

人工智能行业专题

全球云厂商复盘，AI 驱动新一轮资本开支周期

优于大市

核心观点

全球云厂商复盘：政策驱动+战略调整，拉动云厂商资本开支增长。从海外来看，亚马逊最先进入云计算领域，从 IaaS 业务逐步向 PaaS、SaaS 业务拓展，为全球云计算领导者，在 IaaS 领域具有领先优势。微软、谷歌从 PaaS、SaaS 业务逐步切入 IaaS 业务领域，微软抓住混合云行业趋势，市场份额快速提升；谷歌 18 年实施“追赶战略”后，云计算业务开始快速发展。从国内来看，阿里巴巴为国内云计算领导者，得益于移动互联网流量快速增长，云计算业务快速发展。从资本开支增长维度来看：1) **政策驱动**——2010 年美国制定“云优先”的发展战略，全球云厂商资本开支快速增长；2) **战略调整**——14 年微软提出“云为先”战略，18 年谷歌提出“追赶战略”后，公司资本开支出现显著提升。自 23 年以来，全球人工智能快速发展，云厂商大力进行 AI 基础设施建设，驱动新一轮资本开支上升周期。

支出侧：从 ROIC 维度观察云厂商资本开支变动。资本开支持续上行，将对折旧摊销、利润、现金流、ROIC 等财务指标产生影响。根据历史数据，资本开支增加后大约 5 个季度对利润产生影响（从服务器采购到数据中心建成，存在时间差）；折旧及摊销费用占到营业利润 25%左右可以作为一个转折点，此时费用对利润产生很大压力；在转折点之后资本开支预期还会上涨 2-3 个季度，后续会同比下行，预计 FY25Q2 资本开支将进入稳定期。

回报侧：云厂商 AI 云租赁回报率测算。1) **大型云厂商：**仅考虑单卡采购成本和单卡电费，单卡 H100 对应 24-29 年回报率分别为 51%、167%、257%、326%、379%、419%；若考虑智算中心建设的综合成本，则中性预期下 24-29 年对应回报率分别为-14%、52%、102%、140%、168%、189%；2) **小型云厂商：**根据各公司官网统计数据，小型云厂商租赁价格显著低于大厂，仅考虑单卡采购成本和单卡电费，24-29 年回报率分别为-50%、-4%、39%、79%、115%、148%。

设备厂商业绩复盘：云厂商资本开支先验指标。设备厂商业绩波动与云厂商的资本开支直接挂钩，对比主要设备厂商思科、intel 等营收增速与云厂商资本开支增速，观测到设备厂商营收变化领先云厂商资本开支约 2 个季度。

格局演绎：云时代设备厂商群雄逐鹿，AI 时代再迎新王。AI 驱动云设备厂商格局变化，部分厂商市占率快速提升，例如 Arista、Supermicro；AI 芯片需求高速增长，英伟达以芯片性能、生态领先对手，占据统治地位。

风险提示：全球云厂商资本开支不及预期；AI 应用商业化不及预期等。

投资建议：行业趋势+公司战略，AI 硬件厂商率先受益。结合行业维度（AI 云是大势所趋）和公司维度（各家将 AI 发展放到核心地位），AI 驱动本轮云厂商资本开支上升周期。根据复盘硬件厂商同云厂商资本开支的关系，硬件厂商的业绩反应往往领先于云厂商资本开支的投入，建议关注 AI 云计算相关硬件厂商，例如海光信息、浪潮信息等。

重点公司盈利预测及投资评级

公司代码	公司名称	投资评级	昨收盘 (元)	总市值 (百万元)	EPS		PE	
					2024E	2025E	2024E	2025E
688041	海光信息	优于大市	125.02	290,589	0.78	1.00	160.28	125.02
000977	浪潮信息	优于大市	43.85	64,553	1.24	1.45	35.36	30.24

资料来源：Wind、国信证券经济研究所预测

行业研究 · 行业专题

计算机

优于大市 · 维持

证券分析师：熊莉

021-61761067

xiongli1@guosen.com.cn

S0980519030002

证券分析师：艾宪

0755-22941051

aixian@guosen.com.cn

S0980524090001

证券分析师：库宏焱

021-60875168

kuhongyao@guosen.com.cn

S0980520010001

市场走势



资料来源：Wind、国信证券经济研究所整理

相关研究报告

《计算机行业 2024 年 10 月投资策略-化债政策超预期，深度关注 AI 基建》——2024-10-22

《汽车智能化系列专题之决策篇（6）：特斯拉发布 Robotaxi，入华进程有序推进》——2024-10-12

《人工智能行业专题：OpenAI 发布 GPT-01，模型能力持续提升》——2024-10-11

《央行新货币政策点评-央行出台新货币政策支持股市稳定发展，关注金融 IT 投资机会》——2024-09-26

《字节全新发布豆包 AI 视频模型，AI 多模态有望迎来爆发期》——2024-09-24

内容目录

全球云计算厂商复盘：数字时代的引擎	7
亚马逊：全球云计算领导者，AWS 收入持续增长.....	7
微软：IaaS+PaaS+SaaS 组合拳战略，Azure 收入占比持续提升.....	15
谷歌：PaaS 业务先行，开源生态领导者.....	20
阿里巴巴：中国云计算先行者，进入战略转型期.....	24
支出侧：从 ROIC 维度观察云厂商资本开支变动	29
云厂商资本开支直接影响硬件侧公司收入.....	29
微软资本开支与折旧 ROIC 变化.....	30
谷歌云资本开支与折旧 ROIC 变化.....	35
阿里云资本开支与折旧 ROIC 变化.....	37
回报侧：云厂商 AI 云租赁回报率测算	39
大型云厂商：H100 算力卡云租赁回报率测算（以微软为例）.....	39
中小云厂商：H100 算力卡云租赁回报率测算.....	42
设备厂商业绩复盘：云厂商资本开支先验指标	44
思科：全球网络设备领导者.....	45
英特尔：CPU 计算领域全球领导者.....	48
浪潮信息：国产服务器领导者.....	51
格局演绎：云时代设备厂商群雄逐鹿，AI 时代再迎新王	53
服务器、交换机市场持续增长，AI 受益厂商份额快速提升.....	53
AI 时代掀起新一轮资本开支，英伟达仍是 AI 芯片王者.....	55
投资建议	58
风险提示	58

图表目录

图 1: 24Q1 亚马逊云计算业务市占率全球第一	7
图 2: 2023 年全球 IaaS 业务市占率情况	7
图 3: 23Q4 亚马逊为全球 SaaS 业务领导者	7
图 4: 亚马逊是全球 AI 云服务领域领导者	7
图 5: 23 年 AWS 业务收入占比持续增长	8
图 6: 前期 (2001 年-2006 年) 亚马逊资本开支快速增长	9
图 7: 亚马逊 S3 基本存储架构	9
图 8: AWS 私有云	10
图 9: AWS Marketplace workflow	10
图 10: AWS 开始自研芯片 Graviton、Inferentia	12
图 11: AWS 对全球大多数地区完成覆盖	12
图 12: AWS 资本开支逐季增长	13
图 13: AWS 开放 P5 实例 (由英伟达 H100 支持)	13
图 14: 公司 AWS 收入历史变化情况	13
图 15: 亚马逊资本开支与 AWS 收入历史变化情况	14
图 16: 11-18 年亚马逊资本开支季度变化情况	15
图 17: 14-21 年全球公有云市场 (IaaS) 份额排名前五厂商	15
图 18: 微软云收入持续增长	16
图 19: 微软云收入占比持续提升	16
图 20: 微软发布 Windows Azure Platform	16
图 21: 微软 SaaS+PaaS+IaaS 组合拳	16
图 22: 混合云逐步成为主流, 23 年占比 89%	17
图 23: 微软 Azure Stack 成为企业使用最多的私有云	17
图 24: 微软全线产品整合 ChatGPT, 发布 Copilot	17
图 25: 微软 Azure ND H100 v5 VM 正式上线	17
图 26: ND H100 v5 VM 产品	18
图 27: 微软自研 Maia 100 芯片	18
图 28: 微软资本开支与微软云收入历史变化情况	19
图 29: 11-18 年微软资本开支季度同比变化情况	19
图 30: 11-18 年微软资本开支季度环比变化情况	20
图 31: 19 年后谷歌云收入快速增长	20
图 32: 谷歌云业务收入占比持续提升	20
图 33: Google App Engine	21
图 34: Google Compute Engine 产品	21
图 35: 2017 年 69% 企业愿意选择 Kubernetes 作为平台工具	21
图 36: 谷歌云同微软 Azure 云的差距逐步拉大	21
图 37: 18 年后资本开支快速增长	22

图 38: 18-23 年谷歌云市占率逐步提升	22
图 39: 谷歌 TPU 快速迭代	22
图 40: 谷歌发布第六代 TPU-Trillium	22
图 41: Google Cloud 中基于 Intel CPU 的算力实例	23
图 42: Google Cloud 中基于 AMD CPU 的算力实例	23
图 43: 谷歌资本开支与谷歌云收入历史变化情况	24
图 44: 阿里巴巴云业务收入快速增长	24
图 45: 阿里云业务收入占比持续提升	24
图 46: 23H2 中国公有云 IaaS 厂商市场份额情况	25
图 47: 23H2 中国公有云 PaaS 厂商市场份额情况	25
图 48: 14-19 年移动互联网流量快速增长	26
图 49: 阿里云全球布局	26
图 50: 亚太地区云计算 IaaS 市场 (2023 年)	26
图 51: 垂直行业云计算市场成为增长主要动力	27
图 52: 2022 年中国各行业用量占比情况	27
图 53: 中国政务公有云市场份额 (2023 年)	27
图 54: 中国政务专属云市场份额 (2023 年)	27
图 55: 阿里云魔搭社区-中国最大的模型社区	27
图 56: 中国 AI 公有云服务市场 2023 年市场份额	27
图 57: 阿里巴巴资本开支与阿里云收入历史变化情况	28
图 58: 英伟达的收入结构 (亿美元)	29
图 59: 英伟达芯片的客户销售情况	29
图 60: 24Q2 英伟达头部“直接”客户对数据中心收入贡献	29
图 61: 24Q2 英伟达头部“间接”客户对数据中心收入贡献	29
图 63: 固定资产投资设置在总资本开支中的占比	30
图 64: 调整后的 ROIC 计算	30
图 66: 随着折旧及摊销费用对利润影响加剧, ROIC 明显下滑	31
图 67: 上一个周期与本周起 ROIC 下降趋势基本吻合	31
图 68: 不同类型的固定资产折旧年限不同	32
图 69: 基于公司披露的数据计算当期折旧费用	32
图 74: 微软 FY16-20 资本开支 (不含融资租赁)	35
图 75: 微软 FY16-20 单季度营收	35
图 82: “芯片-设备厂商收入-云服务商资本开支-云厂商收入”连续作用机制	44
图 83: 海外头部云厂商资本开支总和变化 (单位: 亿元)	45
图 84: 思科统一计算系统	45
图 85: 思科云安全解决方案	45
图 86: Cisco HyperFlex 提供其他 HCI 的 3 倍性能	46
图 87: 适用于多云环境的云 ACI	46
图 88: Cisco 在 AI 方面的收购与投资	46
图 89: Cisco 宣布与 NVIDIA 合作推出突破性 AI 集群解决方案	46
图 90: 思科单季营收与云厂商资本开支 YoY 比较	47

图 91: 思科股价与营收比较	47
图 92: 思科 2007 年至 2023 年主要产品的市场份额变化 (单位: %)	48
图 93: 英特尔至强处理器 5500 系列服务器	48
图 94: 英特尔虚拟化技术架构	48
图 95: 英特尔以数据为中心的产品技术组合	49
图 96: 第二代英特尔至强可扩展处理器	49
图 97: AI 赋能全新应用场景	49
图 98: 英特尔酷睿 UltraU/H 系列处理器	49
图 99: Intel 单季营收与云厂商资本开支 YoY 比较	50
图 100: Intel 股价与营收比较	50
图 101: 2010 年至 2023 年 Intel 桌面端 CPU 与移动端 CPU 的市场份额变化 (单位: %)	50
图 102: 浪潮“云海”云计算战略架构	51
图 103: 浪潮云海 OS V2.0 三层架构	51
图 104: 浪潮云分布式架构	51
图 105: 浪潮云服务器 ECS 专业性能场景	51
图 106: 浪潮信息 AI 赋能相关产品	52
图 107: 浪潮信息边缘云代表 IEC 产品架构	52
图 108: 浪潮信息营收与国内云厂商资本开支 YoY 比较	52
图 109: 全球服务器市场规模预测 (亿美金)	53
图 110: 2023 年全球服务器市场份额	53
图 111: 近年来全球服务器厂商市场份额变化	54
图 112: 全球 AI 数据中心以太网交换机市场规模 (亿美元)	54
图 113: 2022 年数据中心交换机市场份额	54
图 114: 近年来全球交换机厂商市场份额变化	55
图 115: 各海外互联网厂商资本开支季度性变化	56
图 116: 各海外互联网厂商资本开支季度性增速变化	56
图 117: AI 数据中心芯片市场规模预测	56
图 118: 全球服务器 GPU 市场份额变化	57
图 119: 2023Q3 英伟达各客户采购 AI 芯片数量预估 (千块)	57

表1: 假设资本开支中 38%用于计算设备及软件 (元)	39
表2: H100 算力相关建设成本敏感性测算	39
表3: 假设微软单 GPU 租赁收入为 7.47 美元/小时	40
表4: 微软 H100 全生命周期累计收入测算	40
表5: 单 POD IT 耗能测算	40
表6: 单张 H100 对应智算中心电力消耗测算	41
表7: H100 用电成本测算	41
表8: 微软 H100 回报率测算	42
表9: 假设中小云厂商单 GPU 租赁收入为 2.49 美元/小时	42
表10: 中小云厂商 H100 全生命周期累计收入测算	42
表11: 中小云厂商 H100 回报率测算	43

全球云计算厂商复盘：数字时代的引擎

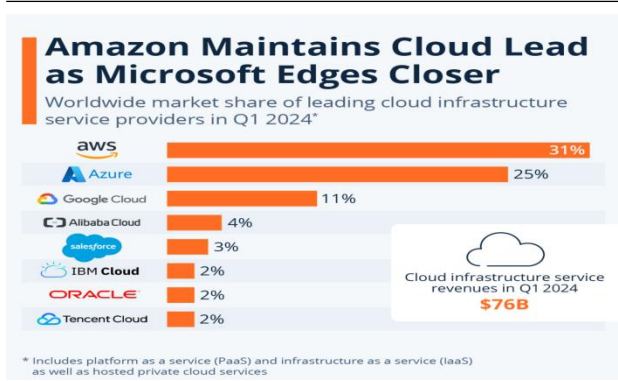
亚马逊：全球云计算领导者，AWS 收入持续增长

亚马逊：全球云计算领导者。根据 Statista、Synergy Research Group 披露数据，24Q1 全球云计算（包括 PaaS 和 SaaS）市场为 760 亿美金，同比+21%，行业快速增长；其中，亚马逊的 AWS 全球市占率为 31%，市占率第一；其次为微软的 Azure 和谷歌的 Google Cloud，市占率分别为 25%、11%，三家合计市占率达到 67%。

细分领域来看：亚马逊 AWS 在 IaaS、SaaS、AI 云服务领域具有领先优势。

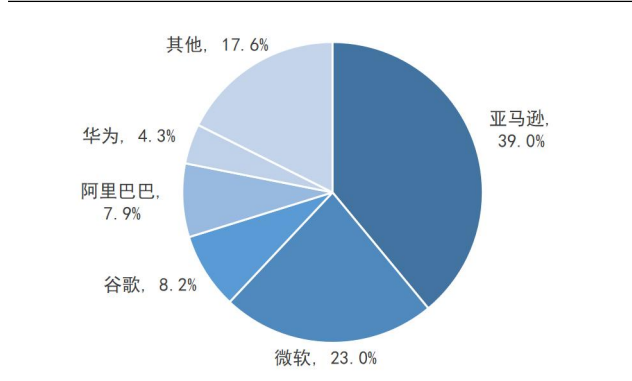
- **IaaS 领域：**根据 Gartner 数据，23 年全球 IaaS 市场 1399.99 亿美金，同比 +16.2%；其中，亚马逊实现收入 546.48 亿美金，市占率为 39.0%，全球第一；
- **SaaS 领域：**AWS 具有广泛的云服务组合，提供 Marketplace 平台，允许独立软件供应商（ISV）在平台提供软件服务，被 Gartner 评为 SaaS 业务领导者；
- **云 AI 开发者服务领域：**AWS 提供完整的生成式 AI 技术堆栈，覆盖算力（英伟达 GPU 实例）、基础模型（Amazon Bedrock 等）和应用（AI 编程助手 Amazon CodeWhisperer 等），被 Gartner 评为云 AI 开发者服务领导者。

图1: 24Q1 亚马逊云计算业务市占率全球第一



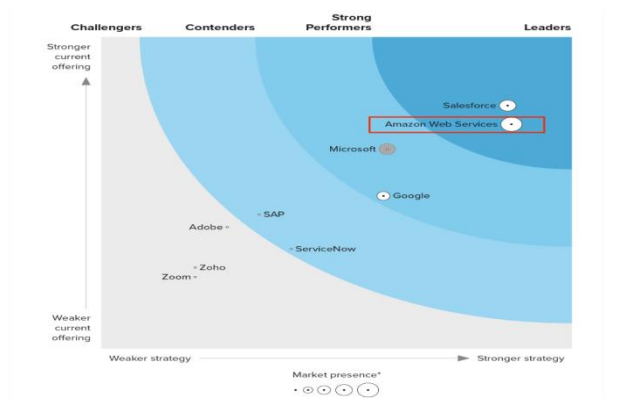
资料来源: Statista, Synergy Research Group, 国信证券经济研究所整理

图2: 2023 年全球 IaaS 业务市占率情况



资料来源: Gartner, 国信证券经济研究所整理

图3: 23Q4 亚马逊为全球 SaaS 业务领导者



资料来源: Gartner, 国信证券经济研究所整理

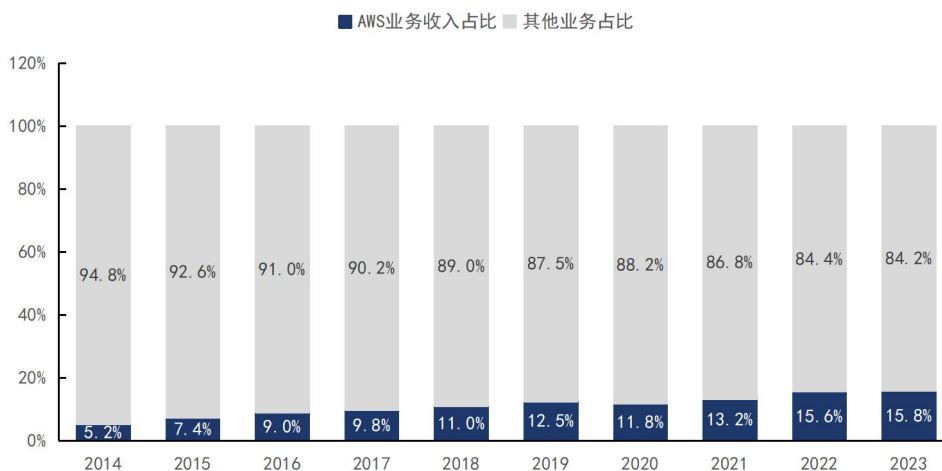
图4: 亚马逊是全球 AI 云服务领域领导者



资料来源: Gartner, 国信证券经济研究所整理

亚马逊：AWS 业务收入占比持续提升。根据公司披露数据，23 年 AWS 业务收入 907.57 亿元，对应 14-23 年 CAGR 为 39.1%，AWS 收入持续提升；23 年 AWS 业务收入占比为 15.8%，相比于 14 年+10.6 个 pct，AWS 业务收入占比持续提升。

图5：23 年 AWS 业务收入占比持续增长



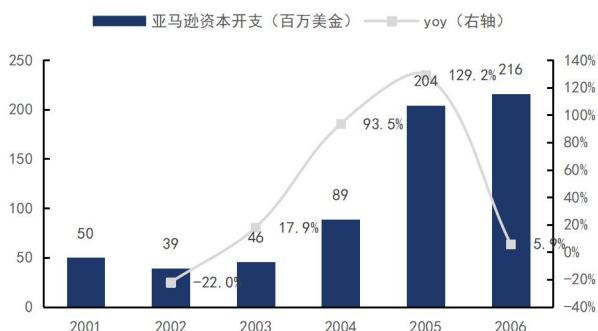
资料来源：公司财报，国信证券经济研究所整理

亚马逊 AWS 业务历史复盘：亚马逊 2006 年发布产品 EC2、S3，标志着 AWS 业务正式启动，后续逐步从单一的 IaaS 业务拓展至 PaaS、SaaS 业务，且产品矩阵持续丰富；同时，受益于美国“云优先”、“云敏捷”战略，云计算业务快速发展。22 年受 AI 大模型对训练需求的提升，积极布局 AI 算力，23 年 7 月 AWS 推出有英伟达 H100 支持 EC2 P5 实例。

- **缘起（2004 年-2006 年）：闲置 IT 资源的再利用。**1) **供给方**：亚马逊以在线零售商起家，为应对圣诞节等流量极速攀升的时间节点，需要购置远超其平时使用量的硬件资源以支撑平台交易稳定运行。一方面，亚马逊的资本开支快速增长，成本上升；另一方面，根据《云计算导论：概念、架构与应用（武志学，人民邮电出版社）》披露数据，亚马逊数据中心在大部分时间只有不到 10% 的利用率，IT 资源大量闲置。2) **需求方**：对于美国大多数中小企业来说，自建数据中心的成本较高，且扩容的灵活性较差。3) **发布产品**：2005 年，亚马逊发布 Amazon Web Service 计算平台，并于 2006 年 3 月、8 月相继推出在线存储服务 S3（Amazon Simple Storage Service）和弹性计算云 EC2（Amazon Elastic Compute Cloud）。
 - **S3 是 AWS 推出的第一款产品**：一个完全针对互联网的数据存储服务，应用程序可以通过一个简单的 Web 服务接口，就可以访问 S3 上的数据，其是一个扁平化的两层结构（存储桶和存储对象），具有耐久性和可用性、弹性和可拓展性、接口简单等优势。
 - **EC2 是 AWS 的核心产品**：1) **更便捷**：EC2 是业界第一个将基础架构大规模开放给公众用户的云计算服务（IaaS），设计目的为让开发人员更容易地进行网络规模计算，开发人员仅通过简单网络服务接口就可以获得和配置服务器容量；2) **经济性和灵活性**：用户只为实际使用的容量付费，可以随时根据需求调整使用容量；3) **协同性**：同 S3 形成协同效应，在

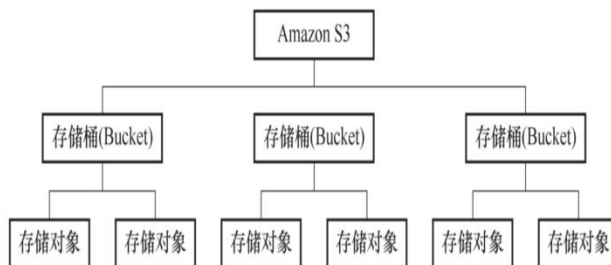
云端调取 S3 存储的数据。

图6: 前期 (2001 年-2006 年) 亚马逊资本开支快速增长



资料来源: Statista, Synergy Research Group, 国信证券经济研究所整理

图7: 亚马逊 S3 基本存储架构



资料来源: 《云图·云途: 云计算技术演进及应用》-汤兵勇主编-机械工业出版社 (2021 年) -P115, 国信证券经济研究所整理

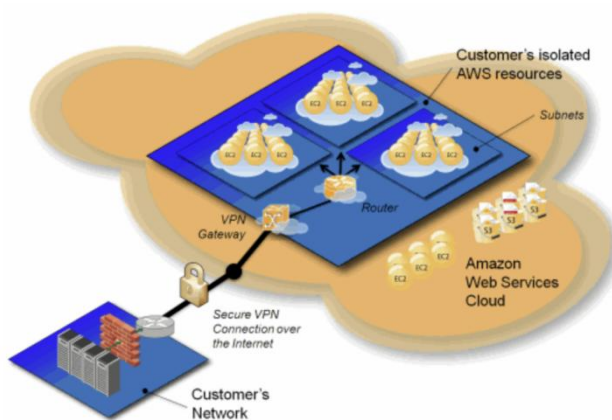
➤ **产品完善期 (2007 年-2012 年): 完善云计算产品, 从单一的 IaaS 逐步拓展至 PaaS、SaaS 领域。**

