

央国企智算创新实践报告

(2025 年)

中国信息通信研究院云计算与大数据研究所

· 中石油（北京）数智研究院有限公司

· 中国移动云能力中心

2025年9月

版权声明

本报告版权属于中国信息通信研究院、中石油（北京）数智研究院有限公司和中国移动云能力中心，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、中石油（北京）数智研究院有限公司和中国移动云能力中心”。违反上述声明者，编者将追究其相关法律责任。

前 言

数字经济时代，算力已成为关键生产力，重塑着经济社会发展格局。伴随人工智能技术迭代加速，尤其是生成式人工智能的突破性发展，智能算力需求呈现爆发式增长态势。作为智能算力的关键载体，算力中心对推动科技创新、产业升级以及提升国家核心竞争力具有战略性、基础性作用。

我国政府持续强化算力基础设施顶层设计，出台了系列政策措施，积极引导和支持智算产业发展。从《新一代人工智能发展规划》到“东数西算”工程，再到《算力基础设施高质量发展行动计划》，从战略布局、资源整合、技术创新等多维度精准施策，为产业发展营造良好的政策环境。在国家政策的引领下，各地方政府因地制宜出台配套措施，形成央地协同推进智算产业发展的工作格局。

央国企积极响应国家号召，凭借自身资源、技术和市场优势，成为智算产业建设和应用的主力军。通过合理布局智算中心，统筹整合内外部资源，提升企业数字化转型能力，推动产业升级和创新发展。同时，在智算平台建设过程中，央国企积极探索创新，构建先进的异构算力资源管理调度体系，既满足企业内部智能化应用需求，又形成可复制、可推广的典型经验和发展模式，为行业发展提供示范引领。

本报告立足于央国企智算领域的创新实践现状，系统梳理智算中心战略布局与平台建设路径，深入挖掘央国企在智算中心布局规划、平台架构设计等核心环节的实践方案。通过提炼典型案例的成功经验与建设成果，为行业提供系统性参考。

时间仓促，报告仍有诸多不足，恳请各界批评指正。后续我们将不断更新完善，如有意见建议请联系中国信通院研究团队：
dceco@caict.ac.cn。



目 录

一、 我国智算中心发展背景.....	1
(一) 政策驱动，加速智算产业高质发展.....	1
(二) 需求激增，推动智算中心加速落地.....	2
(三) 技术革新，激发智算行业创新活力.....	3
(四) 规模扩张，推动智算产业赋能升级.....	4
二、 智算中心规划与建设.....	6
(一) 政策视角：国家智算中心布局与现状.....	6
(二) 行业视角：央国企智算布局与现状.....	7
(三) 企业视角：央国企智算中心布局思路.....	10
(四) 智算中心建设模式.....	12
1. 方式选择.....	13
2. 选址布局.....	15
3. 建设规划.....	16
4. 建设施工.....	16
5. 运营管理.....	17
三、 智算平台建设.....	18
(一) 建设意义.....	18
(二) 技术方案.....	19
1. 智算平台总体架构.....	19
2. 智算平台关键技术.....	22
3. 业界平台建设案例.....	25
(三) 央国企智算平台建设模式.....	26
1. 租用建设模式.....	28
2. 自研建设模式.....	29
3. 共建建设模式.....	31
四、 典型实践案例.....	32
(一) 中国石油昆仑大模型及算力建设.....	32
(二) 中国移动震泽 MaaS 平台	32

（三） 中国建行“建行云”数字基础设施建设	33
（四） 深圳智城翼云科技智慧城市算力统筹调度平台	34
（五） 珠海云上智城国产智算及其调度应用	35
五、 发展建议	37
（一） 明确发展路径，有序推动智算中心建设	37
（二） 优化资源调配，强化智算平台技术创新	37
（三） 深化防护措施，完善智算安全运营体系	38
（四） 强化综合施策，推动绿色低碳集约发展	39

图 目 录

图 1 智算中心建设模式	13
图 2 智算平台总体架构	21
图 3 自研智算平台建设示意图	30

表 目 录

表 1 智算中心建设模式	14
表 2 智算平台建设模式	27

一、我国智算中心发展背景

（一）政策驱动，加速智算产业高质量发展

我国高度重视算力基础设施建设，持续加强全国一体化算力中心协同创新体系顶层设计，不断加大政策引导力度。智能计算中心（以下简称“智算中心”）建设作为其中的关键一环，成为推动智算产业发展、提升国家竞争力的重要动力。2017 年 7 月，国务院发布《新一代人工智能发展规划》，智算中心概念首次被提出。此后，相关部门积极响应，于 2021 年 5 月印发《全国一体化大数据中心协同创新体系算力枢纽实施方案》、2021 年 7 月印发《新型数据中心发展三年行动计划（2021-2023 年）》、2023 年 10 月印发《算力基础设施高质量发展行动计划》、2023 年 12 月印发《深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》、2025 年 2 月印发《关于组织开展算力强基揭榜行动的通知》以及 2025 年 8 月印发《国务院关于深入实施“人工智能+”行动的意见》等一系列政策，为夯实智算基础设施建设提供明确指导，助力智算产业加速发展。

各地政府积极响应国家政策，依据本地特色与优势，出台了系列政策，促进智算产业发展。2023 年 5 月，北京发布《北京市促进通用人工智能创新发展的若干措施》，将新增算力建设项目纳入算力合作伙伴计划，加快推动智算中心建设，形成规模化先进算力供给。2023 年 10 月，内蒙古发布《内蒙古自治区推动数字经济高质量发展工作方案（2023—2025 年）》，提出发展智算产业，加快推进和林格尔新区“多云”算力资源调度示范项目建设，引导算力企业向智算平台转型

发展。2023 年 11 月，山东发布《山东省数字基础设施建设行动方案（2024—2025 年）》，提出要引导通用数据中心、超算中心、智算中心、边缘数据中心等合理梯次布局。2024 年 3 月，上海发布《上海市智能算力基础设施高质量发展“算力浦江”智算行动实施方案（2024-2025 年）》，提出聚焦智能芯片、算法模型、训练框架、计算平台与重点行业应用，逐步构建以自主可控、安全可靠、开源开放、广泛兼容、高性能算力聚合为特征的智能算力基础设施产业体系。2024 年 4 月，江苏发布《江苏省算力基础设施发展专项规划》，计划构建新一代云计算、智算、超算、边缘计算等多元算力供给体系。

（二）需求激增，推动智算中心加速落地

人工智能技术的快速发展和应用，成为推动智能算力需求增长的关键驱动力。AI 大模型创新演进对智能算力产生较大需求，推动大规模智算集群的增长。《智算产业发展研究报告（2024）》¹显示，与语言大模型（GPT-3）相比，Sora 训练阶段的算力需求是大语言模型 170+倍。在推理阶段，算力需求是大语言模型 600+倍。2025 年 2 月，马斯克旗下人工智能公司 xAI 发布了其最新一代大模型 Grok 3，Grok 3 使用了拥有约 20 万张英伟达 H100 GPU 的大型智算中心进行训练，其计算能力是上一代版本 Grok 2 的 10 倍。与此同时，以 DeepSeek 为代表的轻量化模型涌现，降低了算力使用成本，让更多用户高效快捷的使用 AI 服务，激发了更大的智能算力需求。此外，随着市场竞

¹ 中国电信研究院，《智算产业发展研究报告（2024）》。

争的日益加剧，各行业企业积极利用大模型等人工智能技术重塑商业形态和业务模式，推动智能化场景构建，促使智能算力需求持续增长。

为提升高性能智算资源供给、完善算力服务，满足人工智能应用发展需求，各地正加速推进智算中心建设落地。国家发展和改革委员会公布的信息显示，“东数西算”工程的 8 个国家算力枢纽节点建设已全部开工，标志着工程已从系统布局进入全面建设阶段。在区域实践层面，各地紧扣“枢纽-集群-中心”三级架构，形成差异化发展格局。例如四川成都天府智算西南智算中心已在投运，聚焦多模态大模型训练，构建“云-边-端”协同的西南智算枢纽，支撑成渝地区数字经济核心产业发展；中国移动在北京建成大规模训推一体智算中心，支撑高复杂度、高计算需求的百亿、千亿级大模型训练推理；河南郑州智算中心开工建设，总投资超 16 亿元。当前，以智算中心为代表的数字新基建正加速推进、快速落地。

