

AI算力增长与ROI变现测算 考虑芯片、电力、建设能力与资本开支约束

行业研究 · 海外市场专题

互联网 · 互联网 II

投资评级：优于大市（维持）

证券分析师：张伦可

0755-81982651

zhanglunke@guosen.com.cn

S0980521120004

证券分析师：刘子谭

liuzitan@guosen.com.cn

S0980525060001

证券分析师：张昊晨

zhanghaochen1@guosen.com.cn

S0980525010001

- 根据供应链产能与各家增长指引，我们预计2024-2030年全球AI加速器总出货量从736万颗增长至4500万颗，CAGR约35%。对应总功耗需求快速攀升，2023年-2027年每年芯片出货量对应功耗翻倍增长，2030年预计达到152GW。同时，每GW的建设成本快速提升，2027年主力出货产品超500亿美金/GW。当前算力租赁价格较高，今年出货的主力芯片GB系列、Trainium3、TPU V7等毛利率约60%。
- 英伟达表示Hyperscale约贡献了50%的数据中心收入，2026年合计资本开支近8000亿美元，测算全球AI资本开支今年约1.3万亿，2023-2026年累计约2.5万亿，2030年前有望达到3-4万亿美元。CoreWeave、Nebius等NeoCloud厂商快速崛起，通过GPU担保贷款、可转债等多元融资扩产。根据弹性测算，在450-500亿美金/单GW建设成本下，2027年约1.7万亿AI资本开支约可建设35GW-40GW数据中心。
- 从当前数据中心建设进度看，截至2026年上半年，北美累计装机容量约53GW，2023年以来新增近30GW（主要为AI数据中心），当前Hyperscale在北美投运约25GW数据中心，另外扩张激进在建25GW、规划超50GW。结构上，租赁占比逐渐提升，2025年租赁占新增容量提升至约40%、2026超45%。区域布局上，产能集中在弗吉尼亚、得州、俄亥俄等电力资源相对充裕的州。
 - ① 电力端：通过燃气轮机表后供电实现较多电力补充。北美电网年新增仅10-15GW，无法匹配AI算力需求，BTM（表后供电）的燃气轮机成为核心补充方式，2026、2027年燃气轮机可给北美贡献约8GW、14GW增量；燃料电池、核电为中长期补充。
 - ② 建设端：成为核心限制约束。数据中心全建设周期25-36个月，数据中心通电前三类现场安装，必须持证技工，是当前最严重的卡点。2026年上半年北美仅仅落地6.7GW，远低于芯片对应的理论需求。

综合芯片出货、北美电力、北美建设、资本开支四重约束测算，若假设全球实际建设GW受到资本开支规模限制。2027年全球实际落地算力约37GW，仅为芯片理论对应量的60%。若假设北美建设占比全球建设容量比例维持65%，因北美数据中心ROI显著更高，2027年全球实际落地算力约30GW，约芯片对应量的50%。**从27年开始数据中心建设GW远低于芯片出货量GW，甚至不足一半芯片能在当年转化为算力。**

- 云厂商端：微软、谷歌单GW AI云年收入约100-150亿美元/GW/年，亚马逊约67亿美元，差异源于产品结构与定价策略；AI云收入占云总收入比重快速提升，测算2026年谷歌AI云占比达42%。考虑早期CSP与大模型厂商的战略合作、打折供应，当前大模型厂商的平均算力成本80多亿美元/GW/年。
- 大模型厂商端：2026年Anthropic、OpenAI占用超10GW算力，测算每GW算力贡献约100亿美元收入/年（训练推理各占一半算力）。2026年两家公司合计算力支出超千亿美元，随规模效应推理业务毛利率逐步转正。若采用GB200或Rubin推理最新代模型，1GW年收入可达500、1000亿美元。

- 若按假设一全球实际建设GW受到资本开支规模限制，27、28年实际落地约37GW、46GW。若其中一半由SOTA模型厂使用（Anthropic、OpenAI），按当前的变现效率两家27、28年收入约3800、7000亿美元（假设新增算力70%为推理需求）。
- 若按假设二全球实际建设GW受到北美建设限制，27、28年实际落地约30GW、35GW。若其中一半由SOTA模型厂使用（Anthropic、OpenAI），按当前的变现效率两家27、28年收入约2700、4500亿美元（假设新增算力70%为推理需求）。其收入增长更符合公司指引，在当前增长下，北美数据中心仍处于紧平衡，算租价格和算力ROI可持续处于高位。
- **核心结论：**我们预计2027年起数据中心实际建设落地算力将显著持续低于芯片出货对应的理论算力，
 - ① **AI芯片与供应链订单承压，客户待建库存拉长：**芯片订单可能从“抢产能”转向“按项目交付”，供应链增长预期承压。部分芯片会以“待部署库存”形式积压在客户端，形成在建工程或闲置资产，库存积压或导致二级市场折价，并拉低云厂商/建设方ROI。
 - ② **通过“芯片+基建”捆绑模式，芯片厂担保资金、兜底收入后压货给承建方：**英伟达、谷歌等芯片厂商从卖芯片转向“芯片交付+建设担保+收入兜底”的打包方案。例如英伟达联合金融机构担保5000亿美元给承建方（多是新云），并为OpenAI 俄亥俄园区提供千亿级土建电力的残值担保，本质是打通下游环节、确保芯片出货。
 - ③ **算力可维持较高价格，北美区域溢价显著：**落地算力的稀缺性直接延后了市场预期的“算力价格战”，尤其北美算力现货预计能维持现有高ROI和高毛利水平。
- 风险提示：AI商业化落地不及预期导致算力ROI承压；高额发债推升融资成本，侵蚀利润；地缘与供应链约束导致芯片交付延期；芯片快速迭代引发大额资产减值；

表：每年加速器芯片对应GW与数据中心增量（单位：GW）						
	2025	2026E	2027E	2028E	2029E	2030E
芯片出货量对应GW数	13	33	62	105	129	152
假设一：北美区域建设能力为最小约束，建设全球实际建设GW受到资本开支规模限制，按照此前资本开支测算						
北美建设GW	9	16	20	23	26	30
全球建设GW	13	27	37	46	57	60
假设二：若北美建设占比全球建设容量比例维持65%，因北美数据中心ROI显著更高，测算全球算力与资本开支情况						
北美建设GW	9	16	20	23	26	30
全球建设GW	13	24	30	35	40	46

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理测算，详细测算过程见对应章节

请务必阅读正文之后的免责声明及其项下所有内容

- [01] 加速器芯片出货量与对应GW
- [02] 北美数据中心增长与大厂建设计划
- [03] 电力、建设限制与算力增长测算
- [04] 1GW建设成本与CSP、大模型厂收入变现

AI芯片出货量对应功耗2023年到2027年每年翻倍增长

- 根据供应链产能与各家增长指引，我们预计26/27年各类AI芯片出货量如右图，测算对应所需电力容量将分别达到33/62GW。（假设26年英伟达主要供应GB300，27年供应Rubin，谷歌26/27年分别供应TPUv7/v8）。
- 以GB300为例，NVL72一个机架包含72颗GB300 Blackwell Ultra GPU和36颗Grace CPU，整柜满载功耗最高约142kW（含柜内CPU、NVSwitch等主要组件），假设PUE=1.2，对应1GW数据中心大约需要42万张GB300。

图：预期GW数增长对应GB300需求（单位：万颗）

	总功率 (mw)	PUE	单机柜功率 (kw)	机柜数量	单机柜芯片数	所需芯片数
GB300	1000	1.2	142	5869	72	422,535
Rubin	1000	1.2	200	4167	72	300,000
AMD MI400	1000	1.2	130	6410	72	461,538
TPU v7	1000	1.2	90	9259	64	592,593
TPU v8	1000	1.2	150	5556	64	355,556
Trainium3	1000	1.2	67	12438	64	796,020
其他ASIC	1000	1.2	142	5869	72	422,535

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理

图：预期GW数增长对应GB300需求（单位：千颗）

加速器/千颗*	2024	2025	2026E	2027E	2028E	2029E	2030E
GPU	5,234	6,110	8,370	11,500	13,300	14,600	16,000
NVDA	4,850	5,700	7,800	10,500	12,000	13,000	14,000
AMD	384	410	570	1,000	1,300	1,600	2,000
ASIC	2,130	2,850	7,700	14,800	21,000	25,500	29,000
TPU	1,280	1,700	5,000	10,000	15,000	18,000	20,000
Trainium	750	750	1,700	3,300	4,000	5,000	6,000
其他ASIC	100	400	1,000	1,500	2,000	2,500	3,000
Total	7,364	8,960	16,070	26,300	34,300	40,100	45,000
YoY		22%	79%	64%	30%	17%	12%
占比							
GPU	71%	68%	52%	44%	39%	36%	36%
NVDA	66%	64%	49%	40%	35%	32%	31%
AMD	5%	5%	4%	4%	4%	4%	4%
ASIC	29%	32%	48%	56%	61%	64%	64%
TPU	17%	19%	31%	38%	44%	45%	44%
Trainium	10%	8%	11%	13%	12%	12%	13%
其他ASIC	1%	4%	6%	6%	6%	6%	7%
出货量对应所需GW数							
GPU			20	37			
NVDA			18.5	35.0			
AMD			1.2	2.2			
ASIC			13	25			
TPU			8.4	16.9			
Trainium			2.1	4.1			
其他ASIC			2.4	3.6			
总出货对应GW*	6	13	33	62	105	129*	152*

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理 *基于左图各家产品额定负载和PUE假设测算，假设2028年开始单芯片平均功率每年增长5%

各类芯片每小时TCO成本与1GW建设成本

图：GPU、TPU、Trainium芯片TCO比较

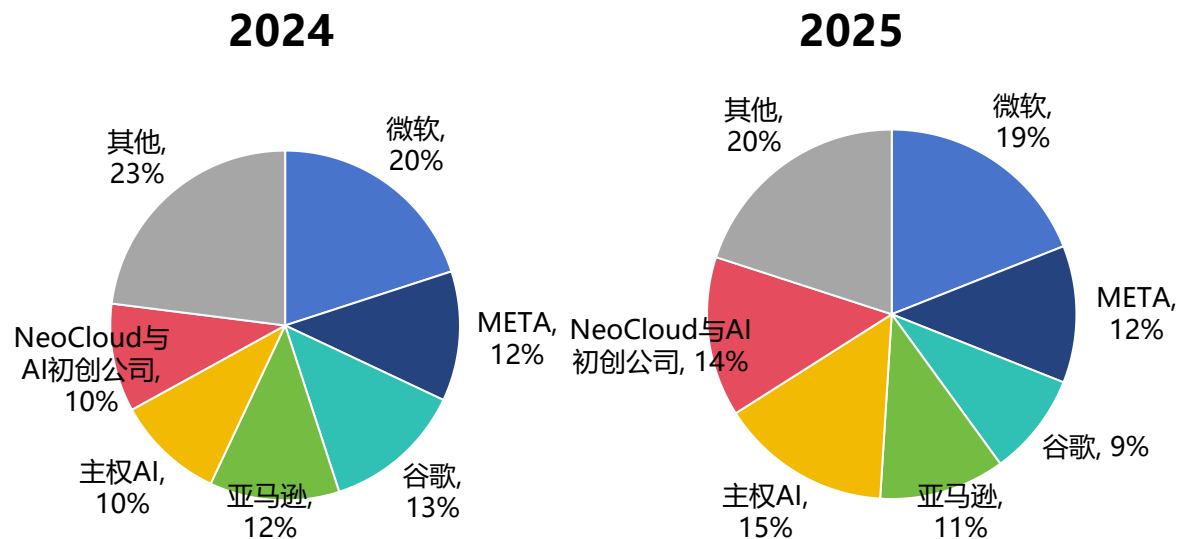
单位: 美元	单位	GB200NVL72(Spectrum)	GB300NVL72(Spectrum)	Trainium2 16卡	Trainium2 64卡	Trainium3 16卡	Trainium3 64卡	TPU V6 64卡	TPU V7 64卡
服务器成本	USD	\$3,179,652	\$3,997,587	\$154,020	\$655,793	\$170,554	\$693,183	\$1,280,000	\$1,920,000
服务器维护服务	USD	\$72,000	\$72,000	\$16,000	\$64,000	\$16,000	\$64,000	\$64,000	\$64,000
网络成本	USD	\$296,733	\$296,733	\$26,300	\$105,202	\$26,300	\$105,202	\$105,200	\$105,200
外置存储成本	USD	\$126,063	\$126,063	\$17,400	\$69,600	\$17,400	\$69,600	\$69,600	\$69,600
软件授权及其他成本	USD	\$20,677	\$23,057	\$2,448	\$9,838	\$2,912	\$11,684	\$12,000	\$15,000
其他安装成本	USD	\$72,000	\$72,000	\$16,000	\$64,000	\$16,000	\$64,000	\$64,000	\$64,000
单服务器总前期集群资本开支	USD	\$3,767,125	\$4,587,440	\$232,168	\$968,433	\$249,166	\$1,007,669	\$1,594,800	\$2,237,800
单服务器加速器芯片数	#	72	72	16	64	16	64	64	64
单芯片每小时资本成本	USD/小时/芯片	\$1.51	\$2.15	\$0.47	\$0.49	\$0.51	\$0.52	\$0.82	\$1.16
电力成本	USD/kWh	\$0.10	\$0.10	\$0.10	\$0.10	\$0.10	\$0.10	\$0.10	\$0.10
每月每kW总托管成本	USD/kW/月	\$185.00	\$185.00	\$185.00	\$185.00	\$185.00	\$185.00	\$185.00	\$185.00
整机功耗	kW	143	160	12	50	16	63	45	64
每月服务器托管成本	USD/月	\$26,464	\$29,609	\$2,305	\$9,280	\$2,917	\$11,718	\$8,288	\$11,840
远程运维与支持工程师成本	USD/月	\$1,179	\$1,179	\$262	\$1,048	\$262	\$1,048	\$1,048	\$1,048
网络连接成本	USD/月	\$351	\$351	\$78	\$312	\$78	\$312	\$312	\$312
单服务器每月总运营成本	USD/月	\$27,994	\$31,139	\$2,645	\$10,640	\$3,257	\$13,078	\$9,648	\$13,200
单芯片每月运营成本	USD/月/芯片	\$389	\$432	\$165	\$166	\$204	\$204	\$151	\$206
单芯片每小时运营成本	USD/小时/芯片	\$0.54	\$0.60	\$0.23	\$0.23	\$0.28	\$0.28	\$0.21	\$0.29
单芯片每小时成本 TCO	USD/小时/芯片	2.28	2.72	0.69	0.71	0.78	0.79	1.02	1.43
1GW 数据中心对应服务器数量	MW	5,108	4,566	58,649	14,569	46,339	11,537	16,312	11,418
1GW 数据中心集群建设成本 (服务器占比50%)	亿美元	385	419	272	282	231	233	520	511
单芯片租赁价格*	USD/小时/芯片	6.5	8.5	1.54	1.54	1.95	1.95	1.9	5.1
毛利率		68%	68%	55%	53%	59%	59%	46%	72%
1GW芯片年租金	亿美元	209	245	127	126	127	126	174	326

