

第3章 电子商务技术基础

3.1 信息处理与计算机网络技术的发展

信息技术是电子商务的基础,信息技术的发展是电子商务发展的基础,也是电子商务的动力。可以说,没有信息技术,就没有今天的电子商务。信息技术的发展是一个加速发展的过程,了解和认识信息技术的发展过程,对于我们了解和掌握电子商务理论、了解信息技术和电子商务的未来发展大有好处。

3.1.1 信息处理技术的发展

人类自进入文明社会以来一直在从事信息处理工作。但是计算机的诞生改变了人们几千年的传统观念和信​​息处理方式,促使人们去进一步研究信息处理技术、信息系统和信息资源充分利用的规律性。20 世纪 50~60 年代由于数据处理技术的突破,导致了商用小型计算机的出现,以计算机为基础的信息技术(Information Technology, IT)全面进入到工商管理的各个领域,各种门类的信息系统发展分支不断涌现,IT 经历了从数值处理、数据处理、知识处理、智能处理和定量分析定性推理到网络资源共享等多个阶段,与之相对应的信息系统也经历了从定量分析到定性推理,到网络资源共享三个阶段。

1. 计算机的产生和以计算机为基础的信息系统形成

正如蒸汽机的出现导致了人类对能量转换规律的研究和工业革命的诞生一样,计算机的出现导致了人类对信息处理规律的研究和信息革命的诞生。

众所周知,1946 年由于战争和军事上的需要,世界上第一台电子计算机诞生了。当时由于技术发展水平所限,只能进行数值处理,导致了最初的计算机应用只限于军事科学、工程计算、数值统计、工业控制、信号处理等这样一些神秘而又狭小的领域。

20 世纪 50 年代由于计算机技术上的发展,计算机不但可以进行数值处理,而且还可以进行数据处理。数据处理技术的突破,导致了以 IBM、NCR 等为代表的一批高科技公司开始形成并向社会推出了商品化的小型计算机系统。商用小型计算机系统的推出,使得计算机的应用迅速跳出了军事和科学计算的神秘圈子,逐步向社会生活的各个方面(特别是工商管理领域)渗透。商用小型计算机的出现,把计算机的应用从单纯的数值运算扩大到适合于数据处理的广泛领域,为计算机和信息技术的繁荣和发展奠定了基础。

2. 计算机技术的发展和广泛应用

20 世纪 50 年代初期,人们开始用计算机进行商务数据处理。60 多年来,数据处理技术随着计算机软、硬件的发展而不断地发展,它大致经历了人工管理、文件系统管理、数据库系统管理和数据仓库等阶段,计算机的应用领域也得到了极大的拓展。

1) 数值处理

数值处理主要指使用计算机等电子设备对物理或数字信号进行运算和处理。数值处理是计算机应用早期最主要的应用特征。正是由于当时的计算机设备只能进行数值处理这样一个技术限制,所以决定了当时的计算机只能应用于军事计算、科学计算和工程计算等一个神秘而又狭小的圈子里。

2) 数据处理

20 世纪 50 年代末,由于计算机技术的发展和应用水平的不断深入,以文件管理技术和数据库技术为代表的数据处理技术出现了。数据处理是对数据的采集、存储、检索、加工、变换和传输。从数值处理到数据处理是计算机应用技术上的一个飞跃,数据处理技术的出现,将计算机应用的主战场推向了工商管理领域。人们开始尝试用电子技术来解决商务管理中的信息处理问题。当时人们把这一时期的系统称为电子数据处理系统(Electric Data Processing system,EDP)。EDP 是信息系统最早的发展分支,同时也是整个信息系统的基础。20 世纪 70 年代初,人们在 EDP 的基础上,提出了管理信息系统(Management Information System,MIS)的概念。较之传统的 EDP,MIS 在处理的方法、手段、技术等方面都有了长足的进步,其概念一经提出立刻受到学术界和企业界的普遍欢迎,MIS 开始大行其道,迅速地进入到企业管理的各个方面。

3) 办公自动化

20 世纪 80 年代初,由于以微处理芯片为基础的微型计算机和微机数字局域网络的出现,使得计算机和网络等硬件设备在价格上大幅度下降,在功能上大幅度提高。自 20 世纪 80 年代以来,各种各样的办公自动化系统开始在世界各地各种各样的办公室中迅速地推广开来。

4) 计算机辅助决策

随着 MIS 技术应用水平的不断深入,人们逐步开始考虑如何将信息系统技术引入到商务活动的最高层,即支持决策的功能上。1971 年,美国麻省理工学院的高瑞(G. A. Gorry)和斯科特·莫顿(M. Scott Morton)提出了决策支持系统(Decision Support System,DSS)的概念,把信息系统在管理领域的应用又推到了一个更高的阶段。

5) 人工智能

计算机在管理领域迅速发展的同时,一些有远见的学者将眼光投向了计算机应用的另一方面,即利用计算机来分析和模拟人脑信息处理和思维过程的人工智能(Artificial Intelligence)方面,并获得了很大的成功。AI 技术在 20 世纪 70~80 年代得到了迅猛发展,逐步产生了知识工程(Knowledge Engineering,KE)、专家系统(Expert System,ES)和基于知识处理的自动推理系统等发展分支。从数据处理到知识处理是计算机在应用技术上的又一次飞跃,它标志着计算机从处理量化问题向着处理定性化问题迈出了关键的一步,也是信息系统从概念、结构到方法、技术上的一次革命性的飞跃。

3. 计算机网络的发展和商务应用

计算机网络技术自 20 世纪 60 年代以来就一直在各种各样的 IS 中被广泛使用,但是传统的网络由于技术上的局限,在网络互联和资源共享方面一直存在着许多问题。直到 20 世纪 90 年代 Internet 的兴起才打破了这种局面,使得 IT 技术进入了网络处理时代,其最主要的技术特征就是网络互联、资源高度共享、时空观念的转变以及物理距离的消失等。这些技

术给企业管理信息系统和各类商务活动带来极大的影响,导致了 IT 应用领域革命性的变化。

作为 MIS 的一部分,人类很早就开始了 IT 在商贸交易领域的应用。例如早在 20 世纪 60~70 年代美国军方和运输部门就利用网络传递报文并从事商贸交易活动。但是由于早期的系统对技术标准和网络的要求比较高,应用的范围一直很有限。进入 20 世纪 90 年代以后,由于网络技术空前发展,情况才有了革命性的改变,人们开始尝试用电子信息技术在网络空间上展开商贸活动。计算机互联网络是电子商务产生的基础。

3.1.2 计算机网络及其种类

1. 什么是计算机网络

计算机网络是将地理上位置不同且功能独立的多个计算机系统通过通信线路连接在一起、由专门的网络操作系统进行管理、能实现资源共享的系统,如图 3.1 所示。简单地说,计算机网络是由两台以上计算机连在一起组成的“计算机群”,再加上相应“通信设施”而组成的综合系统。从而使众多的计算机可以方便地互相传递信息,共享硬件、软件、数据信息等资源。

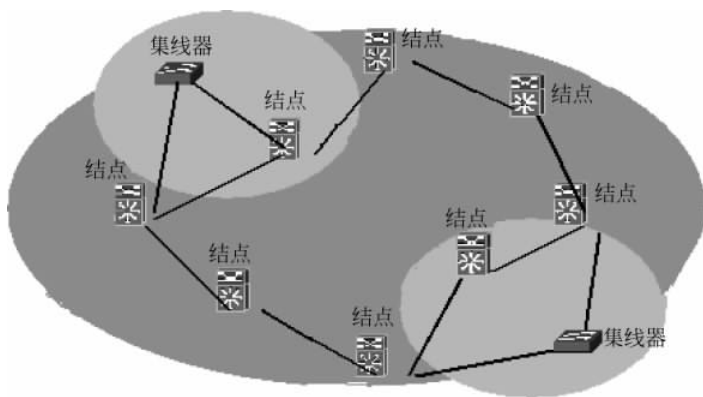


图 3.1 计算机网络示意图

计算机网络不仅可以是计算机的连接,还可以是外部设备(如打印机、硬盘、扫描仪和 CD-ROM 驱动器等)及其他设备的互连。

2. 计算机网络的功能

1) 数据通信

数据通信是计算机网络的基本功能之一,它实现网络系统的信息交换和信息的集中与分散处理。比如说一家公司,它有生产部、仓储部、市场部、财务部等很多地理位置上并不在一起的部门和分公司,各个业务部门需要随时知道其他部门的各种数据,这就需要依托网络才能实现数据通信。

2) 资源共享

资源共享包括共享软件、硬件和数据资源,是计算机网络最具吸引力的功能。所有在计算机网络上的硬件、软件和数据资源可以为所有联网的计算机或用户使用,这就是资源共

享。比如,一个财务系统中某个财务人员输入一个凭证后并记账后,其他财务人员应能在自己的办公计算机上查询账簿情况,实际上,账簿数据应为企业相关财务人员和企业有关领导所共享。

可以在网络上共享的资源除了打印机之外,硬盘、光盘、绘图仪、扫描仪以及各类软件、文本和各种信息资源等也都可以共享。在网络中共享资源既节省了大量的投资和开支,又便于集中管理。

3) 提高计算机系统的可靠性和可用性

提高可靠性表现在网络中的各台计算机可以通过网络彼此互为后备机。提高可用性是指网络中某台计算机负担过重时,网络可将新的任务交给网络中较空闲的计算机。

4) 易于进行分布式处理

对于较大型作业,可通过一定的算法,将作业分解交给不同的计算机,均衡使用网络资源,实现分布式处理的目的。

计算机网络是现代计算机技术和通信技术密切结合的产物,说得准确一些,计算机网络的本质就是利用通信设备和通信线路把不在一起的计算机等设备相互连起来,用相应的软件实现资源共享和信息交换的系统。

早期的网络还只是指一个单位的几台计算机用一根电缆连在一起,实现局部的资源共享和信息交换。今天的网络,是把世界上的上百个国家大小几千万台的计算机连为一体,形成硕大的像蜘蛛网一样的“怪物”,在全世界范围内实现全方位的资源共享和信息交换。这就是国际互联网(Internet),也称为因特网。

3. 计算机网络的产生与发展

早期的计算机都是独立的设备,每台计算机都自主工作,独立于其他计算机。不久人们发现这种方式既不高效,也不经济。人们需要一种新的方案来成功解决存在的问题,即设备和资源的重复建设;不能高效率地进行通信。

解决问题的方法就是构建计算机网络,实现数据通信和资源共享。计算机网络是计算机技术与通信技术逐步发展并相互结合的产物,计算机网络的发展历史不长,但发展速度很快。在 40 多年的时间里,其演变过程大致可概括为三个阶段:具有通信功能的单机系统阶段;以共享资源为主的计算机网络阶段;以局域网及其互连为主要支撑环境的分布式计算阶段。

1) 具有通信功能的单机系统阶段

第一代计算机网络是以单个计算机为中心具有通信功能的远程联机系统,如图 3.2 所示。早期的计算机系统是高度集中的,所连接的多个终端必须紧接着主计算机。所谓终端,是指一台计算机的外部设备,包括 CRT 控制器和键盘,无 CPU 和内存。所有数据以批处理或/和分时处理方式进行。20 世纪 50 年代中后期诞生了远程终端技术,将地理上分散的多个终端通过通信线路连接到一台中心计算机上,这样就出现了第一代计算机网络。第一代计算机网络系统又称终端-计算机网络,是早期计算机网的主要形式。20 世纪 50 年代初,美国建立的半自动地面防空系统 SAGE 就是将远距离的雷达和其他测量控制设备的信息,通过通信线路汇集到一台中心计算机进行集中处理,从而首次实现了计算机技术与通信技术的结合。第一代计算机网络的典型应用是由一台计算机和全美范围内 2000 多个终端组成的飞机订票系统。

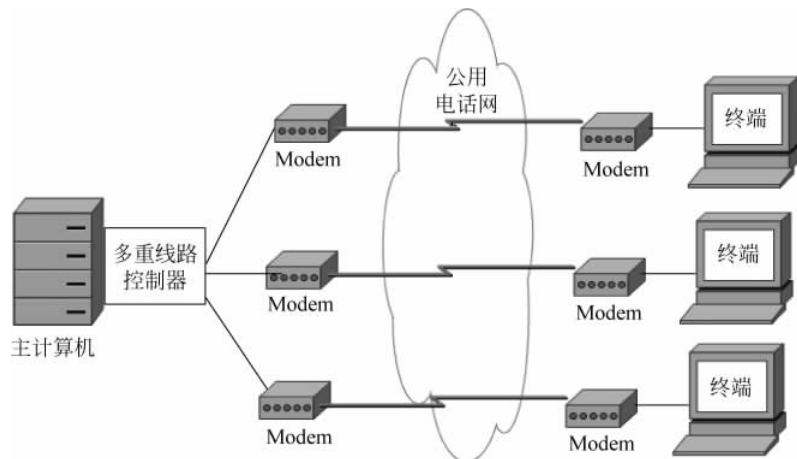


图 3.2 第一代计算机网络（远程联机系统）

在上述简单的“终端—通信线路—计算机”系统中，中央计算机负担较重，既要进行数据处理，又要承担通信控制，为了减轻主机负担，20 世纪 60 年代研制出了通信控制处理机（CCP）或叫前端处理机（FEP）专门负责通信控制，此外，在终端聚集处设置多路器或集中器，用低速线路将各终端汇集到集中器，再通过高速线路与计算机相连，如图 3.3 所示。20 世纪 60 年代初，此网络在军事、银行、铁路、民航和教育等部门都有应用。

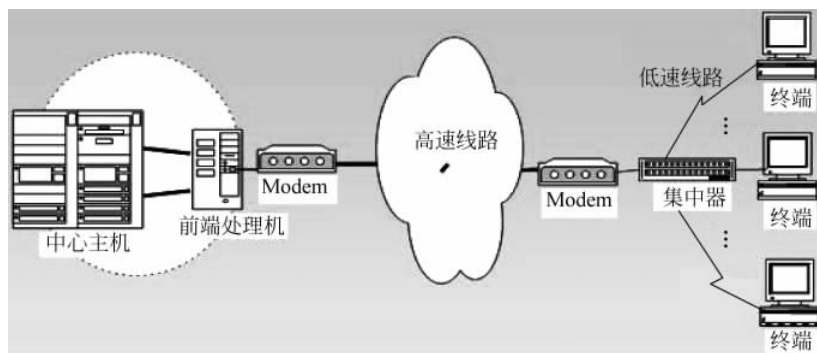


图 3.3 带前置处理机的远程联机系统

2) 计算机—计算机网络

20 世纪 60 年代中期，出现了由若干个计算机互连的系统，开创了“计算机—计算机”通信的时代，并呈现出多处理中心的特点，即利用通信线路将多台计算机连接起来，实现了计算机之间的通信，如图 3.4 所示。20 世纪 60 年代后期，美国国防部高级研究计划局所研制的 ARPANET 网是该网络的典型代表。它的主要目标是借助于通信系统，使网内各计算机系统间能够共享资源。ARPANET 是一个成功的系

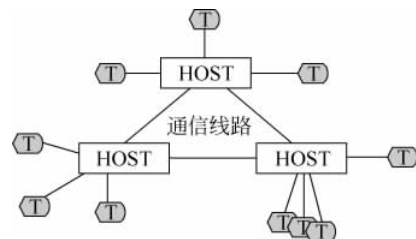


图 3.4 以资源为主要目的的计算机网络阶段

统,它在概念、结构和网络设计方面都为今后计算机网络的发展奠定了基础。

在 ARPA 网中,将协议按功能分成了若干层次,如何分层,以及各层中具体采用的协议的总和,称为网络体系结构,体系结构是个抽象的概念,其具体实现是通过特定的硬件和软件来完成的。

第二代计算机网络是具有统一的网络体系结构,并遵循国际标准的开放式和标准化的网络。ISO 在 1984 年颁布了 OSI/RM,该模型分为七个层次,也称为 OSI 七层模型,公认为新一代计算机网络体系结构的基础,为普及局域网奠定了基础。

3) 分布式计算阶段

自 20 世纪 70 年代开始,随着大规模集成电路技术和计算机技术的飞速发展,硬件价格急剧下降,微机广泛应用,局域网技术得到迅速发展。20 世纪 80 年代后,为了适应办公自动化的需要,各机关、企业迫切要求将自己拥有的为数众多的微机、工作站、小型机等连接起来,从而达到资源共享和互相传递信息等目的。在这种背景下,局域网技术发展呈日新月异之势。

局域网的发展也导致计算模式的变革。早期的计算机网络是以主计算机为中心的,计算机网络控制和管理功能都是集中式的,也称为集中式计算机模式。随着个人计算机(PC)功能的增强,用户一个人就可在微机上完成所需要的作业,PC 方式呈现出的计算机能力已发展成为独立的平台,这就导致了一种新的计算结构—分布式计算模式的诞生,如图 3.5 所示。

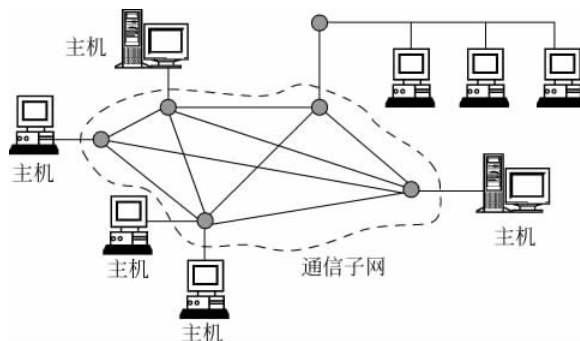


图 3.5 分布式计算机网络

目前计算机网络的发展正处于分布式计算阶段。这一阶段计算机网络发展的特点是:互连、高速、智能与更为广泛的应用,整个网络就像一个对用户透明的大的计算机系统,发展为以 Internet 为代表的互联网。

4. OSI 层次模型

在计算机网络中,不同的计算机拥有不同的操作系统,使用不同的软件,为保证彼此之间正确地交流信息,在通信时必须遵守一定的约定和规程,这些约定和规程就是网络协议。网络协议表示网络相同部件间传送数据时应遵循的一些规程或统一规定。不同的协议描述了不同的技术规范,它们使计算机之间能够进行信息交换,而不管计算机的种类以及连接所采用的技术是否相同。网络协议对网络系统的性能有关键性影响,遵循标准协议的软件能使不同系统间进行通信。常用的网络协议有 IPX/SPX、TCP/IP、NetBEUI 等,TCP/IP 是

Internet 使用的协议,允许在不同结构和操作系统的计算机网络间传送数据。

设计计算机网络结构非常复杂,为简化其复杂程度,通常把计算机网络功能分成若干个层次,每一层次建立在下层之上完成一定的功能,同时为其上层服务。层和协议的集合构成了网络体系结构。由于计算机网络技术的发展十分迅速,再加上不同利益集团之间的商业竞争,目前世界上存在着多种网络体系结构,不同类型的计算机网络同时存在,当然,许多协议之间有很大的兼容性。比较流行和有名的网络协议有国际标准组织(ISO)提出的“开放系统互连参考模型(OSI/RM)”、美国国防部提出的“TCP/IP 协议族”、国际电报电话咨询委员会(CCITT)就公用分组交换网制定的“x. 25 协议”、电气和电子工程师协会(IEEE)提出的“IEEE 802 标准组”等。

1) OSI 参考模型简介

1984 年 10 月 15 日,国际标准化组织(ISO)公布了 OSI 参考模型,这是指导信息处理系统互连、互通和协作的国际标准。OSI 参考模型定义了一个计算机网络功能的 7 层模型,由上至下分别是应用层、表示层、会话层、传输层、网络层、数据链路层和物理层。各层的基本功能如下:

- (1) 物理层——利用物理传输介质为数据链路层提供物理连接,以透明地传输比特流。
- (2) 数据链路层——在通信的实体之间建立数据链路连接,负责在相邻两个结点的链路上进行二进制数据流的传送,并进行差错监测和流量控制,保证信息能到达目的地。
- (3) 网络层——使用分组网、线路交换网、电话网等各种通信网,与通信对方系统之间建立通信路径,进行中断、路由选择管理。
- (4) 传输层——向用户提供可靠的端到端(源主机到目的主机)的数据传输服务。
- (5) 会话层——组织和同步两个通信的会话服务用户之间的对话,并管理数据的交换。
- (6) 表示层——解决在两个通信系统中交换信息时,不同的数据格式的编码之间的转换。
- (7) 应用层——负责向用户提供各种网络应用服务,如文件传输、电子邮件、远程访问等。

局域网协议标准中,目前用得较多的是 IEEE 802.3 和 IEEE 802.5。IEEE 802.3 为以太(Ethernet)网标准,IEEE 802.5 为令牌环(Token Ring)网标准。

5. 计算机网络的分类

计算机网络的分类方式有很多种,可以按地理范围、拓扑结构、传输速率和传输介质等分类。

1) 按地理范围分类

按地理范围可以分为局域网、城域网和广域网。局域网(Local Area Network, LAN)是地理范围一般几百米到 10 千米之内,小范围内的连网。如一个建筑物内、一个学校内、一个工厂的厂区内等。局域网的组建简单、灵活,使用方便。城域网(Metropolitan Area Network, MAN)地理范围可从几万米到近十万里,可覆盖一个城市或地区,是一种中等形式的网络。广域网(Wide Area Network, WAN)地理范围一般在几百万公里左右,属于大范围连网。如几个城市,一个或几个国家,是网络系统中的最大型的网络,能实现大范围的资源共享,如国际性的 Internet 网络。

2) 按传输速率分类

按传输速率可以分为高速网和低速网。网络的传输速率有快有慢,传输速率快的称高

速网,传输速率慢的称低速网。传输速率的单位是 b/s(每秒比特数,英文缩写为 bps)。一般将传输速率在 Kb/s~Mb/s 范围的网络称低速网,在 Mb/s~Gb/s 范围内的网称高速网。也可以将 Kb/s 网称为低速网,将 Mb/s 网称为中速网,将 Gb/s 网称为高速网。

网络的传输速率与网络的带宽有直接关系。带宽(Bandwidth)指的是电子电路中存在一个固有通频带,是指电路可以保持稳定工作的频率范围,即传输信道的宽度。带宽的单位是 Hz(赫兹)。按照传输信道的宽度可分为窄带网和宽带网。一般将 kHz~MHz 带宽的网称为窄带网,将 MHz~GHz 的网称为宽带网,也可以将 kHz 带宽的网称窄带网,将 MHz 带宽的网称中带网,将 GHz 带宽的网称宽带网。通常情况下,高速网就是宽带网,低速网就是窄带网。

3) 按传输介质分类

按传输介质可以分为有线网和无线网。传输介质是指数据传输系统中发送装置和接收装置间的物理媒体,按其物理形态可以划分为有线和无线两大类。传输介质采用有线介质连接的网络称为有线网,常用的有线传输介质有双绞线、同轴电缆和光导纤维。采用无线介质连接的网络称为无线网。目前无线网主要采用三种技术,即微波通信、红外线通信和激光通信。这三种技术都是以大气为介质的。其中微波通信用途最广,目前的卫星网就是一种特殊形式的微波通信,它利用地球同步卫星作中继站来转发微波信号,一个同步卫星可以覆盖地球的三分之一以上表面,三个同步卫星就可以覆盖地球上全部通信区域。