- **行业层面:** 2010 年美国联邦政府制定了“云优先”的发展战略, 为政府机构提供了广泛采用基于云解决方案的权利, 并积极鼓励其下各部门积极采用私营机构云服务以提升政府决策能力, 同时白宫在预算申请文件中将云计算列为促进美国政府技术设施的重要技术, 云计算开始进入快速发展期。
- **IaaS 产品持续完善:** 1) **EC2 实例:** 07 年推出多种 EC2 实例类型, 更加灵活; 08 年推出弹性 IP 地址 (用户可以快速将地址重新映射到账户中的另一个实例, 屏蔽实例故障) 和 EBS (弹性块存储, 开发人员可以通过编程方式穿件独立于 EC2 的存储卷); 09 年推出 EC2 新功能云监控, 开发者可以随时监控 AWS 计算资源使用情况; 10 年发布 EC2 集群计算实例, 网络吞吐量大幅提升; 12 年推出高 I/O 以及高存储 EC2 实例; 2) **S3 产品:** 10 年推出应用于 S3 的导入与导出服务, 通过 web 服务接口, 可以让用户方便地管理数据传输和迁移; 12 年发布 AWS 存储网关, 简化了本地 IT 环境与云端存储的数据移动。
- **PaaS 产品快速发展:** 亚马逊分别于 08 年、09 年、11 年、12 年发布分布式数据库 SimpleDB、RDS (亚马逊关系数据库服务)、Elastic Beanstalk 服务、非关系型数据库 DynamoDB 和 SWF (简单工作流服务), 亚马逊快速进军 PaaS 市场。
- **成立 SaaS 产品平台:** 12 年 AWS 发布 AWS Marketplace (软件在线商城), 独立软件供应商 (ISV)、增值分销商 (VAR) 和系统集成商 (SI) 可以面向 AWS 客户宣传和出售软件。
- **开辟私有云:** 09 年, AWS 推出虚拟私有云 (VPC), 并对 EC2 和 VPC 进行整合; 11 年推出政府专用云服务 (AWS GovCloud), 满足政府机构和承包商的具体监管和合规要求。
- **全球化战略-向欧洲、亚洲、南美洲、澳洲拓展:** 1) **欧洲:** 07 年推出

S3 欧洲版本，08 年 EC2 拓展至欧洲地区，09 年进一步在欧盟地区推出运行微软 Windows 服务和 SQL 服务器的 EC2；10 年在欧盟推出亚马逊私有云。2) **亚洲**：09 年进军新加坡市场，11 年在日本推出基础设施网络服务。3) **南美洲**：11 年新增南美洲圣保罗地区。4) **澳洲**：12 年新增悉尼地区。

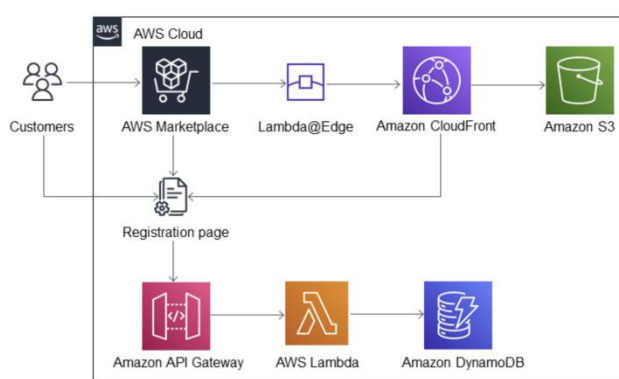
- **价格策略**：1) **分级定价**：08 年推出分级定价，提供大量折扣；2) **保留实例**：09 年推出保留实例定价策略，即用户可以选择通过一次性支付低价保留容量，进一步降低每小时使用成本；3) **竞价实例**：09 年推出竞价实例策略，用户可对 EC2 中闲置的计算能力进行竞价，并在竞价超过当前价格是运行这些实例；4) **降价策略**：截至 12 年，AWS 已经降价 27 次。

图8: AWS 私有云



资料来源：亚马逊，国信证券经济研究所整理

图9: AWS Marketplace workflow



资料来源：亚马逊，国信证券经济研究所整理

- **快速发展期（2013 年-2022 年）**：实例向更高性能、更多种类发展，PaaS 和 SaaS 产品持续完善，发布在研硬件芯片，对全球大多数地区完成覆盖。
 - **行业层面**：2018 年美国联邦政府制定“云敏捷”战略，让各机构采购可以简化转型流程并拥抱具有现代化能力的私营机构云解决方案，协助政府更好地进行信息化决策。
 - **实例向更高性能、更多种类发展**：1) **EC2**：13 年发布高内存集群实例，并对高 I/O 实例进行升级，发布 I2 实例；14 年发布 T2 实例，仅适用于做网络服务器或小型数据库，但成本最低；15 年发布 EC2 通用 M4 实例；16 年发布 X1 实例，配置 2TB 内存，同时为 EBS 提供速度 10GB/s 的专用带宽，支持大规模内存数据库、大数据处理和高性能计算；17 年发布 C5 实例，适用于计算密集型工作负载，应用场景包括高性能网络服务器、高性能计算、科学建模等；18 年发布 EC2 内存增强型实例，专用于运行大型内存数据库；20 年新增 AWS Nitro Enclaves 功能，实现在同一物理主机的不同实例之间创建额外的 CPU 和内存隔离，进一步保护 EC2 中高度敏感数据；21 年推出由 Graviton 3 支持的 C7g 实例，AMD EPYC 支持的 M6a 实例，Intel Ice Lake 支持的 R6i、C6i、M6i 实例；22 年推出由 Intel 支持的 I4i、X2iezn、C6id、M6id、R6id、C6in、Hpc6id、R7iz 实例，由 AMD EPYC 支持的 M6a、Hpc6a、R6a 实例，以及由 Graviton 支

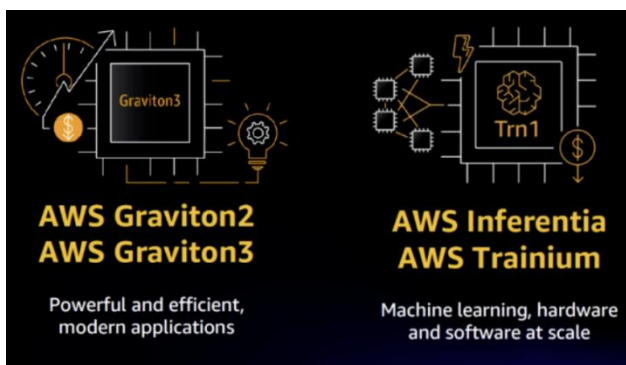
持的 C7gn 实例；2) **S3**: 14 年发布企业级云存储服务 Zocalo, 正式进入云存储企业终端市场；16 年发布 EFS 服务（弹性文件系统），用户可以在 AWS 云中设置和拓展文件存储的全托管式服务；17 年发布 AWS Macie 服务，通过 ML 和模式匹配扫描和分类 S3 数据，确保 S3 数据不包含敏感信息；19 年发布 S3 GlacierDeep Archive 存储服务，用于长期保存很少被访问的数据，价格优势明显；21 年发布 Amazon PrivateLink for S3, 用户可以使用虚拟网络中的 IP 在 S3 和本地资源之间提供私有连接，是对 VPC 网关方案的进一步优化；21 年推出 EFS 单曲存储类，同之前的 EFS 相比，在提供相同特征和优势的情况下，存储成本降低 47%；22 年推出 AWS Backup for Amazon S3, 将 Amazon S3 加入 AWS Backup 的支持服务集，同时，发布 AWS Data Exchange for Amazon S3, 使数据订阅者能够直接从数据提供者的 Amazon Simple Storage Service (Amazon S3) 桶访问第三方数据文件；3) **GPU 实例**: 13 年发布专为 3D 图形功能而设计的 G2 实例（由英伟达 Grid 处理器支持）；16 年发布 P2 实例，配置 16 个英伟达 Tesla k80 GPU, 适用于人工智能、计算流体力学等领域，同年推出 Lex、Polly 以及 Rekognition 三项人工智能服务；17 年发布 G3 图像加速实例、P3 实例，可以将机器学习训练时间大幅缩小；19 年发布 G4 实例，配置英伟达 T4 TensorCore GPU, 帮助提升机器学习推理和图形密集型工作负载的速度；20 年发布配置 NVIDIA A100 GPU 的 P4d 实例；21 年发布由自研 Inferentia 支撑的 Inf1 实例；22 年发布由自研 Inferentia 2 支撑的 Inf2 实例。

- **PaaS 产品不断完善**: 14 年发布 AWS Lambda（用户无需预先配置或管理服务器即可运行任何代码的计算服务）；15 年正式发布数据库引擎 Aurora, 兼容 MySQL 数据库价格仅有传统商业级数据库引擎的 1/10；16 年发布 AWS 数据库迁移服务，可以在不停机的情况下将 Oracle、SQLServer、MySQL、MariaDB 和 PostgreSQL 数据库从本地数据中心迁移到 AWS；19 年发布全托管式文档数据库服务 DocumentDB, 以支持现有的 MongoDB 数据库服务工作负载；20 年发布 Aurora Serverless v2；21 年发布 Amazon EMR Studio、AWS App Runner；22 年发布 Amazon CodeWhisperer（预览版），可以根据开发人员用自然语言编写的注释和集成式开发环境（IDE）中的代码生成代码建议，帮助开发人员提高工作效率，同时 Amazon Aurora 支持集群导出到 S3。
- **SaaS 产品持续丰富**: 15 年发布 WorkMail 云端上午电子邮件和日历服务、AWS Marketplace for Desktop Apps、Amazon WorkSpace Application Manager 功能，给虚拟桌面用户提供访问 Marketplace 的通道，并对其进行管理；16 年宣布用户可以直接通过 AWS Marketplace 订阅 SaaS 应用程序，简化众多 SaaS 产品的访问和计费流程；17 年发布 AWS Glue, 是完全托管式的数据目录和 ETL 服务，可以简化和自动进行数据发现和转化；19 年发布全托管式远程办公 WorkLink 服务，员工可以一键访问内部网站和网络应用程序，无需连接到 VPN；20 年发布 AppFlow 数据集成服务，开发人员可以通过其创建、管理在 AWS 和第三方 SaaS 应用程序之间的双向数据流，而无需编写定制的集成代码，解决数据孤岛问题；21 年推出 AWS Glue 自定义蓝图预览版；22 年推出 AWS Glue 中的个人信息检测和修复功能，同时，发布 Amazon Connect Cases, 为客户提供内置案例管理功能；此外，在通信安全领域，发布端到端加密的企业通信服务 AWS Wickr。
- **从私有云到混合云**: 16 年 AWS 同威睿合作，建立无缝集成混合云服务

VMware Cloud on AWS，为用户提供 SDDC 服务；17 年 AWS 和威睿发布威睿云（VMwareCloud），用户仅需在 AWS 上利用现有的威睿工具，便可在威睿 vSphere 私有、公用和混合云环境中运行应用程序；18 年同威睿发布 Amazon RDS on VMware，将亚马逊关系数据库优势融入本地虚拟化环境、混合环境以及混合云服务；22 年发布 Private 5G，可帮助企业在其设施内设置与扩展私有移动网络，同时虚拟私有云（VPC）支持面向 Amazon Polly 全面推出。

- **自研硬件芯片，收购无线路由器制造商：**在亚马逊数据中心发展到一定规模后，从成本和适配性角度考虑，亚马逊 18 年发布基于 Arm 架构 Graviton 处理器和 AI 芯片 Inferentia；19 年收购无线路由器制造商 Eero，同时发布基于 Arm 架构的 Graviton 2 处理器。
- **全球大多数地区完成覆盖：**1) **亚洲：**13 年中国区有限预览版发布；14 年新增印度地区；16 年新增韩国首尔、印度孟买地区；17 年新增中东巴林、中国宁夏地区；19 年新增中国香港地区；20 年新增印度特伦甘地地区；21 年新增日本大阪地区；2) **欧洲：**14 年新增法兰克福地区；18 年新增欧洲斯德哥尔摩地区；20 年新增米兰地区；3) **北美洲：**16 年新增加拿大地区；4) **非洲：**20 年新增非洲开普敦地区。

图10: AWS 开始自研芯片 Graviton、Inferentia



资料来源：亚马逊，国信证券经济研究所整理

图11: AWS 对全球大多数地区完成覆盖



资料来源：亚马逊，国信证券经济研究所整理

- **大力布局 AI 算力（2023 年-现在）：**在 AI 浪潮的背景下，公司大力拓展以 AI 算力（GPU 为主）支撑的智算中心，资本开支逐季度增长。
 - **行业层面：**2022 年 11 月 30 日，OpenAI 发布 ChatGPT，引发一轮由 AI 大模型驱动的 AI 浪潮，各互联网大厂及部分初创公司开始训练 AI 大模型，对 AI 算力的需求快速增长。
 - **资本开支逐季增长：**自 23Q2 开支，亚马逊资本开支环比稳步增长，24Q2 单季度资本开支为 164 亿美金，环比+10.1%、同比+57.7%，且全年指引乐观，预计 24 年 H2 资本开支要高于 24 年 H1，环比有望持续增长。
 - **大力布局 AI 算力，开放更多 AI 实例：**1) **硬件：**23 年 7 月 AWS 推出 EC2 P5 实例，由英伟达 H100 算力卡支撑；23 年 11 月，AWS 推出由英伟达 H200 支持的 P5e 实例，且成为 GH200 首家云服务供应商；另一方面，持续发力自研算力卡，23 年发布 AI 算力卡 Trainium 2，相比于第一代产品，效能提高 4 倍、能源效率提高 2 倍。2) **软件：**23 年 4 月发布 Amazon

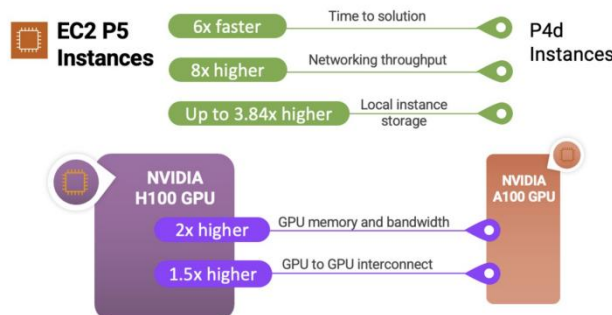
Bedrock，其是一项完全托管的服务，提供各种来自领先 AI 公司（包括 AI21 Labs、Anthropic、Cohere、Stability AI 和 Amazon 等）的高性能基础模型（FM），目前 Claude 3.5 Sonnet、Llama 3、Mistral Large 已登录 Amazon Bedrock；同时，AWS Neuron、Aurora PostgreSQL 等工具全面支持 AI 功能。

图12: AWS 资本开支逐季增长



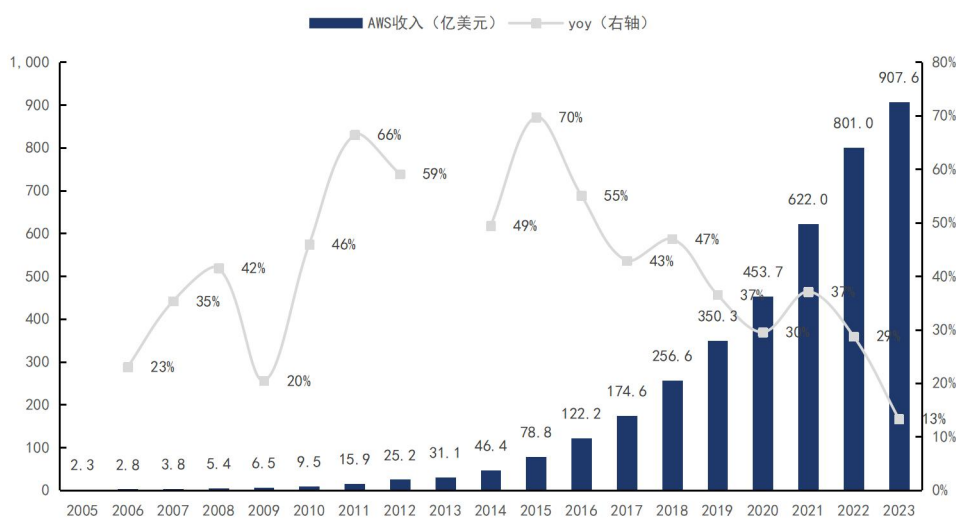
资料来源：亚马逊，国信证券经济研究所整理

图13: AWS 开放 P5 实例（由英伟达 H100 支持）



资料来源：亚马逊，国信证券经济研究所整理

图14: 公司 AWS 收入历史变化情况



资料来源：公司财报，国信证券经济研究所整理（注：05-12年 AWS 归入“其他”类收入，未单独披露，上图 05-12 年 AWS 收入为“其他”类收入）

硬件维度：从单一供应商逐步向多供应商发展。

- **CPU 处理器**：1) Intel：AWS 成立之初便采用 Intel CPU，且一直为 AWS 主流处理器。2) AMD：18 年底，AMD 发布全球首款 7nm 数据中心处理器，相比于 Intel 在处理器工艺制程占据了领先地位，AWS 于 18 年 11 月宣布采用 AMD EPYC 处理器的 m5a、r5a 实例。3) 自研芯片：18 年 AWS 发布基于 ARM 架构的自研数据中心处理器 Graviton，并发布由 Graviton 支持的实例 EC2 A1。
- **GPU 处理器（AI 算力卡）**：1) 英伟达：13 年发布专为 3D 图形功能而设计的 G2 实例（由英伟达 Grid 处理器支持），16 年发布 P2 实例，配置 16 个英伟达 Tesla k80 GPU，19 年发布 G4 实例，配置英伟达 T4 TensorCore GPU，20

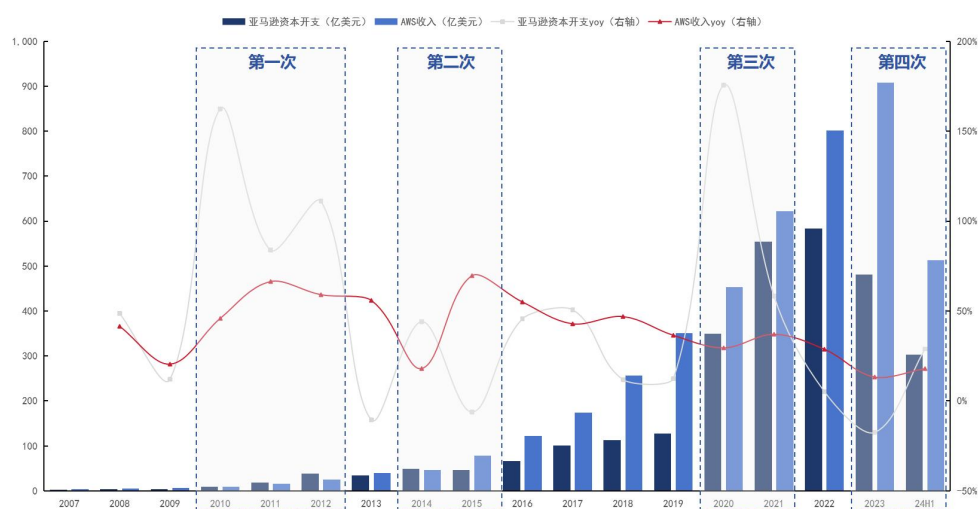
年发布配置 NVIDIA A100 GPU 的 P4d 实例，23 年 7 月 AWS 推出 EC 2 P5 实例，由英伟达 H100 算力卡支撑；23 年 11 月，AWS 推出由英伟达 H200 支持的 P5e 实例，且成为 GH200 首家云服务供应商。2) **AMD**：20 年发布基于 AMD Radeon Pro V520 GPU 的 G4ad 实例，用于图形优化的工作负载。3) **自研芯片**：21 年发布由自研 Inferentia 支撑的 Inf1 实例；18 年发布自研 AI 芯片 Inferentia，21 年发布由自研 Inferentia 支撑的 Inf1 实例，22 年发布由自研 Inferentia 2 支撑的 Inf2 实例；此外，AWS 于 20 年发布机器学习定制训练芯片 Trainium，并于 22 年上线基于 Trainium 的实例 EC2 Trn1。

亚马逊资本开支与 AWS 收入增长复盘：资本开支持续投入，AWS 收入快速增长。

从亚马逊历史数据来看，亚马逊共有 4 轮资本开支大幅增长，具体如下：

- **第一轮（2010 年-2012 年）**：10 年美国联邦政府制定了“云优先”的发展战略，云计算开始进入快速发展期，各厂商加大资本开支投入，拉动云业务收入快速增长。根据历史数据显示，AWS 云收入 10 年开始增长，11 年增速达到短期高值，且在 12、13 年维持相对高位。
- **第二轮（2014 年）**：14 年资本开支出现反弹（AWS 投入增长以及物流履约中心的建设），15 年 AWS 收入增速也随之提升，此外获得大客户 Netflix 的订单也是 15 年 AWS 收入增长的原因。
- **第三轮（2020 年-2021 年）**：疫情期间，亚马逊资本开支大幅增长，但对 AWS 收入拉动不明显，因为资本开支主要用于仓储配送中心的建设，亚马逊履约中心的规模扩大了约 1 倍，很大程度上摆脱了对第三方物流的依赖，同时在部分人口稠密地区实现了物流配送 1 天达。
- **第四轮（2024 年-至今）**：受制于 20-21 年资本开支大幅增长对财务数据的影响，亚马逊 23 年对资本开支总额保持相对谨慎态度，但资本开支中更大比例用于 AWS 基础设施，24 年资本开支同比大幅增长，发力智算中心建设。

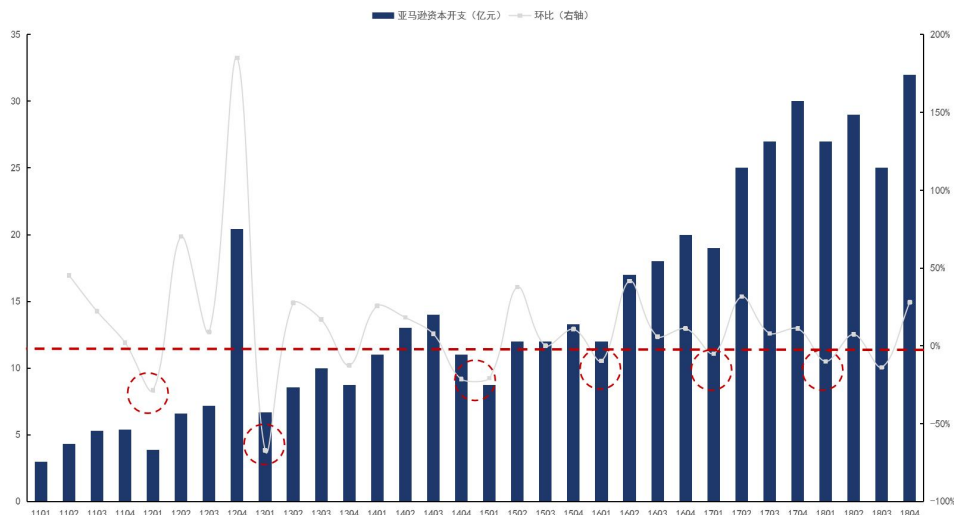
图15: 亚马逊资本开支与 AWS 收入历史变化情况



资料来源：公司财报，国信证券经济研究所整理（注：05-12 年 AWS 归入“其他”类收入，未单独披露，上图 05-12 年 AWS 收入为“其他”类收入）

资本开支季度变化维度：复盘亚马逊 11-18 年季度资本开支变化情况，平均季度资本开支环比正增长 3 个季度后，会出现资本开支环比下降（即资本开支环比负增长），主要由于资本开支持续上行会拉升折旧，进而公司业绩承压。

图16: 11-18 年亚马逊资本开支季度变化情况

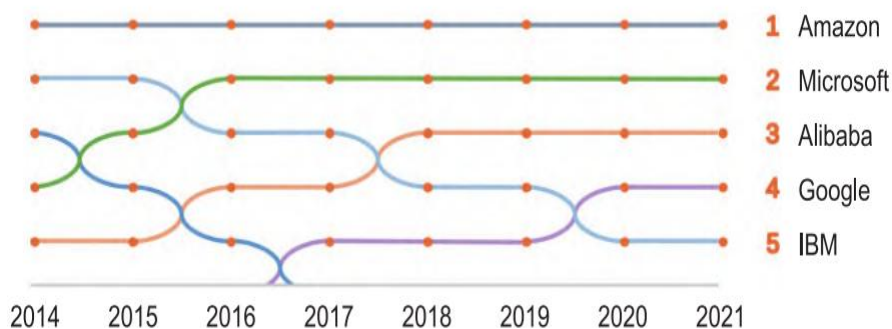


资料来源：公司财报，国信证券经济研究所整理

微软：IaaS+PaaS+SaaS 组合拳战略，Azure 收入占比持续提升

微软：Azure 收入快速增长，收入占比持续提升。 1) 从行业来看：微软 IaaS 市场份额 14 年开始持续提升，16 年后稳居全球第二；2) 从公司来看：根据公司财报披露，24 财年微软云收入 1053.6 亿元，同比+19.9%，对应 14-24 年 CAGR 为 17.1%，云业务收入快速增长。24 财年，微软云业务收入占比为 43.0%，同比+1.5 个 pct，相较于 14 财年+18.0 个 pct，收入占比持续提升。

图17: 14-21 年全球公有云市场（IaaS）份额排名前五厂商



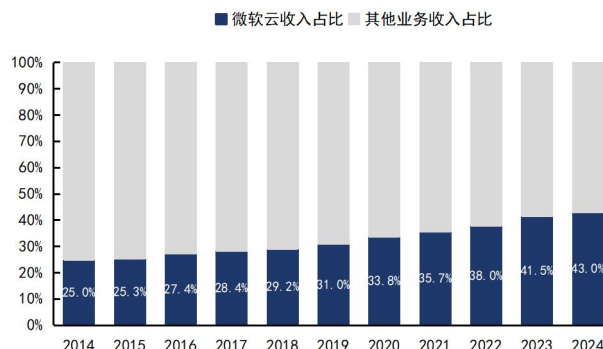
资料来源：《云计算“成功转型者”与“革命引领者”的较量-AWS 与 Azure 对比》-陈赞成-通信企业管理（2022）-P43，国信证券经济研究所整理