资料来源：SemiAnalysis、国信证券经济研究所整理测算 *租赁价格为散单市场Reserved 1-3年预留租赁价格，未考虑折扣因素

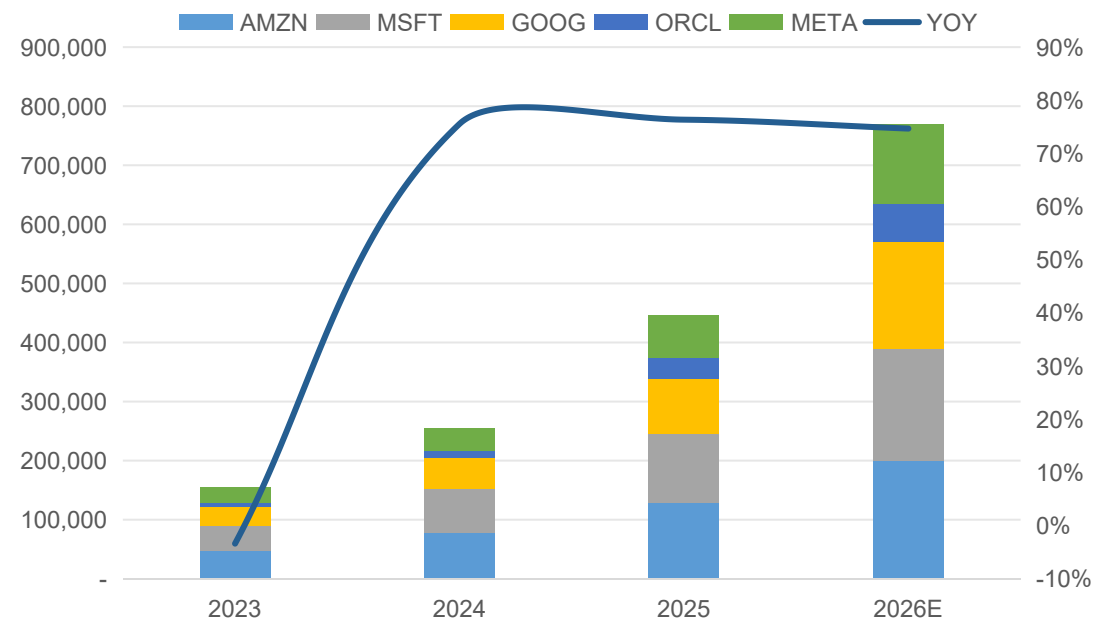
请务必阅读正文之后的免责声明及其项下所有内容

- 海外Hyperscale（北美四大CSP+Meta）2023年到2026年资本开支分别约为1555、2558、4250、7550亿美元，23-25年累计投入8363亿美元，若假设AI占比75%、则AI资本开支为7300亿美元，与26年Hyperscale的AI资本开支接近。
- 英伟达26FYQ4（截至2026年1月）业绩会表示Hyperscale约贡献了50%的数据中心收入，测算全球AI资本开支今年约1.3万亿，2023-2026年累计约2.5万亿。英伟达26FYQ2业绩会表示2025年全球数据中心基础设施投入预计6000亿美元（两年内翻倍），2030年前有望达到3-4万亿美元。

图：海外科技厂贡献英伟达数据中心芯片收入占比测算



图：各互联网与云厂资本开支变化情况（百万美元、%）



资本开支与GW数弹性测算



- 测算明年Hyperscale资本开支约1.1万亿美元，英伟达26FYQ4（截至2026年1月）业绩会表示Hyperscale约贡献了50%的数据中心收入，测算全球AI资本开支约1.7-1.8万亿。
- 根据弹性测算，在450-500亿美金/单GW建设成本下，2027年约1.7万亿AI资本开支约可建设35GW-40GW数据中心；
- 英伟达27FYQ2业绩会表示与Apollo、贝莱德、黑石等六家机构合作，计划撬动5000亿美元第三方资本用于AI基础设施建设；并对OpenAI所在俄亥俄PORTS Pike算力园区提供担保上限1050亿美元，园区首期4.25GW，单代可部署约150万GPU，20年租期。

表：2027年 数据中心容量、建设成本与资本开支弹性测算（亿美元/GW）

单GW建设成本（亿美元）/全球AI数据中心容量（GW）	350	400	450	500	550
25	8,750	10,000	11,250	12,500	13,750
30	10,500	12,000	13,500	15,000	16,500
35	12,250	14,000	15,750	17,500	19,250
40	14,000	16,000	18,000	20,000	22,000
50	17,500	20,000	22,500	25,000	27,500
60	21,000	24,000	27,000	30,000	33,000

表：北美科技公司经营性现金流与债务对应

亿美元	经营性现金流	净债务杠杆率 1.2倍对应债务	每年新增 0.35x债务
2026	6,500	7,800	2,275
2027	8,125	9,750	2,844
2028	10,156	12,188	3,555
2029	12,695	15,234	4,443
2030	15,869	19,043	5,554
总计	53,346		18,671

资料来源：英伟达官网、英伟达业绩会，国信证券经济研究所整理测算

资料来源：公司官网、国信证券经济研究所整理测算（采用彭博一致性预期，假设经营性现金流年化增速25%，每年资本开支为8倍经营性现金流的1.35x）

Hyper scale的资本杠杆: 预计2026年发债3000亿, 当前净债务约为0



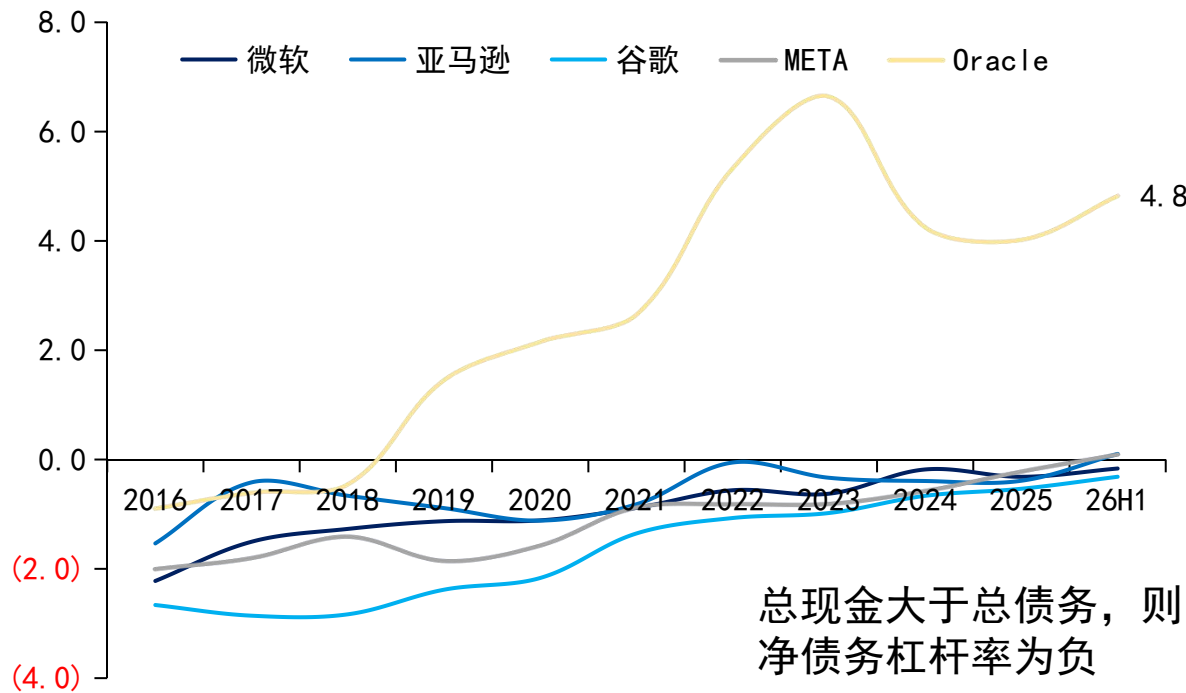
- 2025年Hyperscale全年发行1210亿美元公司债, 2026年前5个月已发1590亿美元, 根据华尔街日报预期全年预计2500 - 3000亿美元。
- 当前五大Hyperscaler平均总债务/EBITDA约1.3x, 26H1净杠杆约0.3x (不考虑Oracle约0附近), Oracle为个例, 净杠杆约4.8x。
- 对标IG基准: 北美IG整体中位净债务/EBITDA, A级: 1.24x、BBB级: 2.38x; 投资级中位利息保障倍数(EBITA/当期利息支出)5.8x。

净债务杠杆率= 净债务 /EBITDA

= (短期债务+长期债务-总现金)

(营业利润+折旧摊销)

图：海外互联网净债务杠杆率变化情况



表：主要北美科技公司发债情况（亿美元）

时间区间	Amazon (AWS)	Microsoft	Alphabet	Meta
26H2总债务	1219	402	795	837
26H2总现金	1067	670	1268	750
2025-2026年6月累计发行总额	820 亿美元以上	约 420 亿美元	约815 亿美元	约550 亿美元
2025年Q4	单笔 150 亿美元多期限美元债（5-30 年期，票面利率 4.6%-4.9%）	两笔合计 145 亿美元，增设超长 30 年期品种	250 亿美元美元债	BlueOwl 合资 SPV 发行 270 亿美元债务（表外负债，Meta 仅持股 20%）
2026年	500 亿美元主力美元债，初始询价 370-420 亿；140 亿加元等值 100 亿美元加元计价债券，同步发行欧元中长期债券	单笔 70 亿美元美元优先无抵押债券 90 亿 + 欧元等值 20 亿，合计 110 亿美元	315 亿美元大额债券，内含 10 亿英镑 100 年期超长债（行业稀缺品种）	追加发行 250 亿美元中长期债券，配套数十亿美元私募信贷额度

资料来源：公司财报、彭博、国信证券经济研究所整理

资料来源：公司官网、彭博社、华尔街日报，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

- NeoCloud融资已不再是早期的少量私募债或设备租赁，而是形成了“GPU及合同支持贷款+可转债+项目融资/SPV+客户预付款+ATM/股权融资+厂商/服务器OEM融资”的多层资本结构。

- ① CoreWeave：大量高息担保DDTL（GPU资产抵押贷款）拉高综合利率；
- ② Nebius：依赖可转债（低现金票息，成本体现在到期赎回溢价/有效利率），债务加权成本更低。

表：公司核心业绩对比

公司	CoreWeave	Nebius
RPO	Q2末1040亿美元，同比+246%；Q3初新增超250亿美元净新客户承诺 Q2末336亿美元，同比+364%；29%合约收入未来24个月兑现	
资本开支	Q2资本开支94亿美元、同比218%；全年指引350-390亿美元（前值300-350亿）	Q2资本开支56.6亿美元、同比1008%；维持全年指引200-250亿美元
债务情况	截至2026年6月30日总债务约350亿美元，包括： DDTL 约150亿美元（包括GPU担保融资），可转债66亿美元，高收益债100亿美元。	截至2026年6月30日总债务约150亿美元，绝大部分为可转换高级票据（2029、2031 到期），无大额传统抵押银团贷。 并在2026年8月新增45亿美元可转债发行与7.75亿美元资产担保融资
26Q2净债务杠杆率	约4x	约5x
26Q2负债利率	26Q2测算综合加权利率≈8.4%、过去一年加权债务成本下行近300bp；（区间：老DDTL最高15%、存量高级债9%-11%、可转债约2%；新DDTL4.0固定档5.9%、浮动档SOFR+2.25%）	整体综合有效债务成本约4%~6%，主体负债以低票息可转债为主（票面0.5%~3%，但要计入折价摊销，有效利率显著高于票面）；新增7.75亿资产担保贷款定价SOFR+2.5%。
算力情况	Q2末活跃算力1.5GW，Q2末签约3.7GW（8月更新至4.2GW）；2026年末活跃算力超1.85GW；2030年目标8GW；海外签约超1GW	2026年算力0.8-1GW、签约电力≥4GW；2027年起每年稳定新增1GW以上算力 2027年电力新增3.4GW；年末签约电力产能上调至5GW；

资料来源：各公司业绩会、彭博社、华尔街新闻，国信证券经济研究所整理测算

- 英伟达当前对Neocloud的支持已经明显超出传统供应商金融。已有合作形式显示，这套机制至少包含四类形式。其中**英伟达与大模型厂商的合作，往往引入新云作为数据中心的承建方和所有方**（大模型厂商受限于现金流和人力，往往租赁数据中心代替建设，不参与ROE较低的数据中心建设任务）。