4) 按拓扑结构分类

按拓扑结构可以分为总线拓扑结构、星型拓扑结构、环型拓扑结构和树型拓扑结构等几种。连接在计算机网络上的计算机、大容量的外存、高速打印机等设备均可看作是网络上的一个结点,也称为工作站。计算机网络的物理连接形式叫做网络的物理拓扑结构,常用的拓扑结构有总线型、星型、环型等。

(1) 总线拓扑结构是一种共享通路的物理结构,普遍用于局域网的连接。这种结构中总线具有信息的双向传输功能,用于网络中信息的传递,其他设备以平行的关系连接在总线上,各个结点共用一个总线作为数据通路。

(2) 星型拓扑结构是一种以中央结点为中心,把若干外围结点连接起来的辐射式互连结构。这种结构适用于局域网,特别是近年来连接的局域网大都采用这种连接方式。这种连接方式以双绞线或同轴电缆作连接线路。

(3) 环型拓扑结构是将网络结点连接成闭合结构,信号顺着从一个方向从一台设备传到另一台设备,每一台设备都配有一个收发器,信息在每台设备上的延时时间是固定的。这种结构特别适用于实时控制的局域网系统。

(4) 树型拓扑结构就像一棵“根”朝上的树,与总线拓扑结构相比,主要区别在于总线拓扑结构中没有“根”。这种拓扑结构的网络一般采用同轴电缆,用于军事单位、政府部门等上、下界限相当严格和层次分明的部门。

3.1.3 计算机组网技术和网络互联技术

所谓网络互联就是利用网络互联设备,将两个或者两个以上具有独立自治能力的计算机网络连接起来,通过数据通信,扩大资源共享和信息交流的范围,以容纳更多的用户。20 世纪 90 年代以来,局域网迅速发展并被广泛地应用,许多单位和部门都建立了局域网,网络

的应用和信息的共享促进了网络向外延伸的需求。网络互联成了 20 世纪 90 年代计算机网络发展的标志,越来越多的人开始意识到,如果没有网络互联技术的支持,用于信息传输的计算机网络也会形成一个个“信息孤岛”,网络互联是计算机网络发展到一定阶段的必然结果。

计算机网络主要由网络硬件和网络软件两部分组成。通常组建局域网需要的网络硬件主要是服务器、网络工作站、网络接口设备(如网卡)、集线器(HUB)及传输介质等。

为了实现不同网络之间的数据传输和资源共享,通常用网络互联设备将若干个网络相互连接成更大的网络,这种网络的典型代表就是互联网。在互联网中,组成互联网的单个网络称为子网,用于连接子网的设备叫做中间系统。中间系统的作用是在不改变原来网络结构和服务功能的条件下,协调各个网络,使得网络的通信得以实现。事实上中间系统可以是一个单独的设备,也可以是一个网络。这些设备主要是中继器、网桥、路由器和网关。

1. 网络接口设备

1) 调制解调器

调制解调器(MODEM)即调制器和解调器的总称,俗称“猫”,它是利用电话线在用户计算机与 Internet 服务提供商的拨号服务器之间建立网络连接的一种设备。其功能是对信号进行调制和解调,将发送端计算机的数字信号转换成模拟信号称为“调制”,将模拟信号转换为数字信号,送入接收端计算机,称为“解调”。

调制解调器按调制方法分类,可分为调幅、调频、调相;按传输速率分类,可分为低速 MODEM(传输速率在 9600b/s 以下)、中速 MODEM、高速 MODEM(传输速率超过 19.2kb/s;按操作状态分类,可分为同步 MODEM 和异步 MODEM;按产品外形分类,可分为外置式和内置式。内置式调制解调器是一块与计算机内部的 ISA 或 PCI 插槽相匹配的电路板。内置调制解调器可能需要跳线和配置参数。外置式调制解调器是一台使用串行端口与计算机相连的设备。

调制解调器与计算机的硬连接完成后,还需安装驱动程序。驱动程序是硬件设备与操作系统之间起通信作用的软件程序,通常由生产商提供。这时操作系统能识别调制解调器,并设置其 IRQ、I/O 地址和 COM 端口,最后还需将调制解调器配制成上网并保持连接。

2) 网络适配器(网卡)

组建一个计算机网络时,通常要在每台计算机上安装一块专门的接口卡,然后将接口卡与通信电缆相连并运行相应的驱动程序,网络中的计算机之间才能进行相互通信。这样的一块接口卡就是网络适配器,俗称网卡。每一块网卡上都存储有一个物理地址,它是每块网卡的编号,称为 MAC 地址,MAC 地址是唯一的。

网卡的种类非常多,按照不同的标准,可以作不同的分类。最常见的是按传输速率、总线接口标准和连接器接口标准进行分类。

(1) 按传输速率分,网卡可分为 10Mb/s 网卡、100Mb/s 网卡、10M/100Mb/s 自适应网卡和 1000Mb/s 网卡。10Mb/s 和 100Mb/s 网卡是指网卡能够提供高达 10Mb/s 和 100Mb/s 的传输速率,可以比较完美地实现多媒体信号的传输。不过,目前只提供单一速率的 100Mb/s 网卡也比较少见,常见的是 10M/100Mb/s 自适应网卡是目前用得最广泛的一种网卡,所谓自适应网卡,是指网卡具有一定的智能,可以与远端网络设备(集线器或交换机)自动协商,以确定当前可以使用的速率。自适应的优点在于不需要用户设定,自动以最

高速率连接到远端网络设备上。1000Mb/s 网卡的传输速率可以高达 1000Mb/s,能满足高清晰视频数据传输的要求,但是,由于其本身的价格因素以及交换机和传输介质的高昂价格,使得千兆位网卡仍只用于服务器端。

(2) 按总线接口分类,可分为 ISA 网卡、PCI 网卡及专门用于笔记本电脑的 PCMCIA 网卡。ISA 总线的网卡多为 10Mb/s,常用于早期的 486DX 或低档的 Pentium 计算机。由于 ISA 总线传输速率低,且不能很好地支持即插即用,并需要人工设置中断和 I/O 地址,所以此类网卡已很少使用。

10M/100Mb/s 自适应网卡几乎是清一色的 PCI 总线,支持即插即用,而且由系统自动分配中断和 I/O 地址,从而使网卡的安装简单轻松。

PCMCIA 网卡也称 PC 卡,专门用于笔记本电脑,即插即用,并支持热插拔。

另外,USB 网络适配器也是一种新型网卡,通用串行总线(Universal Serial Bus,USB)作为一种新型的总线技术,已经被广泛应用于鼠标、键盘、打印机、扫描仪、Modem、音箱等各种各样的设备。由于其传输速率远远大于传统的并行口和串行口,且设备安装简单并且支持热插拔。USB 设备一旦接入,就能够立即被计算机所承认,当装入所需要的驱动程序后,不必重新启动系统就可立即投入使用。当不再需要某台设备时,可以随时将其拔除。

(3) 按照与连接的传输介质相连接的接口分,有 RJ-45 接口网卡、AUI 接口网卡、BNC 接口网卡和 FX 接口网卡。不同的接口适应于不同传输介质的网络,RJ-45 接口适用于以双绞线为传输介质的网络,AUI 接口网卡适用于以粗缆为传输介质的网络,BNC 接口网卡适用于以细缆为传输介质的网络,FX 接口适用于以光纤为传输介质的网络。

2. 网络传输介质

网络传输介质是物理上把计算机相互连接起来的介质,是网络中传输数据、信号的实体。在局域网中常见的网络传输介质有以下四种,即双绞线、同轴电缆、光缆、无线介质。其中,双绞线是经常使用的传输介质,它一般用于星型网络中,同轴电缆一般用于总线型网络,光缆一般用于主干网的连接。

1) 双绞线(TP)

目前,在以太网中大量使用双绞线作为主要传输介质。双绞线是一种常用做电信电缆的铜导线。由于铜是非常好的导体,所以它对电磁信号无任何限制。当两根非常接近的铜线都在传导电信号时,就会出现一定的电磁干扰,这种干扰称作串扰。此外,电磁现象无所不在,这便导致 TP 传输或接收到了一些从其他地方传来干扰信号。如果把电极相反的一根铜线相互绞在一起,可以减少串扰以及信号放射程度,每一根缠绕着的导线在导电时,发出的电磁辐射被绞合的另一根线上发出的电磁辐射所抵消,随着单位长度电缆中所缠绕线的对数的增加,防止串扰的能力也增加。

双绞线由两根 22~26 号绝缘铜线相互缠绕而成。而一对或多对双绞线安置在一个套桶中时,便形成了双绞线电缆。有两种类型的 TP 电缆:非屏蔽双绞线电缆(unshielded twisted-pair,UTP)和屏蔽双绞线电缆(shielded twisted-pair,STP)。目前组网中常用的所谓 5 类线是使用了特殊的绝缘材料的非屏蔽双绞线,最高传输频率为 100MHz,可用于语音和最高传输速度为 100Mbps 的数据传输。

在基于 UTP 的网络中,每台计算机都用一根长电缆将自身与中心集线器连接在一起,每条电缆通过两端安装的 RJ-45 连接器(水晶头)与网卡和集线器(或交换机)相连,最大网

线长度为 100m。如果要加大网络范围,在两段双绞线电缆间可安装中继器,但最多可安装 4 个中继器,使网络最大范围达到 500m。在线缆两头安装 RJ-45 连接器时,应保证两个 RJ-45 连接器中每条导线的排线顺序固定、统一。

双绞线分为直通线和交叉线,如图 3.6 所示。直通线适用于计算机到集线设备(如集线器,交换机)的连接。直接连接两台计算机或直接连接两台集线设备要用交叉线。

2) 同轴电缆

同轴电缆有一条被包围在绝缘层中的铜芯(单股或多股)。它的特点是抗干扰能力好,传输数据稳定,价格也便宜,同样被广泛使用,如闭路电视线等。信号在这条铜芯中传输,包围绝缘层的是另一种由金属薄膜或者金属线做成的导线,这条导线的长度与电缆同长,因而被称作同轴电缆。这种电缆具有两条不同的物理通道,一条用来传输信号,一条用作地线,它能减少干扰,外面的屏蔽物也是电缆的护套,如图 3.7 所示。

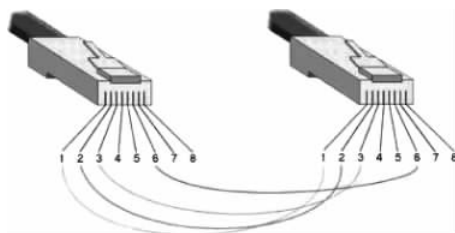


图 3.6 双绞线示意图

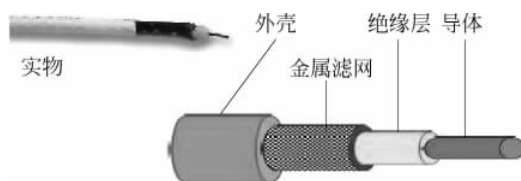


图 3.7 同轴电缆示意图

根据电缆的直径分,同轴电缆主要有两种基本类型,即细同轴电缆(细缆)和粗同轴电缆(粗缆)。细缆直径约为 1/4in,比较柔韧,它使用在 10Base2 以太网中,最大传输距离为 185m,其阻抗为 50Ω。粗缆直径约为 1/2in,因其铜芯较粗,屏蔽层较坚硬,因此安装较难,费用也较高,但它的传输距离约为 500m。

现在计算机局域网中一般都使用细缆组网。细缆一般用于总线型网络布线连接,利用 T 型 BNC 接口连接器连接 BNC 接口网卡,同轴电缆的两端需安装 50Ω 终端电阻器。细缆网络每段干线长度最大为 185m,每段干线最多可接入 30 个用户。细缆安装较容易,而且造价较低,但因受网络布线结构的限制,其日常维护不是很方便,一旦一个用户出故障,便会影响其他用户的正常工作。粗缆适用于较大局域网的网络干线,布线距离较长,可靠性较好。用户通常采用外部收发器与网络干线连接。粗缆局域网中每段长度可达 500m。用粗缆组建的局域网虽然各项性能较高,具有较大的传输距离,但是网络安装、维护等方面比较困难,且造价较高。

3) 光缆

光缆又称光纤电缆,是由一组光导纤维组成的、用来传播光束的、细小而柔韧的传输介质。与其他传输介质相比较,光缆的电磁绝缘性能好,信号衰变小,频带较宽,传输距离较大,是一种较新型的、数据传输速度较高、价格也较高的传输介质,主要用于传输距离较长的主干网的连接中。光缆被一层称作包层的玻璃包住,最外一层用坚硬的外壳包住。中心提供光通路(或者称波导),而包层由多层反射玻璃构成,玻璃层设计可以将光折射到中间芯上。

光缆通信由光发送机产生光束,将电信号转变为光信号,再把光信号导入光纤,在光缆

的另一端由光接收机接收光纤上传来的光信号,并将它转变成电信号,经解码后再处理。光缆的传输距离远、传输速度快,是计算机网络中传输介质的佼佼者。

现在有两种光缆,即单模光缆和多模光缆。单模光缆的纤芯直径很小,只提供一条光通道。在给定的工作波长上只能以单一模式传输,传输频带宽,传输容量大。多模光缆提供多条通道,是在给定的工作波长上,能以多个模式同时传输的光纤。

3. 无线传输

无线技术逐渐发展成为一种时兴的网络技术。一个“无线”网络并不是完全不需要电缆,但结合了传统的有线网络通信的无线设备,通常被称为访问结点的收发器用来在无线设备与有线网络之间收发数据。

1) 微波

微波数据通信系统有两种形式,即地面(基于地球表面)系统和卫星系统。它们使用同样的频率,功能比较相似,但是,它们能力上有些差别。

(1) 地面微波。地面微波一般采用定向式抛物面形天线,这要求与其他地点之间的通路没有障碍或视线能及。

(2) 卫星微波。通过地球上的定向抛物面天线向地球同步卫星发送直线波束。

2) 红外线

红外连接采用发光二极管、激光二极管或者光电二极管来进行站与站之间的数据交换,通过使用位于红外频率波谱中的锥形或者线型光束来传输数据信号,将红外技术应用到无线局域网中,传输信息可以直接或经过墙面、天花板反射后被接收装置收到。然而,红外信号没有能力穿透墙壁和一些其他固体,也容易被强光源盖住。红外线传输技术可以分为点到点红外线系统和广播式红外线系统。

3) 激光

激光网络利用激光脉冲来表示数据信号进行工作。激光是一种视线技术,这种技术意味着在信号的路由和接收设备之间必须有一条无阻隔的通路,这也正是激光无线通信的缺点。

4. 中继器

同网络介质一样,中继器存在于 OSI 模型的第一层,即物理层中。当数据离开信源、进入网络时,是被转换为电脉冲或光脉冲的形式在网络中传输的,这些脉冲称为信号。当信号最先从一个传输站出发时,它们是清晰的且易于识别,但信号经过物理传输介质时,电缆越长,信号就越来越弱越发难于识别。例如,第 5 类双绞线以太网电缆的技术规范中规定信号沿网络传输的最大距离是 100m,如果一个信号传输的距离超过了这一长度,就不能保证网络接口卡能识别它。

中继器为此问题提供简单的解决办法,它可使网络覆盖的距离得到扩展。中继器接收到变弱的信号,将其去噪音并放大,再送回原路径,继续沿着网络传送,这样使网络可操作的距离得到扩展。中继器信号放大功能的另一个好处是可以增加网络结点数目,克服结点数目多造成的信号衰减的影响。

但一根电缆上使用的中继器数量有一定的限制,因为中继器会引起网络中数据传输的延迟,严重影响网络性能。在以太网中,网络中的中继器一般不应超过 4 个。

1) 集线器

集线器(HUB)是一种特殊的中继器,它将分散的用于连接网络设备的线路集中在一起,便于管理和维护。在集线器上有固定数目的端口,如8端口、12端口、16端口、24端口等。每个网络设备都可以通过电缆线连接到一个端口上。在集线器上增加或减少网络设备不影响网络运行,也可以利用它来互连LAN,甚至可以互连不同类型的网络。一般来说,当中继器用作星型拓扑结构的网络中心时,我们称其为集线器而不是中继器。

集线器的缺点是它不能过滤网络业务量。过滤是一个过程或设备,它以一定的特征,如源地址、目的地址或使用的网络协议来屏蔽网络信息,并根据已有的标准确定是将信息转发还是丢弃。在集线器上,数据到达一个端口,然后被发送到其他的所有端口。即经过集线器的数据将发向网络上的其他所有的局域网段,不论它是否需要去那儿。

集线器按带宽分类有10Mb/s、100Mb/s、10M/100Mb/s双速集线器三种。所谓10Mb/s集线器,是指该集线器中的所有端口均只能提供10Mb/s带宽。所谓双速集线器,是指该集线器中内置10Mb/s和100Mb/s两条内部总线,其传输速率可在10Mb/s和100Mb/s之间进行切换。目前,几乎所有的双速集线器均为自适应,每个端口都能自动判断与之相连接设备所能提供的连接速率,并自动调整至与之相适应的最高速率。

2) 网桥

如果一个网络的所有设备都仅仅是由一根电缆连接而成,或者网络的网段由集线器之类无过滤能力的设备连接而成,可能会有不止一个用户同时向网络发送数据。如果多个结点试图同时发送数据,那么就会发生冲突。解决网络上出现太多业务量及太多冲突的办法之一是使用网桥。

网桥又称为简单透明桥,它将两个网段连接在一起,并根据分组上的介质访问控制地址(MAC)实施流量过滤。当正确使用网桥时,可以确保减少拥塞。通过使用网桥可将网络分割成两个网段,减少网络上不必要的业务量并将冲突发生的可能性降至最低。

每个网桥都将其连接的两个网段的地址建立起一个地址表,称为路由表。数据沿网络介质传输时,网桥将数据携带的目的MAC地址与客观存在的表格中的MAC地址进行比较。如果处于同一网段,则不再把数据转发到网络的其他网段。否则就把数据转发到网络的所有其他网段中。这样,大大除去不必要的业务量。

3) 交换机

交换机是一种基于MAC地址识别,能完成封装转发数据包功能的网络设备。交换机拥有一条很高带宽的背部总线和内部交换矩阵,交换机所有的端口都挂接在这条背部总线上。控制电路收到数据包以后,处理端口会查找内存中的MAC地址(网卡的硬件地址)对照表以确定目的MAC的网卡挂接在哪个端口上,通过内部交换矩阵直接将数据迅速包传送到目的结点,而不是所有结点,当目的MAC不存在(或不能确定时)才将数据广播到所有的端口。

交换机作为一种廉价而且高效的集线设备,已经取代传统集线器在网络连接中的霸主地位。交换机(SWITCH)大多数是工作在OSI模型中第2层(数据链路层)的设备,它的作用是对封装数据包进行转发,在端口之间建立并行的连接,缩小冲突域,并隔离广播风暴。

另外,交换机还可以用来建立虚拟局域网(VLAN),将与交换机相连接的物理网络分割成多个逻辑网络,这可增加网络的性能与安全性。

4) 路由器

路由器用来连接不同的网络以及接入因特网。当路由器使用在局域网中时,被连接的网络称为子网,用于连接广域网时,一般是非相关的两个网络。

专用的路由器设备实际上是一些有特殊用途的计算机,这些路由器有微处理器,并运行自有的操作系统。个人计算机也可以通过配置当成路由器使用,只要操作系统支持 IP 或者 IPX 路由。

在路由器中保存着各种传输路径的相关数据——路径表(Routing Table),供路由选择时使用。路径表中保存着子网的标志信息、网上路由器的个数和下一个路由器的名字等内容,可以是由系统管理员固定设置好的,也可以由系统动态修改,可以由路由器自动调整,也可以由主机控制。路由器的主要作用是为不同网络之间的用户提供最佳的通信路径,使用最少时间算法或最优路径算法进行信息传递路径的调节。此外,路由器可进行数据包格式的转换,是不同协议网络间互连的必要设备。

路由器基于网络协议或第三层的信息在不同的网络间传递数据包、路由数据流量而提供端到端的路由。路由器具有在网上传递数据时选择最佳路径的能力,过量和广播流量问题可以通过使用路由器来解决,一般情况下路由器不向前传送广播帧。

5. 网络软件

网络的使用和维护要依赖于网络软件。网上使用的网络软件主要是指网络操作系统、网络数据库管理系统和网络应用软件。

1) 网络操作系统

在网络硬件提供数据传输能力的基础上,为网络用户管理共享资源、提供网络服务功能的系统软件被定义为网络操作系统。目前较流行的网络操作系统有 UNIX、Novell Netware、Windows NT 与 Linux 等。它们在技术、性能、功能方面各有所长,支持多种工作环境,支持多种网络协议,能够满足不同用户的需要,为网络的广泛应用奠定了良好的基础。

(1) UNIX 网络操作系统。

标准 UNIX 操作系统是一个交互式的分时系统,提供了一个支持程序开发全过程的基础和环境,可以支持 40 个终端用户。UNIX 系统是由美国电报电话公司(AT&T)下属的 Bell 实验室的两名程序员 K. 汤普逊(Ken Thompson)和 D. 里奇(Dennis Ritchie)于 1969—1970 年研制出来的。UNIX 问世以来十分流行,它运行在从微机到大型机等各种具有不同处理能力的机器上。近年来,几乎所有的 16 位机、32 位微型计算机都竞相移植 UNIX。

UNIX 的主要特点是技术成熟、可靠性高,其结构简练,便于移植。许多 UNIX 主机和服务器都是每天 24 小时,每年 365 天不间断运行。UNIX 系统是世界上唯一能在笔记本电脑、PC、工作站直至巨型机上运行的操作系统,而且能在所有体系结构上运行。

网络功能强大是其又一特点。作为 Internet 网络技术基础和异种机连接重要手段的 TCP/IP 协议就是在 UNIX 上开发和发展起来的。TCP/IP 是所有 UNIX 系统不可分割的组成部分。此外,UNIX 还支持所有需用的网络通信协议,包括 NFS、DCE、IPX/SPX、SLIP、PPP 等,使得 UNIX 系统能方便地与已有的主机系统以及各种广域网和局域网相连接,这也是 UNIX 具有出色的互操作性的根本原因。

(2) Linux 操作系统。

Linux 操作系统是一种类 UNIX 的操作系统,Linux 操作系统是一种自由的、没有版权

限制的免费软件。它在 20 世纪 90 年代初由芬兰赫尔辛基大学的学生 Linus Torvalds 开发出来,并公布了它的源程序。来自全世界的志愿者都参加了 Linux 的修改工作。领先的系统应用供应商有 Caldera 公司和 Red Hat Software 公司。

(3) Windows NT 操作系统。

Windows NT 是目前最流行的网络操作系统之一,它具有强大的功能。Windows NT 的主要优点在于其技术较为先进,与 Windows 有统一的界面,且两者结合紧密。这样用户使用起来非常直观方便。它能很好地兼容 Windows 的丰富的应用软件,也有利于鼓励软件厂商开发新的应用,因而能很好地利用 Windows 上的优势。Windows NT 拥有可伸缩的解决方案,完全排除操作系统的人为限制。

能够安全简单地访问 Internet 是 Windows NT 最引人注目的地方。它提供了对等的 Web 服务(Peer web Service,PWS)功能,使企业内部网的用户可以创建个人网页,向内部用户发布信息。Windows NT 还提供点对点通信协议(Peer to Peer Tunneling Protocol,PPTP)的支持,它使用户可以通过 Internet 远程访问企业内部网。

