

传媒行业点评报告

NeoCloud 业绩追踪：供需紧张延续，盈利能力提升

增持（维持）

2026 年 08 月 19 日

证券分析师 张良卫

执业证书：S0600516070001

021-60199793

zhanglw@dwzq.com.cn

投资要点

■ **NeoCloud 供需紧张延续，竞争重点正由资源获取进一步转向交付效率与资本回报。** FY26Q2 CoreWeave 与 Nebius 收入分别同比增长 112% 和 454%，新增合同定价及项目经济性继续改善，AI 算力仍处于供不应求状态；同时，行业供给约束已不再局限于 GPU，而是进一步延伸至电力、数据中心、网络及集群部署等全链条环节。需求端，算力需求正由前沿模型训练向推理、企业 AI 及更多垂直行业扩散，训练、推理、评估与模型迭代逐步形成持续循环，进一步拓宽 NeoCloud 长期需求基础。随着客户类型和工作负载更加多元，行业合同结构也由长期锁定产能逐步向长期、中期、短期及按需服务并存演进，并通过存量 GPU 续租及向推理等场景延伸提升资产全生命周期回报。当前头部厂商较高的 EBITDA 利润率已初步验证 AI 算力业务的单位经济性，未来行业关注重点将逐步从收入、订单和签约容量增长，转向投运效率、盈利能力、现金流及资本回报的持续兑现。

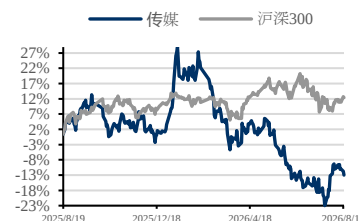
■ **Nebius 需求、定价与盈利能力同步验证，高增长产能储备叠加全栈平台能力打开后续成长空间。** FY26Q2 公司实现收入 5.82 亿美元，同比增长 454%，其中 AI 云收入 5.75 亿美元，同比增长 514%，公司及 AI 云调整后 EBITDA 利润率分别达 40.6% 和 49.7%。公司通过 1-3 年核心 AI 云合同、3-6 个月短期容量合同和投资级客户长期战略合同兼顾利用率、定价弹性与融资能力，其中中期合同 ACV/MW 超过 2,000 万美元，短期容量合同报价超过 4,000 万美元/MW。产能端，公司将 FY26 末签约电力目标提高至 5GW，并引入轻资产合作模式加快容量复制；平台端，持续向“算力+软件栈+模型服务”延伸，通过 Token 工厂、Tavily 及开源模型 Day-0 支持强化差异化竞争力。

■ **CoreWeave 盈利能力持续改善，规模化交付、订单储备与融资能力进一步强化。** FY26Q2 公司实现收入 25.75 亿美元，同比增长 112%，调整后 EBITDA 达到 15.1 亿美元，对应利润率 59%，调整后经营利润率由 FY26Q1 的 1% 提升至 5%，收入保持快速增长的同时盈利能力显著提升，规模效应逐步显现。FY26Q2 公司收入积压达到 1,042 亿美元，同比增长 246%，且尚未计入 FY26Q3 初新增的超过 250 亿美元客户承诺，新签合同利润率及 SKU 价格也继续提升，算力需求与定价仍保持强势。产能端，FY26Q2 已投运电力容量达到 1.5GW，截至 2026 年 8 月 11 日签约电力约 4.2GW，公司进一步将 FY26 末投运目标上调至超过 1.85GW。随着长期 Take-or-Pay 合同、短期容量、托管推理及存量 GPU 续租等多种变现方式逐步完善，公司有望进一步提升 GPU 利用率和资产全生命周期回报。

■ **投资建议：** AI 产业正处于模型能力持续迭代与算力需求加速扩张的共振阶段，建议关注大模型及 AI 算力基础设施产业链投资机会。随着模型能力持续提升、推理成本下降及 Agent 等应用加速落地，模型迭代与算力需求增长有望形成正向循环。其中，模型端建议关注智谱（2513.HK）、MiniMax-W（0100.HK）；算力端建议关注海外 CoreWeave（CRWV）、Nebius（NBIS），以及国内算力租赁及 AI 基础设施产业链相关公司。

■ **风险提示：** AI 算力需求增长不及预期，新增算力产能交付不及预期，资本开支及融资成本高于预期。

行业走势



相关研究

《传媒 ETF 广发（512980）：传统景气修复与 AI 应用价值兑现共振》

2026-08-14

《AI 投入进入回报验证期，云厂商率先兑现增长闭环》

2026-08-04

内容目录

1. NeoCloud：供需紧张延续，行业竞争由资源获取走向综合运营能力	4
1.1. 需求侧：算力需求由训练向推理和企业 AI 扩散，供需紧张仍未缓解	4
1.2. 商业模式：合同期限走向分层，经营目标由锁定利用率转向提升资产全生命周期回报	5
1.3. 盈利与资本：算力单位经济性得到验证，下一阶段核心是将高 EBITDA 转化为可持续资本回报	6
1.4. 龙头分化：CoreWeave 以规模、交付与融资领先，Nebius 在开源模型部署与平台灵活性上寻求差异化	7
2. Nebius：需求、定价与盈利能力同步验证，商业模式向全栈平台演进	9
3. CoreWeave：盈利拐点显现，资产效率持续提升	13
4. 投资建议	16
5. 风险提示	16

