VAST + Lenovo

UNLEASH AI TRAINING AND INFERENCING

Artificial intelligence (AI) workloads, including Machine Learning, Deep Learning (DL), and Inference workloads present a new set of storage considerations requiring architects to rethink data center infrastructure. Modern DL training workloads such as Computer Vision and Natural Language Processing (NLP) predominantly rely on unstructured data, including video, images, and sound, that can only be optimized by training them on the largest data sets possible.

These data-intensive neural networks require performant flash infrastructure to randomly read across massive amounts of files. Unfortunately, legacy NAS solutions and parallel file systems built on spinning-disks fail to deliver the performance necessary for random data access required across the lifecycle of an AI application. And although all-flash configurations for these same storage systems do exist, they are cost-prohibitive at petabyte-plus scale, forcing organizations to implement multiple storage tiers, introducing additional data management complexity.

ENTER VAST DATA & LENOVO AI SOLUTIONS

VAST Data's Universal Storage and Lenovo SR670 V2 systems equipped with NVIDIA A100 Tensor Core GPUs provides a turnkey, high-performance all-flash solution for training and inferencing workloads across 100s of petabytes. Jointly built and tested by Lenovo and VAST Data, this solution delivers predictable performance to meet the needs of your applications in real time, across the entire AI lifecycle – from data ingestion to model development and validation before launching to production. Now, you can finally consolidate ALL your data on a single tier of affordable flash, thereby accelerating inference and training and unlocking previously impossible insights.

WHY VAST+LENOVO

UNPARALLELED PERFORMANCE AT SCALE

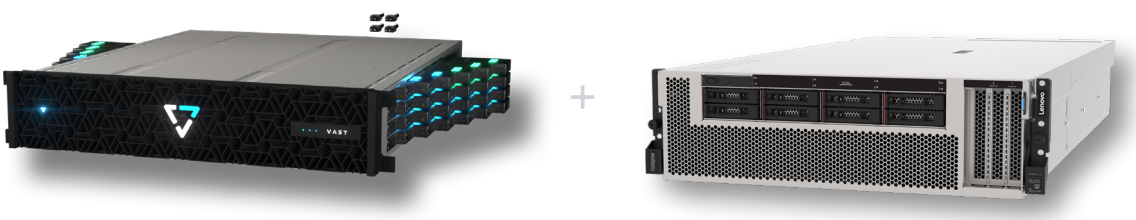
Scale to TB/s, millions of IOPS; 100GB/s+ per AI client over Ethernet or InfiniBand

OPERATIONAL SIMPLICITY

Turnkey Lenovo Scalable Infrastructure (LeSI) with VAST's all-flash Universal Storage NAS appliance delivers integrated environment with no client software dependency or complex management

ARCHIVE ECONOMICS

Radical flash storage economics to make flash affordable for all AI datasets at petabyte scale