Windows NT 非常方便使用于 Internet 环境,它捆绑了 DNS、DHCP、Gopher、Web、FTP 以及 LPD 服务器。它的系统管理工具也值得称道。允许对网络数据输送在包的水平上进行监测,并可将监测数据进行存储以供分析。“任务管理器”提供了网络工作站上应用程序与进程的详尽信息,并可强行终止那些没有响应的任务。

Internet 信息服务器(Internet Information Server,IIS)是 Microsoft 公司的一种集成了多种 Internet 服务(WWW 服务、FTP 服务等)的服务器软件。IIS 和 Windows NT Server 紧密地集成在一起,可以很容易地构造 Web 站点,可以充分利用 Windows NT 的多种功能,其安全机制也以 Windows NT 的 NTFS 安全机制为基础,因此可以实现用 IIS 构建的 Web 站点的安全性。

由于微软公司在桌面操作系统上的长期垄断地位,很多应用程序开发商专门开发各种基于 Windows 系列的操作系统上的应用程序。因此,使用 Windows NT 操作系统能够更容易地得到各种服务软件,这是它得以流行的一个重要原因。

(4) Netware 操作系统。

Novell Netware 是 Internet 进入我国之前,最为流行的一种网络操作系统。它一开始是为 MS-DOS 网络设计的比较专用的文件服务器操作系统。它能很好地处理从客户工作站发出的远程 I/O 请求。但是,由于 20 世纪 90 年代计算系统逐渐由大型变为小型而且多数都转移到了客户机/服务器计算结构上,所以网络服务器的作用也已经随之发生了相应的变化。越来越多的公司都把网络服务器看作是一个平台,希望它能支持内部事务处理系统,这些系统原来是在小型机和大型机上运行的。在一个客户机/服务器环境中,网络服务器必须能向大型机那样管理多个使用大量资源的服务器进程,而且保证同样的完整性、安全性和可靠性。

为了适应新的形式,Novell 公司推出了新一代的智能网络平台——Netware 5。Netware 5 将继续发挥 Novell 公司在网络计算方面的优势,完全支持 IPX 和基于 IPX 的应用程序,同时又提供了运行纯粹 IP 网络的能力。它配备了先进的 Internet 服务器,支持多

种脚本语言,使用起来非常方便。

2) 网络数据库管理系统

网络数据库管理系统是一种可以将网上的各种形式的数据组织起来,科学、高效地进行存储、处理、传输和使用的系统软件。如 SQL Server、Oracle、Sysbase、Informix 等。

3) 网络应用软件

软件开发者根据网络用户的需要,用开发工具开发出来各种应用软件。例如,常见的在局域网环境中使用的 Office 办公套件、商品流转、收银台收款软件等。

3.1.4 局域网

1. 基本概念

根据网络地理覆盖范围的大小和互连距离划分,一般在 10 千米以内,属于一个部门或单位组建的小范围内,例如一所学校、一个公司甚至一幢建筑物等的计算机网络,就是局域网。局域网用来连接计算机、外部设备、终端和其他设备。它使人们高效地共享资源,如共享文件和打印机。局域网组建方便、使用灵活,是目前计算机网络发展中最活跃的分支。如图 3.8 所示,就是一个简单的局域网。

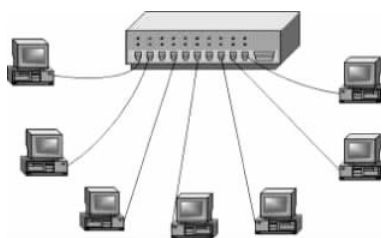


图 3.8 简单的局域网

局域网的性能主要由三个要素决定,即拓扑结构、传输介质和介质访问控制方式。以太网是当今最流行的一种局域网,是世界上第一个成熟的局域网产品。

(1) 拓扑构型就是网络的物理连接形式。如果不考虑实际网络的地理位置,把网络中的计算机看作一个结点,把通信线路看作一根连线,这就抽象出计算机网络的拓扑构型。局域网的拓扑构型主要有总线型、环型和星型三种。

(2) 传输介质是网络连接设备间的中间介质,也是信号传输的媒体。局域网中常见的传输介质有光纤、同轴电缆、双绞线几种。

(3) 介质访问控制方式是计算机网络上控制多个结点利用公共传输介质发送和接收数据的方式。

2. 总线型以太网

以太网是由 Xeros 公司开发的一种基带局域网技术,使用同轴电缆或光纤作为网络传输介质,采用载波多路访问和碰撞检测(CSMA/CD)机制,数据传输速率达到 10Mbps。虽然以太网是由 Xeros 公司早在 20 世纪 70 年代最先研制成功,但是如今以太网一词更多地被用来指各种采用 CSMA/CD 技术的局域网。

总线型拓扑结构采用单根数据传输线作为通信介质,所有的站点都通过相应的硬件接口直接连接到通信介质,而且能被所有其他的站点接受。以太网通常使用专门的网络接口卡或通过系统主电路板上的电路实现。以太网使用收发器与网络介质进行连接。收发器可以完成多种物理层功能,其中包括对网络碰撞进行检测。收发器可以作为独立的设备通过电缆与终端站连接,也可以直接被集成到终端站的网卡当中。如图 3.9 所示为总线型以太

网拓扑结构。

以太网采用广播机制,所有与网络连接的工作站都可以看到网络上传递的数据。通过查看包含在帧中的目标地址,确定是否进行接收或放弃。如果证明数据确实是发给自己的,工作站将会接收数据并传递给高层协议进行处理。

以太网采用 CSMA/CD 媒体访问机制,任何工作站都可以在任何时间访问网络。在发送数据之前,工作站首先需要侦听网络是否空闲,如果网络上没有任何数据传送,工作站就会把所要发送的信息投放到网络当中。

作为一种基于竞争机制的网络环境,由于没有任何集中式的管理措施,所以非常有可能出现多台工作站同时检测到网络处于空闲状态,进而同时向网络发送数据的情况。这时,发出的信息会相互碰撞而导致损坏。工作站必须随机等待一段时间之后,重新发送数据。

总线型网络结构中的结点为服务器或工作站,通信介质为同轴电缆。由于所有的结点共享一条公用的传输链路,所以一次只能由一个设备传输。这样就需要某种形式的访问控制策略,来决定下一次哪一个结点可以发送。一般情况下,总线型网络采用载波监听多路访问/冲突检测(CSMA/CD)控制策略。所有结点都直接连到一条主干电缆上,这条主干电缆就称为总线。该类结构没有关键性结点,任何一个结点可以通过主干电缆与连接到总线上的所有结点通信,这种结构的优点是电缆长度短,布线容易;结构简单,可靠性高;增加新结点时只须在总线的任何点接入,易于扩充。总线结构的缺点是故障检测需要在各个结点进行,故障诊断困难、隔离也困难,尤其是总线故障会引起整个网络的瘫痪。

总线型网络发送时,发送结点对报文进行分组,然后一次一个地址依次发送这些分组,有时要与其他工作站传来的分组交替地在通信介质上传输。当分组经过各结点时,目标结点将识别分组的地址,然后将属于自己的分组内容复制下来。

总线型拓扑结构在局域网中得到广泛的应用,具有布线容易、电缆用量小;可靠性高、易于扩充、易于安装的优点。也有一些局限性,如故障诊断困难、故障隔离困难等。

3. 令牌环网

令牌环是 IBM 公司于 20 世纪 80 年代初开发成功的一种网络技术。目前已经发展成为最为流行的局域网组网技术之一。该网的拓扑结构是传输媒体首尾连接的环,介质访问控制工具是令牌(Token),所以叫令牌环网。环上有多个站逐个与环相连,相邻站之间是一种点对点的链路,因此令牌环与广播方式的 Ethernet 不同,它是一种顺序向下一站广播的 LAN。令牌环所遵循的标准是 IEEE 802.5,它规定了三种操作速率,即 1Mb/s、4Mb/s 和 16Mb/s。开始时,UTP 电缆只能在 1Mb/s 的速率下操作,STP 电缆可操作在 4Mb/s 和 16Mb/s,现已有多家厂商的产品突破了这种限制。通常来说,令牌环网指的就是 IBM 公司的令牌环网和 IEEE 802.5 网络。

各结点形成闭合的环,信息在环中作单向流动,可实现环上任意两结点间的通信。环形结构的优点是电缆长度短、成本低。该结构的缺点是某一结点故障会引起全网故障且故障诊断涉及到每一个结点,故障诊断困难;若要扩充环的配置,就需要关掉部分已接入网中的结点,重新配置网络困难。

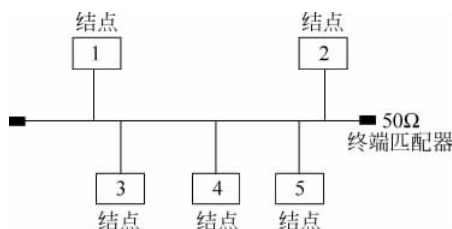


图 3.9 总线型以太网拓扑结构

1) 令牌环网的拓扑结构

令牌环的基本结构如图 3.10 所示,工作站以串行方式顺序相连,形成一个封闭的环路结构。数据顺序通过每一个工作站,直至到达数据的原发者才停止。构成令牌环物理结构的传输介质有屏蔽双绞线(STP)和无屏蔽双绞线(UTP)。

2) 令牌环网的网络运行原理

令牌环上传输的数据格式(帧)有两种:一种是令牌,另一种是常规帧;也即命令/数据帧。令牌是占有发送权的标志,占有令牌的站才能发送,常规帧用来发送数据或控制信息。获得令牌的结点可以向网络发送数据,其他结点只能复制和转发数据。

当环上的一个工作站希望发送帧时,必须首先等待令牌,一旦收到令牌,工作站便可启动发送帧。帧中包括接收站的地址,以标识哪一站应接收此帧。信息帧沿环行网络传递直到到达目标接收站,后者对所传递的信息进行复制,提交至下一站处理。

帧在环上传送时,不管帧是否是针对自己工作站的,所有工作站都进行转发,直到待回到帧的始发站。帧的目标接收者除转发帧外,应在自身站维持一个帧的副本,并通过在帧的尾部设置“响应比特”来指示已收到此副本,发送方可以通过检查返回的信息帧判断信息是否已经被目标接收方接收和复制。

工作站在发送完一帧后,应该释放令牌,以便出让给它站使用。如果接收到令牌的结点不需要发送任何数据,将会把接收到的令牌传递给网络中下一台终端站。每台工作站保留令牌的时间不得超过网络规定的最大时限,其他需要发送数据的工作站必须等待信息帧传递完毕,令牌帧重新释放之后才有机会获得令牌帧发送自己的信息。因此,采用令牌环技术的网络不会发生像以太网那样的信息碰撞。

与 CSMA/CD 网络不同,基于令牌传递技术的网络可以计算出任何一台终端站在能够传递数据之前所需要等待的最长时间。这一特点结合令牌环网本身所具有的较高的可靠性使令牌环网非常适合在能够对网络延时做出准确预测和需要健壮的网络运行能力的环境下使用。例如,生产厂房中的自动化系统多采用令牌环技术组网。

令牌环网采用了一整套错误检测和修复机制,用户可以指定令牌环网中的任何一台工作站作为主动监控器,为其他工作站提供时钟信息,同时执行各种环路维护功能。例如,当信息帧的发送设备失效时,将无法清除已经发出的信息帧,使信息帧在网络环路中持续传递下去,影响其他工作站正常发送信息,并有可能最终导致整个网络瘫痪。为了避免上述情况的发生,令牌环网中的主动监控器可以及时检测并清除网络中出现的错误帧,并自动生成新的令牌发送到网络当中。

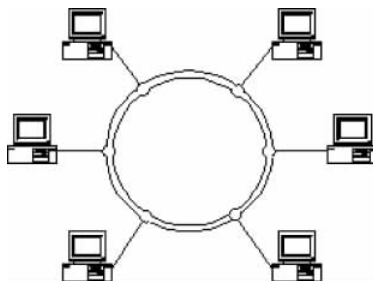


图 3.10 令牌环网的拓扑结构

3.1.5 广域网与计算机通信技术

随着远程办公室和居家工作者的迅猛发展,越来越需要在地理分散的多个地点之间建立安全高速的连接,广域网成为满足这一需要的重要手段。

1. 广域网基本概念

广域网(Wide Area Network, WAN)是在一个广泛地理范围内所建立的计算机通信

网,其范围可以超越城市和国家以至全球。因而对通信的要求及复杂性都比较高。

从广义上讲,广域网(WAN)是将远距离的网络和资源连接起来的系统。广域网通过电话线和卫星提供跨国或全球范围的联系。有区域性或全球性事务的大公司可以使用广域网进行网络互连,并从远距离媒体租用线路来提供系统间的全天候连接。都市网(MAN)从广义上讲也是一种广域网,通常为高速的光纤网络,在一个特定的范围内将局域网段,如校园、工业区等连接起来。都市网使用特殊的高速主干电缆(通常为光纤)直接连接服务器。最典型的广域网为 Internet,其分支如图 3.11 所示。

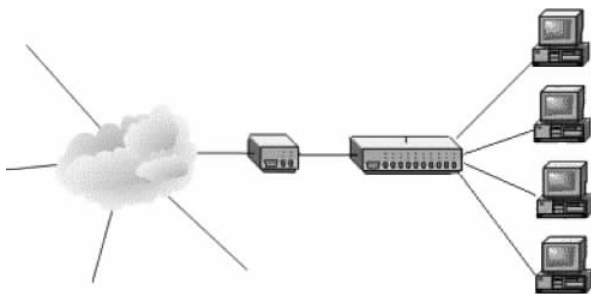


图 3.11 广域网某一分支的典型结构

WAN 由通信子网与资源子网两个部分组成,通信子网实际上是一个数据网,可以是一个专用网(交换网或非交换网)或一公用网(交换网);资源子系统是联在网上的各种计算机、终端、数据库等。这不仅指硬件,也包括软件和数据资源。

WAN 的实现都是按照一定的网路体系结构和相应的协议进行的。为了实现不同系统的互连和相互协同工作,必须建立开放系统互连。参考模型及相应的一系列国际标准协议对于 WAN 的实现、建立和应用有重要的指导作用。

广域网用于通信的传输装置,一般是由公司或电信部门提供的。互连主要采用公用网络和专用网络两种,如果连接的次数有限,要求不固定,通用性好,可选择公用数据网或增值网;如果连接次数很多,且要 24 小时畅通无阻,则采用专用网络为好。

在实际应用中,LAN 可与 WAN 互联,或通过 WAN 与位于其他地点的 WAN 互联,这时 LAN 就成为 WAN 上的一个端系统,各个局域网通过接入服务器接入广域网。

2. 数据交换技术

数据交换是指在计算机通信网络中,通过网络结点的某种转换方式实现任意两个或多个系统之间的连接,它是多结点网络中实现数据传输的有效手段。数据经编码后在通信线路上进行传输,按数据交换技术划分,可分为电路交换、报文交换和分组交换等几种。

1) 电路交换

电路交换是数据通信领域最早使用的交换方式,多用于电话网络交换,如图 3.12 所示。电路交换就是要通过中间交换结点在两个站点之间建立一条专用的通信线路。利用线路交换进行通信,包括建立线路、传输数据和拆除线路三个阶段。

(1) 建立线路。源点向网络发送带目的点地址的请求连接信号,该信号先到达连接源点的第一个交换结点,该结点根据请求中的目的点地址,按一定的规则将请求传送到下一个结点;依此类推,直到目的点。目的点接到请求信号后,若同意通信,从刚才的来路返回一

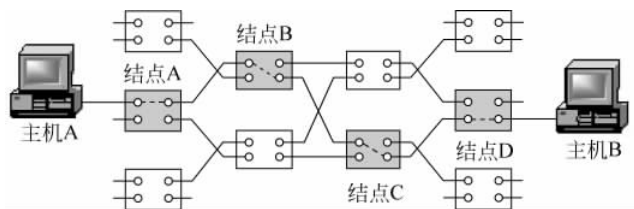


图 3.12 电路交换过程示意图

个应答信号,此时,源目的结点之间的线路即已建成。

(2) 传输数据。源点在已建立的线路上发送数据和控制信息,直至全部发送完毕。在整个数据传输过程中,所建立的电路必须始终保持连接状态。

(3) 拆除线路。源点数据发送完毕,且目的点也正确接收完毕,就可由某一点提出拆线请求,拆除原来建立的线路。

2) 报文交换的工作原理

当端点间交换的数据具有随机性和突发性时,采用电路交换方法的缺点是信道容量和有效时间的浪费。采用报文交换则不存在这种问题。

报文交换方式的数据传输单位是报文,报文就是站点一次性要发送的数据块,其长度不限且可变。当一个站要发送报文时,它将一个目的地址附加到报文上,网络结点根据报文上的目的地址信息,把报文发送到下一个结点,一直逐个结点的转送到目的结点。如图 3.13 所示。

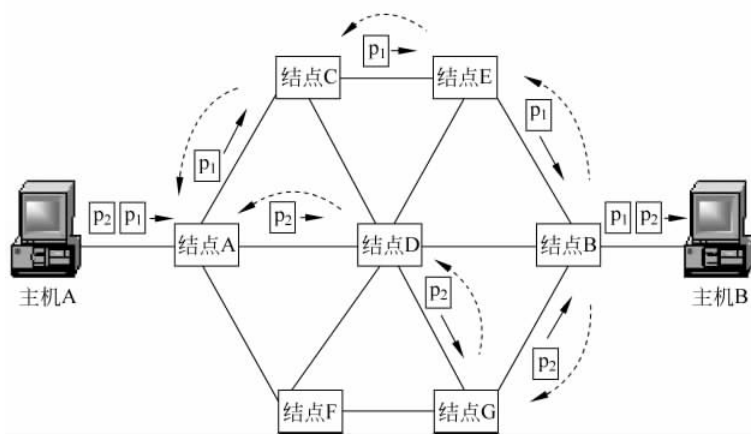


图 3.13 报文交换示意图

每个结点在收到整个报文并检查无误后,就暂存这个报文,然后利用路由信息找出下一个结点的地址,再把整个报文传送给下一个结点。因此,端与端之间无须先通过呼叫建立连接。

报文从源点传送到目的地采用“存储——转发”方式,在传送报文时,一个时刻仅占用一段通道。它的电路利用率高。由于许多报文可以分时共享两个结点之间的通道,所以对于同样的通信量来说,对电路的传输能力要求较低。

3) 分组交换的工作原理

分组交换是报文交换的一种改进,它将报文分成若干个分组,每个分组的长度有一个上限,有限长度的分组使得每个结点所需的存储能力降低了,分组可以存储到内存中,提高了交换速度。它适用于交互式通信,如终端与主机通信。分组交换有虚电路分组交换和数据报分组交换两种。它是计算机网络中使用最广泛的一种交换技术。

(1) 虚电路分组交换。在虚电路分组交换中,为了进行数据传输,网络的源结点和目的结点之间要先建一条逻辑通路。每个分组除了包含数据之外还包含一个虚电路标识符,在预先建好的路径上的每个结点都知道把这些分组引导到哪里去,不再需要路由选择判定。最后,由某一个站用清除请求分组来结束这次连接。它之所以是“虚”的,是因为这条电路不是专用的,如图 3.14 所示。

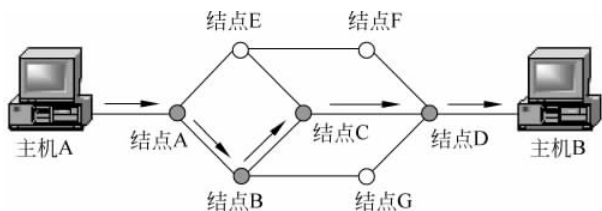


图 3.14 虚电路分组交换示意图

虚电路分组交换的主要特点是:在数据传送之前必须通过虚呼叫设置一条虚电路。但并不像电路交换那样有一条专用通路,分组在每个结点上仍然需要缓冲,并在线路上进行排队等待输出。

(2) 数据报分组交换。在数据报分组交换中,每个分组的传送是被单独处理的。每个分组称为一个数据报,每个数据报自身携带足够的地址信息。一个结点收到一个数据报后,根据数据报中的地址信息和结点所存储的路由信息,找出一个合适的出路,把数据报原样地发送到下一结点。由于各数据报所走的路径不一定相同,因此不能保证各个数据报按顺序到达目的地,有的数据报甚至会中途丢失。整个过程中,没有虚电路建立,但要为每个数据报做路由选择。

3.2 Internet

电子商务的基础是 Internet,也是 Internet 的最大应用。它们相互影响、相互作用,正是 Internet 的迅猛发展,推动了电子商务发展到一个新的时期,而电子商务的发展又对 Internet 提出了新的要求。

3.2.1 Internet 概述

Internet 即国际计算机互联网,又叫国际计算机信息资源网,它是位于世界各地并且彼此相互通信的一个大型计算机网络。组成 Internet 的计算机网络包括小规模局域网(LAN)、城市规模的区域网(MAN)以及大规模的广域网(WAN)。这些网络通过普通电话

线、高速率专用线路、卫星、微波和光缆等把不同国家的大学、公司、科研部门以及军事和政府组织连接起来。Internet 网络互联采用的协议是 TCP/IP 协议。

Internet 具有这样的能力,它能将不同的网络互联起来、构成一个统一的整体。所以 Internet 又称作是网络的网络,它将各种各样的网络连在一起,而不论其网络规模的大小、主机数量的多少、地理位置的异同,以及网络的类型。所谓把网络互联起来,也就是把网络的资源组合起来,这就是 Internet 的精华及其迅速发展的原因。

Internet 也是一个面向公众的社会性组织,世界各地数以百万计的人们可以通过 Internet 进行信息交流和资源共享。

1. Internet 的起源与发展

Internet 的前身 ARPAnet 是美国国防部高级研究计划署在 1969 年建立的实验网络,它最初只有 4 台主机,其设计目标是当网络的一部分因战争或其他特殊原因而遭到破坏时,其余部分仍能正常工作。1986 年美国国家科学基金会(NSF)以 5 个科研教育服务的计算机中心为基础建立了 NSFnet 网络,以便在全美国实现资源共享。由于 NSFnet 的成功设计,在它建成后不久便成为 Internet 的主干网。到 1991 年并入 Internet 的子网已达 3000 多个。

Internet 的第一次快速发展出现在 20 世纪 80 年代中期。由于美国国家科学基金的鼓励和资助,很多大学、政府资助的研究机构,甚至私营的研究机构纷纷把自己的局域网并入 NSFnet 中。此时 NSFnet 已经成为一个“网中网”,在这个“网中网”中各个子网分别负责自己的架设和运作费用,而这些子网又通过 NSFnet 互联起来。

Internet 历史上的第二次飞速发展归功于 Internet 的商业化。20 世纪 90 年代,商业机构开始踏入 Internet,它很快就发现了 Internet 在通信、资料检索、客户服务等方面的巨大潜力。世界各地无数的企业及个人纷纷涌入,带来了 Internet 发展史上一个新的飞跃。1991 年美国的通用原子公司、国际高性能系统公司和 Unnet 技术公司组成了“商用 Internet 协会”,宣布用户可以把他们的 Internet 子网用于任何商业用途。在此之后,世界各地无数的企业及个人纷纷涌入 Internet,这将 Internet 的发展推上了新的高潮。Internet 成熟了,1995 年 4 月 30 日 NSFnet 宣布停止运作,由美国政府指定的三家私营企业代替它。

