

综合

2026年06月09日

AI 驱动变革，云计算进入新一轮扩张周期

——港股行业深度报告

投资评级：看好（首次）

初敏（分析师）

荀月（分析师）

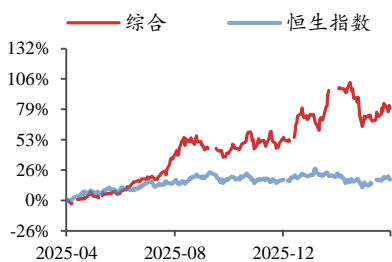
chumin@kysec.cn

xunyu@kysec.cn

证书编号：S0790522080008

证书编号：S0790524110001

行业走势图



数据来源：聚源

相关研究报告

● AI 驱动云计算进入新一轮扩张周期，优选头部全栈平台兼顾高弹性标的

AI 驱动云计算行业主线从渗透率提升切换为 AI 工作负载渗透+MaaS 商业化+CAPEX 转化效率。云厂商竞争核心由资源售卖转向算力、模型、工具链、应用生态的全栈整合。根据中国信通院数据，中国云计算市场 2024 年 8288 亿元、2030 年有望突破 3 万亿元，对应 CAGR 约 23%—24%；其中 GenAI IaaS+MaaS 2024—2029 年 CAGR 达 72%，核心关注 AI 收入占比、推理需求放量、CAPEX 转化效率与毛利率修复。**受益标的：**头部全栈平台，兼顾高弹性标的，**海外：**谷歌、微软、亚马逊，**国内：**阿里巴巴、百度集团、金山云，高端 GPU 长期供给偏紧背景下，NeoCloud 崛起，受益标的 Coreweave、Nebius。

● IaaS 重资产扩张，MaaS 重构云商业模式

AI 推动云计算从弹性资源生意转向重资产基础设施+高附加值模型服务。当前 AI 云收入主要由三部分组成：GPU 租赁、MaaS/API 服务、PaaS/SaaS 工具链。Capex 投入短期压制云业务毛利率，中期抬升行业门槛，长期将进一步强化头部厂商在供给、融资与调度上的寡头优势。与此同时，硬件路线变化也会影响 IaaS 盈利中枢。MaaS 侧，商业模式从模型调用收费逐步演化为 API 引流+PaaS 工具链变现+云资源绑定的盈利路径。当前 MaaS 本身仍普遍承担获客与生态扩张功能，直接利润率并不高，但伴随推理成本下降、企业调用量快速增长、Agent/多模态应用扩散，MaaS 从 AI 云中的边缘收入项升级为增长最快的核心驱动之一。

● 海外云复盘：估值与 Capex 及收入增速显著相关

(1) 云计算从 Capex 提速到收入兑现存在约 1-1.5 年时间差。参考历史数据，云计算资本开支增加后约 3 个季度对现金流产生影响，约 5 个季度对收入及折旧摊销费用形成显著影响。(2) 复盘微软云过去两年发展过程，估值变化与资本开支和云收入显著相关。第一阶段估值与资本开支增速正相关。前期高投入阶段，资本开支加速推动估值上升，后期投入放缓阶段，资本开支增速下降带来估值调整，短期可能引发需求不足的担忧，导致估值阶段性调整、股价波动。第二阶段：资本开支兑现成业绩，云收入加速成为估值核心驱动，将进一步抬升估值。

● 国内云：总量扩容仍将延续，互联网厂商份额有望提升

国内 AI IaaS 市场格局已基本收敛，市场集中度进一步提升。阿里云通过持续加大 AI 基础设施资本支出、加码算力储备，通过提供灵骏 AI 集群、GPU 弹性算力等多元化 AI IaaS 服务，满足客户不同 AI 场景下的旺盛训推需求，市场份额稳居第一；(2) 火山引擎背靠字节跳动算力规模带来的成本优势，搭配相对激进的销售策略，在 AI 市场提供高性价比算力资源；(3) 百度基于昆仑芯片与百舸平台，在算力、网络、训推集群形成技术优势，2025H1 聚焦头部客户算力需求，在汽车及电商、音视频等多个互联网行业有所斩获，进一步扩大市场占比。

● **风险提示：**宏观经济增长放缓、模型进步不及预期、竞争加剧、监管变动

目 录

1、 云计算：从资源和池化到 MaaS 的代际演进	4
1.1、 云计算的本质：从服务器租赁到统一算力调度	4
1.2、 发展历程：IaaS、PaaS、SaaS 到 MaaS	4
1.3、 行业空间：AI 驱动云进入新一轮扩张周期	6
2、 IaaS 重资产扩张，MaaS 重构云商业模式	7
2.1、 IaaS 进入 AI 驱动的重资产扩张周期	8
2.1.1、 北美四大云厂商资本开支显著抬升且上修幅度持续超预期	8
2.1.2、 IaaS 毛利率短期承压，但竞争壁垒同比上升	9
2.1.3、 高端 GPU 长期供给偏紧背景下，NeoCloud 崛起	10
2.2、 从云资源租赁到 AI 交付，MaaS 重构云商业模式	13
3、 海外云复盘：估值与资本开支及收入增速显著相关	14
3.1、 基本面：云计算从 Capex 提速到收入兑现存在约 1-1.5 年时间差	14
3.2、 估值：与资本开支及收入增速显著相关	15
4、 国内云：AI 驱动增长，竞争格局基本收敛	17
4.1、 国内总量扩容仍将延续，AI 成为行业增长的第一驱动力	17
4.2、 格局基本收敛，市场集中度进一步提升	19
5、 受益标的	20
5.1、 海外：传统云龙头走向 AI 全栈，Neocloud 更具弹性	20
5.1.1、 亚马逊、谷歌、微软主导全球云服务市场，AI 需求拉动下微软谷歌更具弹性	20
5.1.2、 Neocloud：Coreweave 规模更大，Nebius 客户多元化	21
5.2、 国内：互联网云厂商有望迎来份额提升	24
5.2.1、 阿里巴巴：具备芯片-云-模型-C 端入口的平台龙头	24
5.2.2、 百度集团：短期搜索业务承压，昆仑芯拆分上市有望驱动估值提升	26
5.2.3、 金山云：小米生态提供确定性，中立定位具备优势	27
6、 风险提示	29

图表目录

图 1： 云计算沿资源服务化-平台化-软件化-智能化路径逐步演进	5
图 2： MaaS 推动云计算整体架构体系演进升级	6
图 3： 罗兰贝格预计到 2030 年全球 AI 将占全球云计算市场的 1/4	6
图 4： 中国云支出结构与美国相比呈现倒挂	7
图 5： 国际云厂商利润率显著高于国内	7
图 6： 目前 GPU 英伟达市场份额仍绝对领先	8
图 7： 云业务利润率随 Capex 投入短期承压长期有望随规模效应企稳并改善	9
图 8： 北美云厂商资本开支/经营现金流比值持续上升	10
图 9： Coreweave 的成本效率显著高于传统云	11
图 10： CoreWeave 与英伟达深度绑定	12
图 11： Coreweave 资产负债率偏高	13
图 12： Coreweave 资本开支占收入比大于 100%	13
图 13： MaaS 竞争力高度依赖工程化能力与云侧调度	13
图 14： 亚马逊 CAPEX 增加后约 3 个季度对现金流产生影响	14

图 15: 复盘 2015-2020 年, 云计算收入增速滞后 CAPEX 增速 1-1.5 年	15
图 16: 微软 PS 估值与资本开支及收入增速显著相关	16
图 17: 2024 年中国云计算市场规模 YOY+34.4%	17
图 18: 中国公有云市场规模 (亿元) 迅速增长	17
图 19: 中国 AI IaaS 2025H1 快速增长	18
图 20: 汽车、金融行业位于中国 Gen AI IaaS 市场需求端前列	19
图 21: IDC 2025H1 中国 Gen AI IaaS 市场中阿里、火山引擎、百度居前	19
图 22: Omedia 中国 AI 云市场 2025H1 市场格局中阿里巴巴、火山引擎、华为云居前	19
图 23: 2026Q1 全球云服务商收入 YOY+35%	20
图 24: 微软及谷歌云收入增长更具弹性	21
图 25: 谷歌云份额显著提升	21
图 26: H100 租赁价格自 2026 年 2 月以来持续走高	22
图 27: 阿里巴巴具备 AI 全栈布局	25
图 28: 阿里云收入同比增速加快	25
图 29: 阿里云利润率 2026Q1 基本稳定	25
图 30: 百度智能云基础设施收入同比增速加快	26
图 31: 2026Q1 金山云收入 YOY+37%	28
图 32: 2026Q1 金山云公有云收入 YOY+47%	28
图 33: 公有云占比提升带动金山云 2025Q3 经调整经营利润率转正	28
表 1: 海外云厂商资本开支维持高增	8
表 2: 谷歌自有业务具备 C 端流量入口	21
表 3: Coreweave 规模更大	22
表 4: 2026Q1 Coreweave 收入同比高增但营业利润率有所下滑	23
表 5: Nebius 2026Q1 收入 YOY+621.5%	23
表 6: 昆仑芯 2024-2025 年订单及合作项目客户由内部驱动向全行业扩张 (不完全统计)	26
表 7: 受益标的盈利预测与估值一览	29

1、云计算：从资源和池化到 MaaS 的代际演进

云计算核心仍是通过统一资源池与弹性调度提升计算资源利用率、降低企业 IT 使用门槛，但其产业形态正因生成式 AI 而发生代际跃迁：从早期以 IaaS 为核心的资源池化，到 PaaS 推动的云原生开发平台化，再到当前以 MaaS (Model as a Service) 为代表的模型能力服务化，云厂商输出的已不再只是算力、存储和数据库，而是进一步演化为可直接调用、可持续优化、可嵌入业务流程的智能能力。这一变化意味着，云计算行业的投资逻辑正在从上云渗透率提升转向“AI 工作负载占比提升+模型调用量扩张+云厂商价值链上移”。从行业阶段看，传统云计算解决的是企业 IT 成本项问题，而 AI 云尤其是 MaaS 开始切入企业增长项和生产力项，推动云服务从资源定价逐步走向价值定价，产业链利润中心也有望从单纯硬件堆栈，逐步转向兼具生态、平台与模型能力的头部云厂商。

1.1、云计算的本质：从服务器租赁到统一算力调度

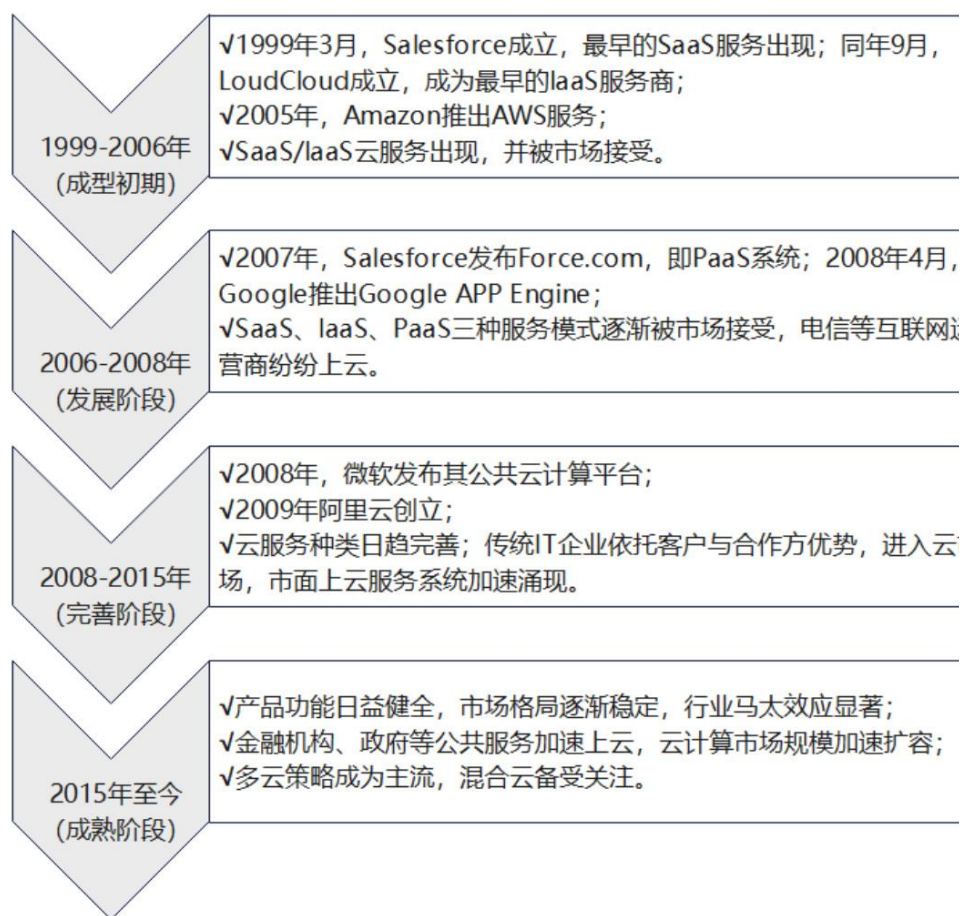
云计算首先是一种资源组织方式的变革。传统本地部署模式下，企业需要预先购买服务器、存储和网络设备，并按峰值需求做冗余建设，导致硬件利用率偏低、扩容周期长、运维复杂；而云计算通过虚拟化与统一调度，将分散的计算、存储、网络资源纳入资源池，实现按需使用、按量付费和弹性伸缩，这构成了其相对传统 IT 体系最核心的增量价值。更关键的是，云计算并不是简单把 IDC 搬到线上，而是把企业原重资产、自维护、低弹性的 IT 能力重构为标准化服务。其底层依赖物理服务器、网络和存储，中间通过虚拟化、容器化和调度系统形成统一资源池，上层再向用户交付 IaaS、PaaS、SaaS 等不同抽象层次的能力。

