



新世纪高职高专信息管理与信息技术专业教材

赵龙强 张雪凤 编著

数据库原理与应用

上海财经大学出版社

新世纪高职高专信息管理与信息技术专业教材

数据库原理与应用

赵龙强 张雪凤 编著

 上海财经大学出版社

图书在版编目(CIP)数据

数据库原理与应用/赵龙强,张雪凤编著. —上海:上海财经大学出版社,2001.2

新世纪高职高专信息管理与信息技术专业教材

ISBN 7-81049-540-2/TP·04

I. 数… II. ①赵… ②张… III. 数据库系统-高等学校:技术学校-教材 IV. TP311.13

中国版本图书馆 CIP 数据核字(2001)第 05052 号

SHUJUKU YUAN LI YU YINGYONG
数据库原理与应用

赵龙强 张雪凤 编著

责任编辑 麻俊生 封面设计 周卫民

上海财经大学出版社出版发行
(上海市中山北一路 369 号 邮编 200083)

网 址: <http://www.sufep.com>

电子邮件: webmaster@sufep.com

全国新华书店经销

上海崇明晨光印刷厂印刷装订

2001 年 4 月第 1 版 2001 年 4 月第 1 次印刷

787mm×960mm 1/16 21.75 印张 462 千字
印数:0 001—4 000 定价:34.00 元

本版图书如有印刷装订错误,请向承印厂调换

前 言

数据库技术是计算机学科的一个重要分支,作为数据管理最有效的手段,它极大地促进了计算机应用的发展。数据库技术是计算机信息系统与应用系统的核心技术,它是信息管理系统开发的重要基础。

中文 Access 2000(简称 Access)是 Windows 环境中功能强大最为流行的关系数据库管理系统,它提供了与其他数据库管理软件的良好接口,并与 Internet 紧密联系,拥有完整的程序开发工具 Visual Basic for Applications(简称 VBA),用户可以借此方便地使用和管理数据库。

本书共十四章,由两大部分组成。第一部分讲述数据库的基本概念、基本理论和基本技术。第二部分讲述应用技能,介绍了 Access 关系数据库的操作。

本书是高职高专信息管理专业和计算机专业学生的基础教材,它在讲述数据库基本概念和原理的基础上,强调理论与实践相结合。为了方便读者学习,每章后面都附有一定量的思考题或习题。

本书由张雪凤编写第 1 章至第 7 章,赵龙强编写第 8 章至第 14 章,并由赵龙强、张雪凤统稿和审定。诚挚感谢上海财经大学冯关源教授审阅本书,并提出许多宝贵意见。由于我们学识浅陋,水平有限,不当之处在所难免,恳请广大读者批评指正。

编者
2001 年 2 月

第一章 数据库系统简介

数据库技术是计算机软件学科的一个重要的分支,也是数据管理的最新技术。数据库技术自 20 世纪 60 年代诞生以来,为计算机收集、存储、加工和利用数据提供了全面的支持,并起到了十分关键的作用,它是当今计算机信息系统的核心技术。

随着 Internet 的发展和经济全球化、市场一体化进程的不断加快,对信息资源的利用已成为各企业或组织在激烈的竞争环境下生存和发展的关键,因此,如何建立一个行之有效的信息系统,保存和充分利用企业或组织内部信息(如企业在生产、销售、财务、库存、人事管理等各种业务活动中所发生的各种数据)和外部信息,以帮助企业做好经营管理工作、增强竞争力,是每个企业或组织都在考虑的问题。而信息系统的核心和基础就是数据库。世界上许多著名企业和组织的实践,都说明了利用数据库技术进行信息管理和处理对企业有着十分重要的意义。

第一节 数据库概述

一、数据库基本概念

数据库基本概念主要包括信息和数据、数据处理和数据管理、数据库、数据库管理系统和数据库系统。

1. 信息和数据

信息是许多学科广泛使用的概念,在不同的领域有不同的含义。在数据处理领域,信息是对现实世界事物存在方式或运动状态的反映。

获取、积累、分析和利用信息是人类认识世界、改造世界的必要过程,而对一个具体的企业或组织来说,对信息资源的合理利用是正确决策的重要依据,也是提高效率、增强竞争力的最主要手段。举例来说,有一家网上书店,希望给其会员提供优质的客户服务,该书店保留了每个会员及其网上买书的信息,由此可以推断不同会员的喜好,并有针对性地给会员提供在线新书信息,以提高网上图书的销售量。

