



中国通信工业协会数据中心委员会  
China Communications Industry Association Data Center Committee

An abstract graphic featuring a stylized city skyline with several skyscrapers in shades of blue and purple. The buildings are composed of horizontal lines, giving them a digital or data-like appearance. In the foreground, there are large, flowing, blue wavy lines that sweep across the page, creating a sense of motion and connectivity.

# 中国智算中心产业 发展白皮书

(2024年)

中国通信工业协会数据中心委员会

- 2024年12月 -

# 前言

智算中心作为信息基础设施的重要组成部分，通过算力的生产、聚合、调度和释放，能够为快速增长的人工智能算力需求提供基础支撑，在推进AI产业化、赋能产业AI化、助力治理智能化、促进产业集群化等方面具有显著作用，是数字经济时代促进科技创新、优化产业结构、提升国家竞争力的重要支撑。

中国智算中心产业迎来重要发展机遇。工业和信息化部、中央网信办、教育部、国家卫生健康委、中国人民银行、国务院国资委等六部门联合印发的《算力基础设施高质量发展行动计划》明确提出，到2025年，我国算力规模超过300EFLOPS，智能算力占比达到35%。在政策指引与市场需求的共同推动下，智算中心产业进入快速发展期，大批智算中心项目落地，智算算力规模迅速增长。未来，随着人工智能、大数据等新兴技术驱动下的应用场景日益丰富，AI硬件、算力服务、模型应用等产业链重要环节的商业模式不断创新发展，智算中心产业生态将加速形成，并向供给优化、数实融合、绿色低碳、产业协同等方向演进。

为深入剖析中国智算中心产业发展现状，精准把握智算中心产业未来发展态势，中国通信工业协会数据中心委员会集合智算中心产业专家资源和产业链力量，联合发布《中国智算中心产业发展白皮书（2024）》。本白皮书立足于我国智算中心产业实际，展望未来，为共创智能计算新时代、建设数字中国贡献力量。本白皮书全面梳理我国智算中心产业发展的现状、趋势、机遇与挑战，系统剖析智算中心建设、运营面临的新形势、新要求，希望为智算中心产业相关管理机构、行业协会、产业从业者等提供参考。

# 参编单位

中国通信工业协会数据中心委员会

张家口市人民政府

芜湖市人民政府

韶关市人民政府

乌兰察布市人民政府

庆阳市人民政府

大同市人民政府

内蒙古和林格尔新区管理委员会

中国长江三峡集团有限公司

中科院建筑设计研究院有限公司

中国建设银行股份有限公司

北京科智咨询有限公司

上海宝信数据中心有限公司

秦淮数据集团

北京数道智算科技有限公司

北京世纪互联宽带数据中心有限公司

科华数据股份有限公司

科大讯飞股份有限公司

河南昆仑技术有限公司

润泽科技发展有限公司

博大数据服务（深圳）有限公司

联想（北京）有限公司

北京中科合盈数据科技有限公司

南方电网大数据服务有限公司

联通（广东）产业互联网有限公司

杭州星临科技有限责任公司

广东美的暖通设备有限公司

四川省天府云数据科技有限责任公司

北京泰豪智能工程有限公司

潍柴重机股份有限公司

中电系统建设工程有限公司

恒华数字科技集团有限公司

深圳市中电电力技术股份有限公司

广州市汇源通信建设监理有限公司

中建三局第一建设安装有限公司

双登集团股份有限公司

兰洋(宁波)科技有限公司

广州尚航信息科技股份有限公司

华章数据技术有限公司

中世顺科技（北京）股份有限公司

润建股份有限公司

浙江洁普智汇能源科技有限公司

广东申菱环境系统股份有限公司

鹏博士电信传媒集团股份有限公司

盛达全球畅服（深圳）科技有限公司

公诚管理咨询有限公司

# 参编专家

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| 金和平 | 黄 超 | 张福林 | 张晓雪 | 潘建初 | 王 芳 | 张 雷 |
| 夏 亮 | 张炳华 | 甄 鉴 | 张泽锐 | 田春青 | 赵文博 | 方虹珺 |
| 周超辉 | 马 羚 | 孙东旺 | 沈 诚 | 严 瀚 | 陈 亮 | 陈智柔 |
| 许志德 | 邓国峰 | 王元月 | 于景淇 | 郭俊峰 | 刘 运 | 刘晶晶 |
| 鞠昌斌 | 闫 昆 | 任国斌 | 杨 薇 | 何 明 | 杨彦超 | 张学超 |
| 汤 莉 | 王 侃 | 唐巨光 | 刘大闯 | 周洪聪 | 周科翔 | 徐冬明 |
| 白瑞晨 | 兰满桔 | 韦统辉 | 张劲松 | 陈 盼 | 邵南军 | 陈 华 |
| 侯兴泽 | 李 佳 | 陈经纬 | 张 兴 |     |     |     |



# 目录

- 一、智算中心产业发展背景及内涵.....8
  - (一) 智算中心产业发展背景.....9
  - (二) 智算中心产业内涵.....10
- 二、全球智算产业步入快速发展新阶段.....13
  - (一) 人工智能驱动智算规模扩张.....14
  - (二) AI芯片技术迭代奠定智算基础.....14
  - (三) 大模型加速人工智能与产业融合.....15
- 三、中国智算中心产业迎来黄金发展期.....16
  - (一) 新型基础设施赋能数字经济发展.....17
  - (二) 政策推动智算中心合理有序建设.....17
  - (三) 大模型应用深化行业数字化转型.....20
- 四、中国智算中心产业发展现状.....21
  - (一) 智算中心市场规模持续扩大.....22
  - (二) 各类投资主体积极投建智算中心.....24
  - (三) 智算中心商业模式不断演进.....28
  - (四) 产业发展持续践行绿色低碳理念.....31
- 五、AI时代算力中心建设开启新篇章.....33
  - (一) 人工智能发展对算力中心建设提出更高要求.....34
  - (二) 技术创新加速AI硬件国产替代.....37
  - (三) 高效网络架构推动算网融合.....40
  - (四) 功率密度提升促进高效制冷实践.....42
  - (五) 新型供配电系统重塑能源应用.....45

|                          |    |
|--------------------------|----|
| 六、智算中心运营服务向高效安全方向发展..... | 47 |
| （一）算力资源高效调度.....         | 48 |
| （二）数据安全多重保障.....         | 49 |
| （三）运维平台智能可视.....         | 50 |
| 七、中国智算中心产业发展展望与倡议.....   | 52 |
| （一）中国智算中心产业发展展望.....     | 53 |
| （二）中国智算中心产业发展倡议.....     | 56 |
| 附：中国智算中心典型项目案例.....      | 59 |

## 图表目录

|                                    |    |
|------------------------------------|----|
| 图表 1：中国智算中心产业链.....                | 11 |
| 图表 2：中国智算中心相关政策梳理.....             | 18 |
| 图表 3：2020-2028年中国智算中心市场规模及预测.....  | 22 |
| 图表 4：中国投运/在建/规划智算中心项目区域分布.....     | 23 |
| 图表 5：中国智算中心项目主体分布（按项目数量）.....      | 25 |
| 图表 6：中国智算中心项目主体分布（按算力规模）.....      | 26 |
| 图表 7：智算中心建设主体类别及优劣势.....           | 27 |
| 图表 8：智算中心商业模式类别.....               | 28 |
| 图表 9：智算中心总体架构.....                 | 35 |
| 图表 10：AI芯片（GPU、FPGA、ASIC）主要特点..... | 38 |
| 图表 11：主流高性能芯片特性对比.....             | 39 |
| 图表 12：InfiniBand和RoCE方案比较.....     | 41 |
| 图表 13：数据中心液冷技术方案对比.....            | 43 |
| 图表 14：中国高性能算力中心先进制冷方案应用案例.....     | 44 |
| 图表 15：算力资源调度与管理架构.....             | 49 |

The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, overlapping curves that sweep across the page from the bottom left towards the top right.

# 第一章 智算中心产业发展 背景及内涵

# 一、智算中心产业发展背景及内涵

## （一）智算中心产业发展背景

人工智能作为引领未来的战略性技术，逐步成为衡量国家国际竞争力的重要领域，高性能算力是人工智能发展的重要组成部分。从全球范围看，各国纷纷制定人工智能相关的战略和政策，推动高性能算力发展。如美国成立智算中心基础设施特别工作组、欧盟出台《欧洲高性能计算共同计划》、日本发布《人工智能战略2022》等。在当前复杂的国际竞争形势下，发展人工智能产业，有利于在新一轮国际科技竞争中掌握主导权。

我国高度重视人工智能发展，党的十八大以来，不断完善人工智能发展顶层设计。自2017年，国务院印发《新一代人工智能发展规划》，明确坚持人工智能研发攻关、产品应用和产业培育“三位一体”推进，全面支撑科技、经济、社会发展和国家安全，部署构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国。到2024年，全国两会政府工作报告强调人工智能在推动经济社会发展中的重要作用，提出强化基础研究、拓展应用领域、加强人才培养等，促进人工智能与实体经济深度融合。国家在政策层面全方位推进人工智能产业发展。

智算中心是人工智能发展的关键基础设施，是带动社会科技创新、经济发展以及社会治理水平提升，赋能全产业链发展的重要力量。智算中心能够为基础研究和应用开发提供先进计算能力，同时加速5G、大数据等新一代信息技术的迭代发展；推动人工智能与各行业融合创新，推动制造、教育、金融、能源等

领域的数字化转型，提升产业发展数字化智能化水平；促进人才培养和资源整合，加快地区人工智能产业集聚发展；助力智慧城市建设、公共安全保障和政务服务优化，提升社会治理水平。智算中心产业成为我国提升国际竞争力、建设创新型国家的重要支撑。

自生成式人工智能问世以来，国内人工智能大模型快速发展。截至2024年11月，国家网信办备案的生成式人工智能大模型已达到252个，通用大模型和垂类应用大模型训练的智算算力需求迅速增加。目前，通用大模型和垂类应用大模型的已经逐步商用，开始赋能千行百业。未来，人工智能大模型与产业必将走向深度融合，产业智算算力需求将得到充分释放。在人工智能快速发展的带动下，智算中心产业投资迅速增加。截至2024年11月，全国已投运的智算中心项目近150个，在建及规划建设的智算中心项目近400个。

## （二）智算中心产业内涵

智算中心是基于人工智能计算架构，提供人工智能应用所需算力服务、数据服务和算法服务的算力基础设施，融合高性能计算设备、高速网络以及先进的软件系统，为人工智能训练和推理提供高效、稳定的计算环境。

智算中心的主要功能包括：

- 提供强大的计算能力：智算中心采用专门的AI算力硬件，如GPU、NPU、TPU等，以支持高效的AI计算任务。
- 高效的数据处理：智算中心融合了高性能计算设备和高速网络，能够处理大规模的数据集和复杂的计算任务。
- 支持多种AI应用：智算中心适用于计算机视觉、自然语言处理、机器学习



等领域，处理图像识别、语音识别、文本分析、模型训练推理等任务

狭义上，智算中心是通用算力中心的升级，在传统机房的基础上部署AI算力以提供服务。广义地说，智算中心是提供人工智能应用所需算力服务、数据服务和算法服务的新型算力基础设施，包含基础层、平台层和应用层。其中，基础部分是支撑智算中心建设与应用的先进人工智能理论和计算架构；平台部分围绕智算中心算力生产、聚合、调度、释放的作业逻辑展开；应用层提供算力生产供应、数据开放共享、智能生态建设和产业创新聚集。

