

高等学校计算机系列教材

# 数据库系统概论

史嘉权 编著



清华大学出版社

高等学校计算机系列教材

# 数据库系统概论

史嘉权 编著

清华大学出版社

北京

## 内 容 简 介

本书围绕数据库的设计、编程与实现,系统、全面地介绍了数据库系统的基本概念、基本原理、基本方法以及应用技术。主要内容包括数据库建模、关系模型和关系运算、数据库语言 SQL(包括最新标准 SQL2 和 SQL3)及其系统环境、关系数据库设计理论及数据库设计、面向对象的对象定义语言和对象查询语言、以及查询优化和并发控制。

本书以关系数据库为基础,以数据库的设计与编程为重点,以引进面向对象的数据库技术为特色。主要特点是:内容新颖、系统全面;突出重点、注重总结;概念清晰、分析深入;例题丰富、实用性强;叙述深入浅出、语言流畅生动。

本书是高等院校计算机专业本科生数据库课程的教材,也可作为其他专业本科生数据库课程的教材。对于从事数据库研制、开发和应用的有关人员,本书也是一本很好的参考书。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

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。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

### 图书在版编目(CIP)数据

数据库系统概论/史嘉权编著. —北京:清华大学出版社,2006.1

(高等学校计算机系列教材)

ISBN 7-302-10267-8

I. 数… II. 史… 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(2004)第 142638 号

出 版 者: 清华大学出版社

<http://www.tup.com.cn>

社 总 机: 010-62770175

地 址: 北京清华大学学研大厦

邮 编: 100084

客 户 服 务: 010-62776969

责任编辑: 马琰珺

印 刷 者: 北京鑫海金澳胶印有限公司

装 订 者: 三河市金元印装有限公司

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 17 字数: 417 千字

版 次: 2006 年 1 月第 1 版 2006 年 1 月第 1 次印刷

书 号: ISBN 7-302-10267-8/TP·6996

印 数: 1~5000

定 价: 22.00 元

# 出版说明

清华大学出版社推出的这套《高等学校计算机系列教材》是《清华大学计算机系列教材》的姊妹篇。

《清华大学计算机系列教材》出版以来,多次获得国家 and 部级奖项。我们经常收到一些师生热情洋溢的来信,强烈感受到他们对新的知识与教育模式的渴求,同时也感受到广大师生对清华大学计算学科教学工作的关注和信任。

随着高等教育规模的持续扩大和高等教育改革的不断深入,不同院校对于计算学科的教学工作提出了新的要求,突出体现在:理论课时的压缩、实践能力的要求提高,以及学科教育与行业需求的不断结合。根据这些发展趋势,清华大学一批学术水平高、教学经验丰富的教授总结了他们几十年的教学和科研经验,有针对性地编写了《高等学校计算机系列教材》。这套教材的特点体现在:

1. 课程内容在《清华大学计算机系列教材》的基础上,进行了适时的修订更新,并且明确了教学基本要求,区分应该熟练掌握和只需一般了解的内容。

2. 强调加强基础理论教育,重视学生实践能力的培养。课程内容为进一步的实践教学既提供了基础知识,又留出了足够的时间。

另外,本套丛书同时出版了相关辅导用书,并为教师免费提供电子课件,便于师生的教学使用。

清华大学计算学科坚持推行具有启发性的、富于创造性的教学工作,为国家源源不断地培养出一批又一批优秀人才。从《清华大学计算机系列教材》中就可以体会到这些艰辛的探索历程,希望作为姊妹篇的《高等学校计算机系列教材》也能得到师生的认可。

清华大学出版社

2005年9月

# 前 言

数据库技术作为计算机软件领域的一个重要分支,是计算机科学技术中发展最快的领域之一,也是应用最广的技术之一。数据库技术发展到今天已成为以计算机为中心的信息系统与应用系统的核心技术和重要基础。

人类在 21 世纪将进入以知识经济为主要基础的信息社会,而数据库正是信息社会信息资源管理与开发利用的基础。

众所周知,“3C”即计算机、通信和信息内容(computer, communication and contents)已成为信息技术的核心,而信息内容则主要存放在数据库中。因此,数据库的建设规模和使用水平便成为衡量一个国家信息化程度的重要标志。

