

模块 1 计算机基础知识

现代计算机的诞生是 20 世纪人类最伟大的发明之一。人类社会正在进行第三次产业革命，即信息革命。信息革命的标志就是计算机技术和通信技术的发展、融合与普及。随着人类进入 21 世纪，计算机已成为了各行业普遍使用的基本工具之一。掌握以现代计算机为核心的信息技术的基础知识，提高计算机应用能力，特别是利用计算机为自身专业服务的能力，是当代大学生必备的基本素质之一。本模块讲解计算机的一些基本知识。

学习目标

认知目标	情感目标	技能目标
认识信息、数据、信息技术与计算机等基本概念。 了解信息技术与计算机的发展过程。 认识计算机的分类与主要应用。 认识信息技术组成，且针对一些具体的信息系统或控制系统能分析所需要采用的信息技术。 掌握计算机中数据的表示方法与二进制之间的运算规则。	了解信息技术对人类的影响，正确认识计算机的价值，树立科学管理、正确使用信息资源的意识。 培养学生的计算思维能力，能够主动地利用计算机为自身的学习与工作服务。	通过本模块的学习，能够快速实现各种进制数据之间的转化。 能快速计算音频文件与位图文件的容量。 掌握各种文字信息的输入方法。

单元知识、活动与任务

单元知识	活动设计	任务设计
信息技术的基础知识 计算机的发展与应用 数据的表示方法 数的表示方法	计算机应用的讨论 矢量图与位图的比较 十进制与二进制之间快速转化	书面作业 文字输入

单元 1 信息技术基本知识

随着信息化在全球的快速进展，世界各国对信息的需求快速地增长，信息产品和信息服务已经成为各个国家、地区、企事业单位、家庭与个人不可缺少的部分。因此，信息技术已成为支撑当今经济活动和社会生活的基石。信息技术代表着当今先进生产力的发展方向，信息技术的广泛应用使信息的重要生产要素和战略资源的作用得以充分发挥，使人们能更高效地进行

资源优化配置，从而推动传统产业不断升级，提高社会劳动生产率和社会运行效率。本单元学习信息技术的基础知识。

知识 1 信息的定义及特征

目前，关于信息作为科学概念和范畴的定义，在学术界尚未取得一致。据有关文献统计，目前，关于信息的定义有数百种，以下是其中几种较典型的定义：

（1）信息是指可以用语言、文字、数据、图表、图形或其他可以让使用者识别的符号来表示的，并可以进行存储、加工、传递、处理及应用的对象。

（2）信息是事物运动的状态与方式，是物质的一种属性。在这里，“事物”泛指一切可能的研究对象，既包括外部世界的物质客体，也包括主观世界的精神现象；“运动”泛指一切意义上的变化，包括机械运动、化学运动、思维运动和社会运动；“运动方式”是指事物运动时在时间上所呈现的过程和规律；“运动状态”则是事物运动在空间上所展示的形状与态势。

（3）信息泛指人类传播的一切内容。人通过获取、识别自然界和社会的不同信息来区别不同事物，从而认识世界和改造世界。

.....

从信息的定义可以看出，信息是一个相当深刻且非常丰富的概念。信息对人类社会具有非常重要的意义，如何有效地获取、识别、传递与处理信息已成为人们关注与研究的重要课题。可识别、可存储、可扩充、可压缩、可传递、可转换、在特定的范围有效构成了信息的主要特征。

首先，信息是可以识别的。信息的识别分为直接识别和间接识别。直接识别是指通过人体感觉器官的识别，间接识别是指通过各种测试手段的识别。不同的信息有不同的识别方法。例如，正常人能利用视觉识别物体的颜色，驾驶员能利用视觉与听觉识别交通警察的交通指令，这些识别方式是直接识别。又如，人们能使用温度计测量物体的温度，利用雷达波来侦测移动物体的速度等，这些识别方式就是间接识别。

其次，信息可以使用各种方法存储，且随着时间的变化可以不断扩充。人们可以将信息记录在纸上或存储在大脑中，也可以存储在计算机的存储器中。为减少存储空间，人们通常使用一些方法对信息数据进行加工、整理、概括，使之精练、浓缩，即信息具有可压缩性。

第三，信息可以通过传输媒介传递，也可以从一种形态转换成另一种形态。如信息可通过报纸、电话、电视及计算机网络传递，又如人们可将报纸上看到的文字新闻转换为语言传递给他人，也可把视觉可识别的电影转化为硬盘能存储的磁信号。可传递性是信息的本质特征。

最后，信息在特定的范围内才有效。所谓特定范围有时间范围和空间范围之分。例如，当天的天气预报信息只有当天有效；一个单位在特定时间内的招聘信息只有在招聘时间范围内有效。又如，人的体重在地球上为 70kg，到了太空，此信息为无效信息。

知识 2 信息的概念

信息技术通常简称为 IT 技术（Information Technology, IT）。什么是信息技术？不同的时期有不同的理解，且随着时代的发展与技术进步在不断地变化着。目前，对信息技术的定义大致有如下三种：

（1）信息技术是指有关信息的收集、识别、提取、变换、存储、处理、检索、检测、分析和利用等的技术。

(2) 信息技术是指利用计算机和现代通信手段获取、传递、存储、处理、显示和分配信息的技术。

(3) 国内有些专家学者认为, 信息技术是指研究信息如何产生、获取、传输、变换、识别和应用的技术。

信息技术体现了一个时代的技术特征。时代不同, 其采用的信息技术也不同, 信息技术可能是机械的, 也可能是激光的, 可能是电子的, 也可能是生物的。

而在当今的数字化时代, 信息技术是以微电子和光电技术为基础, 以计算机和通信技术为支撑, 以信息处理技术为主体的技术的总称, 是一门综合性的技术。计算机技术和通信技术的紧密结合, 标志着数字化信息时代的到来。

知识3 信息技术的发展

人类通信的历史已很悠久。远古时期, 人们就通过简单的语言、壁画等方式交换信息。千百年来, 人们一直在用语言、图符、钟鼓、烟火、竹简、纸书等传递信息, 烽火狼烟、飞鸽传书、驿马邮递就是古代通信的例子。在现代社会中, 交警的指挥手语、航海中的旗语等都是古老通信方式进一步发展的结果。这些通信方式依靠人的视觉与听觉。

19 世纪中叶以后, 随着电报、电话的发明, 电磁波的发现, 人类通信领域产生了根本性的变革, 实现了利用金属导线与光纤来传递信息, 通过电磁波来进行无线通信, 使人类的信息传递脱离常规的视、听、觉方式, 用电信号与光信号作为新的载体, 由此, 带来了一系列的技术革新, 开启了人类通信的新时代。

1837 年, 美国人塞缪·摩尔斯 (Samuel Morse) 成功地研制出世界上第一台电磁式电报机, 为人们使用电报通信提供了基础。

1864 年, 英国物理学家麦克斯韦 (J.C.Maxwel) 预言了电磁波的存在, 建立了电磁理论。1875 年, 苏格兰青年亚历山大·贝尔 (A.G.Bell) 发明了世界上第一台电话机, 这就引领了电话通信的出现。

1888 年, 德国青年物理学家海因里斯·赫兹 (H.R.Hertz) 开展了一系列实验, 发现了电磁波的存在, 他用实验证明了麦克斯韦的电磁理论, 导致了无线通信技术的出现与发展。

20 世纪初, 二极管与三极管的发明, 推动了广播通信在世界各国的蓬勃发展。

1945 年, 美国无线电公司在三基色工作原理的基础上研制了世界上第一台全电子管彩色电视机。随着无线传播与接收技术的研究成功, 电视通信迅速普及开来。

随着电子技术的高速发展, 军事、科学研究迫切需要解决的计算工具也大大改进。1946 年, 美国宾夕法尼亚大学的埃克特 (Eckert) 和莫希里 (Moche) 研制出世界上第一台电子计算机, 这样带来了计算机的发展及应用。1948 年, 美国贝尔实验室的肖克莱、巴丁和布拉坦发明了晶体三极管, 于是晶体管收音机、晶体管电视、晶体管计算机很快代替了各式各样的真空电子管产品。1959 年, 美国的基尔比和诺伊斯发明了集成电路, 从此微电子技术诞生了。1967 年, 大规模集成电路诞生了, 一块米粒般大小的硅晶片上可以集成 1 千多个晶体管线路。1977 年, 美国、日本科学家联合制成超大规模集成电路, 30 平方毫米的硅晶片上集成了 13 万个晶体管。微电子技术极大地推动了电子计算机的更新换代, 使电子计算机具有了前所未有的信息处理功能, 成为现代高新科技的重要标志。

为了解决资源共享问题, 单一计算机很快发展成为计算机网络, 实现计算机与计算机之间的数据通信、数据共享。通信介质从普通导线、同轴电缆发展到双绞线、光纤导线、光缆;

电子计算机的输入、输出设备也飞速发展起来，扫描仪、绘图仪、音频视频设备等，使计算机如虎添翼，可以处理更加复杂的问题。20 世纪 80 年代末多媒体技术的兴起，使计算机具备了综合处理文字、声音、图像、影视等各种形式信息的能力，日益成为信息处理最重要和必不可少的工具。

20 世纪 90 年代初期，随着 WWW 的发展，Internet 逐渐趋向民用。由于 WWW 良好的界面大大简化了 Internet 操作的难度，使得用户的数量急剧增加，许多政府机构、商业公司意识到 Internet 巨大的潜力，于是纷纷加入，这样 Internet 规模越来越庞大，如今 Internet 已经深入到人们生活的各个部分。互联网上的信息多种多样，通过 WWW 浏览、电子邮件等方式，人们可以及时地获得所需的信息、方便的传播信息。可以说 Internet 是继电报、电话发明以来人类通信方式的新的一次革命。

随着人类步入 21 世纪，信息技术发展到第三代移动通信技术（3rd Generation, 3G）。3G 是指将无线通信与互联网等多媒体通信结合的新一代移动通信系统，有人简称之为“无线互联网”。3G 提供的许多精彩业务，如可视电话、视频留言、手机购物、手机搜索、QQ 与微信等已成为人们生活中必不可少的部分。目前，信息技术发展到了第四代移动通信，虽然我们无法说出第四代移动通信的标志是什么，但第四代通信的速度将大幅度地提升，至少提升 10 倍以上，让手机视频在线直播、视频聊天等海量数据传输非常流畅。

知识 4 信息技术的组成

信息技术是用于管理和处理信息所采用的各种技术的总称。具体来讲，信息技术主要包括以下几方面：

（1）感测与识别技术。感测技术的作用是扩展人获取信息的感觉器官功能，包括信息识别、信息提取、信息检测等技术。信息识别主要包括文字识别、语音识别和图形识别等。

（2）信息传递技术。该技术实现信息快速、可靠、安全的转移，通信技术就属于这个范畴。

（3）信息处理与再生技术。信息处理的过程主要包括信息的获取、储存、加工、发布和表示。在对信息进行处理的基础上，还可形成一些新的更深层次的决策信息，这称为信息的“再生”。信息的处理与再生都有赖于现代电子计算机的超凡功能。

目前，遥控、遥测和遥感技术也是非常重要的技术。遥控是利用通信线路对远处被控对象进行控制的一种技术，用于电气、输油管道、化学工业、军事和航天；遥测是将远处需要测量的物理量如电压、电流、气压、温度、流量等转换成电量，利用通信线路传送到观察点的一种测量技术，用于气象、军事和航空航天业；遥感是一门综合性的测量技术，在高空或远处利用传感器接收物体辐射的电磁波信息，经过加工处理，用能够识别的图像或电子计算机用的记录磁带提示被测物体的性质、形状和变化动态，主要用于气象、军事和航空航天等领域。

计算机技术同通信技术一样属于信息技术的核心内容。随着电子技术的发展，计算机技术获得了飞速的发展，计算机体积由大变小，功能由弱变强，处理速度也越来越快，已成为信息处理的重要工具。

单元 2 计算机的发展与应用

现代计算机是一种能够存储程序，并能按照程序自动、高速、精确地进行大量计算和信

息处理的智能电子设备。它是科学技术发展的象征,也是促进科学技术和生产力高速发展的有力工具。目前,计算机的发展程度与应用水平已成为衡量一个国家或一个地区的科学技术发展水平和经济实力的重要标志。

知识 1 计算机的发展

计算机的发展动力就是人类应用的需求,通常人们把计算机发展过程划分为以下几个阶段。

1. 机械计算机(1930 年以前)

在这一阶段,人们发明了一些用来计算的机器。较典型的有:17 世纪由法国著名数学家和哲学家 Blaise Pascal 发明的用来进行加减运算的机器 Pascaline;1890 年,在美国国家统计局工作的 Herman Hollerith 设计并制造出具有编程能力的机器,该机器可以自动阅读、计数和排列存储在穿孔卡上的数据。虽然这些机械计算机在当时算很先进,但还看不到现代计算机的影子。

2. 电子计算机(1930 年至 1950 年)

