

计算机大专教材系列

数据库系统教程

袁晓洁 徐 雷 编著



南开大学出版社

内 容 提 要

本书系统、完整地讲述了当前数据库技术由基本原理到应用实践的主要内容。全书分为两大部分:第一部分是数据库系统的基本原理,由第一章到第七章组成,内容包括:数据模型、数据库系统构成、数据库管理系统、关系数据库、SQL 语言、关系数据理论、关系系统的查询优化、数据库环境和数据库设计;第二部分是 Fox BASE+ 程序设计语言,由第八章组成,介绍了 Fox BASE+ 基本命令和程序设计技术。

本书是在总结了成功的教学方法和经验的基础上,综合数据库领域里重要的基础知识,并理论联系实际,将基本原理融汇贯通于数据库实用技术中编写而成的。它特别适合作为大学、大专、中专及成人教育等计算机专业的数据库教材,同时也可作为非计算机专业的工程技术人员和管理人员学习计算机基础知识的教科书。

数据库系统教程

袁晓洁 徐 雷 编著

南开大学出版社

(天津八里台南开大学校内)

邮编 300071 电话 3509318

新华书店天津发行所发行

天津宝坻县印刷厂印刷

1995 年 11 月第 1 版 1995 年 11 月第 1 次印刷

开本:787×1092 1/16 印张:12.625

字数:314 千 印数:1—8000

ISBN 7-310-00889-8

TP·47 定价:13.50 元

“计算机大专教材系列”编委会

主 编
副主编
编 委

陈有祺
朱瑞香
朱耀庭
袁晓洁
伍颖文
张 蓓

吴功宜
于春凡
周玉龙
李正明

王家骅
孙桂茹
幸运帙
何志红

李 信
刘 军
裴志明

出版说明

随着计算机应用的日益深入、普及,目前我国正在兴起学习计算机专业知识的高潮,各种有关计算机的书籍如雨后春笋般涌现出来,使广大读者大有应接不暇之势。但是,已经出版的这些书籍中,有的偏深偏专,取材偏多偏全,适合有一定基础的计算机专业人员阅读参考;有的则是普及性读物,只适合急于入门的计算机爱好者使用;在为数不多的教材中,大都是为计算机专业本科生使用而编写的,不适合成人教育和大专类学生的需要。鉴于这种形势,我们决定编写一套适合于计算机类各专业大专学生和成人教育使用的教材。这套教材共有十种,虽然它还不能完全覆盖上述办学层次教学计划中的所有课程,但是它包含了培养一个计算机类专科生的主要教学内容。其中入门的教材有《计算机应用基础》和《C 语言程序设计》;属于专业基础的教材有《16 位微型计算机原理与接口》,《汇编语言程序设计》,《数据结构》和《操作系统》;应用性较强的《单片机及其应用》,《数据库系统教程》,《计算机网络基础》和《软件工程引论》。

这套教材贯彻了理论联系实际、学以致用原则。在取材方面,不追求包罗万象、面面俱到,而着力保证把最基本、最实用的部分包含进来。在叙述方面,力求做到深入浅出,尽量用实例来说明基本概念和基本方法。我们希望这套教材不仅能适合课堂讲授的需要,也便于广大读者自学。这套教材由南开大学计算机与系统科学系的教师们编写而成,他们之中既有教学经验十分丰富的教授、副教授,也有活跃在计算机应用最前沿的青年教师。这些教师不仅具有教本科生、研究生的教学经验,也具有教大专生和成人教育的教学经验,这就使这套教材的质量有了基本的保证。但是由于我们初次编写这类教材,尚未经过实践的检验,缺点和不足之处在所难免,敬希同行专家和广大使用者批评指正。

075245/10

前言

目前,计算机应用已普及到各个领域,掌握计算机的基础知识,将计算机运用于各个行业,尽最大可能地发挥其潜在能力,已是一种大趋势。在计算机领域里,应用最广、实用性最强的一种技术就是数据库技术,它已渗透到许许多多的应用行业中,成为人们掌握计算机知识所必修的一门课程。为此我们编写了这本《数据库系统教程》,作为人们学习这门技术所必需的一本教材。

到目前为止,有关数据库方面的教材已有许多种,这些书从内容结构上可分为两类:一类是讲数据库理论知识的,并不涉及某个具体的实用数据库系统;另一类是讲述某一个或几个实用数据库系统的技术及使用方法,而不过多地阐述数据库的基础理论。因此,使用上述教材,很难使读者理论联系实际,也很难全面理解和掌握数据库的完整知识和概念。鉴于此,作者在多年的数据库教学实践中,吸取其他多种教材的优点,总结自己成功的教学经验,并结合目前数据库系统的特点,将理论知识和实用技术综合起来编写成这本教材。本教材涉及面广,难度适中,结构编排合理,适用性良好。

