SOLUTION BRIEF

Pure Storage FlashBlade//EXA

Powering AI factories with the industry's most powerful data storage platform for AI.

Traditional storage architectures aren't built for the demands of modern AI and HPC. FlashBlade//EXA™ from Pure Storage® is the industry's most powerful data storage platform, built to power AI factories by delivering extreme throughput, low-latency metadata performance, and the ability to scale seamlessly. Its disaggregated metadata architecture eliminates constraints, enabling enterprises to train larger models, process multimodal datasets, and efficiently manage high-performance environments. By removing the constraints of legacy storage, FlashBlade//EXA as a true next-gen AI platform ensures AI-driven organizations can innovate without limits.

Industry-leading Performance

10+TB/sec read in a single namespace[1]

Write performance as high as 50% of reads[1]

Unmatched Performance Density

3.4TB/sec per rack

Proven Metadata Core

8+ years of proven metadata innovation

The Changing Landscape of AI and HPC

Artificial intelligence (AI) and high-performance computing (HPC) are at the forefront of technological advancement, driven by advancements in GPU technology and the diversification of workload requirements. As GPUs, including those like NVIDIA's Grace Blackwell architecture, become increasingly powerful, models grow larger, driving massive data growth across training, testing, and inference workflows. The escalating demands of handling ever-growing data at scale and diverse modalities—ranging from text and images to audio and video—are pushing existing storage architectures to their design limits in large-scale HPC and AI environments (see Table 1). These evolving workload requirements demand storage solutions that can provide the performance and scalability needed for modern computational challenges while avoiding additional complexity. A fresh approach to storage and data management is required to keep pace with these dynamic workload requirements in large-scale HPC and AI, and to fuel the next generation of innovators.

	Enterprise AI Workloads	AI Factories	Hyperscalers
Throughput	100GB/sec–1TB/sec	1TB/sec–50TB/sec	>50TB/sec
Capacity	50TBs–100PBs	100PB–Multiple EBs	10s to 100s of EBs
Total # of GPUs	<1,000 GPUs	1,000–10s of thousands of GPUs	> 10s of thousands of GPUs

TABLE 1 Evolution of scale requirements for AI and HPC

Current Storage System Limitations

Existing storage systems, including traditional parallel file systems and the first-gen disaggregation approach that decouples compute from data and metadata may be usable in some scenarios. However, massive parallel and concurrent reads and writes, metadata performance, ease of management and scalability, asynchronous checkpointing, and predictable high throughput are all key elements required for large-scale AI and HPC environments. Existing architectural approaches are hampered by limitations and bottlenecks that fail to deliver consistent performance at scale, while their ability to handle metadata efficiently and scale seamlessly is insufficient to support cutting-edge GPU clusters and meet the evolving needs of high-performance HPC and AI environments.

Building on a Proven Track Record in AI

Pure Storage has a proven track record of helping customers at any stage of their AI journey. Our commitment to innovation in AI and HPC is evident throughout our history of delivering robust solutions starting with the announcement of AIRI (AI-ready Infrastructure) in 2018, continuing with NVIDIA DGX BasePOD™ certification, NVIDIA SuperPOD™ certification, and introducing turnkey GenAI Pods. Pure Storage FlashBlade® has built a strong reputation in the enterprise AI market helping hundreds of enterprise customers such as Meta on their AI journeys. By integrating the lessons learned from our success with hyperscalers and leveraging our proven, innovative metadata architecture, Pure Storage is able to combine the positive elements of existing high-performance storage systems and create a compelling solution that meets the modern needs of large AI and HPC environments.

FlashBlade//EXA: Redefining Scalability and Performance for Large HPC and AI Environments

FlashBlade//EXA is a revolutionary offering that ensures large-scale AI and HPC customers have access to a modern data storage platform that no longer limits the pace of AI evolution, but instead accelerates it. Built to power AI factories, FlashBlade//EXA extends the Pure Storage FlashBlade family, offering a massively parallel processing architecture that disaggregates data and metadata, eliminating the constraints legacy systems impose on HPC and AI workloads and delivering extreme performance and scalability. It leverages the foundational strengths of FlashBlade and integrates Purity//FB's proven metadata architecture. FlashBlade//EXA complements the FlashBlade family that consists of the high speed FlashBlade//S™ and high density FlashBlade//E™ for enterprise workloads. Delivering unmatched throughput and simplicity at scale, FlashBlade//EXA addresses the requirements for the most demanding customer environments including AI Natives, Tech Titans, AI Enterprises, Specialized GPU Clouds, HPC labs and Research Centers. FlashBlade//EXA powers the next-generation data storage platform that can handle production, inference, and training even for the most demanding AI workloads.

