

产业研究深度|科技



谁拿走了算力的利润？ 全国一体化算力网下的商业 模式与企业出路



为企业客户提供全方位的“产业+金融”研究服务！

核心要点

“十五五”开局，全国一体化算力网纳入国家规划纲要 109 项重大工程，与水电等并列升格为国家级基础设施，行业从“东数西算”物理布局迈入“算网融合、全域调度、算电协同”新阶段。截至 2026 年一季度，“1+M+N”调度体系接入 155 家企业、163EFLOPS 算力，八大枢纽智算规模占全国超八成，枢纽间时延低于 20 毫秒。

定价体系重构是本轮变革的主线：日均词元调用量从 2024 年初的 1000 亿跃升至 2025 年底的 100 万亿，2026 年 3 月突破 140 万亿，两年增长超 1000 倍；三大运营商率先推出词元资费套餐，计费由“卖机柜”的硬件计价转向“卖词元”的按量计量，GaaS（算力即服务）成为运营高阶形态。

利润与估值逻辑随之迁移：价值分配重心自上游硬件转向中游调度运营与下游应用，利润来源转为调度溢价与词元分成；估值锚从机柜规模、GPU 库存转向算力利用率、词元产出能力、调度技术与合规资质，具备跨域调度与行业运营能力的主体将获得更高利润份额。

但行业仍受六类结构性矛盾约束：东西部供需空间、高低端算力层级、软硬件生态、投入与回报、异构标准、算力与能耗的六类错配，叠加政策合规、供应链、技术供需、绿色运维四类风险，行业将呈矛盾博弈、螺旋升级而非线性增长。

技术层面，2025 年国产 AI 芯片出货占中国市场 41%；液冷渗透率 2026 年预计跃升至 37%；800G 光模块升级为标配，1.6T 进入规模化商用初期；算电协同在西部枢纽先行落地。竞争层面，公有云、运营商、第三方 IDC、地方国资四阵营壁垒分化、头部集中加剧，中小主体需向行业专属与边缘算力等垂直赛道突围。

对于产业链的企业决策者而言，产业链价值分配逐步由硬件向服务端转移，核心命题在于重新确立自身在 AI 产业链中的定位，推动业务从单纯算力资源出租向高附加值价值服务升级。经营主体应当把握词元计价、国产算力替代、液冷与算电协同三大发展红利；对资本市场投资人而言，需要结合产业节奏分阶段布局赛道，主动规避结构性产能过剩领域，聚焦高价值垂直细分方向挖掘机会。

分析师及联系人

分析师：李忠宇

SAC:S1130524100002
Lizhongyu01@gjzq.com.cn

内容目录

第一章 算力网升格：国家基础设施定位升级与行业痛点	7
1.1 “十五五”算力网顶层政策：从“东数西算”到“一体化调度” ..	7
1.2 算力网作为新基建核心，对经济增长的“乘数效应”更加明显 ..	9
1.3 “1+M+N”算力调度体系：让算力像水电一样流动	11
1.4 传统 AI 算力行业核心痛点：四大结构性困境	13
第二章 模式之争：四种运营模式的盈利逻辑与竞争格局	16
2.1 基础算力资源租赁：硬件计费模式的增长天花板	16
2.2 分层算力订阅：训练/推理/异构混合算力包的规模效应	17
2.3 GaaS（算力即服务）：词元计价下的全链路解决方案	18
2.4 行业专属算力定制：高合规算力集群的盈利模型	20
2.5 四大运营阵营商业模式对比与竞争格局	22
第三章 技术升级：调度、液冷、光通信、芯片与算电协同的底层冲击 ..	24
3.1 算力调度技术：异构统一纳管与跨域低时延调度需求迫切	24
3.2 高密度机房配套技术：液冷散热与高速光通信	25
3.3 算力芯片生态：国产替代加速与软件生态短板	27
3.4 算电协同技术：绿电配套与算力柔性负荷调节	30
第四章 利润去哪儿了？价值重分配、交易市场、竞争格局与估值重构 ..	31
4.1 全产业链价值重分配：从“卖硬件”到“卖服务”的利润转移 ..	31
4.2 行业生态协同：全国统一算力交易市场的形成路径	32
4.3 市场竞争格局演变：头部整合、中小承压与国产替代	33

4.4 全新估值体系：从机柜规模到词元产出能力的范式转换	34
第五章 赛道掘金：产业链高景气环节与高价值落地场景	36
5.1 产业链全景图谱与中游核心节点研判	36
5.1.1 智算中心建设：从“规模扩张”到“效率优先”	38
5.1.2 算力运营与行业格局重构	39
5.2 上游：算力硬件支撑层五大高景气配套赛道研判	44
5.2.1 液冷赛道：从“可选配置”升级为“主流配套方案”	45
5.2.2 高速光模块赛道：AI 算力传输的“神经中枢”	46
5.2.3 算力核心芯片赛道：算力迭代与互联协同的双重突破	47
5.2.4 存储系统赛道：解决算力快速增长的“存墙”瓶颈	51
5.3 下游：高算力需求垂直行业落地研判	52
5.3.1 核心增量逻辑：推理侧成为算力消耗主体	52
5.3.2 高价值行业场景落地	53
5.4 本章小结：底层矛盾约束下的行业确定性演进方向	56
第六章 战略抉择：行业风险、机会分析及行动建议	57
6.1 六类结构性矛盾：行业发展的底层逻辑	57
6.2 九条核心分析：矛盾传导下的综合判断	59
6.3 四类行业性风险：矛盾外化的约束边界	61
6.3.1 政策合规风险制约行业上行天花板	62
6.3.2 供应链外部风险约束行业技术升级空间	62
6.3.3 技术与供需错配风险形成行业产能效率瓶颈	62
6.3.4 绿色运维与数据安全风险考验行业长期运营的稳定性	62

6.4 八大确定性较高的趋势：基于矛盾演化的方向前瞻推演	63
6.5 分主体的战略建议：趋势下的实操路径建议	65
6.5.1 面向算力产业经营企业的管理层：锚定布局、升级业务、抢占红利、筑牢壁垒	65
6.5.2 面向一二级产业投资人：规避过剩赛道、重构估值逻辑、分周期布局、防控明确风险	66
6.5.3 面向政策制定者：统筹标准、引导布局、设计交易、完善激励	67
6.6 长期产业不确定性与针对性对冲策略	68
6.7 本章小结	69
风险提示	70

图表目录

图表 1：中国算力规模 2020-2028 年：智能算力（FP16）与通用算力（FP64）并行扩张	8
图表 2：八大枢纽贡献全国八成智算，核心压舱石地位稳固	9
图表 3：标准机架两年增六成，智算需求驱动扩张	11
图表 4：全国一体化算力网“1+M+N”三级调度架构	12
图表 5：日均词元调用量两年增长超千倍	18
图表 6：国内四大算力运营阵营核心能力对比	23
图表 7：异构算力统一纳管与跨域调度技术架构	24
图表 8：液冷服务器渗透率正经历从“政策驱动”到“技术刚需”的加速跃迁	26

图表 9: 2026 年全球光模块总市场规模有望达到 228 亿美元(LightCounting 口径)	27
图表 10: 国产 AI 芯片份额首超四成, 龙头厂商领跑	28
图表 11: 国内新旧算力运营估值体系对比	36
图表 12: 中国 AI 算力产业链全景图谱	38
图表 13: 中国智能算力服务体系示意图	40
图表 14: 驱动算力产业发展的六类底层结构性矛盾	59
图表 15: 算力行业四类行业风险影响矩阵	61

第一章 算力网升格：国家基础设施定位升级与行业痛点

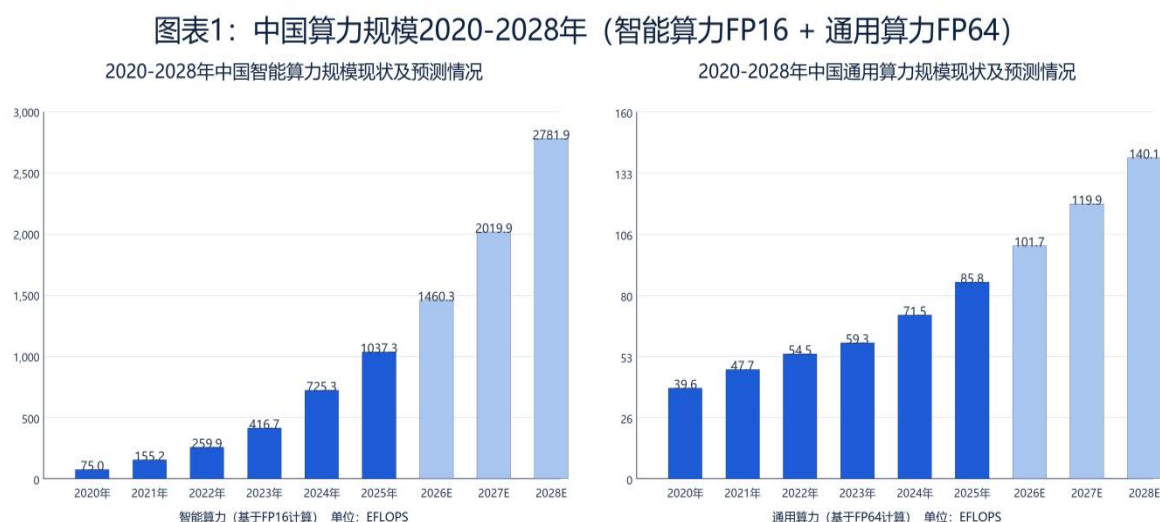
1.1 “十五五”算力网顶层政策：从“东数西算”到“一体化调度”

- 2026 年是中国算力产业发展史上具有分水岭意义的一年。“十五五”规划纲要正式将全国一体化算力网纳入 109 项重大工程，标志着算力基础设施建设从“东数西算”的物理空间布局阶段，转向“算网融合、全域调度、算电协同”的效率新阶段，我们认为这是一次政策底层逻辑的系统性升级，有利于我国中长期算力产业持续发展。
- ❖ **回溯政策演进脉络：**2022 年 2 月“东数西算”工程正式启动，在京津冀、长三角、粤港澳大湾区、成渝、内蒙古、贵州、甘肃、宁夏 8 地布局国家算力枢纽节点，规划 10 个国家数据中心集群。2023 年 12 月，国家发展改革委、国家数据局联合印发《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》（发改数据〔2023〕1779 号），首次提出从单一地理布局优化向“算力网络化、调度一体化、资源市场化”转型。2026 年全国两会政府工作报告首次将“算电协同”纳入顶层新基建部署，原文明确提出实施超大规模智算集群、算电配套工程，同步推进全国一体化算力监测调度体系搭建、扶持公共云产业发展。同年 4 月 28 日中央政治局会议统筹规划水网、新型电网、算力网等六大国家级基础设施，正式将算力网与水电等传统公共基建摆在同等战略定位，算力公共资源属性获得顶层确认。
- ❖ **从政策体系维度观察，**本报告对国内算力政策体系归纳概括出“1+1+N”政策体系，非官方文件原生专有名词。即国家“十五五”规划纲要构筑了“1+1+N”算力政策框架：第一个“1”是规划纲要本身作为顶层设计纲领；第二个“1”

是以《实施意见》为核心的实施框架；“N”是涵盖算力互联互通、绿色算力、异构算力调度等领域的配套文件体系。2026 年密集出台的配套文件建立了八大算力枢纽差异化管控机制，国家枢纽节点外原则上不得新建各类大型或超大型数据中心，进一步推动各类新增算力向国家枢纽节点集聚，将国家枢纽节点打造成为国家算力高地，行业算力供给格局正在经历深刻重塑。

- ❖ 在政策和需求的双强驱动下，中国算力规模快速扩张（参考图表 1）：2025 年智能算力规模达 1037.3EFLOPS（FP16 口径），通用算力规模达 85.8EFLOPS（FP64 口径），预计 2028 年智能算力将突破 2781.9EFLOPS，2023—2028 年复合增速约 46%。国家发改委政策研究室副主任李超在 6 月新闻发布会上表示，“十五五”时期将更加注重供需适配，加强算力网与新型电网、新一代通信网规划建设的协同联动，在“硬投资”层面探索算电协同模式，做到“以电强算、以算促电”软建设”层面强化算力资源监测与市场化调度。

图表 1：中国算力规模 2020-2028 年：智能算力（FP16）与通用算力（FP64）并行扩张



来源：观研天下《中国算力基础设施行业现状深度分析与发展前景预测报告（2026-

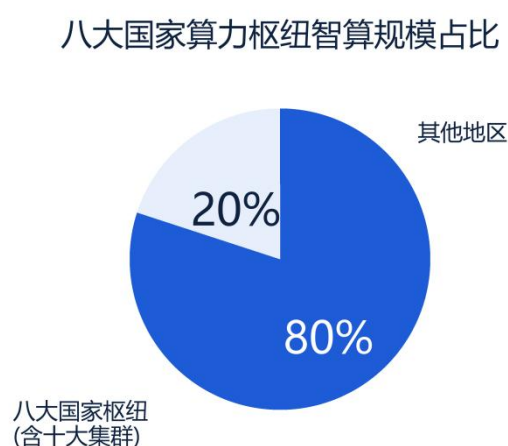
2033 年）》、国金证券研究所

注：数据为商业咨询机构预测值，仅供产业研判，不等同政府统计目标。

1.2 算力网作为新基建核心，对经济增长的“乘数效应”更加明显

- “十五五”规划周期内，算力网作为新基建核心，对经济增长的“乘数效应”更加明显。国家数据局局长刘烈宏在中国发展高层论坛 2026 年年会上透露，预计到“十五五”末，中国人工智能相关产业规模有望达到 10 万亿元区间，算力网建设不仅是基础设施投资，更是撬动万亿级智能经济生态的战略支点。
- ❖ 八大算力枢纽差异化布局与东西部算力分工明确。全国一体化算力网已形成以 8 大国家算力枢纽（如图表 2 所示）、10 个国家算力集群、3 个算电协同发展区域为重点的“8+10+3”算力空间布局。截至 2025 年底，八大国家枢纽（含十大集群）已建成智算规模达 138.8 万 PFlops，占全国比重超过 80%，是算力供给的核心压舱石。

图表 2：八大枢纽贡献全国八成智算，核心压舱石地位稳固



来源：国家数据局、国金证券研究所

- ❖ 东西部算力分工格局清晰。东部地区依托人才、技术、产业优势，聚焦低时延推理算力、前沿算力技术与行业应用创新，是算力需求的发源地与价值高地；

西部地区依托土地、绿电、气候优势，重点承接大规模模型训练、数据存储、离线分析等非实时算力需求，构成全国算力供给的大后方。具体来看：

- ❖ **京津冀枢纽以张家口集群为核心腹地，聚焦政务算力、金融算力等高合规场景，同时承担前沿 AI 技术研发算力供给。**长三角枢纽以芜湖集群为依托，服务金融科技、智能制造等低时延实时业务，是全国算力消费最密集区域。粤港澳大湾区以韶关集群为载体，聚焦互联网、消费电子、智慧城市等应用场景，算力需求增长较快。成渝枢纽以天府数据中心集群为核心，辐射西南地区数字经济发展，同步扩容省级算力调度平台。
- ❖ **西部五大枢纽进入建设高峰。**内蒙古枢纽作为京津冀枢纽的算力腹地，和林格尔新区规划多座规模化智算集群项目，预期打造区域规模较大的算力产业集群。甘肃庆阳依托风光资源与低温气候优势，重点承接东部离线训练、数据存储等非实时算力需求，成为西北算力走廊核心节点。贵州、宁夏枢纽持续深化绿色数据中心建设，依托水电、风电等可再生能源优势，打造低 PUE、高绿电占比的算力承载区。
- ❖ 根据工信部披露的数据显示，全国对外商用算力中心标准机架规模快速扩张，从 2024 年底的 880 万架增长至 2025 年 9 月的 1250 万架，并预计在 2026 年 3 月达到 1445 万架（参考图表 3），两年内增幅超六成，直观印证了智算需求对基础设施的强劲驱动。

