

当前，在新冠肺炎疫情大流行的背景下，社会加速进入以数字化、网络化、智能化为特征的信息社会，经济社会发展范式正在经历变迁。在经济结构范式方面，经济已经从高速增长转向高质量发展，人类对数字产业化、产业数字化能够带来的经济倍增、叠加、放大效应给予厚望，数字经济也因此成为当下全球经济发展的主旋律；在技术应用范式方面，信息技术已由单点突破进入了协同推进、群体性演进的爆发期，对于经济发展的作用从原来的基础动力向核心引擎转变；在商业竞争范式方面，科技创新成为构筑企业竞争优势的关键支撑；在大众消费范式方面，新生代群体逐步成为数字化消费主流人群，传统的规模经营方式向基于规模的价值经营范式演变，信息通信服务逐渐从满足需求向引领群众需求、创造需求发展。

伴随着经济范式的革新，信息基础设施被视为推动经济高质量发展的重要引擎，国家的核心竞争力逐渐发展为算力能力和水平的竞争。在中国，国家高层高度重视新时代基础设施的建设，在 2018 年年底召开的中央经济工作会议上明确 5G 基建、特高压、城际高速铁路和城际轨道交通、充电桩、大数据中心、人工智能、工业互联网作为“新型基础设施建设”，并在随后的国家战略指导性文件、国家政策中均给予了大力支持。

基础设施资源的建设和供给结构经历了不同的发展阶段。在早期，财力雄厚的大型企业倾向于自建自有的数据中心、算力资源以及网络基础设施。但是，以云计算为代表的算力供给模式，以其灵活的资源获取和使用方式、低廉的基础设施使用成本等优势，成为了当下非常受欢迎的一种算力供给模式。随着业务类型的丰富和多样化，雾计算、算力网络等方案也被提出。近年来提出的“算力网络”更是引发了业界的广泛关注。ITU-T、CCSA 等国内外标准化组织也已经开始进行深入的研究和标准化工作。

算力网络是随着 5G 网络、云计算、边缘计算等技术发展而出现的一种新型的资源整合方案，它将属于不同所有方的计算、存储等资源通过网络整合起来，按照用户业务的不同需求提供不同的资源组合。算力网络是在技术契机和商业背景的双重驱动下，得到了产业界的诸多关注。

1.1 算力网络概念

1.1.1 算力网络的定义

算力网络（本书中也简称“算网”）定义的标准化研究当前主要集中在 ITU-T SG13。2019 年 10 月，在 ITU-T SG13 全会上，算力网络的概念被正式提出，相关的研究正式拉开序幕。在该会议上，中国电信和中国联通主导的“算力网络—需求与架构”立项（Y2501 Computing Power Network – framework and architecture）获得全会通过，成为算力感知网络首个国际标准项目，目前仍然在更新中。2021 年 7 月中国电信、中国联通成功立项 Y.NGNe-O-CPN-reqts 标准和 Y.ASA-CPN 标准，分别研究算力网络的 NGNe 编排增强要求及框架、算力网络认证调度架构。2021 年 11 月中国移动进一步提出算力网络需求与场景（CFN），在同期的 ITU-TSG13 全会上，ITU-T 组织者对先后出现的 3 类术语进行了研究范围界定，并确认了算力网络的共同研究目标是 CNC（Computing and Network Convergence），未来考虑以 CNC 作为统一术语研究。由于算力网络标准工作的积极开展，该领域得到了 ITU 的重点关注，同时 ITU-T 与中国通信标准化协会（CCSA）算力网络系列标准相互呼应。首个算力网络国际标准的通过及国际系列标准的形成，是算力网络从国内走向国际的重要一步，在算力网络发展中具有里程碑式的意义。究竟什么是算力网络？目前业内其实并没有给出一个完整、规范的定义，可以从标准组织的研究中寻找一些启发。从 2019—2021 年，ITU 中涉及算力网络的概念和定义有：

- CPN（Computing Power Network）：算力网络是一种新型的网络结构，通过一个集中式的或者分布式的网络控制平面，获取服务节点相关的计算、存储、网络等资源信息，并最终实现资源的最优化分配。
- CAN（Computing Aware Networking）：面向云计算环境，CAN 提供

了一种云算力和网络资源的集成优化方案，使得CSP能够提升云服务能力。

- **CFN (Computing Force Network)**：在IMT-2020以及其演进技术背景下，CFN倡导在全算力感知和集中管理的基础上，引入AI/ML相关技术和能力，实现算力、网络资源的联合优化。

尽管不同的概念侧重点不一样，但是，三者的目标还是一致的，即从全局的视角下，实现算力资源和网络资源的统筹调度，达到联合资源效用的最优。因此，为了概念和后续研究工作的统一，ITU在2021年的时候，将算力网络的定义统一为：

- **Computing and Network Convergence (CNC)**：在网络、算力资源统一感知和控制的基础上，实现算力和网络资源的联合优化。

在国内，产业界依托CCSA，也给出了算力网络的概念：

- 面向算网融合演进的新型网络架构，通过算力资源与网络资源状态的协同调度，将不同应用的业务通过最优路径，调度到最优的计算节点，实现用户体验最优的同时，保证运营商网络资源和计算资源利用率最优化。

除此之外，IETF、ETSI、6GANA等组织中，也或多或少地涉及算力网络的概念。

1.1.2 算力网络与电力网络

算力网络的理念提出以来，业界一直把算力网络类比为电力网络。电力时代构建了一张“电网”，有电就可以用电话、洗衣机、电饭煲、电风扇、电视机，终端的用户无须关心所使用的电来自哪里，只需要保证家里的插头标准、电力电压标准满足要求即可。在算力时代，算力网络将发挥同样重要的作用，为数字经济蓬勃发展提供必不可少的支撑，自动驾驶、人脸识别、游戏渲染等业务申请需要的算力，而无须关注来自哪个物理位置。

