

21世纪高等学校计算机**基础**实用规划教材

# 数据库原理与实践

陆黎明 王玉善 陈军华 编著

清华大学出版社



21世纪高等学校计算机**基础**实用规划教材

# 数据库原理与实践

陆黎明 王玉善 陈军华 编著

清华大学出版社  
北京

## 内 容 简 介

本书全面地阐述了数据库系统的基本概念、基本原理、基本方法和基本技术,内容包括数据库的基本概念、数据模型、关系数据库、结构化查询语言 SQL、完整性约束的实现、索引与视图、数据库安全技术、事务及其特性、并发控制、数据库恢复技术、关系数据库设计理论、数据库设计和实验指导等。

本书注重实践,强调理论联系实际,在有关章节介绍了 SQL Server 2005 中的索引、视图、安全性、事务、并发控制和恢复技术,并用单独一章来介绍 SQL Server 2005 中的函数、游标、存储过程和触发器,最后配有与教学同步的实验指导;另外,在第 9 章以学校管理信息系统为例完整地介绍了数据库设计的全过程,所有这些都利于读者对数据库系统基本原理、基本方法和基本技术的理解和掌握。

本书结构合理、概念清晰、图文并茂并且例题丰富,关注新概念和新技术,配套的练习题不但量多而且题型丰富,有利于读者对数据库原理的理解和掌握。本书适合作为高等学校计算机、软件工程、电子商务、信息与计算科学和地理信息系统等相关专业本科生数据库课程的教材,也可作为电子信息类专业研究生数据库课程的教材,还可作为计算机等级考试(三级数据库技术)的参考书和自学教材。

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。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

### 图书在版编目(CIP)数据

数据库原理与实践/陆黎明,王玉善,陈军华编著.--北京:清华大学出版社,2016

21 世纪高等学校计算机基础实用规划教材

ISBN 978-7-302-41674-6

I. ①数… II. ①陆… ②王… ③陈… III. ①关系数据库系统 IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2015)第 237809 号

责任编辑:黄 芝 薛 阳

封面设计:常雪影

责任校对:时翠兰

责任印制:何 芊

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:保定市中画美凯印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:16.5

字 数:400 千字

版 次:2016 年 1 月第 1 版

印 次:2016 年 1 月第 1 次印刷

印 数:1~2000

定 价:29.50 元

产品编号:065819-01

# 出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程(简称‘质量工程’)”,通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

本系列教材立足于计算机公共课程领域,以公共基础课为主、专业基础课为辅,横向满足高校多层次教学的需要。在规划过程中体现了如下一些基本原则和特点。

(1) 面向多层次、多学科专业,强调计算机在各专业中的应用。教材内容坚持基本理论适度,反映各层次对基本理论和原理的需求,同时加强实践和应用环节。

(2) 反映教学需要,促进教学发展。教材要适应多样化的教学需要,正确把握教学内容和课程体系的改革方向,在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养,为学生的知识、能力、素质协调发展创造条件。

(3) 实施精品战略,突出重点,保证质量。规划教材把重点放在公共基础课和专业基础课的教材建设上;特别注意选择并安排一部分原来基础比较好的优秀教材或讲义修订再版,逐步形成精品教材;提倡并鼓励编写体现教学质量和教学改革成果的教材。

(4) 主张一纲多本,合理配套。基础课和专业基础课教材配套,同一门课程可以有针对不同层次、面向不同专业的多本具有各自内容特点的教材。处理好教材统一性与多样化,基本教材与辅助教材、教学参考书,文字教材与软件教材的关系,实现教材系列资源配套。

(5) 依靠专家,择优选用。在制定教材规划时依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主题。书稿完成后要认真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平教材编写梯队才能保证教材的编写质量和建设力度,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

### 21 世纪高等学校计算机基础实用规划教材

联系人:魏江江 weijj@tup.tsinghua.edu.cn

# 前 言

数据库技术始于 20 世纪 60 年代末,经过 40 多年的发展,已经形成了完整的理论与技术体系,成为计算机科学与技术学科中的一个重要分支。数据库技术已经广泛地应用于国民经济、科教文化和社会生活等各个领域,成为计算机信息系统的核心技术和重要基础,也是各相关领域专业技术人员和管理人员的必备知识和基本技能。

