

建设新型城市算力中心 是无锡发展数字经济的必要实施路径

□大数据集团 黄乐飞

随着新一轮科技革命和产业变革深入推进,数字经济蓬勃发展,数字技术激发的新模式、新业态、新理念不断涌现,正深刻改变着经济社会各领域。在此进程中,算力服务作为数字技术能力的主要输出方式之一,以多样性算力资源为基础,以算力网络为连接,日益成为支撑数字经济发展的关键。算力俨然已成为集信息计算力、网络运载力、数据存储力于一体的新型生产力。而随着大模型算法推动人工智能向更通用方向发展,我国算力需求,尤其是智能算力需求激增,以AIGC为代表的人工智能应用、大模型训练等新应用、新需求快速崛起对算力提出更高要求。

算力是以计算能力为核心,包含算力规模、经济效益和供需情况在内的综合能力。在数字革命背景下,算力是决胜信息时代的关键实力。算力作为数字经济时代的关键生产力要素,已经成为挖掘数据要素价值,推动数字经济发展的核心支撑力和驱动力。

国内算力发展现状

(一)我国算力运营产业步入快速发展的黄金阶段

数据显示,在算力产业中每投入1元,平均将带动3—4元经济产出。鉴于算力产业的巨大潜力,近年来,无论是国家还是各类市场主体,都在大力推进算力资源布局。算力供给方面,2022年我国计算设备算力总规模达到302EFlops,位居全球第二(占比约为33%),连续两年增速超过50%,高于全球增速。算力需求方面,据OpenAI测算,从2012年开始,全球AI训练所用的计算量呈指数级增长,平均3.43个月翻一倍,目前计算量已扩大30万倍,远超算力增长速度。截至2023年7月,中国已发布79个10亿参数规模以上的大模型,仅2023年1月至7月就有64个大模型发布,新一轮“大炼模型”和“炼大模型”热潮,同时硬件供应困难,多重因素叠加导致国内“算力荒”。算力运营行业作为算力资源供需失衡现状下的商业模式创新,通过算力全面统筹与合理配置实现算力资源的“区域一盘棋”“全国一盘棋”,能够有效缓解算力资源的供需矛盾问题,正步入快速发展的黄金阶段。

(二)各地积极探索入局算力运营市场

当前,国家及各级政府相继制定了一系列政策支持算力网络的发展,在需求与政策的双重驱动下,各地政府以及各类企业主体纷纷布局算力运营业务。2023年2月,国内首个一体化算力交易调度平台——东数西算一体化算力服务平台正式上线;4月,全国首个算力交易平台——上海市算力交易平台启动试运行;8月,苏州市公共算力服务平台上线运营;10月,浙江省一体化算力

服务平台正式发布。2023年或成为算力服务平台建设元年。今年以来,各地积极起步建设算力服务平台,助力算力供需匹配,积极探索算力交易与服务一体化模式,竞速算力运营赛道。

无锡算力发展优势及挑战

无锡算力服务发展先天土壤与后发优势并举。在硬件能力方面,早在2006年,无锡已成立国家级超级计算中心,作为全球最大的超级计算机中心之一,拥有完全自主可控的申威通用算力芯片并基于申威芯片研发出太初智算算力芯片,由全国产系统“申威26010”处理器构建,峰值性能超过125PFlops。依托国家超算中心建设,无锡聚力发展自研芯片,大力发展申威、太初等一众本土芯片自研企业,具备股实的硬件配套能力。在算力服务方面,无锡算力公共服务平台积极联合太初团队,以及阿里云、曙光智算等国内头部云厂商、先进计算厂商开展战略合作,力求打造国内领先的算力运营服务平台。同时,澳江等无锡本土企业将有效助力绿色低碳算力供给。

加快新型城市算力中心建设,是当前政策背景下推动无锡数字经济快速发展、加速产业数字化转型的必行之举,但仍存在供需均衡、市场化模式等诸多方面问题。

一是算力芯片短期供不应求,长期风险巨大。受地缘政治因素影响,美国持续升级对华芯片出口管制,英伟达A100和H100已被禁止向中国、中东供货。今年10月,美国持续收紧尖端AI芯片对华出口,中国特供版A800、H800禁售在即,中国算力市场,尤其是智算市场随之受到威胁。当前英伟达A800芯片供不应求,价格飙升至10多万元/片。短期内,算力芯片供不应求,价格暴涨,不利于国内算力基础设施的建设布局。长期而言,国际主流AI芯片的缺位将对我国算力资源均衡供给、自主可控带来巨大风险。

二是算力市场供需结构失衡。在“新基建”的大背景下,国家启动实施“东数西算”工程,助力我国全面推进算力基础设施化,以满足高校科研实验室、企业和研究所高速增长的计算需求。然而,国家超算中心主要为国家科研院所、高等院校提供超算算力资源服务,计算项目排队情况普遍严重,其中深圳超算的计算资源利用率已经严重饱和,资源需求是现有资源的140%以上。随着云游戏、云电竞、元宇宙、智慧城市、虚拟现实和自动驾驶等需要实时决策、快速响应的数字场景高速发展,以图形计算和AI为核心的智能算力需求也呈指数级增长。

