

第 1 章 计算机基础知识



思政课堂

先来看一则消息:2020年12月4日,中国科学技术大学宣布该校潘建伟等人成功构建76个光子的量子计算原型机“九章”,求解数学算法高斯玻色取样只需200秒,而目前世界最快的超级计算机要用6亿年。这一突破使我国成为全球第二个实现“量子优越性”的国家。

请你对上述消息进行思考,阐述科技发展对我国在世界上的地位有着怎样的意义,并讨论当我国科技力量处于世界领先地位时,对我国国民会产生怎样的影响。请将你的思考写在下面。

1.1 计算机基础

1.1.1 计算机的诞生与发展

1. 计算机的诞生

世界上第一台电子数字式计算机于1946年2月15日在美国宾夕法尼亚大学正式投入运行,它的名称为ENIAC(埃尼阿克),是电子数值积分计算机(the electronic numerical intergrator and computer)的英文简称。美国国防部用它来进行炮弹弹道轨迹计算。它使用了17 468个真空电子管,耗电174 kW,占地170 m²,重达30 t,每秒钟可进行5 000次加法运算。虽然它的功能远比不上今天一台普通的微型计算机,但在当时它已是运算速度的绝对冠军,并且其运算的精确度和准确度也是史无前例的。

ENIAC奠定了电子计算机的发展基础,开辟了计算机科学技术的新纪元。有人将其称为人类第三次工业革命开始的标志。

ENIAC诞生后,数学家冯·诺依曼提出了重大的改进理论——存储程序原理,主要内容有3点:其一是电子计算机应该以二进制为运算基础;其二是电子计算机应采用存储程序方式工作;其三是进一步明确指出了整个计算机的结构应由5个部分组成,即运算器、控制器、存储器、输入装置和输出装置。冯·诺依曼存储程序原理的提出,解决了计算机运算自动化和速度配合问题,对后来计算机的发展起到了决定性的作用。直至今日,绝大部分的计算机还是采用冯·诺依曼的方式工作。

早期按冯·诺依曼体系结构设计的计算机有如下一些。

(1)电子离散变量自动计算机(electronic discrete variable automatic computer,EDVAC)。它是第一个按照存储程序原理设计的计算机,该机1952年投入运行,用于核武器理论计算。

(2)电子延迟存储自动计算机(electronic delay storage automatic calculator,EDSAC)。它是第一次实现大型存储程序的计算机,1949年投入运行。

(3)通用自动计算机(universal automatic computer,UNIVAC)。1951年,该机作为商品计算机投入使用,开创了用于数据处理的计算机新时代。

1965年,中国科学院计算技术研究所研制成功了我国第一台大型晶体管计算机109乙机;对109乙机加以改进,两年后又推出了109丙机,在中国“两弹”试制中发挥了重要作用,被用户誉为“功勋机”。

2. 计算机的发展

计算机的发展到目前为止共经历了如下4个时代。

(1)从1946年到1959年,这段时期称为电子管计算机时代。第一代计算机的内部元件使用的是电子管。由于一台计算机需要几千个电子管,每个电子管都会散发大量的热量,因此,如何散热是一个令人头疼的问题。电子管的寿命最长只有3 000小时,计算机运行时常常发生由于电子管被烧坏而使计算机死机的现象。第一代计算机主要用于科学研究和工程计算。

(2)从1960年到1964年,由于在计算机中采用了比电子管更先进的晶体管,所以将这段时期称为晶体管计算机时代。晶体管比电子管小得多,不需要暖机时间,消耗能量较少,处理更迅速、更可靠。第二代计算机的程序语言从机器语言发展到汇编语言。接着,高级语言 FORTRAN 和 COBOL 相继被开发出来并被广泛使用。这时,开始使用磁盘和磁带作为辅助存储器。第二代计算机的体积和价格都下降了,使用的人也多起来,计算机工业迅速发展。第二代计算机主要用于商业、大学教学和政府机关。

(3)从1965年到1970年,集成电路(integrated circuit, IC)被应用到计算机中,因此这段时期被称为中小规模集成电路计算机时代。集成电路是集成在晶片上的一个完整的电子电路,这个晶片比指甲还小,却包含几千个晶体管元件。第三代计算机的特点是体积更小、价格更低、可靠性更高、计算速度更快。第三代计算机的代表是 IBM 公司花了 50 亿美元开发的 IBM 360 系列。这一阶段最主要的发展是在第三代计算机中出现了操作系统,代表着计算机系统的形成和完善。

(4)从1971年到现在,被称为大规模集成电路计算机时代。20世纪80年代以来,ALU 和控制单元(二者合成中央处理器,即 CPU)逐渐被整合到一块集成电路上,称为微处理器。这类计算机的工作模式十分直观:在一个时钟周期内,计算机先从存储器中获取指令和数据,然后执行指令,存储数据,再获取下一条指令。这个过程被反复执行,直至得到一个终止指令。

计算机微型处理器(CPU)以晶体管为基本元件,随着处理器的不断完善和更新换代速度的加快,计算机结构和元件也会发生很大的变化。光电技术、量子技术和生物技术的发展,对新型计算机的发展具有极大的推动作用。

第四代计算机使用的元件依然是集成电路,不过,这种集成电路已经大大改善,它包含几十万个到上百万个晶体管,人们称之为大规模集成电路(large scale integrated circuit, LSI)和超大规模集成电路(very large scale integrated circuit, VLSI)。1975年,美国 IBM 公司推出了个人计算机(personal computer, PC),从此,人们对计算机不再陌生,计算机开始深入人类生活的各个方面。在第四代出现了 CPU,使得计算机普及成为现实,计算机开始在各个领域得到应用。

1.1.2 计算机的分类、特点与应用

1. 计算机的分类

计算机的分类方法很多,可从不同角度对其进行分类,如可以按照计算机的体积、速度、处理能力等特性分类,也可根据使用范围、使用方式分类。通常情况下,计算机采用如下 3 种分类标准。

(1)按发展趋势分类。随着科技的进步,各种计算机技术、网络技术的飞速发展,计算机的发展已经进入一个快速而又崭新的时代,计算机已经从功能单一、体积较大发展到了功能复杂、体积微小、资源网络化等。计算机的未来充满了变数,性能的大幅度提高是毋庸置疑的,而实现性能的飞跃却有多种途径。不过性能的大幅提升并不是计算机发展的唯一路线,计算机的发展还应当变得越来越人性化,同时也要注意环保等。

计算机从出现至今,经历了机器语言、程序语言、简单操作系统和 Linux、Mac OS、BSD、Windows 等现代操作系统四代,运行速度也得到了极大的提升,第四代计算机的运算速度已

经达到几十亿次每秒。计算机也由原来的仅供军事科研使用发展到人人拥有,计算机强大的应用功能产生了巨大的市场需要,未来计算机性能应向着巨型化、微型化、多媒体化、网络化、人工智能化的方向发展。

①巨型化。巨型化是指为了适应尖端科学技术的需要,发展高速度、大存储容量和功能强大的超级计算机。随着人们对计算机的依赖性越来越强,特别是在军事和科研教育方面对计算机的存储空间和运行速度等要求会越来越高。此外计算机的功能更加多元化。

②微型化。随着微型处理器的出现,计算机中开始使用微型处理器,使计算机体积缩小了、成本降低了。另外,软件行业的飞速发展提高了计算机内部操作系统的便捷度,计算机外部设备也趋于完善。计算机理论和技术上的不断完善促使微型计算机很快渗透到全社会的各个行业和部门中,并成为人们生活和学习的必需品。计算机的体积不断缩小,台式计算机、笔记本电脑、掌上电脑、平板电脑体积逐步微型化,为人们提供便捷的服务。因此,未来计算机仍会不断趋于微型化,体积将越来越小。

③多媒体化。传统的计算机处理的信息主要是字符和数字。事实上,人们更习惯的是图片、文字、声音、图像等多种形式的多媒体信息。多媒体技术可以集图形、图像、音频、视频、文字为一体,使信息处理的对象和内容更加接近真实世界。

④网络化。互联网将世界各地的计算机连接在一起,人类社会进入互联网时代。计算机网络化彻底改变了人类世界,人们通过互联网进行沟通、交流(QQ、微博、微信等),共享教育资源(文献查阅、远程教育等)、共享信息查阅(百度、谷歌)等,特别是无线网络的出现,极大地提高了人们使用网络的便捷性,未来计算机将会进一步向网络化方面发展。

⑤人工智能化。计算机人工智能化是未来发展的必然趋势。现代计算机具有强大的功能和极快的运行速度,但与人脑相比,其智能化和逻辑能力仍有待提高。人类在不断探索如何让计算机能够更好地反映人类思维,使计算机能够具有人类的逻辑思维判断能力,可以通过思考与人类进行沟通交流,抛弃以往的通过编码程序来运行计算机的方法,直接对计算机发出指令。

(2)按性能规模分类。计算机性能规模主要指计算机的体积、速度、处理能力等特性。按性能规模的不同,计算机可分为巨型机、大型机、中小型机、微型机和工作站。

①巨型机。巨型机的特点是运算速度快、存储容量大,每秒可执行几十亿条指令,可容纳上百个用户同时使用,可同时完成多项任务。研究巨型机是现代科学技术,尤其是国防尖端技术发展的需要。目前世界上只有少数几个国家能生产巨型机。我国自主研发的天河二号(每秒运算 3.39 亿亿次以上)、神威·太湖之光(每秒运算 9.3 亿亿次)都是巨型机。

②大型机。大型机的特点表现在通用性强、具有很强的综合处理能力、性能覆盖面广等方面,每秒可执行几亿条指令,主要应用在公司、银行、政府部门、社会管理机构和制造厂家等,通常称大型机为企业计算机。大型机在未来将被赋予更多的使命,如大型事务处理、企业内部的信息管理与安全保护、科学计算等。

③中小型机。中小型机是介于大型机和微型机之间的一种机型,每秒可执行千万条指令。中小型机规模小、结构简单、设计周期短,便于及时采用先进工艺。这类机器可靠性高,对运行环境要求低,易于操作且便于维护。中小型机符合部门性的要求,为中小型企事业单位所常用。

④微型机。微型机又称个人计算机,它是日常生活中使用最多、最普遍的计算机,具有

价格低廉、性能强、体积小、功耗低等特点,每秒可执行百万条指令。现在微型机已进入千家万户,成为人们工作、生活的重要工具。微型机可分为台式机和便携机两类,便携机又分为笔记本电脑和个人数字助理(俗称掌上电脑)两种。

⑤工作站。工作站是一种高档的微机系统,它具有较高的运算速度,具有多任务、多用户功能,且兼具微型机的操作便利性和良好的人机界面,可以连接到多种输入/输出设备,具有易于联网、处理功能强等特点。工作站的应用领域已从最初的计算机辅助设计扩展到商业、金融、办公领域,并充当网络服务器的角色。

(3)按功能和用途分类。按功能和用途,计算机可分为通用计算机和专用计算机。通用计算机具有功能强、兼容性强、应用面广、操作方便等优点,通常使用的计算机都是通用计算机。专用计算机一般功能单一、操作复杂,用于完成特定的工作任务。

2. 计算机的特点

计算机的特点可以归纳为以下7点,其中前5点为信息技术界共识,属于计算机的基本特点,后两点尚有学者未将其归纳进来。

(1)支持人机交互。计算机具有多种输入/输出设备,配上适当的软件后,可支持用户方便地进行人机交互。以广泛使用的鼠标为例,当用户手握鼠标时,只需将手指轻轻一点,计算机便可完成某种操作功能,真可谓“得心应手,心想事成”。当这种交互性与声像技术结合形成多媒体用户界面时,更可使用户的操作自然、方便、丰富多彩。

(2)数据处理速度快。计算机由电子器件构成,具有很高的处理速度。目前世界上最快的计算机每秒可运算千万亿次,普通PC每秒也可处理上百万条指令。这不仅极大地提高了工作效率,而且使时限性强的复杂处理可在限定的时间内完成。

(3)数据“记忆”能力强。计算机的存储器类似于人的大脑,可以记忆大量的数据和计算机程序,随时提供信息查询、处理等服务。早期的计算机,由于存储容量小,存储器常常成为限制计算机应用的“瓶颈”。今天,一台普通的PC内存可达8GB或更高,能支持运行大多数窗口应用程序。当然,有些数据量特别大的应用,如大型情报检索、卫星图像处理等,仍需要使用具有更大存储容量的计算机,如大型计算机和巨型计算机。

(4)具有逻辑判断能力。逻辑判断是计算机的又一重要特点,是计算机能实现信息处理自动化的重要原因。冯·诺依曼型计算机的基本思想,就是将程序预先存储在计算机中。在程序执行过程中,计算机根据上一步的处理结果,能运用逻辑判断能力自动决定下一步应该执行哪一条指令。这样,计算机的计算能力、逻辑判断能力和记忆能力三者相结合,使得计算机的能力远远超过任何一种工具而成为人类脑力延伸的有力助手。

(5)具有很高的计算精度。由于计算机采用二进制数字进行计算,因此可以用增加表示数字的设备和运用计算技巧等手段,使数值计算的精度越来越高,可根据需要获得千分之一到几百万分之一的计算精度。

(6)具有自动控制能力。计算机是由程序控制其操作过程的。只要根据应用的需要,事先编制好程序并输入计算机,计算机就能自动、连续地工作,完成预定的处理任务。计算机中可以存储大量的程序和数据。存储程序是计算机工作的一个重要原则,这是计算机能自动处理的基础。

(7)通用性强。计算机能够在各行各业得到广泛的应用,原因之一就是其具有很强的通用性。计算机可以将任何复杂的信息处理任务分解成一系列的基本算术运算和逻辑运算,

反映在计算机的指令操作中。按照各种规律要求的先后次序把它们组织成各种不同的程序,存入存储器中。在计算机的工作过程中,这种存储指挥和控制计算机进行自动、快速的信息处理,并且十分灵活、方便、易于变更,这就使计算机具有极大的通用性。

3. 计算机的应用

计算机的应用领域已渗透到社会的各行各业,正在改变着人们传统的工作、学习和生活方式,推动着社会的发展。从应用领域角度分析,大多数教科书上的说法都不完全一致,主要是对领域的划分不一致,但一致认可的是科学计算。(数值计算)、过程监控、信息管理(数据处理)、辅助技术(计算机辅助设计与制造)、人工智能(智能模拟)等方面。

(1)科学计算(数值计算)。早期的计算机主要用于科学计算。科学计算仍然是计算机应用的一个重要领域,如高能物理、工程设计、地震预测、气象预报、航天技术等。由于计算机具有高运算速度和精度及逻辑判断能力,因此出现了计算力学、计算物理、计算化学、生物控制论等新的学科。

(2)过程监控。利用计算机对工业生产过程中的某些信号自动进行检测,并把检测到的数据存入计算机,再根据需要对这些数据进行处理,这样的系统称为计算机检测系统。特别是仪器仪表引进计算机技术后所构成的智能化仪器仪表,将工业自动化推向一个更高的水平。

(3)信息管理(数据处理)。信息管理是计算机应用最广泛的一个领域。利用计算机来加工、管理与操作任何形式的数据资料,如企业管理、物资管理、报表统计、账目计算、信息情报检索等。国内许多机构纷纷建设自己的管理信息系统(MIS);生产企业也开始采用制造资源规划软件(MRP),商业流通领域则逐步使用电子信息交换系统(EDI),即所谓无纸贸易。

(4)辅助技术(计算机辅助设计与制造)。计算机辅助技术包括以下几种。

- ①计算机辅助设计(computer aided design,CAD)。
- ②计算机辅助制造(computer aided manufacturing,CAM)。
- ③计算机辅助教学(computer aided instruction,CAI)。
- ④计算机辅助测试(computer aided testing,CAT)。
- ⑤计算机集成制造系统(computer integreted manufacturing system,CIMS)。

(5)人工智能(智能模拟)。开发一些具有人类某些智能的应用系统,用计算机来模拟人的思维判断、推理等智能活动,使计算机具有自学习适应和逻辑推理的功能,如计算机推理、智能学习系统、专家系统、机器人等,帮助人们学习和完成某些推理工作。

(6)语言翻译。1947年,美国数学家、工程师沃伦·韦弗与英国物理学家、工程师安德鲁·布思提出了以计算机进行翻译(简称机译)的设想,机译从此登上历史舞台,并走过了一条曲折而漫长的发展道路。机译分为文字机译和语音机译。机译消除了不同文字和语言间的隔阂,堪称高科技造福人类之举。但机译的质量长期以来一直是个问题,尤其是译文质量,离理想目标仍相差甚远。中国数学家、语言学家周海中教授认为,在人类尚未明了大脑是如何进行语言的模糊识别和逻辑判断的情况下,机译要想达到信、达、雅的程度是不可能的。这一观点恐怕道出了制约译文质量的“瓶颈”所在。

1.2 计算机系统的基本组成及主要技术指标

1.2.1 计算机系统的基本组成

1. 计算机硬件系统

冯·诺依曼模型决定了计算机的硬件系统由运算器(arithmetic and logic unit, ALU)、控制器(control unit, CU)、存储器(memory)、输入设备(input devices)和输出设备(output devices)5个部分组成。

(1)运算器。运算器是计算机处理数据形成信息的加工厂,它的主要功能是对二进制数码进行算术运算或逻辑运算。

(2)控制器。控制器是计算机的神经中枢,由它指挥全机各个部件自动、协调地工作。控制器的主要部件包括指令寄存器、译码器、时序节拍发生器、操作控制部件和指令计数器(也称程序计数器)等。