智算中心产业链涵盖从AI芯片/服务器等设计制造、基础设施建设，到智算服务提供，以及生成式大模型研发及基于大模型的行业应用。

图表 1：中国智算中心产业链



信息来源：科智咨询

在上游环节，主要包括土建基础设施和IT基础架构的建设。土建基础设施涵盖土建施工、制冷系统、供配电系统和电信运营等，为智算中心提供稳定可靠的物理环境；IT基础架构包括芯片设计制造、AI服务器、网络设备、存储设备

和数据中心管理系统等，它们共同构成智算中心的计算、存储和通信基础。AI芯片作为智能算力的核心，技术壁垒高，在GPU、FPGA、TPU、NPU等细分领域，英伟达、赛灵思、英特尔、谷歌、华为、阿里、寒武纪等企业是主要参与者。

在中游环节，智算服务提供商、云服务供应商和IDC服务商等基于自身优势，提供智算服务及运维解决方案。主流云服务供应商不仅自建大型智算中心，还加速布局AI大模型，以提供更高效、更智能的服务。IDC服务商也依托云网资源优势参与智算建设，提供智算、超算、通算等多样化的算力服务及一体化运维解决方案。部分云服务商与科技公司利用自身技术壁垒提供大模型及平台服务，为下游用户提供更加丰富、智能的算力服务。

在下游环节，互联网、金融、电信、交通等行业的人工智能应用需求带动自动驾驶、机器人、元宇宙、智慧医疗、文娱创作、智慧科研等相关产业发展。通过提供智能算力和算法，加速新技术的研发和商业化进程，推动产业创新进步。随着各行业积极探索生成式人工智能应用创新点，大模型应用场景逐步拓展。在海外市场，大模型主要在传媒、游戏、机器人、办公等领域落地应用。国内市场大模型应用则主要聚焦于金融、医疗、传媒、游戏、汽车等领域。

The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, parallel strokes that curve and sweep across the page, primarily concentrated on the right side and bottom, leaving the left side more open for the text.

## 第二章

# 全球智算产业步入 快速发展新阶段

## 二、全球智算产业步入快速发展新阶段

### （一）人工智能驱动智算规模扩张

大模型参数量实现指数级增长，推理和训练以海量数据为基础，全球范围内大模型应用推动数据量激增，数据类型和来源更加丰富。在数据量、大模型复杂算法以及应用场景的推动下，承载AI计算能力的智算中心进入快速扩张阶段。

国际科技巨头微软、谷歌、亚马逊，国内互联网企业阿里、腾讯、字节跳动，以及基础电信运营商等，纷纷在全球多个区域建设智能计算中心以拓展AI云服务或满足自身AI业务发展需求，其他各类主体也积极投资建设智算中心。2023年全球智能算力规模增速超过算力总规模增速，随着技术的不断进步和需求的持续扩大，智算中心建设加速推进，智能算力将继续保持快速增长势头。

### （二）AI芯片技术迭代奠定智算基础

高性能芯片技术快速迭代创新，为人工智能发展提供保障，进而带动智算中心发展。Nvidia作为全球GPU算力芯片市场领导者，代表性产品V100、A100、H100技术指标处于领先水平，最新产品Blackwell GPU采用先进的4纳米工艺，提供高达20 petaflops的FP4运算能力。其他科技巨头如AMD、英特尔、微软、亚马逊和谷歌也在AI芯片领域展开竞争。同时，中国AI芯片国产化进程正加速发展，华为、寒武纪、海光信息、景嘉微以及阿里、百度等企业不仅在自研AI芯片技术上取得重要进展，还通过产品集成、行业解决方案及生态伙伴合作等方式推动国产AI芯片商业化应用，为智能算力发展提供坚实基础。



随着人工智能技术的快速发展，以及人工智能与5G、物联网、边缘计算等技术的不断融合，高性能芯片也将迎来技术创新的新高潮。

### （三）大模型加速人工智能与产业融合

生成式人工智能应用蓬勃发展。随着AIGC技术不断进步，大模型商业化应用场景从文生文、文生图应用逐步扩展至音频生成、视频生成、多模态生成类，以及面向不同行业领域或用户群体的工具类应用，如代码生成、Copilot、数字人、营销工具、聊天助手等。通用大模型适用性广，包括ChatGPT、讯飞星火认知、百度文心一言、阿里通义千问等；行业大模型专业性强，国内头部科技企业、基础电信运营商、大型央国企等正加速推动垂类大模型研发落地。

据测算，2024年全球生成式人工智能市场规模将达到400亿美元，到2030年有望增长到1.5万亿美元，2022-2030年复合年增长率高达83%。未来AIGC将加速在更多场景商业化落地，如广告、游戏、自媒体等内容创作领域以及金融、电商、软件开发、教育等行业，呈现多模态、专业化、集成化的特点。



The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, overlapping curves that sweep across the page from the bottom left towards the top right.

# 第三章 中国智算中心产业 迎来黄金发展期

## 三、中国智算中心产业迎来黄金发展期

### （一）新型基础设施赋能数字经济发展

智算中心是基于最新人工智能理论，采用领先的人工智能计算架构，提供人工智能应用所需算力服务、数据服务和算法服务的新型算力基础设施。智算中心以算力技术和算法模型为关键核心，算力技术以AI芯片、AI服务器、AI集群为载体，算法模型则以AI大模型为主要代表。

智算中心在提供高效算力资源和服务的基础上，聚焦区域产业创新发展。智算中心项目具有良好的社会经济效益，对技术创新、产业发展的带动作用明显。根据国家信息中心数据，在智算中心实现80%应用水平的情况下，城市/地区在智算中心建设投入的增长量对创新产出的贡献率约为14%-17%；预计2020年至2030年我国人工智能核心产业规模的年均复合增长率达20.9%，带动相关产业规模的年均复合增长率达25.9%。

### （二）政策推动智算中心合理有序建设

政府高度重视智算中心建设，密集出台相关政策。为推动智算中心有序发展，国家出台多项政策统筹建设面向人工智能领域的算力和算法中心，打造智能算力、通用算法和开发平台一体化的新型智能基础设施。各地政府积极响应，纷纷发布相关政策加快智算中心产业发展。智算政策聚焦关键领域，如高端高效智能经济、安全便捷智能社会建设等，着力打造人工智能重点应用场景，推动智算中心在关键领域的应用和普及，促进数字化转型。

2017年《新一代人工智能发展规划》出台，首次将智能计算中心提升到国家战略层面，明确提出“建立人工智能超级计算中心”的战略目标，为智能计算中心的发展奠定政策基础。2020年，国家发改委将智能计算中心纳入新型基础设施建设范畴，激发各地智算中心投资热情。在政策推动下，多地政府开始积极落地智能计算中心项目，智算中心进入快速扩张期。2022年“东数西算”工程实施以来，国家政策明确对全国算力资源进行统筹和智能调度，对智能计算中心发展的引导力度进一步加强。2024年9月，工信部等十一部门发布《关于推动新型信息基础设施协调发展有关事项的通知》，强调逐步提升智能算力占比，东部发达地区先行先试、探索5G-A、人工智能等建设和应用新模式，西部地区在综合成本优势明显地区合理布局重大算力设施，探索建设超大型人工智能训练算力设施。

图表 2：中国智算中心相关政策梳理

| 时间     | 发文单位  | 政策/会议名称         | 主要内容                                                                                 |
|--------|-------|-----------------|--------------------------------------------------------------------------------------|
| 2017.8 | 国务院   | 《新一代人工智能发展规划》   | 布局前沿基础理论研究，建立新一代人工智能关键共性技术体系，统筹布局人工智能创新平台，建立人工智能超级计算中心、大规模超级智能计算支撑环境等。               |
| 2020.4 | 国家发改委 | 国家发改委新闻发布会      | 官方首次明确了“新基建”的范围，包括信息基础设施、融合基础设施、创新基础设施三个方面。其中，信息基础设施包括以数据中心、智能计算中心为代表的算力基础设施等。       |
| 2021.1 | 国务院   | 《建设高标准市场体系行动方案》 | 加大新型基础设施投资力度，推动第五代移动通信、物联网、工业互联网等通信网络基础设施，人工智能、云计算、区块链等新技术基础设施，数据中心、智能计算中心等算力基础设施建设。 |

|         |               |                              |                                                                                                                                                                          |
|---------|---------------|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2021.7  | 工信部           | 《新型数据中心发展三年行动计划（2021-2023年）》 | 加快提升算力算效水平。引导新型数据中心集约化、高密化、智能化建设，稳步提高数据中心单体规模、单机架功率，加快高性能、智能计算中心部署，推动CPU、GPU等异构算力提升，逐步提高自主研发算力的部署比例，推进新型数据中心算力供应多元化，支撑各类智能应用。强化标准支撑引领。加快推进边缘数据中心、智能计算中心等标准建设，支撑新技术新应用落地。 |
| 2021.12 | 中央网络安全和信息化委员会 | 《“十四五”国家信息化规划》               | 加强信息技术专利创新。围绕大数据中心、智能计算中心等领域加强高价值专利培育。构建云网融合的新型算力设施。统筹建设面向区块链和人工智能等的算力和算法中心。                                                                                             |
| 2021.12 | 国务院           | 《“十四五”数字经济发展规划》              | 推动智能计算中心有序发展，打造智能算力、通用算法和开发平台一体化的新型智能基础设施，面向政务服务、智慧城市、智能制造、自动驾驶、语言智能等重点新兴领域，提供体系化的人工智能服务。                                                                                |
| 2022.1  | 国务院           | 《关于印发全国一体化政务大数据体系建设指南的通知》    | 鼓励各地区各部门推进数据基础能力建设，积极构建数据安全存储、数据存证、隐私计算等支撑体系，推动大数据挖掘分析、智能计算、数据安全与隐私保护等核心技术攻关。                                                                                            |
| 2022.8  | 科技部、财政部       | 《企业技术创新能力提升行动方案（2022-2023年）》 | 推动国家超算中心，智能计算中心等面向企业提供低成本算力服务。                                                                                                                                           |
| 2022.11 | 国务院           | 《关于数字经济发展的情况的报告》             | 算力基础设施达到世界领先水平。建成一批国家新一代人工智能公共算力开放创新平台，以低成本算力服务支撑中小企业发展需求。                                                                                                               |
| 2023.2  | 中共中央<br>国务院   | 《数字中国建设整体布局规划》               | 系统优化算力基础设施布局，促进东西部算力高效互补和协同联动，引导通用数据中心、超算中心、智能计算中心、边缘数据中心等合理梯次布局。                                                                                                        |
| 2024.9  | 工信部等十一部门      | 《关于推动新型信息基础设施协调发展的通知》        | 优化布局算力基础设施，引导面向全国、区域提供服务的大型及超大型数据中心、智能计算中心、超算中心在枢纽节点部署。逐步提升智能算力占比。东部发达地区先行先试、探索5G-A、人工智能等建设和应用新模式，西部地区在综合成本优势明显地区合理布局重大算力设施，探索建设超大型人工智能训练算力设施。                           |

信息来源：科智咨询整理



### （三）大模型应用深化行业数字化转型

人工智能大模型在数量、类别、参数方面不断发展，效率和能力不断提升，为千行百业数字化转型提供人工智能解决方案。截至2024年7月底，全国范围内已有197个生成式人工智能服务完成备案，10亿参数规模以上大模型数量已超100个。通用大模型任务处理能力不断提升，文本理解和生成能力增强，图像识别更加精准；行业大模型深耕专业场景，百度、腾讯、华为、京东、网易等企业已发布适用于特定领域的行业大模型，提高业务效率。在参数方面，大语言模型的参数规模已从亿增长到千亿，甚至达到万亿级别，模型的计算效率和性能也逐步提高，在推动经济社会数字化转型方面发挥重要作用。

目前，人工智能大模型正在由开发阶段步入行业应用阶段，为金融、医疗、教育、制造、交通、能源等多个行业的数字化转型提供了有力支撑。企业可利用大模型对数据进行深度挖掘，优化业务流程，提高运营效率。如制造业大模型可以用于生产线故障预测、生产计划优化等，从而降低生产成本，提高生产效率。金融机构可利用大模型创新业务模式，制定金融产品开发策略，并为客户提供分析和建议等。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。

The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, overlapping strokes, giving them a textured, almost ethereal appearance. They sweep across the page from the top left towards the bottom right, framing the central text.

