



数据库基础及

FoxPro
编程

(第二版)

孙志挥 刘国兴 赵翠霞

11.13
b(2)

东南大学出版社

592 TP311.13
G986(2)

数据库基础及 FoxPro 编程

(第二版)

孙志挥 刘国兴 赵翠霞



A0925709

东南大学出版社

内 容 提 要

本书分两篇共 17 章,系统介绍微型计算机数据库技术的基本原理(包括理论、方法、设计和维护)和先进的关系数据库管理系统 FoxPro(包括基本命令、数据库的建立、使用和管理、程序设计)。全书结构合理,理论与实践结合,论述清楚,便于自学。

本书可供大专院校有关专业作为数据库技术课程的教材或参考书,也可供企事业单位从事数据库管理应用系统开发的科技人员参考。

图书在版编目(CIP)数据

数据库基础及 FoxPro 编程/孙志挥等编. - 南京:
东南大学出版社,1999.6
ISBN 7-81023-696-2

I. 数… II. 孙… III. ①数据库-基本知识②关系数
据库-数据库管理系统, FoxPro-程序设计 IV. TP311.13

中国版本图书馆 CIP 数据核字(1999)第 18100 号

数 据 库 基 础 及 FoxPro 编 程

孙志挥 刘国兴 赵翠霞

*

东南大学出版社出版发行

(南京四牌楼 2 号 邮编 210018)

出版人:洪焕兴

江苏省新华书店经销 武进第三印刷厂印装

*

开本:787mm×1092mm 1/16 印张:15.5 字数:383 千字

1999 年 6 月第 2 版 1999 年 6 月第 6 次印刷

印数:33001-36000 册

ISBN 7-81023-696-2/TP·44

定价:19.50 元

(凡因印装质量问题,可直接向承印厂调换)

前 言

当前,人类社会已进入“信息社会”的新时代。作为信息处理最新技术的微机数据库也已广泛渗透到各行业的管理信息系统、办公自动化和辅助决策的应用领域中。众多需要微机数据库技术的部门都程度不同地面临着一个共同的问题:如何更有效地应用数据库技术?如何设计一个满足需求的数据库应用系统?解决这一问题的关键之一是:必须进一步提高数据库设计开发和维护人员的水平,使他们通过学习,系统地掌握有关微机数据库技术的基本理论和先进的软件工具。为此,我们曾经于 1992 年编写出版了《数据库基础及 FoxBASE + 编程》一书,作为江苏省计经委人才培训的教材。由于该书对数据库基础、原理和设计都作了较为全面简洁的阐述,因此在使用过程中,受到不少高校师生的欢迎,而被选作相应课程的教材,致使该书一印再印,这是编者始料未及的。鉴于目前 FoxPro 系统已取代 FoxBASE 而大量使用,广大读者提出了在该书基础上介绍 FoxPro 的要求,为此我们对原书进行了扩充和改写,并在内容组织上参照了江苏省普通高校计算机等级考试的大纲,以使编写的新书有更大的适用性。

数据库从它诞生起,就横跨了理论、应用和系统三大领域。学习数据库技术也必须在这三方面有所提高,做到数据库理论指导下开发和维护应用系统,以保证开发工作的质量,提高系统运行的效率。由此,本书编写的宗旨为:以微机关系数据库为重点,对数据库技术在理论、应用和系统方面作比较全面的介绍。本书第一篇以导论方式说明数据库的基本知识,分为 8 章来介绍数据库原理、关系数据库理论基础、数据库规范化方法、数据库设计及其实施和维护方面的内容。本书第二篇介绍一个先进的微机数据库应用系统开发软件,即关系数据库管理系统 FoxPro,分 9 章系统介绍 FoxPro 的语法、数据库文件建立和管理、数据组织和检索、多重数据库文件操作、生成器使用、RQBE 及 FoxPro 应用程序设计等。在本书编写工作中,我们力求体现下列特点:①在内容全面完整的基础上,能够为不同需求的读者服务。全书两大篇的内容既相互结合,又自成体系。第一篇可作为数据库概论的教材,第二篇也可作为编程工具的独立教材。②理论与实践结合。数据库设计和维护是一项实践性很强的工作,必须做到多练多用、理论与实践结合。书中讲述有关数据库原理和基本命令时,都配以必要的实例加以说明,并在每章后附有习题。③围绕重点,由浅入深,便于自学。本书以介绍微机关系数据库的原理和工具为主,叙述力求简洁。书中加“*”号的章节,是加深部分,在一般学习时可跳过其中内容,需要时再详细阅读。

