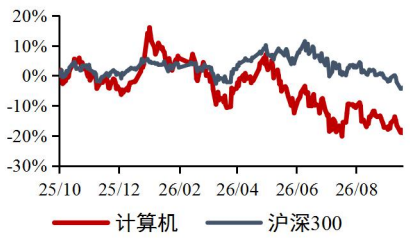
下载日志已记录，仅供内部

计算机 算力租赁专题报告 领先大市-A(维持)

供需持续严重错配，商业模式向 Token 分成跃迁

2026 年 10 月 14 日 行业研究/行业专题报告

计算机板块近一年市场表现



资料来源：常闻

首选股票		评级
300857.SZ	协创数据	买入-A
301248.SZ	杰创智能	买入-A
002990.SZ	盛视科技	买入-A

相关报告：

【山证计算机】北美四大云厂业绩复盘：AI 逐步兑现业绩，Capex 加速上行 2026.6.12

【山证计算机】国产算力全面突破，AI 应用即将爆发 2026.3.16

分析师：

盖斌赫

执业登记编码：S0760522050003

邮箱：gaibinhe@sxzq.com

邹昕宸

执业登记编码：S0760526060002

邮箱：zouxinchen@sxzq.com

投资要点：

- **国内 AI 算力严重供不应求推动今年以来算租行业持续高景气。**在需求端，AI 推理需求呈爆发式增长，叠加企业级 AI Agent 加速落地，单客户算力消耗的乘数效应将进一步抬升算力需求曲线斜率，同时国内外头部云厂商持续增加 AI 投入，推动整体算力市场空间和增长确定性明显提升；而在供给端，英伟达高端 AI 芯片面临多重产能瓶颈约束，叠加美国对高端芯片的出口管制持续收紧，国内可获取的高端芯片增量极为有限。供需矛盾下今年上半年高端 AI 芯片租赁价格保持加速上行，我们预计下半年算力供不应求状况仍无法有效缓解，推动算力租赁价格将维持高位。
- **算力租赁行业市场集中度有望持续提升，重点看好具备供应链优势的厂商。**目前行业参与者多为跨界厂商，已分化为三个梯队，其中，第一梯队已完成大规模算力资源池搭建且锁定长期订单，第二梯队在区域客户资源和本土机房资源上具备差异化优势，而第三梯队仅能提供零散的算力零售服务。随着美国芯片出口管制收紧导致高端卡供货渠道遭封堵，以及中小厂商出租率较低使其面临折旧和利息的固定成本困境，叠加中小厂商缺乏强业务背书推高融资成本，未来中小厂商大概率将逐步出清，市场份额向头部聚拢。基于算租行业的四大核心竞争维度，具备显著供应链能力优势的厂商有望率先脱颖而出，其次在客户资源、资金实力、组网及运维能力上有积累的厂商也将有更强竞争力。目前协创数据综合实力保持领先，同时行云科技、东阳光、盛视科技、杰创智能等厂商算力扩张迅速。
- **算力租赁商业模式正经历从算力资源出租向 Token 分成的跃迁，推理优化能力强且与头部模型厂绑定的公司将充分受益。**卖方市场中大模型厂商等下游客户为抢购算力愿意出让部分 Token 收益，同时算租厂商也有意愿突破传统租赁模式的利润率天花板，于是今年以来算租厂商开始持续探索落地 Token 分成新模式，伴随而来的是更深的产业链嵌入和更高的价值分配比例。参考海外亚马逊与 Anthropic 的合作实践，预计典型的 Token 分成合同将采用“保底租金+超额分成”的双层设计或者纯分成模式，且分成比例设计更偏向于算力提供方。目前国内 Token 分成模式仍处于小规模试点阶段，但未来有望成为算租厂商重要的收入和利润来源，根据我们测算，在中性假设下，到 2028 年算租厂商的 Token 分成收入有望达到 592 亿元。而在 Token 分成模式下，算租厂商除拿卡渠道和资金实力外，还需要优异的推理优化能力及与头部模型厂的深度绑定，因此随着 Token 分成模式的持续渗透，具备强推理优化技术并与智谱等头部模型厂绑定的算租厂商将充分受益。

下载日志已记录，仅供内部参考

➢ **投资建议：**国内高端 AI 算力供给缺口短期内难以弥补，算租行业高景气预计将延续全年。一方面，随着美国芯片出口管制收紧导致高端卡供货渠道遭封堵，以及中小厂商出租率较低使其面临折旧和利息的固定成本困境，未来中小厂商大概率将逐步出清，市场份额向头部聚拢，具备显著供应链能力优势的厂商有望率先脱颖而出，建议关注协创数据、利通电子、杰创智能、盛视科技等；另一方面，算力租赁商业模式正经历从算力资源出租向 Token 分成的跃迁，未来 Token 分成有望成为国内算力租赁厂商的重要收入和利润来源，而在 Token 分成模式下，算租厂商需要优异的推理优化能力及与头部模型厂的深度绑定，建议关注宏景科技、软通动力、中软国际等。

风险提示：AI 算力需求增长不及预期风险，GPU 租赁价格大幅下降风险，地缘政治风险，行业竞争加剧风险，行业标准与监管不确定性风险。

下载日志已记录，仅供内部参考

目录

1. 供需缺口巨大，算租价格持续上行 6

1.1 企业级 AI Agent 加速落地，Token 用量需求爆发式增长 6

1.2 芯片供给刚性约束叠加美国出口管制，国内高端卡供给紧缺 8

1.3 高端卡租赁价格加速上行，下半年将维持高位运行 11

2. 竞争格局和商业模式双重变革 13

2.1 竞争格局：渠道封堵下市场集中度有望提升 13

2.2 商业模式从服务器租赁向 Token 工厂进化，行业估值逻辑重构 18

3. 主要标的 25

3.1 协创数据：算租绝对龙头，步入业绩加速兑现期 25

3.2 宏景科技：智慧城市转型算力服务，利润向上空间打开 27

3.3 利通电子：英伟达 Preferred AI 云伙伴，轻资产运营提高利润弹性 28

4. 投资逻辑 30

5. 风险提示 32

图表目录

图 1： 全球顶尖模型的合计 Token 使用量持续高增 6

图 2： AI 工作负载从“训练主导”转向“推理主导” 6

图 3： 2026-2031 年中国企业智能体市场将迎来陡峭的增长曲线 7

图 4： 全球前九大 CSP 今年资本开支总额预计将达到 8300 亿美元左右，同比增长 79% 8

图 5： 23Q1-27Q4 台积电不同制程晶圆产能情况预测 9

图 6： 全球 HBM 供需缺口持续扩大（以 Bit 计） 9

图 7： 今年 5 月以来 B300/B200/H200/H100 租赁价格均明显上涨 11

下载日志已记录，仅供内部参考

图 8： 全球 AI 芯片供需缺口将延续.....12
图 9： 算力租赁位于算力产业链中游环节.....13
图 10： 算力租赁行业参与者可大致分为三个梯队.....14
图 11： 算力租赁厂商的四大核心竞争维度.....15
图 12： 主要厂商算力租赁业务相关的固定资产增加额对比（亿元）.....17
图 13： 主要厂商算租相关业务收入及增速对比（亿元）.....18
图 14： 主要厂商算租相关业务毛利率对比.....18
图 15： Nebius AI Token Factory 架构.....19
图 16： 中软国际正式与月之暗面展开 Token 分成相关合作.....21
图 17： 在与 Anthropic 的 TaaS 合作模式中，亚马逊的收入由基础设施收入和 Token 分成收入构成.....22
图 18： AWS Bedrock 利润率显著提升.....22
图 19： 2023-2026H1 协创数据收入及增速.....25
图 20： 2023-2026H1 协创数据归母净利润及增速.....25
图 21： 协创数据已形成“算-连-存”的全产业链布局.....27
图 22： 2023-2026H1 宏景科技收入及增速.....28
图 23： 2023-2026H1 宏景科技归母净利润及增速.....28
图 24： 2023-2026H1 利通电子收入及增速.....29
图 25： 2023-2026H1 利通电子归母净利润及增速.....29
图 26： 世纪利通具有英伟达 Preferred 级 AI 云伙伴资质.....29

表 1： 美国高端芯片出口管制政策持续加码.....10
表 2： 主要算租厂商设备采购及在手订单情况.....16
表 3： Token 分成模式下算租厂商的收益弹性更大且客户绑定更深.....20

下载日志已记录，仅供内部参考

表 4： 敏感性测试：预计在中性假设下到 2028 年算租厂商的 Token 分成收入可达 592 亿元..... 24

表 5： 协创数据已累计公告采购 331 亿元服务器设备..... 26

表 6： 主要公司盈利预测和估值情况..... 30

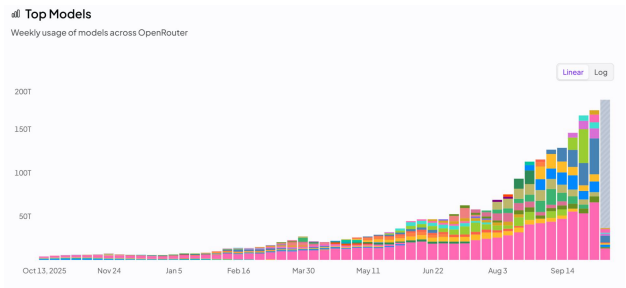
下载日志已记录，仅供内部参考

1. 供需缺口巨大，算租价格持续上行

1.1 企业级 AI Agent 加速落地，Token 用量需求爆发式增长

AI 工作负载已从“训练主导”转向“推理主导”，今年以来 AI 推理需求呈爆发式增长，并且由于 AI 推理需求具备持续、高频、碎片化等特征，相比于自建算力，客户将更倾向于采用租赁模式。随着智能助手、自动化工作流等逐步嵌入企业日常运营，AI 调用已不再是单次问答，而是覆盖规划、检索、工具调用、代码执行、验证反馈等完整链路，26 年以来 AI 推理调用量持续快速增长，根据 OpenRouter 数据，全球顶尖模型的合计周度 Token 使用量已从 26 年初的 6.42 万亿提升至今年 10 月初的 172 万亿，而根据 IDC 数据，2025 年推理算力支出首次接近训练算力支出，并且未来推理将成为 AI 算力市场最主要的新增需求来源。而推理需求天然更适合租赁模式，具体而言，与 AI 训练脉冲式采购不同，推理需求具有持续在线的特征，推理算力一旦上线，就会变成长期运营成本，并且，推理算力需求将伴随用户调用频次及复杂度的增长而增长，其持续性和强度均超过训练；同时，随着 AI 从模型训练走向应用推理，需求将从头部大厂扩散到更多企业，其中存在大量场景不适合自建算力，如高峰时段弹性扩容、短期测试和迁移等。

图 1：全球顶尖模型的合计 Token 使用量持续高增



资料来源：OpenRouter，山西证券研究所

图 2：AI 工作负载从“训练主导”转向“推理主导”



资料来源：IDC，山西证券研究所

随着企业级 AI Agent 加速落地，单客户算力消耗的乘数效应将系统性抬升算力需求曲线斜率。目前中国企业活跃智能体规模正步入加速期，根据 IDC 预测，随着本土模型能力持续升级、Agent 技术与应用生态快速成熟以及产业政策叠加共振，中国企业活跃智能体数量将在 2031 年突破 3.5 亿个，年复合增长率超过 135%；同时，AI Agent 的任务执行量将迎来爆发式增长，年执行任务数将从 2026 年 178 亿次提升至 2031 年 190 万亿次左右，意味着 AI Agent 正持续深度嵌入企业工作流中。由于单个 Agent 任务涉及多步推理、

下载日志已记录，仅供内部参考

工具调用与自我纠错，单次 Token 调用量远高于传统对话模式，并且随着 Agent 处理的任务越来越复杂，所需推理深度与调用链路不断加长，底层 Token 消耗也将以更快的速度增长，根据 IDC 预测，2026-2031 年国内企业智能体年度 Token 消耗量有望实现年均超 30 倍的指数级跃升。

