

第5章

信息资源管理技术及其资源化

5.1 信息资源采集技术

5.1.1 文本生成

1. 键盘输入

人工键盘输入是指用手工击键方式按照一定的规则把汉字输入到计算机,目前已有数百种输入法。

2. 语音识别

目前主流的语音识别技术是基于统计模式识别的基本理论。语音识别过程实际上是一种认识过程。就像人们听语音时,并不把语音和语言的语法结构、语义结构分开,因为当语音发音模糊时,人们可以用这些知识来指导对语言的理解过程。对机器来说,识别系统也要利用这些方面的知识。

3. 手写输入

手写输入方法,就是把要输入的汉字写在一块书写板的设备(如手机屏)上。这种设备将笔尖、手指走过的轨迹按时间采样后发送到计算机中,由计算机软件自动完成识别,并用机器内部的方式保存、显示。

4. OCR

OCR是Optical Character Recognition的简称,指光学字符识别。OCR技术在个人信息管理、办公自动化、电子出版物、网络资源以及各种大型文献资料管理数据库、数字化图书馆等领域应用广泛。

5.1.2 图像、音频和视频采集

1. 图像获取

计算机中的图像是由特殊的数字化设备,将光信号量化为数值,并按一定的格式组织而得到的。这些数字化设备常用的有扫描仪、图像采集卡、数码相机等。扫描仪对已有的照片、图片等进行扫描,将图像数字化为一组数据存储。图像采集卡可以对录像带、电视上的

信号进行“抓图(capture)”,对其中选定的帧进行捕获并数字化。数码相机是一种与计算机配套使用的、新型的数码影像设备,它将被景物以数字信号方式直接记录在存储介质(存储器、存储卡)中,可以很方便地在计算机中进行处理。

2. 音频采集

音频是一种典型的连续时间信号。话筒把声音的机械振动转换为电信号,模拟音频技术中以模拟电压的幅度表示声音的强弱。这种模拟信号是一个在时间轴上的连续平滑的波形,对这样一个在时间上连续的信号,计算机每隔固定的时间对波形的幅值进行采样,用得到的一系列数字化量来表示声音。在某一个特定的时刻对音频信号的测量叫作采样。计算机必须有相应的输入输出设备才能进行声音信号的处理。波形声音的获取是通过声音数字化接口进行的,输入的声音经过数字化后进入计算机中。输出的过程正好与输入相反,声音数字流经解码、逆压扩变换后,通过数模转换电路把离散的数字序列转换为模拟电压波形送到扬声器中播放。

3. 视频采集

视频采集是将模拟摄像机、录像机、LD视盘机、电视机输出的视频信号等输出的视频数据或者视频音频的混合数据输入计算机,并转换成计算机可辨别的数字数据,存储在计算机中,使其成为可编辑处理的视频数据文件。视频采集卡,又称视频捕捉卡,英文名为 video capture card,其功能是将视频信号采集到计算机中,以数据文件的形式保存在硬盘上。它是进行视频处理必不可少的硬件设备,通过它,就可以把摄像机拍摄的视频信号从摄像带转存到计算机中,利用相关的视频编辑软件,对数字化的视频信号进行后期编辑处理,如剪切画面,添加滤镜、字幕和音效,设置转场效果以及加入各种视频特效等,最后将编辑完成的视频信号转换成标准的 VCD、DVD 以及网上流媒体等格式,方便传播和保存。

5.1.3 自动识别技术

自动识别技术是一种高度自动化的信息或者数据采集技术。已经形成了一个包括条码技术、磁条磁(卡)技术、IC卡技术、光学字符识别、射频技术、声音识别及视觉识别等集计算机、光、磁、物理、机电、通信技术为一体的技术。

1. 自动识别技术的种类

自动识别系统根据识别对象的特征可以分为两大类,分别是数据采集技术和特征提取技术。这两大类自动识别技术的基本功能都是完成物品的自动识别和数据的自动采集。数据采集技术的基本特征是需要被识别物体具有特定的识别特征载体(如标签等,仅光学字符识别例外),而特征提取技术则根据被识别物体的本身的行为特征(包括静态、动态和属性特征)来完成数据的自动采集。

1) 数据采集技术

光存储器包括条码(一维、二维)、矩阵码、光标阅读器、光学字符识别(OCR)。

磁存储器包括磁条、非接触磁卡、磁光存储、微波。

电存储器包括触摸式存储、RFID射频识别(无芯片、有芯片)、存储卡(智能卡、非接触

式智能卡)、视觉识别、能量扰动识别。

2) 特征提取技术

动态特征包括声音(语音)、键盘敲击、其他感觉特征。

属性特征包括化学感觉特征、物理感觉特征、生物抗体病毒特征、联合感觉系统。

2. 条码技术

条码由一组规则排列的条、空以及相应的数字组成。这种用条、空以及相应的数字组成的数据编码可以供条码阅读器识读,而且很容易转换成二进制数和十进制数。这些条、空以及相应的数字可以有各种不同的组合方法,构成不同的图形符号,即各种符号体系(也称码制),适用于不同的场合。目前使用频率最高的几种码制是 EAN、UPC、39 码、交叉 25 码和 EAN128 码,其中 UPC 条码主要用于北美地区,EAN 条码是国际通用符号体系,EAN128 条码是由国际物品编码协会(EAN International)和美国统一代码委员会(UCC)联合开发、共同采用的一种特定的条码符号。上述条码都是一维条码。为了提高一定面积上的条码信息密度和信息量又发展了一种新的条码编码形式——二维条码。从结构上讲,二维条码分为两类:一类由矩阵代码和点代码组成,其数据是以二维空间的形态编码的;另一类是包含重叠的或多行条码符号,其数据以成串的数据行显示。重叠的符号标记法有 CODE 49、CODE 16K 和 PDF417。

3. 磁条(卡)技术

磁条技术应用了物理学和磁力学的基本原理。对自动识别设备制造商来说,磁条就是一层薄薄的由定向排列的铁性氧化粒子组成的材料(也称为涂料),用树脂黏合在一起并粘在诸如纸或者塑料这样的非磁性基片上。

磁条技术的优点是数据可读写,即具有现场改写数据的能力;数据存储量能满足大多数需求,便于使用,成本低廉,还具有一定的数据安全性;能黏附于许多不同规格和形式的基材上。这些优点,使之在很多领域得到了广泛应用,如信用卡、银行 ATM 卡、机票、公共汽车票、自动售货卡、会员卡、现金卡(如电话磁卡)、地铁 AFC 等。

4. IC 识别技术

IC(Integrated Card)是 1970 年由法国人 Roland Moreno 发明的。其优点是安全,存储容量大,便于应用,方便保管;防磁、防一定强度的静电,抗干扰能力强,可靠性比磁卡高,使用寿命长,一般可重复读写 10 万次以上。由于它的触点暴露在外面,有可能因人为的原因或静电而遭到损坏。在我们的生活中,IC 的应用比较广,如电话 IC、购电(气)卡、手机 SIM 卡、交通卡以及智能水表、智能气表等。

5. 声音识别技术

声音识别技术是对基于生理学和行为特征的说话者嗓音和语言学模式的运用。声音识别技术的迅速发展以及高效可靠的应用软件的开发,使声音识别系统在很多方面得到了应用,如汉字的语音输入系统就是典型的聲音识别技术。

6. 视觉识别技术

视觉识别系统可以看作是这样的系统:它能获取视觉图像,而且通过一个特征抽取和分析的过程,能自动识别限定的标志、字符、编码结构或可作为确切识别的基础呈现在图像

内的其他特征。

7. 射频识别技术

射频技术(RFID)是利用无线电波来进行通信的一种自动识别技术。其基本原理是通过读头和黏附在物体上的标签之间的电磁耦合或者电感耦合来进行数据通信以达到对标签物品的自动识别。射频系统的优点是不局限于视线,识别距离比光学系统远。射频识别卡具有读写能力,可携带大量数据,难以伪造,智能性较高等。射频识别和条码一样是非接触式识别技术,由于无线电波能“扫描”数据,射频标签最大的优点就在于非接触,因此完成识别工作时无须人工干预,适用于实现自动化且不易损坏,可识别高速运动物体并可同时识别多个射频标签,操作快捷方便。长距离的产品多用于交通上,可达几十米,如自动收费或识别车辆身份。

5.1.4 数据采集系统

1. 数据采集的概念

数据采集指从传感器和其他待测设备等模拟和数字被测单元中自动采集非电量或者电量信号,送到上位机中进行分析 and 处理。数据采集的目的是测量电压、电流、温度、压力或声音等物理现象。基于计算机的数据采集,通过模块化硬件、应用软件和计算机的结合进行测量。

被采集数据是已被转换为电信号的各种物理量,如温度、水位、风速、压力等,可以是模拟量,也可以是数字量。采集一般用采样方式,即每隔一定时间(称采样周期)对同一数据重复采集。采集的数据大多是瞬时值,也可以是某段时间内的一个特征值。准确的数据量测是数据采集的基础。数据量测方法有接触式和非接触式,检测元件多种多样。无论用哪种方法和元件,均以不影响被测对象状态、测量环境和保证数据的正确性为前提。数据采集含义很广,包括对面状连续物理量的采集。在计算机辅助制图、测图、设计中,对图形或图像的数字化的过程也可称为数据采集,此时被采集的是几何量数据。

数据采集是计算机与外部物理世界连接的桥梁。利用串行通信方式,实现对移动数据采集器的应用软件升级,通过制定上位机与移动数据采集器的通信协议,实现两者之间阻塞式通信交互过程。

2. 数据采集系统组成

数据采集系统由硬件和软件两部分组成。从硬件来看,数据采集系统的结构形式主要有两种:第一种是微型计算机数据采集系统,由传感器、模拟多路开关、程控放大器、采样持器、A/D转换器、计算机及外设等部分组成;第二种是集散型数据采集系统,由若干个数据采集站和一台上位机及通信线路组成。数据采集站由单片机数据采集装置组成,位于生产设备附近,可独立完成数据采集和预处理任务,还可将数据以数字信号的形式传送给上位机。上位机用来将各个数据采集站传送来的数据集中显示在显示器上或用打印机打印成各种报表,或以文件形式存储在磁盘上。

数据采集系统整合了信号、传感器、激励器、信号调理、数据采集设备和应用软件。数据采集系统包括了可视化的报表定义、审核关系的定义、报表的审批和发布、数据填报、数据预处理、数据评审、综合查询统计等功能模块。通过信息采集网络化和数字化,扩大了数据采

集的覆盖范围,提高审核工作的全面性、及时性和准确性,最终实现相关业务工作管理现代化、程序规范化、决策科学化和服务网络化。

5.2 信息资源存储技术

5.2.1 存储器及其种类

1. 存储器概述

存储器是计算机系统记忆设备,用来存放程序和数据。计算机中的全部信息,包括输入的原始数据、计算机程序、中间运行结果和最终运行结果,都保存在存储器中。它根据控制器指定的位置存入和取出信息。按用途划分,存储器可分为主存储器(内存)和辅助存储器(外存);也有外部存储器和内部存储器的分类方法。外存通常是磁性介质或光盘等,能长期保存信息。内存指主板上的存储部件,用来存放当前正在执行的数据和程序,但仅用于暂时存放程序和数据,关闭电源或断电后,数据会丢失。

2. 存储器的构成

构成存储器的存储介质主要采用半导体器件和磁性材料。存储器中最小的存储单位就是一个双稳态半导体电路、一个 CMOS 晶体管或磁性材料的存储元,它可存储一个二进制代码。由若干个存储元组成一个存储单元,然后由许多存储单元组成一个存储器。一个存储器包含许多存储单元,每个存储单元可存放 1 字节(按字节编址)。每个存储单元的位置都有一个编号,即地址,一般用十六进制表示。一个存储器中所有存储单元可存放数据的总和,称为存储容量。假设一个存储器的地址码由 20 位二进制数(即 5 位十六进制数)组成,则可表示为 2^{20} ,即 1MB 个存储单元地址。每个存储单元存放 1 字节,则该存储器的存储容量为 1MB。

存储器的主要功能是存储程序和各种数据,并能在计算机运行过程中高速、自动地完成程序或数据的存取。由于在计算机中采用只有两个数字“0”和“1”的二进制来表示数据,记忆元件的两种稳定状态分别表示为“0”和“1”。因此,日常使用的十进制数必须转换成二进制数才能存入存储器中;计算机中处理的各种字符,例如英文字母、运算符号等,也要转换成二进制代码才能存储和操作。

3. 存储器的用途

根据在计算机系统中所起的作用,存储器可分为主存储器、辅助存储器、高速缓冲存储器、控制存储器等。为了解决对存储器要求容量大、速度快、成本低三者之间的矛盾,通常采用多级存储器体系结构,即使用高速缓冲存储器、主存储器和外存储器。

(1) 高速缓冲存储器:用于高速存取指令和数据,存取速度快,但存储容量小。

(2) 主存储器:用于内存存放计算机运行期间的大量程序和数据,存取速度较快,存储容量不大。

(3) 外存储器:用于外存存放系统程序和大型数据文件及数据库,存储容量大,成本低。

4. 常用存储器

(1) 硬盘是计算机主要的存储设备,由一个或多个铝制(或者玻璃制)的碟片组成。这些碟片外覆盖有铁磁性材料。绝大多数硬盘都是固定硬盘,被永久性地密封并固定在硬盘驱动器中。物理结构:①磁头,是读写合一的电磁感应式磁头;②磁道,当磁盘旋转时,磁头若保持在一个位置上,则每个磁头都会在磁盘表面划出一个圆形轨迹,这些圆形轨迹就称为磁道;③扇区,磁盘上的每个磁道被等分为若干个弧段,这些弧段便是磁盘的扇区,每个扇区可以存放 512 字节的信息,磁盘驱动器在从磁盘读取或向磁盘写入数据时,要以扇区为单位;④柱面,硬盘通常由重叠的一组盘片构成,每个盘面都被划分为数目相等的磁道,并从外缘的“0”开始编号,具有相同编号的磁道形成一个圆柱,称为磁盘的柱面。

