

第3章

管理信息系统的技术基础

本章包括网络技术基础、数据库技术、多媒体技术、软件工程技术、信息安全技术基础和人工智能等内容。

3.1 网络技术基础

本节从网络的基本概念入手,介绍互联网的基本概念及企业网络的几种模式。

3.1.1 网络的基本概念

计算机网络是计算机及其应用技术与通信技术结合的产物,它经历了由简单到复杂、由低级到高级的发展历程。计算机网络发展历史经历了四个阶段:第一代计算机网络——远程终端联机阶段;第二代计算机网络——计算机网络形成阶段;第三代计算机网络——计算机网络互联阶段;第四代计算机网络——国际互联网与信息高速公路阶段。

1. 基本概念

1) 计算机网络

计算机网络是指将地理位置不同的具有独立功能的多台计算机及其外部设备,通过通信线路连接起来,在网络操作系统、网络管理软件及网络通信协议的管理和协调下,实现资源共享和信息传递的计算机系统。

2) 协议

协议是用来描述进程之间信息交换数据时的规则术语。在计算机网络中,为了使不同结构、不同型号的计算机之间能够正确地传送信息,必须有一套关于信息传输顺序、信息格式和信息内容等的约定,这一整套约定称为协议。在计算机网络中,两个相互通信的实体处在不同的地理位置,其上的两个进程相互通信,需要通过交换信息来协调它们的动作、达到同步,而信息的交换必须按照预先共同约定好的过程进行。网络协议一般是由网络系统决定的。网络系统不同,网络协议也就不同。例如,NetWare 系统的协议是 IPX/SPX, Windows 2000 系统则支持 TCP/IP 等多种协议。

3) 计算机网络的功能

计算机网络的功能主要是硬件资源共享、软件资源共享和用户间信息交换三方面。

(1) 硬件资源共享。可以在全网范围内提供对处理资源、存储资源、输入输出资源等昂贵设备的共享,使用户节省投资,也便于集中管理和均衡分担负荷。

(2) 软件资源共享。允许互联网上的用户远程访问各类大型数据库,可以得到网络文件传送服务、远程进程管理服务和远程文件访问服务,从而避免软件研制上的重复劳动及数据资源的重复存储,也便于集中管理。

(3) 用户间信息交换。计算机网络为分布各地的用户提供了强有力的通信手段。用户可以通过计算机网络传送电子邮件、发布新闻消息和进行电子商务活动。

2. OSI 网络模型

OSI(Open System Interconnection,开放系统互连)参考模型称为开放系统互连参考模型,是一个逻辑上的定义、一个规范,它把网络从逻辑上分为了七层。每一层都有相关、相对应的物理设备,如路由器、交换机。OSI 参考模型是一种框架性的设计方法,建立七层模型的主要目的是为解决异种网络互联时所遇到的兼容性问题,其最主要的功能就是帮助不同类型的主机实现数据传输。它的最大优点是将服务、接口和协议这三个概念明确地区分开来,通过层次化的结构模型使不同的系统、不同的网络之间实现可靠的通信。OSI 参考模型如图 3-1 所示。

1) 物理层

物理层(Physical Layer)位于 OSI 参考模型的最低层或第一层,该层包括物理联网媒介,如电缆连线连接器。物理层的协议产生并检测电压以便发送和接收携带数据的信号。物理层的任务就是为它的上一层提供一个物理连接,以及它们的机械、电气、功能和过程特性,如规定使用电缆和接头的类型、传送信号的电压等。在这一层,数据还没有被组织,仅作为原始的位流或电气电压处理,单位是比特(b)。

2) 数据链路层

数据链路层(Datalink Layer)位于 OSI 参考模型的第二层,它控制网络层与物理层之间的通信。它的主要功能是如何在不可靠的物理线路上进行数据的可靠传递。为了保证传输,从网络层接收到的数据被分割成特定的可被物理层传输的帧(Frame)。帧是用来移动数据的结构包,它不仅包括原始数据,还包括发送方和接收方的物理地址及纠错和控制信息。其中的地址确定了帧将发送到何处,而纠错和控制信息则确保帧无差错地传输。如果在传送数据时,接收点检测到所传数据中有差错,就要通知发送方重发这一帧。数据链路层在物理层提供比特流服务的基础上,建立相邻节点之间的数据链路,通过差错控制提供数据帧在信道上无差错地传输,并进行各电路上的动作系列。数据链路层在不可靠的物理介质上提供可靠的传输。该层的作用包括物理地址寻址、数据的成帧、流量控制、数据的检错、重发等。数据链路层协议的代表包括 SDLC(同步数据链路控制)、HDLC(高级数据链路控制)、PPP(点对点协议)、STP(生成树协议)、帧中继等。

3) 网络层

网络层(Network Layer)位于 OSI 参考模型的第三层,其主要功能是将网络地址翻译成对应的物理地址,并决定如何将数据从发送方路由到接收方。网络层通过综合考虑发送优先权、网络拥塞程度、服务质量及可选路由的花费来决定从一个网络节点 A 到另一个网络节点 B 的最佳路径。网络层的主要功能是处理路由,而路由器其作用是连接网络各段,



图 3-1 OSI 参考模型

并智能化引导数据传送,属于网络层。在网络中,“路由”是基于编址方案、使用模式及可达性来指引数据的传送。网络层负责在源机器和目标机器之间建立它们所使用的路由。这一层本身没有任何错误检测和修正机制,因此,网络层必须依赖于端端之间的由 DLL(动态链接库)提供的可靠传输服务。

4) 传输层

传输层(Transport Layer)是 OSI 参考模型中最重要的一层。传输协议同时进行流量控制或基于接收方可接收数据的快慢程度规定适当的发送速率。除此之外,传输层按照网络能处理的最大尺寸将较长的数据包进行强制分割。例如,以太网无法接收大于 1500 字节的数据包。发送方节点的传输层将数据分割成较小的数据片,同时对每个数据片安排一个序列号,以便数据到达接收方节点的传输层时,能以正确的顺序重组,该过程被称为排序。工作在传输层的一种服务是 TCP/IP 协议簇中的 TCP(传输控制协议),另一项传输层服务是 IPX/SPX 协议集的 SPX(序列包交换)。

5) 会话层

会话层(Session Layer)负责在网络中的两节点之间建立、维持和终止通信。会话层的功能包括:建立通信链接,保持会话过程通信链接的畅通,同步两个节点之间的对话,决定通信是否被中断及通信中断时决定从何处重新发送。当通过拨号向 ISP(因特网服务提供方)请求连接到因特网时,ISP 服务器上的会话层向用户与用户 PC 上的会话层进行协商连接。若电话线偶然从墙上插孔脱落,则终端机上的会话层将检测到连接中断并重新发起连接。会话层通过决定节点通信的优先级和通信时间的长短来设置通信期限。

6) 表示层

表示层(Presentation Layer)是应用程序和网络之间的“翻译官”。在表示层,数据将按照网络能理解的方案进行格式化,这种格式化也因所使用网络的类型不同而不同。表示层管理数据的解密与加密,如系统口令的处理。例如,在 Internet 查询银行账户,使用的是一种安全连接。账户数据在发送前被加密,在网络的另一端,表示层将对接收到的数据解密。除此之外,表示层协议还对图片和文件格式信息进行编码和解码。

7) 应用层

应用层(Application Layer)负责对软件提供接口以使程序能使用网络服务。术语“应用层”并不是指运行在网络上的某个特别应用程序。应用层提供的服务包括文件传输、文件管理及电子邮件的信息处理。

3. 网络拓扑结构

网络拓扑结构指的是网络上的通信链路及各个计算机之间的相互连接的几何排列或物理布局形式。网络拓扑就是指网络形状,即网络中各个节点相互连接的方法和形式。拓扑结构通常有五种主要类型,即星型、环型、总线型、树型和网状,如图 3-2 所示。

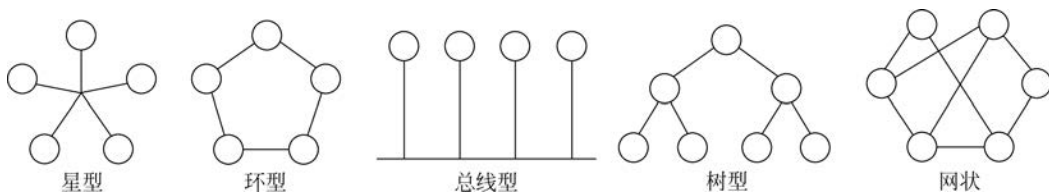


图 3-2 网络拓扑结构

1) 星型拓扑结构

星型拓扑结构的中央节点到其余各节点之间呈辐射状连接,由中央节点完成集中式通信控制。星型拓扑结构的节点有两类,即中心节点和外围节点。中心节点只有一个,每个外围节点都通过独立的通信线路与中心节点相连,外围节点之间没有连线。星型拓扑结构的优点是结构简单,访问协议简单,单个故障不影响整个网络;缺点是可靠性较低,如果中央节点有故障,整个网络就无法工作,全网将瘫痪,且系统扩展较困难。

2) 环型拓扑结构

环型拓扑结构中每个节点连接形成一个闭合回路,数据可以沿环单向传输,也可以设置两个环路实现双向通信。环型拓扑结构扩充方便,传输率较高,但网络中一旦某个节点发生故障,则可能导致整个网络停止工作。

3) 总线型拓扑结构

在总线型拓扑结构中,所有工作节点都连在一条总线上,通过这条总线实现通信。总线结构是目前局域网采用最多的一种拓扑结构。它连接简单,易于扩充节点和删除节点,节点的故障不会引起系统的瘫痪,但是,总线出问题会使整个网络停止工作,故障检测困难。

4) 树型拓扑结构

在树型拓扑结构中,有一个根节点和若干枝节点,最末端是叶节点。其形状像倒立的树“根”。它与总线型比较,总线型没有“根”。根节点的功能较强,常常是高档微机,或中小型机,叶节点可以是微型机。这种结构的优点是扩展容易,易分离故障节点,易维护,特别适合等级严格的行业或部门;缺点是整个网络对根节点的依赖性较大,这对整个网络系统的安全性是一个障碍,若根节点发生故障,整个网络的工作就会受到致命影响。

5) 网状拓扑结构

网状拓扑结构实际是由上述四种拓扑结构中的两种或多种简单组合而成,形状像网一样;网状拓扑结构中计算机之间的通信有多条线路可供选择。它继承了各种结构的优点,但是,其结构复杂,维护难度加大。

4. 几种主要网络结构

虽然网络类型的划分标准各种各样,但是从地理范围划分是一种大家都认可的通用网络划分标准。按这种标准可以把各种网络类型划分为局域网、城域网和广域网。

1) 局域网

所谓局域网(Local Area Network, LAN),就是在局部地区范围内的网络,它所覆盖的地区范围较小。局域网在计算机数量配置上没有太多的限制,少的可以只有两台,多的可达几百台。一般来说,在企业局域网中,工作站的数量在几十到两百台。在网络所涉及的地理距离上一般来说可以是几米至 10km。局域网一般位于一个建筑物或一个单位内,不存在寻径问题,不包括网络层的应用。

2) 城域网

城域网(Metropolitan Area Network, MAN)一般来说是在一个城市,但不在同一地理小区范围内的计算机互联。这种网络的连接距离可以为 10~100km,它采用的是 IEEE 802.6 标准。MAN 与 LAN 相比扩展的距离更长,连接的计算机数量更多,在地理范围上可以说是 LAN 网络的延伸。在一个大型城市或都市地区,一个 MAN 通常连接着多个 LAN。如

连接政府机构的 LAN、医院的 LAN、电信的 LAN、公司企业的 LAN 等。由于光纤连接的引入,使 MAN 中高速的 LAN 互联成为可能。

3) 广域网

广域网(Wide Area Network, WAN)也称为远程网,所覆盖的范围比城域网(MAN)更广,它一般是在不同城市之间的 LAN 或者 MAN 之间互联,地理范围可从几百千米到几千千米。因为距离较远,信息衰减比较严重,所以这种网络一般是租用专线,通过 IMP(接口信息处理)协议和线路连接起来,构成网状结构,解决寻径问题。这种城域网因为所连接的用户多,总出口带宽有限,所以用户的终端连接速率一般较低,通常为 9.6kb/s~45Mb/s。

3.1.2 互联网的基本概念

1. 互联网

互联网,即广域网、局域网及单机按照一定的通信协议组成的国际计算机网络。互联网是指将两台计算机或者是两台以上的计算机终端、客户端、服务端通过计算机信息技术的手段互相联系起来的结果,人们可以与远在千里之外的朋友相互发送邮件、共同完成一项工作、共同娱乐。

1995 年 10 月的 24 日,联合网络委员会(The Federal Networking Council, FNC)通过了一项关于“互联网定义”的决议:联合网络委员会认为,下述语言反映了对“互联网”这个词的定义。

(1) 通过全球唯一的网络逻辑地址在网络媒介基础之上逻辑地链接在一起。这个地址是建立在“互联网协议”(IP)基础之上的。

(2) 可以通过“传输控制协议”和“互联网协议”(TCP/IP),或者今后其他接替的协议或与“互联网协议”(IP)兼容的协议来进行通信。

(3) 以让公共用户或者私人用户享受现代计算机信息技术带来的高水平、全方位的服务。这种服务是建立在上述通信及相关的基础设施之上的。

这是从技术的角度来定义互联网。这个定义至少揭示了三方面的内容:首先,互联网是全球性的;其次,互联网上的每一台主机都需要有“地址”;最后,这些主机必须按照共同的规则(协议)连接在一起。

2. 互联网、因特网、万维网的关系

互联网、因特网、万维网三者的关系是:互联网包含因特网;因特网包含万维网。

凡是能彼此通信的设备组成的网络就叫互联网。所以,即使仅有两台机器,不论用何种技术使其彼此通信,也叫互联网。国际标准的互联网写法是 internet,字母 i 要小写。

因特网是互联网的一种,它是由成千上万台设备组成的互联网。因特网使用 TCP/IP 让不同的设备可以彼此通信。但使用 TCP/IP 的网络并不一定是因特网,一个局域网也可以使用 TCP/IP。判断自己是否接入的是因特网,首先是看自己的计算机是否安装了 TCP/IP,其次看是否拥有一个公网地址(所谓公网地址,就是所有私网地址以外的地址)。国际标准的因特网写法是 Internet,字母 I 要大写。

因特网是基于 TCP/IP 实现的,TCP/IP 由很多协议组成,不同类型的协议又被放在不

同的层,其中,位于应用层的协议就有很多,如 FTP、SMTP、HTTP。只要应用层使用的是 HTTP,就称为万维网(World Wide Web,WWW)。之所以在浏览器里输入百度网址时,能看见百度网提供的网页,就是因为浏览器和百度网的服务器之间使用的是 HTTP 在交流。

3. TCP/IP

TCP/IP(Transfer Control Protocol/Internet Protocol)叫作传输控制协议/网际协议,又叫网络通信协议,它包括上百个各种功能的协议,如远程登录、文件传输和电子邮件等,而 TCP 和 IP 是保证数据完整传输的两个基本的重要协议。通常说 TCP/IP 是 Internet 协议族,而不单单是 TCP 和 IP。TCP/IP 的基本传输单位是数据包(Datagram)。

TCP 负责把数据分成若干数据包,并给每个数据包加上包头;IP 在每个包头上再加上接收端主机地址,这样数据就可以找到自己要去的方。如果传输过程中出现数据丢失、数据失真等情况,TCP 会自动要求数据重新传输,并重新组包。总之,IP 保证数据的传输,TCP 保证数据传输的质量。

TCP/IP 是一组用于实现网络互联的通信协议。Internet 网络体系结构以 TCP/IP 为核心。基于 TCP/IP 的参考模型将协议分成四个层次,分别是网络访问层、网际互联层、传输层(主机到主机)和应用层。

1) 应用层

应用层对应于 OSI 参考模型的高层,为用户提供所需要的各种服务,如 FTP、Telnet、DNS、SMTP 等。

2) 传输层

传输层对应于 OSI 参考模型的传输层,为应用层实体提供端到端的通信功能。该层定义了两个主要的协议:传输控制协议(TCP)和用户数据报协议(UDP)。TCP 提供的是一种可靠的、面向连接的数据传输服务;而 UDP 提供的是不可靠的、无连接的数据传输服务。

3) 网际互联层

网际互联层对应于 OSI 参考模型的网络层,主要解决主机到主机的通信问题。该层有四个主要协议:网际协议(IP)、地址解析协议(ARP)、互联网组管理协议(IGMP)和互联网控制报文协议(ICMP)。IP 是网际互联层最重要的协议,它提供的是一个不可靠、无连接的数据报传递服务。

4) 网络访问层

网络访问层与 OSI 参考模型中的物理层和数据链路层相对应。事实上,TCP/IP 本身并未定义该层的协议,而由参与互联的各网络使用自己的物理层和数据链路层协议,然后与 TCP/IP 的网络访问层进行连接。

3.1.3 企业网络

1. 以太网

以太网(Ethernet)最早是由 Xerox(施乐)公司创建的,在 1980 年由 DEC、Intel 和 Xerox 三家公司联合开发为一个标准。以太网是应用最为广泛的局域网,包括标准以太网

(10Mb/s)、快速以太网(100Mb/s)、千兆以太网(1000Mb/s)和 10Gb/s 以太网。

1) 标准以太网

最开始以太网只有 10Mb/s 的吞吐量,它使用的是 CSMA/CD(带有冲突检测的载波侦听多路访问)的访问控制方法,通常把这种最早期的 10Mb/s 以太网称为标准以太网。以太网主要有两种传输介质,那就是双绞线和同轴电缆。所有的以太网都遵循 IEEE 802.3 标准。

2) 快速以太网

随着网络的发展,传统标准的以太网技术已难以满足日益增长的网络数据流量速度需求。在 1993 年 10 月以前,对于要求 10Mb/s 以上数据流量的 LAN 应用,只有光纤分布式数据接口(FDDI)可供选择,但它是一种价格非常昂贵的、基于 100Mb/s 光缆的 LAN。1993 年 10 月,Grand Junction 公司推出了世界上第一台快速以太网集线器 FastSwitch10/100 和网络接口卡 FastNIC100,快速以太网技术正式得以应用。随后 Intel、SynOptics、3COM、BayNetworks 等公司也相继推出自己的快速以太网装置。与此同时,IEEE 802 工作组也对 100Mb/s 以太网的各种标准,如 100Base-TX、100Base-T4、MII、中继器、全双工等标准进行了研究。1995 年 3 月,IEEE 宣布了 IEEE 802.3u 100Base-T 快速以太网标准(Fast Ethernet),网络进入快速以太网的时代。