在这个时期,电子计算机出现了。较典型的是 1939 年,John V. Atanasoff 和他的助手 Clifford Berry 一起发明了用于解决一些线性方程系统的 ABC (Atanasoff Berry Computer)。第一台通用的完全电子化的计算机是在 1946 年 2 月由 John Mauchly 和 J. Presper Eckert 设计完成。这台计算机当时被命名为电子数字集成器和计算器 (Electronic Numerical Integrator And Calculator, ENIAC),它共用了 18000 多个真空管和 1500 多个继电器,占地面积 170 平方米,重达 30 吨,每秒钟能完成 5000 次加法计算。ENIAC 计算机的主要缺点是存储容量小以及用线路连接的方法编排程序等。这台机器服役了 9 年左右,于 1955 年 10 月光荣退役。

为了解决人工改接线路这一缺点,1946 年 6 月,美籍匈牙利数学家冯·诺依曼 (John Von Neumann) 提出了“存储程序”的计算机方案,即程序和数据都应该存储在计算机的存储器中。第一台基于冯氏思想的计算机于 1950 年在宾夕法尼亚大学诞生,当时被命名为电子离散变量自动计算机 (Electronic Discrete Variable Automatic Computer, EDVAC)。EDVAC 在两个方面进行了关键性的改进:一是把计算机要执行的指令和要处理的数据都用二进制表示;二是把要执行的指令和要处理的数据按照顺序编成程序存储到计算机内部且让它自动执行(存储程序控制原理)。冯氏思想为现代计算机的发展奠定了坚实的基础。

3. 现代计算机(1950 年至现在)

这一阶段的计算机几乎都基于冯·诺依曼“程序存储”理论模型,速度更快、体积更小、价格更便宜。这一时期的计算机按使用的电子元件划分为电子管、晶体管、中小规模集成电路与大规模集成电路四代。每一代计算机的改进主要体现在硬件或软件方面的变化。

(1) 电子管计算机时代。

1950 年到 1959 年,这时期的计算机称为电子管计算机。这一代计算机的主要特征是用电子管作为运算与逻辑元件,用机器语言和汇编语言编写程序。这样的计算机体积庞大,造价昂贵,运算速度低,存储容量小,可靠性与稳定性差,主要用于科学与工程计算。

(2) 晶体管计算机时代。

从 1959 年到 1965 年间,计算机中采用了比电子管先进的晶体管。晶体管与电子管相比,具有体积小、能量消耗低、可靠性与稳定性高的特点。晶体管时代计算机的程序语言从机器语言发展到汇编语言,高级语言 FORTRAN 和 COBOL 相继开发并被广泛使用,同时开始使用磁

盘和磁带作为辅助存储器。由于第二代计算机的体积和价格下降，人们开始接受并使用计算机，因此，计算机工业开始迅速发展。

（3）中小规模集成电路计算机时代。

从 1965 年到 1975 年间，集成电路技术开始发展，并被广泛应用到计算机中来。因此，在这个时期生产的计算机被称为中小规模集成电路计算机。集成电路（Integrated Circuit, IC）是制作在一块晶片上的完整的电子电路。这个晶片看似非常小，却集成了上千个晶体管元件。第三代计算机与第二代计算机相比，它的主要特点是体积更小，价格更低，可靠性与稳定性更高，计算速度更快；用集成电路代替了分立的晶体管元件，用半导体存储器取代了磁芯存储器，操作系统软件也日益成熟。这一时期计算机的设计已逐步走向标准化、规模化和系列化，在科学计算、数据处理与过程控制等领域得到了广泛的应用。

（4）大规模与超大规模集成电路计算机时代。

从 1975 年至今，计算机使用的元件依然是集成电路，但集成度大幅提高，达到几十万甚至上百万个电子元件，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）。第四代计算机最重要的成就表现在微处理器（Microprocessor）技术上，微处理器是一种小型化的电子产品，把计算机的运算和控制等核心部件集成在一块芯片上。1975 年，美国 IBM 公司推出了个人计算机（Personal Computer, PC），从此，计算机开始深入到人们生活的各个领域。

20 世纪 80 年代以后，人们着手研制新一代计算机。新一代的计算机的特点是以人工智能原理为基础，把信息采集、存储、加工、传输和人工智能结合在一起，使计算机具有形式推理、联想、学习和解释能力。这样的计算机突破了原有的冯·诺依曼体系结构，实现了高度的并行处理，主要着眼于机器的智能化，使计算机具有智能接口，可以模拟或部分代替人的智能活动，并具有自然的人机通信能力。具有这种能力的计算机被称为智能计算机，也称作“智能机器人”。2013 年 9 月 4 日，日本就利用小型智能机器人“MIRATA”在东京 2020 年奥运会申办委员会召开的记者会上展示民族的创新力成为宣传的亮点。MIRATA 能用日语与人沟通，模仿人体动作，身高约 34 厘米、体重约 1 公斤，由 Dentsu 公司与东京大学联合研发。

4. 未来计算机

从计算机未来发展的角度来看，科学界看好的计算机除了现代的计算机外，还看好 DNA 生物计算机、光子计算机和量子计算机三类计算机。

（1）DNA 生物计算机。DNA 生物计算机是美国南加州大学伦纳德·阿德拉曼博士在 1994 年提出的设想，该设想是通过控制 DNA 分子间的生化反应来完成数据的运算。这种计算机使用酶做为计算机的“硬件”，DNA 作为计算机的“软件”，输入和输出的“数据”都是 DNA 链。把溶有这些“数据”的溶液恰当地混合，就可以在试管中自动发生反应，这种反应被称之为“运算”。这样的计算机由一堆装着有机液体的试管组成。

（2）光子计算机。光子计算机与传统硅芯片组成的计算机的差异在于用光子来代替电子实现数据的运算和存储，它用不同波长的光来代表不同的数据。1990 年，美国贝尔实验室宣布研制出了世界上第一台光学计算机，该计算机采用砷化镓光学开关，运算速度达每秒 10 亿次。尽管这台光学计算机与理论上的光学计算机还有一定距离，但已显示出它强大的威力。

（3）量子计算机。把量子力学和计算机结合起来的思想是在 1982 年由美国著名物理学家理查德·菲利普·费因曼（Richard Phillips Feynman）提出的。1985 年，英国牛津大学物理学家戴维·多伊奇（David Deutsch）初步阐述了量子计算机的概念，并提出量子并行处理技术

能使量子计算机比传统的图灵计算机功能更强大。

虽然分子、光子和量子计算机的研究还处在实验室阶段，但科学家认为这些计算机具有很高的应用价值。在最近几年里，美国、欧洲和日本政府一直投入巨资资助相关的研究，预计在未来一二十年内，这些新型计算机将取得突破性进展。

知识 2 计算机的分类

计算机的分类方法很多，按其内部逻辑结构可分为单处理机与多处理机（并行机），按 CPU 的数据处理能力可分为 16 位机、32 位机和 64 位机等，按其用途可分为通用计算机与专用计算机。通用计算机是指适用于一般科学计算、工程设计和数据处理等方面的计算机，人们平常所说的计算机就属于通用计算机。专用计算机是指为某种特殊应用而设计，其运行程序固定，执行效率较高，处理速度快，运算精确度高，如用于飞机的自动控制与导航的计算机，坦克上的火控系统所使用的计算机。

从计算机的演变过程和发展趋势来看，通用计算机按其规模、速度和功能等分为超级计算机、大/中型计算机、工作站与微型计算机等。它们之间主要的区别在于计算机体积、功耗、运算速度、数据的存储容量、指令系统的规模等。但随着技术的发展，这种分类已渐渐淡化。下面介绍目前一些主流计算机。

1. 超级计算机

超级计算机又称高性能计算机，通常是指体积大、运算速度快、存储容量高的计算机，主要应用于国防或航天等尖端技术和现代科学计算中。超级计算机的研究是世界公认的高新技术制高点，也是 21 世纪最重要的科学研究领域之一。超级计算机的研制能力已是衡量一个国家经济实力和科学技术水平的重要标志之一。2008 年，美国的走鹃（Roadrunner）和美洲豹（Jaguar）超级计算机的峰值突破千万亿次（Petaflops）大关，标志着计算机已进入千万亿次计算时代。2010 年，我国由曙光公司天津产业基地研制生产的超级计算机“星云”落户深圳，其系统峰值为每秒 3000 万亿次，是当时中国第一台、世界第三台实测双精度浮点计算超过千万亿次的超级计算机，在 2010 年 5 月 31 日发布的第 35 届全球超级计算机 500 强排名中，名列第二。目前，由国防科技大学研制的我国高性能计算机天河一号超级计算机二期工程系统峰值性能达到每秒 4700 万亿次，其运算速度与能效达到国际领先水平。

2. 大/中型计算机

大/中型计算机也具有较高的运算速度，而且有较大的存储容量，主要用于科学计算、数据处理或作为网络服务器。服务器（Server）是指在网络环境下运行相应的服务软件，为网上用户提供信息资源共享和各种应用服务的一种高性能计算机。它的高性能主要体现在高速度的运算能力、长时间的可靠运行、强大的外部数据吞吐能力等方面，是网络的中枢和信息化的核心。

3. 工作站

工作站是一种以个人计算和分布式网络计算为基础，主要面向专业应用领域，具备强大的数据运算与图形、图像处理能力，主要应用于工程设计、动画制作、科学研究、软件开发、金融管理、信息服务、模拟仿真等专业领域。

4. 微型计算机

微型计算机的中央处理单元（CPU）采用微处理器芯片，体积小巧，已广泛用于商业、服务业与工厂的自动控制、办公自动化以及大众化的信息处理之中。微型计算机也被称为 PC

（Personal Computer）机。目前，我国是 PC 生产的主要国家之一。2005 年 5 月 1 日，联想集团宣布完成了对 IBM 全球个人电脑业务的收购，标志着联想已成为全球第三大个人电脑生产企业。

知识 3 计算机的主要应用

计算机发展的动力源于计算机的应用。目前，计算机的应用主要包括以下几个方面。

1. 科学计算

科学计算也称数值计算。早期的计算机主要用于科学计算。目前，科学计算仍然是计算机应用的一个重要领域，如高能物理计算、工程设计、地震预测、气象预报、航空航天技术中的计算问题都要依靠计算机。由于计算机具有高运算速度、高精度及逻辑判断能力，由此也产生了计算力学、计算化学、生物控制论等新型学科。

目前，随着传统计算机技术和网络技术的发展与融合，云计算（Cloud Computing）进入了深入研究与广泛应用阶段。所谓云计算，是将网格计算（Grid Computing）、分布式计算（Distributed Computing）、并行计算（Parallel Computing）、效用计算（Utility Computing）、网络存储技术（Network Storage Technologies）、虚拟化（Virtualization）、负载均衡（Load Balance）等技术融合，旨在通过网络把多个成本相对较低的计算实体整合成一个具有强大计算能力的系统，并借助一些先进的商业模式把这强大的计算能力分布到终端用户手中。简单地说，云计算就是基于互联网的计算。云是指网络资源的分布形态，如处理云、存储云。比尔·盖茨说过：“把你的计算机当做接入口，一切都交给互联网吧。”“用户只需要 640K 的内存就足够了。”就是说未来的 PC 机可以没有硬盘、无需安装 Office、CorelDRAW、Photoshop 等诸多软件，可以简单到只剩下显示器，数据的处理与存储都交给互联网，用户需要完成相应工作时，只需要向供应商付费购买服务即可，这就好比用电不需要每家每户都有发电机，由电网供电，用户向电力公司购买。因此，云是资源分布的形态。

云计算的一个核心理念就是通过不断提高“云”的处理能力，进而减少用户终端的处理负担，最终使用户终端简化成一个单纯的输入输出设备，并能按需享受“云”的强大计算处理能力。

云计算为众多用户提供了一种新的高效计算模式，兼有互联网服务的便利、廉价和大型机的能力。它的目的是将资源集中于互联网上的数据中心，由这种云中心提供应用层、平台层和基础设施层的集中服务，以解决传统 IT 系统的零散性带来的低效问题。云计算是信息化发展进程中的一个阶段，强调信息资源的聚集、优化、动态分配和回收，旨在节约信息化成本、降低能耗、减轻用户信息化的负担，提高数据中心的效率。云计算由第三方提供集中式互联网服务，它是一种新型的应用模式。

2. 数据处理

数据处理是对数据的采集、存储、检索、加工、变换和传输等一系列活动的总称。数据处理的基本目的是从大量的、可能是杂乱无章的、难以理解的数据中抽取并推导出对于某些特定的人们来说有价值、有意义的数据。数据处理是目前计算机应用最广泛的一个领域。据统计，目前 80% 以上的计算机用于数据处理。数据处理技术的发展从简单到复杂经历了以下四个发展阶段：

（1）电子数据处理（Electronic Data Processing, EDP），是以文件系统为手段，实现一个部门内的数据处理。

(2) 管理信息系统 (Management Information System, MIS), 是以数据库技术为工具, 实现一个部门的全面管理, 提高工作效率。

