NVIDIA DGX SuperPOD: Next Generation Scalable Infrastructure for AI Factories

Reference Architecture

Featuring NVIDIA DGX GB300 Systems

Abstract

The NVIDIA DGX SuperPOD architecture has been designed to power the next-generation AI factories with unparalleled performance, scalability, and innovation that supports all customers in the enterprise, higher education, research, and the public sectors. It is a physical twin of NVIDIA's own system used by research and development teams, meaning the DGX SuperPOD customer's infrastructure software, applications, and support have already been tested and vetted on the same architecture through NVIDIA's internal usage.

This DGX SuperPOD Reference Architecture (RA) is based on DGX GB300 systems powered by Grace CPUs and Blackwell Ultra GPUs. The RA discusses the components that define the scalable and modular architecture of DGX SuperPOD. DGX SuperPOD is built on the concept of scalable units (SU); each SU contains 8 DGX GB300 systems, which enables rapid deployment of DGX SuperPOD of any size. The RA also includes details regarding the SU design and specifics of InfiniBand, NVLink network, Ethernet fabric topologies, storage system specifications, recommended rack layouts, and wiring guidelines.

This RA combines the latest NVIDIA technologies to help companies and industries develop their own AI factories. To achieve the most scalability, DGX SuperPOD is powered by several key NVIDIA technologies and solutions, including:

- NVIDIA DGX GB300 systems: Powerful computational building block for AI and HPC.
- NVIDIA QUANTUM-X800 (XDR/800 Gbps) InfiniBand: High performance, low latency, and scalable network interconnect.
- NVIDIA Spectrum-4 (800 Gbps) Ethernet: High performance, low latency, and scalable ethernet connectivity for storage.
- NVIDIA fifth-generation NVLink® technology: a high-speed interconnect for CPU and GPU processors in accelerated compute systems to provide unprecedented performance for most demanding communication patterns.
- NVIDIA Mission Control: a unified operations and orchestration software stack for managing AI factories.

DGX SuperPOD requires a supported storage solution from one of the validated and certified storage vendor from the DGX Storage Program. A full list of supported storage vendors are listed on the DGX SuperPOD landing page.

Contents

DGX SuperPOD Architecture

The DGX SuperPOD architecture is a combination of DGX systems, InfiniBand and Ethernet networking, management nodes, and storage. Figure 1 shows the rack layout of a single SU (Scalable Unit). Each SU requires a Thermal Design Power (TDP) of 1.2 Megawatts (MW). Generally, the data center should meet or exceed Uptime Institute Tier 3 design standards, or alternatively the TIA942-B Rated 3 or EN50600 Availability Class 3 design standards, including concurrent maintainability and no single point of failure.

Each DGX GB300 leverages hybrid cooling with both direct liquid cooling and air cooling to manage the amount of heat produced by each rack. Given DGX SuperPOD with DGX GB300 is a turnkey AI data center scale product, NVIDIA has provided datacenter design guidance and fully integrated operational technology (OT) integration with NVIDIA Mission Control software stack.

Figure 1. Sample Design for DGX SuperPOD GB300

This reference architecture is showcases the scale-out capabilities of the DGX SuperPOD up to configurations up to 128 racks with 9216 GPUs.

Contact your NVIDIA representative for information regarding DGX SuperPOD solutions of four SUs or more.

Key Components of DGX SuperPOD

The DGX SuperPOD architecture is designed to meet the high demands of AI factories for the era of reasoning. AI factories require specialized components like high-performance GPUs and CPUs, advanced networking, and cooling systems to support the intensive computational needs of AI workloads. These factories excel at AI reasoning—enabling faster, more accurate decision-making across industries. And using NVIDIA's end-to-end accelerated computing platform, they're optimized for energy efficiency while accelerating AI inference performance, helping enterprises deploy secure, future-ready AI infrastructure.

DGX SuperPOD is the integration of key NVIDIA components, as well as storage solutions from partners certified to work in a DGX SuperPOD environment. By leveraging the concept of Scalable Units (SUs, as defined below in this document), DGX SuperPOD reduces AI factory deployment times from months to weeks, which in turn reduces time-to-solution and time-to-market of next generation models and applications.

DGX SuperPOD is to be deployed on-premises, meaning the customer owns and manages the hardware. This can be within a customer's data center or co-located at a commercial data center, but the customer owns the hardware, the service it provides, and is responsible for their cluster infrastructure.

The key components of DGX GB300 SuperPOD are shown below, each component will be discussed in detail:

- NVIDIA DGX GB300 NVL72 rack system
- NVIDIA QUANTUM-X800 InfiniBand
- NVIDIA Spectrum Ethernet
- NVIDIA Mission Control software

NVIDIA DGX GB300 Rack System

NVIDIA DGX GB300 (Figure 2) is an AI powerhouse that enables enterprises to expand the frontiers of business innovation and optimization. Each DGX GB300 delivers breakthrough AI performance in a rack-scale, 72 GPU configuration. The NVIDIA Blackwell GPU architecture provides the latest technologies that brings

months of computational effort down to days and hours, on some of the largest AI/ML workloads.

To accommodate the high compute density per rack and at the same time to more efficiently use datacenter space, the DGX GB300 rack system employs a sophisticated hybrid cooling solution to manage the substantial heat generated by the powerful GPUs and other components. The most power-intensive components, such as GPUs and CPUs, are liquid cooled. Other components are air cooled. NVIDIA's Infrastructure Specialist team can provide end-to-end assistance for datacenter planning, system deployment, and bring-up services.