图18: 微软云收入持续增长



资料来源: 公司财报, 国信证券经济研究所整理

图19: 微软云收入占比持续提升

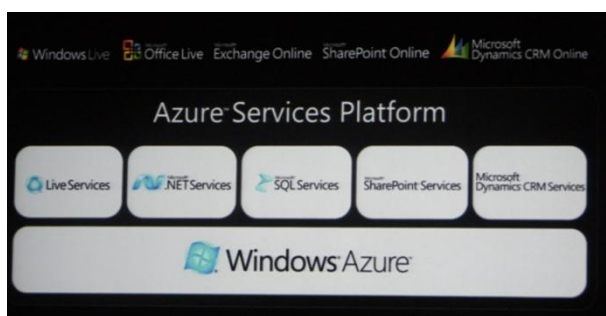


资料来源: 公司财报, 国信证券经济研究所整理

微软 Azure 云业务历史复盘: 微软利用自身在 PC 操作系统、办公软件、数据库等传统 IT 产品领域优势, 由 SaaS+PaaS 切入 IaaS 业务。14 年后确定“云为先”战略, 积极拥抱开源生态; 同时, 积极发展混合云业务, 云业务收入快速增长。23 年开始全面拥抱 AI, 将自身产品同 AI 深度融合, 同时发布 H100 实例供客户使用。

- **发展初期 (2008 年-2013 年): 由 SaaS+PaaS 切入 IaaS 业务, 打云计算“组合拳”。**2008 年, 微软发布云计算战略和平台 Windows Azure Platform, 2010 年正式推出公开服务版 (General Availability), 包含 Azure Cloud Service、Azure Storage、SQL Azure 与 AppFabric 四种服务 (均为 PaaS 产品), 标志着微软切入云计算业务; 同年, 微软发布 Office 365, 将之前的 Live@Edu 及 BPOS 两个平台归入 Office 365, 成为微软旗舰级的 SaaS 产品。2013 年微软宣布 Windows Azure IaaS 公开服务版 (General Availability) 发布, 包括虚拟机和虚拟网络功能。微软充分发挥自身在 PC 操作系统、办公软件、数据库等传统 IT 产品优势, 由 SaaS、PaaS 业务逐步切入 IaaS 业务, 使出“组合拳”快速提升全球市场份额。

图20: 微软发布 Windows Azure Platform



资料来源: 微软官网, 国信证券经济研究所整理

图21: 微软 SaaS+PaaS+IaaS 组合拳

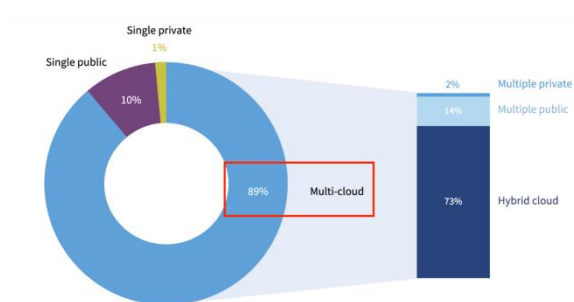


资料来源: 微软官网, 国信证券经济研究所整理

- **快速发展期 (2014 年-2022 年): “云为先”战略+开放心态双轮驱动成长, 混合云优势显现。**1) 战略定向: 2014 年, 纳德拉开始担任微软 CEO, 提出“移

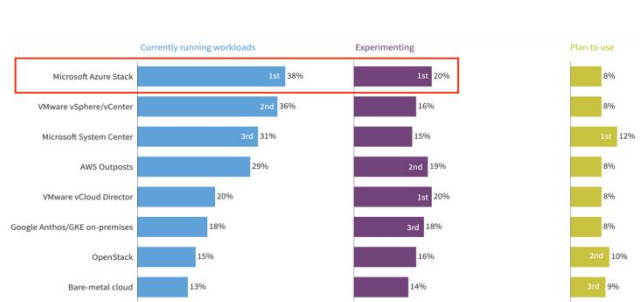
动为先，云为先”战略，微软重新定义为提供生产力和平台的专家，更名为 Microsoft Azure, 对全球用户赋能；同时，微软向竞争对手 Linux 开放 Azure、解绑 Office 与 Windows 并推广 SaaS 订阅服务，以开放的心态加入全球竞争；18 年收购开源公司 Github，拥抱开源战略，客户渗透率快速提升。2) 混合云优势显现：微软 2010 年发布私有云产品 Hyper-V Cloud, 2015 年发布 Azure Stack, 其是业内第一个可以将公有云服务、功能拓展到用户所选环境的产品，其同 HP、思科、DELL、华为、联想等厂商合作，采取“一体机”交付模式（出厂前完成软硬件及功能配置），可以有效减少部署时间、确保系统稳定性、满足客户数据主权等需求；同时，Azure Stack 与 Azure 打通，共享生态，将私有云、边缘云用户转化为混合云和公有云潜在客户，23 年微软 Azure Stack 已成为企业使用最多的私有云。

图22: 混合云逐步成为主流，23 年占比 89%



资料来源：《2024 State of the Cloud Report-Flexera》，国信证券经济研究所整理

图23: 微软 Azure Stack 成为企业使用最多的私有云



资料来源：《2024 State of the Cloud Report-Flexera》，国信证券经济研究所整理

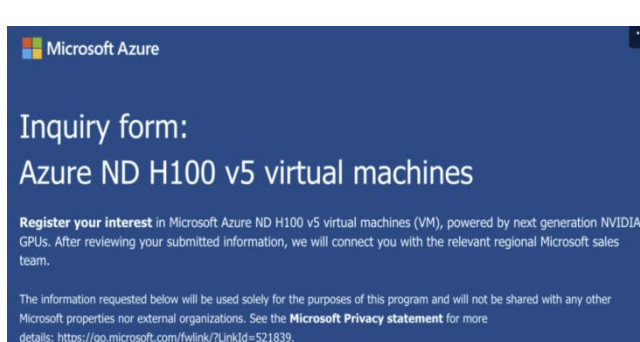
- **全面拥抱 AI 期（2023 年-现在）：**1) **软件层面：**23 年 1 月微软同 OpenAI 合作升级，OpenAI 将长期向微软云提供独家技术支持；同时，微软加速拥抱 ChatGPT, 旗下产品全线整合 ChatGPT(包括 Office 组件、Azure 云服务、Teams 等)，并对搜索引擎 Bing 进行升级，推出 Bing Chat；此外，微软云推出 Azure OpenAI, 提供 GPT-3、GPT-4 等服务供客户调用。2) **硬件层面：**23 年 8 月 Azure ND H100 v5 VM 正式上线，配置英伟达 H100 GPU 和 Intel 第四代 Xeon Scalable 处理器。

图24: 微软全线产品整合 ChatGPT，发布 Copilot



资料来源：微软官网，国信证券经济研究所整理

图25: 微软 Azure ND H100 v5 VM 正式上线

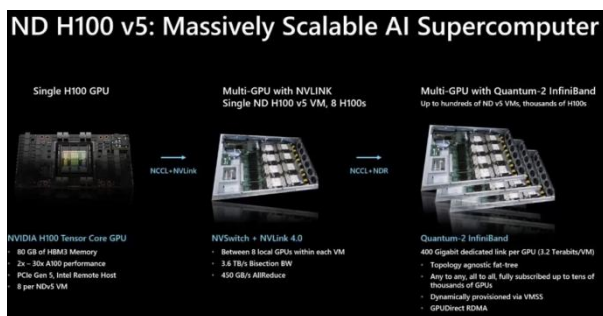


资料来源：微软官网，国信证券经济研究所整理

硬件维度：逐步拓展供应商范围。

- **CPU 处理器**：1) **Intel**：Azure 成立之初便采用 Intel CPU，且一直为 Azure 主流处理器。2) **AMD**：17 年，Azure 推出基于 AMD EPYC 处理器的 Lv 2VM 实例。3) **自研芯片**：23 年 11 月，在 Microsoft Ignite 大会发布 Cobalt 100 CPU，基于 Arm 架构，拥有 128 个核心，以自用为主。
- **GPU 处理器（AI 算力卡）**：1) **英伟达**：20 年 8 月，Azure 发布基于英伟达 A100 GPU 的 ND A100 v4 VM 实例，23 年 8 月基于英伟达 H100 GPU 的 ND H100 v5 VM 正式上线。2) **AMD**：20 年 3 月推出基于 AMD MI25 的实例 Azure NVv4 VM，22 年 5 月推出基于 MI200 的实例 HBv3，23 年 11 月，微软宣布推出基于 MI300 的云实例，24 年 5 月推出基于 MI300X 的 ND MI300X v5 VM 实例。3) **自研芯片**：23 年 11 月，发布自研 AI 芯片 Maia100，采用台积电 5nm 工艺，在 MXInt8 格式下，算力可以达到 1600TFLOPS，在 MXFP4 格式下则为 3200TFLOPS，而显存带宽仅 1.6TB/s，目前以自用为主。

图26: ND H100 v5 VM 产品



资料来源：微软发布会，国信证券经济研究所整理

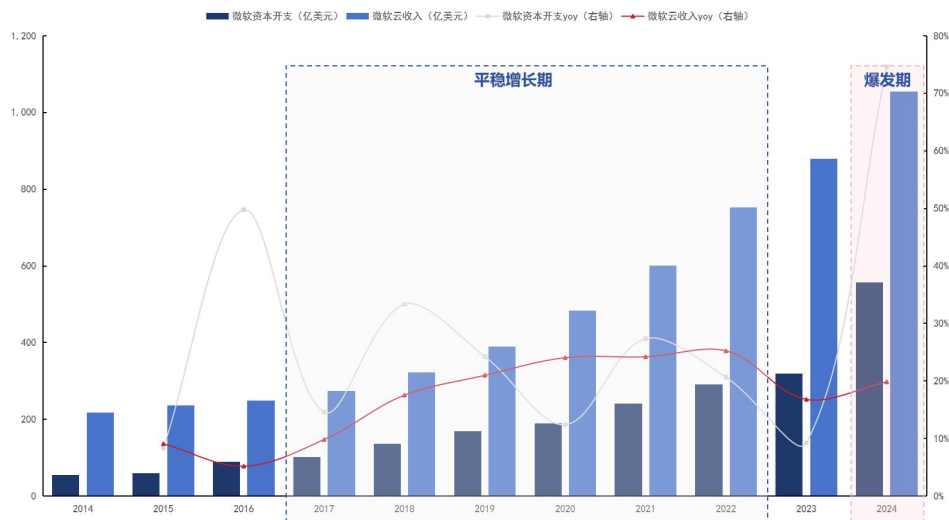
图27: 微软自研 Maia 100 芯片



资料来源：微软官网，国信证券经济研究所整理

微软资本开支与云收入增长复盘：资本开支持续增长，微软云增速稳定，24 年财年发力智算中心。从微软历史数据来看，公司资本开支波动相对较小，微软云收入增速在 16-19 年稳步提升，且在 20-22 年维持稳定。微软 24 年财年资本开支出现显著上升，大力建设 AI 数据中心。

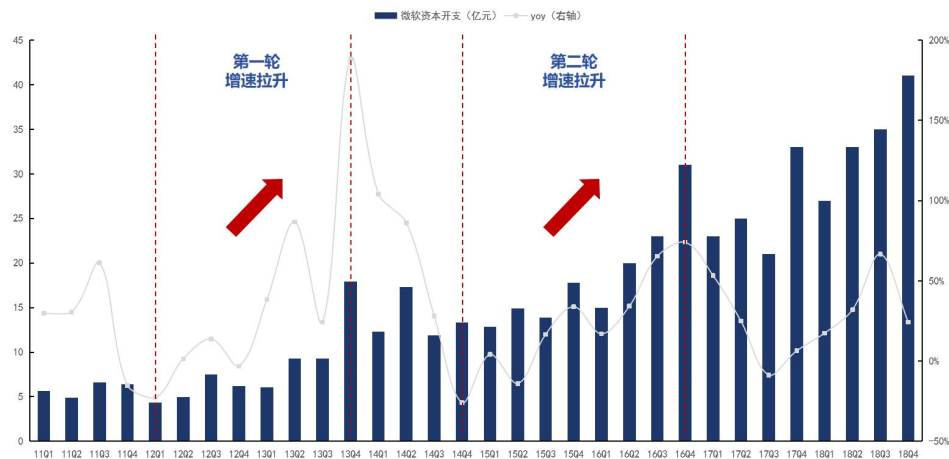
图28: 微软资本开支与微软云收入历史变化情况



资料来源:公司财报,Bloomberg,微软官网,国信证券经济研究所整理(注:14-15年资本开支引用Bloomberg数据 16-24年资本开支引用微软官网数据,以上为微软财年)

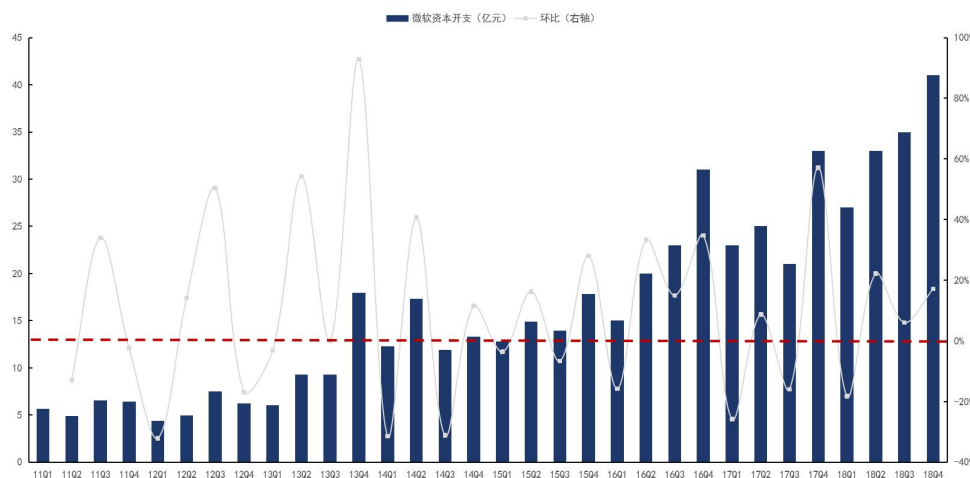
资本开支季度变化维度: 复盘微软 11-18 年季度资本开支变化情况, 从同比数据来看, 平均 8-9 个季度为一轮同比增速拉升周期(例如 12Q1-13Q3、14Q4-16Q4), 在拉升周期中, 表现为微软资本开支同比增速震荡式抬升; 从环比数据看, 微软季度资本开支环比波动较大, 多数情况下呈现钟摆式波动(即上季度资本开支环比增长, 则本季度资本开支环比下滑)。

图29: 11-18 年微软资本开支季度同比变化情况



资料来源:公司财报, Bloomberg, 微软官网, 国信证券经济研究所整理

图30: 11-18 年微软资本开支季度环比变化情况

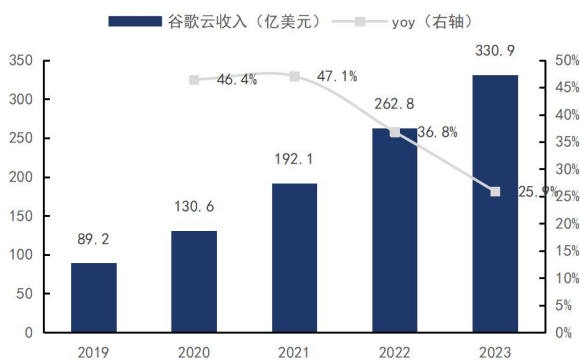


资料来源：公司财报，Bloomberg，微软官网，国信证券经济研究所整理

谷歌：PaaS 业务先行，开源生态领导者

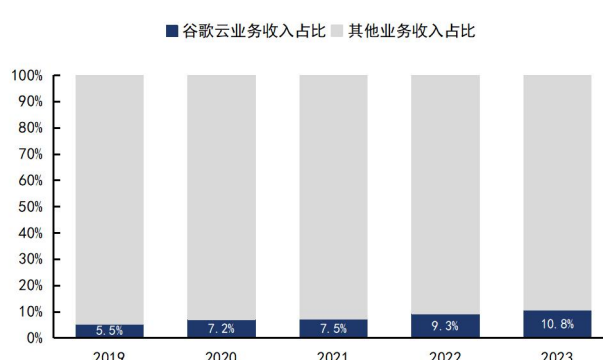
谷歌：19年后谷歌云收入快速增长，收入占比持续提升。19年，前甲骨文 CEO Kurian 出任谷歌云首席执行官，并决定了一项为期5年200亿美元的预算计划，用于大力扩张谷歌云数据中心，谷歌云收入开始快速增长；23年谷歌云业务收入330.9亿美元，同比+25.9%，对应19-23年CAGR为38.8%。同时，2019年谷歌云收入占比仅为5.5%，截止23年，谷歌云收入占比提升至10.8%，提升5.3个pct。

图31: 19年后谷歌云收入快速增长



资料来源：公司财报，国信证券经济研究所整理

图32: 谷歌云业务收入占比持续提升



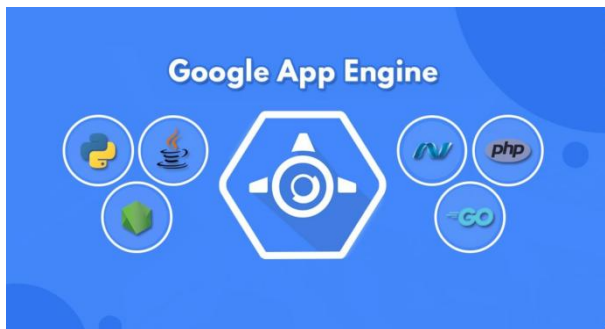
资料来源：公司财报，国信证券经济研究所整理

谷歌云业务历史复盘：最早提出云计算概念，由于前期战略摇摆、To B 经验不足，同微软 Azure 云差距逐渐拉大；谷歌自19年开始发力云业务，通过大幅增加资本开支、开源生态等方式，谷歌云市占率逐步提升，同 AWS、Azure 差距缩小。

- 业务探索期（2006年-2013年）：PaaS 产品为先，逐步完善产品矩阵。2006年谷歌在搜索引擎大会正式提出云计算概念，并在公司内部使用云计算平台；

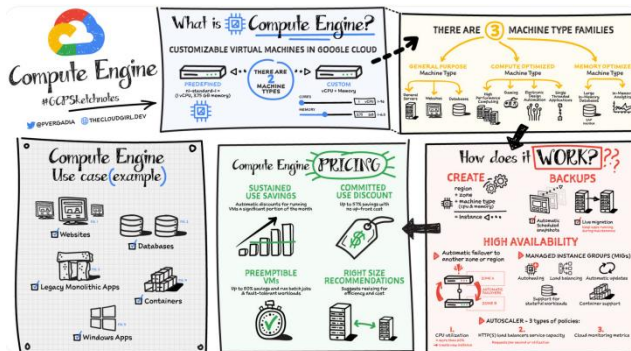
08年发布标志性 PaaS 产品——Google App Engine，其是一个开发、托管网络应用程序的平台，可提供快速开发和部署，管理简单，可轻松实现拓展性。后续谷歌陆续推出 Google Docs、Google Calendar 等 SaaS 产品。2012 年发布 Google Compute Engine 预览版，谷歌开始进入 IaaS 业务，2013 年 5 月产品正式发布，谷歌完成 PaaS+SaaS+IaaS 全产品布局。

图33: Google App Engine



资料来源：谷歌云官网，国信证券经济研究所整理

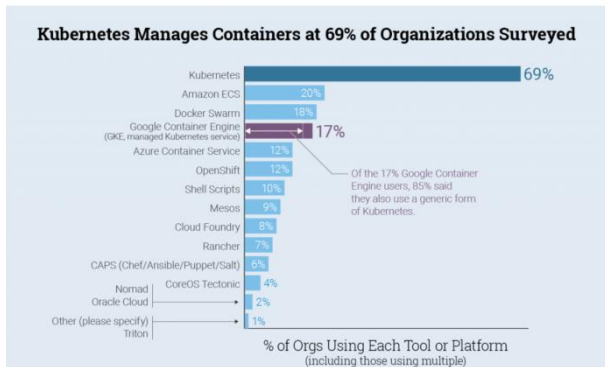
图34: Google Compute Engine 产品



资料来源：谷歌云官网，国信证券经济研究所整理

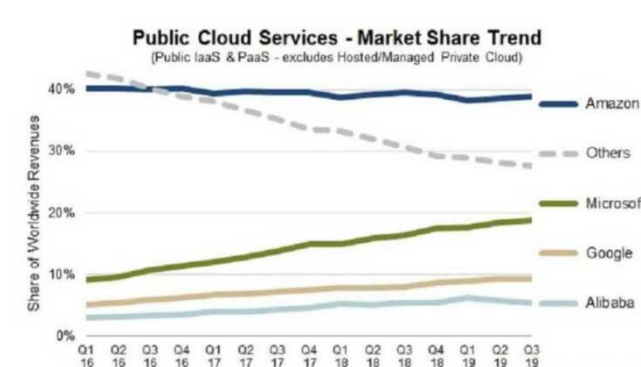
- **增长放缓期（2014 年-2018 年）：1）战略上摇摆，服务意识不足：**谷歌从战略维度上，重视搜索广告业务，没有把云业务放到战略重要位置；同时，公司做 To C 搜索业务起家，而云计算业务主要面向 B 端客户，公司缺少经验；此外，谷歌工程师文化盛行，同 AWS、微软相比，对客户服务意识不强。2）**技术主导，发布开源容器编排工具 Kubernetes：**谷歌 2014 年发布 Kubernetes，其是一个全新的基于容器技术的分布式架构领先方案，目的是实现资源管理自动化，以及跨多个数据中心的资源利用率最大化；同时，Kubernetes 是一个开源的平台，对现有的编程语言、编程框架、中间件没有任何侵入性，产品发布后成为当时市场标准，根据 The New Stack 披露数据，2017 年 Kubernetes 在平台工具选择意愿中达 69%。此外，谷歌分别在托管服务、基础设施服务中率先实现了虚拟机实时迁移、分层网络等技术。整体来看，该时期谷歌云技术持续进步，但由于业务战略定位以及商业化变现（缺乏 To B 经验）等问题，谷歌云同微软 Azure 云的差距逐步拉大。

图35: 2017 年 69%企业愿意选择 Kubernetes 作为平台工具



资料来源：The New Stack，国信证券经济研究所整理

图36: 谷歌云同微软 Azure 云的差距逐步拉大



资料来源：Synergy Research Group，国信证券经济研究所整理

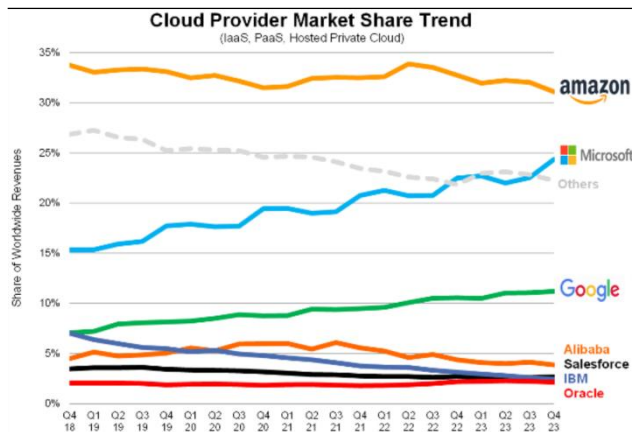
- **转型追赶期（2019年-现在）：实施“追赶战略”，加大资本投入，开放生态，积极拥抱AI浪潮。**
- 1) 战略层面：**18年制定“追赶”战略，云业务战略重要性进一步提升，19年公司任命前甲骨文 CEO Kurian 为谷歌云 CEO，大力发展云计算业务。
 - 2) 硬件层面：**谷歌决定了一项为期5年200亿美元的预算计划，用于大力扩张谷歌云数据中心，同时自研 TPU 芯片，谷歌云收入开始快速增长。
 - 3) 开源生态领导者：**谷歌宣布了基于开源软件、工作负载可移植性，以及与竞争对手坦诚合作的“开放云”远景，目标是让谷歌成为软件厂商们最容易合作的科技巨头，采用的方式包括更低的佣金、联合营销、合作销售等，并承诺不与伙伴们直接展开业务竞争。
 - 4) 积极拥抱 AI：**一方面，公司积极自研大模型（例如 PaLM 2、Gemini、Gemini 1.5 Pro 等产品），用户可以通过谷歌云 Vertex 调用模型 API；另一方面，谷歌自研 AI 算力芯片 TPU，24年5月谷歌发布第六代 TPU 芯片——Trillium，较上一代（TPU v5e）峰值计算性能提升 4.7 倍，能效比提升 67%，内存容量、带宽提升 1 倍，主要部署在谷歌云上使用。整体来看，该时期谷歌云业务快速发展，根据 Synergy Research Group 披露数据，23Q4 谷歌云市占率为 11%，相较于 18Q4，提升 4 个 pct。

图37：18年后资本开支快速增长



资料来源：公司财报，国信证券经济研究所整理

图38：18-23年谷歌云市占率逐步提升



资料来源：Synergy Research Group，国信证券经济研究所整理

图39：谷歌 TPU 快速迭代

	v4	v5e	v5p
Chips per pod	4096	256	8,960
Chip Bf16 TFLOPs	275	197	459
Chip Int8 TOPs	N/A	394	918
HBM (GB)	32	16	95
HBM BW (GB/s)	1228	820	2,765
ICI BW per chip (Gb/s)	2,400	1,600	4,800