## 第四章

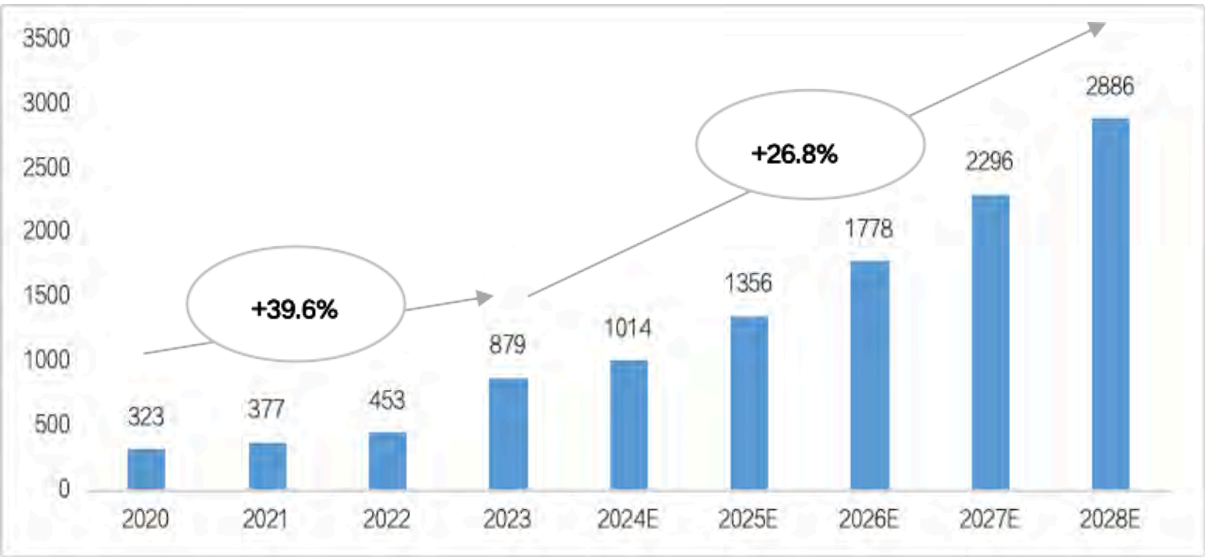
# 中国智算中心产业 发展现状

## 四、中国智算中心产业发展现状

### （一）智算中心市场规模持续扩大

在需求推动下中国智算中心市场投资规模高速增长。2022年生成式人工智能大模型推向市场，在国内引起AIGC发展热潮，大模型训练对智能算力的需求迅速攀升。2023年起国内头部互联网企业及科技公司加速AIGC布局，政府也牵头建设公共智能算力中心，赋能社会数字化转型需求，全国智算中心投资火热，智算中心市场规模大幅增长。2023年中国智算中心市场投资规模达879亿，同比增长90%以上。预计2024年国产化芯片产能提升，智算算力供给瓶颈将逐渐缓解。未来，AI大模型应用场景不断丰富，商用进程加快，智算中心市场增长动力逐渐由训练切换至推理，市场进入平稳增长期，预计2028年中国智算中心市场投资规模有望达到2886亿元。

图表 3：2020-2028年中国智算中心市场规模及预测（亿元）



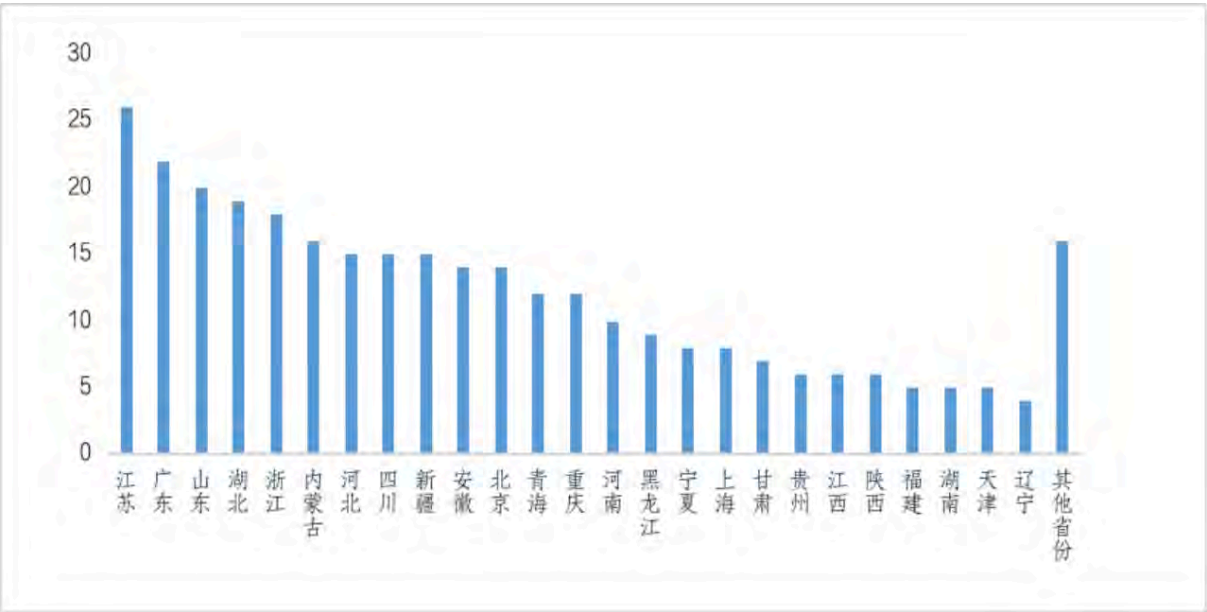
注：智算中心市场规模包含算力（AI服务器）、存储+网络设备、基础设施（机房、风火水电等）、算法（软件平台、数据服务）领域的投资。

数据来源：科智咨询

截至2024年8月，中国智算中心项目超过300个，已公布算力规模超50万PFlops。从已投用、在建、规划的智算中心项目来看，全国各省智算中心总计300余个，约三分之一智算中心项目规划算力大于500PFlops，主要为政府或基础电信运营商投建项目。2024年当年投运项目数量超过50个，60%以上为地方政府、国资平台及电信运营商项目，总计新增算力规模约为2万PFlops。

在区域分布上，智算中心项目主要集中在东部地区。中国智算中心产业发展具有明显区域特征，东部地区集聚大量高科技企业，“AI+行业”应用进程较快，为人工智能发展提供丰富的应用场景和数据资源，智算需求集中。同时，大模型训练和推理通常涉及多卡构成的算力集群，集群节点之间网络通信压力大，出于算力高效性和维护便利性考虑，当前大量智算中心项目落地东部发达地区。截至2024年8月底，江苏省智算中心项目数量最多，其次为广东及山东省；京津冀、长三角及广东省智算中心项目合计超过120个，占比超40%。

图表 4：中国投运/在建/规划智算中心项目区域分布（截至2024年8月）



数据来源：科智咨询



大模型与行业深入融合发展，AI推理需求将快速提升，带动一线城市、区域经济中心及周边地区等智算中心部署。AI训练需求对于网络时延要求相对推理较低，在大规模部署阶段将更加注重经济性，将带动基于AI训练的智算中心项目落地成本更低的“东数西算”西部集群地区。同时，近两年政策强调统筹算力资源布局，推动区域协同发展。2024年8月《工业和信息化部等十一部门关于推动新型信息基础设施协调发展有关事项的通知》提出，东部发达地区先行先试、探索5G-A、人工智能等建设和应用新模式，中西部和东北地区加快千兆城市建设，实现5G、千兆光网均衡发展，西部地区在综合成本优势明显地区合理布局重大算力设施，探索建设超大型人工智能训练算力设施。长期来看，智算中心市场发展重心将逐渐向中西部经济中心及东数西算集群地区转移。

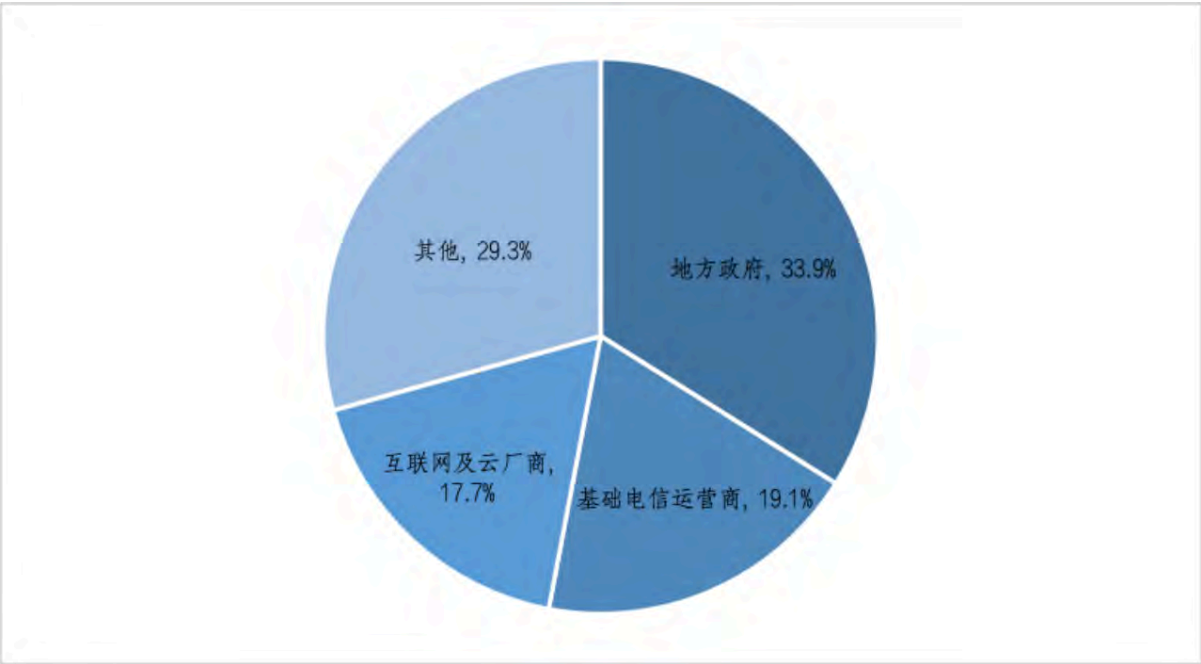
## （二）各类投资主体积极投建智算中心

根据智算中心项目数量统计，截至2024年8月，全国投运、在建及规划的智算中心中，地方政府和基础电信运营商主导建设的智算中心项目占比超过50%，互联网及云厂商项目数量占比约为17.7%，地方政府及基础电信运营商是智算中心主要参与方。

各地方政府一方面响应国家战略，优化算力资源配置，实现东西部算力资源的高效互补和协同联动；另一方面，智算中心作为数字时代区域基础设施的重要组成部分，能够为区域技术创新、产业转型升级和数字经济发展提供重要支撑。基础电信运营商兼具网络提供商与算力中心服务商双重身份，“东数西算”工程启动后，基础电信运营加快建设智算中心，提供网络及算力基础设施支持。从业务发展层面看，电信运营商在传统业务增长放缓的情况下加大在算

