



PRODUCT MANUAL

# 产品手册 耀慧科技

SHENZHEN YAOHUI AI TECHNOLOGY CO., LTD



时间: 2025.09





# CONTENTS



01

**公司理念**

CORPORATE PHILOSOPHY

02

**产品型号**

PRODUCT MODEL

03

**GPU维保**

GPU M&R





01  
PART

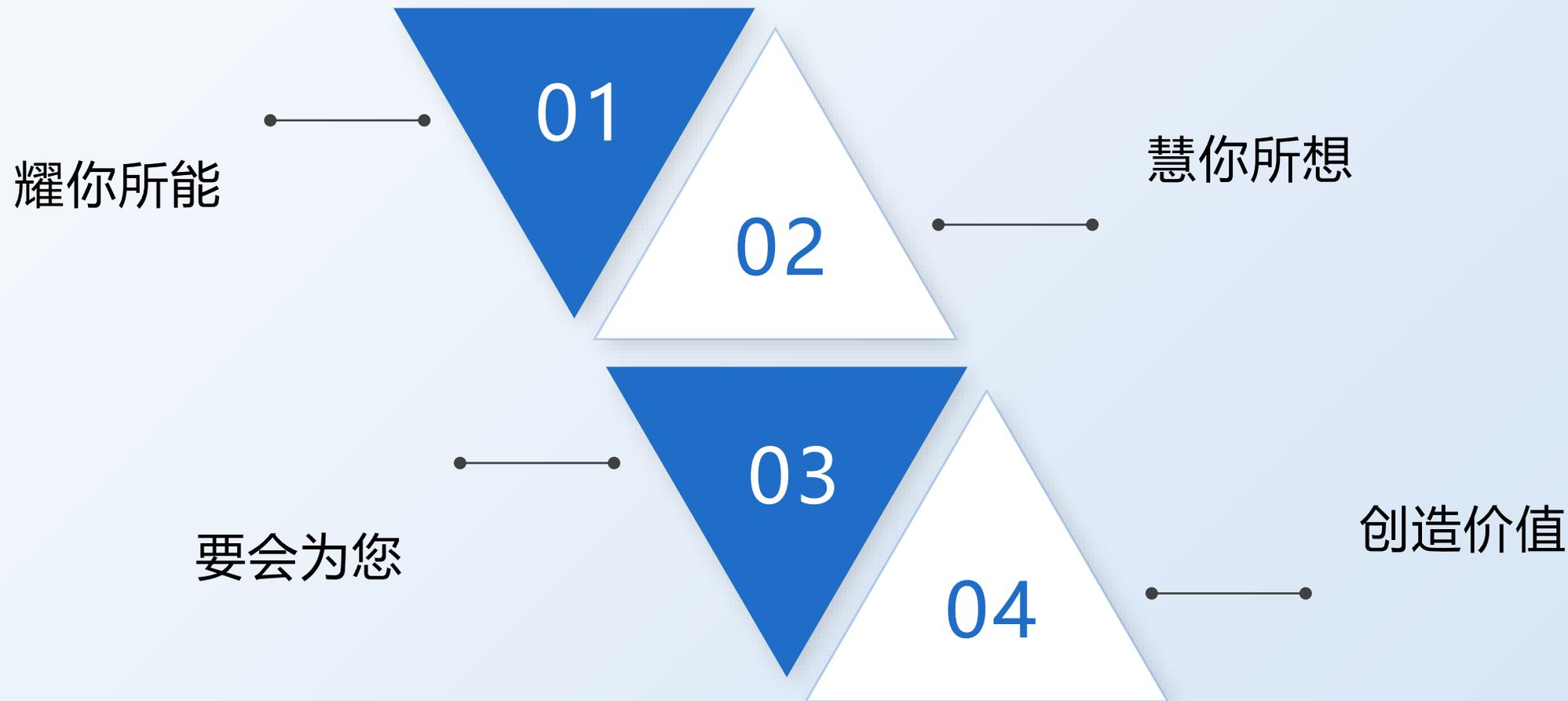


第一部分

公司理念

PART I  
CORPORATE PHILOSOPHY







02  
PART



第二部分

产品型号

PART II  
PRODUCT MODEL





## 5090 8卡 7U服务器

一款具有广泛用途的8卡GPU服务器,基于AMD SP5处理器平台,支持9004/9005系列CPU,支持8张600W GPU,具备性能卓越、扩展性强和可靠性高等特点,适用于人工智能、大模型推理、图形渲染、云游戏等应用场景。