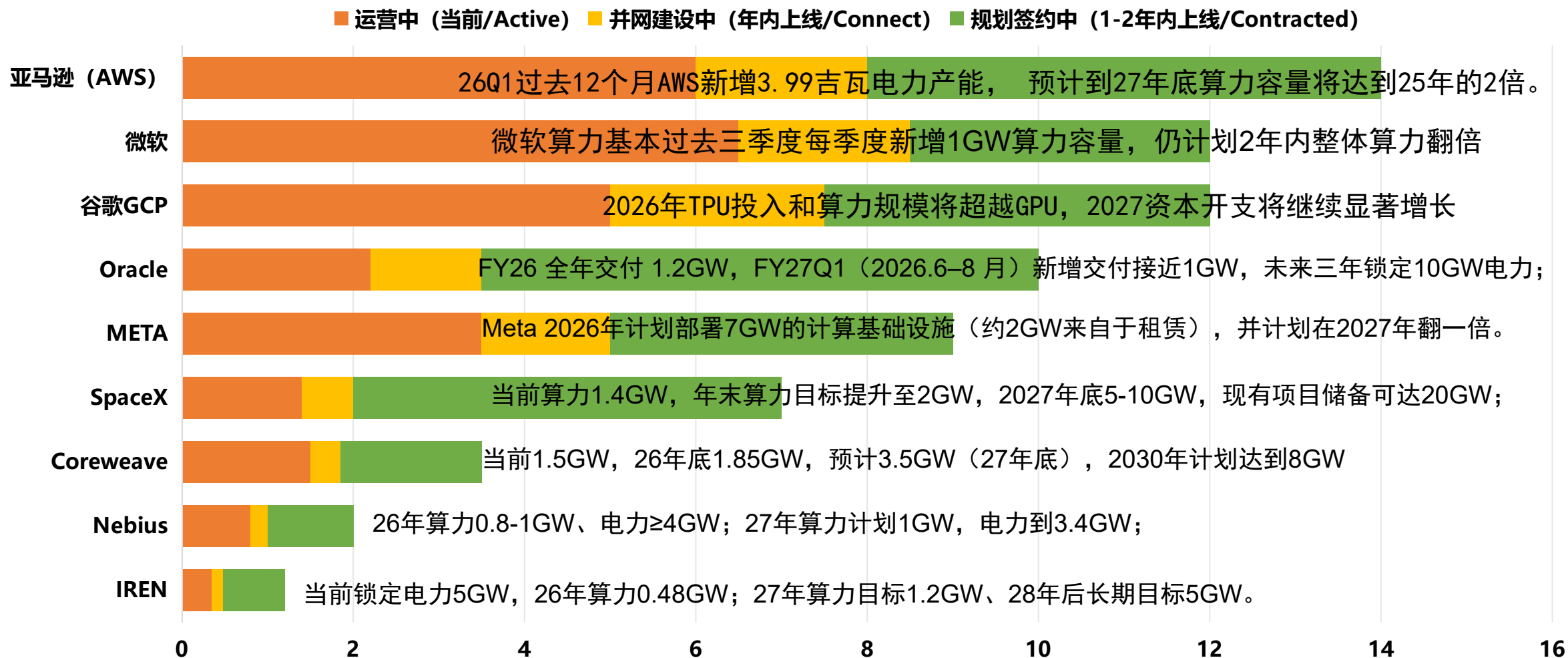
表：英伟达与新云和大模型厂相关的资本操作

说明		与Neocloud和大模型相关
股权投资/认股权证	英伟达现金入股新云，拿到少数股权（一般是战略 minority stake），锁定GPU优先供货权、绑定DSX架构落地、参与公司长期价值增值。	- Nebius：英伟达20亿美元入股，获得约8.3%股权。 - CoreWeave：1亿美元战略入股+配套供货/兜底协议。 - IREN：34亿美元云合同+21亿美元购股权。 - OpenAI：300亿美元股权投资，随算力容量落地分批出资。 - Anthropic：100亿美元股权投资，联合微软共同投资。锁定初始1GW 算力，通过Lambda、Nscale等第三方算力运营商交付。
未售出算力回租或兜底机制	- 若对外市场化出租的利用率不足，英伟达承诺按预先约定单价，回租这部分闲置GPU算力； - 租金≤保底价，触发兜底：英伟达补足差额，保证新云拿到最低收入；市场租金>保底价：超出保底价价格的超额收入，英伟达按约定比例收入分成	SharonAI（澳大利亚）：6年期协议，最多部署4万GB300；如果Sharon对外客户租不掉，英伟达按合同价格租用。依靠这个条款，Sharon完成16亿美元融资，落地DSXAI工厂。
平台化信用担保	英伟达作为设备供应商，向债权方（黑石、KKR、高盛等机构资金）出具平台层面信用背书，证明项目算力需求、GPU供给、架构合规	8月10日联合Apollo、BlackRock、Blackstone、Brookfield、Goldman Sachs、KKR，目标调动5000亿美元 第三方资本 ，英伟达在部分项目上最多承担约25%残值支持/贷款切片担保。
三方残值担保与收入分成	英伟达用自身高信用评级，为AI数据中心「土地/电力/机房壳体（LPS）」做信用增信，换取20年独家算力独家供货权，把一次性芯片收入转型为硬件迭代+持续分成收入；	26Q2 英伟达业绩会，英伟达对OpenAI所在俄亥俄PORTS Pike算力园区提供担保上限1050亿美元（机房壳体+土地电力残值担保，不含GPU，为表外或有负债），单代可部署约150万GPU，20年租期。软银SBEnergy承建重资产，OpenAI作为核心租户。

资料来源：各公司业绩会、彭博社、华尔街新闻，国信证券经济研究所整理测算

- [01] 加速器芯片出货量与对应GW**
- [02] 北美数据中心增长与大厂建设计划**
- [03] 电力、建设限制与算力增长测算**
- [04] 1GW建设成本与CSP、大模型厂收入变现**

图：AI数据中心功率增长计划（GW）

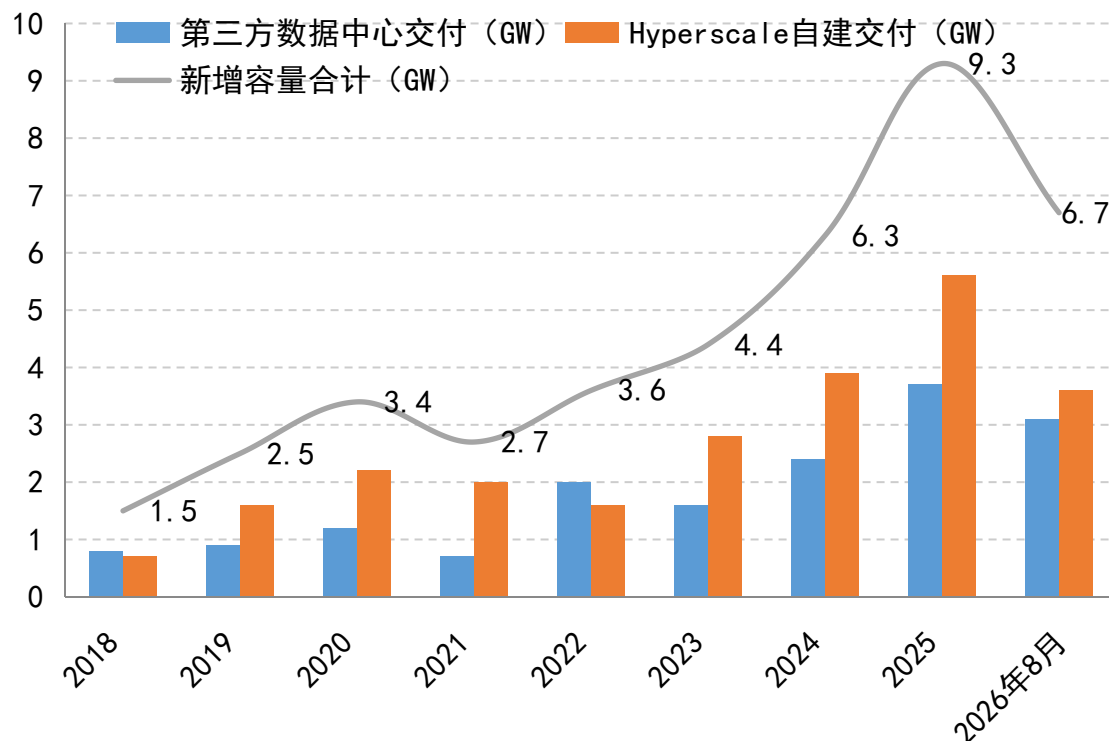


资料来源：各公司官网与业绩会、彭博社、Uptime Institute、华尔街新闻，国信证券经济研究所整理

北美数据中心年度新增和累计装机容量测算（包含传统数据中心）

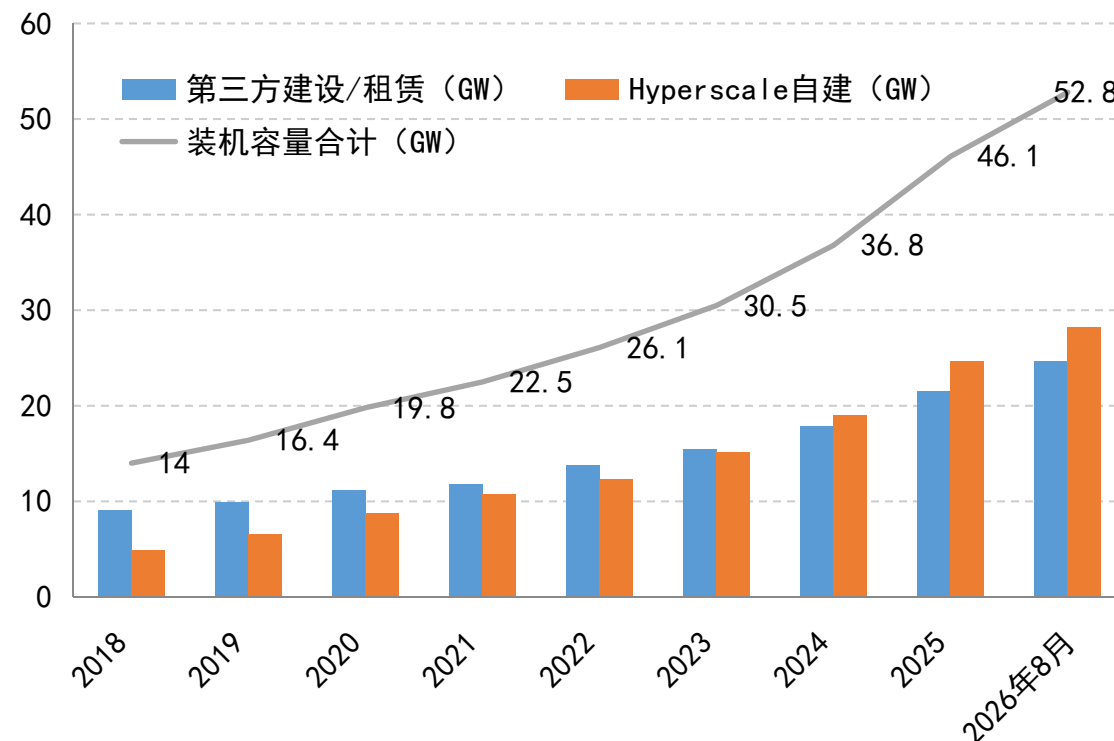
- 截止26H1北美数据中心累计装机容量约52GW（包含传统数据中心），受限于建设进度、26H1北美实际新增投产约6.5GW。
- **2023年至今，北美数据数据中心增长近30GW（主要为AI数据中心）、过去三年每年新增容量复合增速超40%。**
- 第三方交付/租赁占比持续提升。大模型技术发展早期，2023-2024年租赁占新增容量约35%，以Hyperscale自建为主、满足模型训练需求。伴随算力需求转向推理、算力ROI更为明确时，2025年租赁占新增容量提升至约40%、2026超45%。