参加本书编写工作的有:孙志挥(第一篇 1~8 章),刘国兴(第二篇 12~17 章)和赵翠霞(第二篇 9~11 章)。全书由孙志挥修改定稿。

在本书编写过程中,得到南京航空航天大学林钧海教授的指导,他仔细审阅了本书初稿,并提出了宝贵意见,在此谨向帮助与支持本书出版的同志表示衷心的感谢。

限于时间和编者水平,书中错误和不妥之处在所难免,敬希读者和同行指正为幸。

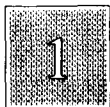
编 者

1999 年 4 月

第一篇

数据库基础

随着现代科学技术的迅猛发展,人类社会已经进入到一个充分利用信息资源的信息社会。诸如企业管理、情报检索、病理分析、计划统计等领域中的信息量之大、范围之广,变化之快,致使信息工程成为一门崭新的学科。这门学科中的数据库技术则是当今信息工程中最重要成果与工具之一。数据库能够有效合理地存储各种数据,为各个信息领域的信息处理提供准确、快速的数据信息。数据库是管理信息系统(MIS—Management Information System)、办公自动化系统(OA—Office Automation)和决策支持系统(DSS—Decision Support System)等应用系统的核心部分。为了更好地应用数据库技术解决各种实际问题,必须学习和掌握数据库系统的基本原理和技术。本篇将系统地介绍数据库的基本概念、各种数据模型,详细说明关系数据库的原理、理论和设计方法,最后介绍几个实用关系数据库管理系统。通过本篇学习,使读者能够掌握数据库的基本原理,为使用数据库设计工具,合理设计应用系统和有效维护数据库的工作打下基础。



数据库的基本概念

近 20 年来,由于计算机技术的发展和实用数据库管理系统的出现,我国各部门的计算机普及应用出现了一个崭新的局面,目前各种类型的计算机已广泛地应用在企业管理信息系统、办公信息系统、地理或地质信息系统、情报检索系统、银行信息系统、民航订票系统等部门。人们通常将以上应用系统称为管理信息系统,它是最大的计算机应用领域。与国外情况类似,管理信息系统在我国约占整个计算机应用的 70% ~ 80%,管理信息系统是计算机硬件、软件和管理科学技术综合应用的反映,也是促进计算机技术进一步发展的主要动力之一。

在管理信息系统中有一个共同的也是极为重要的问题,即如何对这个系统中大量数据进行有效、合理的存储,并为有关实际应用准确、快速地提供所需的信息。数据库及数据库技术就是针对解决这一问题而产生和发展的。目前数据库不但成为数据处理的重要工具,也是管理信息系统的核心技术,因此,从事信息系统设计、使用和维护的人员学习和掌握数据库的原理和技术是十分必要的。

1.1 计算机数据管理技术的进展

在计算机管理信息中,信息的具体反映是数据,而数据则是表达和传递信息的工具。因此,在这个意义上,我们所要讨论的信息管理系统的技术也就是管理数据的技术。

利用计算机进行数据管理的过程如图 1-1 所示。

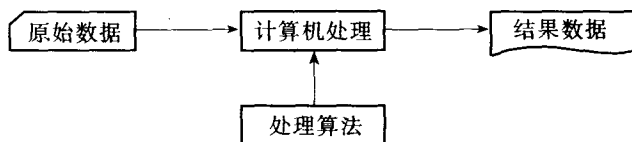


图 1-1 数据处理过程

在这个过程中,首先必须把原始数据和要对数据进行处理的算法输入到计算机,然后由计算机对有关数据按照给定的算法进行加工处理,最后按一定格式输出相应的结果。一般说,数据管理的基本内容包括:数据收集、数据转换、数据存储、分组和排序、数据筛选、数据检索和数据输出。由于在数据处理过程中所遇到的数据是有组织的,相互之间存在一定的联系,因此数据管理的效率往往与数据的存储结构和处理方式有密切的联系。从数据的存储结构和处理方式的角度而言,我们可以把计算机数据管理技术的进展分为四个阶段,并由此产生了四种不同的技术。

1) 自由处理阶段

这一阶段是计算机数据管理的初级阶段。当时的计算机用户在编制应用程序时,不仅没有软件工具可以利用,而且还要考虑很多与数据组织和存储有关的问题,例如数据的存储

结构、存取方式、输入/输出格式等。在这种程序设计中,数据和程序被紧密地结合为一个整体,程序员虽然可以“自由”编写对数据处理的指令,但数据已经成为程序不可分割的一部分,根本谈不上数据的“独立性”。由于一组数据对应于一个程序,程序之间又是相互孤立的,因而当程序之间出现重复数据时,这些数据也不能共享(如图 1-2 所示)。

在这种“程序-数据”的整体方式下,如果存储结构和设备发生变化时,相应的“程序-数据”的整体也要随之重新修改,因此严重地影响了计算机使用效率的发挥和提高。

2) 独立文件阶段

20 世纪 60 年代初期,出现了初等文件系统的构思和技术。在数据的文件组织中,把数据按一定的规定和标准组织起来,成为一个有效的数据组合体,每个数据组合体用一个“文件名”标识。用户在程序中需要引用文件中数据时,可以按照“文件名”来指定相应的文件。

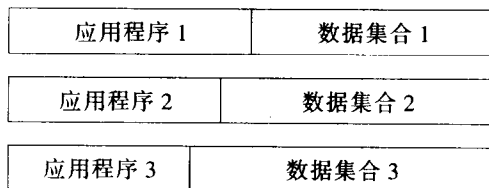


图 1-2 程序与数据为一个整体

在一般情况下,文件对数据的组织是这样约定的:文件是一组具有相同性质的记录的集合,记录由某些相关数据项组成。数据项又可以是一个组合项或初等项,初等项是数据的最小组成单位。由此可见,记录是相关初等项的有意义的组合,它能反映现实世界中的某一事件。当数据被组织成文件之后,就可以离开处理它的程序而单独存在,并将由一个称为“文件管理系统”的程序来专门对文件进行管理和维护。实质上,可以把文件管理系统看作是应用程序和文件之间的一个接口,图 1-3 说明了它们之间的关系,即应用程序通过文件管理系统建立和存储文件,同时通过文件管理系统来存取文件中的数据。在图 1-3 中,应用程序 1 通过文件管理系统可以使用数据文件 A1 和 A2;应用程序 2 通过文件管理系统可以使用数据文件 B;同样,应用程序 3 通过文件管理系统可以使用数据文件 C1、C2 和 C3。