本书共分八章。第一章以介绍数据库基本概念和思想为主,是全书的基础。

第二章到第五章重点介绍目前最流行的关系数据库的有关理论知识,包括关系数据库基本概念、SQL 语言、关系代数、关系数据理论和关系系统的查询优化。

第六章和第七章介绍数据库的环境和数据库的设计知识。

第八章介绍目前应用最广的关系数据库管理系统 Fox BASE⁺。它以应用为主线,在介绍基础知识和基本方法的同时,强化了对实用技术的介绍。

为方便读者起见,还在附录中给出了 Fox BASE⁺ 2.10 版的全部命令和函数索引及说明。

本书第一、二、三、四、五、六、七章由袁晓洁执笔编写,第八章由徐雷执笔编写,并由袁晓洁对全书进行了统编和审阅。在本书的编写过程中得到了南开大学计算机与系统科学系的陈有祺教授、朱瑞香教授和韩维桓教授的指导与帮助,他们为本书提出了许多宝贵意见,对此我们表示深深的谢意。另外南开大学出版社对该书的编辑、校对、出版等工作付出了大量艰辛的劳动,给予了很大的支持和帮助,在此一并表示感谢。

由于作者水平所限,编写的时间仓促,本书错误之处在所难免,敬请专家和读者批评指正。

作者

目 录

第1章 数据库技术概念

1.1 文件技术与数据库技术	(1)
1.1.1 文件技术的局限	(1)
1.1.2 数据库技术的发展	(3)
1.1.3 数据库技术的特点	(5)
1.1.4 数据库基本概念	(7)
1.2 数据模型	(8)
1.2.1 概念模型	(9)
1.2.2 数据模型三要素	(12)
1.2.3 层次模型	(13)
1.2.4 网状模型	(17)
1.3 数据库系统结构	(21)
1.3.1 数据库的三级模式结构	(21)
1.3.2 数据库系统组成	(23)
1.4 数据库管理系统	(25)
1.4.1 DBMS 基本功能	(26)
1.4.2 DBMS 的组成	(27)
1.4.3 DBMS 的评价与选择	(28)
习题	(30)

第2章 关系数据库

2.1 关系模型的基本概念	(31)
2.1.1 关系的通俗解释	(31)
2.1.2 关系的数学定义	(32)
2.1.3 关系模型	(33)
2.1.4 关系数据库管理系统	(35)
2.2 关系代数	(35)
2.2.1 传统的集合运算	(36)
2.2.2 专门的关系运算	(37)
2.2.3 关系代数表达式	(38)
习题	(40)

第3章 关系数据库标准语言—SQL

3.1 SQL 特点	(41)
3.2 SQL 数据定义功能	(42)
3.2.1 建立基本表	(42)
3.2.2 扩充修改基本表	(43)
3.2.3 撤销基本表	(44)
3.3 SQL 数据查询功能	(44)
3.3.1 简单查询	(45)
3.3.2 连接查询	(47)
3.3.3 嵌套查询	(48)
3.3.4 组函数	(49)
3.4 SQL 数据更新功能	(51)
3.4.1 插入数据	(51)
3.4.2 修改数据	(51)
3.4.3 删除数据	(52)
3.5 建立和使用视图	(52)
3.5.1 视图的建立与删除	(53)
3.5.2 视图查询	(54)
3.5.3 视图的更新	(54)
3.6 索引的建立与删除	(55)
3.6.1 索引的建立	(55)
3.6.2 索引的删除	(55)
3.7 SQL 数据控制功能	(56)
3.7.1 使用权的授予	(56)
3.7.2 使用权的收回	(57)
习题	(57)

第4章 关系数据库设计理论

4.1 关系模式的存储异常与数据依赖	(59)
4.2 函数依赖的基本概念	(60)
4.3 关系的规范化	(62)
4.3.1 第一范式	(62)
4.3.2 第二范式	(63)
4.3.3 第三范式	(64)
4.3.4 BCNF 范式	(64)
4.3.5 多值依赖	(66)
4.3.6 第四范式	(68)
4.4 模式分解的优劣	(69)
习题	(70)

第5章 数据库查询的实现与优化

5.1 概述	(72)
--------------	------

5.2 优化的一般策略	(75)
5.2.1 逻辑层优化的一般策略	(75)
5.2.2 物理层优化的一般方法	(76)
习题	(76)

第6章 数据库环境

6.1 关系完整性及安全性	(77)
6.1.1 安全性控制的一般方法	(77)
6.1.2 SQL 系统的安全性	(79)
6.1.3 完整性约束条件	(81)
6.1.4 ORACLE 系统的完整性控制	(81)
6.2 事务处理	(83)
6.2.1 事务的基本概念	(83)
6.2.2 三个并行问题	(84)
6.2.3 并发控制	(84)
6.2.4 恢复策略	(85)
习题	(86)

