算工高性能AI存储

SG-NV7520

应用领域： 大语言模型训练、 AI推理服务、 计算机视觉等

算工SG-NV7520是专为AI而生的新一代高性能AI存储，加速大模型训练/推理。极致性能，创新数控分离架构，单框2400万IOPS，500GB/s带宽，训练集加载效率业界8倍，断点续训速度业界4倍。 广电五舟Pstor K700-A高性能AI存储支持灵活扩展，单集群支持512控Scale-out，EB级存储容量扩展，匹配万亿/十万亿参数多模态大模型平滑演进诉求。同时存储原生支持向量等全新数据范式，内置高性能向量检索引擎，加速大库容向量知识库检索，减少推理“幻觉”，一套存储满足AI训推全流程数据处理需求。

极致性能提升AI集群算力可用度

SG-NV7520通过NDS直通NPU，构建NPU直通数据存储的最优性能路径，绕过复杂协议栈，消除数据拷贝，性能提升40%。SG-NV7520新一代AI存储，通过DataTurbo高性能文件加速引擎，数据读写内核态直通存储，同时数据缓存基于内核态，加速内存高效流动，内存占用率减少50%。实现500GB/s单框大带宽，面对万亿参数大模型TB级Checkpoint实现分钟级读写，加速大模型断点续训。

多维扩展更灵活

SG-NV7520新一代AI存储基于全新硬件架构，支持Scale-out & Scale-up双向融合的弹性扩展能力，单集群支持最大512控扩展，EB级存储容量，兼顾高性能和高容量密度。同时支持阵列后端插DPU、GPU加速卡等多种算力卡，存储集群最大支持Scale-up 4096张卡，提供1.44万TFLOPS算力，加速数据处理，提供加密、压缩、向量检索等多种功能卸载。

向量新范式推理更智能

SPstor K700-A新一代AI存储从传统的NAS、对象到向量、图自适应检索引擎等全新数据范式，支持多模与海量知识高性能检索能力，构建基于RAG知识库和Unified Cache推理场景关键竞争力，知识库QPS性能领先业界3倍，生产系统AI Agent提供亿级大库容知识库存储，查询更精准。

产品参数

硬件架构	盘控一体
每框最大裸容量	983.04TB
每框高度	8U
每框控制器数	2
每框硬盘数	64
每框处理器	2*鲲鹏920
每框最大内存	1024GB
数据盘类型	Palm NVMe SSD
网络类型	25/100 Gb/s TCP/IP；25/100/200 Gb/s RoCE
关键特性	集群，配额，分级存储， 服务质量，多协议互通，端到端数据完整性校验（DIF）
机箱尺寸	352.8mm x 447mm x 952.2mm
工作环境温度	海拔-60~+1800m时的环境温度为5℃~35℃； 海拔1800m~3000m时，海拔每升高220m，环境温度（上限）降低1℃。
工作温度	10%~90%R.H.