PURESTORAGE®

REFERENCE ARCHITECTURE

Storage Reference Design for NVIDIA DGX B300 SuperPOD with FlashBlade//S R2

Contents

Executive overview

NVIDIA DGX SuperPOD with DGX B300 systems and Pure Storage® FlashBlade® is a turnkey data center solution for the AI enterprise, delivering scale-out storage optimized for AI. It is a multi-user system optimized for large-scale AI and high-performance computing (HPC) applications. It functions as a unified system, managing simultaneous users with advanced access controls for resource scheduling.

NVIDIA DGX SuperPOD with DGX B300 systems represents the next generation of scale-out compute clusters that demand the most from their storage fabrics. Newer protocols like RDMA over Converged Ethernet (RoCE) have further increased the need for simple and performant solutions that can handle the complexity and high-performance demands of today's AI workloads.

The data ingestion, processing, and storage requirements for large-scale enterprise AI require the high-performance, highly efficient storage that Pure Storage FlashBlade was architected to deliver.

Scaling to meet the demands of AI workloads is increasingly critical to pushing the boundaries of what's possible to achieve the next level of AI capability. Key features of this solution include:

- **Scalability**: It can scale up to thousands of NVIDIA GPUs and multiple petabytes of DirectFlash® storage, making it suitable for large-scale AI model training and inference.
- **Performance**: NVIDIA DGX SuperPOD with DGX B300 systems is based on the infrastructure built at NVIDIA for internal research purposes and is designed to solve the most challenging computational problems of today.
- **Software integration**: The solution leverages NVIDIA Mission Control for AI infrastructure operations and workload orchestration. The software simplifies AI operations for enterprises by delivering the tools for agility, resilience, and hyperscale efficiency.
- **Enterprise support**: Get dedicated expertise and services to help enterprises deploy and manage their AI infrastructure effectively.

This reference design includes next-generation **SR2 blades,** which are optimized for high-performance and unstructured data workloads. It delivers up to 50% better performance than the first generation—without any forklift upgrades. Ideal for AI/ML, electronic design automation, genomics, cyber resilience, and parallel analytics use cases, FlashBlade//S R2 helps enterprises stay ahead of data growth, maintain SLAs, and simplify operations through modular design and Evergreen® economics. Additionally, FlashBlade//S R2 provides enhanced platform security.

This NVIDIA DGX SuperPOD with DGX B300 systems reference design comes in two variations. The **standard DGX SuperPOD reference design** (without InfiniBand XDR) centers on a design leveraging **Spectrum™-4 Ethernet**, **DC busbar power**, and scalable units (SUs) of **64 NVIDIA DGX™ B300 systems**, emphasizing efficient power delivery via MGX racks and standard high-performance storage across InfiniBand or RoCE-enabled Ethernet fabrics. In contrast, the **B300-XDR reference design** explicitly integrates the **NVIDIA Quantum-X800 800Gb/s InfiniBand** platform as its core compute fabric, supports **72 DGX B300 systems per SU**, and focuses on ultra-high-speed connectivity and latency-optimized scaling for AI factories.

Solution overview

Integrating Pure Storage FlashBlade//S™ into the NVIDIA DGX SuperPOD with DGX B300 systems storage fabric provides an AI data storage platform with enterprise-grade data services, accelerating AI data pipelines through a parallel architecture and multidimensional performance. FlashBlade allows capacity and performance to be scaled independently and continuously improved over time without disruption to meet the changing and growing needs of an NVIDIA DGX SuperPOD with DGX B300 systems environment. This includes benefits such as:

- **Data-in-place, nondisruptive upgrades:** Nondisruptively upgrade storage processor and/or storage capacity resources independently to either scale them within a technology generation or upgrade them to newer generations. With FlashBlade, customers can enjoy continuous, data-in-place hardware and software innovation without planned downtime and disruption windows.
- **Ease of use and management**: Get a single integrated solution for all stages of an AI data pipeline with a simple architecture that is easy to deploy, manage, and use.
- **Environmental efficiency**: Gain a more powerful, ESG-friendly AI infrastructure that delivers high density, reduced power consumption, and unmatched data center efficiency.
- **Performance at any scale**: Start your AI journey at any scale and seamlessly grow and adapt as needed.
- **Agile AI platform**: Accelerate your advantage with a platform built for future innovations by Pure Storage and NVIDIA, empowering you to adapt and lead as the AI landscape and customer requirements evolve.
- **Trusted industry leaders**: Partner with innovation leaders in the market that have jointly delivered state-of-the-art, flash-based, AI-ready infrastructure since the introduction of the Pure Storage AIRI® solution in 2018.

This platform handles a multitude of use cases like large language models (LLMs), generative AI, computer vision, deep learning recommendation systems, HPC, and more.

Pure Storage FlashBlade DGX B300 SuperPOD solution integration

FlashBlade connects to an NVIDIA DGX SuperPOD environment via 16 400GbE ports. This allows the FlashBlade to seamlessly scale from a minimal configuration into multiple chassis, blades, and DirectFlash Module configurations in a simple, nondisruptive manner.

NVIDIA DGX SuperPOD clusters are transforming how data is processed and analyzed, especially when integrated with advanced storage solutions. These storage systems play a crucial role in managing the large amounts of data that AI applications require, enabling seamless access and sharing for distributed computing environments. This combination allows for efficient collaboration among nodes within a compute cluster, ensuring that AI models can be trained on large data sets without delays.

The integration of Pure Storage FlashBlade with NVIDIA DGX SuperPOD compute clusters significantly enhances these capabilities. For example, in healthcare, AI can analyze extensive medical records and imaging data, with centralized storage providing quick retrieval and processing. In the financial sector, these systems support the storage of historical market data, allowing AI models to rapidly analyze trends and make predictions. Additionally, industries like media and entertainment benefit from this integration by managing large volumes of video and audio data for AI-driven content creation and editing.

Ethernet networking is a cornerstone of enterprise data centers due to its low total cost of ownership, extensive interoperability, and proven reliability. When FlashBlade//S is connected via NVIDIA Spectrum-X Ethernet to the DGX storage system fabric, it delivers exceptional performance for large AI clusters. Additionally, the use of RoCE further enhances this setup by minimizing latency and offloading CPU workloads, resulting in faster data transfers and improved overall efficiency for demanding AI applications.

As the demand for AI-driven insights continues to grow, the reliance on powerful compute clusters alongside efficient storage solutions will expand, paving the way for innovation across diverse AI fields.

