

AI Infra 研究专题 1：溯源中美云厂盈利差，AI 时代共迎新机遇——行业深度报告

投资要点

- 过去中美云厂商盈利产生较大差距的核心源于：1) 产业链价值差异；2) 规模差异；3) 商业模式差异。展望未来，我们认为：1) GPU 云时代的云厂商利润率天花板会因为产业链价值提升而提高；2) 中美云厂差异也会随着 AI 驱动景气度提升和商业模式的变更而大幅收窄。
- 展望 GPU 云时代：1) 中美云厂商网络传输层的价值链差异仍客观存在，但双方产业链均向上游布局 AI 芯片、向下游拓展 MaaS 与 Agent 业务，利润率天花板抬升；2) 规模效应中美差距收窄。美国云厂商起步较早，头部 2 家厂商在 CPU 时代已经处于规模经济的收获期。伴随 AI 带来新的基础设施需求，Google 云也在过去 2 年迎来了规模经济的收获期。我们认为：伴随国内 AI 需求增加、与阿里云等企业主动积极出海，未来同样会进入规模经济的收获期，规模效应带来的利润率差距会逐步收窄；3) 大模型更适合公有云部署，中美云厂商业模式差异同样收窄。目前基于大模型的企业级应用以 MaaS 模式为主要落地路径，并主要依托公有云平台部署。
- 产业链价值差异：北美云厂自建“全球骨干网”，国内云厂租用运营商带宽服务。
 1. 国内运营商建设骨干网。国内云厂商不具备基础电信业务的经营资质，不能建设底层公共网络基础设施，只能向获得许可证的四大运营商租用带宽服务，带宽费成为云厂商的重要成本项之一。
 2. 北美云厂商自建海陆光缆打造全球网络，网络传输层价值包含在云业务内。以高度依赖自建骨干网经营业务的 Cogent、Lumen 两家企业盈利能力为参照，我们认为：美国三大云厂骨干网相关业务平均 15% 的 OPM 是中美云厂利润率差异的主要来源之一。
- 规模差异：云业务是有显著规模效应的生意。美国云厂商起步较早，厂商处于规模经济的收获期，具体表现为：
 1. 更低的采购成本。厂商凭借大体量的采购规模议价压价，同时 ODM 直采省去中间环节，有效降低硬件采购开支。2015-2020 年，ODM 厂商广达、超微平均 OPM 仅 2%、4%，低于 OEM 厂商戴尔、慧与服务器业务 10%、11% 的利润率水平。
 2. 更大的研发投入构筑产品领先性，提高资源利用率、降低客户转移成本。2013-2025 年，尽管研发开支均逐年增长，但阿里研发总额及研发费率均不及美国同行。
 3. 全球客户错峰使用带来更高的资源利用率分摊投入成本。1) 算力资源利用率高。2022 年美国云数据中心上架率（65%）高于国内（58%），更少的算力闲置使得单位算力成本更低；2) 更好地分摊企业人员成本。2023-2025 年谷歌云利润率上升，人员相关成本开支营收占比分别下降了 3.8、10.1、9.9 个百分点。
- 商业模式差异：客户付费意愿和商业模式不同，主要体现在公私有云部署意愿上。
 1. 从总量上看，过去中国企业 IT 付费意愿低于全球平均。2024 年全球 IT 支出市场规模达到 5.04 万亿美元，占 GDP4.5%；中国 IT 支出规模近 2.7 万亿元，占 GDP 2.0%。
 2. 从结构上看，美国公有云部署比例高，客户对公有云付费意愿强；国内国央企受合规约束偏好私有云，中小企业虽偏好公有云，但议价敏感、付费意愿偏弱。而私有云因为缺乏规模效应和规模调度，商业模式不如公有云。
- 投资建议

行业评级：看好(维持)

分析师：徐紫薇
执业证书号：S1230524120005
xuziwei@stocke.com.cn

相关报告

- 1 《世界模型渐成共识，行业从“多模态生成”走向“可交互工作流”》 2026.06.13
- 2 《AI 驱动北美云厂资本开支快速上行，云业务仍处于投入期》 2026.05.09

我们看好 GPU 云时代下云厂商价值链扩张带来的长期投资机会。生成式 AI 推动云计算竞争从单纯资源供给转向“算力基础设施、模型服务和企业 AI 平台”的综合竞争，云厂商能够凭借 GPU 资源池、数据中心规模和 AI 生态优势，持续向价值链上游模型服务和下游企业应用拓展，提升单位资源价值和盈利空间。

重点推荐：阿里巴巴；建议关注：亚马逊、微软、谷歌等。

- **风险提示：**产业竞争加剧；AI 商业化落地不及预期；监管政策约束；技术变革替代

目录

1 产业链价值差异：北美云厂自建“全球骨干网”，国内云厂租用运营商带宽服务	6
1.1 国内：运营商建设、云厂商租用骨干网	6
1.1.1 政策与牌照壁垒决定中国的底层公共网络基础设施建设方为运营商	6
1.1.2 带宽费作为中国云厂商的成本项	8
1.2 美国云厂商自建“全球骨干网”，网络传输层价值包含在云业务内	8
2 规模差异：公有云业务是有显著规模效应的生意	10
2.1 更低的采购成本	10
2.2 更大的研发投入构筑产品领先性，提高资源利用率，提升客户转移成本	11
2.3 全球客户错峰使用带来更高的资源利用率分摊投入成本	12
2.3.1 算力资源利用率高	12
2.3.2 更好地分摊企业人员成本	12
3 商业模式差异：客户付费意愿和商业模式不同	13
3.1 中国企业 IT 付费意愿过去低于美国企业	13
3.2 商业模式的不同	14
3.2.1 中国私有云部署比例高	15
3.2.2 中小企业价格敏感度高	17
3.2.3 美国大客户对公有云接受度高	18
4 展望和投资建议	19
4.1 GPU 云时代展望	19
4.1.1 中美云厂商网络传输层的差距仍在，但全产业链价值延展，利润率天花板变高	19
4.1.2 规模效应继续发挥效益，中国云厂商竞争格局改善	21
4.1.3 基于大模型的企业应用以 MaaS 模式落地，商业模式更偏好公有云	22
4.2 投资建议	23
5 风险提示	24