资料来源：谷歌云官网，国信证券经济研究所整理

图40：谷歌发布第六代 TPU-Trillium



资料来源：谷歌云官网，国信证券经济研究所整理

硬件维度：自研芯片云厂商领导者。 Google Cloud 在使用 Intel、AMD 等大厂芯片的同时，发力自研芯片，15年开始发力自研 TPU 芯片，后持续迭代，24年发布第六代 TPU-Trillium；同时，24年发布基于 Arm 的 CPU 芯片 Axion，全面布局

服务器处理器芯片领域。

- **CPU 处理器**：1) **Intel**：Google Cloud 成立之初便采用 Intel CPU，且一直为 Google Cloud 主流处理器，包括 C3、M3、VE2 等实例。2) **AMD**：19 年，Google Cloud 基于 AMD EPYC 系列芯片的算力实例面向客户开放，后陆续发布 C2D、T2D、C3D 等产品。3) **自研芯片**：2024 年发布 Arm Neoverse 2 架构的 CPU 芯片——Axion，专为数据中心设计，旨在提高广泛应用的处理效率，如开源数据库、Web 和应用程序服务器、内存缓存、数据分析引擎、媒体处理及 AI 训练等，与同时期基于 X86 的同类实例相比，性能提高了 50%，能源效率提高了 60%；此外，Axion 的开放架构使得客户可以无需任何修改，即可将现有的 Arm 工作负载迁移到谷歌云。
- **GPU 处理器（AI 算力卡）**：1) **英伟达**：20 年 7 月，Google Cloud 发布基于英伟达 A100 GPU 的 VM A2 实例，是公有云中第一款基于 A100 GPU 的产品；23 年 5 月，Google Cloud 发布基于英伟达 H100 GPU 的 A3 实例，9 月正式上线。2) **自研芯片**：谷歌 2015 年发布 TPU v1，与使用通用 CPU 和 GPU 的神经网络计算相比，TPU v1 带来了 15~30 倍的性能提升和 30~80 倍的能效提升，其以较低成本支持谷歌的很多服务，仅可用于推理；17 年发布 TPU v2，用于加速大量的机器学习和人工智能工作负载，包括训练和推理；18 年发布 TPU v3，算力和功率大幅增长，其采用了当时最新的液冷技术；20 年和 21 年分别发布 TPU v4i 和 v4，应用 7nm 工艺，晶体管数大幅提升，算力提升，功耗下降；23 年 8 月和 12 月分别发布 TPU v5e、TPU v5p，24 年 5 月谷歌发布第六代 TPU 芯片——Trillium，较上一代（TPU v5e）峰值计算性能提升 4.7 倍，能效比提升 67%，内存容量、带宽提升 1 倍，产品持续迭代，性能稳步提升。

图41: Google Cloud 中基于 Intel CPU 的算力实例

INSTANCE TYPE	C3	N2	N1	C2	H3	M3	M2	M1	VE1	O2
	General Purpose	General Purpose			Compute Optimized		Memory Optimized			Google Cloud VMware Engine
	4th Gen Intel® Xeon® processor	3rd or 2nd Gen Intel® Xeon® processor	Intel® Xeon® processor	2nd Gen Intel® Xeon® processor	4th Gen Intel® Xeon® Scalable & Intel IPU	3rd Gen Intel® Xeon® processor	2nd Gen Intel® Xeon® processor	Intel® Xeon® processor	Intel® Xeon® processor	2nd Gen Intel® Xeon® processor
BENEFITS	<ul style="list-style-type: none"> Highly consistent performance Intel AMX to boost AI/ML perf Up to 200 Gbps 	<ul style="list-style-type: none"> Most flexibility and balanced price/performance Custom machine types Widely available 	<ul style="list-style-type: none"> Highest frequency CPU (3.8GHz) Widely available 	<ul style="list-style-type: none"> Enhanced execution resources for compute-bound workloads: data science, ML/AI inference, gaming, high-performance computing 	<ul style="list-style-type: none"> Up to 4TB VMs Run very large databases (SAP HANA) 	<ul style="list-style-type: none"> Up to 12TB VMs Run very large databases (SAP HANA) 	<ul style="list-style-type: none"> Up to 4TB VMs Run very large databases (SAP HANA) 	<ul style="list-style-type: none"> Run VMware natively 	<ul style="list-style-type: none"> Run Oracle and large SAP HANA (up to 24TB) 	

资料来源：谷歌云官网，国信证券经济研究所整理

图42: Google Cloud 中基于 AMD CPU 的算力实例

AMD EPYC™ Powered Machine Families on Google Cloud			
VM Series	CPU	Key Benefit	Aligned Workloads
C3D NEW	4th Gen AMD EPYC "Genoa" Up to 360 vCPUs/VM	Consistent high-performance	High traffic web, app, and ad servers Medium-to-large databases Game servers Media streaming and transcoding Data analytics CPU-based inferencing
N2D Confidential Computing Option	3rd Gen AMD EPYC "Milan" & "Rome" Up to 224 vCPUs/VM	Most flexible VM shapes	Low-medium traffic web and app servers Small-medium databases Business intelligence applications Desktop virtualization CRM applications Dev/test environment
T2D	3rd Gen AMD EPYC "Milan" Up to 60 vCPUs/VM	Single-threaded price/performance for scale-out	Containerized microservices Compression/decompression Image processing Data logging/processing Large-scale Java applications
C2D Confidential Computing Option	3rd Gen AMD EPYC "Milan" Up to 112 vCPUs/VM	Performance per core	High-performance computing (HPC) EDA/EACFD Modeling and simulation Media transcoding High-performance game servers AI/ML

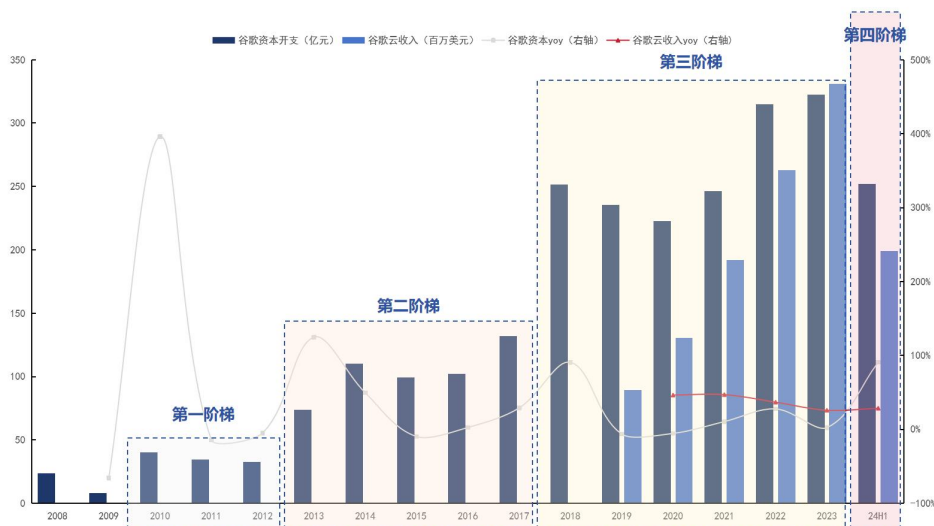
资料来源：谷歌云官网，国信证券经济研究所整理

谷歌资本开支与 AWS 收入增长复盘：资本开支呈现阶梯状增长，谷歌云收入稳定增长。从谷歌历史数据来看，公司资本开支增长可分为 4 个阶梯，具体如下：

- **第一阶梯（2010 年-2012 年）**：10 年美国联邦政府制定了“云优先”的发展战略，云计算开始进入快速发展期，谷歌亦开始加大资本开支，抢占云计算领先地位。
- **第二阶梯（2013 年-2017 年）**：13 年谷歌将 App Engine 重新命名为 Cloud Platform，正式有 SaaS、PaaS 切入 IaaS 领域，开始大幅增加资本开支，用于 IaaS 业务基础设施建设，拉动谷歌云业务增长。

- **第三阶梯（2018年-2023年）**：18年制定“追赶”战略，云业务战略重要性进一步提升，同时谷歌决定了一项为期5年200亿美金的预算计划，用于大力扩张谷歌云数据中心，同时自研TPU芯片，谷歌云收入开始快速增长。
- **第四阶梯（2024年-至今）**：谷歌发力AI算力建设，采用英伟达GPU+自研TPU双算力卡方案，同时训练自研大模型（Palm 2、Gemini等），24年H1谷歌资本开支为252亿美金，同比+91%。

图43: 谷歌资本开支与谷歌云收入历史变化情况



资料来源：公司财报，国信证券经济研究所整理

阿里巴巴：中国云计算先行者，进入战略转型期

阿里云：16-22年云业务收入快速增长，收入占比持续提升。22年阿里云业务收入745.7亿元，对应16-22年CAGR为70.7%，云业务收入快速增长。从收入占比来看，22年阿里云业务收入占比为8.7%，相比于2016年提升+5.7个pct，阿里云收入占比持续提升。从公有云市场份额来看，23H2阿里云在IaaS、PaaS业务中市占率分别为27.1%、24.7%，市占率均为国内第一。

图44: 阿里巴巴云业务收入快速增长



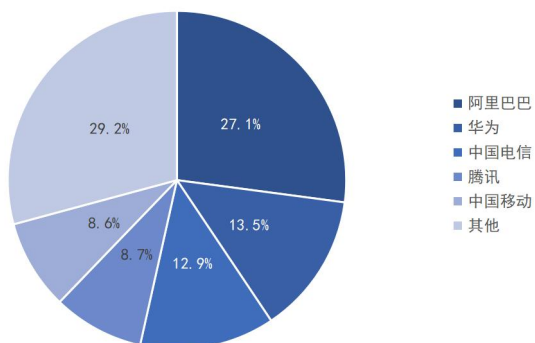
资料来源：公司财报，国信证券经济研究所整理

图45: 阿里云业务收入占比持续提升



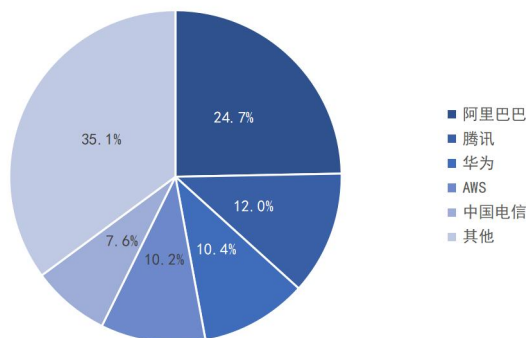
资料来源：公司财报，国信证券经济研究所整理

图46: 23H2 中国公有云 IaaS 厂商市场份额情况



资料来源: IDC, 国信证券经济研究所整理

图47: 23H2 中国公有云 PaaS 厂商市场份额情况



资料来源: IDC, 国信证券经济研究所整理

阿里云业务历史复盘：国内云计算先行者，率先进行去“IOE”，完成自研操作系统“飞天”，受移动互联网发展+海外市场拓展双轮驱动，云业务收入快速增长；21年后，华为云、运营商云发力，在政务云领域受到挤压，增长放缓；23年确定“AI推动、公有云优先”战略，进入战略转型期。

- **业务探索期（2009年-2013年）：中国云计算先行者，率先完成去“IOE”化。**2009年，阿里软件在南京建立首个“电子商务云计算中心”，标志着阿里云开始起步。同时，阿里云发力软硬件一体化全栈研发，完成了“飞天”云操作系统、“神龙”架构服务器、“盘古”分布式存储系统等产品的自研，率先完成去“IOE”（即IBM的小型机+Oracle商业数据库+EMC集中式存储）。2013年8月，阿里云正式运营服务器规模达到5000的“飞天”集群，成为中国第一个独立研发拥有大规模通用计算平台的公司，亦是世界上第一个对外提供5000云计算服务能力的公司。
- **快速增长期（2014年-2019年）：移动互联网+云计算出海双轮驱动。**1) **受益于移动互联网的快速增长：**随着智能手机、4G/5G更快通信方式的普及，14-19年移动互联网流量快速增长，根据工信部数据，19年中国移动互联网流量为1220亿GB，对应14-19年CAGR为126.2%，阿里云作为中国云计算先行者，拥有相对成熟、完善的云服务能力，充分受益于移动互联网浪潮，云计算收入快速增长。2) **云计算出海：**阿里云14年开始出海，成为中国第一家提供全球云计算服务的公司；15年开放新加坡数据中心，并设立国际业务部；17年在马来西亚首建第一个数据中心，18年建立马来西亚第二个数据中心、印度尼西亚第一个数据中心；19年开通印度尼西亚第二个数据中心。截至目前，阿里云在全球30个地区开放了89个可用区，为全球数十亿用户提供可靠的计算支持；根据Gartner数据，23年阿里云在亚太区云计算IaaS市场市占率为22.2%，位列亚太第一。

图48: 14-19 年移动互联网流量快速增长



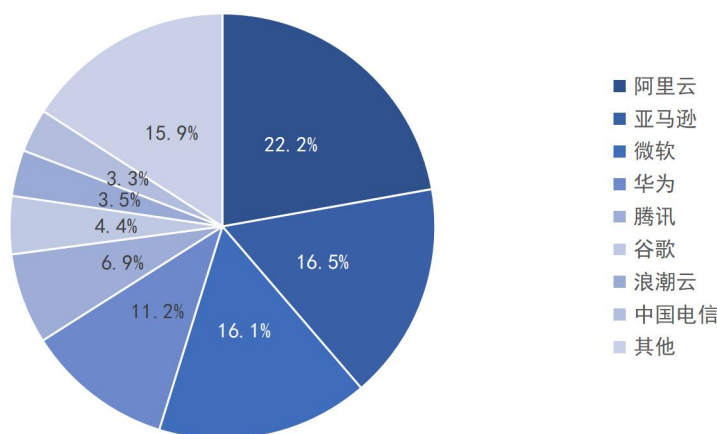
资料来源: 工信部, 国信证券经济研究所整理

图49: 阿里云全球布局



资料来源: 阿里云官网, 国信证券经济研究所整理

图50: 亚太地区云计算 IaaS 市场 (2023 年)

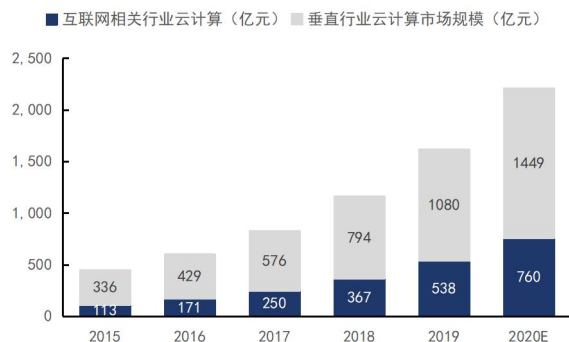


资料来源: Gartner, 国信证券经济研究所整理

- **战略转型期 (2022 年-现在): 政务云领域受挤压, 确定“AI 推动、公有云优先”战略。**
 - 1) 垂直行业云成为增长主动力:** 根据上文数据, 移动互联网流量自 20 年开始增速放缓, 根据前瞻产业研究院数据, 20 年中国云计算市场互联网行业贡献度相较于 18 年减少 16%, 以政府、国央企云为代表的垂直行业成为云计算增长的主要拉动力, 特别是疫情以来, 政府上云需求大幅增长, 根据中商产业研究院数据, 22 年政府占中国云计算的 24.9%, 仅次于互联网行业。
 - 2) 政务云领域受挤压:** 一方面, 政务云的要求同传统互联网云不同, 其更看中云厂商的定制能力, 华为、运营商云在该领域更具备优势; 另一方面, 华为等云厂商对政府等领域的销售能力更强。根据 IDC 披露数据, 23 年中国政府公有云市场, 阿里云占比仅为 13.3%, 位列第三; 在中国政务专属云领域, 主要市场被华为云、运营商云占据。
 - 3) 确定“AI 推动、公有云优先”战略:** 23 年底确定“AI 推动、公有云优先”战略, 从“公有云优先”方面来看, 公司成立公有云业务事业部, 以规模优先、扩大市场占有率为目标, 同时, 兼顾政企市场, 成立混合云业务事业部; 从“AI 推动”方面来看, 公司要做“AI 时代最开放的云”, 其多款自研模型开源, 在升级 AI 基础设施、产品结构、AI 开源社区均贯彻“开放”理念, 其中 AI 开源社区魔搭聚集了

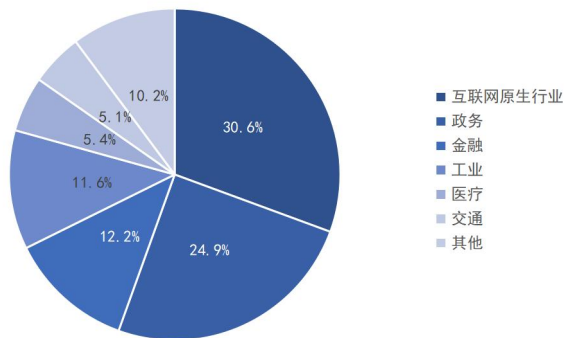
280 万开发者、2300+个模型；同时，发力智算中心建设，根据 IDC 披露数据，23 年 AI 公有云服务市场规模达 126.1 亿元，同比+58.2%，其中，阿里云、百度智能云并列第一，市占率均为 26.4%，其次为腾讯云、华为云，市占率分别为 20.7%、17.8%。

图51: 垂直行业云计算市场成为增长主要动力



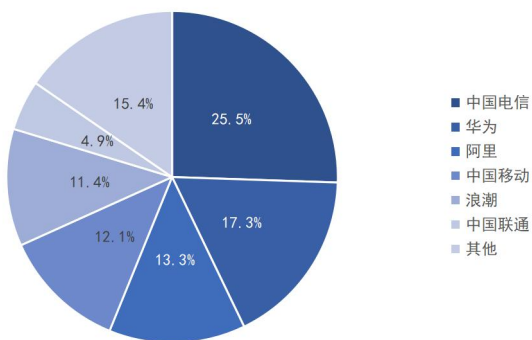
资料来源: 前瞻产业研究院, 国信证券经济研究所整理

图52: 2022 年中国各行业用量占比情况



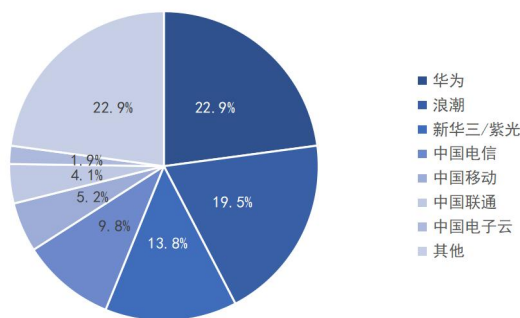
资料来源: 中商产业研究院, 国信证券经济研究所整理

图53: 中国政务公有云市场份额 (2023 年)



资料来源: IDC, 国信证券经济研究所整理

图54: 中国政务专属云市场份额 (2023 年)



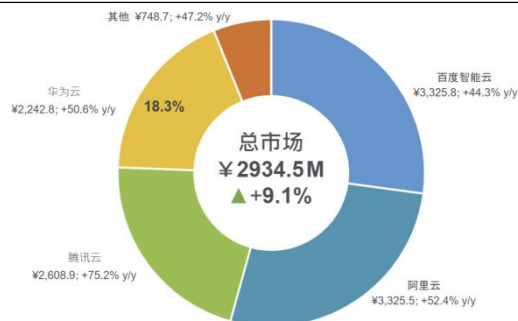
资料来源: IDC, 国信证券经济研究所整理

图55: 阿里云魔搭社区-中国最大的模型社区



资料来源: 魔搭社区, 国信证券经济研究所整理

图56: 中国 AI 公有云服务市场 2023 年市场份额



注: 2023 年厂商份额 (%), 收入 (百万元人民币), 总体市场 (亿元人民币), 增长率 (%)

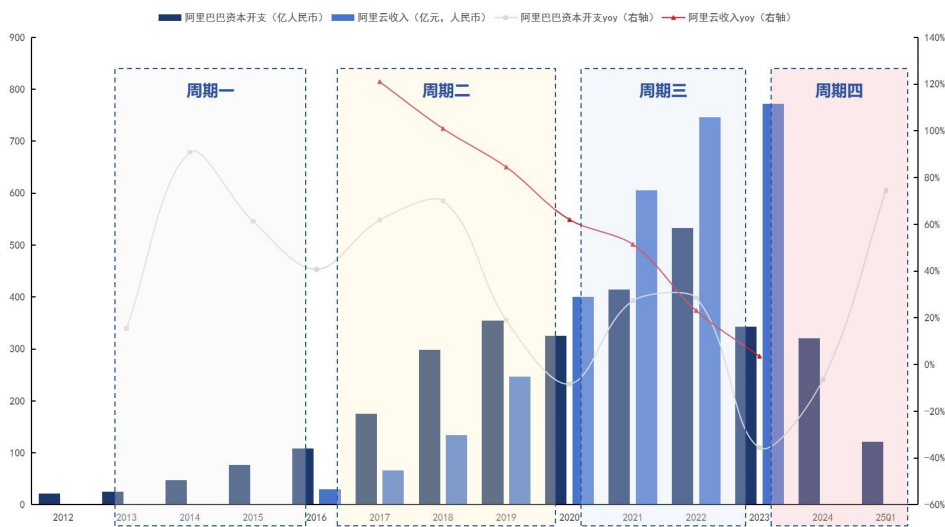
资料来源: IDC, 国信证券经济研究所整理

硬件维度：自研 GPU+AI 芯片，双轮驱动成长。

- **CPU 处理器：**1) **Intel：**阿里云成立之初便采用 Intel CPU，且一直为阿里云主流处理器。2) **AMD：**20 年，阿里云推出基于 AMD EPYC™ ROME 处理器的 c6a 实例，依托神龙架构，将大量虚拟化功能卸载到专用硬件，降低虚拟化开销，提供稳定可预期的高性能。3) **自研芯片：**21 年，阿里云在云栖大会上推出首款自研云原生处理器——倚天 710，基于 ARM 架构，拥有 128 个核心、3.2GHz 主频、DDR5 8 通道、96 个 PCIe 5.0，根据阿里云披露数据，在核心应用场景，相较于普通实例，性价比提升 30%，单位算力功耗降低 60%；阿里云在 22 年云栖大会上宣布倚天 710 已经大规模应用，未来 2 年阿里云 20% 的新增算力将使用自研芯片。
- **GPU 处理器（AI 算力卡）：**1) **英伟达：**20 年阿里云推出基于英伟达 A100 GPU 的 gn7 系列服务器，亦是国内首款应用 A100 GPU 商业化云服务器。2) **自研芯片：**18 年阿里收购中天微，并将达摩院芯片研发团队与中天微团队合并，成立平头哥半导体；19 年阿里云首款自研服务器 AI 芯片含光 800 发布，目前已在阿里云内部规模使用。

阿里巴巴资本开支与阿里云收入增长复盘：资本开支增速呈现周期性波动，发力 AI 数据中心建设。从阿里巴巴历史数据来看，公司资本开支增速呈现周期性波动，平均 3-4 年为一个周期，阿里 FY24 财年进入新一轮资本开始上升周期。FY25Q1 阿里巴巴资本开支为 120.94 亿人民币，同比+75%，大力建设智算中心，资本开支有望持续增长。从阿里云收入端来看，阿里云收入增速受资本开支影响较小，主要受垂直行业领域竞争影响。

图57：阿里巴巴资本开支与阿里云收入历史变化情况



资料来源：公司财报，国信证券经济研究所整理

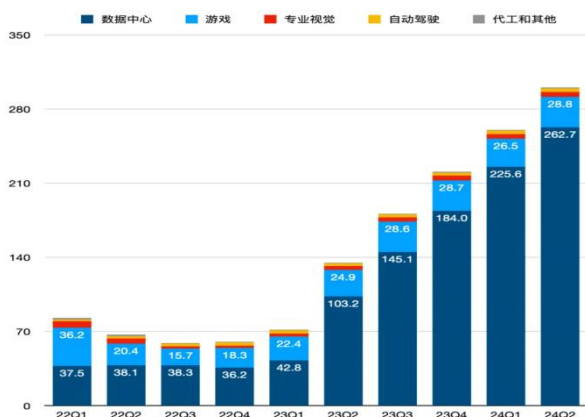
支出侧：从 ROIC 维度观察云厂商资本开支变动

云厂商资本开支直接影响硬件侧公司收入

英伟达等硬件厂商营收直接受到云厂商资本开支影响。以英伟达为例，24Q2 英伟达的数据中心业务营收 262.7 亿美元（同比+154.5%），营收占比 87.5%；根据英伟达财报会议披露数据，云计算厂商（如微软、亚马逊、谷歌等）约占数据中心收入 45%，互联网、消费级科技公司（如 Meta、特斯拉等）占数据中心收入超过 50%，云厂商资本开支变动将直接影响硬件侧公司营收。