对信息的保存、分析和处理不可能都用人工的方法来解决,必须借助于计算机的帮助。也就是说,将收集到的信息存放到计算机中,由计算机来分析和处理,并得到新的信息,而信息在计算机中是以数据的形式存在的。所谓数据是指用符号记录下来的、可以识别的信息。描述信息的符号有数字、文字、图形、图像、声音,等等,它们都可以经过数字化后存入计算机。

例如,某网上书店非常关心其会员的如下信息:姓名、年龄、性别、工作单位、联系电话、E-mail等,在计算机中就可以这样来表示一个会员:

(“赵晴”,20,“女”,“上海新华医院”,“(021)65335678”,zq@online.sh.cn)

上述数据被赋予的含义是一位名为赵晴的会员,年龄20岁,女,工作单位是上海新华医院,联系电话是(021)65335678,E-mail是zq@online.sh.cn。

综上所述,数据与信息既有区别又是密切相关的,数据是信息的符号表示或载体,信息则是数据的内涵,是对数据语义的解释。但在很多场合下人们并不对数据与信息加以区分,常常将这两个名词混用。

2. 数据处理和数据管理

数据处理包括对各种数据进行的收集、存储、加工与传播等一系列活动。其中数据的收集是指在数据的发生处将它们读入到计算机中。例如,在各种预售火车票或飞机票等的订票点都配备了计算机终端,并由售票员将售票数据输入到计算机中;而在各类百货公司的收银台上,收银员也往往会将商品的销售数据通过计算机终端输入到公司的计算机中;又如大部分的超市一般都配备了条形码阅读器,以便于快速地将商品的销售数据输入到计算机中,等等。数据的存储是指将收集到的数据经过整理后用计算机的存储介质保存起来以备今后使用。数据的加工是指从某些已知的数据出发,推导加工出一些新的数据的过程。数据的传播就是利用计算机通讯设备将数据传送给需要的人员。

数据处理也称为信息处理,因为当把客观事物表示成数据后,这些数据便被赋予了特定的含义,而对这些数据进行处理后又可以形成新的数据,这些新的数据又表示了新的信息,从而为人们提供了不必直接观察和度量事物就可以获得有关信息的手段。在数据处理(或信息处理)的各种活动中,数据加工是最重要的数据处理活动。这些对数据进行加工处理的活动,有的只涉及到简单的数学运算或逻辑运算,如根据每一产品的价格计算所有产品的平均价格,又如将商品单价与销售数量相乘得到销售金额,等等。这些活动通常在企业经营管理的事务处理活动过程中就能完成,并完全可以在预先设计好的计算机程序的指挥下自动实现,这种比较简单的数据加工处理活动称为联机事务处理(Online Transaction Processing,简称OLTP)。另外一些数据加工活动涉及的计算比较复杂,因而必须在人的参与和支配下才能完成,这种加工处理活动称为联机分析处理(Online Analytical Processing,简称OLAP)。

随着数据处理量的日渐增大,需要对数据进行分类、组织、编码、存储、检索和维护,这

样的一个过程称为数据管理,它是数据处理业务的基本环节。数据管理技术的优劣,将直接影响到数据处理的效率。

3. 数据库(DataBase,简称 DB)

所谓数据库就是存放在计算机中的、以一种合理的方法组织起来的、与公司或组织的业务活动和组织结构相对应的各种相关数据的集合,该集合中的数据可以为公司的各级经过授权的人员所共享。

该定义具有以下三种含义:

第一,数据库是存在于计算机中的、与公司或组织的业务活动和组织结构相对应的各种相关数据的一个仓库。

第二,存放在数据库中的数据是按一定的方式组织起来的,而不是杂乱无章地存放的。

第三,数据库是一个共享的信息资源,它可以被企业或组织中的多个经过授权的职员使用。

为便于读者进一步理解数据库,我们可以把数据库比喻成一个图书馆:

- 图书馆是用来存放图书的,而数据库是用来存放数据的;
- 图书馆中的书是存放在许多房间中的,而数据库中的数据是存放在计算机文件中的;
- 图书馆的图书是按照一定的规则存放在书架上的(例如按图书类别存放),数据库中的数据也是按照一定的方式组织起来的;
- 图书馆中的书可以被不同的读者借阅,数据库中的数据也可以被不同的用户使用。

可见,数据库与图书馆有很多相似之处,只是图书馆保管的对象是书,而数据库的对象是数据。当然,图书馆与数据库毕竟是不同的。例如,图书馆中的某一本书在某一时刻只能借给一个读者,其他读者想借阅的话必须等到前面的读者归还了该书以后才可以,而数据库中的某一数据可以同时被多个用户查看。又如,图书馆中可以增加新书,也可以报废一些旧书,但却不可以修改某一本书的内容,而数据库中不光可以增、删数据,还可以修改数据。