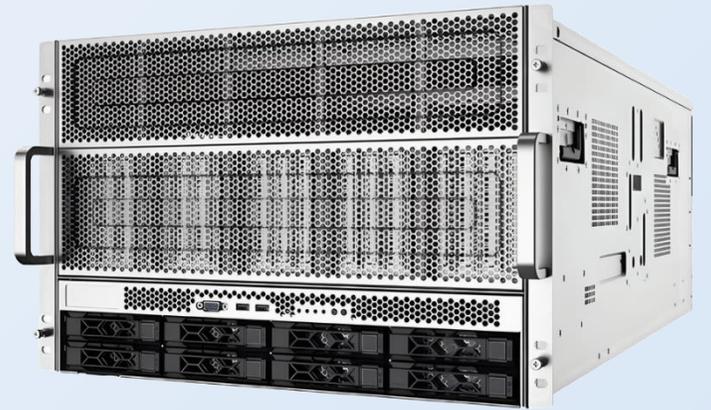
卓越性能 高性价比

兼容风扇卡 涡轮卡

灵活配置 按需选择

稳定可靠 智能管理

形态	标准7U 机架式
CPU	支持2颗 AMD EPYC™9004/9005系列处理器, 最大功率500W
内存	支持24个DDR5 内存插槽,1DPC,最高速率6400MT/s
RAID卡	可选支持12Gb/s SAS HBA 或RAID卡
GPU	支持8个600W GPU, 可选支持风扇卡或涡轮卡
PCIe扩展	最大支持11个标准PCIe5.0插槽 可选支持1个OCP 3.0网卡, PCIe5.0x8, 可选4x1GbE/ 2x10Gb SFP+/ 2x25Gb SFP28
本地存储	支持8个3.5" /2.5*SAS/SATA硬盘(可选支持2/4个U.2 NVMe SSD)
I/O端口	支持1个板载 M.2.(PCIe 3.0x4,2280&22110) 前置:2个USB 3.0, 1个VGA接口 后置:1个串口, 2个USB 3.0接口, 1个VGA接口, 1个RJ45 管理口
系统风扇	散热优化设计, CPU计算模块与 GPU 计算模块隔离散热, 风扇支持热插拔、支持智能调速
电源	可选配4个CRPS 电源, 支持2+2/3+1冗余; 可选配 6个CRPS 电源, 支持3+3冗余; 可选 2000W/2700W/3200W/3600W电源模块; 电源模块支持热插拔;
管理功能	集成BMC管理芯片 AST2600, 支持IPMI2.0.Redfish.SOL.KVM、虚拟媒介等功能 提供1个1Gbps RJ45专用管理口
安全功能	可选TPM 安全模块, 支持机箱开盖入侵检测、加锁机箱上盖板(免工具)
机箱尺寸	支持 BMC 双 flash冗余设计
温度	宽448mmx高306mmx深896mm 工作温度:5C-35°C 存储温度:-40°C-65°C
湿度	工作相对湿度:8% t90%(无冷凝) 存储相对湿度:5%to95%(无冷凝)
操作系统支持	支持Microsoft Windows Server, Red Hat Enterprise Linux,CentOS.Ubuntu等主流操作系统



## 5090 8卡 7U服务器

一款具有广泛用途的GPU服务器,支持第四代或第五代英特尔® 至强® 可扩展处理器,支持8张三风扇版显卡,具备性能卓越、扩展性强和可靠性高等特点,适用于人工智能、大模型推理、图形渲染、云游戏等多种应用场景。