2002年3月，国务院正式批准了以“厂网分开，竞价上网，打破垄断，引入竞争”为宗旨的《电力体制改革方案》（国务院5号文件），国家电力公司分拆成两家电网公司、五家发电集团公司和四家辅业集团公司。电力的生产和供给由五家发电集团公司负责；电力的输送和对外供给由电网公司负责。算力网络时代，运营商的网络能够感知算力的需求和供给，是实现供需高效对接的

有利武器。类比于电力，算力网络要实现算力资源的开放性纳入，构建公平、开放、透明的算力供给对接市场机制。

1.2 新型数字化经济需求

1.2.1 人工智能产业化需求

人工智能是数字经济高质量发展的引擎，也是新一轮科技革命和产业变革的重要驱动力量。作为引领这一轮科技革命和产业变革的战略性技术，加快发展新一代人工智能是赢得全球科技竞争主动权的重要战略抓手。《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》也做出了相关部署。

随着人工智能技术的发展和在各行各业的落地，数据呈爆炸式增长趋势，模型也愈加复杂和庞大，算力成为人工智能发展的重要制约因素。据 OpenAI 公布的 AI 算力报告显示（如图 1-1 所示），自 2012 年以来最先进的 AI 模型的算力需求每 3~4 个月翻一番，也就是每年增长 10 倍，比摩尔定律两年增长一倍的速率快得多。

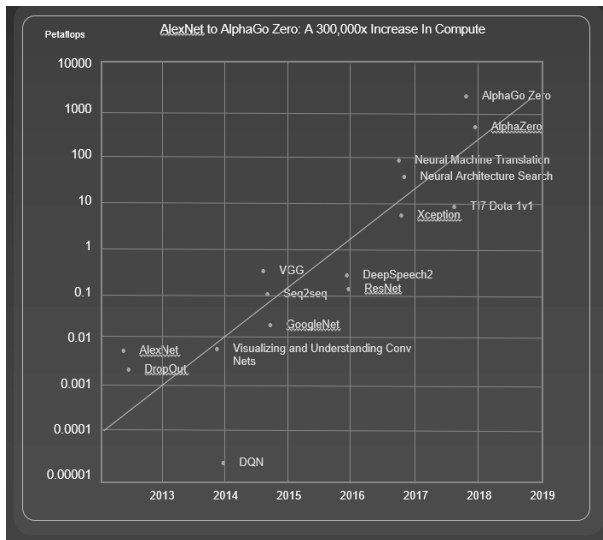


图 1-1 OpenAI 公布的 AI 算力报告

中国作为较早布局人工智能战略的国家，充分重视科技创新对于社会经济发展的助推价值，并通过政策引导，布局人工智能算力基础设施发展，夯实 AI 产业化的根基。

以人工智能新型计算算力为代表的人工智能计算中心（以下简称“智算中心”），是目前人工智能基础设施的重要组成部分。人工智能计算中心是以基于人工智能芯片构建的人工智能计算机集群为基础，主要应用于人工智能深度学习模型开发、模型训练和模型推理等场景。目前，智算中心通过地方的产业政策引导，并依托本地化的特色产业生态，构建规模化的算力中心，向企业提供普惠的公共算力服务、特色的数据和算法服务，支撑科研创新和人才培养，让算力服务更易用，也为未来人工智能计算算力如何布局提供参考。截至 2021 年 4 月，中国先后批准设立了北京、上海、合肥等 15 个国家新一代人工智能创新发展试验区。智算中心被越来越多的地方政府视为支撑和引领数字经济、智能产业、智慧城市和智慧社会发展的关键性信息基础设施。

当前，人工智能计算中心的发展面临新的形势。一方面，人工智能的发展对算力的需求持续攀升；另一方面，在国家“双碳”战略下，需要计算中心加强统筹建设和提升利用率，进一步减排降耗。智算中心逐渐走向网络化和集约化，实现算力、大模型、数据集、行业应用等人工智能要素流动共享。因此，智算中心不再作为独立的系统，而是逐步走向相互连接的算力网络，深化智算中心的高质量建设，是智算中心下一步发展的新形态和新范式。新型网络技术将各地分布的智算中心节点连接起来，构成感知、分配、调度人工智能算力的网络，可以更好地汇聚和共享算力、数据、算法资源，更好地满足我国经济社会高质量发展的新需求。

1.2.2 分布式多级算力需求

伴随着算力需求与日俱增，对算力的要求也开始呈现多面性特征，使得原有的算力横向扩展方式不再使用，分布式云、MEC、泛在计算等主张多级算力的算力方案走进了大家的视野。据 Gartner 的报告显示，分布式云在 2020 年和 2021 年，连续两年被列为了战略性技术。

CCSA 在 2020 年启动了《面向业务体验的算力需求量化与建模研究》的研究项目。在该项目中，对当下比较典型的业务的算力需求进行了量化分析。

1. 训练类业务

训练类业务是指通过大数据训练出一个复杂的神经网络模型，即用大量标记过的数据来“训练”相应的系统，使之可以适应特定的功能。训练需要极高的计算性能、较高的精度、能处理海量的数据、有一定的通用性，以便完成各种各样的学习任务（参见图 1-2）。训练类业务目前主要集中在云端，算力需求主要体现为算力的量级、异构的算力、算力的效率。