Architectural Overview

The logical disaggregation of data and metadata in the architecture of FlashBlade//EXA means that it is effectively comprised of two scale-out clusters: Metadata Core and Data Nodes (see Figure 1). The metadata core is key to delivering scalable, low-latency metadata access and data nodes are key to delivering extreme throughput.

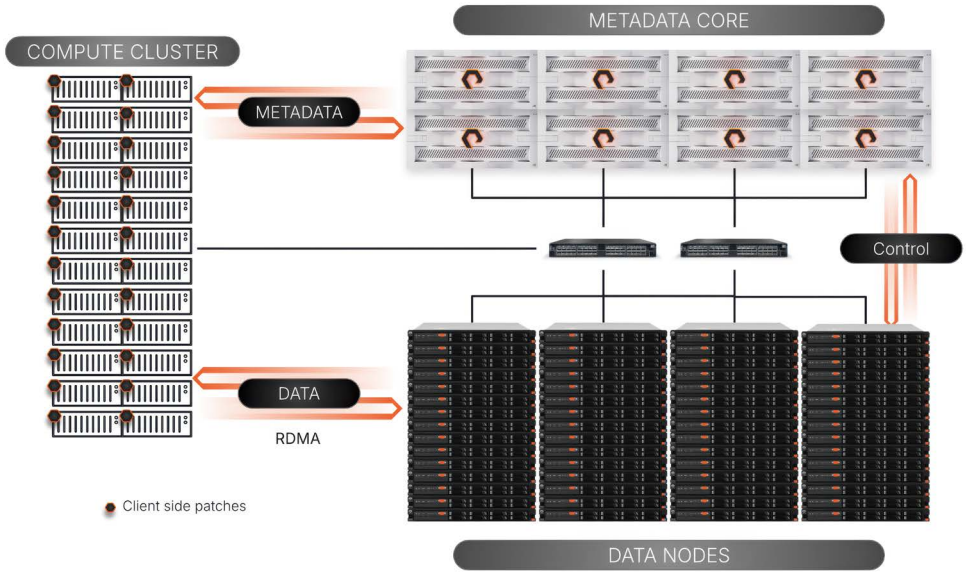

FIGURE 1 FlashBlade//EXA architecture overview

Metadata core: The metadata core in FlashBlade//EXA is designed to provide multidimensional performance and support the highest levels of scalability and resilience. This cluster leverages the proven Pure Storage metadata core to handle massive volumes of metadata requests across multiple metadata nodes efficiently (Figure 2). It eliminates bottlenecks associated with metadata scalability and concurrent metadata operations, ensuring synchronization across distributed metadata copies and optimizing complex file system operations while maintaining performance. The metadata nodes are not in the data path; however, they do interact with data nodes to monitor capacity and health. Leveraging pNFS, the metadata core communicates with the compute cluster using NFSv4.1 over TCP. The metadata cluster is built on proven FlashBlade technology and the Purity//FB operating system, which has been optimized with pNFS for FlashBlade//EXA. Its massively distributed transactional database and key value store technology, ensure high metadata availability and efficient scaling.

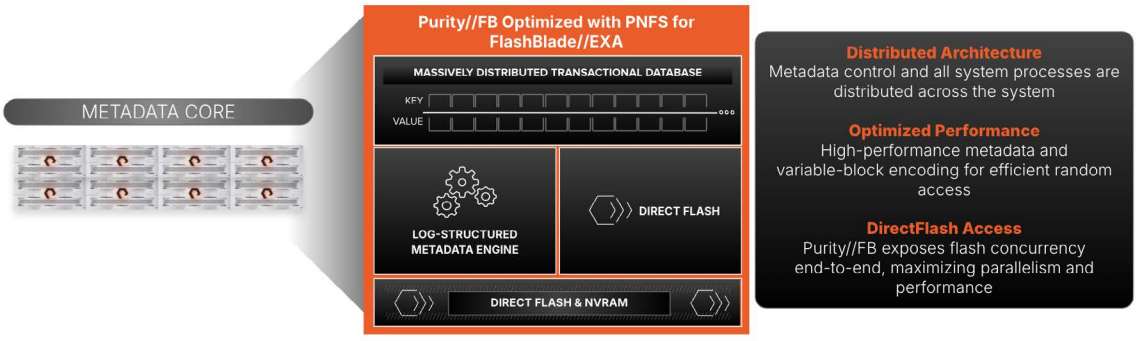

FIGURE 2 FlashBlade//EXA metadata core overview

Data nodes: FlashBlade//EXA data nodes are off-the-shelf (OTS) servers in the initial release, with plans to incorporate DirectFlash® Module (DFM) technology in future iterations. These generic servers will run a thin, Linux-based OS and kernel with volume management and RDMA target services optimized to work with metadata residing on the FlashBlade//EXA metadata nodes. This segregation allows for non-blocking data access with a direct data path between the compute cluster and data nodes. This means the overall system performance is limited only by network bandwidth and the aggregate rate at which data nodes can read and write data. This architecture supports the flexible scaling of storage performance to match evolving AI and HPC workload requirements.