第7章 数据库设计概论

7.1 规划	(87)
7.2 需求分析	(88)
7.3 概念结构设计	(90)
7.4 逻辑结构设计	(91)
7.5 数据库物理设计	(91)
7.6 数据库实现	(94)
7.7 数据库运行和维护	(94)
习题	(95)

第8章 Fox BASE+数据库管理系统

8.1 Fox BASE+概述	(96)
8.1.1 Fox BASE+的配置与运行环境	(97)
8.1.2 Fox BASE+文件的结构	(98)
8.1.3 Fox BASE+命令语句规则	(99)
8.1.4 Fox BASE+的技术指标	(100)
8.2 Fox BASE+数据库操作	(101)
8.2.1 对数据库结构的操作	(102)
8.2.2 对数据库记录的操作	(105)
8.2.3 对数据库运算的操作	(117)
8.2.4 数据库之间的操作	(118)
8.2.5 对数据库文件的操作	(123)
8.3 数据类型、常量变量、表达式和函数	(125)
8.3.1 数据类型	(125)

8.3.2	常量和变量.....	(126)
8.3.3	表达式.....	(127)
8.3.4	有关内存变量的操作.....	(129)
8.3.5	对内存变量数组与数据库变量之间的操作.....	(132)
8.3.6	对函数的操作.....	(134)
8.4	Fox BASE+ 程序设计	(143)
8.4.1	命令文件的建立与运行.....	(144)
8.4.2	交互式命令.....	(147)
8.4.3	屏幕显示命令.....	(148)
8.4.4	选择结构命令.....	(153)
8.4.5	循环结构命令.....	(155)
8.4.6	事件处理命令.....	(157)
8.4.7	过程文件.....	(159)
8.4.8	内存环境.....	(161)
8.4.9	菜单程序的设计.....	(164)
8.4.10	工作方式和状态的设置	(170)
	习题	(171)
附录 A	Fox BASE+ 2.1 命令一览表(按字母顺序排列)	(175)
附录 B	Fox BASE+ 2.1 函数一览表(按字母顺序排列)	(185)

第1章

数据库技术概论

电子计算机在刚刚问世的前10年中,主要是用于解决科学研究和工程设计中的数值计算问题。这个时期它的技术与应用只被少数科学家所拥有。随着计算机自身不断地发展,计算机的应用由科学研究部门扩展到各种企业和行政部门。因此,在计算机的三大主要应用领域,即科学计算、数据处理和过程控制中,数据处理迅速上升为计算机应用的主要方面,数据库技术就是作为数据处理中的一门新技术而发展起来的。

数据处理是指对各种形式的数据进行收集、储存、加工和传播的一系列活动的总和。其目的之一是从大量的、原始的数据中抽取、推导出对人们有价值的信息以作为行动和决策的依据;目的之二是为了借助计算机科学地保存和管理复杂的大量的数据,以便人们能方便而充分地利用这些宝贵的信息资源。

数据库技术所研究的问题就是如何科学地组织和储存数据,如何高效地获取和处理数据。数据库技术作为数据管理的最新技术目前已广泛应用于各个领域,数据库系统已成为当代计算机系统的重要组成部分。

本章主要介绍数据库的基本概念与技术,它们包括:数据模型、数据库体系结构和数据库管理系统等。这些内容是学习后面几章的基础。

1.1 文件技术与数据库技术

在50年代到60年代之间,计算机不仅用于科学计算,而且已经大量用于数据处理方面。此时计算机外存储器有了磁鼓、磁盘等直接存取的存储设备,为计算机进行事务管理奠定了硬件基础;另外软件技术主要是数据结构设计和数据管理技术的研究也得到了迅速的发展,从而出现了专门的管理数据的软件,这就是所谓的“文件技术”阶段。

1.1.1 文件技术的局限

在文件技术阶段,数据管理的主要特点可概括为如下几点:

1. 外存储器成为计算机系统不可缺少的组成部分,用户可以随时通过程序对文件进行查询、修改和增删等处理;
2. 文件组织形式日益多样化,索引文件、直接存取文件、链接文件纷纷出现,既可以满足批处理应用的需要,又能有效地实现记录的随机存取;
3. 出现了专用数据管理系统的软件,即文件系统,它能对驻留在外存储器上的数据文件实施统一管理。这种专用数据管理软件构成了操作系统的一个重要组成部分。由于应用程序不再需要了解数据在存储介质上的实际地址,因而大大减少了程序设计的工作量;