3) 千兆以太网

随着以太网技术的深入应用和发展,企业用户对网络连接速度的要求越来越高,1995 年 11 月,IEEE 802.3 工作组委任了一个高速研究组(Higher Speed Study Group),研究将快速以太网速度增至更高。该研究组研究了将快速以太网速度增至 1000Mb/s 的可行性和方法。1996 年 6 月,IEEE 标准委员会批准了千兆以太网方案授权申请(Gigabit Ethernet Project Authorization Request)。随后 IEEE 802.3 工作组成立了 IEEE 802.3z 工作委员会。IEEE 802.3z 委员会的目的是建立千兆以太网标准:包括在 1000Mb/s 通信速率的情况下的全双工和半双工操作、802.3 以太网帧格式、载波侦听多路访问和冲突检测(CSMA/CD)技术、在一个冲突域中支持一个中继器(Repeater)、10Base-T 和 100Base-T 向下兼容技术。千兆以太网具有以太网的易移植、易管理特性。千兆以太网在处理新应用和新数据类型方面具有灵活性,它是在赢得了巨大成功的 10Mb/s 和 100Mb/s IEEE 802.3 以太网标准的基础上的延伸,提供了 1000Mb/s 的数据带宽。这使得千兆以太网成为高速、宽带网络应用的战略性选择。

4) 10Gb/s 以太网

现在 10Gb/s 的以太网标准已经由 IEEE 802.3 工作组于 2000 年正式制定,10Gb/s 以太网仍使用与以往 10Mb/s 和 100Mb/s 以太网相同的形式,它允许直接升级到高速网络。同样使用 IEEE 802.3 标准的帧格式、全双工业务和流量控制方式。在半双工方式下,10Gb/s 以太网使用基本的 CSMA/CD 访问方式来解决共享介质的冲突问题。此外,10Gb/s 以太网使用由 IEEE 802.3 工作组定义的和以太网相同的管理对象。总之,10Gb/s 以太网仍然是以太网,只不过更快。

5) 全光网

随着 Internet 业务和多媒体应用的快速发展,网络的业务量正在以指数级的速度迅速膨胀,这就要求网络必须具有高比特率数据传输能力和大吞吐量的交叉能力。光纤通信技

术出现以后,特别是在提出信息高速公路以来,光技术开始渗透于整个通信网,光纤通信有向全光网推进的趋势。

所谓全光网,是指信号只是在进出网络时才进行电-光和光-电转换,而在网络中传输和交换的过程信号始终以光的形式存在,不需要经过光-电、电-光转换。

为此,网络的交换功能应当直接在光层中完成,这样的网络称为全光网。它需要新型的全光交换器件,如光交叉连接(OXC)、光分插复用(OADM)和光保护倒换等。全光网是以光节点取代现有网络的电节点,并用光线将光节点互联成网,采用光波完成信号的传输、交换等功能,克服了现有网络在传输和交换时的瓶颈,减少信息传输的拥塞、延时,提高网络的吞吐量。全光网已被认为是未来通信网向宽带、大容量发展的优选方案。

2. 网络设备

1) 网桥

网桥这种设备看上去有点像中继器。它具有单个的输入端口和输出端口。它与中继器的不同之处就在于它能够解析它收发的数据。网桥属于 OSI 参考模型的数据链路层;数据链路层能够进行流控制、纠错处理及地址分配。网桥能够解析它所接受的帧,并能指导如何把数据传送到目的地。特别是它能够读取目标地址信息(MAC),并决定是否向网络的其他段转发(重发)数据包,而且,如果数据包的目标地址与源地址位于同一段,就可以把它过滤掉。当节点通过网桥传输数据时,网桥就会根据已知的 MAC 地址和它们在网络中的位置建立过滤数据库(也就是人们熟知的转发表)。网桥利用过滤数据库来决定是转发数据包还是把它过滤掉。

2) 网关

网关不能完全归为一种网络硬件。用概括性的术语来讲,它们应该是能够连接不同网络的软件和硬件的结合产品。特别地,它们可以使用不同的格式、通信协议或结构连接起两个系统。和本章前面讨论的不一样,网关实际上通过重新封装信息以使它们能被另一个系统读取。为了完成这项任务,网关必须能运行在 OSI 参考模型的几个层上。网关必须同应用通信,建立和管理会话,传输已经编码的数据,并解析逻辑和物理地址数据。

3) 交换机

交换(Switching)是按照通信两端传输信息的需要,用人工或设备自动完成的方法,把要传输的信息送到符合要求的相应路由上的技术的统称。交换机是一种用于电信号转发的网络设备。它可以为接入交换机的任意两个网络节点提供独享的电信号通路。最常见的交换机是以太网交换机。其他常见的还有电话语音交换机、光纤交换机等。

4) 路由器

路由器是一种多端口设备,它可以连接不同传输速率并运行于各种环境的局域网和广域网,也可以采用不同的协议。路由器属于 OSI 参考模型的第三层。网络层指导从一个网段到另一个网段的数据传输,也能指导从一种网络向另一种网络的数据传输。过去,由于过多地注意第三层或更高层的数据,如协议或逻辑地址,路由器曾经比交换机和网桥的速度慢。因此,不像网桥和第二层交换机,路由器是依赖于协议的。在它们使用某种协议转发数据前,它们必须要被设计或配置成能识别该协议。

3. 企业级网络

企业级网络是按照整个企业的信息体系要求,由整个企业的计算机硬件、软件、通信系统及数据集合按照系统化要求建立的网络。除了那些极小的组织,大多数企业的网络都是由多个网络组成的,一个容量大的主干网连接着许多局域网和设备,主干网与许多外部网络(如 Internet)进行连接,这往往是企业网的主要形式。

1) 企业内联网

企业内联网(Intranet)是指采用 Internet 技术、建立在 TCP/IP 基础上的企业内部网。它可以在企业内部享有 Internet 的各种功能。它采用统一的 WWW 浏览器技术去开发用户端软件,并能方便地与 Internet 相连。运用防火墙、访问控制等技术,实现企业内、外的可控连接。这使得它具有比传统的企业内部网更加灵活、适应能力更强的特点。

Internet 的普及,使许多组织纷纷建立 Web 服务,将组织的简介、新闻、产品信息、文件档案等放在 Web 主页上供用户浏览,有些还运用 Internet 提供客户服务,接受订单等,并将此视为重要的对外联络窗口。随着组织应用 Internet 技术的成熟,它们逐渐认识到 Internet 带来的好处,开始将 Internet 引入组织的内部作业环境,建立组织内部使用的 Web 服务。因而 Intranet 迅速发展起来,甚至作为组织改革的一种方法。值得注意的是,Intranet 针对的是组织内部信息系统结构,其服务的对象是组织内部的员工,而且是以联络组织内部工作群体为主,以促进组织内部沟通、提高作业效率、强化组织竞争力为前提。

2) 企业外联网

基于 Internet 技术建立企业级网络 Intranet 的技术思想可以进一步向外扩展,即在一组关系密切的企业之间共享 Internet 带来的好处,使这些企业之间能够在可控的安全环境下彼此之间实现信息共享。这便是扩展了的企业级网络——Extranet。纳入 Extranet 的各企业之间,除了各自原有的内部安全共享机制外,还要制定企业之间的安全共享机制,以及它们共同面对 Internet 时的安全机制。

例如,康柏计算机公司和金融服务集团 Vanguard 协作提供投资信息。康柏的员工可以利用这个外部网访问 Vanguard 的服务器,以获取有关账户的信息,以及一些金融方面的特殊消息、教育资料等。少数公司允许特定客户和供应商真正向公司的事务处理系统输入数据。通常,在 Intranet 上提供给这些用户一些表格,让其输入订单或请购单等信息。

Extranet 是使企业与其他企业或客户联系起来,完成共同目标的合作网络,是 Intranet 与 Internet 之间的桥梁。Extranet 可以有选择地向公众或合作者开放其服务,为电子商贸或其他商业应用提供有用的工具。通常 Extranet 只是 Intranet 和 Internet 基础设施上的逻辑覆盖,而不是物理网络的重构。

3.1.4 无线网

1. 无线网的定义

无线网络(Wireless Network)指的是任何形式的无线电计算机网络,普遍和电信网络

结合在一起,无须电缆即可在节点之间相互链接。无线电信网络一般被应用在使用电磁波的遥控信息传输系统,像是无线电波作为载波和物理层的网络。如 TD-LTE,CDMA2000,WCDMA,TD-SCDMA,cdmaOne,GPRS,EDGE,GSM,UMTS,Wi-Fi,WiMAX,ZigBee。

2. 无线网的类型

(1) 无线个人网(WPAN)是在小范围内相互连接数个设备所形成的无线网络,通常是个人可及的范围内,例如蓝牙。

(2) 无线局域网(WLAN)类似其他无线设备,利用无线电而非电缆在同一个网络上传送数据,甚至无线上网,是 IEEE 802.11 系列标准。

(3) 无线城域网是连接数个无线局域网的无线网络类型。

(4) 移动设备网络。最典型的代表是全球移动通信系统(Global System of Mobile Communication,GSM)。GSM 网络分为三个主要系统:转接系统、基地系统、操作和支持系统。移动电话连接到基地系统,然后连接到操作和支持系统;在连接到转接系统后,电话就会被转到要到达的地方。GSM 是大多数手机最常见的使用标准。

3. GSM

GSM 是当前应用最为广泛的移动电话标准,由欧洲电信标准组织 ETSI 制定。它的空中接口采用时分多址技术。GSM 标准的设备占据当前全球蜂窝移动通信设备市场份额 80%以上。GSM 是一种蜂窝网络,也就是说移动电话要连接到它能搜索到的最近的蜂窝单元区域。GSM 网络运行在多个不同的无线电频率上。

4. 5G 通信

5G 是第五代移动通信网络,其峰值理论传输速度可达每秒数 Gb,比 4G 网络的传输速度快数百倍。5G 网络的主要目标是让终端用户始终处于联网状态。5G 网络将来支持的设备远远不止智能手机,它还支持智能手表、健身腕带、智能家庭设备如鸟巢式室内恒温器等。

5G 的具体特征参数如下。

(1) 传输速率:其 5G 网络已成功在 28 吉赫(GHz)波段下达到了 1Gb/s,相比之下,当前的第四代长期演进(4G LTE)服务的传输速率仅为 75Mb/s。

(2) 智能设备:5G 网络中看到的最大改进之处是它能够灵活支持各种不同的设备。除了支持手机和平板电脑外,5G 网络还将支持可佩戴式设备。在一个给定的区域内支持无数台设备,这是设计的目标。

(3) 网络链接:5G 网络改善端到端性能将是另一个重大的课题。端到端性能是指智能手机的无线网络与搜索信息的服务器之间保持连接的状况。

5. 6G 通信

6G 指的是第六代移动通信技术。6G 网络属于概念性技术,是 5G 的延伸,理论下载速度可达每秒 1TB,目前已有机构开始对其进行研发,预计 2026 年正式投入商用。

6G 将迈进太赫兹(THz,1THz=1000GHz)时代。太赫兹波的波长为 $3\sim 1000\mu\text{m}$,它被认为是 6G 的关键技术之一。其特点是频率高、通信速率高,理论上能够达到太字节每秒

(TB/s),但太赫兹有明显的缺点,那就是传输距离短,易受障碍物干扰,现在能做到的通信距离只有 10m 左右,也就是说,只有解决通信距离问题,才能用于现有的移动通信蜂窝网络。此外,通信频率越高对硬件设备的要求越高,需要更好的性能和加工工艺。这些是目前必须在短时间内解决的问题。由于 6G 要求更短的网络延迟时间、更大的带宽、更广的覆盖和更高资源利用率,因此 6G 除了要求高密度组网、全双工技术外,将卫星通信技术、平流层通信技术与地面技术的融合使此前大量未被通信信号覆盖的地方(如无法建基站的海洋、难以铺设光纤的偏远无人地区)都有可能收发信号。除陆地通信覆盖外,水下通信覆盖也有望在 6G 时代启动。6G 将实现地面无线与卫星通信集成的全连接。通过将卫星通信整合到 6G 移动通信,实现永远在线的全球无缝覆盖。

3.1.5 云计算

1. 云计算的概念

狭义云计算指 IT 基础设施的交付和使用模式,指通过网络以按需、易扩展的方式获得所需资源;广义云计算指服务的交付和使用模式,指通过网络以按需、易扩展的方式获得所需服务。这种服务可以是 IT 和软件、互联网相关,也可以是其他服务。云计算的核心思想是将大量用网络连接的计算资源统一管理和调度,构成一个计算资源池为用户按需服务,提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的,并且可以随时获取,按需使用;随时扩展,按使用付费。

云计算是网格计算、分布式计算、并行计算、效用计算、网络存储、虚拟化和负载均衡等传统计算机和网络技术发展融合的产物。事实上,许多云计算部署依赖于计算机集群(但与网络的组成、体系机构、目的、工作方式大相径庭),也吸收了自主计算和效用计算的特点。通过使计算分布在大量的分布式计算机上,而非本地计算机或远程服务器中,企业数据中心的运行将与互联网更相似。这使得企业能够将资源切换到需要的应用上,根据需求访问计算机和存储系统。好比是从古老的单台发电机模式转向了电厂集中供电的模式,它意味着计算能力也可以作为一种商品进行流通,就像煤、气、水、电一样,取用方便,费用低廉。最大的不同在于,它是通过互联网进行传输的。

2. 云计算服务

云计算可以认为包括以下几个层次的服务。

(1) IaaS(Infrastructure-as-a-Service): 基础设施即服务。消费者通过 Internet 可以从完善的计算机基础设施获得服务。

(2) PaaS(Platform-as-a-Service): 平台即服务。PaaS 实际上是指将软件研发的平台作为一种服务,以 SaaS 的模式提交给用户。因此,PaaS 也是 SaaS 模式的一种应用。但是,PaaS 的出现可以加快 SaaS 的发展,尤其是加快 SaaS 应用的开发速度。

(3) SaaS(Software-as-a-Service): 软件即服务。它是一种通过 Internet 提供软件的模式,用户无须购买软件,而是向提供商租用基于 Web 的软件,来管理企业经营活动。相对于传统的软件,SaaS 解决方案有明显的优势,包括较低的前期成本、便于维护、快速展开使用等。

3. 云计算体系架构

云计算的三级分层：①上层云软件，打破以往大厂垄断的局面，所有人都可以在上面自由挥洒创意，提供各式各样的软件服务。参与者是世界各地的软件开发人员。②中层云平台，打造程序开发平台与操作系统平台，让开发人员可以通过网络撰写程序与服务，一般消费者也可以在上面运行程序。参与者有 Google、微软、苹果和 Yahoo 等。③下层云设备，将基础设备（如 IT 系统、数据库等）集成起来，像旅馆一样，分隔成不同的房间供企业租用。参与者有英业达、IBM、戴尔、升阳、惠普和亚马逊等。

大部分的云计算基础构架由通过数据中心传送的可信赖的服务和创建在服务器上的不同层次的虚拟化技术组成。人们可以在任何有提供网络基础设施的地方使用这些服务。“云”通常表现为对所有用户计算需求的单一访问点。人们通常希望商业化的产品能够满足服务质量(QoS)的要求，并且一般情况下要提供服务水平协议。开放标准对于云计算的发展是至关重要的，并且开源软件已经为众多的云计算实例提供了基础。

云是通过网络将庞大的计算处理程序自动分拆成无数个较小的子程序，再由多部服务器所组成的庞大系统搜索、计算分析之后将处理结果回传给用户。通过这项技术，远程的服务供应商可以在数秒之内，达成处理数以千万计甚至亿计的信息，达到和“超级计算机”同样强大性能的网络服务。

3.1.6 区块链

1. 区块链的概念

从科技层面看，区块链(Blockchain)涉及数学、密码学、互联网和计算机编程等很多科学技术问题。从应用视角看，区块链是一个分布式的共享账本和数据库，具有去中心化、不可篡改、全程留痕、可以追溯、集体维护、公开透明等特点。这些特点保证了区块链的“诚实”与“透明”，为区块链创造信任奠定基础。而区块链丰富的应用场景，基本上都基于区块链能够解决信息不对称问题，实现多个主体之间的协作信任与一致行动。

区块链提供了分布式数据存储、点对点传输、共识机制、加密算法等计算机技术新的应用模式。本质上它是一个去中心化的数据库，同时作为比特币的底层技术，是一串使用密码学方法相关联产生的数据块，每个数据块中包含了一批次交易的信息，用于验证其信息的有效性(防伪)和生成下一个区块。但区块链的安全风险是制约其健康发展的短板。因此，区块链的安全保障体系探索迫切需要加快进行。

中本聪(英语：Satoshi Nakamoto)是一个第一次提出区块链概念的日裔美国人。2008年，他发表了一篇名为《比特币：一种点对点式的电子现金系统》(Bitcoin: A Peer-to-Peer Electronic Cash System)的论文，描述了一种被他称为“比特币”的电子货币及其算法。2009年，他发布了首款比特币软件 Bitcoin-Qt，并正式启动了比特币金融系统。

2014年，区块链 2.0 成为一个关于去中心化区块链数据库的术语。对这个第二代可编程区块链，经济学家们认为它是一种编程语言，可以允许用户写出更精密和智能的协议。因此，当利润达到一定程度时，就能够从完成的货运订单或者共享证书的分红中获得收益。区块链 2.0 技术跳过了交易和“价值交换中担任金钱和信息仲裁的中介机构”，使人们远离全

球化经济,使隐私得到保护,使人们“将掌握的信息兑换成货币”,并且有能力保证知识产权的所有者得到收益。第二代区块链技术使存储个人的“永久数字 ID 和形象”成为可能,并且对“潜在的社会财富分配”不平等提供解决方案。

2017 年,日本经济产业省(METI)公布了区块链平台评估方法。2018 年,23 个欧洲国家签署了区块链合作协议,法国为此建立了区块链加速器。2018 年,美国联邦政府和各州政府出台了区块链相关立法,美国国会、商务部、国家标准与技术研究院(NIST)等部门先后发布了《2018 年联合经济报告》《区块链:背景和政策问题》《区块链和在政府应用中的适用性》《区块链技术概述》等报告,初步阐明了美国政府对区块链的监管和发展思路。2019 年 1 月 10 日,中国国家互联网信息办公室发布《区块链信息服务管理规定》,自 2019 年 2 月 15 日起施行。

2. 区块链的类型

1) 公有区块链

公有区块链(Public Blockchain)是指世界上任何个体或者团体都可以发送交易,且交易能够获得该区块链的有效确认,任何人都可以参与其共识过程。公有区块链是最早的区块链,也是应用最广泛的区块链,各大 Bitcoin 系列的虚拟数字货币均基于公有区块链,世界上有且仅有一条该币种对应的区块链。

2) 联合(行业)区块链

行业区块链(Consortium Blockchain)由某个群体内部指定多个预选的节点为记账人,每块的生成由所有的预选节点共同决定(预选节点参与共识过程),其他接入节点可以参与交易,但不过问记账过程(本质上还是托管记账,只是变成分布式记账,预选节点的多少、如何决定每个块的记账者成为该区块链的主要风险点),其他任何人可以通过该区块链开放的 API 进行限定查询。