4. 数据库管理系统(DataBase Management System,简称 DBMS)

图书馆的书本经常需要管理,如整理书架上的书,修补一些被损坏的书本,增加一些新书,去掉一些旧书,同时也需防止不法分子盗窃书本,等等。同样,存放在数据库中的数据也是需要管理的,这一任务是由数据库管理系统(DBMS)来完成的。

数据库管理系统是位于用户与操作系统之间的一个数据管理软件,主要负责数据库的建立、查询、插入、删除、修改及各种数据控制功能。

5. 数据库系统(DataBase System,简称 DBS)

为了充分利用一个企业或组织的数据资源,除了要有一个数据库和一个功能强大的数据库管理系统软件之外,还必须建立数据库应用系统以供企业或组织中的人员(即最终用户)使用,并委派专职的数据库管理人员(即数据库管理员 DataBase Administrator,简称 DBA)负责数据库的创建与维护工作,规定严密的制度来保护作为企业或组织极为重要的资源的业务数据。我们把由数据库、操作系统、数据库管理系统(及其开发工具)、数据库应用系统、数据库管理员及其他人员所组成的系统称为数据库系统。

数据库系统并不是一开始就有的,它是应数据管理的需要而逐步发展并日益完善起来的。

二、数据管理技术的三个阶段

从第一台计算机诞生到现在,计算机的硬件和软件以一种近乎神奇的速度发展着,计算机的应用范围也从原来的科学计算扩展到了很多新的领域,如计算机辅助设计(CAD)、计算机集成制造(CIM)、办公信息系统(OIS)等。而数据管理技术也经历了人工管理、文件系统和数据库系统三个阶段。

1. 人工管理阶段(20 世纪 50 年代中期以前)

这一阶段,计算机主要应用于科学计算。在硬件方面,外存没有直接存取存储设备,只有纸带、卡片和磁带等;在软件方面,只有汇编语言,没有操作系统和数据管理方面的软件;数据处理方式是批处理。人工管理阶段数据管理的主要特点是:

(1)数据不保存。由于计算机主要用于科学计算,所以数据是不需要保存的。当用户需要计算某一课题时,就临时将有关数据输入内存,计算完毕后输出运算结果,并释放相应内存空间。

(2)数据由应用程序进行管理,数据与程序不具独立性。由于没有专门的软件进行数据的管理,数据是由应用程序自己管理的。程序员不仅要在程序中设计数据的逻辑结构,还要规定包括数据的存储结构、存取方法和输入输出方式在内的物理结构。而当数据的物理结构或逻辑结构改变时,就必须修改相应的程序。

(3)数据面向应用,不能共享。每个应用程序都只使用自己定义的一组数据,即使有些应用程序使用了相同的一部分数据,也都必须在应用程序中各自定义。

2. 文件系统阶段(20 世纪 50 年代后期~60 年代中期)

这一阶段,计算机不仅用于科学计算,也用于数据管理。在硬件方面,已经有了直接存取存储设备,如磁盘、磁鼓等;软件方面,出现了高级语言和操作系统,并在操作系统中包含了专门的数据管理软件——文件系统;数据处理方式有批处理和联机处理。文件系统阶段数据管理的主要特点是:

(1)数据长期保存在外存的数据文件中,可以进行查询、插入、删除和修改操作。

(2)数据文件由文件系统进行管理,但程序与数据的独立性仍较差。由于有了文件系

统对文件进行管理,程序员在程序中只需用文件名就可与数据打交道,因此不必关心数据的物理存放位置,但文件的逻辑结构仍然是在程序中定义的,所以,当文件的逻辑结构改变时,仍要修改相应的程序,程序与数据的独立性仍较差。

(3)数据文件是面向应用的,相互之间缺乏联系,共享性差。即使不同的应用程序需使用部分相同的数据,也必须存放在各自的一组文件中,不能共享数据。

3. 数据库系统阶段(20 世纪 60 年代末至今)

这一阶段,计算机的应用更加广泛,管理规模日益扩大,数据量不断增加,而且多个应用要求能共享数据集合,这就给数据管理技术提出了更高的要求。与此同时,磁盘技术突飞猛进,出现了数百兆字节容量和快速存取的磁盘,而且价格也不高。但是由于文件系统的缺点,软件的编制与维护却越来越困难,成本也不断上涨,价格不断攀升。需要有新的数据管理技术来减轻程序员进行软件编制和维护的负担,降低软件的成本,这就导致了数据库系统的产生。数据库系统阶段数据管理的主要特点是:

(1)数据存放在数据库中,并用数据模型来描述数据本身及其数据间的联系。主要的数据库模型有层次模型、网状模型、关系模型和面向对象模型,当然这些数据模型是逐步发展起来的,并不是一开始就全部都有的。

(2)数据库中的数据由数据库管理系统进行管理,数据与程序的独立性很高。

在数据库系统阶段,应用程序与数据的对应关系如图 1-1 所示,应用程序不是直接与数据库打交道的,而是通过 DBMS(数据库管理系统)对数据库进行操作的。数据的定义、查询、删除、插入、修改都是由 DBMS 负责的,应用程序员不需要在程序中定义数据,当然也就不需要定义数据的逻辑结构和物理结构了。因此,当数据的逻辑结构或物理结构发生改变时,一般都不需要修改程序,数据和程序的独立性很强。

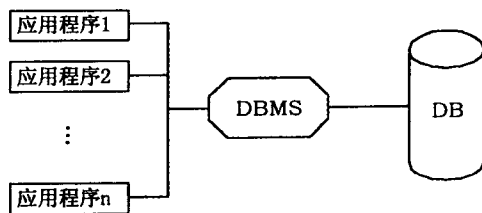


图 1-1

(3)数据是面向系统的,冗余度低,可以共享。

(4)数据库管理系统提供了以下几个方面的数据控制功能:

数据安全性控制:保护数据,防止对数据库的非法操作所引起的数据的丢失、泄露和破坏。

数据完整性控制:保证数据库中的数据永远是正确的、有效的和相容的。

并发控制:避免因多个用户并发进程同时存取、修改数据库时所引起的相互干扰,保证数据的正确性。

数据库的恢复:当数据库中的数据由于种种原因(如系统故障、介质故障、计算机病毒等)而变得不正确,或部分甚至全部丢失时,DBMS有能力将数据库恢复到最近某时刻的一个正确状态。

第二节 数据模型

模型是对现实世界的模拟和抽象。在现实世界中我们经常会接触到模型,例如,在购买房屋时,就会看到房产商为来访者展示的房屋设计模型;在参观市政规划馆时,会看到很多城市规划模型;有时,我们还会观看船模、航模比赛。这些房屋模型、城市规划模型、船模和航模都是对现实世界事物的一种模拟。

数据模型也是模型,它是对现实世界数据特征的描述。一般来说,数据模型应能比较真实地模拟现实世界,比较容易理解,也便于计算机实现。

数据模型的种类有很多,常用的数据模型可分为两种类型:概念模型(即概念数据模型,又称信息模型)和数据模型(即基本数据模型或结构数据模型)。概念模型是一种独立于计算机系统的数据模型,它是按用户的观点来描述某个企业或组织所关心的信息结构。而数据模型是基于计算机系统和数据库的数据模型,它是按某种数据结构来描述数据及相互间的联系。概念模型和数据模型间的关系可参见图 1-2。

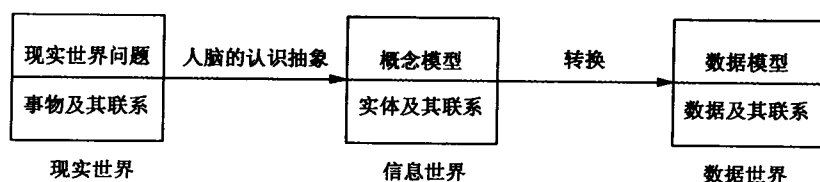


图 1-2

图 1-2 中的现实世界是指存在于人们头脑之外的客观世界,信息世界是指现实世界在人的头脑中的反应,而数据世界是指信息世界中的信息在计算机中的数据存储。现实世界中大量存在的是事物及其联系,例如,一个公司有很多职工,每个职工都在某一个具体的部门工作,职工与职工之间、职工和部门之间、部门与部门之间都是有联系的。为了用数据库系统来解决现实世界中的问题,必须先深入实际,把要解决的现实问题调查清楚,分析与问题有关的事物及其联系,并将其抽象为一种信息结构,即概念模型,它与具体的计算机系统是无关的。这样,现实世界问题中的事物及其联系,就转换成了信息世界中的实体及其联系。而后,还需将信息世界中的实体及其联系转换成数据世界中的数据及