(2) 光盘以光信息作为存储载体,用来存储数据,采用聚焦的氢离子激光束处理记录介质的方法存储和再生信息。激光光盘分为不可擦写光盘和可擦写光盘。高密度光盘是近代发展起来的不同于磁性载体的光学存储介质。常见光盘非常薄,只有 1.2mm 厚,分为 5 层,包括基板、记录层、反射层、保护层和印刷层等。

(3) U 盘的全称“USB 闪存盘”,英文名为 USB Flash Disk。它是一种拥有 USB 接口的、无须物理驱动器的微型大容量移动存储产品,可以通过 USB 接口与计算机连接,实现即插即用。U 盘的优点是体积小、便于携带、存储容量大、性能可靠、价格便宜。

5.2.2 数据库与数据仓库

1. 数据库的定义

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。它是一个按数据结构来存储和管理数据的计算机软件系统。数据库的概念实际包括两层意思:①数据库是一个实体,它是能够合理保管数据的仓库,用户在该仓库中存放要管理的事务数据,数据和库两个概念结合成为数据库。②数据库是数据管理的新方法和技术,它能更合适地组织数据、更方便地维护数据、更严密地控制数据和更有效地利用数据。

2. 技术发展

数据库发展大致划分为人工管理阶段、文件系统阶段、数据库系统阶段、高级数据库阶段。

(1) 人工管理阶段。20 世纪 50 年代之前,计算机主要用于科学计算,程序员在程序中不仅要规定数据结构,还要设计物理结构,包括存储结构、存取方法、输入输出方式等。

(2) 文件系统阶段。操作系统的出现使文件管理成为可能。这时,数据管理步入文件系统阶段。数据以文件为单位存储在外存,且由操作系统统一管理。

(3) 数据库系统阶段。20 世纪 60 年代后,随着计算机在数据管理领域的普遍应用,人们对数据管理技术提出了更高的要求:希望面向企业或部门,以数据为中心组织数据,减少数据的冗余,提供更高的数据共享能力,同时要求程序和数据具有较高的独立性,当数据的逻辑结构改变时,不涉及数据的物理结构,也不影响应用程序,以降低应用程序研制与维护的费用。数据库技术正是在这样一个应用需求的基础上发展起来的。

(4) 高级数据库阶段。随着信息管理内容的不断扩展,出现了丰富多样的数据模型(层次模型、网状模型、关系模型、面向对象模型、半结构化模型等),新技术也层出不穷(数据流、

Web 数据管理、数据挖掘等)。

3. 数据结构模型

数据模型有三种,即层次结构模型、网状结构模型和关系结构模型。

(1) 层次结构模型。按照层次模型建立的数据库系统称为层次模型数据库系统。

(2) 网状结构模型。按照网状数据结构建立的数据库系统称为网状数据库系统。

(3) 关系结构模型。由关系数据结构组成的数据库系统称为关系数据库系统。在关系数据库中,对数据的操作几乎全部建立在一个或多个关系表格上,通过对这些关系表格的分类、合并、连接或选取等运算来实现数据的管理。

4. 联邦数据库

联邦数据库系统(FDBS)是一个彼此协作又相互独立的单元数据库(CDBS)集合,它将单元数据库系统按不同程度进行集成,对该系统整体提供控制和协同操作的软件叫作联邦数据库管理系统(FDBMS)。一个单元数据库可以加入若干个联邦系统,每个单元数据库系统的数据可以是集中式的,也可以是分布式的。联邦数据库技术可以实现对相互独立运行的多个数据库的互操作。相互独立运行的数据库系统为单元数据库系统。

5. 数据仓库

数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合,用于支持管理决策。面向主题指用户使用数据仓库进行决策时所关心的重点方面,如收入、客户、销售渠道等;所谓面向主题,是指数据仓库内的信息是按主题进行组织的,而不是像业务支撑系统那样是按照业务功能进行组织的。这里的集成指数据仓库中的信息不是从各个业务系统中简单抽取出来的,而是经过一系列加工、整理和汇总的过程,因此数据仓库中的信息是关于整个企业的一致全局信息。这里的随时间变化指数据仓库内的信息并不只是反映企业当前的状态,而是记录了从过去某一时点到当前各个阶段的信息。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。

数据库与数据仓库属于比较容易混淆的两个概念。其本质上存在较大差异,主要表现在以下几方面:用途、数据类型、数据结构、访问模式、数据来源和应用场景等。

5.2.3 数据中心智能化

1. 数据中心

数据中心是基于 Internet 网络的基础设施,包括服务器、网络设备、存储设备以及管理软件等,可用于数据存储、处理、服务和管理。数据中心是与人力资源、自然资源一样重要的战略资源。

数据中心已成为现代社会基础设施的一部分,进而对其产生影响。数据中心与各行各业融合,并在其中发挥作用,促使社会快速变革。

海量数据的产生促使数据收集与处理方式发生了重大转变,无论是企事业单位还是政府机构,已从实体服务走向了数据服务。计算机行业从追求计算能力转变为重视数据处理能力,软件业也从编程为主转向以数据为主。数据中心存储的数据越多,其价值也就越高,

这也驱使很多行业 and 部门开始尝试由下向上汇聚数据。

2. 数据中心智能化

数据中心智能化,就是数据中心的智能升级。数据中心智能化指通过人工智能技术的应用提高数据中心的运营效率,降低运营风险和成本,使整体管理水平和持续发展能力有比较大的提升。数字化后的数据中心即为数智中心。

信息管理机构经历了计算站(电算站)、信息中心和数据中心三个阶段,并将逐步进入数智中心阶段。数智中心是数据中心与数字技术(特别是人工智能技术)深度融合的产物。功能上,数智中心不仅具有传统计算站辅助用户利用科学计算的能力,也具有信息中心为用户利用网络提供信息资源共享的能力,还具有数据中心协助用户进行大数据获取和分析的能力,除此之外,数智中心还应该具有较强的智慧特征和能力性,并能利用人工智能的技术手段辅助用户获取、处理、加工、应用其所需的大数据信息资源。以上这4个概念的差异如表5-1所示。

表 5-1 概念差异比较

类 型	功 能 特 点
计算站	提供计算机硬件和软件进行科学计算服务
信息中心	利用服务器和网络进行信息资源服务
数据中心	利用云和大数据平台进行数据资源服务
数智中心	利用智能技术提供大数据资源服务

数据中心由关键主设备和基础支撑设备两部分组成。关键主设备包括服务器、存储设备、路由器、交换机、防火墙和负载均衡器,以及数据中心提供的各种应用程序。基础支撑设备包括供配电系统、不间断电源系统、终端配电系统、电源辅助系统和空调系统等。除此之外,数据中心还包括数据中心的数据技术管理和数据中心的运维。

根据以上分类,我们认为,数据中心智能化工作可分为以下4个分项。

第一,数据中心关键主设备智能化分项工作。关键主设备智能化可分解为服务器、存储设备、路由器、交换机、防火墙、负载均衡器、数据中心提供的各种应用程序7个方面的智能化。

第二,数据中心基础支撑设备智能化分项工作。基础支撑设备智能化可分解为供配电系统、不间断电源系统、终端配电系统、电源辅助系统、空调系统、安防设备和消防设备7个方面的智能化。

第三,数据中心数据智能化分项工作。数据智能化可分解为自动数据采集、智能数据存储(具体分为存储的自动化、存储扩容的自动化、存储收回的自动化)、智能数据治理、智能数据分析、智能数据汇聚与交换、智能数据库、知识库与管理系统、智能数据应用8个方面的智能化。

第四,数据中心运维智能化分项工作。数据中心运维智能化可分解为数据中心自动运维、网络安全自动检测、用户智能统一认证、人工智能运维系统部署、计算资源服务的自动化5个方面的智能化。

5.3 信息资源传输技术

5.3.1 计算机网络及其未来

1. 计算机网络概述

1) 定义

计算机网络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。

计算机网络的简单定义是一些相互连接的以共享资源为目的的自治的计算机的集合。广义上,计算机网络是以传输信息为基础目的,用通信线路将多个计算机连接起来的计算机系统的集合。从用户角度看,计算机网络是可以调用用户所需资源的系统。

2) 功能

计算机网络的主要功能是硬件资源共享、软件资源共享和用户间信息交换。

(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。计算机网络为分布在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。

3) 网络硬件

(1) 服务器。提供网络资源服务的设备,一般为高性能计算机。

(2) 终端。可以是工作站、微机、笔记本、平板电脑、手机等固定或移动设备。

(3) 连网部件。包括网卡、适配器、调制解调器、连接器、收发器、终端匹配器、FAX卡、中继器、集线器、网桥、路由器、桥由器、网关、集线器、交换机等。

(4) 通信介质。包括双绞线、同轴电缆和光纤等有线介质,以及短波、卫星等无线介质。

(5) 中继器。工作在物理层上的连接设备。适用于完全相同的两类网络的互连,主要功能是通过数据信号的重新发送或者转发,来扩大网络传输的距离。

(6) 集线器。其主要功能是对接收到的信号进行再生整形放大,以扩大网络的传输距离,同时把所有节点集中在以它为中心的节点上。

(7) 网桥。一个局域网与另一个局域网之间建立连接的桥梁。属于网络层设备。

(8) 交换器。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。

(9) 路由器。工作在网络层,可以在多个网络上交换数据和路由数据包。

(10) 网关。当连接不同类型而协议差别又较大的网络时,则要选用网关设备。网关工作在应用层。

2. 互联网和第二代互联网

20世纪60年代美国开始互联网的研究,到80年代中后期建成第一代互联网。由互联

网兴起的新经济,引起了世界经济的飞速发展。1996年10月,美国政府宣布启动“下一代互联网 NGI”研究计划,其核心是互联网协议和路由器。它的主要目标是:建设高性能的边缘网络,为科研提供基础设施;开发具有革命性的 Internet 应用技术;促进新的网络服务及应用在 Internet 上的推广。英、德、法、日、加等发达国家及中国均参与这项计划。Internet2 项目使单独的计算机工作站接入带宽达 10Gb/s,Internet2 骨干网的连接速度达 100Gb/s。第一代互联网的 IP 地址协议是 IPv4,其网络地址编码是 32 位,大约 43 亿个,2019 年 12 月已经全部用完。第二代互联网采用 IPv6,地址编码为 128 位,能产生 2^{128} 个 IP 地址,其地址可以满足人类未来的发展。

3. 未来网络

1) 提出背景

目前互联网已取得巨大成功,且技术还在不断演进,但其正面临着新的挑战。主要有以下三个问题。

第一,由于传统网络架构不够灵活,不能适应不断涌现的新业务需求,网络服务质量难以保证,产业价值链难以为继。

第二,互联网流量飞速增长,信息冗余传输严重,网络难以适应未来信息海量增长的需求。当前,以移动互联网为代表的全球 IP 流量高速增长,给电信运营商网络基础设施带来巨大压力及挑战,使运营商不得不持续增加投入进行频繁的网络扩容。

第三,互联网与实体经济深度融合是发展趋势,而且还有巨大发展空间,但现阶段互联网在实时性、安全性、灵活性等方面仍旧满足不了应用需求。

2) 未来网络定义

未来网络(Future Network)又称为新一代网络。业内暂时没有未来网络的定义。刘韵洁院士认为,未来网络就是更快捷、更简单、更便宜、更安全的新一代互联网,以用户为中心,让上网的人感觉更好。邬贺铨院士认为,未来网络是面向普通用户的互联网,应具有可扩展性、可管理性、更安全、更宽的移动带宽;也应该是面向各类企业的互联网,具有大带宽、大连接、高可靠、低时延的特点。智能化是未来网络的一大特点。具有进化性,从现在这种冰冷的物理连接到主动感知各类连接设备,再到分析了解网络运行的各种数据,知晓如何适应外界的变化,最后网络将有智能思考的能力,形成自身的知识,会策略地选择和推演,实现人、网络 and 应用的融合体。未来网络是集连接、感知、计算和数据服务于一体的网络,实现超级安全、自主优化、进化更新等三大能力。从 2010 年开始,世界已进入互联网 3.0 时代,这个时代是融合的时代,在该阶段,工业互联网、能源网、车联网等新的需求对互联网架构提出新的挑战。

3) 未来网络技术发展过程

从 2000 年到 2015 年美国国家自然科学基金会相继资助的未来网络项目包括 NewArch 项目,未来互联网网络设计 FIND 计划,未来互联网架构 FIA 计划,未来互联网架构——下一阶段 FIA-NP 计划。2007 年欧盟在第七研发框架 FP7 中设立未来互联网研究和试验 FIRE 项目。另外,德国、澳大利亚、日本和韩国也展开了这方面的研究。

在中国,2009 年未来网络课题被提出,2013 年开始未来网络小规模试验,2016 年未来网络试验设施 CENI 项目获发改委批准在全国 40 个城市实验,2021 年国家重大科技基础设施 CENI 开放合作启动。

