

算力互联网体系架构 研究报告

(2025 年)

中国信息通信研究院

2026年1月

版权声明

本报告版权属于中国信息通信研究院，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院”。违反上述声明者，本院将追究其相关法律责任。



前 言

2025年8月，国务院印发《关于深入实施“人工智能+”行动的意见》，提出要“加强智能算力互联互通和供需匹配”。算力成为支撑数字经济发展的关键基础设施，是智能时代全球性的紧缺战略资源。算力资源分布不均、技术体系多元、调度能力薄弱等问题日益凸显，跨区域、跨主体、跨架构的算力协同仍面临“找不着、调不动、用不好”的现实挑战。当前，我国产业各方均在积极探索算力互联成网路径，如运营商算力并网、各地算力平台、中国算力网、超算互联网等，推动算力供需调配，形成行业“条状”、区域“块状”算力局域网。

工业和信息化部印发《算力互联互通行动计划》，形成了“先互联、再成网、同步建市场”的总体思路。中国信通院牵头，联合基础电信运营商、云服务提供商、行业用户、高校等产学研单位，以及多地新型互联网交换中心和大数据集团，开展算力互联网体系架构的研究工作。通过广泛调研国内外算力互联的实践路径和架构，分析算力互联网的组网架构、业务生态、核心技术等关键环节，编制《算力互联网体系架构研究报告（2025年）》。

本报告阐述了算力互联网发展背景、概述、体系架构、组网模式、业务及生态、技术体系和发展建议，旨在提供一个全面、深入、前瞻性的参考框架，促进业界对算力互联网形成统一共识，赋能人工智能、科学计算、工业计算等业务发展。同时，本报告也将为后

续的技术创新、标准制定、测试验证及推广应用等工作提供参考，推动算力互联网健康、快速、可持续发展。

算力互联网的发展和演进是一个持续不断的过程，编制组将密切关注国内外算力互联网的发展动态，积极听取产业界的意见与建议，不断完善和优化算力互联网体系架构的研究内容，适时修订并发布报告的新版本，以更好地推动算力互联网发展。



目 录

一、 算力互联网发展背景.....	1
(一) 人工智能产业快速发展, 算力需求极大扩展.....	1
(二) 全球布局算力互联成网, 多措并举抢占高地.....	2
(三) 我国探索算力互联路径, 体系研究亟待完善.....	3
二、 算力互联网概述.....	4
(一) 算力互联网概念定义.....	4
(二) 算力互联网流程机理.....	5
(三) 算力互联网场景需求.....	7
三、 算力互联网体系架构.....	10
(一) 体系架构简介.....	10
(二) 体系架构释义.....	11
四、 算力互联网组网模式.....	14
(一) 算网设施层组网.....	14
(二) 互联资源层组网.....	16
(三) 应用服务层组网.....	18
五、 算力互联网业务及生态.....	19
(一) 算力互联网业务概况.....	19
(二) 算力互联网生态主体.....	21
六、 算力互联网关键技术.....	24
(一) 算网设施层技术.....	24
(二) 互联资源层技术.....	27
(三) 应用服务层技术.....	30
七、 算力互联网建设进展.....	31
(一) 算网设施层: 算网统筹协同初见成效.....	31
(二) 互联资源层: 多级平台体系初步建成.....	32
(三) 应用服务层: 多元算力服务初显活力.....	33
八、 算力互联网发展建议.....	34

（一）推进设施互联，构建高速传输网络.....	34
（二）强化资源互用，实现高效供需匹配.....	35
（三）推动业务互通，加速融合贯通调度.....	35
（四）培育场景赋能，提升普惠便捷体验.....	36



图 目 录

图 1 算力互联网示意图.....	5
图 2 算力互联网机理示意图.....	6
图 3 算力互联网与传统互联网资源使用对比示意图.....	8
图 4 算力互联网与 NGNe 组网架构对比图.....	10
图 5 算力互联网协议架构与互联网对比图.....	11
图 6 算力互联网体系示意图.....	12
图 7 算力互联网架构示意图.....	13
图 8 算力互联网组网架构示意图.....	14
图 9 算网设施层示意图.....	15
图 10 互连资源层与其他层关系示意图.....	17
图 11 算力互联网各节点和平台关系图.....	18
图 12 应用服务层与其他层关系示意图.....	19
图 13 算力互联网体系相关主体图.....	23

一、算力互联网发展背景

近年来，人工智能大模型技术取得突破性进展，随着“人工智能+”行动的持续推广，预计未来几年人工智能将广泛应用于数字经济和生产生活的各个领域，人工智能应用需求将极大扩展。因此，用户需要接入优质高效的人工智能模型，同时也需要随时、随地、按需、低成本的算力资源支持。我国算力产业发展迅速，国家“东数西算”工程深入实施，八大枢纽十大集群建设稳步推进，公共算力基础设施规模持续扩大，智能算力供给能力大幅提升，算力总规模稳居全球前列。我国算力按照区域资源禀赋优化布局和集中建设，但同时存在多种异构芯片技术路线，算力提供主体也较多元化。为满足人工智能、科学计算等任务实时调用合适算力资源的需求，实现全国计算资源优化配置和高效服务，需要在统筹全国算力设施区域优化布局的基础上，实现跨主体、跨架构、跨地域的算力资源互联。

（一）人工智能产业快速发展，算力需求极大扩展

人工智能应用场景持续深化。Gartner 预测，2025 年全球人工智能总支出近 1.5 万亿美元，并将在 2029 年增至 3.3 万亿美元，五年复合年增长率为 27.5%¹。2024 年我国人工智能产业规模突破 9000 亿元，同比增长 24%²。生成式人工智能技术不断向具体应用场景纵深渗透。人工智能大模型在参数规模、多模态能力等方面不断突破，

¹ 数据来源：Gartner 《Forecast Alert: AI in IT Spending, 2Q25》

² 数据来源：中国信息通信研究院测算

与办公协同、教育普惠、工业设计、内容创作等场景深度融合，构建了覆盖多个领域的智能应用生态。

算力基础设施需求持续增长。随着 DeepSeek 等大模型的应用普及，其处理复杂任务（如数据整理、研究报告生成）时需消耗远超对话场景的 token 量，推动算力需求陡增。同时，企业端因 AI 应用扩展（如智能制造、金融分析）开始自建模型训练，进一步加剧算力压力，如 DeepSeek 单日推理调用 2500 万次，需 1814 张 H800 芯片支持，而企业级训练场景对算力冗余和效率的要求更高。这种双重驱动下，算力需求从训练端向推理端迁移，推动全球基础设施需求持续扩张。

（二）全球布局算力互联成网，多措并举抢占高地

主要国家发布算力互联相关政策。美国能源局发布《高性能数据基础设施计划》，推动实现东西海岸 40 余个机构算力互联互通。欧盟委员会发布《2023—2024 年数字欧洲工作计划》，通过政策支持与专项投资，增强欧洲范围内算力系统的互操作性，推动构建互联可信的统一算力设施。英国科学、创新与技术部发布《人工智能赋能科学战略》，提出将探索一套系统，协同英国国家研究与教育网络内的（包括 CPU、GPU 或其他加速器）闲置计算资源赋能科研。

头部企业创新算力互联调度技术。谷歌积极布局多数据中心分布式训练，加速推动康瑟尔布拉夫斯、奥马哈、帕皮隆和林肯市四个园区互联互通，打造 GW 级的人工智能训练互联集群；OpenAI