图 3：2026-2031 年中国企业智能体市场将迎来陡峭的增长曲线

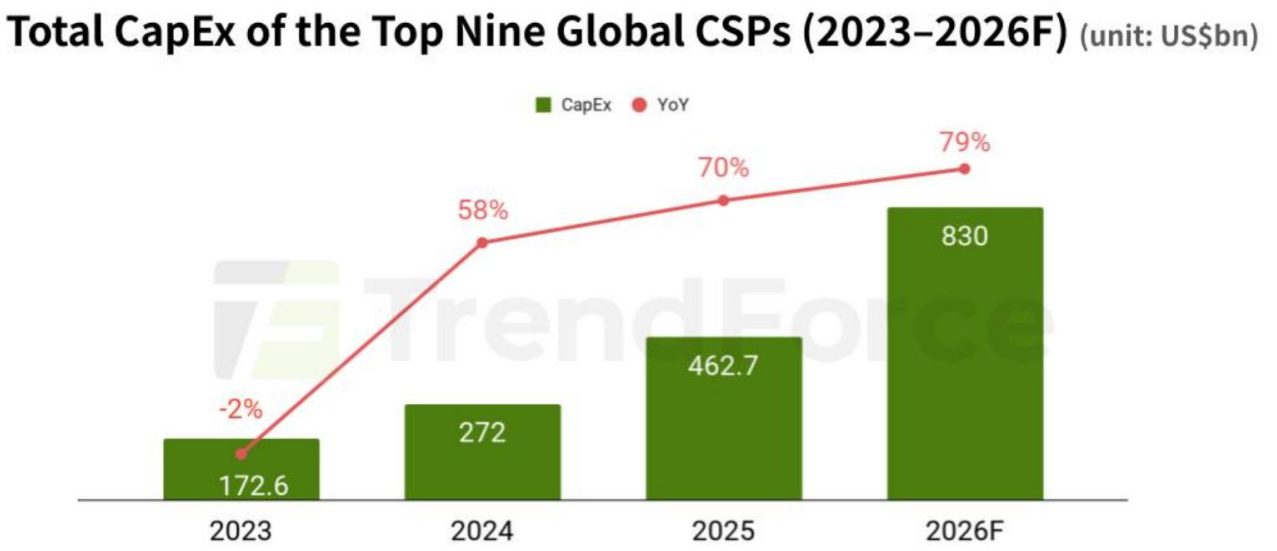


资料来源：IDC，山西证券研究所

在持续高景气的 AI 需求刺激下，国内外头部云厂商仍在加速增加 AI 投入，将推动整体算力市场的空间和增长确定性明显提升。根据 Trendforce 预测，全球前九大 CSP（谷歌、亚马逊、Meta、微软、Oracle、字节、腾讯、阿里和百度）2026 年的资本开支总额将达到 8300 亿美元左右，同比增长 79%，较 25 年进一步加速，其中，北美头部云厂商已将 AI 基础设施投资作为长期战略重点，微软、谷歌、Meta 均上调了今年的资本开支指引；而在国内云服务商中，阿里和字节的 AI 扩张最为激进，阿里专注于通过阿里云提供本地化节点和自主云服务，以渗透新兴市场，自宣布 2025 年扩张计划以来，公司已在巴西和法国建立了新的区域，使其全球业务版图扩展至 31 个区域和 106 个可用区，字节通过 TikTok 积极拓展海外市场，并在欧洲、泰国和马来西亚进行了大量投资。目前大厂的算力投资正日益集中于高性能 GPU 集群的部署、自主 ASIC 芯片的研发及旨在支持高功率密度计算的下一代数据中心等，虽然大厂自建算力会分流部分算租市场需求，但一方面整体算力需求增长更快，另一方面大厂高强度进行算力投入也会产生溢出效应，推动生态伙伴采购、中小模型厂商跟投等，我们认为最终将为算力租赁市场带来天花板的显著抬升。

下载日志已记录，仅供内部参考

图 4：全球前九大 CSP 今年资本开支总额预计将达到 8300 亿美元左右，同比增长 79%



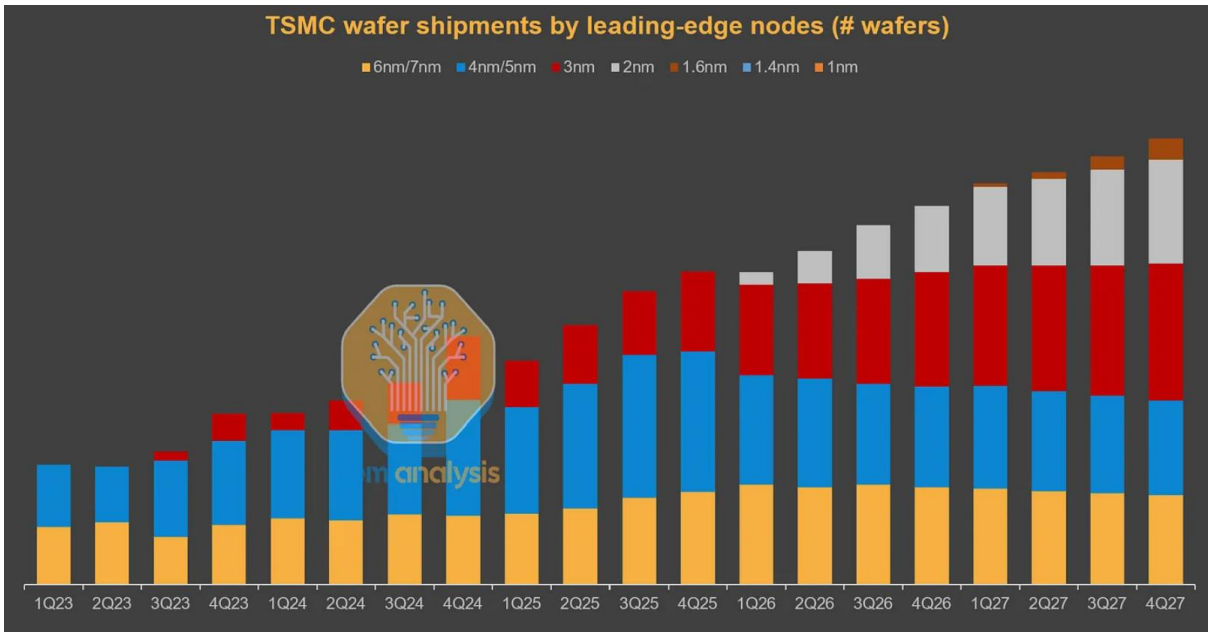
资料来源：Trendforce，山西证券研究所

1.2 芯片供给刚性约束叠加美国出口管制，国内高端卡供给紧缺

英伟达主力芯片 Blackwell 系列面临多重产能瓶颈叠加，交付周期显著拉长，对全球高端算力供给形成刚性约束。其中，1）在晶圆制造环节，Blackwell 芯片集成 2080 亿个晶体管，较前代 Hopper 系列提升 2.6 倍，导致前道制造复杂度呈指数级上升，Blackwell 芯片需要依靠台积电 4NP 高端制造工艺。根据黄仁勋披露，2025 年四季度英伟达给台积电追加的晶圆订单直接把台积电 4NP 工厂产能拉到了满负荷，而晶圆厂扩产周期需要 1.5-2 年，在需求持续高增背景下存在供需缺口；2）先进封装仍是制约 AI 芯片快速上量的重要瓶颈，据台积电 CEO 称，目前先进制程需求与可供应产能之间的缺口“真的很大”，为此台积电也将 2026 年资本开支上调至 600-640 亿美元。并且为在单片晶圆上放下更多 GPU Die，英伟达自 Blackwell 芯片开始便采用更先进的 CoWoS-L 封装技术，供给约束甚至更强；3）在内存方面，随着单个加速器中 HBM 容量快速增长，HBM 需求（以 Bit 计）持续攀升，导致 HBM 的供需缺口不断扩大。

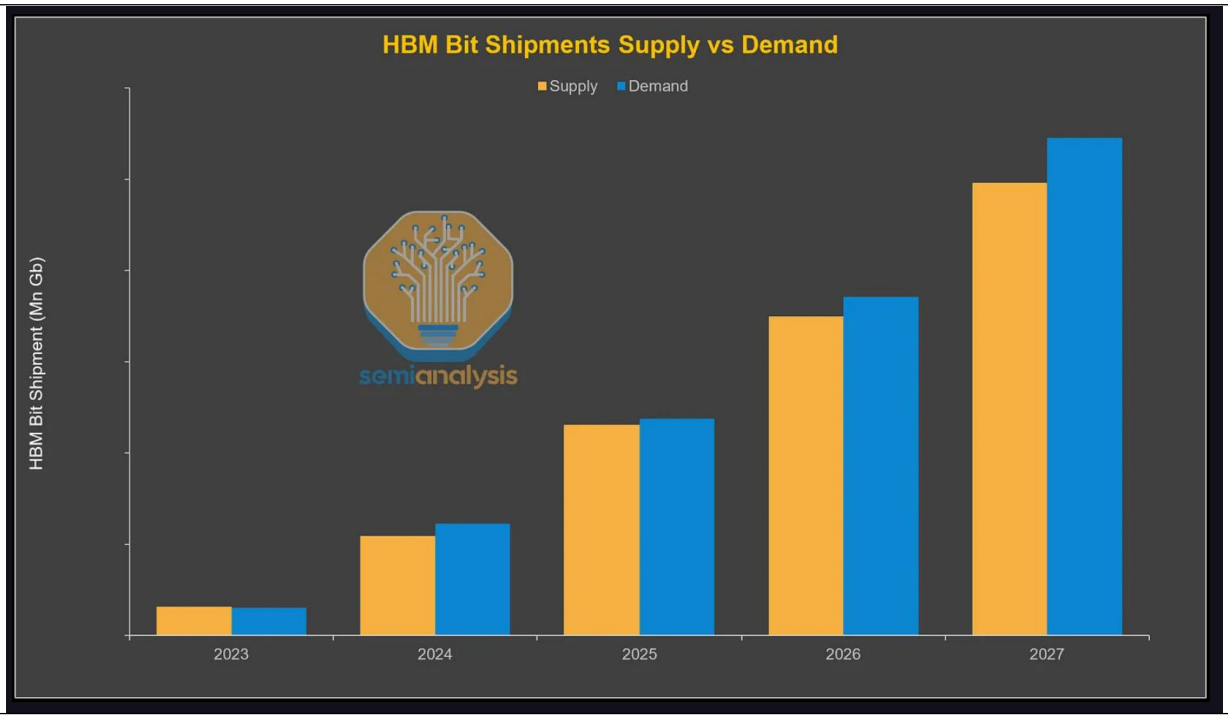
下载日志已记录，仅供内部参考

图 5：23Q1-27Q4 台积电不同制程晶圆产能情况预测



资料来源：Semianalysis，山西证券研究所

图 6：全球 HBM 供需缺口持续扩大（以 Bit 计）



资料来源：Semianalysis，山西证券研究所

下载日志已记录，仅供内部参考

近年来美国对高端芯片的出口管制持续收紧，叠加供给刚性约束，国内可获取的 Blackwell 增量极为有限。虽然美国对次高端芯片（如 H20、MI308）出口到中国的管制有所放松，但对于高端芯片（如 Blackwell 和 Rubin 系列）的出口管制仍在不断加码，今年 5 月底美国商务部为弥补出口管制审查中的漏洞，周日便发布通知称，对先进芯片的出口管制要求适用于所有总部和母公司位于中国的企业，防止高性能芯片绕过监管供应中国公司。考虑到中美科技博弈持续进行，国内 Blackwell 芯片的获取难度更大，存在巨大的供给缺口。目前，B200 现货已近枯竭，而 B300 交付排期也有显著延长。