数据库原理是高等学校计算机专业的核心专业课程,也是软件工程、电子商务、信息与计算科学和地理信息系统等专业的专业课程。通过本课程的学习,不仅要求学生掌握数据库系统的基本概念、基本原理、基本方法和基本技术,更重要的是要培养学生对这些原理、方法和技术的应用能力和问题的求解能力,为今后从事信息系统的开发和管理打下良好的基础。本教材的作者长期从事数据库原理的教学与科研工作,教材的内容选择和组织结构凝聚了作者近 30 年数据库课程教学工作的实践经验。本教材具有以下几个特点。

(1) 在内容选择上,删去了一些过于理论或过于专业或应用性不强的知识,如关系演算、数据库管理系统实现技术、对象关系数据库系统等。增加了一些对新概念和新技术的介绍,如关系代数中的聚集运算,SQL 语言中用 Natural 或 Using 表达连接条件、WITH 子句、物化视图、事务的隔离级别,E-R 图中的聚集概念等。

(2) 强调理论联系实际,注重实践操作。选择主流的商用关系型数据库管理系统 SQL Server 作为实践平台,在介绍数据库系统的基本方法和基本技术的同时,穿插介绍 SQL Server 中的相应方法和技术,如 SQL Server 中数据库和基本表的创建及管理、索引和视图的创建及应用、数据库安全技术、并发控制技术以及备份和恢复技术。第 6 章系统地介绍 SQL Server 2005 中函数、游标、存储过程和触发器的应用。第 9 章以“学校管理信息系统”为例完整地介绍了数据库设计的全过程。附录中的实验指导与教学同步,实验内容有针对性、启发性和综合设计性,通过实际操作有利于学生真正掌握数据库系统的基本方法和基本技术。

(3) 数据库系统的概念众多,原理比较抽象,不易理解。作者精心选择和编写了数量多而且题型丰富的练习题,这些练习题不但与教学同步,而且覆盖不同的知识点,相互之间不重复。通过这些练习题的解答有利于学生对数据库原理的理解、掌握和复习。

本书的适应性十分广泛,从初学者到有一定基础的读者,从师生到专业技术人员,具体来说有以下 4 点。

(1) 对于初学者可以从本书中了解到数据库技术的基本概念、基本方法和基本技术,即数据库基本原理。

(2) 对于有一定基础的读者可以从本书中了解到有关数据库的一些新概念和新技术。

(3) 对于教师和学生,本书除了可以作为教材外,也是一本很好的教学参考书。第 9 章

中的数据库设计实例、附录中的实验指导以及书中的练习题,都是很好的教学参考资料。

(4) 对于了解数据库原理的专业技术人员,可从本书中了解到主流的商用关系型数据库管理系统 SQL Server 的基本使用方法,从而在实践中掌握数据库技术。

## IV

本书的例题全部在 SQL Server 2005 环境下调试通过,当然也可以在 SQL Server 2008 等环境下调试这些例题,读者可根据自己的具体情况选用。本书在写作过程中得到了上海师范大学副校长高建华教授和上海市高等学校计算机等级考试专家上海电力学院顾春华教授的悉心指导。初稿完成后常州大学信息学院的马正华教授进行了认真的审阅,提出了许多宝贵意见。另外,清华大学出版社的编辑进行了认真负责的编辑工作,保证了本书的顺利出版。在此一并向他们表示衷心的感谢。

本书结构合理、概念清晰、图文并茂、例题丰富,适合作为高等学校计算机、软件工程、电子商务、信息与计算科学、地理信息系统等相关专业本科生数据库课程的教材,也可作为电子信息类专业研究生数据库课程的教材,还可作为计算机等级考试(三级数据库技术)的参考书和自学教材。

