White Paper

Choosing the Right Storage for Modern Workloads

From big-data analytics to artificial intelligence, the most exciting applications in many data centers today glean information from large pools of unstructured data and therefore require new scale-out file storage solutions. This report compares the most common approaches to providing large, high-performance file storage with representatives of each.

Storage for Today's Applications

As organizations focus their IT efforts on extracting value from ever-larger datasets, they soon discover that these new applications operate very differently compared to legacy applications. Enterprise data centers have traditionally provided a small amount of high-performance block storage for VMs and databases, and file storage as a lower performance tier for archives and user home directories.

From time-series trade analysis on Wall Street, to genomics, to GPUs training their deep learning model over millions of photos, these applications need fast, shared file storage and they need it to scale to multiple petabytes.

As organizations focus their IT efforts on extracting value from ever-larger datasets, they soon discover that these new applications operate very differently compared to legacy applications.

File Storage Evolves with Applications/Workloads

By the mid-'90s the corporate IT world settled on the dual-controller NAS appliances that still dominate corporate data centers today as the standard file storage solution. While dual-controller systems addressed many applications' needs, users in the petroleum, entertainment, and scientific communities found themselves needing bigger, faster file storage.

Scale-out file storage systems allow these users to scale their storage systems by adding additional compute power along with additional capacity, thereby scaling to capacities and performance levels well beyond dual-controller architecture's capabilities. Over time, the market has settled on two primary architectures for scaleout file systems: shared-nothing NAS and parallel file systems.

Shared-Nothing NAS

The shared-nothing model, exemplified by Dell EMC's PowerScale, formerly Isilon, combines the controller, typically an x86 server, and storage media into a single basic building block, the node. Shared-nothing systems protect data by replicating or erasure-coding across multiple nodes in a cluster. Shared-nothing systems combine compute, and capacity into similar nodes requiring users to scale compute performance and capacity together even when they need one or the other.

Shared-nothing storage systems have gained traction in applications from media and entertainment to genomic analysis and AI by providing the capacity these users need and a simple administration model. Shared-nothing systems create large volumes of network traffic between the nodes of a cluster, which ultimately limits cluster sizes.

Parallel File Systems

Where shared-nothing NAS developed as commercial products, software-defined parallel file systems can trace their origins to the research labs and HPC centers where high-performance file services are at the center of simulations from weather to "things we could tell you, but we'd have to kill you." Parallel file systems run on standard x86 servers using a variety of data protection models including shared-nothing, SAN, and software RAID shared-media models.

Parallel file systems evolved to support scientific applications with large numbers of simultaneous writers to a single file, something the NFS implementations of the day did not handle well. A POSIX file system agent, installed in all the servers that will access the parallel file system, provides the synchronous access and locking that parallel users require. Traditional NAS protocols are included, but are frequently lower performance than the POSIX client.

Parallel file systems like IBM's Spectrum Scale (GPFS) give storage architects enormous flexibility designing a system using hard drives, SSDs, shared-nothing, and shared-media storage. That flexibility creates complexity both making hundreds of decisions during the design phase and running a one-of-a-kind storage system that lacks the fit and finish of polished storage appliances.

Disaggregated, Shared Everything

The latest development in scale-out storage system design is the Disaggregated, Shared Everything (DASE) architecture used in the VAST Data Platform. DASE disaggregates the compute power and controller function of a scale-out NAS, which runs in stateless, containerized, frontend servers, from the Storage Class Memory and Hyperscale Flash SSD media in highly available containers.

This disaggregation is enabled by the NVMe fabric that connects all the front-end servers and all the enclosures in a cluster with just a few microseconds latency. This allows users to scale the performance of a cluster by adding and removing front-end servers from pools independently from the cluster's capacity, managed by adding storage enclosures.

Solution Comparison

The products selected for this comparison are the leading representatives of the scale-out storage architectures used for modern workloads.

Dell EMC's PowerScale, formerly Isilon, is a 20-year-old shared-nothing, scale-out NAS that supports all-flash, HDD/flash hybrid, and all HDD nodes in heterogeneous clusters. As one would expect from a 20-year-old product, Isilon is feature-rich, including automated tiering between Isilon pools and to the cloud.

IBM's Spectrum Scale, formerly GPFS, is a parallel file system first released in 1998. Spectrum Scale provides many configuration and operations options that make planning for and installing Spectrum Scale complex, requiring a high level of expertise.

Pure Storage's Flashblade and Flashblade//S repackage the shared-nothing, scale-out model into a blade server like chassis with 10 or 15 blades per chassis. Flashblade supports the S3 object protocol in addition to NFS and SMB, but files and objects occupy independent namespaces.

