



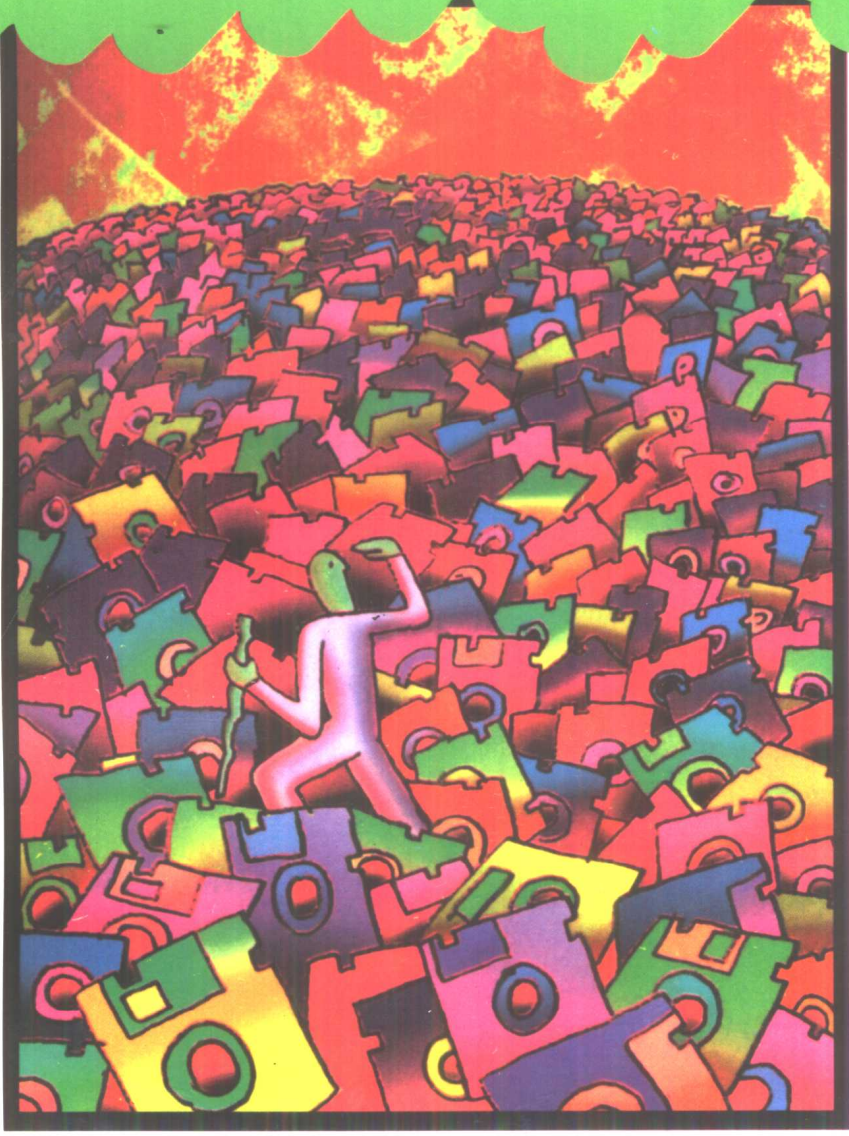
电脑入门普及丛书

数据库入门

电脑入门普及丛书

周屹 刘仁普 谢艳 编著

数据库入门



TP311.13



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
URL: <http://www.phei.co.cn>

电脑入门普及丛书

数据库入门

周 屹 刘仁普 谢 艳 编著

电子工业出版社
Publishing House of Electronics Industry

内 容 简 介

本丛书以普及电脑知识和入门为重点,具有主题突出,编排合理,深入浅出,便于自学,既有基本知识,又有各种使用功能和实际操作和指导,把知识的传授和实践结合起来,由易到难,逐步深入,使学习者能按本书循序渐进的,系统地学完各个专题,并能很好地掌握使用要领,使计算机走入社会各个阶层,充分发挥其效力,真正成为人们学习、工作和生活的好帮手。其内容包括:数据库系统的基本概念;DBASE 和 FoxBASE 数据库基本概述;开机工作方式和简单操作;如何建立数据库;数据库文件的编辑;DBASE 的函数和表达式;常用的数据处理(索引、排序、查询及统计);文件操作;工作方式和运行状态的设置;命令文件的建立和运行;数据库应用程序的常用设计方法和技巧。

