

第 1 章 计算机数据科学

“计算机科学并不只是关于计算机，就像天文学并不只是关于望远镜一样。”

“Computer science is no more about computers than astronomy is about telescopes.”

——Edsger Dijkstra

[艾兹格·迪科斯彻 荷兰（1930-2002）著名计算机科学家]

21 世纪是崭新的数字化信息时代，以大数据为代表的信息密集型科学已成为新一次技术变革的基石。随着信息技术的飞速发展和社会竞争的日趋激烈，特别是信息化进程的日益推进，使信息管理活动日渐活跃，各种各样的信息管理系统应运而生。计算机与信息科学的基础知识已成为人们必须掌握的基本技能，无论是信息的获取和存储，还是信息加工、传输和发布，均通过计算机进行处理，并通过计算机网络有效地传送。

1.1 计算与计算机

人类的计算机科学与其他学科紧密相关（如数学、信息科学、软件工程、信息系统、信息安全、密码学、语言学、逻辑学、统计学和经济学等），这些学科之间有明显的交叉领域，但也有各自的差异。计算机硬件的研究被看成是计算机工程的一部分，商业计算机系统的研究和部署被称为信息技术或者信息系统。

1.1.1 计算的概念

计算（Compute）是数学领域的基础技能，也是认识整个自然科学的工具。人们学习知识最初都是从计算开始，掌握计算更是基本的生存技能。计算在诸多的基础教育及相关学科得以广泛的应用，尤其在计算科学、高性能计算及相关先进技术方面都发挥着其精悍的作用。

狭义的计算包括数据、计算式、运算符或算子以及计算结果。计算中的关系是计算原理中必须阐明的理论基础，包括自然数据、人工数据、人工处理等。

广义的计算则包括数学计算、逻辑推理、文法产生式、集合论相关函数、组合数学置换、变量代换、图形图像变换、数理统计等。还包括人工智能、解空间的遍历、问题求解、图论路径、网络安全、代数系统理论、上下文表示、感知与推理、智能空间、建筑设计等。

计算是对特定数据元的计算，因此数据元的性质对运算符的选择、计算的实现有决定性作用。计算表达式常有不同的形式，包括：代数式、方程、函数、行列式、微积分或者数理统计计算式等。相同的计算常有不同的性质，如线性代数中同样是三矩阵乘法，在相似变换中只要求可逆矩阵，而在二次型的标准型变换中，则要求正交矩阵。

1.1.2 计算科学

计算科学（Computational Science）是围绕着数据和数据处理的科学，精确比较强调两者之间区别，涉及诸多领域。计算科学（或者科学计算）是关注构建数学模型和量化分析技术

的研究领域，实际研究通常是计算机模拟和计算等形式在各个学科问题中的应用。自然科学研究通常可用各种类型的数学方程式表达，计算科学将寻找这些方程式的数值解，其计算涉及庞大的运算量，是简单的计算工具是难以胜任的。在计算机出现之前，科学研究和工程设计主要依靠实验或试验提供数据，计算仅处于辅助地位。随着计算机的迅速发展，使越来越多的复杂计算成为可能。利用计算机进行科学计算为社会带来了巨大的经济效益，同时也使科学技术本身发生了根本性的变化，传统的科学技术只包括理论和实验两个组成部分，使用计算机后，计算已成为与理论和实验同等重要的第三个组成部分。

1.1.3 计算机科学

计算机科学（Computer Science）是研究计算机和可计算系统及相关信息处理的科学。计算机科学涵盖了从算法的理论研究和计算的极限到如何通过硬件和软件实现计算系统，通常应用于创造、描述及转换信息算法处理的系统研究中。计算机科学植根于电子工程、数学和语言学，是科学、工程和艺术的结晶。计算机科学分为理论计算机科学和实验计算机科学两个部分：理论计算机科学一般指信息、数学、自然科学文献中所说的计算机科学；实验计算机科学主要指计算科学领域新的应用及验证的研究。计算机科学主要围绕着：①计算机程序能做什么和不能做什么（可计算性）；②使程序更高效地执行特定任务（算法和复杂性理论）；③程序构造存取不同类型的数据（数据结构和数据库）；④程序实现更智能（人工智能）；⑤便捷人类与程序沟通（人机互动和人机界面）等方面，是计算机系统中实现与应用的实用技术的研究。

1.1.4 计算机产生

计算机是实现算术和逻辑运算的机器，处理的对象都是信息。20世纪30年代，英国数学家图灵和美国数学家波斯特几乎同时提出了理想计算机的概念。到了40年代数字电子计算机产生后，计算技术和有关计算机的理论研究开始得到发展。这方面构成了所说的理论计算机科学，现实计算机及其程序的数学模型性质的研究及计算复杂性（早期称作计算难度）的研究迅速发展起来，形成自动机论、形式语言理论、程序设计理论、算法设计与分析和计算复杂性理论等。40年代后期形成存储程序型，即“冯·诺伊曼型”机器。自50年代以来，计算机的性能在计算速度和编址空间方面提高了几个数量级。60年代出现了大程序，但大程序的可靠性很难保证，西方国家出现了“软件危机”（指有些程序过于庞大，包含几十万条以至几百万条指令，成本过高而可靠性则比较差）。70年代程序设计发展起来，逐渐机器可以模拟人用探索法解题的思维活动。21世纪，计算机成为用于高速计算的电子计算机，实现数值计算、逻辑计算，还有存储记忆功能，程序运行时自动、高速地处理海量数据。出现了超级计算机、工业控制计算机、网络计算机、个人计算机、嵌入式计算机，较先进的计算机还有生物计算机、光子计算机、量子计算机等（将在1.3节详细介绍）。

大数据时代越来越多地融合了各类计算机相关学科的思想，计算机对人类的生产活动和社会活动产生了极其重要的影响，并以强大的生命力飞速发展，带动了全球范围的技术进步，引发了深刻的社会变革。目前计算机已遍及一般学校、企事业单位，进入寻常百姓家，成为信息社会中必不可少的工具，然而其是否也推进了人类的思维方式呢？

1.2 计算思维

思维的本质是心里计算的过程，是由一系列知识所构成的完整解决问题的思路，思维具有普适性、联想性、启迪性及拓展性等。人类所学到的知识和技能具有时间局限性，然而思维则可跨越时间。随着时间的推移，知识和技能可能被遗忘，但思维会逐渐融入到未来的创新活动中，产生新理念。人们与生俱来的思维能力，同时会应用到读、写、算等能力中。计算思维同样是人们理解现实世界中事物间利用计算系统实现控制和处理的过程。