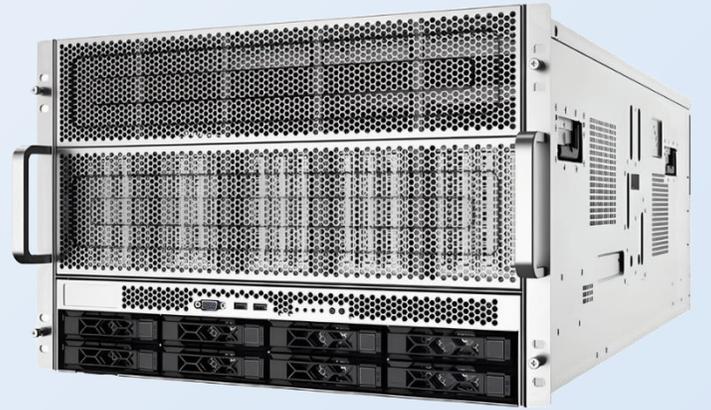
全新平台 卓越性能

原生支持 三风扇显卡

灵活配置 按需选择

稳定可靠 智能管理

形态	标准 7U机架式
CPU	支持2个第四代或第五代英特尔® 至强®可扩展处理器,最高支持350W
内存	支持32个 DDR5内存,最高5600MHz, RDIMM/3DS RDIMM
RAID卡	可选支持12Gb/s SAS HBA及12Gb/s SAS RAID 卡
GPU	支持8张 GeForce GPU卡(三风扇版, TDP 600W), 可选支持10张显卡
PCIe扩展	最大支持11个标准PCIe.5.0插槽 可选支持1个OCP 3.0网卡, PCIe5.0x8, 可选4x1GbE/ 2x10b SFP+/ 2x25Gb SFP28
本地存储	支持8个3.5"/2.5"SAS/SATA硬盘(可选支持2/4个U.2 NVMe SSD)
I/O端口	支持2个板载M.2(SATA 3.0/ PCIe 4.0 x4,2280&22110) 前置:2个USB 3.0, 1个VGA接口 后置:1个串口, 2个USB3.0接口, 1个VGA接口, 1个RJ45 管理口
系统风扇	散热优化设计,热插拔冗余风扇
电源	4个白金级 CRPS 电源模块 (2000W/2700W/3200W/3600W), 支持热插拔, 支持2+2/3+1冗余模式
管理功能	集成BMC管理芯片 AST2600,支持IPMI2.0、Redfish、SOL、KVM、虚拟媒介等功能 提供1个1Gbps RJ45专用管理口
安全功能	可选TPM/TCM安全模块, 机箱开盖入侵检测, 加锁机箱上盖板(免工具)
机箱尺寸	支持 BIOS/BMC 双 flash冗余设计 宽 448mm x高306mmx深886mm(含挂耳924mm)
温度	工作相对湿度:8% to 90%(无冷凝) 存储相对湿度:5%to 95%(无冷凝)
操作系统支持	支持 Microsoft Windows Server、 Red Hat Enterprise Linux、CentOS、Ubuntu等主流操作系统





## 4090 8卡 服务器

专为大规模模型训练、实时推理及高性能计算设计，以消费级硬件实现专业级性能突破。

卓越性能 高性价比

灵活配置 按需选择

稳定可靠 智能管理

形态	标准4U机架式
CPU	支持2颗AMD EPYC™7002/7003处理器，TDP 280W
内存	支持32个DDR4内存,3200MT/s, RDIMM/LRDIMM
RAID 卡	可选支持12Gb/s SAS HBA及12Gb/s SAS RAID卡
GPU	支持8张双宽GPU, TDP 450W
PCIe 扩展	最大支持8+2个标准PCIe4.0插槽 可选支持1个OCP3.0网卡, PCIe4.0 x8, 可选4x1GbE/ 2x10Gb SFP+/ 2x25Gb, SFP28
本地存储	支持8个3.5"/2.5"SAS/SATA硬盘(可选支持2/4个U.2 NVMe SSD) 支持1个板载M.2SSD, 2280&22110尺寸, PCIe2.0 x1
I/O 端口	前置:2个USB 3.0, 1个VGA接口 后置:1个串口, 2个USB3.0接口, 1个VGA接口, 1个RJ45管理口
系统风扇	8个8056高转速风扇, N+1冗余, 支持热插拔
电源	4个白金级CRPS电源模块(2000W/2700W), 支持热插拔, 支持2+2/3+1冗余模式
管理功能	集成BMC管理芯片AST2500, 支持IPMI2.0、SOL、KVM、虚拟媒介等功能 提供1个1Gbps RJ45专用管理口
安全功能	可选TPM/TCM安全模块, 机箱开盖入侵检测, 加锁机箱上盖板(免工具)
机箱尺寸	宽448mmx高175mmx深822mm(含挂耳842mm)
温度	工作温度:5°C-350°C 存储温度:-40C-65°C
湿度	工作相对湿度:8%to 90%(无冷凝) 存储相对湿度:5%to95%(无冷凝) 支持Microsoft
操作系统支持	Windows Server.Red Hat Enterprise Linux、CentOS、Ubuntu等主流操作系统 具体版本请咨询





