Hewlett Packard Enterprise

Technical white paper

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

HPE STORAGE SOLUTIONS FOR AI AND ANALYTICS

Matching datastore architecture and solutions with data pipeline requirements

CONTENTS

Artificial intelligence (AI) and analytics data pipelines are inherently different from those of traditional applications, with distinct requirements at each stage. Storage needs grow increasingly complex as artificial intelligence (AI) is used more broadly as datasets and models get larger and spread across clouds, data centers, and to the network edge. A range of storage architectures and HPE solutions described in this paper can effectively balance these competing needs.

EXECUTIVE SUMMARY

Each stage of an AI and analytics process has distinct storage requirements. Data ingestion and AI training demand high throughput, ETL (Extract, Transform, Load) requires mixed read/write handling, while inference requires low latency and high throughput. The entire data pipeline must use a single namespace to avoid creating silos and make all data and insights visible everywhere. AI at the edge is driving the extension of the data pipelines to cover edge-to-core-to-cloud.

AI storage solutions must meet all these varied requirements and offer scope for future development too.

HPE has extensive experience in creating advanced data pipeline architectures and integrated engineering solutions. We offer an expansive portfolio of AI storage solutions. This paper reviews key aspects of AI storage requirements, and it describes how Cray ClusterStor E1000 Storage System, HPE Ezmeral Data Fabric, HPE Parallel File System Storage, and HPE solutions for Qumulo is uniquely positioned for your specific AI workloads.

COMPLEXITY IMPACTS STORAGE NEEDS

Business leaders know they need to leverage data to accelerate competitiveness and profitable revenue growth. This strategy incorporates the use of analytics and artificial intelligence (AI) tools, including AI subsets, machine learning, and deep learning (ML and DL).

Adopting AI and analytics is challenging. AI workloads have significantly greater data storage and compute needs than traditional workloads. There are many tools and frameworks available on the market to simplify delivery of your first project, but things become increasingly complicated when workloads move into production:

- AI models get bigger and more complex

Bigger infrastructure is needed to support them, including more powerful AI server clusters, more GPUs in the servers, and storage that delivers more throughput.

- Datasets grow to tens or hundreds of petabytes

Infrastructure needs to scale quickly to support this growth.

- Data spread across clouds, data centers, and edge devices

Data can be everywhere, and datastores must run all these scenarios together.

As we can see, AI project complexity has multiple dimensions. These complexity dimensions have a profound impact on datastore architecture requirements.

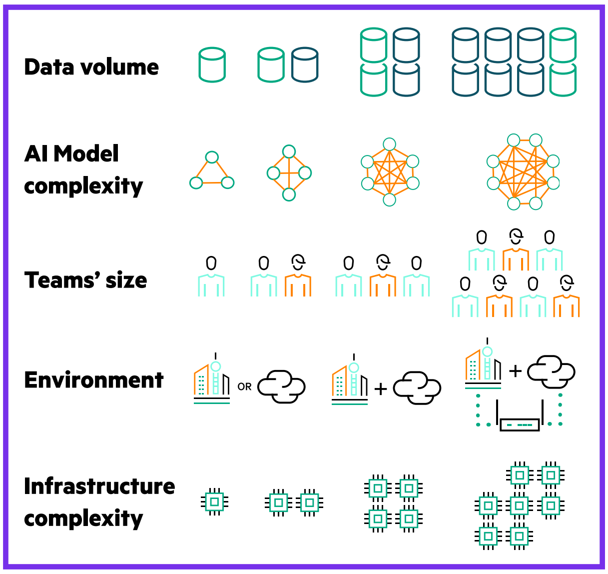


FIGURE 1. Infrastructure complexity grows in correlation with AI complexity

Types of AI and analytics

Storage and compute needs vary depending on the kinds of machine learning (ML), deep learning (DL), or neural network algorithms being employed. Storage requirements are also different at each phase of AI model training.

