长江证券 CHANGJIANG SECURITIES

证券研究报告



行业研究丨深度报告丨软件与服务

Agentic Cloud：云计算的下一次范式迁移

请阅读最后评级说明和重要声明

报告要点

云计算的本质是 Runtime：回顾云计算发展历程，每一次计算范式的迁移都会推动云基础设施和产品形态的重构：从 PC 互联网时代的虚拟化基础设施，到互联网时代面向应用开发的平台云，再到 AI 时代面向模型训练与推理的智能计算云，云的价值始终围绕“如何更高效地承载新的计算主体和软件形态”展开。进入 Agentic 时代，软件运行主体正在从传统应用程序转向具备自主规划和执行能力的智能体，Agent 将成为云环境中的新型运行主体。国内外核心云厂商正加速布局 Agent 基础设施，推动云计算从传统资源供给模式向智能体运行平台演进。

分析师及联系人



杨洋

SAC：S0490517070012

SFC：BUW100



郭敬超

SAC：S0490525120002



卢之晗

SAC：S0490525120001

| 证券研究报告 |

2026-08-31

行业研究 | 深度报告

投资评级 看好 | 维持

Agentic Cloud：云计算的下一次范式迁移

当 Agent 成为云的新用户，Agentic 云时代开启

互联网流量正在从 Human-centric 向 Machine-centric 迁移。过去互联网的核心交互模式是“人访问网站”：用户通过浏览器主动搜索、浏览和交易，网站围绕人的视觉交互和点击行为进行设计。然而，近年来自动化流量快速增长，互联网正在逐渐演变为由人类和机器共同参与的生态。根据 Imperva（Thales）连续多年的《Bad Bot Report》，全球互联网流量中自动化访问占比持续提升：2023 年 Bot 流量已经接近一半（49.6%），2024 年进一步超过人类访问，占比达到 51%；最新 2026 版报告显示，2025 年自动化流量占比进一步提升至 53%，人类访问下降至 47%。我们认为，种种数据证明未来互联网基建需要与时俱进，从“以人为中心”的交互体验优化，转向“以 Agent 为中心”的运行效率保障。

新的 Agentic 范式下，云计算产品迭代与商业模式更新需齐头并进

Agentic 云时代，网站需要新的“保安”和“收银员”。过去云计算主要面向开发者和应用程序提供弹性计算、存储和数据库等基础资源，商业模式以资源使用量（CPU/GPU 小时、存储容量、API 调用次数）计费；但在 Agent 时代，云的服务对象将逐渐从“人和应用”转向“智能体”，云平台不仅需要提供底层算力资源，更需要提供支撑 Agent 持续运行的 Runtime 环境，包括任务编排、工具调用、身份认证、状态管理、安全治理和多 Agent 协同等能力。与此同时，新的 Agentic 云的商业模式也将从资源计费逐步向任务完成价值、Agent 调用次数和智能服务结果付费迁移。未来云厂商的竞争不再只是比拼算力规模，而是比拼谁能够更高效、更安全地承载大规模 Agent 运行，并将基础设施能力转化为企业智能化生产力。

云的本质是 Runtime，历史上每一次范式变革都将带来云计算的更迭

云计算的本质是 Runtime，即为不同范式下的软件运行提供基础环境。回顾云计算发展历程，每一次计算范式的迁移都会推动云基础设施和产品形态的重构：从 PC 互联网时代的虚拟化基础设施，到移动互联网时代面向应用开发的平台云，再到 AI 时代面向模型训练与推理的智能计算云，云的价值始终围绕“如何更高效地承载新的计算主体和软件形态”展开。进入 Agentic 时代，软件运行主体正在从传统应用程序转向具备自主规划和执行能力的智能体，Agent 不再只是调用云资源的应用，而将成为云环境中的新型运行主体。伴随 Agent 逐渐成为企业智能化应用的新型入口，国内外核心云厂商如 AWS 等正在加速布局 Agent 基础设施与运行平台，推动云计算从传统资源供给模式向智能体运行平台演进。

风险提示

1、 大模型发展不及预期风险；

2、 大厂转型速度不及预期风险；

3、 竞争加剧风险；

4、 上游硬件涨价幅度超预期风险。

市场表现对比图(近 12 个月)



资料来源：Wind

相关研究

- 《DeepSeek 涨价，关注国产大模型及国产算力 2026 年第 27 周计算机行业周报》2026-07-14
- 《AI 应用进行时：Claude Science 开启科研工作流新范式，AI4S 场景有望加速渗透》2026-07-12
- 《脑机接口医疗器械迎“硬标准”，产业有望进入规范化发展阶段》2026-07-06

更多研报请访问
长江研究小程序

目录

人类之外，Agent 已经成为增长最快的新型流量来源 6
AI Bot 访问快速增长，成为重要的网站新增访问主体 6
企业 AI 正在经历从“助手”向“Agent”的范式迁移，未来将带动 Agent 流量大幅上升 7
全新 Agentic 云范式下，产品迭代+商业模式进化迫在眉睫 8
产品重构：Agent 对云服务的需求与传统云存在系统性差异 8
商业模式：注意力分配机制失效，Agent 时代网站需要新的“保安”与“收银员” 9
云的本质是 Runtime，每一轮范式迁移都伴随着产品演绎 12
云 Runtime 演进：从代码执行到任务闭环 12
行业大势之下，国内外核心云厂商纷纷开启 Agentic 云布局 14
风险提示 24

图表目录

图 1：2025 年人类与 Bot 访问网页请求占比（单位：%） 6
图 2：2025 年 AI 爬虫行为增长超过 15 倍 7
图 3：2026-05-25 至 2026-06-22 统计周期内 AI 爬虫主要用途 7
图 4：截至 2025 年底真正规模化部署 Agent 的企业比例有限 8
图 5：Gartner 的调研显示仅有 15%企业 IT 负责人正在考虑、试点或部署完全自主 Agent（2025.9） 8
图 6：AI Agent 市场规模预测（亿美元） 8
图 7：2025 年硬广投放费用占比 top 媒介类型 9
图 8：2025 年硬广投放费用占比 TOP 媒介 9
图 9：AI Agent 执行用户任务，用户不访问源网页 10
图 10：AI 执行用户任务，用户访问源网页 10
图 11：注意力机制失效 11
图 12：Agent 时代三种商业模式 11
图 13：传统 Runtime 能力对比 Agent Runtime 能力 13
图 14：2018 年全球 SaaS、IaaS、PaaS 占比 13
图 15：2022 年全球 SaaS、IaaS、PaaS 占比 13
图 16：Google Cloud、Azure、AWS 经营利润率 14
图 17：Cloudflare 营收（亿美元）-股价（美元）复盘及云阶段发展历程 16
图 18：Google Cloud 营收（亿美元）-谷歌股价（美元）复盘及云阶段发展历程 18
图 19：阿里云营收（亿美元）-股价（美元）复盘及云阶段发展历程 19
图 20：阿里云产品生态 20
图 21：AWS 营收（亿美元）-亚马逊股价（美元）复盘及云阶段发展历程 22
图 22：Azure 营收（亿美元）-微软股价（美元）复盘及云阶段发展历程 23

表 1：Agent 云供需缺口 9
表 2：Bot 管理与访问控制需求 11
表 3：云发展三阶段 14

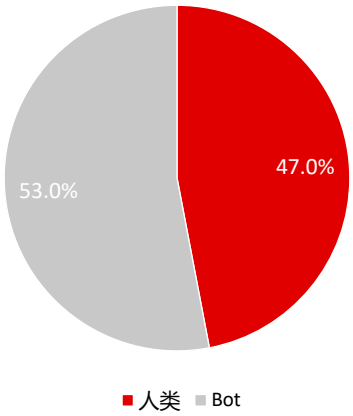
表 4：Cloudflare 边缘 Agent 时代重要事件 15
表 5：Google Cloud Agent 时代重要事件 17
表 6：阿里云 Agent 时代重要事件 19
表 7：AWS Agent 时代重要事件 21
表 8：Azure Agent 时代重要事件.....23

人类之外，Agent 已经成为增长最快的新型流量来源

AI Bot 访问快速增长，成为重要的网站新增访问主体

互联网流量正在从 Human-centric 向 Machine-centric 迁移。过去互联网的核心交互模式是“人访问网站”：用户通过浏览器主动搜索、浏览和交易，网站围绕人的视觉交互和点击行为进行设计。然而，近年来自动化流量快速增长，互联网正在逐渐演变为由人类和机器共同参与的生态。根据 Imperva（Thales）连续多年的《Bad Bot Report》，全球互联网流量中自动化访问占比持续提升：2023 年 Bot 流量已经接近一半（49.6%），2024 年进一步超过人类访问，占比达到 51%；最新 2026 版报告显示，2025 年自动化流量占比进一步提升至 53%，人类访问下降至 47%，同时在过去的 12 个月里，人工智能驱动的机器人活动增加了十倍以上，每天被拦截的请求从 200 万增加到 2025 年的 2500 万。

