

市场研究部

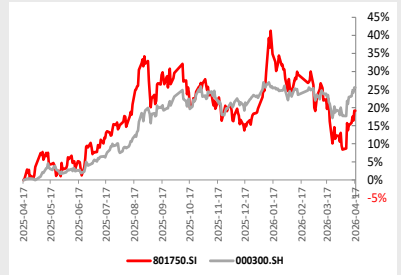
2026 年 4 月 17 日

算力租赁：供需刚性错配下的价值重估

看好

市场表现截至

2026.4.16



数据来源：Wind，国新证券整理

相关研究

分析师：钟哲元
登记编码：S1490523030001
邮箱：zhongzheyuan@crsec.com.cn

证券研究报告

事件

全球算力租赁市场因 AI 需求爆发与硬件供给瓶颈而持续紧张。特别是高端 GPU 租赁价格在短期内快速上涨，反映了供需的严重错配。在国内，智能算力供给与需求之间也存在巨大缺口，推动高端资源出租率居高不下。为应对此局面，主要云服务商已相继提高 AI 算力服务的价格。

核心观点

当前，全球 AI 算力需求的核心驱动已从模型训练转向推理端的规模化消耗。数据显示，AI 应用的爆发推动 Token 调用量呈指数级增长，中国市场增长尤为强劲。这导致推理算力需求已显著超越训练需求，成为行业增长的主引擎。需求激增已传导至全产业链，推高了从核心元器件到算力服务的价格。高端 GPU 租赁市场因供应极度紧张而价格大幅上涨。

然而，需求的爆发与供给端的多重刚性约束形成尖锐矛盾。算力租赁是高度资本密集的行业，自建集群面临巨额投入、漫长周期与严格审批，使得租赁成为获取高端算力，特别是应对弹性需求的主流选择。供给端则受限於高端 GPU 现货稀缺、交付周期延长以及被海外巨头提前锁定的产能。行业具有高固定成本特征，静态成本回收周期约为 3 至 3.5 年，且普遍面临因预付款锁定芯片而导致的“有利润、缺现金流”压力。这些因素共同导致供给释放缓慢，行业进入“有卡为王”的强卖方市场，并加速了市场向头部集中。

中美市场呈现出不同的发展路径。美国市场已形成以英伟达生态为核心，以“照付不议”的 5-7 年超长期合同为主导的成熟模式，为服务商提供了高度确定的现金流。以 CoreWeave 为例，其通过与芯片巨头和下游客户深度绑定，锁定了巨额未来收入。中国市场则因芯片供给不确定性和客户策略等因素，主流合同多为“3 年固定价格+2 年弹性价格”的中期协议。当前最具突破性的趋势是商业模式的升级，行业正从简单的裸算力出租，转向“基础服务费+按 Token 消耗量分成”的价值共享模式，使租赁商得以分享 AI 应用创造的价值。

投资线索

在供需持续错配的背景下，算力租赁行业景气度明确。投资筛选应聚焦四大核心能力：获取高端芯片的渠道能力、健康的资金实力与现金流、高质量的长协订单占比，以及降低运营成本的技术能力。当前，合约期限相对较短、拥有存量算力资源并能有新产能释放的服务商，有望更充分地受益于此轮行业高景气度。

风险提示

1、技术发展不及预期；2、市场竞争加剧；3、地缘政治影响。

目录

一、需求侧：推理需求主导算力指数级扩容.....	3
二、供给侧：重资产属性下的刚性约束.....	3
三、美国市场：英伟达生态主导的长单闭环.....	4
四、中国市场：差异化发展，商业模式升级.....	4
五、投资建议.....	5
六、风险提示.....	5

一、需求侧：推理需求主导算力指数级扩容

全球 AI 产业已完成从技术验证到商业化落地的关键跨越，算力需求驱动逻辑发生根本性切换，从大模型训练端的脉冲式投入转向推理端的常态化、规模化消耗。OpenRouter 数据显示，全球周度 Token 调用量从 1 月初的 6.4 万亿次快速攀升至 3 月底的 27 万亿次，单季度增幅达 321.9%，中国市场表现出更强的增长韧性，国家数据局局长刘烈宏在中国发展高层论坛 2026 年年会上披露，今年 3 月份，中国日均 Token 调用量已突破 140 万亿，较 2024 年初的 1000 亿增长超千倍，超越美国成为全球最大算力消费市场。其中多模态生成与 AI Agent 任务贡献主要增量，代码开发、数据处理等场景的算力消耗呈抛物线式增长。当前全球推理算力总需求已达到训练算力的 5-10 倍，成为拉动行业增长的核心引擎。