图表目录

图 1: Nebius/CoreWeave 收入对比 (百万美元)4

图 2: Nebius/CoreWeave EBITDA 对比 (百万美元)4

图 3: 可用 AI 算力取决于全链条交付能力5

图 4: Nebius 利润率 (%)7

图 5: CoreWeave 利润率 (%)7

图 6: Nebius 收入和同比增速10

图 7: Nebius 调整后 EBITDA 和 Margin.....10

图 8: Nebius 全球算力基础设施持续扩张11

图 9: Nebius 轻资产合作模式12

图 10: CoreWeave 收入和同比增速13

图 11: CoreWeave 调整后 EBITDA 和 Margin.....13

图 12: CoreWeave 收入积压 (FY25Q2 和 FY26Q2)14

图 13: CoreWeave 全球产能布局 (截至 FY26Q2)15

表 1: CoreWeave 与 Nebius 处于不同规模化阶段, 竞争优势各有侧重.....9

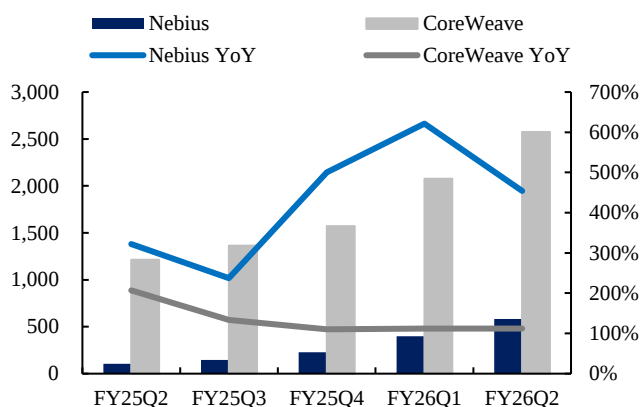
表 2: Nebius 通过三类合同兼顾产能利用率、定价弹性与融资能力11

1. NeoCloud：供需紧张延续，行业竞争由资源获取走向综合运营能力

1.1. 需求侧：算力需求由训练向推理和企业 AI 扩散，供需紧张仍未缓解

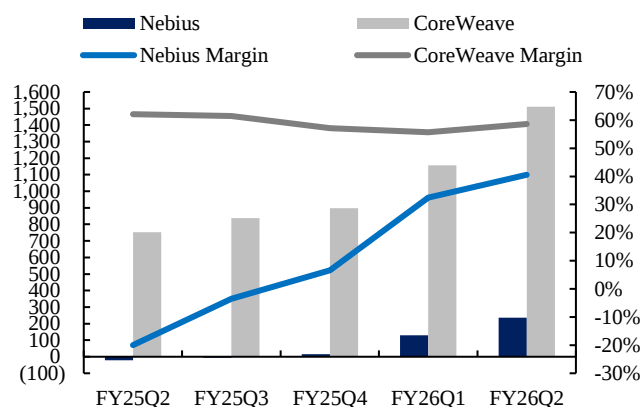
FY26Q2 进一步验证 NeoCloud 算力资源仍处于供不应求阶段，但行业瓶颈已经从单一 GPU 获取转向电力、数据中心、网络和集群交付等全链条能力。CoreWeave FY26Q2 收入同比增长 112%至 25.75 亿美元，收入积压达到 1,042 亿美元，且尚未计入 FY26Q3 初新增超过 250 亿美元客户承诺；Nebius 同期收入同比增长 454%至 5.82 亿美元，其中 AI 云收入同比增长 514%至 5.75 亿美元。与此同时，两家公司均观察到新增算力价格和合同经济性继续改善：CoreWeave FY26Q2 新签合同预计贡献利润率提升，并于 2026 年 7 月将 SKU 价格整体上调约 25%；Nebius FY26Q2 中期合同年平均价值超过 2,000 万美元/MW，短期容量合同超过 4,000 万美元/MW，首次 Blackwell 容量拍卖价格较此前历史最高水平进一步提升 15%。

图1：Nebius/CoreWeave 收入对比（百万美元）



数据来源：Bloomberg，东吴证券研究所

图2：Nebius/CoreWeave EBITDA 对比（百万美元）

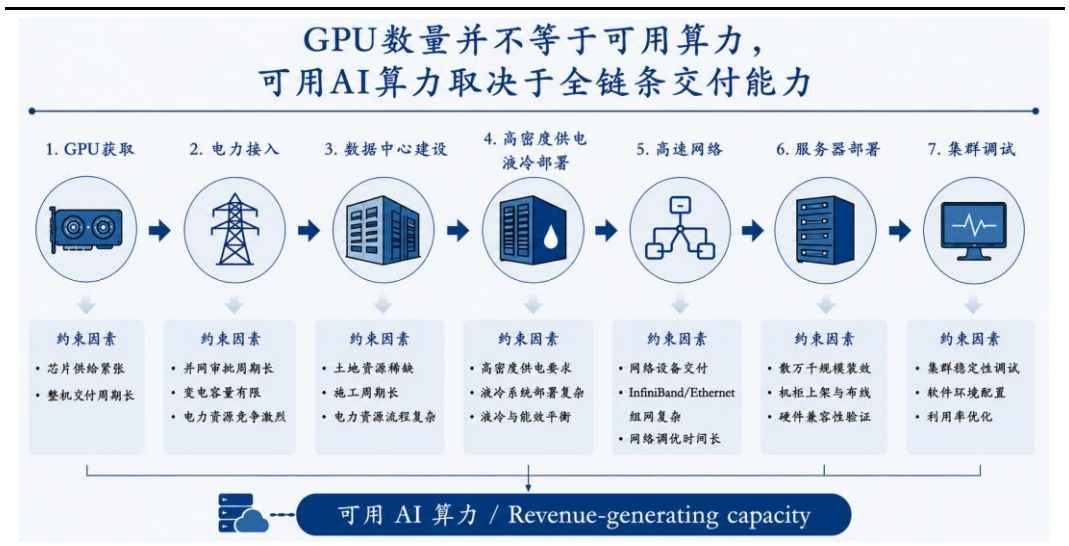


数据来源：Bloomberg，东吴证券研究所

NeoCloud 的核心优势在于更快地将资源转化为可交付算力。大规模 AI 集群需要同步解决电力接入、高密度液冷、机房建设、高速网络、服务器部署及集群调试等环节，任何一项约束都可能推迟算力变现。McKinsey 数据显示，无明显资源约束市场的数据中心建设周期约 12-18 个月，部分电力受限市场甚至超过 3 年。相比之下，NeoCloud 通过提前锁定电力、GPU 和机房资源，并围绕 AI 负载建立标准化工程、网络及软件体系，能够显著压缩部署周期。CoreWeave 在英国投资承诺后仅 6 个月即上线两座 AI 数据中心，公司称这一速度快于竞争对手；Nebius 亦明确将执行速度视为 AI 云的核心壁垒。因此，NeoCloud 的价值并非简单转租 GPU，而是凭借更强的全链条交付能力，更快将

