

VL20

数据库应用

主审 徐升华

编著 万常选 凌传繁 曾雅琳

中国商业出版社

内 容 提 要

本书由三个部分组成,第一部分是数据库系统的有关内容,介绍了数据库技术的基本概念、基本原理及其新发展等。第二部分是本书的主体,它系统地介绍了 FoxPro 2.5/2.6 for Windows 的有关内容,作者在内容的组织和剪裁上进行了大胆的尝试,力求实现:突出重点、启发思维、循序渐进、便于自学。第三部分对大型数据库 Sybase 的基本功能和使用方法,以及面向对象的可视化的数据库前台开发工具 PowerBuilder 的基本功能和使用方法进行了介绍,使读者能领略一下客户机/服务器(Client/Server)模式下的管理信息系统的开发方法,为读者进一步的学习和工作奠定一定的基础。

本书取材和编排勇于创新,力求做到理论联系实际,突出重点,讲求实用,概念清晰,叙述简练,通俗易懂,便于自学。本书内容充实,不仅有字典式的命令格式,更重要的是有丰富的从实际应用中提炼出来的例题,还有适量的习题,所有例题和习题都在计算机上进行了实现,并且配有源程序的软盘。

本书可作为大专院校各专业以及计算机培训班的教材或教学参考书,也适合于初学者自学。

图书在版编目(CIP)数据

数据库应用/万常选等编著。—北京:中国商业出版社,2000.12

ISBN 7-5044-4057-4

I. 数… II. 万… III. 数据库系统—基本知识

IV. TP311.131

中国版本图书馆 CIP 数据核字(2000)第 82081 号

主 审:徐升华

编 著:万常选等

责任编辑:刘毕林

*

中国商业出版社出版发行

(北京市宣武区广安门内报国寺1号 邮政编码:100053)

江西教育印刷厂印刷

*

开本:787×1092 1/16 印张:21.5 字数:554千字

2000年12月第1版 2000年12月第1次印刷

ISBN 7-5044-4057-4/TP·158

定价:38.50元

第一章 数据库系统概述

数据库是 20 世纪 60 年代末 70 年代初发展起来的一门新技术,它的出现使数据处理进入了一个崭新的时代。数据库能把大量的数据按照一定的结构存储起来,在数据库管理系统的集中管理下,实现数据共享,它具有数据结构化、最低冗余度、较高的程序与数据独立性、数据共享、易于扩充、易于编制应用程序等优点,被广泛地应用于各种管理信息系统。在当今信息化的社会里,数据库已成为管理和利用信息资源不可缺少的工具。

计算机应用从科学计算进入数据处理是一个划时代的转折,使计算机从少数科学家手中的珍品成为广大科技人员和管理人员工作中的得力助手和有力工具。

数据处理是指对各种形式的数据进行收集、储存、加工和传播的一系列活动的总和。其目的是从大量的、原始的数据中抽取、推导出对人们有价值的信息,以作为行动和决策的依据;是为了借助计算机有效地保存和管理复杂的大量的数据,以便人们能方便而充分地利用这些宝贵的信息资源。

数据库技术所研究的问题就是如何科学地组织和储存数据,如何高效地获取和处理数据。数据库技术是数据管理的最新技术。数据库系统是当代计算机系统的重要组成部分。

1.1 数据库技术的进展

1.1.1 数据管理的进展

从 20 世纪 50 年代初开始,计算机已经开始在数据处理领域应用。数据处理的中心问题是数据管理,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。40 多年来,数据管理技术随着计算机软、硬件的发展在不断的的发展,它大致经历了如下三个阶段:

- (1) 人工管理阶段(50 年代中期以前)
- (2) 文件系统阶段(50 年代后期到 60 年代中期)
- (3) 数据库系统阶段(60 年代后期开始)

一、人工管理阶段

这一阶段,计算机除硬件设备外没有任何软件可用。在应用程序中除了要规定数据的逻辑结构外,还要考虑数据在计算机中如何存储和组织,并为数据分配空间、决定存取方法。应用程序完全依赖于数据。数据结构的改变,存储与取存方法的变化,都会使应用程序跟着改变。这一阶段,应用程序与数据是一一对应的,如图 1-1 所示。如果几个应用程序要用到同一数据,这些数据也要对应每个程序重复存储,数据的冗余度很大。这一阶段,数据的管理还是手工性质的、分散的,处理效率很低。

二、文件系统阶段

50 年代末,计算机在数据处理领域的应用日渐增多。随着计算机硬件和软件的发展,硬件方面,出现了磁盘、磁鼓等直接存取的外存储设备;软件方面,操作系统中已经有了专

门的管理数据软件——文件管理系统。文件的应用,是计算机数据处理的重大进展。数据文件可以按名引用,应用程序通过文件管理系统与数据文件发生关系,如图 1-2 所示。

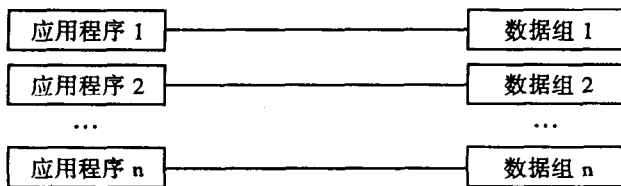


图 1-1 手工管理阶段数据与程序之间的关系

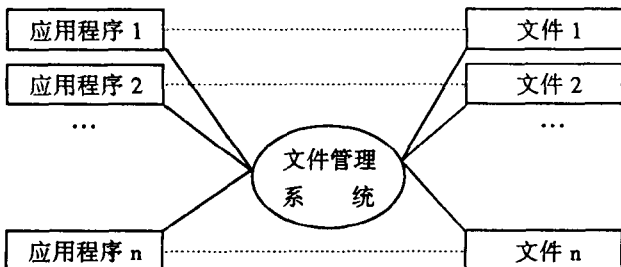


图 1-2 文件系统阶段数据与程序之间的关系

有了文件管理系统,数据的物理结构与逻辑结构之间的转换由数据管理软件完成,使程序与数据之间有了一定的独立性,程序员在应用程序中可以不必过多地考虑数据的物理存储细节,简化了程序员的数据管理工作。同时,一个应用程序可以和几个数据文件发生联系,增加了数据处理的灵活性。但是,数据仍然是分散的,是面向应用的,用文件系统处理数据仍然存在着以下几个问题:

数据冗余度大。由于数据是面向应用程序的,一个数据文件只为某一用户的特定应用程序服务,其它应用程序所需要的相同数据只能重复存储,不能共享数据,造成了大量的数据冗余,浪费存储空间。并且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。

应用程序与数据结构互相依赖。数据文件是按具体应用要求建立的,而应用程序的编制直接依赖于数据的逻辑结构。一旦由于数据的逻辑结构改变,相应的应用程序也得进行修改,增加了程序的编制和维护的工作量。而应用程序的改变,如应用程序所使用的高级语言的变化等,也将影响文件的数据结构的改变。数据与程序之间缺乏独立性。这种应用程序与数据结构的密切相关的情况称为“数据依赖”。

另外,文件系统缺乏对数据进行统一的控制方法,数据的完整性、安全性等无法得到保证,完全要应用程序自己管理,增加了程序的复杂性。

三、数据库系统阶段

文件系统处理数据存在着诸多不足,而计算机用于管理的规模却越来越大,应用越来越广泛,数据量急剧增长,而且数据的共享要求越来越强。这种共享的含义是多种应用、多种语言互相覆盖地共享数据集合,如图 1-3 所示。

这一时期,磁盘技术取得了重要进展,大容量和快速存取的磁盘开始进入市场,给数据库系统的研究提供了良好的物质基础。

为了解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,就出现了数据库管理技术。数据库技术解决了数据冗余和数据依赖问题,实现了数据共享,数据库中

的数据以一定的方式组织起来,由一个软件共同管理,作为应用程序与数据的接口。这一阶段数据与程序的关系如图 1-4 所示。

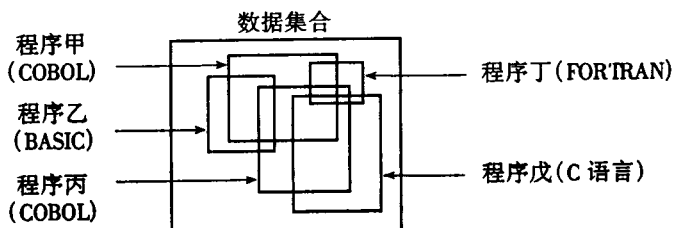


图 1-3 数据共享

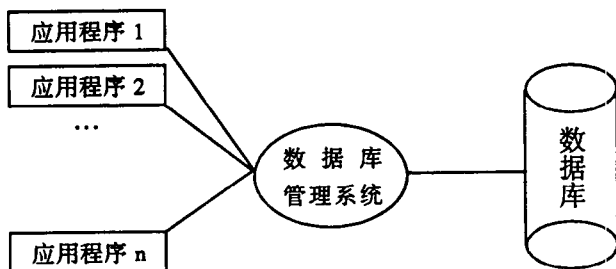


图 1-4 数据库系统阶段数据与程序之间的关系

文件系统中数据是分散的,应用程序对应着各自的数据文件,而在数据库系统中,数据被集中进行管理,就象货物仓库中的物资一样,用户需要什么数据就去库中提取。因此,有人形象地把这样的系统称为“数据库”(Data Base)。

数据库概念的一个重要贡献就是应用系统中的所有数据独立于各个应用而由 DBMS (数据库管理系统)统一管理,实现了数据资源的整体管理。

那么,什么是数据库呢?数据库是一个复杂的系统,给它下一个确切的定义是困难的,目前还没有一个公认的、统一的定义。

但具体地,对一个特定数据库来说,它是集中、统一地保存、管理着某一单位或某一领域内所有有用信息的系统,这个系统根据数据间的自然联系结构而成,数据较少冗余,且具有较高的数据独立性,能为多种应用服务。

从以上叙述可以看出,数据库系统是管理数据且为不同应用服务的工具。它所管理的数据是大量的有关某一方面的信息,需要较长时间的保存;它所提供的数据应该是正确的、可靠的、高效率的。

这样一个数据库系统至少应该包括以下三个部分:

(1) 数据库。一个结构化的相关数据的集合,包括数据本身和数据间的联系。它独立于应用程序而存在,是数据库系统的核心和管理对象。

(2) 物理存储设备。这是保存数据的硬件介质,一般为磁盘等大容量的存储器。

(3) 数据库软件。负责对数据库管理和维护的软件。具体完成对数据定义、描述、操作、维护等功能,接受并完成用户程序及终端命令对数据库的不同请求,并负责保护数据免受各种干扰和破坏。数据库软件的核心是数据库管理系统(Data Base Management System——DBMS)。

另外,数据库系统需要有专门的人员负责数据库的设计、建立、执行和维护,这些人员称为数据库管理员(Data Base Administrator——DBA)。

1.1.2 数据库系统的发展简史

1969年,美国IBM公司研制了世界上第一个数据库管理系统IMS(Information Management System),IMS的数据模型是层次结构的;美国CODASYL委员会(Conference On Data System Language)的DBTG(Data Base Task Group)小组对数据库方法进行了系统的研究、讨论,于60年代末至70年代初公布了若干研究报告,称为DBTG报告,它所提议的方法是基于网状结构的;1970年IBM公司的研究员E.F.Codd博士发表了题为“大型共享数据库数据的关系模型”的著名论文,提出了数据库的关系模型,开创了数据库关系方法和关系数据理论的研究,为数据库技术奠定了理论基础,被公认为是数据库发展历史上的一个重要里程碑。由于E.F.Codd的杰出工作,他于1981年获得了ACM图灵奖。这三大事件标志着数据处理已进入了数据库技术的新时代。

70年代,数据库技术得到了迅速发展。

首先,数据库方法,特别是DBTG方法和思想应用于各种计算机系统,出现了许许多多商品化数据库系统。它们大都是基于网状模型和层次模型的。

其次,这些商用系统的运行使数据库技术日益广泛地应用到企业管理、交通运输、情报检索、军事指挥、政府管理和辅助决策等各个方面。深入到人类生产和生活的各个领域。数据库技术成为实现和优化信息系统的基本技术。