由于编者水平有限,虽然力求精准,但疏漏与不足之处仍在所难免,敬请专家和读者指正。

编者

2015年6月

# 目 录

第 1 章 概述	1
1.1 数据库技术的产生与发展	1
1.1.1 数据处理和数据管理	1
1.1.2 数据管理技术的发展	2
1.1.3 数据库技术的特点和展望	5
1.2 数据库系统的组成	6
1.2.1 数据库	7
1.2.2 数据库管理系统等软件	7
1.2.3 数据库管理员	8
1.3 数据库的体系结构	8
1.3.1 模式的概念	8
1.3.2 三级模式结构	9
1.3.3 二级映像和二级独立性	11
习题 1	12
第 2 章 关系数据库	14
2.1 数据模型	14
2.1.1 数据模型的定义和基本要求	14
2.1.2 数据模型的三个层次	14
2.1.3 数据模型的三个组成要素	15
2.2 概念模型概述	16
2.2.1 信息世界中的基本概念	16
2.2.2 实体间的联系	16
2.2.3 概念模型的表示方法——E-R 图	17
2.3 逻辑数据模型概述	18
2.3.1 层次模型	19
2.3.2 网状模型	19
2.3.3 关系模型	20
2.4 关系模型	20
2.4.1 关系数据结构	21



2.4.2	关系的完整性 .....	24
2.4.3	关系操作的特点与分类 .....	26
2.5	关系代数 .....	26
2.5.1	传统的集合运算 .....	27
2.5.2	专门的关系运算 .....	28
2.5.3	扩展的关系代数运算 .....	32
2.5.4	关系代数运算的应用实例 .....	35
习题 2	.....	37
<b>第 3 章</b>	<b>结构化查询语言 SQL</b> .....	<b>41</b>
3.1	SQL 语言概述 .....	41
3.1.1	SQL 的产生与发展 .....	41
3.1.2	SQL 的功能与特点 .....	42
3.2	数据定义 .....	43
3.2.1	数据库的创建、修改与删除 .....	43
3.2.2	SQL 中的数据类型 .....	48
3.2.3	基本表的创建、修改与删除 .....	52
3.3	完整性约束的实现 .....	53
3.3.1	数据库完整性的概念 .....	53
3.3.2	各类完整性约束的实现 .....	54
3.3.3	表中完整性约束的增加与删除 .....	57
3.4	数据查询 .....	58
3.4.1	单表查询 .....	58
3.4.2	连接查询 .....	64
3.4.3	嵌套查询 .....	66
3.4.4	集合查询 .....	72
3.5	数据更新 .....	73
3.5.1	插入数据 .....	73
3.5.2	修改数据 .....	74
3.5.3	删除数据 .....	76
习题 3	.....	76
<b>第 4 章</b>	<b>索引与视图</b> .....	<b>81</b>
4.1	索引 .....	81
4.1.1	索引的概念 .....	81
4.1.2	创建索引的基本原则 .....	85
4.1.3	索引的创建与删除 .....	86
4.1.4	SQL Server 2005 中的索引 .....	87
4.2	视图 .....	90

4.2.1	视图的定义与删除 .....	90
4.2.2	查询视图 .....	92
4.2.3	更新视图 .....	93
4.2.4	视图的作用 .....	95
4.2.5	SQL Server 2005 中的索引视图 .....	95
习题 4	.....	96
<b>第 5 章</b>	<b>数据库安全技术 .....</b>	<b>99</b>
5.1	安全性控制技术概述 .....	99
5.1.1	计算机系统的三类安全性问题 .....	99
5.1.2	安全标准简介 .....	99
5.1.3	数据库安全性控制概述 .....	101
5.2	用户管理和角色管理 .....	103
5.2.1	用户管理 .....	103
5.2.2	角色管理 .....	105
5.3	权限管理 .....	106
5.3.1	授予权限 .....	106
5.3.2	收回权限 .....	107
5.4	SQL Server 的安全机制 .....	109
5.4.1	SQL Server 2005 的身份验证模式 .....	109
5.4.2	SQL Server 2005 的固定角色 .....	110
5.4.3	拒绝权限 .....	111
习题 5	.....	112
<b>第 6 章</b>	<b>函数、游标、存储过程和触发器 .....</b>	<b>114</b>
6.1	脚本、批和注释符 .....	114
6.1.1	脚本 .....	114
6.1.2	批 .....	114
6.1.3	注释符 .....	115
6.2	常量、变量和运算符 .....	115
6.2.1	常量 .....	115
6.2.2	变量 .....	115
6.2.3	运算符 .....	117
6.3	流程控制语句 .....	117
6.3.1	块语句 .....	117
6.3.2	分支语句 .....	118
6.3.3	循环语句 .....	119
6.3.4	其他控制语句 .....	120
6.4	函数 .....	121