控制器的基本功能是根据指令计数器中指定的地址从内存取出一条指令,对其操作码进行译码,再由操作控制部件有序地控制各部件完成操作码规定的功能。

在微型计算机中,运算器和控制器集成在一起构成了中央处理器,它是计算机系统的核心,能够处理的数据位数是CPU的一个重要性能指标。

(3)存储器。存储器是计算机的记忆装置,用来存储当前要执行的程序、数据及结果。存储器应该具备存数和取数的功能。

存储器分为内存储器(简称内存)和外存储器(简称外存)两类。

(4)输入设备。输入设备是用来向计算机输入命令、程序、数据、文本等信息的,目前比较常见的输入设备是键盘和鼠标。

(5)输出设备。输出设备的主要功能是将计算机处理后的各种内部格式的信息转换为人们所能识别的形式并表达出来。目前,较常见的输出设备有打印机和显示器。

输入/输出设备简称I/O设备,有时也称为外部设备,是计算机系统中不可或缺的组成部分,是计算机与外部世界进行信息交换的中介,是人与计算机联系的桥梁。

2. 计算机软件系统

计算机软件可以分为系统软件(system software)和应用软件(application software)两类。

(1)系统软件。系统软件主要包括操作系统(operating system, OS)、语言处理系统、系统性能检测和实用工具软件等。其中最主要的是操作系统,它提供了一个软件运行的环境,如在计算机中使用最广泛的是微软公司的Windows系列操作系统。

①操作系统的功能。操作系统的功能主要有如下几个。

a. 处理机管理。处理机就是指CPU。如何管理好CPU、提高CPU的使用效率是操作系统的核心任务。

b. 内存管理。内存管理主要包括内存空间的分配、保护和扩充。在内存管理中,操作系

还通过虚拟存储器技术为用户提供一个比实际内存大得多的虚拟内存,以解决物理内存空间不足的问题。

c. 设备管理。设备管理的任务是根据预定的分配策略,将设备接口及外设分配给请求输入/输出的程序,并启动设备完成输入/输出操作。

d. 信息管理。在计算机的外存上存储着大量信息(包括程序和数据),如何组织和管理好这些信息,方便用户使用,就是操作系统信息管理的内容。

e. 用户接口。除了上述 4 项对计算机软硬件资源的管理外,操作系统的另一个主要功能就是为用户提供一个友好的用户接口。用户接口有以下两个层次。

一个是程序级的接口,即系统提供了一组系统调用供用户在编程时使用。通过这些调用,用户可以在程序中访问系统的一些资源,或要求操作系统完成一些预定功能。

另一个是作业级接口,也就是操作系统用户界面,如 Windows 界面、DOS 界面、UNIX 的 shell 命令等都是这种接口的具体体现。

②操作系统的分类。按照操作系统的发展过程,可以将其分为以下 5 类。

a. 单用户操作系统(single user operating system)。单用户操作系统的主要特征是计算机系统内一次只能支持运行一个用户程序。这类系统的最大缺点是计算机系统的资源不能被充分利用。其典型代表是 DOS 操作系统。

b. 批处理操作系统(batch processing operating system)。批处理操作系统是 20 世纪 70 年代运行于大中型计算机上的操作系统,又被称为多任务操作系统。其典型代表是 IBM 的 DOS/VSE 系统。

c. 分时操作系统(time-sharing operating system)。分时操作系统具有如下特征:在一台计算机周围挂上若干台近程或远程终端,每个用户可以在各自终端上以交互方式控制作业运行。

d. 实时操作系统(real-time operating system)。实时操作系统按照其使用方式可以分为两类:一类是广泛用于钢铁、炼油、化工生产过程控制,武器制导等各个领域中的实时控制系统;另一类是广泛应用于飞机票、火车票自动订购系统,情报检索系统,银行业务系统,超级市场销售系统中的实时数据处理系统。

e. 网络操作系统(network operating system)。网络是将物理上分散的具有独立功能的多个计算机系统连接起来,通过网络协议在不同的计算机之间进行信息交换、网络管理、资源共享等,通信及系统安全等方面都是按照各自的标准协议进行开发的。用户可以突破地理条件的限制,方便地使用远程计算机的资源,提供网络通信资源共享功能的操作系统称为网络操作系统。

(2)应用软件。将为解决特定问题而与计算机本身关联不多或者说其使用与计算机硬件基本无关的软件统称为应用软件。

常用的应用软件有如下几类。

①办公软件。办公软件是日常办公需要的一些软件,一般包括文字处理软件、电子表格处理软件、演示文稿制作软件、个人数据库、个人信息管理软件等。常见的办公软件包括 Microsoft Office、金山公司的 WPS 等。

②多媒体处理软件。多媒体处理软件主要包括图形处理软件、图像处理软件、动画制作软件、音频处理软件、桌面排版软件等,如 Adobe 公司的 Photoshop 等。

③Internet 工具软件。随着计算机网络技术的发展和 Internet 的普及,涌现出许许多多基于 Internet 环境的应用软件,如 Web 服务软件、Web 浏览器、文件传送工具 FTP、远程访问工具 Telnet、下载工具迅雷等。

1.2.2 计算机的主要技术指标

计算机的性能涉及体系结构、软硬件配置、指令系统等多种因素,一般说来,主要有下列技术指标。

(1)字长。字长是指计算机运算部件一次能同时处理的二进制数据的位数,字长越长,则计算机的运算精度就越高。通常,字长为字节的整倍数(8 的整倍数)。

(2)时钟主频。时钟主频是指 CPU 的时钟频率,它的高低在一定程度上决定了计算机速度的高低。主频以 GHz 为单位,一般来说,主频越高,速度越快。

(3)运算速度。计算机的运算速度通常是指每秒钟所能执行的加法指令数目,常用百万次/秒(million instructions per second, MIPS)来表示,这个指标能更直观地反映出机器的速度。

(4)存储容量。存储容量分为内存容量和外存容量,这里主要指内存储器的容量。内存容量越大,机器所能运行的程序就越多,处理能力就越强。

(5)存取周期。内存储器的存取周期也是影响整个计算机系统性能的主要指标之一。简单地说,存取周期就是 CPU 从内存储器中存取数据所需的时间,目前,内存的存取周期为 7~70 ns。

此外,可靠性、可维护性、平均无故障时间和性价比也是计算机的技术指标。

1.3 计算机中信息的表示与存储

1.3.1 计算机中的数制

1. 进位计数制

数制也称计数制,是指用一组固定的符号和统一的规则来表示数值的方法。按进位的原则进行计数的方法称为进位计数制。例如,在十进位计数制中,是按照“逢十进一”的原则进行计数的。

常用进位计数制有十进制(decimal notation)、二进制(binary notation)、八进制(octal notation)、十六进制数(hexadecimal notation)。

计算机中数的书写规则如下:

二进制数的书写通常在数的右下方注上基数 2 或后面加 B 表示。

八进制数的书写通常在数的右下方注上基数 8 或后面加 O 表示。

十进制数的书写通常在数的右下方注上基数 10 或后面加 D 表示。一般约定 D 可省略。

十六进制数的书写通常在数的右下方注上基数 16 或后面加 H 表示。

2. 进位计数制的基数与位权

基数和位权是进位计数制的两个要素。

(1) 基数。基数就是进位计数制的每位数上可能有的数码的个数。例如，十进制数每位上的数码有 0、1、2、…、9 十个数码，所以基数为 10。

(2) 位权。位权是指一个数值每一位上的数字的权值的大小。例如，十进制数 4567 从低位到高位位权分别为 10^0 、 10^1 、 10^2 、 10^3 。因此 4567 按位权展开为

$$4567 = 4 \times 10^3 + 5 \times 10^2 + 6 \times 10^1 + 7 \times 10^0$$

(3) 数的位权表示。任何一种数制的数都可以表示成按位权展开的多项式之和。

例如，十进制数的 435.05 可表示为

$$435.05 = 4 \times 10^2 + 3 \times 10^1 + 5 \times 10^0 + 0 \times 10^{-1} + 5 \times 10^{-2}$$

位权表示法的特点是：每一项 = 某位上的数字 × 基数的若干幂次，而幂次的大小由该数字所在的位置决定。

3. 二进制数

计算机中为何采用二进制？因为二进制运算简单，电路简单可靠、容易实现，逻辑性强。

(1) 定义：按“逢二进一”的原则进行计数，称为二进制数，即每位上计满 2 时向高位进一。

(2) 特点：每个数的数位上只能是 0、1 两个数字；二进制数中的最大数字是 1，最小数字是 0；基数为 2。例如，10011010 与 00101011 是两个二进制数。

(3) 二进制数的位权表示如下。

$$(1101.101)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}$$

(4) 二进制数的运算规则。

① 加法运算。

a. $0 + 0 = 0$ 。

b. $0 + 1 = 1 + 0 = 1$ 。

c. $1 + 1 = 10$ 。

② 乘法运算。

a. $0 \times 0 = 0$ 。

b. $0 \times 1 = 1 \times 0 = 0$ 。

c. $1 \times 1 = 1$ 。

4. 八进制数

(1) 定义：按“逢八进一”的原则进行计数，称为八进制数，即每位上计满 8 时向高位进一。

(2) 特点：每个数的数位上只能是 0、1、2、3、4、5、6、7 八个数字；八进制数中的最大数字是 7，最小数字是 0；基数为 8。例如， $(1347)_8$ 与 $(62435)_8$ 是两个八进制数。

(3) 八进制数的位权表示如下。

$$(107.13)_8 = 1 \times 8^2 + 0 \times 8^1 + 7 \times 8^0 + 1 \times 8^{-1} + 3 \times 8^{-2}$$

5. 十六进制数

(1) 定义：按“逢十六进一”的原则进行计数，称为十六进制数，即每位上计满 16 时向高位进一。

(2)特点:每个数的数位上只能是0、1、2、3、4、5、6、7、8、9、A、B、C、D、E、F十六个数码;十六进制数中的最大数字是F,即15,最小数字是0;基数为16。例如, $(109)_{16}$ 与 $(2FDE)_{16}$ 是两个十六进制数。

(3)十六进制数的位权表示如下。

$$(109.13)_{16} = 1 \times 16^2 + 0 \times 16^1 + 9 \times 16^0 + 1 \times 16^{-1} + 3 \times 16^{-2}$$

$$(2FDE)_{16} = 2 \times 16^3 + 15 \times 16^2 + 13 \times 16^1 + 14 \times 16^0$$

6. 常用计数制间的对应关系

表1-1列出了常用计数制间的对应关系。

表 1-1 常用计数制间的对应关系

十进制	二进制	八进制	十六进制
1	1	1	1
2	10	2	2
4	100	4	4
8	1000	10	8
10	1010	12	A
15	1111	17	F
16	10000	20	10

7. 数制间的转换

将数由一种数制转换成另一种数制称为数制间的转换。因为日常生活中经常使用的是十进制数,而在计算机中采用的则是二进制数,所以在使用计算机时就必须把输入的十进制数换算成计算机所能够接受的二进制数。计算机在运行结束后,再把二进制数换算成人们习惯的十进制数输出。这两个换算过程完全由计算机自动完成。

(1)十进制数转换成非十进制数。

①十进制整数转换成非十进制整数。十进制整数转换成非十进制整数采用“余数法”,即除基数取余数。将十进制整数逐次用任意非十制数的基数去除,一直到商是0为止,然后将所得到的余数由下而上排列即可。

【例 1-1】 将 $(213)_{10}$ 转换成二进制数。

解:计算过程如下。

$2 \overline{)213}$	(K_0)	1	↑	低
$2 \overline{)106}$	(K_1)	0		
$2 \overline{)53}$	(K_2)	1		
$2 \overline{)26}$	(K_3)	0		
$2 \overline{)13}$	(K_4)	1		
$2 \overline{)6}$	(K_5)	0		
$2 \overline{)3}$	(K_6)	1		
$2 \overline{)1}$	(K_7)	1		高

计算结果为 $(213)_{10} = (11010101)_2$ 。

②十进制小数转换成非十进制小数。十进制小数转换成非十进制小数采用“进位法”，即乘基数取整数。将十进制小数不断地用其他进制的基数去乘，直到小数的当前值等于 0 或满足要求的精度为止，最后得到的积的整数部分由上而下排列即为所求。

【例 1-2】 将十进制数 $(0.514)_{10}$ 转换成相应的二进制数。

解：计算过程如下。

0.514		
×2		
1.0281	(K_1)
×2		
0.0560	(K_2)
×2		
0.1120	(K_3)
×2		
0.2240	(K_4)
×2		
0.4480	(K_5)
×2		
0.8960	(K_6)
×2		
1.7921	(K_7)

高

↓

低

计算结果为 $(0.514)_{10} = (0.1000001)_2$ 。

(2)非十进制数转换成十进制数。非十进制数转换成十进制数采用“位权法”，即把各非十进制数按位权展开，然后求和。

【例 1-3】 将 $(1101.101)_2$ 转换成十进制数。

$$\begin{aligned}
 \text{解：} (1101.101)_2 &= 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \\
 &= 8 + 4 + 1 + 0.5 + 0.125 \\
 &= (13.625)_{10}
 \end{aligned}$$

(3)二、八、十进制数之间的转换。

①二进制数与八进制数之间的转换方法。

a. 把二进制数转换为八进制数时，按“三位并一位”的方法进行。

以小数点为界，将整数部分从右向左每 3 位一组，最高位不足 3 位时添 0 补足 3 位；小数部分从左向右，每 3 位一组，最低位不足 3 位时添 0 补足 3 位。然后，将各组的 3 位二进制数按权展开后相加，得到一位八进制数。

【例 1-4】 将 $(101110001.11001)_2$ 转换成八进制数。

解：计算过程如下。

1015
1106
0011
.
1106

010.....2

计算结果为 $(101110001.11001)_2 = (561.62)_8$ 。

b. 将八进制数转换成二进制数时,采用“一位拆三位”的方法进行,即把八进制数每位上的数用相应的3位二进制数表示。

【例 1-5】 将 $(54.14)_8$ 转换为二进制数。

解:计算过程如下。

5.....101

4.....100

.

1.....001

4.....100

计算结果为 $(54.14)_8 = (101100.001100)_2$ 。

②二进制数与十六进制数之间的转换。

a. 把二进制数转换为十六进制数时,按“四位并一位”的方法进行。

以小数点为界,将整数部分从右向左每4位一组,最高位不足4位时,添0补足4位;小数部分从左向右,每4位一组,最低有效位不足4位时,添0补足4位。然后,将各组的4位二进制数按权展开后相加,得到一位十六进制数。

【例 1-6】 将 $(101100.0001101)_2$ 转换成十六进制数。

解:计算过程如下。

0010.....2

1100.....C

.

0001.....1

1010.....A

计算结果为 $(101100.0001101)_2 = (2C.1A)_{16}$ 。

b. 将十六进制数转换成二进制数时,采用“一位拆四位”的方法进行,即把十六进制数每位上的数用相应的4位二进制数表示。

【例 1-7】 将 $(C1.B7)_{16}$ 转换成二进制数。

解:计算过程如下。

C.....1100

1.....0001

.