(3) 决策支持系统 (Decision Support System, DSS), 是以数据库、模型库和方法库为基础, 帮助管理决策者们提高决策水平, 改善运营策略的正确性与有效性。

(4) 大数据。随着时代的发展, 世界正变得越来越数字化。目前, 大数据作为一项颠覆性的技术, 它正在以各种各样的方式影响着每个人的生活。所谓大数据是指利用常用软件工具捕获、管理和处理数据所耗时间超过可容忍时间的数据集。维克托·迈尔·舍恩伯格在《大数据时代》一书中指出, 大数据带来的信息风暴正在改变我们的生活、工作和思维, 大数据开启了一次重大的时代转型。他也明确指出, 大数据时代最大的转变就是放弃对因果关系的渴求, 取而代之关注数据的相关关系。维克托最具洞见之处在于我们仅要知道“是什么”, 而不需要知道“为什么”, 它颠覆了千百年来人类的思维惯例, 对人类的认知和与世界交流的方式提出了全新的挑战。全球知名咨询公司麦肯锡认为: “数据已经渗透到当今每一个行业和业务职能领域, 成为重要的生产因素。人们对于海量数据的挖掘和运用, 预示着新一波生产率增长和消费者盈余浪潮的到来。”大数据的核心就是预测。2012年, 美国政府以“Big Data Is a Big Deal”为题发布新闻, 宣布启动“大数据研究与发展计划”, 大力推动和改善与大数据相关的收集、组织和分析工具及技术, 以推进从大量的复杂的数据集合中获取知识的洞见能力。2012年5月, 联合国发布了大数据白皮书, 列举了大数据在交通、经济、疫情预测、社交网络等方面的应用, 同时指出大数据对联合国和各个政府来说是一个历史性的发展机遇。大数据技术是指将来自于互联网、移动互联网、社交网与物联网的数据进行聚集与整合, 然后根据不同和需求对数据进行分析产生报表或图表等来供用户进行相关的预测。

大数据主要特征体现在巨量性 (Volume)、多样性 (Variety)、快速性 (Velocity)、价值低密度性 (Veracity) 4个方面。

巨量性是指大数据拥有巨大的数据量以及完整的规模。据统计, 2010 全球正式进入 ZB (ZettaByte, 泽字节) 数据时代。据 IDC (Internet Data Center, 互联网数据中心) 预测, 到 2020 年, 全球将拥有 35ZB 的数据量。多样性指大数据有多种途径来源的关系型和非关系型数据。在当前的信息网络中, 数据类型早已不是单一的数据表, 日志、Blog、微博、音频、视频等结构化、半结构化与非结构化数据等多种数据类型混杂在网络上。这样, 对数据的处理能力提出了更高的要求。速度性主要表现为数据流和大数据的移动性。现实生活中主要体现为数据的实时性。随着移动网络的发展, 人们对数据的实时应用需求更加普遍, 比如通过手持终端设备关注天气、交通、物流等信息。快速性要求具有时间敏感性和决策性的分析——能在第一时间抓住重要事件发生的信息。因此, 在海量数据的处理过程中, 数据处理效率是大数据技术的生命线。价值低密度性是指大数据其价值具有稀疏性、不确定性和多样性。“互联网女皇”Mary Meeker 在 2012 年互联网发展趋势中, 用一幅生动的图像来描述大数据。一张是整整齐齐的稻草堆, 另外一张是稻草中缝衣针的特写。寓意通过大数据技术的帮助, 可以在稻草堆中找到你需要的东西, 哪怕是一枚小小的缝衣针。这两幅图揭示了大数据技术一个很重要的特点, 价值的稀疏性。在信息网络中, 大数据的价值很难被完全挖掘利用, 因此, 如何通过强大的算法更迅速地完成数据的提纯是目前大数据汹涌背景下亟待解决的问题。

3. 过程检测与控制

目前, 在工业生产过程中, 利用计算机对温度、压力、流量、液位和成分等参数自动检

测，并把检测到的数据存入计算机，再根据需要对检测到的数据进行处理，然后再对工业生产过程进行自动调整控制，这样的系统被称为计算机检测与控制系统。现在，随着嵌入式技术的发展，在仪器仪表中引入计算机技术后研究出的智能化仪器仪表，将工业自动化推向了一个更高的水平。

4. 计算机辅助系统

在飞机、汽车、船舶、机械、建筑工程与集成电路等行业中，为了提高产品（工程）质量，缩短生产周期，降低成本，设计与制造人员借助计算机自动或半自动地完成设计和产品制造的技术称为计算机辅助设计（Computer Aided Design, CAD）和计算机辅助制造（Computer Aided Manufacturing, CAM）。现在，CAD/CAM 技术发展非常迅速，应用范围不断扩大又派生出很多新的分支，如计算机辅助测试（Computer Aided Testing, CAT），计算机辅助教学（Computer Aided Instruction, CAI）等。

5. 智能模拟

人工智能（Artificial Intelligence, AI）学科研究的主要内容包括知识表示、自动推理和搜索方法、机器学习和知识获取、知识处理系统、自然语言理解、计算机视觉、智能机器人、自动程序设计等。智能模拟是用计算机模拟人类的某些智能行为（如感知、思维、推理等）的应用。人工智能的应用领域有问题求解、逻辑推理与定理证明、自然语言处理、智能信息检索技术与专家系统等。

6. 计算机网络

计算机网络是指将地理位置不同的具有独立功能的多台计算机及其外部设备，通过通信线路（有线或无线）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，实现计算机之间的资源共享和通信的计算机系统，是计算机技术与通信技术发展的产物。在日常生活与工作中，人们到处可以感受到网络给我们带来的好处。例如，利用计算机网络收发邮件，使用计算机网络开展网上办公，使用计算机网络实现远程教育，使用计算机网络开展电子商务等都是计算机网络的应用。又如网络电影、网络电视、网络游戏等为人们带来了无限的快乐。

随着人类需求的发展与变化，人们已开始研究物联网与泛在网。所谓物联网，是指在互联网概念的基础上，将其用户端延伸和扩展到任何物品与物品之间，进行信息交换和通信的一种网络。物联网的目的是实现物物相联，从而融合物理信息的感知、传输、处理、控制，提供高智能的应用服务。物联网与传统互联网的主要区别是其包含了物物互联与物机相联，而不是局限于机机相联，图 1-1 中用户端延伸和扩展到出租车、电梯、空调、浴缸、冰箱与电话等，依靠互联网来实现手机与这些物品之间的通信，并使用手机对这些物品进行控制。



图 1-1 物联网实物图

泛在网即广泛存在的网络，它以“无所不在、无所不包、无所不能”为基本特征，以实现在任何时间、任何地点、任何人、任何物都能顺畅地通信为目标。泛在网是指基于个人和社会需求，实现人与人、人与物、物与物之间按需进行的信息获取、传递、存储、认知、决策与使用等服务，具有超强的环境感知、内容感知及智能性，为个人和社会提供无所不在的信息服务和应用。当前的泛在网研究强调自然而无所不在的人机交互和物机交互，被学术界总为 5A（Any Time, Any Place, By Any Person, Using Any Device, Connecting Any Objects）。目前，随着经济发展和社会信息化水平的日益提高，构建泛在网络社会、带动信息产业的整体发展已经成为一些发达国家和城市追求的目标。

当前，计算机的应用领域已渗透到社会的各行各业，正在改变着传统的工作、学习和生活方式，推动着社会的发展与进步。

单元 3 数据的表示方法

正如前面所说，计算机是实现数据处理的机器。那计算机能处理哪些数据？这些数据在计算机中如何表示？本单元就来讨论这些问题。

知识 1 信息的计量

在计算机中，0 与 1 是组成信息的两个基本符号，通常称二进制的一位为比特（bit）。例如，信息 10101100 10010101 为 16 比特。比特作为信息的计量单位显得太小，在计算机中，稍大一点的二进制计量单位是字节（Byte），是信息存储的基本单位，一般用大写“B”表示。八位二进制合称为 1 字节，存储一个西文字符需要一个字节空间，即 8 比特的存储空间。

在存储器中，还用到 KB（千字节）、MB（兆字节）、GB（吉字节）与 TB（太字节）、PB（拍字节）、EB（艾字节）、ZB（泽字节）与 YB（尧字节）等单位，它们之间的换算关系是：

千字节（KiloByte, KB）， $1\text{KB}=2^{10}$ 字节=1024B

兆字节（MegaByte, MB）， $1\text{MB}=2^{20}$ 字节=1024KB

吉字节（GigaByte, GB）， $1\text{GB}=2^{30}$ 字节=1024MB

太字节（TeraByte, TB）， $1\text{TB}=2^{40}$ 字节=1024GB

拍字节（PetaByte, PB）， $1\text{PB}=250$ 字节=1024TB

艾字节（ExaByte, EB）， $1\text{EB}=260$ 字节=1024PB

泽字节（ZettaByte, ZB）， $1\text{ZB}=270$ 字节=1024EB

尧字节（YottaByte, YB）， $1\text{YB}=280$ 字节=1024ZB

知识 2 数据类型

在计算机中，数据以不同形式出现。目前，计算机能处理的数据包括数值（Number）、文本（Text）、图像（Image）、音频（Audio）和视频（Video）。用计算机求圆周率就是处理数值数据，用文字处理软件进行文字录入、文字移动、文字删除、文字打印等工作就是用计算机来处理文本数据，用 Photoshop、CorelDRAW 等软件对图像进行放大、缩小、旋转等编辑就是用计算机来处理图像数据，用计算机播放音乐就是处理音频数据，用计算机编辑视频和播放电影就是处理视频数据。在 IT 行业中，用术语“多媒体（Multimedia）”来定义包含数值、文本、图像、音频和视频等信息的信息。

知识 3 计算机内部的数据

在计算机内部，数字、文本、图像、声音与视频信息等数据该如何表示？1940 年，美国著名的数学家与控制论学者维纳（Norbert Wiener, 1894~1964 年）首先提出使用二进制编码表示数据。1949 年，英国剑桥大学的 M.V.Wikes 教授和他的学生采用存储程序原理研制了电子自动延迟存储计算机（Electronic Delayed Storage Automatic Computer, EDSAC），在该计算机中首次采用二进制编码解决了数据在计算机中的表示问题，且验证了这种计算机工作可靠、稳定与高速。

在计算机内部采用二进制表示数据的主要原因是构成计算机的电子元件有两种稳定状态，如晶体管工作时具有导通与截止两种状态，磁芯磁化有两个方向，电容器有充电和放电两种状态，开关有开启与关闭两种状态，脉冲有电位的高与低。如果用 0 与 1 来描述这两种状态，在技术上实现较为容易，同时二进制的运算较简单，这样的运算电路设计也容易实现。

计算机实现数据处理的过程是：首先把所有计算机外部数据转换成二进制数据，这种转换被称为编码（Coding）。经过编码后的数据存入计算机中，由计算机的中央处理器完成处理，当数据从计算机输出时再还原回来，还原过程被称为解码（Decoding）。计算机处理各种数据的过程如图 1-2 所示。

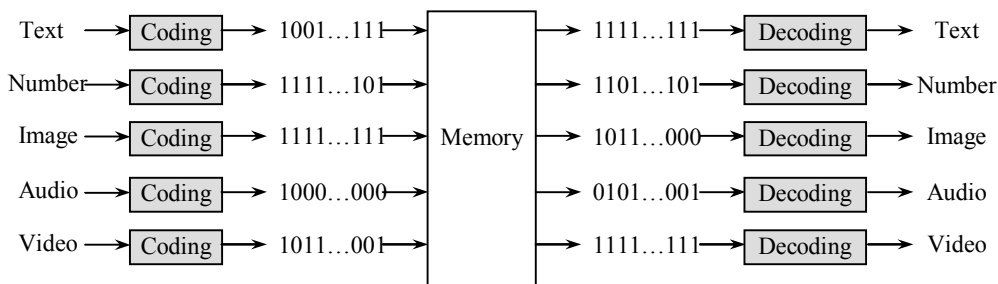


图 1-2 计算机处理各种数据的过程

知识 4 数据的表示

在计算机内部，数值、文本、图像、声音与视频数据都是用二进制来表示的。这些数据如何表示成二进制数据呢？本知识点将回答这一问题。

1. 文本

在任何语言中，文本的片断是用来表示该语言中某个意义的一系列符号。例如，在英文中使用 26 个符号（A、B、C、...、Z）表示大写字母，使用 26 个符号（a、b、c、...、z）表示小写字母，使用 10 个符号（0、1、2、...、9）表示数字符号（非数值数字，后面将看到它们的不同之处），以及使用符号（、！、...、？、”）表示标点，另外还有空格、换行符、制表符等。

问题是在一种语言中，一个符号到底用多少位二进制来表示？这个问题取决于该语言集中有多少个不同的符号。例如，某种语言有 16 个符号，用二进制来表示这些符号必须能够区分这 16 个符号，对另外一些语言，如英文、中文，有更多的符号，需要的二进位就更多一些。

