

关系数据库 及其应用

—— FoxPro2.5 及其应用

主 编

邱家武

副主编

朱少林

肖慎勇

屈振新

韩 波

科 学 出 版 社

关系数据库及其应用

——FoxPro2.5 及其应用

主 编 邱家武
副主编 朱少林 肖慎勇
屈振新 韩 波

科 学 出 版 社

2000

内 容 简 介

本书是为大、中专非计算机专业学生编写的关系数据库教材。本书以 FoxPro 2.5 为例详尽地阐述了关系数据库的基本理论、基本知识及相关应用,内容详实,层次严谨,辅以大量实例,可读性强,尤其注意到关系数据库在财经领域中的应用。

本书亦适于广大计算机爱好者自学之用。

图书在版编目(CIP)数据

关系数据库及其应用:FoxPro2.5 及其应用/邱家武主编:朱少林等编著.
—北京:科学出版社,2000
ISBN 7-03-008297-4

I. 关… II. ①邱… ②朱… III. 关系数据库-数据库管理系统, Fox-Pro2.5 IV. TP311.132%3

中国版本图书馆 CIP 数据核字(2000)第 02196 号

科学出版社 出版

北京东黄城根北街 16 号
邮政编码:100717

北京双青印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

2000 年 3 月第 一 版 开本: 787×1092 1/16
2000 年 3 月第一次印刷 印张: 14.5
印数: 1—5 100 字数: 335 000

定价: 22.00 元

(如有印装质量问题, 我社负责调换(环伟))

前 言

数据库技术产生于 20 世纪 60 年代中后期,这是一种对数据进行组织和管理的最新技术。它的出现极大地促进了计算机向各行各业的渗透。可以说,数据库技术是开发各类信息系统、办公自动化、计算机数据处理必不可少的工具。

目前,关系数据库得到了广泛的应用和发展。为了提高学生的分析能力和动手能力,使学生在毕业后能熟练地从事计算机信息处理工作,本书以当前广泛使用的 FoxPro2.5 版本为主线来阐明关系数据库的基本知识及其应用。特别注意到用 FoxPro 关系数据库管理系统去解决财经领域中的实际问题。

本书共分 11 章:第一章介绍数据库基本知识,主要是关系数据库的基本知识及数据规范化理论;第二章至第六章是 FoxPro 的基础篇,重点介绍 FoxPro 的基本知识及其应用;第七章至第十章是 FoxPro 的高级篇,介绍 SQL 查询语言的用法以及辅助编程工具;第十一章讲述在信息系统开发过程中关系数据库实例的设计及应用,并给出了主要模块的程序。

本书由邱家武任主编,朱少林、肖慎勇、屈振新、韩波任副主编。第一章由邱家武编写;第二章至第四章由屈振新编写;第五章、第六章由朱少林编写,第七章至第十章由韩波编写;第十一章由肖慎勇编写;最后由邱家武统稿。刘莉萍为本书的编辑整理做了大量的文字及图形工作。

由于我们水平所限,书中难免有许多不足之处,敬请读者批评指正。

邱家武

2000 年元旦于中南财经大学

目 录

前 言

第一章 数据库	1
第一节 概述	1
第二节 实体模型	5
第三节 数据库的数据模型	6
第四节 关系的规范化理论	10
第五节 数据库管理系统	12
习题一	14
第二章 FoxPro 概述	16
第一节 FoxPro 简介	16
第二节 FoxPro2.5b 的安装和运行环境	18
习题二	20
第三章 FoxPro 的命令结构和数据形式	20
第一节 FoxPro 命令与结构	21
第二节 变量与常量	23
第三节 表达式	25
第四节 内部函数	31
第五节 宏替换	47
习题三	49
第四章 命令文件	52
第一节 命令文件的建立与执行	52
第二节 交互式命令和屏幕输入输出设计	55
第三节 FoxPro 的程序结构	66
第四节 过程与自定义函数	84
第五节 调试和跟踪	90
第六节 数组	93
习题四	97
第五章 数据库的建立与修改	100
第一节 建立和更改数据库的结构	100
第二节 输入数据库记录	106
第三节 查看和修改数据库的记录	107
第四节 定位记录及修改记录	111

