神威遠通
sunwaysaga

神威

北京神威远通科技有限公司
Beijing Sunway Saga Science & Technology Co.,Ltd.

公司介绍 全部自主产权 / 自带原始防火墙 / 安全·保密·耐用
Company Introduction

国家自主安全的政治背景

关键核心技术是要不来、买不来、讨不来的，只有把关键核心掌握在自己手中，才能从根本上保障国家经济安全、国防安全和其他安全。

加快科技自立自强步伐，解决外国“卡脖子”问题，健全新型举国体制，强化国家战略科技力量。

—— 2023年1月中共中央政治局第二次集体学习

神威遠通
sunwaysaga

企业简介

北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于2003年，自2023年开始致力于研发生产基于神威太湖之光技术体系的申威系列CPU芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

32核服务器处理器

8核终端处理器

套 片

64核服务器处理器

全自主发展模式

自主创新

安全可控

多项自主知识产权

高校合作协同发展

打造国产自主可控产品

企业资质

公司自成立，注重自身素质提高。近年来，在市场准入及资质认证方面均取得突破性进展。

取得资质证书包括：

中国国家强制性产品认证证书

证书编号：2024010911632610

认证委托人名称及地址
北京神威远通科技有限公司
北京市海淀区上地信息路1号1号楼205室

生产者名称及地址
北京神威远通科技有限公司
北京市海淀区上地信息路1号1号楼205室

生产企业名称及地址
北京中数申威技术有限公司
北京市昌平区回龙观街道立业路6号院1号楼6层1628室、1层北1102-3室及1102-4东侧（昌平示范园）

产品名称和系列、型号、规格
申威服务器
SUNWAY*******（其中*号为 A-Z、a-z、0-9、"-"或为空，仅区分内存、硬盘、不同软件及客户，不影响产品安全及 EMC 性能）：220-240VAC，5.0A，60/50Hz（×2）

产品标准和技术要求
GB 17625.1-2022;GB 4943.1-2022;GB/T 9254.1-2021（A 级）

上述产品符合《强制性产品认证实施规则 电子产品及安全附件》（CNCA-C09-01:2023）的要求，特发此证。

发证日期：2024 年 05 月 23 日　有效期至：2029 年 03 月 14 日

证书信息和有效性可扫描下方二维码或登录发证机构网站查验，也可在认监委网站（www.cnca.gov.cn）查询。

经中国合格评定国家认可委员会认可 CNAS C001-P

签发：谢肇煦

中国质量认证中心
CHINA QUALITY CERTIFICATION CENTRE

中国质量认证中心有限公司

http://www.cqc.com.cn　北京市丰台区南四环西路188号九区5号楼 100070　电话：+86 10 83886666

软件企业认定证书

经审核，北京神威远通科技有限公司　符合《鼓励软件产业和集成电路产业发展的若干政策》和《软件企业认定标准及管理办法》（试行）的有关规定，认定为软件企业，特发此证。

证书编号：京 R-2003-0465　发证机关：

2003 年 12 月 31 日

高新技术企业
证书

企业名称：北京神威远通科技有限公司　证书编号：GR202311005636

发证时间：2023年11月30日　有效期：三年

批准机关：

质量管理体系认证证书

兹证明

北京神威远通科技有限公司

商标注册证

核定使用商品/服务项目（国际分类：35）

注册人　北京神威远通科技有限公司

注册人地址　北京市海淀区上地信息路1号1号楼205室

注册日期　2024年02月28日　有效期至　2034年02月27日

局长　发证机关

商标注册证

核定使用商品/服务项目（国际分类：9）

注册人　北京神威远通科技有限公司

注册人地址　北京市海淀区上地信息路1号1号楼205室

注册日期　2024年03月07日　有效期至　2034年03月06日

局长　发证机关

神威遠通
sunwaysaga

生产能力

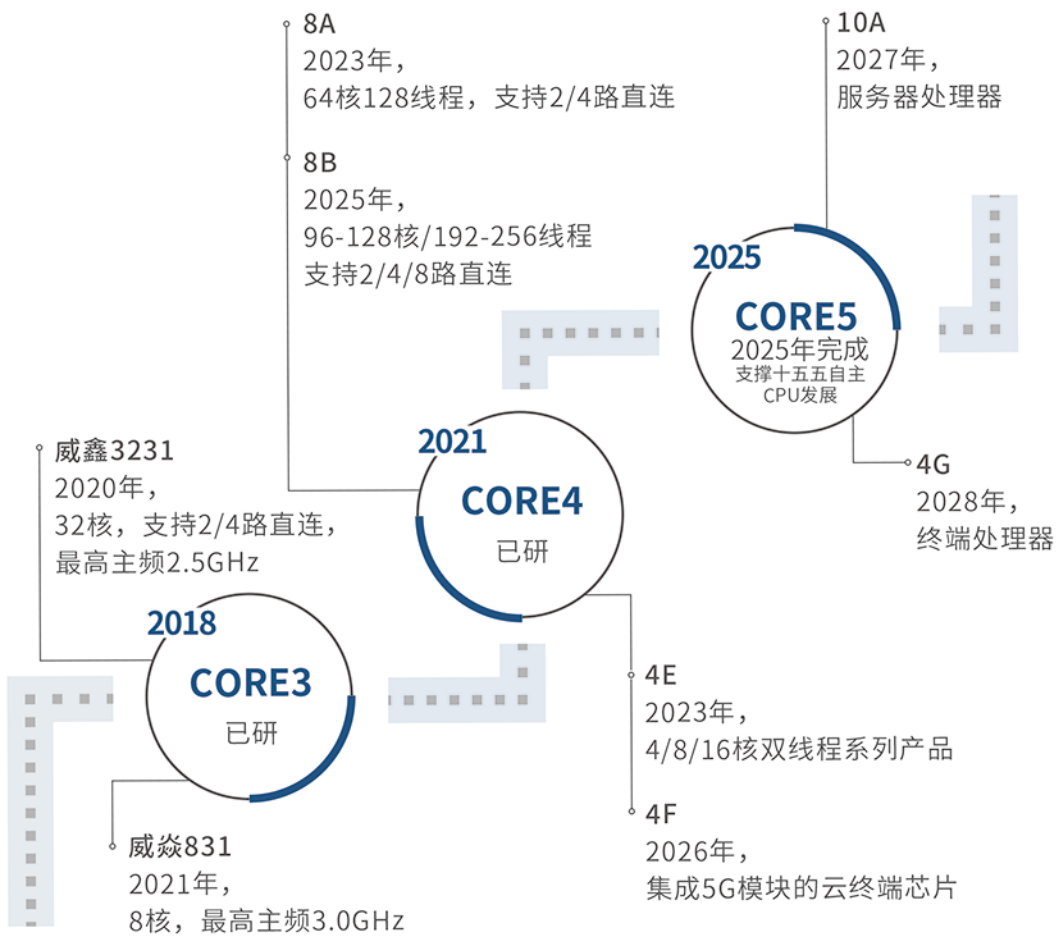
产品涉及军工科技、电脑板卡、汽车电路、工业控制、仪器仪表、网络通信、消费电子等领域。

工厂具备高精度表面贴装、大规模插贴混装电子产品的制造能力，从电子产品工艺设计、元器件采购、贴装、BGA返修、插装、测试、组装至发运的完整技术体系。

产品方案
Product Service

申威处理器技术路线

申威芯片技术路线图



自主指令系统

- “申威64”指令系统，32位定长指令，64位字长
- 整数：支持8位、16位、32位、64位和部分256位整数运算
- 浮点：支持IEEE754单精度和双精度沸点运算，支持浮点除法和浮点平方根运算
- 向量：支持256位的整数短向量、浮点短向量运算
- 密码：支持AES、SHA、SM3、SM4等密码运算

产品特点

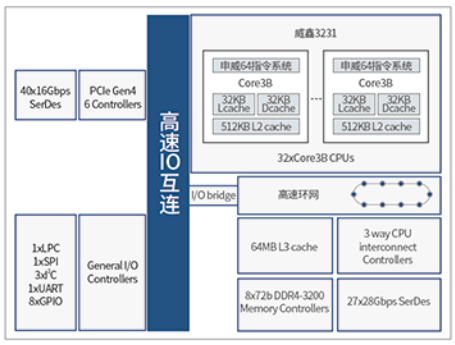
基于“申威64”自主指令系统和自主研发体系
64位通用处理器，采用同构/异购多核结构
采用SoC集成技术，片上集成多路DDR3/DDR4存储控制器借口和多种I/O接口
采用多层次低功耗设计与管理技术，有效降低芯片运行功耗
提供灵活的配置选项，同一处理器产品的核心数、存储控制器路数、PCI-E接口路数等可裁剪，以适应多种应用场景，并降低用户开发成本

神威遠通 sunwaysaga

申威核心处理器介绍

申威3231 服务器处理器

申威3231（SW-WX3231)是一款用于通用服务器、存储服务器、安全设备等产品的高性能处理器。基于申威自主指令集（SW64),集成32个64位RTSC结构的申威处理器核心、8路DDR4存储控制器和40lane PCI-E4.0接口，支持2-4路处理器直连。