## 4090 8卡 服务器

高效工作 降低负载

磁盘阵列 多重备份

远程管理 便携控制

ECC内存 稳定运行

可扩展性 领域应用

芯片组	Intel C621A芯片组 CPU TDP 支持≤270W 功耗
处理器	支持2颗IntelXeon3代处理器  可选Xeon 6330/8368Q处理器支持2颗IntelXeon3代处理器
内存	支持32个DDR43200MHz内存插槽  可扩展至8TB内存容量
硬盘	支持2*M.2接口, 12个3.5/2.5英寸SAS/SATA硬盘  标配企业级960G SATA读密集型
扩展槽	支持8个PCIE 4.0x16, 1个PCIE 3.0x4  支持8张双宽显卡
电源	可选2000W[2+2] CRPS标准电源  支持N+N冗余/N+1冗余
网口	支持1个0CP3.0网卡  可选配1/10/25/50/100/200/400GE PCIe标准网卡
尺寸	4U机架式服务器机箱  175mmx447mmX898mm(高*宽*深)
系统	支持windows、Linux(Ubuntu/CentOS...)等

可协助安装操作系统, 如需正版授权需额外购买, 详情请咨询

# AI智能

AI INTELLIGENCE





## H100 服务器

**H100: 适合常规AI训练、实时推理及中小规模模型。**

### 生成式AI

动态并行计算技术使  
Transformer模型训练速度提升  
9倍，推理速度提升30倍

### HPC

FP64算力达30 TFLOPS，适用  
于气象模拟、分子动力学等科学  
计算

### 实时推理

400G InfiniBand网络支持低延迟  
响应

### 高性价比

小型推理中使用

cpu 2\*Intel8558/  
2.1GHz/260M/48C/96T/330W/5600MHz

内存 32\*64GB/DDR5/4400或4800MHz/ECC/REG

SSD 2\*SSD/960GB/SATA 6Gb/2.5寸/读取型 <1DWPD

SSD 2\*SSD/7.68TB/2.5寸/U.2/NVME/PCIE4.0/读取型

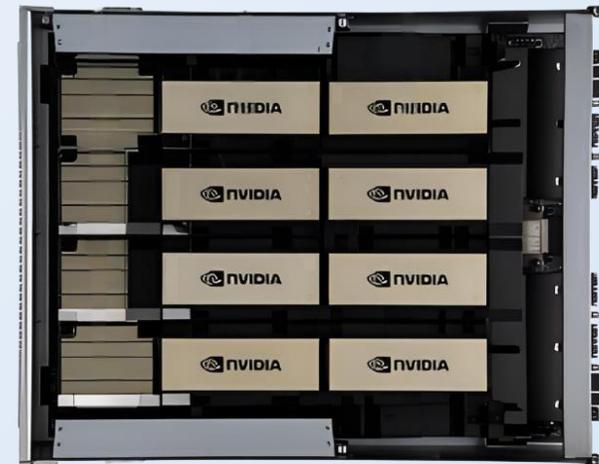
显卡 1\*HGX H100 80GB 8-GPU

IB网络 8\*IB网卡 Mellanox CX-7 NDR IB VPI 400Gb/s 单口OSFP

IB网络 1 X IB网卡 Mellanox CX-6 VPI HDR IB 200Gb/s 单口  
QSFP56 PCIe4.0x16 MCX653105A-HDAT

