

半导体行业深度报告

Agentic AI 时代的算力重构：CPU，从“旁观者”到“总指挥”的价值回归

增持（维持）

2026 年 01 月 28 日

证券分析师 陈海进

执业证书：S0600525020001

chenhj@dwzq.com.cn

证券分析师 李雅文

执业证书：S0600526010002

liyw@dwzq.com.cn

投资要点

■ **从对话模型到行动智能体：Agent 架构重塑 CPU 与 GPU 分工：**市场认为 AI 算力=GPU，CPU 只是负责调度。但 AI 落地方向来到 Agent，由“纯对话”转向“执行任务”时，算力需求发生了结构性分化。**1) 执行控制流 CPU 化：**Agent 的行动阶段是大量 if/else 判断与系统调用。Manus 等主流 Agent 架构为每个 Agent/任务分配一个隔离的云端虚拟机，不同任务在逻辑上高度异构（如网页浏览、代码修改、环境部署等），任务之间可并行但控制流完全不同。分支类任务不适合由 GPU 执行，因控制流发散导致算力利用率急剧下降。而分支类任务却正是 CPU 微架构长期优化的主战场。**2) 记忆体系去 GPU 化：**在 Agent 场景下，长上下文推理会产生巨大的 KV cache，其占用随对话轮次与上下文长度线性增长，快速耗尽 GPU HBM 容量。而 CPU 搭配大容量 DDR5 / LPDDR5（并通过 CXL 扩展）承载 KV Cache 与部分参数，正在成为兼顾吞吐、扩展性与成本效率的主流架构选择。

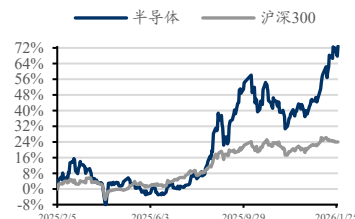
■ **Agent 软件基础设施加速落地，CPU 多核化与需求放量进入确定性通道：**从实验数据复盘来看，CPU 成为 Agentic AI 的真实瓶颈。研究显示，在完整的 Agent 执行链路中，工具处理相关环节（检索、Python/Bash 执行、Web 请求等）在 CPU 上消耗的时间占端到端延迟的比例最高可达 90.6%。在高并发场景下（Batch Size 提升至 128），CPU 端到端延迟从 2.9 秒跃升至 6.3 秒以上。我们认为研究结果揭示了在大量 Agentic 场景中，系统吞吐受限的并非 GPU 计算能力，而是 CPU 的核心数并发调度问题。我们由此判断，进入 Agent 时代后，CPU 侧的工具执行与调度能力将从 GPU 的附属角色，演化为需要被单独规划与优化的核心资源池。

产业端推进方面，AWS 和 Google Cloud 等头部 CSP 正在加速建设面向 Agent 的沙盒环境软硬件基础设施，率先在软件层面强化 Agent Sandbox 的隔离与编排能力，通过运行时与调度体系的完善，为后续 CPU 侧基础设施规模化部署奠定基础。与此同时，CPU 龙头也在 Agent 驱动下向超多核架构演进：AMD 推出的 Turin 最高可达 192 核；Intel 的 Sierra Forest 采用纯能效核设计，核心数可达 144 甚至 288 核。我们认为，随着 Agent 商业化推进，厂商必须持续压低每次任务执行成本。在这一目标下，超多核 CPU 以更高的并行度、更低的单位功耗，支撑大规模、长期运行的 Agent 执行环境。英伟达亦在新架构中提升 CPU 配比，CPU 的重要性抬升，挂钩 Agentic 的叙事。我们认为 NVIDIA 主动提升 CPU 权重，等同于在系统层面确认：在长上下文与高并发 Agent 场景中，大内存 CPU 是承载海量 KV Cache 的最优容器。

■ **投资建议：**我们看好海光信息与澜起科技，前者受益于 Agentic AI 推动 CPU 从调度走向高并发执行的核心算力地位抬升，后者受益于长上下文与高并发场景下 CXL 驱动的“存力+算力”协同升级。建议关注龙芯中科，万通发展（数渡科技）。

■ **风险提示：**Agent 落地节奏不及预期风险；新型硬件架构替代风险；CSP 资本开支波动风险。

行业走势



相关研究

《关注 AI 算力需求快速发展，看好国产设备商充分受益》

2025-07-19

《长鑫存储启动上市辅导，看好国内先进制程扩产利好国产设备商》

2025-07-08

内容目录

1. 从对话模型到行动智能体：Agent 架构重塑 CPU 与 GPU 分工	4
1.1. 主流 Agent 形态从模型推理走向“浏览器+代码执行”	4
1.2. 为何 CPU 在 Agent 时代具有不可替代性	4
1.2.1. Agent 执行阶段的离散控制流决定 CPU 优于 GPU	5
1.2.2. 长上下文推理亦推升 CPU 侧内存连接需求	5
2. Agent 软件基础设施加速落地，CPU 多核化与需求放量进入确定性通道	7
2.1. Agentic AI 的系统瓶颈已由 GPU 计算转向 CPU 并发调度	7
2.2. CPU 龙头迭代方向明确：Agent 驱动下的超多核架构演进	8
2.3. 自研 CPU 继续加码，挂钩 Agentic 的叙事	9
2.4. Engram 架构重构记忆路径，CPU 重要性进一步强化	10
3. 相关公司	11
4. 风险提示	13