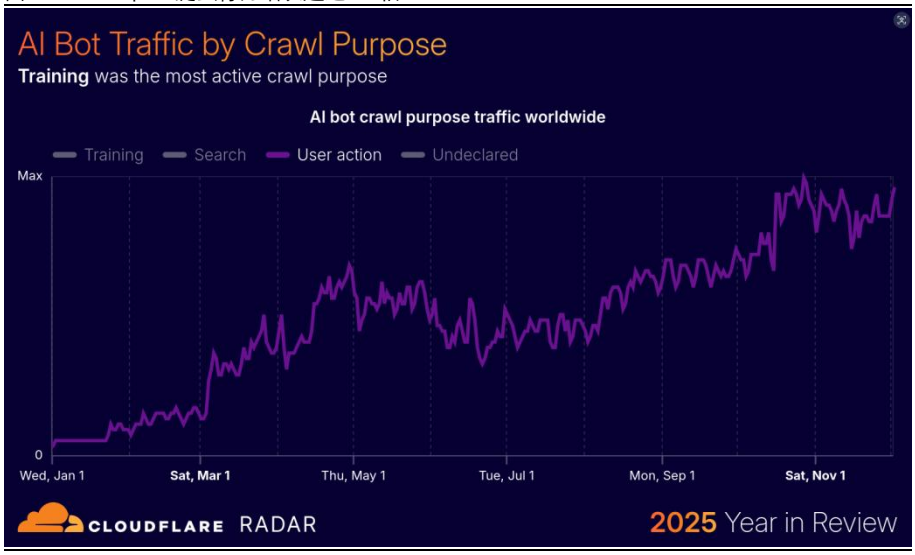
图 1：2025 年人类与 Bot 访问网页请求占比（单位：%）



资料来源：《Bad Bot Report》Thales（2026），长江证券研究所

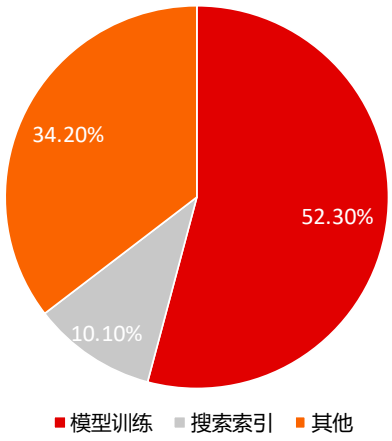
Cloudflare 数据显示，人类之外 AI Bot 正在成为网站新增访问主体，互联网正在迎来“Agent 消费者时代”。Cloudflare 数据显示，AI Bot 访问正在快速增长，其中面向用户任务执行的 AI crawling（AI 爬虫）在 2025 年增长超过 15 倍。未来，Agent 可能替代人类成为互联网服务的主要调用者：它们不再浏览网页，而是直接调用网站、API 和数字服务完成任务。从具体的用途看，AI 爬虫的主要用途分为模型训练、搜索索引等方向。

图 2：2025 年 AI 爬虫行为增长超过 15 倍



资料来源：Cloudflare，长江证券研究所

图 3：2026-05-25 至 2026-06-22 统计周期内 AI 爬虫主要用途



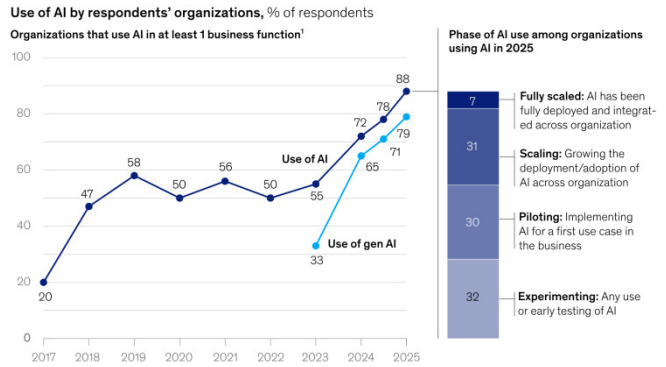
资料来源：TechnologyChecker，长江证券研究所

企业 AI 正在经历从“助手”向“Agent”的范式迁移，未来将带动 Agent 流量大幅上升

随着企业侧的 AI 使用逐渐从单点应用转变为系统重塑，Agent 的使用量可能在届时迎来大幅增长。海外数据看，企业 AI 正在经历从“助手”向“Agent”的范式迁移，但目前仍处于早期阶段。虽然生成式 AI 已经快速渗透企业流程，McKinsey 数据显示 2025 年底 88%的企业已在至少一个业务环节使用 AI，但多数企业仍处于试验和局部应用阶段，真正规模化部署 Agent 的比例有限。当前企业 AI 主要集中于信息检索、内容生成和辅助决策，而能够自主规划任务、调用外部工具并跨系统执行工作的 Agent 仍未普及。Gartner 数据显示，2025 年仅 15%的企业 IT 负责人正在考虑、试点或部署完全自主 Agent，而具备任务型 Agent 能力的企业应用在 2025 年仍不足 5%。

因此，未来随着 Agent 从“辅助工具”向“数字员工”演进，围绕 Agent 运行所需的 Runtime、调度、安全、数据连接和服务编排能力，将成为云计算新的基础设施需求。

图 4：截至 2025 年底真正规模化部署 Agent 的企业比例有限



资料来源：McKinsey，长江证券研究所

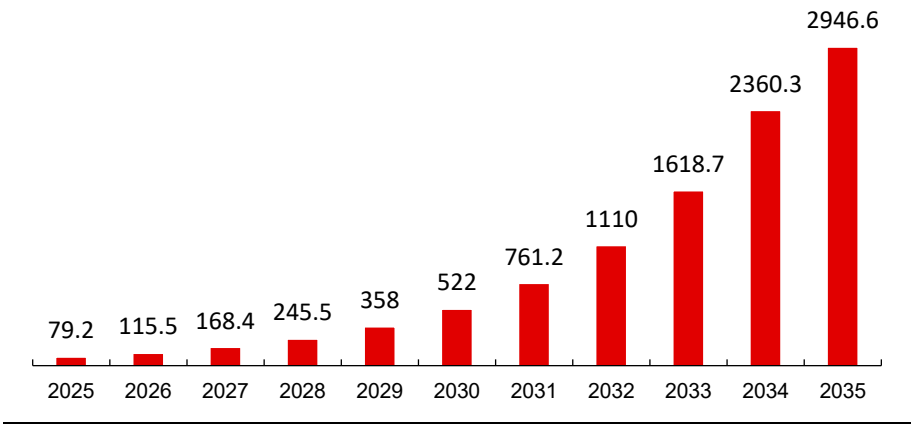
图 5：Gartner 的调研显示仅有 15%企业 IT 负责人正在考虑、试点或部署完全自主 Agent（2025.9）



资料来源：Gartner，长江证券研究所

随着企业 Agent 的渗透率逐渐提升，市场规模预计呈爆发式增长。 根据 Precedence Research 预测，全球 AI Agent 市场规模将从 2025 年的 79.2 亿美元跃升至 2035 年的 2946.6 亿美元，复合年增长率达 43.57%。当前可见的 Agent 的变现模式包括工具、算力、效果三种交付方式，以面向不同客群获得利润。

图 6：AI Agent 市场规模预测（亿美元）



资料来源：Precedence Research，长江证券研究所

全新 Agentic 云范式下，产品迭代+商业模式进化迫在眉睫

产品重构：Agent 对云服务的需求与传统云存在系统性差异

企业侧当前对于全面布局 Agent 存在迟疑，除去模型能力之外，同时也会考虑到基建的不足和数据安全等问题。Google Cloud 的一项调研显示，83%的企业认为现有基础设施无法满足生产级自主 AI 系统需求，同时约 80%的企业将安全、治理和 MLOps 视为 Agent 规模化部署的重要挑战。

Agent 对云服务的需求与传统云存在系统性差异，供需缺口正在倒逼 Runtime 架构重构。 Agent 的调用模式在工具发现、状态管理、安全模型、可观测性与多 Agent 协作五个维度上全面超出传统云的设计假设。供需缺口正在倒逼云架构重构。Agent 云要求机器可发现的工具、可持久化的状态与可审计的执行链，云厂商必须在产品层重建一套面向 Agent 的原生能力体系，而非在既有产品上叠加功能模块。

表 1：Agent 云供需缺口

需求维度	传统云	Agent 云
工具发现与调用	人类读文档，硬编码 API 集成	Agent 自主发现、语义化理解、机器可读的工具描述
状态管理与记忆	无状态、请求—响应即时模式	原生持久化：跨会话记忆、向量存储、上下文维持
安全模型	按“角色”粗粒度权限	面向 Agent 的身份体系、细粒度工具权限、操作审计链、沙箱隔离
可观测性	应用级日志	平台级 Agent 调用链路追踪、推理过程可视化
多 Agent 协作编排	单体应用	规划 Agent、执行 Agent、审核 Agent 之间的通信、调度、冲突解决