综上所述,在计算机专业的教学中,数据库课程的地位和作用是显而易见的。

在编写过程中,作者结合长年对清华大学计算机系本科生讲授多门软件课程的教学体会和经验,对书中的重点和难点进行了深入的分析,并结合典型例题使抽象的概念具体化,然后在此基础上进行总结归纳,以使读者准确理解、熟练掌握相关的知识点。比如,第 4 章嵌套查询部分,采用不同的解题方法,从解题思路、查询过程以及不同解法的本质差别等方面对典型例题进行了详细的分析,有助于开阔思路,深入理解并灵活运用所学知识。比如,第 6 章关系数据库设计理论有一定难度,作者从关系模式设计中可能出现的问题入手,分析产生的根源,提出解决的途径,进而总结了分解的原则和方法。又比如,第 5 章查询优化部分,涉及到关系代数的多个等价变换规则,比较抽象,作者通过典型例题使常用的等价变换规则变得具体而直观。

本书以当前的主流数据库——关系数据库——为基础,以数据库系统最常用最基本的内容——数据库的设计与编程——为重点,以引进数据库领域的最新成果——面向对象数据库的对象定义语言 ODL 和对象查询语言 OQL 以及结构化查询语言 SQL 的最新标准 SQL2 和 SQL3——为特色。

本书在编写时之所以把内容新颖作为考虑的一个重要因素,是希望在数据库技术发展很快的情况下,本书作为教材在 10 年之内都能基本适用。这样有利于教师熟悉教材、习题及实验内容,并能融会贯通,从而达到最好的教学效果。

本书的主要特点是:内容新颖、系统全面;突出重点、注重总结;概念清晰、分析深入;例题丰富、实用性强;叙述深入浅出、语言流畅生动。

本书主要围绕数据库的设计、编程与实现,讨论数据库系统的基本概念、基本原理、基本方法以及有关的应用。全书共分 10 章。第 1 章介绍关系模型、关系数据库、数据库体系结构、数据库管理系统、数据库运行过程等基础知识。第 2 章讲解数据库建模的两种基本方法:对象定义语言 ODL 和实体-联系模型(E-R 图)。第 3 章首先讲解关系模型,包括从 ODL 设计或 E-R 图设计转换为关系设计的方法;然后讲解三种关系运算:关系代数、关系演算和关系逻辑。第 4 章讲解作为数据库标准语言的结构化查询语言 SQL,包括查询和更新数据库、定义关系模式等语句的基本格式及其应用。第 5 章讨论查询优化和并发控制,本章

内容属于数据库实现的范畴。第6章讨论关系数据库设计理论,提出规范化的模式分解方法。第7章介绍数据库设计的全过程。第8章结合嵌入式SQL讲解SQL的系统环境,讨论了数据库的完整性和安全性。第9章讲解面向对象的查询语言,包括对象查询语言OQL和结构化查询语言的最新标准SQL3所扩充的面向对象的功能。第10章介绍数据库技术的最新发展动态。

为便于学习,在每章的前面都列出了本章要点,指出本章的主要内容以及应该理解和掌握的知识点;在每章的最后都给出本章的小结,对本章的主要知识点进行归纳总结;在每章的后面附有习题,帮助读者巩固所学的知识。

教学实验是学习数据库课程的重要环节。本书的最后是和教学实验有关的内容,包括:MS SQL Server介绍,本课程以微软公司开发的数据库管理系统MS SQL Server为实验环境;MS SQL Server 2000的有关语法,供读者编程时查阅。

常迪、赵鹏等同学参加了部分章节的编写。清华大学戴梅萼教授审阅了全稿,并提出了宝贵的建议和意见,在此表示衷心的感谢。

另外,《数据库系统概论——习题、实验与考试辅导》作为与本书配套的辅助教材,将与本书同时出版,主要内容包括:《数据库系统概论》要点;习题及其解答(共508题,部分难题有2种解法,个别难题有深入分析);实验环境与基础知识(系统介绍微软公司的SQL Server 2000及与教学实验有关的基础知识);教学实验内容及3个综合实验报告示例;考试大纲(给出要求熟练掌握、基本掌握、初步掌握和理解的知识);组卷方案、模拟试题及参考答案(各2套)。