和微软计划将多个超大型园区互连，并在全国范围内进行大规模的分布式训练；英伟达推出 Lepton 调度平台，加速构建覆盖其全球 GPU 算力的一张虚拟网络，使开发者能够在全球任意地方灵活接入英伟达的 GPU 算力资源，便捷高效地开展模型部署、训练和推理任务，并实现资源的弹性扩展与按需调度。

（三）我国探索算力互联路径，体系研究亟待完善

地方政府发起算力调度平台的建设。平台侧重点是为各区域汇聚算力，服务本地区以及周边地区的算力需求。地方政府各自建设并运营编排调度系统，将区域内算力资源汇聚形成资源池，进行供需对接、交易购买、使用调度。由于各平台接口等方面未实现统一，区域算力调度平台之间相对独立，形成了算力区域“块状网络”。

电信运营商提出算力网络和算力并网方案。算力网络的侧重点是让网络层融合算力信息，将任意数据中心之间数据流动的最优路由设计成带有段标识符的算力路由标签，由每个数据中心出口路由器利用基于 IPv6 转发平面的段路由技术封装在流量包里，在承载网进行路由转发。算力并网侧重点是在调度层将社会其他闲散算力或者自身不同类型的算力接入自身资源池统一管理，如中国移动“算网大脑”、中国电信“息壤”平台等，形成移动、电信各自一张算力“条状网络”。

体系化的算力与网络融合研究仍待完善。研究问题和对象方面，算力网络等研究大多是从通信模型和承载网视角出发，以“网”为

主，强调在承载网层面增强对算力需求的支持。由运营商主导数据分发与链路调度，可利用既有网络基础设施进行算力感知，提升网络能力，但仍需推动更多“算”企互联参与整体架构。基础理论方面，多数文章为工程架构，缺少面向算力互连网络的概念内涵和理论基础，尚无科学理论基础支撑，难以支撑进一步的创新发展。产业方面，目前实践仍以单点突破为主，难以形成全球化广泛的资源互连。

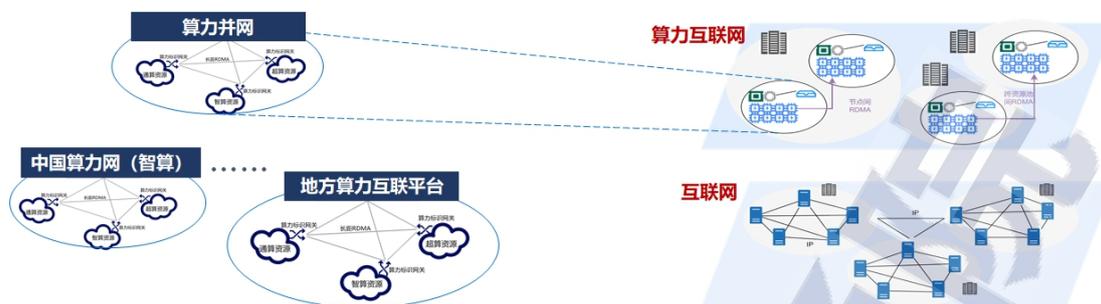
综上所述，亟须为算力互连成网提供有效的理论支撑，加快各个体系的算力网互连，形成算力互连网。

二、算力互连网概述

（一）算力互连网概念定义

算力互连网（Internet of Compute）是指互连网面向计算任务与调度需求进行能力增强和系统升级，通过在互连网上构建统一算力标识和协议接口、增强异构计算和跨域高性能网络传输能力等方式，使计算任务及其相关数据可精准寻找相适应的算力资源并高效执行，形成算力之间智能感知、实时发现和按需获取的一张逻辑互连网络。³

³ 栗蔚，余晓晖，张博圣，孙松林等.算力互连网架构：基于熵平衡支持算力资源跨域互连的下一代网络架构[J].通信学报,2025,46(09):1-16.



来源：中国信息通信研究院

图 1 算力互联网示意图

（二）算力互联网流程机理

算力互联网流程机理是指算力应用（即计算任务及其数据）在不同算力资源之间流动并完成计算的全过程。与传统应用计算任务上行数据小、连接持续时间长的特点不同，典型算力互联网应用的计算任务具有上行数据量大、计算密集且时间集中的特征，例如大模型推理等人工智能任务，以及气象预测等科学计算任务。

算力互联网中，传输的对象是“计算任务”。任务经编码识别后，通过最优路径信道传输，在目标资源池中部署执行。该过程涵盖四个重要环节，即计算任务识别、任务算力匹配、数据传输流动和任务计算执行，如图 2 所示。当计算任务需要从一个资源池流动到另一个资源池时，首先进行任务定义，明确任务类型以及对算力、存储、网络等关键资源的需求规格；其次是进行资源感知，基于算力标识体系查询并匹配满足任务需求的目标资源池；随后进入传输流动阶段，将任务相关的数据高效、可靠地传输至目标资源池；最后进入计算执行阶段，在目标资源池中进行计算处理，并将结果返

回至请求方。



来源：中国信息通信研究院

图 2 算力互联网机理示意图

1. 计算任务识别

平台系统基于统一的规则，根据应用的类型、性能要求等属性对计算任务进行识别。此环节中，计算任务及其相关属性被标准化定义，平台通过统一的识别机制，提取任务的类型、性能要求、数据位置、地理偏好等关键参数，进而评估任务的调度复杂度与资源需求边界，为后续算力匹配调度决策提供依据。

2. 任务算力匹配

平台系统依托算力互联网中的算力统一标识体系，结合计算任务的识别结果，将任务与跨区域、跨资源池的适配算力资源进行匹配。此环节中，算力资源被抽象为结构化的算力标识符，对其位置分布、计算能力、存储状态、网络带宽等关键属性进行标准化表征，进而构建“任务-资源”的映射关系，支撑算力资源的精准发现与高效调度。

3. 数据传输流动

计算任务及其相关数据通过全光网络、弹性网络等高性能网络基础设施，从人工智能计算中心（AIDC）等算力资源节点高效传输至目标算力资源。此环节需要构建高性能、低时延的网络通道，保障调度指令与状态反馈的时效性与可靠性，支撑跨域算力协同的连续性与稳定性。

4.任务计算执行

计算任务基于增强型异构计算框架完成部署，并在目标算力节点上高效执行。此环节中，平台系统对任务信息进行解析，在选定的算力节点上完成任务的执行与结果输出，同时反馈资源的使用状态和执行进度，实现算力执行过程的可监控、可管理。

（三）算力互联网场景需求

当本地算力资源不足或需要调用异构类型算力时，计算任务需迁移至其他的算力资源池执行，该过程需要多种算力资源跨区域互联互通。与水电等传统公共基础设施通过网络输送到用户侧不同，算力互联网中的计算任务及其关联数据需要传输至具备匹配能力的目标算力资源池完成计算，并将结果返回请求方。区别于传统基于固定区域、固定资源配置的算力计量方式，算力互联网面向跨域互联的动态资源环境，采用“卡时”“核时”等细粒度、可度量的资源使用单位，实现对算力消耗的统一计量与服务化核算。