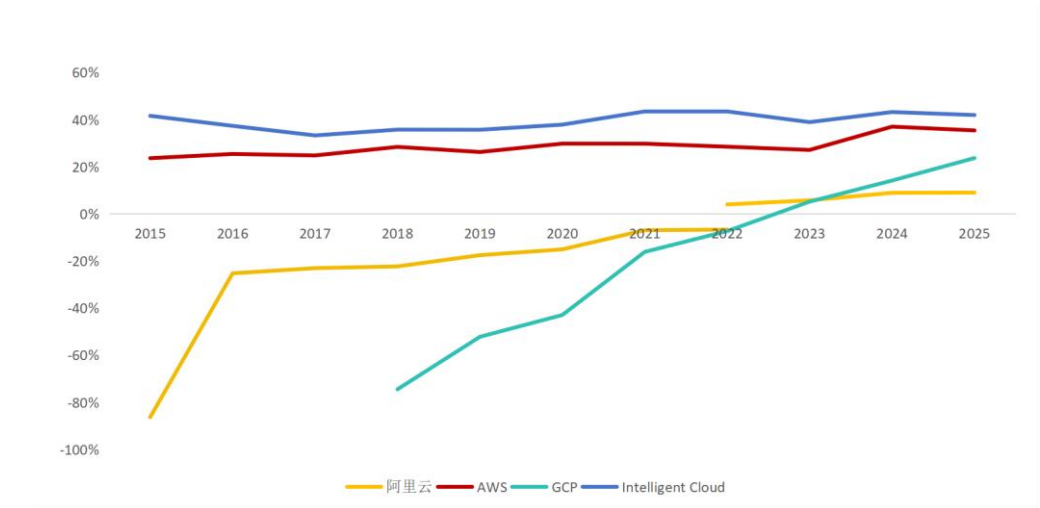
图表目录

图 1:	2025 年, AWS、智能微软云 (Intelligent Cloud) 和谷歌云的利润率显著高于阿里云	5
图 2:	骨干网是 CPU 云服务的底座支撑	6
图 3:	2018-2025 年, 中国电信/中国移动/中国联通净利率情况	7
图 4:	2011-2022 年 Cogent/Lumen OPM 情况	9
图 5:	美国云厂商资本性支出规模远高于阿里巴巴	10
图 6:	2015-2020 年台湾广达、SMCI 营业利润率低于 Dell 和 HPE 服务器相关业务营业利润率	11
图 7:	2013-2025 年阿里巴巴研发支出规模低于美国三大云厂商	12
图 8:	2013-2025 年阿里巴巴研发费用率低于美国三大云厂商	12
图 9:	谷歌云利润弹性主要驱动来自人员成本的规模效应	12
图 10:	2018-2027E 中国和美国 IT 支出对比	13
图 11:	2022 年中国、美国及全球 IT 支出占 GDP 比例	13
图 12:	云计算分类	14
图 13:	公有云和私有云对比	14
图 14:	2025 年中国公有云市场总额仅为美国市场的 9.8%	15
图 15:	2024 年, 国有企业营业收入占比超过 25%	15
图 16:	2024 年, 国有企业资产占比 38%	15
图 17:	2015-2023 年, 三大运营商云业务和阿里云业务规模相对份额对比	16
图 18:	2015-2023 年, 三大运营商云业务和阿里云营业收入增速对比	16
图 19:	2015-2023 年中国电信/中国移动/中国联通应收账款增速	17
图 20:	GPU 云时代产业链延伸为了“芯片+IAAS+PAAS+MaaS+Agent”全产业链	19
图 21:	GPU 云时代对 GPU 的依赖增强, 英伟达数据中心营业利润率大幅提升	20
图 22:	AWS、智能微软云 (Intelligent Cloud)、谷歌云和阿里云的业务规模不断扩大	21
图 23:	阿里云中国 IaaS 公有云市场份额 (%)	21
图 24:	阿里云全球 IaaS 公有云市场份额 (%)	21
图 25:	中国云厂商 25 年 H1 智能云市场份额 (%)	22
图 26:	中国云厂商 25 年智能云市场份额 (%)	22
图 27:	预期 2026 年中国公有云大模型调用 Token 消耗量将升至约 40000 万亿	23
表 1:	中国四大运营商骨干网	7
表 2:	阿里巴巴/金山云营业成本具体内容	8
表 3:	美国三大云厂商骨干网	8
表 4:	2022 年美国云厂商计算机设备及软件资本性支出/云业务收入占比高	8
表 5:	Cogent/Lumen 业务构成	9
表 6:	AWS、微软云、谷歌云头部行业用户	18
表 7:	中美云厂商自研芯片情况	19
表 8:	中美云厂商 MaaS 平台	20
表 9:	各大厂商盈利预测 (单位: 百万美元)	23

报告导语:

中美云厂商在 CPU 时代,利润率差距显著。2013-2025 年,AWS 和智能微软云(Intelligent Cloud)的平均营业利润率分别为 29%和 39%,且自 2015 年之后两大厂商利润率持续维持在 25%以上。2025 年,AWS、谷歌云和微软云的营业利润率为 35%、24%、42%,而阿里云 EBITA Margin 仅 9%。

图1: 2025 年, AWS、智能微软云(Intelligent Cloud)和谷歌云的利润率显著高于阿里云



资料来源:公司公告、浙商证券研究所

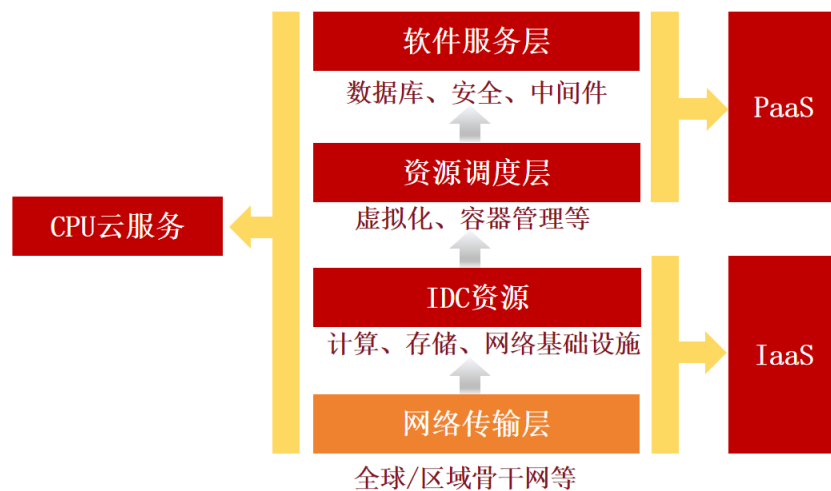
备注: 2015-2022 年阿里云数据为 OPM, 2023-2025 年数据为 EBITA Margin; 阿里巴巴、微软按照各自财年核算, 图中标注的 2025 年分别对应阿里巴巴 FY2026 (2025.4.1-2026.3.31)、微软 FY2025 (2024.7.1-2025.6.30), 两家公司历年财务数据均遵循该统计口径。

产生较大盈利差距的核心源于: 1) 产业链价值差异; 2) 规模差异; 3) 商业模式差异。

1 产业链价值差异：北美云厂自建“全球骨干网”，国内云厂租用运营商带宽服务

CPU 云服务包含 IaaS 和 PaaS。骨干网作为底层网络传输底座，实现多地域 IDC 机房互联，是分布式算力规模化调度、云体系数据全域流通的基础支撑，也是中美产业链价值的差异点。

图2：骨干网是 CPU 云服务的底座支撑



资料来源：浙商证券研究所

1.1 国内：运营商建设、云厂商租用骨干网

中国云厂商受制度、监管以及合规限制，依赖中国电信、中国移动、中国联通等持牌主体运营商的底层公共网络物理设施，需向运营商缴纳带宽使用费。

1.1.1 政策与牌照壁垒决定中国的底层公共网络基础设施建设方为运营商

国内所有云端流量的公网传输、跨省调度、跨境互通，必须且只能依托四大运营商的公共电信基础设施实现。

2008 年，六大通信运营商重组，整合为中国电信、中国移动、中国联通三家公司。2009 年，工业和信息化部出台《电信业务经营许可管理办法》，三大运营商取得基础电信业务经营许可证。中国广电则是在 2016 年 5 月成为第四家获得经营许可的运营商。

表1：中国四大运营商骨干网

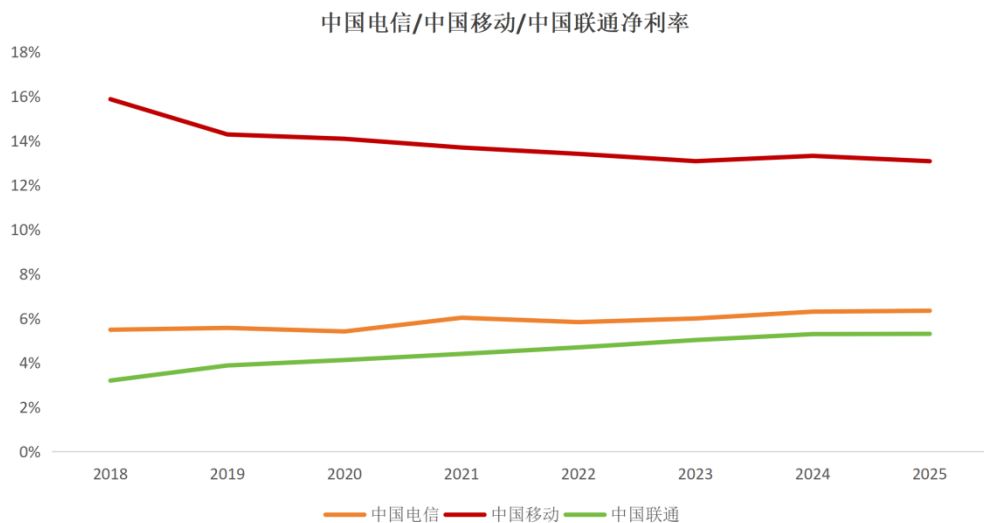
运营商	骨干网名称	建成年份	服务对象
中国电信	中国公用计算机互联网（ChinaNet）	1996 年	公众
	中国电信下一代承载网（CNCN）	2004 年	政企大客户
中国移动	中国移动互联网（CMNET）	2000 年	公众
	IP 专用承载网	2004 年	政企大客户
中国联通	China169	1999 年	公众
	CUII（原 CNCNET）	2000 年	政企大客户
中国广电	CBNNET	2018 年	公众

资料来源：CSDN、维基百科、浙商证券研究所

对于国内云厂商而言，通常自建园区内网基础设施建设，包括机房内交换机、路由器、服务器组网等，用于数据中心内部算力互联、云上私有网络调度；但无资质和权限建设和运营底层网络基础设施，包括跨省骨干光缆、公共骨干传输网、城域网、跨境海缆/陆缆、公共互联网出口带宽等基建。国内云厂商通过向运营商批量采购公网带宽资源、租用骨干网端口与互联网出口的方式，为客户提供上层云网络配套服务。

2025 年，中国电信、中国移动和中国联通三大运营商净利率分别为 6%、13%和 5%。

图3：2018-2025 年，中国电信/中国移动/中国联通净利率情况



资料来源：Wind、浙商证券研究所

1.1.2 带宽费作为中国云厂商的成本项

从公司财务报表的披露内容来看，以阿里巴巴、金山云为代表的中国云厂商明确将带宽费列为营业成本。

表2： 阿里巴巴/金山云营业成本具体内容

公司	
阿里巴巴	包括存货成本、物流成本，与运营我们的移动平台和网站相关的支出（例如服务器、电脑、呼叫中心及其他设备的折旧和维护费，还有 带宽费与主机托管费 ）；负责客户服务、移动平台及平台运营的员工、支付处理顾问的工资、奖金、福利及股权激励费用；以固定定价或收入分成模式向第三方营销机构支付的流量获取成本；向第三方支付的内容采购成本，以及线上媒体资源相关的原创内容制作成本；支付给支付宝及其他金融机构的支付手续费；其他杂项成本。
金山云	主要包括 带宽及互联网数据中心成本 （互联网数据中心成本）、电子设备、数据中心机器及设备的折旧费用、直接参与创收活动的员工的薪金及福利，以及提供服务直接应占的其他费用。