（三）技术革新，激发智算行业创新活力

随着大模型参数量持续增大，数据处理难度和模型训推复杂度不断提升，持续推动智算技术革新，激发了行业创新活力。在芯片方面，我国芯片研发与美国相比，虽起步较晚，但在技术研发和产业生态建设方面，正逐渐展现出强劲发展势头。目前，我国部分企业已实现 28 纳米及以上、14 纳米及以上等工艺芯片的大规模量产并应用，以中芯国际为代表的部分企业正加速研发基于 DUV 技术的 7 纳米工艺芯片。在大模型方面，大模型正在加速人工智能产业化进程，推动算力服务普惠普适。DeepSeek、百度“文心一言”、阿里巴巴“通义千问”等国内

超大规模模型正加速演进，认知能力不断提升。同时，“预训练大模型+下游任务微调”的新范式，已成为破解人工智能技术落地难题的重要突破口。国内厂商纷纷加码大模型投入与研发，如 DeepSeek 也通过与企业合作构建垂类模型，助力中小企业搭建智能客服等应用，加速其数字化进程。阿里云通过 MaaS 开放大模型能力，服务各行各业数字化转型；各类科技公司也积极参与人工智能技术竞争，智算行业发展迎来新局面，算力普惠普适将成为重要支撑。在智算平台方面，通过多层级、多主体、异构算力节点的资源调度和交易，促进了智算资源的高效利用，如中国移动的震泽智算平台，通过分布式云原生 Kosmos 技术和异构混训解决方案，支持大模型跨域、跨芯训练，并通过算网大脑进行统一调度，推动通、智、边一体化调度，实现“中训边推”。

我国积极倡导发展绿色算力，借助“双碳”战略的推进，助力算力产业向绿色低碳转型。绿色算力技术的发展涉及多个层面，其中设施层和设备层的技术创新尤为关键。设施层技术创新聚焦于降低供电传输损耗、创新储能部署方式和应用自然冷却技术等。如内蒙古依托丰富的风能和太阳能资源，建设多个绿色算力中心，通过直接利用可再生能源供电，减少了碳排放，实现了绿色低碳高质量发展。设备层技术创新聚焦于服务器硬件节能、存储材料及工艺优化，以及无损网络介质的应用。如中国移动、中国联通等利用存算分离技术，大幅地降低了计算和存储服务器的能耗。

（四）规模扩张，推动智算产业赋能升级

我国算力总体规模持续扩大，智算中心建设加速推进，智算占比

不断提升。2023 年 10 月，工业和信息化部等六部门联合印发的《算力基础设施高质量发展行动计划》中强调要稳步提升算力综合供给能力，逐步合理提升智能算力占比。我国算力结构不断调整，智能算力加快布局。建设规模层面，截至 2025 年 6 月底，我国在用算力中心机架总规模达 1085 万标准机架，智算规模达 788 EFLOPS（FP16），算力供需呈现动态均衡态势。技术架构层面，截至 2024 年底，我国在用算力中心超 50%连接骨干网，网络互联质量显著提升。区域协同层面，截至 2025 年 6 月底，国家算力枢纽节点省市内在用标准机架占全国总量的 70%以上，枢纽节点整体效能与辐射范围持续优化。产业发展层面，我国算网基础设施升级优化，万卡集群建设加速推进，带动生成式人工智能应用加速涌现。

在智算规模持续扩大的背景下，基于智算的服务场景不断涌现，涵盖能源、工业、金融、教育、交通等多个行业，智算成为推动行业发展的关键驱动力。在能源领域，智算技术正深度赋能产业智能化升级，助力实现绿色低碳转型。中国石油打造的昆仑大模型应用场景覆盖油气勘探、开发、炼化、销售产业链以及产业金融、装备制造等领域，推进关键业务生产过程从劳动密集型向技术密集型转变，促进客户服务质量、员工工作效率持续提升，以点带面赋能产业焕新初见成效并持续深入。在工业领域，智算中心通过实时数据优化生产与资源配置。例如，中国宝武钢铁集团依托智算中心构建“钢铁产业大脑”，实现高炉能效优化和板带缺陷毫秒级检测，显著降低能耗并提升质量。在金融领域，金融机构可借助大模型创新业务模式、制定金融产品开

发策略，评估信贷、市场等风险，在风险事件发生前预警并提供应对策略，减少潜在的损失。如中国工商银行开发千亿级 AI 大模型，构建智慧化风险管理大数据应用生态链，推动风险管理从“事后分析”向“事前甄别”和“事中干预”转变。在教育、娱乐等领域，大模型可提供智能问答、个性化推荐等服务，满足用户需求。

二、智算中心规划与建设

（一）政策视角：国家智算中心布局与现状

我国持续加强智算中心顶层引导，对智算中心在内的算力基础设施进行优化部署，合理有序提升智算占比。《全国一体化大数据中心协同创新体系算力枢纽实施方案》明确提出构建八大枢纽节点，引导算力中心向枢纽节点聚集。《关于深入实施“人工智能+”行动的意见》明确提出优化国家智算资源布局，完善全国一体化算力网，充分发挥“东数西算”国家枢纽作用。“东数西算”工程进一步明确了东西部枢纽节点功能定位，其中东部枢纽节点重点支持工业互联网、金融证券、灾害预警、人工智能推理等对网络要求较高的业务。西部枢纽节点则重点处理后台加工、离线分析、存储备份等对网络要求不高的业务。此外，《算力基础设施高质量发展行动计划》明确提出，结合人工智能产业发展和业务需求，重点在西部算力枢纽及人工智能发展基础较好地区集约化开展智算中心建设，逐步合理提升智算占比。

从发展现状看，我国东西部地区在智算领域呈现出互补的发展态势。东部地区由于数字经济活跃，吸引大量大模型企业聚集，智算需求相对较高，促使智算中心建设加速。根据国家网信办发布的数据，

截至 2025 年 3 月，共有 346 款生成式人工智能服务在国家网信办完成备案²。其中，北上广的大模型数量约占全国总数的 65%，北京、上海、广东已经成为国产大模型的重要聚集地。从智算中心的建设看，东部地区数字经济发展活跃，各类大模型应用快速涌现，智算需求将持续增长，促使智算中心建设不断加速。**西部地区则依托自身能源、资源优势，在政策引导下积极推动智算中心布局。**如乌兰察布、呼和浩特、庆阳、贵阳等部分西部枢纽省市，以及郑州、合肥、长沙等中部省会城市，不断完善政策保障体系，优化算力中心建设运营环境，对智算中心项目给予补贴和奖励。同时，依托自身能源、资源优势，引导智算中心项目投资建设，智算中心数量和规模不断提升。

（二）行业视角：央国企智算布局与现状

我国大型央国企主要分布在涉及国家安全和国民经济命脉的关键行业，如能源、通信、交通、工业、金融等，央国业务体量较大，服务用户众多，亟需通过数字化转型降低生产经营成本，提高运营服务效率。为了进一步提高对人工智能技术的应用水平，加快推动业务变革，促进产业升级，央国企正在积极推动智算建设布局。不同行业央国企业务覆盖范围、经营规模、数据处理实时性以及信息安全保障等方面存在显著差异，其智算布局也呈现出多样化的特点。在智算中心的选址和建设过程中，央国企普遍遵循贴近业务场景和用户集中区域的原则，以提供更优质的算力服务。同时，充分考虑国家宏观政策导向和区域协同发展的战略要求，合理规划在中西部地区资源、能源

2 国家互联网信息办公室发布关于发布生成式人工智能服务已备案信息的公告(2025 年 1 月至 3 月)。

富集区域的智算中心布局，充分发挥当地的资源优势，实现资源的优化配置，形成集团级、区域级等多层级的智算布局体系。

能源行业在推动全过程数字化、智能化转型的过程中，大多选择在西部国家枢纽节点部署智算节点，以降低成本。中国石化在呼和浩特构建了包含高速计算资源、大容量存储、高效数据流转网络的异构智算资源池，目前总的智能算力约为 76 PFLOPS，预计到 2025 年将达到 100 PFLOPS³，为人工智能应用提供所需算力服务、数据服务和算法服务，支撑中训边推和行业智能生态建设。国家能源集团在宁夏布局大型智算中心，依托当地丰富煤电与绿电资源，降低算力成本。该智算中心聚焦煤矿智能开采、电力负荷预测等场景，为集团数字化转型提供算力支撑，也助力西部枢纽节点算力生态完善。

