

## 第 2 章 管理信息系统的技术基础

### 本章学习要点

- 了解信息技术的相关知识。
- 掌握计算机系统、计算机硬件系统、计算机软件系统的结构和特点。
- 了解计算机网络技术的基本概念。
- 掌握计算机网络的拓扑结构、分类及应用。
- 了解数据库技术的基本概念。
- 掌握数据库系统结构、关系数据库标准语言 SQL 及数据库的设计步骤。

管理信息系统是基于计算机的系统，同时也是基于网络的系统。本章主要介绍管理信息系统相关技术的基础知识，包括信息技术、计算机网络技术和数据库技术。

### 2.1 信息技术

一般地，信息技术是计算机硬件系统、软件系统和通信技术的总称。信息技术是管理信息系统的基础，只有把信息技术与管理结合起来，才能真正发挥管理信息系统的作用。

#### 2.1.1 计算机系统

计算机系统是现代管理信息系统的物质基础，可以说没有计算机系统就没有现代意义的管理信息系统，因此，有必要对计算机系统的基本知识有所了解。

计算机技术是实现现代管理信息系统的重要手段。世界上第一台电子计算机诞生于 1946 年。计算机以其处理速度快、运算精度高、有记忆和逻辑判断能力、数据传输速度快等特点而广泛应用于商业中的数据处理、科学技术中的数值计算和工业中的过程控制等领域，已经成为人类生活不可缺少的智能工具。自第一台电子计算机诞生以来，计算机科学与技术已成为 20 世纪发展最快的一门学科，尤其是微型计算机的出现和计算机网络的发展，使计算机的应用渗透到社会的各个领域，有力地推动了信息社会的发展。

##### 1. 计算机的诞生及发展

多年来，人们以计算机物理器件的变革作为标志，把计算机的发展划分为四代。

第一代（1946~1954 年）是电子管计算机，计算机使用的主要逻辑元件是电子管，也称电子管时代。主存储器先采用延迟线，后采用磁鼓磁芯，外存储器使用磁带。软件方面，用机器语言和汇编语言编写程序。这个时期计算机的特点是：体积庞大、运算速度低（一般每秒几千次到几万次）、成本高、可靠性差、内存容量小。这个时期的计算机主要用于科学计算，从事军事和科学研究方面的工作。

第二代（1954~1964 年）是晶体管计算机，这个时期的计算机使用的主要逻辑元件是晶体管，也称晶体管时代。主存储器采用磁芯，外存储器使用磁带和磁盘。软件方面，开始使用

管理程序，后期使用操作系统，并出现了 FORTRAN、COBOL、ALGOL 等一系列高级程序设计语言。这个时期计算机的应用扩展到数据处理、自动控制等方面。计算机的运行速度已提高到每秒几十万次，体积已大大减小，可靠性和内存容量也有较大的提高。其代表机型有：IBM 7090、IBM 7094、CDC 7600 等。

第三代（1965~1970 年）是集成电路计算机，这个时期的计算机用中小规模集成电路代替了分立元件，用半导体存储器代替了磁芯存储器，外存储器使用磁盘。软件方面，操作系统进一步完善，高级语言数量增多，出现了并行处理、多处理机、虚拟存储系统以及面向用户的应用软件。计算机的运行速度也提高到每秒几十万次到几百万次，可靠性和存储容量进一步提高，外部设备种类繁多，计算机和通信密切结合起来，广泛地应用到科学计算、数据处理、事务管理、工业控制等领域。其代表机器有 IBM 360 系列、富士通 F230 系列等。

第四代（1971 年以后）是大规模和超大规模集成电路计算机。这个时期的计算机的主要逻辑元件是大规模和超大规模集成电路，一般称大规模集成电路时代。主存储器采用半导体存储器，外存储器采用大容量的软、硬磁盘，并开始引入光盘。软件方面，操作系统不断发展和完善，同时发展了数据库管理系统、通信软件等。计算机的发展进入了以计算机网络为特征的时代。计算机的运行速度可达到每秒上千万次到亿万次，计算机的存储容量和可靠性又有了很大提高，功能更加完备。这个时期计算机的类型除小型、中型、大型机外，开始向巨型机和微型机（个人计算机）两方面发展，使计算机开始进入了办公室、学校和家庭。

目前，新一代计算机正处在设想和研制阶段。新一代计算机是把信息采集、存储处理、通信和人工智能结合在一起的计算机系统，也就是说，新一代计算机由处理数据信息为主转向处理知识信息为主，如获取、表达、存储及应用知识等，并有推理、联想和学习（如理解能力、适应能力、思维能力等）等人工智能方面的能力，能帮助人类开拓未知的领域和获取新的知识。

## 2. 计算机的工作特点

通常所说计算机是电子数字计算机的简称。电子数字计算机是一种快速、自动进行数值计算和信息处理的计算工具。其主要特点是：

（1）计算机能快速计算。计算机是一种可以快速计算的工具有，其运算速度的一种直观衡量标准是用每秒钟执行基本运算操作的次数来表示。现代计算机每秒的运算次数可以达到几十万次甚至几万亿次，可以完成过去人工无法完成的计算工作。如短期气象预报，人工计算需要数天甚至更长时间，而用计算机只需几分钟甚至更短的时间即可完成。我国最新研制的曙光 4000 计算机，每秒运算速度可达 3 万亿次。

（2）计算机具有通用性。由于计算机是把任何复杂的信息处理都分解为大量的基本算术操作和逻辑操作的组合，所以，计算机可以处理任何复杂的数学问题和逻辑问题，不仅对数值数据，而且可以对非数值数据，如图形、图像、文字和声音等数据进行处理。正因为如此，我们说计算机不是针对特定计算问题，而是适合各种计算问题的求解。

（3）计算机具有高精度性。计算机由程序控制其操作过程，它根据事先编制的程序自动、连续地工作，完成预定的计算任务。这可避免人工计算可能产生的诸如疲劳、粗心等所导致的各种错误，而且机器和算法的设计在理论上可以保证达到任意要求的精确度。

（4）计算机具有逻辑判断能力。逻辑运算与理解判断是计算机基本的功能，也是重要的功能。计算机的逻辑判断能力能实现计算机工作的自动化，并赋予计算机某些智能处理能力，从而奠定了计算机作为一种智能工具的基础。正由于计算机具有快速、准确、通用和逻辑判断

功能,决定了它能解决任何复杂的、大运算量的数学问题和逻辑问题。

### 3. 计算机的分类

按计算机的规模和性能,计算机可以分为六大类。

(1) 巨型机 (Supercomputer)。人们把价格昂贵、功能强大,计算速度在每秒千亿次以上的计算机叫做巨型机或超级计算机,世界上只有少数几个国家能够生产巨型机。2003年12月上海超级计算中心和曙光信息产业(北京)有限公司共同研制的曙光4000A,是目前国产最高性能的超级计算机,它由2000多个AMD 64位处理器构成,浮点运算速度达每秒10万亿次。据美国能源部劳伦斯·伯克利国家实验室2003年11月公布的排行,日本NEC公司的超级计算机“地球模拟器(Earth Simulator)”蝉联榜首,运算速度达每秒35.6万亿次,美国惠普为洛斯阿拉莫斯国家实验室制造的超级计算机(ASCI Q)以每秒13.8万亿次的运算速度列居第二。

(2) 小巨型机 (Minisupercomputer)。这是最新发展起来的一种价格相对便宜的小规模巨型机,也叫桌上型超级计算机。它们由并行的多个微处理器组成。目前这一机型发展迅速,直接对巨型机提出了挑战。

(3) 大型机和中型机 (Mainframe)。大型机和中型机的计算速度为每秒几亿到几百亿次,一般是作为大型计算中心的主机。我国许多部门装备的大型机有IBM4381、IBM ES-9000等。

(4) 小型机 (Minicomputer)。小型机的计算速度为每秒几千万到几亿次。它一般为中小型企事业单位使用,可满足部门性的要求,如DEC公司的VAX机、ALEFA机等。

(5) 个人机 (Personal Computer)。个人计算机也叫微型计算机或PC机,这是一种面向个人使用的计算机。微型机的计算速度差别较大,最新的高档微型机的速度可达每秒几亿次。它是目前最为普遍的计算机。

(6) 工作站 (Workstation)。这是介于高档微机和小型机之间的机型。它一般都配备有大屏幕显示器和大容量存储器,而且速度快,通信功能强,主要用于图像处理或计算机辅助设计等专业领域。

### 4. 计算机系统的组成

一个完整的计算机系统应包括硬件系统和软件系统两部分。

硬件系统就是以实物形式存在的计算机设备,包括运算器、控制器、内存储器、外存储器、输入/输出设备等。对硬件系统的要求是如何提高计算机的性能和安全性。

软件系统是指各种计算机程序运行时需要的数据及有关文档资料的总称。研究和开发软件的主要目的是如何管理好计算机,如何使用户更好地使用计算机,如何有效地维护计算机,如何更好地发挥计算机硬件资源的效能。

计算机系统的组成如图2-1所示。

#### 2.1.2 计算机硬件系统

硬件是计算机系统的物质基础,硬件系统是指组成计算机的所有电子的、电磁的、机械的、光的元件和装置。

计算机硬件系统由运算器、控制器、存储器和输入/输出设备等五个基本部分构成。其中运算器用于完成算术运算和逻辑运算;控制器用于控制计算机本身的各个部分,使之有条不紊地工作;存储器用于保存计算方法、原始数据、中间结果和最终结果;输入设备用于把原始数据、程序及操作命令输入计算机;输出设备用于输出计算机的处理结果和其他一些信息。计算

机系统的硬件结构如图 2-2 所示。

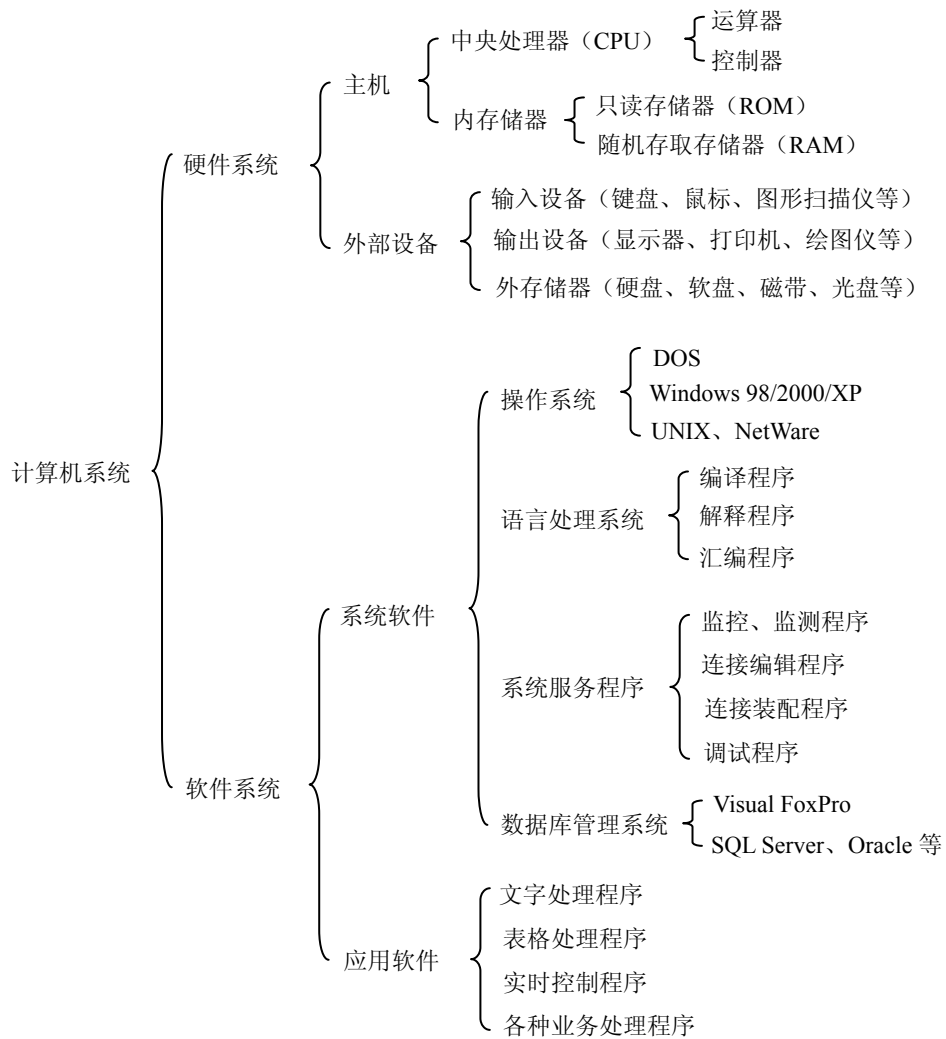


图 2-1 计算机系统的组成

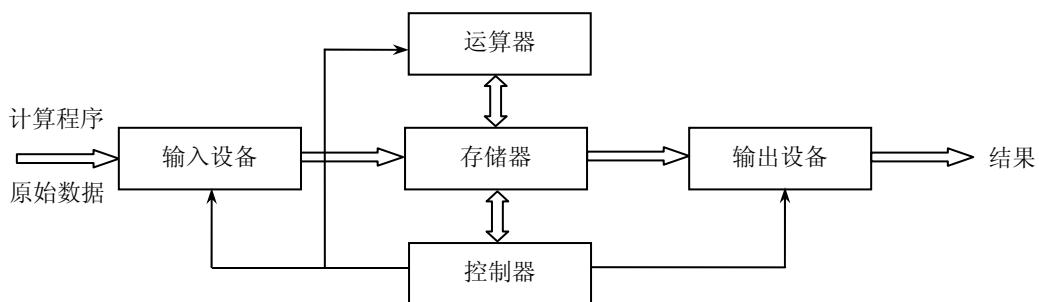


图 2-2 计算机系统的硬件结构

这样的硬件结构就是有名的冯·诺依曼体系结构。计算机的工作过程是：人们首先把操作指令和原始数据通过输入设备送入计算机的存储器；当计算开始时，指令被逐条送入控制器；

控制器向存储器和运算器发出存数取数命令和运算命令,经过运算器计算并把计算结果传送到存储器;在控制器的取数和输出命令作用下,通过输出设备输出计算结果。

### 1. 中央处理器 CPU

CPU 是微型计算机的核心,它由运算器、控制器和一些寄存器组成。运算器主要是执行算术运算和逻辑运算;控制器则规定了计算机执行指令的顺序,并根据指令的具体含义,控制计算机各部件协调地工作。通俗地讲,控制器就是计算机的指挥机关,它指挥计算机各部分的动作,完成计算机的各种操作。控制器本身也是按指令的要求来实现和指挥其他部件的。指令是由人编制并输入机器内,以保证计算机正确工作的命令。运算器在控制器的指示下执行各种操作,如取数、送数、相加、移位等,换句话说,运算器按照控制器发出的一系列命令来完成上述各种操作。寄存器可以暂存指令和数据。