核心

- 指令集架构：SW64
- 主频：2.0~2.5GHz
- 核心微结构：Core3B
- 核心数量：32

高速缓存

- 每核心集成32KB指令和32KB数据一级Cache
- 每核心集成512KB二级Cache
- 全片共享64MB三级Cache

存储接口

- 集成8路64位DDR4-3200存储器通道，每个通道支持2根内存条
- 支持ECC校验
- 支持UDIMM、RDIMM和LRDIMM等类型内存条
- 单芯片最大支持2TB内存，四路最大支持8TB内存
- 总带宽204.8GB/s

PCI-E 4.0接口

- 集成40 Lane PCI-E 4.0接口，支持x4、x8和x16灵活配置
- 集成6个PCI-E 控制器
- IO双向聚合带宽160GB/s

多路扩展

- 集成3个直连接口，支持构建2路和4路服务器产品
- 直连接口带宽84GB/s

通用接口

- 3*I²C
- 1*UART
- 1*LPC
- 1*SPI
- 8*GPTO
- 1*JTAG

产品特色

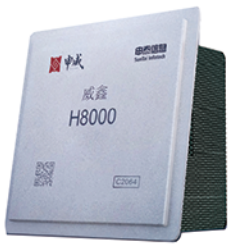
- 高存储容量
- 高存储带宽
- 高IO带宽
- 低功耗
- 低封装热阻
- 支持电压和温都监测
- 支持多模式运行功耗控制

封装与尺寸

- 封装类型：FCLBGA
- 引脚数：4041
- 尺寸：65mm*65mm*4.62mm

产品分类	标准版	低功能版
产品型号	WX3231	WX3231L
核心数量	32	28
典型主频	2.4GHz	2.0GHz
典型功耗	80W	60W

申威H8000 高性能通用处理器



申威 H8000 是一款基于申威自主指令集架构（SW64）的高性能通用处理器，面向大数据、云计算领域，适用于高性能计算服务器、存储服务器和AI服务器等整机产品，可服务于党政、能源、交通、电信和金融等重点行业。

支持硬件双线程　支持国密SM2/3/4　超大三级Cache

核心

- 指令集架构：SW64
- 核心微结构：申威第四代核心（Core4）
- 主频：2.0~2.2GHz
- 核心数量：64个物理核心，128个逻辑核心

存储接口

- 集成8个72位DDR4-3200存储器通道，每个通道支持2根内存条
- 支持ECC校验
- 支持UDIMM、RDIMM、LRDIMM和NVDIMM等类型内存条
- 单芯片支持最大内存容量位4TB，多路系统最大支持8TB

多路扩展

- 集成3个直连接口，支持构建2路和4路服务器产品
- 直连接口双向峰值带宽189GB/s

高速缓存

- 每核心集成64KB指令和64KB数据一级Cache
- 每核心集成512KB二级Cache
- 全片共享128MB三级Cache

PCIe 4.0接口

- 集成40 Lane PCIe 4.0接口，支持x4、x8、x16灵活配置
- IO双向聚合带宽160GB/s

产品特色

- 高存储容量
- 高存储带宽
- 高IO带宽

通用接口

- 3*I^2C
- 1*UART
- 1*LPC
- 1*SPI
- 8*GPTO
- 1*JTAG

封装与尺寸

- 封装类型：FCLBGA
- 尺寸：65mm*65mm*3.978mm

通用服务器
单路、双路、四路

存储服务器
分布式、阵列、双控

行业定制
CPCI、VPX、工控机

人工智能
AI服务器、GPU服务器

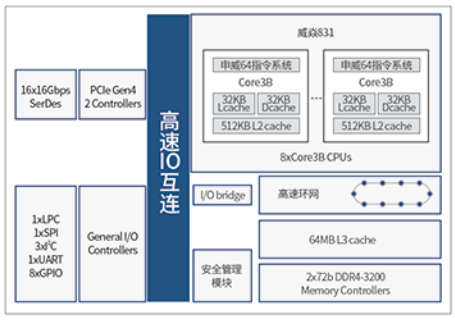
网安设备
网关、密码设备、入侵检测

网安设备
超融合、边缘服务器

神威遠通 sunwaysaga

申威831 终端处理器

申威831（SW-WX831)是一款用于桌面、个人终端、便携计算机等产品的终端处理器，可用于信息安全、工业控制、低端服务器等领域。基于申威自主指令集（SW64),集成8个64位RTSC结构的申威处理器核心、1个主动安全管控模块、2个DDR4存储控制器和16lane PCI-E4.0接口。



核心

- 指令集架构：SW64
- 主频：2.0~3.0GHz
- 核心微结构：Core3B
- 核心数量：8

高速缓存

- 每核心集成32KB指令和32KB数据一级Cache
- 每核心集成512KB二级Cache
- 全片共享16MB三级Cache

存储接口

- 集成2路72位DDR4-3200存储器通道
- 支持ECC校验
- 支持UDIMM、RDIMM和SOMM等类型内存条
- 单芯片最大支持256GB内存
- 存储接口带宽51.2GB/s

PCI-E 4.0接口

- 集成16 Lane PCI-E 4.0接口，支持x8和x16灵活配置
- 集成2个PCI-E 控制器
- IO双向聚合带宽64GB/s

安全模块

- 支持SM2/3/4密码算法
- 支持可行根及密钥管理
- 支持可信引导
- 支持物理内存和IO监管

通用接口

- 3*I²C
- 1*UART
- 1*LPC
- 1*SPI
- 8*GPTO
- 1*JTAG

封装与尺寸

- 封装类型：FCLBGA
- 引脚数：1048
- 尺寸：35mm*35mm*3.16mm

产品特色

- 高性能，最高主频为3.0GHz
- 高存储容量，最大支持256GB
- 内置安全，支持可信引导，主存和IO管控
- 低功耗
- 功耗随核心动态裁剪而降低，能适应各种应用场合

产品分类	标准版	移动版	低功能版
产品型号	WY831-8A	WY831C2	WY831C3
核心数量	8	4	2
典型主频	2.5GHz	2.2GHz	1.2GHz
典型功耗	15W	8W	5W

产品主方向

为教育、金融、党政机关等各部门做国产化替代产品，提供PC机、服务器、存储、交换机、路由器、各种云和相关服务内容，信息安全等。

基于申威CPU研发的包括路由器、交换机、光传输、板卡、网络安全、PC机、服务器等产品。

为军工军队研发定制专装、特装类产品，提供安全，自主可控的AI服务器、交换机、一体机等产品，为特定应用场景提供一体化解决方案。

神威远通存储服务器

基于国产化神威芯片的硬件设备

神威远通 双路计算服务器

神威远通桌面终端PC

神威远通边缘计算服务器

神威远通分布式存储系统

三层交换机

神威远通 sunwaysaga

信创服务器（国产化替代）

神威远通 服务器

神威远通双路服务器 /Sunway3231SL

- 处理器： SW3231*2
- 处理器核数：32 核，含BMC 控制卡
- 内存 128G DDR4：1 块 512GB M.2 SATA，内置2 块 4TB 的 机械盘，配置 RAID 卡
  4 个千兆自适应网口，两路万兆网接口
- 1 个维护千兆自适应网口
- 6*USB3.0 type-A 接口
- 2*VGA 扩展接口：PCIE 扩 展插槽 5 个

神威远通 分布式存储系统

神威远通存储服务器/Sunway3231CC

- 处理器： SW3231*2
- 处理器核数：32 核， 含 BMC 控制卡
- 内存 512G DDR4：2 块 512GB M.2 SATA，内置 12 块 10TB 的机械盘，配置 RAID 卡
  4 个千兆自适应网口，两路万兆网接口
- 1 个维护千兆自适应网口
- 6*USB3.0 type-A 接口
- 2*VGA 扩展接口：PCIE 扩展插槽 5 个

- **集群规模：** 2-10240 节点
- **数据保护：** 多副本 、纠删码
- **协议支持：** POSIX、CIFS、NFS、FTP、S3、iSCSI 、HDFS、cinder等
- **访问控制：** 支持 AD、LDAP、本地用户/组，存储私有权限
- **单文件系统容量：** EB
- 高级功能： WORM、无限快照、配额、存储分层 ，屏蔽操作系统管理员的存储私有权限
- **配置及维护：** Web、CLI、Restful API
- **事件通知：** Log、E-mail、SNMP、IPMI、短信