6.4.1	系统函数	121
6.4.2	用户自定义函数	123
6.5	游标	127
6.5.1	定义游标	127
6.5.2	打开游标	129
6.5.3	通过游标读取数据	130
6.5.4	通过游标修改和删除数据	130
6.5.5	关闭游标和释放游标	131
6.6	存储过程	132
6.6.1	系统存储过程	133
6.6.2	用户自定义存储过程	133
6.6.3	使用存储过程实现封装业务逻辑	136
6.7	触发器	137
6.7.1	触发器概述	137
6.7.2	触发器的创建和管理	138
6.7.3	使用触发器实现强制业务规则	141
	习题 6	143
<b>第 7 章</b>	<b>事务管理</b>	<b>146</b>
7.1	事务	146
7.1.1	事务的定义	146
7.1.2	事务的 ACID 性质	146
7.1.3	事务的状态	147
7.1.4	SQL Server 2005 中的事务	148
7.2	并发控制	149
7.2.1	并发操作与数据的不一致性	150
7.2.2	可串行性	151
7.2.3	封锁及封锁协议	153
7.2.4	两段锁协议	156
7.2.5	锁的粒度	158
7.2.6	事务的隔离级别	160
7.2.7	SQL Server 2005 中的并发控制	161
7.3	数据库恢复技术	165
7.3.1	恢复的实现技术	165
7.3.2	数据库备份和登记日志方法	166
7.3.3	故障类型及恢复策略	167
7.3.4	具有检查点的恢复技术	169
7.3.5	SQL Server 2005 中的数据库备份与恢复	169
	习题 7	176

<b>第 8 章 关系数据库设计理论</b> .....	180
8.1 关系模式规范化设计的必要性 .....	180
8.2 函数依赖与码 .....	181
8.2.1 函数依赖的定义及分类 .....	182
8.2.2 函数依赖的公理系统和推理规则 .....	183
8.2.3 属性集 $X$ 关于函数依赖集 $F$ 的闭包 .....	184
8.2.4 码 .....	185
8.2.5 候选码的快速求解方法 .....	185
8.3 关系模式的规范化 .....	186
8.3.1 第一范式 .....	187
8.3.2 第二范式 .....	187
8.3.3 第三范式 .....	188
8.3.4 BC 范式 .....	189
8.3.5 规范化小结 .....	190
8.4 关系模式的分解 .....	192
8.4.1 模式分解的概念 .....	192
8.4.2 分解的无损连接性和函数依赖保持性 .....	193
8.4.3 模式分解的算法 .....	196
习题 8 .....	198
<b>第 9 章 数据库设计</b> .....	201
9.1 数据库设计概述 .....	201
9.1.1 数据库设计在软件开发中的地位和作用 .....	201
9.1.2 数据库设计的基本步骤 .....	201
9.2 基于 E-R 图的概念结构设计 .....	203
9.2.1 实体与用户需求 .....	203
9.2.2 属性及其分类 .....	204
9.2.3 联系及其分类 .....	205
9.2.4 弱实体类型 .....	207
9.2.5 扩展的 E-R 特性 .....	208
9.2.6 E-R 图实例 .....	208
9.2.7 概念结构设计的方法和步骤 .....	209
9.3 逻辑结构设计 .....	210
9.3.1 强实体的表示 .....	210
9.3.2 联系的表示 .....	211
9.3.3 弱实体的表示 .....	212
9.3.4 聚集的表示 .....	212
9.3.5 逻辑结构设计的步骤 .....	213

9.4	E-R 模型设计问题 .....	213
9.5	物理结构设计 .....	215
9.6	数据库设计实例——学校管理信息系统 .....	217
9.6.1	系统概述 .....	217
9.6.2	学籍管理子系统的需求与概念设计 .....	217
9.6.3	行政管理子系统的需求与概念设计 .....	218
9.6.4	课表制定子系统的需求与概念设计 .....	219
9.6.5	子系统视图的集成 .....	222
9.6.6	逻辑结构设计 .....	224
习题 9	.....	228
<b>附录 A</b>	<b>实验指导</b> .....	<b>230</b>
实验 1	创建和管理数据库和数据表 .....	231
实验 2	实现数据完整性 .....	232
实验 3	查询数据 .....	233
实验 4	视图的定义与使用 .....	237
实验 5	数据库安全技术 .....	238
实验 6	存储过程与触发器 .....	239
实验 7	并发控制 .....	240
实验 8	数据库恢复技术 .....	243
实验 9	索引与查询优化 .....	244
实验 10	导入导出及数据库的分离与附加 .....	247
<b>参考文献</b>	.....	<b>249</b>