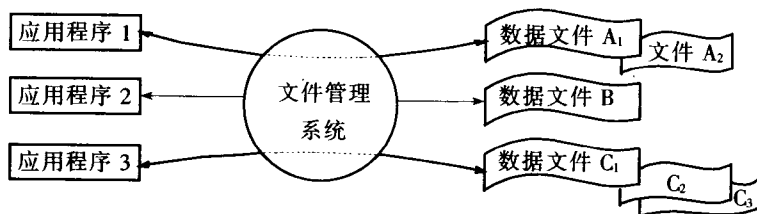


图 1-3 文件系统的接口使用

在数据管理上,通过文件管理系统对数据文件实行统一管理。这样,用户在应用程序设计时只需考虑数据的逻辑结构和特征,并按规定的组织方式和存储方法来建立和使用相应的数据文件。由于用户能够较少考虑数据物理存储方面的各种实现的细节问题,简化了应用程序设计的复杂性,因而对数据的管理也较“自由处理”方式更加合理、方便、有效。这是数据管理技术的一大进步。

但是,以上所管理的文件仍是用户专用的。它可能仅属于个别拥有该文件的用户,使得数据文件处于分散的状态,仍然可能造成大量数据的重复,数据不但冗余同时也浪费存储空间。此外,文件与应用程序之间还存在着相互依赖的关系,具有密切相关性,文件一旦离开它所依赖的程序便失去它存在的意义;数据文件之间彼此无联系,使得文件仅反映它所对

应事物的特征,却不能表达各事物之间客观存在的本质联系;文件之间还会由此产生不相容性;由于这是一种以程序为中心的处理方式,数据面向应用程序,当文件或程序有一方发生变化时就导致另一方发生相应的变化,否则程序与文件之间便会因不能适应而引起程序运行的各种错误。

3) 共享文件阶段

这一阶段中的共享文件方式是在独立文件基础上为解决数据共享而发展产生的一种新的文件方式,它的工作方式如图 1-4 所示。在这种文件方式中,所建立的数据文件是多个用户可能用到的数据,任何一个应用程序都可通过图 1-4 中所示的专门公用程序去访问任何一个文件中的数据,这样使得文件中的数据可为多个用户应用程序所共享。

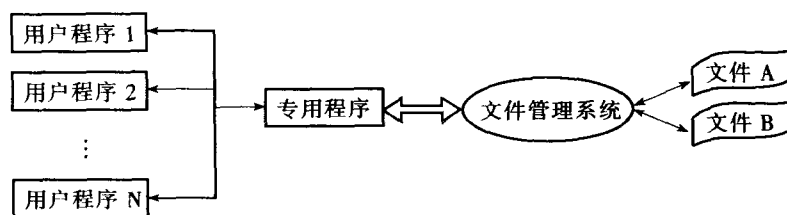


图 1-4 共享文件方式

以上所介绍的两种数据管理方式(独立文件和共享文件)均属于文件管理方法范畴。它们存在下列缺点:(1) 各数据文件之间存在着重复数据;(2) 应用程序在一定程度上依赖于数据;(3) 更新重复数据会产生数据不一致性。这些缺点使得建立在文件系统上的各个应用程序系统难于修改、扩充,数据的一致性难于维护,存储空间的有效利用率不高。随着社会的发展,客观世界各事物联系的日益复杂和信息量的不断增长,很显然,文件系统已不能适应和满足现代社会对计算机使用的要求。例如,美国 Rockwell 公司为执行阿波罗登月计划,于 20 世纪 60 年代研制的一个零部件生产计划管理系统是一个基于磁带的文件系统,其中有 60% 的数据是冗余的,它只能以批处理方式(先将一批数据记录在磁带上,再送入计算机处理的方式)运行,加之维护工作的困难,该系统的状况曾一度成为实现阿波罗计划的严重障碍。

4) 数据库方法阶段

计算机的软件工作者为了克服上述数据管理方法的缺点,经过长期不懈的努力,于 20 世纪 70 年代初提出数据库的概念。数据库技术为数据管理提供了一种完善高级的管理方式,实现了对所有数据实行统一的、集中的、独立的管理,使得数据的存储不依赖于使用数据的应用程序,从而极大地提高了应用程序的生产和运行效率。

数据库方式的出现有着多方面的促进因素,一是社会对于数据管理技术发展提出了新的需要,二是计算机技术的迅速发展(如大容量的快速存取的磁盘装置的出现,计算机性能/价格比的提高等)促使了数据库技术的产生和发展。数据管理技术进入数据库阶段有三件标志性的大事,它们是:

① 1969 年美国 IBM 公司的信息管理系统 IMS(Information Management System)作为最早的数据库系统的代表问世;

② 1969 年 10 月美国 CODASYL 的数据库任务组(DBTG)提出网络数据模型的数据库规范,并于 1971 年 4 月公布了它的研究成果《DBTG 报告》,确定了数据库设计的 DBTG 方法,真正把数据库与文件区别开来;