由于水平有限,难免有不妥之处,敬请读者批评指正。

史嘉权

2005年9月12日 于清华园

# 目 录

<b>第 1 章 数据库系统概述</b> .....	1
1.1 数据管理技术的发展 .....	1
1.1.1 人工管理阶段.....	1
1.1.2 文件系统阶段.....	2
1.1.3 数据库系统阶段.....	2
1.2 有关数据库的基本术语 .....	4
1.2.1 数据.....	4
1.2.2 数据模型.....	5
1.2.3 数据库.....	5
1.2.4 数据库管理系统.....	5
1.2.5 数据库系统.....	5
1.3 关系数据库系统 .....	6
1.3.1 什么是关系.....	6
1.3.2 关系模型.....	7
1.3.3 关系数据库系统.....	7
1.4 数据库系统的体系结构 .....	7
1.4.1 三层模式结构.....	7
1.4.2 两层映像功能.....	8
1.5 DBMS 的体系结构 .....	9
1.5.1 DBMS 的组成概述 .....	9
1.5.2 查询处理程序 .....	10
1.5.3 存储管理程序 .....	10
1.5.4 事务管理程序 .....	11
1.5.5 客户程序/服务程序体系结构.....	12
1.6 数据库系统运行过程.....	12
1.7 本书导读.....	13
1.7.1 设计 .....	14
1.7.2 编程 .....	14
1.7.3 实现 .....	15
小结 .....	15
习题 .....	16
<b>第 2 章 数据库建模</b> .....	17
2.1 对象定义语言.....	18
2.1.1 面向对象的设计 .....	18

2.1.2	类的说明 .....	18
2.1.3	ODL 中的属性 .....	18
2.1.4	ODL 中的联系和反向联系 .....	20
2.1.5	联系的三种类型 .....	22
2.2	实体-联系模型(E-R 图) .....	23
2.2.1	E-R 图中联系的三种类型 .....	25
2.2.2	联系中的角色 .....	25
2.2.3	联系的多向性 .....	25
2.3	设计原则 .....	26
2.3.1	真实性 .....	26
2.3.2	避免冗余 .....	27
2.3.3	简单性 .....	27
2.3.4	合理选择元素类型 .....	27
2.4	子类 .....	28
2.4.1	ODL 中的子类 .....	28
2.4.2	ODL 中的多重继承 .....	29
2.4.3	E-R 图中的子类 .....	30
2.4.4	E-R 图中的继承 .....	31
2.5	对约束的建模 .....	32
2.5.1	键码 .....	33
2.5.2	单值约束 .....	34
2.5.3	引用完整性 .....	35
2.5.4	其他类型的约束 .....	36
小结	.....	36
习题	.....	37
<b>第 3 章</b>	<b>关系模型和关系运算 .....</b>	<b>39</b>
3.1	关系模型的基本概念 .....	39
3.1.1	属性 .....	40
3.1.2	模式 .....	40
3.1.3	元组 .....	40
3.1.4	域 .....	41
3.1.5	关系的等价表示法 .....	41
3.1.6	关系的实例 .....	41
3.2	从 ODL 设计到关系设计 .....	42
3.2.1	从 ODL 属性到关系属性 .....	42
3.2.2	非原子属性的表示 .....	42
3.2.3	单值联系的表示 .....	44
3.2.4	多值联系的表示 .....	45
3.2.5	联系与反向联系的表示 .....	45



