

第 1 章 概 述

多年来, FoxPro 是我国数据库技术普及教育的软件之一, 具有广泛的应用基础和用户群。该软件不仅可以用于开发小型数据库系统, 而且已被广泛用于大型数据库的前端开发。Visual FoxPro 6.0 具有完整的数据库管理系统功能, 同时具有面向对象程序设计的各类开发工具, 因此可大大简化应用系统的开发过程, 并且简单易学、操作方便。

从应用的角度来看, 读者需要掌握数据库系统的基础理论知识, 以便指导应用实践。为此, 本章将数据库系统的基础理论知识中非常实用的内容通过提取、精炼汇集在一起进行介绍, 以便读者在开发应用系统时知其然, 更能知其所以然, 从而更好地学好用好 Visual FoxPro。

1.1 数据库系统概述

数据库技术是数据管理的技术, 它应数据管理任务的需要而产生于 20 世纪 60 年代中期, 是一门综合性技术, 涉及操作系统、数据结构、算法设计和程序设计等知识。下面简要介绍数据库技术及数据库系统的理论知识。

1.1.1 数据库系统基础知识

早期的计算机主要用于科学计算, 当计算机应用于生产管理、商业财贸、情报检索等领域时, 所面对的是数量惊人的各种类型的数据。为了有效地管理和利用这些数据, 就产生了数据库技术。

1. 数据管理的基本概念

1) 数据

数据是一种物理符号序列, 用来记录事物的情况, 用型和值来表征。不同数据类型, 记录的事物性质是不一样的。例如数值型数据(1, 2, 3……)可以用来记录事物的多少。

2) 信息

信息是经过加工的数据, 这种数据对人类社会实践、生产及经营活动能产生决策性影响。也就是说, 信息是一种数据, 是经过加工的有用数据。

所有的信息都是数据, 而只有经过提炼和抽象之后具有使用价值的信息才能成为信息。经过加工所得到的信息仍以数据的形式表现, 此时的数据是信息的载体, 是人们认识信息的一种媒体。

3) 数据处理

在计算机近六十年的发展史中, 其应用的主要方面从最初的数值计算发展到数据处理。数据处理是指对各种类型的数据进行收集、存储、分类、计算、加工、检索及传输的过程。数据处理的目的是得到信息。数据处理有时(不严格)也称为信息处理。

2. 数据管理的发展过程

数据处理的核心问题是数据管理。数据管理指的是对数据的分类、组织、编码、储存、检索和维护等。在计算机软硬件发展的基础上,在应用需求的推动下,数据管理技术得到了很大的发展,它经历了人工管理、文件系统和数据库系统 3 个阶段。

1) 人工管理阶段(20 世纪 50 年代中期以前)

其特征是数据和程序一一对应,即一组数据对应一个程序,数据面向应用,用户必须掌握数据在计算机内部的存储地点和方式。不同的应用程序之间不能共享数据。

人工管理数据有两个缺点,一是应用程序与数据之间依赖性太强,不独立;二是数据组和数据组之间可能有许多重复数据,造成数据冗余,数据结构性差。

2) 文件系统阶段(20 世纪 50 年代后期至 60 年代中期)

其特征是把数据组织在一个个独立的数据文件中,实现了“按文件名进行访问、按记录进行存取”的管理技术。在文件系统中,按一定的规则将数据组织成为一个文件,应用程序通过文件对文件中的数据进行存取和加工。至今文件系统仍是一般高级语言普遍采用的数据管理方式。文件系统对数据的管理,实际上是通过应用程序和数据之间的一种接口实现的,如图 1.1 所示。

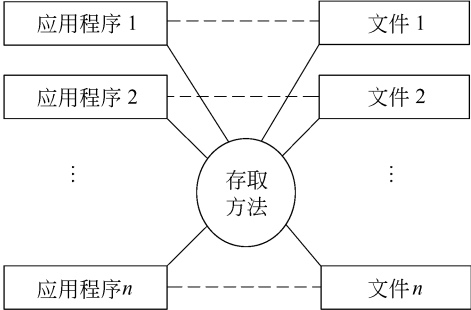


图 1.1 应用程序与数据的对应关系(文件系统)