使用算法	应用场景描述	算力需求估算	网络需求估算	存储需求估算	备注
VGGNet	在数据集上训练网络模型，提升检测效果；以训练迭代一次为例(ImageNet中的ILSVRC2010数据集，包含了1000类，共1.2百万的训练图像，50000张验证集，150000张测试集)	19 PFLOPS (采用VGG16结构为例，在ImageNet数据集上训练1次所需的算力：1200000*16GFLOPS；数据来源: https://github.com/Lyken17/pytorch-OpCounter)	非实时性业务	VGG16模型权重大小138.37M；ImageNet数据集大小1TB	VGG16在ImageNet数据集为例
	在数据集上训练网络模型，提升检测效果；以训练迭代一次为例；(COCO数据集；包含80个对象类别，33万图像)	6 PFLOPS (采用VGG16结构为例，在CoCo数据集上训练1次所需的算力：330000*16GFLOPS；数据来源: https://github.com/Lyken17/pytorch-OpCounter)		VGG16模型权重大小138.37M；CoCo数据集大小20G	VGG16在CoCo数据集为例
Resnet50	在数据集上训练网络模型，提升检测效果；以训练迭代一次为例 (ImageNet中的ILSVRC2010数据集，包含了1000类，共1.2百万的训练图像)	5 PFLOPS (采用ResNet50结构为例，在ImageNet数据集上训练1次所需的算力：1200000*4.14GFLOPS；数据来源: https://github.com/Lyken17/pytorch-OpCounter)		ResNet50模型权重大小25.56M；ImageNet数据集大小1TB	ResNet50在ImageNet数据集为例
	在数据集上训练网络模型，提升检测效果；以训练迭代一次为例	2 PFLOPS (采用ResNet50结构为例，在CoCo数据集上训练1次所需的算力：330000*4.14GFLOPS；数据来源: https://github.com/Lyken17/pytorch-OpCounter)		ResNet50模型权重大小25.56M；CoCo数据集大小20G	ResNet50在CoCo数据集为例

图 1-2 训练类业务相关参数

2. 推理类场景

推理类业务是指利用训练好的模型，使用新数据推理出各种结论。即借助现有神经网络模型进行运算，利用新的输入数据来一次性获得正确结论的过程。也叫作预测或推断。推理对性能要求不高，更注重综合指标，包括单位能耗算力、时延、成本等（参见图 1-3），因此，越来越多的推理过程逐渐从云端转移到边缘端或者终端，算力从集中走向了分布、泛在化。

使用算法	应用场景描述	算力需求估算	网络需求估算	存储需求估算	备注
CNN	单张图片的人脸识别任务	10GFLOPS 输入尺寸为2560*1920，MTCNN所需要的算力约为10GFLOPS (数据来源: 源:https://zhuanlan.zhihu.com/p/92177798)			
	单路单流对人脸图像进行识别;应用在实验室环境	13 GFLOPS (由于CNN特征提取占80%的算力,所以可以推测出检测一张分辨率2560*1920的图片所需要的算力: 10/0.8=13)	时延<60ms	MTCNN模型权重大小186MB	MTCNN人脸识别算法为例(CNN约占80%算力需求)
	单路多流对人脸图像进行识别;应用在写字楼等场景,实现开发(300张图片开发为例)人脸识别功能	4 TFLOPS (300张并发检测所需要的算力: 300*0.013TFLOPS/每张)	时延<60ms		
	多路多流对人脸图像进行识别(16路,300张图像并发);应用在城市街道、闹市区	64 TFLOPS (16路监控视频高并发检测所需要的算力: 16*4TFLOPS/每路)	时延<200ms		
RNN	对一条语音进行语音识别	2 GFLOPS (数据来源: BIG-LITTLE NET: AN EFFICIENT MULTI-SCALE FEATURE REPRESENTATION FOR VISUAL AND SPEECH RECOGNITION 中的论文实验)	时延<60ms	DeepSpeech2普通话语音识别模型权重大小216MB	DeepSpeech2语音识别算法为例
	实现开发(500条语音识别为例)语音识别任务	1 TFLOPS (500条语音高并发所需算力:2GFLOPS*500)	时延<60ms		

图 1-3 推理类业务相关参数

3. 云 AR/VR 场景

传统 AR/VR 业务由于开发成本高、硬件受限、场景受限等弊端，未能得到大规模的普及。随着算力的普及增长以及 5G 的商用，移动式云 AR/VR 的业务场景将会大受欢迎。如图 1-4 所示，移动 AR/VR 业务是一种云和端结合的方式，本质是一种交互式在线视频流。对于云侧拥有超强算力和低时延的网络，该算力网络根据业务需求，一键部署与之匹配的业务能力，更多的渲染工作在云端完成，然后通过网络传送给用户端，如手机、PC、Pad、机顶盒等终端设备，用户通过输入设备（虚拟键盘、手柄等）对业务进行实时的操作。

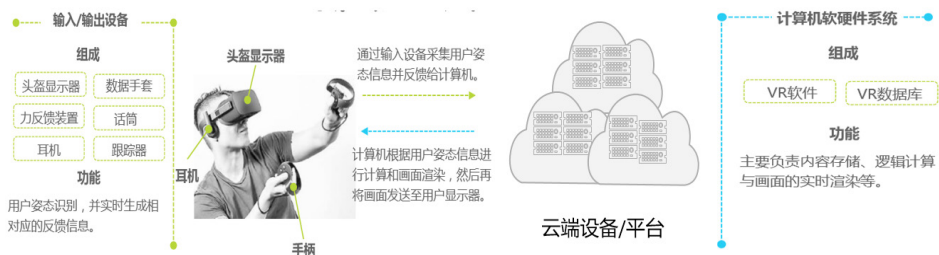


图 1-4 VR 系统组成及交互示意图

另外，云 AR/VR 业务在高铁、地铁等高速移动的场景下，用户侧终端设

备将会在多个基站甚至多个地域进行网络切换，这样与初始连接的云端节点网络之间的延迟增加，因为需要在多个云端节点进行切换，根据用户的实际情况进行统一调度和管理，将计算能力在多个节点之间无缝迁移，且保障流畅、切换无感的用户体验。这就要求算力网络节点能够快速调用计算能力，设计灵活的架构，进行弹性伸缩，满足用户的密集需求，如图 1-5 所示。

使用算法	应用场景描述	算力需求估算	网络需求估算	存储需求估算
	PC VR,移动VR 2DAR 动作本地闭环, 全景云端下载 远程办公, 购物等	40 EFLOPS 算力需求来自视频解码以及视频内容语义感知 和环境感知(数据来源:《泛在算力: 智能社会的 基石》立场文件-华为)	20Mbps 时延<50ms	运行环境: 内存 4G, 存储32G
	VR 新零售: 利用计算机 3D 建模等技术, 将商品的 3D 模型还原, 身临其境地 浏览商品, 更细致地观察商品		40Mbps 时延<20ms	