图表目录

图 1: OpenAI 的 Agent 模式集成多种工具调用形态.....	4
图 2: Manus 沙盘.....	4
图 3: 不同 GPU 在不同 Warp 分支发散程度下性能表现 (单位: 毫秒)	5
图 4: CPU 在复杂任务编排中具备同步与调度能力.....	5
图 5: Agent 任务中, CPU 占据端到端总延迟的绝大部分	7
图 6: Intel Xeon 6700/6900 代数据中心 CPU 的核心数最高可达 288 核.....	8
图 7: AMD 数据中心 CPU 核心数持续上行	9
图 8: 英伟达 GB200 Grace Blackwell Superchip 结构图	10
图 9: DeepSeek 明确由 CPU 调度 Engram 数据库, CPU 重要性凸显	10
图 10: AMD 营收拆分及盈利预测 (亿美元)	11
图 11: Intel 营收拆分及盈利预测 (亿美元)	11
图 12: 海光信息 CPU 代际梳理.....	12

1. 从对话模型到行动智能体：Agent 架构重塑 CPU 与 GPU 分工

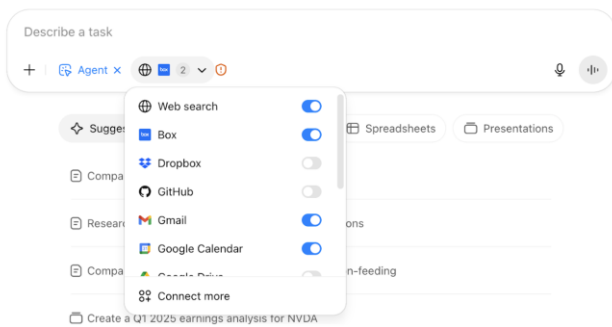
1.1. 主流 Agent 形态从模型推理走向“浏览器+代码执行”

Agent 的本质是运行在一个具备操作系统能力的云端沙盒计算环境中。我们认为 Agent 的能力跃迁，并非来自在 GPU 上堆叠更多 token 进行推理，而是来自调用外部系统（浏览器、文件系统、终端、网络）完成真实世界任务。在成熟 Agent 架构中，GPU 只是被操作系统调度的加速器，Agent 的“身体”是一台完整的云端计算机。OpenAI 官方披露，ChatGPT Agent 通过视觉浏览器、代码解释器等工具完成复杂多步骤任务，能力不局限于模型内部推理。Manus 将产品目标明确表述为：不仅思考，更要替用户行动。其核心基础设施是 Manus 沙盒——一台真实的云端电脑/虚拟机：

- **任务级隔离：**每个 Agent/任务分配一台独立云端虚拟机，任务间完全隔离，可并行执行。
- **完整计算环境：**VM 具备操作系统级能力，包括网络、文件系统、浏览器、终端与各类软件工具。

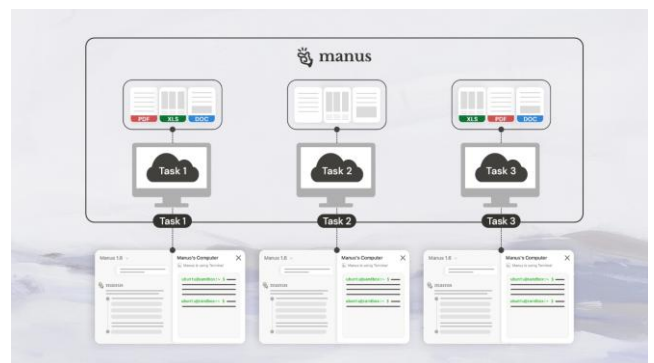
主流 Agent 架构为每个 Agent 配备一台云端电脑，Agent 并非直接运行在 GPU Kernel 上，而是运行在具备完整操作系统能力的云端计算环境中。Logto 的技术分析指出，Manus 云端浏览器本质是运行在云端的私有浏览器环境，可维持登录状态与会话连续性。据 E2B，Manus 多 Agent 系统需要完成网页浏览、文件搜索、终端命令等复杂任务，因此需要一台完整的云端电脑。LatentSpace 更是指出，包括 Mistral、Perplexity、HuggingFace、LMArena、Gumloop 等在内的 Agent 系统，均采用“每个 Agent 配一台云端电脑”的架构范式。