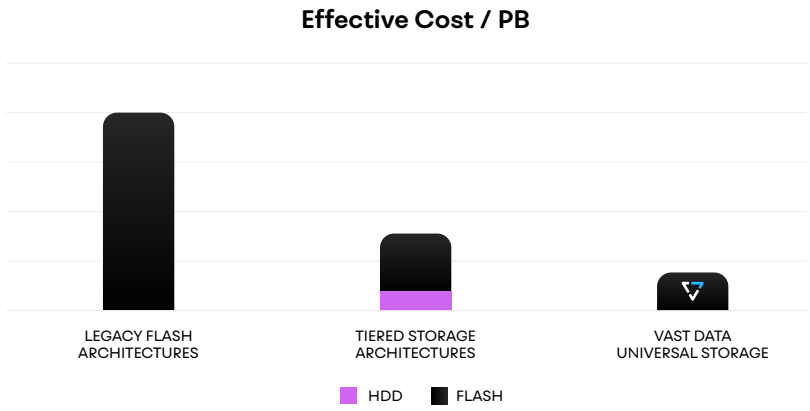
LENOVO SR670 SYSTEM

The Lenovo ThinkSystem SR670 V2 is a versatile GPU-rich 3U rack server that supports eight double-wide NVIDIA A100 Tensor Core GPUs, or the NVIDIA HGX A100 4-GPU offering with NVLink switch all-to-all GPU connection and Lenovo Neptune hybrid liquid-to-air cooling. The server is based on the new third-generation Intel Xeon Scalable processor family (formerly codenamed "Ice Lake") and the new Intel Optane Persistent Memory 200 Series. By using the processing power of the NVIDIA GPUs the server delivers optimal performance for machine learning (ML) and deep learning (DL) to extract greater insights and drive innovation across an array of industries. Also, the SR670 V2 employs Lenovo Neptune liquid-to-air (L2A) hybrid cooling technology to remove the heat of the HGX NVIDIA A100 GPUs through a unique closed loop liquid-to-air heat exchanger that delivers the benefits of liquid cooling such as higher density, lower power consumption, quiet operation, and higher performance without adding plumbing.

Dense & Flexible	3U Form Factor, up 8 double wide GPUs
Cooling with no datacenter changes	Hybrid Liquid-to-Air
NVLink	Supported
GPU Direct Storage (GDS)	Supported
NFS Multipath & RDMA	Supported

VAST DATA UNIVERSAL STORAGE

VAST Data, the pioneer of the Universal Storage concept, delivers all-flash storage systems for a wide range of workloads and use-cases. While built for high-performance, VAST eliminates the tradeoff between storage capacity and performance by delivering high-capacity and high-performance, at a price-point significantly lower than traditional all-flash storage systems.



Universal Storage brings flash TCO in line with tiered storage approaches to de-risk customer storage decisions all while freeing up HW capital for AI processors.

Our formula for compounded flash savings: QLC + 2.5% Erasure Codes + Similarity-Based Data Reduction

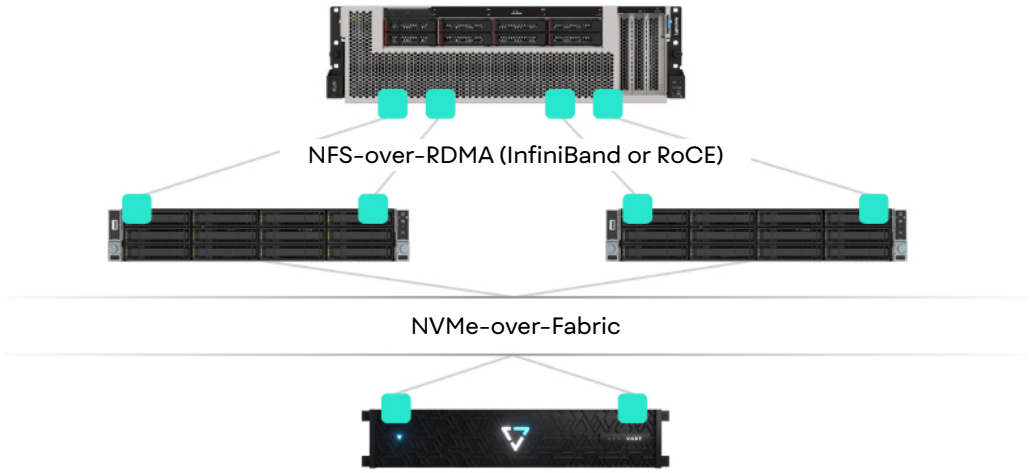
VAST Data's LightSpeed, both a product and storage philosophy for AI-based workloads, is optimized for both simplicity and speed. New VAST Data Universal Storage clusters built using LightSpeed enclosures deliver 2X the I/O throughput compared to previous enclosures. In addition, VAST supports several NFS client-side enhancements to achieve best-in-class throughput for AI applications and GPU servers without requiring complex parallel file systems.

NFS-over-RDMA

The Network File System (NFS) is a popular protocol used for accessing files remotely over networks and is widely used in enterprise environments for numerous applications because of its simplicity. In specialized environments, however, NFS has historically been deemed too slow for high-performance requirements. But by adding support for Remote Direct Memory Access (RDMA), NFS can now deliver greater efficiency of data transfers by bypassing client-side CPU and memory when leveraging InfiniBand or RDMA-over-Converged Ethernet (RoCE) networks. This results in higher throughput and lower latency while preserving the operational simplicity and familiarity with this industry standard storage protocol.

NFS Multipath

While NFS-over-RDMA optimizes throughput over a single client-side network interface, some applications, such as those running on GPUs, need more throughput than a single network connection can provide. By default, NFS uses one IP address and TCP connection for I/O limiting the total throughput of most NAS systems to ~2 GB/s. NFS multipath facilitates using multiple IP addresses and network connections for I/O, in effect aggregating different network links to provide access to a single file system while increasing total throughput.