海量统一

- 横向扩展，单文件系统EB级海量空间 、千亿文件规模
- 同时支持文件/对象/块协议，大数据、对象、文件多协议数据透明共享

安全可靠

- 全局高可用架构集群
- 支持副本/纠删码冗余方式
- 支持文件、块、对象无限快照备份保护

数据访问兜底安全

- 支持存储私有权限，屏蔽操作系统管理员权限

性能优化

- 高性能POSIX RDMA高速网络
- 元数据集群加速海量文件访问效率
- 单文件系统可在线扩展至TB/s吞吐带宽

软硬融合

- 申威3231处理器
- 万兆高速网络
- 统信OS
- 自研分布式存储系统

神威遠通 sunwaysaga

神威远通 元数服务器

神威远通元数服务器/Sunway3231YSJ

- 处理器：SW3231*2
- 处理器核数：32 核，含 BMC 控制卡
- 内存 512G DDR4： 1 块 512GB M.2 SATA, 2 块 960GB SSD SATA, 内置 10 块 10TB 机械盘, 配置 RAID 卡,4 个千兆自适应网口, 两路万兆网接口
- 1 个维护千兆自适应网口
- 6*USB3.0 type-A 接口
- 2*VGA 扩 展接口：PCIE 扩展插槽 5 个

神威远通 文件服务器

神威远通文件服务器/Sunway3231WJ

- 处理器： SW3231*2
- 处理器 核数：32 核， 含 BMC 控制卡
- 内存 256G DDR4: 1 块 512GB M.2 SATA，内置 6 块 8TB 的机械 盘，配置 RAID 卡 4 个千兆自适应网口，两路万兆网接口
- 1 个维护千兆自适应网口
- 6*USB3.0 type-A 接口
- 2*VGA 扩展 口：PCIE 扩展插槽 5 个

神威远通 服务器

神威远通AI算力服务器/Sunway3231AIA

- 处理器：SW3231*2
- 处理器核数：32 核
- 最大支持32路DDR4内存条，单条内存模组支持最大 128GB存储容量，含BMC 控制卡
- 内存 128GDDR4： 1 块 512GB M.2 SATA，内置2 块 4TB 的机械盘，配置 RAID 卡
- 核心计算GPU：R200*1
- 指标：

  TDP: 150W

  FP16: 128T Flops

  FP32: 32T Flops

  INT8: 256Tops

  INT16:128Tops

  显存: 16G
- 4路10/100/1000Mbps自适应RJ45以太网口
- 1路10/100/1000Mbps自适应RJ45维护以太网口

  （可选万兆双口RJ45、25G双口光纤、万兆双口光纤等多种网络接口）
- 6路USB3.0TypeA接口（后出4路，前出2路）
- 3路PCIeGen4x8扩展插槽，2路PCIeGen4x16扩展插槽
- 2路DB15VGA接口（前后面板各出1路）
- 工作温度：+10°C~+35°C
- 储存温度：-40°C~+55°C
- 工作湿度：35%~80%
- 电源输入：输入电源为100-240V/50-60Hz

神威远通 sunwaysaga

神威远通 服务器

神威远通区块链一体机 / Sunway3231BC

- 机箱：2U机架式
- 处理器：32核SW3231 SW64自主指令集架构
- 内存：64GB DDR4，单条内存模组支持最大128GB容量
- 存储：2块512GB M.2 SSD，2块2TB机械硬盘，配备RAID卡
- 网络接口：2路千兆自适应网口，2路万兆光纤接口，1路Console接口
- 电源：550W 1+1冗余电源
- 扩展插槽：3路PCIeGen4×8插槽，2路PCIeGen4×16插槽
- 外部接口：6路USB3.0TypeA接口，1路VGA接口
- 软件功能特点如下：

开箱即用，快速部署：

从硬件支持、平台搭建、应用定制全面一体化

即插即用，支持一键部署区块链，实现秒级的区块链快速组网

一键建链，便捷上链：

用户无需了解区块链技术，开机即可进行引导式快速建链、简单上链，实现数据存证、数据验证、数据溯源、数据共享、业务协同等功能，外部系统通过提供的多语言SDK/API即可实现适配开发

自主可控，安全可靠：

采用威鑫3231高性能信创处理器，兼容麒麟、统信等信创操作系统，兼容达梦、人大金仓等信创数据库，兼容东方通、宝兰德等信创中间件；

支持全国密多级加密机制，支持信创密码卡，保证链上数据、身份、节点通信安全可靠

可选配区块链BaaS平台、跨链平台、安全监测平台：

区块链BaaS平台提供联盟链多链纳管、运维监控、分析展示等丰富的可视化交互功能

跨链平台提供跨链准入与交互机制、跨链身份体系、跨链监管等功能，兼容主流区块链

安全监测平台提供区块链全链路安全监测服务，确保数据链上链下流转全程可溯、安全可控

神威远通 边缘计算服务器

神威远通边缘计算服务器/SunwayTPUBY

- 处理器：主控：8核 ARM A53 @2.3GHz
- 峰值算力： 17.6~211T TOPS (INT8) / 2.2 ~26.4TFLOPS (FP32)
- 操作系统：Linux系统
- 视频解码类型：H.264 & H.265
- 数据采集设备：设备需支持RTSP标准协议
- 智能参数：视频解析，支持8路~192路1080p高清视频智能分析，每路视频可支持3种算法同时运算
- 包含算法如下：

  行为分析：抽烟检测、安全帽检测、反光衣检测、长袖检测、工服检测、攀爬检测、人员倒地检测、人员离岗、人员睡岗、玩手机打电话检测、客流统计、人员扭打、口罩检测

  物品警戒：烟/火检测、电瓶车检测、通道堵塞

  周界警戒：人员徘徊、区域入侵、越线检测、翻越围栏、车辆违停、含车牌识别、小动物识别、人脸识别
- 软件功能如下：

  异常报警：当发生异常情况时，同时进行预警上报和消息记录

  事件管理：异常事件可上传或保存至指定位置，灵活设置留存时间，并提供查阅功能
- 转码输出：支持将接入的原始视频流以GB28181协议输出，支持将原始视频流叠加智能分析框后，以GB28181协议输出
- 内存：标准12GB
- eMMC：标准32GB
- 工作温度：-20°C ~ 60°C
- 网络接口：10/100/1000Mbps自适应*2
- 外部接口：支持USB/HDMI/SATA/RS-232/RS-485 / TF,LTE可选
- 电源接口：直流12V

神威远通 sunwaysaga

神威远通 PC机

硬件规格

- 申威831处理器，8核，主频2.5GHz
- 16GB DDR4内存
- 512GB SSD系统盘
- 1个PCIEX16插槽
- 1个PCIEX4插槽
- 1个千兆以太网口(RJ45)
- 8个USB接口
- 独立显卡，支持VGA、HDMI或DVI接口
- 配置DVD刻录光驱

神威远通 三层交换机

物理端口

- 48个10/100/1000Base-T电接口、8个SFP+光口(支持千兆、万兆模式)

管理接口

- 1个Console配置接口、1个管理以太接口、1个USB接口

外形尺寸

- 机箱高度为1U

电源

- 输入电压（AC）：100～240V，50/60Hz，支持1+1冗余电源

功耗 ≤100W

交换容量 ≥176Gbps

包转发率 ≥64Mpps

MAC地址表 64K

温度

温度工作：0℃到50℃　　存储温度：-40℃到70℃

高性能算力服务器

申威众核AI 推训一体机/SunwayZH1118

· 业界领先的高密液冷一体化设计
· 超强性能、超高密度、超低功耗

- CPU：2颗 Intel 最新一代处理器
- 加速卡： 8 张申威众核AI OAM 模组
- 系统尺寸：2U 机架式
- 电源：Power Shelf 集中供电
- 散热：液冷
- 内存：DDR5 RDIMM
- 硬盘：NVMe SSD
- 网络：Infiniband / Ethernet
- 软件：基础软件、计算加速库、通信加速库、训练框架、推理引擎、预置模型库、企业级 AI 平台
- 计算性能：80 TFLOPS FP32； 1920 TFLOPS FP16； 3840 TOPS INT8

申威众核AI 推训一体机/SunwayZH1008

· 4U 8卡设计，软硬一体优化
· 大模型应用专属底座

- CPU：2颗 Intel / AMD 最新一代处理器
- 加速卡： 8 张申威众核AI PCIe 加速卡
- 系统尺寸：4U 机架式
- 电源：4*2000W 交流 & 240V 高压直流电源模块
- 散热：风冷
- 内存：DDR5 RDIMM
- 硬盘：NVMe SSD
- 网络：Infiniband / Ethernet
- 软件：基础软件、计算加速库、通信加速库、训练框架、推理引擎、预置模型库、企业级 AI 平台
- 计算性能：80 TFLOPS FP32； 1920 TFLOPS FP16； 3840 TOPS INT8