图1: OpenAI 的 Agent 模式集成多种工具调用形态



数据来源：OpenAI，东吴证券研究所

图2: Manus 沙盒



数据来源：Manus，东吴证券研究所

1.2. 为何 CPU 在 Agent 时代具有不可替代性

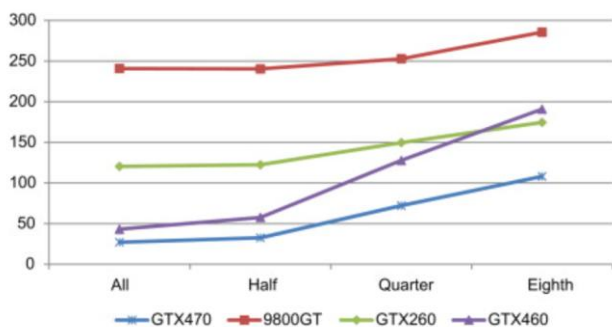
1.2.1. Agent 执行阶段的离散控制流决定 CPU 优于 GPU

Agent 运行在沙盒化的通用计算机环境中，这一执行形态决定了其工作负载天然呈现高度分支化与强控制流特征。Manus 为每个 Agent 任务分配一个隔离的云端虚拟机，不同任务在逻辑上高度异构（如网页浏览、代码修改、环境部署等），任务之间可并行但控制流完全不同。

分支类任务不适合由 GPU 执行，控制流发散导致算力利用率急剧下降。英伟达在其 CUDA 官方文档中明确指出，GPU 采用 SIMT 执行模型：线程以 32 个为一组(warp)，在同一周期内执行同一条指令。一旦同一 warp 内线程进入不同分支路径 (Warp Divergence)，GPU 会将不同分支串行执行，不属于当前分支的线程被 mask 掉，直接导致吞吐下降。实验结果显示，即便是有限程度的分支发散，也会显著拉长执行时间。在 32 路完全发散的极端情况下，性能下降可达 27-125 倍，等效算力利用率仅剩个位数百分比。因此，分支类任务不适合由 GPU 执行，是 GPU 编程中需要避免的情形。

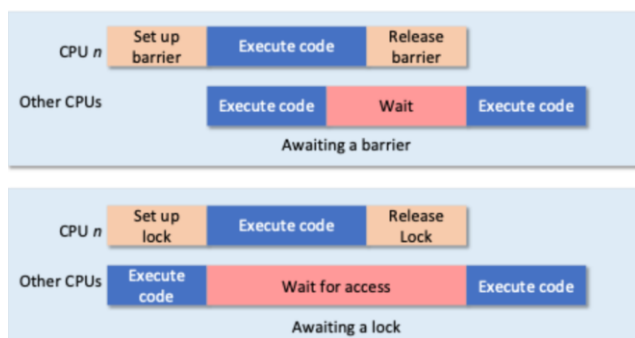
CPU 为复杂控制流而生，天然适配 Agent 的执行特征。现代服务器 CPU (如 AMD EPYC、Intel Xeon) 属于典型 MIMD 架构：每个核心拥有独立的程序计数器、指令缓存与解码逻辑；不同核心可完全独立执行不同控制流。微架构层面，CPU 专门为复杂逻辑分支进行了长期演进：真实通用程序中，平均每 5-7 条指令就包含一次条件分支；若要维持性能，分支预测准确率需达到 95%+；现代 CPU 通过复杂分支预测器 (如 TAGE、感知器)、乱序执行、重排序缓冲区 (ROB) 等机制，在通用代码中可实现 1-2.5 IPC 的稳定吞吐。因此，Agent 的“行动”阶段本质是大量 if/else 判断、系统调用与串行决策，正是 CPU 微架构长期优化的主战场。

图3：不同 GPU 在不同 Warp 分支发散程度下性能表现 (单位：毫秒)



数据来源：ScienceDirect，东吴证券研究所

图4：CPU 在复杂任务编排中具备同步与调度能力



数据来源：SemiEngineering，东吴证券研究所

1.2.2. 长上下文推理亦推升 CPU 侧内存连接需求

除复杂控制流外，长上下文推理对 KV Cache 的内存承载需求，构成 Agent 场景下

推升 CPU 需求的另一条逻辑。在 Agent 场景下,长上下文推理会产生巨大的 KV cache,其占用随对话轮次与上下文长度线性增长,快速耗尽 GPU HBM 容量。业界论文如 HiFC 等指出 KV cache 已成为长上下文 LLM 推理的首要内存瓶颈,并探索通过 CPU DRAM 等来承载 KV cache。