5.3.2 万兆以太网与高速无线网

1. 万兆以太网

以太网采用 CSMA/CD 机制,即带碰撞检测的载波监听多重访问。千兆以太网接口基本应用在点到点线路,不再共享带宽。碰撞检测、载波监听和多重访问已不再重要。千兆以太网与传统低速以太网最大的相似之处在于采用相同的以太网帧结构。万兆以太网技术与千兆以太网类似,仍然保留了以太网帧结构。通过不同的编码方式或波分复用提供 10Gb/s 传输速度。所以就其本质而言,10G 以太网仍是以太网的一种类型。

10G 以太网于 2002 年 7 月在 IEEE 通过。10G 以太网包括 10GBASE-X、10GBASE-R、10GBASE-W 以及基于铜缆的 10GBASE-T 等(2006 年通过)。10GBASE-X 使用一种特紧凑包装,含有一个较简单的密集波分复用器件、4 个接收器和 4 个在 1300nm 波长附近以大约 25nm 为间隔工作的激光器,每一对发送器/接收器在 3.125Gb/s 速度(数据流速度为 2.5Gb/s)下工作。10GBASE-R 是一种使用 64B/66B 编码(不是在千兆以太网中所用的 8B/10B)的串行接口,数据流为 10Gb/s,因而产生的时钟速率为 10.3Gb/s。10GBASE-W 是广域网接口,与 SONET OC-192 兼容,其时钟为 9.953Gb/s,数据流为 9.585Gb/s。

万兆以太网技术已经成熟,适用领域十分广阔。各种迅速增长的带宽密集型项目,像高带宽园区骨干、数据中心汇聚、集群和网格计算、合一(语音、视频、图像和数据)的通信、存储组网、金融交易以及政府、医疗卫生领域、研究单位和大学的超级计算研究等,都离不开万兆以太网技术。

2. 高速无线网

无线网既包括允许用户建立远距离无线连接的全球语音和数据网络,也包括为近距离无线连接进行优化的红外线技术及射频技术。

1) 通信载体分类

(1) 地面微波。地面微波系统主要用于长途电信服务,可代替同轴电缆和光纤,通过地面接力站中继,还可用于建筑物之间的点对点线路。常见的用于传输的频率范围为 2~40GHz。频率越高,带宽越宽,数据传输速率也就越高。

(2) 卫星微波。通信卫星实际上是一个微波接力站,用于将两个或多个称为地球站或地面站的地面微波发送器/接收器连接起来。卫星使用上、下行两个频段,接收一个频段(上行)上的传输信号,放大或再生信号后,再在另一个频段(下行)上将其发送出去。卫星主要应用于电视广播、长途电话传输和个人用商业网络,其传输的最佳频率范围为 1~10GHz。

(3) 广播无线电波。广播无线电波是全向性的,不要求使用碟形天线,天线也无须严格地安装到一个精确的校准位置上。无线电波(radio)是笼统术语,频率范围为 3kHz~300GHz。广播无线电波(broadcast radio)是非正式术语,包括 VHF 频段和部分 UHF 频段,范围为 30MHz~1GHz。

(4) 红外线。红外线传输不能超过视线范围,且距离短的红外线传输无法穿透墙体。微波系统中遇到的安全和干扰问题在红外线传输中都不存在,而且红外线不需要频率分配许可。

(5) 光波。频率更高的光波,主要指非导向光波,而非用于光纤的导向光波。

2) GSM 全球移动通信系统

全球移动通信系统(Global System of Mobile Communication, 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 作为开放标准提供了更简易的互操作性。这样,标准就允许网络运营商提供漫游服务,用户就可以在全球使用他们的移动电话。

GSM 小组创立于 1982 年,其技术在 1987 年被提出,1990 年第一个 GSM 规范说明完成,文本长 6000 多页。商业运营开始于 1991 年,地点是芬兰的 Radiolinja。1998 年,3G 项目启动。4G 集 3G 与 WLAN 于一体,能够快速传输数据,高质量音频、视频和图像等。4G 下载速度为 100Mb/s,比家用宽带 ADSL(4Mb/s)快 25 倍,并能够满足无线服务的要求。5G 是 4G 的延伸,5G 网络的理论下行速度为 10Gb/s(相当于下载速度 1.25GB/s)。2017 年 12 月 21 日,在国际电信标准组织 3GPP RAN 第 78 次全体会议上,5G NR 首发版本正式冻结并发布。2018 年 2 月 23 日,沃达丰和华为完成首次 5G 通话测试。2018 年 8 月 3 日,美国联邦通信委员会(FCC)发布高频段频谱的竞拍规定,这些频谱将用于开发下一代 5G 无线网络。2018 年 12 月 1 日,韩国三大运营商 SK、KT 与 LG U+同步在韩国部分地区推出 5G 服务,这也是新一代移动通信服务在全球首次实现商用。同年 12 月 10 日我国工业和信息化部正式对外公布,已向中国电信、中国移动、中国联通发放了 5G 系统中低频段试验频率使用许可。2019 年 6 月 6 日,我国工业和信息化部向中国电信、中国移动、中国联通、中国广电发放 5G 商用牌照,这意味着 5G 正式商用。

3) 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吉字节,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是智能手机,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 具体特征参数如下。

传输速率: 其 5G 网络已成功在 28 千兆赫(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。而此前这一传输瓶颈被业界普遍认为是一个技术难题,而三星电子则利用 64 个天线单元的自适应阵列传输技术破解了这一难题。

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 网络还将支持可穿戴式设备。在一个给定的区域内支持无数台设备,这是设计的目标。在未来,每个人将拥有 10~100 台设备为其服务。

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。在发送短信或浏览网页的时候,在观看网络视频时,如果发现视频播放不流畅甚至停滞,这很可能就是因为端到端网络连接较差的缘故。

4) 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 1TB/s,目前已有机构开始其研发,预计 2026 年正式投入商用。

2018 年 3 月 9 日我国工业和信息化部部长苗圩表示,中国已经开始着手 6G 研究。

2019年3月15日美国联邦通信委员会投票通过了开放95GHz~3THz频段的决定,以供6G实验使用。纽约大学教授泰德·拉帕波特称:“联邦通信委员会已经启动了6G的全球竞赛。”美国总统特朗普发推特说:“我希望5G甚至6G的技术能尽快在美国普及。这比当前的标准要更强、更快、更智能。美国公司必须加紧努力,否则就会落后。我们没有理由落后……。”除中美两国外,欧盟、俄罗斯等也正在紧锣密鼓地开展相关工作。

因为中国华为公司在5G方面的技术领先优势,美国出台了一系列限制中国华为发展的政策,这使5G和6G已经附带了很多政治色彩。5G和6G已经远远超越了技术层面的发展和创新,它已上升成为国家层面的技术竞争。实际上,5G的发展需求源自高速视频图像的传输。随着人们对视频体验要求的提升,视频在媒介中占据着越来越重要的地位。除了更高的清晰度之外,一些新技术,如增强现实、虚拟现实等的融入,要求视频技术必须具有更快的传输速度和处理能力,这是6G发展的原动力。

从1G到5G,有一个“诡异”现象在不断出现,即移动通信每次更新换代时,每逢奇数G,都会出现“短命”的现象。由于1G只能语音不能上网,1971年12月被AT&T提出并实施后,很快被2G取代。尽管3G在处理图像、音乐、视频流等方面有一定优势,但4G以广带接入和分布网络为基础且50倍于3G速度实现三维图像高质量传输,而迅速将其代替。目前的5G似乎也有类似的开端现象,因为6G似乎在各方面都有较多的优势。这也提醒移动通信厂商在加紧部署5G应用推广的同时,也需尽快展开6G技术的开发和应用研究。

频率范围为95GHz~3THz的“太赫兹波”频谱被开放,供实验使用,使下一代6G无线网络的研发有了技术政策层面的许可。曾经被认为无用的太赫兹(THz,1THz=1000GHz)频谱,或将成为未来高速通信的频段。6G将迈进太赫兹时代。通常,太赫兹波指0.1~3THz的电磁波,如图5-1所示。

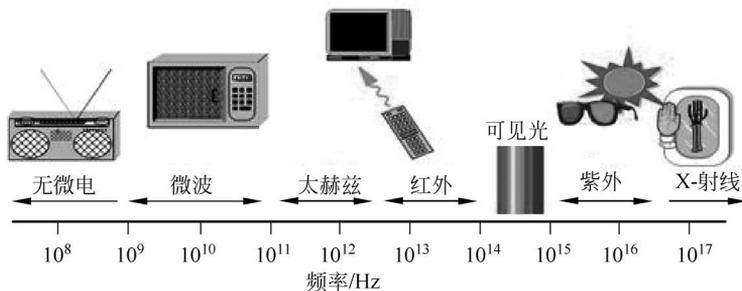


图 5-1 频率范围及其应用(<http://www.mwrf.net/news/suppliers/2012/5282.html>)

4G主要依托正交频分复用技术,而5G主要依托天线技术和高频段技术。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。

2018年6月泰克科技公司及法国著名的研究实验室IEMN已经实现了300GHz频段

中使用单载波无线链路实现 100Gb/s 数据传输。芬兰的奥卢大学无线通信中心早已展开 6G 研发,中国华为、韩国 SK 集团信息通信技术中心、美国贝尔实验室均已展开这方面的研究。预计 2025 年,6G 将进入商业应用。

华为的 6G 技术方案是通信感知一体化、星地融合和 RIS 可变信道三个维度上的技术融合,方案已完成阶段性验证,并逐步从理论走向实现。其中,①通信感知一体化包括 E-band、太赫兹、光频谱等 6G 潜在使能频段,以及定位跟踪、成像与环境重构、健康监测等场景的应用;②星地融合将整合地面网络和非地面网络,由低轨或超低轨卫星构建的大型超低轨星座;③RIS 可变信道。可重构智能超表面(reconfigurable intelligent surface)是一种用于设计无线网络和无线传输模式的潜在新型技术,可为电磁波辐射增强提供解决方案。由于控制灵活,超表面通过数字平台实现“可编程”。每个单元都是一个特定形态的小辐射器,这使来自所有单元的组合波束可按需成形并指向目标位置;从宏观上看,多个单元可以设计成一个电磁互连网络,建设性地朝着特定功能(如波吸收、表面波消除、天线去耦和波束赋形等)运行。因此,也可把在环境中布置的 RIS 阵列看作是信道的一部分,并可通过控制 RIS 参数来动态改变无线信道的传播特性。

5.3.3 移动互联网

1. 移动互联网的基本概念

移动互联网是 PC 互联网发展的产物,是移动通信和互联网二者融合的“共同体”,是互联网的技术、平台、商业模式及应用与移动通信技术结合后的应用总称。

移动互联网继承了移动随时、随地、随身和互联网开放、分享、互动的优势,是一个全域性的、以宽带 IP 技术为核心的,可同时提供语音、传真、数据、图像、多媒体等高品质电信服务的新一代开放的电信基础网络,由运营商提供无线接入,互联网企业提供各种成熟的应用。

在移动互联网中,用户使用手机、PAD 或其他无线终端设备,通过速率较高的移动网络,在移动状态下(如在地铁、公交车等)随时、随地访问 Internet 以获取信息,使用商务、娱乐等各种网络服务。通过移动互联网,人们可以使用手机、平板电脑等移动终端设备浏览新闻,还可以使用各种移动互联网应用,例如在线搜索、在线聊天、移动网游、手机电视、在线阅读、网络社区、收听及下载音乐等。在移动环境下可进行网页浏览、文件下载、位置服务、在线游戏、视频浏览和下载等应用。

移动互联网正逐渐渗透到人们生活、工作的各个领域,微信、支付宝、位置服务等丰富多彩的移动互联网应用迅猛发展,并深刻改变信息时代的社会生活。近几年,已实现了从 4G 到 5G 的跨越式发展。即将到来的 6G 全球网络覆盖将使身处大洋和沙漠中的用户可随时随地保持与世界的联系。

2. 移动互联网的特征

(1) 交互性: 用户可以用随身携带的移动终端,在移动状态下接入和使用移动互联网应用服务。

(2) 便携性: 由于移动终端小巧轻便且可随身携带的特点,人们可以将其放进随身携带的书包和手袋,并在任意场合接入网络。

(3) 隐私性: 移动终端设备的隐私性远高于 PC, 移动互联网信息保护程度要求也比较高。高隐私性要求决定了移动互联网终端应用的安全特点, 数据共享时既要保障认证客户的有效性, 也要保证信息的安全性。

(4) 定位性: 移动互联网的典型应用是位置服务应用, 比如位置签到、位置分享及基于位置的社交应用。

(5) 娱乐性: 移动互联网上的丰富应用, 如图片分享、视频播放、音乐欣赏、电子邮件等, 为用户的工作、生活带来更多的便利和乐趣。

(6) 局限性: 移动互联网应用在提供便捷服务的同时, 也受其移动通信能力和终端硬件能力的限制。

(7) 身份统一性: 其指移动互联用户自然身份、社会身份、交易身份、支付身份通过移动互联网平台得以统一。

3. 移动互联网的发展历程

整个移动互联网发展历史可以归纳为 4 个阶段: 萌芽阶段、培育成长阶段、高速发展阶段和全面发展阶段。

(1) 萌芽阶段(2000—2007 年)。其主要是基于 WAP(无线应用协议)的应用模式。受限 2G 网速和非智能手机, 移动互联网只是 WAP 把互联网 HTML 信息转换后推送到手机屏上。

(2) 培育成长阶段(2008—2011 年)。2009 年 1 月 7 日我国工业和信息化部发放 3G 网络牌照。再加上智能手机的出现, 移动网速的大幅提升初步破解了手机上网带宽瓶颈, 移动智能终端丰富的应用软件让移动上网的娱乐性得到大幅提升。

