



FROM TSINGHUA UNIVERSITY
CONTRIBUTOR TO SOCIETY

源于清华 服务社会



超强 A800I A2 推理服务器

超强 A800I A2 推理服务器采用8模组高效推理方式，提供强劲 AI 推理能力，在算力、内存带宽和互联能力方面具有优势，可广泛应用于生成式大模型推理，例如智能客服、文案生成、知识沉淀等内容生成类场景，支持业界领先的推理算力和高速 NPU 互联技术，提升大模型推理效率。



更高算力密度

4U 高度最高提供 业界领先算力

搭载8颗超强算力的昇腾 AI 模组，高效加速应用



极致能效比

提供超高能效比，节省运营费用



高速网络带宽

8*200G RoCE v2高速接口

8路 NPU 通过 HCCS 或 PCIe 高速互联

应用场景

部署在数据中心机房，使能中心推理



内容生成



智能问答



搜索NLP



内容审核



互联网推荐



智能客服

功能	产品规格
形态	4U AI 服务器
NPU	8 * 昇腾 AI 模组, 支持HCCS或PCIe高速互联
CPU	4颗鲲鹏920处理器, 支持64核 3.0GHz/48核 2.6GHz
HBM	512 GB/256 GB
内存	32个 DDR4 内存插槽, 最高 3200 MT/s 单根内存条容量支持 16 / 32 / 64 GB
本地存储	8 * 2.5 SATA + 2 * 2.5 NVMe 4 * 2.5 SATA + 6 * 2.5 NVMe, 支持U.2接口
RAID支持	支持 RAID 0 / 1 / 10 / 5 / 50 / 6 / 60
网络	灵活网卡, 4 * 25GE NPU 8 * 200 GE QSFP 接口直出, RoCE 协议
PCIe接口	最多支持3个 PCIe 4.0 扩展插槽
电源	4个热插拔 2.6 kW 电源模块, 支持2+2冗余
供电	200 ~ 240V AC 240V DC
散热方式	风冷
风扇	支持8个热插拔风扇模组, 支持 N+1 冗余
工作环境温度	5°C ~ 40°C
结构尺寸	175 mm * 447 mm * 790 mm

以业务目标为基础的数据中心解决方案

让您更充分地利用IT和企业解决方案, 帮助降低IT复杂性、削减成本并提高效率。我们利用您当前技术的同时将针对您的需求量身定制的创新技术无缝地融入进来, 以您应用的业务为基础, 为您构建高效, 敏捷, 开放, 智慧的数据中心软硬件一体解决方案, 简化您的IT, 解决您的业务难题, 并持续为您创造价值。有关详情, 请与您的软通计算销售代表联系。

