第一篇
算力网络与大数据

第 1 章 算力概述

从原始社会的手动式计算到古代的机械式计算、近现代的电子计算，再到现在的数字计算，算力指代了人类对数据的处理能力，也集中代表了人类智慧的发展水平。

1998 年图灵奖得主——著名的计算机科学家吉姆·格雷（Jim Gray）将科学研究的范式分为四类——除了之前的物理实验、模型推演和计算仿真三大范式，新的信息技术已经促使新的第四范式出现，即数据密集型科学发现（Data-Intensive Scientific Discovery）。第四范式的基本思想是把数据看成现实世界的事物、现象和行为在数字空间的映射，认为数据自然蕴含了现实世界的运行规律；进而以数据作为媒介，利用数据驱动及数据分析方法揭示物理世界现象所蕴含的科学规律。推动其发展的三大要素是数据、算法和算力，将数据作为原料，基于算力基础研究能够从其中挖掘智慧的算法。

在此趋势下，数据能力和算力需求呈现循环增强的状态，数据量的不断增加要求算力的配套进化。据 IDC 预测，到 2025 年，全球数据总量预计达到 180ZB。1ZB 相当于 1.1 万亿吉字节（GB），如果把 180ZB 全部存在 DVD 光盘中，这些光盘叠起来大概可以绕地球 222 圈。而与此同时，集成电路的发展仍然遵循着“摩尔定律”——1965 年，英特尔创始人之一戈登·摩尔提出，集成电路上可容纳的元器件数目，每隔 18 ～ 24 个月便会增加一倍，性能也会提升一倍。而集成电路会直接影响到中央处理器（CPU）的性能，进而影响到计算机的计算能力。换言之，算力始终处于一个稳步上升的状态，而且成本会越来越低。

同时，算力的发展强力推动数字经济持续向前。在 2016—2020 年，全球算力规模平均每年增长 30%，数字经济规模和 GDP 每年分别增长 5% 和 2.4%。2020 年，我国产业数字化规模达 31.7 万亿元，占 GDP 比重为 31.2%，同比增长 10.3%。未来，算力与数字经济共促共进的关系将进一步强化，算力不断发

展推动数字经济持续向前，数字经济持续向前加重对算力的支撑依赖，“算力时代”正在到来。

1.1　算力是什么

1.1.1　计算工具的发展历史

人类从茹毛饮血的原始社会到万物互联的数字智能社会，伴随人类智慧进步的需求，计算工具经历了从简单到复杂、从低级到高级、从手动到自动、从自动到电子的发展过程。

1．手动式计算

人类学会用工具辅助自己计算和记录，从最早的结绳记事到算筹、算盘的使用。早在公元前 3000 年，古埃及人用结绳来记录土地面积和谷物收获的情况。这种传统的结绳记事方式至今仍被一些没有文字的遗存文明所使用。而早期的计算器为纯手动方式，如算筹、算盘等。算筹通过短棍摆放的位置表示数字。算盘是中国祖先创造发明的一种简便的计算工具，通常由可滑动的珠子制成，如图 1-1 所示。

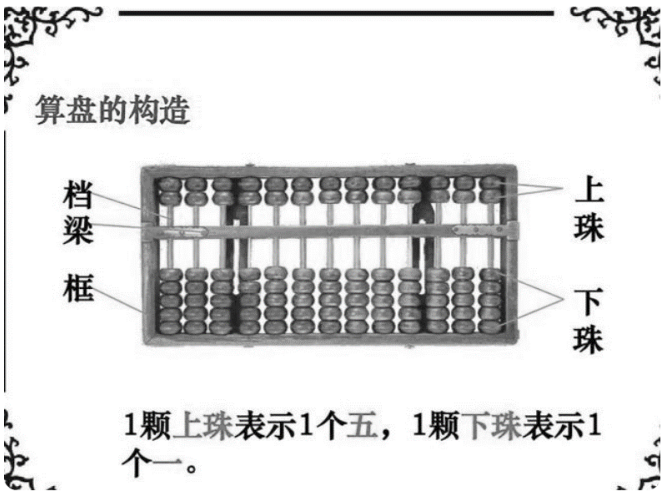

图 1-1　算盘的构造

2．机械式计算

17 世纪，欧洲出现了利用齿轮技术的计算工具。

- 1642年，法国数学家帕斯卡利用齿轮技术，发明人类第一台机械式计算工具“帕斯卡加法器”。
- 1673年，德国数据家莱布尼兹发明“四则运算器”。
- 1804年，法国机械师约瑟夫·雅各发明可编程织布机，创造“穿孔卡片”这一输入方式。
- 1832年，英国数学家巴贝奇发明差分机，采用有存储、运算、控制三种具有现代意义的装置，人类从手动机械时代进入自动机械时代。