表 1：美国高端芯片出口管制政策持续加码

政策	颁布时间	颁布机构	管制内容	涉及 AI 芯片
出口管理条例（更新）	2022.10.7	美国商务部	禁止：TPP≥4800 且 I/O 带宽≥600GB/s	A100、H100
出口管理条例（更新）	2023.10.17	美国商务部	禁止：1）TPP≥4800 或 2）TPP≥1600 且 PD≥5.92 受限：1）2400≤TPP＜4800 且 1.6≤PD＜5.92 或 2）TPP≥1600 且 3.2≤PD＜5.92	A800、H800、RTX4090、MI250、MI250X、MI300X
出口管理条例（更新）	2024.3.29	美国商务部	AI 芯片出口限制同样适用于内载此类芯片的所有电子设备	-
人工智能扩散框架的临时性最终规则	2025.1.15	美国商务部	建立三级国家框架和计算配额，第三层级（中国等）完全禁止进口（包括从其他层级）高端 AI 芯片	-
美国政府通知	2025.4.9	美国政府	H20、MI308 出口中国需许可证	H20、MI308
出口管理条例（更新）	2025.5.13	美国商务部	废除 AI 扩散条例，推出三项新指导意见：1）全球禁用华为 Ascend 芯片；2）限制 AI 芯片用于中国 AI 模型；3）供应链反制指南	-
美国政府通知	2025.7	美国政府	批准 H20、MI308 出口，销往中国的芯片征 15%关税	H20、MI308
出口管理条例（更新）	2026.1.13	美国商务部	将 H200 及同等性能芯片对华出口的许可审查政策从“推定拒绝”调整为“逐案审查”	H200、MI325X
关于芯片限制的指导意见	2026.5.31	美国商务部	对出口先进人工智能芯片的许可要求适用于所有总部或母公司位于中国的企业	-

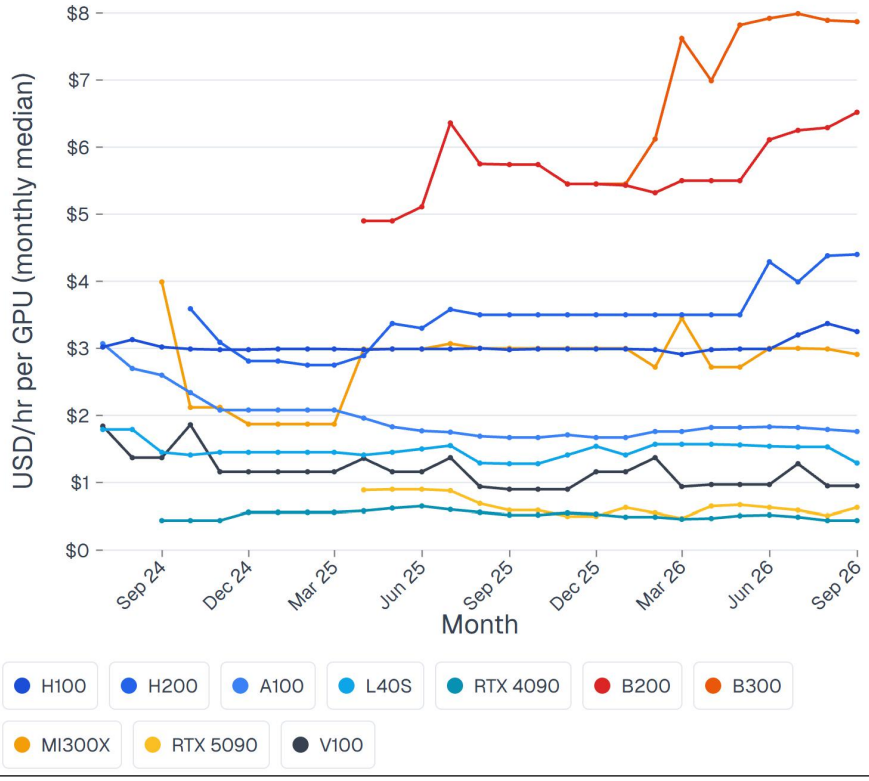
下载日志已记录，仅供内部参考

资料来源：美国驻华大使馆官网，《产业经济评论》2024 年 02 期《美对华半导体管制的趋势、实施要点与中国因应》，深圳市司法局官网，上海公平贸易服务网，美国商务部，Semianalysis，Reuters，复旦发展研究院《美国观察丨美国 AI 芯片出口管制政策的博弈：维护国家安全和维持技术栈主导的拉锯》，Institute for AI Policy and Strategy，电子工程专辑，山西证券研究所

1.3 高端卡租赁价格加速上行，下半年将维持高位运行

供需矛盾推动今年上半年高端 GPU 租赁价格加速上行，海内外形成共振，而国内由于供给更紧张弹性更大。今年 5 月，Nebius 发布全线调价公告，预计 6 月 1 日生效，所有 GPU 型号涨幅均超过 25%，部分型号创历史新高，其中，按需容量下 H100 从 2.95 美元/小时涨至 3.85 美元/小时，涨幅 31%，H200 从 3.50 美元/小时涨至 4.50 美元/小时，涨幅 29%，B200 从 5.50 美元/小时涨至 7.15 美元/小时，涨幅 30%，B300 涨幅也有 29%；而在抢占式容量下 H100、H200、B200 和 B300 的涨幅分别为 72%、69%、36%、26%；CoreWeave 也在 7 月对主流型号 GPU 租赁价格进行上调，上调幅度约 25%。根据 AIMultiple 统计，今年 5 月开始 B200、B300、H200、H100 的全球综合租赁价格均出现明显上涨，B300 租赁单价已接近 8 美元/小时。而在国内，今年上半年阿里云等云厂商也已上调 AI 云服务价格。

图 7：今年 5 月以来 B300/B200/H200/H100 租赁价格均明显上涨

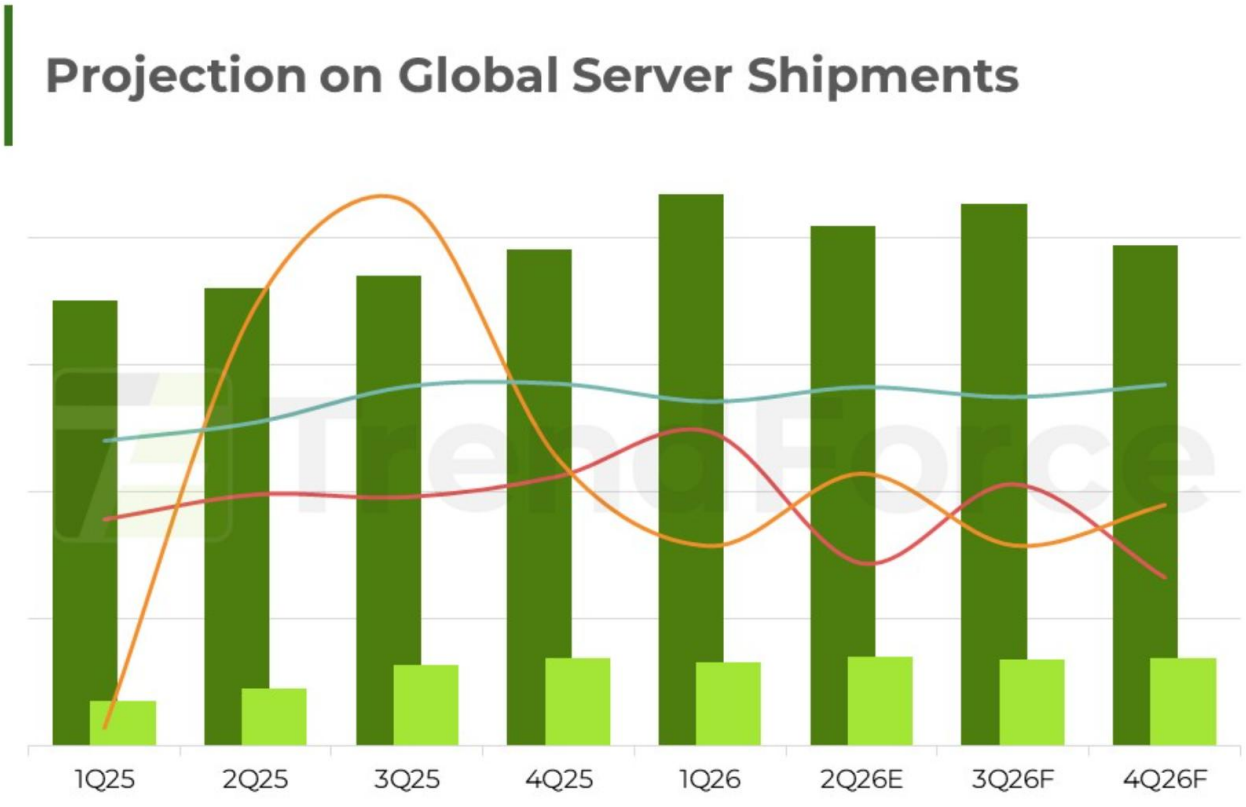


下载日志已记录，仅供内部参考

资料来源：AIMultiple，山西证券研究所

下半年算力供不应求状况预计无法有效缓解，算租价格将维持高位。展望 2026 下半年：1）在需求侧，大厂资本开支下半年集中释放，同时 AI Agent 快速渗透推动 Token 用量持续爆发，根据 IDC 数据，2025 全年中国企业级 MaaS 市场 Token 消耗量为 1944 万亿，同比增长约 16 倍，预计 2026 年 Token 消耗量将达到约 40000 万亿次，较 2025 年进一步增长约 20 倍；2）在供给端，一方面，英伟达高端 GPU 供给约束未见放松，国内可获得的增量芯片仍将十分有限，另一方面，以昇腾为代表的国产算力芯片上量仍需时间。因此，预计下半年供需缺口将延续甚至进一步扩大，H200、B300 等高端卡的租赁价格将维持高位，同时 A100、A800 等旧版本芯片有望受益于客户降配替代而涨价。

图 8：全球 AI 芯片供需缺口将延续



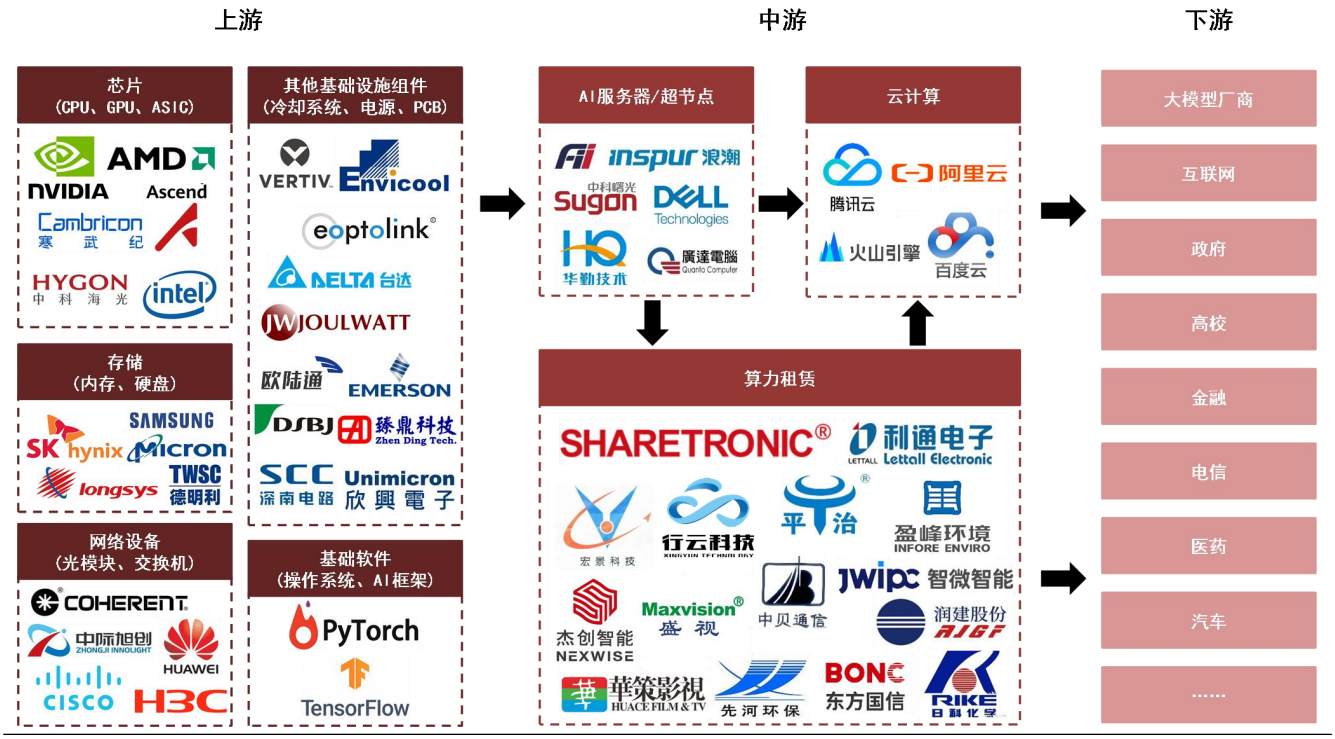
资料来源：Trendforce，山西证券研究所

下载日志已记录，仅供内部参考

2. 竞争格局和商业模式双重变革

算力租赁位于算力产业链中游环节，是连接算力供给和算力需求的重要枢纽。算力产业链上游主要提供基础软硬件，包括芯片、网络设备、电源、冷却设备、IDC 基础设施、基础软件等，中游为算力集成及服务环节，包括服务器、超节点、IDC、算力租赁和云计算等，下游对接互联网厂商、大模型厂商及政企单位，为其模型训练和应用推理场景提供算力。中游的算力租赁行业正凭借其低成本、灵活、高效等优势，打破算力资源配置壁垒，成为 AI 产业爆发的关键支撑，并且受益于目前算力供需严重错配，算力租赁行业的定价权正显著增强，带来利润分配的结构性上移。