Agent 场景的最优解正在收敛为“GPU 负责计算、CPU 以大容量 DDR5/LPDDR5 (并可通过 CXL 扩展)承载上下文与参数”的混合架构。在成本端, TrendForce 数据显示, HBM3e 单位容量价格曾是 DDR5 的 4-5 倍,即便到 2026 年价差仍在 1-2 倍。相比之下, x86/Arm 服务器 CPU 侧可通过 DDR5/LPDDR5 部署 TB 级低成本内存:例如 AMD EPYC 9004 单 socket 支持最高 6TB DDR5; Grace Hopper/GB200 则通过 LPDDR5X 提供最高 480-960GB CPU 内存,并与 GPU HBM 组成统一地址空间。我们判断,在后台 Agent 场景下,单纯追求极致 GPU 延迟并非最优解。以 GPU 承担计算、以 CPU 搭配大容量 DDR5 / LPDDR5 (并通过 CXL 扩展)承载 KV Cache 与部分参数,正在成为兼顾吞吐、扩展性与成本效率的主流架构选择。

2. Agent 软件基础设施加速落地，CPU 多核化与需求放量进入确定性通道

2.1. Agentic AI 的系统瓶颈已由 GPU 计算转向 CPU 并发调度

CPU 成为 Agentic AI 的真实瓶颈。在 Agentic AI 场景下，CPU 侧工作负载呈现爆发式增长，并已成为系统性能的核心约束。《A CPU-Centric Perspective on Agentic AI》对 Haystack RAG、Toolformer、ChemCrow、LangChain、SWE-Agent 等 Agentic workflows 进行了系统级测试。研究显示，在完整的 Agent 执行链路中，工具处理相关环节（检索、Python/Bash 执行、Web 请求等）在 CPU 上消耗的时间占端到端延迟的比例最高可达 90.6%。在高并发场景下（Batch Size 提升至 128），CPU 端到端延迟从 2.9 秒跃升至 6.3 秒以上，增幅超过 2 倍。其根本原因是并发进程数超出物理核心数后，引发严重的上下文切换与同步开销，导致现有 CPU 资源被迅速耗尽。

Agentic AI 的性能瓶颈正转向 CPU 的并发调度能力。我们认为研究结果揭示了在大量 Agentic 场景中，系统吞吐受限的并非 GPU 计算能力，而是 CPU 的核心数并发调度问题。我们由此判断，进入 Agent 时代后，CPU 侧的工具执行与调度能力将从 GPU 的附属角色，演化为需要被单独规划与优化的核心资源池。CPU 不再只是推理调度器，而是决定 Agent 系统并发上限与服务能力的关键基础设施，其配置与规模将直接约束 Agentic AI 的实际落地效率与商业化能力。

图5: Agent 任务中，CPU 占据端到端总延迟的绝大部分

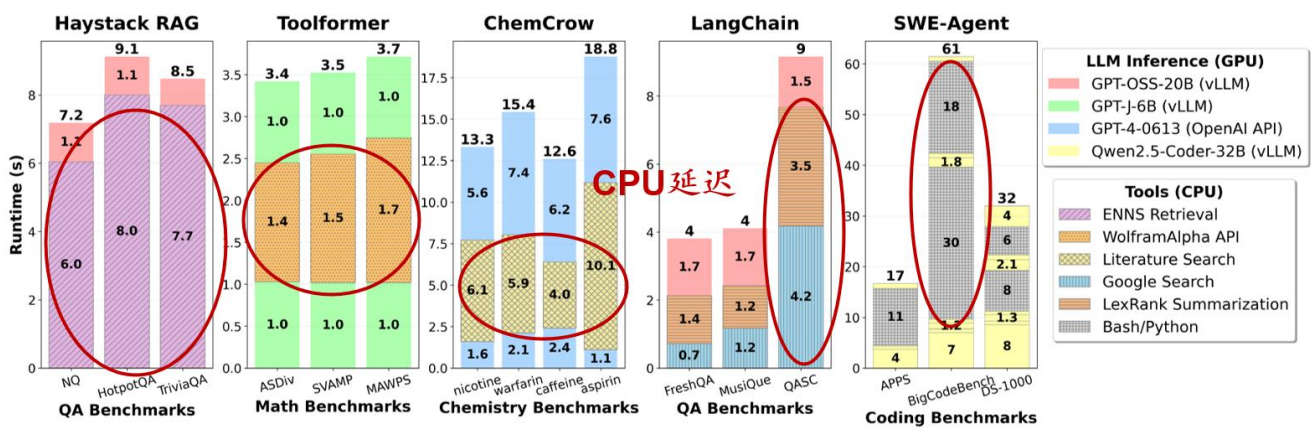


Figure 2. (a) Haystack with ENNS retrieval on QA benchmarks (b) Toolformer with WolframAlpha API on Math benchmarks (c) Chemcrow with literature (Arxiv/Pubmed) search tool on Chemistry benchmarks (d) Langchain with web search and LexRank summarization tools on QA benchmarks (e) Mini-SWE-Agent with bash/Python execution tools on coding benchmarks