本丛书是面向中学生及电脑爱好者自学,也可作为计算机培训班的培训教材。

丛 书 名: 电脑入门普及丛书

书 名: 数据库入门

编 著 者: 周 屹 刘仁普 谢 艳

责任编辑: 张荣琴 金 秋

特约编辑: 张成全

排版制作: 电子工业出版社计算机排版室

印 刷 者: 北京大中印刷厂

出版发行: 电子工业出版社出版、发行 URL: <http://www.phei.co.cn>

北京市海淀区万寿路 173 信箱 邮编 100036 发行部电话: 68214070

经 销: 各地新华书店经销

开 本: 787 × 1092 1/16 印张: 11.25 字数: 296 千字

版 次: 1998 年 1 月第一版 1998 年 1 月第一次印刷

书 号: ISBN 7-5053-4179-0
TP·1862

定 价: 14.50 元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页者,本社发行部负责调换

版权所有·翻印必究

目 录

第 1 章 数据库系统的基本概念	(1)
第 1 节 信息与数据	(1)
第 2 节 什么是数据库	(4)
第 3 节 数据库系统的构成和功能特点	(6)
第 4 节 关系数据库及建立过程	(8)
第 5 节 关系型数据库的基本概念	(10)
第 2 章 DBASE、FoxBASE 数据库的基本概述	(12)
第 1 节 DBASE 和 XBASE 系列数据库的历史与主要产品	(12)
第 2 节 XBASE 系统的运行环境和主要技术指标	(15)
第 3 节 XBASE 命令的语法规则	(16)
第 4 节 XBASE 系统的文件类型	(18)
第 3 章 开机工作方法和简单操作	(21)
第 1 节 开机工作步骤	(21)
第 2 节 进入和退出 DBASE	(22)
第 3 节 XBASE 的基本字符集	(23)
第 4 节 用 XBASE 进行简单的数学运算	(24)
第 4 章 数据库的建立	(26)
第 1 节 数据库文件的建立	(26)
第 2 节 数据库文件的打开和关闭	(30)
第 3 节 记录的输入和添加	(31)
第 4 节 记录的定位和插入	(35)
第 5 章 数据库文件的编辑	(38)
第 1 节 数据库结构和记录的显示查询	(38)
第 2 节 数据的编辑	(40)
第 3 节 数据库结构的修改	(46)
第 4 节 记录的删除和恢复	(46)
第 6 章 XBASE 的函数和表达式	(51)
第 1 节 XBASE 的常量和变量	(51)
第 2 节 XBASE 的运算符和表达式	(57)
第 3 节 XBASE 的常用函数	(61)
第 7 章 常用数据处理——索引、排序、查询及统计	(73)
第 1 节 排序	(73)
第 2 节 索引	(75)
第 3 节 数据的查询	(82)
第 4 节 几种常用的统计	(88)

第 8 章 文件操作	(96)
第 1 节 文件目录和文件复制	(96)
第 2 节 文件的改名和删除	(108)
第 3 节 数据库文件的连接和组合	(110)
第 9 章 工作方式和运行状态的设置	(122)
第 10 章 数据的输入和输出	(136)
第 1 节 单个字符的输入	(136)
第 2 节 键盘输入命令	(137)
第 3 节 字符串的输入	(138)
第 4 节 非格式化输出	(138)
第 5 节 格式化的输入输出	(139)
第 11 章 命令文件的建立和运行	(143)
第 1 节 应用程序的建立	(143)
第 2 节 命令文件的运行	(144)
第 3 节 条件语句	(145)
第 4 节 多分支选择语句	(147)
第 5 节 循环语句	(149)
第 6 节 子程序与过程	(151)
第 12 章 数据库应用程序的常用设计方法和技巧	(155)
第 1 节 应用程序的基本设计方法	(155)
第 2 节 应用程序的具体设计和常用技巧	(155)
附录 1:ASCII 码及控制字符表	(166)
附录 2:DBASE-III、FoxBASE 常用命令表	(168)

第 1 章 数据库系统的基本概念

本世纪 40 年代中期,第一台电子计算机在美国问世,从此数据处理的理由原来的手工处理阶段进入了自动化处理阶段。在开始的十余年里,计算机大多用于科学研究中的数值运算。到 50 年代末期,计算机开始用于事务管理。但是在整个 60 年代,由于技术的局限性和计算机本身昂贵的价格,计算机并没有在各种管理工作中得到广泛的应用。进入 70 年代,随着数据库技术的迅速发展和廉价的微型计算机的诞生和普及,计算机逐步进入了各行各业的管理领域并且越来越多地改变着世界的面貌。发展到今天,在全世界现有的计算机系统当中,用于数值运算的仅占很少一部分,大多数计算机都用于各种管理工作。计算机应用先进的数据处理技术,能够快速高效地处理大量信息,又可以将各种数据长期保存,这些特点都是其他技术无法取代的,计算机已经进入了日常生活和经济建设中的各个领域,为了学习计算机管理技术,我们首先要了解以下的基本理论和概念。

第 1 节 信息与数据

一、什么是信息

现在各行各业的人们都在越来越多地谈论“信息”这一概念。信息已经被列入了社会发展的三大科学支柱(材料、能源、信息)之一,成为现代社会文明和科学技术发展的重要标志。对于信息这个概念,到底应该如何理解呢?

信息是一个抽象概念,对于它的定义也有很多说法。从计算机管理理论的角度来讲,应该作如下定义:信息是人们的头脑中对现实世界各种事物的抽象反映,它是反映客观世界里各种事物特征和变化的知识。例如:河流的水位高低,飞机的航向、航速,学校中的在校学生总数,工厂的产品产量等等,都可以称作信息。

信息的产生是通过人的感官(眼、耳、鼻、舌等)的感觉和头脑的加工而形成的对于某件事物的概念。这里所说的事物,既可以是有形的具体物质,例如人、书本、桌椅等;也可以是无形的各种抽象概念和理论,例如“三角形三个角的角度之和等于 180 度”,这也是一条关于三角形的信息。

由于世界在不断运动和变化,而人类对于世界的研究和认识也在不断地发展和深入,所以反映事物特征与变化的信息也在不断地产生和更新。信息的作用在于它是人类从物质到意识的反映过程和人类直接和间接获取各种知识的重要中介,也是人类认识世界改造世界不可缺少的重要工具。从宏观的太阳系到微观的分子结构,从人类的悠久历史到现代的社会文明,所有这些都要通过信息去认识,信息普遍存在于自然界、人类社会和人的思维认识过程当中。信息也是一种宝贵的软资源。对于它的有效利用程度,已经成为衡量社会发展水平的重要标志之一。