3) 私有区块链

私有区块链(Private Blockchain)仅仅使用区块链的总账技术进行记账,可以是一个公司,也可以是个人,独享该区块链的写入权限,本链与其他分布式存储方案没有太大区别。传统金融都是想实验尝试私有区块链,而公有区块链的应用例如 Bitcoin 已经工业化,私有区块链的应用产品还在摸索当中。

3. 区块链的特征

(1) 去中心化。区块链技术不依赖额外的第三方管理机构或硬件设施,没有中心管制,除了自成一体的区块链本身,通过分布式核算和存储,各个节点实现了信息自我验证、传递和管理。去中心化是区块链最突出、最本质的特征。

(2) 开放性。区块链技术基础是开源的,除了交易各方的私有信息被加密外,区块链的数据对所有人开放,任何人都可以通过公开的接口查询区块链数据和开发相关应用,因此整个系统信息高度透明。

(3) 独立性。基于协商一致的规范和协议(类似比特币采用的哈希算法等各种数学算法),整个区块链系统不依赖其他第三方,所有节点能够在系统内自动安全地验证、交换数据,不需要任何人为的干预。

(4) 安全性。只要不能掌控全部数据节点的 51%，就无法肆意操控修改网络数据，这使区块链本身变得相对安全，避免了主观人为的数据变更。

(5) 匿名性。除非有法律规范要求，单从技术上来讲，各区块节点的身份信息不需要公开或验证，信息传递可以匿名进行。

4. 区块链的架构模型

区块链的架构模型自下而上由数据层、网络层、共识层、激励层、合约层和应用层组成。其中，数据层封装了底层数据区块以及相关的数据加密和时间戳等基础数据和基本算法；网络层则包括分布式组网机制、数据传播机制和数据验证机制等；共识层主要封装网络节点的各类共识算法；激励层将经济因素集成到区块链技术体系中来，主要包括经济激励的发行机制和分配机制等；合约层主要封装各类脚本、算法和智能合约，是区块链可编程特性的基础；应用层则封装了区块链的各种应用场景和案例。该模型中，基于时间戳的链式区块结构、分布式节点的共识机制、基于共识算力的经济激励和灵活可编程的智能合约是区块链技术最具代表性的创新点。

5. 区块链的核心技术

1) 分布式账本

分布式账本指的是交易记账由分布在不同地方的多个节点共同完成，而且每个节点记录的是完整的账目，因此它们都可以参与监督交易合法性，同时也可以共同为其作证。

与传统分布式存储不同，区块链分布式存储的独特性主要有两点：一是区块链每个节点都按照块链式结构存储完整的数据，传统分布式存储一般是将数据按照一定的规则分成多份进行存储。二是区块链每个节点存储都是独立的、地位等同的，依靠共识机制保证存储的一致性，而传统分布式存储一般是通过中心节点往其他备份节点同步数据。没有任何一个节点可以单独记录账本数据，从而避免了单一记账人被控制或者被贿赂而记假账的可能性。由于记账节点足够多，理论上讲除非所有的节点被破坏，否则账目就不会丢失，从而保证了账目数据的安全性。

2) 非对称加密

存储在区块链上的交易信息是公开的，但是账户身份信息是高度加密的，只有在数据拥有者授权的情况下才能访问到，从而保证了数据的安全和个人的隐私。其使用计算机密码学的非对称加密方法。

3) 共识机制

共识机制就是所有记账节点之间怎么达成共识，去认定一个记录的有效性。这既是认定的手段，也是防止篡改的手段。区块链提出了四种不同的共识机制，适用于不同的应用场景，在效率和安全性之间取得平衡。

区块链的共识机制具备“少数服从多数”以及“人人平等”的特点，其中“少数服从多数”并不完全指节点个数，也可以是计算能力、股权数或者其他的计算机可以比较的特征量。“人人平等”是当节点满足条件时，所有节点都有权优先提出共识结果、直接被其他节点认同后并最后有可能成为最终共识结果。以比特币为例，采用的是工作量证明，只有在控制了全网超过 51% 的记账节点的情况下，才有可能伪造出一条不存在的记录。当加入区块链的节

点足够多时,这基本上不可能,从而杜绝了造假的可能。

4) 智能合约

智能合约是基于这些可信的不可篡改的数据,可以自动化地执行一些预先定义好的规则和条款。以保险为例,如果说每个人的信息(包括医疗信息和风险发生的信息)都是真实可信的,在一些标准化的保险产品中,就能很容易地进行自动化的理赔。在保险公司的日常业务中,虽然交易不像银行和证券行业那样频繁,但是对可信数据的依赖有增无减。因此,利用区块链技术,从数据管理的角度切入,能够有效地帮助保险公司提高风险管理能力。具体来讲主要分投保人风险管理和保险公司的风险监督。

3.2 数据库技术

本节介绍数据库的基本概念、数据描述、数据库组织、关系数据库系统、典型的数据库软件、数据库技术的新发展、大数据技术和可视化。

3.2.1 数据库的基本概念

1. 数据库的定义

定义 1: 数据库(Data Base,DB)是“按照数据结构来组织、存储和管理数据的仓库”。

定义 2: 数据库是依照某种数据模型组织起来并存放在二级存储器中的数据集合。这种数据集合具有如下特点:尽可能不重复,以最优方式为某个特定组织的多种应用服务,其数据结构独立于使用它的应用程序,对数据的增加、删除、修改和查询由统一软件进行管理和控制。从发展的历史看,数据库是数据管理的高级阶段,它是由文件管理系统发展起来的。

定义 3: 数据库是一个长期存储在计算机内的、有组织的、有共享的、统一管理的数据集合。

2. 数据库系统的定义

数据库系统(DataBase System,DBS)是由数据库及其管理软件组成的系统。它是为适应数据处理的需要而发展起来的一种较为理想的数据处理的核心机构。它是一个为实际可运行的存储、维护和应用系统提供数据的软件系统,是存储介质、处理对象和管理系统的集合体。它通常由软件、数据库和数据管理员组成。其软件主要包括操作系统、各种宿主语言、实用程序及数据库管理系统。数据库由数据库管理系统统一管理,数据的插入、修改和查询均要通过数据库管理系统进行。

3. 数据库管理系统的定义

数据库管理系统(DataBase Management System,DBMS)是一种操纵和管理数据库的大型软件,用于建立、使用和维护数据库。它对数据库进行统一的管理和控制,以保证数据库的安全性和完整性。用户通过 DBMS 访问数据库中的数据,数据库管理员也通过 DBMS

进行数据库的维护工作。它提供多种功能,可使多个应用程序和用户用不同的方法在同一时刻或不同时刻去建立、修改和询问数据库。它使用户能方便地定义和操纵数据,维护数据的安全性和完整性,以及进行多用户下的并发控制和恢复数据库。

3.2.2 数据描述

一个信息管理系统中的信息总是从客观事物出发,经过人的综合归纳,抽象成计算机能够接受的信息,流经数据库,通过控制决策机构,最后用来指导客观事物。信息的这一循环经历了三个领域:现实世界、信息世界和计算机世界。在这三个领域中对信息的描述采用不同的术语,三个领域的联系如图 3-3 所示。

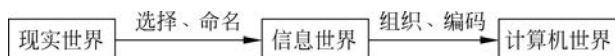


图 3-3 现实世界、信息世界、计算机世界三个领域的联系

1. 现实世界

现实世界是存在于人们头脑之外的客观世界,由客观事物及其相互联系组成。其使用的术语有:

(1) 客观事物。实际存在的人和事物,如销售人员、销售部门、销售客户等;也可以是事物与事物间的联系,如销售合同、退换货处理等。

(2) 事物特征。每一个事物都具有特性,事物通过自身特性与其他事物相区别。例如,销售人员的姓名、性别、学历等,事物特征有名和值之分,具有相同特性的事物属于同一个事物类。

2. 信息世界

信息世界中信息是客观世界中实体的特性在人们头脑中的反映,它用一种人为的文字、符号、标记来表示。其使用的术语有:

(1) 实体。现实世界中客观存在并且可以相互区分的事物称为实体。

(2) 实体集。现实世界中的事物类,在信息世界中称为实体集,它是同类实体的集合。

(3) 属性。现实世界中事物的特征就是实体的属性。属性也有名和值之分,属性名用来划分实体所属的实体集,属性值则是某个实体在该属性下的具体表现。属性值的集合称为属性的域。

(4) 实体标识符。用于和同类实体相互区分的属性集合(不含多余的属性)称为实体标识符。

3. 计算机世界

计算机世界又称为数据世界,由于计算机只能处理数据化的信息,因此必须对信息进行数据化处理。其使用的术语有以下四种。

(1) 记录值。简称记录,表示实体。

(2) 文件。文件是记录的集合。

(3) 字段或数据项。即计算机世界的属性。

(4) 关键字。即计算机世界的实体标识符。

三个领域术语的联系关系如表 3-1 所示。

表 3-1 三个领域术语的联系关系

现实世界	信息世界	计算机世界
客观事物	实体	记录
事物类	实体集	文件
事物特征	属性	字段或数据项
区分事物特征	实体标识符	关键字

3.2.3 数据库组织

数据库组织是指从全局出发,对数据库中的数据、数据之间的联系,以及用户的要求进行综合平衡考虑,从而提出数据模型及数据库系统的结构形式。数据库组织的好坏会影响系统的效率和用户对数据库使用的方便程度。

1. 数据模型

数据模型描述了数据库中的数据内容及其联系方式,体现了数据库的逻辑结构。不同的数据模型就是用不同的数据组织形式来表述实体及其联系。

目前最常用的数据模型有层次模型、网状模型和关系模型。其中层次模型和网状模型统称为非关系模型。

1) 层次模型

层次模型是数据库系统中最早出现的数据模型,它用树状结构表示各类实体以及实体间的联系。层次模型数据库系统的典型代表是 IBM 公司的信息管理系统,这是一个曾经广泛使用的数据库管理系统。在数据库中,满足以下两个条件的数据模型称为层次模型:①有且仅有一个节点无双亲,这个节点称为“根节点”。②其他节点有且仅有一个双亲。若用图来表示,层次模型是一棵倒立的树。节点层次(Level)从根开始定义,根为第一层,根的孩子称为第二层,根称为其孩子的双亲,同一双亲的孩子称为兄弟。

层次模型对具有一对多的层次关系的描述非常自然、直观、容易理解,这是层次数据库的突出优点。

2) 网状模型

在数据库中,对满足以下两个条件的数据模型称为网状模型:

- (1) 允许一个以上的节点无双亲。
- (2) 一个节点可以有多于一个的双亲。

网状数据模型的典型代表是 DBTG 系统,也称为 CODASYL 系统,它是 20 世纪 70 年代数据系统语言研究会(Conference On Data Systems Language,CODASYL)下属的数据库任务组(Data Base Task Group,DBTG)提出的一个系统方案。若用图表示,网状模型是一个网络。

自然界中实体型间的联系更多的是非层次关系,用层次模型表示非树形结构是很不直接的,网状模型则可以克服这一弊病。

3) 关系模型

关系模型是目前最重要的一种模型。美国 IBM 公司的研究员 E. F. Codd 于 1970 年发表题为《大型共享系统的关系数据库的关系模型》的论文,文中首次提出了数据库系统的关系模型。20 世纪 80 年代以来,计算机厂商新推出的数据库管理系统几乎都支持关系模型,非关系系统的产品也大都加上了关系接口。数据库领域当前的研究工作都是以关系方法为基础的。

一个关系模型的逻辑结构是一张二维表,它由行和列组成。每一行称为一个元组,每一列称为一个字段。

关系数据模型的操纵主要包括查询、插入、删除和更新数据。这些操作必须满足关系的完整性约束条件。关系的完整性约束条件包括三大类:实体完整性、参照完整性和用户定义的完整性。其具体含义将在后面介绍。

在关系数据模型中,实体及实体间的联系都用表来表示。在数据库的物理组织中,表以文件形式存储,每张表通常对应一种文件结构。

关系模型与非关系模型不同,它是建立在严格的数学概念的基础上的。关系模型的概念单一,无论实体还是实体之间的联系都用关系来表示,对数据的检索结果也是关系(即表),因此结构简单、清晰,用户易懂易用。关系模型的存取路径对用户透明,从而具有更高的数据独立性,更好的安全保密性,也简化了程序员的工作和数据库开发建立的工作。因此关系数据模型诞生以后发展迅速,深受用户的喜爱。

当然,关系数据模型也有缺点,其中最主要的缺点是,由于存取路径对用户透明,查询效率往往不如非关系数据模型。因此,为了提高性能,必须对用户的查询请求进行优化,这却增加了开发数据库管理系统的负担。

2. 数据库的结构

数据库系统的三级模式结构是指数据库系统是由外模式、模式和内模式三级组成,如图 3-4 所示。

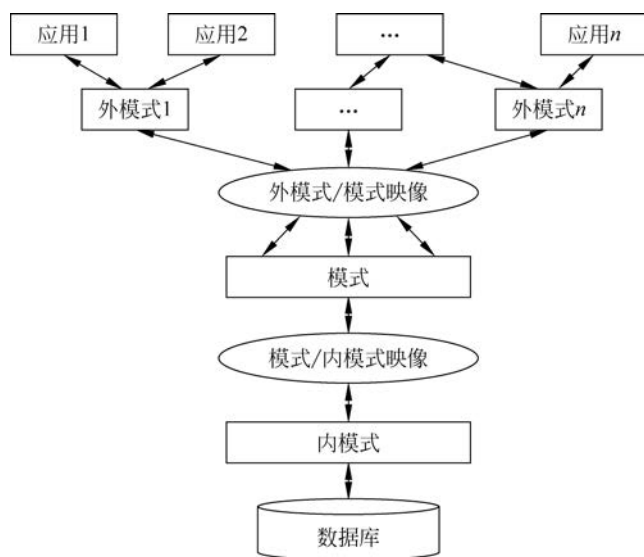


图 3-4 数据库系统的三级模式结构

(1) 外模式。外模式也称为子模式或用户模式,它是数据库用户(包括应用程序员和最终用户)对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。一个数据库可以有多个外模式。

(2) 模式。模式也称为逻辑模式,是数据库中全体数据的逻辑结构和特征的描述,是所有用户的公用数据视图。一个数据库只有一个模式。

(3) 内模式。内模式也称为存储模式,它是数据物理和存储结构的描述,是数据在数据库内部的表示方式。一个数据库只有一个内模式。

数据库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的数据能够具有较高的逻辑独立性和物理独立性。

模式描述的是数据的全局逻辑结构,外模式描述的是数据的局部逻辑结构。对应于同一个模式可以有任意多个外模式。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。当模式改变时(例如,增加新的数据类型、新的数据项、新的关系等),由数据库管理员对各个外模式/模式的映像做相应改变,可以使外模式保持不变,从而不必修改应用程序,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,它定义了数据全局逻辑结构与存储结构之间的对应关系。当数据库的存储结构改变时(例如,采用了更先进的存储结构),由数据库管理员对模式/内模式映像做相应改变,可以使模式保持不变,从而保证了数据的物理独立性。

图 3-4 中的概念模型是现实世界到计算机世界的一个中间层次。现实世界的事物反映到人的脑子中来,人们把这些事物抽象为一种既不依赖于具体的计算机系统又不为某一数据库管理系统支持的概念模型,然后再把概念模型转换为计算机上某一数据库管理系统支持的数据模型。

3. 数据库设计

数据库设计(DataBase Design)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,使之能够有效地存储数据,满足各种用户的应用需求(信息要求和处理要求)。在数据库领域内,常常把使用数据库的各类系统统称为数据库应用系统。

数据库设计包含两方面内容:一是数据模型与数据库结构的设计;二是应用程序的设计。在数据模型与数据库结构的设计上,要汇总各用户的要求,尽量减少冗余,实现数据共享,设计出满足各用户的统一的数据模型,可以分为需求分析、逻辑设计、物理设计、应用程序设计及测试、性能测试及企业确认、装配数据库等几个步骤。

3.2.4 关系数据库系统

1. 关系数据库的特点

关系数据库是以二维平面表作为数据模型的数据库系统。它与层次与网状数据库系统比较,有下列优点。

(1) 数据结构简单。层次与网状模型均使用指针实现实体之间的联系,错综复杂的指针会使程序员眼花缭乱,而关系模型均为表格框架,结构简单。

(2) 可以直接处理多对多的关系。不论实体间的联系是一对多还是多对多,在关系数据库模型中均可用表格形式表示。

(3) 能够一次获取多个记录数据。在层次和网状模型数据库中,每次操作只能得到一个记录值,如果要得到多个记录值,则要借助高级语言的循环、条件等语句才能实现。而在关系数据库中,每一个查找命令可以得到满足该条件的所有记录。

(4) 数据独立性较高。在层次和网状数据库中,对于数据的物理组织要进行一定的干预,而在关系数据库中,用户一般只需指出要存放的数据类型、长度等特性,而不必涉及数据的物理存放。

(5) 有较坚实的理论基础。层次和网状数据库的设计在很大程度上凭设计者的经验和技术水平,关系数据库是以关系数学理论为基础,这样使关系模型的研制和应用设计有理论指导,能保证数据库的质量。

2. 关系数据库的规范化

范式是关系数据库模型设计的基本理论,一个关系模型可以从第一范式到第五范式进行无损分解,这个过程也称为规范化(Normalize)。在数据仓库的模型设计中目前一般采用第三范式,它有非常严格的数学定义。如果从其表达的含义来看,一个符合第三范式的关系必须具有以下三个条件。

(1) 每个属性的值是唯一的,不具有多义性。

(2) 每个非主属性必须完全依赖于整个主键,而非主键的一部分。

(3) 每个非主属性不能依赖于其他关系中的属性。

第三范式的定义基本上是围绕主键与非主属性之间的关系而做出的。如果只满足第一个条件,则称为第一范式;如果满足前面两个条件,则称为第二范式,以此类推。因此,各级范式是向下兼容的。

3. 关系数据库语言 SQL

在关系数据库中,提供给用户对数据进行操作的语言称为关系数据语言,简称数据语言。它以关系运算和关系演算(谓词演算)为基础,结构简单,是一种十分方便的用户接口。SQL(Structured Query Language,结构化查询语言)是1974年IBM公司的圣约瑟研究实验室为关系数据库SystemR研制的,当时称为SEQUEL语言,以后不断进行改进。

20世纪80年代初,先后由Oracle公司与IBM公司推出基于SQL的关系数据库系统。1986年美国国家标准协会(ANSI)批准SQL为数据库语言的美国标准,不久国际标准化组织(ISO)也批准SQL作为关系数据库的公共语言。此后各数据库产品公司纷纷推出各自支持SQL的软件或者与SQL的接口。因此不管微型计算机、小型计算机或大型计算机,不管是哪种数据库,都采用SQL作为共同的数据存取语言,从而使未来的数据库世界可以连接成为一个统一的整体。因此SQL在未来相当长的一段时期内,将成为关系数据库领域中的一个主流语言。