4. 数据不再仅仅属于某个特定的程序,而可以重复使用。

文件技术的上述特点使得这项技术在 20 世纪 60 年代中得到了充分的发展,把计算机应用推向了一个新的高潮。但由于文件结构的设计仍然是基于特定用途的,程序仍然是基于文件的特定物理结构和存取方法编制的,因此,数据结构与程序之间的依赖关系并未根本改变,因而限制了它的进一步发展。

60 年代中期以后,计算机在数据处理领域的应用迅速发展,由个别部门的应用逐步发展成多个部门的普遍应用,由简单孤立的单项应用发展为彼此相关的复杂应用,从而使管理的规模更加庞大,数据量急剧增长,共享性也更强了。这就带来了数据管理上的一些新问题。现举一例说明。

某学校的学生处、教务处和卫生院均要使用计算机对学生的有关信息进行管理,但其各自处理的内容不同,如用文件系统实现,可按如下方式进行组织。

学生处要处理的信息包括:

学号,姓名,系名,年级,专业,年龄,性别,籍贯,政治面目,家庭住址,个人履历,社会关系
.....

为此,学生处的应用程序员必须定义一个文件 F1,该文件结构中的记录应包括上述几个数据项。

教务处要处理的信息包括:

学号,姓名,系名,年级,专业,课名,成绩,学分.....

显然,教务处的应用程序员需定义一个文件 F2,该文件结构中的每一记录包括以上几个数据项。

类似地,当卫生院要记录和处理学生的有关健康情况信息时,其创建的文件 F3 应包括下列数据项:

学号,姓名,系名,年级,专业,年龄,性别,身高,体重,健康状况.....

这样当上述三个部门共用某台计算机时,该计算机的外存中要同时保存 F1,F2,F3 三种文件,可这三种文件中均有学生的学号、姓名、系名、年级和专业等信息,因此重复的数据项达到了 1/3 以上,严重冗余。

数据冗余不仅浪费存储空间,更严重的是带来潜在的不一致性。由于数据存在多个副本,所以当发生数据更新时,就很可能发生某些副本被修改而另一些副本被遗漏的情况,从而使数据发生不一致,影响数据的正确性和可靠性。比如,某学生由于特殊原因需要转系,他从数学系转到计算机系,当学生处得到该信息后,将该学生所属的系名由数学系改为计算机系,因而 F1 文件中保存了正确的信息。但如果教务处和卫生院没有得到该学生转系的信息,或者没有及时更改 F2 和 F3 文件,这就造成了不一致,使得一个学生属于两个系,人们也就无法确认哪一个是正确的信息。由于数据的使用价值在很大程度上依赖其可靠性,所以,这种不一致的后果是不可忽视的。可以想象,当这种情况发生在军事、航天、金融等行业部门时,其后果是非常严重的。

另一方面,在传统文件技术阶段,文件是为某一特定应用服务的,应用程序和文件是一对一的(见图 1-1),这就造成了应用程序与数据结构过分地互相依赖,而且系统很难扩充。一旦数据的逻辑结构改变,就必须修改应用程序和文件结构的定义;反之,应用程序的改变,也就影响文件的数据结构的改变。比如,学生处在管理学生信息时发现,只登记学生年龄是不够的,还应记录下学生的出生年月,因此需在 F1 的文件中加一个数据项,这就必须修改 F1 的文件结

构,同时还要修改涉及 F1 文件的应用程序;另一方面,若用户原先是用 Pascal 语言编制的访问 F1 文件的应用程序,现想改用 C 语言,那么,F1 文件的数据结构也必须随之改变。由此可看出在文件技术阶段,扩充已有的系统是很困难的,因此,用户迫切希望应用程序能够不受数据结构变化的影响。

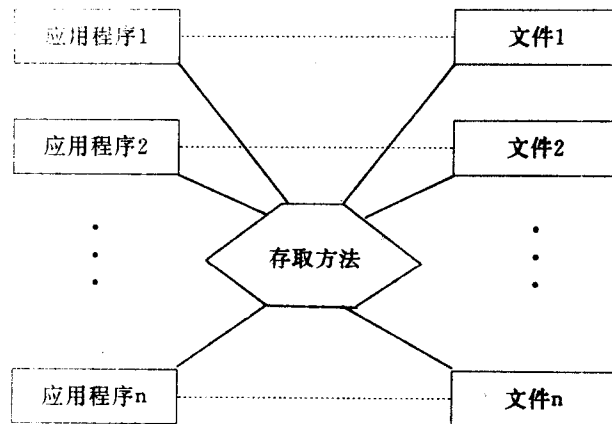


图1-1 应用程序与文件之间一对一关系

此外,文件系统缺乏对数据操作进行控制的方法,对于数据的安全性、保密性和正确性等方面的控制,完全要用户程序自己负责,这使得应用程序的编制相当繁琐。

综上所述,传统的文件技术有许多缺点,不能满足人们的要求,因此迫切需要新的数据管理技术来实现对数据的共享,实现数据与程序的独立性,并提供安全性和完整性,也就是说在操作系统之上必须有一个软件系统——数据库管理系统 DBMS(Data Base Management System),在建立、运用和维护时,对数据库进行统一控制,这就是数据库技术。