③ 1970年6月IBM公司SANJUSE研究所的高级研究员E.F.Codd发表论文《大型共享数据库数据的关系模型》,奠定了关系数据库的理论基础,该研究所研制的SYSTEMR是关系数据库系统的典型代表。

一般认为,从1970年到1981年是数据库的发展时期,而从1981年开始,数据库技术进入成熟时期。近20年来,数据库技术作为计算机科学和技术中的一个重要分支得到了惊人的发展,已经成为各种计算机应用系统的核心部分。

关于数据库方法的详细内容将在下一节讨论。

1.2 什么是数据库系统

在数据库的发展过程中,人们对“数据库”的定义和解释是众说纷纭的,从不同的角度对数据库有不同的理解。例如:

James Martin 强调数据的共享性,把数据库比喻为“各种数据用户都可去钓鱼的水库”;

J.D.Ullman 强调计算机系统,认为“数据库是存储在计算机系统的数据”;

C.J.Date 强调数据组织,在他的论文中说:“数据库是一个计算机化的簿记系统”;

T.J.Teorrry 强调可操作性,认为“数据库是可操作的数据集合”。

我们不再列举有关数据库的种种定义。总之,数据库是一门新的科学技术,在一般情况下,可以把数据库理解为是在一个或多个企业组织中,以最佳方式为多个用户服务的、在计算机上可运行的、有结构业务数据的集合。所谓企业组织,是指一个独立存在的单位,如公司、工厂、学校、银行、工程管理机构等。所谓业务数据,是指关于企事业营运的数据,如公司和工厂的生产计划调度数据、成本核算数据、银行账务数据、学校的教学管理数据等。这些数据,可以通过统计、调查、测试或从原始单据上得到。

数据库系统则是把有关计算机硬件、软件、数据集合和人员组合起来为用户提供信息服务的系统。这个系统可用图1-5来说明。

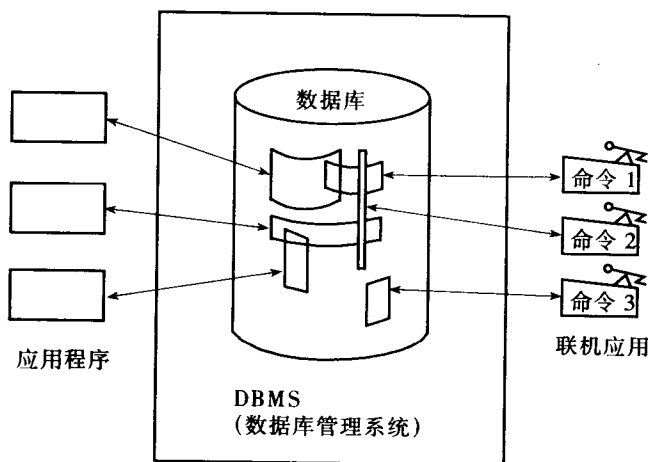


图 1-5 数据库系统示意图

(1) 数据集合

数据集合是某一企业组织中各种应用所需业务数据的集成,并为所有这些应用所共享。

这里“集成”是指若干不同性质数据文件的统一化。也就是说,为了便于管理和处理,数据存入数据库时必须具有一定的结构和组织方式。在数据库系统中,数据的定义与应用程序分开,对数据库的描述是独立的,因此数据库可以为多种任务(应用程序)所使用,达到共享数据的目的。此外,在构造数据库时,我们可以完全地或部分地消除有关文件中的数据重复,减少数据的冗余存储。例如,一个用于学校管理的数据库可能同时包括职工文件(编号、姓名、部门、工资等)和业务档案文件(编号、姓名、部门、技术级别、奖惩情况等),在这两个文件中的编号、姓名、部门三项数据没有必要在两个文件中重复存储,可将它们的对照关系存储在一个文件中(需要时可从该文件中查询到),供其它场合使用。由此可见,数据集成是减少冗余并实现共享的基础。

必须指出,“相关数据”、“一定的组织方式”和“共享”是关系数据库本质的三个要点。我们在后面还将详细讨论。

(2) 计算机系统

计算机系统这里指用于数据库管理的计算机硬件资源和基本软件资源。硬件资源包括中央处理机CPU、用以存放系统软件和数据文件的大容量存储器(目前主要指硬磁盘)和计算机网络等;软件资源包括操作系统、数据库管理系统、各种宿主语言(如C、COBOL、FORTRAN等)、软件开发工具以及应用程序等。

(3) 数据库管理系统(DBMS - Database Management System)

它是物理数据库本身(也就是实际存储的数据)和系统用户之间的界面。DBMS是管理和维护数据库数据的一组软件,其作用类似于仓库系统的管理机构。DBMS负责处理用户(应用程序)存取数据库的各种请求,如检索、修改和存储数据等操作要求。也就是说,在DBMS的控制下用户不能直接接触数据库,而只能通过DBMS存取数据,用户不必注重于数据的逻辑或物理的表达细节,而只需注意于数据的信息内容。数据库管理系统的另一功能是为管理员提供对数据库的维护手段,保护数据库中数据不受破坏。我们将在第2章中详细讨论数据库的功能和组成。

一个企业组织的所有数据资源如果被集中存储在某个大容量的硬盘设备上,这种数据库系统被称作集中式数据库系统;如果所有数据资源存放在位于多个不同地理位置的计算机的硬盘上,用户通过网络通信设备来访问这些数据,这样的系统被称作分布式数据库系统。