B.....1011

7.....0111

计算结果为 $(C1.B7)_{16} = (11000001.10110111)_2$ 。

1.3.2 计算机中的数据单位

为衡量计算机中的数据量,人们规定了二进制的常用单位。

1. 位

位也称为比特(bit),即“一个二进制位”的意思,只能表示0和1。位是计算机存储数据的最小单位。

2. 字节

字节(byte,B)是表示信息含义的最小单位,也是计算机存储容量的基本单位。

换算关系如下。

$$1 \text{ B} = 8 \text{ bit}$$

$$1 \text{ KB} = 2^{10} \text{ B} = 1\,024 \text{ B}$$

$$1 \text{ MB} = 1\,024 \text{ KB} = 2^{20} \text{ B}$$

$$1 \text{ GB} = 1\,024 \text{ MB} = 2^{30} \text{ B}$$

$$1 \text{ TB} = 1\,024 \text{ GB} = 2^{40} \text{ B}$$

3. 字

字(word)是计算机进行数据处理时一次存取、加工和传送的一组二进制位。字的长度称为字长,字长是衡量计算机性能的一个重要指标。字长越长,精度越高。不同微处理器的字长不同,常见的有32位、64位。

1.3.3 数字、字符与汉字的编码

1. 数字编码

计算机基础的信息处理系统是利用电子元件(硬件)的不同状态组合来存储和处理信息的。元件的不同状态组合能代表数字系统的数字,因此编码就是将符号转换为计算机可以接受的数字系统的数,称为数字代码。BCD(binary-coded decimal)码也称“二-十进制编码”,它是专门解决用二进制数表示十进制数的问题的编码方法。它既具有二进制数的形式,以满足数字系统的要求,又具有十进制的特点(只有10种有效状态)。在某些情况下,计算机也可以对这种形式的数直接进行运算。二-十进制编码方法很多,有8421码、2421码等,最常用的是8421编码。

8421码是使用最广的BCD码,是一种有权码,其各位的权分别是(从最高有效位开始到最低有效位)8、4、2、1。

【例 1-8】 写出十进制数 563.97D 对应的 8421BCD 码。

【解】

$$563.97\text{D} = (0101\ 0110\ 0011.1001\ 0111)_{8421\text{BCD}}$$

【例 1-9】 写出 8421BCD 码(1101001.01011)_{8421BCD} 对应的十进制数。

【解】

$$(1101001.01011)_{8421\text{BCD}} = (0110\ 1001.0101\ 1000)_{8421\text{BCD}} = 69.58\text{D}$$

在使用8421BCD码时一定要注意其有效的编码仅10个,即0000~1001。4位二进制数的其余6个编码1010、1011、1100、1101、1110、1111不是有效编码。表1-2所示为8421编码。

表 1-2 8421 编码

十进制数	8421BCD 码	十进制数	8421BCD 码
0	0000	8	1000
1	0001	9	1001
2	0010	10	0001 0000
3	0011	11	0001 0001
4	0100	12	0001 0010
5	0101	13	0001 0011
6	0110	14	0001 0100
7	0111	15	0001 0101

BCD 又分为非压缩式和压缩式两种:非压缩式 BCD 码只有低四位有效,压缩式 BCD 码则将高四位也用上了,即一字节有两个 BCD 码。

2. 字符编码

字符是各种文字和符号的总称,包括各国的文字、标点符号、图形符号、数字等。字符集是多个字符的集合,字符集种类较多,每个字符集包含的字符数不同,常见的字符集有 ASCII 字符集、GB2312 字符集、BIG5 字符集、GB18030 字符集、Unicode 字符集等。计算机要准确地处理各种字符集文字,必须进行字符编码,以便能够识别和存储各种文字。

ASCII(American standard code for information interchange)即美国信息互换标准代码字符集,是基于罗马字母表的一套计算机编码系统。它主要用于显示现代英语和其他西欧语言。它是现今最通用的单字节编码系统,并等同于国际标准 ISO 646。

ASCII 字符集包括控制字符(Enter 键、Backspace 键等)与可显示字符(英文大小写字母、阿拉伯数字和西文符号)两大类,即 32 个通用控制字符、10 个十进制数码、52 个英文大小写字母和 34 个专用字符。通常用 7 位(bits)表示一个字符,共 128 个字符。7 位编码的字符集只能支持 128 个字符,为了表示更多的欧洲常用字符,在计算机发展过程中对 ASCII 进行了扩展,ASCII 扩展字符集使用 8 位(bits)表示一个字符,共 256 个字符。ASCII 扩展字符集比 ASCII 字符集扩充出来的符号包括表格符号、计算符号、希腊字母和特殊的拉丁符号。表 1-3 列出了 7 位 ASCII 字符编码。

表 1-3 7 位 ASCII 字符编码

ASCII 值	控制字符	ASCII 值	控制字符	ASCII 值	控制字符	ASCII 值	控制字符
000	NUL	032	空格	064	@	096	`
001	SOH	033	!	065	A	097	a
002	STX	034	"	066	B	098	b
003	ETX	035	#	067	C	099	c
004	EOT	036	\$	068	D	100	d
005	ENQ	037	%	069	E	101	e
006	ACK	038	&	070	F	102	f

(续表)

ASCII 值	控制字符	ASCII 值	控制字符	ASCII 值	控制字符	ASCII 值	控制字符
007	BEL	039	'	071	G	103	g
008	BS	040	(072	H	104	h
009	HT	041)	073	I	105	i
010	LF	042	*	074	J	106	j
011	VT	043	+	075	K	107	k
012	FF	044	,	076	L	108	l
013	CR	045	—	077	M	109	m
014	SO	046	.	078	N	110	n
015	SI	047	/	079	O	111	o
016	DLE	048	0	080	P	112	p
017	DC1	049	1	081	Q	113	q
018	DC2	050	2	082	R	114	r
019	DC3	051	3	083	S	115	s
020	DC4	052	4	084	T	116	t
021	NAK	053	5	085	U	117	u
022	SYN	054	6	086	V	118	v
023	ETB	055	7	087	W	119	w
024	CAN	056	8	088	X	120	x
025	EM	057	9	089	Y	121	y
026	SUB	058	:	090	Z	122	z
027	ESC	059	;	091	[123	{
028	FS	060	<	092	\	124	
029	GS	061	=	093]	125	}
030	RS	062	>	094	^	126	~
031	US	063	?	095	_		

3. 汉字编码

汉字信息的输入、输出及其处理,远比西文困难得多,原因是汉字的编码和处理实在太复杂了。经过多年的努力,我国在汉字信息处理的研制和开发方面取得了突破性的进展,使我国的汉字信息处理技术处于世界领先地位。

(1) 国标码和汉字内码。汉字也是一种字符,常用的汉字就有 3 000~5 000 个,显然无法用一字节的编码来区分。所以,汉字通常用两字节进行编码。1981 年我国公布的《信息交换用汉字编码字符集 基本集》(GB 2312—1980)共收集了 7 445 个图形字符,其中汉字字符 6 763 个,并分为两级,即常用的一级汉字 3 755 个(按汉语拼音排序)和次常用汉字 3 008 个(按偏旁部首排序),其他图形符号 682 个。

GB 2312—1980 编码简称国标码,它规定每个图形字符由两个 7 位二进制编码表示,即每个编码需要占用两字节,每字节内占用 7 位信息,最高位补 0。例如,汉字“啊”的国标码为 3021H,即 00110000 00100001。

GBK 编码是在 GB 2312—1980 标准基础上的内码扩展规范,使用了双字节编码方案,其编码范围为 8140~FEFE(剔除 xx7F),共 23 940 个码位,共收录了 21 003 个汉字,完全兼容 GB 2312—1980 标准,支持国际标准 ISO/IEC 10646-1 和国家标准 GB 13000—2010 中的全部中日韩汉字,并包含 BIG5 编码中的所有汉字。GBK 编码方案于 1995 年 10 月制定,1995 年 12 月正式发布,目前中文版的 Windows 7、Windows 10 及之前版本的中文版 Windows 操作系统等都支持 GBK 编码方案。

汉字内码是汉字在计算机内部存储、处理和传输用的信息代码,要求它与 ASCII 码兼容但又不能相同,以便实现汉字和西文的并存兼容。通常将国标码 2 字节的最高位置“1”作为汉字的内码。以汉字“啊”为例,其内码为 B0A1H,即 10110000 10100001。

(2)汉字输入码。在计算机系统处理汉字时,首先遇到的问题是输入汉字。汉字输入码又称为外码,是指从键盘输入汉字时采用的编码,主要有以下几类。

①数字编码。用一串数字代表一个汉字,最常用的是国标区位码,它实际上是国标码的一种简单变形。把 GB 2312—1980 全部字符集分为 94 区,其中 1~15 区是字母、数字和图形符号区,16~55 区是一级汉字区,56~87 区是二级汉字和偏旁部首区,每个区又分为 94 位,编号也是 01~94。这样,每一个字符便具有一个区码和一个位码。将区码置前、位码置后,组合在一起就成为区位码。国标码与区位码是一一对应的。可以这样认为:区位码是十进制表示的国标码,国标码是十六进制表示的区位码。将某个汉字的区码和位码分别转换成十六进制后再分别加 20H 即可得到相应的国标码。使用区位码输入汉字或字符,方法简单且没有重码,能输入所有的汉字,但是用户不可能把区位码背诵下来,查找区位码也不方便,所以难以实现快速输入汉字或字符,通常仅用于输入一些特殊字符或图形符号。

②拼音码。拼音码是一种以汉语读音为基础的输入方法,由于汉字同音字较多,因此重码率较高,输入速度较慢。

③形码。形码指根据汉字形状确定的编码。尽管汉字总量很多,但构成汉字的部件和笔画是有限的。因此,把汉字的笔画部件用字母或数字进行编码,按笔画书写顺序依次输入,就能表示一个汉字。常用的五笔字型码就是采用这种编码方法。

④音形码。音形码根据汉字的读音和字形进行编码。它的编码规则既与音素有关,又与形素有关,即取音码实现简单、易于接受的优点和形码形象、直观之所长,从而得到较好的输入效果。例如,双拼码、五十字元等。

不同的汉字输入方法有不同的汉字外码,即汉字的外码可以有多个,但内码只能有一个。目前已有的汉字输入编码方法有数百种,如首尾码、拼音码、表形码、五笔字型码等。一种好的汉字输入编码方法应该具备规则简单、易于记忆、操作方便、编码容量大、编码短和重码率低等特征。

(3)汉字字形码。汉字字形码用在输出时产生汉字的字形,通常采用点阵形式产生。所谓汉字字形码,就是确定一个汉字字形点阵的代码。全点阵字形中的每一点用一个二进制位来表示,随着字形点阵的不同,它们需要的二进制位数也不同。例如,24×24 的字形点阵,每字需要 72 字节;32×32 的字形点阵,每字共需 128 字节。与每个汉字对应的这一串字

节,就是汉字的字形码。不同字形码的集合就形成不同的字库。

(4)汉字编码之间的关系。汉字通常通过汉字输入码,并借助输入设备输入计算机内,再由汉字系统的输入管理模块进行查表或计算,将输入码(外码)转换成机器内码存入计算机存储器中。当存储在计算机内的汉字需要在屏幕上显示或在打印机上输出时,要借助汉字机内码在字库中找出汉字的字形码,这种代码转换过程如图 1-1 所示。



图 1-1 汉字在计算机中的处理过程及编码之间的关系

1.4 计算机病毒及防治

1.4.1 计算机病毒的定义及特征

1. 计算机病毒的定义

计算机病毒是一组以通过复制自身实现感染其他软件为目的的程序。当程序运行时,嵌入的病毒也随之运行并感染其他程序。一些病毒不带有恶意攻击性编码,但更多的病毒携带毒码,一旦被事先设定好的环境激发,即可感染和破坏。《中华人民共和国计算机信息系统安全保护条例》中明确将计算机病毒定义为:“编制或者在计算机程序中插入的破坏计算机功能或者毁坏数据,影响计算机使用,并能自我复制的一组计算机指令或者程序代码。”

2. 计算机病毒的特征

计算机病毒具有如下一些特征。

(1)破坏性。凡是由软件手段能触及计算机资源的地方均可能受到计算机病毒的破坏。其表现为:占用 CPU 的时间和内存,造成进程堵塞;对数据或文件进行破坏;打乱屏幕的显示等。

(2)隐蔽性。病毒程序大多附着在正常程序之中,很难被发现。

(3)潜伏性。病毒侵入后,一般不会立即活动,须等一段时间,条件成熟后才发生作用。

(4)传染性。传染性是病毒的一个重要特性,它通过修改别的程序,把自身复制进去,从而达到扩散的目的。

(5)可执行性。病毒与其他合法程序一样,是一段可执行程序,但它不是一个完整的程序,而是寄生在其他可执行程序中,因此,它享有一般程序所能得到的权力。病毒运行时,会与合法程序争夺系统的控制权。

1.4.2 计算机病毒的分类

从第一个病毒问世以来,究竟世界上有多少种病毒,说法不一。无论多少种,病毒的数

量仍在不断增加。按照计算机病毒的特点及特性,计算机病毒的分类方法有许多种。因此,同一种病毒可能有多种不同的分法。

(1)按照计算机病毒攻击的系统分类,主要有攻击 DOS 系统的病毒、攻击 Windows 系统的病毒、攻击 UNIX 系统的病毒、攻击 OS/2 系统的病毒。

(2)按照病毒的攻击机型分类,主要有攻击微型计算机的病毒、攻击小型计算机的病毒、攻击工作站的病毒。

(3)按照计算机病毒的链结方式分类,主要有源码型病毒、嵌入型病毒、外壳型病毒、操作系统型病毒。

(4)按照计算机病毒的破坏情况分类,主要有良性计算机病毒和恶性计算机病毒。

(5)按照计算机病毒的寄生部位或传染对象分类,主要有磁盘引导区传染的计算机病毒、操作系统传染的计算机病毒、可执行程序传染的计算机病毒。

(6)按照计算机病毒激活的时间分类,主要有定时病毒和随机病毒两种。

(7)按照传播媒介分类,主要有单机病毒和网络病毒。

(8)按照寄生方式和传染途径分类,主要有引导型病毒、文件型病毒和混合型病毒。

1.4.3 计算机病毒的防范

(1)预防第一。利用 Windows Update 确保操作系统及时更新,防止利用系统漏洞传播的病毒有机可乘;确定系统登录密码已设定为强密码;关闭不必要的共享或将共享资源设为“只读”状态。留意病毒和安全警告信息,做好相应的预防措施。

(2)选择优秀的反病毒软件。建议安装优秀的反病毒软件,推荐选择金山毒霸系列杀毒软件产品。

(3)定期扫描系统。如果是第一次启动反病毒软件,最好让它扫描整个系统。通常,反病毒程序都能够设置成在计算机每次启动时扫描系统或在定期计划的基础上运行。

(4)定期更新反病毒软件。既然安装了病毒防护软件,就应该确保它是最新的。优秀的反病毒程序带有自动连接到互联网上,并且只要软件厂商发现了一种新的威胁就会添加新的病毒探测代码的功能,而目前金山毒霸的主动实时升级正是这种技术的代表。

(5)不轻易执行附件中的 EXE 和 COM 等可执行程序。这些附件极有可能带有计算机病毒或黑客程序,轻易运行,很可能带来不可预测的结果。对于认识的朋友和陌生人发过来的电子邮件中的可执行程序附件都必须检查,确定无异后才可使用。

(6)不轻易打开附件中的文档文件。对方发送过来的电子邮件及相关附件的文档,首先要用“另存为...”命令(“Save As...”)保存到本地硬盘,待查杀计算机病毒软件检查无毒后才可以打开使用。如果直接双击 DOC、XLS 等附件文档,会自动启用 Word 或 Excel,如附件中有计算机病毒,则会立刻传染;如有“是否启用宏”的提示,不要轻易打开,否则极有可能染上宏病毒。

(7)不直接运行附件。对于文件扩展名比较特殊的附件,或者是带有脚本文件(如 *.VBS、*.SHS 等)的附件,不要直接打开,一般可以删除包含这些附件的电子邮件,以保证计算机系统不受计算机病毒的危害。

(8)慎用预览功能。如果使用 Outlook Express 作为收发电子邮件软件,也要进行设置。执行“工具”菜单中的“选项”命令,在“阅读”中不选中“在预览窗格中自动显示新闻邮件”和

“自动显示新闻邮件中的图片附件”。这样可以防止有些电子邮件计算机病毒利用 Outlook Express 的默认设置自动运行,破坏系统。

(9)警惕发送出去的邮件。对于本机往外传送的邮件,也一定要仔细检查,确定无毒后方可发送,否则将给接收邮件的计算机用户带来病毒危害。

1.4.4 计算机职业道德与软件知识产权

1. 计算机职业道德

计算机职业作为一种不同于其他职业的特殊职业,有着自己与众不同的职业道德和行为准则,这些职业道德和行为准则是每一个计算机行业从业人员都要遵守的。

(1)职业道德的概念。所谓职业道德,就是同人们的职业活动紧密联系的符合职业特点所要求的道德准则、道德情操与道德品质的总和。

每个从业人员,不论是从事哪种职业,在职业活动中都要遵守道德。例如,教师要遵守教书育人、为人师表的职业道德,医生要遵守救死扶伤的职业道德,等等。职业道德不仅是从业人员在职业活动中的行为标准和要求,而且是本行业对社会所承担的道德责任和义务。职业道德是社会道德在职业生活中的具体化。

(2)职业道德的特点。职业道德作为一种特殊的道德规范,具有以下 4 个主要特点。

①在内容方面,职业道德总是要鲜明地表达职业义务、职业责任及职业行为上的道德准则。

②在表现形式方面,职业道德往往比较具体、灵活、多样。它总是从本职业的实际交流活动出发,采用制度、守则、公约、承诺、誓言及标语口号的形式加以表现。

③从调节范围来看,职业道德一方面是用来调节从业人员内部关系,加强职业、行业内部人员的凝聚力;另一方面也是用来调节从业人员与其服务对象之间的关系,以塑造本职业从业人员的形象。

④从产生的效果来看,职业道德既能使一定的社会或阶级的道德原则和规范“职业化”,又能使个人道德品质“成熟化”。

(3)计算机行业从业人员职业道德的最基本要求。法律是道德的底线,计算机行业从业人员职业道德的最基本要求就是国家关于计算机管理方面的法律法规。我国目前还没有一部统一的计算机信息法,但是全国人大、国务院和国务院的各部委等具有立法权的政府机关还是制定了一批管理计算机行业的法律法规,比较常见的有《全国人民代表大会常务委员会关于维护互联网安全的决定》《计算机软件保护条例》《互联网信息服务管理办法》《互联网电子公告服务管理办法》等,这些法律法规应当被每一位计算机行业从业人员所牢记,严格遵守这些法律法规是计算机行业从业人员遵守职业道德的最基本要求。

2. 计算机软件知识产权

随着计算机科学技术的飞速发展,我国法制建设进程不断加快,计算机软件知识产权保护问题越来越受到产业界和法学界的关注。

(1)计算机软件和知识产权保护。计算机软件是相对于硬件而言的,在我国,计算机软件保护的内容是计算机程序和计算机文档。20 世纪 60 年代,德国最早提出了保护计算机软件。此后,世界各国,特别是发达国家都在寻找保护计算机软件的方法。1978 年,世界知识

产权组织颁布了《保护计算机软件示范条款》，为世界各国保护计算机软件提供了导向。我国也顺应国际趋势，为了能较完备地保护程序所有者的权益，在《中华人民共和国著作权法》（以下简称《著作权法》）第三条第八款中明确地将计算机软件列为受保护的作品范围。同时又针对计算机软件不同于其他作品的特点，我国于1991年制定了《计算机软件保护条例》单行法规对计算机软件进行保护。计算机软件与知识产权有着密切的关系，有关软件外在表现的专有权属于著作权，可以通过《著作权法》加以保护；有关软件内容思想方面的专有权属于专利权或商业秘密，可以通过《中华人民共和国专利法》或《中华人民共和国反不正当竞争法》加以保护；有关软件名称的使用权属于商标权，可以通过《中华人民共和国商标法》得以保护。可见，随着一项软件的开发完成，有可能产生4种知识产权，也就是说有可能涉及知识产权的所有领域。