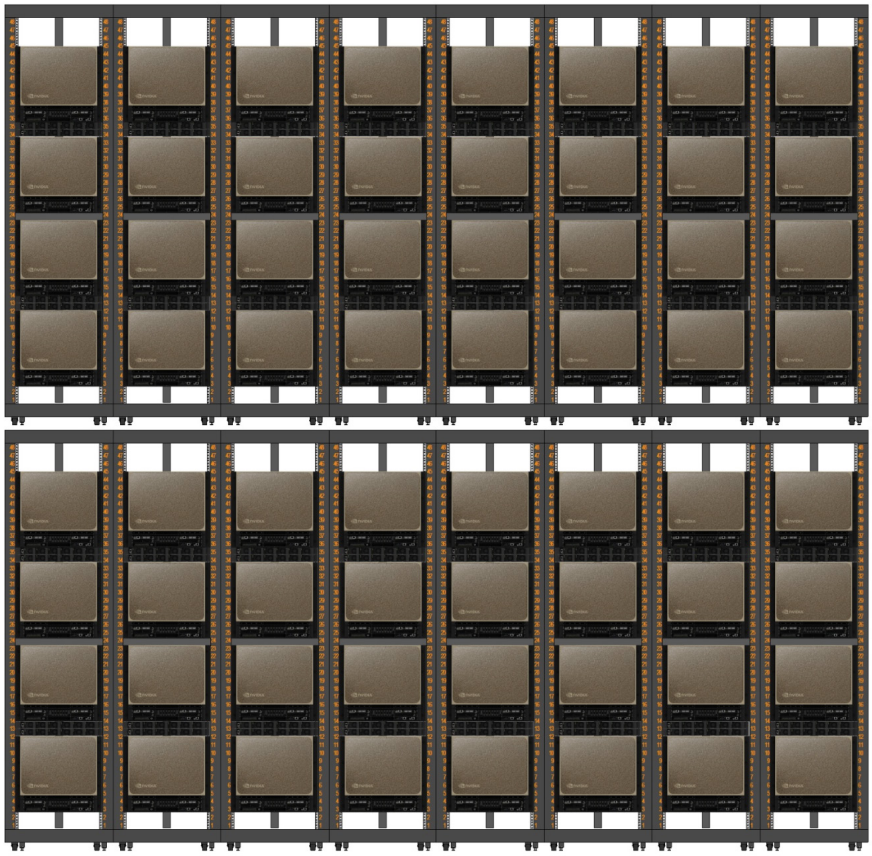

FIGURE 1 64 NVIDIA DGX B300 Busbar systems in MGX racks each with DC power shelves and MGX rack stiffeners (NVIDIA Spectrum-X Ethernet-based compute fabric)

FIGURE 2 Example management rack configuration with networking switches, management servers, storage arrays, and UFM appliances (sizes and quantities will vary depending upon models used)

FIGURE 3 72 NVIDIA DGX B300 PS systems in standard racks each with three 2U rack power distribution units (PDUs) for maximum redundancy (you may need to reduce the number of DGXs hosted on the same rack depending on your data center's capability)

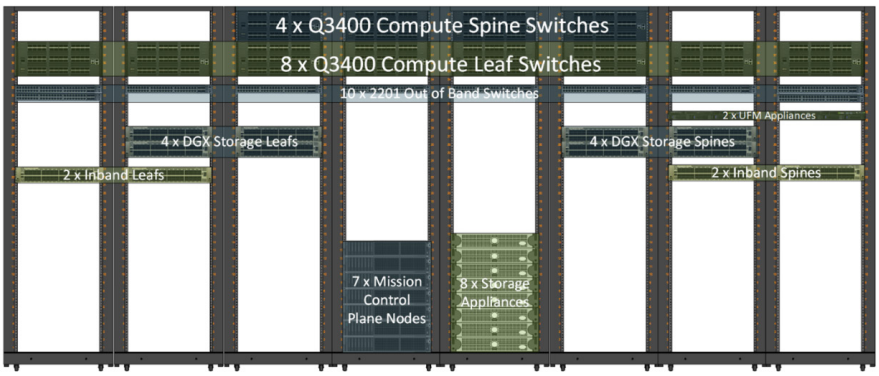

FIGURE 4 Example management rack configuration for 1SU with networking switches, management servers, storage arrays, and UFM appliances (sizes and quantities will vary depending upon models used)

About FlashBlade

FlashBlade is an industry-leading, unified fast file and object storage platform that Pure Storage engineered from the ground up for flash to deliver the cost-effective reliability, simplicity, and multidimensional performance needed to address the demands of exabyte-level modern data sets.

A high-performance, consolidated storage system for both file and object workloads, FlashBlade delivers a simplified experience for infrastructure and data management. It's a system that's easy to buy, set up, change, and improve—in moments, not months.