当前的 Internet 约有数十万个子网通过自愿原则互联起来,全球 Internet 的用户超过 1 亿个。当今世界上,没有一家公司叫 Internet 公司,也没有任何机构完全拥有 Internet。

我国从 1994 年 3 月起正式加入 Internet。同年 5 月,在中国科学院高能物理研究所实现联网,从此我国成为 Internet 网络大家庭中的正式成员。

2. TCP/IP 协议

TCP/IP(Transmission Control Protocol/Internet Protocol)是传输控制协议和网际互联协议的英文缩写。世界上有各种不同类型的计算机,例如 IBM 兼容机和苹果的 MAC;相同的机器也存在着不同的操作系统,如 UNIX、Windows 和 OS2 等。这些不同的机器要互相通信,就必须有同一个标准。TCP/IP 协议具有与操作系统无关,与物理网络无关,并且寻址标准的特点,在 Internet 中较好地解决了不同类型、不同操作系统计算机互联的问题,目前已经成为网际互联的事实上的工业标准。

1) TCP/IP 协议体系模型

TCP/IP 协议通常采用四层模型,即将 TCP/IP 结构由上至下分成应用层、传输层、网

间层和网络访问层。

(1) 网络访问层。该层将 IP 报文封装成被网络传输的帧,并将 IP 地址映射为网络使用的物理地址。

(2) 网间层。网际互联协议 IP 是 TCP/IP 的核心,IP 可提供基本的分组传输服务。在互联网,主机物理地址的差异被 IP 层所屏蔽。IP 层所用到的地址叫做万联网地址,又称 IP 地址。IP 地址共分两部分即网络号与主机号。其中网络号表示某个主机是属于哪一个网络的,该网络号在互联网中具有唯一性;主机号则表示该主机在所属网络中的地址号。

2) TCP 协议

线路的调和,让网络中每一台计算机每次只传送一定的数据量,每次所能传送的数据量的单位称为一个数据分组(或数据小包)。数据总量分割传送、设备轮流服务的原则称为分组交换。计算机网络用来保证每台计算机平等地使用网络资源的技术称为分组交换技术。分组交换可以有效地避开延迟。当某台计算机发送较长信息时,它可以分为若干个分组;另一台计算机发送较短信息,可以不分组或少分组。长信息发送出一个分组后,短信息有机会发送自己的分组,结果是短信息无须等待长信息发送完就可以发送,从而避开了延迟。计算机很容易做数据分组的工作,而且在分组交换网络中,分组传输得很快,常常达到每秒传输 1000 个以上的分组,当几个人同时将信息发送到一个共享网络时,千分之几秒的时间间隔是感觉不到的,所以可以认为延迟是不存在的。分组交换允许任何一台计算机在任何时候都能发送数据。当只有一台计算机需要使用网络时,它就可以连续发送分组。一旦另一台计算机准备开始发送数据,共享就开始了,两台计算机轮流地发送数据,公平地分享资源。如果又有一台计算机准备开始发送,那么网络将对多台计算机重新分配网络资源。分组交换技术能够在有计算机准备发送数据和有计算机停止发送数据时立即进行自动调整,因而每台计算机在任何时刻都能够公平地分享网络资源。与大多数网络一样,Internet 也采用分组交换技术,数据以分组方式传送。发送方是将信息分组再通过 Internet 传送;接收方在接收到一个信息的分组后,重新组装成原来完整的信息。在 Internet 上同一时刻流动着来自各个方向的多台计算机的分组信息。

3) IP 协议

(1) IP 地址。

Internet 网络中,我们经常会遇到 IP 地址这个概念,这也是网络中的一个重要概念。为了在 Internet 网络环境中实现计算机之间的通信,网络中的任何一台计算机都必须有一个地址,而且同一个网上的地址不允许重复。在进行数据传输时,通信协议一般需要在所要传输的数据中增加某些信息,而其中最重要的就是发送信息的计算机的地址(源地址)和接收信息的计算机的地址(目标地址)。所谓 IP 地址就是 Internet 中主机的地址,是给每个连接在 Internet 上的主机分配一个在全世界范围唯一的 32bit 编号。TCP/IP 规定,每个互联网地址长 32 位,由网络号和主机号两部分组成。IP 地址的结构使我们可以在 Internet 上很方便地寻址。

(2) IP 地址分类。

由于网络号与主机号的位数决定了整个互联网中能包含多少个网络即每个网络中能容纳多少台主机。然而在互联网中的网络数是难以确定的数字,只能估计每一个网络的预期规模。按照网络规模大小,将互联网地址分为主要三类即 A 类、B 类和 C 类,这三类地址的

容量如表 3.1 所示。

表 3.1 互联网中网络类型

类 别	网络号位数	网络规模	主机号位数	主 机 规 模
A 类	8 位	126 个	24 位	1 677 214
B 类	16 位	16 382 个	16 位	65 534
C 类	24 位	200 万	8 位	254

① A 类为能够为 1600 万个主机服务的超大型网络。A 类网络地址被分配给主要的网络服务供应商,即大型的地区网或国家网。A 类网络地址的第一个 8 位位组的值为 1~126,即全世界最多有 126 个 A 类网络,但每个网络可以拥有的主机数目相当大,最多可以容纳 1 677 214 个主机。

② B 类为大型网络。B 类地址被分配给拥有大型网络的机构,如大学和大型企业。B 类地址的第一个 8 位位组的值应在 127~191 之间,该 8 位位组用于描述网络本身,后两个 8 位位组用于表示主机,共含有 16 382 个可能的网络标识符,每个 B 类网络可以拥有 65 534 个主机。

③ C 类为大量的小型网络。C 类地址的第一个 8 位位组的值应为 192~223; 前三个 8 位位组用于描述网络号码,最后一个 8 位位组指定了主机的号码。由于小型网络数量众多,因而大多数网络将拥有 C 类地址。C 类地址有超过 200 万个可能的网络号码,但每个 C 类网络只能支持 254 个主机。以上地址定义方式既适应了大网量少、小网量大、大网主机多、小网主机少的特点,又方便网络号和地址号的提取。

除了以上三类主要地址外,互联网地址还有另外两类地址,即 D 类和 E 类地址,其中 D 类地址的第一个 8 位位组为 224~239,该组地址是为多路广播(Multicasting)而保留的,它是一种成组通信的方法,使用这种技术可以将报文信息立即发送给组内的所有成员,E 类地址的第一个 8 位位组的值为 240~247,该类地址保留给将来使用。

(3) 互联网地址的直观表示法。

在协议软件中,互联网地址是以二进制形式表示的,这种形式较适用于软件处理,但令人感到头痛。在面向人的文档中,互联网地址被直观地表示为四个以小数点隔开的十进制整数,其中每个整数对应一个字节,这种表示方法被称为“点分十进制标识”,例如,202. 120. 101. 23。IP 地址通常用更直观的、以圆点分隔号的 4 个十进制数字表示,每一个数字对应于 8 个二进制的比特串,如某一台主机的 IP 地址为: 128. 20. 4. 1。

4) Internet 域名系统(DNS)

域名指明了网络连接所在的国家,拥有网络连接的机构类型,在一些情况下,域名被定义得更细致。

(1) 什么是域名。

Internet 域名是 Internet 网络上的一个服务器或一个网络系统的名字。域名是由“·”分隔两段或三段字符串,如,sohu. com 或 sina. com. cn。其中“sohu”或“sina”这一段字符串是服务器或网络的名字,由其所有者定义,由若干个英文字母或数字组成;其后由类别域名和行政区域名组成,由“·”分隔。

(2) 行政区域名。

行政区域名表示服务器或网络系统的地域属性,是基于地理位置的顶层域,采用了国际标准化组织在 ISO 3166 文档中指定的二字符国家名称。例如,.cn 代表中国,.us 代表美国,.uk 代表联合国等。

(3) 类别域名。

类别域名体现服务器或网络系统的性质,主要有六个域名:

.com 代表商业组织,如 netcom.com,microsoft.com。

.net 代表网络机构,如 Internic.net。

.gov 代表政府机构,如 whitehouse.gov。

.edu 代表教育机构,如 mit.edu。

.mil 代表非保密的军事网络,如 Army.mil。

.org 代表非商业或教育型的其他机构。

由于非地理域的域名是根据政治属性而不是根据简单的地理属性确定的,所以有同一机构拥有但位置相距甚远的机器,可能有相似的域名,此时就不能根据域名来推测主机所在的位置。另外,由于美国的行政区域名可以省略。

(4) 域名的意义。

从技术上讲,域名只是一个 Internet 中用于解决地址对应问题的一种方法,可以说只是一个技术名词。但是,由于 Internet 已经成为了全世界人的 Internet,域名也自然地成为了一个社会科学名词。

从社会科学的角度看,域名已成为了 Internet 文化的组成部分。从商界看,域名已被誉为“企业的网上商标”。没有一家企业不重视自己产品的标识——商标,而域名的重要性和其价值,也已经被全世界的企业所认识。

Internet 已经为越来越多的人所认识,电子商务、网上销售、网络广告已成为商界关注的热点。由于域名和商标都在各自的范畴内具有唯一性,从企业树立形象的角度看,域名从某种意义上讲,就和商标有着潜移默化的联系。事实上许多企业在选择域名时,也往往希望用和自己企业商标一致的域名。

3. IPv6 协议

IPv6 是 IP 协议的新版本,是在 IPv4 的基础上诞生的,它巨大的地址容量能够满足 Internet 飞速发展的需要,被认为是建设移动信息社会的一个重要基石,它集移动性、安全性和服务质量保证于一体,是建设未来 Internet 的最佳方案。

现有 Internet 的基础是 IPv4,到目前为止有近 30 年的历史了。由于 Internet 的迅猛发展,据统计平均每年 Internet 的规模就扩大一倍。尽管 IPv4 有 40 多亿个地址,但由于 Internet 用户的快速增长及地址分配的不均匀性,IPv4 的局限性就越来越明显。个人电脑市场的急剧扩大,还有个人移动计算设备的上网、网上娱乐服务的增加、多媒体数据流的加入,以及出于安全性等方面的需求都迫切要求新一代 IP 协议的出现。

同 IPv4 相比,IPv6 的主要优势体现在以下几方面:

1) 更大的地址空间

IPv4 中规定 IP 地址长度为 32,即有 $2^{32}-1$ 个地址;而 IPv6 中 IP 地址的长度为 128,

即有 $2^{128} - 1$ 个地址。

2) 更小的路由表

IPv6 的地址分配一开始就遵循聚类原则,这使得路由器能在路由表中用一条记录表示一片子网,大大减小了路由器中路由表的长度,提高了路由器转发数据包的速度。

3) 增强的组播支持以及对流支持

这使得网络上的多媒体应用有了长足发展的机会,为服务质量控制提供了良好的网络平台。

4) 加入了对自动配置的支持

这是对 DHCP 协议的改进和扩展,使得网络(尤其是局域网)的管理更加方便和快捷。

5) 更高的安全性

在使用 IPv6 网络中用户可以对网络层的数据进行加密并对 IP 报文进行校验,这极大地增强了网络安全。

3.2.2 Internet 接入技术

目前,我国的接入网建设主要涉及以下几个技术领域:

1. 光纤接入技术

光纤是适合宽带业务发展的理想接入传输介质。在一定的距离与规模条件下,光纤接入网络的综合成本已经与铜缆相近。随着用户环路新技术的开发与应用,光纤用户环路技术已经日趋成熟,而且价格在稳步下降。为了进一步降低成本,各种光纤混合网将应运而生。

2. 无线接入技术

无线本地环路是用无线通信技术连接交换机与用户终端设备的通信系统。它以微波、卫星、小天线地球站(VEST)以及无线蜂房通信技术作为无线传输手段,为用户提供一种本地接入的通信方法。以码分多址(CDMA)技术为基础的无线通信是实现无线本地环路的最佳方案。码分多址技术以独特的数字传输方式,为各种新型通信工具与通信服务提供了可靠的技术手段。利用码分多址技术构成的无线本地环路具有以下几个特点:

- (1) 传输距离远,可以在更大覆盖范围内为更多的用户服务。
- (2) 使用灵活,系统安装容易。
- (3) 通信质量高。
- (4) 服务管理方便。

通信由有线到无线、电视传输从无线到有线是目前发展的趋势。随着无线接入成本的不断下降,无线接入方式将会与光纤接入方式相辅相成。

3. 铜缆接入技术

铜缆接入是目前电话业务的主要接入手段,电信企业已经为之投入了大量的资金,因此在建设接入网时要充分考虑铜缆的使用价值。随着数字传输技术的不断发展,铜缆不仅将在电话接入中继续发挥主要作用,而且在光纤接入网尚未形成或用户急需时,也可以通过数字传输技术满足用户接入的需要。

接入网所涉及的技术领域宽,业务面广,并且多数技术还处于不断发展与完善的过程中。随着数据与多媒体业务的发展,用户对接入网提出了更高的要求。传统的以模拟铜缆

为传输手段的用户接入方式已经无法满足社会需求,接入网正在向数字化与宽带化方向发展。

在接入网的建设中,很多国家都确定了以光纤为主、其他手段为辅的方针。我国的国情是人口密度大,用户线路长,各地区之间的发展极不平衡。针对我国国情,在接入网建设中提出了“以光纤为主,以无线为辅,加快接入网光纤化进程”的发展策略,同时还强调“对于接入网的发展建设,既应兼顾目前的实际情况,更应具有长远的眼光,即兼顾当前电话主业的发展与开展宽带业务的需求,以适应未来网络线路升级的要求”。因此,采用光纤作为连接日益加宽传输带宽的接入设备的传输媒介,同时采用异步传输模式 ATM,将是我国接入网建设发展的主要模式。

3.2.3 Internet 的功能与应用

Internet 发展迅猛,提供的服务在不断增加,应用领域也在不断扩大,而且日益渗透到人们的生活和工作中,成为日常交流中不可缺少的组成部分。Internet 上的信息资源非常丰富,信息应用的种类也是多种多样的。以下列出的是 Internet 提供的基本服务。

1. 电子邮件(E-mail)

电子邮件利用计算机的存储、转发原理,克服时间、地理上的差距,通过计算机终端和通信网络进行文字、声音、图像等信息的传递。它是 Internet 的一项重要功能。Internet 电子邮件地址格式为:用户名@域名。例如:zhigang@hotmail.com。

电子邮件采用电子化的方式收发消息,这些消息可以包括文本(text)、带格式的文本(formatted text)或多媒体。此外各种文件还可以附加在消息里传输。各种 E-mail 软件具有不同的功能,如常用的软件有 Internet Mail、Outlook、Outlook Express、Eudora 等。

2. 文件传输(FTP)

文件传输服务允许 Internet 上的客户将一台计算机上的文件传送至另一台计算机上。它可以传送所有类型的文件:文本文件、二进制可执行文件、图像文件、声音文件、数据压缩文件等。FTP 比任何其他方式(如电子邮件)交换数据都要快得多。

3. WWW

WWW(万维网)或叫 Web,是目前最受用户欢迎的一种服务,它把 Internet 上不同的地点的相关数据信息有机地组织起来,供用户查询。它基于超文本系统在 Internet 上传递信息。超文本的功能是通过一个网页链接到另外的资源上,如另一个网页、电子邮件、远程登录、FTP 等,这些资源提供了关于指明的(在网页上是加亮的)字或词组的更多的信息,这叫做超级链接。用户只需单击网页上加亮的词语,客户浏览器就会去寻找相链接的服务。所以,Web 是 Internet 上功能最强的一种服务,它可以链接各种服务。

4. 电子公告牌(BBS)

BBS 也是一项受广大用户欢迎的服务项目,用户可以在 BBS 上留言、发表文章、阅读文章等。

BBS 的英文全称是 Bulletin Board System,翻译为中文就是“电子公告板”。BBS 最早是用来公布股市价格等类信息的,与一般街头和校园内的公告板性质相同,只不过是计算机来传播或获得消息而已。一直到个人计算机开始普及之后,有些人尝试将苹果计算机

上的 BBS 转移到个人计算机上,BBS 才开始渐渐普及开来。近些年来,由于爱好者们的努力,BBS 的功能得到了很大的扩充。

目前,通过 BBS 系统可随时取得各种最新的信息;也可以通过 BBS 系统来和别人讨论计算机软件、硬件、Internet、多媒体、程序设计以及生物学、医学等各种有趣的话题;还可以利用 BBS 系统来发布一些“廉价转让”、“人才招聘”及“求职应聘”等启事;更可以召集亲朋好友到聊天室内高谈阔论。

5. 新闻组(Usenet)

Usenet 是讨论组的集合,它是一种分布式的网上论坛,由散布在全球的 N 个服务器结点所支撑。它有上万个版块,称为组(NewsGroup)。顾名思义起初的组可能主要是用来交流新闻的,但现在一般都是讨论组的性质。组分类别组织起来,例如 sci. chem. electrochem 表示电化学讨论组。常见的顶级类名有 comp(计算机)、news(关于 Usenet 本身话题)、sci(科技)、humanities(人文)、soc(社会)、biz(商业)、rec(娱乐)、talk(聊天)、misc(其他)等。所有的骨干结点都具有这“八大类”,所以又称为主流组。结点管理员可以自行选择开通一些非主流组如 alt 类的组、地域组和本地组。cn. * 类就是中文的地域组。讨论组在国外由于历史长,专业组的水平一般都较高,具有很好的氛围。相比之下中文组要差一点。

新闻组与 BBS 类似,它也是提供一个场所,让对某个问题感兴趣的各个用户之间进行提问、回答、新闻和评论以及其他信息交流。

6. Telnet

在 Internet 中,用户可以通过远程登录使自己的计算机成为远程计算机的终端,然后在其上运行程序,或使用它的软件和硬件资源。

远程登录可以使本地的计算机连到 Internet 上的一个远程的计算机上,但前提是必须有对远程计算机的使用权限,即必须知道账号和密码。登录以后本地计算机就成为这个远程计算机的终端,就像在当地一样,可以使用主机允许使用的各种功能,如执行各种命令、编制程序等。

7. MUD

MUD 是一种网络游戏,是用 Telnet 方式登录到对方服务器上,进行聊天、练功、角色扮演等活动。

3.3 B/S 计算模式

随着计算机技术的不断发展与应用,计算模式从集中式转向了分布式,尤为典型的是客户机/服务器模式(Client/Server,C/S 结构)。两层结构 C/S 模式,在 20 世纪 80 年代及 20 世纪 90 年代初得到了大量应用,最直接的原因是可视化开发工具的推广。之后,它开始向三层结构发展。近年来,随着网络技术不断发展,尤其是基于 Web 的信息发布和检索技术、Java 计算技术以及网络分布式对象技术的飞速发展,导致了很多应用系统的体系结构从 C/S 结构向更加灵活的多级分布结构演变,使得软件系统的网络体系结构跨入一个新阶段,即浏览器/服务器模式(Browser/Server,B/S 体系结构)。

3.3.1 C/S 模式

客户机/服务器(Client/Server,C/S)模式又称 C/S 结构,是软件系统体系结构的一种。这种体系结构是将一个由多台计算机分别执行的应用系统分成两大部分,使它们有机地结合在一起,协同完成整个系统的应用,从而达到系统中软、硬件资源最大限度的利用。C/S 模式简单地讲就是基于企业内部网络的应用系统,它由一个或多个称作服务器的主计算机系统和多个客户计算机系统构成,服务器是整个应用系统资源的存储与管理中心,多台客户机则各自处理相应的功能,共同实现完整的应用。应用程序分为两大部分:一部分是由多个用户共享的信息与功能部分即服务器部分;另一部分是为每个用户所专有,称为客户部分。客户部分负责执行前台功能,如管理用户接口、数据处理和报告请求等。而服务器部分则执行后台服务,如管理共享外设、控制对共享数据库的操纵、接受并应答客户机的请求等。

用户通过客户机上专用的客户程序提供的交互界面向服务器上的服务程序发出某项操作请求;服务程序完成操作,并返回结果或予以答复;再通过客户程序的交互界面向用户显示有关信息和结果,从而让用户获得服务器提供的信息服务,如图 3.15

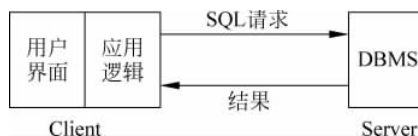


图 3.15 二层 C/S 模式

所示。在这种工作模式下,整个过程在客户程序与服务程序之间交互进行,用户可以获取服务器提供的文本与多媒体信息服务。

客户机/服务器应用系统基本运行关系体现为“请求——响应”的应答模式。每当用户需要访问服务器时就由客户机发出“请求”,服务器接受“请求”并“响应”,然后执行相应的服务,把执行结果送回客户机,由它进一步处理后再提交给用户。由于这种结构被设计成两层模式,显示逻辑和事务处理逻辑部分均被放在客户端,数据处理逻辑和数据库放在服务器端,从而使客户端变得很胖,成为胖客户机,相对来说,服务器端的任务则较轻,成为瘦服务器。

二层 C/S 模式具有开发简单、技术要求较低等特点;随着 C/S 结构应用范围的不断扩大和计算机网络技术的发展,这种结构带来的问题日益明显,主要表现在以下几个方面。

1. 系统的可靠性有所降低

一个 Client/Server 系统是由各自独立开发、制造和管理的各种硬件和软件的混合体,出现问题时,很难立即获得帮助。

2. 维护费用较高

业务逻辑都在客户端控制,从而形成“胖客户”;一旦企业的需求产生变化,客户端的变化往往也很大,常常不得不对客户端动“大手术”,严重时甚至丢弃原来的客户端重新再做。

3. 客户端和服务端之间的依赖性较强

由于现代企业的规模越来越大,ERP 产品数据和应用的分布性特点越来越明显,而传统的 C/S 二层模型客户端和服务端对于对方的依赖较强,对于构建企业的信息平台是一个障碍。

4. 系统缺乏灵活性

客户机/服务器模式需要对每一应用独立地开发应用程序,消耗了大量资源。由于信息

量的迅速增大,专用的客户端已经无法满足多功能的需求。

3.3.2 B/S 模式

将二层 C/S 结构中客户端的功能分离,即将客户端的业务逻辑和用户界面分离,形成客户端表示层、中间应用逻辑层和后端数据服务层这样的体系结构,二层 C/S 结构就演变成三层 C/S 结构,如图 3.16 所示。在保证客户端功能的前提下,为用户提供了一个简洁的界面。所有应用处理逻辑放在业务逻辑层,这意味着如果需要修改应用程序代码,只需要对中间层应用服务器进行修改,而不用修改大量的客户端应用程序。从而使开发人员可以专注于应用系统核心业务逻辑的分析、设计和开发,简化了应用系统的开发。



图 3.16 三层 C/S 模型

1. 表示层

表示层是应用的用户接口部分,它担负着用户与应用间的对话功能。它用于检查用户从键盘等输入的数据,显示应用输出的数据。为使用户能直观地进行操作,一般要使用图形用户接口(GUI),操作简单、易学易用。在变更用户接口时,只需改写显示控制和数据检查程序,而不影响其他两层。检查的内容也只限于数据的形式和值的范围,不包括有关业务本身的处理逻辑。