3．机电式计算

19 世纪，采用机电技术代替机械装置。

- 1886年，美国统计学家赫尔曼·霍勒瑞斯借鉴穿孔卡片原理发明制表机，可自动进行四则运算、累计存档、制作报表。
- 1938年起，德国工程师朱斯采用继电器，研制出人类二进制计算机Z-1、Z-2、Z-3、Z-4。
- 1944年，美国哈佛大学应用数学教授霍华德·艾肯使用大量继电器，发明机电式计算机Mark I。

4．电子计算机

20 世纪 30 年代已经具备了制造电子计算机的技术能力。

1939 年，美国爱荷华州立大学数学物理学教授约翰·阿塔纳索夫（John Atanasoff）及其研究生克利福德·贝瑞（Clifford Berry）用电子技术来提高计算机的运算速度，研制出 ABC（Atanasoff-Berry Computer），为模拟计算向数字计算的跨越奠基。

二战期间，“计算机之父”阿兰·图灵发明“图灵机”模型，为数字计算机的诞生奠定理论基石。

1946 年，冯·诺依曼基于图灵机，发明第一台通用计算机、第一代电子管计算机 ENIAC，如图 1-2 所示，确立了数字计算机整体架构，人类电子计算机进入数字时代。

继 ENIAC 之后，人类先后经历第二代晶体管计算机、第三代集成电路计算机、第四代大规模集成电路计算机三个阶段。

图 1-2　世界上第一台通用计算机 ENIAC

可以说，正因为人类对算力的不懈追求，我们才从原始的结绳记数、算盘，走进个人电脑、手机、人工智能，进入快捷、便利、多姿多彩的数字生活。放眼未来，对算力极限的不断挑战，还将推动着我们进一步靠近重塑未来生活的历史性机遇。

1.1.2　算力的内涵和外延

从狭义上看，算力是设备通过处理数据，实现特定结果输出的计算能力。2018 年，诺贝尔经济学奖获得者威廉 • 诺德豪斯（William D. Nordhaus）在《计算过程》一文中提出，“算力是设备根据内部状态的改变，每秒可处理的信息数据量”。算力实现的核心是 CPU、GPU、FPGA、ASIC 等各类计算芯片，并由计算机、服务器、高性能计算集群和各类智能终端等承载，海量数据处理和各种数字化应用都离不开算力的加工和计算。

简单来说，就是电子计算机处理器计算和处理任务的计算能力，是衡量一台电子计算机、数据中心或云计算中心在一定的网络消耗下处理的数据容量与所花时间之比的单位总计算能力。

从科学的角度来说，算力是指计算机系统的理论峰值速度，其计算公式如下：

理论峰值速度 = MHz × 每个时钟周期执行浮点运算的次数（Flops）× CPU 数目

其中 MHz 是指 CPU 的主频，每个时钟周期执行浮点运算的次数是由处理器中浮点运算单元的个数，以及每个浮点运算单元在每个时钟周期能处理几条浮点运算来决定的。算力数值越大代表综合计算能力越强。

从广义上看，算力也可以包括具备存储能力的各类独立存储或分布式存储，以及通过操作系统逻辑化的各种数据处理能力。从更宏观的层面上看，算力是数字经济时代的新生产力，是推动数字经济进一步发展的坚实基础。数字经济时代的关键资源是数据、算力和算法，其中数据是新生产资料，算力是新生产力，算法是新生产关系，这三者构成数字经济时代最基本的生产基石。现阶段，5G、云计算、大数据、物联网、人工智能等技术的高速发展，推动数据的爆炸式增长和算法的复杂程度不断提高，带来了对算力规模、算力能力等需求的快速提升，算力的进步又反向支撑了应用的创新，从而实现了技术的升级换代、应用的创新发展、产业规模的不断壮大和经济社会的持续进步。随着 5G 商用步伐的加快，物与物之间的连接不断深化，算力在自动驾驶、智慧安防、智慧城市等领域的应用不断扩大，边缘计算以及雾计算的需求日益增加，算力的范畴和边界仍在不断扩展。