以太网网络 1 X 以太网卡 Mellanox CX-5 EN 25GbE 双口SFP28  
PCIe3.0 x8 MCX512A-ACAT

训练 支持671b,支持70b





## H200 服务器

**H200: 专为生成式AI、高性能计算（如科学模拟）及云服务设计。**

### 能耗优化

H200能耗比H100降低50%，支持更高负载运行

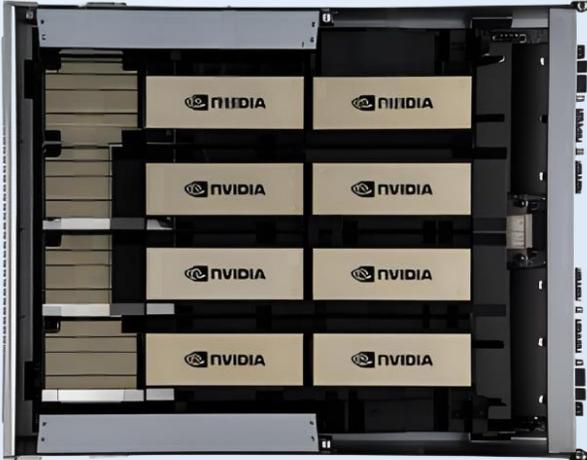
### 扩展性

支持多卡协同（单服务器最多256张卡），适用于大规模分布式训练

### 兼容性

与H100完全兼容，可无缝升级现有系统

- cpu 2\*Intel8558/  
2.1GHz/260M/48C/96T/330W/5600MHz
- 内存 32\*64GB/DDR5/4400或4800MHz/ECC/REG
- SSD 2\*SSD/960GB/SATA 6Gb/2.5寸/读取型 <1DWPD
- SSD 2\*SSD/7.68TB/2.5寸/U.2/NVME/PCIE4.0/读取型
- 显卡 1\*HGX H200 141GB 8-GPU
- IB网络 8\*IB网卡 Mellanox CX-7 NDR IB VPI 400Gb/s 单口OSFP
- IB网络 1 X IB网卡 Mellanox CX-6 VPI HDR IB 200Gb/s 单口QSFP56 PCIe4.0x16 MCX653105A-HDAT
- 以太网网络 1 X 以太网卡 Mellanox CX-5 EN 25GbE 双口SFP28 PCIe3.0 x8 MCX512A-ACAT
- 训练 支持671b
  
- cpu 2\*Intel8558/  
2.1GHz/260M/48C/96T/330W/5600MHz
- 内存 32\*64GB/DDR5/4400或4800MHz/ECC/REG
- SSD 2\*SSD/960GB/SATA 6Gb/2.5寸/读取型 <1DWPD
- SSD 2\*SSD/7.68TB/2.5寸/U.2/NVME/PCIE4.0/读取型
- 显卡 1\*HGX H20 141GB 8-GPU
- IB网络 4\*IB网卡 Mellanox CX-7 NDR IB VPI 400Gb/s 单口OSFP
- IB网络 1 X IB网卡 Mellanox CX-6 VPI HDR IB 200Gb/s 单口QSFP56 PCIe4.0x16 MCX653105A-HDAT
- 以太网网络 1 X 以太网卡 Mellanox CX-5 EN 25GbE 双口SFP28 PCIe3.0 x8 MCX512A-ACAT
- 训练 支持70b



## YH205KI



形态	2U AI服务器
CPU	支持1颗鲲鹏920
内存	4个DDR4内存插槽，最高3200MT/s
AI加速卡	最大支持3张Atlas 300I/VPro或4张Atlas 300I加速卡
AI算力	最大420 TOPS INT8
本地存储	(8~12)*3.5 SAS/SATA
RAID支持	RAID 1/5/6/10等
PCIe	最多4个PCIe 4.0x8标准扩展槽位
板载网卡	4*10GE/25GE(光口)+2*GE(电口)
电源	·2个550W或900W交流热插拔电源，支持AC 220V/DC 240V或者2个1200W直流热插拔电源，支持DC-48V ·支持1+1冗余
风扇	4个热拔插风扇，支持N+1冗余
环境条件	长期:5°C~50°C.短期:0°C~55°C
结构尺寸	86.1 mm * 447 mm * 475 mm



## YH210KI



形态	2U AI服务器
CPU	支持2颗 鲲鹏920
内存	32个DDR4内存插槽，最高2933 MT/s
AI加速卡	最大支持8张Atlas 300I/300V， Atlas 300I/ V Pro或4张Atlas 300I Duo
AI算力	最大1120 TOPS INT8
本地存储	·25*2.5 SAS/SATA ·12*3.5 SAS/SATA ·8*2.5 SAS/SATA+12x2.5 NVMe
RAID支持	RAID 0/1/10/5/50/6/60等
PCIe	最多支持9个PCIe4.0 接口
电源	2个热插拔900 W或2000W交流电源模块，支持1+1冗余备份
风扇	4个热拔插风扇，支持N+1冗余备份
环境温度	5°C~ 40°C
结构尺寸	447 mm * 790 mm*86.1 mm