图：北美数据中心新增容量（GW）



资料来源：Aterio、UptimeInstitute、华尔街新闻，国信证券经济研究所整理测算

图：北美数据中心累计装机容量（GW）

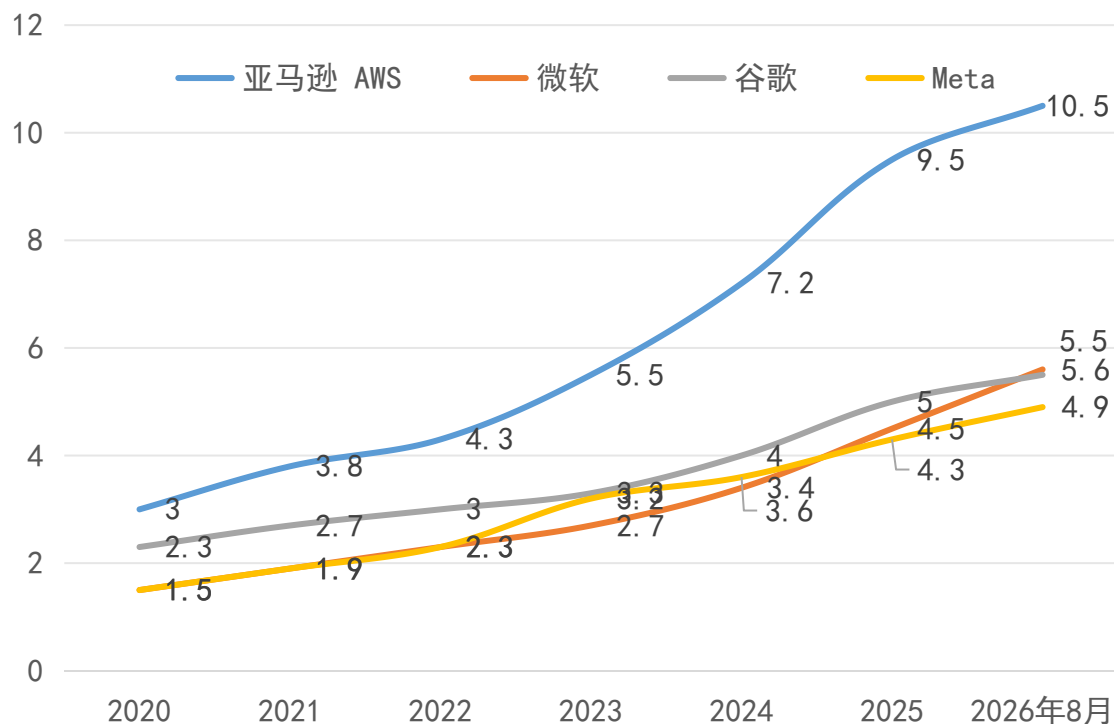


资料来源：Aterio、UptimeInstitute、华尔街新闻，国信证券经济研究所整理测算

北美数据中心容量变化（包含传统数据中心）

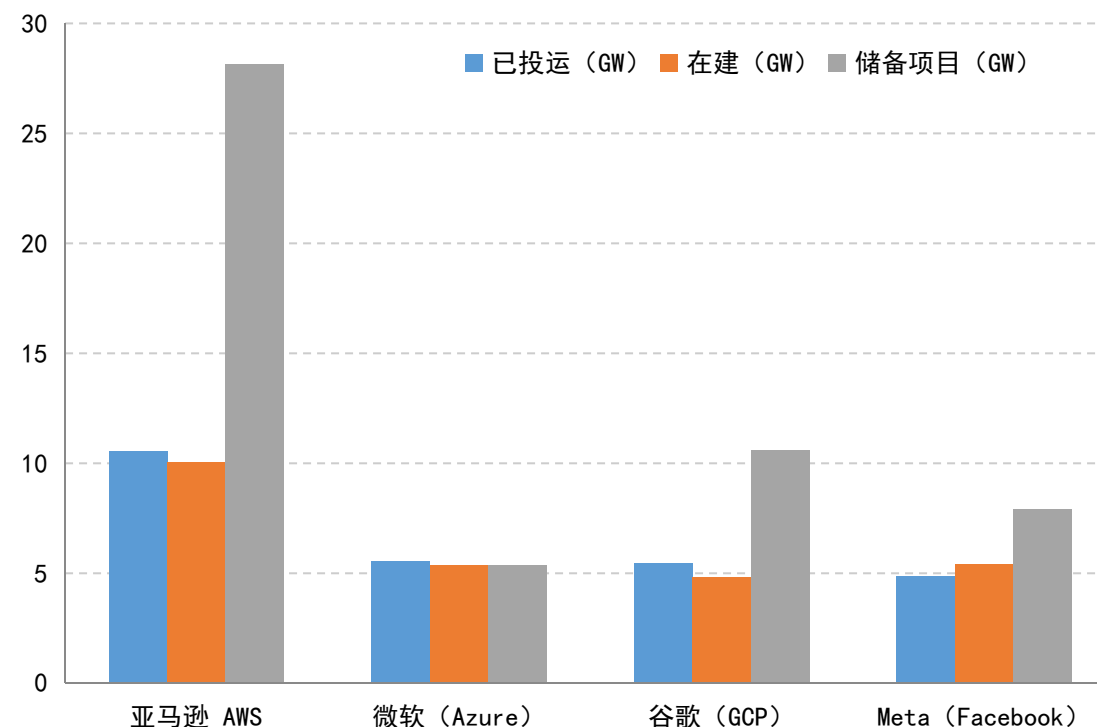
- 当前Hyperscale在北美投运约25GW数据中心、占据北美半壁江山。另外在建25GW、规划超50GW。2026Q2，MSFT、AWS、GOOGL业绩会披露 未履约云订单RPO超1.5万亿、同比翻倍以上增长，微软RPO加权平均履约周期2.3年，Hyperscaler签约交付周期将由当前12-18个月拉长至24个月以上。
- 亚马逊AWS在北美拥有超10GW的数据中心(含传统数据中心)、超25GW在建，在Hyperscale中占比最高且在过去三年增长显著。亚马逊2025Q4业绩会披露，过去12个月新增3.9GW电力容量，预计到27年年底将翻倍，仅25Q4就新增1.2GW。

图：Hyperscaler北美数据中心容量变化（GW）



资料来源：Aterio、omdia、公司业绩会，国信证券经济研究所整理测算

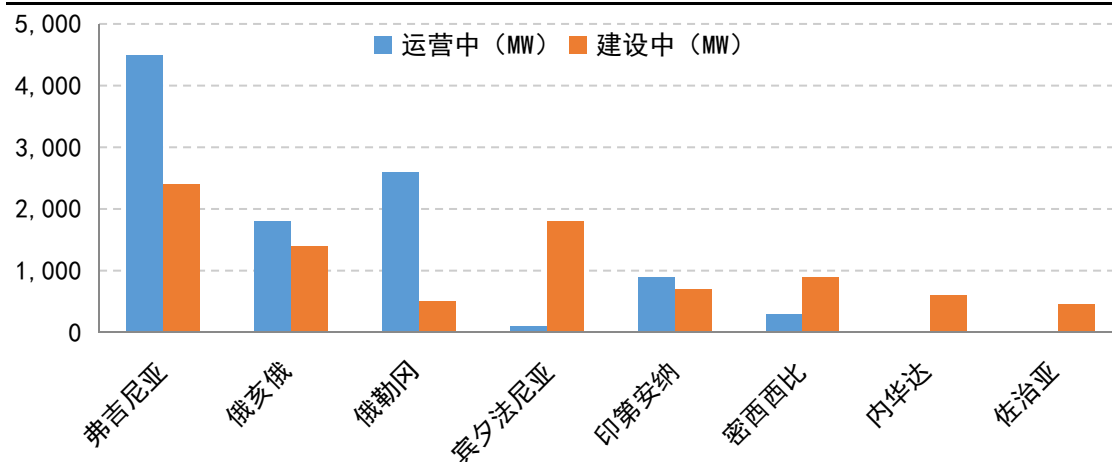
图：北美大科技公司各阶段容量（GW）



资料来源：Aterio、omdia、公司业绩会，国信证券经济研究所整理测算

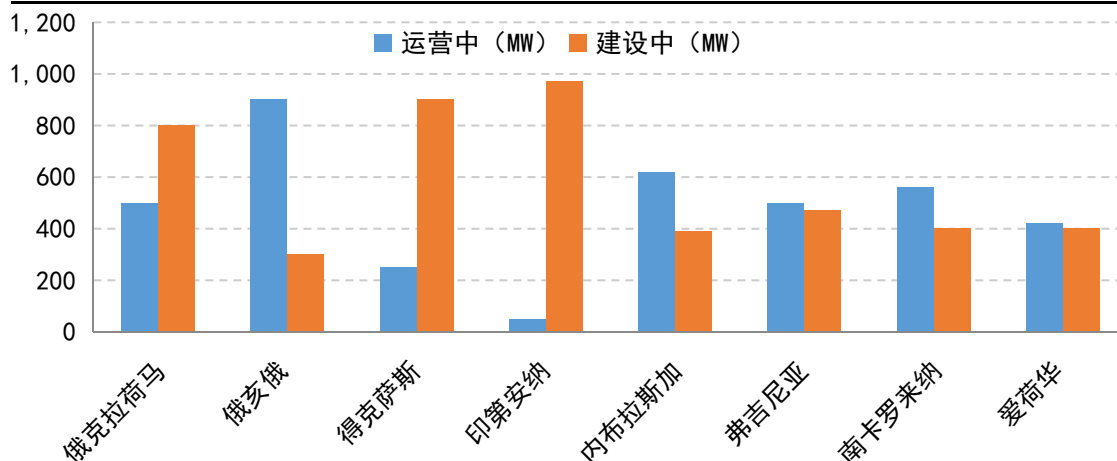
当前Hyperscale北美主要数据中心

图：亚马逊AWS主要州运营中与建设中容量（MW）



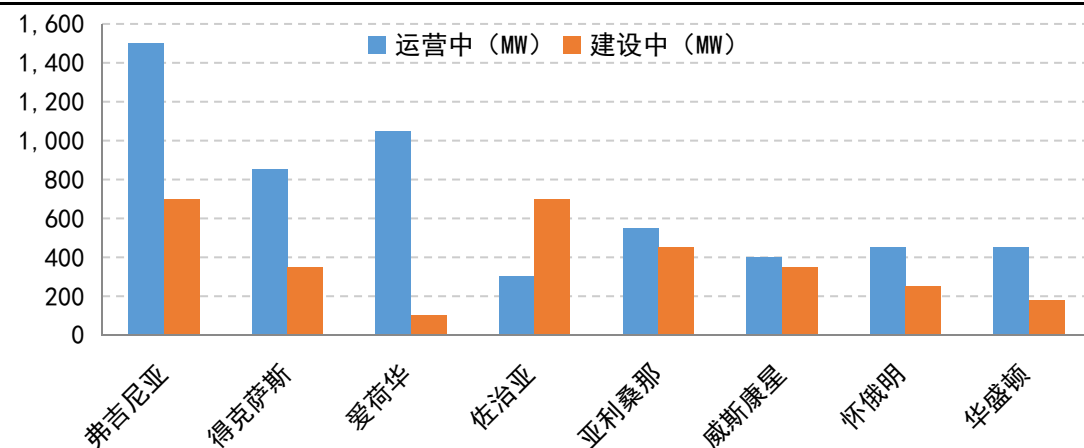
资料来源：公司官网，华尔街新闻，国信证券经济研究所整理测算

图：谷歌（GCP）：主要州运营中与建设中容量（MW）



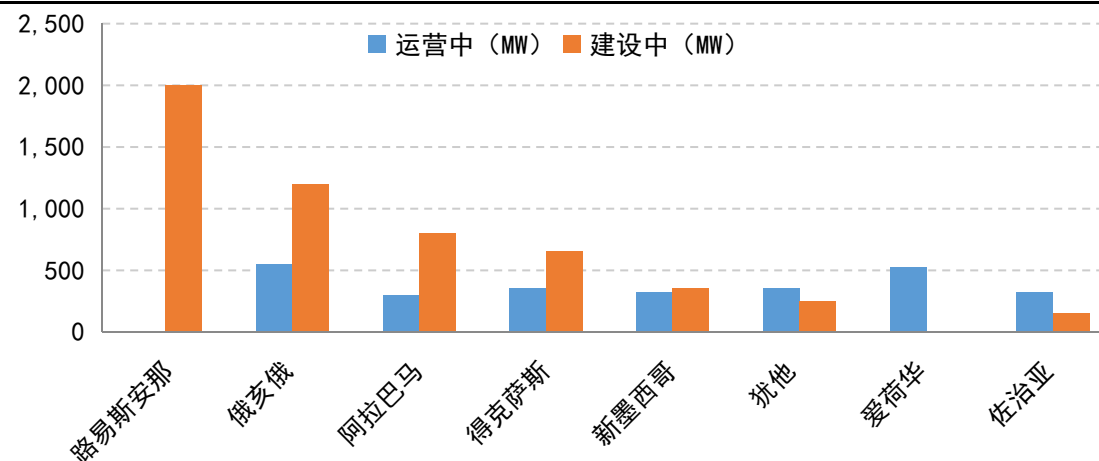
资料来源：公司官网，华尔街新闻，国信证券经济研究所整理测算

图：微软（Azure口径）：主要州运营中与建设中容量（MW）



资料来源：公司官网，华尔街新闻，国信证券经济研究所整理测算

图：Meta：主要州运营中与建设中容量（MW）

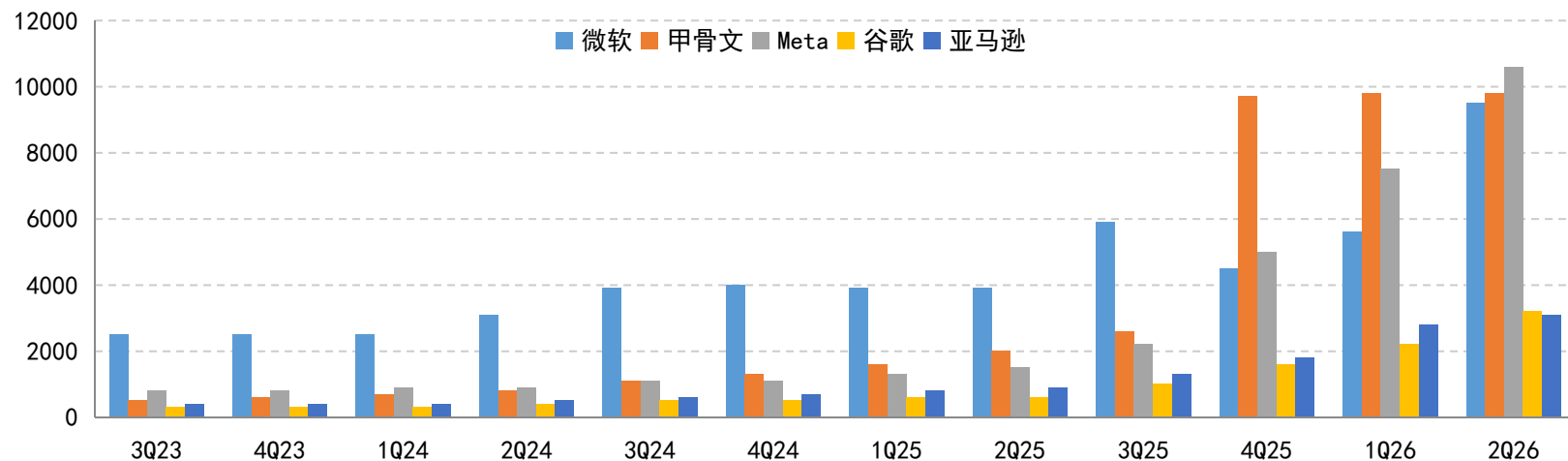


资料来源：公司官网，华尔街新闻，国信证券经济研究所整理测算

2026年Hyperscale外部租赁承诺增长超14GW

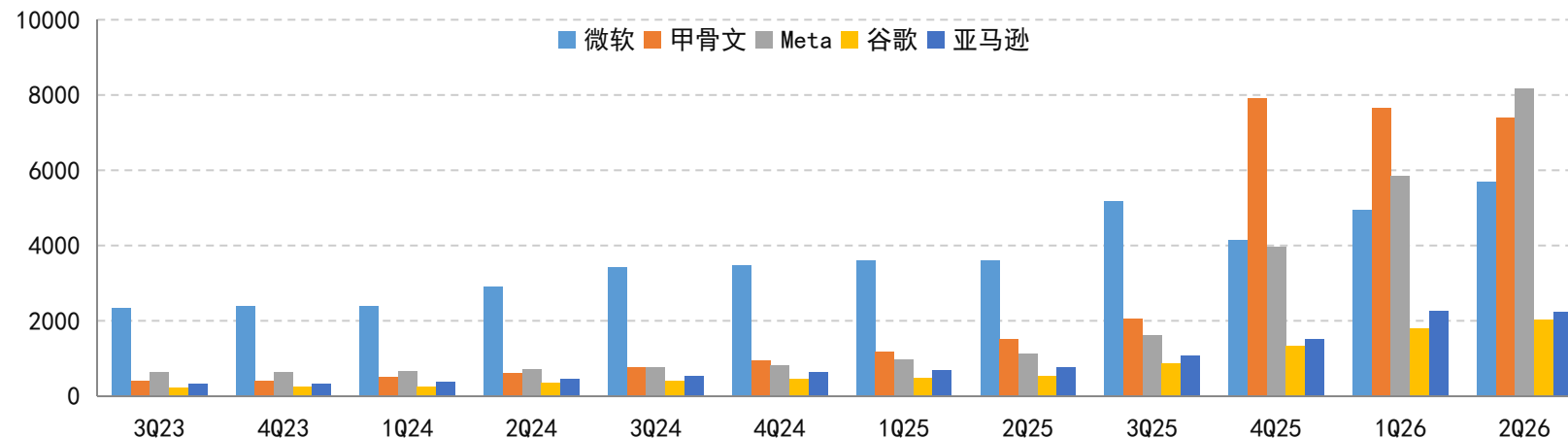
- 26Q2除微软保持正向现金流外，META接近0、其余谷歌、亚马逊、Oracle等科技巨头现金流已转负，为了缓解现金流压力更多采用租赁实现算力增长。
- 截止2026Q2，测算五家合计租赁承诺超35GW，其中北美区域约25GW。微软、Oracle、META的租赁承诺均约10GW，亚马逊、谷歌租赁承诺约3GW。
- 2026年，Hyperscale租赁承诺增长约14GW、其中北美约增长7GW。
- 截止26Q2，NeoCloud如CoreWeave、Nebius、IREN等在手待履约订单快速增长到约2000亿美元，约是今年云收入的十倍。