1.2　无处不在的现代算力

在万物互联的时代背景下，算力以各种各样的设备形态出现在人们的日常生活和工作中，作为承载算力的底层核心芯片，其处理能力也达到了空前的高度。基于不同行业的不同特点，CPU 通用算力逐渐从单一的 x86 架构向 ARM、RISC-V 等多种架构扩展，芯片性能也在不断向更强、更高效的方向演进；GPU、FPGA 等异构计算突破了通用算力的性能瓶颈，技术发展聚焦云游戏、AI 等场景，为人们提供更为极致的算力服务。算力的发展已经呈现出多架构共存、多技术协同、多领域协同的局面。

1.2.1　算力的分类

算力有不同的分类方法。按照使用主体，可以分为个人算力和企业算力；按照算力类型，可以分为基础算力、智能算力和超算算力，如图 1-3 所示，

分别提供基础通用计算、人工智能计算和科学工程计算。其中，基础通用算力主要基于 CPU 芯片的服务器所提供的计算能力，智能算力主要基于 GPU、FPGA、ASIC 等芯片的加速计算平台提供人工智能训练和推理的计算能力，超算算力主要基于超级计算机等高性能计算集群所提供的计算能力。

图 1-3　算力分类

1．基础算力

由基于 CPU 芯片的服务器所提供的算力，主要用于基础通用计算。日常提到的云计算、边缘计算等都属于基础算力，它为移动计算、物联网等提供计算支持。基础算力占整体算力的比重由 2016 年的 95% 下降至 2020 年的 57%，但其依旧是算力主力。

2．智能算力

由基于 GPU、FPGA、ASIC 等 AI 芯片的加速计算平台提供的算力，主要用于人工智能的训练和推理计算，比如语音、图像和视频的处理。在技术架构上，人工智能的核心计算能力由训练、推理等专用计算芯片提供，注重单精度、半精度等多样化计算能力。在应用方面，人工智能计算中心主要支持人工智能与传统行业的融合创新与应用，提升传统行业的生产效率，在自动驾驶、辅助诊断、智能制造等方面大显身手。近年来，智能算力规模增长迅速，在整体算力中的占比已由 2016 年的 3% 提升至 2020 年的 41%，预计到 2023 年，智能算力的占比将提升至 70%。

3．超算算力

由超级计算机等高性能计算集群所提供的算力，主要用于尖端科学领域的计算，比如行星模拟、药物分子设计、基因分析等。在技术架构上，超算算力的核心计算能力由高性能 CPU 或协处理器提供，注重双精度通用计算能力，追求精确的数值计算。在应用方面，超算中心主要应用于重大工程或科学计算领域的通用和大规模科学计算，如新材料、新能源、新药设计，高端装备制造，航空航天飞行器设计等领域的研究。超算算力在整体算力中的占比较为稳定，约为 2%。

1.2.2　典型算力平台

算力泛在化促使智能终端呈现多元化发展态势。随着 5G 网络、边缘计算的规模建设，新兴应用将加速驱动数据处理由云端向边侧、端侧扩散，边端计算能力持续增长，算力泛在化已成趋势，带动各种计算设备的巨大需求。未来随着边端设备种类的丰富，个人 PC 甚至家庭网关都将可能作为算力的节点，手机、智能汽车等智能终端的普及形成了数据就近处理和泛在计算处理的场景，由此也将促进用户周边信息化空间内，不同距离、不同规模的算力相互协同和联动，呈现“云—边—端”三级计算架构。目前，5G 泛终端已达 20 类，涵盖 VR/AR 头显、CPE、工业级路由器 / 网关、无人机、机器人、车辆 OBU 等众多品类，将在工业、医疗等非成本敏感领域普及并迭代演进，并对文化教育、休闲娱乐方式等产生颠覆性变革。

1．云计算

云计算就是把一个个服务器或者计算机连接起来构成一个庞大的资源池，以获得超级计算机的性能，同时又保证了较低的成本，如图 1-4 所示。云计算是分布式计算技术的一种，它的原理是通过网络“云”，将所运行的巨大的数据计算处理程序分解成无数个小程序，再交由计算资源共享池进行搜寻、计算及分析后，将处理结果回传给用户。云连接着网络的另一端，为用户提供了可以按需获取的弹性资源和架构。用户按需付费，从云上获得需要的计算资源，包括存储、数据库、服务器、应用软件及网络等，大大降低了使用成本。

云计算的本质是从资源到架构的全面弹性，这种具有创新性和灵活性的资源降低了运营成本，更加契合变化的业务需求。

图 1-4　云计算

2．边缘计算

边缘计算是一种计算资源与用户接近、计算过程与用户协同、整体计算性能高于用户本地计算和云计算的计算模式，是实现无处不在的“泛在算力”的重要手段。其中，边缘设备可以是任意形式，其计算能力通常高于前端设备，且前端设备与边缘设备之间应当具有相对稳定、低延迟的网络连接。