第五节 删除与恢复记录	115
习题五	118
第六章 数据库操作	121
第一节 数据汇总	121
第二节 索引与索引查找	122
第三节 数据库结构与记录的复制	126
第四节 多重数据库操作	129
习题六	135
第七章 查询语言 SQL	138
第一节 查询语言简述	138
第二节 SELECT 命令的基本用法	140
第三节 复杂查询	144
第四节 查询结果处理	146
习题七	150
第八章 菜单生成器	151
第一节 菜单生成器综述	151
第二节 创建应用系统的菜单	153
第三节 菜单生成器的使用	153
第四节 建立学生成绩管理系统菜单	161
习题八	163
第九章 屏幕生成器	164
第一节 对象及事件驱动简介	164
第二节 屏幕生成器综述	165
第三节 使用屏幕生成器创建屏幕	167
第四节 生成屏幕程序的方法	173
第五节 创建学生管理系统屏幕文件	176
习题九	179
第十章 报表生成器	180
第一节 报表生成器概述	180
第二节 使用报表生成器	183
第三节 创建报表实例	185
习题十	188
第十一章 数据库应用实例	189
第一节 管理信息系统及开发概述	189
第二节 一个实例	193
习题十一	225
参考文献	226

第一章 数据库

近 30 年来,数据管理技术上的一大突破就是数据库方法的提出。数据库方法的研究,始于 20 世纪 60 年代中期。1969 年美国的 CODASYL(Conference on Data System Language)委员会、美国数据系统委员会及 IBM 公司,先后提出了 DBTG(Data Base Task Group,数据库任务组)报告和 IMS 数据库系统。接着,在 1970 年美国 IBM 公司的研究员 E. F. CODD 发表了题为“大型共享系统的关系数据库的关系模型”的论文,首次提出了关系模型数据库的报告。几十年来,国际上相继开展了对数据库方法的深入研究,并将其应用于各类计算机信息管理系统中。如今,一个软件系统是否使用现代的数据库方法,已经成为衡量这个软件系统在数据管理上是否先进的一个重要标志。从某种意义上说,数据库方法的产生,已经引起了计算机信息处理领域中的一场革命,同时极大地推动了社会管理科学的进步。

本章简单介绍数据管理的三个发展阶段:实体模型、数据库的数据模型、关系的规范化理论及数据库管理系统等基本知识,为学好本书的重点 FoxPro 关系数据库及其应用打下基础。

第一节 概 述

一、数据、信息、数据库、数据库系统、数据库管理系统

1. 数据(data)

说起数据,人们首先想到的是数字。其实数字只是最简单的一种数据。数据的种类很多,在日常生活中数据无处不在:文字、图形、图像、声音、学生的档案记录、货物的运输情况等,这些都是数据。

为了认识世界、交流信息,人们需要描述事物。数据实际上是描述事物的符号记录。在日常生活中人们直接用自然语言(如汉语)描述事物。在计算机中,为了存储和处理这些事物,就要抽出对这些事物感兴趣的特征组成一个记录来描述。例如,在学生档案中,如果人们最感兴趣的是学生的姓名、性别、出生时间、籍贯、所在系别、入学时间(大学),那么可以这样描述:李明、男、1980、江苏、信息系,1998。

2. 信息

信息是经过处理、加工提炼而用于决策制定或其应用活动的的数据。该定义隐含了多种意义。其一是信息接近于知识、智能,与“原始的”、“粗糙的”数据相反,它是“精炼的”、“平滑的”,更便于应用者使用。其二是信息的相对性,即它是相对于决策制定或其他应用活动,相对于决策制定的人及其背景与经历等的。因为,在某种情况下是很重要的信息而在另一种情况下可能根本无用;对于一个决策者或应用者在特定时候非常有价值的信息,

而对于另一决策者或者甚至同一决策者在另外时候或另外的场合很可能没什么用处。其三是信息与决策制定紧密地联系在一起,收集信息就是为了决策制定,而决策制定者仅在掌握了有效信息时才能制定正确的决策。

在实践中,数据和信息常常又是很难区分的,因为在制定专门决策或解决特定问题时,数据即变成信息。数据库系统可使组织在目标范围内作出各种决策和解决各种应用问题所需的信息能通过已存储在数据库的数据获得。因此,数据库必须包含和存储尽可能多的、需要的(也仅仅是需要的)数据。