数据来源：《A CPU-Centric Perspective on Agentic AI》，东吴证券研究所

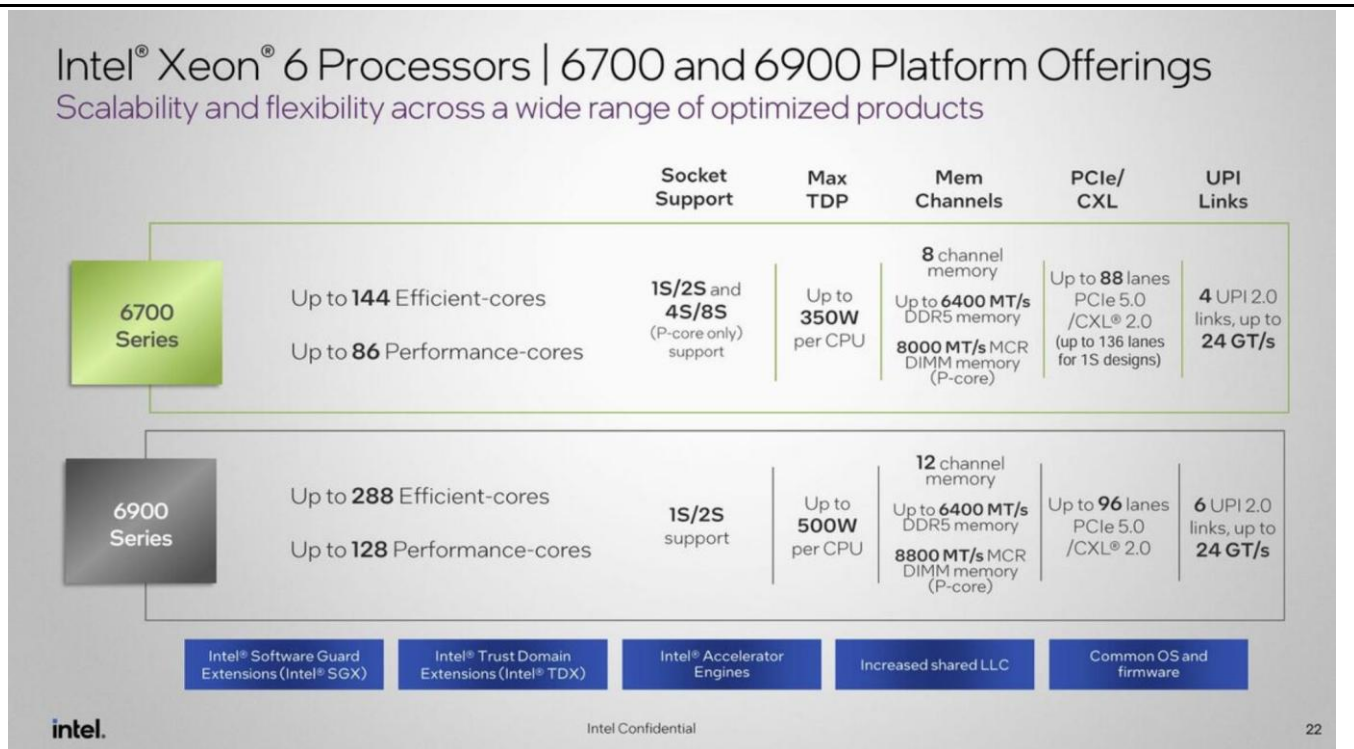
注：图中数据为各类 Agentic 工作流在不同基准任务下的端到端运行时间拆分（单位：秒），其中堆叠部分区分 CPU 侧工具执行耗时与 GPU 侧模型推理耗时

头部 CSP 正在加速建设面向 Agent 的沙盒环境软硬件基础设施。AWS 在 2025 年推出 Amazon Bedrock AgentCore Runtime / Code Interpreter，支持在安全隔离的沙盒中执行代码，并对每个 session 设定明确的 CPU 与内存资源边界，实现对 agent 执行面的托管化与资源池化管理。Google Cloud 在 2025 年明确提出 Agent Sandbox on Kubernetes 架构，用于承载 agent 的代码执行等任务，官方表述可以调配数千个沙盒，并通过 gVisor / Kata 等技术实现强隔离与规模化调度。我们认为，主流 CSP 已率先在软件层面强化 Agent Sandbox 的隔离与编排能力，通过运行时与调度体系的完善，为后续 CPU 侧基础设施规模化部署奠定基础。

