

APRIL 2025

Pure Storage Deepens AI-at-scale Focus With FlashBlade//EXA

Simon Robinson, Principal Analyst

Overview

Pure Storage is doubling down on the enormous potential market around providing storage and data management for running AI at scale with the recent unveiling of FlashBlade//EXA, a high-performance storage solution designed to handle the most demanding AI and high-performance computing (HPC) workloads. The announcement underscores Pure's intent to become a force among the large- and hyper-scale players creating storage and data solutions for next-generation AI workloads.

Analysis

The almost limitless potential of AI continues to create vast opportunities for unleashing new value for almost every organization. Nearly half (44%) of respondents to recent research from Enterprise Strategy Group, now part of Omdia, said AI, data science and machine learning have become significantly more important to their organization's future over the past two years. Additionally, 29% of respondents said that supporting generative AI initiatives would be one of the most important initiatives to justify IT investments in 2025.¹

Yet, deploying AI at scale also presents unprecedented challenges to IT organizations charged with building an infrastructure that is capable of fully taking advantage of AI. New technology initiatives such as AI are already viewed as compounding an already-complex IT environment, with almost half of respondents (47%) reporting that they are experiencing a problematic shortage of AI and machine learning skills.²

Moreover, many organizations are finding plenty of devils in the details of running AI at scale. While a great deal of the focus around AI was necessarily concentrated on the compute environment, this is now expanding to incorporate the broader data environment.

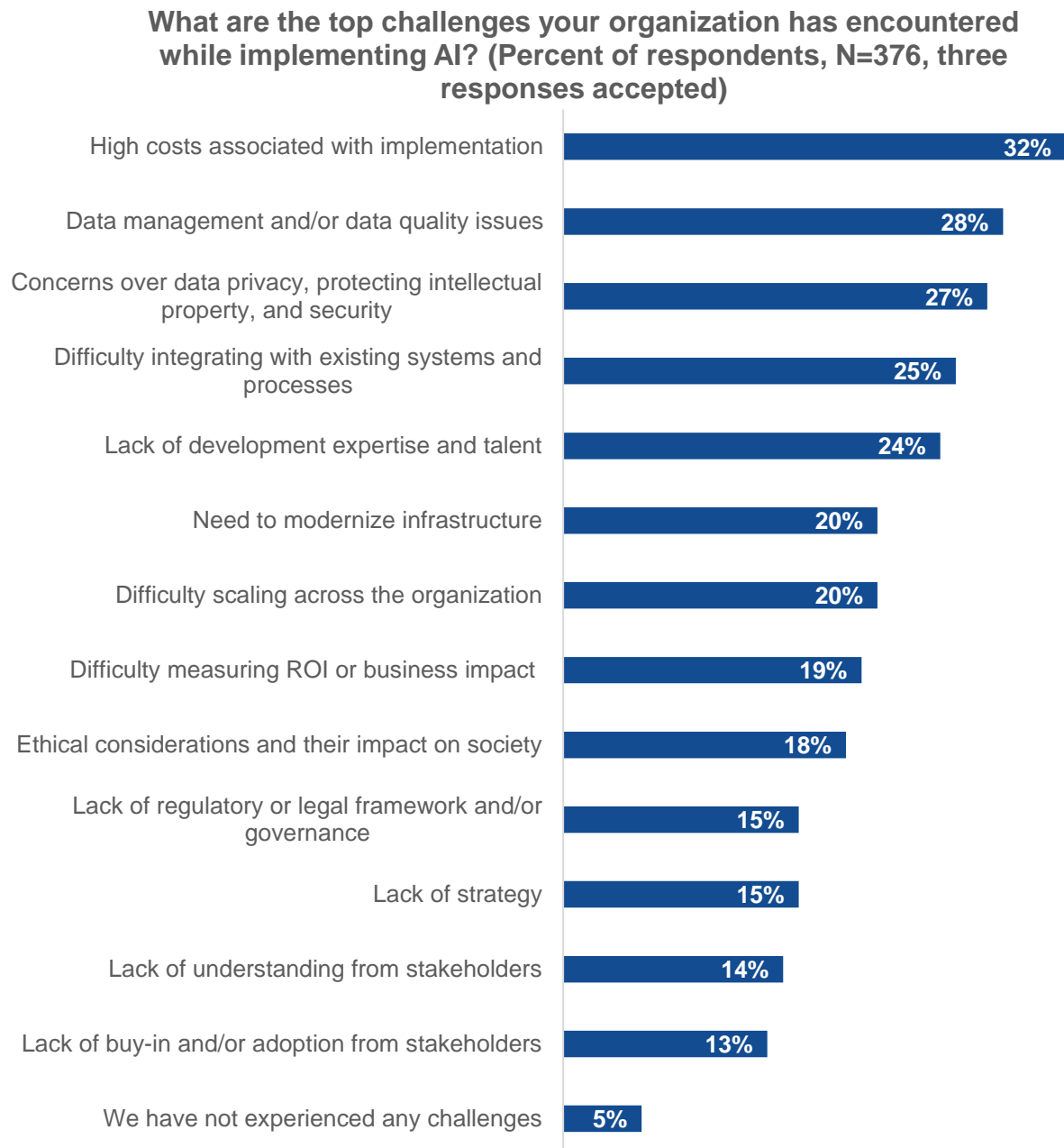
Data is increasingly viewed as the lifeblood of any AI initiative—the one aspect that can make the difference between success and failure. Getting the data aspect right at scale presents numerous substantial challenges across the broader data environment. A recent Enterprise Strategy Group research study found that data management and/or data quality issues were the second most frequently cited challenge associated with implanting AI, behind overall cost issues (see Figure 1).³ Concerns over data privacy, protecting IP, and security were also frequently cited, along with integration issues and the need to modernize infrastructure.