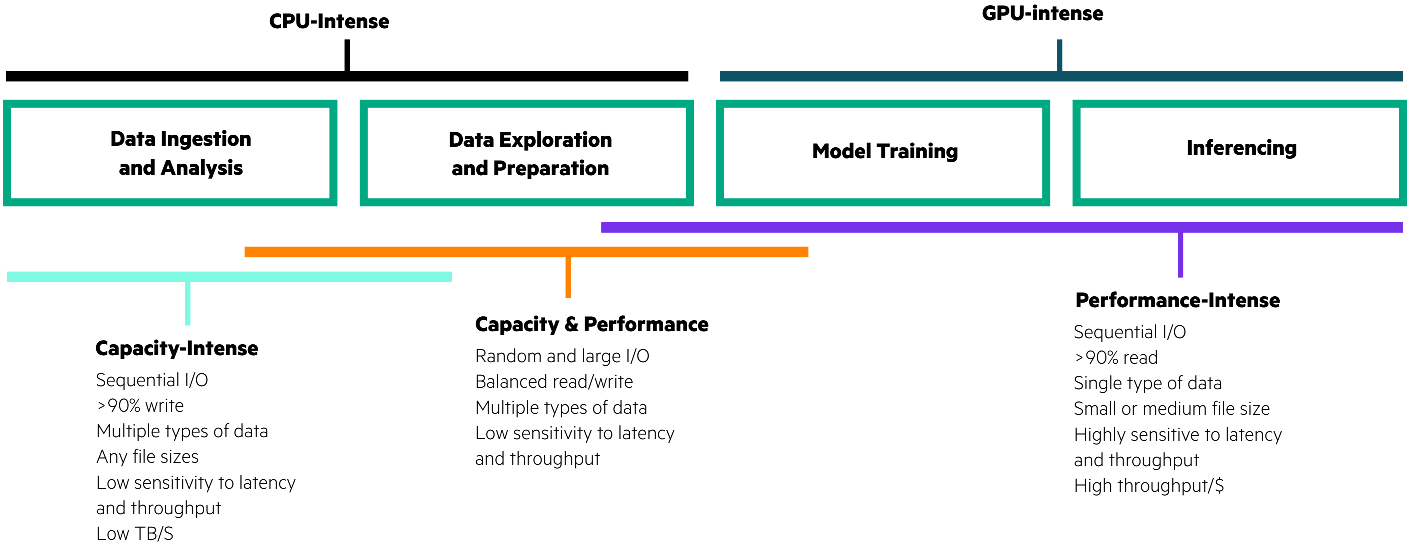


FIGURE 2. Input/output profiles (I/O) for different AI model training phases

Data ingestion workloads have sequential I/O patterns that use files of any kind and size and have very low sensitivity to latency (except in streaming). Writes account for 90% of workload input/output (I/O) activity.

Traditional analytics, data exploration, and data preparation use both random and sequential I/O. The workloads have balanced read-and-write patterns, use multiple data types and sources, and manage files of all sizes from small to large. They have low sensitivity to latency and throughput requirements. Data preparation is an integral part of the AI modeling process.

In deep learning and neural network spaces the requirements change. Algorithms typically have sequential I/O patterns, process a single type of data in small files, and require low-latency and high-throughput solutions. These workloads run significantly faster with GPU acceleration, whereas data analytics algorithms run faster on traditional CPUs.

Other Factors to consider

Data set size, AI model characteristics, storage throughput, and edge-to-core connection (discussed above) are critical factors to consider in selecting the right data pipeline. Other relevant factors include:

1. **File system type and data access:** There are multiple options for file systems with different data access services. Scale-out file systems such as Lustre or GPFS and Hadoop Distributed File System (HDFS) are very popular for AI, big data and HPC workloads. Network File System (NFS) remains the most widely deployed shared file system for technical applications for the last 30 years. Alternatively, new GPU-accelerated protocols like RDMA (Remote Data Memory Access) and GDS (GPU Direct System) are now being used to eliminate CPU bottlenecks.

2. **File size and dataset size:** The number of files and their size play an important role in the datastore's performance.

3. **Ease of Management:** This includes install, configuration, and tuning. For example, graphical user interface and automatic policy-based data management and tiering simplify operations. The technical skill necessary to use a storage solution is a crucial consideration when selecting the datastore.

4. **Capacity planning:** Storage environments can be complicated and dynamic, making capacity planning difficult. Storage solutions have built-in system monitoring tools, log files, system management to help with capacity scale-out.