其联系,即把概念模型转换成具体的 DBMS 支持的数据模型。可见,概念模型是从现实世界到数据世界的一个中间层次,主要用于数据库的设计,而数据模型是现实世界问题中的事物及其联系在计算机中的最终表示,主要用于 DBMS 的实现。下面分别介绍这两种数据模型。

一、概念模型

概念模型是对现实世界的第一层抽象,是用户与数据库设计人员之间进行交流的工具。下面先介绍与概念模型有关的概念,而后再介绍用于表示概念模型的实体—联系方法。

1. 基本概念

(1)实体(Entity)。实体是指客观存在、可相互区分的事物。实体可以是具体的对象,如一件产品、一座仓库、一艘船、一幢房子、一个学生、一门课程等;实体也可以是抽象的事件,如一次订货、一次购房、一次选课等。

(2)实体集(Entity Set)。实体集是指同类实体的集合。例如,某个公司的所有产品、某个公司的所有仓库、某个学校的所有学生等。一个实体集的范围可大可小,主要取决于要解决的应用问题所涉及的环境的大小。例如,为解决某个学校的应用问题,那么该校全体学生组成的集合就是一个学生实体集,但如果应用问题与上海市所有的学校有关,那么学生实体集包含的就是上海市的所有学生。

(3)属性(Attribute)和码(Key)。实体集中的所有实体都具有一组相同的特性,如学生实体集中的每个实体都有学号、姓名、年龄、性别、籍贯、系等特性,我们把实体所具有的某一特性称为属性。而把可以唯一标识实体的属性组称为实体的码。例如,在学生实体集中,一个学生的学号可以唯一地对应一个学生,即一个学号有唯一的姓名、年龄、性别、籍贯和系的值与之相对应,这样学号就是学生实体的码。

(4)实体型(Entity Type)和实体值(Entity Value)。实体有类型和值之分。用于描述和抽象同一实体集共同特征的实体名及其属性名的集合称为实体型,如学生(学号,姓名,年龄,性别,籍贯,系)就是一个实体型。相应地,实体集中的某个实体的值即为实体值,如(“99011028”,“徐芳”,19,“女”,“江苏”,“会计系”)就是一个实体值。

(5)实体间的联系(Relationship Between Entity)。实体间的联系就是实体集与实体集之间的联系,这种联系共有以下三种:

- 一对一联系。如果对于实体集 E1 中的每个实体,在实体集 E2 中至多只有一个实体与之相对应,反之亦然,则称实体集 E1 与 E2 之间的联系是一对一联系(如图 1-3),记为“1:1”。例如,电影院中观众和座位之间就具有一对一的联系,因为在一个座位上最多坐一个观众,而一个观众也只能坐在一个座位上。

- 一对多联系。如果对于实体集 E1 中的每个实体,在实体集 E2 中有任意个(零个或多个)实体与之相对应,而对于 E2 中的每个实体却至多和 E1 中的一个实体相对应,则

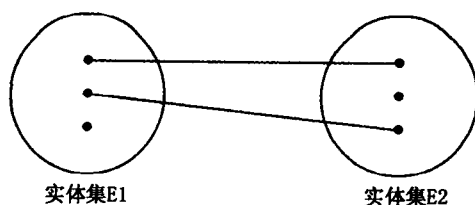


图 1-3

称实体集 E1 与 E2 之间的联系是一对多联系(如图 1-4),记为“1:n”。例如,公司的部门与其职工之间、学校的专业系与学生之间、球队与球员之间都具有一对多的联系。

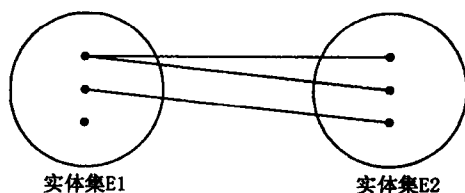


图 1-4

● 多对多联系。如果对于实体集 E1 中的每个实体,在实体集 E2 中有任意个(零个或多个)实体与之相对应,反之亦然,则称实体集 E1 与 E2 之间的联系是多对多联系(如图 1-5),记为“m:n”。例如,学校的学生与课程之间就具有多对多的联系,因为一个学生可以选修多门课,一门课程也可被多个学生选修。公司的产品与其客户之间也具有多对多联系,因为一个产品可以被多个客户订购,一个客户也可以订购多个产品。