FlashBlade enterprise features

FlashBlade//S enables enterprise-level data management for AI at scale. Using a distributed metadata architecture, it offers multidimensional performance on a consolidated platform with Network File System (NFS), Server Message Block (SMB), and Simple Storage Service (S3) protocol access. The cloud-based Pure1® data management platform provides a single view to monitor, analyze, and optimize your storage from anywhere. Key capabilities provided by FlashBlade//S and the Purity operating system for AI and other workloads include:

Multidimensional performance: This is the ability to deliver very high throughput and high performance to support multiple workloads simultaneously, including workloads with:

- Any file size (small or large)
- Sequential or random I/O
- Batched or real-time jobs
- A very large number of files
- Metadata performance for file access (10s of millions of file operations per second)

Intelligent architecture: FlashBlade was built from the ground up to truly leverage the performance and efficiencies of flash via our DirectFlash architecture. It is simple to deploy, manage, and upgrade without requiring constant tuning.

Nondisruptive upgrades: Maintenance operations, hardware and software upgrades, and performance and capacity expansions are completed without disruption.

Always-on availability: Data is the lifeblood of your AI pipeline. FlashBlade provides enterprise-grade availability and resiliency to ensure high availability every day without planned downtime.

Dynamic scalability: Seamlessly scale not only capacity but also performance, metadata, number of files and objects, and more.

Cloud extensibility: Connect workflows from on-premises to public cloud environments.

Zero move tiering: This enables intelligent data access across heterogeneous all-flash clusters that optimize both performance and cost.

Multiprotocol support: FlashBlade in DGX SuperPOD is tuned for NFS over RoCE; FlashBlade also natively supports S3 and SMB.

Reliability: This solution keeps organizations up and running with:

- Zero planned downtime
- 99.9999% uptime SLA with Evergreen//One™
- Fewer manual interventions, which leads to less human errors

FlashBlade NVIDIA DGX SuperPOD networking

FlashBlade integrates effortlessly with NVIDIA DGX SuperPOD clusters using established Ethernet connectivity. It further enhances performance by supporting RoCE communication between DGX systems and FlashBlade, ensuring low-latency data transfer. Connecting FlashBlade to a pair of NVIDIA Spectrum-X SN5600 Ethernet switches follows a standard installation process that has been successfully implemented in thousands of installations worldwide.

A FlashBlade array is composed of four key components: external fabric modules (XFMs), chassis, blades, and direct fabric modules (DFMs). The XFMs aggregate all traffic from the SN5600 switches to the blades, as illustrated in Figure 5.

The chassis houses the blades and associated DFMs, facilitating connectivity among the blades and to the broader network. Each blade can accommodate up to four DFMs and serves as an endpoint for client communications. It is important that all storage provided by the DFMs is accessible to every blade within the cluster, with no file system constraints on individual blades. This architecture allows any blade in the array to access and service client requests as needed, maximizing efficiency and resource utilization.

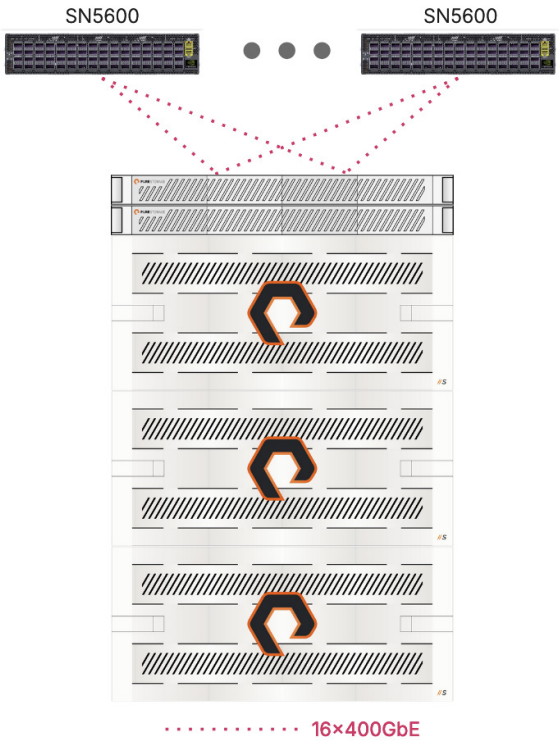

FIGURE 5 NVIDIA DGX SuperPOD storage fabric network integration

The FlashBlade NVIDIA DGX SuperPOD system components and FlashBlade DGX SuperPOD Accessory Kit bill of materials can be found in Appendix A.