1.1.2 数据库技术的发展

在 50 年代初,数据处理进入文件处理阶段,实际上是数据库系统的前奏,到了 60 年代,数据库技术开始成形,从四个方面可反映出数据库技术的形成过程。

1. CODASYL 的建立与发展

1959 年 5 月,美国国防部召集了有计算机厂家、用户和政府其它部门等的专家 40 余人,讨论开发事务处理通用语言,并组成了数据系统语言协会(The Conference On Data Systems Languages 简称 CODASYL)。该协会在随后的十年里做了大量的事务处理软件开发和论证工作,其中主要的成果有:

(1)1960 年,设计并发表了新的语言 COBOL(Common Business Oriented Language)。

(2)1967 年,CODASYL 的 COBOL 协会专门成立了相应的组织,定名为数据库任务组 DBTG(Data Base Task Group),对语言问题进行研究。DBTG 小组于 1968 年发表了第一个工作报告:有数据库操作功能的 COBOL 的扩展。

(3)1969 年,CODASYL 委员会公布其研究成果 DBTG 报告,该报告确定并建立了数据库系统的许多概念、方法和技术。

(4)1971 年,CODASYL 之下又成立了新的组织——数据定义语言(DDL)委员会,并提出了数据定义的语言报告。随后 DBTG 对子模式定义语言和数据操纵语言作为 COBOL 语言扩

展内容作了研究并予以完成。

DBTG 所提议的方法是基于网状结构的,它是数据库网状模型的基础和典型代表。

2. 商用数据库管理系统的发展

60 年代初,许多公司开发了用于商用管理的软件系统,这些系统虽然还称不上是数据库系统,但却具有了现代数据库的特色。这些系统有:GE 公司的 IDS 商用系统;IBM Informatics 公司在 IBM System/360 上进行的 MARK IV 系统;MITRE 公司接受美国空军的委托开发的实验系统 ADAM(Advanced Data Management System);SCD 公司开发的 TDMS(Time-Shared Data Management System)系统等等。

与此同时,IBM 公司却迟迟没有对此项问题进行研究,直至 1965 年才开始提出 GIS(Generalized Information System)。后来又于 1969 年发表了 IMS(Information Management System),它是一个 DB/DC(Data Base/Data Communication)的庞大系统。IBM 与洛克菲尔公司共同开发提出了基于自己的数据库模型:一种以层次数据结构为基础的数据模型及与其相适应的语言 DL/1(Data Language/1)。该产品有极大的竞争力。人们把 IMS 作为当前流行的三大数据库模型之一——层次模型数据库的代表。

3. 关系方法的发展

1970 年,IBM 公司 San Jose 研究所的 E. F. Codd 在美国计算机学会会刊“Communication of the ACM”上发表了题为“大型共享数据库数据的关系模型”(“A Relational Model of Data for Shared Data Banks”)的著名论文,为关系模型的讨论奠定了基础,并将数据库系统的研究推向新的历史时期,即从只以应用为主转向理论研究发展的新时期。

继 1971 年后,Codd 又继续发表了多篇关于关系模型的文章,把数学方法引入到数据库系统,使数据库系统有了坚实的理论基础。Codd 本人作为关系数据库的创始人和奠基人,获得了 1981 年 ACM 图灵奖。

自 Codd 提出关系模型数据库方法以后,1974 年 IBM 就提出了 SQL(Structured Query Language)语言,这是一种基于关系方法实现对数据库存取的语言。大约从 1975 年到 1979 年的五年期间,IBM 公司 San Jose 实验室成功地研制了一个实现 SQL 语言的关系数据库系统原型 System R,该系统在 IBM 370 上运行。之后,IBM 又将 SQL 语言引入到 DB2(IBM Data Base 2)中,配置在 MVS 上运行,并于 1983 年推出 DB2 产品,表明了关系模型数据库系统终于被人们所接受,并得到最终的肯定。

目前,有影响的关系模型数据库系统有许多,在此值得一提的是 ORACLE 公司的 ORACLE。从 1977 年开始该公司就准备实现 Codd 的思想,准备研制横跨在大范围计算机上可运行的数据库系统,它采用 SQL 语言,用 C 语言开发,使 ORACLE 不仅使用方便,速度快且移植性好。1979 年开始推出 ORACLE 第一版,它是由核心和一组支持软件组成的软件包。从开始研制并经过 10 年的改进,到 1986 年推出第 5 版,增加了与各种语言、决策支持工具以及网络通信软件的接口,形成了开放式体系结构,是当前唯一可以通用于大、中、小和微型计算机的产品;为不同类型计算机提供整体化和标准化软件环境,是 ORACLE 的最大特点。