### 2. 存储器

微型计算机中的存储器按用途和特性可以分为三类:主存储器、辅助存储器和高速缓冲存储器。

(1) 主存储器。微型计算机中主存储器(也叫内存)是 CPU 直接访问的存储器,可存储程序和原始数据,也可以存储计算结果、中间结果等。微型机的内存都是采用大规模集成电路制成的半导体芯片,这种存储器具有密度大、体积小、重量轻、存取速度较快、使用灵活等优点。内存芯片的存储容量指标是字 $\times$ 位,这里的“字”表示一个芯片内存的字数,“位”表示芯片上每个字的位数,如芯片容量为  $1K \times 8$ ,它表示该芯片上可存储  $1K$  字节,每个字节为 8 位。若用这种芯片组成  $2KB$  的存储器(2048 个字节),则需要两个这样的芯片并联组成。

主存储器一般可分为只读存储器(ROM)和随机存储器(RAM),前者中的数据只能被读取,后者中的数据可被读取和重写。

(2) 辅助存储器。辅助存储器也叫外存,其特点是容量大、价格低、速度慢。最常用的辅助存储器有软磁盘存储器、硬磁盘存储器以及光盘存储器等。

软磁盘存储器是一种磁表面介质存储器,它包括软盘驱动器和软盘。软盘由可弯曲的柔软塑料圆盘作基材,上面喷涂一层磁性材料而制成,具有质轻、不易损坏、便于携带的特点。软盘必须通过磁盘驱动器进行读写。常用的软盘是 3.5 英寸盘(简称 3 寸盘)。

硬磁盘存储器是计算机系统中重要的外部设备,它集磁盘驱动器和磁盘片为一体(简称硬盘)。其存储的方式与软盘相似。随着应用软件的系列化和大型化,硬盘的作用越来越重要,许多软件都要求有硬盘才能运行。硬盘的盘片是由铝合金制成,盘表面喷涂磁粉以记录信息。一个硬盘由多个盘片组成,其磁头和盘片封装在一个密闭的空间中,一般用户是不可以拆卸的,所以也叫固定磁盘。每个盘片与软盘类似,分为若干磁道(也叫柱面,即圆柱面),而每一个磁道又可分为若干个扇区,一般为 17 个扇区或更多一些,每个扇区又可以具体设定保存字节数,一般在 DOS 系统中每扇区设定为 512 字节的结构。这样对于一个具体的硬盘,如果知道了它的物理参数,即该硬盘有多少个磁头,盘面分为多少柱面,每一柱面有多少扇区及每一个扇区的字节数,就可以计算出该硬盘实际容量的大小。

光盘存储器是利用激光技术存储信息的计算机外部存储设备。它的基本工作原理是利用光的强弱来读写数据。光盘存储器按其记录方式一般分为如下两种:一是只读光盘存储器(CD-ROM),指只能读出光盘片中数据的存储器;二是可改写光盘存储器,指可对光盘片进行读和写操作的存储器。

光盘存储器与其他存储器相比,具有存储容量大(一张 12 厘米的光盘可存储大约 680MB

数据)、记录数据可保存时间长以及安全可靠等优点,但其存取数据的速度目前比硬盘慢。

(3) 高速缓冲存储器。因为 CPU 的速度要高于主存储器的存取速度,为了解决数据传输时的拥堵问题,在 CPU 和主存储器之间插入高速缓冲存储器,可以利用 Cache 技术将部分数据暂存在缓冲存储器中,以提高 CPU 的利用率。

### 3. 输入设备

输入设备负责将信息(数据和程序)通过人工键入或磁盘自动输入计算机,常见的输入设备是键盘、鼠标器、触摸屏等。

键盘是一种向计算机输入数据的基本输入设备,它通过电缆与主机相连,使用舒适方便。键盘内装有一块低档微处理器芯片,该芯片既可控制键盘工作,还能对键盘进行自检。如果击键较快而系统的 CPU 来不及处理,则键入的内容可以暂时保存在该芯片的缓冲区内(可保存 20 次击键内容)。该芯片的作用是把键盘上的按键动作转换成相应的 ASCII 码并输送给计算机主机。

鼠标器是一种小巧的人机交互式输入设备,由于其形似老鼠而得名。鼠标器通过专用电缆与计算机主机连接,当手持鼠标器在平板上滑动时,鼠标器自身的机构可测出滑动的方向和距离,配合其软件可控制屏幕上的光标移动。它是便利、灵活的人机对话和交互式作图工具,也是图形操作系统(Windows)必不可少的操作设备。根据鼠标的工作原理可分机械式鼠标器和光电式鼠标器。机械式鼠标器由机械装置在平板上滑动来定位;光电式鼠标则由光线感应于平板上的滑动来定位。

### 4. 输出设备

输出设备是负责将计算机加工处理后的结果输出,以供人们查阅和使用的设备。常见的输出设备有显示器、打印机等。

显示器是显示字符和图形的设备。显示器屏幕上的一个画面称为一帧;每帧有若干条扫描线,如 200 线、512 线等;每线又分为若干个点,形成许多像素。若每个像素用一位二进制数存储信息,则只能表示该像素是黑还是白;当需要表示每个像素的亮度等级(即灰度)或彩色时,就得用几位二进制数的存储信息来描述一个像素。如用 3 位二进制数,就可以表示 8 个灰度等级或 8 种不同颜色。如果需要 256 种颜色,则每个像素至少需要 8 位(一个字节)二进制数表示。

显示器有单色显示器和彩色显示器两种。单色显示器又分字符显示器和图形显示器,字符显示器每屏可显示 80 列×25 行字符,每个字符块的大小为 9×14 点,字符块中的字符由 7×9 点组成,能显示(或处理)8 位的 256 种不同编码的字符输出。彩色显示器经过多年来的发展已经形成了系列化产品,其发展趋势是高分辨率、多色彩的新型显示器。

显示适配器是驱动显示器工作的硬件插卡,它的主要功能是使计算机主机与显示器相互协调。显示适配器分为如下几种:

(1) CGA: 彩色图形适配器(Color Graphic Adapter),是最早的彩色显示器适配器,适用于低分辨率的彩色图形显示器。它支持 4 种颜色的彩色显示,其图形分辨率为 320×200,就是说水平方向有 320 像素,垂直方向有 200 像素。

(2) EGA: 增强图形适配器(Enhance Graphic Adapter),是改进型的彩色图形适配器,适用于中分辨率的彩色图形显示器。它支持 16 种颜色的彩色显示,其图形分辨率为 640×350。

(3) VGA: 视频图形阵列(Video Graphic Array),是 IBM 引入的一种新的图形标准,适用于高分辨率的彩色图形显示器。它支持 256 种颜色的彩色显示,其图形分辨率为 640×480。

此外还有超级视频图形阵列 (SVGA)、扩展图形阵列适配卡 (XGA) 等。

计算机常用的打印机有针式打印机 (或称点阵式打印机)、激光打印机和喷墨式打印机三类。目前喷墨式打印机使用最广泛。

(1) 激光打印机 (Laser Printer): 是一种高质量的非击打式印字输出打印机。激光打印机采用了激光和电子照相技术, 其主要优点是噪声低、质量高。

(2) 喷墨打印机 (Ink Vapor Printer): 是一种可以使用普通打印纸的非击打式打印机, 具有低成本、低噪声的优点。这种打印机是将墨水通过喷墨管喷射到普通打印纸上实现印字, 喷墨打印机还具有印字质量好等优点。如果把红、绿、蓝三基色的墨水喷头安装在一台喷墨打印机上, 便可以实现彩色图像的打印。

(3) 针式打印机 (Dot Matrix Stylus Printer): 是利用钢针撞击色带和记录纸, 从而用点阵组成字符图形的一种打印机。其特点是运行可靠、操作简单、耗材价格低廉, 但速度慢、噪声大。按打印头钢针的多少, 分为 9 针打印机和 24 针打印机。

汉字打印有两种方式: 一种方式是打印机直接接收汉字图形 (字形), 按图形方式打印, 如 EPSON LQ100 打印机; 另一种是打印机本身存储有汉字点阵字库, 计算机只需传递汉字内码到打印机, 由打印机转换为汉字字型打印, 如 EPSON LQ1600K 打印机。第一种方法的打印机简单、造价低、打印速度慢, 第二种方法的打印机需配置汉字库、成本高、打印速度快。

### 2.1.3 计算机软件系统

计算机软件是计算机程序、程序所使用的数据以及有关的文档资料的集合。软件的作用在于确定计算机做什么以及如何做。从用户的角度来看, 软件是用户与硬件之间的使用界面。为了让计算机完成某种处理或计算, 用户必须编制软件或程序来指定硬件所应采取的动作或算法。软件的发展使计算机的应用领域变得非常宽广, 也使用户能够更方便、更有效地使用计算机硬件资源。

根据计算机软件的总体结构和表现形式, 软件一般可分为系统软件和应用软件两大类, 它们之间的关系如图 2-1 所示。

系统软件是直接控制和协调计算机、通信设备及其他外部设备的软件。这类软件一般紧靠硬件, 是用户与计算机之间的第一层界面。它们与具体应用无关, 只是在系统一级提供服务。操作系统就是典型的系统软件。

系统软件中还包括语言处理程序 (如把汇编语言转换为机器语言的汇编程序, 把高级语言转换为机器语言的编译程序或解释程序)、作为软件研制开发工具的编辑程序、调试程序、装配和链接程序、测试程序, 以及适应事务处理需要的数据库管理程序等。

应用软件是直接完成某种具体应用的软件。这类软件的范围最广, 产品最多, 从简单的工资管理程序到某一单位的管理信息系统, 都是应用软件的范畴。应用软件一般是供最终用户使用的, 如图形软件、文字处理软件、财会软件、计划报表软件、辅助设计软件、程序开发软件及模拟仿真软件等。

#### 1. 操作系统 (Operation System, 简称 OS)

操作系统是控制和管理计算机硬、软件资源, 合理组织计算机工作, 并为用户使用计算机提供服务的软件。计算机系统中各部件之间相互配合、协调一致的工作是靠操作系统的统一控制和管理才能得以实现的, 同时它又是用户与计算机硬件之间的重要界面。任何一个用户都不可能直接操纵计算机硬件, 而是通过操作系统来使用计算机。

微型计算机的操作系统以前是使用命令执行方式的 DOS，目前使用的基本上都是微软公司生产的图形界面的 Windows。操作系统因用户个数、处理方式、所处环境的不同有各种不同的特点。通常有如下类型：

(1) 单用户操作系统。单用户操作系统指支持一个用户作业的操作系统，这个用户独占计算机系统的全部硬、软件资源，一般作为微型计算机操作系统。单用户操作系统还分为单任务和多任务两类，现在单用户操作系统几乎都具有多任务的处理功能。

(2) 分时操作系统。即多用户操作系统，它是一种允许多个联机用户同时使用一台计算机（通常称作主机）的操作系统，主机的 CPU 按照预先分配给各个终端的时间片轮流为各个用户服务，即各个联机用户终端在各自的时间片内占有 CPU，分时共享计算机系统的资源。分时系统具有交互会话的功能，可以随时进行人机对话。当用户不是太多时，用户感受不到分时，每个用户就像是自己独占整台计算机一样。常见的分时操作系统有 UNIX 等。

(3) 实时操作系统。实时操作系统是一种时间性强、反应迅速快的操作系统，是对外部输入的信息在规定的时间内能处理完并输出结果的系统。对它的要求一个是实时性，一个是可靠性。此类系统多用于实时控制和自动控制系统中，如订票系统，当接到各售票窗口发来的订票信息后，能即时检索出是否还有旅客要订的票，并给出回答。

(4) 网络操作系统。前面介绍的三种操作系统，都是控制和管理一台计算机工作的，而网络操作系统是用来管理连接在计算机网络上的多个计算机的操作系统。它是比单机操作系统更高一级的系统软件。

## 2. 计算机程序设计语言与开发工具

程序设计语言用来书写计算机可以执行的程序。由于计算机不能识别人类的自然语言(只能接受和处理二进制代码所表示的数据)，所以为了实现对计算机的有效控制，人类发明了各种计算机程序设计语言来编制程序。常用的计算机程序设计语言如下：

(1) 机器语言。机器语言是直接用二进制代码指令表达的计算机编程语言，它可被计算机直接识别和执行。由于每条语句都是一条二进制形式的指令代码，故执行速度快、效率高。但是人阅读困难，编程烦琐，容易出错，而且不能移植。

(2) 汇编语言。汇编语言用能够反映指令功能的助记符表示二进制形式的指令代码，它是符号化的机器语言。汇编语言程序虽然比机器语言程序直观，也比较容易编写和修改，但是汇编语言程序在计算机上不能直接运行，必须用特殊的翻译程序把它翻译成机器语言后，计算机才能够执行。这个翻译过程就叫做汇编，原来的程序叫做源程序，汇编后的程序叫做目标程序。

(3) 高级语言。高级语言的描述方法接近于人们的自然语言和数学语言，是自然语言化的计算机程序设计语言。目前计算机高级语言有许多种，如 FORTRAN、PASCAL、C、PROLOG、BASIC 等。高级语言直观方便，但是高级语言也不能在计算机中直接运行，必须翻译成机器语言后计算机才能够识别。高级语言一般有两种翻译方法：一种是编译，另一种是解释。编译是把源程序先翻译成目标程序，以后再执行目标程序；而解释是边翻译，边执行，逐句完成。