图 1-5 云 VR/AR 业务参数表

各类计算密集型业务对于算力的需求与日俱增，如工业自动化、机器视觉等应用，需要大量的算力资源。同时，与传统应用相比，诸如 AI 类的应用在时延上是非常敏感的，单纯地靠增加算力的能力或者扩充网络传输通道，已经不能很好地满足业务场景的需求。因此，为了满足当下新型业务对算力的多维度需求，业内的一致观点是，需要结合算力能力与网络能力，统筹全局调配基础设施资源，这也就是算力网络的宗旨和目标。

1.3 算力网络发展的基础

1.3.1 算力网络的技术底座

算力技术和网络技术的大力发展为算力网络提供了良好的技术基础。

1. 基础能力跨越式发展

一方面，算力已经从原来集中的、大型的数据中心走向了分散，中小企业甚至个人都拥有一定规模的算力。虽然单个算力的能力可能不是很高，但是算力的量级非常大。早在 1999 年，加州大学伯克利分校的几位科学家，就通过网络把世界上无数的计算机连接在一起，组成巨大的算力网络，来处理从浩瀚

的宇宙中收集的信号，并找寻地外生命（SETI）存在的证据，也就是 SETI@home 项目。如今，这个项目已经过去了 20 多年，算力的水平在遵循摩尔定律的基础上，有了长足的提升，社会上闲散算力聚集起来的能量将不可估量。在这个过程中，算力经历了大型机、分布式模式、云计算模式等由合到分，由分到合的过程，同时伴随着算力的管理、维护技术的日趋成熟。

另一方面，网络连接的能力、范围等都得到了扩展，资源的获取更加的无边界和无障碍。无论是移动网络还是数据网络，网络带宽越来越大。网络经历了 Mbps、Gbps、10Gbps、100Gbps 等阶段，网络这条公路越来越宽，能够承载的 bit 位越来越多，同时，移动通信网络也在 2G/3G/4G/5G 越来越宽的道路上演进。网络的运维管理智能化水平也逐渐提高。同时，网络的成本近几年也在不断下调。因此，对于资源需求方，已经无须太过于关注资源是从何处获取，资源的供给方也无须担心网络供给不足的问题。

2. 算网融合发展初现

算力和网络一直是计算机领域发展的两条主线。在早期的大型机时代，用户终端仅仅作为了显示器，并通过通信线路连接到大型主机，使用集中点的算力资源。那时候的网络仅起到了连接的作用，功能单一。随着个人 PC 的普及，计算资源逐渐分散，计算机网络逐渐从满足连接需求向满足计算协同的需求发展，比较典型的有 P2P 网络和网格计算。随着互联网业务的飞速发展，云计算高可靠且高弹性的资源供给模式与以 Web 为代表的互联网业务需求高度吻合，从而迎来了黄金发展期，也成为互联网经济发展的核心推动力。

在过去，算力和网络都沿着各自的轨道，往前迭代。虽然在网格计算的时代两者短暂地融合过，但是由于当时技术的局限以及商业模式的不成熟，网格计算最终还是成为了实验室的产物，并没有走向成熟的商业化。但是，时至今日，两者的融合趋势越来越显现。一方面，繁荣的移动互联网业务催生了移动通信网络、固定通信网络、云计算服务的连接需求，并且诸如 AR/VR、车联网等新形态的业务，从业务侧给网络提出了安全、可靠、可确定等要求。这就要求网络从原来默默支撑、哑管道的角色，逐渐走向前台业务，走进千差万别、需求各异的业务。另一方面，随着个人业务、企业业务向云的迁移，算力供给模式由传统的自有 IDC 提供向公有云、分布式云、混合云等多种形态演进，网络如何去适配新型算力结构的需求，也成为当下网络演进的课题。因此，在数字化

经济日益繁荣的时代，云网融合、算力网络、云网络、多云网络、NaaS 等代表了云网融合演进方向的词汇频频出现在从业人员的视野中。

算力网络作为资源分配与网络连接的一体化方案，将改变未来基础设施资源的供给模式，也会激发更多的商业模式。

1.3.2 “双碳”社会的国际共识

工业社会，工业革命和技术创新促使人类生产力水平不断提高，人类在改造自然界、创造极大财富的同时，也在不断地破坏着赖以生存的自然环境。能源的不合理使用、废物和有害物的排放造成的环境问题不断为人们所认识。其中，包括二氧化碳在内的温室气体的过度排放带来全球气候变化，也已被确认为不争的事实。在此背景下，“碳排放”碳足迹、低碳经济、低碳技术、低碳生活、低碳社会等一系列新概念、新倡议应运而生。

《联合国气候变化框架公约》是世界上第一个为全面控制二氧化碳等温室气体排放，以应对全球气候变暖给人类经济和社会带来不利影响的国际公约，也是国际社会在应对全球气候变化问题上进行国际合作的一个基本框架。1997年12月，联合国气候变化公约第三届缔约方大会在日本京都举行，会议通过了具有法律约束力的《京都议定书》，对2012年前主要发达国家减排温室气体的种类、减排时间表和额度等做出了具体的规定。2009年，在丹麦首都哥本哈根举行的联合国气候峰会上，欧美发达国家提议以现在各国的碳排放规模为基准，按照相应的比例来各自承担减排的责任。2015年12月，在第二十一届缔约方会议上，欧美国家又提出了“碳排放权”，这是一个全新的、有针对性的概念。“碳排放权”的提出，意味着碳排放行为开始被资本化。也就是说，在自己配额不够的情况下，可以通过向别人购买配额的形式获得，后者在替前者完成减排任务的同时，也可以获得收益。发展中国家想要发展经济，发展工业是不二的选择，然而发展工业就必然会产生碳排放，这就会受限于欧美发达国家。