二、信息的特点

按照不同的应用领域,信息可以划分为经济信息、社会信息、科技信息、军事信息等不同的种类。

各类信息都具有相同的特点。首先,信息是可以识别的。这种识别既可以是由人的感官实现的直接识别,也可以是通过各种探测手段和工具进行的间接识别。对于物体的大小、形状的认识仅通过肉眼就能实现,但是河水的流量,金属的硬度等信息的获取就必须通过特殊的手段间接完成。

其次,信息是可以转换的。信息可以以多种不同的形式存在,并且能够在不同的形式之间进行转换。例如,具体的物质信息可以转换成语言、文字、图象等形式,而语言文字也可以转换成广播电视中的电信号或计算机里的各种编码。

第三,信息是可以存储的。人对事物的记忆就是大脑对信息的存储。而计算机可以把大量的信息存储在磁盘、磁带、光盘或半导体芯片等多种存储介质当中,可以长期保存,反复使用。这些都是人脑无法实现的。

第四,信息是可以处理的。人类的思维活动就是人脑对于信息进行处理的过程。而计算机则要通过由人编制的各种系统软件来完成对信息的处理。

第五,信息是可以传递的。人们在日常生活中通过语言、动作、表情进行的相互交流就是信息的传递。在现代电子世界中,信息可以通过代码的形式在电报、电话、光纤、计算机等网络中快捷地进行传递、交换,实现信息资源的充分共享。

第六,信息是可以再生的。存在于人脑中的信息可以通过语言、文字、图表等形式再生成。同样,存储在计算机里的信息也可以通过显示、打印等方式实现再生。

三、什么是数据

任何具体事物,都要通过信息来反映和认识。但信息作为一种抽象的反映,不可能直接被计算机所接收。处理任何信息,都要通过数据来完成。在计算机科学中,“数据”一词的含义远远不是仅局限于原始意义的数值数据,任何可以由计算机处理的对象,包括文字、声音、图象等等,都可以称为计算机的数据。它是人们为了反映客观世界而记录下来的、可以被鉴别的符号。

数据一般可以分为表示一定数量概念的数值数据(内容为数字)和非数值数据(内容为文字,符号等)两大类。所有数据在计算机中都是以若干位二进制代码的形式来表示的。另外,同样的信息可以用不同的数据形式表现出来,但是它们所表示的信息内容却不会改变。例如:“这个班里有 40 个学生”和“*There are forty students in this class.*”就是以两种不同的数据形式表现出来的同样的信息。

信息和数据之间存在着内容和形式上的联系,数据是用来负载信息的数字、字母和符号,是表达和传播、交换信息所必需的工具;而信息则是对数据的解释,信息要通过数据的形式来表现出来。它只能依靠数据而存在,不可能独立出来。二者之间也有比较明显的不同,数据本身只是一些可以识别的符号,并不具有任何实际意义,只有对数据赋予了某些具体含义之后,数据才能成为信息,即信息是经过加工处理后并对客观世界产生影响的数据。

四、数据处理

在日常工作中,每天都有大量的关于不同问题的数据产生出来。由于这些数据涉及不同的领域,而数据之间也没有必要的联系,所以显得毫无规律,杂乱无章。这些表面化的数据并不具有任何价值。但是,通过对这些具体数据进行收集、转化、汇总、分析、计算等数据处理,就可以把大量的表面数据简化,提炼成能够反映事物本质和内在联系的有价值的信息。这些数据会为人类的生产、经营等社会活动提供强有力的参考。

例如,对于购物中心等商业企业,每天都要面对商品、销售、收入、支出等各种信息,产生成千上万的具体数据。通过对这些具体数据进行收集整理,分析计算,统计输出等处理,就可以得到准确反应企业经营、财务、商品情况的各种报表。经过处理后的总销售额,某种商品的销售数量及库存量,总的销售总收入和销售利润,税金等数据,将为企业制定销售策略,决定各种商品的进货渠道和数量,根据当前的财务状况确定财务计划等各种经营活动提供准确而可靠的依据。

通过上面的例子就可以看到,数据处理的主要目的包括以下几项:

首先,把收集到的各种原始数据经过分类整理和格式转换变换成为易于观察、分析,并且可以进行进一步处理的有规律的数据。

其次,要把大量的具体数据经过加工变为可以反映某种事物本质的、比较精炼的数据。只有这种数据才能够对人类的决策和行动产生影响,因而这也是数据处理的关键。

第三,还要把已有的数据进行储存,以备今后继续使用。

人类社会产生时,也就出现了原始的数据处理。到现在数据处理已经经历了手工、机械和电子化数据处理三个阶段。现在已经有大量的数据是通过计算机来处理的了。一般的电子化数据处理包括以下几个过程。

第一,数据的收集。根据用户的需要,及时地从产生数据的各个环节收集所有原始数据。有时还要对收集到的数据进行必要的分类和检验,以保证原始数据的正确性,完整性和时效性。

第二,数据的转换。为了使各种原始数据能够被计算机承认并接收,需要用格式变换或代码化的处理方法进行数据转换。对于某些与计算机所要求的格式有差异的数据,可以通过对其数据格式进行转换来达到要求。而对于一些文字信息则可以用比较简单的代码来表示。例如,可以用“T”来代表老师,而用“S”代表学生等等。