2.2. CPU 龙头迭代方向明确：Agent 驱动下的超多核架构演进

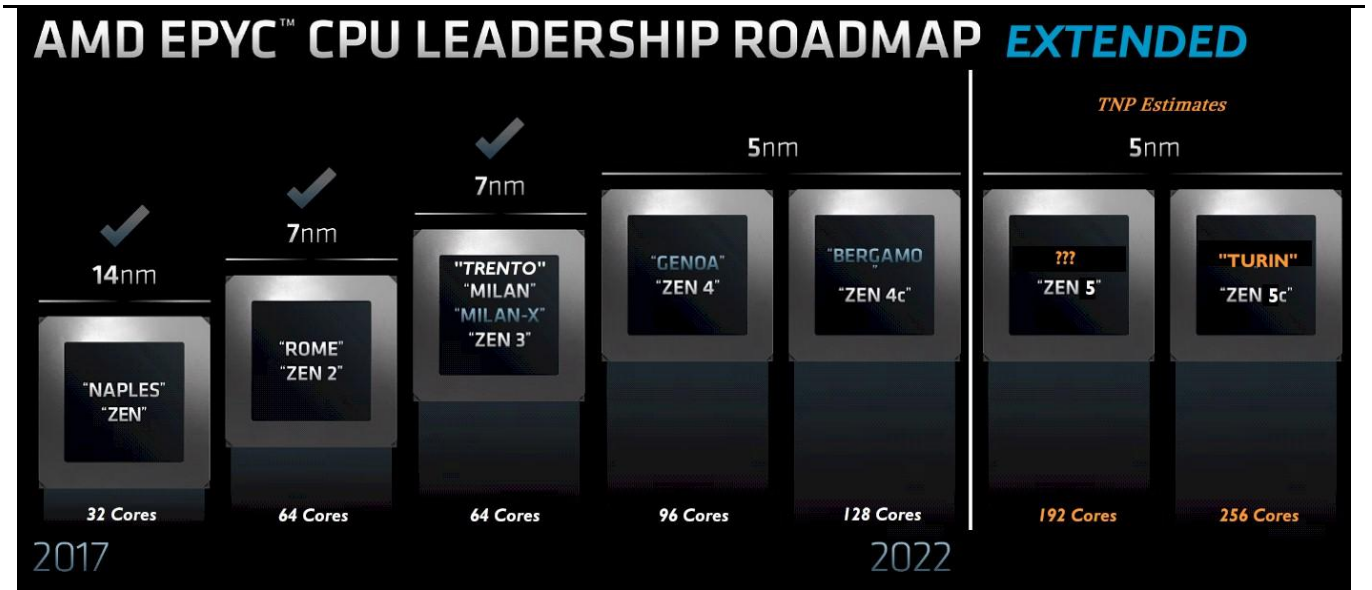
Agent 执行范式转向“长任务、强隔离”，促使服务器 CPU 向超多核架构演进。过去的互联网应用是“高并发、短连接”（如抢票、刷抖音），注重单核响应速度；Agent 时代的执行负载则呈现出高并发、长任务、重隔离的结构化变化。为同时承载成千上万的 Sandbox（沙箱/虚拟机），服务器侧更需要的是并行吞吐与单位任务成本，而非极致单核频率。因此，CPU 架构的重心正在转向“超多核、能效核（E-Core）”路线。这一趋势已在头部厂商产品规划中得到验证：AMD 推出的 Turin 最高可达 192 核，主打高密度计算与并行吞吐；Intel 的 Sierra Forest 核心数可达 144 甚至 288 核。我们认为，随着 Agent 商业化推进，厂商必须持续压低每次任务执行成本。在这一目标下，超多核 CPU 以更高的并行度、更低的单位功耗，支撑大规模、长期运行的 Agent 执行环境。

图6：Intel Xeon 6700/6900 代数据中心 CPU 的核心数最高可达 288 核



数据来源：HotHardware, Intel, 东吴证券研究所

图7: AMD 数据中心 CPU 核心数持续上行



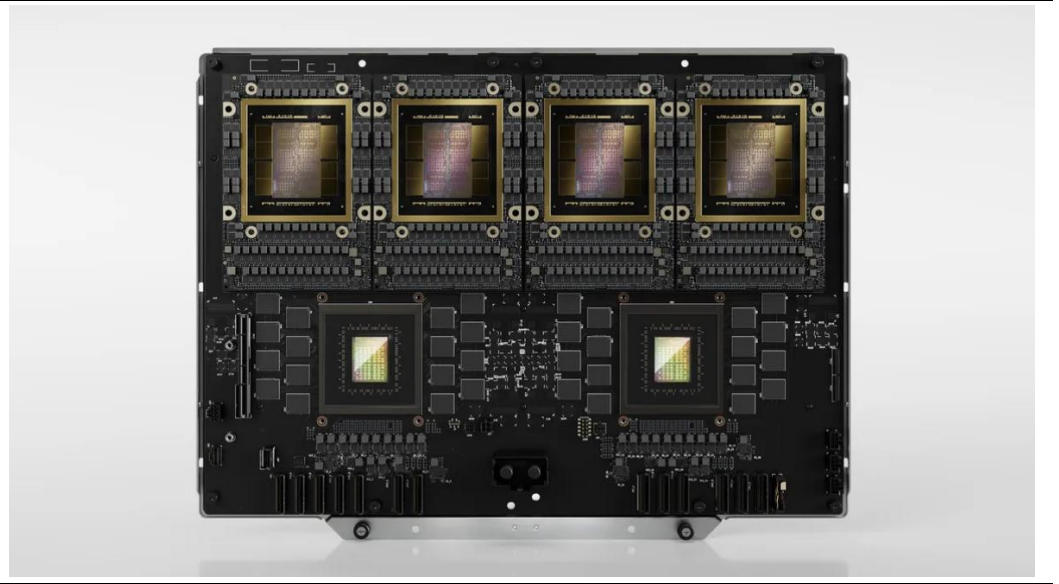
数据来源: TheNextPlatform, AMD, 东吴证券研究所

2.3. 自研 CPU 继续加码，挂钩 Agentic 的叙事

英伟达新架构提升 CPU 配比，CPU 的重要性抬升。在英伟达的 GB200 NVL72 新架构中，Grace CPU 与 Blackwell GPU 维持了 1:2 的高 CPU:GPU 配比，并被放置在统一内存体系之下。作为对比，传统 GPU 服务器架构（如 DGX H100）通常采用 2 颗 x86 CPU + 8 颗 GPU 的配置，对应约 1:4 的 CPU:GPU 配比，CPU 角色更多限于调度与外围支撑。在 GB200 架构中，Grace CPU 侧最高可配置 480GB LPDDR5X，并通过 NVLink-C2C 与 GPU 侧 HBM3e 组成统一地址空间，使 GPU 能够直接访问 CPU 内存。根据英伟达技术文档，该设计目标正是让 GPU 可以 oversubscribe 本地显存，将超出 HBM 容量的工作集放置于 CPU 内存中继续执行。

CPU 将成为 KV Cache 的核心承载体。我们认为，英伟达的架构调整直指 Agent 场景的核心痛点——上下文持续拉长，KV Cache 规模快速膨胀。在该模式下，Grace CPU 所搭载的大容量、高带宽内存实际承担了 Agent 的“短期记忆”角色。我们认为 NVIDIA 主动提升 CPU 权重，等同于在系统层面确认：在长上下文与高并发 Agent 场景中，大内存 CPU 是承载海量 KV Cache 的最优容器，而 GPU 更适合专注于计算本身。

图8: 英伟达 GB200 Grace Blackwell Superchip 结构图



数据来源: 英伟达官网, 东吴证券研究所

2.4. Engram 架构重构记忆路径, CPU 重要性进一步强化

DeepSeek 新模型架构将部分参数外置并由 CPU 调度。DeepSeek 在 2026 年论文中将部分知识重构为可显式查表的参数化记忆 (Engram 架构), 并明确由 CPU 负责调度 Engram 数据库, 从模型架构本身验证了随着 AI 系统走向长时记忆, CPU 正从“推理调度器”升级为下一代 AI 系统的核心控制与记忆管理中枢。

图9: DeepSeek 明确由 CPU 调度 Engram 数据库, CPU 重要性凸显

Table 4 | End-to-end Inference Throughput. We measure inference throughput with a 100B-parameter Engram layer entirely offloaded to host memory.

Experimental Setup		
Hardware	NVIDIA H800	
Workload	512 Sequences	
Sequence Length	Uniform(100, 1024)	
Throughput Results		
Base Model	Configuration	Throughput (tok/s)
4B-Dense	Baseline	9,031.62
	+ 100B Engram (CPU Offload)	8,858.28
8B-Dense	Baseline	6,315.52
	+ 100B Engram (CPU Offload)	6,140.02

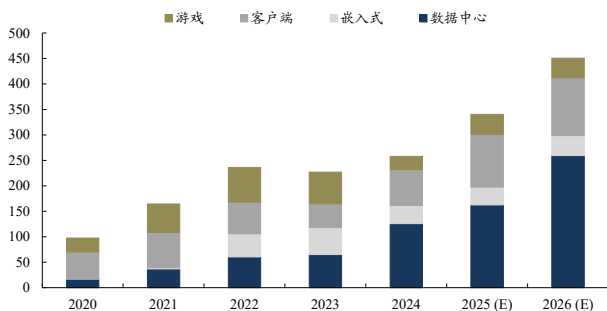
数据来源: DeepSeek, 东吴证券研究所

3. 相关公司

英特尔：AI 驱动的服务器 CPU 需求已进入供不应求阶段。在英特尔 25Q4 业绩法说会上，管理层明确表示，为 AI 芯片配套的服务器中央处理器需求出现快速飙升，公司对激增的需求措手不及，即便在产能全开情况下仍难以满足市场。管理层进一步指出，过去两个季度服务器 CPU 强劲需求超出预期；若供给更充足，当季数据中心收入将显著高于已披露水平，供给约束已成为主要掣肘。