图 9：算力租赁位于算力产业链中游环节



资料来源：华经产业研究院，山西证券研究所绘制

2.1 竞争格局：渠道封堵下市场集中度有望提升

目前算力租赁行业参与者多为跨界厂商，可大致分为三个梯队。其中，1）第一梯队：已完成大规模算力资源池搭建并锁定长期订单，拥有优质、稳定的客户资源，并且卡源供应稳定、资金实力雄厚、合规资质齐全，产能利用率可长期维持在 90%以上，包括协创数据、宏景科技、行云科技等；2）第二梯队：在区域客户资源、本土机房资源等方面具有差异化优势，但在手算力规模相对较小，同时拿货能力和集群交付

下载日志已记录，仅供内部参考

能力也弱于第一梯队厂商，包括华策影视、先河环保、锦鸡股份、日科化学等；3）第三梯队：多为零散中小服务商，无稳定卡源、无核心合规资质，仅能提供零散的算力零售服务，资金实力、客户资源、在手算力均处于劣势。

图 10：算力租赁行业参与者可大致分为三个梯队



资料来源：山西证券研究所绘制

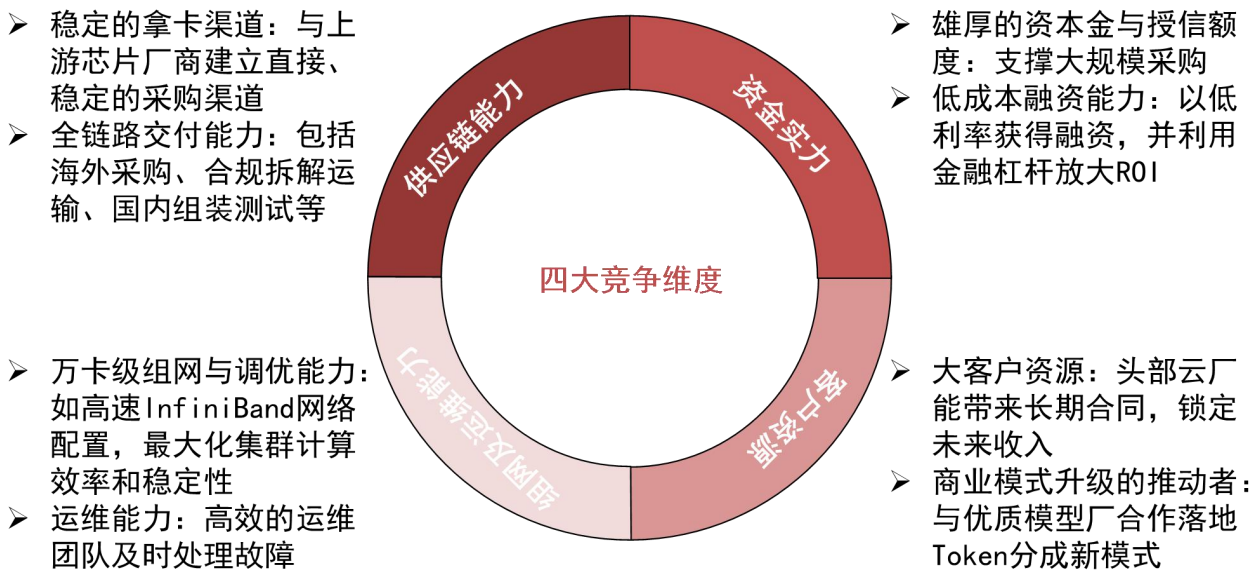
渠道封堵叠加出租率困境下中小算租厂商大概率将逐步出清，市场份额有望向头部集中。一方面，今年 5 月底美国 BIS 法案更新后，非正规渠道的配额高度向头部公司集中，同时，B300 八卡服务器现货价格自年初每台 400 万级别到现在已有数倍的涨幅，小体量买家的采购成本溢价持续扩大，在已锁定低成本存量资产且资金实力雄厚的头部公司面前缺乏竞争力；另一方面，在卖方市场格局下，头部公司的大客户已通过长协形成锁定，形成“大订单→大采购→更强渠道”的正反馈，中小厂商即便以更低租金也难以撬动这一客户群。而若高端算力集群在建设后不能快速达到 70-80%以上出租率，折旧和利息的固定成本将使其陷入持续亏损。此外，相较于头部公司拥有广泛、低价的融资渠道，部分中小供应商为缺乏实际业务背书的跨界者，如需完全通过外部融资获取采购资金，将面临银行授信风险溢价偏高、担保条款苛刻等困境，导致与头部公司之间存在巨大的融资成本差异，进而对其盈亏平衡点产生实质性影响。

算力租赁厂商的核心竞争维度包括供应链能力、资金实力、客户资源和组网及运维能力，在这些维度

下载日志已记录，仅供内部参考

具备明显优势的厂商将有更大机会脱颖而出。具体而言，1）供应链能力：这是当前行业进入门槛最高的维度，高端 GPU 全球供应紧张叠加出口管制趋严的背景下，与上游芯片厂商建立稳定、直接、低成本的购货渠道是重中之重，而具备 NCP 等资质的企业，能以更低成本、更稳定地获取货源，形成显著优势；2）资金实力：算力租赁是典型的重资产模式，前期需要投入大量资金采购 AI 服务器、网络设备等硬件设备，因此，资金实力及融资能力决定了企业的规模天花板和抗风险能力；3）客户资源：拥有优质、稳定且大体量的客户将支撑未来订单的稳定性，并带来稳定的现金流，是企业持续扩张的基础；4）组网及运维能力：在硬件采购落地基础上，需要将其高效的组织、调度起来。而构建和运维千卡/万卡级大规模 GPU 集群具有较高的技术门槛，企业需要具备优秀的组网能力和系统调优能力，同时运维团队能及时处理服务器等故障。

图 11：算力租赁厂商的四大核心竞争维度



资料来源：中研网，山西证券研究所绘制

目前协创数据在算力储备、订单规模、拿卡渠道、资金实力等方面保持领先地位，而以行云科技、东阳光、杰创智能、盛视科技等为代表的厂商算力扩张迅速。具体而言，自 2024 年 10 月公告首个 AI 服务器采购以来，截至目前，协创数据已公告合计不超过 331 亿元的服务器采购，并且随着采购的服务器持续到货并交付客户，2025 年公司机器设备项目的固定资产增加 90.79 亿元，2026 上半年进一步增加 138.74 亿元，推动公司固定资产总额达到 220.24 亿元。与此同时，协创数据申请的银行授信额度已达 1415 亿元，并辅以融资租赁、抵押借贷、定增等融资方式，公司强大的融资能力将支撑算力业务进一步扩张。而在其他厂商中，行云科技、东阳光、杰创智能、盛视科技等已公告大体量的服务器采购或算力服务订单，一定程度上

下载日志已记录，仅供内部参考

也彰显出其在供货渠道和客户资源上具备较强竞争力。

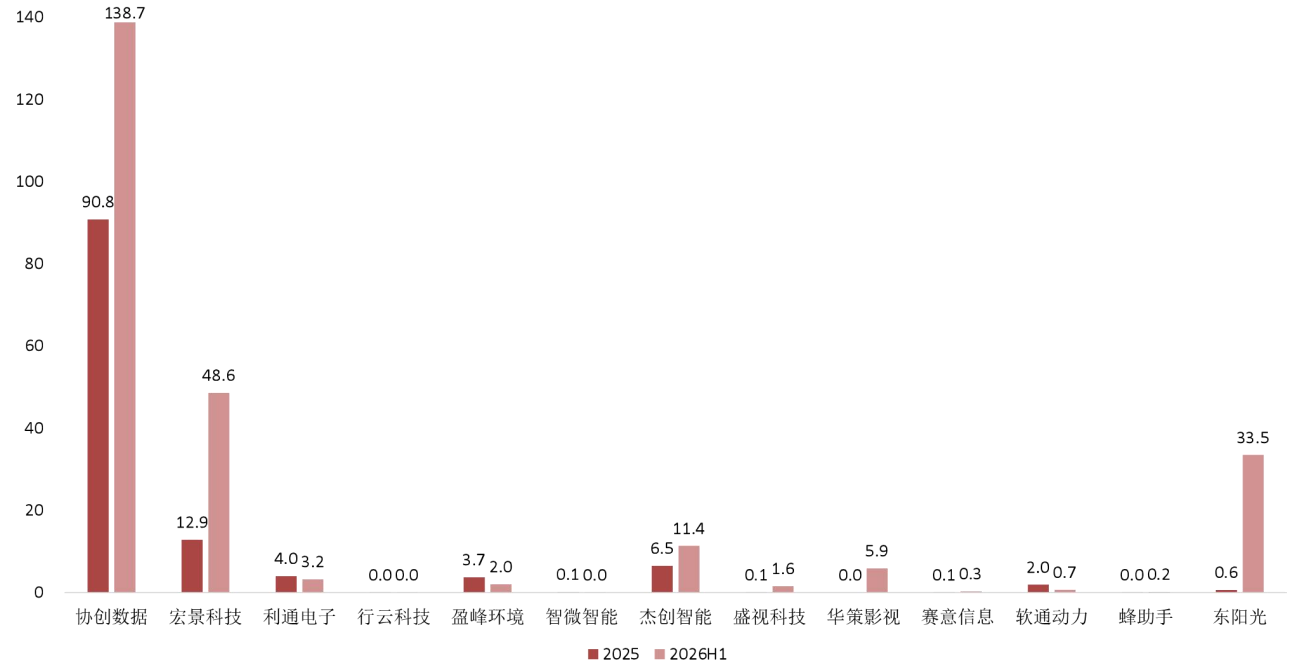
表 2：主要算租厂商设备采购及在手订单情况

公司	在手算力/在手订单/设备采购情况
协创数据	已公告不超过 331 亿元服务器采购
宏景科技	已公告 34.96 亿元算力服务订单
利通电子	截至 26 年 6 月底，在手 3.8 万 P AI 算力，其中直租算力 1.6 万 P，转租算力 2.2 万 P
行云科技	在手 171.82 亿元算力长期框架订单
盈峰环境	-
智微智能	已公告不超过 40 亿元服务器采购
杰创智能	已公告不超过 140 亿元设备采购，已建成约 9000P 算力，形成集群设备资产超 25 亿元，在建算力超 4 万 P
盛视科技	已公告不超过 71 亿元 IT 设备及零配件采购
华策影视	已公告不超过 33 亿元服务器采购，对应需交付算力 2.9 万 P，项目完成后在手算力 3.7 万 P
赛意信息	已公告 67.2 亿元高性能算力服务销售合同
软通动力	“北京壹号词元工厂”在京点亮，并与头部大模型厂商签署智算服务协议
蜂助手	已公告 70.22 亿元服务器采购及 103.25 亿元算力服务订单
东阳光	已公告 390-460 亿元算力服务订单；截至 26 年 6 月底，已落地 41.14 亿元 IT 设备
中软国际	与月之暗面（Kimi）正式签署“登月计划”Token 分成及联合创新合作协议