1.2.1 计算思维概念

计算思维 (Computational Thinking) 是运用计算机科学的基础概念去求解问题、设计系统和理解人类行为的一系列思维活动的统称，2006 年由美国卡内基·梅隆大学周以真教授首次提出。计算思维与理论思维、实验思维并称三大科学思维，与理论思维、实验思维一样，是一种抽象思维。如同所有人都具备是非判断、文字读写和进行算术运算能力一样，如同印刷出版促进了 3R 的传播一样，计算思维涵盖了反映计算机科学在内的一系列广泛的思维活动。计算思维涉及运用计算机科学的基础概念去求解问题、设计系统和理解人类的行为，例如计算博弈理论正改变着经济学家的思考方式，再如纳米计算改变着化学家的思考方式，还有量子计算改变着物理学家的思考方式。当接触函数时，指定一个输入，就会有对应的输出关系。举一个现实中的场景：在餐厅吃饭，服务员点菜，随后服务员把做好的菜端上来。运用计算思维分析餐厅的厨房可以看成是一个函数，点菜单是传递给这个函数的参数；厨师在厨房里做菜的过程是这个函数的执行过程；做好的菜是返回结果，回显到餐桌上。尽管计算机学科研究涉及面广，但其共同特征还是基于不同层次的计算环境下专业系统问题的求解过程，需要潜移默化地培养用计算思维理念解决自身专业问题的能力。

1.2.2 计算思维特征

计算思维特征 (Computational Thinking Characteristic) 是人机结合的思维特征，其 6 大特征表现为抽象性、构造性、数字化、系统化、网络化和虚拟化。随着问题复杂度的提高和问题规模的增大，计算工具逐渐由单机变成了网络。使用系统化思维和网络化思维来解决问题，最终要在虚拟的计算机世界中实现。以大数据时代为背景产生的计算思维特征有如下 6 个方面：

- ① 计算思维具有概念化，并不是程序化；
- ② 计算思维是人类自身所固有的，并不是计算机的思维方式；
- ③ 计算思维是思维的体现，并不是人所造出的产物；
- ④ 计算思维是数学和工程思维的互补与融合的产物；
- ⑤ 计算思维是根本的思维方式，并不是刻板的技术演变；
- ⑥ 计算思维面向所有人、所有地方，放之四海而皆准。

1.2.3 计算思维养成

计算机实现了计算执行的自动化，人们将解决问题的步骤和方法以算法的形式表示出来，

并用计算机能够直接识别和运行的语言告诉计算机如何工作，逐渐成为思维的执行者。解决问题的方式逐渐演变为以计算机能够识别的方式操作，问题的解决变成了算法的描述及程序的编写。利用计算机结合数学、物理、经济、军事、社会、生活等各方面解决问题，形成以计算机为工具来解决问题的思维方式，与各个学科交叉融合，形成了新的独特的解决问题的计算思维方式。

抽象性是计算思维的基础，在形成一切计算的发展过程中，抽象性满足如下 3 个必要前提条件：

- ① 问题能形式化，即可用完备的方法建立问题的模型；
- ② 问题可计算，即可找到求解问题的算法并能用程序实现；
- ③ 问题有复杂度，即程序可在有限计算空间和时间内运行出结果。

构造性是运用计算机对抽象模型进行自动化求解。计算机算法是一个有穷规则的集合，其规则规定了一个解决某一特定类型问题的运算序列，构造出实现高效算法的步骤和方法，准确描述解题。

数字化是符号化在计算机中的特定表现，符号化是各种抽象的基本特征。计算思维是一种符号化的抽象思维。计算思维中符号抽象概念比理论思维和实验思维中的意义要丰富和复杂。

系统化是思维首要的世界观，世界由无数复杂的系统组成。采用系统化的思维全面地审视面临的问题，就是把认识对象作为系统，从系统与要素、要素与要素、系统与环境的相互联系和相互作用中综合考察认识对象。计算思维的系统化来源于人类的系统化思维，人类运用系统化思维对复杂世界系统的事物和问题进行抽象、建模和构造算法。

网络化是思维源于信息网络的人机结合，人脑就是复杂的网络化结构。解题思维需要在计算机网络环境得以有效执行，计算工具影响人类思维，促使了人类需要采取网络化的思维来描述解决问题。

虚拟化是将人类解题的思维以计算机能够识别的方式一步一步显性地描述出来，并在一个虚拟的计算机世界中执行人类的思维。虚拟思维把人们头脑里看不见、摸不着却又实实在在存在的思维过程在虚拟环境中展现，使实践和思维具有了实现感。

1.2.4 计算思维应用

理解计算思维的关键是把握计算思维的学科特征和计算思维的属性。为了培养计算思维能力，首先要树立计算思维的世界观，在这种观念的指导下才能更好、更快地理解计算思维和提高计算思维的能力。计算工具影响思维方式，思维方式体现计算工具的特点。在计算思维中，计算的本质是抽象的自动执行。确定合适的抽象条件，选择合适的某类计算机去自动解释执行该抽象。以计算机识别的方式建模并构造解决问题的高效算法，然后在计算机上运行并调试算法。将问题及其解法或算法采用符号化的形式抽象出来，与特定的计算机或计算机语言相关。优化是直接触及程序本身，提高程序运行效率，降低复杂度，降低冗余，合理选择数据类型等的优化规则，自动实现一些相关性能，在具体的运行环境中采用合适的计算机语言完成抽象的解释和自动运行。计算机硬件系统和软件系统在不同的抽象层次上提供了问题求解的计算环境。计算思维应用通常有 5 个基本流程步骤，如图 1-1 所示，是通过操作过程的构造与实施，来解决给定问题的思维方式。

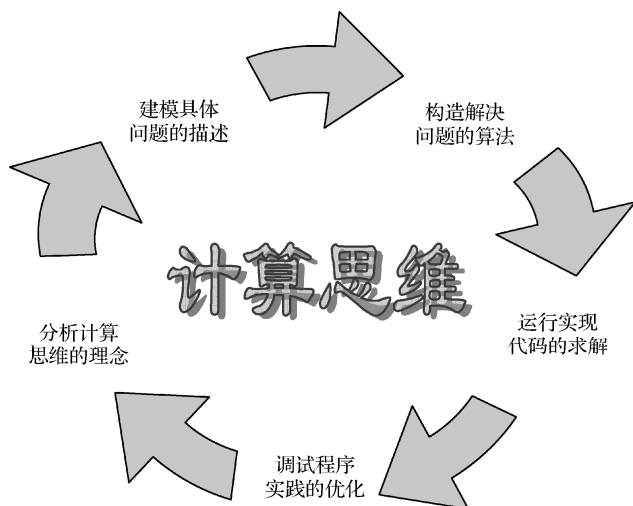


图 1-1 计算思维应用流程

1.3 信息技术与计算机文化

人类自进入文明社会以来，利用大脑存储信息，使用语言交流和传播信息，人类的信息活动从具体到抽象，使人同动物彻底分离。文字的产生和使用可以记载、传递及交流信息，纸张和印刷术的发明成为信息记载和信息传递的载体，电报、电话、广播和电视的发明和普及大大缩短了人们交流信息的时空界限。当代的信息数字化依托于互联网、物联网、电子计算机、现代通信及控制技术的发展和运用，计算机的诞生改变了人们几千年的信息处理方式。计算机目前已成为各行各业必不可少的、最基本和最通用的工具之一。

