

教育部人才培养模式改革和开放教育试点教材

# 数据库系统 教程

史嘉权 等 编著



清华大学出版社

<http://www.tup.tsinghua.edu.cn>

教育部人才培养模式改革和开放教育试点教材

# 数据库系统教程

史嘉权 等编著

清华大学出版社

**(京)新登字 158 号**

### 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 SQL.2 和 SQL.3)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。主要特点是新颖、系统、全面、实用。书中内容深入浅出,通俗易懂,概念清晰,例题丰富,并注重总结归纳。

本书是中央广播电视大学开放教育计算机科学与技术专业本科生数据库课程的教材,也可作为其他高等院校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。**

**本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。**

书 名:数据库系统教程

作 者:史嘉权 等编著

出版者:清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者:北京通州大中印刷厂

发行者:新华书店总店北京发行所

开 本:787×1092 1/16 印张:17.25 字数:396 千字

版 次:2001 年 8 月第 1 版 2001 年 9 月第 2 次印刷

书 号:ISBN 7-302-04699-9/TP·2794

印 数:5001~10000

定 价:22.00 元

## 序

我们正处在跨越世纪的门槛上,人类社会在一股股变革性力量的推动下发生着根本性的变化。知识经济时代的到来向我们显示,一个国家最重要的资源已经不再是土地、劳动力或资本,而是其国民的知识和创造力;国与国的竞争虽然常常表现为政治、经济或军事实力的较量,但归根到底已是一场教育和科技的竞争。换言之,国家的综合实力将主要由其国民的教育水平来决定。一时间,世界各国的校长们、跨国企业的巨头们乃至许多的政府首脑们都在纷纷议论 21 世纪的教育,以迎接知识经济的挑战。我们中华民族有着蜿蜒几千年的文明,为在世界民族之林重振雄风,再展辉煌,发出了时代的特强音:实施科教兴国,提高全民素质。从中央领导到广大群众,都对教育提出了更高的要求,寄予了更大的希望,同时也给予了更多的支持。人们在这方面的思想观念和实践探索正在以空前的速度发展着。

中国的高等教育已经走完了一个世纪的路程。已经过去的 20 世纪正是它从无到有、从小到大、由产生到发展的一段百年历史。中国人民在短短的数十年时间里构筑了资本主义国家好几百年才形成的高等教育体系,涌现出一批高水平的学校,培养了一大批高层次优秀人才,取得了辉煌的成就。但是在新时期,教育不适应现代化建设需要的矛盾不断显露,我国劳动者受教育水平普遍较低的现象无法面对新世纪的机遇和挑战,我国高等教育的发展现状也难以满足广大人民群众空前强烈的受教育愿望。一代伟人邓小平早在十年前就一针见血地指出,我们的最大失误是教育,一是放松了对青少年的思想道德教育,二是教育规模发展不够快。现在看来,这两个问题依然是症结所在。一个十二亿人口的泱泱大国,高等学校的毛入学率仅 10% 左右,实在很不相称。我国的高等教育已经面临着大力发展、高速发展、从根本上改变落后状态的紧迫问题。

令人欣慰和鼓舞的是中国有一所全世界最大的大学——中国广播电视大学,上百万的学生遍布在九百六十万平方公里的辽阔土地上。它突破传统教育在空间上的限制,不断减弱时间上的束缚,以覆盖面广、全方位为各类社会成员提供教育服务的优势,成为中国高等教育体系中的一个重要组成部分。二十多年来,它为实现高等教育大众化,为提高我国劳动者的整体素质,为变巨大的人口包袱为巨大的人力资源,以形成浩浩荡荡的高水平建设大军,发挥了不可磨灭的作用。最近,中央电大又有重大改革举措,进一步面向社会开展了“开放教育”等项试点工作,在教育思想、招生对象、培养模式、管理机制方面进行新的探索。尤其引人注目的是中央电大与国内的一些重点高校形成了紧密的合作关系,携手为我国现代远程教育开拓新路。重点高校有学科和教学上的优势,它们的加盟有利于电大提高教学质量、办出特色;而中央电大有很丰富的教育资源,有完整的办学系统,有一支富有经验的教学与管理队伍,特别是有较强的社会服务意识和人才市场意识,这对于需要进一步向社会开放的普通高校而言,又有许多值得学习和借鉴之处。我们完全有理由相信,中央电大和重点高校的结合,不仅可以在现阶段实现优势互补、资源共享,而且有