(3) 高速发展阶段(2012—2013 年)。进入 2012 年移动上网需求大增, 手机厂商效仿苹果模式, 推出了触摸屏智能手机, 这使手机上网浏览更方便, 移动应用更丰富。

(4) 全面发展阶段(2014 年至今)。2013 年 12 月 4 日我国工业和信息化部发放 4G 网络牌照。4G 网速提高, 促进了实时性增强、流量增大、速度加快, 使移动视频应用更流畅。5G 进一步提速, 倍增了手机视频应用。

5.4 信息资源处理技术

5.4.1 分布式计算与并行计算

1. 分布式计算

1) 分布式计算概述

分布式计算是一种计算方法, 与集中式计算相对。随着计算技术的发展, 有些应用需要非常巨大的计算能力才能完成。如果采用集中式计算, 需要耗费相当长的时间来完成。分布式计算将该应用分解成许多小的部分, 分配给多台计算机处理。这样可以节约整体计算时间, 大大提高计算效率。

分布式计算具有以下优点: 稀有资源共享; 通过分布式计算在多台计算机上平衡计算负载; 把程序放在最适合运行它的计算机上。

随着计算机的普及,个人计算机进入千家万户。随之出现的问题是越来越多的计算机设备处于闲置状态,即使在开机状态下中央处理器的潜力也远远不能被完全利用。可以想象,一台家用的计算机将大多数的时间花费在“等待”上面。即便是使用者实际使用他们的计算机时,处理器依然是寂静地消费,依然是不计其数的等待(等待输入,但实际上并没有做什么)。互联网的出现,使得连接调用所有这些拥有限制计算资源的计算机系统成为了现实。

如果将闲置状态的计算资源“整合”起来,其计算速度将变得非常迅速,而且被实践证明是的确可行的。目前一些较大的分布式计算实验证明,其处理能力已经可以达到甚至超过目前世界上速度最快的巨型计算机。

2) 网络计算技术

网络计算是目前最重要的分布式计算技术之一,它通过网络系统将分布在不同地点或区域的计算机资源(包括各种硬件和软件以及信息数据等)连接成一个巨大的“异构计算机”,虽然这些计算资源分布在各自不同的计算机上,这些计算机可能有不同的操作系统、不同的技术协议,但是通过网络技术组建的这个系统却可以像一台计算机一样对这些资源进行管理和利用,从而完成一些计算规模巨大的复杂运算和数据处理任务。

网络计算(grid computing)通过网络连接地理上分布的各类计算机(包括机群)、数据库、设备等,形成对用户相对透明的虚拟的高性能计算环境,应用包括分布式计算、高吞吐量计算、协同工程和数据查询等诸多功能。网络计算被定义为一个广域范围的“无缝的集成和协同计算环境”。

从另一个意义上说,这种计算资源的统一管理和共享,不仅仅为提供复杂计算提供支持,还可以在很大的区域范围内打破企业、组织和国家界限,避免重复资源的投资和浪费,充分利用自己的计算资源。

网络是把整个因特网整合成一台巨大的超级计算机,实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源的全面共享。当然,网络并不一定非要这么大,可以构造地区性的网络,比如区域网络、局域网络、企事业单位内部网络、家庭计算机网格等。事实上,网格的根本特征是资源共享,而不是它的规模。由于网格是一种新技术,因此具有新技术的两个特征:其一,不同的群体有不同的名词称谓;其二,网格的精确含义和内容还没有固定,而是在不断变化。

2. 并行计算

1) 并行计算产生的背景

早期微处理器的计算性能以平均每年 50% 的速度提升。受摩尔定律的限制,微处理器性能提升速度渐渐变慢。进入 21 世纪之后,主流的微处理器制造商已决定通过并行处理来快速提升微处理器的计算性能,以求进一步提升计算能力。CPU 的核数越来越多将成为一种趋势,2019 年普通桌面 PC 已达 128 核。

并行计算与串行计算的区别在于:串行计算只在单个 CPU 上进行求解,而并行计算则是同一个时间段内在多个 CPU 上求解;从硬件角度上来讲,串行计算就是在普通计算机上求解,并行计算则是在并行计算机上求解。需要并行计算求解的问题能分成很多并行子问题解决。

随着大数据时代的到来,大规模数据处理对高性能计算(high performance computing)

提出了需求。为将一个需要大规模处理的应用部署至高性能计算机上,需要开展并行计算(parallel computing)。广义上的并行计算可分为时间上的并行和空间上的并行,时间上的并行泛指流水线技术,而空间上的并行是指多个处理器并发地进行计算。并行计算科学中主要研究空间上的并行问题,即先将计算任务分解为若干个计算子任务,然后同时使用多个计算资源以快速解决一个大型且复杂的计算问题。

2) 并行计算概述

并行计算或称平行计算,是一种一次可执行多个指令的算法,目的是提高计算速度,及通过扩大问题求解规模,解决大型而复杂的计算问题。

并行计算是指同时使用多种计算资源解决计算问题的过程,是提高计算机系统计算速度和处理能力的一种有效手段。它的基本思想是用多个处理器来协同求解同一问题,即将被求解的问题分解成若干个部分,各部分均由一个独立的处理机来并行计算。并行计算系统既可以是专门设计的、含有多个处理器的超级计算机,也可以是以某种方式互连的若干台的独立计算机构成的集群。通过并行计算集群完成数据的处理,再将处理的结果返回给用户。

并行计算的特点主要表现在:①并行计算由相对独立的模块独立分开管理控制;②在并行计算中会存在多个模块同时异步进行,该并行能有效降低运行耗费的时间;③所有的模块与模块之间存在一定的交互;④并行计算可能由于并发和交互导致错误发生。

并行体系的主要特点为:①并行性,多流水、超标量是提升 CPU 并行性能的重中之重。在新型并行体系中,根据并行的程度分级。②可扩展性,是指增加资源时,性能和功能得到提升。认识可扩展性应该认识到成本伸缩的比例应该小于线性比例。

并行计算采用的方法主要包括:①资源共享,是指用软件方法实现资源在某段时间内可以同时访问;②资源重复,是指通过重复设置和利用硬件与软件资源,以此来提高计算机系统的性能;③时间重叠,例如不同的进程在时间上相同资源的使用相互错开或者轮流重复利用某硬件资源,来提升速度。并行遗传算法是并行计算的一个分支应用,将会大大减少得到最优解所需要的时间。

3) 并行计算的发展趋势

首先,计算机体系结构中,多核逐渐成为主流。要想进一步提升系统的运算能力,仅仅依靠提升晶体管的集成度是不够的,因为系统运算能力还会受到材料物理性能等因素的影响,因此,系统中应用多核技术已经成为一种必然趋势。所谓多核技术,就是将多个计算内核集成在一个处理器中,每一个内核都能完成一个计算指令,这样一个处理器就能够完成并行计算指令。多核 CPU 的计算密度更高,并行处理能力更强,在相同计算条件下所消耗的功率更低,可以满足实际需要。

其次,异构众核集成技术应用越来越广泛。一般情况下模型中都采用 CPC(协同并行计算)与 MIC(集成众核)组合构架模式,相对复杂的逻辑计算部分由 CPU 负责,而一些密集运算则由 CPU 或者是 MIC 负责,这些密集运算的典型特征就是分支较少且并行度高,这种异构方式为超级计算机的发展奠定了基础。

再次,服务器逐渐向着大规模集群化且廉价化的方向发展,越来越多的互联网企业和网络运营商选择将大规模服务廉价集群作为系统硬件设施。这种集群式服务器的典型特征就是会自动将故障状态视为常态,因为即使集群中的一部分组间发生故障,也不会对系统整体

造成太大影响。同时,该集群会为异构硬件扩容提供支持,或是在系统中加入存储资源,或是直接加入新的机器,系统根据实际情况对这些资源或者机器进行自动调取,这一过程不会对系统运行产生任何影响。

最后,随着计算机技术的发展,大数据编程模型已经取得了不错的研究成果,其主要被应用在数据分析以及数据处理上。为了提高计算能力,大数据编程模型会为系统资源提供横向扩展支持,同时程序中自带容错机制,一旦出现节点失效问题能够及时应对。大数据的应用效率会受到很多因素的影响,包括并行性级别、通信问题以及存储问题等,性能优化大多数都是针对某个系统框架或者是某个模型而言的,没有一个完整、统一的理论,而面向大数据处理的并行计算模型就要解决这一问题,统一优化理论,制定出可以面向所有模型的优化方法。

5.4.2 服务器刀片化与虚拟化

1. 刀片化

1) 刀片服务器概述

刀片服务器(blade server)是指在标准高度的机架式机箱内可插装多个卡式的服务器单元,是一种实现高可用高密度(High Availability High Density, HAHD)的低成本服务器平台,为特殊应用行业和高密度计算环境专门设计。刀片服务器就像“刀片”一样,每一块“刀片”实际上就是一块系统主板。它们可以通过“板载”硬盘启动自己的操作系统,类似于一个个独立的服务器,在这种模式下,每一块主板运行自己的系统,服务于指定的不同用户群,相互之间没有关联。不过,管理员可以使用系统软件将这些母板集成成一个服务器集群。在集群模式下,所有的母板可以连接起来提供高速的网络环境,并同时共享资源,为相同的用户群服务。在集群中插入新的“刀片”,就可以提高整体性能。而由于每块“刀片”都是热插拔的,所以,系统可以轻松地替换,并且将维护时间减少到最小。

刀片服务器在设计之初都具有低功耗、空间小、单机售价低等特点,同时它还继承发扬了传统服务器的一些技术指标,比如把热插拔和冗余运用到刀片服务器中,这些设计满足了密集计算环境对服务器性能的需求;有的还通过内置的负载均衡技术,有效地提高了服务器的稳定性和核心网络性能。而从外表看,与传统的机架式服务器/塔式服务器相比,刀片服务器能够最大限度地节约服务器的使用空间和费用,并为用户提供灵活、便捷的扩展升级手段。

刀片服务器比机架式服务器更节省空间,同时,散热问题也更突出,往往要在机箱内装上大型强力风扇来散热。此型服务器虽然空间较节省,但是其机柜与刀片价格都不低,一般应用于大型的数据中心或者需要大规模计算的领域,如银行电信金融行业以及互联网数据中心等。

2) 刀片服务器的发展趋势

刀片服务器由于节约空间、便于集中管理、易于扩展和提供不间断的服务,成为下一代服务器的新要求。结合目前推出的各种新技术,可大大提高刀片服务器的性能。

(1) 高性能的处理器:未来的服务器可通过采用更高性能的处理器、内存等硬件的方式来提高单个刀片的处理能力,同时提高系统的计算能力。

(2) 虚拟化: 可以采用虚拟化和云计算的方式, 根据实时的数据处理要求来调度不同数据中心的服务器进行运算。虚拟化技术是一种比较实用的技术。服务器的虚拟化特性将在实践中得到更多的应用。

(3) 单芯片多处理器: 随着处理器技术的发展, 在单个刀片上可以集成多个 CPU, 这样在能耗、散热上都会比传统的刀片服务器更具有优势。目前各个厂商都在进行这方面服务器的开发。

2. 虚拟化技术概述

虚拟化(virtualization)是一种资源管理技术, 是将计算机的各种实体资源, 如服务器、网络、内存及存储等, 予以抽象、转换后呈现出来, 打破实体结构间的不可切割的障碍, 使用户可以比原本的组态更好的方式来应用这些资源。这些资源的新虚拟部分是不受现有资源的架设方式、地域或物理组态所限制的。一般所指的虚拟化资源包括计算能力和存储能力。在实际的生产环境中, 虚拟化技术主要用来解决高性能的物理硬件产能过剩和老旧硬件产能过低的重组重用, 透明化底层物理硬件, 从而最大化地利用物理硬件。

3. 刀片化 + 虚拟化

IT 开发商围绕着服务器以下几方面的研发: 一是刀片服务器硬件系统的投资; 第二是围绕刀片服务器的生态系统投资, 如 RDMA 解决方案、存储器的扩展解决方案等; 第三是在服务器虚拟化方面以及管理软件方面。刀片化 + 虚拟化将成为未来服务器的技术发展趋势。

1) 刀片将取代主机

由于刀片服务器具有对向外扩展(scale-out)和向上扩展(scale-up)的优点, 刀片服务器替代主机将成为趋势。这是因为除了硬件小巧、高性能的优势外, 软件自动管理工具和部署工具非常简单。刀片服务器价格不高、运营成本低、设备维护简单、组件更新容易、运行速度快等特点正在改变人们对此的思维方式。

2) 通过虚拟提高使用率

服务器虚拟化技术已经成熟。比如, 虚拟分区的技术包括硬件分区、软件分区、资源分区等分区技术, 不同的分区可以运行在包括 UNIX、Linux、Windows 等不同的操作系统中, 提高整合环境中的系统利用率。负载管理器(work load manager)可以根据客户工作负载的不同需求来分配资源, 甚至可以细化到 CPU、内存、磁盘带宽等资源。例如, 当客户同时运行多个业务时, 可为每种业务设置优先级, WLM 可根据业务优先级、响应时间、内存资源需求等服务级目标(SLO), 在虚拟分区间调配 CPU 实现计算资源的自动调节。

4. 虚拟化应用

1) 服务器虚拟化

服务器虚拟化是利用虚拟化技术在物理服务器上划分出 $N(N > 10)$ 台虚拟逻辑服务器, 这些虚拟出的逻辑服务器以独立个体的形式运行。逻辑服务器云计算中的虚拟化关键技术应用计算机物理资源管理和使用方式的根本改变是云计算模式突破的关键。通过虚拟化技术, 可以快速虚拟出一个按需配置、独立的虚拟计算机资源供用户使用。计算机资源的虚拟化可以是系统虚拟、硬件虚拟、软件虚拟、存储虚拟等, 使用属性不同决定了其虚拟方式不同, 其目的是根据应用的具体负荷情况对资源进行调度, 充分整合计算与存储资源, 使得