语言的符号数与二进制数位数的关系不是线性关系，而是对数关系。如果某种语言的符号仅有 2 个，就用 1 ($\log_2 2=1$) 位二进制数就能区分，用 0 表示其中一个符号，用 1 表示另一个符号。如果某种语言需要 16 个符号，就要用 4 ($\log_2 16=4$) 位二进制数来表示。从表 1-1 中

可以很容易地看出语言符号数与二进制位数之间的关系。2 位二进制能表示 4 种不同的形式：00、01、10 和 11。4 种形式中的任一种都可代表一个符号。同样的，3 位二进制有 8 种不同的形式：000、001、010、011、100、101、110 和 111，可以表示 8 种不同的符号。

表 1-1 语言符号数与二进制位数的关系

符号数量	二进制位数
2	1
4	2
8	3
16	4
...	...
128	7
256	8
...	...
65536	16

(1) ASCII。ASCII 码是美国标准信息交换码 (American Standard Code for Information Interchange, ASCII)，它由美国国家标准局 (American National Standards Institute, ANSI) 发布的用于表示英文字符集的代码集，这种代码同时也被国际标准化组织 (International Organization for Standardization, ISO) 批准作为国际字符标准。ASCII 码采用 7 位二进制数表示 128 个不同的符号，如表 1-2 所示。编码的二进制数排列顺序为 $b_7b_6b_5b_4b_3b_2b_1$ 。

表 1-2 ASCII 字符与编码对照表

$b_7b_6b_5$ $b_4b_3b_2b_1$	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P		p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	”	2	B	R	b	r
0011	EXT	DC3	#	3	C	S	c	s
0100	EOT	DC4	\$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	‘	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	↑	n	~
1111	SI	US	/	?	O	↓	o	DEL

基本的 ASCII 字符集共有 128 个字符，其中有 96 个可打印字符，包括常用的字母、数字符、标点符号等，另外还有 32 个控制字符。如英文单词“BYTE”在计算机中表示方法如图 1-3 所示。

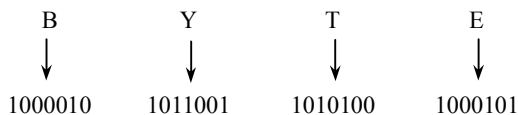


图 1-3 用 ASCII 码表示英文单词“BYTE”

(2) 扩展 ASCII 码。在计算机中，为了使 7 位 ASCII 码统一成为一个字节（Byte，8 位二进制），采用在 ASCII 码左边增加额外的 0 来进行扩充，得到的代码就是扩展 ASCII 码。扩展 ASCII 码是 8 位码，范围从 00000000~01111111。

(3) GB2312-80 编码。1980 年，为了使每一个汉字有一个统一的计算机代码，我国颁布了第一个汉字编码的国家标准 GB2312-80：《信息交换用汉字编码字符集基本集》，这个字符集是我国中文信息处理技术的基础。

GB2312-80 国家标准采用了扩充编码的办法，使用两个字节（16 位二进制数）表示一个汉字的编码，由于 8 位扩展 ASCII 码最高位为 0，为了确保汉字的每个字节的编码与扩展 ASCII 码有所区别，GB2312-80 汉字编码标准中汉字的每个字节的最高位置为 1。遵循这样的编码规则的编码通常称为汉字的国标码或机内码，是汉字信息在计算机内部进行存储、交换、检索等操作的代码。《信息交换用汉字编码字符集基本集》GB2312-80 收入汉字信息交换用的基本图形字符，采用一字一码的原则，具体包括：一般符号、序号、数字、拉丁字母、日文假名、希腊字母、俄文字母、汉语拼音符号、汉语注音字母及简化汉字 6763 个，总计 7445 个图形字符。

(4) Unicode 码。随着计算机的发展与广泛使用，人们要求计算机处理的语言符号越来越多，为了确保世界上不同语言符号编码的统一性与唯一性，一些国际知名硬件和软件商联合共同设计了一种名为 Unicode（Universal Code）的编码。这种编码使用 16 位二进制表示一个符号，最多能表示 65536 (2^{16}) 个符号，代码的不同部分被分配用于表示世界上的不同语言符号。Java 语言使用这种代码来表示字符。Unicode 编码就像它的名字一样，是一种世界通用的符号编码。

(5) ISO 码。ISO 码是由国际标准化组织设计的一种使用 32 位二进制表示一个语言符号的编码，如 ISO10646。这种代码最多能表示 4294967296 (2^{32}) 个符号，足以表示当今世界上的所有的语言符号。

2. 数值

在计算机中，数值是使用二进制数来表示的。如何用二进制数表示数值数据将在下一单元阐述。

3. 图像

目前，图像在计算机中有位图与矢量图两种。位图是通过图像获取设备获得现实景物/对象的映像，矢量图是使用矢量绘图/设计软件以交互方式制作而成。

(1) 位图图形。在位图图形中，图像被分为像素（Pixel）矩阵，每一个像素对应图像上的一个点。像素的大小取决于分辨率。分辨率是一个表示平面图像精细程度的概念，通常它是以横向和纵向点的数量来衡量的，表示成“水平点数×垂直点数”的形式。图 1-4 (a) 的分辨率为 8×8，图 1-4 (b) 分辨率为 16×16。

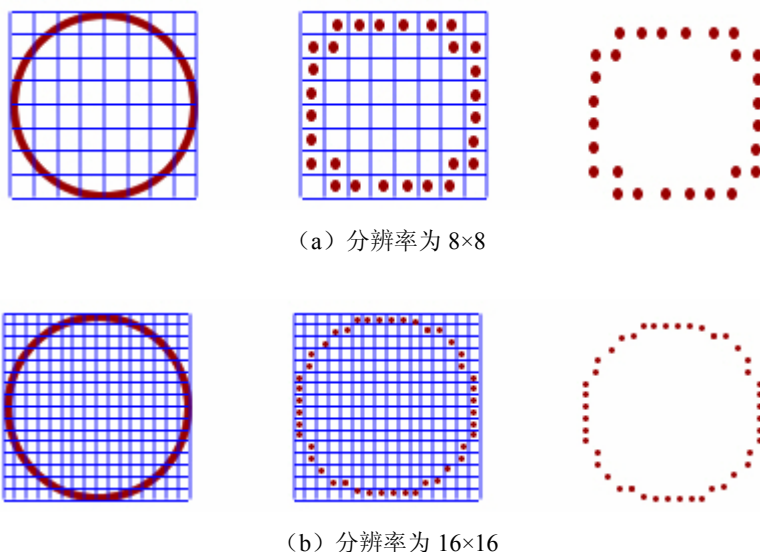


图 1-4 图像分辨率

在一个固定的平面内，分辨率越高，意味着可使用的点数越多，像素越高，图像越细致。位图是像素的集合，把图像分成像素之后，每一个像素点就用一定的二进制位来描述。例如，对于仅由黑白点组成的图像，1 位二进制就足够描述一个像素，用 0 表示黑色像素点，用 1 表示白色像素点。图像中的每个像素点被一个一个记录下来存储在计算机中。图 1-5 显示了这种黑白图像以及它的表示方法。

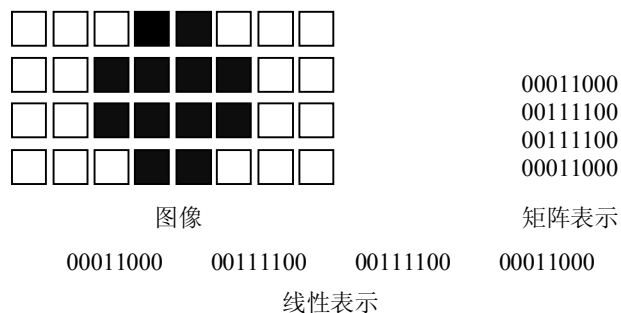


图 1-5 黑白图像的位图图形表示方法

如果一幅图像不单由纯黑、纯白像素组成，而包括黑白过渡的灰度，就可以通过增加每个像素点的二进制位来表示灰度。例如，可以分别使用 2 位二进制 00（表示黑色像素）、01（表示深灰度像素）、10（表示浅灰度像素）、11（表示白色像素）来显示四重灰度级。

如果图像是彩色图像，位图用红、绿、蓝三原色的光学强度来表示像素的颜色。具体的处理方法是：每一种彩色像素被分解成红、绿、蓝三种主色（通常把这种图像称为 RGB 图像），然后测出每个像素点三种颜色的强度，每种颜色的强度被分配固定的二进制位（通常为 8 位）。也就是说，每一个像素用 24 位二进制描述它的颜色成分的强度，前 8 位二进制表示红色（Red）的强度，中间 8 位二进制表示绿色（Green）的强度，后 8 位表示蓝色（Blue）的强度。图 1-6 显示了四种颜色的像素点的二进制的表示方法。

	R	G	B
红（100%的强度）	11111111	00000000	00000000
绿（100%的强度）	00000000	11111111	00000000
蓝（100%的强度）	00000000	00000000	11111111
白（100%的强度）	11111111	11111111	11111111

图 1-6 彩色图像中的红色、绿色、蓝色、白色像素点的表示

完成了上述学习内容后，大家思考这样一个问题：假如有一个分辨率为 320×240 的位图图像，如果是 256 级灰度图或 RGB 彩色图像，存储这样的图像分别需要多大存储空间？

如果是 256 级灰度，保存一个像素点则需要 8 位二进制，也就是一个字节的存储空间；如果是 RGB 彩色图像，保存一个像素点则需要 24 位二进制，即 3 个字节的存储空间。

位图图像文件的大小可以通过以下的方法计算得出：

文件的字节数 = 图像分辨率 $\times$ 图像位深度 / 8

其中，图像位深度是二进制颜色位数。

因此，通过计算，保存分辨率为 320×240 的 256 级位图灰度图像，存储空间为 $320 \times 240 \times 8 / 8 = 76800\text{B}$ ，保存分辨率为 320×240 的 RGB 彩色位图图像，存储空间为 $320 \times 240 \times 24 / 8 = 230400\text{B}$ 。

（2）矢量图。由上可知，位图图形表示法存在的问题之一是一幅图像采用精确的二进制数表示后存储在计算机中，由于需要存储每个像素点的值，存储这样的图像需要较大的存储空间；问题二是如果想重新调整图像的大小，就必须改变像素的大小，这将产生波纹状或颗粒状图像。矢量图很好地解决了这些问题，矢量图使用直线和曲线来描述图形，图形的元素是一些点、线、矩形、多边形、圆和弧线等，它们都是通过存储在计算机中的数学公式计算获得的。由于矢量图形是通过公式计算获得，所以矢量图形文件体积一般较小，同时，由于矢量图形放大、缩小或旋转后的图形也是通过计算公式重新生成，因此，矢量图放大、缩小或旋转后不会失真。矢量图最大的缺点是难以表现色彩层次丰富的图像效果。

从上可知计算机接受的数字图像有位图和矢量图 2 种类型，其中位图能够制作出色彩和色调变化丰富的图像，但文件占用存储空间大，传输速度慢，不方便传递和应用；而矢量图以指令集合记录图像，内容以线条和色块为主，文件较小，容易进行变换，不会失真，但不易制作色调丰富或色彩范围太多的图像。折衷的方法是在人眼可以接受的范围内，对位图文件进行有损压缩，以损失一部分颜色、细节的代价来换取文件存储空间的减少以方便应用。在实际应用中，大家要根据应用的目标选择合适的表示形式。

4. 音频

音频是多媒体技术的重要特征，是携带信息的重要媒体之一。大家知道，声波是随时间而连续变化的物理量，通过能量转换装置，可用随声波变化而改变的电压或电流信号来模拟。假设以模拟电压的幅度来表示声音的强弱。为使计算机能处理音频，必须对声音信号数字化。音频信号的数字化过程包括采样（Sampling）、量化（Quantization）、编码（Coding）等过程。具体转化过程如下：

（1）首先是采样。采样（或称取样、抽样）是把时间连续（如图 1-7（a）所示）的模拟信号转换成时间离散、幅度连续的采样信号，如图 1-7（b）所示。每隔一个时间间隔在模拟

音频波形上取一个幅度值，这样，由一个连续的模拟音频波形，产生了一组离散的数值序列。其中，时间间隔被称为采样周期，单位时间采样的次数就是采样频率。在多媒体中，对于音频信号，最常用的有三种采样频率：44.1kHz；22.05kHz；11.025kHz。