第三,关系方法的理论研究和软件系统的研制取得了很大成果。IBM公司在IBM370系列机上研制关系数据库实验系统System R获得成功(1974—1979)。1981年IBM公司又宣布了具有System R全部特征的新的数据库软件产品SQL/DS问世。与此同时,美国加州大学柏克利分校也研制了INGRES关系数据库实验系统。并紧接着出现了商用INGRES系统。使关系方法从实验室走向社会。因此,在计算机领域中,有人把70年代称为数据库时代是不无道理的。

进入80年代,关系数据库由于结构简单、操作方便、并有扎实的理论基础,而得到长足的发展,有关关系型数据库的研究和产品日益增多,80年代后几乎所有新开发的数据库系统均是关系型的,关系型数据库成为数据库发展的主流。较有代表的商品化关系数据库软件还有:ORACLE、SYBASE、INFORMIX、DB2等。同时,各种微机关系型数据库系统不断出现,性能越来越好,功能越来越强,并得到了广泛应用。最有代表性的产品就是FOX系列产品,它从DBASE发展到FOXBASE,又发展到FOXPRO。

数据库系统已从第一代网状、层次数据库系统发展到第二代关系数据库系统和第三代以面向对象为主要特征的数据库系统。数据库技术与网络通信技术、面向对象技术、并行计算技术、多媒体技术、人工智能技术等互相渗透,互相结合,成为当前数据库技术发展的主要特征。它使数据库领域中新的技术内容层出不穷,新的学科分支不断涌现,形成了新一代数据库系统的大家族。例如,分布式数据库、面向对象数据库、数据仓库、并行数据库、实时数据库、时态数据库、演绎数据库、知识数据库、智能数据库、专家数据库、多介质数据库、工程数据库、空间数据库、主动数据库、模糊数据库、科学统计数据库、移动数据库以及电子商务数据库等技术。这些数据库技术大多处于数据库研究前沿,由于很有应用前景,吸引了国内外很多学者正在从事这些方面的研究。

回顾数据库技术和系统的发展历史,数据库领域所取得的成就主要体现在关系数据库、事务管理和查询优化等方面。关系数据库由于其简单性和清晰的概念基础,得到了研究人员、数据库厂商和最终用户的青睐,形成了数据库目前的繁荣局面。事务管理是DBMS

支持数据共享和多用户操作的关键,是 DBMS 保持数据正确性及简化应用编程人员工作(无须关心其它并发使用同一数据的应用程序的干扰)的基本措施。查询优化是数据库系统性能提高的基础,尤其是在关系数据库系统中,由于系统性能主要由系统自身负责(这是关系系统之所以简单易用的原因之一),查询优化显得更为重要。从某种意义上说,关系数据库得以取代层次和网状型数据库而成为市场主宰,查询优化技术突破是一个重要因素。

1.2 数据模型

数据库是一个结构化的数据集合,这个结构不是人们随心所欲确定的,而是根据现实世界中事物本身的属性及其之间的联系来确定的。现实世界中的事物是彼此关联的,任何一事物都不可能孤立存在,因此反映客观世界事物本身及其联系的数据库结构也必然横竖交错。文件系统忽略了这种联系,因而简单,但同时存在各种弊端。数据库系统不仅考虑事物本身,更要考虑事物之间的联系,因而复杂,但同时获得各种优点。

在数据库中,用数据模型这个工具来对现实世界进行抽象,数据模型是数据库系统中用于提供信息表示和操作手段的形式构架。不同的数据模型是提供给我们模型化数据和信息的不同工具。根据模型应用的不同目的,可以将数据模型分为两种类型:一种是独立于计算机之外的,如实体-联系模型,这种模型不涉及信息在计算机中如何表示,而是用来描述某一特定范围内人们所关心的信息结构,是按用户的观点来对数据和信息建模,这种模型常称为概念模型或信息模型;另一类模型是直接面向计算机的,是按计算机系统的观点对数据进行建模,常称为基本数据模型或数据模型,数据库中常用的基本数据模型有层次、网状和关系模型。

1.2.1 数据模型的三要素

数据模型是现实世界中的各种事物及其间的联系用数据和数据间的联系来表示的一种方法。一般地讲,数据模型是严格定义的概念的集合,这些概念精确地描述系统的静态特性、动态特性和完整性约束条件。因此,数据模型通常由数据结构、数据操作和完整性约束三部分组成。

一、数据结构

数据结构是所研究对象和对象具有的特性、对象间的联系的集合,它是对数据静态特性的描述。这些对象是数据库的组成部分。如网状模型中的数据项、记录、系型,关系模型中的域、属性、关系等。

在数据库系统中,通常按照数据结构的类型来命名数据模型,如层次结构、网状结构和关系结构的模型分别命名为层次模型、网状模型和关系模型。

二、数据操作

数据操作是指对数据库中各种对象(型)的实例(值)允许执行的操作的集合,包括操作及有关的操作规则。通常对数据库的操作有检索和更新(包括插入、删除和修改)两大类,这些操作反映了数据的动态特性,因为现实世界中的实体及实体间的联系是在不断变化的,数据模型应能反映出这种变化。

三、数据的约束条件

数据的约束条件是完整性规则的集合。完整性规则是给定的数据模型中数据及其联

系所具有的制约和依存规则,用以限定符合数据模型的数据库状态以及状态的变化,以保证数据的正确、有效、相容。

数据模型的这三个方面内容完整地描述了一个数据模型。而数据的结构是刻画模型性质最重要的方面。

1.2.2 概念模型与实体—联系方法

数据模型是数据库系统的核心和基础。为了把现实世界中的具体事物抽象、组织为 DBMS 支持的数据模型,人们常常首先将现实世界抽象为信息世界,然后将信息世界转换为机器世界,如图 1-5 所示。因此概念模型是现实世界到机器世界的一个中间层次。