通信行业业务覆盖范围广，大规模、高并发、实时性等不同类型业务众多，通常需要构建覆盖全国主要节点多层级智算体系，以更好的满足广泛、差异化智算需求。以三大运营商为代表的通信行业企业在全国主要节点推动智算中心建设，形成多层级智算体系。根据《智算中心综合评价报告（2024 年）》⁴，中国移动正在统筹规划“N+X”智算中心布局体系，目前，已在京津冀、长三角、粤港澳大湾区、成渝、贵州、内蒙古、宁夏、黑龙江、湖北、山东等地区投产首批 12 个智算中心节点，实现“东数西算”枢纽节点基本覆盖。中国联通在通用算力“5+4+31+X”广泛布局的基础上，积极推进智算“1+N+X”梯次布

³ 2024 石油石化行业人工智能技术交流大会发布数据。

⁴ 中国信息通信研究院，《智算中心综合评价报告（2024 年）》。

局，构建以算力为核心的一体化算网融合生态体系，打造数字经济“第一算力引擎”。中国电信规划并建设了全国“2+3+7+X”公共智算云池，尤其在京津冀、长三角建设液冷单集群万卡智算池。

金融行业注重数据处理的安全性、实时性，智算中心大多部署在一线城市及省会城市，以满足全国或区域性的智算应用需求。金融机构每日产生海量的交易数据、客户信息数据等，对数据处理的安全性、实时性要求极高。智算中心具备强大的数据计算、分析和决策支持能力，有效支持金融结算、风险评估、市场分析、合规监管、智能客服等各类业务，为智慧金融发展提供重要支撑。在智算中心布局选址上，金融行业通常在一线城市及省会城市开展智算中心布局，城市网络性能优异、靠近终端用户，并且数字化专业人才聚集，保障了金融行业智算中心的高效运营。在建设运营模式上，金融机构与电信运营商形成优势互补的合作范式。金融机构作为主要出资方，从行业特性出发，对智算中心项目进行整体把关，确保其安全性、可靠性等符合金融行业数据计算、存储需求，电信运营商凭借技术与资源优势，承担提供智算中心建设及集成服务。双方协同推进金融智算中心建设，助力金融行业智算应用的全国化布局与区域化深耕。

工业领域聚焦现场实时数据处理和云边端协同能力构建，智算中心大多部署在靠近工业现场的地区，或通过云边协同构建“远端智算训练中心+边缘推理中心”的架构。随着工业 4.0 的推进，工业企业深化数字化、智能化技术的应用，对数据处理的实时性、准确性等提出较高要求。为更好地响应工业现场实时数据处理需求，工业智算中心

常采用云边协同方式部署。云端智算中心凭借强大的计算资源与先进的算法模型库，对边缘端上传的复杂数据进行深度挖掘与训练优化，再推送至边缘推理中心。边缘推理中心依据模型对工业现场实时数据进行决策判断，实现工业生产智能化闭环控制，提升生产效率，降低运营成本。例如，2024 年 4 月宝钢股份 AI 算力中心投运，依托宝信软件宝之云与华为昇腾 AI 平台搭建算力基础环境，现已成为国内钢铁行业领先的人工智能算力中心，有力支撑宝钢股份智能化转型。

（三）企业视角：央国企智算中心布局思路

通过对国家政策、行业趋势的研究，以及对基础电信运营商、石油石化、银行证券及地方国企等典型央国企智算建设布局现状的剖析，进一步明确了央国企智算中心布局的基本思路——**遵循政策导向，结合行业趋势，兼顾市场与责任，科学有序推动智算布局，打造性能卓越、绿色低碳、安全可靠、运营高效的智算中心底座。**

从国家政策来看，央国企在智算中心布局中应充分遵循国家、地方政策要求，参考我国智算中心总体布局现状和趋势，优先支持在枢纽节点、集群等关键区域开展智算中心建设。同时，应综合考虑土地、新能源、电力资源充裕度等环境因素，以及人工智能、大数据等新技术政策和产业环境，对各地区进行综合评估，选择合适的地区开展智算中心建设。央国企作为国民经济的支柱，应积极履行社会责任，在响应和落实国家政策要求重点方面发挥“排头兵”作用，为国家总体战略布局和产业协同发展贡献力量。

从行业趋势来看，不同行业的央国企业务覆盖范围、规模、响应速度、安全合规等要求存在差异，不同行业央国企在开展智算中心布局建设的过程中，应充分参考本行业内其他央国企先进的智算布局经验。同时，结合自身业务需求开展智算中心选址布局，推动智算中心建设。

从央国企现有实践来看，央国企在智算中心建设过程中既要满足市场规律，服务更多用户，提高运营收益，又要综合考虑社会责任与可持续发展，通过智算中心建设带动地区经济发展。对于终端用户较多，且致力于为其他行业用户提供智算服务的央国企，如基础电信运营商，可加强在全国主要枢纽节点和算力需求较高的非枢纽节点布局，加快构建覆盖全国的智算网络体系。对于能效管理要求较高，且主要满足本企业智算应用需求的企业，可重点在西部枢纽节点和集群开展智算建设布局。

与地方政府及民营企业主导建设的智算中心相比，央国企在智算中心布局建设过程中更加关注政策合规性、技术自主性、资源协同性、安全性和可持续发展能力。在政策合规性方面，央国企注重遵循国家宏观智算中心建设布局战略，确保智算中心建设与国家发展战略、区域发展规划相协调，避免盲目建设和资源浪费。在技术自主性方面，央国企更加注重核心技术和关键设备的自主可控，确保智算中心的安全稳定运行，降低对外部技术的依赖。在资源协同性方面，央国企资金实力雄厚，行业数据丰富，智算建设过程中，不仅只关注企业自身盈利能力，同时关注行业发展，会积极整合内外部资源，推动产业链

上下游企业的协同发展，提高行业整体算力应用和服务水平。在**安全性方面**，央国企在智算中心的建设和运营中，更加注重安全稳定和合规运营，会严格遵守国家相关法律法规和政策要求，加强网络安全和数据安全管理，确保智算中心的安全稳定运行。在**可持续发展方面**，央国企会更加注重环境保护和节能减排，采用绿色节能的技术和设备，降低智算中心的能耗和排放。同时，央国企还会关注社会事业服务能力建设，利用智算中心的技术优势为社会提供优质的公共服务和产品，推动社会进步和发展。

（四）智算中心建设模式

央国企智算中心建设总体上遵循方式选择、选址布局、建设规划、建设施工和运营管理这一基本流程，其中方式选择是央国企智算中心项目建设的关键，对后期项目投资和建设施工产生深远影响。选址布局可通过由面到点，先圈定区域后确定选址的方式进行位置选择。建设规划需要完成项目报批、风火水电基础设施以及算存运设备的规划设计，确保项目合规，并满足业务承载需求。建设施工过程主要选择施工方式，并形成完善的施工质量、进度、成本管理机制。运营管理主要是构建统一的制度体系、资源管理、资产管理和有偿使用机制，确保智算资源得到高效有序合理使用。



来源：中国信息通信研究院

图 1 智算中心建设模式

1.方式选择

智算中心建设方式主要包括改造升级、自建、租赁和租建结合。改造升级主要针对建成时间较早、业务承载能力较弱，但位置较为关键的算力中心。通过对供配电、制冷、机柜等机房关键系统进行改造，提高原有机柜供电能力和功率密度，并在此基础上，推进智算设备部署，优化运营服务，提升对智算业务的支撑和承载能力。自建是未来智算中心服务发展的主要方向。该模式允许央国企根据业务需求和技术趋势，量身定制智算中心的规模、架构和技术选型。尽管初期资金投入较大，但通过优化资源配置、提高设备利用率和降低运维成本，能降低长期经营成本。同时，新建智算中心能够使企业掌握核心技术和数据资源，增强技术自主可控能力。智算租赁是央国企获取智算资源的快速通道和有效补充。租赁模式允许企业根据实际需求灵活调整

