的要求。

我们本着对高等院校的计算机专业状况的认识,主要面向与上述第二、第三两个层次对应的院校及与之相近的软件学院,总结多年的计算机专业的教改经验,在一定程度上溶入了 ACM & IEEE CC2001 和 CCC2002(中国计算机科学与技术学科教程)的教改思路,组织我国一直投身于计算机教学和科研的教师,编写了这套“高等院校计算机专业教育改革推荐教材”(以下简称“推荐教材”)。自然,“推荐教材”中所贯穿的改革思路和做法,也是针对上述第二、第三两个层次对应院校的计算机专业学生。这些思路和做法可概括成以下三句话:

- 适度调整电子技术基础、计算机理论基础和系统软件的教学内容。
- 全面强化计算机工具软件、应用软件的教学要求。
- 以应用为目标大力展开软件工程的教学与实践。

电子技术基础、计算机理论基础、系统软件教学关系到学生的基本素质、发展潜力和日后

的应变能力。“推荐教材”在调整它们的教学内容时的做法是:适度压缩电子线路、数字电路和信号系统的教学内容,变三门课程为两门,并插入数字信号处理的基础内容;合并“计算机组成原理”、“微型计算机接口技术”和“汇编语言”为“计算机硬件技术基础”一门课程;注意适当放宽“离散数学”课程的知识面,使之与 CCC2002 的要求基本接轨,但适度降低其深度要求;更新系统软件课程的教学内容,以开放代码的 Linux 作为操作系统原理的讲授载体,更加关注系统软件的实践性和实用性。

为了提高计算机专业人才的计算机应用能力,全面强化计算机工具软件、实用软件的教学要求是十分重要的,这也是上述改革思路的核心。为此,“系列教材”的做法是:强化程序设计技术,强化人机接口技术,强化网络应用技术。

为强化程序设计技术,“推荐教材”支持在单片机环境、微机平台、网络平台的编程训练;支持运用程序设计语言、程序设计工具以及分布式对象技术的编程训练。大大加强面向对象程序设计课程的组合(设计了三门课程:面向对象的程序设计语言 C++,面向对象的程序设计语言 JAVA 和分布式对象技术),方便教师和读者的选择。

为强化人机接口技术,“推荐教材”设计了“人机交互教程”,“计算机图形学”和“多媒体应用技术”等可供选择的、有层次特色的课程组合。

为强化网络应用技术,“推荐教材”设计了“计算机网络技术”,“计算机网络程序设计”,“计算机网络实验教程”和“因特网技术及其应用”等可供选择的、新颖丰富的课程组合。

将软件工程课程作为专业教学重点,以应用为目标大力展开软件工程的教学与实践,是“推荐教材”改革思路的又一亮点。为改变以往软件工程课程纸上谈兵的老毛病,“推荐教材”从工程应用出发,理论联系实际,突出建模语言及其实现工具的运用,设计了“软件工程的方法与实践”,“统一建模语言 UML 导论”和“ROSE 对象建模方法与技术”等可供选择的、创新独特的软件工程课程组合。对于各类软件学院,“推荐教材”的这一特色无疑是很有吸引力的。

强调实践也是计算机学科永恒的主题,对计算机应用专业的学生来说更是如此。重应用和重实践是“推荐教材”的一个整体特点。这一特点,一方面有利于解决本文开始所指出的计算机专业学生较之非计算机专业学生,在应用开发工作中上手慢的问题;另一方面,使计算机专业的学生能在更大范围内、更高层面上掌握计算机应用技术。这一特点正是许多高等院校计算机专业教育改革追求的一个目标,也是国家教育部倡导软件学院的初衷之一。

“推荐教材”由基础知识、程序设计、应用技术、软件工程和实践环节等五个模块组成。各模块有其对应的培养目标与功能,从而构架出一个创新的、完整的计算机应用专业的课程体系。模块化的设计,使各学校可根据学生及学校的特点做自由的选择和组合,既能达到本专业的总体要求,又能体现具有特色的个性发展。整套教材的改革脉络清晰,结构特色鲜明,值得各高等院校在改革教学内容、编制教学计划、挑选教材书目时借鉴和参考。当然,很多书目也适合很多相关学科的计算机课程用作教材。

“推荐教材”的组成模块和书目详见封底。显然它不能说是完备的(实践环节模块更是如此),其改革的思路、改革的举措也可能有值得探讨的地方。我们衷心希望得到计算机教育界同仁和广大读者的批评指正。