资料来源：协创数据关于公司算力业务进展情况的自愿性披露公告、关于公司购买资产的公告，宏景科技 2025 年报、关于签署算力业务合同的公告，利通电子 2026 半年报，行云科技投资者纪要 20261013、关于全资子公司深圳行云签订重大合同的进展公告，智微智能关于公司拟购买资产的公告，杰创智能 2026 半年报，盛视科技关于购买资产的公告，华策影视投资者纪要 20260512，赛意信息关于签订高性能算力服务销售合同的进展公告，软通动力 2026 半年报，蜂助手 2026 半年报、关于签署算力业务日常经营重大合同的公告，东阳光 2026 半年报，中软国际 2026 半年报，山西证券研究所

下载日志已记录，仅供内部参考

图 12：主要厂商算力租赁业务相关的固定资产增加额对比（亿元）

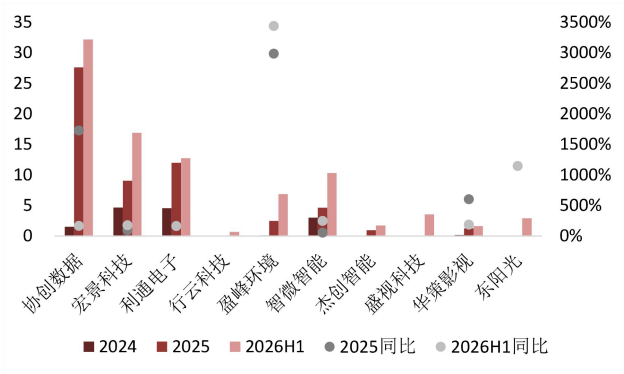


资料来源：各公司 2025 年报和 2026 半年报，山西证券研究所；注：考虑到各家公司的不同情况，协创数据、华策影视统计机器设备项目口径，宏景科技、杰创智能统计算力设备项目口径，盈峰环境统计专用设备项目口径，其他公司统计电子设备相关项目口径

目前算租厂商业绩已开始有所分化。其中，协创数据作为行业龙头，2025 年算租相关收入已达 27.61 亿元，同比高增 1727%，2026 上半年进一步增长至 32.16 亿元，同比增长 163%，收入的绝对体量大幅领先于其他厂商；宏景科技、利通电子、智微智能 2026 上半年算租相关收入均超过 10 亿元，同比增速均超过 150%；盈峰环境和东阳光 2026 上半年算租相关收入的同比增速均在 1000%以上，并且根据两家公司的在手订单情况，预计今年算租相关业务有望继续放量；行云科技、杰创智能、盛视科技等今年开始布局算力租赁，目前进展迅速，上半年均已贡献超过 5000 万收入。同时，考虑到 GPU 折旧是刚性固定成本，高出租率情况下增量营收能够高效转化为利润增量，而协创数据、宏景科技等头部算租厂商出租率维持高位，盈利能力已有明显改善，26H1 协创数据净利润同比增长 326%，远超 153%的营收增速，宏景科技净利率也从 24 年-11.58%提升至 26H1 16.54%。

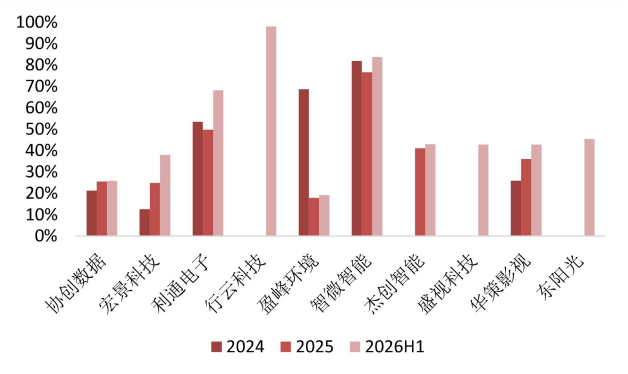
下载日志已记录，仅供内部参考

图 13：主要厂商算租相关业务收入及增速对比（亿元）



资料来源：各公司 2025 年报和 2026 半年报，山西证券研究所

图 14：主要厂商算租相关业务毛利率对比



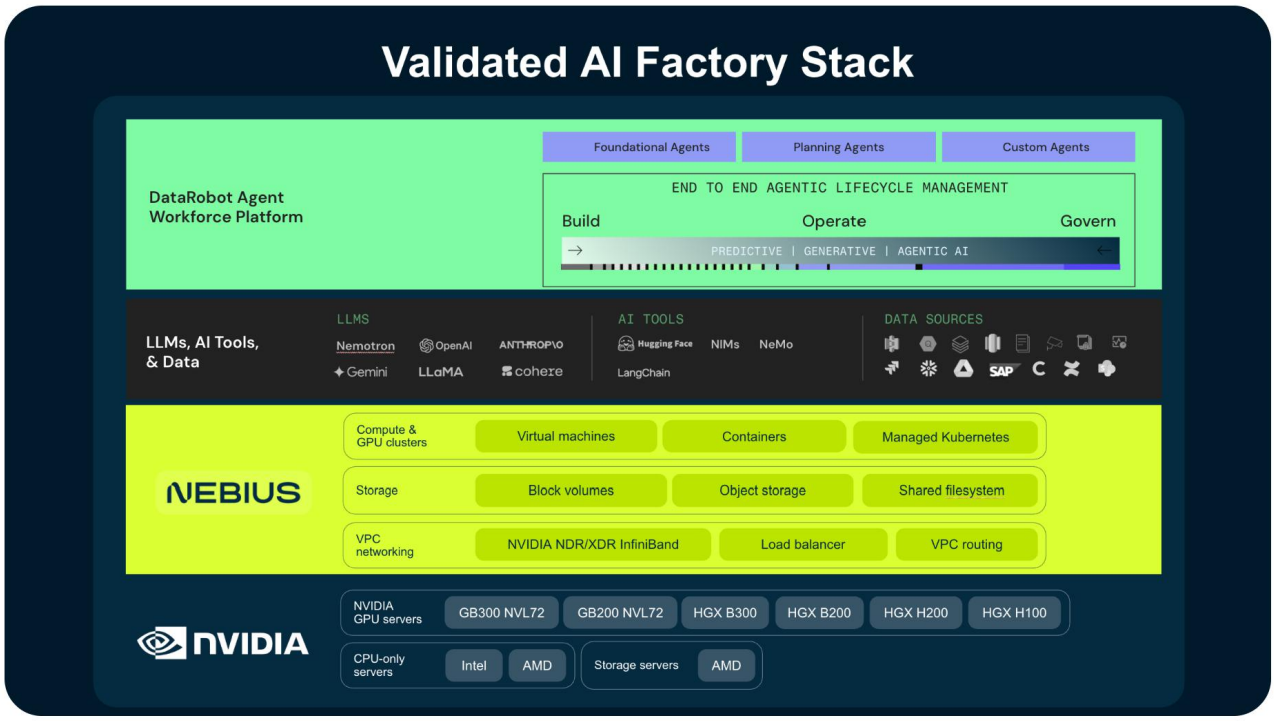
资料来源：各公司 2025 年报和 2026 半年报，山西证券研究所

2.2 商业模式从服务器租赁向 Token 工厂进化，行业估值逻辑重构

通过复盘，算力租赁行业从传统裸金属租赁演进为卖方市场下的算力集群出租，目前正经历从算力资源出租向 Token 工厂模式的升级。具体而言，1）2023 年及之前，算力租赁行业商业逻辑非常简单，即买卡、采购机柜、按 GPU 时长出租，收入上限由硬件存量决定，利润弹性低。行业竞争力的核心是渠道关系（能否拿到货）和融资能力（能否买得起），没有太多技术含量；2）2024-2025 年，随着 AI 推理需求爆发，Token 调用量开始呈指数级增长，需求端爆发叠加 GPU 供给刚性约束，算力租赁厂商从价格接受者变成了价格制定者。同时，算租行业开始从裸金属出租向完整集群交付演变，对算租厂商的大规模组网和运维能力产生要求；3）2026 年起，英伟达黄仁勋在 GTC 大会上正式提出“AI Factory”概念，将未来数据中心定义为“以 token throughput 衡量产出的专用计算基础设施”，完成了行业叙事框架的顶层定义，而海外算力租赁商 Nebius 也率先推出 Token Factory 推理平台，成为算租厂从出租资源向卖 Token 产出转型的标志性事件。

下载日志已记录，仅供内部参考

图 15：Nebius AI Token Factory 架构



资料来源：Nebius 官网，山西证券研究所

对于算力提供方而言，Token 工厂的收费模式可大致分为收入分成、按量计费和订阅制三种，而算租厂商预计将主要采用前两种。其中，1）收入分成：算租厂商通过与以大模型厂商为代表的客户签订分成协议，按照实际生成的 Token 收入进行分成，也成为算租厂商绑定客户的重要手段；2）按量计费：作为目前 Token 工厂的主流收费模式，下游客户（开发者、企业用户等）通过 API 调用部署在 Token 工厂的各类模型，系统实时计量所消耗的 Token 数量，并按约定的单价进行结算。收入与真实需求强挂钩，避免了传统算力租赁中的空置损耗，同时为客户提供了极大的使用弹性；3）订阅制：为了满足不同客户的需求，Token 工厂也推出了类似“电话套餐”的订阅制模式，客户按月或按年支付固定费用，获得一定额度的 Token 使用权。订阅制能够降低客户的心理门槛，并为算力供给方提供稳定的现金流，尤其适合 C 端用户和中小企业，目前三大运营商均已推出相应的 Token 套餐。我们认为，算租厂商一方面对于 C 端及中小 B 客户的销售渠道覆盖能力相对较弱，另一方面订阅制的收益弹性有限，因此，算租厂商大概率将重点布局收入分成和按量计费两类模式。

Token 分成模式下算租厂商的产业链嵌入更深且价值分配比例更高，有望获得更多青睐。在传统的 GPU 服务器裸金属和集群出租模式中，用户可直接访问 GPU 芯片，避免虚拟化性能损耗，计价方式为按 GPU

下载日志已记录，仅供内部参考

卡时/按月固定收租金，算租厂商处于产业链资源出租环节，议价空间上限明确。而在 Token 工厂模式下，算租厂商将算力、电力等资源转化为 Token 产能并提供给模型厂商等客户，其收入来源从固定租金转为按 Token 调用量收费（包括分润），直接受益于模型推理需求爆发，盈利弹性大幅拓展。并且，相比于按量计费，采用 Token 分成收费模式的算力租赁商将从产业链的成本端向收入端跃升，从被动卖资源升级为主动参与价值分配，与客户形成深度绑定。

表 3：Token 分成模式下算租厂商的收益弹性更大且客户绑定更深

	算力资源租赁	Token 工厂（按量计费）	Token 工厂（Token 分成）
收入模式	按 GPU*小时/月/年收固定租金	按 Token 使用量收费	固定租金+Token 收入分成或纯 Token 收入分成
计价单位	GPU*小时/月/年	百万 Token 产出	GPU*小时/月/年，百万 Token 产出
算租厂商角色	算力资源供应商	Token 生产服务商	Token 生产服务商
收入弹性	低	高	高
毛利率	低	高	高
产业链位置	资源层	运营层	运营+产出层