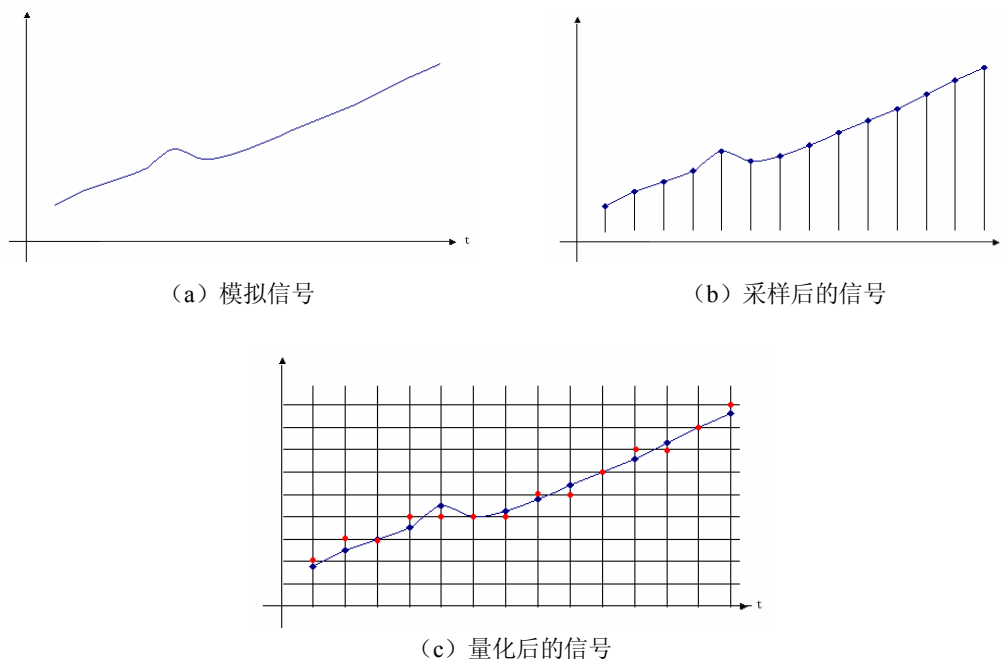


图 1-7 音频的采样与量化

(2) 其次是量化。量化是将时间离散、幅度连续的采样信号转换成时间离散、幅度离散的数字信号，如图 1-7 (c) 所示。将模拟音频信号的电压幅度值划分为若干个级数，每个级数对应一个二进制数字；将各个采样结果提升或下降到级数值，形成一组二进制数字序列。例如量化位数采用 16 位，它对应 65536 个量化级。

(3) 接着是编码，所谓编码是将量化值转化为计算机可存储的二进制数。

(4) 最后是存储，存储就是将编码后的二进制数存储到计算机中。

一般来说，要求声音的质量越高，则量化级数和采样频率越高，保存这一段声音的相应文件也就越大。

声音文件存储空间（字节）= 采样频率 × 量化位 × 声道 × 时间 / 8

例如，采样频率为 44.1kHz，量化位是 16 位的立体声 1 分钟声音所需字节是：

$$44.1 \times 1000 \times 16 \times 2 \times 60 / 8 \approx 10\text{MB}$$

根据采样原理，采样的频率至少高于信号最高频率的 2 倍，采样的频率越高，声音“回放”出来的质量也越高，但是要求的存储容量也越大。

从上述可见，数字化音频的信息量是比较大的，为了使音频信息能更有效地存储和传输，就必须对它进行压缩处理，音频压缩标准比较成熟的有“MPEG 音频”。

音频信号的另一种处理方法是分析与合成的方式。用合成的方式计算机也能制造出音频信号，人们可以使用计算机进行作曲，或能按照人的要求发出人所需要的声音。

音乐合成的方式是按一定的协议标准，采用音乐符号记录方法来记录和解释乐谱，并合成相应的音乐信号，这也就是 MIDI 方式。MIDI (Musical Instrument Digital Interface) 是乐器

数字接口的缩写，是数字音乐的一个国际标准。MIDI 标准规定了电子乐器与计算机连接的电缆硬件以及电子乐器之间、乐器与计算机之间传送数据的通信协议等规范。MIDI 标准使不同厂家生产的电子合成乐器可以互相发送和接收音乐数据。MIDI 文件记录的是一系列指令而不是数字化后的波形数据，所以它占用存储空间较小。

MIDI 音质与设备有关。如果硬件设备不理想，MIDI 的音质、音色不能充分发挥，反之，MIDI 文件的音质甚至会高于 WAV 文件。

在数字音频回放时，由数字到模拟的转化器（数/模转换器）解码可将二进制编码恢复成原始的声音信号，再通过音响设备输出。

5. 视频

视频是图像（帧）在时间上的表示，如图 1-8 所示。电影就是一系列的帧，播放软件通过一张一张地播放而形成运动图像。所以，如果知道如何将图像数据存储在计算机中，也就知道了在计算机中如何存储视频数据；每一幅图像或帧被转化成二进制数并存储，这些图像组合起来可表示视频。由于视频数据较大，因此，视频数据通常被压缩存储。压缩方法分为无损压缩与有损压缩。

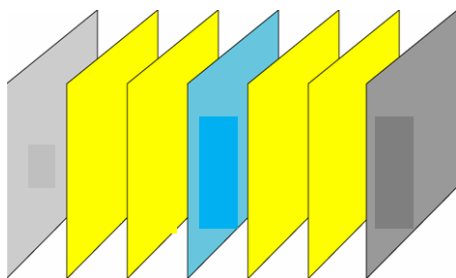


图 1-8 视频的图像构成

无损压缩是将相同的或相似的数据归类，使用较少的数据量描述原始数据，压缩率不高，一般为 2:1 到 5:1，例如哈夫曼编码。有损压缩是利用人类视觉和听觉感观对图像或声音中某些频率成分不敏感的特性，允许在压缩过程中损失一定的信息以减少数据量，压缩比可达 100:1 到 200:1。目前人们常混用这两种压缩方法。

单元 4 数的表示方法

在计算机中，数同样也是用二进制数来表示的。本单元首先将介绍与计算机相关的进制，进制之间的相互转换，然后阐述计算机中数的表示方法。

知识 1 十进制、二进制、八进制与十六进制

进位计数制是一种数的表示方法，它按进位的方式来计数，简称为进制。在计算机中使用的数制有十进制（Decimal）、二进制（Binary）、八进制（Octal）与十六进制（Hexadecimal）。每种进制数据有共同的特征。

1. 十进制数

- 十进制所采用的计数符号有十个，即 0、1、2、3、4、5、6、7、8、9。
- 十进制的基数为 10，计数的规则：逢 10 进 1，借 1 当 10。
- 数码处在数的位置不同，则数值的大小不同。例如：

1	2	3	4	.	5	7
千位	百位	十位	个位		十分位	百分位

每个数码分别代表的数值为：

千位： 1×10^3	百位： 2×10^2	十位： 3×10^1
个位： 4×10^0	十分位： 5×10^{-1}	百分位： 7×10^{-2}

我们常常把 1234.57 展开成 $1 \times 10^3 + 2 \times 10^2 + 3 \times 10^1 + 4 \times 10^0 + 5 \times 10^{-1} + 7 \times 10^{-2}$ 的求和式。

一般地, 任意一个十进制数

$$\begin{array}{ccc} a_n a_{n-1} a_{n-2} \dots a_0 & . & b_1 b_2 \dots b_m \\ \text{整数部分} & & \text{小数部分} \end{array}$$

都可以看作是下面多项式的组合求和式:

$$a_n a_{n-1} a_{n-2} \dots a_0 . b_1 \dots b_m = a_n \times 10^n + a_{n-1} \times 10^{n-1} + a_{n-2} \times 10^{n-2} + \dots + a_0 \times 10^0 + b_1 \times 10^{-1} + b_2 \times 10^{-2} + \dots + b_m \times 10^{-m}$$

其中: a_i ($i=0, 1, 2, \dots, n$) 和 b_j ($j=1, 2, \dots, m$) 均为数码 0, 1, ..., 9 中的任一个; n 、 m 为整数, 表示数码的位置; 10 为基数, 即允许选用的基本数码的个数; 10^n 是基数的权, 它确定了该数码所在位, 如 10^2 是对应百位, 10^0 是对应个位等。

2. 二进制数

- 二进制有两个计数符号(数码): 0、1, 基数 $N=2$ 。
- 二进制的进位规则: 逢 2 进 1, 借 1 当 2。
- 任一个二进制数可写成多项式表示形式。

如: $(1101)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$ 。

其中: 2^3 、 2^2 、 2^1 、 2^0 表示各位的权, 即标明对应二进制数码所在的位。

3. 八进制数

- 八进制有八个数码: 0、1、2、3、4、5、6、7, 基数 $N=8$ 。
- 八进制的进位规则: 逢 8 进 1, 借 1 当 8。
- 任一个八进制数都可写成多项式的表示形式。

如: $(563)_8 = 5 \times 8^2 + 6 \times 8^1 + 3 \times 8^0$ 。

其中: 8^2 、 8^1 、 8^0 为八进制中各位的权。

4. 十六进制数

- 十六进制有十六个数码: 0~9、A、B、C、D、E、F, 基数 $N=16$ 。
- 十六进制的进位规则: 逢 16 进 1, 借 1 当 16。
- 任一个十六进制数都可写成多项式表示形式。

如: $(FA5)_{16} = F \times 16^2 + A \times 16^1 + 5 \times 16^0$ 。

注意: 十六进制数 A、B、C、D、E、F 分别对应于十进制的 10、11、12、13、14、15。

为了区分进制数据, 在书写时有习惯的表示方法。如二进制 101101.101 写成 $(101101.101)_2$ 或 101101.101B; 八进制 37.6 写成 $(37.6)_8$ 或 37.6O; 十进制 205.8 写成 $(205.8)_{10}$ 或 205.8D, 也可以直接写成 205.8; 十六进制 3FC.6D 写成 $(3FC.6D)_{16}$ 或 3FC.6DH, 如果十六进制数第一个数字是字母, 书写时在前面加一个数字“0”, 如 FFFE_H 写成 0FFFE_H。

知识2 数制之间的相互转换

1. 非十进制数转换成十进制数(这里的“非十进制”指的是二进制、八进制与十六进制)

非十进制转换成十进制的方法是根据进制数的第三个特点, 即按权展开求和。

例 1.1 将 $(1011.101)_2$ 转换成十进制数。

$$(1011.101)_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} = 8 + 2 + 1 + 0.5 + 0.125 = (11.625)_{10}。$$

例 1.2 将 $(A3.2C)_{16}$ 转换成十进制数。

$$\begin{aligned}
 (A3.2C)_{16} &= A \times 16^1 + 3 \times 16^0 + 2 \times 16^{-1} + C \times 16^{-2} \\
 &= 10 \times 16^1 + 3 \times 16^0 + 2 \times 16^{-1} + 12 \times 16^{-2} \\
 &= 160 + 3 + 1/8 + 3/64 = (163.172)_{10}
 \end{aligned}$$

例 1.3 将 $(1657)_8$ 转换成十进制数。

$$(1657)_8 = 1 \times 8^3 + 6 \times 8^2 + 5 \times 8^1 + 7 \times 8^0 = (943)_{10}$$

2. 十进制数转换成非十进制数

十进制数转换成非十进制数分为两种情况，一种是十进制整数转化为非十进制；另一种是十进制小数（纯小数）转化为非十进制。

十进制整数化为非十进制的方法是除基取余法；十进制小数转换为非十进制则采用乘基取整法。具体转化过程如下所示。

例 1.4 将 $(25)_{10}$ 转换成二进制数。

$$\begin{array}{r}
 2 \overline{) 25} \quad \cdots \cdots 1 \uparrow \\
 \underline{2 12} \quad \cdots \cdots 0 \\
 2 \overline{) 6} \quad \cdots \cdots 0 \\
 \underline{2 3} \quad \cdots \cdots 1 \\
 2 \overline{) 1} \quad \cdots \cdots 1 \\
 \underline{0}
 \end{array}$$

$$(25)_{10} = (11001)_2$$

例 1.5 将 $(125)_{10}$ 转换成八进制数。

$$\begin{array}{r}
 8 \overline{) 125} \quad \cdots \cdots 5 \uparrow \\
 \underline{8 15} \quad \cdots \cdots 7 \\
 8 \overline{) 1} \quad \cdots \cdots 1 \\
 \underline{0}
 \end{array}$$

$$(125)_{10} = (175)_8$$

例 1.6 将 $(0.125)_{10}$ 转换成二进制数。

$$\begin{array}{r}
 0.125 \\
 \underline{2} \\
 0.250 \quad \cdots \cdots 0 \downarrow \\
 \underline{2} \\
 0.500 \quad \cdots \cdots 0 \\
 \underline{2} \\
 1.000 \quad \cdots \cdots 1
 \end{array}$$

$$(0.125)_{10} = (0.001)_2$$

例 1.7 将 $(0.625)_{10}$ 转换成十六进制数。

$$\begin{array}{r}
 0.625 \\
 \underline{16} \\
 3750 \\
 \underline{625} \\
 10.000 \quad \cdots \cdots 10 \\
 (0.625)_{10} = (0.A)_{16}
 \end{array}$$