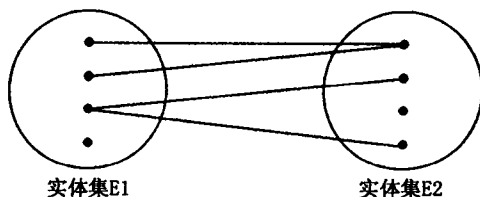


图 1-5

上面介绍了实体之间存在的三种类型的联系,需要说明的是联系可以存在于两个实体之间,也可存在于多个实体之间;不同实体集的实体间有联系,同一实体集的实体间也可以有联系。

2. 实体—联系方法

实体—联系方法是 P. P. S. Chen 于 1976 年提出的用于表示概念模型的方法,该方法

直接从现实世界抽象出实体型及其相互间的联系,并用实体—联系图(E—R图)来表示概念模型。E—R图又称为E—R模型,在E—R图中,实体型、属性及实体间的联系是这样来表示的:

(1)实体型:用标有实体名的矩形框表示。

(2)属性:用标有属性名的椭圆形框表示,并用一条直线与其对应的实体型相连接。

例如,一个客户实体及其属性(包括客户代号、公司名称、城市和电话)可以用图1-6来表示。

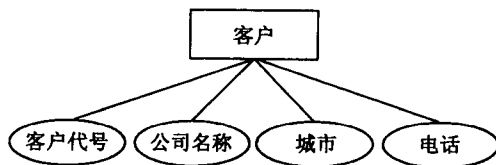


图 1-6

(3)实体间的联系:用标有联系名的菱形框表示,并用直线将联系与相应的实体型相连接,且在直线靠近实体型的那端标上1或n等,以表明联系的类型(1:1、1:n或m:n)。

例如,客户实体、产品实体及相互间的多对多联系可以用如图1-7所示的E—R图表示,其中产品实体具有产品代号、产品名称和单价等属性。由于客户在订购产品时必定会涉及到数量,所以数量可以作为客户和产品间联系(订货联系)的属性,该数量属性也用椭圆型表示,并用直线与订货联系相连接。

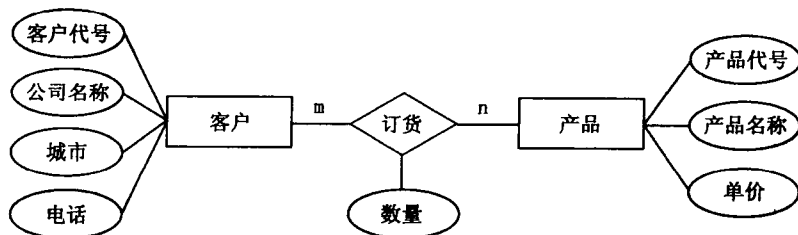


图 1-7

从上面的E—R图中我们可以得到的信息是:客户信息、产品信息以及什么客户订购了什么产品,数量为多少。但我们无法了解客户在每一次订货时所发出的订单的信息,如每张订单的代号是什么?何时订购的?该订单订购了什么产品?数量为多少?运费为多少?等等。如果所罗列的问题都是用户关心的问题的话,必须在E—R图中体现出来。因此,E—R图中可以增加“订单”实体型,其中包含订单代号、订购日期、运费等属性。另外,还要增加下述两个联系:

- 客户实体与订单实体之间的一对多联系：一个客户可以有多个订单，但一张订单只对应一个客户。

- 订单实体与产品实体之间的多对多联系：一张订单上可以有多种产品，一种产品也可以出现在不同的订单上。

这样，图 1-7 中的 E-R 图可以改为如图 1-8 所示的 E-R 图。

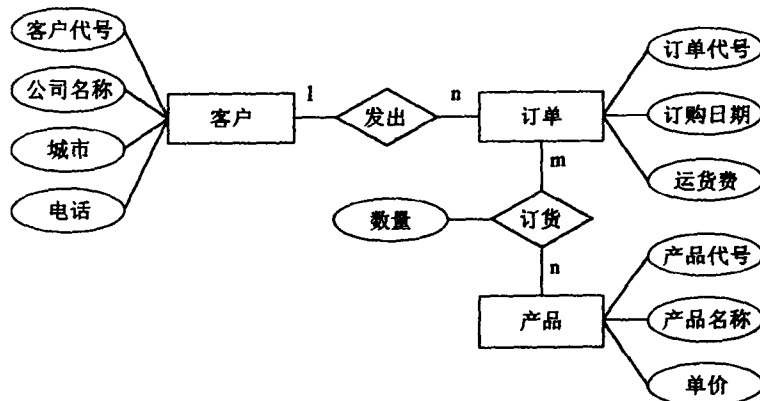


图 1-8

可见，在用 E-R 图描述概念模型时，必须根据具体的应用环境来决定图中包含哪些实体，实体间又包含哪些联系。

从上面 E-R 图也可以看出，用 E-R 图来描述概念模型非常接近人的思维，易被人理解，而且 E-R 图与具体的计算机系统无关，易被不具有计算机知识的最终用户接受。因此，E-R 图是数据库设计人员与用户进行交互的最有效工具。