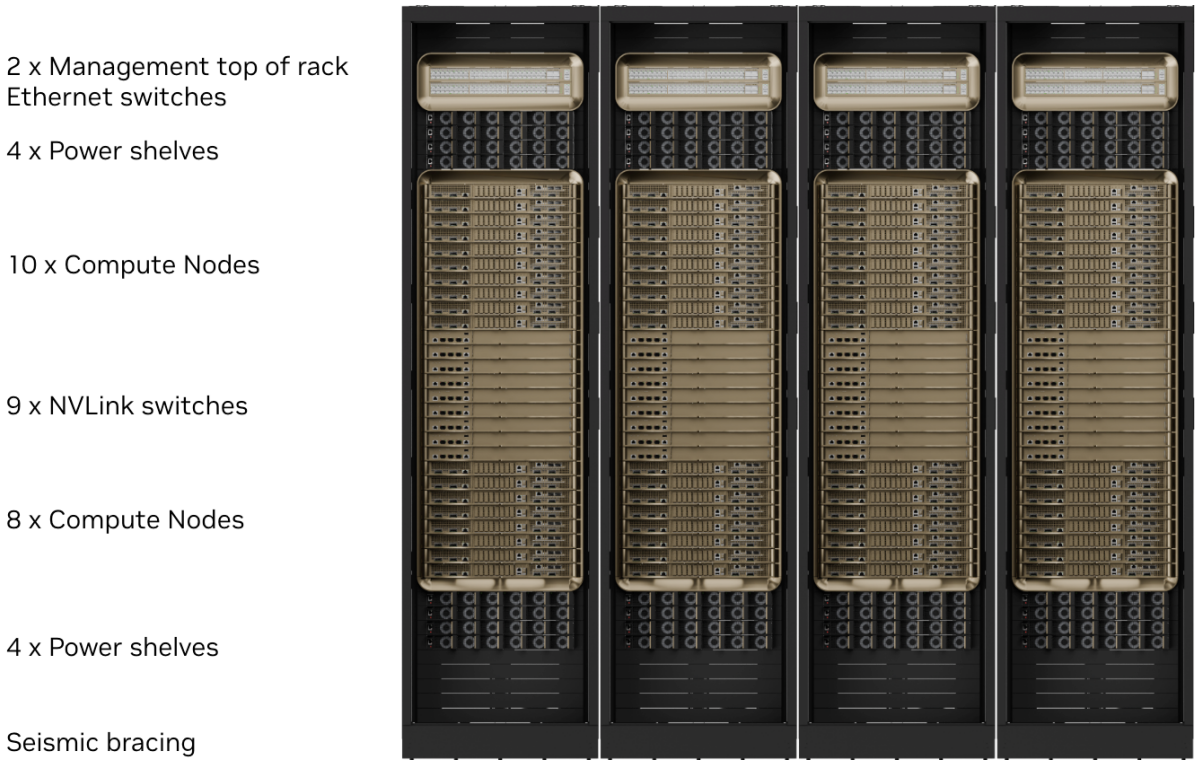

Figure 2. DGX GB300 (4 racks are shown here)

DGX GB300 Compute Tray

The compute nodes for DGX GB300 rack system are using the NVL32 topology, which contains 72 GPUs in a single NVLink domain. There are 18 compute nodes (aka. compute trays) in a DGX GB300 rack system. Each compute tray contains two GB300 Superchips, and each Superchip has two B300 GPUs and one Grace CPU. A coherent chip-to-chip interconnect link, called NVLink-C2C, bridges the

two Superchips and enables them to act as a single logical unit for a single OS instance to operate.

The compute tray integrates four ConnectX-8 NICs to support InfiniBand QUANTUM-X800 (800Gbps) connectivity for the inter-racks Compute network, and one BlueFiled-3 NICs to support 2x400Gbps connectivity for the In-band Management and Storage networks. All network ports are located at the front-side of the rack, intentionally for the cold (or hot) aisle.

Each compute tray also provides a total of 4x 3.84TB E1.S NVMe as RAID0 for local storage and 1x 1.92TB M.2 NVMe for the OS image.

Figures 3 and 4 show the front and rear of a DGX GB300 compute tray respectively.

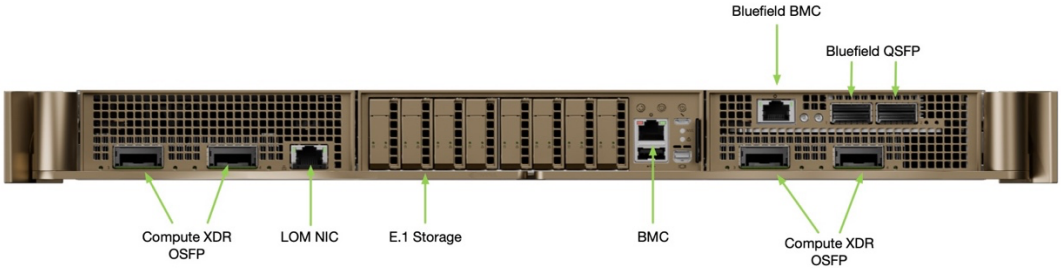

Figure 3. GB300 Compute Tray Front

Figure 4. GB300 Compute Tray Rear

Figure 5 shows the block diagram of the GB300 compute tray with two GB300 Superchips, each chip combines two NVIDIA B300 Tensor Core GPUs and one NVIDIA Grace CPU, connected over a 900GB/s ultra-low-power NVLink-C2C interconnect.