# AI智能

AI INTELLIGENCE

### 最优能效

同等算力，更低能耗。能效比较业界产品提升15%

### 边云协同

云端训练，模型实时更新。云端统一管理

### 强劲算力

满足多个场景推理需求。支持384路高清视频实时分析

### 卓越算力

双处理器架构，搭载8张推理卡、或者2张训练加速卡。最高支持1024路高清视频实时分析

### 极致能效

全方位优化的系统散热设计，降低系统散热能耗;提供功率封顶和功率控制措施

## YH2715E



形态	2U机架式
CPU	1/2个第四代英特尔®至强®可扩展系列处理器，TDP 350W，可支持澜起津逮处理器
AI加速卡	最大支持8张Atlas 3001/300V，Atlas 3001/V Pro或4张Atlas 300I Duo
内存	32DIMM插槽,支持DDR5 RDIMM内存
AI算力	最大支持8个Atlas 300I/V Pro加速卡，最大1120TOPS INT8
本地存储	前置:8/12块3.5/2.5英寸SATA/SAS/NVMe硬盘; 后置:2/4块2.5英寸硬盘 前置:24块2.5英寸SATA/SAS/NVMe硬盘 后置:2/4块2.5英寸硬盘支持内置1*M.2
PCIe扩展槽	默认提供3个PCI-E5.0标准插槽，1个OCP3.0专用插槽
PCIe	最多支持9个PCIe4.0接口
电源	支持1+0单电源模式、1+1冗余电源模式;电源模块可选800W/1200W/1600W/2000W铂金交流电源模块
风扇	6个热插拔冗余风扇
物理尺寸	高87mm*宽438mm*深735mm

## YH4910E2



形态	4U 机架式
CPU	最大支持两颗第四代英特尔®至强®可扩展处理器，TDP最大350W
内存	最大支持10张全高全长双宽加速卡:支持昇腾Atlas 3001 Pro/Atlas 300V Pro/Atlas 300V/Atlas300I Duo加速卡
AI算力	默认12个3.5寸SAS/SATA/NVMe硬盘，可选8/16/24块2.5SAS/SATA/NVMe硬盘;板载支持2个M.2NVMe或M.2 SATA3硬盘:
本地存储	后置:13个PCI-E5.0x16;默认支持一个OCP3.0，可扩展至2个;
PCIe扩展槽	前置:4个PCI-E5.0x8;可选SAS RAID、HBA、HCA、HFI卡
电源	4个2000W&2700W&3000W N+1/N+N冗余电源模块，支持热插拔
风扇	N+1冗余重型风扇;电源模块标配独立风扇
物理尺寸	高889mm*宽447mm*深177mm

# AI智能

AI INTELLIGENCE

基于第四代英特尔®至强®可扩展系列处理器打造的AI服务器;“Intel®+昇腾”提供超强算力广泛应用于AI推理和训练场景中

### 超强性能

采用高性能低功耗的Ascend 310P 910B处理器，最大支持2颗第四代英特尔®至强®可扩展处理器，最大功率350W;