资料来源：韶关算力产业协会公众号，西部智算公众号，Semianalysis，山西证券研究所整理

今年以来，Token 分成开始成为越来越多算租厂商的新布局方向。一方面，在卖方市场中，大模型厂商等下游客户为抢购算力愿意出让一部分 Token 收益，另一方面，算租厂商也有意愿突破传统租赁模式的利润率天花板，于是今年以来算租厂商开始持续探索 Token 分成新模式。截至目前，中软国际已率先与月之暗面（Kimi）正式签署“登月计划”Token 分成及联合创新合作协议，即以 AllMeta 平台叠加 Kimi 大模型为技术底座，承接 Kimi 标品之外的全部定制开发、行业实施与私有化部署，并且更重要的是，中软国际将按 Token 调用量参与收益分成，即对于在合作过程中 Kimi 大模型产生的 Token 消耗收益和其他衍生收入，双方将按照协议约定比例进行分成，中软国际得以分享模型被持续使用所产生的长期价值。

下载日志已记录，仅供内部参考

图 16：中软国际正式与月之暗面展开 Token 分成相关合作

中軟國際與月之暗面簽署Token分成及聯合創新合作協議
共建FDE創新實驗室
佈局企業級Agentic AI業務

中軟國際有限公司（以下簡稱「**中軟國際**」或「**本公司**」）自願發佈本公告，向公眾披露本公司最新經營合作進展。

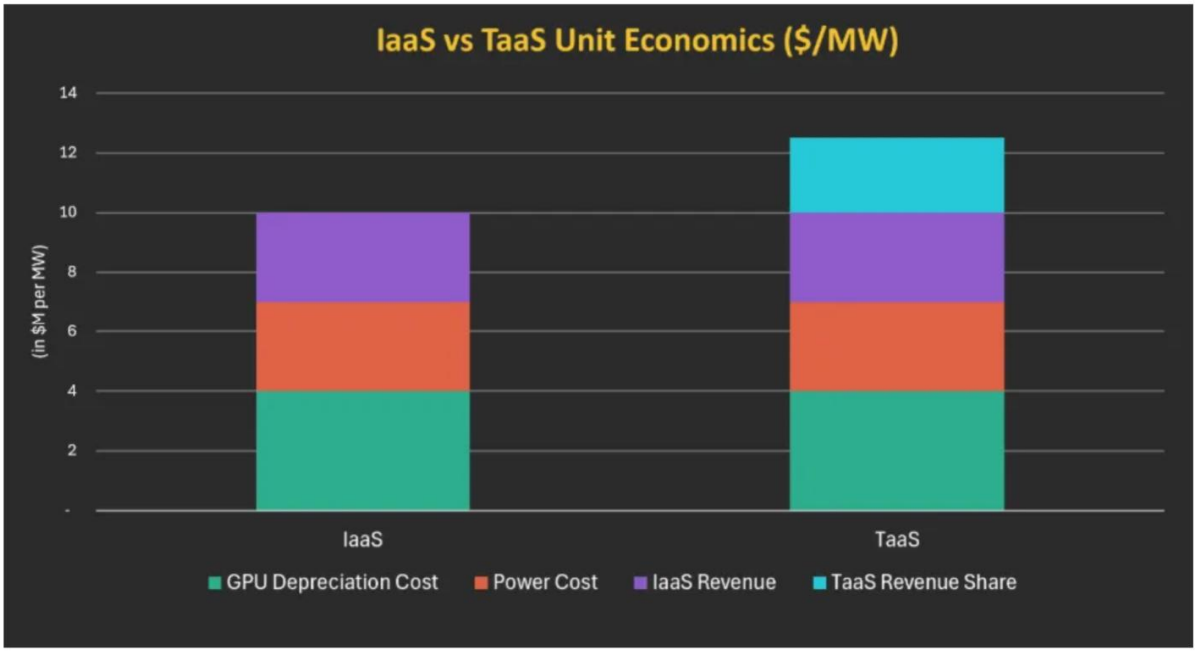
本公司欣然宣佈，近日，公司已與北京月之暗面科技有限公司（以下簡稱「**月之暗面**」）正式簽署「登月計劃」之Token分成及聯合創新合作協議，是公司面向高專業價值Token運營商轉型的關鍵戰略舉措。雙方將基於各自核心優勢，圍繞企業服務領域Agentic AI開展全方位深度合作，共同打通大模型商業化落地最後一公里。

资料来源：中软国际自愿性公告，山西证券研究所

参考海外亚马逊与 Anthropic 的合作实践，预计典型的 Token 分成合同会采用“保底租金+超额分成”的双层设计或者纯分成模式。亚马逊与 Anthropic 的合作模式正从“算力小时”进入“Token 账单”阶段，即亚马逊与 Anthropic 按 Token as a Service 模式分成。具体而言，AWS Bedrock 平台成为 Anthropic 模型的分发渠道，亚马逊向支付 AWS 账单的用户分发 Claude 代币，但实际销售方为 Anthropic，模型使用也受 Anthropic 约束。作为卖方，Anthropic 将销售代币的全部收入入账；而作为平台提供方，AWS 既能获得基础设施费用，又能获得 Token 销售收入的分成。对于 Anthropic 而言，无需签订昂贵的多年期 IaaS 合同即可获得算力资源，同时可接入云厂商的客户群体；而对于亚马逊而言，虽然相比于锁定收入的五年期 IaaS 合同风险更高，但利润率却有显著改善。目前亚马逊凭借其交易架构设计及执行能力已成为该领域的早期赢家，Bedrock 平台规模远超竞争对手。基于亚马逊与 Anthropic 的成功合作案例，我们预计国内算租厂商 Token 分成模式合同会倾向于采取类似的“保底租金+超额分成”的设计，以满足算租厂商和模型厂商的不同诉求，同时在模型厂商收入高增长持续性较强的情况下也不排除采用纯分成模式。

下载日志已记录，仅供内部参考

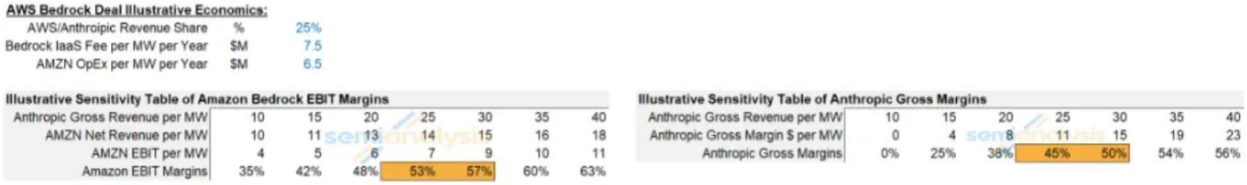
图 17：在与 Anthropic 的 TaaS 合作模式中，亚马逊的收入由基础设施收入和 Token 分成收入构成



资料来源：Semianalysis，山西证券研究所

Token 分成模式下 AWS 利润率显著提升，并且将随着 Token 用量的增长而继续向上。根据 Semianalysis 测算，当 AWS 分成比例为 25%、IaaS 费用为 750 万美元/MW/年、亚马逊运营成本为 650 万美元/MW/年时，若在 Bedrock 计算资源上每兆瓦 Anthropic ARR 为 1000 万美元，则亚马逊 Bedrock EBIT 利润率为 35%，若 Anthropic ARR 提升 500 万美元至 1500 万美元，则 Bedrock EBIT 利润率将提升至 42%，而传统 IaaS 服务的利润率仅为 13%（（7.5-6.5）/6.5）。根据 Semianalysis 估算，26Q1 Anthropic 在 Bedrock 每兆瓦计算能力上的 ARR 收入约 2600 万美元，对应 Bedrock EBIT 利润率可达 53-57%，成为 26Q1 AWS 利润增量的重要贡献来源；而 Semianalysis 预计 26Q2 Anthropic ARR 收入可达 4200 万美元左右，对应 Bedrock EBIT 利润率将超过 63%。

图 18：AWS Bedrock 利润率显著提升



AWS Bedrock Deal Illustrative Economics:

AWS/Anthroipic Revenue Share	%	25%
Bedrock IaaS Fee per MW per Year	$M	7.5
AMZN OpEx per MW per Year	$M	6.5

Illustrative Sensitivity Table of Amazon Bedrock EBIT Margins

Anthropic Gross Revenue per MW	10	15	20	25	30	35	40
AMZN Net Revenue per MW	10	11	13	14	15	16	18
AMZN EBIT per MW	4	5	6	7	9	10	11
Amazon EBIT Margins	35%	42%	48%	53%	57%	60%	63%

Illustrative Sensitivity Table of Anthropic Gross Margins

Anthropic Gross Revenue per MW	10	15	20	25	30	35	40
Anthropic Gross Margin $ per MW	0	4	8	11	15	19	23
Anthropic Gross Margins	0%	25%	38%	45%	50%	54%	56%

下载日志已记录，仅供内部参考

资料来源：Semianalysis，山西证券研究所

未来 Token 分成有望为国内算力租赁厂商贡献重要收入和利润增量，我们测算，到 2028 年，中性预测下算租厂商的 Token 分成收入将达到 592 亿元。

1）Token 总收入测算：根据 IDC 预测，按 Tokens 规模算，中国模型即服务（MaaS）市场规模将从 2025 年 1941 万亿 tokens 提升至 2028 年 3608000 万亿 tokens，年复合增长率高达 1130%；而按照收入口径算，2025 年中国 MaaS 市场规模为 30.7 亿元，2026 年提升至 186 亿元，计算可得 2025/2026 年每百万 Token 的价格为 1.6/0.5 元，26 年 Token 价格同比下降 71%。我们认为，随着模型推理效率持续优化以及底层算力硬件性价比日益凸显，Token 综合价格下降是大势所趋，假设 27-28 年 Token 综合价格将分别下降 30%和 20%，则 2028 年国内 Token 价格将降至 0.26 元/百万 tokens。因此，预计到 2028 年国内 Token 总收入可达 9395 亿元。

2）算租厂商 Token 分成收入测算：我们基于 Token 分成模式渗透率以及分成比例两大变量进行敏感性分析。一方面，由于 Token 分成模式并不一定适用于所有类型的客户，比如资金实力强的互联网大厂，目前看主要尝试的是大模型厂商，参考 IDC 发布的 2025H2 国内 GenAI IaaS 市场行业分布数据，基模厂商对 GPU 集群的采购占比为 21.5%，假设 Token 分成模式的渗透率天花板在 30%左右。同时，大模型厂商等下游客户仍将采用传统租赁模式获取部分算力，因此，在悲观情况下，我们预计 28 年大模型厂商等客户采用 Token 分成模式获取算力的比例仅为 10%，即对应 Token 分成的渗透率约 3%，而乐观情况下可达到 50%，对应 Token 分成的渗透率约 15%；另一方面，由于目前算力租赁市场仍呈现明显的卖方市场特征，且智能算力供应短缺问题短期内无法有效缓解，因此，我们认为近三年算租厂商将持续拥有较强的议价权。考虑到算力资源是模型厂等客户产生 Token 收入的基础，分成比例设计将更偏向于算力提供方，因此，卖方市场下乐观假设 28 年算租厂商的 Token 收入分成比例可达 85%（按纯分成模式算），而悲观假设其分成比例在 55%。最终测算得到 2028 年国内算租厂商 Token 分成收入在中性假设下为 592 亿元，乐观假设下为 1198 亿元，悲观假设下为 155 亿元。

下载日志已记录，仅供内部参考

表 4：敏感性测试：预计在中性假设下到 2028 年算租厂商的 Token 分成收入可达 592 亿元

算租厂商的 Token 分成收入		Token 分成模式渗透率						
		3%	5%	7%	9%	11%	13%	15%
算租厂商的 Token 分成比例	55%	155	258	362	465	568	672	775
	60%	169	282	395	507	620	733	846
	65%	183	305	427	550	672	794	916
	70%	197	329	460	592	723	855	986
	75%	211	352	493	634	775	916	1057
	80%	225	376	526	676	827	977	1127
	85%	240	399	559	719	878	1038	1198