图表 3: 标准机架两年增六成, 智算需求驱动扩张



来源: 工业和信息化部, 国金证券研究所

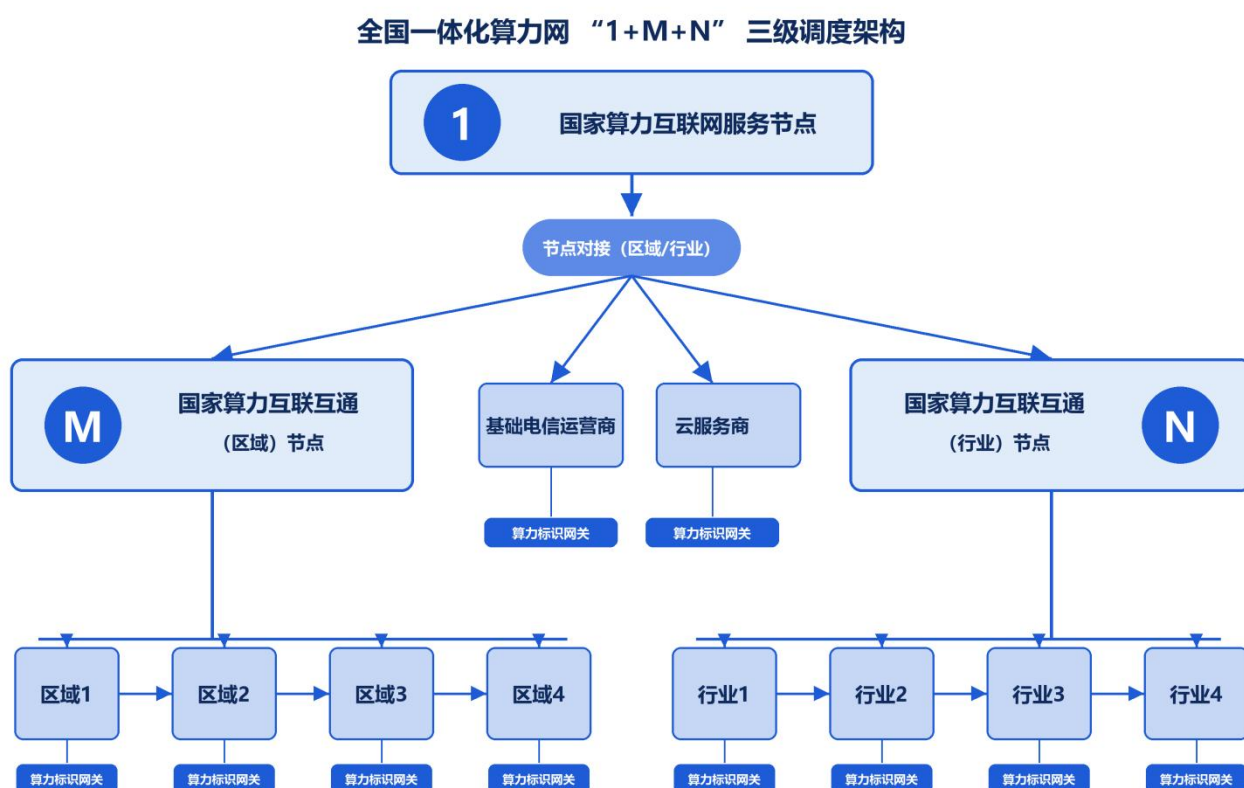
注: 数据为全国对外商用算力中心标准机架, 不含企业自用机房机架

❖ 在网络互联方面, 围绕算力枢纽已建成干线光缆 145 条, 将 8 大算力枢纽间互联, 全国各省份至 8 大枢纽的网络传输时延均降至 20 毫秒以内。中国信息通信研究院总工程师敖立表示, 我国已基本建成国家枢纽的两级时延圈: 国家枢纽间传输时延不超过 20 毫秒, 区域集群到周边主要城市不超过 5 毫秒。这意味着, 位于上海的 AI 企业可以实时调用内蒙古的智算中心进行模型训练, 网络时延已不再是制约跨区域算力调度的瓶颈。

1.3 “1+M+N” 算力调度体系 算力像水电一样流动

国家工业和信息化部于 2026 年 2 月印发《关于组织开展国家算力互联互通节点建设工作的通知》, 正式启动国家算力互联互通节点体系建设, “1+M+N” 体系架构 (如图表 4 所示), 已然成为全国一体化算力网的核心支撑。

图表 4: 全国一体化算力网 “1+M+N” 三级调度架构



来源：工业和信息化部、国金证券研究所

❖ “1”指国家算力互联网服务平台，负责全国算力资源的统一标识分配与运行监测；“M”指区域节点（如京津冀、长三角、粤港澳、成渝等算力需求旺盛地区），面向所在区域提供标识注册、资源汇聚等综合性支撑服务；“N”指行业节点（如金融、工业、通信、科研等行业算力平台），聚焦重点行业提供市场化服务并接入区域节点。截至 2026 年一季度，顶层平台直接接入 163EFLOPS 算力资源；全国各级区域节点、行业节点累计汇集 316EFLOPS 智算资源，已接入 31 个省（区、市）的 155 家企业（含区域、行业节点接入企业）的 578 个资源池、72 万张 GPU 加速卡，部署完成 1032 万条算力标识（含地方节点全量标识，与国家平台 662 万条口径不同），开通数据专线 52 条，月均调度近 300 次（注：统计口径为跨区域大型算力调度任务，小型本地调度不

计入统计）。

- ❖ 该体系采取“统一标识、统一标准、统一规则”的“三统一”运行机制：各节点通过国家节点获取唯一标识编码，为算力资源赋予“数字身份证”，实现全国算力标识一码通行；依托算力资源汇聚系统、算力选择系统等六大核心系统，实现算力资源接口的标准化适配；明确算力交易、互联调度、安全管控等全流程规则，各类算力资源按统一规则参与市场交易。
- ❖ 国家级监测调度平台已成功接入 137 万 P 的智算资源，以 72% 的接入率将分散资源统一拉齐。这一调度体系的核心价值在于将分散各地的“算力孤岛”通过标准化接口和统一调度平台连接起来，实现算力资源的跨区域、跨架构、跨主体高效流转。中国工程院院士高文将其比作电力网的升级版：“当需要用算力时，只需插上插头，不用关心算力来自哪里，系统会自动匹配性价比最高的算力中心。”
- ❖ 在市场化机制方面，多地出台算力采购补贴、算力券激励政策，各地补贴比例、年度额度存在差异。三大运营商推出的词元资费套餐开创了算力商品化新模式：上海电信推出词元算力套餐，1 元对应 25 万额度点；上海移动联合头部互联网平台企业推出 AI 原生工作台，1 元可购 40 万 Tokens（注：额度点与 Tokens 计量口径不同，不直接可比），实现“一个额度、一个价格、任选模型”。

1.4 传统 AI 算力行业核心痛点：四大结构性困境

- 全国一体化算力网建设持续提速，但 AI 算力运营领域仍存在供需错配、定价碎片化、异构生态割裂、盈利高度分化四大结构性难题。空间与层级供需失衡

造成部分算力资产利用率偏低；传统硬件计价缺乏统一标准，词元新模式尚未完善；多芯片软硬件互不打通、CUDA 生态构筑高迁移门槛；产业进入 GaaS（算力即服务）服务新阶段后，调度、枢纽、长单三重壁垒拉开企业差距，中小服务商差异化生存赛道收缩。这四大痛点是顶层政策优化、行业商业模式迭代需要核心解决的关键问题。

❖ **痛点一：算力供需多重资源错配，低效闲置问题突出。**国内算力产业存在三层结构性资源错配矛盾。一是区域供需空间割裂：东部 AI 产业集聚、算力需求旺盛，但能耗收紧，机房扩容受限；西部绿电、能耗资源充裕，但本地算力需求薄弱，跨区域数据传输成本与时延约束难以承接东部海量业务，区域资源供需错配。二是算力层级结构性失衡：适配超大模型训练的万卡级液冷高端智算持续供给紧张，而仅能承载基础业务的传统低端通用 IDC 产能过剩，出现“高端算力一卡难求、低端算力资源未能充分释放”分化格局。三是基建规划与真实需求脱节，IDC 行业监测数据显示部分新建智算中心 GPU 综合利用率低于合理运营区间，重资产投入持续空转。盲目扩产造成资源浪费，“以用定建、按需规划”已成为全行业共识。

❖ **痛点二：行业定价体系碎片化，词元标准化模式尚待成熟。**传统算力租赁行业以机柜、单 GPU 卡、服务器时长作为核心计费单位，市场价格体系高度混乱：不同服务商、不同芯片架构、不同地域机房之间报价差距悬殊，行业缺少统一的算力效能折算基准，客户难以横向对比真实算力性价比，低价恶性竞争常态化。词元按量计费体系为行业标准化提供解决方案，将计价标的从硬件资源转向 AI 实际产出，实现成本与业务需求匹配。但当前 Token 商业化仍处于早期探索阶段，行业统一分词计量规则、分级定价机制、算力-模型收益

分成模式均未落地，计量口径不统一、调价波动大、多场景适配不足等问题显著，完整标准化定价体系仍需长期完善。

❖ **痛点三：算力异构孤岛问题突出，软件生态形成核心壁垒。**当前算力市场多架构芯片并行共存，英伟达 GPU、主流国产 NPU 及多款国产 AI 芯片等硬件底层架构、专属编程栈相互独立，各厂商算子库、互联协议、运维工具互不打通，形成相互隔离的算力孤岛。多芯片混合集群无法实现统一资源纳管、跨架构智能调度，算力闲置率居高不下，资源利用效率持续走低。同时跨芯片业务迁移成本极高：CUDA 生态历经近 20 年持续迭代，沉淀海量专用优化算子、全球数百万开发者，主流大模型、AI 工具原生适配 CUDA 体系；国产 AI 芯片硬件纸面性能持续追赶，但软件生态积累严重不足，模型移植需大规模改写代码、长期调参优化，实际运行吞吐量存在明显折损，生态壁垒成为国产算力规模化落地、混合算力统一运营的突出短板。

❖ **痛点四：行业盈利高度分化，头部集聚效应逐步显现。**当前算力运营行业呈现明显分化态势，头部厂商依托大规模智算集群、核心长期客户资源、低成本融资能力持续扩产，算力上架率、毛利率、现金流持续优化；中小算力服务商则持续承受客户流失、行业价格内卷、重资产刚性折旧三重挤压，盈利持续收窄甚至亏损。行业正告别单纯囤卡转租的硬件租赁粗放阶段，加速迈入 GaaS 一体化算力服务时代。新阶段竞争核心不再是 GPU 库存规模，而是三大不可复制壁垒：自研全局算力调度系统、国家级大型智算枢纽节点、与头部 AI/政企客户的长期稳定合作。缺少调度能力无法优化资源利用率、无枢纽资产缺乏成本优势、无长期订单抵御周期波动的中小厂商，竞争压力加大，部分主体转向细分赛道。行业资源禀赋分化效应逐步显现，各类算力运营主体资源禀

赋分化，集中格局的最终形态仍有待观察。

- **中美算力产业发展路线对比：**海外算力产业以单集群极致硬件优化为主，依托成熟芯片软件生态，通过超大规模集中式智算集群追求单点算力密度最大化，如海外某头部厂商推出的整机柜算力架构，通过高密度液冷集成和专用互联协议实现单柜功率的大幅提升，该路线的优势在于硬件效率极高，但对芯片供给和先进制程高度依赖。我国受供应链外部约束，走出了一条**跨区域异构调度、算电协同的分布式一体化路径：**通过全国一体化算力网将分散在各地的异构算力资源统一纳管、按需调度，以系统层面的协同优化弥补单点硬件性能差距，同时依托西部绿电优势降低算力综合成本。两类路线由各自产业基础、外部约束条件决定，不存在绝对优劣，反映了不同产业基础下的差异化技术选择。

第二章 模式之争：四种运营模式的盈利逻辑与竞争格局

2.1 基础算力资源租赁：硬件计费模式的增长天花板

- 基础算力资源租赁是当前较为成熟、占比较高的算力运营模式，以机柜、GPU卡或服务器为计价单位，按时间（小时/月/年）向客户收取租金。据中国信通院数据，2026 年一季度国内算力租赁市场规模已达 680 亿元，同比增长 62%，预计全年将突破 2600 亿元。
- ❖ **基础租赁模式具备现实业务优势：**对需求方而言，无需承担算力集群巨额硬件采购、机房建设、运维团队搭建的一次性资本开支，将重资产投入转化为分期运营成本，算力获取门槛显著降低，同时支持算力资源弹性伸缩，能够快速响应短期爆发式训练、推理业务需求；对供给方而言，可以通过多客户错峰复

用硬件资源，提升机房与 GPU 资产综合利用率，依托长周期租赁合约获得相对稳定的现金流回报。

- ❖ **基础租赁模式的核心局限在于三重天花板：**收入天花板受限于物理机柜规模和芯片数量，难以实现指数级增长；客户粘性弱，价格战频发；无法捕获 AI 应用层的增值收益。随着算力网推进和词元计价体系成熟，纯硬件租赁模式的利润空间将被持续压缩，行业正在向更高附加值的模式升级。

2.2 分层算力订阅：训练/推理/异构混合算力包的规模效应

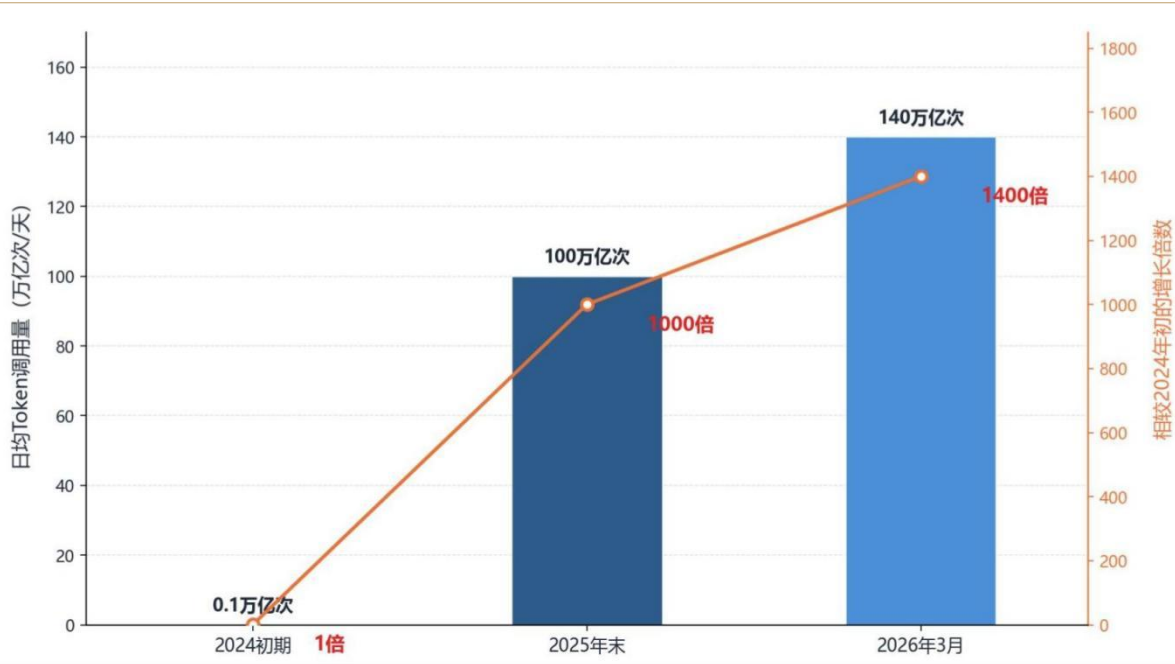
- 分层算力订阅模式是在基础租赁之上的升级，核心是算力资源按场景（训练/推理/异构混合）打包为标准化“算力包”，客户按订阅周期付费，类似 SaaS 的商业模式。其关键优势在于三重效应：通过长期合约锁定客户，提升 ARPU 值和收入可预测性；通过资源池化调度提升算力利用率；通过分层定价满足不同客户群体需求。
- ❖ **典型的分层结构：**训练算力包（高功率液冷集群，面向大模型训练，单价最高）、推理算力包（中低功率 GPU 集群，面向 API 调用和模型推理，需求量较大）、异构混合算力包（国产+海外芯片混合调度，面向成本敏感型客户）。运营商和云厂商在这一模式上具有天然优势：三大电信运营商已启动算力互联网试验网建设，将算力包嵌入其政企客户服务体系。
- ❖ **该模式的增长驱动力来自 AI 推理需求的快速释放。**随着大模型从“训练阶段”进入“大规模应用阶段”，推理算力需求增速远超训练算力。由国际数据公司（IDC）和浪潮信息联合发布的《2025 年中国人工智能算力发展评估报告》内容提到：2025 年中国智能算力规模达 1037.3EFLOPS，其中推理算力占比快

速攀升。分层订阅模式能更精准匹配推理场景的弹性需求特征，是未来 3-5 年增长确定性较高的商业模式之一。

2.3 GaaS（算力即服务）：词元计价下的全链路解决方案

- GaaS（算力即服务，GPU-as-a-Service）是算力运营的高阶形态，核心特征是“任务级调度、全链路解决方案、词元计价”。在这一模式下，客户不再关心底层硬件型号和数量，而是直接按词元消耗量或计算任务结果付费，算力运营商承担从资源调度、任务拆分、异构适配到结果交付的全流程服务。