可能成长出一种符合我国国情发展教育的最具潜力的新型教育模式。

现在摆在我们面前的这套中央广播电视大学本科(专科起点)“计算机科学与技术”专业教材,就是中央电大和清华大学合作的产物。在开放教育试点启动之际,在计算机及其网络技术日新月异、其爆炸式发展和神话般应用使人们眼花缭乱、不知所措之时,在我国至少缺乏数十万计算机软件及网络技术人才的当口,这套教材像雪里送炭,像清风送爽,终于在人们的企盼和惊喜中问世了。它确实及时和解渴。教材的编者是清华大学计算机系一批学术水平高、教学经验丰富的教授,他们以知识、能力和素质的全面训练为目标,将教材的先进性、实用性和可读性融为一体。教材纲目清楚,重点突出,深入浅出,便于自学。书中每章有小结,章章有习题,有的还配有实验指导和习题解答,不仅对计算机专业学生适用,其他专业的学生也可以从此入门。清华大学的老师们还准备为这套教材制作多媒体导读光盘和网络辅导教材,指明教学基本要求,区分应该熟练掌握和只需一般了解的内容,并进行重点难点分析和讲解。这全套的教材称得上是难得的好书。

对于中国广播电视大学我是颇有感情的,不只是因为它过去的功绩和带给人们未来的曙光,还因为我本人二十年前也曾参与过中央电大《电子技术基础》课程的教学工作。那时我收到许多电大学生热情洋溢的来信,强烈感受到他们对知识与教育的渴求,感受到他们学习的艰辛和坚韧不拔的毅力,同时也感受到了广大学生对我的信任和鼓励。当年的电大学生如今多数已成为我国经济建设和社会发展中的骨干,一些人后来获得了博士学位,有的已成为我国重点大学的教授。中央电大的成功实践已在社会上赢得了很好的声誉,而当前扩大教育规模、构建终身学习体系的社会呼唤又给电大今后的发展提供了新的难得的机遇。近年来,信息网络与多媒体技术突飞猛进,也使电大的远程教育形式跃上了现代化的新台阶。这次中央电大和清华大学合作,共同在计算机专业开放教育改革试点中付出了辛勤的劳动,播下了希望的种子。我期待着中央电大有更多的创新,更大的发展,更加充满活力。我也殷切希望电大的学生们为中华民族的强盛而自强不息,学有所成。

努力吧,中国广播电视大学一定能成为中国教育界一颗璀璨的明珠。

清华大学副校长、教授 **胡东成**

二〇〇〇年八月于北京

# 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。

本书作为中央电大专升本的教材,在编写时注意了以下原则:首先,在深度和广度上要完全达到计算机专业大学本科生的要求;其次,起步基点较低,以适应从未学过数据库的读者(包括非计算机专科的学生);最后,对于原来计算机专科的学生,能温故而知新,进而在更高的层次上理解和掌握本教材的内容。

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 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。

本书的主要特点是新颖、系统、全面、实用。内容深入浅出,注意总结归纳。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:综合实验,给出三个综合实验课题和实验要求,并且给出其中一个实验“零件交易中心管理系统”的实验报告,作为完整的实验报告供读者参考;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 6.5/7.0 的有关语法,供读者编程时查阅。