神威遠通
sunwaysaga

异构众核为摆脱GPU追随路径最有前景的技术路线

	ASIC 架构	GPU 架构	异构众核架构
通用性	很差，聚焦单独特定场景	高 ,AI领域高度通用	更高，天然适配HPC+AI
算力性能	特定场景下，平均算力最高	通用场景及最高运算力上，更具优势	系统级应用及复杂科学计算，算力更强
国外玩家	Google 无创业公司挑战	NVIDIA 无创业公司挑战	富岳 Esperanto TECHNOLOGIES tenstorrent 海外技术大牛初创企业，多采用众核架构
国内玩家	Cambricon 寒武纪 Enflame 燧原科技	摩尔线程 壁仞科技 METAX 沐曦 海博半导体 奇芯科技 AZURENGINE	Tecorigin 太初元碁
应用场景	更多用于挖矿、音视频解码等 AI 推理专用场景	图形渲染 起家 向自动驾驶、医疗设备、游戏设备、加密货币等AI推理+AI 训练通用场景不断拓展	复杂科学计算+AI训练+AI推理的 HPC+AI高度通用场景 计算精度要求明显更高
软件生态	自研 不通用	CUDA及CUDA-like体系	自主可控 自研通用体系及工具链
	✕	✕	✓
	在特定场景可以做到性能最优，有应用落地场景舒适区，更适合做推理芯片；由于其低延迟、低功耗与高运算效率，在训练领域也有发展空间（如寒武纪、燧原）；但由于灵活性差，算法如果迭代则需要重新开发。	国内玩家采用类CUDA体系，生态成熟， 迭代也更容易；但大部分企业无AI芯片原创设计能力，普遍通过购买成熟IP，追随英伟达GPU迭代并走类CUDA路线，因此十分“烧钱”且上限被英伟达封死。	太初的异构众核架构源于申威体系，核心能力体现在调度与通信上。超算体系的多年积累，具备更好的系统能力。自主可控的高通用计算平台，完全独立的AI芯片设计能力，架构更好支持大规模、高并行度的计算，符合未来发展方向。

申威众核异构芯片架构，高性能系统的底层支撑

【一主四核】的众核异构架构

融合异构的体系结构

一主四核

轻量级片上通信机制

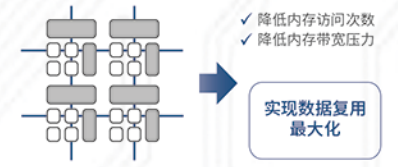
提升运算核心间的协同效率

灵活的片上存储体系

支持数据在核心存储和主存间的高效异步数据传输

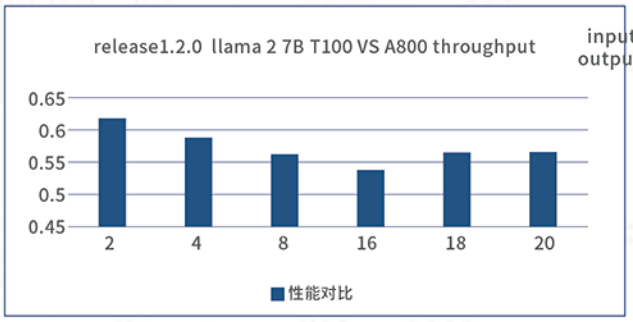
高效的智能加速核心

智能运算加速部件采用脉动阵列

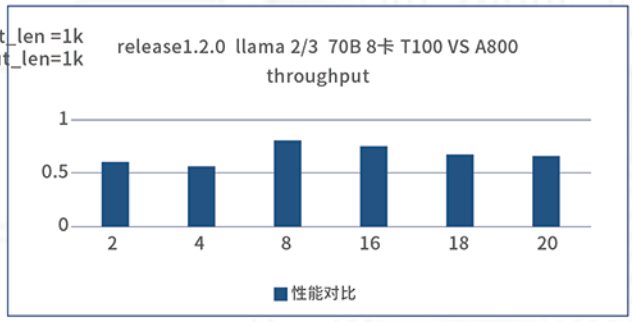


【推理】申威众核AI大模型推理性能接近A800 70%

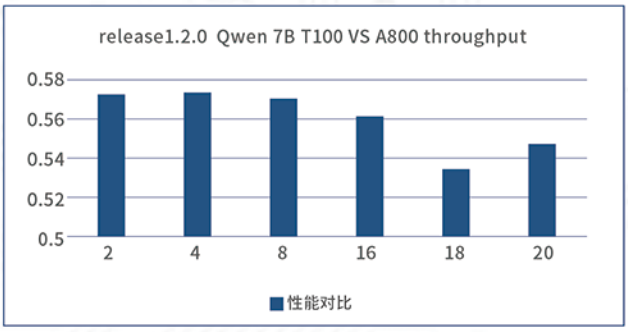
input_len =1k
output_len=1k



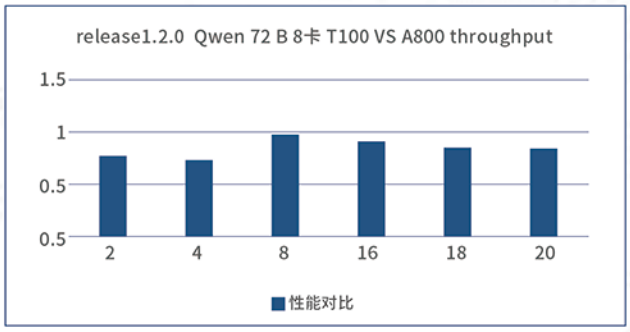
llama 7B 单卡 吞吐平均是A800 的57.2%



llama 70B 8卡 吞吐平均是A800 的71%



Qwen-7B-Chat 单卡 吞吐平均是A800 的56%



Qwen-72B 8卡 吞吐平均是A800 的84.2%

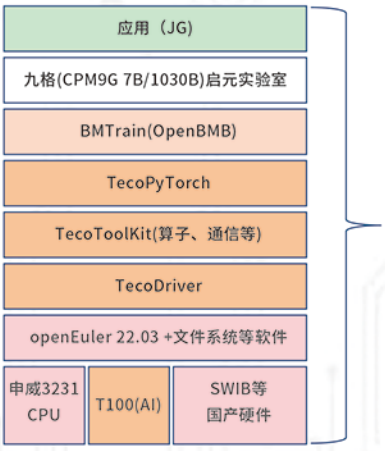
测试机器说明：

1）NVIDIA A800 80GB PCIe，Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz * 2，Ubuntu 20.04.5 LTS

2）单卡：T100-64G @2.5GHz，AMD EPYC 7453 28-Core Processor@2.75GHz，NF5468A5，Ubuntu22.04

3）多卡：T100-64G @2.5GHz*8，AMD EPYC 9654 96-Core Processor@3.7GHz，H3C UniServer R5350 G6，Ubuntu2204软件栈版本：V1.2.0 release

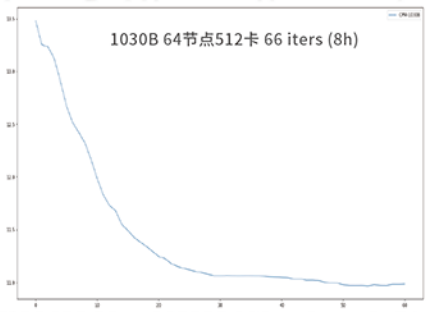
【大模型预训练】纯国产环境，完成512卡集群万亿模型预训练适配



✓ 功能点1：功能: 前向，反向，loss计算(对比GPU 1e-5)验收通过；
产品功能性测例17/17通过

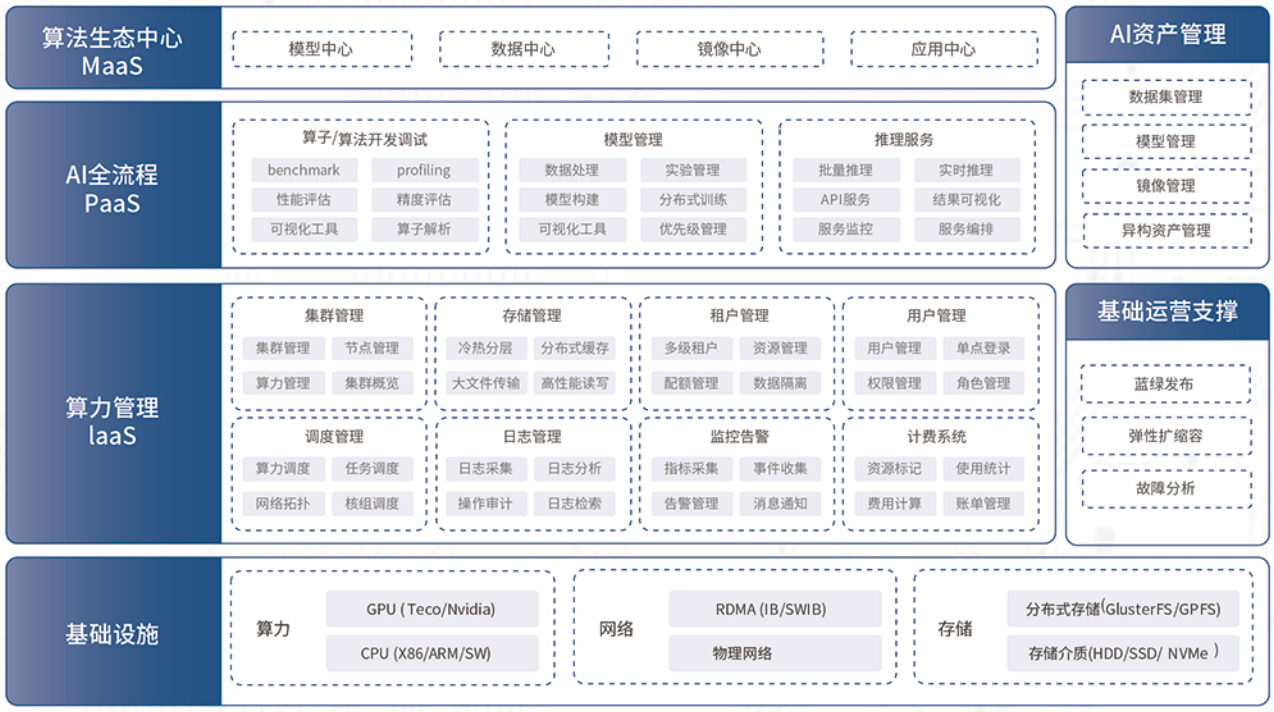
```
finish testing nccl_backward
 82%|____________________________________________________| 14/17 [08:31<01:03, 21.18s/it]
finish testing column_parallel_linear
 88%|_____________________________________________________| 15/17 [08:37<00:33, 16.74s/it]
finish testing row_parallel_linear
 94%|______________________________________________________| 16/17 [08:43<00:13, 13.64s/it]
                                        finish testing training
100%|_______________________________________________________| 17/17 [1:44:37<00:00, 369.27s/it]
```