从部署模式看，云计算又可分为公有云、私有云与混合云。其中，公有云凭借低成本与高扩展性成为主流增长引擎，私有云满足高安全和强合规场景，混合云则适合需要在灵活性与安全性之间平衡的政企客户。

1.2、发展历程：IaaS、PaaS、SaaS 到 MaaS

从全球产业历程看，云计算并非一蹴而就，而是沿着资源服务化—平台化—软件化—智能化的路径逐步演进。(1) 1999 年-2006 年：成型初期，Salesforce 与 LoudCloud 分别推动 SaaS 和 IaaS 雏形落地；(2) 2006 年-2008 年，App Engine 推动 PaaS 兴起；(3) 2008 年-2015 年，微软 Azure、阿里云等平台完善了公共云供给体系；(4) 2015 年至今，行业进入成熟期，产品体系日趋完善，格局向头部集中，多云和混合云逐渐成为主流。

图1：云计算沿资源服务化-平台化-软件化-智能化路径逐步演进



资料来源：智研咨询

生成式 AI 推动云计算架构发生了两层变化。第一，基础设施层从 CPU 导向转向异构算力导向 GPU、NPU、TPU 等成为关键资产，网络、存储与调度系统也需要围绕大模型训练和推理重构；第二，平台与应用层之间新增了 MaaS 层，使模型成为独立且可计量的服务单元。传统云计算的核心是部署应用，AI 云的核心则转为调用模型并编排智能 workflow，这意味着 PaaS 层将从数据库、中间件、操作系统，进一步扩展到向量数据库、Agent 框架、检索增强生成、多模态 workflow 等 AI 原生组件；SaaS 层也从标准化软件逐步演化为 AI 原生应用、助手和智能体。这一架构变化直接改变了云厂商的商业模式。过去 IaaS 更多依赖资源售卖，价格竞争显著；而 MaaS 一旦深度嵌入客户业务流，切换成本会显著提升，绑定层次从基础设施下沉到技术、数据、流程，迁移不仅涉及 IT 重构，更触及业务核心能力与产出逻辑，竞争焦点也从单纯价格转向模型效果、工具链完备度、行业 Know-how 与生态绑定能力。

图2: MaaS 推动云计算整体架构体系演进升级



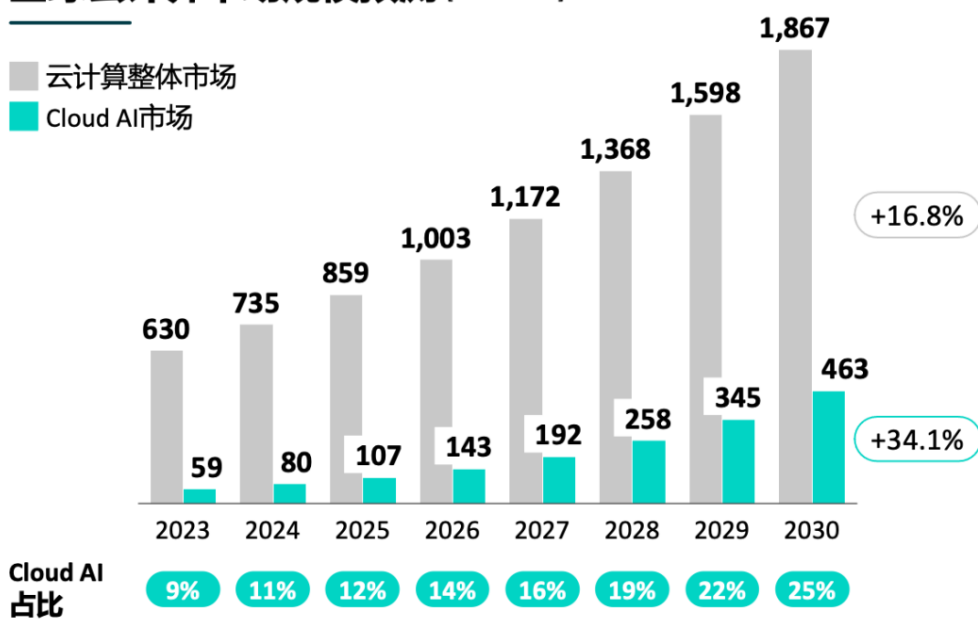
资料来源: 中国信息通信研究院

1.3、行业空间: AI 驱动云进入新一轮扩张周期

根据 Gartner 数据, 2024 年以 IaaS、PaaS、SaaS 为代表的全球云计算市场规模达 6929 亿美元, 同比增速 20.3%, 未来随着 AI 模型训练对 IaaS 消费的拉动以及应用类服务在 MaaS 的推理需求, 预计云计算行业持续增长。根据罗兰贝格数据, 预计 2024 年至 2030 年, 全球云计算市场复合增速为 16.8%, 而 AI 云将保持 34% 的增速, 至 2030 年将占全球云计算市场的四分之一。

图3: 罗兰贝格预计到 2030 年将占据全球云计算市场的 1/4

全球云计算市场规模预测 [USD Billion]



资料来源: 罗兰贝格

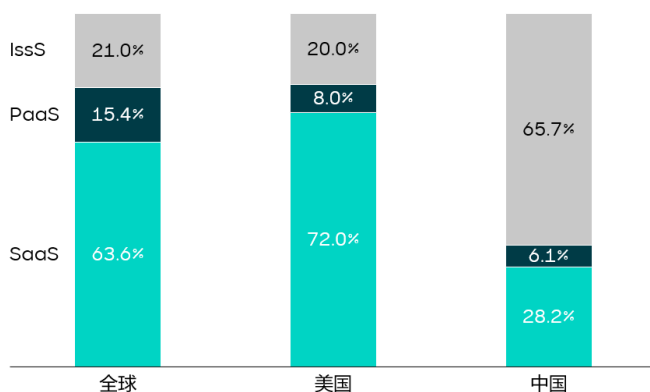
2、IaaS 重资产扩张，MaaS 重构云商业模式

AI 推动云计算从弹性资源生意转向重资产基础设施+高附加值模型服务。当前 AI 云收入主要由三部分组成：GPU 租赁、MaaS/API 服务、PaaS/SaaS 工具链。其中，GPU 租赁仍是现阶段 AI 云收入大头，约占 50%左右；MaaS 收入占比为低个位数；剩余来自数据库、监控、开发框架、插件等 PaaS/SaaS 工具链产品。由于工具链产品边际成本更低，其毛利率可达 70%-80%，反而是 AI 云盈利改善的关键支撑。

(1) IaaS 侧，北美头部云厂商资本开支已从传统通用云扩容切换至 AI 数据中心、GPU 服务器、网络与供电散热体系的系统性重构，行业特征由轻资产软件平台明显转向高 CAPEX、长回收期、强供给约束的基础设施周期。GPU 租赁虽毛利率高于传统 IaaS，但并非高利润率生意，国内 GPU 租赁毛利率约 25%，高于传统 IaaS，但仍显著低于成熟软件层产品。海外口径传统云毛利率可达 70%-80%，而 AI 云毛利率约 40%左右。这意味着，在 AI 需求快速增长初期，IaaS 更多体现为高投入、高周转、利润中性偏低的资源平台，真正决定利润率的是能否在 GPU 租赁之上叠加更多高附加值服务。

图4：中国云支出结构与美国相比呈现倒挂

云计算的IaaS/PaaS/SaaS结构对比



资料来源：罗兰贝格

图5：国际云厂商利润率显著高于国内

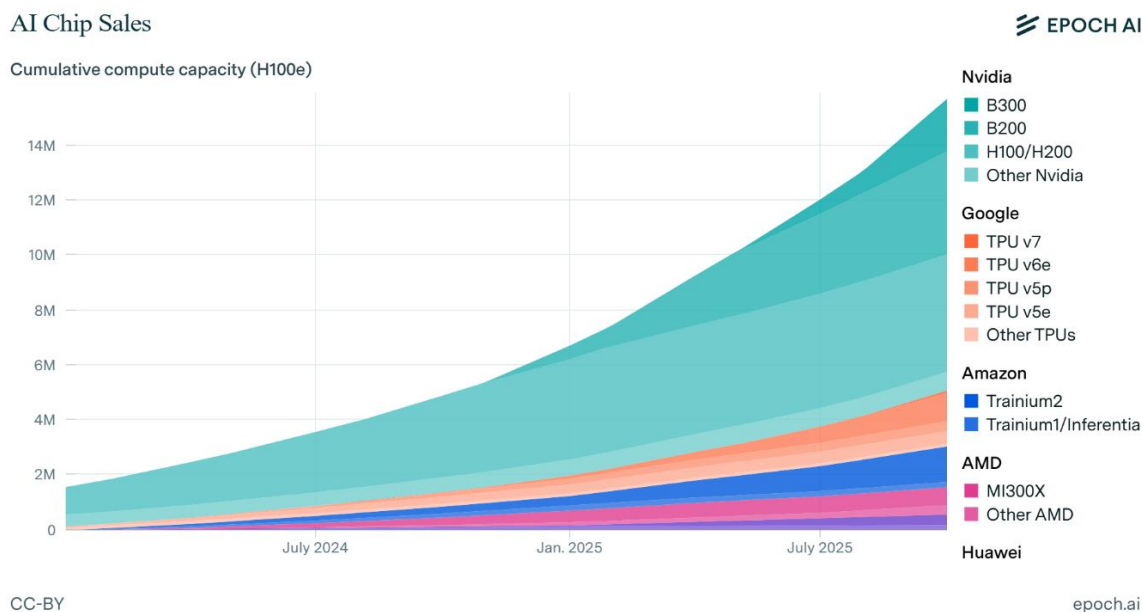
2020年以来全球四大云厂商营业利润率对比



资料来源：罗兰贝格

与此同时，硬件路线变化也会影响 IaaS 盈利中枢。随着推理需求占比提升，云厂对 ASIC 的兴趣明显增强，因为推理侧负载更稳定、标准化程度更高，适合以自研或定制芯片降低单位算力成本。但 ASIC 并非立即替代 GPU，而是形成训练以 GPU 为主、推理向 ASIC 倾斜的并行架构。

图6：目前 GPU 英伟达市场份额仍绝对领先



资料来源：EPOCH AI

综上，Capex 投入短期压制云业务毛利率，中期抬升行业门槛，长期将进一步强化头部厂商在供给、融资与调度上的寡头优势。与此同时，硬件路线变化也会影响 IaaS 盈利中枢。

(2) MaaS 侧，商业模式从模型调用收费逐步演化为 API 引流+PaaS 工具链变现+云资源绑定的盈利路径。当前 MaaS 本身仍普遍承担获客与生态扩张功能，直接利润率并不高，但伴随推理成本下降、企业调用量快速增长、Agent/多模态应用扩散，MaaS 正从 AI 云中的边缘收入项升级为增长最快的核心驱动之一。