(4) 用户

用户是系统的服务对象。一般而言,一个数据库系统有三类用户:应用程序设计员、终端用户和数据库管理员(DBA - Database Administrator)。

① 应用程序设计员。他们是专业用户,具有较多的计算机专业知识,能够使用计算机系统所提供的各种程序设计语言(包括嵌入式DBMS语言)和开发工具编写应用程序,按应用需要对数据库进行检索、插入、删除和更新等操作(如工资计算程序、库存管理程序等)。

② 终端用户。他们可通过联机终端设备,使用由基本数据库命令语句组成的查询语句,对数据库进行检索、插入、删除和更新操作。这类用户通常是使用数据库的职能部门人员,如工程技术人员、企业各级管理人员等。有些用户常常存取数据库中的大量数据,处理并产生大量输出数据,故通常又称为批处理用户。

③ 数据库管理员DBA。他们是数据库的系统工程人员,负责对整个数据库系统进行维护,以保证数据库系统的正常运行,DBA在数据库开发、设计、运行、控制及维护等不同阶段

有不同的任务和职责,其详细内容我们在第7章中介绍。

1.3 数据库系统的主要特征

目前,数据库系统已成为数据处理业务部门用作管理其内部数据的主要工具,从微型计算机到大型计算机的厂商都竞相提供数据库管理系统的商品软件供用户选用,其原因正是由于数据库在数据处理中具有许多独特的优点。数据库不仅解决了文件管理方法中存在的主要问题,而且提供了强有力的数据管理能力和维护能力,显示了强大的生命力。数据库系统的主要特征有下列6点:

1) 数据的独立性

在数据处理中,数据库是在传统的文件系统基础上发展起来的,但是数据库中的数据与应用程序之间不存在相互依赖关系。也就是说,数据的逻辑结构、存储结构和存取方法等,不因应用程序的修改而修改,反之亦然。数据独立性通常分为两级:第一级为物理独立性,数据库的物理性质发生改变时(如更换存储设备、改变文件组织方式和存取策略等),不影响数据的逻辑结构,也不致引起应用程序的修改;第二级是逻辑独立性,这是指数据库总体逻辑结构的改变(如修改数据的定义、增加新的数据类型、改变数据间的联系等)无须修改原来的应用程序。同样,当用户需要修改应用程序时,也不强求数据结构作相应的改变。由此可见,数据独立性就是数据与应用程序之间互不依存性,我们把一个满足数据独立性要求的系统,称之为“以数据为中心”的系统,或称之为“面向数据”的系统,数据库系统就是具有这种特征的系统。

一个理想的数据库系统应该做到应用程序和数据库结构完全独立,以使得数据资源共享真正成为可能。但是,到目前为止,数据的逻辑独立性尚未能完全地实现,因为要达到这样的目标,将使数据库管理系统的设计变得十分复杂,设计难度和费用增加,因此实际的数据库系统往往是只具有一定程度的数据独立性。

2) 最小的数据冗余度

最小冗余度是指存储在数据库中的数据重复尽可能减少。在非数据库系统中,每个应用程序有它自己的数据文件,从而造成存储数据的大量重复,这给数据管理带来下列弊病:(1) 占用了计算机存储设备的大量存储空间;(2) 为了更新某些冗余复本,保持数据的一致性,必须执行多次更新操作,增加不必要的操作时间;(3) 由于数据的不同复本可能处于不同的更新阶段,因而可能给出数据不一致的信息。在数据库系统方式下,则是从整体观点来组织和存储数据的,数据是集成化、结构化的,数据统一存放、集中控制,因而能减少数据的冗余。必须指出,在数据库系统方式下,冗余度并不能完全消除。在实际应用中,有时为了某种原因而使用同一数据,在一定控制条件下的多次存储还是必要的。这时的冗余度被称作受控冗余度。

3) 最大的数据共享性

数据共享是促成数据库技术发展的重要原因之一,也是数据库最本质的特征。数据共享性可在下面几方面得到体现:

(1) 不同的应用程序可以使用同一个数据库,从图1-5可以看到,一个特定的应用总是只存取数据库中的一个数据子集,而不同的应用程序对应的数据子集可以相互重叠;

(2) 不同的应用程序可以在同一时刻去存取同一数据,这在数据库中称作“并发控制”

使用;

(3) 数据库中的数据不但可供现有的各个应用程序共享,还可以为新开发的应用程序使用,无需附加新的数据,实现新、老应用程序共享数据库中的数据;

(4) 应用程序可以用不同的程序设计语言或软件工具编写,它们可以访问同一数据库。终端命令语言是常用的一种语言,一个商品化的 DBMS 通常提供了与多种程序设计语言的接口。

4) 避免数据的不一致性

当本应相同的数据项在不同应用中出现了不同的值时,称之为数据的不一致性。例如同一职工的工资,如果在工资管理文件和人事管理文件中具有不同值,即出现了不一致性。数据的不一致性主要是由于数据冗余引起的,数据库在理论上可以消除数据的冗余,因而也可以避免数据的不一致性,即使存在某些冗余,数据库系统也提供了对数据的各种控制和检查,以保证在更新数据时同时更新所有的复本数据,从而保证数据的一致性。