关系数据库系统的出现,促进了数据库的小型化和普及化,使得在微型机上配置数据库系统成为可能。较流行的 dBASE-Ⅱ可在 8 位微型机上运行,经过几年完善已形成 dBASE-Ⅲ、Ⅲ plus、Ⅳ 等系列以及 FoxBASE,可分别运行于各类微机和高档微机上。ORACLE、UNIFY 等较完善的关系数据库系统也都有其微机版本。

4. 标准化和学术活动的影响

1969~1971 年期间, CODASYL 提案发表后, 对数据库系统标准化问题引起了很多讨论。1972 年美国标准化协会计算机信息处理部(ANSI/X3)下设的标准化计划委员会(SPARC), 成立了数据库控制系统研究班, 分别于 1975 年和 1978 年发表了两个报告。这些报告研究了当时数据库系统的现状, 从整个信息系统的接口观点出发进行分析, 对数据库系统开发和研究有很深影响。其中提出的数据库结构标准化建议, 即, 将数据库划分为: 用户层(外层)、概念层、内层(存储层)等三层已被人们所接受。

1986 年 ANSI 宣布 SQL 为关系数据库系统的标准化语言。结构的标准化和语言的标准化无疑对数据库技术的发展有着积极的意义。

我国数据库技术的发展大致始于 70 年代中、后期, 其标志是 1977 年的第一次数据库技术研讨会。自 1982 年起, 几乎每年由该学会下设的数据库学组召开一次全国性的学术交流会, 会议起到了交流我国数据库应用与有关科研成果、指导我国数据库学术发展的作用。经过十多年的努力, 我国数据库技术水平从初期的学习、理解、试探阶段发展到消化、改造、创新阶段。

1.1.3 数据库技术的特点

与文件系统相比, 数据库技术提供了对数据的更高级更有效的管理, 它有如下主要特点:

1. 数据共享性。数据共享的意义是多种应用、多种语言互相覆盖的共享数据集合。在传统的文件组织中, 每个数据文件是特定的应用所私有的, 如图 1-1 所示。而数据库是从整体观点来看待和描述数据的, 数据不再是面向某一应用, 而是面向整个系统, 如图 1-2 所示。这可以大大减少数据冗余度, 既节约存储空间, 减少存取时间, 又可避免数据之间的不相容性和不一致性。前面所举文件系统例子, 若从数据库角度进行组织, 则如图 1-3 所示。

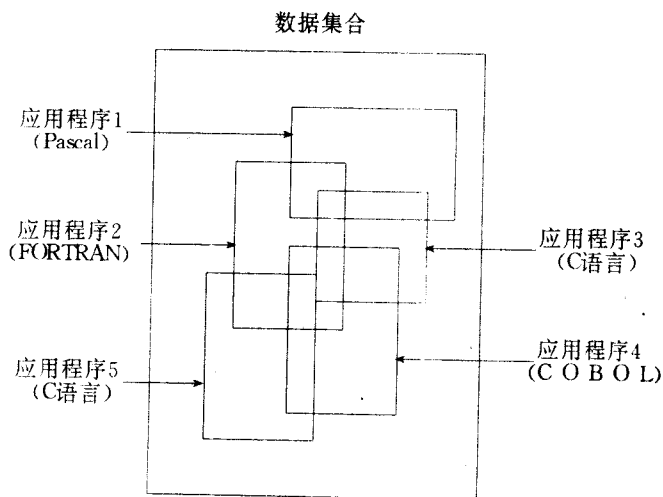


图1-2 应用程序共享数据集合

由该结构可看出, 数据不再分属于各个应用程序, 而是集中存储在一起。对一个特定的组织而言, 除了由于需要保密而强加的某些限制以外, 这些数据是为该组织的各个下属部门和各个方面的应用所共享的, 这样就大大提高了数据的使用价值。在图 1-3 中, 学生记录是为教务处、学生处和卫生院所共享的, 若某个学生需要转系, 则只要修改学生记录中的系名属性即可, 这样就不会出现不一致性的情况。另外, 除了共享的数据以外, 各部门还可以有自己的私有数