2.1、IaaS 进入 AI 驱动的重资产扩张周期

2.1.1、北美四大云厂商资本开支显著抬升且上修幅度持续超预期

北美四大云厂商资本开支显著抬升，且上修幅度持续超预期。2025 年微软、谷歌、亚马逊资本开支合计预计达到 1911 亿美元；2025 年进一步升至 3141 亿美元，同比增长 64%，按各公司指引 2026 年预计进一步抬升至 4975 亿美元。从资本开支结构看，谷歌 2026 年指引中约 60%投向服务器，40%投向数据中心和网络等长周期资产；微软也指出短生命周期资产如 CPU、GPU 服务器，仍将占资本开支约 2/3。

表1：海外云厂商资本开支维持高增

资本开支 (亿美元)	2024Q1	2024Q2	2024Q3	2024Q4	2025Q1	2025Q2	2025Q3	2025Q4	2026Q1
亚马逊	149	176	226	278	250	322	351	395	442
YOY (%)	5%	54%	81%	91%	68%	83%	55%	62%	77%

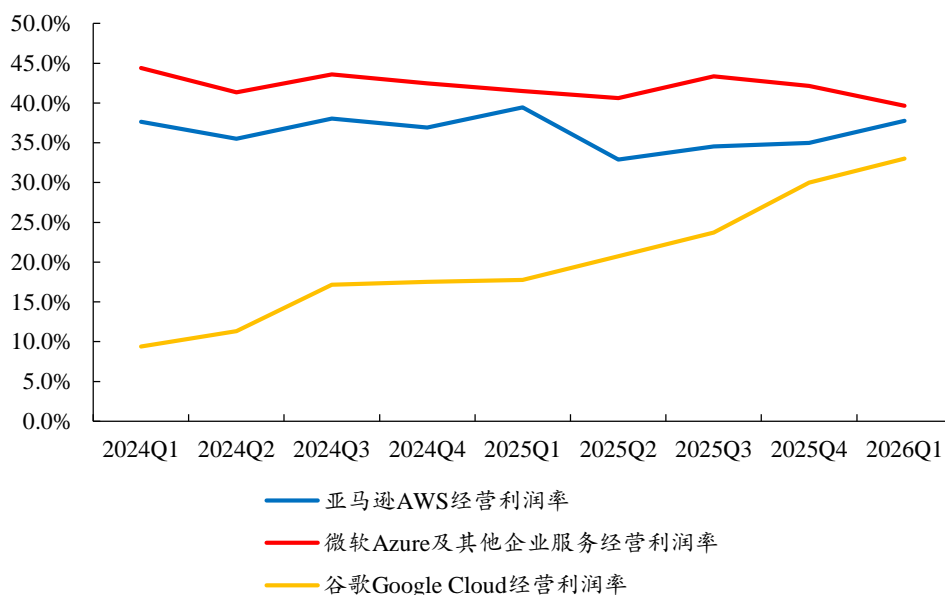
资本开支 (亿美元)	2024Q1	2024Q2	2024Q3	2024Q4	2025Q1	2025Q2	2025Q3	2025Q4	2026Q1
微软	110	139	149	158	167	171	194	375	308
YOY (%)	66%	55%	50%	62%	53%	23%	30%	66%	84%
谷歌	120	132	131	143	172	224	240	280	357
YOY (%)	91%	91%	62%	30%	43%	70%	83%	75%	108%

数据来源：公司公告、开源证券研究所

2.1.2、IaaS 毛利率短期承压，但竞争壁垒同比上升

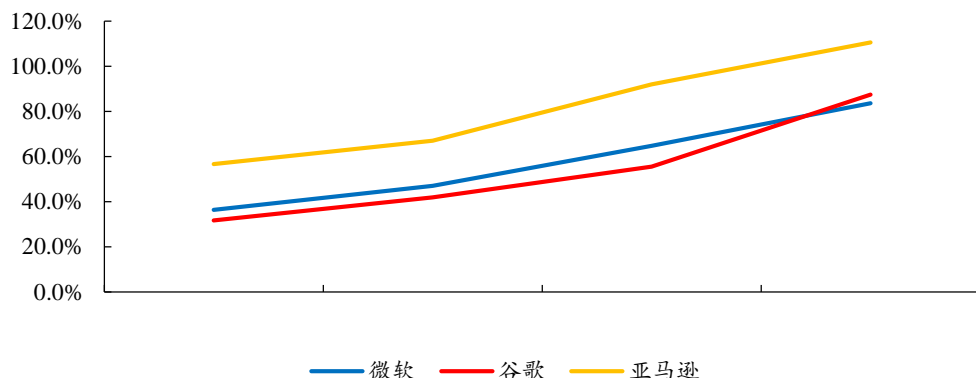
AI 对 IaaS 最直接的影响，是收入增长与成本压力同步上升，前期拉低成熟云平台利润率，尤其是折旧和推理成本尚未被足够收入摊薄时；不同厂商的利润弹性高度依赖其业务结构，尤其是高毛利 PaaS/SaaS 收入占比与 AI 收入转化效率。短期内，大规模 GPU 采购、服务器折旧、网络改造和供电散热升级推高折旧与营业成本，导致云业务毛利率承压；但中期随着利用率提升、推理负载上升和调度效率优化，毛利率有望企稳甚至改善。

图7：云业务利润率随 Capex 投入短期承压长期有望随规模效应企稳并改善



数据来源：公司公告、开源证券研究所

从资本密集度看，亚马逊、微软、谷歌资本开支占经营现金流比重已由 2023 年的 41% 升至 2026 年预测的 83%，这意味着 AI 云基础设施建设正在吞噬传统互联网平台的自由现金流空间，行业资产负债表将承受更高压力，但也将显著抬升中小云厂的进入门槛。

图8：北美云厂商资本开支/经营现金流比值持续上升


数据来源：公司公告、Bloomberg、开源证券研究所

2.1.3、高端 GPU 长期供给偏紧背景下，NeoCloud 崛起

在超大云厂之外，CoreWeave、Nebius 等 NeoCloud 玩家正在重构 IaaS 竞争格局。其商业逻辑是随 AI 从模型训练向大规模推理应用渗透，作为专为 AI 负载打造的云平台，以 GPU 即服务 (GPU as a Service, GaaS) 切入，围绕高 GPU 算力集群、高吞吐网络，以及适配大规模模型训练和推理流程的存储系统，相较于通用云服务，聚焦人工智能专属架构的专业云服务商，往往具备架构精简、GPU 任务性能更优异的优势，同时采用贴合 AI 算力使用场景的定制化计费模式，为模型厂商和大型企业提供更**高性价比**的专用算力服务。2025 年 CoreWeave 收入 51.3 亿美元，同比增长 168%，截至 2025 年底合约积压订单达 668 亿美元，毛利率 71.7%。

基于 Signal65 2026 年 4 月发布的 3 年 TCO 分析报告，CoreWeave 在 AI 工作负载上的成本效率优势主要源于以下 3 个维度：

(1) 核心 GPU 算力成本更低且效率更高。GPU 是 AI 云成本的最大组成部分，CoreWeave 在基础定价和成本结构上均占优：单 GPU 小时价低 36%，NVIDIA H100 GPU 按需定价为 \$6.155/小时，而超大规模云厂商加权平均价为 \$9.606/小时，GPU 成本占 CoreWeave 总合同成本的 78.88%，而超大规模仅为 65.85%。这意味着用户在 CoreWeave 的支出中，更多直接流向最有价值的算力资源，而非被各类附加服务稀释。此外，按等效算力调整 GPU 数量（完成相同工作量），CoreWeave 的 GPU 成本优势从 36% 提升至 48%，总 TCO 优势进一步扩大至 54%。

(2) AI 原生存储架构将存储成本降低 60% 以上。CoreWeave 通过 LOTA（本地对象传输加速器）技术，将高性能缓存直接集成在对象存储层，免费提供与 Lustre 相当的 IO 性能，而超大规模云厂商需额外购买 Managed Lustre (\$0.60/TB/月)，且为满足带宽要求需过度存储容量。

(3) 透明定价模式，附加成本接近于 0，消除预算不确定性。CoreWeave 将 AI 工作负载必需的基础服务全部纳入基础定价，而超大规模云厂商均作为增值服务收费，此外不存在 API 调用费、数据出口费、检索费等难以预测的交易成本。

图9: Coreweave 的成本效率显著高于传统云

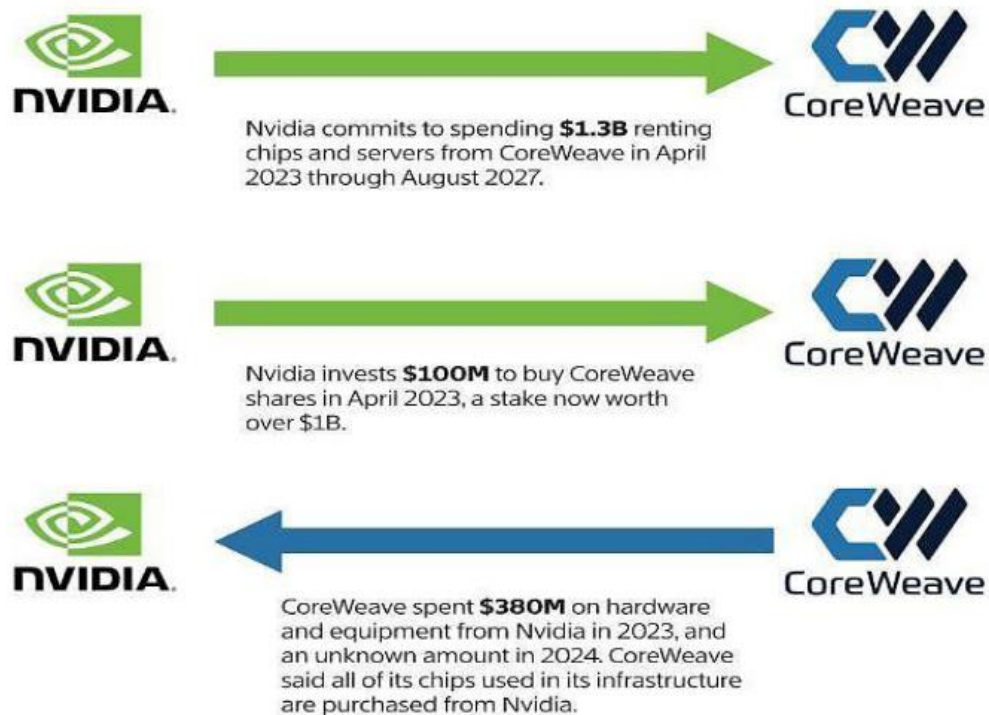

资料来源: Coreweave 官网、Signal65

CoreWeave 的核心壁垒在于 GPU 获取能力+融资能力+万卡集群调度能力的组合。(1) 英伟达通过投资、供货保障、反向租赁等方式深度绑定 CoreWeave, 使其能够率先向市场推出最新一代的 GPU 集群, 并提供比通用云厂商更优的性能和效率: 2023 年 4 月投资 1 亿美元, 并承诺在 2023 年 4 月至 2027 年 8 月期间支付 13 亿美元租回芯片与服务器。(2) 创新融资模式与规模效应驱动的成本优势: 公司开创性地采用与客户长期合同绑定的资产支持融资模式 (DDTLs), 将高信用的客户合同转化为低成本的资金来源。随着业务规模扩大和运营记录的积累, 其融资成本持续显著下降, 这使得其商业模式从单纯现货租赁, 升级为长期保底合同+按需补充的类 IDC 现金流模式。

图10: CoreWeave 与英伟达深度绑定

How Nvidia Funded CoreWeave's Rise

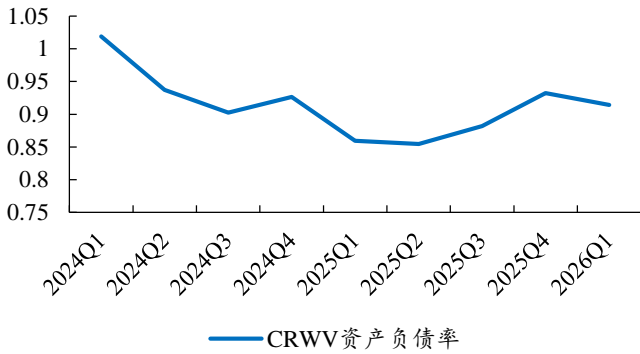
When CoreWeave began experiencing massive growth in early 2023, it wasn't just because it had access to Nvidia GPUs. It was also because Nvidia signed a major sales contract with the former cryptocurrency miner