5) 统一管理和控制

上一节已经说明了数据集成化的概念。数据集成化又称数据结构化,它要求数据库按照一定的数据模型(在下一章中介绍)来组织和存储数据,结构化的数据不但反映了数据之间的自然联系,而且可实现对数据的集中控制和管理。由此解决了数据资源共享、减少冗余等诸方面的问题,同时这种工作方式也带来了文件系统所不具有的一系列特性,如数据的安全性、完整性等。下面对这些概念作一简要说明:

(1) 安全性保护

数据的安全性主要指数据保密和防止数据的不合法使用。上面提到的数据资源共享并不意味着向一切用户开放所有的数据资源,如对一个应用用户,数据库系统将对他的权限进行安全性检查,只有合法的用户才能以规定的方式(如检索、修改等)对允许他使用的那一部分数据进行相应的操作。

(2) 并发控制

并发控制是指不同应用(应用程序、用户终端)在同一时刻期望存取同一数据时发生的事件。例如,在铁路或航空售票系统中,设有多个售票点分别为 X、Y、Z,若数据库中有某车次(或航班)余票项 n 为 15。当 X、Y 两点同时售票时,他们可以同时检索到有余票 n ,X 点按需要售出 10 张,将修改余票 n 为 5,但尚未将新值送入数据库时,Y 点又按 n 为 15 的情况售出 8 张,此时该售票业务必然出现差错,送入的 n 也成为不可预期的值,从而使 n 值失去了完整性。为了保证数据库的准确性,数据库管理系统的并发控制设施必须采用相应的封锁技术来避免上述错误的发生。

(3) 完整性

数据的完整性也就是数据的正确性,数据的不一致性是失去完整性的例子。前已述及数据冗余可能引起数据的不完整性,但是没有冗余的数据项同样可能出现不正确的值,例如表示月份的数据项出现大于“12”的值或职工的每天工作时间超过“24”。数据库管理系统的集中控制可以避免此类情况出现,它通过 DBA 定义完整性检查,对每一次更新的数据实施相应的检查,保证数据的完整性。必须指出,数据完整性检查对于多用户系统尤为重要,因为多个用户同时使用数据所引起对数据的干扰是经常大量发生的,数据库系统必须在这种情况下对数据的完整性进行必要的维护。

(4) 故障的发现和恢复正常状态

数据库系统在运行过程中,由于硬件/软件的故障及用户操作的失误,随时都有可能使数据库遭到局部性的或全局性的破坏,并且其所涉及的数据往往是无法用重新人工录入而复原的。数据库系统应提供一套设施和方法,来发现和警惕由于上述各种因素所引起的故障,并迅速地把数据库恢复到故障以前的正确状态,使系统正常运行。

(5) 可修改和可扩充性

一个实用数据库系统建立之后,需要逐步地扩充,因此要求整个数据库系统在结构和组织技术上具有易于修改和扩充的特点。

6) 可按标准化的规定实施数据库系统

随着信息处理自动化技术的发展,信息的标准化越来越得到有关方面的重视,各个团体、行业、部门、国家和国际标准正在不断制定和产生,标准化不仅能简化信息管理的过程,提高处理速度,而且使信息更具有通用性,能在更广泛的范围内实现通讯。标准化不论是在一个组织内部的还是国际性的,都使数据库系统内集中控制的实施成为可能。当然标准化的程度和层次越高,系统的实用范围越广范,在一定程度上,数据库技术和标准化技术是相互促进着发展的。

1.4 数据库的体系结构

数据库中用数据模型对现实世界的某一部分进行模拟,称之数据模式。对此,数据库有相应严谨的体系结构。本节简要介绍关于数据模式体系结构的概念。

1975年美国ANSI/SPARC数据库管理系统研制组提出了数据库三级结构的建议,即一个数据库的数据模式结构从逻辑上可分成外部级(External Level)、概念级(Conceptual Level)和内部级(Internal Level)三级(如图1-6所示)。由于该建议所提出的结构清晰,符合人们对客观世界特征的认识规律,因此三级结构已成为构架目前世界上各种类型和各种应用数据库系统的体系结构的依据。