3.2.6	ODL 子类的表示 .....	46
3.3	从 E-R 图到关系设计 .....	46
3.3.1	实体集到关系的转换 .....	46
3.3.2	E-R 联系到关系的转换 .....	47
3.3.3	“属于”联系到关系的转换 .....	48
3.4	关系代数 .....	48
3.4.1	关系的集合运算 .....	49
3.4.2	投影 .....	50
3.4.3	选择 .....	50
3.4.4	笛卡儿积 .....	51
3.4.5	自然连接 .....	52
3.4.6	$\theta$ 连接 .....	53
3.4.7	改名 .....	53
3.4.8	复合运算 .....	54
3.4.9	基本运算和导出运算 .....	55
3.5	关系演算 .....	56
3.5.1	元组关系演算 .....	56
3.5.2	域关系演算 .....	59
3.6	关系逻辑 .....	61
3.6.1	谓词和原子 .....	61
3.6.2	规则和查询 .....	61
3.6.3	从关系代数到数据逻辑 .....	62
小结	.....	66
习题	.....	67
<b>第 4 章</b>	<b>数据库语言 SQL .....</b>	<b>70</b>
4.1	SQL 的特点 .....	70
4.2	简单查询 .....	71
4.2.1	选择条件的构成 .....	73
4.2.2	字符串的比较 .....	74
4.2.3	日期和时间的比较 .....	75
4.2.4	输出的排序 .....	75
4.2.5	聚合运算符 .....	76
4.2.6	分组 .....	76
4.3	连接查询 .....	77
4.3.1	查询的并、交、差 .....	77
4.3.2	自然连接与笛卡儿积 .....	78
4.3.3	元组变量 .....	79
4.4	SQL2 中的连接查询 .....	79
4.4.1	笛卡儿积 .....	80

4.4.2	自然连接 .....	80
4.4.3	等值连接 .....	81
4.4.4	外部连接 .....	82
4.5	嵌套查询 .....	85
4.5.1	产生单值的子查询 .....	85
4.5.2	涉及到关系的选择条件 .....	86
4.5.3	涉及到元组的选择条件 .....	87
4.5.4	相关子查询 .....	88
4.5.5	例题详解 .....	91
4.5.6	查询语句格式小结 .....	94
4.6	数据库更新 .....	95
4.6.1	插入 .....	95
4.6.2	删除 .....	97
4.6.3	修改 .....	97
4.7	定义关系模式 .....	98
4.7.1	属性的数据类型 .....	98
4.7.2	定义表 .....	99
4.7.3	撤销表 .....	100
4.7.4	更改关系模式 .....	100
4.7.5	建立和撤销索引 .....	102
4.8	视图的定义和查询 .....	103
4.8.1	定义视图 .....	103
4.8.2	查询视图 .....	105
4.8.3	更新视图 .....	106
4.8.4	撤销视图 .....	107
	小结 .....	107
	习题 .....	109
<b>第5章</b>	<b>查询优化与并发控制 .....</b>	<b>111</b>
5.1	查询优化的一般策略 .....	111
5.2	关系代数的等价变换 .....	113
5.2.1	变换规则 .....	113
5.2.2	应用举例 .....	115
5.3	查询优化步骤 .....	117
5.4	并发调度 .....	118
5.4.1	事务 .....	119
5.4.2	数据不一致性 .....	119
5.4.3	可串行化调度 .....	120
5.5	封锁管理 .....	121
5.5.1	封锁机制中的主要概念 .....	121

5.5.2	封锁协议	122
小结		124
习题		125
<b>第6章</b>	<b>关系数据库设计理论</b>	<b>127</b>
6.1	函数依赖	127
6.1.1	函数依赖的定义	127
6.1.2	关系的键码	129
6.1.3	超键码	129
6.1.4	函数依赖规则	129
6.1.5	计算属性的封闭集	131
6.2	模式设计	132
6.2.1	问题的提出	132
6.2.2	问题的根源	133
6.2.3	解决的途径	135
6.2.4	分解的原则	139
6.2.5	分解的方法	141
6.2.6	关系模式规范化小结	144
6.3	多值依赖	145
6.3.1	属性独立性带来的冗余	145
6.3.2	多值依赖的定义	146
6.3.3	第4范式	146
6.3.4	分解成第4范式	147
小结		148
习题		149
<b>第7章</b>	<b>数据库设计</b>	<b>151</b>
7.1	概述	151
7.1.1	数据库设计的任务	151
7.1.2	数据库设计的特点	153
7.1.3	数据库设计的步骤	153
7.2	需求分析	155
7.2.1	应用领域的调查	156
7.2.2	定义信息与应用	157
7.2.3	定义操作任务	158
7.2.4	定义数据项	158
7.2.5	预测未来的改变	160
7.3	概念设计	161
7.3.1	概念设计的基本方法	161
7.3.2	视图设计的基本策略	162