稀缺资源转化为可用算力，并将时间价值转化为商业溢价。

图3：可用 AI 算力取决于全链条交付能力



数据来源：Nebius，东吴证券研究所

需求来源也正在由少数前沿训练客户向推理、企业 AI 和更多垂直行业扩散，降低了 NeoCloud 长期增长仅依赖大模型训练的单一性。过去 NeoCloud 主要承接 Hyperscaler 自有产能不足和前沿模型实验室的大规模训练需求，而随着 AI 逐步进入软件、工业、金融、生命科学和企业工作流，训练、推理、评估与模型迭代开始形成持续循环，计算需求也由模型开发初期的一次性投入逐步转变为伴随应用运行持续产生的基础设施需求。CoreWeave FY26Q2 新增及拓展客户已覆盖工业、生命科学、金融、公共部门等，其托管推理业务预订 ARR 在数月内由约 100 万美元提升至超过 1 亿美元；Nebius 客户亦覆盖 AI 原生企业、模型实验室、Agent 相关公司及大型企业。我们认为，NeoCloud 正在由 Hyperscaler 的阶段性能补充，逐步演变为 AI 产业独立的基础设施供给层，其长期市场空间将更多由 AI 应用规模而非单纯训练集群数量决定。

1.2. 商业模式：合同期限走向分层，经营目标由锁定利用率转向提升资产全生命周期回报

随着客户和工作负载更加多元，NeoCloud 合同模式正由长期锁量逐步向长期、中期、短期及按需服务并存演进。行业发展初期，长期 Take-or-Pay 或容量承购合同的核心作用是锁定 GPU 利用率和未来收入，在需求不确定性较高、资本投入巨大的阶段提供经营确定性。随着算力供需持续紧张、企业客户和推理需求增加，不同客户对期限、弹性和价格的需求开始分化。Nebius 已明确形成三类合同：1-3 年中期合同构成核心 AI

云业务，3-6 个月短期容量用于满足时间敏感需求并获取更高价格，投资级客户长期合同则更多承担锁定需求和支持扩张的功能；CoreWeave 此前以长期 Take-or-Pay 合同为主，FY26 其融资体系也开始支持更短期限客户合同，为企业市场和托管推理业务扩张创造条件。

合同期限分层意味着行业经营目标正在由“尽可能提前售罄产能”转向在收入确定性与价格弹性之间进行动态配置。长期合同能够提供稳定利用率和收入能见度，但提前锁价也可能牺牲未来供需进一步收紧带来的价格上行；短期合同确定性较低，却能够更快反映市场价格，并通常获得更高单位收益。FY26Q2，Nebius 表示即使能够按中期合同条件提前出售全部 FY27 容量，仍主动保留部分产能用于短期合同；CoreWeave 亦指出较短期限合同通常具有更高 ASP 和利润率。对头部厂商而言，未来最优经营策略并非单纯追求最高签约率，而是在长期客户、中短期容量和按需推理之间配置资源，使单位 MW 的收入和利润最大化。

旧代 GPU 的再利用进一步拓宽了资产收益来源，使 NeoCloud 商业模型延伸至 GPU 全生命周期。市场此前对 NeoCloud 的重要担忧之一是 GPU 快速迭代可能导致旧卡价格和残值迅速下降，但 CoreWeave FY26Q2 披露，2020 年推出的 A100 仍能够以具备吸引力的价格签署延续至 2029 年的合同，上一代 GPU 价格亦维持在较高水平。随着推理、微调及企业 AI 应用增加，不同代际 GPU 正在形成更清晰的工作负载分层：最新 GPU 用于高性能训练和前沿推理，成熟 GPU 则可承接成本敏感型推理及稳定工作负载。若旧代 GPU 能够持续续租或转入推理业务，NeoCloud 的经济回报便不再完全依赖首次长期合同，资产使用年限和全生命周期 ROIC 均有进一步提升空间。

1.3. 盈利与资本：算力单位经济性得到验证，下一阶段核心是将高 EBITDA 转化为可持续资本回报

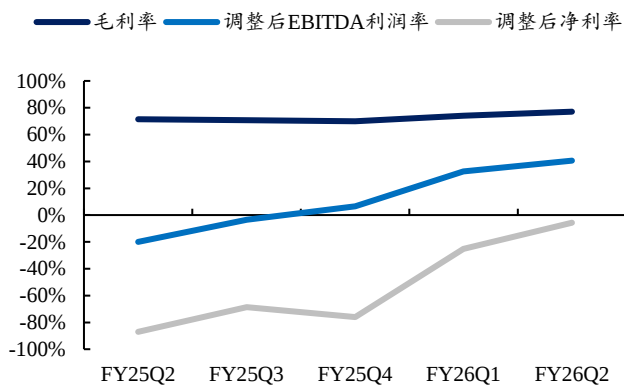
FY26Q2 已经较为清晰地验证头部 NeoCloud 的 AI 算力业务具备较强单位经济性，行业关注点正在由“能否赚钱”转向“资本回报能否持续”。CoreWeave FY26Q2 调整后 EBITDA 利润率达到 59%，调整后经营利润率由 FY26Q1 的 1% 提升至 5%；Nebius 调整后 EBITDA 利润率达到 40.6%，其中 AI 云业务调整后 EBITDA 利润率达到 49.7%。在利用率较高、算力价格保持强势的背景下，GPU 云基础设施本身并非低毛利业务，规模扩大后数据中心运营和人员等固定成本亦能够逐步被摊薄。

但 EBITDA 并不能完整反映 NeoCloud 的商业质量，折旧和融资成本仍决定最终股东回报。NeoCloud 需要在收入形成之前大规模采购 GPU、建设数据中心和配置网络设备，资本支出显著前置；随着规模持续扩大，虽然收入可以摊薄固定成本，但折旧绝对额也会同步增长。同时，相较于依靠经营现金流和投资级信用扩张的 Hyperscaler，NeoCloud 更依赖客户预付款、GPU 资产融资、合同支持贷款、可转债及股权融资，资金成本普遍更高。因此，较高 EBITDA 利润率只能证明算力项目本身具有较好经营回报，