FlashBlade network connectivity

Aggregate network connectivity to the NVIDIA DGX SuperPOD storage fabric is provided by dual FlashBlade XFM-8400 switches, shown in Figure 6 as ports 21–28. Each XFM uplinks to the NVIDIA Spectrum-X SN5600 Ethernet switches with eight connections each for a total of 16 400Gb/E connections.

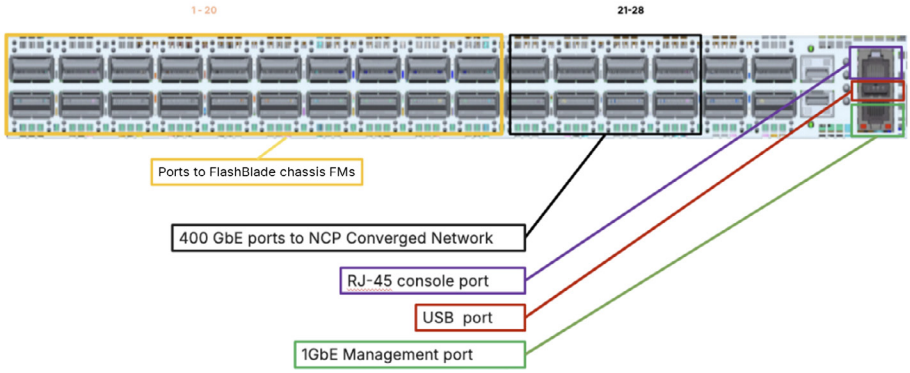

FIGURE 6 XFM port map

Aggregate data traffic connections to the array are passed to the chassis and then blades within each chassis, fully distributing all connections across all blades in the array. All blades within the array have access to all storage available across the array and can respond to any client request.

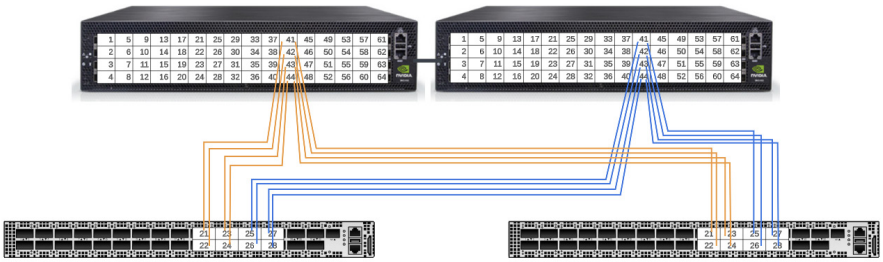

FIGURE 7 Example of aggregate connection uplinks from XFMs to the NVIDIA DGX SuperPOD SN5600 storage fabric

NVIDIA DGX SuperPOD scaling

DGX SuperPOD to FlashBlade scaling is based on a single rack to start. As the cluster grows, the FlashBlade can be expanded to provide additional storage and I/O capabilities. Tables 1 and 2 capture sample FlashBlade deployments for DGX SuperPOD SUs 1 to 8. Contact your Pure Storage representative for complete NVIDIA DGX SuperPOD sizing options.

SU	Namespaces	FlashBlade chassis per namespace	Capacity (raw) in TB	RU	Racks	Power (kW)	FlashBlade (NSs x total blades x DFMs per blade x DFM capacity
1	1	2	1651.6	12	1	4.996	1×15×3×37T
2	1	3	3303.2	17	1	8.941	1×30×3×37T
4	1	7	7707.4	37	1	20.025	1×70×3×37T
8	1	10	11010.5	52	2	28.338	1×100×3×37T

TABLE 1 Sizing with SR2 blades (standard performance)—DGX B300.