图表 5：日均词元调用量两年增长超千倍



来源：国家数据局，国金证券研究所

- ❖ 词元经济的快速增长是 GaaS 模式落地的核心催化剂。国家数据局数据显示（如图表 5 所示），中国日均词元调用量从 2024 年初的 1000 亿跃升至 2025 年底的 100 万亿，到 2026 年 3 月已突破 140 万亿，两年增长超 1000 倍。仅 2025 年底至 2026 年 3 月的三个月内就增长 40%。2026 年 5 月 22 日，国家数据局局长刘烈宏主持召开词元经济座谈会，明确将推动词元经济发展纳入工

作体系。

- ❖ 关于词元经济、词元计价模式对 AI 产业链的影响深度分析请参见国金证券产业研究中心近期已发布的产业深度报告《Token 计费重塑 AI 产业链：你的企业定价权还在吗？》
- ❖ **GaaS 模式核心优势：**对需求侧客户，实现算力消费“按效付费”，屏蔽底层硬件差异，无需关注 GPU 型号、机房位置、集群运维，降低大模型应用落地的技术门槛；可根据业务波峰波谷弹性消耗词元资源，避免闲置硬件资源带来的沉没成本，尤其适配推理业务高并发、瞬时爆发的业务特征。对供给侧运营商，打破机柜物理数量对收入的硬性约束，依托跨域异构调度能力盘活分散、异构算力资源；可以混合调度国产 NPU 与海外 GPU，实现成本与性能的动态配比；业务收入从硬件租金升级为任务服务收入，能够获取调度、适配、运维带来的服务增值收益。
- ❖ **GaaS 模式现存不足与现实约束：**第一，行业缺乏全国统一的词元计量、折算、结算标准，不同厂商“额度点、Token”统计口径不一致，跨平台比价困难；第二，异构算力适配改造成本较高，要实现多芯片架构统一纳管、任务拆分调度，需要投入大量调度软件与算子适配开发；第三，计费、计量、审计体系尚处于试点阶段，跨主体、跨省份算力调用的结算、对账机制不健全；第四，客户使用习惯仍在培育，部分政企客户传统预算体系仍适配机柜、服务器硬件采购模式，对按词元按量付费的采购模式接受度有限；第五，任务调度带来大量跨域数据传输，会产生额外网络时延与带宽成本，极端场景下会抵消调度带来的成本节约。

- ❖ **GaaS 模式的盈利结构与传统租赁有本质区别：**收入与词元消耗量直接挂钩，不受物理机柜数量限制；通过智能调度实现异构算力成本优化（将低成本国产芯片用于推理、高端芯片用于训练）；通过全链路服务捕获应用层增值收益。长期来看，GaaS 计费模式渗透率有望持续提升，推动算力运营商业形态迭代，行业经营重心将由单纯硬件机柜、GPU 裸卡租赁，转向面向客户提供一体化算力服务。

2.4 行业专属算力定制：高合规算力集群的盈利模型

- **行业专属算力定制模式面向金融、自动驾驶、工业、医疗等高合规要求行业，**提供从物理隔离、数据安全、行业适配到运维保障的全栈定制化算力服务。其核心壁垒在于行业资质（如金融行业等保三级）、数据安全合规能力、以及行业模型适配经验的三重门槛。行业专属算力定制商业模式核心特征体现为高溢价、高粘性、高壁垒。客户更换算力服务商涉及海量数据迁移、模型重新训练、合规体系再次认证，整体切换成本较高。服务商与客户建立深度合作后，能够持续挖掘长期客户价值。与此同时，该模式对运营主体的垂直行业认知、定制化项目交付能力门槛较高，更适合具备垂直行业积累的专业算力运营商。
- ❖ **行业专属算力定制模式的核心优势：**对客户侧，可实现硬件、网络、存储物理隔离，满足数据不出域、操作全程可审计等高等级监管合规硬性要求；可针对行业业务特征做硬件选型、网络时延、安全策略深度定制，适配金融交易、自动驾驶仿真、医疗影像等特殊业务指标；集群资源专属独占，不受多租户共享环境的资源抢占、性能波动影响，业务稳定性、SLA 服务等级更容易保障。对服务商侧，项目客户粘性极强，替换迁移成本高，容易形成多年长周期框架订

单；相比通用算力租赁，可以获取合规服务、行业适配带来的服务溢价；通过深度绑定行业客户，沉淀垂直领域模型适配、安全运维能力，构筑竞争对手难以复制的行业壁垒。

- ❖ **金融行业是行业专属算力的典型场景：**银行、证券、保险等金融机构在 AI 大模型应用中，面临严格的监管要求（数据不出域、模型可解释、交易可追溯），需要建设物理隔离的专属算力集群。
- ❖ **自动驾驶行业对算力的需求特征是“高吞吐、低时延、海量数据”。**自动驾驶研发属于数据密集型业务，对低时延分布式算力调度存在强需求。依托一体化算力网络，企业可以就近调度区域内多元算力资源，加速自动驾驶模型迭代。相比通用算力需求，自动驾驶研发客户项目周期更长，付费能力更强。
- ❖ **行业专属算力模式现存短板与现实约束：**第一，前期建设与定制改造成本高，硬件、安全设备、合规体系、行业适配开发叠加，项目初始资本投入显著高于通用算力租赁；第二，项目交付周期长，从方案设计、安全测评、等保测评到上线投产往往需要数周乃至数月，无法做到开箱即用；第三，客户行业属性强，项目高度非标化，不同行业合规条款、业务指标差异巨大，难以标准化批量复制，规模化扩张难度大；第四，资源利用率存在不确定性，专属集群完全服务单一客户，如果客户业务量波动，容易出现算力闲置；第五，行业合规要求迭代频繁，需要持续投入人力、资金更新安全策略与审计体系，持续运维成本较高。上海《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，到 2025 年本市新建智算中心国产算力芯片使用占比超过 50%，该约束仅适用于上海本地新建商用智算项目，不属于全国层面统一强制标准；北京市《北京市算力基础设施建设实施方案（2024-2027

年）》明确，到 2027 年具备 100%自主可控智算中心建设能力，该目标指向本土智算全流程工程技术储备攻关能力。随着多地算力基础设施国产化导向政策落地，有望持续拉动自主架构专属算力的市场需求。

2.5 四大运营阵营商业模式对比与竞争格局

- **当前中国 AI 算力运营市场形成四大阵营竞争格局：**公有云厂商、电信运营商、第三方 IDC 服务商、地方国资算力平台。各阵营资源禀赋、核心壁垒和发展路径差异显著。
- ❖ **公有云厂商（公有云服务商）：**技术能力突出，在跨域调度、异构算力融合、GaaS 服务方面走在行业前列。核心优势是庞大的开发者生态和 PaaS/SaaS 全栈能力，能够提供从算力到模型到应用的一体化服务。但公有云的“多租户”模式在金融、政务等高合规场景面临信任挑战。
- ❖ **电信运营商：**资源禀赋突出，拥有完善的基础网络、机房资源和政企客户关系。三大运营商已启动算力互联网试验网建设，并率先推出词元资费套餐。运营商的核心战略是将算力嵌入其“云网融合”体系，打造“网络+算力+AI”的一体化服务。其优势在于政企客户渠道和政策资源，劣势在于技术创新效率和生态开放度。
- ❖ **第三方 IDC 服务商（第三方 IDC 运营企业）：**市场化和运营效率突出，但受限于重资产模式和规模瓶颈。在 AI 算力转型中，第三方 IDC 面临两难处境：液冷改造成本高昂，但不上液冷则无法承接 AI 算力需求。行业正在经历深度洗牌，头部企业通过并购整合扩大规模，中小服务商差异化生存赛道收缩，部分主体探索细分垂直赛道。

- ❖ **地方国资算力平台：政策优势突出，在土地、电力、能耗指标等方面享受地方政府的全方位支持。**多地地方政府设立算力专项扶持基金，建设地方智算中心。但国资平台在技术能力、运营效率和市场竞争力方面存在短板，未来更可能作为“算力资源提供方”参与生态，而非独立运营。
- 从中长期竞争格局演变观察，行业将形成“运营商守网络和政企渠道、云厂商做技术和生态、第三方 IDC 做精细化运营、国资平台做资源供给”的差异化分工格局（如图表 6 所示）。行业资源禀赋分化逐步显现，头部运营主体依托枢纽、调度、长单资源具备综合优势，中小服务商可深耕垂直细分赛道形成差异化生存空间。

图表 6：国内四大算力运营阵营核心能力对比

维度	公有云厂商	电信运营商	第三方 IDC 服务商	地方国资算力平台
资源禀赋	开发者生态与 PaaS/SaaS 全栈能力	基础网络、机房资源与政企客户关系	市场化运营经验与存量客户资源	土地、电力、能耗指标等政策支持
技术能力	跨域调度、异构融合、GaaS 服务领先	云网融合能力强，词元套餐先行	液冷改造与 AIDC 升级能力	技术能力相对薄弱
客户渠道	互联网企业与中小企业	政企与行业大客户	互联网、金融等行业客户	政府与国有企业
合规优势	等保与安全认证体系完善	等保三级资质与国资背景	等保三级与行业认证	数据本地化与政务合规优势
核心短板	重资产扩张受能耗指标约束	市场化运营效率待提升	重资产模式与规模瓶颈	运营效率与市场竞争力不足
发展方向	技术与生态引领者	网络与渠道守成者	精细化运营转型	资源供给与政策引导

来源：国金证券研究所

第三章 技术升级：调度、液冷、光通信、芯片与算电协同的底层冲击

3.1 算力调度技术：异构统一纳管与跨域低时延调度需求迫切

- 算力调度技术是全国一体化算力网的技术核心，解决的是“如何将分散在各地的异构算力资源统一管理、按需调度、高效利用”这一根本问题。当前调度技术面临三大挑战：异构算力统一纳管（英伟达 GPU、主流国产 NPU 及多款国产 AI 芯片等架构各异的芯片需要统一接口）、跨区域低时延调度（训练任务跨域拆分与协同）、调度平台效能优化（资源利用率改善空间显著）。

图表 7：异构算力统一纳管与跨域调度技术架构



来源：国金证券研究所整理

- ❖ 在异构算力统一纳管方面，国家超算互联网平台已做出成功实践。该平台接入 14 个地区、超过 30 家国家级超算中心与智算中心，整合形成总算力规模超过 15 万加速卡与 200 万核心的大型异构算力资源池，为用户提供近 70 种

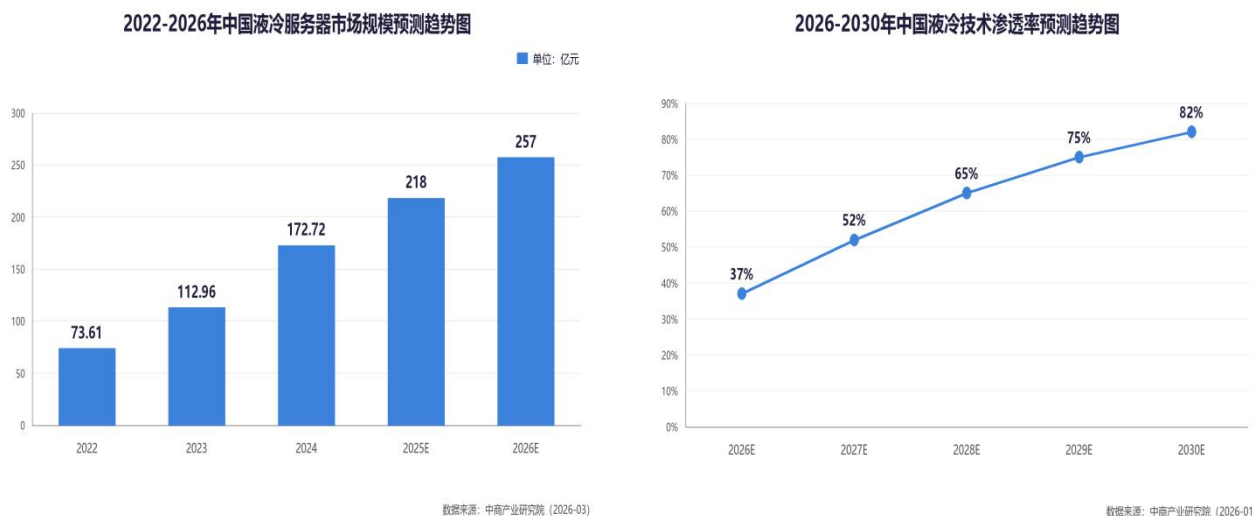
规格的算力。2026 年 2 月上线的郑州国家级超算节点搭载大规模国产算力集群，可对外提供大容量异构算力供给，能够将不同地区、不同类型的异构算力资源统一接入、统一调度。

- ❖ 在跨域调度方面，2025 年以来，江苏省未来网络集团的全国算网协同调度中心已完成跨省市远距离大模型调度工程验证，跨域协同可缩短模型迭代周期，为全国一体化算力网落地提供了关键技术验证。
- ❖ 调度平台效能优化方面，通过跨中心优化调度、削峰填谷和作业排队机制，全网算力资源平均利用率较此前有所改善，但不同集群、不同业务场景的优化效果各不相同。北京算力互联互通和运行服务平台已接入大量算力服务商，覆盖京津冀蒙晋等地区；长三角（上海）算力互联互通平台同样接入大量服务商主体。
- ❖ 总体而言，跨域调度已验证可行，但异构统一纳管与调度效能优化仍处于持续迭代阶段。

3.2 高密度机房配套技术：液冷散热与高速光通信

- AI 算力规模快速扩张对数据中心物理基础设施提出了更高要求。传统风冷散热已逼近物理极限，单机柜功率从传统的 5-8kW 飙升至 30-100kW 甚至更高，液冷是当前高密智算主流散热方案，风冷、液冷混合架构长期并行落地。

图表 8：液冷服务器渗透率正经历从“政策驱动”到“技术刚需”的加速跃迁



来源：中商产业研究院，国金证券研究所

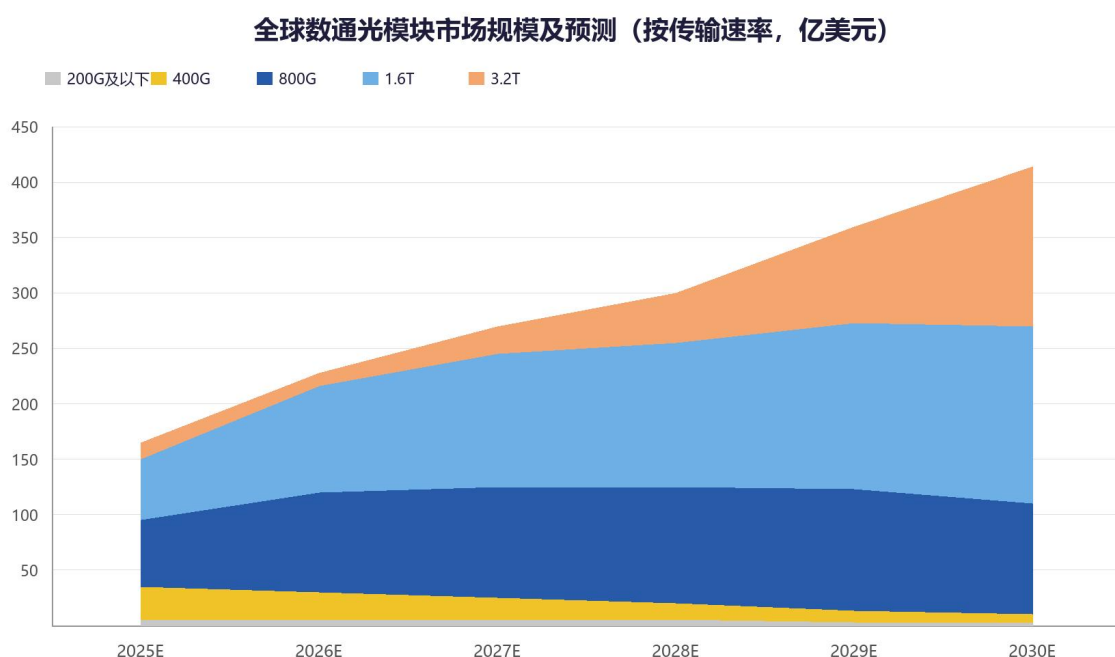
注：数据口径为中商产业研究院 2026 年 3 月预测液冷服务器整机市场，不含单独液冷散热设备配套市场。2027-2030 为乐观情景预测，受成本、政策、产能会存在较大偏差风险。