文件系统的最大特点是解决了应用程序和数据之间的公共接口问题,使得应用程序采用统一的存取方法来操作数据。同时,应用程序和数据不再是直接的对对应关系。但是,文件系统只是简单地存放数据,它们相互之间并没有有机的联系。数据的存放依赖于应用程序的使用方法,不同的应用程序仍然很难共享同一数据文件,这就使得数据的独立性较差。另外,文件系统对数据存储没有一个相应的模型约束,数据冗余性较大。因此,文件系统难以适应大数据量信息处理的需要。

3) 数据库管理阶段(20 世纪 60 年代后期以来)

在 20 世纪 60 年代后期,计算机性能得到很大提高,人们克服了文件系统的不足,开发出一种软件系统,称之为数据库管理系统。从而将传统的数据管理技术推向一个新阶段,即数据库系统阶段。

一般而言,数据库系统由计算机软硬件资源组成。它实现了有组织地、动态地存储大量关联数据,并且方便多用户访问。它与文件系统的重要区别是数据的充分共享、交叉访问及应用程序的高度独立性。通俗地讲,数据库系统可把日常一些表格、卡片等数据有组织地集合在一起,输入到计算机中,然后通过计算机进行处理,再按一定要求输出结果。所以,数据库相对于文件系统来说,主要解决了 3 个问题。

- ① 有效地组织数据,这主要指对数据进行合理设计,以便计算机存取;
- ② 将数据方便地输入到计算机中;
- ③ 根据用户的要求将数据从计算机中抽取出来(这是人们处理数据的最终目的)。

数据库也是以文件方式存储数据的,但它是数据的一种高级组织形式。在应用程序和数据库之间有一个新的数据管理软件(database management system,DBMS),即数据库管理系统。数据库管理系统对数据的处理方式和文件系统不同,它把所有应用程序中使用的数据汇集在一起,并以记录为单位存储起来,便于应用程序查询和使用,其关系如图 1.2 所示。

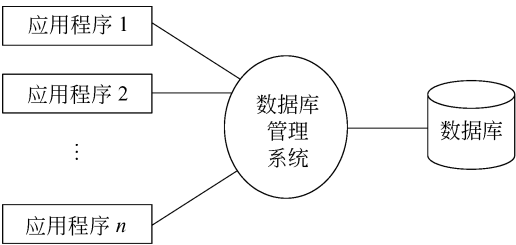


图 1.2 应用程序与数据的对应关系(数据库系统)

数据库系统和文件系统的区别是,数据库对数据的存储是按照同一结构进行的,不同的应用程序都可以直接操作这些数据(即应用程序的高度独立性)。数据库系统对数据的完整性、惟一性和安全性都提供一套有效的管理手段(即数据的充分共享性)。数据库系统还提供管理和控制数据的各种简单操作命令,使用户编写程序时容易掌握(即操作方便性)。

3. 数据库系统的组成

数据库系统实际上是一个应用系统,它是在计算机硬、软件系统支持下,由用户、数据库管理系统、存储在存储设备上的数据和数据库应用程序构成的数据处理系统。

1) 数据

这里的数据是指数据库系统中存储在存储设备上的数据,它是数据库系统操作的对象。存储在数据库中的数据具有集中性和共享性。

2) 数据库管理系统

是指负责数据库存取、维护 and 管理的软件系统,它提供对数据库中数据资源进行统一管理和控制的功能,起着用户程序和数据库数据之间相互隔离的作用。数据库管理系统是数据库系统的核心,其功能强弱是衡量数据库系统性能优劣的主要方面。数据库管理系统一般由计算机软件公司提供。

