



超强算力 极致性能

KunLun G5680 V2 AI训练推理服务器



KunLun G5680 V2

KunLun G5680 V2 是一款4U4路AI服务器，是基于鲲鹏920与昇腾910芯片的AI训练推理服务器，AI算力可达3 PFLOPS FP16，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推理，适用于智慧城市、智慧金融、教育科研、运营商等需要大算力的行业领域。



超强算力密度

- 支持4颗鲲鹏920处理器
- 支持8颗昇腾910 AI处理器，4U高度最高提供3 PFLOPS FP16超强算力，0.79 PFLOPS FP32业界领先算力



超高能效

- 高效能、低功耗的AI算力基础设施
- 鲲鹏920芯片能效比优于业界标杆30%
- 提供3 PFLOPS/5.2 kW超高能效比



高速网络带宽

- 8*200G RoCE v2高速接口
- 8路NPU通过HCCS Full mesh互联，双向互联带宽达392GB/s



技术规格

形态	4U AI服务器
处理器	支持4颗鲲鹏920系列处理器（48/64核），CPU主频2.6GHz/3.0GHz
内存	最多32个DDR4内存插槽，支持RDIMM，内存速率最高3200 MT/s，最大内存容量2TB
NPU	支持8颗昇腾910 AI处理器
HBM	512GB，带宽1600GB/s；256GB，带宽800GB/s
AI算力	3 / 2.5 / 2.24 PFLOPS FP16，0.79 / 0.64 / 0.6 PFLOPS FP32
本地存储	8 * 2.5 SATA/SAS+2 * 2.5 NVMe 4 * 2.5 SATA/SAS+6 * 2.5 NVMe/SAS/SATA
RAID支持	可选配支持RAID 0、1、10、5、50、6、60
网络	8 * 200GE QSFP接口直出，RoCE协议
PCIe扩展	最多支持3个PCIe4.0扩展槽位
风扇	支持8个热拔插风扇模组，支持N+1冗余
管理	BMC集成1个专用管理GE网口，提供全面的故障诊断、自动化运维、硬件安全加固等管理特性 •BMC支持Redfish、SNMP、IPMI2.0等标准接口；提供基于HTML5/VNC KVM的远程管理界面；支持监测、诊断、配置、Agentless及远程控制等带外管理功能，简化管理复杂度 •可选配管理软件，提供无状态计算、OS批量部署、固件自动升级等高级管理特性，实现全生命周期智能化、自动化管理
电源	支持4个2600W、3000W AC白金/钛金电源模块，支持2+2冗余，支持热插拔（依据整机功耗和单电源功率配置）
操作系统	FusionOS、BCLinux、SUSE、Ubuntu、CentOS、统信、银河麒麟等操作系统
工作温度	5°C ~ 35°C（41°F ~ 95°F）（符合ASHRAE CLASS A1/A2）
尺寸(高×宽×深)	175 mm×447 mm×790 mm

河南昆仑技术有限公司


技术咨询电话：400-080-9000 技术支持邮箱：support@kunlunit.com

地址：河南省郑州市郑东新区龙子湖智慧岛中道东路时埂街北创智天地大厦10层

网址：www.kunlunit.com

版权所有 © 河南昆仑技术有限公司 2023。保留一切权利。
非经河南昆仑技术有限公司书面同意，任何单位和个人不得擅自摘抄、复制本手册内容的部分或全部，并不得以任何形式传播。

商标声明

 KunLun KunLun是河南昆仑技术有限公司的商标或者注册商标。在本手册中以及本手册描述的产品中，出现的其他商标、产品名称、服务名称以及公司名称，由其各自的所有人拥有。

免责声明

您购买的产品、服务或特性等应受河南昆仑技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，河南昆仑技术有限公司对本文档内容不做任何明示或默示的声明或保证。由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。