高等院校计算机专业教育改革推荐教材

编委会

前 言

数据库应用广泛,其理论和技术是计算机及其相关专业大学生的必修内容,也是计算机工程技术人员必备的知识和技能。为适应国家教育部推出的“高等教育面向 21 世纪教学内容和课程体系改革计划”,满足高等院校学生素质教育的需要,作者根据多年的教学实践和数据库系统的开发经验,编写了本教材。

本书覆盖了数据库的基本理论和知识、数据库管理系统的实现原理,以及数据库设计开发方法,有助于读者全面理解和掌握数据库的应用技术;内容选择注重理论的基础性和技术的实用性;在实践和技术的相关章节中,对应用技术的阐述,既结合具体产品,切实可行,又注意不失同类产品和技术的普遍性,具有广泛的指导作用。全书共分为 3 篇:

第 1 篇:数据库理论知识篇,包括 4 章。

第 1 章 介绍数据库的基本概念、数据库系统结构和数据库系统。

第 2 章 介绍了层次、网状、关系、对象和对象关系等数据模型。

第 3 章 介绍关系数据库理论,包括关系代数和关系演算。

第 4 章 介绍关系数据库标准语言 SQL,含数据定义、查询、更新、视图等操作。

第 2 篇:数据库管理系统篇,包括 3 章。

第 5 章 介绍了数据库管理系统的功能、结构和执行过程,对数据字典和分布式数据库管理也进行了阐述。

第 6 章 介绍了数据库存储管理、数据组织管理和数据查询的实现技术。

第 7 章 介绍了数据库中事务管理、安全管理和完整性机制。

第 3 篇:数据库应用技术篇,包括 4 章。

第 8 章 介绍了关系模式规范化设计理论、方法和应用。

第 9 章 介绍了数据库设计的工程化方法和技术,覆盖了需求分析、概念设计、逻辑设计、物理设计、数据库实施、维护和性能测评。

第 10 章 介绍了数据库应用系统的层次结构、数据库与应用系统的连接和网络环境下数据库应用系统的新技术。

第 11 章 介绍了数据库技术和应用的新发展,包括对象数据库、对象关系数据库、主存数据库、工程数据库、移动数据库、数据挖掘、数据仓库和决策支持系统等。

本书内容丰富、通俗易懂,全书以 SQL Server 2000 为背景,并附有习题和上机实验,能帮助读者在掌握理论知识的同时,提高应用能力。

全书由黄志球负责内容的组织、统稿和审定。第 1~5、8 章由李清执笔,第 6、10、11 章由柳学涛、陶剑青执笔,第 7 章由张定会执笔,第 9 章由沈国华执笔。杨涛参加了图形表格的编辑和上机实验的调试工作。王元元教授为本书的出版做了大量的工作,在此表示感谢。

限于时间和水平,书中错误之处在所难免,敬请广大读者和专家批评指正。

编 者

目 录

编者的话

前言

第 1 篇 数据库理论知识篇

第 1 章 数据库引论	1
1.1 数据库技术的发展历史	1
1.1.1 数据处理的发展	1
1.1.2 数据管理的发展	2
1.1.3 数据库技术	4
1.2 数据库系统结构	6
1.2.1 数据库的三级结构	6
1.2.2 模式、存储模式与子模式	7
1.2.3 数据库及其特点	8
1.3 数据库系统	10
1.3.1 数据库系统的构成	10
1.3.2 数据库管理系统	12
1.3.3 数据操纵接口——数据库语言	13
1.3.4 数据库存取的执行过程	15
1.4 小结	15
1.5 习题	16
第 2 章 数据模型	17
2.1 现实世界到计算机的映射	17
2.1.1 三个世界	17
2.1.2 两种模型	20
2.2 概念模型	21
2.2.1 基本元素	21
2.2.2 连接	22
2.2.3 E-R 图	24
2.3 层次模型	25
2.3.1 基本概念和结构	26
2.3.2 操作	29
2.3.3 约束	30
2.4 网状模型	31
2.4.1 基本概念和结构	31