力领域的布局，着力发展业务第二增长曲线。

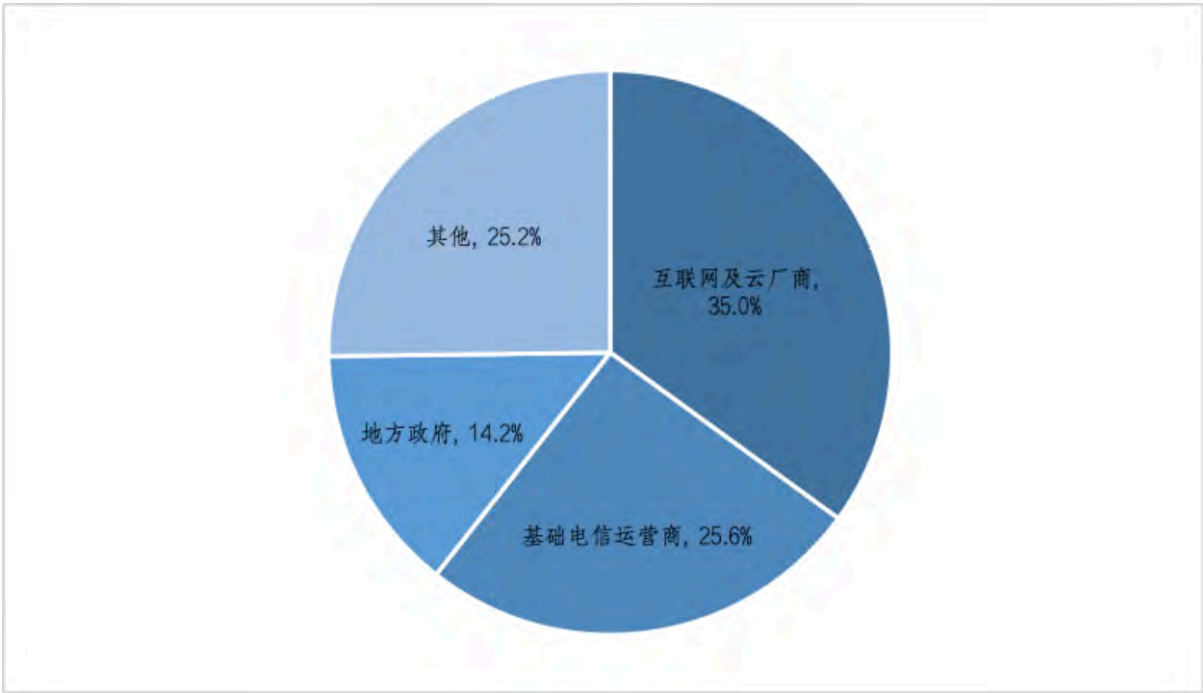
图表 5：中国智算中心项目主体分布（按项目数量，截至2024年8月）



数据来源：科智咨询

从智算中心算力规模来看，互联网及云厂商在智算中心投资建设中占据重要地位。科技巨头在人工智能计算中心的建设中扮演重要角色，许多大型科技公司如阿里巴巴和腾讯，都在人工智能计算中心投入大量的资源。互联网及云厂商建设的智算中心规模较大，多为万卡集群，智算中心具备大规模、可扩展性、绿色化等特征，满足互联网及云业务长远发展需求。企业主体在智算中心建设中可以提供强大的资金、技术支持和市场应用，以及市场化的管理运营机制，相比于政府等国资平台更注重效率和效益，有助于提高智算中心的运营效率和服务质量。截至2024年8月，全国投运、在建及规划的智算中心中，互联网及云厂商建设的智算中心规模占比超过30%，其次为基础电信运营商，占比约为25.6%。

图表 6：中国智算中心项目主体分布（按算力规模，截至2024年8月）



数据来源：科智咨询

不同主体的智算中心功能定位及布局诉求存在差异。政府主导建设的人工智能计算中心主要服务于地方人工智能大模型研发应用及相关领域数字化转型，加速产业发展、城市治理、公共服务等各方面智能化进程，推动智慧医疗、智慧交通、金融科技、无人驾驶等各领域数智化水平。基础电信运营商投资建设的智算中心成为政府算力基础设施的良好补充，承担国家算力网络建设任务的同时满足自身算力需要及算力业务拓展需求。大型云厂商如腾讯云、阿里云、百度云等，具有完善的供应链资源、扎实的技术能力以及丰富的客户资源，出于满足集团AI发展需求及云业务拓展考虑，建设智算中心。此外，浪潮、协鑫等产业上游供应商，商汤等AI科创企业，理想、小鹏等下游应用企业等基于产业链价值延伸和日常业务需要布局智算中心。

图表 7：智算中心建设主体类别及优劣势

| 主体              | 战略目标                   | 优劣势                                     | 代表案例                                    |
|-----------------|------------------------|-----------------------------------------|-----------------------------------------|
| 地方政府            | 响应国家战略，推动政策落地；促进区域产业发展 | 优势：政策、资金、产业园区客户                         | 北京/上海/广州/武汉/天津/重庆/长沙/南京/西安/成都/合肥等30多个城市 |
|                 |                        | 劣势：缺乏IT技术、供应链经验、运营能力不足                  |                                         |
| 互联网及云厂商、基础电信运营商 | 满足自身大模型训练需求；拓展算力业务     | 优势：充足的资金、技术和客户资源，丰富的软件供应链资源，较成熟的云算力业务模式 | · 腾讯合肥智算中心/百度阳泉智算中心/阿里飞天云智能华东算力中心等；     |
|                 |                        | 劣势：自有大模型与大模型训练客户形成竞争互斥                  | · 中国电信武清智算中心/中国联通芜湖智算中心/中国移动武汉智算中心等     |
| 第三方IDC服务商       | 寻求业务增长                 | 优势：丰富的IDC机房资源，IDC一体化建设运营能力              | · 润泽国际信息港A-11云数据中心                      |
|                 |                        | 劣势：缺乏IT供应链资源和算力客户                       | · 博大数据深圳前海智算中心                          |
| 服务器厂商/芯片渠道商     | 纵向一体化                  | 优势：AI芯片、服务器等硬件资源获取能力                    | · 协鑫智算（上海）中心                            |
|                 |                        | 劣势：缺乏IDC供应链和机房建设运营能力                    | · 浪潮新疆克拉玛依智算中心                          |
| AI企业、应用企业       | 纵向一体化                  | 优势：算法及相关软件能力，客户资源和应用场景积累                | · 商汤临港AIDC                              |
|                 |                        | 劣势：缺乏AI芯片货源、IDC资源                       | · 理想汽车智算中心<br>· 小鹏“扶摇”智算中心              |
| 跨界企业            | 战略转型/发展新业务             | 优势：一般与芯片厂商、渠道商合作，获得芯片货源                 | · 英博数科北京AI创新赋能中心                        |
|                 |                        | 劣势：缺乏IDC资源和持续稳定的客户                      | · 威星智能贵安智算中心<br>· 恒润股份芜湖智算中心            |

信息来源：科智咨询

现阶段，智能算力在快速增长的同时存在供需匹配问题，部分智能算力资源利用率较低。规划和实际需求之间存在差距，如算力架构设计不合理，或技术更新迭代后设备无法满足新的计算要求，可能导致智算中心资源闲置。部分智



算中心由于市场推广不足或服务类型单一等，算力资源未能得到充分利用。

（三）智算中心商业模式不断演进

1、智算中心商业模式类别

智算需求具有场景多样化、高度定制化特征，同时智算中心涵盖从底层基础设施到上层应用各个方面，衍生出多元化的智算中心服务体系。智算中心服务包括机房托管服务、算力租赁服务、AI平台服务、模型定制服务及AI应用服务。

图表 8：智算中心商业模式类别

| 类型              | 商业模式   | 介绍                            | 供应方                      | 需求方                      |
|-----------------|--------|-------------------------------|--------------------------|--------------------------|
| IaaS<br>基础设施即服务 | 机房托管服务 | 在传统数据中心机房托管基础上，提供更高功耗、配电和网络定制 | 智算转型的数据中心服务商、中立的智算中心服务商等 | 头部云商及AI公司、大型央国企等         |
|                 | 算力租赁服务 | 以云服务形式租赁智能算力，按使用时间和规模收费       | 云厂转型的智算服务商、中立的智算中心服务商等   | 中小型科技公司、IT公司、非连续需求的科研机构等 |
| PaaS<br>平台即服务   | AI平台服务 | 提供人工智能应用开发工具和平台               | 头部IT公司                   | 中小企业和开发者                 |
| MaaS<br>模型即服务   | 模型定制服务 | 模型定制、精调、部署等AI大模型全流程服务         | 成熟的大模型供应商（具有AI大模型技术能力）   | 中小垂直行业企业                 |
| SaaS<br>软件即服务   | AI应用服务 | 直接应用于企业业务，提供人工智能分析、决策等服务      | 具有AI能力的垂直行业头部企业          | 小型垂直行业企业                 |

信息来源：科智咨询

机房托管服务与传统数据中心服务模式类型相同，但需要面向智算的特点进行深度定制。例如，为了满足AI训练对高功耗、高密度计算的需求，智算中心的机房托管服务在功耗管理、配电系统、网络架构等方面进行了优化升级。

算力租赁服务是指通过云服务的形式，将算力以按需付费的方式提供给用户，降低了算力使用的门槛和成本。这种服务模式不仅满足了中小企业在研发、测试阶段的算力需求，也为科研机构提供了灵活、高效的算力支持。

AI平台服务提供基于智算能力的PaaS（平台即服务）服务，主要包括人工智能应用开发工具和平台。这些工具和平台通常包含了数据预处理、模型训练、模型评估、模型部署等全生命周期的管理功能，帮助开发者快速构建、优化和部署AI应用。

模型定制服务提供基于智算能力的MaaS（模型即服务）层服务，主要提供从模型定制、精调到部署的全方位服务，这种服务模式不仅提高了AI模型的准确性和效率，还为企业带来了更大的商业价值。

AI应用服务是基于智算能力的SaaS（软件即服务）层服务，SaaS层直接应用于企业业务场景中，提供人工智能分析、决策等智能化服务，供应方通常为具有AI能力的垂直行业头部企业，这类企业基于自身在特定领域的深厚积累和专业知识，构建出了一系列针对行业痛点的AI应用解决方案。

## 2、机柜托管与算力租赁是当前主流服务模式

机柜托管以及算力租赁服务是当前智算中心市场的主流商业模式。在产业发展初期，智算中心一般为算力用户和传统IDC客户提供机柜托管服务。算力用户

的服务器由用户自行解决，这类用户多为云厂、AI企业、高校或科研机构，通常单次采购规模较小。

算力租赁市场目前处于起步期，在投入使用的智算算力中，用于租赁的部分占比35%左右，其它均为自用。在租赁的算力中，约80%被用于大型模型的训练，约20%则用于小模型的训练和推理任务。在区域分布上，60%的智算需求集中在北上广深一线城市。智算算力租赁业务的兴起一是由于短期内智算资源供给能力不足，特别是以GPU为代表的AI芯片的紧缺；二是自建智算中心投资规模大、运维能力要求较高，算力使用综合成本高。大多数中小企业在发展AI模型和应用的过程中面临算力瓶颈，算力租赁业务迎来发展机遇。

从客户群体来看，AI企业、行业应用企业、高校科研机构、解决方案提供商发布的大模型数量多，但互联网企业模型落地最快。AI企业以及高校科研机构掌握AI算法技术，但极少拥有算力，通常以租赁算力为主；互联网企业AI技术积累深、平台数据和客户丰富、通常有自己的智算云；华为、小米、浪潮等硬件厂商具有硬件终端解决方案和用户数据，部分有智算设施。目前来看，AI企业和高校科研机构是智算租赁市场的主要客户群体，此外各行业头部企业也可能成为重要的潜在客户。