❖ **在液冷服务器市场方面**，根据中商产业研究院发布的《2024-2030 年中国液冷服务器行业市场发展现状及潜力分析研究报告》显示，中国液冷服务器市场在 2024 年继续保持快速增长，市场规模约达 172.72 亿元，同比增长 67.0%，2025 年市场规模约达 218 亿元。中商产业研究院预测，2026 年中国液冷服务器市场规模有望达 257 亿元。中国液冷服务器渗透率正经历从政策驱动到技术刚需的加速跃迁，2021-2025 年从不足 3%快速突破至 20%，2026 年随着 AI 算力爆发和 GB200 等强制液冷方案落地将跃升至 37%，此后进入高速增长期，预计 2027 年突破 50%临界点、2028 年达 65%、2030 年攀升至 82%接近饱和，其中冷板式液冷占据主导，浸没式在超高密度场景持续渗透，前快后稳，最终形成新建数据中心全面液冷、存量场景风冷补充的稳态格局。

❖ **在高速光通信市场方面**，围绕算力枢纽已建成干线光缆 145 条，基础电信运营商持续推进 400G/800G 高速光传输网络部署。根据行业研究机构

LightCounting 预测（国内光模块企业 2025 年年报转述），2026 年全球光模块总市场规模有望达到 228 亿美元，其中 800G 和 1.6T 光模块合计市场规模约 146 亿美元占整体市场约 64%。

图表 9：2026 年全球光模块总市场规模有望达到 228 亿美元（LightCounting 口径）



来源：国内光模块企业 2025 年年报中转述行业研究机构 LightCounting 的预测图，国金证券研究所

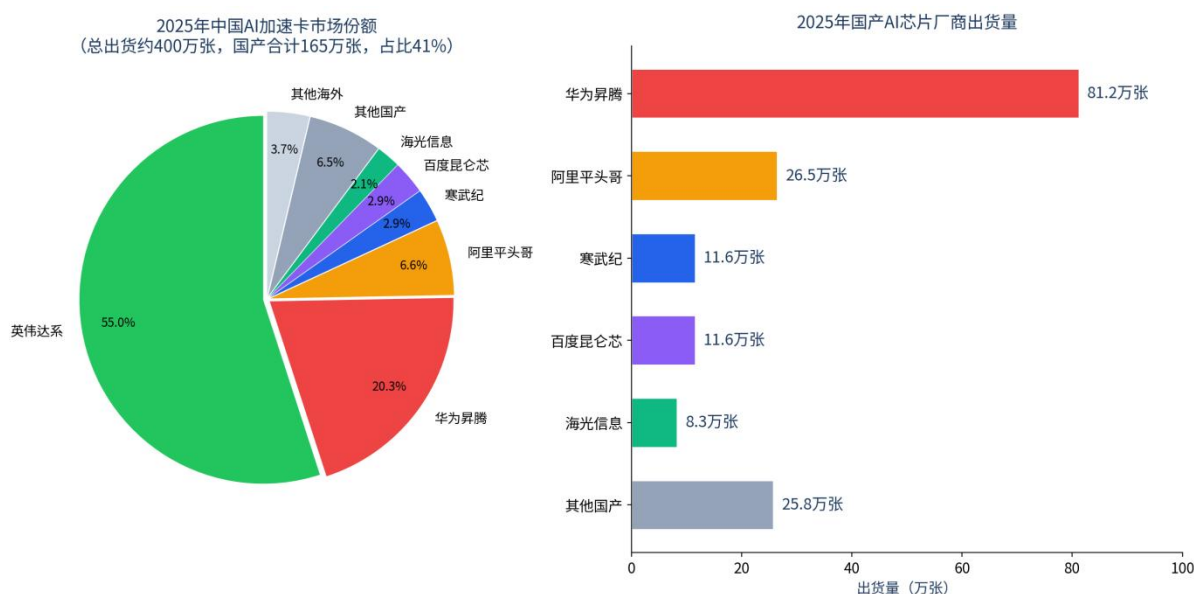
❖ 从产业链来看，中国光模块企业已在全球市场位居行业前列。国内光模块企业在 800G、1.6T 等高端产品上已具备规模化交付能力，是全球 AI 算力基础设施建设的核心参与者。总体而言，液冷与高速光通信技术迭代节奏快，但存量改造成本与技术换代周期是规模化落地的核心约束。

3.3 算力芯片生态：国产替代加速与软件生态短板

国产 AI 芯片正在经历从“样品”到“商品”、从“补充”到“主流”的历史性跨

越。2025 年中国市场 AI 加速卡总出货量达 400 万张，其中国产半导体厂商出货 165 万张，拿下 41% 的市场份额，英伟达市场份额占到 55%。

图表 10：国产 AI 芯片份额首超四成，龙头厂商领跑



来源：IDC 商业报告，国金证券研究所

注：统计口径为国产 AI 芯片及中国大陆市场 AI 加速卡出货，不含出口；数据为新闻媒体引自 IDC 付费商业报告后对外转述，不属于政府官方统计。华为昇腾 81.2 万张、平头哥 26.5 万张、昆仑芯 11.6 万张、寒武纪 11.6 万张等细分出货数据均来自该新闻转述。

❖ **国产阵营已形成清晰的梯队格局：**头部国产芯片厂商出货规模位居国产阵营前列；多家国产芯片厂商紧随其后，出货量存在显著梯度差异；其余厂商构成差异化补充。参考 IDC 商业报告结合产能与地方国产化政策给出多情景推演：乐观情景下 2026 全年国产 AI 加速卡出货占比有望进一步提升，中性情景下国产份额维持上行态势。考虑到份额测算受芯片管制、产能释放节奏等多重外部变量扰动，我们认为第三方的统计及预测数值仅作参考。

❖ **国产 AI 芯片头部阵营已取得关键节点进展。**其一，华为于 2026 中国合作伙伴大会（2026 年 3 月）发布 Atlas350 加速卡，该卡搭载昇腾 950PR 处理器，芯片端首次配置华为自研 HBM 高带宽内存，mxFP4 精度算力达 1.56 PFLOPS；

在官方公布的 FP4 推理测试场景下，加速卡性能为英伟达对华特供版 H20 的 2.87 倍（注：H20 为面向中国市场特供版本，与海外旗舰产品定位不同；该对比为厂商公布测试场景结果，非全场景通用算力对比）。其二，2026 年以深度求索（DeepSeek）为代表的头部大模型企业完成国产芯片架构推理侧的全栈深度适配，实现 DeepSeek-V4 系列模型在华为昇腾 CANN 软件体系下推理业务的完整落地，成为国内国产生态规模化落地的标杆案例。该类适配改造目前仍要承担较高的算子重写与工程调优成本，当前成熟成果集中于推理环节，训练环节仍存在较多适配难点；现阶段该项目属于定制化联合攻坚案例，短期暂不确认具备全行业复制推广的条件。

- ❖ 2026 年上半年部分头部厂商首次实现单季度经营现金流转正，商业化造血能力得到验证。国产 GPU 厂商持续推进与主流开源大模型的深度适配，软硬件协同能力持续提升。我们看到现实存在的挑战依然严峻，虽然国产 AI 芯片单卡峰值算力已接近国际主流水平，但在软件生态成熟度、先进制程工艺、系统级优化能力方面仍有差距。我们认为当下存在三大核心考验：第一是海外生态壁垒尚未根本破解，尽管兼容 CUDA，但复杂模型适配和开发者生态完善度仍有差距；第二是国内 7nm 及以下先进制程产能供给存在阶段性约束，HBM 高带宽存储等关键配套介质供给亦集中于海外少数原厂并面临出口管制限制，制约性能天花板提升；第三是如果美国持续加码芯片出口管制，海外高端 GPU 供给路径将 ~~会持续~~ 受阻。
- ❖ 目前已经形成业界共识：国产算力是保障国内供应链自主可控的核心路线，与海外算力产品形成场景差异化互补体系。尽管国产 AI 芯片硬件指标持续追赶，但软件生态割裂仍是长期落地核心瓶颈。我们认为国产化份额提升依赖

国内开源生态共建的效果，既需要产业链各厂商不断努力、更需要举全国之力系统性攻关。

3.4 算电协同技术：绿电配套与算力柔性负荷调节

- “算电协同” 2026 年首次写入政府工作报告，并纳入“十五五”规划纲要，标志着算力与电力的深度融合从概念走向落地。算力基础设施属于高载能领域，万卡级大型智算集群电力负荷规模巨大，硬件配置不同集群功耗差异显著，绿色电力供给能力直接影响算力中心的运营成本与长期可持续运行能力。
- ❖ 国家政策引导算力枢纽节点提高新建数据中心绿电配套比例，80%为国家算力枢纽新建数据中心的中长期引导目标，并非面向全国所有项目的统一硬性考核指标，各地落地执行标准存在差异。在实践层面，行业已探索形成“源网荷储”一体化、“光伏直连+风电长协”等绿电多元利用模式。2026 年，广东阳江 GW 级海底算力中心正式启动，项目依托当地千万千瓦级海上风电基地规划开展绿电直供试点；根据项目公开方案测算，结合海水自然冷却与全链路液冷技术，相较传统数据中心可实现能效提升 30%以上。内蒙古和林格尔新区持续深化算电协同实践，布局多座规模化智算集群项目，配套建设“风光储”一体化绿电供给示范项目，打造全国算电协同标杆集群。
- ❖ 根据国家发展改革委、国家能源局、工业和信息化部、国家数据局联合印发的《关于促进人工智能与能源双向赋能的行动方案》（国能发科技〔2026〕34 号），到 2027 年，支撑人工智能创新发展的安全、绿色、经济的能源保障体系初步构建，清洁能源与算力设施互动能力显著提升（注：该政策目标优先在算力枢纽节点开展试点先行，向全国落地推广存在区域时间差）。

- ❖ **算力柔性负荷调节是算电协同的高级形态**，即将算力中心作为电力系统的可调节负荷，在用电高峰时适度压降算力负载、用电低谷时满载运行，实现“算随电走、削峰填谷”。该机制一方面可以降低算力中心用电成本，另一方面可为电网提供灵活调节资源，是算电协同的核心发展方向。
- ❖ **需辩证看待算电协同的落地挑战**：（1）风光等可再生能源天然存在发电波动性强、可调度性弱的特征，大规模绿电直供仍需配套储能设施，储能投入会显著抬升算力中心综合建设成本；（2）同时算力-电力跨系统协同调度，对电网调度系统、算力调度系统实时联动能力要求极高，当前相关技术模式尚处在试点验证阶段，调节响应时延、控制精度与运行可靠性仍有待大规模工程检验。算电协同规模化推广，需要在绿电供给稳定性、储能项目经济性、跨系统协同运行效率三者之间取得综合平衡。

第四章 利润去哪儿了？价值重分配、交易市场、竞争格局与估值重构

4.1 全产业链价值重分配：从“卖硬件”到“卖服务”的利润转移

- ❖ 全国一体化算力网建设正在引发算力产业链全链条的价值重分配。传统模式下，利润高度集中在上游硬件环节（芯片、服务器），中游算力运营商和下游应用方的利润空间有限。随着词元计价体系成熟和 GaaS 模式普及，价值正在从“硬件”向“服务”迁移，中游算力运营商和应用服务商的利润占比将显著提升。
- ❖ 从产业规模看，中国信通院测算，2024 年中国人工智能核心产业规模已突破 9000 亿元，2025 年预计突破 1.2 万亿元。国家数据发展研究院测算，2024 年

面向 AI 数据集相关的数据产业产值规模已超过 2 万亿元，预计 2025 年增长 15%，产值规模超过 2.3 万亿元。刘烈宏在中国发展高层论坛 2026 年年会上推演，到“十五五”末，我国人工智能相关产业规模有望达到 10 万亿元区间；产业链上下游都将迎来巨大增量空间，但价值分配格局将出现明显调整（注：10 万亿元为论坛公开推演结果，非国家规划约束指标，口径为 AI 相关全产业链，不等同于 AI 核心产业）。

- ❖ **利润迁移逻辑体现在三个维度：**一是从“芯片差价”向“调度溢价”转移，具备异构算力统一纳管和智能调度能力的算力运营主体，通过资源优化获取一定服务溢价，溢价水平随场景波动；二是从“机柜租金”向“词元分成”转移。GaaS 模式下，算力运营主体不仅收取算力使用费，还能从模型推理的词元消耗中获取分成；三是从“一次性建设”向“持续服务”转移，行业专属算力定制的长单模式带来持续稳定现金流，行业定制客户迁移成本高、长期合作稳定性更强。
- ❖ **这一价值迁移还将在供给约束背景下进一步强化：**HBM 等关键硬件的全球性紧缺推高了硬件采购与持有成本，重资产囤卡转租模式的盈利弹性受到压缩，而具备跨域调度、存量资源盘活能力的算力运营主体相对议价权反而提升，轻资产运营、服务创造价值的产业逻辑在硬件紧缺环境中更强的现实支撑。

4.2 行业生态协同：全国统一算力交易市场的形成路径

- ❖ **全国一体化算力网的终极目标是构建全国统一的算力交易市场，让算力像水电一样按需取用、按量计费。**这一目标的实现路径包括三个阶段：**第一阶段**

“资源入网”，通过“1+M+N”节点体系将分散算力资源标准化接入；第二阶段“智能调度”，通过统一标识和调度平台实现跨域资源匹配；第三阶段“市场交易”，形成算力的市场化定价和交易机制。

- ❖ 当前我国正处于第一阶段向第二阶段过渡的关键期。如本报告第一章 1.3 节所述，截至 2026 年一季度，国家算力互联网服务顶层平台直接接入 163EFLOPS 算力资源；全国各级区域节点、行业节点累计汇集 316EFLOPS 智算资源，覆盖 31 个省（区、市）。“三统一”运行机制的可行性与实用性，已在实际跨域调度任务中得到有效验证。
- ❖ 跨厂商算力互通是生态协同的另一关键维度。“模芯生态创新联盟”等产业组织正在推动从大模型开发阶段就去适配国产 AI 芯片，为国产芯片产业协同打开新思路。国产芯片厂商软件栈持续开源开放、多家国产 GPU 厂商软件平台实现与主流开源大模型的“Day0 适配”，都是生态协同的重要进展。
- ❖ 三大运营商联合中国信通院牵头建设行业算力服务平台，航空航天、金融科技、工业制造等领域的重点单位正结合场景需求推进算力服务平台建设。未来，全国将形成“核心节点统筹、区域节点中转、边缘节点覆盖”的三级算力网络。算力交易市场有望参照电力市场运行逻辑，逐步探索构建“中长期合约+现货交易+辅助服务”的多层次交易体系（该模式目前处于产业研究与试点阶段，尚未全国落地）。

4.3 市场竞争格局演变：头部整合、中小承压与国产替代

- ❖ 算力运营行业的竞争格局正在经历深刻重塑，三大趋势并行推进：头部整合、中小服务商承压、国产算力替代深化。

- ❖ **头部整合方面，具备枢纽资源、调度技术和长单壁垒的头部企业正在通过并购整合扩大规模优势。**根据国新办 2026 年 7 月 20 日工信部发布会公开数据，截至 2026 年 6 月底，国内已建成 52 个万卡以上智算设施。从产业实践来看，该类超大规模智算设施大多由头部市场主体运营。算力产业的高速增长吸引了大量资本进入，但最终能够穿越洗牌周期的，只有那些具备差异化竞争力的企业。
- ❖ **中小算力企业面临三重压力：价格战，头部企业凭借规模优势以更低价格提供算力服务；技术门槛，液冷改造、异构调度、词元计价等技术升级需要大量资本投入；客户流失，大客户更倾向于选择全栈能力的头部服务商。**不具备枢纽资源、调度能力和行业壁垒的中小算力企业经营压力持续加大。
- ❖ **国产算力替代是影响竞争格局的核心变量。**如报告第三章 3.3 节所述，2025 年国产 AI 芯片出货份额已超四成，2026 年有望进一步提升。各地国产化导向政策已陆续落地，政策口径存在明确区域差异。2026 年 5 月 26 日，中国信息安全测评中心完成首批国产 AI 训练推理芯片专项安全可靠测评工作，多家国产 AI 芯片厂商的 9 款产品获得“安全可靠 I 级”资质；获得认证的厂商在央企、金融、运营商领域的市场竞争力得到提升，相关领域订单预期将实现明显增长。

4.4 全新估值体系：从机柜规模到词元产出能力的范式转换

- ❖ **算力运营企业的估值框架正在经历体系性重构。**传统估值以机柜规模、服务器数量、电力容量等“物理资产”为核心指标，但这一体系已无法准确反映算力企业的真实价值。新的估值体系需要纳入四个新维度：算力利用率、词元产