租用算力资源规模，无需承担高昂的硬件投资和维护费用，减轻资金压力。此外，租赁模式提供了短期和长期选择，企业可根据项目需求灵活调整租赁期限。根据企业实践，也可按以上方式组合建设，即采用租建结合的方式开展智算中心建设。在初期，企业可通过租赁智算资源快速启动上层应用开发，同时规划自建智算中心以满足长期需求，不仅能减少对外部供应商的过度依赖，还能兼顾灵活性和成本效益。

表 1 智算中心建设模式

模式	内容	优缺点	适用对象
改造升级	对已有基础设施供电、制冷、机柜等机房关键系统进行改造，同时更新 IT 软硬件设备，使其具备智算服务能力	优点：可充分利用已有资源，建设成本相对较低 缺点：对技术要求较高，通常难以实现大规模、超大规模算力升级	适用于已有一定规模、智算需求并非超大规模，且对成本较为敏感的企业
自建	全新规划、选址、设计并构建智算中心	优点：可根据业务需求进行定制化设计，自主性较强，安全性较高 缺点：建设成本相对较高，建设周期长	适用于对算力有较高定制化需求、需要长期稳定运行，且对数据安全有严格要求的场景，如一些大型金融机构、科研机构
租赁	根据业务需要，从专业服务商处租赁智算资源和服务	优点：控制建设成本，可根据业务变化灵活调整租赁、自建资源比例和使用方式，兼顾灵活性和自主性 缺点：管理复杂性增加，面临技术整合挑战	适用于业务多元化的大中型企业，既有相对稳定的核心智算业务需求，又有弹性较大的非核心业务或临时性算力需求

租建结合	根据算力需求层次与变化规律，采取“自建核心算力设施 + 租赁弹性算力资源”的组合模式	优点：兼顾成本与核心需求，灵活性与扩展性强，平衡技术门槛与服务质量 缺点：跨平台、跨资源协调管理复杂度较高	适用于希望平衡成本投入与技术自主可控能力的企业或机构
------	--	--	----------------------------

来源：中国信息通信研究院

2.选址布局

在进行智算中心选址布局时，央国企可采用“优化布局划区域、综合评估定选址”的策略，由面及点，重点考虑当前及未来业务需求，同时综合考虑政策要求、能源供应、网络连接、环境条件等因素，初步圈定备选区域，并通过多维度综合量化评估，最终确定智算中心选址，以确保智算中心的稳定运行和可持续发展。具体步骤如下：

区域圈定。需重点考虑集团上中下游业务分布，满足业务部署、交互需求及技术需求，评估各业务部门工作负载特性、数据处理量、实时性要求等因素。同时，考虑与集团公司已有智算中心的数据交互能力。通过精准定位业务需求，初步评估智算中心规模、数量、位置，确保智算中心的设计、配置、选址能够有效支撑核心业务流程。

确定选址。应考虑国家政策、地理气候、能源、业务分布等因素。遵守法律法规，遵循国家布局规划，优先向枢纽节点、政策支持地区聚集。较好的选址区位应具备充沛的电力、充足的冷源，并避免自然灾害风险，确保智算中心安全、高效运行。除满足当前需求，还需前瞻性的预测未来业务趋势，预留扩展空间，适应业务规模的增长。为了更好地支持各类用户的算力需求，建议采取“中训边推”的布局模式。

在智算中心布局过程中，应参考其他央国企智算布局的成功案例。通过案例分析，为本企业的智算中心选址提供有益参考。此外，选址时需综合评估地区的网络环境、科研环境和人才基础等因素。良好的网络环境、科研环境和人才基础，能够为智算中心的建设和运营提供重要支撑。

3.建设规划

智算中心的规划设计需从土木工程到硬件配置、从模型构建到业务运营进行全面统筹设计。建设前期需完成土地证、房产证、地规证、建规证、施工许可证等许可手续，建设过程中还需办理消防、用水、通信等相关手续。同时，应站在全局视角，从基础设施层、平台层、模型层和安全治理等角度进行整体规划，重点是对机电、制冷、安防消防、算力设备等系统和设备进行规划设计。机电设计通常采用高密度方式布置，通过提高功率增强对智算服务器等相关设备的供电支持。在制冷系统方案设计中，从气候环境、功率密度、机房改造量、算力密度（单柜）以及花费成本等多维度，对风冷、水冷、液冷或联合制冷方案进行比选。在安防消防方面，需构建全面、多级联动的安防消防机制，重点选用一些支持场景感知、智能辅助和协同联动功能的设备，以更好地提供安防、消防支持。在算力设备部署过程中，央国企应加强国产算力设备的应用，减少对国外技术、产品的依赖。探索构建多元异构的算力设备方案，提升异构算力设备高效互联和协同计算能力，并针对不同算力应用场景提供适配方案。

4.建设施工

智算中心的建设施工可以采用 EPC（Engineering, Procurement, Construction）总包模式或分包模式。在 EPC 总包模式下，总包商对项目的设计、采购和施工全程负责。此模式有助于避免责任推诿，提高项目效率，并通过整体规划和实施，实现各环节的衔接和集成，降低项目整体风险。分包模式则将项目分解为多个专业部分，由不同专业分包商负责实施，提高项目质量。分包模式有助于分散风险，降低单一环节问题对整个项目的影响。

土建、设备采购安装、系统集成等环节是智算中心建设的核心组成部分，其质量直接关系到整个项目的稳定性和可靠性。在项目实施过程中，应重点做好以下工作。一是制定完善的质量管理制度和流程，明确各环节的质量标准和验收要求。二是派遣专业的质量管理人员和监理人员，对施工现场进行全程监督。通过集中监督检查、定期检查、抽样检测等方式，及时发现并纠正施工过程中的质量问题。三是根据项目的实际情况和需求，合理安排施工进度，优化施工流程，提高施工效率。四是建立有效反馈机制，及时收集和处理施工过程中出现的问题和意见，确保项目顺利进行。

5.运营管理

智算中心的运营管理是一个持续的过程，目标是实现资产可知、分配可管、运营可控以及优化可得，以支持其高效稳定运行，释放算力应用价值。智算中心运营管理包括统一的制度体系、资源管理、资产管理及有偿使用机制。统一制度体系旨在确保运营管理过程遵循统一的管理要求。集团总部、事业部、子公司以及作业现场需根据各自

的管理范围及权限，制定管理制度。集团总部应关注长期发展战略和短期运营目标，事业部及子公司应更加注重具体设施的管理规范，作业现场直接面向具体运维操作及运营服务流程，需为运营管理人员提供具体的工作指导。**统一资源管理旨在确保智算中心资源得到高效、灵活的配置和利用。**具体措施包括建立全面的算力产品目录，为不同租户构建资源空间和管理界面，实现多租户隔离。通过自动化运维、资源调度等技术，实现资源的动态调度和优化。**统一资产管理旨在确保智算中心资产得到准确、全面的管理和维护。**通过资产数据标准对资产采购、入库、调拨、报废等流程进行规范，确保资产的合规性、可追溯性，以及资产数据的准确性和完整性。同时，加强低效资产的治理，提高资产利用率和经济效益。**统一有偿使用机制旨在确保智算中心的资源得到合理、公平的定价和分配。**通过建立统一的计价标准和优惠政策，明确服务内容、质量标准、故障处理、统计结算流程等，确保收费的公平性和透明度。同时，结合市场需求，灵活调整定价策略和服务内容，提高智算中心的市场竞争力和盈利能力。

三、智算平台建设

（一）建设意义

智算平台是一种支持异构智算资源高效管理的平台，能够根据用户算力需求，对智算资源进行高效管理、智能编排和灵活调度。随着大模型参数规模的快速增长，大模型训练对智算资源的需求持续攀升，对智算平台在资源高效管理、异构协同和高效节能等方面提出了更高要求。智算平台已成为智算中心在资源管理、服务运营、创新发展中