2. 业务逻辑层

业务逻辑层又称应用逻辑层、功能层,功能相当于应用的本体。它是将具体的业务处理过程逻辑地编入程序中。例如,在制作订购合同的时要计算合同金额,按照定好的格式配置数据、打印订购合同,而处理所需的数据则要从表示层或数据层取得。表示层和逻辑层之间的数据交往要尽可能简洁。

通常,在逻辑层中包含有确认用户对应用和数据库存取权限的功能以及记录系统处理日志的功能,这层的程序多半是用可视化编程工具开发的,也有使用 COBOL 和 C 语言的。

3. 数据层

数据层由 DBMS 和数据库构成,DBMS 负责管理对数据库数据的操作,必须能迅速执行大量数据的更新和检索。现在的主流是关系数据库管理系统(RDBMS)。因此,一般从功能层传送到数据层的要求大都使用 SQL 语言。

三层 C/S 结构具有许多二层 C/S 结构无法比拟的优点。

1) 具有灵活的硬件系统构成

对于各个层可以选择与其处理负荷和处理特性相适应的硬件。这是一个与系统可缩放性直接相关的问题。例如,最初用一台 UNIX 工作站作为服务器,将数据层和功能层都配置在这台服务器上。随着业务的发展,用户数和数据量逐渐增加,这时就可以将 UNIX 工作站作为功能层的专用服务器,另外追加一台专用于数据层的服务器。若业务进一步扩大,

用户数进一步增加,则可以继续增加功能层的服务器数目,用以分割数据库。清晰、合理地分割三层结构并使其独立,可以使系统构成的变更非常简单。因此,被分成三层的应用基本上不需要修正。

2) 提高程序的可维护性

三层 C/S 结构中,应用的各层可以并行开发,各层也可以选择各自最适合的开发语言。

3) 利于变更和维护应用技术规范

因为是按层分割功能,所以各个程序的处理逻辑变得十分简单。

4) 进行严密的安全管理

越关键的应用,用户的识别和存取权限设定愈重要。在三层 C/S 结构中,识别用户的机构是按层来构筑的,对应用和数据的存取权限也可以按层进行设定。例如,即使外部的入侵者突破了表示层的安全防线,若在功能层中备有另外的安全机构,系统也可以阻止入侵者进入其他部分。

5) 系统管理简单,可支持异种数据库,有很高的可用性

近年来,随着 Internet 的广泛应用和相关技术的发展,人们在三层 C/S 结构的软件配置时,将表示层用浏览器实现,将业务逻辑层划分为 Web 服务器和应用服务器两部分,这样又演变成一种新的系统结构——浏览器/服务器(Browser/Server)结构,简称 B/S 结构。一个典型的 B/S 模式应用系统结构如图 3.17 所示。



图 3.17 B/S 计算模型

B/S 模式是对三层 C/S 模式应用的扩展。一方面,基于“客户机/服务器”工作模式的服务越来越多地转向 Web 形式;另一方面,浏览器作为访问因特网各种应用信息服务的通用客户程序而发展起来,并逐渐成为新一代操作系统的人机交互界面,用户越来越多地使用浏览器访问因特网上的大多数各类信息服务,因此,“客户机/服务器”工作模式逐步转为“浏览器/服务器”工作模式。B/S 模式最大的好处是运行维护比较简便,能实现不同的人员,从不同的地点,以不同的接入方式(比如 LAN, WAN, Internet/Intranet 等)访问和操作共同的数据;最大的缺点是对企业外网环境依赖性太强,由于各种原因引起企业外网中断都会造成系统瘫痪。

3.3.3 中间件

计算机技术在迅猛发展,从硬件技术看,CPU 速度越来越高,处理能力越来越强;从软件技术看,应用程序的规模不断扩大,特别是 Internet 及 WWW 的出现,使计算机的应用范围更为广阔,许多应用程序需在网络环境的异构平台上运行。这一切都对新一代的软件开发提出了新的需求。在这种分布式异构环境中,通常存在多种硬件系统平台(如 PC,工作站,小型机等),在这些硬件平台上又存在各种各样的系统软件(如不同的操作系统、数据库、语言编译器等),以及多种风格各异的用户界面,这些硬件系统平台还可能采用不同的网络

协议和网络体系结构连接。如何把这些系统集成起来并开发新的应用是一个非常现实而困难的问题。为解决分布异构问题,人们提出了中间件(middleware)的概念。

1. 中间件的概念

在计算机和网络技术应用中,各种各样的应用软件需要在各种平台之间进行移植,或者一个平台需要支持多种应用软件和管理多种应用系统,软、硬件平台和应用系统之间需要可靠和高效的数据传递或转换,使系统的协同性得以保证。这些,都需要一种构筑于软、硬件平台之上,同时对更上层的应用软件提供支持的软件系统,而中间件正是在这个环境下应运而生。

在众多关于中间件的定义中,比较普遍被接受的是 IDC 的表述,即中间件(middleware)是一种独立的系统软件或服务程序,分布式应用软件借助这种软件在不同的技术之间共享资源,中间件位于客户机和服务器的操作系统之上,管理计算资源和网络通信。

从中间件的定义可以看出,中间件是一类软件,而非一种软件;中间件不仅仅实现互连,还要实现应用之间的互操作;中间件是基于分布式处理的软件,定义中特别强调了其网络通信功能。

中间件是基础软件的一大类,属于可复用软件的范畴。顾名思义,中间件处于操作系统软件与用户的应用软件的中间。中间件在操作系统、网络和数据库之上,应用软件的下层,总的作用是为处于自己上层的应用软件提供运行与开发的环境,帮助用户灵活、高效地开发和集成复杂的应用软件。

中间件是位于平台(硬件和操作系统)和应用之间的通用服务,如图 3.18 所示,这些服务具有标准的程序接口和协议。针对不同的操作系统和硬件平台,它们可以有符合接口和协议规范的多种实现。

由于标准接口对于可移植性和标准协议对于互操作性的重要性,中间件已成为许多标准化工作的主要部分。对于应用软件开发,中间件远比操作系统和网络服务更为重要,中间件提供的程序接口定义了一个相对稳定的高层应用环境,不管底层的

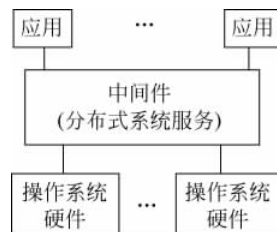


图 3.18 中间件

计算机硬件和系统软件怎样更新换代,只要将中间件升级更新,并保持中间件对外的接口定义不变,应用软件几乎不须任何修改,从而保护了企业在应用软件开发和维护中的重大投资。

中间件在分布式的客户和服务之间扮演着承上启下的角色,如事务管理、负载均衡以及基于 Web 的计算等。利用这些技术有助于减轻应用软件开发者的负担,使他们利用现有的硬件设备、操作系统、网络、数据库管理系统以及对象模型创建分布式应用软件时更加得心应手。由于中间件能够保护企业的投资,保证应用软件的相对稳定,实现应用软件的功能扩展;同时中间件产品在很大程度上简化了一个由不同硬件构成的分布式处理环境的复杂性,所以它的出现正日益引起用户的关注。

2. 中间件的种类

中间件所包含的范围十分广泛,针对不同的应用需求涌现出多种各具特色的中间件产

品。由于中间件需要屏蔽分布环境中异构的操作系统和网络协议,它必须能够提供分布环境下的通信服务,我们将这种通信服务称之为平台。基于目的和实现机制的不同,我们将平台分为远程过程调用(Remote Procedure Call)、面向消息的中间件(Message-Oriented Middleware)和对象请求代理(Object Request Brokers)等三类。下面针对几类主要的中间件分别加以简要的介绍。

1) 远程过程调用

远程过程调用是一种广泛使用的分布式应用程序处理方法。一个应用程序使用 RPC 来“远程”执行一个位于不同地址空间里的过程,并且从效果上看和执行本地调用相同。RPC 为 Client/Server 分布式计算提供了有力的支持。同时,远程过程调用 RPC 所提供的是基于过程的服务访问,Client 与 Server 进行直接连接,没有中间机构来处理请求,因此也具有一定的局限性。

2) 面向消息的中间件

面向消息的中间件(MOM)指的是利用高效可靠的消息传递机制进行平台无关的数据交流,并基于数据通信来进行分布式系统的集成。通过提供消息传递和消息排队模型,它可在分布环境下扩展进程间的通信,并支持多种通信协议、语言、应用程序、硬件和软件平台。目前流行的 MOM 中间件产品有 IBM 的 MQSeries、BEA 的 MessageQ 等。消息传递和排队技术有以下三个主要特点,即通信程序可在不同的时间运行、对应用程序的结构没有约束、程序与网络复杂性相隔离。

3) 对象请求代理

随着对象技术与分布式计算技术的发展,两者相互结合形成了分布对象计算,并发展为当今软件技术的主流方向。1990 年年底,对象管理集团 OMG 首次推出对象管理结构(Object Management Architecture,OMA),对象请求代理(Object Request Broker)是这个模型的核心组件。它的作用在于提供一个通信框架,透明地在异构的分布计算环境中传递对象请求。CORBA 规范包括了 ORB 的所有标准接口。1991 年推出的 CORBA 1.1 定义了接口描述语言 OMG IDL 和支持 Client/Server 对象在具体的 ORB 上进行互操作的 API。CORBA 2.0 规范描述的是不同厂商提供的 ORB 之间的互操作。

对象请求代理(ORB)是对象总线,它在 CORBA 规范中处于核心地位,定义异构环境下对象透明地发送请求和接收响应的基本机制,是建立对象之间 Client/Server 关系的中间件。ORB 使得对象可以透明地向其他对象发出请求或接受其他对象的响应,这些对象可以位于本地也可以位于远程机器。ORB 拦截请求调用,并负责找到可以实现请求的对象、传送参数、调用相应的方法、返回结果等。Client 对象并不知道同 Server 对象通信、激活或存储 Server 对象的机制,也不必知道 Server 对象位于何处、它是用何种语言实现的、使用什么操作系统或其他不属于对象接口的系统成分。

值得指出的是,Client 和 Server 角色只是用来协调对象之间的相互作用,根据相应的场合,ORB 上的对象可以是 Client,也可以是 Server,甚至兼有两者。当对象发出一个请求时,它是处于 Client 角色;当它在接受请求时,它就处于 Server 角色。大部分的对象都是既扮演 Client 角色又扮演 Server 角色。

4) 事务处理监控

事务处理监控(Transaction Processing Monitors)最早出现在大型机上,为其提供支持

大规模事务处理的可靠运行环境。随着分布计算技术的发展,分布应用系统对大规模的事务处理提出了需求,比如商业活动中大量的关键事务处理。事务处理监控介于 Client 和 Server 之间,进行事务管理与协调、负载平衡、失败恢复等,以提高系统的整体性能。它可以被看作是事务处理应用程序的“操作系统”。总体来说,事务处理监控有以下功能:

- (1) 进程管理,包括启动 Server 进程、为其分配任务、监控其执行并对负载进行平衡。
- (2) 事务管理,即保证在其监控下的事务处理的原子性、一致性、独立性和持久性。
- (3) 通信管理,为 Client 和 Server 之间提供了多种通信机制,包括请求响应、会话、排队、订阅发布和广播等。

中间件能够屏蔽操作系统和网络协议的差异,为应用程序提供多种通信机制,并提供相应的平台以满足不同领域的需要。因此,中间件为应用程序构建了一个相对稳定的高层应用环境。然而,中间件服务也并非“万能药”,中间件所应遵循的一些原则离实际还有很大距离。多数流行的中间件服务使用专有的 API 和专有的协议,使得应用建立于单一厂家的产品,来自不同厂家的实现很难互操作。有些中间件服务只提供某些平台的实现,从而限制了应用在异构系统之间的移植。应用开发者在这些中间件服务之上建立自己的应用还要承担相当大的风险,随着技术的发展他们往往还需重写他们的系统。尽管中间件服务提高了分布计算的抽象化程度,但应用开发者还需面临许多艰难的设计选择,例如,开发者还需决定分布应用在 Client 方和 Server 方的功能分配。

3. 主流中间件技术平台

考察当前主流的分布计算技术平台,主要有 OMG 的 CORBA、Sun 的 J2EE 和 Microsoft DNA 2000。它们都是支持服务器端中间件技术开发的平台,但都有其各自的特点,将其分别阐述如下。

1) OMG 的 CORBA

CORBA 是 Common Object Request Broker Architecture 的缩写,简称公共对象请求代理结构,它由国际对象管理组织 OMG 制定。CORBA 分布计算技术是 OMG 组织基于众多开放系统平台厂商提交的分布对象互操作内容的基础上制定的公共对象请求代理体系规范。

CORBA 分布计算技术,是由绝大多数分布计算平台厂商所支持和遵循的系统规范技术,具有模型完整、先进,独立于系统平台和开发语言,被支持程度广泛的特点,已逐渐成为分布计算技术的标准。CORBA 标准主要分为 3 个层次,即对象请求代理、公共对象服务和公共设施,如图 3.19 所示。最底层是对象请求代理 ORB,规定了分布对象的定义(接口)和语言映射,实现对象间的通信和互操作,是分布对象系统中的“软总线”;在 ORB 之上定义了很多公共服务,可以提供诸如并发服务、名字服务、事务(交易)服务、安全服务等各种各样的服务;最上层的公共设施则定义了组件框架,提供可直接为业务对象使用的服务,规定业务对象有效协作所需的协定规则。目前,CORBA 兼容的分布计算产品层出不穷,其中有中间件厂商的 ORB 产品,如 BEAM3、IBM Component Broker;有分布对象厂商推出的产品,如 IONAObix 和 OOCObacus 等。

总之,CORBA 的特点是大而全,互操作性和开放性非常好。CORBA 的缺点是庞大而复杂,并且技术和标准的更新相对较慢,CORBA 规范从 1.0 升级到 2.0 所花的时间非常短,而再往上的版本的发布就相对十分缓慢了。

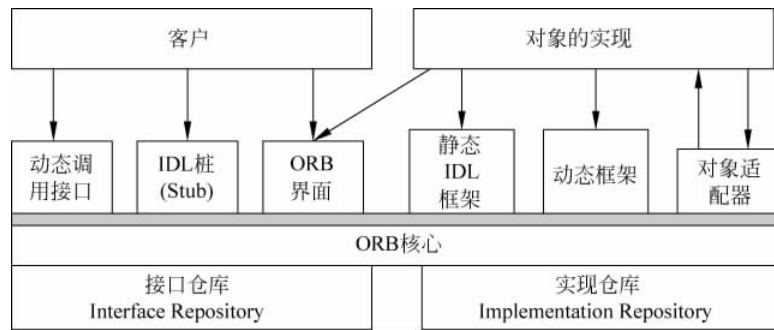


图 3.19 CORBA 的体系结构

2) Sun 的 J2EE

J2EE 是 Java 2 Platform Enterprise Edition 的缩写,是一种多层应用模式的结构体系。整个规范由 Sun 公司提出,它将业务逻辑从系统服务功能和用户界面中分离出去,放置在客户层和应用基础设施这两层之间的中间层,是目前应用的最为广泛的面向 Web 的应用系统结构规范。

为了推动基于 Java 的服务器端应用开发, Sun 于是在 1999 年底推出了 Java2 技术及相关的 J2EE 规范, J2EE 的目标是: 提供平台无关的、可移植的、支持并发访问和安全的,完全基于 Java 的开发服务器端中间件的标准。

在 J2EE 中, Sun 给出了完整的基于 Java 语言开发面向企业分布应用规范, 其中, 在分布式互操作协议上, J2EE 同时支持 RMI 和 IIOP, 而在服务器端分布式应用的构造形式, 则包括了 Java Servlet、JSP(Java Server Page)、EJB 等多种形式, 以支持不同的业务需求, 而且 Java 应用程序具有“Write once, run anywhere”的特性, 使得 J2EE 技术在发布计算领域得到了快速发展。

EJB 是 Sun 推出的基于 Java 的服务器端构件规范 J2EE 的一部分, 自从 J2EE 推出之后, 得到了广泛的发展, 已经成为应用服务器端的标准技术。Sun EJB 技术是在 Java Bean 本地构件基础上, 发展的面向服务器端分布应用构件技术。它基于 Java 语言, 提供了基于 Java 二进制字节代码的重用方式。EJB 给出了系统的服务器端分布构件规范, 这包括了构件、构件容器的接口规范以及构件打包、构件配置等的标准规范内容。EJB 技术的推出, 使得用 Java 基于构件方法开发服务器端分布式应用成为可能。从企业应用多层结构的角度, EJB 是业务逻辑层的中间件技术, 与 JavaBeans 不同, 它提供了事务处理的能力, 从分布式计算的角度, EJB 像 CORBA 一样, 提供了分布式技术的基础。提供了对象之间的通信手段。从 Internet 技术应用的角度, EJB 和 Servlet, JSP 一起成为新一代应用服务器的技术标准, EJB 中的 Bean 可以分为会话 Bean 和实体 Bean, 前者维护会话, 后者处理事务, 现在 Servlet 负责与客户端通信, 访问 EJB, 并把结果通过 JSP 产生页面传回客户端。

J2EE 的优点是, 服务器市场的主流还是大型机和 UNIX 平台, 这意味着以 Java 开发构件, 能够做到“Write once, run anywhere”, 开发的应用可以配置到包括 Windows 平台在内的任何服务器端环境中去。

3) Microsoft DNA 2000

Microsoft DNA 2000(Distributed interNet Applications)是 Microsoft 在推出 Windows

2000 系列操作系统平台基础上,在扩展了分布计算模型,以及改造 Back Office 系列服务器端分布计算产品后发布的新的分布计算体系结构和规范。

在服务器端,DNA 2000 提供了 ASP、COM、Cluster 等的应用支持。目前,DNA 2000 在技术结构上有着巨大的优越性。一方面,由于 Microsoft 是操作系统平台厂商,因此 DNA 2000 技术得到了底层操作系统平台的强大支持;另一方面,由于 Microsoft 的操作系统平台应用广泛,支持该系统平台的应用开发厂商数目众多,因此在实际应用中,DNA 2000 得到了众多应用开发商的采用和支持。

DNA 2000 融合了当今最先进的分布计算理论和思想,如事务处理、可伸缩性、异步消息队列、集群等内容。DNA 使得开发可以基于 Microsoft 平台的服务器构件应用,其中,如数据库事务服务、异步通信服务和安全服务等,都由底层的分布对象系统提供。以 Microsoft 为首的 DCOM/COM/COM+ 阵营,从 DDE、OLE 到 ActiveX 等,提供了中间件开发的基础,如 VC、VB、Delphi 等都支持 DCOM,包括 OLE DB 在内新的数据库存取技术,随着 Windows 2000 的发布,Microsoft 的 DCOM/COM/COM+ 技术,在 DNA2000 分布计算结构基础上,展现了一个全新的分布构件应用模型。首先,DCOM/COM/COM+ 的构件仍然采用普通的 COM(Component Object Model)模型。COM 最初作为 Microsoft 桌面系统的构件技术,主要为本地的 OLE 应用服务,但是随着 Microsoft 服务器操作系统 NT 和 DCOM 的发布,COM 通过底层的远程支持使得构件技术延伸到了分布应用领域。DCOM/COM/COM+ 更将其扩充为面向服务器端分布应用的业务逻辑中间件。通过 COM+ 的相关服务设施,如负载均衡、内存数据库、对象池、构件管理与配置等,DCOM/COM/COM+ 将 COM、DCOM、MTS 的功能有机地统一在一起,形成了一个概念、功能强的构件应用体系结构。而且,DNA 2000 是单一厂家提供的分布对象构件模型,开发者使用的是同一厂家提供的系列开发工具,这比组合多家开发工具更有吸引力。

但是它的不足是依赖于 Microsoft 的操作系统平台,因而在其他开发系统平台(如 UNIX、Linux)上不能发挥作用。

虽然这三种平台因为其形成的历史背景和商业背景有所不同,各自有自己的侧重和特点,其实在它们之间也有很大的相通性和互补性。

4. 中间件是实现电子商务的基础软件

电子商务的本质就是对处于分布环境中的各种计算机系统进行交流协调,从而开创新的商业运作模式。网络通信,尤其是互联网技术,是电子商务的通信基础,而管理和传输系统之间的业务信息、协调各个系统的处理模块的中间管理服务系统,是保证电子商务应用成功的关键。电子商务应用服务器、通用业务网关、支付网关、通信平台和安全平台,统一纳入电子商务中间件构架的范畴。

从技术角度看,电子商务将由 Internet/Intranet 技术、传统 IT 技术以及具体的业务处理所构成。但是,系统的建立将会面临许多新的问题,包括应用系统能不能快速地建立,能不能适应大用户数、高处理量要求,能不能提供高效率、高可靠性、高可用性等关键任务的要求,能不能满足安全需要等。

以上这些问题,只是依靠简单的 Web 技术是不够的。目前常用的 Web 技术由于早期更多的是面向信息发布,因此存在并发访问瓶颈、难扩展、效率低、安全等诸多问题,不能满足电子商务的需要。为了很好地解决这些问题,需要以 Web 的低层技术为基础,规划出一

个整体的应用框架,并提供一个支持平台,用于 Internet 应用的开发、部署和管理,并能借此解决上述各种问题。这已经发展成为一个能广泛适应的标准的支撑层,成为 Internet 应用的基础设施(Infrastructure),这一支撑层实际上是基于 Internet 的中间件,也就是应用服务器。

同时,由于企业并不能把业务一步跨到 Internet 上,而必须同传统的应用系统结合,因此也必须通过中间件来集成 Web 应用和传统应用,实现完整的电子商务。

电子商务中间件构架是一种电子商务应用集成的关键件,不管电子商务应用分布在什么硬件平台上,使用了什么数据库系统,透过什么复杂的网络,电子商务应用的互连和互操作是电子商务中间件构架首先要解决的问题。在通信方面,电子商务中间件构架要支持各种通信协议和通信服务模式,传输各种数据内容,数据格式翻译、流量控制、数据加密、数据压缩等;在电子商务中间件构架核心,要解决名字服务、安全控制、并发控制、可靠性和效率保证等;在电子商务应用开发方面,要能提供基于不同平台的丰富的开发接口,支持流行的开发工具和异构互连接口标准等;在管理方面,解决电子商务中间件构架本身的配置、监控、调谐,为电子商务应用的易用易管理提供保证。

其次,针对不同的 Web 应用环境,对电子商务中间件构架有各种不同的要求。对 workflow 应用,需要根据条件以及条件满足状态,将信息、响应状态从一个应用传递到另一个应用;对联机事务处理,需要保证分布式的数据一致性、不停机作业、大量并发的高效率;对于一个数据采集系统需要保证可靠传输等。

3.4 数据处理技术

电子计算机诞生于 1946 年,在其后的一段时间里,计算机主要用于科学和工程技术领域的数值计算。但随着社会生产力和文明的不断发展,信息在人类社会活动中起着越来越重要的作用。20 世纪 50 年代初期,人们开始用计算机进行数据处理。数据处理是对各种形式的数据进行收集、存储、加工和传输等活动的总称,数据管理是指数据收集、分类、组织、编码、存储、检索、传输和维护等环节,是数据处理的基本操作,称为数据管理。数据管理是数据处理的核心问题。