2.4.2 操作	33
2.4.3 约束	34
2.5 关系模型	34
2.6 其他数据模型	36
2.6.1 面向对象的数据模型	36
2.6.2 对象关系数据模型	39
2.7 小结	40
2.8 习题	41
第3章 关系数据库理论	43
3.1 关系模型概述	43
3.1.1 关系的数学描述	43
3.1.2 基本术语	43
3.2 关系代数	45
3.2.1 五种基本关系代数运算	45
3.2.2 四种组合关系代数运算	47
3.2.3 七种扩充关系代数运算	49
3.3 关系演算	51
3.3.1 关系的表示	52
3.3.2 元组关系演算	52
3.3.3 域关系演算	55
3.4 关系模型总结	57
3.4.1 关系模型的三要素	57
3.4.2 关系模型的完整性约束	57
3.4.3 关系模型的特点	58
3.5 小结	58
3.6 习题	59
第4章 关系数据库标准查询语言 SQL	62
4.1 SQL 语言概述	62
4.1.1 SQL 语言的基本组成	62
4.1.2 SQL 语言的特点	63
4.1.3 SQL 语言的数据类型	64
4.2 SQL 中的数据定义	64
4.2.1 模式的定义、修改和撤销	65
4.2.2 索引的建立和撤销	67
4.3 SQL 中的数据查询	68
4.3.1 SQL 的单表查询	69
4.3.2 SQL 的计算与统计函数	71
4.3.3 SQL 的查询条件	71
4.3.4 SQL 查询结果的聚集与排序	73

4.3.5 SQL 的连接、嵌套和集合查询	74
4.4 SQL 中的数据更新	77
4.4.1 SQL 插入数据语句	77
4.4.2 SQL 修改数据语句	78
4.4.3 SQL 删除数据语句	79
4.5 SQL 中的视图	80
4.5.1 视图的定义、查询和更新	80
4.5.2 视图的用途	82
4.6 SQL 中的特殊操作	84
4.6.1 空值	84
4.6.2 外部连接	87
4.6.3 递归查询	89
4.7 小结	91
4.8 习题	92

第 2 篇 数据库管理系统篇

第 5 章 数据库管理系统	94
5.1 数据库管理系统概述	94
5.1.1 DBMS 的功能	94
5.1.2 常用 DBMS	96
5.1.3 SQL Server 2000 的功能特点	98
5.2 数据库管理系统结构	99
5.2.1 DBMS 的系统结构	99
5.2.2 DBMS 的解释执行过程	102
5.2.3 SQL Server 2000 的结构	103
5.3 数据字典	105
5.3.1 数据字典在 DBMS 中的作用	105
5.3.2 数据字典的内容	106
5.3.3 数据字典与 DBMS	107
5.3.4 数据字典的使用	108
5.3.5 SQL Server 的数据字典	109
5.4 分布式数据库管理系统	110
5.4.1 分布式数据库系统及其结构	110
5.4.2 数据分布策略	113
5.5 小结	114
5.6 习题	115
第 6 章 数据存储与查询技术	117
6.1 数据库基本元素	117
6.2 数据库存储管理	117

6.2.1 存储介质	118
6.2.2 基本元素存储	119
6.2.3 数据压缩技术	124
6.3 数据组织管理	125
6.3.1 文件结构与存取路径	125
6.3.2 索引技术	126
6.3.3 散列技术	131
6.4 数据查询技术	131
6.4.1 数据库系统查询的实现	132
6.4.2 查询编译器	132
6.4.3 查询执行	135
6.4.4 查询优化	135
6.5 SQL Server 的数据查询技术	136
6.5.1 SQL Server 服务器系统结构	136
6.5.2 SQL Server 查询技术	137
6.5.3 SQL Server 查询优化技术	137
6.6 小结	139
6.7 习题	139
第7章 事务、安全与完整性约束	141
7.1 事务管理	141
7.1.1 事务	141
7.1.2 事务的恢复	145
7.1.3 事务的并发控制	148
7.1.4 分布式数据库的并发控制与恢复	150
7.1.5 SQL Server 的事务管理	153
7.2 安全管理	156
7.2.1 数据库安全控制的方法	156
7.2.2 SQL Server 的数据库的安全管理	159
7.3 完整性约束	160
7.3.1 完整性约束的形态	160
7.3.2 SQL 中的完整性约束	161
7.3.3 SQL Server 的完整性约束机制	164
7.4 小结	165
7.5 习题	166