资料来源：51CTO，稀土掘金，github，博客园，CSDN，长江证券研究所

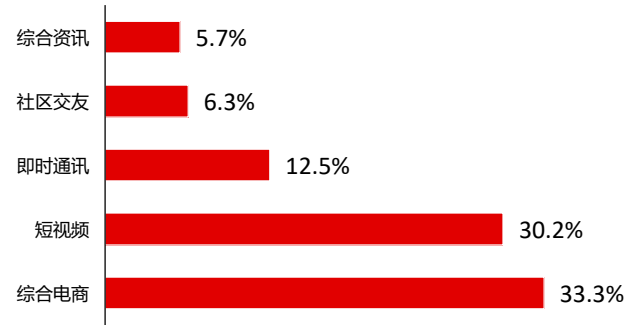
商业模式：注意力分配机制失效，Agent 时代网站需要新的“保安”与“收银员”

传统互联网的商业模式构建在注意力分配机制之上

在获取流量增长的背景下，“注意力—广告”模式是过去二十年互联网内容变现的底层逻辑，其定价根基在于注意力时长的稀缺性。 用户通过搜索发现信息、访问源网页并停留，广告曝光由此产生，内容创作者据此获得收益。根据 Grand View Research 数据，全球数字广告市场在 2025 年达到约 5679 亿美元规模，其竞价机制建立在用户注意力稀缺且可量化分配的假设之上；用户在平台上停留的时间越长、定位越精准，广告价格越高。

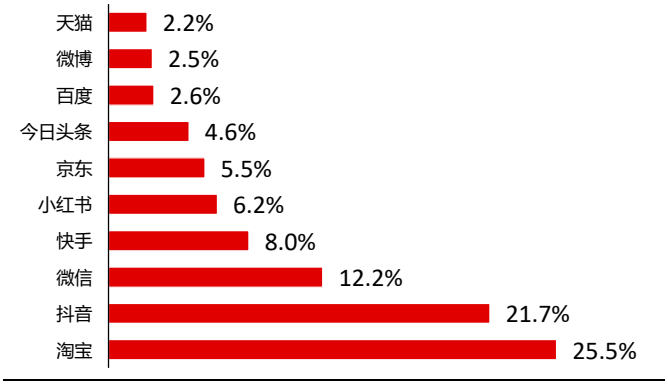
注意力时长直接决定广告库存深度与变现效率。YouTube 观众平均单次视频停留 8 至 12 分钟，TikTok 约 30 秒；更长的观看时长意味着单场会话可承载包括前插播、中插播及结尾广告在内的 3 至 5 个广告位，形成显著的广告库存优势。由此，拥有 10 万订阅者的 YouTube 创作者在金融科技、B2B SaaS 等垂直领域的收入，可能超过百万粉丝的 TikTok 账号。

图 7：2025 年硬广投放费用占比 top 媒介类型



资料来源：QuestMobile，长江证券研究所

图 8：2025 年硬广投放费用占比 TOP 媒介



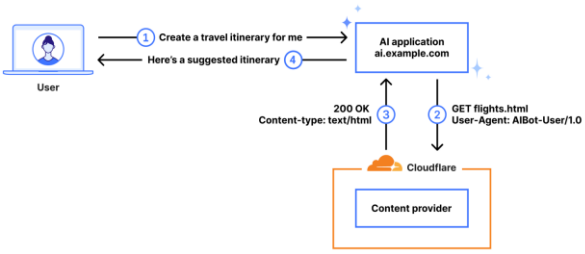
资料来源：QuestMobile，长江证券研究所

Agent 时代网站零点击率攀升，注意力机制面临失效

流量分发弱化，传统搜索引擎作为流量分发中枢的角色正被系统性解构。传统搜索引擎谷歌曾长期维持全球搜索市场 90%以上的份额，2024 年第四季度谷歌搜索份额首次跌破 90%至 89.34%，为 2015 年以来首次，此后持续徘徊于 90%。Gartner 预测，2026 年传统搜索引擎流量规模将萎缩 25%，份额被 AI 及 Agent 取代。据 HUMAN Security 报告显示，2025 年 AI Agent 及 Agentic Browser 流量同比增长 7,851%；全部自动化流量同比增长 23.51%，约为人类流量 3.10%增速的 8 倍。

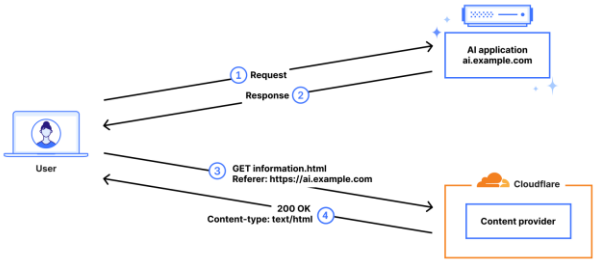
点击与访问减少，基于页面停留时长的广告曝光机制与内容创作者之间的商业闭环面临断裂，注意力机制失效。截至 2026 年 7 月，Cloudflare Radar 数据显示，Anthropic 抓取-回访比为 4,580:1，OpenAI 为 848:1，Perplexity 为 186:1，而 Google 传统搜索仅为 5:1。超过 50%的 AI 爬虫活动用于模型训练；在有用户意图的搜索索引中，AI Bot 大规模抓取内容后直接生成合成答案，点击与访问环节丢失。

图 9：AI Agent 执行用户任务，用户不访问源网页



资料来源：Cloudflare，长江证券研究所

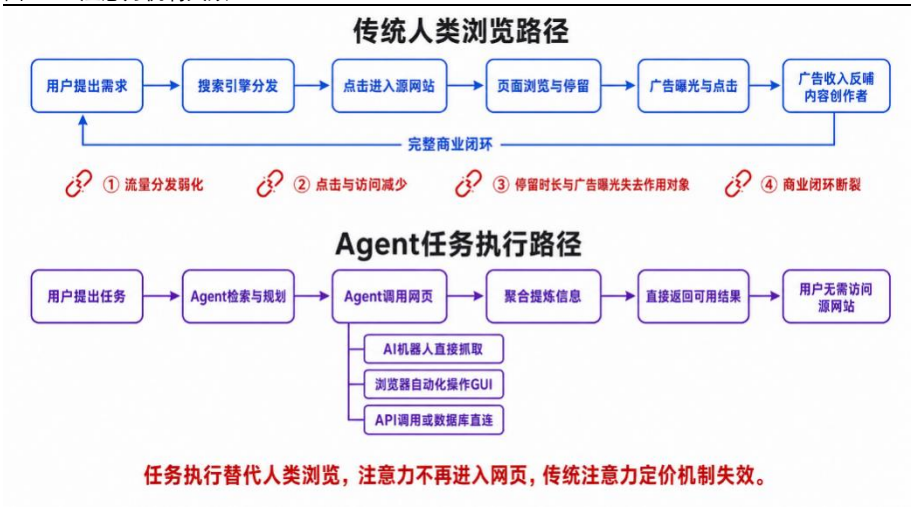
图 10：AI 执行用户任务，用户访问源网页



资料来源：Cloudflare，长江证券研究所

过去互联网以注意力分配为核心的商业模式逐渐失效，Agent 时代需要新的“保安”和“收银员”——前者保护网站的价值，后者定义新的收费模式。传统互联网价值链为用户→浏览器→网站→云→收入（广告/交易），云处于价值链最远端，仅收取资源费用；Agent 时代价值链演变为用户→Agent/代理浏览器→网站→云→收入。原商业模式通过流量换取曝光、收取广告费；新商业模式将收费节点转移至浏览器与云基础设施，以按量计费与订阅制为主要形式。收费节点的基础设施化是云厂商议价能力提升的结构性来源。

图 11：注意力机制失效



资料来源：长江证券研究所（备注：逻辑框架图由 AI 辅助生成，人工核对。）

图 12：Agent 时代三种商业模式



资料来源：长江证券研究所（备注：逻辑框架图由 AI 辅助生成，人工核对。）

➢ **保安——Bot 管理与访问控制**

网站让渡内容换取流量失效，网页对机器访问者的态度从“开放欢迎”转变为“选择性收费”。网站防护手段由从“协议自律”升级为“基础设施强制”。根据 Cloudflare 数据，2025 年 Q2 至 2026 年 Q2，网站对爬虫的封锁率从 10.2%飙升至 35.8%，成功响应率从 80.5%降至 49.2%。robots.txt 本质是自愿遵守的君子协定，2025 年 7 月起，承载全球约 20%网站流量的 Cloudflare 对全网默认封锁 AI 爬虫，在爬虫触达源站前即完成拦截。2026 年 7 月，其进一步将 AI 流量分为 Search（搜索索引）、Agent（代用户实时执行任务）、Training（模型训练）差异化管控，访问控制进入细粒度时代。

表 2：Bot 管理与访问控制需求

问题	传统方案	Agent 时代新需求
谁可以访问我的内容?	robots.txt（自愿遵守）	面向 Agent 的细粒度访问控制、付费墙、API 密钥认证
如何防止恶意爬取?	IP 封禁、验证码	Agent 身份验证、行为审计、沙箱隔离

