

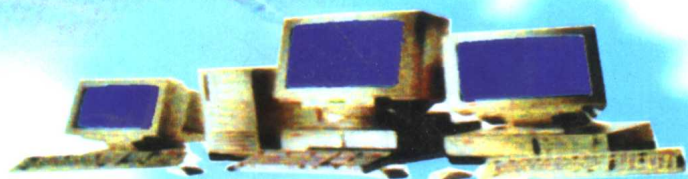
计算机专业大专系列教材

(第二版)

数据库原理及应用

——数据库基本理论与FoxPro的应用

李大友 主编



清华大学出版社

计算机专业大专系列教材

数据库原理及应用

——数据库基本理论与 FoxPro 的应用

(第二版)

李大友 主编

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书是根据计算机专业大专教学计划要求编写的。主要内容包括：数据库系统基础和数据库系统应用,共两篇。

其中基础篇包括：数据库系统导论、关系数据库模型、结构化查询语言 SQL、关系数据库设计理论和数据库设计。

应用篇包括：FoxPro 概述、FoxPro 基本操作、用“RQBE”进行快速查询、多重数据库操作、数据库信息查询、设计和生成报表、应用程序编制和应用示例。

全书相关部分均附有适量习题。

本书结构严谨、层次清晰、深入浅出、理论与实践紧密结合,是一本很好的大专教材。

本书不仅可作为计算机专业大专教材,而且可作为非计算机相关专业本科教材。

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：数据库原理及应用(第二版)

作 者：李大友

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：北京市昌平环球印刷厂

发行者：新华书店总店北京发行所

开 本：787×1092 1/16 印张：13.25 字数：304 千字

版 次：2000 年 5 月第 2 版 2000 年 12 月第 3 次印刷

书 号：ISBN 7-302-01139-7/TP·412

印 数：75001~90000

定 价：14.50 元

序 言

为什么要组织编写这套计算机专业大专使用的教材?根据什么来组织这套教材?这套教材的特点是什么?它能起到哪些作用?这就是我们在这篇序言中,要回答,也必须回答的问题。

计算机专业大专教育发展非常迅速,它反映了社会对这个层次人材的需求。在数量上已经超过了对本科人材的需求。大专这个层次有自己的特殊性,时间只有3年,要学习的内容很多,怎样精选教学内容,就成为十分重要的问题,他们又不同于中专层次,要求既有相当坚实的理论基础,又要运用理论解决实际问题,因此,如何处理好理论和实践的关系就十分重要。

大专这个层次人材的重要性是不言而喻的。但是,在培养大专这个层次人材的过程中,突出的矛盾之一,就是缺乏合适的大专教材。目前,不是用本科教材代用,就是很难及时获得所需教材。这就是组织这套专为大专使用的教材的起因。

那么,我们组织编写这套教材以什么为依据呢?中国计算机学会教育委员会与全国高等学校计算机研究会联合推荐的《计算机学科教学计划1993》是我们这次组织编写这套教材的主要依据。

“93计划”所提供的指导思想和学科内容不仅适合大学本科,也适合于大专的需要。

“93计划”明确规定了计划实施的目标:1. 要为“计算机学科”的毕业生提供一个广泛坚实的基础;2. 在培养人材的过程中,必须反映培养目标的差异;3. 要为学生毕业后,进一步学习新的知识和迎接新的工作挑战,做好理论和实践上的准备;4. 要学生能够把在校学到的知识,用到解决实际问题的过程中去。

在学科内容方面“93计划”概括了9个科目领域。9个科目领域组成“计算机学科”的主科目。每个科目领域都有重要的理论基础、重要的抽象(实验科学)、重要的设计和实现的成就。

这9个科目领域作为教学计划的公共要求,它们是:算法与数据结构、计算机体系结构、人工智能与机器人学、数据库与信息检索、人-机通信、数值与符号计算、操作系统、程序设计语言、软件方法学和工程。