注意：有些十进制小数在转换为非十进制数的时候，使用乘基取整法无法得到精确值。

例 1.8 将十进制小数 0.6 转换成二进制数。

$$\begin{array}{rcl}
 0.6 & & \\
 \underline{2} & & \\
 1.2 & \cdots & 1 \\
 \underline{2} & & \\
 0.4 & \cdots & 0 \\
 \underline{2} & & \\
 0.8 & \cdots & 0 \\
 \underline{2} & & \\
 1.6 & \cdots & 1 \\
 \underline{2} & & \\
 \vdots & & \vdots
 \end{array}
 \quad \downarrow$$

因此，十进制数 0.6 转化为二进制数近似等于 0.1001B。

另外，如果一个十进制数既有整数部分又有小数部分，化为非十进制，采用的转化方法就是整数采用整数转化方法，小数采用小数转化方法，然后把两部分合并。

例 1.9 把十进制数 25.125 化为二进制。转化时先把整数 25 按例 1.3 的方法转化为二进制 11001，小数 0.125 按例 1.5 的方法转化为 0.001，然后，把整数部分与小数部分合并得到 $(11001.001)_2$ 。

3. 非十进制数之间的相互转换

(1) 二进制数转换成八进制数。

规则：以小数点为中心，分别向左、向右每三位为一组，首尾组不足三位时，首尾用 0 补足，再将每组二进制数转换成一位八进制数，此方法也被称为三位分组法。

例 1.10 将 $(1010011.01011)_2$ 转换成八进制数。

$$\begin{array}{ccccccc}
 001 & 010 & 011 & . & 010 & 110 & \\
 1 & 2 & 3 & . & 2 & 6 &
 \end{array}$$

即 $(1010011.01011)_2 = (123.26)_8$ 。

(2) 八进制数转换成二进制数。

规则：将每位八进制数用三位二进制数表示即可。

例 1.11 将 $(617.34)_8$ 转换成二进制数。

$$\begin{array}{ccccccc}
 6 & 1 & 7 & . & 3 & 4 & \\
 110 & 001 & 111 & . & 011 & 100 &
 \end{array}$$

即 $(617.34)_8 = (110001111.011100)_2$

(3) 二进制数转换成十六进制数。

规则：以小数点为中心，分别向左、向右每四位为一组，首尾组不足四位时，首尾用 0 补足，再将每组二进制数转换成一位十六进制数，此方法也被称为四位分组法。

例 1.12 将 $(1101111100111.1001111101)_2$ 转换成十六进制数。

$$\begin{array}{ccccccc}
 0001 & 1011 & 1110 & 0111 & . & 1001 & 1111 & 0100 \\
 1 & B & E & 7 & . & 9 & F & 4
 \end{array}$$

即 $(1101111100111.1001111101)_2 = (1BE7.9F4)_{16}$ 。

(4) 十六进制数转换成二进制数。

规则：将每位十六进制数用四位二进制数表示即可。

例 1.13 将十六进制 B6E.9 转换成二进制数。

$$\begin{array}{ccccccc} & \text{B} & & 6 & & \text{E} & & . & & 9 \\ & 1011 & & 0110 & & 1110 & & . & & 1001 \end{array}$$

即 $(\text{B6E.9})_{16} = (1011\ 0110\ 1110.1001)_2$ 。

知识 3 二进制数之间的运算

在计算机中，二进制数之间可进行算术运算和逻辑运算，且运算规则简单。

1. 二进制的算术运算

(1) 加法运算规则。

$$0 + 0 = 0 \quad 0 + 1 = 1 \quad 1 + 0 = 1 \quad 1 + 1 = 0 \text{ (产生进位)}$$

(2) 减法运算规则。

$$0 - 0 = 0 \quad 0 - 1 = 1 \text{ (产生借位)} \quad 1 - 0 = 1 \quad 1 - 1 = 0$$

(3) 乘法运算规则。

$$0 \times 0 = 0 \quad 0 \times 1 = 0 \quad 1 \times 0 = 0 \quad 1 \times 1 = 1$$

例 1.14 完成二进制数 1101 与 1001 之间的加法、减法与乘法运算。

$$\begin{array}{r} 1101 \\ +1001 \\ \hline \text{结果: } 10110 \end{array} \quad \begin{array}{r} 1101 \\ -1001 \\ \hline \text{结果: } 0100 \end{array} \quad \begin{array}{r} 1101 \\ \times1001 \\ \hline 1101 \\ 0000 \\ 0000 \\ 1101 \\ \hline \text{结果: } 1110101 \end{array}$$

2. 二进制的逻辑运算

逻辑运算是指对因果关系进行分析的一种运算。逻辑运算的结果并不表示数值大小，而是表示一种逻辑概念，若成立，用真或 1 表示；若不成立，用假或 0 表示。二进制数的逻辑运算有“与”、“或”、“非”和“异或”四种。

(1) “与”运算 (AND)。

“与”运算又称逻辑乘，用符号“ $\wedge$ ”来表示。运算规则如下：

$$0 \wedge 0 = 0 \quad 0 \wedge 1 = 0 \quad 1 \wedge 0 = 0 \quad 1 \wedge 1 = 1。$$

即当两个参与运算的数的对应码位中有一个数为 0，则运算结果为 0，只有两码位对应的数都为 1，结果才为 1。这与前面介绍的二进制数乘法运算是一样的。

例 1.15 分别求 $10111001 \wedge 11110011$ 与 $100010101 \wedge 101111100$ 的结果。

$$\begin{array}{r} 10111001 \\ \wedge 11110011 \\ \hline 10110001 \end{array} \quad \begin{array}{r} 100010101 \\ \wedge 101111100 \\ \hline 100010100 \end{array}$$

(2) “或”运算 (OR)。

“或”运算又称逻辑加，用符号“ $\vee$ ”来表示。运算规则如下：

$$0 \vee 0 = 0 \quad 0 \vee 1 = 1 \quad 1 \vee 0 = 1 \quad 1 \vee 1 = 1。$$

即当两个参与运算的数相应码位中有一个数为 1，则运算结果为 1，只有两码位对应的数均为 0，结果才为 0。

例 1.16 分别求 $10111001 \vee 11110011$ 与 $100010101 \vee 101111100$ 的结果。

$$\begin{array}{r} 10111001 \\ \vee \quad 11110011 \\ \hline 11111011 \end{array} \quad \begin{array}{r} 100010101 \\ \vee \quad 101111100 \\ \hline 101111101 \end{array}$$

(3) “非”运算 (NOT)。

“非”运算实现逻辑否定，即进行求反运算。逻辑非通常表示为在逻辑变量上方加一横线。“非”运算规则： $\bar{0} = 1$ ， $\bar{1} = 0$ 。注意“非”运算只是针对一个数所进行的“运算”，这与前面的“与”和“或”运算不一样。它的实质意义就是取反。如 10111101，进行“非”运算后就得到 01000010，对比相应位即可验证以上运算规则了。

(4) “异或”运算 (XOR)。

“异或”运算用符号“ $\oplus$ ”来表示。其运算规则如下：

$$0 \oplus 0 = 0 \quad 0 \oplus 1 = 1 \quad 1 \oplus 0 = 1 \quad 1 \oplus 1 = 0$$

即当两个参与运算的数取值相异时，运算结果为 1，否则为 0。

例 1.17 分别求 $10111001 \oplus 11110011$ 与 $100010101 \oplus 101111100$ 的结果。

$$\begin{array}{r} 10111001 \\ \oplus \quad 11110011 \\ \hline 01001010 \end{array} \quad \begin{array}{r} 100010101 \\ \oplus \quad 101111100 \\ \hline 001101001 \end{array}$$

知识 4 数的表示方法

计算机中数的表示方法有定点格式与浮点格式两种。一般来说，定点格式允许的数值范围有限，但处理定点数的硬件比较简单，而浮点格式表示的数的范围较大，但要求的硬件比较复杂。为了清楚了解计算机中数的表示方法，首先认识计算机中有关数的 3 个概念，然后阐述计算机中定点数与浮点数的表示方法。

1. 计算机中数的有关概念

在计算机中描述一个数值型数据，需考虑数的长度、数的符号以及小数点三个问题。

(1) 数的长度。在数学中，数的长度是指它用十进制表示时所占用的实际位数，如十进制数 3456 的长度是 4。但在计算机中，数的长度按“比特”(bit，二进制位英文词 binary digit 的缩写，二进制的一位就是 1bit)来计算。同时，由于计算机的存储容量以“字节”(Byte，等于 8bit)为计量单位，所以数据长度也以字节为单位计算。在计算机中存储的数据的位都是 8 的整数倍，最小长度为 8bit。

值得注意的是：数学中的数的长度有长有短，如 753 的长度为 3，4569 的长度为 4，有几位就写几位。但在计算机中，同类型的数据（如同属整型的两个数据）的长度是统一的，不足的部分用 0 填充，这样的做法是为了简化数据的处理。换句话说，计算机中同一类型的数据具有相同的数据长度。

(2) 数的符号。由于数有正负之分，我们平常是用“+”表示正，用“-”表示负。在计算机中也需采用一种方法来描述数的符号，一般用数的最高位（左边第一位）来表示数的符号，

且约定以 0 表示正，以 1 表示负。

(3) 小数点。在现实记数中，我们使用“.”来表示小数点。在计算机中如何表示小数点的问题就同数的符号一样，也值得我们思考。在计算机中表示数值型数据，其小数点的位置是隐含的，即约定小数点的位置。

2. 定点数的表示方法

在定点数的表示方法中，小数点的位置一旦约定，就不再改变。常用的定点数表示方法有定点整数（也称纯整数）和定点小数（也称纯小数）两种。

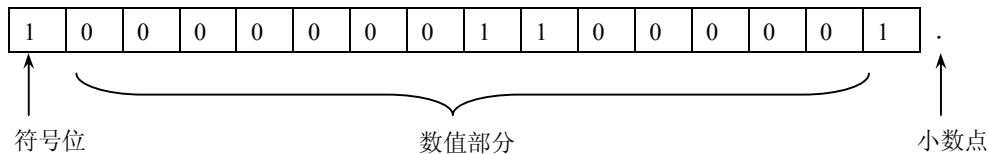
(1) 定点整数。

定点整数的小数点的位置约定在最低数值位的后面。数据存放的格式如下：

S	a_n	a_{n-1}	a_{n-2}	...	a_2	a_1	a_0
符号	量值部分						

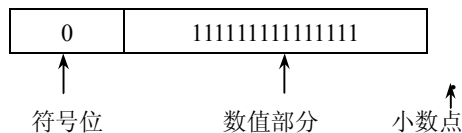
如果符号位为 0，表示该数是一个正整数，如果符号位为 1，表示该数是一个负整数。

假设这种表示方法中的量值部分存放的是数的绝对值，则用这种表示方法表示的整数的范围为： $0 \leq |x| \leq 2^{n+1}-1$ 。以 8 位存储单元为例，最高位为符号位，存放整数的范围为-1111111 到+1111111。用 16 位存储单元存放符号加绝对值定点整数-193 的格式如下：



注意：193D=(11000001)₂，由于 11000001 不足 15 位，故前面补足 7 个 0，最高位用 1 表示负数（数的符号）。

假设机器字长为 16 位，符号占一位，数值部分占 15 位，故与下面机器数等效的十进制数为+32767。

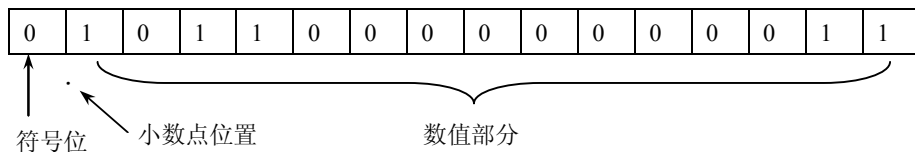


(2) 定点小数。定点小数的小数点位置约定在符号位和数值部分的最高位之间，用以表示小于 1 的纯小数。数据存放的格式如下：

S	b_1	b_2	b_3	...	b_{m-2}	b_{m-1}	b_m
符号	量值部分						

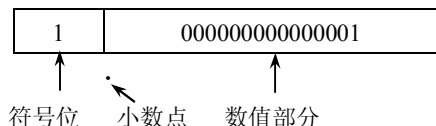
以 8 位存储单元为例，最高位为符号位，存放小数的范围为-0.1111111 到+0.1111111。

假如定点数的长度仍为 2 个字节，则十进制小数 0.6876 在计算机内用定点小数表示的形式如下：



实际上, $0.6876 = (0.101100000000011)_2$, 由于最高位表示数的符号, 故 2 个字节可以精确到小数点后第 15 位。

假定机器字长为 16 位, 符号位占 1 位, 数值部分占 15 位, 故与下面机器数等效的十进制数为 -2^{-15} 。



目前, 计算机中多采用定点纯整数表示, 因此, 将定点数表示的运算简称为整数运算。

*3. 浮点数的表示方法

大家在学物理学时就已经知道, 电子的质量为 9×10^{-28} 克, 太阳的质量为 2×10^{33} 克, 电子的质量很小, 太阳的质量很大。由于在计算机中用定点数来直接存储这些很小的数或很大的数非常困难, 因此, 必须寻找另一种数的表示方法, 这种方法就是浮点表示法。