在计算机软件法律保护的现有模式上，世界上多数国家都将计算机软件纳入著作权保护对象，我国也是如此。1972年，菲律宾在其著作权法中明确将计算机软件列入被保护对象，成为世界上第一个用著作权法保护计算机软件的国家。1980年，美国继20世纪60年代前后在司法上将计算机软件作为著作权保护对象之后，修订版权法，并增加了保护软件的条款。但是，用著作权法对计算机软件予以保护，并非最好的模式。曾经有人主张用专利法保护软件，也有人主张用商标法来保护软件，但都因专利法、商标法不适合软件保护的要求而作罢。最后，多数国家采用著作权保护模式，这是一种不得已的过渡方式，因为用著作权法保护软件存在一些问题，如传统的著作权法是否能适应具有实用工具性的软件的保护需要等。所以有人预言，迟早会有一种新的、前所未有的法律来保护软件。

(2) 计算机软件的特点及其保护的特殊性。包括我国在内的许多国家中，计算机软件都是被作为《著作权法》的保护对象之一，是被作为文字作品进行保护的。但是，计算机软件与一般文字作品相比，具有自己的特点。

① 计算机软件使用的语言具有抽象性。与传统的文字作品不同，计算机软件所表现的是人类为解决某个问题而设计的逻辑步骤，使用人工设计的程序语言，只有少数受过专门训练的人才能使用。

② 计算机软件兼有作品和工具的双重性质。计算机程序可以借助数字、文字和符号表现出来，并可以用物质载体，如光盘、硬盘固定下来。计算机文档是用自然语言或者形式化语言编写的。所以，计算机软件有与文字作品相同的性质，同时，计算机程序及其文档能够驱动计算机硬件完成一定的功能，以取得预期的结果。因此，计算机软件是一种工具，具有实用性，而传统的文字作品不具备这种性质。

③ 计算机软件的更新比较快。计算机软件具有实用工具性质，随着科技的飞速发展，软件的更新较快，一旦有新的同类软件产生，原有软件的实际效用就大大降低，很快就会被市场淘汰。目前，计算机软件的使用价值一般在10年左右。

④ 计算机软件具有较大的经济价值。由于计算机软件是一种实用工具，因此，与一般文字作品比较，它在实际应用中会产生较大的经济价值，其财产权应当受到特殊的保护。

⑤ 就权利而言，计算机软件著作权不具有修改权和保护作品完整权。这由软件的实用工具性所决定。人们购买软件并不是为了阅读或欣赏，而是安装在计算机上，使其运行，得到预期的结果。为了适应不同的计算机或不同的环境，必须对购置的软件稍加修改。这一点在世界各国的版权法中均有反映。

(3)我国计算机软件知识产权保护现状。国内软件市场欠规范主要表现在两个方面:一是用户版权意识淡薄,盗版风行;二是少数厂商以无序为竞争手段,扩大市场份额,进一步助长了不良风气的蔓延,给规范市场增添了障碍。对软件知识产权尊重的侵蚀是保护这种产权的法律问题。不幸的是,借用、移植、复制软件程序在国内已经不是什么引人注目的事了。这些猖獗地侵犯著作权的事件致使许多公司蒙受了严重的财产损失,并大大阻碍了新技术的发展。复制诸如数据库和电子表格之类的软件程序是非法的,可许多用户还是无所顾忌。这使得保护软件知识产权的问题变得更复杂。软件公司靠出售产品使用权盈利,盗版直接侵害厂家的经济利益,软件商们与盗版行为做斗争义不容辞,但在实践中却有个别厂家听任乃至鼓励盗版,有悖常理。这是缺少“游戏规则”的市场竞争中滋生的“怪胎”,对软件行业的总体健康发展有百害而无一利。

1.5 多媒体技术

1.5.1 多媒体技术的概念

1. 媒体

媒体是指人们用于传播和表示各种信息的载体。在计算机领域,媒体一般有两种含义:一是指交流、传播信息的载体,如数字、文本、声音、图像等;二是指用于存储信息的实体,如磁盘、光盘、磁带等。

按照国际电报电话咨询委员会的定义,媒体可分为感觉媒体、表示媒体、表现媒体、存储媒体和传输媒体 5 类。各种媒体之间的关系如图 1-2 所示。

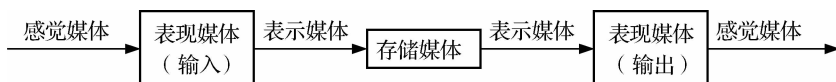


图 1-2 各种媒体之间的关系

2. 多媒体

多媒体(multimedia)是由 multiple 和 media 复合而成的,简单地理解为两个或两个以上的单媒体的有机组合。计算机中的多媒体是指将文本、数字、图形、图像、音频、视频和动画等基本媒体以不同的形式组合所构成的复合体。

3. 多媒体技术

多媒体技术是指用计算机综合处理文本、图形、图像、声音、视频等多种媒体信息,使它们建立一种逻辑链接,并集成为一个交互性系统的技术。它实际上是一种信息处理技术,是把多媒体信息通过计算机进行数字化采集、获取、压缩/解压缩、编辑、存储等加工处理,再以单独或合成形式表现出来的一体化技术。

1.5.2 多媒体技术的特征

多媒体技术具有以下 4 个特征。

(1)多维性。多维性是指多媒体扩展和放大了计算机处理的信息范围和空间,不再局限于数值和文本,而是广泛采用图像、图形、视频、音频等形式来增加输出信息的表现能力,丰富显示效果。多媒体技术使人们不仅能看到文字、图像,还能让人们听到声音,使人有身临其境的感觉。

(2)集成性。多媒体技术的集成性是指将多种媒体有机地组织在一起,并建立起不同媒体之间的联系,做到图、文、声、像一体化。

(3)交互性。多媒体技术的交互性是指除了播放以外,还具有与使用者交互沟通的特点,也就是说,人们可通过软件系统的支持,对多媒体进行控制。

(4)实时性。音频信息和视频信息都是与时间有关的媒体,在加工、存储和播放时,需要考虑时间特性。例如,在播放音频文件时,应该保证声音的连续性。这就对存取数据的速度、解压缩的速度及最后播放的速度提出了很高的要求,这就是媒体的实时性。

1.5.3 多媒体信息中的媒体元素

从多媒体技术来看,多媒体由文本、图形、图像、动画、音频、视频等基本要素组成。

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,是人 与计算机之间进行信息交换的主要媒体。它主要用于对知识的描述性表示,如阐述概念、定义、原理和问题及显示标题、菜单等内容。

2. 图形

图形是由各种绘图工具绘制的几何图形,也称为矢量图形,由线、形、体、文字等图元构成,具有占用存储空间小、缩放不失真等优点,但难以表现真实感场景。

3. 图像

图像是由输入设备捕获的实际场景或以数字化形式存储的画面。图像能表现对象的细节和质感,图像占用的存储空间大,常采用压缩技术实现图像的存储和传输。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。

4. 动画

动画就是运动的图画,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。

5. 音频

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。现实世界中的各种声音必须转换成数字信号并经过压缩编码,计算机才能接受和处理。多媒体计算机处理音频信号时,通过采样、量化和编码过程把模拟信号转化为离散的数字信号进行存储、处理和传输。这种数字化的声音信息以文件形式保存,即通常所说的音频文件或声音文件。

6. 视频

若干幅内容互相联系的图像连续播放就形成了视频。视频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。视频主要来自摄像机拍摄连续自然场景画面。多媒体计算机处理的视频信号必须是全数字化的信号。

1.5.4 多媒体技术的应用

多媒体技术主要应用于以下几个方面。

(1)数据压缩,图像处理技术。图像、音频、视频信号的数字化,给存储器的存储容量、通信干线的信道传输率及计算机的速度都带来了极大的压力。数据压缩技术为图像、音频和视频信号的压缩提供了解决方法,从而提高计算机实时处理音频、视频信息的能力,同时也保证音频、视频的高质量播放。为此,国际标准化组织在20世纪90年代制定了JPEG标准、H.261标准、MPEG标准这三个重要的视频图像压缩编码的国际标准。

①压缩编码JPEG技术。压缩编码JPEG技术为国际上彩色、灰度、静止图像的第一个国际标准,它主要用于对静态图像的压缩,包括电视图像序列的帧内图像的压缩。

提示: 图片文件格式一般有JPG、PSD、GIF(网络模式)、PNG、BMP、TIFF、PCX、FLC、DIB等,很多格式可以相互转换。

②视频图像压缩编码261技术。视频图像压缩编码261技术为视频图像压缩编码的国际标准,主要适用于视频电话和电视会议。

③MPEG视频压缩技术。MPEG视频压缩技术是针对运动图像的数据压缩技术。它分为MPEG-I、MPEG-II、MPEG-IV、MPEG-7和MPEG-21。

MPEG-I用于数字存储上活动图像及伴音的编码,数据速率为1.5 Mb/s,采用SIF格式图像,音质接近CD,MPEG-I压缩技术相当成熟,广泛地应用在VCD制作、图像监控领域。

MPEG-II是在MPEG-I基础上进行扩充和完善的。MPEG-II的视频数据速率为4~5 Mb/s,可提供分辨率为720×480(NTSC)或720×576(PAL)的广播级质量的视频,用于包括宽屏幕和高清晰度电视(HDTV)在内的高质量电视和广播。

提示: 视频文件格式一般有AVI、MPEG1-4、MOV、RM、WMV、MTV、DAT、WMV、ASF等,很多格式可以相互转换。

(2)音频信息处理。在多媒体技术中,常见的声音文件格式有WAV文件、VOC文件、MIDI文件、AIF文件、SON文件、RMI文件等。

①音频信息录制编辑软件。常用的音频信息录制编辑软件有WaveEditor软件、Cool Edit软件、Wave Editor软件、WaveStudio及专业声卡自带的编辑软件。例如,Sound Blaster卡的VEdit2软件、Microsoft SoundSystem卡的Quick Recorder软件。

②语音识别。随着计算机技术的迅猛发展,语音识别技术将人机语音通信变成了现实,并且成为新一代智能计算机的一大亮点。例如,QQ、微信的语音聊天,以及机器人的大量应用。而多媒体技术正向网络化和智能化两个方面快速发展,为人们的工作、学习和生活提供方便。

1.6 计算思维与互联网思维

1.6.1 计算思维

1. 计算思维的定义

计算机和信息科技的普及实际上是在全社会传播一种计算思维(computational thinking), 计算思维是当前一个颇受关注的涉及计算机科学本质问题和未来走向的基础性概念。计算思维作为一种重要的思维方式, 与理论思维、实验思维共同构成了科学思维的全部内容。美国卡内基梅隆大学的周以真教授(Jeannette M. Wing)首先给出了计算思维的定义: 运用计算机科学的基础概念进行问题求解、系统设计及人类行为理解等涵盖计算机科学广度的一系列思维活动。

提示: 计算思维虽然有着计算机科学的许多特征, 但是计算思维本身却并不是计算机科学的专属。实际上, 即使没有计算机, 计算思维也在逐步地发展, 并且有些内容与计算机并没有关系。但是, 正是计算机的出现, 给计算思维的研究和发展带来了根本性的变化。

计算思维是一种普适的思维, 是概念化思维、人的思维、数学和工程互补融合的思维、面向所有人的思维, 同时是现代信息社会中每个人的基本技能。

2. 计算思维的本质

计算思维强调一切皆可计算, 从物理世界到人类社会模拟, 从人类社会再到智能活动, 都可认为是计算的某种形式。计算思维最根本的内容, 即其本质(essence)是抽象(abstraction)和自动化(automation)。

3. 计算思维的特性

计算思维的特性主要表现在以下几个方面。

(1) 概念化, 不是程序化。计算机科学不是计算机编程。像计算机科学家那样去思维意味着远远不只能为计算机编程, 还要求能够在抽象的多个层次上思维。

(2) 基础的, 不是机械的技能。基础的技能是每一个人为了在现代社会中发挥职能所必须掌握的。生搬硬套之机械的技能意味着机械的重复。具有讽刺意味的是, 只有当计算机科学解决了人工智能的宏伟挑战——使计算机像人类一样思考之后, 思维才会变成机械的生搬硬套。

(3) 人的, 不是计算机的思维。计算思维是人类求解问题的一条途径, 但绝非试图使人类像计算机那样思考。与枯燥且沉闷的计算机相比, 人类聪颖且富有想象力。使用计算思维控制计算设备, 用人类自己的智慧去解决那些计算时代之前不敢尝试的问题, 实现一种“只有想不到, 没有做不到”的境界。

(4) 数学和工程思维的互补与融合。计算机科学在本质上源自数学思维, 因为像所有的科学一样, 它的形式化解析基础筑于数学之上。计算机科学又从本质上源自工程思维。基本计算设备的限制迫使计算机学家必须计算性地思考, 不能只是数学性地思考。

1.6.2 互联网思维

1. 认识互联网思维

随着互联网技术作为工具的逐步发展,越来越多的商业形态受到互联网的冲击。当这种冲击不断加深和变革不断加剧时,互联网就不再仅仅是一种技术,而是逐渐演变成为一种思维范式,即互联网思维。这种思维是在(移动)互联网、云计算、大数据等科技不断发展的背景下,对市场、用户、产品、企业价值链乃至整个商业生态进行重新审视的思考方式。

提示: 不是因为有了互联网,才有了互联网思维,而是因为互联网的出现和发展,使得这些思维得以集中爆发。

互联网正在成为现代社会真正的基础设施之一,就像电力和道路一样。互联网不仅仅是用来提高效率的工具,是构建未来生产方式和生活方式的基础设施,更重要的是,互联网思维成为一切商业思维的起点。

2. 互联网思维的本质

互联网思维的本质是商业回归人性。互联网的发展让互动变得更加高效,包括人与人之间的互动,也包括人机互动。Web 1.0 时代是门户时代,具体代表有新浪、搜狐、网易等门户网站;Web 2.0 时代是搜索、社交时代,典型产品有博客、微博等;Web 3.0 时代是大互联时代,典型特点是多对多交互。从目前发展来看,现在仅仅是大互联时代的初期,真正的3.0时代是基于物联网、大数据和云计算的智能生活时代,是一个在“以人为本”的思维指引下的新商业文明时代。

互联网思维更注重人的价值。基于互联网的商业模式建立在平等、开放的基础上,是真正的以人为本的经济,是一种人性的回归,让商业真正回归人性。

3. 互联网思维模型

互联网的发展使得大数据、云计算、社会化网络等技术成为基础设施,用户和品牌之间得以更加便捷地连接和互动,不再只是销售或服务人员去面对终端用户,用户越来越多地参与到厂商的价值链条各个环节。因此,在互联网时代,为了更快、更好地满足用户需求,传统的价值链模型就会被互联网技术和思维重构,在重构过程中,产生了互联网思维模型,具体体现在以下几个方面。

(1)用户思维。在价值链各个环节中都要“以用户为中心”去考虑问题,只有深度理解用户才能生存,没有认同,就没有合同。在“以用户为中心”的互联网时代,消费者的话语权日益增大,并且影响着企业各环节的决策,以小米为代表的新经济企业,使得用户越来越广泛地参与到产品研发和品牌建设环节之中。可以说,用户思维是所有互联网思维的核心,没有用户思维,也就不可能领悟好其他思维。

(2)简约思维。互联网时代,信息爆炸,用户的耐心越来越不足,所以,必须在短时间内抓住用户。简约思维也即大道至简,越简单的东西越容易传播,同时,在产品的设计方面,要做减法,即外观要简洁,内在的操作流程要简化。

(3)极致思维。极致就是超越用户想象,极致思维就是把产品、服务和用户体验做到极致,超越用户预期。在互联网时代,只有产品和服务给消费者带来的体验足够好,才可能真正地抓住消费者,赢得人心,这就是一种极致思维的体现。

(4)迭代思维。“敏捷开发”是互联网产品开发的典型方法论,以用户为核心、迭代、循序渐进,允许有所不足,不断试错,在持续迭代中完善产品。迭代思维中有两个点,一个是“微”,一个是“快”。“微”,要从细微的用户需求入手,贴近用户心理,在用户参与和反馈中逐步改进;“快”,要求及时、实时关注消费者需求,快速地对消费者需求做出反应,从而使产品更贴近消费者。

(5)流量思维。流量意味着体量,体量意味着分量。互联网产品,大多用免费策略极力争取用户、锁定用户,只要用户活跃数量达到临界点,就会质变,从而带来商机或价值。注意力经济时代,先把流量做上去,才有机会思考后面的问题,否则连生存的机会都没有。

(6)社会化思维。社会化商业的核心是网,公司面对的客户以网的形式存在,这将改变企业生产、销售、营销等整个形态。社会化思维要求利用好社会化媒体,一定要站在用户的角度,以用户的方式和用户沟通,同时,关注众包协作。

提示: 众包是以“蜂群思维”和层级架构为核心的互联网协作模式。维基百科就是典型的众包产品。另外,小米手机在研发中让用户深度参与,实际上也是一种众包模式。

(7)大数据思维。大数据思维是对大数据的认识,对企业资产、关键竞争要素的理解。用户在网络上一般会产生信息、行为、关系3个层面的数据,这些数据的沉淀,有助于企业进行预测和决策。一切皆可被数据化,企业必须构建自己的大数据平台,小企业,也要有大数据。大数据核心不在大,而在于数据挖掘和预测。