第3篇 数据库应用技术篇

第8章 关系模式的规范化设计	168
8.1 关系模式的设计问题	168
8.2 规范化理论	169

8.2.1	函数依赖	169
8.2.2	基于函数依赖的范式	173
8.2.3	多值依赖与第四范式	175
8.2.4	连接依赖与第五范式	177
8.3	关系模式的分解	178
8.3.1	模式分解	179
8.3.2	无损连接分解和保持函数依赖	179
8.3.3	模式分解算法	182
8.4	规范化理论应用	186
8.4.1	规范化的优缺点	186
8.4.2	反规范化	188
8.4.3	规范化示例	188
8.5	小结	190
8.6	习题	191
第9章	数据库设计	194
9.1	数据库设计概述	194
9.1.1	数据库设计的内容	194
9.1.2	数据库设计的方法	194
9.1.3	数据库设计的步骤和特点	195
9.2	需求分析	198
9.3	概念设计	198
9.3.1	概念模型设计的方法与步骤	198
9.3.2	数据抽象与局部视图设计	200
9.3.3	视图的集成	201
9.4	逻辑设计	205
9.4.1	概念模型向数据模型的转换	205
9.4.2	设计用户模式	208
9.4.3	模型的评价与优化	208
9.5	物理设计	209
9.5.1	物理设计的要求和内容	209
9.5.2	创建索引	211
9.5.3	性能评价	211
9.6	数据库的实施与维护	211
9.6.1	数据库数据的装入	211
9.6.2	数据库的运行和维护	212
9.6.3	DBA 的组织	212
9.6.4	数据库的性能评价与监视	214
9.6.5	日志与备份恢复	215
9.7	小结	215

9.8 习题	216
第 10 章 数据库与应用系统	218
10.1 应用系统的层次结构	218
10.1.1 单机应用系统	218
10.1.2 C/S 应用系统	219
10.1.3 三层体系结构	221
10.1.4 B/S 应用系统	222
10.2 数据库与应用系统	223
10.2.1 宿主系统与自含系统	223
10.2.2 嵌入式 SQL 语句	224
10.2.3 数据库访问中间件	227
10.2.4 数据库应用系统开发工具	232
10.3 网络环境下数据库系统的新技术	234
10.3.1 XML	235
10.3.2 数据复制技术	237
10.3.3 网络数据库	239
10.3.4 分布环境下的信息集成技术	242
10.4 小结	244
10.5 习题	245
第 11 章 数据库技术和应用的新发展	246
11.1 对象数据库和对象关系数据库	246
11.1.1 对象数据库	247
11.1.2 对象关系数据库	248
11.2 主存数据库	250
11.3 其他数据库	250
11.3.1 工程数据库/空间数据库	250
11.3.2 移动数据库	251
11.4 数据库在应用系统中的新应用	252
11.4.1 决策支持系统	252
11.4.2 数据分析	253
11.4.3 数据挖掘	254
11.4.4 数据仓库系统	254
11.5 小结	256
11.6 习题	257
附录 A	258
A.1 实验 1 了解 SQL Server 2000 的运行环境	258
A.2 实验 2 SQL Server 的注册	260
A.3 实验 3 数据库的创建和删除	261
A.4 实验 4 用 SQL 语句创建数据库	264

A.5	实验 5	表的创建和删除	267
A.6	实验 6	索引的创建	267
A.7	实验 7	简单 SELECT 语句的使用	270
A.8	实验 8	集合函数与行集函数	273
A.9	实验 9	带子句的 SELECT 语句	275
A.10	实验 10	子查询	279
A.11	实验 11	面向用户的视图	281
A.12	实验 12	许可管理	283
A.13	实验 13	SQL Server 的监控	286
A.14	实验 14	数据库的维护	288
A.15	实验 15	数据库的备份	290
A.16	实验 16	数据库的恢复	291
A.17	实验 17	创建维护计划	293
参考文献			297

第 1 篇 数据库理论知识篇

第 1 章 数据库引论

随着计算机软、硬件的发展,计算机应用不再局限于以往单一的数值计算领域,而是越来越多地应用于非数值计算的领域。从纯科学计算进入非数值的数据处理是一个划时代的转折,使计算机成为广大科技人员和管理人员工作中的得力助手和有力工具。

数据库技术研究如何科学地组织和存储数据,如何高效地获取和处理数据,是数据管理的最新技术。数据库系统是管理信息系统、办公自动化系统和决策支持系统等诸多应用系统的核心部分,在当代计算机系统中占有重要的地位。

1.1 数据库技术的发展历史

1.1.1 数据处理的发展

数据是人类社会赖以生存和发展的一项重要资源,人们的社会生活和日常生活都离不开数据。数据处理是指对各种形式的数据进行收集、整理、存储、加工和传播等一系列工作的总称。其目的是为了从大量的、原始的数据中提取、推导出有价值的信息,用来作为人们行动和决策的依据。随着社会生产和文明的日益发展,数据处理在人类的整个社会生活中起着越来越重要的作用,数据处理技术也不断地从低级向高级发展,它大致可分为三个阶段:手工处理、机械处理和电子处理。

