

电子

AI 算力需求持续释放，重点看好 AI 服务器产业链

证券研究报告

2023 年 06 月 14 日

投资评级

行业评级

强于大市(维持评级)

上次评级

强于大市

作者

潘暕

分析师

SAC 执业证书编号: S1110517070005

panjian@tfzq.com

俞文静

分析师

SAC 执业证书编号: S1110521070003

yuwenjing@tfzq.com

行业走势图



资料来源: 聚源数据

相关报告

- 1 《电子行业点评: 苹果 MR 发布在即, 重点推荐相关产业链》 2023-05-31
- 2 《电子行业深度研究: 电子行业 23Q1 总结: 有望进入复苏周期》 2023-05-21
- 3 《电子行业深度研究: AI 赋能制造业道路, 传统安防龙头估值逻辑切换》 2023-04-09

AI 服务器直接受益于需求端增长, Nvidia 高速互联助力 AI 运算。2018 年至今, 模型升级带来参数量指数级增加, AI 训练+推理模型需求催生 AI 服务器海量需求 (训练阶段算力需求=6×模型参数数量×训练集规模), 按照 A100 640GB 服务器参数假设, 据测算, 训练端需要服务器数量为 3423 台/日, 推理端按照 1 亿活跃用户测算对应成本为 4000 万美元, 并随着用户量和访问互动量上升, 算力需求持续提升。此外, Nvidia 高速互联助力 AI 运算, 800G 交换机陆续发布, 下一代超宽互联蓄势待发。

AI 重塑服务器行业格局, ODM 模式/厂商受益于 AI 服务器增长。根据 Statista 数据, 2021 年全球服务器市场规模达到 831.7 亿美元, 同比增长 6.97%, 其中 AI 服务器市场达到 156.3 亿美元, 同比增长 39.1%。AI 服务器有望成为增速最快的细分板块, 预计 AI 服务器市场将在 2026 年达到 347.1 亿美元, 5 年 CAGR 达到 17.3%。ODM 模式/厂商受益于 CSP 客户增量, 2022 年出货的服务器中 ODM 生产的市占率或达 50%。全球 ODM 厂商竞争格局集中度高, CR5 高达 94.4%, 分别为鸿海/工业富联 (43%)、广达 (17%)、纬创 (14%)、英业达 (12.8%) 和超微 (7.6%)。

AI 服务器放量预期利好上游核心部件。AI 服务器销量增加将拓宽上游核心零部件的增量市场, 尤其是 AI 芯片 (GPU、ASIC 和 FPGA)、存储器、固态硬盘等。其中, 国内上游环节的国产替代程度不一。我们认为国内处于乐观机遇期的有 AI 服务器制造厂商、服务器用 PCB 厂商、DRAM 厂商和散热厂商, PCB 和 DRAM 的服务器领域均是行业实现扩产和开拓市场的重点所在; 制约瓶颈为人工智能芯片, 仍被国外厂商垄断, 有望实现突破的环节为先进封装 Chiplet, 或成为我国算力困境的关键转折点。

投资建议: 建议关注 AI 服务器及上游产业链相关标的: 1)AI 服务器龙头: 工业富联; 2)服务器 PCB: 鹏鼎控股; 3)服务器线束与连接器: 电连技术、兆龙互连; 4)算力芯片: 寒武纪、海光信息 (天风计算机团队覆盖)、景嘉微 (天风计算机团队联合覆盖); 5)存储供应链: 兆易创新、北京君正、江波龙 (天风计算机团队联合覆盖)、澜起科技、雅克科技、鼎龙股份 (天风化工团队联合覆盖)、华懋科技 (天风汽车团队联合覆盖)、华特气体; 6)边缘 AI: 瑞芯微、晶晨股份、全志科技、恒玄科技、富瀚微、中科蓝讯、乐鑫科技; 7)AI to B/机器视觉: 大华股份、海康威视、鼎捷软件 (天风计算机团队覆盖)、凌云光、天准科技、舜宇光学、海康威视、奥普特 (天风机械军工团队覆盖); 8)Chiplet: 长电科技、通富微电、华天科技、长川科技 (天风机械团队覆盖)、华峰测控 (天风机械团队覆盖)、利扬芯片、芯碁微装、伟测科技

风险提示: 中美贸易摩擦导致上游原材料断供、AI 服务器出货不及预期、技术瓶颈仍未摆脱

内容目录

1. 人工智能服务器为算力支撑，助燃 AI 产业化.....	4
1.1. 从通用服务器到 AI 服务器的过渡	4
1.2. AI 服务器与普通服务器相比有着明显的性能优势	4
1.3. AI 服务器的进步源于芯片，未来将关注更多功能性	5
2. 庞大算力需求是 AI 服务器未来放量的关键驱动力.....	6
2.1. AI 模型优化迭代引发算力缺口，未来推理功能将需更多算力	6
2.2. AI 服务器作为承载算力主体将受益，成服务器市场的增长主力	8
2.3. Nvidia 高速互联助力 AI 运算，多 GPU 通信成为关键技术	11
3. AI 服务器放量预期利好上游核心部件，挑战机遇共存	19
3.1. 危机四伏：上游供应危机尚未解除，国产替代必需提上日程.....	19
3.2. 柳暗花明：以 Chiplet 工艺打开我国对算力的想象空间	22
3.3. 适逢其会：为上游部分元件打开增量空间.....	23
3.3.1. AI 服务器或将打开 PCB 增量市场.....	23
3.3.2. 高算力使服务器芯片散热成为难题，散热模组应需求提高散热效率.....	24
3.3.3. 数据存力作为算力进阶需求迭起，看好国产存储器后续发展.....	24
3.4. 下游指明服务器发展方向，人工智能渗透率提高将扩展 AI 服务器应用	25
4. 投资建议.....	27
5. 风险提示.....	28

图表目录

图 1：AI 服务器的 CPU+架构.....	4
图 2：中国智能算力规模.....	7
图 3：全球服务器市场规模及增速	8
图 4：中国服务器市场规模及增速	8
图 5：全球 AI 服务器市场规模及增速.....	8
图 6：中国 AI 服务器市场规模及增速.....	8
图 7：ODM 和品牌服务器厂商市占率.....	9
图 8：预计 2022 年 ODM 厂商竞争格局	9
图 9：2021 年 H1 全球 AI 服务器竞争格局	9
图 10：2021 年 H1 中国 AI 服务器竞争格局.....	9
图 11：2021 年全球服务器竞争格局	10
图 12：2021 年中国服务器竞争格局	10
图 13：GPT 模型参数对应 GPU 数量与 DGX A100 POD 配置.....	11
图 14：NVIDIA DGX Systems	12
图 15：NVIDIA DGX A100 配置	12
图 16：NVIDIA DGX H100 配置	13
图 17：NVIDIA DGX H100（H100、NVLink、NVSwitch 配套）	13

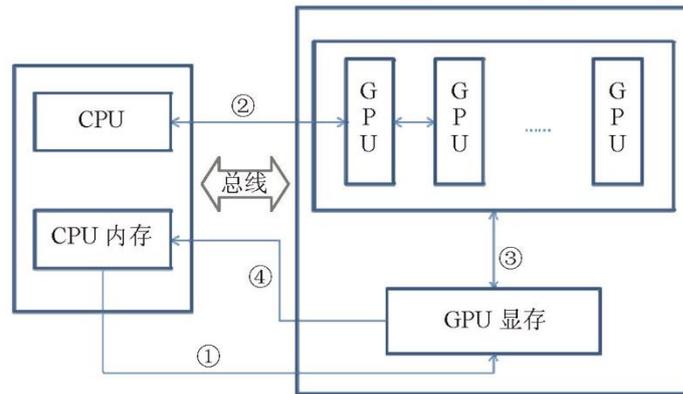
图 18: DGX A100 与 DGX H100 32 节点 256 GPU NVIDIA SuperPOD 架构比较.....	14
图 19: PCIe 与 NVLink 架构比较.....	14
图 20: NVIDIA NVLink 迭代规格.....	14
图 21: 以 P100 为例, NVLink 组成架构.....	15
图 22: NVIDIA NVSwitch 参数.....	15
图 23: NVSwitch 拓扑图 (以 16 个 GPU 为例)	16
图 24: NVSwitch 芯片.....	16
图 25: NVIDIA Mellanox 400G InfiniBand 组成.....	16
图 26: NVIDIA ConnectX-7 智能主通道配接器 (HCA)示意图.....	16
图 27: NVIDIA Quantum-2 QM9700 Series 示意图.....	17
图 28: 光模块工作原理.....	17
图 29: 光模块成本结构.....	17
图 30: 思科第一款 800G Nexus.....	18
图 31: 相关产业链.....	19
图 32: 2018 年 IDC 公布的服务器成本拆解.....	19
图 33: 中国 AI 芯片市场规模 (亿元)	20
图 34: 2019/2020 年全球服务器应用场景.....	26
图 35: 2021 年我国服务器下游应用场景.....	26
图 36: 主要的 AI 服务器采购商 2019-2022 年 Capex (亿元)	27
表 1: 主流服务器类型 (不完全统计)	4
表 2: AI 服务器与普通服务器相比具有更好的技术优势.....	5
表 3: AI 服务器的技术迭代核心在于硬件架构.....	5
表 4: AI 服务器未来发展方向.....	5
表 5: 2018-2020 年模型升级带来的参数量激增.....	6
表 6: 训练所需服务器测算.....	6
表 7: 以 ChatGPT 产品测算所需算力.....	7
表 8: AI 服务器厂商.....	10
表 9: 光芯片可分为激光器芯片和探测器芯片.....	18
表 10: 主流 AI 芯片的对比.....	20
表 11: 国内 AI 服务器所搭载的 GPU 厂商主要以英伟达为主.....	21
表 12: 英伟达 A100 与 A800GPU 性能相差不大.....	21
表 13: 国内生产供应服务器的 GPU 厂商.....	22
表 14: 璧仞科技的 BR100、寒武纪的思元 370 与 NVIDIA A100 性能比较.....	23
表 15: Chiplet 实现量产的国内公司.....	23
表 16: PCB 领域的国内厂商.....	23
表 17: 散热模组领域的国内厂商.....	24
表 18: 国内企业的存储器产品.....	25
表 19: 提供企业级 SSD 的企业.....	25
表 20: 相关公司盈利预测与估值.....	27

1. 人工智能服务器为算力支撑，助燃 AI 产业化

1.1. 从通用服务器到 AI 服务器的过渡

AI 服务器在众多服务器中脱颖而出源于其架构优势。AI 服务器是承载智慧计算中 AI 计算的核心基础设施，是一种能够提供人工智能的数据服务器，既可以用于支持本地应用程序和网页，也可以为云和本地服务器提供复杂的 AI 模型和服务，通过异构形式适应不同应用范围以及提升服务器的数据处理能力，异构方式包括 CPU+GPU/TPU/ASIC/FPGA。

图 1: AI 服务器的 CPU+架构



资料来源:《人工智能服务器技术研究》王峰, 天风证券研究所

AI 服务器的发展脱胎自通用服务器的性能提升需求。复盘主流服务器的发展历程，随着数据量激增、数据场景复杂化，诞生了适用于不同场景的服务器类型：通用服务器、云计算服务器、边缘计算服务器、AI 服务器。随着大数据、云计算、人工智能及物联网等网络技术的普及，充斥在互联网中的数据呈现几何倍数的增长，使得以 CPU 为主要算力来源的传统服务器承受着越来越大的压力，并且对于目前 CPU 的制程工艺而言，单个 CPU 的核心数已经接近极限，但数据的增加却还在继续，因此服务器数据处理能力必须得到新的提升，在这种环境下，AI 服务器应运而生。面对 ChatGPT 所引出的大规模预训练模型，AI 服务器以其架构优势带来的大吞吐量特点，有望在一众服务器中脱颖而出。

表 1: 主流服务器类型（不完全统计）

	特点	配置	应用场景
通用服务器	物理服务器，独立存在，拥有完全管理员权限和独立 IP 地址，安全稳定性高。	CPU、硬盘、内存等。	
云计算服务器	通过虚拟化技术，将一台/多台服务器虚拟化成一个个可以切分的资源池，客户按需灵活配置与扩展，管理便捷，费用相对低廉。	按客户需求配置 CPU、内存、数据盘等。	适合对业务弹性扩展需求和易用性的需求：电商、IT 行业、教育、移动应用、游戏等。
边缘计算服务器	承担 70% 以上的计算任务，需支持 ARM/GPU/NPU 等异构计算，针对不同业务场景开发，远程控制运维。		工业互联网、车联网、医疗保健、AR/VR、智慧城市等。
AI 服务器	采用异构形式服务器，承担大量计算；大规模并行运算、多重向量/张量运算、计算效率高。	GPU/FPGA/ASIC 等加速芯片、CPU、内存等。	融合深度学习、机器视觉、知识图谱等人工智能技术的应用：医疗影像智能分析、人脸/语音/指纹识别、安防监控场景等。

资料来源: 人工智能与创新公众号、皖云数科公众号、物联网世界官网、高升数据公众号、机智云物联网公众号、天风证券研究所

1.2. AI 服务器与普通服务器相比有着明显的性能优势

AI 服务器相较于传统服务器算力上大幅跃升。AI 服务器利用 CPU+的架构模式，CPU 仍作为 CPU 的数据处理主要模块，同时植入并行式计算加速部件，如 ASIC、FPGA、GPU 等，负责人工智能计算负载加速。总而言之，在 CPU+架构下，AI 服务器的技术选型和部件配置针对不同的业务场景做相应的调整优化，通过合理的负载分担实现计算能力的提升。

表 2: AI 服务器与普通服务器相比具有更好的技术优势

	AI 服务器	普通服务器
卡的数量	以加速卡为主导，基础要求为四块以上的 GPU 卡，甚至需要搭建外部服务器作为支持。	以 CPU 为主导，单卡/双卡 CPU。
P2P 通讯	GPU 卡间需要大量的参数通信，模型越复杂，通信量越大： SXM3 协议下，P2P 带宽高值 300GB/s； SXM2 协议下，P2P 带宽高值 50GB/s； PCI3.0 协议下，P2P 带宽高值 32GB/s。	普通 GPU 服务器一般只要求单卡性能。
特有设计	全面考虑对存储、通信、网络等相关领域的技术方案进行合理配置，使之与计算部件的计算能力相匹配，避免出现性能瓶颈。	-
专有技术	Purley 平台更大内存带宽； NVlink 提供更大的互联带宽； TensorCore 提供更强的 AI 算力。	-

资料来源：南京锟前官网，天风证券研究所

1.3. AI 服务器的进步源于芯片，未来将关注更多功能性

我们认为 AI 服务器的技术迭代取决于硬件中 AI 芯片的选择。传统普通服务器数据处理核心单一，以 CPU 为主，AI 服务器则采取 CPU+的异构方式完成数据处理能力提升。参考英伟达 HGX 的产品迭代路径，核心差异在于搭载的 GPU，HGX-1、HGX-2 和 HGX A100 分别对应 Tesla P100 GPU、Tesla V100 和 A100，在深度学习和高性能计算领域都展现出更高的性能。因此，我们认为可以从 AI 芯片发展窥见 AI 服务器的技术迭代历程，主要以 CPU+GPU 为主的 AI 服务器，结合云服务趋势发展至 CPU+GPU/ASIC/FPGA 的多功能+优性能+灵活性+高性价比 AI 服务器。

表 3: AI 服务器的技术迭代核心在于硬件架构

阶段性进展	硬件架构发展路径
汇聚主流 2010-2014	2010 年后随着云计算、大数据为代表的新一代信息技术高速发展并逐渐开始普及，云端采用“CPU+GPU”混合计算模式，推动人工智能算法的演进和人工智能芯片的广泛使用。
百花齐放 2015-2020	随着以深度学习为核心的人工智能技术得到全球范围内的关注，业界对人工智能算力的要求越来越高。更多高性能的定制化芯片开始出现，部署不同场景的服务器中，2015 年 Google 推出基于 ASIC 架构的 TPU 用于云端训练和推理，2017 年微软发布基于 FPGA 芯片组建的 Project Brainwave 低时延深度学习系统，2018 年亚马逊发布高性能推理芯片 AWS Inferentia 等。
融合发展 2021-至今	上云需求+数据规模膨胀，云原生+人工智能服务器逐渐成为一种趋势，既能够应对人工智能计算密集型工作，也能兼顾高弹性、高敏捷和高性价比的优势。

资料来源：《人工智能芯片产业技术发展研究》商惠敏，澎湃，驱动科技公众号，Kyligence 公众号，天风证券研究所

结合整个人工智能技术和服的发展，未来人工智能服务器将重点发展软硬件平台融合、低功耗设计、智能边缘计算等领域。特别是随着量子计算、类脑芯片等新一代人工智能计算加速技术的兴起，人工智能服务器的设计与实现可能会产生颠覆性变化。总之，人工智能服务器作为提供计算能力的核心要素，都是人工智能研发应用体系中不可或缺的组成部分，它的良性发展势必会为整个人工智能产业的成长奠定坚实的基础。

表 4: AI 服务器未来发展方向

发展趋势	主要内容
软硬件融合	通过更多考虑软硬件协同，以系统应用为导向驱动芯片等硬件设计，让用户得到更好的体验。
低功耗发展	在节能减排要求下，重点探索液冷技术以实现更高的散热效率和更少的电能使用。
推理工作负载	伴随企业人工智能应用成熟度逐步递增，企业将人工智能训练负载转移到推理工作负载，这意味着人工智能模型将逐步进入广泛投产模式。

资料来源：半导体行业观察公众号，《2022-2023 中国人工智能算力发展评估报告》IDC 与浪潮信息，天风证券研究所

2. 庞大算力需求是 AI 服务器未来放量的关键驱动力

2.1. AI 模型优化迭代引发算力缺口，未来推理功能将需更多算力

算力是人工智能研究的生产力。在当今以深度学习为中心的人工智能发展中，AI 模型的进步主要依赖于模型的规模化扩展。在 AlexNet 网络模型出现后，ResNet、Transformer、BERT 等优秀模型也在不断涌现，尤其在图像、语音、机器翻译、自然语言处理等领域带来了跨越式提升。AI 模型智能程度不断发展的同时，AI 模型的数据量、结构的复杂程度也在不断增加，其带来了模型的参数量增长，模型尺寸呈指数级增加。随着模型尺寸不断膨胀，实现高效 AI 模型训练的重要支撑即更快的算力，即在短时间内完成大规模 AI 计算。

表 5：2018-2020 年模型升级带来的参数量激增

时间	公司	模型	参数量（亿）
2018	Open AI	GPT-1	1.5
2018	Google	Bert-Large	3.4
2019	微软	MT-DNN	3.3
2019	Open AI	GPT-2	15.42
2019	阿里巴巴	PERSEUS-BERT	1.1
2019	NVIDIA	Pmojert Megatron	83
2019	Facebook	RoBERTa	3.35
2019	Facebook	XILM	6.65
2019	NVIDIA	Megaltron-Scaled Version of OpenAI GPT-2	83
2020	微软	NLG	172
2020	Open AI	GPT-3	1750

资料来源：ittbank 公众号，天风证券研究所

AI 训练模型的需求或将释放 AI 服务器海量产能。根据 OpenAI 在 2020 年发表的论文，训练阶段算力需求与模型参数数量、训练数据集规模等有关，且为两者乘积的 6 倍：训练阶段算力需求=6×模型参数数量×训练集规模。

GPT-3 模型参数约 1750 亿个，预训练数据量为 45 TB，折合成训练集约为 3000 亿 tokens。即训练阶段算力需求=6×1.75×10¹¹×3×10¹¹=3.15×10²³ FLOPS=3.15×10⁹ PFLOPS

依据谷歌论文，OpenAI 公司训练 GPT-3 采用英伟达 V100 GPU，有效算力比率为 21.3%。GPT-3 的实际算力需求应为 1.48×10⁹ PFLOPS(17117 PFLOPS-day)。

假设应用 A100 640GB 服务器进行训练，该服务器 AI 算力性能为 5 PFLOPS，最大功率为 6.5 kw，则我们测算训练阶段需要服务器数量=训练阶段算力需求÷服务器 AI 算力性能=2.96×10⁸ 台（同时工作 1 秒），即 3423 台服务器工作 1 日。

表 6：训练所需服务器测算

项目	值
----	---

参数量	1750 亿
预训练数据量	45TB
算力需求	3.15×10^8 PFLOPS
有效算力比率	21.3%
实际算力需求	1.48×10^9 PFLOPS
A100 服务器算力性能	5 PFLOPS
工作 1 天所需服务器 (台)	3423

资料来源：英伟达公司官网，天翼智库微信公众号，天风证券研究所

后期推理侧算力需求将成为主力。据 IDC 数据，2021 年中国人工智能服务器工作负载中，57.6% 的负载用于推理，42.4% 用于模型训练，预计至 2026 年 AI 推理的负载比例将进一步提升至 62.2%。具体落到 ChatGPT 的算力需求场景中，可以将预训练、Finetune 归类为训练，日常运营归类为推理。ChatGPT 的更新迭代已展示出对算力的高需求：从训练端来看，GPT1-3 经历了从 1.5 亿到 1750 亿参数的增长过程，预计需要 3640 PFlop/s-day；从推理端来看，据 SimilarWeb 数据，2023 年 1 月 ChatGPT 官网日访问量突破千万级别，1 月 31 日达到 2800 万访问量，以 3.4% 的日增长速度持续扩展，其中独立访客数量为 1570 万。据 CNBC，达到 1 亿级别活跃用户将产生成本约 4000 万美元，则该 1570 万用户所耗费互动成本约为 628 万美元，推算得相对算力处理量为 5714.8 PFlop/s-day，随着新用户不断进入并多次访问，该算力需求将不断推高，是未来算力的重点需求端。

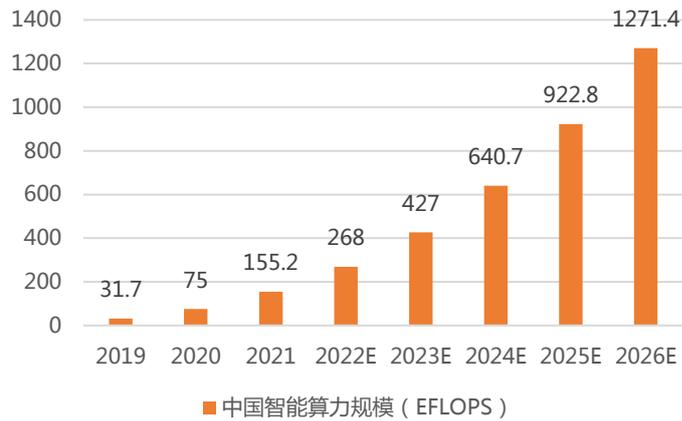
表 7：以 ChatGPT 产品测算所需算力

应用场景	ChatGPT 具体应用	功能	算力要求 (PFlop/s-day) (以 GPT-3 模型为例)
训练	预训练	通过大量无标准的纯文本数据，训练模型基础语言能力，得到 GPT-1/2/3 基础大模型。	3640
	Finetune	在预训练的大模型基础上，进行监督学习、强化学习等，实现对模型参数数量的优化调整。	-
推理	日常运营	基于用户的输入进行推理计算，并输出反馈结果。	5714.8

资料来源：《2022-2023 中国人工智能算力发展评估报告》IDC 与浪潮信息，Similarweb，ittbank 公众号，蓝海大脑官网，天风证券研究所

国内大厂在 AI 元年坚定入局+国内算力基建化，或将成为国内 AI 服务器需求来源。在 ChatGPT 带动的浪潮中，国内大厂开始密切发布类 ChatGPT 产品，3 月 16 日百度正式推出国内首款生成式 AI 产品“文心一言”、3 月 30 日腾讯推出“混元”AI 大模型、4 月 7 日阿里的“通义千语”开启内测邀请，华为 4 月 8 日推出盘古大模型等。据 IDC 数据，2021 年中国智能算力规模为 155.2 EFLOPS，随着 AI 模型日益复杂、计算数据量快速增加、人工智能应用场景不断深化，未来国内智能算力规模将持续增长，预计 2026 年智能算力规模将达 1271.4 EFLOPS，5 年 CAGR 达到 52.3%。对于国内人工智能算力需求的高增长，我国正牵头东数西算、智能计算中心，通过算力基础设施完成从点到网的升级，据 IDC 观点，中国市场的人工智能硬件支出占人工智能总支出的比例将在未来 5 年保持在 65% 左右，其中服务器是硬件中的重要部分。我们认为，AI 服务器作为智能算力运算的主要载体，看好未来持续放量的高预期走势。

图 2：中国智能算力规模



资料来源:《2022-2023 中国人工智能算力发展评估报告》IDC 与浪潮信息, 贵州省大数据发展管理局, 天风证券研究所

2.2. AI 服务器作为承载算力主体将受益, 成服务器市场的增长主力

AI 服务器市场扩张表现亮眼, 中国市场领跑全球。根据 Statista 数据, 2021 年全球服务器市场规模达到 831.7 亿美元, 同比增长 6.97%, 其中 AI 服务器市场达到 156.3 亿美元, 同比增长 39.1%, 在整体服务器市场中占比 18.79%, 同比提升 4.34pct, 这主要系随着人工智能所需算力扩大, AI 服务器作为新型算力基础设施的主体将直接影响 AI 创新迭代和产业落地, 我们认为市场将需更多、更强算力的 AI 服务器作为核心解决方案, 据 IDC 与浪潮信息, 预计 AI 服务器市场将在 2026 年达到 347.1 亿美元, 5 年 CAGR 达到 17.3%。再看中国市场, 2021 年国内服务器市场规模达到 250.9 亿美元, 同比增长 15.9%, 高于全球增长速度; 其中 AI 服务器市场达到 59.2 亿美元, 同比增长 68.2%, 在国内服务器市场中占比 23.6%, 这主要得益于国内人工智能应用的加速落地, 浪潮信息、新华三、宁畅等厂商助推人工智能基础设施产品的优化更新, IDC 调研显示, 超过 80% 的中国厂商表示在未来将增加人工智能服务器的投资规模, 预计在 2026 年中国 AI 服务器市场规模将达到 123.4 亿美元, 5 年 CAGR 为 15.82%。

图 3: 全球服务器市场规模及增速



资料来源: Statista, 天风证券研究所

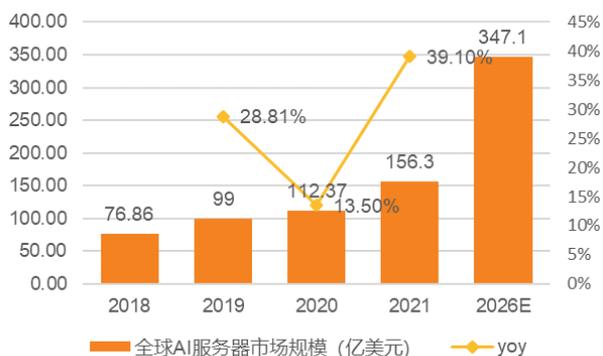
图 5: 全球 AI 服务器市场规模及增速

图 4: 中国服务器市场规模及增速

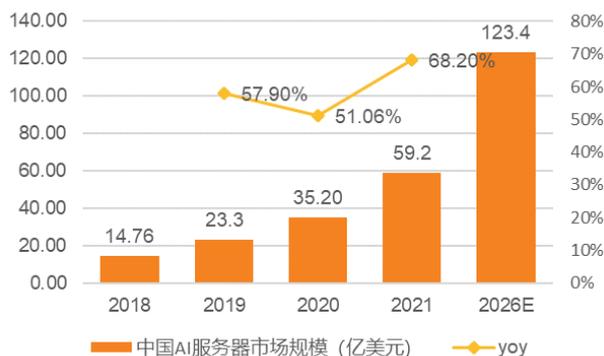


资料来源: 中商产业研究所公众号, 天风证券研究所

图 6: 中国 AI 服务器市场规模及增速



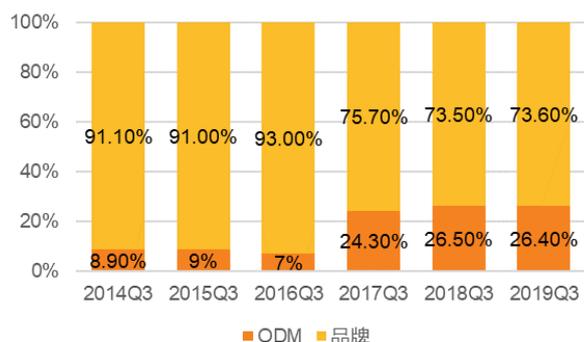
资料来源：量子位公众号，《2022-2023 中国人工智能算力发展评估报告》
IDC 与浪潮信息，天风证券研究所



资料来源：量子位公众号，《2022-2023 中国人工智能算力发展评估报告》IDC
与浪潮信息，天风证券研究所

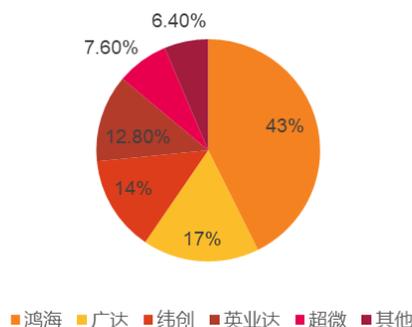
全球服务器行业生产模式趋向白牌化，ODM 厂商市占率集中。服务器厂商可分为 ODM 厂商和品牌厂商两种类型。品牌服务器厂商以浪潮、华为、新华三为代表，ODM 厂商以超微、广达为代表，根据品牌服务器厂商的委托完成硬件生产，加贴委托方商标并交付给品牌持有者进行销售，近年不少客户会绕过品牌商向 ODM 厂商直接订购服务器成品，白牌服务器生产模式的兴起对传统品牌服务器厂商造成一定冲击。在下游业务业务+成本+时间的需求考虑下，Facebook 在 2011 年率先主导成立 OCP 联盟，向内部成员统一硬件标准与硬件开源，成员中的白牌厂商开始获得服务器设计方案，为白牌厂商绕过品牌厂商提供技术支持，这将进一步推高服务器白牌厂商的市占率。据重磅数据，ODM 市场份额从 2014Q3 的 6.6% 快速攀升至 2019Q3 的 26.4%，根据 2022 年全球服务器销售量（848.7 亿美元）与全球服务器代工市值（4557 亿元）测算，2022 年出货的服务器中 ODM 生产市占率或达 50%，我们认为这主要得益 CSP 客户（如微软、谷歌）的发展，服务器需求随着业务成长而持续增长，采购量不断增加。根据 Digitimes Research，全球 ODM 厂商竞争格局集中度高，CR5 高达 94.4%，分别为鸿海（43%）、广达（17%）、纬创（14%）、英业达（12.8%）和超微（7.6%），主要客户涵盖惠普、戴尔、联想、亚马逊、微软、谷歌等国际巨头。随着超大型资料中心需求强劲+下游云计算领域 Capex 扩大，造就白牌厂商的强劲发展势头。

图 7: ODM 和品牌服务器厂商市占率



资料来源：重磅数据，天风证券研究所

图 8: 预计 2022 年 ODM 厂商竞争格局

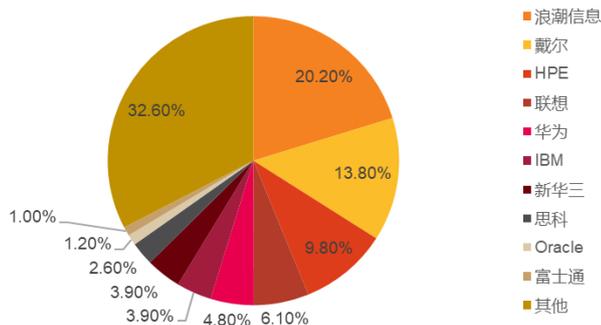


资料来源：芯智讯，天风证券研究所

国内服务器品牌厂商在 AI 服务器市场占优，未来有望放大优势。从全球 AI 服务器市场格局来看，2021 年上半年数据显示，中国厂商在 TOP5 厂商中占据过半席位（浪潮、联想、华为），累计占比为 31.1%，其中浪潮信息以 20% 的市占率占据榜首。从国内 AI 服务器市场竞争格局来看，市场集中度较高，浪潮信息占据近半市场，CR5 超过 80%。未来，随着国产厂商在 AI 服务器的持续深耕，有望在既有市场优势中吸收更大来自 AI 产业的发展机遇。对比全球/中国 AI 服务器竞争格局，我们认为国产服务器厂商在 AI 服务器市场优势较为明显，未来有望利用既有市场优势+推出具有竞争力的 AI 服务器进一步拓宽市场份额，吸收来自 AI 产业的发展机遇，有望以技术优势打造盈利护城河，在国际市场上争取更高话语权。

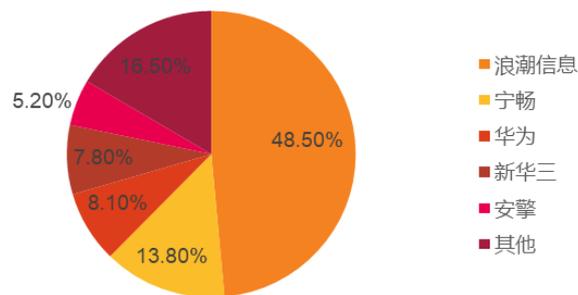
图 9: 2021 年 H1 全球 AI 服务器竞争格局

图 10: 2021 年 H1 中国 AI 服务器竞争格局



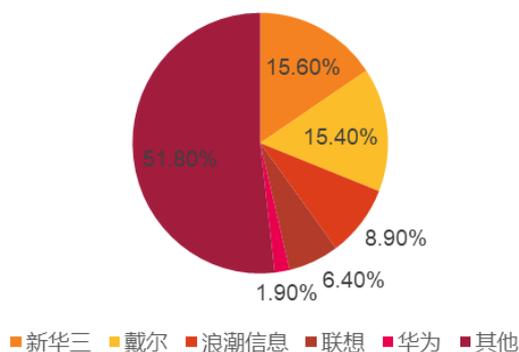
资料来源：机器人技术与应用公众号，天风证券研究所

图 11：2021 年全球服务器竞争格局

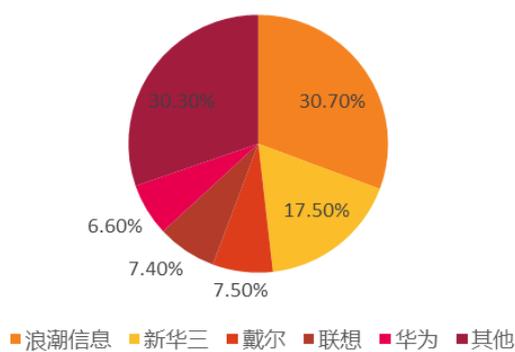


资料来源：IDC 公众号，天风证券研究所

图 12：2021 年中国服务器竞争格局



资料来源：智研咨询公众号，天风证券研究所



资料来源：智研咨询公众号，天风证券研究所

表 8：AI 服务器厂商

公司	主要情况
浪潮信息	浪潮信息是全球领先的 IT 基础设施产品、方案和服务提供商，通过场景优化设计形成了丰富的产品线，涵盖计算型、存储型、多节点、关键应用、整机柜等各类服务器，支持全场景高效计算。2022 年上半年，浪潮信息服务器出货量保持全球第二，中国第一。
工业富联	工业富联在云计算服务器出货量持续全球第一，与全球主要服务器品牌商、国内外 CSP 客户深化合作，推出新一代云计算基础设施解决方案，包括模块化服务器、高效运算 HPC 等，重点解决因 ChatGPT 持续升温而引发 AIGC 算力井喷需求。公司多年来为数家第一梯队云服务商 AI 服务器（加速器）与 AI 存储器供应商，产品已经开发至第四代。伴随着 AI 硬件市场迅速成长，公司相关产品 2022 年出货加倍，AI 服务器及 HPC 出货增长迅速，在 2022 年云服务商产品中，占比增至约 20%，持续维持增长态势。
纬颖科技	纬颖科技专注于提供超大型数据中心及云端基础构架各项产品及系统的解决方案，藉由传承纬创资通丰富的创新设计技术，以及多年的制造经验，纬颖科技的顶尖团队为客户提供巨型数据中心、运营商及企业的云计算解决方案，其中包括高性价比的服务器、储存设备、网络系统、机房基础设备及软硬件整合的私有云解决方案。
中兴通讯	中兴通讯服务器产品拥有该领域多项自主知识产权及专利，涵盖机架服务器、刀片服务器和云计算节点服务器，满足互联网、高性能计算、数据库、数据中心等多种场景需求，并配合中兴通讯其它产品形成丰富的产品组合，满足客户的各种业务需要。服务于政府、金融、交通、教育、医疗、电信运营商和互联网等多个行业。
中科曙光	中科曙光以 IT 核心设备研发、生产制造为基础，对外提供高端计算机、存储、安全数据中心等 ICT 基础设施产品，大力发展云计算、大数据、人工智能、边缘计算等先进计算业务，为用户提供全方位的信息系统服务解决方案。主要产品为高端计算机、存储产品、软件开发、系统集成、技术服务。
超聚变	超聚变聚焦发展算力与生态，致力成为全球领先的算力基础设施与服务提供者，凭借在多样性算力、液冷技术等方面的优势，成功入选 Gartner 全球服务器代表性厂商名录。
新华三	新华三集团作为数字化解决方案领导者，致力于成为客户业务创新、数字化转型值得信赖的合作伙伴。作为紫

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：电子行业深度研究-AI算力需求持续释放-重点看好AI服务器产业链-天风证券.pdf

请登录 <https://shgis.com/post/1796.html> 下载完整文档。

手机端请扫码查看：