3. 数据库(database,简称 DB)

收集并抽取出一个应用活动所需要的大量数据之后,应将其保存起来以供进一步加工处理。保存方法有很多种:人工保存、存放在文件里、存放在数据库里,其中数据库是存放数据的最佳场所。

所谓数据库就是长期储存在计算机内的、有组织的、可共享的数据集合。数据库中的数据按一定的数据模型组织、描述和储存,具有较小的冗余度,较高的数据独立性和易扩展性,并可为多个用户共享。

4. 数据库管理系统(database management system,简称 DBMS)

如何科学地组织数据并将其存储在数据库中,又如何高效地处理这些数据呢?完成这个任务的是一个软件系统——数据库管理系统。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件。

数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制。数据库管理系统使用户能方便地定义数据和操纵数据,并能够保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

5. 数据库系统(database system,简称 DBS)

数据库系统是指在计算机系统中引入数据库后的系统构成,一般由数据库、数据库管理系统(及其开发工具)、应用系统、数据库管理员和用户构成。应当指出的是:数据库的建立、使用和维护等工作只靠一个 DBMS 远远不够,还要有专门的人员来完成,这些人被称为数据库管理员。

二、数据管理的三个发展阶段

1. 人工管理阶段

在 20 世纪 50 年代中期以前,计算机主要用于科学计算。当时的硬件状况是:外存只有纸带、卡片、磁带,没有磁盘等直接存取的存储设备;软件状况是:没有操作系统,没有管理数据的软件;数据只能靠人工管理,其处理方式是批处理。

人工管理数据具有如下特点:

(1) 数据不保存。由于当时计算机主要用于科学计算,一般不需要将数据长期保存,只是在计算某一课题时将数据输入,用完就撤走。不仅对用户数据如此处置,对系统软件

有时也是这样。

(2) 数据需要由应用程序自己管理,没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构,而且要设计物理结构,包括存储结构、存取方法、输入方式等。因此程序员负担很重。

(3) 数据不共享。数据是面向应用的,一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时,由于必须各自定义,无法互相利用、互相参照,因此程序与程序之间有大量的冗余数据。

(4) 数据不具有独立性。数据的逻辑结构或物理结构发生变化后,必须对应用程序做相应的修改,这就进一步加重了程序员的负担。

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期,计算机的应用范围逐渐扩大,计算机不仅用于科学计算,而且还大量用于数据处理。这时硬件方面有了磁盘、磁鼓等直接存取设备;软件方面,操作系统中已经有了专门的数据管理软件,一般称为文件系统;处理方式上不仅有文件批处理,而且能够联机实时处理。

用文件系统管理数据具有如下特点:

(1) 数据可以长期保存。由于计算机大量用于数据处理,数据需要长期保留在外存上,反复进行查询、修改、插入和删除等操作。

(2) 由专门的软件即文件系统进行管理,程序和数据之间由软件提供的存取方法进行转换,使应用程序与数据之间有了一定的独立性,程序员可以不必过多地考虑物理细节,将精力集中于算法。而且数据在存储上的改变不一定反映在程序上,大大节省了维护程序的工作量。

(3) 数据共享性差。在文件系统中,一个文件基本上对应于一个应用程序,即文件仍然是面向应用的。当不同的应用程序具有部分相同的数据时,也必须建立各自的文件,而不能共享相同的数据,因此数据的冗余度大,浪费存储空间。同时由于相同数据的重复存储、各自管理,给数据的修改和维护带来了困难,容易造成数据的不一致性。

(4) 数据独立性低。文件系统文件是为某一特定应用服务的,文件的逻辑结构对该应用程序来说是优化的,因此,要想对现有的数据再增加一些新的应用会很困难,系统不容易扩充。一旦数据的逻辑结构改变,必须修改应用程序,修改文件结构的定义。而应用程序的改变也将引起文件的数据结构的改变。因此数据与程序之间仍缺乏独立性。