Many organizations on their AI journeys are, therefore, concluding that modernizing the infrastructure—right down to the storage environment—may be a necessary step to fully take advantage of AI.

¹ Source: Enterprise Strategy Group Research Report, [2025 Technology Spending Intentions Survey](#), December 2024.

² Ibid.

³ Source: Enterprise Strategy Group Research Report, [Navigating Build-versus-buy Dynamics for Enterprise-ready AI](#), January 2025.

Figure 1. Top Challenges With Implementing AI

Source: Enterprise Strategy Group, now part of Omdia

The Need to Scale AI Infrastructure: Introducing Pure Storage FlashBlade//EXA

The potential for a wide variety of organizations to build and deploy AI applications to support and run various aspects of their business is significant. Accordingly, storage solution providers, including Pure Storage, continue to focus on creating products and solutions aimed at enabling such environments. Pure's portfolio includes a range of AI-enabling capabilities; the company's FlashBlade//S product line is certified with NVIDIA DXG SuperPOD, for example.

Right now, however, the most consequential build-out of AI infrastructure is taking place among a relatively small number of hyperscalers, GPU-as-a-service specialists, and other “AI factories” such as advanced research facilities and government agencies. This is the space Pure is targeting with its latest offering, FlashBlade//EXA.

It’s a significant announcement for a couple of reasons. First, it’s a statement of intent that Pure believes its underlying technology is capable of serving not just the needs of mainstream enterprises, but also the taxing demands of organizations building AI at extreme scale.

Second, the FlashBlade//EXA architecture emphasizes the power and extensibility of its software. Though Pure is often associated with hardware-level innovations such as its DirectFlash Modules, it is a software company at heart. FlashBlade//EXA’s architecture emphasizes this, in part because it’s a modified variant of its Purity//FB software that runs on a combination of third-party servers and SSDs, as well as Pure’s own hardware.

Pure’s logic here is based on understanding the specific needs of these types of buyers and knowing where it can add the most value. Organizations building large- and hyper-scale AI factories have typically made massive investments in compute/GPUs and related hardware; what they really struggle with is building an infrastructure that can scale to extreme levels within a cost model that makes sense.