资料来源: The Information

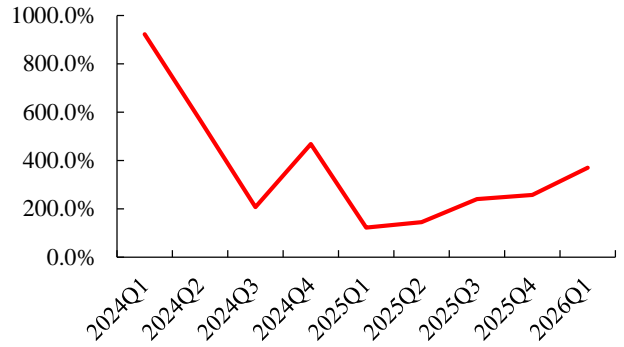
但 NeoCloud 模式也存在明显约束。截至 2026Q1, CoreWeave 资产负债率达 91% 且客户集中度高, 2025 年微软贡献收入占比达 67%, 因此这一模式更像是 AI 基础设施繁荣期的高弹性资产平台, 而非天然稳态高 ROE 模型。对传统云厂的意义则在于, CoreWeave 证明了 AI 时代算力交付可从综合云中拆分出来, 形成更垂直、更高周转的供给形态, 传统云厂必须在标准化公有云与高性能 AI 专属算力之间找到新的产品边界。

图11: Coreweave 资产负债率偏高



数据来源: Bloomberg、开源证券研究所

图12: Coreweave 资本开支占收入比大于 100%

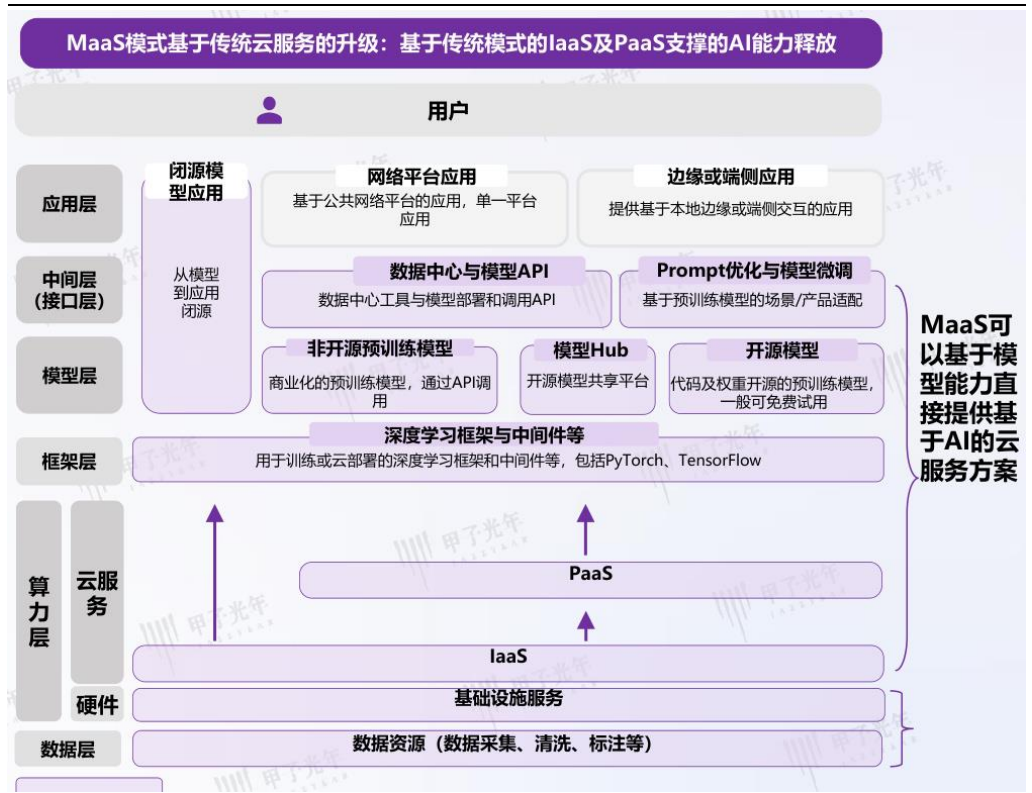


数据来源: Bloomberg、开源证券研究所

2.2、从云资源租赁到 AI 交付, MaaS 重构云商业模式

MaaS 本质是把模型能力封装为 API、私有化部署或混合云服务,按照 Token 调用量、算力消耗或功能模块收费。其价值不只是收费接口,而是通过部署优化、推理框架适配、微调工具链和应用编排能力,降低企业落地门槛,加速 AI 在具体行业中形成可计费场景。从云计算架构看, MaaS 位于 IaaS/PaaS 之上、应用之下,承接底层算力、框架与模型,同时向上输出标准化 API、调优与微服务能力。

图13: MaaS 竞争力高度依赖工程化能力与云侧调度



资料来源: 甲子光年

当前 MaaS 对 AI 云收入占比较低, 调用量的增长远快于收入占比提升, 但是增长最快且具备生态粘性的部分。中国日均 Token 调用量已从 2024 年初的 1000 亿升

至 2025 年底的 100 万亿，并在 2026 年 3 月突破 140 万亿，两年增长超千倍。

海外谷歌 2025 年云业务中 AI 相关收入约 130 亿美元，其中约 100 亿美来自租卡，约 20 亿+美元来自 API 调用，预计 2026 年 MaaS 收入有 4-5 倍增长。推理时代到来后，云厂商的价值不再只是出售 GPU 时长，而是提升单位 token 背后的平台收入密度。一旦客户从单纯租卡，转向在云上完成模型部署、知识库接入、Agent 编排和业务应用开发，云厂商的收入结构就会从低附加值 IaaS，转向更高附加值的平台层与应用层。

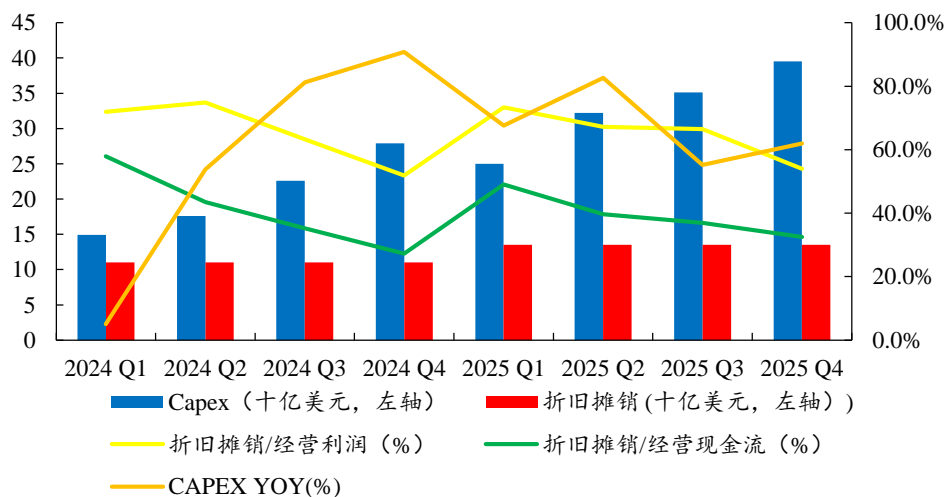
3、海外云复盘：估值与资本开支及收入增速显著相关

3.1、基本面：云计算从 Capex 提速到收入兑现存在约 1-1.5 年时间差

参考历史数据，云计算资本开支增加后约 3 个季度对现金流产生影响，约 5 个季度对收入及折旧摊销费用形成显著影响，原因在于从服务器采购到数据中心建成、资产转固存在时间差，但此次 AI 投资杠杆更高、供需错配度更高、建设周期更长、芯片迭代速度更快、商业模式亦有所变化，或导致收入兑现节奏更慢。

(1) 第一阶段：Capex 先行，资本支出斜率陡增，厂商必须先建设数据中心、采购 AI 芯片。此时，Capex 的增速远高于收入增速。早期会面临巨大的折旧压力。通常服务器折旧周期为 5-6 年，初期毛利会受损，但头部厂商通过延长折旧年限来平滑利润表。

图14：亚马逊 CAPEX 增加后约 3 个季度对现金流产生影响

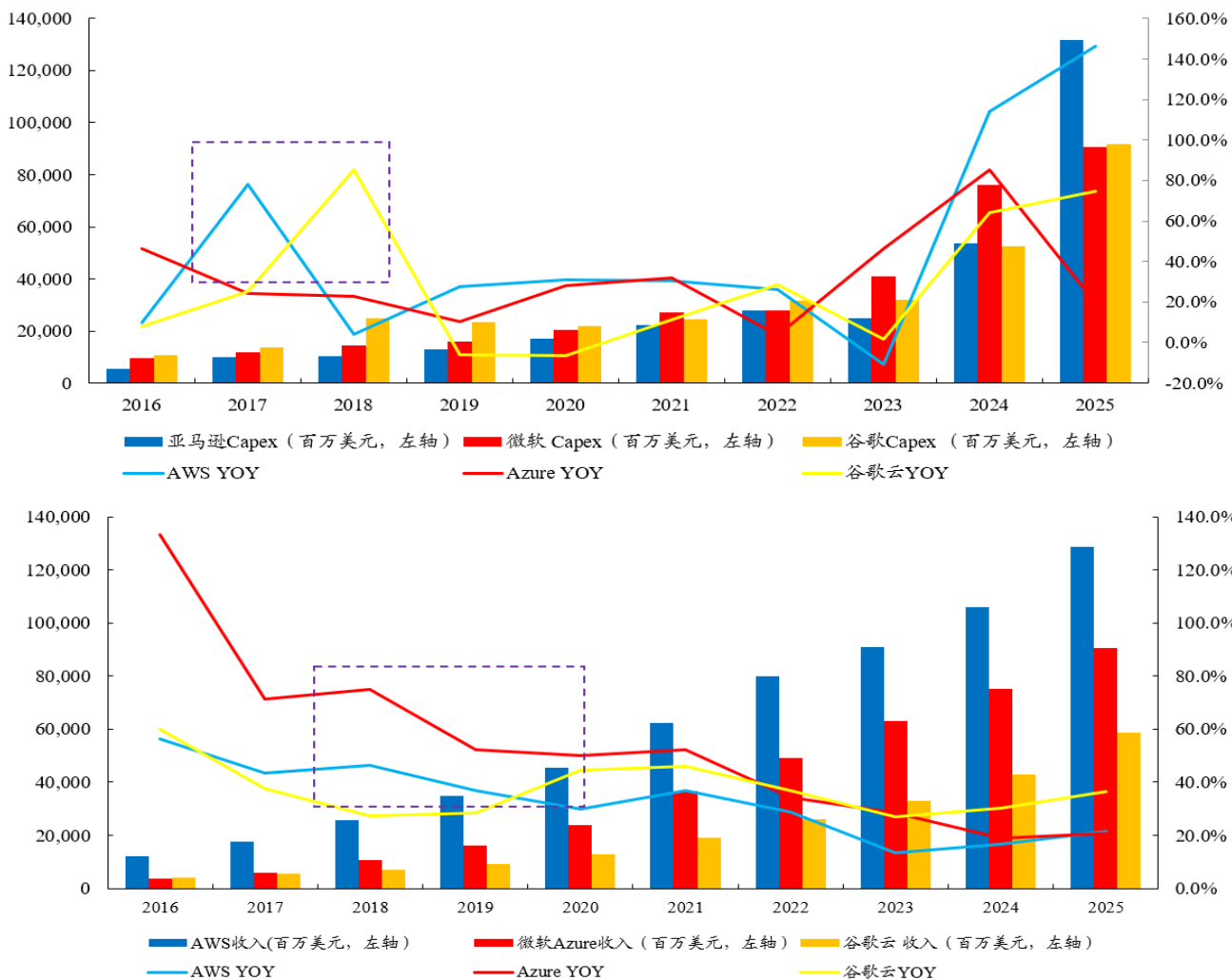


数据来源：公司公告、开源证券研究所

(2) 第二阶段：收入加速，收入增速曲线开始追赶 Capex 曲线，算力转租和 AI 模型 API 服务开始大规模变现。

(3) 利润率扩张：规模效应显现，利用率爬坡：基建填满后，边际成本快速下降；模型层和应用层 (MaaS) 的利润率远高于基础算力租赁；自研芯片如 AWS 的 Trainium 或 Google 的 TPU 大规模替代外购 GPU，显著降低成本。