(8)平台思维。互联网的平台思维就是开放、共享和共赢的思维,平台模式最有可能成就产业巨头。平台模式的精髓,在于打造一个多主体共赢互利的生态圈,将来的平台之争,一定是生态圈之间的竞争。当企业不具备构建生态平台的实力时,就必须思考怎样利用现有的平台。

在企业内部,平台思维同样会影响企业的组织变革,将企业内部打造成“平台型组织”。例如,腾讯六大事业群的调整,都旨在发挥内部组织的平台化作用,内部平台化就是要变成自组织而非他组织,自组织是自己来创新。

(9)跨界思维。互联网和新科技的发展,导致很多产业边界变得模糊,互联网企业已无孔不入,如零售、图书、金融、电信、娱乐、交通、媒体等。掌握了用户和数据资产,就可以参与到跨界竞争,跨界变得越来越普遍。腾讯相继牵头设立银行,小米做手机、又做电视,是因为它们一方面掌握着用户数据,另一方面又具备用户思维,携“用户”以令诸侯,从而赢得跨界竞争。

1.7 计算机领域的研究热点

1.7.1 云计算

1. 云计算的定义

什么是“云计算”呢? 维基百科给出的定义是:云计算将 IT 相关的能力以服务的方式提供给用户,允许用户在不了解提供服务的技术、没有相关知识和设备操作能力的情况下,通

过 Internet 获取需要的服务。百度百科给出的定义是:云计算平台也称为云平台。云计算平台可以划分为 3 类:以数据存储为主的存储型云平台,以数据处理为主的计算型云平台,以及计算和数据存储处理兼顾的综合云平台。美国国家标准与技术研究院(NIST)的定义是:云计算是一种按使用量付费的模式,这种模式提供可用的、便捷的、按需的网络访问,进入可配置的计算资源共享池(资源包括网络、服务器、存储、应用软件、服务),这些资源能够被快速提供,只需要投入很少的管理工作或服务供应商进行很少的交互。人们看了这样的定义后,可能还是会对云计算一头雾水,因为“云”和“计算”根本就是两个不相关的概念,不像分布式计算、并行计算等概念相互之间是有联系的。

那么“云计算”这个概念是由谁提出来的呢?2006年,Google 高级工程师克里斯托弗·比希利亚第一次向 Google 董事长兼 CEO(首席执行官)施密特提出“云计算”的想法,在施密特的支持下,Google 推出了“Google 101 计划”,并正式提出“云”的概念。

“云计算”概念刚提出时,被很多业内的专家认为是在炒作概念,因为从本质上说,“云计算”与当时的分布式计算、网格计算很相似。但经过多年的发展,编者认为“云计算”与分布式计算或网格计算还是有很大区别的,而且正是由于提出了这个全新的名词,其商业应用价值得到大大提高,吸引了众多厂商和用户的追捧,其在不同领域的发展也得到了飞速的发展。虚拟化、多租户、高可用性、虚拟机迁移等概念或应用也随着云计算的发展而逐渐被人们所了解和认识。

下面给出本书对云计算的定义:云计算是指将多台计算机系统的资源(计算、存储、网络等)进行统一管理,对多租户提供按需使用、简化管理、方便部署的计算机应用系统。

其实要想更深入地理解云计算的概念,可以从不同的角度来看,下面分别从用户、技术提供商和技术开发人员的角度来解读云计算。

(1)从用户角度看云计算。从用户的角度来看,主要根据用户的体验和效果来描述,云计算可以总结为:云计算系统是一个信息基础设施,包含硬件设备、软件平台、系统管理的数据,以及相应的信息服务。用户使用该系统时,可以实现“按需索取、按用计费、无限扩展、网络访问”的效果。

简单地说,用户可以根据自己的需要,通过网络去获得自己需要的计算资源和软件服务。这些计算资源和软件服务是直接供用户使用而不需要用户做进一步的定制开发、管理与维护等工作。同时,这些计算资源和软件服务的规模可以根据用户业务与需求的变化,随时调整到足够大的规模。用户使用这些计算资源和软件服务,只需要按照使用量来支付租用的费用。

(2)从技术提供商角度看云计算。技术提供商对云计算的理解为通过调度优化的技术,管理和协同大量的计算资源;针对用户的需求,通过互联网发布与提供用户所需的计算资源和软件服务;基于租用模式以按用计费的方法进行收费。

技术提供商强调云计算系统需要组织与协同大量的计算资源来提供强大的 IT 能力和丰富的软件服务,利用调度优化的技术来提高资源的利用效率。云计算系统提供的 IT 能力和软件服务针对用户的直接需求,并且这些 IT 能力和软件服务都在互联网上发布,允许用户直接利用互联网来使用这些 IT 能力和软件服务。用户对资源的使用,按照其使用量来进行计费,实现云计算系统运营的盈利。

(3)从技术开发人员角度看云计算。技术开发人员作为云计算系统的设计和开发人员,

认为云计算是一个大型集中的信息系统,该系统通过虚拟化技术和面向服务的系统设计等手段来完成资源与能力的封装和交互,并且通过互联网来发布这些封装好的资源和能力。

所谓大型集中的信息系统,指的是包含大量的软硬件资源,并且通过技术和网络等对其进行集中式管理的信息系统。通常这些软硬件资源在物理上或者在网络连接上是集中或者相邻的,能够协同来完成同一个任务。

信息系统包含软硬件和很多软件功能,这些软硬件和软件功能如果需要被访问与使用,就必须有一种把相关资源和软件模块打包在一起且能够呈现给用户的方式。虚拟化技术和Web服务是最常见的封装与呈现技术,可以把硬件资源和软件功能等打包,并以虚拟计算机和网络服务的形式呈现给用户使用。

云计算作为一种技术手段和实现模式,使得计算资源成为向大众提供服务的社会基础设施,将对信息技术本身及其应用产生深远影响,软件工程方法、网络和终端设备的资源配置、获取信息和知识的方式等,无不因为云计算的出现而产生重要的变化。与此同时,云计算也深刻改变着信息产业的现有业态,催生了新型的产业和服务。云计算带来社会计算资源利用率的提高和计算资源获得的便利性,推动以互联网为基础的物联网迅速发展,将更加有效地提升人类感知世界、认识世界的能力,促进经济发展和社会进步。

2. 云计算的发展

云计算是继大型计算机到客户端/服务器模式之后的又一次发展过程,了解云计算的发展历史,有利于读者理解云计算的基本概念和掌握相关技术。

云计算的发展过程如图 1-3 所示。

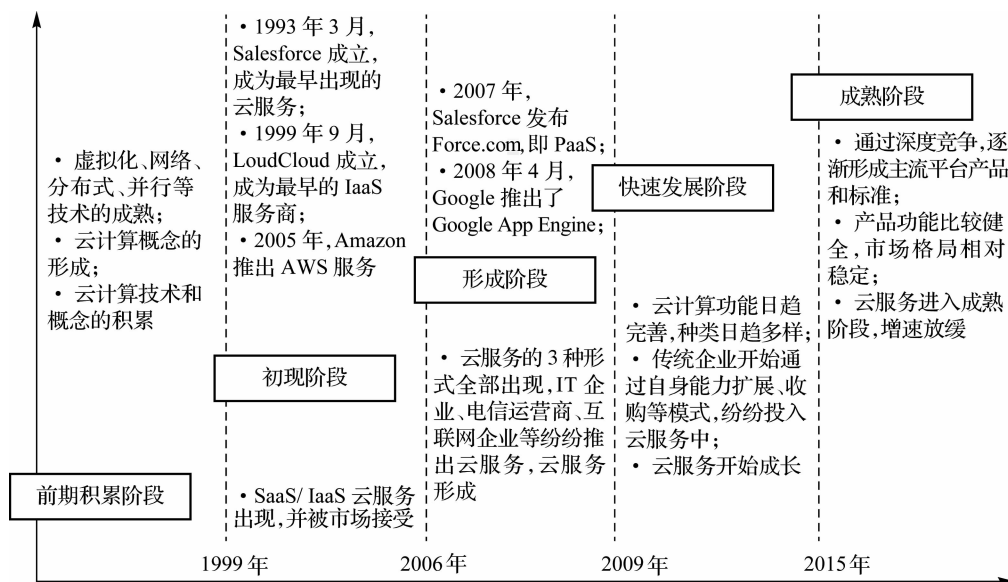


图 1-3 云计算的发展过程

3. 云计算的关键技术

(1) 高性能计算技术 (high performance computing, HPC)。高性能计算通常指使用很多处理器 (作为单个机器的一部分) 或者某一集群组织中的几台计算机 (作为单个计算资源

操作)的计算系统和环境,这是计算机科学的一个分支,主要研究并行算法和开发相关软件,致力于开发高性能计算机。

高性能计算的本质是支持全面分析、快速决策,即通过收集、分析和处理全面的材料、大量原始资料即模拟自然现象或产品,以最快的速度得到最终分析结果,提示客观规律,支持科学决策。对科研工作者来说,这意味着减少科学突破的时间,增加突破的深度;对工程师来说,这意味着缩短新产品上市的时间,增加复杂设计的可信度;对国家来说,这意味着提高综合国力和参与全球竞争的实力。

高性能计算机的发展趋势主要表现在网络化、体系结构主流化、开放和标准化、应用的多样化等方面。网络化的趋势将是高性能计算机最重要的趋势。高性能计算机的主要用途是作为网络计算环境中的主机。以后越来越多的应用是在网络环境下的应用,会出现数以十亿计的客户端设备,所有重要的数据及应用都会放在高性能服务器上,client/server模式会进入第二代,即服务器聚集的模式,这是一个发展趋势。

随着计算机技术的飞速发展,高性能计算速度不断提高,其标准也不断变化更新,对称多处理(symmetrical multi-processing, SMP)、大规模并行处理机、集群系统、网格计算、消息传递接口等都是高性能计算技术的内容。

(2)分布式数据存储技术。分布式数据存储是指将数据分散存储到多个数据存储服务器上。分布式数据存储目前很多都借鉴了 Google 的经验,在众多的服务器搭建一个分布式文件系统,再在这个分布式文件系统中实现相关的数据存储业务,甚至是再实现二级存储业务。

分布式数据存储技术包含非结构化数据存储和结构化数据存储。其中,非结构化数据存储主要采用文件存储和对象存储技术,而结构化数据存储主要采用分布式数据库技术,特别是 NoSQL 数据库。

(3)虚拟化技术。云计算的本质是服务,服务意味着一种可按需取用的状态。虚拟化(virtualization)是从单一的逻辑角度来看待和使用不同物理资源的方法,是物理资源的逻辑抽象。有人把云计算等同于虚拟化,这个说法不够准确。不采用虚拟化技术,也可以提供云计算服务,但要想真正发挥云计算的优势,虚拟化技术必不可少。虚拟化技术是云计算中的核心技术之一,它可以让 IT 基础更加灵活化,更易于调度,且能更加有效地进行虚拟机之间的隔离。

所谓虚拟化,是指将信息系统的各种物理资源,如服务器、网络、内存及数据等进行抽象、转换后呈现出来,打破实体结构间的不可切割的障碍,使用户可以更好地应用这些资源。这些新虚拟出来的资源不受现有资源的架设方法、地域或物理配置的限制。虚拟化的本质就是将原来运行在真实环境上的计算系统或组件运行在虚拟出来的环境中,其工作原理如图 1-4 所示。

(4)用户交互技术。随着云计算的逐步普及,浏览器已经不仅仅是一个客户端的软件,而逐步演变为承载着互联网的平台。浏览器与云计算的整合技术主要体现在两个方面:浏览器网络化与浏览器云服务。

国内各家浏览器都将网络化作为其功能的标配之一,主要功能体现在用户可以登录浏览器,并通过账号将个性化数据同步到服务端。用户在任何地方,只需要登录自己的账号,就能够同步更新所有的个性内容,包括浏览器选项配置、收藏夹、网址记录、智能填表和密码保存等。

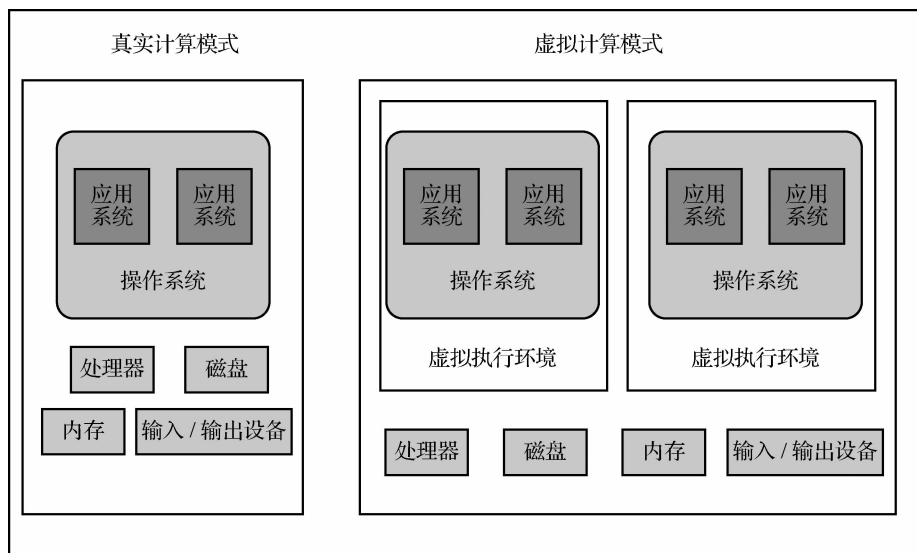


图 1-4 虚拟化工作原理

目前的浏览器云服务主要体现在 P2P 下载、视频加速等单独的客户端软件中，主要的应用研究方向包括基于浏览器的 P2P 下载、视频加速、分布式计算和多任务协同工作等。在多任务协同工作方面，AJAX(Asynchronous JavaScript and XML, 异步 JavaScript 和 XML)是一种创建交互式网页应用的网页开发技术，改变了传统网页的交互方式，改进了交互体验。

(5)安全管理技术。安全问题是用户是否选择云计算的主要顾虑之一。传统集中管理方式下也有安全问题，云计算的多租户、分布性、对网络和服务提供者的依赖性，为安全问题带来新的挑战。其中，主要的数据安全问题和风险内容如下。

①数据存储及访问控制。数据存储及访问控制包括如何有效存储数据以避免数据丢失或损坏，如何避免数据被非法访问和篡改，如何对多租户应用进行数据隔离，如何避免数据服务被阻塞，如何确保云端退役(at rest)数据的妥善保管或销毁，等等。

②数据传输保护。数据传输保护包括如何避免数据被窃取或攻击，如何保证数据在分布式应用中有效传递等。

③数据隐私及敏感信息保护。数据隐私及敏感信息保护包括如何保护数据所有权，并可根据需要提供给受信方使用，如何将个人身份信息及第三方数据移动到云端使用等。

④数据可用性。数据可用性包括如何提供稳定可靠的数据服务以保证业务的持续性，如何进行有效的数据容灾及恢复等。

⑤依从性管理。依从性管理包括如何保证数据服务及管理符合法律及政策的要求等。相应的数据安全管理工作如下。

①数据保护及隐私(data protection and privacy)包括虚拟镜像安全、数据加密及解密、数据验证、密钥管理、数据恢复、云迁移的数据安全等。

②身份及访问管理(identity and access management, IAM)包括身份验证、目录服务、身份鉴别/单点登录(single sign on, SSO)、个人身份信息保护、安全断言标记语言、虚拟资源访问、多租用数据授权、基于角色的数据访问和云防火墙技术等。

③数据传输(data transportation)包括传输加密及解密、密钥管理、信息管理等。

④可用性管理(availability management)包括单点失败(single point of failure, SPoF)、

主机防攻击和容灾保护等。

⑤日志管理(log management)包括日志系统监控、可用性监控、流量监控、数据完整性监控和网络入侵监控等。

⑥审计管理(audit management)包括审计信任管理、审计数据加密等。

⑦依从性管理(compliance management)包括确保数据存储和使用等符合相关的风险管理 and 安全管理的规定要求。

4. 云计算的应用

云计算的应用主要有以下几个方面。

(1)云物联。随着物联网业务量的增加,对数据存储和计算量的需求将带来对云计算能力的要求。

(2)云安全(cloud security)。云安全是一个从云计算演变而来的新名词。云安全的策略构想是:使用者越多,每个使用者就越安全,因为如此庞大的用户群,足以覆盖互联网的每一个角落,只要某个网站被挂马或某个新木马病毒出现,就会立刻被截获。云安全通过网状的大量客户端对网络中软件行为的异常监测,获取互联网中木马、恶意程序的最新信息,推送到服务器端进行自动分析和处理,再把病毒和木马的解决方案分发到每一个客户端。

(3)云存储。云存储是在云计算概念上延伸和发展出来的一个新的概念,是指通过集群应用、网格技术或分布式文件系统等功能,将网络中大量各种不同类型的存储设备通过应用软件集合起来协同工作,共同对外提供数据存储和业务访问功能的一个系统。当云计算系统运算和处理的核心是大量数据的存储和管理时,云计算系统中就需要配置大量的存储设备,那么云计算系统就转变成为一个云存储系统,所以云存储是一个以数据存储和管理为核心的云计算系统。

1.7.2 物联网

1. 物联网的定义

物联网是新一代信息技术的重要组成部分,其英文名称是 The Internet of things。顾名思义,物联网就是物物相连的互联网。这有以下两层意思。

(1)物联网的核心和基础仍然是互联网,是在互联网基础上延伸和扩展的网络。

(2)物联网的用户端延伸和扩展到了任何物品与物品之间,进行信息交换和通信。

因此,物联网的定义是通过射频识别(RFID)、红外感应器、全球定位系统、激光扫描器等信息传感设备,按约定的协议把任何物品与互联网相连接,进行信息交换和通信,以实现物品的智能化识别、定位、跟踪、监控和管理的一种网络。图 1-5 所示就是物联网与世界的形象描述。