来源：中国信息通信研究院

图 3 算力互联网与传统互联网资源使用对比示意图

1. 人工智能推理任务

人工智能推理任务需要根据应用时延敏感性、数据规模、请求来源位置等实际需求，灵活接入不同地域、不同类型的算力资源，以满足用户对算力资源“随时、随地、按需”调用的服务诉求。基于算力互联网架构，人工智能推理模型的服务可用性较单一节点算力部署有显著提升。当推理任务被调度至用户近端的边缘或区域算力节点时，首 token 响应时间与增量 token 生成时延均显著降低，用户体验得到明显改善。如推理加速服务商，依托互联算力池动态调度模型服务实例，实现服务能力的弹性扩展与性能优化，切实满足多样化、高实时性的算力接入需求。

2. 游戏、视频渲染等计算任务

游戏、视频渲染等计算密集型任务，需根据应用的数据流特征、视频流稳定性及低时延传输等实际需求，将渲染任务部署在靠近用户侧的算力节点，实现视频流的就近生成和高效分发。该场景依赖跨域算力资源的协同调度与灵活部署，以保障用户获得高质量的游戏交互体验和流畅的视频观看服务。基于算力互联网架构，系统可

将视频流原始数据或游戏渲染任务动态调度至临近用户的边缘或区域算力节点，在本地完成画面渲染与帧缓存，显著降低端到端传输时延，提升服务响应速度与用户体验。如云游戏服务，随着用户地理位置的变化，系统需实时调度相应的渲染实例至就近算力资源池，确保画面连续性、操作低延迟与运行稳定性，从而保障沉浸式、无卡顿的游玩体验。

3.本地算力资源不足时的调度任务

当本地算力资源无法满足计算任务需求时，要根据任务特征和数据处理需求，在跨域或跨服务商的算力资源中寻址并匹配适配的算力节点，完成任务的远程调度与执行，以保障计算任务的连续性与效率。基于算力互联网架构，系统可根据任务的算力需求，动态构建最优算网路径，在互联的异构算力资源池中实现算力资源的弹性调度。如科学计算任务，通常由多个子任务构成，各子任务对计算能力、加速器类型、内存带宽等资源需求各不相同。通过将不同资源需求的子任务精准调度至匹配的算力资源池执行，可有效提升整体任务的并行处理能力与执行效率，缩短计算周期。

4.人工智能分布式协同训练任务

人工智能分布式协同训练任务需要根据基础大模型参数规模、算力需求、数据中心电力供应等关键因素，在邻近多算力中心间进行互联调度，实现跨节点的协同计算与参数同步，支撑超大规模参

数模型的高效训练。该场景对网络性能要求极高，依赖高带宽、低时延、无损传输的直连网络基础设施，以保障大规模任务并行与梯度同步的稳定性，因而技术复杂度与建设成本较高，目前尚在短距离的协同试验阶段。

三、算力互联网体系架构

（一）体系架构简介

算力互联网体系架构充分参考主流设计方法论，结合算力互联互通试验验证工作和试验网实践经验，构建科学、可扩展的架构。架构重点参考下一代网络演进 NGNe（Next Generation Network evolution）的理念和路径，将算力互联网整体划分为算网设施层、互联资源层和应用服务层三个逻辑层级。各层功能设计以业务需求为导向，逐层解耦、协同演进，支撑算力资源的高效互联、统一调配与服务化供给。



来源：中国信息通信研究院

图 4 算力互联网与 NGNe 组网架构对比图

算力互联网协议架构在充分兼容传统互联网 OSI 五层模型的基础上进行了增强和演进。协议架构包括物理层、数据链路层、网络层、传输层和应用层，继承了现有网络体系的分层结构。在此基础上，针对算力协同对高性能通信的需求，在传输层引入 RDMA（Remote Direct Memory Access，远程直接内存访问）等低时延、高吞吐传输协议，显著提升大规模数据交换场景下的传输效率，如图 5 所示。

	功能	算力互联网协议簇	互联网协议簇
应用层	感知算力标识、获取可用算力资源，发送任务信息 封装任务数据，发起数据传输	TCP/IP应用层协议： HTTP/TFTP/DNS	HTTP/TFTP/DNS/Telnet/SMTP
传输层	实现对数据包的传输控制，例如分段和重组	RDMA	TCP/UDP
网络层	实现数据包在网络间的高速传输和路由选择	IP(IPv4/IPv6寻址)	IP(IPv4/IPv6寻址)
数据链路层	识别主机唯一标识，提供错误检测盒纠正功能	Ethernet (MAC地址寻址)	Ethernet (MAC地址寻址)
物理层	实现数据在物理介质上的传输	10GBASE-T/光纤以太网	10GBASE-T/光纤以太网

来源：中国信息通信研究院

图 5 算力互联网协议架构与互联网对比图

（二）体系架构释义

算力互联网体系架构的核心分层涵盖算网设施层、互联资源层、应用服务三层架构，以核心架构为基础，从组网模式、业务生态、技术体系等维度对各层级进行系统性剖析，明确各层的功能定位、系统能力边界及实施部署路径。

算力互联网体系架构重点关注互联资源层、应用服务层的研究，

算网设施层由基础电信运营商、数据中心服务商等建设，从完整性考虑纳入算力互联网研究范围。



来源：中国信息通信研究院

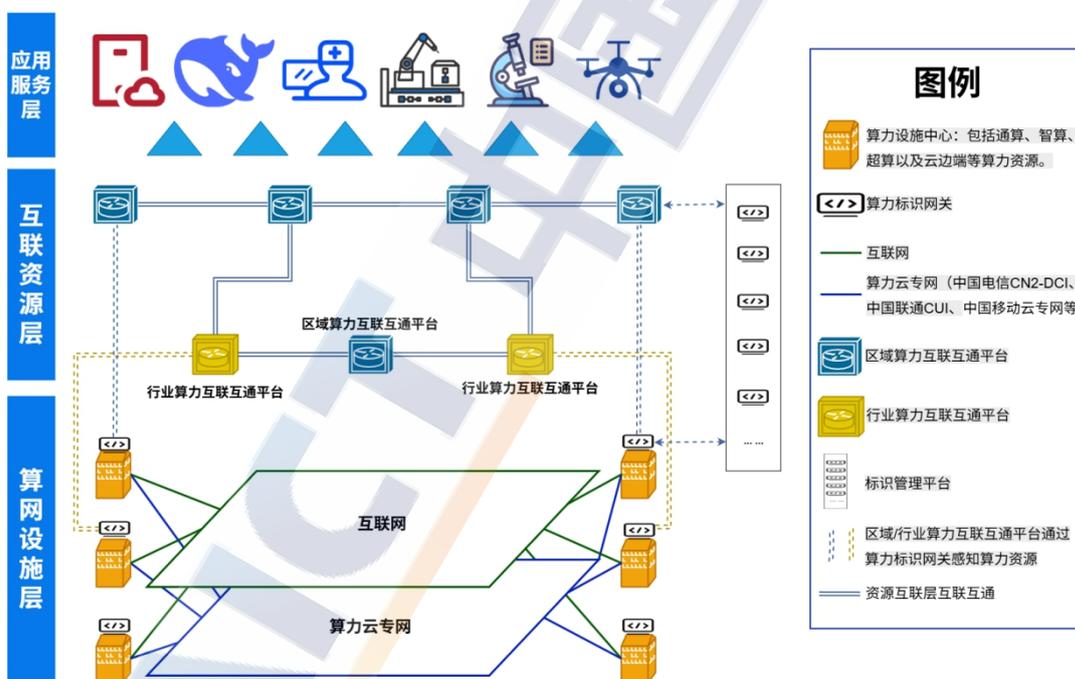
图 6 算力互联网体系示意图

算网设施层由计算与传输两类物理基础设施组成，是算力互联网的底层资源载体。计算设施包括通用计算、智能计算、超级计算等数据中心以及云、边、端等算力资源。传输设施包括公共互联网、企业算力专网等，承载算力节点间数据高效、可靠传输的承载功能。