资料来源：IDC，山西证券研究所测算

Token 分成模式下算租厂商除拿卡渠道和资金实力外，还需要优异的推理优化能力及与头部模型厂的深度绑定，因此随着 Token 分成模式的持续渗透，具备强推理优化技术并与智谱等头部模型厂绑定的算租厂商将充分受益。具体而言，1）Token 分成与传统租赁模式的核心区别在于定价从“GPU 卡时”切换为“Token 产出”，意味着算力利用率能否通过 KV Cache 极致复用、调度算法优化、模型服务封装等技术大幅提升，直接决定单位 GPU 的 Token 产量，进而决定分成收益；2）由于算租厂商承担了 100%的硬件投入成本，而模型厂商理论上可做到零前置投入，因此算租厂商必须选择头部模型，只有模型厂商 Token 销售旺盛，算租厂商才有可能拿到高分成收益，否则算力硬件折旧将形成亏损。

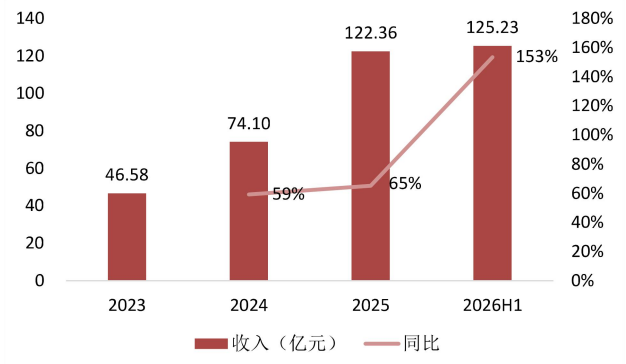
下载日志已记录，仅供内部参考

3. 主要标的

3.1 协创数据：算租绝对龙头，步入业绩加速兑现期

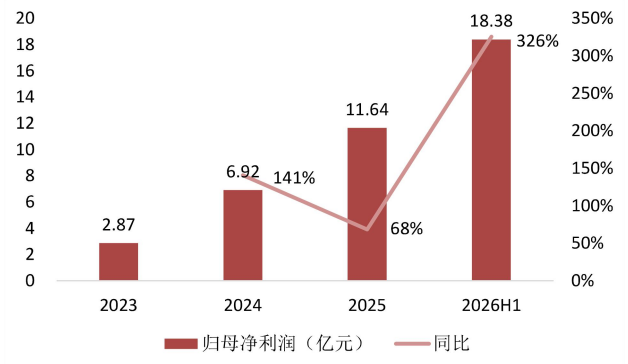
协创数据作为国内算租龙头，已进入利润加速释放期。近年来，协创数据收入保持加速增长趋势，25年收入同比增长65%，26H1收入增速进一步提高至153%，我们认为其增长核心来源在于多个大型算力集群项目完成验收并进入稳定计费阶段，叠加今年以来服务器再制造和数据存储业务同样增长强劲。同时，26H1公司归母净利润同比增速超过营收增速的两倍，规模效应开始体现。

图19：2023-2026H1协创数据收入及增速



资料来源：Wind，山西证券研究所

图20：2023-2026H1协创数据归母净利润及增速



资料来源：Wind，山西证券研究所

算力集群及在手订单规模国内领先，低成本融资为未来进一步扩张奠定坚实基础。根据梳理，自24年10月公司发布首个9亿元H20服务器采购公告以来，至今已累计公告采购331亿元服务器设备。一方面，在以销定产模式下展现了公司在手订单充沛；另一方面，随着服务器集群逐步交付落地，公司未来的确定性高增长具备有力支撑。而在资金方面，公司通过申请金融机构授信、融资租赁、抵押借贷及定增等多种方式补充资金，25年以来公司已累计申请1415亿元的授信额度，并且，近期公司拟定增募资80亿元，其中46.86亿元将投入智算中心建设项目，在展示在手订单充裕的同时，为算力集群进一步扩张补充低成本资金。

下载日志已记录，仅供内部参考

表 5：协创数据已累计公告采购 331 亿元服务器设备

公告时间	公告	决策会议	采购金额	采购内容	采购方	供应方
2024.10.24	《关于公司算力业务进展情况的自愿性披露公告》	-	9 亿元	H20 NVLINK 服务器	广州奥佳	上海域允
2025.3.8	《关于公司购买资产的公告》	第三届董事会第三十二次会议和第三届监事会第二十九次会议	30 亿元	服务器	协创数据	多家供应商
2025.5.28	《关于公司购买资产的公告》	第三届董事会第三十五次会议和第三届监事会第三十二次会议	40 亿元	服务器	协创数据	多家供应商
2025.8.16	《关于公司购买资产的公告》	第四届董事会第三次会议	12 亿元	服务器	协创数据	多家供应商
2025.10.22	《关于公司购买资产的公告》	第四届董事会第六次会议	40 亿元	服务器	协创数据	多家供应商
2025.12.17	《关于公司购买资产的公告》	第四届董事会第九次会议	90 亿元	服务器	协创数据/子公司	多家供应商
2026.2.13	《关于公司购买资产的公告》	第四届董事会第十一次会议	110 亿元	服务器	协创数据/子公司	多家供应商

资料来源：协创数据关于公司算力业务进展情况的自愿性披露公告，协创数据关于公司购买资产的公告，山西证券研究所

多元业务协同，已形成“算-连-存”全产业链布局。历经近 20 年存储业务积累，协创数据在 NAS、SSD、服务器存储领域持有多项发明专利，形成了丰富的供应链管理与企业级存储运维经验。算力集群在训练数据读写、模型检查点存储等环节对高速存储有较高需求，“算存一体”的综合服务能力形成了区别于纯算力运营商的差异化护城河。此外，公司通过增资取得光为科技 51%股权，正式切入 400G/800G/1.6T 光模块赛道，向上游核心器件延伸，进一步完善“算-连-存”的全产业链布局。

下载日志已记录，仅供内部参考

图 21：协创数据已形成“算-连-存”的全产业链布局



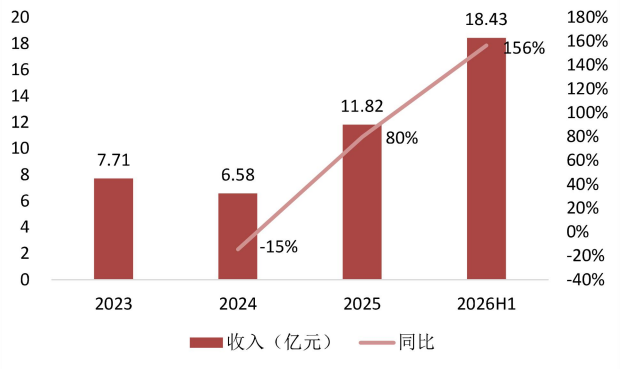
资料来源：协创数据公众号，山西证券研究所

3.2 宏景科技：智慧城市转型算力服务，利润向上空间打开

从智慧城市成功转型算力服务，宏景科技开始快速兑现利润。公司战略重心已从智慧城市解决方案全面转向以高性能算力服务为核心的数字基础设施业务，研发人才结构调整（强化 AI 基础设施、高算力、算网融合等方向的团队建设）的同时完成了算力相关子公司的布局（设立全资子公司深圳数字纵横科技有限公司、控股子公司深圳宏景纵横科技有限公司，并注销全资子公司贵州新瓴科技有限公司），25 年公司算力服务业务收入占比已达到 77%，26H1 占比进一步提升。受益于业务结构性升级，公司利润率显著提升，推动公司 25 年归母净利润实现扭亏为盈，达到 0.26 亿元，同比增长 135%，26H1 归母净利润达 1.37 亿元，同比增长 127%。

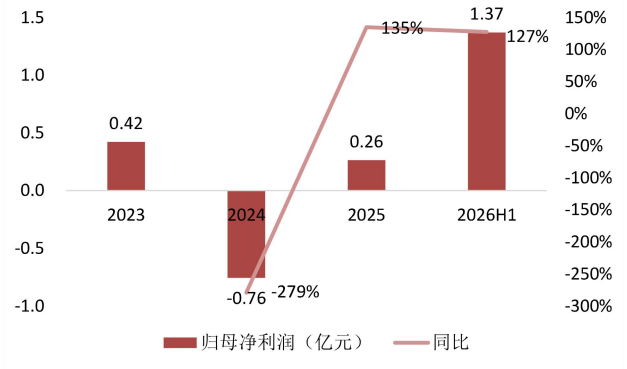
下载日志已记录，仅供内部参考

图 22：2023-2026H1 宏景科技收入及增速



资料来源：Wind，山西证券研究所

图 23：2023-2026H1 宏景科技归母净利润及增速



资料来源：Wind，山西证券研究所

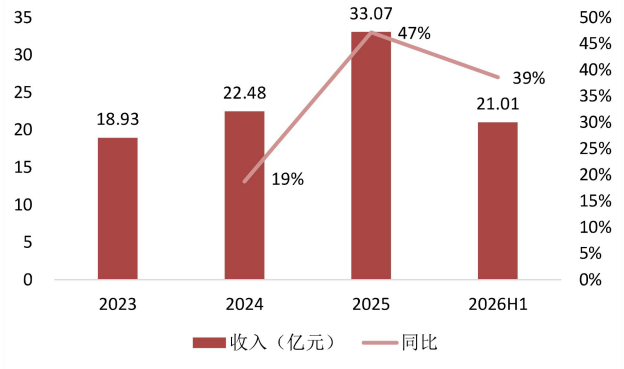
重点打造 Token 工厂，开启利润弹性窗口。公司在持续推进传统服务器集群租赁的同时，通过升级算力调度、分布式存储和高效训推能力，快速打造高产出、低成本的 Token 工厂，将算力规模优势转化为产能与盈利优势。相较于传统的服务器租赁，Token 工厂模式的利润弹性更大，2025 年公司毛利率已提升至 20%以上，2026 上半年毛利率进一步提升至 35%，未来公司毛利率有望随着 Token 工厂模式渗透率提高而持续提高。

3.3 利通电子：英伟达 Preferred AI 云伙伴，轻资产运营提高利润弹性

传统制造主业扭亏为盈计划持续推进，同时，主要以轻资产模式运营的算力租赁业务快速上量将打开长期盈利空间。在制造业务方面，利通电子管理层明确核心目标为扭亏为盈，目前正通过管理优化（变革销售/采购/运营策略等）、海外产能盘活（推动越南工厂持续盈利、墨西哥工厂减亏扭亏等）、新增长点布局推进（向高端智能制造转型等）三维度不断推进，为公司利润率改善奠定基础；在算力业务方面，公司采用直租+转租的模式，并以转租为主，截至 26 年 6 月底，公司在手算力为 3.8 万 P，其中直租算力 1.6 万 P，转租算力 2.2 万 P。由于算力转租以轻资产模式运营，并采用净额法结算，ROE 和毛利率较高，因此随着租赁业务放量，公司的利润弹性将高于业内其他多数厂商。