7.3.3	视图综合设计方法	162
7.4	逻辑设计	164
7.4.1	E-R 图到关系模式的转换	164
7.4.2	逻辑模式的规范化和优化	168
7.5	物理设计	170
7.5.1	影响物理设计的因素	170
7.5.2	选择存取方法	171
7.5.3	设计存储结构	172
7.6	数据库的实施、运行和维护	173
7.6.1	数据库的实施	173
7.6.2	数据库的运行和维护	173
	小结	174
	习题	175
<b>第 8 章</b>	<b>SQL 系统环境</b>	<b>176</b>
8.1	嵌入式 SQL	176
8.1.1	什么是嵌入式 SQL	176
8.1.2	如何实现嵌入式 SQL	176
8.1.3	使用游标的 SQL 编程	179
8.1.4	嵌入式 SQL 的处理过程	180
8.2	有关事务的嵌入式 SQL	181
8.3	SQL 环境	184
8.3.1	数据库元素的层次结构	184
8.3.2	客户程序和服务程序系统	186
8.3.3	Client/Server 体系结构	188
8.4	数据库的完整性	189
8.4.1	说明键码约束	189
8.4.2	说明外键码约束	191
8.4.3	说明检验约束	192
8.4.4	更改约束	193
8.4.5	触发与触发程序	194
8.5	SQL 中的安全和用户权限	197
8.5.1	数据库的安全	197
8.5.2	数据库的访问控制	197
8.5.3	权限	198
8.5.4	用户的标识与鉴别	201
	小结	202
	习题	203
<b>第 9 章</b>	<b>面向对象查询语言</b>	<b>205</b>
9.1	对象查询语言 OQL 概述	205

9.1.1	ODL 中的方法与范围 .....	205
9.1.2	OQL 中的类型 .....	208
9.1.3	一个 OQL 的实例 .....	208
9.2	OQL 表达式 .....	209
9.2.1	路径表达式 .....	210
9.2.2	基本表达式 .....	210
9.2.3	表达式的附加格式 .....	212
9.3	OQL 及其宿主语言编程 .....	216
9.3.1	给宿主语言变量赋值 .....	216
9.3.2	从聚集中提取元素 .....	217
9.3.3	访问集中的每个元素 .....	217
9.4	SQL3 中的元组对象 .....	218
9.4.1	行类型及其说明 .....	219
9.4.2	引用及其使用 .....	220
9.4.3	作为值的对象标识 .....	222
9.5	SQL3 中的抽象数据类型 .....	223
9.5.1	ADT 的定义 .....	224
9.5.2	ADT 方法的定义 .....	225
9.6	ODL/OQL 和 SQL3 的比较 .....	226
	小结 .....	227
	习题 .....	228
<b>第 10 章</b>	<b>数据库技术发展动态 .....</b>	<b>231</b>
10.1	分布式数据库 .....	231
10.1.1	分布式数据库系统简介 .....	231
10.1.2	分布式数据库系统举例 .....	232
10.2	并行数据库 .....	233
10.2.1	并行数据库简介 .....	233
10.2.2	并行数据库系统结构 .....	234
10.3	多媒体数据库 .....	235
10.3.1	多媒体数据及其特点 .....	236
10.3.2	多媒体数据库简介 .....	237
10.4	主动数据库 .....	238
10.5	数据仓库 .....	239
10.5.1	数据仓库简介 .....	239
10.5.2	数据仓库的结构 .....	240
	小结 .....	241
	习题 .....	241
<b>附录 A</b>	<b>MS SQL Server 介绍 .....</b>	<b>242</b>
A.1	SQL Server 简介 .....	242

A. 2	SQL Server 管理工具和实用程序 .....	242
<b>附录 B</b>	<b>MS SQL Server 2000——Transact-SQL 语法 .....</b>	<b>245</b>
B. 1	简单说明 .....	245
B. 1. 1	语法约定 .....	245
B. 1. 2	表达式 .....	245
B. 1. 3	搜索条件 .....	245
B. 2	建立、更改和撤销基本表 .....	246
B. 2. 1	建表 .....	246
B. 2. 2	更改表 .....	249
B. 2. 3	撤销表 .....	249
B. 3	更新数据 .....	250
B. 3. 1	插入数据 .....	250
B. 3. 2	删除数据 .....	251
B. 3. 3	修改数据 .....	251
B. 4	查询 .....	252
<b>参考文献</b>	.....	<b>254</b>