资料来源：公司公告、浙商证券研究所

1.2 美国云厂商自建“全球骨干网”，网络传输层价值包含在云业务内

美国云厂商通过资本开支自建全球骨干网，形成相比于国内云厂更高的底层网络基础设施价值。

表3： 美国三大云厂商骨干网

维度	AWS	GCP	Azure
网络形态	自建全球私有骨干网	自建全球私有骨干网	自建全球私有骨干网
物理层	600 万+公里陆地及海底光纤电缆	775 万公里的陆地和海底光缆	165000 英里的光纤和海底电缆系统
网络控制权	完全自控	完全自控	完全自控

资料来源：公司公告、浙商证券研究所

“全球骨干网”的建设意味着美国云厂商的资本性支出相对云收入的比例较高，印证北美云厂“算力+公共网络”双基建自持的商业模式。资本性支出中，美国云厂商大量资金流向服务器和网络设备等技术基础设施用于支持云业务发展。2022 年，亚马逊和谷歌“服务器和网络设备”资本性支出占云收入比例为 36%和 41%。

表4： 2022 年美国云厂商计算机设备及软件资本性支出/云业务收入占比高

公司	2022 年计算机设备及软件资本性支出/云业务收入
亚马逊（Amazon）	36%
谷歌（Google）	41%
微软（Microsoft）	12%
阿里巴巴	8%

资料来源：公司公告、浙商证券研究所

备注：阿里巴巴、微软按照各自财年核算，表中 2022 年分别对应阿里巴巴 FY2023（2022.4.1-2023.3.31）、微软 FY2022（2021.7.1-2022.6.30）。

Cogent 与 Lumen 为运营全球骨干网络的厂商，公司的业务收入高度依赖其在全球建设的骨干网。2011-2022 年间，两者平均 OPM 水平为 15%左右。

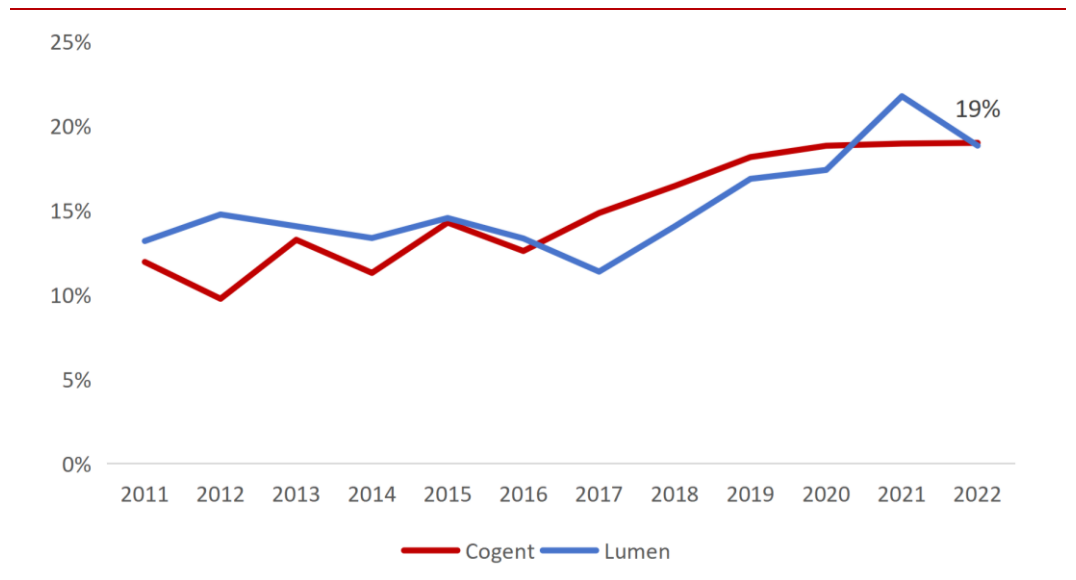
表5: Cogent/Lumen 业务构成

公司	业务类型	具体内容
Cogent	网内服务	网内服务面向已通过本公司自有光纤物理直连至我方网络的楼宇客户提供。交付网内互联网接入与专网服务时，无需依赖本地固网运营商、有线电视商提供接入线路；带宽覆盖 100Mbps 至 400Gbps，同时包含光波传输、机房托管、IPv4 地址租赁等全部依托自有光纤基础设施交付的业务。
	网外服务	为所在楼宇未与我方网络实现光纤直连的客户 提供互联网接入及专网服务。该类网外服务主要面向企业客户开展，需租用其他运营商线路搭建客户场地至我方骨干网络的“最后一公里”链路。
	非核心收入	并购承接而来的非核心业务。仅维持存量客户服务履约，不再主动对外推广、销售该类业务；预计非核心业务收入将持续下滑。
	波长收入（2022 年收购 Sprint 后新增）	为有大带宽专属点对点传输需求的网络型客户提供 10Gbps、100Gbps 及 400Gbps 光波与光传输服务。
Lumen	Business	面向大型企业、中型企业、公共事业政府客户、运营商批发客户、海外客户提供服务；营收按照产品生命周期分为增长型、培育型、存量传统、其他四类。依赖建立在全球的大约 163,000 座光纤网内建筑和 34,000 英里的光缆线路。
	Mass Market	面向家庭用户与小型商户，包含光纤宽带、铜缆宽带、语音及其他传统业务。

资料来源：公司公告、浙商证券研究所

2022 年，Cogent、Lumen 的 OPM 均为 19%。以两家成熟骨干网运营商盈利能力作为参照，我们可以认为：美国三大云厂商骨干网相关业务平均 15%的利润率是中美云厂利润率差异的主要来源之一。

图4: 2011-2022 年 Cogent/Lumen OPM 情况



资料来源：Wind、浙商证券研究所

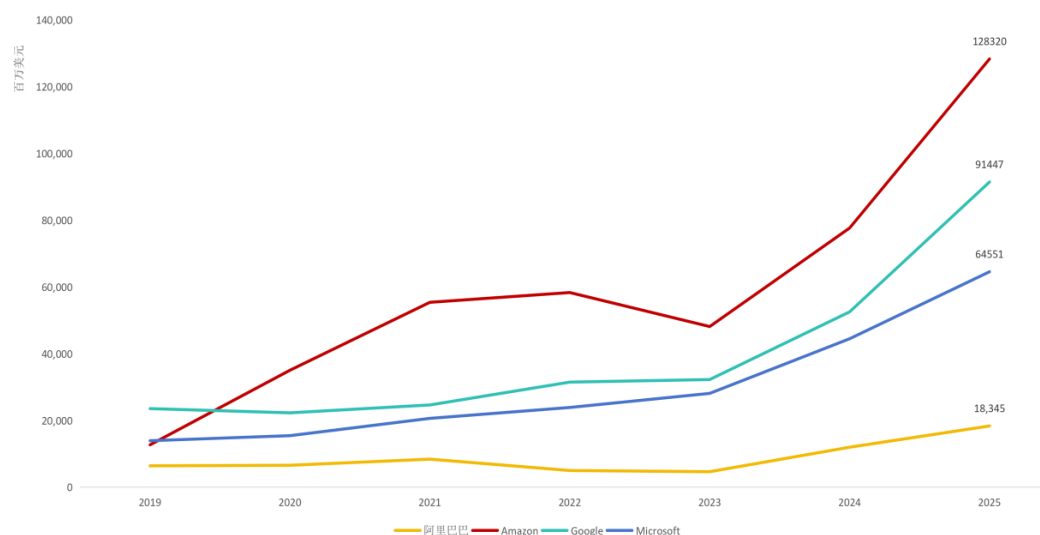
2 规模差异：公有云业务是有显著规模效应的生意

规模差异不仅是中美云厂利润率差异的重要原因，也是过去 AWS、微软智能云利润率持续领先 Google 云业务的主要原因。以 AWS 为代表的美国头部云厂商发展起步更早，且全球化布局，在行业发展初期即完成大规模基础设施的集中建设，庞大的硬件资产与业务体量充分发挥规模效应，更早进入了规模经济的收获期，主要体现为三点：1）更低的采购成本；2）更大的研发投入构筑产品领先性，提高资源利用率、提升客户转移成本；3）全球客户错峰使用带来更高的资源利用率分摊投入成本。