1) 手工处理阶段。从原始社会到 19 世纪末,由于社会生产力和科学技术水平的限制,数据处理处于手工处理阶段。在这一阶段中,人们从最原始的用石块和木棍等来计数到后来发明了算盘、基于齿轮的六位加法器、计算尺、微分机以及对数和二进制系统等,虽然在计算工具和计算方法上取得了一系列的成就,但总的说来,这一时期用于数据处理的计算工具精确度差、处理能力低,数据处理处于较低级的手工操作阶段,效率很低。

2) 机械处理阶段。这一阶段最显著的标志便是将卡片制表机用于数据处理。1890 年,美国中央统计局的 H. Holler 为了编制人口统计表的需要,发明了卡片制表机,该卡片制表机由穿孔机、验孔机、分类机、卡片整理机、复孔机和制表机等几部分组成,能以半自动方式进行卡片的穿孔、校验、分类、整理和制表等工作,与手工处理阶段相比,其中部分手工操作被机械取代,大大提高了数据处理的能力和效率,使数据处理跨入了机械处理阶段,是数据处理的一个重大突破,但处理效率仍然受到机械设备的限制。

3) 电子处理阶段。1946 年第一台电子计算机 ENIAC 的诞生,使数据处理进入了一个崭新的阶段——电子处理阶段。电子计算机以其自动、快速的处理,大容量的数据存储,方便灵

活的输入和输出,彻底改变了数据处理手工操作多,效率低下,可靠性差这些与社会生产力发展不相适应的落后状态,使人们真正从大量手工操作的数据处理中解放出来。20 世纪 60 年代以来,社会生产力高速发展,新技术层出不穷,计算机软、硬件不断发展,特别是大容量存储器的生产和使用,使计算机不仅能进行一般的数据处理,而且还能对图片、图像、声音等复杂的多媒体数据进行处理,这时期的数据处理技术得到了突飞猛进的发展。

1.1.2 数据管理的发展

数据处理的中心问题是数据管理。数据管理指的是对数据的分类、组织、编码、存储、检索和维护。数据管理技术和数据处理方式有着密切的联系,直接影响着数据处理的效率。在数据处理的手工阶段和机械阶段,其数据管理也是手工的。随着数据处理进入电子处理阶段,数据管理也随着计算机硬件和软件的发展而不断发展,经历了人工管理、文件系统和数据库系统三个阶段。

1) 人工管理阶段。20 世纪 50 年代末以前,计算机主要用于科学计算。从硬件看,外存只有磁带、卡片和纸带,还没有磁盘等直接存取的存储设备;从软件看,没有操作系统,没有统一的数据管理软件,数据处理是批处理方式。这个阶段的数据管理具有如下特点:

① 不保存数据。由于此阶段计算机主要用于科学计算,因此只是在计算时将所需数据输入,用完并不保存,不仅对用户数据如此,对系统软件所需的数据也如此。

② 无统一的数据管理软件。对数据的管理完全由各个程序员进行。程序员在编制程序时,不仅要规定数据的逻辑结构,而且还必须设计数据的物理结构,包括数据的逻辑定义和组织,数据存放的存储设备,存取方法,输入输出方式等,表示处理流程的程序与其所处理的数据相互结合成一个整体,数据的存储位置发生改变时,相应程序也必须进行修改,程序和数据不具有独立性。

③ 没有文件的概念。数据的组织完全由程序员设计。

④ 程序和数据一一对应。每一组数据对应一个应用程序,即使两个应用程序所需处理的数据相同,也必须各自定义和使用,数据无法共享,数据是面向应用的。

这一阶段的数据管理是分散的,没有充分发挥计算机在数据管理中的作用,严重影响了计算机的使用效率,如图 1-1 所示。

2) 文件系统阶段。20 世纪 50 年代末至 60 年代中期,计算机开始大量用于数据管理,硬件方面出现了磁盘、磁鼓等直接存取的存储设备;软件方面,操作系统中有了多道程序、分时系统,出现了专门的管理数据的软件,这就是文件系统。这个阶段对数据不仅能进行文件批处理,还能联机实时处理。在文件系统中,数据按其内容、结构和用途组成若干命名的文件。文件一般为某一应用程序或一组应用程序所有,但可供指定的其他应用程序共享。应用程序通过文件管理系统进行数据文件的建立、打开、读、写、关闭和撤销等操作。以文件形式来进行数据管理使得计算机在数据管理中直接发挥作用,是数据管理技术的重大发展。虽然文件系统允许多个应用程序共享一组数据,但实际上,由于数据文件是各应用程序根据其自身的需要而建立的,因此这些文件中的数据是很难为其他应用程序所用的,文件系统所管理的数据基本上还是分散的、相互独立的数据文件,如图 1-2 所示。以此为基础的数据处理存在以下缺点:

① 编写应用程序不方便。应用程序的编写者必须对所用文件的逻辑和物理结构有一定的了解,文件系统只提供打开、关闭、读、写等几个低级的文件操作,对文件的查询、修改等处理

都必须由应用程序完成,若有多个应用程序有查询处理,则它们都必须具有查询功能,不可避免地造成功能上的重复;若有新的要求,则必须由熟悉文件的程序员再来编写应用程序,非常低效。

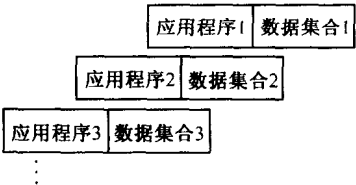


图 1-1 人工管理阶段的数据处理

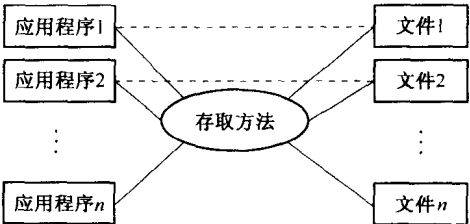


图 1-2 文件系统阶段的数据处理

② 数据冗余度大。一个数据文件通常只为某个特定的应用程序服务,不同的应用程序使用相互独立的数据文件,因此相同的数据会同时出现在几个数据文件中,造成数据的冗余,有时甚至是大量的数据冗余。例如,职工的姓名、所在部门、工资等数据会同时出现在职工的人事档案文件和工资文件中;学生的姓名、所在系也会同时出现在学生档案文件和成绩文件中。数据冗余不仅浪费存储空间,而且会带来数据的不一致。在文件系统中,无维护数据一致性的监控机制,数据的一致性完全由应用程序完成,给数据的修改和维护带来了困难。

③ 程序和数据的相互依赖,简称数据依赖。在文件管理方式下,数据文件由使用它的应用程序建立,一旦应用程序发生改变,原来的数据文件也必须作相应的改变,否则就无法使用。反之,若原来的数据文件,其数据的存储结构、存储方式,存储内容发生改变,则对应的应用程序也要作相应修改,否则,应用程序就无法正常运行。这表明数据文件 and 应用程序是相互依赖的,或者说文件系统的独立性不好。程序与数据的相互依赖使得建立在文件系统上的各个应用系统难以修改、扩充。

④ 不支持对文件的并发访问。现代计算机系统中,为了有效地利用计算机资源,一般允许多个应用程序并发地运行。文件系统一般不支持多个应用程序对同一文件的并发访问。为解决多个应用程序同时访问同一个数据文件而引起的错误,虽然文件系统可限制文件的并发操作,或为每个应用程序增加一个文件副本,但这样做,前者降低程序的并发度,后者增加数据冗余,会引起数据的不一致。

⑤ 缺少统一的数据管理。文件系统在数据的结构、编码、表示格式、命名以及输入输出格式等方面不易形成规范化和标准化,难以采取有效的措施保障数据的安全和保密。

这些缺点在规模较大的应用系统中尤其明显。针对这些缺点,计算机软件工作者经过长期不懈的努力,得出了数据库的概念。

3) 数据库系统阶段。20 世纪 60 年代后期以来,社会生产力高速发展,新技术层出不穷,信息量急剧膨胀,使整个人类社会成为信息化的社会,人们对信息和数据的处理已进入自动化、网络化和社会化的阶段。诸如管理信息系统、办公信息系统、民航订票系统、情报检索系统、智能决策系统等等都要求对数据进行快速地处理,及时地得出结果。管理这种大量的、持久的、共享的数据是计算机所面临的共同问题。数据库技术为数据管理提供了一种较完善的高级管理方式,它克服了文件系统方式下分散管理的缺点,对所有的数据实行统一、集中的管理,使数据的存储独立于使用它的程序,从而实现数据共享。硬件价格的大幅度下降为数据库