互联资源层由国家算力互联网服务平台、区域和行业算力互联互通平台组成，位于算网设施层和业务应用层之间，是算力互联网的核心节点。该层通过互联互通平台提供一系列开放的资源互联能力，包括算力标识注册、标识解析、查询检索、路径调度参数生成、居间结算、算力度量等关键功能。其中，国家平台负责管理部署在算网设施层的算力标识网关，构建全国统一的算力资源寻址体系；

区域和行业平台基于标识网关实现异构算力的接入汇聚，支撑本地化、专业化算力服务运营，提供经营性服务。区域和行业算力互联互通平台可以采用新建独立系统或对现有平台实施标准化升级改造两种方式，实现服务能力的对外开放与互联互通。

应用服务层面向用户提供多元化的算力服务，主要包括传统算力资源式服务、算力调配接入服务以及算力应用服务等不同算力服务业务。该层通过统一的服务接口与服务目录，实现算力能力的封装与开放，为用户提供建立在统一调度体系之上的便捷化、普惠化算力使用模式，助力实现“像用电一样使用算力”的服务愿景。

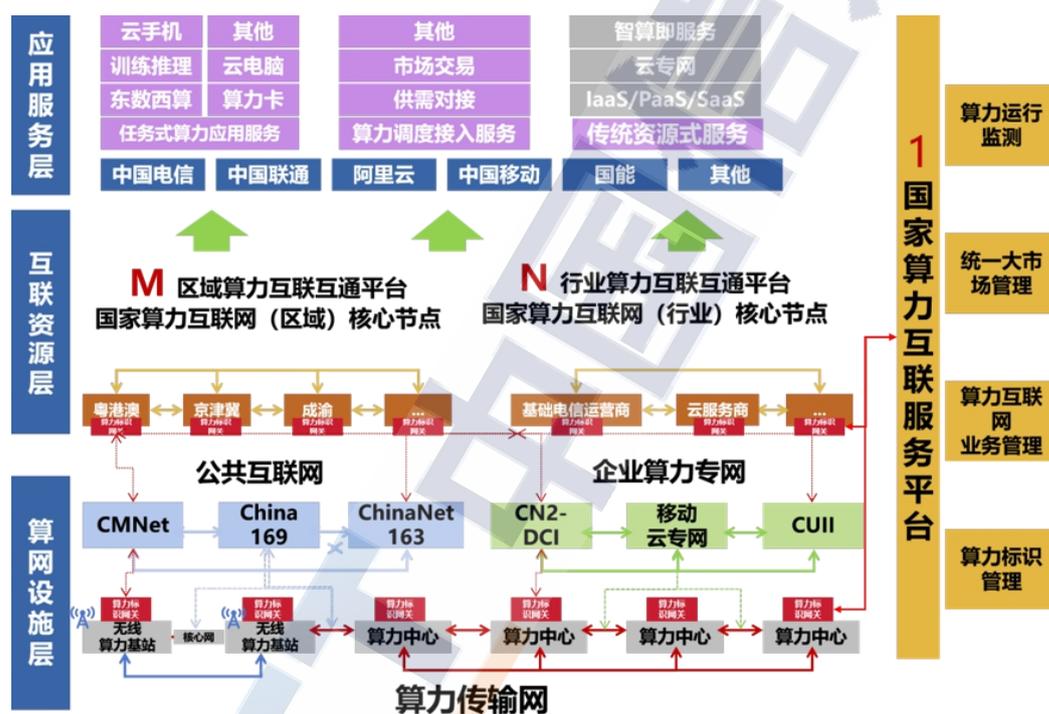


来源：中国信息通信研究院

图 7 算力互联网架构示意图

四、算力互联网组网模式

算力互联网整体组网架构分为算网设施层、互联资源层及应用服务层三个层级，具体架构如图 8 所示。从实际组网需求出发，实施单位可基于该参考架构，对现有系统进行标准化升级改造，或选择新建符合规范的算力互联平台，实现算力资源的统一接入、高效调度与服务化供给。



来源：中国信息通信研究院

图 8 算力互联网组网架构示意图

（一）算网设施层组网

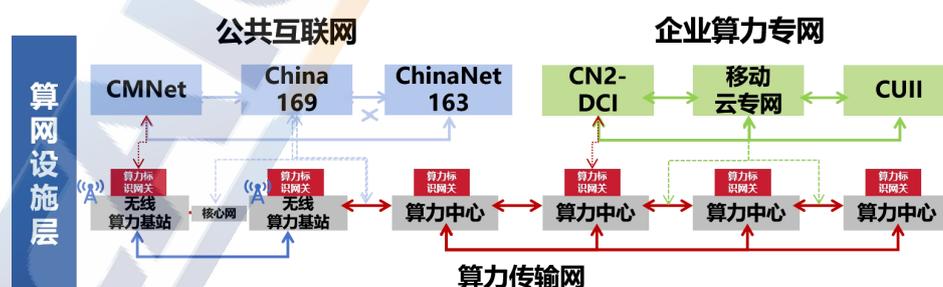
算网设施层组网是指连接各类算力中心的传输互联网络软硬件设施体系，涵盖了算力资源间的多种组网模式。从架构层级来看，该层底层为算力传输网，提供高速、可靠的数据通路支撑；上层则

根据算力服务在时延敏感型、带宽需求、安全隔离等方面存在差异，形成公共互联网、企业算力专网等多种模式并存的融合组网格局。

算力传输网是基础电信运营商为支撑算力任务与数据的高效、低成本传输，依托全光网、400G/800G 等高性能网络技术，在不同算力设施间构建的高性能物理承载网络。该层级涵盖固定网络设施与无线移动网络设施两大类型，是算力互联的底层传输底座。

公共互联网是面向社会公开开放、实现互联网节点大规模互联的基础性网络基础设施，主要是为提供普适性、广覆盖的互联服务。公众需求提供基础性互联网络服务，如中国电信 ChinaNet163、中国联通 CHINA169、中国移动 CMNet 等。

企业算力专网是面向低时延、高带宽、低抖动等业务需求构建的专用网络，具备高质量连接能力。该类网络由电信运营商基于现有高品质网络基础设施升级而成，典型代表包括中国电信 CN2-DCI、中国联通 CUII、中国移动云专网等。在此基础上，云服务商可依托物理专网，结合多种灵活连接方式，构建按需定制的虚拟化网络服务，即虚拟算力云专网。



来源：中国信息通信研究院

图 9 算网设施层示意图

（二）互联资源层组网

互联资源层组网是互联互通平台的核心组成部分，其主要功能是在控制层面实现标准化算力标识注册、算网参数解析、度量计量等关键能力，为上层算力调配接入及应用服务提供统一的算力资源汇聚与互联互通基础支撑能力。该层级通过构建算力互联网核心节点，支撑跨区域、跨服务商的算力资源互联。