我们根据上述指导思想和学科要求精选了13门课,作为大专用的主干教材。它们是:《数据结构》、《数字电路逻辑设计》、《计算机组成原理》、《微型计算机原理》、《微型计算机接口技术》、《计算机网络》、《数据库原理及应用》、《操作系统基础》、《汇编语言程序设计》、《C语言程序设计》、《软件工程概论》、《微机系统应用基础》和《离散数学》。

这13门教材大体上反映了除人工智能与机器人学和数值与符号计算之外的全部要求,足以满足大专主干课程教学的需求。

这套教材我们都是聘请大专院校有丰富教学实践经验的、工作在第一线的专家、教授编写。在编写过程中,充分考虑了大专的特点,在选材上贯彻少而精的原则,在处理上贯彻理论密切联系实际的原则。力求深入浅出,便于教学。并且在主要章节后均附有适量的习题。

这套教材适合于计算机专业大专使用,也可供非计算机专业的本科生使用。

李大友

1996.6

前 言

数据库技术是计算机学科中最重要的技术之一,已是计算机软件技术的一个独立分支,其应用领域十分广泛。

数据库技术包括数据库系统原理与数据库应用系统两大部分。前者是后者的理论基础,只有在正确理论的指导下,才能设计出好的数据库应用系统。

比较遗憾的是,多年来在数据库技术的教学实践中,两者结合得不甚理想。要么只讲理论,使读者感到很抽象;要么只讲使用方法,使读者感到缺乏理论指导,不知如何下手。

本教材试图使两者紧密结合在一起。首先给出必要的理论基础,然后在理论指导下进行实践。在数据库系统基础中,将数据库理论的最基础的部分提炼出来,做了深入浅出的论述,以指导应用系统的设计。在数据库应用系统中,以应用十分广泛的 FoxPro 数据库管理系统为工具,讲解了数据库应用系统的基本操作和程序设计的基本方法。

数据库系统基础篇主要介绍了计算机数据管理的发展概况、数据库系统的组成、数据和数据联系的描述、关系数据模型和关系运算、结构化查询语言——SQL 的查询方法和集合运算方法、数据定义、数据操纵、数据控制和数据字典、关系数据库设计理论:规范化问题、函数依赖和关系范式、数据库设计、需求分析、结构和实施。

数据库系统应用篇主要介绍了 FoxPro 概述、FoxPro 的基本操作,用“RQBE”进行快速查询的方法、多重数据库的基本概念和操作、数据库信息的查询方法、设计和生成报表、应用程序的编制和 FoxPro 应用实例。

本书层次清晰、结构严谨、深入浅出、理论与实践紧密结合,并附有适量习题。

本书不仅可作为计算机专业大专教材,而且可供非计算机专业本科生使用。

本书由李大友教授主编,参加编写的还有王利、于长云、王哲明、王定基和张颖等。

编 者

2000. 1

目 录

序言	I
前言	III

第一篇 数据库系统基础

第 1 章 数据库系统导论	1
1.1 数据与信息	1
1.2 计算机数据管理的发展	4
1.3 数据库系统的组成	8
1.4 数据和数据联系的描述	11
习题	16
第 2 章 关系数据模型	17
2.1 非关系模型	17
2.2 关系模型	20
2.3 关系运算	22
习题	28
第 3 章 结构化查询语言——SQL	29
3.1 SQL 查询	29
3.2 SQL 数据定义	38
3.3 SQL 数据操纵	42
3.4 SQL 数据控制	44
3.5 数据字典	45
习题	47
第 4 章 关系数据库设计理论	49
4.1 规范化问题	49
4.2 函数依赖	51
4.3 关系范式	55
习题	62
第 5 章 数据库设计	64
5.1 需求分析	64
5.2 概念结构设计	67
5.3 逻辑结构设计	71