2.1 更低的采购成本

云厂商运营业务需进行大量服务器、数据中心、网络等硬件设备及配套软件的采购。相较国内云厂商，美国头部云企业资本开支规模显著更高，依托庞大的集采体量形成规模效应。

图5：美国云厂商资本性支出规模远高于阿里巴巴



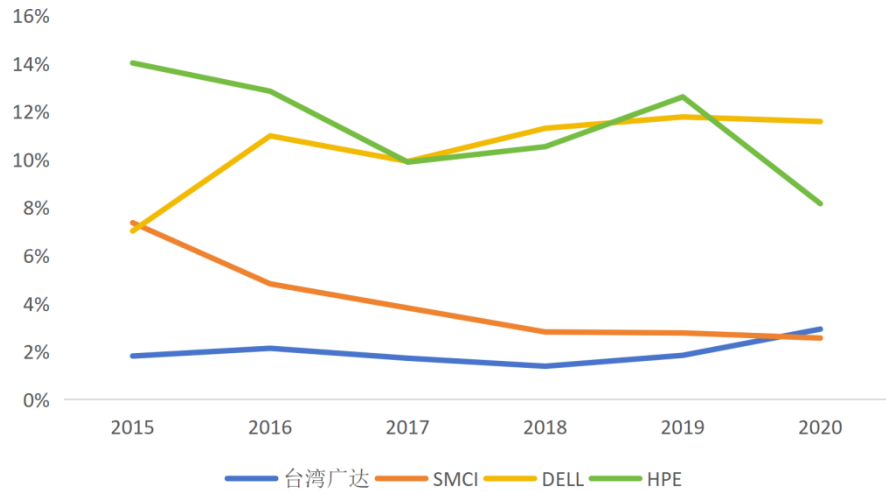
资料来源：Wind、公司公告、浙商证券研究所

备注：阿里巴巴、微软按照各自财年核算，图中标注的 2025 年分别对应阿里巴巴 FY2026（2025.4.1-2026.3.31）、微软 FY2025（2024.7.1-2025.6.30），两家企业历年财务数据均遵循该统计口径。

企业的大体量订单拥有较强议价能力，能够获取阶梯批量折扣、直供定制硬件，直接压低硬件与软件的单位采购成本。英伟达作为云厂商核心上游供应商，其超大规模云客户贡献数据中心营收占比约 50%；2025 年英伟达美国市场收入占比达 46.94%，中国市场为 28.88%，直观印证中美云厂商在设备采购体量上的差距。

规模较大的云厂商能够采用 ODM 直采模式，降低采购成本。2015-2020 年，以销售服务器为主营业收入的 ODM 企业台湾广达、超微电脑（SMCI）的平均营业利润率为 2%、4%；而主打企业级市场的 OEM 品牌厂商戴尔（Dell）和慧与（HPE）服务器相关业务的平均营业利润率为 10%、11%。

图6：2015-2020 年台湾广达、SMCI 营业利润率低于 Dell 和 HPE 服务器相关业务营业利润率



资料来源：公司公告、Wind、浙商证券研究所

备注：超微电脑、戴尔按照各自财年核算，图中标注的 2020 年分别对应超微电脑 FY2020（2019.7.1-2020.6.30）、戴尔 FY2021（2020.2.1-2021.1.29），两家企业历年财务数据均遵循该统计口径。

2.2 更大的研发投入构筑产品领先性，提高资源利用率，提升客户转移成本

美国头部云厂商凭借持续且高强度的研发投入，深耕自研芯片、云原生架构、智能调度与全栈云服务等技术，构筑起全方位的产品领先壁垒。领先的底层技术架构与精细化运维体系，大幅提升算力、存储等资源的整体利用效率，摊薄单位服务成本。

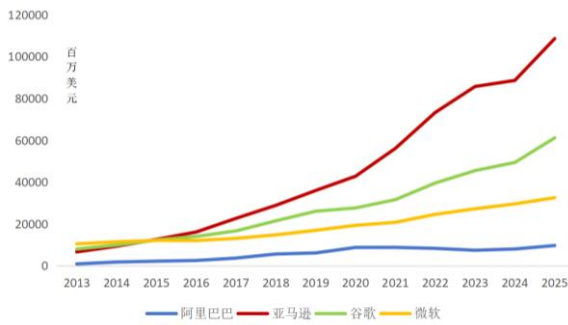
同时，美国云厂商依托大量研发布局的托管数据库、消息中间件、Serverless 函数等 PaaS 服务成熟稳定、品类丰富、易用性突出，更加全面渗透进企业应用研发、数据治理、业务运维等流程。

完善的开发运维体系和良好的使用体验使得客户主观上缺乏迁移意愿；即便考虑更换云服务商，相较于可灵活迁移的 IaaS 基础资源，企业依托各大云厂商 PaaS 搭建的体系和单一云平台深度耦合。不仅要完成海量数据迁移，还需大规模改写业务代码、重构分布式架构、重建整套开发运维流水线，极高的转移成本构筑起客户锁定壁垒，持续增强用户粘性。

2013-2025 年数据显示，国内外云厂商研发支出均不断上升，但阿里巴巴研发支出规模和研发费用率明显低于美国三大云厂商。2025 年，亚马逊、微软、谷歌、阿里巴巴的研发费用率分别为 15%、12%、15%、6%。



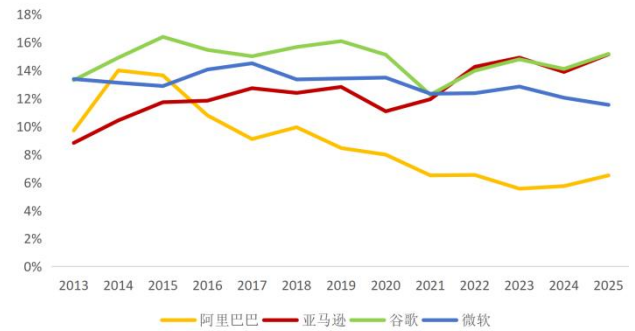
图7：2013-2025 年阿里巴巴研发支出规模低于美国三大云厂商



资料来源：Wind、浙商证券研究所

备注：阿里巴巴、微软按照各自财年核算，图中标注的 2025 年分别对应阿里巴巴 FY2026(2025.4.1-2026.3.31)、微软 FY2025(2024.7.1-2025.6.30)，两家企业历年财务数据均遵循该统计口径。

图8：2013-2025 年阿里巴巴研发费用率低于美国三大云厂商



资料来源：Wind、浙商证券研究所

备注：阿里巴巴、微软按照各自财年核算，图中标注的 2025 年分别对应阿里巴巴 FY2026(2025.4.1-2026.3.31)、微软 FY2025(2024.7.1-2025.6.30)，两家企业历年财务数据均遵循该统计口径。

2.3 全球客户错峰使用带来更高的资源利用率分摊投入成本

不同地区企业客户的业务负载存在明显的时间差异和地域差异。云厂商能够将分散于不同地区、不同时间段的大量客户需求纳入统一资源池，通过跨区域的弹性调度实现计算、存储和网络资源的动态匹配，提高服务器、CPU 等硬件设备的运行效率。

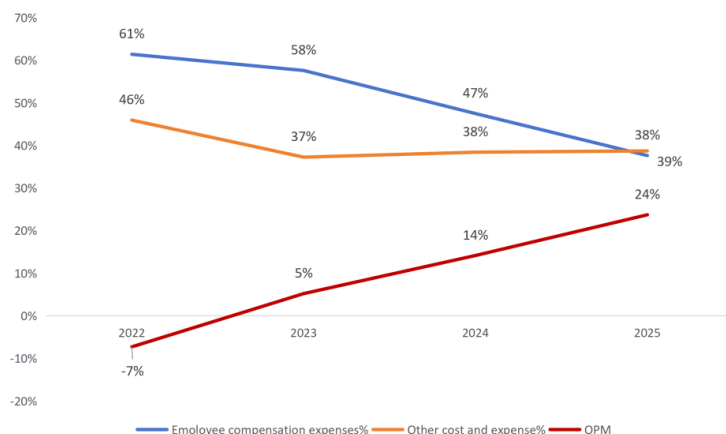
2.3.1 算力资源利用率高

美国三大云厂商利用大体量的业务规模实现了更高的算力资源利用率，降低了单位投入成本。国家信息中心数据显示，2018-2022 年，我国在数据中心机架总规模年均增速超过 30%，算力基建扩张速度迅猛，但资源利用效率偏低。2022 年国内数据中心上架率仅为 58%，低于全球平均水平（60%）和欧美发达国家平均水平（65%）。业务规模的限制使得机架闲置，从而带来算力资源浪费、投入分摊效率不足的问题，单位算力的投入成本居高不下。

2.3.2 更好地分摊企业人员成本

云厂商的人员规模与薪酬体系相对固定。头部云厂商凭借较高的规模效应，更好的分摊人力成本。随着谷歌云规模的上升，2023、2024、2025 年，谷歌云人员相关成本开支占收入的比例分别下降了 3.8 个百分点、10.1 个百分点、9.9 个百分点。人员成本的规模效应是利润释放弹性的重要驱动力。

图9：谷歌云利润弹性主要驱动来自人员成本的规模效应



资料来源：公司公告、浙商证券研究所

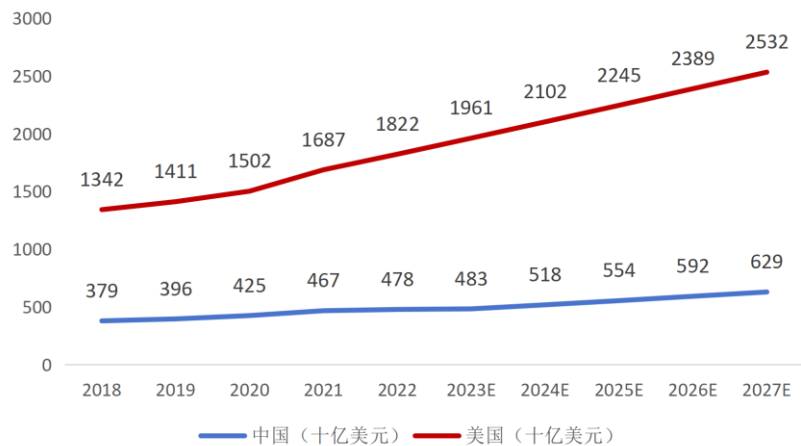
3 商业模式差异：客户付费意愿和商业模式不同

一方面，从总量上看，中国企业 IT 付费意愿低于全球平均；另一方面，从结构上看，美国公有云部署比例高，客户对公有云付费意愿强；中国央企大量上云需求以私有云为主，虽然中小企业对公有云接受度高，但中小企业价格敏感度高，付费意愿相对较低；而私有云因为缺乏规模效应和规模调度，商业模式天然不如公有云。