NFS Multipath aggregates multiple SR670 v2 network interfaces to provide better remote storage throughput with built-in high-availability

NVIDIA GPUDirect Storage (GDS) Support

Eliminating potential system bottlenecks increases the performance and efficiency of AI workload environments. NVIDIA GPUDirect™ Storage (GDS), one of a family of technologies that are part of NVIDIA Magnum IO™, enables a direct path for direct memory access (DMA) transfers between storage and GPU memory without burdening CPUs. This results in dramatically lower CPU utilization within the GPU server, but also increased storage bandwidth and decreased latency. VAST Data supports GDS via LightSpeed v3.4+. Using NFS-over-RDMA as the storage protocol, GDS-enabled applications can now benefit from the increased performance between Lenovo SR670 v2 systems and VAST Data's all-flash storage.

WORKLOADS AND DATASET SIZE CONSIDERATIONS

An ideal AI Compute + Storage architecture should strike a good balance between cost and performance across a range of common deep learning model types, including Image Processing, Recommendation, and Natural Language Processing.

Storage Performance Level	Work Description	Dataset Size
Good	Natural Language Processing	Most all datasets fit in cache
Better	Image processing with Compressed images, ImageNet / ResNet-50	Many to most datasets can Fit within the local node's cache
Best	Training with 1080p, 4K, or uncompressed images, offline inference, ETL	Datasets are too large to fit into cache, massive first epoch I/O requirements, workflows that only read the dataset once

Therefore, storage infrastructure design typically implies tradeoffs for optimizing and balancing performance, capacity and cost. A wrong design might lead to a negative impact on training performance, as it affects the rate of data reading (and re-reading) from storage.

Traditionally, a storage and caching hierarchy is employed to counter the tradeoffs typically found across performance, capacity, and cost. The table below shows a typical hierarchy illustrating the capacity and performance tradeoffs common to a SR670-v2 HGX A100 system.

Storage Hierarchy Level	Technology	Total Capacity	Performance
RAM	DDR4	up to 4TB	> 200 GB/s per node
Local Cache	NVMe	up to 30TB	> 55 GB/s per node
Traditional External Storage (NFS)	Varies	Varies depending on specific needs	Varies, but typically ~2GB/s per node
VAST Data (NFS)	**NVMe**	**300TB to multi-PB+**	**> 80 GB/s per node**

Caching data in local RAM provides the best performance for reads. This caching is transparent once the data are read from a storage filesystem. However, the size of RAM is limited and less cost effective than other storage and memory technologies. Local NVMe storage is a more cost-effective way to provide caching close to the GPUs. However, manually replicating datasets to the local disk can be tedious and time consuming and is not a viable long-term storage location for datasets. While external storage solutions vary wildly in terms of performance, VAST Data's LightSpeed storage systems can deliver performance that meets or exceeds the local cache eliminating the need for complicated, and sometimes costly, caching strategies.

And with the industry's only Disaggregated and Shared Everything (DASE) architecture, VAST Data is capable of scaling both performance and capacity from day one, accommodating changing workload requirements much more feasibly.

SUMMARY

As AI becomes increasingly pervasive in our world - across all industries and verticals - organizations need a simpler way of providing modern IT infrastructure. Legacy storage architectures are no longer adequate in meeting the I/O performance needed to keep GPU equipped systems, such as the Lenovo ThinkSystem SR670 V2, busy.

VAST Data's innovative approach to all-flash storage blurs the performance lines between local cache and external storage. With key industry-first innovations, VAST Data overcomes traditional NAS appliance limitations to deliver the performance and capacity required of any AI project, without the need for complex parallel file system alternatives. The combination of Lenovo ThinkSystem SR670 V2 and VAST Data´s solutions provide a flexible scale-out compute and storage architecture that reduces infrastructure overhead, eliminates data tiering complexity, and lowers overall TCO while simultaneously improving performance and streamlining data management.

RESOURCES

Explore the Lenovo HPC and AI Innovation and Briefing Center
Lenovo Validated Design for AI Infrastructure on ThinkSystem Servers
Learn more about VAST Data's Universal Storage for Deep Learning & Machine Learning Workloads

© 2022 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.