计算机资源的利用率最大化。虚拟化技术的实质是使用虚拟监控器管理底层硬件资源,将计算机资源逻辑抽象化,把单一的存储、计算、应用与服务都变成可跨域使用、动态分配、伸缩与扩展自由的资源,也能让故障独立隔离,在逻辑上以独立整体服务模式提供给用户使用,以满足灵活多变的用户需求。云计算核心和关键的技术原动力即是虚拟化技术的广泛应用。

2) 存储虚拟化

存储虚拟化技术旨在提高设备存储效率,整合不同类型存储资源,为用户提供统一访问接口,从而解决异构存储系统的扩展性、兼容性、容错性等问题。存储虚拟化结构有两种模式:对称结构与非对称结构。对称结构也称作“带内存储虚拟化技术”。存储设备的虚拟化主要在存储设备和应用服务器的数据路径上实现。它的数据与控制信息使用相同的传输路径,使用虚拟化管理软件(运行在虚拟化控制器上)实现虚拟化功能。非对称结构亦称“带外存储虚拟化技术”,将虚拟化管理软件安装在存储网络中的独立服务器上,以此实现存储设备的逻辑映射、存储分配、数据安全保障等功能;先访问映射后的虚拟设备,而后通过数据通路直接访问存储设备,数据和指令不并在同一个路径上。

云环境下的存储虚拟化结构可设计成三层(物理层、逻辑层、虚拟层)模式。虚拟层与逻辑层之间的映射表存放着虚拟卷和逻辑卷的关系信息,映射表信息的建立与更新,可实现虚拟卷存储容量的动态扩充与缩减,满足存储容量的实时需求。逻辑层与物理层的映射表存放着逻辑存储池与存储节点的关系信息,通过映射表可以准确地定位逻辑存储池在存储节点中的物理地址。层与层之间通过映射表链接,以实现存储资源的统一管理 with 动态分配。用户无须关心存储设备所在的具体位置,只管放心存储与访问数据。存储虚拟化技术的应用有着两大优势:首先减少了在云存储中物理存储介质之间因厂家原因而存在的差异性;其次是存储空间的灵活伸缩性,可动态扩展存储空间,按需动态分配存储空间,大大减少了设备费用的投入。

3) 桌面虚拟化

随着云计算的高速发展,在传统企业终端和资源整合管理领域中产生了新型的典型应用——桌面云。桌面云采用虚拟化技术将个人计算机终端与用户的桌面工作环境分离开来进而迁移,每个用户的操作系统、应用和用户配置文件等数据以整体打包的方式存储在云服务器上,以镜像方式配置专属的虚拟桌面。以浏览器或专业程序为介质平台,访问存储在云服务器上的虚拟桌面以及各种应用程序,所有操作的数据结果将最终保留在云计算中心,用户无须额外再配置应用程序和文件,可随时更换地点和客户端,使用所产生的体验仿佛就像用户使用自己的个人计算机一样,并无差异。

5.4.3 云计算与区块链

1. 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向用户按需服务。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、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。通过使计算分布在不同的分布式计算机上,而不是本地计算机或远程服务器中,企业的数据中心运行将与互联网更相似。这将使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三个层次的服务:基础设施即服务(IaaS)、平台即服务(PaaS)和软件即服务(SaaS)。云计算服务通常提供通用的通过浏览器访问的在线商业应用,软件和数据可存储在数据中心。

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。

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。

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

云计算的三级分层为云软件、云平台、云设备。云软件 SaaS 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。云平台 PaaS 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编写程序,提供服务,一般消费者也可以在上面运行程序。云设备 IaaS 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。

2. 区块链

1) 区块链的概念

区块链(blockchain),从技术层面看涉及数学、密码学、互联网和计算机编程等技术。从应用视角看,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。这些特点保证了其“诚实”与“透明”,为区块链创造信任奠定基础。区块链丰富的应用场景,针对信息不对称问题,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。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,每一个数据块中包含了一批交易的信息,用于验证其信息的有效性

(防伪)和生成下一个区块。区块链的安全风险是制约其健康发展的短板。

2) 类型

(1) 公有区块链: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。

(2) 联合(行业)区块链:由某个群体内部指定多个预选的节点为记账人,每个块的生成由所有的预选节点共同决定,其他接入节点可以参与交易,但不过问记账过程,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: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,本链与其他的分布式存储方案没有太大区别。

3) 特征

(1) 去中心化。区块链技术不依赖额外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。

(2) 开放性。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,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) 架构模型

区块链系统架构模型,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。

5.4.4 超级计算机与量子计算机

1. 超级计算机

1) 超级计算机的概念

超级计算机(super computer)在《计算机科学技术百科全书》中的解释是:具有非常高的运算速度,有非常快而容量又非常大的主存储器 and 辅助存储器,并充分使用并行结构软件的计算机。

超级计算机是计算机中功能最强、运算速度最快、存储容量最大的一类计算机。一般是指由数以万计处理器(机)组成的、能计算普通 PC 和服务 器不能完成的大型复杂课题的计算机。超级计算机两个主要特点是:存储容量非常大和计算速度非常快,即数据处理能力

特别强。

2) 超级计算机的结构

超级计算机的体系结构有以下几种。

第一,对称多处理(Symmetric Multi-Processor, SMP)。其结构是在单个机柜中部署多个处理器,各处理器完全相同,平等访问软硬件资源,处理器间通过总线或者交叉开关相连,共享存储器,但各自有独立的 Cache。

第二,大规模并行处理(Massively Parallel Processing, MPP)。MPP 指在同一地点由大量处理器构成的并行计算机,以处理器作为节点,采用分布式存储,节点间通信用消息传递方式,其规模可扩展到数千万个节点。

第三,机群(cluster)。其为互相连接的多个独立计算机的集合,采用高速互连网络连接起来的一组微机或工作站,各节点彼此独立(有独立完整的内存和操作系统)。

第四,群聚集(constellations)。指以大型 SMP(处理器数目不少于 16 个)为节点构成的机群,各节点间通过高速专用网络互连,也称为机群 SMP(Cluster-SMP 或 CSMP)。

3) 超级计算机之争

1978 年我国开始“银河”超级计算机的研制,1983 年完成。2010 年我国超级计算机获得全球第一名以后又多次获得第一名,随后中美超级计算机之争日趋白热化。2018 年 10 月 24 日神威 E 级原型机、“天河三号”E 级原型机和曙光 E 级原型机系统完成交付。接着,中国暂停了超级计算机的投入,并转向量子计算机,因为量子计算机速度更快。

2. 量子计算机

1) 定义

量子计算机(quantum computer)是一类遵循量子力学规律进行高速数学和逻辑运算、存储及处理信息的物理装置。

量子计算机对每一个叠加分量进行变换,以实现并行计算,也称为量子优越性。量子计算机由许多硬件和软件组成。其中,软件包括量子算法、量子编码等;硬件包括量子晶体管、量子存储器、量子效应器等。量子计算机,以量子态为记忆单元和信息存储形式,以量子动力学演化为信息传递与加工基础的量子通信与量子计算。量子计算机的基本单位量子比特(qubit)用 0 或 1 表示,其分别对应量子力学体系中的两个态。量子晶体管通过电子高速运动来突破物理的能量界限,从而实现晶体管的开关作用,这种晶体管控制开关的速度非常快,晶体管比普通芯片的运算能力更强,而且对使用环境条件适应能力更强。量子存储器是一种存储信息效率很高的存储器,它能够在非常短的时间里对计算信息进行赋值。量子计算机的效应器是量子计算机的控制系统,主要用于各部件的运行控制。

2) 技术突破

2007 年加拿大 DWave 公司成功研制出一台具有 16 量子比特的“猎户星座”量子计算机。2019 年 10 月 23 日谷歌模仿“猎户星座”研制出了包含 53 个可用量子比特可编程的超导量子计算机。2020 年 12 月 4 日中国成功构建 76 个光子的量子计算原型机“九章”。2021 年 10 月 26 日,中国完成光子数 113 个的量子计算原型机“九章二号”和 66 比特可编程超导量子计算原型机“祖冲之二号”。2023 年 10 月 11 日,中国成功构建 255 个光子的量子计算原型机“九章三号”,其速度比最强的超级计算机“前沿”快一亿亿倍。目前中国在量子计算机领域已经处于世界领先地位。

3) 量子计算机技术的现状和趋势

第一阶段,开发 50~100 个量子比特的高精度专用量子计算机,可解决超级计算机无法解决的高复杂度特定问题,以实现“量子计算优越性”。

第二阶段,通过对规模化多量子体系的精确制备、操控与探测,研制可相干操纵数百个量子比特的量子模拟机,用于解决超级计算机无法胜任的工作。

第三阶段,通过积累在专用量子计算与模拟机研制过程中的各种技术,提高量子比特的操纵精度使之达到能超越量子计算苛刻的容错阈值(大于 99.9%),大幅度提高可集成的量子比特数目至百万甚至千万量级,实现容错量子逻辑门,研制可编程的通用量子计算原型机。

第四阶段,大规模部署应用百万甚至千万量级量子计算机,以取代传统计算机。目前大数据和人工智能的应用,特别是类似 ChatGPT 的应用对算力提出了巨大需求。为解决大数据和人工智能应用对算力需求的“瓶颈”问题,量子计算机的快速部署应用是解决该问题的“最佳方案”。

截至目前,前两个阶段的工作已经完成,后两个阶段是近未来必须尽快解决的问题。

5.5 信息资源输出技术

5.5.1 打印机及其分类

打印机是计算机的输出设备之一,用于将计算机处理结果打印在相关介质上。

1. 打印机质量

衡量打印机好坏的指标有三项:打印分辨率、打印速度和噪声。下面主要介绍打印分辨率和打印速度。

1) 打印分辨率

打印机分辨率又称为输出分辨率,是指在打印输出时横向和纵向两个方向上每英寸最多能够打印的点数,通常以“点/英寸”即 dpi(dot per inch)表示。而所谓最高分辨率就是指打印机所能打印的最大分辨率,也就是所说的打印输出的极限分辨率。平时所说的打印机分辨率一般指打印机的最大分辨率,目前一般激光打印机的分辨率均在 600dpi×600dpi 以上。

打印分辨率是衡量打印机打印质量的重要指标,它决定了打印机打印图像时所能表现的精细程度,它的高低对输出质量有重要的影响,因此在一定程度上来说,打印分辨率也就决定了该打印机的输出质量。分辨率越高,其反映出来可显示的像素个数也就越多,可呈现出更多的信息和更好、更清晰的图像。

打印分辨率一般包括纵向和横向两个方向,它的具体数值大小决定了打印效果的好坏与否,一般情况下激光打印机在纵向和横向两个方向上的输出分辨率几乎是相同的,但是也可以人为来进行调整控制;而喷墨打印机在纵向和横向两个方向上的输出分辨率相差很大,一般情况下我们所说的喷墨打印机分辨率就是指横向喷墨表现力。如 800×600dpi,其中 800 表示打印幅面上横向方向显示的点数,600 则表示纵向方向显示的点数。分辨率不

仅与显示打印幅面的尺寸有关,还要受打印点距和打印尺寸等因素的影响,打印尺寸相同,点距越小,分辨率越高。

打印机分辨率越高,输出的效果就越精密。但是,并不是每种打印需求都需要最高精度的打印。对于文本打印而言,600dpi 已经达到相当出色的线条质量。但在现代的办公中,打印文档的类型日益多样化,图像、照片、CAD、GIS 等需要高精度打印的内容越来越多,在这个时候,除了打印负荷量和打印速度外,用户必须仔细考虑打印机的打印质量能否满足自己的需求。对于照片打印而言,更高的分辨率意味着更加丰富的色彩层次和更平滑的中间色调过渡,经常需要 1200dpi 以上的分辨率才可以实现。

2) 打印速度

打印速度是指打印文稿所需要的时间,一般分为彩色文稿打印速度和黑白文稿打印速度。打印速度取决于打印机所采用的打印方式。热敏打印方式由于要加热热敏打印头,速度相对比较慢,而喷墨打印方式采用直接打印方式相对来说比较快,激光打印在三种打印方式中是最快的,一般激光打印的彩色和黑白打印速度都在每秒十几页。

三种打印方式中虽然激光和喷墨打印速度都比較快,但是由于采用的墨粉和墨盒耗材价格较高,因而打印成本较高。热敏打印采用价格比较低廉的热敏纸,所以打印成本比较低。

评价一台打印机是否优劣,不仅要看打印图像的品质,还要看打印速度。打印速度用每分钟打印多少页纸(PPM)来衡量。另外,打印速度与打印时设定的分辨率有直接的关系,打印分辨率越高,打印速度越慢。一般打印机速度为每分钟出纸 70 张(黑白双面打印、普通 A4 复印纸)和每分钟出纸 50 张(彩色双面打印、彩激 A4 复印纸)。

2. 打印机分类

1) 针式打印机

针式打印机从 9 针到 24 针,有窄行和宽行之分,几十年来流行不衰,与它极低的打印成本和很好的易用性以及单据打印的特殊用途是分不开的。当然,它很低的打印质量、很大的工作噪声也是它无法适应高质量、高速度的商用打印需要的根本原因,所以现在只有在银行、超市等可以看见它的踪迹。针式打印机色带分为宽带和窄带。部分色带可以单独更换,部分色带须连色带架一起更换。可以根据需要,更换不同颜色的色带。针式打印机如图 5-2 所示。