并不意味着净利润和自由现金流会同步快速改善。

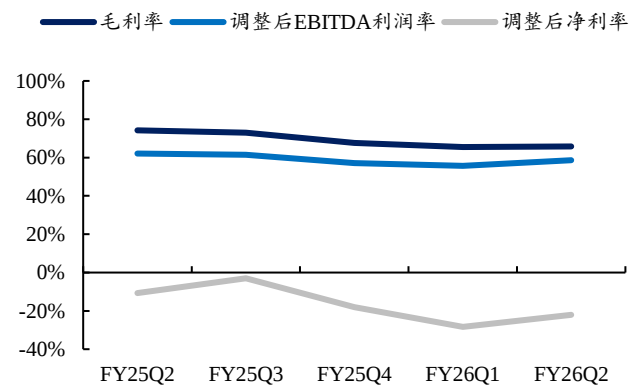
决定 NeoCloud 长期价值的核心变量，最终是资产收益率与资本成本之间的差额。从投入端看，关键在于单位 MW 建设成本、融资成本和建设周期；从运营端看，则取决于产能利用率、单位算力定价及非 GPU 服务附加值；从资产生命周期看，还需要观察合同到期后的续租价格和旧代 GPU 残值。在强需求环境下，高杠杆可以放大产能增长和股东回报，但若新增产能交付延迟、利用率下降或价格回落，折旧和利息同样会放大盈利和现金流波动。因此，评价 NeoCloud 不应只观察收入增速、Backlog 或签约 GW，更需要关注投运 GW、单位 MW 收入与利润、资本回收周期、融资成本及自由现金流等指标。

图4: Nebius 利润率 (%)



数据来源: Bloomberg, 东吴证券研究所

图5: CoreWeave 利润率 (%)



数据来源: Bloomberg, 东吴证券研究所

1.4. 龙头分化: CoreWeave 以规模、交付与融资领先, Nebius 在开源模型部署与平台灵活性上寻求差异化

CoreWeave 已经进入更成熟的规模化交付阶段，其核心优势在于基础设施规模、客户订单和融资能力之间形成了较完整的正向循环。FY26Q2 公司已投运电力容量达到 1.5GW，截至 2026 年 8 月 11 日签约电力容量进一步提升至约 4.2GW，收入积压达到 1,042 亿美元，实际投运规模和收入体量均明显领先于 Nebius。公司通过获取大客户长期合同锁定需求，再以客户合同现金流、GPU 资产及基础设施项目进行融资，并将新增资本继续投入 GPU 采购和数据中心建设，形成订单、融资、交付和收入相互强化的扩张闭环。随着融资工具持续丰富，公司过去一年债务加权平均成本下降近 300bps，资本获取能力已经成为规模化交付的重要竞争优势。对 CoreWeave 而言，下一阶段的核心验证点已由需求和订单转向经营利润和现金流能否随资产规模持续改善。

Nebius 仍处于从大规模资源储备向投运产能转化的阶段。公司将 FY26 末签约电力

目标提高至 5GW，并计划自 FY27 起每年投运超过 1GW 算力，但签约电力仍需经过数据中心建设、接电、网络集成、设备部署及客户验收后才能转化为收入。与 CoreWeave 相比，Nebius 现阶段基础设施规模更小，但合同期限和客户结构更加灵活，同时客户预付款及较轻的企业层债务负担为后续扩张保留了更大的资本结构调整空间。

两家公司均在向全栈 AI 云深化，但平台差异更多体现在软件能力的侧重点。 CoreWeave 的平台体系已经覆盖训练、推理、模型开发、Agent、编排和可观测性，其托管推理业务预订 ARR 超过 1 亿美元，平台能力已经形成较明确的商业化收入。Nebius 则更突出开源模型生态和模型部署优化，公司为前沿开源模型提供 Day-0 支持，并持续针对模型的质量、成本和性能进行部署优化；Token 工厂进一步覆盖模型托管、推理、后训练和部署，公司不仅提供 GPU 容量，也帮助客户更快将开源及定制模型投入生产环境。Nebius FY26Q2 四笔大型 AI 云订单的客户也在实际 POC 中验证了其技术能力，部分客户已进一步讨论后训练和推理合作，说明基础设施客户向上层模型服务交叉销售具备一定基础。

因此，两家公司的后续核心观察点存在明显差异。 CoreWeave 已经较充分验证规模化交付和资本获取能力，后续重点在于高资本开支和杠杆环境下，规模增长能否持续转化为经营利润、现金流及资产回报；Nebius 则需要继续验证签约容量向投运容量的转化，同时观察其在开源模型快速部署、推理优化和软件服务上的技术差异化能否进一步转化为客户黏性、更高单位算力收入和实际软件收入。我们认为，CoreWeave 的核心主线更偏向规模和资本回报兑现，Nebius 的核心主线则是产能兑现与开源模型平台能力商业化。