5.4 物理设计与实施	74
习题	77

第二篇 数据库系统应用

第 6 章 FoxPro 概述	78
6.1 FoxPro for Windows 的运行环境	79
6.2 FoxPro for Windows 的启动和退出	79
6.3 FoxPro for Windows 的菜单系统	80
6.4 FoxPro for Windows 屏幕上的基本操作	81
第 7 章 FoxPro 的基本操作	86
7.1 数据库的建立、修改与浏览	86
7.2 数据库的排序和索引	96
7.3 数据库信息的统计和计算	100
第 8 章 用“RQBE”进行快速查询	106
8.1 通过“RQBE”窗口进行查询	106
8.2 保存查询	109
8.3 多种条件的组合查询	109
8.4 运行查询	111
第 9 章 多重数据库	113
9.1 “View”窗口	113
9.2 工作面板	114
9.3 命令按钮	116
9.4 利用“View”窗口建立相关数据库	117
9.5 浏览多个数据库的字段	118
9.6 视图文件	119
9.7 一对多关系的相关数据库	120
第 10 章 数据库信息的查询与统计	122
10.1 非索引查询	122
10.2 索引查询	122
10.3 多重记录查询	122
10.4 设置过滤器	123
10.5 在索引中建立查询	124
第 11 章 设计和生成报表	125
11.1 报表布局窗口	125
11.2 对象对话框	126
11.3 快速生成报表	131

11.4 移动对象和改变带区	133
11.5 存储报表设计	133
11.6 输出报表	134
第 12 章 应用程序的编制	135
12.1 用户编写程序的方法	135
12.2 屏幕生成器的使用	140
12.3 菜单生成器的使用	148
第 13 章 FoxPro 应用示例	156
习题	162
答案	166
附录 A FoxPro 命令汇总	167
附录 B FoxPro 函数汇总	178
附录 C FoxPro 2.5 错误信息清单	186
附录 D FoxPro 2.6 新增命令和函数	195

第一篇 数据库系统基础

第 1 章 数据库系统导论

数据库(database)是通用化的相关数据的集合,它不仅包括数据本身,而且包括关于数据之间的联系。数据库技术产生于 20 世纪 60 年代末、70 年代初。在计算机的三大主要应用领域(科学计算、数据处理与过程控制)中,数据处理约占 70%。计算机作为信息处理的工具,为适应数据处理需求的迅速提高,满足各类信息系统对数据管理的要求,在文件系统的基础上发展了数据库系统。数据库技术主要研究如何存储、使用和管理数据,是计算机数据管理技术发展的最新阶段。30 多年来,数据库在理论上、实现技术上,均得到很大的发展,研制出许多商用数据库管理系统,使得计算机应用渗透到各行各业的各类管理工作中。管理信息系统、办公自动化系统、决策支持系统等都是使用了数据库管理系统或数据库技术的计算机应用系统。

现在,数据库技术已成为计算机领域中最重要技术之一,是软件学科的一个独立分支。数据库方法原本是针对事务处理中的大量数据管理,但是它的应用范围不断扩大,不仅应用于事务处理,而且进一步应用到情报检索、人工智能、专家系统、计算机辅助设计等,涉及到非数值计算各方面的应用。应用范围的扩大又进一步促进了数据库技术的深入发展。可以说,数据库系统已成为当代计算机系统的重要组成部分。

我国自从第一次全国人口普查使用了数据库技术以来,随着微机数据库系统的推广,数据库应用逐渐普及,深受广大用户欢迎。计算机应用人员只有了解数据库系统的基本原理,掌握数据库设计的基础理论,熟悉数据库管理系统的特点,才能开发出好的数据库应用系统。

1.1 数据与信息

1.1.1 数据、信息与数据处理

在计算机应用中,数据处理和以数据处理为基础的信息系统所占的比重最大。一个国家现代化水平越高,科学管理、自动化服务的需求就越大。

人类的一切活动都离不开数据,离不开信息。在不同的领域里,信息的含义有所不同。一般认为信息(information)是数据和消息中所包含的意义。数据和信息有时可以混用,例如,数据处理也称为信息处理。有时必须分清数据和信息,例如,不能把信息系统称为数据

