VAST + Mellanox TECHNOLOGIES

MELLANOX AND VAST CONQUER STORAGE CHALLENGES IN THE MACHINE LEARNING ERA OF COMPUTING

Universal Storage and Ethernet Storage Fabric Overcome Vast Amounts of Data and Put an End to the Storage Pyramid

THE CHALLENGE

The introduction of flash to data centers a decade ago has forced storage administrators to trade-off the performance of all-flash storage against the lower cost of disk based secondary storage systems. These compromises force most organization into the familiar pyramid of storage tiers where active data is on an all flash array while less active data is relegated to slower, less expensive hard disk drives, cloud or even tape. Today's artificial intelligence (AI) and machine learning (ML), algorithms access storage very differently than the enterprise applications the pyramid was created to support. These systems use very expensive CPUs and GPUs to train on vast amounts of data including

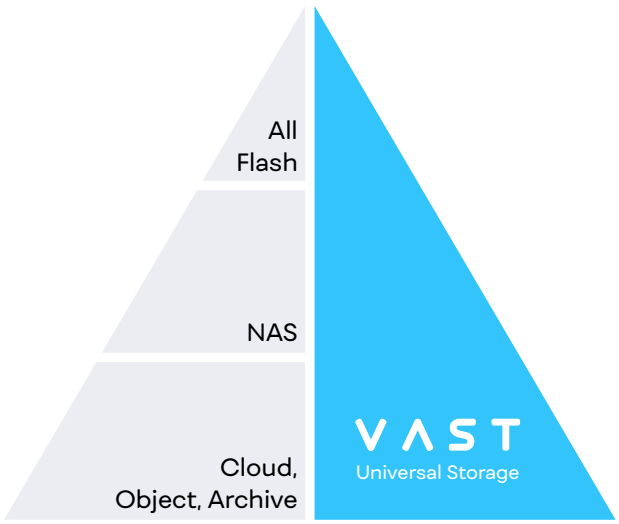
HIGHLIGHTS

- The fastest storage meets the fastest network, providing an answer for the most demanding applications
- Eliminating the storage tiering pyramid with a single tier of economical flash also eliminates worries about where data resides
- RDMA offloads boost NFSoRDMA performance to 18.7 GB/s per client
- Enables remote NVMe devices to be accessed with direct attached performance
- Up to Exabyte capacity with linear performance scalability
- Non-blocking, line-rate performance with dynamically shared buffers delivers extreme packet processing
- Simple, shared infrastructure with tenant-level dedicated performance

older data. This breaks the basic assumption the pyramid is based on, that there is less active data to relegate to lower tiers. AL/ML workloads require high performance access to petabytes, or exabytes, of data. These workloads demand an entirely new storage architecture to address the modern deep learning, training & inference era of computing. This new architecture must provide a single tier of storage, that scales to exabytes, provides the performance to keep the CPUs and GPUs fed and has to compete with spinning disks for affordability. Of course the gains of this new storage model would be all but lost without a high-performance, low-latency storage fabric that allows the storage architecture to take advantage of the speed of flash while offing linear scalability.

THE PYRAMID IS DEAD

Data scientists have always struggled to balance I/O performance with the quantity of data ingested by inference and training applications. They've been forced by economics to tier their data across a complex pyramid of storage systems each designed for either fast access or inexpensive capacity. Since it's impossible to analyze data on slow capacity-oriented storage systems, by tiering data off to capacity- oriented storage, organizations lose out on additional opportunities to find correlations in the data they've tiered down, correlations that could be used for a competitive advantage. VAST Data's Universal Storage system provides a single tier of all solid-state storage that is fast enough for the most demanding applications, scales to exabytes, and is affordable enough to compete with spinning disks, even for inactive data. With costs comparable to HDD, there is no longer a need for tiering of data.

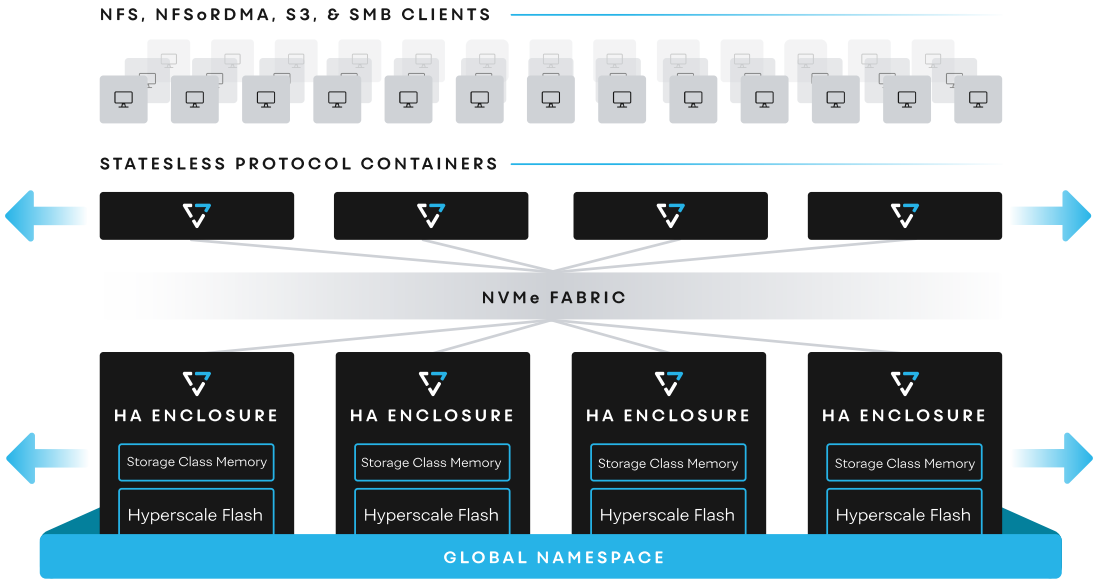


VAST Data's Universal Storage system ends the storage pyramid.