因此，为了我国可持续发展的内在要求，履行国际责任、推动构建人类命运共同体的责任担当，中国政府提出了“2030年达到碳高峰，到2060年实现碳中和”的承诺，也就是国内的“双碳”战略。

作为高耗能行业，做好绿色转型、实现可持续发展成为数据中心响应国家“双碳”目标的必修课。据统计，数据中心能源侧的碳排放主要来自IT设备

及基础设施的电力消耗。在数据中心能耗结构中，主设备耗电占45%~55%，空调设备占35%~45%，电源及其他设备占10%~15%。空调设备的冷却系统能耗是数据中心的消耗大头。因此，优化数据中心的物理分布位置或者引入新的冷却技术等成为了数据中心绿色转型的主要方式。

“东数西算”工程可谓数据中心建筑布局优化的最典型例子。将新数据中心的建设地址选在能源丰富的西部地区，承接东部算力需求。以贵州为例，其年平均气温仅有15℃，常年无风沙，且拥有众多大型水电工程，电价便宜。这种得天独厚的气候和能源优势，不仅有利于服务器散热，而且有利于电路板保持洁净，以及用电成本与设备维护成本的降低，整体运营成本随之也保持在最优水平。

算力网络依托高速、移动、安全、泛在的网络连接，整合多层次算力资源，提供数据感知、传输、存储、运算等一体化服务的数字基础设施，能够实现东西部一体化的算力服务，是使能“东数西算”工程落地的重要技术。

1.3.3 算力网络发展的国家政策

算力网络是响应国家战略、加速技术创新、顺应产业发展、推动公司转型的必然要求，将为社会数智化转型带来全新机遇。算力网络是支撑国家网络强国、数字中国、智慧社会战略的根本要求，是对接国家规划，落实“东数西算”工程部署的重要支撑，是推动国家新型基础设施走向纵深的全新路径，将有力地推动算力经济的持续健康发展。

算力网络是信息科技创新的新赛道，是加快大数据、AI、区块链、物联网等多要素融合的重要载体。发展算力网络必然会引发大量跨领域融合技术和原创技术的突破。算力网络推动网络与计算两大技术体系融合发展，以网络创新优势带动算力网络创新发展，占据新的技术制高点。

算力网络是行业发展的新引擎，是行业价值重构的重大机遇。通过发展算力网络，可以有效融通多元业态、提供多元供给、形成多元服务，催生全新的商业模式，极大地拓宽行业边界、提升行业价值，促进产业高速发展。

算力网络是对云网融合的深化和新升级，主要体现在：一是对象升级，云是算的一种载体，算力将更加立体泛在，包含边端等更丰富的形态；二是融合升级，算力网络不仅是编排管理的融合，更强调算力和网络在形态和协议上的一体融合，同时也强化了以算为中心，ABCDNETS等多种技术的融合共生。三是

运营升级,算力网络对网络运营管理的要求更高,从一站式向一体化、智慧化演进四是服务升级,算力网络是以算力为载体,多要素融合的新型一体化服务。

算力网络也将成为6G网络发展的基础。在网络和计算深度融合发展的大趋势下,6G的核心愿景和应用场景呼唤基础设施的创新,要求网络 and 计算相互感知、高度协同,实现泛在计算互联,提升网络资源效率。

算力网络是运营商面向下一个十年的网络战略,从“算”和“网”两个角度来看,运营商的垄断性优势是网。从“算”的角度来看,在最近刚认定的44个2021年国家绿色数据中心中,运营商有14个,接近1/3。其他的数据中心来自阿里巴巴、百度、腾讯、字节跳动、华为等大厂,还有世纪互联等数据中心企业,以及像银行、能源这样的适配于垂直行业的数据中心。这也是我国算力来源的现状。算力网络,价值在于算力与网络的融合,现在的算力大户是互联网大厂,运营商的这场“借东风”,如何构建以及能否如愿,仍然是当下的一个重要课题。

1.4 算力网络的新应用形态介绍

1.4.1 新型算力组合

从成本和技术角度考虑,云计算服务向边缘计算领域扩展已经成为了业界的共识。传统的云计算服务通过规模效应降低成本的商业模式,迎来了蓬勃的发展。在云服务时代,云服务商建设超大规模的数据中心,并研发自动化、智能化的管理手段、数据中心节能技术、资源调配技术等,来降低基础设施资源的使用成本,提升灵活性。但是,随着边缘计算的兴起,云商通过技术以及资金投入构筑的市场准入门槛被大大降低。因此,市场的算力供给结构将出现非常大的变化。

算力网络作为一个中立的算力调配平台,为最终的业务需求匹配的算力组合,位置、供应商、算力类型等均可能存在差异,与当前的云环境有较大的差别。

1.4.2 新型算力模式

5G时代的到来,可以为更多的地方提供不受限制的高速网络接入,提供

更多的网络连接的可能性。接入网络的带宽、便利性、稳定性和安全性等方面，5G 都是一种网络连接方式的最好选择。尤其是随着 SD-WAN、智能网络、自智网络等新技术的引入，网络已经不再是瓶颈。并且随着确定性网络等技术的引入，网络的可保障性也得到了提高。因此，归属于不同所有方的资源，可以通过网络整合起来，形成一个整体供给。但是，在这个过程中，如何整合各方资源，给最终用户一个无差别的感知，成为一个新的问题。这个无差别的感知包括了业务体验、运营体验、管理及维护体验等方面的无差别。

为了实现这样一种状态，算力供给的商业模式将会被重塑。

首先，对于算力资源的整合。在过去，算力资源整合是以云服务商为中心的，云服务商通过规模效应来节省各类支出，包括建设、管理和维护等各方面。但是，对于边缘算力，在 DC 的建设、设备的管理维护、DC 节点的分布、DC 内网络资源的规划等方面都和传统的云数据中心存在一定的差异。如果为了整合边缘算力，需要雇用大量的人力，来进行点位的洽谈、后期离散设备的维护与运营，那么云服务商可能很难有意愿去维护和推进。