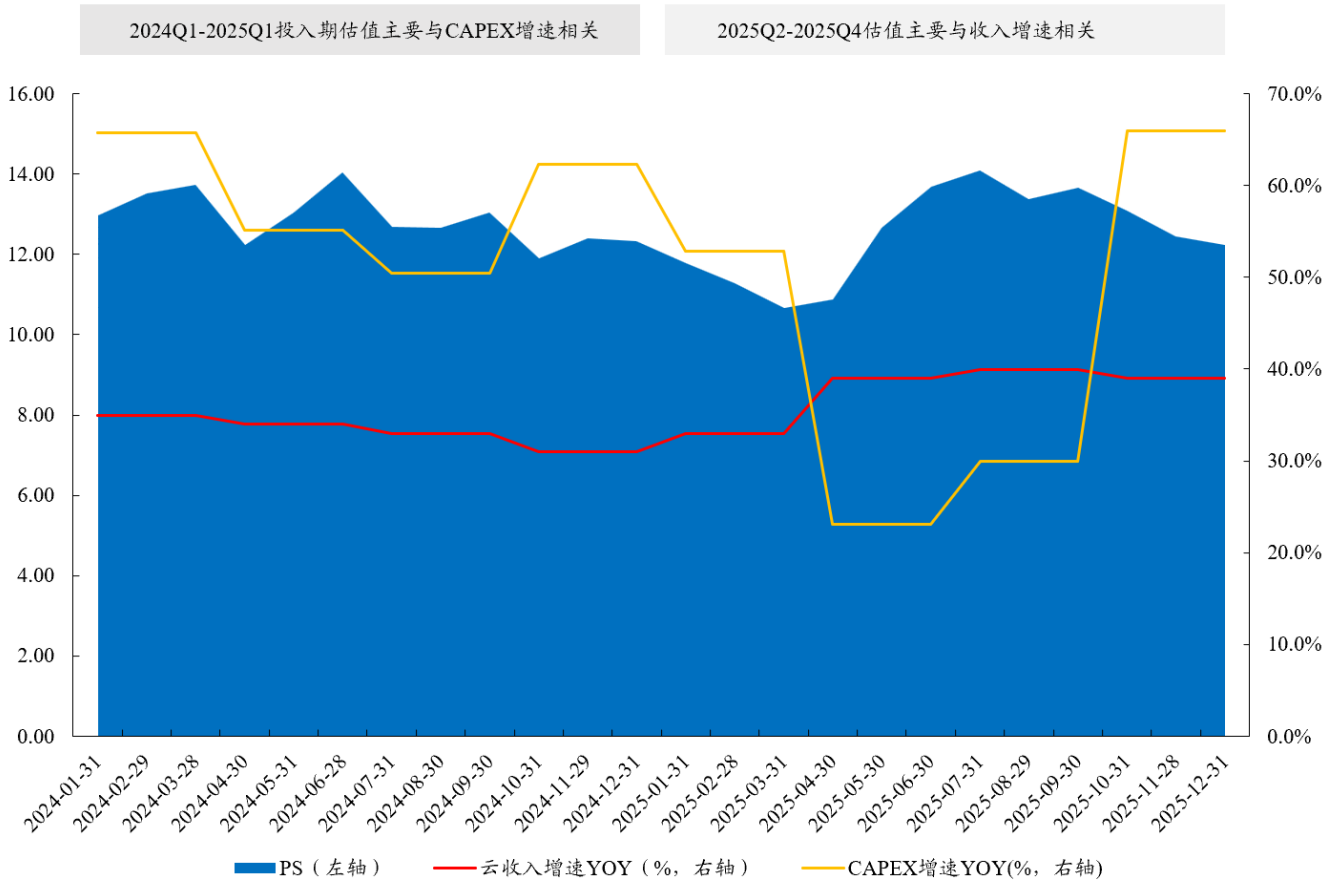
图15：复盘 2015-2020 年，云计算收入增速滞后 CAPEX 增速 1-1.5 年



数据来源：公司公告、开源证券研究所

3.2、估值：与资本开支及收入增速显著相关

复盘微软云过去两年发展过程，估值变化与资本开支和云收入显著相关。(1) 第一阶段：估值与资本开支增速正相关。前期高投入阶段，资本开支加速推动估值上升，加大云基础设施、配合云收入增速的稳定性，市场会强化“投入→未来增长加速”的预期，估值上升、进而放大股价涨幅。后期投入放缓阶段，资本开支增速下降带来估值调整，短期可能引发需求不足的担忧，导致估值阶段性调整、股价波动。(2) 第二阶段：资本开支兑现成业绩，云收入加速、利润率提升、规模效应显现成为估值核心驱动，将进一步抬升估值。

图16: 微软 PS 估值与资本开支及收入增速显著相关


数据来源: Wind、公司公告、开源证券研究所

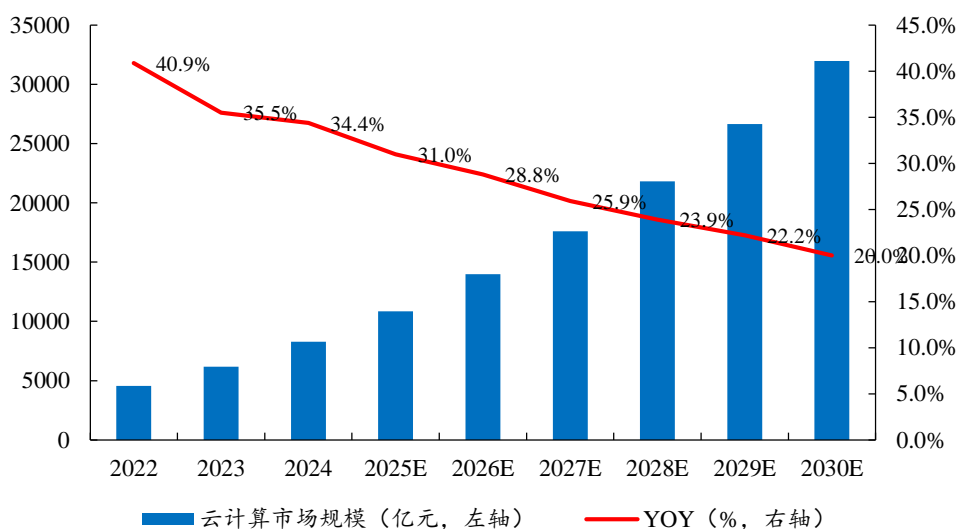
总体而言,当 CAPEX 持续高增且云业务需求未见回落时,市场会逐步从担忧投入过大转向接受超级云平台的长期垄断溢价。IaaS 的核心矛盾是重资产投入与回收周期拉长, MaaS 的核心矛盾是高速增长与低直接利润率并存。但二者结合后,反而构成 AI 云时代最具壁垒的商业模型:前者提供稀缺供给,后者构建需求黏性与价值定价权。**IaaS 层面,重点关注:一是资本开支结构是否从土建转向服务器与短周期资产,二是云业务毛利率能否在 AI 收入放量后企稳, MaaS 层面则关注:调用量增长是否转化为付费客户增长、MaaS 能否带动 PaaS/SaaS 渗透、企业级本地化部署是否成为稳定利润来源。**

4、国内云：AI 驱动增长，竞争格局基本收敛

4.1、国内总量扩容仍将延续，AI 成为行业增长的第一驱动力

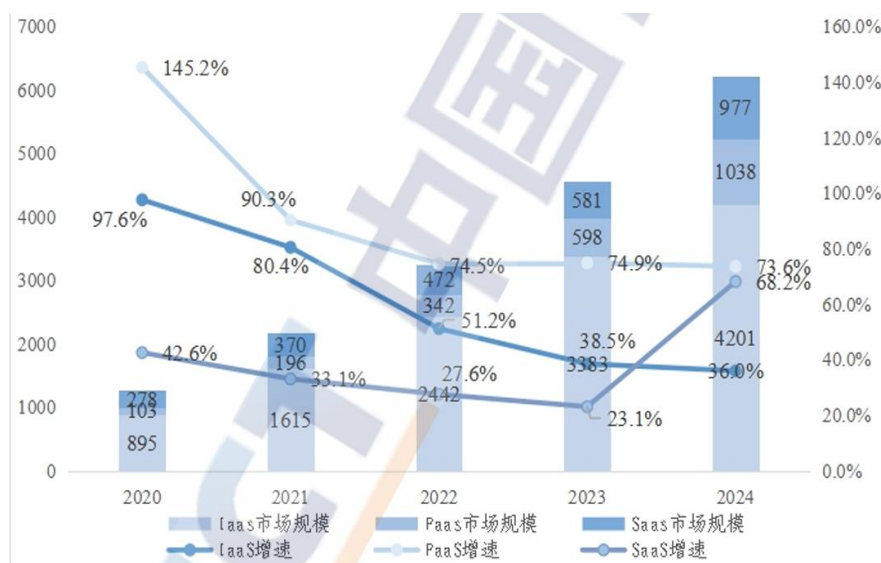
云计算行业整体受 AI 驱动保持增长。据中国信通院统计，2024 年我国云计算市场规模 8288 亿元，同比增长 34.4%，以云计算基座的数智市场格局初步形成。其中，公有云市场规模 6216 亿元，同比增长 36.6%，私有云市场规模 2072 亿元，同比增长 29.3%。从细分领域看，受益企业级 AI 应用及智能体带动，2024 年国内公有云 IaaS 市场规模 4201 亿，YOY+68.2%；AI 开发平台及出海业务带动公有云 PaaS 市场规模突破千亿元。

图17：2024 年中国云计算市场规模 YOY+34.4%



数据来源：中国信通院、开源证券研究所

图18：中国公有云市场规模（亿元）快速增长

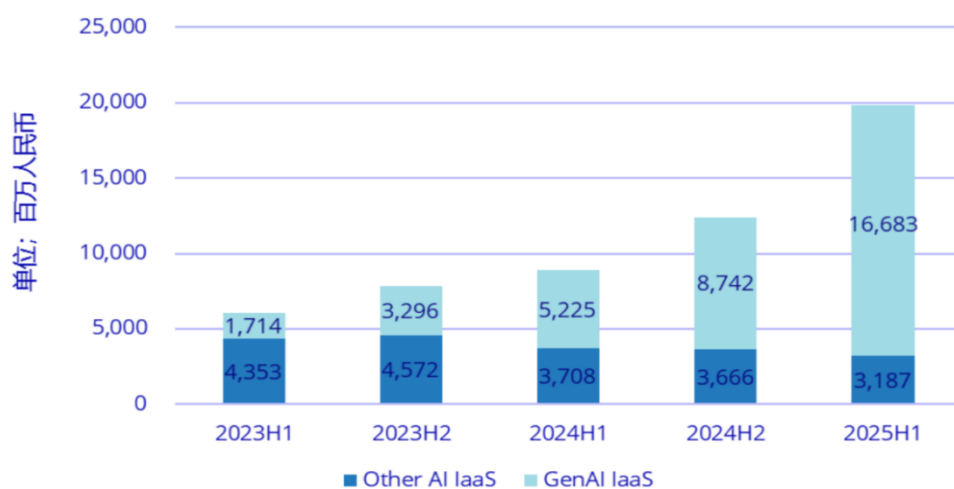


资料来源：中国信通院

从结构看，中国公有云市场仍然偏 IaaS 化。根据 IDC 数据，2025 上半年中国 AI IaaS 整体市场同比增长 122.4%，市场规模达到 198.7 亿元人民币（注：IDC 所定义的智算云基础设施市场是指以 GPU、

AI IaaS 市场，包括面向生成式 AI 场景的 GenAI IaaS 服务市场（服务于 AIGC、大语言模型训练与推理等需求）和面向非生成式 AI 场景的 Other AI IaaS 服务市场（服务于传统机器学习模型训练、渲染仿真、计算机视觉、数据分析等需求）。其中，GenAI IaaS 市场同比增长 219.3%，市场规模达 166.8 亿元人民币；Other AI IaaS 市场同比缩减 14.1%，市场规模达 31.9 亿元人民币。

