Hewlett Packard Enterprise

Business white paper

Check if the document is available in the language of your choice.

SPEND LESS ON HPC/AI STORAGE AND MORE ON CPU/GPU COMPUTE

Accelerate your digital transformation with parallel file storage from the leading high-performance computing vendor[1]

[1] Hyperion Research, HPE Market Update for HPC User Forum, September 2021

TABLE OF CONTENTS

EXECUTIVE SUMMARY

According to Gartner, the best-run organizations are constantly searching for opportunities to realize efficiencies in the Run category of Gartner's Run—Grow—Transform model in order to shift budgets toward accelerating digital transformation.

If you are looking to realize cost efficiencies or prevent a looming cost explosion, one opportunity to examine is fast file storage designed to feed unstructured data to high-performance clusters of GPU- and/or CPU-powered rack servers built on HPE Apollo 2000 systems and HPE Apollo 6500 systems or HPE ProLiant DL rack servers.

The long list of typical enterprise workloads running on these high-performance clusters includes training of machine learning, computer-aided engineering, crash test simulations, wind tunnel simulations, automated driving systems, digital twins, mechanical design, chemical engineering, electronic design automation, precision medicine, genome sequencing, drug discovery, cryogenic electron microscopy, fraud and anomaly protection, quantitative pretrade analytics, algorithmic trading, cybersecurity, seismic exploration, reservoir simulation, and many more.

All these workloads have one thing in common: they all require massive data movement of massive data sets during processing. For that, they all depend on low-latency, high-speed 200 Gbps networks such as HPE Slingshot, InfiniBand, or Gigabit Ethernet to connect compute nodes with each other and with shared high-speed storage.

The convergence of computational modeling and simulation (often referred to as **classic high-performance computing, or classic HPC**) and artificial intelligence (AI) is changing everything. The impact of this convergence is especially disruptive on the HPC storage layer, as the I/O profiles of both methods could not be more different. The storage architectures that have served us well in the past, when simulation and AI were deployed on separate infrastructure stacks, are breaking architecturally and economically in the new era.

The costs of staying with legacy storage architectures in the new era add up quickly:

- **Delayed new product introduction and development** due to long job run times in business-critical R&D business processes
- **Low asset utilization** of expensive CPU/GPU compute infrastructure due to I/O bottlenecks that keep the compute nodes idling waiting for their data
- **Regrettable attrition of top talent** such as key engineers and data scientists due to frustration over job pipeline congestion that prevents them from working productively
- **Frequent, unbudgeted spending requests** due to ad hoc upgrades of the file storage infrastructure needed to cope with the architectural deficiencies—at the expense of CPU/GPU compute budgets

As the market share leader in HPC servers,[2] Hewlett Packard Enterprise saw the convergence of classic modeling and simulation with AI methods such as machine learning and deep learning coming and now offers you a new portfolio of parallel HPC/AI storage systems that are purpose engineered to address all the previously mentioned challenges—in a cost-effective way.

Interested in learning about the three approaches that can help you to feed your CPU- and GPU-accelerated compute nodes without I/O bottlenecks while creating efficiencies in Gartner's Run category? Keep reading.

[2] Hyperion Research, HPE Market Update for HPC User Forum, September 2021

A NEW EXASCALE ERA REQUIRES NEW STORAGE FOR SYSTEMS OF INNOVATION AND INSIGHT

Advanced computing is transforming. Explosive data growth that by far exceeds price or performance improvements in underlying storage media technologies—coupled with the convergence of modeling and simulation and AI workloads—has created new requirements for compute, software, networking, and storage. Called the exascale era, it's far more than a speed milestone. It's a technology inflection point characterized by a new set of capabilities for converged workloads running in mission- or business-critical workflows. And it affects every industry and field of inquiry.

In addition to constantly increasing model sizes and data sets, these new realities put new demands on high-performance storage. It is happening right now. For example, a recent study of the independent analyst firm Intersect360 found out that the majority (63%) of the HPC users today already are running machine learning programs.[3] And an additional 7% of the respondents stated that they plan to do so until the end of this year.

Storage technology that worked for petascale era workloads cannot power this new era's converged workflows because the I/O patterns and the characteristics of converged simulation and AI workloads could not be more different. As a result, keeping your current storage infrastructure will leave you unable to keep up—both in terms of performance and budget.