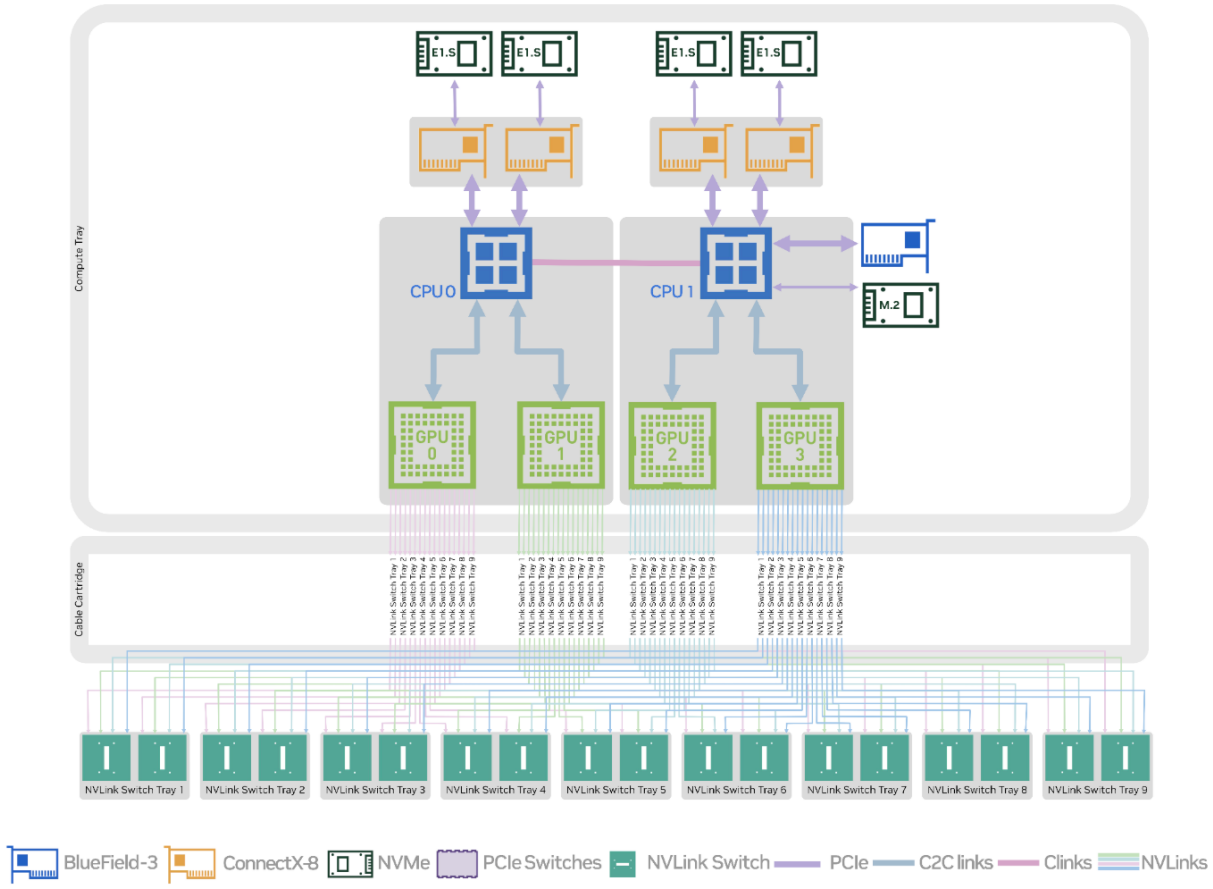

Figure 5. GB300 Compute Tray Block Diagram

NVLink Switch Tray

Each DGX GB300 NVL72system rack also features nine NVLink switches (aka. NVLink switch tray). All NVLink interconnections are done through the rear-side, blind-mate connectors to the cable cartridges within the rack.

Each switch tray features two COMe RJ45 ports for NVOS and one BMC module for out-of-band management with one RJ45 port. Figure 4 showcases the front design of a NVLink Switch Tray.

Figure 6. NVLink Switch Tray

DGX Power Shelves

The power shelf used for DGX GB300 SuperPOD has six 5.5kW PSUs configured as N redundant and can deliver up to 33kW of power. There are eight total power shelves in a single DGX GB300 NVL72 rack system. In comparison to previous generation DGX Power Shelves, the DGX GB300 power shelf is equipped with larger energy storage to reduce the peak load requirement on datacenter power supply.

At the rear of the power shelf is a set of RJ45 ports used for power brakes and other out-of-band management capabilities. At the front of the power shelf is the BMC port. Figure 7 shows the front of the power shelf.