1.3.1 信息技术

信息技术 (Information Technology) 是在信息科学的基本原理和方法的指导下扩展人类信息功能的技术，是实现信息化的核心手段。信息技术是以电子计算机和现代通信为主要手段实现信息的获取、加工、传递和利用等功能的技术总和。人的信息功能包括感觉器官承担的信息获取功能、神经网络承担的信息传递功能、思维器官承担的信息认知功能和信息再生功能、效应器官承担的信息执行功能。按扩展人的信息功能分类，信息技术可分为以下 4 个方面技术。

1. 传感技术

传感技术是信息的采集技术，对应于人的感觉器官，它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。信息识别包括文字识别、语音识别和图形识别等，通常采用一种叫做“模式识别”的方法。传感技术、测量技术与通信技术相结合而产生的遥感技术，能使人感知信息的能力得到进一步的加强。

2. 通信技术

通信技术是信息的传递技术，对应于人的神经系统的功能。它的主要功能是实现信息快速、可靠、安全地传递，各种通信技术都属于这个范畴，如广播技术。由于存储、记录可以

看成是从“现在”向“未来”或从“过去”向“现在”传递信息的一种活动，因此也可将其看成是信息传递技术的一种。

3. 计算机技术

计算机技术是信息的处理和存储技术，对应于人的思维器官。计算机信息处理技术主要包括对信息的编码、压缩、加密和再生等技术。计算机存储技术主要包括计算机存储器的读写速度、存储容量及稳定性的内存储技术和外存储技术。

4. 控制技术

控制技术是信息的使用技术，是信息过程的最后环节，对应于人的效应器官。它包括调控技术、显示技术等。

由上可见，传感技术、通信技术、计算机技术和控制技术是信息技术的四大基本技术，其主要支柱是通信（Communication）技术、计算机（Computer）技术和控制（Control）技术，即“3C”技术。信息技术的四种技术划分只是相对的、大致的，没有截然的界限，如传感系统里也有信息的处理和收集，而计算机系统里既有信息传递，也有信息收集的问题等。

1.3.2 计算机发展

电子计算机的诞生是科学技术史上的里程碑，它是一种能够自动、高速、精确地进行各种信息处理的电子设备。电子数字计算机是一种不需要人的干预，能够自动连续地、快速、准确地完成信息存储、数值计算、数据处理和过程控制等多种功能的电子机器。电子逻辑器件是电子机器的物质基础，其基本功能是进行数字化信息处理，人们常称其为“计算机”，又因其工作方式与人的思维过程十分类似，亦被叫做“电脑”。

现代计算机孕育于英国，诞生于美国。1936年英国科学家图灵于伦敦权威的数学杂志发表了一篇著名的论文《理想计算机》，其中提出了著名的“图灵机”（Turing Machine）的设想。图灵机由3部分组成：一条带子，一个读写头和一个控制装置，阐述了“图灵机”不是一种具体的机器，而是一种理论模型，用来制造一种十分简单，但运算能力极强的计算装置，后续人们称图灵为“计算机理论之父”，如图1-2所示。

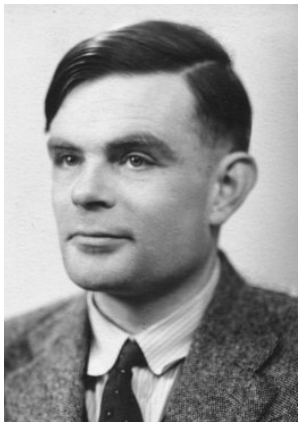


图1-2 “计算机理论之父”图灵

世界上第一台电子数字计算机是1946年2月14日在美国宾夕法尼亚大学由John Mauchly和J.P. Eckert为导弹设计服务小组制成的ENIAC（Electronic Numerical Integrator and Computer），是电子数字积分计算机。它使用了18800个电子管，150多个继电器，耗电150kW，占地面积170m²，重量达30t，每秒钟只能完成5000次加法运算，运算精确度和准确率是史无前例的，以圆周率（ π ）的计算为例，中国古代科学家祖冲之利用算筹，耗费15年心血，才把圆周率计算到小数点后7位数，一千多年后，英国人威廉·山克斯以毕生精力计算圆周率，计算到小数点后707位，而使用ENIAC进行计算仅用了40s就达到了这个记录，还发现威廉·山克斯计算结果中的第528位计算有误。虽然ENIAC体积大、速度慢、能耗大，但它却为发展电子计算机奠定了技术基础，开辟了计算机科学技术的新纪元。

在 ENIAC 计算机研制的同时，冯·诺依曼与莫尔合作研制了 EDVAC 计算机。它采用存储程序方案，此种方案沿用至今，所以现在的计算机都被称为以存储程序原理为基础的冯·诺依曼型计算机。计算机发展经历了 4 个阶段，在推动计算机发展的诸多因素中，电子器件的发展起着决定性的作用，同时计算机系统结构和计算机软件的发展也起着至关重要的作用。

1. 第一代计算机

第一代计算机称为电子管计算机，从 1946 年到 1958 年。其特征是采用电子管作为计算的逻辑元件；计算机体积庞大，可靠性差，输入/输出设备有限，使用穿孔卡片；主存容量为数百字节到数千字节，主要以单机方式完成科学计算；数据表示主要是定点数；用机器语言或汇编语言编写程序，体积大、能耗高、速度慢、容量小、价格昂贵，应用也仅限于科学计算和军事方面。

2. 第二代计算机

第二代计算机称为晶体管计算机，从 1958 年到 1964 年。其特征是采用晶体管代替了电子管；用磁芯和磁盘作为主存储器；体积、重量和功耗方面都比电子管计算机小很多，运算速度进一步提高，主存容量进一步扩大；软件有了很大发展，出现了 FORTRAN、COBOL、ALGOL 等高级语言，以简化程序设计；计算机不但用于科学计算，而且用于数据处理，并开始用于工业控制。有代表性的计算机是 IBM 公司生产的 IBM-7094 计算机和 CDC 公司的 CDC 1604 计算机。

3. 第三代计算机

第三代计算机称为中、小规模集成电路计算机，从 1964 年到 1975 年。其特征是集成电路 IC (Integrated Circuit) 代替了分立元件；用半导体存储器逐渐取代了磁芯存储器；采用了微程序控制技术；在软件方面，操作系统日益成熟，其功能日益强化，多处理机、虚拟存储器系统，以及面向用户的应用软件的发展，大大丰富了计算机软件资源。