4. 关系系统

一个系统当且仅当它支持如下条件时可以定义为关系系统。

(1) 具有关系数据结构。也就是说,从用户观点看,数据库是由表构成的,并且系统中只有表这种结构。

(2) 支持选择、投影和(自然)连接运算。对这些运算不要求用户定义任何物理存取路径。关系模型中并非每一部分都是同等重要的,所以并不苛求一个实际的关系系统必须完全支持关系模型。

不支持关系数据结构的系统显然不能称为关系系统。仅支持关系数据结构,但没有选择、投影和连接运算功能的系统,用户使用起来仍不方便,这种系统仍不能算作关系系统。支持选择、投影和连接运算,但要求定义物理存取路径,这样就降低或丧失了数据的物理独立性,这种系统也不能算作真正的关系系统。选择、投影、连接运算是最有用的运算,能解决绝大部分实际问题,所以要求关系系统只要支持这三种最主要的运算即可,并不要求它必须提供关系代数的全部运算功能。

3.2.5 典型的数据库软件

1. Oracle 数据库

Oracle 数据库是一种大型数据库系统,一般应用于商业、政府部门。其功能很强大,能够处理大批量的数据。目前它已演化成为大数据管理平台。Oracle 数据库管理系统是一个以关系型和面向对象为中心管理数据的数据库管理软件系统,其在管理信息系统、企业数据处理、因特网及电子商务等领域有着非常广泛的应用。因其在数据安全性与数据完整性方面的优越性能,以及跨操作系统、跨硬件平台的数据互操作能力,已有越来越多的用户将 Oracle 作为其应用数据的处理系统。Oracle 数据库是基于“客户端-服务器”(C/S)模式结构。客户端应用程序执行与用户进行交互的活动。其接收用户信息,并向服务器端发送请求。服务器系统负责管理数据信息和各种操作数据的活动。

2. SQL Server 数据库

SQL Server 是一个关系数据库管理系统。它最初是由 Microsoft、Sybase 和 Ashton-Tate 三家公司共同开发的,于 1988 年推出了第一个 OS/2 版本。Microsoft SQL Server 2005 是一个全面的数据库平台,使用集成的商业智能(BI)工具提供了企业级的数据管理,它的数据库引擎为关系型数据和结构化数据提供了更安全可靠存储功能,使用户可以构建和管理用于业务的高可用和高性能的数据应用程序。Microsoft SQL Server 2005 数据引擎是本企业数据管理解决方案的核心。SQL Server 2008 是一个可信任的、高效的、智能的数据平台,旨在满足目前和将来管理与使用数据的需求。SQL Server 2008 是一个重要的产品版本,它推出了许多新的特性和关键的改进。SQL Server 2008 进一步增加了部分特性和安全性。SQL Server 2014 已经整合了云的功能。SQL Server 2019(15. x)在早期版本的基础上构建,旨在将 SQL Server 发展成一个平台,以提供开发语言、数据类型、本地或云环境以及操作系统选项。

3. DB2 数据库

DB2 数据库是 IBM 公司研制的一种关系型数据库系统,主要用于大型应用系统,具有

较好的可伸缩性,可支持从大型机到单用户环境,应用于 OS/2、Windows 等平台下。DB2 提供了高层次的数据利用性、完整性、安全性、可恢复性,以及小规模到大规模应用程序的执行能力,具有与平台无关的基本功能和 SQL 命令。DB2 采用了数据分级技术,能够使大型机数据很方便地下载到 LAN 数据库服务器,使得 C/S 用户和基于 LAN 的应用程序可以访问大型机数据,并使数据库本地化及远程连接透明化。它以拥有一个非常完备的查询优化器而著称,其外部连接改善了查询性能,并支持多任务并行查询。DB2 具有很好的网络支持能力,每个子系统可以连接十几万个分布式用户,可同时激活上千个活动线程,对大型分布式应用系统尤为适用。除了它可以提供主流的 OS/390 和 VM 操作系统,以及中等规模的 AS/400 系统之外,IBM 公司还提供了跨平台(包括基于 UNIX 的 Linux, HP-UX, Sun Solaris,以及 SCO UnixWare; 还有用于 PC 的 OS/2 操作系统,以及微软的 Windows 2000 和其早期的系统)的 DB2 产品。DB2 数据库可以通过使用微软的开放数据库连接(ODBC)接口、Java 数据库连接(JDBC)接口或者 CORBA 接口代理被任何的应用程序访问。

4. Sybase 数据库

1984 年,Mark B. Hiffman 和 Robert Epstern 创建了 Sybase 公司,并在 1987 年推出了 Sybase 数据库产品。Sybase 有三种版本:一是 UNIX 操作系统下运行的版本;二是 Novell NetWare 环境下运行的版本;三是 Windows NT 环境下运行的版本。UNIX 操作系统环境下广泛应用的是 Sybase 10 及 Sybase 11 for SCO UNIX。Sybase 数据库主要由三部分组成:进行数据库管理和维护的一个联机的关系数据库管理系统 Sybase SQL Server;支持数据库应用系统的建立与开发的一组前端工具 Sybase SQL Toolset;可把异构环境下其他厂商的应用软件 and 任何类型的数据连接在一起的接口 Sybase Open Client/Open Server。

5. Informix 数据库

Informix 在 1980 年成立,目的是为 UNIX 等开放操作系统提供专业的关系数据库产品。公司的名称 Informix 便是取自 Information 和 UNIX 的结合。Informix 第一个真正支持 SQL 的关系数据库产品是 Informix SE(Standard Engine)。Informix SE 是在当时的微机 UNIX 环境下的主要数据库产品。它也是第一个被移植到 Linux 上的商业数据库产品。该产品在国内不太常见。

6. MySQL 数据库

MySQL 是一个小型关系数据库管理系统,开发者为瑞典 MySQL AB 公司。2008 年 1 月 16 日 MySQL 被 Sun 公司收购。目前 MySQL 被广泛地应用在 Internet 上的中小型网站中。由于其体积小、速度快、总体拥有成本低,尤其是开放源码这一特点,许多中小型网站为了降低网站总体拥有成本而选择了 MySQL 作为网站数据库。

7. Access 数据库

Access 是微软公司推出的基于 Windows 的桌面关系数据库管理系统,是 Office 系列应用软件之一。它提供了表、查询、窗体、报表、页、宏、模块七种用来建立数据库系统的对

象;提供了多种向导、生成器、模板,把数据存储、数据查询、界面设计、报表生成等操作规范化;为建立功能完善的数据库管理系统提供了方便,也使得普通用户不必编写代码,就可以完成大部分数据管理的任务。

8. Visual FoxPro 数据库

Visual FoxPro 原名 FoxBase,最初是由美国 Fox Software 公司于 1988 年推出的数据库产品,在 DOS 上运行 Visual FoxPro,与 xBase 系列兼容。FoxPro 是 FoxBase 的加强版,最高版本曾出过 2.6。之后于 1992 年,Fox Software 公司被 Microsoft 公司收购,并加以发展,使其可以在 Windows 上运行,并且更名为 Visual FoxPro。Visual FoxPro 比 FoxBASE 在功能和性能上又有了很大的改进,主要是引入了窗口、按钮、列表框和文本框等控件,进一步提高了系统的开发能力。最新的版本是 9.0 (发布于 2007 年),在微软官方网站发布了一份公告,未来将不会再推出 VFP 10。

3.2.6 数据库技术的新发展

数据库系统在过去的几十年中经历了第一代(层次数据库和网状数据库)和第二代(关系数据库)两个发展阶段,在各行各业得到了广泛应用。它的主要成就是数据模型的建立和有关数据库的理论基础研究。现在无论是集中式或分布式数据库,其存取技术已非常成熟,且具备数据恢复、流量控制、完整性和一致性、查询优化等技术。尤其是建立在牢固的关系数学基础上的关系模型和语言,为用户提供了建模、查询和操作数据库的命令和语言,同时数据库应用开发工具也取得了长足的发展和进步。

近年来,随着管理信息系统应用领域的扩大,数据库在办公自动化、计算机辅助设计与制造(CAD/CAM)、医学辅助诊断(MAD)等方面得到应用,它们的数据除了数值和文本形式外,还采用声音、图形、图像、视频等多种媒体,对数据库系统提出了新的要求。

1. 面向对象的数据库技术

面向对象技术中描述对象及其属性的方法与关系数据库中的关系描述非常一致,它能精确地处理现实世界中复杂的目标对象。面向对象中属性的继承性可以实现在对象中共享数据和操作。在面向对象的数据库系统中把程序和方法也作为对象由面向对象数据库管理系统(OODBMS)统一管理。这样使得数据库中的程序和数据能真正共享,任何被开发的应用程序都作为对象目标库的一部分,被用户及开发者共享,这样就大大缩小了数据库和应用程序之间的距离,降低了应用系统开发费用,提高了系统的可靠性。面向对象数据库系统支持面向对象数据模型。对象数据模型是用面向对象观点来描述现实世界实体(对象)的逻辑组织、对象间限制、联系等的模型。

2. 数据仓库技术

数据仓库是面向主题的、集成的、稳定的、不同时间的数据集合,用以支持经营管理中的决策制定过程。数据仓库是基于知识的数据库管理,知识不仅是传统的统计资料和数据,它也以真实信息和能帮助决策者做出正确决策的专家知识的规则形式存在。

数据仓库在技术上可以根据它的工作过程分为数据的抽取、数据的存储和管理、数据展

现以及数据仓库的维护技术。数据仓库的组织过程是数据的抽取、数据的存储和管理、数据展现。数据的抽取是数据进入仓库的入口,其技术发展将集中在系统集成化方面。数据的存储和管理方式是数据仓库有别于传统数据库的特性,也是知识组织的关键。并行处理技术是数据仓库组织管理的重要技术。数据表现是数据仓库的门面,主要集中在多维分析、数理统计和数据挖掘三方面。

3. 多媒体数据库系统

多媒体数据库是数据库技术与多媒体技术结合的产物。多媒体数据库不是对现有的数据进行界面上的包装,而是从多媒体数据与信息本身的特性出发,考虑将其引入数据库之后而带来的有关问题。多媒体数据库从本质上来说,要解决三个难题:第一是信息媒体的多样化,不仅仅是数值数据和字符数据,要扩大到多媒体数据的存储、组织、使用和管理;第二要解决多媒体数据集成或表现集成,实现多媒体数据之间的交叉调用和融合,集成粒度越细,多媒体一体化表现才越强,应用的价值也才越大;第三是多媒体数据与人之间的交互性。

4. 分布式数据库系统

计算机网络虽然可以实现地域分散情况下的数据传输,以达到数据共享的目的,但它只是全文件的复制,缺乏对数据的管理。因此,人们设想以“数据库系统+计算机网络”来实现分布式数据库系统,既达到对数据的集中管理与共享,又能使地域的分散性被系统隐蔽起来。分布式数据库系统是地理上分布在网络的不同节点而逻辑上属于同一个系统的数据库系统,它通过计算机网络连接在一起。其中每个场地都有一个完全的数据库系统,所有场地都可以协同工作,使得任何场地上的用户都可以访问网络上任何地方的数据,就好像数据是存储在用户自己的场地上一样。从以上叙述中不难看出,分布式数据库系统有分布性、场地自治性、逻辑整体性三个特点。

3.2.7 大数据技术

1. 大数据的定义

根据维基百科的定义,大数据是指无法在可承受的时间范围内用常规软件进行捕捉、管理和处理的数据集合。研究机构 Gartner 关于大数据的定义是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。麦肯锡全球研究所给出的定义是一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合,具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。

大数据技术的战略意义不在于掌握庞大的数据信息,而在于对这些含有意义的数据进行专业化处理。换言之,如果把大数据比作一种产业,那么这种产业实现盈利的关键在于提高对数据的加工能力,通过加工实现数据的增值。

2. 大数据的特点

大数据的 4V 特点是指数据量巨大 (Volume)、数据类型多样 (Variety)、数据流动快

(Velocity)和数据潜在价值大(Value)。

(1) 数据量巨大。大数据是互联网时代发展到一定段时期所必经的过程。伴随着现代社交工具的不断发展,及信息技术领域的不断突破,可以记录的互联网数据正在爆发式地增长。

(2) 数据类型多样。结构和非结构化数据及半结构化数据构成了总的数据。以金融为例,包括股市曲线图、嘉宾专家股市视频、QQ聊天、手机微信、炒股网络日志、关系数据库格式、Word文档、Excel表格、PDF文本、JGP图像、网页等。

(3) 数据流动快。在大数据的构成中,实时数据占到了相当的比例。及时、有效地进行数据处理会涉及交流、传输、感应、决策等。大数据流动快,意味着数据产生速度快,传输速率快,处理速度快。

(4) 数据潜在价值大。当然,大数据中并不全是有价值的数据,需要进行剥离和分析,尤其是涉及科技、教育和经济领域的重要数据。因此,可以理解为数据的价值大小与数据总量的大小成反比。潜在价值的发现将是大数据挖掘的重要研究方向,也会带来高额回报。

3. 大数据解决问题的观念变化

(1) 大数据研究的不是随机样本,而是全部数据。在大数据时代,可以获得和分析某个问题或对象背后的全部数据;有时甚至可以处理与某个特殊现象相关的所有数据,而不再依赖于随机采样。

(2) 大数据研究的不是精确性,而是大体方向。之前需要分析的数据很少,所以必须尽可能精确地量化数据;随着研究数据规模的扩大,研究者对数据精确度的痴迷将减弱。拥有了大数据,不必过于对某一个现象刨根问底,只需掌握大体的发展方向即可,适当忽略微观层面上的精确度,在宏观层面易于获得更好的洞察力。

(3) 大数据研究的不是因果关系,而是相关关系。寻找因果关系是人类长久以来的习惯,而在大数据时代,人们无须再紧盯事物之间的因果关系,而应该寻找事物之间的相关关系;相关关系也许不能准确揭示某件事情为何会发生,但是它会提醒人们这件事情正在怎么发生。

4. 大数据技术的核心

大数据技术有四个核心部分,它们是大数据采集与预处理、大数据存储与管理、大数据计算模式与系统及大数据分析可视化。

1) 大数据采集与预处理

在大数据的生命周期中,数据采集处于第一个环节。大数据的采集主要有四种来源:管理信息系统、Web信息系统、物理信息系统和科学实验系统。对于不同的数据集,可能存在不同的结构和模式,如文件、XML树、关系表等表现为数据的异构性。对多个异构的数据集需要做进一步集成处理或整合处理,将来自不同数据集的数据收集、整理、清洗、转换后生成一个新的数据集,为后续查询和分析处理提供统一的数据视图。人们针对管理信息系统中的异构数据库集成技术、Web信息系统中的实体识别技术和Deep Web集成技术、传感器网络中的数据融合技术进行了很多研究工作,取得了较大的进展,已经推出了多种数据清洗和质量控制工具。

2) 大数据存储与管理

大数据多半是以半结构化和非结构化数据为主,而大数据应用通常是对不同类型的数据进行内容检索、交叉比对、深度挖掘和综合分析。面对这种应用需求,传统数据库无论在技术上还是在功能上,都难以为继。因此,近几年出现了 OldSQL、NoSQL 与 NewSQL 并存的局面。按数据类型的不同,大数据的存储和管理可采用不同的技术路线,大致可以分为三类。

第一类主要面对的是大规模的结构化数据。针对这类大数据,通常采用新型数据库集群。它们通过列存储或行列混合存储及粗粒度索引等技术,结合 MPP(Massive Parallel Processing,大规模并行处理)架构高效的分布式计算模式,实现对 PB 量级数据的存储和管理。

第二类主要面对的是半结构化和非结构化数据。对此,基于 Hadoop 开源体系的系统平台更为擅长。它们通过对 Hadoop 生态体系的技术扩展和封装,实现对半结构化和非结构化数据的存储和管理。

第三类主要面对的是结构化和非结构化混合的大数据,对此,采用 MPP 并行数据库集群与 Hadoop 集群的混合来实现对 EB 量级数据的存储和管理。

3) 大数据计算模式与系统

大数据计算模式就是根据大数据的不同数据特征和计算特征,从多样性的大数据计算问题和需求中提炼并建立的各种高层抽象或模型。例如,MapReduce 是一个并行计算抽象、加州大学伯克利分校的 Spark 系统中的“分布内存抽象 RDD”、CMU 的图计算系统 Graph Lab 中的“图并行抽象”(Graph Parallel Abstraction)等。大数据处理多样性的需求驱动了多种大数据计算模式出现。与这些计算模式相对应,出现了很多对应的大数据计算系统和工具。例如,大数据查询分析计算模式,其工具为 HBase、Hive、Cassandra、Impala、Shark;批处理计算模式,其工具为 MapReduce、Spark;流式计算模式,其工具为 Flume、Storm、SparkStreaming;迭代计算模式,其工具为 Hadoop、MapReduce、Twister、Spark;图计算模式,其工具为 GraphX;内存计算模式,其工具为 Spark。

4) 大数据分析可视化

大数据分析是指对规模巨大的数据进行分析。一时间,数据仓库、数据安全、数据分析、数据挖掘等技术逐渐成为行业追捧的焦点。大数据分析包括六方面:可视化分析、数据挖掘算法、预测性分析、语义引擎(从文档中智能提取信息)、数据质量与数据管理及数据仓库与商业智能。

3.2.8 可视化

近年来,随着计算机应用的广泛和科学技术的迅速发展,来自超级计算机、卫星遥感、CT、天气预报及地震勘测等领域的数据量越来越大,科学计算可视化技术已经成为科学研究中必不可少的手段。它是科学工作者及工程技术人员洞察数据内含信息、确定内在关系与规律的有效方法,使科学家和工程师以直观形象的方式揭示、理解抽象科学数据中包含的客观规律,从而摆脱直接面对大量无法理解的抽象数据的被动局面。

1. 科学计算可视化的定义

可视化就是将科学计算的中间数据或结果数据,转换为人们容易理解的图形图像形式。随着计算机、图形图像技术的飞速发展,人们现在已经可以用丰富的色彩、动画技术、三维立体显示及仿真(虚拟现实)等手段,形象地显示各种地形特征和植被特征模型,也可以模拟某些还未发生的物理过程(如天气预报)、自然现象及产品外形(如新型飞机)。

目前,科学计算可视化已广泛应用于流体计算力学、有限元分析、医学图像处理、分子结构模型、天体物理、空间探测、地球科学、数学等领域。从可视化的数据上划分,有点数据、标量场、矢量场等;从可视化的维度上划分,有二维、三维,以至多维;从可视化实现层次划分,有简单的结果后处理、实时跟踪显示、实时交互处理等。通常,一个可视化过程包括数据预处理、构造模型、绘图及显示等几个步骤。随着科学技术的发展,人们对可视化的要求不断提高,可视化技术也向着实时、交互、多维、虚拟现实及因特网应用等方面不断发展。