✓ 功能点1：功能点2：万亿: cpm9g-1030B在国产系统(申威3231+欧拉)64节点512卡正常跑通

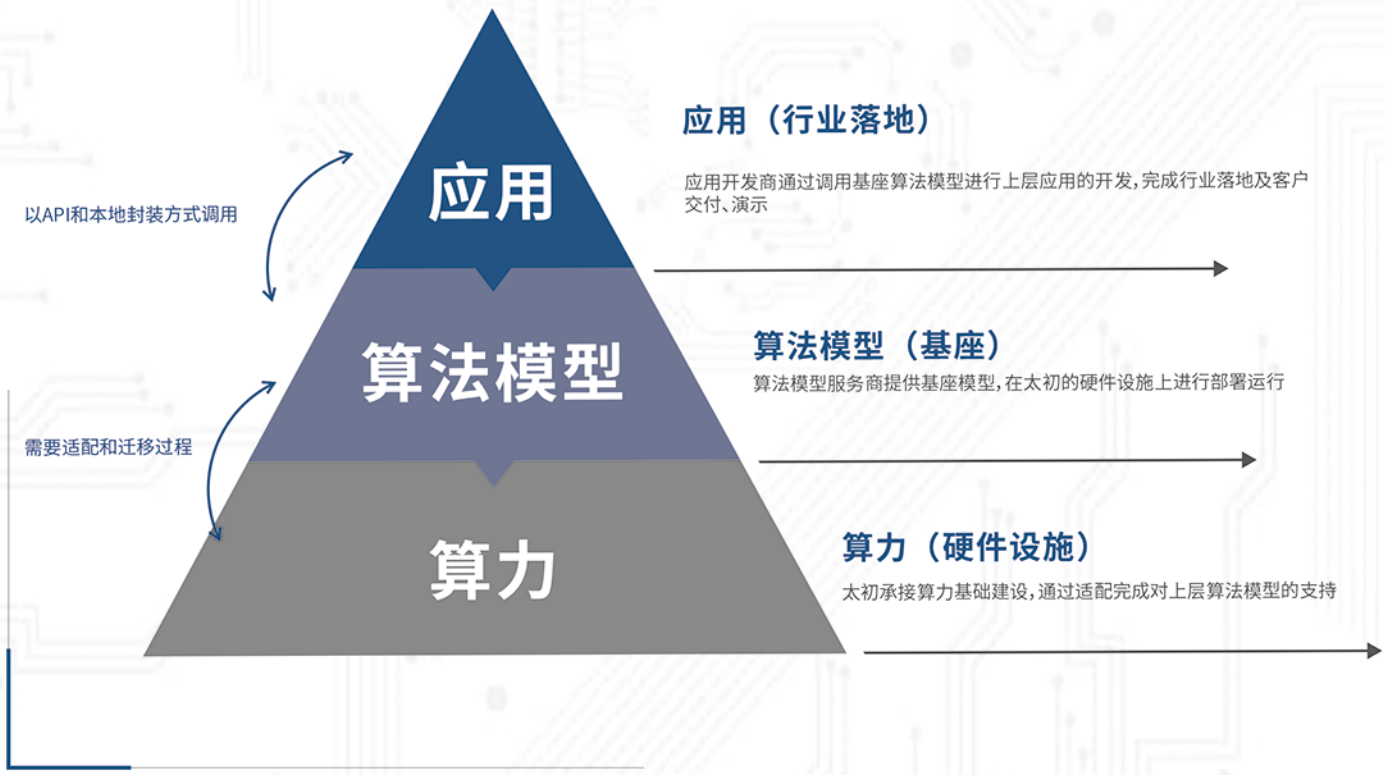


神威遠通 sunwaysaga

智算中心解决方案



行业解决方案



案例：盐城智算中心

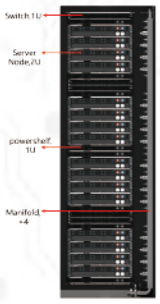
- 规划建设800P，国产化率超60%，一期建设206P，**2023年获批国家新一代人工智能公共算力开放创新平台（首批）。**
- 22年6月启动运营至今，近满负载运行，运营收入超亿元。
- 在人工智能、航空航天、医疗医药等领域，围绕智算产业化建设达成产研合作，与高校科研机构合作成立产业化人才培养项目，打造“产教融合管理体系”。

国家新一代人工智能公共算力开放创新平台

1	天河智能计算开放创新平台 (国家超级计算天津中心)	天津
2	智能计算新一代人工智能开放创新服务平台 (中科可控信息产业有限公司)	昆山
3	国家新一代人工智能公共算力开放创新平台 (山西数据技术有限公司)	长治
4	超大规模混合算力海洋人工智能公共算力开放创新平台 (青岛国实科技集团有限公司)	青岛
5	国家新一代人工智能公共算力开放创新平台 (湖南大学国家超级计算长沙中心)	长沙
6	成都超算中心人工智能公共算力开放创新平台 (成都超算中心运营管理有限公司)	成都
7	面向科学发现的人工智能开放平台 (中国科学院计算机网络信息中心)	北京
8	“乌镇之光”超算人工智能公共算力平台 (乌镇之光(桐乡)超算中心有限公司)	乌镇
9	“嵩山” 人工智能公共算力开放创新平台 (郑州大学国家超级计算郑州中心)	郑州
10	合肥人工智能计算中心 (合肥市大数据资产运营有限公司)	合肥
11	**盐城市人工智能开放创新平台 (盐城神威超级计算技术有限公司)**	**盐城**
12	宁波人工智能算力创新平台 (宁波中科信息技术应用研究院)	宁波
13	福建人工智能计算中心 (福州市电子信息集团有限公司	福州
14	中原人工智能公共算力开放创新平台 (许昌市鲲鹏人工智能计算有限公司)	许昌
15	河北人工智能计算中心 (河北智 能计算科技有限公司)	廊坊
16	广州人工智能公共算力中心 (广州广电研究院有限公司)	广州
	… …	

神威遠通 sunwaysaga

案例：无锡超算二期AI预训练集群



- 2024年Q2交付
- 总建设200P FP16全国产智能算力
- 自研**业内最高密度**机柜
- 单机柜配置**128张**申威众核AI加速卡
- 单机柜算力达到**32PFlops**，为国内最高算力密度
- 采用集中式供电，单机柜功耗达**100KW**

案例：延安智算中心

- 2024年Q1交付
- 总建设200P FP16智能算力，包含通用+国产
- 采用四链路400Gbps计算网络互联
- 支撑大模型预训练、微调和推理
- 目标是支撑加快延安**数字经济产业转型**，率先在全省实现智算中数字经济引擎的大布局

案例：蠡湖未来城智算中心

- 2024年Q3交付
- 总建设300P FP16智能算力，包含通用+国产
- 采用四链路400Gbps计算网络互联
- 支撑大模型预训练、微调和推理
- 目标是以算力基础设施为核心，搭建配备大模型的公共服务平台，支持无锡蠡湖未来城孵化算法软件层企业，打通AIGC产业全产业链