算力租赁市场竞争较为激烈，存在上游芯片厂商、下游AI企业、地方政府、上市公司新进入者、云厂商五大力量。AI企业现阶段集中力量开发大模型产品，市场给新进入者留下窗口期，芯片厂商、地方政府、上市公司等纷纷进入智算领域，但相较之下，云厂商、AI企业拥有客户、数据、MaaS业务，竞争力明显强于地方政府与上市公司。随着大模型竞争格局逐步明朗，云厂商将依托

平台、生态力量、规模效应全面拓展智算市场，小规模智算租赁企业将被收购，市场资源进一步整合。

#### （四）产业发展持续践行绿色低碳理念

算力绿色化发展已成为我国数字经济发展的课题。“双碳”目标提出后，国家更加重视数据中心能耗，《贯彻落实碳达峰碳中和目标要求 推动数据中心和5G等新型基础设施绿色高质量发展实施方案》、《数据中心节能诊断服务指南（2023）》、《数据中心绿色低碳发展专项行动计划》等进一步落实数据中心能效监管，对PUE等指标提出明确要求，推动产业绿色低碳发展。同时，东数西算工程的实施促进了西部地区的风能、太阳能等可再生能源的消纳，通过优化算力基础设施布局，提升算力中心绿色能源应用比例，实现能源的合理分配和利用。

2023年度国家绿色数据中心名单重点新增智算中心领域，引导产业加快绿色智算中心建设。当前，我国智算中心逐步加强绿色节能技术应用，提升算力性能和能源利用效率。在绿色IT设备方面，通过部署整机柜服务器、冷板式液冷服务器等，降低服务器能源消耗。如深圳百旺信智算中心应用一体化芯片仿真设计整机柜服务器，实现单柜算力40PFlops，提升服务器使用效率，同时达到节能效果；中国电信京津冀智能算力中心冷板式液冷万卡资源池全年PUE值为1.15，单位算力能耗为1.5kW/P。在绿色能源利用上，智算中心利用当地的绿电和自然冷源，提升可再生能源应用比例，降低智算中心能耗。例如，燧弘华创智算中心依托庆阳创新的“源网荷储碳数”电力系统和新风自然冷却技术，将PUE稳定控制在1.2以下。在智算中心运营上，采用高效间接蒸发冷却、液冷、高压



离心式变频水冷、封闭冷通道等先进制冷技术，实现智算中心散热能耗降低50%以上。部分智算中心建立智能化管理平台，实时监控机房运行状态，精确控制制冷和供电系统，推动制冷耗电量下降8%-15%，保障智算中心安全稳定与低碳运行。如中国电信安徽智算中心A1楼通过优化任务调度、负载均衡等技术，提高信息设备利用率。

面对巨大耗电量，智算中心绿色化水平仍有较大提升空间。高性能服务器单位能耗远高于传统服务器，千亿参数大模型训练阶段用电或超过百万度。在政策指导下，智算中心产业继续贯彻落实绿色低碳要求，统筹推进算力与电力融合发展，充分考虑与电力基础设施的协同布局，加强能源需求侧管理和供给侧优化，夯实国家数字基础设施，推动绿色算力发展。

The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, parallel strokes that curve and sweep across the page, primarily concentrated on the right side and bottom, leaving the left side relatively clear for the text.

第五章

AI时代算力中心建设  
开启新篇章

## 五、AI时代算力中心建设开启新篇章

### （一）人工智能发展对算力中心建设提出更高要求

高度耦合是智算中心与传统数据中心建设架构的根本区别。传统数据中心在数据中心基础设施、算力、网络通信、存储、软件与平台等环节可采用独立采购建设的模式，各环节之间协同性较弱。智算中心主要业务场景要求高并行，单一训练任务以整个算力集群为基础，因此算力基础层内部的设计和运行是高度耦合的，计算、存储及网络须紧密协同。在生产算力环节，IT规划设计须综合考虑算力用途、算力类型、算力规模，以架构的确定性应对各类异构算力需求，提升计算效率和灵活性。存储与数据方面，存储系统不仅要能够处理PB级的数据量，还要在数据采集、清洗、训练及推理的每个环节，都能提供高速、可靠的数据访问。另外，通信网络对智算集群的吞吐量和性能起关键作用，智算中心核心业务场景——大模型训练中的大规模参数对算力和显存都提出了更高的要求，智算中心需要配套建设低时延、大带宽、长期稳定性、大规模扩展性和可运维的高性能网络。

图表 9：智算中心总体架构



信息来源：国家信息中心

1、基础设施

高密机柜对智算中心基础设施整体布局规划、供电、制冷、承重等带来挑战。高密部署导致IT机房面积占比大幅缩小，为支持多元算力的混合部署，应规划好空间布局以容纳不同冷却技术和不同功率密度的算力设备。供电方面，高功率、大电流输电损耗凸显，对变压器、母排、线缆等电力传输设备要求更高，可采用电力模块替代变压器、低压配电、UPS、输出配电等多个独立的产品，或采用高压直流方式提升输配电效率。制冷方案选择要综合成本、PUE、可靠性等多方因素，风冷仍占据重要地位，液冷方案是未来主流方向。在承重方面，由于单机柜密度提升及制冷方案的变化，在设计智算中心时需要提高建



筑承重负载，合理布局功能区域和设备，并考虑未来可能的扩展和设备更新。

## 2、AI算力

智算中心常见应用场景为训练和推理，根据其对算力精度的需求的差异分为FP32、TF32、FP16、BF16、INT8等。进行智算中心算力规划应根据具体的模型参数量，计算出训练、推理场景对GPU显存大小的需求；根据GPU单卡算力和卡间通信带宽设计合理的并行方案，并用于芯片服务器的选型和集群网络方案设计。

AI芯片技术迭代速度快，远超过基础设施建设速度，智算中心算力规划应充分考虑核心硬件设备技术更新，保持与时俱进，满足实际应用需要。

## 3、高速互连网络

在满足大规模计算集群的连接需求方面，智算中心网络正迅速向十万乃至数十万卡互联演进，参数面网络的接入速率已从200GE提升至400GE乃至800GE。高性能网络有利于保障集群智算节点间的通信效率，以满足更高性能和更大规模的算力需求。低延时需求：RDMA技术通过绕过操作系统内核，让一台主机可以直接访问另外一台主机的内存，从而降低多机多卡间端到端通信时延。实现RDMA的方式有InfiniBand、RoCEv1、RoCEv2、iWARP四种，当前智算中心的RDMA技术主要采用的方案为InfiniBand和RoCEv2两种。高带宽需求：智算中心需要处理海量数据，网络设备的吞吐能力、端口带宽及密度需要全面提升，确保数据在训练和推理过程中的快速传输。稳定性需求：大模型分布式训练任务有可能需要数天或数周，网络不稳定将会影响整个训练任务的进度。可运维



需求：在成百上千张GPU卡的智算集群中，网络架构的复杂性、运维是否简单稳定并且有足够的工具来管理集群是需要重点考虑的维度。

#### 4、存储与数据

海量文件存储系统需要具备几方面的特点，一是要求弹性扩展，满足存储集群增加节点的需要，二是横向扩展时合理的数据分布，三是扩缩容稳定性、易运维。在AI训练过程中，高效的数据读有利于提升整体训练效率。大模型训练的主要文件包括样本文件、Checkpoint文件、模型文件，整体数据量很大，对读写的要求也非常高。NPU/GPU直通存储技术简化CPU内存缓冲和复制过程，大幅缩短数据读写时间。在AI推理阶段，面对高并发、长序列的推理场景，以KVCache（键值缓存）为中心的多级缓存加速技术能够优化数据访问路径，提升系统的整体性能和响应速度，从而更好地满足大模型应用的需求。

### （二）技术创新加速AI硬件国产替代

智能算力的核心是CPU、GPU、FPGA、ASIC等各类计算芯片。AI芯片内核数量多，擅长并行计算，满足AI算法所需的大量并行处理能力，并显著提升计算效率和灵活性。智算服务器是智算中心的主要算力硬件，一般采用CPU+GPU、CPU+FPGA、CPU+ASIC等异构形式，以充分发挥不同算力芯片在性能、成本和能耗上的优势。

图表 10：AI芯片（GPU、FPGA、ASIC）主要特点

|          | GPU                                       | FPGA                                     | ASIC                          |
|----------|-------------------------------------------|------------------------------------------|-------------------------------|
| 定义       | 图形处理器，用于图形渲染、并行计算、AI训练和推理                 | 现场可编程门阵列，可编程逻辑解决方案，适用于快速原型设计和特定应用        | 专用集成电路，为特定应用定制设计的芯片           |
| 能耗       | 相对较高，尤其是高性能GPU                            | 较低，根据设计可优化能耗                             | 低，专为特定任务设计，可优化能耗              |
| 并行处理能力   | 采用数量众多的计算单元和超长流水线，具备高并行结构，通过多核并行计算支撑大算力需求 | 具有高度实时性和灵活性，可以同时进行数据并行和任务并行计算            | 专为特定任务设计，高度并行化。               |
| 灵活性和可拓展性 | 灵活性较高，可通过软件编程实现不同功能；可拓展性好，可通过多GPU扩展计算能力   | 灵活性很高，可现场编程改变硬件逻辑；可拓展性好，可通过外部接口和其他系统组件连接 | 灵活性低，制造完成后功能固定；可拓展性差          |
| 可定制性     | 通用性强，可通过软件更新实现一定程度的定制；已有成熟产品线             | 半定制化，可编程灵活度高。                            | 全定制化，完全根据特定应用定制，需要较长的设计和制造周期。 |
| 成本       | 取决于性能需求                                   | 中等，取决于FPGA的规模和复杂度                        | 高，需要定制设计和制造                   |
| 应用场景     | 游戏、深度学习、数据分析等                             | 通信、军事、航空等                                | 加密货币挖矿、高性能计算、特定算法加速等          |
| 主要厂商     | NVIDIA、AMD                                | Xilinx、Intel                             | Google、华为、寒武纪                 |

信息来源：科智咨询整理

在高端芯片限制不断收紧的背景下，国产化芯片加速自主发展。AI服务器核心在于高性能芯片，参数规模、训练数据随着AI大模型的发展呈现指数级增长，相较于大量叠加服务器台数，提升芯片性能同样不可或缺。目前国内自研芯片领军企业主要包括寒武纪、海光信息、景嘉微、华为海思等，寒武纪思元370芯片、昆仑芯科技昆仑二代AI芯片以及阿里平头哥含光800芯片等在性能上进

一步缩小与国际先进水平的差距，并在特定应用场景中展现出优势；华为AI芯片昇腾910在算力上和英伟达A100性能基本相当。整体而言国内AI芯片与海外龙头企业产品仍存在较大差距，长期来看国产自研AI芯片有望逐步成为国内AI产业发展的重要支撑，有望围绕产品形成算力、算法、应用一体化的开放生态。未来智算中心将集成更多的人工智能处理能力，推动技术融合与创新。