NVIDIA DGX Power shelves includes deep integration

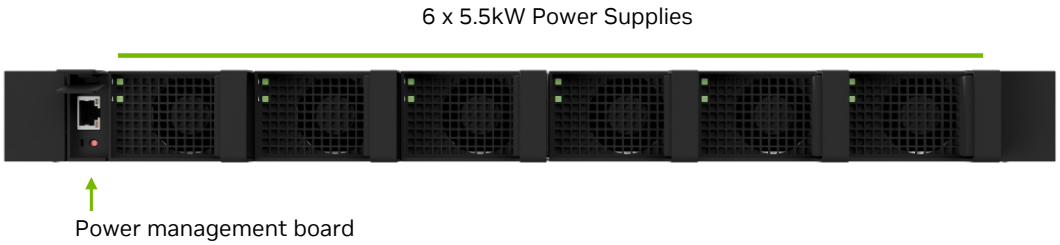

Figure 7. DGX Power Shelf

NVIDIA InfiniBand Technology

InfiniBand is a high-performance, low latency, RDMA capable networking technology, proven over 20 years in the harshest compute environments to provide the best inter-node network performance. InfiniBand continues to evolve and lead data center network performance.

The latest generation InfiniBand, Quantum-X800, has a peak speed of 800 Gbps per direction with an extremely low port-to-port latency. It is backwards compatible with the previous generations of InfiniBand specifications. InfiniBand is more than just peak bandwidth and low latency. It provides additional features to optimize performance including adaptive routing (AR), collective communication with SHARP™, dynamic network healing with SHIELD™, and supports several network topologies including fat-tree, Dragonfly, and multi-dimensional Torus to build the largest network fabrics and computer systems possible.

NVIDIA Mission Control

The DGX GB300 SuperPOD Reference Architecture represents the best practices for building high-performance AI factories. There is flexibility in how these systems can be presented to customers and users. NVIDIA Mission Control software is used to manage all DGX GB300 SuperPOD deployments.

NVIDIA Mission Control is a sophisticated full-stack software solution. As an essential part of the DGX SuperPOD experience, it optimizes developer workload performance and resiliency, ensures continuous uptime with automated failure handling, and provides unified cluster-scale telemetry and manageability. Key features include full-stack resiliency, predictive maintenance, unified error reporting, data center optimizations, cluster health checks, and automated node management.

Mission Control software incorporates the same technology that NVIDIA uses to manage thousands of systems for our award-winning data scientists and provides an immediate path to a TOP500 supercomputer for organizations that need the best of the best.

DGX SuperPOD is to be deployed on-premises, meaning the customer owns and manages the hardware. This can be within a customer's data center or co-located at a commercial data center, but the customer owns the hardware, the service it provides, and is responsible for their cluster infrastructure as well as providing the building management system for integration.

Components

The hardware components of DGX GB300 SuperPOD are described in

Table 1. The software components are shown in Table 2**.**

Table 1. DGX SuperPOD / 1 Scalable Unit infrastructure

Component	Technology	Description
8x Compute Racks	NVIDIA DGX GB300 NVL72	The world's premier rack-scale, purpose-built AI systems featuring NVIDIA Grace-Blackwell Ultra GB300 modules. GB300 Superchips. NVL72 scale-out GPU interconnect and integrated NVLink switch trays.
NVLink Fabric	NVIDIA NVLink 5	NVLink Switches supports fast, direct memory access between GPUs on the same compute rack.
Compute Fabric	NVIDIA Q3400 InfiniBand Switch	Rail-optimized, non-blocking, full fat-tree network with four Quantum X-800 connections per system for inter-rack GPU communications.
Storage and In-band Management Fabric	NVIDIA Spectrum 4 SN5600 Ethernet switches	The fabric is optimized to match peak performance of the configured storage array, built with 64 port 800 Gbps Ethernet switch providing high port density with high performance.
InfiniBand management	NVIDIA Unified Fabric Manager Appliance, Enterprise Edition	NVIDIA UFM combines enhanced, real-time network telemetry with AI powered cyber intelligence and analytics to manage scale-out InfiniBand data centers
NVLink Management	NVIDIA Network Manager eXperience (NMX-M)	NVIDIA NMX Manager manages, operates the NVLink switches and provides real-time network telemetry to manage all NVLink infrastructure
Out-of-band (OOB) management network	NVIDIA SN2201 switch	48 x 1 Gbps Ethernet switch leveraging copper ports to minimize complexity

Table 2. DGX SuperPOD software components

Component	Description

Mission Control Software	NVIDIA Mission Control software delivers a full stack data center solution engineered for NVIDIA DGX SuperPOD with DGX GB300 infrastructure deployments. It integrates essential management and operational capabilities into a unified platform, thereby providing enterprise customers with simplified control over their NVIDIA DGX SuperPOD infrastructure deployments at scale. NVIDIA Mission Control also leverages NVIDIA Run;ai functionality, providing seamless workload orchestration.
NVIDIA AI Enterprise	NVIDIA AI Enterprise is an end-to-end, cloud-native software platform that accelerates data science pipelines and streamlines development and deployment of production-grade LLMs and other generative AI applications.
Magnum IO	Enables increased performance for AI and HPC
NVIDIA NGC	The NGC catalog provides a collection of GPU-optimized containers for AI and HPC
Slurm	A classic workload manager used to manage complex workloads in a multi-node, batch-style, compute environment
Run:AI	NVIDIA Run:ai accelerates AI operations with dynamic orchestration across the AI lifecycle. It maximizes GPU utilization, scales AI workloads efficiently, and integrates seamlessly into hybrid infrastructure with minimal overhead. Browse the documentation to install and monitor your environment, manage resources and organizations, run and scale AI workloads, and integrate NVIDIA Run:ai into your workflows using APIs.