内容如何被合理使用?	版权声明	训练数据许可协议、实时调用计费

资料来源：稀土掘金，Cloudflare，区块周刊，MaxAEO，长江证券研究所

➢ **收银员——Agent 调用计费**

封锁成为默认选项后，“收费放行”是内容方与 AI 方之间的新均衡。基础设施角色也从保安延伸为收银员。通过 B2A（Business to Agent）模式，网站或内容平台直接向 AI Agent 收取费用，以换取内容访问权、数据使用权或 API 调用权的商业模式。Agent 时代的变现路径收敛为三类：

1.按调用计费：每次 Agent 调用 API/抓取页面收取费用。

2.按任务计费：Agent 完成特定任务（如比价、预订）后收取佣金。

3.数据许可费：允许 Agent 使用网站数据训练模型，收取许可费。

云的本质是 Runtime，每一轮范式迁移都伴随着产品演绎

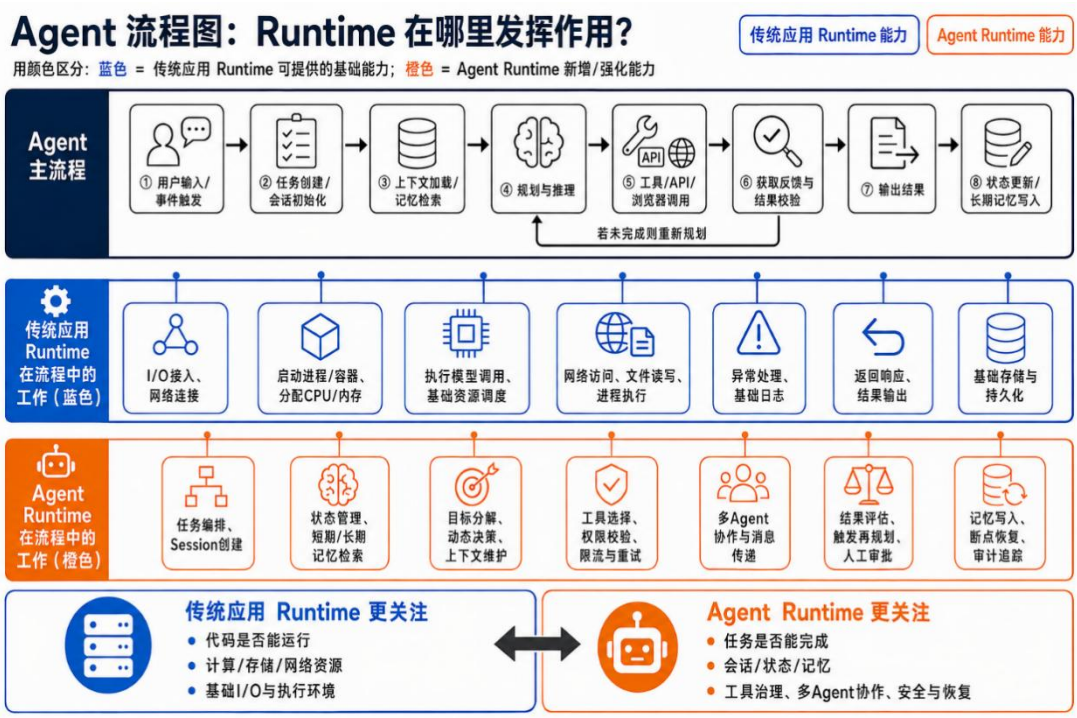
云 Runtime 演进：从代码执行到任务闭环

回归本质，云计算的核心是 Runtime，即为当下需求提供运行环境。传统应用 Runtime（基础层）解决“代码能否运行”的问题，定价锚定资源规格。 其执行链路为静态代码→可执行程序→资源分配→执行退出，关注代码能否运行、资源是否隔离。定价按资源规格（vCPU、内存、存储容量）×使用时长，本质上是对标准化资源的批发转售。

2023-至今：AI Runtime（平台层）解决“模型能否高效训练与推理”的问题，定价锚定算力消耗、芯片能力。 其执行链路为数据→张量→计算图→分布式训练/推理→模型输出，关注吞吐、延迟与显存效率，按 GPU 卡时、Token 消耗量或推理次数计费。

2025-未来：Agent Runtime（应用层），核心是为了 Agent 的顺畅运行提供需要的匹配环境。 阿里云提出 AI 正在经历从“模型即产品”向“Agent 即应用”的范式跃迁：模型需要更强的训练和推理基础设施来高效生产 Token，而智能体则需要全新的云原生环境来承载其短生命周期、突发负载、动态的环境依赖等独特特征，这对云厂商提出了全新的要求。

图 13：传统 Runtime 能力对比 Agent Runtime 能力

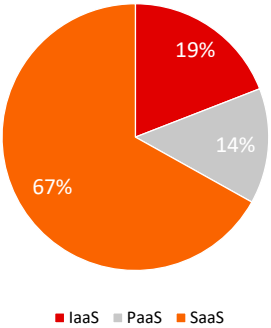


资料来源：长江证券研究所（备注：逻辑框架图由 AI 辅助生成，人工核对。）

为适应不同时代的任务需求，云厂商的产品实际上是在灵活迭代的。

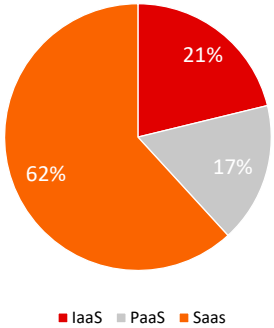
第一阶段（2006－2020）：资源调度与规模化供给，云处于价值链最远端。 云厂商以 IaaS 与 PaaS 为核心，向电商、社交、游戏等泛互联网企业提供弹性基础设施。利润来源于硬件采购的规模成本差与资源利用率提升，定价锚点为云资源性能成本。在这一阶段，云完全不参与上层业务利润分配，其收入弹性受制于下游客户的 IT 预算而非业务增长本身。

图 14：2018 年全球 SaaS、IaaS、PaaS 占比



资料来源：IDC，长江证券研究所

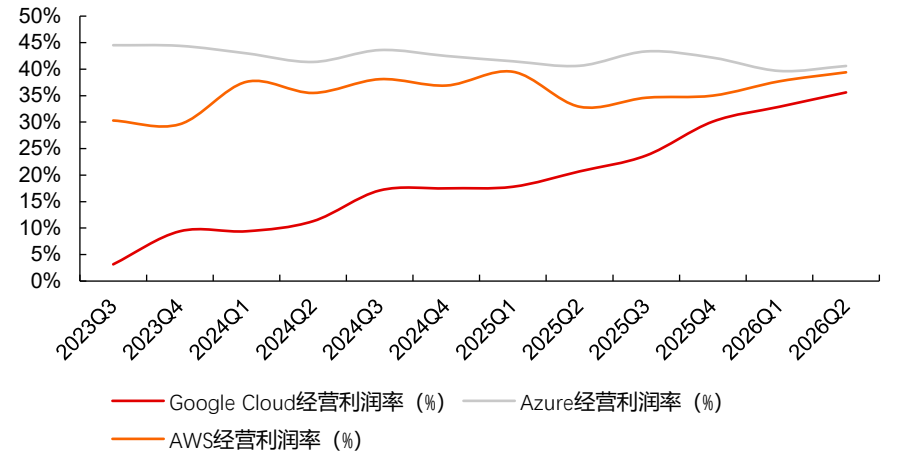
图 15：2022 年全球 SaaS、IaaS、PaaS 占比



资料来源：IDC，长江证券研究所

第二阶段（2023-至今）：算力与 Token 成为核心生产资料，云从管道层升级为能源层。 大模型爆发使 GPU/NPU 集群、推理优化与 API 服务成为新的利润核心，MaaS 与 AI Infra 推动定价锚点从资源规格转向 Token 消耗，利润来源扩展至算力调度效率、模型服务溢价及软硬协同降本技术。云厂商的议价能力显著增强，但算力价格的长期下行决定了单纯卖算力的模式难以支撑利润率持续扩张。

图 16：Google Cloud、Azure、AWS 经营利润率



资料来源：Bloomberg，长江证券研究所

第三阶段（2025-未来）：向上穿透至应用层，甚至远期可能参与利润分成，云有望进一步成为价值分配的核心参与者之一。 Agent 时代，云厂商通过 Agent Runtime、Skills Marketplace 与协议标准（MCP/A2A）直接嵌入业务流程，定价锚点进一步上移至任务完成次数、节省的人力工时与 GMV 增量，利润来源纳入平台佣金、高价值 Skill 授权费及企业级安全合规增值服务。利润与下游业务 GMV 直接挂钩，生态粘性与协议标准主导权成为核心竞争壁垒。

表 3：云发展三阶段

	传统应用时代	AI 云时代	Agent 云时代
时间	2005-2023	2023-至今	2025-未来
核心定位	资源基础设施	智能能力平台	任务执行平台
主要产品	计算、存储、网络	算力、模型 API、数据工具	运行环境、工具、技能生态
客户购买什么	IT 资源	模型与推理能力	可完成的业务任务
如何收费	按资源用量收费	按 Token 与算力收费	订阅 + 成功费 + 生态抽成
核心壁垒	规模与稳定性	模型、算力与推理效率	任务完成率与生态粘性