3. 数据库系统阶段

20 世纪 60 年代中期以来,计算机用于管理的规模更为庞大,应用越来越广泛,数据量急剧增长,同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。这时硬件已有大容量磁盘,硬件价格下降;而软件价格上升,为编制和维护系统软件及应用程序所需的成本相对增加;在处理方式上,联机实时处理要求更多,并开始提出和考虑分布处理。在这种背景下,以文件系统作为数据管理手段已经不能满足应用的需求。为解决多用户、多应用共享数据的需要,使数据为尽可能多的应用服务,数据库技术应运而生,出现了统一管理数据的专门软件系统——数据库管理系统。

数据库管理系统有以下特点:

(1) 数据结构化

数据结构化是数据库与文件的根本区别。

在文件系统中,相互独立的文件的记录内部是有结构的。传统文件的最简单形式是等长同格式的记录集合,但每个记录的长度必须等于信息量最多的记录的长度,因而会浪费大量的存储空间。所以最好是采用变长记录或主记录与详细记录相结合的形式建立文件。文件系统尽管其记录内部已有了某些结构,但记录之间没有联系。数据库系统实现整体数据的结构化,是数据库的主要特征之一,也是数据库系统与文件系统的本质区别。

(2) 数据的共享性好,冗余度低

数据的共享程度直接关系到数据的冗余度。数据库系统从整体角度描述和组织数据,数据不再面向某个应用而是面向整个系统。这样既可以大大减少数据冗余,节约存储空间,又能够避免数据之间的不相容性与不一致性。所谓数据的不一致性是指同一数据不同拷贝的值不一样。采用人工管理或文件系统管理时,由于数据被重复存储,当修改不同的拷贝时就易造成数据的不一致。

(3) 数据独立性高

数据库系统提供了两方面的映象功能,从而使数据既具有物理独立性,又有逻辑独立性。数据库系统的一个映象功能是数据的总体逻辑结构与某类应用所涉及的局部逻辑结构之间的映象或转换功能。这一映象功能保证了当数据的总体逻辑结构改变时,通过对映象的相应改变可以保持数据的局部逻辑结构不变,由于应用程序是依据数据的局部逻辑结构编写的,所以应用程序不必修改。这就是数据与程序的逻辑独立性,简称数据的逻辑独立性。

数据库系统的另一个映象功能是数据的存储结构与逻辑结构之间的映象或转换功能。这一映象功能保证了当数据的存储结构(或物理结构)改变时,通过对映象的相应改变可以保持数据的逻辑结构不变,因而应用程序也不必改变。这就是数据与程序的物理独立性,简称数据的物理独立性。

由于数据与程序之间的独立性,便可以把数据的定义和描述从应用程序中分离出来。另外,由于数据的存取由 DBMS 管理,用户不必考虑存取路径等细节,从而简化了应用程序的编制,大大减少了应用程序的维护和修改。

(4) 数据由 DBMS 统一管理和控制

由于对数据实行了统一管理,而且所管理的是结构化的数据,因此使用数据的方式很灵活。可以取整体数据的各个合理子集用于不同的应用系统,而且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,便可以满足新的要求。因此使数据库系统弹性大,易于扩充。

除了管理功能外,为了适应数据共享的环境,DBMS 还必须提供以下几方面的数据控制功能:

(1) 数据的安全性(security)

数据的安全性是指保护数据,防止不合法使用数据造成数据的泄密和破坏,而每个用户只能按规定对某些数据以某种方式进行访问和处理。

(2) 数据的完整性(integrity)

数据的完整性指数据的正确性、有效性和相容性。即将数据控制在有效的范围内,或要求数据之间满足一定的关系。

(3) 并发控制(concurrency)

当多个用户的并发进程同时存取、修改数据库时,可能会发生相互干扰而得到错误的结果,并使得数据库的完整性遭到破坏,因此必须对多用户的并发操作加以控制和协调。

(4) 数据库恢复(recovery)

计算机系统的硬件故障、软件故障、操作员的失误以及故意的破坏也会影响数据库中数据的正确性,甚至造成数据库部分或全部数据的丢失。DBMS 必须具有将数据库从错误状态恢复到某一已知的正确状态(也称为完整状态或一致状态)的功能,这就是数据库的恢复功能。

综上所述,数据库是长期存储在计算机内有组织的、大量的、共享的数据集合。它可以供各种用户共享,具有最小冗余度和较高的数据独立性。DBMS 在数据库建立、运用和维护时对数据库进行统一控制,以保证数据的完整性、安全性,并在多用户同时使用数据库时进行并发控制,在发生故障后能对系统进行恢复。