2. 科学计算可视化的应用

可视化技术自 1987 年 2 月诞生之日起,便受到了各行各业的欢迎。在过去的十年里,可视化的应用范围已从最初的科研领域走到了生产领域;到今天,它几乎涉及了所有能应用计算机的领域。

(1) 医学。在医学上由核磁共振、CT 扫描等设备产生的人体器官密度场,对于不同的器官组织,表现出不同的密度值。通过在多个方向、多个剖面来表现病变区域,或者重建为具有不同细节程度的三维真实图像,医生对病灶部位的大小、位置不仅有定性的认识,而且有定量的认识,尤其是对大脑等复杂区域,数据场可视化所带来的效果相当明显。借助虚拟现实的手段,医生可以对病变的部位进行确诊,制定出有效的手术方案,并在手术之前模拟手术。在临床上也可应用在放射诊断、制订放射治疗计划等领域。医学成像设备和检查结果如图 3-5 所示。

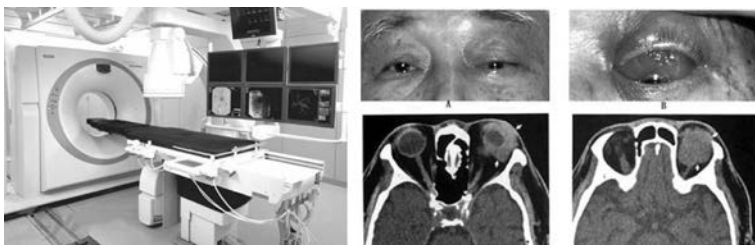


图 3-5 医学成像设备和检查结果

(2) 生物、分子学。在对蛋白质和 DNA 分子等复杂结构进行研究时,可以利用电镜、光镜等辅助设备对其剖片进行分析、采样,获得剖片信息;利用这些剖片构成的体数据可以对其原形态进行定性和定量分析,因此可视化是研究分子结构不可或缺的工具。分子模拟图如图 3-6 所示。

(3) 航天工业。在可视化技术下,飞行器运动情况和飞行器的表现,可以非常直观地展现出来。借助可视化技术,许多意想不到的监测困难等都可以迎刃而解了。卫星运动飞行图和宇宙星座模拟图如图 3-7 所示。

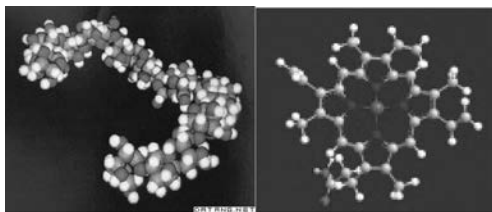


图 3-6 分子模拟图

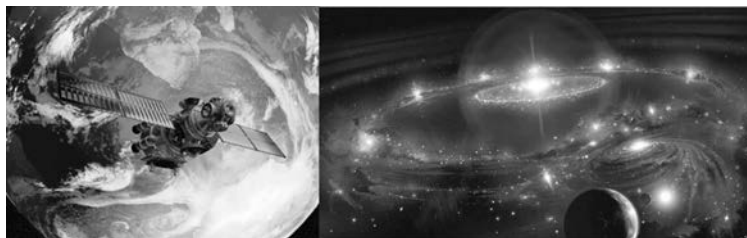


图 3-7 卫星运动飞行图和宇宙星座模拟图

(4) 地质勘探。利用模拟人工地震的方法,可以获得地质岩层信息。通过数据特征的抽取和匹配,可以确定地下的矿藏资源。用可视化方法对模拟地震数据的解释,可以大大地提高地质勘探的效率和安全性。数字城市地图和模拟地形图如图 3-8 所示。

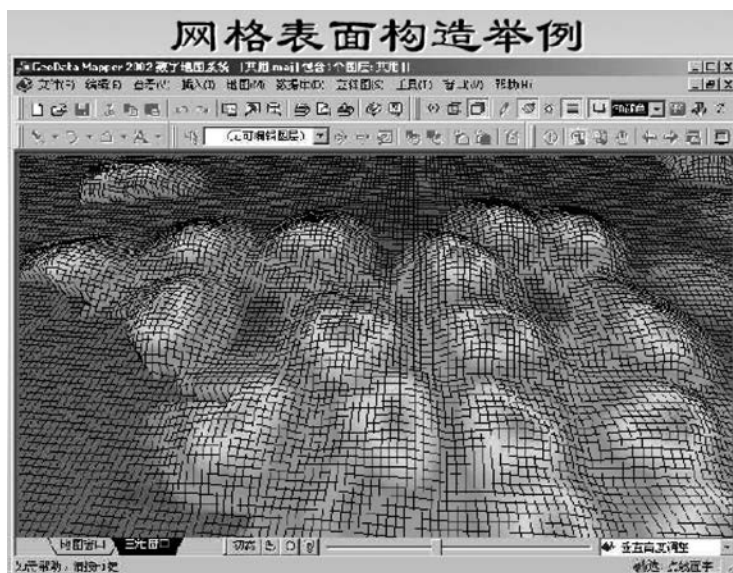


图 3-8 数字城市地图和模拟地形图

(5) 立体云图显示。气象分析和预报需要处理大量的测量或计算数据,气象云图是其中一种非常重要的气象数据,也常用于发布天气预报。气象研究中,地形和云层的高度是影响天气的重要因素。运用可视化技术,将三维立体地形图和三维立体云图合成显示、输出,能给人更形象、直观的认识。合成云图如图 3-9 所示。

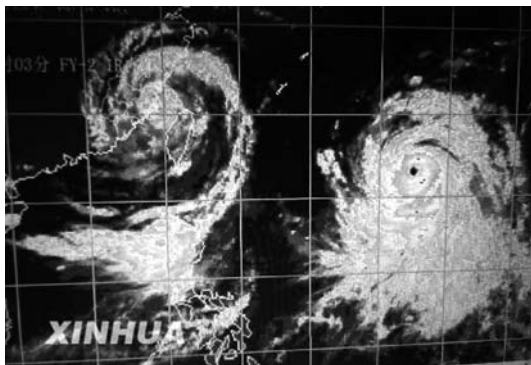


图 3-9 合成云图

3.3 多媒体技术

本节介绍多媒体技术的基本概念和具体的多媒体技术,包括音频技术、视频技术、数字图像技术,以及多媒体系统。

3.3.1 多媒体技术的基本概念

1. 多媒体的概念

(1) 媒体(Media)就是人与人之间实现信息交流的中介,简单地说,就是信息的载体,也称为媒介。在计算机行业里,媒体有两种含义:其一指传播信息的载体,如语言、文字、图像、视频、音频等;其二是指存储信息的载体,如 ROM、RAM、磁带、磁盘、光盘等,目前,主要的载体有 CD-ROM、VCD、网页等。

(2) 多媒体一般理解为多种媒体的综合,是计算机和视频技术的结合。可以理解为直接作用于人感官的文字、图形、图像、动画、声音和视频等各种媒体的统称,即多种信息载体的表现形式和传递方式。

2. 多媒体的基本形式

(1) 文本。文本是以文字和各种专用符号表达的信息形式,它是现实生活中使用得最多的一种信息存储和传递方式。用文本表达信息给人充分的想象空间,它主要用于对知识的描述性表示,如阐述概念、定义、原理和问题以及显示标题、菜单等内容。

(2) 图像。图像是多媒体软件中最重要的信息表现形式之一,它是决定一个多媒体软件视觉效果的关键因素。

(3) 动画。动画是利用人的视觉暂留特性,快速播放一系列连续运动变化的图形图像,也包括画面的缩放、旋转、变换、淡入淡出等特殊效果。通过动画可以把抽象的内容形象化,使许多难以理解的教学内容变得生动有趣。合理使用动画可以达到事半功倍的效果。

(4) 声音。声音是人们用来传递信息、交流感情最方便、最熟悉的方式之一。在多媒体课件中,按其表达形式,可将声音分为讲解、音乐、效果三类。

(5) 视频影像。视频影像具有时序性与丰富的信息内涵,常用于交代事物的发展过程。视频非常类似于我们熟知的电影和电视,有声有色,在多媒体中充当起重要的角色。

3. 多媒体技术的定义

多媒体技术是计算机技术和视频技术的结合。多媒体由硬件和软件组成。多媒体是数字控制和数字媒体的汇合,计算机负责数字控制系统,数字媒体是音频和视频先进技术的结合。

多媒体技术是多种信息类型技术的综合。这些媒体可以是图形、图像、声音、文字、视频、动画等信息表示形式,也可以是显示器、扬声器、电视机等信息的展示设备,传递信息的光纤、电缆、电磁波、计算机等中介介质,还可以是存储信息的磁盘、光盘、磁带等存储实体。多媒体技术应该包括音频技术、视频技术、图像技术、通信技术、存储技术等。

4. 多媒体技术的特点

(1) 集成性。能够对信息进行多通道统一获取、存储、组织与合成。

(2) 控制性。多媒体技术是以计算机为中心,综合处理和控制在多媒体信息,并按人的要求以多种媒体形式表现出来,同时作用于人的多种感官。

(3) 交互性。交互性是多媒体应用有别于传统信息交流媒体的主要特点之一。传统信息交流媒体只能单向地、被动地传播信息,而多媒体技术则可以实现人对信息的主动选择和控制。

(4) 非线性。多媒体技术的非线性特点将改变人们传统循序性的读写模式。以往人们读写方式大都采用章、节、页的框架,循序渐进地获取知识,而多媒体技术将借助超文本链接的方法,把内容以一种更灵活、更具变化的方式呈现给读者。

(5) 实时性。当用户给出操作命令时,相应的多媒体信息都能够得到实时控制。

(6) 互动性。它可以形成人与机器、人与人及机器间的互动、互相交流的操作环境及身临其境的场景,人们根据需要进行控制。人机相互交流是多媒体最大的特点。

(7) 方便性。用户可以按照自己的需要、兴趣、任务要求、偏爱和认知特点来使用信息,任取图、文、声等信息表现形式。

(8) 动态性。“多媒体是一部永远读不完的书”,用户可以按照自己的目的和认知特征重新组织信息,增加、删除或修改节点,重新建立链接。

3.3.2 多媒体技术介绍

1. 音频技术

1) 音频处理技术

多媒体涉及多方面的音频处理技术,如音频采集、语音编码/解码、文语转换、音乐合成、语音识别与理解、音频数据传输、音频视频同步、音频效果与编辑等。其中数字音频是个关键的概念,它指的是一个用来表示声音强弱的数据序列,它是由模拟声音经抽样(即每隔一

个时间间隔在模拟声音波形上取一个幅度值)、量化和编码(把声音数据写成计算机的数据格式)后得到的。计算机数字 CD、数字磁带(DAT)中存储的都是数字声音。模数转换器把模拟声音变成数字声音;数模转换器可以恢复出模拟声音。

2) 乐器数字接口

MIDI(Musical Instrument Digital Interface)是乐器数字接口,用它来泛指数字音乐的国际标准。由于它定义了计算机音乐程序、合成器及其他电子设备交换信息和电子信号的方式,因此可以解决不同电子乐器之间不兼容的问题。另外,标准的多媒体 PC 平台能够通过内部合成器或连接到计算机 MIDI 端口的外部合成器播放 MIDI 文件,利用 MIDI 文件演奏音乐,所需的存储量最少。MIDI 文件是指存放 MIDI 信息的标准文件格式。MIDI 文件中包含音符、定时和多达 16 个通道的演奏定义。文件包括每个通道的演奏音符信息:键通道号、音长、音量和力度(击键时,键达到最低位置的速度)。由于 MIDI 文件是一系列指令,而不是波形,它需要的磁盘空间非常少;并且现装载 MIDI 文件比波形文件容易得多。

3) 常见声音文件格式

(1) WAV。该格式记录声音的波形,故只要采样率高、采样字节长、机器速度快,利用该格式记录的声音文件能够和原声基本一致,质量非常高,但这样做的代价就是文件太大。

(2) MP3。现在最流行的声音文件格式,因其压缩率大,在网络可视电话通信方面应用广泛,但和 CD 唱片相比,音质不能令人非常满意。

(3) RA。这种格式真可谓是网络的灵魂,强大的压缩量和极小的失真使其在众多格式中脱颖而出。和 MP3 相同,它也是为了解决网络传输带宽资源而设计的,因此主要目标是压缩比和容错性,其次才是音质。

(4) CMF。Creative 公司的专用音乐格式,和 MIDI 差不多,只是音色、效果上有些特色,专用于 FM 声卡,但其兼容性也很差。

(5) MIDI。目前最成熟的音乐格式,实际上已经成为一种产业标准,其科学性、兼容性、复杂程度等各方面远远超过本书前面介绍的所有标准(除交响乐 CD、Unplug CD 外,其他 CD 往往都是利用 MIDI 制作出来的),它的 General MIDI 就是最常见的通行标准。作为音乐工业的数据通信标准,MIDI 能指挥各音乐设备的运转,而且具有统一的标准格式,能够模仿原始乐器的各种演奏技巧,甚至无法演奏的效果,而且文件的长度非常小。

2. 视频技术

1) 动态图像

动态图像包括动画和视频信息,是连续渐变的静态图像或图形序列,沿时间轴顺次更换显示,从而构成运动视感的媒体。当序列中每帧图像都是由人工或计算机产生的图像时,常称为动画;当序列中每帧图像都是通过实时摄取自然景象或活动对象时,常称为影像视频,或简称为视频。动态图像演示常常与声音媒体配合进行,二者的共同基础是时间连续性。一般意义上谈到视频时,往往也包含声音媒体。但在这里,视频(动画)特指不包含声音媒体的动态图像。

2) 常见视频文件格式

(1) AVI(Audio Video Interleaved,音频视频交错)。该文件格式是 Microsoft 公司开

发的一种符合 RIFF 文件规范的数字音频与视频文件格式,AVI 格式允许视频和音频交错在一起同步播放,支持 256 色和 RLE 压缩,但 AVI 文件并未限定压缩标准。

(2) MPEG(Moving Pictures Experts Group,动态图像专家组)。该文件格式是运动图像压缩算法的国际标准,它采用有损压缩方法减少运动图像中的冗余信息,同时保证每秒 30 帧的图像动态刷新率,已被几乎所有的计算机平台共同支持。MPEG 标准包括 MPEG 视频、MPEG 音频和 MPEG 系统(视频、音频同步)三部分。

(3) RA/.RM/.RMVB。该文件格式是 RealVideo 文件,是 RealNetworks 公司开发的一种新型流式视频文件格式,它包含在 RealNetworks 公司所制定的音频视频压缩规范 RealMedia 中,主要用来在低速率的广域网上实时传输活动视频影像,可以根据网络数据传输速率的不同而采用不同的压缩比率,从而实现影像数据的实时传送和实时播放。

(4) MOV/QT(QuickTime)。该文件格式是 Apple 计算机公司开发的一种音频、视频文件格式,用于保存音频和视频信息,具有先进的视频和音频功能。QuickTime 文件格式支持 25 位彩色,支持 RLE、JPEG 等领先的集成压缩技术,提供 150 多种视频效果,并配有提供了 200 多种 MIDI 兼容音响和设备的声音装置。

(5) ASF/WMV。Microsoft 公司推出的 ASF(Advanced Streaming Format,高级流格式)的主要优点包括本地或网络回放、可扩充的媒体类型、部件下载以及扩展性等。ASF 应用的主要部件是 NetShow 服务器和 NetShow 播放器。WMV 是一种独立于编码方式,在 Internet 上实时传播多媒体的技术标准,WMV 的主要优点包括本地或网络回放、可扩充的媒体类型、部件下载、可伸缩的媒体类型、流的优先级化、多语言支持、环境独立性、丰富的流间关系以及扩展性等。

3. 数字图像技术

1) 数字图像分类

数字图像分两大类:一类为位图;另一类为矢量图。前者是以点阵形式描述的图,后者是以数学方法描述的一种由几何元素组成的图形。后者对图像的表达细致、真实,缩放后图形的分辨率不变,在专业级的图形处理中运用较多。

2) 常见图像文件格式

(1) BMP(Bit Map Picture)。计算机上最常用的位图格式,有压缩和不压缩两种形式。该格式可表现从 2 位到 24 位的色彩,分辨率也可从 480×320 像素至 1024×768 像素。该格式在 Windows 环境下相当稳定,在文件大小没有限制的场合中运用极为广泛。

(2) WMF(Windows Metafile Format)。Microsoft Windows 图元文件,具有文件短小、图案造型化的特点。该类图形比较粗糙,并只能在 Microsoft Office 中调用编辑。

(3) GIF(Graphics Interchange Format)。在各种平台的各种图形处理软件上均可处理的经过压缩的图形格式。缺点是存储色彩最高只能达到 256 种。

(4) JPG(Joint Photographic Experts Group)。可以大幅度地压缩图形文件的一种图形格式。对于同一幅画面,JPG 格式存储的文件是其他类型图形文件的 $1/20 \sim 1/10$,而且色彩数最高可达到 24 位,所以它被广泛应用于 Internet 上的 Homepage 或 Internet 上的图片库。

(5) TIF(Tagged Image File Format)。文件体积庞大,但存储信息量也巨大,细微层次的信息较多,有利于原稿阶调与色彩的复制。该格式有压缩和非压缩两种形式。

(6) EPS(Encapsulated PostScript)。用 PostScript 语言描述的 ASCII 图形文件,在 PostScript 图形打印机上能打印出高品质的图形(图像),最高能表示 32 位图形(图像)。该格式分为 Photoshop EPS 格式、Adobe Illustrator EPS 和标准 EPS 格式,其中后者又可以分为图形格式和图像格式。

(7) PSD(Photoshop Standard)。Photoshop 中的标准文件格式,专门为 Photoshop 而优化的格式。

(8) TGA(Tagged Graphic)。True Vision 公司为其显示卡开发的图形文件格式,创建时期较早,最高色彩数可达 32 位。

3.3.3 多媒体系统

1. 多媒体系统的组成

一般的多媒体系统由如下四部分内容组成:多媒体硬件系统、多媒体操作系统、媒体处理系统工具和用户应用软件。

1) 多媒体硬件系统

多媒体硬件系统包括计算机硬件、声音/视频处理器、多种媒体输入输出设备及信号转换装置、通信传输设备及接口装置等。其中,最重要的是根据多媒体技术标准而研制生成的多媒体信息处理芯片和板卡、光盘驱动器等。

2) 多媒体操作系统

多媒体操作系统也称为多媒体核心系统(Multimedia Kernel System),具有实时任务调度、多媒体数据转换和同步控制、对多媒体设备的驱动和控制,以及图形用户界面管理等。