第三,数据的组织。对数据按照其内部联系和计算机软件的要求进行整理和编排,使数据在计算机中占用尽量少的内存和其他资源,并能够较快地被软件调用。

第四,数据的输入。将经过数据转换并已组织好的数据,按照设定的格式输入到计算机里。

第五,数据的处理。利用各种计算机应用软件,对输入的数据进行各种处理,包括索引、排序、统计、计算、修改、更新等操作,从而得到用户需要的结果。

第六,数据的输出。将数据处理的结果,按照用户要求的格式从计算机中输出,提供给用户。输出的结果可以是简单的数据,也可以是报表、图形等其他的形式。

第七,数据的存储。所有的相关数据,包括输入的原始数据,处理后的结果数据以及处理过程中的中间数据都会被计算机记录到磁盘、磁带等存储介质上,以备今后继续使用。

第2节 什么是数据库

一、数据库的概念

数据库是一种用于数据处理的比较先进的软件工程。简单地说,数据库就是“存放数据的仓库”,也就是提供数据的基地。数据库完成对所有数据的保存并负责用户或应用程序对存储数据的访问。利用数据库对数据进行结构化存储、集中管理和统一使用,具有数据存储结构化、数据冗余小、数据充分共享、程序与数据独立性高、数据操纵语言高级、易学易用等优点。目前数据库技术广泛应用于事务性管理,计算机辅助设计,工程计算等许多方面。

数据库里保存的内容,既不是输入数据也不是输出数据。收集到的各种原始数据经过格式转换并输入计算机以后,并不是毫无规律地全部堆放在数据库里,而是按照数据库的组织形式和数据结构的定义,以一种相对稳定的中间数据形式存放在数据库里。同样,输出的结果也不是数据库里存储的数据,而是根据存储的数据或者其中的一个子集处理生成的规定的结果数据。

数据库是有组织的信息集合,其中存储的数据通常都具有集成性和共享性。所谓集成性是指把与某些特定应用环境中的各种应用相关的数据以及数据之间的关系全部集中地按照一定的结构进行存储;而共享性是指数据库中的所有数据都可以为多个不同的用户所共享。即多个不同的用户,使用不同的语言,为了实现不同的应用目的,而同时存取数据库,甚至同时存取同一块数据。支持共享的数据库系统,称为多用户系统。大型机上的数据库系统一般都是多用户系统,而微机上的数据库系统大多是单用户系统。

此外,存储在数据库系统中的数据整体可以是一个数据库也可以划分成性质不同的多个数据库,前者称为集中式数据库,后者称为分布式数据库。

为了便于应用程序对数据的调用和访问,应当以最少的数据冗余(即对相同数据的重复保存)和最佳的组织方式将数据存储到数据库文件里。为此要根据不同数据项之间的相互联系和数据库的设计要求,对每个库文件的数据结构进行定义和对每项数据的内容格式进行详细描述;这一过程称为库结构的定义,任何数据都要严格地遵照库结构的定义存储在数据库文件里,并可以按照定义好的结构,向库文件里增加新的数据。由于这一定义过程是独立的并且是公开的,所以各种应用程序或用户都可以依照库结构自由的访问库文件里的所有数据,完全做到了用户共享。数据库文件与库结构定义和用户访问的关系可见图 1-1。

数据的管理还要涉及多种文件的组织方式,常用的文件包括数据库文件、排序文件、索引文件、标签文件等等。这些内容在今后的有关章节当中,都会加以介绍。

二、数据库管理系统

数据库管理系统(Data Base Management System,简称 DBMS),是一组用于管理和维护数据库中数据的软件的集合,是连接数据库和用户的桥梁。数据库管理系统能够维持数据库的正常运行,接受并完成用户对数据的直接访问和对数据库的各种应用要求。也就是说,用户不能直接操作数据库里的数据,而必须利用 DBMS 来实现对数据的存取访问或修改更新。通过 DBMS,用户可以完全不考虑数据库的物理或逻辑结构要求,直接找到数据的具体内容。同样,对于数据库的应用软件也离不开 DBMS,应用软件的产生要利用 DBMS 提供的各种工具,而应

用软件的执行也要由 DBMS 来解释和处理。

归纳起来,数据库管理系统主要完成以下几项主要功能:

- (1) 定义数据库
- (2) 管理数据库
- (3) 建立和维护数据库
- (4) 应用数据库

在数据库管理系统当中,以上功能都是由不同的系统程序完成的,每组系统程序都只完成相应的功能。下面介绍一下这些程序组的系统功能。

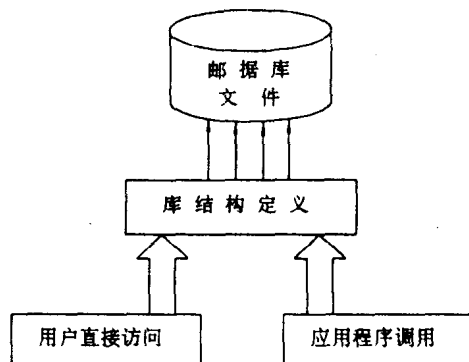


图 1-1 数据库文件库结构定义和用户访问的关系

1. 定义数据库

利用 DBMS 提供的工具,可以定义数据库的全部逻辑数据结构和局部逻辑数据结构,以保证其一致性和完整性;还可以完成数据的物理存储结构定义和索引,排序方式的设定。也就是说,数据库所有最基本的结构都要通过 DBMS 来定义生成。

2. 管理数据库

DBMS 对数据库的管理包括两方面的内容:一方面是对数据库文件中存储的数据进行管理,另一方面,还要对整个数据库系统的运行和工作进行控制。

在管理具体数据时,DBMS 要接收来自应用程序或用户直接发出的操作命令或访问请求,并据此对数据进行存取访问、插入、修改、添加、删除等操作。在网络数据库里,还要实现对数据安全性和完整性的控制,以及并发控制等。而在控制整个数据库系统时,则要借助系统总控程序来完成。由系统总控程序来调度,协调数据库中所有程序的运行活动,使它们能够按照一定的顺序依次执行,从而保证数据库系统的正常工作。在多个程序运行时,系统总控程序还要完成不同程序之间和程序与数据库管理系统之间的通信和数据交换。