TABLE 1. Comparison of traditional petascale era infrastructures

Characteristics	Traditional HPC cluster	Traditional AI POD
Primary workload	Modeling and simulation	Machine learning/deep learning
Compute node type	CPU nodes such as HPE Apollo 2000 Gen10 Plus systems or Dell PowerEdge rack servers	GPU-accelerated nodes such as HPE Apollo 6500 Gen10 Plus systems or NVIDIA DGX A100 systems
Number of compute nodes	Hundreds to thousands	A few to tens
Typical interconnect	InfiniBand	Gigabit Ethernet
Typical I/O intensity	Write intensive	Read intensive
Typical storage	HDD-based parallel storage such as Cray ClusterStor L300, DDN EXAScaler, or DDN GRIDScaler	All-flash enterprise file storage such as Dell EMC Isilon F-Series, NetApp AFF, or Pure Storage FlashBlade
Capacity measured in	Petabytes	Terabytes
Scalability in single file system under heavy load	Up to hundreds of petabytes	Up to single digit petabytes
Ideal uses	• Mainly writing • Mainly large files • Mainly sequential order • Speeds up to double or triple digits GB/s	• Reading and writing • Files of all sizes • Sequential and random order • Speeds up to low double digits gigabytes per second
Price per terabyte	Low	High

[3] "HPC User Budget Map Survey: Machine Learning's Impact on HPC Environments," Intersect360, March 2020

It seems that in the new converged era, users are oftentimes stuck between a rock and a hard place:

- If you continue with traditional HPC storage systems, most likely you will experience I/O bottlenecks for AI/machine learning workloads because traditional HDD-based HPC storage is not well suited to serve the large number of files of all sizes that machine learning needs to read in the training phase. That can lead to job pipeline congestion, missed deadlines, unsatisfied data scientists, and constant escalations.
- If you try to scale traditional AI storage to the multi-petabyte requirements of the converged workloads, you most likely will experience scalability issues and exploding storage costs.

Therefore, the new exascale era really needs new HPC storage.

THE HPE APPROACH

HPE saw this coming. And that is why we have developed two new HPC/AI storage systems that are:

- Nearly as cost-effective and scalable as traditional HPC storage
- Clearly well suited to serve files of all sizes in any access patterns such as all-flash enterprise file storage does today
- Ultimately not requiring additional licensing of the file system by terabyte capacity or per storage drive (HDD and/or SSD)

These characteristics tie directly to these HPE products:

- HPE Parallel File System Storage that embeds the leading parallel file system in the enterprise—IBM Spectrum Scale (AKA GPFS)[4]
- Cray ClusterStor E1000 Storage System that embeds the leading parallel file system for scientific use cases—Lustre[5]

TABLE 2. Comparison of traditional petascale era infrastructures with converged exascale era infrastructure

Characteristics	Traditional HPC cluster	New era converged	Traditional AI POD
Primary workload	Modeling and simulation	Both	Machine Learning/Deep Learning
Compute node type	CPU nodes	Both	GPU-accelerated nodes
Number of compute nodes	Hundreds to thousands	Hundreds to thousands	A few to tens
Typical interconnect	InfiniBand	Both (plus HPE Slingshot)	Gigabit Ethernet
Typical I/O intensity	Write intensive	Both	Read intensive
Typical storage	HDD-based parallel storage	Parallel storage that supports fast NVMe flash pools and cost-effective HDD pools in the same file system: HPE Parallel File System Storage and Cray ClusterStor E1000 Storage System	All-flash enterprise file storage
Capacity measured in	Petabytes	Petabytes	Terabytes
Scalability in single namespace	Up to exabytes	Up to exabytes	Up to low double-digit petabytes
Ideal uses	• Mainly writing • Large files • Mainly sequential order • Speeds up to double- or triple-digit GB/s	• Reading and writing • Files of all sizes • Both sequential and random order • Speeds up to double- or triple-digit GB/s	• Reading and writing • Files of all sizes • Both sequential and random order • Speeds up to low double-digit GB/s
Price per terabyte	Low	Medium	High

[4, 5] "Shifts Are Occurring in the File System Landscape," Hyperion Research, Special Study, June 2020

Seymour Cray once said:
"Anyone can build a fast CPU. The trick is to build a fast system."[7]

It is a given that one size does not fit all. That's why HPE created a portfolio of fast, scalable, and cost-effective parallel storage systems that embed the leading parallel file system of respective target markets in the HPE factory. Operational support services for the whole system—hardware and software—from HPE Pointnext Services are included. This means that you have singular accountability. No need to contact IBM in the case of support for HPE Parallel File System Storage, and no need to contact the Lustre community in the case of Cray ClusterStor E1000 support.