3) 多媒体处理系统工具

多媒体处理系统工具也称为多媒体系统开发工具软件,是多媒体系统的重要组成部分。

4) 用户应用软件

根据多媒体系统终端用户要求而定制的应用软件,或面向某一领域的用户应用软件系统,它是面向大规模用户的系统产品。

2. 多媒体计算机的硬件

多媒体计算机的主要硬件除了常规的硬件如主机、软盘驱动器、硬盘驱动器、显示器、网卡之外,还有音频信息处理硬件、视频信息处理硬件及光盘驱动器等部分。

1) 音频卡

音频卡(Sound Card)用于处理音频信息,它可以把话筒、录音机、电子乐器等输入的声音信息进行模数(A/D)转换、压缩等处理,也可以把经过计算机处理的数字化的声音信号通过还原(解压缩)、数模(D/A)转换后用音箱播放出来,或者用录音设备记录下来。

2) 视频卡

视频卡(Video Card)用来支持视频信号(如电视)的输入与输出。

3) 采集卡

采集卡能将电视信号转换为计算机的数字信号,便于使用软件对转换后的数字信号进行剪辑处理、加工和色彩控制。还可将处理后的数字信号输出到录像带中。

4) 扫描仪

扫描仪将摄影作品、绘画作品或其他印刷材料上的文字和图像,甚至实物,扫描到计算机中,以便进行加工处理。

5) 光驱

光驱分为只读光驱(如 CD-ROM)和可读写光驱(如 CD-R,CD-RW),可读写光驱又称为刻录机,用于读取或存储大容量的多媒体信息。

3. 多媒体创作工具

多媒体创作系统是支持应用开发人员进行多媒体应用软件创作的工具。它能够用来集成各种媒体,并可设计阅读信息内容方式的软件。借助这种工具,应用人员可以不用编程也能做出很优秀的多媒体软件产品,极大地方便了用户。与之对应,多媒体创作工具必须担当起可视化编程的责任,它必须具有概念清晰、界面简洁、操作简单、功能伸缩性强等特点。

1) 媒体创作软件工具

应用较广泛的有三维图形视觉空间的设计和创作软件,如 Macromedia 公司的 Extreme 3D,它能提供包括建模、动画、渲染以及后期制作等诸多功能,直至专业级视频制作。Autodesk 公司的 2D Animation 和 3D Studio(包括 3ds Max)等也是媒体创作工具。而用于 MIDI 文件(数字化音乐接口标准)处理的音序器软件有 Music Time、Recording Session、Master Track Pro 和 Studio for Windows 等;至于波形声音工具,在 MDK(多媒体开放平台)中的 Wave Edit、Wave Studio 等。

2) 多媒体节目写作工具

第一种是基于脚本语言的写作工具,典型的如 Toolbook,它能帮助创作者控制各种媒体数据的播放。第二类是基于流程图的写作工具,典型的如 Authorware 和 IconAuthor,它们使用流程图来安排节目,每个流程图由许多图标组成,这些图标扮演脚本命令的角色,并与一个对话框对应,在对话框输入相应的内容即可。第三类写作工具是基于时序的,典型的如 Action,它们是通过将元素和时间轴线安排来达到使多媒体内容演示的同步控制。

3) 媒体播放工具

这一类软件非常多,其中 Video for Windows 就可以对视频序列(包括伴音)进行一系列处理,实现软件播放功能。而 Intel 公司推出的 Indeo 在技术上更进了一步,在纯软件视频播放上,还提供了功能先进的制作工具。

4) 其他各类媒体处理工具

还有其他几类软件,如多媒体数据库管理系统、Video CD 制作节目工具、基于多媒体板卡(如 MPEG 卡)的工具软件、多媒体出版系统工具软件、多媒体 CAI 制作工具、各式 MDK 等。它们在各领域中都受到很大欢迎。

4. 多媒体应用工具

下面从家用计算机的角度介绍几款现在最为流行的多媒体应用软件。

(1) 图形制作和图像浏览工具。Adobe Photoshop 和 3ds Max 是很出色的图像工具软件,CorelDRAW 和 Freehand 也是专业级的图像处理软件。

(2) 媒体播放和音频工具。Video for Windows 和 Multimedia Explorer 等都是不错的软件。

(3) 视频播放工具。视频播放可用豪杰超级解霸,也可用 WinDVD 和 Power DVD。

3.4 软件工程技术

软件工程(Software Engineering)是一门研究用工程化方法构建和维护有效的、实用的和高质量的软件的学科。它涉及程序设计语言、数据库、软件开发工具、系统平台、标准、设计模式等方面。

3.4.1 软件工程的定义

就软件工程的定义,很多学者、组织机构都分别给出了自己的定义。

(1) Barry Boehm: 运用现代科学技术知识来设计并构造计算机程序及为开发、运行和维护这些程序所必需的相关文件资料。

(2) IEEE 在软件工程专业术语汇编中的定义: 软件工程是将系统化的、严格约束的、量化的方法应用于软件的开发、运行和维护,即将工程化应用于软件。

(3) Fritz Bauer 在 NATO 会议上给出的定义: 建立并使用完善的工程化原则,以较经济的手段获得能在实际机器上有效运行的可靠软件的一系列方法。

(4) 《计算机科学技术百科全书》中的定义: 软件工程是应用计算机科学、数学及管理科学等原理,开发软件的工程。软件工程借鉴传统工程的原则、方法,以提高质量、降低成本。其中,计算机科学、数学用于构建模型与算法,工程科学用于制定规范、设计范型(Paradigm)、评估成本及确定权衡,管理科学用于计划、资源、质量、成本等管理。

3.4.2 软件工程的目标

软件工程的目标: 在给定成本、进度的前提下,开发出具有可修改性、有效性、可靠性、可理解性、可维护性、可重用性、可适应性、可移植性、可追踪性和可互操作性并且满足用户需求的软件产品。追求这些目标有助于提高软件产品的质量和开发效率,减少维护的困难。

软件工程活动是“生产一个最终满足需求且达到工程目标的软件产品所需要的步骤”,主要包括需求、设计、实现、确认以及支持等活动。需求活动包括问题分析和需求分析。问题分析获取需求定义,又称为软件需求规约。需求分析生成功能规约。设计活动一般包括概要设计和详细设计。概要设计建立整个软件体系结构,包括子系统、模块以及相关层次的说明、每一模块接口定义。详细设计产生程序员可用的模块说明,包括每一模块中数据结构说明及加工描述。实现活动把设计结果转换为可执行的程序代码。确认活动贯穿于整个开发过程,实现完成后的确认,保证最终产品满足用户的要求。

3.4.3 软件工程的过程

1. 软件过程

软件过程可概括为三类：基本过程类、支持过程类和组织过程类。

(1) 基本过程类包括获取过程、供应过程、开发过程、运作过程、维护过程和管理过程。

(2) 支持过程类包括文档过程、配置管理过程、质量保证过程、验证过程、确认过程、联合评审过程、审计过程以及问题解决过程。

(3) 组织过程类包括基础设施过程、改进过程以及培训过程。

2. 基本过程

软件过程主要针对软件生产和管理进行研究。为了获得满足工程目标的软件,不仅涉及工程开发,而且还涉及工程支持和工程管理。对于一个特定的项目,可以通过剪裁过程定义所需的活动和任务,并可使活动并发执行。与软件有关的单位,根据需求和目标,可采用不同的过程、活动和任务,生产一个最终能满足需求且达到工程目标的软件产品所需要的步骤。

软件工程过程主要包括开发过程、运作过程、维护过程。它们覆盖了需求、设计、实现、确认以及维护等活动。

3.4.4 新的软件开发技术

1. 软件复用

软件复用(Software Reuse)就是将已有的软件成分用于构造新的软件系统,以缩减软件开发和维护的花费。无论对可复用构件原封不动地使用还是做适当的修改后再使用,只要是用来构造新软件,则都可称为复用。被复用的软件成分一般称为可复用构件。软件复用是提高软件生产力和质量的一种重要技术。早期的软件复用主要是代码级复用,后来扩大到包括领域知识、开发经验、项目计划、可行性报告、体系结构、需求、设计、测试用例和文档等一切有关方面。对一个软件进行修改,使它运行于新的软硬件平台不称为复用,而称为软件移植。软件复用可以划分为如下复用级别。

1) 代码的复用

代码的复用包括目标代码和源代码的复用。其中目标代码的复用级别最低,历史也最久,当前大部分编程语言的运行支持系统都提供了连接(Link)、绑定(Binding)等功能来支持这种复用。源代码的复用级别略高于目标代码的复用,程序员在编程时把一些想复用的代码段复制到自己的程序中,但这样往往会产生一些新旧代码不匹配的错误。想大规模地实现源程序的复用只有依靠含有大量可复用构件的构件库。如“对象链接及嵌入”(OLE)技术,既支持在源程序级定义构件并用以构造新的系统,又使这些构件在目标代码的级别上仍然是一些独立的可复用构件,能够在运行时被灵活地重新组合为各种不同的应用。

2) 设计的复用

设计结果比源程序的抽象级别更高,因此它的复用受实现环境的影响较少,从而使可复

用构件被复用的机会更多,并且所需的修改更少。这种复用有三种途径:第一种途径是从现有系统的设计结果中提取一些可复用的设计构件,并把这些构件应用于新系统的设计;第二种途径是把一个现有系统的全部设计文档在新的软硬件平台上重新实现,也就是把一个设计运用于多个具体的实现;第三种途径是独立于任何具体的应用,有计划地开发一些可复用的设计构件。

3) 分析的可复用

这是比设计结果更高级别的复用,可复用的分析构件是针对问题域的某些事物或某些问题的抽象程度更高的解法,受设计技术及实现条件的影响很少,所以可复用的机会更大。复用的途径也有三种,即从现有系统的分析结果中提取可复用构件用于新系统的分析;用一份完整的分析文档作为输入产生针对不同软硬件平台和其他实现条件的多项设计;独立于具体应用,专门开发一些可复用的分析构件。

4) 测试信息的复用

测试信息的复用主要包括测试用例的复用和测试过程信息的复用。前者是把一个软件的测试用例在新的软件测试中使用,或者在软件做出修改时在新的一轮测试中使用。后者是在测试过程中通过软件工具自动地记录测试的过程信息,包括测试员的每一个操作、输入参数、测试用例及运行环境等一切信息。这种复用的级别,不便和分析、设计、编程的复用级别做准确的比较,因为被复用的不是同一事物的不同抽象层次,而是另一种信息,但从这些信息的形态看,大体处于与程序代码相当的级别。

2. 软件产品线

1) 软件产品线的概念

软件产品线是一组具有共同体系构架和可复用组件的软件系统,它们共同构建支持特定领域内产品开发的软件平台。一个软件产品线由一个产品线体系结构、一个可重用构件集合和一个源自共享资源的产品集合组成,是组织一组相关软件产品开发的方式。软件产品线的产品则是根据基本用户需求对产品线架构进行定制,将可复用部分和系统独特部分集成而得到。软件产品线方法集中体现一种大规模、大粒度软件复用实践,是软件工程领域中软件体系结构和软件重用技术发展的结果。

1997年,由北京大学主持的国家重大科技攻关项目“青鸟工程”是软件产品线方法的原型平台。进入21世纪,为了适应Internet应用及信息技术方面的重大变革,软件系统开始呈现出一种柔性可演化、连续反应式、多目标自适应的新系统形态。从技术的角度看,在面向对象、软件构件等技术支持下的软件实体以主体化的软件服务形式存在于Internet的各个节点之上,各个软件实体相互间通过协同机制进行跨网络的互联、互通、协作和联盟,从而形成一种与WWW相类似的软件Web(Software Web)。将这样一种Internet环境下的新的软件形态称为网构软件(Internetware)。

2) 软件产品线的流程

软件产品线的开发有四个技术特点:过程驱动、特定领域、技术支持和以架构为中心。与其他软件开发方法相比,选择软件产品线的宏观原因有:对产品线及其实现所需的专家知识领域的清楚界定,对产品线的长期远景进行了策略性规划。软件生产线的概念和思想,将软件的生产过程分到三类不同的生产车间进行,即应用体系结构生产车间、构件生产车间

和基于构件、体系结构复用的应用集成(组装)车间,从而形成软件产业内部的合理分工,实现软件的产业化生产。软件生产线如图 3-10 所示。

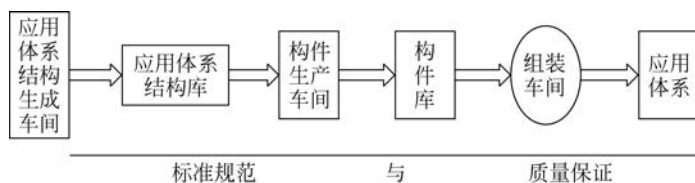


图 3-10 软件生产线

(1) 软件产品线工程。软件产品线是一种基于架构的软件复用技术,它的理论基础是:特定领域(产品线)内的相似产品具有大量的公共部分和特征,通过识别和描述这些公共部分和特征,可以开发需求规范、测试用例、软件组件等产品线的公共资源。而这些公共资源可以直接应用或适当调整后应用于产品线内产品的开发,从而不再从草图开始开发产品。因此典型的产品线开发过程包括两个关键过程:领域工程和应用工程。

(2) 软件产品线的组织结构。软件产品线开发过程分为领域工程和应用工程,相应的软件开发的组织结构也有两部分:负责核心资源的小组和负责产品的小组。在 EMS 系统开发过程中采用的产品线方法中,主要有三个关键小组:平台组、配置管理组和产品组。

(3) 软件产品线构件。软件产品线构件是用于支持产品线中产品开发的可复用资源的统称。这些构件远不是一般意义上的软件构件,它们包括领域模型、领域知识、产品线构件、测试计划及过程、通信协议描述、需求描述、用户界面描述、配置管理计划及工具、代码构件、性能模型与度量、 workflow 结构、预算与调度、应用程序生成器、原型系统、过程构件(方法、工具)、产品说明、设计标准、设计决策、测试脚本等。在产品线系统的每个开发周期都可以对这些构件进行精化。

3. 敏捷开发

2001 年 2 月 11—13 日,在犹他州 Wasatch 山的滑雪胜地,17 个计算机专家在两天的聚会中,签署了“敏捷软件开发宣言”(The Manifesto for Agile Software Development),宣告:“我们通过实践寻找开发软件的更好方法,并帮助其他人使用这些方法。通过这一工作我们得到以下结论:个体和交流胜于过程和工具;工作软件胜于综合文档;客户协作胜于洽谈协议;回应变革胜于照计划行事。”

1) 方法类型

敏捷过程(Agile Process)来源于敏捷开发。敏捷开发是一种应对快速变化的需求的一种软件开发能力。相对于“非敏捷”更强调沟通、变化、产品效益,也更注重作为软件开发中人的作用。敏捷开发包括一系列的方法,主流的有如下七种。

(1) XP。XP(极限编程)的思想源自 Kent Beck 和 Ward Cunningham 在软件项目中的合作经历。XP 注重的核心是沟通、简明、反馈和勇气。因为知道计划永远赶不上变化,XP 无须开发人员在软件开始初期做出很多的文档。XP 提倡测试先行,为了将以后出现 bug(缺陷)的概率降到最低。

(2) SCRUM 方法。SCRUM 是一种迭代的增量化过程,用于产品开发或工作管理。它是一种可以集合各种开发实践的经验化过程框架。SCRUM 中发布产品的重要性高于一切。该方法由 Ken Schwaber 和 Jeff Sutherland 提出,旨在寻求充分发挥面向对象和构件技术的开发方法,是对迭代式面向对象方法的改进。

(3) Crystal Methods。Crystal Methods(水晶方法族)由 Alistair Cockburn 在 20 世纪 90 年代末提出。之所以是个系列,是因为他相信不同类型的项目需要不同的方法。虽然水晶系列没有 XP 那样的产出效率,但会有更多的人能够接受并遵循它。

(4) FDD。FDD(Feature-Driven Development,特性驱动开发)由 Peter Coad、Jeff de Luca、Eric Lefebvre 共同开发,是一套针对中小型软件开发项目的开发模式。此外,FDD 是一个模型驱动的快速迭代开发过程,它强调的是简化、实用,易于被开发团队接受,适用于需求经常变动的项目。

(5) ASD。ASD(Adaptive Software Development,自适应软件开发)由 Jim Highsmith 在 1999 年正式提出。ASD 强调开发方法的适应性(Adaptive),这一思想来源于复杂系统的混沌理论。ASD 不像其他方法那样有很多具体的实践做法,它更侧重为 ASD 的重要性提供最根本的基础,并从更高的组织和管理层次来阐述开发方法为什么要具备适应性。

(6) DSDM。DSDM(Dynamic System Development Method,动态系统开发方法)是众多敏捷开发方法中的一种,它倡导以业务为核心,快速而有效地进行系统开发。实践证明 DSDM 是成功的敏捷开发方法之一。在英国,由于其在各种规模的软件组织中的成功应用,它已成为应用最为广泛的快速应用开发方法。DSDM 不但遵循了敏捷方法的原理,而且也适合那些成熟的传统开发方法及有坚实基础的软件组织。

(7) 轻量型 RUP 框架。轻量型 RUP 框架其实是一个过程的框架,它可以包允许许多不同类型的过程,Craig Larman 极力主张以敏捷型方式来使用 RUP。他的观点是:目前如此众多的努力以推进敏捷型方法,只不过是接受能被视为 RUP 的主流 OO 开发方法而已。

2) 敏捷开发的工作方式

大数据技术的这四个核心部分会导致高度迭代式的、增量式的软件开发过程,并在每次迭代结束时交付经过编码与测试的软件。敏捷开发小组的主要工作方式包括:增量与迭代式开发;作为一个整体工作;按短迭代周期工作;每次迭代交付一些成果;关注业务优先级;检查与调整。

(1) 增量与迭代。增量开发,意思是每次递增地添加软件功能。每一次增量都会添加更多的软件功能。迭代式开发允许在每次迭代过程中需求可能有变化,通过不断细化来加深对问题的理解。

(2) 敏捷小组的整体工作。项目取得成功的关键在于,所有的项目参与者都把自己看作朝向一个共同目标前进的团队的一员。一个成功的敏捷开发小组应该具有“我们一起参与其中”的思想。虽然敏捷开发小组是以小组整体进行工作,但是小组中仍然有一些特定的角色。有必要指出和阐明那些在敏捷估计和规划中承担一定任务的角色。

(3) 敏捷小组的短迭代周期。迭代是受时间框(Timebox)限制的,意味着即使放弃一些功能,也必须按时结束迭代。时间框一般很短。大部分敏捷开发小组采用 2~4 周的迭代,但也有一些小组采用长达 3 个月的迭代周期仍能维持敏捷性。大多数小组采用相

对稳定的迭代周期长度,但是也有一些小组在每次迭代开始的时候选择合适的周期长度。

(4) 敏捷小组每次迭代交付。在每次迭代结束时让产品达到潜在可交付状态是很重要的。实际上,这并不是说小组必须全部完成发布所需的所有工作,因为它们通常并不会每次迭代都真的发布产品。由于单次迭代并不总能提供足够的时间来完成足够满足用户或客户需要的新功能,因此需要引入更广义的发布(Release)概念。一次发布由一次或以上(通常是以上)相关功能的迭代组成。最常见的迭代一般是2~4周,一次发布通常是2~6个月。