(4) 第四代语言(4GL)。第四代语言是为降低程序开发工作难度和提高程序开发效率而设计的通用语言。用第四代语言编写程序时，往往只要用类似于自然语言的交互方式描述用户的信息处理需求，编写出来的程序是非过程化的，即程序指令只要告诉计算机需要“做什么”，而不必详述“怎么做”的具体细节，这使得非计算机专业的用户无须借助技术人员的力量也能够自行开发所需的应用软件。第四代语言开发工具往往包括一些可以直接为最终用户使用的软件包，为用户提供一个功能强大且方便使用的软件开发环境。如某些数据库系统的查询语言和



一些应用软件包的宏语言就具有第四代语言的特征。由于第四代语言易学易用、开发速度快、效率高、便于维护等特点，它将成为管理信息系统开发的一个重要途径。

(5) 面向对象的程序设计语言。这是 20 世纪 80 年代以来新发展的程序设计语言，它不同于以往的高级语言将数据与对数据的操作相分离，而是将它们合成为对象。对象包含数据和对数据的操作，这样的对象可以重用，从而大大提高编程效率。目前应用最为广泛的面向对象语言是 C++ 和 Java，在管理信息系统的开发中也常有应用。

(6) 标记语言。由于 Internet 的风靡全球，标记语言也开始引起人们的注意，其中超文本标记语言 HTML (Hyper Text Markup Language) 应用最为广泛，它已经成为 Web 的通用语言；几乎所有的 Web 页面都是用 HTML 编写的。HTML 简单易学、简明紧凑，能够对文字、图表以及图像、声音、动画等多媒体数据进行统一处理。但是 HTML 中表示文件格式的标签集是固定的，在处理许多需要专门格式的文件（如数学公式和化学分子式等）时显得无能为力。

这就促使人们进行进一步的研究，从而产生了可扩展标记语言 XML (Extensible Markup Language)。XML 侧重于数据本身，它的标签集不是固定的，用户可以根据自己的需要定义任何一种标签来描述自己文档中的数据元素，它的出现使网上的信息查询、数据交换更加便利，有助于人们更加有效地利用网络。

(7) 开发工具。随着人们对信息处理能力的要求越来越高，以及以计算机为代表的现代信息技术迅速地应用到社会生活的各个角落，社会对各种软件的需求也日益紧迫。为了满足这种需求，各个软件厂商都在争先恐后地开发新的软件开发工具。软件开发工具以计算机自身处理信息的强大能力为基础，在软件开发的各个阶段，对软件开发的各个方面提供各种帮助，是软件工作人员必须具备的重要技术手段。软件开发工具的概念大致可以描述为：在高级程序设计语言（第三代语言）的基础上，为提高软件开发的质量和效率，从规划、分析、设计、测试、成文和管理各方面，对软件开发者提供各种不同程度的帮助的一类广泛的软件。

对于软件开发工具这个概念可以从以下几个方面来理解：它是在高级程序设计语言（第三代语言）之后，软件技术进一步发展的产物；它的目的在于在开发软件过程中给予人们各种不同方面、不同程度的支持或帮助；它支持软件开发的全过程，而不是仅限于编码或其他的特定工作阶段。

目前，在市场上就有很多开发工具以满足不同的开发要求。除了功能较全的集成化开发工具 PowerBuilder、Visual Basic、Delphi、InterDev 等，还涌现出了一批专用开发工具，比如网页制作工具 Dreamweaver、Fireworks、FrontPage、HomePage 等；多媒体制作工具 Authorware 等；基于网页的开发工具 ASP (Active Server Pages)、JSP (Java Server Pages)、PHP (Personal Home Page) 以及计算机辅助软件工程 (Computer Aided Software Engineering, 简称 CASE) 工具 Rational Rose 等。

### 3. 数据库管理系统

数据库管理系统是专门对数据记录进行综合管理的软件，对数据文件结构的定义、数据记录的更新、数据记录的查询以及对数据记录的各种运算提供全面的支持。许多数据库系统都提供了功能强大的数据库操作语言和应用开发工具，可直接用来开发信息处理和数据管理应用系统。数据库管理系统是信息系统的主要技术基础，本书将在下一节作专门介绍。

### 4. 应用软件

应用软件是直接面向最终用户的具体应用的软件。应用软件以操作系统为基础，用程序设计语言编写，或用数据库管理系统构造，用于满足用户对计算机应用的各种具体要求。由于计

算机应用领域十分广泛，因此应用软件的种类很多，但总体来看，应用软件主要分为两大类。

(1) 通用应用软件(包)。它是指提供某些通用信息处理功能的商品化软件。这种软件对某类应用具有通用性，因此可以被许多具有此类应用需求的用户所使用。通用应用软件一般具有较强的通用性，所提供的功能或处理往往可以选择、设置和调配，便于用户可以灵活地配置软件以满足其特定的需求。例如文字处理软件、表格处理软件、数值统计分析软件、财务核算软件、人事档案管理软件等都是通用应用软件。通用应用软件一般是由计算机软件开发商开发的商品化软件，用户购买这类软件后，要经过一定的配置过程才能满足用户的特定要求。某些大型、复杂的通用应用软件要由专业技术人员进行安装、配置和调试，而大多数通用应用软件，特别是微型计算机的应用软件，其安装和调配往往是由最终用户经过简单的培训就可以进行的，他们只要按照软件说明书的要求输入所需的基础数据就可以直接应用于具体业务了。例如，用户购买了财务核算软件包后，只要先用软件中的建账功能设置各级各类科目，建立起账户，然后就可以输入记账凭证，进行电算化的财务管理了。

(2) 专用应用软件。有时也被称为用户定制软件。在许多应用场合中，用户对数据处理的功能需求具有很大特殊性，通用软件不能满足。在这种情况下，就要由专业人员采取管理信息系统开发的方法与技术，为用户单独开发具有特定功能的专用应用软件。

通用软件大多是面向广大用户的商品化软件，而专用软件往往是定点开发，用户数和应用范围比较小，但通用应用软件与专用应用软件的区别并不是绝对的，只是通用性与专用性的程度不同而已。通用软件也要通过配置过程实现用户的具体功能需求，同时还提供一些特殊的专用功能以扩大其应用范围；而专用软件往往也有许多通用性考虑，以适应应用环境的调整与变化，并且可能随着应用对象的增加和范围的扩大而发展成为通用软件。

## 2.2 计算机网络技术

### 2.2.1 计算机网络概述

随着时代的发展，人们对更复杂的信息处理手段的需求增长得很快，而计算机技术和通信技术的发展与融合，为解决这些需求提供了可能，计算机和通信技术结合的产物便是计算机网络。

从广义上说，计算机网络是地理上分散的多台独立自主的计算机遵循约定的通信协议，它通过软、硬件互联以实现交互通信、信息交换以及资源共享的目的。它出现在 20 世纪 50 年代，至今虽然发展时间不长，但是发展很快，经历了从简单到复杂、从单个局域网通信到全球网络互联的发展过程。如今，计算机网络已成为人们社会生活、工作、经济、贸易等各个方面不可缺少的重要组成部分。从应用的角度来看，计算机网络主要有如下几个方面的作用：

(1) 远程的资源共享。许多机构都有一定的计算机在运行，这些计算机的地理范围可能相距甚远，最初每台计算机都独立工作，但是管理部门可能需要将这些计算机连接起来，以获取和核对整个公司的信息。通过计算机网络，可以让网络上的用户，无论他处在何方，也无论资源的物理位置，都能够在计算机间进行文件传送，使用远程计算机上的数据以及运行或设置程序等，从而取得很好的经济效益和社会效益。

(2) 网络间的通信和合作。现代社会信息量激增，信息的交换也日益增多，每年有几百万吨信件要快递。利用计算机网络传递信件是一种全新的电子传递方式，它与现在的通信工具

相比有很多优点,它不像电话,对话者需直接同时连接;也不像广播系统只是单向地传递信息;更不像信件,它几乎是即时地通过网络传递到接收者。电子邮件在办公室自动化以及提高人们的工作效率上有着十分重要的作用。

(3) 提高计算机的可靠性和可用性。提高可靠性表现在计算机网络中各台计算机可以通过网络彼此互为后备机,一旦某台计算机出现故障,那么其任务可以通过其他计算机代为处理,避免了单机在无后备机使用的情况下由于计算机故障导致的系统瘫痪的现象,这样就大大地提高了可靠性。

提高可用性在于当网络中某台计算机负载过重时,网络可将新的任务转交给网中较空闲的计算机完成,这样就可以均衡各台计算机的负载,提高了每一台计算机的可用性。例如,在军事、银行、航空、交通管制、核反应堆安全设备和其他许多的应用中,提高可靠性和可用性是极其重要的。

(4) 易于分布处理。在计算机网络中,各用户可根据情况合理选择网内资源,以就近、快速地处理。对于较大型的综合问题,通过一定的算法将任务分交给网络中的计算机,达到分布处理的目的。这样就可以利用网络技术将多台计算机连成具有高性能的计算机系统,以共同完成对一个复杂问题的处理,这就是当今称之为协同式计算的一种网络计算方式。

## 2.2.2 网络的拓扑结构与分类

### 1. 计算机网络的拓扑结构

为进一步分析网络单元彼此互联的形状与其性能的关系,采用拓扑学中一种研究与大小形状无关的点、线特性的方法,把网络单元定义为节点,两个节点间的连线称为链路。这样,从拓扑学观点看,计算机网络则是由节点和链路组成的。

网络节点和链路的几何位置就是网络的拓扑结构。网络中共有两类节点:转接节点和访问节点。节点计算机、交换机和终端控制器等属转接节点,它们在网络中只是转接和交换传送的信息;主计算机和终端等是访问节点,它们是信息交换的源节点和目标节点。

通信子网的拓扑类型较多,主要有以下六种:星型、树型、环型、总线结构、全部互联和不规则型,如图 2-3 所示。

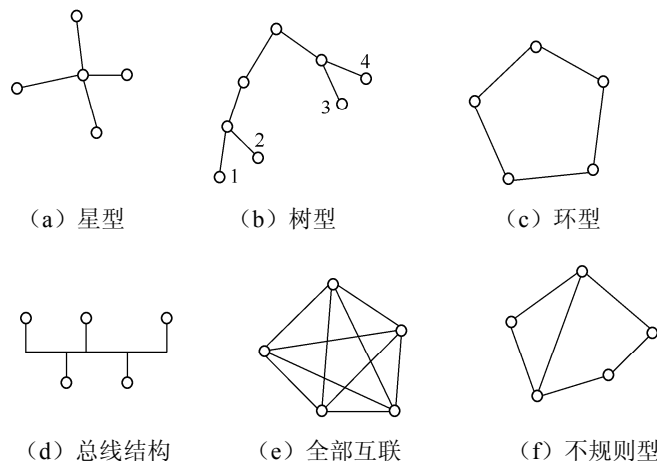


图 2-3 网络的拓扑结构

(1) 星型。图 2-3 (a) 所示是星型结构。星型的中心节点是主节点，它接收各分散节点的信息再转发给相应节点，具有中继交换和数据处理功能。星型网的结构简单，建网容易，但可靠性差，中心节点是网络的瓶颈，一旦出现故障则全网瘫痪。

(2) 树型。树型网络是分层结构，适用于分级管理和控制系统中。与星型结构相比，由于通信线路总长度较短，故它的成本低，易推广，但结构较星型复杂。网络中除叶节点（图 2-3 (b) 中标有数字 1、2、3 和 4 的各节点）及其连线外，任一节点或连线的故障均影响其在支路网络的正常工作。

(3) 环型。网络中节点计算机连成环形，就成为环型网络，如图 2-3 (c) 所示。环路上，信息单向从一个节点传输到另一个节点，传输路径固定，没有路径选择问题。环型网络实现简单，适用于传输信息量不大的场合。由于信息从源节点到目的节点都要经过环路中的每个节点，故任何节点的故障均导致环路不能正常工作，可靠性较差。由于环型网络具有结构简单、容易实现、无路径选择和建网投资少等优点，使它在多机系统和局部计算机网络中占有重要地位，是使用较多的结构。

(4) 总线结构。如图 2-3 (d) 所示，总线结构中，各节点通过一个或多个通信线路与公共总线连接。总线型结构简单、扩展容易。网络中任何节点的故障都不会造成全网的故障，故相对上述几种结构可靠性高。

(5) 全部互联。如图 2-3 (e) 所示，网络中任意两节点间都有直接通路相连，故通信速度快，可靠性高；但建网投资大，灵活性差，适用于对可靠性有特殊要求的场合。

(6) 不规则型。网络中各节点的连接没有一定的规则，一般当节点地理分散，而通信线路是设计中主要考虑因素时，采用不规则型网络，其结构如图 2-3 (f) 所示。

上述结构中，星型和树型均属集中控制方式，它们的主要缺点是可靠性差，主节点的故障会导致全网瘫痪。环型和总线型主要使用分布式控制方式，在局域网络中多被采用。不规则型网络则主要用在远程网络中。如何确定网络的拓扑结构，这是网络设计中首要考虑的问题，需根据应用场合、任务要求和费用等诸因素综合分析比较后确定。

## 2. 计算机网络的分类

计算机网络可以按照不同的方法进行分类。

(1) 按地域范围划分，主要有局域网、城域网和广域网三类。局域网是指传输距离在 0.1~10km，传送速率在 1Mbps~10Mbps 的范围较小的一种网络。局域网是计算机网络发展最快的一个分支，经过 20 世纪 60 年代的技术准备、70 年代的技术开发和 80 年代的商品化阶段，现在已在企事业单位的计算机应用中发挥着重要作用，目前正朝着多平台、多协议、异机种方向发展，数据速率和带宽也在不断提高；城域网则是在局域网上的扩展，它的区域范围相比局域网要大，通常指的是一个城市范围内的区域；广域网从广义上讲是指将远距离的网络和资源连接起来的任何系统，它一般有相距较远的局域网通过公共电信网络互联，提供跨国或全球范围的联系。