图表 11：主流高性能芯片特性对比

| 企业   | 产品型号       | 运算能力<br>(理论峰值)     | 性能功耗比     | 峰值性能<br>(FP16,<br>TFLOPs) | 架构                 | 显存带宽      |
|------|------------|--------------------|-----------|---------------------------|--------------------|-----------|
| 英伟达  | A100       | 624TOPS<br>@INT8   | 2TOPS/W   | 312                       | Ampere             | 1935GB/s  |
| 英伟达  | H100       | 3958TOPS<br>@INT8  | ——        | 1979                      | Hopper             | 3TB/s     |
| AMD  | MI 100     | 184.6TOPS<br>@INT8 | 0.6TOPS/W | 184.6                     | CDNA               | 1.2TB/s   |
| 华为海思 | Ascend 910 | 512TOPS<br>@INT8   | 2TOPS/W   | 320                       | HUAWEI<br>Da Vinci | ——        |
| 寒武纪  | 思元370      | 256TOPS<br>@INT8   | ——        | 96                        | MLUarch03          | 614.4GB/s |
| 海光信息 | 深算一号       | ——                 | ——        | ——                        | GPGPU              | 1024GB/s  |
| 百度昆仑 | 昆仑2        | 256TOPS<br>@INT8   | 2.1TOPS/W | 128                       | XPU-R              | ——        |
| 平头哥  | 含光800      | 825TOPS<br>@INT8   | ——        | ——                        | 自研                 | ——        |

信息来源：科智咨询整理

目前国产化AI芯片仍处起步发展阶段，在科学研究和产业应用方面具有广阔的创新空间。下一代AI芯片开发需着力在以下方面获得突破：架构创新，可重构芯片、存算一体、类脑智能等新兴架构的研究突破有望提升芯片性能；感存

算一体，在传感器端开始对信号做初步处理，减少后端芯片的压力；通过融合计算单元与存储单元，减少数据访问和搬运，显著提升能效并降低功耗；新型存储器，为了降低数据频繁交换导致的延迟和功耗，以存储为中心的计算架构逐渐兴起，成为AI芯片的一大新兴技术路线；复杂AI算子支持，探索能够高效分解和映射Transformer等复杂结构的芯片架构，实现算法与硬件的协同优化；稀疏化计算能力，利用神经网络中的稀疏性，减少无效计算，优化功耗；云边AI芯片协同，边缘计算节省云端服务器带宽和算力成本，带来更多交互方式、更强的安全保护。

### （三）高效网络架构推动算网融合

提升智算中心算力利用率和集群可用度，需要构建超高速、超大规模的组网架构。超高速网络技术通过提供更高的带宽，减少网络延迟，确保大规模集群中的数据传输更加高效。超大规模组网架构技术确保规模集群中的数据传输稳定性和可靠性，能够实现十万卡以上的超高速网络互联。

当前智算中心的RDMA技术主要采用InfiniBand和RoCEv2协议。RDMA（Remote Direct Memory Access，远程直接内存访问）是一种用于高性能网络通信的技术，相比传统的网络通信方式，RDMA具有更低的延迟、更高的带宽和更低的CPU利用率等优点，可显著提高网络通信的性能和效率。RoCE方案相较InfiniBand方案通用性更强，除用于构建高性能网络外，还可以在传统的以太网网络中使用，但在交换机的Headroom、PFC、ECN等相关参数的配置比较复杂，因此在万卡以上的超大规模场景下RoCE网络吞吐性能较InfiniBand方案相对更弱。RoCE提供了一种成本效益高的解决方案，可以利用现有的以太网基础



设施，价格相对更低；InfiniBand性能突出，适合对网络性能有极高要求的应用场景。总体而言，智算中心网络架构的选择取决于具体的应用需求、预算和现有基础设施。

图表 12: InfiniBand和RoCE方案比较

|          | InfiniBand                     | RoCE                                           |
|----------|--------------------------------|------------------------------------------------|
| 协议基础     | 独立于以太网，专为高性能计算（HPC）环境设计        | 基于以太网，RoCEv2版本支持跨VLAN和IP子网通信，克服了RoCEv1的局限      |
| 性能       | 提供更高的带宽和更低的延迟，最高可达200Gbps      | 提供低延迟和高吞吐量，但通常低于InfiniBand。RoCEv2的最大带宽为100Gbps |
| 易用性和可扩展性 | 通常需要专门的硬件支持，部署成本较高             | 可以利用现有的网络基础设施，易于部署和扩展                          |
| 容错性和可靠性  | 使用星型拓扑结构和链路聚合，提供冗余路径和高级的容错机制   | 提供多路径技术和纠错重传机制，确保数据传输的可靠性                      |
| 硬件支持     | 主要由Mellanox提供支持，提供全面的硬件和软件解决方案 | 有多家供应商提供支持RoCE的网络适配器和交换机                       |

信息来源：科智咨询

“东数西算”工程实施以来，国家政策持续推动全国一体化算力网络建设。2023年12月发布的《国家发展改革委等部门关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》指出，加快建设跨区域、多层次算力高速直连网络，积极推进算网深度融合；建立跨区域算力资源调度机制，构建立体联动的算力调度体系，促进东中西部算力资源实现供需平衡。为推动算网协同、加快全国一体化算力网络建设，算力网络建设需要做到弹性、敏捷、无

损、安全、感知。算力网络需具备弹性带宽需求，适应不同计算场景的带宽变化。具备泛在算力敏捷接入的能力，提升用户的算力获取效率。

#### （四）功率密度提升促进高效制冷实践

AI芯片性能提升带动服务器及单机柜功耗大幅增加，液冷方案成为智算中心制冷更优选择。传统风冷系统的散热效率无法与高功率机柜的散热需求相匹配，一方面，受数据中心建筑面积与单位运营成本等因素的影响，数据中心单机柜功率密度将持续上升，在AIGC的推动下单机柜功率上升速度有望较预期更快。另一方面，风冷方式移热速率较低，相对于风冷散热，液冷能更好地支持20kW以上高密机柜冷却，提高换热效率，实现数据中心PUE低至1.2以下，符合机房高密度演进趋势，同时促进数据中心减少碳排放。

液冷技术类型根据冷却液和发热设备接触换热方式的不同，可划分为冷板式液冷、浸没式液冷和喷淋式液冷。相较于其他液冷技术路线，冷板式液冷应用更加普遍。据科智咨询统计，冷板式液冷方案应用比例达到91%，是现阶段及未来较长一段时间的主流液冷技术形式。2022年，冷板式数据中心市场规模达到90.5亿元，同比增长45.4%。目前冷板式液冷数据中心已形成相对成熟的解决方案，通过冷板和CDU带走IT设备超过80%的热量；不需要对数据中心机房进行大规模改造；散热效率高，可降低PUE至1.2以下；液体不与设备接触，可靠性高、易展开维护性设计，噪音低，空间利用率高、且热能可回收。与浸没式液冷相比，热交换受到冷板的限制，节能效果相对降低。但浸没式液冷和喷淋式液冷直接接触冷却液，兼容性差，专用机柜对于管路要求高，维护复杂，且存在冷却液挥发的问题，运行成本较高。

图表 13：数据中心液冷技术方案对比

|         | 冷板式液冷               | 浸没式液冷                   | 喷淋式液冷                          |
|---------|---------------------|-------------------------|--------------------------------|
| 接触形式    | 冷却液不接触发热体，采用导热板传热   | 冷却液浸泡发热体                | 冷却液喷淋发热体                       |
| 建设成本    | 主要成本在换热系统和冷却系统，成本适中 | 冷却液用量较多，与冷板式相比，成本高      | 通过改造机柜增加必须装置，成本较小              |
| 服务器改造成本 | 增加冷板，成本较低           | 改造成本高                   | 改动小，成本中等                       |
| 运维成本    | 运维成本低               | 运维成本适中                  | 运维成本适中                         |
| 冷却液     | 冷却液用量小，要求低          | 冷却液使用量大，对冷却液安全性要求高      | 冷却液用量适中，要求高                    |
| 服务器兼容性  | 根据服务器定制冷板，兼容所有机器    | 根据冷却液不同，需进行兼容性测试        | 根据冷却液不同，需进行兼容性测试               |
| 空间利用率   | 较高                  | 中等                      | 中等                             |
| 噪音程度    | 较低                  | 低                       | 较低                             |
| 环境影响    | 无                   | 冷却液相变过程可能会导致气体蒸发外散      | 冷却液雾滴和气体可能散发到机箱外               |
| 冷却效果    | 较好                  | 优秀                      | 优秀                             |
| 使用场景    | 无限制                 | 有限制                     | 有一定限制                          |
| 应用程度    | 目前应用最广泛             | 适用于对功率密度、节能性要求较高的大型数据中心 | 不适合高密度服务器和超大规模数据中心，现阶段落地应用相对较少 |
| 主流供应商   | 超聚变、浪潮              | 曙光数创                    | 广东合一                           |

信息来源：科智咨询

液冷方案在算力中心的规模化应用仍处于初级阶段，目前为小规模部署。经前期积淀，液冷产品及技术解决方案基本成型，但最终技术选型还需要经过一定周期的验证和市场筛选。超算中心及部分头部互联网企业开展液冷试点应

用，积极寻求技术验证、产品适配等方面的生态协作，为液冷大规模商用提供宝贵经验和良好的示范带动作用，加速液冷产业发展。

图表 14：中国高性能算力中心先进制冷方案应用案例（部分）

| 项目              | 制冷方案  | 说明                                                                                                                                               |
|-----------------|-------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| 天府智算西南算力中心      | 风冷    | 项目使用42kW智算风冷算力仓，采用冷热通道全密封+列间空调高温进水模式，将冷热气流循环与外界环境进行隔离。相比传统数据中心，进水温度提高至18℃，制冷系统能效比提升10%以上。通过AI智能管控平台实现制冷系统联动调优，整体节能25%以上，风冷散热PUE大幅降低。             |
| 中国建设银行武汉智算中心    | 风冷+液冷 | 采用风冷+液冷混合散热方案，其中对主要发热源CPU和内存进行液冷设计，解决服务器90%的散热问题。在机房空调和冷板液冷技术双重影响下，CPU核心的温度低于常规风冷机型，保障CPU在低温条件下实现超频运行。                                           |
| 厦门大学嘉庚实验室智算数据中心 | 风冷+液冷 | 采用冷板式液冷系统+风冷列间空调的制冷系统架构，风液比80%，整系统按2N架构进行配置。PUE实时智能自动化调优，降低数据中心运行维护成本，实现节能减排。                                                                    |
| 京东云华北（廊坊）数据中心   | 冷板式液冷 | 整机柜冷板式液冷技术将室外冷源直接注入服务器主板芯片进行冷却，达成系统去冷机化。CPU散热由冷却液通过室内板式换热器换热后，可直接利用室外冷却塔进行散热，实现全年运行PUE低于1.1，基础设施能耗节省可达30%，对应碳排放总量减少10%以上。                        |
| 中国联通石家庄绿色智算中心   | 浸没式液冷 | 芯片满功耗平均核温 $65\pm2^{\circ}\text{C}$ ， $\text{PUE}\leq 1.1$ ，在灵活扩容、模块组装、安全可靠的前提下，结合先进的温控系统和管路布局等设计，有效解决服务器高热流密度的散热难题。在复杂环境和空间受限等非理想条件下实现全链路集成快速部署。 |
| 中国电信中部智算中心      | 浸没式液冷 | 通过嵌入式液冷单元，盲插总线架构，实现全场景 $\text{PUE}\leq 1.15$ ，实现绿色低碳，可支持部署万亿参数大模型。                                                                               |

信息来源：科智咨询整理

未来，随着液冷技术的逐渐成熟，浸没式冷却带来的液体泄漏风险和运维成本预计逐步降低，液冷系统的其他配套设施也有望因技术进步和规模生产推动成本进一步下降。全栈式液冷有望在更多的领域得到应用，从硬件到软件，整

个系统设计更加高效节能。企业和研究机构将积极开展长周期、大规模的液冷部署试验，收集和分析实际运行数据以评估液冷技术的性能和可靠性，发现并解决潜在的问题和风险。更多行业标准的制定与落地带动产业链有序发展，提升产业价值链地位。各设备和服务厂商正积极推动行业标准化，全力打造更开放的液冷生态。