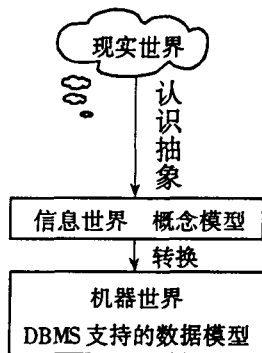


图 1-5 数据管理的三个世界

现实世界的事物通过认识和抽象反映到人的脑子中来形成信息世界。所谓信息,是客观世界中存在的事物在人们头脑中的反映,人们把这种反映用文字、图形等形式记录下来,经过命名、整理、分类就形成了信息。

在信息领域中,客观存在并可相互区分的事物叫实体,如一个人、一本书、一次活动等都是实体。属性表示实体所具有的特征,如姓名、年龄、性别、身高等是表示人这个实体所具有的属性,一个实体是由它的若干个属性来表示的。唯一标识实体的属性集称为码,例如,学号是学生实体的码。具有相同属性的实体具有共同的特征和性质,用实体名及若干属性名的集合来描述和刻划同类实体,形成实体的类型,简称实体型,例如,学生(学号、姓名、年龄、性别、成绩)是一个实体型。同型实体的集合称为实体集,例如,全体学生就是一个实体集。

现实世界的事物之间是有联系的,即实体间的联系。这种联系必然要在信息世界中加以反映,一般存在两类联系:一类是实体内部的联系,即实体中属性间的联系;另一类是实体间的联系。在数据模型中不但要考虑实体属性间的联系,更重要的是实体与实体间的联系。下面主要讨论实体间的联系。实体间的联系是错综复杂的,但就两个实体型间的联系来说,有以下三类:

(1) 一对一的联系(1:1)。若对于实体集 A 中的每一个实体,实体集 B 中至多有一个实体与之联系,反之亦然,则称实体集 A 与实体集 B 具有一对一联系,记为 1:1。

(2) 一对多的联系(1:n)。若对于实体集 A 中的每一个实体,实体集 B 中有 n 个实体($n \geq 0$)与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中至多有一个实体与之联系,则称实体集 A 与实体集 B 具有一对多联系,记为 1:n。

(3) 多对多的联系(m:n)。若对于实体集 A 中的每一个实体,实体集 B 中有 n 个实体($n \geq 0$)与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中也有 m 个实体($m \geq 0$)与

之联系,则称实体集 A 与实体集 B 具有多对多联系,记为 $m:n$ 。

例如,一个班级有一个班长,而每个班长只在一个班任职,则班级与班长之间具有一对一的联系。一个班级有若干个学生,而每个学生只在一个班级,则班级与学生之间是一对多的联系。一个学生可以选修一门课程,一门课程可以被若干个学生选修,则学生与课程之间是多对多的联系。

概念模型的表示方法最常用的是实体-联系方法(Entity-Relationship Approach),通常称为实体-联系模型,简称为 E-R 模型,它是 P.P.S. Chen 于 1976 年提出来的。这个方法是用 E-R 图来描述某一组织的概念模型,即用 E-R 图来表示某一组织的信息结构,它在数据库设计时非常有效,曾在计算机界引起较大反响。这里仅介绍 E-R 图的要点。

E-R 图中使用的基本符号有:

(1) 用长方形表示实体型,在框内写上实体名。

(2) 用椭圆形表示实体的属性,椭圆内写上属性名,并用无方向线段把实体与其属性连接起来。如,职工实体具有编号、姓名、性别、年龄、婚否、工作日期、职称等七个属性,则用 E-R 表示如图 1-6 所示。

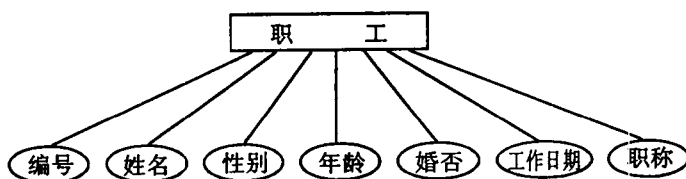


图 1-6 职工实体及属性

(3) 用菱形表示实体间的联系,菱形框内写上联系的名称,并用无方向线段将菱形分别与有关实体相连接,在线段旁标上联系的类型。若实体之间联系也具有属性,则把属性和菱形也用线段连接上。如图 1-7 表示了班级与班长、班级与学生、学生与课程及教师之间的联系。

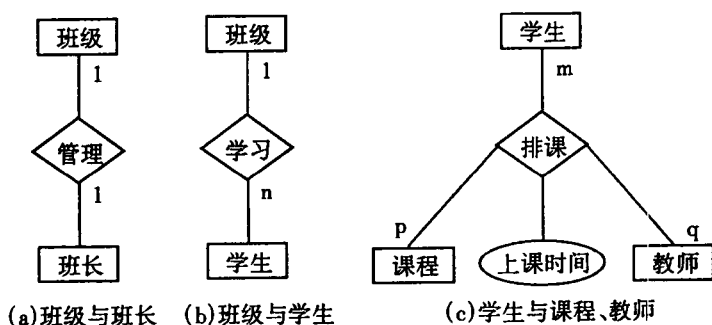


图 1-7 实体与实体间的联系

【例 1.1】用 E-R 图来表示某个学校教学管理的概念模型。

教学管理涉及的实体有:

- (1) 学生。属性有:学号、姓名、性别、籍贯。
- (2) 班级。属性有:班级号、班级名。
- (3) 课程。属性有:课程号、课程名、课时、学分。
- (4) 教师。属性有:编号、姓名、职称。
- (5) 教室。属性有:教室号、地点、容量、电话。

(6) 系。属性有:系编号、系名、办公地点、电话。

这些实体之间的联系如下:

(1) 学生、课程与教师之间的联系。每个学生可以选修若干门课程,每门课程同时有若干个学生选修;每个学生同时可以选修若干个教师讲授的课,每个教师同时教若干个学生;每门课程同时可能有多个教师讲课,每个教师同时可能讲授多门课程。因此,学生、课程和教师三者之间是多对多的联系,且可以用“上课时间”来描述“排课”联系的属性。它们之间的联系如图 1-7(c)所示。

(2) 教室与课程之间的联系。同一时间一个教室只能上一门课,一门课也只能在同一个教室上课,因此在某给定时间教室与课程之间是一对一的联系,且用“上课时间”来描述“上课”联系的属性。