其次，中小型算力资源的提供方也很难开展相关业务。由于技术实力的限制，能够运营的产品和提供的服务相当有限，尤其当用户提出新的需求之后，很难及时提供有效的技术支持。同时，受限于企业规模，市场推广力度小，很难将自身的资源信息告知潜在用户群。

因此，算力网络可能是这些问题的一类解决方案。通过算力平台，构建中立的、公正的算力供需对接平台，提供方和消费方无须感知对方的存在，按照一致的标准和规则，进行交易，催生一种新的算力供给模式。

1.5 算力网络产业规划

1.5.1 算力网络的标准规划

在产、学、研、政等各方的努力下，算力网络标准的体系化工作在 ITU 和 CCSA 中同步开展。ITU 更侧重于算力网络的架构、应用场景，CCSA 则覆盖了端到端的算力网络体系，包含了算力网络的技术架构、运营架构等，对于算力网络未来的商业模式、商品形态等都进行了前瞻性的研究。

1. ITU-T 中的算力网络标准

ITU-T 的算力网络研究有以下主要方向：

- 以CPN为定义衍生了5种标准。“Y.2501 Computing Power Network – framework and architecture”标准将算力网络定义为一种资源分配优化的网络CPN。其核心要义在于基于用户需求和算网环境，通过网络控制面实现算力资源节点、存储资源节点、网络资源节点的最佳分配。该标准给出了CPN的4层架构，即南北的服务层（需求信息、计费、交易等模块）、控制层（信息采集、资源分配、网络调度等模块）和资源层，以及支撑实现业务和资源编排管理的编排层（资源和业务编排、算力建模、安全、算网OAM等模块）。同时该标准提出了保障算网可行性需求，包括算力资源统一度量、算力的采集和监控等。“Signaling requirements for Computing power Network”标准定义了CPN架构涉及的资源数据模型、接口数据模型和故障的数据模型等，同时定义了向上的各类资源采集及汇聚接口、向下的执行策略及资源配置等接口功能，给出了资源的采集与部署、故障的处理等信令流程。“Requirements and signaling of intelligence control for the border network gateway in computing power network”标准给出了边界智能网关设备集成了CPN功能实现基于业务需求的资源调度，包括域内的资源调度和跨域的资源信息交换实现的资源调度。“算力网络认证和调度体系架构”“NGNe编排增强的架构和要求支持算力网络”两份标准于2021年11月完成立项，拟对算力网络架构中两个关键功能块的架构、功能和需求展开研究。
- 以CAN为定义共开展了4项标准研究。“Functional requirements of computing-aware networking”项目给出了基于云网融合资源优化的基本实现需求，保留资源的统一度量、算力感知网络的广播、路由、算力感知网络服务合约、算力感知网络OAM等。
- 以CNC为术语的标准当前衍生了3份标准，“Management and orchestration related requirements and framework for computing and network convergence in IMT-2020 networks and beyond”标准拟制定算网融合编排模块的需求与架构。“QoS assurance-related requirements and framework for computing and network convergence supported by IMT-2020 and beyond”标准拟使用CPN架构对架构各控制层进行增强，实现QoS保障。“Requirements of

computing and network convergence for IMT-2020 and beyond”标准给出了算网融合的资源感知、资源与服务映射、资源管理、资源度量、服务连续性、融合AI/ML等的算网能力要求、交易要求、接口要求等。

2. CCSA 中的算力网络标准

自 2019 年至今，CCSA 已完成 14 项算力网络标准立项，涉及算网总体技术要求、算网设备和协议技术要求、算网关键模块技术要求、算网运营管理技术要求等。在已经开展的《算力网络 总体技术要求》项目中，主要研究了算力网络的 4 层架构中的算力服务技术要求“研究算力网络的算力服务层负责算力网络与用户服务的信息交互，将业务或者应用的服务请求，映射为服务应用信息及用户业务请求，包括算力请求等参数，发送给算力路由节点”，算力网络的路由技术要求“研究算力的控制和转发”，管理技术要求“包括算力服务编排、算力运营，算力建模与度量、算力节点管理、算力 OAM、算网安全等”。在《算力网络运营管理总体技术要求》项目中的研究主要包含：是面向用户的层面，研究算网的源自能力和服务定义；面向按需、灵活的业务需求，研究算网编排及其他各层的能力抽象、组合编排和封装、人工智能技术在算网资源和能力的辅助建模、产品模型的智能化组装方面的应用；网络控制层重点研究算力网络中的全局网络资源收集、分发；算力管理层研究全域多层级的算力资源注册、度量等，为算力交易提供支持；基础设施层需要研究如何能更快捷地部署新的业务，进一步向用户提供端到端资源可视、确定性时延保障的服务能力。在《面向算网融合的算力度量与算力建模研究》项目中，研究涉及算力资源建模、服务能力建模、综合能力建模。算力资源建模包括算力建模、通信能力建模和存储能力建模等；服务能力建模是从安全性、定位功能、移动性支持功能三个方面对服务能力进行建模；综合能力建模指使用算力节点动态算力综合性能计算方法，采用指标评价相似度对多维指标进行处理并得到动态算力综合评价指标。