资料来源：钛媒体，中国信通院，阿里云，长江证券研究所

行业大势之下，国内外核心云厂商纷纷开启 Agentic 云布局

Cloudflare：新兴云厂商边缘网络入口垄断

➢ **云核心优势**

Cloudflare 是全球反向代理服务的龙头，依靠其网络规模成为 AI 边缘云计算基础设施平台。公司网络覆盖全球超过 350 个城市、1000 个数据中心，每日处理数万亿级请求，全球边缘网络能够使约 95%的互联网连接人口在约 50 毫秒内访问其网络节点。2025 年公司实现收入 21.7 亿美元，同比增长 30%。

➢ **云发展历程**

阶段一：云计算边缘基建时代。移动互联网应用爆发，用户对页面加载、API 响应的容忍度降至毫秒级，传统中心化数据中心的长距离回传成为瓶颈。Cloudflare 率先将轻量级计算 runtime 部署至全球边缘节点，把计算能力无限贴近终端用户，实现零冷启动、就近执行；随后推出前端托管与自动化构建能力，完成从网络加速到边缘计算平台的转型，使用户无需自建全球基础设施即可获得低延迟体验。

阶段二：云计算边缘全栈时代。边缘计算场景从简单的请求处理扩展到复杂应用，用户需要在边缘侧完成数据持久化、结构化存储与 AI 推理。Cloudflare 相继补齐对象存储、Serverless 数据库、向量检索与 AI 推理服务，形成边缘全栈云，打破了传统 CSP 中心强、边缘弱的架构惯性。

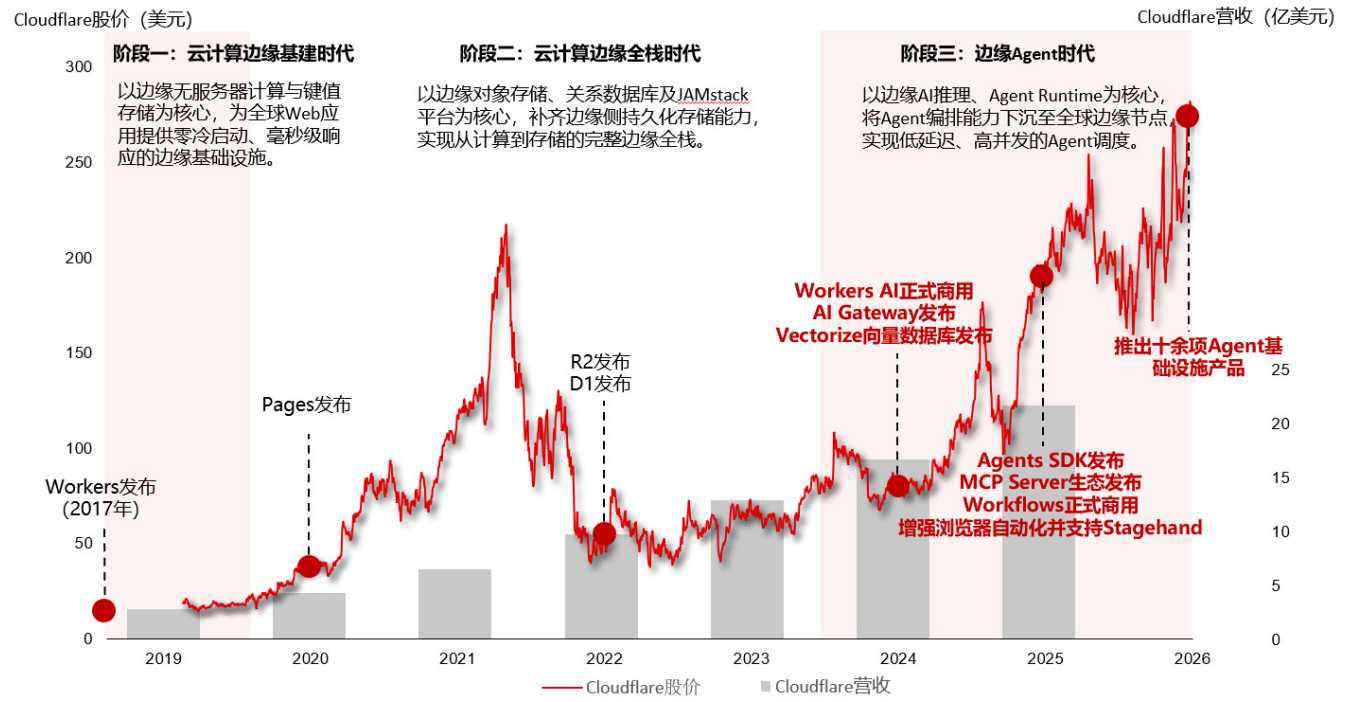
阶段三：边缘 Agent 时代。AI Agent 浪潮下，智能体对低延迟推理、长时状态保持、多工具链调度的需求远超传统互联网时代。Cloudflare 将 AI 推理、状态持久化、工作流编排与工具调用协议深度整合，让边缘节点具备 Agent 原生能力，实现 Agent 的低延迟、高并发、全球化调度。

表 4：Cloudflare 边缘 Agent 时代重要事件

时间	重要事件	战略意义	商业验证
2024-04	Workers AI 正式商用	建立全球边缘推理基础设施	-
2025-04	Agents SDK、 MCP server 生态发布	形成 Agent 运行、状态、记忆和工具连接	-
2025-09	增强浏览器自动化并支持 Stagehand	让 Agent 通过浏览器操作缺少 API 的网站和业务流程	Playwright 支持正式商用；Stagehand 进入 Cloudflare 开发平台
2026-04	推出十余项 Agent 基础设施产品	扩展至 Agent 沙箱、身份与网络控制，升级为 Agent 网络与安全基础设施	AI 与 Agent 推动 Workers、安全和网络需求

资料来源：Cloudflare 官网，长江证券研究所

图 17：Cloudflare 营收（亿美元）-股价（美元）复盘及云阶段发展历程



资料来源：Bloomberg，Cloudflare 官网，长江证券研究所

➢ Agentic Cloud 战略

Agentic AI-first 战略为核心。6 月，Cloudflare 联合创始人兼 CEO Matthew Prince 在访谈中认为，如果 Bot 访问网页趋势持续下去，五年内，互联网流量可能会是今天的 1000 倍。Agent 大规模普及后，互联网“每周都会爆出一个 Log4j 级别的漏洞”，安全风险正在急剧攀升。服务器、网络基础设施、CPU、GPU、内存等为 Bot 服务的基础设施都会变得至关重要。Cloudflare 正通过裁员、建 Agent 基础设施、推微支付重构商业规则、锚定大客户提客单价四重举措，从网络安全厂商转型为 Agentic Internet 的底层操作系统。

目前，凭借高覆盖网关代理优势，将 AI Bot 流量纳入“拦截—缓存—计费”（Pay Per Crawl）商业闭环。2024 年 7 月，公司推出一键屏蔽爬虫选项。次年 7 月，推出 Pay Per Crawl 内测：默认阻止 AI 爬虫，发布商可设置按次收费，AI 公司需付费或协商后才能爬取，Cloudflare 承担交易中介角色，向 AI 公司收费并向内容方结算抽成。截至 2025 年 8 月，超过 250 万个网站在 Cloudflare 上选择完全禁止 AI 训练爬虫。

基于 Pay Per Crawl 模式，Cloudflare 围绕 HTTP 402 状态码构建微支付基础设施，意在协议层面固化 AI 时代的内容定价权。公司正构建每秒 1000 万次交易的处理能力，较 Visa 高两个数量级，为 AI 时代的内容微支付提供底层协议。

低延迟边缘 AI 推理（Workers AI）抢占超大规模云推理负载。Cloudflare 把 AI 模型的推理计算从中心化云服务器下沉到全球的边缘节点，让用户的 AI 请求在离终端最近的节点完成计算并返回结果，以降低延迟并减少出口费用。2026Q1， Workers AI 被管理层明确称为“Q1 业绩超预期的显著贡献者”，边缘推理的差异化定位已开始兑现收入。

Google Cloud：TPU+Gemini 垂直闭环

➢ **云核心优势**

Google Cloud 是唯一不依赖英伟达也能供给顶级 AI 算力的云厂商，模型能力正最直接地转化为云业务竞争力。从 TPU 到 Gemini 的全栈自研能力带来的成本与供给自主权，使 Google Cloud 成为传统云厂商毛利率改善弹性最大，具备“卖芯片+卖模型+卖云”三重变现路径的厂商。2025 年营收约 587 亿美元、同比增长约 36%；2026Q2 积压订单环比接近翻倍、突破 5000 亿美元。

➢ **云发展历程**

阶段一：云计算应用时代。移动互联网与电商爆发，企业对高并发、分布式存储和实时数据分析的需求激增。Google Cloud 依托搜索和广告业务积累的分布式计算能力，将 Google 的数据处理、分布式数据库和容器技术产品化，形成区别于 AWS 基础设施广度路线的“数据云+混合云”能力，为互联网企业提供可弹性扩展的基础设施底座。