3.1 中国企业 IT 付费意愿过去低于美国企业

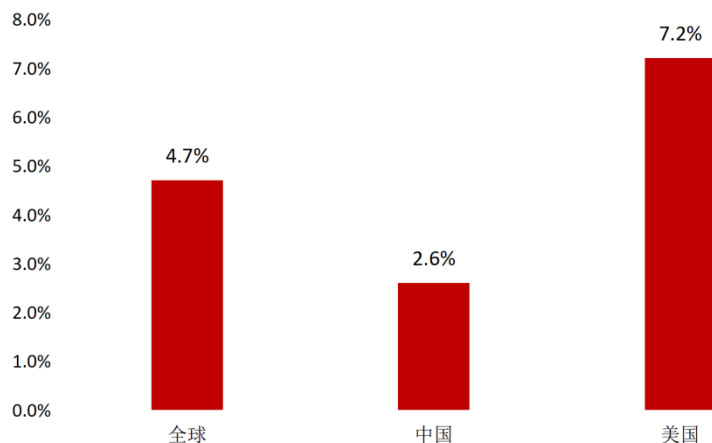
中国企业 IT 支出过去低于美国。根据沙利文数据，2022 年，中国 IT 支出为 4780 亿美元，占 GDP 比例为 2.6%；而美国 IT 支出 18220 亿美元，占 GDP7.2%。Gartner 数据显示，2024 年全球 IT 支出市场规模达到 5.04 万亿美元，占 GDP4.5%；中国信通院数据则表明，2024 年我国 IT 支出规模近 2.7 万亿元，占 GDP2.0%。我国企业 IT 支出规模和意愿低于全球平均。随着中国企业对数字化、智能化的重视程度不断提升，预计未来差距会逐渐减少。

图10： 2018-2027E 中国和美国 IT 支出对比



资料来源：沙利文、浙商证券研究所

图11： 2022 年中国、美国及全球 IT 支出占 GDP 比例



资料来源：沙利文、浙商证券研究所

3.2 商业模式的不同

云计算主要分为公有云、私有云和混合云三种形态。

图 12：云计算分类

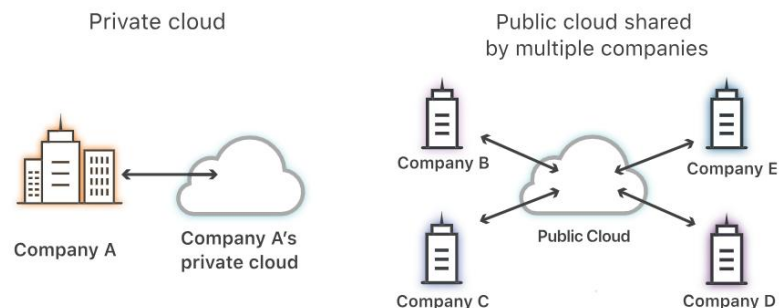
私有云	公有云	混合云
➢ 定义：用户自己开发或运营商针对用户开发的云计算环境； ➢ 特点：服务、规则、布置可控，控制力强；安全性高	➢ 定义：第三方提供商通过公共互联网为用户提供的云服务； ➢ 特点：响应标准化、自动化；按需申请、按量付费；可无限扩容	➢ 定义：将公有云提供的存储、计算能力同企业自有资源结合的云； ➢ 特点：兼具公有云和私有云特点，安全、可弹性扩容及敏捷开发部署

资料来源：浙商证券研究所

私有云偏项目制，难以获取更高的剩余价值，资源利用率低，通常为 20%-40%。私有云部署的服务场景具有极强的封闭性，其资源搭建、服务器运维、数据存储均服务于单一企业或特定闭环体系，资源调度范围被严格局限在固定圈层之内，无法形成跨区域、跨行业的规模化资源流转。

公有云通过大规模集群管理和智能调度，资源利用率可提升至 60%以上，理想情况可达 100%。公有云依托全球化的基础设施布局，可在全球范围内算力、存储、带宽等各类基础资源的统一整合、动态调配与按需复用，既能高效承接全球不同地区、不同行业用户的多元化需求，又能精准盘活分散在全球各地的闲置云端资源。在这种全球化调度体系下，云端基础资源不再是单一主体的专属资产，而是可循环、可复用、可规模化变现的公共数字化资源。运营方通过统一调度海量分散资源、匹配全球市场需求、优化资源配置效率，不断压缩边际成本，放大服务溢价，持续从资源供需的差值和规模化服务的流量差值中积累剩余价值，构建起可持续、高增量的云端商业盈利体系。

图 13：公有云和私有云对比

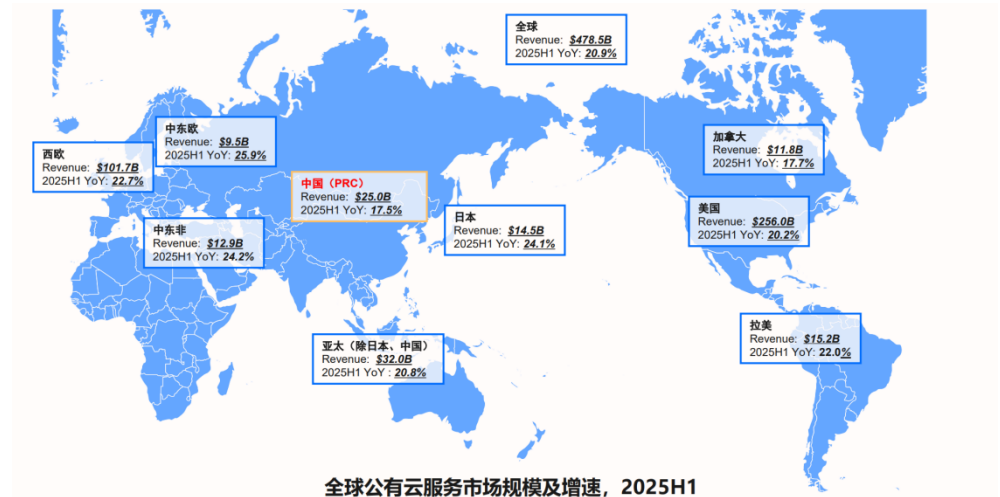


资料来源：Cloudflare、浙商证券研究所

3.2.1 中国私有云部署比例高

根据麦肯锡的研究，中国与其他国家与地区的区别在于其私有云部署比例较高。中国信通院数据显示，2022 年中国服务器公有云的输出比例仅有 28%，而美国超过 60%。2025 年，中国公有云市场规模为 250.1 亿美元，为美国市场的 9.8%。产生差异的原因在于中国国央企偏好部署私有云。

图 14：2025 年中国公有云市场总额仅为美国市场的 9.8%



资料来源：IDC、浙商证券研究所

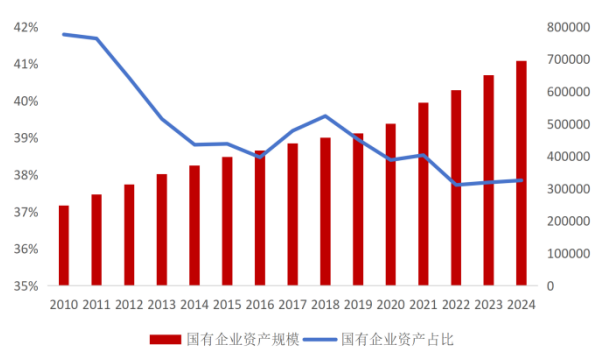
国央企在国民经济中占据重要地位。2024 年，国有企业资产规模占比 38%，营业收入占比 28%。出于合规的要求，国央企对本地化部署的要求高，私有云需求更好。