3) 应用程序

应用程序是指为适合用户操作、满足用户需求而编写的数据库应用程序。

4) 用户

用户是指使用数据库的人员。数据库系统中的用户主要有终端用户、应用程序员和管理员 3 类。

终端用户是指计算机知识不多的工程技术人员及管理人员,他们只能通过数据库系统所提供的命令语言、表格语言以及菜单等交互式对话手段使用数据库中的数据。

应用程序员是指为终端用户编写应用程序的软件人员,他们设计的应用程序主要用途是使用和维护数据库。

数据库管理员(database administrator,DBA)是指全面负责数据库系统正常运转的高级人员,他们负责对数据库系统本身的深入研究。

4. 数据库系统特点

数据库系统的出现是计算机数据处理技术的重大进步,其特点如下所述。

1) 实现数据共享

数据共享允许多个用户同时存取数据而互不影响。数据共享包括3个方面:首先,所有用户可以同时存取数据;其次,数据库不仅可以为当前的用户服务,也可以为将来的新用户服务;最后,可以使用多种语言完成与数据库的接口。

2) 实现数据独立

所谓数据独立是指应用程序不随数据存储结构的改变而变动。数据独立包括物理数据独立和逻辑数据独立两个方面。

物理数据独立是当数据的存储格式和组织方法改变时,不影响数据库的逻辑结构,从而不影响应用程序。

逻辑数据独立是当数据库逻辑结构变化时(如数据定义的修改、数据间联系的变更等),不会影响用户的应用程序,即用户应用程序无须修改。

数据独立性提高了数据处理系统的稳定性,从而提高了程序维护的效益。

3) 减少了数据冗余度

在数据库系统中,用户的逻辑数据文件和具体的物理数据文件不必一一对应,存在着“多对一”的重叠关系,有效地节省了存储资源。

4) 避免了数据不一致性

由于数据只有一个物理备份,所以数据的访问不会出现不一致的情况。

5) 加强了对数据的保护

数据库中加入安全保密机制,可以防止对数据的非法存取。由于进行集中控制,故有利于控制数据的完整性。数据库系统采取了并发访问控制,保证了数据的正确性。另外,数据库系统还采取了一系列措施,实现了对数据库破坏的恢复。

5. 数据库应用系统

数据库应用系统(database application systems,DBAS)是指开发人员利用数据库系统资源开发出来的、面向某一类实际应用的软件系统,分为两大类。

1) 管理信息系统

例如,财务管理系统、人事管理系统、教学管理系统、图书管理系统、生产管理系统等,它们是面向机构内部业务和管理的数据库应用系统。

2) 开放式信息服务系统

这是面向外部、能够提供动态信息查询功能,以满足用户的不同信息需求的数据库应用系统。例如,大型综合的科技情报系统、经济信息系统和专业的证券实时行情、商品信息等均属于这类系统。

一个数据库应用系统通常由数据库和应用程序两部分组成,它们是在数据库管理系统支持下设计和开发出来的。数据库设计和应用程序开发的有关内容将在后续章节中讲解。

1.1.2 现实世界的描述

现实世界是存在于人脑之外的客观世界,如何使用数据来解释和认识现实世界,则需要使用相应手段进行数据描述。下面首先讨论数据描述所涉及的3个领域,然后讨论概念模型和数据模型。

1. 数据描述

计算机数据管理的对象是现实生活中的客观事物,人们在实施对客观事物的管理过程中,首先要经历了解熟悉的过程,从观测中得到大量描述具体事物的数据。但是这些数据是无法送入计算机的,必须进一步整理和归类,进行数据的规范化,然后才能将规范化数据送入计算机的数据库中保存起来。这一过程经历了3个领域——现实世界、信息世界和数据世界。

1) 现实世界

现实世界是存在于人脑之外的客观世界,事物及其相互联系就处于现实世界之中。事物可用“对象”与“性质”来描述,又有“特殊事物”与“共同事物”之分。

2) 信息世界

信息世界是现实世界在人们头脑中的反映。客观事物在信息世界中称为实体,反映事物联系的是实体模型。

3) 数据世界

数据是信息世界中信息的数据化,现实世界中的事物及其相互联系在这里用数据模型描述。因此,客观事物系信息之源,是设计数据库的出发点,也是使用数据库的最终归宿。实体模型与数据模型是对客观事物及其相互联系的两种抽象描述。数据库的核心问题是数据模型,为了得到正确的模型,首先要充分了解客观事物。