与云计算数据中心相比，边缘计算中直接为用户提供服务的计算实体（如移动通信基站、WLAN 网络、家用网关等）距离用户很近，通常只有一跳的距离，即直接相连。这些与用户直接相连的计算服务设备称为网络的“边缘设备”。

边缘计算模型将原有云计算中心的部分或全部计算任务迁移到数据源附近，相比于传统的云计算模型，边缘计算模型具有实时数据处理和分析、安全性高、隐私保护、可扩展性强、位置感知以及低流量的优势。

3．终端计算

终端设备大部分时间都在扮演数据消费者的角色，比如使用智能手机观看视频、浏览新闻等。然而，随着 5G 泛终端的普及，终端设备也有了生产数据和计算数据的能力，且计算能力日益增强。终端计算将终端自身的算力进行整合和管理，不仅能够实现真正的算力泛在，同时还能够为用户提供更高质量的计算服务。

1.3　算力，新时代的生产力

算力从根本上改造、升级了生产力的三要素，最终驱动着人类社会的转型升级。我们谈到的智慧农业、智能工厂、无人配送等智能场景，背后都是计算的力量，它将劳动者由传统的人变成了“人 +AI”，“新劳动者”可以超越生物极限，呈现指数级增长。数据作为计算的处理对象，成为一种新的劳动对象、新的生产资料，从有形到无形，生生不息，越用越多。劳动资料也由传统的机械升级为计算力驱动的信息化设备，劳动效率同样是指数级增长，生产力得到了前所未有的解放。

1.3.1　算力成为数据经济新引擎

新一轮科技革命和产业变革正在重塑全球经济结构，算力作为数字经济时代新的生产力，是支撑数字经济发展的坚实基础，对推动科技进步、促进行业数字化转型以及支撑经济社会发展发挥重要的作用。

“算力已成为我国当前最具活力、最具创新力、辐射最广泛的信息基础设施，算力成为衡量数字经济活力的关键指标。”中国信息通信研究院《中国算力发展指数白皮书》显示，2016—2020 年，我国算力规模平均每年增长 42%，数字经济规模增长 16%，GDP 增长 8%，算力对数字经济和 GDP 增长的拉动作用显著。2020 年，我国以计算机为代表的算力产业规模达到 2 万亿元，直接带动经济总产出 1.7 万亿元，间接带动经济总产出 6.3 万亿元，即在算力中每投入 1 元，将带动 3 ～ 4 元的经济产出。

一方面，算力直接带动数字产业化的发展。电子信息制造业、电信业、软件和数字技术服务业、互联网行业等数字核心产业的发展与算力的发展息息相关。互联网行业是最大的算力投资行业和算力需求行业。国外，亚马逊、微软和谷歌通常每个季度投入的资本支出总额超过 250 亿美元，其中大部分用于布局超大规模数据中心；国内，互联网行业算力占整体算力近 50% 的份额，以阿里巴巴、腾讯、百度、字节跳动为代表的互联网巨头对算力的需求更加迫切，同时算力的集中部署也使互联网行业成为先进生产力的代表。电信业、软件和数字技术服务业数据化起步较早，也是我国算力应用较大的行业，对算力的应用处于行业领先水平。

另一方面，算力助推产业数字化的转型升级。在算力基础设施的支撑下，电子商务、平台经济、共享经济等数字化新模式接替涌现，工业互联网、智能制造等全面加速，为我国产业数字化的持续健康发展输出强劲动力。根据中国信息通信研究院《中国算力发展指数白皮书》统计，2020 年，我国产业数字化规模达 31.7 万亿元，占 GDP 比重为 31.2%，同比名义增长 10.3%，占数字经济比重由 2015 年的 74.3% 提升至 2020 年的 80.9%。算力投入带来的数字化智能技术不仅为制造、交通、零售等多个行业带来产业产值增长，还带来了生产效率提升、商业模式创新、用户体验优化等延伸性效益，对经济增长的拉动作用愈加凸显。以制造业领域为例，以云计算、边缘计算、智能计算为代表的算力投入和规模应用可以显著提升生产效率，从需求洞察、研发、采购、生产、营销和售后等产业链环节对制造业进行赋能和重构，打造高度协同的智能制造生态体系。