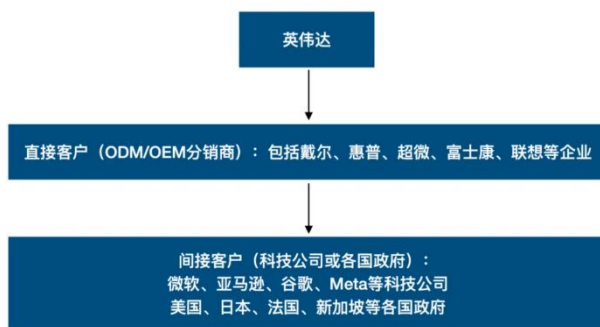
云厂商为英伟达“间接客户”。1) **直接客户**：主要包括戴尔、惠普、超微、富士康、联想等渠道商（或代工商）。2) **间接客户**：渠道商（或代工商）向大型科技公司（如微软、亚马逊、谷歌、Meta 等）、各国政府机构（如美国、日本、法国、新加坡等）销售搭载 AI 芯片的服务器，科技公司、各国政府为英伟达“间接客户”。“间接客户”的算力需求拉动“直接客户”向英伟达采购 AI 芯片，即“间接客户”的需求是英伟达业绩增长的根本驱动力。

图58: 英伟达的收入结构（亿美元）



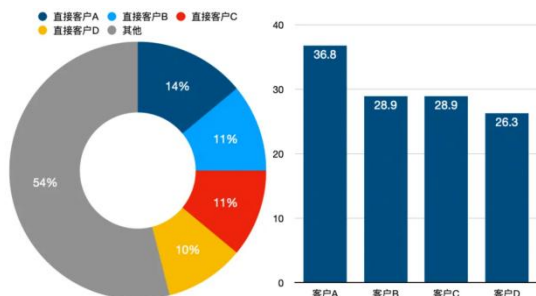
资料来源：公司财报、国信证券经济研究所整理

图59: 英伟达芯片的客户销售情况



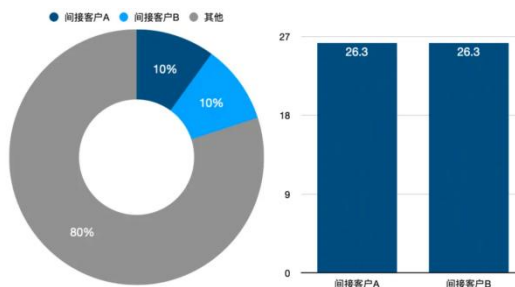
资料来源：公司财报、国信证券经济研究所整理

图60: 24Q2 英伟达头部“直接”客户对数据中心收入贡献



资料来源：公司财报、国信证券经济研究所整理

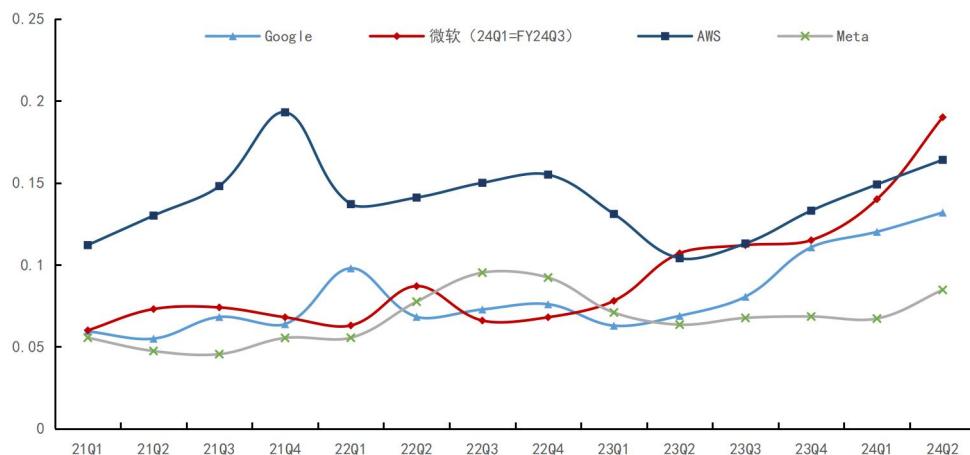
图61: 24Q2 英伟达头部“间接”客户对数据中心收入贡献



资料来源：Gartner、国信证券经济研究所整理

24 年北美云厂商持续加大资本开支投入，指引乐观。根据财报披露数据，24Q2 微软、谷歌、亚马逊、Meta 四大云巨头资本开支显著增长，合计 571 亿美元，其中微软资本开支 190 亿美元，同比+78%、环比+36%；谷歌资本开支 132 亿美元，同比+91%、环比+10%；亚马逊资本开支 164 亿美元，同比+58%、环比+10%；Meta 资本开支 85 亿美元，同比+33%、环比+26%。云厂商对全年资本开支展望乐观，通过加大资本开支进一步扩张算力、迭代技术，加快 AI 基础设施建设和技术研发进度。

图62: 全球各大厂资本开支（亿美元，自然年指标）



资料来源：各公司财报，国信证券经济研究所整理

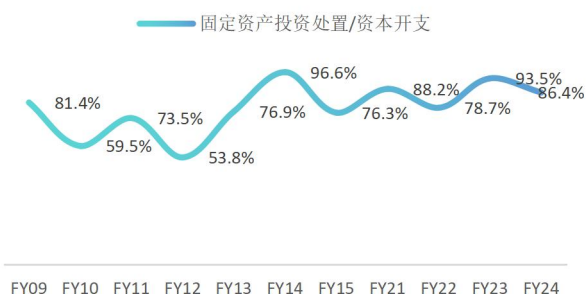
微软资本开支与折旧 ROIC 变化

➤ **微软的 ROIC 变动规律：短期稳定，长期下滑**

为了提升模型预测的准确性，减少基于假设的误差，我们采用以下计算方式：

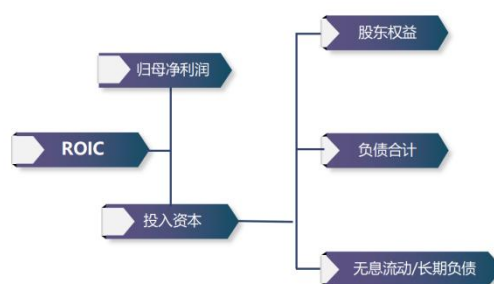
ROIC 计算=归属于母公司股东净利润*2/(期初投入资本+期末投入资本)；投入资本=股东权益（不含少数股东权益）+负债合计-无息流动负债-无息长期负债。

图63: 固定资产投资设置在总资本开支中的占比



资料来源：微软财报、国信证券经济研究所整理

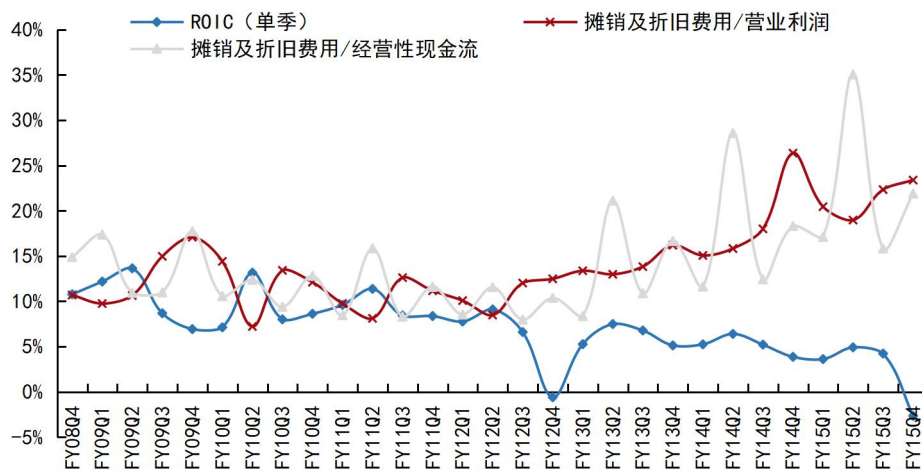
图64: 调整后的 ROIC 计算



资料来源：微软财报、国信证券经济研究所整理

ROIC 历史数据复盘：短期波动，长期下滑。2013 年微软 IaaS 产品（Windows Azure IaaS）正式发布，开始大力投入算力基础设施建设，折旧及摊销费用持续上升，ROIC 承压，长期呈下滑趋势。

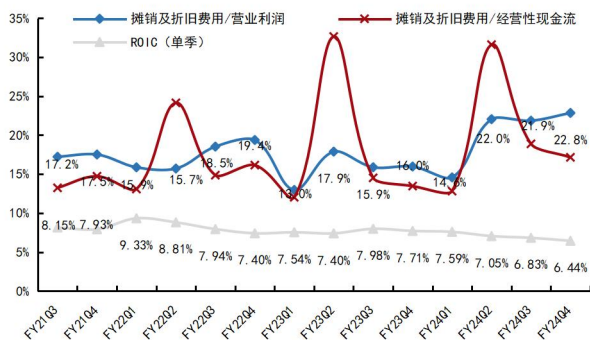
图65: ROIC（单季度）历史数据，长期呈现下滑趋势



资料来源：微软财报，国信证券经济研究所整理

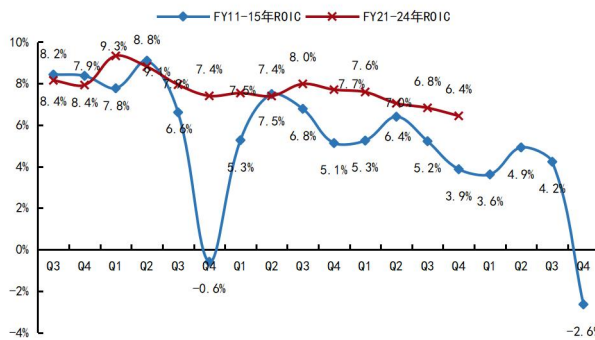
ROIC 周期趋势：因资本开支加大而持续下滑。微软加大资本投入可能导致 25 财年 ROIC 下降，增加折旧及摊销对净利润的负面影响，目前 ROIC 已呈现平稳下降态势。从历史数据来看，在资本开支大幅增加的周期中，ROIC 往往会进一步降低，预计 2025 财年微软的 ROIC 可能继续下降，折旧及摊销费用对净利润的影响将逐步增大。

图66: 随着折旧及摊销费用对利润影响加剧，ROIC 明显下滑



资料来源：微软财报、国信证券经济研究所整理

图67: 上一个周期与本周起 ROIC 下降趋势基本吻合



资料来源：微软财报、国信证券经济研究所整理

➢ **折旧对业绩的影响：对现金流和利润的影响具有滞后性**

为了提升模型预测的准确性，减少基于假设的误差，我们基于历史数据做了以下假设：

假设一（当期折旧费用的计算）：依据公司披露的现金流量表中的折旧及摊销费

用，并从中扣除无形资产摊销费用。

传统计算当期折旧费用的方法误差较大。传统计算当期折旧费用的方法为固定资产总值除以折旧年限。存在以下问题：1) 不同类型的固定资产（如建筑物、租赁改良、家具及设备、网络及设备）有不同的折旧年限，这对计算每期的折旧费用造成困难；2) 不是所有的资本开支都用于服务器和网络设备，且公司不经常披露详细的资产分类数据；3) 微软在 2023 财年将服务器和网络设备的折旧年限从 4 年延长至 6 年，这导致模型需要大幅调整当期的折旧费用。新的假设可以增加模型与实际情况的拟合度因为折旧及摊销费用、摊销费用完全是公司披露的历史数据，所以当期计算的折旧费用也比较准确，预期跟公司自身记录的详细折旧费用没有太大区别。

假设二：折旧摊销费用与公司资本开支的季度环比增速一致。通过分析历史数据，我们发现资本开支的增长趋势与折旧及摊销费用的增长趋势大致一致。资本开支的增加通常意味着对固定资产和无形资产的更多投资，这自然会导致折旧和摊销费用的增加。基于这一逻辑，我们可以利用公司提供的资本开支指引来预测 2025 财年的折旧和摊销费用。

假设三：资本开支扣除费用后 80%资本化为固定资产。基于历史数据分析，除了个别年份，大部分年份资本开支的 70%~90%转化为固定资产投资或处置，因此在做模型预测时假设这部分在资本开支中占比 8 成。

假设四：公司无形资产基本维持稳定。

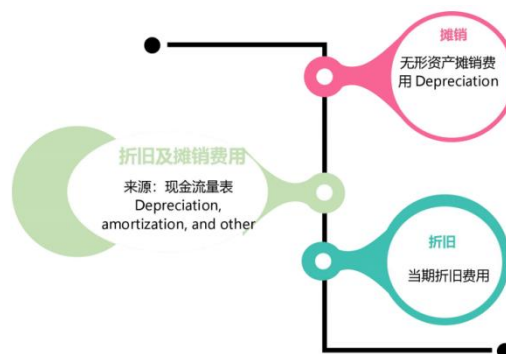
假设五：公司预计由于云和人工智能基础设施的投资，资本支出将在季度基础上大幅增加。 FY25 的资本支出将大于 FY24。

图 68: 不同类型的固定资产折旧年限不同

固定资产类型	折旧年限
内部使用的计算机软件	3-5年
计算机设备	2-6年
建筑物及改良	5-10年
租赁改良	3-20年
家具和设备	1-10年
土地	不进行折旧

资料来源：微软财报、国信证券经济研究所整理

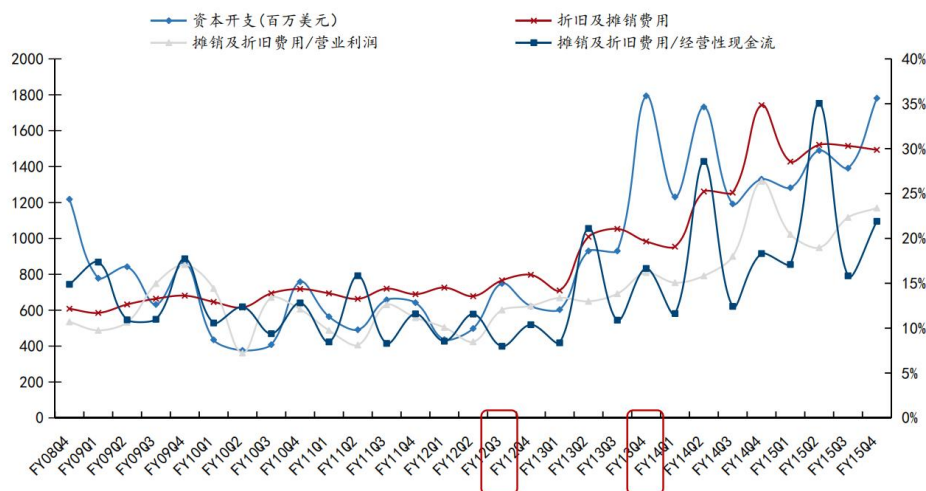
图 69: 基于公司披露的数据计算当期折旧费用



资料来源：微软财报、国信证券经济研究所整理

资本开支上行会提升技术和产能，亦带来短期财务压力，主要体现在折旧对业绩的影响上。资本开支增加后大约五个季度对利润产生影响，三个季度后影响现金流。尽管资本开支的增长与折旧及摊销费用的增长大致相同，但折旧计提具有时间上的滞后性，主要系根据在建工程的会计处理准则，即服务器已经购置，但数据中心尚未建成，这些服务器将继续计入在建工程；由于数据中心建设需要时间（12-16 个月），从资本开支增加到折旧费用开始显著上升，存在一个时间滞后。

图70: 13年微软 IaaS 产品发布后, 资本开支攀升, 折旧及摊销费用对利润影响加剧



资料来源: 微软财报, 国信证券经济研究所整理

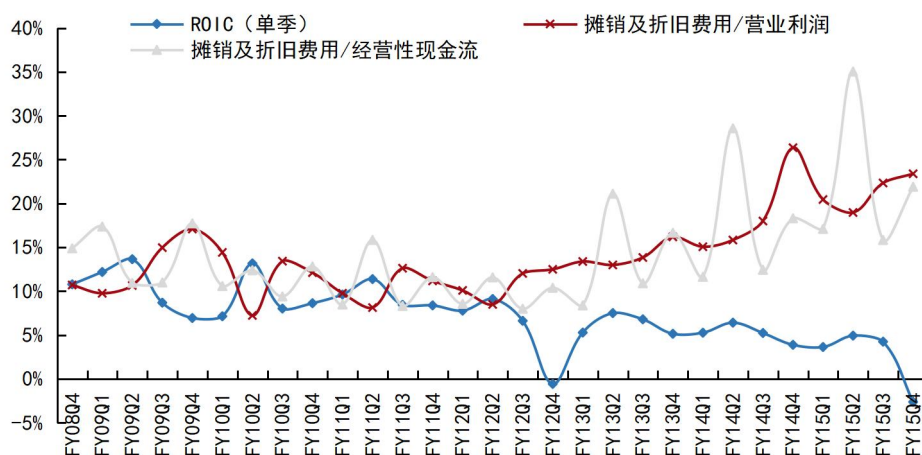
微软资本开支及其对业绩的影响复盘如下:

1) 资本开支对利润的影响: 微软在 FY12Q3 首次大幅增加资本开支, 之后经历两个季度的下降又再次上升。折旧及摊销费用在营业利润中的占比直到 FY13Q4 才显著上升, 显示出大约五个季度的滞后效应。由于净利润计算时要减去折旧及摊销费用, 该费用占比上升会导致净利润下降。

2) 资本开支对现金流的影响: 资本开支对现金流的影响更快, 折旧及摊销费用在现金流中占比大幅上行出现在 2012 年 9 月, 中间间隔三个季度。

此外, 折旧摊销对 ROIC 也有影响, 稳中下滑。从历史数据来看, ROIC (单季度) 与折旧及摊销费用占营业利润比值的走势成反比。近期微软再次加大了资本投入, 目前已呈现出平稳的逐渐下降的趋势, 后续可能会进一步下滑。

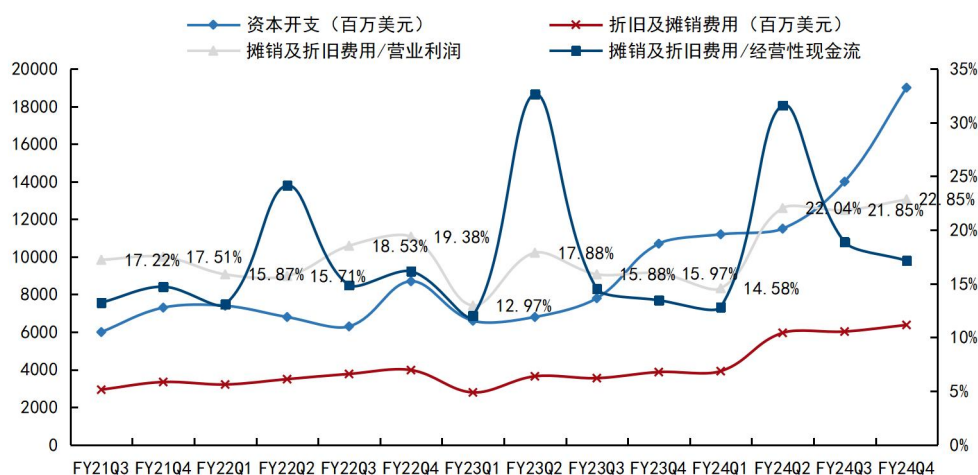
图71: ROIC (单季度) 历史数据, 长期呈现下滑趋势



资料来源: 微软财报, 国信证券经济研究所整理

资本开支持续上行，将对折旧摊销、利润、现金流、ROIC 等财务指标产生影响。根据历史数据，折旧及摊销费用占到营业利润 25%左右可以作为一个转折点，此时费用对利润产生很大压力。在转折点之后资本开支预期还会大幅上涨 2-3 个季度，后续会同比下行。因此可预测这次周期的转折点可能会出现在 FY25Q2（即 2024 年 Q4）。再过两个季度（FY25Q4），资本开支可能会达到稳定期，后续可能同比下降。

图72: FY21-24 年资本开支及折旧与摊销费用等

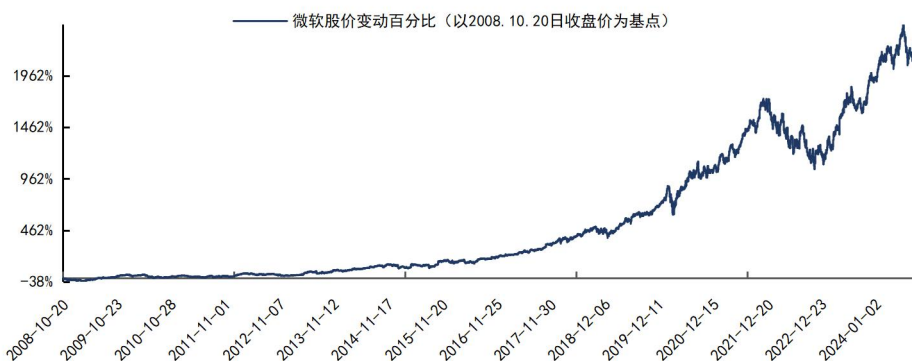


资料来源：微软财报，国信证券经济研究所整理

➤ 微软股价表现复盘：云计算打开第二成长曲线，拉动公司股价上涨

云业务打开第二增长曲线，拉动公司股价上涨。2019Q2，微软智能云部门收入达 114 亿美元，首次超过办公软件部门收入，FY24 微软云计算部门收入达 1053.6 亿美元，同比增长 19.9%，对应 14-24 年 CAGR 为 17.1%。云业务打开公司第二增长曲线，拉动公司整体业绩持续增长，公司股价不断上涨。

图73: 微软 2008 年至今股价涨幅（以 2008. 10. 20 为基点）



资料来源：微软财报，国信证券经济研究所整理

业绩向好可消除资本开支带来的负面反馈。微软在 2016–2019 年周期也面临资本开支的上行,但是营收和利润的积极情况打消了市场的担忧。根据 FY19Q2 的财报,微软的营收和利润增长主要得益于 Office 365 和 Azure 的稳健财务业绩,以及 Surface 和 Xbox 游戏业务的出色表现。

图74: 微软 FY16–20 资本开支 (不含融资租赁)



资料来源: 微软财报、国信证券经济研究所整理

图75: 微软 FY16–20 单季度营收



资料来源: 微软财报、国信证券经济研究所整理

谷歌云资本开支与折旧 R01C 变化

回顾 2021–2024 年期间,谷歌资本上行趋势明显。21Q1–23Q1 期间,谷歌资本开支规模基本稳定,22Q1 谷歌资本开支大幅波动,同比环比大幅上行,并在后续季度回落。23Q2 后,受全球 AI 计算需求推动,谷歌资本开支开始呈现明显上行趋势,并在 24Q2 增长至近三年峰值 131.86 亿美元,同比、环比增速趋于平稳。公司预计全年季度资本支出大于或等于 120 亿美元,后续季度资本支出或将维持高位。

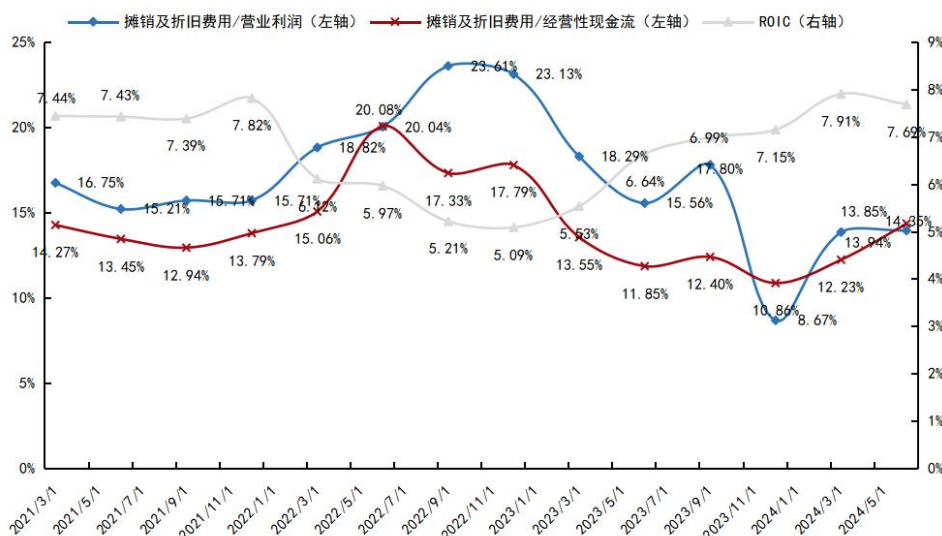
图76: 谷歌近期资本开支变化



资料来源: 公司财报, 国信证券经济研究所整理

谷歌折旧及摊销费用对利润、现金流影响上行，ROIC 或维持下滑趋势。谷歌折旧及摊销费用占经营利润、经营性现金流比重呈现明显同向变动趋势，21Q1-22Q2 期间，折旧及摊销费用对利润、现金流影响程度均呈现明显上行趋势，并在 22Q3-Q4 维持高位后逐步下行。公司折旧及摊销费用对利润、现金流影响在 22Q3-Q4 呈现明显波峰，并在后续下降，于 23Q4 呈现波谷，公司 ROIC 则在 22Q3-Q4 呈现明显波谷，并在后续回升，与两者呈明显负相关关系。从变动趋势来看，公司折旧及摊销费用对利润、现金流影响上行时，公司 ROIC 将呈现明显下降趋势，24 年上半年，公司折旧及摊销费用对利润、现金流影响明显上行，公司 ROIC 在 24Q2 显著下降，后续季度资本开支维持高位，折旧及摊销费用对利润、现金流影响或将保持上行，公司 ROIC 将进一步下滑。