出能力、调度技术水平、合规资质壁垒。

- ❖ **算力利用率是衡量运营效率的核心指标。**算力资源闲置会推高折旧、电力等刚性开销，在低利用率运行场景下，经产业测算，单位算力成本会有不同程度的抬升。而跨域统一调度可缓解算力资源闲置现象，不同机房、业务场景优化效果各不相同，同等物理资产产生更多收入和利润，是区分优质算力企业与低效算力企业的关键分水岭。
- ❖ **词元产出能力反映算力转化为商业价值的效率。**在 GaaS 模式下，日均词元调用量、词元单价、词元毛利率成为核心运营指标。同等资产规模下，词元产出效率更优的运营企业，其估值具备向上溢价空间，倍数应高于传统机柜租赁企业。
- ❖ **调度技术决定了算力运营商的“技术溢价”。**能够实现异构算力统一纳管、跨域低时延调度、智能任务拆分的企业，能够通过资源优化获取一定服务溢价，溢价水平随场景波动。这一技术壁垒难以在短期内复制，是算力运营企业的核心护城河。
- ❖ **合规资质壁垒在金融、政务等高价值场景中至关重要。**等保三级、安全可靠认证、数据安全合规等资质是进入高溢价市场的“入场券”。投资者在评估算力企业时，应将合规资质作为重要估值因子纳入考量。

图表 11：国内新旧算力运营估值体系对比

维度	旧估值体系	新估值体系
核心指标	机柜规模、服务器数量、电力容量	算力利用率、产出能力、调度技术水平
收入逻辑	物理资源出租，线性增长	词元消耗与服务溢价，非线性增长
壁垒来源	土地、电力、资金	调度技术、合规资质、客户粘性
估值锚点	重资产 PB/EV-EBITDA	效率指标与成长溢价
风险因子	上架率、电价、折旧	利用率波动、技术迭代、合规变化

来源：国金证券研究所

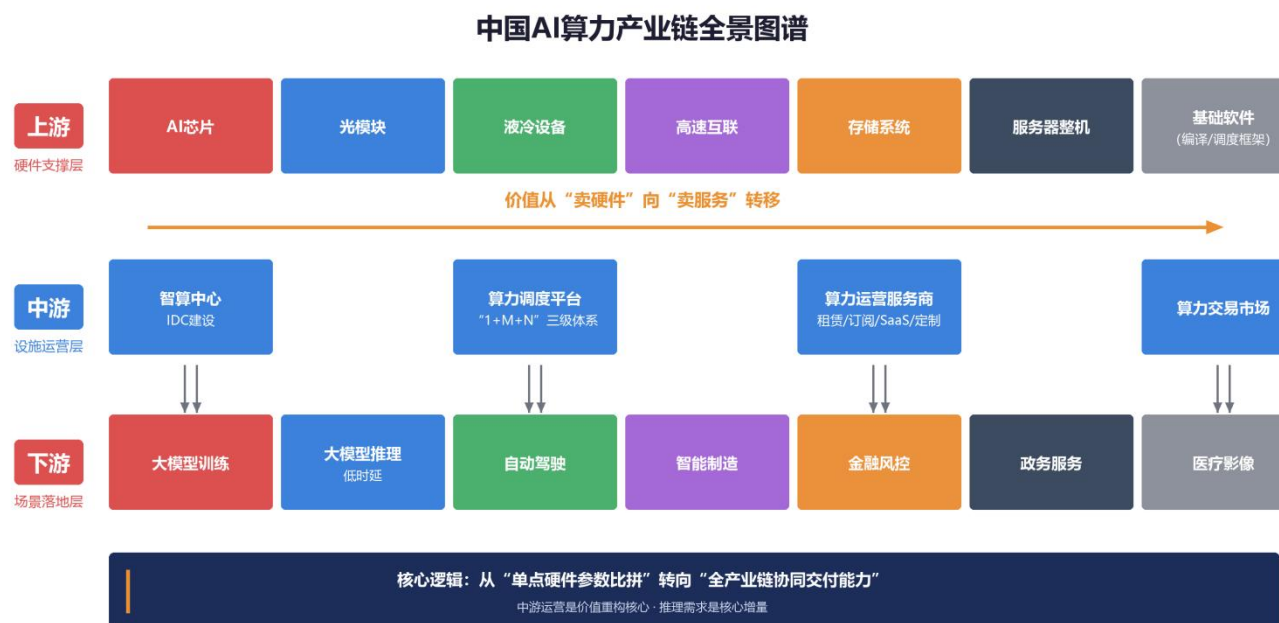
第五章 赛道掘金：产业链高景气环节与高价值落地场景

5.1 产业链全景图谱与中游核心节点研判

- ❖ 前文从顶层政策设计、商业模式演进、技术底层冲击、产业价值重构四个维度，系统推演了全国一体化算力网对产业价值分配与竞争格局的深层影响。本章将视角落地至算力全产业链实体层面，从中游智算中心与算力运营、上游五大硬件配套赛道、下游高价值行业需求三个层次，逐一拆解各赛道的增长逻辑、核心约束与结构性矛盾，为后续底层矛盾提炼与趋势研判提供产业事实基础。
- ❖ 算力产业链呈现清晰的塔状价值分布结构，各层级环环相扣、传导联动，任一环节的细微变化都会对其他环节产生蝴蝶效应，需要从系统层面协同解决。参考行业通用划分标准，产业链可分为上、中、下游三层，价值逻辑与竞争属性差异显著：

- ❖ **上游为算力硬件支撑层，是产业链价值壁垒最高的环节，** 占据整个产业链较高价值份额，核心供给包括 AI 芯片、光模块、服务器整机、存储系统、液冷机柜、基础软件等，为中游算力整合与下游应用提供底层物理支撑。
- ❖ **中游为基础设施与运营服务层，是连接上游硬件与下游应用的核心桥梁，** 承担算力资源整合、调度、交付的关键职能，涵盖智算中心 IDC 建设、一体化算力调度平台、算力租赁交易等业态，行业本质是 AI 时代的“算力电厂”，核心是将上游零散硬件转化为下游可直接使用的标准化算力服务。
- ❖ **下游为算力场景落地层，是产业价值的最终验证端，** 覆盖大模型训练、大模型推理、政务服务、金融风控、医疗影像、工业仿真等全场景算力需求，直接决定中上游算力资源的配置节奏与优先级，牵引整个产业的迭代方向。
- ❖ **当前行业的核心逻辑变化在于：产业竞争已从早期“单点硬件参数比拼”，** 转向“全产业链条协同交付能力”的系统竞争，单一环节的技术优势无法形成长期壁垒，必须实现上中下游多环节的高效联动。其中中游是当前产业模式重构的核心环节，行业经营属性正从“重资产建设”转向“重专业运营”。

图表 12：中国 AI 算力产业链全景图谱



来源：国金证券研究所整理

注：本图为产业逻辑示意图，仅展示产业链层级与业务逻辑；“1+M+N”为国家算力互联网平台官方节点体系。

5.1.1 智算中心建设：从“规模扩张”到“效率优先”

- ❖ 智算中心是算力资源的核心物理载体，行业存量规模与增量结构，直接反映产业建设重心的变化趋势。从公开数据看，据中商产业研究院测算，2025 年全国数据中心整体市场规模达 3180 亿元，2026 年预计达 3621 亿元。行业内部正在发生结构性变化，传统 IDC 托管业务占比持续下行，算力调度、智算服务等高附加值业务占比不断抬升，这一变化背后，是算力需求质变所牵引的行业逻辑重构。
- ❖ 作为 AI 算力的核心载体，智算中心的增长动能显著超过传统 IDC，算力建设重心持续向高密智算集群倾斜，国内智能算力规模呈现加速扩张态势。截至 2026 年 3 月底，据国家数据局公开披露，全国智能算力总规模已达 1882EFLOPS

(FP16)，其中八大国家算力枢纽节点占比超过 80%，智算中心已成为支撑我国 AI 产业发展的核心底盘。

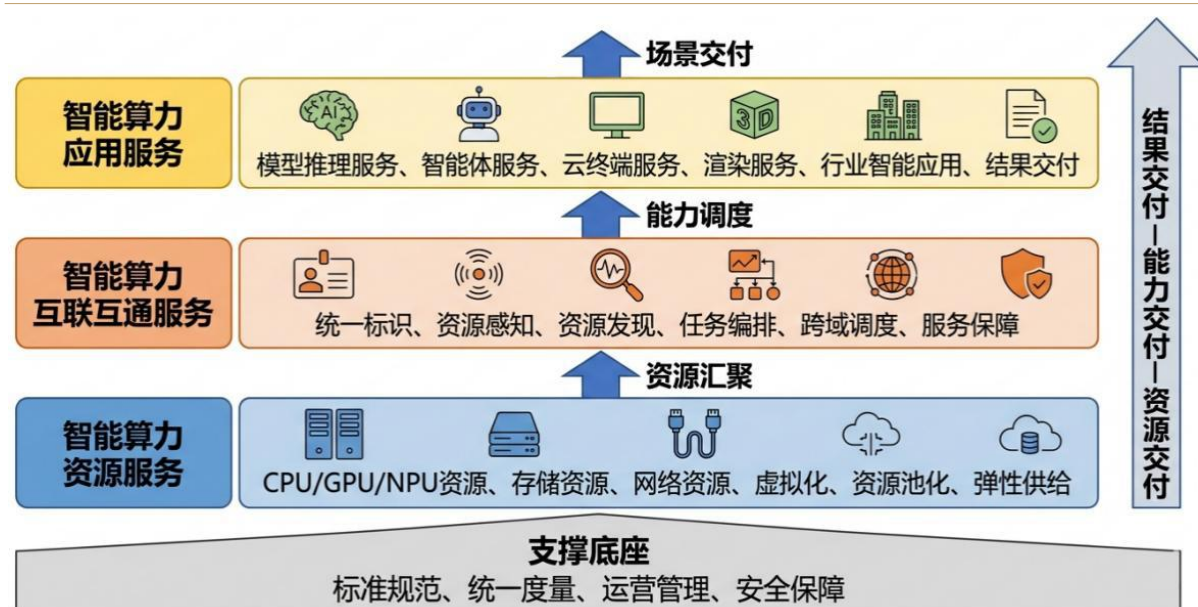
- ❖ 行业的突出矛盾，仍是“装机规模增长与有效算力供给不足”的结构性失衡。国内部分区域智算中心存在资源利用效率偏低的结构性现象，高端 GPU 实际负载偏低；同时东部产业集聚区域的高端算力供给紧张，总量增长的背后，是严重的区域供需错配，存量资源的价值无法充分释放。造成这一现象的根源，是行业长期存在的“重建设、轻运营”思维，此前多地布局智算中心时，以硬件装机规模为核心考核指标，忽视了算力资源的区域协同性与场景适配性，导致项目建成后缺乏持续的真实业务支撑，无法形成“建设、运营、收益、再投入”的正向商业循环。
- ❖ 随着行业逐步从硬件建设期转入效率优化期，智算中心的核心评价标准，正从“装机规模大小”转向“实际有效算力供给能力”，即集群能否稳定支撑下游场景的真实业务需求，而非单纯的硬件参数堆砌。行业布局逻辑也随之调整：不再单纯追求机架数量扩张，而是以产业需求为导向进行精准布局，八大算力枢纽间的功能分工边界正在清晰化。

5.1.2 算力运营与行业格局重构

- ✧ 算力运营是连接智算中心硬件资源与下游客户需求的关键环节。服务商的核心价值已不再是“出租机房空间”，而是依托专业化能力将零散硬件资源转化为下游可低成本、稳定、规模化使用的标准化算力服务。正如中国信通院对“智能算力服务”的定义，即基于互联网汇聚 GPU、NPU 等异构资源，通过统一服务接口按需提供可度量的计算、存储、网络服务。这一行业价值迁

移，本质上由下游算力需求的结构性变化驱动，适配 AI 业务场景的高功率机柜成为行业核心增量，彻底重塑了中游算力服务市场的竞争规则。

图表 13：中国智能算力服务体系示意图



来源：中国信息通信研究院、国金证券研究所

✧ 从行业格局看，中游市场已由“运营商主导”转向“云算力运营商+头部第三方”双轮驱动，竞争逻辑从规模扩张转向有效交付与运营效率。传统 IDC 时代，三大运营商凭借网络、带宽和机房资源占据优势；智算需求爆发后，市场主导权逐步向具备算力纳管、液冷高密交付和长周期客户运营能力的第三方与云厂商迁移。据信通院《中国第三方数据中心服务商分析报告（2023 年）》，2022 年第三方服务商市场份额约 51.68%，首次超过三大运营商合计的 48.32%（电信 19.27%、移动 14.70%、联通 14.35%），这一拐点确认了第三方阵营的主导地位。此后的集中化不再主要表现为“整体份额”的线性抬升，而是体现为头部能力分层与资源向领先者聚集：据信通院《算力中心服务商分析报告（2025 年）》，服务商评价已从规模导向转向“建设布局广度-技术

创新高度-市场经营能力”三维效能导向，并形成领航者（万国数据、润泽科技、世纪互联、中金数据、普洛斯）、拓局者（数据港、中联数据、有孚、浩云长盛、博浩数据）、稳进者（合盈数据、光环新网、奥飞数据、博大数据、尚航科技）三梯队格局。供给端看，科智咨询数据显示，截至 2025 年底全国在运营 IDC 机柜 IT 总容量达 19,933MW，智能算力规模达 159 万 PFlops(FP16)，行业总量继续扩张，但高密机柜、液冷适配与订单化交付能力主要集中在头部服务商；贝恩口径下第三方市场 CR5 仍仅 35%-40%，说明头部化与长尾并存，行业仍有并购整合空间。运营商系 IDC 整体份额保持稳定，但自用机柜投放规模持续收缩，行业定位逐步摆脱传统硬件资源出租，向综合化、一体化算力服务提供商转型。

✧ 当前国内智能算力服务已形成千亿级、市场提升的格局，云算力运营

商占据核心主导地位。据中国信通院《智能算力服务研究报告（2026 年）》测算，2025 年我国智能算力服务市场规模突破 1300 亿元，其中 AI 云企业级服务约 800 亿元、以云电脑、云手机卡时为代表的消费级市场约 500 亿元。从竞争主体看，市场呈现云巨头、运营商、独立 IDC 三足鼎立的态势。公有云 IaaS 领域，据 IDC《中国公有云服务市场（2025 上半年）跟踪》，阿里云以 26.8% 份额居首，华为云（12.9%）、天翼云（12.3%）、移动云（9.4%）、腾讯云（7.9%）紧随其后；在更高价值的 AI 云赛道，据 Omdia《中国 AI 云市场份额 2025》，2025 年中国 AI 云市场总规模达 567 亿元，阿里云以 38.1% 居首，火山引擎（20.4%）、百度云（9.4%）、腾讯云（6.3%）、天翼云（5.9%）分列其后，前五家合计占比超过八成。整体看，阿里云、天翼云、移动云、华为云、腾讯云等头部云厂商稳居行业前列，充分体现出云算力运营商在市场

中的核心主导地位。

- 行业各类市场主体的竞争逻辑已发生根本性转变，不同类型服务商的核心能力壁垒呈现明显分化，行业分层竞争格局基本成型：
- ✧ 传统 IDC 服务商普遍面临从基础机房托管向专业化算力服务升级的转型压力，行业内部分化差距持续拉大。具备前瞻布局能力的头部企业已率先完成人工智能数据中心的迭代升级。以润泽科技为典型代表，企业完成京津冀、长三角、粤港澳大湾区、成渝地区双城经济圈、海南自由贸易港、甘肃六大区域布局，建成 7 个人工智能计算中心基础设施集群，构建起覆盖基础设施、算法工具、智能平台到解决方案的算力服务体系。2023 年 7 月，其廊坊基地 A-11 项目交付，为全国首例整栋纯液冷智算中心，项目采用自研智能低碳全域制冷系统，实现服务器液体制冷与全域自然冷却。与此同时，中国电信京津冀智能算力中心建成京津冀地区唯一国产液冷单集群万卡智算资源池，液冷单池算力规模达 3760PFLOPS，芯片、服务器、高速互联网络等核心硬件均采用国内信创厂商产品，核心环节国产化率达到 100%，实现硬件、软件、配套技术全链路自主可控，验证了头部算力园区绑定国家算力枢纽、布局液冷高密算力、推进全链条国产化替代的差异化发展路径。而多数转型滞后的中小 IDC 服务商，仍局限于低毛利的传统机房托管业务，受制于行业阶段性产能过剩、客户议价能力提升等因素，经营压力持续加剧。
- ✧ 云算力运营商依托全栈技术能力构筑深厚竞争壁垒，核心优势体现在底层异构算力统一纳管、中层资源智能调度、上层行业场景适配的全链条一体化整合能力。依据信通院智能算力服务体系框架（本章节图表 13），行业整体分为资源服务、互联互通服务、应用服务三大核心层级。头部云厂商通过算力资

