

ChatGPT:国内算力，拐点之时

AIGC行业深度报告(10)

华西计算机团队

2023年8月14日

分析师：刘泽晶

SAC NO: S1120520020002

邮箱：liuzj1@hx168.com.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

核心逻辑:

- ◆ **英伟达，全球AI算力王者归来:** 巨头指引，海外科技龙头厂商持续布局AI，持续加大资本支出储备AI相关产品，涉及数据中心和服务器，其中包括META、Google、微软，同时海外AI应用持续落地，其商业模式得到验证，例如Microsoft 365 Copilot的定价超预期。AI大模型时代拉开帷幕，对AI芯片需求量明显增多。英伟达实为全球AI算力王者，英伟达相关AI芯片三个月内两度涨价，其根本原因在于算力芯片供不应求，稀缺性强，**算力芯片依旧为大模型时代的稀缺要素。**
- ◆ **AI有望成为台积电未来支柱，英伟达相关服务器厂商业绩有望高度景气:** 台积电表示**AI服务器需求强劲，有望成为未来支柱领域**，公司预计预计AI服务器未来5年收入将实现50%的CAGR增长。同时应对AI芯片需求，台积电加大资本支出。公司为应对英伟达、AMD等厂商对AI芯片 CoWoS 等先进封装的需求，扩大对先进封装制造厂的建设。此举同样验证算力芯片高度景气。英伟达芯片相继落地，AI服务器厂商高度景气，中国台湾多家电子代工厂争相出货有望业绩兑现，例如鸿海、纬创、英业达、广达、超微电脑等。
- ◆ **国内算力，拐点之时: 算力是AIGC落地的先行指标，**我们认为全球正处于AIGC的爆发期，如果说AIGC是科技企业开启第二轮业绩增长曲线的“流量入口”，那么算力即是科技厂商开启AI争夺战的“入场券”，目前我国相关AI应用目前已有雏形，我国正处于智算中心建设的加速期，因此相关AI算力产品有望成为本轮科技浪潮下的先行指标，其中包括算力服务器、光模块、交换机等产品。我们认为相关公司在本轮科技浪潮中业绩有望高度景气，目前相关公司业绩已经得到验证，例如中际旭创、中科曙光等。
- ✓ **投资建议:** 关注三条投资主线：**1)具备算力服务器的厂商，受益标的为浪潮信息、神州数码、中科曙光、紫光股份、拓维信息、工业富联、四川长虹、中国长城等；2)具备算力芯片的厂商，受益标的为寒武纪、海光信息、龙芯中科、景嘉微等；3)相关AI云厂商，受益标的为首都在线、鸿博股份、云赛智联、青云科技、优刻得、光环新网、新炬网络、真视通、恒润股份等。**
- ◆ **风险提示:** 核心技术水平升级不及预期的风险、AI伦理风险、政策推进不及预期的风险、中美贸易摩擦升级的风险。



目录

01 全球科技巨头指引，算力持续验证

02 国内算力，拐点之时

03 投资建议：梳理AIGC相关受益厂商

04 风险提示

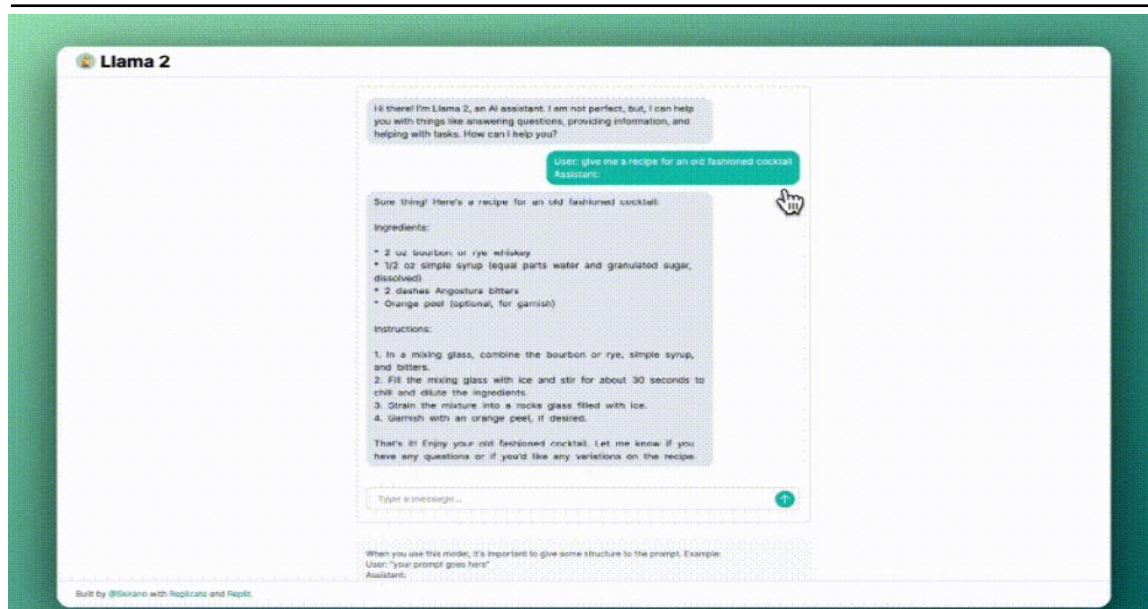


01 全球科技巨头指引，算力持续验证

1.1.1 巨头指引: Meta AI+ 广告业务加速回暖，加大AI投资

- ◆ **应用程序业务为绝对收入来源，拓宽元宇宙业务：**Meta主营业务共分为应用程序系列部门（FoA）和现实实验室（Reality Labs）两部分。其中，应用程序系列主要包括其四大社交平台Facebook、Instagram、Messenger、WhatsApp，以及其他服务；Reality Labs主要包括增强和虚拟现实相关的消费硬件、软件和内容。
- ◆ **AI+ 广告业务加速回暖，业绩增长超预期：**当地时间7月26日，Meta公布截至2023年6月30日的第二季度业绩，营收创一年半最大同比增幅，EPS止降回升、增长21%。《华尔街日报》称，广告业务的反弹、Reels的变现以及**人工智能驱动的广告定向投放**和衡量都对第二财季的业绩做出了贡献。此外，Meta三季度营收指引高于预期，今年费用指引上调20亿美元，今年资本支出指引下调30亿美元，对非AI服务器节约成本，称继续投资AI和元宇宙将导致明年费用增加，元宇宙业务明年亏损将大增。
- ◆ **将加大AI投资，挖掘大模型价值：**今年2月，Meta在官网公布了一款新的人工智能大型语言模型Llama，并在7月宣布将与包括微软在内的主要云计算服务提供商合作，将其AI大语言模型Llama 2投入商业应用。Meta表示，不会对其大模型的访问或使用收费。扎克伯格提到，今年的一大重要事项就是继续挖掘Llama的价值，同时继续专注AI和元宇宙的结合。

Llama 2 大模型应用展示



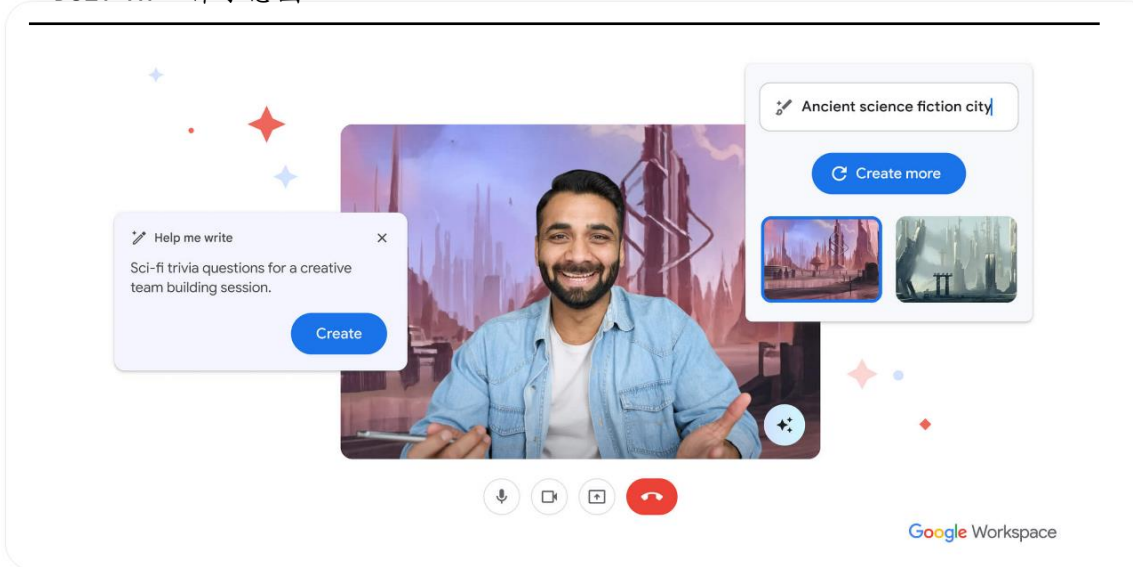
Meta元宇宙部门产品 Quest 3



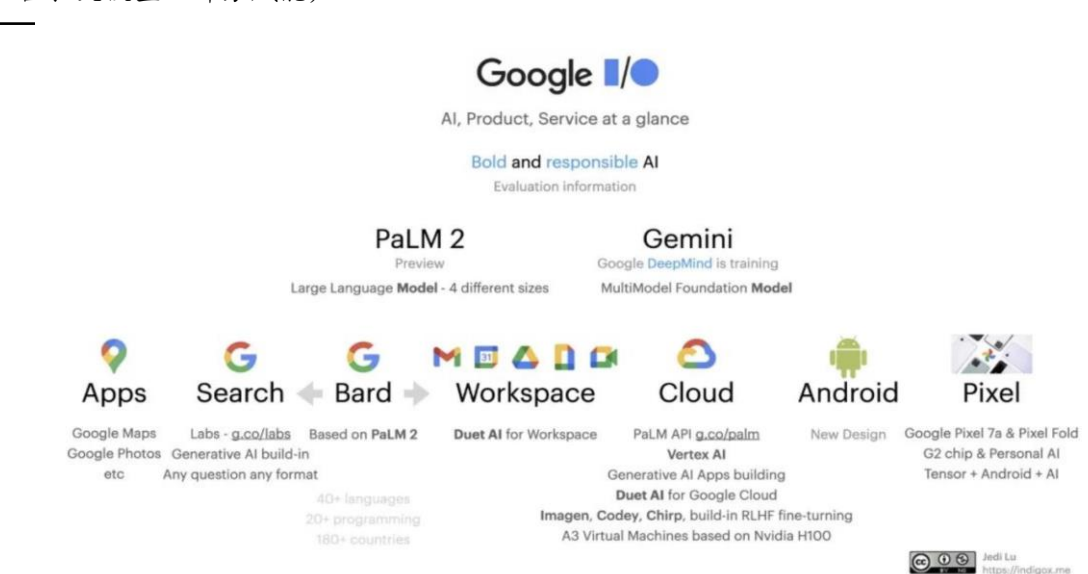
1.1.2 巨头指引: 谷歌持续加大AI投资，涉及数据中心和服务器

- ◆ **谷歌第二季度财报发布，广告和云业务双丰收：**谷歌第二季度实现营收746.04亿美元，同比增长7%，超出市场预期4.4%的增长，和727.7亿美元的营收预期。其中广告业务创收381.42美元，同比增长3.3%，我们认为其高于市场预期，其中最大预期差来自于YouTube广告(同比增速4.4%)。云相关业务营收80.31美元，同比增长约28%，保持高增长速度。
- ◆ **谷歌将持续加大对生成式AI的探索：**根据微软公开财报会，谷歌会持续加大对生成式AI的探索，其中相关管理层已经看到了生成式AI的巨大潜能，Bard的相关数据非常喜人，Duet AI、DeepMind、搜索业务、Gemini(多模态)等也有望引入谷歌云服务之中。此外，目前AI在帮助客户进行大规模广告覆盖中起到了基础作用，其中包括出价、广告针对性、创意、创新等服务，谷歌高管宣称未来有望引入生成式AI新产品，构建AI广告功能。
- ◆ **未来将继续加大AI领域投资，涉及数据中心和服务器：**谷歌公布本季度的资本支出68.9亿美元，谷歌表示在人工智能领域现阶段已处于领先地位，正在推动搜索引擎的下一步发展。根据公司公开财报会表示，公司将持续扩大资本开支，其中包括TPU、GPU、及数据中心扩容方面。

DUET AI工作示意图



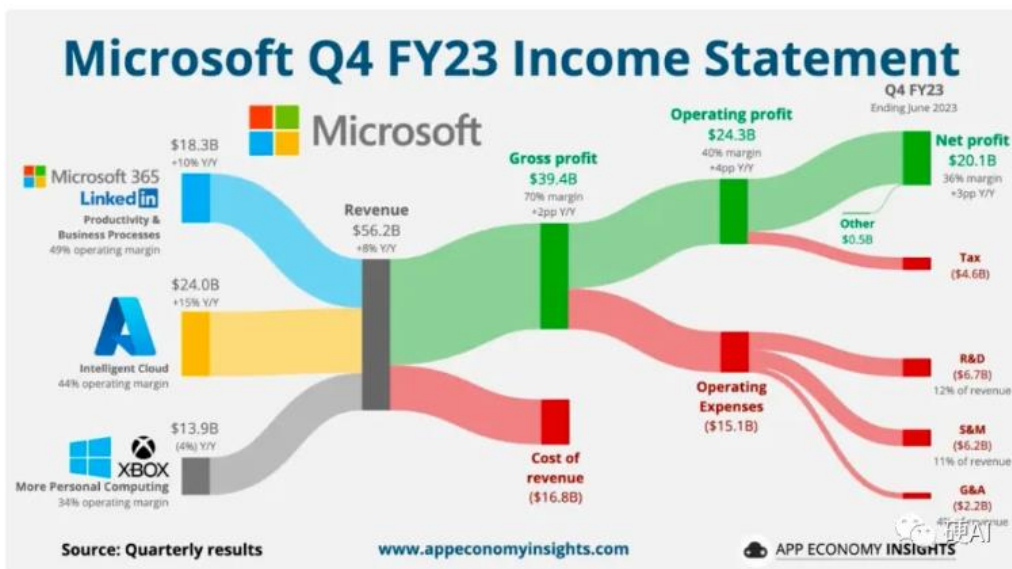
谷歌大模型及部分赋能产品



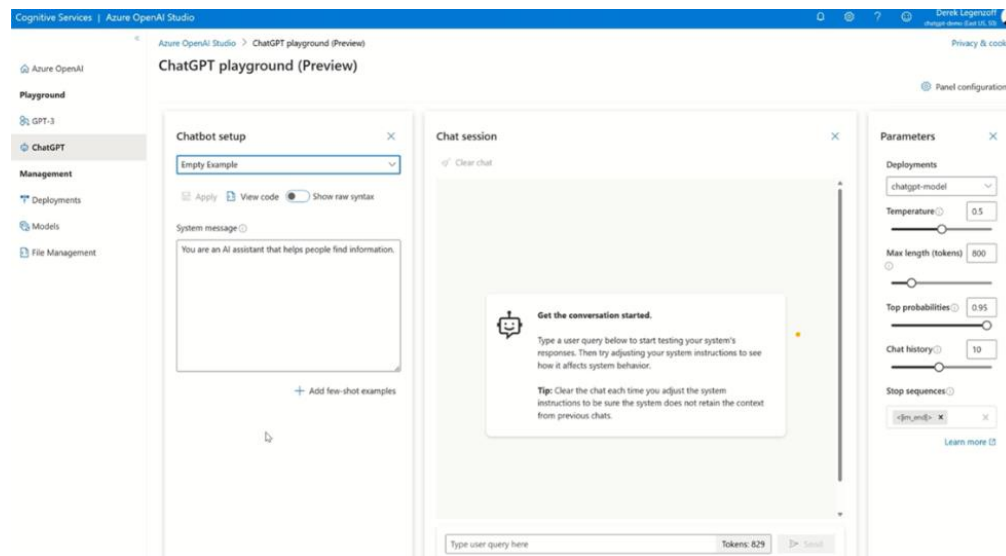
1.1.3 巨头指引：微软持续加大资本支出，AI实为微软关键胜负手

- ◆ **营收超预期，AI影响尚未显现：** 北京时间7月26日，微软发布2023年第二季度财报。其中第二季度营收562亿美元，同比增长8%。具体来看，云服务营收240亿美元，同比增长15%。其中，Azure和其他云服务收入同比上涨26%；微软预计，下财季 Azure营收增长率将为25%至26%。生产力和业务流程营收183亿美元，同比增长10%；个人计算业务收入139亿美元，同比下降4%，Office365系列软件营收同比增长15%，微软365消费者订阅数增长至6700万；Dynamics365软件营收同比增长26%；然而其中AI影响尚未显现。
- ◆ **微软破局之道，AI呼之欲出：** 根据微软公开电话会议，微软宣布为满足人工智能信号，整个财年中每个季度的资本支出都将连续增长，资本支出年增长率约为50%，将达到450亿至500亿美元；同时，Office等产品的定价也成为公司未来业绩增长的保障；此外，微软宣布聊天机器人Bing Chat将在第三方浏览器上推出，包括谷歌的Chrome和苹果的Safari；我们判断微软AI实为微软的第二波成长曲线，也为其破局之路，微软选择一种非常坚定的姿态，持续投资AI建设。

2023财年第四财季及全年报表



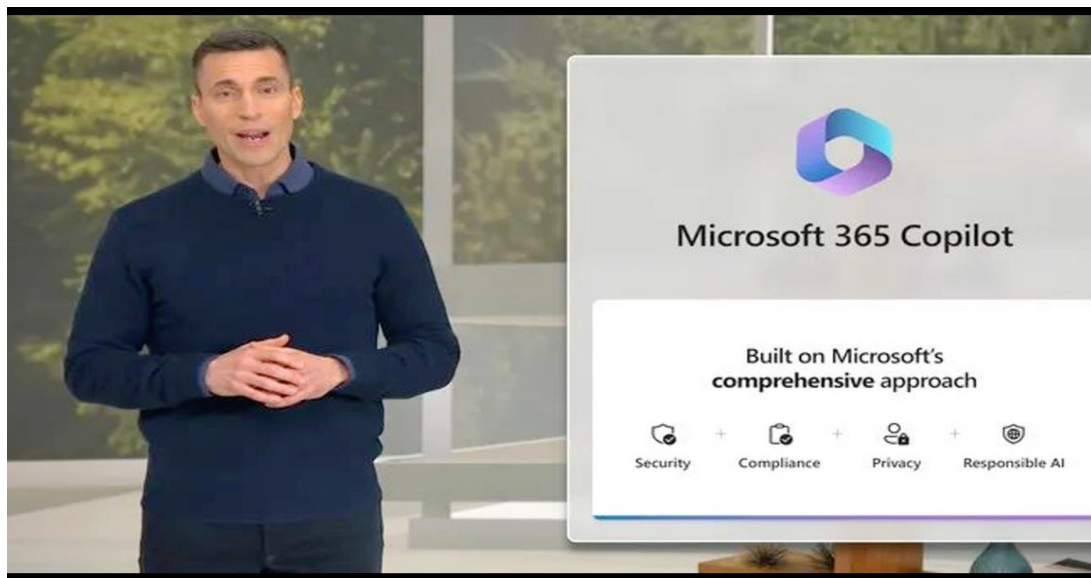
微软云Open AI解决方案示意图



1.2 海外AI应用持续落地，商业模式得到验证

- ◆ **Microsoft 365 Copilot定价超预期，AI市场潜力的巨大。** 7月19日，在Microsoft Inspire 活动中，微软宣布了 Microsoft 365 Copilot 的定价：企业账户每用户每月售价为 30 美元，远超市场预期，彰显人们对AI工具提升工作效率的认可以及产品需求强劲。受此影响，上周二微软股价收涨3.98%，报359.49美元，创历史新高。据SensorTower近日发布的《2023年AI应用市场洞察》报告，2023年上半年，全球AI应用市场迅猛增长，下载量同比增长114%，达到3亿次；AI应用内购收入同比增长175%，接近4亿美元，AI市场潜力的巨大。
- ◆ **从云到端，国外大厂进入新一轮的大模型周期。** 今年4月，微软正式开源DeepSpeedChat训练模型，Meta开源LLaMA，在AI开发者中快速兴起基于LLaMA的训练风潮；5月，谷歌发布了适用边缘的大模型PaLM2，打通从云到边的AI全链路；6月，Meta 推出“类人”AI 模型I-JEPA，能比现有模型更准确地分析和完成未完成的图像；7月，苹果正式宣布启动开发AppleGPT，Meta发布最新AI大语言模型LLaMa 2系列，模型信息和起始代码全部开源，并支持免费可商用等。都表明国外大厂正在进入新一轮的大模型周期。

Microsoft 365 Copilot



Meta发布LLaMa 2

Large language model

Llama 2: open source, free for research and commercial use

We're unlocking the power of these large language models. Our latest version of Llama – Llama 2 – is now accessible to individuals, creators, researchers, and businesses so they can experiment, innovate, and scale their ideas responsibly.

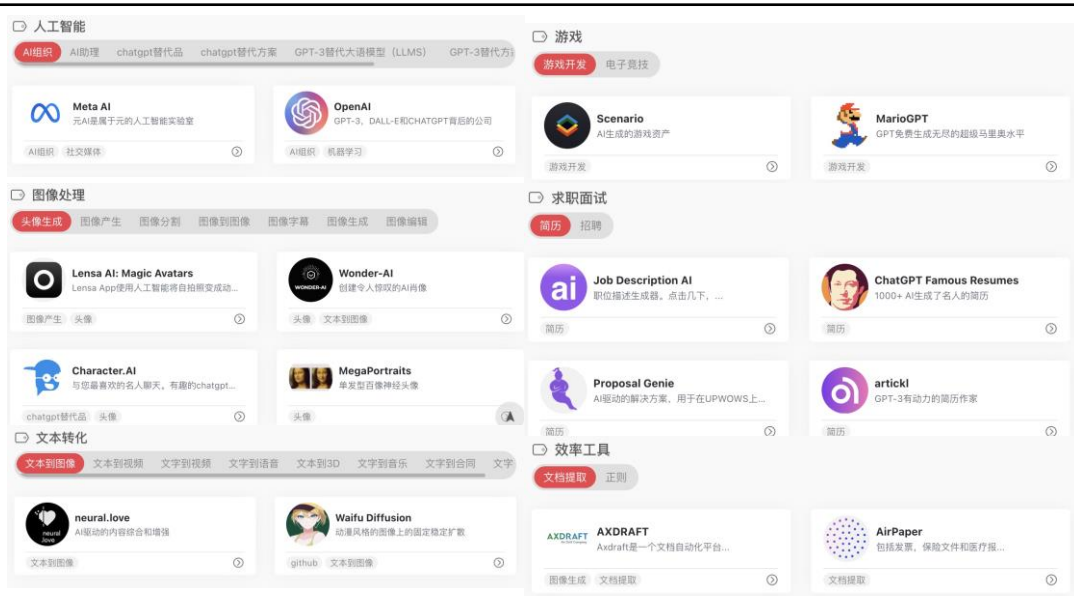
Download the Model



1.2 海外AI应用持续落地，商业模式得到验证

- ◆ **海外大模型龙头企业开启AI建设，落地应用爆发性增加。**例如：1) 众多应用已接入GPT-4，涉及文案处理、AI助理、游戏开发、图像处理、效率工具等多个领域。2) 谷歌将承载PaLM 2大模型的聊天机器人Bard整合到办公软件“全家桶”中，应用包含Gmail、谷歌Docs、Slides等。3) Cohere与甲骨文合作帮助企业组建端对端的业务流程。4) Meta LLaMA2**大模型**融入Amazon Web Services，进一步助力商业化人工智能应用落地。
- ◆ **海外AI应用落地趋势明显，部分应用活跃度高速上涨。**根据SimilarWeb数据，七月中旬（即7月9日-7月15日）多款AI应用访问量呈现高速增长速度。其中Tome（教育辅助）、NovelAI（小说生成工具）、Jasper（文案生成及处理）、Midjourney（图像生成）、CodeWhisperer（编程辅助）、CharacterAI（互动聊天）、Shopify（商家客服咨询）的周访问量**同比增长**18.73%、16.00%、12.93%、10.24%、5.60%、5.58%、3.44%。

部分接入GPT-4的应用



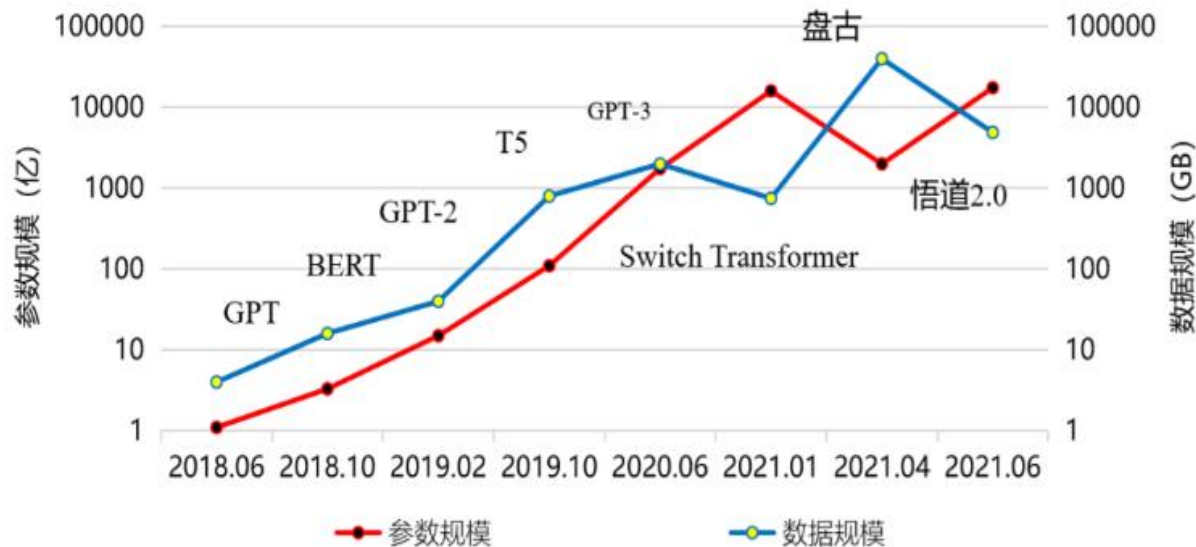
七月第二周高速增长的AI应用盘点

应用名称	应用功能	7月9-7月15访问量 (百万)	同比增长	7月2-7月8访问量 (百万)
Tome	教育辅助	3.626	18.73%	3.054
NovelAI	小说生成工具	1.102	16.00%	0.950
Jasper	文案生成工具	1.022	12.93%	0.905
Midjourney	图像生成工具	6.780	10.24%	6.150
CodeWhisperer	编程辅助工具	15.840	5.60%	15.000
Character AI	互动式聊天	45.810	5.58%	43.390
Shopify	客服咨询（销售）	25.550	3.44%	24.700

1.3 ChatGPT开启大模型“军备赛”，算力呈现明显缺口

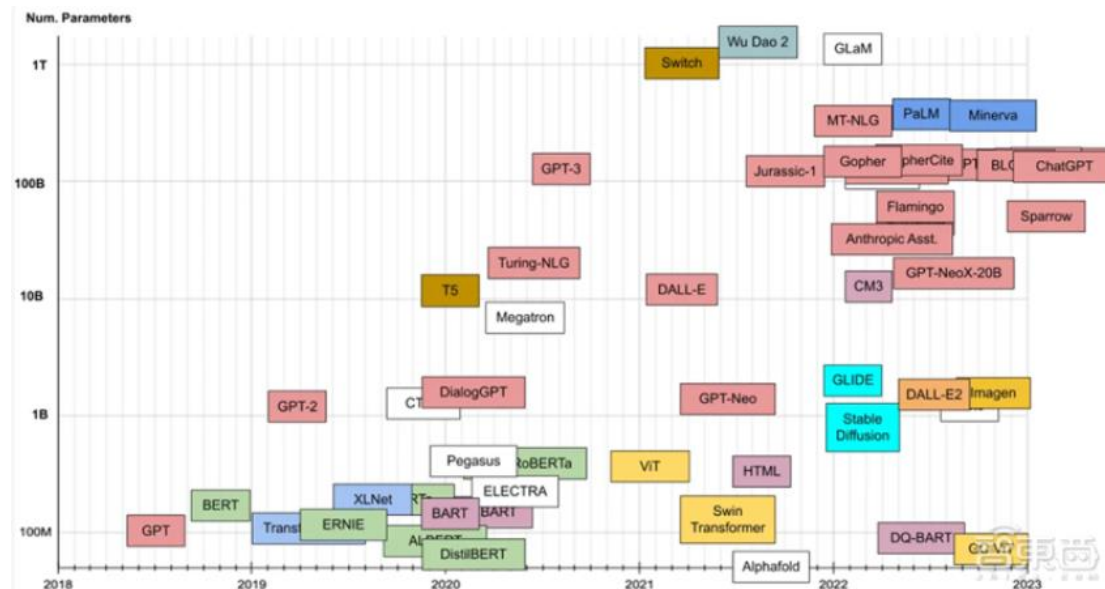
- ◆ **ChatGPT开启算力军备赛**：我们认为数据、平台、算力是打造大模型生态的必备基础，且算力是训练大模型的底层动力源泉，一个优秀的算力底座在大模型(AI算法)的训练和推理具备效率优势；同时，算力是AI技术角逐“入场券”，其中AI服务器、AI芯片等为核心产品；此外，以英伟达为代表的科技公司正在快速补足全球AI算力需求，为大模型增添必备“燃料”。**我们认为目前全球科技企业正处于算法和应用正在快速落地的时间内，算力短期呈现爆发性。**
- ◆ **大模型参数呈现指数规模，引爆海量算力需求**：根据财联社和OpenAI数据，ChatGPT浪潮下算力缺口巨大，根据OpenAI数据，模型计算量增长速度远超人工智能硬件算力增长速度，存在万倍差距。运算规模的增长，带动了对AI训练芯片单点算力提升的需求，并对数据传输速度提出了更高的要求。根据智东西数据，过去五年，大模型发展呈现指数级别，部分大模型已达万亿级别，因此对算力需求也随之攀升。

大模型参数数量和训练数据规模快速增长



资料来源：新浪，智东西，可创办日报，华

近年大模型的参数规模增长趋势



1.4 英伟达，全球AI算力王者归来

- ◆ **AI大模型时代拉开帷幕，对AI芯片需求量明显增多。** AI从2012年发展至今，随着模型参数量越来越大，对于计算量的要求也逐步增高。2022年11月份OpenAI发布ChatGPT，以ChatGPT为代表的大模型持续引爆市场。早在2019年，微软斥资几亿美元为Open AI的训练打造一台超级服务器，其中包括上万张英伟达A100，旨意为ChatGPT和New Bing提供算力基础，而相关大模型的火爆，AI硬件竞争持续升温，芯片遭“哄抢”导致价格大涨，根据界面新闻，英伟达AI旗舰芯片H100已明显提价。
- ◆ **AI芯片业务大超预期：**根据公司2024财年发布会，公司AI芯片所在数据中心业务收入达42.8亿美元(环比+18%，同比+14%)，创历史新高,二季度营收预期约110亿美元，同比增近33%；毛利率超64%。业绩会上英伟达称，众多云公司竞相部署AI芯片，其锁定了数据中心芯片的大幅增长，计划下半年大幅增加供应。
- ◆ **英伟达持续发布全新产品，助力全球算力:** 英伟达今年陆续发布AI Foundations，DGX超级计算机，全新RTX4070，以及DGX GH200 AI超级计算机等全新AI产品。我们认为此次AI浪潮不同于元宇宙阶段，大模型已经产生相关落地应用，相关大模型的火热势必对算力产生超高需求，英伟达作为全球算力龙头深度受益。

英伟达2023年相关AI产品



1.4 英伟达，全球AI算力王者归来

- ◆ 2023年3月23日GTC会议，英伟达全新AI相关产品助力全球AI生态。1) 基础软件：推出全新加速库；2) 芯片方面：推出数据中心 Grace CPU，具备高能效、高运行速度等优势；3) 服务器：推出DGX超级计算机；4) **全新AI服务平台（DGX云与生成式AI服务）**，AI的“Iphone”时刻已经来临，**AI foundations 云服务**能够构建、改进和操作定制的大型语言模型和生成式 AI 模型，助力初创企业具备拥有生成式AI的能力，且已经具备多种生成式AI模型和相应案例。
- ◆ 平台实为模型和算力之间的“桥梁”，是AIGC或大模型生成的必备要素，不论是数据库还是编译器，都需要通过平台来实现资源的合理分配以达到软硬件的最优组合，从而大幅提升模型效率。平台通过调用数据包来适配软硬件之间的结构，来达到模型的最优组合，从而提升模型乃至整个虚拟机的效率。

英伟达加速计算库



英伟达DGX H100



英伟达AI Foundations能力示意图

文本生成

用于营销文案和故事情节创作

总结

用于新闻、电子邮件、会议纪要和信息合成

聊天机器人

实现智能 Q&A 和实时客户支持

图像生成

用于品牌创建和游戏角色

编码

用于动态评论和函数生成

翻译

用于多种语言的全球协作

蛋白质预测

用于 3D 蛋白质结构和属性预测

生物分子生成

用于新型蛋白质序列和小分子生成

分子对接

用于小分子和目标蛋白之间相互作用的建模

1.4 英伟达，全球AI算力王者归来

- ◆ **2023年5月30日COMPUTEX主题演讲，英伟达展示多款新系统、软件及服务，让生成式AI触手可及，革新了广告、制造、电信等行业**
- ✓ **DGX GH200 AI超级计算机**：由NVIDIA GH200 Grace Hopper超级芯片和NVIDIA NVLink Switch System驱动，相比上一代将NVLink带宽提升了48倍以上。据腾讯网消息，谷歌云、Meta和微软是首批有望接入DGX GH200探索其用于生成式AI工作负载能力的公司。英伟达还计划将DGX GH200设计作为蓝图提供给云服务提供商等。
- ✓ **模块化参考架构NVIDIA MGX**：能够灵活地兼容多代NVIDIA产品，制造商可以使用通用架构和模块化组件构建CPU和加速服务器。
- ✓ **NVIDIA Spectrum-X网络平台**：是全球首款专为AI网络打造的51Tb/s以太网交换机，提高了基于以太网AI云的性能与效率，与现有以太网的堆栈实现互通。单台交换机即可实现突破性的256个200Gb/s端口的连接，以支持AI云的增长和扩展。
- ✓ **ACE游戏开发版**：在Omniverse的基础上，“ACE游戏开发版”为语音、对话和角色动画提供优化的AI基础模型。其代工服务协助开发者微调游戏模型，然后通过 NVIDIA DGX Cloud，GeForce RTX PC 或现场加以部署，以实时进行推理。在客户端侧，NVIDIA和微软合作，将加强1亿台配备RTX GPU的PC性能，这些GPU中的Tensor Core可大幅提升400多个AI加速应用和游戏的性能。
- ✓ **NVIDIA Omniverse的虚实融合**：在数字广告领域，全球最大营销服务机构WPP正与NVIDIA一起在Omniverse Cloud上构建首个生成式AI内容引擎；在工业制造领域，企业可通过Omniverse和生成式AI 的API接口，构建工厂数字孪生。其他已运用相关产品的企业包括和硕、富士康、Techman Robot等

英伟达DGX GH200 AI超级计算机



英伟达“ACE游戏开发版”构建NPC场景



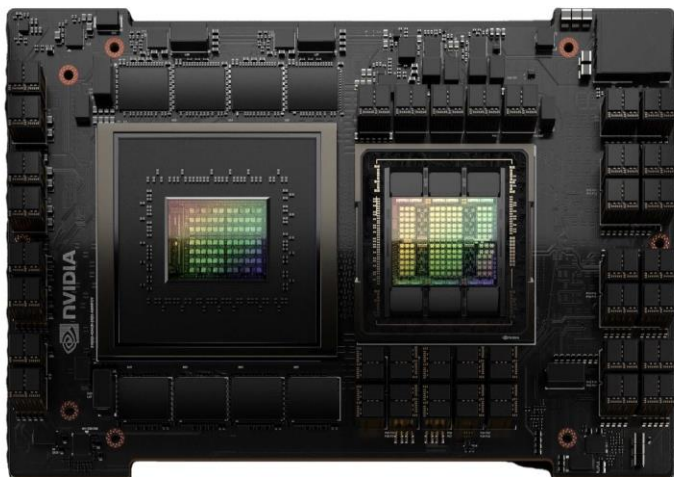
英伟达最新组件Metropolis for Factories



1.4 英伟达，全球AI算力王者归来

- ◆ **芯片方向**：发布应用于大规模人工智能和高性能计算应用程序的突破性加速CPU+GPU—NVIDIA Grace Hopper超级芯片，该超级芯片可以为运行TB级数据的应用程序提供高达10倍的性能提升，使科学家和研究人员有能力解决更为复杂的问题。
- ◆ **显卡方向**：NVIDIA TITIAN RTX显卡，它由Turing（图灵）架构提供支持，为用户的PC带来130 Tensor TFLOPs的性能，576个Tensor核心和24GB超快GDDR6内存，并且加入了加速AI和光线追踪的最新Tensor Core和RT Core技术。
- ◆ **服务器方向**：DGX超级计算机，配有 8 个 H100 GPU 模组，H100 配有 Transformer 引擎，旨在处理类似令人惊叹的 ChatGPT 模型，8 个 H100模组通过 NVLINK Switch 彼此相连，以实现全面无阻塞通信。
- ◆ **云服务方向**：NVIDIA DGX平台为企业AI而构建，将NVIDIA软件，基础设施和专业知识的精华结合在一个从云端到本地数据中心的现代统一AI开发解决方案中。DRX平台具有（1）**先进的人工智能开发平台**：除常规应用，包含用于训练模型、优化框架和加速数据科学软件库的NVIDIA AI Enterprise 软件（2）**注入NVIDIA AI 专业知识**：可直接访问NVIDIA DGXperts帮助优化企业的AI工作负责，以获得更高的投资回报率。（3）**前所未有的性能**：为AI基础设施提供清晰的可预测的成本模型。

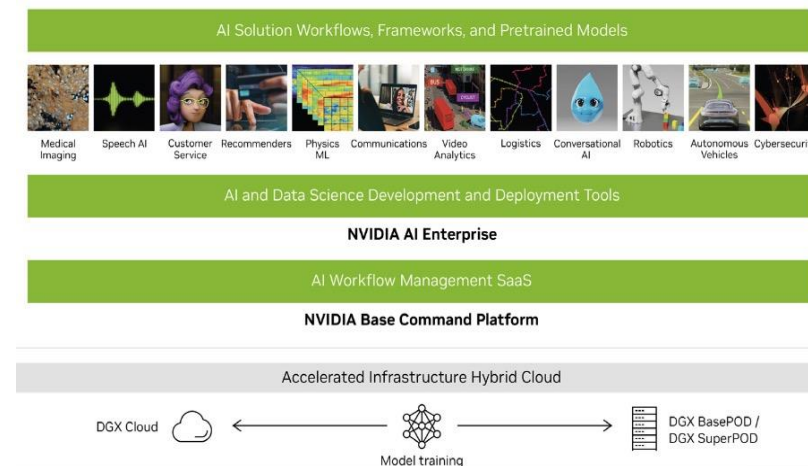
NVIDIA Grace Hopper超级芯片



NVIDIA TITIAN RTX 显卡



NVIDIA DGX云平台概念图



1.5 三个月内两度涨价，英伟达涨价逻辑拆析

- ◆ 5月14日，据集微网消息，英伟达A100价格从去年12月开始上涨，截至今年4月上半月，其5个月价格累计涨幅达到37.5%；同期A800价格累计涨幅达20.0%。
- ◆ 我们认为，第一轮涨价逻辑如下：
 - ✓ 1) GPT催生AI算力需求爆发。随着ChatGPT带来AI产业大热，相关产业对AI算力硬件的需求也同步高涨。据前人工智能NLP企业首席科学家、千芯科技董事长陈巍测算，国内想要直接训练出一个GPT-3级别的大模型，研发训练层面最少需要3000到5000枚A100级别的AI芯片。随着国内诸如阿里巴巴、商汤科技等各类公司对大模型的持续深耕，保守估计国内A100级别的AI芯片缺口在30万枚左右。
 - ✓ 2) 台积电供应侧芯片产能受限。由于台积电独有CoWoS封装技术，英伟达目前将GPU旗舰产品芯片独家由台积电代工生产，仅其余少量产品由三星代工完成。但由于AI浪潮引发了上游高端芯片需求的暴增，以往英伟达GPU的交付时间仅需3个月左右，目前新订单要推迟到年底，等待时间超过半年，甚至1年。同时目前英伟达芯片已经没有库存，H100和A100高算力GPU供不应求。

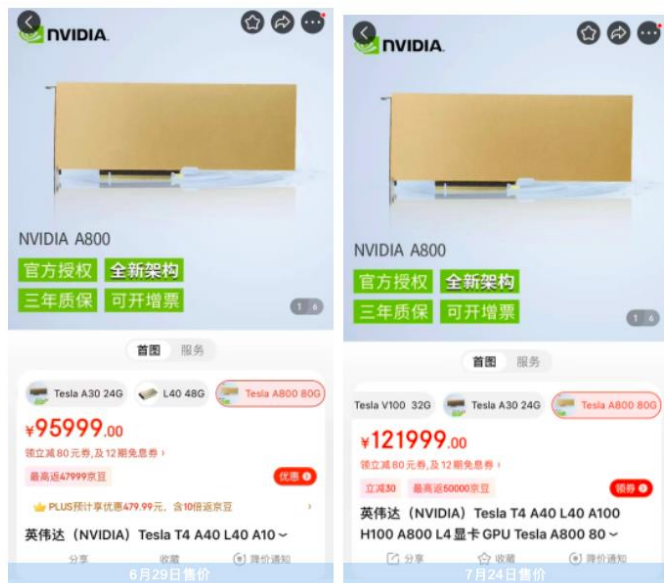
英伟达产品实现云、边、端全面布局



1.5 三个月内两度涨价，英伟达涨价逻辑拆析

- ◆ 7月13日，据财联社消息，英伟达A800的售价已较一周前上涨约30%，单卡现货已达到近13万元一颗。同时今年6月开始，H800正式大规模推广，目前一颗H800的单卡GPU价格已高达20余万元，远高于涨价后的A800。
- ◆ 我们认为，第二轮涨价原因主要在于A800、H800芯片供不应求，稀缺性强。根据Techweb消息，去年美国BIS颁发《出口管制条例》，为满足条例规定，英伟达调整技术参数，为中国厂商特供两者的替代品A800和H800。而由于H800 今年 3 月才开始投产，以当下的排产进度预计，部分 H800 要到今年底才能交货，导致目前芯片稀缺。政策是影响A800类芯片供应紧张程度的关键，可以预见，国内AI浪潮将推动A800类芯片的行情持续上涨。
- ◆ 综合英伟达近期两度涨价逻辑可以看出，英伟达A800等高端芯片价格持续大幅涨价象征着以英伟达为首的算力芯片依旧是供不应求，算力芯片依旧为大模型时代的稀缺要素。

英伟达A800持续涨价



美国商务部工业和安全局发布半导体出口管制新规



FOR IMMEDIATE RELEASE
October 7, 2022
www.bis.doc.gov

BUREAU OF INDUSTRY AND SECURITY
Office of Congressional and Public Affairs
OCPA@bis.doc.gov

Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)

Washington, D.C.—The Department of Commerce’s Bureau of Industry and Security (BIS) is implementing a series of targeted updates to its export controls as part of BIS’s ongoing efforts to protect U.S. national security and foreign policy interests. These updates will restrict the People’s Republic of China’s (PRC’s) ability to both purchase and manufacture certain high-end chips used in military applications and build on prior policies, company-specific actions, and less public regulatory, legal, and enforcement actions taken by BIS.

1.6 台积电产能供不应求，AI有望成为其未来支柱

◆ 英伟达与台积电共创AI时代硬件基础。

- ✓ 在ChatGPT掀起的AI“淘金热”之下，英伟达的AI GPU几乎是所有大模型厂商的选择，而这背后也离不开台积电的高水平制造工艺。
- ✓ 技术上最核心的原因是：V100、A100 以及最新的 H100 AI算卡无一不采用台积电 CoWoS 先进封装技术，用来解决高算力 AI 背景下芯片的存算一体。而这项CoWoS 技术也无法离开台积电的支持。

◆ AI浪潮下，台积电几乎独占英伟达AI芯片制造。

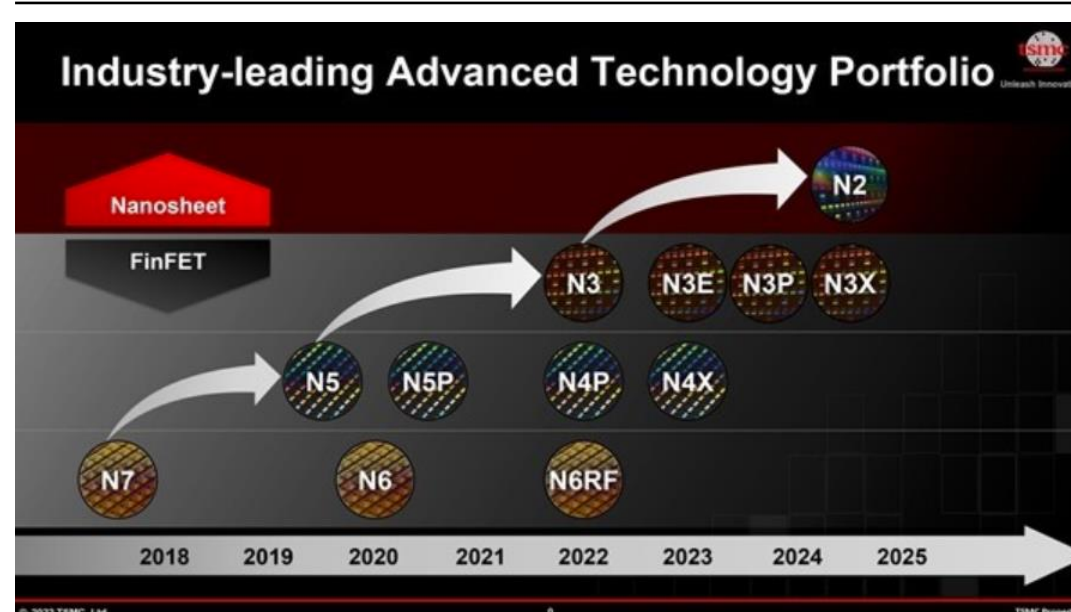
- ✓ 根据36氪报道，英伟达对台积电的订单需求急剧增加，其中包括基于台积电7nm的A100芯片和4nm工艺的H100芯片。并且，台积电的5/4纳米工艺生产线几乎已经满负荷运转，而英伟达大量的最急件订单将持续一年时间。

英伟达芯片发展和制程迭代

年份	架构	制程	晶体管数量
2010年	Fermi	40纳米	30亿
2012年	Kepler	28纳米	71亿
2014年	Maxwell	28纳米	80亿
2016年	Pascal	16纳米	153亿
2017年	Volta	12纳米	211亿
2018年	Turing	12纳米	186亿
2020年	Ampere	7纳米	283亿
2022年	Hopper	4纳米	800亿

资料来源：华西证券研究所

台积电制程技术迭代历程



1.6 台积电产能供不应求，AI有望成为其未来支柱

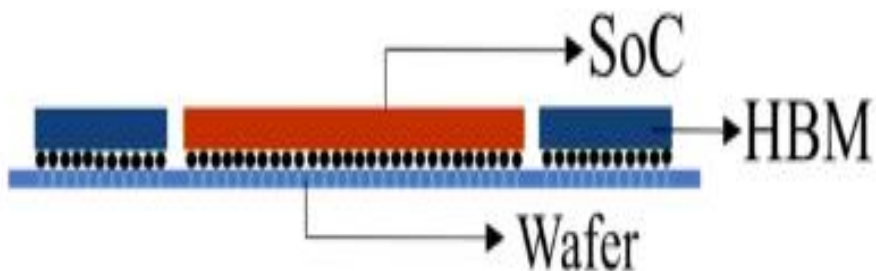
◆ 台积电具备先进封装技术的先发优势。

- ✓ 2012 年，台积电首次推出独家的 CoWoS 先进封装技术，实现了从晶圆代工到终端封装的一条龙服务，包括英伟达、苹果和 AMD 的核心产品都依赖台积电先进制程及其先进封装技术。

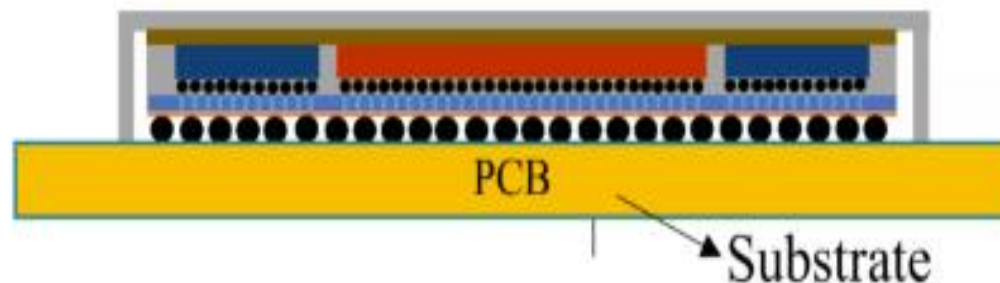
◆ CoWoS是台积电最强护城河技术。

- ✓ "CoWoS"是一种半导体封装技术，属于"2.5D"封装技术之一。该技术可以将多个有源硅芯片（通常包括逻辑和HBM堆栈）集成在无源硅中介层上，实现高度集成和优良性能。中介层充当顶部有源芯片的通信层，实现不同芯片之间的数据传输。然后再将内插器和有源硅连接到封装基板上，将芯片与系统PCB上的I/O连接起来，实现完整的半导体封装。
- ✓ 在封装过程中，芯片以 3D 方式堆叠在单个封装中，可以缩短芯片之间的间距，使得芯片间的连接速度更快，可以带来高达 50%或更多的巨大性能提升。

Chip on Wafer示意图



CoW芯片与基板（Substrate）连接示意图



1.6 台积电产能供不应求，AI有望成为其未来支柱

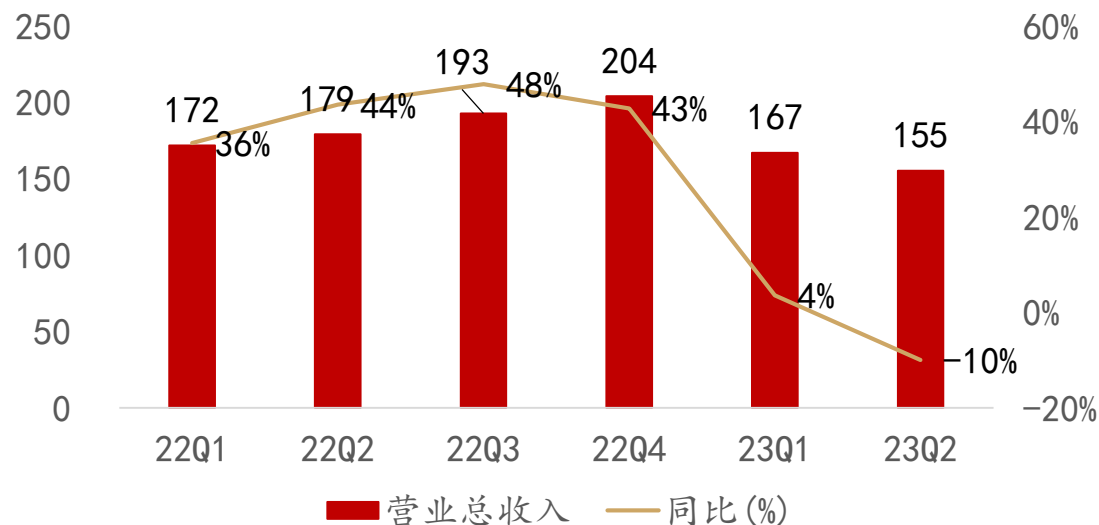
◆ 受宏观经济影响，台积电下调23全年营收预期。

- ✓ 7月20日，台积电发布业绩快报，23H1实现营收319.47亿美元，同比下降7.1%；净利润125.53亿美元，同比下降14.9%。对此，台积电下调全年营收增速指引至-10%。

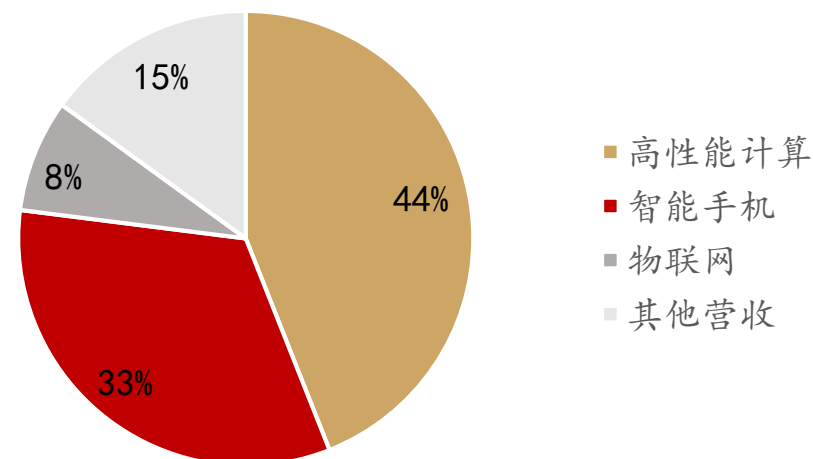
◆ AI服务器需求强劲，有望成为未来支柱领域。

- ✓ 台积电表示，5G和HPC行业的增长趋势推动了计算需求的大规模结构性增长，同时对性能和能效计算提出了更高的要求，这促使前沿技术成为必要条件，公司预计21-26年将实现营收15-20%的CAGR增长。其中，23Q2AI服务器处理器相关需求约占总收入的6%左右，预计未来5年收入将实现**50%的CAGR增长**，收入占比将扩大到10%左右。

22Q1-23Q2营收情况（亿元）



23Q2台积电营收占比（按平台）



1.6 台积电产能供不应求，AI有望成为其未来支柱

- ◆ **台积电5/4nm产线接近满载，产能不足是关键难题。**
 - ✓ 台积电总裁魏哲家在公司法说会上表示，人工智能迸发式的发展为台积电带来了强劲的需求。尤其是在先进封装领域，特别是台积电的晶圆衬底上芯片（CoWoS）产能相当紧张。预计这种紧缩措施将在明年年底得到缓解。
 - ✓ 此外，台积电在亚利桑那州的圆晶厂由于能熟练安装设备的专业人员数量不足，量产时间推迟至2025年。
- ◆ **应对AI芯片需求，台积电加大资本支出。**公司为应对英伟达、AMD 等厂商对 AI 芯片 CoWoS 等先进封装的需求，扩大对先进封装制造厂的建设。此外，台积电也决定在资本支出中加重在 CoWoS 先进封装产能的建设。

台积电近年建厂进度汇总

地区	建设情况
美国	台积电在2020年5月15日宣布在美国亚利桑那州建造一座芯片工厂的该工厂于2021年6月份开始动工建设，2022年夏天封顶，计划2024年开始量产4纳米芯片，而不是之前计划的5纳米芯片，规划月产能为2万片晶圆。由于工程和设备安装进度延迟、人力资源短缺和成本紧张，该工厂不太可能在2024年全面投产，可能会推迟到2025年。
日本	2022年4月台积电与日本索尼及日本电装合资的晶圆代工厂JASM，已于21日在日本熊本县菊阳町开始动工，新工厂预定投资86亿美元，满产后每月可量产5.5万片28nm及更先进制程12吋晶圆。计划的初期预估资本支出约70亿美，2022年动工，2024年底开始生产22/28nm制程12吋晶圆，月产能约为4.5万片。随着日本电装的加入，该晶圆厂的预定资本支出提高到了86亿美元，同时提供12 / 16nm鳍式场效晶体管（FinFET）制程的专业的芯片制造服务，并将月产能提高至5.5万片12吋晶圆。
欧洲	台积电计划在德国德累斯顿市建立其第一家欧洲工厂。
中国南京	2023年7月25日，台积电正式确认，因应英伟达、AMD 等厂商对 AI 芯片 CoWoS 等先进封装的需求，将斥资 900 亿新台币（约 206.1 亿元人民币）在苗栗县铜锣乡兴建先进封装厂。新竹科学园区管理局已同意核拨土 7 公顷土地，预定 2026 年底完成建厂，2027 年第 3 季量产这将是台积电继龙潭、竹南、南科后第六座封装生产据点。
中国台湾	2023年7月20日，台积电在法说会上表示正按计划在南京扩充28nm制程产能。

资料来源：36氪，TechWeb，腾讯新闻，澎湃

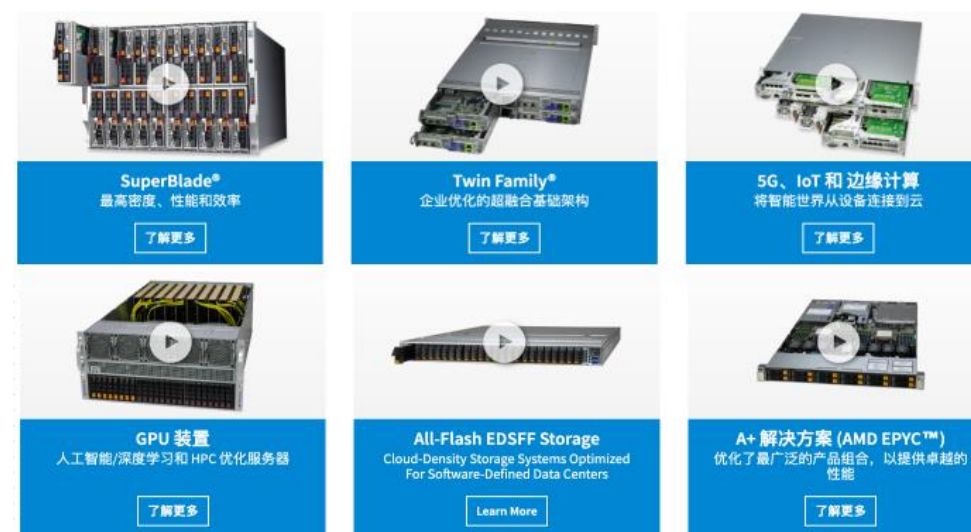
1.7 英伟达芯片相继落地，相关服务器厂商业绩有望高度景气

- ◆ **乘AI之风，超微电脑2023第四财季业绩超预期**：根据财报狗消息，公司第四财季营收21.8亿美元，同比增长34%，环比增长70%；超出了公司17-19亿美元的最初指导范围。2023财年收入为71.2亿美元，同比增长37%。根据财报会议，超微电脑明确表示，第四季度是公司强劲的AI/GPU和机架规模解决方案推动的，占总体收入的52%，且根据财报狗消息，公司强调24财年盈利能力预计持续增长，预计数据中心、新兴通信服务器提供商、企业人工智能扩建等将退工公司多元化增长。
- ◆ **与英伟达深度合作，开发新款AI服务器平台**：超微电脑是一家应用优化服务器解决方案全球领导厂商，涵盖的应用市场包括人工智慧(AI)、物联网(IOT)、服务器、储存、软件以及绿色运算解决方案。同时，据科创板日报6月2日消息，超微电脑创始人、CEO梁见后指出，市场对AI需求强劲，公司正扩大美国、荷兰等地产能，公司在马来西亚及日本都设有服务器生产基地，预计年底前要把4000台机柜产能拉升至5000台。

超微电脑公司官网



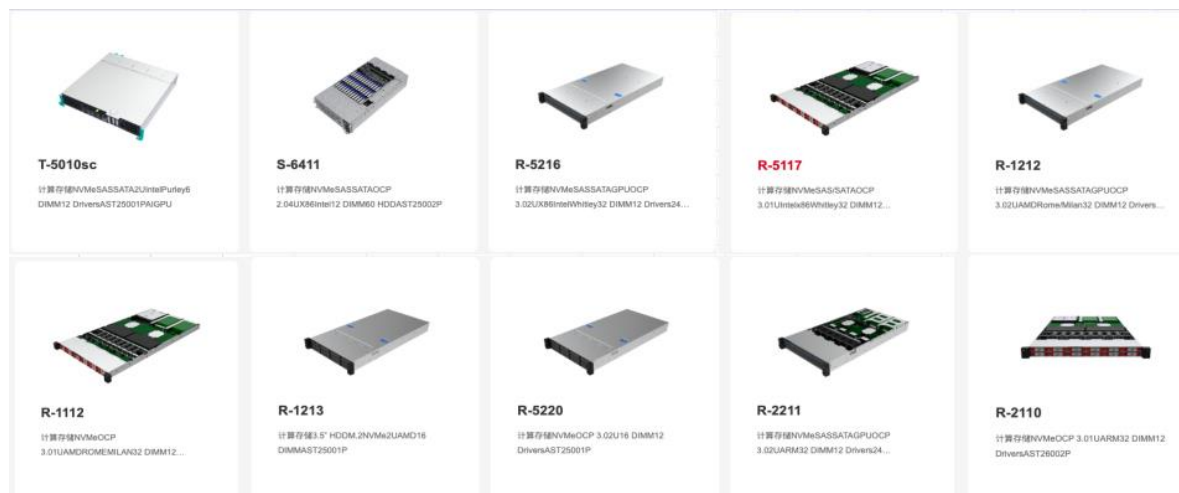
超微电脑产品



1.7 英伟达芯片相继落地，相关服务器厂商业绩有望高度景气

- ◆ **AI服务器需求旺盛，中国台湾多家电子代工厂争相出货有望业绩兑现。**我们认为随着英伟达芯片下半年持续供货，中国台湾相应AI服务器厂商有望拉动出货量上涨，使得业绩得到兑现。
- ✓ **鸿海有望利用AI服务器重振增长。**目前鸿海全球服务器市占率超过四成，是全球第一大AI服务器供应商。鸿海主要通过旗下工业富联与鸿佰科技进行AI服务器业务。鸿海董事长刘扬伟认为，随着ChatGPT使用率的上升，AI服务器市场下半年可能实现三位数增长。目前，公司由于智能手机市场的下滑，已经连续三个季度报告利润不及预期。董事长刘扬伟今年5月在年度股东大会上表示，鸿海正与英伟达在自动驾驶应用方面展开合作，预计下半年AI服务器销售收入将至少增长一倍。
- ✓ **纬创AI服务器出货持续带动盈利高升。**纬创7月3日宣布，公司6月单月营收赚赢今年第1季度，达到810.99 亿元。据电子时报报道，相关供应链表示，作为英伟达芯片的主要系统代工厂商，下半年纬创在A100的代工比重还将上升。

工业富联高性能服务器产品



纬创集团旗下纬颖服务器产品



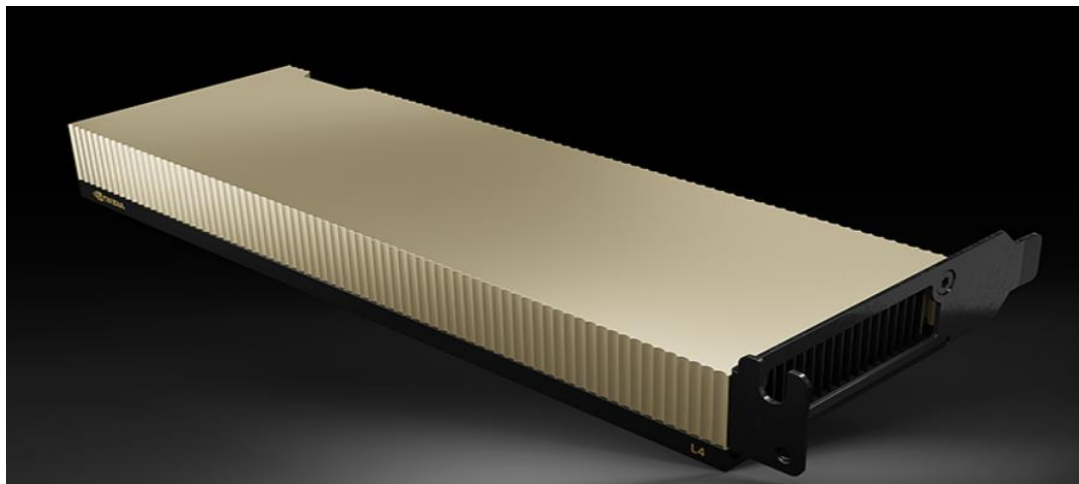


02 国内算力，拐点之时

2.1 新一轮AIGC爆发，英伟达芯片持续落地

- ◆ **英伟达芯片在全球范围内受到了广泛的关注和认可，具有极强的竞争力，长期占据高端GPU市场的领导地位**，TrendForce指出，目前AI服务器市场上，英伟达的GPU已成为主流芯片，市场份额约占60-70%。
- ◆ **与海外大厂合作，芯片落地应用范围广泛**。22年11月，微软和英伟达宣布建立多年的合作关系，为 Azure 客户开发新型的 AI人工智能云端超级计算机，由英伟达 GPU 技术驱动；23年3月，英伟达与谷歌合作推出一款新的云硬件产品，L4平台，作为通用 GPU，L4 提供视频解码以及转码和视频流功能，经过优化可以运行以视频为中心的应用程序；5月，NVIDIA宣布，GH200 Grace Hopper超级芯片已经全面投产，将为复杂AI、HPC工作负载提供澎湃的动力，欧洲和美国的超大规模云服务商、超算中心，将会成为接入GH200系统的首批客户。7月，特斯拉首席执行官马斯克表示，特斯拉对英伟达硬件的需求十分巨大，以至于其无法满足所需。

NVIDIA L4 Tensor Core GPU



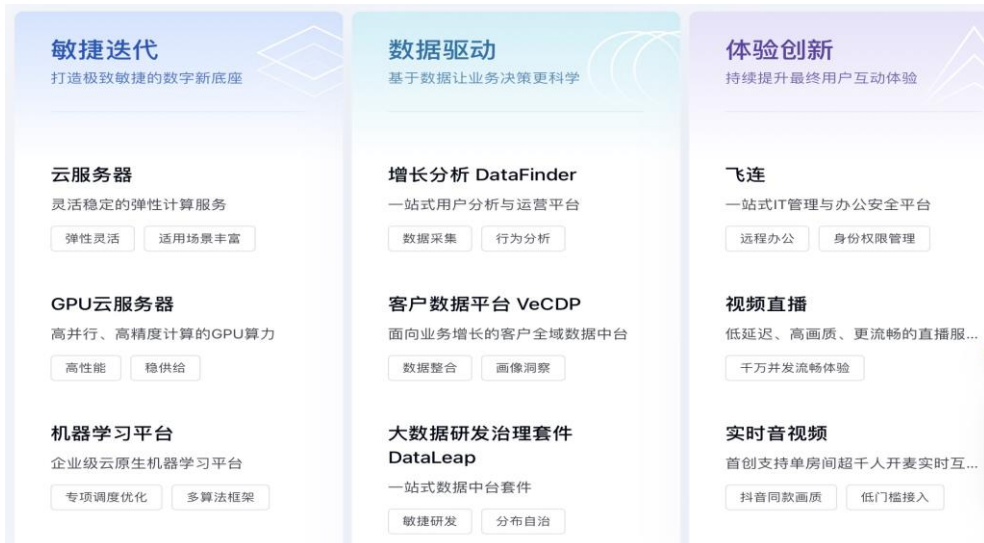
NVIDIA DGX GH200



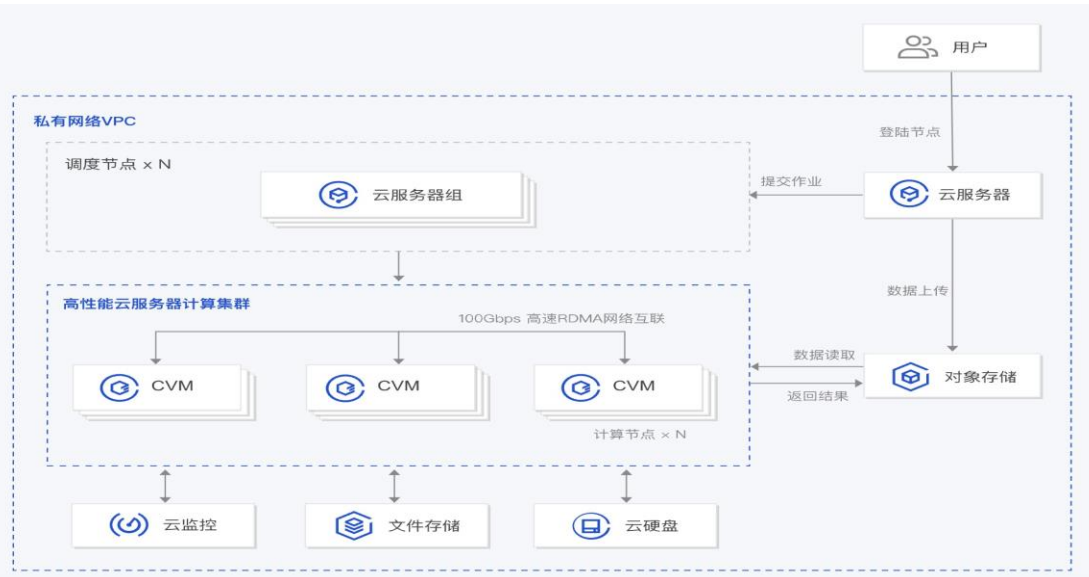
2.1 新一轮AIGC爆发，英伟达芯片持续落地

- ◆ 根据ICC讯消息，英伟达提供专供中国市场的A800和H800的芯片，互联网厂商正争相采购。
- ◆ 腾讯云是腾讯公司旗下的产品，为开发者及企业提供云服务、云数据、云运营等整体一站式服务方案。今年腾讯云发布搭载英伟达H800 Tensor Core GPU的新一代腾讯云星海自研服务器。进而推出国内首个嵌入英伟达H800芯片的面向大模型训练的新一代HCC高性能计算集群。根据实测结果显示，该集群的算力性能较前代提升高达3倍，是目前国内性能最强的大模型计算集群。
- ◆ 字节跳动，据《晚点LatePost》报道，今年向英伟达订购了超过10亿美元的GPU，AI相关加速卡总计10万块，将会应用到字节旗下产品火山引擎中，为火山引擎的企业客户提供充足的算力支持，提高大模型的研发效率，降低业务成本。
- ◆ 优刻得致力于云计算服务平台的研发与应用。根据Ucloud公众号，公司表示为大模型训练提供充足算力，未来或会将A800部署到自建的数据中心中。另外，还表示已从英伟达采购其他相关产品包含网卡、交换机等。

字节跳动火山引擎部分功能



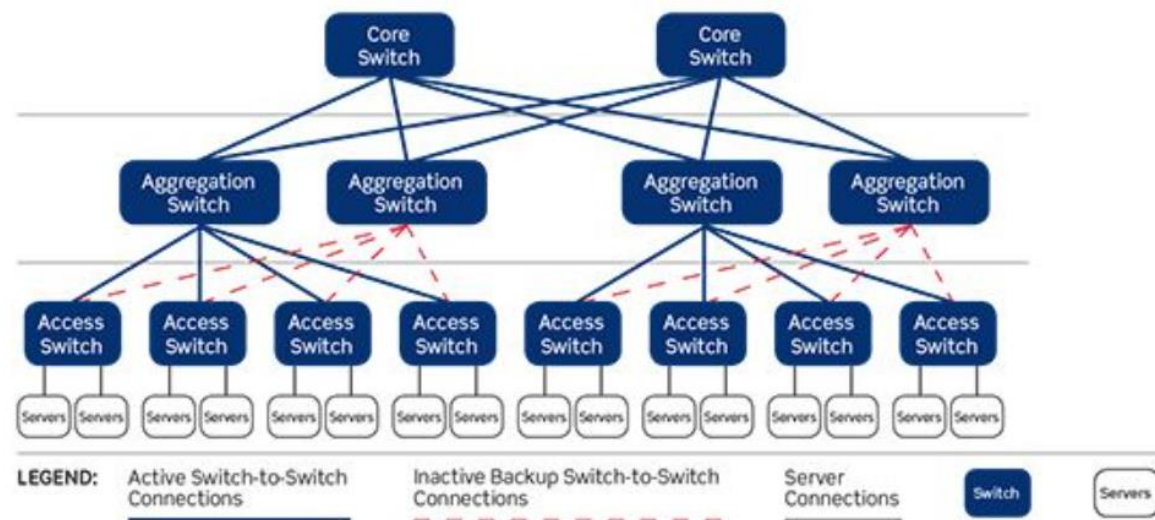
腾讯云新一代HCC高性能计算集群相关产品架构



2.2 算力是AIGC落地的先行指标

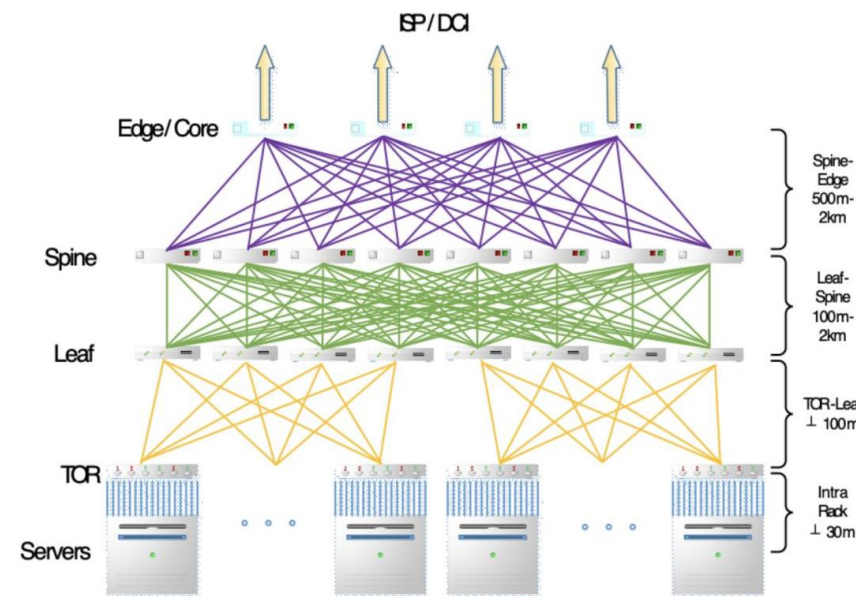
- ◆ **算力是AIGC落地的先行指标**：我们认为全球正处于AIGC的爆发期，如果说AIGC是科技企业开启第二轮业绩增长曲线的“流量入口”，那么算力即是科技厂商开启AI争夺战的“入场券”，目前国内外科技巨头持续布局算力储备，此外从英伟达芯片持续落地、涨价与台积电扩产的逻辑背后来看，AI算力芯片依旧是供不应求，是AI新时代的稀缺产品。
- ◆ **我国正处于处于智算中心建设的加速期**：以海外科技巨头为例，相关AI产品和模型部分已经落地，商业模式已得到验证，而我国相关AI应用目前已有雏形。此外，在我国AIGC短期应用爆发的趋势下，算力建设同样也是印证本轮景气度的先行指标，目前我国正处于从“智算”到“超算”跨越的新时代，我国正处于智能算力建设的加速期。
- ◆ **相关AI算力产品有望成为本轮AI浪潮下的先行指标**：我们认为无论是英伟达算力芯片落地角度还是我国智能算力建设加速角度，相关算力产品有望成为本轮景气度的先行指标，从数据中心架构示意图中看，相关的算力产品包括算力服务器、光模块、交换机等产品，同时我们认为相关公司在本轮科技浪潮中业绩有望高度景气。

数据中心架构示意图



资料来源：京东云，华西证券研究所

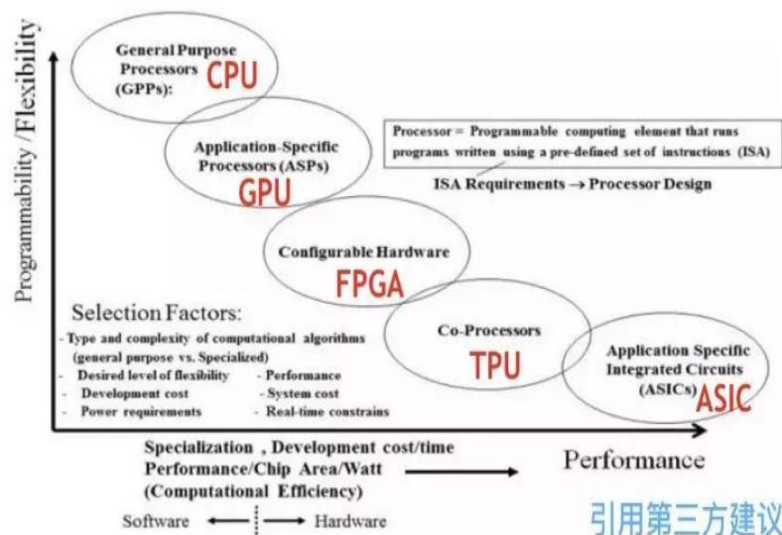
京东云数据中心网络结构图



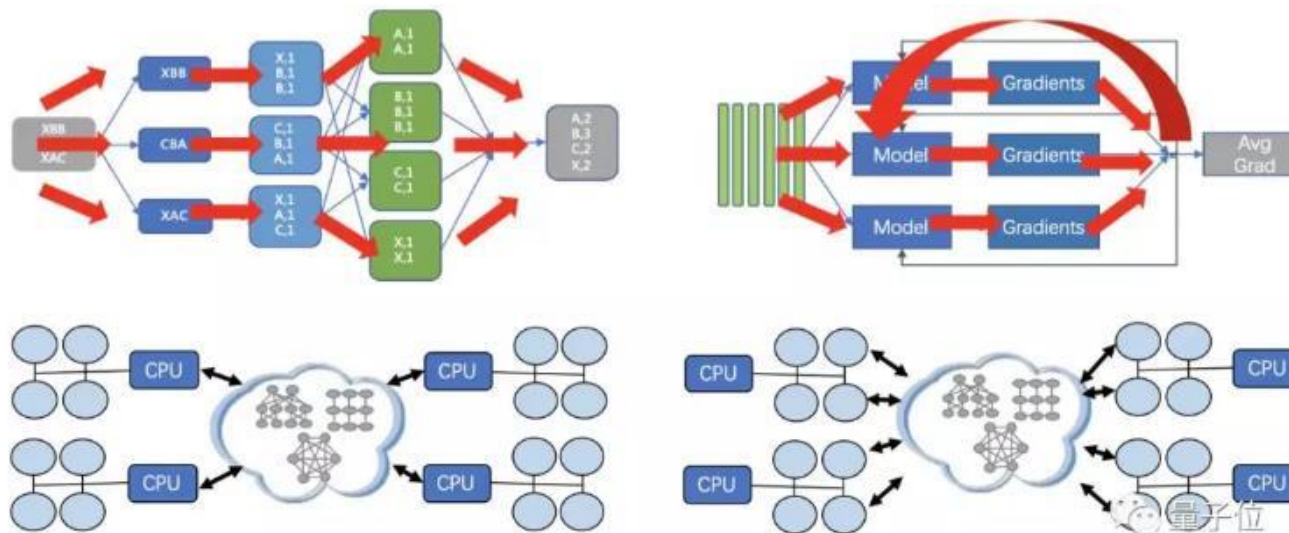
2.2.1 AI加速服务器时代已经到来

- ◆ **AI芯片是AI算力的“心脏”，GPU价值凸显：**伴随数据海量增长，算法模型趋向复杂，处理对象异构，计算性能要求高，AI 芯片在人工智能的算法和应用上做针对性设计，可高效处理人工智能应用中日渐多样繁杂的计算任务。在人工智能不断扩大渗透的数字时代，芯片多元化展现出广阔的应用前景，通过不断演进的架构，为下一代计算提供源源不断的动力源泉。
- ◆ **GPU作为AI芯片的主力军，正在大放异彩：**AI芯片主要包括图形处理器(GPU)、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、神经拟态芯片(NPU)等。人工智能深度学习需要异常强大的并行处理能力，GPU相比于CPU更擅长于并行计算能力，正在大放异彩。根据IDC的数据，2021年H1中国人工智能芯片，GPU占比最多为91.90%。
- ◆ **GPU服务器优势显著：**GPU服务器超强的计算功能可应用于海量数据处理方面的运算，如搜索、大数据推荐、智能输入法等，相较于通用服务器，在数据量和计算量方面具有成倍的效率优势。此外，GPU可作为深度学习的训练平台，优势在于1、GPU 服务器可直接加速计算服务，亦可直接与外界连接通信；2、GPU服务器和云服务器搭配使用，云服务器为主，GPU服务器负责提供计算平台；3、对象存储 COS 可以为 GPU 服务器提供大数据量的云存储服务。

GPU、FPGA、ASIC对比(纵轴代表灵活性、横轴代表性能)



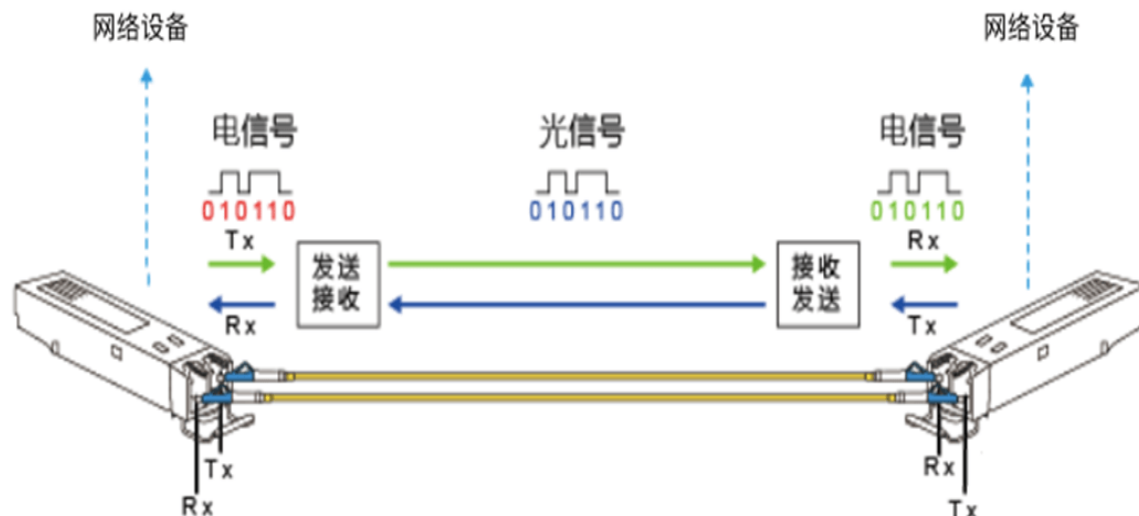
阿里云多路AI集群服务器示意图



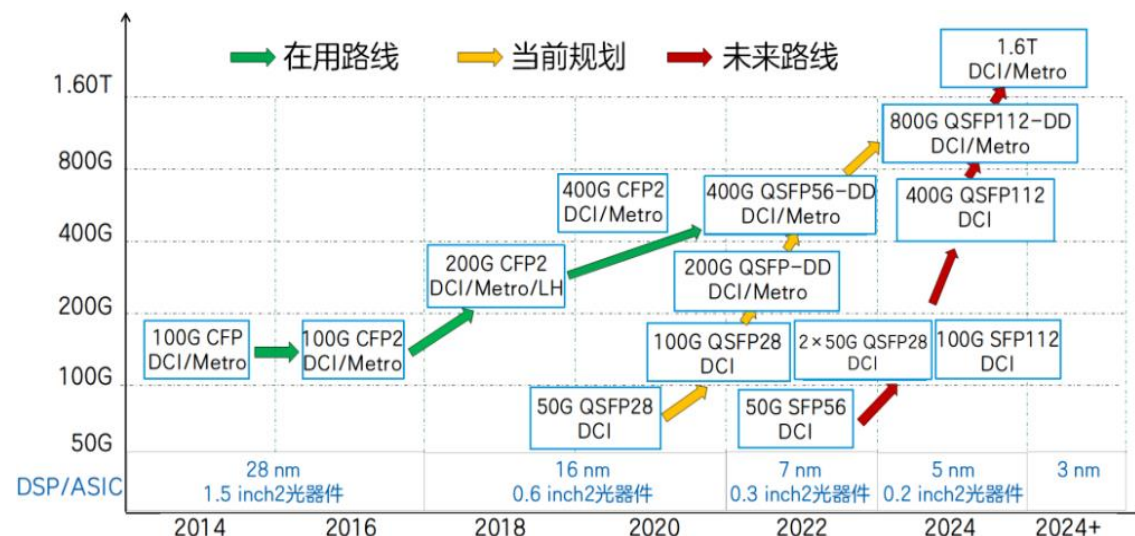
2.2.2 AIGC爆发，驱动光模块持续迭代及放量

- ◆ **光模块是光纤通信中重要组成部分**：是实现光信号传输过程中光电转换和电光转换功能的光电子器件。光模块工作在OSI模型的物理层，是光纤通信系统中的核心器件之一。它主要由光电子器件（光发射器、光接收器）、功能电路和光接口等部分组成，主要作用就是实现光纤通信中的光电转换和电光转换功能。
- ◆ **AIGC爆发，数据中心向智算中心跨跃，光模块有望快速放量**：为满足AI智算中心的高效处理诉求，网络架构持续向低延时高速率演进，同时由于AI大模型是高性能计算，因此强调低延时和高传输速率，因此我们判断在AIGC大背景下，高速率光模块有望快速放量。
- ◆ **AIGC爆发，促使光模块持续迭代**：光模块广泛应用于无线及传输设备，传输速度需配套满足相关算力需求，我们认为伴随着英伟达DGX GH200 AI超级计算机重磅发布，全球算力供不应求的大背景下，AI算力需求呈现指数级增长，因此对光模块的带宽要求和传输速率、覆盖度等指标有更高的要求，光模块有望持续迭代升级和快速放量。

光模块原理图



光模块发展示意图



- 中国移动2023-2024年数据中心交换机产品新建采集情况



2.4.1 算力先行，相关公司业绩已经得到验证——中际旭创

- ◆ 我们认为受益于AIGC爆发，相关科技公司有望业绩持续爆发，我们认为算力是本轮AIGC的先行指标，其景气度有望得到验证。
- ◆ 中际旭创是光模块龙头公司：公司主营业务为高端光通信收发模块以及光器件的研发、生产及销售。其全资子公司苏州旭创致力于高端光通信收发模块的研发、设计、封装、测试和销售，产品服务于云计算数据中心、数据通信、5G 无线网络、电信传输和固网接入等领域的国内外客户，为云数据中心客户提供 100G、200G、400G 和 800G 的高速光模块。
- ◆ Q2业绩稳步增长，AI拉动800G光模块需求高增： 根据公司7月12日业绩快报，公司预计实现归属于上市公司股东的净利润55,000 万元 – 67,000 万元，相较于同期增加11.69% – 36.05%。其中因确认股权激励费用减少公司合并净利润约 6,400 万元。我们认为公司业绩稳步增长的原因在于AIGC爆发，进而拉动了包括光模块在内的通信产品需求的显著增长，并加速了光模块向 800G 及以上产品的迭代。报告期内，得益于 800G 等高端产品出货比重的逐渐增加以及持续降本增效，公司产品毛利率、净利润率进一步得到提升。我们判断，从交付节奏来看，公司下半年和明年相关高端产品出货量有望明显加。

中际旭创业绩快报

项目	本期报告	上年同期
归属于上市公司股东的净利润	比上年同期增加： 11.69%-36.05%	盈利： 49245.52万元
	盈利：55000万元-67000万元	
扣除非经常性损益后的净利润	比上年同期增加： 17.31%-44.38%	盈利： 44327.64万元
	盈利：52000万元-64000万元	

中际旭创光模块相关产品



2.4.2 算力先行，相关公司业绩已经得到验证——中科曙光

- ◆ **中科曙光，AI智算王者**: 中科曙光是我国核心信息基础设施领军企业，公司专注域高端计算机相关业务，打造完整全套的IT设施解决方案。公司基于自身多年技术积累，在高端计算机、存储、云等方面形成了浓厚的技术壁垒。我国正处于“超算”向“智算”跨越的新时代，智算建设有望明显加速，中科曙光作为高性能服务器领军企业，旗下多款产品深度受益。
- ◆ **H1业绩符合预期，AI智算王者迈入高速成长期**: 我们认为公司业绩符合预期，在高基数的同时实现稳步增长，彰显公司在AI浪潮下的强大算力优势，我们认为其主要原因系公司持续专注主营业务的生产经营并不断提高公司整体竞争力，促进业绩持续增长，同时我们预计H2，随着公司上游半导体供应问题有望得到改善，同时伴随AIGC爆发，公司业绩有望迈入高速成长期，并持续高度景气。
- ◆ **公司经营情况稳重向好，业绩环比显著提高**: 公司Q2预计营收32亿元，预计环比增长39.25%，我们判断公司传统服务器业务未受到明显影响受益于信创等相关行业，我们判断公司传统服务器依旧需求强劲。此外，公司属于智能算力建设“主力军”，未来相关智算订单有望快速形成收入，公司业绩有望迈入高速成长期。

中科曙光2023H1业绩快报

项目	本期报告	上年同期	增减变动幅度(%)
营业总收入	5,400,697,692.86	5,031,681,576.96	7.33
营业利润	729,710,770.53	586,807,353.11	24.35
利润总额	733,173,419.35	592,268,447.81	23.79
归属于上市公司股东的净利润	543,960,978.38	473,270,518.55	14.94
归属于上市公司股东的扣除非经常性损益的净利润	318,229,355.91	273,202,650.85	16.48
基本每股收益	0.37	0.33	12.12
加权平均净资产收益率(%)	3.14	3.68	-0.54
	本报告期末	本报告期初	增减变动幅度(%)
总资产	30,373,527,942.66	31,816,791,344.19	-4.54
归属于上市公司股东的所有者权益	17,363,392,197.81	17,017,174,614.14	2.03
股本	1,464,013,974.00	1,464,013,974.00	-
归属于上市公司股东的每股净资产	11.86	11.62	2.07

公司解决方案示意图



资料来源：，中科曙光官网，华西证券研究



03 投资建议：梳理AIGC相关受益厂商

3.1 投资建议: 梳理AIGC的受益厂商

- ◆ 我们认为海外AI应用持续落地，商业模式得到验证，我国相关大模型和应用已有雏形，如果说AIGC是科技企业开启第二轮业绩增长曲线的“流量入口”，那么算力即是科技厂商开启AI争夺战的“入场券”。以科技巨头英伟达为例，目前全球相关AI加速卡供不应求，我国亦处于智算中心建设加速期，**因此算力有望成为本轮AIGC先行指标，目前已有相关公司兑现业绩，其业绩望持续兑现，积极的推荐以下三条投资主线：**
 - ✓ 1)具备算力服务器的厂商，受益标的为**浪潮信息、神州数码、中科曙光、紫光股份、拓维信息、工业富联、四川长虹、中国长城等；**
 - ✓ 2)具备算力芯片的厂商，受益标的为**寒武纪、海光信息、龙芯中科、景嘉微等；**
 - ✓ 3)相关AI云厂商，受益标的为**首都在线、鸿博股份、云赛智联、青云科技、优刻得、光环新网、新炬网络、真视通、恒润股份等。**

注：*来自wind一致预测

资料来源：， 华西证券研究所

3.1 投资建议: 梳理AIGC的受益厂商

AIGC的A股受益标的

公司名称	股票代码	收盘价	市值(亿元)	EPS (元)			PE (倍)		
		2023/8/14	2023/8/14	2022	2023E	2024E	2022	2023E	2024E
寒武纪*	688256. SH	159. 40	664. 05	-3. 14	-1. 95	-1. 34	-	-	-
海光信息	688041. SH	53. 80	1250. 49	0. 38	0. 64	0. 87	141. 6	84. 1	61. 8
景嘉微*	300474. SZ	74. 61	341. 26	0. 64	0. 89	1. 26	116. 6	83. 6	59. 2
龙芯中科*	688047. SH	94. 40	378. 54	0. 14	0. 48	0. 92	674. 3	195. 8	102. 8
中科曙光	603019. SH	41. 80	611. 78	1. 06	1. 38	1. 81	39. 4	30. 3	23. 1
神州数码*	000034. SZ	29. 79	199. 36	1. 57	1. 84	2. 18	19. 0	16. 2	13. 7
拓维信息*	002261. SZ	16. 58	207. 89	-0. 81	0. 10	0. 13	-	171. 8	123. 7
工业富联*	601138. SH	21. 40	4251. 42	1. 02	1. 20	1. 41	21. 0	17. 8	15. 2
四川长虹*	600839. SH	4. 11	189. 73	0. 10	0. 15	0. 19	40. 5	28. 2	21. 7
中国长城*	000066. SZ	12. 25	395. 16	0. 04	0. 17	0. 24	322. 4	72. 1	51. 3
鸿博股份	002229. SZ	37. 35	186. 13	-0. 15	-0. 03	0. 12	-	-	300. 2
首都在线	300846. SZ	14. 49	67. 64	-0. 41	0. 16	0. 26	-	91. 4	55. 0
鸿博股份*	002229. SZ	37. 35	186. 13	-0. 15	-0. 03	0. 12	-	-	300. 2
光环新网*	300383. SZ	10. 49	188. 57	-0. 49	0. 35	0. 43	-	29. 8	24. 4
新炬网络*	605398. SH	26. 76	31. 21	0. 68	0. 61	0. 74	39. 4	44. 0	36. 3
优刻得*	688158. SH	17. 35	78. 61	-0. 92	-0. 63	-0. 55	-	-	-
云赛智联*	600602. SH	13. 94	162. 59	0. 13	0. 16	0. 20	105. 6	85. 7	69. 3

注: *来自wind一致预测

3.2.1 浪潮信息：中国服务器/AI服务器市占率稳居榜首

- ◆ **浪潮信息是全球领先的新型IT基础架构产品、方案及服务提供商：**公司是全球领先的 AI 基础设施供应商，拥有业内最全的人工智能计算全堆栈解决方案，涉及训练、推理、边缘等全栈 AI 场景，构建起领先的 AI 算法模型、AI 框架优化、AI 开发管理和应用优化等全栈 AI 能力，为智慧时代提供坚实的基础设施支撑。
- ◆ **公司算力技术壁垒浓厚：**生产算力方面，公司拥有业内最强最全的 AI 计算产品阵列，业界性能最好的Transformer 训练服务器 NF5488、全球首个 AI 开放加速计算系统 MX1、自研 AI 大模型计算框架 LMS。聚合算力层面，公司针对高并发训练推理集群进行架构优化，构建了高性能的NVMe 存储池，深度优化了软件栈，性能提升 3.5 倍以上。调度算力层面，浪潮信息 Aistation 计算资源平台可支持 AI 训练和推理，是业界功能最全的 AI 管理平台；同时，浪潮信息还有自动机器学习平台 AutoML Suite，可实现自动建模，加速产业化应用。

浪潮信息智算中心



浪潮信息智算中心



3.2.2 神州数码: 华为生态核心践行者

- ◆ **神州数码领先的数字化转型：**神州数码围绕企业数字化转型的关键要素，开创性的提出“数云融合”战略和技术体系框架，着力在云原生、数字原生、数云融合关键技术和信创产业上架构产品和服务能力，为处在不同数字化转型阶段的快消零售、汽车、金融、医疗、政企、教育、运营商等行业客户提供泛在的敏捷IT能力和融合的数据驱动能力。
- ◆ **神州数码为华为生态核心践行者：**公司旗下的神州鲲泰基于华为鲲鹏处理器多款不同种类的服务器产品，包括1、单路服务器：R222、R224；2、双路服务器：R522、R524、R722、R724、R2240、R2260、R2280。3、四路服务器：R822。此外，公司基于华为鲲鹏920处理器与昇腾Atlas AI加速卡，神州数码开发了采用ARM架构的一系列AI服务器。

神州数码服务器及相关参数

名称	示意图	形态	处理器	内存支持	AI加速卡/AI处理器	AI算力
KunTai A222		2U单路边缘机架式服务器	1*鲲鹏920处理器，24核，主频2.6GHz	4个DDR4 RDIMM，最高速率3200MT/s内存保护支持ECC、SEC/DED、SDDC、Patrol scrubbing功能；单根内存条容量支持16GB/32GB/64GB/128GB	最大支持3张Atlas 300V 视频解析卡或Atlas 300I Pro 推理卡或Atlas 300V Pro 视频解析卡	最大420 TOPS INT8
KunTai A722		2U 双路推理型 AI 机架式服务器	2*鲲鹏920处理器，支持32、48、64核可选，主频2.6GHz	16个或32个DDR4 RDIMM，最高速率2933MT/s内存保护支持ECC、SEC/DED、SDDC、Patrol scrubbing功能；单根内存条容量支持16GB/32GB/64GB/128GB	最大支持8张，Atlas 300V 视频解析卡或Atlas 300I Pro 推理卡或Atlas 300V Pro 视频解析卡	最大1120 TOPS INT8
KunTai A924		4U四路训练型AI机架式服务器	4*鲲鹏920处理器，支持48核，主频 2.6GHz	支持32个DDR4内存插槽，速率最高2933MT/s内存保护支持ECC、SEC/DED、SDDC、Patrol scrubbing功能；单根内存条容量支持32GB/64GB/128GB	8*昇腾910，支持直出100G RoCE网络接口	最大512Tops Int8或256Tops FP16

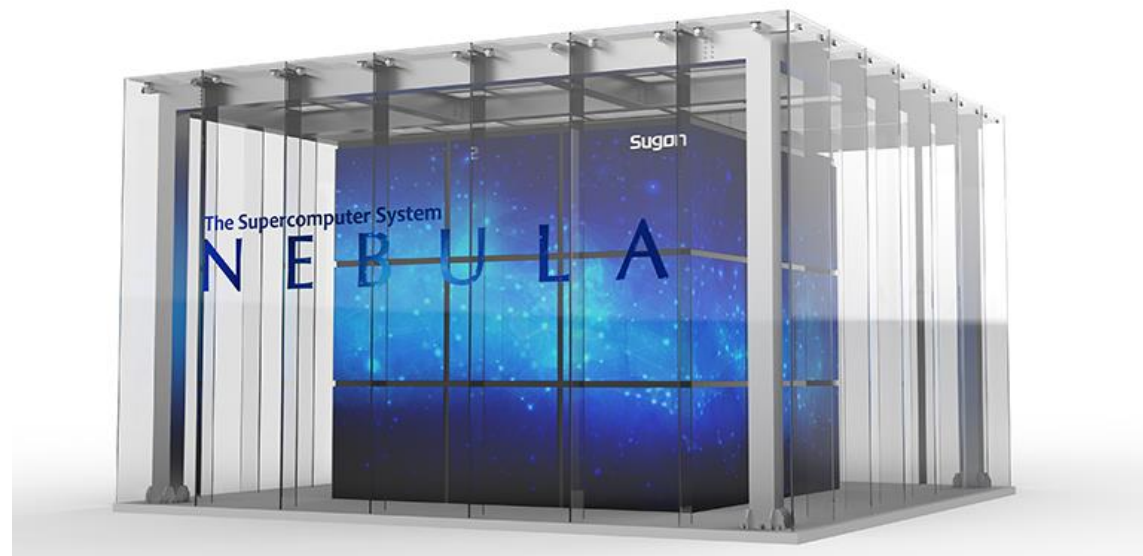
3.2.3 中科曙光：我国高性能计算、智能计算领军企业

- ◆ **中科曙光作我国核心信息基础设施领军企业：**在高端计算、存储、安全、数据中心等领域拥有深厚的技术积淀和领先的市场份额，并充分发挥高端计算优势，布局智能计算、云计算、大数据等领域的技术研发，打造计算产业生态，为科研探索创新、行业信息化建设、产业转型升级、数字经济发展提供了坚实可信的支撑。
- ◆ **依托先进计算领域的先发优势和技术细节，中科曙光全面布局智能计算：**完成了包括AI核心组件、人工智能服务器、人工智能管理平台、软件等多项创新，构建了完整的AI计算服务体系。并积极响应时代需求，在智能计算中心建设浪潮下，形成了5A级智能计算中心整体方案。目前，曙光5A智能计算中心已在广东、安徽、浙江等地建成，江苏、湖北、湖南等地已进入建设阶段，其他地区也在紧张筹备和规划中。

中科曙光主要产品

通用服务器 机架式服务器 高密度服务器 刀片服务器 核心应用服务器	智能计算服务器 深度学习训练 智能应用推理	终端&工作站 微型计算机 工作站	高性能计算机 通用高性能计算机 高性能计算机系统组件 高性能计算机的服务支撑
机房冷却设施 微模块产品 液冷基础设施产品	存储产品 分布式统一存储 多控统一存储 高密度存储服务器 备份一体机	网络安全产品 数据中心安全产品 汇聚分流设备 智能加速卡 网络内容识别分析系统 网络态势感知系统	大数据平台软件 大数据智能引擎系列 数据工程服务系列 视频智能分析系列 大数据与人工智能实训平台
云计算平台软件 云计算操作系统 超融合一体机 云桌面 云容灾	计算服务 弹性计算服务 混合计算服务 专有计算服务 API 托管、运营	云计算服务 云服务器 ECS 裸金属 BMS 对象存储 OSS 云容器实例 CCI 人工智能服务	城市云 智慧城市 国资云 交通云 医疗云
5A级智算中心		数据开发 DDS 数据治理中心 DGS 数据服务 DSS 数据可视化 DAV 数据集成 Data Integration	

中科曙光硅立方液体相变冷却计算机



3.2.4 紫光股份：全栈ICT基础设施及服务商

- ◆ **紫光股份全栈ICT基础设施及服务商。**公司提供技术领先的网络、计算、存储、云计算、安全和智能终端等全栈ICT基础设施及服务，包括网络设备、存储产品、云计算与云服务、网络安全产品及服务智能终端六大块；提供行业智慧应用一站式数字化解决方案，以及顶层设计、软件开发、系统集成及运维管理等全生命周期服务。2023 NAVIGATE领航者峰会上，紫光股份旗下新华三集团重磅推出全球首发51.2T 800G CPO硅光数据中心交换机（H3C S9827系列），助力AIGC时代极致算力释放。
- ◆ **2022年多项产品市场占有率持续领先。**2022年公司在中国以太网交换机、企业网交换机、数据中心交换机市场，分别以33.8%、35.3%、31.0%的市场份额排名第二；中国企业网路由器市场份额31.8%，持续位列市场第二；中国企业级WLAN市场份额28.0%，连续14年蝉联市场第一；中国X86服务器市场份额18.0%，保持市场第二；中国刀片服务器市场份额54.8%，蝉联市场第一；中国存储市场份额12.5%，连续位居市场第二。2022年前三季度，在中国UTM防火墙市场以23.7%的市场份额位居市场第二；中国超融合市场份额23.9%，位列市场第一。在2023年《财富》中国上市公司500强紫光股份位列总榜单第184位，排名稳步提升，持续领跑ICT行业。

H3C S9827系列



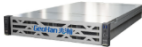
智能数字平台



3.2.5 拓维信息: 华为生态重要参与者

- ◆ **拓维信息是领先的软硬一体化解决方案提供商:** 公司1996年成立，业务涵盖政企数字化、智能计算、鸿蒙生态，覆盖全国31个省级行政区、海外10+国家，聚焦数字政府、运营商、考试、交通、制造、教育等重点领域和行业，服务超过1500家政企客户，为其提供全栈国产数字化解决方案和一站式全生命周期的综合服务。
- ◆ **拓维信息为华为生态重要参与者:** “兆瀚”系列通用服务器是基于ARM架构，搭载鲲鹏920处理器设计开发的机架式型服务器，拥有高的性能、可靠性、高效环保、兼容性强等特点；“兆瀚”系列AI服务器能够满足当前各类主流AI场景与AI大模型的训练需求，已经在国内多个区域人工智能计算中心、城市人工智能中枢、通用AI服务器场景中得到了应用，已经在国内多家头部互联网企业开展适配测试。

拓维信息旗下“兆瀚”系列服务器产品介绍

种类	名称	示意图	形态	处理器	内存支持	AI加速卡/AI处理器	AI算力
通用服务器	兆瀚RH220系列		2U双路机架	支持两颗华为鲲鹏920处理器，CPU主频2.6GHz。单CPU最多64个内核，最大功率180w。	最多支持32个DDR4内存DIMM插槽，最高速率2933MT/s	/	/
	兆瀚RH520系列		4U机架服务器	支持两颗华为鲲鹏920处理器，CPU主频2.6GHz。单CPU最多64个内核，最大功率180w。	最多支持32个DDR4内存DIMM插槽，最高速率2933MT/s	/	/
AI服务器	兆瀚RA2300-A		2U推理服务器	支持两颗华为鲲鹏920处理器，CPU主频2.6GHz。单CPU最多64个内核，最大功率180w。	最多支持32个DDR4内存DIMM插槽，最高速率2933MT/s	支持Atlas 300I Pro推理卡和Atlas 300V Pro视频解析卡	最大1.12 POPS INT8；最大560 TFLOPS PF16
	兆瀚SA300		2U智能边缘服务器	支持一颗华为鲲鹏920处理器，CPU主频2.6GHz。单CPU最多64个内核，最大功率181w。	最多支持4个DDR4内存DIMM插槽，最高速率2934MT/s	支持Atlas 300I Pro推理卡/Atlas 300V Pro视频解析卡	最大420 TOPS INT8 或 384路1080P 30 FPS视频解析（硬件解码能力）
	兆瀚RA5900-A		4U训练服务器	支持四颗华为鲲鹏920处理器，CPU主频2.6GHz。单CPU最多64个内核，最大功率182w。	最多32个DDR4内存插槽，支持RDIMM。单根内存条容量支持32 GB/64GB	8*昇腾910	/
	兆瀚RA2302-B		2U AI 服务器	2*64核青松处理器	32个DDR4内存插槽，最高3200 MT/s，支持ECC	最大支持4个Atlas 300I/V Pro	最大560 TPOS INT8

资料来源：公司官网，华西证券研究所

3.2.6 四川长虹：全球化综合型企业

- ◆ **四川长虹电器股份有限公司是一家全球化综合型企业**，主营“以电视、冰箱(柜)、家用空调、洗衣机等为代表的消费电器业务，以冰箱压缩机为代表的部品业务，以ICT产品分销和专业ICT解决方案提供为代表的ICT综合服务业务，以电子制造(EMS)为代表的精益制造服务业务以及其他相关特种业务”等。23年6月，四川长虹正式签约成为成都大运会唯一指定家电品牌。
- ◆ **子公司华鲲振宇是国内信创产业的领军企业**。华鲲振宇于7月26日在北京举办产品发布会，分享基于“鲲鹏+昇腾”的算力创新，发布全新一代算力基础设施与解决方案——“天擎”系列高规格通用计算平台，“天智”系列高密AI算力平台，“天恒”超融合移动平台解决方案和“天极”系列液冷解决方案。公司融合了华为鲲鹏和昇腾根技术软硬件能力，以及长虹集团航天领域的制造能力。基于鲲鹏和昇腾已推出30多种算力产品。在北京、深圳和成都都设立了研发中心，建立了覆盖全国32个省和主要城市的销售服务网络。是华为第一个“鲲鹏+昇腾”双领先能力的认证伙伴。

成都大运会唯一指定家电品牌——四川长虹



华鲲振宇产品发布会



3.2.7 中国长城：国产自主安全计算产业的引领者

- ◆ **自主安全计算产业的引领者**:中国长城是中国电子旗下“安全、先进、绿色自主计算产业专业子集团”，是中国“PKS”自主计算体系建设主力军和网信科技自主创新生力军。作为我国自主安全计算产业的引领者，中国长城曾研发出我国第一台具有自主知识产权的中文微型电脑、第一块电脑硬盘、第一款终端ASIC芯片、第一台显示器、第一台光纤转换器、第一台光笔图形显示终端等，是我国电脑、电源、高新电子、金融信息化、医疗信息化等领域系列国家和行业标准的起草单位之一。
- ◆ **产品覆盖面广，PC端及服务器产品线完善**。依托“PKS”自主计算体系，中国长城构建了从芯片、台式机、笔记本、服务器、网络交换设备到应用系统等具有完整自主知识产权的产品谱系。服务器端种类包含通用性、AI型、存储型和塔式，PC端包含笔记本、一体机以及台式机。

中国长城产品体系



3.3.1 海光信息：支持全精度，GPU实现规模量产

- ◆ **海光信息主要从事高端处理器、加速器等计算芯片产品和系统的研究、开发，主要产品包括海光CPU和海光DCU:**2018年10月，公司启动深算一号DCU产品设计，海光8100采用先进的FinFET工艺，典型应用场景下性能指标可以达到国际同类型高端产品的同期水平。2020年1月，公司启动DCU深算二号的产品研发。
- ◆ **海光DCU性能强大:**海光DCU基于大规模并行计算微结构进行设计，不但具备强大的双精度浮点计算能力，同时在单精度、半精度、整型计算方面表现同样优异，是一款计算性能强大、能效比较高的通用协处理器。海光DCU集成片上高带宽内存芯片，可以在大规模数据计算过程中提供优异的数据处理能力。

海光信息主要产品



系列	7000系列CPU	5000系列CPU	3000系列CPU
核心规格	最大32个物理核心	最大16个物理核心	最大8个物理核心
应用领域	高端通用服务器、先进计算系统	通用服务器	个人工作站、工控设备等终端产品

系列	8000系列DCU
核心规格	60-64个深度计算单元
应用领域	先进计算系统、人工智能

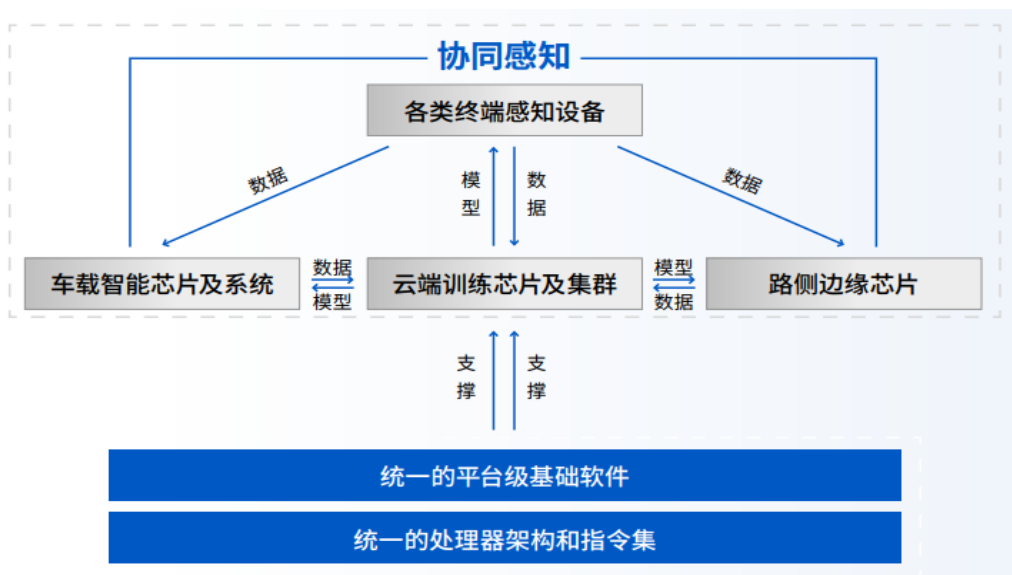
海光深算一号性能达到国际同类产品水平

项目	海光	NVIDIA	AMD
品牌	深算一号	Ampere 100	MI100
生产工艺	7nm FinFET	7nm FinFET	7nm FinFET
核心数量	4096 (64CUs)	2560 CUDA processors 640 Tensor processors	120CUs
内核频率	Up to 1.5GHz (FP64) Up to 1.7Ghz (FP32)	Up to 1.53Ghz	Up to 1.5GHz (FP64) Up to 1.7Ghz (FP32)
显存容量	32GB HBM2	80GB HBM2e	32GB HBM2
显存位宽	4096 bit	5120 bit	4096bit
显存频率	2.0 GHz	3.2 GHz	2.4 GHz
显存带宽	1024 GB/s	2039 GB/s	1228 GB/s
TDP	350 W	400 W	300 W
CPU to GPU 互联	PCIe Gen4 x 16	PCIe Gen4 x 16	PCIe GEN4 x 16
GPU to GPU 互联	xGMI x 2, Up to 184 GB/s	NVLink up to 600 GB/s	Infinity Fabric x 3, up to 276 GB/s

3.3.2 寒武纪：少数全面掌握AI芯片技术的企业之一

- ◆ **寒武纪是目前国际上少数几家全面系统掌握了通用型智能芯片及其基础系统软件研发和产品化核心技术的企业之一：**寒武纪主营业务是应用于各类云服务器、边缘计算设备、终端设备中人工智能核心芯片的研发和销售。公司的主要产品包括终端智能处理器IP、云端智能芯片及加速卡、边缘智能芯片及加速卡以及与上述产品配套的基础系统软件平台。
- ◆ **公司AI技术积累浓厚：**能提供云边端一体、软硬件协同、训练推理融合、具备统一生态的系列化智能芯片产品和平台化基础系统软件。2022年3月，寒武纪正式发布了新款训练加速卡“MLU370-X8”，搭载双芯片四芯粒封装的思元370，集成寒武纪MLU-Link多芯互联技术，主要面向AI训练任务。

寒武纪“云边端车”协同



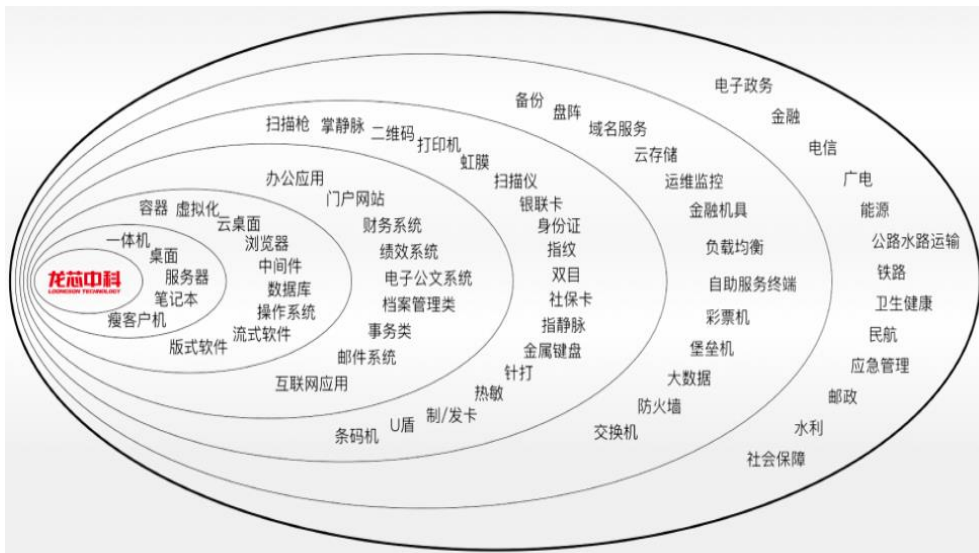
寒武纪产品技术图谱



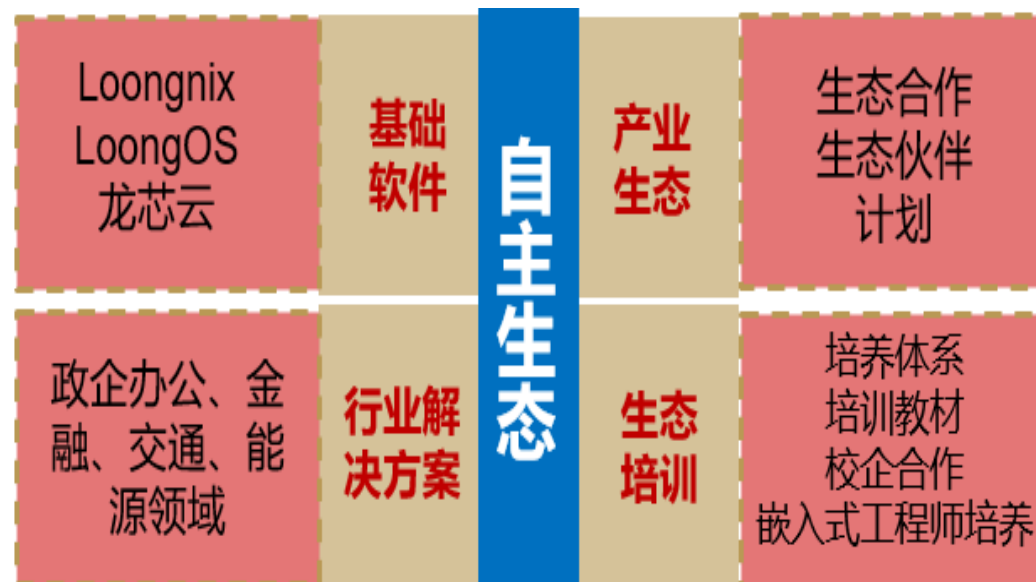
3.3.3 龙芯中科：2K2000系列集成自主GPU

- ◆ **龙芯中科主营业务为处理器及配套芯片的研制、销售及服务：**主要产品与服务包括处理器及配套芯片产品与基础软硬件解决方案业务。公司基于信息系统和工控系统两条主线开展产业生态建设，面向网络安全、办公与业务信息化、工控及物联网等领域与合作伙伴保持全面合作，产品在电子政务、能源、交通、金融、电信、教育等行业领域已获得广泛应用。
- ◆ **公司自主研发2K200系列GPU：**2022年12月，龙芯2K2000完成了初步功能调试及性能测试，达到其设计目标，2023年将推出试用。龙芯2K2000集成了两个LA364处理器核，典型工作频率为1.5GHz，共享2MB的L2缓存，SPEC2006INT（base）单核定/浮点分值达到13.5/14.9分。龙芯2K2000芯片集成了龙芯自主研发的GPU，并优化了图形算法和性能。

龙芯中科生态合作示意图



龙芯中科自主生态



3.3.4 景嘉微：新一代JM9系列有望打开商用市场

- ◆ **国产GPU龙头企业:** 公司成立于2006年，主要从事军用电子产品的研发、生产、销售，目前形成了三大业务板块分别是图形线控模块、小型专用雷达和芯片业务。GPU方面，2014年首推JM5400实现了军用GPU的国产替代；第二款芯片JM7200于2018年研发成功，具备了PC端的功能；日前，公司9系列芯片研发成功，具备高性能计算能力。
- ◆ **新一代JM9系列有望打开商用市场:** 日前，公司JM9系列图形处理芯片已顺利发布，应用领域涵盖地理信息系统、媒体处理、CAD辅助设计、游戏、虚拟化等高性能显示和人工智能计算领域。目前，信创市场为公司提供了新的业务增长点，JM9系列图形处理芯片的成功发布将为公司未来进一步拓展通用市场提供强有力的产品支撑。

景嘉微GPU系列产品



景嘉微7系列GPU示意图



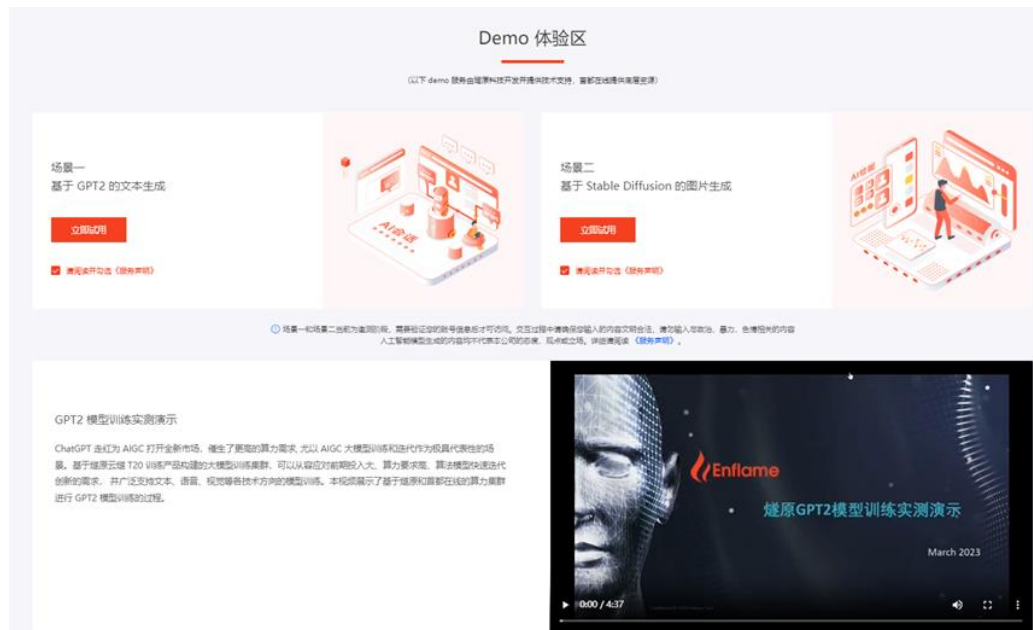
3.4.1 首都在线: AI算力云龙头, AIGC “挖井人”

- ◆ **公司绑定英伟达、燧原, AI云开启第二波成长曲线。**公司已摆脱单一的IaaS 公有云, 重点转向AI算力云转型, 有望借助底层英伟达GPU算力储备, 以AI云为抓手, 开启第二波成长曲线, 我们认为算力网络以及边缘节点是公司AI云的核心壁垒之一。公司首云星图云算力平台已经震撼发布, 深度绑定英伟达, 算力平台采用A100、A40、A5000, 为全球数字世界多场景提供澎湃算力。同时, 公司携手燧原科技, 开启AIGC芯征程, 重点针对大模型 MaaS开展联合攻关, 正式推出云燧i20支撑的AIGC实时推理应用。
- ◆ **海外游戏具备竞争优势, AI算力云赋能千行百业。**我们认为公司AI云平台产品发布与公司底层算力储备密不可分, 借助通过算力、网络、存储等核心能力构建“云-网-数”一体的边缘计算平台, 就近为高算力业务场景如**云游戏、AI、XR、数字人、数字孪生、元宇宙、智能制造**等各领域提供了算力支持。其中云游戏方面, 我们判断云游戏市场处于“技术成熟走向商业可行”与“商业可行走向商业腾飞”的交替阶段, 公司坐拥算力和算力网络双重竞争优势。此外, 公司传统IDC和云业务积极布局海外, 也将会是公司另一个业绩爆发点。

公司星图云底层算例示意图

公司与遂原科技合作示意图

高性能算力供给



3.4.2 云赛智联：城市数字化转型建设主力军

- ◆ **云数赋能，智联未来。**云赛智联成立于1986年12月，作为上海仪电（集团）有限公司旗下的上市公司，是一家以云计算与大数据、行业解决方案及智能化产品为核心业务的专业化信息技术服务企业。公司加快电子制造与信息技术相融合，不仅拥有坚实的物联网硬件产品研发制造基础，而且形成了以自主建设的高等级大规模数据中心为平台，集云计算业务架构和云计算服务能力为一体。全面向智能化、网络化、信息化转型升级。
- ◆ **建设超算中心，助力城市数字化转型：**今年3月，由云赛智联所属上海南洋万邦软件技术有限公司，承建的上海超级计算中心市公共算力服务平台正式落地，人工智能计算机集群正式上电开机。根据南洋万邦公众号，该项目的计算峰值能力将达到100PFLOPS，以满足上海科技创新和产业发展对人工智能公共算力和服务需求，成为上海数字化转型智能底座。6月29日，云赛智联召开监事会十二届二次会议。会议同意云赛智联与上海仪电集团、上海数据集团、上海市信息投资股份有限公司和其他非关联股东共同出资成立算力设施公司——上海智能算力科技有限公司。公司在超算方面的持续布局将助力公司云计算大数据板块高质量发展。

云赛智联云计算和大数据业务介绍

云计算与大数据

CLOUD COMPUTING AND BIG DATA

云计算与大数据

IaaS

专业级安全系统服务，访问控制、入侵监测与防护、紧急响应服务等；应用外包服务，负载均衡、主机管理、网络监测、高速缓存、内容分发（CDN）等。

PaaS

通过数据中心的云存储、云灾备、云主机和云安全等服务以及政务网、政务网和城域网等硬件系统建设和系统运维，同时利用软件系统能力整合智慧城市的各种信息系统，有效解决信息孤岛问题，实现大数据资源挖掘和共享，在智慧城市中的交通、医疗、食品、照明等等各个领域开拓创新。

SaaS

依据客户不同的应用场景，以及在该场景下的问题点分析，提供适合客户需求的解决方案。让用户的信息系统更有价值。

上海超算人工智能算法平台功能介绍





04 风险提示

- ◆ **核心技术水平升级不及预期的风险:** AIGC相关产业技术壁垒较高，公司核心技术难以突破，进程低于预期，影响整体进度。
- ◆ **AI伦理风险:** AI可能会生产违反道德、常规、法律等内容。
- ◆ **政策推进不及预期的风险:** 受到宏观经济、财政、疫情影响，政策推进节奏不及预期。
- ◆ **中美贸易摩擦升级的风险:** 供应链存在部分海外提供商，容易受到美国制裁，导致产品研发不及预期。

分析师与研究助理简介

刘泽晶（首席分析师）2014-2015年新财富计算机行业团队第三、第五名，水晶球第三名，10年证券从业经验。

分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，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，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资评级	说明
以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过15%
	增持	分析师预测在此期间股价相对强于上证指数在5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过15%
行业评级标准		
以报告发布日后的6个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层

网址：<http://www.hx168.com.cn/hxzq/hxindex.html>

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本公司签约客户使用。本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司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或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。

所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，如需引用、刊发或转载本报告，需注明出处为华西证券研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。

THANKS