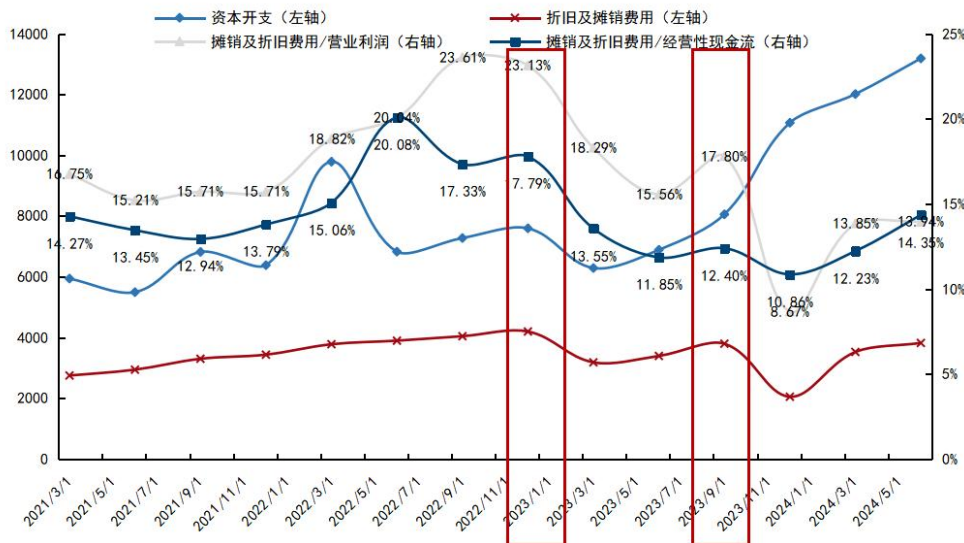
图 77: 谷歌 ROIC 数据复盘



资料来源：公司财报，国信证券经济研究所整理

谷歌资本开支上行将在 3 个季度后显著影响利润及现金流。22Q1、22Q4 谷歌资本开支呈现波峰，公司折旧及摊销费用对利润、现金流影响分别在 22Q4 及 23Q3 呈现出波峰，期间变动趋势与资本开支变动趋势基本一致。综合历史情况，公司折旧及摊销费用对营业利润、经营性现金流产生影响与资本开支变动相比有明显的滞后效应，一般在资本开支上行 3 个季度后，公司的利润及现金流会同时受到折旧及摊销费用的显著影响。公司净利润计算时需扣除折旧及摊销费用，该费用占比上升会导致净利润下行，因此资本开支继续上行或将导致公司后期业绩承压。23Q1 开始公司资本开支显著上行，公司折旧及摊销费用对利润、现金流影响在 23Q4 开始稳定上行，进入 24 年后公司资本开支进一步提高，预计 24-25 年期间公司净利润将受到折旧及摊销费用的显著影响。

图78: 谷歌资本开支变化影响

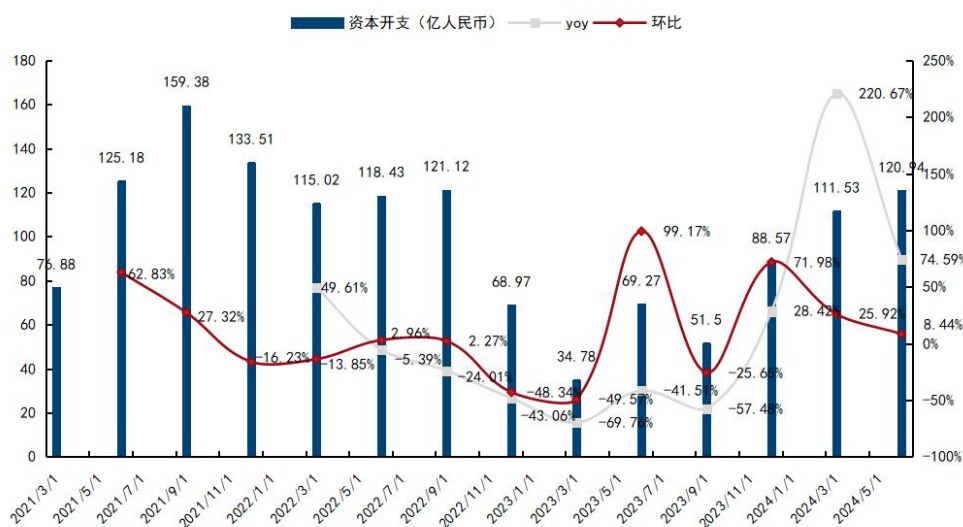


资料来源: 公司财报, 国信证券经济研究所整理

阿里云资本开支与折旧 ROIC 变化

回顾 2021-2024 年期间, 阿里巴巴资本开支大幅波动。自然年 21Q1 (FY21Q4), 阿里巴巴资本开支环比大幅上行, 并在 21Q2 达到峰值, 后续公司资本开支维持高位, 在 22Q4 开始环比下滑。2023 年后, 阿里巴巴资本开支再次环比上行, 并在 24Q2 增长至 120.94 亿人民币。

图79: 阿里巴巴近期资本开支变化

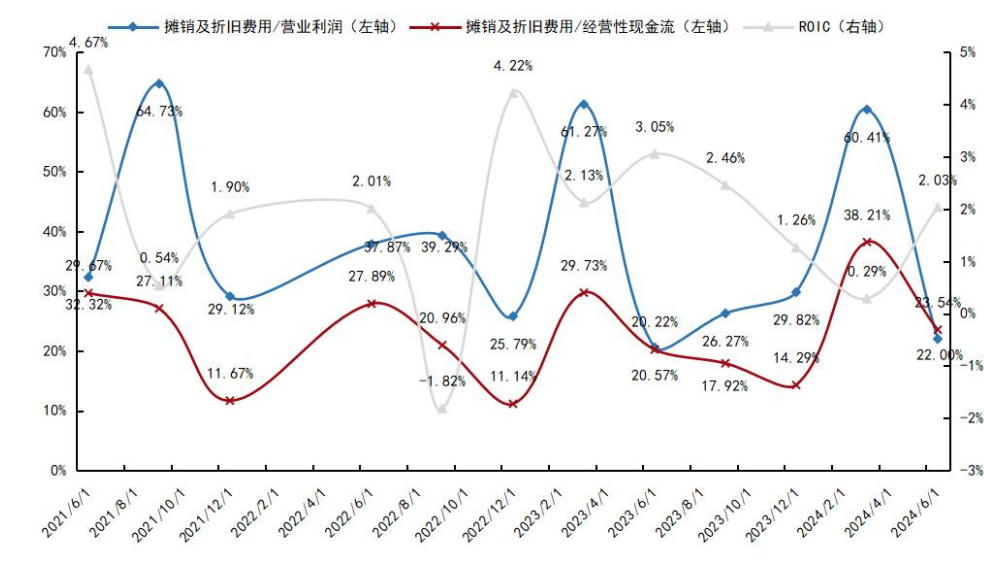


资料来源: 公司财报, 国信证券经济研究所整理

折旧摊销费用对利润影响呈度与 ROIC 显著负相关。21Q1 公司利润受罚款显著影响, ROIC 呈现异常波动。剔除受罚款影响的异常季度后, 公司折旧及摊销费用占

营业利润、经营性现金流比重呈明显同向变动，23Q1 及 24Q1 均呈现明显波峰，公司 ROIC 与两者呈现显著负相关关系，在 23Q1 及 24Q1 呈现明显波谷。从趋势来看，公司折旧及摊销费用对利润影响上行时，公司 ROIC 呈现出明显下行趋势。

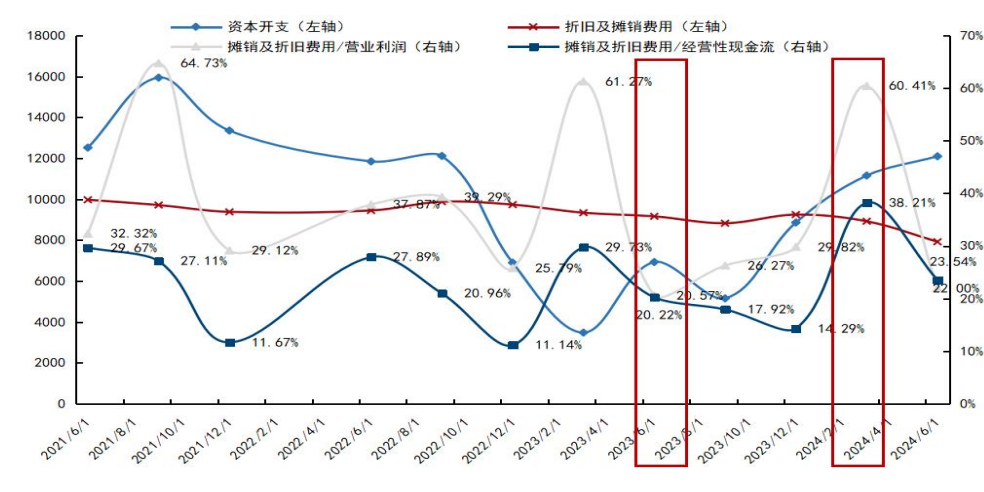
图80: 阿里巴巴 ROIC 数据复盘



资料来源: 公司财报, 国信证券经济研究所整理

阿里巴巴资本开支上行将在 2-3 个季度后显著影响利润及现金流。23Q2 公司资本开支显著上行，并在后续季度环比下降，公司折旧及摊销费用占营业利润、经营性现金流比重在 24Q1 呈现明显上升，并在后续季度显著下行，与资本开支呈现相同变化趋势。综合历史情况，公司折旧及摊销费用对营业利润、经营性现金流产生影响与资本开支波动相比有明显的滞后效应，一般在资本开支上行 2-3 个季度后，公司的利润及现金流会同时期受到折旧及摊销费用的显著影响。公司净利润计算时需扣除折旧及摊销费用，该费用占比上升会导致净利润下行，因此资本开支继续上行或将导致公司后期业绩承压。

图81: 阿里巴巴资本开支变化影响



资料来源: 公司财报, 国信证券经济研究所整理

回报侧：云厂商 AI 云租赁回报率测算

大型云厂商：H100 算力卡云租赁回报率测算（以微软为例）

假设一：资本开支中 38%用于计算设备及软件。云厂商资本开支主要用于土地、建筑物、计算设备及软件等的购置，根据微软财报披露数据，每年用于购置计算机设备及软件的费用在资本开支中占比基本在 32%-45%间波动，即我们假设资本开支中 38%用于计算设备及软件的购置。同时，我们对建设 AI 算力相关成本占计算设备及软件的开支比例进行敏感性分析，假设 H100 算力相关建设成本占计算设备及软件的开支比例分别为 40%、50%、60%，同时根据 Omdia Research 披露数据，23 年微软预计购买 15 万张 H100（自然年），则单张 H100 对应综合建设成本分别为 4.17、5.22、6.26 万美元。

表1: 假设资本开支中 38%用于计算设备及软件（元）

财年	2020	2021	2022	2023	2024	
土地	1823	3660	4734	5683	8163	
建筑物	33995	43928	55014	68465	93943	
租赁改良	5487	6884	7819	8537	9594	
计算设备及软件	41261	51250	60631	74961	93780	
家具及设备	4782	5344	5860	6246	6532	
总计	87348	111066	134058	163892	212012	
累计折旧	43197	51351	59660	68251	76421	
净值	44151	59715	74398	95641	135591	
计算机设备及软件增量		9989	9381	14330	18819	
资本开支		242	292	319	557	平均
计算机设备及软件增量/资本开支		41%	32%	45%	34%	38%

资料来源：微软财报，国信证券经济研究所整理

表2: H100 算力相关建设成本敏感性测算

	23Q1	23Q2	23Q3	23Q4
季度资本开支（自然年，亿美金）	78	107	112	115
2023 年资本开支（自然年，亿美金）		412		
计算机设备及软件开支/资本开支		38%		
计算机设备及软件开支		156.56		
2023 年 H100 采购数量（自然年，颗）		150,000		
假设 H100 算力相关建设成本占计算设备及软件开支的比例				单颗 H100 对应综合建设成本（万美元）
40%				4.17
50%				5.22
60%				6.26

资料来源：微软财报，国信证券经济研究所整理

假设二：微软单 GPU 租赁收入为 7.47 美元/小时。根据微软 Azure 云官网统计数据，选取 3 个微软 H100 租赁实例，根据 24 年 8 月 28 日租赁价格，可计算对应单颗 GPU 的租赁价格分别为 6.98、7.24、8.19 美元/小时，取平均 7.47 美元/小时作为微软单颗 GPU 租赁收入。

表3: 假设微软单 GPU 租赁收入为 7.47 美元/小时

微软 H100 实例名称	GPU 数量	微软 H100 实例租赁价格 (美元/小时)								单 GPU 租赁收入 (美元/小时)	
		2024. 5. 28	2024. 6. 14	2024. 6. 28	2024. 7. 14	2024. 7. 28	2024. 8. 14	2024. 8. 28	对应单 GPU		
Standard_NC40ads_H100_v5	1	6.98	6.98	6.98	6.98	6.98	6.98	6.98	6.98	6.98	7.47
Standard_NC80adis_H100_v5	2	17.64	13.96	13.96	13.96	13.96	13.96	13.96	13.96	7.24	
Standard_ND96isr_H100_v5	12	98.32	98.32	98.32	98.32	98.32	98.32	98.32	98.32	8.19	

资料来源: 微软 Azure 云官网, 国信证券经济研究所预测

单张 H100 全生命周期累计收入计算: 1) 生命周期: 根据微软服务器及网络设备的折旧年限, 假设单张 H100 全生命周期为 6 年; 2) 租赁价格: 根据现有统计的租赁价格趋势, 预计单张 H100 租赁价格呈现逐步下降趋势; 3) 租赁利用率: 假设微软的租赁利用率为 70%。根据上述假设条件, 可以得到单张 H100 卡 24、25、26、27、28、29 年累计收入贡献分别为 4.58、8.25、11.18、13.53、15.40、16.90 万美金。

表4: 微软 H100 全生命周期累计收入测算

	2024	2025	2026	2027	2028	2029
H100 租赁价格 (美元/小时)	7.47	5.98	4.78	3.83	3.06	2.45
租赁利用率	70%	70%	70%	70%	70%	70%
H100 全生命周期累计收入 (万美金)	4.58	8.25	11.18	13.53	15.40	16.90

资料来源: 微软 Azure 云官网, 国信证券经济研究所预测

电力成本计算: 1) **单 POD IT 耗能测算:** 根据英伟达披露数据, 127 节点 POD 功率合计为 1411.176kw, 其中计算部分占比最高 (为 89.83%, 即 DGX H100 服务器), 其次为计算用交换机 (为 5.72%); 2) **单台服务器对应 IT 设备功率测算:** 以英伟达 DGX H100 服务器为例, 单台服务器功率为 10,200w; 如前文所述, 已知 127 节点 SuperPOD 中服务器部分 (存储、网络、管理) 和交换机部分 (计算、存储、In-Band 管理、Out-of-Band 管理) 功率, 则对应单台服务器为 911.62w, 加总可知单台服务器对应 IT 设备功率要求为 11,112w; 3) **单台服务器对应智算中心用电量:** 假设利用率为 70%, PUE 为 1.25, 可知单台服务器对应智算中心功率消耗为 11,112w, 对应年度用电量为 85,171KWh; 4) **单卡 H100 对应智算中心用电量:** 单台 H100 服务器对应 8 张 H100 卡, 单张 H100 对应智算中心用电量为 10,646KWh; 5) **用电成本测算:** 微软在全球建设智算中心, 根据欧盟统计局数据, 23 年欧盟居民用电在 0.2-0.35 欧元/kWh, 工业用电价格通常低于居民用电, 根据美国能源信息署 (EIA) 数据, 23 年美国居民用电平均价格为 0.14-0.15 美元/kWh, 平均工业电价约 0.07-0.08 美元/kWh, 此处我们假设平均工业用电电价为 0.1 美元/kWh, 则单张 H100 年度电力相关费用为 0.04 万美金, 单张 H100 对应智算中心年度电力相关费用为 0.11 万美金。

表5: 单 POD IT 耗能测算

对应产品	数量	服务器				交换机			
		计算 (Compute)	存储 (Storage)	管理 (Mgmt)	网络 (Fabric)	计算 (Compute)	存储 (Storage)	In-Band 管理	Out-of-Band 管理
		DGX H100	多种	PowerEdge R750	NVIDIA UFM 3.1	QM9700	QM9700	SN4600C	SN2201
平均功率 (Watts)	单一	10,200	2880	704	600	1,376	1,376	466	98
	加总	1,295,400	17,280	3,520	2,400	66,048	22,016	3,728	784
峰值功率 (Watts)	单一	10,200	3,600	880	750	1,720	1,720	820	135
	加总	1,295,400	21,600	4,400	3,000	82,560	27,520	6,560	1080
峰值热负荷 (BTU/h)	单一	34,804	12,284	3,003	2,559	5,869	5,869	2,798	461
	加总	4,420,088	73,702	15,013	10,236	281,706	93,902	22,384	3,685
系统占比		89.83%	1.50%	0.31%	0.21%	5.72%	1.91%	0.45%	0.07%

资料来源: 《NVIDIA DGX SuperPOD Data Center Design》, 国信证券经济研究所预测

表6: 单张 H100 对应智算中心电力消耗测算

单张 H100 对应智算中心电力消耗测算		功率 (W)	
DGX H100 服务器		10,200	
	AI 算力-GPU	5600	
	H100 SXM	700	
	数量 (颗)	8	
其他-CPU、网络模组、PCIe Switch、NVSwitch 等		4600	
127 节点服务器-存储部分		17,280	
	对应单节点服务器存储	136	
127 节点服务器-管理部分		3,520	
	对应单节点服务器管理	28	
127 节点服务器-网络部分		2,400	
	对应单节点服务器网络	19	
127 节点对应交换机-计算部分		66,048	
	对应单节点服务器对应交换机-计算部分	520	
127 节点对应交换机-存储部分		22,016	
	对应单节点服务器对应交换机-存储部分	173	
127 节点对应交换机-In-Band 管理部分		3,728	
	对应单节点服务器对应交换机-In-Band 管理部分	29	
127 节点对应交换机-Out-of-Band 管理部分		784	
	对应单节点服务器对应交换机-Out-of-Band 管理部分	6	
单台服务器对应 IT 设备功率要求		11,112	(加总)
	利用率	70%	(假设)
单台服务器对应 IT 设备功率消耗		7,778	
PUE (=数据中心总能耗/IT 设备能耗)		1.25	(假设)
单台服务器对应数据中心功率消耗		9,723	
单台服务器对应数据中心年度用电量		85,171	(KWh)
单张 H100 对应智算中心电力消耗测算		10,646	(KWh)

资料来源: 微软 Azure 云官网, 国信证券经济研究所预测

表7: H100 用电成本测算

	2024	2025	2026	2027	2028	2029
单张 H100 年度耗电量 (kWh)	4,292	8,585	12,877	17,170	21,462	25,754
电价 (美元/h)	0.1	0.1	0.1	0.1	0.1	0.1
累计用电成本 (万美元)	0.04	0.09	0.13	0.17	0.21	0.26
单张 H100 对应智算中心耗电量(kWh)	10,646	21,293	31,939	42,585	53,232	63,878
电价 (美元/h)	0.1	0.1	0.1	0.1	0.1	0.1
累计用电成本 (万美元)	0.11	0.21	0.32	0.43	0.53	0.64

资料来源: 英伟达, 国信证券经济研究所预测

回报率测算: 1) 单张 H100 回报率测算: 假设单张 H100 采购价格为 3 万美金, 仅考虑单卡采购成本和单卡电费, 24-29 年回报率分别为 51%、167%、257%、326%、379%、419%; **2) 单张 H100 对应智算中心综合成本回报率测算:** 在中性预期下 (即假设 H100 算力相关建设成本占计算设备即软件开支的 50%), 24-29 年回报率分别为-14%、52%、102%、140%、168%、189%。

表8: 微软 H100 回报率测算

单张 H100 回报率测算 (仅考虑单卡购买成本及电费)						
	2024	2025	2026	2027	2028	2029
单卡采购成本 (万美元)	3					
累计用电成本 (万美元)	0.04	0.09	0.13	0.17	0.21	0.26
成本 (万美元)	3.04	3.09	3.13	3.17	3.21	3.26
投资回报率 (%)	51%	167%	257%	326%	379%	419%
单张 H100 综合成本回报率测算 (考虑单卡购买成本、其他配套设备购买成本、智算中心运营电费)						
假设 H100 算力相关建设成本占计算设备及软件开支的比例为 40%	单颗 H100 对应综合建设成本 (万元)					4.17
累计用电成本 (万美元)	0.11	0.21	0.32	0.43	0.53	0.64
成本加总	4.28	4.39	4.49	4.60	4.71	4.81
投资回报率	7%	88%	149%	194%	227%	251%
假设 H100 算力相关建设成本占计算设备及软件开支的比例为 50%	单颗 H100 对应综合建设成本 (万元)					5.22
累计用电成本 (万美元)	0.11	0.21	0.32	0.43	0.53	0.64
成本加总	5.33	5.43	5.54	5.64	5.75	5.86
投资回报率	-14%	52%	102%	140%	168%	189%
假设 H100 算力相关建设成本占计算设备及软件开支的比例为 60%	单颗 H100 对应综合建设成本 (万元)					6.26
累计用电成本 (万美元)	0.11	0.21	0.32	0.43	0.53	0.64
成本加总	6.37	6.48	6.58	6.69	6.79	6.90
投资回报率	-28%	27%	70%	102%	127%	145%

资料来源: 英伟达, 国信证券经济研究所测算

中小云厂商: H100 算力卡云租赁回报率测算

单张 H100 全生命周期累计收入计算: 1) 生命周期: 维持上文假设, 假设单张 H100 全生命周期为 6 年; 2) 租赁价格: 根据各中小云厂商租赁价格统计结果, 中小云厂商租赁价格明显低于大厂, 假设 24 年单颗 GPU 租赁价格为 2.49 美元/小时, 同时假设未来租赁价格呈现下降趋势; 3) 租赁利用率: 假设微软的租赁利用率为 70%。根据上述假设条件, 可以得到单张 H100 卡 24、25、26、27、28、29 年累计收入贡献分别为 1.53、2.98、4.36、5.66、6.91、8.09 万美金。

表9: 假设中小云厂商单 GPU 租赁收入为 2.49 美元/小时

中小云厂商实例租赁价格 (美元/小时)								单 GPU 租赁收入 (美元/小时)
日期	2024.7.28	2024.8.14	2024.8.28	2024.9.14	2024.9.23	2024.9.28	2024.10.14	
Lambda	2.49	2.49	2.49	2.49	2.49	2.49	2.49	2.49
FluidStack	3.75	2.89	2.89	2.89	2.89	2.89	2.89	
RunPod	3.29	3.29	3.29	2.89	2.89	2.89	2.69	
Hyperstack	3.44	3.44	3.44	1.9	1.9	1.9	1.9	
平均价格	3.24	3.03	3.03	2.54	2.54	2.54	2.49	

资料来源: Lambda、FluidStack、RunPod、Hyperstack 官网, 国信证券经济研究所预测

表10: 中小云厂商 H100 全生命周期累计收入测算

	2024	2025	2026	2027	2028	2029
H100 租赁价格 (美元/小时)	2.49	2.37	2.25	2.13	2.03	1.93
租赁利用率	70%	70%	70%	70%	70%	70%
H100 全生命周期累计收入 (万美元)	1.53	2.98	4.36	5.66	6.91	8.09

资料来源: Lambda、FluidStack、RunPod、Hyperstack 官网, 国信证券经济研究所预测

回报率测算: 假设用电累计成本同上文微软假设相同, 假设单张 H100 采购价格为 3 万美金, 仅考虑单卡采购成本和单卡电费, 24-29 年回报率分别为-50%、-4%、

39%、79%、115%、148%。

表11: 中小云厂商 H100 回报率测算

单张 H100 回报率测算 (仅考虑单卡购买成本及电费)						
	2024	2025	2026	2027	2028	2029
单卡采购成本 (万美元)	3					
累计用电成本 (万美元)	0.04	0.09	0.13	0.17	0.21	0.26
成本 (万美元)	3.04	3.09	3.13	3.17	3.21	3.26
H100 全生命周期累计收入 (万美元)	1.53	2.98	4.36	5.66	6.91	8.09
投资回报率 (%)	-50%	-4%	39%	79%	115%	148%

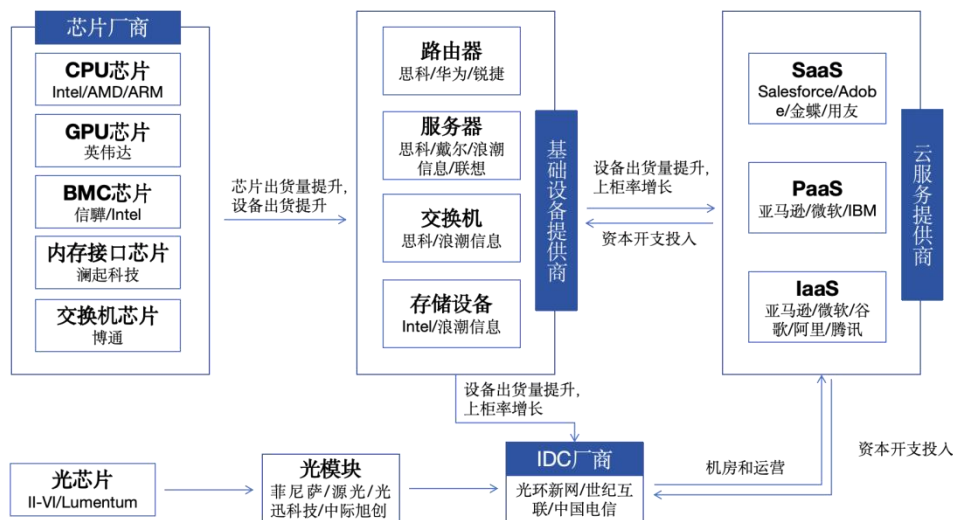
资料来源: 英伟达, 国信证券经济研究所测算

设备厂商业绩复盘：云厂商资本开支先验指标

在云服务商的资本开支中，设备厂商的业绩变化可以作为需求的先验指标。各设备厂商提供服务器、路由器、交换机等设备给 IDC 厂商或直接将设备出售给云厂商，因此复盘设备厂商的业绩变化与云服务商的资本开支的对应关系可以有助于判断云服务商的拐点。