4. 第四代计算机

第四代计算机称为大规模和超大规模集成电路计算机，从 1975 年至今。其特征是以大规模集成电路 LSI (Large-Scale Integration) 或超大规模集成电路 VLSI 为计算机主要功能部件；主存储器也采用集成度很高的半导体存储器；软件方面发展了数据库系统、分布式操作系统等。此时出现了微型机，由于微型机体积小、功耗低、成本低，其性价比优于其他类型的计算机，目前 1 个超大规模集成电路芯片包含 14 亿以上的晶体管，因此得到广泛的应用。

5. 新一代计算机

神经计算机是模仿人的大脑判断能力和适应能力，并具有可并行处理多种数据功能的神经网络计算机。它本身可以判断对象的性质与状态，并采取相应的行动，同时并行处理实时变化的大量数据并引出结论。神经计算机类似于智能生物的大脑，可以完成类似于生物大脑的复杂计算，甚至可以完成了类似于写作等复杂功能。

目前，世界上各先进国家正在加紧研制新一代计算机，其不仅是在原有结构的基础上进行器件的更新换代，未来将突破冯·诺依曼型计算机的结构，研制具有知识库管理功能的、高度并行的智能计算机。计算机正向以下 5 个方面发展。

(1) 巨型化。天文、军事和仿真等领域需要进行大量的计算，要求计算机有更高的运算速度和更大的存储容量，这就需要研制功能更强的巨型计算机。

(2) 微型化。微型计算机已经广泛应用于仪器、仪表和家用电器中，并大量进入办公室和家庭。但人们需要体积更小、更轻便、易于携带的微型计算机，以便出门在外或在旅途中使用，便携式微型计算机和掌上微型计算机正在适应用户的需求，迅速普及。

(3) 多媒体化。多媒体计算机是利用计算机技术、通信技术和大众传播技术综合处理多种媒体技术信息的计算机，这些信息包括数字、文本、声音、视频和图形图像等，使多种信息建立了有机的联系，集成为一个系统，并具有交互性。多媒体计算机将改善人机界面，使计算机朝着人类接收和处理信息的最自然的方式发展。

(4) 网络化。网络可以使分散的各种资源得到共享，使计算机的实际效用提高了很多。计算机连网不再是可有可无的事，人们足不出户就可获得所需的信息和服务，与世界各地快捷通信，网上贸易是计算机应用的重要组成部分。

(5) 智能化。目前计算机已能够部分代替人脑劳动，但是人们希望计算机具有更多的类似人的功能。科学家们正在研制不使用集成电路的计算机，如生物计算机、光子计算机和量子计算机等。

1.3.3 计算机的分类

随着计算机技术的发展和应用的普及，尤其是微处理器的发展，使得计算机的类型越来越多样化，其按照不同的原则有如下分类方法。

1. 按信息在计算机中处理方式分类

(1) 数字计算机

数字计算机是当今世界电子计算机行业中的主流，其内部处理的是一种称为符号信号或数字信号的电信号。它采用二进制运算，其主要特点是：离散在相邻的两个符号之间不可能有第三种符号存在；解题精度高，便于存储，是通用性很强的计算工具，既能胜任科学计算和数字处理，也能进行过程控制和 CAD/CAM 等工作。由于存在这种处理信号的差异，使得它的组成结构和性能优于模拟式电子计算机。

(2) 模拟计算机

模拟计算机问世较早，用内部所使用的电信号模拟自然界的实际信号，因而称为模拟电信号。模拟计算机处理问题的精度差；所有的处理过程均需模拟电路来实现，电路结构复杂，抗外界干扰能力较差。

模拟计算机的机器变量是连续变化的电压变量，对于变量的运算是基于电路中电压、电流、元件等电特性的相似运算关系。通用电子模拟计算机的组成包括：线性运算部件（比例器、加法器、积分器等）、非线性运算部件（函数产生器、乘法器等）、控制电路、电源、排线接线板、输出显示与记录装置。

模拟计算机特别适合于求解常微分方程，也被称为模拟微分分析器。物理系统的动态过程多数是以微分方程的数学形式表示的，所以模拟计算机很适用于动态系统的仿真研究。模拟计算机在工作时是把各种运算部件按照系统的数学模型连接起来，并行地进行运算，各运算部件的输出电压分别代表系统中相应的变量。因此，模拟计算机具有处理速度高，可以直观表示出系统内部关系的特点。

(3) 数字模拟混合计算机

数字模拟混合计算机是取数字、模拟计算机之长，既能高速运算，又便于存储信息。但这类计算机造价昂贵，目前所使用的大部分计算机属于数字计算机。

(4) 光计算机

激光计算机的核心部分处理机是用激光产生的光波代替电波进行 0 和 1 的转换。处理机是计算机的心脏，接收各种信号或资料，根据程式指令加以处理，然后以新的形式输出，图 1-3 为激光计算机。光子计算机是一种由光信号进行数字运算、逻辑操作、信息存储和处理的新型计算机。不同于电子计算机对电子的控制，光子计算机的运行依靠激光器、光学反射镜、透镜、滤波器等光学元件和设备对光子的控制完成光运算，目前美国的贝尔实验室已经研发出了世界上第一台光子计算机。

2. 按功能分类

(1) 专用计算机

专用计算机用于解决某个特定方面的问题，配有为解决某问题使用的软件和硬件。专用计算机功能单一，可靠性高，结构简单，适应性差。但在特定用途下最有效，最经济，最快速，是其他类型计算机无法替代的。如军事系统专用计算机、银行系统专用计算机、生产过程的自动化控制、数控机床等。

(2) 通用计算机

通用计算机功能齐全，适应性强，用于解决各类问题。它既可以进行科学计算，也可以用于数据处理，通用性较强。目前所使用的大部分计算机属于通用计算机。

3. 按规模分类

按照计算机规模，并参考其运算速度、输入/输出能力、存储能力等因素划分，通常将计算机分为巨型机、大型机、小型机、微型机等。

(1) 巨型机

巨型机运算速度快，存储量大，结构复杂，价格昂贵，主要用于尖端科学研究领域，如 IBM390 系列、银河机等。1994 年，银河 II 超级计算机在国家气象局正式投入运行，用于天气中期预报。图 1-4 为 10 亿次银河 II 巨型计算机。目前还有速度达到千万亿次的曙光机。



图 1-3 激光计算机



图 1-4 银河 II 10 亿次并行巨型计算机

(2) 大型机

大型机规模次于巨型机，有比较完善的指令系统和丰富的外部设备，主要用于计算机网

络和大型计算中心，图 1-5 为 IBM 大型计算机。

大型机一直都是服务器的创新之源，随着其技术不断下移，Power 平台、x86 平台都得到了前所未有的强化。大型机不仅没有走向弱势，而且形成了更为丰富的外延产品圈，可以全方位地满足不同类型的客户需要。