Adapting the quote on the left to our HPC/AI storage topic—anyone can build fast storage with NVMe flash, but the trick is to not only build a fast but also a scalable and cost-effective system.

To achieve the latter, HPE leverages medianomics in both our parallel storage systems. What is medianomics? It is a portmanteau word that combines the words media and economics into a single word. It means to leverage the strengths of the different storage media—for NVMe SSD performance, and for HDD cost-effective capacity. And it means to do this while minimizing the weaknesses—for NVMe SSD, it's high cost per terabyte; and for HDD, it's the inability to efficiently serve small, random I/O. IDC forecasts[6] that even in the year 2024 the price per terabyte of NVMe SSD will be seven times higher than the price for terabyte of HDDs. For this economic reason, we believe that hybrid file systems will become the norm for the foreseeable future.

TABLE 3. Storage media: Strengths and weaknesses

Storage media	Strengths	Weaknesses
SSD	Performance	Cost per terabyte
HDD	Cost per terabyte	Not well suited for random I/O of small files

[6] "WW 2020–2024 Enterprise SSD and HDD Combined Market Overview," IDC, July 2020

[7] hpe.com/info/hpc-storage

Both our parallel storage systems enable you to have two different media pools in the same file system:

- NVMe flash pools to drive the required performance (throughput in GB/s or IOPS)
- HDD pools to provide most of the required storage capacity in the most cost-effective way

It is important to state that we do this within the same file system and do not tier different files systems to leverage medianomics. Table 4 gives a high-level overview storage explaining which storage system fits where.

TABLE 4. Positioning of exascale era HPC/AI storage system

Characteristics	HPE Parallel File System Storage	Cray ClusterStor E1000 Storage System
Typically attached to	Clusters of HPE Apollo systems or HPE ProLiant DL rack servers with HPE IB HDR/Ethernet 200Gb adapters	HPE Cray EX supercomputers or large clusters of HPE Apollo systems and HPE ProLiant DL servers
Number of clients (compute nodes)	A few to tens	Hundreds to thousands
Typical interconnect	InfiniBand HDR or 100/200GbE	200 Gbps HPE Slingshot, InfiniBand HDR, or 100/200GbE
Embedded parallel file system	IBM Spectrum Scale	Lustre
Set of enterprise-grade storage features including availability, redundancy, security, compliance, data protection, and replication	Comprehensive	Basic
Protocol support	POSIX, NFS, SMB, HDFS, Object	POSIX
Scale-out typically in increments of	Terabytes (starts at 10 TB)	Petabytes
Best suited for HPC/AI in	Manufacturing, healthcare and life sciences, financial services, hi-tech, pharma, telco, retail, and for home directories in organizations listed in the right column	Research laboratories and universities, defense and intelligence agencies, space agencies, climate research institutions, weather forecasting, geosciences (including oil and gas)

THREE APPROACHES TO DRIVE BUSINESS OUTCOMES WITH OUR NEW STORAGE

These three scenarios address organizations that are today using third-party file storage for HPC/AI clusters built on HPE Apollo 2000 systems and HPE Apollo 6500 systems or HPE ProLiant DL rack servers with HPE InfiniBand/200GbE adapters.

Tactical: Departmental scope

In this example, one department today is using an HPE Apollo 2000 cluster with IBM Spectrum Scale/HDD-based DDN GRIDScaler storage for simulation workloads and another HPE Apollo 6500 cluster with all-flash NAS storage from Dell EMC for training machine learning. Table 5 shows the **before** and **after** views:

TABLE 5. Example for departmental storage consolidation

	Before		After
Characteristics	**Storage "stovepipe" for HPE Apollo 2000 cluster**	**Storage "stovepipe" for HPE Apollo 6500 cluster**	**Shared storage for both clusters**
Storage system	DDN GRIDScaler	Dell EMC Isilon F-Series	HPE Parallel File System Storage
File system	IBM Spectrum Scale	Dell EMC OneFS	IBM Spectrum Scale
User data storage on	HDD	SSD	Both
End-to-end support for both compute and storage	No	No	Yes
License fee per terabyte capacity	Yes	Yes	No

Top business benefits and outcomes include:

- **HPC/AI storage cost savings** that can either be redeployed to expand CPU/GPU compute nodes or to other business-critical initiatives:
  - Elimination of license fees per terabyte capacity that can lead to exploding storage costs when storage capacity requirements grow fast
  - Cost-effective application of medianomics within the same file system
  - Cost-effective standard x86 rack servers (HPE ProLiant DL) as storage servers versus expensive custom storage controllers
- **Faster problem identification and resolution** for issues that are spanning both CPU/GPU compute and storage:
  - Elimination of unproductive and frustrating vendor finger-pointing
  - Unified accountability of HPE Pointnext Services for the full infrastructure stack
  - Full spectrum of operational support services (HPE Pointnext Tech Care, HPE Pointnext Complete Care) available for the full stack

Strategic: Multidepartmental HPC/AI consolidation of storage islands

In this example, several departments in an organization are using HPE Apollo 2000 clusters with IBM Spectrum Scale HDD-based storage from multiple vendors for simulation workloads and several HPE Apollo 6500 clusters with all-flash NAS storage from multiple vendors for training machine learning. Table 6 shows the before and after views:

TABLE 6. Example for multi-departmental storage consolidation

	Before					After
Characteristics	**Department A: Storage "stovepipe" for HPE Apollo 2000 cluster**	**Department B: Storage "stovepipe" for HPE Apollo 2000 cluster**	**Department A: Storage "stovepipe" for HPE Apollo 6500 cluster**	**Department B: Storage "stovepipe" for HPE Apollo 6500 cluster**	**Department C: Storage "stovepipe" for HPE Apollo 6500 cluster**	**Shared storage for all clusters**
Storage system	DDN GRIDScaler	Lenovo DSS-G	Dell EMC Isilon F-Series	NetApp AFF	Pure Storage FlashBlade	HPE Parallel File System Storage
File system	IBM Spectrum Scale	IBM Spectrum Scale	Dell EMC OneFS	NetApp ONTAP	Purity//FB	IBM Spectrum Scale
User data stored on	HDD	HDD	SSD	SSD	Flash	HDD and SSD
End-to-end support for both compute and storage	No	No	No	No	No	Yes
License fee per TB capacity	Yes	Yes	Yes	Yes	Yes	No
Asset utilization example	60%	90% (expensive upgrade on order)	50%	90% (expensive upgrade on order)	40%	80% (no upgrade on order)
Usable capacity example	500 TB	600 TB	200 TB	300 TB	200 TB	1,600 TB

Top business benefits and outcomes include:

- **HPC/AI storage cost savings** that can either be redeployed to expand CPU/GPU compute nodes or to other business-critical initiatives:
  - Elimination of license fees per TB capacity that can lead to exploding storage costs when storage capacity requirements grow fast
  - Cost-effective application of medianomics within the same file system
  - Cost-effective standard x86 rack servers (HPE ProLiant DL) as storage servers versus expensive custom storage controllers
  - Cost saving effect of running a shared infrastructure at higher average utilization with less total capacity (1,600 TB after versus 1,800 TB before) with an overall lower average cost per TB
- **Faster problem identification and resolution** for issues that are spanning both CPU/GPU compute and storage:
  - Elimination of unproductive and frustrating vendor finger-pointing
  - Unified accountability of HPE Pointnext Services for the full infrastructure stack
  - Full spectrum of operational support services (HPE Pointnext Tech Care and HPE Pointnext Complete Care) available for the full stack

HPE GreenLake for HPC
The cloud that comes to you

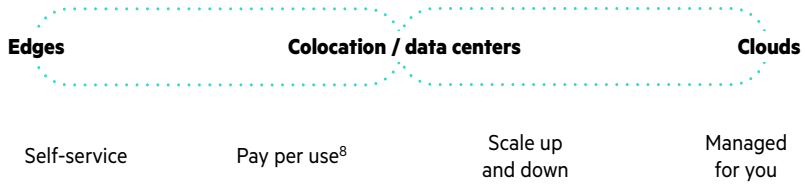

FIGURE 1. The cloud that comes to you

Transformational: Your first fully managed private HPC cloud
Everybody who ever has done a comparative full cost analysis comparing your own HPC/AI equipment run by your own team with high utilization 24x7 in your own data center versus running the same workloads in the public cloud can identify with this the simple truth:

From a cost perspective, the public cloud is not competitive for high-performance workloads that need fast data movement of massive data sets during processing and that run with high utilization.

However, if you are in one of those organizations with a C-level mandate to move everything out of your own data centers to the cloud as your organization no longer is investing in owned data centers, then HPE GreenLake for HPC is for you.

You can meet this mandate by getting your first fully managed HPC private cloud with a true self-service experience for your users in a colocation facility near you.