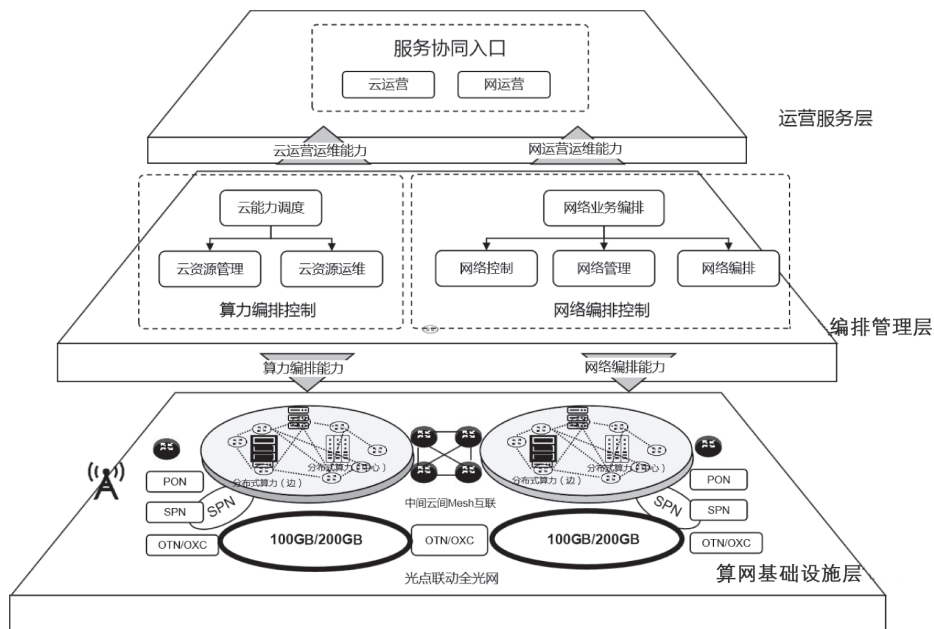
1.5.2 中国运营商的算力网络规划

算力网络能够由中国运营商提出，并在中国这块大地上发展和逐步完善，背后的驱动力有技术的发展，同时也有商业市场的需求。过去，运营商通过兴建网络、IDC 等基础设施资源，并以标准化的方式来开展一体化的服务，这种

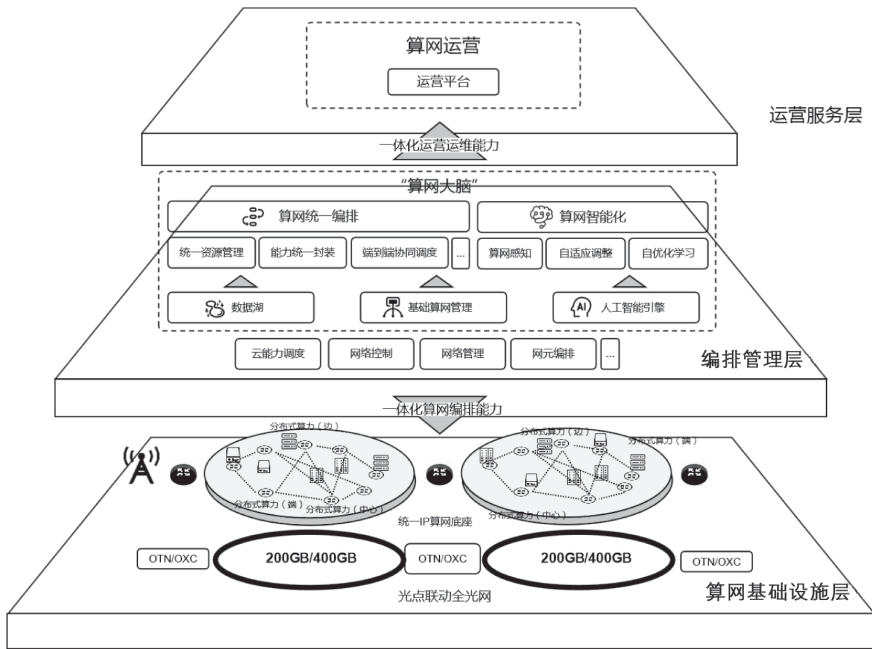
通过规模效应来降低成本的商业模式运转良好，给运营商带来了丰厚的利益。在云时代，以阿里云、腾讯云为代表的云厂商以这种商业模式，占据了云市场的半壁江山。运营商一向将自己定位为基础设施的供应商，投入了大量的资源想要与云厂商一较高下。但是，在过去的十几年中，运营商的斩获并不太理想。随着政企云、国资云的提出，运营商在云市场上逐渐有上升之势。同时，随着边缘云、私有云、混合云等算力架构的引入，运营商云网兼备的优势日渐突出。因此，算力网络作为一种云网边端协同的资源分配方式，成为了运营商进军算力供给阵列的有力武器。

2021年11月，中国移动董事长在移动合作伙伴大会上，正式将算力网络上升为移动的集团战略，引爆了产业界对于这项技术的极大关注。据悉，中国电信和中国联通也在筹划启动集团级的战略规划，对于算力网络的落地部署进行指导。从目前的公开资料中，能了解到三家运营商对于算力网络的一些观点和看法。

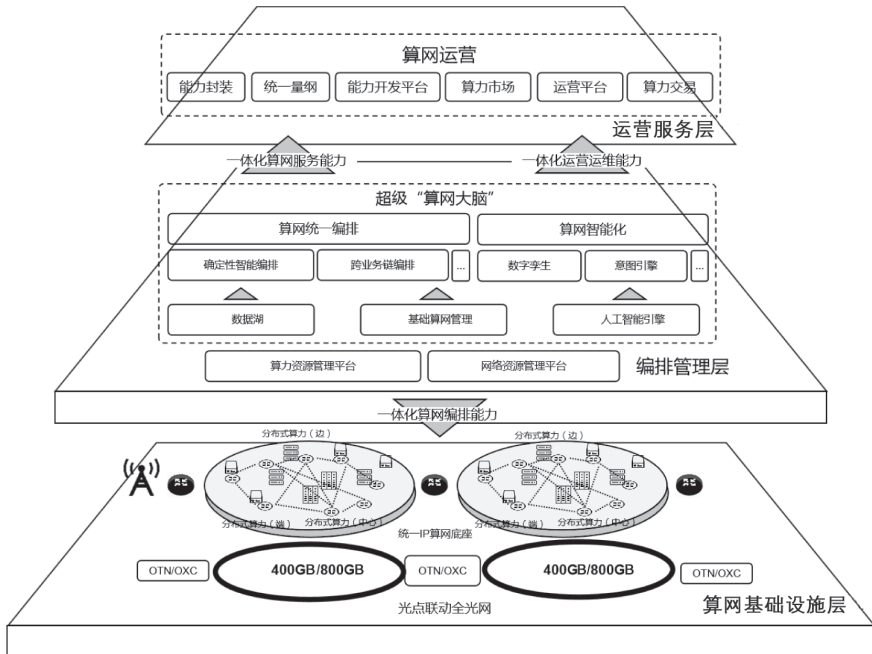
中国移动算力网络规划出现在《算力网络白皮书》中。在该方案中，中国移动将算力网络的最终发展目标定义为“算力泛在、算网共生、智能编排、一体服务”，同时，给出了“泛在协同、融合统一、一体内生”三个阶段（图 1-6 所示），并对各阶段的重点工作、关注重点进行了描述。