二、数据模型

存放在数据库里的数据是某个企业、组织或部门的业务活动所涉及的各种数据，这些数据相互之间是有联系的，必须用一定的结构将其组织起来。在数据库中引入了数据模型来描述数据及它们之间的联系。针对不同的对象和应用目的可以采用不同的数据模型。常用的数据模型包括：层次模型 (Hierarchical Model)、网状模型 (Network Model)、关系模型 (Relational Model) 和面向对象模型 (Object Oriented Model) 等。

1. 层次模型

用树形结构来表示实体型及其联系的数据模型称为层次模型。采用层次模型的数据库系统称为层次型数据库系统。层次型数据库系统的典型代表是 1968 年由 IBM 公司推出的 IMS (Information Management System) 商用数据库管理系统。

在层次模型中，实体型称为记录型，实体的属性称为数据项或字段。实体型在层次模

型中是用结点表示的,实体的属性也在结点中列出;实体型间的联系用结点间的有向连线表示。图 1-9 就是某医院的一个层次模型。我们把位于上层的结点称为双亲结点,而把下层的结点称为子女结点。例如,图 1-9 中的“病房”结点就是“病人”结点的双亲结点,而“病人”结点是子女结点。在层次模型中,双亲结点与子女结点间的联系都是一对多(或一对一)的。

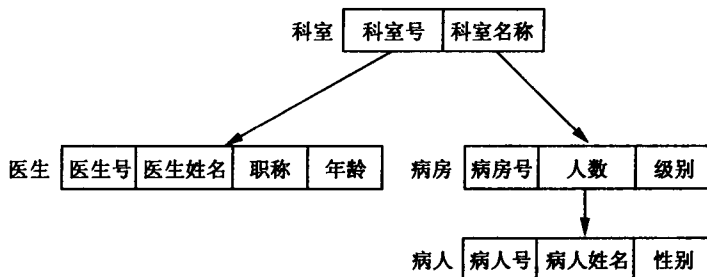


图 1-9

层次模型还必须满足如下两个条件:

- (1)有且只有一个结点没有双亲结点,这个结点称为根结点。
- (2)根以外的其他结点有且只有一个双亲结点。

2. 网状模型

用网状结构来表示实体型及其联系的数据模型称为网状模型。采用网状模型的数据数据库系统称为网状型数据库系统。网状型数据库系统的典型代表是 DBTG 系统,或称 CODASYL 系统。它是由数据系统语言研究会 CODASYL (Conference On Data System Language) 下属的数据库任务组 (Data Base Task Group) 于 20 世纪 70 年代提出的一个系统方案。该方案中提出的很多基本概念、方法和技术对网状数据库系统的研制和发展起着非常重要的指导作用。20 世纪 70 年代产生了很多成功的网状型 DBMS 产品,如 HP 公司的 IMAGE/3000、Univac 公司的 DMS1100、Honeywell 公司的 IDS/II 等。

与层次模型一样,网状模型中实体型也称为记录型,实体的属性也称为数据项或字段。记录型及其联系也是用结点及结点间的有向连线表示的。同样,网状模型中,双亲结点与子女结点间的联系也是一对多(或一对一)的。

与层次模型不同的是,网状模型去除了层次模型的两个限制,它允许:

- (1)一个以上的结点可以没有双亲结点。
- (2)一个结点可以有多个的双亲结点。

因此,层次模型实际上是网状模型的一个特例,网状模型比层次模型更具普遍性,更容易表示现实世界中事物间的复杂联系。图 1-8 的 E-R 图可以转换成如图 1-10 所示的网状模型。

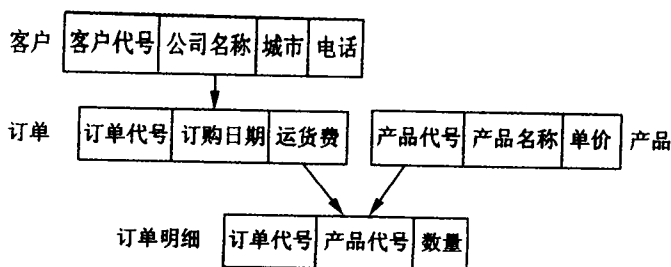


图 1-10

在图 1-10 的网状模型中,客户和产品结点都没有双亲结点,订单明细结点有两个双亲结点,所以该数据模型不属于层次模型,而属于网状模型。在该网状模型中,为表示原来的 E-R 图中订单和产品实体间的多对多联系,引入了另一个记录型“订单明细”,该记录型的属性是由订单记录型的码“订单代号”、产品记录型的码“产品代号”及联系的属性“数量”组成的。这样,订单与订单明细之间、产品与订单明细之间的联系都是一对多的了。