MAXIMIZING EFFICIENCY AND SCALE

Eliminate the trade-off of multiple storage tiers and moving to an all solid-state solution is the idea behind VAST's new Disaggregated, Shared Everything (DASE) storage architecture.

DASE breaks free from the idea that scalable storage needs to be built from non-shared data clusters. VAST uses 3D X-Point as a metadata store and write buffer for performance. The cost is kept down by using QLC flash for capacity and NVMe over Fabric (NVMe-oF) to connect it all together. The VAST storage system is cost competitive with HDDs, while providing terabytes per second and millions of IOPS of performance over NFS for files and/or the S3 API for object storage. The DASE architecture connects VAST Servers to the SSDs in VAST Enclosures via NVMe-oF. Servers can scale independently of capacity without the east-west internode cross-talk that is often challenging within a shared-nothing architecture. VAST servers can be containerized and embedded into application servers to bring NVMe over fabric performance to every host. The VAST DASE enables a storage system with 80% less cost than traditional enterprise flash memory and allows for all data to be available for fast access, all the time. As a result, mining data for new insights becomes ultimately simple and affordable.



Disaggregated, Shared Everything (DASE) Storage Architecture.

BOOSTING NETWORK PERFORMANCE

We've discussed that the adoption of AI and ML requires enormous amounts of data for proper training techniques. Likewise, this data must also be fed to farms of GPUs with maximum throughput, both to access shared data and in transferring this data to GPUs. This requires a network that can handle fast and efficient data delivery. Due to massive amount of distributed processing needed for AI/ML, an efficient and high-performant networking solution is required. The Mellanox Spectrum Ethernet switches and ConnectX® adapters

bring great performance into these environments with 25, 50 or 100GE speeds ensuring an industry leading end-to-end, high bandwidth, low latency Ethernet fabric capable of supporting the enormous demand placed on the network infrastructure by computational computing.

A PROPER ETHERNET OR INFINIBAND FABRIC IS ESSENTIAL FOR PERFORMANCE

AI applications are used on the training of computers, which require complex computations and fast and efficient data delivery. This is accomplished through the use of fast, efficient and light weight protocols that can streamline the communication and delivery process. Mellanox's heritage has been in meeting these needs, and they are therefore uniquely positioned to provide end-to-end solutions for organizations seeking a competitive advantage for their data. Mellanox enables smart offloading such as RDMA and GPUDirect®, and in-network computing capabilities to dramatically improve computer networks that assist in providing the performance attributes required to power AI workloads. The ConnectX® family of RDMA enabled adapters provide the low latency communication for VAST's newly architected NVMe-oF based backend, which has been disaggregated from the servers and moved to a resilient, shared storage enclosure. NVMe-oF is used to speed up throughput and lower latency and achieves near local performance for the remote enclosure.

AI workloads require intensive processing and without the proper network adapter, CPU cycles can be wasted and applications choked. Mellanox ConnectX adapters enable near-native performance through stateless offloads while consuming very little to no CPU cycles. Extended hardware resources which support 64 physical functions (PF) and 512 virtual functions (VF), along with SR-IOV help isolate PCI Express resources while RoCE can be used to offload network functions from the CPU to the network adapter. This alleviates CPU loads by implementing an intelligent network that can ease CPU strain, increase network bandwidth and enable scale and efficiencies.

VAST Data leverages the ConnectX adapter's ROCE offload features to bypass the bottlenecks TCP creates

	NFS over TCP	NFS over RDMA
Single client, single mount	2.1 GB/s	8.7 GB/s
Single client, multiple mounts	8 GB/s	20.5 GB/s

on high bandwidth networks. Using NFSoRDMA boosts throughput significantly without the need for user agents or kernel patches.

Mellanox Spectrum Ethernet switches offer a scalable and efficient Ethernet fabric. The underlying Mellanox Spectrum hardware delivers ultra-low latency, and non-blocking line-rate speed, while delivering packet processing with buffer fairness through a single shared buffer. Ensurng each port has the same fast access to the buffer eliminates the need for port mapping –which greatly simplifies deployments. In a training and inferencing algorithm environment such as AI and ML, fluid resource pools greatly benefit from fair load balancing, ensuring fast access to data no matter where it resides. In addition, the use of overlay technologies can greatly increase efficiencies allowing for a much faster and dynamic approach compared to a conventional network architecture. As a result, Mellanox Spectrum switches deliver an optimal and predictable network performance.

> Zebra is transforming patient care and radiology with the power of AI. To achieve our mission, our GPU infrastructure needs high-speed accelerated file access to shared storage that is faster than what traditional scale out file systems can deliver. That said - we're also a fast-growing company and we don't have the resources to become HPC storage technicians. VAST provides Zebra a solution to all of our A.I. storage challenges by delivering performance superior to what is possible with traditional NAS while also providing asimple, scalable appliance that requires no effort to deploy and manage.
>
> zebra MEDICAL VISION
>
> **Eyal Toledano, CTO**

CONCLUSION

The VAST DASE architecture and Universal Storage concept lays the foundation for dramatic efficiency gains at every level. Mellanox Ethernet high-performance and low latency interconnects are the backbone. The combination allows for easier administration with less data center impact when everything is stored on high-density flash. With the proper network in place for the data to traverse, pipelines run faster thanks to the most efficient data movement via offloads, accelerators and RDMA-enabled capabilities. All this contributes to the most efficient handling of complex computations and fast and efficient data delivery that enable world leading AI platforms to unleash life science insights.