Data and Storage Emerge as Key Bottlenecks to Achieving AI at Scale

Many of the challenges with building this type of cost model fall squarely on the storage infrastructure, which is rapidly emerging as a key bottleneck in the entire AI process. The specific issues vary depending on the stage of the AI lifecycle and the type of AI workload, but some of the most consequential problems include:

- **Checkpointing.** Having GPUs sit idle as the system waits for checkpoints to be written to storage elongates model training times and can hugely impact the economics of the overall process.
- **Metadata performance.** Massive structured data “maps” of underlying unstructured data need to be constructed for an AI model to function at a scale that is simply unprecedented.
- **The general unpredictability of resource requirements.** The scale of AI workloads scale render tightly coupled approaches highly inefficient, especially when scale reaches extreme levels.

Pure’s argument is that existing HPC storage approaches, whether historically built around traditional parallel file systems or newer disaggregated approaches, are a sub-optimal fit because they were not fundamentally designed to address these challenges. Though they can meet high-scale performance requirements, they do so on the basis that the workload is well understood and predictable; even then, they are still not linearly scalable, Pure believes.

FlashBlade//EXA Overview

By contrast, Pure is designing FlashBlade//EXA with the assumption of extreme scale and unpredictability. It’s based on a disaggregated approach that separates the metadata functions from the physical data storage layer, such that both are independently scalable to extreme levels. The “Metadata Core” runs on a Pure FlashBlade array to store and manage metadata and is based on a massively scalable transactional database and key value store. This communicates with the GPU cluster, servicing metadata queries and directing the compute nodes to where the relevant data is stored.

Then, separately, the actual data is stored on a cluster of third-party data nodes, which runs a “thin” Linux-based OS with volume management and target services and is customized to work with the Metadata Core. These nodes are accessed directly by the GPU cluster at high speed via Remote Direct Memory Access (RDMA), using standard industry networking protocols (initially supporting NFS for file-level access at GA, though Pure also plans to add S3 support for object access in the future). In this way, customers can leverage their likely large pre-existing investments of 1U and 2U servers running NVMe SSDs, which can then independently scale separately from the Metadata Core as requirements change.

So what does “extreme scale” actually mean? Pure says that, while typical enterprise scale-out network-attached storage (NAS) taps out at around 1TB/sec of throughput and up to 100PB of data, FlashBlade//EXA has been designed to scale to 50TB/sec or more, supporting multi-exabytes of data and tens or even hundreds of thousands of GPUs. By any measure, these are simply enormous levels of scale.

Conclusion

The announcement of FlashBlade//EXA is a strong statement of intent by Pure and represents a substantial expansion of its strategic intent over the longer term. The total number of opportunities among AI factories and hyperscalers may be relatively small, but the opportunity within each is potentially huge. FlashBlade//EXA, therefore, represents a potentially significant increase in Pure’s TAM. Additionally, in targeting this segment, Pure will gain substantial experience and understanding of large-scale requirements, which may be crucial as more mainstream organizations contemplate building their own large-scale AI infrastructures.

The question, then, is whether Pure can establish itself as a credible player in this space. On this point, it already has; in recent months, the company has been talking up a design win with a “top four hyperscaler.” Pure recently revealed this win to be with Meta, with whom it is already extensively engaged in supporting its efforts in AI and beyond. The timing of FlashBlade//EXA, therefore, couldn’t be better. The market for AI storage infrastructure may just be heating up, but Pure is already off to the races.

©2025 TechTarget, Inc. All rights reserved. The Informa TechTarget name and logo are subject to license. All other logos are trademarks of their respective owners. Informa TechTarget reserves the right to make changes in specifications and other information contained in this document without prior notice.

Information contained in this publication has been obtained by sources Informa TechTarget considers to be reliable but is not warranted by Informa TechTarget. This publication may contain opinions of Informa TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent Informa TechTarget’s assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, Informa TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of Informa TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at cr@esg-global.com.

About Enterprise Strategy Group

Enterprise Strategy Group, now part of Omdia, 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