# 第 1 章 数据库系统概述

## 本章要点

1. 了解数据管理技术的发展,其中重点是数据库系统阶段。理解数据库系统的主要特点,初步了解数据库系统。
2. 关系数据库系统是当前数据库系统的主流,也是本书的重点。初步理解关系、关系模型、关系数据库系统等基本概念。
3. 了解数据库管理系统的组成和各部分的基本功能。
4. 初步理解本章提到的有关数据库的专业术语,为后续各章的学习打下好的基础。

数据库技术是计算机应用领域中非常重要的技术,它产生于 20 世纪 60 年代末,是数据管理的最新技术,也是软件科学的一个重要分支。本章首先回顾数据管理技术的发展过程,然后介绍数据库技术的基本术语,并在此基础上介绍关系数据库系统和数据库系统结构以及数据库管理系统的体系结构。

## 1.1 数据管理技术的发展

数据管理指的是如何对数据进行分类、组织、储存、检索及维护。要注意,这里所说的数据,不仅是指数字,还包括文字、图形、图像、声音等。凡是计算机中用来描述事物的记录,统称为数据。

随着计算机软硬件的发展,数据管理技术不断地完善,经历了如下三个阶段:

- (1) 人工管理阶段;
- (2) 文件系统阶段;
- (3) 数据库系统阶段。

### 1.1.1 人工管理阶段

20 世纪 50 年代中期以前,计算机主要用于科学计算。那时的计算机硬件方面,外存只有卡片、纸带及磁带,没有磁盘等直接存取的存储设备;软件方面,只有汇编语言,没有操作系统和高级语言,更没有管理数据的软件;数据处理的方式是批处理。这些决定了当时的数据管理只能依赖人工来进行。

人工管理阶段的特点是:

- (1) 数据不进行保存。当时的计算机主要用于科学计算,一个程序对应一组数据。在计算某一问题时,把程序和对应的数据装入,计算完就退出,没有将数据长期保存的必要。
- (2) 没有专门的数据管理软件。数据需要由应用程序自己管理,因此应用程序的设计者不仅要考虑数据的逻辑结构,还要考虑数据的物理结构,比如存储结构、存取方法、输入输

出方式等等。一旦存储结构发生变化,应用程序也要做相应的修改,程序员的负担非常重,数据的独立性也很差。

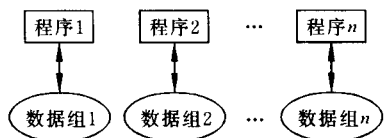


图 1.1 人工管理阶段的特征

(3) 数据面向应用。一组数据对应一个程序。倘若多个程序使用相同的数据,必须各自定义,不能共享。所以程序之间存在大量的数据冗余。

(4) 只有程序的概念,基本上没有文件的概念。

人工管理阶段的特征如图 1.1 所示。

### 1.1.2 文件系统阶段

20 世纪 50 年代末到 60 年代中期,随着科学技术的进步,计算机技术有了很大提高,计算机的应用范围也不断扩大,不仅用于科学计算,还大量用于管理。这时计算机硬件已经有了磁盘、磁鼓等直接存取的外存设备;软件则有了操作系统、高级语言,操作系统中的文件系统是专门用于数据管理的软件;处理方式不仅有批处理,还增加了联机实时处理。

文件系统阶段的特点如下:

(1) 数据可以长期保存在磁盘上。用户可以反复对文件进行查询、修改、插入和删除等操作。

(2) 文件系统提供了数据与程序之间的存取方法。应用程序和数据有了一定的独立性。数据物理结构的改变也不一定反映在程序上,大大减轻了程序员的负担。

(3) 数据冗余量大。文件系统中,文件仍然是面向应用的,一个文件基本上对应于一个应用程序。即使多个程序使用了一部分相同的数据,也必须建立各自的文件,不能对数据项进行共享,因此数据冗余大,存储空间浪费。由于数据可能有多个副本,对其中之一进行修改时还容易造成数据的不一致性。