HPE GreenLake for HPC provides:

- **Scalable HPC solutions with self-service capabilities**—With more workloads and applications driving the use of high-performing compute environments, HPE GreenLake edge-to-cloud platform can help address demand with a scalable solution that offers transparency into usage with self-service capabilities for users to get the resources they need on demand.
- **Efficient, pay-per-use model**—HPC workload demands can fluctuate. With HPE GreenLake platform, you can avoid the need to overprovision by rightsizing your environment. You pay only for the resources you consume, and built-in buffer capacity enables you to handle both steady growth and unexpected spikes in demand.
- **Industry-leading technology and services**—HPE GreenLake for HPC combines the power of our industry-leading HPC infrastructure and expertise with the as-a-service cloud experience from HPE GreenLake platform. From one partner, HPE, customers can get the dense compute, high-speed storage, interconnects, and the software needed to run and manage HPC clusters including the new HPC/AI parallel storage systems that were described earlier in this paper.
- **Managed for you**—HPE GreenLake Management Services, securely delivered from our world-class IT Operation Centers (ITOCs) around the globe, help you fill skills gaps and free up your resources for more productive tasks. HPE experts handle the performance tuning, capacity planning, lifecycle management, firmware updating, and

[8] Above a reserve

patch management while monitoring critical KPIs of performance, uptime, time to resolution, and ticket status.

THE HPE ADVANTAGE

Compare HPE parallel HPC/AI storage systems with other vendor's approaches to HPC/AI storage, and it's clear that we are doing something different. That is because HPE approaches next-generation HPC/AI storage from a different

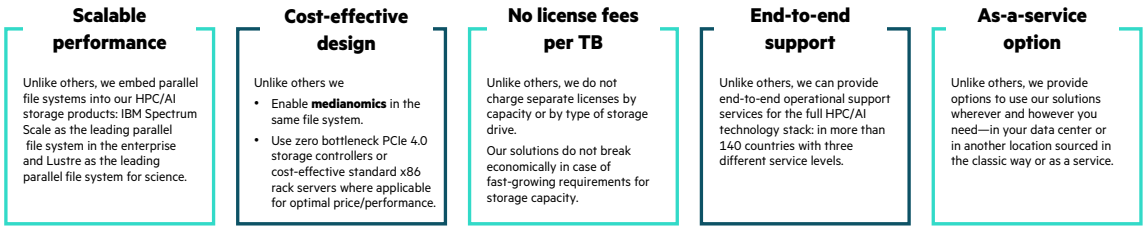

FIGURE 2. Pillars of differentiation

perspective than others in this marketplace. Here is a summary of the HPE advantage.

If you are struggling with one or more of these typical HPC/AI storage challenges, reach out to your HPE representative to set up a meeting with your team and our experts:

- **Delayed new product introduction and development** due to long job run times in business-critical R&D business processes
- **Low asset utilization** of expensive CPU/GPU compute infrastructure due to I/O bottlenecks that keep the compute nodes idling waiting for their data
- **Regrettable attrition of top talent** such as key engineers and data scientists due to frustration over job pipeline congestion that prevents them from working productively
- **Frequent, unbudgeted spending requests** due ad hoc upgrades of the file storage infrastructure to cope with the architectural deficiencies

The HPE portfolio of parallel HPC/AI storage products is designed to solve these challenges. We are ready to partner with you now.

ACCELERATE YOUR TRANSFORMATION

If you want to accelerate the realization of cost savings and business outcomes by switching to our new HPC/AI parallel storage systems, you might want to take a look at the HPE Accelerated Migration offering from HPE Financial Services.

HPE Accelerated Migration can help you fast-track your HPC/AI storage transformation with minimal disruption to your current environment. Unlock the hidden value in your existing HPC/AI storage assets as you transition to new, more cost-effective solutions. Shift existing, owned assets to a flexible usage payment model during the transition and free up cash for new investment.

With the majority of your IT budget consumed to keep the business running, funding for transformation—both for infrastructure transformation and digital transformation—is often insufficient. Finding incremental resources today can help you close the gap. HPE Accelerated Migration helps you access the value in your legacy equipment and move forward sooner. Contact your HPE representative to learn more.

Resources

HPE Accelerated Migration

HPE GreenLake for High Performance Computing

LEARN MORE AT

hpe.com/us/en/solutions/hpc-high-performance-computing/storage.html

Make the right purchase decision.
Contact our presales specialists.

Get updates

a50003566ENW, November 2021, Rev. 1

Hewlett Packard Enterprise