证券研究报告 | 行业深度报告

信息技术 | 计算机

算力：AIGC时代的“卖铲人”

——AIGC系列报告（二）

刘玉萍

liuyuping@cmschina.com.cn

S1090518120002

周翔宇

zhouxiangyu@cmschina.com.cn

S1090518050001

CMS 招商证券

2023.4.2

要点概览

本篇报告系统地梳理了大模型训练及推理需要多少算力。

大模型参数量快速提升，Transformer架构成为发展趋势。根据最新论文对“涌现”效应的研究，当模型训练量超过10^{22}后，模型准确率有了很大的提升，近年来，NLP模型的发展十分迅速，模型的参数量每年以5至10倍的速度在提升，背后的推动力即大模型可以带来更强大更精准的语言语义理解和推理能力。Transformer架构通过计算数据之间的关系提取信息，相较于卷积神经网络具有更强大的运算效率，更适合参数和数据集庞大的自然语言处理学习。

基于GPT3大模型的训练/推理所需的算力及金额测算。

- 训练端，以GPT3为例，完成一次大模型训练所需的算力需求量为3646PF·Days，若用10000张英伟达V100/A100训练则分别需要14.59/3.34天，对应训练费用分别为4.72/1.89百万美元。
- 推理端，以GPT3为例，1000个token的推理算力需求约为350TFLOPS，对应推理成本约为0.15美分。

英伟达GPU是当前最适合做训练的AI芯片。GPU提供多核并行计算的基础，且核心数众多，可以支撑大量数据的并行运算，英伟达Tensor Core技术能够通过降低精度，在性能方面实现数量级的提高。此外，针对大规模AI训练，英伟达推出DGX系统，包括A100、H100、BasePOD、SuperPOD四款产品，其中，DGX A100、DGX H100为英伟达当前服务于AI领域的服务器产品。

投资建议：算力是AIGC时代的“卖铲人”。我们认为发展算力基础设施是AIGC产业发展中必不可少的环节，我国在算力领域仍有较大成长空间。其中，国产AI芯片领域重点推荐寒武纪（电子联合覆盖）、海光信息（电子联合覆盖）；服务器领域重点推荐中科曙光、浪潮信息。

风险提示：AI服务器供应链风险；AI芯片研发不及预期风险；AI相关上市公司短期涨幅过大风险。

目录

1.1 模型不断增大，Transformer架构成为发展趋势

AI模型的参数量及数据集不断增大，Transformer架构成为最适合大模型的架构。

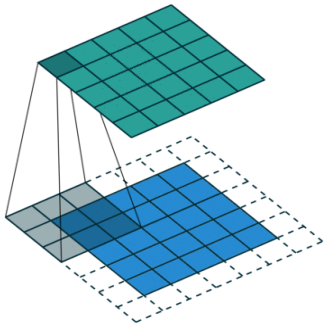
- 1998年LeCun提出了第一个卷积神经网络，随即被用于美国邮政系统的手写邮编识别。但由于此类方法需要较大的数据集和较强的算力，此类方法在之后的十几年里的发展缓慢。
- 由于通信和计算领域基础设施的完善，卷积神经网络在2012年之后迎来了爆发式的发展，模型和数据集都扩大了几个数量级。
- 2018年，研究发现当模型和数据集到达一定规模时，继续扩大模型和数据集给卷积神经网络带来的收益有限。Transformer架构通过计算数据之间的关系提取信息，更适合参数和数据集庞大的自然语言处理学习。

表：深度学习发展过程中代表性的经典模型介绍

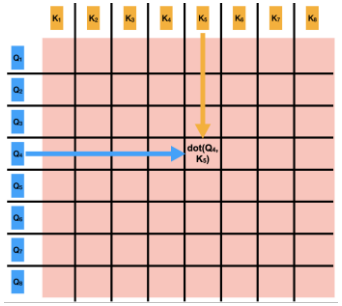
	时间	模型	数据集	模型架构	参数	当时的算力
图像	1998	LeNet	MNIST，6万张28*28的手写数字黑白图片	卷积神经网络	6万	CPU
	2015	ResNet	ImageNet，1500万张224*224的彩色图片		6000万	NVIDIA V100
NLP	2018	BERT	33亿 token 的NLP数据集	Transformer	3亿	NVIDIA A100
	2020	GPT-3	3000亿 token 的NLP数据集		1750亿	NVIDIA H100