3. 建立和维护数据库

借助于 DBMS 的工具,设计人员可以按照总体设计的要求建立某个具体的数据库,也可以对某个数据库文件建立其索引文件和排序文件。另外,为了保证数据的安全性和系统的正常工作,还要定期对数据进行一些经常性的维护工作,如数据的备份,重新装入和重新组织等;这些功能也是由 DBMS 负责完成的。

4. 应用数据库

用于微机上的 DBMS,一般都比较简单。所以都提供了一种非常高级(即非常类似于某种自然语言,如英语)的操作语言的解释系统。利用这些高级操作语言所支持的各种命令,函数和语句,不论是专业的程序设计人员还是普通用户,都可以设计出自己的应用程序,并通过 DBMS 解释执行。也就是说,借助于 DBMS 才能把数据库中的数据利用起来;否则,离开了 DBMS 的数据库不具有任何实用价值。

从上面的介绍我们可以看到,数据库管理系统要完成从数据库的设计、建立到管理、维护、应用的一切工作。DBMS 是整个数据库系统的核心,离开了它,数据库也就不存在了。

三、应用程序

就像程序设计人员利用 Pascal、Fortran、C 等高级语言编写各种程序完成科学计算一样,仅

靠数据库管理系统是不够的。用户还要通过 DBMS 提供的程序设计工具设计出自己需要的应用程序,经过调试或编译之后运行,直到生成最后的执行结果。在每个数据库系统当中,都会有多个应用程序共享当前的数据,完成各种任务,这是数据库系统的最终用途。

第 3 节 数据库系统的构成和功能特点

实现信息的电子化管理仅靠数据库还不够,它要以计算机系统为基础,同时还时刻离不开程序设计人员和用户,三者的有机结合才形成了用于信息电子化管理的完整系统。

一、计算机系统

这里讲的计算机系统包括用于数据库管理的所有硬件资源和基本软件。

硬件资源包括:进行数据处理和运算的中央处理机(CPU),存放基础软件和执行程序的各内部存储器(RAM 或 ROM),传递数据的总线(Bus),存放数据和各种软件的外部存储器、磁盘机、键盘等输入设备和屏幕、打印机等输出设备,以及各种接口,电源设备等,这些硬件资源都是不可缺少的。未配备各种软件的计算机硬件系统一般称为裸机。

计算机内的各种软件资源按照其功能和作用可以划分为不同的层次。最基础的软件是硬件的自检、诊断程序、基本的装入加载程序以及一些外部设备的低级驱动程序。这些都是使计算机正常运行的最基本的要求,虽然我们不能直接使用这些软件,但是任何时候计算机中都必须有这些软件存在。在第二层的是操作系统,(Operate System,简称 OS)这是一类非常重要的系统软件,是一组用来控制和管理计算机资源,便于用户使用计算机程序的组合。它要完成管理所有计算机资源,用户资源共享,软件输入输出,多道并行处理,进程调度,发现及处理错误等工作,是所有系统软件当中最基础的部分。其次是一些比较低级的应用程序,包括汇编程序,编辑程序和一些低级服务程序等,这些应用程序普通用户也很少用到,一般由程序设计人员用于一些系统软件的设计开发。最后一层是较为高级的应用程序,包括高级语言的编译系统,数据库系统,图形软件等应用程序。

计算机系统的总体结构可见图 1-2。

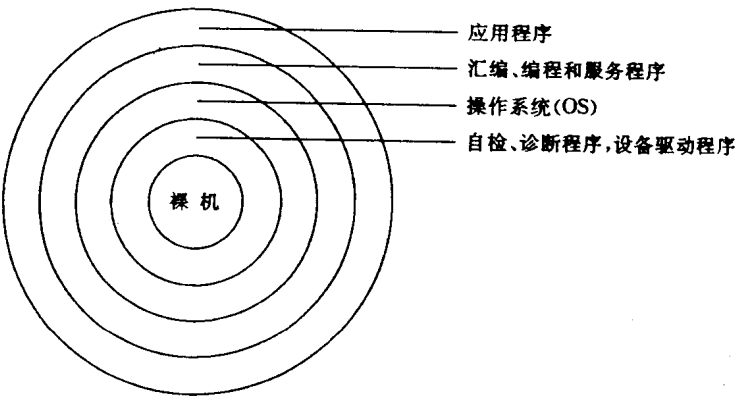


图 1-2 计算机系统结构图

二、数据库系统

数据库系统是由数据库文件、数据库管理系统和应用程序三部分组成的,这三部分的功能和关系前面已经做过介绍。三个部分相互配合,就可以完成从数据的收集整理,数据库的建立到数据的输入,维护和数据应用的整个信息管理和应用过程。这也就是本书所要介绍的全部内容。

整个数据库系统的结构和相互关系可以见图 1-3。

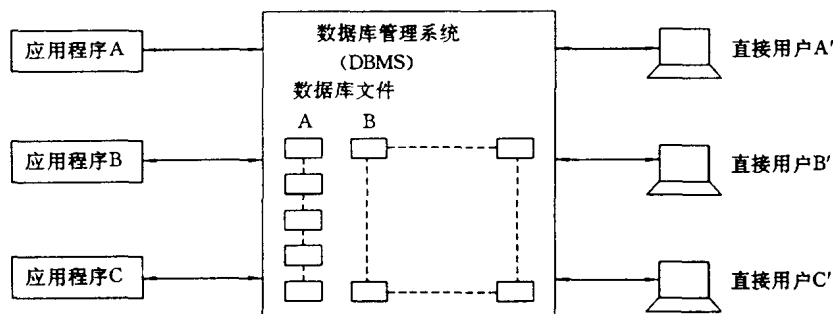


图 1-3 数据库系统的结构及相互关系

三、数据库系统的有关人员

任何先进的计算机系统和先进的软件都不能离开人的设计和使用,因而一套完整的数据库系统也必须包括设计和使用数据库的有关人员。

1. 数据库管理员

数据库管理员对于整个数据库系统是非常重要的,他们应当对于计算机的软硬件知识和实际应用具有相当深入的了解,而且对于他们所使用的数据库系统的性质、功能、系统结构、运行过程等方面非常熟悉。他们要负责对整个数据库系统进行全面的管理和维护,还要在数据库出现错误时排除故障,保证数据库的正常运行。任何一个数据库的运行都不能离开数据库管理员。