# 第 1 章

# 概 述

当今社会是一个信息社会,信息正以惊人的速度增长,信息资源是各个部门的重要财富和资源。建立一个行之有效的信息系统成为一个企业或组织生存和发展的重要条件,信息化程度的高低也成为衡量一个国家现代化程度的重要指标。数据库技术诞生于 20 世纪 60 年代末,它的出现使得计算机应用渗透到了人类社会的每个角落,并正改变着人们的工作方式和生活方式。电子商务系统、计算机集成制造系统、办公自动化系统、地球信息系统和决策支持系统等都使用了数据库技术,数据库技术作为信息系统的基础和核心得到了越来越广泛的应用。现在,数据库技术有了比较完整的理论体系和实用技术,已成为计算机科学中最重要的技术之一。

本章介绍数据库技术中的数据管理、数据库、数据库管理系统、数据库系统以及数据库的体系结构等基本概念。

## 1.1 数据库技术的产生与发展

数据库技术的发展与应用是人类信息处理活动的客观要求,它极大地提高了信息处理的能力,已成为信息时代重要的特征之一。

### 1.1.1 数据处理和数据管理

#### 1. 信息与数据

信息(Information)已是现代社会中普遍使用的概念。一种通俗的解释为:信息是人们关心的事情的消息。例如,市场对某种商品的需求量,对于生产商和经销商来说,是很重要的消息,这就是信息。企业的产量、产值和利润等经济指标的统计数字,对企业管理者来说,也是信息。电视台播出的气象预报、股票指数、外币兑换率以及其他的经济新闻、科技新闻、文化新闻和体育新闻等,对于关心这些消息的个人或群体,都是信息。

信息从发生者通过传播信息的媒介被接收者接收,传播信息的媒介称为载体。即信息可以脱离发生者而借助于载体传输,并使接收者可以感知。这种反映信息内容,并可被接收者识别的符号(即载体)称为数据(Data)。数据的效用就在于它能够反映信息的内容并可被接收者识别。这里的数据不限于数字,文字、图形、图像、音频和视频等都是数据各种不同的形式,它们都可以经过数字化后存入计算机。

因此,数据是信息的具体表现形式,是信息的载体,而信息是数据的内涵,是对数据语义的解释。例如,7.8%是一个数据,它可以是一个公司某一年利润的增长率,也可以是一个股票某一天的涨幅,数据只有经过解释并赋予一定的语义后才能成为信息。

信息始于数据,数据被赋予语义而转换为信息。信息和数据“形影不离”,在不影响对问题理解的情况下,常把信息和数据这两个术语不加区分地使用,如信息处理和数据处理有时并没有严格的语义区别。但有时却必须加以区分,如数据文件不能说成信息文件。

## 2. 数据处理与数据管理

信息借助于数据可在一定条件下存储起来,存储的信息在适当的条件下可以进行传输。信息还可以通过一定的手段进行加工,如压缩、分解、综合、抽取和排序等。信息的可加工性为人类利用信息认识与改造客观世界和主观世界开辟了广阔的前景。人们可以从某些已知的信息(即数据)出发,加工推导出一些新的数据,而这些新的数据又表示了新的信息。通常把信息的收集、管理、加工、传播等一系列活动的总和称为信息处理或数据处理。

在用计算机进行数据处理时,相对来说,加工和计算比较简单,而数据的管理比较复杂。数据管理是指对数据进行分类、组织、编码、存储、检索和维护,这部分操作是数据处理的基本环节,而且是任何数据处理业务中必不可少的共有部分,是数据处理的中心问题。因此,数据处理是与数据管理相联系的,数据管理技术的优劣将直接影响数据处理的效率。显然,一个高效、通用、使用方便的数据管理软件,必将大大地减轻信息系统开发者的负担,极大地提高数据处理的能力。数据库技术就是专门研究数据管理的技术。