CREATING THE DATA PIPELINE

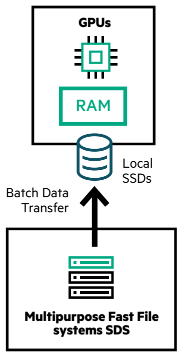
The main storage challenge when supporting GPU-based servers is the creation of AI and analytics data pipelines that efficiently support different IO patterns, multiple parallel project execution, edge-to-cloud-to-core strategies, and accessibility to datasets that grow in size and complexity.

Modularity, scalability, and data architecture composability are essential because AI, ML, and analytics workload profiles change over time. The architecture must allow easy addition of storage and compute "building blocks" to meet new node profile needs. The architecture must enable combinations of heterogeneous compute and storage blocks (e.g., standard compute, AI optimized, high-storage density, optimized memory, and optimized latency for real-time analytics) to create specialized pipeline architectures for specific workloads. These architectures operate within the same cluster sharing the same data.

Data pipeline architecture complexity can range from a single scale-out file system to a combination of a fast parallel file system, data lakes, and object stores. There are four main high-level architectural categories, each addressing incremental AI/ML environment complexity:

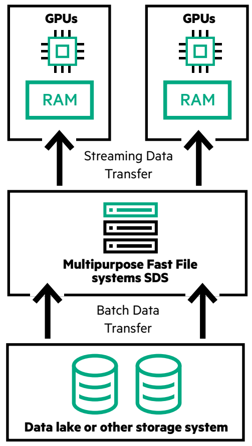
1. No data pipeline

- Datasets are copied (batch) in the AI server disks. AI training starts using data loaded in the local disks
- High Throughput (SSD disk IO), but a single AI model can't run on clustered servers
- Data is moved manually between AI Server and data storage
- No long-term viability – dataset will rapidly outgrow local disk capability



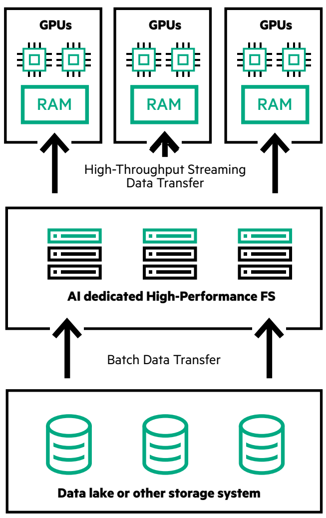
2. Multi-purpose datastore

- A single platform supports both traditional file server and AI/DI workload
- The data platform transfers in real-time the dataset (using standard protocols like NFS) to the AI server(s) during the AI model training or inference sessions
- Limited-throughput for a single client
- Cost-effective solution for hybrid small and mid-size AL, DL and traditional workloads



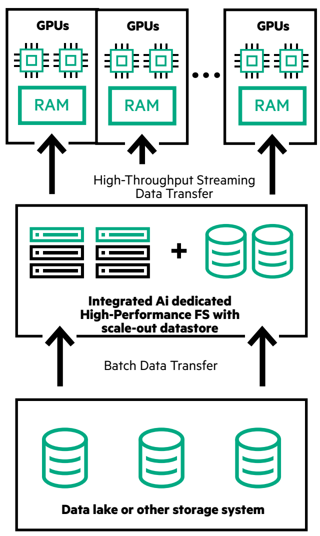
3. Integrated datastore

- High-performance parallel datastore is logically and physically integrated with the large-capacity datastore
- High-aggregated throughput
- Simplifies infrastructure and operations by hiding internal data movement and tiering
- Easy to expand, tune, and manage but lacks the flexibility of two-tier systems in hybrid installations



4. Disaggregated datastores

- High-performance and high-capacity datastores are physically and logically separated. Architecture can effectively manage AI and ML workloads of any size
- Performance store is a scale-out parallel file system containing all active data used for front-end operations (e.g., streaming data to AI compute nodes). Capacity store is a scale-out object store or data lake for all data
- High-aggregated throughput (>300 GB/s) and for a single client (>40 GB/s)
- Data movement between stores is managed manually or with data management systems
- Physical separation of components allows scaling of the two platforms independently but can make operations more complicated



HPE SCALE-OUT DATASTORE SOLUTIONS FOR AI

The HPE Elastic Platform for Analytics (EPA) is designed to simplify data pipeline architecture design. HPE EPA leverages servers from the HPE Apollo and HPE ProLiant DL families, which can be further extended with scale-out datastore capabilities.