(3) 系与教师之间的联系。一个教师只能属于一个系,一个系可以包括多个教师,因此系与教师之间是一对多的联系。

(4) 班级与学生之间的联系。同理,班级与学生之间也是一对多的联系。

(5) 系与班级之间的联系。同理,系与班级之间也是一对多的联系。

(6) 教师之间具有领导与被领导关系,即系主任领导若干个教师,因此教师实体集中具有一对多的联系。

(7) 学生之间也具有领导与被领导关系,即班长领导若干个学生,因此学生实体集中具有一对多的联系。

图 1-8 给出了某学校教学管理概念模型的 E-R 图,其中图 1-8(a)中给出的是该概念模型中涉及的实体及实体的属性,图 1-8(b)中给出的是该概念模型中涉及的实体以及实体间的联系。

E-R 模型是抽象和描述现实世界的有力工具,这种模型的关键是如何划分实体属性和实体间的关系。有些信息可用属性表示,有些信息可用实体关系表示。例如,班级与学生这两个实体集是一对多的关系,这里有三种划分方法:

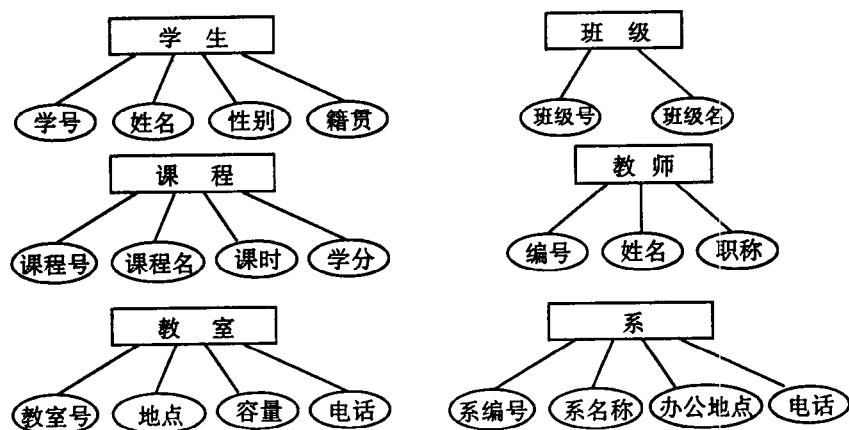
- (1) 学生作为实体,班级作为它的属性;
- (2) 学生和班级都作为实体,再建立两实体间的联系;
- (3) 班级作为实体,学生作为它的属性。

究竟采用哪种方法合适?这要看具体情况。如果班级本身只有一个信息表示,如班级号或班级名,则方法(1)较好;如果班级还有其它信息表示,则方法(2)较好。但不管什么情况,方法(3)总是不好的,而方法(1)、(2)总是可以的。由此可见,在 1-1 关系和 1-n 关系中,表示 1 的实体如果只有一个信息需要表示,则把它作为另一个实体的属性比较合适。其它情况下都应作为两个实体来考虑,并用两实体间的联系表示所需信息。

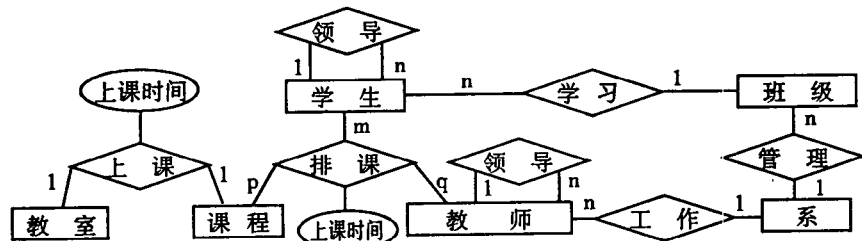
1.2.3 三种主要的数据库模型

一、层次模型

用树形结构来表示实体之间联系的模型叫层次模型。在这种树形结构中,树是由结点和连线组成的。结点表示实体集(记录型),连线表示两实体之间的联系。这种联系只能是 1:m 联系(含 1:1 联系)。通常把表示 1 的实体放在上方,称为双亲结点,或称为父结点;表示 m 的实体放在下方,称为子女结点,或称为孩子结点。树的最高位置上只有一个结点,称为根。根以外的其它结点都有一个双亲结点与它相连,可能同时有一个或多个子女结点与它相连,同一双亲的子女结点称为兄弟结点,没有子女结点的结点称为叶结点。



(a) 实体及其属性图



(b) 实体及其实体间的联系图

图 1-8 某学校教学管理 E-R 图

层次模型必须满足以下两个条件:

- (1) 有且仅有一个结点无双亲,这个结点就是根结点;
- (2) 其它结点有且仅有一个双亲。

对于多对多的关系,在层次模型中不能直接表示,必须先将该关系分解为两个一对多的关系,然后再用层次模型来表示。

图 1-9 是学校管理信息系统的层次模型(部分)。

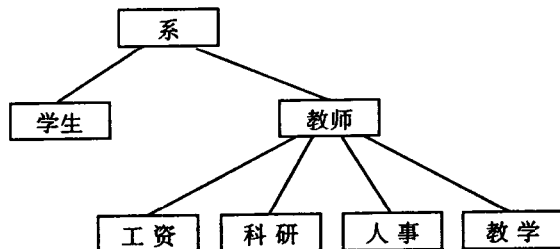


图 1-9 学校管理信息系统的层次模型(部分)

对于层次模型,实体与实体之间的联系在计算机中是通过指针来实现的。

二、网状模型

如果取消层次模型中的两个限制,即允许有一个以上的结点没有双亲且每一个结点可以有多个双亲结点,便形成了网结构。用网结构来表示的实体及其联系的模型叫网状模型。如图 1-10(a)、(b)给出的都是网状模型。

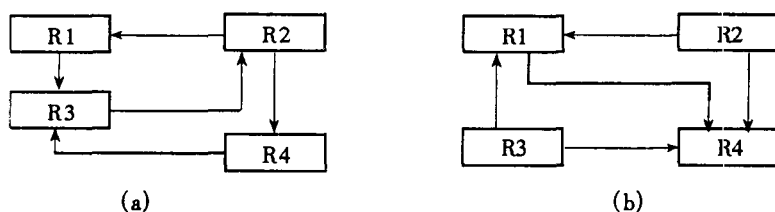


图 1-10 网状模型

层次模型与网状模型不同之处主要有两点：

- (1) 层次模型中从子女到双亲的联系是唯一的，而网状模型则可以不唯一。
- (2) 网状模型中允许复合链，而层次模型是不可以的。所谓复合链是指两个结点之间存在着不只一种联系。