我们的解决方案

全国产化、高密度全目标解析，充分利旧现有监控

单台最大支持
144路全目标结构化分析或192 路人脸分析或192路人车非分析

全目标解析
单机支持人脸+结构化

超高性能引擎
以低成本助力全量解析

轻量化应用
以最小化系统解决找人/车痛点

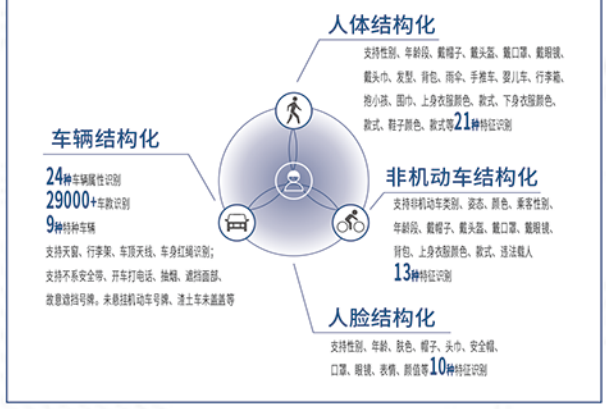
多算法解析：灵活调度、自由选择

客户根据不同的业务类型，灵活配置解析算法

一台起步
以计算卡为单位，每卡执行一类人物，可自由切换，随需调配

灵活调度
提供灵活的任务调度接口，客户可根据业务情况，灵活分配任务

丰富的结构化属性：四大类60+种描述项



视觉感知类算法

车辆目标属性识别

非机动车目标属性识别

人体目标属性识别

人脸目标属性

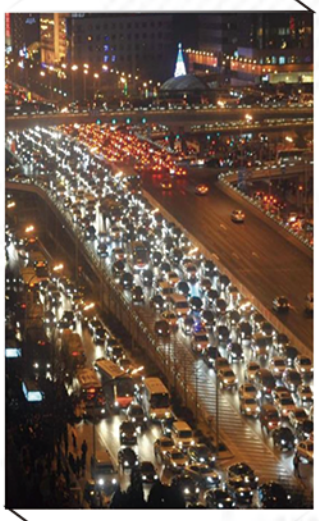
年龄、性别、眼镜
帽子、口罩、表情

神威遠通
sunwaysaga

全目标结构化解决方案

人脸　人体　非机动车　机动车　行为

(SE6-1U生态智算服务器x格灵深瞳算法)



解决了什么问题

本方案主要解决了对海量视频进行的**全目标结构化分析的问题。**

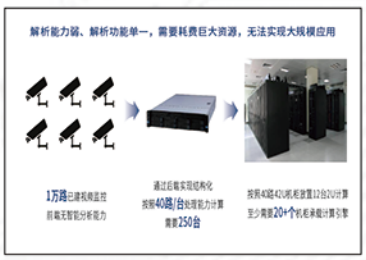
如何解决问题

采用**格灵深瞳**的算法和**应用**，通过**ES6-1U生态智算服务器**提供基础算力对海量视频进行人脸和人车非结构化分析。

方案的优势和亮点

广泛识别　全面分析　强化实战　自主可控

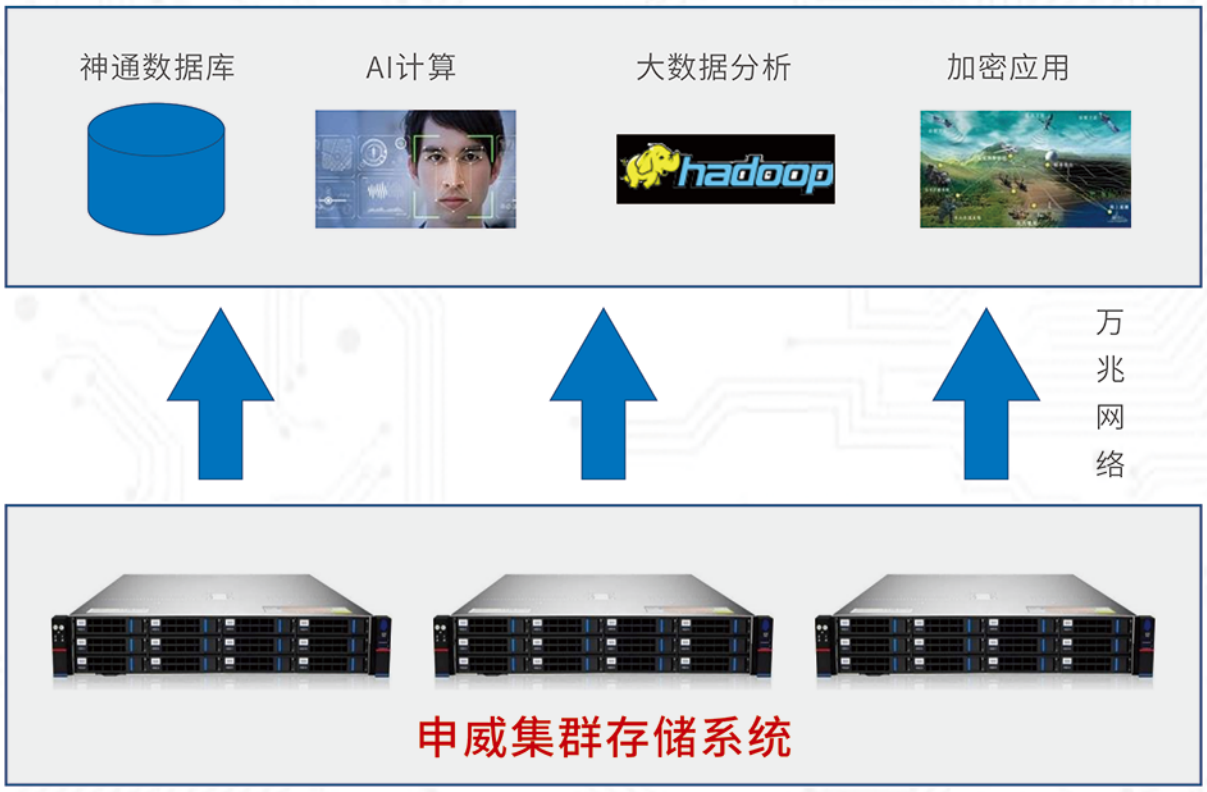
视频存储难



平台整合难

视频分析难

某研究所超算中心



- 70台12块12TB的申威平台存储服务器提供10080TB容量，有效支撑了超算应用。
- 支撑全国产平台优化国产化关系型分布式数据库平台。
- 系统内置双缓存机制，有效屏蔽掉视频点播中常见热点数据问题，减轻系统整体访问压力。
- 灵活的集群存储架构，具有故障恢复、故障自动切换等功能；可不影响用户体验的前提下进行应用节点数量、传输带宽、数据迁移、存储容量等方面的扩展，降低运维成本 。

神威遠通 sunwaysaga

大模型赋能电网各业务环节

应用、模型、样本、算力、机制等要素，聚焦业务流程全环节，推动人工智能规模化应用

安全监测

智能巡检

智能决策

作业管控

智能工单

智能客服

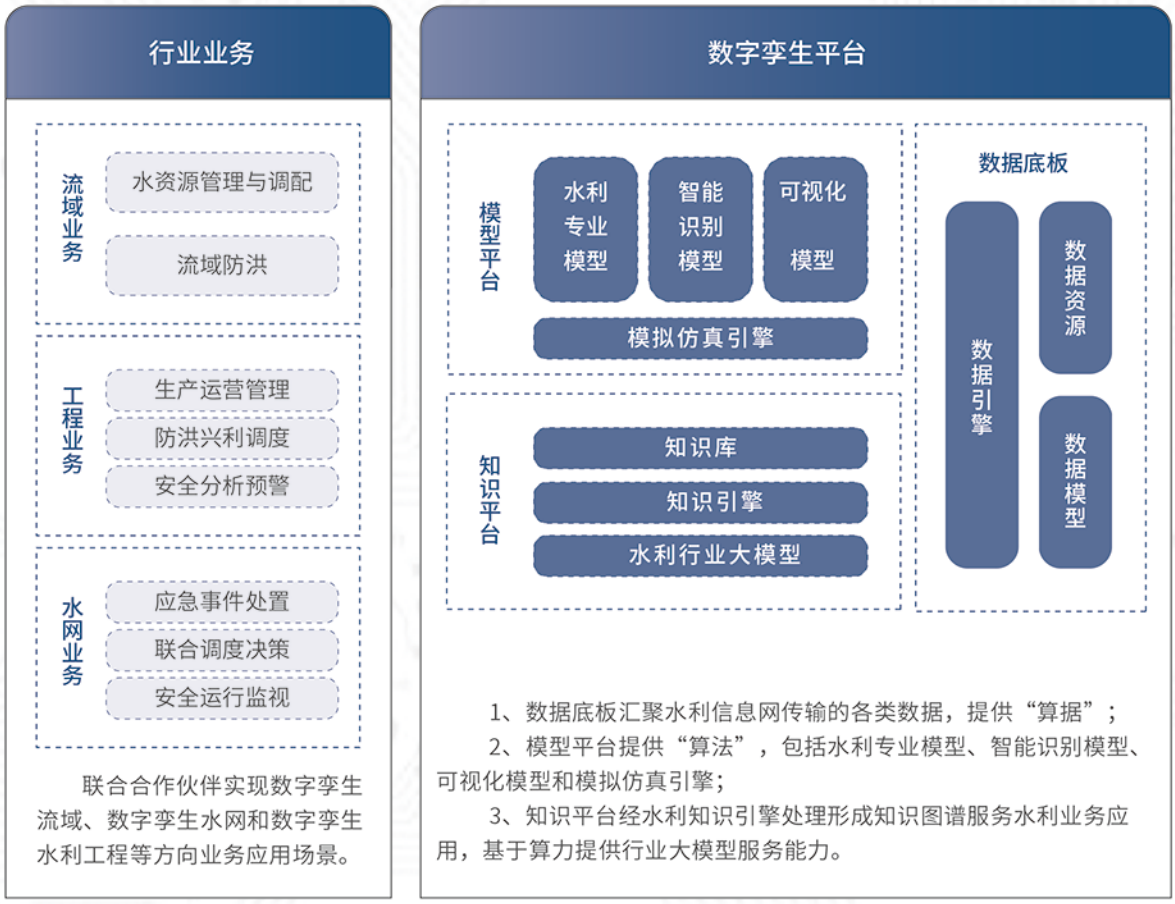
仿真推演

流程优化

多模态大模型围绕计算机视觉、智能语音、自然语言处理、专家系统、智能决策、动作控制等融合A能力，支持输电无人机智能巡检、变电站智能巡视、智能客服、行业问答、现场作业智能管控等各类应用。