图 15：2024 年，国有企业营业收入占比超过 25%



资料来源：《中国统计年鉴 2025》、浙商证券研究所

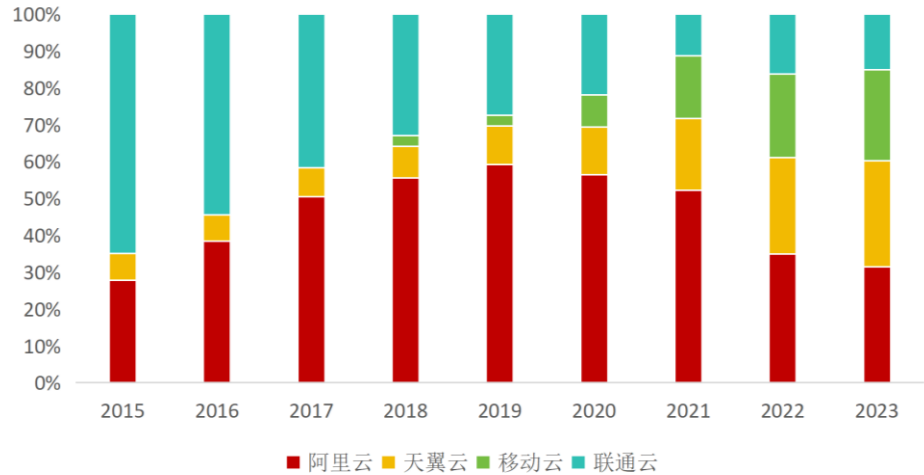
图 16：2024 年，国有企业资产占比 38%



资料来源：《中国统计年鉴 2025》、浙商证券研究所

2015-2023 年，受益于国央企的大量上云需求，以私有云、政务云为核心的运营商云业务保持高速增长，而阿里云营业收入增长率呈现下降趋势。尤其 2020 年后，天翼云、移动云和联通云云业务收入规模和相对份额不断扩大，增长速度均高于阿里云。

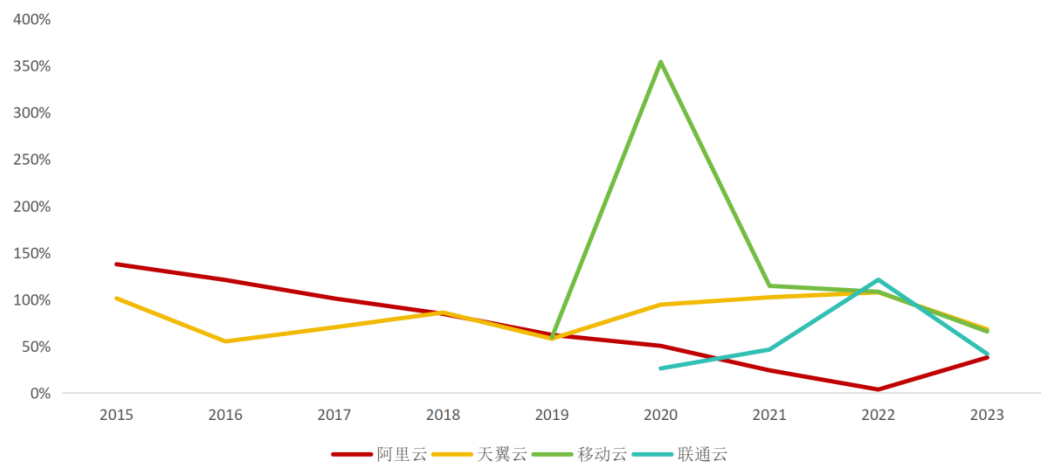
图 17： 2015-2023 年，三大运营商云业务和阿里云业务规模相对份额对比



资料来源：公司公告、浙商证券研究所

备注：阿里巴巴按照财年核算，图中标注的 2023 年对应阿里巴巴 FY2024（2023.4.1-2024.3.31），历年财务数据均遵循该统计口径。

图 18： 2015-2023 年，三大运营商云业务和阿里云营业收入增速对比

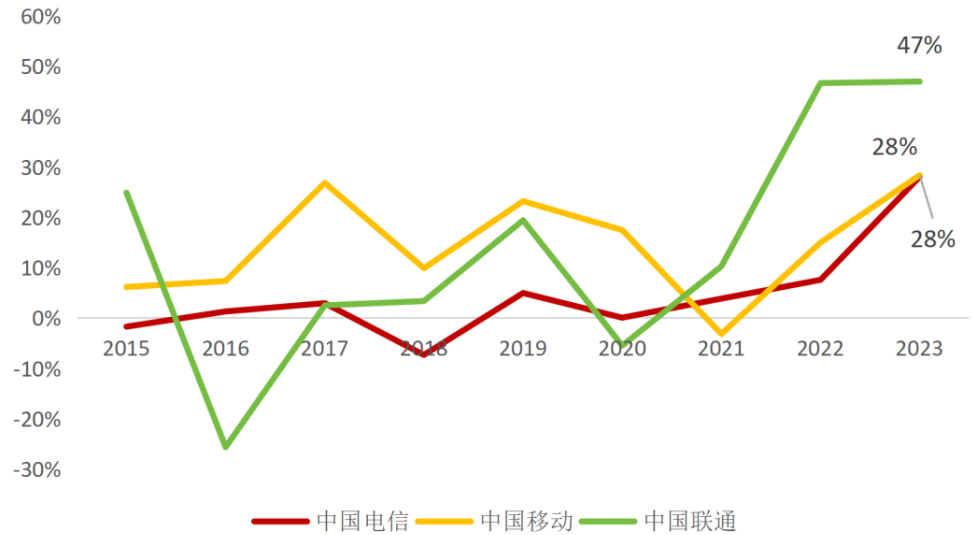


资料来源：公司公告、浙商证券研究所

备注：阿里巴巴按照财年核算，图中标注的 2023 年对应阿里巴巴 FY2024（2023.4.1-2024.3.31），历年财务数据均遵循该统计口径。

然而国央企账期通常较长。对运营商而言，业务规模的扩张将带来公司应收账款不断累积。尤其在运营商云业务规模迅速扩张的 2020-2023 年，应收账款规模也迅速扩大，2023 年中国移动和中国电信应收账款增长率均为 28%，中国联通达到 47%。

图 19： 2015-2023 年中国电信/中国移动/中国联通应收账款增速



资料来源：公司公告、浙商证券研究所

3.2.2 中小企业价格敏感度高

中国中小企业普遍更青睐公有云服务。公有云采用按需付费、弹性扩容的订阅制运营模式，企业无需高额的前期大量资本投入，大幅降低企业数字化转型的门槛与试错成本。

但与此同时，中小企业现金流相对薄弱、利润空间狭窄，IT 预算规模小。企业在选用公有云服务时具备极强的价格敏感度，不会过度追求高端性能与增值服务，更侧重性价比。他们会通过比价、降配、关停闲置资源等方式控制云服务成本，对云厂商定价调整的变化反应十分敏锐。

3.2.3 美国大客户对公有云接受度高

以 AWS、Azure 为代表的美国云厂商面向全球客户，核心集中北美和欧洲市场，企业大客户对公有云接受度高。中国信通院数据显示，2023 年美国企业和欧盟企业的上云率已经分别达到 85%和 70%。Fortune business insights 研究报告进一步指出，2025 年超过 68%的美国大型企业通过公有云运行关键工作任务，客户付费意愿强。

表6: AWS、微软云、谷歌云头部行业用户

	亚马逊 AWS	微软云	谷歌云
互联网	Facebook、Instagram、WhatsApp、Netflix、Discovery+、Zoom、Disney+	小米	Spotify
零售	Lululemon、Adidas、百惠买	玛氏、宝洁、雀巢、欧莱雅、可口可乐	苹果公司、高露洁
金融	高盛、哥伦比亚银行、蒙特利尔银行、纳斯达克证券交易所、永明金融集团	摩根士丹利、富国银行、加拿大皇家银行、软银、贝塔斯曼	德意志银行
医疗	辉瑞、拜耳、百特国际、罗氏、吉利德	GE 医疗、英国国民保健署、Adaptive 公司	
制造	西门子、斯特兰蒂斯、法拉利、恩智浦半导体	惠普、3M、日产、戴姆勒	法国电气企业泰雷兹集团
服务	美联航、温德姆酒店集团、德甲联赛、美国冰球联盟	希尔顿酒店、空中客车	
政府	美国国防部、中央情报局、美国国家海洋和大气管理局、美国医疗保险和医疗补助服务中心、美国退伍军人事务所	美国联邦政府、美国小企业管理局、美国疾病管制与预防中心、美国国防部、中央情报局、美国陆军、西澳大利亚州政府	美国国际开发署、美国空军研究实验室、美国农业部、美国马萨诸塞州政府
电信	爱立信、英国电信、欧洲运营商 Telefonica、贝尔公司、瑞士电信	AT&T、Verizon、欧洲运营商 Telefonica 和 BT、日本电信、澳洲电信、新加坡电信	德国电信子公司 T-Systems

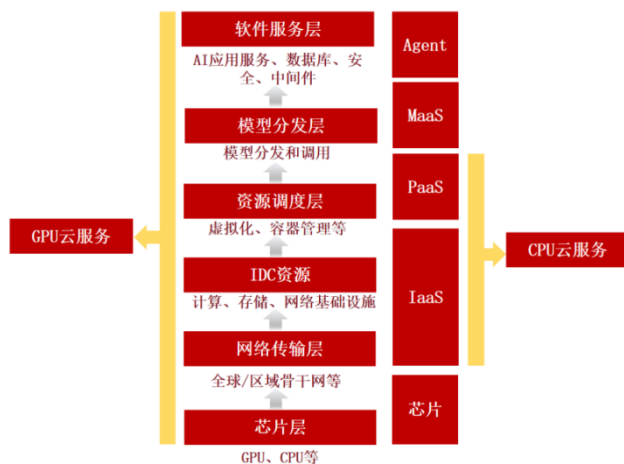
资料来源：电子工程专辑、浙商证券研究所

4 展望和投资建议

GPU 时代，云厂商产业价值链变迁，由 CPU 云时代“IaaS+PaaS”层延伸为了“芯片+IaaS+PaaS+MaaS+Agent”全产业链；因此利润率的天花板应高于 CPU 云时代。