3.4.1 数据处理技术的发展过程

数据处理技术是对数据的分类、组织、存储、操作和维护的技术。在计算机数据处理中,简单地说计算机是数据处理机,输入原始数据,经过计算机的处理,获得我们所需要的信息。40 多年来,数据处理技术随着计算机软、硬件的发展而不断地发展,它大致经历了人工管理、文件系统和数据库系统三个阶段。

1. 人工管理阶段

这是计算机用于数据处理的初级阶段。在该阶段,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的存储和组织,要为数据分配空间,决定存取方法,应用程序完全依赖于数据。数据结构的改变,数据存取的物理地址或存储设备的变化,都会要求修改相应的应用程序。也就是说,应用程序和数据一一对应,数据和处理它的应用程序混为一

个整体,如图 3.20 所示。由于数据的物理组织是由应用程序员根据应用的要求设计的,故很难实现多个应用程序共享数据资源,造成数据的大量重复。这一时期数据的处理主要是手工性质的,其数据处理的缺点是显而易见的,即数据独立性差、冗余度很高等,从而造成数据分散、处理效率低,维护困难。

2. 文件系统管理阶段

计算机操作系统中引入文件管理系统后,如图 3.21 所示,对上述问题有了较大的改进。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联系,数据的物理结构和逻辑结构间实现了转换,从而提高了数据的物理独立性。在文件系统中,还提供了多种文件组织形式,如顺序文件组织、索引文件组织和直接存取文件组织等。



图 3.20 人工管理方式的数据与程序的关系

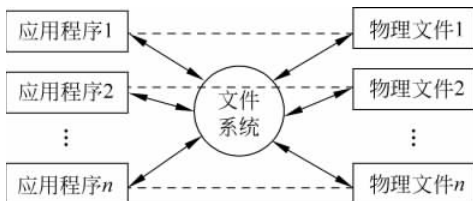


图 3.21 文件系统方式数据与程序的关系

在这一阶段,实现了以文件为单位的数据共享。但未能实现以记录或数据项为单位的数据共享,数据仍然是分散的,是面向应用的,所以数据还存在大量的冗余,应用程序和数据结构之间相互依赖程度高,数据的完整性和安全性等无法得到保证。

3. 数据库系统管理阶段

由于文件管理方式的弊端,20 世纪 60 年代后期出现了数据库技术,数据库技术的目标是,克服程序与数据文件的相互依存,力求数据独立,尽量克服数据的冗余,还要解决数据的安全和维护问题。

数据库系统是在文件系统的基础上发展起来的新技术,它克服了文件系统的缺点。解决了冗余和数据依赖问题,提供了更广泛的数据共享,为应用程序提供了更高的独立性,保证了数据的完整性和安全性,并为用户提供了方便的用户接口,如图 3.22 所示。

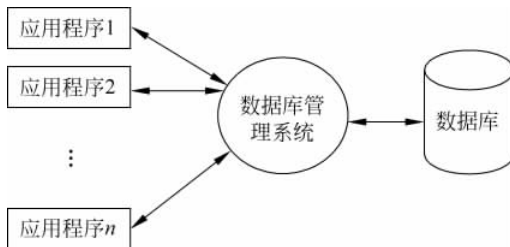


图 3.22 应用程序与数据库的关系

数据处理技术发展的三个阶段最明显的特征是数据与应用程序的联系程度不同,数据处理技术发展的目标是增加数据与应用程序的独立性,减少数据和应用程序的联系,从而增强数据共享性,减少数据的冗余;增强数据之间的联系,保证数据的一致性和安全性。

3.4.2 数据库技术概述

随着信息社会发展,要求计算机不但能进行科学计算,而且能进行大量数据的简单处理(如数据的查询和更新),使计算机的应用从科学研究部门逐步扩展到企业、行政部门,因而产生了数据库技术。数据库技术是计算机应用领域的重要分支,其核心任务是进行数据管理。数据库技术产生于20世纪60年代末,现已形成相当规模的理论体系和使用技术。

1. 数据库基本概念

1) 数据库

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。而在数据库系统中的数据被集中进行管理,就像仓库中的货物一样,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数据系统称为“数据库”(Data Base)。严格地说,数据库是以一定组织方式存储在一起的相关数据的集合,它以最优的方式为一个或多个应用服务;数据的存储独立于使用它的程序;对数据的输入、输出和修改均能按一种公用的和可控制的方法进行;数据被结构化,为今后的应用研究提供基础。

2) 数据库管理系统

数据库管理系统(Data Base Management System,DBMS)是处理数据库存取和各种管理控制的软件系统,是数据库系统的中心枢纽。用户对数据库进行的各种操作——数据库的建立、使用和维护,都是在DBMS的统一管理和控制下进行的。因为有了DBMS负责处理数据库和用户程序间的接口,所以用户不必注重数据的逻辑和物理表达细节,只需注意数据的内容就可以了。Oracle便是这样的数据库管理系统。数据库管理系统通常由三部分组成:

(1) 描述语言及其翻译程序。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存储描述语言。主要用来定义数据库的逻辑结构,描述物理数据库的结构。

例如,某人事部门要将本单位人员的基本信息管理起来,这些信息包括编号、部门名、姓名、性别等。将本单位所有职工的相应信息汇总起来存储在计算机中之后,就形成了一个人事数据库。但是这些数据在计算机中的存储格式并不是杂乱的,为了有效地对它们进行管理、维护和使用,必须建立一个结构,并设计出数据在计算机中的物理存储方法。

(2) 数据操纵语言及其翻译程序。它完成对数据库的基本操作,即检索、插入、删除、修改等。我们建立数据库的主要目的就是为了通过这些操作来存取相关信息。

(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。它包括系统主控程序、存取控制程序、并发控制程序和数据库服务程序等。

3) 数据库应用系统

数据库应用系统是指利用计算机和数据库技术建立起来的、面向各种数据处理应用的计算机软件系统,如人力资源管理系统、企业管理信息系统等。

4) 数据库系统

数据库系统(Data Base System,DBS)是指计算机系统引入数据库之后组成的系统,是用来组织和存取大量数据的管理系统。它是由计算机系统(硬件和基本软件)、数据库、数据库管理系统、数据库应用系统和有关人员(数据库管理员、应用设计人员、最终用户)组成的具有高度组织性的总体。数据库系统管理数据的方式较文件系统有许多不同,它应包括如

下特征:

(1) 数据共享。数据库中的数据可以供多个用户所使用,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。应用程序不再同物理存储器上具体的文件相对应,每个用户所使用的数据有其自身的逻辑结构。数据独立性表现在物理独立性和逻辑独立性两个方面。它给数据库的使用、调整、维护和扩充带来了方便,提高了数据库应用系统的稳定性,减轻了程序员的负担。

(3) 减少数据冗余。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。

(4) 数据的结构化。数据库系统中的数据是相互关联的,这种关联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。整个数据库是以一定的形式构成的。

(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、一致性和并发控制问题。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。同时,还可以规定密码和用户权限。另外,不同用户同时使用数据库,可能造成数据的不一致,数据库系统具有并发控制功能,以保证数据的正确性。

此外,系统还提供其他的数据保护措施,如数据的有效性检查、故障恢复等来保证数据的正确性。

2. 数据库技术的应用与发展

随着计算机技术、通信技术和网络技术的飞速发展,信息系统渗透到社会的各个领域,作为其核心和基础的数据库技术也得到了越来越广泛的应用。数据库的建设规模、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。

近年来,随着多媒体技术、空间数据库技术和计算机网络的飞速发展,数据库系统的发展十分迅速,应用领域愈来愈广,企事业单位、政府部门的行政管理、办公自动化;企业生产计划管理;军队物资管理;银行财务管理;铁路、民航飞机票预订系统;铁路车次调度系统;宾馆、酒店房间预订系统;百货公司订购与销售系统;医院病房、病历管理;学校师资、教学、学生学籍管理;图书馆管理;政府部门的计划和统计系统;人口普查;气象预报;地震,勘探等大量数据的存储和统计分析;公安部门罪犯档案管理;地理信息系统(GIS);电子商务等都离不开数据库系统。

数据库处理在信息系统的研究中一直是非常重要的主题,数据库系统仍然处于迅速发展的青春时期,充满了活力。一方面,一些较成熟的技术,如各种大、中、小和微型计算机数据库管理系统和一些传统的数据库设计方法已投入实用;另一方面,尚有许多理论及实际问题亟待完善、开发和探索,如空间数据库、多媒体数据库、网络数据库、智能数据库等。特别是网络数据库的一些理论和技术问题正在成为数据库研究的一个热点问题,传统的数据库技术和系统显得力不从心,这对传统的技术和研究开发工作提出了挑战。为应付这些挑战,一方面需要将原有的思想和技术进行扩充、推广和转移来解决面临的问题。另一方面,需要研究新的技术,提出新的数据管理概念。这两方面相互结合,可为21世纪数据库技术的研究开发开辟新的局面。数据库系统这门学科目前处于既需普及,又待提高的状态。

3.4.3 数据仓库与数据挖掘

1. 数据仓库

目前,数据仓库一词尚没有一个统一的定义,著名的数据仓库专家 W. H. Inmon 在其著作《Building the Data Warehouse》一书中给予如下描述:数据仓库(Data Warehouse)是一个面向主题的(Subject Oriented)、集成的(Integrate)、相对稳定的(Non-Volatile)、反映历史变化(Time Variant)的数据集合,用于支持管理决策。

对于数据仓库的概念我们可以从两个层次予以理解,首先,数据仓库用于支持决策,面向分析型数据处理,它不同于企业现有的操作型数据库;其次,数据仓库是对多个异构的数据源有效集成,集成后按照主题进行了重组,并包含历史数据,而且存放在数据仓库中的数据一般不再修改。由数据仓库概念的含义可知,数据仓库拥有以下四个特点:

1) 面向主题

操作型数据库的数据组织面向事务处理任务,各个业务系统之间各自分离,而数据仓库中的数据是按照一定的主题域进行组织。主题是一个抽象的概念,是指用户使用数据仓库进行决策时所关心的重点方面,一个主题通常与多个操作型信息系统相关。

2) 集成的

面向事务处理的操作型数据库通常与某些特定的应用相关,数据库之间相互独立,并且往往是异构的。而数据仓库中的数据是在对原有分散的数据库数据抽取、清理的基础上经过系统加工、汇总和整理得到的,必须消除源数据中的不一致性,以保证数据仓库内的信息是关于整个企业的一致全局信息。

3) 相对稳定的

操作型数据库中的数据通常实时更新,数据根据需要及时发生变化。数据仓库的数据主要供企业决策分析之用,所涉及的数据操作主要是数据查询,一旦某个数据进入数据仓库以后,一般情况下将被长期保留,也就是数据仓库中一般有大量的查询操作,但修改和删除操作很少,通常只需要定期地加载、刷新。

4) 反映历史变化

操作型数据库主要关心当前某一个时间段内的数据,而数据仓库中的数据通常包含历史信息,系统记录了企业从过去某一时点(如开始应用数据仓库的时点)到目前的各个阶段的信息,通过这些信息,可以对企业的发展历程和未来趋势做出定量分析和预测。

整个数据仓库系统是一个包含四个层次的体系结构,具体如图 3.23 所示。

数据仓库不同于数据库。数据库是一种通用的平台,用来管理企业的数据;而数据仓库是一种概念,在此概念下进行的构造过程,我们叫它数据仓库处理。企业数据仓库的建设,是以现有企业业务系统和大量业务数据的积累为基础。数据仓库不是静态的概念,只有把信息及时交给需要这些信息的使用者,供他们做出改善其业务经营的决策,信息才能发挥作用,信息才有意义。而把信息加以整理归纳和重组,并及时提供给相应的管理决策人员,是数据仓库的根本任务。因此,从产业界的角度看,数据仓库不是花钱可以购买的现成产品(而数据库是),数据仓库建设是一个工程,是一个过程。

2. 数据挖掘

当今数据库的容量已经达到上万亿的水平(T)——1 000 000 000 000 个字节。在这些

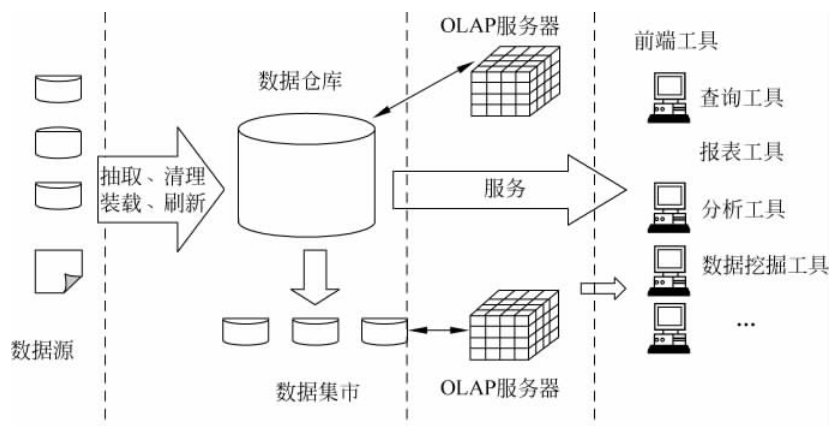


图 3.23 数据仓库的体系结构

大量数据的背后隐藏了很多具有决策意义的信息,那么怎么得到这些“知识”呢?也就是怎样通过一棵棵的树木了解到整个森林的情况?那就是数据挖掘。数据挖掘是一个利用各种分析工具在海量数据中发现模型和数据间关系的过程,这些模型和关系可以用来做出预测。数据挖掘的目的是提高市场决策能力;检测异常模式;在过去的经验基础上预言未来趋势等。

数据挖掘,在“数据矿山”中找到蕴藏的“知识金块”,帮助企业减少不必要投资的同时提高资金回报。数据挖掘给企业带来的潜在的投资回报几乎是无止境的。世界范围内具有创新性的公司都开始采用数据挖掘技术来判断哪些是他们的最有价值客户、重新制定他们的产品推广策略(把产品推广给最需要他们的人),以用最小的花费得到最好的销售。

1) 数据挖掘的思路

构建数据仓库的过程就是根据预先设计好的逻辑模式从分布在企业内部各处的数据库中提取数据并对经过必要的变换最终形成全企业统一模式数据的过程。主要有如下三个步骤。

(1) 描述数据。计算统计变量(比如平均值、均方差等),再用图表或图片直观地表示出来,进而可以看出一些变量之间的相关性(比如有一些值经常同时出现)。

(2) 建立预言模型。用历史数据建立一个预言模型,然后再用另外一些数据对这个模型进行测试。

(3) 验证模型。先对一小部分客户做一个实际的测试,然后再决定。

2) 数据挖掘和数据仓库

大部分情况下,数据挖掘都要先把数据从数据仓库中拿到数据挖掘库或数据集中,数据挖掘库可能是数据仓库的一个逻辑上的子集,而不一定非得是物理上单独的数据库。

3) 数据挖掘应用

由于数据挖掘带来的显著的经济效益,使数据挖掘越来越普及。它不仅能用于控制成本,也能给企业带来效益。很多企业都在利用数据挖掘技术帮助管理客户生命周期的各个阶段,包括争取新的客户、在已有客户的身上赚更多的钱和保持住好的客户。如果能够确定客户的特点,就能为客户提供针对性的服务。比如,已经发现了购买某一商品的客户的特

征,那么就可以向那些具有这些特征但还没有购买此商品的客户推销这个商品;找到流失的客户的特征就可以,在那些具有相似特征的客户还未流失之前进行针对性的弥补,因为保留一个客户要比争取一个客户便宜得多。

数据挖掘可以应用在各个不同的领域。电信公司和信用卡公司是用数据挖掘检测欺诈行为的先行者。保险公司和证券公司也开始采用数据挖掘来减少欺诈。医疗应用是另一个前景广阔的产业,即数据挖掘可以用来预测外科手术、医疗试验和药物治疗的效果。零销商更多地使用数据挖掘来决定每种商品在不同地点的库存,通过数据挖掘更灵活地使用促销和优惠券手段。制药公司通过挖掘巨大的化学物质和基因对疾病影响的数据库来判断哪些物质可能治疗某种疾病产生效果。

3. 联机分析处理

随着数据库技术的广泛应用,企业信息系统产生了大量的数据,如何从这些海量数据中提取对企业决策分析有用的信息成为企业决策管理人员所面临的重要难题。传统的企业数据库系统(管理信息系统)作为数据管理手段,主要用于事务处理,但它对分析处理的支持一直不能令人满意。因此,人们逐渐尝试对数据库中的数据进行再加工,形成一个综合的、面向分析的、更好地支持决策制定的决策支持系统(Decision Support System,DSS)。

1) 联机分析处理的概念

联机分析处理的概念最早由关系数据库之父 E. F. Codd 于 1993 年提出的。Codd 认为用户的决策分析需要对关系数据库进行大量计算才能得到结果,而查询的结果并不能满足决策者提出的需求。因此,Codd 提出了多维数据库和多维分析的概念,即 OLAP。OLAP 委员会对联机分析处理的定义为:使分析人员、管理人员或执行人员能够从多种角度对从原始数据中转化出来的、能够真正为用户所理解的、并真实反映企业维特性的信息进行快速、一致、交互地存取,从而获得对数据的更深入了解的一类软件技术。OLAP 的目标是满足决策支持或多维环境特定的查询和报表需求,它的技术核心是“维”这个概念,因此 OLAP 也可以说是多维数据分析工具的集合。

2) 数据挖掘与联机分析处理

一个经常问的问题是,数据挖掘和 OLAP 到底有何不同。它们的区别是使用完全不同的工具,基于的技术也大相径庭。

OLAP 是决策支持领域的一部分。传统的查询和报表工具是告诉你数据库中都有什么(What happened),OLAP 则更进一步告诉你下一步会怎么样(What next),和如果我采取这样的措施又会怎么样(What if)。用户首先建立一个假设,然后用 OLAP 检索数据库来验证这个假设是否正确。也就是说,OLAP 分析师是建立一系列的假设,然后通过 OLAP 来证实或推翻这些假设来最终得到自己的结论。OLAP 分析过程在本质上是一个演绎推理的过程。但是如果分析的变量达到几十或上百个,那么再用 OLAP 手动分析验证这些假设将是一件非常困难和痛苦的事情。

数据挖掘与 OLAP 不同的地方是,数据挖掘不是用于验证某个假定的模式(模型)的正确性,而是在数据库中自己寻找模型。它在本质上是一个归纳的过程。

数据挖掘和 OLAP 具有一定的互补性。在利用数据挖掘出来的结论采取行动之前,你也许要验证一下如果采取这样的行动会给公司带来什么样的影响,那么 OLAP 工具能回答你的这些问题。

3.4.4 数据处理技术与电子商务

20 世纪 90 年代以来,Internet 日益普及,Web 成为最流行、最大的网络信息系统,并以惊人的速度继续发展。Web 技术和数据库技术的结合,产生了互联网数据库这一新兴的数据库应用领域。

数据库技术是电子商务平台建设的一项支撑技术,它提供了对电子商务的全方位的支持。数据库对电子商务的支持主要表现在:提供电子商务中各种数据的存储和管理,为电子商务提供决策支持等几方面。

1. 数据的收集、存储和组织

数据是企业的重要资源,特别是在电子商务时代,数据的重要性就更显突出,它是企业各类生产经营活动和日常管理的基础,是企业决策的依据。而数据的收集、组织和存储是数据库系统的主要功能。

2. 对 EDI 的支持

EDI 在企业间自动传递商务信息,是企业电子商务的一个重要组成部分。EDI 以企业信息系统为基础,而企业信息系统的核心是数据库系统。

3. 决策支持

企业电子商务所面对的是一个全球市场,在 Internet 上企业可以获得更多信息,这是一个机遇,也是一个挑战,企业应该充分利用电子商务的海量数据进行分析,并依据分析结果做出正确决策。电子商务系统如果缺少好的决策支持,一方面是信息资源的巨大浪费,另一方面也是企业的一大损失。

4. 对 Web 的支持

早期 WWW 上大多数信息是静态的,信息的内容完全由信息的提供者决定,信息的变化只能有网络管理人员来进行更动,用户往往在被动状态,因此缺少互动性。

电子商务活动需要大量的动态数据,这就使网站从处理“文件型”的数据,进而发展到需要结合数据库系统,建立网上联机数据库,以满足多方位的需求。通过数据库技术,不仅可以采用数据库管理,通过更新数据库中的数据,自动达到更新网页;还要将网站的内容存储在数据库中,利用数据库管理,通过更新数据库的搜索功能,增强网上搜索功能;又能使使用者借助浏览器,通过 Internet 或 Intranet,存取 Web 数据库的数据,以实现各种基于 Web 数据库的应用。

3.5 Web 开发技术

Web 是 WWW(World Wide Web)的简称,中文意思是万维网。我们给 Web 下的定义是:Web 是建立在客户机/服务器模型之上,以 HTML 语言和 HTTP 协议为基础,能够提供面向各种 Internet 服务的、一致的用户界面的信息浏览系统。在 WWW 中可以传送声音、图形、图像、动画,而且这些媒体可以同步出现。

3.5.1 Web 概述

1. Web 的产生与发展

在 WWW 创立之前,Internet 上的信息传递和发布都是通过 E-mail、FPT 等实现的,但它们对信息的交流存在各种各样的限制,Internet 的作用还不能完全发挥。人们希望寻求一种好的信息结构框架、一种独立于各种平台的方法,以便在 Internet 上高效地传递信息。

20 世纪 80 年代末,出现了许多帮助人们分类查找信息的工具,其中最大的突破性工具就是 Web。Web 是 1989 年在日内瓦的欧洲粒子物理实验室首先提出来的。一位名叫 Tim Berners Lee 的物理学家为了让物理学家们快速地进行交流,特别是能让大家共享他们随时的实验进展报告,他想建立文件连接网络,以便让读者随意地在文件间跳跃。于是诞生了超文本的概念。

在超文本系统中,信息被组织成称作网页的信息文件,信息之间的关系不是层次的,不靠菜单链接,而是一种新的链接关系。链接是信息之间的一种逻辑对应关系,表示了信息之间的关联性。它使得文本不再像一本书一样是固定的线性的,而是可以从一个位置跳到另外的位置。你可以从中获取更多的信息,可以转到别的主题上,想要了解某一个主题的内容只要在这个主题上点一下,就可以跳转到包含这一主题的文档上。例如,一个文件中的某个词可以与这个词的解释文件相关联。今天的网页中不仅有文本信息,还有图形、声音以及影像等多媒体信息,所以它们又称为超媒体文件。

Web 就是一种超文本信息系统,它采用超文本和超媒体的信息组织方式,将信息的链接扩展到整个 Internet 上。在 Internet 中,存放网页并提供网页浏览服务的服务器主机称作 Web 服务器。

运行在客户端用于浏览 Web 网页的软件就是浏览器,它是浏览 Internet 上的文本、图像、声音的主要工具。在 1993 年,伊利诺斯州立大学的超级计算应用国家中心发布了 Mosaic 第一个图形浏览器,随后诞生了众多的浏览器,我们今天常用的有 Microsoft Internet Explorer 等。

2. Web 的特点

Web 对信息的逻辑组织方式,以及它所有追求的目标,使 Web 具有显著特色,在它诞生不长的时间内就得到了广泛的应用,今天,WWW 几乎已成为 Internet 的代名词,也是如它的特点和功能分不开的。Web 有如下特点。