水利

以国产GPU及智算云平台为核心，提供信息化基础设施解决方案，构建数字孪生平台（数据底座、模型平台、知识平台），实现流域、水网和水利工程等方向业务应用场景，持续提升智慧水利建设。



1、水利部已印发智慧水利建设、数字孪生流域建设技术要求等系列文件，推进智慧水利建设是推动新阶段水利高质量发展的六条路径之一，数字孪生流域是智慧水利建设的核心和关键，七大江河数字孪生流域建设是“十四五”期间的重要任务 。

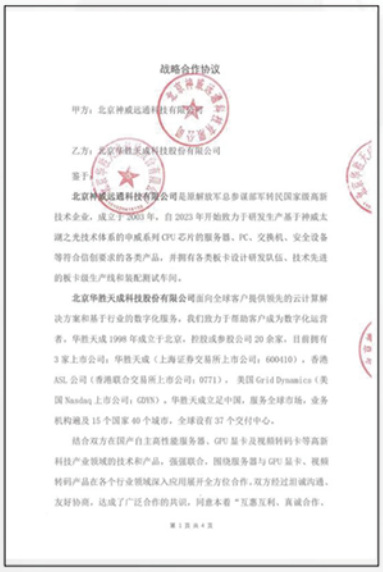
2、以国产信创GPU及云平台为核心，研发构建数字孪生信息化基础设施产品和解决方案，搭建数字孪生平台（数据底座、模型平台、知识平台），联合合作伙伴拓展数字孪生流域、数字孪生水网和数字孪生水利工程等方向等业务应用场景；

神威遠通 sunwaysaga

信创市场战略合作伙伴

核心级合作伙伴

北京华胜天成科技股份有限公司



战略合作协议

甲方：北京神威远通科技有限公司

乙方：北京华胜天成科技股份有限公司

鉴于：

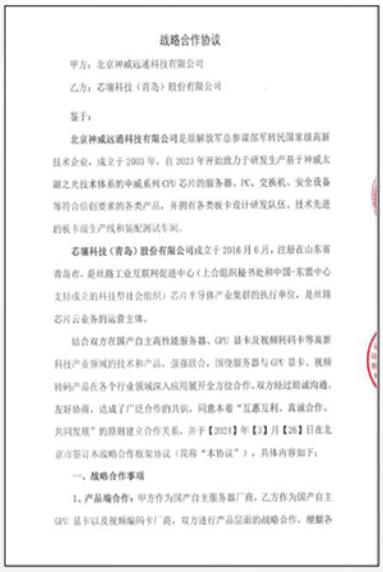
北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于2003年，自2023年开始致力于研发生产基于神威太湖之光技术体系的申威系列CPU芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

北京华胜天成科技股份有限公司面向全球客户提供领先的云计算解决方案和基于行业的数字化服务，我们致力于帮助客户成为数字化运营者，华胜天成1998年成立于北京，控股或参股公司20余家，目前拥有3家上市公司：华胜天成（上海证券交易所上市公司：600410），香港ASL公司（香港联合交易所上市公司：0771），美国Grid Dynamics（美国Nasdaq上市公司：GDYN），华胜天成立足中国，服务全球市场，业务机构遍及15个国家40个城市，全球设有37个交付中心。

结合双方在国产自主高性能服务器、GPU显卡及视频转码卡等高新科技产业领域的技术和产品，强强联合，围绕服务器与GPU显卡、视频转码产品在各个行业领域深入应用展开全方位合作，双方经过坦诚沟通、友好协商，达成了广泛合作的共识，同意本着“互惠互利、真诚合作、

第1页共4页

芯壤科技（青岛）股份有限公司



战略合作协议

甲方：北京神威远通科技有限公司

乙方：芯壤科技（青岛）股份有限公司

鉴于：

北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于2003年，自2023年开始致力于研发生产基于神威太湖之光技术体系的申威系列CPU芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

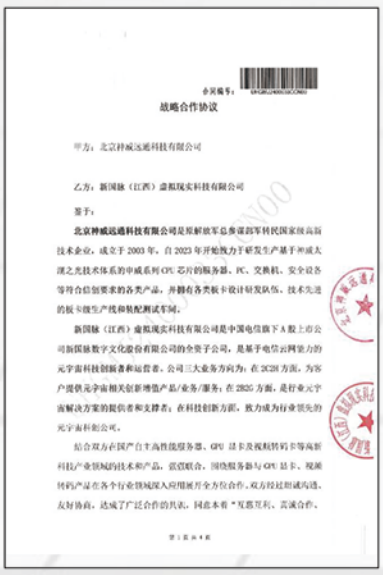
芯壤科技（青岛）股份有限公司成立于2016月6月，注册在山东省青岛市，是丝路工业互联网促进中心（上合组织秘书处和中国-东盟中心支持成立的科技型社会组织）芯片半导体产业集群的执行单位，是丝路芯片云业务的运营主体。

结合双方在国产自主高性能服务器、GPU显卡及视频转码卡等高新科技产业领域的技术和产品，强强联合，围绕服务器与GPU显卡、视频转码产品在各个行业领域深入应用展开全方位合作，双方经过坦诚沟通、友好协商，达成了广泛合作的共识，同意本着“互惠互利、真诚合作、共同发展”的原则建立合作关系，并于【2024】年【3】月【26】日在北京市签订本战略合作框架协议（简称“本协议”），具体内容如下：

一、**战略合作事项**

1、**产品端合作：**甲方作为国产自主服务器厂商，乙方作为国产自主GPU显卡以及视频编码卡厂商，双方进行产品层面的战略合作，根据各

新国脉（江西）虚拟现实科技有限公司



办文编号：

战略合作协议

甲方：北京神威远通科技有限公司

乙方：新国脉（江西）虚拟现实科技有限公司

鉴于：

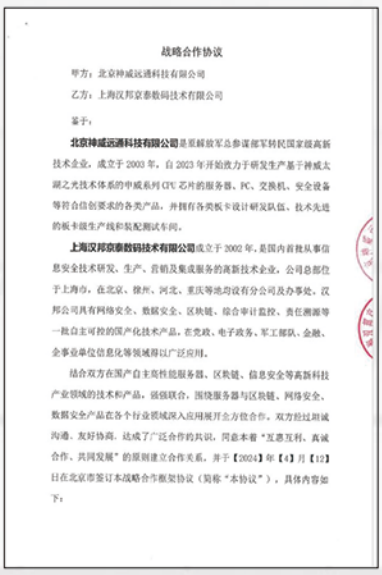
北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于2003年，自2023年开始致力于研发生产基于神威太湖之光技术体系的申威系列CPU芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

新国脉（江西）虚拟现实科技有限公司是中国电信旗下A股上市公司新国脉数字文化股份有限公司的全资子公司，是基于电信云网能力的元宇宙科技创新赛和运营者。公司三大业务方向为：在2C2H方面，为客户提供元宇宙相关创新增值产品/业务/服务；在2B2G方面，是行业元宇宙解决方案的提供者和支撑者；在科技创新方面，致力成为行业领先的元宇宙科创公司。

结合双方在国产自主高性能服务器、GPU显卡及视频转码卡等高新科技产业领域的技术和产品，强强联合，围绕服务器与GPU显卡、视频转码产品在各个行业领域深入应用展开全方位合作，双方经过坦诚沟通、友好协商，达成了广泛合作的共识，同意本着“互惠互利、真诚合作、