(2) 按拓扑结构划分，可分为总线型、星型、环状、网状网等。

(3) 按交换方式划分，可分为电路交换网、分组交换网、信元交换网、帧中继交换网等。

(4) 按网络协议划分，可分为 TCP/IP、SNA、SPA/IPX、AppleTalk 等。

(5) 按网络数据传输和系统的拥有者划分，可分为专用网和公用网。公用网由电信部门组建，一般由国家和政府部门控制和管理；专用网则由某部门或公司组建。

### 2.2.3 通信传输介质与网络设备

#### 1. 通信传输介质

通信传输介质是通信系统中发送端与接收端之间的信道通路，主要有以下几种。

(1) 双绞线。双绞线是指按一定规则螺旋缠绕在一起的两根绝缘铜线，它是最传统、应用最普遍的传输介质，如电话线。两条线绞扭在一起的目的是为了减少导线之间的电磁干扰。双绞线的线路损失大，传输速率低，并且抗干扰能力较弱，但由于其价格便宜，易于安装实现结构化布线，传输数字信号的距离可达几百米，因此在局域网中应用得很普遍。

(2) 同轴电缆。同轴电缆由内外两条导线构成，内导线是单股粗铜线或多股细铜线，外导线是一条网状空心圆柱导体，内外导线之间隔有一层绝缘材料，最外层是保护性塑料外皮。同轴电缆可以在较宽的频率范围内工作，抗干扰能力强，传输距离可达几公里，在计算机网络中被广泛采用。

(3) 光导纤维（光纤）。光导纤维是由高折射率的细玻璃或塑料纤维外包低折射率的外壳构成。其基本工作原理是：在发送端通过发光二极管将电脉冲信号转换成光脉冲信号，在光纤中以全反射的方式传输，在接收端通过光电二极管将光脉冲信号转换还原成电脉冲信号。

由于光波的频率范围很宽，所以光纤具有很宽的频带，光可以在光纤中进行几乎无损耗的传播，因此可以实现远距离高速数据传输；此外，由于是非电磁传输，无辐射，光纤的抗干扰能力强，保密性好，误码率低。但光纤传输系统价格较贵（尽管光纤本身并不贵，但光端设备复杂），因此一般用作网络通信的主干线。

(4) 微波。微波是利用高频无线电波在空气中的传播进行通信，发送站将数据信号载波到高频微波信号上定向发射，接收站将信号下载进行接收处理或转发。微波是直线传播的，具有高度的方向性，因此传输距离要受到地球表面曲率所造成的视线距离的限制。如果传输超过一定距离（最长不能超过 50 公里），就要通过中继站进行接力传输。

微波传输频带较宽，成本比同轴电缆和光纤低，但误码率高。微波传输安装迅速、见效快，易于实现，是在不能铺设线路条件下的远程传输、移动网络通信等场合中最经济、便利的通信手段。

(5) 卫星通信。卫星通信是利用地球同步卫星做微波中继站进行远距离传输。地球同步卫星位于地面上方 36000 公里的高空，其发射角度可以覆盖地球的 1/3 地区，三颗同步卫星就可以覆盖整个地球表面。通过同步卫星上的转发设备，将来自地面的微波信号发送给所覆盖的区域并转发给其他同步卫星。因此传输距离不受视线距离的限制，可以发送给全球任何一个区域。卫星通信传输的突出特点是，具有一发多收的传输功能，覆盖面积大，传输距离远，并且传输成本不随传输距离的增加而提高，特别适合于广域网络远程互联。但卫星通信成本高，传播延迟较长，并且存在安全保密等方面的问题。

#### 2. 网络设备

(1) 局域网的主要通信设备。局域网的主要通信设备有网卡、中继器、集成器、路由器等。

1) 网卡（NIC, Net Interface Card）。网卡是将各计算机连接成网络的接口部件，是网络接口卡，又称网络适配器。使用时插在服务器和客户机的主板扩展槽中，通信线路通过它与计算机相连接。网卡的功能是将计算机数据转换成能在通信介质中传输的信号。

2) 服务器（Server）。服务器本身是一台功能强大的计算机。它可由高档微机、工作站或专门设计的计算机（专用服务器）充当。

服务器的功能是为网络上的其他计算机提供信息资源。按照服务器在网络中所起的作用又可分为文件服务器、打印服务器、通信服务器等。文件服务器可提供大容量磁盘存储空间,为网上各微机用户共享;打印服务器负责接收来自客户机的打印任务,管理安排打印队列和控制打印机的打印输出;通信服务器负责网络中各客户机对主计算机的联系,以及网与网之间的通信等。各类服务器主要是提供各种网络上的服务,并实施网络的各种管理,一般不作用户的操作使用。

3) 客户机 (Client)。客户机是网络中的用户使用的计算机。它可使用服务器所提供的各类服务,从而提高单机的功能。它是用户直接操作的网络的一端。

4) 传输介质。传输介质也称为通信介质。网络中的数据信息可通过两类传输介质来传输。一类是有线传输介质,在局域网中常用的有线传输介质有双绞线、同轴电缆和光缆。另一类是无线传输介质,即微波、红外线、激光等。从网络的发展趋势来看,网络上使用的传输介质将会逐渐由有线介质向无线介质的方向发展。

5) 中继器 (Repeater)。中继器用于同一网络中两个相同网络段的连接。它对传输中的数字信号进行再生放大,用以扩展局域网中连接设备的传输距离。

6) 集线器 (Hub)。集线器用于局域网内部多个工作站与服务器之间的连接,可以提供多个微机连接端口。在工作站集中的地方使用集线器,便于网络布线,也便于故障的定位与排除。集线器还具有再生放大和管理多路通信的能力。

(2) 广域网的常用设备。广域网中的常用设备有调制解调器、网桥、路由器、网关等。这些设备的概况介绍如下:

1) 调制解调器 (Modem)。若计算机要利用电话线联网,可使用调制解调器连接。其功能是将计算机输出的数字信息转换成可在远程通信线路上传输的模拟信号,并能将从通信线路上接收到的模拟信号转换成数字信息传送给计算机。

2) 网桥 (Bridge)。网桥是用于较低层的网络连接设备,主要用来连接两个同类网络,也可用来连接不同协议网络或不同拓扑结构的网络。

网桥可将一个网的帧格式转换为另一个网的帧格式而进入另一网中。网桥的操作在网络数据链路层进行。网桥可以将大范围的网络分成几个相互独立的网段,使得某一网段的传输效率提高。各网段之间还可以通过网桥进行通信和访问。

3) 路由器 (Router)。路由器是属于比网桥高一层的网络间连接设备。它不仅具有网桥的功能,还具有路由寻址、数据转发及数据过滤的功能。它能在复杂的网络互联环境中建立非常灵活的连接。

路由器工作在网络层,它在接收到数据链路层的数据包时,都要拆包,查看网络层的 IP 地址,确定数据包的路由,然后再对数据链路层的信息打包,最后将该数据包转发。路由器实际是一台计算机,它的处理功能程序固化在硬件中,以提高对数据包的处理速度。也可用一台带有两个以上网卡的普通计算机配上相应的软件来实现路由器的功能。

4) 网关 (Gateway)。网关是与规模更大的不同操作系统的网络进行连接时使用的设备。网关不仅具有路由器的全部功能,还能进行不同协议的转换。

利用网关可将局域网连接到广域网上(如连到 Internet 上),从而使用户省去与大型计算机连接的接口设备和电缆,却能共享大型计算机的资源。

#### 2.2.4 典型的计算机网络应用

随着互联网的发展,网络已开始使人们能跨越时间和空间的限制,成为人们随时随地获取

信息的重要手段。随之,计算机网络的应用也有了进一步的发展。网络从传输单一的正文数据,发展为传输语音和视频数据,以及集成声、文、图、像的多媒体传输;从传统的点到点交互,发展到单点到多点交互,实现如视频广播和远程教育的应用,进而发展到多点到多点间的交互应用,如计算机视频会议系统;从集中控制结构发展到分布控制结构,并进一步发展到联邦(具有不同管理域)控制结构;从传统的没有服务质量要求的服务模式,发展到具有服务质量保证的服务模式——高性能网络服务。

高传输速率、高带宽、高服务质量(传输正确性、传输安全性、可靠性)是高性能计算机网络的重要指标,也是21世纪计算机网络发展的主要方向。为此,网络的发展,将会要求和带动如下网络技术和应用。

#### 1. 综合业务数字网(ISDN)技术

ISDN(Integrated Service Digital Network)是全数字的通信网。由于现有电话网一般是模拟网或模拟数字混合网,其传输速率低,最高仅56kbps(bits per second,位/秒),不能传输多媒体信息。ISDN支持文字、语音、数据、图形、活动图像的通信,可以实现全数字化的信息传输,是今后一段时间内多媒体通信的基本传输手段。随着对高清晰度电视和高速率业务的要求,光纤传输发展为宽带ISDN,即B-ISDN。它利用光纤做传输介质,采用异步传输模式,其传输速率从150Mbps到10Gbps,可支持多种媒体通信业务,满足多媒体通信对网络总带宽的要求,是实现多媒体有效通信的基本条件。由于ISDN技术上的成熟、稳定性和使用的经济性,已成为中小商业用户和家庭用户最经济、快速的数字接入互联网技术。

#### 2. 高速交换网技术

利用网段微化技术,通过在网段间建立多个并行连接,提供单独网段上的专用频带,有效地提高了网络的吞吐量,提高了传输效率。交换技术是计算机网络的核心技术。采用新的交换方式,变集中交换为分布交换、多层交换(三层交换和高层交换);使用高速交换设备,如光交换机等,以提高网间交互速度。

#### 3. 宽带网技术

宽带网能实现高速上网,且在一条线路同时支持语音、数据和视像等多种业务服务的网络。宽带网技术将从根本上改变全球范围信息传播手段。随着大容量光缆、数字交换机、卫星通信链路等世界范围的公共通信基础设施的建立,可在全球范围内高速、有效地传输多种类型的信息。

(1) 异步传输模式ATM。ATM是一种能同时满足文字、图形、图像、音频、视频等传输要求的技术,具有高速、大容量、实时等特点,真正实现无缝地高速传送多媒体信息。

(2) 电缆调制解调器。电缆调制解调器产品正逐渐成熟,电缆局部网已开始步入家庭和中小型企业,通过电缆插头可以方便地将有线电视直接接入因特网。

(3) 10G位以太网。为适应网络数据流量的迅速增加,2000年初,美国电气和电子工程师学会IEEE(Institute for Electrical and Electronic Engineers)进行了10G位以太网的开发,并发布了10G位以太网的802.3ae标准规范。10G位以太网不仅具有很高的带宽(传输速率达10亿位/秒,即10Gbps),且扩展了网络范围,使其超越局域网,进入广域网或城域网范围。此外,实时影视点播、远程医疗、远程教育和虚拟专用网等应用也需在宽带网发展的基础上实现。

#### 4. 网络标准化

网络标准化是时代的要求。国际标准化组织(ISO)在这方面做了大量的工作,其开放系统互连参考模型(OSI/RM)是标准化的基础,随着宽带网、光纤骨干网、移动通信和无线通

信等技术的应用和发展,新的标准也不断产生宽带标准化生产的产品,具有互联方便、可维护性好等优点。

#### 5. 移动通信技术

笔记本电脑和智能手持设备 SHD (Smart Handheld Device) 的发展,使得对移动无线网的要求日益增加。将固定综合宽带网通信业务用于移动环境中,采用新技术,提高数字传输速率,实现个人终端用户能在全球范围内的任何时间、任何地点与任何人用任何方式高质量地完成任何信息的移动通信与传输。预计,至 2010 年可实现 2Mbps~20Mbps 的传输速率,最高达 156Mbps,2020 年将最高达 600Mbps,最终实现人类通信的梦想——无线信息高速公路。

#### 6. 全球智能网的构筑

全球智能网的构筑,即把全球局域网与 Internet 融为一体,处理亿万个连接点,提供智能服务,包括确认网上用户身份、位置、需求和服务方法等。网络提供优良的可访问性和广泛的兼容性,是用户的智能助手。用户可在任何时候、任何地点访问网络,网络具有自动故障检测、诊断和排除功能。

#### 7. 复播技术

传统的网络应用局限在两台计算机之间进行相互操作。目前,一些新的应用,如网络电视、视频会议、协同计算等,它们均需在一组计算机之间进行通信,即多点通信。采用复播技术,打破传统的广播方式,实现信包投递信包组方式,即把同样信息复制多次,投递到组内每一个要接收此信息的成员。

#### 8. 宽带接入技术

Internet 时代的到来,使人们对信息的要求比以往任何时候都强烈。局域网、小型办公室以及家庭用户都迫切要求快速接入 Internet,这即是“最后一公里接入技术”。传统的公用电话网通过 Modem 拨号上网的最高传输速率为 56kbps,已远远无法满足人们快速上网的要求。随着信息技术的发展,特别是多媒体信息需求的提高,促进了宽带化、数字化接入技术的研究,并有了极大的发展。其宽带接入的方式有:有线电视电缆(Cable Modem)接入、电信网接入、卫星接入、固定无线接入和移动无线接入。其中,卫星接入、固定无线接入和移动无线接入是今后一段时间内需研究的内容。有线电视电缆接入,由于有线电视网具有频带宽(高达 1GHz)的特点,已被世界各国公认为是“信息高速公路最后一公里”接入具有巨大开发价值的方式,故有线电视电缆接入技术很有应用前景。

电信网接入中的基于双绞线的不对称数字用户线 ADSL (Asymmetric Digital Subscriber Line) 接入,由于双绞线用户在全世界占据了全部用户线的 90% 以上,充分利用该部分资源,开发新的宽带业务是中近期接入网发展的重要任务,ADSL 技术是近期有广阔应用前景的接入技术之一,它将代替传统的 Modem 方式,成为家庭和小型商务应用的主流接入技术。随着光纤价格的不断下降以及企业、家庭用户对宽带要求的不断增长,光纤系统将逐步从骨干网扩大到接入网。如混合光纤同轴(HFC)接入和基于光纤的无源光网络(PON)方式等,均是光纤接入技术今后研究的内容。