系统。

1. 数据的概念

所谓数据,通常指用符号记录下来的可加以鉴别的信息。数据的概念包括两个方面:其一,数据内容是事物特性的反映或描述;其二,数据是符号的集合。

由于描述事物特性必须借助一定的符号,这些符号就是数据的表示形式。例如,某人的出生日期是“一九七八年六月二十五日”,当然也可以将以上的汉字形式改为用“06/25/78”来表示。

数据的概念在数据处理领域中比在科学计算领域中已经大大地拓宽了。所谓“符号”,不仅仅指数字、字母、文字和其他特殊字符,而且还包括图形、图像、声音等多媒体数据;所谓“记录下来”也不仅是指印在纸上,而且包括记录在磁介质、光介质上和半导体存储器里。

数据在空间上的传递称为通信,在时间上的传递称为存储。

2. 信息的概念

信息是关于现实世界事物的存在方式或运动形态反映的综合,是人们进行各种活动所需要的知识。数据与信息既有联系又有区别。数据是载荷信息的物理符号或称为载体。数据能表示信息,但并非任何数据都能表示信息,正如人们常说的“如果计算机输入的是垃圾,输出的也会是垃圾”。同一数据也可能有不同的解释。由此,信息只是人们消化理解了的数据。信息是抽象的,不随数据设备所决定的数据形式而改变,而数据的表示方式却具有可选择性。

信息是反映客观现实世界的知识,用不同的数据形式可以表示同样的信息。例如,同样一条新闻在报纸上以文字的形式刊登,在电台以声音的形式广播,在电视上以图像的形式放映以及在计算机网络上以通信形式传播,其信息内容可以相同。当然,由于信息载体不同,喜闻乐见的程度不同,接收对象不同,所产生的效果或作用各异,又将另当别论。

3. 数据处理

数据处理是指将数据转换成信息的过程。广义地讲,它包括对数据的收集、存储、加工、分类、检索、传播等一系列活动。狭义地讲,它是指对所输入的数据进行加工整理。其基本目的是从大量的、已知的数据出发,根据事物之间的固有联系和运动规律,通过分析归纳、演绎推导等手段,萃取出对人们有价值、有意义的信息,作为决策的依据。由此可见,信息是一种被加工成特定形式的数据,这种数据形式对于数据接收者来说是有意义的。对数据的加工可以相对比较简单,也可以相当复杂。简单加工包括组织、编码、分类、排序等;复杂加工可以复杂到使用统计学方法、数学模型等对数据进行深层次的加工。

可以用下面的式子简单地表示出信息与数据之间的关系:

$$\text{信息} = \text{数据} + \text{处理}$$

数据是原料,是输入,而信息是产出,是输出结果。当两个或两个以上的数据处理过程前后相继时,前一过程称为预处理。预处理的输出作为二次数据,成为后面处理过程的输入,此时信息和数据的概念就产生了交叉,表现出相对性,如图 1.1 所示。“信息处理”的真正含义应该是为了产生信息而处理数据。

例如,一个人的“出生日期”是有生以来不可改变的基本特征之一,属于原始数据,而

“年龄”是用现年与出生日期相减而得到的数字,具有相对性,可视为二次数据。同样道理,“生产日期”与“购置日期”是产品和设备的原始数据,“失效日期”和“资产折旧”是经过简单计算得出的结果。

又如,用手工或计算机填写的发货单,对于发货部门的工作人员来说,即照单发货的信息,但对于仓储部门的管理者来说,它只是核算盘点库存量的原始数据。由于数据与信息之间存在着这种关系,这两个词有时被交替使用,其根本区别在于信息对当前或将来的行动或决策有价值。