2) 喷墨打印机

喷墨打印机因其有着良好的打印效果与较低价位的优点而占领了广大中低端市场。喷墨打印机在打印介质的选择上有一定的优势:既可以打印信封、信纸等普通介质,还可以打印各种胶片、照片纸、光盘封面、卷纸、T 恤转印纸等特殊介质。喷墨打印机墨水有 4 种颜色、5 种颜色或 6 种颜色等几种,打印机墨水一般可以单独更换其中一种颜色的墨水,打印喷头可以永久使用。这种打印机好处是换墨水的成本较低,不足之处是如果打印头多次使用后,那么打印质量会有所下降,也容易出现堵塞喷嘴的问题,严重的话打印机要维修或报废;有些打印机喷嘴和墨盒是一体的,更换墨盒时,连同墨盒底部的喷嘴也一同被换下来,这种墨盒的成本比较贵,好处是这种打印机不会出现喷嘴堵塞的问题,如果堵塞,那么换掉墨盒后,打印机还能用,打印质量可以保持精美。喷墨打印机如图 5-3 所示。



图 5-2 针式打印机



图 5-3 喷墨打印机

3) 激光打印机

激光打印机是高科技的产物,有望代替喷墨打印机,有黑白和彩色两种,具有高质量、更快速、更低成本的特点。其中,低端黑白激光打印机的价格目前已经降到了几百元。其原理是利用光栅图像处理器产生要打印页面的位图,然后将其转换为电信号等一系列的脉冲送往激光发射器,在这一系列脉冲的控制下,激光被有规律地放出。与此同时,反射光束被接收的感光鼓所感光。激光发射时就产生一个点,激光不发射时就是空白,这样就在接收器上印出一行点来。然后接收器转动一小段固定的距离继续重复上述操作。当纸张经过感光鼓时,鼓上的着色剂就会转移到纸上,印成了页面的位图。最后当纸张经过一对加热辊后,着色剂被加热融化,固定在了纸上,就完成打印的全过程,整个过程准确而且高效。与喷墨打印机比,其单页打印成本要便宜很多。彩色激光打印机的价位较高,但其打印结果和打印成本都无法与彩色喷墨打印机比。有些激光打印机的墨粉和硒鼓是可以分离的,墨粉用完后,可以方便地填充墨粉,然后继续使用,直到硒鼓老化更换;有些激光打印机墨粉和硒鼓是一体的,墨粉用完后,硒鼓要弃掉,造成一定的浪费。硒鼓的成本在整机成本中占很大一部分比例。激光打印机如图 5-4 所示。



图 5-4 激光打印机

4) 其他打印机

除了以上三种最为常见的打印机外,还有热转印打印机和大幅面打印机等几种应用于专业方面的打印机机型。热转印打印机是利用透明染料进行打印的,它的优势在于专业、高质量的图像打印,可以打印出近似照片的连续色调的图片来,一般用于印前及专业图形输出。大幅面打印机的打印原理与喷墨打印机基本相同。它的主要用途一直集中在工程与建筑领域。随着其墨水耐久性的提高和图形解析度的增加,大幅面打印机也开始被越来越多的应用于广告制作、大幅摄影、艺术写真和室内装潢等装饰宣传的领域。

5.5.2 数码复印机与数码印刷机

1. 数码复印机

数码复印机如图 5-5 所示。数码复印机相当于把扫描仪和激光打印机功能相融合的设备。数码复印机与模拟复印机的主要区别主要是工作原理不同。模拟复印机的工作原理是:通过曝光、扫描将原稿的光学模拟图像通过光学系统直接投射到已被充电的感光鼓上

产生静电潜像(latent image, 又称潜影), 再经过显影、转印、定影等步骤, 完成复印过程。数码复印机的工作原理是: 首先通过 CCD(电荷耦合器件)传感器对通过曝光、扫描产生的原稿的光学模拟图像信号进行光电转换, 然后将经过数字技术处理的图像信号输入到激光调制器, 调制后的激光束对被充电的感光鼓进行扫描, 在感光鼓上产生由点组成的静电潜像, 再经过显影、转印、定影等步骤, 完成复印过程。

2. 数码印刷机概述

1) 数码印刷与传统印刷的区别

数码印刷就是借助计算机处理的数字文件直接进行印刷, 因此也称计算机直接印刷, 英文为 computer to press。数码印刷和印刷数码化是两个不同的概念。印刷数码化泛指全过程的部分或全部的数码化。例如, 激光照排、远程传版、数码打样、计算机直接制版、数字化工作流程、印刷厂 ERP 等都属于印刷数码化的范畴。数码印刷则指具体的数码印刷技术、过程和应用。后者比前者的概念范畴小而窄。

数码印刷是将计算机文件直接印刷在纸张上, 也是有别于传统印刷烦琐的工艺过程的一种全新印刷方式。它的特点是: 一张起印, 无须制版, 立等可取, 即时纠错, 可变印刷, 按需印刷。数码印刷是在打印技术基础上发展起来的一种综合技术, 以电子文本为载体, 通过网络传递给数码印刷设备, 实现直接印刷。印刷生产过程中无版和信息可变是最大特征, 涵盖印刷、电子、计算机、网络、通信等多种技术领域。

2) 数码印刷设备的分类

目前市场上现有的数码印刷设备大致可分为以下三大类。

(1) 数码印刷一体机。

这类数码印刷设备由复印机技术加数码技术发展而来, 包括三大部分: 扫描部分、激光打印部分和折页装订部分。其功能比较全面, 可以进行扫描、复印、打印和数码发布。典型代表有东芝、柯尼卡、理光、佳能、施乐、奥西等。该类设备属于多功能数码印刷系统, 是数码印刷的一类, 由以前的单一的输出设备转变了数码输入、输出一体的设备。数码印刷一体机的特点与优势: 输出速度有每分钟 20、30、40、50、60、70、80、100 页甚至更高, 适应市场面广; 由于数码技术的加入而保证了质量稳定; 设备操作简单、工作环境清洁; 数码“集团”服务系统变成现实, 用户更方便; 系统造价低, 具有较强的竞争力。

(2) 数码多功能一体机。

这类数码印刷设备以打印机技术为基础发展而来。这类设备从原理上讲与数码印刷一体机相同, 但由于并未能掌握高速激光打印机芯技术(佳能、施乐除外), 所以多数厂家的数码一体机只能在低速(每分钟 30 页以下)市场中竞争, 甚至争夺传真机的市场和打印机的市场。这类设备很难进入数码印刷市场。数码多功能一体机如图 5-6 所示。

(3) 数码印刷机。

这类数码印刷设备由印刷机技术和数码技术发展而来。其典型代表为 HP INDIGO、海德堡 DI。根据不同的数码技术, 可以设计出不同功能的系统: ①简易 CTP 技术和小胶印



图 5-5 数码复印机

技术的结合形成非一体；②机银盐制版技术和小胶印技术的结合实现；③高速喷墨技术和小胶印技术相结合实现。数码印刷机如图 5-7 所示。



图 5-6 数码多功能一体机



图 5-7 数码印刷机

5.5.3 多模态显示技术

多模态显示指在显示时采用的一种显示模式。常见的显示模式,比如文本显示模式,图像或视频显示模式。组合显示模式有图片+文字、图像+音频、图片+文字+音频、文字+音频等混合显示模式。多媒体和虚拟现实属于由多种硬件、软件以及图文声像视等参与的多模态显示模式。

1. 静态显示技术

1) 新型显示器

近几年,PDP 等离子显示器、有机 EL 显示器、FED 场致发射显示器、LED 显示器、立体 3D 显示器等各种领域的技术有比较大的发展。

2) 电子纸

电子纸技术是可以实现像纸一样阅读舒适、超薄轻便、可弯曲、超低耗电的显示技术。电子纸是一种类似纸张的电子显示器,其兼有纸的优点(如视觉感观几乎完全和纸一样等),又可以像常见的液晶显示器一样不断转换、刷新显示内容,并且比液晶显示器省电得多。电子纸如图 5-8 所示。实现电子纸技术的途径主要包括胆固醇液晶显示技术、电泳显示技术(EPD)以及电润湿显示技术等。



图 5-8 电子纸

3) 电子书

根据新闻出版总署的定义,电子书是指将文字、图片、声音、影像等内容数字化的出版物以及植入或下载数字化文字、图片、声音、影像等内容的集存储介质和显示终端于一体的手持阅读器。电子书如图 5-9 所示。



图 5-9 电子书

2. 多媒体技术

1) 多媒体技术概述

多媒体技术是计算机技术和视频技术的结合,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体。

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、计算机等中介媒质,还可以是存储信息的磁盘、光盘、磁带等存储实体。多媒体技术包括音频技术、视频技术、图像技术、通信技术、存储技术等。

2) 多媒体系统

一般的多媒体系统由以下四部分组成。

(1) 多媒体硬件系统:包括计算机硬件、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。

(2) 多媒体操作系统:或称为多媒体核心系统,具有实时任务调度、多媒体数据转换、对多媒体设备的驱动和控制及图形用户界面管理等功能。

(3) 媒体处理系统工具:或称为多媒体系统开发工具,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用户的应用软件系统,它是面向大规模用户的系统产品。

3. 虚拟现实

1) 虚拟现实概述

虚拟现实技术将计算机、传感器、图文声像等多种设置结合在一起,创造出一个虚拟的“真实世界”。在这个世界,人们看到、听到和触摸到的是一个并不存在的虚幻景象,是模拟技术使人们产生的“身临其境”的感觉。虚拟现实是一种三维的、由计算机制造的模拟环境。在这个环境中,用户可以操控机器,与机器相互影响,并完全沉浸其中。因此,从这个定义上看,“虚拟”是从计算机的“虚拟记忆”这个概念派生出来的。虚拟现实为人们提供了一个与我们现实生活极为相似的虚幻世界。虚拟现实不仅仅是一种设计,还是一个表达和交流的媒体。借助头盔、数字手套和其他传感设备,一个人可以与另一个“虚拟人”进行交流。虚拟现实中的虚拟人可以是机器,也可以是现实人的“虚影”(见图 5-10)。

2) 虚拟现实系统

虚拟现实(Virtual Reality, VR)是近年来出现的高新技术。虚拟现实是一项综合集成技术,涉及计算机图形学、人机交互技术、传感技术、人工智能等领域,它用计算机生成逼真



图 5-10 虚拟现实系统、三维头盔和数字手套

的三维视觉、听觉、嗅觉等感觉,使人作为参与者通过适当装置,自然地与虚拟世界进行体验和交互作用。虚拟现实主要有三方面的含义:第一,虚拟现实是借助于计算机生成逼真的实体,“实体”是对于人的感觉(视、听、触、嗅)而言的;第二,用户可以通过人的自然技能与这个环境交互,“自然技能”是指人的头部转动、眼动、手势等其他人体的动作;第三,虚拟现实往往要借助于一些三维设备和传感设备来完成交互操作。虚拟现实系统示意图如图 5-11 所示。

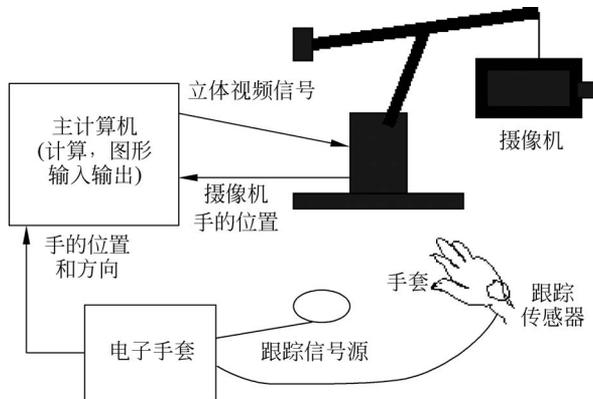


图 5-11 虚拟现实系统示意图(1992 年 Bryson)

5.6 信息资源安全技术标准

5.6.1 硬件安全技术

1. 物理安全的概念

物理安全是为保证信息系统的安全、可靠运行,降低或阻止人为或自然因素从物理层面对信息系统的保密性、完整性、可用性带来的安全威胁,从系统的角度采取的适当安全措施。

物理安全是以一定的方式运行在一些物理设备之上的、保障物理设备安全的第一道防线。因为物理安全会导致系统存在风险。例如,环境事故造成的整个系统毁灭;电源故障造成的设备断电以至操作系统引导失败或数据库信息丢失;设备被盗、被毁造成数据丢失或信息泄露;电磁辐射可能造成数据信息被窃取或偷阅;报警系统的设计不足或失灵可能造成的事故等。

设备安全技术主要是指保障构成信息网络的各种设备、网络线路、供电连接、媒体数据本身及其存储介质等安全的技术,主要包括设备的防盗、防电磁泄漏、防电磁干扰等,是对可用性的要求。所有的物理设备都是运行在一定的物理环境之中的。

物理环境安全是物理安全的最基本保障,是整个安全系统不可缺少和忽视的组成部分。环境安全技术主要是指保障信息网络所处环境安全的技术,主要技术规范是对场地和机房的约束,强调对于地震、水灾、火灾等自然灾害的预防措施,包括场地安全、防火、防水、防静电、防雷击、电磁防护、线路安全等。

2. 物理安全分类

(1) 信息系统物理安全。为了保证信息系统安全、可靠运行,确保信息系统在对信息进行采集、处理、传输、存储过程中,不致受到人为或自然因素的危害,而使信息丢失、泄露或破坏,对计算机设备、设施(包括机房建筑、供电、空调)、环境人员、系统等采取适当的安全措施。

