

宝德自强推理服务器 PR425KI G2



PR425KI G2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。

更高算力密度

- 4U高度，搭载8颗超强算力的AI模组，高效加速应用

极致能效比

- 提供0.6 PFLOPS/ kW超高能效比，节省运营费用

高速网络带宽

- 8*200GE RoCE v2高速接口
- 8路NPU通过HCCS Full mesh互联，双向互联带宽

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

产品规格

形态	4U AI服务器
NPU	8 * AI处理器 (具体性能请联系宝德客户经理)
CPU	4 * 鲲鹏920
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
RAID支持	支持 RAID 0 / 1 / 10 / 5 / 50 / 6 / 60
网络	8 * 200GE 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 ~ 240 V AC 240 V 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~35°C (41°F~95°F)
结构尺寸	175 mm (高) * 447 mm (宽) * 790 mm (深)