阶段二：云 AI 应用时代。AI 产业化加速，大模型训练对专用算力和 MLOps 全栈工具提出全新要求。Google Cloud 以 TPU 算力、Vertex AI 开发平台和 Gemini 模型逐步形成覆盖 AI 训练、部署、推理及应用开发的云端技术栈，将 AI 算力、模型开发、MLOps 及基础模型服务整合到统一云平台。

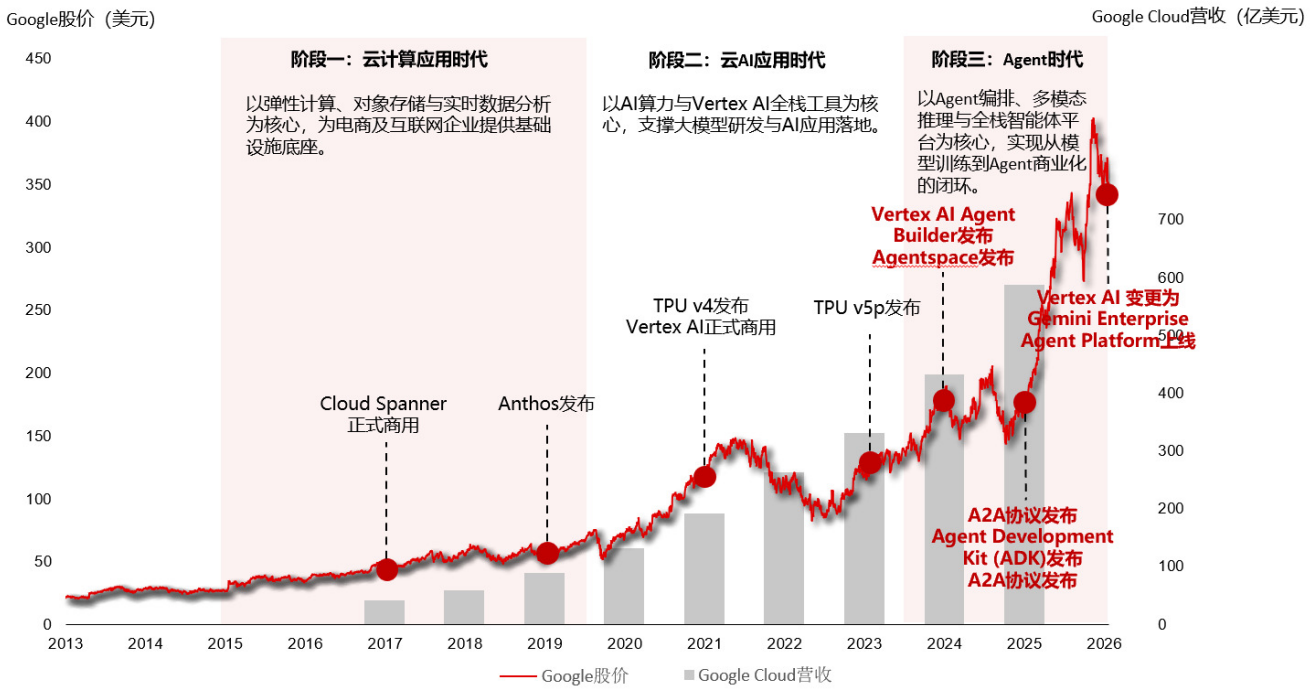
阶段三：Agent 时代。Google Cloud 以 Gemini 和 Vertex AI 作为能力底座，以开放协议扩大云上 Agent 生态，提供低门槛的 Agent 构建、多模态推理与跨系统协作能力，推动不同框架、模型和平台上的 Agent 互联。

表 5：Google Cloud Agent 时代重要事件

时间	重要事件	战略意义	商业验证
2024-04	Vertex AI Agent Builder 发布	建立 Agent 构建、数据连接和托管运行平台	-
2024-12	Google Agentspace 发布	把企业搜索、Gemini 和专业 Agent 整合为员工工作入口	后续演进为 Gemini Enterprise Agent Platform
2025-04	开源 ADK、A2A 协议发布	布局开发框架与跨 Agent 通信标准	A2A 发布时获得 50 多家企业参与
2026-04	Vertex AI 变更为 Gemini Enterprise Agent Platform 上线	统一模型、Agent 开发、运行、集成、安全和治理	-

资料来源：Google 官网，长江证券研究所

图 18：Google Cloud 营收（亿美元）-谷歌股价（美元）复盘及云阶段发展历程



资料来源：Bloomberg，Google 官网，长江证券研究所

➢ **Agentic Cloud 战略**

2026 年 4 月，Cloud Next 大会提出“Agentic Enterprise”概念，判断企业 AI 正从试点进入规模化生产。CEO Thomas Kurian 认为真正的价值无法通过拼凑碎片化的芯片和脱节的模型来创造，应当通过“芯片—模型—数据—智能体”垂直一体化架构构建全栈 Agent 平台。芯片层，Google Cloud 发布第八代 TPU 并引入 NVIDIA 芯片，形成专为 Agent 优化的双栈算力；模型层，以 Gemini Enterprise Agent Platform 聚合自研与第三方模型，统一推理引擎保障企业级输出质量；数据层，将企业数据转化为 Agent 可理解的上下文，打通数据到决策的闭环；智能体层，构建 A2A 协议，实现跨厂商 Agent 自主协作。生态端，Google Cloud 投入 7.5 亿美元加速 Agent 部署，与 Merck 达成战略合作，并以 320 亿美元收购 Wiz 补强云安全；标杆客户涵盖 NASA、Apple。

阿里云：全栈自研体系加码企业场景

➢ **云核心优势**

阿里云是亚太地区最大的 IaaS 云厂商，实现国产全栈自研体系打通丰富企业场景及生态。阿里云构建了从倚天芯片、飞天云操作系统、云基础设施，到通义千问大模型、百炼 Agent 平台及钉钉企业应用的国产全栈体系，依托电商、金融、物流和办公等大规模业务场景，将 AI 能力直接嵌入企业研发、协同、客服、营销与经营流程。2026Q1，云智能集团收入 416.26 亿元，同比增长 38%，外部商业化收入增速提升至 40%；AI 相关产品收入达 89.71 亿元，占云外部商业化收入 30%，实现连续第十一个季度三位数同比增长。

➢ **云发展历程**

阶段一：云计算应用时代。阿里云以自研操作系统、计算架构强化虚拟化性能和云服务器资源效率，将阿里集团电商业务和双十一大促中的高并发、高可用及弹性扩展能力转化为通用云服务。

阶段二：云 AI 应用时代。阿里云以全栈自研系统强化云计算与 AI 算力的软硬协同能力，推动核心系统全面云原生化，支撑 AI 训练与产业智能化落地。完善国产云基础设施，为模型训练、推理及企业核心系统上云提供统一的计算平台。

阶段三：Agent 时代。阿里云以通义千问模型和百炼平台为基础，提供模型调用、RAG、工作流、Agent 开发、部署和应用管理能力，构建全栈 AI 基础设施和 Agent 服务平台商业化闭环。

表 6：阿里云 Agent 时代重要事件

时间	重要事件	战略意义	商业验证
2025	宣布未来三年投入至少 3800 亿元用于 AI 与云基础设施	从传统云服务商全面转型为 AI 基础设施提供商	Token 计费，按输入/输出 Token 分别计价
2025-09	ModelStudio-ADK、ModelStudio-ADP 发布	抢占企业级 Agent 开发入口	已有超 20 万开发者开发了 80 多万个 Agent
2026-03	企业级 AI 原生智能体平台悟空发布	Agent 能力以企业级产品形态规模化落地	-
2026-05	Qwen3.7-Max、真武 M890、磐久 AL128 发布；新加坡千问大会 Qwen Cloud 发布；MuleRun Team 版、Enterprise 版发布	模型、芯片、服务器全栈自研迭代，Qwen Cloud 出海	席位订阅（Team 版 140 元/席位/月、1350 元/席位/年，已成交 1019 笔；Enterprise 版 2,500 元/席位/年，已成交 180 笔）