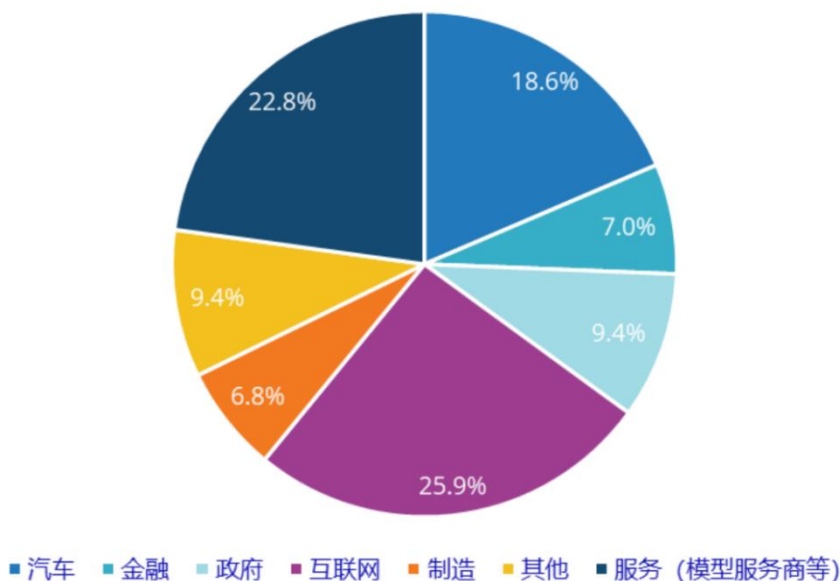
图19：中国 AI IaaS 2025H1 快速增长



数据来源：IDC 中国

- (1) 供给端，自 2024 年起，云服务商对 AI 基础设施的高速资本投入逐步落地，2025H1 通过公有云、边缘云、专属云等多元云服务部署模式，快速触达并服务最终用户，受此影响，2025 年算力市场的资源供给与价格整体稳定，仅伴随小幅波动；
- (2) 市场需求方面，需求端同样表现强劲，2025H1 互联网、汽车、手机制造、金融、泛政府等多个行业，持续追加智能算力及 AI 应用相关支出：其中车企及智驾供应商竞争加剧，不断加码算力投入以抢占智驾方案领先地位；泛政府行业受 2025 年初 DeepSeek 催化，算力支出大幅增长；同时具身智能行业正加速落地，预计其相关算力需求将在 2025 年下半年后分阶段释放。

图20：汽车、金融行业位于中国 Gen AI IaaS 市场需求端前列

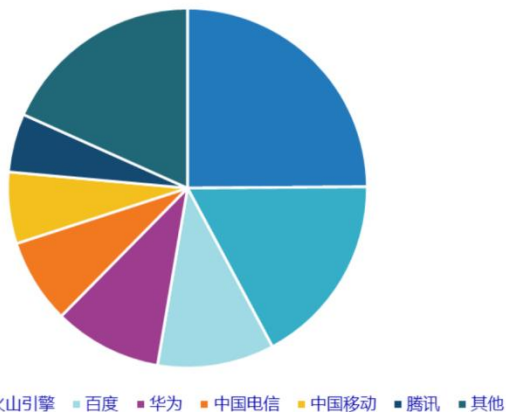


资料来源：IDC

4.2、格局基本收敛，市场集中度进一步提升

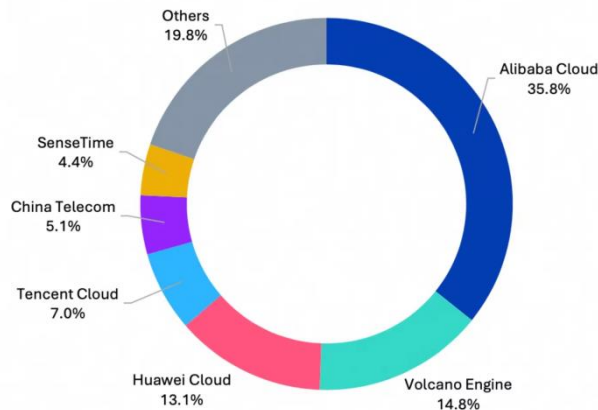
国内 AI IaaS 市场格局已基本收敛，市场集中度进一步提升。(1) 阿里云通过持续加大 AI 基础设施资本支出、加码算力储备，通过提供灵骏 AI 集群、GPU 弹性算力等多元化 AI IaaS 服务，满足客户不同 AI 场景下的旺盛训推需求，市场份额稳居第一；(2) 火山引擎背靠字节跳动算力规模带来的成本优势，搭配相对激进的销售策略，在 AI 市场提供高性价比算力资源；(3) 百度基于昆仑芯片与百舸平台，在算力、网络、训推集群形成技术优势，2025H1 聚焦头部客户算力需求，在汽车及电商、音视频等多个互联网行业有所斩获，进一步扩大市场占比。

图21：IDC 2025H1 中国 Gen AI IaaS 市场中阿里、火山引擎、百度居前



资料来源：IDC

图22：Omedia 中国 AI 云市场 2025H1 市场格局中阿里巴巴、火山引擎、华为云居前



资料来源：Omedia (注：Omedia 口径包含 AI IaaS、AI PaaS、AI MaaS)

5、受益标的

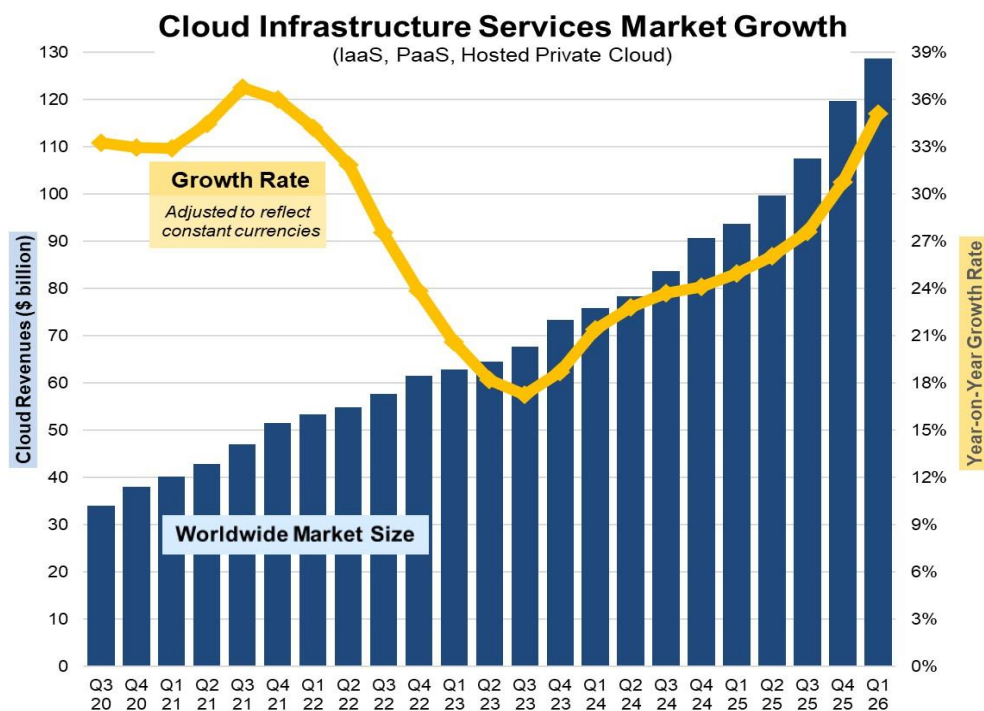
当前云计算行业竞争从传统 IaaS 价格战转向谁能更高效地承接训练和推理需求，并把底层算力能力向模型调用、Agent 开发和行业解决方案持续外溢，核心指标包括 CAPEX 投入强度、全栈技术闭环、AI 收入兑现速度、估值重估弹性。在这一框架下，重点标的可分为两类：(1) 具备全栈 AI 能力的核心平台型公司，代表为微软、谷歌、亚马逊、阿里巴巴、百度集团；(2) 受益于推理需求扩散与生态协同的弹性型云厂商，代表为 CoreWeave、Nebius、金山云。

5.1、海外：传统云龙头走向 AI 全栈，Neocloud 更具弹性

5.1.1、亚马逊、谷歌、微软主导全球云服务市场，AI 需求拉动下微软谷歌更具弹性

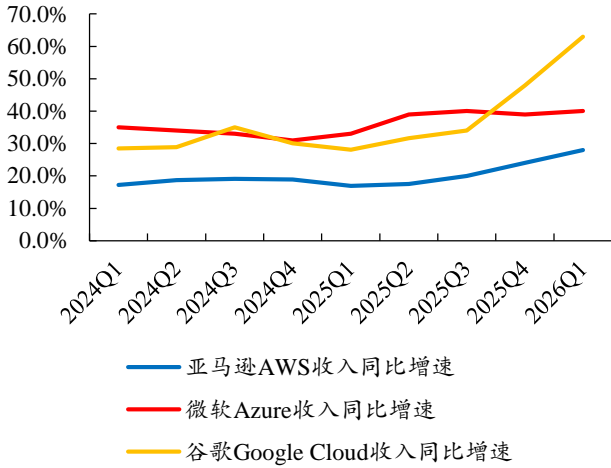
全球云竞争格局仍由 AWS、Azure、Google Cloud 三强主导，但 AI 时代的领先不再只取决于 IaaS 规模，而取决于谁能更快把 GPU、模型、平台工具和应用变现能力整合成闭环。根据 Synergy Research Group 数据，2026Q1 全球云服务商收入规模达 1290 亿美元，同比增长 35%，据此我们估算，AWS 份额 29%，Azure 份额 21%，Google Cloud 份额 16%，AI 需求拉动下谷歌及微软增长弹性强于传统云周期。

图23：2026Q1 全球云服务商收入 YOY+35%



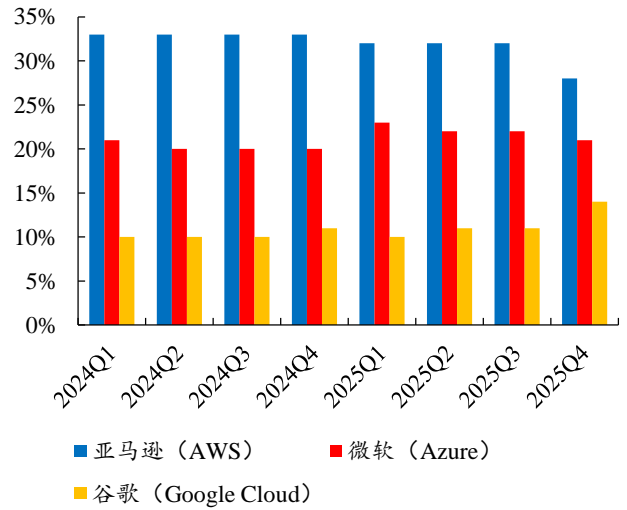
资料来源：Synergy Research Group

图24：微软及谷歌云收入增长更具弹性



数据来源：公司公告、开源证券研究所

图25：谷歌云份额显著提升



数据来源：Canalys、Synergy Research Group、公司公告、开源证券研究所

表2：谷歌自有业务具备 C 端流量入口

		微软	亚马逊	谷歌
应用	自有应用	Microsoft Copilot	Amazon Q	Gemini for workspace
		Github Copilot	Rufus	搜索
	投资公司	Open AI	Anthropic	
工具	平台	Azure AI	Bedrock	Vertex AI
模型		GPT	Titan	Gemini for workspace
算力	自研	Maia 100	Trainium	TPU
	外购		NVIDIA、AMD 等	

资料来源：开源证券研究所

5.1.2、Neocloud：Coreweave 规模更大，Nebius 客户多元化

CoreWeave 与 Nebius 都提供 GPU 云算力、网络与存储能力，也都与 NVIDIA、Microsoft、Meta 保持较深合作，需求端共同受益于 AI 算力紧缺与训练/推理扩张。但两者的战略重心并不相同：CoreWeave 更强调抢规模、抢速度、抢头部客户，Nebius 则更强调客户多元化、企业化拓展与长期经营弹性。