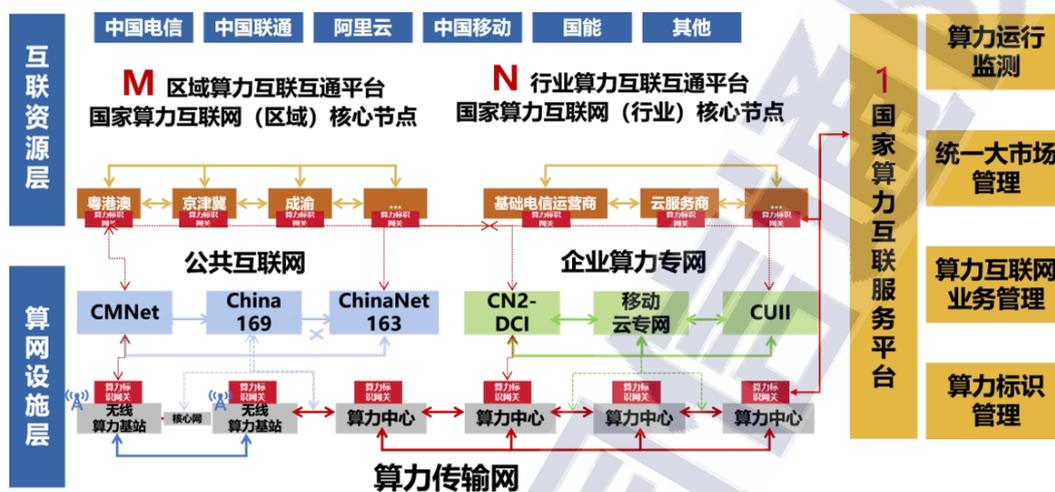
结合当前各地及各企业的算力互联实践基础，互联资源层采用“1+M+N”节点架构，构建具备灵活调度能力的逻辑网络体系。其中“1”代表国家总体平台，负责标准化算力标识的分配与管理；“M”代表区域核心节点，负责汇聚区域内的算力资源；“N”代表行业核心节点，负责汇聚行业内的算力资源。区域和行业节点平台建议采用共建共享、对等连接的方式，实现彼此间标准化互联。

国家算力互联网服务平台，承担算力资源标识标准化分配、运行监测、信息发布等综合管理职能，实现全国算力资源统一调配和管理。

国家算力互联网（区域）核心节点，整合统筹区域内跨服务商的多元异构算力资源，对接国家平台获取唯一标识编码，承担算力资源入网入市管理、数据快递线路开通、算力市场监测等功能。

国家算力互联网（行业）核心节点，当前主要面向基础电信运营商、云服务商等具备全国算力资源布局且正在开展算力并网的企业，结合算力专网或云专网形成算网一体化调度平台，依据算力互

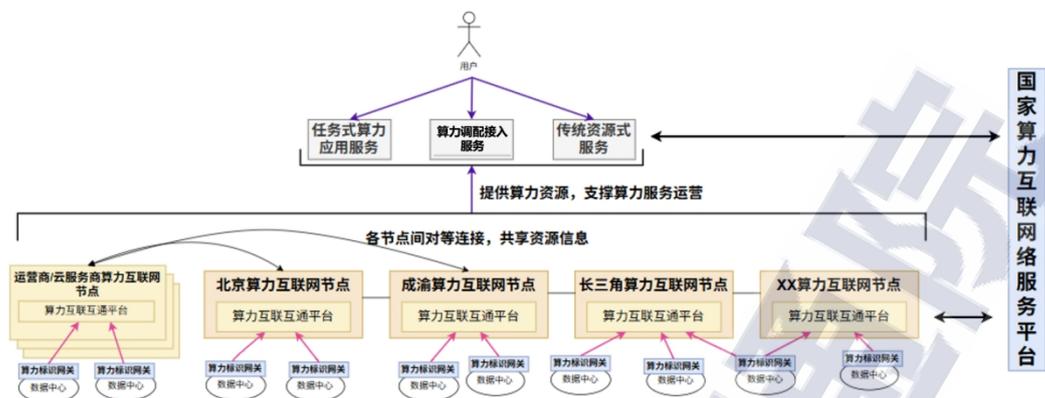
联网建设相关标准规范升级已有平台，接入国家算力互联网平台体系，承担加速算力供需匹配、培育算力服务业态等功能。未来其他行业可与核心节点合作，开展面向本行业的算力服务。



来源：中国信息通信研究院

图 10 互连资源层与其他层关系示意图

算力互联网各级节点之间具有对等互联关系。区域及行业核心节点通过算力互联互通平台接入公共算力资源，并基于对等交换机制实现算力资源信息共享与协同。一方面该机制向上为算力调配接入及应用服务层提供基础支撑；另一方面各节点需同步向国家算力互联网服务平台报送算力标识信息，以实现全网资源的统一调配和管理。

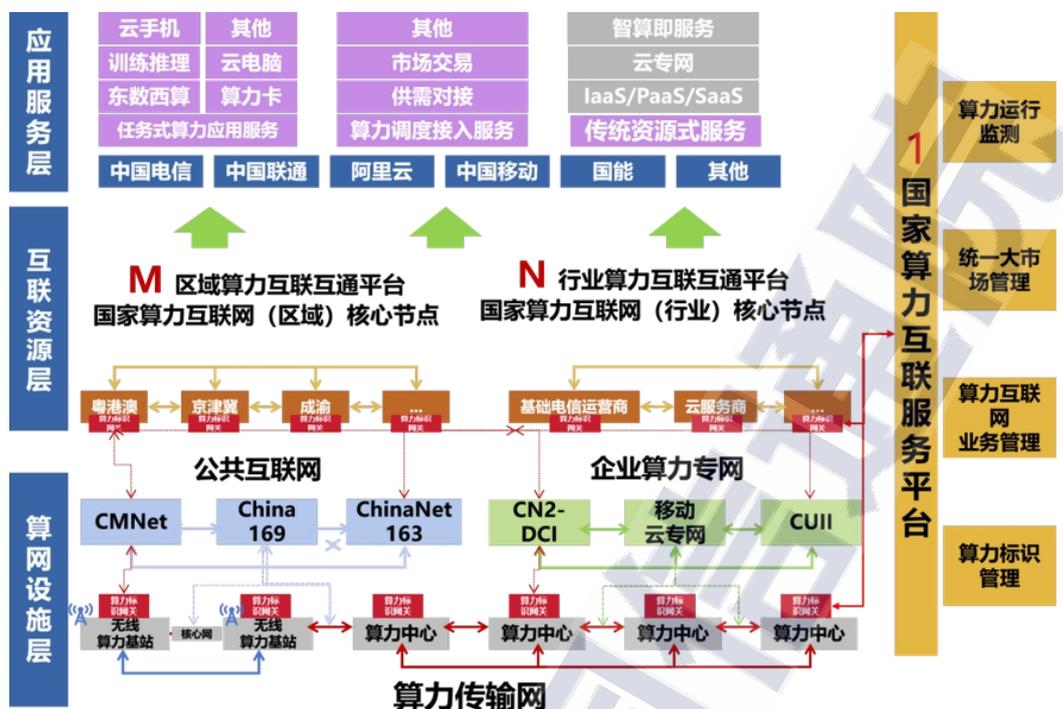


来源：中国信息通信研究院

图 11 算力互联网各节点和平台关系图

（三）应用服务层组网

应用服务层面向终端用户提供多元化算力服务，主要包括任务式算力应用服务、算力调配接入服务、传统资源式服务等业务类型。基础电信运营商、云服务企业、数据中心服务商、多云管理服务商及各类软件厂商等算力服务相关主体，可通过国家算力互联网（区域 / 行业）核心节点，获取可调度算力资源的算网参数与接口信息，并结合自身技术能力构建逻辑层面的算力资源调度入口，实现算力服务的灵活接入与按需供给。