图：全球超大规模企业数据中心租赁承诺



资料来源：Aterio、omdia、公司业绩会，国信证券经济研究所整理测算

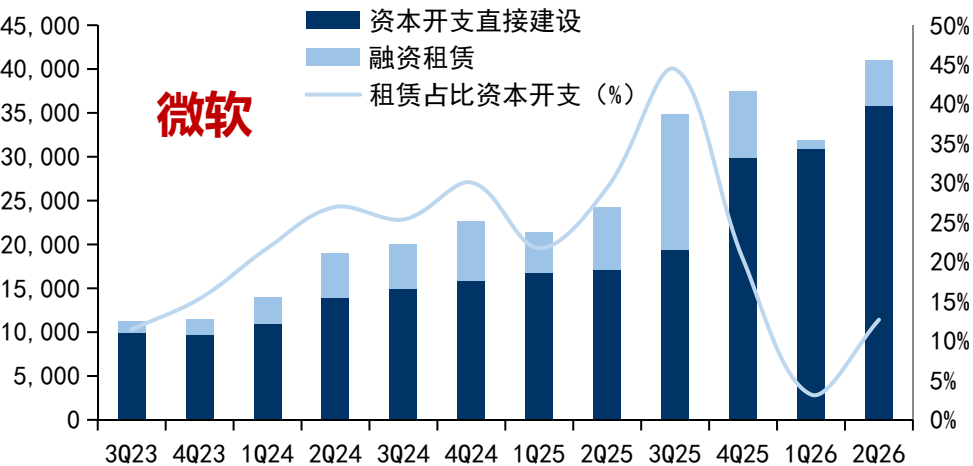
图：北美数据中心租赁承诺



资料来源：Aterio、omdia、公司业绩会，国信证券经济研究所整理测算

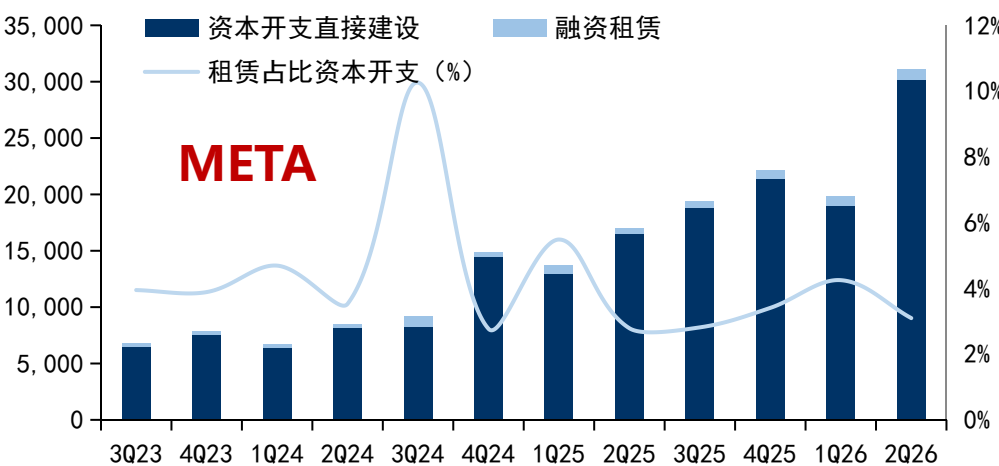
自由现金流承压下，租赁算力（经营/融资租赁）将提升

图：微软资本开支与融资租赁变化情况（百万美元、%）



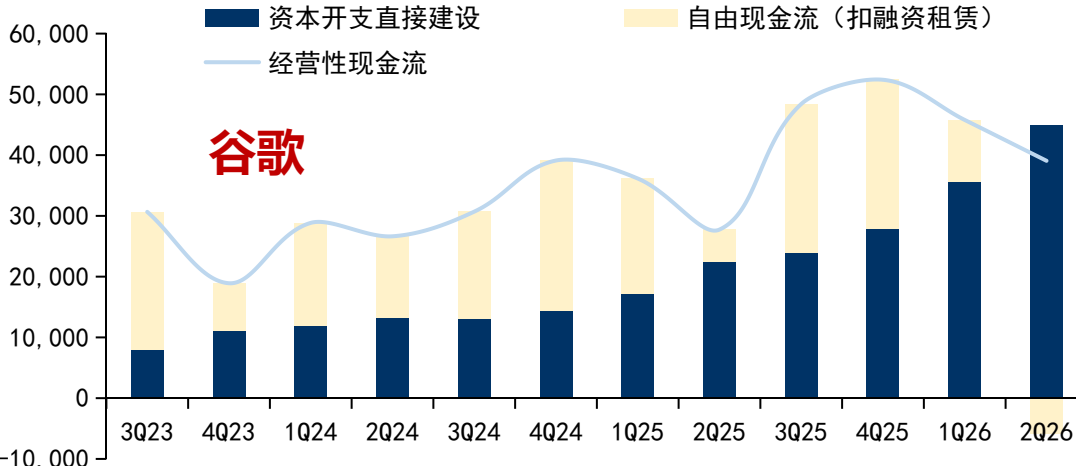
资料来源：公司财报、国信证券经济研究所整理

图：META资本开支与融资租赁变化情况（百万美元、%）



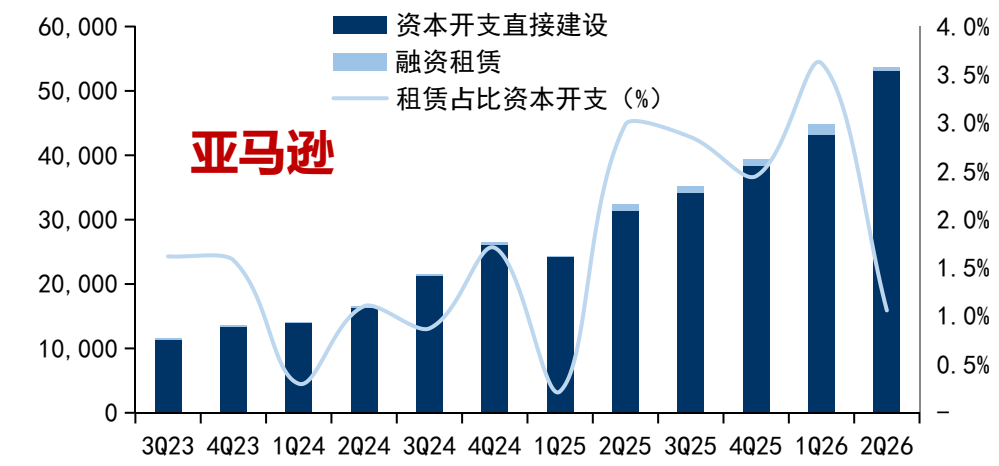
资料来源：公司财报、国信证券经济研究所整理

图：谷歌资本开支与融资租赁变化情况（百万美元、%）



资料来源：公司财报、国信证券经济研究所整理

图：亚马逊资本开支与融资租赁变化情况（百万美元、%）



资料来源：公司财报、国信证券经济研究所整理

微软数据中心建设计划

- 微软维持两年内总数据中心/AI容量翻倍目标，FY2026新增88座数据中心、最近连续三个季度各新增约1GW容量，2026年核心大项目集中在美国威斯康星Fairwater、得州Abilene、得州Pecos，以及欧洲/亚太多国主权云与AI基础设施扩建。

表：微软美国与欧洲算力规模预测

区域	项目	容量/GW	当前进度	关键时间点
威斯康星	Fairwater (Mount Pleasant)	约2GW口径最常见，更激进口径可到3GW+	首座已投运，园区继续扩建	2026年初首期投运；2028年第二设施完成
佐治亚	Fairwater Atlanta	0.5-0.86GW	已运行/扩容中	2025年10月开始运营
得州Abilene	Crusoe租赁	0.9GW	土建推进中	首栋2027年中通电
得州Pecos	自建+Chevron供电	约2GW DC + 2.67GW BTM发电	规划开发中	首电2028年，建设5-7年
怀俄明	新土地储备扩建	未披露	多年规划	2026年启动土地获取
Cheyenne				
美国大亚特兰大都会区	East US 3区域	未披露	处于规划/建设推进中	2027年初开放
葡萄牙锡尼什	Sines项目（与Nscale Start Campus合作）	1.2GW，100亿美元	已披露合作与规划，属于中长期AI枢纽项目	2026-2030年推进；规划含1.26万块BW芯片+6.6万块Rubin芯片
英国洛顿	Loughton项目（与Nscale合作）	150亿美元	规划中	2025-2028年；部署2.3万块英伟达GPU
西班牙	Zaragoza数据中心园区	21亿美元	2026年开工	2024-2025为前期，2026年启动建设
德国	德国项目	38亿美元	2024-2025推进	已在执行期内
法国	法国项目	46亿美元	推进中	截至2025年底最多部署2.5万块GPU
瑞典	瑞典项目	1GW，32亿美元	推进中	2024-2026年
丹麦	丹麦云区域/扩建	数亿美元	2026年3月丹麦东部云区域已开放	西丹麦瓦德/埃斯比约项目2026-2028年推进
挪威	挪威Narvik（租赁）	前期承诺62亿美元	部署中	将额外部署3万块NVIDIA Vera Rubin芯片

表：微软算力规模预测

国家/地区	项目特征	投资金额	当前进度	后续关键时间点
印度尼西亚	云与AI基础设施建设	17亿美元	推进中	2024下半年-2027年
印度（首批）	区域扩建	30亿美元	规划/推进中	2025-2027年
印度（新增）	新一轮扩容	175亿美元	远期规划	2027-2029年
马来西亚	数据中心/云扩容	22亿美元	规划中	2025-2029年
新加坡	云与AI基础设施	55亿美元	已宣布投资	2025-2029/2030年持续投入
日本	数据中心与AI设施	100亿美元	已宣布	至2029年四年投资
泰国	数据中心建设	超10亿美元	规划中	2026-2029年
澳大利亚	云与AI基础设施	180亿美元	已宣布投资	2026-2030年
加拿大	基础设施扩建	54.8亿美元	规划中	2026-2028年
阿联酋（与G42）	Khazna数据中心相关	101亿美元	推进中	2026年底达200MW，整体周期至2029年
沙特阿拉伯东部	新云区域	未披露	建设中	2026年Q4上线，含3个AZ

资料来源：路透社、彭博、公司官网、国信证券经济研究所整理测算

资料来源：路透社、彭博、公司官网、国信证券经济研究所整理测算

META数据中心建设计划

- Meta当前围绕俄亥俄Prometheus、路易斯安那Hyperion、得州El Paso、印第安纳Lebanon、加拿大阿尔伯塔等少数GW级园区快速上量，并以“**2026年底约7GW、2027年底约14GW（含租赁）**”的内部目标牵引未来两年资本开支和外部合作。

表：META算力规模预测

项目名称	建设区域	规模/GW	投资金额	当前建设进度	后续关键时间点
Prometheus	美国俄亥俄州 New Albany	1GW+	多篇外部资料/融资口径指向约100亿美元级建设体量，且有30亿美元建设贷款	已进入投运爬坡，行业资料称500MW+已活跃；建有5栋永久建筑+5栋快速部署结构，并配套400MW离网燃气发电	最初预期2026年上线；第三方跟踪显示通电里程碑推迟至2026年9月30日，目标仍为2026年底接近/达到1GW
Hyperion	美国路易斯安那州 Richland Parish	远期5GW；2030年前2GW	初始100亿美元，后Blue Owl JV披露270亿美元，2026年7月上调至超500亿美元	大规模施工中；2026年已有3700+工人在原地块施工，并新购1400英亩相邻土地	2030年前2GW，2032年前完成5GW；地方报道口径全面完工可至2036年
El Paso	美国得州 El Paso	1GW	2026年3月Meta披露超100亿美元；与BlackRock合作后总开发成本约140亿美元	2025年10月开工，2026年中已2300+工人在场；最新申报扩展至12栋建筑、约2590万平方英尺	2028年开始上线/投运，工程申报显示2029年2月完成
Lebanon	美国印第安纳州 Lebanon (LEAP)	1GW	超100亿美元	2026年2月正式开工；规划13栋建筑，其中10栋数据中心，1500英亩，分期交付	首期预计2027年底或2028年初上线
Alberta	加拿大阿尔伯塔省 Sturgeon County	1GW	130亿加元，约92亿美元	2026年7月宣布，为Meta在美国境外最大数据中心，290万平方英尺	配套Greenlight发电设施，预计2030年下半年投运

资料来源：路透社、彭博、公司官网、国信证券经济研究所整理测算

表：META租赁的算力规模预测

合作方	合同金额	容量口径	2026年状态	关键点
Nebius（首份）	~30亿美元/5年	测算约0.04GW	2026年2月全部上线	两期交付，分别确认12/11个月收入
Nebius（新约）	270亿美元/5年	测算约0.36GW	2027年初起交付，	
CoreWeave	140亿美元	测算约0.19GW	续行	2025年9月签订的长期供给框架。
CoreWeave	210亿美元/至2032年12月	测算约0.23GW	2026H2起交付，含Rubin初始部署	2026年4月新合同
GoogleCloud	~100亿美元/6年	测算约0.11GW	续行	2025年8月签订，2026年持续供给
Crusoe	—	~1.6GW	报道披露，细节待确证	为2026-2027提供大体量的新增功率池
总计				2026年约交付略超1GW算力，2027年约交付2.5GW算力