1) 分布式的信息资源

Internet 的信息资源具有极强的分布特征,这些资源通过信息指针可以引导用户端的客户机程序从一台计算机转到另一台计算机,这种转移对用户是透明的,为用户访问 Internet 资源提供了一个简单的、直观的界面。

2) 统一的用户界面

用于采用客户机/服务器的工作方式,利用应用层网络协议进行信息交换,而且提供了 CGI 通用网关接口,因此可以支持各种计算机、各种操作系统、各种用户界面以及各种信息服务。

3) 支持各种信息资源和各种媒体的演播

Internet 信息资源具有不同的信息结构,WWW 可提供包括文本、图像、声音、动画和视

频等多种类型的信息服务。

4) 强大的功能

今天,Web 上的服务日益增多,在 Web 数据库等技术的支持下,它已把 E-mail、FTP、BBS 等各种功能融合在一起,并且其功能还在不断地发展。

5) 广泛的用途

包括各种组织介绍;电子出版物、电子图书馆和博物馆;虚拟现实;个人信息发布等。

3. Web 相关概念

Web 是一个新生的事物,有许多相关的新概念,在此,让我们先来了解几个相关的概念。以便于进一步的学习。

1) Web Page(网页)

网页主要是由文字和图形构成的,让使用者在家中的电脑用浏览器观看图文并茂的网上作品,这传统书籍的表现形式极为相似,但其实不只是如此,网页还可以结合动画、音乐构成一页页活动的电子图书。实际上,Internet 的广受欢迎和使用者可以看到丰富多彩的网页有着密切的关系。

当浏览某个网站的时候,通常会看到这个网站的第一页,这个网站的第一页就被称为这个网站的主页(Home Page)。

2) Web Site(网站)

Web Site 形如那些放置网页供大家登录浏览的电脑。

3) Browser(浏览器)

浏览器是浏览 Internet 上的文本、图像、声音的主要工具,从通信协议看其提供响应和服务。Mosaic 是 WWW 中第一个流行的浏览程序。但由于其不能支持某些较高级的 HTML 特性,现已被 Netscape 和 Microsoft Internet Explorer 所取代。

4) Hyperlink(超链接)

超链接是通过统一资源定位符,在 Web 信息之间构造了一种逻辑联系。超链接的概念就是在编写网页过程中,编写者预先设定一些关键字,然后再提供这些关键字想要链接的目的地,这些目的地可以是一段文字,一张图片,一个网页或一个网站,这取决于文件编写者如何去达到其文件的表现形式。在网页设计上,大量使用这种超链接,使得网页变成一种立体化的文件,并更加生动有趣,这些都是传统文字、书籍做不到的事。

5) URL(Uniform Resource Locator,统一资源定位符)

URL 是 WWW 页的地址,就是 Internet 中某主机上文件的唯一地址。它是按照 Internet 的一般定义方式,它从左到右由下述四部分组成,例如,http://www.yahoo.com/。

(1) Internet 资源类型(scheme)。指定使用的传输协议,指出 WWW 客户程序用来操作的工具。如“http”表示 WWW 服务器,“ftp”表示 FTP 服务器,“gopher”表示 Gopher 服务器,其后用“//”分隔。

(2) 服务器地址(host)。指出 Web 页面所在的服务器主机名和域名。如 www.yahoo.com 表示商业机构(一级域名 com)下企业域 yahoo 用作传递网页的 www 主机。

(3) 端口(port)。对某些资源的访问来说,需给出相应的服务器的端口号。如,http://210.77.146.154:8080/iTCMedu/index.jsp,8080 就是端口号。端口可以默认。

(4) 路径(path)。指明服务器上某资源的位置,通常由“目录/子目录/文件名”这样结

构组成。与端口一样,路径并非总是需要的。

URL 地址格式排列为: scheme: //host: port/path, 例如 http: //www. iTCMedu. com/index. html 就是一个典型的 URL 地址。

6) HTML(HyperText Markup Language,超文本标记语言)

HTML 是一种专门的编程语言,用于创建存储在 WWW 服务器上的网页,并能由 Mosaic、Netscape、Microsoft Explorer 等浏览器进行浏览。

7) HTTP 协议

HTTP(Hypertext Transfer Protocol),即超文本传输协议。是 WWW 浏览器和 WWW 服务器之间的应用层通信协议,定义服务器端和客户端之间文件传输的沟通方式。HTTP 是最广为人知的协议,它位于 TCP/IP 协议的应用层,也是互联网中最核心的协议之一。它不仅保证正确传输超文本文档,还确定传输文档中的哪一部分,以及哪一部分内容首先显示(如文本先与图形)等。

8) SMTP 和 POP3

简单邮件传输协议(Simple Mail Transfer Protocol,SMTP)和邮局协议(Post Office Protocol version 3,POP3)是基于 TCP/IP 网络的协议,用于主机与主机之间的电子邮件交换。

3.5.2 Web 工作原理

WWW 上的每个网页都对应一个文件。我们浏览一个页面,要先把页面所对应的文件从提供这个文件的计算机里,通过 Internet 传送到我们自己的计算机中,再由 WWW 浏览器翻译成为我们见到的有文字、有图形甚至有声音的页面。

1. Web 工作原理

Web 的工作基于客户机/服务器计算模型,由 Web 浏览器(客户机)和 Web 服务器(服务器)构成,两者之间采用超文本传送协议(HTTP)进行通信。HTTP 协议是 Web 浏览器和 Web 服务器之间的应用层协议,是通用的、无状态的、面向对象的协议。

Web 工作的基本原理是:由浏览器向 WWW 服务器发出 HTTP 请求,WWW 服务器接到请求后,进行相应的处理,将处理结果以 HTML 文件的形式返回给浏览器,客户浏览器对其进行解释并显示给用户。

一个完整的 HTTP 协议会话过程包括四个步骤(如图 3.24 所示):

1) 连接

Web 浏览器与 Web 服务器建立连接,打开一个称为 Socket(套接字)的虚拟文件,此文件的建立标志着连接建立成功。

2) 请求

Web 浏览器通过 Socket 向 Web 服务器提交请求。HTTP 的请求一般是 GET 或 POST 命令(POST 用于 FORM 参数的传递)。

3) 应答

Web 浏览器提交请求后,通过 HTTP 协议传送给 Web 服务器。Web 服务器接到后,进行事务处理,处理结果又通过 HTTP 传回给 Web 浏览器,从而在 Web 浏览器上显示出所请求的页面。

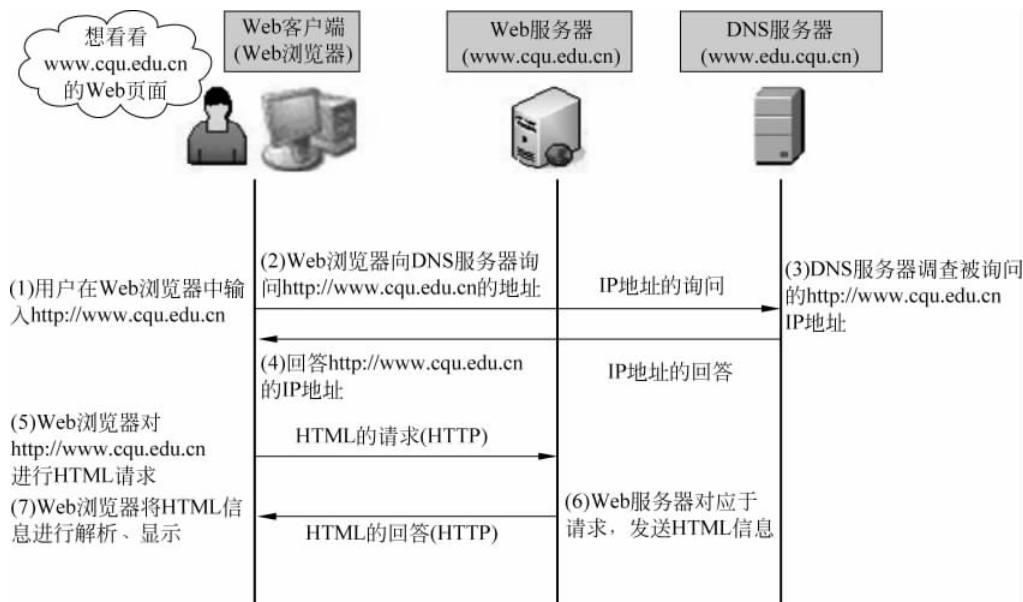


图 3.24 Web 工作原理与流程

4) 关闭连接

应答结束后 Web 浏览器与 Web 服务器必须断开,以保证其他 Web 浏览器能够与 Web 服务器建立连接。

在这一过程中,要涉及两个不同的程序:一个程序安装在客户机上,它执行用户命令,发出 HTTP 请求;接到响应后,立即显示相应网页内容,确保用户的要求得到执行。这个程序叫做 WWW 客户机程序,比如上网所使用的浏览器(IE 或 Netscape)。另一个程序在服务器上,称作 Web 服务器程序,如 IIS 或“阿帕奇(Apache)”Web 服务器软件,它对 WWW 客户机所要求的一切进行满足,也就是接到 HTTP 请求后,发出响应。

2. Web 服务器

Web 服务器是指驻留于互联网上某种类型计算机的程序。通俗地讲,Web 服务器专门处理浏览器发来的 HTTP 请求(request),并传送(serves)页面使浏览器可以浏览。当 Web 浏览器(客户端)连到服务器上并请求文件时,服务器将处理该请求并将文件发送到该浏览器上,附带的信息会告诉浏览器如何查看该文件(即文件类型)。服务器使用 HTTP(超文本传输协议)进行信息交流。

Web 服务器不仅能够存储信息,还能在用户通过 Web 浏览器提供的信息的基础上运行脚本和程序。例如,假设你要提供免费公司资讯,只要建立一张免费请求表单,它就会要求你的读者输入邮寄及公司信息;读者填完表后,单击提交按钮,该表单将送至服务器计算机上的某一程序,它负责处理该请求,并用 E-mail 给读者发一份免费资讯。你还可以让该程序把客户提供的信息发给你,以便用于某数据库上。用于执行这些功能的程序或脚本称为网关脚本/程序,或称为 CGI(通用网关界面)脚本。在 Web 上大多数表单和搜索引擎上都使用了该技术。

Web 服务器可驻留于各种类型的计算机,从常见的 PC 到巨型的 UNIX 网络,以及其他

各种类型的计算机。它们通常经过一条高速线路与因特网连接,如果对性能无所谓,则也可使用低速连接(甚至是调制解调器),但对于架设电子商店来说,性能绝对是要考虑的问题。

如今互联网的 Web 平台种类繁多,各种软硬件组合的 Web 系统更是数不胜数,在 UNIX 和 Linux 平台下使用最广泛的免费 HTTP 服务器是 W3C、NCSA 和 APACHE 服务器,而 Windows 平台 NT/2000/2003 使用 IIS 的 Web 服务器。在选择使用 Web 服务器应考虑本身特性因素有性能、安全性、日志和统计、虚拟主机、代理服务器、缓冲服务和集成应用程序等,下面介绍几种常用的 Web 服务器。

1) Microsoft IIS

Microsoft 的 Web 服务器产品为 Internet Information Server(IIS),IIS 是允许在公共 Intranet 或 Internet 上发布信息的 Web 服务器。IIS 是目前最流行的 Web 服务器产品之一,很多著名的网站都是建立在 IIS 的平台上。IIS 提供了一个图形界面的管理工具,称为 Internet 服务管理器,可用于监视配置和控制 Internet 服务。

IIS 是一种 Web 服务组件,其中包括 Web 服务器、FTP 服务器、NNTP 服务器和 SMTP 服务器,分别用于网页浏览、文件传输、新闻服务和邮件发送等方面,它使得在网络(包括互联网和局域网)上发布信息成了一件很容易的事。它提供 ISAPI(IntranetServerAPI)作为扩展 Web 服务器功能的编程接口;同时,它还提供一个 Internet 数据库连接器,可以实现对数据库的查询和更新。

2) IBM WebSphere

WebSphere Application Server 是一种功能完善、开放的 Web 应用程序服务器,是 IBM 电子商务计划的核心部分,它是基于 Java 的应用环境,用于建立、部署和管理 Internet 和 Intranet Web 应用程序。这一整套产品进行了扩展,以适应 Web 应用程序服务器的需要,范围从简单到高级直到企业级。

WebSphere 针对以 Web 为中心的开发人员,他们都是在基本 HTTP 服务器和 CGI 编程技术上成长起来的。IBM 将提供 WebSphere 产品系列,通过提供综合资源、可重复使用的组件、功能强大并易于使用的工具,以及支持 HTTP 和 IIOP 通信的可伸缩运行时环境,来帮助这些用户从简单的 Web 应用程序转移到电子商务世界。

3) BEA WebLogic Server

BEA WebLogic Server 是一种多功能、基于标准的 Web 应用服务器,为企业构建自己的应用提供了坚实的基础。各种应用开发、部署所有关键性的任务,无论是集成各种系统和数据库,还是提交服务、跨 Internet 协作,起始点都是 BEA WebLogic Server。由于它具有全面的功能、对开放标准的遵从性、多层架构、支持基于组件的开发,基于 Internet 的企业都选择它来开发、部署最佳的应用。

BEA WebLogic Server 在使应用服务器成为企业应用架构的基础方面继续处于领先地位。BEA WebLogic Server 为构建集成化的企业级应用提供了稳固的基础,它们以 Internet 的容量和速度,在连网的企业之间共享信息、提交服务,实现协作自动化。BEA WebLogic Server 的遵从 J2EE、面向服务的架构,以及丰富的工具集支持,便于实现业务逻辑、数据和表达的分离,提供开发和部署各种业务驱动应用所必需的底层核心功能。

4) iPlanet Application Server

作为 Sun 与 Netscape 联盟产物的 iPlanet 公司生产的 iPlanet Application Server 满足

最新 J2EE 规范的要求。它是一种完整的 Web 服务器应用解决方案,它允许企业以便捷的方式,开发、部署和管理关键任务 Internet 应用。该解决方案集高性能、高度可伸缩和高度可用性于一体,可以支持大量的具有多种客户机类型与数据源的事务。

iPlanet Application Server 的基本核心服务包括事务监控器、多负载平衡选项、对集群和故障转移全面的支持、集成的 XML 解析器和可扩展格式语言转换(XLST)引擎以及对国际化的全面支持。iPlanet Application Server 企业版所提供的全部特性和功能,并得益于 J2EE 系统构架,拥有更好的商业工作流程管理工具和应用集成功能。

5) OracleIAS

OracleIAS 的英文全称是 Oracle Internet Application Server,即 Internet 应用服务器,OracleIAS 是基于 Java 的应用服务器,通过与 Oracle 数据库等产品的结合,OracleIAS 能够满足 Internet 应用对可靠性、可用性和可伸缩性的要求。

OracleIAS 最大的优势是其集成性和通用性,它是一个集成的、通用的中间件产品。在集成性方面,OracleIAS 将业界最流行的 HTTP 服务器 Apache 集成到系统中,集成了 Apache 的 OracleIAS 通信服务层可以处理多种客户请求,包括来自 Web 浏览器、胖客户端和手持设备的请求,并且根据请求的具体内容,将它们分发给不同的应用服务进行处理。在通用性方面,OracleIAS 支持各种业界标准,包括 JavaBeans、CORBA、Servlets 以及 XML 标准等,这种对标准的全面支持使得用户很容易将在其他系统平台上开发的应用移植到 Oracle 平台上。

6) Apache

Apache 源于 NCSAhttpd 服务器,经过多次修改,成为世界上最流行的 Web 服务器软件之一。Apache 是自由软件,所以不断有人来为它开发新的功能、新的特性、修改原来的缺陷。Apache 的特点是简单、速度快、性能稳定,并可做代理服务器来使用。本来它只用于小型或试验 Internet 网络,后来逐步扩充到各种 UNIX 系统中,尤其对 Linux 的支持相当完美。

Apache 是以进程为基础的结构,进程要比线程消耗更多的系统开支,不太适合于多处理器环境,因此,在一个 ApacheWeb 站点扩容时,通常是增加服务器或扩充群集结点而不是增加处理器。到目前为止 Apache 仍然是世界上用得最多的 Web 服务器,世界上很多著名的网站都是 Apache 的产物,它的成功之处主要在于它的源代码开放、有一支开放的开发队伍、支持跨平台的应用(可以运行在几乎所有的 UNIX、Windows、Linux 系统平台上)以及它的可移植性等方面。

7) Tomcat

Tomcat 是一个开放源代码、运行 Servlet 和 JSP Web 应用程序的基于 Java 的 Web 应用软件容器。TomcatServer 是根据 Servlet 和 JSP 规范进行执行的,因此我们就可以说 TomcatServer 也实行了 Apache-Jakarta 规范且比绝大多数商业应用软件服务器要好。

Tomcat 是 JavaServlet 2.2 和 JavaServerPages 1.1 技术的标准实现,是基于 Apache 许可证下开发的自由软件。Tomcat 是完全重写的 ServletAPI 2.2 和 JSP 1.1 兼容的 Servlet/JSP 容器。Tomcat 使用了 JServ 的一些代码,特别是 Apache 服务适配器。随着 CatalinaServlet 引擎的出现,Tomcat 第四版号的性能得到提升,使得它成为一个值得考虑的 Servlet/JSP 容器,因此目前许多 Web 服务器都是采用 Tomcat。

3. Web 浏览器

Web 实际上是一个由文件、图片和声音构成的巨大的信息集合,这些信息存储在遍布全球的各种各样的计算机上,要访问 Web 并把它的全部特征都反映到用户的计算机上,必须使用专用的软件,这就是 Web 浏览器。它是运行于用户计算机上的一种软件。目前有适合不同操作系统以及用户界面的 Web 浏览器。

WWW 浏览器(Browser)是一种 WWW 客户程序,其最基本目的在于让用户在自己的客户机上检索、查询、获取 WWW 上的各种资源。一般来说,浏览器应具备以下几种基本功能:

1) 检索查询功能

浏览器读入 HTML 文档,解释 HTML 所描述的图表、声音、动画、表格,以及进一步的链接信息,可在任意 WWW 服务器上畅游。

2) 文件服务功能

能在下载文档时实时查阅该文档,并可利用 HTTP 去跟踪感兴趣的链接。当感到正在下载的文档不需要时,可以随时中止下载过程。可对正在查阅的文档随时保存、打印、前后浏览等。

3) 热表管理

浏览器应能够自动记住用户刚刚访问过的 WWW 地址,这称为“热表”。当用户想要回到刚才曾访问过的某一网页中,用户可以从热表中快速地切换。

4) 建立自己的主页(Home Page)

当用浏览器启动 Internet 上某一 URL 地址上的某一文档文件时,由浏览器首先显示的那个文档,叫做主页。在主页中,可以加入表示用户特点的图形或图像,列出最常用的一些链接。浏览器提供了很好的接口,可以利用 HTML 和 HTTP 在 WWW 服务器上方便地制作出自己的主页来。

5) 提供其他 Internet 服务

浏览器除了完成自己基本的查询浏览信息功能外,正努力提供越来越多的 Internet 服务,如 FTP、Gopher、WAIS、Telnet、Usenet 及 E-mail 等。

目前市面上已有几十种浏览器,功能有强有弱,它们大多为免费或共享软件,可以在 Internet 上方便地获取。其中流行的和具有代表性的浏览器有 Netscape 公司的 Communicator 和 Microsoft 公司的 Internet Explorer(IE)。

3.5.3 Web 客户端开发技术

Web 编程语言的发展随着人们对 Web 主页的交互性要求越来越高,以完成现在 C/S 架构所能完成的大部分功能而不断发展。最初的 HTML 只能提供静态的页面,没有交互,也没有复杂的动画,这样的页面只能完成简单的信息发布功能。而加入嵌入式语言,如 Java Script、VB Script 等以及使用 Java Applet 技术,使页面有了一些简单的交互功能,如分层菜单等,另外还提供了动态主页和动画的功能,使得信息发布方式更加丰富多彩。但是,由于这些语言都是在客户端浏览器上运行的,规模不能太大,否则下载时间会很长,也不可能支持很多高级的功能,因为这样会占用过多的客户机资源,因此,无法实现诸如数据库操作等复杂操作,自然也就无法在 Intranet 和网上购物中应用了。CGI(Common Gateway

Interface)是一种在服务器端运行的应用程序,用户通过浏览器调用 CGI 程序并传递参数,CGI 程序运行后将结果返回用户的浏览器,这样就允许在 Web 上实现更多更复杂的任务,使得 Web 上的电子商务应用成为可能。但是,CGI 程序编写比较复杂,不同平台上的 CGI 程序也无法重用,因此,各种应用服务器应运而生。应用服务器支持使用功能更强大甚至是人们以前用于 C/S 架构的编程语言,如 VB、PL/SQL 等编写 Web 上的应用程序,由应用服务器负责对这些程序进行解释,并转换成 HTML 在 Web 上发布,且能够从客户的浏览器处获得参数运行 Web 应用程序。应用服务器往往提供更好的与数据库连接的性能,使用户可以更容易地编写更复杂的 Web 应用程序。

客户端技术是 Web 程序中最重要技术之一。客户端技术主要用来描述在浏览器中显示的页面,以及利用 JavaScript 技术对页面进行控制,与服务端进行通信等。常用的客户端技术主要包括 HTML、JavaScript 等。通过学习这些客户端技术,可以很容易地编写具有丰富用户体验的 Web 程序。

1. HTML

HTML(Hyperlink Text Markup Language),即超文本标记语言,是标准通用标记语言(Standard Generalized Markup Language,SGML)的一个简化版本。它是 Web 上的专用表达语言,是开发 Internet 和 Intranet 最有用的工具。利用 HTML 创建的网页只是纯文本代码,任何人都可在计算机平台上的任何文本编辑器中创建、编辑或浏览页面的 HTML 代码。HTML 标准是麻省理工学院(MIT)的 World Wide Web Consortium(W3C)来定义和建立的。大多数浏览器都支持 HTML 标记符。HTML 是一种跨平台文本格式语言、它是以 .html 或 .htm 为扩展名的文件。HTML 语言是在 WWW 上的出版语言,而不能描述实际的表现形式。HTML 语言使用描述性的标记符(称为标签)来指明文档的不同内容。标签是区分文本各个组成部分的分界符,用来把 HTML 文档划分成不同的逻辑部分(结构),如段落、标题和表格等。标签描述了文档结构,它向浏览器提供该文档的格式化信息,以传送文档的外观特征。由服务器传递过来的 HTML 文档、或由 HTML 调用的音频和视频文件经过浏览器解释后,以带有声音、文字、图像的生动页面显示在我们面前。

HTML 文档是由标记符和元素组成的。HTML 标记符确定了浏览器所显示文档元素的格式,它由左尖括号(<)、标记名和右尖括号(>)组成。大多数 HTML 标记是成对出现的,它们分别用作起始标记和结束标记,HTML 的结束标记与起始标记的唯一区别是多了个斜杠(/),例如标记 <HTML> 标明 HTML 文档的开始,标记 </HTML> 指示 HTML 文档的结束。