(a) 阶段一：泛在协同



(b) 阶段二：融合统一



(c) 阶段三：一体内生

图 1-6 中国移动算力网络的三阶段

从发展目标来看，中国移动的算力网络中，算力和网络是基础设施底座，其上会承载各类创新的 IT 技术和能力，同时，利用网络对于算力供给和需求的感知能力，实现多级资源的智能化统筹调度。资源层面，会关注云、边缘算力布局以及各算力之家的网络传输资源布局。能力层面，以云、边为底座，对外提供 CT 能力开放、数据智能服务、区块链服务等能力。

从《算力网络白皮书》描述的愿景以及近期的算力网络交流情况来看，目前移动处于泛在协同阶段，瞄准融合统一阶段，一体内生则是理想阶段。泛在协同阶段，重点推进算力资源、网络资源的协同，运营统一，管理分离。此阶段的技术攻关包含了云网编排系统升级、网络基础设施和算力基础设施的统一管理。融合统一阶段需要重点攻关算网智能化技术，推进算力、网络的融合管理、调配，实现资源统一管理、统一编排和调度。

中国电信算力网络的体系化研究成果输出至 ITU-T 的 Y.2500 系列标准中。其中，项目“Y.2501 Computing Power Network – framework and architecture”是中国电信的首个算力网络标准项目，在这个项目中给出了算力网络的定义、参考架构。从生产侧来说，电信并没有给出明确的算力网络规划。在中国电信的《云网融合 2030 技术白皮书》中，将算力网络作为云网融合演进过程中所需的一种关键技术。因此，亚信科技认为电信的算力网络演进将沿着新一代云网运营系统向前推进（如图 1-7 所示）。这一观点也得到了电信集团的确认。

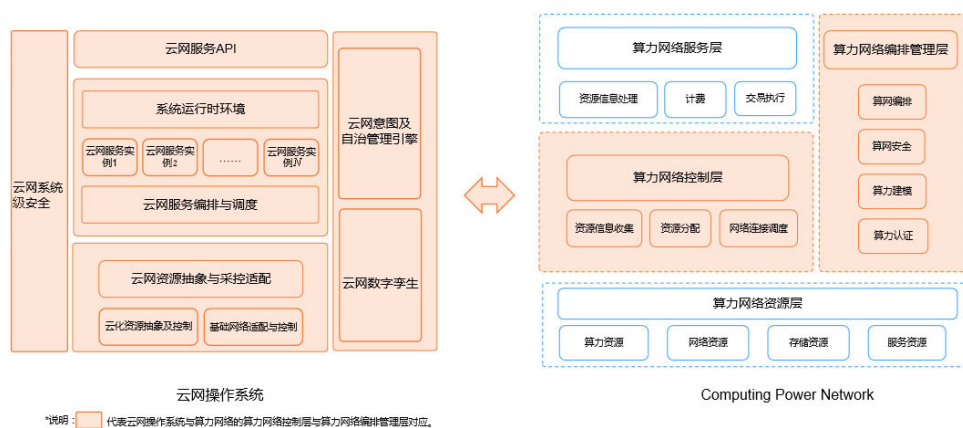


图 1-7 中国电信云网操作系统与算力网络

中国联通在《云网融合向算网一体技术演进白皮书》中，描述了中国联通认为的算力网络一体化架构。该白皮书中，明确地将算力网络定义为云网融合 2.0，是一种新型的、可以调配算力的网络架构（如图 1-8 所示）。与中国移动和中国电信的算力资源储备相比，中国联通算力资源布局较弱，因此，对于算力网络的研究还是以研究院为主体，重点考虑不同的算力供给模式对网络架构演进的需求。

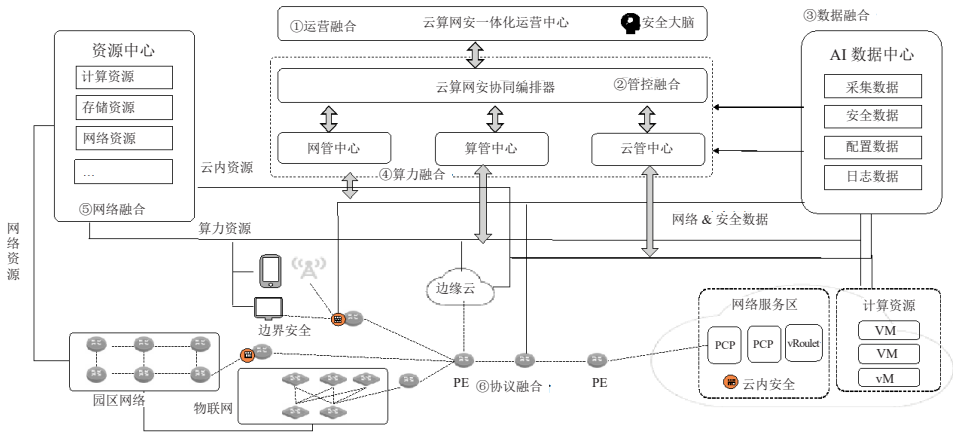


图 1-8 中国联通算力网络规划