第1页共4页

汉邦软科集团



战略合作协议

甲方：北京神威远通科技有限公司

乙方：上海汉邦京泰数码技术有限公司

鉴于：

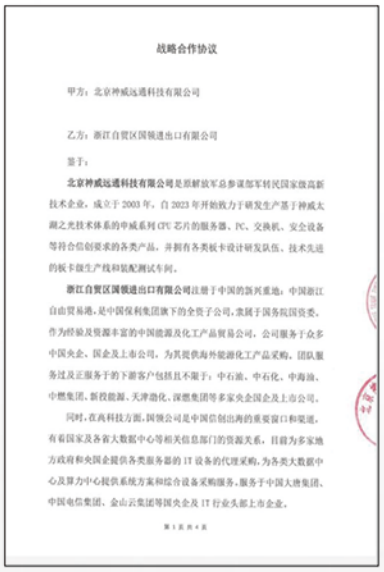
北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于2003年，自2023年开始致力于研发生产基于神威太湖之光技术体系的申威系列CPU芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

上海汉邦京泰数码技术有限公司成立于2002年，是国内首批从事信息安全技术研发、生产、营销及集成服务的高新技术企业，公司总部位于上海市，在北京、徐州、河北、重庆等地均设有分公司及办事处，汉邦公司具有网络安全、数据安全、区块链、综合审计监控、责任溯源等一批自主可控的国产化技术产品，在党政、电子政务、军工部队、金融、企事业单位信息化等领域得以广泛应用。

结合双方在国产自主高性能服务器、区块链、信息安全等高新科技产业领域的技术和产品，强强联合，围绕服务器与区块链、网络安全、数据安全产品在各个行业领域深入应用展开全方位合作，双方经过坦诚沟通、友好协商，达成了广泛合作的共识，同意本着“互惠互利、真诚合作、共同发展”的原则建立合作关系，并于【2024】年【4】月【12】日在北京市签订本战略合作框架协议（简称“本协议”），具体内容如下：

核心级合作伙伴

浙江自贸区国领进出口有限公司



战略合作协议

甲方：北京神威远通科技有限公司

乙方：浙江自贸区国领进出口有限公司

鉴于：

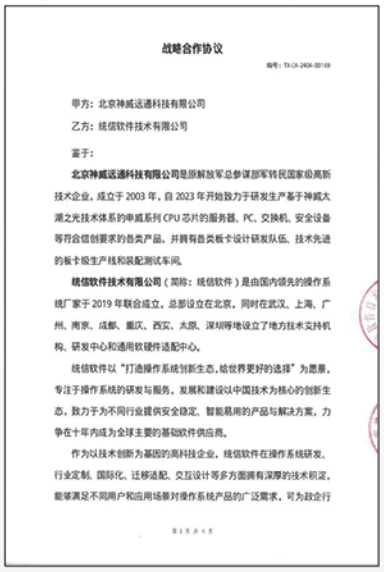
北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于 2003 年，自 2023 年开始致力于研发生产基于神威太湖之光技术体系的申威系列 CPU 芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

浙江自贸区国领进出口有限公司注册于中国的新兴重地，中国浙江自由贸易港，是中国保利集团旗下的全资子公司，隶属于国务院国资委，作为经验及资源丰富的中国能源及化工产品贸易公司，公司服务于众多中国央企、国企及上市公司，为其提供海外能源化工产品采购，团队服务过及正服务于的下游客户包括且不限于：中石油、中石化、中海油、中燃集团、新投能源、天津渤化、深燃集团等多家央企国企及上市公司。

同时，在高科技方面，国领公司是中国信创出海的重要窗口和渠道，有着国家及各省大数据中心等相关信息部门的资源关系，目前为多家地方政府和央国企提供各类服务器的 IT 设备的代理采购，为各类大数据中心及算力中心提供系统方案和综合设备采购服务，服务于中国大唐集团、中国电信集团、金山云集团等国央企及 IT 行业头部上市企业。

第 1 页 共 4 页

统信软件技术有限公司



战略合作协议

编号：TS-CA-2404-00169

甲方：北京神威远通科技有限公司

乙方：统信软件技术有限公司

鉴于：

北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于 2003 年，自 2023 年开始致力于研发生产基于神威太湖之光技术体系的申威系列 CPU 芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

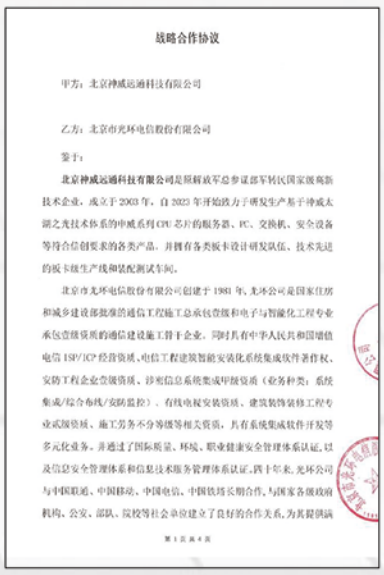
统信软件技术有限公司（简称：统信软件）是由国内领先的操作系统厂家于 2019 年联合成立，总部设立在北京，同时在武汉、上海、广州、南京、成都、重庆、西安、太原、深圳等地设立了地方技术支持机构、研发中心和通用软硬件适配中心。

统信软件以“打造操作系统创新生态，给世界更好的选择”为愿景，专注于操作系统的研发与服务，发展和建设以中国技术为核心的创新生态，致力于为不同行业提供安全稳定、智能易用的产品与解决方案，力争在十年内成为全球主要的基础软件供应商。

作为以技术创新为基因的高科技企业，统信软件在操作系统研发、行业定制、国际化、迁移适配、交互设计等多方面拥有深厚的技术积淀，能够满足不同用户和应用场景对操作系统产品的广泛需求，可为政企行

第 1 页 共 4 页

北京市光环电信股份有限公司



战略合作协议

甲方：北京神威远通科技有限公司

乙方：北京市光环电信股份有限公司

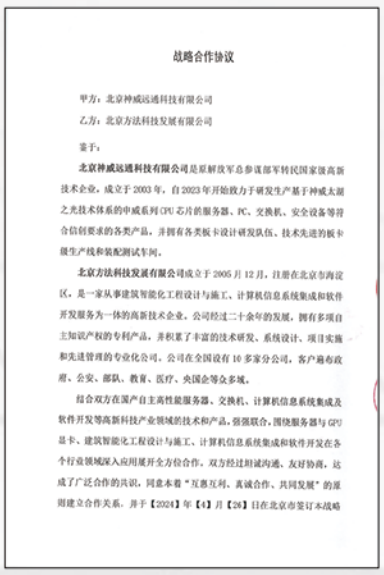
鉴于：

北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于 2003 年，自 2023 年开始致力于研发生产基于神威太湖之光技术体系的申威系列 CPU 芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

北京市光环电信股份有限公司创建于 1981 年，光环公司是国家住房和城乡建设部批准的通信工程施工总承包资质和电子与智能化工程专业承包壹级资质的通信建设施工特十企业，同时具有中华人民共和国增值电信 ISP/ICP 经营资质、电信工程建筑智能安装化系统集成软件著作权、安防工程企业壹级资质、涉密信息系统集成甲级资质（业务种类：系统集成/综合布线/安防监控）、有线电视安装资质、建筑装饰装修工程专业贰级资质、施工劳务不分等级等相关资质，具有系统集成软件开发等多元化业务。并通过了国际质量、环境、职业健康安全管理体系认证，以及信息安全管理体系和信息技术服务管理体系认证。四十年来，光环公司与中国联通、中国移动、中国电信、中国铁塔长期合作，与国家各级政府机构、公安、部队、院校等社会单位建立了良好的合作关系，为其提供涵

第 1 页 共 4 页

北京方法科技发展有限公司



战略合作协议

甲方：北京神威远通科技有限公司

乙方：北京方法科技发展有限公司

鉴于：

北京神威远通科技有限公司是原解放军总参谋部军转民国家级高新技术企业，成立于 2003 年，自 2023 年开始致力于研发生产基于神威太湖之光技术体系的申威系列 CPU 芯片的服务器、PC、交换机、安全设备等符合信创要求的各类产品，并拥有各类板卡设计研发队伍、技术先进的板卡级生产线和装配测试车间。

北京方法科技发展有限公司成立于 2005 月 12 月，注册在北京市海淀区，是一家从事建筑智能化工程设计与施工、计算机信息系统集成和软件开发服务为一体的高新技术企业，公司经过二十余年的发展，拥有多项自主知识产权的专利产品，并积累了丰富的技术研发、系统设计、项目实施和先进管理的专业化公司，公司在全国设有 10 多家分公司，客户遍布政府、公安、部队、教育、医疗、央国企等众多域。

综合双方在国产自主高性能服务器、交换机、计算机信息系统集成及软件开发等高新科技产业领域的技术和产品，强强联合，围绕服务器与 GPU 显卡、建筑智能化工程设计与施工、计算机信息系统集成和软件开发在各个行业领域深入应用展开全方位合作，双方经过坦诚沟通、友好协商，达成了广泛合作的共识，同意本着“互惠互利、真诚合作、共同发展”的原则建立合作关系，并于【2024】年【4】月【26】日在北京市签订本战略

神威遠通 sunwaysaga

神威遠通
sunwaysaga

中国芯
中国架构
中国操作系统

0086-82895308

北京市海淀区上地创业园1座205室

https://www.sunwaysaga.com/