### 2.2.5 管理信息系统的运行模式

随着网络的发展,其体系结构按其发展过程,经历了文件/服务器模式、客户端/服务器(C/S)模式和浏览器/服务器(B/S)模式等阶段。



### 1. 文件/服务器模式

20世纪80年代,个人计算机进入了商用舞台,同时计算机应用的范围和领域也日趋广泛。这对那些没有能力实现大型机方案的企业来说,个人计算机无疑就有了用武之地。在个人计算机进入商用领域不久,局域网也问世了,同时也诞生了初期的主机—终端模式,数据处理和数据库应用全部集中在主机上,终端没有处理能力,这样所有的处理都在主机上执行,这样,终端用户过多时,主机负担过重,导致处理能力下降。造成了主机瓶颈。

到80年代以后文件/服务器模式开始流行,在文件服务器系统结构中,应用程序是在客户工作站上运行的,而不是在服务器上运行的,文件服务器只提供了资源(数据)的集中管理和访问途径。这种结构的特点是将共享数据资源集中管理,而将应用程序分散安排在各个客户工作站上。文件服务器结构的优点在于实现的费用比较低廉,而且配置非常灵活,在一个局域网中可以方便地增减客户端工作站。但是,文件服务器结构的缺点也非常明显,由于文件服务器只提供对数据的共享访问和文件管理服务,所有的应用处理都要在客户端完成,这就意味着客户端的个人计算机必须要有足够的处理能力,以便执行需要的任何程序,或能完成任何必要的任务。这可能经常需要客户端的计算机升级,否则改进应用程序的功能、提高应用程序的性能等都会成为一句空话。

另外,虽然应用程序可以存放在网络文件服务器的硬盘上,但它每次都要传送到客户端的个人计算机的内存中执行。因为所有的处理都是在客户端完成的,网络上就要经常传送大量无用的数据,使网络的负载过重,又导致了传输瓶颈。

### 2. 客户端/服务器(C/S)模式

虽然文件/服务器结构的费用低廉,但是和大型机的集中式相比,它缺乏足够的计算和处理能力。为了解决费用和性能的矛盾,客户端/服务器结构就应运而生了。这种结构允许应用程序分别在客户工作站和服务器(注意:不再是文件服务器)上执行,可以合理划分应用逻辑,充分发挥客户工作站和服务器两方面的性能。图2-4示意了客户端/服务器的结构。

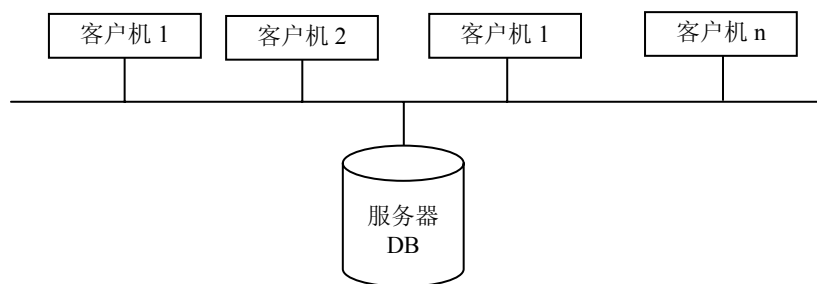


图2-4 客户端/服务器的结构

在客户端/服务器结构中,应用程序或应用逻辑可以根据需要划分在服务器和客户工作站中。这样,为了完成一个特定的任务,客户工作站上的程序和服务器上的程序可以协同工作。从图2-4可以看出客户端/服务器结构和文件/服务器结构的区别:客户端/服务器结构的客户工作站向服务器发送的是处理请求,而不是文件请求;服务器返回的是处理的结果,而不是整个文件。

当今,客户端/服务器最流行的领域就是数据库应用领域。比较著名的数据库厂商都提供了支持客户端/服务器结构的数据库管理系统,如Microsoft的SQL Server、Sybase的Adaptive Server和Oracle等。

从以上可以看出,大型机集中式结构的所有程序都在主机内执行,而文件服务器局域网结构的所有程序都在客户端执行,这两种结构都不能提供真正的可伸缩应用系统框架。而客户端/服务器结构则可以将应用逻辑分布在客户工作站和服务器之间,可以提供更快、更有效的应用程序性能。

### 3. 分布式模式

早期的客户端/服务器是基于两层结构的,即一层是客户层,一层是服务器层,这样的结构存在着一些弱点,第一,在网络计算机系统的中心位置不能插接应用程序组件,而在实际的应用中,很多企业往往需要经常为每个客户机插接一个应用程序,这样就给程序的维护带来很大的困难,在这种环境下,即使是一些小的修改,也会有很多困难;第二,由于客户机经常要处理一些敏感的数据,而在非中心化的计算环境中,无法对客户的信息存取进行控制,容易产生安全漏洞;第三,由于应用程序的业务逻辑存储在客户机上,所以就给客户机带来了沉重的负担。

为了适应企业应用,又出现了一种新的三层客户端/服务器模式。所谓三层,实际就是在客户层和服务器层间加入了一个中间层,客户层面向用户,服务器层提供数据服务,中间层面向商业或企业的需求,它可以是一个方案,通过相应的软件支持。中间层强调的是组件开发,需要什么功能就通过开发组件提供相应的功能,而不需更改客户端或服务器端。

### 4. 浏览器/服务器(B/S)模式

随着 Internet 的发展,分布在全世界各地的各种计算机系统及网络用户全部可以连接起来,他们可以通过共同的通信协议在不同的网络和操作系统间交换数据。

因此,人们正试图把 WWW 上的数据源集成为一个完整的 Web 数据库,从而使这些数据资源得到充分利用。因此,将 Web 技术与数据库技术相结合,开发动态的 Web 数据库应用势在必行。它是在 Web 服务器端提供中间件来连接 Web 服务器和数据库服务器。中间件负责管理 Web 服务器和数据库服务器之间的通信并提供应用程序服务,它能够直接访问数据库或调用外部程序或利用程序代码来访问数据库,因此它可以提供与数据库相关的动态 HTML 页面,或执行用户查询,并将查询结果格式化 HTML 页面,然后通过 Web 服务器返回给用户的浏览器。

## 2.3 数据库技术

### 2.3.1 数据库技术概述

随着数据管理规模的不断扩大,数据量急剧增加,用文件系统来组织和管理数据显露出许多缺陷。20 世纪 60 年代后期开始,计算机用于信息处理的规模越来越大,应用越来越广泛,对数据管理技术提出了更高的要求,与此同时出现了内存大、运行速度快的主机和大容量的磁盘,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,使数据为尽可能多的应用服务,新的数据管理技术——数据库技术产生了。而数据管理技术进入数据库阶段的标志是 20 世纪 60 年代末发生的三件大事:

(1) 1969 年,美国 IBM 公司研制并开发了基于层次结构的数据库管理系统——商品化软件 IMS (Information Management System)。

(2) 1969 年 10 月,美国数据系统语言协会 CODASYL (Conference On Data System Langure) 下属的数据库任务组 DBTG (Data Base Task Group),对数据库方法进行系统的研究

和讨论后，发表了关于网状数据模型的 DBTG 报告（1971 年通过）。

（3）1970 年，IBM 公司 San Jose 研究实验室的研究员 E.F.codd 发表了题为“大型共享数据库的数据关系模型”的论文。文中提出了数据库的关系模型，从而开创了关系数据库的研究领域，奠定了关系数据模型的理论基础。E.F.codd 因此在 1981 年获得了 ACM 图灵奖。

20 世纪 70 年代以来，数据库技术是计算机科学技术中发展最快的领域之一，也是应用最广泛的技术之一，它已经成为计算机信息系统和应用系统的核心技术和重要基础。数据库技术同时也带动了一个巨大的软件产业产品及其相关工具和解决方案，大量有效的商业化产品正在各行各业得到广泛应用。

数据库系统克服了文件系统的缺陷，提供了对数据更高级、更有效的管理，数据库系统具有如下特点：

### 1. 数据结构化

数据结构化是数据库与文件系统的根本区别。

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。

在数据库系统中，不仅要考虑某个应用的数据结构，还要考虑整个组织的数据结构。这种数据组织方式为各部分的管理提供了必要的记录，使数据结构化了。这就要求在描述数据时不仅要描述数据本身，还要描述数据之间的结构。例如一个学校涉及的多种应用，形成复杂的整体数据结构，如图 2-5 所示。

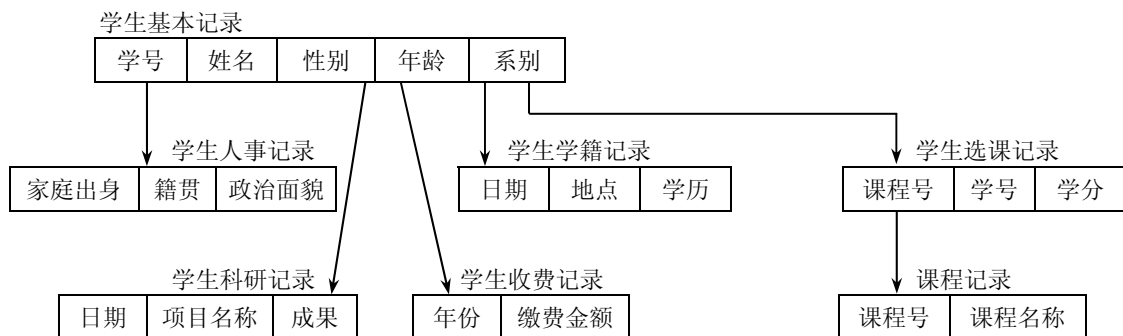


图 2-5 数据结构化

在数据库系统中，数据不再针对某一应用，而是面向全组织，具有整体的结构化。不仅数据是结构化的，而且存取数据的方式也很灵活，可以存取数据库中的某个数据项、一组数据项、一个记录或一组记录。

### 2. 数据的共享性高，冗余度低，易扩充

数据库系统从整体角度看待和描述数据，数据不再面向某个应用而是面向整个系统，因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余，节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。

由于数据面向整个系统，是有结构的数据，不仅可以被多个应用共享使用，而且容易增加新的应用。这就使得数据库系统弹性大，易于扩充，可以适应各种用户的要求。

### 3. 数据独立性高

数据独立性是由 DBMS 的二级映像功能来保证的。数据与程序的独立，把数据的定义从

程序中分离出去，加上数据的存取又由 DBMS 负责，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

在数据库系统中，对数据进行统一的管理和控制，数据库系统为用户提供了定义、检索和更新数据的手段，同时还提供了数据的安全性、完整性和并发控制等功能。

数据库系统中应用程序与数据之间的对应关系如图 2-6 所示。

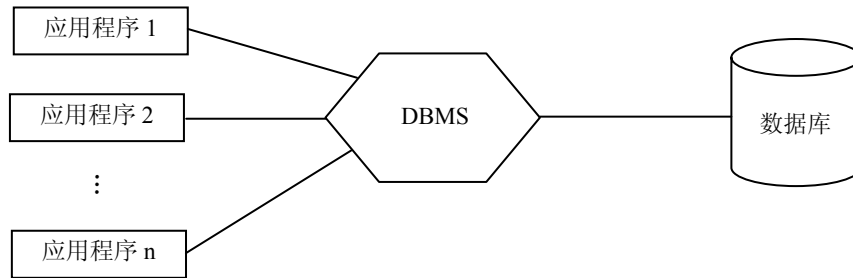


图 2-6 数据库系统应用程序与数据之间的对应关系

下面简单介绍数据库系统中的一些常用术语。

(1) 数据 (Data)。数据是数据库中存储的基本对象。描述事物的符号记录称为数据。描述事物的符号可以是数字，也可以是文字、图形、图像、声音、语言等，数据有多种表现形式，它们都可以经过数字化后存入计算机。

(2) 数据库 (DataBase, 简称 DB)。数据库是存放数据的仓库。只不过这个仓库是在计算机存储设备上，而且数据是按一定的格式存放的。

所谓数据库是指长期存储在计算机内的、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、描述和存储，具有较小的冗余度、较高的数据独立性和易扩展性，并可为各种用户共享。

(3) 数据库管理系统 (DataBase Management System, 简称 DBMS)。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，主要功能有：数据定义功能、数据操纵功能、数据库的运行管理和数据库的建立与维护功能。

(4) 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，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。

### 2.3.2 数据模型

#### 1. 信息抽象

信息是人们对客观世界各种事物特征的反映，而数据则是表示信息的一种符号。从客观事物到信息，再到数据，是人们对现实世界的认识和描述过程，这里经过了三个层次的变换，分别是现实世界、信息世界和数据世界。

(1) 现实世界。现实世界是客观存在的事物以及事物间的联系，是由事物本身的性质决定的。例如学校的教学管理系统中有学生、课程和教师，教师教授课程，学生选修课程等。

(2) 信息世界。信息世界是客观存在的现实世界在人们头脑中的反映，是对客观事物及

其联系的一种抽象描述。它是现实世界到数据世界必须经过的中间层次。信息世界涉及的概念主要有：

1) 实体 (Entity)。客观存在并可相互区别的事物称为实体。实体可以是具体的对象，例如，一个职工、一个学生、一个部门、一门课；也可以是抽象事件，例如学生的一次选课、部门的一次订货等。

2) 属性 (Attribute)。实体所具有的某一特性称为属性。一个实体可以由若干个属性来刻画。例如，学生实体可由学号、姓名、性别、年龄、系别等属性来描述。属性的具体值称为属性值，用来刻画一个具体实体。如属性值的组合 (020611, 李明, 男, 19, 计算机系) 就表示一名学生。

3) 码 (Key)。唯一标识实体的属性集称为码。例如学号是学生实体的码。

4) 域 (Domain)。属性的取值范围称为该属性的域。例如，学号的域为 6 位整数，姓名的域为字符串集合等。

5) 实体型 (Entity Type)。具有相同属性的实体必然具有共同的特征和性质。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, 系别) 就是一个实体型。