源池化技术，将物理分散的计算、存储、网络硬件资源进行抽象整合，形成逻辑统一的算力资源池，结合容器化、虚拟化技术实现算力资源的弹性调配与多租户隔离，有效提升 GPU 等高端异构算力资源的利用效率。目前头部公有云厂商已在全国多省市落地国产化 AI 算力集群，深度承接各地政企智能化改造项目，同时在低碳算力集群建设、私有化算力专属部署等细分赛道形成差异化竞争优势。

✧ 从产业演进趋势来看，国内算力服务正沿着资源层、平台层、应用层持续迭代升级，算力硬件、智能调度软件、行业场景应用相结合的一体化交付模式，已经成为行业主流发展方向。底层算力服务以裸金属、GPU 卡时租赁为主，属于重资产运营模式，盈利空间主要取决于硬件折旧周期与机柜、算力出租率，盈利稳定性较弱；中层平台服务以容器化算力调度为核心，通过算力池化、智能编排技术实现异构算力的精细化切分与弹性供给，大幅提升资源利用效率；顶层应用服务以模型推理、Token 计量计费为核心形态，围绕具体业务任务交付服务成果，商业模式更贴合下游实际需求，服务附加值与市场定价权显著提升。

✧ 这一服务模式的迭代升级，为产业链中上游市场主体带来明确的发展机遇。硬件端层面，异构芯片兼容适配、全域算力资源池化、高速网络互联、规模化集群组网等技术能力，成为设备厂商的核心竞争门槛，具备全栈液冷算力解决方案落地能力的企业，将充分受益于智算中心规模化建设浪潮。软件端层面，算力智能调度、虚拟化管控、标准化计量计费等平台能力，是行业从硬件资源售卖转向专业化服务输出的核心抓手，平台运营主体可通过交易撮合佣金、流量分成、增值技术服务实现多元变现，商业模式相较重资产硬件租赁更

轻量化、盈利弹性更强。场景端层面，模型托管、行业定制化应用交付、算力调用分成等高阶服务，能够进一步打开行业盈利空间，但业务落地成效高度依赖下游 AI 应用场景的商业化释放速度。

- ✧ **同时也需正视行业发展的潜在风险，底层算力租赁模式存在算力单卡利用率波动、硬件折旧成本前置的问题。**若企业无法锁定长期稳定客户、算力出租率持续偏低，重资产运营模式将面临资产减值、现金流承压等经营风险。而 MaaS 等高阶算力服务虽具备更高附加值，但行业商业化落地尚处初期，收入兑现节奏受下游产业景气度影响较大，存在落地进度不及预期的不确定性。
- ✧ **整体而言，算力行业的价值分配重心已彻底从传统资源持有端，转向专业化运营服务端。**单纯的机房空间、硬件租赁业务收入占比持续回落，而算力智能调度、专业运维保障、场景化定制适配等高附加值综合服务收入占比稳步提升。现阶段，定制化项目交付能力、长期稳定运营能力、跨区域算力协同调度能力，已成为区分头部优质服务商与中小市场参与者的核心壁垒。

5.2 上游：算力硬件支撑层五大高景气配套赛道研判

- ❖ **算力上游为产业提供硬件与基础软件底座，核心赛道围绕算力生成、传输、散热、存储四大需求展开。**行业的核心特征是“供给约束与技术迭代并行”。下游 AI 业务的快速增长需求，推动各配套赛道技术快速迭代，产业链价值量占比持续提升。其中高景气度的五大细分赛道分别为 AI 计算芯片、高速互联芯片、高速光模块、液冷、AI 存储系统，各细分赛道在统计口径、行业成熟度、技术壁垒、头部竞争格局层面均存在显著差异。

5.2.1 液冷赛道：从“可选配置”升级为“主流配套方案”

- 随着智算中心算力密度大幅提升，传统风冷方案已无法满足高密机柜的散热需求，液冷技术成为支撑算力硬件稳定运行的核心保障，行业进入加速放量阶段。
- ❖ 从市场规模看，据赛迪顾问电子信息产业研究中心《2025-2026 年中国液冷数据中心市场报告》测算，2025 年中国液冷数据中心市场规模已达 159.8 亿元，同比增长 45.2%，连续三年保持 30%以上高增长；报告预测未来三年行业将维持 35%以上增速，2028 年市场规模有望达到 470.4 亿元，实现三年翻倍以上增长。从技术路径演进方向看，冷板式液冷凭借改造成本低、兼容现有架构的优势占据绝对主导地位，浸没式液冷为超高热密度场景下的技术发展方向，智算与互联网是液冷落地的两大核心应用场景。行业长期看，液冷正在从可选的散热方案逐步转变为高密智算场景的必选项，竞争焦点也由技术方案验证转向产能交付与工程落地能力。
- ❖ 行业壁垒正从单一组件制造，向一体化集成方案快速迁移，市场集中度较高：第一梯队为具备全栈液冷能力的头部科技企业，合计占据较高市场份额；其中头部厂商凭借浸没式液冷全栈解决方案位居行业前列，产品已部署于全国多个超大规模智算中心；另有多家厂商在开放式冷板液冷领域形成差异化产品布局，适配国内主流智算中心的高密机柜需求。**在核心组件环节，国内企业已具备全球竞争优势：**冷板/CDU 环节由多家液冷系统集成企业引领；浸没式液冷技术领域有专业厂商布局；冷却液环节的国产氟化液供给厂商已实现核心耗材的国产自主可控。

- ❖ **赛道核心结构性矛盾：改造成本与能效收益的博弈。**存量数据中心液冷改造涉及机房结构重构、管路铺设、冷却液循环系统安装等高成本工程，单机柜改造成本显著高于新建项目，而改造后的能效收益回收周期受当地电价、PUE 监管力度和实际业务负载影响，存在较大不确定性。这一矛盾使得液冷推广呈现新建项目快于存量改造的特征。

5.2.2 高速光模块赛道：AI 算力传输的“神经中枢”

- **光模块承担算力集群内部、跨集群之间服务器的数据互联传输功能，是整个算力网络的关键硬件组件。**算力集群高并发的数据交互特点，直接牵引光模块产品迭代节奏，下游大模型训练产生的海量数据交换需求，是拉动行业升级的重要动因。
- ❖ **行业当前同时出现产品快速迭代、量价同步上行的发展态势。**根据观研报告网《中国光模块行业现状深度研究与投资前景预测报告（2026-2033 年）》网页公开摘要测算，2025 年全球光模块市场规模达到 268 亿美元，同比增长 74.03%；预计 2030 年全球市场规模将达到 1086 亿美元，2026-2030 年复合年增长率 32.3%（观研报告口径）。AI 大模型带来 token 调用规模指数级扩张，海内外云厂商持续加大资本开支投入，数通市场规模已经超过传统电信市场，成为光模块行业最主要增长来源。国内算力基础设施建设持续落地，国内光模块市场增长动能已经由传统通信业务转向算力相关场景；测算显示，2024 年中国光模块市场规模约 329 亿元，同比增长 52.31%，2025 年接近 449 亿元，同比增长 36.47%，预计 2030 年国内市场规模有望突破 1520 亿元。
- ❖ **这一增长的核心驱动逻辑，是 AI 算力的快速增长重构了行业需求边界：**大模

型训练单次迭代所产生的数据交互量呈指数级上升，驱动光模块向更高速率、更高传输效率升级。在技术端，我们认为行业产品结构加速迭代，800G 产品已从“过渡选项”升级为数据中心标配，1.6T 产品则在头部算力厂商的推动下，踏入规模化商用初期。竞争格局方面，国内光模块厂商在全球供应链中的份额持续提升，国内多家光模块企业已成为全球头部算力集群的核心供应商；但产业链短板集中在上游环节，高端光芯片、DSP 芯片对外依存度较高，成为制约行业进一步扩张的供给瓶颈。

- ❖ 行业现实挑战集中体现在迭代速度与库存管理之间的冲突。光模块迭代周期远短于传统通信器件，从 800G 向 1.6T、3.2T 升级节奏紧凑。企业需要平衡研发端的前瞻投入和存量产品的库存周转，迭代加速带来上代产品贬值风险，对供应链管控和需求预判能力提出较高要求。叠加客户集中度偏高、关键物料产能约束、进出口与地缘因素扰动，行业经营不确定性进一步加大。

5.2.3 算力核心芯片赛道：算力迭代与互联协同的双重突破

- 算力核心芯片是 AI 算力基础设施的核心硬件底座，整体分为算力计算与高速互联两大核心细分领域，二者分工明确、协同配套，共同决定算力集群的性能上限、运行效率与规模化落地能力。其中 AI 芯片承担核心算力运算职能，是单点算力输出的核心载体；高速互联芯片及核心 IP 负责集群内多节点、多芯片的数据交互与协同调度，是异构算力融合、算力池化扩容的关键支撑。随着大模型训练、推理场景持续扩容，算力集群从单服务器运算走向大规模异构集群协同，算力芯片的综合竞争维度也持续升级，不再局限于单芯片算力参数比拼，延伸至芯片性能、软件生态、互联适配、集群协同、综合成本的全方位竞争，国产替代成为行业长期核心发展主线。

❖**AI 计算芯片：性能追赶与生态突破；**AI 计算芯片是算力硬件体系的核心载体，现阶段产业定位已从传统服务器单一计算器件，演进为决定整套 AI 算力基础设施性能、部署效率与运营成本的核心平台。该类芯片广泛部署于大型数据中心与机柜级算力系统，覆盖大模型预训练、微调、推理、多模态处理及高性能科学计算等核心场景，产品形态包含通用 GPU、专用 AI 加速器、云厂商自研 ASIC 及具备 AI 加速能力的高端服务器 CPU。在算力集群运行链路中，CPU 负责任务调度与数据加载，AI 芯片承接海量矩阵运算、注意力计算与张量处理，搭配 HBM 高带宽内存存储模型权重与中间数据，依托高速互联器件完成多芯片并行同步，构建完整算力运转体系。当前行业竞争格局呈现二元分化特征，海外头部厂商构建的软硬件生态占据高端市场主导地位，国内厂商持续技术迭代，依托本土化算力建设机遇加速实现国产替代。

❖**市场维度高增长确定性明确，行业长期扩容趋势显著。**根据 QYResearch《2026-2032 全球及中国服务器 AI 芯片行业研究及十五五规划分析报告》公开测算数据，2025 年全球服务器 AI 芯片市场销售额达到 570.0 亿美元，预计 2032 年市场规模将增长至 2326.7 亿美元，2026 至 2032 年复合增速可达 20.5%，全球赛道持续维持高景气度。

❖**从行业发展痛点来看，硬件性能与软件生态的错配是核心制约因素。**经过多年技术攻坚，国内头部 AI 芯片单卡理论算力已趋近国际主流产品，硬件层面可基本满足大模型训练、推理的基础算力需求。但行业长期存在生态割裂问题，各厂商独立搭建算子库、编译架构与开发工具链，软硬件适配体系互不兼容，尚未形成行业通用的标准化开发接口，导致下游应用企业跨芯片平台的模型迁移、场景适配成本居高不下，大幅制约了国产芯片硬件性能的实际释放。目

前行业主流破局思路为依托开源生态搭建统一适配中间层，将多厂商、多平台的多对多适配模式优化为通用一对多模式，系统性降低全行业技术适配成本。

❖ **整体而言，国内 AI 计算芯片赛道面临双重结构性约束。**一方面，软件生态建设周期长、投入大、壁垒高，碎片化的生态体系难以短期整合完善，是制约产品规模化商业化落地的核心瓶颈；另一方面，7nm 及以下先进制程受外部环境约束，芯片工艺迭代与性能升级空间受限，压制了国产高端 AI 芯片的性能天花板，形成“硬件快速追赶、生态持续滞后”的长期发展格局。

➤ **高速互联芯片：异构算力融合的底层支撑。**高速互联芯片是衔接算力节点、实现集群协同的核心底层硬件，与 AI 计算芯片形成互补配套关系，核心作用是解决多服务器、多芯片之间的数据传输与同步调度问题，其带宽、延迟、兼容性指标直接决定异构算力的整合效率与集群规模化运行能力，是实现算力池化、统一调度的关键技术底座。如果说 AI 芯片决定单点算力上限，高速互联芯片则决定整体算力集群的协同效率与扩容上限，是大规模异构算力集群落地的必备核心器件。

➤ **受益于全球 AI 算力集群持续扩容，高速互联芯片赛道增长势能强劲。**根据 PWConsulting (PMarketResearch) 《AIServerInterconnectChipMarket》统计与预测,2025 年全球 AI 服务器互联芯片市场规模为 65.0 亿美元,预计 2032 年将达到 365.9 亿美元,2025-2032 年复合增长率达 27.95%。作为高速互联芯片的核心基础组件,SerDesIP 核支撑高速数据传输功能;根据 PWConsulting 《WorldwideSERDESIPCoresMarket》统计与预测,2025 年全球 SerDesIP 核市场规模为 14.1 亿美元,其中数据中心与云计算场景需求占比 49.9%;预计 2032

年该市场将增至 33.5 亿美元，2025-2032 年复合增速为 13.15%。其成长持续受益高端 AI 服务器与算力集群设备迭代升级。

- **技术落地层面，部分国内头部厂商在高速互联领域已实现阶段性技术突破。**
 自主可控的高速互联芯片及 SerDesIP 产品，基本适配当前主流 AI 芯片互联技术标准，可支撑大规模算力集群内多芯片并行数据交互与协同调度，一定程度缓解了大规模算力组网的数据传输压力。对于跨品牌、跨架构异构算力的融合编排，国产高速互联产品可提供底层技术支撑，助力本土算力集群逐步实现资源统筹调度与协同运行。
- **行业结构性矛盾主要体现为技术路线碎片化与大规模集群协同需求的不匹配。**
 当前全球高速互联赛道并行存在 NVLink、CXL 及厂商自研协议等多条技术路线，不同技术体系在硬件接口、传输协议、交互逻辑层面存在差异，整体兼容性偏弱。行业尚未形成统一通用的互联标准，使得异构算力节点互通难度较高、跨平台适配成本偏大，多数混合架构算力集群的协同效率、数据吞吐能力及运行稳定性，普遍弱于同构算力集群，一定程度制约了国内异构算力规模化、一体化落地进程。
- **整体来看，算力核心芯片赛道呈现“计算端追赶性能、互联端完善协同”的差异化发展特征，两大细分领域均存在显著结构性短板。** AI 计算芯片的主要约束集中在软件生态碎片化与先进制程受限，硬件参数已逐步趋近国际同类产品，但生态短板制约规模化落地；高速互联芯片的核心瓶颈在于行业技术标准碎片化，影响异构算力集群整体协同效能释放。长期来看，硬件技术持续迭代、软件生态逐步统一、行业标准规范化建设，有望成为国内算力芯片实现国产替代、支撑算力产业扩容发展的核心突破方向。

5.2.4 存储系统赛道：解决算力快速增长的“存墙”瓶颈

- 大模型训练需要海量数据的高吞吐率、低时延存储访问，存储系统已成为决定算力集群实际性能释放的关键瓶颈，可称之为 AI 时代的“存墙”问题。其成因主要有三：其一，模型参数规模、训练数据量与上下文长度持续增长，训练过程中的数据读写与检查点（Checkpoint）存储消耗随之快速放大；其二，高带宽、低时延的专用存储介质需求快速增长，以 HBM 高带宽存储器为代表的高端介质已成为 AI 计算芯片的标配；其三，算力集群的规模扩张，带来存储容量的配套式增长。
- ❖ 从技术演进方向看，存储介质迭代与架构变革同步推进。传统集中式存储架构，在 AI 场景高并发读写压力下容易出现控制器瓶颈，难以匹配大规模智算集群的业务特征，分布式直连存储方案逐步获得更多落地应用。国内部分头部存储厂商相关产品已经适配国内主流算力集群业务场景，可在保障数据高吞吐、低时延访问的基础上，帮助算力集群优化整体存储部署成本。
- ❖ 供给层面，HBM 市场呈现需求快速扩张、供给高度集中、交付周期承压的格局。全球 HBM 产能主要集中于 SK 海力士、三星、美光三家原厂；Counterpoint 数据显示，2026 年第一季度 SK 海力士 HBM 市场份额约 58%，位居行业首位。美光管理层在财报电话会中披露，2026 自然年度的 HBM 产能已经通过长期协议基本锁定，AI 服务器配套内存交付周期明显拉长。SK 海力士 CEO 于 2026 年 8 月对外表态，AI 带来的存储紧张局面或将延续至 2030 年底；该判断属于企业管理层的前瞻性判断，实际供需格局会受产能扩产节奏、下游 AI 资本开支变化扰动，具备不确定性。