资料来源：路透社、彭博、公司官网、国信证券经济研究所整理测算

亚马逊数据中心规划

图：亚马逊数据中心规划情况

项目	数据中心	目前容量	规划容量	概况	节点	配套电力
亚马逊宾夕法尼亚AI创 新园区			1.9GW	亚马逊计划在宾夕法尼亚州投资逾200亿美元， 扩建面向AI与云计算的数据中心基础设施	AWS计划在 2026/2027/2028/2029年 年中前，至少逐步提升至 240/360/480/640 MW	Talen已承诺通过 Susquehanna核电站向 AWS提供1.9 GW电力， 期限至2042年
Project Rainier	Anthropic-Amazon New Carlisle	0.91GW	2.2GW	计划在印第安纳州投资约110亿美元	预计2026年10月达 1.037GW，2028年 10月达1.925GW	-
亚马逊密西西比项目	Amazon Madison Mega Site	0.284GW	预计到2028年3月 达0.683GW	计划在密西西比州投资100亿美元；AWS将与 MCEDA合作，在Madison County两个工业园 区建设多个数据中心综合体	-	-
亚马逊密西西比项目	Amazon Ridgeland	0.284GW	预计到2028年9月 达0.84GW	-	-	-
	AWS Berwick	0.108GW	预计到27年8月达 0.169GW			
亚马逊俄亥俄项目			-	计划投资100亿美元，以满足快速增长的云计算 和AI需求	计划于2030年底前最多扩 展8个新站点	-
亚马逊佐治亚项目			-	计划投资110亿美元，扩建佐治亚州基础设施， 以支持云计算与AI技术	-	-
亚马逊北印第安纳			2.4GW	计划在印第安纳州北部投资约150亿美元建设新 数据园区		
亚马逊北卡罗莱纳园区			-	计划在北卡罗来纳州投资100亿美元，扩建云计 算基础设施并推进AI创新	计划于2025年10月破土动 工	-
合计		1.586GW	8.192GW			

资料来源：公司公告、Epoch AI、国信证券经济研究所整理

图：谷歌数据中心规划情况

项目	规划容量	配套电力	节点	投资	要点
德州云与AI基础设施计划	-	已签约6.2 GW; Haskell 0.64 GW太阳能	至2027	400亿美元	Armstrong/Haskell/Wilbarger; Meitner配套1+GW能源
Intersect共址项目组合	多GW	同步新建清洁电力	首期2026投运; 2027完工	目标200亿美元	第三方能源基建目标; 与德州计划可能重叠
Google Pryor	-	-	2025-27	90亿美元	新建Stillwater并扩建Pryor
弗吉尼亚州 (Chesterfield等)	-	-	至2026	90亿美元	新建Chesterfield并扩充现有设施
南卡罗来纳州园区组合	-	-	至2027	90亿美元	扩建Berkeley; 继续建设Dorchester两处站点
阿肯色州West Memphis园区	-	0.6 GW太阳能	至2027	40亿美元	Google在该州首个数据中心
爱荷华州云与AI基础设施	-	-	-	70亿美元	州级云、AI与劳动力计划
Broadwing能源 / 伊利诺伊 Decatur	0.6 GW	>0.4 GW CCS燃气	2029末发电; 2030初商业化	-	Google购入大部分电力; 约90%碳捕集

资料来源：公司公告、国信证券经济研究所整理

海内外科技厂商关于资本开支与算力容量的相关表述

表：海外科技厂商关于资本开支与算力容量的相关表述

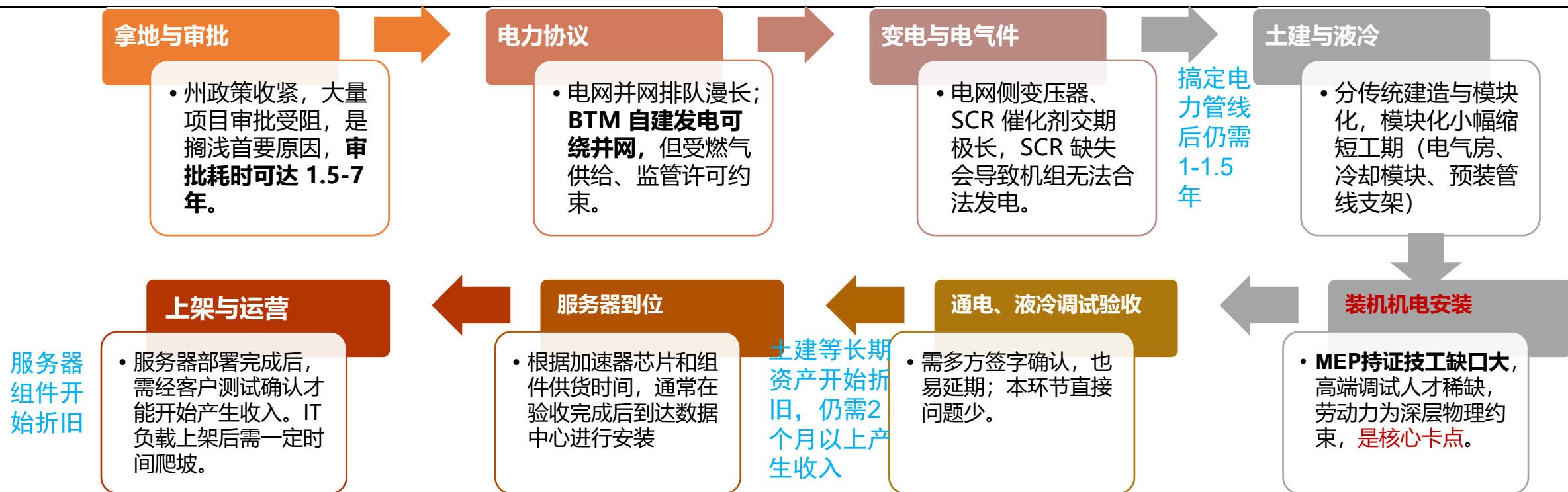
	资本开支表述	AI算力表述	数据中心算力表述	数据中心功耗表述	电力供应表述
微软	2026自然年资本开支约1750亿美元（采购量不变、口径调整） - 微软每年在AI基建上的固定投资超过800亿美元，这是十年规划。	公司计划FY26（25Q3-26Q2）将AI总容量提升超80%。	26Q2业绩会：过去三季度每季度新增1GW算力容量，仍计划2年内整体算力翻倍	25Q2业绩会：数据中心超400个，过去一年新增超2GW； - 微软CEO25年10月发言：微软云全球数据中心总功耗5GW；	
谷歌	26Q2：全年1950-2050亿美元(同比约+100%，上季度指引1800 - 1900亿) 2027资本开支将继续显著增长	2026年TPU投入和算力规模将超越GPU			
META	上调2026年资本开支至1300 - 1450亿美元(同比约+90%，上季度1250-1450亿)	24年底计拥有等效60万H100GPU的计算能力；	2026年计划部署7GW的基础设施，并计划在2027年翻一倍。		
亚马逊	预计26年全年将达到2200亿(同比+71%、因存储涨价，上季度指引2000亿美元)	26Q1：芯片业务ARR200亿美元，独立对外销售，ARR将500亿美元	26Q2业绩会：预计到27年底算力容量将达到25年的2倍。		26Q1业绩会：过去12个月AWS新增3.99吉瓦电力产能，
OpenAI	- 星际之门计划投资超5000亿美元； - OpenAI承诺在未来多年内向微软Azure云服务采购约2500亿美元的服务；	英伟达未来供应10GW数据中心的芯片；AMD未来供应6GW芯片；博通未来供应10GW芯片；		2025：未来四年内星际之门Stargate计划容量达到近10GW；	

- [01] 加速器芯片出货量与对应GW
- [02] 北美数据中心增长与大厂建设计划
- [03] 电力、建设限制与算力增长测算
- [04] 1GW建设成本与CSP、大模型厂收入变现

数据中心建设流程与核心卡点

- 一个数据中心从园区方案定型到IT上线/产生收入，全周期约25-36个月。
- 从搞定电力管线到完成IT调试仍有较长周期，一般为13-18个月。如亚马逊2025Q4披露一年新增3.9GW电力容量，仍需一年以上才能变为算力容量。当前电网增长较慢，主要通过BTM（表后供电）实现绕线并网。
- **数据中心通电前三类现场安装，必须持证技工，是当前最严重的卡点。**现场安装MEP包括：机械（冷却管路、CDU、风侧设备）、电气（配电、母线、开关柜接线、保护整定）、管道（水系统）。

图：数据中心建设流程



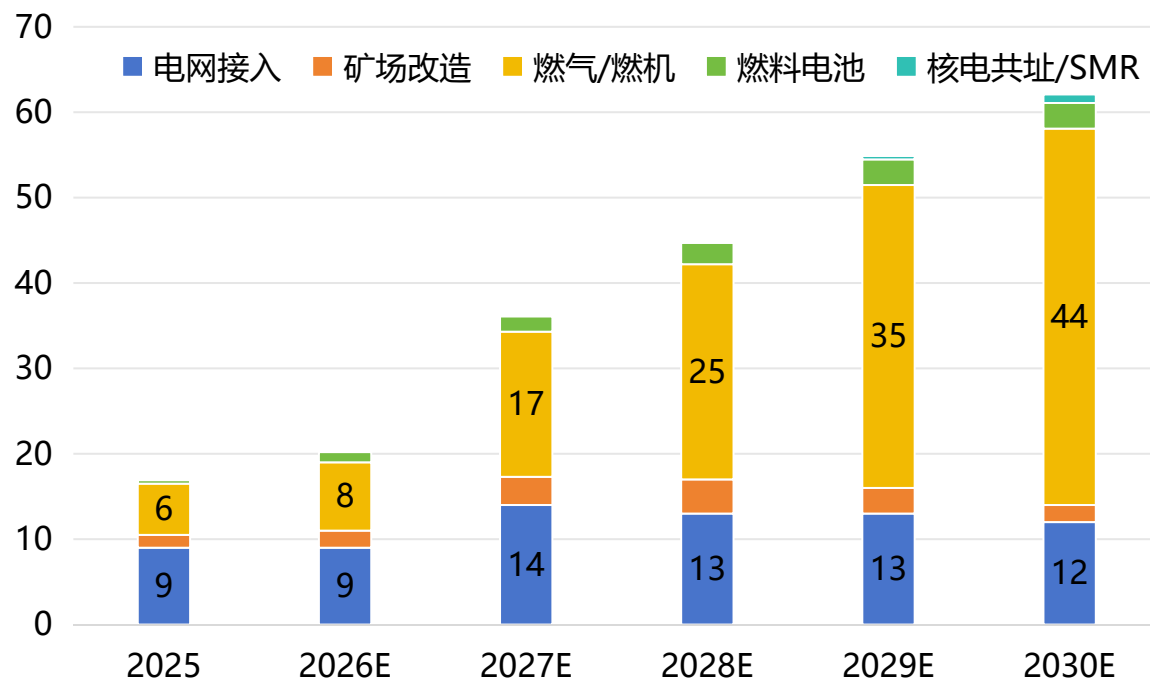
北美电力扩充方式：电网扩产与BTM（表后供电）并行

- 北美AI电力并非“总发电量不够”，而是“可在2-4年内交付、可并网/可离网、可匹配数据中心负荷特征的确定性电力不够”。因此BTM（behind-the-meter，表后供电）之所以成为核心途径，本质是因为超大数据中心已经不再完全依赖电网，而是转向“自己带电来建数据中心”。
- 2025-2030年北美最现实的增量主线是“气电+燃料电池+矿场改造”三件套，核电更多是中长期资产锁定而非2028年前主供给。

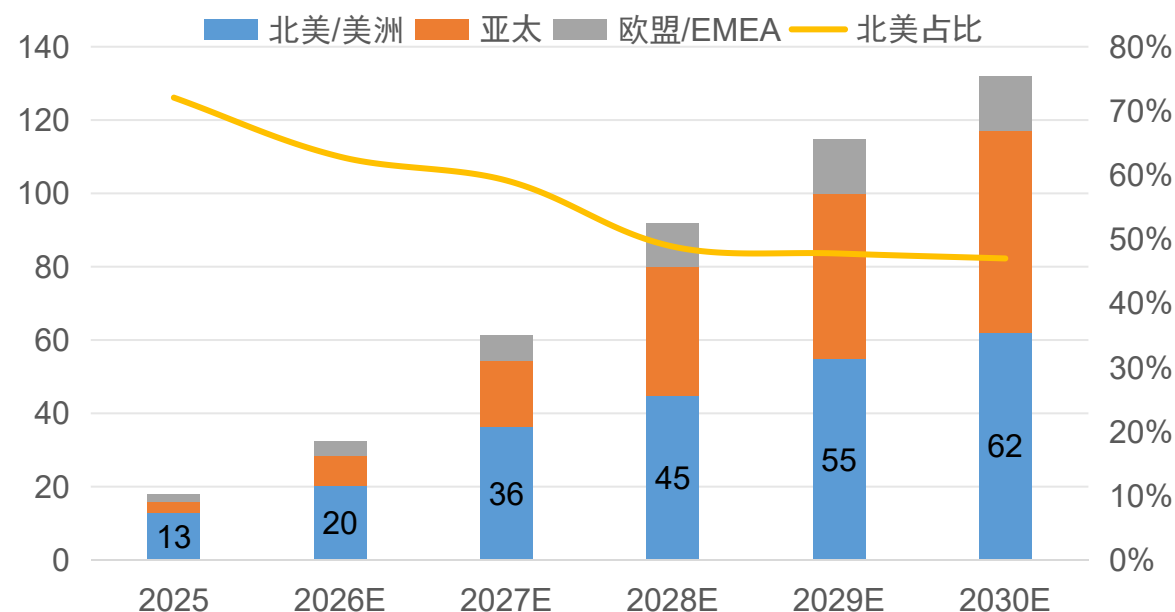
表：北美各类型电力容量扩充方式说明					
	说明	供应量	主要方式	特点	相关公司
电网在建	受电网限制，北美每年新增约15GW	10-15GW/年	原有电网扩充	输电、变电、互联审批与主设备交付周期长	变压器、GIS、HVDC、UPS
矿场改造		约2-4GW/年	将已经具备电力接入、高压设施、许可的资产，升级为HPC/AI数据中心	需要冷却系统升级、供电冗余等级不足	CORZ、WULF、APLD、CIFR、MARA、HUT、IREN、
燃气轮机	BTM（behind-the-meter，表后供电），自发电	28年前5-10GW/年，后15-20GW/年	重型/中型燃机（天然气）作为主电源	唯一兼具“大规模、稳定基荷、技术成熟、可在2030前显著贡献GW”的主方案	GE Vernova、Siemens Energy、Mitsubishi、Doosan、
燃料电池		26年0.5GW，后每年增加0.5GW	SOFC（固体氧化物燃料电池）现场供电/微电网方案	部署速度快，典型可在90天到一年内。模块化、对社区阻力更低。	Bloom Energy
核电/SMR		28年前较少，30年后约1-5GW/年	核发电，稳定、无碳、高容量因子、天然适合7x24基荷	2030年后落地。兑现，监管、并网、许可与建设进度比市场最初想象更慢	Constellation、Talen、Vistra；远期Oklo、TerraPower、Kairos；

