Enterprise Strategy Group by TechTarget

WHITE PAPER

Simplify and Accelerate AI Adoption With Future-proof AI Storage

By Mike Leone, Practice Director and Principal Analyst
Enterprise Strategy Group

January 2025

Contents

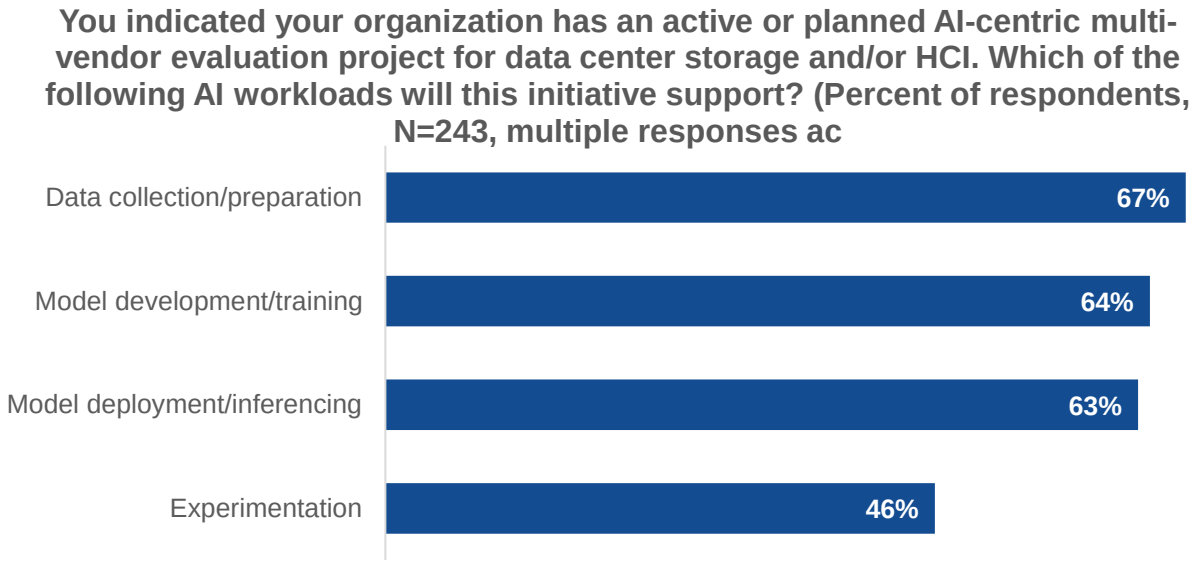
Storage Is Vital for the Performance, Scalability, Resilience, and Efficiency of AI Workloads

When it comes to AI use cases and applications, data storage isn't optional—it's foundational to progress and value. Of course, storage has always mattered greatly for all IT initiatives, but the technology has advanced substantially past the days of spindles and block and file storage. In fact, it's fair to say that AI has dramatically changed the rules of the game when it comes to storage.

There are several key reasons for this, including the extensive volumes of data both created by different AI versions—generative, predictive, causal—and used in the creation of AI training models at the heart of these initiatives. But that's not all: data collection, data preparation, training, experimentation, and inferencing/production all are pushing the limits when it comes to data volumes.

According to research by Informa TechTarget's Enterprise Strategy Group, of those organizations that have an active or planned AI-centric multi-vendor evaluation project that requires data center storage and/or HCI, 67% cited data collection/preparation as the workload that is or will be supported by the initiative, followed by model development/training (64%), model deployment/inferencing (63%), and experimentation (46%).[1]

Figure 1. AI Workload Support



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

But massive capacity—and easy and fast scaling of data—isn't all organizations need. They also need a future-proof AI data storage strategy to deliver the multidimensional performance, ease of deployment, and energy efficiency required for storage-intensive AI workloads. Use cases such as medical imaging, telecom infrastructure management, cybersecurity threat analysis, retail fraud detection, credit risk analysis, manufacturing quality control, and predictive maintenance are all examples of AI use cases that demand reliable, secure, high-performance storage.

[1] Source: Enterprise Strategy Group Research Report, *Navigating the Cloud and AI Revolution: The State of Enterprise Storage and HCI*, March 2024. All Enterprise Strategy Group research references and charts in this white paper are from this research report.

Identifying and Understanding Data Storage Challenges for the AI-ready Organization

Building, deploying, and benefitting from AI workloads for today's and tomorrow's demanding business scenarios requires careful evaluation, planning, and management of a wide range of options. With that comes a number of challenges, data silos, bottlenecks, and issues that make storage difficult for AI workloads.