Design Requirements

DGX SuperPOD is designed to minimize system bottlenecks throughout the tightly coupled configuration to provide the best performance and application scalability. Each subsystem has been thoughtfully designed to meet this goal. In addition, the overall design remains flexible so that SuperPOD can be tailored to better integrate into existing data centers.

System Design

DGX SuperPOD is optimized for a customers' particular workload of multi-node AI and HPC applications:

- A modular architecture based on Scalable Units (SUs) of 8 DGX GB300 rack-scale systems.

> A fully tested system scales to two SUs, but larger deployments can be built based on customer requirements

- A fully tested system scales for a single SU, seamlessly.
- Rack-level integrated design which allows rapid installation and deployment of liquid cooled, high density compute racks.

- Storage partner equipment that has been certified to work in a DGX SuperPOD environment.
  - Storage partner equipment that has been certified to work in DGX SuperPOD environments.
  - Full system support—including compute, storage, network, and software—is provided by NVIDIA Enterprise Support (NVEX).

Compute Fabric

- The compute fabric is rail-optimized to the top layer of the fabric.
- The compute fabric is a balanced, full-fat tree.
- The compute fabric is designed with current state-of-the-art, top performing, high-performance, low-latency network switches, and supports future generation of networking hardware.
- Managed NDR switches are used throughout the design to provide better management of the fabric.
- The fabric is designed to support the latest SHARPv3 features.

Storage Fabric (High Speed Storage)

The storage fabric provides high bandwidth to shared storage. It ~~also~~ has the following characteristics:

- It is independent of the compute fabric to maximize performance of both storage and application performance.
- Provides single-node line-rate of 400Gbps to each DGX GB300 compute tray.
- Storage is provided over RDMA over Converged Ethernet to provide maximum performance and minimize CPU overhead.
- User-accessible management nodes provide access to shared storage.

In-Band Management Network

- The in-band management network fabric is Ethernet-based and is used for node provisioning, data movement, Internet access, and other services that must be accessible by the users.
- The in-band management network connections for compute and management servers operate at 200 Gbps and are bonded for resiliency.

Out-of-Band Management Network

The OOB management network connects all the baseboard management controllers (BMC), BlueField baseboard management controllers, NVSwitch Management Interfaces (COMe), as well as other devices that should be physically isolated from system users.

Storage Requirements

The DGX SuperPOD compute architecture must be paired with a high-performance, balanced, storage system to maximize overall system performance. DGX SuperPOD is designed to use two separate storage systems, high-performance storage (HPS) and user storage, optimized for key operations of throughput, parallel I/O, as well as higher IOPS and metadata workloads.

High-Performance Storage

High-Performance Storage is provided via RDMA over Converged Ethernet v2 (RoCEv2) connected storage from a DGX SuperPOD certified storage partner, and is engineered and tested with the following attributes in mind:

- High-performance, resilient, POSIX-style file system optimized for multi-threaded read and write operations across multiple nodes.
- Certified for Grace-based systems.
- Native RoCE support.
- Local system RAM for transparent caching of data.
- Leverage local flash device transparently for read and write caching.

The specific storage fabric topology, capacity, and components are determined by the DGX SuperPOD certified storage partner as part of the DGX SuperPOD design process.

User Storage

User Storage differs from High-Performance storage in that it exposes an NFS share on the in-band management fabric for multiple uses. It is typically used for “home directory” type usage (especially with clusters deployed with SLURM), administrative scratch space, and shared storage as needed by DGX SuperPOD components in a High Availability configuration (e.g., Mission Control), the requirement for log files collection, and system configuration files.

With that in mind, User Storage has the following recommended attributes:

- Designed for high metadata performance, IOPS, and key enterprise features such as log collection, data capacity. This is different than the HPS, which is optimized for parallel I/O and large capacity.
- Communicate over Ethernet, using NFS.
- 100GbE DR1 Connectivity at least.

User storage in a DGX SuperPOD is often satisfied with existing NFS servers already deployed, such that a new export is created and made accessible to the

DGX SuperPOD's in-band management network. User Storage is therefore not described in detail in this DGX SuperPOD reference architecture.

Network Fabrics

Building systems by scalable units (SUs) provides the most efficient designs. However, if a different node count is required due to budgetary constraints, data center constraints, or other needs, the fabric should be designed to support the full SU, including leaf switches and leaf-spine cables, and leave the portion of the fabric unused where these nodes would be located. This will ensure optimal traffic routing and ensure that performance is consistent across all portions of the fabric.

DGX SuperPOD configurations utilize five network segments, these segments are:

- NVLink 5
- Compute Network
- Storage Network
- In-Band Management Network
- Out-of-Band Management Network

These network segments are carried by four different physical fabrics:

- Multi-node NVLink Fabric (MN-NVL)
- Compute InfiniBand Fabric
- Storage and In-band Ethernet Fabric
- Out-of-Band network

Each of these fabrics are discussed in this section.

Multi-node NVLink Fabric (NVL5)

Each DGX GB300 rack is built with 18 compute trays and 9 NVLink switch trays. Each NVLink switch tray is equipped with 2 NVLink switch chips and are responsible for the full-mesh connectivity between all 72 GPUs within the same DGX GB300 rack. Each B300 GPU features 18 NVL5 links and has one dedicated NVL5 link connectivity to each one of the 18 switch chips, delivering a total bandwidth of 1.8 TB/s low latency bandwidth.

Figure 7 visualizes the topology of the Multi-node NVLink topology.