(3) 小型机

小型机比大型机成本低，容易维护。小型机是指采用 8-32 颗处理器，性能和价格介于 PC 服务器和大型主机之间的一种高性能 64 位计算机。国外小型机对应英文名是 Minicomputer 和 Midrange Computer。Midrange Computer 是相对于大型主机和微型机而言，该词汇经常被国内一些教材误译为中型机，Minicomputer 一词是由 DEC 公司（数字设备公司）于 1965 年创造。国内小型机习惯上用来指 UNIX 服务器。1971 年贝尔实验室发布多任务多用户操作系统 UNIX，随后被一些商业公司采用，成为后来服务器的主流操作系统。在国外，小型机是一个已经过时的名词，60 年代由 DEC 公司首先开发，并于 90 年代消失。小型机用途广泛，现可用于科学计算和数据处理，也可用于生产过程自动控制和数据采集及分析处理等，图 1-6 为小型机。



图 1-5 IBM 大型计算机



图 1-6 小型机

(4) 微型机

微型机由微处理器、半导体存储器 and 输入/输出接口等芯片组成，使得它比小型机的体积更小、价格更低、灵活性更好、可靠性更高、使用更加方便。

4. 按工作模式分类

(1) 服务器

服务器 (Server) 是一种可供网络用户共享的高性能计算机。服务器一般具有大容量的存储设备和丰富的外部设备、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，是网络的中枢和信息化的核心。由于要运行网络操作系统，所以需要其具有较高的运行速度、处理能力强、稳定性好、可靠性高、安全性好、易扩展性和可管理性等。服务器所面对的是整个网络的用户，需要 7×24 小时不间断工作，对稳定性要求极高。另一方面，为了实现高速以满足众多用户的需求，服务器通过采用对称多处理器 (SMP) 安装，插入大量的高速内存来保证工作。其主板可以同时安装几个甚至几十、上百个 CPU (厂商专门为服务器开发生产的)，图 1-7 为浪潮系列服务器。

(2) 工作站

工作站 (Workstation) 是以个人计算机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, 为满足工程设计、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需求而设计开发的高性能计算机。图 1-8 为一体化工作站。

无盘工作站是指无软盘、无硬盘、无光驱连入局域网的计算机。在网络系统中, 把工作站端使用的操作系统和应用软件全部放在服务器上, 系统管理员只要完成服务器上的管理和维护即可, 软件的升级和安装也只需要配置一次, 整个网络中的所有计算机就都可以使用新软件了。所以无盘工作站具有节省费用、系统的安全性高、易管理性和易维护性等优点。



图 1-7 浪潮系列服务器



图 1-8 一体化工作站

1.3.4 计算机的特点和应用

早期的计算机仅仅用于数值计算, 随着计算机技术的发展, 计算能力的提高, 计算机广泛应用到各个领域。计算机主要具有处理速度快、运算精度高、存储能力强、逻辑判断好、可靠性高、通用性强和自动化等特点。

(1) 处理速度快。处理速度是计算机的一个重要性能指标, 计算机的处理速度可以用每秒钟执行加法的次数来衡量。计算机的运算速度已由早期的每秒几千次发展到现在的每秒几千亿次甚至万亿次。其高速度处理能力把人们从浩繁的脑力劳动中解放出来, “瞬间”即可完成人类旷日持久的运算工作, 这也是计算机被广泛使用的主要原因之一。

(2) 运算精度高。计算机对数据运算的结果精度可达到十几位甚至几十位有效数字, 根据需要可达到更高的精度。

(3) 存储能力强。计算机超大的存储容量可存储海量信息, 并且可存储当时没有做完的工作, 放到计算机的“记忆”中, 在任意时间再拿出来使用或继续完成。

(4) 逻辑判断好。计算机不但能完成各类算术运算, 而且还具有进行比较和判断等逻辑运算的功能, 使计算机能够处理逻辑推理问题, 是实现信息处理自动化的前提。

(5) 可靠性高。由于采用集成电路技术, 计算机具有非常高的可靠性, 可连续无故障地运行几个月甚至几年。

自 1946 年第一台电子数字计算机诞生以来, 人们一直在探索计算机的应用模式, 尝试利用计算机去解决各领域中的问题, 计算机的应用主要有以下 8 个方面。

(1) 科学计算。即数值计算, 科学和工程计算的特点是计算量大, 逻辑关系相对复杂。例如: 卫星轨道计算, 导弹发射参数的计算, 宇宙飞船运行轨迹和气动干扰的计算等。

(2) 信息处理。即数据处理或事务处理, 是指对各种信息进行收集、存储、加工、分析和统计, 向使用者提供信息存储、检索等一系列活动的总和。例如: 银行储蓄系统的存款、取款和计息, 办公自动化中利用计算机进行信息处理, 图书、书刊、文献和档案资料的管理和查询等。

(3) 过程控制。即自动控制, 利用计算机对动态的过程进行控制、指挥和协调, 由计算机对采集到的数据按一定方法经过计算, 然后输出到指定执行机构去控制生产的过程。例如: 在化工厂可用来控制化工生产的某些环节或全过程等。

(4) 计算机辅助系统。即设计人员使用计算机进行设计的一项专门技术, 用来完成复杂的设计任务。它不仅应用于产品和工程辅助设计, 还包括辅助制造、辅助测试、辅助教学以及其他许多方面的内容, 这些都统称为计算机辅助系统(也称为 CAS 技术), 举例如下:

- ① 计算机辅助设计 CAD (Computer Aided Design);
- ② 计算机辅助制造 CAM (Computer Aided Manufacture);
- ③ 计算机辅助教学 CAI (Computer Aided Institute);
- ④ 计算机辅助测试 CAT (Computer Aided Test);
- ⑤ 计算机基础教育 CBE (Computer Based Education);
- ⑥ 计算机集成制造系统 CIMS (Computer Intergrated Manufacturing System)。

(5) 人工智能。AI (Artificial Intelligence) 是计算机模拟人类大脑的高级思维活动, 详细介绍见 1.5.2 节。

(6) 电子商务。EC (Electronic Commerce) 广义上指使用各种电子工具从事商务或活动, 狭义上指基于浏览器/服务器应用方式, 利用 Internet 从事商务或活动。电子商务涵盖的范围很广, 一般可分为企业对企业 (Business-to-Business), 或企业对消费者 (Business-to-Consumer) 两种。例如: 消费者的网上购物, 商户间网上交易和在线电子支付等。