资料来源：阿里云官网，阿里云云栖大会资料，长江证券研究所（MuleRun 席位订阅数据截至 2026 年 8 月 30 日，统计近 180 天成交）

图 19：阿里云营收（亿美元）-股价（美元）复盘及云阶段发展历程



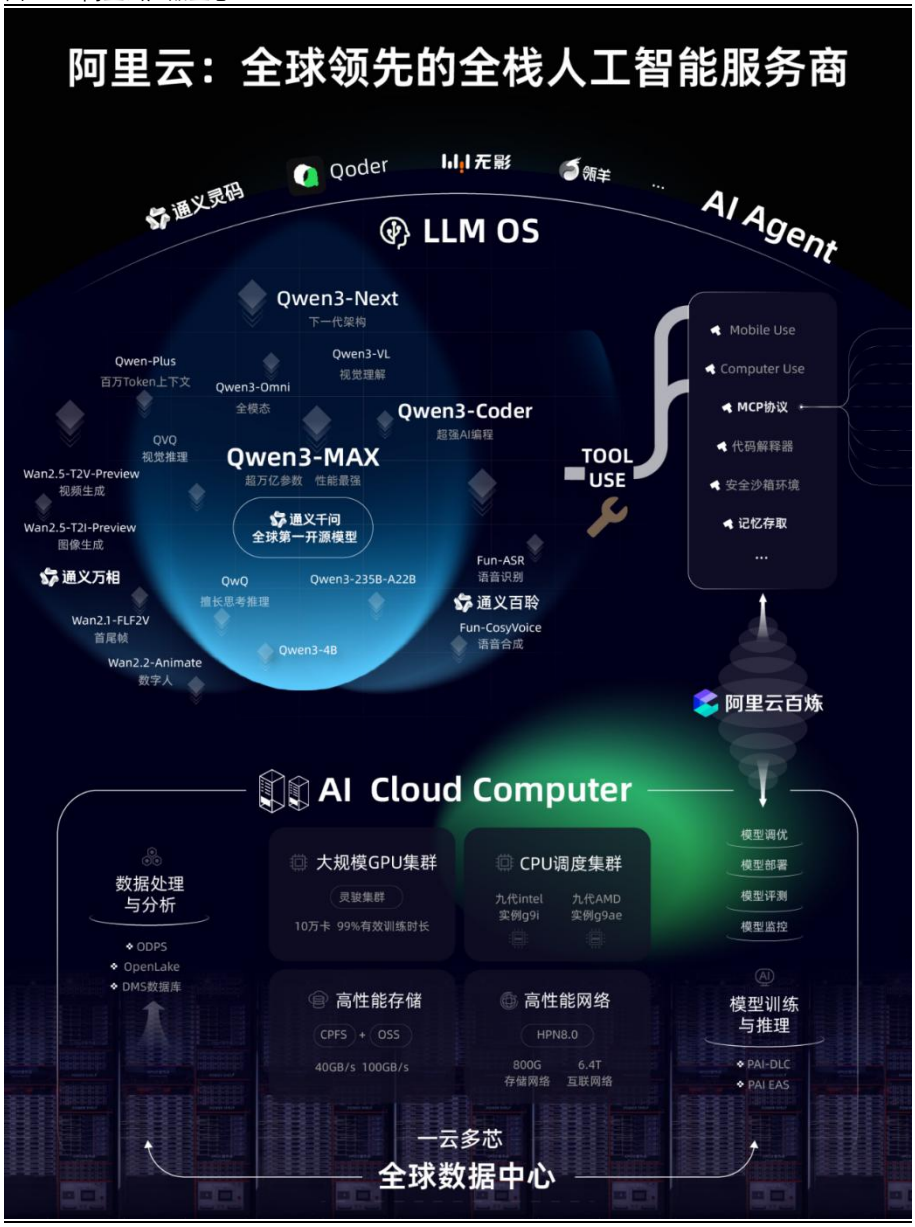
资料来源：Bloomberg，阿里云官网，阿里云云栖大会资料，长江证券研究所

➢ **Agentic Cloud 战略**

2026 年 5 月，阿里云在年度峰会宣布全面进入 Agentic 时代，提出“芯片—云—模型—推理”四位一体的全栈升级路径，并发布新一代旗舰模型 Qwen3.7-Max、自研 AI 芯片真武 M890。7 月，阿里于 WAIC 进一步推出 Agent Native Cloud，进一步完善面向 Agent 的全栈产品布局。阿里云云原生应用平台负责人周琦表示，下一轮竞争，比的不是谁拥有更多 Agent，而是谁能把 Agents 变成可控、可复用、可协作、会进化的组织资产。

目前，阿里云底层依托自研芯片与超节点服务器优化算力成本。平台层，通过百炼平台及 AgentRun、AgentTeams、AgentLoop 等工具链，覆盖 Agent 构建、运行、治理、协作全生命周期；应用层，推出无影 Agentic Computer，为企业 Agent 提供 7×24 小时桌面级运行环境。同时，阿里云将 60 余项核心云产品能力转化为 Skill 化、MCP 兼容格式，使 Agent 可直接调用云资源，数据库、大数据、运维、安全等核心产品线均已开发专用产品级 Agent。

图 20：阿里云产品生态



资料来源：阿里云云栖大会资料，长江证券研究所

AWS：规模与自研芯片构筑成本护城河

- **云核心优势**



AWS 是全球公有云开创者、市场份额占比第一，算力供给规模领先、自研芯片优势具备利润空间。2025 年自研芯片（Trainium/Graviton）年化收入突破两百亿美元且保持三位数增长，Anthropic 已锁定最高 5GW Trainium 算力，发布 Bedrock AgentCore 构成 Agent 编排层。2025 年 AWS 实现营收 1,287 亿美元，同比增长 20%，年末在手订单约 2,440 亿美元。

➢ **云发展历程**

阶段一：云计算应用时代。全球电商与互联网应用高速扩张，企业对弹性计算、Serverless 架构和托管数据库的需求成为刚需。AWS 以基础设施即服务和云原生架构为核心，构建全球覆盖的计算与存储底座，奠定云计算商业化范式。

阶段二：云 AI 应用时代。数字化转型深化，AWS 推出自研芯片降低云上推理训练成本，并为企业提供多种基础模型的统一云端访问、调用和应用开发能力，降低企业绑定单一模型供应商的风险。

阶段三：Agent 时代。AWS 以大模型平台与智能助手为核心，构建从模型训练到 Agent 部署的完整基础设施，将分散的 Agent 基础设施封装成标准化云服务。

表 7：AWS Agent 时代重要事件

时间	重要事件	战略意义	商业验证
2023-07	Agents for Amazon Bedrock 发布	进入 Agent 编排和工具调用市场	-
2025-03	Bedrock 多 Agent 协作正式商用	由单 Agent 编排扩展到 Supervisor 与专业 Agent 协作	预览阶段已覆盖金融、医疗、供应链、制造和客服等行业
2025-07	AgentCore 发布	补齐运行时、记忆、身份、工具、浏览器、代码执行和观测	发布预览版约 5 个月，SDK 下载超过 200 万次
2025-10	Amazon Quick Suite 发布	进入企业员工 Agent 入口	-

资料来源：AWS 官网，长江证券研究所

图 21：AWS 营收（亿美元）-亚马逊股价（美元）复盘及云阶段发展历程



资料来源：AWS 官网，长江证券研究所

➢ **Agentic Cloud 战略**

2025 年 3 月 AWS 重组成立独立 Agent-focused AI 团队，标志 AWS 押注 AI Agent 成为其下一条数十亿美元业务线。产品端，全面增强 Bedrock AgentCore 作为生产级 Agent 执行层，配套推出面向业务用户、面向开发者等工具，投入 10 亿美元成立 Forward Deployed Engineering 组织嵌入客户现场，形成 Agent 全栈体系。生态上，与顶尖模型厂商绑定、合作。AWS 通过追加 50 亿美元投资深度绑定 Anthropic，锁定最高 5GW Trainium 算力，2026 年 4 月已部署超 100 万 Trainium2 芯片；将 OpenAI GPT 最新模型及 Codex 引入 Bedrock，以“自研芯片+双顶级模型”构建 Agent 算力底座。

Azure：绑定 OpenAI，企业级分发能力最强

➢ **云核心优势**

Azure 是全球第二大公有云厂商，其软件生态与企业级客户关系绑定最深，是 AI 能力向企业工作流渗透的最短路径。依托企业级混合云架构及 OpenAI 战略合作，将大模型能力深度嵌入企业级安全框架；Azure OpenAI Service 与 Azure AI Foundry 构成模型托管与 Agent 编排核心层，将 Agent 运行时深度嵌入企业既有 IT 体系。2026 Q1 Intelligent Cloud 收入 347 亿美元，同比增长 28%，其中 Azure 及其他云服务收入同比增长 39%，约 95%财富 500 强企业部署 Azure。

➢ **云发展历程**

阶段一：云计算应用时代。通过 Azure Stack 实现本地数据中心与公有云统一架构，打通多云与边缘资源管理，构建开发者全生命周期云原生工具链，推动传统企业 IT 资产平滑上云。

阶段二：云 AI 应用时代。以 OpenAI 云合作为战略支点，将 GPT 系列大模型封装为企业级安全云服务，叠加低代码开发平台与云数据仓库，实现从模型托管、数据治理到 AI 应用部署的完整云原生 AI 基础设施。