### 1.1.2 数据管理技术的发展

数据管理技术是应数据管理任务的需要而产生的,伴随着计算机硬件和软件的发展,数据管理技术经历了人工管理、文件系统和数据库系统三个发展阶段。数据库技术的出现是信息管理模式的大变革,提高了信息的利用率,加快了信息的传播,缩短了信息系统的开发过程。

#### 1. 人工管理阶段

这一阶段为 20 世纪 50 年代中期以前,这时的计算机主要用于科学计算,其他工作还没有开展。计算机的外部存储器只有纸带、磁带和卡片,没有磁盘等直接存储设备。软件还只有汇编语言,没有操作系统,更没有专门管理数据的软件。

科学计算一般不需要将数据长期保存,在进行某一计算任务时,原始数据随程序一起输入内存,计算结束将结果打印输出后,用户程序退出计算机系统,原始数据与程序所占的内存空间一起被释放。由于一组原始数据只对应一个程序(即某个计算任务),故数据是面向应用的。这种一组数据对应一个程序的应用模式会造成以下两个问题。

(1) 当多个应用程序涉及某些相同的数据时,无法互相利用和互相参照,必须各自定义,因此程序与程序之间有重复的数据,即数据不具有共享性。

(2) 由于没有专门的软件管理数据,数据需要由应用程序自己来管理(即由程序员来管理)。由于数据的物理结构是由应用程序自己设计的,程序中的存取子程序必须随着存储结构的改变而改变,即程序与相应的数据有着很强的依赖性。当数据的存储结构改变时,应用程序必须修改,因此程序与数据之间不具有独立性。

由于科学计算的原始数据一般不会非常庞大,故在这一阶段上述问题尚未被人们充分认识。

#### 2. 文件系统阶段

这一阶段为 20 世纪 50 年代后期至 20 世纪 60 年代中期,这时计算机不仅用于科学计

算,还用于数据管理。外部存储器已经有了磁盘、磁鼓等直接存储设备。软件领域出现了高级语言和操作系统,操作系统中的文件系统是专门管理外存的数据管理软件。下面通过实例说明这一阶段是如何管理数据的。

某高校为了管理日趋复杂的学生数据,先后在文件系统的基础上开发了学生选课管理系统、学生借书管理系统和学生医疗管理系统。与某一特定系统相关的所有数据都存放在一个特定的数据文件中,并由该系统对其进行管理(如图 1.1 所示,图中的虚线表示某系统以及相对应的数据)。从表面上看,该高校的学生管理工作做得井井有条,但随着时间的推移,各个系统管理的数据越来越多,这种各自独立建立在文件系统之上的数据管理方法暴露出来的问题也越来越多,主要有以下三个问题。

### 1) 数据共享性差,冗余度大

由于一个数据文件基本上只对应于一个程序,即文件仍然是面向应用的。当不同的应用程序有部分相同的数据(如本例中学生的学号、姓名、性别等)时,必须各自建立自己的数据文件,而不能共享相同的数据,从而造成大量的重复数据,这种在各个文件中的重复数据就是冗余数据。冗余数据不但浪费了存储空间,而且也容易造成数据的不一致性。数据的不一致性就是指某个文件中的学生数据改变了(如某学生有了新的家庭地址),而另外文件中相同的数据却未相应地改变。例如,借书管理系统已将他的家庭地址更新,而选课管理系统中仍然存放着旧的地址,从而导致教务处(即成绩管理部门)按旧地址发送的成绩单被退回。因此,数据冗余与数据的一致性产生了冲突,将学生的家庭地址放在一个文件中即可消除这种冲突。图 1.1 所示为文件系统阶段应用程序与数据之间的对应关系。