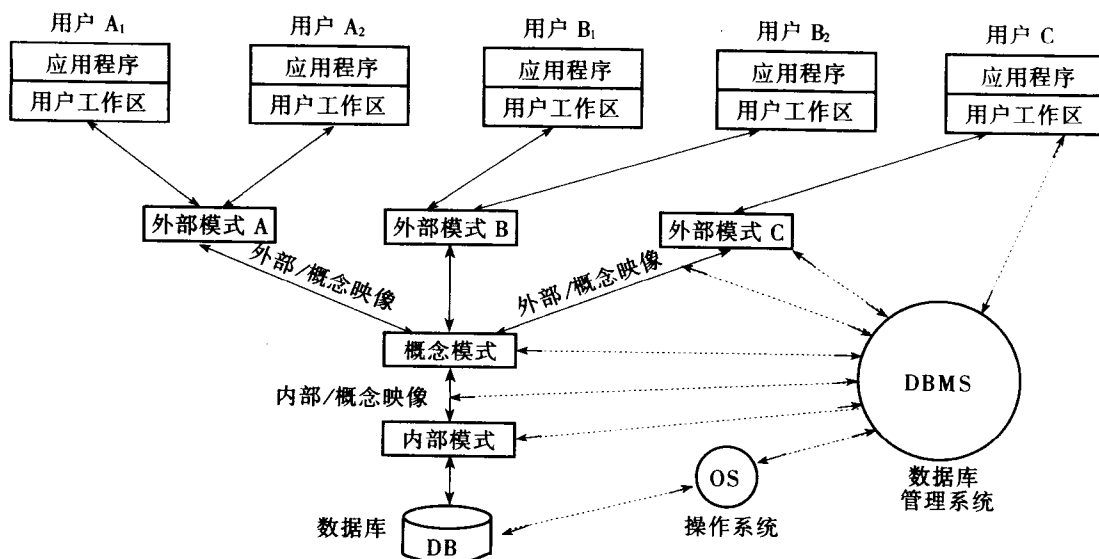


图 1-6 数据库系统结构

1) 外部级

外部级是三级结构中最接近于用户的级,它表示了各个用户对数据库的看法,这种看法称作外部视图或用户视图。每个用户必须以自己的观念使用一个外部模式(又称子模式)来对数据库的数据结构进行逻辑描述,即描述组成用户视图的有关记录的组成、相互联系、数据项的特征、数据的安全性和完整性约束条件等。

每个用户必须使用一个子模式,但不同的用户也可以共用一个子模式,即一个子模式可以为多个用户使用。这里所讲子模式是模式的一个子集,子模式由数据库管理系统提供的描述语言书写。

2) 概念级

概念级是介于外部级和内部级之间的一个中间级,它是整个组织对数据库的视图,即整个数据库的抽象表示,又称概念视图。一个数据库允许有许多不同的外部视图(用户视图),但只有一个概念视图。概念视图由概念模式定义,它描述整个数据库的逻辑结构,包括现实世界中实体的性质与联系、定义记录、数据项、数据的完整性约束及记录之间的联系。概念模式通常简称模式。从一个模式可以导出多个不同的子模式。

3) 内部级

内部级又称内部视图或存储数据库,它表示了数据的实际存储方法,是三级结构中最接近物理存储的级。内部视图由内部模式描述,包括各种存储记录的类型,索引及索引字段、记录的物理存储顺序等。

上述三级结构之间可通过 DBMS 中的软件,按一定的对应规则进行相互转换,从而把这三级结构连结为一个整体。这种对应规则称为“映像”。数据库的三级结构和它们之间的映像是实现数据独立性的有力保证。

1.5 数据库系统的构成

1) 数据库系统

在前面的讨论中,我们已经指出一个数据库系统包含了数据集合、计算机硬件/软件 and 用户等主要组成部分。它们之间的关系可用图 1-7 说明,其中:

- (1) DB 指集成化、结构化的数据集合;
- (2) 硬件包括整个系统运行所必须的硬设备;

(3) 软件包括 DBMS、操作系统(OS)、应用程序和应用软件包等。我们把 DBMS 看成是操作系统的一个特殊用户,它向操作系统申请所需要的硬件/软件资源,并受操作系统的控制和调度。操作系统是 DBMS 与硬件之间的接口,是 DBMS 的基础。

2) 数据库管理系统(DBMS)

DBMS 是一种负责数据库定义、建立、操纵、管理和维护的软件系统。其任务是很明确的,即保证数据安全可靠,提高数据库应用的简明性和方便性。一个功能完善的 DBMS 一般应具有如下的功能:

- (1) 定义和建立数据库;具有定义数据库三级视图的能力,并提供相应的工具以便描述

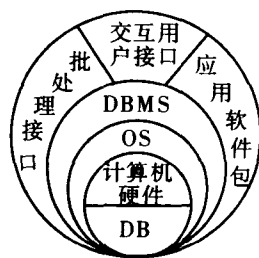


图 1-7 数据库系统的组成

三级模式。把各种源模式翻译成计算机目标代码形式,把关于数据库中数据的描述信息保存在数据字典之中;并把实际数据库存储到物理存储设备上,完成目标数据库的建立工作。

(2) 数据库的操纵:是直接面向用户(应用程序)的功能,即接收、分析和执行应用程序对数据库提出的各种操作请求,完成对数据库的检索、插入、删除和修改等操作。

(3) 数据库的控制:执行对访问数据库时的安全性、完整性检查和数据共享的并发控制,保证数据库的可靠性和可用性。

(4) 数据库的维护:包括运行时记录工作日志,评价分析数据库的运行性能,为保证数据库的运行效率而对数据库进行必要的重新定义和重新组织等。

(5) 故障恢复:当数据库系统在运行中发生故障时,DBMS 提供相应的恢复功能,把故障造成的影响限制在最小范围,并能以最快的速度排除故障、恢复和重新启动系统进入正常运行状态。

3) 数据描述语言和数据操纵语言

(1) 数据描述语言

数据库中的数据是按一定的结构集成的,数据描述语言(DDL - Data Description Language)是定义和描述这种结构的工具。它可以正确地定义和描述数据库中的数据及数据之间的联系,以便 DBMS 根据这些定义和描述既从内部的存储记录导出整体的概念级记录,又从整体的概念级记录导出应用程序所需要的外部记录。或者相反,从应用程序所需要的外部记录推知其所属的整体的概念级记录及对应的内部存储记录。

由于数据库系统在结构上分为三级,对一个数据库的结构描述也应当分为子模式、模式和存储模式三级。但目前并不是所有的 DBMS 都能独立地提供这三种不同的数据描述语言,往往是把三级描述的语言结合在一起,统称为 DDL。