不可或缺的管理工具。在资源管理方面，智算平台通过虚拟化、容器化技术以及高效的集群管理能力，实现各类异构计算、存储、网络等资源的集中管理和高效利用，支持大量智算芯片、服务器的并行计算，以及不同任务间的资源合理分配，提升整体资源利用率。在服务运营方面，智算平台能够实时监控业务系统的运行状态，优化业务处理流程，及时预警并处理潜在问题，保障业务稳定运行。同时，通过能效控制，可以有效降低计算能耗，实现高效计算。在创新发展方面，智算平台通过协同管理底层硬件、资源池、中间件以及各类应用平台，实现各类算力资源的灵活调度、管理和分配，不仅高效满足了各类任务的算力需求，还降低了智算应用门槛，促进人工智能技术的普及应用，为企业技术创新和智能化转型提供支撑。

（二）技术方案

1. 智算平台总体架构

智算资源管理是支撑智算软硬件资源高效分配和使用的关键，但现有智算资源管理存在诸多不足，资源管理和应用效率低，难以适应智算应用快速发展的需要。现有算力资源管理技术框架和平台大多以通用算力为基础，采用云资源管理方式对算力资源进行管理，但未考虑不同厂商智算芯片在编程接口、性能等方面的差异性，难以适应智算管理的需要。据公开数据统计，传统模式下智算中心的 GPU 利用率较低，平均数值低于 30%⁵。尽管部分厂商已经开始研发针对智算业务场景和资源需求特点的算力资源管理平台，但总体来看，现有平

⁵ 中国移动通信有限公司研究院，《NICC 新型智算中心算力池化技术白皮书（2023 年）》。

台和技术仍存在诸多不足，主要体现在以下两个方面。一是**资源分配方式相对粗放**。当前的资源分配通常以整卡为单位，或通过较为简单的虚拟化方式在一台智算服务器上构建多个虚拟机。这种粗粒度的分配模式难以灵活满足各类人工智能任务对资源的差异化需求，导致部分 GPU 资源常处于闲置状态，显著降低了算力资源的整体利用效率。二是**算力资源未能被充分利用**。人工智能算法模型的设计缺陷、算力优化工作不到位，以及模型框架与底层芯片之间的适配性不佳，都可能导致算力性能无法充分发挥，从而造成资源浪费。

智算平台是支撑异构算力高效管理与协同应用的基础性平台，对智算中心应用服务能力塑造具有重要作用。根据管理范围的不同，智算平台可实现对单集群算力资源管理，以及多集群的协同管理，单集群管理方面，可支持单一智算中心内异构算力资源的调度与算力加速，能够根据任务对算力资源的需求，在智算中心内为任务分配、调度相应计算资源，并支持超大规模集群算力加速能力，通过优化调度逻辑，提升软硬件协同效率，强化集群整体计算性能，满足高负载、高算力需求场景的运行效率。多集群协同管理方面，可实现跨域多个异构算力集群资源的统一管理、编排和调度，当接收到业务请求时，平台可对业务所需算力资源进行预估，并对接入和纳管的底层智算中心的计算、存储和网络资源进行感知和编排，预留相关资源，将数据和业务请求调度到相应的智算中心节点，确保业务在跨域环境下的高效运行。



来源：中国信息通信研究院

图 1 智算平台总体架构

智算平台包括硬件层、资源层、框架层、运营层和应用层五大层级。硬件层包含 CPU、GPU、FPGA、NPU 等多种异构智算资源，通过智算中心内网络实现多卡多机的高速互联，为上层提供强大的计算能力支撑。资源层建立在硬件层之上，负责各类硬件资源的感知、虚拟化与池化任务。通过资源层，系统能够动态识别并整合硬件资源，形成统一的资源池，实现资源的弹性分配与高效利用。此外，资源层还负责根据上层编排系统的智能调度算法和指令，将数据、容器动态调度到最适合的智算资源上，在任务执行过程中，会持续监控容器运行状态、节点健康度，一旦出现算力节点故障、数据传输超时等异常情况，会再次触发调度机制，重新评估资源状态，将数据、容器迁移至更优节点，确保任务以最优方式执行。框架层提供训练框架和推理框架，包括 TensorFlow、PyTorch 等主流框架。这些框架封装了大量

高级 API 和工具，使得开发者能够更直观、高效地编写、调试和优化深度学习模型。**运营层**作为智算平台的中枢神经，负责整个平台的监控、调度与优化。通过收集并分析来自各层的实时数据，运营层利用智能算法进行资源调度和任务管理，确保平台能够以最优状态运行。**应用层**涵盖计算机视觉、自然语言处理、跨媒体分析推理等常见智算应用。这些应用依托下层提供的强大计算能力、灵活模型开发环境与高效运营管理机制，实现了从数据输入到结果输出的全自动化处理。

2. 智算平台关键技术

智算平台关键技术主要包括异构算力资源池化、智算资源编排、智算任务调度等。其中：

1) 异构算力资源池化

在传统资源池化技术中，人工智能应用程序需通过访问芯片运行时提供的编程接口，获取对智算芯片资源调度的权限。然而，当前市场上不同厂商所生产的智算芯片，其编程接口存在显著差异，如华为的 CANN(Compute Architecture for Neural Networks)、英伟达的 CUDA (Compute Unified Device Architecture)、AMD 的 ROCm (Radeon Open Compute Platform) 等。因此，现有的智算平台需要针对每类芯片单独构建智算资源池体系，难以实现对各类芯片资源的高效协同调用与统一管理，对提升智算平台整体运行效率和资源综合利用率形成了一定制约。

异构算力资源池化旨在整合不同架构的计算资源(如 CPU、GPU、FPGA、ASIC 等)，形成统一的算力资源池，以满足多样化的应用场

景需求。异构算力资源池化的核心在于实现异构芯片、硬件设备的高速互联、兼容适配和资源抽象，涉及到的关键技术包括异构硬件互联与高速通信技术、资源池化接口协议、资源抽象与虚拟化技术等。其中，**异构硬件互联与高速通信技术**主要实现多卡多机高速互联，卡间互联方面，英伟达采用专用 NVLink、NVSwitch 协议实现多 GPU 互联，并积极推动 NVLink Fusion 技术研发，支持芯片间的高速互联。国外 UALink 联盟推出 UALink 1.0 互联标准，促进各异构芯片简单开放互联。国内形成了华为 UB、阿里 ALS、腾讯 ETH-X、移动 OSIA 等互联协议，支持异构芯片间的高速互联。**资源池化接口协议**是指制定统一的资源调用接口，允许上层应用以标准化方式访问异构算力。

资源抽象与虚拟化技术按照虚拟化的不同层级可分为**硬件抽象层虚拟化**、**操作系统层虚拟化**、**函数库层虚拟化**、**应用程序层虚拟化**等。**硬件抽象层虚拟化**以物理硬件（如整卡 GPU、服务器）为单位，通过 Hypervisor 模拟完整硬件环境，每个虚拟机独立运行操作系统（如 Windows/Linux），资源隔离性最强但灵活性最低。**操作系统层虚拟化**是指基于 Linux 内核实现进程级隔离，共享宿主机内核，资源隔离粒度细化至 CPU 核心、内存 MB 级，支持动态伸缩。如通过 Docker、Kubernetes（K8s）将应用及其依赖封装为容器，共享宿主机内核，支持 CPU 和 GPU 的细粒度切分。**库函数层虚拟化**是指通过劫持或仿真硬件函数库（如 CUDA、OpenCL）的 API 调用，将应用对算力的请求重定向至池化资源。通常情况下，人工智能应用会通过调用芯片提供的运行时 API 来实现对特定智算芯片资源的访问。而在

API 劫持技术下，这些调用会被资源池管理系统捕获、处理。系统根据人工智能应用的需求、算力资源的实时状态及分配策略等因素，动态决定如何分配算力资源。**应用程序层虚拟化**从应用程序运行时行为出发，动态拆分任务至异构资源执行，通常可采用应用程序监视器实现对应用程序行为的监控和预估，包括计算任务的类型、计算复杂度和资源使用情况等，为资源调度、分配提供依据。应用程序监视器可分为前端和后端：前端负责监控指定应用程序的活动，获取其请求并转发至后端处理；后端则根据程序的资源需求进行动态分配，或将应用程序拆分到多台机器上运行，实现资源的细粒度管理和远程调用。