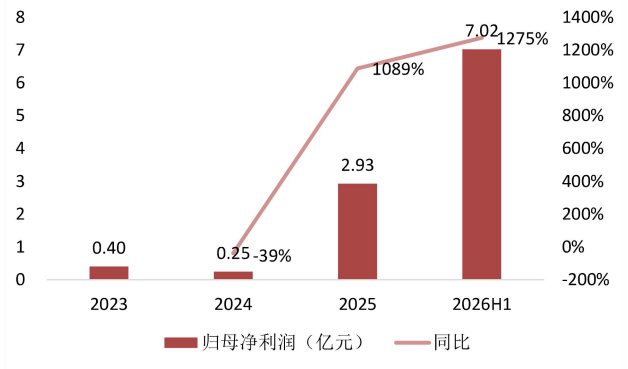
下载日志已记录，仅供内部参考

图 24：2023-2026H1 利通电子收入及增速



资料来源：Wind，山西证券研究所

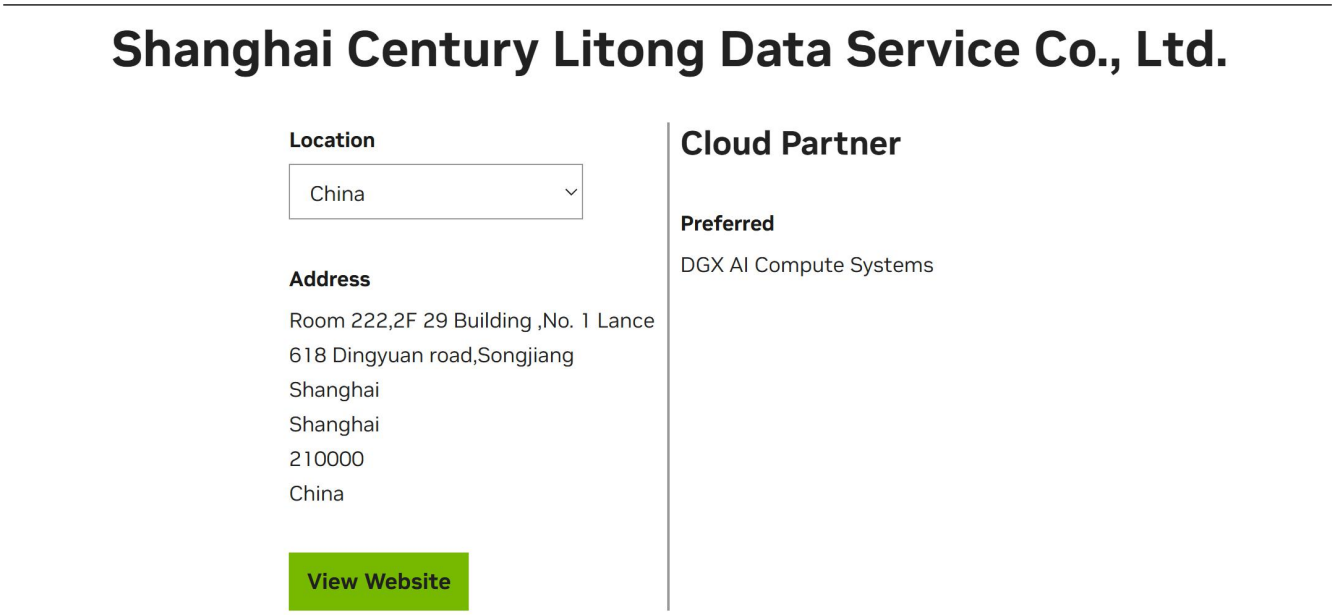
图 25：2023-2026H1 利通电子归母净利润及增速



资料来源：Wind，山西证券研究所

利通电子是中国大陆唯一英伟达 Preferred DGX/AI 云合作伙伴，具备拿卡渠道优势。根据英伟达 NPN 合作伙伴体系，Elite 级为最高等级，Preferred 级为次高级。在中国大陆算力租赁公司中，仅有利通电子（子公司世纪利通）获得 Preferred 级 AI 云伙伴资质，其独享高端 GPU 优先直供、长协锁价及稳定配额，是英伟达在中国大陆高端算力领域的核心官方枢纽，相比于普通渠道，其在拿卡成本、货源稳定性、交付周期、配额量等方面具备明显竞争优势。

图 26：世纪利通具有英伟达 Preferred 级 AI 云伙伴资质



资料来源：英伟达官网，山西证券研究所

下载日志已记录，仅供内部参考

4. 投资逻辑

国内 AI 算力严重供不应求推动今年以来高端 AI 芯片租赁价格持续上行，由于算力需求持续旺盛的同时供给端面临产能瓶颈+海外管制约束，供给缺口短期内难以弥补，算租行业高景气预计将延续全年。**1）市场格局：**随着美国芯片出口管制收紧导致高端卡供货渠道遭封堵，以及中小厂商出租率较低使其面临折旧和利息的固定成本困境，叠加中小厂商缺乏强业务背书推高融资成本，未来中小厂商大概率将逐步出清，市场份额向头部聚拢。基于算租行业的四大核心竞争维度（供应链能力、资金实力、客户资源和组网及运维能力），具备显著供应链能力优势的厂商有望率先脱颖而出，建议关注协创数据、利通电子、杰创智能、盛视科技等；**2）商业模式：**算力租赁商业模式正经历从算力资源出租向 Token 分成的跃迁，目前国内算租厂商的 Token 分成模式已进入试点落地阶段，虽规模较小，但未来有望成为国内算力租赁厂商的重要收入和利润来源，根据我们测算，在中性假设下，到 2028 年算租厂商的 Token 分成收入有望达到 592 亿元，并且 Token 分成模式下利润率将显著提升，因此利润弹性更大。在 Token 分成模式下，算租厂商除拿卡渠道和资金实力外，还需要优异的推理优化能力及与头部模型厂的深度绑定，因此随着 Token 分成模式的持续渗透，具备强推理优化技术并与智谱等头部模型厂绑定的算租厂商将充分受益，建议关注宏景科技、软通动力、中软国际等。

表 6：主要公司盈利预测和估值情况

证券代码	证券名称	收盘价	EPS（26E）	EPS（27E）	EPS（28E）	PE（26E）	PE（27E）	PE（28E）
300857.SZ	协创数据	228.0	11.3	21.8	26.4	20.2	10.4	8.6
301396.SZ	宏景科技	138.1	2.5	5.0	5.0	56.3	27.5	27.5
603629.SH	利通电子	91.5	-	-	-	-	-	-
300209.SZ	行云科技	35.8	-	-	-	-	-	-
000967.SZ	盈峰环境	7.6	0.2	0.3	0.3	33.1	28.6	28.6
001339.SZ	智微智能	64.0	2.0	2.8	2.8	32.1	22.6	22.6
301248.SZ	杰创智能	54.3	1.0	4.4	8.5	53.0	12.3	6.4
002990.SZ	盛视科技	58.1	2.3	3.6	4.3	25.6	16.2	13.6
300133.SZ	华策影视	9.2	0.2	0.2	0.2	50.1	41.7	41.7
300687.SZ	赛意信息	26.2	0.3	0.6	0.6	80.8	44.0	44.0
301236.SZ	软通动力	35.9	0.3	0.5	0.5	107.3	74.1	74.1
301382.SZ	蜂助手	32.5	0.7	0.9	0.9	45.5	35.6	35.6
600673.SH	东阳光	29.3	0.5	0.7	0.7	55.1	43.1	43.1
0354.HK	中软国际	3.2	0.2	0.3	0.4	15.3	11.4	8.9
300137.SZ	先河环保	8.8	0.1	0.5	0.5	85.9	16.5	16.5
300872.SZ	天阳科技	16.4	0.2	0.3	0.3	85.2	58.6	58.6
300093.SZ	金刚光伏	20.9	-0.3	-0.3	-0.3	-63.3	-80.9	-80.9

下载日志已记录，仅供内部参考

资料来源：Wind，山西证券研究所；注：采用 2026 年 10 月 13 日的收盘价，协创数据、杰创智能、盛视科技 26-28 年预测 EPS 基于自己的估值模型，其余公司 26-28 年 EPS 均采用 Wind 一致预期数据，收盘价和 EPS 单位均为人民币元

下载日志已记录，仅供内部参考

5. 风险提示

AI 算力需求增长不及预期风险。若 AI Agent 落地速度放缓或大模型商业化受阻，推理需求增速可能显著低于预期，导致供需格局开始逆转，进而可能引发 GPU 租赁价格系统性下行。

GPU 租赁价格大幅下降风险。若昇腾 950 量产超预期或算力租赁厂商发起价格战，GPU 租赁价格可能连续下行，进而算租厂商的收入和利润增长将受到负面冲击。

地缘政治风险。若美国进一步收紧对华 AI 芯片出口管制，国内对英伟达高端 GPU 的获取难度将更大，可能导致国内下游客户加速转向国产算力，对依靠海外卡拿货渠道优势的厂商经营发展将产生不利影响。

行业竞争加剧风险。随着更多参与者涌入 Token 工厂赛道，行业竞争可能加剧，分成比例面临下行压力，部分算租厂商的先发优势可持续性存疑。

行业标准与监管不确定性风险。Token 分成模式涉及的确权、分配、审计、税务等标准尚未完善，若后续此类标准与监管完善进度落后，可能影响 Token 分成模式的规模化落地进度。

下载日志已记录，仅供内部参考

分析师承诺：

本人已在中国证券业协会登记为证券分析师，本人承诺，以勤勉的职业态度，独立、客观地出具本报告。本人对证券研究报告的内容和观点负责，保证信息来源合法合规，研究方法专业审慎，分析结论具有合理依据。本报告清晰准确地反映本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点直接或间接受到任何形式的补偿。本人承诺不利用自己的身份、地位或执业过程中所掌握的信息为自己或他人谋取私利。

投资评级的说明：

以报告发布日后的 6--12 个月内公司股价（或行业指数）相对同期基准指数的涨跌幅为基准。其中：A 股以沪深 300 指数为基准；新三板以三板成指或三板做市指数为基准；港股以恒生指数为基准；美股以纳斯达克综合指数或标普 500 指数为基准。

无评级：因无法获取必要的资料，或者公司面临无法预见的结果的重大不确定事件，或者其他原因，致使无法给出明确的投资评级。（新股覆盖、新三板覆盖报告及转债报告默认无评级）

评级体系：

——公司评级

买入： 预计涨幅领先相对基准指数 15%以上；
增持： 预计涨幅领先相对基准指数介于 5%-15%之间；
中性： 预计涨幅领先相对基准指数介于-5%-5%之间；
减持： 预计涨幅落后相对基准指数介于-5%- -15%之间；
卖出： 预计涨幅落后相对基准指数-15%以上。

——行业评级

领先大市： 预计涨幅超越相对基准指数 10%以上；
同步大市： 预计涨幅相对基准指数介于-10%-10%之间；
落后大市： 预计涨幅落后相对基准指数-10%以上。

——风险评级

A： 预计波动率小于等于相对基准指数；
B： 预计波动率大于相对基准指数。

下载日志已记录，仅供内部参考

免责声明：

山西证券股份有限公司(以下简称“公司”)具备证券投资咨询业务资格。本报告是基于公司认为可靠的已公开信息，但公司不保证该等信息的准确性和完整性。入市有风险，投资需谨慎。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，公司不对任何人因使用本报告中的任何内容引致的损失负任何责任。本报告所载的资料、意见及推测仅反映发布当日的判断。在不同时期，公司可发出与本报告所载资料、意见及推测不一致的报告。公司或其关联机构在法律许可的情况下可能持有或交易本报告中提到的上市公司发行的证券或投资标的，还可能为或争取为这些公司提供投资银行或财务顾问服务。客户应当考虑到公司可能存在可能影响本报告客观性的利益冲突。公司在知晓范围内履行披露义务。本报告版权归公司所有。公司对本报告保留一切权利。未经公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯公司版权的其他方式使用。否则，公司将保留随时追究其法律责任的权利。

依据《发布证券研究报告执业规范》规定特此声明，禁止公司员工将公司证券研究报告私自提供给未经公司授权的任何媒体或机构；禁止任何媒体或机构未经授权私自刊载或转发公司证券研究报告。刊载或转发公司证券研究报告的授权必须通过签署协议约定，且明确由被授权机构承担相关刊载或者转发责任。

依据《发布证券研究报告执业规范》规定特此提示公司证券研究业务客户不得将公司证券研究报告转发给他人，提示公司证券研究业务客户及公众投资者慎重使用公众媒体刊载的证券研究报告。

依据《证券期货经营机构及其工作人员廉洁从业规定》和《证券经营机构及其工作人员廉洁从业实施细则》规定特此告知公司证券研究业务客户遵守廉洁从业规定。

山西证券研究所：

上海
上海市浦东新区滨江大道 5159 号陆家嘴滨江中心 N5 座 3 楼

深圳
广东省深圳市南山区科苑南路 2700 号华润金融大厦 23 楼

太原
太原市府西街 69 号国贸中心 A 座 28 层
电话：0351-8686981
http://www.i618.com.cn

北京
北京市丰台区金泽西路 2 号院 1 号楼丽泽平安金融中心 A 座 25 层