6) 实体集 (Entity Set)。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。

7) 联系 (Relationship)。实体间的联系通常是指不同实体集之间的联系。两个实体型之间的联系可以分为三类：

- 一对一联系 (1:1)。设 A、B 为两个实体集。如果对于 A 中的每一个实体，B 中至多有一个实体与其发生联系；反之，B 中的每一个实体至多对应 A 中的一个实体，则称 A 与 B 具有一对一联系，记为 1:1。例如班级和班长之间具有一对一联系。
- 一对多联系 (1:n)。如果对于 A 中的每一个实体，B 中有 n 个实体 ( $n \geq 0$ ) 与之联系；反之，对于 B 中的每一个实体，A 中至多只有一个实体与之联系，则称 A 与实体 B 有一对多联系，记为 1:n。例如班级和学生之间具有一对多联系。
- 多对多联系 (m:n)。如果对于 A 中的每一个实体，B 中有 n 个实体 ( $n \geq 0$ ) 与之联系；反之，对于 B 中的每一个实体，A 中也有 m 个实体 ( $m \geq 0$ ) 与之联系，则称 A 与实体 B 具有多对多联系，记为 m:n。例如课程和学生之间具有多对多联系。

## 2. 数据模型

数据模型是现实世界数据特征的抽象，应满足三方面的要求：一是能比较真实地模拟现实世界；二是容易为人所理解；三是便于在计算机上实现。不同的数据模型实际上是提供给人们模型化数据和信息不同工具。根据模型应用的不同目的，可以将这些模型划分为两类，它们分属于两个不同的层次：一类是概念模型，另一类是数据模型。

(1) 概念模型。概念模型用于信息世界的建模，是现实世界到信息世界的第一层抽象，是数据库设计人员进行数据库设计的有力工具，也是数据库设计人员和用户之间进行交流的语言。

概念模型的表示方法很多，其中最为著名、最常用的是 P.P.S.Chen 于 1976 年提出的实体—联系 (Entity-Relationship Approach)。该方法用 E-R 图来描述现实世界的概念模型。

E-R 图提供了表示实体型、属性和联系的方法。

- 实体型：用矩形表示，矩形框内写明实体名。
- 属性：用椭圆形表示，并用无向边将其与相应的实体连接起来。

- 联系：用菱形表示，菱形框内写明联系名，并用无向边分别与有关实体连接起来，同时在无向边旁标上联系的类型（1:1、1:n 或 m:n）。

需注意的是，联系本身也是一种实体型，也可以有属性。如果一个联系具有属性，则这些属性也要用无向边与该联系连接起来。

下面用 E-R 图来表示某学校学籍管理的概念模型。

其中，涉及的实体包括：

- 学校：校名，校长名。
- 系部：系名，系主任，联系电话。
- 教师：工号，姓名，性别，职称。
- 学生：学号，姓名，性别，年龄。
- 课程：课程号，课程名，学分。
- 参考书：书号，书名，作者。

这些实体之间的联系包括：

- 一个学校有多个系部，因此学校和系部之间是 1:n 的联系。
- 一个系部有多名学生，因此系部和学生之间是 1:n 的联系。
- 一名学生可选修多门课程，一门课程可被多名学生选修，因此学生和课程之间是 m:n 的联系。
- 一门课程可以有若干个教师讲授，使用若干本参考书，而每个教师只讲授一门课程，每本参考书只供一门课程使用，则课程与教师、参考书之间的联系是一对多的。

下面给出该学校的学籍管理 E-R 图，如图 2-7 所示。

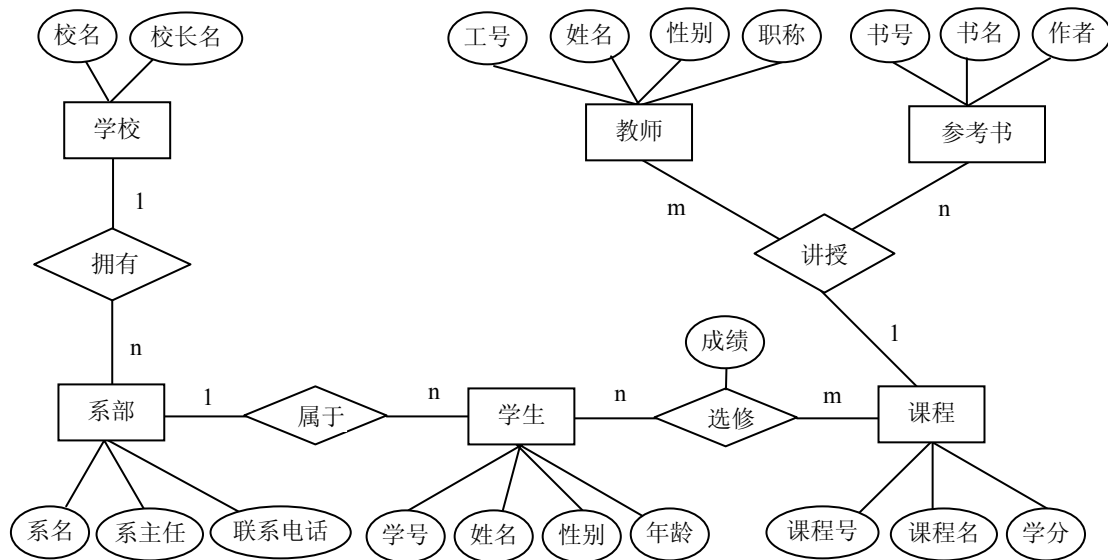


图 2-7 学校学籍管理 E-R 图

(2) 数据模型。把现实世界的客观对象抽象成信息世界的概念模型后，需要再把概念模型转换为计算机支持的数据模型。数据模型是数据库系统设计中用于提供信息表示和操作手段的形式构架，是数据库实现的基础。

目前，在实际数据库系统中支持的数据模型主要有三种：层次模型（Hierarchical Model）、

网状模型 (Network Model) 和关系模型 (Relational Model)。20 世纪 80 年代以来, 计算机厂商推出的数据库管理系统几乎都是支持关系模型的数据库系统。关系模型已经占领市场主导地位。

关系模型建立在严格的数学概念的基础上, 它用二维表来描述实体与实体间的联系。下面以学生信息表 (如表 2-1 所示) 为例, 介绍关系模型中的一些术语。

表 2-1 学生信息表

学号	姓名	年龄	性别	系别
0206611	王勇	18	男	计算机
0206612	李力	19	男	计算机
0206613	张民	18	男	计算机
0206614	刘芳	18	女	计算机

- 关系 (Relation): 对应通常说的一张表, 如表 2-1。
- 元组 (Tuple): 表中的一行即为一个元组。
- 属性 (Attribute): 表中的一列即为一个属性, 给每个属性起一个名称即属性名。如表 2-1 的五个属性 (学号, 姓名, 年龄, 性别, 系别)。
- 主码 (Key): 表中的某个属性组, 它可以唯一确定一个元组。如表 2-1 的学号。
- 域 (Domain): 属性的取值范围。
- 分量: 元组中的一个属性值。
- 关系模式: 对关系的描述, 一般表示为: 关系名(属性 1, 属性 2, ..., 属性 n)。

例如上面的关系可以描述为:

学生(学号, 姓名, 年龄, 性别, 系别)

一个关系模型是若干个关系模式的集合。在关系模型中, 实体以及实体间的联系都是用关系来表示。例如学生、课程、学生与课程之间的多对多联系在关系模型中可以如下所示:

学生(学号, 姓名, 年龄, 性别, 系别)

课程(课程号, 课程名, 学分)

选修(学号, 课程号, 成绩)

由于关系模型概念简单、清晰、易懂、易用, 并有严密的数学基础以及在此基础上发展起来的关系数据理论, 简化了程序开发及数据库建立的工作量, 因而迅速获得了广泛的应用, 并在数据库系统中占据了统治地位。

### 2.3.3 数据库系统结构

从数据库管理系统角度看, 数据库系统通常采用三级模式结构; 从数据库最终用户角度看, 数据库系统的体系结构分为单用户结构、主从式结构、分布式结构和客户/服务器结构。

#### 1. 数据库系统的模式结构

模式 (Schema) 是数据库中全体数据的逻辑结构和特征的描述, 它仅仅涉及到型的描述, 不涉及到具体的值。模式的一个具体值称为模式的一个实例 (instance)。同一个模式可以有很多实例。模式是相对稳定的, 而实例是相对变动的。模式反映的是数据的结构及其关系, 而实例反映的是数据库某一时刻的状态。

(1) 数据库系统的三级模式结构。数据库系统的三级模式结构是指数据库系统是由外模

式、模式和内模式三级构成，如图 2-8 所示。

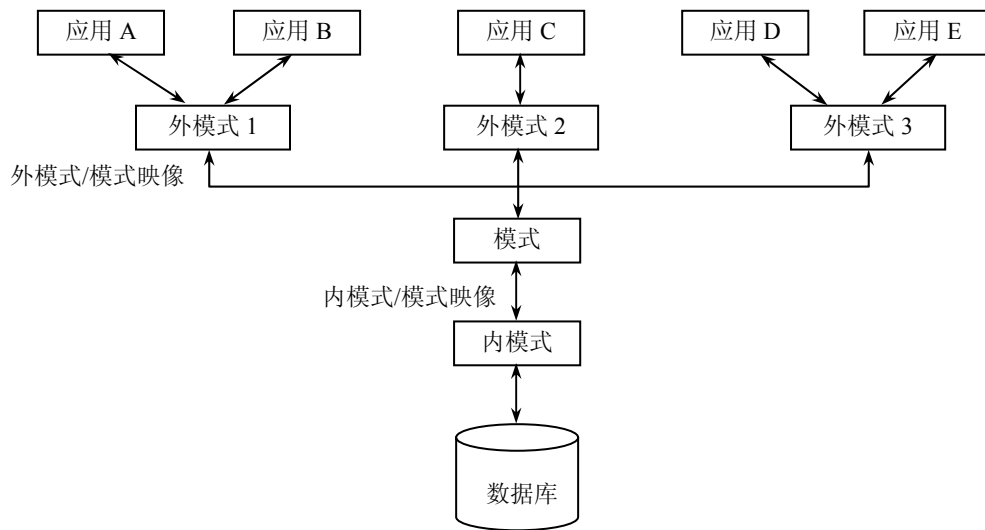


图 2-8 数据库系统的三级模式结构

1) 模式 (Schema)。模式也称逻辑模式，是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图。它是数据库系统模式结构的中间层，即不涉及数据的物理存储细节和硬件环境，也与具体的应用程序、应用开发工具及高级程序设计语言无关。

模式实际上是数据库数据在逻辑级上的视图。一个数据库只有一个模式。数据库模式以某一种数据模型为基础。定义模式时不仅要定义数据的逻辑结构（例如数据记录由哪些数据项构成，数据项的名字、类型、取值范围等），而且要定义与数据有关的安全性、完整性要求，定义这些数据之间的联系。

2) 外模式 (External Schema)。外模式也称子模式或用户模式，它是数据库用户（包括应用程序员和最终用户）能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示。

外模式通常是模式的子集。一个数据库可以有多个外模式。由于它是各个用户的数据视图，如果不同的用户在不同应用需求、看待数据的方式、对数据保密的要求等各方面存在差异，则其外模式描述就是不同的。即使对模式中同一数据，在外模式中的结构、类型、长度、保密级别等都可以不同。另一方面，同一外模式也可以为某一用户的多个应用系统所使用，但一个应用程序只能使用一个外模式。

外模式是保证数据库安全性的一个有力措施。每个用户只能看见和访问所对应的外模式中的数据，数据库中的其余数据是不可见的。

3) 内模式 (Internal Schema)。内模式也称存储模式，一个数据库只有一个内模式。它是数据物理结构和存储结构的描述，是数据在数据库内部的表示方式（例如，记录的存储方式是顺序存储、按照 B 树结构存储还是按 Hash 方法存储；索引按照什么方式组织；数据是否压缩存储，是否加密；数据的存储记录结构有何规定等）。

数据库系统的三级模式是对数据的三个抽象级别，它把数据的具体组织留给 DBMS 管理，使用户能逻辑地抽象地处理数据，而不必关心数据在计算机中的具体表示方式与存储方式。为了能够在内部实现这三个抽象层次的联系和转换，数据库管理系统在这三级模式之间提供了两



层映像：外模式/模式映像和模式/内模式映像。正是这两层映像保证了数据库系统中的数据能够具有较高的逻辑独立性和物理独立性。

(2) 数据库的二级映像功能与数据独立性。对于每一个外模式，数据库系统都有一个外模式/模式映像，它定义了该外模式与模式之间的对应关系。这些映像定义通常包含在各自外模式的描述中。

当模式改变时，由数据库管理员对各个外模式/模式的映像作相应改变，可以使外模式保持不变，从而应用程序不必修改，保证了数据的逻辑独立性。

数据库中只有一个模式，也只有一个内模式，所以模式/内模式映像是唯一的，它定义了数据的全局逻辑结构与存储结构之间的对应关系。该映像定义通常包含在模式描述中。

当数据库的存储结构改变了（例如采用了更先进的存储结构），由数据库管理员对模式/内模式映像作相应改变，可以使模式保持不变，从而保证了数据的物理独立性。

### 2.3.4 关系数据库标准语言 SQL

SQL 是 Structure Query Language 的缩写。SQL 语言是一种类似于英语的数据语言，1986 年美国国家标准化组织 ANSI 确认 SQL 作为数据库系统的工业标准。所以商用的数据库系统都采用 SQL 语言作为数据语言，或者提供对 SQL 语言的支持。

SQL 语言的命令一般分为四类：

#### 1. 查询语句

查询语句是用来对已经存在数据库中的数据按照特定的组合、条件表达式或者次序进行检索，其一般格式为：

```
SELECT[ALL|DISTINCT]<目标列表表达式>[,<目标列表表达式>]...  
FROM<表名或视图名>[,<表名或视图名>]...  
[WHERE<条件表达式>]  
[GROUP BY<列名 1> [HAVING<条件表达式>]]  
[ORDER BY<列名 2> [ASC|DESC]]
```