网状模型是基于图来组织数据，实体与实体之间的联系在计算机中也是通过指针来实现的，对数据的访问和操纵需要遍历数据链来完成。

三、关系模型

在关系模型中，数据是按照数学中关系概念来组织和理解的，对用户而言，数据库中按照关系形式组织的数据可简单直观地理解为二维表。表格是大家所熟悉的数据表示方法，例如学生选课常用到下面三张表格：

- (1) 学生基本情况表(表 1-1)；
- (2) 所开课程一览表(表 1-2)；
- (3) 学生选课表(表 1-3)。

表 1-1 学生基本情况表

学 号	姓 名	性 别	籍 贯	所 在 系
960101	李梅花	女	浙江杭州	法律系
960102	王海南	男	海南三亚	计算机系
960206	万川流	男	四川万县	会计系
960214	张桂花	女	云南昆明	经济系

表 1-2 所开课程一览表

课程号	课 程 名 称	课 时	学 分
A0101	邓小平理论	54	3
I0101	计算机原理	90	5
I0301	计算机网络	72	4

表 1-3 学生选课表

学 号	课 程 号	选 课 学 期
960101	A0101	1996 下
960101	I0101	1997 上
960102	A0101	1996 下
960102	I0301	1998 下
960206	I0101	1997 上
960214	A0101	1996 下
960214	I0101	1997 上
960214	I0301	1998 上

用表格数据来表示实体及实体间联系的模型叫关系模型。在关系模型中,不仅实体集用表格来表示,如表 1-1 表示的是学生实体集,表 1-2 表示的是课程实体集,而且实体间的联系也是用表格来表示,如表 1-3 表示的是学生实体与课程实体之间的联系“选课”的内容。这里的联系是通过存放两实体的关键字(码)来实现的。学生选课表中的学号和课程号分别是学生实体集和课程实体集中的关键字,这个表可以告诉我们“谁选了什么课?”以及“某门课共有哪些人选修?”。这里课程号是课程的标识,要想知道该课程的具体情况,可根据课程号到所开课程一览表中去查找;反之,学号是学生的标识,要想知道该学生的具体情况,可根据学号到学生基本情况表中去查找。

由此可见,关系模型在本质上是不同于层次和网状模型的。其本质之差异就在于,关系模型是通过表格数据而不是通过指针链接来实现两实体间的联系的。

关系模型是建立在数学概念基础上的,这里我们并不去给出关系的确切数学定义,只是给出它的形式化的描述。在关系模型中,把数据看成二维表中的元素,而一张二维表就是一个关系,可以给它取个名字即为关系名。表中的一行称为一个元组,它相当于记录值。表中的一列称为属性,给它命个名即为属性名,它相当于数据项或字段。表中能够唯一标识一个元组的属性组称为主码。主码可以是一个属性,如学号是学生基本情况表的主码;主码也可以是若干个属性的集合,如学号与课程号共同构成学生选课表的主码。属性名的取值范围叫做域。用关系名和它所包括的属性名来对关系进行描述,对关系的描述称为关系模式,如:学生(学号,姓名,性别,籍贯,所在系)。

关系模型具有下列特点:

(1) 关系模型的概念单一。在关系模型中,无论是实体还是实体之间的联系都用关系来表示。如:

学生(学号,姓名,性别,籍贯,所在系)

课程(课程号,课程名称,课时,学分)

学生选课(学号,课程号,选课学期)

(2) 关系必须是规范化的关系。所谓规范化是指关系模型中,每一个关系模式要满足一定的要求或者称为规范条件。规范条件很多,这里不详细讨论,只列出最基本的要求:关系中的每一个属性都是不可再分的基本属性,即不允许表中还有表;各元组相异,不允许重复,即关系中必须要有主码。

1.3 数据库系统的结构与功能特征

在数据库系统中,用户可以逻辑地、抽象地处理数据,而不必考虑数据在计算机中是如何进行组织、存放的。要提供这一功能,数据库系统结构应该是一个多级结构。它应具有既能让用户方便地存取数据,又能高效地组织数据,使数据在物理存储器上以最佳形式进行存放的能力。现有的数据库系统虽然各不相同,但在总的体系结构上都是三级结构。

1.3.1 数据库系统的三级模式结构

数据库系统的结构一般划分为三个层次,叫做三级模式,分别为子模式、模式和存储模式。它们之间的关系如图 1-11。

1. 模式。亦称为概念模式或逻辑模式。概念模式的概念已在上一节介绍过了,它是数据库的总框,是对数据库中全体数据的逻辑结构和特性的描述。

2. 子模式。亦称外模式,是数据库用户的数据视图。它体现了用户的数据观点,是对

用户数据的逻辑描述。其内容与模式描述大致相同。

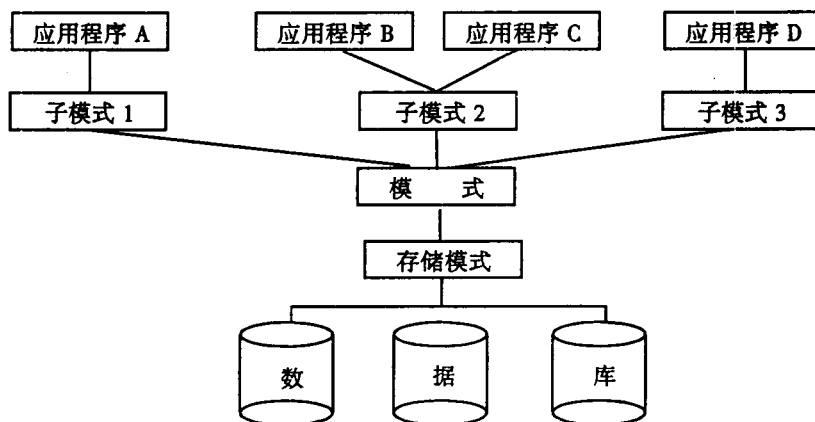


图 1-11 数据库系统的三级模式结构

子模式通常是模式的一个子集,也可以是整个模式。所有的应用程序都是根据子模式中对数据的描述而不是根据模式中对数据的描述编写的。子模式也可以共享,在一个子模式上可以编写多个应用程序,但一个应用程序只能使用一个子模式。根据应用不同,一个模式可以对应多个子模式,子模式可以相互覆盖。