Integrations with industry-standard technology: FlashBlade//EXA supports seamless integration with existing customer network infrastructure and offers flexibility in the data nodes configuration, allowing for use of hardware compatible with NVIDIA Spectrum™-X networking. It utilizes industry-standard Layer3/BGP routing to manage traffic between metadata, data, and workload clients, eliminating the complexity associated with multiple networking segments and IP addresses required by other solutions. The architecture also supports any compute cluster, ensuring compatibility with diverse GPU-driven AI environments. This approach emphasizes the use of industry-standard protocols to simplify management and enhance interoperability with various HPC and AI workloads.

FlashBlade//EXA Benefits

FlashBlade//EXA provides ground-breaking benefits, leveraging Pure Storage's innovative data architecture to address the escalating demands of modern AI and high-performance computing environments. FlashBlade//EXA:

- **Delivers industry-leading performance leveraging the Pure Storage proven metadata core:** FlashBlade//EXA stands out by achieving unprecedented performance levels, specifically designed to meet the challenges of modern large-scale HPC and AI workloads. FlashBlade//EXA can ensure data access with industry-leading read performance exceeding 10TB/sec[1] in a single namespace. The write performance of FlashBlade//EXA can be as high as 50% of reads[1]. This unmatched performance is underpinned by eight years of metadata core innovation. With the ability to support billions of metadata operations and more than 20 times the file systems in a single namespace compared to alternative solutions in the market, the multi-dimensional performance of FlashBlade//EXA powers the most advanced AI models with billions of parameters and multimodal data sets with high confidence. Enterprises can benefit from reduced data processing times and enhanced engagement with AI models, driving more immediate and insightful business outcomes.
- **Improves manageability at scale by eliminating traditional parallel storage system complexity:** FlashBlade//EXA redefines simplicity and efficiency in large-scale data environments through its sophisticated yet straightforward design, minimizing the complexity traditionally associated with parallel file systems. Installation times can be reduced by up to 50% compared to competitive systems. With an impressive performance density of 3.4TB/sec per rack, FlashBlade//EXA can optimize the evergrowing power and cooling costs associated with energy hungry GPU environments.
- **Empowers organizations to keep pace with AI innovation with a highly configurable and disaggregated architecture:** The architecture of FlashBlade//EXA enables seamless adaptation to evolving AI workloads, making it a flexible and future-proof investment for enterprises. The disaggregated architecture allows independent scaling of metadata and data nodes, thus eliminating traditional bottlenecks and ensuring optimal performance irrespective of workload changes. This configurability supports the rapid pace of AI model training and deployment, enabling organizations to stay ahead in the competitive landscape by continuously innovating and upgrading their AI capabilities.

Technical Specifications

FlashBlade//EXA Metadata Core

Scalability	1–10 chassis, 10 blades per chassis
Capacity	1–4 DFMs per blade, 37.5TB DFM
Connectivity with 2 XFMs	16 × 400GbE uplinks
Physical	**Metadata chassis:** Dimensions (per chassis): 5U Power: 2600 W (nominal at full configuration) **Pair of XFMs** Dimensions (per XFM): 1U Power: 310 W (nominal at full configuration)

FlashBlade//EXA Data Nodes

Scalability	Unlimited scalability
Minimum CPU, DRAM requirements	32 cores, 192GB DRAM
Capacity per Node	No. of NVMe drives: 12–16 PCIe Gen4+ Drive capacity: 3.8–61.44TB (PCIe Gen5 drives recommended for best performance)
Connectivity	For best performance 2 × 400Gb Ethernet NICs
Physical	Minimum dimensions: 1U Drive form factor: Determined by data node Power: Determined by data node

Additional Resources

- Discover FlashBlade//EXA for your large scale HPC and AI demands.

1 Based on Pure Storage performance testing with controlled hardware environment.

©2025 Pure Storage, the Pure Storage P Logo, DirectFlash, FlashBlade, FlashBlade//E, FlashBlade//EXA, FlashBlade//S, and the marks in the Pure Storage Trademark List are trademarks or registered trademarks of Pure Storage Inc. in the U.S. and/or other countries. The Trademark List can be found at purestorage.com/trademarks. Other names may be trademarks of their respective owners.

PS2733-01-en 03/25