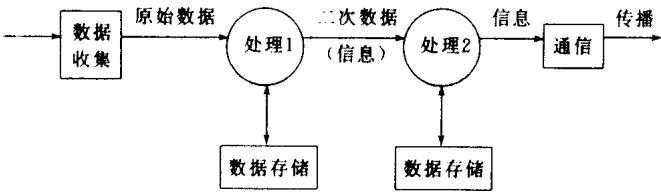


图 1.1 数据与信息的关系

1.1.2 信息系统

信息已经受到社会的广泛重视,被看做社会和科学技术发展的三大支柱之一(材料、能源和信息)。在一个组织里,信息已作为人力、物力、财力之外的第四种资源,占有重要地位。当然,信息是一种非物质的,是可以管理和控制的,是有别于基本资源的新形式的资源。当今社会已进入信息时代,管理人员在工作中多半时间花费在记录、检索和获取信息上。

1. 信息系统

信息系统是指为了某些明确的目的而建立的,它是由人员、设备、程序和数据集合构成的统一整体。信息系统的主要功能是提供信息,以支持一个组织机构的运行、管理和决策。更确切地说,信息系统将不适用的数据形式加工成可利用的形式。

一个信息系统的质量取决于它是否能及时地为用户提供所需要的信息。在一个组织机构中,不同阶层的管理人员因其管理的目标不同,所需要的信息也不相同。信息系统针对各个层次的需求,通过计算机实现信息支持,达到辅助管理的目的。

2. 信息系统分类

信息系统可分为三类:

- (1) 数据处理系统
- (2) 管理信息系统
- (3) 决策支持系统

数据处理系统 EDP(electronic data processing):用计算机代替繁杂的手工事务处理工作,其目的是提高数据处理的准确性、及时性,节约人力,提高工作效率。例如,计算机运行会计核算软件,对会计的“簿记”事务进行常规处理,提供数据查询,会计报表等功能,使会计部门的日常工作自动化。

管理信息系统 MIS(management information system):由若干子系统构成的一个集

成的人机系统,从组织的全局出发,实现数据共享,提供分析、计划、预测、控制等方面的综合信息。其主要目的是发挥系统的综合效益,提高管理水平。

例如,某企业管理信息系统由技术管理子系统、人事管理子系统、财务管理子系统、物资管理子系统、生产管理子系统、设备管理子系统和销售管理子系统组成。实现计算机管理能够迅速、准确地提供有关信息,不仅有力地支持各个职能部门的组织管理,并且通过信息共享加强了各子系统之间的协同,使整个系统有机地联系起来,同时为企业领导制定计划、确定经营目标、指挥生产提供信息支持,从而大大提高企业的综合效益,增强市场竞争能力。

应该指出,管理信息系统的结构和组织机构的影响是可逆的。现存的组织结构可以影响管理信息系统的设计,管理信息系统也可以影响未来组织机构的调整。由于计算机的引进使工作和管理环境为之一新,原来行之有效的制度、方法和监督机制在新的环境可能不再适用了,必须依据新环境的特点制定新制度,采用新方法,以至实施新的业务流程。

决策支持系统 DSS(decision support system):为决策过程提供有效的信息和辅助决策手段的人机系统,其主要目的是帮助决策者提高决策的科学性及有效性。

计算机辅助决策必须积累大量的数据、案例、方法和模型,更进一步还可以利用知识库系统和专家系统。决策支持系统的服务对象是面向某种决策问题的管理人员,它协助决策者在求解问题的过程中方便地检索出相关数据,对多种可选方案进行测试比较,然后做出决定。

这里需要强调指出,决策支持系统只能对决策提供支持,并不能由计算机代替人,自动地做出决定,人是决策行动的主体。例如,不同的管理人员运行同一套决策支持系统软件,结果可能做出不同的决策来。

1.2 计算机数据管理的发展

通过上一节对信息系统的简单介绍可以了解到,各级各类信息系统都需要大量的数据作为基础,数据处理的中心问题是数据管理。数据管理是指对数据的组织、分类、编码、存储、检索和维护。

与任何其他技术的发展一样,计算机数据管理也经历了由低级到高级的发展过程。计算机数据管理随着计算机硬件(主要是外存储器)、软件技术和计算机应用范围的发展而不断发展,多年来大致经历了如下4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
- ④ 分布式数据库系统阶段