短期看，Nebius 对现货 GPU 价格更敏感；CoreWeave 则拥有更高可见度。Nebius 在 Microsoft 和 Meta 之外更多来自即时或较短合同周期业务，因此在当前 GPU 价格走强阶段，具备更强重定价弹性。Agent 及多模态带动 26 年 Blackwell/Hopper 供需紧张，Nebius 具备极强的下游定价转嫁能力。5 月 21 日，Nebius 将 NVIDIA GPU 租赁价格上调近 30%，包括包括 H100、H200、B200 和 B300 等型号。相比之下，CoreWeave 超过 98% 的业务绑定在多年期合同上，收入可见度更高，但对现货价格上行的弹性较小。从客户结构与盈利路径看，CoreWeave 主要服务最大 AI 支出方，收入集中度较高，目前约三分之二收入来自 Microsoft；Nebius 同样先依靠大客户起量，但导入更多企业客户与 AI 初创公司。

图26: H100 租赁价格自 2026 年 2 月以来持续走高



资料来源: SemiAnalysis

表3: Coreweave 规模更大

	Coreweave	Nebius
数据中心数量	近 50 个, 主要为租赁	截至 2025 年底租赁+自有 7 个, 2026 年预计新上线 12 个
电力容量	2026Q1 签约电力已达 3.5GW	2026 年指引签约电力 4GW, 年底上线电力 800MW-1GW
收入	2026 年指引 120-130 亿美元	2026 年指引 30-34 亿美元
ARR	2026 年指引 310-350 亿美元	2026 年指引 ARR70-90 亿美元
利润	2026 年指引经调整经营利润 9-11 亿美元	2026 年指引经调整 EBITDA margin~40%
主要客户	微软、OpenAI、Meta、英伟达	微软、Meta
CAPEX	2026 年指引 180-190 亿美元	2026 年指引 200-250 亿美元

资料来源: 公司公告、开源证券研究所

从 2026Q1 业绩看, CoreWeave 高收入增长延续, 但交付与利润率压力同步上行。整体看, AI 推理需求快速增长阶段, GPU 租赁商业模式已被验证, 但仍需关注订单交付兑现能否支撑利润率改善及融资成本下行速度。(1) 需求端: 2026Q1 CoreWeave 收入 21 亿美元, 同比增长 112%、环比增长同时新增签约承诺超过 400 亿美元, 积压订单升至 994 亿美元, 环比增长近 50%, 36%预计未来 24 个月确认, 75%在未来 4 年确认。全球四大顶级 AI 模型开发商均已使用 CoreWeave 云, 中国以外 10 大 AI 领导者中有 9 家为客户, 所有 GPU 世代(A100、H100、H200、L40) 平均定价环比均上涨。(2) 基础设施: 2026Q1 活跃功率超过 1GW 里程碑, 新增 400MW 已签约功率, 总签约功率超 3.5GW, 预计 2026 年底达到或超过 1.7GW 活跃功率, 2030 年超过 8GW, 运营近 50 个数据中心。(3) 业绩指引: 2026 全年营收 120-130 亿美元、资本开支上调至 310-350 亿美元, 调整后营业利润 9-11 亿美元, 意味着下半年利润释放斜率或显著高于上半年。

表4: 2026Q1 Coreweave 收入同比高增但营业利润率有所下滑

\$ mn	2024Q1	2024Q2	2024Q3	2024Q4	2025Q1	2025Q2	2025Q3	2025Q4	2026Q1
收入	189	395	584	747	982	1213	1365	1572	2078
YOY	2281%	1225%	677%	544%	420%	207%	134%	110%	112%
成本	59	109	143	182	262	313	369	509	716
研发及基础设施开支	93	183	286	399	561	670	747	950	1273
毛利	37	104	155	166	158	230	248	113	89
YOY			873%	993%	331%	122%	60%	-32%	-44%
毛利率	19.4%	26.2%	26.6%	22.2%	16.1%	19.0%	18.2%	7.2%	4.3%
经营利润	17	78	117	113	-27	19	52	-89	-144
经营利润率	8.9%	19.7%	20.1%	15.1%	-2.8%	1.6%	3.8%	-5.7%	-6.9%
经调整经营利润	25	85	125	121	163	200	217	88	21
YOY			1048%	1809%	550%	134%	74%	-27%	-87%
经调整经营利润率	13.3%	21.6%	21.4%	16.1%	16.6%	16.5%	15.9%	5.6%	1.0%
利息开支	41	67	104	149	264	267	311	388	536
YOY			826%	549%	300%	198%	160%	103%	
% O									
积压订单 (十亿)	15.9	16.2	15	15.1	14.7	30.1	50	66.8	99.4
YOY					-8%	86%	233%	342%	578%
环比增加					-0.4	15.4	19.9	16.8	32.6
CAPEX	1113	2574	2237	2400	1858	2937	1850	8200	6800
YOY					67%	14%	-17%	242%	266%
活跃电力 (MW)				360	420	470	590	850	1000
环比增加					60	50	120	260	150
净债务	1221	2828	2651	6153	6272	9062	11070	14478	21755
环比增加		1607	-177	3502	119	2790	2007	3408	7277

数据来源: Bloomberg、开源证券研究所

Nebius 上调 2026 年 Contracted Power 指引至 >4.0GW, 资本储备充足, 短期算力短缺下受益客户多元化及自建数据中心运营优势。(1) 2026Q1 Nebius 收入 3.99 亿美元, YOY+684%, 高于彭博一致预期 (3.92 亿), 源于产能规模扩张驱动, 同时伴随强劲的定价权和高利用率, 新一代 GPU 定价持续走高, 平均交易规模在 GPU 承诺量增加和合同期限延长的带动下显著增长; 核心指标 ARR 在 3 月底达到 19.2 亿美元, 同比增长 674%; 算力规模化经营杠杆持续释放, 核心 AI 云业务的经调整 EBITDA 利润率达 45%。(2) 资金保障: 2026Q1 成功融资 63 亿美元, 包括来自 NVIDIA 的 20 亿美元股权投资, 期末现金储备达 93 亿美元足以支撑后续 Capex 投入和技术并购。(3) 战略重点: 向上游软件栈延伸, 从算力租赁向全栈生态转变。公司近期收购 Tavily (智能体搜索)、Eigen AI (推理优化) 和 Clarifai (系统级优化), 通过并购深度布局推理加速、Token 目前已在宾夕法尼亚州 (1.2GW) 和芬兰 (310MW) 锁定大型站点, 预计 2027 年运营后, 自有产能占比将超过 75%, 为了有望进一步优化长期单位成本。

表5: Nebius 2026Q1 收入 YOY+621.5%

\$	2024 Q1	2024 Q2	2024 Q3	2024 Q4	2025 Q1	2025 Q2	2025 Q3	2025 Q4	2026Q1
收入 (mn)	11	25	43	38	55	105	146	228	399

\$	2024 Q1	2024 Q2	2024 Q3	2024 Q4	2025 Q1	2025 Q2	2025 Q3	2025 Q4	2026Q1
YOY (%)	-99.5%	-98.9%	-98.0%	-98.6%	381.3%	322.1%	237.4%	500.8%	621.5%
营业成本 (mn)	9	18	19	28	30	30	43	69	104
YOY (%)	-99.2%	-98.1%	-98.0%	-97.8%	231.3%	66.3%	127.0%	149.1%	251.9%
毛利润 (mn)	22	27	56	27	47	71	71	70	295
YOY (%)	-57.2%	-51.8%	-1.0%	-49.0%	107.5%	161.4%	25.4%	154.8%	532.7%
毛利率 (%)	195.7%	109.7%	130.1%	72.4%	84.4%	67.9%	48.3%	30.7%	74.0%
研发费用 (mn)	25	35	34	35	40	43	45	53	67
研发费用率	218.3%	139.4%	79.4%	93.4%	72.3%	40.7%	30.7%	23.3%	16.9%
销售及行政费用 (mn)	51	83	54	90	66	68	90	160	144
销售及行政费用率 (%)	442.2%	333.5%	123.6%	237.5%	119.5%	64.9%	61.3%	70.2%	36.0%
EBITDA (mn)	343	(68)	(52)	(76)	(63)	(21)	(5)	15	130
YOY (%)	96.3%	-122.4%	-118.4%	-121.3%	-118.3%	69.2%	90.0%	119.9%	-306.9%
EBITDA margin (%)	2984.0%	-273.6%	-119.9%	-199.2%	-113.2%	-20.0%	-3.6%	6.6%	32.5%
EBIT (mn)	(82)	(122)	(87)	(112)	(130)	(111)	(130)	(235)	(128)
EBIT margin (%)	-715.6%	-491.8%	-200.9%	-294.7%	-234.2%	-105.8%	-89.1%	-103.0%	-32.1%
净利润, non-GAAP (mn)	160	(73)	(47)	(57)	(93)	(92)	(100)	(173)	(100)
净利率 (%)	1393.0%	-292.1%	-109.2%	-151.1%	-167.3%	-87.1%	-68.7%	-76.0%	-25.1%
EPS, non-GAAP	0.49	(0.20)	(0.26)	(0.18)	(0.39)	(0.38)	(0.40)	(0.68)	(0.32)
YOY (%)	421.7%	-162.0%	-374.5%	-151.6%	-179.1%	-94.3%	-51.1%	-285.8%	-17.5%
现金及现金等价物 (mn)	153	2,326	2,288	2,450	1,447	1,679	4,795	3,678	9,298
YOY (%)	-90.1%	188.8%	158.8%	127.2%	847.5%	-27.8%	109.5%	50.2%	542.6%
CAPEX (mn)	-	162	167	418	544	511	956	2056	2473
YOY (%)	-	-	-	-	-	216.0%	472.2%	392.3%	354.6%
ARR (mn)	-	80	120	90	249	431	551	1250	1900
YOY (%)	-	-	-	-	-	437.5%	359.2%	1288.9%	1288.9%
短期与长期债务 (mn)	12	10	28	50	188	213	4512	4973	8450
净债务	(146)	(3884)	(2260)	(2400)	(1259)	(1467)	(283)	1295	(848)

资料来源: Bloomberg、开源证券研究所

5.2、国内：互联网云厂商有望迎来份额提升

5.2.1、阿里巴巴：具备芯片-云-模型-C端入口的平台龙头

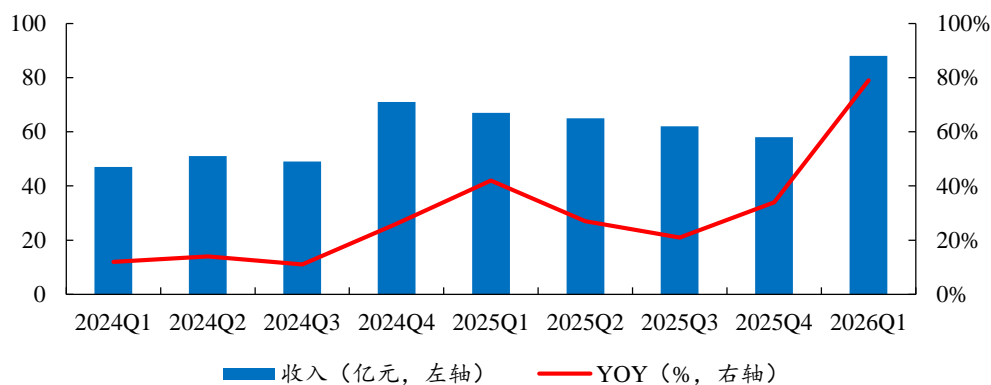
阿里巴巴形成从底层集群到上层应用的统一技术栈,这意味着其 CAPEX 并非孤立投向机房和 GPU,而是能够通过模型调用、Agent 开发、企业部署和生态分发实现多层复用,从而提升单位资本开支的收入产出效率,与谷歌 TPU—Gemini—Vertex

AI—Google Cloud 路径高度相似。管理层明确未来五年云和 AI 商业化收入目标突破 1000 亿美元,2026 年外部收入同比增速有望在未来几个季度持续加速,模型和应用

5.2.2、百度集团：短期搜索业务承压，昆仑芯拆分上市有望驱动估值提升

智能云 MaaS 市场份额提升、昆仑芯受益国产替代趋势。短期搜索业务承压，但随着文心大模型对 C 端应用的重构、智能云在 MaaS 市场的份额提升、昆仑芯在国产算力替代中的战略地位，以及萝卜快跑 (Apollo Go) 在武汉等地商业模式跑通，2026 年有望成为百度 AI 商业化变现的拐点。(1) 百度智能云基础设施 2026Q1 收入 88 亿元，同比增长 79%，GPU 云收入同比增长 184%。根据 IDC 数据，2025 年，中国 AI 应用公有云服务市场规模突破 137 亿元人民币，百度智能云以 30.7% 的市场份额位居第一。(2) 根据 IDC 发布的 2024 年中国加速计算芯片的出货报告，英伟达出货量占比 70% 超过 190 万片。在国产 AI 芯片头部厂商中，华为 64 万片位居第一，其后为昆仑芯 6.9 万、天数 3.8 万、寒武纪 2.6 万、沐曦 2.4 万、燧原 1.3 万片。