(5) 敏捷小组的优先级。敏捷开发小组从两方面显示出它们对业务优先级的关注。首先,它们按照产品所有者所制定的顺序交付功能,而产品所有者一般会按照使机构在项目上的投资回报最大化的方式来确定功能的优先级,并将它们组织到产品发布中。要达到这一目的,需要根据开发小组的能力和所需新功能的优先级建立一个发布计划。其次,敏捷开发小组关注完成和交付具有用户价值的功能,而不是完成孤立的任务(任务最终组合成具有用户价值的功能)。

(6) 敏捷小组的检查和调整。在每次新迭代开始时,敏捷开发小组都会结合在上一次迭代中获得的所有新知识做出相应的调整。如果小组认识到一些可能影响到计划的准确性或是价值的内容,它们就会调整计划。小组可能发现它们过高或过低地估计了自己的进展速度,或者发现某项工作比原来以为的更耗费时间,从而影响到计划的准确性。

4. CMM 方法

CMM 是指“能力成熟度模型”,其英文全称为 Capability Maturity Model for Software,英文缩写为 SW-CMM,简称 CMM。它是对于软件组织在定义、实施、度量、控制和改善其软件过程的实践中各个发展阶段的描述。CMM 的核心是把软件开发视为一个过程,并根据这一原则对软件开发和维护进行过程监控和研究,以使其更加科学化、标准化,使企业能够更好地实现商业目标。

CMM 是一种用于评价软件承包能力并帮助其改善软件质量的方法,侧重于软件开发过程的管理及工程能力的提高与评估。CMM 分为五个等级:一级为初始级;二级为可重复级;三级为已定义级;四级为已管理级;五级为优化级。

CMM 为软件企业的过程能力提供了一个阶梯式的改进框架,它基于过去所有软件工程过程改进的成果,吸取了以往软件工程的经验教训,提供了一个基于过程改进的框架,指明了一个软件组织在软件开发方面需要管理哪些主要工作、这些工作之间的关系,以及以怎样的先后次序,一步一步地做好这些工作而使软件组织走向成熟。

3.5 信息安全技术基础

信息安全本身包括的范围很大。大到国家军事政治等机密安全,小到如防范商业企业机密泄露、防范青少年对不良信息的浏览、个人信息的泄露等。网络环境下的信息安全体系是保证信息安全的關鍵,包括计算机安全操作系统、各种安全协议、安全机制(数字签名、信息认证、数据加密等),直至安全系统,其中任何一个安全漏洞便可以威胁全局安全。信息安

全服务至少应该包括支持信息网络安全服务的基本理论,以及基于新一代信息网络体系结构的网络安全服务体系结构。

3.5.1 信息安全概述

信息安全是指信息网络的硬件、软件及其系统中的数据受到保护,不受偶然的或者恶意的原因而遭到破坏、更改、泄露,系统连续、可靠、正常地运行,信息服务不中断。信息安全的实质就是要保护信息系统或信息网络中的信息资源免受各种类型的威胁、干扰和破坏,即保证信息的安全性。根据国际标准化组织的定义,信息安全性的含义主要是指信息的完整性、可用性、保密性和可靠性。信息安全是任何国家、政府、部门、行业都必须十分重视的问题,是一个不容忽视的国家安全战略。但是,对于不同的部门和行业来说,其对信息安全的要求和重点却是有区别的。

我国的改革开放带来了各方面信息量的急剧增加,并要求大容量、高效率地传输这些信息。为了适应这一形势,通信技术发生了前所未有的爆炸性发展。目前,除有线通信外,短波、超短波、微波、卫星等无线电通信也正在越来越广泛地应用。与此同时,国外敌对势力为了窃取我国的政治、军事、经济、科学技术等方面的秘密信息,运用侦察台、侦察船、卫星等手段,形成固定与移动、远距离与近距离、空中与地面相结合的立体侦察网,截取我国通信传输中的信息。

3.5.2 信息安全威胁

信息系统安全领域存在的挑战:系统太脆弱,太容易受攻击;被攻击时很难及时发现和制止;有组织、有计划地入侵,无论在数量上还是在质量上都呈现快速增长趋势;在规模和复杂程度上不断扩展网络而很少考虑其安全状况的变化情况;因信息系统安全导致的巨大损失并没有得到充分重视,而有组织的犯罪、情报和恐怖组织却深谙这种破坏的威力。几种典型的安全威胁如图 3-11 所示。

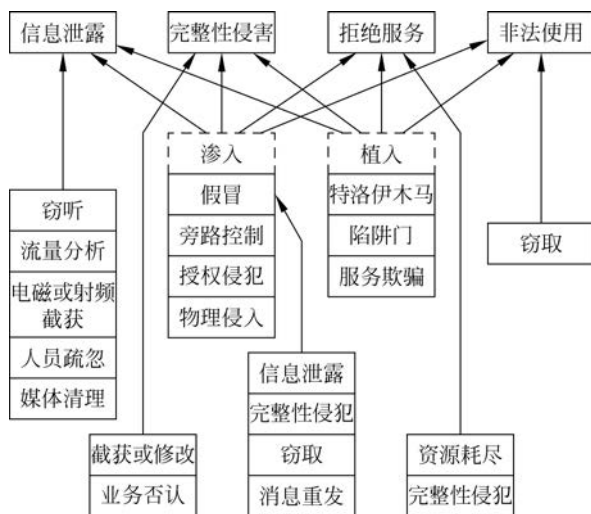


图 3-11 安全威胁

3.5.3 信息安全的目标和原则

1. 信息安全的目标

所有的信息安全技术都是为了达到一定的安全目标,其核心包括保密性、完整性、可用性、可控性和不可否认性共五个安全目标。

1) 保密性

保密性(Confidentiality)是指阻止非授权的主体阅读信息。它是信息安全一诞生就具有的特性,也是信息安全主要的研究内容之一。更通俗地讲,就是说未授权的用户不能够获取敏感信息。对纸质文档信息,我们只需要保护好文件,不被非授权者接触即可。而对计算机及网络环境中的信息,不仅要制止非授权者对信息的阅读,也要阻止授权者将其访问的信息传递给非授权者,以致信息被泄露。

2) 完整性

完整性(Integrity)是指防止信息被未经授权的主体篡改。它使保护信息保持原始的状态,使信息保持其真实性。如果这些信息被蓄意地修改、插入、删除等,形成虚假信息,将带来严重的后果。

3) 可用性

可用性(Usability)是指授权主体在需要信息时能及时得到服务的能力。可用性是在信息安全保护阶段对信息安全提出的新要求,也是在网络化空间中必须满足的一项信息安全要求。

4) 可控性

可控性(Controllability)是指对信息和信息系统实施安全监控管理,防止非法利用信息和信息系统。

5) 不可否认性

不可否认性(Non-repudiation)是指在网络环境中,信息交换的双方不能否认其在交换过程中发送信息或接收信息的行为。

信息安全的保密性、完整性和可用性主要强调对非授权主体的控制。信息安全的可控性和不可否认性恰恰是通过对授权主体的控制,实现对保密性、完整性和可用性的有效补充,主要强调授权用户只能在授权范围内进行合法的访问,并对其行为进行监督和审查。

2. 信息安全的原则

为了达到信息安全的目标,各种信息安全技术的使用必须遵守一些基本的原则。

1) 最小化原则

受保护的敏感信息只能在一定范围内被共享,履行工作职责和职能的安全主体,在法律和相关安全策略允许的前提下,为满足工作需要,仅被授予其访问信息的适当权限,称为最小化原则。敏感信息的“知情权”一定要加以限制,是在“满足工作需要”前提下的一种限制性开放。

2) 分权制衡原则

在信息系统中,对所有权限应该进行适当的划分,使每个授权主体只能拥有其中的一部分权限,使他们之间相互制约、相互监督,共同保证信息系统的安全。如果一个授权主体分

配的权限过大,无人监督和制约,就隐含了“滥用权力”“一言九鼎”的安全隐患。

3) 安全隔离原则

隔离和控制是实现信息安全的基本方法,而隔离是进行控制的基础。信息安全的一个基本策略就是将信息的主体与客体分离,按照一定的安全策略,在可控和安全的前提下实施主体对客体的访问。

在这些基本原则的基础上,人们在生产实践过程中还总结出一些实施原则,它们是基本原则的具体体现和扩展,包括整体保护原则、谁主管谁负责原则、适度保护的等级化原则、分域保护原则、动态保护原则、多级保护原则、深度保护原则和信息流向原则等。

3.5.4 信息安全策略

信息安全策略是指为保证提供一定级别的安全保护所必须遵守的规则。实现信息安全,不但靠先进的技术,而且也得靠严格的安全管理、法律约束和安全教育。

1. 应用先进的信息安全技术

用户对自身面临的威胁进行风险评估,决定其所需要安全服务种类,选择相应的安全机制,然后集成先进的安全技术,形成一个全方位的安全系统,它是网络安全的根本保证。

2. 建立严格的的安全管理制度

计算机网络使用机构应建立相应的网络安全管理办法,加强内部管理,建立合适的网络安全管理系统,加强用户管理和授权管理,建立安全审计和跟踪体系,提高整体网络安全意识。

3. 制定严格的法律、法规

计算机网络是一种新生事物,它的许多行为无法可依,无章可循,导致网络上计算机犯罪处于无序状态。面对日趋严重的网络犯罪,必须建立与网络安全相关的法律、法规,使不法分子慑于法律,不敢轻举妄动。

4. 启用安全操作系统

给系统中的关键服务器提供安全运行平台,构成安全 WWW 服务、安全 FTP 服务、安全 SMTP 服务等,并作为各类网络安全产品的坚实底座,确保这些安全产品的自身安全。

3.5.5 信息安全的相关技术

1. 用户身份认证

用户身份认证是安全的第一道大门,是各种安全措施可以发挥作用的前提。用户身份认证技术包括静态密码、动态密码(短信密码、动态口令牌、手机令牌)、USB Key、IC 卡、数字证书、指纹、虹膜等。

2. 防火墙

防火墙在某种意义上可以说是一种访问控制产品。它在内部网络与不安全的外部网络

之间设置障碍,阻止外界对内部资源的非法访问,防止内部对外部的不安全访问。主要技术有包过滤技术、应用网关技术、代理服务技术。防火墙能够较为有效地防止黑客利用不安全的服务对内部网络的攻击,并且能够实现数据流的监控、过滤、记录和报告功能,较好地隔断内部网络与外部网络的连接。但其本身可能存在安全问题,也可能会是一个潜在的瓶颈。

3. 网络安全隔离

网络安全隔离有两种方式:一种是采用隔离卡实现的;另一种是采用网络安全隔离网闸实现的。隔离卡主要用于对单台机器的隔离,网闸主要用于对于整个网络的隔离。这两者的区别可参考相关资料。网络安全隔离与防火墙的区别可参考相关资料。

4. 安全路由器

由于 WAN 连接需要专用的路由器设备,因此可通过路由器来控制网络传输。通常采用访问控制列表技术来控制网络信息流。

5. 虚拟专用网

虚拟专用网(VPN)是在公共数据网络上,通过采用数据加密技术和访问控制技术,实现两个或多个可信内部网之间的互联。VPN 的构筑通常都要求采用具有加密功能的路由器或防火墙,以实现数据在公共信道上的可信传递。

6. 安全服务器

安全服务器主要针对一个局域网内部信息存储、传输的安全保密问题,其实现功能包括对局域网资源的管理和控制、对局域网内用户的管理,以及局域网中所有安全相关事件的审计和跟踪。

7. 电子签证机构

电子签证机构(CA)作为通信的第三方,为各种服务提供可信任的认证服务。CA 可向用户发行电子签证证书,为用户提供成员身份验证和密钥管理等功能。PKI 产品可以提供更多的功能和更好的服务,将成为所有应用的计算基础结构的核心部件。

8. 安全管理中心

由于网上的安全产品较多,且分布在不同的位置,这就需要建立一套集中管理的机制和设备,即安全管理中心。它用来给各网络安全设备分发密钥,监控网络安全设备的运行状态,负责收集网络安全设备的审计信息等。

9. 入侵检测系统

入侵检测系统(IDS)作为传统保护机制(如访问控制、身份识别等)的有效补充,形成了信息系统中不可或缺的反馈链。

10. 入侵防御系统

入侵防御系统(IPS)作为 IDS 很好的补充,是信息安全发展过程中占据重要位置的计算机网络硬件。

11. 安全数据库

由于大量的信息存储在计算机数据库内,有些信息是有价值的,也是敏感的,需要保护。安全数据库可以确保数据库的完整性、可靠性、有效性、机密性、可审计性及存取控制与用户身份识别等。

12. 安全操作系统

安全操作系统给系统中的关键服务器提供安全运行平台,构成安全 WWW 服务、安全 FTP 服务、安全 SMTP 服务等,并作为各类网络安全产品的坚实底座,确保这些安全产品的自身安全。

13. 信息安全服务

信息安全服务是指为确保信息和信息系统的完整性、保密性和可用性所提供的信息技术专业服务,包括对信息系统安全的咨询、集成、监理、测评、认证、运维。

14. 数据加密

数据加密技术从技术实现上分为软件和硬件两种。按作用不同,数据加密技术主要分为数据传输、数据存储、数据完整性的鉴别以及密钥管理技术这 4 种。

3.6 人工智能

3.6.1 人工智能概述

1. 人工智能的定义

人工智能(Artificial Intelligence, AI)是一门综合计算机科学、生理学、哲学等学科的交叉科学。其研究课题涵盖很广,从机器视觉到专家系统,包括许多不同的领域。其目的就是让机器学会“思考”。

人工智能学科是计算机科学中涉及研究、设计和应用智能机器的一个分支。它的目标是研究用机器模仿和执行人脑的某些智慧功能,并开发相关理论和技术。

人工智能是智能机器所执行的通常与人类智能有关的智能行为,如判断、推理、证明、识别、感知、理解、通信、设计、思考、规划、学习和问题求解等思维活动。

人工智能自诞生以来,从符号主义、联结主义到行为主义的变迁,这些研究从不同角度模拟人类智能,在各自研究中都取得了很大的成就。

2. 人工智能的学派

符号主义又称为逻辑主义、心理学派或计算机学派,其原理主要为物理符号系统假设和

有限合理性原理。符号主义认为,人工智能源于数学逻辑,人的认知基元是符号,而且认知过程即符号操作过程,通过分析人类认知系统所具备的功能和机能,然后用计算机模拟这些功能,来实现人工智能。符号主义主要困难主要表现在机器博弈的困难;机器翻译不完善;人的基本常识问题表现的不足。

联结主义又称为仿生学派或生理学派,其原理主要为神经网络及神经网络间的连接机制与学习算法。联结主义认为,人工智能源于仿生学,特别是人脑模型的研究,人的思维基元是神经元,而不是符号处理过程,因而人工智能应着重于结构模拟,也就是模拟人的生理神经网络结构,功能、结构和智能行为是密切相关的,不同的结构表现出不同的功能和行为。所谓人工神经网络模拟,也即通过改变神经元之间的连接强度来控制神经元的活动,以模拟生物的感知与学习能力,可用于模式识别、联想记忆等。联结主义主要困难主要表现在对知识获取在技术上的困难;模拟人类心智方面的局限。

行为主义又称进化主义或控制论学派。行为主义认为,人工智能源于控制论,智能取决于感知和行动,提出了智能行为的“感知—动作”模式,智能不需要知识、表示和推理;人工智能可以像人类智能一样逐步进化;智能行为只能在现实世界中与周围环境交互作用而表现出来。

3. 人工智能的现状

早年,亚里士多德试图说明人与机器是同一关系。法国哲学家拉·梅特里最早提出“人是机器”理念,1950年,人工智能先祖——图灵提出著名的图灵实验,给出了判断机器人的标准。机械论一直试图说明“人是机器”,但最终未能将心灵与物质统一。大数据时代出现了人机合一的趋势,即物质实体的机器与人身的融合。人工智能延伸了人脑的功能,也扩展了人的本质力量。人工智能可能的突破是人的本质力量延伸,发展出新型智能集成体。

1) 自助式人工智能系统

自主的字面意思指自己做主,不受别人支配。在人工智能中,自主系统(Autonomous System)指能够在无人干预情况下对周围环境进行独立感知和识别,并对下一步行为做出自我判断和决策的人工智能系统。传统认知指人认识外界事物或对外界事物信息加工的过程。其能力指人脑加工、存储和提取信息的能力,包括感觉(对外界的反映)、知觉(对事物整体的认识)、记忆、注意(对某一对象的指向和集中)、思维(对客观事物的概括和间接的反应,其反映是事物本质和事物间规律性的联系)、想象(对已储存的表象进行加工改造形成新形象的心理过程)和语言(人能表达或理解语句,辨析有歧义的语句,判别不同语句的实际语义)能力。

2) 自助式人工智能系统的功能

目前,自主式人工智能系统已具备感知、认知、学习、思维、推理、判断和决策的能力,其智力已到达孩童的水平。

(1) 自主式人工智能系统已经具备了独立感知对象、获取数据、存储、识别对象、推理、判断和决策的能力。自主式人工智能系统对外界的感知是通过智能传感器、探测仪等实现的。其首先是定位和标识目标对象;第二步是获取目标数据,或直接采集数字信号或将模拟信号转换为数字信号;第三步是与信息-物理融合系统(CPS)连接,将数据传给计算机存储;第四步是利用模式识别技术将已获取的数据与对象进行模式匹配,以识别预定目标。

目前模式识别技术在以下几方面表现不俗,如(印刷和手写)文字识别、车牌自动识别、人脸识别、导弹对攻击目标的自动识别等。现以巡航导弹为例,说明自主式人工智能系统的自主能力,当然也包括其认知能力。巡航导弹攻击固定目标发射后,地形匹配制导雷达利用存储的地图与实际地形做比较,确定导弹的位置,校正飞行路线。弹载全球定位系统与至少四颗卫星联系,接收导航信号,确定飞行状况。如果飞行轨道偏离,其人工智能系统会自动纠错。其目标区域末端导航由光学数字场景匹配区域关联系统提供,将存储的目标图像与实际目标图像比较,确定既定目标,及自动发起攻击。