- 美国电网供给每年约10-15GW，形成巨大的需求缺口。但BTM尤其燃机贡献显著，2025年全球燃气轮机行业产能约50GW，其中数据中心占2025年全球燃机订单约22GW，约占总订单的25%，26年预计超15GW。
- 全球算力建设则呈现更清晰的区域分化：2025-2030年北美仍是第一增长极，年化增长从13GW增至62GW；全球AI数据中心增量从18GW增至125GW。如果按逐年装机节奏看，2026-2030年，北美仍贡献其中最大份额，但中国与东南亚、中东会逐步提高边际贡献。

图：北美电力容量分类型增长（GW）



图：全球电力容量分区域增长（GW）



资料来源：Omdia、公司官网，国信证券经济研究所整理测算

资料来源：Omdia、公司官网，国信证券经济研究所整理测算

请务必阅读正文之后的免责声明及其项下所有内容

表：燃气轮机出货量（GW）

分类维度	项目	2024	2025	2026	2027	2028	2029	2030	2024-2030 CAGR	产能增量 ΔGW
总产能	发电设备制造总产能	84	84	107	130	150	174	197	15%	112
按应用场景	灵活型（备用 + 主力）	28	28	35	40	47	55	115	26%	86
	专用备用型	27	27	38	45	50	55	11	-14%	-16
	专用主力型	29	29	35	44	52	64	71	16%	42
按 OEM 厂商	卡特彼勒 Caterpillar	25	25	29	34	40	47	55	14%	30
	康明斯 Cummins	26	26	35	38	42	46	50	12%	24
	MTU（罗尔斯·罗伊斯）	7	7	10	11	13	15	17	15%	10
	瓦锡兰 Wärtsilä	6	6	8	10	13	16	18	20%	12
	潍柴 Weichai	1	1	3	7	8	9	11	55%	10
	INNIO	3	3	3	4	5	7	8	22%	6
	西门子能源 Siemens	4	4	5	6	6	8	9	14%	5
	三菱重工 Mitsubishi	3	3	3	4	4	5	6	11%	3
	埃夫朗（曼）Everllence	2	2	2	3	3	4	4	16%	3
	Bloom Energy	1	1	1	2	3	3	4	25%	3
	贝克休斯 Baker Hughes	3	3	3	3	3	4	4	5%	1
	PROENERGY	2	2	2	2	2	3	3	12%	1
	FTAI Aviation	0	0	0	2	3	3	3	—	3
	GE Vernova	2	2	2	2	2	2	2	2%	0
	Boom Supersonic	0	0	0	1	2	2	2	—	2

资料来源：各公司官网、SemiAnalysis、McCoy、Omida，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

- 持证电工、专业技工短缺，北美26H1仅新增约6GW数据中心，算力订单转化为收入速度慢。头部 EPC（Engineering Procurement Construction 设计-采购-施工总承包）企业人力不足（25年底约26万人，未来按照EPC员工年复合增速10%假设测算），AI建设、电网改造、新能源项目争抢同一批持证电工，尤其北弗吉尼亚、凤凰城、达拉斯瓶颈严重
- 解决方式：OEM系统集成商（Vertiv、Schneider）从卖设备延伸，Meta、AWS养大规模内部工程团队，提供部分EPC交付。模块化把大量工作转移到工厂，用制造业产能替代稀缺的现场EPC施工人力。

表：数据中心状态与未来落地情况

项目状态	未来12个月落地概率	判断依据
建筑完成、电力和制冷已调试 energized/installed capacity	85% - 95%	开始折旧，一般需2个月以上产生收入，剩余风险主要是最终验收和服务器部署
已取得明确并网日期，变电站在建 Firm interconnection	65% - 80%	仍可能受主变、开关设备及调试延期影响
建设中，但依赖新电力 Under Construction	35% - 60%	电力设施通常慢于机房建设
已签约租赁，但电力计划尚未锁定 Leases Signed, shell completed	20% - 40%	租约并不等于具备可用电力
已宣布、未开工 Pipeline/contracted	5% - 20%	项目取消、缩减或延期概率较高

资料来源：Neocloud官网，国信证券经济研究所整理测算

表：建设进度与容量上限测算

年份	北美EPC员工总数 (千人)	EPC员工 YoY*	新增EPC供给 *	建设GW/年	配比(千人/GW/年)
2023A	200			7	29
2024A	222	11%		10	23
2025E	260	17%		12	21
2026E	290	12%	10	15	20
2027E	320	10%	55	20	19
2028E	350	9%	75	23	19
2029E	380	9%	90	26	18
2030E	410	8%	100	30	17

资料来源：各公司官网，国信证券经济研究所整理测算 *假设未来五年EPC员工年复合增速10% *新增供给包括新玩家和模块化提效。

表：北美各州数据中心建设限制

州/市场	当前核心限制	核心要点
弗吉尼亚州 / 北弗吉尼亚	输电走廊、230kV线路及变电站审批；居民区附近选线阻力；专业电工和调试资源紧张	<50MW 接电仍需 12 年；>50MW 需新建输变电，部分只能临时 / 分阶段供电
俄亥俄州	AEP Ohio接电容量、区域输电升级、负荷真实性审查	25MW+ 项目排队做负荷研究，投运受制于输电升级，电费费率四年爬坡
得克萨斯州	超大规模接电队列、动态稳定性、输电通道和可调负荷要求	ERCOT 批量审批 75MW + 负荷，IDC 占 89% 申请，项目面临分阶段、减负荷要求
亚利桑那州 / 凤凰城	高温条件下的峰值电力、发电与输电建设周期、水资源和冷却方案	APS 待承接大负荷 19GW，远超系统峰值，部分项目要等到 2028 年后或自建微网
佐治亚州 / 亚特兰大	新增发电容量、输电扩建、项目负荷真实性及成本分摊	短期电力供给不足，2026- 2028 项目普遍需要分期接电
俄勒冈州	输电接入、柴油备用机空气排放许可、土地和用水约束	俄勒冈 40+IDC 项目，6GW 备用电源，柴油机组环评许可难度随规模抬升
路易斯安那州	专用发电、500kV输电线路、变电站建设与监管审批	Meta 项目配套燃气发电、长距离输电，2028- 2029 才逐步完成电力配套
宾夕法尼亚州	PJM区域输电、发电资源接入和当地电力基础设施	依赖 PJM 跨州输电升级，输变电建设周期长于机房土建
爱荷华州	局部输电容量、风电出力波动、备用电源及水资源	资源禀赋尚可，大型 AI 集群仍需新增变电站与可调电源
印第安纳州	电网扩建、发电资源、地方许可及成本分摊	单一大项目即可触发变电站、输电改造
内华达州	水资源、高温制冷、电力与土地审批	高温带来 PUE 抬升，受峰值电力容量约束
密西西比、阿拉巴马、南卡罗来纳	电力基础设施相对薄弱、变电站及输电扩建周期、专业承包商不足	土地税收友好，本地供应链薄弱，依赖跨州 EPC 施工资源
威斯康星、怀俄明、新墨西哥、犹他	发电与输电接入、水资源、极端气候及设备运输	土地充裕，超大规模园区需配套专用变电站与长距离输电

资料来源：华尔街新闻，各公司官网，公司国信证券经济研究所整理测算

芯片出货量、电力供给、建设能力与资本开支约束

图：芯片出货量、电力供给、建设能力与资本开支约束下 预期GW数增长

加速器/千颗*		2025	2026E	2027E	2028E	2029E	2030E
芯片出货量约束	GPU	6,110	8,370	11,500	13,300	14,600	16,000
	NVDA	5,700	7,800	10,500	12,000	13,000	14,000
	AMD	410	570	1,000	1,300	1,600	2,000
	ASIC	2,850	7,700	14,800	21,000	25,500	29,000
	TPU	1,700	5,000	10,000	15,000	18,000	20,000
	Trainium	750	1,700	3,300	4,000	5,000	6,000
	其他ASIC	400	1,000	1,500	2,000	2,500	3,000
	Total	8,960	16,070	26,300	34,300	40,100	45,000
	YoY	22%	79%	64%	30%	17%	12%
	出货量对应所需GW数	13	33	62	105	129	152
电力约束 (主要北美区域)	北美/美洲	13	20	36	45	55	62
建设进度约束 (主要北美区域)	北美建设能力上限	12	16	20	23	26	30
资本开支约束	全球AI资本开支 (亿美元)	6,000	13,083	18,647	25,137	34,276	40,000
	每GW建设成本 (亿美元/GW)	430	480	500	550	605	666
	资本开支可建设GW*	13	27	37	46	57	60

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理测算 *2027年假设单GW建设成本为500亿美元，后续单GW建设成本每年增长10%。
北美与全球电力增长测算详见具体章节，资本开支增长测算详见具体章节

图：芯片出货量、电力供给、建设能力与资本开支约束下 预期GW数增长

	2025	2026E	2027E	2028E	2029E	2030E
芯片出货量	8,960	16,070	26,300	34,300	40,100	45,000
YoY	22%	79%	64%	30%	17%	12%
出货量对应所需GW数	13	33	62	105	129	152
北美电力约束	13	20	36	45	55	62
北美建设能力上限	12	16	20	23	26	30
全球AI资本开支 (亿美元)	6,000	13,083	18,647	25,137	34,276	40,000
每GW建设成本 (亿美元/GW) *	430	480	500	550	605	666
资本开支可建设GW	13	27	37	46	57	60
假设一：北美区域建设能力为最小约束，假设全球实际建设GW受到资本开支规模限制，按照此前资本开支测算						
实际北美建设GW	9	16	20	23	26	30
实际全球建设GW	13	27	37	46	57	60
YoY		110%	37%	23%	24%	6%
北美区域建设占比	69%	57%	53%	50%	46%	50%
假设二：若北美建设占比全球建设容量比例维持65%，因北美数据中心ROI显著更高，测算全球算力与资本开支情况						
实际北美建设GW	9	16	20	23	26	30
全球建设GW	13	24	30	35	40	46
YoY		83%	27%	16%	14%	15%
全球AI资本开支 (亿美元)	5,590	11,446	15,182	19,439	24,303	30,715
YoY		105%	33%	28%	25%	26%

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理测算

- [01] 加速器芯片出货量与对应GW
- [02] 北美数据中心增长与大厂建设计划
- [03] 电力、建设限制与算力增长测算
- [04] 1GW建设成本与CSP、大模型厂收入变现

微软云1GW变现：当前约在148亿美元/GW/年（不包含OAI采购收入）

国信证券
GUOSEN SECURITIES

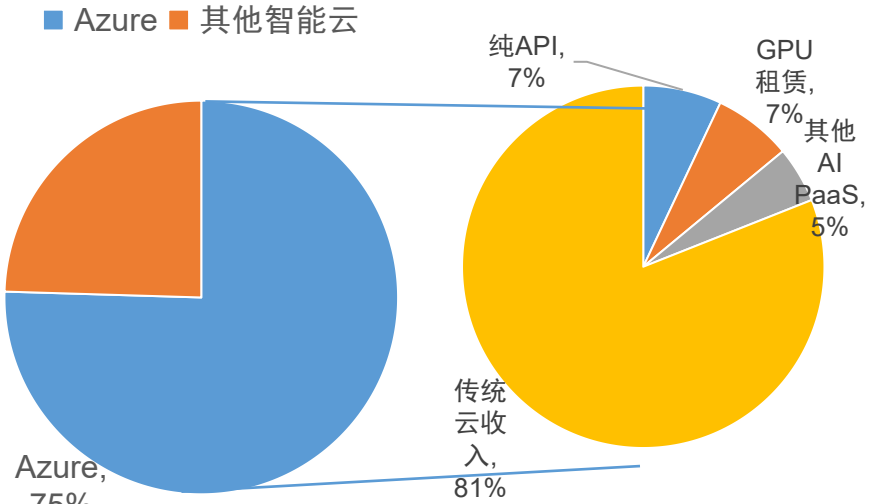
- 微软Azure整体F26财年（20250701-20260630）收入约1040多亿美元，占比微软智能云75%。测算Azure AI板块（右侧小饼图口径）F26财年收入接近200亿（收入不包含OpenAI向微软采购的云服务收入），占Azure整体收入的比例约20%。OpenAI向微软采购的云服务计入Azure云的非AI收入部分，且给与折扣，算力变现低于AI云。

表：微软云拆分与AI贡献汇总（百万美元）

单位：百万美元	2024	2025	2026E	2027E	3Q25	4Q25	1Q26	2Q26
微软云收入	104,859	120,433	156,660	219,324	30,897	32,907	34,681	39,306
YoY	9%	15%	30%	40%	28%	29%	30%	32%
Azure云收入	72,468	88,901	127,786	191,679	22,987	24,875	25,597	32,540
YoY		23%	44%	50%	40%	39%	40%	43%
Azure云占比智能云	69%	74%	82%	87%	74%	76%	74%	83%
微软云OP	46,762	50,447	63,584	83,343	13,391	13,821	14,913	16,115
OPM	45%	42%	41%	38%	43%	42%	43%	41%
每GW AI算力收入 (亿美元) *		135	148	161				
年末AI云算力		4.5	8.5	12				
AI云收入	7,167	15,179	28,917	54,394	3,908	4,478	5,375	7,159
YoY		112%	91%	88%	98%	92%	84%	85%
占微软云	7%	13%	18%	25%	13%	14%	15%	18%
占Azure云比例	10%	17%	23%	28%	17%	18%	21%	22%
AI IaaS收入	2,509	5,313	10,223	17,251	1,368	1,567	1,881	2,506
YoY		112%	92%	69%	98%	92%	84%	85%
MaaS收入	2,365	5,009	10,606	19,168	1,290	1,478	1,774	2,362
YoY		112%	112%	81%	98%	92%	84%	85%

表：微软云拆分与AI贡献汇总（百万美元）

FY2026



资料来源：公司业绩会，彭博，国信证券经济研究所整理测算 *按照当年平均对外AI算力假设计算，平均算力为年初算力与年末算力和的一半，假设微软云所用算力（外部算力）占比2025到2027年分别为40%、30%、35%。外部占比较低是因OAI相关算力租赁不算做AI云收入，26年对外算力占比下降因微软内部模型训练提升、Copilot用量增加、OAI用量占比提升。后续对外算力占比逐步提升。