三是算力运营与服务面临诸多技术门槛。算力运营服务在软硬件

技术、人才、需求理解、生态网络等多方面具备较高技术门槛,核心技术涵盖算力资源接入与池化技术、智能调度技术、算力集群运行数据采集与分析技术、应用全生命周期监控与分析技术、应用软件SaaS化平台技术等众多领域。与此同时,随着人工智能模型训练推理等应用增加,多样化算力协同成为常态,高功率机柜占比提升明显,单IT机柜主流功率密度将从6—8KW/柜提高到12—15KW/柜,超算、智算中心功率密度将提升至30kW以上。算力中心高功率机柜建设面临着电力供应、散热管理、设备运行安全性等诸多难点问题。此外,国内自主可控的液冷技术、算卡组网技术,都对算力资源的供给提出挑战。

四是算力产品矩阵、商业模式探索难度大。算力中心的产品矩阵普遍较为单一,绝大多数以直接提供计算资源,只是重视算力供给量的提升,而忽视了算力供给质量、品类和用途,且缺乏用户侧的需求开发,导致智能算力、智能应用、大模型、解决方案等与智算算力产品相关联的市场资源缺乏,无法弥补用户需求和市场供给间的鸿沟。此外,算力中心商业模式普遍较为单一,以按核时/机时向用户收费为主,虽然部分算力中心开始尝试计算资源租赁、超算云服务作为业务补充与创新,相比阿里云、华为云等云服务商,云服务的配套服务、开发工具还不够成熟,整体的服务履约能力有限。

无锡算力公共服务平台上线

2023年10月,国家六部门联合印发《算力基础设施高质量发展行动计划》,明确“构建通用、智能和超级算力协同发展的供给体系”“推动以云服务方式整合算力资源,充分发挥云计算资源弹性调度优势”。10月21日,2023世界物联网博览会上,“无锡算力公共服务平台”正式上线。无锡数字经济时代全面开启,平台将以算力驱动数字变革,为无锡数字经济创新发展、城市全面数字化转型注入新动能。

无锡建设高性能算力中心,打造城市“算力洼地”,对发展城市新经济意义重大,当前企业面对计算需求,其核心关注的并不是单一芯片或者单一服务器的性能,而是如何获取更低的计算成本、更高的计算算力,来满足自身的计算需求。“算力洼地”的建设,将鼓励中小企业敢于将算力作为先进生产力去积极创新,构建自身数字化竞争力。

建设新型城市算力中心的关键要素

新型城市算力中心的建设不能简单地只依靠政府去推动,而是要靠

技术的创新来打造算力洼地,来实现性能及效能上的突破。从雪浪小镇诞生之日起,无锡就是思想策源地、产业新跑道、资本新天地。无锡物联网产业的发展根植于制造业,而80%的数据来自制造业,物联网产业带动算力需求增长,这是产业发展的必然结果,同时也要求无锡必须要在算力建设上做好提前布局来支撑产业发展需求。

首先要充分发挥本土资源能力,用好无锡“芯”,打造无锡算力公共服务平台“硬核”。支持和应用无锡本土太初等国产算力芯片、AI芯片,打造“一云多芯”的新型城市算力中心,发展算力核心技术,扶持带动无锡先进算力芯片产业发展,提高算力核心软硬件的自主研发与安全可控,提升无锡在算力领域的领先地位和技术创新能力。

同时要做好存量资源统筹与调度,打造无锡“算力资源一张网”。依托算力公共服务平台发挥算力资源“虹吸效应”,汇聚无锡市区两级、不同类型算力资源,通过整合本土自主算力资源,同时汇聚运营商、云厂商算力资源,突破算力资源地域局限,丰富多元算力供给和服务体系。通过建立统一的算力调度平台,将算力供应方和需求方紧密连接,实现算力资源高速互联和无缝对接,促进算力供需的对接交流与精准匹配,再通过实时监测算力资源的负载情况,动态灵活地调整算力资源的供给分配。

此外还要将产品服务快速实现结构化,打造无锡“算力产品特色矩阵”。算力服务的终端用户覆盖航空航天、石油勘探、智能制造、地球环境、生命科学、人工智能等各领域,存在较强的专业化需求。算力公共平台只有深入调研无锡本地用户群体、理解应用场景特征、针对性地进行产品研发并提供差异化的解决方案,才能真正地满足用户需求,并在服务客户的过程中不断总结行业经验和成功案例。无锡算力公共平台将打造集算力基础服务、软件应用服务、大模型服务、社区交流服务于一体的“算力产品特色矩阵”,拓展客户资源、提升用户黏性。

最后还需专业运营团队提供支撑,打造“海量用户+应用平台+优质算力”生态网络。通过组建专业的运营推广团队,建立成熟的市场化销售通路,以“算力+算法”专业服务深度挖掘、引导激发本地企业算力使用需求,开发在生物医药、车联网、高端纺织服装等领域的本地特色产品与服务,强化算力在传统产业数字化转型和重点行业领域中的推广应用。逐步建立起基于算力公共服务平台的“海量用户+应用平台+优质算力”的算力生态网络,开展全域、全场景、全链路和全周期的运营推广,打响无锡算力公共服务平台品牌,持续提升知名度与影响力。