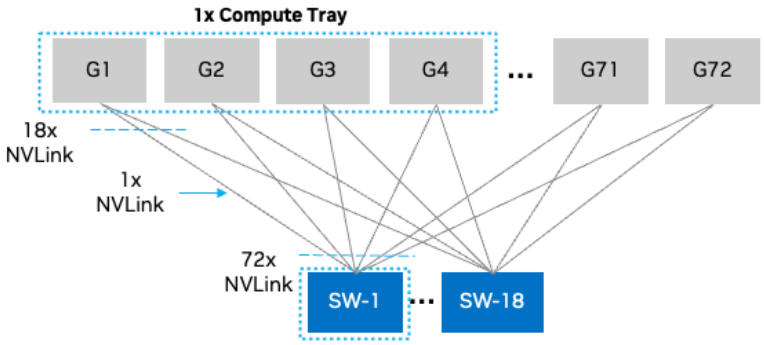

Figure 7: Multi-node NVLink Topology

Compute Fabric

Figure 8 shows the compute fabric layout for the full GB300 SuperPOD Scalable Unit (8 DGX GB300 systems). Each compute rack is rail-aligned. Traffic per rail of each compute tray is always one hop away from other compute trays in the same Scalable Unit. Traffic between different compute racks, or between different rails, traverses the spine layer.

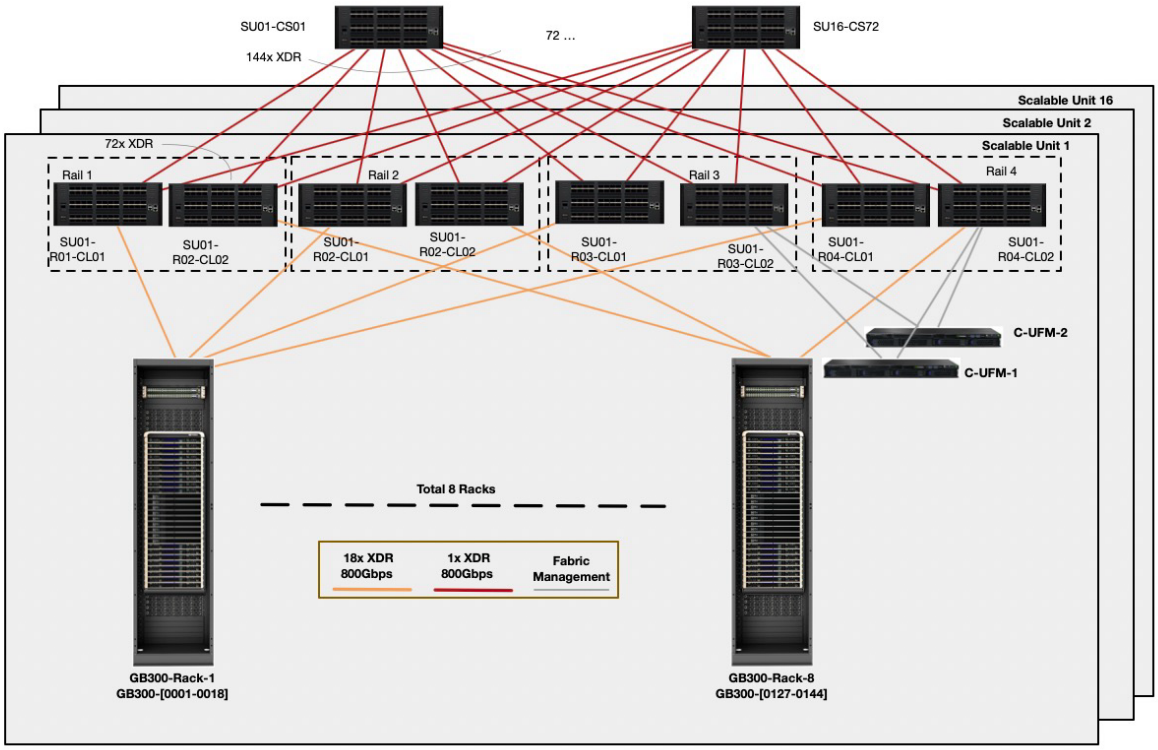

Figure 8. Compute fabric for full 576 GPUs DGX SuperPOD

Thanks to the enhanced port radix if 144 port for Q3400-RD Quantum-X800 switches, we can scale the SuperPOD beyond 1 SU with a two-layer fabric. Each SU contains a total of eight leaf switches distributed in 4 rails, with each rail having 2 leaf switches.
Up to 16 SUs are aggregated in the 2-layer fabric – connecting to 9k GPUs in total.

Figure 8 shows the compute fabric design for DGX SuperPOD with GB300.