- ❖ HBM 已经成为继先进制程之后，国内算力供应链需要重点突破的环节。国产化方面，据产业链信息，部分国内厂商正推进 HBM 国产替代验证，包括与本土 AI 芯片的适配测试。整体看国产 HBM 仍处在技术验证阶段，距离大规模商用尚要跨越客户认证、良率爬坡等多重环节，短期内国内算力集群对海外高端 HBM 产品的依赖难以完全消除。
- ❖ 行业需要在高性能介质的使用需求，与大规模集群部署的成本经济性之间寻找平衡点。HBM 等高性能存储介质采购成本偏高，如果在超大规模算力集群中全域配置，将带来显著的成本压力；分布式直连存储架构性能优势突出，但相比传统集中式方案，部署复杂度、运维难度也有所抬升。综合成本与性能考量，混合存储架构成为现阶段行业较为务实的过渡选择，针对热、温、冷数据分层采用差异化存储介质与架构。

5.3 下游：高算力需求垂直行业落地研判

- 下游是验证全产业价值的最终环节。行业发展重心已从“训练规模竞赛”，转向“推理规模化落地”，推理算力需求的增长持续超过训练算力需求，成为产业核心增量来源。不同行业的算力需求特征、部署逻辑、落地进度存在显著差异，直接牵引中上游算力资源的配置节奏。

5.3.1 核心增量逻辑：推理侧成为算力消耗主体

- ❖ 从行业发展阶段看，AI 产业发展重心已从超大模型预训练的“军备竞赛”，转向千行百业的规模化推理落地，推理算力需求增速已持续高于训练算力，成为行业核心增量来源。这一趋势背后的逻辑是，大模型行业的发展阶段发生了本质变化：在行业初期，企业以训练超大参数模型为核心目标，追求极致

的集群算力规模，带动训练算力需求高速增长；但随着模型技术逐步成熟，行业关注点转向落地变现，推理是实现商业价值的必经环节，各类行业应用的规模化上线，持续拉动实时、高并发的推理算力需求。

- ❖ 从需求属性看，训练、推理两大类需求的特征存在显著差异，对算力资源的配置提出了显著分化的要求：大规模训练业务具备高吞吐、可容忍时延的特征，适合部署在电力、土地资源丰富的西部枢纽节点，以降低算力长期综合成本；而推理业务要求低时延、高并发，需要贴近终端用户部署，以保障用户体验，两类需求的资源配置逻辑天然分化。

5.3.2 高价值行业场景落地

- 高价值行业场景，是算力产业实现商业闭环的关键突破口。该类场景付费能力较强、业务需求相对稳定，能够为算力产业中长期投入提供业务支撑，也深刻影响算力行业的长期商业化天花板。不同行业的算力诉求存在明显差异，对中上游算力软硬件产品的场景适配能力提出更高要求。
- ❖ **自动驾驶代表车-云协同算力的核心增量方向**。自动驾驶是当前算力需求增长确定性相对较高的行业场景，车端与云端算力形成强关联、强协同的业务闭环，对算力体系综合能力提出较高要求。**车端层面**，伴随自动驾驶等级迭代，单车算力需求呈现数量级提升，高阶智驾对车载算力、云端仿真算力均存在刚性需求。**云端维度**，仿真测试是主要算力消耗场景，复杂多智能体仿真任务内存消耗突出；多路摄像头、激光雷达产生的海量感知数据，每小时生成大量待处理缓存数据，对算力集群存储 I/O 能力形成较高压力。
- ❖ 为匹配智驾业务对云端算力的需求，据公开产业信息，国内部分整车企业（如

蔚来、小鹏等）已建设大规模异构训练算力集群，用于端到端大模型训练。依托工信部算力互联互通试点建设，成渝（重庆）枢纽已建成主城都市区 1ms、成渝城市群 5ms 的全光算力时延圈，支撑成渝区域内部算力调度；同时落地“疆算入渝”跨域算力协同创新实践，采用新疆绿电算力承担训练任务、重庆本地算力承接实时推理的模式，实现跨省算网协同。在成渝（重庆）等国家算力枢纽节点，面向车-云协同场景的算力基础设施正加速落地，为区域智能驾驶业务提供低时延网络与算力底座。

❖ **智能制造是工业边缘算力的高增长赛道。**工业 AI 落地对算力提出低时延、高带宽、高安全性要求，业务逻辑与互联网通用算力场景存在明显区别，是工业 AI 实现商业化落地的重要基础。核心需求来源于工业 AI 质检、设备实时控制等业务场景，工业边缘节点需要具备充足实时算力，保障高速高精度产品质检、设备故障诊断、生产流程自动控制等业务的低时延运行。从产业实践来看，智能制造是算力商业化落地推进较快的行业方向之一，国内已积累不少可参考的行业实践案例。例如某制造企业将工业 AI 能力下沉至边缘算力节点后，质检环节的效率、准确率得到改善，检测耗时有所压缩，漏检风险得到降低，项目投资回报表现优于传统方案。

❖ **基础设施层面，国内多个算力枢纽布局面向智能制造的边缘算力节点。**重庆正构建云一边一端协同、算运存融合的多层次算力体系，并在区县与园区、厂区分下沉边缘算力节点；据重庆市“十五五”AI 规划与工业产业大脑指南，区域边缘节点可面向智能检测、故障分析、生产控制等工业场景提供实时算力支撑。重庆搭建多层次算力体系，区域边缘节点可承接工业场景实时业务；福州新区（长乐区）建设面向重点工业园区的工业互联网算力支撑体系，服务本

地制造企业智能化转型。

- ✧ 除自动驾驶、智能制造之外，政务、金融、医疗是算力落地的重点行业场景。该类行业普遍付费意愿较强，同时合规监管门槛高，算力建设的核心约束集中在数据安全与合规管控层面。
- ✧ 政务场景：算力需求集中于政务数据共享、城市综合治理、公共安全防护等领域，对算力安全可控、数据本地留存能力要求严苛。地方政务算力建设较多采用“专属集群+加密传输”模式，政务数据的存储、计算部署于隔离专属集群，业务访问设置权限管控。
- ✧ 金融场景：算力需求集中在实时风控、反欺诈、智能投顾/客服等方向，重点关注系统高可用、低时延、安全合规。头部金融机构核心业务多采用“两地三中心”部署架构，核心业务数据与计算资源运行在隔离专属集群，实施访问权限管理。
- ✧ 医疗场景：算力需求覆盖医学影像辅助诊断、基因组测序、临床辅助决策等，要求算力具备高带宽、低时延能力，严格匹配医疗数据合规规则。行业较多采用“本地存储+远程调度+加密传输”方案，在不迁移核心原始医疗数据的前提下完成跨域算力调度。
- ✧ 上述高合规属性行业的落地实践，印证“数据不动、算力动”是具备可行性的合规业务范式，同时也对算力服务商的合规治理、专业化交付能力提出更高标准。
- ✧ 综合算力上中下游各环节的研判，可梳理当前国内算力产业整体发展逻辑：产业表层的规模扩张、技术迭代、商业模式重构，背后由多组结构性矛盾共同

驱动。行业竞争逻辑已经从早期单纯比拼硬件建设规模，逐步转向全产业链协同、专业化运营、合规化赋能、差异化分工的综合性系统竞争。

5.4 本章小结：底层矛盾约束下的行业确定性演进方向

✧ 分层总结如下：**中游算力运营层面**，智算中心建设从规模扩张转向效率优先，算力运营商的核心竞争力从资源端向运营端迁移，调度技术、枢纽资源和长期客户体系构成三重壁垒；**上游硬件层面**，液冷、光模块、AI 计算芯片、高速互联芯片、存储系统五大高景气赛道并行增长，各赛道核心矛盾差异显著（液冷改造成本、光模块迭代压力、芯片软件生态、互联标准碎片化、存储成本平衡），需差异化研判投资逻辑与产业节奏；**下游场景层面**，推理算力成为核心增量，自动驾驶、智能制造、政务、金融、医疗构成高价值场景梯队，场景落地进度与算力资源配置策略紧密联动。

❖ 从赛道层面看，各环节的确定性趋势已非常清晰：

- **上游**：光模块、液冷、AI 计算芯片、高速互联芯片、存储系统是高景气度支撑赛道，行业增长的核心驱动逻辑，是算力集群向更高带宽、更高密级、更异构协同的方向迭代；
- **中游**：智算中心建设将严格遵循全国一体化算力网的统筹布局，从“规模扩张”转向“效率优先”；算力运营的价值重心，将从资源端进一步向运营端倾斜；
- **下游**：推理侧成为算力需求的核心增量来源，自动驾驶、智能制造、政务、金融、医疗是支撑行业长期增长的核心高价值场景。

❖ 本章梳理的行业逻辑与赛道趋势，本质上是底层矛盾约束下的行业确定性演进方向，推导行业中长期发展的确定性规律，为后续产业机会的筛选提供明

确的锚定依据。

第六章 战略抉择：行业风险、机会分析及行动建议

- 中国的算力产业当前正从“大规模基建布局”转向“高质量运营增效”的关键发展时期。行业所有扩张节奏、技术迭代、模式重构、格局分化均由六类长期并存、相互制衡的底层结构性矛盾驱动：**供需空间错配、算力层级错配、软硬件生态错配、投入回报错配、标准异构错配、算电能耗错配**。基于矛盾演化的确定性传导，行业政策、技术、市场、竞争维度呈现全新约束特征，中长期发展方向清晰可辨。

6.1 六类结构性矛盾：行业发展的底层逻辑

- 我们在 1.4 章节梳理的四大运营痛点，是本节六类产业底层结构性矛盾在企业经营层面的集中外化表现。算力产业发展过程中出现的阶段性发展困境、中长期演化方向，本质上是六类结构性矛盾动态博弈、此消彼长，推动行业螺旋式迭代升级的结果。
- ❖ **供需空间错配**：东部沿海产业集聚旺盛，算力业务需求庞大，但能耗指标趋严、数据中心机房扩容空间受到约束；西部地区绿电资源充沛、土地成本具备优势，但本地 AI 产业应用基础相对薄弱。跨区域开展算力调度时，会受到传输时延、网络成本、数据安全合规多重条件约束，制约算力资源东西向高效流转配置。
- ❖ **算力层级错配**：大模型训练所需万卡级超高密度智算资源供给持续偏紧；与此同时，部分存量通用 IDC 集群的低端通用算力产能出现闲置，行业逐步呈现高端算力资源供给紧张、低端算力资产利用率不足的两极化格局。

- ❖ **软硬件生态错配：**部分国产 AI 芯片硬件单卡理论算力已接近国际主流产品水平，但不同厂商算子库、编译工具链体系相对割裂。跨芯片架构开展模型迁移与适配需要投入较高成本，软件生态碎片化问题，制约国产算力实现更大规模商业化落地。
- ❖ **投入回报错配：**算力基础设施属于重资产、长周期属性赛道，传统机柜租赁商业模式存在盈利天花板。算力标准化计价体系仍处于试点培育阶段，全行业统一的算力交易、结算规则尚未完全建立，资产资本投入与现金流产出之间的错配压力客观存在。
- ❖ **标准异构错配：**建设全国一体化算力调度网络，需要标准化接口、统一算力标识体系作为基础底座。当前多架构芯片、多厂商硬件设备通信协议尚未完全打通，异构算力集群实现统一纳管、跨域协同调度，在技术实现、行业标准层面仍然存在现实壁垒。
- ❖ **算电能耗错配：**AI 智算集群属于高载能业务场景，行业对于 碳排放的监管要求持续收紧；国内风光等可再生能源装机规模快速扩张，但新能源具备间歇性、波动性特征，配套电力储能建设成本较高。算力业务 7×24 小时稳定运行诉求，与绿色低碳发展目标之间，形成需要持续权衡平衡的现实难题。

图表 14：驱动算力产业发展的六类底层结构性矛盾



来源：国金证券研究所

6.2 九条核心分析：矛盾传导下的综合判断

- ❖ **综合六类结构性矛盾的相互作用**，结合当前产业运行现状，梳理形成覆盖行业全局的九条核心结论，为后续风险研判、趋势推演、策略制定提供统一逻辑支撑：
- ❖ **发展阶段梳理**：全国一体化算力网建设，已经从“东数西算”物理基建布局阶段，步入以算网融合、全域调度、算电协同为重点的运营效率提升阶段，调度能力成为行业发展的重要抓手。
- ❖ **政策趋势研判**：大算力枢纽差异化管控机制，是重塑国内算力供给空间格局的核心变量。非枢纽地区从严管控新增大型商业算力项目的导向，将持续约

束行业盲目扩产行为。

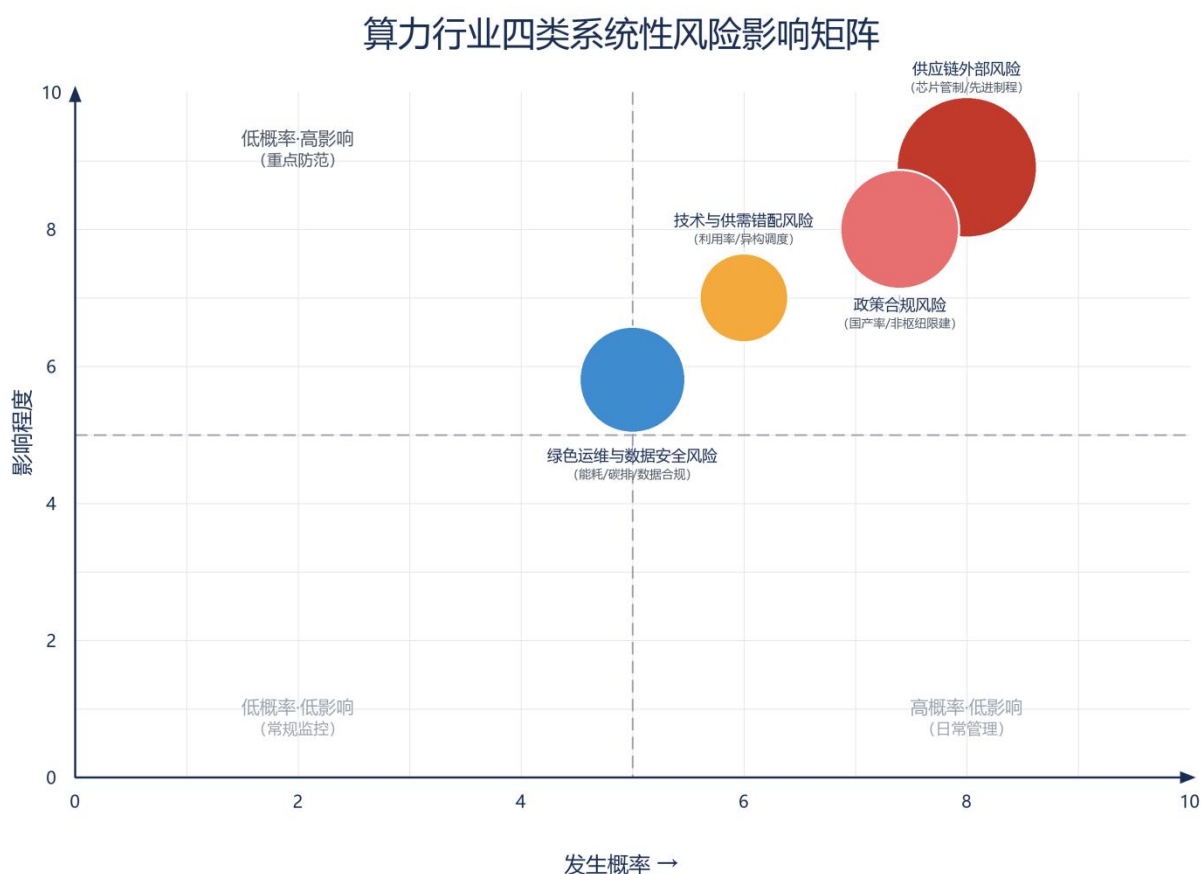
- ❖ **基础设施认知：**工信部提出的“1+M+N”国家算力互联互通节点体系，是破除区域、主体、技术层面壁垒，实现算力跨域高效流通的重要基础，也是提升行业资源配置效率的关键条件。
- ❖ **商业模式理解：**行业盈利逻辑从“硬件规模租赁”向“分层价值服务”逐层升级，形成“基础资源出租+词元 GaaS 一体化服务+行业专属算力定制”的三级盈利架构。
- ❖ **技术路线判断：**液冷技术已经从高密智算场景的可选配置，逐步成为行业主流配套方向。国产 AI 芯片市场占比不断提升，但软件生态割裂的短板，短期内难以从根本上得到解决。
- ❖ **价值分配研判：**产业价值分配重心，由上游硬件制造环节逐步向中游算力运营服务环节转移。具备成熟调度能力、行业落地运营能力的服务商，有望分得更多产业利润。
- ❖ **交易体系总结：**算力市场化交易仍处于试点探索阶段，整体成熟度明显低于电力交易市场。结合产业实际情况看，中长期大概率沿着“先落地长期合约，再发展现货交易，逐步探索辅助服务”的路径推进；该判断属于产业层面推演，并非监管既定政策。
- ❖ **竞争格局考量：**行业头部集中效应逐步显现，头部主体依托枢纽资源资质、自研调度技术、长期行业客户订单的三重壁垒持续巩固优势，市场竞争从“硬件参数比拼”转向“全链条协同交付能力”的综合较量。
- ❖ **赛道选择建议：**推理算力需求增速已经超过训练算力，成为行业主要增量。金