常迪、赵鹏、王欣和李博同学参加了部分章节的编写。清华大学戴梅萼教授、北京理工大学吴鹤龄教授和首都经贸大学盛定宇教授分别审阅了全稿,并提出了宝贵的建议和意见,在此表示衷心的感谢。

由于水平有限,难免有不妥之处,敬请读者惠予批评指正。

史嘉权

2001年6月16日 于清华园

# 目 录

序 .....	I
前言 .....	III
<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 .....	5
1.2.1 数据 .....	5
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 .....	6
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 .....	8
1.4.2 两层映象功能 .....	9
1.5 DBMS 的体系结构 .....	10
1.5.1 DBMS 的组成概述 .....	10
1.5.2 查询处理程序.....	10
1.5.3 存储管理程序.....	11
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.....	13
1.7 本书导读.....	14
1.7.1 设计.....	14
1.7.2 编程.....	15
1.7.3 实现.....	15
小结 .....	15



习题 .....	16
<b>第 2 章 数据库建模 .....</b>	<b>17</b>
2.1 对象定义语言 ODL .....	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 .....	21
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 .....	24
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 .....	26
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 .....	30
2.5 对约束的建模 .....	31
2.5.1 键码 .....	32
2.5.2 单值约束 .....	34
2.5.3 引用完整性 .....	34
2.5.4 其他类型的约束 .....	35
小结 .....	36
习题 .....	37
<b>第 3 章 关系模型和关系运算 .....</b>	<b>38</b>
3.1 关系模型的基本概念 .....	38
3.1.1 属性 .....	39
3.1.2 模式 .....	39
3.1.3 元组 .....	39
3.1.4 域 .....	40
3.1.5 关系的等价表示法 .....	40
3.1.6 关系的实例 .....	40

3.2	从 ODL 设计到关系设计 .....	41
3.2.1	从 ODL 属性到关系属性 .....	41
3.2.2	类中的非原子属性 .....	41
3.2.3	单值联系的表示 .....	43
3.2.4	多值联系的表示 .....	44
3.2.5	联系与反向联系的表示 .....	44
3.2.6	ODL 子类的表示 .....	45
3.3	从 E/R 图到关系设计 .....	45
3.3.1	实体集到关系的转换 .....	45
3.3.2	E/R 联系到关系的转换 .....	46
3.3.3	“属于”联系到关系的转换 .....	47
3.4	关系代数 .....	47
3.4.1	关系的集合运算 .....	48
3.4.2	投影 .....	49
3.4.3	选择 .....	50
3.4.4	笛卡儿积 .....	50
3.4.5	自然连接 .....	51
3.4.6	$\theta$ 连接 .....	52
3.4.7	改名 .....	52
3.4.8	复合运算 .....	53
3.4.9	基本运算和导出运算 .....	54
3.5	关系演算 .....	55
3.5.1	元组关系演算 .....	55
3.5.2	域关系演算 .....	59
3.6	关系逻辑 .....	60
3.6.1	谓词和原子 .....	60
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 .....	65
	习题 .....	67
<b>第 4 章</b>	<b>数据库语言 SQL .....</b>	<b>69</b>
4.1	SQL 的特点 .....	69
4.2	简单查询 .....	70
4.2.1	选择条件的构成 .....	72
4.2.2	字符串的比较 .....	73
4.2.3	日期和时间的比较 .....	74
4.2.4	输出的排序 .....	74
4.2.5	聚合运算符 .....	75