数据库系统的出现使信息系统的研制从以加工数据的程序为中心转向围绕共享的数据库来进行。这样既便于数据的集中管理,又有利于应用程序的研制和维护,提高了数据的利用率和相容性,提高了决策的可靠性。

第二节 实体模型

为一个企业或单位设计数据库,实质上是将企业中的有关数据组织成一个计算机系统所使用的 DBMS 所采用的数据模型,这一过程通常要分两步进行:

第一步,通常采用实体—联系方法(Entity Relationship Approach)简称 E-R 方法,进行数据库的概念设计,以构造实体模型。实体模型是反映现实世界中的客观事物及其联系的抽象信息表示,它反映了用户的现实环境,描述了从用户角度看到的数据库。实体模型是设计数据库的先导,但是它不是某一个 DBMS 支持的数据模型,也不依赖于具体的计算机系统,与数据库将怎么实现也无关。

第二步,将实体模型转换为某一个 DBMS 所支持的数据模型。

建立实体模型时,会用到实体、实体型、实体集、属性、属性值、码以及联系等概念。实体是指可相互区别的客观事物,可以是人、事、物,也可以是抽象的概念或联系。实体由若干属性的属性值组成,属性是实体的某一方面的特性。例如:学号、姓名、年龄、性别等是学生的几个属性。某一学生的学号是“9406”,姓名为“张三”,年龄为“20”,性别为“男”,分别是上面几个属性所取的属性值。属性的取值范围称为该属性的域。用实体名及其属性名集合来抽象和刻画同类实体,称为实体型。例如:学生(学号、姓名、性别、年龄)就是一个实体型。

由若干个属性的属性值所组成的集合表征了一个实体,在上例中,“9406”、“张三”、“20”、“男”,这些属性值的组合就表征 9406 号学生这样一个实体。具有一组公共属性的所有实体的集合称为一个实体集。例如“学生”是一个实体集,它具有学号、姓名、年龄、性别等公共属性。能够唯一标识实体的属性或属性组称为码,例如学号就是学生这个实体

集的码。

不同类型的实体型之间存在着联系,这种联系可以分为以下三类:

(1) 一对一的联系(1:1)

如果实体集 A 中的任意一个实体至多对应于实体集 B 中的一个实体;反之,实体集 B 中的任意一个实体至多对应于 A 中的一个实体,则称实体集 A 与实体集 B 具有一对一的联系。例如:乘车旅客与车票之间、病人与病床之间都是一对一的联系。

(2) 一对多的联系(1:n)

如果实体集 A 中至少有一个实体对应于实体集 B 中一个以上的实体;反之,实体集 B 中的任意一个实体至多对应于 A 中的一个实体,则称实体集 A 对实体集 B 是一对多的联系。例如:省对县、学校对学生都是一对多的联系。

(3) 多对多的联系(m:n)

如果实体集 A 中至少有一个实体对应于实体集 B 中一个以上实体;反之,实体集 B 中也至少有一个实体对应于 A 中的一个以上实体,则称实体集 A 与实体集 B 是多对多的联系。例如:学生与课程、机器产品与标准件之间都是多对多的联系。

在实际情况中,还存在“多对一”的联系,但从本质上说,“多对一”不过是“一对多”的逆转。现实世界中任何事物之间的联系都可用若干组上述联系等价地表示。对于两个以上实体集之间的联系,可用类似于两个实体集之间的表述方法进行。

实体模型用 E-R 图来描述,实体型、联系、属性是 E-R 图的三个基本成份。在 E-R 图中,实体型用矩形框表示,在框内写上实体名;以菱形表示实体型间的联系,在框内写上联系名,并用无向边将菱形分别与有关的实体相连接,同时在连线旁写上联系的类型(1:1,1:n 或 m:n);属性用椭圆形表示,并用无向边将属性与实体连接起来。

【例 1】在考察了某培训班的教学情况后,建立教学情况的实体模型。

教学情况可由教员、学生、课程等方面的情况组成。学生与课程之间有一种联系——学习,它是多对多联系,因为一个学生可以学习多门课程,而一门课又有多个学生学习。教员与课程之间也有一种联系——任课,它是一对多的联系,一个教员可讲授多门课程,但一门课至多只有一个教员任教。用 E-R 图描述教学情况的实体模型如图 1-1 所示。