1.2.1 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。当时在硬件方面,外存储器只有卡片、纸带、磁带,没有像磁盘这样的可以随机访问、直接存取的外部存储设备。软件方

面,没有专门管理数据的软件,数据由计算或处理它的程序自行携带。数据处理方式基本是批处理。数据与应用程序之间的关系如图 1.2 所示。

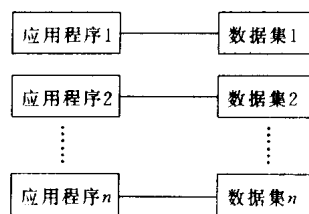


图 1.2 人工管理数据与程序的关系

这一时期数据管理的特点是:

1. 数据与程序不具有独立性

一组数据对应一组程序。这就使得程序依赖于数据,如果数据的类型、格式或者数据量、存取方法、输入输出方式等改变了,程序必须做相应的修改。

2. 数据不长期保存

由于数据是面向应用程序的,在一个程序中定义的数据,无法被其他程序利用,因此程序与程序之间存在大量的重复数据。

3. 系统中没有对数据进行管理的软件

数据管理任务,包括存储结构、存取方法、输入输出方式等完全由程序设计人员自负其责。这就给应用程序设计人员增加了很大的负担。

1.2.2 文件系统阶段

20 世纪 50 年代后期至 60 年代中后期,计算机开始大量地用于管理中的数据处理工作。大量的数据存储、检索和维护成为紧迫的需求。在硬件方面,可直接存取的磁鼓、磁盘成为联机的主要外存。在软件方面,出现了高级语言和操作系统。操作系统中的文件系统(有的也称为信息处理模块)是专门管理外存的数据管理软件。数据处理方式有批处理,也有联机实时处理。

在这一阶段,程序与数据有了一定的独立性,程序和数据分开存储,有了程序文件和数据文件的区别。数据文件可以长期保存在外存储器上多次存取,如进行查询、修改、插入、删除等操作。数据的存取以记录为基本单位,并出现了多种文件组织形式,如顺序文件、索引文件、随机文件等。

在文件系统的支持下,数据的逻辑结构与物理结构之间可以有一定的差别,逻辑结构与物理结构之间的转换由文件系统的存取方法来实现。数据与程序之间有设备独立性,程序只需用文件名访问数据,不必关心数据的物理位置。这样,程序员可以集中精力在数据处理的算法上,而不必考虑数据存储的具体细节。

数据的逻辑结构是指呈现在用户面前的数据结构。数据的物理结构是指数据在物理设备上的实际存储结构。例如,用户看到的记录是按照记录号顺序排列的,而实际上这些记录可能是分散存储在磁盘的不同扇区里,用链接方式组织在一起的。用户访问文件时只需给出文件名、逻辑记录号,而不必关心记录在存储器上的地址和内外存交换数据的过程。

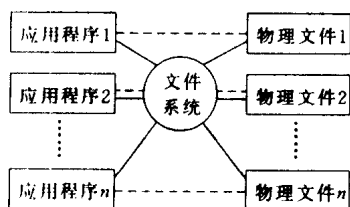


图 1.3 文件系统数据与程序的关系

此阶段数据与程序的关系如图 1.3 所示。

文件系统阶段对数据的管理虽然有了长足的

进步,但一些根本性问题仍然没有彻底解决,主要表现在以下 3 个方面:

1. 数据冗余度大

数据冗余是指不必要的重复存储,同一数据项重复出现在多个文件中。在文件系统下,数据文件基本上与各自的应用程序相对应,数据不能以记录和数据项为单位共享。即使有部分数据相同,只要逻辑结构不同,用户必须各自建立自己的文件,这不仅浪费存储空间,增加更新开销。更严重的是,由于不能统一修改,容易造成数据的不一致性。

2. 缺乏数据独立性

文件系统中的数据文件是为了满足特定业务领域某部门的专门需要而设计的,服务于某一特定应用程序。数据和程序相互依赖,如果改变数据的逻辑结构或文件的组织方法,必须修改相应的应用程序。同样道理,如果修改应用程序,如改用另一种程序设计语言来编写程序,也将影响数据文件的结构。