SU	Namespaces	FlashBlade chassis per namespace	Capacity (raw) per TB	RU	Racks	Power (kW)	FlashBlade (NSs x total blades x DFMs per blade x DFM capacity
1	1	5	5505.3	27	1	14.483	1×50×3×37T
2	1	10	11010.5	52	2	28.338	1×100×3×37T
4	2	10	22021.0	104	3	56.676	2×100×3×37T
8	4	10	44042.0	208	5	113.352	4×100×3×37T

TABLE 2 Sizing with SR2 blades (enhanced performance)—DGX B300.

Storage performance guidance from NVIDIA for DGX B300

NVIDIA provides guidelines for NVIDIA DGX SuperPOD with DGX B300 systems workload I/O based on Tables 3 and 4. The numbers are meant to represent the best case scenario and can be used to help understand the workload that a single rack would generate in terms of I/O load.

SU	Number of DGX systems	Standard performance Read (GB/s)	Standard performance Write (GB/s)	Enhanced performance Read (GB/s)	Enhanced performance Write (GB/s)
1	64	80	40	250	124
2	128	160	80	500	248
4	256	320	160	1000	496
8	512	640	320	2000	992

TABLE 3 DGX SuperPOD (NVIDIA Spectrum-X Ethernet-based compute fabric) performance scaling guidance

SU	Number of DGX systems	Standard performance Read (GB/s)	Standard performance Write (GB/s)	Standard performance Read (GB/s)	Standard performance Write (GB/s)
1	72	80	40	250	124
2	144	160	80	500	248
4	288	320	160	1000	496
8	576	640	320	2000	992

TABLE 4 DGX SuperPOD (NVIDIA Quantum-X800 InfiniBand (IB)/XDR-based compute fabric) performance scaling guidance.

High-resolution computer vision data sets can exceed 30TB, requiring ~4GBps per GPU of read performance. Natural language processing and LLM training usually need less read bandwidth, but checkpointing demands peak read/write throughput since training pauses during this synchronous phase. For best results, allocate at least half of the recommended read performance as write performance in LLM or large-model cases. Because workloads vary, always benchmark and size performance and capacity to your organization's specific needs.

Summary

Pure Storage FlashBlade was purpose-built from the ground up to fully take advantage of flash for scale-out workloads and is optimized to support the data and GPU requirements for large-scale AI initiatives. FlashBlade allows capacity and performance to be scaled independently and continuously improved over time without disruption. This unique modular storage architecture offers flexible consumption choices while minimizing waste by using less power and space. The distributed storage architecture of FlashBlade matches the distributed scale-out compute infrastructure of NVIDIA DGX SuperPOD clusters. FlashBlade delivers simple management; space and energy efficiency; and consistent performance across different data types, I/O profiles, and AI workloads—all while remaining future-proof to support evolving AI requirements.

Additional information

- FlashBlade//S
- Pure Storage platform for AI

Appendix A: Deployment guide for Pure Storage FlashBlade with NVIDIA DGX SuperPOD—FlashBlade components

Tables 5 and 6 show an example configuration for a 1SU NVIDIA DGX SuperPOD.[1] It is recommended to work with a Pure Storage representative to optimally configure the Pure Storage FlashBlade//S for NVIDIA DGX SuperPOD with DGX B300 systems clusters.

FlashBlade component list for 1SU

Item	Quantity
FM-8400R2	1 pair (2 total)
FlashBlade//S500 R2 chassis (1×50×3×37T)	5 (chassis)

TABLE 5 FlashBlade DGX B300 SuperPOD bill of materials

Accessory Kit list for 1SU

Item	Quantity
FB-Higher-Performance-400G-DR4-Accessory-Kit	10
FB-XFM-8400-400G-MC-Accessory-Kit	5

TABLE 6 FlashBlade DGX SuperPOD Accessory Kit bill of materials

1 | https://docs.nvidia.com/dgx-superpod/reference-architecture/scalable-infrastructure-b300/latest/storage-architecture.html

ID1196-01-en-11/25