3. 存储模式。亦称内模式,是对数据库在物理存储器上具体实现的描述。它规定数据在存储介质上的物理组织方式、记录寻址技术,定义物理存储块的大小,溢出处理方法等。它与模式是相对应的。

数据库系统的三级模式结构将数据库的全局逻辑结构同用户的局部逻辑结构和物理组织结构分别开来,给数据库的组织和使用带来了方便。不同用户可以有自己的数据视图,所有用户的数据视图集中起来统一组织,消除数据冗余,得到全局数据视图。全局数据视图经数据存储描述语言定义和描述,得以在设备介质上存储。这中间进行了两次转换,一次是子模式与模式之间的映象,一次是模式与存储模式之间的映象。

子模式与模式之间的映象定义了它们之间的对应关系,通常包含在子模式中。当全局逻辑结构因某种原因改变时,只要修改子模式与模式之间的对应关系,而不必修改局部逻辑结构,相应的应用程序亦可以不必修改,实现了数据的逻辑独立性。

模式与存储模式之间的映象定义了数据逻辑结构和物理存储之间的对应关系。当数据库的物理存储结构改变时,需要修改模式与存储模式之间的对应关系,而保持模式不变。使全局逻辑结构独立于物理数据,提供了数据的物理独立性。

1.3.2 数据库管理系统的组成

数据库管理系统(DBMS)是数据库软件的核心,它建立在操作系统的基础之上,对数据库的所有操作(数据库的建立、使用和维护等)进行统一的管理和控制,是一种综合的通用的大型系统软件。它的最基本功能就是允许用户逻辑地、抽象地处理数据而不必过问这些数据在计算机中是如何处理的细节。

数据库管理系统通常由三部分组成:

- 语言,用来定义和使用数据库;
- 数据库管理控制程序,负责数据库运行时的控制和管理;
- 数据库服务程序,用来建立和维护数据库。

下面简单介绍它们的内容和功能。

(1) 数据描述语言。数据库的数据描述语言包括模式描述语言、子模式描述语言和数据库存储描述语言。

① 数据描述语言(DDL 语言),是用来定义数据库的全局逻辑结构。

② 子模式描述语言(SDDL 语言),是用来定义用户的局部逻辑结构。

③ 数据库存储描述语言(DSDL 语言),是用来描述物理数据库的结构。

(2) 数据操纵语言(DML 语言)。它完成对数据库的基本操作:检索、插入、删除、修改。DML 语言有两大类:自含型 DML 语言和宿主型 DML 语言。

自含型 DML 语言可以交互使用。它一般具有检索、修改、建库等功能,使用方便,易于学习,主要用于查询。

宿主型 DML 语言不能单独执行,它必须嵌入主语言(如 PASCAL、C 等高级语言)中使用。它是应用程序与数据库管理程序之间的接口,DML 语言主要实现对数据库的读、写操作,主语言完成数据处理过程的控制、数据的输入/输出控制等功能。因此 DML 语言是对主语言的扩充。

(3) 数据库管理控制程序。对数据库的所有操作都是在数据库管理控制程序的统一管理下进行的。它大致包括以下几种例程:

① 系统主控程序,它监督 DBMS 的各种活动,使其它例程正确工作;

② 存取控制程序,它检查用户标识、口令、权限,决定是否允许对数据库的访问;

③ 并发控制程序,它处理多用户同时访问数据时的并发操作;

④ 数据有效性检查程序,它根据数据的约束条件检查数据的有效性、一致性;

⑤ 数据保护程序,它负责数据保密、故障恢复等;

⑥ 数据存取程序,它执行数据的读、写操作;

⑦ 数据更新程序;

⑧ 通讯程序,它完成应用程序或终端与 DBMS 间的通讯。

(4) 数据库服务程序。为保证数据库系统的正常运行,除上述管理控制程序外,还有一系列服务程序。它包括:

① 转储程序,它是为了防止系统出现故障而进行定期数据备份的程序;

② 数据装入程序,它用于把大量原始数据成批存储到外存介质上形成数据库;

③ 日志程序,该程序将所有对数据的修改都记录下来,以备出现故障后恢复时使用;

④ 恢复程序,当数据库遭到破坏时,该程序可以利用备份数据和日志文件等进行恢复;

⑤ 统计分析程序,该程序对数据存取情况、空间利用情况进行统计分析;

⑥ 重组程序,它用于重新组织数据库。

此外,还包括对无用数据的收集及空间的再分配等程序。

现代的数据库系统,为了提高开发应用系统的效率,提供了各种应用开发工具。如应用生成器,报表生成器,电子表格软件,图形系统等,这些软件以数据库管理系统为核心,直接支持数据库系统的应用开发。

1.3.3 数据库系统的结构

数据库系统是一个复杂的系统,除了数据库管理系统 DBMS 外,它还包括它的支持系统及工作环境。通常数据库系统由三个部分组成:

(1) 硬件资源及数据库。即计算机设备及其存放在外存上的数据仓库。由于数据库系统的系统软件除操作系统外还有 DBMS,而且数据库数据庞大,因此数据库系统需要硬件能提供足够大的外存和内存空间。

(2) 软件资源。数据库系统的软件资源主要包括:支持 DBMS 运行的操作系统、主语言(即高级语言)、DBMS 及应用软件。

(3) 人员。管理、开发和使用数据库系统的人员主要包括：数据库管理员(DBA)、系统分析员、应用程序员和用户。数据库系统中的不同人员涉及不同的数据抽象级别，具有不同的数据视图，如图 1-12 所示。