3. 数据无集中管理

除了对记录的存取由文件系统承担以外,文件没有统一的管理机制,其安全性与完整性无法保障。数据的维护任务仍然由应用程序来承担。

文件是无弹性、无结构的数据集合。所谓无弹性是指由于记录的内部结构由应用程序自己定义,而不是由系统来统一管理,因此对现有数据文件的应用不易扩充,不易移植,也难以增删数据项以适用新的应用要求。无结构是指各个数据文件之间是独立的,缺乏联系,不能反映现实世界事物之间的联系。

这些问题阻碍了数据处理技术的发展,不能满足日益增长的信息需求,这正是数据库技术产生的原动力,也是数据库系统产生的背景。应用需求和计算机技术的发展促使人们研究一种新的数据管理技术——数据库技术。

1.2.3 数据库系统阶段

从 20 世纪 60 年代后期开始,计算机应用于管理的规模更加庞大,需要计算机管理的数据量急剧增长,并且对数据共享的需求日益增强。大容量磁盘(数百兆字节以上)系统的采用,使计算机联机存取大量数据成为可能;软件价格上升,硬件价格相对下降,使独立开发系统维护软件的成本增加。文件系统的管理方法已无法适应开发应用系统的需要。为解决数据的独立性问题,实现数据的统一管理,达到数据共享的目的,发展了数据库技术。

数据库(database)是通用化的相关数据集合,它不仅包括数据本身,而且包括关于数据之间的联系。数据库中的数据不只面向某一项特定应用,而是面向多种应用,可以被多个用户、多个应用程序共享。例如,某个企业、组织或行业所涉及的全部数据的汇集。其数据结构独立于使用数据的程序,对于数据的增加、删除、修改和检索由系统进行统一的控制,而且数据模型也有利于将来应用的扩展。

为了让多种应用程序并发地使用数据库中具有最小冗余度的共享数据,必须使数据与程序具有较高的独立性。这就需要有一个软件系统对数据实行专门管理,提供安全性和完整性等统一控制机制,方便用户以交互命令或程序方式对数据库进行操作。

为数据库的建立、使用和维护而配置的软件称为数据库管理系统 DBMS(data base management system)。它是在操作系统支持下运行的。目前较流行的数据库管理系统包

括 Oracle, Informix, Sybase, 以上也有微机版本。微机上的数据库管理系统有 dBASE, FoxBASE, FoxPro, FoxPro for Windows, Access, CLIPPER, WINBASE 等。

现在, 数据库已成为各类信息系统的核心基础。在数据库管理系统支持下数据与程序的关系如图 1.4 所示。

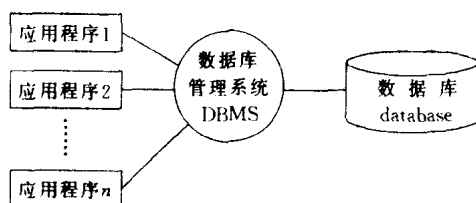


图 1.4 数据库系统数据与程序的关系

数据库系统的主要特点是：

1. 实现数据共享, 减少数据冗余

在数据库系统中, 对数据的定义和描述已经由应用程序中分离开来, 通过数据库管理系统来统一管理。数据的最小访问单位是数据项, 既可以按数据项的名称存取库中某一个或某一组数据项, 也可以存取一条记录或一组记录。

建立数据库时, 应当以面向全局的观点组织库中的数据, 而不能像文件系统那样仅仅考虑某一部门的局部应用。数据库群中存放全组织(如, 企业)通用化的综合性的数据, 某一类应用通常仅使用总体数据的子集, 这样才能发挥数据共享的优势。

2. 采用特定的数据模型

整个组织的数据不是一盘散沙, 必须表示出数据之间所存在的有机的关联, 才能反映现实世界事物之间的联系。也就是说, 数据库中的数据是有结构的, 这种结构由数据模型表示出来。