### 稳定品质

每CPU支持8个内存通道，总共支持32DIMM插槽;PCIe插槽全部支持5.0



03  
PART



第三部分  
GPU维保

PART III  
GPU M&R



## 一站式全栈智算算力服务

- 面向智能化场景，高性能、高性价比的算力服务
- 覆盖从定义、生产、建设到业务运营的全栈服务
- 多环境（私有化环境、互联网环境等）混合部署

### 1 算力定义

#### 目标系统性能

- GPU 数量
- 可用功率
- 训练/推理性能
- 集群策略

#### 可扩展性计划

- 容量弹性
- 系统性能伸缩性

#### 投资回报率

- 技术路线选择
- 价格趋势评估

### 2 生产/建设

#### 服务器定制

- 芯片选型
- GPU服务器定制生产
- 面向模型的硬件优化
- 信创服务器生产
- AI专用集群系统
- 浸没液冷专用服务器

#### 数据中心工程

- 高功率机柜
- 冷板液冷
- 浸没液冷
- 绿色能源方案

### 3 算力运营

#### AI+行业智算平台

- 金融 元宇宙
- 大健康 工业 教育
- AI 开发部署工具

#### 云管平台

- 公有云 混合云 超算云
- GPU池化 服务编排  
集群管理
- 跨IDC算力调度

#### 安全管理

- 云安全
- 数据保护
- 跨境数据通道



# 服务范围

SERVICE SCOPE



## 主动式运维服务

- 在线系统常规检查服务
- 在线技术支持服务
- 补丁与升级服务



## 综合培训服务 (可选)

- 技术指导服务
- 日常运维服务培训
- 员工IT技能普及培训

## 基础运维服务

- 备件更换服务
- 设备维修服务
- 7\*24H 远程问题诊断服务



## 专业技术服务 (可选)

- 系统巡检服务
- 组网及性能调优
- 现场备机备件服务



备注：带“可选”标志，服务需单独购买，不包含在标准维保里。



## 增值部分



### I. 原厂技术支持

- 提供原装备件和对应原厂技术指导

### I. 算力节点运维能力

- 网络实施阶段的联调、验证、调优服务，包括：设备和链路配置、网络拓扑调试、协议测试和验证、网络设备的性能调试和优化、UFM软件的安装和配置及功能验证等

## 维修部分



### I. 设备

- 备件采购通路限制，包括整机采购& MOQ要求、禁售限制，影响备件成本及SLA
- RMA通路限制，包括海外原厂支持限制、国内维修能力限制，同样影响成本及SLA

### I. 工具

- 维修工具，特别是诊断工具准备，影响维SLA及客户满意度

### I. 技术及经验

- 国内维保基本起步阶段，故障率数据积累不足
- 具备单Site 100+ 算力服务器MA支持经验



## GPU 计算检测



- 可识别性检测: HBM 显存检测
- PCIe 协商速率检测
- VPCIe MRRS 检测
- 卡间 GDR 状态检测
- 卡间 GDR 带宽检测
- 稳定性压力测试
- 重负载工况下核心温度检测;
- 重负载工况下功耗检测

## NVSwitch 交换芯片检测

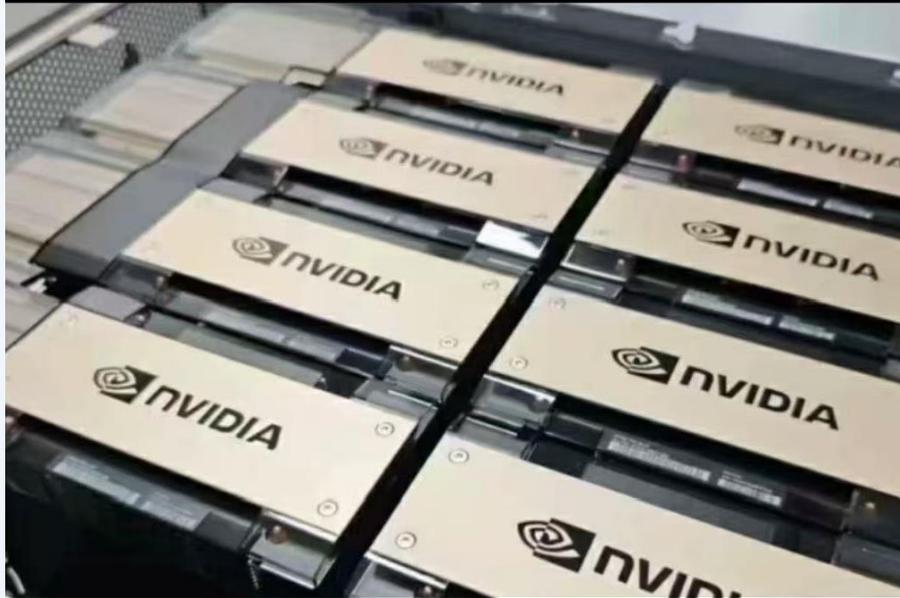


- 可识别性检测
- 工作状态检测
- NVLink 通道数量检测
- NVLink 互联带宽检测
- 重载下温度检测

## 维修故障



- GPU单卡芯片维修
- GPU模组基板维修
- 模组基板元器件维修
- NVSwitch交换芯片维修
- 液冷模组供电模块维修
- 等等...



维修技术文档 (图2)

硬件维修案例 (图1图3图4图5)



THANKS

谢谢观看

Thank you for your attention!