4.2.6 分组	75
4.3 连接查询	76
4.3.1 查询的并、交、差	76
4.3.2 连接与笛卡儿积	77
4.3.3 元组变量	78
4.4 嵌套查询	78
4.4.1 产生单值的子查询	79
4.4.2 涉及到关系的选择条件	80
4.4.3 涉及到元组的选择条件	81
4.4.4 相关子查询	82
4.5 数据库更新	85
4.5.1 插入	85
4.5.2 删除	86
4.5.3 修改	87
4.6 定义关系模式	88
4.6.1 属性的数据类型	88
4.6.2 定义表	89
4.6.3 撤销表	89
4.6.4 更改关系模式	90
4.6.5 建立和撤销索引	91
4.7 视图的定义和查询	93
4.7.1 定义视图	93
4.7.2 查询视图	94
4.7.3 更新视图	95
4.7.4 撤销视图	97
小结	97
习题	99
<b>第5章 查询优化与并发控制</b>	<b>101</b>
5.1 查询优化的一般策略	101
5.2 关系代数的等价变换	103
5.2.1 变换规则	103
5.2.2 应用举例	105
5.3 查询优化步骤	107
5.4 并发调度	109
5.4.1 事务	109
5.4.2 数据不一致性	110
5.4.3 可串行化调度	110
5.5 封锁管理	112

5.5.1	封锁机制中的主要概念 .....	112
5.5.2	封锁协议 .....	112
小结	.....	114
习题	.....	116
<b>第 6 章</b>	<b>关系数据库设计理论</b> .....	<b>117</b>
6.1	函数依赖 .....	117
6.1.1	函数依赖的定义 .....	117
6.1.2	关系的键码 .....	119
6.1.3	超键码 .....	119
6.1.4	函数依赖规则 .....	119
6.1.5	计算属性的封闭集 .....	121
6.2	模式设计 .....	122
6.2.1	问题的提出 .....	123
6.2.2	问题的根源 .....	123
6.2.3	解决的途径 .....	125
6.2.4	分解的原则 .....	129
6.2.5	分解的方法 .....	132
6.2.6	关系模式规范化小结 .....	135
6.3	多值依赖 .....	135
6.3.1	属性独立性带来的冗余 .....	135
6.3.2	多值依赖的定义 .....	136
6.3.3	第四范式 .....	137
6.3.4	分解成第四范式 .....	137
小结	.....	138
习题	.....	140
<b>第 7 章</b>	<b>数据库设计</b> .....	<b>141</b>
7.1	概述 .....	141
7.1.1	数据库设计的任务 .....	141
7.1.2	数据库设计的特点 .....	143
7.1.3	数据库设计的步骤 .....	143
7.2	需求分析 .....	145
7.2.1	应用领域的调查 .....	146
7.2.2	定义信息与应用 .....	147
7.2.3	定义操作任务 .....	148
7.2.4	定义数据项 .....	148
7.2.5	预测未来的改变 .....	150
7.3	概念设计 .....	151

7.3.1	概念设计的基本方法 .....	151
7.3.2	视图设计的基本策略 .....	152
7.3.3	视图综合设计方法 .....	152
7.4	逻辑设计 .....	154
7.4.1	E/R图到关系模式的转换 .....	154
7.4.2	逻辑模式的规范化和优化 .....	158
7.5	物理设计 .....	160
7.5.1	影响物理设计的因素 .....	160
7.5.2	选择存取方法 .....	161
7.5.3	设计存储结构 .....	162
7.6	数据库的实施、运行和维护 .....	163
7.6.1	数据库的实施 .....	163
7.6.2	数据库的运行和维护 .....	164
	小结 .....	165
	习题 .....	165
<b>第8章</b>	<b>SQL系统环境 .....</b>	<b>167</b>
8.1	嵌入式SQL .....	167
8.1.1	什么是嵌入式SQL .....	167
8.1.2	如何实现嵌入式SQL .....	167
8.1.3	使用游标的SQL编程 .....	170
8.1.4	嵌入式SQL的处理过程 .....	171
8.2	有关事务的嵌入式SQL .....	172
8.3	SQL环境 .....	175
8.3.1	数据库元素的层次结构 .....	175
8.3.2	客户程序和服务程序系统 .....	177
8.3.3	Client/Server体系结构 .....	179
8.4	数据库的完整性 .....	180
8.4.1	说明键码约束 .....	181
8.4.2	说明外键码约束 .....	182
8.4.3	说明检验约束 .....	183
8.4.4	更新约束 .....	184
8.4.5	触发与触发程序 .....	185
8.5	SQL中的安全和用户权限 .....	188
8.5.1	数据库的安全 .....	188
8.5.2	数据库的访问控制 .....	189
8.5.3	权限 .....	190
8.5.4	用户的标识与鉴别 .....	193
	小结 .....	194