表1: CoreWeave 与 Nebius 处于不同规模化阶段, 竞争优势各有侧重

对比维度	CoreWeave	Nebius
发展阶段	已进入 GW 级规模化交付和经营杠杆释放阶段	仍处于大规模产能建设与收入快速爬坡阶段
FY26Q2 收入	25.75 亿美元, 同比+112%	5.82 亿美元, 同比+454%
FY26Q2 调整后 EBITDA Margin	59%	40.6%
FY26 收入指引	124-132 亿美元	30-34 亿美元
FY26 末 ARR 指引	185-195 亿美元	70-90 亿美元
合同储备/客户承诺	截至 FY26Q2 收入积压 1,042 亿美元	截至 2026 年 8 月 12 日, 投资级战略客户相关承诺>400 亿美元
FY26 年末电力目标	已投运电力>1.85GW	已投运电力 0.8-1.0GW, 签约电力 5GW
长期产能目标	FY30 已投运电力≥8GW	自 FY27 起每年新增>1GW 已投运容量
FY26Q2 CapEx	94 亿美元	约 57 亿美元
FY26 CapEx 指引	350-390 亿美元	200-250 亿美元
核心合同模式	历史上以约 5 年长期 Take-or-Pay 合同支撑规模扩张, 正增加短期限合同和托管推理	1-3 年中期合同为核心 AI 云业务, 短期合同通常≤6 个月, 另有投资级客户长期战略合同
推理商业化	托管推理已预订 ARR>1 亿美元 (2026/8/11 披露), FY26 末 ARR 目标≥2.5 亿美元; 截至 FY26Q2 末, 存储、CPU、网络 and 软件等非 GPU 产品 ARR>4 亿美元	Token 工厂和 Tavily 已对 FY26Q2 增长产生贡献, 但公司尚未给出独立 ARR 目标
模型部署/推理能力	强调大规模生产推理性能与成本效率; Grace Blackwell 平台创 MLPerf 训练/推理纪录, 公司测试实现较低单位 Token 推理成本	更突出前沿开源模型 Day-0 支持、模型部署、后训练和推理优化; Eigen AI 和 Clarifai 能力已整合进 Token Factory
融资能力	截至 FY26Q2 累计已落实>320 亿美元债务及股权资本; 过去一年债务加权平均成本下降近 300bps, 对应约 11 亿美元年化利息节约	FY26 预计获得>90 亿美元客户预付款; FY26Q2 新签交易相关 CapEx 的 50%-60%可被预付款覆盖; 26 年 7 月完成首笔约 7.75 亿美元资产支持融资
资产负债表特征	大规模债务融资支撑产能扩张, FY26Q2 净利息支出 6.40 亿美元, FY26Q3 利息支出指引 8.6-9.4 亿美元	企业层债务仍较少, 但未来多 GW 建设仍需要持续外部融资
现阶段核心优势	规模化交付、客户订单、资本获取能力、全栈平台商业化	高增长产能储备、容量配置灵活性、开源模型快速部署与推理优化

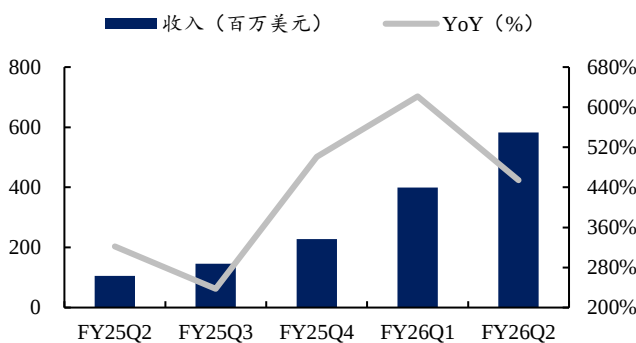
数据来源: CoreWeave, Nebius, 东吴证券研究所

2. Nebius: 需求、定价与盈利能力同步验证, 商业模式向全栈平台演进

AI 云收入持续高速放量, 规模效应推动盈利能力进一步改善。FY26Q2 Nebius 实现收入 5.82 亿美元, 同比增长 454%、环比增长 46%; 其中 AI 云收入 5.75 亿美元, 同比

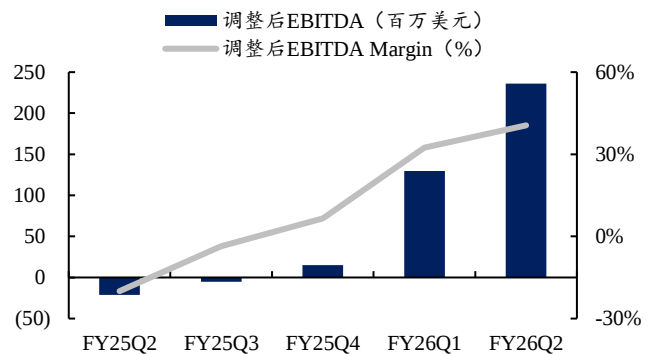
增长 514%，占集团收入接近 99%，期末年化收入运行率达到 30 亿美元，较 FY26Q1 的 19.2 亿美元提升约 56%。FY26Q2 集团调整后 EBITDA 达到 2.36 亿美元，对应利润率 40.6%，环比提升 8pct，并显著高于 Bloomberg 一致预期的 28.3%（2026/8/7）；其中 AI 云调整后 EBITDA 达到 2.86 亿美元，对应利润率 49.7%，已投算力资产的经营盈利能力进一步得到验证。收入增长主要受新增算力容量投产、利用率提升及按需需求增长推动，同时 Token 工厂、Tavily 等软件及平台服务开始贡献增量，公司正在从单纯 GPU 资源转租向全栈 AI 云服务延伸。

图6：Nebius 收入和同比增速



数据来源：Bloomberg，东吴证券研究所

图7：Nebius 调整后 EBITDA 和 Margin



数据来源：Bloomberg，东吴证券研究所

公司正在通过三类合同兼顾基础利用率、短期价格弹性与长期融资能力，供需紧张进一步转化为更优的合同经济性。第一类为 Nebius 核心 AI 云业务的 1-3 年中期合同，主要面向 AI 原生公司、模型实验室及企业客户，FY26Q2 签署 4 笔标志性合同，单笔平均价值超过 10 亿美元，对应 ACV/MW 超过 2000 万美元，部分大单达到 2000-2500 万美元/MW；相关合同通常附带较高比例预付款，可覆盖项目资本开支的 50%-60%。第二类为 3-6 个月短期容量合同，主要满足客户时间敏感的训练或推理需求，由于即时可用 GPU 仍高度紧缺，当前报价已超过 4000 万美元/MW，公司因此并未提前售罄全部 FY27 产能，而是主动保留部分容量以捕捉短期溢价；首次 Blackwell 容量拍卖的成交价格也较公司历史最高价格高出约 15%。第三类为面向投资级客户的长期战略合同，其主要作用并非追求最高单位价格，而是锁定长期需求并支持资产支持融资，加快数据中心建设。此外，FY26Q2 财报电话会上管理层表示当前需求并非增长瓶颈，甚至存在多个买家竞争单个 GPU 的情况，同时新签项目资本及运营成本回收期已由过去 2-3 年缩短至约 1 年 10 个月。