网状模型和层次模型统称为非关系模型。在非关系模型中,实体是用记录来实现的,记录之间的联系是用指针来实现的,因此查询效率较高;但是,应用程序在访问非关系模型的数据时,必须熟悉数据库的逻辑结构,选择合适的存取路径,这大大加重了程序员的负担。因此,从 20 世纪 80 年代中期开始,非关系数据库产品已逐步被关系数据库产品所取代。

3. 关系模型

用表结构来表示实体及实体间联系的模型称为关系模型。采用关系模型的数据库系统称为关系型数据库系统。

关系模型是 1970 年由 IBM 公司 San Jose 研究实验室的研究员 E. F. Codd 提出的,他连续发表了“大型共享数据库数据的关系模型”等一系列论文,提出了关系数据模型,奠定了关系数据库的理论基础。但是,关系数据库产品一直到 80 年代初才形成,此后,关系数据库系统得到了广泛的应用和普及。直到现在,关系数据库产品仍是大多数企业或组织信息化建设的首选产品。典型的关系数据库产品有 Access、Foxpro、DB2、Oracle、Sybase 和 Informix 等。

采用关系模型的数据库由一系列的“二维表”组成,每个表保存着企业业务活动中所涉及的一个特定实体(或者两个实体之间的某种联系)的所有实例的各种属性值数据。例如,表明各个客户向本公司订购产品的状况的“订单”就是一个实体,为了记录订单这一实体所有实例的订单代号、客户代号、订购日期和运货费等属性值数据,就可以在公司数据库中规定一个描述公司所有订单内容的“订单”表(如图 1-11 所示)。

图 1-11 订单表中的每一行数据代表着一份具体的订单。例如,订单表中的第一行

数据表明存在着一份订单,其订单代号为“O001”,订货客户的代号是“C001”,订购日期为1999年12月3日,运货费为20元。

在关系型数据库系统中,图1-11所示的一张二维表格被称为关系,即“订单”关系;表中每个列就是一个属性,列头是相应属性的属性名;表中的每一行称为一个元组,元组中的每一个属性值称为一个分量。

订单代号	客户代号	订购日期	运货费
O001	C001	1999-12-03	20
O002	C002	1999-12-10	50
O003	C001	2000-04-05	20
O004	C002	2000-04-20	20
O005	C003	1999-12-20	100
O006	C003	2000-06-09	50
O007	C004	2000-07-08	50
O008	C005	2000-10-08	100

图 1-11

对关系的描述称为关系模式,一般格式为:

关系名(属性名 1,属性名 2,...,属性名 n)

因此,上面的订单关系可以描述为:订单(订单代号,客户代号,订购日期,运货费)。

与层次模型和网状模型相比,关系模型具有如下几个特点:

- 关系模型概念单一、数据结构简单,实体与实体间的联系都是用关系(二维表)表示的。但是,在网状模型和层次模型中实体间的联系是由记录和记录之间所构成的网状结构和层次结构来表示的,数据结构复杂。

- 在关系模型中,数据操作是集合操作,即操作对象和操作结果都是关系。例如,为查询“上海”的客户,用户只要发出一条命令,系统就会把“上海”客户的信息显示给用户,因此查询结果就是一个“上海”客户的集合,该集合也是一个二维表格,即关系。但是在层次模型和网状模型中,数据操作方式是单记录的。

- 关系数据库语言是非过程化的。用户只要告诉系统“做什么”,而不需要知道“怎么做”。这样就将用户从导航式的检索中解脱了出来,大大降低了用户编程的难度,也易于关系数据库系统的推广和应用。

- 关系模型以关系代数为基础,形式化基础强。

4. 面向对象模型

虽然与层次模型和网状模型相比,关系数据模型有严格的数学基础,概念简单清晰,非过程化程度高,在传统的数据处理领域使用得非常广泛。但是,随着数据库技术的发展,出现了许多如CAD、图像处理等新的应用领域,甚至在传统的数据处理领域也出现了