(2) 设备物理安全。为保证信息系统的安全、可靠运行,降低或阻止人为或自然因素对硬件设备安全可靠运行带来的安全风险,对硬件设备及部件所采取的适当安全措施。

(3) 环境物理安全。为保证信息系统的安全、可靠运行所提供的安全运行环境,使信息系统得到物理上的严密保护,从而降低或避免各种安全风险。

(4) 介质物理安全。为保证信息系统的安全、可靠运行所提供的安全存储的介质,使信息系统的数据库得到物理上的保护,从而降低或避免数据存储的安全风险。

5.6.2 传输安全技术

1. 网络安全概述

计算机网络安全是指利用网络管理控制和技术措施,保证在一个网络环境里,数据的保密性、完整性及可使用性受到保护。计算机网络安全包括两个方面,即物理安全和逻辑安全。物理安全指系统设备及相关设施受到物理保护,免于破坏、丢失等。逻辑安全包括信息的完整性、保密性和可用性。计算机网络安全不仅包括组网的硬件、管理控制网络的软件,也包括共享的资源,快捷的网络服务,所以定义网络安全应考虑涵盖计算机网络所涉及的全部内容。

2. 虚拟专用网络技术

虚拟专用网络(Virtual Private Network, VPN)指的是在公用网络上建立专用网络的技术。之所以称为虚拟网,主要是因为整个 VPN 的任意两个节点之间的连接并没有传统专用网所需的端到端的物理链路,而是架构在公用网络服务商所提供的网络平台,如 Internet、ATM(异步传输模式)、Frame Relay(帧中继)等之上的逻辑网络,用户数据在逻辑链路中传输。它涵盖了跨共享网络或公共网络的封装、加密和身份验证链接的专用网络的扩展。VPN 主要采用了隧道技术、加解密技术、密钥管理技术和使用者与设备身份认证技术。

3. 网络隔离技术

1) 网络隔离技术概述

网络隔离技术是指两个或两个以上的计算机或网络在断开连接的基础上,实现信息交

换和资源共享。也就是说,通过网络隔离技术既可以使两个网络实现物理上的隔离,又能在安全的网络环境下进行数据交换。

面对新型网络攻击手段的出现和高安全度网络对安全的特殊需求,网络隔离技术应运而生。网络隔离技术的目标是确保隔离有害的攻击,在可信网络之外和保证可信网络内部信息不外泄的前提下,完成网间数据的安全交换。网络隔离技术是在原有安全技术的基础上发展起来的,它弥补了原有安全技术的不足,突出了自己的优势。

网络隔离技术的主要目标是将有害的网络安全威胁隔离开,以保障数据信息在可信网络内在进行安全交互。目前,一般的网络隔离技术都是以访问控制思想为策略、物理隔离为基础,并定义相关约束和规则来保障网络的安全强度。

2) 网络隔离技术方案

网络隔离技术主要有如下 5 种。

(1) 双机双网隔离方案,是指通过配置两台计算机来分别连接内网和外网环境,再利用移动存储设备来完成数据交互操作,然而这种技术方案会给后期系统维护带来诸多不便,同时还存在成本上升、占用资源等缺点,而且通常效率也无法达到用户的要求。

(2) 双硬盘隔离技术方案,通过在原有客户机上添加一块硬盘和隔离卡来实现内网和外网的物理隔离,并通过选择启动内网硬盘或外网硬盘来连接内网或外网网络。由于这种隔离技术方案需要多添加一块硬盘,所以对那些配置要求高的网络而言,就造成了成本浪费,同时频繁地关闭、启动硬盘容易造成硬盘的损坏。

(3) 单硬盘隔离技术方案,其原理是从物理层上将客户端的单个硬盘分割为公共和安全分区,并分别安装两套系统来实现内网和外网的隔离,这样就可具有较好的可扩展性,但是也存在数据是否安全界定困难、不能同时访问内外两个网络等缺陷。

(4) 集线器级隔离技术方案,其主要特征是在客户端只需使用一条网络线就可以部署内网和外网,然后通过远端切换器来选择连接内外双网,避免了客户端要用两条网络线来连接内外网络。

(5) 服务器端隔离技术方案,其关键是在物理上没有数据连通的内外网络下,如何快速、分时地处理和传递数据信息。该方案主要是通过采用复杂的软硬件技术手段来在服务器端实现数据信息过滤和传输任务,以达到隔离内外网的目的。

5.6.3 信息资源安全技术

1. 信息安全概述

信息安全指信息网络的硬件、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。

2. 信息安全的目标

信息安全的目标包括保密性、完整性、可用性、可控性和不可否认性。

(1) 保密性(confidentiality)是指阻止非授权的主体阅读信息。

(2) 完整性(integrity)是指防止信息被未经授权地篡改。它是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,形成虚假的信息将带来严重的后果。

(3) 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。

(4) 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(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。

3. 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,不仅要靠先进的技术,而且也要靠严格的安全管理、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,决定其所需要的安全服务种类,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。面对日趋严重的网络安全问题,必须建立与安全相关的法律、法规,使非法分子慑于法律,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。

4. 信息安全技术

(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。身份认证技术包括静态密码、动态密码(短信密码、动态令牌、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。

(3) 安全路由器。由于 WAN 连接需要专用的路由器设备,因而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。

(4) 安全服务器。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。

(5) 电子签证机构(CA)。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。

(6) 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。

(7) 入侵检测系统(IDS)。入侵检测作为传统保护机制(比如访问控制、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。

(8) 入侵防御系统(IPS)。入侵防御系统作为 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。

(9) 安全数据库。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。

(10) 安全操作系统。它是一种满足计算机系统安全技术需求的操作系统,具有自主访问控制、强制访问控制、标记、身份鉴别、客体重用、审计、数据完整性、隐蔽信道分析、可信路径、可信恢复等方面的安全要求。

(11) 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、审计、培训和风险评估、容灾备份、应急响应等工作。

(12) 数据加密。数据加密技术按技术上的实现分为在软件和硬件两方面。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四种。

5.6.4 信息资源安全标准

ISO 和 IEC 是世界范围的标准化组织,各国的相关标准化组织都是其成员,它们通过各技术委员会参与相关标准的制定。ISO/IEC 和西方一些国家开始发布和改版一系列信息安全管理标准,使安全管理标准进入了一个繁忙的改版期。ISO/IEC 联合技术委员会子委员会 27(ISO/IEC JTC1/SC27)是信息安全领域最权威和国际认可的标准化组织,它已经为信息安全保障领域发布了一系列国际标准和技術报告,最主要的标准是 ISO/IEC 13335、ISO/IEC 27000 系列等。

1. 我国信息安全管理

1) 标准组织发展

按照国务院授权,在国家质量监督检验检疫总局管理下,由国家标准化委员会统一管理全国标准化工作,下设有 255 个专业技术委员会。中国标准化工作实行统一管理 with 分工负责相结合的管理体制,有 88 个国务院有关行政主管部门和国务院授权的有关行业协会分工管理本部门、本行业的标准化工作,有 31 个省、直辖市政府、自治区有关行政主管部门分工管理本行政区域内本部门、本行业的标准化工作。

从 20 世纪 80 年代开始,本着积极采用国际标准的原则,转化了一批国际信息安全基础技术标准,制定了一批符合中国国情的信息安全标准,同时一些重点行业还颁布了一批信息安全的行业标准,为我国信息安全技术的发展做出了很大的贡献。

2) 标准化组织

我国有关部门十分关注信息安全标准化工作,早在 1984 年 7 月就组建了数据加密技术

委员会,1990年3月成立了中国信息协会(CIIA)。数据加密技术委员会于1997年8月改组为全国信息技术标准化委员会的信息安全技术分委员会,负责制定信息安全的国家标准。2002年4月成立了全国信息安全标准化技术委员会(TC260)。1999年3月31日经公安部科技局批准,公安部信息系统安全标准化技术委员会正式成立。2002年中国通信标准化协会网络与信息安全技术工作委员会成立。

2. 国内安全标准

1) 相关标准

CC(即中国国内 GB/T 18336—2015 和国际 ISO/IEC 15408—1999)和 ISO/IEC 27001—2005(《信息安全管理体系规范》)标准的共同点表现在以下四个方面:几个标准所涉及的范围从大的角度来说都是信息安全领域;几个标准对信息安全的定义相同,都是指对信息保密性、完整性和可用性的保护;几个标准对信息安全风险的定义基本相同,都是从资产、威胁、薄弱点和影响来考察风险;几个标准都针对不同的风险提出了相应的控制目标和控制措施。几个标准之间最主要的区别在于着眼点的不同。

公安部、国家保密局、国家密码管理局、国务院信息化工作办公室制定的《信息安全等级保护管理办法》由一系列标准文件组成:《信息安全等级保护管理办法》公通字[2007]43号、《计算机信息系统 安全保护等级划分准则》(GB 17859—1999)、《信息安全等级保护实施指南》《信息安全等级保护定级指南》《信息安全等级保护基本要求》《信息安全等级保护测评准则》《信息安全技术网络基础安全技术要求》(GB/T 20270—2006)、《信息安全技术 信息系统通用安全技术要求》(GB/T 20271—2006)、《信息安全技术 操作系统安全技术要求》(GB/T 20272—2019)和《信息安全技术 数据库管理系统安全技术要求》(GB/T 20273—2019)。

2020年,我国信息安全标准发布进入快车道。2020年我国发布的信息安全标准,共27项国家标准、42项行业标准;2021年我国发布的信息安全标准,共32项国家标准、24项行业标准;2022年我国发布的信息安全标准,共36项国家标准、17项行业标准;2023年我国发布的信息安全标准,共57项国家标准、86项行业标准。

2) 等级保护级别

《信息安全等级保护管理办法》中,“等级保护的实施与管理”第十二条明确指出,在信息系统建设过程中,运营、使用单位应当按照《计算机信息系统 安全保护等级划分准则》(GB 17859—1999)、《信息系统安全等级保护基本要求》等技术标准。等级保护共划分为五个级别,当前主要使用1~4级。不同等级的信息系统应具备不同的基本安全保护能力,其能力要求是逐级递增的。国家标准 GB 17859—1999《计算机信息系统 安全保护等级划分准则》将信息系统安全划分为五个等级。

第一级 用户自主保护级。本级的计算机信息系统可信计算基通过隔离用户与数据,使用户具备自主安全保护的能力。它具有多种形式的控制能力,对用户实施访问控制,即为用户提供可行的手段,保护用户和用户组信息,避免其他用户对数据的非法读写与破坏。

第二级 系统审计保护级。本级的计算机信息系统可信计算基实施了粒度更细的自主访问控制,它通过登录规程、审计安全性相关事件和隔离资源,使用户对自己的行为负责。

第三级 安全标记保护级。本级的计算机信息系统可信计算基具有系统审计保护级所有功能。此外,还提供有关安全策略模型、数据标记以及主体对客体强制访问控制的非形式

化描述；具有准确地标记输出信息的能力；消除通过测试发现的任何错误。

第四级 结构化保护级。本级的计算机信息系统可信计算基建立于一个明确定义的形式化安全策略模型之上,它要求将第三级系统中的自主和强制访问控制扩展到所有主体与客体。此外,还要考虑隐蔽通道。本级的计算机信息系统可信计算基必须结构化为关键保护元素和非关键保护元素。计算机信息系统可信计算基的接口也必须明确定义,使其设计与实现能经受更充分的测试和更完整的复审。加强了鉴别机制；支持系统管理员和操作员的职能；提供可信设施管理；增强了配置管理控制。系统具有相当的抗渗透能力。

第五级 访问验证保护级。本级的计算机信息系统可信计算基满足访问监控器需求。访问监控器仲裁主体对客体的全部访问。访问监控器本身是抗篡改的；必须足够小,能够分析和测试。为了满足访问监控器需求,计算机信息系统可信计算基在其构造时,排除那些对实施安全策略来说并非必要的代码；在设计 and 实现时,从系统工程角度将其复杂性降低到最小程度。支持安全管理员职能；扩充审计机制,当发生与安全相关的事件时发出信号；提供系统恢复机制。系统具有很高的抗渗透能力。

5.7 信息技术的几种资源化趋势

5.7.1 计算资源与算力

1. 背景

据 2023 年 8 月 19 日工业和信息化部信息,我国算力总规模 197EFLOPS,位居全球第二。全国数据中心机架总规模超过 760 万,存力总规模超过 1000EB(1 万亿 GB),国家枢纽节点间的网络单向时延降低到 20ms 以内,中国算力产业初具规模,计算类产品产量全球第一。

2. 算力的定义

国外算力的英文表述为 computing power 或 HashRate(哈希率)等,但孙凝晖院士认为算力的英文表述应该是 computility 更准确。

狭义算力定义是一台计算机具备的理论上最大每秒浮点运算次数(FLOPS)。除运算能力外,还应该包括数据存储与访问能力、与外界的数据交换能力、数据显示能力等。广义上,算力是计算机设备或计算/数据中心处理信息的能力,是计算机硬件和软件及网络共同执行计算的能力。

计算能力的度量,一般与被处理的信息类型有关。例如,在高性能计算中用每秒双精度浮点计算来度量其算力；在人工智能场景中用单精度、半精度或整数来度量其算力；在比特币中用挖矿机每秒能做多步哈希(hash)碰撞来度量其算力,挖矿算力即计算哈希函数输出的速度；在高通量场景中,度量算力的标准是每秒处理的字节数。

3. 计算资源

计算资源(resource on the computation)是计算复杂性理论的一个术语。狭义计算资源指计算机程序运行时所需的 CPU 资源、内存资源、硬盘资源和网络资源。实际上这是一种硬件资源的定义。广义计算资源包括硬件、软件、网络、数据,以及其他相关的支撑环境,甚至包括操作人员。计算资源可以看成是一个信息系统,各部分相互联系、相互依赖、相互影