整个 SELECT 语句的含义是，根据 WHERE 子句的条件表达式，从 FROM 子句指定的基本表或视图中找出满足条件的元组，再按 SELECT 子句中的目标列表表达式，选出元组中的属性值形成结果表。如果有 GROUP 子句，则将结果按<列名 1>的值进行分组，该属性列值相等的元组为一个组，每个组产生的结果在表中是一条记录。如果 GROUP 子句带 HAVING 短语，则只有满足指定条件的组才允许输出。如果带有 ORDER 子句，则结果表还要按<列名 2>的值进行升序或降序排序。

#### 2. 数据操纵语句 DML (Data Manipulation Language)

DML 语句用来改变数据库中的数据，包括三种基本形式。

(1) INSERT 语句。INSERT 语句可给数据库中的某个表添加一个或多个新行。INSERT 语句在简单的情况下的形式如下：

```
INSERT [INTO] table_or_view [(column_list)] data_values
```

此语句将使 data\_values 作为一行或者多行插入已命名的表或视图中。column\_list 是由逗号分隔的列列表，用来指定为其提供数据的列。如果没有指定 column\_list，表或者视图中的所有列都将接收数据。

如果 column\_list 没有为表或视图中的所有列命名，将在列表中没有命名的任何列中插入一个 NULL 值（或者是默认情况下，这些列定义的默认值）。在列的列表中没有指定的所有列

都必须允许 NULL 值或者指定的默认值。

所提供的数据值必须与列的列表匹配。数据值的数目必须与列数相同，每个数据值的数据类型、精度和小数位数也必须与相应的列匹配。有两种方法指定数据值：

1) 用 VALUES 子句为一行指定数据值。

```
INSERT INTO table_or_view [(column_list)]
    VALUES (value1,value2...)
```

2) 用 SELECT 子查询为一行或多行指定数据值。

```
INSERT INTO table_or_view [(column_list)]
    SELECT 子查询
```

(2) UPDATE 语句。UPDATE 语句可以更改表或视图中单行、行组或所有行的数据值。还可以用该语句更新远程服务器上的行（使用链接服务器名称或 OPENROWSET、OPENDATASOURCE 和 OPENQUERY 函数），前提是用来访问远程服务器的 OLE DB 提供程序支持更新操作。引用某个表或视图的 UPDATE 语句每次只能更改一个表中的数据。UPDATE 语句在简单的情况下有如下形式：

```
UPDATE {table_name | view_name} SET {column=expression[,...]} WHERE CURRENT OF cursor_name
```

其中：

- SET 子句：包含要更新的列和每个列的新值的列表（用逗号分隔），格式为 column\_name = expression。表达式提供的值包含多个项目，如常量、从其他表或视图的列中选择或使用复杂的表达式计算出来的值。
- FROM 子句：指定为 SET 子句中的表达式提供值的表或视图，以及各个源表或视图之间可选的联接条件。
- WHERE 子句：指定搜索条件，该搜索条件定义源表和视图中可以为 SET 子句中的表达式提供值的行。

(3) DELETE 语句。DELETE 语句可删除表或视图中的一行或多行。DELETE 语法的简化形式为：

```
DELETE table_or_view FROM table_sources WHERE search_condition
```

其中，table\_or\_view 是指定要从中删除行的表或视图。table\_or\_view 中所有符合 WHERE 搜索条件的行都将被删除。如果没有指定 WHERE 子句，将删除 table\_or\_view 中的所有行。FROM 子句是指定删除时用到的额外的表或视图及联接条件，使用 WHERE 子句搜索条件中的谓词来限定要从 table\_or\_view 中删除的行。该语句不从 FROM 子句指定的表中删除行，而只从 table\_or\_view 指定的表中删除行。

任何已删除所有行的表仍会保留在数据库中。DELETE 语句只从表中删除行，要从数据库中删除表，必须使用 DROP TABLE 语句。

### 3. 数据定义语句 DDL (Data Definition Language)

DDL 用来建立数据库中各种数据对象（包括表、视图、索引、存储过程、触发器等），有三种基本形式：

(1) CREATE：新建数据库对象。

(2) ALTER：更新已有数据对象的定义。

(3) DROP：删除已经存在的数据对象。

### 4. 数据控制语句 DCL (Data Control Language)

DCL 用于授予或者收回访问数据库的某种权限和事务控制，主要包括四种基本形式：

- (1) GRANT: 授予权限。
- (2) REVOKE: 收回权限。
- (3) COMMIT: 提交事务。
- (4) ROLLBACK: 回滚事务。

SQL 语言之所以能成为工业标准, 主要因为它是一种综合的、功能强、易学又便于使用的语言。主要特点如下所示:

(1) 高度非过程化。SQL 语言进行数据操作只要提出“做什么”, 不需要说明“如何做”。这大大减轻了数据库用户的负担, 有利于提高生产率。

(2) 面向集合的操作方式。SQL 语句采用集合操作方式, 例如要查询所有男职员的信息, 数据库管理系统不会一次返回一个职工的信息, 而是一次返回所有满足条件的记录。除了查询之外, 插入、删除、修改操作都可以一次处理多行记录, 也就是集合操作。

(3) 语法简单。SQL 语言功能强大, 但是语法极其简单。完成核心功能共有 9 个动词, 如表 2-2 所示。

表 2-2 SQL 语言的动词

SQL 功能	动词
数据查询	SELECT
数据定义	CREATE, DROP, ALTER
数据操纵	INSERT, UPDATE, DELETE
数据控制	GRANT, REVOKE

### 2.3.5 关系的规范化

前面讨论了数据组织的概念、关系数据库的基本概念、关系模型的三个部分以及关系数据库的标准语言。但是还有一个很基本的问题尚未涉及, 针对一个具体问题, 应该如何构造一个适合于它的数据库模式, 即应该构造几个关系模式, 每个关系由哪些属性组成等。关系的属性间存在相互依赖又相互制约的联系称为数据依赖, 其中很重要的一种叫函数依赖, 下面介绍有关函数依赖的概念。

**定义 1** 设  $R(U)$  是属性集  $U$  上的关系模式。  $X$ 、 $Y$  是  $U$  的子集。若对于  $R(U)$  的任意一个可能的关系  $r$ ,  $r$  中不可能存在两个元组在  $X$  上的属性值相等, 而在  $Y$  上的属性值不等, 则称  $X$  函数确定  $Y$  或  $Y$  函数依赖于  $X$ , 记作  $X \rightarrow Y$ 。

**定义 2** 在  $R(U)$  中, 如果  $X \rightarrow Y$ , 并且对于  $X$  的任何一个真子集  $X'$ , 都有  $X' \not\rightarrow Y$ , 则称  $Y$  对  $X$  完全函数依赖, 记作:  $X \xrightarrow{F} Y$ 。

例如, 关系选修(学号,课程号,成绩)中, (学号,课程名)  $\xrightarrow{F}$  成绩。

**定义 3** 若  $X \rightarrow Y$ , 但  $Y$  不完全函数依赖于  $X$ , 称  $Y$  对  $X$  部分函数依赖, 记作:  $X \xrightarrow{P} Y$ 。

例如, 关系学生(学号,姓名,性别,年龄,系别)中, (学号,姓名)  $\xrightarrow{P}$  年龄。

**定义 4** 在  $R(U)$  中, 如果  $X \rightarrow Y$ ,  $Y \subsetneq X$ ,  $Y \rightarrow X$ ,  $Y \rightarrow Z$ , 则称  $Z$  对  $X$  传递函数依赖, 记作:  $X \xrightarrow{T} Y$ 。

例如，设有关系 BOOKS(BNO,TITLE,LOCA)，其中一本书对应唯一书号，并可能为某一个出版社出版；一个出版社一般只有一个唯一名称和地址，但一个出版社可出版多种书。该关系拥有的函数依赖： $BNO \rightarrow TITLE$ ， $TITLE \rightarrow LOCA$ ， $TITLE \rightarrow LOCA$ 。因此，该关系中存在传递函数依赖  $BNO \xrightarrow{T} LOCA$ 。

为了使数据库设计的方法走向完备，人们研究了规范化理论，指导人们设计规范的数据库模式。从 1971 年起 E.F.Codd 就提出了规范化理论，至今有关规范化理论的研究已经取得了很多成果。关系数据库中的关系是要满足一定要求的，满足不同程度要求的为不同范式。满足最低要求的叫第一范式，简称 1NF。在第一范式中满足进一步要求的为第二范式，其余以次类推。关系规范化的程度为第一范式、第二范式、第三范式、第四范式和第五范式等。

### 1. 1NF

任给关系 R，如果 R 中每个列与行的交点处的取值都是不可再分的基本元素，则  $R \in 1NF$ 。

不属于 1NF 的关系称为非规范化关系，如表 2-3 所示。经过转换形成规范化的关系，成为 1NF，如表 2-4 所示。

表 2-3 非规范化关系

个人信息			选课信息	
学号	系别	住址	课程名	成绩
9962001	计算机系	本部	高数	70
9962002	计算机系	本部	英语	85
9992001	医学系	东校区	高数	86
9992001	医学系	东校区	体育	72
9971001	人文系	顺道街	高数	92

表 2-4 规范化关系 S-L-C

学号	系别	住址	课程名	成绩
9962001	计算机系	本部	高数	70
9962002	计算机系	本部	英语	85
9992001	医学系	东校区	高数	86
9992001	医学系	东校区	体育	72
9971001	人文系	顺道街	高数	92

该关系的候选码为： $(学号,课程名)$ ，函数依赖为： $(学号,课程名) \xrightarrow{F} 成绩$ ， $学号 \xrightarrow{F} 系别$ ， $(学号,课程名) \xrightarrow{P} 系别$ ， $系别 \xrightarrow{F} 住址$ 。

1NF 存在的问题：

(1) 插入异常。假若要插入一个学生“学号=9962003，系别=计算机系，住址=本部”，但该生还未选课，即这个学生无课程号，这样的元组就插不进关系模式 S-L-C 中。因为插入该元组时必须给定码值，而这时码值的一部分为空，因而学生的固有信息无法插入。

(2) 删除异常。假定某个学生只选一门课，如 9971001 就选了一门课“高数”。现在“高数”这门课他不选了，那么“高数”这个数据项就要删除。而“高数”是主属性，删除了“高数”，整个元组就必须跟着删除，使得 9971001 的其他信息也被删除了，从而造成删除异常，

即不应删除的信息也删除了。

(3) 修改复杂。某个学生从医学系转到计算机系，这本来只需修改此学生元组中的系别分量。但因为关系模式 S-L-C 中还含有住址属性，学生转系将同时改变住处，因而还必须修改元组中的住址分量。

(4) 冗余度大。如果这个学生选修了 k 门课，系别、住址重复存储了 k 次，存储冗余大。

存在上述问题的原因在于：在关系模式中存在着两种非主属性。一种如成绩，它对码是完全函数依赖。另一种如系别、住址，对码是部分函数依赖。

解决问题的办法是把关系模式 S-L-C 分解为两个关系模式：关系 SC 和关系 S-L。分解后的关系模式 SC 与关系模式 S-L 中不再存在部分函数依赖，从而解决了插入异常与删除异常的问题。

## 2. 2NF

若  $R \in 1NF$ ，且每一个非主属性完全函数依赖于码，则  $R \in 2NF$ 。例如，经过上述 1NF 分解形成两个关系 SC（如表 2-5 所示）和 S-L（如表 2-6 所示）。

表 2-5 关系 SC

学号	课程名	成绩
9962001	高数	70
9962002	英语	85
9992001	高数	86
9992001	体育	72
9971001	高数	92

表 2-6 关系 S-L

学号	系别	住址
9962001	计算机系	本部
9962002	计算机系	本部
9992001	医学系	东校区
9971001	人文系	顺道街

由定义判断关系模式  $SC \in 2NF$ ，关系模式  $S-L \in 2NF$ 。

2NF 关系存在的问题：

(1) 修改复杂。由于每个系的学生住在同一个地方，因此，当一个系的住址改变时，则要修改多个元组的值。

(2) 冗余度大。如一个系有 500 名学生，那么学生的住址则重复存储 500 次，造成存储空间的冗余度大。

在 2NF 中存在问题的原因是：关系模式 S-L 存在非主属性住址对码学号的传递依赖。在关系模式 S-L 中，由于学号  $\rightarrow$  系别，系别  $\rightarrow$  住址，得到传递函数依赖学号  $\xrightarrow{T}$  住址。

解决问题的方法是将 S-L 分解为：关系模式 S-D 和关系模式 D-L。

分解后的关系模式 S-D 与关系模式 D-L 中不再存在传递函数依赖，从而解决了修改复杂与冗余度大的问题。

## 3. 3NF

关系模式  $R \langle U, F \rangle$  中若不存在这样的码 X，属性组 Y 及非主属性 Z ( $Z \not\subseteq Y$ ) 使得  $X \rightarrow Y$ ,  $Y \twoheadrightarrow X$ ,  $Y \rightarrow Z$  成立，则称  $R \langle U, F \rangle \in 3NF$ 。

例如，经过上述 2NF 分解形成两个关系 S-D（如表 2-7 所示）和 D-L（如表 2-8 所示）。

由定义判断：关系模式 S-D  $\in 3NF$ ，关系模式 D-L  $\in 3NF$ 。

3NF 仍可能存在问题：插入异常、删除异常和修改复杂。在这种情况下，可以通过进一步规范化对其进行分解。分解后，原来的插入及删除操作异常问题已不存在。

表 2-7 关系模式 S-D

学号	系别
9962001	计算机系
9962002	计算机系
9992001	医学系
9971001	人文系

表 2-8 关系模式 D-L

系别	住址
计算机系	本部
医学系	东校区
人文系	顺道街

在关系数据库中，对关系模式的基本要求是满足第一范式。这样的关系模式就是合法的、允许的。但是有些关系模式存在插入、删除异常、修改复杂、数据冗余等问题。人们寻求解决这些问题的方法，这就是规范化的目的。

