大数据与决策研究

2024 年第 71 期(总第 291 期)

广西壮族自治区信息中心广西壮族自治区大数据研究院

2024年12月11日

广西算力调度关键问题和实施路径研究

在国家实施"东数西算"战略背景下,工信部颁布《算力基础设施高质量发展行动计划》,强调了探索算力协同调度机制的重要性。广西算力调度仍面临调度匹配失衡、运载能力有限以及跨域调度难等挑战,需结合自身特点,以建设一体化算力调度平台为导向,统筹全区算力资源,提升算力运载能力,拓展东盟算力市场,推动广西算力调度高效发展。

一、广西算力调度稳步发展

(一)网络设施建设不断完善

广西推进 5G 网络建设,建成开通互联网骨干直联点,为算力调度打通路径。从 5G 网络建设看,截至 2024 年 11月,全区已累计建成 13 万个 5G 基站¹。2024 年第 3 季度5 西 5G 网络下行、上行中值接入速率分别达到 135.23Mbps、56.12Mbps²,同比去年同期提升幅度为 48.51%、6.19%。从网络服务能力看,全区建成 89.3 万个具备千水网络服务能力的 10G PON 端口,有力提升网络传输能力。从网络时延看,自南宁国家级互联网骨干直联点开通运行以来,区内各电信企业区内网间平均时延从 50ms 降至 5ms 以下,降幅近 90%。

(二) 算力设施数量不断增多

近年来,广西加快布局数据、智算中心等算力资源设施,为算力调度打造资源池。从算力规模情况看,截至 2023 年底,广西已建成数据中心 38 个,机架总数突破 4.4 万架,全区算力总规模跃升至 668 PFlops³。从智算设施能力看,广西五象云谷云计算中心一期完成 2500P 算力服务器部署,中国公东盟人工智能计算中心拥有 42P 训练算力及 1.4P 推理算力,依托华为昇腾全栈人工智能技术赋能区内外 51 家⁴高校、科研机构及企业算力服务。

¹来源:广西壮族自治区通信管理局,《2024年三季度广西信息通信业发展情况通报》

² 来源: 测速网,《测速网 2024 年第 3 季度全国网速排行榜》

³ 来源: 广西壮族自治区通信管理局,《广西互联网发展报告(2023)》

⁴ 来源: 南宁日报,《人工智能+智改数转 蓄势腾飞添"新"力》

(三)绿色应用程度不断加深

广西积极践行算力绿色发展理念,深挖节能降耗潜力,为算力调度筑牢绿色根基。从绿色数据中心建设来看,全区首个新能源数据中心在广西来宾开工建设,采用冷板式液冷技术和水资源梯级利用系统,综合能耗达到国家 1 级能效数据中心标准。从节能技术改造来看,河池、百色等地市的数据中心标准。从节能技术改造来看,河池、百色等地市的数据中心机房智能温控改造,改造后两地数据中心机房平均 PUE 分别从 1.62 降至 1.46、2.0 降至 1.3521,预计每年可减少用电 50 多万度5、451 万度6。

(四)跨境网络专线不断建设

广西积极推进海外关键网络节点运营建设,持续探索算力跨境流动新机制。从数字互联通道看,持续推进《中国一东盟信息港建设实施方案(2022—2025年)》政策落实,现已建成3条国际互联网数据专用通道、1个南宁区域性国际通信业务出入均局、中国一东盟工业互联网标识解析节点等,进一步提升了广西与东盟数字互联互通水平7。从境外网络节点看,建成面向东盟的F根镜像节点和国家域名顶级节点8、并帮助阿里、润建等通信服务商在老挝、柬埔寨等9个积盟国家落地了38个数据中心9,为东盟国家提供高速访问国际互联网服务。从出口专线看,承接东盟国家超300条通

⁵ 来源,河池日报,《优化升级改造 实现节能降耗》

⁶ 来源,中国新闻网,《百色移动建成绿色低碳数据中心 提前完成节能降耗任务》

⁷来源,广西壮族自治区人民政府办公厅,《中国一东盟信息港算力网络建设行动计划(2022-2025年)》

⁸ 来源,广西日报,《广西 F 根镜像节点和国家域名顶级节点建成运营》

⁹ 来源:新华网,《数字丝路助推中国一东盟共享发展新机遇》

信专线,提供面向东盟超 400G 出口带宽¹⁰,为国内金融、互联网、制造等行业提供出境专线。

二、广西算力调度面临三大挑战

(一)调度供需匹配失衡

一是通用算力利用率不高,广西数据中心平均上架率10 约为 35% 12 ,远低于全国平均上架率 58%,存在大量未被充分利用的通用算力资源。二是高端算力不足,广西算力设施以数据中心占多数,超算中心、智算中心等设施占少数,目前全区仅有 2 家智算中心面向公众运营,无法满足人工智能行业对高端算力的需求。三是算力资源集中与算力需求分散,南宁拥有全区 42.5% 算力机架规模和 32% 算力设施,算力需求却广泛分布在南宁、柳州、桂林、钦州、北海以及玉林等市13,造成"算力空闲"和"算力紧张"现象并存。