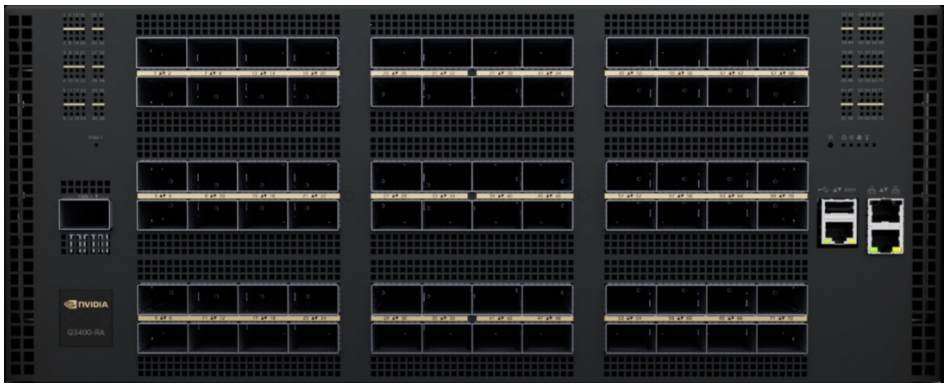

Figure 10. Q3400-RD InfiniBand Quantum-X800 switch

Table 3. Larger SuperPOD component counts

# GPUs	# SUs	IB Leaf Switch		IB Spine Switch
		Per SU	Total	Total
1152	2	8	16	8
2304	4	8	32	18
4608	8	8	64	36
9216	16	8	128	72

Ethernet Fabric

With DGX GB300, we continue to use the previously introduced Ethernet Fabric that was introduced with GB200 SuperPOD for storage and in-band network to enhance cost-efficiency while maintaining the high level of performance required by the storage network in a large-scale AI training cluster.

The storage and in-band fabric use SN5600D switches shown in Figure 10.

Figure 10. SN5600D switch

Figure 12 represents the scale-out design for the Ethernet Fabric. The design follows a 2-layer design with spine-leaf switches. Each SU contains four leaf SN5600 switches. These switches are connected to the BlueField BF3240 dual-port card at a 400Gbps for best performance and redundancy.

The 2-layer fabric with up to 24 spine switches supports DGX SuperPODs with up to 16 SUs (9k GPUs). Each the ethernet features also a storage leaf side – with a configurable number of switches starting at 2 leaves. Finally, the Ethernet fabric is extended by a pair of management leaves to connect the persistent home storage, and the management nodes, as well as a pair of out-of-band-spines to connect the vast amount of SN2201 OOB switches.

For DGX GB300 SuperPOD scale-out, each compute leaf connects via 24 links at 400 Gbps, supporting a 1:3 blocking ratio. The storage leaves connect via 72 non-blocking 400 Gbps links. The management leaves are connected at 24x 400 Gbps as well to ensure best performance delivery for provisioning and home storage.

Figure 12. Storage and In-band Ethernet fabric logical design

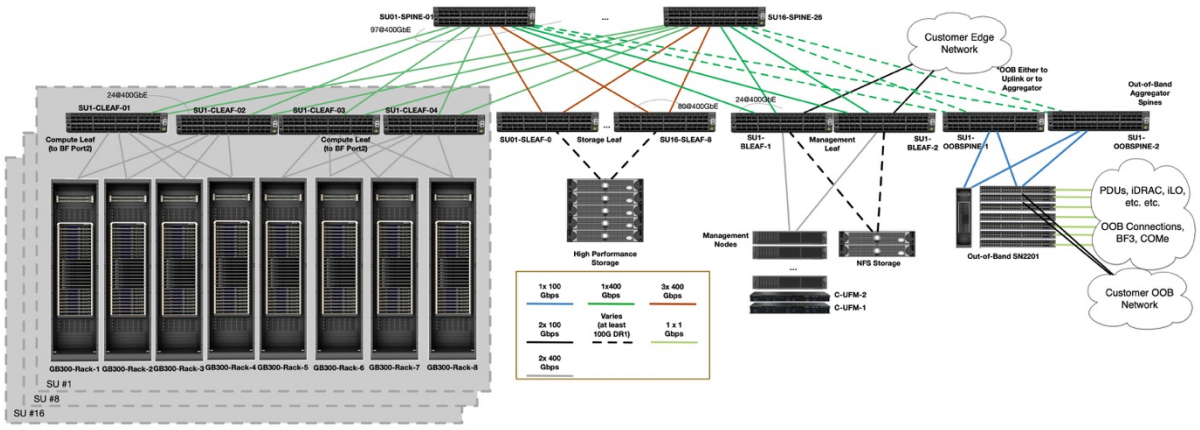

Table 4. Spine and Super Spine Switch Requirements for Scale Out

# GPUs	# SUs	Spines	Compute Leafs	Storage Leafs	Max. Supported Storage Bandwidth	Management Leafs and OOB Spines
1152	2	4	8	2	64x400GbE	2+2
2304	4	8	16	2	128x400GbE	2+2
4608	8	12	32	4	256x400GbE	2+2
9216	16	24	64	8	1024x400GbE	2+2

Network Segmentation of the Ethernet Fabric

The ethernet fabric is segmented into these segments on the SuperPOD:

- Storage Network and In-band Network
- Out-of-Band Management Network

In the following, we introduce these networks in detail.

Storage and In-band Network

As DGX SuperPOD is the optimized for best-in-class performance, the DGX GB300 SuperPOD dedicates a 400Gbps port to the storage network segment and a dedicated link that provides at least 200Gbps to the inband network segment.
This design has the following benefit:

- RoCE traffic is not shared on a bonded interface with In-band traffic, which makes the Storage traffic more predictable and consistent
- Simplicity in setup (dedicated ports) are used for different network segments, avoiding complex, per host routing configuration which are more prone to error.
- Network configuration is compatible with previous, DGX SuperPOD GB200.

For high-performance storage access, the RDMA over Converged Ethernet version 2 (RoCEv2) is used as communication protocol to ensure the performance requirement for storage access.