2) 智算资源编排

智算资源编排是指智算平台根据业务需求，自动化地部署、配置和管理算力资源，为各类业务场景提供高效的算力应用服务。其核心在于制定合适的算力资源分配策略，构建业务需求与资源之间的映射关系，通过对有限算力资源的管理和分配，高效有序地满足各类业务需求。常见的算力编排策略包括整体业务性能最优、高优先级业务性能最优、响应速度最优、能耗最优等。其中，整体业务性能最优策略的核心目标是最大化整个业务系统的整体性能；高优先级业务性能最优策略优先保障高优先级业务的算力需求；响应速度最优策略优先为需要快速响应的业务分配算力资源，以减少用户等待时间；能耗最优策略则综合考虑业务需求和资源使用情况，制定一个既能满足业务需求又能降低能耗的资源分配方案。

3) 智算任务调度

智算任务调度是指在智算资源编排策略的指导下，通过实时监控和分析计算资源状态、网络链路状态、负载以及应用需求变化，动态选择最优网络链路，将数据和业务请求调度到合适的算力资源节点，从而确保任务高效执行，并满足编排策略的要求。智算任务调度重点关注在已确定的策略下，实现最大化的资源利用率，保障业务调度过程弹性高效、稳定安全。算力调度涉及的关键技术包括算力路由、任务分析与分解、弹性调度与故障迁移等。其中，**算力路由**是指实时收集、汇总各计算节点资源使用情况，生成包含网络 and 计算参数信息的新型路由表，为算力调度提供实时、精准的决策依据。**任务分析与分解**是解析任务属性，拆分不同的子业务，并在编排策略指导下将业务调度到各计算节点，如用户采用能耗最优的编排策略，智算平台会选择将业务调度到能效比较高且满足业务需求的计算节点。**弹性调度**是指实时采集节点资源使用状态，根据任务量自动调整资源分配，如任务激增时，申请空闲资源进行业务处理。**故障迁移**是指任务执行过程中，遇到节点故障情况，能够自动迁移到其他节点并恢复中间状态，确保业务稳定、连续。

3. 业界平台建设案例

算力平台主要用于实现对各类计算资源的池化管理和编排调度，相关平台的构建通常基于 K8s。作为一个开源的容器集群管理系统，K8s 提供了容器编排、资源调度、弹性伸缩、部署管理和服务发现等功能，使容器化应用的部署工作更加简单高效。在智算资源管理和调度中，K8s 能够自动监测应用需求，动态调整资源分配，确保系统在

各种负载情况下均可保持稳定运行。同时，K8s 支持多种调度策略，满足高性能计算场景下的资源调度需求。因此，K8s 已成为智能算力管理调度的事实标准⁶。

业界厂商以 K8s 为基础，持续加强异构算力资源适配、编排调度算法等技术研发，推动智算平台建设。中国移动打造“大云震泽”智算品牌，提供数据处理、智算训推、大模型服务三大平台，基于底层“N+X”智算资源，实现万卡任务一键拉起、通智边一体化调度、模型任务异构混训等关键技术，并提供“30+”基础模型及一站式智能体开发平台，为客户提供从资源到模型应用的全链路工具。华为 HCS+CCE+ModelArts 是一套综合性云计算和人工智能解决方案，为企业和开发者提供更加高效、便捷的人工智能开发和应用部署环境。其中，HCS（Huawei Cloud Stack）提供强大的计算能力，支持快速构建和部署应用程序。CCE（Cloud Container Engine）是华为云提供的容器云服务，支持 Kubernetes 等容器编排技术，帮助企业快速构建、部署和管理容器化应用。ModelArts 作为一站式人工智能开发平台，提供海量数据预处理及半自动化标注、大规模分布式训练、自动化模型生成及端-边-云模型按需部署能力。联想发布万全异构智算平台，围绕算力匹配魔方、GPU 内核态虚拟化、集合通信算法库、人工智能高效断点续训技术、人工智能与 HPC 集群超级调度器五个方面进行创新，为用户提供更加简单、高效、便捷的算力应用服务。

（三）央国企智算平台建设模式

⁶ 中国信通院《云计算白皮书（2024 年）》、CCNIS 算网融合产业及标准推进委员会《云原生泛在算力调度研究报告》。

智算平台的建设可采用多种模式，包括租用、自研及共建。租用模式依托云服务提供商，适合需快速扩展且对成本敏感的场景。自研模式适合对算力有高度定制化需求，并强调长期稳定运行和数据安全的企业。共建模式则依托合作伙伴的资源共享与技术协作，适用于联合推进特定项目的场景。各模式主要内容如表 2 所示。

表 2 智算平台建设模式

模式	内容	优缺点	适用场景
租用	租用第三方服务商的算力平台，一般作为算力资源租赁的配套部分。	优点：灵活性高、成本低； 缺点：可能受限于服务商提供的服务范围和质量。	快速扩展算力资源、灵活性要求较高、对成本较为敏感的场景，如金融、教育、医疗等行业。
自研	企业自行研发算力平台，并独立管理和维护算力资源。	优点：控制力强、安全性高； 缺点：研发投入大、维护成本高。	对算力有高度定制化需求、需长期稳定运行、对数据安全有严格要求场景，如大型金融机构、科研机构。
共建	与合作伙伴联合建设算力平台，共享资源和技术。	优点：资源互补、技术共享； 缺点：合作过程中可能存在利益冲突和协调难度。	需要合作推进某个项目或业务的场景，且合作双方都有一定资源和技术优势的情况。如地方政府与企业共建区域算力平台、高校与企业合作共建科研算力平台。

来源：中国信息通信研究院

央国企与民营企业是我国经济社会发展不可或缺的重要市场主

体，两者在行业分布、业务范围、监管要求、信息系统需求等方面存在差异，在智算平台建设模式选择上也各有侧重。央国企业务涉及国民经济的多个关键领域和重要行业，业务覆盖范围广、服务用户多，对信息系统的安全性、稳定性、可靠性、技术自主性等提出较高要求，因此更倾向于成熟稳定、安全且具备技术自主性的建设方案和模式。对于承载核心业务的系统平台应以自研为主。对于一些缺乏资金和技术积累的企业，可与相关厂商合作开展平台共建。对于一些资金相对缺乏，且不涉及核心数据的基础通用业务，可租用第三方服务商算力平台，实现业务快速灵活部署和上线服务。

1.租用建设模式

租用模式适用于央国企数字化转型初期，或面临短期内需要大量算力资源支持，但自身不具备快速部署、维护和管理大规模异构智算资源能力的情况，一般作为算力资源租赁的配套部分。此外，一些对成本较为敏感，或不希望在短期内一次性投入大量资金用于软硬件采购、研发和运维的企业，通常选择租用模式。在租用模式下，企业能够根据自身业务需求灵活购买算力资源和平台服务，实现资源的弹性拓展。

在智算平台租赁过程中，央国企应重点关注租赁期限、平台服务能力、业务迁移、租赁费用等关键问题，并形成明确的租用方案。在租赁期限方面，短期内平台租赁成本较低，但长期费用递增。大型央国企短期内适宜租赁，长期应考虑自研。短期租赁期限通常 1-3 年，长期可达 5 年以上。中小型央国企算力需求挖掘期较长，长期租赁成

本可控。在平台服务能力方面，应选择技术基础扎实、经验丰富且口碑良好的服务提供商。在业务迁移方面，需对现有 IT 系统及数据进行全面评估，了解系统架构、性能、数据量、信息安全等级保护要求等关键信息，制定详细的系统应用迁移方案。在租赁费用方面，需根据数字化转型的算力需求，对租赁算力资源容量及所需算力应用服务进行评估，明确智算平台的费用结构。

部分央国企在算力应用上的投入不足，很多应用尚未上云。通过采用租赁平台模式，央国企避免自研智算平台可能带来的技术实现存在漏洞、人才短缺、资金链断裂等安全隐患，提高效率和安全性。此外，也有部分央国企为加快推动数字化、智能化应用，在自研智算平台的同时，租用外部智算平台及资源。

2. 自研建设模式

自研模式是指企业根据自行或委托外部厂商结合自身平台建设需求研发智算平台，自主管理、维护和使用算力资源，以满足自身业务需求。自研模式具有高度自主性，适用于对自主化要求较高、信息化专业人才较多、信息系统管理和研发经验丰富的企业。