## （五）新型供配电系统重塑能源应用

随着机柜功率密度的增大，智算中心供电架构要求逐渐发生变化，智算中心的规划建设将更加注重供电系统的容错性、可靠性，确保系统正常运行。智算中心供配电架构设计需要统筹考虑相关设备的负荷容量、上下级开关保护配合、供电系统的检修计划、故障隔离等因素，并优化拓扑结构设计，配置完善的自动控制系统，提高系统可用性。

采用更灵活可靠的电力架构提升供配电系统效率。一是一体化电源/电力模块。传统数据中心的供配电系统普遍采用分散式的结构，设备分散部署，占用面积大，线缆线路长、损耗大，后续的线缆检修和系统扩容难度高。一体化电源或者电力模块设备将中压柜、变压器、UPS、HVDC、馈电等预制集成交付，缩短供电链路和施工周期，使供电系统的供电效率提高约1%。二是直流供电模式。2N HVDC供电架构中，每台高压直流电源设备均可独立承担系统的全部负载，提供了最高级别的冗余和可靠性，相比传统的交流电源系统通常具有更高的转换效率和更低的能耗。10kV交流直转240V直流供电架构中，移相变压器柔性集成了10kV交流的配电，隔离变压、模块化整流器和输出配电等环节，优化供电链路，系统容量可根据需求进行灵活配置。三是DR/RR供电架构。传统数



据中心主流的供电架构是2N架构，RR架构为N+1配置，可使智算中心供电系统的利用率从50%提升到 $N/(N+1)$ ，同时占地面积减少 $(N-1)/2N$ 。DR架构中配电设备使用减少接近25%，电源设备成本低，但每套系统需要物理隔离，土建成本较DR架构高。

在人工智能时代，智算中心供配电系统向融合、绿色、智能方向发展。超融合设计下的一体化电源缩短供电链路长度，集约高效、交付周期短，具有重要的应用和推广价值。结合源网荷储、AI调优、智能电网等技术，提升智算中心能源利用效率和新型电力系统需求侧资源优化配置效率。通过源网荷储一体化的方式，将新能源发电的不连续性和用电负荷的稳定性要求相匹配，提升绿电使用比例和系统运行效率。利用AI技术进行电力负荷预测、故障检测和优化调度，结合先进AI算法实现对输电线路的精准分析和故障预警，提升供配电系统管理的科学化与精细化水平，降低维护成本。智能电网技术有利于智算中心供配电系统与外部电网进行更为紧密的互动与协作，优化电力调度，提高电力使用效率和可靠性。

The background of the page is decorated with flowing, wavy lines in a light blue color, creating a sense of movement and modernity.

## 第六章

# 智算中心运营服务 向高效安全方向发展

## 六、智算中心运营服务向高效安全方向发展

### （一）算力资源高效调度

为高效利用各类计算资源，优化算力使用效率并降低成本，需要在算力集群中对计算资源进行分配和管理，即算力调度。算力调度以算力感知为基础，感知并整合算力节点收集的算力资源信息，统一表达为包含计算、网络、存储等多维度资源的综合模型，利用算网编排技术实现复杂算力业务的路径编排，按需分配、实时调度不同区位、不同运营主体、不同架构的算力资源。算力交易平台连接算力买方与卖方，根据用户的差异化需求，实现智能、公平、开放、公平、安全、可溯的算力交易。

算力调度体系强调按需分配和灵活调度计算资源、存储资源及网络资源，并加强网络安全管理。支持弹性分配以适应任务不同、变化多端的动态需求，加强算力、网络和安全系统间的协同防御，利用人工智能提升安全技术手段与应急处置能力。为了提高资源利用率和系统性能，算力调度系统需要实现负载均衡，并利用虚拟化技术实现资源的隔离和共享。

全国算力调度统筹全国不同地区算力，促进东西部算力资源高效配置，达成全国算力基础设施化的目标。从算力调度的范围来看，算力调度涉及区域内算力调度、跨区域算力调度、云边端协同调度及算力和电力的协同调度。以城市算力网建设为基础，继而“连点成片”形成区域算力网，各区域算力网按照统一标准互联互通，最终形成全国一体化的算力网体系，实现渐进式落地。

图表 15：算力资源调度与管理架构



（二）数据安全多重保障

智算中心是提供算力服务、数据服务和算法服务，推动人工智能产业发展的算力基础设施，其安全可信环境的构建关乎个人权益、产业发展和社会运行。智算中心应全面推进数据安全技术应用，通过加密技术、匿名化技术等，加强数据安全保护和管理，实现对多种资源的全方位隔离与防护。利用深度包检测（DPI）技术、数据丢失预防（DLP）等对网络流量、行为日志、数据流转、共享接口等安全监测分析。利用先进的人工智能和机器学习技术提高网络安全防护系统的智能水平，实现智能化分析、防御、自主捕捉及未知风险应对，提供更加动态和前瞻性的安全保护。

智算中心可通过建设内外部安全管理体系保障数据安全。在智算中心内部，加强网络基础设施的可靠性和高效性，同时积极利用数据安全技术如身份识别、数据加密、数据脱敏、数据流动管理等，以及数据服务与内容安全技术如



数据交易技术管理、开放共享管理、人工智能深度防伪、数据内容合规等，提高智算中心从数据采集到算力应用层面的安全可靠。在外部层面，加强数据中心异地容灾备份建设，利用分布的公有云，将新建IT系统和容灾系统同时部署在云上，降低数据丢失风险，保障业务连续性和数据资产安全。

### （三）运维平台智能可视

通过实时监控及时发现和解决设备故障，确保智算中心不间断平稳运行。运维监控需兼顾实时性和准确性两大要素，若监控频率过高，可能会对系统性能造成不利影响，反之，若监控间隔过长则可能导致关键事件的遗漏。智算中心运维监控与故障处理应更加高效，需要进行跨域故障感知和分析。通过对系统运行状态的持续监测，监控智算中心计算、存储、网络、光模块等设备的运行状态、温度、湿度、能耗等信息。借助前沿故障检测技术，如日志分析、分钟级网络流量分析、存储故障和性能分析等，迅速准确确定故障位置，从而缩短故障诊断时间，最大化防止训练任务中断。建立高效故障修复流程，发生故障时能够立即采取相应措施恢复系统的稳定运行，缩短系统停机时间。根据实际需求定制化运维平台配置，包括监控指标、报警规则、数据展示方式等，建立简洁、直观的界面，便于运维人员快速响应并处理问题。

智能运维以数据为基础，以算法为支撑，以场景为导向，为数据中心运维管理提供智能化解决方案，实时分析和处理问题。大型超大型智算中心承载的业务和数据量巨大，智能化运维体系的构建有利于提升智算中心运行的稳定性及能源利用的高效性。自动化平台运维是流程化、标准化运维的进一步发展，自动化运维工具通过平台建设，将大量重复性的运维活动转化为自动化操作，提



高运维效率，增强运维过程的可视化。

智算中心智能化运维解决方案应具备如下特点。一是统一运维，提供智算中心一站式IT服务。二是数据驱动，聚合独立的运维数据形成运维大数据，并结合AI分析能力，实现对智算中心运维态势的分析及可视化，支持运维决策。三是全局可视化，提供从平台层到底层硬件基础设施的全栈监控可视化。四是自动化及智能化，提供自动化工具链，并对智算中心运行态势进行智能化预测，形成故障应对机制，快速修复漏洞。五是安全合规，保障智算中心运维过程中的数据安全、隐私保护、合规管理等方面符合相关法规和标准，降低法律风险。智能运维方案的核心部分为智算中心基础设施管理（DCIM）、数字化运维服务管理平台（DOSM），并在此基础上搭建精细化运维工作管理架构，提升智算中心智能化水平和运营效率。

The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, overlapping curves that sweep across the page from the bottom left towards the top right.

# 第七章 中国智算中心产业 发展展望与倡议

A solid, vertical blue bar is positioned on the right side of the page, aligned with the chapter title.

## 七、中国智算中心产业发展展望与倡议

### （一）中国智算中心产业发展展望

#### 1、中国智算中心产业发展前景

智算中心产业发展环境持续优化。政策推动新型智能基础设施建设，以更好支撑社会经济发展需求，包括AI软硬件、智算服务、大模型应用在内的产业链各环节有望获得更多政策支持。算电协同建设推进，新型电力系统与算力枢纽节点资源将进一步融合衔接，保障电力资源供给，促进算力绿色化高效运行。全国一体化算力网络建设全面提速，为智算中心提供泛在连接、灵活高效的互联网络。

AI应用规模落地激发智算需求增长。AI大模型商业化应用场景不断拓展，前沿数字技术逐步转向大规模行业应用阶段，参数规模快速从亿级突破至万亿级，产生大规模智算算力需求。近年来中国智算中心建设火热，目前全国已有超过30个城市正在积极推进智算中心建设或规划。未来随着人工智能应用场景的持续创新和拓展，智算中心将更好赋能社会经济发展。

综合的智算生态服务将成为主流模式。随着AI技术的不断演进与智算生态的日益成熟，智算中心各服务类型之间将实现更加紧密的协同与融合。综合技术能力强的科技公司或具有IT基础的智算中心集成方有望基于完善的智算中心生态，提供“IaaS+PaaS+MaaS+SaaS”的全栈式综合智算生态服务。深度整合的服务模式不仅能够提升服务效率和用户体验，也将促进技术创新与产业升级的良性循环。

多元普惠的算力生态体系逐步构建。AI软硬件协同性进一步提高，支撑智算服务和应用。大模型应用呈现多元化趋势，从互联网、云计算等领域拓展至传统行业，数实融合加速。智算中心服务模式多样且更加成熟，满足不同类型的智算需求。基于分布式云技术和算力调度平台，实现异构算力资源跨区域灵活整合，算力成本逐渐降低，算力资源获取便利性提高，算力服务更加普惠化。

智算中心产业发展更加绿色安全。产业发展将更加注重可持续，充分利用清洁能源，逐步加大分布式供能、可再生能源应用比例。加强绿色设计，推动软硬件协同联动节能，推广液冷、近端制冷等新技术，降低能源消耗和减少环境污染。加强网络安全及数据安全保障能力，通过强化安全技术手段、实施安全风险评估等方式，提升算力设施运行安全性。

## 2、中国智算中心产业发展机遇与挑战

### （1）发展机遇

随着自动驾驶、工业互联网等数字化场景落地以及AI大型模型商用，前沿数字技术已经从研发阶段转向实际应用阶段。未来，数字经济与实体经济进一步融合发展，推动智算中心产业走向成熟阶段。

技术创新助力智算中心产业升级。智算中心为新兴技术的研究、验证、创新、推广提供高性能的计算和存储能力以及专业的数据处理和分析工具，新兴技术为智算中心产业提供高效、智能、安全的运行方式。在算力基础设施层面，未来更加关注整体系统架构的灵活性和高效性，从单一技术的升级转变为全生命周期技术架构的设计及优化，促进产业创新发展。



复杂场景计算推动智算中心生态完善。为支撑多任务复杂场景行业应用，部署具有多模态数据、知识深度融合特性的重点领域专用大模型，智算中心软硬件适配能力有望不断提高。推动人工智能大模型与自主可控芯片、智算中心与国产开源深度学习框架开展适配，突破异构算力适配、异构算力调度等关键技术，建立可兼容底层硬件差异的异构开发平台，通过开源开放方式推动智算中心产业生态加速完善。