习题	194
<b>第9章 面向对象查询语言</b>	<b>197</b>
9.1 对象查询语言 OQL 概述	197
9.1.1 ODL 中的方法与范围	197
9.1.2 OQL 中的类型	200
9.1.3 一个 OQL 的实例	200
9.2 OQL 表达式	201
9.2.1 路径表达式	202
9.2.2 基本表达式	202
9.2.3 表达式的附加格式	205
9.3 OQL 及其宿主语言编程	209
9.3.1 给宿主语言变量赋值	209
9.3.2 从聚集中提取元素	209
9.3.3 访问聚集中的每个元素	210
9.4 SQL3 中的元组对象	211
9.4.1 行类型及其说明	211
9.4.2 引用及其使用	212
9.4.3 作为值的对象标识	215
9.5 SQL3 中的抽象数据类型	216
9.5.1 ADT 的定义	216
9.5.2 ADT 方法的定义	217
9.6 ODL/OQL 和 SQL3 的比较	219
小结	220
习题	221
<b>第10章 数据库技术发展动态</b>	<b>223</b>
10.1 分布式数据库	223
10.1.1 分布式数据库系统简介	223
10.1.2 分布式数据库系统举例	224
10.2 并行数据库	225
10.2.1 并行数据库简介	225
10.2.2 并行数据库系统结构	226
10.3 多媒体数据库	228
10.3.1 多媒体数据及其特点	228
10.3.2 多媒体数据库简介	230
10.4 主动数据库	230
10.5 数据仓库	231
10.5.1 数据仓库简介	231

10.5.2 数据仓库的结构·····	232
小结·····	233
习题·····	234
<b>附录 A 数据库系统概论课教学实验</b> ·····	<b>235</b>
A.1 综合实验·····	235
A.1.1 实验一 零件交易中心管理系统·····	235
A.1.2 实验二 图书管理系统·····	235
A.1.3 实验三 民航订票管理系统·····	236
A.2 零件交易中心管理系统实验报告(参考答案)·····	236
<b>附录 B MS SQL Server 介绍</b> ·····	<b>248</b>
B.1 SQL Server 简介·····	248
B.2 SQL Server 管理工具和实用程序·····	248
<b>附录 C MS SQL SERVER 6.5/7.0——Transact-SQL 语法</b> ·····	<b>251</b>
C.1 简单说明·····	251
C.1.1 语约定·····	251
C.1.2 表达式·····	251
C.1.3 搜索条件·····	251
C.2 建立、更改和撤销基本表·····	252
C.2.1 建表·····	252
C.2.2 更改表·····	255
C.2.3 撤销表·····	256
C.3 更新数据·····	256
C.3.1 插入数据·····	256
C.3.2 删除数据·····	257
C.3.3 修改数据·····	257
C.4 查询·····	258
<b>参考文献</b> ·····	<b>260</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。

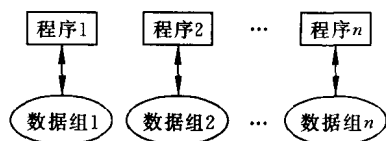


图 1.1 人工管理阶段的特征

(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。

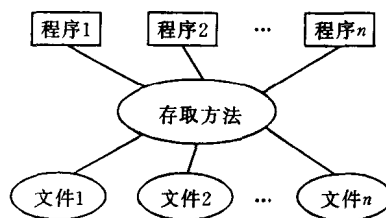


图 1.2 文件系统阶段的特征

### 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