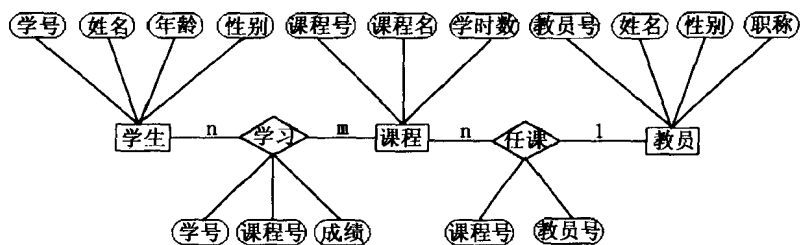


图 1-1 教学实体模型

第三节 数据库的数据模型

数据模型是对客观事物及其联系的数据描述,即实体模型的数据化。

在数据世界中,与实体型、实体、实体集、属性、码等概念相应的术语是记录型、记录、

文件、数据项、主码。

实体内部的联系反映在数据上是记录内数据项之间的联系,实体间的联系反映在数据上是记录间的联系,数据库的数据模型是指记录间联系的描述。

数据模型所有的术语都用在模型一级,整个模型就象一个框架,给它填上具体的数据就得到数据模型的一个实例。

通常,数据库有三种不同的数据模型:层次模型、网状模型、关系模型。根据数据模型的不同,数据库可分为层次数据库、网状数据库、关系数据库。

一、层次模型

数据的层次模型是以记录类型为结点的有向树。这样的树是由结点和连线组成的。其中结点表示记录类型,连线表示记录类型间的联系,这种结构就象一棵倒放的树。它的两个限制是:

- (1) 有且仅有一个结点无双亲,称为根结点;
- (2) 除根结点外,其他结点有且仅有一个双亲。

在这个结构中,父结点与子结点的联系是一对多的联系,即一个父记录对应于多个子记录,而一个子记录只对应一个父记录。

【例2】反映一个学校“系、教研室、课程、教员”情况的数据模型可以设计成层次模型。

假定每个系设置若干教研室,开设若干门课程;每个教研室又有若干名教员。它的实体模型如图1-2所示。

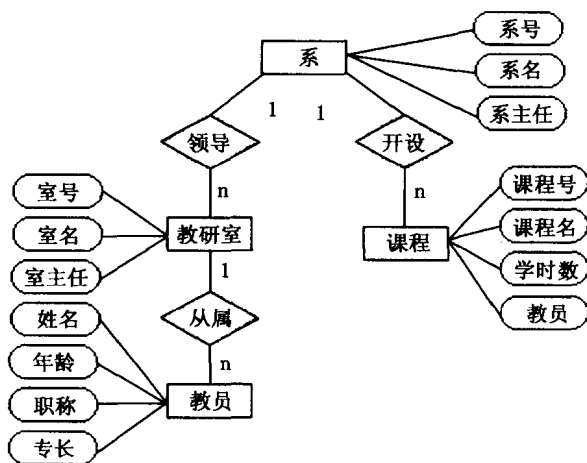


图 1-2 系教课实体模型

根据图1-2的实体模型,就可以构造出如图1-3的层次数据模型。

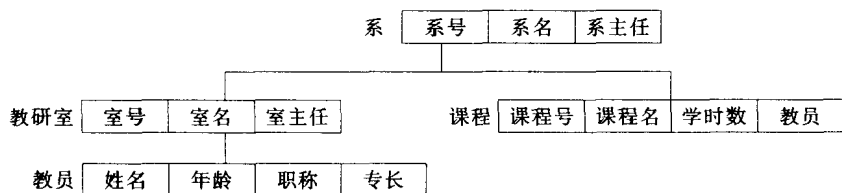


图 1-3 系教课层次模型

在层次模型中,必须按照从根开始的某条路径提出询问,否则就不能直接回答。因此设计层次模型时要细心考虑路径问题,因为路径一经规定就不能更改。

二、网状模型

数据的网状模型是以记录类型为结点的网状结构。网状模型去掉了层次模型的两个限制,允许:

- (1) 可以有一个以上的结点无双亲;
- (2) 至少有一个结点有多于一个的双亲;
- (3) 两个结点之间可以有两种或多种联系。

在网状模型中,用“系”表示记录间的联系,所谓系就是一棵二级树,它的根称为主结点,它的叶称为从结点。系有型与值之分,系型表示记录型之间的联系;系值表示记录值之间的联系。对每一个系类型都必须命名,以相互区别。实际上,任何一个网状模型均可以分解为一系列的“系”。

对应于图 1-1 所示的教学实体模型,可以构造出教学网状数据模型,如图 1-4 所示。

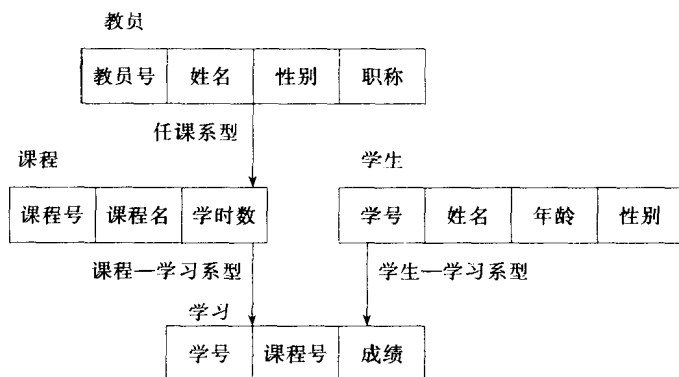


图 1-4 教学网状模型

三、关系模型

关系模型是目前最重要的一种数据模型,数据库领域当前的研究工作都是以关系方法为基础。数据的关系模型是由若干“关系框架”组成的集合,它的实例由若干“关系”组成,而关系又由关系框架和若干“元组”构成。关系框架描述了关系中的属性组成,它相当于记录型,元组相当于记录值,关系相当于文件。

关系模型的最大特点就是描述的一致性,即实体及实体间的联系均用关系来描述。现以图 1-1 的教学实体模型来构造其关系数据模型与实例,如图 1-5。从例中看到,学生与课程之间的联系和教员与课程之间的联系分别用学习关系、任课关系来描述。

学生关系框架

学号	姓名	年龄	性别
----	----	----	----

教员关系框架

教员号	姓名	年龄	性别
-----	----	----	----

课程关系框架

课程号	课程名	学时数
-----	-----	-----

学习关系框架

学号	课程号	成绩
----	-----	----

任课关系框架

课程号	教学号
-----	-----

学习关系

学号	课程号	成绩
9401	K11	91
9401	K12	86
9401	K13	84
9402	K12	71
9403	K11	68
9403	K13	72
9404	K12	85
9404	K13	80

教员关系

教员号	姓名	性别	职称
T01	赵 强	男	讲 师
T02	钱建平	女	讲 师
T03	孙 勇	男	助 教
T04	李志伟	男	副教授

学生关系

学号	姓名	年龄	性别
9401	张山	20	男
9402	李斯	19	女
9403	王武	21	男
9404	王武	20	女

任课关系

课程号	教员号
K11	T04
K12	T01
K13	T04

课程关系

课程号	课程名	学时数
K11	数据结构	70
K12	外 语	74
K13	数据库	60

图 1-5 数据的关系模型与实例

与层次数据模型和网状数据模型相比较,关系数据模型是发展较晚的一种数据模型,但由于关系数据库具有完备的设计理论,且操作语言内容丰富,用户使用方便,最有发展前途,现在大多数管理信息系统都是采用关系数据库。目前我国广泛使用的 FoxBase、FoxPro 等都是关系数据库。