2. 物联网技术的特征

和传统的互联网相比,物联网有其鲜明的特征。

(1)它是各种感知技术的广泛应用。物联网上部署了海量的多种类型传感器,每个传感器都是一个信



图 1-5 物联网与世界

息源,不同类别的传感器所捕获的信息内容和信息格式不同。传感器获得的数据具有实时性,按一定的频率周期性地采集环境信息,不断更新数据。

(2)它是一种建立在互联网上的泛在网络。物联网技术的重要基础和核心仍旧是互联网,通过各种有线和无线网络与互联网融合,将物体的信息实时准确地传递出去。在物联网上的传感器定时采集的信息需要通过网络传输,由于其数量极其庞大,形成了海量信息,在传输过程中,为了保障数据的正确性和及时性,必须适应各种异构网络和协议。

(3)物联网不仅仅提供了传感器的连接,其本身也具有智能处理的能力,能够对物体实施智能控制。物联网将传感器和智能处理相结合,利用云计算、模式识别等各种智能技术,扩充其应用领域。从传感器获得的海量信息中分析、加工和处理出有意义的数据,以适应不同用户的不同需求,发现新的应用领域和应用模式。

3. 具备物联网的条件

这里的物要满足以下条件才能够被纳入物联网的范围。

- (1)要有数据传输通路。
- (2)要有一定的存储功能。
- (3)要有 CPU。
- (4)要有操作系统。
- (5)要有专门的应用程序。
- (6)遵循物联网的通信协议。
- (7)在世界网络中有可被识别的唯一编号。

4. 物联网的应用

随着物联网技术的发展,其应用已经扩展到社会各个方面,如物流、交通、安防、能源、医疗、建筑、制造、家居、零售、农业等。

(1)智慧物流。智慧物流是新技术应用于物流行业的统称,指的是以物联网、大数据、人工智能等信息技术为支撑,在物流的运输、仓储、包装、装卸、配送等各个环节实现系统感知、全面分析及处理等功能。智慧物流的实现能大大地降低各行业运输的成本,提高运输效率,提升整个物流行业的智能化和自动化水平。物联网应用于物流行业中,主要体现在三方面,即仓储管理、运输监测和智能快递柜。

(2)智能交通。交通被认为是物联网所有应用场景中最有前景的应用之一。而智能交通是物联网的体现形式,利用先进的信息技术、数据传输技术及计算机处理技术等,通过集成到交通运输管理体系中,使人、车和路能够紧密配合,改善交通运输环境、保障交通安全及提高资源利用率。行业内应用较多的五大场景包括智能公交车、共享单车、汽车联网、智慧停车、智能红绿灯、高速无感收费等。

(3)智能安防。智能安防的核心在于智能安防系统,系统主要包括门禁、报警和监控三大部分。安防是物联网的一大应用市场,传统安防对人员的依赖性比较大,非常耗费人力,而智能安防能够通过设备实现智能判断。目前,智能安防最核心的部分在于智能安防系统,该系统是对拍摄的图像进行传输与存储,并对其进行分析与处理。一个完整的智能安防系统主要包括三大部分,即门禁、报警和监控,行业中主要以视频监控为主。

由于采集的数据量足够大且延时较低,因此,目前城市中大部分的视频监控采用的是有

线连接方式,而对于偏远地区及移动性的物体监控则采用 4G 等无线技术。

(4)智能能源。物联网应用于能源领域,可用于水、电、燃气等表计及路灯的远程控制上。智慧能源属于智慧城市的一个部分,当前,将物联网技术应用在能源领域,主要用于水、电、燃气等表计及根据外界天气对路灯的远程控制等,基于环境和设备进行物体感知,通过监测,提升利用效率,减少能源损耗。根据实际情况,智慧能源分为四大应用场景:智能水表、智能电表、智能燃气表和智能路灯。

(5)智能医疗。智能医疗的两大主要应用场景为医疗可穿戴和数字化医院。在智能医疗领域,新技术的应用必须以人为中心。而物联网技术是数据获取的主要途径,能有效地帮助医院实现对人的智能化管理和对物的智能化管理。对人的智能化管理指的是通过传感器对人的生理状态(如心跳频率、体力消耗、血压高低等)进行捕捉,将它们记录到电子健康文件中,方便个人或医生查阅。对物的智能化管理指的是通过 RFID 技术对医疗物品进行监控与管理,实现医疗设备、用品可视化。

(6)智能建筑。物联网应用于建筑领域主要体现在用电照明、消防监测及楼宇控制等。建筑是城市的基石,技术的进步促进了建筑的智能化发展,物联网技术的应用让建筑向智慧建筑方向演进。智慧建筑越来越受到人们的关注,是集感知、传输、记忆、判断和决策于一体的综合智能化解决方案。当前的智慧建筑主要体现在用电照明、消防监测及楼宇控制等,将设备进行感知、传输并远程监控,不仅能够节约能源,同时也能减少运维的楼宇人员。而对于古建筑,也可以进行白蚁(以木材为生的一种昆虫)监测,进而达到保护古建筑的目的。

(7)智能制造。物联网技术赋能制造业,实现工厂的数字化和智能化改造。制造领域的市场体量巨大,是物联网的一个重要应用领域,主要体现在数字化及智能化的工厂改造上,包括工厂机械设备监控和工厂的环境监控。通过在设备上加装物联网装备,使设备厂商可以远程随时随地对设备进行监控、升级和维护等操作,更好地了解产品的使用状况,完成产品全生命周期的信息收集,指导产品设计和售后服务;而厂房的环境监控主要包括空气温湿度、烟感等情况。

数字化工厂的核心特点是:产品的智能化、生产的自动化、信息流和物资流合一。目前,从世界范围看,还没有一家企业宣布建成一座完全数字化的工厂。近些年来,一些企业开始给行业内其他企业提供以生产环节为基础的数字化和智能化工厂改造方案。企业的数字化和智能化改造大体分成 4 个阶段:自动化生产线与生产装备,设备联网与数据采集、数据的打通与直接应用、数据智能分析与应用。这 4 个阶段并不按照严格的顺序进行,各阶段也不是孤立的,边界较模糊。

(8)智能家居。智能家居的发展分为 3 个阶段,单品连接、物物联动及平台集成,当前处于单品向物物联动过渡阶段。智能家居指的是使用各种技术和设备,来提高人们的生活方式,使家庭变得更舒适、安全和高效。物联网应用于智能家居领域,能够对家居类产品的位置、状态、变化进行监测,分析其变化特征,同时根据人的需要,在一定程度上进行反馈。

(9)智能零售。智能零售依托于物联网技术,主要体现了两大应用场景,即自动售货机和无人便利店。行业内将零售按照距离分为 3 种不同的形式:远场零售、中场零售和近场零售,三者分别以电商、商场/超市和便利店/自动售货机为代表。物联网技术可以用于近场和中场零售,且主要应用于近场零售,即无人便利店和自动(无人)售货机。

智能零售通过将传统的售货机和便利店进行数字化升级、改造,打造无人零售模式。通过数据分析,并充分运用门店内的客流和活动,为用户提供更好的服务,为商家提供更高的

经营效率。

(10)智慧农业。智慧农业指的是利用物联网、人工智能、大数据等现代信息技术与农业进行深度融合,实现农业生产全过程的信息感知、精准管理和智能控制的一种全新的农业生产方式,可实现农业可视化诊断、远程控制、灾害预警等功能。农业分为农业种植和畜牧养殖两个方面。农业种植分为设施种植(温室大棚)和大田种植,主要包括播种、施肥、灌溉、除草和病虫害防治五个部分,以传感器、摄像头和卫星等收集数据,实现数字化和智能机械化发展。当前,数字化的实现多以数据平台服务来呈现,而智能机械化以农机自动驾驶为代表。畜牧养殖主要是将新技术、新理念应用在生产中,包括繁育、饲养、疾病防疫等,且应用类型较少,因此,用精细化养殖定义整体畜牧养殖环节。

1.7.3 大数据

1. 大数据的定义

大数据又称巨量资料,指的是所涉及的数据资料量规模巨大到无法通过人脑甚至主流软件工具,在合理时间内达到抽取、管理、处理,并整理成为帮助企业经营决策更积极目的的资讯。

简而言之,大数据就是数据量非常大、数据种类繁多、无法用常规归类方法应用计算的数据集成。

大数据具有四大特点,即 volume(容量)代表海量的数据规模;variety(种类)代表数据类型的多样性;value(价值)代表深度的数据价值;velocity(速度)代表数据流转的迅速与体系的动态性,如图 1-6 所示。

(1)volume(数据体量巨大)。目前,人类社会所生产的印刷材料的总和和数据量大约是 200 PB(1 PB=210 TB),人类说过的语言的总和和数据量大约是 5 EB(1 EB=210 PB)。数据的体量决定了它背后的信息价值,随着各种移动端的流行,云存储技术的发展,现代社会的人类活动都可以被记录下来,因此产生了海量的数据。发送的微博、自拍的照片、穿戴的运动手环等通过互联网上传到云端,各种数据聚集到特定地点的存储系统,如政府等机构,最终形成了体量巨大的大数据。



图 1-6 大数据的特点

(2)variety(数据类型呈多样性)。数据类型的多样性使得数据被分为结构化数据与非结构化数据两种。而互联网将网络通过各种移动端形成了整体,人们不仅可以通过互联网获取数据,也是数据的传播者,相对于过去,以文本为主的结构化数据往往是便于存储的,随着非结构化的数据越来越多,如网络小说、拍摄的视频、录制的音频、共享的地理位置等,这些多样性的数据使得对数据处理的能力要求更高。需要对数据进行加工、清洗、分析等步骤,将它们变为易于存储的结构化数据。这需要在海量的数据之间发现它们之间的关联性,把看似毫无关系的数据联系起来,形成有价值的信息。

(3)velocity(处理迅速)。处理迅速是大数据区别于传统数据挖掘的最显著特征,大数据具有实时性。例如,人们出去吃饭,导航餐厅,用移动端的地图查询位置,选择避免堵车的路线,还会从网络上查看餐厅的口味评价如何,吃饭时,也许会拍下食物和餐厅的照片上传到微博,因此,各种网络的连接带来大量的数据交换,对速度的要求更高,要以实时的方式传达给用户。

(4)value(数据价值)。大数据的应用在物联网、云计算、大数据挖掘等技术的迅速发展

带动下,呈现出它的过程:把数据源的信号转换为数据,再把大数据加工成信息,通过获取的信息来做决策。因此,大数据价值的挖掘过程就像大浪淘沙,数据的体量越大,相对有价值的的数据就越少。

2. 大数据的关键技术

大数据的关键性技术主要分为4种:流处理、并行化、可视化和摘要索引。

(1)流处理。随着公司的业务处理流程越来越复杂,流式数据处理技术已成为大数据的重要处理技术,能满足实时处理的数据处理需求。

例如,传统的方法只能计算已经给出具体数据的一组数据的平均值,如果数据是移动的,这样的平均值计算则需要大数据的流处理方法,创建一个数据流统计集,逐步添加数据块,进行移动平均值计算。

(2)并行化。小数据的存储能力通常不到10 GB,中数据的存储能力不到1 TB,大数据的存储则是分布于多台机器上,多达以PB为单位的数据,在分布式数据条件下,需要在极短的时间内处理数据,这时就需要进行并行化处理。

(3)可视化。数据可视化分为信息可视化和科学可视化两种。可视化工具是实现可视化的必要手段,常见的可视化工具有以下两类。

①管理决策者或者数据分析师可以利用探索性可视化工具找出数据之间的关联性,这是可视化工具的洞察力作用,如Tableau、TIBCO QlikeView等。

②叙事性可视化工具挖掘数据的方式较为独特。例如,需要用叙事性可视化工具查看某个时间段内某企业的营销数据,可视化格式将预先被创建,数据会按照时间点被逐年显示,并按照设定的条件排序。

(4)摘要索引。摘要索引是加速查询数据的预计算摘要的过程,这个预计算摘要会被预先创建。摘要索引的作用是为将要进行的查询做计划。现在摘要索引尚没有一个明确的规则,但随着大数据技术的发展,这一问题将会得到解决。

3. 大数据的社会价值

大数据因其背后的价值,被比喻为21世纪的黄金,被看作新兴的经济元素,大数据不仅本身可以看作重要的生产要素,其对产品的形成过程也起到至关重要的作用。大数据的主要价值有如下几个方面。

(1)大数据是新时代信息技术的关键支撑。大数据的热潮在全球的盛行,顺应了现代信息技术发展的趋势。互联网时代为大数据的普及和发展打下了坚实的基础,人们能随时通过移动端使用互联网,伴随着物联网、网上购物、交友网站、云计算的兴起,每个人的数据无处不在,随时随地产生。作为信息技术时代的产物,大数据的应用又反作用于信息技术的发展,促进物联网、云计算等技术的革新,大数据作为融合现代信息技术的关键支撑,为物联网、云计算等现代信息技术的发展提供了依托的平台。

(2)大数据是促进现代社会经济发展的推动力。大数据本身隐含着巨大的经济价值和社会价值。大数据行业的爆发式发展,将带来一批针对大数据市场的新的商业理念、新的营销服务、新的产品和新的技术,推动现代信息产业的加速发展。在国内的城市建设、民生发展等领域,大数据也起着举足轻重的作用。目前,我国着力推行智慧城市的建设,大数据的应用能将城市中方方面面的数据联合起来,分析提取有效数据,依靠它们做出智慧决策。在

红绿灯的时间设置上,可以依照不同的时间段、该条道路的车行流量、拥堵状况数据分析,来合理设置红绿灯的时间,缓解交通难题。伴随着智慧城市在我国不断建设和完善,大数据的出色表现在提升地方政务能力和社会管理能力方面发挥着重要作用,使之形成充满各地特色的、新兴的智能领域应用。

大数据帮助企业深度挖掘客户喜好,助力企业智能决策。大数据对企业洞察用户提供了有力的武器,满足企业针对客户的个性化营销需求,为企业做出正确的市场决策提供更多维度。大数据出现以前,市场调查是通过人工获取方式,采用调研和营销实现,这样的数据具有明显的“人工计划”特征,在市场调查之前,收集数据的样板、调研方式、分析方式和获取数据的目的都有一个清晰的规划,因此,这些产生的数据是“结构化”的。而依靠互联网产生的大数据,其来源是互联网用户行为,包括网页检索、页面浏览、网络交易和网络社交行为等,它并不受人工控制,因此数据的产生、分析过程具有不确定性,这样的大数据是多维度的,360°全方位接近用户,让决策的依据更科学。

(3)大数据将成为科技创新的引擎。人工数据时代,信息化的滞后使得大量的数据彼此分离,闲置在各自的系统空间,技术的落后让传统的信息处理方式无法满足科技发展的需求。新兴的大数据在整合数据、分析数据、存储数据、处理数据、应用数据,解决系统实时性的、并发性的问题,包括云存储、数据价值分析等方面都颠覆了传统。大数据成为各个领域科技创新的引擎。例如,大型家电生产厂家,在产品生产线安装上传感器采集数据,这些生产信息的分析和价值挖掘,能实时提高产品合格率。在电力领域,智能电表的数据采集同样发挥着不可忽视的作用,不仅作为电费收取的依据,还扮演着判断房屋空置与否的角色,延伸开来,可作为城市房价定位的参考指标。再者,电网所采集的耗电量数据可以判断出该部分地区的商业发展情况。在未来,不论是国家政府,还是金融商业,各个数据集中的领域,大数据将成为各企业、各单位、提升竞争力、占领市场的核心,加速企业从“业务驱动”向“数据驱动”转型升级,为企业提高利润、增强实力、研发产品带来新的机遇。

4. 大数据的应用

(1)大数据经典案例。在大数据时代,如何通过大数据的应用来实现企业价值,是很多企业思考的问题,下面介绍一些利用大数据创造价值的成功案例。

①啤酒与尿布。全球零售业巨头沃尔玛在对消费者购物行为分析时发现,男性顾客在购买婴儿尿片时,常常会顺便搭配几瓶啤酒来犒劳自己,于是尝试推出了将啤酒和尿布摆在一起的促销手段。没想到这个举措居然使尿布和啤酒的销量都大幅增加了。如今,“啤酒+尿布”的数据分析成果早已成了大数据技术应用的经典案例,被人津津乐道。

②数据新闻让英国撤军。2010年10月23日,《卫报》利用维基解密的数据做了一篇“数据新闻”。将伊拉克战争中所有的人员伤亡情况均标注于地图之上。地图上一个红点便代表一次死伤事件,用鼠标单击红点后弹出的窗口则有详细的说明:伤亡人数、时间,造成伤亡的具体原因。密布的红点多达39万,显得格外触目惊心,一经刊出立即引起朝野震动,推动英国最终做出撤出驻伊拉克军队的决定。

③Google成功预测冬季流感。2009年,Google通过分析5000万条美国人最频繁检索的词汇,将之与美国疾病中心在2003年到2008年间季节性流感传播时期的数据进行比较,并建立一个特定的数学模型。最终Google成功预测了2009冬季流感的传播甚至可以具体到特定的地区和州。