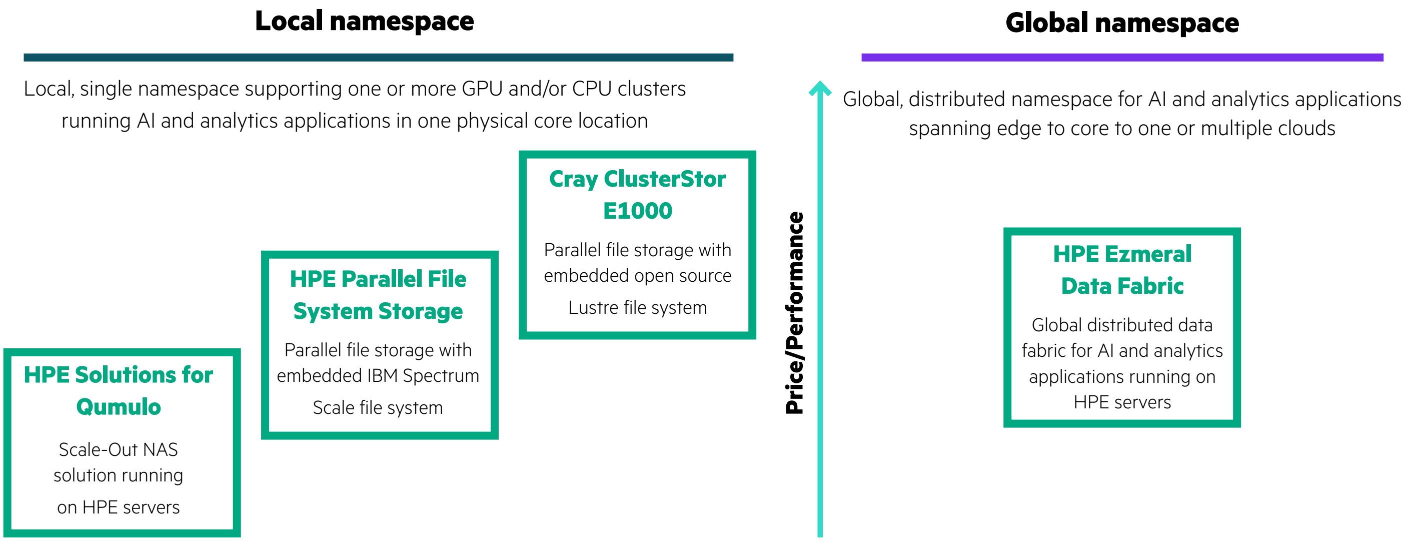


FIGURE 3. HPE portfolio of AI datastore software solutions

Cray ClusterStor E1000

Cray ClusterStor E1000 is a Lustre-based integrated storage solution for medium and large AI applications that require high throughput and leading price/performance. Cray ClusterStor E1000 supports namespaces from one petabyte to exabytes in size with linear performance scaling.

Cray ClusterStor E1000 systems are being for large-scale AI use cases in research centers, defense and intelligence agencies, weather forecasting, climate research and energy as well as and new medical visualization technologies such as cryo-EM, which generate images in the 100 GB range. Using ML to accelerate interpretations of these images requires significant storage capacity and extreme IO.

HPE Parallel File System Storage

The first and only IBM Spectrum Scale-based AI storage system with standard x86 rack servers as storage nodes and no need for separate file system licenses. Available in all flash, all HDD or mixed configurations, cost-effective IBM Spectrum Scale (FKA GPFS) storage can be configured to your workload requirements and budget. HPE Parallel File System currently supports usable capacities in a single file system from 27 TB to 25 PB.

Typical HPE Parallel File System Storage applications include data-intensive modeling and simulation, artificial intelligence, and high-performance data analytics workloads. Due to the comprehensive list of enterprise storage features this system is best suited to AI operations in manufacturing, automotive, healthcare and life sciences, financial services, pharma, retail, telco and hi-tech industries.