2. 实体模型

现实世界中的事物在人们头脑中反映的信息世界是用文字和符号记载下来的,描述事物的术语有下列7种。

1) 实体(entity)

客观事物在信息世界中称为实体。实体可以是具体的,如一个学生、一本书;也可以是抽象的事件,如一场足球比赛。实体用型(type)和值(value)来表征,例如学生是一个实体,学生姓名是实体的型描述,而具体的学生李明、王力是实体值。

2) 属性(attribute)

实体有许多特性,这些特性在信息世界中称为属性。属性用型(type)和值(value)表征,例如学号、姓名、年龄是属性的类型,而具体的值0107039、赵东、18等则是属性值。

每个属性都有一个值域(domain),值域的类型可以是整数、实数或字符型,例如,学生的姓名、年龄都是学生这个实体的属性,姓名的类型为字符型,年龄的类型为整型。

3) 实体型

所谓实体型就是对实体的型的描述,通常使用实体名和属性名的集合。如学生实体,其型的描述为“学生(学号,姓名,年龄)”。

4) 实体值

所谓实体值是实体的具体实例,是属性值的集合,如学生赵东的实体值是 0107039、赵东、18。

5) 实体集(entity set)

性质相同的同类实体的集合称为实体集,如一个班的所有学生、一批书籍等。

6) 实体联系

建立实体模型的一个重要任务就是要找出实体之间的联系。常见的实体联系有以下 3 种(设 A 和 B 为两个实体)。

① 一对一联系(1 : 1) 如果 A 中的任一属性至多对应 B 中的惟一属性,且 B 中的任一属性至多对应 A 中的惟一属性,则称 A 与 B 是一对一联系,例如,电影院中观众与座位之间、乘车旅客与车票之间、病人与病床之间等都是是一对一联系。

② 一对多联系(1 : N) 如果 A 中至少有一属性对应 B 中一个以上的属性,且 B 中的任一属性最多对应 A 中的一个属性,则称 A 对 B 是一对多联系,例如,学校对系、班级对学生等都是是一对多联系。

③ 多对多联系(M : N) 如果 A 中至少对应 B 中一个以上属性,且 B 中的也至少有一个属性对应 A 中一个以上属性,则称 A 与 B 是多对多联系,例如,学生与课程之间、工厂与产品之间、商店与顾客之间等都是多对多联系。

上述 3 种联系是实体之间的基本联系,原则上,许多实体之间的复杂联系都可用若干组基本联系等价地表示。

7) 概念模型(即实体模型)

反映实体之间联系的模型称为实体模型。数据库设计的重要任务就是建立实体模型,它是概念数据库的具体描述。在建立的实体模型中,实体要逐一命名以示区别,并描述其间的各种联系。现以教学管理为例建立实体模型。

教学管理可由教师、学生、课程、专业等实体组成。其中,教师具有属性:教师代号、姓名、性别、职称(当然还有其他属性,为讨论简单这里省略);学生具有属性:学号、姓名、性别、专业,课程具有属性课号和课名;专业具有属性:专业名称和负责人等。

学生对课程是多对多联系,因为一个学生可以学习多门课程,而一门课程又有多个学生学习。教师对课程假设是多对多联系,即一位教师可以讲授多门课程,一门课程可由多位教师讲授。专业对学生是一对多联系,即一个专业有多个学生,而每个学生都只属于一个专业;同样地,专业对教师也都是一对多的联系(假设一个教师只能属于一个教学专业),相应的实体模型参见图 1.3。

图 1.3 是一种静态实体模型,只反映实体的当前状态,而不能反映实体状态的变化过程。

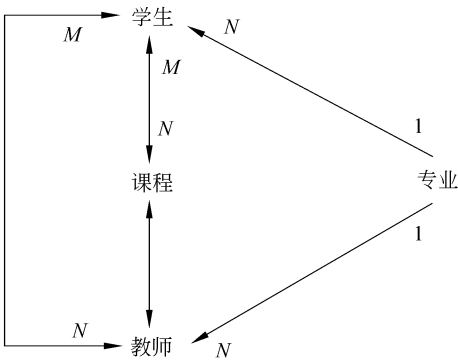


图 1.3 教学实体模型

3. 数据模型分类

数据模型是数据库系统的基石。从创建数据库技术以来,数据模型有 4 种类型,即层

次模型、网状模型、关系模型和关系对象模型。

1) 层次模型(hierarchical model)