规范化的基本思想是逐步消除函数依赖中不合适的部分，使模式中的各关系模式达到某种程度的分离。让一个关系描述一个概念、一个实体或者实体间的一种联系，因此所谓规范化实质上是概念的单一化。关系模式的规范化过程是通过分解来实现的，从认识非主属性的部分函数依赖的危害开始，2NF、3NF、BCNF 的提出是这个认识过程逐步深化的标志。关系模式的规范化过程就是把低一级的关系模式分解为若干个高一级的关系模式的过程。

### 2.3.6 数据库设计

数据库技术是信息资源管理最有效的手段。数据库设计是指对于一个给定的应用环境，构造最优的数据库模式，建立数据库及其应用系统，有效存储数据，满足用户信息要求和处理要求。

按照规范设计的方法，考虑数据库及其应用系统开发全过程，将数据库设计分为以下六个阶段。

(1) 需求分析阶段。需求收集和分析，结果得到数据字典描述的数据需求（和数据流程图描述的处理需求）。

(2) 概念结构设计阶段。通过对用户需求进行综合、归纳与抽象，形成一个独立于具体 DBMS 的概念模型，可以用 E-R 图表示。

(3) 逻辑结构设计阶段。将概念结构转换为某个 DBMS 所支持的数据模型（例如关系模型），并对其进行优化。

(4) 数据库物理设计阶段。为逻辑数据模型选取一个最适合应用环境的物理结构（包括存储结构和存取方法）。

(5) 数据库实施阶段。运用 DBMS 提供的数据库语言（例如 SQL）及其宿主语言（例如 C），根据逻辑设计和物理设计的结果建立数据库，编制与调试应用程序，组织数据入库，并进行试运行。

(6) 数据库运行和维护阶段。数据库应用系统经过试运行后即可投入正式运行。在数据库系统运行过程中必须不断地对其进行评价、调整与修改。

### 2.3.7 数据库保护

为了保证数据的安全可靠和正确有效，DBMS 必须提供统一的数据保护功能，主要包括数据的安全性、完整性、并发控制和数据库恢复等内容。

数据的安全性是指保护数据库以防止不合法的使用所造成的数据泄露、更改和破坏。数据的安全可通过对用户进行标识和鉴定、存取控制、OS 级安全保护等措施得到一定的保障。

数据的完整性是指数据的正确性、有效性与相容性。关系模型的完整性有实体完整性、参

照完整性及用户自定义的完整性。

(1) 实体完整性。指二维表中描述主关键字的属性不能取空值。如学生基本信息表中的属性“学号”被定义为主关键字,则“学号”的值不能为空。

(2) 参照完整性。指具有一对多联系的两个表之间,子表中与主表的主关键字相关联的那个属性(外部码)的值要么为空,要么等于主表中主关键字的某个值。

(3) 用户自定义的完整性。它是针对某一具体数据库的约束条件,由应用环境确定。例如月份必须是1~12的正整数,研究生的年龄应小于45等。

并发控制是指当多个用户同时存取、修改数据库时,可能会发生互相干扰而得到错误的结果并使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、协调。

数据库恢复是指当计算机软、硬件或网络通信线路发生故障而破坏了数据或对数据库的操作失败使数据出现错误或丢失时,系统应能进行应急处理,把数据库恢复到正常状态。

### 2.3.8 数据库系统的发展

数据库技术从20世纪60年代中期发生到现在,仅短短30年已从第一代的网状、层次数据库系统,第二代的关系数据库系统,发展到第三代以面向对象模型为主要特征的数据库系统。数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透、互相结合,成为当前数据库技术发展的主要特征。

#### 1. 数据库技术与其他相关技术相结合

数据库技术与其他学科的内容相结合,是新一代数据库技术的一个显著特征,涌现出各种新型的数据库系统。例如:数据库技术与分布处理技术相结合,出现了分布式数据库系统;数据库技术与并行处理技术相结合,出现了并行数据库系统;数据库技术与人工智能相结合,出现了演绎数据库系统、知识库和主动数据库系统;数据库技术与多媒体处理技术相结合,出现了多媒体数据库系统;数据库技术与模糊技术相结合,出现了模糊数据库系统等。

(1) 分布式数据库系统。随着地理上分散的用户对数据库共享的要求,结合计算机网络技术的发展,在传统的集中式数据库系统基础上产生和发展了分布式数据库系统。

分布式数据库应具有以下特点:

- 数据的物理分布性。
- 数据的逻辑整体性。
- 数据的分布独立性。
- 场地自治和协调。
- 数据的冗余及冗余透明性。

(2) 并行数据库系统。并行数据库系统通过充分利用通用并行计算机的处理机、磁盘等硬件设备的并行数据处理能力来提高数据库系统的性能。并行数据库系统自出现以来,已经取得了一系列引人注目的进展,主要原因有两个:

1) 关系数据模型和SQL语言的广泛采纳。

2) 近十几年来,计算机的处理能力得到了非常迅速的提高,高性能CPU、大容量内存、高速磁盘阵列以及高宽带通信网络的出现为高性能数据处理提供了充分的硬件支持。

国外Intel、NCR、nCUBE、Sequent、Tandem、Teradata公司和国内曙光等许多供应商已经能够提供基于快速、廉价的处理器、内存、磁盘和通信网络的多处理机系统。这些机器以较低的价格提供了比大型机更为强大的整体处理能力。并且,这些并行计算机具有很好的伸缩性,

其模块化体系结构使得系统能够不断增长,通过增加处理机、内存和磁盘使得系统对于给定的任务能够加速完成,或者使得系统在相同的时间内能够处理大规模的任务。

(3) 知识库系统。知识库系统是数据库和人工智能两种技术结合的产物,目前,已成为数据库研究的热门课题之一。数据库和人工智能是计算机科学两个非常重要的领域,它们相互独立发展,在各自的领域取得了突出成就并获得广泛应用。

在数据库技术中引入人工智能技术,多年来是沿着数据库的智能化和智能化的数据库这两个途径发展的。

(4) 多媒体数据库。媒体是信息的载体。多媒体是指多种媒体,如数字、文字、图形、图像和声音的有机集成,而不是简单的组合。其中数字、字符等称为格式化数据,文本、图形、图像、声音、视象等称为非格式化数据。非格式化数据具有大数据量、处理复杂等特点。多媒体数据库实现对格式化和非格式化的多媒体数据的存储、管理和查询,其主要特征有:

- 能够表示多种媒体的数据。
- 能够协调处理各种媒体数据。
- 提供更强的适合非格式化数据查询的搜索功能。
- 多媒体数据库应提供特种事务处理与版本管理能力。

(5) 主动数据库。主动数据库(Active Data Base)是相对于传统数据库的被动性而言的。许多实际的应用领域,如计算机集成制造系统、管理信息系统、办公室自动化系统中常常希望数据库系统在紧急情况下能根据数据库的当前状态,主动适时地做出反应,执行某些操作,向用户提供有关信息。

主动数据库通常采用的方法是在传统数据库系统中嵌入 ECA(即事件—条件—动作)规则,在某一事件发生时引发数据库管理系统去检测数据库当前状态,看是否满足设定的条件,若条件满足,便触发规定动作的执行。

为了有效地支持 ECA 规则,主动数据库的研究主要集中于解决以下问题:

- 主动数据库的数据模型和知识模型。
- 执行模型。
- 条件检测。
- 事务调度。
- 体系结构。
- 系统效率。

(6) 模糊数据库。20 世纪 80 年代中后期开始,国内外出现了一系列对模糊关系数据库的有关概念和方法的研究,建立了一套较合理的理论体系,包括各种模糊数据表示、模糊距离的度量、模糊数据模型、模糊语言、模糊查询方法等,并且摸索了一些较实用的实现技术,在计算机上实现了一些基于上述理论和技术的模糊数据库管理系统的原型。

模糊数据库系统可以定义为:存储、组织、管理和操作模糊数据的数据库系统。它除了具有普通数据库系统的公共特性外,在以下方面体现了模糊性:

- 存储的是以各种形式表示的模糊数据。
- 数据结构和数据之间的联系是模糊的。
- 数据上的运算和操作也是模糊的。
- 对数据的约束,包括完整性和安全性,是模糊的。
- 用户使用数据库的窗口—用户视图是模糊的。



- 数据的一致性和冗余性的定义也是模糊的。

精确数据可以看成是模糊数据的特例。

## 2. 面向应用领域的数据库技术

数据库技术被应用到特定的领域中,出现了工程数据库、地理数据库、统计数据库、科学数据库、空间数据库等多种数据库,使数据库领域中新的技术内容层出不穷。

(1) 数据仓库。传统的数据库技术是以单一的数据资源为中心,进行各种操作型处理。操作型处理也叫事务处理,是指对数据库联机的日常操作,通常是对一个或一组记录的查询和修改,主要是为企业的特定应用服务的,人们关心的是响应时间、数据的安全性和完整性。分析型处理则用于管理人员的决策分析。例如:DSS、EIS 和 多维分析等,经常要访问大量的历史数据。于是,数据库由旧的操作型环境发展为一种新环境——体系化环境。体系化环境由操作型环境和分析型环境(数据仓库级、部门级、个人级)构成。数据仓库是体系化环境的核心,它是建立决策支持系统(DSS)的基础。

W.H.Inmon 给数据仓库作出了如下定义:数据仓库是面向主题的、集成的、稳定的、不同时间的数据集,用以支持经营管理中的决策制定过程。面向主题、集成、稳定和随时间变化是数据仓库四个最主要的特征。

有了数据就如同有了矿藏,而要从大量数据中获得决策所需的数据就如同开采矿藏一样,必须要有工具。分析工具是数据仓库系统的重要组成部分,这包括联机分析处理技术和工具、数据挖掘技术和工具。

技术的进步,不懈的努力使人们终于找到了基于数据库技术 DSS 的解决方案,这就是:DW+OLAP+DM→DSS 的可行方案。

数据仓库、OLAP 和数据挖掘是作为三种独立的信息处理技术出现的。数据仓库用于数据的存储和组织,OLAP 集中于数据的分析,数据挖掘则致力于知识的自动发现。它们都可以分别应用到信息系统的设计和实现中,以提高相应部分的处理能力。但是,由于这三种技术内在的联系性和互补性,将它们结合起来即是一种新的 DSS 构架。

(2) 工程数据库(Engineering Data Base)。工程数据库是一种能存储和管理各种工程图形,并能为工程设计提供各种服务的数据库。它适用于 CAD/CAM、计算机集成制造(CIM)等通称为 CAX 的工程应用领域。工程数据库针对工程应用领域的需求,对工程对象进行处理,并提供相应的管理功能及良好的设计环境。

工程数据库管理系统是用于支持工程数据库的数据库管理系统,主要应具有以下功能:

- 支持复杂多样的工程数据的存储和集成管理。
- 支持复杂对象(如图形数据)的表示和处理。
- 支持变长结构数据实体的处理。
- 支持多种工程应用程序。
- 支持模式的动态修改和扩展。
- 支持设计过程中多个不同数据库版本的存储和管理。
- 支持工程长事务和嵌套事务的处理和恢复。

在工程数据库的设计过程中,由于传统的数据模型难以满足 CAX 应用对数据模型的要求,需要运用当前数据库研究中的一些新的模型技术,如扩展的关系模型、语义模型、面向对象的数据模型。

(3) 统计数据库(Statistical Data Base)。统计数据是人类对现实社会各行各业、科技教

育、国情国力的大量调查数据。采用数据库技术实现对统计数据的管理，对于充分发挥统计信息的作用具有决定性的意义。

统计数据库是一种用来对统计数据进行存储、统计（如求数据的平均值、最大值、最小值、总和等）、分析的数据库系统。

统计数据库具有层次型特点，但并不完全是层次型结构。统计数据也有关系型特点，但关系型也不完全满足需要。概括起来，统计数据具有以下特点：

1) 多维性是统计数据的第一个特点，也是最基本的特点。

2) 统计数据是在一定时间（年度、月度、季度）期末产生大量数据，故入库时总是定时地大批量加载。经过各种条件下的查询以及一定的加工处理，通常又要输出一系列结果报表。这就是统计数据的“大进大出”特点。

3) 统计数据的时间属性是一个最基本的属性，任何统计量都离不开时间因素，而且经常需要研究时间序列值，所以统计数据又有时间向量性。

4) 随着用户对所关心问题的观察角度不同，统计数据查询出来后常有转置的要求。

(4) 空间数据库 (Spatial Data Base)。空间数据库是以描述空间位置和点、线、面、体特征的拓扑结构的位置数据及描述这些特征的性能的属性数据为对象的数据库。其中的位置数据为空间数据，属性数据为非空间数据。其中，空间数据是用于表示空间物体的位置、形状、大小和分布特征等信息的数据，用于描述所有二维、三维和多维分布的关于区域的信息，它不仅具有表示物体本身的空间位置及状态信息，还具有表示物体的空间关系的信息。非空间信息主要包含表示专题属性和质量描述数据，用于表示物体的本质特征，以区别地理实体，对地理物体进行语义定义。

由于传统数据库在空间数据的表示、存储和管理上存在许多问题，从而形成了空间数据库这门多学科交叉的数据库研究领域。目前的空间数据库成果大多数以地理信息系统的形式出现，主要应用于环境和资源管理、土地利用、城市规划、森林保护、人口调查、交通、税收、商业网络等领域的管理与决策。

空间数据库的目的是利用数据库技术实现空间数据的有效存储、管理和检索，为各种空间数据库用户使用。目前，空间数据库的研究主要集中于空间关系与数据结构的形式化定义，空间数据的表示与组织，空间数据查询语言，空间数据库管理系统。

## 习题

1. 计算机系统的发展经历了几个阶段？由哪几部分组成？
2. 计算机硬件系统由哪几部分组成？
3. 什么是系统软件？什么是应用软件？举例说明。
4. 什么是计算机网络？它有什么作用？
5. 计算机网络的分类有几种？拓扑结构如何？
6. 计算机网络的应用有哪些？
7. 简述数据库系统的三级模式结构和两层映像的功能。
8. 数据库设计需要几个阶段？