(7) 多媒体应用。多媒体计算机具有集成性和交互性的特点, 集文字、声音、图像等信息于一体, 并使人机双方通过计算机进行交互。多媒体技术的发展大大拓宽了计算机的应用领域, 视频、音频信息的数字化使得计算机走向家庭, 走向个人。

(8) 网络通信。即融合计算机技术和数字通信技术产生计算机网络, 数字化时代通过网络通信和计算机系统的联合, 实现网上预约、网上购物及订票订餐等, 人们正在享受网络的乐趣, 其正在逐渐改变人们的工作和生活方式。

计算机在社会各领域中的广泛应用, 有力地推动了社会的发展和科学技术水平的提高, 同时也促进了计算机技术的不断更新, 使其向微型化、网络化、智能化的方向不断发展。

1.4 数据科学

信息化是一个生产数据的过程, 现实世界中的事物和现象以数据的形式存储到信息空间中。数据被快速大量地生产并存储在信息空间, 此种现象称为数据爆炸 (Data Explosion)。探索数据的规律和现象是探索宇宙的规律、探索生命的规律、寻找人类行为的规律、寻找社会发展的规律的一种重要手段。

1.4.1 数据科学概念

数据科学 (Data Science) 是数据支持决策的基础数据科学, 采用科学方法, 运用数据挖掘工具对复杂多量的数字、符号、文字、网址、音频或视频等信息进行数字化重现与认识。数据科学主要有两个内涵: 一个是数据本身的研究 (包括各种类型、状态、属性及变化形式和变化规律); 另一个是自然科学和社会科学研究, 作为科学研究的数据方法手段, 其目的在于揭示自然界和人类行为的现象和规律。

1.4.2 数据科学发展

数据科学在 20 世纪 60 年代已被提出, 当时并未获得学术界的注意和认可。1974 年彼得·诺尔出版了《计算机方法的简明调查》, 其中将数据科学定义为: 处理数据的科学, 一旦数据与其代表事物的关系被建立起来, 将为其他领域与科学提供借鉴。2001 年美国统计学教授威廉·S·克利夫兰发表了《数据科学: 拓展统计学的技术领域的行动计划》, 因此有人认为是克利夫兰首次将数据科学作为一个单独的学科, 并把数据科学定义为统计学领域扩展到以数据作为现金计算对象相结合的部分, 奠定了数据科学的理论基础。数据科学涉及数据获取, 数据存储、查询与管理, 数据分析与理解, 数据应用等方面。数据获取主要涉及物理、材料、电子等学科; 数据存储、查询与管理主要涉及计算机学科和诸多的 IT 企业; 数据分析与理解主要是数学、统计学、信息科学的问题; 而数据应用则涉及各行各业。

1.4.3 数据科学相关

数据科学作为一个全新的领域, 是社会科学和统计学、信息与计算机科学等领域的交叉学科, 它的发展同计算机、互联网及近年来大数据的不断发展密不可分。数据科学不但涉及传统静态数据的收集、整理、存储和分析, 更侧重开发处理动态数据的新概念和方法, 如数据获取、数据存储与管理、数据安全、数据分析、可视化等。还需要有基础理论和新技术, 如数据存在性、数据测度、时间、数据代数、数据相似性与簇论、数据分类与数据百科全书、数据伪装与识别、数据实验、数据感知等。数据科学的理论和方法将改进现有的科学研究方法, 形成新型的科学研究方法, 并且针对各个研究领域开发出专门的理论、技术和方法。

1.4.4 数据科学应用

随着大数据成为当前的热点, 信息技术发展的重点从计算转向数据, 数据的有效应用变得至关重要, 数据科学就是在这一背景下产生和发展起来的。数据科学通常指基于计算机科学、统计学、信息系统等学科的理论和技术, 研究数据的收集整理, 以及从海量数据中分析处理, 获得有效知识并加以应用的新兴学科。数据科学着重于数据, 虽然和运筹学与管理科学有交叉之处, 但更强调以数据为导向的分析与决策。数据科学主要应用在如下问题中。

(1) 处理大数据高维问题。数据科学研究维数 p 远远大于样本量 n , 如人类的 DNA 碱基对数量大约 30 多亿, 而城市的人口一般在几千万人左右, 针对研究某个遗传病时, 就算全部对其国民采样, 相对于几十亿的 DNA 碱基对数量, 所能获得的数据量是很小的, 决策要素和数据量可能都趋于无穷式地增长, 而决策要素相比于数据量增长成指数级, 通过对决策空间施加稀疏性约束来做决策。由于高维数据本身常常呈现低维特征, 利用低维特征将高维问题简化到低维值得研究。

(2) 处理重采样问题。数据科学对于数量巨大而又不能集中存储的数据集, 随机拆分是解决问题的可能途径, 拆分成若干小数据集, 通过对每一小数据集进行推断, 然后将基于小数据集上的推断聚合形成一个整体推断, 根据特定的目标设计最好的采样和模型, 尽可能优化证实目标。

(3) 解决分布式计算问题。分布式计算采用多处理器自主处理, 其本质是协同完成计算任务。数据科学研究优化问题时采用随机搜索方法, 可分解思路, 将问题分类以找到数据分布计算的模型。

(4) 完善信息融合问题。数据科学研究内容既包含数量、图形结构化数据，也包含文本、图像、视频、音频的非结构化数据，可正确地抽取和融合不同类型数据所蕴含的互补信息。信息融合无论是数据层次、特征层次还是决策层次都有一系列深刻的融合估计理论与方法问题。

(5) 处理可视分析问题。数据的可视是基本的工具与平台。根据数据本质特征设计可表达的几何空间，建立数据空间与可表达的几何空间的映射等是可视分析的最根本科学问题。

数据科学应用到每一领域都有其各自的侧重点，但始终围绕着一个数据核心流程，即收集、整理、挖掘和分析。数据科学家需具备统计学、决策科学、计算机科学等方面的知识，熟练掌握数据分析软件，具备强大的计算机编程能力和数据分析能力，做好数据与相应产品和服务的结合。

1.5 智能计算科学

计算科学和数据科学正在蓬勃发展，是未来竞争和发展的趋势。数据的巨大价值将被有效地挖掘，数据采集将更便捷、更廉价、规模更大。智能计算科学将融合定量分析，利用计算机来分析解决各领域研究的数据科学问题。智能计算自身的特点是借用自然界、生物界规律的启迪，根据其原理模仿设计求解问题的算法，包括遗传算法、进化算法、蚁群算法、人工智能、神经网络、机器学习、生物计算、量子计算、模式识别、知识发现、数据挖掘等。本小结在继 1.1.2 小结学习计算科学的相关概念后，从计算机的视角阐述高性能计算、人工智能、机器学习，同时在天文计算中利用观测数据中的科学规律结合高性能计算预测，自动发现天体物质运动或复杂现象的演化规律。