HPE Solutions for Qumulo

With its scale-out, flash-first architecture, and a distributed file system purpose-built for massive concurrency across all data types, the Qumulo file system performance, scalability, and easy operations. It keeps configuration requirements and management complexity to a bare minimum, allowing seamless and linear scalability from TBs to PBs all in one namespace. Qumulo also provides the scale-out performance and concurrency needed to accelerate AI workloads at scale.

Qumulo scale-out NAS has been certified and optimized for use with the HPE Apollo 4000 Systems and HPE ProLiant DL325 Gen 10 Plus server family, to deliver an extremely cost-effective, petabyte-scale, high-performance solution for AI-centric workloads.

HPE Ezmeral Data Fabric

HPE Ezmeral Data Fabric provides an exabyte-scale, edge-to-cloud-to-core distributed file system and data platform, supporting a broad range of AI, ML, DL, and analytics applications in a single, global namespace. HPE Ezmeral can ingest, store, manage, process, apply, and analyze all data types from any data source, with a range of different ingestion mechanisms

HOW TO SELECT THE RIGHT DATASTORE FOR YOUR AI NEEDS

It is a common misconception that performance is the only important criterion when choosing AI datastore solutions but other factors also need to be considered to ensure you choose storage optimized for your workloads.

Performance

There are three performance elements to consider in choosing a data platform for AI:

1. Aggregated throughput - The sum of the data transfer rate required by all AI models training and inferencing running simultaneously.

2. Throughput for a single process - many models, scenarios and protocols are designed to operate in a single process configuration.

3. Latency, file size, and data access profile - the time it takes for the IO request to be completed has a significant bearing on throughput.

For the most demanding applications, that need to access very large data sets in a single namespace at very high speeds consider HPE Parallel File System Storage and Cray ClusterStor E1000 which performs exceptionally well in all three categories. HPE Ezmeral Data Fabric and Qumulo are better suited to smaller AI environments – with HPE Ezmeral Data Fabric being the right choice if you need to build a global namespace across many geographically dispersed environments.

AI and Analytics Workload management

HPE offers a solution for six different AI and analytics workloads;

TABLE 1. HPE datastore workload management capabilities

Workload	Recommendation	Current environment
Mixed HPC (modeling & simulation) and AI	Cray Cluster E1000 HPE Parallel File System Storage	• Mature AI operations • Big data running on HPC GPU-powered servers
High-performance AI	HPE Parallel File System Storage Cray Cluster E1000	• Mature AI/ML operations • HPC GPU-powered servers
High-performance analytics	HPE Ezmeral Data Fabric Qumulo All-Flash / Hybrid	• Mature AI operations • HPC GPU-powered servers
Medium performance AI	Qumulo All-Flash / Hybrid HPE Ezmeral Data Fabric	• Developing AI capabilities • Mid-spec GPU-powered servers
Mixed enterprise file sharing and AI	Qumulo All-Flash / Hybrid	• Developing AI capabilities • Mid-spec GPU-powered servers
Mixed analytics and AI	HPE Ezmeral Data Fabric	• Entry-level AI capabilities • Low/Mid-spec GPU servers

Capacity

AI data training sets tend to be massive to deliver better algorithms, but there are differences when it comes to how much usable capacity can be stored and accessed in a single namespace:

Cray ClusterStor 1000 and HPE Ezmeral Data Fabric do not have scalability limits. Both scale to exabytes if required. HPE Parallel File System Storage currently scales to 25 petabytes in a single file system, a testing limitation rather than an architectural barrier. The product road map contains appropriate actions (larger drives, higher density enclosures, etc.) to expand scalability over time. HPE Solutions from Qumulo typically operate in the hundreds of terabyte or single digit petabyte range.

Ease of operation

Parallel filesystems such as Lustre and IBM Spectrum Scale (FKA GPFS) embedded in the Cray SuperStor E1000 Storage System and HPE Parallel File System Storage are more complex to configure, maintain, monitor, and manage than Network File System (NFS) based file storage. In cases where no parallel file system skills are available in an organization and performance requirements are moderate, NFS-based NAS storage like Qumulo may be a more appropriate choice.

Edge to cloud to core support

The support of data distribution across public cloud(s) and on-premises environments is another consideration. Customers interested in managing hybrid cloud and on-prem solution should be interested in having their datastore deployed in both data centers and clouds.