2. 程序设计开发人员

程序设计开发人员应当是专业的计算机软件设计人员。他们的工作是要根据用户的使用要求和实际情况,设计出数据库的总体逻辑和物理结构,并通过数据库提供的各种编程工具或者借助其他高级语言(例如 C, COBOL 等)设计,调试完成各类应用程序,实现检索、删除、更新、统计等数据处理,并输出处理结果。

3. 普通用户

他们是数据库和应用程序的最终使用者,也称终端用户。这类用户一般不是计算机专业人员,而是企事业单位的工程技术人员或管理人员。他们仅仅利用数据库中的一部分数据和应用程序,对于他们感兴趣的数据进行处理,得到他们所需的结果。

由于用在微机上的数据库都比较简单,内容也相对较少,所以以上 3 种数据库系统有关人员的划分也并不严格,数据库管理员的职责可以由程序设计人员和普通用户所取代,数据库的设计开发人员可能同时也是用户,普通用户也可以自行编写、设计应用程序。

四、数据库的主要特点

一个成功的数据库系统应当具有以下几个特点:

1. 最小的数据冗余度

最小数据冗余度的含义是指在数据库里存储的重复数据要尽可能少。在实际应用中,有时同一数据的多次存储是必要的,数据的冗余度只能减少而不会完全消除。但是这种冗余要尽量减少,防止过多的重复数据占据大量存储空间和扰乱数据库运行。

2. 最大的共享性

能够实现数据的共享是数据库最本质的特点。数据的共享体现在以下方面:不同的应用程序可以重叠使用同一个数据库,不同的应用也可以同时存取同一数据,即数据的并发使用;比较先进的数据库还可以提供高级语言的访问接口,用户可以用数据库本身的命令语言或高级语言编程去访问同一个数据库。

3. 数据的独立性

数据的独立性是指数据资源和用户的应用程序是相互分开的,二者之间具有相对的独立性。也就是说,数据库和应用程序之间没有直接的关联。在存储设备变换、文件组织方式发生改变和数据库的逻辑结构(不是具体的物理存储结构)发生改变时,不需要修改已有的应用程序;同样,在修改现有的应用程序和编写新的应用时,也不必修改数据结构,一个完全满足数据独立性的数据库系统,应该是以数据为中心的,而所有的应用都应当面向数据。

4. 数据的安全性和完整性

数据的安全性是指对数据存取的控制。数据资源的共享并不是说所有用户都可以任意地存取或修改任何数据。为了防止数据泄密和非法修改,数据库系统应当提供一套有效的数据安全检查功能和控制措施,以保证数据使用和修改的合法性。数据库管理员将会利用这些功能和措施,维护数据库的安全。数据的完整性是指对数据正确性、合理性和一致性的保证,只有正确、合理、一致的数据才具有实用价值。例如在日期数据中,表示月份的数据应当不大于12,表示日期的数据应当不大于31,只有这样的数据才是正确合理的;否则,如果月份数据大于12,日期的数据大于31,这些错误的数字也就失去了存在的意义。所以数据库管理系统还要提供对于数据完整性的控制功能,使数据库设计人员在设计数据库时能够对数据的完整性的约束和控制条件进行定义。此外,在多用户的网络数据库里还能够进行对数据的并发控制,即多个用户在同一时刻存取同一数据时的处理。通过对数据并行处理时处理方法的控制,可以防止同时处理数据时错误的产生。

5. 数据的可修改性、可扩充性和故障的处理

所有数据库都不可能一次就建立完成,而是逐步建立完善起来的,所以数据库系统应当易于在结构上进行修改和数据扩充。另外数据库还应该具有一套完善的设施和方法,随时监测系统的运行,并且能够在故障发生时,以最快的速度排除故障,恢复数据库的正常运行。

第4节 关系型数据库及其建立过程

一、关系型数据库

在数据库技术的发展过程中,曾经出现过层次型数据库和网型数据库,但目前比较先进的

关系型数据库已经占据了主导地位。

关系型数据库利用了代数学中集合、域关系和笛卡尔积的概念和集合运算方法,把数据组织成表格的形式,这种表就称为关系。

假设有一个 K 列的二维表,分别具有属性 $A_1, A_2, \dots, A_k$, 对应于每列的属性 A_i , 都有一个对应的域 D_i (即属性的取值范围)。整个二维表就应该在值域 $D_1, D_2, \dots, D_k$ 的范围中取值。 K 个值域上的笛卡尔乘积 $D = D_1 \times D_2 \times \dots \times D_k$ 里的任意一个子集就是一个 K 度的关系,用 $R(A_1, A_2, \dots, A_k)$ 来表示。二维表里的每一行都是 K 度关系里的一个元素,这个元素是一个 K 元组 $(d_1, d_2, \dots, d_k)$, 元素的第 i 个分量的值 d_i 应当属于值域 D_i 的范围之中。