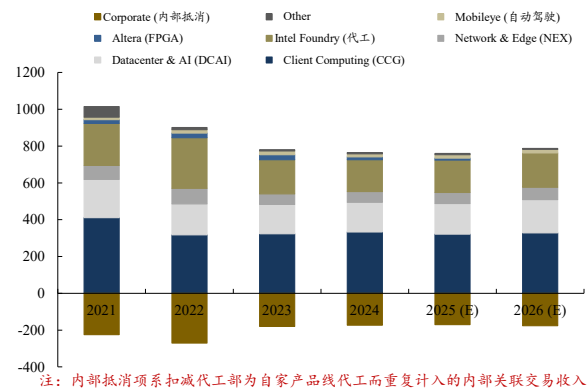
AMD：数据中心业务增长强劲。AMD 在 25Q3 业绩法说会上声称数据中心业务实现创纪录增长，核心由 EPYC 服务器 CPU 与 Instinct AI GPU 的强劲需求共同驱动。管理层多次强调第五代 EPYC (Turin) 正在快速爬坡，市场需求“非常强”，作为公司 Helios / MI400 / MI450 机架级 AI 方案中的关键通用计算与内存承载部件。

图10：AMD 营收拆分及盈利预测（亿美元）



数据来源：Bloomberg，东吴证券研究所
注：盈利预测采用 Bloomberg 一致预期

图11：Intel 营收拆分及盈利预测（亿美元）



数据来源：Bloomberg，东吴证券研究所
注：盈利预测采用 Bloomberg 一致预期

海光信息：响应“高并发、重隔离”的多核趋势。以海光为代表，其产品路线从早期 32/64 核架构，快速演进，并在在研的海光 C86-5G 上迈入 128 核/512 线程时代，同时同步升级 DDR5 与 CXL 2.0。我们认为这一演进路径并非单纯追求单核性能，而是明显围绕多任务并发、虚拟化密度与单位任务成本进行设计，反映出国产 CPU 已准确对齐 Agent 时代“高并发、长任务、强隔离”所要求的超多核、高内存带宽算力范式。

图12: 海光信息 CPU 代际梳理



数据来源: 芯智讯, 东吴证券研究所

澜起科技: CPU 平台迭代+带宽需求上行, 带动内存子系统升级。公司于25年1月推出了第二子代 MRCD/MDB 芯片, 并成功向全球主要内存厂商送样。基于产品性能、产业生态等因素的考虑, 行业普遍更看好第二子代 MRDIMM 渗透率的提升, 从公司的在手订单来看, 这一点也得到印证。截至2025年10月27日, 公司预计在未来六个月内交付的 DDR5 第二子代 MRCD/MDB 芯片在手订单金额已超过人民币1.4亿元。

由于第二子代 MRDIMM 的数据传输速率达到 12800MT/s, 相比第一子代 MRDIMM (支持速率 8800MT/s) 提升 45%, 是第三子代 RDIMM (支持速率 6400MT/s) 的两倍, 这将大幅提升系统性能, 在对内存带宽有较大需求的工作负载下, MRDIMM 有望成为这些应用系统主内存的优选方案; 同时, 业内将有更多的服务器 CPU 平台支持第二子代 MRDIMM, 也包括一些 ARM 架构的 CPU 平台, 有利于 MRDIMM 生态的进一步完善。这些因素将共同推动 MRDIMM 行业渗透率的提升以及 MRCD/MDB 芯片需求的增长。作为 MDB 芯片国际标准的牵头制定者, 公司引领 MRCD/MDB 芯片技术的创新并保持行业领先地位。

4. 风险提示

Agent 落地节奏不及预期风险：Agent 商业化仍依赖应用成熟度、客户使用习惯及配套软件生态完善程度，若实际落地进展慢于市场预期，相关算力需求释放节奏可能延后，从而影响公司产品放量与收入兑现。

新型硬件架构替代风险：随着算力体系持续演进，未来不排除出现针对 Agent 场景高度定制化的新型加速器或异构架构，在部分功能或成本维度上对通用 CPU 形成替代，对现有产品竞争格局造成不确定影响。

CSP 资本开支波动风险：相关公司下游需求高度关联云厂商资本开支周期，若宏观环境变化或云厂商阶段性调整投资节奏，可能导致订单确认与交付进度波动，进而对公司业绩产生影响。

免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中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投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

- 买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；
- 增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；
- 中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；
- 减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；
- 卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

- 增持：预期未来 6 个月内，行业指数相对强于基准 5%以上；
- 中性：预期未来 6 个月内，行业指数相对基准-5%与 5%；
- 减持：预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号
邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>