1.5.1 高性能计算

高性能计算 (High Performance Computing, HPC) 早期是指求解问题速度很快的一类计算，但随着计算历史过程演变衍生出很多新兴计算，如网格计算、效用计算、向量计算、并行与分布式计算、服务计算和云计算等。实现了更快的计算速度，更大的负载能力和更高的可靠性。高性能计算的实现方法有多种，如提高单一处理器的计算性能，集成处理器等，后面即将学习的云计算就是高性能计算的新发展，高性能并行计算技术是藏在云计算内的核心技术。

近年来，我国高性能计算机的发展世界瞩目，其中有代表性的国产高性能计算机包括：神威蓝光、天河 1 号、天河 2 号、联想 6800 和曙光星云等。近几年中国超算的国际地位明显提高，天河系列获得国际广泛的认可后，中国的高性能从业人员由“旁观者”转变为“贡献者”。高性能计算领域的国际合作越来越多，超算市场格局发生了巨大的变化。以前中国的超算市场基本以国外厂商为主，然而由于高性能计算整体技术水平的提升，使得中国厂商占据了主导地位，联想更是其中的典型代表。在 2015 年国际超级计算机大会上公布的最能体现厂商整体实力的《全球高性能计算 TOP 500 排行榜》中，联想以 23 套的份额跃居全球前五，中国第一。天河 1 号在第 36 届世界 500 强排行榜中排名第一位，天河 2 号已蝉联三次冠军。

1.5.2 人工智能

人工智能 (Artificial Intelligence, AI) 是研究人类智能活动的规律，使计算机构造出具有一定智能性的人工系统，让计算机去完成以往需要人的智力才能胜任的工作。人工智能是应

用计算机的软、硬件来模拟人类某些智能行为的基本理论、方法和技术。该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，其理论和技术日益成熟，应用领域也不断扩大，是人类智慧未来研究的必需工作。

专家系统是人工智能研究的一个应用领域，应用专家系统可以对输入的原始数据进行分析、推理，做出判断和决策，例如：智能模拟机器人、医疗诊断、语音识别、金融决策、人机对弈等。人工智能不是人的智能，但能像人一样思考，在处理某些问题上也可能超过人的智能。计算机模拟人的某些思维过程和智能行为（如学习、推理、思考、规划等）主要包括计算机实现智能的原理，制造类似于人脑智能的计算机，计算机实现更高层次的应用。人工智能将涉及计算机科学、心理学、哲学和语言学等学科。从思维观点看，人工智能不仅限于逻辑思维，还要考虑形象思维、灵感思维才能促进人工智能的突破性的发展，人工智能与思维科学的关系是实践和理论的关系，人工智能是处于思维科学的技术应用层次。

1.5.3 机器学习

学习是人类具有的一种重要智能行为，机器学习（Machine Learning, ML）是利用计算机模拟、实现人类的学习行为，以获取新的知识或技能，其能够重新组织已存储的知识结构，改善自身的性能。在机器学习过程中涉及一个非常接近 AI 的新领域，即深度学习（Deep Learning），其意在建立、模拟人脑进行分析学习的神经网络，深度学习模仿人脑的机制来解释数据（如图像，声音和文本）。

学习部分提供对应的信息及数据，利用这些信息修改知识库增进系统执行部分完成任务的效能，执行部分根据知识库完成任务，同时把获得的信息反馈给学习部分。影响学习系统设计最重要的因素是信息质量。学习系统需要在获得足够多的数据之后，删除不必要的细节，进行总结推广，形成指导动作的一般原则，并放入知识库。学习部分的任务比较繁重，设计起来相对困难。知识库是影响学习系统设计的次要因素。知识的表示有多种形式，可选择的表示方式包括表达能力强、易于推理、容易修改的知识库、知识表示易于扩展等。在处理一个问题时，学习系统不能在全然没有任何知识储备的情况下凭空获取知识，每个学习系统对要求具有某些知识理解的环境提供信息、分析、比较、假设、检验并修改对应假设，因此学习系统是对现有知识的扩展和改进。

1.5.4 天文计算

伴随天文大数据信息的飞速扩容，宇宙恒星、星系及类星体等信息数据量急速增加，天文计算（Astronomy Computing）运用了计算科学的方法研究宇宙中天体和天体系统的结构、形成、演化及活动，是天文学与计算机科学交叉结合的应用。在大型天文巡天项目的科学中，我国郭守敬望远镜（Large Sky Area Multi-object Fiber Spectroscopic Telescope, LAMOST）的科学目标集中在河外星系的观测、银河系结构和演化及多波段目标认证（<http://www.lamost.org/>），包括探索暗物质和暗能量、观测太阳系内的小行星（尤其是对地球具有潜在威胁的小行星）、侦测瞬态天文现象及银河系观测等，获取科学使命的数据。科学家们通过对千万数量级的星系、类星体等河外天体及恒星的光谱进行观测，分析出宇宙学模型、宇宙大尺度结构、星系形成和演化、银河系结构与演化等，在研究上做出了重大贡献。图 1-9 为 LAMOST 观测站。在本章 1.8 节拓展实训中将学习数据观测。



图 1-9 LAMOST 观测站

1.6 大数据时代的云计算

PC (Personal Computer) 是人们日常工作生活中常用的工具, 使用它可以完成文档处理, 数据存储, 也可通过电子邮件或 U 盘、移动硬盘等进行数据分享。但若 PC 异常, 可能面临资料丢失, 工作延迟等不利情况。数据已经渗透到当今每一个行业和业务职能领域, 对于海量数据的挖掘和运用成为重要的生产因素。本节要学习的物联网、移动互联网等是大数据的来源, 而大数据分析结果也将反馈于物联网和移动互联网, 为其发展提供更具价值的支撑。

1.6.1 大数据

大数据 (Big Data) 是指为决策问题提供服务的大数据集、大数据技术和大数据应用的总称。大数据集是指一个决策问题所用到的所有可能的数据, 通常数据量巨大、来源多样、类型多样; 大数据技术是指大数据资源获取、存储管理、挖掘分析、可视展现等技术; 大数据应用是指用大数据集和大数据技术来支持决策活动, 是新的决策方法。1996 年 SGI 公司首席科学家 John Mashey 在其题为《Big Data and the Next Wave of InfraStress》的演讲中首次提出了 Big Data 这一概念。2001 年 Doug Laney 最先提出大数据的“3V”模型, 包括容量 (Volume)、速度 (Velocity) 和种类 (Variety)。随后信息业界把 3V 扩展到了 11V, 增加了有效性 (Validity)、真实性 (Veracity)、价值性 (Value)、可行性 (Viable) 及可见性 (Visible) 等。