包含笛卡尔乘积的全部元素的集合是没有意义的,只有取其中的某个子集作为关系才有意义。假设我们构造一个二列的二维表,它具有两个属性,分别是“姓名”和“成绩”,其取值范围分别是姓名 = [王平、李刚、张洪], 成绩 = [0 ~ 100], 取其笛卡尔积的一个子集,就可以构成一个如下的二维表,也就是二度的关系:

表 1.1

姓 名	成 绩
王 平	85
李 刚	77
张 洪	94

从上面的例子可以看出,关系型数据库可以存放两种数据:实体本身的数据和实体间相互联系的数据。在数据库生成时,根据数据本身的性质也就同时建立起了它们之间的联系。这种联系可以是一对一的,也可以是一对多或是多对多的。在一个学校当中,某个班级与班主任之间的关系就是一对一的;一个班级只能有一位班主任,而一个班主任也只能负责一个班级。而班级与学生之间的关系就是一对多的;一个班级里可以有很多个学生,很多个学生也可能属于同一个班级。任课教师和班级之间的关系则是多对多的;每个班级都会有很多个任课教师,而同时每个教师也要完成许多班级的教学。

二、关系型数据库的建立过程

建立一个比较完整的关系型数据库通常要经过以下几个阶段。

1. 需求分析阶段

本阶段的主要工作是详细了解即将开发的数据库的所需功能和设计要求。包括需要数据库完成的所有功能,各个功能的具体要求及实现方法,用户要求的数据输入和结果输出方法,以及所有输入输出数据的类型、范围、长度等内容。一个完整的需求分析结果也就是对将要开发的数据库的各方面要求的详细描述,这一结果最后生成一份书面的需求分析报告,供以后阶段具体开发工作时对照使用。

2. 总体设计阶段

在这一阶段应当根据需求分析的结果,设计出整个库文件部分的逻辑结构和相互的关联,应用程序中主控程序与各个模块的调用关系和各个功能模块的划分等内容,以此建立起整个数据库的总体框架结构,并可以进一步设计出每个库文件的物理结构和数据结构,需要建立的索引、排序等辅助文件,对输入数据的处理方法,以及各个应用程序的具体流程。根据这些设

计就可以完成整个数据库的具体开发。

3. 程序设计阶段

在这个阶段就是要根据应用程序的流程,利用数据库语言或其他高级语言,把程序的设计思想转化成真正的应用程序。程序设计是开发一个数据库系统的重要阶段,整个系统的成功与否在很大程度上取决于程序设计的质量,应该作到尽量简洁高效,正确无误。

4. 调试完成阶段

本阶段的主要工作是在计算机上利用数据库管理系统(DBMS)完成库文件的初建和索引、排序等文件的建立。并把所有应用程序输入到计算机中,进行试运行、改错和程序优化,使应用程序具有尽量少的错误,能够比较可靠地运行。对于规模比较大或比较重要的数据库,还要完成多项测试实验,并编写软件的文档资料。经过这些工作之后,就可以得到一个基本正确的,可以投入实际使用的数据库系统。

以上就是建立一个关系型数据库系统的大致过程。有关数据库建立的具体问题还会在以后的章节里详细介绍。

第5节 关系型数据库的基本概念

上面几节已经对数据库的基本原理与特点进行了介绍,本节要介绍一些关系型数据库里最基本的名词和概念,这些内容都是本书中要经常用到的。

一、字段

字段也称为数据项,是描述某个实体具体属性的字符集。字段是可以命名的最小的信息单位,每个字段都用一个唯一固定的字段名来标识。为了使用户能够理解数据项的含义,字段名的确定应当遵照固定的命名规则,将字段名与所表示的属性名结合起来。一个字段一旦定义之后,其字段名就不能再改变,长度也被固定下来。但是对于某个库文件里的具体数据项,其内容和长度又都是各不相同的。比如,在一个描述某中学学生情况的数据库当中,姓名、性别、学号、成绩等描述学生具体情况的数据集就构成不同的字段,姓名、性别、学号等属性名本身就可以被用作字段名,但每个学生记录里姓名、性别、学号等字段的内容又是各不相同的。

二、记录

记录是比字段更大的数据单位,一般是由多个字段组成的有序集合,具有统一的规格,它可以比较完整地描述某个对象的各种属性。例如,姓名、性别、学号、班级、年龄、成绩等内容可以反映出一个学生各方面的基本情况,而某个具体学生的姓名、性别、学号、成绩等数据字段的组合也就形成了关于这个学生的记录。由于每条记录都是由多个不同的字段构成的,所以在具体使用之前还要对所含每个数据项进行格式描述,即具体规定每个数据项的数据类型(数值型、字符型、日期型等)和数据长度,也就是对库文件结构的定义。记录中每个数据项的具体内容都要完全遵照库结构的定义来产生。在库文件里每个数据项都占据固定数量的存储空间,而所有数据项的存储空间之和也就是一条记录的总长度。

三、文件

文件是相同类型的记录的集合。一条记录可以描述某一个具体实体的属性,而文件则用

来描述具有一定关系的实体集的属性。仍以上一段中的例子来看,在一个在校学生数据库当中,一条记录记录着一个学生的具体情况,而一个班级里所有学生记录的集合,就可以构成一个文件。一个数据库可以根据其相互关系把所有数据划分成多个文件,每个文件都应当有一个唯一固定的文件名。一个数据文件可以含有很多的记录,在微型数据库,每条记录的长度都是固定不变的,并且按照输入库文件的先后顺序连续存储。