资料来源：公司官网，彭博，国信证券经济研究所整理测算

34

谷歌云与亚马逊云1GW变现：当前分别约为146、67亿美元/GW/年



- 26年预计谷歌云收入同比+85%、增长约500亿美元，传统云100亿，400亿来自AI云的TPU、GPU租赁、MaaS增长，TPU外售是主要驱动因素。
- 亚马逊变现较低因AI算力有较多Trainium算力、成本低，Anthropic早期战略订单价格较低。26Q2业绩会亚马逊表示目前服务器和网络设备的回本周期不到3年，寿命5-6年。AWS AI云26年的重要增长驱动来自Bedrock平台 Claude模型贡献的MaaS收入，预计在26Q1 150亿AI云ARR中占比达到1/3。

表：谷歌云拆分与AI贡献汇总（百万美元）

单位：百万美元	2024	2025	2026E	2027E	3Q25	4Q25	1Q26	2Q26
谷歌云收入	43,229	58,705	108,414	211,209	15,157	17,664	20,028	24,768
YoY	31%	36%	85%	95%	34%	48%	63%	82%
OP	6,112	13,910	37,962	75,149	3,594	5,313	6,598	8,814
OPM	14%	24%	35.0%	35.6%	24%	30%	33%	35.6%
每GW AI云算力收入 (亿美元)		125	146	184				
年末AI云算力		3	7.5	13				
AI云	3,541	9,397	45,894	131,687	2,405	4,015	6,274	9,365
YoY	117%	165%	388%	187%	152%	278%	374%	466%
占GCP	13%	23%	52%	70%	23%	32%	40%	47%
占云收入比例	8%	16%	42%	62%	16%	23%	31%	38%
OP	1,173	3,492	16,763	48,240	907	1,554	2,357	3,486
OPM	33.1%	37.2%	36.5%	36.6%	37.7%	38.7%	37.6%	37.2%
AI IaaS收入	3,541	7,049	34,154	96,467	1,700	2,724	4,513	7,017
YoY	117%	99%	385%	182%	78%	156%	274%	395%
OP	1,173	2,318	10,893	30,630	555	908	1,476	2,312
OPM	33.1%	32.9%	31.9%	31.8%	32.6%	33.4%	32.7%	33.0%
TPU收入	2,214	4,059	23,827	80,130	893	1,826	2,448	4,642
YoY	92%	83%	487%	236%	49%	175%	302%	535%
占云收入比例	5%	7%	22%	38%	6%	10%	12%	19%
TPU租赁	2,214	4,059	18,827	50,130	893	1,826	2,448	4,142
YoY	168%	83%	364%	166%	49%	175%	302%	467%
TPU外售			5,000	30,000				500
MaaS收入		2,348	11,740	35,220	704	1,291	1,761	2,348
YoY			400%	200%				
OP		1,174	5,870	17,610	352	646	881	1,174
OPM		50%	50%	50%	50%	50%	50%	50%

资料来源：公司官网，彭博，国信证券经济研究所整理测算 *假设谷歌云所用算力（外部算力）占比2025到2027年分别为50%、60%、70%。25年公司有较多模型训练需求，后续对外算力占比逐步提升。

表：亚马逊云拆分与AI贡献业绩汇总（百万美元）

单位：百万美元	2024	2025	2026E	2027E	3Q25	4Q25	1Q26	2Q26
AWS收入	107,556	128,725	179,225	260,495	33,006	35,579	37,587	42,232
YOY	19%	20%	39%	45%	20.2%	24.0%	28%	37%
OP	39,834	45,606	68,027	98,293	11,434	12,465	14,161	16,621
OPM	37.0%	35.4%	38.0%	37.7%	34.6%	35.0%	37.7%	39.4%
每GW AI云算力收入		56	67	84				
年末AI云算力		3	8	14				
AI云	2,750	6,691	31,548	83,283	1,987	2,554	3,750	6,250
YOY	400%	143%	371%	164%			329%	390%
占比	3%	5%	18%	32%	6%	7%	10%	15%
ARR							15,000	25,000
OP	1,005	2,542	12,313	32,724	755	975	1,450	2,425
OPM	37%	38%	39%	39%	38%	38%	39%	39%
AI IaaS	2750	5,739	15,798	36,158	1,698	2,091	2,500	3,750
YOY	400%	109%	175%	129%			214%	225%
Trn		3,050	9,670	24,378	900	1,150	1,500	2,250
YOY			217%	152%			275%	275%
占AI IaaS比例		53%	61%	67%				
OPM		40%	40%	40%	40%	40%	40%	40%
MaaS (Bedrock)		952	15,750	47,125	289	463	1,250	2,500
YOY			1554%	199%			1487%	1963%
占AI云比例			50%	57%	15%	18%	33%	40%
ARR							5000	10000
OP		381	6,300	18,850	116	185	500	1,000
OPM		40%	40%	40%	40%	40%	40%	40%

资料来源：公司官网，彭博，国信证券经济研究所整理测算 *假设谷歌云所用算力（外部算力）占比2025到2027年分别为80%、85%、90%。公司自有模型训练需求较少、主要支持Anthropic，后续对外算力占比逐步提升。

Anthropic、OpenAI：当前1GW推理算力收入约200亿美元/GW/年



- 测算26年两家头部模型厂商年度算力支出超千亿美元、同比增速5倍以上，其中Anthropic 490亿美元（240亿推理、250亿训练），OpenAI 550亿美元（230亿推理、320亿训练），**2026年两家头部模型厂商占用超10GW算力，测算每GW推理算力当前贡献约200亿美元收入/年。**

表：Anthropic、OpenAI公司算力成本与变现

百万美元	Anthropic			OpenAI			备注
	2024	2025	2026E	2024	2025	2026E	
总收入	524	4,129	66,722	3,600	12,905	47,508	
总算力成本	2,641	6,561	49,140	4,061	15,900	54,640	26年年度算力支出超千亿美元
YOY		148%	649%		292%	244%	
推理算力成本	1,016	2,561	24,144	2,061	8,400	22,640	
模型训练成本	1,625	4,000	24,996	2,000	7,500	32,000	
总算力规模 (MW)	110	779	5963	169	1887	6630	26年占用约12GW算力
推理算力规模	42	304	2930	86	997	2747	
训练算力规模	68	475	3033	83	890	3883	
每GW推理算力年收入 (亿美元)	120	140	230	420	130	170	
YoY		10%	68%		-69%	34%	
每MW推理算力成本 (亿美元)		84	82		84	82	每MW H100算力租赁约2千万美元/年
推理算力毛利率		38%	64%	43%	35%	52%	
推理算力经营利润率	-297%	-198%	11%	-231%	-179%	-111%	

数据来源：Exclusive、Semianalyse，公司官网、新浪新闻，国信证券经济研究所整理测算

请务必阅读正文之后的免责声明及其项下所有内容

从算力租赁与Token定价测算：云厂与模型厂1GW边际收入



- 按照当前算力租赁价格测算，云厂部署1GW英伟达**GB200**、**Rubin**芯片年化收入约200亿美元以上、H系列芯片年收入约150亿美元以上，按照当前tokens定价，大模型厂基于1GW的GB200、Rubin芯片部署**Opus5**可产生500、1000亿美元以上收入（一半训练、一半推理）。
- 假设1：1GW H100/GB200/Rubin/950DT（国内）对应的资本开支分别为370/400/470/400亿美元，对应的芯片数量分别为78/43/26/60万张。
- 假设2：H100、GB200、Rubin、950DT租赁价格分别按2.5/6/10/3美元/h/GPU测算。

1GW数据中心商业闭环					模型厂商	Opus5 GB200	Opus5 Rubin	Opus5 H100	GLM5.2 950DT
单位：亿美元					芯片				
云厂					收入（亿美元/GW/年）	528	1343	179	139
GB200					TPS (Token/GPU)	2037	8636	377	1627
Rubin					Token价格(美元/百万Tokens)	3.9	3.9	3.9	0.9
H100					毛利率	79%	92%	52%	43%
950DT					收入/GW（亿美元）	528	1343	179	139
总成本					算力成本	224	224	171	158
CAPEX					训练成本	112	112	85	79
综合折旧年限					占比	50%	50%	50%	50%
折旧					推理成本	112	112	85	79
运营成本					占比	50%	50%	50%	50%
运营成本/capex					其他运营成本	106	269	36	21
收入（亿美元/GW/年）					占收入比	20%	20%	20%	15%
1GW数据中心对应B200数量					OP	198	850	-28	-40
B200租赁价格（美元/h/GPU）					OPM	37%	63%	-16%	-29%
毛利率									
OP									
OPM									

资料来源：英伟达官网、路透社、彭博，国信证券经济研究所整理测算

核心结论：从27年开始数据中心建设GW远低于芯片出货量GW

- 目前CSP每GW的收入变现在100-150亿美元/GW/年，考虑早期CSP与大模型厂商的战略合作、打折供应，当前大模型厂商的平均算力成本80多亿美元/GW/年。
- 由于硬件产业链涨价、且新产品迭代，算力边际建设价格增长到400-550亿美元/GW，按照五年使用寿命计算，每年折旧价值超100亿美元。
- 2026年两家头部模型厂商（Anthropic、OpenAI）占用超10GW算力，测算每GW算力贡献约100亿美元收入/年（训练推理各占一半算力）。若采用GB200或Rubin推理最新代模型，1GW年收入可达500、1000亿美元。根据大模型变现效果可得，Neocloud 26Q2较高的短单价格甚至超过终端变现能力，不可持续（Nebius 26Q2业绩会表示算力租赁短期1年内订单400-500亿美元/GW/年）。
- 我们预计从27年开始数据中心建设GW远低于芯片出货量GW，甚至不足一半芯片能在当年转化为算力。
- 若按假设一全球实际建设GW受到资本开支规模限制，27、28年实际落地约37GW、46GW。若其中一半由SOTA模型厂使用（Anthropic、OpenAI），按当前的变现效率两家27、28年收入约3800、7000亿美元（假设新增算力70%为推理需求）。

表：每年加速器芯片对应GW与数据中心增量（单位：GW）						
	2025	2026E	2027E	2028E	2029E	2030E
芯片出货量对应GW数	13	33	62	105	129	152
假设一：北美区域建设能力为最小约束，建设全球实际建设GW受到资本开支规模限制，按照此前资本开支测算						
北美建设GW	9	16	20	23	26	30
全球建设GW	13	27	37	46	57	60
假设二：若北美建设占比全球建设容量比例维持65%，因北美数据中心ROI显著更高，测算全球算力与资本开支情况						
北美建设GW	9	16	20	23	26	30
全球建设GW	13	24	30	35	40	46

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理测算，详细测算过程见对应章节

38

核心结论：预计AI芯片与供应链订单承压，客户待建库存拉长



- 若按假设二全球实际建设GW受到北美建设限制，**27、28年实际落地约30GW、35GW**。若其中一半由SOTA模型厂使用（Anthropic、OpenAI），按当前的变现效率两家**27、28年收入约2700、4500亿美元**（假设新增算力70%为推理需求）。其收入增长更符合公司指引，**在当前增长下，北美数据中心仍处于紧平衡，算租价格和算力ROI可持续处于高位**。若建设落地算力超过该水平，则算力ROI将下降（终端收入变现无法显著提升），算租价格尤其非北美区域无法维持。
- ①AI芯片与供应链订单承压，客户待建库存拉长：**芯片订单可能从“抢产能”转向“按项目交付”，供应链增长预期承压。部分芯片会以“待部署库存”形式积压在客户端，形成在建工程或闲置资产，库存积压或导致二级市场折价，并拉低云厂商/建设方ROI。
- ②通过“芯片+基建”捆绑模式，芯片厂担保资金、兜底收入后压货给承建方：**英伟达、谷歌等芯片厂商从卖芯片转向“芯片交付+建设担保+收入兜底”的打包方案。例如英伟达联合金融机构担保5000亿美元给承建方（多是新云），并为OpenAI 俄亥俄园区提供千亿级土建电力的残值担保，本质是打通下游环节、确保芯片出货。
- ③算力可维持较高价格，北美区域溢价显著：**落地算力的稀缺性直接延后了市场预期的“算力价格战”，尤其北美算力现货预计能维持现有高ROI和高毛利水平。

表：每年加速器芯片对应GW与数据中心增量（单位：GW）						
	2025	2026E	2027E	2028E	2029E	2030E
芯片出货量对应GW数	13	33	62	105	129	152
假设一：北美区域建设能力为最小约束，建设全球实际建设GW受到资本开支规模限制，按照此前资本开支测算						
北美建设GW	9	16	20	23	26	30
全球建设GW	13	27	37	46	57	60
假设二：若北美建设占比全球建设容量比例维持65%，因北美数据中心ROI显著更高，测算全球算力与资本开支情况						
北美建设GW	9	16	20	23	26	30
全球建设GW	13	24	30	35	40	46

资料来源：英伟达、各公司官网、SemiAnalysis、国信证券经济研究所整理测算，详细测算过程见对应章节

- ① **AI商业化落地不及预期，算力ROI长期承压：**当前巨头持续加码算力资本开支，但AI付费转化、企业客户付费渗透速度低于乐观测算，算力租赁价格存在下行压力，折旧、电力、运维刚性成本高企，或导致云厂商AI业务长期无法实现正向现金流，资本回报周期拉长。
- ② **高额发债推升融资成本，利息支出侵蚀利润：**2026年科技巨头发债规模创历史新高，市场认购热度边际走弱，融资利率上行；Meta、亚马逊等企业债务规模持续扩张，每年利息支出大幅增加，对冲AI业务增量收益。
- ③ **地缘与供应链约束，芯片交付不及规划：**英伟达高端GPU、AMD定制芯片存在出口管制与产能约束，厂商算力采购交付进度存在不确定性；自研芯片量产爬坡进度慢于预期，无法有效对冲外部芯片采购成本与供给缺口。
- ④ **芯片快速迭代引发大额资产减值风险：**AI芯片迭代周期缩短至2-3年，Blackwell、Rubin新架构性能跨越式提升，上代H100/H200硬件快速贬值；头部企业数千亿GPU资产存在大额减值计提压力，压制企业利润水平。
- ⑤ **算力供给过剩，行业陷入价格内卷：**全球云厂商、第三方算力租赁企业同步大规模扩产，中长期算力供给增速高于AI需求增速，机房利用率下滑，算力租赁价格持续下行，削弱算力租赁业务毛利率。

免责声明

国信证券投资评级			
投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即报告发布日后的6到12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普500指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所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或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。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。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。