产业链动态循环，形成“芯片-设备厂商收入-云服务商资本开支-云厂商收入”相互作用机制。云计算行业的进步主要得益于技术的创新，这要求产业链的各个环节能够相互配合，共同作用，以适应市场对于高效性、稳定性和扩展性云服务的需求。随着数据中心处理能力的增强，云服务提供商致力于推出更多样化、更高效的服务，目的是吸引更多的用户和满足他们的需求。这种日益增长的需求推动了云服务提供商增加资本支出，以扩大数据中心的规模和提升设备的技术水平。与此同时，芯片制造商也在持续推出性能更优越、能效更高的新产品。服务器和网络设备制造商利用这些高性能芯片，生产出更加先进的设备。这些设备随后被数据中心服务提供商部署在他们的数据中心内，目的是提高数据中心的的服务质量，支持云服务提供商在云计算领域的技术升级，并推动他们的业绩增长，形成“芯片-设备厂商收入-云服务商资本开支-云厂商收入”连续作用机制。

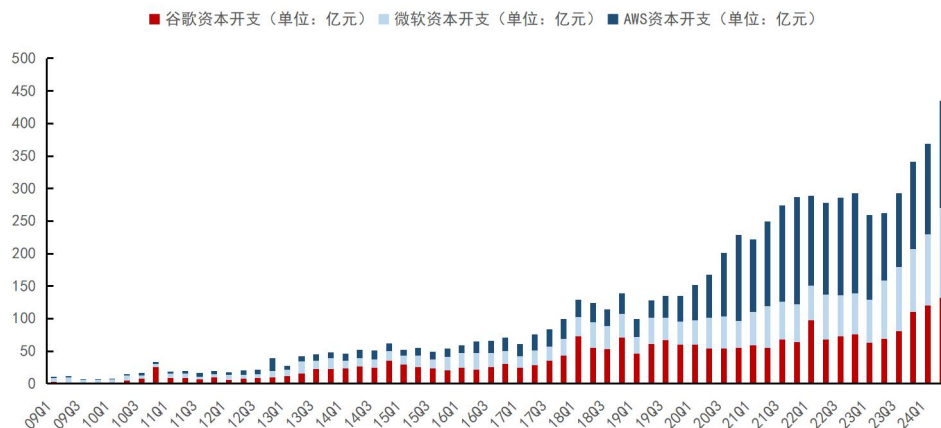
图82：“芯片-设备厂商收入-云服务商资本开支-云厂商收入”连续作用机制



资料来源：各公司官网，国信证券经济研究所整理

设备厂商业绩前置云厂商资本开支约 2 个季度。设备厂商业绩波动与云厂商的资本开支直接挂钩，亚马逊、谷歌、微软等头部云厂商资本开支主要用于购买服务器等基础建设。对比主要设备厂商思科、intel 等营收增速与云厂商资本开支增速，我们观测到设备厂商营收变化领先云厂商资本开支约 2 个季度。

图83: 海外头部云厂商资本开支总和变化 (单位: 亿元)

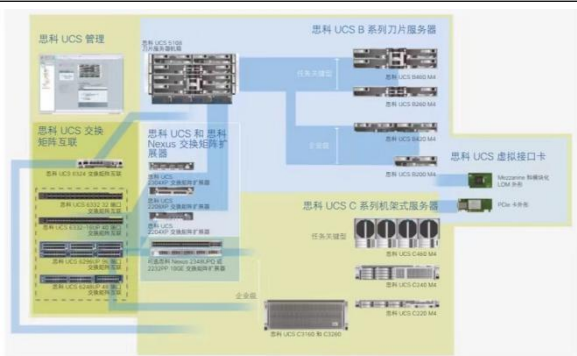


资料来源: 各公司官网, 国信证券经济研究所整理

思科: 全球网络设备领导者

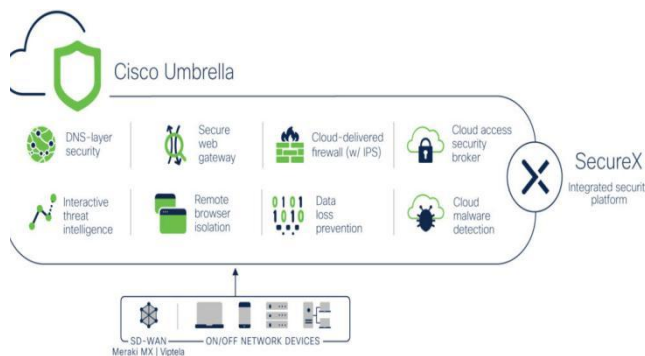
- **发展初期 (2009 年-2011 年):** SaaS 和 IaaS 成为云计算应用的两大主流趋势, PaaS 需求也飞速增长, 极大程度带动思科硬件设备与方案的研发与销售。这一阶段 Cisco 推出统一计算系统 (UCS), 利用高能效统一系统整合服务器、网络、存储访问和虚拟资源, 数据中心整合策略给基础架构市场带来突破性发展; 为应对战略性架构化云计算应用的安全风险, 思科提出包含 SecureX、安全云基础架构、云安全服务、安全云访问组成的云安全解决方案, 与数据中心和无边界网络服务完美结合。

图84: 思科统一计算系统



资料来源: 思科官网, 国信证券经济研究所整理

图85: 思科云安全解决方案

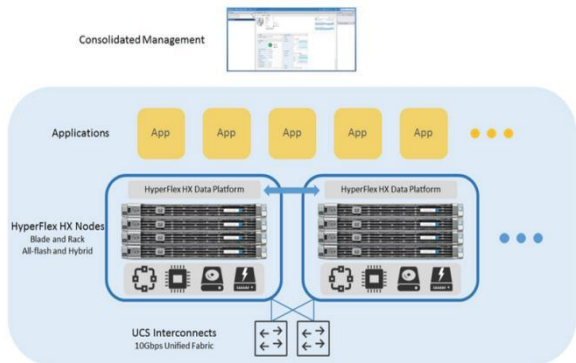


资料来源: 思科官网, 国信证券经济研究所整理

- **快速发展期 (2012 年-2021 年):** 随着云服务的全面发展, 思科向数据中心、安全、服务三大业务转型。在数据中心业务方面, UCS (统一计算系统) 持续发展, 同时推出 HyperFlex 超融合基础设施解决方案, 加入 Cisco Nexus9000 系列交换机与思科云 ACI, 完备的产品矩阵提高了数据中心工作的可扩展性与灵活性。安全业务向应用安全服务拓展, 提供云安全、Email 安全、认证

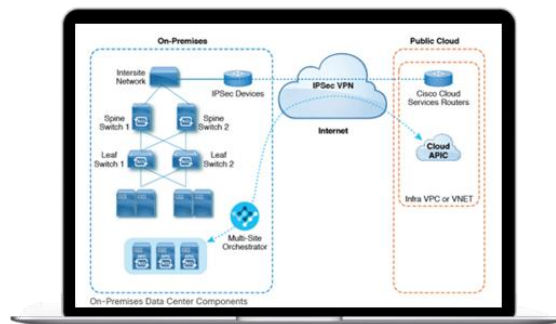
及接入管理、威胁保护与管理系统等系统的解决方案。服务业务方面，思科在 IT 资源托管服务基础上推出包括 Nexus Dashboard 和 Cisco vManage 在内的云服务，帮助云厂商构建、管理混合 IT 环境。

图86: Cisco HyperFlex 提供其他 HCI 的 3 倍性能



资料来源: Cisco 官网, 国信证券经济研究所整理

图87: 适用于多云环境的云 ACI



资料来源: Cisco 官网, 国信证券经济研究所整理

- **全面拥抱 AI 期（2022 年-现在）**：2022 年以来，“AI+云”成为趋势，思科也通过 AI 赋能增强其在网络、安全与服务解决方案上的能力。发布思科 AI 助手，在 Webex 中集成 AI 技术推出自主学习型 Webex Contact Center 并与英伟达携手推出面向生成式 AI 的基础设施解决方案，加速 AI 在云厂商决策与管理中的部署。同时，Networking Cloud 统一网络平台通过 Cisco ThousandEyes 和 Nexus Dashboard 为客户在多云环境提供新的可见性与控制水平。在安全业务上其 Security Cloud 的推出亦通过 AI 简化策略管理的同时提高威胁响应能力。

图88: Cisco 在 AI 方面的收购与投资



资料来源: Cisco 官网, 国信证券经济研究所整理

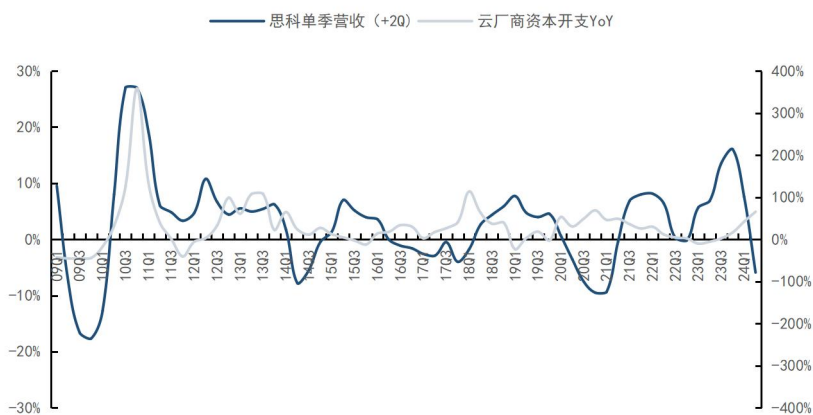
图89: Cisco 宣布与 NVIDIA 合作推出突破性 AI 集群解决方案



资料来源: Cisco Instagram, 国信证券经济研究所整理

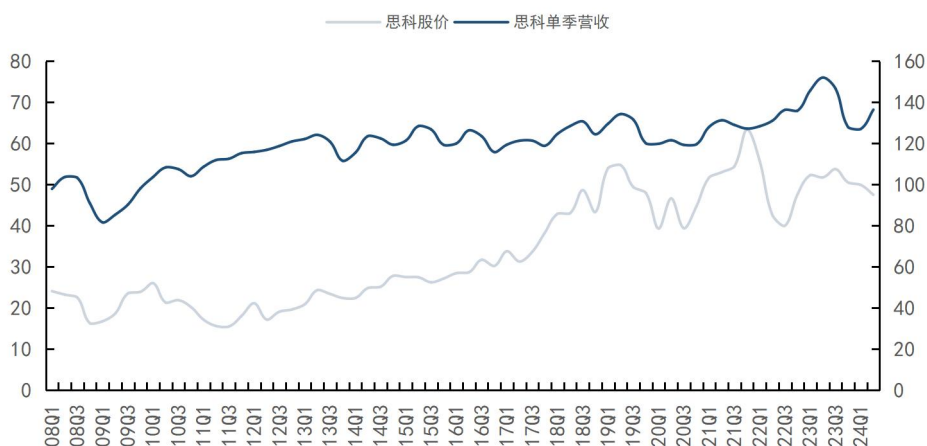
思科设备厂商的业绩变化是下游云厂商资本开支变动的先验指标。通过对比思科单季营收与云厂商资本开支，发现思科单季营收 YoY 领先云厂商资本开支 YoY 两个季度，说明思科设备厂商的业绩变化是下游云厂商资本开支变动的先验指标，对思科的营收波动关注有助于判断云厂商的需求波动。

图90: 思科单季营收与云厂商资本开支 YoY 比较



资料来源: 思科财报, 国信证券经济研究所整理

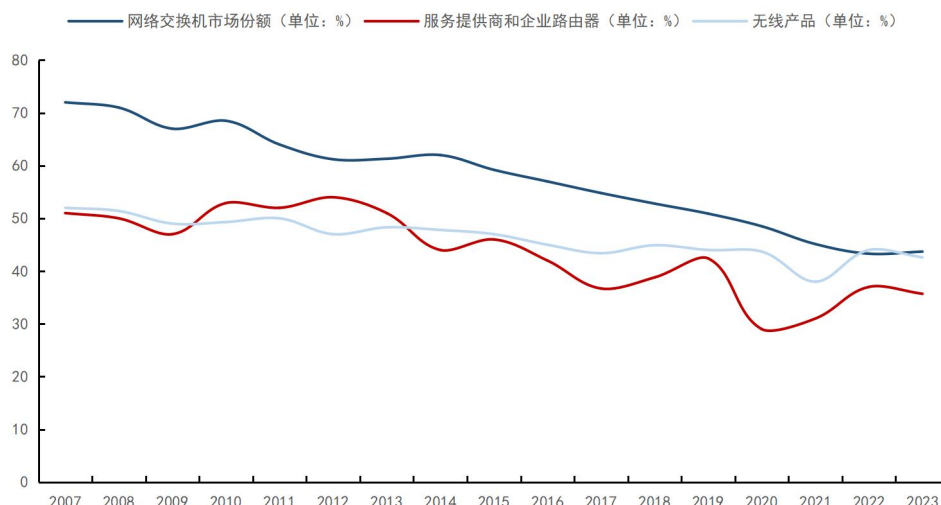
图91: 思科股价与营收比较



资料来源: 思科财报, 国信证券经济研究所整理

思科网络设备在全球仍占据主导地位。从 2007 年以来, 受云服务发展影响, 以网络交换机、网络服务与路由器、无线网络产品为核心产品的思科面临新兴公司的冲击与竞争, 一部分市场份额被蚕食, 但总体来说在全球市场仍占据主导地位。

图92: 思科 2007 年至 2023 年主要产品的市场份额变化 (单位: %)



资料来源: Cisco 财报, IDC, 国信证券经济研究所整理

英特尔: CPU 计算领域全球领导者

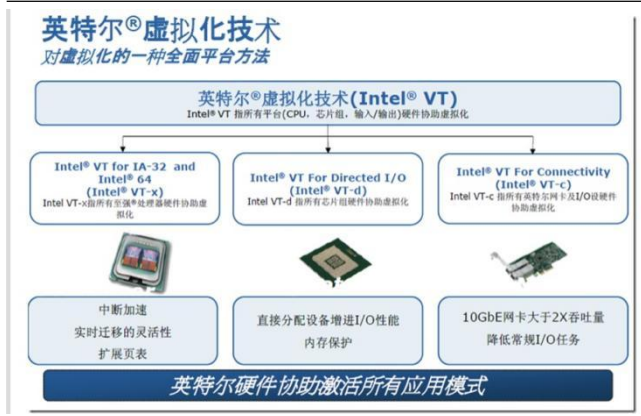
- **积极探索期 (2009 年-2012 年):** 为适应云计算与虚拟化技术的兴起, Intel 推出系列数据中心创新产品: 1) 至强处理器和基于 Nehalem 架构的 Nehalem-EX 处理器适配多数据中心; 2) 实现虚拟化技术 Intel VT 的 2.0、3.0 版本迭代, 其中的 VT-X、VT-D 技术对应解决虚拟机性能损耗、安全和 I/O 资源分配等问题。在此阶段, Intel 启动云构建器计划优化云厂商基础设施建设, 宣布推出云搜索 Intel Cloud Finder, 用于提高搜索和选择云服务商效率。

图93: 英特尔至强处理器 5500 系列服务器



资料来源: 英特尔官网, 国信证券经济研究所整理

图94: 英特尔虚拟化技术架构



资料来源: 英特尔官网, 国信证券经济研究所整理

- **全面发展期 (2013 年-2020 年):** 云计算发展迅速, Intel 进一步推出以数据为中心的产品技术组合。在数据计算方面, Intel 推出第二代至强可扩展

处理器，与基于 FPGA 的加速解决方案结合使用。在数据存储方面，Intel 发布傲腾数据中心级持久内存，推出傲腾固态硬盘 DC P4800X 与英特尔固态硬盘 DC P4510。在数据传输方面 Intel 发布以太网 800 系列、Omni Path 架构和硅光子技术，为云厂商提供端到端互联解决方案。

图95: 英特尔以数据为中心的产品技术组合



资料来源：Intel 官网，国信证券经济研究所整理

图96: 第二代英特尔至强可扩展处理器



资料来源：Intel 官网，国信证券经济研究所整理

- **AI 赋能期（2021 年-至今）**：顺应 AI 赋能云服务的趋势，Intel 将至强可扩展处理器更新至第五代 Emerald Rapids，通过增强型深度学习加速技术增强 AI 推理和训练性能；同时推出 oneAPI 开放跨平台编程模型支持 AI 和数据数据分析工作负载。在此阶段 Intel 提供 Agilex 等系列 AI 芯片与酷睿 Ultra 处理器，结合 CPU、GPU、NPU 提供强大 AI 处理能力。

图97: AI 赋能全新应用场景



资料来源：英特尔官网，国信证券经济研究所整理

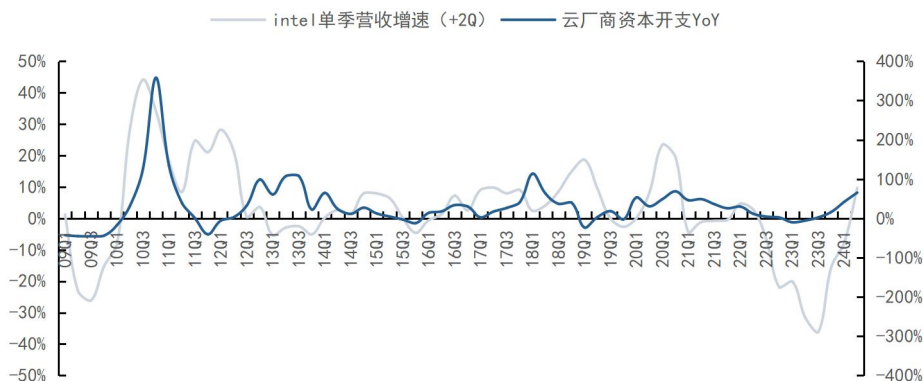
图98: 英特尔酷睿 UltraU/H 系列处理器

Intel®	Core™	Ultra 5 processor 125H
Intel®	Core™	Ultra 5 processor 125U
Intel®	Core™	Ultra 5 processor 134U
Intel®	Core™	Ultra 5 processor 135H
Intel®	Core™	Ultra 5 processor 135U
Intel®	Core™	Ultra 7 processor 155H
Intel®	Core™	Ultra 7 processor 155U
Intel®	Core™	Ultra 7 processor 164U
Intel®	Core™	Ultra 7 processor 165H
Intel®	Core™	Ultra 7 processor 165U
Intel®	Core™	Ultra 9 processor 185H

资料来源：英特尔官网，国信证券经济研究所整理

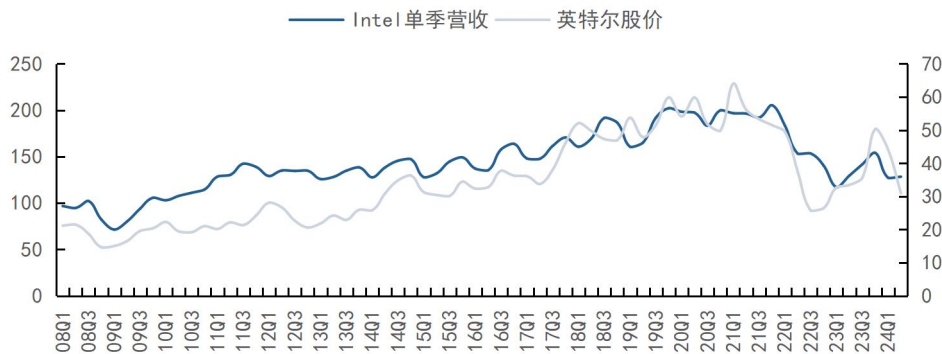
Intel 业绩变化是下游云厂商资本开支变动的先验指标。 Intel 单季营收 YoY 领先云厂商资本开支 YoY 两个季度，可从其业绩变化反映下游云厂商需求的先验指标，对 Intel 的营收波动关注有助于判断云厂商的需求波动。

图99: Intel 单季营收与云厂商资本开支 YoY 比较



资料来源: Intel 财报, 国信证券经济研究所整理

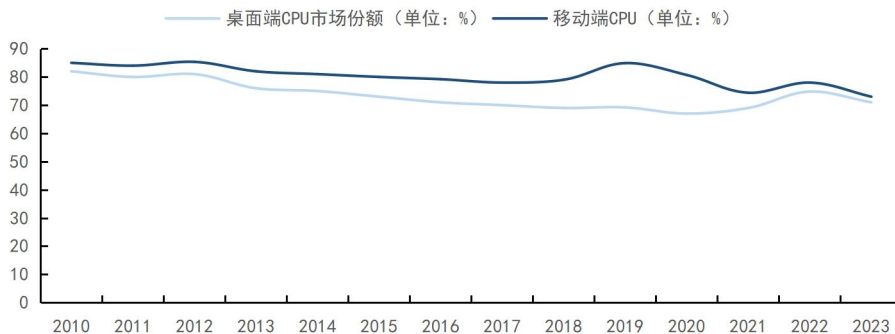
图100: Intel 股价与营收比较



资料来源: Intel 财报, 国信证券经济研究所整理

云服务迅速兴起并发展, AMD 的崛起以及市场对更高性能与更优价格比的处理器的需求增加, 英特尔市场份额受到一定冲击, 但仍保持市场领先。同时, 英特尔也在积极拓展业务范围, 加大对于物联网、人工智能、5G 与智能驾驶业务的开发力度, 力争在保持核心产品市场份额的情况下寻找可替代的增量业务。

图101: 2010 年至 2023 年 Intel 桌面端 CPU 与移动端 CPU 的市场份额变化 (单位: %)



资料来源: Intel 官网, IDC, 国信证券经济研究所整理

浪潮信息：国产服务器领导者

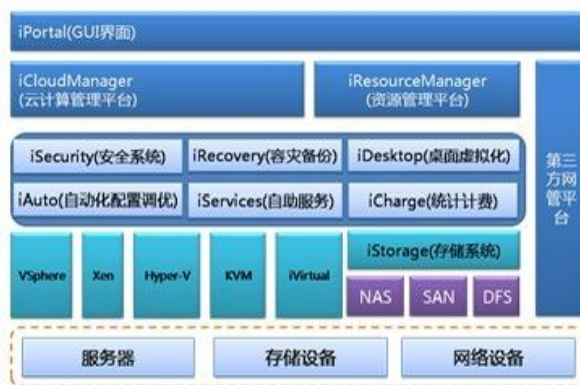
- **业务探索期（2010年-2013年）**：2010年全球云计算蓬勃发展，浪潮信息逐步推动业务向云计算转型。2010年发布“云海”云计算战略并与山东计算中心签署云计算合作协议，在2011年推出我国首款云数据中心操作系统产品浪潮云海OS，并于2012年发布第二代云海OS V2.0。

图102: 浪潮“云海”云计算战略架构



资料来源：浪潮信息官网，国信证券经济研究所整理

图103: 浪潮云海OS V2.0三层架构



资料来源：浪潮信息官网，国信证券经济研究所整理

- **快速增长期（2014年-2021年）**：移动互联网拉动云计算快速发展，浪潮进一步巩固在云计算领域的领先优势。2014年浪潮公布“大数据重构企业智慧”战略和“1+4”企业云战略，同步披露企业云落地路线图。云硬盘EBS、对象存储OSS、云服务器ECS、云物理主机CPS、虚拟私有网络VPC等从计算、存储、网络等方面为云厂商的服务提供强有力保障。

图104: 浪潮云分布式架构



资料来源：浪潮云，国信证券经济研究所整理

图105: 浪潮云服务器ECS专业性场景



资料来源：浪潮云，国信证券经济研究所整理

- **AI赋能期（2022年-至今）**：乘AI赋能云计算之势，浪潮逐步从通用计算服务器向AI服务器生产转型，发布NF5698G7、NF5688G7等AI训练服务器，以及NF5468G7、NF5468A5、NF5468M6等AI推理服务器；同时，浪潮关注边缘云，发布边缘计算“4+3+X”全栈产品家族，携手Intel等合作伙伴共建边缘开放的融携实验室，为多元场景带来新的算力连接、运维与支持。“4”包含

边缘微中心、边缘服务器、便携 AI 服务器、边缘微服务器，如 NE5260G7 与 NE3160G7；“3”指缘视、缘脑、缘智三大软件平台；“X”代表云边协同、边缘 AI、超融合等多类边缘计算解决方案。