表2: Nebius 通过三类合同兼顾产能利用率、定价弹性与融资能力

合同类型	典型期限	主要客户/使用场景	定价及付款特征	对 Nebius 的主要作用
核心 AI 云合同	1-3 年	AI 原生公司、模型实验室、Agent 公司及大型企业；FY26Q2 代表客户包括 Reflection、Cohere、美国模型实验室及大型量化交易机构	FY26Q2 标志性合同对应年合同价值约 2,000-2,500 万美元/MW；通常附带较高比例预付款，约可覆盖相关 CapEx 的 50%-60%；本季度 4 笔大单平均合同价值均超过 10 亿美元	核心业务模式，锁定较稳定的中期需求和基础利用率，同时保持较好的单位经济性，是公司 AI 云收入增长的主要基础
短期容量合同	通常不超过 6 个月	存在紧急、时间敏感型训练或推理需求的客户，需要立即获得稀缺 GPU 容量	定价显著高于中期合同，目前洽谈价格约 4,000-5,000 万美元/MW，部分更高；公司还尝试容量拍卖，首次 Blackwell 拍卖价格较此前历史最高价高约 15%	捕捉短期供需紧张带来的价格溢价，公司即使能够按中期合同条件售罄 FY27 容量，仍主动保留部分产能用于短期交易，以提高单位容量收益和定价弹性
长期战略合同	长期	投资级客户及大型战略客户，典型用途是大规模、长期算力采购	公司未统一披露该类合同的每 MW 定价；核心特点是客户信用质量高、合同期限长、现金流可预测性较强	主要承担需求锁定和融资功能，公司可以利用投资级客户长期合同支持有担保/资产支持融资，从而更快、更低成本地推进数据中心建设，而非单纯追求最高单位价格

数据来源：Nebius，东吴证券研究所

产能储备继续扩张，同时引入轻资产合作模式以缓解资本与交付约束。公司将 FY26 末签约电力目标进一步上调至 5GW，并计划自 FY27 起每年新增超过 1GW 已投运电力容量；FY26 末已连接电力目标仍维持 0.8-1.0GW。在传统自建及托管数据中心之外，公司 FY26Q2 进一步提出轻资产合作模式：基础设施融资、土地、设施、电力及硬件建设可更多由合作伙伴承担，Nebius 主要负责数据中心集群架构、云平台 and 软件能力，并负责导入客户需求和增值服务。轻资产合作有望成为自有建设模式的重要补充，在降低资产负债表资金占用的同时，加快新增容量复制。

图8: Nebius 全球算力基础设施持续扩张



数据来源：Nebius，东吴证券研究所

图9：Nebius 轻资产合作模式



数据来源：Nebius，东吴证券研究所

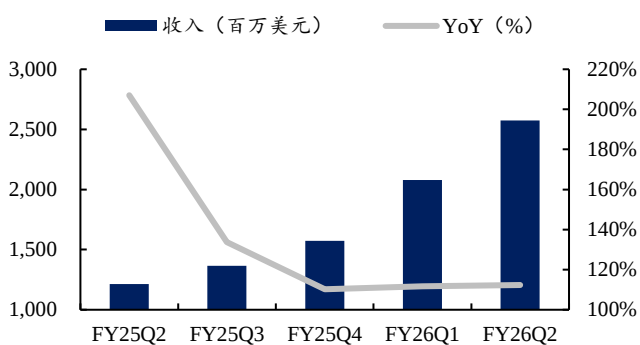
软件及平台能力持续完善，全栈服务强化 AI 云差异化竞争力。Nebius 持续向全栈 AI 云平台延伸,通过快速接入前沿模型和完善软件服务提升客户使用效率与平台黏性。公司以多租户 AI 云和自研软件栈为基础，面向 AI 原生企业、Agent 开发商、模型实验室及大型企业客户提供从 GPU 计算、存储、网络到集群调度、模型部署和推理的一体化服务，并强调通过软件能力帮助客户在规模化使用 AI 时兼顾性能、成本和可靠性。针对前沿开源模型快速迭代，公司提供 Day-0 支持，已陆续接入 Nemotron Light、Nemotron Ultra、GLM5.2、Kimi K3、DeepSeek Flash、MiniMax3 等模型；同时持续优化模型部署架构和运行效率,提升不同模型在平台上的质量、成本及性能表现。产品层面，Token 工厂提供模型推理与部署服务，Tavily 则向 Agent 及搜索场景延伸，二者均是一站式 AI 平台的重要组成部分。我们认为，随着客户需求由单纯获取 GPU 逐步转向高效运行模型和应用，Nebius 的竞争力也将更多由 GPU 资源供给拓展至“算力+软件栈+模型服务”，有助于提升单位算力变现能力和客户黏性，并形成区别于单纯 GPU 租赁的服务溢价。

融资体系逐步完善，资本效率仍待持续验证。随着长期客户合同规模扩大，Nebius 正逐步将客户信用和项目现金流转化为融资能力。公司 2026 年 7 月完成首笔 7.75 亿美元有担保融资，以已部署 GPU 及投资级客户合同现金流为基础，资产支持融资资金亦可在不同 AI 云基础设施项目间灵活调配。同时，客户预付款已成为建设资金的重要来源，从而缩短公司自有资金占用周期。不过，公司仍处于大规模建设阶段，FY26Q2 经营活动净现金流入 22.46 亿美元，而投资活动净现金流出 56.46 亿美元，其中购置物业、设备及无形资产支出 56.57 亿美元。后续更值得关注的是新增资本投入能否及时转化为可运营算力和收入，以及融资成本、自由现金流和潜在股权稀释能否随着业务规模扩大逐步改善。

3. CoreWeave: 盈利拐点显现，资产效率持续提升

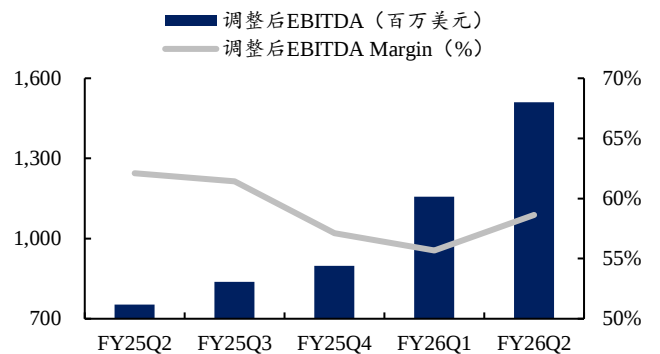
收入保持高速增长，规模效应开始向经营利润率兑现。FY26Q2 CoreWeave 实现收入 25.75 亿美元，同比增长 112%、环比增长 24%，调整后 EBITDA 达到 15.1 亿美元，同比增长约 101%，对应利润率 59%。更值得关注的是，调整后经营利润由 FY26Q1 的 0.21 亿美元提升至 1.28 亿美元，调整后经营利润率由 FY26Q1 的 1% 升至 5%，收入环比增长的同时利润增速明显更快，反映前期大规模产能建设后固定成本摊薄和运营杠杆开始释放。基于当前需求及产能交付进度，公司将 FY26 全年收入指引上调至 124-132 亿美元，调整后经营利润指引升至 9.6-11.5 亿美元，并将 FY26 年末年化收入运行率指引提高至 185-195 亿美元；公司预计调整后经营利润率将在 FY26H2 继续环比改善，并于 FY26Q4 达到低双位数水平。