四、关键字

在实际使用的数据库中,每个库文件里都存储着大量的记录。进行数据处理时,必须通过某一个字段来确定需要处理的记录,这个能够唯一标识一条记录的字段叫做关键字。关键字将用于数据库命令作用对象的确定和库文件的查询与检索。在中学学生管理数据库里,如果所有记录中的学生姓名都各不相同,就可以用“姓名”字段作为关键字;否则,如果学生中有重名现象,则必须以各不相同的学号作为关键字来标识每条记录。

下面我们通过一个 XBASE 数据库文件的实例就可以清楚地理解这些概念:

记录号	姓名	学号	性别	年龄	语文	数学	英语
1	王红	101001	女	14	82	86	79
2	李强	101002	男	15	91	87	85
3	张丽	101003	女	15	93	76	82
4	赵刚	101004	男	14	78	82	81
5	陈雷	101005	男	15	84	85	73
6	周华	101006	女	16	76	79	82
7	吴军	101007	男	15	72	68	71
8	肖铃	101008	女	14	82	77	76
9	郑雯	101009	女	16	92	94	85
10	韩建	101010	男	16	77	74	82

我们可以看到,这个库文件里的每条记录由 7 个字段组成,其字段名分别是“姓名”、“学号”、“年龄”、“性别”、“语文”、“数学”和“英语”,每个字段里的数据类型和长度都是相同的。7 个字段的具体数据就组成了关于一个学生的一条记录,这个名为“STUDENT.DBF”的数据库文件包含了 10 条记录。“学号”是能够唯一标识一条记录的字段,是这个数据库文件的关键字。

数据库里的字段、记录和文件之间具有以下关系;

(1) 数据库文件是用 XBASE 存储在计算机里的二维表。二维表的横行就是记录,纵列就是字段。

(2) 对每个字段应当规定不同的字段名及其数据类型和最大宽度(最多占用几个字符),也就是库文件的结构。

(3) 库文件的记录按照输入顺序自动生成连续的记录号,任意两条记录不能完全相同。

(4) 字段或记录的顺序可以进行交换。

以上的基本概念和特点,在后面会经常用到,应当熟练掌握。

第 2 章 DBASE 和 FoxBASE 数据库基本概述

第 1 节 DBASE 与 XBASE 系列数据库的历史与主要产品

目前在全世界广为应用的 DBASE 系列数据库是美国安信达 (ASHTOH-TATE) 软件公司开发的非常成功的软件产品。80 年代初期, ASHTOH-TATE 公司推出了第一个比较成功的 DBASE 产品——用于 8 位微型计算机上的 DBASE-II。不久又推出了用于 16 位微型计算机的第二代产品, 也就是著名的 DBASE-III, 很快就以其功能强大, 使用简单, 操作灵活的特点风靡了全世界。根据统计, 自推出以来, DBASE 系列产品一直是最畅销的微机软件之一。而我国自 1982 年开始就对 DBASE-II 进行了引进汉化和使用推广, 经过十几年的努力, DBASE 数据库已经进入了越来越多的应用领域, 逐步成为了国内最流行的数据处理软件。

能够取得如此巨大的成绩, 完全是由 DBASE 数据库自身的优点所决定的。

首先, 它的功能非常强大而实用。DBASE 没有像其他大型数据库那样, 提供许多非常专业化的功能, 而使系统变得非常复杂且成本十分昂贵; 它对普通数据具有非常强大的通用处理功能, 能够完成各种企事业单位和学校机关的大量日常管理工作。因此它的功能对于普通用户来讲非常实用, 而其使用又非常简单, 价格也很低廉, 所以很快就被大量中小型单位和一般用户所采用, 占领了微机数据库的广大市场。

其次, 它的操作非常灵活。由于 DBASE 是微型数据库, 仅用一台普通微型计算机就可以正常运行, 所以对运行环境的要求也非常低, 任何用户都可以操作并使用数据库里的所有数据, 而不像大型的专用数据库, 需要昂贵的大中型计算机和多种外围设备等运行环境, 还要设立专门的数据库管理员和程序设计人员来管理数据库的运行。

第三, 它的使用非常简单。DBASE 为用户提供了一种非常类似于其他高级语言的专用数据处理语言, 用户可以通过使用大量的语句、函数和内存变量, 编制具有数据定义说明、数据操作和程序控制等多种功能的应用程序; 还可以在 DBASE 环境里进行工作环境、模式的设定和程序的文本编辑与修改。而这些命令的格式与英语的自然语法结构又非常相似, 程序结构也比较简单明了, 用户可以很快理解其功能与含义。同时一些命令的执行还采用了与 DOS 操作系统相同的人机对话的执行方式。所有这些都使 DBASE 具有简明易学, 特别是便于初学者学习使用的特点。因而 DBASE 数据库自推出以来, 一直受到各界用户的青睐, 并享有“大众数据库”“傻瓜软件”的美誉。

自从 80 年代初, 美国 IBM 公司推出了 IBM-PC 型微机之后, 微机开始大规模进入了人类日常的工作、生活当中, 并且一直保持着极高的发展速度。仅仅经过了十几年的使用, 目前全球的装机总量已经超过了 1 亿台, 并且以每年三、四千万台的速度持续增长。而 DBASE 正是适应了这一趋势, 它的硬件运行环境是在全世界最为普及的微型计算机, 为人们日常工作中大量繁琐的数据收集、处理工作提供了简便高效的解决方法。因而一经推出, 就受到了用户的普遍欢迎, 并很快占据了数据库市场的主导地位。也正是由于这个原因, 国外众多软件公司先后开发出了多种与 DBASE 数据库标准完全一致的兼容型或改进型产品。而我国近十余年来也先