

# 算工高性能AI存储

## SG-NV7510



### 应用领域：数据共享、大数据、备份与归档、中小数据中心等

算工SG-NV7510是专为AI而生的新一代高性能AI存储，提供极致性能，加速大模型训练/推理。单框可提供160GB/s带宽，400万IOPS。算工SG-NV7510高性能AI存储基于全新数控分离架构，支持32控灵活扩展，匹配万亿/十亿参数多模态大模型平滑演进诉求。支持AI原生向量、张量、KVCache等全新数据范式，可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”，提供KV Cache多级缓存能力，降低推理时延，提升推理性价比，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-NV7510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160GB/s单框大带宽，可提供超过TB级集群聚合带宽。SG-NV7510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30%。



### 多维扩展更灵活

SG-NV7510新一代AI存储基于全新硬件架构，2控起配，单集群支持36控扩展，可支持10PB级存储容量，兼顾高性能和高容量密度。Pstor K500-A新一代AI存储一套存储满足多样化生态需求，支持文件、对象、向量等多种协议，实现数据归集、预处理、模型训练、部署推理等AI大模型开发全流程，数据0拷贝加速大模型迭代开发。



### 原生AI数据范式推理更高效

SG-NV7510新一代AI存储支持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推理加速关键竞争力。提供内置多模态AI知识库，QPS性能业界领先，建库更高效，查询更精准。

## 产品参数

硬件架构	盘控一体
每框最大裸容量	1105.92TB
每框高度	2U
每框控制器数	2
每框硬盘数	36
每控处理器	2*鲲鹏920
每控最大内存	512G
数据盘类型	Palm NVMe SSD
网络类型	25 Gb/s TCP/IP; 25/100 Gb/s RoCE
关键特性	配额 (SmartQuota) , 服务质量 (SmartQoS) , 推理加速 (Unified Cache) , 端到端数据完整性校验 (DIF)
机箱尺寸	86.1mm x 447mm x 950mm
工作环境温度	海拔-60~+1800m时的环境温度为5°C~35°C (柜) /35°C (框) ; 海拔1800m~4000m时, 海拔每升高220m, 环境温度 (上限) 降低1°C。
工作湿度	10%~90%R.H