文件系统只表示记录内部的联系, 类似于属性之间的联系, 而不涉及不同文件记录之间的联系。要想在不同文件中查询相关的数据, 必须编写一个程序。

例: 有 3 个文件, 图书(总编号、分类号、书名、作者、出版单位、单价); 读者(借书证号、姓名、性别、单位、职称、地址); 借阅(借书证号、总编号、借阅日期、备注), 要想查找某人所借图书的书名、出版社及借阅者的职称, 则必须编写一段不很简单的程序来实现。

数据库系统不仅表示属性之间的联系, 而且表示实体之间的联系。只要定义好数据模型, 上述询问可以非常容易地联机查到。关于数据模型我们将在第 2 章中详细介绍。

3. 具有较高的数据独立性

在数据库系统中, DBMS 提供映像的功能, 确保应用程序对数据结构和存取方法有较高的独立性。数据的物理存储结构与用户看到的逻辑结构可以有很大差别。用户只以简单的逻辑结构来操作数据, 无需考虑数据在存储器上的物理位置与结构。

4. 有统一的数据控制功能

数据库作为多个用户和应用程序的共享资源, 对数据的存取往往是并发的, 即多个用户同时使用同一个数据库。数据库管理系统必须提供并发控制功能、数据的安全性控制功

能和数据的完整性控制功能。

1.2.4 分布式数据库系统阶段

20 世纪 70 年代后期之前,数据库系统多数是集中式的。分布式数据库系统是数据库技术和计算机网络技术相结合的产物,在 80 年代中期已有商品化产品问世。分布式数据库是一个逻辑上统一、地域上分布的数据集合,是计算机网络环境中各个结点局部数据库的逻辑集合,同时受分布式数据库管理系统的控制和管理,如图 1.5 所示。

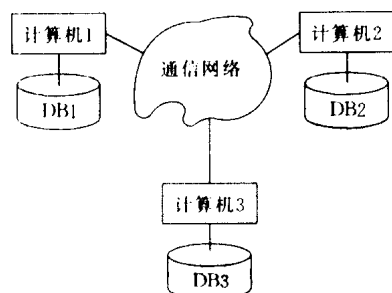


图 1.5 分布式数据库系统

分布式数据库在逻辑上像一个集中式数据库系统,实际上数据存储处于不同地点的计算机网络的各个结点上。每个结点有自己的局部数据库管理系统,它有很高的独立性。用户可以由分布式数据库管理系统(网络数据库管理系统),通过网络通信相互传输数据。分布式数据库系统有高度透明性,每台计算机上的用户并不需要了解他所访问的数据究竟在什么地方,就像在使用集中式数据库一样。其主要优点有:

1. 局部自主

网络上每个结点的数据库系统都具有独立处理本地事务的能力(大量的),而且各局部结点之间也能够互相访问、有效地配合处理更复杂的事务。因此,分布式数据库系统特别适合各个部门的地理位置分散的组织机构。例如,银行业务、飞机订票、企业管理等。

2. 可靠性和可用性

分布式系统比集中式系统有更高的可靠性,在个别结点或个别通信链路发生故障的情况下可以继续工作。一个局部系统发生故障不至于导致整个系统停顿或破坏,只要有一个结点上的数据备份可用,则数据是可用的。可见,支持一定程度的数据冗余是充分发挥分布式数据库系统优点的先决条件之一。

3. 效率和灵活性

分布式系统分散了工作负荷,缓解了单机容量的压力。数据可以存储在临近的常用结点,如果本结点的数据子集包含了要查询的全部内容,显然比集中式数据库在全集上查找节省时间。

系统易于实现扩展。例如,一个单位要增加新的机构,分布式数据库系统能够在对现有系统影响较小的情况下实现扩充。因此,扩大系统规模比集中式系统更加方便、经济、灵活。

1.3 数据库系统的组成

1.3.1 数据库系统的组成

数据库系统是指引进数据库技术后的计算机系统。例如,一个以数据库为基础的管理