数据的层次模型使用树状结构来表示实体的类型和实体间的联系,树的结点为记录的类型,记录类型只有简单的层次关系。

2) 网状模型(network model)

网状模型是层次模型的拓展,广义地讲,任意一个连通的基本层次联系的集合就是一个网状模型。

3) 关系模型(relation model)

1970年,美国的 E. F. Codd 提出了关系模型的概念。关系模型是数据模型中最重要的数据模型,自 20 世纪 80 年代以来,人们一直使用关系模型来描述现实世界。计算机软件厂商推出一系列关系型数据库管理系统。因为关系模型是建立在数学概念基础上的,有其坚实的数学基础,本书主要介绍关系数据模型的应用。

4) 关系对象模型

自 20 世纪 90 年代中期以来,人们看到了关系模型的一些严重缺陷,提出关系对象模型。关系对象模型一方面对数据结构方面的关系结构进行改良,另一方面人们对数据操作引入了对象操作的概念和手段,今天的数据库管理系统基本上都提供了这方面的功能。

1.1.3 关系模型

1. 关系模型中的几个术语

1) 字段(field)

信息世界中的“属性”就是数据世界中的“数据项”;从数据库的角度讲,数据项就是字段;从表格的角度讲,数据项称为列。例如学生的学号、姓名、性别和专业就是字段名。字段和属性一样,也用型和值表示。由此可见,字段、属性、数据项、列这些术语所描述的对象是相同的,只是从不同角度对对象进行描述。

2) 记录(record)

字段的有序集合称为记录。在关系模型中,记录称为元组;在表中,记录称为行;在概念模型中称为实体。换句话说,实体、记录、元组和行分别是不同角度描述同一对象的术语。记录也由型和值来描述:记录型是字段型的集合,记录值是字段值的集合。

3) 表(relation)

记录是字段的集合,表是记录的集合。记录的型和值构成了关系数据库的基本单位,即表。显然,表也分为型和值,表的型也称为关系模式,或称为表结构,由一系列字段型所组成。

4) 关键字(key word)

为了确定一条具体的记录,通常使用一种称为关键字的术语来描述。所谓关键字就是能够惟一确定记录的字段或字段的集合。有了关键字就可以很方便地使用指定的记录。

5) 关系数据库

所谓关系数据库就是由若干个表组成的集合。也就是说,关系数据库至少有一个表,

才能称为数据库。现实中,关系数据库是由若干个表有机地组合在一起,以满足某类应用系统的需要。在关系数据库系统中,关系模式是相对稳定的,而关系数据是随时间不断变化的,因为数据库中的数据在不断更新。

6) 关系组成与性质

一个关系实际上就是一个表,表是由不同的行和列组合而成的,图 1.4 是个学生关系。

8	8	2	10	
C	C	C	C	D
学号	姓名	性别	专业	出生日期
0103001	郑盈莹	女	外贸	1983-3-23
0103002	王小艳	女	外贸	1983-7-20
0101009	赵伟	男	中文	1982-4-2
0008015	和音	女	数学	1982-8-23
9909010	欧阳申强	男	计算机	1981-5-21
0109039	康红	女	计算机	1983-6-3

图 1.4 学生关系

注:假设“学号”由年级(两位)+专业(两位)+序号(3 位)组成。如“0103001”代表该学生是 01 级外贸专业的 001 号。

从图 1.4 中可以看出,学生关系是由表结构和表记录组成的一个表。在这个表中有 5 列,即 5 个字段,有 6 行,即 6 条记录。表结构部分的表示是字段名、字段类型和字段宽度,如学号字段,其名为“学号”、类型为 C(字符型)、宽度为 8。表记录部分表示了一条一条的记录值。

从这个关系中可以得出一些重要的结论:

- 表中每一列元素是类型相同的数据;
- 列不重名,列的顺序可以任意放置;
- 行的顺序也可以任意,表中任意两行不能完全相同(即没有重行);
- 表中元素是不可再分的最小数据项(描述对象属性的数据)。

这就是关系的基本性质,也是用来衡量一个关系的基本要素。在这些性质中,有一点是关系结构的关键,即表中元素是不可再分的最小数据项。最小数据项是基本单元格,不能表中套有表。这样规定的目的是将复杂的问题简单化,但是带来的后果却是使得应用复杂化。例如在学生关系中增加一个家庭地址字段,而该字段由邮政编码、地址两个基本项组成;按照关系的性质,该字段就不是最小数据项,因而就不能组成一个关系,必须再细化之后,才能成为一个关系。然而,在实际应用中,使用家庭地址会更方便。也正是由于这一点,才引入了对象数据库的概念。换句话说,对象数据库中去掉了最小数据项这一性质。

2. 建立关系结构模型

建立关系结构模型实际上是将概念模型数据化。由于实体之间存在着复杂的联系，因此所有描述实体的数据之间也存在着复杂的联系。为使模型能清晰、准确地反映客观事物，并能用于数据库设计，创建关系结构模型应采取如下步骤：

- ① 给关系结构数据模型命名，使不同模型得以区别；
- ② 给每个表命名；
- ③ 给每个字段命名，指出数据项特征，即字段的类型、长度、值域；
- ④ 找出各个表之间的联系。

教学管理中的关系数据结构模型如图 1.5 所示。其中的某一实体(如学生)，表示一



图 1.5 教学管理的关系数据结构模型

个关系表。图 1.5 中的“学生”是该实体(即关系表)的名称,“学号(文本,7)”中的“学号”是字段名,“文本”是数据类型,表示为字符型,“7”是指定该字段的长度为 7 位数,下划线表示该字段为主关键字段。

3. 关系模型优点

从图 1.5 可以看出,关系模型具有如下优点:

- 建立在严格的数据概念基础之上,每个关系是用一张表格来描述的,字段、记录描述得很清楚。更重要的是可用关系的性质来衡量关系。
- 关系规范化,即每一个数据项(即字段)是一个不可分的数据单元。
- 概念简单,数据结构简单、清晰,用户易懂易用。
- 存取路径对用户透明,数据的独立性高,安全保密性强。

总之,在关系模型中使用表格来描述,很接近实际。用户可以在表格中对数据进行操作,简单易行。

1.1.4 关系操作

关系操作是基于关系模型的基础操作,这是数据库操作中的一部分。这里将介绍关系数据库中最常用的 3 种关系操作,即投影、选择和连接。

1. 投影(projection)

投影是对一个关系表的字段进行选择,消去表的某些字段,并按要求重新安排次序。

2. 选择(selection)

选择是对一个关系表的记录进行选择,把符合某个条件的记录集选择出来,重新构建一个原表的子集。

3. 连接(join)

连接操作是将两个关系表,按照两个关系表中相同字段间的一定条件选择记录子集。

1.1.5 关系完整性

数据库中完整性是指数据的正确性和相容性。例如在成绩关系表中,“成绩”被约束为数值型的,其长度为 5,小数位为 1,即其正确的成绩范围为 0~999.9。但对于具体的对象,需要进一步的约束在一个指定的范围之内,如每门课程的成绩为 0~100(假设 100 分满分)。另外,在成绩表中出现的学号,必须是学生表的学号,否则就会造成混乱。为了保证这些关系的正确性,就要求对创建的数据结构模型提供一种约束机制。

关系模型的完整性规则是对关系的某种约束条件。关系模型中有 3 类完整性约束。即实体完整性、参照完整性和用户定义完整性,其中实体完整性和参照完整性是关系模型必须满足的完整性约束条件,称为关系的两个不变性。

1. 关系模式

关系的描述称为关系模式(relation schema),关系可以用表格来描述,也可以用数学式描述。如成绩关系,可以描述为“学号(C,7)、课号(C,6)、平时(N,5.1)、期中(N,5.1)和期末(N,5.1)”关系模式,如图 1.6 所示。