④大数据与乔布斯癌症治疗。乔布斯是世界上第一个对自身所有DNA和肿瘤DNA进行

排序的人。为此,他支付了高达几十万美元的费用。他得到的不是样本,而是包括整个基因的数据文档。医生按照所有基因按需下药,最终这种方式帮助乔布斯延长了好几年的生命。

⑤奥巴马大选连任成功。2012年11月奥巴马大选连任成功的胜利果实也被归功于大数据,因为他的竞选团队进行了大规模与深入的数据挖掘。《时代杂志》更是断言,依靠直觉与经验进行决策的优势急剧下降,在政治领域,大数据的时代已经到来;各色媒体、论坛、专家铺天盖地的宣传让人们对于大数据时代的来临兴奋不已,无数公司和创业者都纷纷跳进了这个狂欢队伍。

⑥微软大数据成功预测奥斯卡21项大奖。2013年,微软纽约研究院的经济学家大卫·罗斯柴尔德(David Rothschild)利用大数据成功预测24个奥斯卡奖项中的19个,成为人们津津乐道的话题。2014年罗斯柴尔德再接再厉,成功预测第86届奥斯卡金像奖颁奖典礼24个奖项中的21个,继续向人们展示现代科技的神奇魔力。

(2)大数据的操作实例。

①中信银行信用卡营销。

a. 实施背景。中信银行信用卡中心是国内银行业为数不多的几家分行级信用卡专营机构之一,也是国内具有竞争力的股份制商业银行信用卡中心之一。近些年来,中信银行信用卡中心的发卡量迅速增长,2008年银行向消费者发卡约500万张,而这个数字在2010年增加了1倍。随着业务的迅猛增长,业务数据规模也急剧膨胀。中信银行信用卡中心无论在数据存储、系统维护等方面,还是在有效利用客户数据方面,都面临着越来越大的压力。

中信银行信用卡中心迫切需要一个可扩展、高性能的数据仓库解决方案,支持其数据分析战略,提升业务的敏捷性。通过建立以数据仓库为核心的分析平台,实现业务数据集中和整合,以支持多样化和复杂化的数据分析,如卡、账户、客户、交易等主题的业务统计和OLAP(联机分析处理)多维分析等,提升卡中心的业务效率;通过从数据仓库提取数据,改进和推动有针对性的营销活动。

b. 技术方案。从2010年4月到2011年5月,中信银行信用卡中心实施了EMC Greenplum数据仓库解决方案。实施EMC Greenplum解决方案之后,中信银行信用卡中心实现了近似实时的商业智能(BI)和秒级营销,运营效率得到全面提升。中信银行大数据应用技术架构图如图1-7所示。

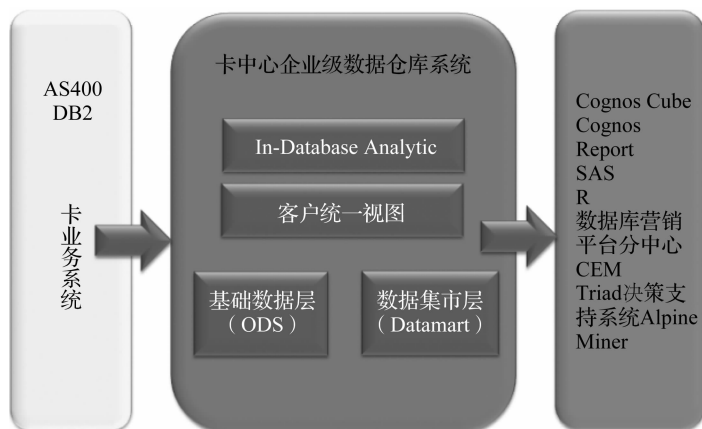


图 1-7 中信银行大数据应用技术架构图

Greenplum 解决方案的一个核心功能是,它采用了“无共享”的开放平台的 MPP 架构,此架构是为 BI 和海量数据分析处理而设计的。目前,最普遍的关系数据库管理系统(如 Or-

acle 或 Microsoft SQL Server), 都是利用“共享磁盘”架构来实现数据处理的, 会牺牲单个查询性能和并行性能。而使用 Greenplum 数据库提供的 MPP 架构, 数据在多个服务器区间会自动分区, 各分区拥有并管理整体数据的不同部分; 所有的通信是通过网络互连完成, 没有磁盘级共享或连接, 使其成为一个“无共享”架构。Greenplum 数据库提供的 MPP 架构为磁盘的每一个环节提供了一个专门的、独立的高带宽通道, 段上的服务器可以以一个完全并行的方式处理每个查询, 并根据查询计划在段之间有效地移动数据, 因此, 相比普通的数据库系统, 该系统提供了更高的可扩展性。

c. 效益提升。2011年, 中信银行信用卡中心通过其数据库营销平台进行了 1 286 个宣传活动, 每个营销活动配置平均时间从 2 周缩短到 2~3 天, 且市场活动中答应客户在刷满一定金额或次数后送给他们的礼品, 可以在客户刚好满足条件的那次刷卡后马上获得, 实现了秒级营销, 而不必像之前那样等待好几个工作日。2011 年的前 3 个季度, 中信银行信用卡中心交易量增加 65%, 比股份制商业银行的平均水平高 14%, 比国内所有银行的平均值高 4%。中信银行信用卡中心迄今已为客户进行了 4 000 万次的信用额度调整。中信银行信用卡中心催收管理团队使用了基于数据仓库的 FICO TRIAD 系统后, 信用卡不良贷款 (NPL) 比率同比减少了 0.76%。中信银行信用卡中心电话销售中心将所有外呼营销历史整合到数据仓库, 通过对大量历史数据分析后调整客户提取和营销策略, 在上线后的第一个月便实现单位工时创收提升 33%、笔均贷款额提升 18%, 目前中信银行正在开发针对每个产品的营销响应模型, 以进一步提升产能。

② 兴业证券客户综合分析管理系统。

a. 实施背景。随着我国证券市场的日益规范和成熟, 证券公司之间的竞争也日趋激烈。证券公司越来越注重对客户的有效服务, 以及对营业部、经纪人的业绩管理, 而现有的 IT 系统通常只是面向业务交易而设计的, 随着市场竞争的日益激烈, 越来越不能满足证券公司的决策分析需要。为了提升证券公司的客户服务及精准营销的能力, 兴业证券采用大数据技术提升自身客户综合分析管理系统能力。

b. 技术方案。吉贝克针对兴业证券所面临的环境, 采用数据仓库和数据挖掘技术, 自主研发了证券公司客户综合分析管理系统, 以满足证券公司日益深化的客户管理需求。客户综合分析管理系统功能架构图和客户生命周期服务管理示意图如图 1-8 和图 1-9 所示。



图 1-8 客户综合分析管理系统功能架构图

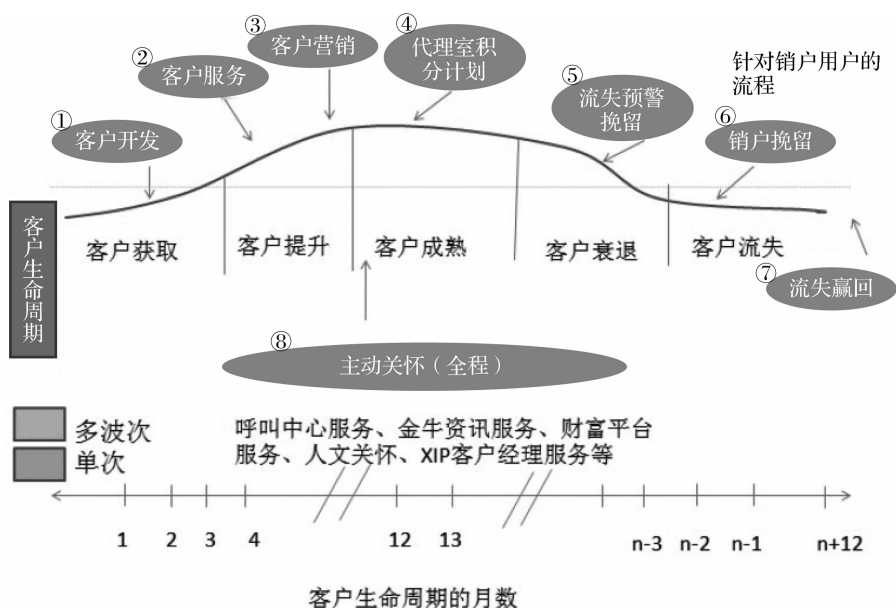


图 1-9 客户生命周期服务管理示意图

围绕客户维护生命周期,在不同的生命周期阶段采用有针对性的方式来降低客户流失率,如主动关怀、客户营销、流失预警挽留、销户挽留等。

c. 效益提升。采用上述系统后,数据加载速度明显提升,目前 100 万行数据入库仅需 6~7 秒,10 GB 的数据加载和导出也可以在 5 分钟内完成。同时,数据处理和查询的效率也显著提升,每天的数据处理时间基本控制在 2 小时以内。对于日常简单查询,在 50 条并发查询的情况下可以实现 1 秒内完成。对于长时间跨度、多条件的复杂查询,也能在 5 秒内完成。

1.7.4 移动互联网

1. 移动互联网的定义

尽管移动互联网是目前 IT 领域最热门的概念之一,然而业界并未就其定义达成共识。这里先介绍几种有代表性的移动互联网的定义。

第一种定义:移动互联网是一种通过智能移动终端,采用移动无线通信方式获取业务和服务的新兴业态,包含终端、软件和应用 3 个层面。终端层包括智能手机、平板电脑、电子书、MID 等;软件层包括操作系统、中间件、数据库和安全软件等;应用层包括休闲娱乐类、工具媒体类、商务财经类等不同应用与服务。

第二种定义:移动互联网是指通过无线智能终端(如智能手机、平板电脑等)使用互联网提供的应用和服务(包括电子邮件、电子商务、即时通信等),保证随时随地的无缝连接的业务模式。

第三种定义:移动互联网利用手机、PDA 或其他手持终端通过各种无线网络进行数据交换。

第四种定义:中兴通讯从通信设备制造商的角度给出了定义,即狭义的移动互联网是指

用户能够利用手机、PDA 或其他手持终端通过无线通信网络接入互联网,广义的移动互联网是指用户能够利用手机、PDA 或其他手持终端以无线的方式通过各种网络(如 WLAN、BWL、GSM、CDMA 等)来接入互联网。可以看到,对于通信设备制造商来说,网络是其看待移动互联网的主要切入点。

第五种定义:移动互联网有广义和狭义之分。广义的移动互联网是指用户可以使用手机、笔记本电脑等移动终端通过协议接入互联网。狭义的移动互联网则是指用户使用手机终端通过无线通信的方式访问采用 WAP(无线通信协议)的网站。

认可度比较高的定义是中华人民共和国工业和信息化部电信研究院在 2011 年的《移动互联网白皮书》中给出的:移动互联网是以移动网络作为接入网络的互联网及服务。它包括 3 个要素:移动终端、移动网络和应用服务。该定义将移动互联网涉及的内容主要概括为以下 3 个层面。

- (1)移动终端,包括手机、专用移动互联网终端和采用数据卡方式的便携计算机等。
- (2)移动通信网络接入,包括 2G、3G、4G 甚至 5G 等。
- (3)公众应用服务,包括 Web、WAP 方式等。

移动终端是移动互联网的前提,移动网络是移动互联网的基础,而应用服务则成为移动互联网的核心。

上述定义给出了移动互联网两方面的含义:一方面,移动互联网是移动通信网络与互联网的融合,用户以移动终端接入无线移动通信网络的方式访问互联网;另一方面,移动互联网还产生了大量新型的应用,这些应用与终端的可移动、可定位和随身携带等特性相结合,为用户提供个性化的与位置相关的服务。

综合以上观点,我们也提出一个参考性定义:移动互联网是指以各种类型的移动终端作为接入设备,使用各种移动网络作为接入网络,从而实现包括传统移动通信、传统互联网及其各种融合创新服务的新型业务模式。

2. 移动互联网的主要特征

移动互联网的主要特征有如下几个。

(1)终端移动性。相对于固定互联网,移动互联网灵活、便捷、高效。移动终端体积小且易于携带;移动互联网里包含了各种适合移动应用的各类信息,用户可以随时随地地进行采购、交易、质询、决策、交流等各类活动。移动性带来接入便捷、无所不在的连接及精确的位置信息,而位置信息与其他信息的结合蕴藏着巨大的业务潜力。

(2)个性化。移动互联网创造了一种全新的个性化服务理念和商业运作模式。对于不同用户群体和个人的不同爱好与需求,为他们量身定制出多种差异化的信息,并通过不受时空地域限制的渠道,随时随地传送给用户。终端用户可以自由自在地控制所享受服务的内容、时间和方式等,移动互联网充分体现了个性化的服务。

(3)业务及时性。用户使用移动互联网能够随时随地获取自身或其他终端的信息,及时获取所需的服务和数据。

(4)服务便利性。由于移动终端的限制,移动互联网服务要求操作简便,响应时间短。

(5)业务使用的私密性。在使用移动互联网业务时,所使用的内容和服务更私密,如手机支付业务等。

(6)融合性。首先,移动语音和移动互联网业务的一体化促进了业务融合;其次,手机终

端趋向于变成人们随身携带的唯一的电子设备,其功能集成度越来越高。

移动互联网相比于传统固定互联网的优势在于:实现了随时随地的通信和服务获取,具有安全、可靠的认证机制,能够及时获取用户及终端信息,业务端到端流程可控等。

移动互联网的劣势主要包括:无线频谱资源的稀缺性;用户数据安全和隐私性;移动终端软硬件缺乏统一标准,业务互通性差等。

总之,移动互联网继承了桌面互联网的开放协作的特性,又继承了移动网的实时性、隐私性、便携性、准确性和可定位的特点,如图 1-10 所示。

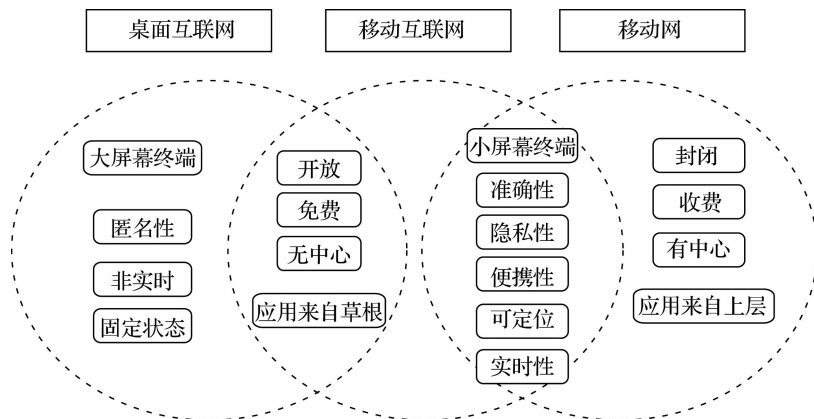


图 1-10 移动互联网的特征

3. 移动互联网的发展

移动互联网的发展可分为以下 4 个主要阶段。

(1) 萌芽期(2000—2007 年)。中国移动互联网发展的初级阶段。2000 年 11 月 10 日,中国移动推出“移动梦网计划”,打造开放、合作、共赢的产业价值链。2002 年 5 月 17 日,中国电信在广州启动的“互联星空”计划,标志着互联网服务提供商(Internet service provider, ISP)和互联网内容提供商(Internet content provider, ICP)开始联合打造宽带互联网产业。2002 年 5 月 17 日,中国移动率先在全国范围内正式推出 GPRS 业务,这个阶段的主要产品有文字信息、图案及铃声。

该时期主要受限于移动 2G 网速和手机智能化程度,中国移动互联网处于一个简单的无线应用协议(wireless application protocol, WAP)应用期。WAP 应用把 Internet 网上 HTML 的信息转换成用无线标记语言(wireless markup language, WML)描述的信息,显示在移动电话的显示屏上。WAP 时期,用户主要在移动互联网上看新闻、读小说、听音乐,这是一个内容为王的移动互联网时代,这个阶段开始出现移动互联网产品经理,如服务提供商(service provider, SP)、产品经理或 WAP 产品经理等。

(2) 成长培育期(2008—2011 年)。随着 3G 的应用,新浪微博等社交网络、基于位置的服务(location based service, LBS)的应用,iPhone 的移动 App、互联网电子商务在手机上广泛应用,互联网上的应用移植开始出现了一个新的名词 SoLo MoCo,即 social(社交的)、local(本地的)、mobile(移动的)和 commercial(商务化)。在这个阶段,移动互联网产品经理得到进一步发展,呈现出发展态势,逐渐受到重视,有的公司设立专门的移动终端部门,负责公司产品在移动终端上的战略布局和发展。

在此期间,各大互联网公司都在摸索如何抢占移动互联网入口,百度、腾讯、奇虎 360 等

一些大型互联网公司都企图推出手机浏览器来抢占移动互联网入口；新浪、优酷等互联网公司则通过与手机制造商合作，在智能手机出厂时就把企业服务应用（如新浪微博、视频播放器等应用）预先安装在手机中。

(3) 高速成长期(2012—2013年)。进入2012年之后，由于移动上网需求大增，iOS、Android(安卓，见图1-11)等智能操作系统的大规模商业化应用，传统功能手机进入一个全面升级换代期，以三星、HTC等为代表的传统手机厂商纷纷效仿苹果模式，普遍推出了触摸屏智能手机和手机应用商店。具有触摸功能的智能手机的大规模普及和应用解决了传统键盘机上网的诸多不便，安卓智能手机操作系统的普遍安装和手机应用程序商店的出现极大地丰富了手机上网功能，移动互联网应用呈现出爆发式增长。