Storage, in general, is more complex, more demanding, and more challenging today than ever before. And it's not just because storage volumes have exploded. Research from Enterprise Strategy Group pointed out that integration issues, performance demands, storage provisioning, and supporting application modernization efforts all make storage more challenging.

Additionally, the growing adoption and implementation of AI has upped the bar even higher for organizations' storage requirements. Consider, for instance, AI's role in accelerating the massive growth of unstructured data, as well as the growing need for linear storage performance at unprecedented scale. Several other issues stand out, including the huge energy efficiency demands AI puts on storage infrastructure, as well as the need to build "responsible AI usage" best practices into storage infrastructure.

Most organizations also have major shortfalls when it comes to in-house storage expertise. This means organizations need access to experienced tech providers and their partner ecosystems in order to plan, deploy, and manage storage, especially for storage-centric AI workloads.

Four major challenges stand out as prime targets for organizations to address:

- **Maximizing GPU utilization.** Ensuring storage is tightly aligned with GPU infrastructure is important to avoid latency, bandwidth limitations, and inefficient use of GPUs.
- **Ensuring proper mixed-media performance.** AI workloads typically create and use data that is stored on multiple devices, such as hard drives, flash memory, solid state disks, and more. An AI-ready data storage platform must deliver consistently high performance.
- **Supporting data volume growth and scalability.** Delivering performance at scale is an essential goal for all storage solutions, but it is particularly tricky as organizations ramp up their use of tools such as generative AI, predictive AI, and causal AI. As a result, scaling must be seamless and non-disruptive.
- **Supporting energy efficiency.** Although storage technologies are far more energy efficient today than they were just a few short years ago, AI workloads put substantial pressure on infrastructure from an energy consumption standpoint.

Finally, organizations must account for the need to future-proof their storage infrastructure against the rapid change of AI, which undoubtedly has made earlier storage investments obsolete before new projects reach production status.

What to Look For in an AI-ready Storage Infrastructure Platform

As organizations consider their options when it comes to putting in place the ideal storage infrastructure platform for AI workloads, it is important to keep in mind both the technical requirements and the capabilities of the solutions provider.

According to research from Enterprise Strategy Group on storage purchase decision criteria across the AI lifecycle, hybrid/multi-cloud capabilities, data security/governance, ease and speed of data movement, and speed of deployment/provisioning are at the top of the list when it comes to the most important storage capabilities related to

AI workloads. These criteria hold true across the full spectrum of AI development and deployment: data collection, data preparation, training, experimentation, and inferencing/production.

Figure 2. Top 10 Areas Influencing Storage Purchases to Support AI Workloads

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

What else matters? Several must-have features for AI-ready storage include:

- **Tight alignment with GPUs for optimized performance.** Maximizing the utilization of GPUs is vitally important because organizations count on uninterrupted, consistent, and high performance of GPUs, especially as AI workloads scale up and out.
- **Support for unified data.** The proliferation of data sources has been well underway for some time, but AI has turbo-charged that trend. Data storage solutions must support a wide range—and a wide diversity—of data sources for AI workloads.
- **Simplified storage management.** This has always been a challenge for organizations in increasingly heterogeneous IT environments, and especially with the move toward hybrid cloud, multi-cloud, and edge computing architectures. AI workloads kick that up a notch.
- **Massive, reliable scalability.** Scalability is essential because AI initiatives often start small but grow rapidly and unpredictably as models get refined, new training data becomes available, and end-user adoption increases concurrency and latency requirements. A scalable data storage platform ensures seamless handling of sudden data spikes, supporting iterative development and long-term growth without compromising performance, reliability, or cost efficiency.
- **Simplified, reliable deployment processes.** In particular, deployment processes must be optimized for different AI workloads, putting pressure on storage architectures that can be implemented quickly, reliably, and consistently.

When it comes to selecting the right AI-ready storage provider, organizations should look for a partner with a deep commitment to storage research and development, an established track record for supporting storage installations on a global basis, the ability to develop and deploy storage platforms for demanding AI workloads, and an understanding of how to align storage infrastructure with essential GPU infrastructure to deliver linear performance at scale.

Pure Storage Delivers on the Promise of a Future-proofed, AI-ready Storage Infrastructure

Since storage has long been a staple of successful IT deployments and a critical element in any workload's infrastructure stack, special attention should be placed on making smart, forward-thinking decisions on both the storage technology platform and on the provider behind the solution.

Pure Storage, a well-established player in the storage industry and a leader in flash storage, has made substantial investments in, and commitments to, supporting AI workloads that now are vital to transformative applications. Pure Storage has adopted a platform mentality to AI-centric storage in an effort to both deliver peak performance for demanding workloads and to ensure that storage solutions are deployed effectively and efficiently.