(4) 文件之间缺乏联系,相互孤立,仍然不能反映现实世界各种事物之间错综复杂的联系。

文件系统阶段的特征如图 1.2 所示。

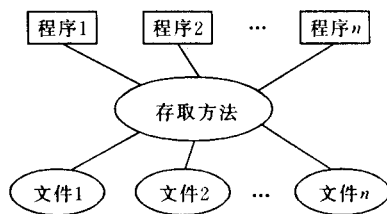


图 1.2 文件系统阶段的特征

### 1.1.3 数据库系统阶段

20 世纪 60 年代末以来,计算机的应用更为广泛,用于数据管理的规模也更为庞大,由此带来数据量的急剧膨胀。计算机磁盘技术有了很大发展,出现了大容量的磁盘。在处理方式上,联机实时处理的要求更多。这种种变化都促进了数据管理手段的进步,数据库技术应运而生。

数据库系统的特点如下:

(1) 数据的结构化

在文件系统阶段,只考虑了同一文件记录内部数据项之间的联系,而不同文件的记录之间是没有联系的,也就是说,从整体上看数据是无结构的。如上一节所述,这样的文件是有局限性的,不能反映现实世界各种事物之间错综复杂的联系。在数据库系统中,实现了整体



数据的结构化,把文件系统中简单的记录结构变成了记录和记录之间的联系所构成的结构化数据。在描述数据的时候,不仅要描述数据本身,还要描述数据之间的联系。数据之间的联系通过存取路径来实现,把相关的数据有机地组织在一起。

例如在学校的管理系统中,不同的部门有不同的要求,人事、医疗、教务等部门分别了解学生的人事情况、医疗保健情况、选课情况等等。传统的文件系统中,不同的应用要使用不同的文件。比较简单的文件形式是等长、同格式记录的集合。比如学生的人事记录文件,可以采用图 1.3 所示的记录格式。

学号	姓名	性别	出生年月	系别	政治面貌	籍贯	家庭成员	简历
----	----	----	------	----	------	----	------	----

图 1.3 学生人事记录

而学生的选课记录文件,则可以采用图 1.4 所示的记录格式。

学号	姓名	性别	出生年月	系别	课程号	课程名	成绩
----	----	----	------	----	-----	-----	----

图 1.4 学生选课记录

由图 1.3 和图 1.4 可见,首先,每个学生的情况不同,其家庭成员、简历、选课的数据量有多有少,如果用等长记录格式存储学生数据,只能按数据量最大的学生记录来安排存储,这样会造成极大的浪费,如果用变长记录来存储,又不便于数据管理;其次,无论是人事记录文件还是选课记录文件,每个文件记录的数据项都包括了学号、姓名、性别和出生年月等,这造成了大量的重复存储。

在数据库系统中,从整体的角度来组织数据,综合考虑各种应用,有效地解决了上述问题。数据组织方式如图 1.5 所示。

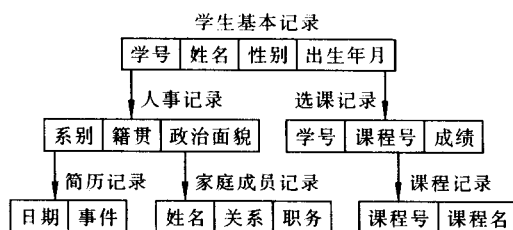


图 1.5 结构化的学生记录

### (2) 数据共享性好

由图 1.5 设计的数据结构可见,人事部门可以据此了解学生的人事情况,教务部门也可以据此了解学生的选课情况,这些数据可以供多个部门使用,实现了数据的共享。各个部门的数据基本上没有重复的存储,数据的冗余量较小。

### (3) 数据独立性好

数据库系统有三层结构:用户(局部)数据的逻辑结构、整体数据的逻辑结构和数据的物理结构。在这三层结构之间数据库系统提供了两层映像功能。首先是用户数据逻辑结构和整体数据逻辑结构之间的映像,这一映像保证了数据的逻辑独立性:当数据库的整体逻辑结构发生变化时,通过修改这层映像可使局部的逻辑结构不受影响,因此不必修改应用程序