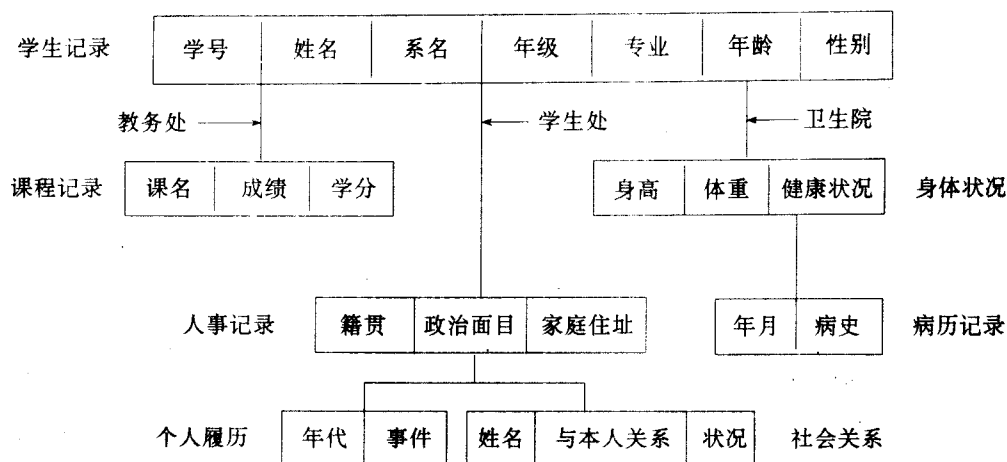


图1-3 共享数据范例

据,这些数据也放在数据库中,但为了防止其它部门对它进行访问,可加上各种保密限制,禁止非法获得这些数据。

由于数据库是高度结构化的,数据库中数据项之间及记录之间是有相互联系的,因此,当应用需求改变或增加时,只要重新选取不同子集或者加上一小部分数据,便可以有更多的用途,满足新的要求,使得系统很容易扩充,而文件系统则很难达到这一点。

2. 具有较高的数据和程序的独立性。在文件系统阶段,对文件的访问实际上是对存储数据的物理介质或设备进行访问,一旦更换物理介质或改变文件存储结构,则势必会影响应用程序。而在使用数据库时,应用程序对存储结构有较高的独立性,这种独立性是由系统在存储结构和逻辑结构之间提供的映像来获得的。当存储结构或者说物理结构改变时,只要相应地改变逻辑结构和物理结构之间的映像,就可以使逻辑结构保持不变,从而使建立在逻辑结构之上的应用程序也保持不变,这称为物理数据独立性。在当前硬件迅猛发展的情况下,存储数据的物理介质及设备不断更新换代,存储方法和存取策略也不断改善,具有物理数据独立性,可以使数据库的物理存储结构随技术发展不断加以调整,而不必修改应用程序。

另一方面,一个数据库系统所拥有的数据比某个特定的应用所需要的数据要多得多,因此对每个应用还要提供局部的逻辑结构,这种局部逻辑结构只是总体逻辑结构的子集,局部逻辑结构和全局逻辑结构之间使用映射进行联系。这样就可以做到当总体逻辑结构改变时,局部逻辑结构可以保持不变,而程序员根据局部逻辑结构编写的程序也可以不变,这就是所谓的逻辑独立性。例如,在图1-3中,假如学生处要登记每个学生的出生年月,那么可以在学生记录中加一个数据项(出生年月)。如果教务处和卫生院不需要访问该数据项,则其原来的应用程序可以不改变。提高数据独立性是数据库所追求的一个主要目标。在数据库中,数据与应用程序的关系如图1-4所示。

3. 对数据实行集中统一的控制。由于数据库中的数据为各种用户所共享,而计算机的共享一般是并发的,即多个用户可以同时使用数据库,因此数据库系统除了提供统一的数据定义、检索及更新操作手段外,还需提供控制数据安全性和完整性的方法,并能保障系统在并发存取数据时的顺利正确执行。

一个较完善的数据库系统通常提供以下三个方面的数据控制功能:

(1)数据的安全性控制。数据的安全性是指保护数据以防止不合法的使用所造成的数据泄

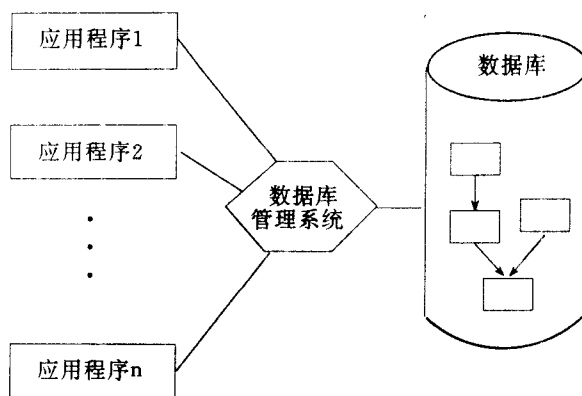


图1-4 数据库中数据与应用程序的关系

密和破坏,这就要采取一定的安全保密措施。例如,系统用检查口令或其他手段来检查用户身份,合格用户才能进入数据库系统;提供用户保密级别和数据存取权限的定义机制,当用户对数据库执行操作时,系统自动检查用户能否执行这些操作,检查通过后才执行允许的操作。