	DELL EMC PowerScale	IBM Spectrum Scale (GPFS)	PURE STORAGE Flashblade Flashblade//S	VAST Data Platform
Class	Scale-Out NAS	HPC File system	Scale-Out NAS/Object	Scale-Out NAS/Object
Application Focus	Media, Life Sciences	HPC	AI, Backup, Media	HPC, Media, AI, Life Sciences, Backup
Architecture	Shared-nothing	Shared-nothing, shared-media, SAN	Shared-nothing	DASE
POSIX Client	✗	✓	✗	✗
Scale Min-Max Effective	72TB-185PB RAW	48TB-8EB	123TB-19.2PB	500TB-EB
HDD/Hybrid	✓ / ✓	✓ / ✓	✗ / ✗	✗ / ✗
All Flash	✓	✓	✓	✓
Low-Cost Hyperscale Flash support	✓	✗	✓	✓
Data Protection	N+1-N+4 (Most customers N+2)	3 & 4 Way Mirror, N+2, N+3	N+2+	N+4
Data Protection Overhead	11-25%	20-75%	35-50%	3-20%
Storage Pools	One Per Node Type Minimum	Several	One	One
Asymmetric Scaling (multiple node types per pool)	✗	✓	✗	✓
Symmetric Scale-Out (similar node type per pool)	✓	✓	✓	✓
Internal (inter-pool) Tiering	Scheduled	Scheduled	✗	✗
Tier to External Storage	✓	✓	✓	✗
Deployment				
On-Premises	✓	✓	✓	✓
In-Cloud	✓	✓	✗	✓
Snap To Cloud	✗	✗	✗	✓
Tier to Cloud	✓	✓	By File	✗
Protocol Support				
NFS	✓	✓	✓	✓
SMB	✓	✓	✓	✓
S3	✓	✓	✓	✓
Multi-Protocol SMB <> NFS	✓	✓	✓	✓
Multi-Protocol File <> Object	✓	✓	✗	✓
Data Reduction				
Duplicate Block Elimination	Inline on Current Models	✗	✗	Inline, Always on
Single Block Compression	Inline on Current Models	✓	✓	Inline, Always on
Global, Cross-Block Compression	✗	✗	✗	Inline, Always on
All Flash Cost	$$$$	$$-$$$$	$$	$

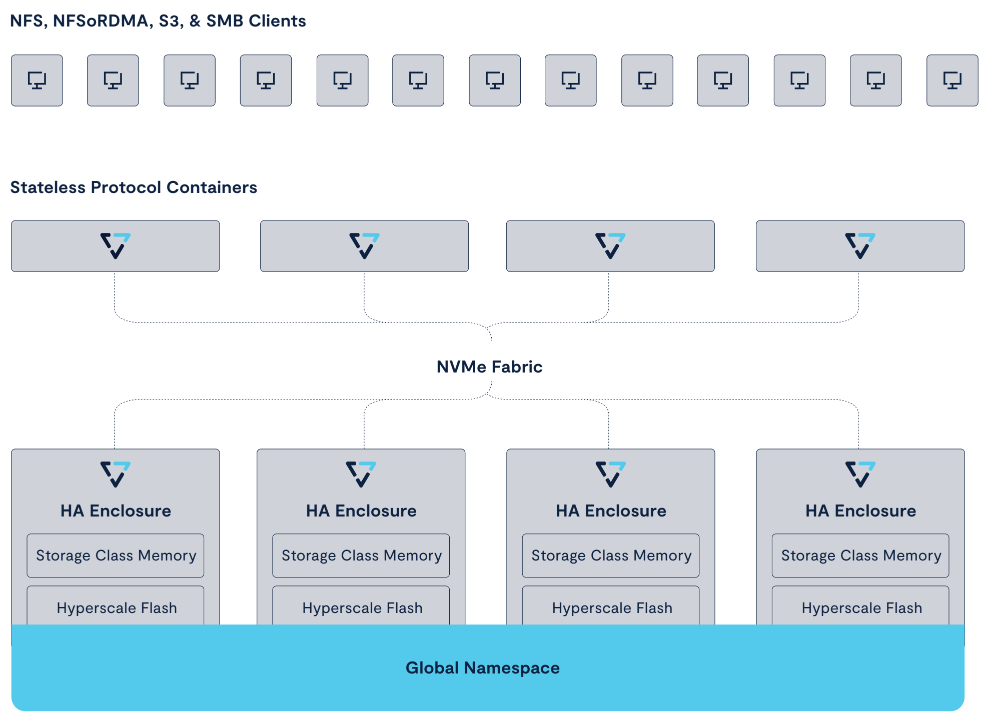
A Brief Look at the VAST Data Platform

VAST's Data Platform reduces the cost of all-flash, file, and object storage to make it affordable for all of an organization's applications, not just a privileged few. VAST is based on the DASE architecture of containerized, stateless protocol servers, connected over NVMe-oF to highly available storage enclosures filled with Storage Class Memory and low-cost Hyperscale Flash SSDs.

VAST software running in the protocol servers includes several innovations to minimize flash wear and maximize efficiency:

- Locally decodable erasure codes provide N+4 data protection with as little as 3% overhead
- Similarity data reduction yields greater reduction than any other storage system
- Global flash translation using Storage Class Memory write buffer minimizes flash wear

The result is an all-flash file and object system that's fast enough for the most demanding AI or media workloads, scales from petabytes to exabytes, and costs no more than a flash/disk hybrid solution based on shared-nothing or parallel filesystem technologies.



For more information on the VAST Data Platform and how it can help you solve your application problems, reach out to us at hello@vastdata.com.