智算中心建设带动设备设施市场快速增长。智算中心建设提速，项目落地带动上游市场规模增长。据科智咨询统计，2023年至2024年上半年中国智算中心项目总计超过300个，投资规模超百亿元，服务器、交换机、存储、安全、供配电、AI平台软件等细分需求采购项目数量快速增加，相关领域市场规模增长前景可观。数据中心绿色低碳发展要求下，智算中心绿色产品和解决方案将进一步推广，法律法规对保障数据安全、网络安全的管理措施和技术措施提出更加严格要求，液冷、供配电、信创、网络安全等多个相关行业发展潜力巨大。

## （2）发展挑战

在算力基础方面，高端算力芯片供给不足影响算力发展进程。美国不断收紧对华AI芯片出口政策，算力端核心环节受限，高端算力面临供给瓶颈。在算力生态方面，软硬件协同适配影响算力资源利用效率，不同厂商和架构的服务器往往不兼容，造成算力使用和应用优化困难等问题。同时，模型开发应用难度较大，算力使用门槛较高，模型与需求的适配能力不利于算力应用的繁荣及算力生态完善。在算力运营方面，人工智能时代数据中心行业用电量大幅增加，智算中心绿色化成为产业高质量发展的重要课题。未来智算中心产业需要平衡



算力发展和绿色低碳，优化能源利用效率。

## （二）中国智算中心产业发展倡议

### 1、积极推进算力基础设施建设

积极引导算力设施建设。适度超前建设数字基础设施，系统化推进5G、千兆光网、算力中心建设，提升信息通信网络设施、算力中心覆盖深度和广度。立足人工智能创新应用增长需求，京津冀、长三角、粤港澳大湾区等面向区域发展战略有序建设智能算力设施，西部地区探索建设大型人工智能训练算力设施。

推动算力供给规模化集约化。支持通用数据中心向智算数据中心转型升级，有序提高智算数据中心规模，全面提升先进算力占比。优化算力区域布局，统筹推进大型智算中心及边缘智算中心建设。推动运营商、云服务商和各类算力平台智能算力与通用算力协同发展，满足各类业务算力需求，加快形成多云互联算力资源调度体系，探索构建算力交易服务机制。

### 2、加强先进技术创新与推广

加强先进技术创新与推广。加大光通信、毫米波、5G增强、量子通信等基础技术研发力度，加速人工智能、区块链、数字孪生、虚拟现实等新型技术创新突破。围绕基础架构、训练算法、调优对齐、推理部署等环节，研发人工智能通用大模型。推动部署重点领域专用大模型，支撑多任务复杂场景行业应用。

加快实现软硬件高质量自主可控。推动ARM、RISC-V、X86等基础架构以及

GPU、ASIC、FPGA等国产异构智能算力研发进程，推动芯片全面兼容国产训练框架。提高算力中心软硬件适配能力，鼓励人工智能大模型与自主可控芯片、智算中心与国产开源深度学习框架开展适配，提高供应链安全性和协同性。

### 3、提升智算中心绿色低碳水平

提升可再生能源利用水平。将智算中心可再生能源利用目标和方案纳入节能审查标准，逐年提升智算中心可再生能源利用率。在气候条件适宜、绿电供应充足的地区充分利用各类冷能资源提高智算中心能效。采用源网荷储等技术，促进智算中心与风电、光伏等可再生能源融合开发、就近消纳，探索开展绿电直供。鼓励参与绿电绿证交易和电力需求响应。

促进绿色节能技术应用。因地制宜推动液冷、蒸发冷却、热管、氟泵等高效制冷技术应用，推广新型超低损耗光纤、碳化硅射频器件、高能效变频空调等高效产品设备。开展高效变压器、直流供电等供配电系统更新换代，鼓励采用模块化机房、预制化电力模块等节能产品。支持应用全闪存、存算分离等存储技术，降低存储系统能耗。积极推进人工智能运维平台建设，实现制冷系统联动调优，落实节能减排目标。

### 4、深化算力赋能行业应用

促进算力与实体经济创新融合发展。推动人工智能在工业制造、医疗健康、交通运输、金融服务、公共服务、能源等领域的应用，利用大数据和AI提高效率、降低成本、增强用户体验。搭建行业应用服务平台，赋能中小企业业务数字化转型。建立包含数字政府、智慧交通、智慧医疗、智慧教育、智慧文旅等

在内的智慧城市体系，通过智慧城市建设推进数字产业整体数智化水平，实现人工智能算力供给端与需求端充分有效对接。

加强算力应用保障。依托人工智能公共算力服务平台，通过算力券等模式降低科研机构、科创企业、初创企业、中小微企业等主体的算力使用成本，激发算力资源需求。提升算力资源利用率，为算力供给方、算力调度运营方、算力需求方等提供优惠支持，全面释放算力价值。优化数据资源开发利用，建立数据流通规则和运营机制，形成存储、共享、交换、训练、加工、交易等数据服务体系。加强数据加密、传输及验证等技术应用，确保数据安全、隐私保护和合规性。

## 5、构建开放协作的产业生态

建立和完善算力产业标准体系。结合我国智算中心产业发展实际情况，制定合理的技术标准、管理标准与服务标准，提高产品质量和服务水平。构建并推动产业相关发展标准在算力基础设施建设中的应用，以实现智算中心产业可持续发展。积极参与国际算力产业标准的制定和修订工作，提升我国在国际算力产业标准领域的话语权和影响力。

构建开放协作的智算中心产业生态。发挥和强化龙头企业引领作用，带动关键技术创新突破和跨领域应用。制定统一的行业标准和规范，实现不同硬件、软件、模型等以及企业、行业间的互联互通。建立开放的交流与合作机制，学习国际先进技术和发展经验。加强人才培养，促进产学研融合，提高产业链各环节的技术水平和创新能力，为人工智能发展提供坚实技术和人才基础。

The background of the page is decorated with abstract, flowing blue lines that create a sense of movement and depth. These lines are composed of many thin, overlapping strokes, giving them a textured, almost ethereal appearance. They sweep across the page from the top left towards the bottom right, framing the central text.

附录

中国智算中心  
典型项目案例



## 附：中国智算中心典型项目案例

### | 世纪互联智算中心投资运营案例

世纪互联怀来国家算力枢纽示范基地位于河北省张家口市怀来县，建筑面积64,000平方米，规划总容量约200,000kVA，规划总IT电力70MW左右，远期具备持续可扩容能力。其中，1#数据中心项目建筑面积28,600平方米，规划设计1,300余架24kW机柜，承载智算服务器约5,000台，搭建人工智能算力平台，可提供10,000P以上算力规模，总投资约15亿元人民币。

项目采用最新的云计算和人工智能技术，提供高效、稳定、安全的算力服务。超大规模万卡集群的人工智能算力平台具备大规模并行处理能力，通过集群管理和算力调度实现资源的高效分配和利用，优化作业调度，提高整体的计算效率。同时，在数据中心的设计和运营中引入绿色节能理念，降低能耗，提高资源利用率。围绕GPU基础设施创新，推出业内首个第三代全新架构的风、液组合高强冷却解决方案，能够大幅降低PUE至区域内领先水平，在控制项目成本的同时极大提高数据中心运营效率。项目融合云计算、大数据和人工智能等多种前沿技术，打造创新的智算产业园模式，获得“全国一体化大数据中心协同创新体系绿色示范项目”称号。



## | 合盈数据智算中心投资运营案例

为适配万卡级超大规模智算集群部署，合盈数据打造集中、绿色、高效、快速的零碳AIDC基础设施解决方案，构建全国首个单体最大绿色算力集群。2024年，合盈数据超大规模智算中心已交付部分客户，项目仍在陆续交付中。

集中：1、算力集群占地1000亩，多个50MW以上独立单元聚合，满足智能算力扩容需求。2、AI算力严格受通信距离限制，以网络为中心持续迭代。3、集群专属市政管廊，自建变电站，超2N高可靠架构内部环网给水系统，保证电力安全。

绿色：环京区域最大“源网荷储一体化”碳中和示范项目-合盈数据（怀来）科技产业园将实现1000MW的负荷规模。在张家口区域，源侧规划了553万千瓦的风电、光伏项目，建成后预计每年将产生不低于100亿度的绿电。总计铺设515.46km外线，确保绿电消纳。电源侧配储，构建共享储能和独立储能系统提高能源利用效率。

高效：弹性设计，打造适配不同GPU特性的多种制冷方案、高负载和高电能利用率的供电架构、AI算力高可用的运维体系，包括间接蒸发自然冷却、一体化电力模块、AI全天自我调优等。液冷产品通过前期的预研和验证，进行大规模部署。全链路智能可视，并配有自动巡检机器人，运维效率提升35%。

快速：采用预制模块化建设方式，建筑垃圾少，无噪声。从工厂生产到最后业务上线仅需6个月，保障AI算力业务快速扩容。

## | 昆仑技术智算中心产品和解决方案案例

河南昆仑技术有限公司（以下简称“昆仑技术”）依托自主创新的软硬件研究中心、先进智能制造以及开放的联合创新实验室等核心能力，聚焦通用计算、AI异构计算、基础软件与数据服务等领域，提供稳定可靠、安全可信、绿色可持续的IT产品与创新解决方案。

昆仑技术为国家石油天然气管网集团有限公司提供“鲲鹏+昇腾”解决方案，打造机器视觉AI中台。采用全国产的KunLun训练/推理服务器及昇腾AI推理卡Atlas 300V Pro，满足自主可控要求；构建工业视频监控统一的基础设施底座，解决碎片化分布、资源分散和调配困难等现有问题，提高系统的效能和资源利用率。通过构建统一视频融合平台，实现统一的视频数据接入、视频数据处理、共享和访问控制，促进业务协同合作。通过强化场景监控，加强施工作业、管道线路和场站等视频监控，保障国家能源管道安全。

- 集团中心节点：针对视频管理，在中心节点部署鲲鹏服务器KunLun 2280；针对算法管理，在中心节点部署具备超高算力的昇腾AI大模型训练服务器（910 Pro B）、昇腾推理服务器G2280+昇腾AI推理卡Atlas 300V Pro。
- 区域边缘节点：针对智能解析，在区域节点部署“昇腾推理服务器G2280+昇腾AI推理卡Atlas 300V Pro”、以高效执行AI推理分析及视频流编解码处理。
- 视频防火墙：对视频数据访问进行控制，满足视频网—办公网跨网安全传输要求。

## | 盛达全球智算中心服务案例

2024年，盛达全球协助某智算中心顺利完成基础环境迁移工作。本次项目规模较大且技术要求较高，涉及国内知名厂商的数千台GPU设备及其相关组件，总计超过12万件设备。最终，项目以0事故、0异常的成绩顺利完成并交付，获得客户多个部门的高度认可。

盛达全球通过多项技术和服务的创新，充分保障设备物理迁移的安全性及高效性。项目采用了以下关键技术和方法，一是最新的自研运输系统，通过5G通信技术实现了对搬迁过程的智能监控和数据安全保障。二是通过物联网(IoT)技术，实现设备智能识别与自动化跟踪，关键指标的直观展示和定制化的数据分析报告显著提高了库存管理和设备定位的效率。此外，低碳定制化包装理念的引入，增强了项目的安全性能并降低碳排放。

在项目风险管理方面，通过建立定制化、综合性智算中心项目风险评估体系，实时评估与跟踪搬迁过程中的各种风险，并提供相应的应对策略。全面、详细的应急预案确保在项目执行过程中出现突发情况时，能够迅速有效地响应。

另外，项目执行全过程中确保物流服务的透明化，客户可以通过盛达全球自研可视化系统，在手机应用程序上实时查看设备状态及视频。数据导出功能也为客户后续对数据进行深入分析提供了便利。



地址：北京市海淀区中关村南大街31号神舟大厦5层

电话：13611072586

网址：[www.cidc.org.cn](http://www.cidc.org.cn)