大数据的数据量已达到 PB 级甚至 EB 级规模。“大”是大数据最直观、最重要的特征, 但远非全部。大数据已经发展为一个较为宽泛的概念, 其包括数据及其采集、处理、分析、解释等在内的一系列相关技术、方法、手段的统称。大数据是人们获得新的认知、创造新价值的源泉。如今的研究表明大数据不仅指数据本身, 也是一种处理数据的工具。如前面章节描述的天文计算, 在大数据分析技术应用中, 将从大数据的视角重新审视宇宙, 获取宇宙领域的海量数据, 并记录、分析、比较、归一化处理历史数据和不断推理产生新数据, 发展新的方法, 充分挖掘其中蕴藏的科学价值。再如社交网络上个人发布的日志、图片、音频、视频, 电商网站记录的交易信息、商品和用户的购物行为, 路网监控拍摄的视频, 设备传感器采集的数据等都是大数据。

1.6.2 云计算

云计算 (Cloud Computing) 是由 Google 提出的网络应用模式, 其采用分布式计算技术,

提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的，并且可以随时获取，按需使用。

狭义云计算是指 IT 基础设施的交付和使用模式，通过网络以按需、易扩展的方式获得所需的资源（包含硬件、平台、软件）。

广义云计算是指服务的交付和使用模式，指通过网络以按需、易扩展的方式获得所需的服务。这种服务可以是 IT 和软件、互联网相关的，也可以是任意其他的服务。

云计算的本质是通过网络将庞大的计算处理程序自动分拆成无数个较小的子程序，移交由多部服务器组成的庞大系统，完成搜寻、计算分析后，将处理结果回传给用户，整个过程将在数秒之内处理数以千万计乃至亿计的信息。未来云计算还能更好地解析癌症细胞、分析 DNA 结构、进行基因图谱定序等。基于多种技术混合演进的云计算，其具备的特点如下。

(1) 超大规模。云计算的先行公司包括 Amazon、Google、IBM、Microsoft 和 Yahoo，还有其他众多成功的公司，包括 Everydo、Salesforce、Facebook、Youtube 和 Myspace 等。云计算管理系统具有相当大的规模，其拥有数百万台服务器。

(2) 虚拟化。云计算支持用户在任意位置，使用各种终端获取应用服务，所请求的资源来自“云”服务器，应用在“云”中运行，用户无须了解也不用担心应用运行的具体位置。

(3) 高可靠性。云计算提供了最可靠、最安全的数据存储中心，用户可排除数据丢失、病毒入侵等麻烦。其使用了数据多副本容错、计算节点同构可互换等措施来保障服务，具有比本地计算机更高的可靠性。

(4) 通用性强。云计算对用户端的设备要求较低，使用方便，可以支撑不同应用运行。

(5) 高可扩展性。云计算可实现不同设备间的数据与应用共享，进行动态伸缩，满足应用和用户规模增长的需要。

(6) 廉价按需服务。云计算的特殊容错措施可以采用极其廉价的节点来构成云，具有高度的自动化，集中式管理，使大量企业无须担负高昂的数据中心管理成本。

云计算提供了无限的廉价存储和计算能力，包括：软件服务（SAAS）、实用计算、网络服务、平台服务、管理服务（MSP）、商业服务和互联网整合。我国云计算发展十分迅猛，2015 年初，国务院下发《关于促进云计算创新发展培育信息产业新业态的意见》，明确指出：着力突破大数据挖掘分析等关键技术，推动大数据挖掘、分析、应用和服务，开展基于云计算的大数据应用示范，支持政府机构和企业创新大数据服务模式。大数据中隐藏着巨大的经济、科学、社会及军事价值，有关大数据的讨论逐渐成为产业界、学术界、政府部门等各界共同关注的热点。大数据正渗透到社会生活的各个方面，将导致人们的工作和生活方式发生巨大变革。

云计算潜在的危险性是所提供的服务垄断在私人机构实现商业信用，对于政府机构、商业机构（如银行、证券等掌控敏感数据），在选择云计算服务时应保持足够的警惕，特别是在选择使用外国机构提供的云计算服务时。

1.6.3 物联网

物联网（Internet of Things, IOT）是指通过射频识别、红外感应器、全球定位系统、激光扫描器等信息传感设备装置获取物体上的各种信息，赋予物体智能，并通过接口与互联网相连而形成了物品与物品相连的巨大分布式协同网络。物联网的 3 大特征是：普遍对象设备化、自治终端互连化和普适服务智能化。

物联网所采集和捕获的各种数据组成了大数据，同时大数据技术使物联网采集到的数据产生了价值。在实际研究中，任何事物和该事物在人们头脑中形成的概念相互依存，该事物的概念是该事物在人们头脑中的本质属性的反映。物联网也是一种事物，其与互联网概念的关系也是相互依存的关系。离开了互联网，物联网概念就是无本之木，无源之水。互联网必然在人们头脑中形成与之相对应的物联网概念。物联网概念是在互联网概念的基础上，其用户端延伸和扩展到物品与物品之间，进行信息交换和通信的一种网络概念。物联网应用的领域很多，如智能家居、智能交通、智能物流、智能医疗、智能电网、智能农业、智能安防及商业智能等。

1.6.4 智慧未来

未来互联网（Future Internet）具有更多的用户、更多的信息、更复杂的结构及交错的互动参与特性。大数据时代，互联网实现用户信息无区域限制访问。未来的互联网将逐步由物联网、内容与知识网、服务互联网和社会网络等构成。服务互联网是指将全球各地提供的不同服务互连，归所有用户使用，即一切皆服务（Everything as a Service, EaaS）。社会网络（Social Network）是指社会个体成员之间由于互动（行为）而形成的相对稳定的关系体系。如在“治霾”方面，污染严重的制造业和传统检测设备的商家会找到互联网企业，洽谈整体解决方案及产品合作方案。未来“互联网+”将促成过去商家做不到或者不敢做的事情，使其融合成为智慧的联盟。

《互联网进化论》一书中提出“互联网的未來功能和结构将与人类大脑高度相似，也将具备互联网虚拟感觉，虚拟运动，虚拟中枢，虚拟记忆神经系统”，互联网虚拟大脑结构图包含的物联网、传统互联网、移动互联网在源源不断地向互联网大数据层汇聚数据并纳入云计算，物联网对应互联网的感觉和运动神经系统；大数据代表互联网的信息层，是互联网智慧和意识产生的基础；云计算是互联网中枢神经系统的萌芽，构成互联网核心硬件层和核心软件层的集合。

1.7 自主实践

计算机相关实验很普遍，正确使用计算机可以延长计算机的寿命，同时促进操作者计算机使用技能与素养的提升。

1.7.1 正确使用计算机

一、预习内容

1. 学习计算机机房上机规则；
2. 了解启动和关闭计算机的常用方法。

二、实践目的

1. 掌握计算机机房上机规则；
2. 正确使用主机箱面板上的启动和重启按钮。