其次，中美云厂核心差异中的规模差异和商业模式差异减弱，中国云厂商的利润弹性更大。

图20： GPU 云时代产业链延伸为了“芯片+IAAS+PAAS+MaaS+Agent”全产业链



资料来源：浙商证券研究所

4.1 GPU 云时代展望

4.1.1 中美云厂商网络传输层的差距仍在，但全产业链价值延展，利润率天花板变高

由于国内政策和牌照壁垒限制，骨干网建设带来的带宽收入仍然成为中美云厂商在GPU 云时代最核心的差异。

与此同时，中美云厂商均向产业链上下游拓展，云厂商未来盈利能力和利润率天花板有望进一步提升：

- 1) 向上游延伸至 AI 芯片，通过掌握核心硬件降低成本、提升资源调度效率；

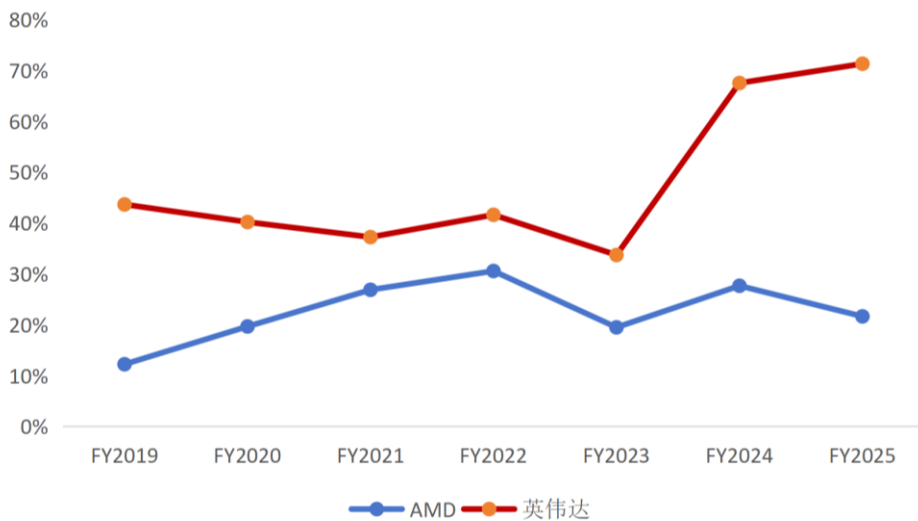
表7： 中美云厂商自研芯片情况

公司	自研芯片	发布时间
亚马逊（Amazon）	CPU: AWS Graviton	2018 年 Graviton1; 2026 年 Graviton5
	AI 加速芯片: Inferentia	2018 年 Inferentia1; 2022 年 Inferentia2
	AI 加速芯片: Trainium	2020 年 Trainium1; 2025 年 Trainium3
谷歌（Google）	ASIC 芯片: TPU	2015 年 TPuv1; 2025 年 TPuv7
微软（Microsoft）	CPU: Cobalt	2023 年 Cobalt100; 2025 年 Cobalt200
	AI 加速芯片 Maia	2023 年 Maia100; 2026 年 Maia200
	AI 加速芯片: 含光 800	2019 年
	AI 加速芯片: 真武 810E	2026 年
阿里巴巴（平头哥）	AI 加速芯片: 真武 M890	2026 年
	服务器 CPU: 倚天 710	2021 年
	智能网卡芯片: 磐脉 920	2026 年
	SSD 主控芯片: 镇岳 510	2023 年

资料来源：公司公告、维基百科、百度百科、浙商证券研究所

英伟达和 AMD 作为当前 AI 计算基础设施中的通用加速芯片供应商，其高性能 GPU 及配套软件生态占据 AI 训练市场核心位置。随着云厂商自研 ASIC/AI 芯片逐渐成熟，将降低对外部 GPU 采购的依赖，并通过芯片定制化提升单位算力成本效率。2025 财年英伟达、AMD 数据中心业务的营业利润率分别为 71%、22%。对于当前云厂商而言，采购成本较高。

图21： GPU 云时代对 GPU 的依赖增强，英伟达数据中心营业利润率大幅提升



资料来源：公司公告、浙商证券研究所

2) 向下游延伸至 MaaS 层和 AI Agent，通过软件化服务提高客户粘性和单位算力变现能力。一方面，AI 应用深度嵌入企业业务流程，提高客户迁移成本和平台粘性；另一方面，通过 API 调用、模型服务订阅收费，云厂商能够突破单纯资源出租的价格约束，提升单位算力的商业价值。

表8： 中美云厂商 MaaS 平台

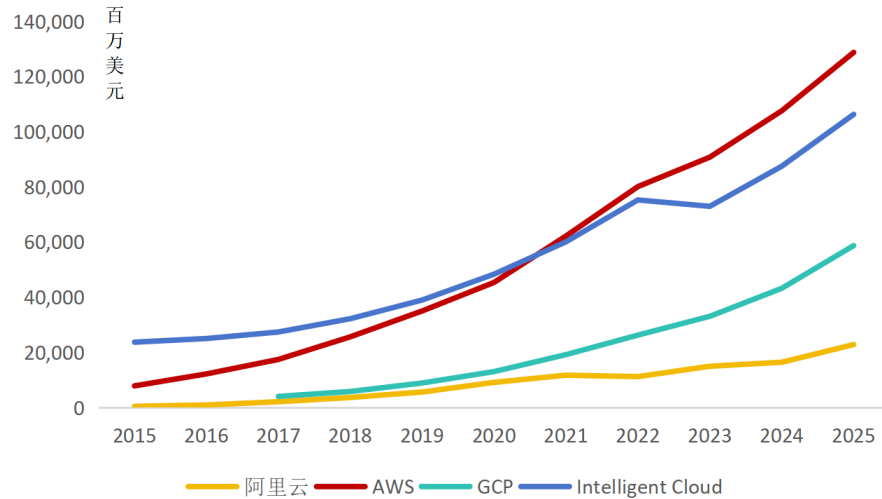
云厂商	MaaS 平台
AWS	Bedrock
谷歌云	Vertex AI
微软 Azure	Azure AI Studio、Azure OpenAI Service
阿里云	百炼
火山引擎	火山方舟
百度智能云	千帆 AppBuilder

资料来源：36 氪、浙商证券研究所

4.1.2 规模效应继续发挥效益，中国云厂商竞争格局改善

2015-2025 年，中美云厂商的业务体量持续扩张，且以阿里巴巴为代表的中国云厂商开始出海进一步扩张规模，预期规模效应带来的利润率差距会减少。

图22: AWS、智能微软云（Intelligent Cloud）、谷歌云和阿里云的业务规模不断扩大

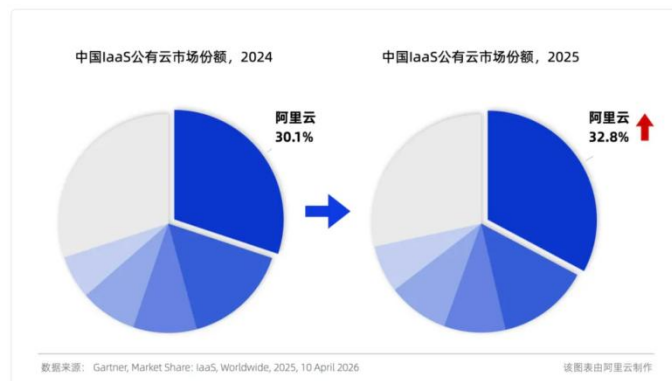


资料来源: Wind、公司公告、浙商证券研究所

备注: 阿里巴巴、微软按照各自财年核算，图中标注的 2025 年分别对应阿里巴巴 FY2026 (2025.4.1-2026.3.31)、微软 FY2025 (2024.7.1-2025.6.30)，两家企业历年财务数据均遵循该统计口径。

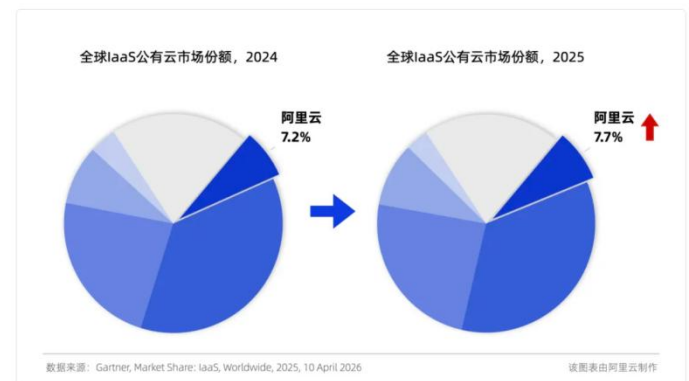
Gartner 数据显示，2024 年至 2025 年，阿里云在中国 IaaS 公有云的市场份额从 30.1% 上升至 32.8%，在全球 IaaS 市场份额也从 7.2% 上升至 7.7%。本土底层算力底盘不断夯实，出海布局推动全球业务体量稳步扩容，阿里云形成了在本土市场的规模优势，在全球范围内形成追赶态势，规模效应带来的利润率差异未来预期会减少。