- Cray ClusterStor E1000 and HPE Parallel File System Storage are designed to support server clusters in one physical location
- HPE Ezmeral Data Fabric can run across edge, core, and public cloud environments (but it has some limits in public cloud deployment such as no global namespace sharing)
- Qumulo supports both on-prem and in cloud deployment

REAL-LIFE SCENARIOS

Large HPC and AI infrastructure

An oil and gas company needs to manage a vast quantity (dozens of PB) of data to run simulations and AI models. To control costs, data is stored in two main systems: the disk-based scratch file system and a tape-based archive with software moving between both. Finding, locating, and copying files from tape is very slow.

Recommended HPE solution

Cray ClusterStor E1000 with an object store controls storage costs, while replacing tape libraries with object storage allows for retrieval and makes data available for analysis faster.

- HPE DMF copies data from Object Store and Cray ClusterStor and vice-versa
- Cray ClusterStor E1000 File Systems with 5 PB
- HPE DMF v.7 with 5 PB usable storage
- 10 PB Object Storage with 9+3 erasure coding scheme
- InfiniBand 100 GbE for the front-end computers (Cray ClusterStor and HPC servers)
- 25 GbE connecting Cray ClusterStor to the Object Store

Massive analytics infrastructure

A global automaker needs an infrastructure to support the development of intelligence for their autonomous cars program. They require a data platform to collect and manage massive amounts of data from test vehicles and make them available to developers worldwide. The key issues are:

- Extracting data from test vehicles quickly and efficiently
- Enabling direct access to data, including from legacy systems, using standard protocols like NFS
- Leveraging deep learning techniques to train algorithms
- Scaling storage easily and effectively
- Ensuring high availability and data resiliency
- Enabling developers to run containerized applications
- Existing integration with AI and analytics ecosystems

Recommended HPE solution

- HPE Ezmeral Data Fabric running on 1000x HPE Apollo 4510 servers with 60x 16 TB HDDs. The platform is capable of ingesting 5 PB/day and offering aggregated throughput of 10 TB/s over 50-100 PB of dataset.

AI-as-a-service infrastructure

A manufacturing company wants an AI environment to support four data science teams running experiments (AI training or inference) in parallel. They have limited experience with AI and containers, favoring bare-metal solutions.

AI's main activity is classification of 1m sensor-captured images (1 MB each, 1 PB total) in the factory lines. The total data volume grows 15% per year. Models use datasets of a max of 50K images each and need to complete in about one hour.

Recommended HPE solution

- 4x HPE Apollo 2000 with two v100 GPUs each
- 10x HPE Apollo 4200 192 TB hybrid HDD/SSD with HPE solution for Qumulo

This solution provides aggregated throughput of 11.2 GB/s, sufficient to execute four experiments in parallel per hour.

All Flash file storage for a cluster of dense GPU-accelerated nodes

A semiconductor company needs a minimum of 500 TB usable storage capacity that can feed a cluster of dense GPU-nodes with 32 x NVIDIA® A100 GPUs with an aggregate read bandwidth of a minimum of 100 GB/second.

Recommended HPE solution

- 4 x HPE Apollo 6500 Gen 10 Plus with Nvidia A100 HGX 40 GB x8 and 1 TB memory each
- Connected via InfiniBand HDR to;
- 1 x HPE Parallel File System Storage
    - 4 rack units with a total of 64 NVMe SSD 15.36 TB for 527 TB usable capacity delivering 102 GB/sec read and 65 GB/sec write aggregate throughput

SUMMARY

Adopting AI and analytics brings challenges that grow with the size and complexity of your projects. Storage options play a vital role in delivering the performance, scale, and mobility of your workloads need. Data is everywhere, and there is growing complexity associated with storage requirements for AI workloads. Challenges are related to GPU-based servers' support and to the multitude of different analytics and AI workloads, challenges that require a solid data pipeline architecture. As you have seen, performance, including throughput and latency, are essential factors to consider for AI storage solutions, but also workload size, capacity, ease of operation, and edge to the cloud to the core support. HPE has a portfolio of datastore solutions to address any data pipeline architecture for AI and the expertise to handle any storage AI and analytics demands.

LEARN MORE AT

HPC Storage Briefcase on Seismic
Competitive analysis

Hewlett Packard Enterprise

a00114847enw, June 2021