图106: 浪潮信息 AI 赋能相关产品



资料来源：浪潮信息官网，国信证券经济研究所整理

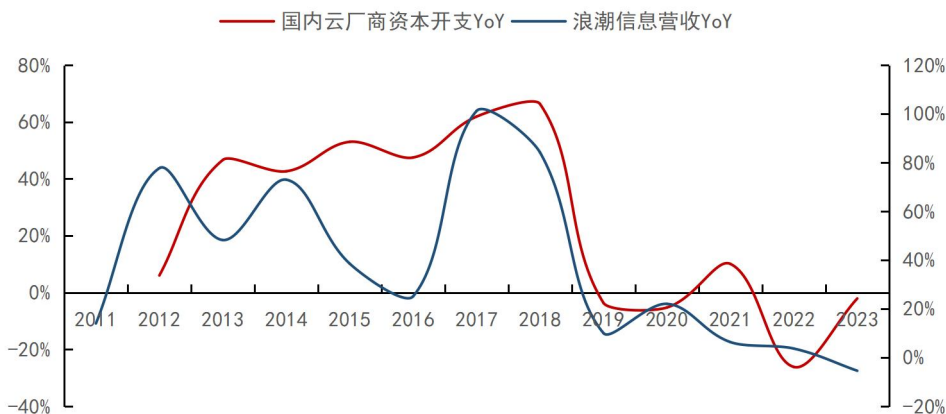
图107: 浪潮信息边缘云代表 IEC 产品架构



资料来源：浪潮云，国信证券经济研究所整理

浪潮信息营收可以作为观测国内云厂商需求的先验数据。国内云厂商资本开支数据取自阿里云、百度云、腾讯云资本开支总和，因阿里云按年披露资本开支，做年度数据观测得出浪潮信息营收 YoY 前置国内云厂商资本开支 YoY 一个年度，浪潮信息营收可以作为观测国内云厂商需求的先验数据。

图108: 浪潮信息营收与国内云厂商资本开支 YoY 比较



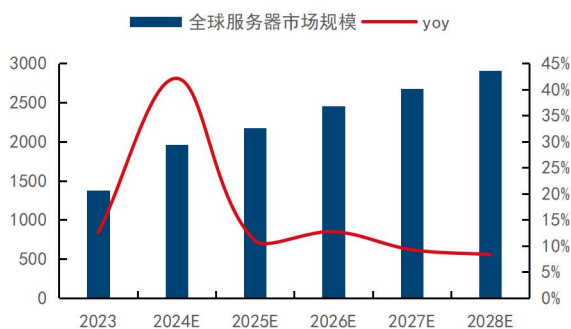
资料来源：阿里云、百度云、腾讯云财报、浪潮信息财报，国信证券经济研究所整理

格局演绎：云时代设备厂商群雄逐鹿，AI 时代再迎新王

服务器、交换机市场持续增长，AI 受益厂商份额快速提升

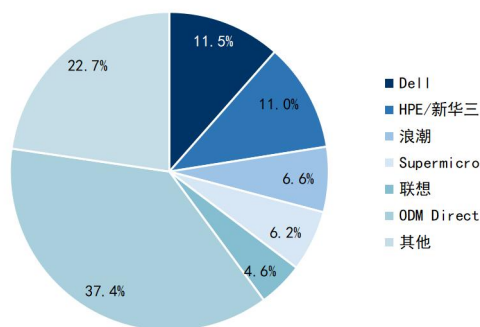
24 年服务器市场有望大幅增长，Dell 仍是全球份额第一。根据 IDC 最新数据，全球服务器市场在 2024 年第二季度的支出增长了 64.1%，主要是由于超大规模企业，以及云厂商对 GPU 服务器采购增加。非 GPU 服务器业务增长有所恢复，整体业务同比增长 16.7%。IDC 预期 2028 年的支出目标提高到 2920 亿美元，未来五年市场复合增速达到 16.0%。AI 是推动市场高速增长的核心原因，尤其是 24 年非 X86 服务器预期保持 58% 以上的增速。从 23 年服务器市场份额来看，虽然 Dell、HPE 和新华三、浪潮稳居前三，但服务器收入均有下滑；前五中只有 Supermicro 增长较快，份额提升明显，主要得益于 GPU 服务器的巨大需求。另一方面，ODM 产品收入和份额也增长较快，成为近年来产业趋势。

图109: 全球服务器市场规模预测（亿美金）



资料来源：IDC，国信证券经济研究所整理

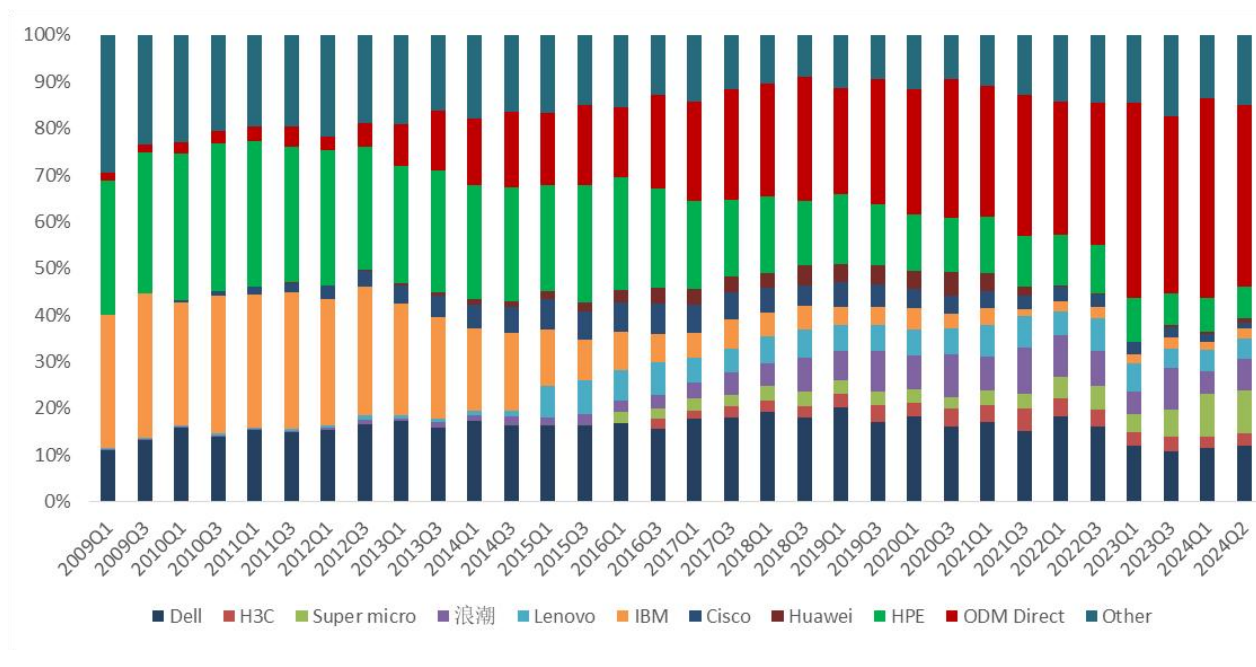
图110: 2023 年全球服务器市场份额



资料来源：IDC，国信证券经济研究所整理

云计算快速增长时代推动 ODM 成趋势，国内浪潮受益明显，AI 时代 Supermicro 持续提升份额。全球服务器市场玩家众多，包括 Dell、IBM、Supermicro、HPE、思科、Oracle、浪潮、华为、联想等。自 2002 年亚马逊推出 AWS 云计算服务以来，市场经过几年的教育后终于迎来产业共识，并形成投资爆发。早期的服务器市场仍以传统大牌厂商 Dell、IBM、HPE 为主，ODM 占比较小。但在各云厂商逐步兴起，并加大投资后，由于大量采购触发了云厂商的定制化需求，以及对价格更为敏感，ODM 成为产业趋势。Dell、HPE 份额均有下滑，IBM 于 2014 年将大部分服务器业务卖给了联想。国内云计算兴起较晚，2015 年左右国内各大云厂商开始发力，并推动了浪潮、华为等厂商的加快成长，目前浪潮是全球第三的服务器厂商，华为则受到了相关制裁影响。2023 年大模型爆发推动计算设备进入 AI 时代，美国的芯片限制对国内服务器厂商产生了较大影响，Dell 和 Supermicro 则较为受益，尤其是 Supermicro 凭借 AI 服务器快速提升份额。同时，作为 AI 芯片龙头的英伟达，其 AI 服务器也在迅速增长，虽然整体体量较小，但经测算 23 年市场份额也达到约 0.81%，份额提升近 2 倍。

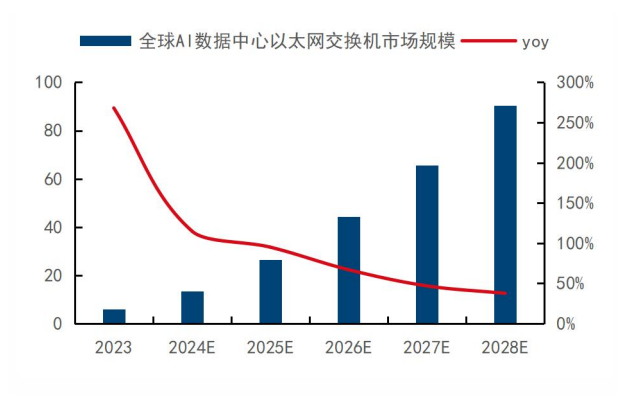
图111: 近年来全球服务器厂商市场份额变化



资料来源: IDC, 国信证券经济研究所整理

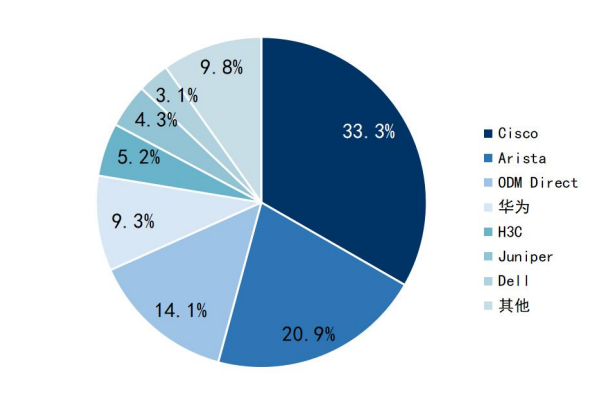
2024年全球以太网交换机市场承压,但AI数据中心交换机仍保持高速增长,Cisco和Arista份额地位稳固。根据IDC《全球以太网交换机季度跟踪报告》,2023年全球以太网交换机市场规模达到442亿美元,同比增长20.1%。但2024年第一二季度市场均出现了下降,24Q2全球以太网交换机收入同比下降14.1%至102亿美元。整体来看数据中心市场依然表现较好,Q2增长7.6%;主要受到AI产业的推动以及对超高速交换机的强烈需求,二季度200/400GbE交换机的总市场收入同比猛增104.3%。根据IDC对生成式AI数据中心交换机市场的预测,市场规模有望从2023年的6.4亿美元增长至2028年的91亿美元,复合增速超过70%。参考2022年数据中心以太网交换机市场份额,Cisco和Arista在数据中心市场仍然表现强势,AI对高速交换需求提升后,二者优势将更为显著。

图112: 全球AI数据中心以太网交换机市场规模(亿美元)



资料来源: IDC, 国信证券经济研究所整理

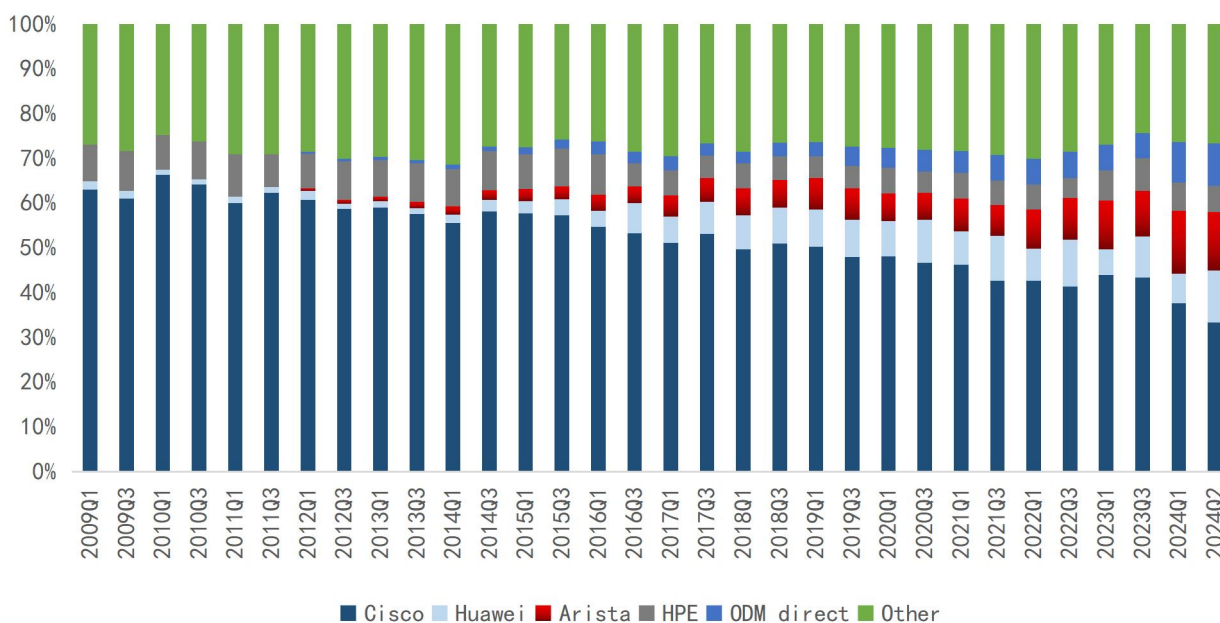
图113: 2022年数据中心交换机市场份额



资料来源: IDC, 国信证券经济研究所整理

Cisco 整体份额有所下降，AI 和高性能网络需求驱动 Arista。云计算时代 Cisco 仍是全球交换机龙头，但整体份额随着华为、Arista、ODM 产品的崛起而持续缓慢下滑。与服务器产业逻辑类似，华为交换机份额在 15 年国内云计算产业爆发后开始增长加快；ODM 也逐步成为各大云厂商偏好的产品模式。Arista 凭借全球高速以太网交换机的产品优势，首先在数据中心崭露头角，并在 AI 时代迎来爆发。尤其是 AI 需求拉动，大量云客户需要从 100G 升级到 200G 或从 200G 升级到 400G 的网络产品。另一方面，英伟达凭借 AI 芯片领域的组网优势，尤其 infiniband 技术在模型训练阶段的领先，其交换机业务近两年也迅速放量，23 年份额超过 1%。

图114: 近年来全球交换机厂商市场份额变化



资料来源：IDC，国信证券经济研究所整理

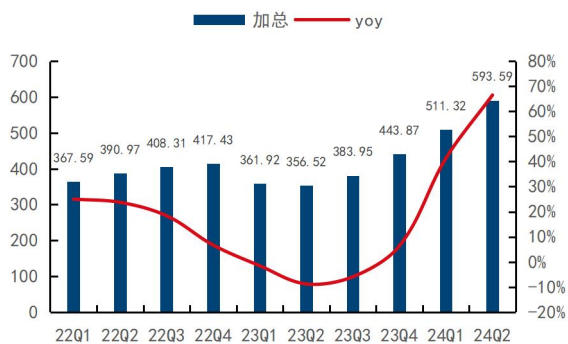
AI 时代掀起新一轮资本开支，英伟达仍是 AI 芯片王者

美国科技巨头持续加大资本开支，加码 AI 投入。随着云时代的日渐成熟，22 年各云厂商资本开支总和增速已经同比放缓，并在 23 年 Q1、Q2 延续了惯性下滑趋势。但是自 23Q1 大模型火爆之后，美国科技巨头均开始加大 AI 领域投资，并在 23Q3 呈现拐点，23Q4 和 24H1 各云厂商均开启了新一轮投资扩张周期。根据对谷歌、微软、AWS、META、Oracle 五家巨头的加总计算，24Q1 资本开支合计达到 511.32 亿美元，同比增长 41.28%；24Q2 资本开支合计达到 593.59 亿美元，同比增长 66.50%。从各家增速来看，谷歌、微软、Oracle 在 24 年资本开支增速提升明显。在本轮 AI 革命背景下，各云巨头均担心“踏空”，因此 24 年开始继续加大 AI 领域的资本开支。谷歌指引全年每季度的资本支出将大致保持在或高于 Q1 资本支出水平（120 亿美元）。META 上调资本开支指引，从之前 350 亿至 400 亿美元的范围，上调至 370 亿至 400 亿美元之间；同时公司预计 2025 年的资本支出将显著增长。微软指引 2025 财年资本支出将高于 2024 财年。

芯片设备等厂商受益于云厂商资本开支周期，但整体呈现上涨趋势。综合服务器、交换机、芯片等厂商来看，在云计算初始投资阶段，基础设施的建设在前期有一

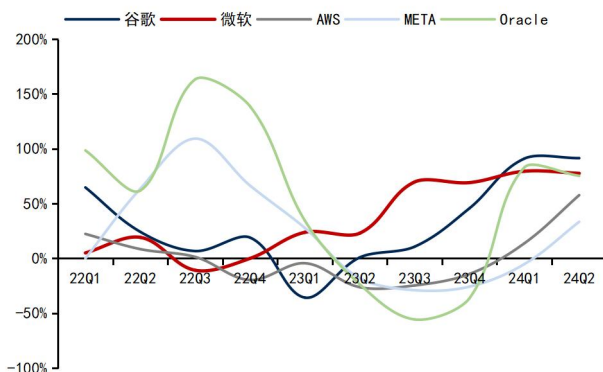
定周期性，主要是云厂商资本开支在新领域加大投入后，会保持一定的投资节奏，以配合下游应用厂商逐步厂商。随着后续 SaaS 应用百花齐放并持续放量，其基础设施投入重回稳定上涨节奏。

图115: 各海外互联网厂商资本开支季度性变化



资料来源：各云厂商公告，国信证券经济研究所整理

图116: 各海外互联网厂商资本开支季度性增速变化

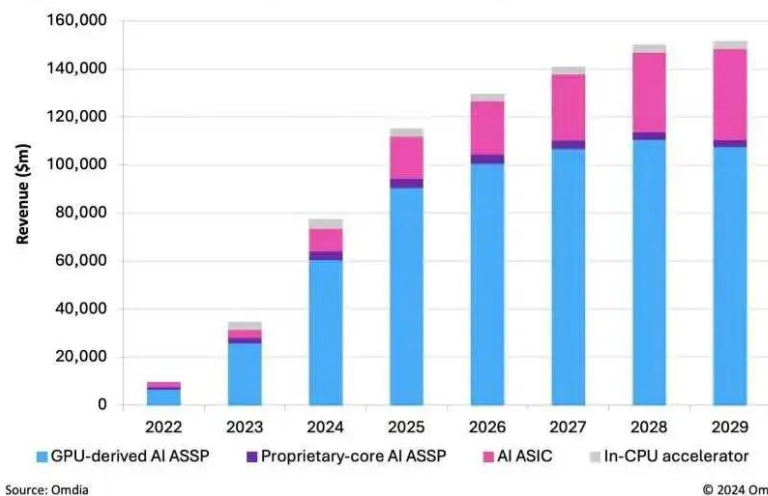


资料来源：各云厂商公告，国信证券经济研究所整理

近两年 AI 芯片预期仍是高速增长。根据研究咨询机构 Omdia 发布的《云计算和数据中心人工智能处理器预测》报告，22 年 AI 数据中心芯片市场规模不到 100 亿，24 年预期将达到 780 亿美元，24-25 年仍保持高速增长，预期 29 年市场将达 1510 亿美元。除了 GPU 市场高增长之外，AI ASIC 也成为当下 AI 芯片产业趋势，并保持快速增长。对比价格高昂的 GPU，AI ASIC 功耗和成本优势明显，且在训练和推理侧也均可使用。根据预测，到 2028 年 AI ASIC 芯片的市场规模有望突破 400 亿美元，年复合增长率高达 45%。

图117: AI 数据中心芯片市场规模预测

AI processors for cloud and enterprise data centers forecast by device, world markets: 2022–29

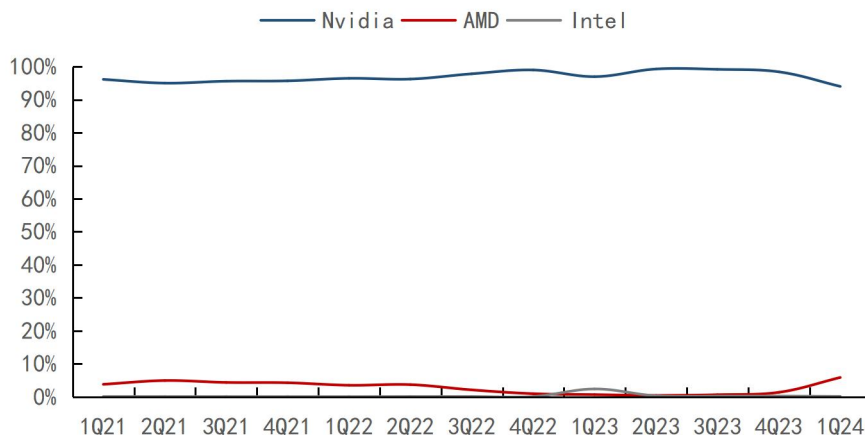


资料来源：Omdia，国信证券经济研究所整理

英伟达在服务器 GPU 市场仍占据极高市场份额。英伟达长期占据 GPU 市场几乎垄

断地位，前期货额均保持在 95%以上，24Q1 受到 AMD 冲击略有下滑，但份额也达到了 94%。AMD 于 23 年底推出新一代 AI/HPC 专用加速器 Instinct MI300 系列，在市场对 AI 算力需求井喷的背景下，MI300 销量增长迅速。MI300 的上市给市场多了一种选择，也相对更具性价比，但目前市场英伟达仍占据统治地位。

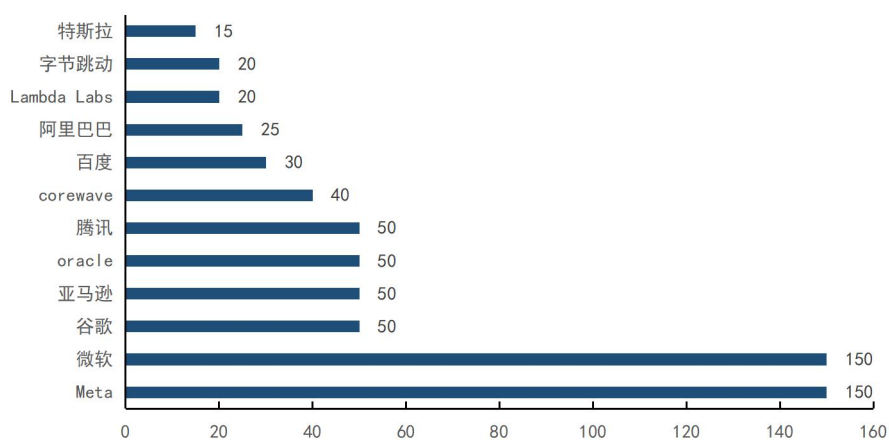
图118: 全球服务器 GPU 市场份额变化



资料来源：IDC，国信证券经济研究所整理

Meta 和微软向英伟达采购 AI 芯片量最大。根据 Omdia Research 的报告，23 年第三季度 Meta 和微软两家公司以 15 万块 H100 GPU 的购买量并列位居第一；谷歌、亚马逊、甲骨文以及腾讯以 5 万块 H100 GPU 的购买量共同位居第三。英伟达将大多数的 GPU 服务器都供应给了大型云服务提供商。根据媒体报道，微软 2023 年在采购英伟达芯片上花费了 45 亿美元，微软内部计划到 2024 年底储备 180 万块 AI 芯片，到 2027 财年，微软预计将在 GPU 和数据中心上投资约 1000 亿美元。Meta 也表示将在今年年底前向英伟达购买 35 万个 H100 GPU 芯片，从而使该公司的 GPU 总量达到约 60 万个。

图119: 2023Q3 英伟达各客户采购 AI 芯片数量预估 (千块)



资料来源：Omdia，国信证券经济研究所整理

投资建议

从复盘云计算厂商发展历史来看，其资本开支往往跟随行业发展趋势、自身战略调整呈现脉冲式增长：从行业维度来看，2010年美国“云优先”的发展战略，云计算开始进入快速发展期，各厂商加大资本开支投入；从公司自身战略来看，14年微软提出“移动为先，云为先”战略，18年谷歌制定“追赶战略”，云厂商资本开支均出现大幅上升。从当前阶段看，行业维度，AI云是大势所趋，公司维度，各家将AI发展放到核心地位，带动本轮资本开支的快速增长。

根据复盘硬件厂商同云厂商资本开支的关系，硬件厂商的业绩反应往往领先于云厂商资本开支的投入，关注AI云计算相关硬件厂商，例如海光信息、浪潮信息等。

风险提示

全球云厂商资本开支不及预期。根据前文所述，资本开支快速增长会拉动折旧上行，进而利润承压，云厂商为保持业绩稳健，后续可能会减少资本开支投入。

AI应用商业化不及预期。AI应用商业化落地不及预期，客户减少对AI云算力的需求，进而减少云厂商在AI领域资本开支的投入，进而影响产业链相关公司的业绩释放。

技术快速迭代，硬件供应商变动风险。全球人工智能快速发展，以英伟达为例，自身芯片快速迭代，技术方案快速革新，若硬件供应链厂商无法满足最新技术的需求，自身业绩有大幅下滑的风险。

免责声明

分析师声明

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。

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 $\pm 10\%$ 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 $\pm 10\%$ 之间
		弱于大市	行业指数表现弱于市场代表性指数 10%以上

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中所提及的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。

国信证券经济研究所

深圳

深圳市福田区福华一路 125 号国信金融大厦 36 层
邮编：518046 总机：0755-82130833

上海

上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层
邮编：200135

北京

北京西城区金融大街兴盛街 6 号国信证券 9 层
邮编：100032