在关系模型中,最主要的关系操作是:筛选、投影和连接。

1. 筛选

筛选是指对关系元组的选择,即根据给定的条件,在指定的关系中选取若干元组。

【例 3】在图 1-5 的教员关系中,要找出性别是男的教员,利用筛选操作可得到新关系,如图 1-6 所示。

男教员			
教员号	姓名	性别	职称
T01	赵 强	男	讲 师
T03	孙 勇	男	助 教
T04	李志伟	男	副教授

图 1-6 筛选操作

姓名	职称
赵 强	讲 师
孙 勇	助 教
李志伟	副教授

图 1-7 投影操作

2. 投影

投影是指对关系属性的选择,即根据给定的条件,在指定的关系中选取若干属性。

【例4】图1-6的男教员关系在姓名、职称属性上投影,得到教师职称关系,如图1-7所示。

3. 连接

连接是指将两个关系对给定的条件按某种规则连接成一个新关系。

【例5】图1-8有关系R和S,按关系R中的属性 A_3 小于关系S中的属性 A_5 为条件连接,就可以得到新关系RS。

R 关系			S 关系		RS 关系				
A_1	A_2	A_3	A_4	A_5	A_1	A_2	A_3	A_4	A_5
a	b	10	a	12	a	b	10	a	12
b	c	28	b	7	a	b	10	c	17
e	d	15	c	17	a	b	10	d	19
f	d	18	d	19	e	d	15	c	17
f	a	20			e	d	15	d	19
					f	d	18	d	19

图1-8 连接操作

第四节 关系的规范化理论

对于给出的一组数据,如何构造一个适合于它们的数据模型,或者说如何决定这些数据的逻辑结构,这就是数据库的逻辑设计问题。由于关系模型有严格的数学理论,并且可以向别的数据模型转化,我们就以关系模型为背景来讨论这个问题,形成数据库逻辑设计的有力工具——关系的规范化理论。

关系数据库中的关系是要满足一定要求的,按照满足要求的不同程度分为第一范式(简称1NF),在第一范式中进一步满足一些要求的为2NF,依此类推,有3NF、4NF、5NF。本书我们只讨论前三种范式。

在一个关系中,能够唯一标识记录的一个或多个属性称为候选码。若候选码多于一个,则选定其中一个作为主码。包含在任何一个候选码中的属性为主属性,否则称为非主属性。

一、第一范式

对关系有一个最起码的要求,它的每一个属性必须是不可分的数据项。满足了这个条件的关系就称为规范关系,或称为第一范式,简记为1NF。

【例6】图1-9所示的表格中,属性课程和成绩有重复组,所以它是非规范关系。

对于非规范关系,如果将重复组所在行其他属性的值也予以重复,就可以得到规范关系,或说是1NF,如图1-10所示。

学号	姓名	课程	成绩	系名	系主任
941001	张山	数学	86	计算机	赵刚
		外语	78		
		计算机	87		
90205	李斯	数学	90	数学	钱志伟
		外语	85		
94208	王武	数学	71	数学	钱志伟
94506	田平	汉语	81	外语	孙丽
		外语	89		

图 1-9 非规范关系

学号	姓名	课程	成绩	系名	系主任
941001	张山	数学	86	计算机	赵刚
941001	张山	外语	78	计算机	赵刚
941001	张山	计算机	87	计算机	赵刚
90205	李斯	数学	90	数学	钱志伟
90205	李斯	外语	85	数学	钱志伟
94208	王武	数学	71	数学	钱志伟
94506	田平	汉语	81	外语	孙丽
94506	田平	外语	89	外语	孙丽

图 1-10 1NF 关系

这个学习关系存在几个问题:

(1) 插入异常。如果一个系刚成立,尚无学生,或者学生尚未安排课程。那么,就无法把这个系及其系主任的信息存入数据库。

(2) 删除异常。如果某个系的学生全部毕业了,我们在删除学生选修课程的信息的同时,把这个系及其系主任的信息也丢掉了。

(3) 冗余度大。每一个系主任的姓名要与该系每一个学生的每门课程的成绩出现的次数一样多。这样一方面浪费空间,另一方面要付出很大的代价来维护数据的完整性。

由于存在上述问题,显然该关系不是一个好的数据模式,问题的原因是关系规范程度低,存在多余的数据相关性。解决的办法是将其分解成多个关系模式,使之达到某种程度的“分离”,达到高一级范式的要求。

二、第二范式

如果关系 R 是满足第一范式的,且它的每一个非主属性都完全函数依赖于码,则关系 R 是第二范式,记为 2NF。

对码的函数依赖,是指对码的每一个值,其他非主属性只有唯一值与之相对应。在码是由多个属性组成时,如果非主属性函数依赖于作为码的属性组就称为完全函数依赖;否则,如果非主属性只函数依赖于码中的部分属性就不是完全函数依赖,或称为部分函数依