图：卷积神经网络

卷积神经网络通过卷积核提取输入数据中的信息



图：Transformer网络



Transformer通过计算输入数据之间的关系来提取信息

1.2 涌现理论：大模型是自然语言处理的核心

自然语言处理任务的准确率与训练量紧密相关，因此大模型在自然语言处理领域不可或缺。

- 根据最新的论文研究，当模型训练量小于10^{22}时，模型在几个自然语言处理任务上的准确率都在0附近，而当模型训练量超过10^{22}后，模型的准确率有了很大的提升，该效应称之为“涌现”。
- 根据OpenAI的官网披露，GPT4大模型在参数量及数据集较GPT3有大幅提升，我们认为从GPT4在自然语言任务处理上所表现出的优异性能进一步表明，通过提高模型参数量、扩大数据集来提高模型性能的方法仍然没有碰到天花板。

图：各种模型Emergent Ability出现时的训练量对比

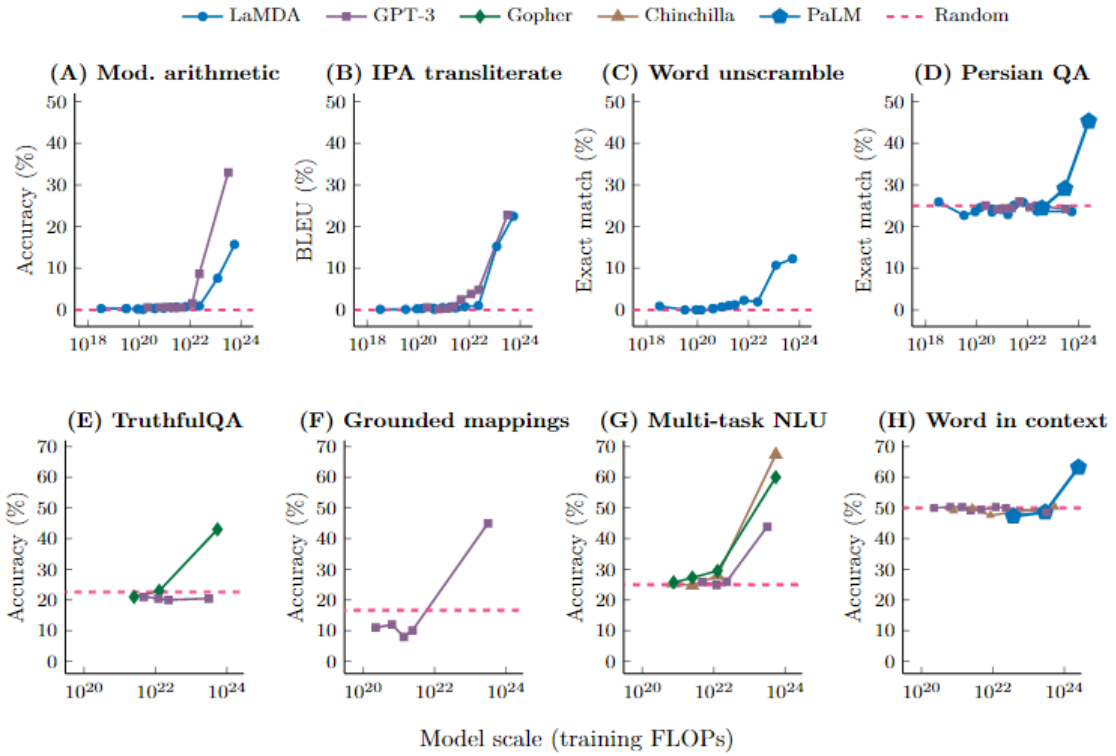


Figure 2: Eight examples of emergence in the few-shot prompting setting. Each point is a separate model. The ability to perform a task via few-shot prompting is emergent when a language model achieves random performance until a certain scale, after which performance significantly increases to well-above random. Note that models that used more training compute also typically have more parameters—hence, we show an analogous figure with number of model parameters instead of training FLOPs as the x-axis in Figure 11. A–D: BIG-Bench (2022), 2-shot. E: Lin et al. (2021) and Rae et al. (2021). F: Patel & Pavlick (2022). G: Hendrycks et al. (2021a), Rae et al. (2021), and Hoffmann et al. (2022). H: Brown et al. (2020), Hoffmann et al. (2022), and Chowdhery et al. (2022) on the WiC benchmark (Pilehvar & Camacho-Collados, 2019).

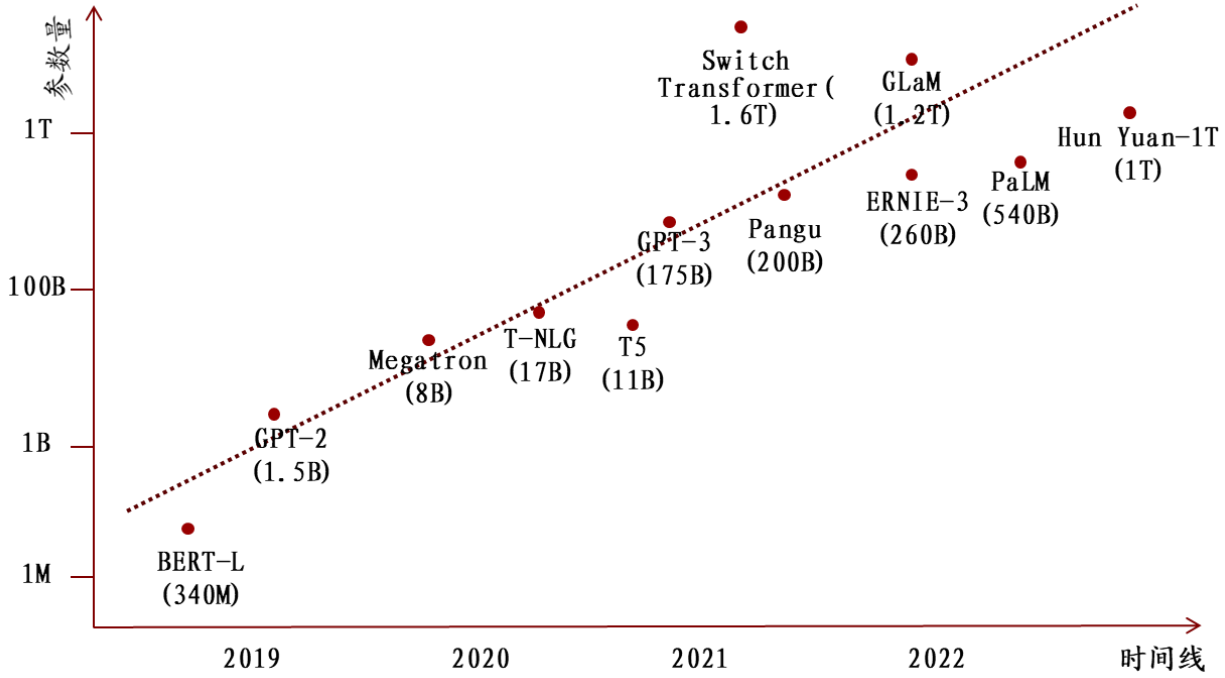
资料来源：“Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models[J]. arXiv preprint arXiv:2206.07682, 2022.”、OpenAI官网、招商证券

1.3 大模型参数量快速提升

目前，大模型参数量快速增长，已经出现万亿级别的大模型。

- 近年来，NLP大模型的发展十分迅速，模型的参数量每年以5至10倍的速度在提升，背后的推动力即大模型可以带来更强大更精准的语言语义理解和推理能力。
- 2020年末，OpenAI发布的GPT-3模型达到了1750亿参数的大小，相比2018年94M的ELMo模型，三年的时间整整增大了1800倍之多。按此趋势，预计两年后，会有100 Trillion参数的模型推出。

图：大模型参数量快速增长



资料来源：腾讯AIGC白皮书、招商证券

1.4 英伟达GPU是当前最适合做训练的AI芯片

革命性的AI训练能，英伟达GPGPU是目前最适合做AI训练的芯片。

- GPU提供多核并行计算的基础，且核心数众多，可以支撑大量数据的并行运算。AI场景训练和推理通常不涉及大量的分支运算与复杂的控制指令，更适合在GPU上进行。
- 英伟达Tensor Core能够通过降低精度，例如Transformer引擎中的8位浮点（FP8）、Tensor Float32（TF32）和FP16，在性能方面实现数量级的提高。此外，通过 CUDA-X库直接支持原生框架，实施可自动完成，从而在保持准确性的同时，大幅缩短从训练到收敛的时间。
- 目前，国内外主流云计算厂商均使用英伟达GPU芯片作为其超级计算能力的底座。

图：英伟达GPGPU架构

表：云计算厂商均采用英伟达芯片

	M60	P4	P40	P100	T4	RTX	V100	A10	A40	A100	NGC
阿里云		✓		✓	✓		✓			✓	✓
AWS	✓				✓		✓	✓		✓	✓
百度云		✓	✓		✓		✓			✓	
Google Cloud		✓		✓	✓		✓			✓	✓
IBM Cloud	✓			✓			✓				
Microsoft Azure	✓		✓	✓	✓		✓	✓		✓	✓
Oracle Cloud				✓			✓			✓	✓
腾讯云		✓	✓		✓		✓			✓	✓
NPN CSPs	✓			✓	✓	✓	✓		✓		

资料来源：英伟达官网、招商证券

1.4 英伟达GPU是当前最适合做训练的AI芯片

英伟达TensorCore已经经历了四代，当前H100被誉为最适合Transformer模型训练的芯片。

- 英伟达H100基于英伟达Hopper Tensor Core架构，综合技术创新可以将大型语言模型的速度提高30倍。

表：英伟达V100/A100/H100算力对比

芯片	CUDA Core			Tensor Core		
	FP32 (TFLOPS)	FP16 (TFLOPS)	INT8 (TOPS)	TF32 (TFLOPS)	FP16 (TFLOPS)	INT8 (TOPS)
V100	15.7	31	62		125	
A100	19.5	39	78	156	312	624
H100	67	134	268	495	990	1979

表：英伟达四代Tensor Core架构梳理

Nvidia Tensor Core		效果
第一代	Volta	NVIDIA Volta中的第一代Tensor Core专为深度学习而设计，通过FP16和FP32下的混合精度矩阵乘法提供了突破性的性能。与NVIDIA Pascal相比，用于训练的峰值TFLOPS性能提升了高达12倍，用于推理的峰值TFLOPS性能提升了高达6倍。这项关键功能使Volta提供了比Pascal高3倍的训练和推理性能。
第二代	Turing	NVIDIA Turing Tensor Core技术能进行多精度计算，可实现高效的AI推理。Turing Tensor Core提供了一系列用于深度学习训练和推理的精度（从FP32到FP16再到INT8和INT4），性能大大超过NVIDIA Pascal GPU。
第三代	Ampere	NVIDIA Ampere Tensor Core基于先前的创新成果而构建，通过使用新的精度（TF32 和FP64）来加速和简化AI采用，并将Tensor Core的强大功能扩展至HPC。第三代Tensor Core支持BFloat16、INT8和INT4，可为AI训练和推理创建高度通用的加速器。
第四代	Hopper	NVIDIA Hopper架构利用Transformer引擎改进第四代Tensor Core，该引擎使用新的8位浮点精度，可为万亿参数模型训练提供比FP16高6倍的性能。Hopper Tensor Core使用TF32、FP64、FP16和INT8精度，将性能提升3倍，能够加速处理各种工作负载。

资料来源：英伟达官网、招商证券

目录

2.1 大模型训练算力总需求测算

根据“Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. and Amodei, D., 2020. Scaling laws for neural language ”的论文，基于transformer的自然语言处理(NLP)大模型可分为三类：Encoder-Only(E)，Decoder-Only(D)和Encoder-Decoder(ED)。模型的训练算力需求可根据以下公式计算：

$$\text{训练算力需求} = \text{模型参数量} \times \text{数据集}token\text{数} \times \text{系数}k$$

其中，k的取值取决于模型种类，如果模型种类为Encoder-Only或Decoder-Only，则k=6；如果模型种类为Encoder-Decoder，则k=3。

以GPT3大模型为例，总参数量（parameters）约等于175B（$175*10^9$）；数据集token数约等于300B（$300*10^9$），GPT3大模型是Decoder-Only（D），因此我们测算GPT3大模型训练算力需求量为：

$$(175*10^9) \times (300*10^9) \times 6 = 3.15 \times 10^{23} = 315\ \text{ZettaFLOPS}$$

转换为单日算力需求：

$$3.15 \times 10^{23} \div 24 \div 365 = 3646\ \text{PF} \cdot \text{Days}$$

资料来源：“Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. and Amodei, D., 2020. Scaling laws for neural language ”、OpenAI官网、招商证券

2.1 大模型训练算力总需求测算

表：目前已知大模型训练算力总需求

	model	Parameters（亿）	token in dataset（亿）	domain		k	Zflops (10^21)	pf*day (10^15)
Google	BERT	4810	33	NLP	encoder-only	6	10	110
	PaLM	5400	7800	NLP	decoder-only	6	2527	29250
	magen	110		多模态（文本+图像）				
	lambda	1370	1680	NLP	decoder-only	6	138	1598
	Parti	200		多模态（文本+图像）	decoder-only	6		
Microsoft	Florence	6.4		多模态（文本+图像）				
	Turing-NLG	170	2700	NLP	encoder-only	6	28	319
Facebook	OPT-175B	1750	1800	NLP	decoder-only	6	189	2188
	M2M-100	150		NLP	encoder-decoder	3		
Deep Mind	Gopher	2800	3000	NLP	encoder-only	6	504	5833
	AlphaCode	414	9670	NLP	encoder-decoder	3	120	1390
OpenAI	GPT3	1750	3000	NLP	decoder-only	6	315	3646
	ChatGPT	1751		NLP	decoder-only	6		
	GPT4	1750-2800		多模态（文本+图像）				
Nvidia	Megatron-Turing NLG	5300	2700	NLP	decoder-only	6	859	9938
百度	ERNIE	2600	3000	NLP	decoder-only	6	468	5417

2.2 大模型训练费用测算

根据单卡峰值算力，我们可以通过以下公式测算训练模型所需时间：

训练时间 = 训练总计算量 ÷ 单卡峰值算力 ÷ 算力利用率 ÷ 芯片卡数

其中，算力利用率与芯片数量成反比，与芯片架构迭代成正比。以GPT3为例，若10000张英伟达V100芯片训练，算力利用率为20%，则训练GPT3所需训练时间为：

3646 PF·Days ÷ 125TFlops ÷ 20% ÷ 10000 = 14.59 Days

表：英伟达V100/A100训练GPT3模型分别需要天数

芯片	单卡峰值算力	算力利用率	训练GPT3所需时间（天）
V100	125TFlops	20%	14.59
A100	312TFlops	35%	3.34

我们认为得出训练GPT3模型一次的资金需求公式为：

训练价格 = 训练总时长 X（单卡价格（$/小时）X 24）X 芯片数

根据目前微软Azure服务器租赁价格测算，我们测算用10000颗V100训练GPT3模型一次的资金需求公式为：

14.59 X（1.350（$/小时）X 24）X 10000 = 4.72 Million USD

表：英伟达V100/A100训练GPT3模型价格

芯片	微软Azure服务器（$/小时）	单卡价格（$/小时）	芯片数	训练时长（天）	训练GPT模型价格（百万美元）
V100	10.796	1.350	10000	14.59	4.723
A100	18.829	2.354	10000	3.34	1.886

资料来源：微软Azure、英伟达官网、OpenAI官网、招商证券

2.2 大模型训练费用测算

表：目前已知大模型通过英伟达A100/V100训练所需金额

公司	模型	算力需求(PF*day)	一万张V100总时间（天）	一万张A100总时间（天）	V100价格（百万美元）	A100价格（百万美元）
谷歌	BERT	110	0.44	0.10	0.143	0.057
	PaLM	29250	117.00	26.79	37.894	15.130
	Lambda	1598	6.39	1.46	2.071	0.827
微软	Turing-NLG	319	1.28	0.29	0.413	0.165
Facebook	OPT-175B	2188	8.75	2.00	2.834	1.132
Deep Mind	Gopher	5833	23.33	5.34	7.557	3.017
	AlphaCode	1390	5.56	1.27	1.801	0.719
OpenAI	GPT3	3646	14.58	3.34	4.723	1.886
英伟达	Megatron-Turing NLG	9938	39.75	9.10	12.874	5.140
百度	ERNIE	5417	21.67	4.96	7.017	2.802

自建一个类GPT3大模型算力基础设施成本超过2.5亿美元。

- 以10000张英伟达A100芯片为例，英伟达DGX A100服务器内涵8张英伟达A100芯片，服务器定价约20万美元，10000张A100对应1250台服务器，总价约为2.5亿美元，约17.17亿元人民币。

2.3 推理所需要的算力需求及成本测算

大模型除训练需求外，在日常问答应用中将产生大量推理需求。根据“Scaling laws for neural language”的论文，模型的推理算力需求可根据以下公式计算：

$$\text{推理算力需求} = \text{模型参数量} \times (\text{“输入+输出”}token\text{数}) \times \text{系数}k$$

其中，k的取值取决于模型种类，如果模型种类为Encoder-Only或Decoder-Only，则k=2；如果模型种类为Encoder-Decoder，则k=1。

以GPT3大模型为例，总参数量（parameters）约等于175B（$175*10^9$）；1000个token的推理算力需求为：

$$(175*10^9) \times (1000) \times 2 = 3.5 \times 10^{14} = 350\ \text{TFLOPS}$$

假设使用一颗A100芯片推理，使用TF32峰值算力156TFLOPS，则消耗时间约为：

$$350\ \text{TFLOPS} \div 156\ \text{TFLOPS} = 2.24\ \text{second}$$

已知A100单卡价格为$2.354/小时，则在GPT3模型中，1000token的推理成本为：

$$\$2.354/\text{h} \div 3600 \times 2.24 = \$0.0015/1\text{k token}$$

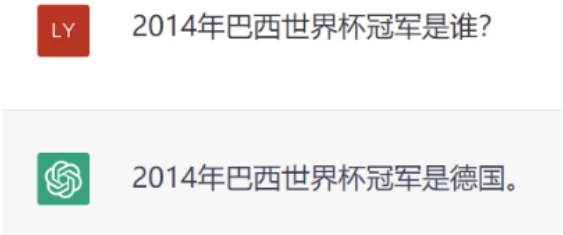
资料来源：“Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. and Amodei, D., 2020. Scaling laws for neural language”、英伟达官网、招商证券

2.4 模型API接口调用价格测算

目前，ChatGPT及GPT4都已上线API接口，其中ChatGPT API价格为0.2美分/1k token，而GPT4 API价格较ChatGPT大幅提升。此外，百度API接口调用价格初定为0.012元/1k token。

表：目前已知模型API接口调用价格

模型	输入端	输出端
ChatGPT	$0.002/1k token	$0.002/1k token
GPT4-8K	$0.03/1k token	$0.06/1k token
GPT4-32K	$0.06/1k token	$0.12/1k token
百度文心一言	￥0.012/1k token	￥0.012/1k token



以左图为例，我们通过询问ChatGPT一个问题并得到答案测算以上模型调用API的费用，其中问题“2014年巴西世界杯冠军是谁？”共计15个token；回答“2014年巴西世界杯冠军是德国”共计16个token（包含标点符号）。

表：目前已知模型API接口调用费用测算

模型	输入端token数	输入端价格	输入端金额	输出端token数	输出端价格	输出端金额	总金额
ChatGPT	15	$0.002/1k token	$0.00003	16	$0.002/1k token	$0.000032	$0.000062
GPT4-8K	15	$0.03/1k token	$0.00045	16	$0.06/1k token	$0.00096	$0.00141
GPT4-32K	15	$0.06/1k token	$0.0009	16	$0.12/1k token	$0.001920	$0.00282
百度文心一言	15	￥0.012/1k token	¥0.00018	16	￥0.012/1k token	¥0.000192	¥0.000372

目录

3.1 针对大规模AI训练，英伟达推出DGX系统

英伟达DGX系统针对企业大规模AI基础架构提供出色的解决方案，专门打造的先进AI系统产品系列。

- 每个NVIDIA DGX系统均配备可提供企业支持的DGX硬件和NVIDIA Base Command软件，其中包含强化的系统软件、优化的 AI 库、出色的集群管理、稳健的工作调度和工作负载编排。
- NVIDIA DGX系统包括A100、H100、BasePOD、SuperPOD四款产品。其中，A100、H100为DGX系统主要服务器产品。

图：英伟达DGX系统组成

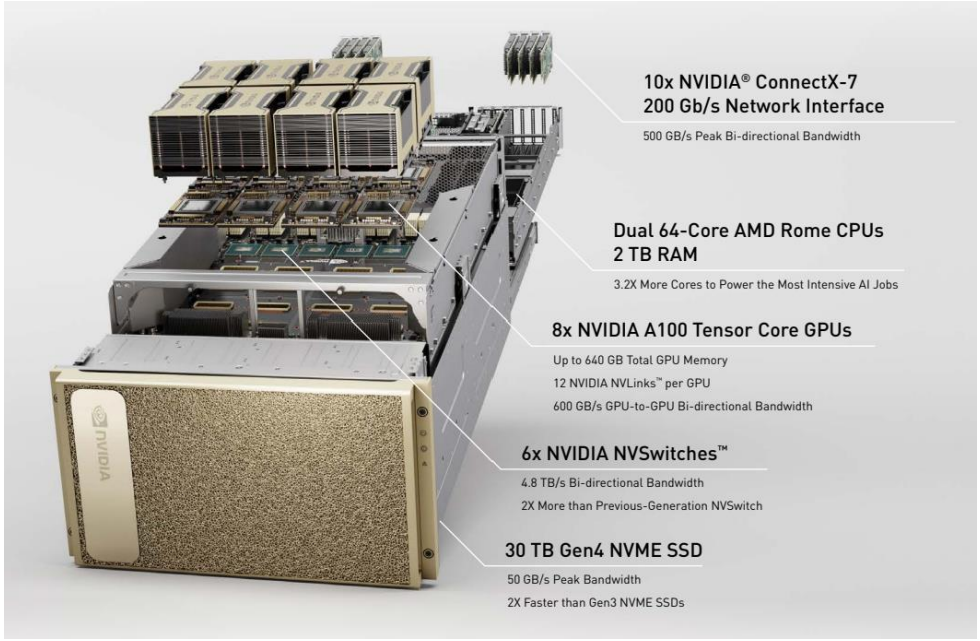
资料来源：英伟达官网、招商证券

3.2 英伟达DGX A100：目前最主流的AI服务器

NVIDIA DGX A100是适用于各种AI工作负载的通用系统，能够为率先推出的5 petaFLOPS AI系统提供之前难以实现的计算密度、性能和灵活性。

- DGX A100采用NVIDIA A100 Tensor Core GPU，使企业能够将训练、推理和分析整合到一个易于部署的统一AI基础架构中。
- NVIDIA DGX A100不仅仅是一台服务器，更是一个完整的软硬件平台。它基于全球最大的DGX集群NVIDIA DGX SATURNV积累的知识经验而建立，背后有NVIDIA数千名AI专家支持。

图：英伟达DGX A100服务器架构



表：英伟达DGX A100服务器参数

系统规格

	NVIDIA DGX A100 640GB
GPU	8 个 NVIDIA A100 80GB Tensor Core GPU
GPU 显存	共 640GB
性能	5 petaFLOPS AI 10 petaOPS INT8
NVIDIA NVSwitch	6
系统功耗	最大 6.5 千瓦
CPU	双路 AMD Rome 7742、共 128 个核心、2.25 GHz（基准频率）、3.4 GHz（最大加速频率）
系统内存	2TB

3.3 英伟达DGX H100：DGX系统的最新迭代

DGX H100是NVIDIA DGX系统的最新迭代，也是NVIDIA DGX SuperPOD的基础。

- DGX H100包含8个NVIDIA H100 GPU，总显存高达640GB，峰值性能高达32petaFLOPS。
- 作为全球首款搭载NVIDIA H100 Tensor Core GPU的系统，NVIDIA DGX H100可带来突破性的AI规模和性能。它搭载NVIDIA ConnectX-7智能网卡和NVIDIA BlueField-3数据处理器（DPU），为NVIDIA DGX SuperPOD带来6倍性能提升，2倍更快的网络，和高速可扩展性。新一代架构可用于自然语言处理和深度学习推荐模型等复杂的大型AI任务。

图：英伟达DGX H100服务器架构

表：英伟达DGX H100服务器参数

规格	
GPU	8 个 NVIDIA H100 Tensor Core GPU
GPU 显存	共 640GB
性能	32 petaFLOPS FP8
NVIDIA® NVSwitch™	4x
系统功耗	最高 10.2kW
CPU	双路 x86
系统内存	2TB

资料来源：英伟达官网、招商证券

目录

4.1 AI芯片稀缺标的寒武纪、海光信息

国产AI芯片标的稀缺，重点推荐寒武纪、海光信息。

- **寒武纪：**公司专注于人工智能芯片产品的研发与技术创新。目前，公司在AI训练/推理领域拥有智能加速卡思源系列产品。在2022年WAIC上，董事长陈天石透露了公司在研全新一代云端智能训练芯片思元590，据介绍，思元590采用MLUarch05全新架构，实测训练性能较在售旗舰产品有大幅提升，将提供更大的内存容量和更高的内存带宽，其IO片间互联接口也较上代实现大幅升级。
- **海光信息：**公司深度计算处理器基于主流通用并行计算架构，可搭配海光CPU使用，广泛应用于科学计算、人工智能模型训练和推理。

表：寒武纪MLU370&海光信息Z100&英伟达A100产品参数对比

产品参数		寒武纪思元MLU370-X8	海光深算Z100	英伟达A100
峰值算力	INT8	256TOPS	49.1TOPS	624TOPS
	INT16	128TOPS	-	-
	FP16	96TFLOPS	24.5TFLOPS	312TFLOPS
	BF16	96TFLOPS	-	312TFLOPS
	FP32	24TFLOPS	12.2TFLOPS	156TFLOPS
显存容量		32GM	32GB	80GB
内存带宽		1228GB/s	1024GB/s	1935GB/s
最大热功耗		250W	280W	300W