The Pure Storage Platform is engineered with a future-proof mentality in order to ensure storage investments can keep up with ever-mounting demands for AI workloads. The platform delivers linear performance at scale to avoid latency and other performance bottlenecks, high reliability, availability, and resiliency, rock-solid security, and integrated hardware acceleration for performance-intensive AI workloads.

Additionally, since the Pure Storage Platform was designed specifically for AI-centric workloads, it is optimized to work closely with industry-leading GPUs, such as those from NVIDIA, in order to deliver peak GPU performance.

The Pure Storage Platform comprises several key components, including:

- **FlashBlade//S,** a scale-out storage system designed to deliver peak performance for AI training and inference requirements. Its modular design supports the introduction of new storage capabilities and requirements to help provide a future-proof model for AI now and in the coming years.

Pure Storage's Storage Ecosystem Includes a Tight NVIDIA Relationship

Having an extensive partner ecosystem is a crucial requirement for today's demanding AI workloads, and Pure Storage's network of hardware, software, and services partners helps in the development and delivery of enterprise-wide AI workload storage.

Pure Storage's relationships with partners in a wide variety of categories helps deliver a more robust, well-rounded, and as-needed set of capabilities. These include traditional value-added resellers, global systems integrators, managed service providers, and service specialization partners. But one of the key leverage points in Pure Storage's partner ecosystem is its Technology Alliance partners, particularly its collaboration with AI infrastructure giant NVIDIA.

For instance, Pure Storage FlashBlade is a certified storage solution for NVIDIA DGX SuperPOD, with the solutions working together to accelerate AI compute demands, deliver and manage AI storage more efficiently, and support an array of software tools for AI workloads.

Additionally, the two companies now can help organizations fast-track their AI workload deployments with the AIRI reference architecture certified with NVIDIA DGX BasePOD. This reference architecture provides organizations with a validated, enterprise-class, full-stack AI infrastructure solution that is faster and easier to implement than a do-it-yourself implementation. It includes NVIDIA Base Command, NVIDIA AI Enterprise, and Pure Storage Purity operating environment as an enterprise AI development foundation.

- **FlashStack** for AI with Cisco Validated Designs, which can help organizations deploy a validated architecture for AI workloads that reduces the design and investment risks when building a data, compute, and storage infrastructure for the AI data pipeline.
- **AI-Ready Infrastructure (AIRI),** a full-stack AI storage solution. It is certified for use on the NVIDIA DGX BasePOD reference architecture, using NVIDIA DGX GPUs, NVIDIA networking, and Pure Storage FlashBlade//S.
- **Evergreen//One for AI,** a storage-as-a-service offering designed and optimized for AI workloads. It provides performance guarantees that are based upon the maximum bandwidth needed for GPUs and accelerators, thus obviating the potential for storage-based performance bottlenecks.
- **Portworx,** a data services platform for Kubernetes-based applications. Portworx is key to creating, orchestrating, securing, and protecting cloud-native applications built using the Kubernetes container, utilizing persistent storage, data availability, data protection, data security, cross-cloud, and data migrations.

Having all these capabilities integrated into the Pure Storage Platform from the start helps organizations simplify AI workload deployment, a critical potential stumbling block in organizations' efforts to generate a fast return on investment and lower their total cost of ownership of AI workloads.

The platform's many native capabilities and AI-ready design helps make Pure Storage an experienced, and future-ready technology partner for the many storage-intensive AI workloads, such as healthcare imaging and diagnostic, financial market analytics, 5G telecommunications networks, and life sciences applications like genomics.

Conclusion

Storage is a major consideration in the development and deployment of AI workloads across a wide range of industries, such as financial services, telecommunications, healthcare, and many others. Making the right choice on storage technologies and, especially, storage solution partners is critical.

Selecting the wrong technology or storage solution provider can have significant negative implications in a range of areas, especially performance as storage needs accelerate and scale, ensuring high GPU performance, ease of deployment in increasingly complex storage environments, and storage capabilities that can keep up with the breakneck pace of innovation throughout the AI landscape.

Organizations should look for an AI-ready storage infrastructure platform, not just an ad hoc, mismatched collection of storage systems, software, and services assembled over time for disparate needs and use cases.

As a long-established, market-proven supplier of a wide range of storage solutions, Pure Storage is an excellent option for organizations looking to match their AI workloads with optimal storage features, capabilities, and performance. Pure Storage's AI storage platform delivers the enhanced functionality needed not only for today's AI workloads and use cases, but especially for the even-more-demanding workloads sure to emerge in the very near future and beyond.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com