图10: CoreWeave 收入和同比增速



数据来源: Bloomberg, 东吴证券研究所

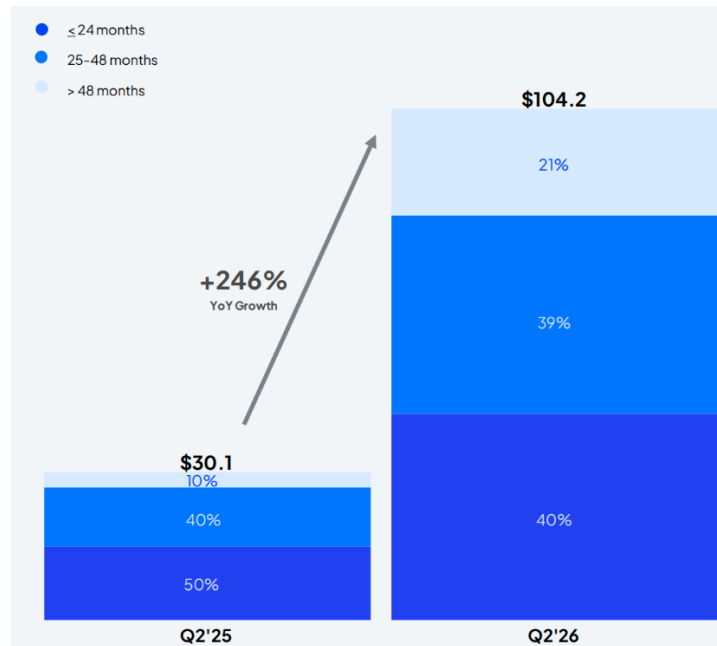
图11: CoreWeave 调整后 EBITDA 和 Margin



数据来源: Bloomberg, 东吴证券研究所

需求持续供不应求，订单储备、价格和新签合同利润率同步改善。FY26Q2 公司收入积压达到 1,042 亿美元，同比增长 246%，其中约 40% 预计在未来 24 个月内确认，企业及 AI 原生客户持续扩张推动订单结构进一步多元化；1,042 亿美元收入积压尚未包括 FY26Q3 初新增的超过 250 亿美元客户承诺。FY26 Q2 电话会上，管理层表示，几乎每一块新上线 GPU 均对应多个潜在客户，近期容量基本售罄，Blackwell 和 Vera Rubin 等新一代产品的价格及利润率均创历史新高，旧代 GPU 价格也维持在较高水平；FY26Q2 新签合同预计贡献利润率较此前几个季度高 5-10pct，26 年 7 月公司又根据供需环境对 SKU 价格整体上调约 25%，反映算力稀缺性、平台性能以及客户自身 AI 应用 ROI 提升正共同转化为更强的定价能力。

图12: CoreWeave 收入积压 (FY25Q2 和 FY26Q2)



数据来源: CoreWeave, 东吴证券研究所

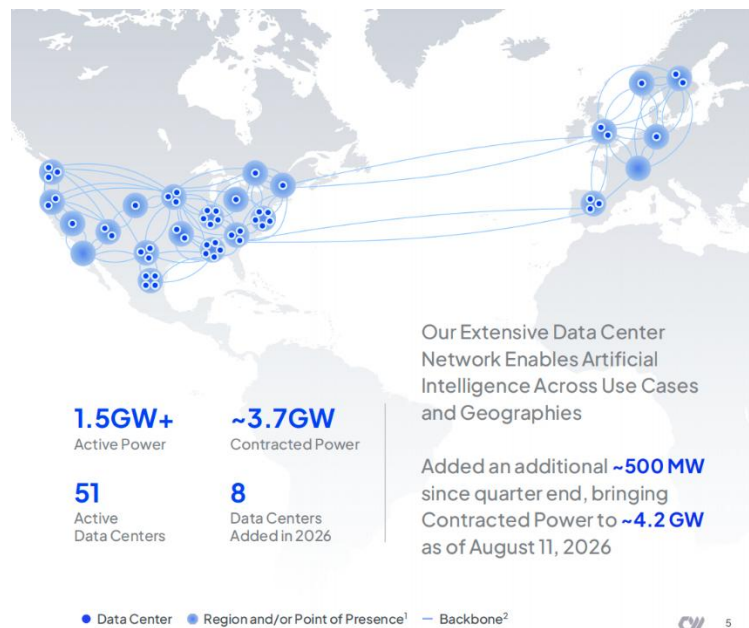
长期 Take-or-Pay 合同仍构成业务基本盘, 短期合同及存量 GPU 再利用进一步提升资产全生命周期回报。CoreWeave 此前主要依靠长期 Take-or-Pay 合同锁定客户需求和基础设施回报, 典型合同期限约 5 年, 资本开支集中于建设初期, 集群交付后合同收入逐步爬坡并形成稳定现金流。FY26Q2 电话会上, 公司表示初始合同期内的项目回报已能够偿还用于建设的资产层债务, 因此合同到期后的续约、转售或转用于其他业务均可形成增量收益。FY26Q2 进入续约阶段的设备占整体机群比例仍较低, 但旧代 GPU 需求保持强劲, 公司已将 2020 年推出的 A100 签署新合同并延续至 2029 年, 体现成熟 GPU 在推理等场景下仍具较高使用价值; 与此同时, 托管推理等业务也为到期 GPU 提供了新的变现渠道。融资端, 公司进一步通过公开银团化的延迟提款定期贷款拓宽融资渠道, 融资体系亦开始支持期限更短的客户合同。随着合同期限和变现方式更加多元, 公司正由以长期锁量为主, 逐步拓展至长期 Take-or-Pay 合同、短期高价容量和托管推理等多种模式, 在维持收入可见度的同时, 有望进一步提升 GPU 利用率与资产全生命周期回报。