自研智算平台的建设是一个复杂且系统化的工程。在需求分析与规划阶段，企业需深入剖析内部各项业务对智算平台在数据处理能力、计算速度、存储等方面的应用服务需求，构建业务算力需求列表。同时，全面研究本地部署、私有云、公有云等部署模式，以及 SaaS、PaaS、IaaS 等多样化的服务模式。通过综合评估，选择最契合企业发展需求的部署和服务模式。基于前期规划，中期进行架构设计与开发

测试阶段。基于业务需求和部署模式，设计平台总体架构。具体而言，依据算力需求采购设备搭建物理资源层，并构建智算资源层。此外，建立平台服务层开发环境，为平台运行提供基础服务。依据测试结果，对平台进行持续优化，同时形成相关的技术文档，为后续的维护和升级提供支撑。在安全管理与运维阶段，要高度重视安全防护措施的设计与实施，加强智能运维（AIOps）、数字孪生与可视化、IT 运维智能体等技术在智算平台安全防护和运营管理中的应用，保障平台具备高可靠性、高安全性和高性能。建立健全涵盖监控、备份、恢复等机制的运维管理体系，制定完善的制度规范和工作手册，确保平台长期、稳定、高效的运行。



来源：中国信息通信研究院

图 3 自研智算平台建设示意图

大型央国企对数据安全性要求较高，通过自研智算平台建设，有

助于强化对核心业务的承载能力，并提升对算力资源的管控能力。中国移动、中国联通、中国电信均积极推动智算平台建设，如中国移动“震泽”MaaS 平台、中国联通“星罗”先进算力调度平台、天翼云息壤一体化智算服务平台，均提供异构算力调度和管理能力，实现从智算资源到算法、数据的全栈化服务，显著提高智算应用和服务能力。

3. 共建建设模式

共建模式适用于具备一定智算技术专业人才储备、计划自行开发大型模型、注重模型训练与推理效率，同时寻求减少计算资源成本投入并控制人力规模的企业。在选择共建模式时，央国企和合作伙伴应明确以下四个问题。一是平台定位与目标。在共建模式下，智算平台不仅要为本企业提供智算服务支持，还需进一步形成支持产业上下游协同的能力，为上下游企业提供信息共享、智算汇聚、调度、交易、行业大模型等基础服务，为产业创新高质量发展提供支撑。二是数据安全与隐私保护。在共建过程中，应重点关注企业数据、用户数据的安全性和隐私性，避免数据外泄。三是技术选型与兼容性。共建平台应选择成熟、稳定、兼容性强的技术架构，确保平台高效稳定运行且易于扩展。四是合作机制与责权划分。合作双方应明确合作模式、资源投入及收益分配机制，合理划分双方责权，确保合作顺畅且可持续。

目前，共建模式主要适用于一些产业服务性平台的建设。如中国信通院与各地方政府合建的中国算力平台，致力于实现对全国各地算力中心基础信息的查询、监测，并提供算力纳管等服务。此外，一些高校也积极与电信运营商合作开展智算平台建设，如复旦大学与阿里

云、中国电信共同打造的智算平台 CFFF（Computer for the Future at Fudan），主要服务于前沿技术研究和科学计算。

四、典型实践案例

（一）中国石油昆仑大模型及算力建设

中国石油围绕“业务发展、管理变革、技术赋能”三大主线，坚持“价值导向、战略引领、创新驱动、平台支撑”总体原则，科学部署，精心组织，全力打造能源化工行业一流大模型和深度应用场景，不断塑造高质量发展新动能新优势。在行业大模型方面，语言大模型参数达 3000 亿，视觉大模型参数达 44 亿，多模态大模型参数达 800 亿；在专业大模型方面，勘探全领域专业大模型持续迭代，问数、代码、物资采购、安全生产等专业大模型不断完善；场景大模型不断升级，为碳酸盐岩走滑断裂识别等关键领域赋新能。在场景应用方面，围绕 26 条业务线、119 个业务域，优化形成“十域百景千应用”的全景视图，发布 100 个应用场景。在 AI 中台建设方面，创新研发融合多家训推工具链的 AI 中台，通过重构与集成，实现了商业大模型在一个中台的统一纳管，支持主流开源大模型的私有化部署，算力、模型、服务的统一管理和协同调度能力显著增强。在智算算力建设方面，搭建了全栈国产化的智算算力环境，算力资源纳入 AI 中台管理，高峰时算力使用规模达到 1950P。

（二）中国移动震泽 MaaS 平台

中国移动积极落实国家“东数西算”战略，构建“4+N+31+X”的算力网络布局，建成覆盖京津冀、长三角、粤港澳等区域的 13 个核心

智算中心节点，算力服务器规模超 100 万台，智算规模达到 61.3EFLOPS，支持 10 万卡 GPU 集群统一纳管，搭建高速泛在网络，实现 400G 骨干网覆盖国家八大枢纽节点，云专网覆盖全国超 300 个城市。在智算中心建设方面，中国移动以“N+X”为目标体系，按照“集中训练、分布推理，统一管控、弹性调度，自主可控、绿色低碳”的原则，构建技术领先、绿色节能、服务全局的智算中心。中国移动的智算中心主要分布在京津冀、长三角、粤港澳、湖北、成渝、贵州和内蒙古等地。其中，N 节点主要承载自研训练及大颗粒算力需求，承接央国企大模型训练任务、科学装置以及头部大企业大颗粒 AI 算力需求，呼和浩特、哈尔滨等大万卡训练场已投入使用，支撑 1+X 大模型训练；X 节点主要满足在边缘场景下，位于地市、县区、园区、业务现场的实时推理需求，在私有化场景下，以项目制方式建设，当前基于移动云底座已规模化部署智算推理资源，累计上线推理卡超过 2.1 万张。在人工智能平台建设方面，中国移动以“震泽 MaaS 服务平台”为核心，依托中国移动 N+X 万卡级智算资源布局，提供基于国产算力的高效数据、训练、推理、智能体及智能运维五大引擎，全链路优化国产算力“数/训/推/用”性能，降低国产算力成本达 35%，MFU 算力超 35%，故障识别准确率超 85%；提供一句话生成智能体能力，搭载百款模型、千款智能体、百款插件&MCP 等多样化生态，助力各行各业“零门槛”构建专属 AI 应用，实现全方位场景和全要素能力的运营与供给。

（三）中国建行“建行云”数字基础设施建设

中国建设银行加速数智化转型，积极践行金融行业科技战略，夯实数字金融基础设施。根据《中国建设银行-2024 年年度报告》，2024 年末算力规模 507.72 PFLOPS，较上年增长 9.58%，其中图形处理器 (GPU) 等新型算力占比超 23.39%。在人工智能平台建设方面，中国建设银行构建了“建行云”数字基础设施，主要聚焦计算机视觉、智能语音、自然语言处理、知识图谱、智能决策等五大领域，助力金融数字化转型。截至 2024 年末，已完成金融大模型的迭代更新 16 次，上线 168 个金融大模型应用场景。在智算中心建设方面，为有效应对大型金融模型训练的算力需求，中国建设银行积极构建高效的算力基础设施。其中位于和林格尔新区的算力中心项目被定位为超大规模金融算力中心建设和运营的标杆项目，该项目总占地面积 569 亩，规划总建筑面积约 52 万平方米，计划总投资额达 100 亿元人民币，项目全部建成后服务器装机能力可达到 30 万台。

（四）深圳智城翼云科技智慧城市算力统筹调度平台

深圳市智城翼云科技有限公司已建设并运营了“深圳开放智算中心”、“深圳市智慧城市算力统筹调度平台”，形成了“一中心+一平台”的组合优势，构建了“算力度量-算网汇聚-算网调度”三位一体技术体系和高性能的算网供给与调度能力。“一中心+一平台”将强化发力支撑“算力+”场景落地，通过算力供给及调度满足深圳科研、科技创新、人工智能产业企业的多样算力需求，同时引领国内技术发展。

“深圳开放智算中心”规划建设 4000PFLOPS 算力规模，已完成建设 2500PFLOPS 算力规模。目前深圳开放智算中心为全市政务服务、

科学研究、人工智能企业提供稳定可靠的智能算力服务。依托河套深圳园区，积极推动打造跨境算力枢纽，推进香港算力融入全国一体化算力网络，谋划构建跨境算力协同体系，通过“算力供港、港数深算”新模式，为香港高校、科创企业提供高性价比算力支持，加速大湾区人工智能企业享有普惠算力服务。