HTML 文本是由 HTML 命令组成的描述性文本,由具有一定语法结构的标记符和普通文档组成。HTML 命令可以说明文字、图形、动画、声音、表格、链接等。HTML 的结构包括头部(Head)、主体(Body)两大部分,其中头部描述浏览器所需的信息,而主体则包含所要说明的具体内容。

在这里,我们仅给出 HTML 文件的一个简单例子,细节请参考有关书籍。

例 1. 一个 HTML 文档的例子。

```
<html>
<head>
<title>电子商务技术讲座</title>
```

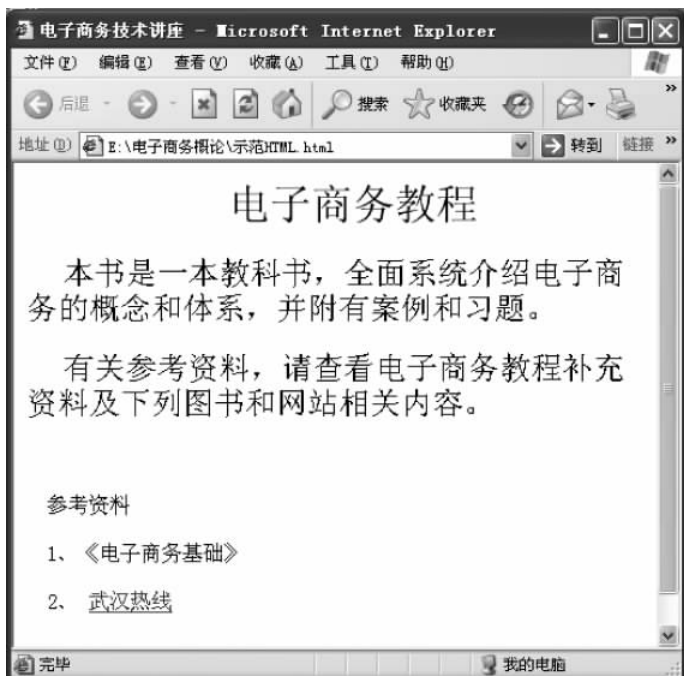



图 3.25 例 1 HTML 文本的显示结果

所需的标记,用户也就没有办法了。这时只好等待标记语言的下一个版本,希望在新版本中能够包括所需的标记,但是这样一来就得依赖于软件开发商的选择了。实际上,用固定数量的各种标记来刻画丰富多彩的现实世界,再多的标记也是不够的。

XML 是一种元标记语言。用户可以定义自己需要的标记。这些标记必须根据某些通用的原理来创建,但是在标记的意义上,也具有相当的灵活性。例如,假如用户正在处理与家谱有关的事情,需要描述人的出生、死亡、埋葬地、家庭、结婚、离婚等,这就必须创建用于每项的标记。新创建的标记可在文档类型定义(Document Type Definition, DTD)中加以描述。DTD 可看作是一本词汇表和某类文档的句法。

XML 定义了一套元句法,与特定领域有关的标记语言(如 MusicML、MathML 和 CML)都必须遵守。如果一个应用程序可以理解这一元句法,那么它也就自动地能够理解所有的由此元语言建立起来的语言。浏览器不必事先了解多种不同的标记语言使用的每个标记。事实是,浏览器在读入文档或是它的 DTD 时才了解了给定文档使用的标记。

例 2. XML 语言文档例。

(1) DTD 格式文件。

```
<!ELEMENT myfile (title, author)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
```

(2) XML 文档。

```
<?xml version = "1.0" encoding = "GB2312" ?>
<!DOCTYPE myfile (View Source for full doctype...)>
```

```
<myfile>
<title>XML 轻松学习手册</title>
<author>ajie</author>
</myfile>
```

(3) 显示 XML 文档的 HTML 文件。

```
<html>
<head>
<script language="javascript" for="window" event="onload">
var xmlDoc = new ActiveXObject("Microsoft.XMLDOM");
xmlDoc.async = "false";
xmlDoc.load("myfile.xml");
nodes = xmlDoc.documentElement.childNodes;
title.innerText = nodes.item(0).text;
author.innerText = nodes.item(1).text;
</script>
<title>在 HTML 中调用 XML 数据</title>
</head>
<body bgcolor="#FFFFFF">
<b>标题: </b>
<span id="title"></span><br>
<b>作者: </b>
<span id="author"></span><br>
</body>
</html>
```

3. JavaScript 和 VBScript

JavaScript 是一种解释性的、基于对象的脚本语言(an interpreted, object-based scripting language)。

HTML 网页在互动性方面能力较弱,例如下拉菜单,就是用户单击某一菜单项时,自动会出现该菜单项的所有子菜单,用纯 HTML 网页无法实现;又如验证 HTML 表单(Form)提交信息的有效性,用户名不能为空,密码不能少于 4 位,邮政编码只能是数字之类,用纯 HTML 网页也无法实现。要实现这些功能,就需要用到 JavaScript。JavaScript 语言程序可以被嵌入 HTML 的文件之中。通过 JavaScript 可以做到回应使用者的需求事件(如 form 的输入)而不用在网络上来回传输任何信息。JavaScript 和 Java 很类似,但许多 Java 的特性在 JavaScript 中并不支持。

JavaScript 主要是基于客户端运行的,用户单击带有 JavaScript 的网页,网页里的 JavaScript 就传到浏览器,由浏览器对此作处理。前面提到的下拉菜单、验证表单有效性等大量互动性功能,都是在客户端完成的,不需要和 Web Server 发生任何数据交换,因此不会增加 Web Server 的负担。

几乎所有浏览器都支持 JavaScript,如 Internet Explorer(IE)、Firefox、Netscape、Mozilla、Opera 等。下面来看一个最简单的例子,代码如下:

```
<html>
<head><title>一个最简单的 JavaScript 示例(仅使用了 document.write)</title></head>
<body>
```

```
<script type="text/javascript">
    document.write("Hello, World!");
</script>
</body>
</html>
```

VBScript 是微软编程语言 Visual Basic 家族中的一个成员,也是一种脚本语言,可以用于微软 IE 浏览器的客户端脚本和微软 IIS (Internet Information Service) 的服务器端脚本。

VBScript 是 VB 的一个子集,它提供的各种语句和语法、常量和变量、函数和过程的规则与 VB 完全相同,并且也提供了许多基本的计算、处理函数。VBScript 是标准的脚本语言,广泛应用于动态网页、大型电子商务系统、Windows 系统管理等领域。下面是一个简单的 VBScript 的例子。

```
<html>
<head><title>一个简单的 VBScript 代码示例</title></head>
<body>
<p>你会看到一个消息框,在此行文字显示之后弹出.</p>
<script type="text/vbscript">
    MsgBox("Hello, World!")
</script>
</body>
</html>
```

3.5.4 Web 服务器端开发技术

1. Java 技术

Java 是由 Sun 微系统公司开发出来的。他们对 Java 的定义是“一种简单、面向对象、分布式、解释执行、功能强大、安全、体系结构中立、可移植、高性能、多线程、动态的语言。”

Java 提供了一些独特的性能,它使 Web 实现多媒体交互。可以使我们不必跟随那些文本的超链接,从动画、音频、视频就可以获得丰富的信息。非专业程序员不必学习 Java 语言,请求需要的就是具有能验证 Java 功能的支持 Java 运行时间系统的 Web 浏览器。

Java 是一种与平台无关的面向对象的语言,平台无关性是指 Java 能运行于不同的平台。Java 引入虚拟机原理,并运行于虚拟机。使用 Java 编写的程序能在世界范围内共享。Java 的数据类型与机器无关,Java 虚拟机(Java Virtual Machine)是建立在硬件和操作系统之上,实现 Java 二进制代码的解释执行功能,提供于不同平台的接口的。Java 程序被放置在 Internet 服务器上,当用户访问服务器时,Java 程序被下载到本地的用户机上,由浏览器解释运行。

Java Applet 是嵌入在 HTML 文档中的 Java 程序;它是一种非常适合在 WWW 上发布的应用程序,由于 Java 语言的安全机制,用户一旦载入 Applet,就可以放心地生成多媒体的用户界面或完成复杂的计算而不必担心病毒的入侵。虽然 Applet 可以和图像、声音、动画等一样从网络上下载,但它并不同于这些多媒体文件格式,它可以接受用户的输入,动态地进行改变,而不仅仅是动画的显示和声音的播放。

Java 是 Sun 公司开发的一种面向对象的新一代网络编程语言,它可以在各种不同的机器、操作系统的网络环境中进行开发,具有解释型语言(如 Basic 语言)和编译型语言(如 C 语言)的特性。Java 摒弃了 C++ 中各种弊大于利的功能和许多很少用到的功能,用 Java 开发的程序可以在网络上传输,并运行于任何客户机上。

电子商务要求程序代码安全、可靠,同时要求能与运行于不同平台的机器的全世界客户开展业务,传统的编程语言难以胜任电子商务的要求。Java 以其强安全性、平台无关性、硬件结构无关性、语言简洁同时面向对象,在网络编程语言中占据无可比拟的优势,成为实现电子商务系统的首选语言。

Java 虽出现的时间不长,但已被业界接受,IBM、Apple、DEC、Adobe、Silicon Graphics、HP、Oracle、Toshiba、Netscap 和 Microsoft 等大公司已经购买了 Java 的许可证。Microsoft 还在其 Web 浏览器 Explorer 3.0 版中增加了对 Java 的支持。

另外,众多的软件开发商也开发了许多支持 Java 的软件产品。如: Borland 公司的基于 Java 的快速应用程序开发环境 Latte; Metrowerks 公司和 Natural Intelligence 公司分别开发的基于 Machintosh 的 Java 开发工具; Sun 公司的 Java 开发环境 Java Workshop; Microsoft 也开发出系列 Java 产品。数据库厂商如 Illustra、Sybase、Versant、Oracle 都在开发支持 HTML 和 Java 的 CGI(Common Gateway Interface)。在以网络为中心的计算时代,不支持 HTML 和 Java,就意味着应用程序的应用范围只能限于同质的环境。

2. JSP

JSP(Java Server Pages)是以 Java 为基础开发的,所以它沿用 Java 强大的 API 功能, JSP 页面中的 HTML 代码用来显示静态内容部分,嵌入到页面中的 Java 代码与 JSP 标记来生成动态内容部分。JSP 可以被预编译,从而提高了程序的运行速度。另外 JSP 开发的应用程序经过一次编译后,可以随时随地地运行,所以在大部分系统平台中,代码无须做修改就可以在支持 JSP 的任何服务器中运行。如图 3.26 下面是一个最简单的一段 JSP 代码。

```
<% @ page contentType = "text/html; charset = GBK" language = "java" errorPage = "" %>
<html>
<head>
<title>欢迎</title>
</head>
<body>
欢迎使用 Java JSP! 现在是: <% out.println(new java.util.Date()); %>
</body>
</html>
```

页面中粗体字代码放在<%和%>之间,表明这些是 Java 脚本,通过这种方式就可以把 Java 代码嵌入 HTML 页面中,就变成了动态的 JSP 页面。

当客户端浏览器向服务器请求一个 JSP 页面时,服务器收到该请求后,首先检查所请求的这个 JSP 文件内容(代码)是否已经被更新,或者是否是 JSP 文件创建后的第一次被访问,如果是,这个 JSP 文件就会在服务器端的 JSP 引擎作用下转化为一个 Servlet 类的 Java 源代码文件,如果不是则不需要重复转换。紧接着,这个 Servlet 类会在 Java 编译器的作用下被编译成一个字节码文件,并装载到 jvm(Java 虚拟机)解释执行。

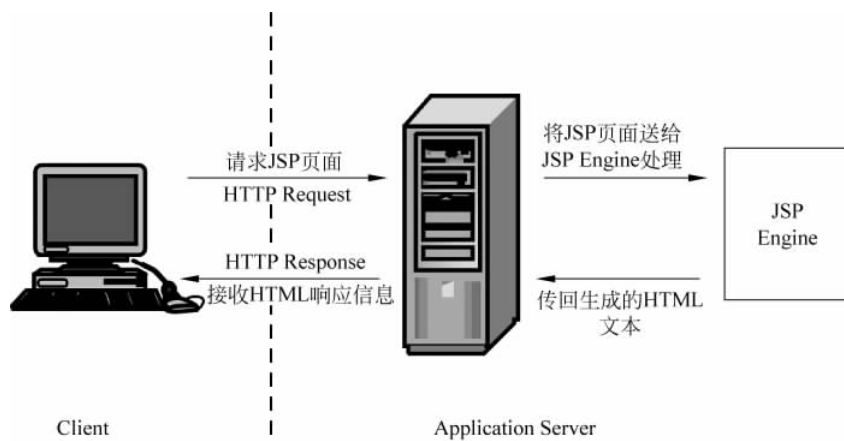


图 3.26 JSP 工作原理图

3. ASP 技术

ASP 即 Microsoft Active Server Pages 的简称。ASP 是微软开发的一套服务器端运行的脚本平台,ASP 内含于 IIS,目前 IIS 已经发展到第 6 版。

通过 ASP 我们可以结合 HTML 网页、ASP 指令和 ActiveX 元件建立动态、交互且高效的 Web 服务器应用程序。同时,ASP 也支持 VBScript 和 JavaScript 等脚本语言,默认为 VbScript。

ASP 是经过服务器解析之后再向浏览器返回数据,所以有了 ASP 就不必担心客户的浏览器是否能运行你所编写的代码。因为所有的程序都将在服务器端执行,包括所有嵌在普通 HTML 中的脚本程序。当程序执行完毕后,服务器仅将执行的结果返回给客户浏览器,这样也就减轻了客户端浏览器的负担,大大提高了交互的速度。

但是这样也导致一个问题,运行 ASP 页面相对于普通的 HTML 页面要慢一点,因为普通的 HTML 页面只需要浏览器就能够解析,而 ASP 则必须是服务器将整页的代码都执行一遍之后再发送数据。

ASP 的安全性还是要提一提,由于代码是需要经过服务器执行之后才向浏览器发送的,所以在客户端看到的只能是经过解析之后的数据,而无法获得源代码,故编写者不用担心自己的代码会被别人剽窃。由于所有代码在服务器运行,所以,可以在同一文件中使用多种不同的脚本语言,下面是一个典型的在同一 .asp 文件中使用两种脚本语言的例子:

```

< HTML >
< BODY >
< TABLE >
< % Call Callme % >
< /TABLE >
< % Call ViewDate % >
< /BODY >
< /HTML >
< SCRIPT LANGUAGE = VBScript RUNAT = Server >
Sub Callme
    Response.Write "< TR>< TD> Call < /TD>< TD> Me < /TD>< /TR>"
  
```

```
End Sub
< /SCRIPT >
< SCRIPT LANGUAGE = JScript RUNAT = Server >
function ViewDate()
{
    var x
    x = new Date()
    Response.Write(x.toString())
}
< /SCRIPT >
```

4. ASP.NET

这种建立动态 Web 应用程序的技术,是 .NET 框架的一部分,可以使用任何 .NET 兼容的语言来编写 ASP.NET 应用程序。使用 Visual Basic .NET、C#、J#、ASP.NET 页面 (Web Forms) 进行编译可以提供比脚本语言更出色的性能。Web Forms 允许在网页基础上建立强大的窗体。当建立页面时,可以使用 ASP.NET 服务端控件来建立常用的 UI 元素,并对它们编程来完成一般的任务。这些控件允许开发者使用内建可重用的组建和自定义组建来快速建立 Web Forms,使代码简单化。

可以将 ASP.NET 页面简单地看成一般的 HTML 页面,页面包含标记有特殊处理方式的一些代码段。从技术上讲,ASP.NET 模块分析 ASPX 文件的内容,并将文件内容分解成单独的命令以建立代码的整体结构。完成此工作后,ASP.NET 模块将各命令放置到预定义的类定义中。然后使用这个类定义一个特殊的 ASP.NET 对象 Page。该对象要完成的任务之一就是生成 HTML 流,这些 HTML 流可以返回到 IIS,再从 IIS 返回到客户。简而言之,在用户请求 IIS 服务器提供一个页面时,IIS 服务器就根据页面上的文本、HTML 和代码建立该页面,如图 3.27 所示。

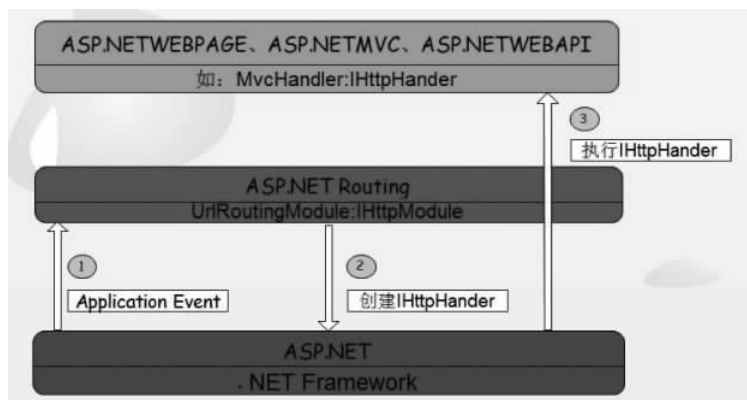


图 3.27 ASP.NET 工作原理图

5. PHP

超文本预处理器 (Hypertext Preprocessor, PHP) 的语法类似于 C, 并且混合了 Perl、C++ 和 Java 的一些特性,它是一种开源的 Web 服务器脚本语言,与 ASP 一样可以在页面中加入脚本代码来生成动态内容。对于一些复杂的操作可以封装到类或函数中。在 PHP 中

提供了许多已经定义好的函数,例如提供的标准数据库接口,使得数据库连接方便,扩展性强。PHP 可以被多个平台支持,但被应用最广泛的还是 UNIX/Linux 平台。由于 PHP 本身的代码对外开放,经过了许多软件工程师的检测,因此,该技术具有公认的安全性能。

PHP 的所有应用程序都是通过 Web 服务器(如 IIS 或 Apache)和 PHP 引擎程序解释执行完成的,工作过程如图 3.28 所示。

(1) 当用户在浏览器地址中输入要访问的 PHP 页面文件名,然后回车就会触发这个 PHP 请求,并将请求传送给支持 PHP 的 Web 服务器。

(2) Web 服务器接受这个请求,并根据其后缀进行判断如果是一个 PHP 请求,Web 服务器从硬盘或内存中取出用户要访问的 PHP 应用程序,并将其发送给 PHP 引擎程序。

(3) PHP 引擎程序将会对 Web 服务器传送过来的文件从头到尾进行扫描并根据命令从后台读取,处理数据,并动态地生成相应的 HTML 页面。

(4) PHP 引擎将生成 HTML 页面返回给 Web 服务器。Web 服务器再将 HTML 页面返回给客户端浏览器。

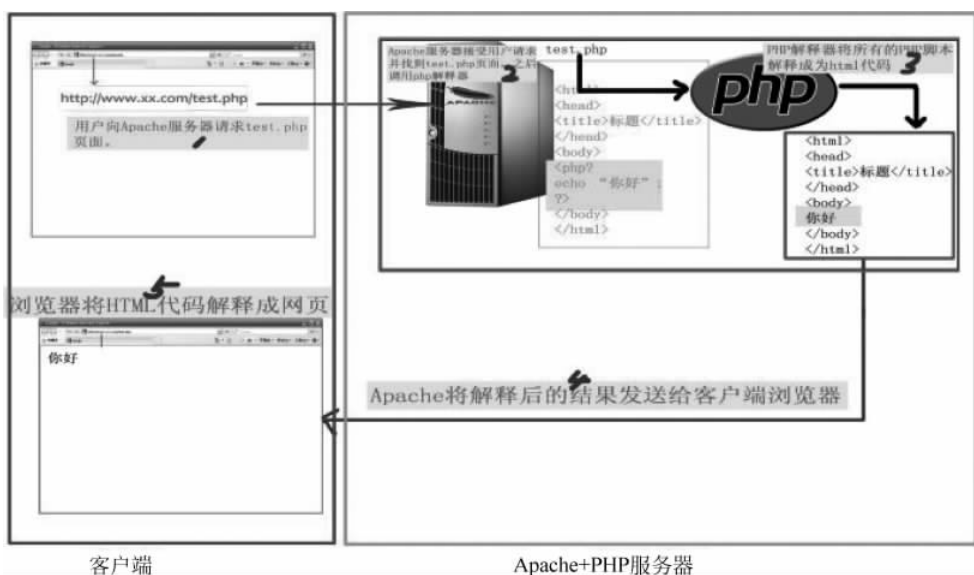


图 3.28 PHP 工作原理图

6. CGI

CGI(Common Gateway Interface)即公共网关接口,它是程序和应用软件用来与网络服务器软件连接的通用标准,CGI 提供网络浏览器和数据信息之间的传送业务。它为 HTTP 服务器定义了一种与外部应用程序共享信息的方法。

CGI 是一段程序,它运行在 Server 上,提供同客户端 HTML 页面的接口。根据客户端在进行请求时所采用的方法,服务器负责收集由客户端提供的信息,并根据客户端请求的 CGI 脚本程序文件名,执行该脚本程序。所谓脚本程序,就是该程序在 CGI 控制下运行的程序。当服务器接收到来自某一客户机的请求,要求启动一个网关程序(即 CGI 脚本)时,它把有关该请求的信息综合到一个环境变量集合中。CGI 脚本程序启动后,服务器将客户

端的信息(环境变量)传给此程序,然后 CGI 脚本程序将检查这些环境变量,试图找到那些为响应请求所必需的信息。该程序对客户的信息处理后,将运行结果交给服务器,由服务器负责再向客户端传递。

CGI 主要运用于处理搜索引擎和一般表单,其主要用途在于使用户能够编写用于与浏览器相交互的程序。

CGI 脚本程序可以用 HTML 文件中的 URL 指明,就像普通的 URL 一样可以在任何地方出现。例如,一个指向 CGI 脚本的连接,它的 HTML 是这样的。

```
< A HREF = 'http://www.cgi.com/cgi-bin/date.cgi' > Date < /A >
```

在许多服务器中的 cgi-bin 目录是仅能够放置 CGI 脚本的目录。当你选择这个连接时,你的浏览器将向 www.cgi.com 服务器提出请求。服务器接收这个请求并计算出 URL 处的脚本文件名,然后执行这个脚本。

本章小结

本章讲解了电子商务的技术基础相关的知识。首先本章介绍了信息技术及其发展过程;介绍了计算机通信和计算机网络相关的基础知识,包括局域网、广域网的概念、计算机组网技术和网络互联技术;重点介绍了 Internet 的基本概念,论述了它的接入技术、功能和应用;讨论了 C/S 和 B/S 计算模式,介绍了中间件的概念和工作原理;最后简要介绍了数据处理技术和 Web 开发技术,介绍了 Web 的工作原理和 Web 开发语言。

思 考 题

1. WWW 的工作原理是什么?
2. 一条信息在 Internet 上是经过唯一的路径到达目的地的吗?
3. 什么是客户/服务器结构? 什么是 B/S 结构? 二者有什么区别?
4. 什么是局域网络? 什么是广域网络? 它们各自有什么特点?
5. 局域网的主要特点是什么?
6. 什么是网络协议? 它的作用是什么? 举例说明。
7. 比较两层和三层客户机/服务器结构并说明各自的用途。一个电子商务网站最有可能采用何种结构?
8. 什么是中间件? 说明中间件的工作原理。
9. 试说明 Web 工作原理。