(2)数据的完整性控制。数据的完整性指数据的正确性、有效性与相容性。数据库中的数据是对客观世界中某些实体性质的反映,它有一定的语义含义,如性别的值只可能是“男”或“女”;会计记帐时,收支应当平衡,即收入之和减支出之和等于剩余数;月份是1~12之间的正整数;学生的学号是唯一的等等。数据库系统可以提供必要的功能,保证数据库中的数据在输入、修改过程中始终符合原来的定义和规定。

(3)并发控制。当多个用户的并发进程同时存取、修改数据库时,可能会发生互相干扰而得到错误的结果,并使数据库完整性遭到破坏,因此必须对多用户的并发存取动作加以控制和协调。

鉴于数据库系统的上述特点,大型复杂的信息系统大多以数据库为核心,因而数据库系统在计算机应用中起着越来越重要的作用。

1.1.4 数据库基本概念

本节主要给出与数据库相关联的几个基本概念。

1. 数据库(Data Base)

数据库是指在辅助存储器中的存储数据。这些数据是现实世界中的一些相关信息,它们在特定的组织(企、事业)中能为多种应用(或用户)服务。它们以数据模型所确定的数据结构方式存储,并能以有效的存取方法为操纵数据的语言提供快速响应。

2. 数据库管理系统(Data Base Management System)

数据库管理系统(DBMS)是数据库的一个软件,它允许一个或多个使用者对数据库中的抽象数据提出请求(包括询问和修改),并以合乎使用者要求的格式提供给使用者。抽象数据的含义是指数据由数据模型表示。若数据以关系数据模型存储,则相应的管理系统称为关系数据库管理系统。

3. 数据库系统

大量的经过加工整理的、存储在称作数据库中的数据,由数据库管理系统管理,为多个不同的应用(或用户)共同使用的数据处理系统,可称为数据库系统。

由此可见,数据库系统有两个主要组成部分:数据库管理系统和数据库。其中,当然也包括存放数据的存储介质及其设备和使用、维护数据库系统的人员。

数据库系统的研究包括三个方面:

(1)数据库管理系统的研究

包括 DBMS 应具有什么样功能的原理性问题和如何实现的技术性问题研究。当前, DBMS 的研究已从集中式数据库管理系统向分布式数据库管理系统 (DDBMS)、知识库管理系统 (KBMS) 等方面延伸,以及延伸到适应各种应用领域。

(2)数据库理论的研究

主要围绕关系数据库理论、事务理论、逻辑与数据库(演绎数据库)、面向对象的数据库、知识库等方面的研究,探索新思想的表达、提炼、简化,最后使其为人们所理解;研究新算法以提高数据库效率。

(3)数据库设计方法及工具的研究

数据库设计的主要含义是在数据库管理系统的支持下,按照应用要求为某一部门或组织设计一个结构良好、使用方便、效率较高的数据库及其应用系统。目前在这一领域是进行数据库设计方法、设计工具和理论的研究;数据模型和数据建模的研究;计算机辅助数据库设计方法及其软件系统的研究;数据库设计规范和标准的研究等等。

4. 数据模型

数据库的数据模型可以看作是一种形式化描述数据、数据之间的联系以及有关的语义约束规则的抽象方法。它规定数据如何结构化和一体化,以及规定对这种结构化数据进行何种操作。

当前流行的数据库系统中有三种主要的数据模型:层次模型、网状模型、关系模型。

5. 数据库语言

数据库语言由数据定义语言 (DDL) 和数据操纵语言 (DML) 组成,它为用户提供了交互式使用数据库的方法。

数据定义语言负责描述和定义数据的各种特性,用户通过使用数据定义语言可将数据库的结构以及数据的特性通知给相应的 DBMS,从而生成存储数据的框架,DDL 用来定义数据库的模式。

数据操纵语言实现对数据库数据的基本操作:检索、插入、修改和删除。

DML 有两类:一类是嵌入主语言中的,如嵌入 COBOL、C 等高级语言中,这类 DML 语言本身不能独立使用,因此称为宿主型的;另一类是交互式命令语言,它们语法简单,可以独立使用,所以称为自主型或自含型的。

1.2 数据模型

数据库中存储的是数据,这些数据反映了现实世界中有意义、有价值的信息,它不仅反映数据本身的内容,而且反映数据之间的联系。那么如何抽象表示、处理现实世界中的数据和信息呢?这就需要使用数据模型这个工具。数据模型是数据库中用于提供信息表示和操作手段的形式框架,它是我们将现实世界转换为数据世界的桥梁。

目前广为使用的模型大致可分为两种:一种称为概念模型(信息模型),它的建立独立于任