需求爆发呈现全产业链传导特征，上游核心元器件与配套服务价格同步大幅上涨。本轮算力需求激增已推高从 DRAM、NAND 闪存、光纤光缆到数据中心托管服务、燃气轮机的全链条价格，GPU 租赁是最新出现供应极度紧张、价格大幅上涨的品类。SemiAnalysis 数据显示，英伟达 H100 GPU 一年期租赁价格出现大幅飙升，截至 2026 年 3 月，一年期合约租金已从 2025 年 10 月的每 GPU 每小时 1.70 美元上涨至 2.35 美元，涨幅接近 40%，全行业 GPU 算力资源已基本全线售罄。服务器厂商为对冲成本风险大幅上调 AI 服务器售价，且涨幅远超元器件成本增幅，今年一季度以来海内外云厂商集体涨价，但当前 AI 工具的投资回报率极高，企业付费意愿强，即便算力价格持续上涨也未能显著抑制需求。此外，需求结构呈现多元化演进趋势，逐步形成“头部厂商训练+中小厂商推理+传统行业数字化改造”的多层次需求格局。

当前，高端 AI 算力租赁已成为市场需求的主流选择。这一趋势主要由几个结构性因素驱动。在合规层面，受严格的出口管制政策限制，企业通过正规渠道自购海外高端算力芯片的额度极为有限，使得租赁成为获取合规算力的关键途径。从资本投入角度看，自建大规模算力集群涉及巨额的一次性硬件投资、漫长的建设周期以及复杂的能评、环评审批，难以快速响应市场变化。此外，大模型训练与推理应用本身存在显著的波峰波谷，自建固定容量的模式缺乏弹性。相比之下，算力租赁能够实现资源的快速交付与弹性伸缩，有效匹配企业动态的业务需求。因此，国内头部科技企业普遍采用了“以国产算力自建为基础底座，以租赁海外高端算力满足弹性需求”的混合策略。

二、供给侧：重资产属性下的刚性约束

然而，需求的爆发式增长与供给端的多重刚性约束形成了尖锐矛盾，导致全球高端算力持续短缺。一方面，高端 GPU 现货资源几近枯竭，官方渠道的交付周期不断延长；另一方面，新一代芯片的产能已被海外云巨头以长期协议提前锁定。这

使得国内市场的增量供给极为有限，行业进入“有卡为王”的强卖方市场。供需的严重失衡直接推动算力租赁价格进入快速上行通道。目前，海内外高端 GPU 的租金价格持续攀升，头部服务商的算力上架率普遍超过 95%，已锁定的算力资源即使面对涨价，客户也因稀缺性不愿释放。

算力租赁是高度资本密集和资源密集的行业，产能扩张受到硬性约束。行业具有高固定成本的显著特征，前期硬件采购与数据中心建设需投入巨额资本。目前一个算力集群的静态成本回收周期约为 3 至 3.5 年。这要求企业必须具备强大的资金实力和融资能力。更关键的是，行业普遍面临“增收不增现”的财务压力。企业需要提前支付大额预付款以锁定芯片，但集群建设、上架调试周期长，客户回款存在账期，导致多家企业出现净利润可观但经营性现金流为负的情况，现金流缺口持续扩大，这严重制约了其扩张步伐。

此外，上游元器件的涨价进一步加剧了供给紧张。AI 服务器价格的上涨侵蚀了项目的投资回报率，部分中小型运营商被迫放缓或取消了部署计划。同时，获取高端 GPU 的合规渠道、以及匹配高功耗芯片的液冷基础设施能力，构成了更高的行业壁垒。多重压力下，行业洗牌加速。拥有早期低价芯片库存、资金充裕的头部服务商抗风险能力强，而缺乏稳定芯片来源和成本优势的厂商则面临出清风险。市场集中度由此快速提升，市场份额向前几名头部企业集中的趋势日益明显。

三、美国市场：英伟达生态主导的长单闭环

对比国际市场，美国算力租赁市场已发展出较为成熟的生态与商业模式。其核心特征在于，以英伟达为代表的芯片巨头通过股权投资与优先供应，与头部算力租赁商深度绑定。比如 CoreWeave 作为独立 AI 算力云服务商，其商业模式深度绑定英伟达。英伟达通过持股和兜底协议确保了 CoreWeave 在芯片供应上拥有优先权，能率先获得最新 GPU。2026 年 4 月，CoreWeave 在短期内接连与 Meta 和 Anthropic 达成重大合作，其累计在手未交付订单高达 878 亿美元，其主要业务由多年期锁价合同覆盖，未来收入确定性极高。这些租赁商再与下游的云厂商和 AI 实验室签订长达 5 至 7 年的“照付不议”长期合同。这种合同规定，无论客户是否实际使用，都需支付最低费用，从而为租赁商提供了高度确定的未来收入和现金流，使其能够基于长期订单获得低成本融资以持续扩产。海外云服务商正处于激进扩张期，将大量现金投入研发和基础设施建设，表现为战略性亏损；而国内服务商在当前高景气度下，更侧重于兑现早期存量资产的利润。