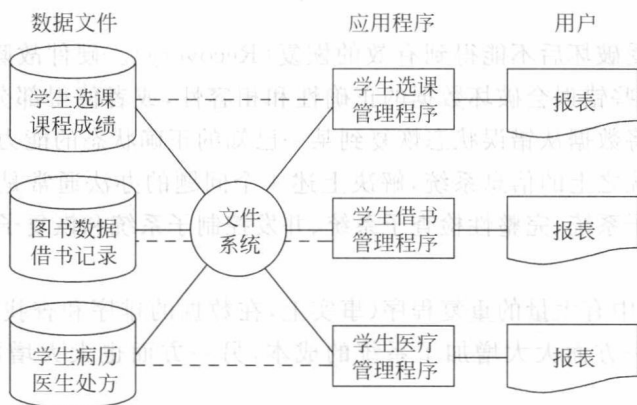


图 1.1 文件系统阶段应用程序与数据之间的对应关系

### 2) 程序与数据之间的独立性不高

由于在程序和数据之间由文件系统提供存取方法进行转换,使得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(例如,数据存放在哪个磁道和哪个扇区),而且存储结构上的改变不一定反映在程序上,这有效地节省了维护程序的工作量。但由于文件仍然是面向应用的,文件之间是独立的,为某个应用开发及组织的程序和数据与其他应用开发及组织的程序和数据可能不兼容(例如,选课管理系统中用一个英文字母来表示学生的性别,即用字母 M 或 F 表示男或女,而医疗管理系统中用一个汉字来表示学生的性别,即用男或女来表示),程序还依赖于对应数据的逻辑结构,程序的维护工作仍然是一项



艰巨且代价昂贵的任务。另外,当系统中要增加一个新的应用时,必须修改数据的逻辑结构,相应地也必须修改应用程序。因此程序与数据之间的独立性仍然不高。

### 3) 数据缺乏统一的管理和控制

这主要表现在以下 4 个方面。

第一是数据的安全性(Security)不能得到有效的保护。数据是有价值的,系统必须保护数据以防止不合法的用户(或者合法用户的非法使用)所造成的数据泄漏、更改或破坏。例如,在学生选课管理系统中,系统必须保证学生的考试成绩相应的任课老师或教务员有权修改,而一般人员(特别是学生本人)是不能修改的。

第二是数据的完整性(Integrity)不能得到有效的检查。数据的完整性是指数据的正确性和相容性。数据必须正确,例如,学生的性别只能是男或女,某学生某门课程的考试成绩只能在 0~100 分之间等。另外,数据也必须相容(即一致性),否则不同系统对相同的数据就会有不同的输出,这会使用户无所适从,这一点问题一中已经有了论述。系统必须进行数据的完整性检查,以防错误数据的输入和输出,即所谓垃圾进垃圾出(garbage in garbage out)造成的错误结果。

第三是并发操作(Concurrency)不能得到有效的控制。当多个用户同时存取或修改某项数据时会相互干扰而得到错误的结果或破坏数据的完整性。例如,在学生选课管理系统中,当一个用户修改了某个数据(5 改为 4),另一个用户也同时修改了该数据(也是 5 改为 4),从而使得前一个用户的修改操作丢失。如果该数据表示某门课程剩余的选修名额数,那么就出现了一个名额给了两个学生的问题。系统必须对并发操作进行控制,以防这种现象的出现。

第四是数据遭受破坏后不能得到有效的恢复(Recovery)。硬件故障、软件故障或误操作是不可避免的,这些错误会破坏数据的正确性和相容性,或者使得部分或全部数据丢失。系统必须具有迅速将数据从错误状态恢复到某一已知的正确状态的能力。

建立在文件系统之上的信息系统,解决上述 4 个问题的办法通常是:在各自独立的系统中加上安全保密子系统、完整性检查子系统、并发控制子系统和恢复子系统。但这会产生两个新的问题。

第一,各个系统中有大量的重复程序(事实上,在数据的排序和查找等环节上也存在大量的重复程序),这一方面大大增加了系统的成本,另一方面也大大增加了程序维护的工作量。

第二,由于开发上述 4 个子系统有相当的难度加上开发应用程序的程序员的能力有限,各个系统对数据的管理和控制往往很不完善。无法想象一个数据的安全性不能得到有效的保护、数据破坏后不能得到及时的恢复、多个用户不能同时对数据进行操作、数据中又存在大量错误的系统有多大的实际使用价值?又有谁会去使用呢?

## 3. 数据库系统阶段

20 世纪 60 年代后期,计算机已经大量地应用于管理,硬件方面出现了大容量硬盘,且价格下降,而软件价格上升,编写和维护软件所需的成本相对增加。这时,由于管理的规模越来越大,多种应用程序共享数据的要求也越来越强烈。在这种背景下,以文件系统作为数据管理方法已经不能满足应用的需求。于是数据库技术便应运而生,从文件系统到数据库系统,标志着数据管理技术的飞跃。