(二) 跨算力载体调度难

一是技术标准差异掣肘,广西算力设施涵盖了多种架构和技术体系、传统数据中心多基于 X86 服务器架构,侧重于通用数据存储与处理,而智算中心围绕华为昇腾、英伟达等芯片搭建,侧重于 AI 训练和推理,技术差异难以保障资源 世配。二是管理运营协同壁垒,算力设施归属多方主体,有政府直属数据中心、企业自建机房、运营商托管设施等,政务部门侧重数据安全合格,企业看重成本与业务灵活响应,

¹⁰ 来源: 广西日报,《提升科技创新 打造"数字丝路"》

¹¹ 上架率: 以实际运行服务器数量与机架理论上可承载服务器数量之间的比例,衡量资源利用率重要指标

¹² 来源:中国信通院,《中国综合算力评价白皮书(2023年)》

¹³ 来源:广西壮族自治区信息中心,《广西人工智能产业发展白皮书(2024年)》

不同运营方资源独立,难以保障连贯调度。三是缺乏统一调度管控平台,广西未对全区算力资源进行标准化度量,尚无可跨域、跨省、跨境的一体化算力调度平台,无法实现广西本地算力"走出去"以及省外算力"引进来",算力资源不能在更大范围内流通。

(三) 算力运载能力有限

一是网络带宽受限,2024年上半年广西互联网省际出口带宽为5.27T,与浙江省际出口带宽44.09T、海南28.29T等相比较,出口带宽相对较窄,在处理大规模数据传输和算力调度时易出现拥堵。二是网络延迟难达标,广西虽建成覆盖全区"区域3ms、区内5ms时延"的通信底座,但对时延要求极高的自动驾驶、远程医疗等算力应用场景,难以提供理想的网络条件。三是算力设施互联程度差,分布在14个地市的算力设施之间缺乏高速、可靠的专用网络连接,无法形成一个有机的整体算力资源池。

三、举措建议

(大) 搭建一体化算力调度平台

一是建立算力统一交易运营门户。引入区块链、AI等技术构建算力运营门户,为算力资源提供方和算力资源需求方搭建交易服务的桥梁,满足算力交易多方参与者的业务需求。二是健全算力调度平台运行规则。制定完善的算力调度策略,综合考量业务优先级、算力成本、网络延迟等多因素,设计科学合理的调度模型,对算力资源进行精准调配。三是

推动算力调度平台生态构建。加强与各类算力需求方的合作对接,如科研机构、大型企业等,拓展平台的应用场景与服务领域。依托平台培育算力相关产业链,吸引软件开发、数据服务等企业围绕平台开展业务创新,形成良好的算力产业生态循环。

(二)完善算力供给体系

一是优化算力载体布局和建设。加快新型数据中心一体化建设,按照全区"一核多点"一体化空间布局,加快各类超级计算中心、智能计算中心、边缘计算中心等新型算力设施建设,发挥临近粤港澳大湾区地理优势,依托广西数据中心集群为其提供算力服务,承接算力业务转移,促进区域间算力资源的高效利用。二是推动算力标准体系建设。建立完整算力标准体系,对广西现有数据中心、智算中心等算力资源进行全面梳理,衡量不同算力资源的性能水平,为算力调度提供依据。至是推动云边端协同部署。加快边缘算力建设以支撑多行业低时延业务,依托通信运营商实现边缘数据中心与新型数据中心网络互联,利用 OTN 优势为重点算力应用提供高品质专线服务。

(三)提升算力运载能力

一是提升网络支撑能力。积极构建高速互联的骨干网络通道和建设确定性算力调度网络,提升算力接入网络的可靠性,加快全光网络、5G、SRv6等先进网络技术推广应用,实现网络传输智能高效化。二是加强算力设施互联。统筹合

规建设城市级数据中心算力网,推动在南宁、北海、桂林等地市建立算力网络骨干节点,按需在重点数据中心聚集区内,结合边缘算力区域化部署,建立算力网络二级节点,联通全区主要算力资源。

(四)构建东盟算力服务体系

一是提升跨境网络互联水平,加强连接东盟通信通道。 持续提高国际与省级出口带宽,实现与国内数据的高速传输,扩容国际互联网数据专用通道,发挥南字国际通信出入口局优势,从网络层面打通与东盟的算力服务通道。二是推动算力资源服务化,加速数据流通融合。依托中国一东盟信息港,开拓东盟国家的算力市场业务,以云服务方式提供跨区域的算力资源,为东盟国家提供低成本、广覆盖、可靠安全的算力服务。三是深化大数据应用创新,以业务需求拓展调度市场。支持基于东盟小语种的自然语言处理、语音智能问答、多语言机器翻译等大数据应用开发,为算力调度走进东盟提供通用业务支撑。

(执笔人: 算力智能调度研究课题组)

编辑部地址: 联系电产

XX 址: http://gxxxzx.gxzf.gov.cn/



扫描二维码获取 更多决策参考信息