(2) 自主式人工智能系统已经初步具备了独立认知能力。其具体表现在:感知(如传感器对外界的反映)能力、知觉(如导弹中模式识别对目标对象的整体识别)能力、记忆(计算机的存储)能力、注意(如巡航导弹对被攻击对象的指向和寻找)能力、思维(如谷歌自动驾驶汽车对外界环境中人和物体的识别,及人和物体关系的识别。再如 2016 年 Facebook 人脸识别技术的识别率已经达到了 97.25%。不仅如此,它还可以对人与人关系进行识别,及对人与人关系大数据规律进行分析)能力、想象(如《阿凡达》电影的拍摄,首先是人物形体的动作捕捉,在底片扫描完成后,使用诸如 BOUJOU、PFTRACK 等软件跟踪镜头,捕捉摄影机运动轨迹并将其送入三维软件合成和渲染,就可以得到电影人物阿凡达的画面。电影中新的人物形象是原拍摄人物在计算机中存储图像的更新和替换)能力和语言(如机器能听懂人的语言,并能自动应答其问题)能力。

(3) 自主式人工智能系统也表现出较强的学习能力。机器学习学科主要研究计算机模拟人类学习行为和过程,获得新的数据、信息、知识或技能,然后重组知识结构,以便改善自身性能。例如人机曾经经过多次象棋大战,以前多是机器败北。但 2016 年 3 月 15 日韩国围棋选手李世石 1:4 憾负人工智能阿尔法狗(AlphaGo)。这说明机器具有较强的学习能力。十八岁人工智能美少女“小冰”,就是微软打造的一款人工智能交互主体人物。她是诗人、歌手、主持人,也是画家和设计师。与其他人工智能不同,“小冰”注重人工智能在拟合人类情商维度的发展,强调人工智能情商,而非任务完成,并不断学习人类创造者的能力。2021 年 9 月 22 日第九代“小冰”发布会称,她已实现仅用 200 个对话样本达到 4.19 的高分,这已非常接近人类。目前深度内嵌“小冰”的智能设备已累计超过 10 亿台,已成为全球规模最大的第三方跨设备人工智能系统。ChatGPT 是美国人工智能研究实验室 OpenAI 推出的一种人工智能技术驱动的自然语言处理工具,使用 Transformer 神经网络架构,拥有语言理解和文本生成能力,其连接大量语料库来训练模型,语料库包含真实世界中的对话,使 ChatGPT 具备上知天文下知地理,还能根据聊天的上下文进行互动,做到与真正人类几乎无异的聊天交流。ChatGPT 不单是聊天机器人,还能撰写邮件、视频脚本、文案、翻译、代码等。ChatGPT 于 2022 年 11 月 30 日发布,2023 年 1 月末其月活用户已突破 1 亿,成为史上增长最快的消费者应用。清华大学的原创虚拟学生“华智冰”与微软的“小冰”类似,其也展现了国人在人工智能方面的智慧。人工智能在语音识别、图像分类、机器翻译、可穿戴设备、无人驾驶汽车、医疗诊断等方面均取得了突破性进展,标志着机器已具备了初级学习的能力。

(4) 自主式人工智能系统的智力水平已经达到了一个新的高度。2011 年超级计算机沃森与人类鏖战智力问答完胜,这一结果表明,计算机已经具备了对人类提问的自主判断和自动应答的图灵智能实验测试水平,实验结果证明计算机已具备了人的智能。近年计算机

的深度学习能力异军突起,它借助神经网络模拟人脑进行分析学习,使强人工智能水平大增,特别适合于解析大数据问题。2014年年末谷歌宣布对人类神经大脑模拟系统研究成果:递归神经网络已进一步实现更强的“逻辑推理能力”。2015年10月美国伊利诺伊大学的研究小组完成的测试发现,目前最先进人工智能系统的智力相当于4岁孩子的水平。2016年谷歌智商测试表明,其人工智能水平已经接近6岁儿童。

3.6.2 人工智能相关学科

1. 机器学习概述

机器学习(Machine Learning)主要研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。它是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域,它主要使用归纳、综合而不是演绎。

机器学习在人工智能的研究中具有十分重要的地位。机器学习逐渐成为人工智能研究的核心之一。它的应用已遍及人工智能的各个分支,如专家系统、自动推理、自然语言理解、模式识别、计算机视觉、智能机器人等领域。其中尤其典型的是专家系统中的知识获取瓶颈问题,人们一直在努力试图采用机器学习的方法加以克服。

机器学习的研究是根据生理学、认知科学等对人类学习机理的了解,建立人类学习过程的计算模型或认识模型,发展各种学习理论和学习方法,研究通用的学习算法并进行理论上的分析,建立面向任务的具有特定应用的学习系统。这些研究目标相互影响,相互促进。

机器学习已经有了十分广泛的应用,例如专家系统、认知模拟、规划和问题求解、数据挖掘、计算机视觉、自然语言处理、生物特征识别、搜索引擎、医学诊断、网络信息服务、图像识别、故障诊断、自然语言理解、机器人和博弈、检测信用卡欺诈、DNA序列测序、语音和手写识别、证券市场分析和计算机游戏等。

2. 决策支持系统概述

决策支持系统(Decision Support System, DSS)是辅助决策者通过数据、模型和知识,以人机交互方式进行半结构化或非结构化决策的计算机应用系统。它是管理信息系统向更高一级发展而产生的先进信息管理系统。它为决策者提供分析问题、建立模型、模拟决策过程和方案的环境,调用各种信息资源和分析工具,帮助决策者提高决策水平和质量。

决策支持系统基本结构主要由四部分组成,即数据部分、模型部分、推理部分和人机交互部分。数据部分是一个数据库系统;模型部分包括模型库及其管理系统;推理部分由知识库、知识库管理系统和推理机组成;人机交互部分是决策支持系统的人机交互界面,用以接收和检验用户请求,调用系统内部功能软件为决策服务,使模型运行、数据调用和知识推理达到有机的统一,有效地解决决策问题。

3. 专家系统

专家系统是一种智能计算机程序系统,其内部含有大量的某个领域专家水平的知识与经验,能够利用人类专家的知识和解决问题的方法来处理该领域问题。也就是说,专家系统

是一个具有大量的专门知识与经验的程序系统,它应用人工智能技术和计算机技术,根据某领域一个或多个专家提供的知识和经验,进行推理和判断,模拟人类专家的决策过程,以便解决那些需要人类专家处理的复杂问题。简而言之,专家系统是一种模拟人类专家解决领域问题的计算机程序系统。

专家系统是人工智能中最重要的也最活跃的一个应用领域,它实现了人工智能从理论研究走向实际应用、从一般推理策略探讨转向运用专门知识的重大突破。自从1977年以来,知识工程的研究,专家系统的理论和技术不断发展,应用渗透到几乎各个领域,包括化学、数学、物理、生物、医学、农业、气象、地质勘探、军事、工程技术、法律、商业、空间技术、自动控制、计算机设计和制造等众多领域,开发了几千个专家系统,其中不少在功能上已达到,甚至超过同领域中人类专家的水平,并在实际应用中产生了巨大的经济效益。

4. 深度学习

深度学习的概念源于人工神经网络的研究。含多隐层的感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,以发现数据的分布式特征表示。深度学习是机器学习中一种基于对数据进行表征学习的方法。观测值可以使用多种方式来表示,如每个图像像素强度值的向量,或者更抽象地表示成一系列边、特定形状的区域等。而使用某些特定的表示方法更容易从实例中学习任务(如人脸识别或面部表情识别)。深度学习的好处是用非监督式或半监督式的特征学习和分层特征提取等高效算法来替代手工获取特征。

从一个输入中产生一个输出所涉及的计算可以通过一个流向图来表示。流向图是一种能够表示计算的图,在这种图中每个节点表示一个基本的计算及一个计算的值,计算的结果被应用到这个节点的子节点的值。考虑这样一个计算集合,它可以被允许在每个节点和可能的图结构中,并定义了一个函数族。输入节点没有父节点,输出节点没有子节点。这种流向图的一个特别属性是深度,即从一个输入到一个输出的最长路径的长度。

深度机器学习方法分有监督学习与无监督学习。不同的学习框架下建立的学习模型很是不同。例如,卷积神经网络(Convolutional Neural Network, CNN)就是一种深度的监督学习下的机器学习模型,而深度置信网(Deep Belief Network, DBN)就是一种无监督学习下的机器学习模型。

深度学习的概念由Hinton等人于2006年提出。基于深度置信网络提出非监督贪心逐层训练算法,为解决深层结构相关的优化难题带来希望,随后提出多层自动编码器深层结构。此外LeCun等人提出的卷积神经网络是第一个真正多层结构学习算法,它利用空间相对关系减少参数数目以提高训练性能。深度学习是机器学习研究中的一个新的领域,其动机在于建立、模拟人脑进行分析学习的神经网络,模仿人脑的机制来解释数据,例如图像、声音和文本。

5. 个性化推荐系统

个性化推荐系统(简称推荐系统)是互联网和电子商务发展的产物,它是建立在海量数据挖掘基础上的一种高级商务智能平台,向顾客提供个性化的信息服务和决策支持。近年来已经出现了许多非常成功的大型推荐系统实例,与此同时,个性化推荐系统也逐渐成为学

术界的研究热点之一。

推荐系统是利用电子商务网站向客户提供商品信息和建议,帮助用户决定应该购买什么产品,模拟销售人员帮助客户完成购买过程。个性化推荐是根据用户的兴趣特点和购买行为,向用户推荐用户感兴趣的信息和商品。

随着电子商务规模的不断扩大,商品个数和种类快速增长,顾客需要花费大量的时间才能找到自己想买的商品。这种浏览大量无关的信息和产品过程无疑会使淹没在信息过载问题中的消费者不断流失。

为了解决这些问题,个性化推荐系统应运而生。个性化推荐系统是建立在海量数据挖掘基础上的一种高级商务智能平台,以帮助电子商务网站为其顾客购物提供完全个性化的决策支持和信息服务。

互联网的出现和普及给用户带来了大量的信息,满足了用户在信息时代对信息的需求,但随着网络的迅速发展而带来的网上信息量的大幅增长,使得用户在面对大量信息时无法从中获得对自己真正有用的那部分信息,对信息的使用效率反而降低了,这就是所谓的信息超载(Information Overload)问题。

解决信息超载问题的一个非常有潜力的办法是推荐系统,它是根据用户的信息需求、兴趣等,将用户感兴趣的信息、产品等推荐给用户的个性化信息推荐系统。和搜索引擎相比,推荐系统通过研究用户的兴趣偏好,进行个性化计算,由系统发现用户的兴趣点,从而引导用户发现自己的信息需求。一个好的推荐系统不仅能为用户提供个性化的服务,还能和用户之间建立密切关系,让用户对推荐产生依赖。

推荐系统现已广泛应用于很多领域,其中最典型并具有良好的发展和应用前景的领域就是电子商务领域。同时学术界对推荐系统的研究热度一直很高,逐步形成了一门独立的学科。

推荐系统有三个重要的模块:用户建模模块、推荐对象建模模块、推荐算法模块。推荐系统把用户模型中兴趣需求信息和推荐对象模型中的特征信息匹配,同时使用相应的推荐算法进行计算筛选,找到用户可能感兴趣的推荐对象,然后推荐给用户。

随着推荐技术的研究和发展,其应用领域也越来越多。例如,新闻推荐、商务推荐、娱乐推荐、学习推荐、生活推荐、决策支持等。推荐方法的创新性、实用性、实时性、简单性也越来越强。例如,上下文感知推荐、移动应用推荐、从服务推荐到应用推荐。

3.6.3 人工智能应用

1. 智慧地球

智慧地球就是把感应器嵌入和装备到电网、铁路、桥梁、隧道、公路、建筑、供水系统、大坝、油气管道等各种物体中,并且被普遍连接,形成所谓的物联网,然后将物联网与现有的互联网整合起来,实现人类社会与物理系统的整合。这一概念由 IBM 公司首席执行官彭明盛首次提出。

2. 智慧城市

智慧城市是指充分借助物联网、传感网,涉及智能楼宇、智能家居、路网监控、智能医院、

城市生命线管理、食品药品管理、票证管理、家庭护理、个人健康与数字生活等诸多领域。应把握新一轮科技创新革命和信息产业浪潮的重大机遇,充分发挥信息通信产业发达、RFID 相关技术领先、电信业务及信息化基础设施优良等优势,通过建设信息通信基础设施、认证、安全等平台和示范工程,加快产业关键技术攻关,构建城市发展的智慧环境,形成基于海量信息和智能过滤处理的新的生活、产业发展、社会管理等模式,面向未来构建全新的城市形态。

3. 智能交通系统

智能交通是一个基于现代电子信息技术面向交通运输的服务系统。智能交通系统(ITS)是指将先进的信息技术、数据通信传输技术、电子控制技术、计算机处理技术等应用于交通运输行业从而形成的一种信息化、智能化、社会化的新型运输系统,它使交通基础设施能发挥最大效能。智能交通系统是 21 世纪交通事业发展的必然选择。

1) 智能车辆

智能车辆是一个集环境感知、规划决策、多等级辅助驾驶等功能于一体的综合系统,它集中运用了计算机、现代传感、信息融合、通信、人工智能及自动控制等技术,是典型的高新技术综合体。智能汽车与一般所说的自动驾驶有所不同,它指的是利用多种传感器和智能公路技术实现的汽车自动驾驶。智能汽车不需要人去驾驶,人只需要舒服地坐在车上享受这高科技的成果就行了,因为这种汽车上装有相当于汽车的“眼睛”“大脑”“脚”的电视摄像机、电子计算机和自动操纵系统之类的装置。汽车能和人一样会“思考”“判断”“行走”,可以自动启动、加速、刹车,可以自动绕过地面障碍物。在复杂多变的情况下,它的“大脑”能随机应变,自动选择最佳方案,指挥汽车正常、顺利地行驶。

2) 无人驾驶

近年来,无人驾驶技术飞速发展,以无人机、无人舰艇等为代表。

无人机是自动控制、具有自动导航和执行特殊任务的无人飞行器。它具有全天候、大纵深、长时间作战、快速侦察的能力。

无人舰艇是一种无人操作的舰艇。主要有无人水面舰艇和无人潜航器两种。主要用于执行危险及不适于有人船只执行的任务。无人舰艇有望在未来 10 年内彻底变革海军的军事行动和战争。

4. 智能电网

智能电网的核心在于构建具备智能判断与自适应调节能力的多种能源统一入网和分布式管理的智能化网络系统,可对电网与客户用电信息进行实时监控和采集,且采用最经济最安全的输配电方式将电能输送给终端用户,实现对电能的最优配置与利用,提高电网运行的可靠性和能源利用效率。智能电网的本质是能源替代和兼容利用,它需要在开放的系统和共享信息模式的基础上,整合系统中的数据,优化电网的运行和管理。

2009 年 5 月,国家电网公司首次公布了智能电网计划,全面建设以特高压电网为骨干网架、各级电网协调发展的,以坚强电网为基础,信息化、自动化、互动化为特征的自主创新、国际领先的坚强智能电网。国家电网公司还提出在物联网发展的三个阶段:信息汇聚阶段、协同感知阶段、泛在聚合阶段。

国家电网公司提出了“面向应用、立足创新、形成标准、建立示范”的研究指导思想,在物

联网的专用芯片、标准体系、信息安全、软件平台、测试技术、实验技术、应用系统开发、无线宽带通信等方面进行了全面部署,力争在未来三年内实现物联网技术应用在电力系统的多项核心技术突破。形成若干项有重大影响的创新性科研成果,成为在国内外有重要影响的从事智能电网物联网技术研究和应用的研发中心和产业化基地。

2011年3月2日国家电网宣布,2011年智能电网将进入全面建设阶段,并在示范工程、新能源接纳、居民智能用电等方面大力推进。智能电网是一个涵盖广泛的工程,信息网络传输能力只是其中之一,如果智能电网全面建成,它将对现有信息网络具有完全的可替代性,而且能力甚至更为强大。

5. 智能楼宇

智能楼宇也称智能建筑,其核心是5A系统,即建筑设备自动化系统(BA)、通信自动化系统(CA)、办公自动化系统(OA)、火灾报警与消防联动自动化系统(FA)和安全防范自动化系统(SA)。通过综合布线将五个系统进行有机综合,使建筑物具有了安全、便利、高效、节能的特点。

智能楼宇最早是对讲系统,源于欧美等发达国家,20世纪80年代末进入我国,20世纪90年代初期,国外楼宇对讲系统生产制造商陆续进入中国市场,20世纪90年代末楼宇对讲产品进入第二个高速发展期,大型社区联网及综合性智能楼宇对讲设备开始涌现。目前,一个比较完善的智能楼宇至少包括视频监控系统、安防报警系统、楼宇对讲系统、门禁一卡通系统、火灾报警系统、公共广播系统、多媒体会议系统、有线电视和卫星电视系统、多媒体信息发布系统、机房系统、楼宇BA系统及IBS系统集成11部分。

6. 智能家居

智能家居又称智能住宅,它是融合了自动化控制系统、计算机网络系统和网络通信技术于一体的网络化智能化的家居控制系统。智能家居将让用户有更方便的手段来管理家庭设备,如,通过家触摸屏、无线遥控器、电话、互联网或者语音识别控制家用设备,更可以执行场景操作,使多个设备形成联动;另外,智能家居内的各种设备相互间可以通信,不需要用户指挥也能根据不同的状态互动运行,从而给用户带来最大程度的高效、便利、舒适与安全。

智能家居集成是利用综合布线技术、网络通信技术、安全防范技术、自动控制技术、音视频技术将家居生活有关的设备集成。网络通信技术是智能家居集成中关键的技术之一。安全防范技术是智能家居系统中必不可少的技术,在小区及户内可视对讲、家庭监控、家庭防盗报警、与家庭有关的小区一卡通等领域都有广泛应用。自动控制技术是智能家居系统的核心技术,广泛应用在智能家居控制中心、家居设备自动控制模块中,对于家庭能源的科学管理、家庭设备的日程管理都有十分重要的作用。音视频技术是实现家庭环境舒适性、艺术性的重要技术,体现在音视频集中分配、背景音乐、家庭影院等方面。

智能家居系统包含的主要子系统有家居布线系统、家庭网络系统、智能家居(中央)控制管理系统、家居照明控制系统、家庭安防系统、背景音乐系统、家庭影院与多媒体系统、家庭环境控制系统八个子系统。

习题

一、选择题

1. C/S 是一种重要的网络计算机模式,其含义是()。
A. 客户-服务器模式
B. 文件-服务器模式
C. 分时-共享模式
D. 浏览器-服务器模式
2. 如果在一个数据模型中有一个节点,它有多于一个的父节点,那么这个模型是()。
A. 网状模型
B. 树状模型
C. 层次模型
D. 关系模型
3. TCP/IP 是指()。
A. 传输控制协议
B. 传输控制协议和网间协议
C. 网间协议
D. 系统网络结构和数字网络体系结构

二、填空题

1. 计算机数据管理分为三个阶段,即_____、_____和_____。
2. 计算机网络系统是由_____和_____组成的。

三、名称解释

1. 计算机网络
2. 数据库
3. 多媒体
4. 信息安全

四、简答题

1. 数据库管理阶段在管理数据方面有哪些特点?
2. 什么是网络协议?它的作用是什么?