“深圳市智慧城市算力统筹调度平台”紧密耦合超算中心、鹏城云脑等大科学装置，连接贵州、韶关等国家算力枢纽节点，充分调动汇集各方算力资源，实现了“算力一网化、统筹一体化、调度一站式”，是深圳市在线算力的“操作系统”。已汇聚鹏城云脑II、天翼云、华为云、腾讯云等超过 140 个资源池，汇聚智能算力超 19000PFLOPS，到 2025 年底平台供给调度规模将超 40000PFLOPS，主要面向深圳市以及粤港澳大湾区各产业链、国资国企及中小微创企业提供算力服务，并逐步承接公共服务职能，打造商业运作及公共服务双线并行的算力运营服务模式。

在算力产业布局方面，“一中心+一平台”将继续强化发力支撑“算力+”场景落地，围绕深圳推动升级工业制造、“20+8”产业集群等，结合多领域算力赋能产业激励政策，承载算力资金补贴、优惠等措施，引导人工智能、工业制造、教育、医疗、金融、交通、低空等多产业积极使用市统筹算力资源，加快算力与产业深度融合，促使算力与深圳市主要支柱产业相辅相成，共同发展。

（五）珠海云上智城国产智算及其调度应用

为贯彻国家新质生产力发展战略，珠海市于 2023 年提出“云上智

城”建设规划，并于 2024 年授权珠海云上智城投资运营控股有限公司作为算力基础设施建设运营主体。

云上智城公司及其投资建设的珠海市大数据中心，是珠江口西岸规模领先的 A 级数据中心服务商，持有约 4000 个标准机柜资源，在此资源基础上积极推进国产算力布局。公司所建设的 400P 国产算力已经上线，即将投入正式运营。400P 算力包含适用于训练、推理、微调的多种国产算力，可以服务于多种应用场景。在此基础上构建四大核心平台：算力互联平台实现跨区域算力共享；智能体平台提供多模态 AI 决策支持；数据要素平台建立数据流通机制；产业赋能平台推动传统行业转型。通过整合包括 ChatGLM4.5、DeepSeek-V3、Qwen 等主流大模型，为珠海本地政企提供一站式数字化服务，并辐射粤港澳大湾区。其中，算力互联平台是智算充分发挥其作用的底座级平台，构建“本地+远程”的两级调度服务体系，打造粤港澳大湾区高性价比算力洼地，加速区域内中小企业转型与制造业升级，吸引产业生态集聚，提升城市数字治理水平和核心竞争力，为珠海数字经济高质量发展提供持续动力。该平台围绕六大能力建设：聚合多元算力，实现跨域统一纳管；增强调度机制，支持异构单卡调度；推行一体化运营，优化资源匹配；落地算力券政策，降低企业成本；提供大模型云服务，促进 AI 普及；实现多向接入与系统集成，平台建成后接入韶关算力集群及全国算力网，成为国家“东数西算”战略的一个毗邻港澳的桥头堡阵地。

五、发展建议

（一）明确发展路径，有序推动智算中心建设

央国企作为国民经济的龙头，在规划方面，应遵循国家刚性约束，将智算中心建设提升至战略高度。央国企具有雄厚的资金实力和较强的资源整合能力，有能力承担大型智算中心的建设和运营任务，应依据企业业务需求和市场定位，规划智算中心的规模、功能及服务对象。如能源类央国企侧重于利用智算中心优化能源勘探、生产和调配流程。通过精准的战略规划，央国企能够明晰智算中心建设的发展路径，确保资源投入的针对性和有效性。

在明确各阶段的目标、任务，以及当前及未来智算需求基础上，央国企应分阶段合理有序推动智算中心基础设施建设。应摸清企业内现有智算中心建设、高性能 GPU 芯片采购及使用等情况，为后续资源整合和优化提供坚实的数据支持。持续开展智算产业态势、前沿技术趋势研究研判，形成动态跟踪机制，及时调整技术路线和建设规划，分阶段推动项目建设。同时，要建立科学的评估体系，定期对智算中心建设成效进行量化分析，确保资源投入与当前及未来长期业务需求相匹配，避免资源浪费。此外，大型央国企可考虑与其他相关企业开展合作，整合算力技术、优化高性能算力资源配置，优先支持具备基础条件、能力成熟、场景必要的大模型项目。

（二）优化资源调配，强化智算平台技术创新

整合与优化国家算力资源调配。央国企应落实国家战略部署，加强通用计算、智能计算和超级计算等多元算力资源的科学布局，与国

家算力基础设施实现更紧密的对接与整合，通过高效的资源调度动态分配算力资源，促进各关键领域良性发展、重点项目有序开展。

构建统一的资源共享标准体系，强化算力资源灵活共享与协同合作。央国企应共同制定、完善智算资源共享的标准与规范，明确资源的分类、计量、定价、安全等方面的要求，确保资源共享的公平性、透明性和可操作性，打破企业之间的壁垒，实现资源的自由流通和高效共享。此外，央国企可联合基础电信运营商，依托大数据、人工智能等先进技术构建智算平台，实现算力资源的共享和动态调配。

与科技企业合作推动智算平台技术创新。央国企应与科技企业加强合作，借助其在人工智能、大数据、云计算、区块链等前沿技术领域的专业优势，联合开展技术研发与创新应用，优化智算平台性能。

（三）深化防护措施，完善智算安全运营体系

遵循国家及行业的安全标准和法规。央国企应严格遵守《网络安全法》、《个人信息保护法》等法律法规，确保智算中心的合法合规运营。同时，应建立快速响应的应急机制，覆盖事件报告、初步处置、影响评估及恢复重建等环节，以高效应对各类突发事件。

构建多层次防护体系。央国企应打造包含物理安全、网络安全、系统安全及应用安全在内的立体化安全屏障，有效抵御多样化的安全威胁。在硬件与软件的选择上，应考虑减少供应链安全风险。

建立全面且动态更新的安全管理及运维体系。安全管理体系方面，央国企应制定安全策略、管理制度及明确的操作规程，确保所有操作均遵循严格的规范指引，并采用国际先进的加密技术对传输及存储的数据实施高强度加密，有效防止数据遭受非法窃取或篡改。运维体系

方面，央国企应加强核心技术研发，通过自动化配置管理工具等确保系统配置一致性，减少人为错误，以创新驱动智能化运维持续发展。

（四）强化综合施策，推动绿色低碳集约发展

为有序推动智算中心的绿色低碳发展，央国企在规划建设中应以国家战略指导方针为依据，从强化算力集约布局、推动绿色能源利用、加强节能降碳技术应用等方面综合施策，加强对智算中心建设与应用的引导和约束，确保资源的合理配置与高效利用。

在强化算力集约布局方面，央国企应强化算力集约供给，依据国家、区域算力布局相关政策要求，科学规划企业智算中心节点布局，推动算力资源向能源富集、气候适宜、土地成本适宜的区域集中，满足均衡型、计算和存储密集型等多样化业务需求。通过与基础电信运营商、云服务提供商及各类算力平台单位合作，推动智算与通算协同发展。同时，应加强跨地区智算中心间的资源协同，实现负载均衡和资源动态调配，提升整个系统的弹性和灵活性。

在推动绿色能源利用方面，央国企需建立智算中心绿色能源利用的刚性约束机制，在智算中心的建筑设计、施工和运维阶段，广泛应用人工智能、大数据、云计算及物联网等数字化技术优势，实现智算中心全生命周期的智能高效管理。

在加强节能降碳技术应用方面，央国企应加强节能降碳设备利用，优先采购通过国家节能产品认证、绿色等级认证的产品和设备。积极推动老旧小散算力中心改造，加快推动高压直流、储能、蓄冷、液冷、

余热回收等节能降碳技术应用，结合人工智能、数字孪生等技术实现对制冷、供电、设备运行状态、能耗等数据的监测及全局能效优化，实现从单点节能向系统智能降碳转变。积极参与绿色低碳等级认证，不断提高节能降碳水平，助力“双碳”目标落地见效。

中国信息通信研究院 云计算与大数据研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62300095

传真：010-62304980

网址：www.caict.ac.cn