来源：中国信息通信研究院

图 12 应用服务层与其他层关系示意图

五、算力互联网业务及生态

（一）算力互联网业务概况

算力互联网业务包括算力资源提供、算力调配接入、算力应用服务、网络传输、算力互联网五种业务。

1. 算力资源提供业务

算力资源提供业务是指利用相应的机房设施，以外包出租的方式为用户的服务器等互联网或其他网络相关设备提供放置、代理维护、系统配置及管理服务，以及提供数据库系统或服务器等设备的出租及其存储空间的出租、通信线路和出口带宽的代理租用和其他应用服务。

2. 算力调配接入业务

算力调配接入业务是指基于算力互联互通平台，整合多源异构算力资源信息，通过资源编排、任务调度等能力，为用户提供供需对接、市场交易等服务。

3. 算力应用业务

算力应用业务是指将计算、存储、网络等基础资源按任务分类或用户需求进行封装，以用户资源无感知方式提供的新型算力服务。服务以卡时、带宽、token 等任务用量为主要计量计费单位。常见的服务如人工智能应用、算力卡、云手机、云电脑等，其中人工智能应用、科学计算等服务通常以 token 计量计费，算力卡、云手机等任务式服务通常以卡时计量计费，数据快递等任务式服务以带宽用量计量计费。

4. 网络传输业务

网络传输业务是指依托骨干通信网络资源，为用户提供固定通信、移动通信、企业算力专网等基础网络连接服务。

5. 算力互联业务

算力互联业务是指基于算力互联互通平台和企业算力专网（包括虚拟企业算力云专网），提供算力资源统一调配、网络资源协同的一体化算网融合服务。

（二）算力互联网生态主体

算力互联网生态体系由多元主体协同构成，涵盖从算力资源供给、网络传输、平台支撑到应用服务的全链条参与方。主要生态主体包括：算力资源提供商、网络运营商、算力互联互通平台主体、算力调配接入服务商、算力应用服务商以及算力互联运营商等，各主体在生态中承担不同角色，共同推动算力资源的高效整合与服务化演进。

1. 算力资源提供商

算力资源提供商是指为用户提供通算、智算、超算等软硬件计算能力的主体，其核心业务包括裸机租赁、云服务等，是算力服务的“最初一公里”提供者。

主体类型：数据中心服务商、云服务商等，如提供云主机、GPU 云主机等计算服务的云服务商，提供裸机等各类硬件计算资源的数据中心服务商。

2. 算力调配接入服务商

算力调配接入服务商是指面向用户算力任务需求，整合多源异构算力资源，并通过资源编排、任务调度等能力，为用户提供算力供需对接、任务分发与执行保障服务的主体，是实现算力服务的“最后一公里”的关键角色。

主体类型：云服务商、基础电信运营商、各地调度平台运营主

体等，如提供中国电信“息壤”平台服务、中国移动“算网大脑”等调度服务的基础电信运营商，贵州、宁夏等各地算力平台运营主体。

3. 算力应用服务商

算力应用服务商是面向终端用户需求，提供基于算力资源封装的各类应用层服务的主体，服务形态包括人工智能应用、科学计算、算力卡、云手机、云电脑等。

主体类型：云服务商、基础电信运营商、软件服务商等，如提供模型推理服务的软件服务商；提供算力卡、云电脑、云手机等服务的电信运营商和云服务商。

4. 网络运营商

网络运营商是指面向用户提供基于国家骨干通信网络的传输服务的企业，是保障算力资源高效调度与低时延连接的基础支撑力量。

主体类型：基础电信运营商，如中国电信、中国移动、中国联通。

5. 算力互联运营商

算力互联运营商是指同时具备网络传输服务能力和算力互联调配能力，能够提供算网融合一体化服务的企业，是推动“算力+网络”深度融合发展的核心运营主体。当前基础电信运营商、云服务商都在探索算网融合服务，主体的边界和定义随业务形态发展调整更新。

主体类型：同时具备算、网服务能力的主体，如中国电信、中

中国移动、中国联通等基础电信运营商，以及具备云专线等虚拟企业算力云专网服务能力且建设运营算力互联互通平台的头部企业。

6. 算力互联互通平台主体

算力互联互通平台主体是指负责建设和运营算力互联互通平台的机构，通过向算力调配接入服务商开放互联调度接口，提供算力标识解析、算网参数调度、算力计费度量等基础公共服务，是实现全国算力互联一张网的核心支撑平台。

主体类型：地方企事业单位以及行业龙头企业，如各地和各行业算力互联互通平台主体。当前算力互联互通平台定位于提供公共服务。



来源：中国信息通信研究院

图 13 算力互联网体系相关主体图

六、算力互联网关键技术

（一）算网设施层技术

1. 全光网络

全光网络（All-Optical Network, AON）是指在网络传输与交换过程中全程采用光信号处理的网络架构，仅在用户接入和出口端进行光电/电光转换。其核心优势在于实现高速率、低时延、高安全性与智能化的网络传输能力。

核心技术能力包括：

（1）超大带宽技术

通过采用高阶调制、新型光纤材料及先进接入技术（如 50G-PON、100G+高速光模块），实现从接入到骨干的端到端大带宽传输，满足 AI 训练、超高清视频、科学计算等高带宽业务需求。

（2）超低时延技术

通过优化光层传输路径、减少中间电层处理环节，显著降低传输时延与抖动，支撑实时控制、远程操作等时延敏感型业务。

（3）智能管控技术

依托统一的智能管控平台，实现光网络资源的集中调度、性能优化与故障自愈，推动光网络向“可感知、可调度、可优化”的智能化方向演进。

目前，全光网络已在重点行业和区域实现初步部署。骨干网方面，100G/200G 高速光传输系统已广泛商用；接入网方面，50G-PON

标准逐步落地，部分城市已启动试点建设。此外，智能光网络（如 ASON、SDN 光网络）也在部分运营商网络中实现部署，初步具备资源调度与故障恢复能力。

2.长距 RDMA

长距 RDMA(Remote Direct Memory Access over Long Distance) 是指利用 RDMA 技术实现跨广域网络的高效数据传输机制。其核心在于通过绕过操作系统和 CPU 干预,实现远程主机内存之间的直接读写,从而显著降低数据传输的时延与系统开销,提升网络吞吐效率。该技术为跨区域数据中心互联、分布式计算、高性能存储等场景提供了底层网络支撑。

核心技术能力包括:

(1) 零复制传输技术

长距 RDMA 通过绕过操作系统内核与协议栈,实现数据在应用缓冲区之间的直接传输,消除了传统 TCP/IP 协议栈中多次内存拷贝与 CPU 中断的开销,提升了数据传输效率。