融、政务、医疗等高合规要求领域，是行业专属算力的重要落地突破口；光模块、液冷、AI 芯片、高速互联、存储系统具备较好的产业景气度。

6.3 四类行业性风险：矛盾外化的约束边界

- 结构性矛盾在现阶段集中形成行业不可忽视的系统性约束。按照风险来源、影响范围，将行业系统性风险划分为四大类别，各类风险相互关联，共同划定行业扩张的现实边界。

图表 15：算力行业四类行业风险影响矩阵



来源：国金证券研究所

注：风险定位为研究判断，基于当前政策环境与产业现状，随外部条件变化动态调整

6.3.1 政策合规风险压低行业上行天花板

- ❖ 八大算力枢纽差异化布局管理、非枢纽地区严控新建大型算力项目、机房 PUE 硬性指标管理，三重政策要求划定算力设施的布局边界。部分地区对国有资本投资的智算项目，设置国产硬件采购的相关引导要求，各地执行尺度存在差异。高度依赖海外芯片、布局在非枢纽区域的算力项目，将面临审批、能耗指标获取、合规整改、资产利用率不及预期等现实压力。

6.3.2 供应链外部风险约束行业技术升级空间

- ❖ 海外高端 AI 芯片出口管制持续收紧，国内 7nm 及以下先进制程产能存在阶段性供给压力；HBM 等关键存储介质供给高度集中于海外厂商，同样受外部管制因素影响，会限制高密度智算集群的装机规模。液冷核心零部件、高速光芯片等部分元器件需要进口，存在供应链扰动的可能性，会挤压中小算力运营商盈利空间，赛道内部市场竞争进一步加剧。

6.3.3 技术与供需错配风险形成行业产能效率瓶颈

- ❖ 算力层级错配矛盾长期存在，部分地方盲目上马智算项目，易出现“建得起、用不起、调不动”的经营困境；异构算力统一纳管、跨域低时延调度的技术瓶颈，短期难以完全突破；国产芯片与大模型适配成本较高、生态效率低于行业预期，直接制约集群有效算力释放，行业资源配置效率提升存在明显技术障碍。

6.3.4 绿色运维与数据安全风险考验行业长期运营的稳定性

- ❖ 算力高能耗属性匹配日趋严格的碳排放、能耗监管标准，绿电供应稳定性、余热回收利用经济性，都是长期运营的刚性约束；液冷系统运维、跨区域数据传

输，对安全管控提出更高要求，数据泄露、传输合规的潜在风险，持续抬高行业运营门槛。

6.4 八大确定性较高的趋势：基于矛盾演化的方向前瞻推演

- ❖ 综合前述六类结构性矛盾以及四大系统性风险的约束边界，产业现实约束与技术、政策变量相互交织，既对算力行业扩张形成客观限制，也在重塑产业演进路径。立足于当前产业发展现状，结合政策导向、技术迭代、供需格局变化，我们进一步推演产业中长期演化方向，梳理出八大确定性较高的趋势。

趋势一：算力网络完成从“物理连接”到“逻辑统一”的升级

- ❖ 全国一体化算力网将按“资源入网→智能调度→市场交易”的顺序分阶段成熟，全域调度能力实现算力跨区域、跨主体、跨架构的自由调配；市场化交易机制成熟周期较长，中长期将形成“中长期合约+现货交易+辅助服务”的多层次交易框架，与电力交易体系存在本质场景差异。

趋势二：词元计价成为行业标准化底座，与传统模式长期并存

- ❖ 算力商品化进程持续推进，词元计价的行业渗透率稳步提升，逐步成为行业标准化定价核心方向；但机柜租赁、词元订阅两种模式将长期共存，行业统一词元计量口径、跨平台结算规则的落地节奏存在不确定性，市场客户认知仍需要持续培育。

趋势三：国产 AI 芯片市场份额持续提升，实现分层级行业替代

- ❖ 国产 AI 加速卡出货占比持续上行，在政务、工业、自动驾驶等行业专属算力场景中，逐步形成规模化替代优势；但受软件生态割裂、先进制程产能约束、海外技术封锁三重变量限制，短期无法实现全行业、大中小算力场景的全覆

盖替代。

趋势四：液冷技术渗透率快速提升，风冷液冷形成混合适配格局

- ❖ 高功率密度智算场景中，液冷技术的标配地位进一步巩固；存量机房受改造成本高、投资回收周期长等因素制约，液冷改造进度存在显著区域差异，风冷、液冷技术将在不同算力场景中长期共存，形成分层适配格局。

趋势五：行业头部集中格局加剧，中小主体向垂直赛道差异化转型

- ❖ 行业资源禀赋分化效应持续放大，头部运营主体凭借枢纽节点资源、核心调度系统与长期客户深度绑定所形成的复合壁垒，持续收拢市场份额；中小算力服务商难以在全链条层面与头部企业竞争，将重点深耕行业专属算力、边缘算力等垂直细分赛道，行业集中格局的最终形态仍有待观察。

趋势六：算电协同在西部枢纽先行落地，成为绿色发展标准配置

- ❖ 依托西部枢纽的绿电供给优势，源网荷储、算电协同模式将在西部训练算力集群中率先试点推广，配套大规模储能、柔性负荷调节技术，成为智算中心标准配置；但风光发电的间歇性特征、储能项目高投入成本，仍是算电协同规模化推广的核心约束。

趋势七：高速光模块、CPO 技术迭代升级，破解网络传输瓶颈

- ❖ 作为算力网络核心互联器件，1.6T 光模块当前进入小规模商用试点，3.2T、CPO 技术是长期迭代方向；技术迭代将有效突破算力集群内部、跨地域算力调度的网络时延和带宽瓶颈，量产落地时点存在一定技术不确定性。

趋势八：行业专属算力、边缘算力成为核心增量，实现双层协同落地

- ❖ 金融、政务、医疗等高合规行业专属算力需求持续扩容，定制化服务成为行业新的增长极；城市边缘算力节点逐步布局，承接低时延推理需求，与西部枢纽训练算力、东部枢纽行业专属算力形成分层协同架构，边缘算力商业化盈利模式仍在持续验证。

6.5 分主体的战略建议：趋势下的实操路径建议

- 针对产业经营主体、资本市场投资人、政策制定主体的不同定位，结合行业中长期确定性趋势与核心风险约束，我们国金证券产业研究中心团队为三类主体分别适配了可落地的战略建议方案：

6.5.1 面向算力产业经营企业的管理层：锚定布局、升级业务、抢占红利、筑牢壁垒

- ❖ 行业企业需主动遵循顶层政策约束，依托确定性趋势构建长期竞争力，重点落地四项核心策略：
- ❖ **精准选址布局，匹配区域分工：**严格遵循八大枢纽差异化布局导向，非枢纽区域原则上不再新上大型商业算力项目；结合业务属性匹配区域资源禀赋，大模型训练算力优先布局西部低电价、高绿电占比区域，推理算力、行业专属算力优先布局东部产业集聚区域，贴近终端客户、降低传输时延。
- ❖ **升级业务模式，锚定高附加值赛道：**主动降低传统机柜租赁收入占比，加快业务分层升级：基础资源层提供标准化算力出租服务；中间层打造词元订阅、弹性算力等标准化产品；高价值层深耕行业专属算力定制、全链路算力运营服务，以高毛利业务对冲重资产折旧压力。
- ❖ **把握三大行业红利窗口，提前布局核心资源：**一是词元计价红利，率先搭建词

元计量、结算体系，参与行业标准制定，抢占算力商品化先机；二是国产算力替代红利，提前与国产芯片厂商、模型企业协同适配，搭建全栈国产化算力服务能力，匹配地方国产化政策要求；三是液冷绿电配套红利，新建高密智算中心优先配套液冷技术、绿电储能系统，满足 PUE 监管和头部客户 ESG 采购标准。

- ❖ **筑牢行业客户壁垒，稳定长期收入盘：**重点拓展金融、政务、工业制造、自动驾驶等高合规、高粘性行业客户，以专属算力定制方案、跨域调度服务能力为核心，签订长期框架订单，锁定基础算力需求，降低行业周期波动影响。

6.5.2 面向一二级市场投资人：规避过剩赛道、重构估值逻辑、分周期布局、防控明确风险

- ❖ **投资布局需聚焦行业价值转移主线，规避结构性产能过剩赛道，建立贴合行业真实发展逻辑的研判框架：**
- ❖ **明确赛道筛选标准，规避结构性产能过剩风险：**弱化“机柜规模、芯片算力”等单一量化指标，重点筛选具备三重核心壁垒的优质标的：拥有八大枢纽地区合规算力资源、具备自研异构算力调度平台技术能力、绑定大型行业长期客户订单；传统无差异化竞争力的通用 IDC 项目，属于明确产能过剩赛道，应主动规避。
- ❖ **重构资产估值逻辑，聚焦效率与合规指标：**从传统“机柜数量、硬件规模”的重资产估值导向，转向以“算力资源利用率、词元产出效率、异构调度技术能力、行业合规资质”为核心的多维估值框架；同等资产规模下，算力利用率、词元产出效率更优的运营企业，具备明确估值溢价空间。

- ❖ **分周期布局高景气配套赛道，匹配行业成长节奏：**结合行业技术落地节奏，分阶段配置标的、把握结构性机会：短期（2026-2027 年）重点关注液冷配套、800G/1.6T 高速光模块行业，直接受益高密智算建设；中期（2028-2029 年）布局国产 AI 芯片、算力调度软件企业，受益国产化替代、算力规模化交易；长期（2030 年及以后）聚焦算电协同配套、边缘算力运营服务商，匹配行业绿色化、协同化发展趋势。
- ❖ **重点规避四类明确投资风险：**政策端规避非枢纽地区算力项目、不符合国产化率和 PUE 管控标准的标的；供应链端规避依赖单一海外芯片供给、无国产化替代预案的运营企业；技术端缺乏调度技术储备、未提前适配国产算力生态的主体；经营端规避重资产投入规模过大、长期订单不足、液冷运维能力薄弱的标的。

6.5.3 面向政策制定者：统筹标准、引导布局、设计交易、完善激励

- ❖ **政策规划聚焦打通行业底层发展壁垒，以标准建设、梯度布局、政策激励为抓手，推动行业高质量发展，避免同质化、低水平重复建设：**
- ❖ **完善全国统一标准体系，打通跨域调度底层壁垒：**加快制定算力资源池化、并网、监测、安全管控、异构调度统一技术规范，统一算力资源计量、评价、定价标准；重点推动“1+M+N”三级调度体系全国落地，打通不同架构芯片、不同厂商设备、不同区域算力资源的通信接口，从技术层面破除行业异构流通壁垒。
- ❖ **差异化规划枢纽产业布局，引导区域协同分工：**强化八大枢纽区域差异化定位，引导西部枢纽重点承接训练算力、数据存储业务，发挥其绿电资源优势；

东部枢纽聚焦推理算力、行业专属算力定制，贴近产业应用端；建立全国算力项目统筹协调机制，严控非枢纽地区新建大型商业算力项目，避免区域间同质化、低水平重复建设。

- ❖ **设计分层算力交易制度，培育市场化流通机制：**借鉴电力市场成熟交易经验，结合算力行业场景特性，构建“中长期合约+现货交易+辅助服务”的多层次算力交易体系；牵头制定全国统一词元计量标准、跨区域结算规则、数据安全治理规范，逐步培育算力商品流通市场，以交易机制倒逼行业运营效率提升。
- ❖ **组合配套激励政策，引导行业绿色化、高质量发展：**从四个维度完善配套支持体系：绿色端出台绿电应用补贴、PUE 达标奖励政策；资产端支持算力基础设施类 REITs 项目落地，盘活存量资产；国产化端对适配国产芯片、构建全栈国产化能力的项目给予政策扶持；低效资产端推动高 PUE 老旧 IDC 项目出清，优化行业整体资源供给效率。

6.6 长期产业不确定性与针对性对冲策略

- ❖ 行业长期发展仍存在四类关键外部变量，国金证券产业研究中心团队给出相应的应对预案，仅供参考：
- ✧ **不确定性一：国际高端 AI 芯片出口管制进一步收紧，直接约束高密智算集群装机规模，扰动行业供给节奏。应对策略：**加快推动国产算力全栈生态建设，推行“模芯协同”开发模式，引导大模型企业与国产芯片厂商同步适配、联合优化；在政务、金融等政策驱动行业场景中，优先采用国产化算力资源，储备本土算力资源池，降低对海外高端芯片的依赖。同时加快推动国产 HBM 等关键存储配套能力建设，支持国产存储原厂扩产与下游认证导入，缓解高端算

力对海外存储的依赖。

- ✧ 不确定性二：词元计价、算力市场化落地节奏不及行业预期，标准化交易体系进展缓慢，制约行业规模化流通。应对策略：由行业主管部门牵头制定全国统一词元计量技术标准，明确算力资源量化评价口径；先在政务、工业、金融等长期合作行业客户间试点推广词元计价套餐，逐步完善定价、结算机制，以行业试点培育市场认知，稳步推进算力商品化。
- ✧ 不确定性三：异构算力调度技术突破速度不及预期，算力利用率提升存在瓶颈；跨域、跨架构调度效率无法满足行业实际需求，制约资源配置效率提升。应对策略：加大算力调度软件、异构适配技术的产学研联合攻关投入，搭建国家级算力调度技术验证平台、行业测试中心，在真实场景中持续优化调度算法；鼓励头部调度技术企业开放标准化接口，推动不同厂商设备间的协议适配，分阶段提升异构资源纳管效率。
- ✧ 不确定性四：跨区域算力调度的数据安全、隐私合规约束趋严，数据跨区域传输面临更严格监管要求，阻碍东西部算力流通。应对策略：推广“数据不动、算力动”的核心理念，普及数据沙箱、水印等安全技术；建立跨区域数据安全协同治理体系，明确数据存储、传输、使用全链条责任边界，在合规前提下推进算力资源跨域调配。

6.7 本章小结

- ❖ 中国的算力产业是支撑我国数字经济高质量发展、推动 AI 行业应用落地的核心底层基础设施。当前行业正处于从“规模扩张”转向“效率提升”的关键战略窗口期，六类底层结构性矛盾划定行业长期发展边界；四大类系统性风险

构成行业经营的硬性约束；八大确定性趋势指明行业演化的核心方向。基于以上研究内容，国金证券产研中心为三类主体提供差异化落地策略，输出贴合行业实际的决策参考建议。

- ❖ 尽管行业发展面临供应链技术约束、生态建设短板、合规监管要求等多重挑战，但在政策端统筹规划、产业端协同攻关、资本端持续布局的共同推动下，“算力像水电一样普惠易用”的行业发展图景具备清晰可预期性。对于产业经营主体、资本市场投资人、政策制定主体而言，把握行业确定性发展趋势，提前布局词元经济、国产化适配、算电协同三大核心主线，针对性防范系统性风险，是未来 3-5 年抢占行业领先地位的关键路径。

风险提示

政策落地不及预期、技术更新迭代不及预期、供应链及外部风险失控、下游客户预算不及预期等风险；提到的公司及业务布局仅为产业案例分析，不构成个股投资推荐。

特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、传播、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于C3级（含C3级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路1088号
紫竹国际大厦5楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建内大街26号
新闻大厦8层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路2028号
皇岗商务中心18楼1806