所谓浮点表示法, 就是把一个数的有效数字和数的范围在计算机的存储单元中分别予以表示, 把数的范围和精度分别表示, 而数的小数点位置会随比例因子的不同而在一定范围内自由浮动的表示法。

大家也知道: 任意一个十进制数 N 可以写成 $N = 10^e \cdot m$, 同样, 在计算机中任意一个二进制数 N 也可以写成 $N = 2^e \cdot m$ 。其中 m 为尾数 (mantissa), 是一个纯小数, e 为指数 (exponent), 在计算机中也称作阶码。在这种记数法中, m 与 e 是有符号的, m 的符号被称为数符, e 的符号被称为阶符。

为了在计算机中存储这样的一个数, 只需存储这个数的阶符、阶码、数符与尾数就可以实现这个浮点数的存储。存储的格式如下:

e_s	$e_1 e_2 e_3 \dots e_m$	m_s	$m_1 m_2 m_3 \dots m_n$
-------	-------------------------	-------	-------------------------

其中: e_s 为阶符, 0 表示正, 1 表示负, $e_1 e_2 e_3 \dots e_m$ 为阶码, 用定点整数表示, m_s 为数符, 0 表示正, 1 表示负, $m_1 m_2 m_3 \dots m_n$ 为尾数, 用定点小数表示。这种表示方法存在一个问题, 一个数要识别两次符号, 给数据的处理带来了很大的不便。

为了规范浮点数的存储, 同时也便于软件移植, 电气与电子工程师协会 (Institute of Electrical and Electronics Engineers, IEEE) 制定了标准 IEEE 754, 在该标准中定义了 32 位与 64 位浮点数的存储格式, 一种是单精度浮点数, 用 32 位存储; 另一种是双精度浮点数, 用 64 位存储。且规定基数为 2, 阶码 e 用移码表示, 尾数 m 用原码表示, 根据原码的规格化要求最高位总是 1, 且将 1 默认存储。所谓移码, 是为了避开阶码的符号, 对每个阶码都加上一个正的常数, 称为偏移常数 (也称为 Excess 数), 使能表示的所有阶码都为正整数, 变成 “偏移” 了的阶码, 又称 “增码”。

32 位浮点数和 64 位浮点数的标准格式分别为:

单精度:	s(1bit)	e(8bit)	m(23bit)
------	---------	---------	----------

双精度:	s(1bit)	e(11bit)	m(52bit)
------	---------	----------	----------

在两种浮点数中, s 为浮点数的符号位, 0 表示正数, 1 表示负数。 m 为尾数, 用小数表

示， e 为阶码，用整数表示，小数点放在尾数域的最前面。

一个规格化的浮点数的真值必须表示为： $x=s2^e \times (1.m)$

例 1.18 $+1000111.0101$ 不是一个规范的浮点数。为了规范化，需要把它表示成 $+1.0001110101 \times 2^6$ ，这样的数就是一个规范化数。

这个数在计算机中如何存储呢？如果存储为单精度浮点数，符号位 s 为 0，尾数为 00011101010000000000000，为尾数后补 0，确保尾数为 23 位， e 为 6，但在阶码部分不是存储 00000110，而是存储 $6+127$ ，即 133，它的二进制为 10000101。

这里 127 是一个 Excess 数。这样， $+1000111.0101$ 以单精度浮点数存储格式为：

0	10000101	00011101010000000000000
符号	阶码 (Excess_127)	尾数

注意： m 前的 1 省略。

如果存储为双精度浮点数，符号位 s 为 0，尾数为 0001110101000000...0000000，为尾数后补 0，确保尾数为 52 位， e 为 6，但在阶码部分不是存储 0000000110，而是存储 $6+1023$ ，即 1029，它的二进制为 10000000101。

这里 1023 也是一个 Excess 数。这样， $+1000111.0101$ 以双精度浮点数存储格式为：

0	10000000101	000111010100000000000...000000000000000000000000
符号	阶码 (Excess_1023)	尾数

注意： m 前的 1 也省略。

浮点数以这样的方式存储，计算机是如何解译呢？也就是说，计算机是怎么知道存储的数是一个什么数呢？这里以单精度 11000000111000011000000000000000 为例来说明。

计算机将内存存储的浮点数解译的过程是：

- (1) 取最左边 1 位做符号，该浮点数最左边 1 位为 1，所以是负数。
- (2) 将符号位后的 8 位转换成十进制 (129) 并减去 127，结果为 2，也就是说指数为 2。
- (3) 在剩下的 23 位前加一个 1 和小数点，可以忽略右边所有多余的 0。本例为 1.11000011。
- (4) 根据指数的值将小数点移动到正确的位置。本例向右移动 2 位，为 111.000011。
- (5) 将整数部分化为十进制。本例整数部分的十进制为 7。
- (6) 将小数部分化为十进制。本例小数部分的十进制为 0.03125。
- (7) 将符号、整数和小数部分组合起来。因此本例存储的十进制数为 -7.03125。

从以上可以看出，计算机中浮点表示法是把一个数的有效数字和数的范围在计算机的一个存储单元中分别予以表示。在这种把数的范围和精度分别表示的方法中，数的小数点位置随比例因子的不同而在一定范围内自由浮动，这样能扩大计算机中数的表示范围。Java 平台上的浮点数类型 `float` 和 `double` 就采用了 IEEE 754 标准中所定义的单精度 32 位浮点数和双精度 64 位浮点数的格式。

知识小结

信息具有抽象性，反映客观世界，数据是具体的，用于描述客观世界。

可识别、可存储、可扩充、可压缩、可传递、可转换、在特定的范围有效是信息的主要特征。

信息技术简称为 IT 技术，现代计算机技术与通信技术是 IT 技术的两大支柱。
通用计算机按其规模、速度和功能等划分为超级计算机、大/中型计算机、工作站、微型计算机等。
超级计算机的研制能力是衡量一个国家经济实力和科学技术水平的重要标志之一。
计算机应用主要包括科学计算、过程控制、数据处理、计算机辅助、智能模拟与计算机网络等方面。
云计算，就是基于互联网的计算，云是指网络资源的分布型态。
计算机技术与通信技术的结合产生了网络技术。
互联网是计算机与计算机网络按照一定的通讯协议通信组成的庞大的交互式计算机网络。
物联网就是物物相连的互联网，是一种实现对物品的智能化识别、定位、跟踪、监控和管理的一种网络。
泛在网以无所不在、无所不包、无所不能为基本特征，以实现在任何时间、任何地点、任何人、任何物都能顺畅地通信为目标。
二进制的 1 位称为比特，字节是计算机存储数据的基本单位。
计算机处理的数据有文本、图形、图像、音频、视频信息。在计算机中这些数据都以二进制的形式存储与处理。
ASCII 码是美国标准信息交换码，采用 7 位二进制数表示 128 个不同的符号。
图像分为位图和矢量图。
未压缩的位图文件的字节数=图像分辨率×图像位深度/8
音频信号的数字化过程包括采样、量化、编码等过程。
音频采样频率越高，波形失真越小；音频量化位数越高，波形失真越小。
未压缩的音频文件大小计算方法为采样频率×量化位×声道×时间 / 8。
视频是图像（帧）在时间上的表示。
在计算机中，二进制可进行算术运算与逻辑运算，算术运算规则简单，实现较容易。逻辑运算包括“与”、“或”、“非”与“异或”运算。
计算机中的数采用二进制表示，存储数的方法有定点格式与浮点格式两种。
浮点数分为单精度与双精度两种。

活动设计

活动 1 计算机应用的讨论

目标

充分认识信息技术的作用，认识与理解计算机的应用。

组织

请把学生分成几个组，每组相互讨论。

要求

讨论形成结论，每组派代表做 PPT 在课堂上发言，教师要对讨论做总结。

内容

2013 年 6 月 9 日上午，在北京飞控中心的组织下，“神十”发射前的最后一次演练有序进行，飞船、火箭、发射场、航天员、搜救、着陆场、通讯等系统参练，旨在检验指挥系统和包括渭南、喀什、厦门、青岛、东风测控站在内的测控系统的准备情况。同时，指挥大厅也与各参练测控站间联系进行了测试，结果显示畅通，画面显示清晰。

大家知道，“神舟十号”在火箭的推动下，从地面到达 300 多公里高的太空轨道。一路上，不仅有厚厚的大气层，还有数以万计、高速飞行的太空垃圾。在“神舟十号”发射前，需要提前探知飞行途中的“障碍物”计算轨道，规避潜在危险。在“神舟十号”发射升空过程中，还需要实时观测飞行轨道调整飞行姿态。这种观测工作是由谁来完成的呢？地面雷达测量站就成了“神舟十号”的眼睛。太原卫星发射中心担负飞船上升阶段一半以上的跟踪测量任务。当前的太空不空，太空中不仅有飞行器，也有太空垃圾。目前，地球外围的太空垃圾有 1 到 2 万。这些垃圾小的可能是人造卫星的碎片、漆皮，大的可能是火箭发动机的残骸。这些没用的垃圾对火箭和飞船却是巨大的危险。因此，太原卫星发射中心需持续对太空进行科学观察，“盯紧”这些太空垃圾，并掌握它们的运行规律，为“神舟十号”安全飞行提供一条安全、干净的轨道。

对轨道环境情况的监测贯穿整个“神舟十号”任务，从发射前期到发射时，再到入轨，直到其安全返回地面。如果发现威胁，相关雷达站点就会迅速通知飞行控制中心，微调“神舟十号”的飞行姿态或轨道，以避免危险。

太原卫星发射中心还担任着外测数据测量、遥测参数测量和光学测量。所谓外测数据测量是指通过雷达等设备测量火箭发射上升段的方位、俯仰、速度等相关飞行数据。通过计算可以得知火箭是否按既定路线飞行，是否需要调整，以及如何调整。

遥测参数测量是地面遥测装置接收火箭上发出的实时数据，比如：箭体设备工作状态、温度、湿度、抖动情况等内外环境参数，让发射中心及时了解火箭的状况，同时，为今后科研及改进火箭的设计，提供重要参数。包括我们在电视直播中看到航天员在飞船内部的画面，也是通过遥测设备接收到的。还有光学测量一般是使用摄像机，拍摄火箭在空中飞行的画面。多种设备和手段，形成了一个立体式的无缝测量网络。

由于地球表面是个圆弧，一个雷达站只能测量所在水平面以上的空间，因此，“神舟十号”整个发射过程的测量，需要多个雷达站接力完成。第一棒是酒泉卫星发射中心，从发射到上升阶段的 200 多秒，第二棒就是太原卫星发射中心，从起飞后的 200 多秒到船箭分离准确进入预定轨道，最后是青岛和远洋号同时“接棒”。

问题

问题一：发射“神十”使用了哪些信息技术，主要完成哪些工作？

问题二：“神十”发射过程中要用到哪些性能的计算机，为什么？

问题三：在“神十”这项发射工程中，计算机有哪些主要应用？

活动 2 矢量图与位图的比较**目标**

认识计算机中位图与矢量图信息的表示方法，且对矢量图与位图各自的优缺点有全面的认识。

组织

请把学生分成几个组，每组相互讨论。

要求

讨论形成结论，每组派代表在课堂上发言，教师要对讨论做总结。

准备

在计算机一安装图像处理软件，请准备矢量图与位图各一个，首先记录每个文件的大小。

过程

教师使用图像处理软件把矢量图转化为同分辨率的位图或把位图转化为矢量图，观察文件大小的变化。

问题

解释观察大小变化的原因，且完成下表的填写。

比较内容	矢量图	位图
生成方法		
表现内容		
文件大小		
放大缩小效果变化情况		
计算机显示时间		

活动 3 十进制与二进制之间快速转化

目标

快速实现十进制与二进制之间的转化，进而实现十进制与八进制、十六进制之间的快速转化。

组织

请把学生分成多个组，交叉对抗，一组派代表写一个二进制或十进制，另一组派代表按顺序摆好十进制板且写出二进制或十进制。

要求

学生先学习活动引导内容或教师讲解引导内容。教师要为学生准备一副内容中提到的纸牌。对抗时使用的数不宜太大与太小。

引导内容

通过前面的介绍，十进制与非十进制之间的转换方法可以用图 1-8 来描述。

从图中可以看出，在十进制与非十进制之间的转换中，十进制与二进制之间的转换非常重要，如果能快速实现二进制与十进制之间的转换，那二进制与八进制之间、二进制与十六进制之间转换相对容易一些。但对于初学者按前面介绍的方法完成十进制与二进制之间的转换还真不是一件容易的事。那是否存在容易掌握且能快速转换的方法呢？回答是肯定的。在此介绍一种快速完成十进制与二进制之间快速转换的方法。