四、中国市场：差异化发展，商业模式升级

中国市场则呈现出不同的发展路径。在合同模式上，以“3 年固定价格+2 年弹性价格”的中期协议为主流，不同于美国的超长单。这主要源于国内客户资本开支

更为谨慎、租赁方面面临芯片供应不确定性等因素。同时，国内客户通常需要支付一定比例的预付款，这在一定程度上缓解了服务商的现金流压力。当前最具突破性的趋势在于商业模式的升级，行业从过去简单地出租物理算力资源，转变为采用“基础服务费+按实际 Token 消耗量分成”的价值共享模式。这使得算力租赁商得以分享 AI 应用创造的价值，而不仅仅是提供底层资源。比如国内主要电信运营商已将其确立为主要战略方向，同时，阿里云、腾讯云等主流云厂商推出的大模型服务（MaaS），其核心计费方式正是按调用 Token 量结算。这标志着行业的盈利逻辑正从提供底层硬件资源，转向深度参与 AI 应用的价值创造，并按产生的智能处理量进行收益分成。此外，帮助客户将受管制的算力资源转化为合规可用的服务，是国内市场的核心价值之一。

五、投资建议

建议围绕四大核心能力筛选投资标的。第一，拿卡能力，这是行业最核心壁垒，优先选择具有高端算力芯片渠道的企业；第二，资金实力，优先选择融资成本低、资产负债率合理、现金流健康的企业，规避高杠杆流动性风险；第三，订单质量，重点关注长协占比高、客户结构多元的企业；第四，技术运营能力，液冷技术与智能调度系统可降低运营成本、提升 GPU 利用率，具备相关优势的企业将获得超额收益。当前阶段，合约期限短、拥有大量 H100 存量、近期有新增容量释放的服务商，将更充分受益于行业景气度。

六、风险提示

- 1、技术发展不及预期；
- 2、市场竞争加剧；
- 3、地缘政治影响。

投资评级定义

公司评级		行业评级	
强烈推荐	预期未来 6 个月内股价相对市场基准指数升幅在 15%以上	看好	预期未来 6 个月内行业指数优于市场指数 5%以上
推荐	预期未来 6 个月内股价相对市场基准指数升幅在 5%到 15%	中性	预期未来 6 个月内行业指数相对市场指数持平
中性	预期未来 6 个月内股价相对市场基准指数变动在-5%到 5%内	看淡	预期未来 6 个月内行业指数弱于市场指数 5%以上
卖出	预期未来 6 个月内股价相对市场基准指数跌幅在 15%以上		

免责声明

钟哲元，在此声明，本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。

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而直接或间接收到任何形式的补偿等。在本人所知情的范围内，本人所在机构、本人以及本人的利害关系人与本报告所评价或推荐的证券不存在任何利害关系。国新证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，以下简称本公司）已在知晓范围内按照相关法律规定履行披露义务。本公司的资产管理和证券自营部门以及其他投资业务部门可能独立做出与本报告中的意见和建议不一致的投资决策。本报告仅提供给本公司客户有偿使用。

本公司不会因接收人收到本报告而视其为客户。本公司会授权相关媒体刊登研究报告，但相关媒体客户并不视为本公司客户。本报告版权归本公司所有。未获得本公司书面授权，任何人不得对本报告进行任何形式的发布、复制、传播，不得以任何形式侵害该报告版权及所有相关权利。

本报告中的信息、建议等均仅供本公司客户参考之用，不构成所述证券买卖的出价或征价。本报告并未考虑到客户的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时可就研究报告相关问题咨询本公司的投资顾问。本公司市场研究部及其分析师认为本报告所载资料来源可靠，但本公司对这些信息的准确性和完整性均不作任何保证，也不承担任何投资者因使用本报告而产生的任何责任。本公司及其关联方可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务，敬请投资者注意可能存在的利益冲突及由此造成的对本报告客观性的影响。

国新证券股份有限公司市场研究部

地址：北京市朝阳区朝阳门北大街 18 号中国人保寿险大厦 11 层（100020）

传真：010-85556155 网址：www.crsec.com.cn