客户需求正由前沿模型训练向企业生产环境和更多行业扩散, 客户结构进一步拓宽。FY26Q2 公司新增及拓展客户覆盖工业、生命科学、金融、公共部门等: Caterpillar 将利用 Vera Rubin 平台开展工业级 Physical AI 训练和推理, Isomorphic Labs 用于 AI 药物研发, Flow Traders 和 IMC 用于量化交易模型开发, Leidos 合作延伸至国防和国家安全, IBM 则在 CoreWeave 上开展强化学习、Agent 工具使用及模型评估。公司认为, AI 工作负载正从过去相对集中的模型训练逐渐形成“训练—推理—评估—再训练”的持续循环,

算力需求因此不再是模型开发初期的一次性投入，而会伴随 AI 应用进入生产环境持续产生，这也使企业客户和推理业务逐渐成为继 Hyperscaler 及前沿模型公司之后新的需求来源。

产能扩张继续提速，电力资源和全球数据中心布局为后续收入增长提供基础。 FY26Q2 公司已投运电力容量达到 1.5GW，单季度新增接近 500MW；季度末签约电力达到 3.7GW，截至 8 月 11 日进一步增至约 4.2GW。此外，公司还拥有超过 1.5GW 来自已储备带电土地、现有站点扩建权及已签署意向书的潜在电力资源。公司将 2026 年末已投运电力容量目标由超过 1.7GW 上调至超过 1.85GW，并维持 2030 年超过 8GW 目标；海外方面，目前已签约超过 1GW 美国以外电力资源，包括近期进入印度尼西亚的 360MW 项目，同时首批自建数据中心亦在推进。我们认为，在近端需求已经显著超过供给的情况下，CoreWeave 未来收入增长的主要约束仍在于电力、数据中心、GPU 和供应链资源能否按计划转化为实际可运营算力。

图13: CoreWeave 全球产能布局 (截至 FY26Q2)



数据来源：CoreWeave，东吴证券研究所

资本结构持续优化，但高强度扩张下的资金成本仍需关注。 FY26Q2 公司资本开支达到 94 亿美元，随着新增算力交付提速，FY26 全年资本开支指引进一步提升至 350-390 亿美元，基础设施建设仍处于高投入阶段。与扩张节奏相匹配，公司持续优化债务结构和融资成本，过去一年债务加权平均成本下降近 300bps，按 FY26Q2 期末债务规模计算，对应约 11 亿美元年化利息节约。但债务规模仍将随产能建设继续上升，公司预计 FY26Q3 利息支出达到 8.6-9.4 亿美元，反映融资成本下降尚不足以完全抵消债务规

模扩张带来的利息压力。因此，后续更值得关注的是新增资本开支能否按计划转化为投运算力和收入，并通过利润率提升及现金流增长逐步消化高杠杆带来的财务负担。

4. 投资建议

AI 产业正处于模型能力持续迭代与算力需求加速扩张的共振阶段，建议关注大模型及 AI 算力基础设施产业链投资机会。随着模型能力持续提升、推理成本下降及 Agent 等应用加速落地，模型迭代与算力需求增长有望形成正向循环。其中，模型端建议关注智谱（2513.HK）、MiniMax-W（0100.HK）；算力端建议关注海外 CoreWeave（CRWV）、Nebius（NBIS），以及国内算力租赁及 AI 基础设施产业链相关公司。

5. 风险提示

- 1) AI 算力需求增长不及预期：**若大模型训练、推理及企业 AI 应用落地速度不及预期，可能导致算力需求、产能利用率及合同价格下降，影响 NeoCloud 收入增长和盈利能力；
- 2) 新增算力产能交付不及预期：**AI 算力投产涉及 GPU、电力、数据中心、液冷、网络及集群部署等多个环节，若建设、并网或设备交付延迟，可能影响收入确认及资本回收节奏；
- 3) 资本开支及融资成本高于预期：**NeoCloud 业务资本投入较大，若融资成本上升、融资渠道收紧或资本开支持续超预期，可能增加利息负担和自由现金流压力，并带来潜在股权稀释风险。

免责声明

东吴证券股份有限公司(以下简称“本公司”)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。

本研究报告仅供符合中国证监会《证券期货投资者适当性管理办法》规定的专业投资者使用,且接收人须为本公司认定并建立服务关系的专业机构投资者使用,本公司不会因接收人收到本报告而视其为客户。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,本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。本报告不面向普通投资者发布或传播。任何非专业机构投资者收到本报告后,应立即退回并销毁,不得依据本报告内容做出任何投资决策。

在法律许可的情况下,东吴证券及其所属关联机构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。

市场有风险,投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息,本公司力求但不保证这些信息的准确性和完整性,也不保证文中观点或陈述不会发生任何变更,在不同时期,本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有,未经书面许可,任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的,应当注明出处为东吴证券研究所,并注明本报告发布人和发布日期,提示使用本报告的风险,且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的,应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期(A 股市场基准为沪深 300 指数,香港市场基准为恒生指数,美国市场基准为标普 500 指数,新三板基准指数为三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的),北交所基准指数为北证 50 指数),具体如下:

公司投资评级:

买入:预期未来 6 个月个股涨跌幅相对基准在 15%以上;

增持:预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间;

中性:预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间;

减持:预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间;

卖出:预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级:

增持:预期未来 6 个月内,行业指数相对强于基准 5%以上;

中性:预期未来 6 个月内,行业指数相对基准-5%与 5%;

减持:预期未来 6 个月内,行业指数相对弱于基准 5%以上。

我们在此提醒您,不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系,表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况,如具体投资目的、财务状况以及特定需求等,并完整理解和使用本报告内容,不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号
邮政编码: 215021
传真: (0512) 62938527
公司网址: <http://www.dwzq.com.cn>