首先，来分析一下二进制的展开式：

$$a_n a_{n-1} a_{n-2} \dots a_0 \cdot b_1 \dots b_m = a_n \times 2^n + a_{n-1} \times 2^{n-1} + a_{n-2} \times 2^{n-2} + \dots + a_0 \times 2^0 + a_{-1} \times 2^{-1} + a_{-2} \times 2^{-2} + \dots + a_{-m} \times 2^{-m}$$

其中： a_i ($i=-m, \dots, -1, 0, 1, 2, \dots, n$) 要么为 0，要么为 1。

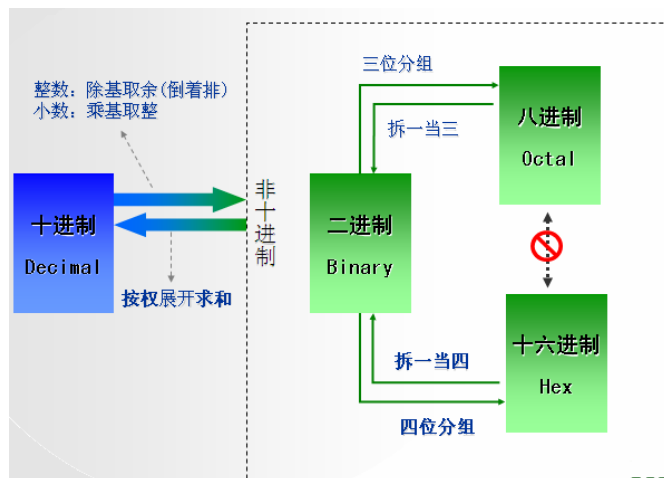


图 1-8 进制之间的转换关系

各位的位权构成数列：……1024,512,256,128,64,32,16,8,4,2,1,0.5,0.25,0.125……，这是一个等比数列，从右向左看，公比为 2。

从该数列可以看出，任何二进制都是由二进制位乘以它的位权的和来得到它的十进制。反过来说，任何二进制也可表示成上述数列中的多个数（也可以是一个）之和来得到。

例 1.19 求二进制 101110001 对应的十进制。

二进制	1	0	1	1	1	0	0	0	1
位权	256	128	64	32	16	8	4	2	1
位与位权乘积	256	0	64	32	16	0	0	0	1

二进制 101110001 的十进制为 $256+64+32+16+1=369$ 。

例 1.20 求十进制 279 的二进制。

首先把十进制数分成展开式的位权之和，即 $279=256+16+4+2+1$ 。

再按位权顺序补充如下：

256	0	0	0	16	0	4	2	1
1	0	0	0	1	0	1	1	1

则该十进制对应的二进制数为 100010111。

通过这样讲解后，在引入一种二进制与十进制相互转换的纸牌法。有这样一副纸牌，每一张牌上写有某一位二进制的位权，且对应二进制位权的牌有且仅有一张，当然还有值为 0 的空牌多张，如图 1-9 所示。

有了这样的纸牌后，用于实现十进制与二进制之间的转化后就非常容易了。

例 1.21 把二进制 101001.01 化为十进制，方法如图 1-10 所示。

该二进制对应的十进制为牌面数字之和，即 41.25。

例 1.22 把二进制 130 化为十进制，方法如图 1-11 所示。

16	32	64	128	256	512	8
0.125	0.25	0.5	1	2	4	8
0	0	0	0	0	0	0

图 1-9 二进制位权纸牌

1	0	1	0	0	1	0	1
32		8			1		0.25

图 1-10 二进制化十进制

128						2	
1	0	0	0	0	0	1	0

图 1-11 十进制化二进制

任务设计

任务 1

书面作业

目标

通过对任务的完成，系统掌握与巩固本模块所学的知识，同时检验自己对本模块知识的掌握程度。

过程

请先看题，如果你觉得该题做不好，请认真参阅教材中有关内容，把这些题写在活页纸上上交。

内容一

选择题

1. 计算机最主要的工作特点是（ ）。

- A. 存储程序与自动控制 B. 高速度与高精度
C. 可靠性与可用性 D. 有记忆能力
2. 第一台基于冯氏思想的计算机于 1950 年在宾夕法尼亚大学诞生，这台计算机被称为（ ）。
A. EDVAC B. ENIAC C. Pascal D. ABC
3. 现代计算机按（ ）划分为四代，每一代计算机的改进主要体现在硬件或软件方面变化，而不基于模型的改变。
A. 硬件 B. 软件 C. 电子元件 D. 程序
4. 目前，科学界非常看好未来计算机的发展，主要看好的未来计算机有（ ）、光子计算机和量子计算机。
A. DNA 生物计算机 B. 机械计算机
C. 光能计算机 D. 人工智能机
5. 电子计算机的发展已经历了四代，四代计算机的主要元器件分别是（ ）。
A. 电子管、晶体管、集成电路、激光器件
B. 电子管、晶体管、集成电路、大规模与超大规模集成电路
C. 晶体管、集成电路、激光器件、光介质
D. 电子管、数码管、集成电路、激光器件
6. 大规模和超大规模集成电路芯片组成的微型计算机属于现代计算机阶段的（ ）。
A. 第一代产品 B. 第二代产品
C. 第三代产品 D. 第四代产品
7. 决策支持系统简称（ ），它是以数据库、模型库和方法库为基础，帮助管理决策者提高决策水平，改善运营策略的正确性与有效性。
A. MIS B. CAD C. DSS D. EDP
8. 利用计算机进行数学定理的证明，是计算机在（ ）方面的应用。
A. 科学计算机 B. 数据处理
C. 人工智能 D. 计算机辅助
9. 利用计算机进行资料检索属于计算机应用中的（ ）。
A. 科学计算 B. 数据处理 C. 实时控制 D. 人工智能
10. 服务器是指在网络环境下运行相应的服务软件，为网上用户提供（ ）的一种高性能计算机。
A. 共享信息资源 B. 各种服务
C. 打印 D. 共享信息资源与服务
11. （ ）的研制能力是衡量一个国家经济实力和科学技术水平的重要标志之一。
A. 超级计算机 B. 图形工作站
C. 物联网 D. 微型计算机
12. （ ）的目的是将资源集中于互联网上的数据中心，由中心提供应用层、平台层和基础设施层的集中服务，以解决传统 IT 系统的零散性带来的低效问题。
A. 服务器 B. 云计算 C. 物联网 D. 泛在网
13. 在计算机应用中，“计算机辅助设计”的英文缩写为（ ）。
A. CAD B. CAM C. CAE D. CAT

14. 工厂的仓库管理软件属于()。
- A. 应用软件 B. 系统软件 C. 工具软件 D. 字处理软件
15. 下列关于“多媒体数据”说法正确的是()。
- A. 多媒体数据指的是音频数据
B. 多媒体数据指的是视频数据
C. 文本与数字不属于多媒体数据
D. 多媒体数据包含数字、文本、图像、音频和视频的信息
16. 下列字符中, ASCII 码值最小的是()。
- A. a B. A C. x D. Y
17. 英文字母“A”的十进制 ASCII 值为 65, 则英文字母“D”的十六进制 ASCII 值为()。
- A. 44 B. 81 C. 73 D. 94
18. 汉字国标码(GB2312-80)规定, 在计算机内部, 每个汉字用()。
- A. 一个字节表示 B. 两个字节表示
C. 三个字节表示 D. 四个字节表示
19. 按 16×16 点阵存放国标 GB2312-80 中一级汉字(共 3755 个)的汉字库, 大约需占()存储空间。
- A. 1MB B. 512KB C. 256KB D. 128KB
20. 在微型计算机中, 应用最普遍的字符编码是()。
- A. BCD 码 B. ASCII 码 C. 汉字编码 D. 补码
21. 下列英文单词中, () 是“十六进制”的意思。
- A. Decimal B. Binary C. Octal D. Hexadecimal
22. 以国标 GB2312-80 为基础的汉字机内码是两个字节的编码, 每个字节的最高位为()。
- A. 1 B. 0 C. 不能确定 D. 根据具体情况而定
23. 汉字国标码中共收集了()个汉字及符号。
- A. 3755 B. 3008 C. 6763 D. 7445
24. 某汉字的区位码是 5448, 它的机内码是()。
- A. D6D0H B. E5E0H C. E5D0H D. D5E0H
- 说明: 国际码=区位码 + 2020H, 汉字机内码=国际码 + 8080H。首先将区位码转换成国际码, 然后将国际码加上 8080H, 即得机内码。
25. 中国国家标准汉字信息交换编码是()。
- A. GB 231280 B. GBK C. UCS D. BIG5
26. 十进制整数 100 化为二进制数是()。
- A. 1100100 B. 1101000
C. 1000100 D. 1110100
27. 在计算机中, 数 0111111111111111 的值是()。
- A. +32767 B. 整数 C. 小数 D. 不能确定
28. 在计算机中, 一个浮点数由三部分组成, 它们是()。
- A. 符号、阶码和尾数 B. 符号、基数和尾数
C. 符号、阶码和基数 D. 符号、整数和小数

29. 下列用不同数制表示的无符号数中, 数值最大的是 ()。
- A. 二进制数 11011101 B. 八进制数 334
- C. 十进制数 219 D. 十六进制数 DA
30. 与十六进制数值 CD 等值的十进制数是 ()。
- A. 204 B. 205 C. 206 D. 203
31. 任何进位计数制都有的两要素是 ()。
- A. 整数和小数 B. 定点数和浮点数
- C. 数码的个数和进位基数 D. 阶码和尾码
32. 计算机中能实现的运算是 ()。
- A. 加法 B. 减法
- C. 乘法 D. 算法运算与逻辑运算
33. 执行二进制逻辑“与”运算 $01011001 \wedge 10100111$, 其运算结果是 ()。
- A. 00000000 B. 11111111 C. 00000001 D. 11111110
34. 执行二进制算术“加”运算 $11001001+00100111$, 其运算结果是 ()。
- A. 11101111 B. 11110000
- C. 00000001 D. 10100010
35. 一个非零的无符号二进制整数, 若在其右边末尾加上两个 0 形成一个新的无符号二进制整数, 则新的数是原来数的 () 倍。
- A. 2 B. 4 C. 8 D. 1

内容二

简答题

1. 什么是信息? 信息有哪些基本特征?
2. 什么是信息技术? 信息技术主要由哪些技术构成?
3. 目前, 计算机主要应用在哪些方面?
4. 什么是物联网? 什么是泛在网? 它们有何特征?
5. 在计算机中保存分辨率为 477×365 的 RGB 彩色位图图像, 需要存储空间是多少?
6. 请完成如下表

采样频率	量化位	立体声或单声道	1 分钟所需的字节
44.1kHz	16	立体声	
22.05kHz	8	单声道	
44.1kHz	8	单声道	
22.05kHz	16	立体声	

7. 什么是进制? 各种进制间有何共同特征?
 8. 把下列十进制数化为二进制数。
- A. 23 B. 119 C. 0.625 D. 136.125
9. 把下列二进制数分别化为十六进制数与八进制数。
- A. 1010101 B. 1001
- C. 110110110110 D. 1110101001.11011

10. 把下列二进制数化为十进制数。

A. 10101011	B. 11011
C. 1010101.101	D. 0.1101
11. 将下列二进制数分别表示成单精度与双精度浮点数。

A. $2^{-5} \times 101101.00000110011000$	B. $-2^{-5} \times 1.0110100$
--	-------------------------------
12. 把下列十进制数分别表示成单精度与双精度浮点数。

A. 12.625	B. 127.125
-----------	------------
13. 请完成下列二进制的计算。

A. $1010+1001$	B. $1111-1010$
C. $1111+1111$	D. $1010-1111$
14. 何谓定点数？定点数分为哪几种？
15. 何谓浮点数？IEEE 定义了哪两种用来在计算机内存中存储浮点数值的标准？

任务 2 文字输入

目标

熟悉键盘与指法，掌握一至两种汉字的输入方法。

内容一

完成下列英文输入：

“What student wants from our teachers?” offers us a perspective from our consumers, that is, our students. In this video, a research team conducted a series of interviews in our university campus. Students provided us thoughtful, candid, open comments on what they want from teachers. No matter who you are, as a teacher, a university administrator, or a researcher, we will learn from these interviews...

内容二

完成下列汉字输入：

当宇航员从太空中俯视我们这颗云蒸霞蔚、生机勃勃的行星时，当月球上的摄影机拍下一轮巨大的地球从月平线上升起时，我们都会为眼前的景象砰然心动。这就是我们地球母亲美丽的容颜，这就是我们人类永远的故乡。由于大气和水较多吸收太阳光谱中的红色，这颗玲珑剔透的行星便静静焕发出独特的、梦幻般的蔚蓝。

地球的年龄究竟有多大？这个难题曾经考验过许多科学家的智慧。有人想出用沉积岩形成的时间来测定，有人主张用海水含盐浓度的增加来推算，而最精确可靠、量程最大的宇宙计时器，显然要数放射性元素的蜕变了。根据对月球岩石和太阳系陨星的测定和比较，我们地球的高寿应该是 46 亿岁了。