(2) 无损网络传输技术

该技术通常基于 InfiniBand 或 RoCEv2 (RDMA over Converged Ethernet version 2) 等高性能网络协议栈,要求底层网络具备拥塞控制与流量调度能力,以确保数据在长距传输过程中不丢包、不重传,保障传输的稳定性和可靠性。

当前,长距 RDMA 技术已在部分高性能计算 (HPC)、AI 训

练集群、分布式存储等场景中开始试点应用。部分头部云服务商和科研机构已部署基于 RoCEv2 的长距 RDMA 网络，实现跨数据中心的数据高效同步与计算协同。此外，部分运营商和设备厂商也在探索将长距 RDMA 与算力网络结合，推动其在跨域资源调度中的应用落地。

3.弹性网络

弹性网络（Elastic Network）是指在网络接入侧具备带宽动态调整能力的网络架构，能够根据业务需求实现端到端的带宽弹性供给与流量智能调度。其核心目标是提升网络资源利用率、保障业务服务质量（QoS），并实现网络资源的按需分配与灵活调度，是支撑算力互联网、云网协同、边缘计算等新型业务形态的重要网络能力。

核心技术能力包括：

（1）带宽弹性供给技术

弹性网络通过实时感知业务流量变化，动态调整接入与骨干网络带宽资源。在业务高峰期自动扩容，保障高带宽需求下的服务质量；在低谷期则释放冗余资源，提升资源利用效率并降低运营成本。

（2）流量多路负载技术

通过多路径传输与负载均衡技术，将业务流量合理分散至多个链路或设备上，提升网络吞吐能力和链路可靠性，避免单点故障导致的业务中断，实现网络的高可用性与智能化调度。

当前，弹性网络技术已在部分云服务商、运营商及行业专网中

初步应用。例如，部分云平台已实现基于 SD-WAN 的弹性带宽服务，支持按需带宽调整；部分基础电信运营商也在探索将弹性网络能力与算力网络调度平台结合，实现跨区域资源的动态带宽分配。此外，面向工业互联网、远程医疗等场景的弹性网络试点也在逐步推进。

（二）互联资源层技术

1. 算力标识

算力标识是算力互联网中用于唯一标识、描述和定位算力资源的基础性技术要素，是实现算力资源可寻址、可调度、可交易的关键支撑能力。其核心作用在于对异构、分布式的算力资源进行统一命名与描述，实现跨区域、跨服务商、跨层级的资源发现与调度。

核心技术能力包括：

（1）资源描述

通过标准化元数据对算力资源进行多维度描述，涵盖资源主体、地理位置、资源规模（如 CPU、GPU、内存）、服务类型（如 AI 训练、科学计算、云桌面）等关键参数，为资源匹配与调度提供数据支撑。

（2）精准寻址

算力标识支持从城市、可用区、服务器到芯片级的多级寻址机制，结合 IP 地址、软件标识、硬件地址等信息，构建“算力路径”（Computing Path），实现对最小计算单元的精准定位和访问。

（3）动态解析

算力标识可由算力互联互通平台进行解析，将用户输入的资源需求（如计算能力、服务类型、地理位置等）转换为对应的算力标识，匹配并返回符合要求的算力资源列表，支撑资源调度与交易。

目前，算力标识技术已在部分区域和行业试点中初步应用。国家算力互联网服务平台正在推动统一标识体系的构建，部分区域级、行业级算力互联互通平台也已开始探索基于标识的资源注册、解析与调度机制。

2.算力资源并网

算力资源并网是指将异属、异构、异地的算力资源通过网络连接实现可达、可用、可管、可调度的技术过程。其核心在于通过统一接口与协议实现资源的注册、纳管与封装，支撑跨区域、跨服务商、跨类型的算力资源协同调配与共享使用，是构建算力互联网体系的基础能力之一。

核心技术包括：

（1）算力注册

对拟接入的算力资源进行统一注册，验证其真实性、完整性与唯一性，确保资源信息准确、可信。注册信息包括资源类型、性能参数、地理位置、服务状态等，是后续资源调度与交易的基础数据支撑。

（2）算力纳管

根据不同资源类型与管理需求，提供多层级的纳管模式，包括

运营纳管、管理纳管、集群纳管与裸算力纳管等，实现对资源权限、状态、使用情况的统一管理与调度。

（3）算力封装

通过统一接口设计、通信协议与接入插件，将不同架构、不同平台、不同服务模式的算力资源进行标准化封装，屏蔽底层差异，对外提供统一调用接口，实现算力资源的互联互通与调度调用。

目前，算力资源并网技术已在多地平台建设中初步落地。例如，国家算力互联网平台已启动多区域、多服务商的资源并网试验，初步实现跨数据中心、跨云平台的资源注册与调度。中国算力网（China 算力网）、超算互联网等项目也在探索异构算力资源的统一接入与调度机制。此外，部分云服务商和调度平台也已具备初步的异构资源纳管与封装能力。

3.任务编排调度

任务编排调度是指在分布式、多计算节点的异构算力环境中，依据任务优先级、资源需求、实时负载、地理位置等因素，动态调配计算任务到相应算力资源，以实现利用效率最大化、任务响应最优化的综合调度技术。该技术是支撑算力互联网实现跨域、跨架构、跨主体资源协同调配的核心能力。

核心技术能力包括：

（1）多任务统一编排

面向人工智能、大模型训练、科学计算等典型智算任务，以及

通用计算、超算任务，实现任务、资源、数据的协同编排，提升资源利用率，合理匹配任务负载与资源能力，保障关键业务服务质量（QoS）。

（2）多类资源调度控制

构建统一的调度控制组件，支持智算、通算、超算等多类异构资源的统一调度，兼容多种开发架构与操作系统环境。具备跨网络、跨服务器、跨芯片的数据传输与任务调度能力，实现卡间、节点间、跨域间的数据协同与任务部署。

（3）多样化调度策略

基于不同应用场景与资源特性，提供丰富的调度策略，包括 GPU 独占与共享调度、卡间传输监测调度、节点与应用资源感知调度、资源预留调度等，支持在复杂环境下实现精细化资源调度与动态优化。

（三）应用服务层技术

1. 任务定义

任务定义是算力互联网中对计算任务进行标准化描述与封装的基础能力，旨在通过结构化配置文件对任务的主体信息、资源需求、执行环境、调度偏好等进行统一定义，为后续任务调度、资源匹配与执行提供数据支撑。任务通常以应用进程的形式进行封装，包含任务代码、运行环境、依赖库等要素，并被赋予唯一任务 ID，作为调度的最小单元。

（1）结构化任务描述

基于算力标识体系，任务定义通过配置文件对任务的主体、位置、所需资源类型（CPU/GPU/内存）、芯片类型、调度偏好等关键属性进行标准化描述，形成统一的任务元数据模型。

（2）任务 ID 与调度映射

每个任务定义对应唯一的任务 ID，并通过调度平台与资源管理模块进行映射，实现任务的注册、调度、监控与追踪，支撑端到端的调度闭环管理。

2. 算力度量

应用服务层应具备面向服务使用侧的算力动态衡量技术，在传统算力度量基础上增强对经济性、服务性、先进性、一体性的特性考量，围绕新型算力互联网架构特征，提供综合考虑算力设施、算力资源、算力应用服务、互联成网核心能力与层级支撑关系的四级度量模型，分别从成本、能力、效益三维立体角度进行算力度量，根据面向对象、场景不同进行模型调参。