图23: 阿里云中国 IaaS 公有云市场份额 (%)



资料来源: Gartner、浙商证券研究所

图24: 阿里云全球 IaaS 公有云市场份额 (%)



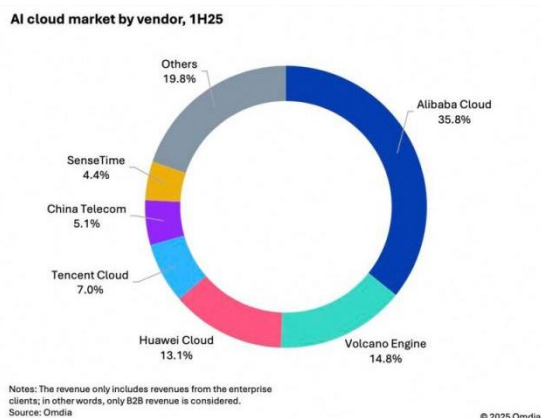
资料来源: Gartner、浙商证券研究所

在 AI 云市场上，2025 年国内竞争格局持续优化，行业集中度显著攀升：

1) 阿里云市场份额由 35.8% 提升至 38.1%，龙头壁垒持续夯实；2) 零散中小厂商合计占比从 19.8% 大幅收缩至 9.4%，尾部玩家加速出清，同质化低价内卷逐步消退。3) 大模型能力成为云厂必备能力——2025 年前 4 玩家合计 74.2% 市场份额，均为头部大模型厂商。

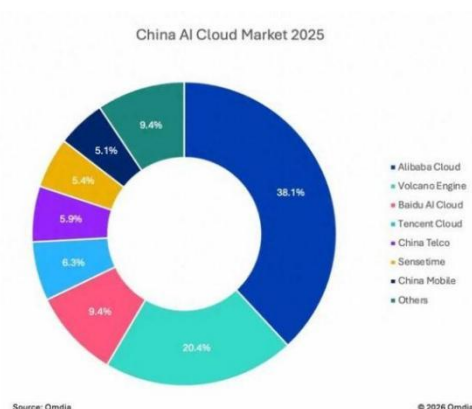
Omdia 发布的数据表明，国内云计算赛道差异化分工愈发清晰，行业马太效应凸显。预期未来阿里云等头部厂商将依托持续扩大的市场体量不断摊薄算力采购、机房运维、底层研发等固定成本，正向兑现算力基建、调度体系、生态服务层面的规模效应，规模优势将顺利转化为成本与盈利壁垒。

图25：中国云厂商 25 年 H1 智能云市场份额（%）



资料来源：Omdia、浙商证券研究所

图26：中国云厂商 25 年智能云市场份额（%）



资料来源：Omdia、浙商证券研究所

4.1.3 基于大模型的企业应用以 MaaS 模式落地，商业模式更偏好公有云

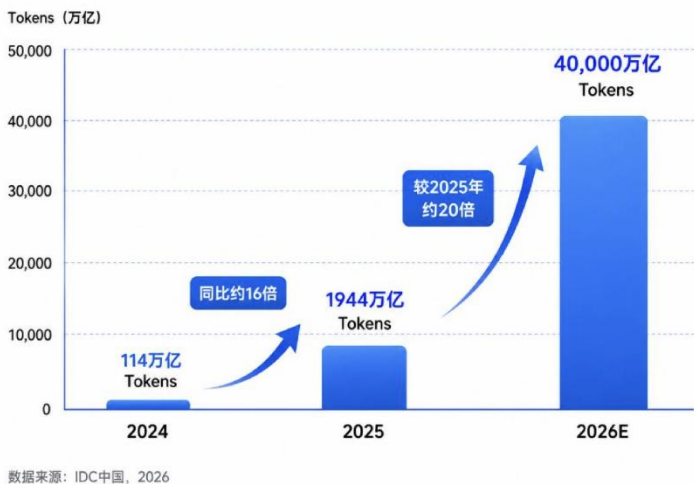
当前基于大模型的企业级应用以 MaaS 模式为主要落地路径，并主要依托公有云平台部署。

从需求端看，大模型训练与推理需要大规模 GPU 集群、数据资源以及模型运维能力，企业自行建设基础模型基础设施面临较高的资本投入和技术门槛，因此企业更倾向于通过云厂商提供的模型服务获取 AI 能力。

从供给端看，当前国内主流 MaaS 平台火山方舟、阿里云百炼、百度云千帆等核心基础均搭建在自身公有云基础设施之上，主打公有云标准化调用服务。通过 API 调用和模型托管方式向企业开放基础模型能力，是当前企业获取生成式 AI 能力的重要商业入口。

IDC 的数据表明，2025 年中国公有云大模型调用量已达 1944 万亿 Tokens，相比 2024 年同比增长 16 倍，预计 2026 年全年 Token 消耗量将增长约 20 倍，升至约 40000 万亿。Token 消耗量是衡量 MaaS 业务增长的重要指标之一，伴随企业级 AI 应用规模化落地渗透，公有云算力租赁、大模型调用服务需求将持续攀升，国内公有云整体市场规模有望进一步扩张。

图27： 预期 2026 年中国公有云大模型调用 Token 消耗量将升至约 40000 万亿



资料来源：IDC、浙商证券研究所

4.2 投资建议

我们看好 GPU 云时代下云厂商价值链扩张带来的长期投资机会。生成式 AI 推动云计算竞争从单纯资源供给转向“算力基础设施、模型服务和企业 AI 平台”的综合竞争，云厂商能够凭借 GPU 资源池、数据中心规模和 AI 生态优势，持续向价值链上游模型服务和下游企业应用拓展，提升单位资源价值和盈利空间。

建议关注：亚马逊、微软、谷歌、阿里巴巴等。

表9： 各大厂商盈利预测（单位：百万美元）

代码	公司	现价 (USD)	2026 Adj. 营收	2027 Adj. 营收	2028 Adj. 营收	2026 Adj.净利润	2027 Adj.净利润	2028 Adj.净利润
AMZN US Equity	亚马逊	232.11	824893.5	934733.7	1040533.4	114032.0	129416.9	162111.3
MSFT US Equity	微软	381.70	329510.3	384947.4	455272.2	126502.2	144888.8	169303.7
GOOGL US Equity	谷歌	319.74	432325.9	537826.3	645185.5	217669.2	187958.5	223061.0
9988 HK Equity	阿里巴巴	112.14	165878.1	185873.9	210175.5	15802.6	21460.0	26994.2

资料来源：Bloomberg，浙商证券研究所

备注：数据日期为 2026/7/24 日收盘价，盈利预测与估值统一使用 Bloomberg 一致预期；阿里巴巴、微软按照各自财年核算，图中标注的 2026 年分别对应阿里巴巴 FY2027（2026.4.1-2027.3.31）、微软 FY2026（2025.7.1-2026.6.30），两家企业财务数据均遵循该统计口径。

5 风险提示

1、产业竞争加剧：科技巨头争相投入资本开支抢市场份额，如果出现某个厂商发起价格战，则可能在某个阶段全行业增收不增利。

2、AI 应用商业化落地不及预期：目前仍处于 AI 产业发展初期，MoE 模型的发展一定程度降低了大模型的使用门槛，但 AI 应用商业化落地仍需较大的资本开支投入，在短期 AI 应用商业化落地有可能不及市场预期。

3、监管政策约束：云厂商需遵守本国及全球多区域监管规则，叠加地缘局势持续扰动，数据安全、跨境数据流动、算力出口管制、反垄断等监管政策处于动态调整阶段。跨境经营壁垒持续抬升，合规投入不断增加，若企业无法及时适配各地监管要求，或将遭遇海外业务扩张受阻、合规处罚、项目推进不及预期等问题，进而对经营业绩形成负面影响。

4、技术变革替代：云计算、AI 底层技术持续快速演进，芯片、系统架构、模型范式不断革新。技术路线存在不确定性，若厂商未能精准把握技术演进方向、产品迭代节奏滞后，存量软硬件资产可能加速贬值，现有业务面临新技术方案替代冲击，削弱竞争壁垒。

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 买 入：相对于沪深 300 指数表现 + 20% 以上；
2. 增 持：相对于沪深 300 指数表现 + 10% ~ + 20%；
3. 中 性：相对于沪深 300 指数表现 - 10% ~ + 10% 之间波动；
4. 减 持：相对于沪深 300 指数表现 - 10% 以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 看 好：行业指数相对于沪深 300 指数表现 + 10% 以上；
2. 中 性：行业指数相对于沪深 300 指数表现 - 10% ~ + 10% 以上；
3. 看 淡：行业指数相对于沪深 300 指数表现 - 10% 以下。

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。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者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。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、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或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。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。

浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层

北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层

深圳地址：广东省深圳市福田区广电金融中心 33 层

上海总部邮政编码：200127

上海总部电话：(8621) 80108518

上海总部传真：(8621) 80106010

浙商证券研究所：https://www.stocke.com.cn