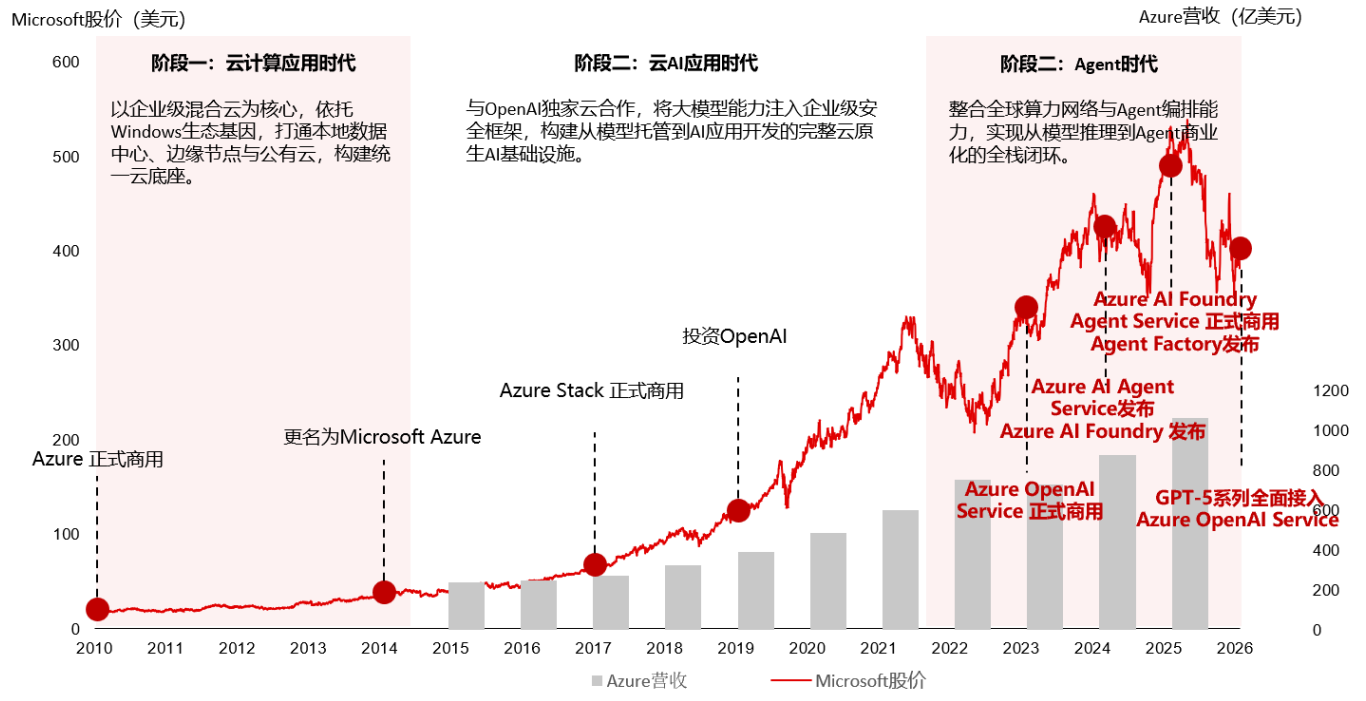
阶段三：以 Azure AI Foundry 为统一控制平面，向下整合自研 Maia AI 芯片、Azure OpenAI 算力与全球数据中心网络，向上通过 Azure AI Agent Service 提供多 Agent 编排、企业级安全治理与跨云 API 调用，构建全栈云原生 Agent 平台。

表 8：Azure Agent 时代重要事件

时间	重要事件	战略意义	商业验证
2024-12	Azure AI Agent Service 发布	建立专业开发者 Agent 开发与托管服务	2025 年初 Foundry 月活超过 20 万
2025-05	Azure AI Foundry Agent Service 正式商用	补齐生产级多 Agent 编排、开放协议和全生命周期治理能力	FY2025 创建约 300 万个 Agent
2025-06	Foundry Agent Service 推出 Deep Research	把复杂检索、推理和研究流程封装为企业可调用 Agent 能力	80%的财富 500 强使用 Foundry
2025-11	Agent Factory 发布	整合 Work IQ、Fabric IQ、Foundry IQ，统一计量模式支持企业构建、部署和规模化 Agent	-

资料来源：Microsoft 官网，长江证券研究所

图 22：Azure 营收（亿美元）-微软股价（美元）复盘及云阶段发展历程



资料来源：Bloomberg，Microsoft 官网，长江证券研究所

➢ Agentic Cloud 尝试及战略规划

2026 年 6 月，微软 Build 大会中 CEO Satya Nadella 提出企业 AI 的下一章不再是演示或问答，而是能够自主执行多步骤任务并交付实际工作的 Agent。为支撑这一范式转移，Azure 以 Azure AI Foundry 为核心构建企业级 Agent 基础设施。平台聚合 1900+款主流模型并通过托管服务降低企业算力门槛，配套推出面向 Agent 的安全测试与身份治理体系；基础设施层提供弹性运行时与长时程工作流调度，支持 Agent 按需扩缩容。

风险提示

1、 大模型发展不及预期风险。若大模型发展不及预期则可能使得整个 AI 行业进展不及预期。

2、 大厂转型速度不及预期风险。新产品的推出与公司整体的组织效率、研发进度等多要素相关。

3、 竞争加剧风险。若竞争加剧则有可能影响行业玩家的利润水平。

4、 上游硬件涨价幅度超预期风险。上游硬件涨价可能会影响云商的利润空间。

投资评级说明

行业评级	报告发布日后的 12 个月内行业股票指数的涨跌幅相对同期相关证券市场代表性指数的涨跌幅为基准，投资建议的评级标准为：	
	看　　好：	相对表现优于同期相关证券市场代表性指数
	中　　性：	相对表现与同期相关证券市场代表性指数持平
	看　　淡：	相对表现弱于同期相关证券市场代表性指数
公司评级	报告发布日后的 12 个月内公司的涨跌幅相对同期相关证券市场代表性指数的涨跌幅为基准，投资建议的评级标准为：	
	买　　入：	相对同期相关证券市场代表性指数涨幅大于 10%
	增　　持：	相对同期相关证券市场代表性指数涨幅在 5%～10%之间
	中　　性：	相对同期相关证券市场代表性指数涨幅在-5%～5%之间
	减　　持：	相对同期相关证券市场代表性指数涨幅小于-5%
	无投资评级：	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级。

相关证券市场代表性指数说明：A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准。

办公地址

上海
Add /虹口区新建路 200 号国华金融中心 B 栋 22、23 层
P.C /（200080）

武汉
Add /武汉市江汉区淮海路 88 号长江证券大厦 37 楼
P.C /（430023）

北京
Add /朝阳区景辉街 16 号院 1 号楼泰康集团大厦 23 层
P.C /（100020）

深圳
Add /深圳市福田区中心四路 1 号嘉里建设广场 3 期 36 楼
P.C /（518048）

分析师声明

本报告署名分析师以勤勉的职业态度，独立、客观地出具本报告。分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点。作者所得报酬的任何部分不曾与，不与，也不将与本报告中的具体推荐意见或观点而有直接或间接联系，特此声明。

法律主体声明

本报告由长江证券股份有限公司及/或其附属机构（以下简称「长江证券」或「本公司」）制作，由长江证券股份有限公司在中华人民共和国大陆地区发行。长江证券股份有限公司具有中国证监会许可的投资咨询业务资格，经营证券业务许可证编号为：10060000。本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格书编号已披露在报告首页的作者姓名旁。

在遵守适用的法律法规情况下，本报告亦可能由长江证券经纪（香港）有限公司在香港地区发行。长江证券经纪（香港）有限公司具有香港证券及期货事务监察委员会核准的“就证券提供意见”业务资格（第四类牌照的受监管活动），中央编号为：AXY608。本报告作者所持香港证监会牌照的中央编号已披露在报告首页的作者姓名旁。

其他声明

本报告并非针对或意图发送、发布给在当地法律或监管规则下不允许该报告发送、发布的人员。本公司不会因接收人收到本报告而视其为客户。本报告的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证，也不保证所包含信息和建议不发生任何变更。本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。本公司已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不包含作者对证券价格涨跌或市场走势的确定性判断。报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。本研究报告并不构成本公司对购入、购买或认购证券的邀请或要约。本公司有可能会与本报告涉及的公司进行投资银行业务或投资服务等其他业务(例如:配售代理、牵头经办人、保荐人、承销商或自营投资)。

本报告所包含的观点及建议不适用于所有投资者，且并未考虑个别客户的特殊情况、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。投资者不应以本报告取代其独立判断或仅依据本报告做出决策，并在需要时咨询专业意见。

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、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可以发出其他与本报告所载信息不一致及有不同结论的报告；本报告所反映研究人员的不同观点、见解及分析方法，并不代表本公司或其他附属机构的立场；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本公司及作者在自身所知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

本报告版权仅为本公司所有，本报告仅供意向收件人使用。未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布给其他机构及/或人士（无论整份和部分）。如引用须注明出处为本公司研究所，且不得对本报告进行有悖原意的引用、删节和修改。刊载或者转发本证券研究报告或者摘要的，应当注明本报告的发布人和发布日期，提示使用证券研究报告的风险。本公司不为转发人及/或其客户因使用本报告或报告载明的内容产生的直接或间接损失承担任何责任。未经授权刊载或者转发本报告的，本公司将保留向其追究法律责任的权利。

本公司保留一切权利。