图 1-11 智能手机操作系统

智能手机的大规模普及和应用，激发了手机越过竞争对手(over the top, OTT)应用，以微信为代表的手机移动应用开始呈现大规模爆发式增长。

小米等互联网公司更是创新了智能手机的营销模式，打出了不靠手机硬件、靠手机服务挣钱的“智能手机+互联网服务”新商业模式，依托高性价比的智能手机载体来加大公司互联网服务应用的推广力度。

另外，滴滴出行、今日头条等一大批基于移动互联网应用服务创新和商业模式创新的应用在此期间大量涌现，极大地激发了投资界对移动互联网应用的投资兴趣。

(4) 全面发展期(2014年至今)。随着4G网络的部署，移动上网网速得到极大提高，上网网速瓶颈限制得到根本破除，移动应用场景得到极大丰富。2013年12月4日，中华人民共和国工业和信息化部正式向中国移动、中国电信和中国联通三大运营商发放了TD-LTE-4G牌照，中国4G网络正式大规模铺开。2015年2月27日，中华人民共和国工业和信息化部又向中国联通发放“LTE/第四代数字蜂窝移动通信业务(LTE-FDD)”经营许可证。4G网络建设让中国移动互联网发展走上了快速发展轨道。

由于网速、上网便捷性、手机应用等移动互联网发展外在环境基本得到全部解决，移动互联网应用开始全面发展。在桌面互联网时代，门户网站是企业开展业务的标配。在移动互联网时代，手机App应用是企业开展业务的标配，4G网络催生了许多利用移动互联网开展业务的公司，特别是由于4G网速大大提高和5G技术的推进，促进了实时性要求更高、流量需求较大类型移动应用的快速发展，许多手机应用开始大力推广移动视频应用，涌现出了秒拍、快手、抖音等一大批基于移动互联网的手机视频和直播应用。

4. 移动互联网的应用

产品存在的意义在于满足人的需求，而人的需求符合马斯洛理论，也就是说，人的需求有一个从低到高的发展层次，低层次的需要是生理需要，向上依次是安全、爱与归属、被尊重和自我实现的需要。其实，我们可以看到国内互联网巨头满足的就是人们的这些基本需求。

互联网是这样，移动互联网同样如此。在互联网时代被验证的商业模式有广告、游戏、

电子商务等；而在移动互联网时代，目前被验证的有游戏和广告等。对于游戏而言，其是大众娱乐需求，对于不同设备，只要用户愿意使用这些不同的设备，游戏的需求将一直存在；对于电子商务，其在移动设备上更像是互联网的一个延伸，电子商务站其实更像是内容站，对于内容站而言，不同设备只是终端而已。

移动互联网应用到目前为止主要体现在智能终端的应用。由于智能终端除了通话功能外还有 Wi-Fi、GPS、速度传感器、NFC、相机、多点触控、手写输入等诸多先进功能，所以基于这些功能的应用就应运而生了。其主要体现在以下几个方面。

(1) 移动游戏。随着智能终端 CPU 处理能力及手机内置显卡功能的增强，智能手机及智能终端的显示效果与之前相比不可同日而语。现在国内的移动互联网游戏发展非常快速，据网络调查数据显示，截至 2020 年 3 月，我国手机游戏用户规模达 5.29 亿。从全球市场来看，游戏依然是智能机平台最受欢迎的应用，美国知名市场调查机构尼尔森(Nielsen)针对美国智能机用户进行了一项调查，结果显示，64%的用户下载了游戏应用程序，居用户下载应用程序比例首位，这些用户平均每月玩游戏的时间达 8 小时，这些数据显示出来的是继计算机游戏后，手机游戏背后出现的诱人的潜在市场与价值。

(2) 移动视频。移动视频业务是通过移动网络和移动终端为移动用户传送视频内容的新型移动业务。随着 4G 网络的部署和终端设备性能的提高，使用移动视频业务的用户越来越多。苹果公司在发布备受全球关注的第四代 iPhone 时出现了一个小插曲：当乔布斯在为现场观众演示 iPhone 4 的视频通话功能时，由于网络拥塞引起该项业务无法进行演示，以至于乔布斯不得不要求观众暂时关闭手机。这从侧面反映出视频流的迅猛增长给通信网络带来了巨大挑战，同时也说明了越来越多的用户在使用移动视频业务。目前，国内各大视频网站都推出了基于智能移动终端的应用。

(3) 移动社交。移动社交是指用户以手机、平板电脑等移动终端为载体，以在线识别用户及交换信息技术为基础，按照流量计费，通过移动网络来实现的社交应用功能。移动社交不包括打电话、发短信等通信业务。与传统的 PC 端社交相比，移动社交具有人机交互、实时场景等特点，能够让用户随时随地地创造并分享内容，让网络最大程度地服务于人们的现实生活。

移动社交的形式有以下几种。

① 文字。文字作为最早的移动社交的形式具备简单、快速、及时等特点，在即时通信领域，文字的地位不可动摇，即使在移动社交高速发展的今天，其依然是人们最为常用的沟通形式。虽然新型的交流模式不断出现，但基于文字而形成的传统移动社交形式依然会保持长久优势。

② 声音。很难想象，基于手机应运而生的移动社交，声音作为传播载体正在被人们广泛使用。与文字不同，声音在进行沟通时让人感觉更为直观、清晰。无论是国外的 Kiki 还是国内的微信，其语音使用在移动社交领域都已经很普及。

③ 图片。图片社交形式的发展很大程度上要归功于智能手机拍照功能的逐步强大及 Instagram 等应用的成功。在以文字为核心的第一代社交网络引爆市场后，用户急需一个专注于图片分享的移动社交形式来更加简单、生动地分享生活碎片。

④ 视频。视频作为移动化分享的传播形式，其发展与前 3 种形式相比还有很大差距，这主要归因于其本身对于网络及硬件的要求较高。国外有很多应用正在专注于 15 秒视频分

享的移动社交网络,如 Shoutz 等。

随着智能手机的高速发展,社交网络已开始向移动化转移,以 Facebook 为代表的业界巨头正围绕着移动领域进行布局。而在国内,腾讯开始加快对该领域的布局。与传统社交应用相比,移动社交多了两个非常重要的特性,即实时性和位置性。移动智能手机便于随身携带,而绝大多数智能手机都有 GPS(全球定位系统)功能,针对这两个特性而开发的社交应用不胜枚举,如米聊、微信、陌陌等。

(4)移动广告。移动广告的定义为通过移动媒体传播的付费信息,旨在通过这些商业信息影响受众的态度、意图和行为。移动广告实际上就是一种支持互动的网络广告,它由移动通信网承载,具有网络媒体的一切特征,同时由于移动性使用户能够随时随地接收信息,比互联网广告更具优势。

(5)应用商店。在线应用程序商店作为新型软件交易平台首先由苹果公司于 2008 年 7 月推出,依托苹果的庞大市场取得了极大成功。应用商店目前是移动互联网应用中最重要、最新的发展趋势之一。

1.7.5 人工智能

1. 人工智能的定义

人工智能,其字面含义就是人造智能,即指用计算机模拟或实现的智能,因此人工智能又称机器智能。当然,这只是对人工智能的字面解释或一般解释,关于人工智能的科学定义,同智能一样,至今学术界还没有一个统一的认识。下面是几位著名的人工智能方面的科学家分别在不同的年代对人工智能做出的定义。

人工智能是研究使计算机更灵活有用,了解使智能的实现成为可能的原理。因此,人工智能研究结果不仅是使计算机模拟智能,而且是了解如何帮助人们变得更有智能。

——P. H. Winston(1978 年)

人工智能是计算机科学的一个分支,它关心的是设计智能计算机系统,该系统具有通常与人的行为相联系的智能特征,如了解语言、学习、推理、问题求解等。

——A. Barr 和 E. A. Feigenbaum(1981 年)

人工智能是研究怎样让计算机模拟人脑从事推理、规划、设计、思考、学习等思维活动,解决至今认为需要由专家才能处理的复杂问题。

——Elaine Rich(1983 年)

人工智能是研究智能行为的科学。它的最终目的是建立关于自然智能实体行为的理论和指导创造具有智能行为的人工制品,这样一来,人工智能有两个分支,一个为科学人工智能,另一个为工程人工智能。

——Michael R. Genesereth 和 Nils J. Nilsson(1987 年)

根据 Michael 和 Nilsson 关于人工智能的定义,人工智能既是一门工程技术学科,也是一门理论研究学科。作为工程技术学科,人工智能的目的是提出建造人工智能系统的新技术、新方法和新理论,并在此基础上研制出具有智能行为的计算机系统。例如,在医学和地质学领域能够给出专家水平建议和劝告的专家系统,能够进行棋类游戏与比赛的博弈系统,能够解答数学问题的问题求解系统,等等。作为理论研究学科,人工智能的目的是提出能够描述和解释智能行为的概念与理论,为建立人工智能系统提供理论依据。

以上是人工智能的一些比较权威的定义。人工智能还有一个比较模糊的定义,那就是“如果某个问题在计算机上没有解决,那么这个问题就是人工智能问题”,因为一旦解决了某个问题,也就有了解决这个问题的模型或算法,因而也就被划分到某个学科或某个学科的分支中。因此,从这种意义上讲,人工智能永远是一个深奥而永无止境的追求目标。

2. 人工智能的发展

人工智能的发展大致经过了下面3个阶段。

(1)第一阶段(20世纪50—80年代)。这一阶段人工智能刚诞生,基于抽象数学推理的可编程数字计算机已经出现,符号主义快速发展,但由于很多事物不能形式化表达,建立的模型存在一定的局限性。此外,随着计算任务的复杂性不断加大,计算能力远远不能满足需求,人工智能发展一度遇到瓶颈。

(2)第二阶段(20世纪80—90年代末)。在这一阶段,专家系统得到快速发展,数学模型有重大突破,但由于专家系统在知识获取、推理能力等方面的不足,以及开发成本高等原因,人工智能的发展又一次进入低谷期。

(3)第三阶段(21世纪初至今)。大数据的积累、理论算法的革新、计算能力的提升,为人工智能的发展提供了丰富的数据资源,协助训练出更加智能化的算法模型。人工智能的发展模式也从过去追求“用计算机模拟人工智能”,逐步转向以机器与人结合而成的增强型混合智能系统,用机器、人、网络结合成新的群智系统,以及用机器、人、网络和物结合成的更加复杂的智能系统。

3. 人工智能的关键技术

人工智能技术关系到人工智能产品是否可以顺利应用到人们的生活场景中。在人工智能领域,它普遍包含了机器学习(machine learning)、知识图谱、自然语言处理、人机交互、计算机视觉、生物特征识别、AR/VR七个关键技术。

(1)机器学习。机器学习是一门涉及统计学、系统辨识、逼近理论、神经网络、优化理论、计算机科学、脑科学等诸多领域的交叉学科,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,是人工智能技术的核心。基于数据的机器学习是现代智能技术中的重要方法之一,研究从观测数据(样本)出发寻找规律,利用这些规律对未来数据或无法观测的数据进行预测。根据学习模式将机器学习分类为监督学习、非监督学习和强化学习等。

(2)知识图谱。知识图谱本质上是结构化的语义知识库,是一种由节点和边组成的图数据结构,以符号形式描述物理世界中的概念及其相互关系,其基本组成单位是“实体-关系-实体”三元组,以及实体及其相关“属性-值”对。不同实体之间通过关系相互连接,构成网状的知识结构。在知识图谱中,每个节点表示现实世界的“实体”,每条边为实体与实体之间的“关系”。通俗地讲,知识图谱就是把所有不同种类的信息连接在一起而得到的一个关系网络,提供了从“关系”的角度去分析问题的能力。

知识图谱可用于反欺诈、不一致性验证、组团欺诈等公共安全保障领域,需要用到异常分析、静态分析、动态分析等数据挖掘方法。特别地,知识图谱在搜索引擎、可视化展示和精准营销方面有很大的优势,已成为业界的热门工具。

(3)自然语言处理。自然语言处理是计算机科学领域与人工智能领域中的一个重要方

向,研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法,涉及的领域较多,主要包括机器翻译、机器阅读理解和问答系统等。

(4)人机交互。人机交互主要研究人和计算机之间的信息交换,主要包括人到计算机和计算机到人的两部分信息交换,是人工智能领域重要的外围技术。人机交互是与认知心理学、人机工程学、多媒体技术、虚拟现实技术等密切相关的综合学科。传统的人与计算机之间的信息交换主要依靠交互设备进行,主要包括键盘、鼠标、操纵杆、数据服装、眼动跟踪器、位置跟踪器、数据手套、压力笔等输入设备,以及打印机、绘图仪、显示器、头盔式显示器、音箱等输出设备。人机交互技术除了传统的基本交互和图形交互外,还包括语音交互、情感交互、体感交互及脑机交互等技术。

(5)计算机视觉。计算机视觉是使用计算机模仿人类视觉系统的科学,让计算机拥有类似人类提取、处理、理解和分析图像及图像序列的能力。自动驾驶、机器人、智能医疗等领域均需要通过计算机视觉技术从视觉信号中提取并处理信息。近来随着深度学习的发展,预处理、特征提取与算法处理渐渐融合,形成端到端的人工智能算法技术。根据解决的问题,计算机视觉可分为计算成像学、图像理解、三维视觉、动态视觉和视频编解码五大类。

(6)生物特征识别。生物特征识别技术是指通过个体生理特征或行为特征对个体身份进行识别认证的技术。从应用流程看,生物特征识别通常分为注册和识别两个阶段。注册阶段通过传感器对人体的生物表征信息进行采集,如利用图像传感器对指纹和人脸等光学信息、麦克风对说话声等声学信息进行采集,利用数据预处理及特征提取技术对采集的数据进行处理,得到相应的特征进行存储。

生物特征识别技术涉及的内容十分广泛,包括指纹、掌纹、人脸、虹膜、指静脉、声纹、步态等多种生物特征,其识别过程涉及图像处理、计算机视觉、语音识别、机器学习等多项技术。目前生物特征识别作为重要的智能化身份认证技术,在金融、公共安全、教育、交通等领域得到广泛的应用。

4. 人工智能的应用

人工智能的飞速发展,为制造、家居、教育、交通、安防、医疗、物流等各行各业的发展和社会服务带来了前所未有的变化,深刻改变着人们的社会生活,让人们的学习更个性、工作更便捷、生活更美好。

(1)智能制造。智能制造是基于新一代信息通信技术与先进制造技术深度融合,贯穿于设计、生产、管理、服务等制造活动的各个环节,具有自感知、自学习、自决策、自执行、自适应等功能的新型生产方式。智能制造对人工智能的需求主要体现在以下3个方面。

①智能装备,包括自动识别设备、人机交互系统、工业机器人及数控机床等具体设备,涉及跨媒体分析推理、自然语言处理、虚拟现实智能建模及自主无人系统等关键技术。

②智能工厂,包括智能设计、智能生产、智能管理及集成优化等具体内容,涉及跨媒体分析推理、大数据智能、机器学习等关键技术。

③智能服务,包括大规模个性化定制、远程维护及预测性维护等具体服务模式,涉及跨媒体分析推理、自然语言处理、大数据智能、高级机器学习等关键技术。

(2)智能家居。智能家居以住宅为平台,基于物联网技术,是由硬件(智能家电、智能硬件、安防控制设备、家具等)、软件系统、云计算平台构成的家居生态圈,实现人远程控制设备、设备间互连互通、设备自我学习等功能,并通过收集、分析用户行为数据为用户提供个性

化生活服务,使家具生活安全、节能、便捷等。

(3)智能教育。随着人工智能技术的发展,越来越多的人工智能工具被应用于教育领域,称为教师教学和学生学习的得力助手。智能导师是人工智能在教育领域的一个重要应用,它主要通过自然语言处理和语音识别技术,由计算机模拟教师教学的经验和方法,对学生实施一对一的教学,并向具有不同需求和特征的学习者传递知识。它能够在学习者学习的过程中实时跟踪、记录和分析学习者的学习过程与结果,以了解其个性化的学习特点,并根据这一特点为每一名学习者选择合适的学习资源,制定个性化的学习方案。

(4)智能交通。智能交通系统是通信、信息和控制技术在交通系统中集成应用的产物。智能交通系统借助现代科技手段和设备,将各核心交通元素联通,实现信息互通与共享及各交通元素的彼此协调、优化配置和高效使用,形成人、车和交通的一个信息采集系统,采集道路中的车辆流量、行车速度等信息,信息分析处理系统处理后形成实时路况,决策系统据此调整道路红绿灯时长,调整可变车道或潮汐车道的通行方向等,信息发布系统将路况推送到导航软件和广播中,让人们合理规划行驶路线。通过不停车电子收费系统(ETC),实现对通过ETC入口站的车辆身份及信息自动采集、处理、收费和放行,有效提高通行能力、简化收费管理、降低环境污染。

(5)智能安防。智能安防技术是一种利用人工智能对视频、图像进行存储和分析,从中识别安全隐患并对其进行处理的技术。智能安防与传统安防的最大区别在于智能化,传统安防对人的依赖性比较强,非常耗费人力,而智能安防能够通过机器实现智能判断,从而尽可能实现实时的安全防范和处理。

(6)智能医疗。人工智能的快速发展为医疗健康领域向更高的智能化方向发展提供了非常有力的技术条件。近年来,智能医疗在辅助诊疗、疾病预测、医疗影像辅助诊断、药物开发等方面发挥了重要的作用。

(7)智能物流。智能物流就是利用条形码、射频识别技术、传感器、全球定位系统等方面优化改善运输、仓储、包装、装卸等物流业基本活动,同时也在尝试使用智能搜索、推理规划、计算机视觉及智能机器人等技术,实现获取运输过程的自动化运作和高效率优化管理,提高物流效率。