1.3.2　信息消费持续释放算力需求

算力的重要性不言而喻，这与我们观察到的社会对算力的需求呈现爆发式增长也是相吻合的。信息消费已成为创新最活跃、增长最迅猛、辐射最广泛的经济领域之一。信息消费包括直接信息消费和间接信息消费。既包括传统产品消费到新型信息产品消费的演变，也包括基于语音、互联网接入与服务、信息内容与应用、软件应用等信息服务消费，已经渗透到衣、食、住、行、娱各个层面。

信息消费的形态正在发生深刻变化。信息产品从手机、电脑向数字家

庭、虚拟现实 / 增强现实、智能网联汽车等产品延伸，可穿戴设备、消费级无人机等新产品不断涌现。信息服务从通信需求转向应用服务和数字内容消费。

智能终端消费和移动数据流量消费规模继续扩大。当前我国移动电话普及率稳步提升，4G 用户渗透率超八成，用户数及用户占比持续扩大，截至 2020 年年底，我国 5G 网络用户数超过 1.6 亿，约占全球 5G 总用户数的 89%。我国 5G 手机出货量和占比不断提升，2020 年，国内 5G 手机总计出货量为 1.63 亿部，占同期手机出货量的 52.9%。与此同时，移动数据流量消费规模继续扩大，2020 年，移动互联网月户均接入流量（DOU）跨上 10 GB 区间，接入流量消费达 1656 亿 GB，比上年增长 35.7%，其中手机上网流量达到 1568 亿 GB，比上年增长 29.6%，在总流量中占 94.7%。

算力泛在化促使智能终端呈现多元化发展态势，算力的需求同样从终端爆发。新兴应用加速驱动数据处理由云端向边侧、端侧扩散，边端计算能力持续增长，带动各种计算设备的巨大需求。边缘智能时代，任何一个智能设备都可能是一个“数据中心”，很可能在一个芯片上实现连接、传感、存储和计算的功能。以无人驾驶为例，因为车内部署了大量传感器，且需要实时计算这些复杂的异构数据，并给出可靠的决策依据，有人测算，自动泊车功能就需要 100 TOPS 性能；完全自动驾驶功能必须有 500 TOPS 以上的算力支持。作为对比，驾驶员监测功能只需要 10 TOPS 的算力即可处理。

一个动力充沛、多样性的算力时代正在来临。未来，这个新的算力时代不但将催生一个万亿级产业，更将通过新基建，源源不断地为中国经济注入澎湃的动力。

1.3.3　算力助力数字世界和物理世界无缝融合

未来，数字世界和物理世界将无缝融合，如图 1-5 所示，算力正成为社会发展的基石。预计到 2030 年，全球数据年新增 1YB，通用算力增长 10 倍达到 3.3 ZFlops，人工智能算力增长 500 倍（超过 100 ZFlops）。全球连接总数将达到 2000 亿，传感器的数量达到百万亿级，大量感知计算将在边缘完成，处理大约 80% 的数据。

多维度建立从物理世界到数字世界的映射

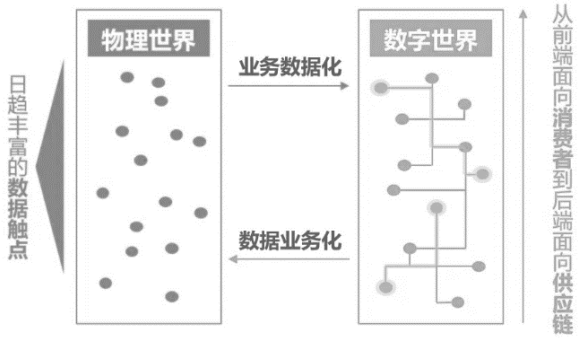

图 1-5　物理世界和数字世界融合

在智慧交通领域，每辆汽车平均行驶 2 小时，行驶中每秒上传的压缩数据将从现在的 10KB 升至 1MB，10 万辆智能网联汽车每天需要传输的数据量大约为 720TB。L4+ 自动驾驶汽车单车算力将达到 5000TOPS，单个车厂的云端至少需要 10 EFlops 以上的算力。

智慧城市领域，智慧城市中的物联网传感器持续生成城市运行的环境数据，每一个物理实体都将有一个数字孪生，如城市楼宇、水资源、基础设施等将组成城市数字孪生，实现更加智能的城市管理。城市智慧治理将带来 100 倍的社会数据聚集，实现高效城市治理。

企业数智化领域，超过 30% 的企业在数字世界中运营与创新，各种 AR/VR 用户数达到 10 亿。工业制造等传统企业将在更加复杂的产业链上下游环境中实现由软件定义的高效运营，每万名制造业员工将与 390 个机器人共同工作，机器准确理解人的指令，准确感知环境，做出决策建议与行动。