图30：百度智能云基础设施收入同比增速加快



数据来源：公司公告、开源证券研究所（注：2025Q3 起收入披露口径有调整）

表6：昆仑芯 2024-2025 年订单及合作项目客户由内部驱动向全行业扩张（不完全统计）

时间	合作公司	项目/核心内容	合作类型	订单规模/关键成果
2024	百度集团 (内部业务)	文心大模型训练/推理、百度搜索、Apollo 自动驾驶、核心业务算力支撑	内部场景落地	P800 芯片支持万卡集群部署
	国芯科技	边缘 AI 计算、车规功能安全 SoC 技术联合研发	技术战略合作	签署《战略合作框架协议》，聚焦汽车电子与边缘计算
	南方电网	电网负荷预测、安全生产监控、能源巡检 AI 解决方案落地	行业场景合作	规模化部署昆仑芯算力设备，覆盖能源电力核心
	吉利汽车	自动驾驶算力支撑、车规级 AI 芯片技术适配	汽车领域合作	交付规模从几十卡到万卡级，助力自动驾驶技术迭代
	vivo	消费电子 AI 算力支撑，适配终端智能应用	消费电子合作	纳入昆仑芯上百家核心客户，实现规模化交付
	中国移动	AI 计算设备（推理型）集采项目，含 3 个标包	运营商集采订单	十亿级订单，多标包高份额中标
2025	中国电信	中国电信研究院联合包括昆仑芯科技有限公司在内的多家产业合作伙伴，完成了业界首次面向大模型推理的异构算力协同	/	2025 年 10 月，中国电信研究院联合多家产业合作伙伴成功完成业界首次基于英伟达算力+国产算力交叉组合的异构算力协同技术验证，旨在通过混合使用不同厂商的算力资源，实现大模型推理

	技术验证		集群的降本增效
招商银行	AI 芯片资源项目，昆仑芯 P800 服务器采购	金融行业算力合作	规模化部署，支持智能客服、多模态数据分析等场景，部分模型推理性能行业领先
格灵深瞳	共建国产化 AI 算力底座，联合开发政务专用 AI 一体机	政务领域战略合作	覆盖智能巡检，支持智能客服、多模态数据分析等场景，部分模型性能行业领先
中国钢研	全栈智能基础设施搭建，专属智算平台建设	工业制造合作	基于昆仑芯 P800 构建算力底座，赋能工业质检与流程优化
同济大学/北京大学	科研算力平台建设、大模型训练与信创教育设备适配	教育科研合作	千万级-亿元级部署，支持高校科研算力普惠化
民生银行	金融大模型全栈架构落地，涵盖算力-平台-模型-应用四层	金融行业深度合作	金融领域首个端到端大模型落地项目

资料来源：光锥智能公众号、开源证券研究所

5.2.3、金山云：小米生态提供确定性，中立定位具备优势

短期受益于 AI 智算业务的快速增长和云服务行业进入涨价周期，中长期小米生态提供强确定性，生态外从传统的云服务商向 AI 算力核心提供商转型。2025 年 AI 相关收入已占公有云收入的近一半，且增速远超传统业务。未来，随着 AI 应用从训练阶段大规模进入推理阶段，对算力的需求将从一次性、项目性转变为持续性、高频次的需求。金山云已前瞻性布局，MaaS 模式具备更强的规模弹性和更高的利润空间。

生态内：作为小米和金山生态唯一的战略云平台，公司深度绑定小米的 AI 战略 7。随着小米在 AI 大模型 (MiMo)、智能汽车、人车家全生态等领域的持续投入，金山云将获得稳定且高确定性的算力订单。2026 年 4 月，金山云公告上调 2026-2027 年为小米集团提供云服务的年度上限 (由 2026-2027 年的 31.4/40.0 亿元上调至 40.0/60.0 亿元)，金山云继续向小米集团提供公有云存储、云计算、算力服务及综合云解决方案等云服务，全面支持其“人车家全生态”战略、智能电动汽车及 MiMo 系列大模型 API 服务。

生态外：金山云的中立第三方定位构筑了独特的竞争优势。与拥有自有大模型的阿里云、腾讯云不同，金山云不与客户 (尤其是其他大模型公司) 构成直接竞争，更能保障客户数据和模型的安全。这一差异化定位使其吸引了包括 Kimi (月之暗面)、智谱、MiniMax 等在内大模型客户。

图31: 2026Q1 金山云收入 YOY+37%

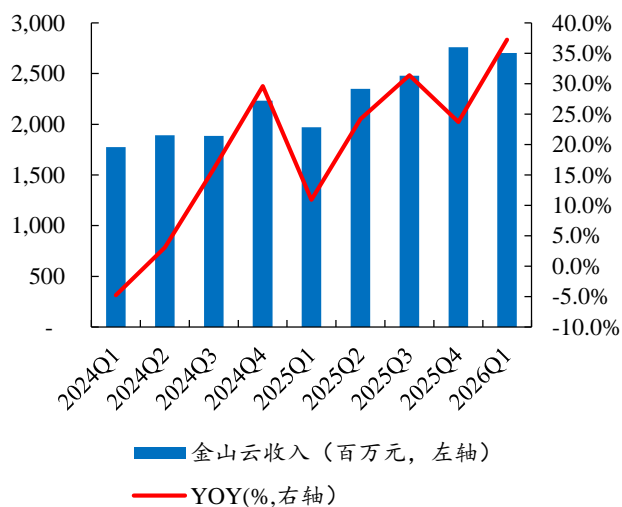
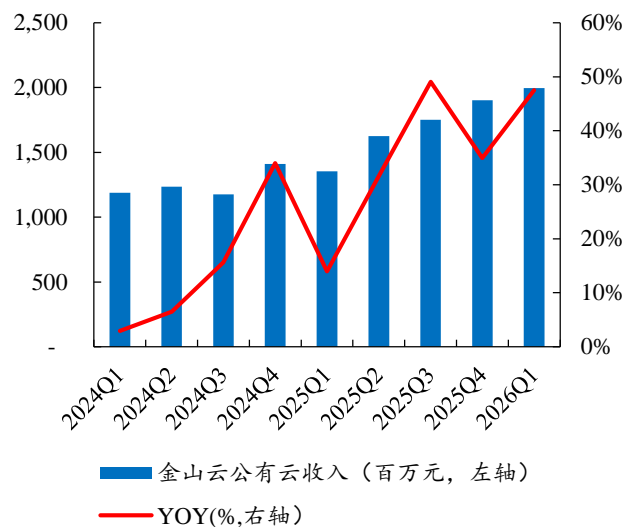


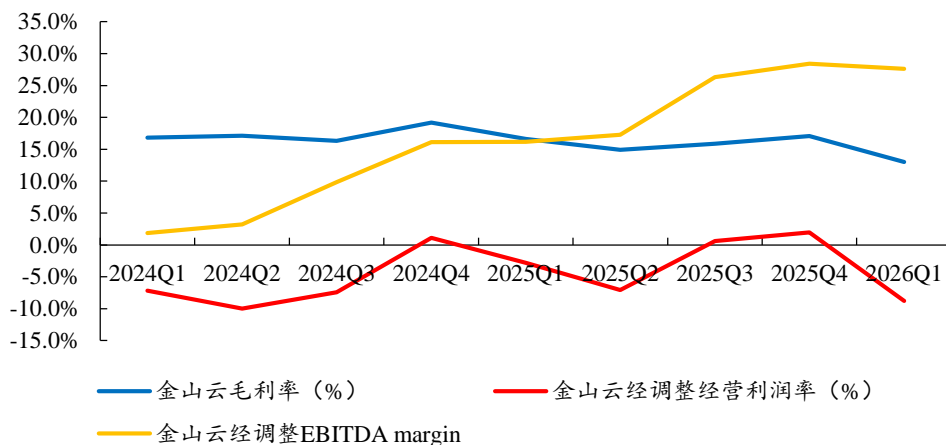
图32: 2026Q1 金山云公有云收入 YOY+47%



数据来源: 公司公告、开源证券研究所

数据来源: 公司公告、开源证券研究所

图33: 公有云占比提升带动金山云 2025Q3 经调整经营利润率转正



数据来源: 公司公告、开源证券研究所

表7：受益标的盈利预测与估值一览

代码	公司名称	货币单位	市值(亿)	收入 (亿)				归母净利润 (亿)				2026E 利润率	估值方法	估值倍数			评级
				2026E	2027E	2028E	CAGR	2026E	2027E	2028E	CAGR			2026E	2027E	2028E	
AMZN	amazon	USD	26,379	8,247	9,331	10,350	12%	1,136	1,283	1,611	19%	13.7%	PE	23.2	20.6	16.4	未评级
GOOG	google	USD	43,902	4,193	5,032	6,008	20%	1,768	1,823	2,159	10%	36.2%	PE	24.8	24.1	20.3	未评级
MS																	未评级
9988.HK	阿里巴巴-W	CNY	19,848	10,865	11,996	14,573	16%	1,087	1,439	1,894	32%	9.9%	PE	18.3	13.8	10.5	买入
9888.HK	百度集团-SW	CNY	2,749	1,313	1,391	1,498	7%	182	205	229	12%	14.7%	PE	15.1	13.4	12.0	买入

代码	公司名称	货币单位	市值(亿)	收入 (亿)				EBITDA (亿)				2026E 利润率	估值方法	估值倍数			评级
				2026E	2027E	2028E	CAGR	2026E	2027E	2028E	CAGR			2026E	2027E	2028E	
CRWV	CoreWeave	USD	559	126	249	391	76%	74	158	249	84%	63.3%	EV/EBITDA	4.4	2.2	1.4	未评级
NBIS	Nebius	USD	553	36	119	207	141%	15	67	129	191%	56.1%	EV/EBITDA	36.4	8.3	4.3	未评级
3896.HK	金山云	CNY	234	124	153	176	19%	36	46	67	37%	29.8%	EV/EBITDA	6.6	5.1	3.5	未评级

数据来源：Bloomberg、开源证券研究所（注：阿里巴巴-W、百度集团-SW 盈利预测来自开源证券研究所，其余公司盈利预测来自Bloomberg 一致预期，数据截至 2026 年 6 月 8 日，1HKD=0.8658CNY，1USD=6.7833CNY）

6、风险提示

宏观经济增长放缓：经济增长放缓导致企业 IT 开支收紧从而影响 AI 需求。

竞争加剧：国内移动互联网用户流量见顶，移动互联网用户数量与时长竞争加剧；云面临海内外巨头竞争压力，客户扩展或不及预期。

监管政策变动：AI 各项监管政策存在不确定性。

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已正式实施。根据上述规定，开源证券评定此研报的风险等级为R4（中高风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为境内专业投资者及风险承受能力为C4、C5的普通投资者。若您并非境内专业投资者及风险承受能力为C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。

因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

分析师声明

本研究报告的署名人员具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告，并对内容和观点负责。本报告清晰准确地反映了署名人员的研究观点，所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。本报告署名人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中的具体的推荐意见或观点有直接或间接的联系。

股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现 20%以上；
	增持（outperform）	预计相对强于市场表现 5%~20%；
	中性（Neutral）	预计相对市场表现在-5%~+5%之间波动；
	减持（underperform）	预计相对弱于市场表现 5%以下。
行业评级	看好（overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡（underperform）	预计行业弱于整体市场表现。

备注：评级标准为以报告日后的6~12个月内，证券相对于市场基准指数的涨跌幅表现，其中A股基准指数为沪深300指数（北交所基准指数为北证50指数）、港股基准指数为恒生指数、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）、美股基准指数为标普500或纳斯达克综合指数。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取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动，过往的业绩表现不应作为其日后表现的预示。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。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户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。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，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。投资者应自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。未经本公司事先书面授权，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，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

开源证券研究所

上海

地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层
邮编：200120
邮箱：research@kysec.cn

深圳

地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

北京

地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

西安

地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn