

# 华为算力编年史

## AIGC行业深度报告(9)

华西计算机团队

2023年6月14日

分析师：刘泽晶

SAC NO: S1120520020002

邮箱：liuzj1@hx168.com.cn

## 核心逻辑:

- ◆ **华为算力编年史，国产芯片之光。**我们复盘了华为国产芯片发家史，从手机端到AI算力，远期来看，**2009年-2012年**，手机芯片不断革新；**2013年-2017年**，麒麟芯片横空出世；**2018年-至今**，昇腾+鲲鹏打造AI算力生态；近期来看，**19年**，中国计算产业千亿蓝海，新兴产业链助推行业高增，万亿市场呈待挖掘，华为鲲鹏应运而生；**20-21年**，华为持续加码国产软硬件，以一云两翼双引擎，构筑开放产业生态，“硬件开放、软件开源、使能伙伴”，助力计算产业长远发展。**22年**，行业信创元年，华为深度受益；**23年**，我们判断算力呈现爆发模式，华为正依托昇腾AI算力，开启新一轮成长周期。
- ◆ **算力大爆发，华为AI开启新一轮成长周期:**再三强调，大模型背景下算力势必迎来爆发，我国超算规模持续高增。在智算大趋势下，华为为加快千行百业数字化转型和智能化升级，计划构建面向通用计算的鲲鹏计算产业和面向人工智能的昇腾AI计算产业。根据新华网财经，华为昇腾计算业务总裁张迪焯表示，“昇腾AI”基础软硬件平台已孵化和适配了30多个主流大模型，已有25个城市基于昇腾建设人工智能计算中心，其中14个已经上线并饱和运行，根据电子信息产业网，华为计算产品线总裁邓泰华在华为全联接大会2022上表示，截至2022年11月，昇腾AI已有20多个算力节点接入中国算力网。华为AI产品种类丰富，包括AI芯片、AI服务器、AI云、全栈全场景AI解决方案等，我们判断华为依托昇腾AI算力开启新一轮成长周期，相关产业链有望深度受益。
- ◆ **AI硬件自主可控势在必行:**如果说产品是AI赋能、企业开启第二轮业绩增长曲线的“流量入口”，那么算力即是大厂开启算力争夺战的“入场券”。近年来国际环境发生较大变化，因此自主可控势在必行，我国**政策端**持续发力，加速推动国产自主可控进程，我国短期发布多条政策助力AI发展，工作方向主要瞄准推动国产AI芯片突破等。此外，我国**产业端**积极响应，智能算力建设正处于持续提速阶段。**我们再次重申观点，短期算卡为王，长期自主可控！**
- ◆ **投资建议:** AI算力，自立自强，AI+信创势在必行，**我们判断华为正以昇腾AI开启新一轮成长周期**，关注四条投资主线：1) **华为AI服务器厂商**，相关受益标的为：**拓维信息、神州数码、广电运通、四川长虹、东华软件、同方股份**等；2) **国产AI芯片厂商**，相关受益标的为：**寒武纪、海光信息、景嘉微、龙芯中科**等；3) **AI服务器厂商**，相关受益标的为：**中科曙光、工业富联、浪潮信息**等；4) **AI云厂商**，相关受益标的为：**首都在线、鸿博股份、青云科技、优刻得、光环新网、新炬网络**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 华为算力编年史

02 AI硬件自主可控势在必行

03 投资建议: 梳理算力相关受益厂商

04 风险提示



## 01 华为算力编年史

## 1.1 国产芯片之光——华为海思

- ◆ **海思是全球领先的Fabless半导体与器件设计公司。**公司前身为华为集成电路设计中心，1991年启动集成电路设计及研发业务，为汇聚行业人才、发挥产业集成优势，2004年华为将海思注册全资子公司，提供海思芯片对外销售及服务。海思产品覆盖智慧视觉、智慧IoT、智慧媒体、智慧出行、显示交互、手机终端、数据中心及光收发器等多个领域。
- ◆ **产品服务辐射全球，持续研发专利丰厚。**海思在中国、新加坡、韩国、日本、欧洲等地设有12个办事处和研发中心，产品和服务遍布全球100多个国家和地区。海思有着20余年的技术积累，目前拥有200+自主知识产权的芯片，8,000+专利。

使能万物互联的智能终端



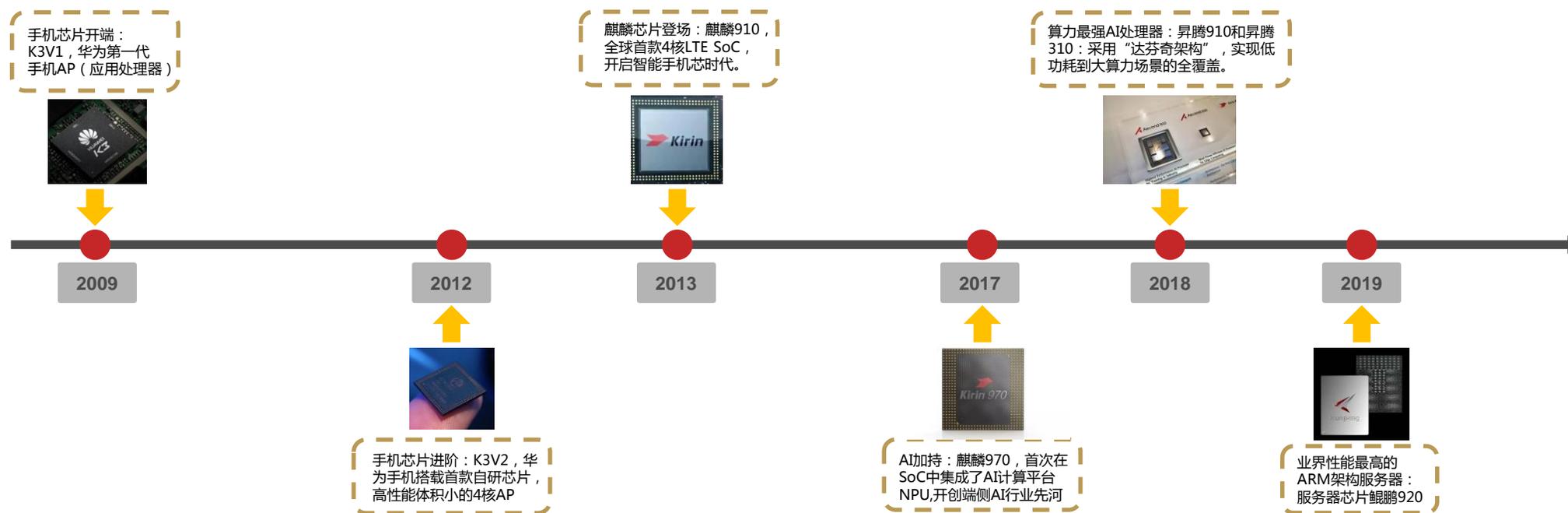
海思产品服务九大领域



## 1.2 国产芯片发家史，从手机端到AI算力

- ◆ **手机芯片不断革新（2009年-2012年）**：2009年，华为海思推出了第一款面向公开市场的手机终端处理器K3V1，此后又于2012年推出K3V2处理器并搭载在定位旗舰的Mate 1、P6等机型上。虽然K3V2功耗高，兼容性差，但为后续型号革新奠定基础；
- ◆ **麒麟芯片横空出世（2013年-2017年）**：2013年底，华为海思推出了第一款SoC芯片麒麟910。随着麒麟芯片不断演进，2017年麒麟系列芯片已经发展到麒麟970，并用在P20等华为旗舰机型上，这是首次在SoC中集成了AI计算平台NPU，开创端侧AI行业先河；
- ◆ **昇腾+鲲鹏打造AI算力生态（2018年-至今）**：2018年华为海思发布昇腾310，并在2018年发布昇腾910：采用“达芬奇架构”，实现低功耗到大算力场景的全覆盖。2019年亦发布业界性能最高的ARM架构服务器芯片--鲲鹏920。

华为海思芯片重点发展历程



资料来源：海思官网，华为公众号，恒昌公司官微，威尔克通信实验室，全球半导体观察，华西证券研究所整理

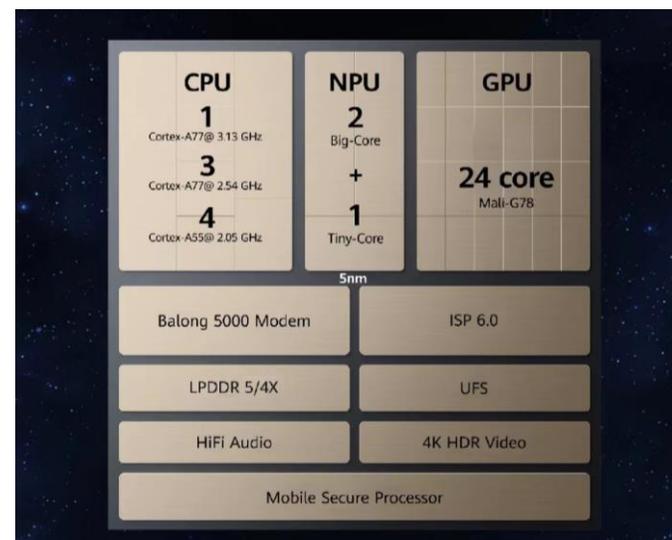
# 1.3.1 麒麟芯片：从K3V1到麒麟9000，中国手机芯片崛起之路

- ◆ 麒麟系列是华为在手机上搭载的CPU处理器芯片。2020年，麒麟系列更新到麒麟9000。麒麟9000是业界最高集成度5nm 5G SoC，采用全球顶级5nm工艺制程，集成153亿个晶体管的手機芯片。
- ◆ 麒麟9000——更快、更高、更强
- ✓ **刷新5G速度**：麒麟9000通过支持5G SA双载波聚合，使其在Sub-6G下行理论峰值速率达4.6Gbps，上行理论峰值速率达2.5Gbps，带来更快的5G体验。
- ✓ **高效能、高性能**：麒麟9000全新升级Cortex-A77 CPU，采用1+3+4三档能效架构CPU，大核主频突破3.1GHz。同时搭载全球首个24核Mali-G78 GPU，图像处理能力极大提升，能做到长时间运行大型游戏更流畅更省电。
- ✓ **AI+AR**：AI方面，麒麟9000解锁更多AI功能，例如让AI实时处理从照片走向视频；AR方面，基于SLAM和实时语义理解，以及专属的AR加速器硬化模块，华为构建的实时环境感知平台能够随人的视觉同步感知、测量，并达到比人眼更高的精度。

麒麟芯片发展历程



麒麟9000芯片性能



## 1.3.2 鲲鹏CPU：硬件开放、软件开源、使能伙伴、发展人才

- ◆ 鲲鹏包括服务器和PC机芯片：
  - ✓ 芯片端——鲲鹏920面向数据中心，主打低功耗高性能。鲲鹏处理器是华为自主研发的基于ARM架构的企业级系列处理器产品，包含“算、存、传、管、智”五个产品系统体系。规格方面，鲲鹏920可以支持64个内核，主频可达2.6GHz，集成8通道DDR4，集成100G RoCE以太网卡。与此同时，鲲鹏920支持PCIe4.0及CCIX接口，可提供640Gbps总带宽。根据华为云官微，华为表示鲲鹏920大部分性能提升来自优化的分支预测算法和增加的OP单元数量，以及改进的内存子系统架构。
  - ✓ 服务器端——打造TaiShan服务器，实现高效能计算。华为还推出基于鲲鹏920的三款ARM TaiShan200 服务器，新服务器专为高性能、高效率场景而设计，主要应用于大数据、分布式存储、ARM原生应用等场景。
- ◆ 构建自主生态，打造国产算力。目前，华为正基于“硬件开放、软件开源、使能伙伴、发展人才”的策略推动鲲鹏计算产业的发展。

鲲鹏芯片产品体系演进



鲲鹏全栈开放，使能全产业链伙伴创新



资料来源：鲲鹏计算产业白皮书，鲲鹏官网，慧博，华西证券研究所

### 1.3.3 昇腾AI芯片：面向“端、边、云”的全场景AI基础设施

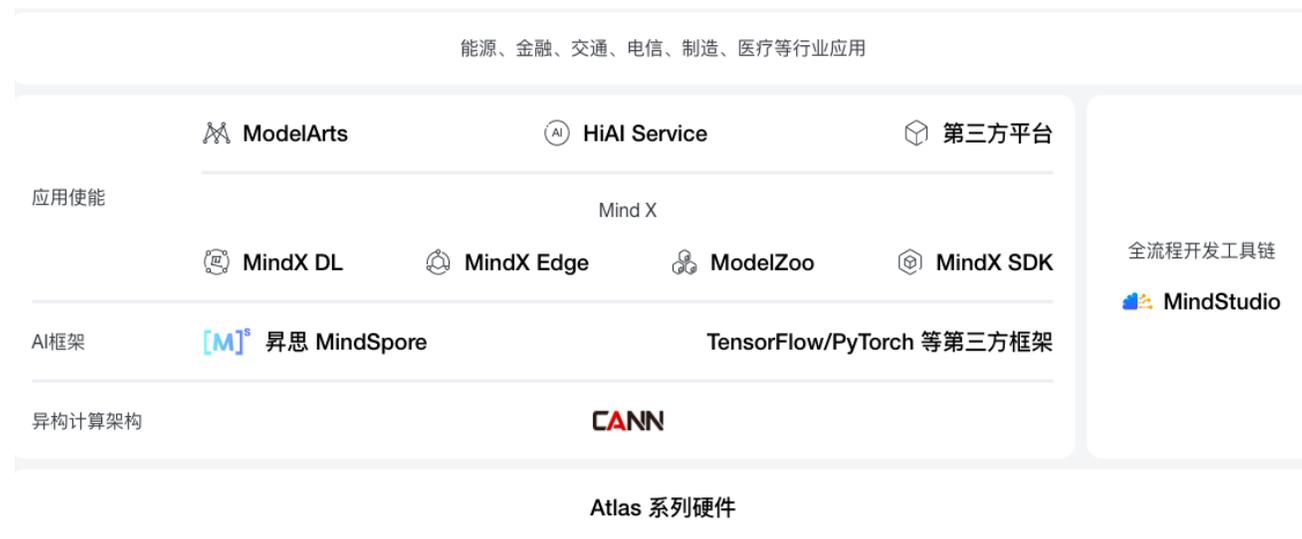
#### ◆ 全栈全场景AI芯片，构建智能计算架构核心

- ✓ 昇腾 910 和昇腾 310 两款AI 芯片均使用华为的达芬奇架构，每个 AI 核心可以在 1 个周期内完成 4096 次 MAC 计算，集成了张量、向量、标量等多种运算单元，支持多种混合精度计算，支持训练及推理两种场景的数据精度运算。
- ✓ 基于昇腾系列AI处理器和基础软件构建Atlas人工智能计算解决方案，包括Atlas系列模块、板卡、小站、服务器、集群等丰富的产品形态，打造面向“端、边、云”的全场景AI基础设施方案，覆盖深度学习领域推理和训练全流程，可以满足不同场景的大模型计算需求。
- ◆ **华为的全栈全场景AI解决方案日臻完善。** 主要基于昇腾系列处理器和基础软件构建的全栈AI计算基础设施、行业应用及服务，包括昇腾系列芯片、Atlas系列硬件、芯片使能、CANN（异构计算架构）、AI计算框架、应用使能等。

昇腾910与昇腾310对比

	昇腾310	昇腾910
特征	华为首款全栈全场景AI芯片	业界算力最强的AI处理器
架构	达芬奇架构3D Cube技术	达芬奇架构3D Cube技术
算力	16 TOPS@INT8, 8 TOPS@FP16	640 TOPS@INT8 , 320 TFLOPS@FP16
最大功耗	8W	310W
应用场景	视频分析、智能监控、自动驾驶	自然语言处理、计算机视觉

昇腾全栈 AI 软硬件平台，构筑智能世界的基石

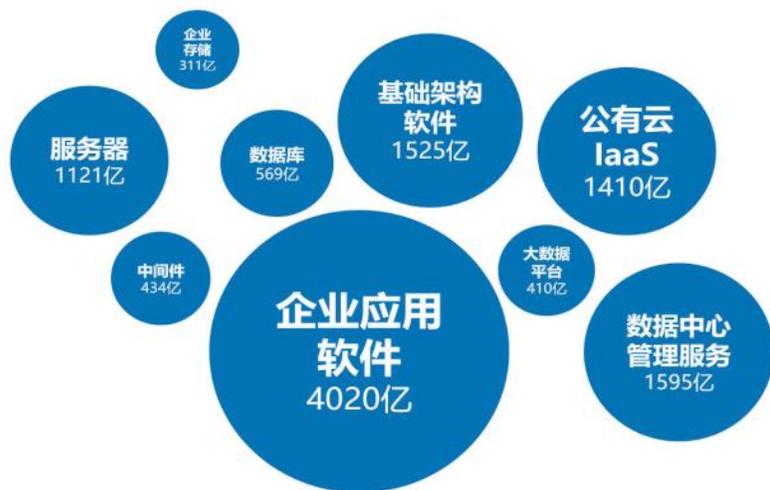


资料来源：昇腾官网，华为云公众号，华西证券研究所

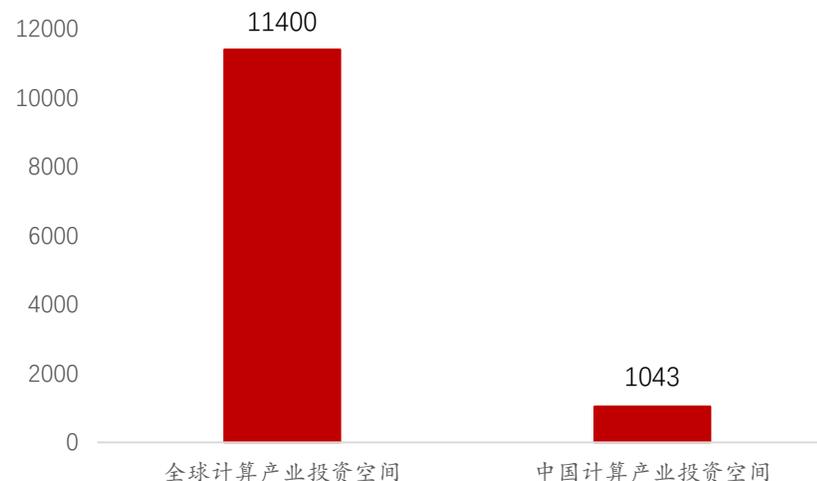
## 1.4 19年：万亿市场呈待挖掘，华为鲲鹏应运而生

- ◆ **技术持续迭代，重塑ICT产业体系。**新应用、新技术、新计算架构，百亿级联接、爆炸式数据增长将重塑ICT产业新格局，催生新的计算产业链条，涌现出新的厂家和新的生态体系：
  - ✓ (1) 硬件：服务器及部件、企业存储设备
  - ✓ (2) 软件：操作系统和虚拟化软件、数据库、中间件、大数据平台、企业应用软件
  - ✓ (3) 云服务、数据中心管理服务
- ◆ **中国计算产业千亿蓝海，新兴产业链助推行业高增。**根据《鲲鹏计算产业白皮书》，新的计算产业链将推动全球计算产业快速发展，带动全球数字经济走向繁荣。IDC预测，到2023年，全球计算产业投资空间1.14万亿美元。中国计算产业投资空间1043亿美元，接近全球的10%，是全球计算产业发展的主要推动力和增长引擎。IDC预测，在数字化浪潮的推动下，基于新一代计算平台，中国的服务器、存储、操作系统、数据库等IT企业将迎来巨大的发展机遇。

2023年全球计算产业投资额（美元）



2023年计算产业投资空间（亿美元）



# 1.4 19年：万亿市场呈待挖掘，华为鲲鹏应运而生

## ◆ 2019鲲鹏元年，华为看准时机抢先入局

- ✓ 2019年1月，华为宣布推出鲲鹏920以及基于鲲鹏920的TaiShan系列服务器。基于ARM架构，鲲鹏处理器能解决云端、边缘端、计算密集型和数据密集型等多种场景。
- ✓ 2019年5月，华为发布全球首个基于自研ARM架构的分布式缓存鲲鹏Redis
- ✓ 2019年8月，华为全面开放鲲鹏主板和整机参考架构，华为聚焦做好鲲鹏处理器，使能合作伙伴优先发展自有品牌的计算产品整机。
- ✓ 2019年9月，华为宣布将欧拉 OS 正式开源，并发布鲲鹏开发者套件DevKit。

鲲鹏920芯片



TaiShan200服务器



OpenEuler服务器开源

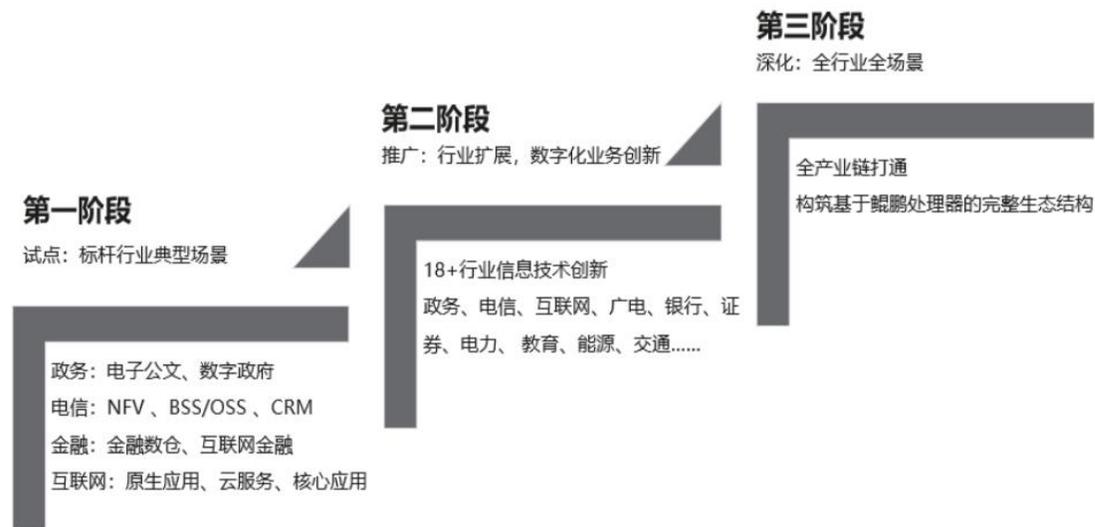


## 1.4 19年：万亿市场呈待挖掘，华为鲲鹏应运而生

- ◆ **鲲鹏全栈开放，致力打造国产算力生态。** 鲲鹏计算产业是基于Kunpeng处理器构建的全栈IT基础设施、行业应用及服务，包括PC、服务器、存储、操作系统、中间件、虚拟化、数据库、云服务、行业应用以及咨询管理服务。华为聚焦于发展Kunpeng处理器的核心能力，通过战略性、长周期的研发投入，吸纳全球计算产业的优秀人才和先进技术，构筑Kunpeng处理器的业界领先地位，为产业提供绿色节能、安全可靠、极致性能的算力底座。
- ◆ **分三个阶段构建全行业、全场景鲲鹏计算产业体系。** 完成鲲鹏计算产业从关键行业试点到全行业、全场景产业链建设目标需分为三个阶段：
  - ✓ 第一阶段：通过在政务、电信、金融和互联网等行业选取典型场景进行产业使能、孵化和试点，通过试点建立产业上下游厂家和用户信心。
  - ✓ 第二阶段：面向政务、电信、互联网、广电、金融证券、电力、能源、交通等行业全面打通产业体系，为行业数字化业务创新提供基础。
  - ✓ 第三阶段：面向全行业、全场景，打通产业链，构筑基于Kunpeng处理器的产业体系。

鲲鹏计算产业定义

鲲鹏计算产业发展阶段



# 1.5 20-21年: 持续加码国产软硬件，鲲鹏+昇腾构筑最强生态

- ◆ **一云两翼双引擎，构筑开放产业生态。** 2019年华为指出，计算产业呈现四大发展趋势：**趋势1:**从数据中心到计算中心；**趋势2:**端和边正驱动计算架构的创新；**趋势3:**计算应能源与环境友好；**趋势4:**计算架构的持续创新。为满足世界高速增长算力需求，华为推出“一云两翼、双引擎”的产业布局，构筑开放的产业生态。
  - ✓ **一云：**指华为云，通过全栈创新，提供安全可靠的混合云，成为生态伙伴的黑土地，为世界提供普惠算力。
  - ✓ **两翼：**指智能计算业务以及智能数据与存储业务。在智能计算领域，面向端、边、云，提供“鲲鹏+昇腾+x86+GPU”的多样性算力。在智能数据与存储领域，融合了存储、大数据、数据库、AI,围绕数据的全生命周期，让数据的每比特成本最优、让数据的每比特价值最大。
  - ✓ **双引擎：**指围绕“鲲鹏”与“昇腾”打造的两个基础芯片族，构筑异构的计算架构。
  - ✓ **开放生态：**指通过硬件开放和软件开源，使能广大合作伙伴，形成一个开放的产业生态。

一云两翼双引擎+开放生态



鲲鹏+昇腾，战略投入“双引擎”



## 1.5 20-21年: 持续加码国产软硬件，鲲鹏+昇腾构筑最强生态

- ◆ **昇腾方面——发布最强算力Atlas。** 基于昇腾910和310 AI处理器，华为推出了全球最快的AI训练集群Atlas 900、AI训练服务器Atlas 800、智能小站Atlas 500、AI推理与训练卡Atlas 300和AI加速模块Atlas 200,完成了Atlas全系列产品布局，覆盖云、边、端全场景，面向训练和推理提供强劲算力。基于统一的达芬奇架构和全场景AI计算框架，可实现云边端协同，加速全行业的智能化再造。
- ◆ **鲲鹏方面——持续完善鲲鹏生态。** 2020年，华为宣布开源数据库openGauss，openGauss为全球首款基于鲲鹏的企业级数据库，可覆盖企业70%以上的数据库业务场景。同时也是全球首款AI-Native数据库，具备AI-Native自调优能力，调优效果比DBA（数据库管理员）提升85%。2021年4月，鲲鹏社区正式上线，覆盖鲲鹏全栈产业开发者。

Atlas全系列产品



openGauss开源



鲲鹏社区正式上线



# 1.5 20-21年：持续加码国产软硬件，鲲鹏+昇腾构筑最强生态

- ◆ “硬件开放、软件开源、使能伙伴”，助力计算产业长远发展。
- ✓ **鲲鹏方面**：华为聚焦计算架构创新，处理器的研发，以及华为云的服务。硬件上，华为对外提供主板、SSD、网卡、RAID卡、Atlas模组和板卡，优先支持合作伙伴发展服务器和PC等计算产品。软件上，华为将开源操作系统、数据库和AI计算框架，使能伙伴发展自己品牌的产品并为开发者提供全场景开发框架。同时，华为已构建在线鲲鹏社区，提供加速库、编译器、工具链、开源操作系统等，帮助合作伙伴和开发者快速掌握操作系统、编译器以及应用的迁移调优等能力。
- ✓ **昇腾方面**：华为基于昇腾系列处理器，通过模组、板卡、小站、服务器、集群等丰富的产品形态，打造面向“端、边、云”的全场景AtlasAI基础设施方案。同时，华为提供Atlas模组、板卡等部件，使能伙伴发展基于Atlas部件的智能端侧，智能边缘、服务器、集群等AI设备。华为提供开源AI计算框架MindSpore，同时也支持业界主流AI框架如TensorFlow PyTorch Caffe等，同时华为提供模型转换工具支持主流模型便捷的转换到MindSpore。

鲲鹏全栈开放



昇腾全栈+多样性算力创新



## 1.5.1 硬件开放：服务器主板及整机

- ◆ **鲲鹏服务器主板及整机：高性能+高可靠+高效能。** 鲲鹏服务器主板是基于鲲鹏处理器的数据中心服务器主板，具有多核、超大内存带宽、支持PCIe 4.0和100GE网络等特点。主板内置BMC芯片和BIOS软件，兼容业界主流内存、硬盘、网卡、操作系统等软硬件。华为开放鲲鹏服务器主板接口规范和设备管理规范，提供机箱、散热、供电、背板等参考设计指南，帮助合作伙伴快速开发出自有品牌的服务器产品。
- ◆ **硬件开放，全力加码鲲鹏产业生态。** 2019年华为正式宣布“硬件开放、软件开源、使能合作伙伴”发展战略。在硬件方面，开放主板和部件，使能合作伙伴打造自有品牌整机产品。根据搜狐新闻，到2020年6月，伙伴生产的鲲鹏服务器销量已经超过了华为自产的TaiShan服务器，真正做到市场伙伴优先；根据华为官网，截至2020年8月，已有**12家整机厂商**基于鲲鹏服务器主板推出自有品牌的服务器整机。
- **“伙伴优先”战略案例：**根据搜狐新闻，中国移动2020年PC服务器集采项目中，基于鲲鹏处理器的服务器中标达19563台，其中中移系统集成13475台，华为TaiShan服务器仅6088台，即在合作伙伴的服务器获取了近70%的份额，华为自有品牌的TaiShan服务器只占较小比

鲲鹏服务器主板产品

主板型号	形态	CPU	内存	PCIe扩展
S920X05	2路鲲鹏服务器主板	支持2*鲲鹏920，支持48核/64核，2.6GHz	最多支持32个DDR4 DIMM插槽	最多支持9个的PCIe 4.0插槽 包括一个RAID扣卡或OCP专用的PCIe扩展槽位和8个标准的PCIe扩展槽位
S920X00	2路鲲鹏服务器主板	2*鲲鹏920处理器	最多支持32个DDR4-2933内存插槽	最多支持8个PCIe 4.0 x8或者3个PCIe 4.0 x16 + 2个PCIe 4.0 x8
S920X01	1路鲲鹏服务器主板	1*鲲鹏920处理器	最多支持16个DDR4-2933内存插槽	最多支持3个PCIe 4.0 x8 或者1个PCIe 4.0 x16 + 1个PCIe 4.0 x8
S920S00	2路鲲鹏服务器主板	2*鲲鹏920处理器 (5220或3210)	最多支持16个DDR4-2933内存插槽	最多支持8个PCIe 4.0 x8 或者3个PCIe 4.0 x16 + 2个PCIe 4.0 x8
S920X02	4路鲲鹏服务器主板	4*鲲鹏920 5230/5250/7260可选	最多支持32个DDR4-2933内存插槽，最大内存2TB	最多支持4个PCIe 4.0 x8 + 2个PCIe 4.0 x16标准插槽
S920X03	2路鲲鹏服务器主板	2*鲲鹏920 5230/5250可选	最多支持16个DDR4 DIMM，最高2933MT/s 最大内存1TB	最多支持6个PCIe 4.0 x8标准插槽 或2个全高半长PCIe 4.0 x16和2个PCIe 4.0 x8标准插槽

鲲鹏服务器伙伴整机产品



## 1.5.1 硬件开放：服务器主板及整机

- ◆ **TaiShan服务器：高效能计算+安全可靠+开放生态。** TaiShan服务器是华为新一代数据中心服务器，基于多核华为鲲鹏处理器，适合为大数据、分布式存储、原生应用、高性能计算和数据库等应用高效加速。服务器实现算-存-传-管-智，计算芯片全自研，开放计算平台，支持业界主流软硬件。我们认为Taishan服务器是为华为鲲鹏生态打造的样板。
- ◆ **广泛应用于各行业。** TaiShan服务器在政府、金融、运营商、电力、交通等行业已规模部署和应用。一个典型的应用案例：根据华为官网，华为采用20台TaiShan服务器构建基于鲲鹏架构的计算集群，为中科大打造千万亿次“瀚海20超级计算系统”。该系统是当前中国高校最大高性能计算校级平台之一，采用华为全液冷的解决方案和高密服务器，规模高达30480颗CPU核，为生物信息学计算软件和流体力学提供超强算力。

TaiShan服务器



TaiShan服务器主要产品型号

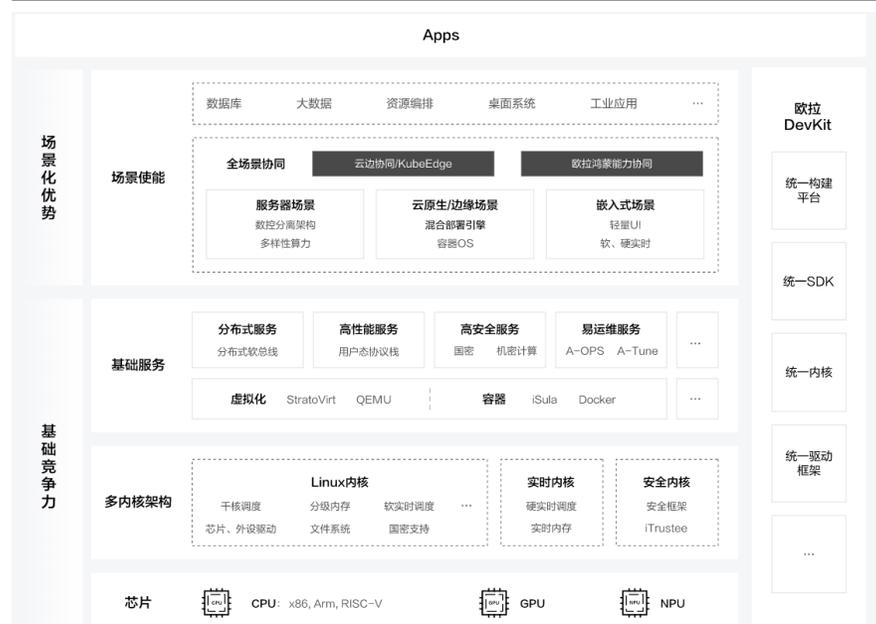


## 1.5.2 软件开源：操作系统

- ◆ **操作系统欧拉openEuler：鲲鹏开源生态的灵魂。** 2019年华为宣布开源服务器操作系统，在openEuler开源社区开放源代码、镜像及开发测试环境；华为全联接2021上，面向数字基础设施的欧拉全新升级。当前openEuler内核源于Linux，支持鲲鹏等多种处理器，适用数据库、大数据、云计算、人工智能等场景。据华为2022年报，已有九家主流操作系统伙伴发布基于openEuler 22.03 LTS版本的商业发行版。
- ◆ **openEuler架构构筑全场景协同领先优势，“欧拉+鸿蒙”生态互通。** 根据鲲鹏官网，openEuler架构具有高性能、高安全和易运维基础能力的特征，其分布式软总线和KubeEdge支撑构筑全场景协同领先优势。根据华为官网，欧拉和鸿蒙已经实现了内核技术共享，2021年升级后续将让搭载欧拉操作系统的设备可以自动识别和连接鸿蒙终端，进一步在安全OS、设备驱动框架、新编程语言等方面实现共享。根据华为2022年年报，openEuler社区吸引超过1.2万名开发者，企业成员已达770余家，合作伙伴基于openEuler社区版推出累计商用超过

300万套

基于 openEuler 的 OS 发行版



### 架构简介

- **全场景协同**  
分布式软总线和KubeEdge，支撑构筑全场景协同领先优势
- **服务器**  
主流多样性算力覆盖全面，生态完善的数字基础设施底座
- **云原生/边缘**  
打造降本增效、敏捷易用云原生/边缘基础设施体验
- **嵌入式场景**  
软、硬实时解决方案，满足工控领域多层次时延诉求
- **高性能**  
内核架构优化、用户态协议栈和内存分级架构，使能最优性能
- **高安全**  
机密计算框架和国密全栈，软硬协同持续构建自主可控安全优势
- **易运维**  
A-OPS智能运维，A-Tune智能调优效率倍级提升



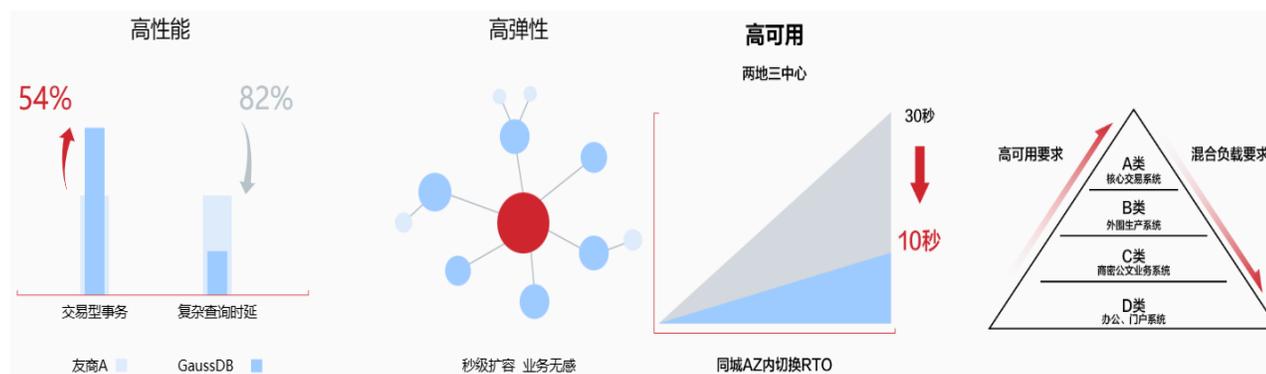
## 1.5.3 软件开源：数据库

- ◆ **OPEN GAUSS：共享企业级数据库能力。** 2019年5月华为GaussDB全球发布；2020年6月华为正式宣布开源数据库能力，开放openGauss数据库源代码，并成立openGauss开源社区。openGauss是一款高性能（两路鲲鹏性能150万tpmC）、高安全（端到端全方位安全防护）、高可靠（故障切换时间RTO<10s）的企业级开源关系型数据库。适用于如大并发、大数据量、以联机事务处理为主的交易型应用和操作与分析并重的物联网数据场景，可覆盖企业70%以上数据库业务场景。据华为2022年报，openGauss联合伙伴完成500多个行业解决方案适配。
- ◆ **最新发布全栈自主、金融级分布式数据库GaussDB，性能提升54%。** 根据观察者网，华为全球智慧金融峰会2023上发布新一代分布式数据库GaussDB，未来将深耕金融场景。根据华为云官网，该产品支持分布式事务，支持1000+节点的扩展能力，PB级海量存储，目前产品已支撑多个国有大行核心业务。根据观察者网，在与中国邮政储蓄银行的合作中，邮储银行向华为开放了6.5亿用户的银行分布式新核心系统建设机会，系统中使用了GaussDB，目前该系统已全面投产上线，可实现日均20亿笔交易、峰值6.7万笔/秒的能力，新系统效率平均提升40%。

华为openGauss产品：商用+自用+开源



华为金融级分布式数据库GaussDB产品



## 1.5.4 软硬件联手共建生态：合作伙伴一览

### 华为产业链生态合作伙伴

#### 硬件及配件

##### 服务器/整机

神州数码	子公司神州鲲泰，2019年对华为鲲鹏系列产品进行软件适配。2020年与麒麟软件、统信软件等15家合作伙伴签署合作协议；同时神州数码是第一批基于昇腾的合作伙伴之一
东华软件	2020年6月，东华软件与华为联合宣布基于鲲鹏处理器的“鹏霄”服务器，在宁波高新区正式投产下线
拓维信息	华为鲲鹏/昇腾AI+开源鸿蒙的软硬一体战略合作伙伴，子公司湘江鲲鹏
广电运通	2022年，旗下广电五舟与华为举行了PoweredbyAscend(PBA)昇腾产业生态品牌授权签约仪式，后续公司成为昇腾AI生态的重要参与者和共建者
远东传动	2019年12月，远东传动参股的黄河科技公司基于鲲鹏芯片的服务器与PC机生产线已投产
四川长虹	2020年7月成立子公司华鲲振宇，负责西南地区基于华为鲲鹏芯片的服务器、PC机、视觉等系列产品。2022年位列“鲲鹏+昇腾”生态销量规模第一，是华为生态伙伴中唯一获得“鲲鹏+昇腾”双领先级认证的企业
同方股份	在华为鲲鹏、昇腾、存储产品线全面合作。针对AI产品的算力部分，得益于华为Ascend芯片的算力能力，目前在满足训练场景的FP16及满足推理场景的INT8的算力能力上，可以对标英伟达的部分产品

#### BOIS

卓易信息	大陆唯一X86架构BIOS独立供应商，为华为（ARM架构）提供BIOS固件技术服务
------	---

#### 基础软件

##### 操作系统

诚迈科技	2019年，子公司深之度和华为合作参与UOS统一操作系统筹建
润和软件	2020年，公司携手华为鲲鹏发布HopeEdge V1.0（首款面向边缘的物联网操作系统）；同年，依托华为昇腾Atlas 200DK发布智慧交通解决方案
中国软件	2021年1月，中软旗下麒麟软件和华为鲲鹏联合实验室成立，强化“华为鲲鹏芯片+麒麟操作系统”基础平台

#### 中间件

东方通	东方通是华为第一批重点合作的中间件厂商，2019年9月其应用服务器中间件Tongweb产品入驻华为云商城；2023年5月东方通网络内容风控产品V2.0顺利通过华为技术认证，与华为昇腾Atlas系列完成兼容性测试
宝兰德	宝兰德率先获得华为Validated认证首家完成中间件领域华为鲲鹏技术全栈优化，鲲鹏计算产业生态重要伙伴
普元信息	2022年，普元EOS应用开发平台凭借基于鲲鹏基础技术架构体系进行兼容性验证和全栈调优的多项优势，入选《鲲鹏精选解决方案》，湖南省鲲鹏生态创新中心推荐的EOS开发平台，是普元的全栈式信创中间件产品之一

#### 数据库

太极股份	2013年太极股份和华为签署战略合作伙伴协议成为华为公司多产品一级经销商，子公司人大金仓是数据库龙头
海量数据	2020年openGauss开源后,海量数据基于openGauss推出数据库Vastbase G100
爱可生	2021年华为联合爱可生发布面向分布式数据库的创新存储方案
东方国信	2019年华为与东方国信联合发布了基于GaussDB数据库的政务数据治理联合解决方案。公司自主研发的AI深度学习应用层产品，与华为昇思MindSpore（盘古大模型的底层AI框架）完成兼容认证，加入华为昇思MindSpore开源社区
美亚柏科	公司的“乾坤”大数据操作系统已同华为云鲲鹏云服务器完成兼容性互认证

## 1.5.4 软硬件联手共建生态：合作伙伴一览

### 华为昇腾四大行业生态合作伙伴

#### 昇腾智造

昇腾智造是面向制造行业打造的一站式、交钥匙的人工智能质检解决方案。

- ◆ **智能检测**：基于昇腾的AI质检解决方案首先在华为南方工厂上线，面向电子器件组装错漏检测、工业OCR文字识别、标签印刷缺陷检测等典型场景，华为持续技术积累，将行业know-how沉淀于算法和软件中，实现AI质检应用的快速开发；
- ◆ **精准判别**：通过制造行业SDK帮助合作伙伴和客户提升AI应用开发效率，降低开发门槛。目前，该方案已在南方工厂190+产线实现了稳定高效运用，将质检工作量降低了超60%，准确率超99.9%

合作伙伴：

宝德、拓维信息、长江计算、依瞳科技、中科弘云、华睿科技、沃德普、凌华科技、研华科技、东声智能

#### 昇腾智城

昇腾智城是面向智慧城市，通过人工智能加速视频分析，帮助城市治理更聪明、更精细化的解决方案。

- ◆ **城市治理**：聚焦城管、应急、水务等场景，构建统一、共享的AI服务平台。建立城市事件库，提供自动发现、智能分拨、全域感知等功能；
- ◆ **政务服务**：将AI技术融入到智慧导办、智慧热线、智慧审批、智慧问答等服务中，使能智慧服务全流程，打造“秒批秒办”的政务服务；
- ◆ 通过易用MindX SDK套件的开放，有效降低AI应用开发门槛，协同行业伙伴共同打造繁荣ISV生态，赋能智慧城市全场景。

合作伙伴：

云天励飞、安擎、云从科技、宝德、芯翌科技、黄河、千视通、华鲲振宇、极视角、神州鲲泰、文安智能、以萨技术、中科弘云

#### 昇腾智行

昇腾智行是面向智慧交通的解决方案，用人工智能技术对车辆、轨迹等进行智能分析。

- ◆ **高速自由流通行**：依托昇腾AI车道控制器“AI使能”、“云边协同”、“安全可靠”等关键特性，帮助高速路网通行效率提升高达5倍；
- ◆ 在节能减排的同时，实现道路的安全控制和协同管理；·**收费稽核，精准计费**：打造AI大数据稽核平台和边缘AI一体机，实现车辆精准识别和计费；
- ◆ **视频云联网**：通过“AI+云网协同”，提供视频转码共享和云边协同智能分析服务，助力高速公路交通监测管控效率的全面提升。该方案已部署在京珠南、广深、广惠等20多条高速公路
- ◆ **车路协同**：通过多元信息采集融合感知处理，实时监测高速车辆行驶状态，实现车辆编队行驶、路面异常预警、网联信息发布。在北京延崇等地试点，通信效率提升10%-30%，事故率降低30%-50%。

合作伙伴：

拓维信息、中远海运、千视通、格灵深瞳、国交润万、爱笔智能、北联国芯、中科视语卓视智通、麦图科技

#### 昇腾智巡

昇腾智巡是面向能源电力行业智慧巡检解决方案，用人工智能的分析取代了传统的人工巡检，让巡检更安全，效率更高。

- ◆ **远程自动巡视，边缘实时分析**：通过边缘计算与AI的结合，实现输电线、变电站、配电房电力设备状态和周界环境的智能分析和实时预警，助力构建智能电网。
- ◆ **提升电网运营效率，降低运维成本**：输电线巡检效率相比传统方案提升80倍；
- ◆ **电网站房巡检更加实时和精准**，人工巡检工作量大幅降低，供电可靠性显著提高。

合作伙伴：

智洋创新、南瑞继远、云化能源、华雁智能金三立、亚联发展、佳华科技、旷视科技

## 1.5.5 软硬件联手共建生态：BoostKit 加速库

- ◆ **加速库：释放硬件算力潜能，实现应用软件性能。**加速库介于硬件资源和解决方案之间，对上应用于各种解决方案，如大数据、分布式存储、数据库等，从而支撑于政府、运营商、金融等领域；对下可适配主流操作系统如openEuler等，通过充分发挥硬件的能力，或者内嵌于操作系统，以及单独作为一个函数库去提升解决方案的性能。我们认为加速库是软硬件结合构造一体化生态的重要桥梁。
- ◆ **鲲鹏 BoostKit 加速库：提供基于 ARM 指令深度优化和基于鲲鹏KAE（鲲鹏硬件加速引擎）开发的加速库。**对常见的开源库做进一步深度性能优化，同时补齐鲲鹏生态链上缺失的关键性能库，为鲲鹏应用提供极致的性能保障。覆盖系统库、压缩、加解密、媒体、数学库、存储、网络等7类加速库，为大数据加解密、分布式存储压缩、视频转码等应用场景提供高性能加速。

鲲鹏BoostKit加速库解决方案



鲲鹏BoostKit加速库

系统库	压缩	加解密	媒体	数学库	存储	网络
Glibc 指令加速	Gzip 指令加速	KAE加解密 OpenSSL SM3/SM4/RSA /AES/MD5	HMPP 超媒体性能库	KML KML_FFT KML_BLAS KML_SPBLAS KML_MATH KML_VML KML_LAPACK KML_SVML KML_SOLVER KML_SCALAPACK KML_VSL	Smart Prefetch 智能预取	XPF OVS流表加速库
HyperScan 指令加速	ZSTD 指令加速		X265 开源H.265视频转码		SPDK SSD用户态驱动	DPDK 用户态网络驱动
AVX2Neon 异构生态迁移	Snappy 异构生态迁移		X264 H.264视频编解码		ISA-L 存储加速库	
HAF 同构加速框架	KAEzip 压缩硬加速		KVSIP 鲲鹏矢量信号处理库			

— 软加速库 — 硬加速库

## 1.5.6 软硬件联手共建生态：应用使能套件 BoostKit

- ◆ **鲲鹏应用使能套件 BoostKit，提供“数据亲和”组件：**根据搜狐新闻，2021年9月在华为全联接2021上，鲲鹏应用使能套件BoostKit正式升级为2.0。鲲鹏 BoostKit 基于硬件、基础软件和应用软件的全栈优化，提供高性能开源组件、基础加速软件包和应用加速软件包 BoostKit 2.0提供四类“数据亲和”加速组件，包括数据就近计算，数据加速传输，数据并行化处理，数据安全等。**我们认为“数据亲和”使得数据与应用间绑定更紧密，是计算创新的大趋势。**
- ◆ **BoostKit深入行业应用。**根据搜狐新闻，1) 在**运营商行业**，联通（广东）产业互联网有限公司基于鲲鹏自主研发天玑安全平台，鲲鹏 BoostKit将等保测评响应时间缩短50%，安全日志数据吞吐能力提升30%；2) 在**金融行业**，中金金融认证中心有限公司打造基于鲲鹏的安全一体机解决方案，鲲鹏BootsKit KAE加速引擎将加密性能提升10倍以上；3) 在**医疗行业**，中山大学第一附属医院依托鲲鹏全栈平台开展精准医学尖端科研，生物信息分析软件性能提升20%以上。4) 在**交通行业**，南方航空打造基于鲲鹏的航空运行管理平台-南航E家平台，将平台的差旅吞吐量提升1倍

华为鲲鹏应用使能套件 BoostKit



BoostKit 2.0 “数据亲和”加速组件



## 1.5.7 软硬件联手共建生态：开发者社区

- ◆ **沃土计划2.0全面升级，计划五年发展500万开发者。**根据华为官网新闻，2015年首届华为开发者大会宣布沃土1.0计划；2019年华为全联接大会上，发布最新沃土计划2.0，宣布未来5年投入15亿美金，将从产品（基于鲲鹏+昇腾计算处理器）、赋能、联盟、社区（按行业建设应用生态，按区域建设产业生态）、激励等5大方面升级，计划五年内在全球发展500万开发者。2020年3月华为开发者大会2020（Cloud）上，华为宣布沃土计划2.0的进一步举措，为开发者打造了1个云上开发平台+1套开发工具+1个开发者社区。
- ◆ **华为开发者社区：一站式知识共享平台。**华为推出的“开发者社区”平台，包含众多与开发者直接相关的Offering，提供从了解、学习、开发、测试、认证、发布到商业化的端到端技术支持。据华为2022年年报，鲲鹏已有4,200多家合作伙伴、190万名开发者、12,000多个解决方案通过鲲鹏兼容性测试认证。昇腾已发展超120万开发者，社区开源贡献者超过8,700人，与1,000多家ISV联合推出2,000多个行业AI解决方案。

鲲鹏开源社区数据总览（截至2023年6月11日）

鲲鹏开源社区	openEuler社区	openGauss社区	openLooKeng社区
社区用户	1,376,358	1,905,869	152,614
贡献者	14,605	4,938	3,084
单位成员	905	289	27
商用	18	12	
请求合并 (K)	109.7	11.8	2.2

昇腾 昇思MindSpore社区数据总览（截至2023年6月11日）



## 1.6 22年: 行业信创延展元年，华为有望优先受益

- ◆ **2022年是党政信创向行业信创延展落地的元年**：2021年信创产业逐步走向应用落地阶段，“适配生态”、“竞标中标”、“产品迭代”等关键词彰显了信创产业在“2+8”行业应用中的良好态势。2022年是信创正式开启规模化应用，一个关键的变化是从党政信创向行业信创延展。
- ◆ **“二十大”明确“安全”的重要战略地位**。2022年10月，中共“二十大”正式召开，根据大众日报统计，党的二十大报告中“安全”一词贯穿全篇，共出现了91次，创历年之最。报告第十一部分用专章对维护国家安全和社会稳定进行安排，突出了国家安全在党和国家工作大局中的重要地位。
- ◆ **我国信创相关政策加速落地**。据零壹智库、数字化讲习所不完全统计，截至2022年10月底，中央及各地方各级政府共发布信创相关政策166部，其中，中央层面政策35条，地方层面政策131条。

### “二十大”报告独立章节系统阐述安全

#### 十一、推进国家安全体系和能力现代化，坚决维护国家安全和社会稳定

国家安全是民族复兴的根基，社会稳定是国家强盛的前提。必须坚定不移贯彻总体国家安全观，把维护国家安全贯穿党和国家工作各方面全过程，确保国家安全和社会稳定。

我们要坚持以人民安全为宗旨、以政治安全为根本、以经济安全为基础、以军事科技文化社会安全为保障、以促进国际安全为依托，统筹外部安全和内部安全、国土安全和国民安全、传统安全和非传统安全、自身安全和共同安全，统筹维护和塑造国家安全，夯实国家安全和社会稳定基层基础，完善参与全球安全治理机制，建设更高水平的平安中国，以新安全格局保障新发展格局。

(一) 健全国家安全体系。坚持党中央对国家安全工作的集中统一领导，完善高效权威的国家安全领导体制。强化国家安全工作协调机制，完善国家安全法治体系、战略体系、政策体系、风险监测预警体系、国家应急管理体系，完善重点领域安全保障体系和重要专项协调指挥体系，强化经济、重大基础设施、金融、网络、数据、生物、资源、核、太空、海洋等安全保障体系建设。健全反制裁、反干涉、反“长臂管辖”机制。完善国家安全力量布局，构建全域联动、立体高效的国家安全防护体系。

(二) 增强维护国家安全能力。坚定维护国家政权安全、制度安全、意识形态安全，加强重点领域安全能力建设，确保粮食、能源资源、重要产业供应链安全，加强海外安全保障能力建设，维护我国公民、法人在海外合法权益，维护海洋权益，坚定捍卫国家主权、安全、发展利益。提高防范化解重大风险能力，严密防范系统性安全风险，严厉打击敌对势力渗透、破坏、颠覆、分裂活动。全面加强国家安全教育，提高各级领导干部统筹发展和安全能力，增强全民国家安全意识和素养，筑牢国家安全人民防线。

### 2006-2022 年信创相关政策数量变化趋势



## 1.6 22年: 行业信创延展元年，华为有望优先受益

- ◆ **信创“专项激励”政策持续落地，向区县下沉。** 2022年11月，金华市发布《支持信息技术应用产业发展的若干政策意见》，提出“在充分竞争的前提下，鼓励在政府采购中积极采购信息技术应用企业所开发的产品”。我国地方性信创专项激励政策陆续出台，目前已有广州、天津、黄陂、苏州、太原等地发布相关政策，将促进党政和行业信创规划持续落地。

地方性信创专项激励政策陆续出台

时间	发布单位	政策名称	重点内容
2022年11月	浙江省金华市	《支持信息技术应用产业发展的若干政策意见》	步及信息技术应用产业投资、人才团队引进、企业发展上台阶、信息技术应用产品推广四大方面。明确：在充分竞争的前提下，鼓励在政府采购中积极采购信息技术应用企业所开发的产品。
2022年11月	山西省太原市小店区	《小店区“信创产业集聚区”专项奖励办法》	支持基于国产自主芯片的硬件、软件产品研发生产、应用迁移、适配测试、信息服务在能源、金融、政务、教育、医疗、电力、交通、旅游、司法等领域及“水暖煤农”等北方特色行业的应用项目，鼓励企业申报太原市相关部门遴选产品应用示范项目，在获得市级奖补的基础上，按照市奖励金额的20%予以奖励，最高不超过100万元。
2022年	江苏省苏州市吴中区	《苏州吴中经济技术开发区关于促进信息技术应用创新产业发展的若干政策》	15条专项政策对信创项目引进、龙头企业培育、产业基础提升、信创产业集聚、研发能力提升、标准体系建设、学术交流等给与相应的补贴支持。
2022年5月	湖北省武汉市黄陂区	《黄陂“信创8条”专项政策措施》	每年设立5000万元信创产业专项支持资金，用于对从事信息技术应用创新产业发展的相关项目引入、自主研发、应用创新、生产销售、人才引育、融资服务等奖励政策的兑现保障，引导推动信创产业发展。

## 1.6 22年: 行业信创延展元年，华为有望优先受益

- ◆ **信创招投标持续加速。**2022年12月，建设银行发布《国产芯片服务器采购项目》中标公示：飞腾芯片服务器 7411.07 万元，鲲鹏芯片服务器 36425.06 万元、海光芯片服务器 15847.07 万元。2022 年 11 月，建设银行发布《国产操作系统软件》中标公示。麒麟软件 9900 万中标。2022 年以来，运营商，银行等行业招标结果陆续公布，国产化占比持续提升，信创订单加速放量。

国产芯片、服务器采购项目

时间	发布单位	项目名称	中标情况
2022年12月	中国建设银行	《全行台式电脑（国芯）采购项目》	中标人：联想、浪潮信息、紫光恒越。采购价格12440.76万元（预估金额）
2022年12月	中国建设银行	《国产芯片服务器采购项目》	1) 飞腾芯片服务器7411.07万元；2) 鲲鹏芯片服务器36425.06万元；3) 海光芯片服务器15847.07万元。
2022年12月	中国建设银行	《国产操作系统软件采购项目》	麒麟软件9900万元
2022年6月	中国农业银行	《2022年第一批PC服务器项目》	1) 中科可控7.14亿元（主选供货商）。2) 浪潮信息6.61亿元（主选供货商）；3) 中兴通讯7.39亿元（备选供货商）

运营商信创招投标加速

运营商	采购项目	总数量(台)	国产CPU系列数量(台)	国产CPU系列占比(%)
中国电信	2020年服务器采购	56314	H系列(国产CPU: 海光或者鲲鹏)11185台	19.86%
中国电信	2020-2021年服务器集采	69952	H系列(国产CPU: 海光或者鲲鹏)24823台	35.50%
中国电信	2021-2022年服务器集采	200000	G系列(国产CPU)53401台	26.70%(其中, 鲲鹏: 28%, 海光: 72%)
中国移动	2020年PC服务器集采	138272	鲲鹏CPU: 19563台	14.15%
中国移动	2021-2022年PC服务器集采第1批及补充采购	204696	国产CPU: 85251台(鲲鹏 CPU: 44837台, 海光CPU: 40414台)	国产CPU: 41.65%(鲲鹏: 21.90%, 海光: 19.74%)
中国移动	2021-2022年PC服务器集采第2批(标包1-9)	82245	国产CPU: 33632台(鲲鹏 CPU: 14064台、海光CPU: 19568台)	国产CPU: 40.89%(鲲鹏: 7.10%, 海光: 23.79%)

## 1.6 22年: 行业信创延展元年，华为优先受益

- ◆ **华为鲲鹏CPU是信创的核心供应商:** Kunpeng处理器是鲲鹏计算产业基座，华为持续保持重点投入，秉承量产一代、研发一代、规划一代的演进节奏。2019年，华为旗下海思半导体自研芯片推出数据中心高性能处理器产品鲲鹏 920，鲲鹏 920 处理器兼容 ARM 架构，采用 7nm 工艺制造，可以支持 32/48/64 个内核，主频可达 2.6GHz，综合性能世界领先。我们认为22年在信创的大背景下，以鲲鹏为CPU的服务器有望快速放量。
- ◆ **华为信创产业生态持续完善:** 根据华为全联接大会 2022 披露数据，目前，搭载 HarmonyOS的华为终端设备超过 3.2 亿；HMS 海外 Top 3000 应用集成度达 80%；华为云开发者总数突破 350 万；鲲鹏和昇腾伙伴数量超过 5000 家、解决方案数量达 13000 个、开发者总数超过 260 万。欧拉实现从企业主导到产业共建，累计装机量达 245 万套，我们预计 2022 年在中国服务器操作系统领域新增市场份额将达 25%。

华为生态稳步发展



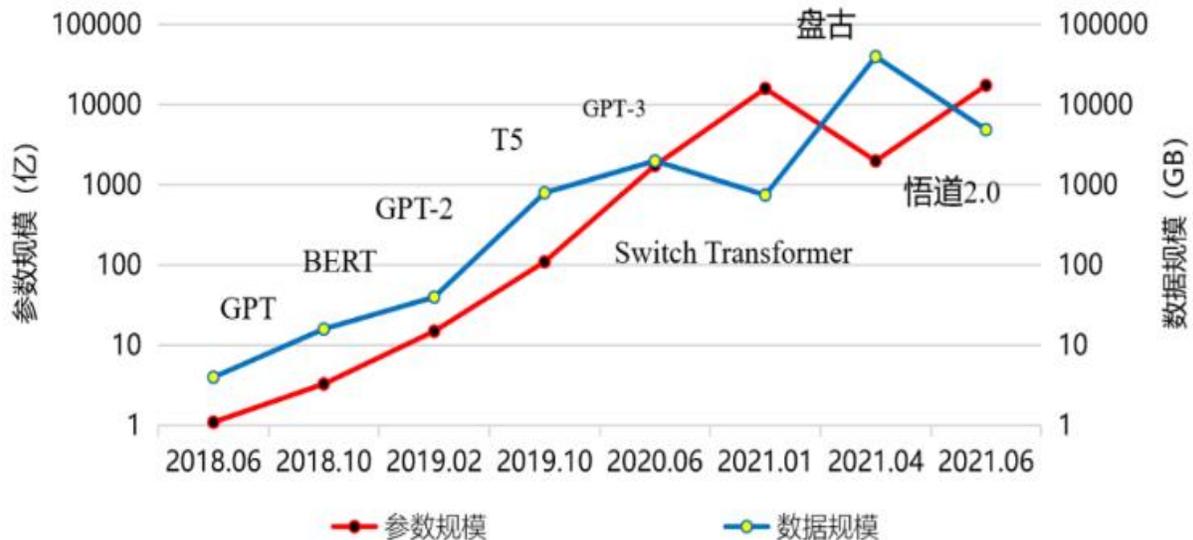
华为兼容开放的基础软件生态



## 1.7.1 23年：再三强调，大模型背景下算力势必迎来爆发

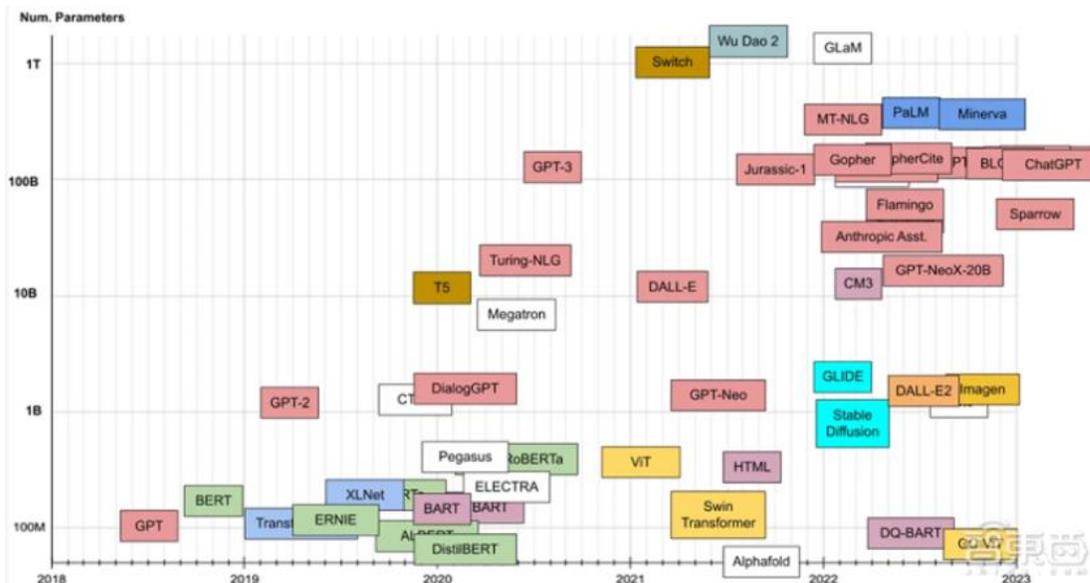
- ◆ **ChatGPT开启算力军备赛**：我们已经在《ChatGPT：百度文心一言畅想》中证明数据、平台、算力是打造大模型生态的必备基础，且算力是训练大模型的底层动力源泉，一个优秀的算力底座在大模型(AI算法)的训练和推理具备效率优势；同时，我们在《ChatGPT打响AI算力“军备战”》中证明算力是AI技术角逐“入场券”，其中AI服务器、AI芯片等为核心产品；此外，我们还在《ChatGPT，英伟达DGX引爆AI“核聚变”》中证明以英伟达为代表的科技公司正在快速补足全球AI算力需求，为大模型增添必备“燃料”。
- ◆ **大模型参数呈现指数规模，引爆海量算力需求**：根据财联社和OpenAI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，华西证券研究所

近年大模型的参数规模增长趋势



## 1.7.2 23年：华为盘古大模型赋能千行百业

- ◆ **华为盘古大模型：千亿参数，赋能多行业。**根据《华为技术》，2021年4月，华为盘古大模型正式发布，为此团队**调用超2000块昇腾910，进行超过2个月训练**。其中，盘古NLP大模型是业界首个千亿参数的中文预训练大模型，训练过程中使用了40TB的文本数据，包含大量通用知识与行业经验。盘古CV大模型也在业界首次实现不同部署场景下按需抽取模型，从特定的小场景到综合性的复杂大场景均能覆盖，同时其提出的基于样本相似度的对比学习，实现了在ImageNet上小样本学习能力上的业界第一。截至2022年1月，盘古大模型已经在100多个行业场景完成验证。
- ◆ **华为PanGu-Σ模型参数达1.085万亿，语言能力突出。**据2023年3月华为发布的PANGU-Σ论文，PanGu-Σ模型参数为1.085万亿，基于华为自研的MindSpore框架开发；团队收集了4个域的数据集，其中来自CLUECorpus2020，C4的双语数据集包括77.51B ( Chinese (38.75) + English(38.76B) )；在对话能力，PanGu-Σ可以生成多样化、语义一致、知识丰富的结果，回答事实性问题具有高准确率。

华为云盘古预训练大模型架构



华为PanGu-Σ大模型语言能力评估示例

Table 9: A case of topic-grounded dialog

Dialog history	你最近有看中国排球比赛吗? (Have you watched the Chinese volleyball match recently?) 就是啊,当然有在看,中国女排这几次的比赛都很优秀。 (Of course. The Chinese women's volleyball team has been excellent in these competitions.) 我觉得中国女排能有现在的成功也离不开教练郎平的教导。 (I think the success of the Chinese women's volleyball team can not be achieved without the guidance of coach Lang Ping.)
EVA2.0	就是讲说啊,我觉得还是很优秀的能够这么成功不容易啊。 (Yes, they are excellent, and it is not easy to be so successful.) 背后一定付出很多,我以前打过排球,很容易受伤。 (They must have payed a lot of effort. I've played volleyball before, and it's easy to get hurt.)
PanGu-Bot 2.6B	是啊是啊!真的不容易。 (Yeah yeah! it's really not easy.) 所以我们要好好的保护自己,不让自己受伤。 (So we need to protect ourselves from getting hurt.)
PanGu-Σ	郎平的执教也是很有一套的,她还入选过世界排球名人堂呢。 (Lang Ping is also a good coach. She has also been selected into the World Volleyball Hall of Fame.)

Table 11: A case of open domain question-answering

Question:	中国的四大发明是哪四个?	What are China's four major inventions?
EVA2.0:	我觉得是发明了“发明”这个词的人	I think it's the man who invented the word 'invented'.
PanGu-Bot 2.6B:	造纸术、火药术、火药武器、火药武器	Papermaking, gunpowder, gunpowder weapons, gunpowder weapons
PanGu-Σ:	造纸术、印刷术、指南针、火药	Papermaking, printing, compass, gunpowder

## 1.7.3 23年：我国超算规模持续高增

- ◆ **超算到智算，“AI+”步入“+AI”时代。**超算中心立足于科学研究，是支撑国家科研体系的大科学装置，主要解决大系统、大工程、大科学的问题。而人工智能计算中心立足于赋能产业，是支撑数字经济的基础设施，可为大规模AI算法和模型研究形成条件支撑，实现数字经济高质量发展。区别于传统的超算中心，智算中心建设在当前AI高速发展背景下大势所趋。
- ◆ **从政策到产业，推动智算规模高增。**2023年1月，国家信息中心发布《智能计算中心创新发展指南》，提出构建智算中心的“四化”技术路线，强调算法基建化的重要性。根据智东西，2021年到2022年初，全国有不低于20座城市建成或在建智算中心，智算中心数量达到27个，有12个位于八大枢纽。据IDC统计，2022年我国智能算力规模达268EFLOPS，预计到2026年，或将增至1271.4EFLOPS。

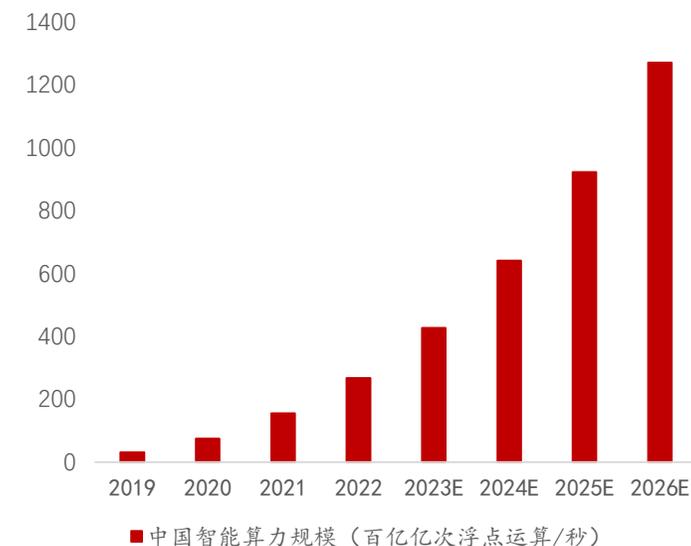
我国智算产业相关政策

时间	政策/产业部署	发布主体	内容
2023年1月	《智能计算中心创新发展指南》	国家信息中心	提出构建智算中心的“四化”技术路线：以算力基建化为主体，使得AI算力成为城市的公共基础资源，供政府、企业、公众按需使用；以算法基建化为引领，服务模式从提供算力为主向提供“算法+算力”转变；以算法基建化为引领，以低代码甚至无代码开发的模式，为用户使用便捷的智能算力；以设施绿色化为支撑；通过采用液冷技术等节能降碳技术
2020年11月	《智能计算中心规划建设指南》	国家信息中心	提升AI算力生产供应，智算中心基于新型硬件架构和人工智能算法模型，保证规划建设的技术领先性；促进数据开放共享，汇聚各行业领域数据资源，全面提升AI算法训练数据质量；培育区域智能生态；推动AI产业创新聚集推动AI产业创新聚集，加速AI应用场景落地
2017年7月	《新一代人工智能发展规划》	国务院	首次提及智算中心概念。强调建设布局人工智能创新平台，重点突破人机协同的感知与执行一体化模型等核心技术，建立人工智能超级计算中心、大规模超级智能计算支撑环境、在线智能教育平台等

8大国家算力枢纽智算中心（2021.1.1-2022.2.15）

国家算力枢纽节点	建成或正在建设的智算中心
甘肃枢纽	庆阳智算中心
京津冀枢纽	中国电信京津冀大数据智能算力中心
	河北人工智能计算中心
长三角枢纽	商汤科技人工智能计算中心
	南京智能计算中心
	昆山智算中心
	杭州人工智能计算中心
	腾讯智慧产业长三角（合肥）智算中心
粤港澳大湾区枢纽	合肥先进计算中心
	广州人工智能公共算力中心
	深圳市人工智能公共算力中心
成渝枢纽	成都智算中心

中国智能算力规模预测（百亿亿次浮点运算/秒）



## 1.7.4 23年：算力大爆发，华为昇腾开启新一轮成长曲线

- ◆ **华为以AI算力为底座，开启新一波成长曲线。**在智算大趋势下，华为为加快千行百业数字化转型和智能化升级，计划构建面向通用计算的鲲鹏计算产业和面向人工智能的昇腾AI计算产业。
- **鲲鹏计算**：根据新华网财经，截至2023年5月，已有**超过4500家合作伙伴、200多万名开发者加入鲲鹏，13500多个解决方案通过鲲鹏适配认证**。鲲鹏已在政府、金融、电力、运营商等多个国计民生领域得到规模应用，逐步成为数字基础设施算力底座的首选。一方面，华为从硬件创新走向基础软硬件协同全栈创新；另一方面，从单点创新走向系统级架构创新，通过计算、存储、交换的一体化设计、全液冷技术的引入，比如打造天成全液冷多样性算力平台，相比传统机柜实现4-8倍的算力密度提升，机房面积节省约70%。

鲲鹏计算产业生态策略



鲲鹏生态发展情况（截至2022年11月）



## 1.7.4 23年：算力大爆发，华为昇腾开启新一轮成长曲线

- ◆ **华为以AI算力为底座，开启新一波成长曲线。**在智算大趋势下，华为为加快千行百业数字化转型和智能化升级，计划构建面向通用计算的鲲鹏计算产业和面向人工智能的昇腾AI计算产业。
- **昇腾智算：**根据新华网财经，华为昇腾计算业务总裁张迪煊表示，“昇腾AI”基础软硬件平台已孵化和适配了30多个主流大模型，我国一半以上的原生大模型是基于“昇腾AI”基础软硬件平台打造，包括鹏程系列、紫东系列、华为云盘古系列等。已有25个城市基于昇腾建设人工智能计算中心，其中14个已经上线并饱和运行，为本地数字经济发展提供坚实支撑；同时，华为还与全国280多所高校深度合作，探索人才培养与产教协同模式创新，昇腾AI已赋能3000多名老师，累计培养16万高校人才。根据电子信息产业网，华为计算产品线总裁邓泰华在华为全联接大会2022上表示，截至2022年11月，昇腾AI已有20多个算力节点接入中国算力网。

昇腾提供平台原生力



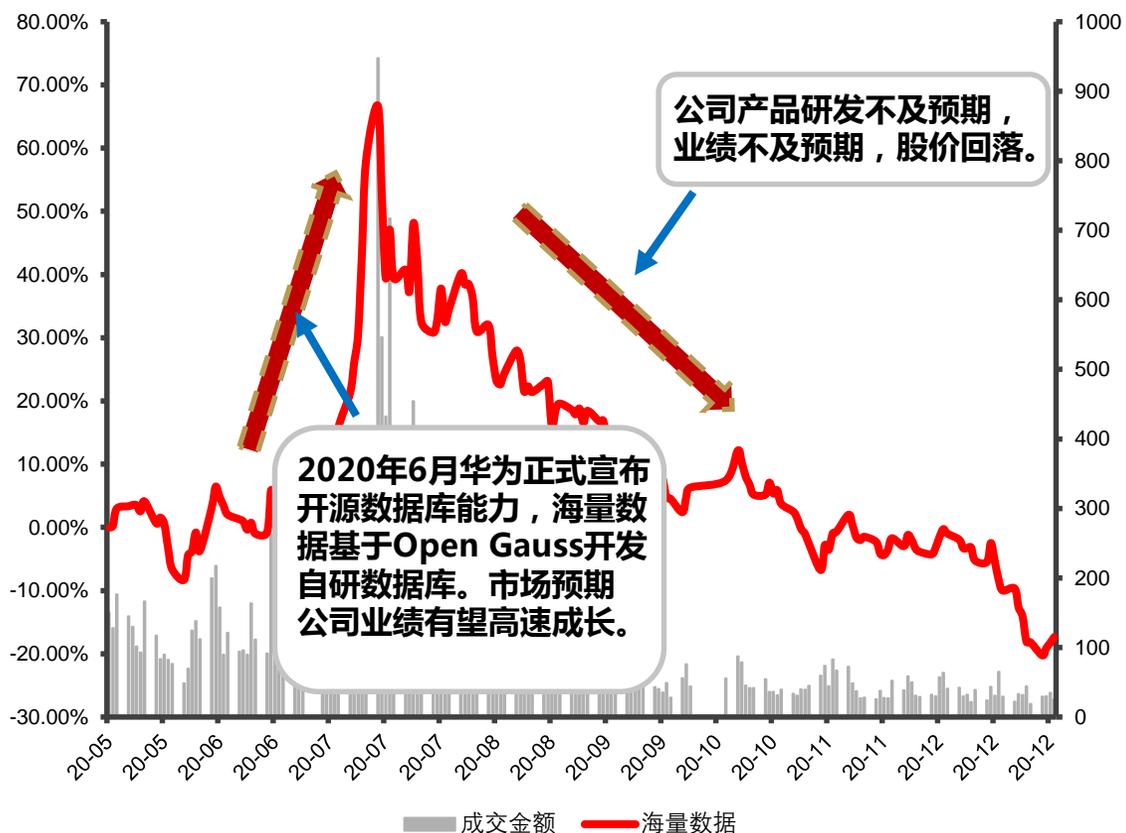
华为基于昇腾处理器的人工智能计算中心方案架构



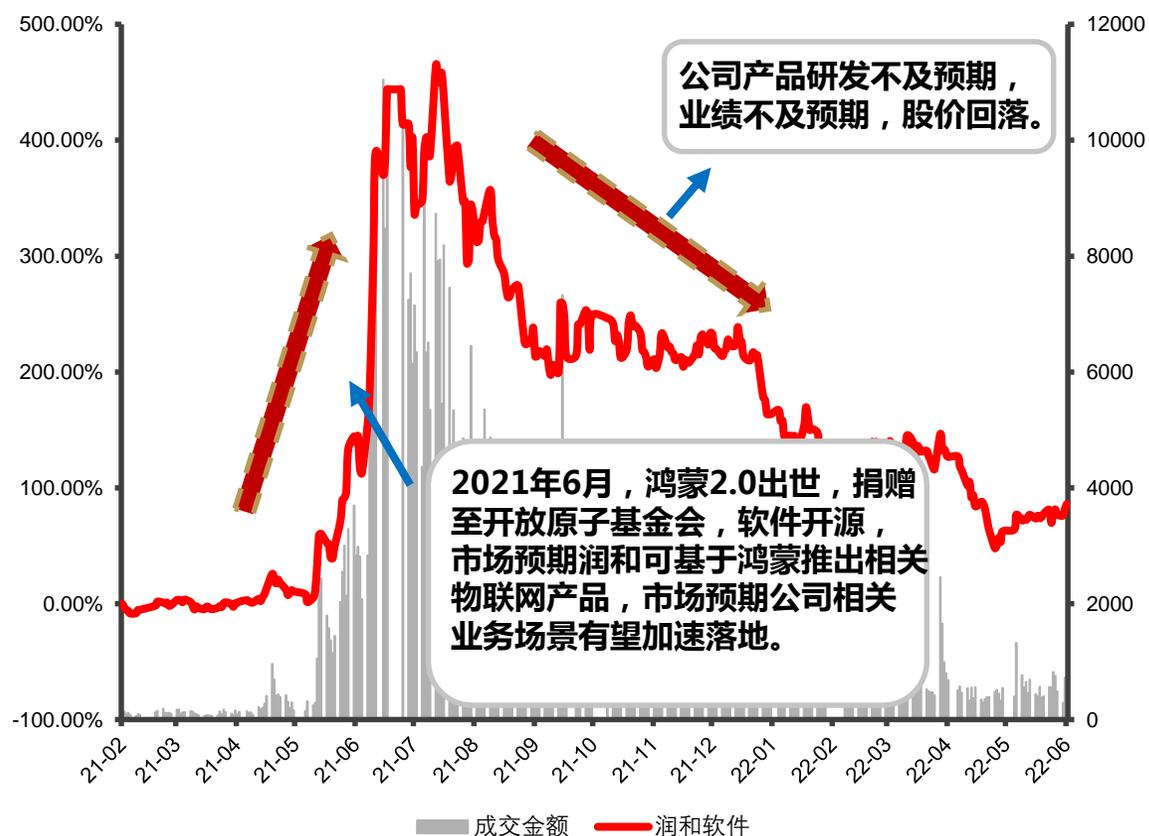
## 1.8 华为重点合作厂商复盘，依托昇腾AI开启新一轮成长曲线

- ◆ 我们以华为生态合作伙伴为例，复盘华为产业链成长曲线，我们判断华为产业链正依托昇腾AI开启新一轮成长曲线。

海量数据20年5月至20年12月股价复盘



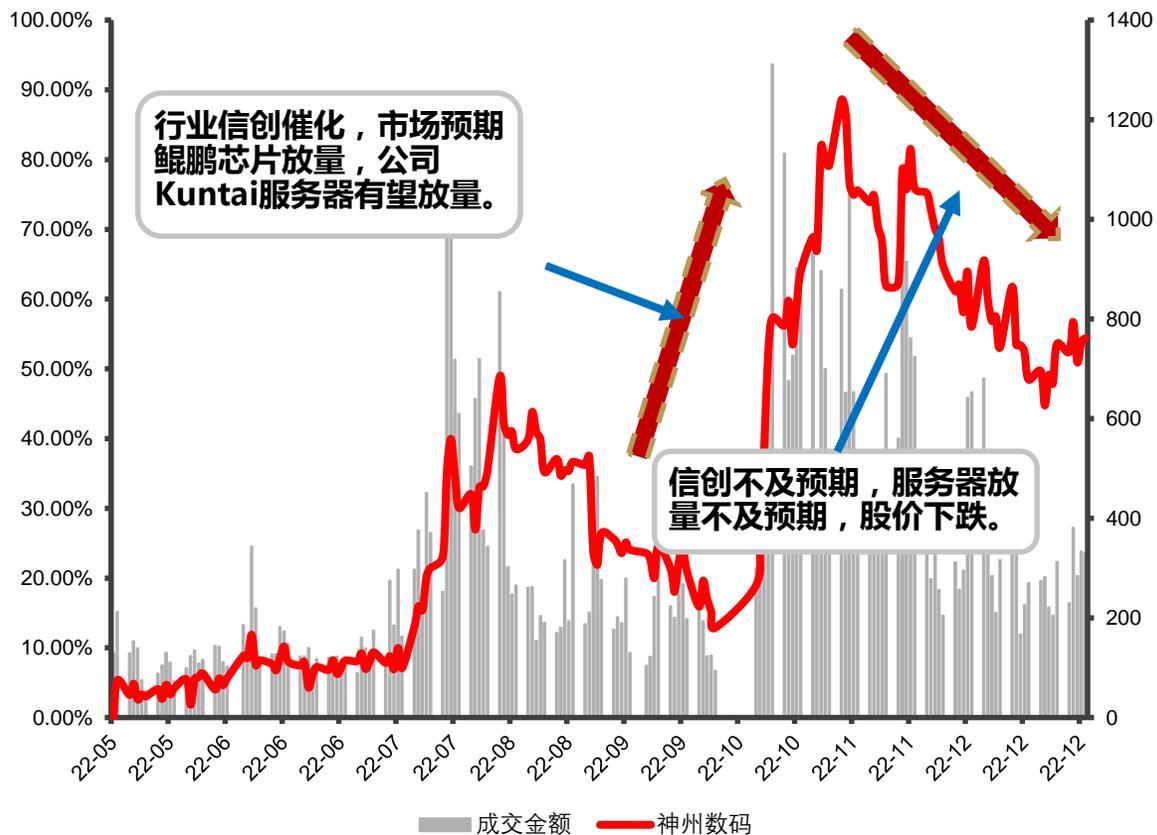
润和软件21年2月至22年6月股价复盘



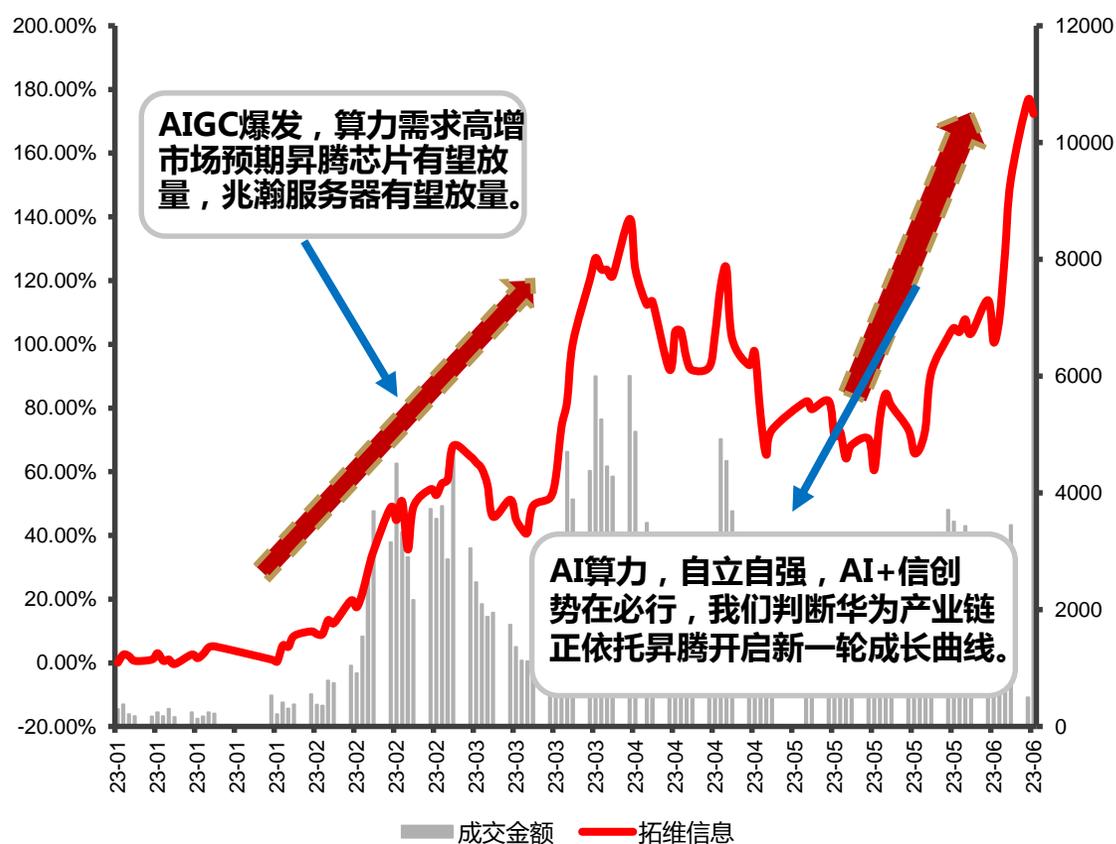
## 1.8 华为重点合作厂商复盘，依托昇腾AI开启新一轮成长曲线

- ◆ 我们以华为生态合作伙伴为例，复盘华为产业链成长曲线，我们判断华为产业链正依托昇腾AI开启新一轮成长曲线。

神州数码22年5月至22年12月行情复盘



拓维信息23年初至今行情复盘



## 1.9.1 AI产品交付-基础硬件层：昇腾芯片

- ◆ 华为海思昇腾芯片是两款人工智能处理器，包括昇腾910和昇腾310处理器，采用自己的达芬奇架构。昇腾910支持全场景人工智能应用，而昇腾310主要用在边缘计算等低功耗的领域。
- **昇腾310：高效能+灵活可编程**，在典型配置下可以输出16TOPS@INT8, 8TOPS@FP16，功耗仅为8W。采用自研华为达芬奇架构，集成丰富的计算单元, 提高AI计算完备度和效率，进而扩展该芯片的适用性。全AI业务流程加速,大幅提高AI全系统的性能，有效降低部署成本。
- **昇腾 910：业界算力最强的AI处理器**，基于自研华为达芬奇架构3D Cube技术，实现业界最佳AI性能与能效，架构灵活伸缩，支持云边端全栈全场景应用。算力方面，昇腾910完全达到设计规格，半精度（FP16）算力达到320 TFLOPS，整数精度（INT8）算力达到640 TOPS，功耗310W。

昇腾 310 关键特性

自研华为达芬奇架构NPU	在8W数据精度下算力可达16TOPS	高性能3D Cube
Architecture	• HUAWEI Da Vinci	
Computing Engine	• 3D Cube	
Performance	• 16 TOPS@INT8 and 8 TOPS@FP16	
Max Power	• 8W	
Process	• 12nm FFC	

昇腾 910 关键特性

自研华为达芬奇架构NPU	640 TOPS@INT8, 320TFLOPS@FP16	最大功耗310W
Architecture	• HUAWEI Da Vinci	
Computing Engine	• 3D Cube	
Performance	• 320 TFLOPS @FP16 and 640 TOPS @INT8	
Max Power	• 310W	
Process	• N7+	

## 1.9.2 AI产品交付-基础硬件层：昇腾推理服务器

- ◆ 华为昇腾系列AI处理器和基础软件构建Atlas人工智能计算解决方案，其中昇腾AI服务器具有超强计算性能，可广泛应用于中心侧AI推理、深度学习模型开发和训练场景，目前主要包括4个型号的产品：
  - **Atlas 800 推理服务器（型号：3000）：高算力+高能效。**基于昇腾310芯片的推理服务器，最大可支持8个Atlas 300I 推理卡，提供强大的实时推理能力，广泛应用于中心侧AI推理场景。
  - **Atlas 800 推理服务器（型号：3010）：灵活配置，适配多项负载。**基于Intel处理器的推理服务器，最多可支持7个Atlas 300I 推理加速卡，支持560路高清视频实时分析，广泛应用于中心侧AI推理场景。
  - **Atlas 800 训练服务器（型号：9000）：超强算力密度。**基于华为鲲鹏920+昇腾910处理器的AI训练服务器，具有超强算力密度、超高能效与高速网络带宽等特点。该服务器广泛应用于深度学习模型开发和训练，适用智慧城市、智慧医疗、天文探索、石油勘探等需要大算力的行业领域。
  - **Atlas 800 训练服务器（型号：9010）：超强算力密度。**基于Intel处理器+华为昇腾910芯片的AI训练服务器，具有超强算力密度、高速网络带宽等特点。该服务器广泛应用于深度学习模型开发和训练，适用于智慧城市、智慧医疗、天文探索、石油勘探等需要大算力的行业领域。

Atlas 800 推理服务器（型号：3000）



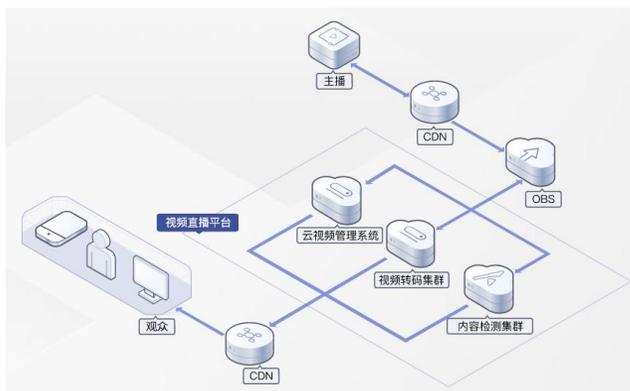
Atlas 800 推理服务器参数

Atlas 800 推理服务器型号	形态	CPU	CPU内存	AI加速卡/AI处理器	AI算力
3000	2U AI服务器	2 * 鲲鹏920	32个DDR4内存插槽，最高3200 MT/s	最大支持8个Atlas 300I 推理卡	最大704 TOPS INT8
3010	2U AI服务器	1/2个Intel® Xeon® SP Skylake 或 Cascade Lake处理器，最高205W	24个DDR4内存插槽，最高3200 MT/s	最大支持7个Atlas 300I 推理卡	最大616 TOPS INT8
9000	4U AI服务器	4 * 鲲鹏920	最多32个DDR4内存插槽，支持RDIMM 内存速率最高3200 MT/s 单根内存条容量支持16GB/32GB/64GB	8 * 昇腾910处理器	2.24 PFLOPS FP16 1.76 PFLOPS FP16
9010	4U AI服务器	2 * Intel V5 Cascaded Lake处理器	最多24个DDR4内存插槽，支持RDIMM	8 * 昇腾910处理器	2.24 PFLOPS FP16 1.76 PFLOPS FP16

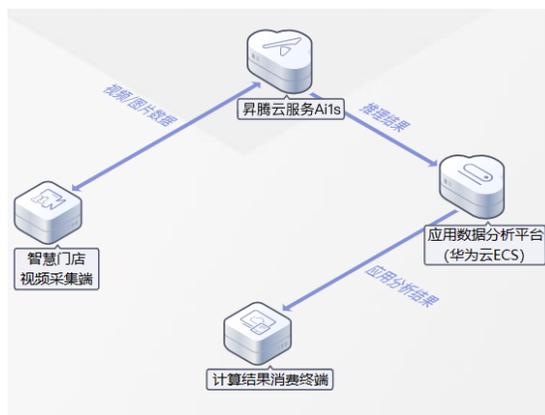
## 1.9.3 AI产品交付-基础硬件层：昇腾AI云服务器

- ◆ 昇腾AI云服务器 ( AI Accelerated Cloud Server ) 是一种可随时自助获取、可弹性伸缩、应用于AI训练/AI推理加速的云服务器，可为多种AI应用场景提供高效算力。具有多种应用场景，包括内容检测、智慧零售、视频转码、画质增强。服务器具有**高性能（推理提供最高128TFLOPS@FP16算力，性能领先业界主流GPU40%；训练提供最高2PFLOPS@FP16算力）**、简单易用、生态兼容、规格丰富、应用广泛、网络加速（高性能智能网卡，最大带宽25Gbps）的产品优势。
- ◆ **基于昇腾AI云服务器，微晟科技智慧门店客户案例：**根据华为云官网，微晟科技与华为云的合作，有效提高在人工智训练方面的效率，借助华为云技术体系，为行业输出更完备的高性能智能零售解决方案。实现推理成本下降30%；使用AI容器计算集群，性能提升30%；分布式AI训练集群，边云同态部署；兼容Tensorflow、Caffe等业界主流算子，降低智能防损系统迁移至昇腾云服务时间。

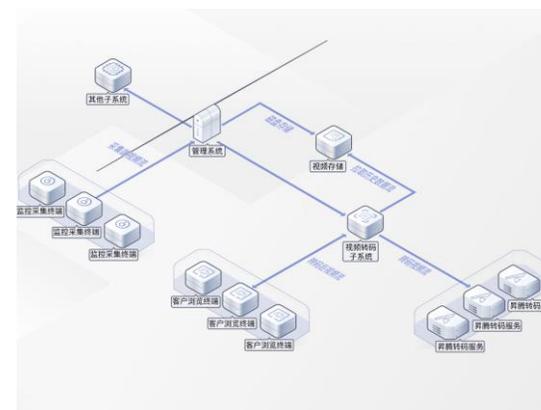
昇腾AI云服务器 内容检测应用场景



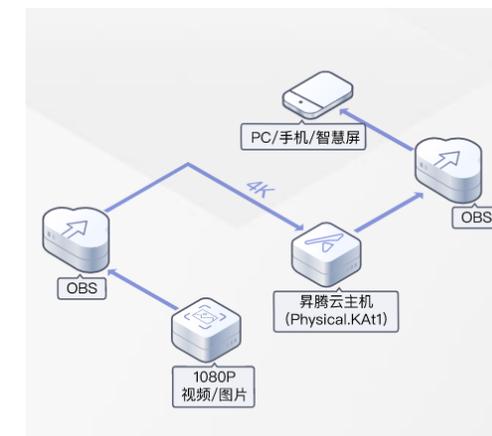
昇腾AI云服务器 智慧零售应用场景



昇腾AI云服务器 视频转码应用场景



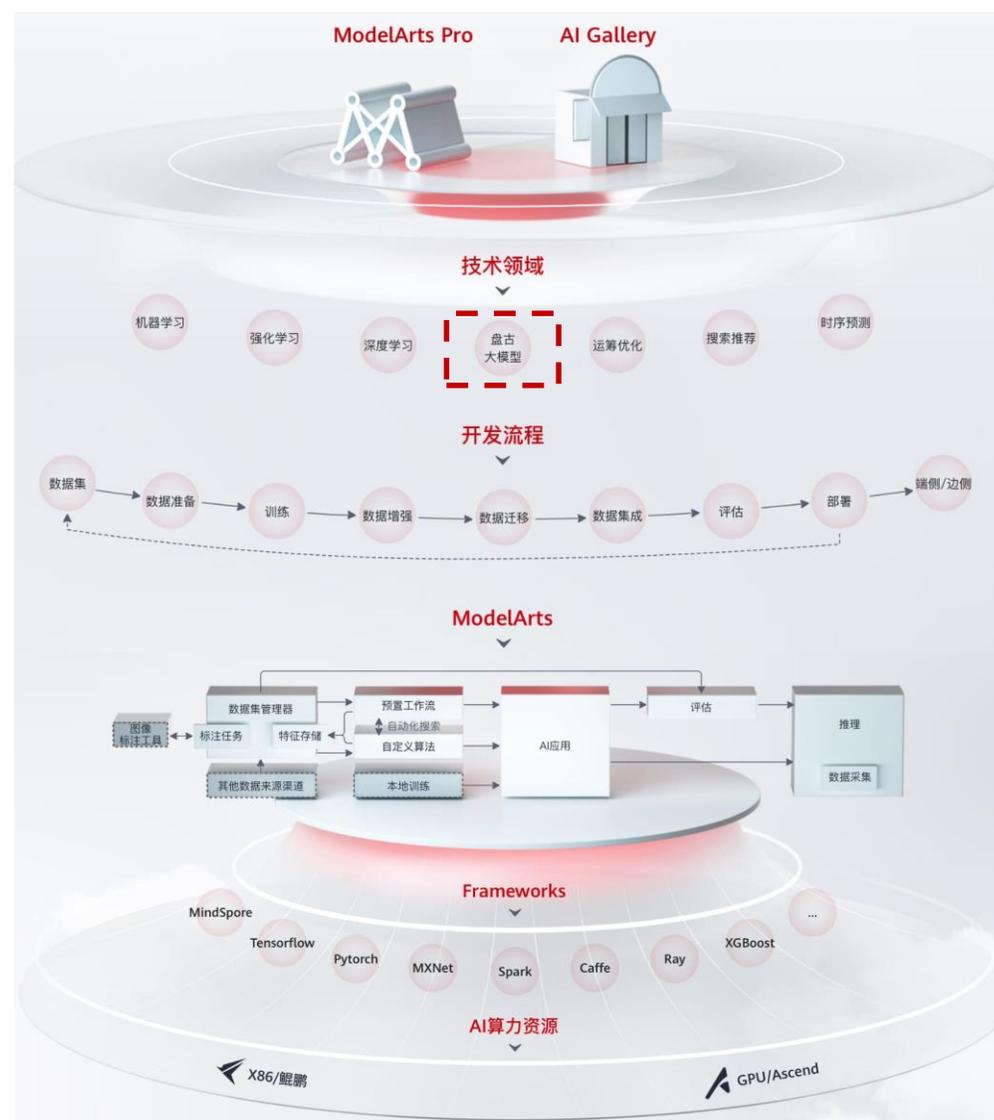
昇腾AI云服务器 画质增强应用场景



## 1.9.4 AI产品交付-开发服务层：华为云AI平台

- ◆ **华为云AI开发平台ModelArts：面向开发者的一站式 AI 平台。**为机器学习与深度学习提供海量数据预处理及交互式智能标注、大规模分布式训练、自动化模型生成，及端-边-云模型按需部署能力，帮助用户快速创建和部署模型，管理全周期 AI workflow。
- **盘古大模型背后的支撑：**华为ModelArts平台提供E级算力调度，同时结合物理网络拓扑，提供动态路由规划能力，为大模型训练提供了最优的网络通信能力。此外，借助ModelArts平台的高效处理海量数据能力，仅用7天就完成了40TB文本数据处理。
- ◆ **华为云D-Plan AI解决方案助力千行百业智能升级。**该解决方案为企业业务人员提供可视化的操作界面来扩展对机器学习的访问，只需简单配置就能自行生成准确的销量预测模型；而对于初阶的算法工程师，该方案还提供了对应的高阶功能，允许进行模型调参，从而达到更佳模型效果。
- **方案架构：**基于华为云ModelArts搭建一套端到端的销量预测 workflow，考虑企业级客户的工业化部署使用，支持定时调用功能与资源统一管理功能。
- **应用场景：**工业质检、零售-CV、出行-预测及调度、游戏行业、电商-推搜及预测、物流行业等。

ModelArts产品介绍



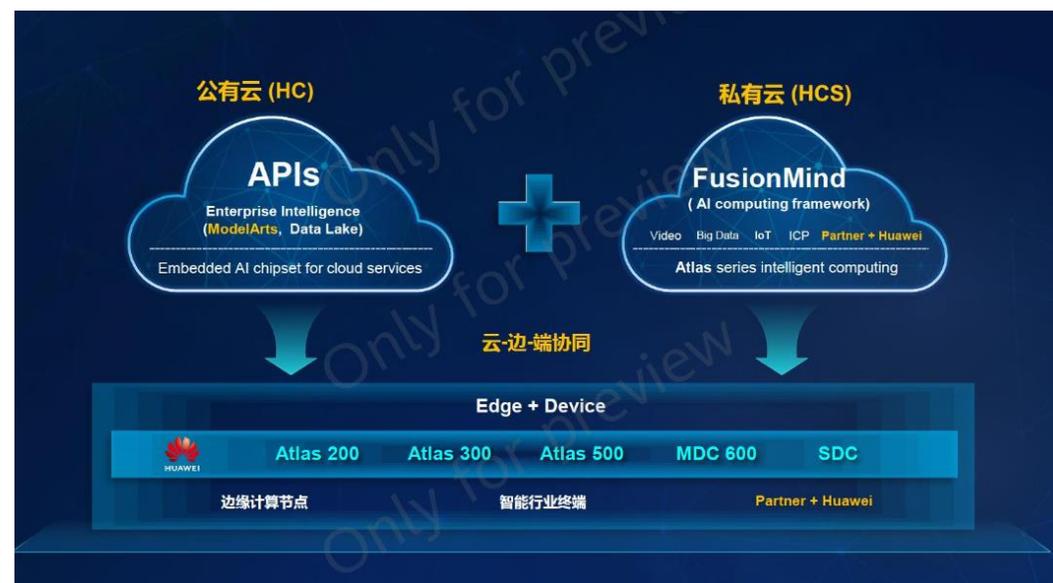
## 1.9.5 AI 产品交付: 全栈全场景AI解决方案

- ◆ **全栈全场景AI解决方案**：根据华为官网新闻，华为全联接大会2018上，发布了AI战略与全栈全场景AI解决方案，包括全球首个覆盖全场景AI昇腾系列芯片及产品云服务；芯片算子库和高度自动化算子开发工具CANN；支持端、边、云独立的和协同的统一训练和推理框架MindSpore；提供全流程服务（ModelArts），分层API和预集成方案的应用使能。其目标是实现“普惠AI”。2019年，凭借领先的全栈全场景AI能力，华为入选担纲建设基础软硬件国家新一代AI开放创新平台。
  - **全栈**：技术功能视角，是包括芯片、芯片使能、训练和推理框架和应用使能在内的全堆栈方案；
  - **全场景**：是包括公有云、私有云、各种边缘计算、物联网行业终端以及消费类终端等全场景的部署环境；
  - **支持HiAI+EI**：基于该解决方案，华为云EI（适用于行业）能为企业、政府提供全栈人工智能解决方案；HiAI（适用于消费者业务）能为智能终端提供全栈解决方案，且HiAI Service是基于华为云EI部署的。

华为全栈全场景AI解决方案



华为普惠AI方案





## **02 AI硬件自主可控势在必行**

## 2.1.1 国际局势波动较大，自主可控势在必行

- ◆ **自2018年来，国际局势波动较大，严重限制我国高科技领域发展。**
- ✓ 根据美国提出的《国家量子倡议法》（2018）、《美国人工智能发展倡议》（2019）以及《出口管制改革法案》（2018）等相关法案和计划，美国已对我国在14类新兴和基础技术领域，包括AI技术、人工智能芯片、机器人、量子计算、脑机接口和先进材料等方面实施出口和技术合作限制措施。
- ✓ 2022年8月，拜登正式签署《芯片与科学法案》，其中提到禁止接受联邦奖励资金的企业，在中国扩建或新建先进半导体的新产能；同年10月，美国政府进一步紧缩半导体产品对华出口的政策，主要包括限制英伟达、AMD等公司向中国出售高算力人工智能芯片；限制应用材料、泛林、科磊等美国设备厂商向任何中国公司出售半导体设备；将31家中国公司、研究机构及其他团体列入所谓“未经核实的名单”（UVL清单），限制它们获得某些受监管的美国半导体技术能力

美国制裁、限制事件汇总

时间	事件
2018/11/1	美国商务部发布涉及人工智能和机器学习技术、先进计算技术、数据分析技术等14项新兴和前沿技术的对华出口管制框架
2019/5/1	“布拉格5G安全大会”召开：联合发布了“布拉格提案”，该提案从政策、安全、技术、经济四个方面探讨如何排除中国5G技术产品。
2020/1/1	特朗普政府发布限制人工智能软件出口新规，应用于智能化传感器、无人机和卫星的目标识别软件都在限制范围之内。
2020/2/1	推动42个加入《瓦森纳协定》的国家扩大半导体对华出口管制范围，旨在加强防备相关技术外流到中国。
2020/2/1	美国商务部更新《出口管制条例》，将“用于自动分析地理空间图像的软件”列入对华管制清单中，应用于智能化传感器、无人机、卫星和其他自动化设备的目标识别软件。
2020/5/1	美国宣布将加入七国集团“人工智能全球合作伙伴组织”，力图以霸权力量主导构成不利于中国的全球人工智能管理规则，限制中国人工智能技术发展。
2020/5/1	发起七国集团（G7）加澳大利亚、韩国和印度的“D10俱乐部”（D10 Club），以减少对中国电信技术的依赖。
2020/10/1	美国家人工智能安全委员会提出通过多边合作、数字联盟等形式与北约、印度等建立国际联盟，推广美国标准和规则，形成对我人工智能的封锁围堵之势。
2022/7/1	美国半导体设备制造商收到美商务部的通知，拟要求禁止向中国大陆供应用于14nm或以下芯片制造的设备。
2022/8/1	美国总统拜登正式签署《芯片与科学法案》，以补贴美国的半导体产业。关于补贴资助对象资格的内容里，明确写到，禁止接受联邦奖励资金的企业，在中国等对美国国家安全构成威胁的特定国家扩建或新建某些先进半导体的新产能，期限为10年，违反禁令或未能修正违规状况的公司，可能需要全额退还联邦补助款。
2022/9/1	美国两大芯片制造巨头英伟达（NVIDIA）与AMD同时发布公告，声称均已接到美国拜登政府下达的最新命令，要求停止向中国出口用于人工智能的最先进芯片。制裁主要针对的两个芯片是Nvidia A100和H100图形处理单元以及AMD的MI250人工智能芯片。
2022/10/1	BIS修订《出口管理条例》：美国从多方面加强对出口到中国的半导体的管制措施。新的管控措施主要涉和先进计算及半导体制造业以及超级计算机和半导体最终用途。

## 2.1.2 美国限制高端芯片流入中国，严重干扰国内大模型发展生态

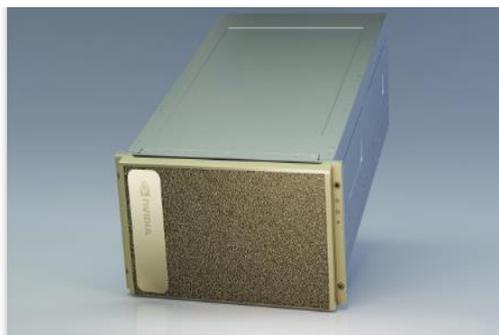
- ◆ **美国政府禁止英伟达、AMD向中国出口用于人工智能的顶级计算芯片。**
- ✓ 根据钛媒体，2022年9月，美国商务部宣布限制英伟达（NVIDIA）和AMD等美国公司向中国出口先进计算机图像处理器（GPU），该禁令主要限制了英伟达的A100、H100高端芯片以及AMD的MI250出口中国，目的是瞄准国内先进计算进行遏制，影响国内人工智能领域发展。
- ✓ **浪潮集团被加入“实体清单”**：根据钛媒体，2023年3月美国商务部发布浪潮被加入被加入美国“实体清单”，限制了美国科技公司对浪潮的技术、产品支持。浪潮的服务器业务在CPU、GPU等关键芯片技术商高度依赖外国厂商，此外，截至2022年末，浪潮服务器及部件占总营收99.17%，若此次制裁被严格落实，其服务器业务将严重停滞。

英伟达H100、A100芯片



DGX H100

AI supercomputer optimized for large generative AI and other transformer-based workloads.



DGX A100

AI supercomputer delivering world-class performance for mainstream AI workloads.

浪潮AI服务器



## 2.1.3 英伟达应对制裁，提出特供版A800芯片

- ◆ **A800是A100的下位替代版。**
  - ✓ 根据快科技，在美国限制英伟达向中国出售高算力芯片A100、H100后，英伟达发布公告，确认发布新款中国特供版A800 GPU芯片来替换A100，以满足制裁政策。A800完全符合美国政府有关出口管制的测试。
- ◆ **A800在带宽性能方面劣于A100。**
  - ✓ A800芯片的数据传输速率为400GB/s，低于A100芯片的600GB/s，而其他参数变化不大。这也说明A800相比于A100在整体通信带宽性能上低了33%左右，影响了多卡互联性能，但是单卡性能保持不变。在一定程度上，这种削弱会导致在AI大模型训练上消耗更长的时间。
- ◆ **英伟达表示会继续推出H100的替代版H800。**
  - ✓ 根据快科技，2023年3月21日，英伟达在GTC 2023春季图形大会上，NVIDIA近日宣布为中国市场开发了第二个特供版H800，该产品是在已有的H100基础上进行了调整，以符合美国政府的规定。

A100、A800对比

NVIDIA A100 TENSOR CORE GPU SPECIFICATIONS  
(SXM4 AND PCIE FORM FACTORS)

	A100 80GB PCIe	A100 80GB SXM
FP64	9.7 TFLOPS	
FP64 Tensor Core	19.5 TFLOPS	
FP32	19.5 TFLOPS	
Tensor Float 32 (TF32)	156 TFLOPS   312 TFLOPS*	
BFLOAT16 Tensor Core	312 TFLOPS   624 TFLOPS*	
FP16 Tensor Core	312 TFLOPS   624 TFLOPS*	
INT8 Tensor Core	624 TOPS   1248 TOPS*	
GPU Memory	80GB HBM2e	80GB HBM2e
GPU Memory Bandwidth	1,935GB/s	2,039GB/s
Max Thermal Design Power (TDP)	300W	400W***
Multi-Instance GPU	Up to 7 MIGs @ 10GB	Up to 7 MIGs @ 10GB
Form Factor	PCIe dual-slot air cooled or single-slot liquid cooled	SXM
Interconnect	NVIDIA® NVLink® Bridge for 2 GPUs: 600GB/s ** PCIe Gen4: 64GB/s	NVLink: 600GB/s PCIe Gen4: 64GB/s
Server Options	Partner and NVIDIA-Certified Systems™ with 1-8 GPUs	NVIDIA HGX™ A100-Partner and NVIDIA-Certified Systems with 4, 8, or 16 GPUs NVIDIA DGX™ A100 with 8 GPUs

NVIDIA A800 TENSOR CORE GPU SPECIFICATIONS  
(SXM4 AND PCIE FORM FACTORS)

	A800 40GB PCIe	A800 80GB PCIe	A800 80GB SXM
FP64	9.7 TFLOPS		
FP64 Tensor Core	19.5 TFLOPS		
FP32	19.5 TFLOPS		
Tensor Float 32 (TF32)	156 TFLOPS   312 TFLOPS*		
BFLOAT16 Tensor Core	312 TFLOPS   624 TFLOPS*		
FP16 Tensor Core	312 TFLOPS   624 TFLOPS*		
INT8 Tensor Core	624 TOPS   1248 TOPS*		
GPU Memory	40GB HBM2	80GB HBM2e	80GB HBM2e
GPU Memory Bandwidth	1,555GB/s	1,935GB/s	2,039GB/s
Max Thermal Design Power (TDP)	250W	300W	400W***
Multi-Instance GPU	Up to 7 MIGs @ 5GB	Up to 7 MIGs @ 10GB	Up to 7 MIGs @ 10GB
Form Factor	PCIe (dual-slot air cooled or single-slot liquid cooled)		SXM
Interconnect	NVIDIA® NVLink® Bridge for 2 GPUs: 400GB/s ** PCIe Gen4: 64GB/s		NVLink: 400GB/s PCIe Gen4: 64GB/s
Server Options	Partner and NVIDIA-Certified Systems™ with 1-8 GPUs		NVIDIA HGX™ A800-Partner and NVIDIA-Certified Systems with 4 or 8 GPUs

## 2.2.1 政策端持续发力，加速推动国产自主可控进程

### ◆ 政府聚焦人工智能产业，发布多条政策助力AI发展。

- ✓ 2023年5月30日，北京市发布《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025年）》和《北京市促进通用人工智能创新发展的若干措施》。
- ✓ **工作方向主要瞄准**：突破基础理论，引领关键核心技术创新，强化可信人工智能技术；推动国产AI芯片突破，研发通用高算力训练芯片、低功耗边缘端芯片和创新架构；加强自主开源深度学习框架研发，实现软硬件深度协同；提升算力供给能力，建设公共算力中心，实施算力伙伴计划；加强公共数据开放共享，推动数据融合创新；构建高效协同的大模型技术产业生态，加强人工智能企业梯度培育，强化企业多维服务。构建人工智能生态系统。

北京市人民政府官网最新政策界面



北京政府关于AI政策概述

时间	部门	政策	主要内容
2023. 5. 30	北京市人民政府	《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025年）》	1. 夯实人工智能底层基础，聚焦突破人工智能关键技术。2. 积极引导国产大模型研发应用国产人工智能芯片，加速提升算力供给的国产化率。3. 深化国产芯片部署应用，推动自主可控软硬件算力生态建设
2023. 5. 30	北京市人民政府	《北京市促进通用人工智能创新发展的若干措施》	1. 推动算力基础建设，针对算力需求，打造多云算力调度平台。2. 优化相关数据要素质量，助力大模型训练。3. 构筑大模型等通用人工智能技术体系，推动下游AI创新场景应用。

## 2.2.1 政策端持续发力，加速推动国产自主可控进程

- ◆ **加快推动人工智能高质量发展、创新应用场景。**
- ◆ 2023年5月31日，深圳市政府发布《深圳市加快推动人工智能高质量发展高水平应用行动方案（2023—2024年）》。
- ✓ **工作方向主要瞄准：**建设城市级智能算力平台，打造大湾区智能算力枢纽，建设企业级智能算力平台；加强科技研发攻关，支持创新产品研发；规划建设产业集聚区，大力培育企业梯队，依托鹏城云脑搭建城市级人工智能生态孵化平台，为中小企业提供低成本智能算力资源；推进人工智能产业发展和应用，包括搭建供需对接平台、推进公共服务和城市治理的人工智能应用，培育企业梯队，建设产业集聚区，以及推动各行业的人工智能应用和创新发展；优化数据提供和高技术人员构成；加强组织领导，并成立对应的工作专班。
- ◆ **北深两地政策发布有望加速我国人工智能发展，推动算力供给环节国产化替代进程。**
- ✓ 1) 算力是AI发展的基础，决定处理数据的能力和模型的性能，但是在多个环节仍然受限于海外制裁、存在技术“卡脖子”等问题。2) 关键技术自主可控仍然是国内发展AI的重中之重，在当前国际形势下，只有大力发展自主可控才可以保障数据安全和占据主动权。

深圳市人民政府官网最新政策界面



5月31日，深圳正式印发《深圳市加快推动人工智能高质量发展高水平应用行动方案（2023—2024年）》（以下简称《行动方案》），同步发布首批“城市+AI”应用场景清单，统筹设立规模1000亿元的人工智能基金群，以最充足的算力、最大的政策支持、最优的产业生态、最好的人才环境、最丰富的场景应用，积极打造国家新一代人工智能创新发展试验

深圳政府关于AI政策概述

时间	部门	政策	内容
2023年5月30日	中共深圳市委办公厅、深圳市人民政府办公厅	《深圳市加快推动人工智能高质量发展高水平应用行动方案（2023—2024年）》	1. 加强智能算力集群供给。2. 聚焦核心技术和产品创新能力。3. 增强产业聚集。4. 优化下游应用。5. 培育高质量数据要素市场，汇集高水平研发人员。

## 2.2.1 政策端持续发力，加速推动国产自主可控进程

- ◆ **针对算力产业，我国激励政策陆续出台。**
  - ✓ 当前我国已进入《新型数据中心发展三年行动计划（2021-2023年）》落地见效的关键年。《行动计划》主要目标为用3年时间，基本形成布局合理、技术先进、绿色低碳、算力规模与数字经济增长相适应的新型数据中心发展格局。到2023年底，全国数据中心机架规模年均增速保持在20%左右，平均利用率力争提升到60%以上，总算力超过200 EFLOPS，高性能算力占比达到10%。
- ◆ **各地全力保障数字基础设施建设，积极带动关联产业集聚发展。**
  - ✓ 在第七届世界智能大会上，中国电子董事长曾毅表示没有强大的算力，新一代人工智能将是无本之木。此外，5月12日，《北京市促进通用人工智能创新发展的若干措施（2023-2025年）（征求意见稿）》，在“加强算力资源统筹供给能力”等5个方面提出21项具体措施推动人工智能创新落地。同时，在5月19日，北京市启动通用人工智能产业创新伙伴计划推动大模型产业加速落地，该计划提出八大任务作为支撑，分别为加快满足近期迫切算力需求、提升中长期算力供给能力、推出一批高质量训练数据、谋划建设国家级数据训练基地、大模型应用创新标杆试点工程、推动大模型赋能千行百业等。

工信部印发相关计划



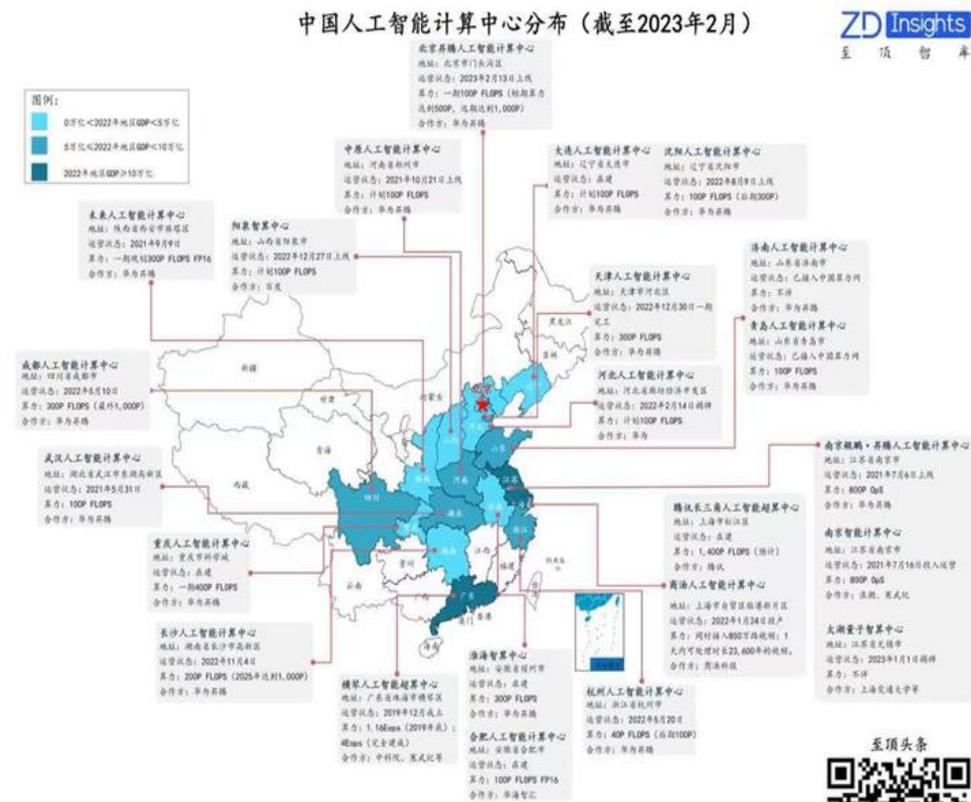
北京市通用人工智能产业创新伙伴计划成员名单（第一批）



## 2.2.2 产业端积极响应，智能算力建设持续提速

- ◆ **北京昇腾人工智能计算中心正式点亮**：北京昇腾人工智能计算中心正式点亮，将推动北京人工智能产业高质量发展。该智能计算中心采用昇腾AI基础软硬件，充分释放硬件算力，加速人工智能企业创新应用和模型孵化。
- ◆ **贵州省大数据局印发《面向全国的算力保障基地建设规划》**：总体目标是到2025年，面向全国的算力保障基地建设任务全面完成，贵州超大规模数据中心集群的地位更加巩固，存算比更加合理，优化基础设施布局、结构、功能和系统集成，数据中心实现集约化、规模化、绿色化发展，网络互联互通、能源安全可靠提高到新的水平，打造具有国际竞争力的数字产业集群。
- ◆ **上海市经济信息化委印发《上海市推进算力资源统一调度指导意见》**：主要目标为到2023年底，依托本市人工智能公共算力服务平台，接入并调度4个以上算力基础设施，可调度智能算力达到1,000 PFLOPS (FP16) 以上；到2025年，市人工智能公共算力服务平台能级跃升，完善算力交易机制，实现跨地域算力智能调度，通过高效算力调度，推动算力供需均衡，带动产业发展作用显著增强。
- ◆ **惠州首个超大规模数据及算力中心力争年内投产**：2023年年初，作为大数据及关联产业发展的重要支撑点的粤港澳大湾区（惠州）数据产业园建设取得明显成效。落户该园区的润泽（惠州）国际信息港一期项目试运行工作进展顺利，预计年内正式投产。其目标是构建具有国际领先技术水平的算力基础设施，带动数据服务及硬件研发制造等关联产业集聚发展。

中国人工智能计算中心分布图（截至2023年2月）



## 2.2.2 产业端积极响应，智能算力建设持续提速

- ◆ **山东首个人工智能计算中心上线运行，竞逐人工智能赛道**：2023年3月17日青岛市人工智能产业园正式开园，同步上线的青岛人工智能计算中心，成为山东首个上线运行的人工智能计算中心。中心首期具备100P算力，相当于5万台高性能PC的算力，将面向青岛乃至胶东地区的企业、高校和科研机构提供普惠公共算力服务。
- ◆ **河南省数字化转型战略工作方案出炉，推进郑州、洛阳构建超大型绿色数据中心集群**：2023年3月30日河南省制造强省建设领导小组办公室印发《2023年河南省数字化转型战略工作方案》，目标今年电子信息制造业营业收入力争突破8000亿元，先进计算、软件产业规模均超过500亿元。
- ◆ **天津市人工智能计算中心揭牌，加快打造天津数字经济发展新动能**：2023年3月18日，天津市人工智能计算中心正式揭牌上线，助力人工智能产业创新发展。人工智能中心不仅提供基础算力服务，还提供应用创新服务、产业孵化服务等，把算力、算法、数据、应用场景和人才进行5要素的聚集，帮助企业在人工智能科研创新上降本增效。

2023年河南省数字化转型战略目标任务分解

序号	城市	智能工厂/智能车间	贯标升级版/对标升级版(家)	数字化转型项目(个)	企业上云
1	郑州市	22	36/360	116	10740
2	开封市	7	15/150	48	1120
3	洛阳市	12	24/240	80	2080
4	平顶山市	7	15/150	52	1540
5	安阳市	7	15/150	52	460
6	鹤壁市	6	12/120	40	290
7	新乡市	12	21/210	72	1840
8	焦作市	8	18/180	60	440
9	濮阳市	5	12/120	40	570
10	许昌市	9	18/180	60	1610
11	漯河市	6	12/120	40	730
12	三门峡市	5	12/120	40	350
13	南阳市	12	21/210	72	2730
14	商丘市	9	18/180	56	1710
15	信阳市	4	15/150	52	1510
16	周口市	8	15/150	52	1360
17	驻马店市	7	15/150	48	770
18	济源示范区	4	6/150	20	150
合计		150	300/3000	1000	30000

天津市人工智能计算中心内的算力服务器



## 2.3.1 重申强调，算力在大模型的背景下势必迎来爆发

- ◆ **ChatGPT用户数量暴增，同样侧面证明了AI产业革命下，对于算力基础的高度需求。**根据SimilarWeb的数据，2023年2月，ChatGPT访问数量为10亿次/每月，而2023年4月，ChatGPT的访问量增加至17.56亿次/每月。
- ◆ **根据我们的测算，目前ChatsGPT产品运营需英伟达A100 GPU约71296片，预计投入算力成本达17.73亿美元。**据SimilarWeb数据，2023年5月（至5月24日）ChatGPT官网（chat.openai.com）总访问量为14.08亿次。据环球零碳研究中心数据，每次用户与ChatGPT互动，ChatGPT的每个响应词在A100 GPU上需要350毫秒。英伟达DGXA100服务器单机售价约为19.9万美元/台，每台大约可搭载8片A100 GPU。根据测算结果，从今年2月至今，对英伟达A100 GPU需求持续增长，我们预计后续该趋势有望保持。
- ◆ **短期算卡为王，长期自主可控：**重申强调，算力在大模型的背景下势必迎来爆发，而算卡作为算力的心脏其重要性不言而喻。**长期来看，随着国际形势不断变化，我国高科技及AI的科技发展可能受阻，因此发展自主可控算力芯片势在必行。**

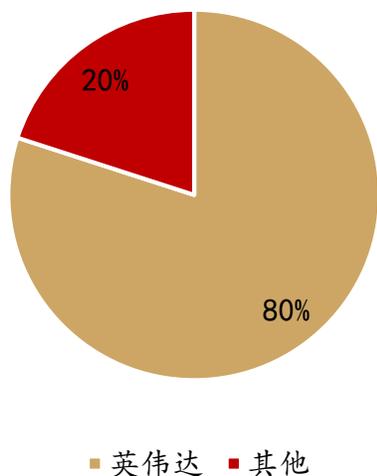
ChatGPT运行的算力需求及成本测算

	单位	2023年2月	2023年3月	2023年4月	2023年5月 (至5.24)
ChatGPT访问量	亿次/月	10.00	15.59	17.56	14.08
月均天数	天	28	31	30	24
日均访问量	百万次/天	35.71	50.29	58.53	58.67
咨询量	字/次/天	300	300	300	300
总咨询量	亿字/天	107.14	150.87	175.60	176.00
A100 GPU算力耗量	ms/字	350	350	350	350
<b>A100需求量</b>	<b>片/天</b>	<b>43,403</b>	<b>61,117</b>	<b>71,134</b>	<b>71,296</b>
A100 GPU售价	万美元/片	1.5	1.5	1.5	1.5
DGXA100服务器搭载A100数	片/台	8	8	8	8
DGXA100服务器需求量	台	5,425	7,640	8,892	8,912
DGXA100系统售价	万美元/台	19.9	19.9	19.9	19.9
<b>初始算力成本</b>	<b>亿美元</b>	<b>10.80</b>	<b>15.20</b>	<b>17.69</b>	<b>17.73</b>

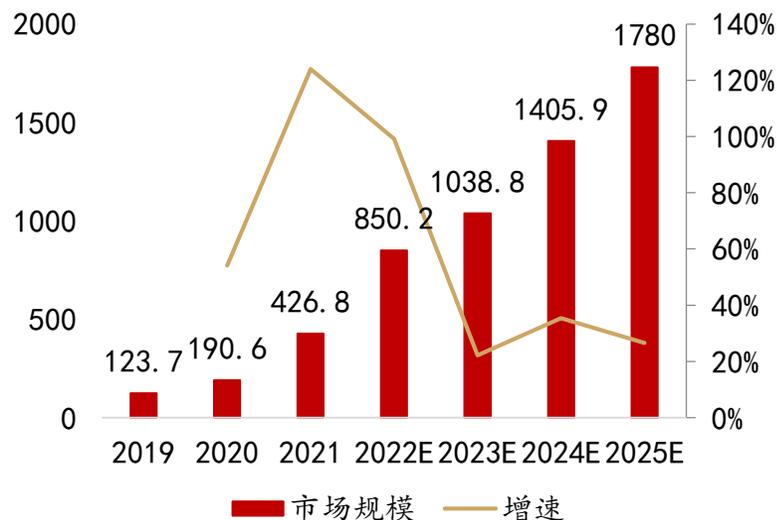
## 2.3.2 国产AI算力芯片自主可控势在必行

- ◆ **我国AI芯片方面仍处于“垄断局面”，高端AI芯片仍需自主可控，我国相关企业已初具竞争实力：**根据IDC数据，2021年，中国加速卡数量出货超过80万片，其中Nvidia占据超过80%市场份额。此外还包括AMD、百度、寒武纪、燧原科技、新华三、华为、Intel和赛灵思等。
- ◆ **人工智能逐渐成为主流的发展趋势，中国人工智能市场投资规模呈上升趋势。**在中国市场，IDC预测，2026年中国人工智能投资有望达到266.9亿美元，约占全球投资的8.9%，在其他国家中排名世界第二。
- ◆ **AI算力规模的快速增长将催生更大的AI芯片需求：**根据亿欧智库的数据，预计2023年中国AI芯片市场规模将达到1039亿元，2025年中国AI芯片市场规模将达到1780亿元，三年GAGR为19.66%。

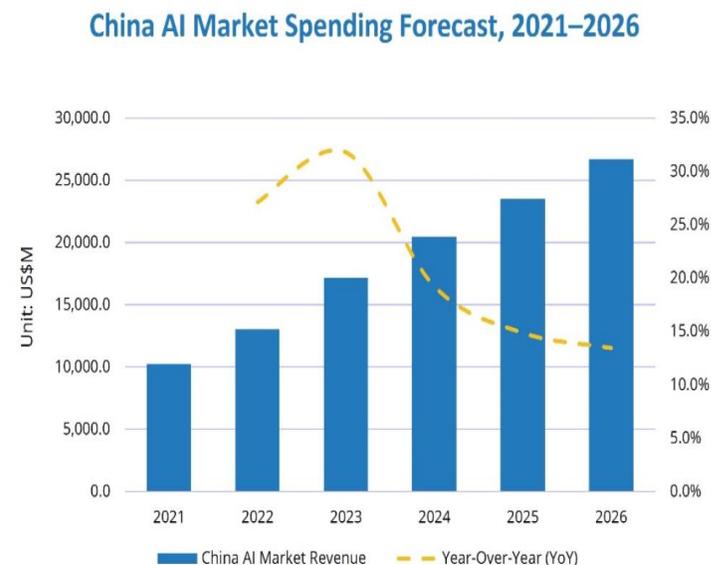
2021年我国服务器加速卡市场份额



中国AI芯片市场规模及其预测（亿元）



中国人工智能市场投资情况及其预测



## 2.3.2 国产AI算力芯片自主可控势在必行

- ◆ **发展国产AI芯片势在必行**：我国AI芯片已经呈现百舸争流的情况，国产化AI芯片势在必行，相关厂商积极加速推进AI芯片布局，促进AI芯片市场发展。

中国AI芯片主要厂商及其产品情况

公司	产品型号	应用（训练或推理）	算力	频率	功耗	制程
寒武纪	思元370 思元290 思元270	训练+推理 训练 推理	256TOPS (INT8) 512TOPS (INT8) 128TOPS (INT8)	/	/	7nm
燧原科技	T20 T21 i20	训练 训练 推理	256TOPS (INT8) 256TOPS (INT8) 256TOPS (INT8)	1.5GHz	300w 300w 150w	/
昆仑芯	昆仑芯2代AI芯片	训练+推理	256TOPS (INT8)	/	/	7nm
平头哥	含光800	推理	820TOPS	/	/	12nm
沐曦	MAN100	训练+推理	/	/	/	7nm
华为海思	HUAWEI Ascend310 HUAWEI Ascend910	边缘计算AI	16TOPS (INT8) 640TOPS (INT8)	/	8w 310w	12nm N7+
紫光展锐	SC9863A	边缘计算AI	/	1.6GHz	/	/
后摩智能	™H30	感存算AI	156TPOS	/	35w	/
云天励飞	DeepEye 2000	安防/人脸识别AI	/	/	/	22nm
地平线	征程®5	自动驾驶AI	128TPOS	/	/	16nm
景嘉微	JM9	图形处理	1.5TFLOPS (FP32)	1.5GHz	/	14nm
龙芯中科	2K1000LA	可信+边缘计算AI	/	/	5w	/
海光信息	深算一号	计算AI	/	2GHz	350w	7nm
清微智能	TX8	训练+推理	/	/	/	/

## 2.3.3 国产加速计算服务器时代到来

- ◆ **人工智能应用场景下的加速计算服务器是中国服务器的核心驱动力:** AI服务器作为AI芯片的载体景气度上行，大模型的出现带动AI服务器呈现加速状态，根据IDC的数据，在2021年的统计，预计到2025年中国加速服务器市场规模将达到108.6亿美元，且2023年仍处于中高速增长期，增长率约为20%。
- ◆ **AI服务器作为算力载体为数字经济时代提供广阔动力源泉:** AI服务器更专精于海量数据处理和运算方面，我们认为其可以为人工智能、深度学习、神经网络、大模型等场景提供广阔的动力源泉，并广泛应用于医学、材料、金融、科技等千行百业。

2021-2025年中国服务器市场规模及增速(亿美元)



中国半年度加速计算市场预测，2021-2025



资料来源: IDC, 各公司官网, 华西证券研究所

公司名称	名称/型号	处理器	内存支持	AI加速卡/AI处理器	AI算力
浪潮科技	NF5468M6	2颗第三代Intel® Xeon®可扩展处理器(Ice Lake), TDP 270W, 支持3条UPI互联	支持32条DDR4 RDIMM/LRDIMM内存, 速率最高支持3200MT/s	/	/
神州数码	KunTai A222	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	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	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
拓维信息	兆瀚RA2300-A	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	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	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	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
龙芯中科	KU 2208-L2	支持2颗国产龙芯3B4000处理器; 8核心 主频1.8GHz-2.0GHz, 单颗功耗达50W	板载8个DDR4 DIMM扩展插槽, 支持DDR4 ECC RDIMM; 内存单根内存容量: 8GB、16GB, 内存可达: 128GB	/	/
	KU 2208-L3	支持2颗国产龙芯3C5000L处理器, 高达32核心; 主频2.2GHz, 功耗高达130W	板载8个DDR4 DIMM插槽, 支持双通道DDR4 ECC RDIMM, 支持内存单根内存容量: 8GB、16GB、32GB、64GB, 支持至: 512GB内存	/	/

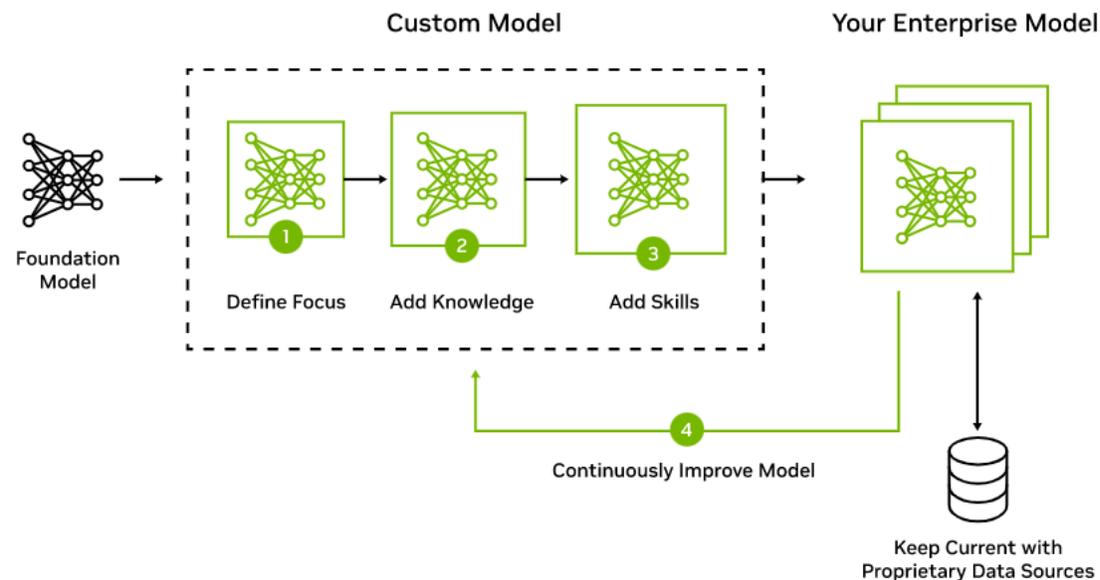
## 2.3.4 “共享AI能力与算力”，AI云需求高增

- ◆ **部署生成式 AI 应用难度较高，AI云提供平台定制化能力。** 随着大模型带来的人工智能产业崛起，AI应用如文本生成、自动客服、自动驾驶等领域快速扩张。对于大多数企业自己部署这样的能力是非常困难的，我们认为一是由于目前的算力缺口及训练的边际成本较高，二是由于全链部署应用需要深厚的软硬件结合生态技术。而AI云可将如英伟达等专业供应商的AI能力整合到云上，让企业能够直接接入应用或从基础层进行预训练，进而形成自己的模型和应用。
- ◆ **AI云需求快速增长，云算力革命开启。** 在企业对大模型训练、人工智能应用部署等AI能力需求持续上行的态势下，AI云产品受到市场的青睐。以阿里、腾讯为代表的平台型公司在云端市场布局上，更多地关注的是通用云的打造；而华为、曙光更多地是从硬件的角度着手加入云市场布局。我们认为软硬结合的AI云玩家（英伟达、首都在线）符合市场的需求。同时，AI云需求高增也意味着未来掌握智算卡的企业将继续占领市场高地，万变不离其宗，我们坚持认为拥有算力的企业在下一阶段进行应用和平台化竞争的过程中具有先发优势。

首都在线云游戏解决方案



英伟达NeMo云服务





## **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

- ◆ 我们认为大模型有望赋能千行百业，算力作为“底层燃料”其重要性不言而喻，AI算力，自立自强，AI+信创势在必行，**我们判断华为正依托昇腾AI开启新一轮成长周期，相关产业链有望深度受益**，此外拥有算卡的厂商有望开启新一轮成长曲线，而长期自主可控为大势所趋，积极的推荐：
- ◆ 1) **华为AI服务器厂商**，相关受益标的为：**拓维信息、神州数码、广电运通、四川长虹、东华软件、同方股份**等；
- ◆ 2) **国产AI芯片厂商**，相关受益标的为：**寒武纪、海光信息、景嘉微、龙芯中科**等；
- ◆ 3) **AI服务器厂商**，相关受益标的为：**中科曙光、工业富联、浪潮信息**等；
- ◆ 4) **AI云厂商**，相关受益标的为：**首都在线、鸿博股份、青云科技、优刻得、光环新网、新炬网络**等。

## 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)		
		2023/6/13	2023/6/13	2022	2023E	2024E	2022	2023E	2024E
赛武纪*	688256.SH	232.90	970.25	-3.14	-1.95	-1.34	-	-	-
海光信息	688041.SH	80.78	1877.60	0.38	0.64	0.87	212.6	126.2	92.9
景嘉微*	300474.SZ	95.51	434.74	0.64	0.90	1.26	149.2	106.3	75.6
龙芯中科*	688047.SH	142.88	572.95	0.14	0.52	0.98	1020.6	277.0	145.5
中科曙光	603019.SH	53.78	787.35	1.06	1.38	1.81	50.7	39.1	29.7
神州数码*	000034.SZ	29.19	195.35	1.57	1.84	2.20	18.6	15.8	13.3
拓维信息*	002261.SZ	19.09	239.83	-0.82	0.10	0.13	-	198.2	142.8
工业富联*	601138.SH	22.02	4373.81	1.02	1.20	1.35	21.6	18.4	16.4
首都在线	300846.SZ	19.23	89.77	-0.41	0.16	0.37	-	121.3	51.6
鸿博股份*	002229.SZ	37.63	187.53	-0.15	-0.03	0.12	-	-	302.5
光环新网*	300383.SZ	12.39	222.72	-0.49	0.36	0.45	-	34.0	27.3
新炬网络*	605398.SH	39.78	33.14	0.68	0.85	1.03	58.5	46.7	38.5
优刻得*	688158.SH	23.20	105.12	-0.92	-0.53	-0.33	-	-	-
广电运通*	002152.SZ	12.53	311.17	0.33	0.40	0.45	38.0	31.4	27.9
四川长虹*	600839.SH	4.92	227.12	0.10	0.15	0.19	48.5	33.7	26.0
东华软件*	002065.SZ	8.06	258.36	0.13	0.21	0.30	64.4	37.6	27.0
同方股份*	600100.SH	9.18	307.56	-0.23	-	-	-	-	-

注: \*均来自wind一致预测

资料来源: WIND, 华西证券研究所

## 3.2.1 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

## 3.2.2 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，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 广电运通: 华为生态重要参与者与共建者

- ◆ **广电运通科技赋能传统产业升级:** 公司是领先的人工智能行业应用企业，业务覆盖智能金融、公共安全、智能交通、数字政府、大文旅、新零售及智慧教育等领域。公司以科技赋能传统产业升级，聚焦金融科技和城市智能，布局算法、算力、数据、场景四大人工智能要素。同时，以aiCore System大数据底座平台为核心，以“智能终端+大数据”的路径实现价值闭环，构筑万物互联的数字经济生态
- ◆ **广电运通为华为生态重要参与者与共建者:** 公司与华为联合攻关研发基于鲲鹏生态开发资源的“端-边-云”系列产品：针对华为鲲鹏系列处理器指令集和微架构的特点优化产品技术架构，围绕RISC通用架构处理器+AI芯片（NPU、TPU）的异构计算平台性能最大化目标设计端-边-云的协同算法和计算策略。2022年7月广电五舟成为“华为昇腾万里伙伴”，未来联合打造有竞争力的人工智能解决方案。

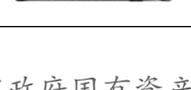
广电运通服务器鲲鹏系列

名称	示意图	性能和扩展特点	可用性和可服务性特点	可管理性和安全性特点	能源效率
GRG200-RK2280		支持华为64 bits高性能多核Kunpeng 920处理器 最大可支持64cores, 2.6GHz, 可支持多种核数量和频率的型号搭配	单板硬件采用电信级器件和加工工艺流程 支持SAS/SATA/NVMe硬盘 BMC集成管理模块（iBMC） 能够持续监控系统参数	集成在服务器上的iBMC管理模块可用来监控系统运行状态，提供远程管理功能 集成了业界标准的统一可扩展固件接口（UEFI）	提供白金电源模块，50%负载下电源模块效率高达94% 高效率的单板VRD电源 支持主备供电
GRG200-RK2480		支持华为64 bits高性能多核Kunpeng 920处理器 最大可支持256 cores（四路），2.6GHz，单台服务器支持4个处理器、最大256个内核	单板硬件采用电信级器件和加工工艺流程 支持SAS/SATA/NVMe硬盘 BMC集成管理模块（iBMC） 能够持续监控系统参数	集成在服务器上的iBMC管理模块可用来监控系统运行状态 集成了业界标准的统一可扩展固件接口（UEFI） 支持带锁的服务器机箱安全面板	提供白金电源模块，50%负载下电源模块效率高达94% 高效率的单板VRD电源 支持主备供电
GRG200-RK2180		支持华为64 bits高性能多核Kunpeng 920处理器 最大可支持64cores, 2.6GHz，单台服务器支持1个处理器、最大64个内核	单板硬件采用电信级器件和加工工艺流程 支持SAS/SATA/NVMe硬盘 BMC集成管理模块（iBMC） 能够持续监控系统参数	集成在服务器上的iBMC管理模块可用来监控系统运行状态，提供远程管理功能 集成了业界标准的统一可扩展固件接口（UEFI）	提供白金电源模块，50%负载下电源模块效率高达94% 高效率的单板VRD电源 支持主备供电
GRG200-RK5280		支持华为64 bits高性能多核Kunpeng 920处理器 最大可支持64cores, 2.6GHz，单台服务器支持1个处理器、最大64个内核	单板硬件采用电信级器件和加工工艺流程 支持SAS/SATA/NVMe硬盘 BMC集成管理模块（iBMC） 能够持续监控系统参数	集成在服务器上的iBMC管理模块可用来监控系统运行状态，提供远程管理功能 集成了业界标准的统一可扩展固件接口（UEFI）	提供白金电源模块，50%负载下电源模块效率高达94% 高效率的单板VRD电源 支持主备供电

### 3.2.4 四川长虹: 华为生态关键底座

- ◆ **四川长虹“家电大王”转型智能制造:** 创建于1958年，历经数十年的发展，从期初立业、彩电兴业，到如今的信息电子相关多元拓展，已成为集消费电子、核心器件研发与制造为一体的综合型跨国企业集团。主营家用电器业务，以IT产品分销和专业IT解决方案提供为代表的IT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。
- ◆ **四川长虹为华为生态的重要底座:** 2020年子公司四川华鲲振宇成立，负责西南地区基于华为鲲鹏芯片的服务器、PC机、视觉等产品业务，推动基于鲲鹏技术体系的整体解决方案在9大领域开展应用示范。2022年5月华鲲振宇正式成为华为昇腾万里认证级整机硬件合作伙伴，9月其顺利晋升为“优选级整机硬件伙伴”。同年，华鲲振宇基于昇腾AI打造成都智算中心，一期算力达 300 PFLOPS FP16。

华鲲振宇基于华为处理器/芯片的部分产品介绍

种类	名称	示意图	形态	处理器/芯片	内存支持	AI加速卡/AI处理器	AI算力
AI服务器	AT800 (Model 9000)		4U AI服务器	华为鲲鹏920+昇腾910处理器	最多32个DDR4内存插槽，支持RDIMM；内存设计速率最大可达2933 MT/s；单根内存条容量支持32 GB/64 GB	8 * 昇腾910	2.24 / 1.76 PFLOPS FP16(单颗AI处理器算力280T/220 TFLOPS FP16, 可选)
智能计算设备	AT210 A1		桌面型 边缘推理小盒子	Atlas 200 模块 (昇腾 310 处理器)	1 个 Micro SIM 卡槽 2 个 TF 卡槽(1 路 TF1 安装系统 64G, 1 路 TF2 存储数据最大支持 1T)	昇腾 310 处理器	/
AI服务器	AT800 (Model 9010)		4U AI服务器	Intel处理器+华为昇腾910芯片	最多24个DDR4内存插槽，支持RDIMM	8 * 昇腾910	2.56 / 2.24 / 2 / 1.76 PFLOPS FP16
均衡通用服务器	TG215 B1		2U单路机架服务器	鲲鹏920处理器	16个DDR4-2933 DIMM插槽	/	/
存储型通用服务器	TG525 B1		4U双路机架服务器	2个鲲鹏920处理器 (24/32/48/64核 2.6GHz)	4/32核机型16个DDR4-2933 DIMM插槽；48/64核机型32个DDR4-2933 DIMM插槽	/	/
高性能型通用服务器	TG245 B1		2U四路机架服务器	4个鲲鹏920处理器 (48/64核 2.6GHz)	32个DDR4-2933 DIMM插槽	/	/
高密通用服务器	TG125 B1		1U双路机架服务器	2个鲲鹏920处理器 (48/64核 2.6GHz)	32个DDR4-2933 DIMM插槽	/	/

### 3.2.5 东华软件: 华为深度合作伙伴，打造AI产业生态闭环

- ◆ **优秀的软件和信息系统解决方案提供商:** 东华软件自2001年1月成立以来，不断加大技术创新、产品创新和业务模式创新的力度，已为数千个用户提供了优秀的软件和信息系统解决方案，涵盖多种应用与技术平台，用户涉及医疗、金融、政务、能源、水利、公安、电信等多个行业。公司是是国家规划布局内重点软件企业、国家火炬计划重点高新技术企业、计算机服务及软件特大型企业。
- ◆ **华为深度合作伙伴，打造AI产业生态闭环:** 在大型算力中心建设业务方面，神州新桥与华为展开紧密合作，逐渐成为了华为紧密的合作伙伴。近年来，公司及公司全资子公司累计中标智能计算中心项目超40亿。早在2019年，公司全资子公司神州新桥中标4.12亿武汉人工智能计算中心（国内首个基于昇腾生态的省级人工智能计算中心），标志着公司承接大型人工智能计算中心项目的开端。

武汉人工智能计算中心



公司及全资子公司近年中标智能计算中心项目及中标金额（亿元）

项目名称	中标金额
武汉人工智能计算中心项目	4.12亿
沈阳人工智能计算中心项目	5.01亿
河北人工智能计算中心项目	5.1亿
大连人工智能计算中心项目	5.39亿
国家超级计算长沙中心升级建设项目	0.67亿
武汉超算中心项目	8.7亿
长沙人工智能创新中心项目	4.01亿
宁波人工智能超算中心项目	4.86亿
北京昇腾人工智能计算中心项目	2.35亿

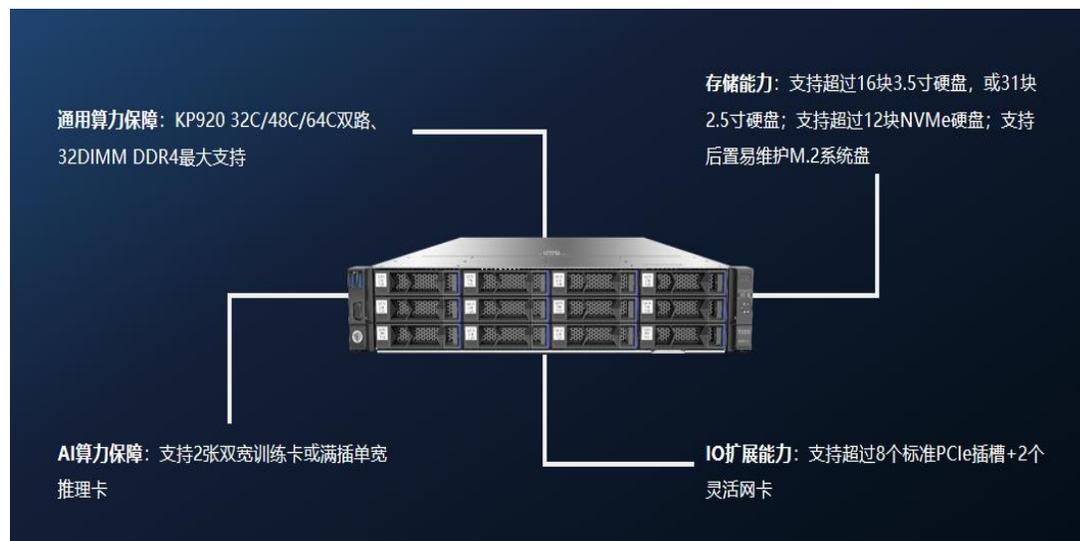
### 3.2.6 同方股份: 践行华为生态，引领信创产业

- ◆ **合志同方，无问西东**：同方股份于1997年由清华大学出资成立并同年上市的高科技企业。同方股份坚持“技术+资本”双轮驱动的发展战略，形成了核技术应用、智慧能源、数字信息、科技创新主干产业集群，打造“智慧生活的创造者”的品牌形象，目前公司的技术、产品和服务已遍及五大洲一百余个国家和地区。
- ◆ **践行华为生态，引领信创产业**：在信创领域，公司推出了自主研发的基于华为鲲鹏处理器的服务器产品，产品从核心芯片、整机架构到固件和操作系统均为国内自主知识产权，不仅具备高安全、高可控的产品优势，而且通过与诸多国内基础软件产品的适配，能够满足数据中心多样性计算、绿色计算的需求。同时，公司还推出了基于华为昇腾芯片的 AI 人工智能服务器，并先后在北京、天津、青岛、宁波、大连等多地落地人工智能超算中心项目。

同方鲲鹏系列通用计算机一览



同方昇腾系列AI服务器





## 04 风险提示

## 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“卡脖子”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/hxqz/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。本报告仅供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**