(2) 数据操纵语言

数据操纵语言(DML - Data Manipulation Language)是用户的应用程序设计语言与 DBMS 之间的接口。它是 DBMS 向用户提供的一组宏指令或调用语句,用户可以使用 DML 向数据库管理系统提出对数据库中的数据进行各种操作请求。如:

①从数据库中检索满足给定条件的数据,它可以是一个或几个数据项,一个记录或一组记录;

②把新的数据插入到数据库中的指定位置上;

③删除数据库中没有必要保存的原有数据和记录;

④更新某些已经发生变化的数据项的值,以便正确反映当前的数据状态。

DML 一般具有操作表达明确,功能齐全,使用方便的特点。

DDL 和 DML 分别说明数据库对象的定义和操纵,它们是数据库的用户设计应用程序时所用的程序设计语言的一个子集,称为数据子语言(DSL)。本书第二篇将详细介绍数据库管理系统 FoxPro 的 DDL 和 DML。

4) 内含系统与宿主系统

一般说,提供给用户进行应用程序设计的 DML 有宿主型和内含型两类,与其相应的数据库管理系统分别称为宿主系统和内含系统。

(1) 宿主系统(Host - based System)

在这种系统中,数据库应用程序的程序设计语言由一般的程序设计语言和数据操纵语言(DML)两部分组成。数据操纵语言嵌入到应用程序需要对数据库操作之处。包含 DML

语言的程序设计语言称为主语言。由于主语言是一般的程序设计语言(如 C、PASCAL、FORTRAN等),因此可设置相应全局变量、局部变量、中间变量,If - then - else 逻辑、循环和各种运算等非数据库设施,完成应用程序所需的各种过程性处理。由于 DML 是整个语言的一个子集,有时也称为子语言。

(2) 自含系统(Self-contained System)

自含系统使用自含型数据操纵语言,它由一组命令组成,每条命令可以独立地执行一种处理,其操作不但有数据库的操作,而且还有计算等非数据库的操作。事先把这些处理算法编成程序放在计算机中,用户使用这些命令时不用考虑何时访问及如何访问数据库等一系列数据操纵问题,只须提供相应的调用参数。自含操纵语言是非过程性的语言,这种语言一般由用户在联机终端上以交互方式使用。现有的微型计算机数据库管理系统大多采用自含方式,如 Informix 和本书第二篇要重点介绍的 FoxPro 等。

习 题

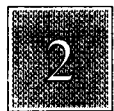
P1.1 什么是数据库和数据库系统,两者的差别何在?

P1.2 简述数据库系统的主要特点。

P1.3 什么叫数据的独立性,数据库系统如何体现数据的独立性?

P1.4 解释下列术语:概念模式、外模式、内模式,DBMS、DDL、DML、宿主系统、自含系统。

P1.5 画出数据库的三级体系结构并说明各级之间的关系。



数据库数据模型

上一章概括地介绍了数据库系统的基本知识,建立了数据库“是一个统一集中的数据管理机构”的概念。这一章我们进一步讨论数据库系统的基本原理,主要介绍数据库系统如何反映并转换现实世界中的各种复杂数据及其相互之间关系的有关问题。

2.1 现实世界的抽象描述

在数据库的理论研究和工程应用中,人们习惯于把数据处理中的信息转换分为三个范畴,即现实世界、信息世界和数据世界,并在此基础上讨论客观存在的信息及各种复杂关系反映到计算机及数据库中的过程。

2.1.1 现实世界

现实世界即客观存在的世界,它是人类社会存在和发展的环境。现实世界中实际存在着许多千差万别但又息息相关的“事件”,我们将一个实际存在的且又可以被识别的事件称为现实世界的“个体”。个体可以是具体事物,也可以是抽象概念。区分不同个体的依据是他们所具有的不同特征,如人的特征可以有社会(身份证)号、姓名、性别、年龄、职务、工资级别、社会关系等。我们正是利用这些特征来区分不同的人。每个个体可能有许多特征,他们从各个不同的侧面来表征特征,但是人们研究个体时总是选择感兴趣的那部分特征。仍以人为例,在人事管理中往往选择姓名、性别、年龄、政治状况、履历、社会关系等特征来作为区分个体的依据;而在财务工资管理中,则选择人的另外一些特征,如姓名、工资级别、工龄等特征。

在同一研究范围内的同种个体总是具有相同的特征项目,如上述人事管理和财务工资管理中分别所列特征项目。通常把具有相同特征项目要求的个体称为同类个体,所有同类个体构成一个集合,称为“全体”。例如所有的“职工”,所有的“学生”,所有的“课程”都各自构成一个全体。不论是个体或全体,他们之间还存在着各种关联,人们分析客观世界的个体或全体时,也是仅选择那些有兴趣的关联,没有必要也不需要去选择所有的关联。如教师与学生之间的关联有多种:教师“指导”学生(学习)、学生“评价”教师(授课质量)、教师“安排”学生(生活)……但在教务管理系统中仅选择前两种来反映教学活动管理的关联。

2.1.2 信息世界

信息世界是现实世界在人们头脑中的反映,是一种较为抽象与概念性的世界。在信息世界中用到一些基本术语,现介绍如下:

(1) 实体(Entity)

实体是被认识的个体在信息世界中的抽象表示。实体是信息世界中的基本单位。

(2) 属性(Attribute)

个体的每一特征在人脑中形成的认识称为属性。在信息世界中正是用属性来刻画相应