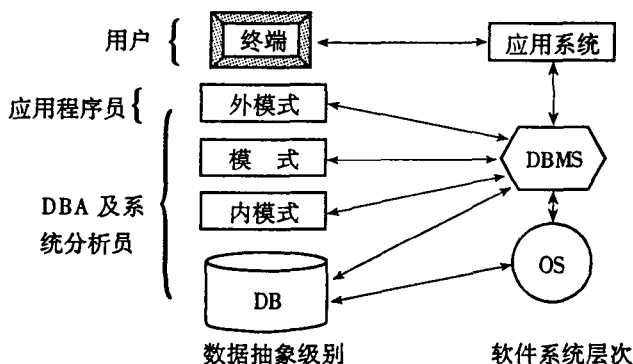


图 1-12 各种人员的数据视图

在数据库系统中，DBA 是数据库管理机构的一组人员，他们负责全面地管理和控制数据库系统，具体的职责包括：

- ① 决定数据库的信息内容和结构。究竟在数据库中存放哪些信息最终由 DBA 决定。为此，他们必须参与数据库设计的全过程，和用户、应用程序员、系统分析员紧密配合，设计概念模式。
- ② 决定数据库的存储结构和存取策略。DBA 要综合各用户的应用要求，和数据库设计人员共同决定数据的存储结构和存取策略，以使存储空间利用率和存取效率达到较优。
- ③ 定义数据的安全性要求和完整性约束条件。
- ④ 监督和控制数据库的使用和运行。
- ⑤ 数据库系统的改进和重组。

1.3.4 数据库系统的功能特征

数据库技术聚集了数据处理最精华的思想，是管理数据最先进的工具。与文件系统相比，它有如下一些主要特征：

(1) 数据共享。实现数据共享是数据库技术发展的一个主要原因，也是数据库系统的一个重要特征。数据库中的数据可以供多个用户所使用，每个用户只与库中的一部分数据打交道；用户数据可以是重叠的，在同一时刻不同的用户可以同时存取数据而互不影响，大大地提高了数据的利用率。

(2) 数据独立性。在文件系统中，应用程序不但与数据文件相互对应，而且与数据的存储和存取方式密切相关。在数据库系统中，应用程序不再与物理存储器上存储的具体文件相对应，每个用户所使用的数据有其自身的逻辑结构。在数据库方式下，数据独立性表现在物理独立性和逻辑独立性两个方面。这已经在 1.3.1 中介绍过。数据独立性给数据库的使用、调整、优化和进一步扩充带来了方便，提高了数据库应用系统的稳定性，减少了程序员的负担。

(3) 减少数据冗余。在数据库系统管理下的数据，不再是面向应用，而是面向系统。数据集中管理，统一进行组织、定义和存储，避免了不必要的冗余，因而也避免了数据的不一致性。在实际应用系统中，为提高数据的存取效率，少量的数据冗余还是存在的，但这些冗余的数据将受到控制，系统负责对冗余数据进行检查和维护。

(4) 数据的结构化。文件系统中的数据从整体上看是无结构的，数据文件之间不存在

任何联系。而数据库系统中的数据是相互关联的,这种联系不仅表现在记录内部,更重要的是记录类型之间的相互联系。

(5) 统一的数据保护功能。多个用户共享数据资源,需要解决数据的安全性、一致性及并发控制问题。为使数据安全、可靠,系统对用户使用数据有严格检查,对非法用户将拒绝进入数据库。不同用户同时使用数据库时,可能会相互干扰而出现数据的不一致,数据库系统具有并发控制功能,以保证数据的正确性。

1.4 数据库的新进展

1.4.1 分布式数据库系统

分布式数据库系统(Distributed Database System)的研究始于20世纪70年代中期,这个时期以来,由于计算机网络通信技术的迅速发展,以及地理上分散的公司、团体和组织对于数据库更为广泛应用的需求,在集中式数据库系统成熟技术的基础上产生和发展了分布式数据库系统。分布式数据库系统是数据库技术和计算机网络技术两者相互渗透和有机结合的结果。现在,在该领域里一些工作已经取得了显著的成果,许多基本问题已得到解决,一批原型系统已经研究成功并获得了许多经验,一些商品化产品正在试制或已经推出,可以说分布式数据库系统的技术已相当成熟,产品化的时代已经到来。

什么叫分布式数据库? 分布式数据库是一个逻辑上属于同一整体而物理上分布存放在计算机网络不同结点(亦称为场地)上的数据集合。这里要强调两点:

(1) 分布性。数据库中的数据不是存储在同一场地,而是分布存放在网络的各结点上。这一点很容易与集中式数据库相区别。

(2) 逻辑整体性。这些数据逻辑上是互相联系的,是一个整体(逻辑上如同集中数据库)。这就可以和分散在计算机网络不同结点上的数据库或文件的集合相区别。后者各结点的数据之间没有内在的逻辑联系。所以在讨论分布式数据库时就有了全局数据库(逻辑上)和局部数据库(物理上)的概念。

一、分布式数据库的特点

1. 集中控制与局部自治

数据库是用户共享的资源。在集中式数据库中,为了保证数据库的安全性和完整性,对共享数据库的控制是集中的,并设有 DBA 来负责监督和维护系统的正常运行。

在分布式数据库中,数据的共享有两个层次:一是局部共享,即在局部数据库中存储局部场地上各用户共享的数据,这些数据是本场地用户常用的;二是全局共享,即在分布式数据库的各个场地也存储可供网中其它场地的用户共享的数据,支持系统中的全局应用。因此,相应的控制机构也有两个层次:集中和自治。各局部的 DBMS 可以独立地管理局部数据库,具有自治的功能。同时,系统又设有集中控制机构,协调各局部 DBMS 的工作,执行全局应用。当然,不同的系统,集中和自治的程度不尽相同,有的集中控制的功能更强一些,有的高度自治,连全局应用事务的协调也由局部 DBMS、局部 DBA 共同承担。

2. 数据独立性

数据独立性是数据库系统最主要的优点之一。它使应用程序不受实际数据结构的影响,因而,当数据库逻辑结构和物理结构改变时不需要修改应用程序,即实际结构对应用程序是透明的。

在分布式数据库系统中,数据独立性除了集中式数据库所具有的逻辑和物理独立性