七、算力互联网建设进展

（一）算网设施层：算网统筹协同初见成效

算力设施统筹布局。在全国一体化算力网建设、“东数西算”战略等大背景下，新建智能算力中心逐渐融入八大枢纽节点建设。截至 2025 年一季度，我国“东数西算”八大枢纽节点算力总规模达

到 215.5EFlops，智能算力规模占枢纽节点算力规模的 80.8%，枢纽节点间 20 毫秒时延圈已基本实现⁴。

网络设施迭代升级。算力传输网关键技术升级，400G 技术体系基本完善，运营商逐步启动干线场景规模部署，对于更高速率的传输技术，800G/1.6T 技术标准正在加紧研制。此外，G.654.E 光纤、空芯光纤等新型传输媒介加快应用步伐，百公里级 RDMA 技术在存算拉远场景中得到验证，都为构建高性能智算网络提供有力技术支撑。

（二）互联资源层：多级平台体系初步建成

建成国家算力互联网服务平台。平台具备算力标识查询管理、算力供需对接、算力市场服务、数据快递专线等服务功能。截止 2025 年 12 月，基于算力标识网关，已采集获取包括全量算力标识及可用算力标识在内共 102w 条标识，接入 136 家算力提供商，共 117EFLOPS 的算力资源。

建设重点区域算力互联互通平台。截至目前，京津冀、长三角、粤港澳、成渝等算力产业发展重点区域基于算力互联网体系架构开展公共算力标准化接入汇聚、跨服务商算力交易居间结算、第三方算力调度平台适配接入、跨域互联协同等技术验证，探索新技术催生的新形态算力服务业务发展。

建设通信行业算力互联互通平台。三大运营商启动算力互联网试验网建设，中国电信推动智算互联网试验网建设，中国移动依托

⁴ 数据来源：国家数据局

算力网络试验示范网（CFITI）开展试验网工作，中国联通启动算力互联网试验网建设。运营商基于算力互联网标准体系，建设试验网、试验新型算力业务，协同开展新技术研发探索，提高算力互联网服务能力和性能，形成覆盖全国的算力互联网。截止目前，三大运营商试验网已接入社会算力超 60 家，总算力规模超 60EFlops，累计服务用户超 700 万，总服务卡时超 1600 万。

（三）应用服务层：多元算力服务初显活力

任务式算力服务市场崛起。云电脑基于用户地理位置与网络状态，智能调度计算任务至延迟最优的算力节点，实现高效、低延时的服务响应，促进算力服务的泛在化与均等化。云游戏依托云端 GPU 集群完成游戏渲染，并根据实时负载动态分配至适宜的算力节点，通过视频流方式回传至终端，彻底摆脱对本地高性能硬件的依赖。机构数据显示，2024 年全球云游戏市场规模已达 97.1 亿美元。未来该市场预计将从 2025 年的 157.4 亿美元稳步扩张至 2032 年的 1217.7 亿美元⁵。

调度接入服务市场快速发展。算力调度平台通过智能编排算法，将用户任务动态匹配至成本最优或性能最优的异构算力资源，实现跨域、跨服务商的资源一体化供给，正成为释放算力价值的关键枢纽。除运营商、云服务商外，八大枢纽、地方政府、产业园区等均启动建设算力调度相关平台，探索调度接入的商业运营，如贵州算

⁵ 数据来源：Fortune Business Insights 《云游戏市场规模、股票和行业分析》

力调度平台，已整合全省算力资源，实现跨区域、跨平台的高效调配，2024 年算力年交易额突破 100 亿元⁶。

八、算力互联网发展建议

我国算力互联网建设尚处于起步阶段，面向未来发展，应遵循“先试点后推广、先互联再成网”的总体思路，以设施互联、资源互用、业务互通为着力点，以标准规则为基础、应用场景为牵引、安全保障为支撑，稳步推进算力互联网建设。

（一）推进设施互联，构建高速传输网络

设施互联是算力互联网建设的物理基础，主要指智算、通算等异构算力设施之间通过高性能网络实现的底层连接。我国算网设施取得显著发展，算力基础设施布局持续优化，智能计算中心、边缘计算节点加速部署，算力与网络融合不断深化，初步形成通算、智算、超算协同发展的多元算力体系，为算力互联提供了坚实支撑。

为进一步优化算力节点间互联协同能力，增强数据传输效率与网络服务质量，需从以下两个方面推进：一是提升算力节点内互联能力。加快新型高速互联总线协议的设计开发与应用推广，推广 RDMA 等高性能传输协议在广域场景中的部署应用。二是强化网络间互联能力。升级国家级互联网骨干直联点和新型互联网交换中心，增强运营商算力专网能力，完善网间质量监测体系，提升协同调度水平。

⁶ 数据来源：贵州省大数据发展管理局

（二）强化资源互用，实现高效供需匹配

资源互用是设施之上计算等逻辑资源之间的互联，包括通算、智算、超算等算力资源之间的共享互用，实现跨主体、跨架构、跨地域“算力资源池”的智能感知、实时发现和按需使用的供需匹配能力。当前，我国算力主体超 7000 家，其技术体系、基础架构、调用接口等存在差异，没有形成全国范围内可感知调度的算力服务，用户难以快速找到位置、成本、性能均合适的算力资源，供需尚未实现高效对接。

为统一汇聚公共算力，实现算力企业资源的高效互联和有效利用，需重点关注四方面：一是建立统一的算力标识体系。制定算力标识描述规范、算力标识分配和使用规范。二是提升接口互操作能力。利用算力标识网关等，推动不同算力主体使用统一的调用接口和通信协议。三是建设多级算力互联互通平台。建设国家算力互联网服务平台，建设区域、行业算力互联互通平台。四是保障算力互联互通平稳运行。开展算力设施和服务的运行、安全、性能和质量等监测。

（三）推动业务互通，加速融合贯通调度

业务互通是算力互联网实现应用服务层融合的核心能力，包括计算、存储、网络等业务的协同调度。目前，运营商、云服务商等通过云计算操作系统、算网大脑等技术产品，实现计算、数据和网络的融合应用。

为推动算、存、网多种业务互通，实现跨主体、跨架构、跨地域算力调配，需重点关注三方面：一是提升应用调度互通能力。构建算子库与开发框架体系，创新算力资源检索方式和算力调配系统。二是提升数据与存储互通能力。推动全局文件系统等存储技术应用，实现跨平台、跨区域的数据高效流动与存储共享。建立统一的数据接口规范与访问机制，提升数据在不同算力资源间的可迁移性与可调度性。三是提升算网融合能力。基于云原生、AI 技术升级网络，推进算网融合管理平台建设。

（四）培育场景赋能，提升普惠便捷体验

当前，在企业级应用领域，针对大模型训练、仿真计算等场景，算力互联互通可实现数据高效流动计算。在消费级应用领域，对于云电脑、智能体调度等应用，用户可以像按“千瓦时”用电一样，按“卡时”灵活购买算力资源，满足使用人工智能的需求。

为进一步深化算力互联网创新与场景应用，需重点关注两方面：一是探索构建算力互联网体系。加速公共算力资源互联成网，形成智能感知、实时发现、按需获取的算力互联网。二是赋能典型应用场景。推动算力互联在新业态、企业级、消费级场景的应用，以及与多种行业互联网的融合创新。

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62300072

传真：010-62304980

网址：www.caict.ac.cn