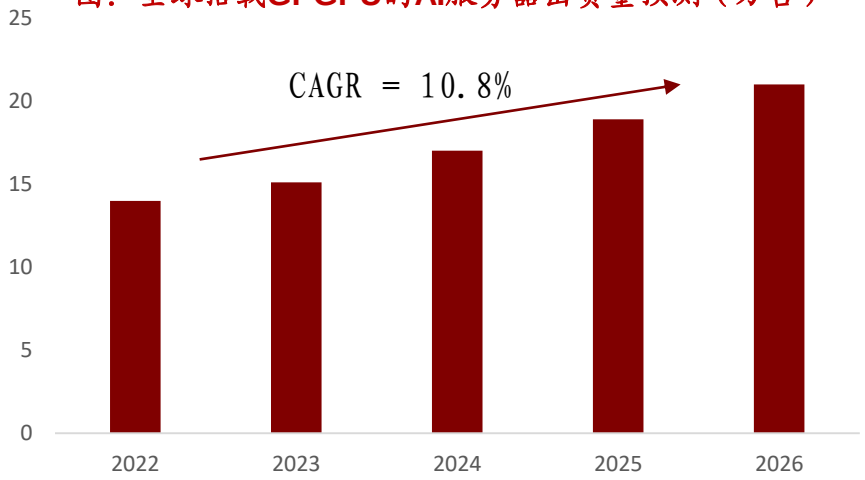
资料来源：英伟达官网、寒武纪官网、海光信息招股说明书、招商证券

4.2 服务器重点推荐中科曙光、浪潮信息

服务器领域重点推荐中科曙光、浪潮信息。

- 据TrendForce数据，预估2022年搭载GPGPU的AI服务器年出货量占整体服务器比重近1%，即约14万台。预计2023年出货量年成长8%，2022至2026年年复合增长率达10.8%。

图：全球搭载GPGPU的AI服务器出货量预测（万台）



图：浪潮信息AI服务器

图：中科曙光AI服务器

表：中科曙光&浪潮信息AI服务器相关业务梳理

标的	AI服务器相关业务
中科曙光	公司多年布局人工智能生态，以丰富的产品覆盖、创新的架构设计，打造高效易维护的人工智能加速平台。公司联营企业海光信息是国内AI芯片龙头，拥有深算一号AI芯片，与曙光AI服务器有较大协同效应。
浪潮信息	浪潮AI服务器的中国市场份额连续四年保持在50%以上，并与人工智能领先科技公司保持在系统与应用方面的深入紧密合作2021 年，人工智能服务器全球市场份额 20.9%，保持全球第一，中国市场份额超过50%。

资料来源：TrendForce、中科曙光官网、浪潮信息官网、招商证券

风险提示

AI服务器供应链风险：北美时间3月2日，美国商务部发布公告，将28个中国实体列入实体清单，其中包括浪潮集团。如果未来我国AI服务器相关企业无法向美国购买核心零部件产品则对公司未来业务发展有较大不利影响。

AI芯片研发不及预期风险：目前我国AI芯片与英伟达仍存在较大差距，若我国AI芯片研发进度不及预期，则对我国AIGC产业发展有较大不利影响。

AI相关上市公司短期涨幅过大风险：目前AI相关上市公司短期涨幅较快，股价波动较大。

参考报告

1、《微软引领AI+办公应用史诗级革命——AI+系列报告三》2023-03-18

2、《ChatGPT快速流行，重构AI商业模式——AIGC投资机会梳理》2023-02-08

分析师承诺

负责本研究报告的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与，未来也将不会与本报告中的具体推荐或观点直接或间接相关。

团队介绍：

刘玉萍：计算机行业首席分析师，北京大学汇丰商学院金融学硕士。2022年水晶球最佳分析师第一名。

周翔宇：计算机行业分析师，三年中小盘研究经历，获得2016/17年新财富中小市值团队第五、第二名。

孟林：计算机行业分析师，中科院信息工程研究所硕士，两年四大行技术部工作经验，两年一级市场投资经验，2020年加入招商证券。

投资评级定义

报告中所涉及的投资评级采用相对评级体系，基于报告发布日后6-12个月内公司股价（或行业指数）相对同期当地市场基准指数的市场表现预期。其中，A股市场以沪深300指数为基准；香港市场以恒生指数为基准；美国市场以标普500指数为基准。具体标准如下：

股票评级

强烈推荐：预期公司股价涨幅超越基准指数20%以上

增持：预期公司股价涨幅超越基准指数5-20%之间

中性：预期公司股价变动幅度相对基准指数介于±5%之间

减持：预期公司股价表现弱于基准指数5%以上

行业评级

推荐：行业基本面向好，预期行业指数超越基准指数

中性：行业基本面稳定，预期行业指数跟随基准指数

回避：行业基本面转弱，预期行业指数弱于基准指数

重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价，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。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

感谢您宝贵的时间

Thank You

前沿报告库
wk.askci.com

前沿报告库是中国新经济产业咨询报告共享平台。行业范围涵盖新一代信息技术、5G、物联网、新能源、新材料、新消费、大健康、大数据、智能制造等新兴领域。为企事业单位、科研院所、投融资机构等提供研究和决策参考。

扫一扫免费
获取海量报告