响、相互作用,并以系统整体方式发挥作用。软件资源包括操作系统、应用软件、系统工具等。数据资源指存储在计算设备中的各种信息,如文档与超文本、图形与图像、音频与视频、非结构数据等。进入 21 世纪,大数据、人工智能和算力将计算资源的外延内涵进行了扩展。其中,智能算法对软件资源进行了扩展。算力将硬件资源进行了扩展,其不仅包括超级计算机、云计算,甚至量子计算机和生物计算等内容。

4. 大数据、智能算法与算力资源三者融合

大数据是人工智能算法的“燃料”。没有大数据,人工智能算法无法“燃烧”。而算力是人工智能算法运行的“物质基础”。算力与大数据和人工智能密不可分。三者相辅相成,协调发展。就目前的形势看,算力的发展相对滞后,因为 ChatGPT-4 迭代一次,耗时较长,开销太大。

5. 国家算力资源

21 世纪进入数字经济时代,算力已成为综合国力的核心指标。在新一代信息技术的加持下,融合国力已经成为综合国力的“升级版”。该视角下国家算力可分解为 5 方面的能力:一是计算速度,包括芯片、服务器、计算机、超算等都能反映这方面的能力。二是算法能力,比如大模型的规模及其处理数据的能力。三是大数据存储量,即一个国家拥有或可支配的数据总量。四是通信能力,包括 5G 基站的数量、通信的速度、延滞、带宽、可靠性、能耗等。五是云计算服务能力,包括数据处理中心服务器的数量。全球大国算力竞争已把计算速度、计算方法、通信能力、存储能力、数据总量 5 个方面列为其争夺的“制高点”。

5.7.2 边缘计算与雾计算

资源视角,边缘计算与雾计算均可以看作一种计算资源,两者强调将计算资源迁移到离用户更近的地方进行计算服务。

1. 边缘计算

1) 边缘计算的定义

边缘计算是一种致力于使计算尽可能靠近数据源,以减少延迟和带宽使用的网络理念。这样使云端计算量减少,并将进程迁移到本地或靠近客户端。将计算放到网络边缘可以最大程度地减少客户端和服务器之间的“长”距离通信。

边缘计算指在靠近客户或数据源头的一侧,采用网络、计算、存储、应用核心能力为一体的开放平台,就近提供端服务。其应用程序在边缘侧发起,产生更快的网络服务响应,满足实时业务、应用智能、安全隐私保护等方面的基本需求。

2) 边缘计算的优势

(1) 节省成本。其有助于最大程度地减少带宽使用量和服务器资源消耗。因为带宽和云资源是有限的,且花费较大。

(2) 减少延迟。将进程移至边缘的一个重要原因是减少服务延迟,因为设备每次需要与某处的远程服务器通信时均会花费时间。

(3) 增加新功能。边缘计算可以提供以前无法提供的新功能。例如可以使用边缘计算在网络边缘处理和分析数据,使得实时处理成为可能。

2. 雾计算

1) 雾计算概述

雾计算(fog computing)是云计算概念的延伸。在雾计算模式中数据、数据处理和应用

程序集中在网络边缘的设备中,而不是几乎全部保存在云中。雾计算没有很强的计算能力,其只有一些较弱的,零散的或简单的计算设备。

雾计算 2011 年由哥伦比亚大学斯特尔佛教授首先提出。其目的是利用“雾”来阻挡黑客入侵,后来思科公司采用,并赋予雾计算新的含义。雾计算是一种面向物联网的分布式计算基础设施,可将计算能力和数据分析应用扩展至网络“边缘”,它使客户能够在本地分析和 管理数据,从而通过连接获得即时服务,如图 5-12 所示。

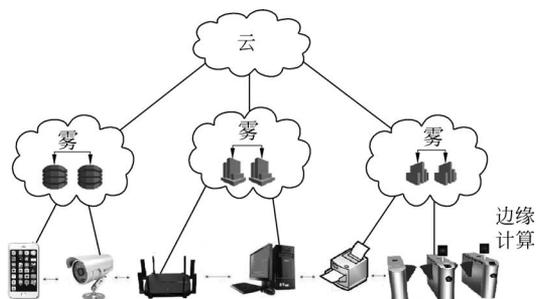


图 5-12 云计算与雾计算

2) 雾计算的特点

雾计算主要使用的是网络边缘的设备,数据传递时延很低。雾计算可以处理地理分布“较远”的计算,比如带有大量网络节点的大规模传感器网络。雾计算移动性好,手机和其他移动设备可以互相之间直接通信,信号不必传到云端甚至基站去绕一圈,具有较好的移动性。

雾计算不需要功能强大的服务器,它由性能较弱且分散的各种计算设备组成,雾计算是介于云计算和个人计算之间的技术,属于半虚拟化的服务计算架构模型,强调数量,不管单个计算节点能力多么弱都要发挥作用。与云计算相比,雾计算所采用的架构属于分布式边缘网络结构。雾计算将数据、数据处理和应用程序集中在网络边缘的设备中,而不像云计算那样将几乎全部计算资源都放在云中,雾计算数据存储及处理更依赖本地设备,而非服务器。雾计算是新一代分布式计算,符合互联网的“去中心化”特征。

5.7.3 信息技术资源池化

1. 池化技术

池化技术就是将对象放入池子,使用时从池中取,用完之后再交还给池子管理。通过优化资源分配的效率,达到性能的调优。池可以理解为由可重复使用的对象集合。使用池的原因很简单,即某些对象的创建成本很高,用完就抛弃的模式不利于提高应用程序的整体性能。使用池化技术可减少系统开销,提升系统性能。简单来说,就是通过复用提升性能。以下是目前比较常见的池化技术,作为技术资源,已在很多场合得到了应用。

2. 内存池

内存池提供了一个提高内存使用率的方法。在创建内存池的过程中,会预先分配足够大的内存形成一个初步的内存池,然后每次用户请求内存的时候就会返回内存池中的一块

空闲的内存,并将这块内存的标志置为已使用。当内存使用完毕释放的时候,也不是真正地调用 free 或 delete,而是把内存放回内存池,并将标志置为空闲。最后,待应用程序结束才将内存池销毁,将内存池中的每一块内存释放。

3. 存储池

存储池可以理解成为协调本地存储设备(可能有多个磁盘)的整体功能,并把这些设备的空间进行整合、综合分配。存储池可以帮助用户节省时间并最大限度地降低磁盘资源的读取和写入时长,节约损耗量。存储池可以减少磁盘空间碎片闲置,并提高硬盘组的利用率,避免事故灾难和系统故障,确保数据的可用性。

4. 应用池

应用池也称应用程序池,它是微软的一个全新概念,应用程序池是将一个或多个应用程序链接到一个或多个工作进程集合的配置。因为应用程序池中的应用程序与其他应用程序被工作进程边界分隔,所以某个应用程序池中的应用程序不会受到其他应用程序池中应用程序所产生的问题的影响。

5. 数据池

数据池主要用于数据暂留,比如大数据量的前处理、缓存、备查,区别于业务流程数据和用于分析的数据中台,它是业务处理系统的辅助数据库。当数据架构中设置了数据池组件,其架构就有了弹性,数据处理的方案会增加。

6. 线程池

线程池的原理类似于操作系统中缓冲区的概念。线程池设置时会预先启动若干数量的线程,这些线程一般处于睡眠状态。当客户端有新的请求时,就会唤醒线程池中的某一个睡眠的线程,让其处理客户端的这个请求,当处理完这个请求之后,线程又处于睡眠的状态。线程池可以提升程序的性能。线程池除了简单地复用已存在的线程,可以控制并发之外,还提供了对线程更丰富的控制。

7. 计算池

计算池可以为群集操作提供计算资源。在 SQL Server Pod 上运行的节点中,计算池可为大数据集群提供横向扩展计算资源。计算池由一个或多个计算 Pod 组成,SQL Server 主实例负责协调 Pod 自动创建和管理过程,每个 Pod 包含一组基本服务和一个 SQL Server 数据库引擎的实例。

8. 云计算资源池

云计算资源池是云计算平台的一种功能,其为一组集中管理的计算资源,包括计算节点、存储节点和网络节点等。云计算资源池可以为多个应用程序提供计算资源,这些应用程序可以根据客户的需求和资源使用情况动态选择和分配资源,从而提高资源的利用率、降低服务成本和提高可扩展性。在云计算中,资源池化指的是将多种计算和存储资源整合在一起,形成一个统一的资源库,以便进行动态分配和管理。资源池化能够实现资源的高度共享,提高资源使用率,简化资源管理,并为用户提供按需分配的灵活服务。

9. 对象池

对象池可以复用池中对象,避免对象频繁地创建和销毁。一方面减少了创建销毁时资

源的开销,同时也减少了垃圾回收器执行回收的次数。在 Java 中,连接和线程均为对象。对 Java 来说,连接池、线程池也是对象池。

10. 连接池

当服务器与外部服务器需要交互时,就需要创建连接。以数据库连接为例,每次连接都需要创建连接,执行完命令后再销毁,等下次执行再创建。数据库建立连接需要与数据库服务器通过网络请求建立通信,这让数据库耗费大量资源去响应连接请求。显然其会消耗计算资源、网络资源和数据库资源。创建连接过程比较耗时,会影响并发能力。

5.7.4 组件化、颗粒化与词条化

1. 组件化

模块化是比较古老的概念。在面向对象技术中,组件是模块概念的“升级版”。“知件”“数件”等是近年出现的新概念。

2008 年陆汝钐院士和金芝教授对“知件”展开了研究,其为一种知识模块。知件是独立的、计算机可操作的、商品化的、符合某种工业标准的、有完备文档的、可被某一类软件或硬件访问的知识模块。知件是一种标准的部件,更换知件就像更换计算机上的插件一样方便。作为一种知识单元,可以商品化出售。通过知件的形式,可以把软件中的知识含量分离出来,使软件和知件成为两种不同的研究对象和两种不同的商品,使硬件、软件和知件在 IT 产业中三足鼎立。从软件工程视角,知件开发过程可以进行工程管理,即知件工程。如果计算机生成的是规范化的、包装好的、商品化的知件,那么它就可以大规模生产。

数据时代,大数据、数据要素等概念的出现,预示着数据在经济社会中的广泛应用。为了使数据使用规范化,数据模块也称“数件”被提出。它是一种数据模块,类似“知件”。可以与硬件、软件和知件一起共同构筑计算机系统。无论是硬件、软件、知件和数件等,均可以作为一种技术资源使用。

2. 颗粒化与词条化

本书作者曾经在 2020 年第 6 期《中国图书馆学报》上发布文章介绍过文献“颗粒化”的概念。知识颗粒化是指馆藏文献的“模块化”“小块化”“单元化”等。文献“颗粒化”工作很重要,它可以根据读者需求,有针对性地提供“最小化”的文献知识服务,以缓解读者的阅读疲劳。为了区分馆藏文献“知识单元”的规模和大小,也可以给出知识单元的粒度定义,比如知识块、知识点、知识颗粒、知识子粒、知识微粒等方面的概念。

关于知识颗粒、知识元和知识基因的相关研究已经有学者展开。这方面研究有两个方向,一是知识元或知识基因的相关研究。知识元研究包括基于知识元的检索(也就是词条检索,“知网”已开始提供词条检索服务),其中包括知识元的抽取方法和知识元服务;二是知识颗粒的相关研究。比如档案文本信息“颗粒化”,以及细粒度知识标准体系的构建。人工智能抽取知识颗粒的方法是一种比较新的研究思路,涉及文本、图像、音视、视频或多媒体形式文献的颗粒化。

无论是组件化,还是颗粒化与词条化,都为计算资源的粒度研究和使用的指明了一个新的思路和发展方向。

练习题

一、名词解释

1. 数据库
2. 联邦数据库
3. 数据仓库
4. 计算机网络
5. 分布式计算
6. 并行计算
7. 云计算
8. 信息安全
9. 数据中心
10. 超级计算机
11. 量子计算机
12. 算力

二、简答题

1. 简述数据采集系统。
2. 简述几种自动识别技术。
3. 简述 5G 和 6G 技术。
4. 池化技术的目的是什么?
5. 简述几种显示技术。
6. 简述信息安全等级标准。
7. 介绍虚拟化技术。
8. 简述几种打印机的特点。

参考文献

- [1] 张凯. 计算机导论[M]. 北京: 清华大学出版社, 2010.
- [2] 张凯. 信息安全导论[M]. 北京: 清华大学出版社, 2018.
- [3] 张凯. 物联网导论[M]. 北京: 清华大学出版社, 2012.
- [4] 段孝国. 分布式计算技术介绍[J]. 电脑知识与技术, 2011, 7(22): 5463-5465.
- [5] 刘师范. 并行计算方法研究与应用[J]. 数字技术与应用, 2014(01): 109-110.
- [6] 文娟. 面向大数据处理的并行计算模型及性能优化[J]. 中国高新技术企业, 2016(06): 33-34.
- [7] 刘芬. 刀片服务器的发展[J]. 科技视界, 2014(24): 97.
- [8] 王利. 集群计算[J]. 计算机教育, 2004(12): 48-49.
- [9] 马天蔚. 刀片化+虚拟化——惠普谈服务器技术发展趋势[J]. 每周电脑报, 2004(13): 48.
- [10] 欧志亮, 林雄光. 云计算中的虚拟化关键技术应用[J]. 廊坊师范学院学报(自然科学版), 2019, 19(2): 30-33.