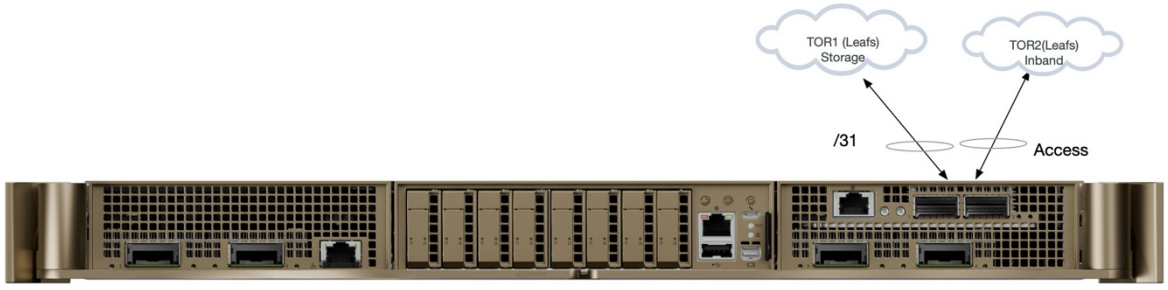

Figure 16. Storage and Inband Access

The in-band management network provides several key functions:

- Connects all the services that manage the cluster.
- Enables access to the lower-speed, NFS tier of storage.

- Provides uplink (border) connectivity for the in-cluster services such as Mission Control, Base Command Manager, Slurm, and Kubernetes to other services outside of the cluster such as the NGC registry, code repositories, and data sources.
- Provides end user access to the Slurm head nodes and Kubernetes services.

This new network design keeps the SuperPOD's performance promise while allowing backwards compatibility in networking design. To resolve the missing high availability feature, NVIDIA aims to enable HBN on the BlueField DPUs to enable innovative storage solutions in near future.

Out-of-Band Management Network

Figure 16 shows the OOB Ethernet fabric. It connects the management ports of all devices including DGX GB300 compute trays, switch trays, and management servers, storage, networking gear, rack PDUs, and all other devices. These are separated onto their own network. There is no use-case where a non-privileged user needs direct access to these ports and are secured using logical network separation.

The OOB network carries all IPMI related control traffic. In DGX SuperPOD, it also serves as the network for fabric and switch management of the compute InfiniBand fabric and compute NVLink fabric.

The OOB network is physically rolled up into the aggregation layer (spine layer) of each SU as a dedicated VXLAN. The OOB management network use SN2201 switches, shown in Figure 17.

With DGX SuperPOD GB300 , we have added the aggregation spines for all out-of-band switch, which allows one to compose a physically independent fabric for all the IPMI management traffic.

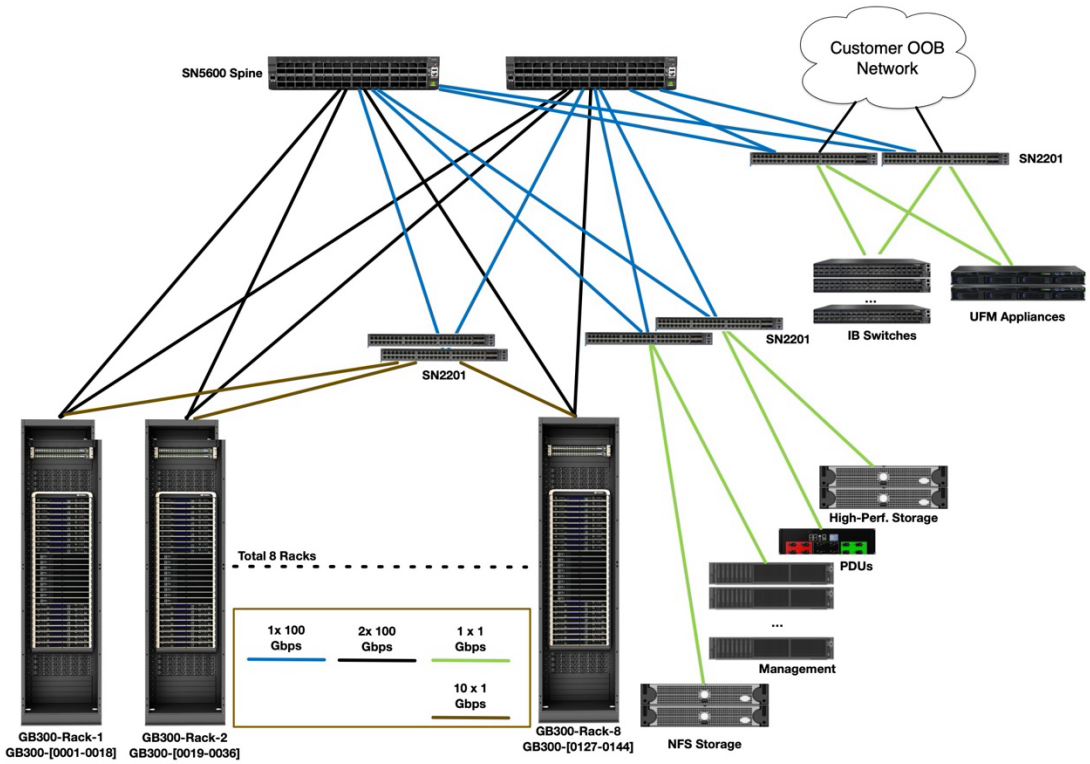

Figure 16. Logical OOB management network layout

Figure 17. SN2201 switch

Customer Edge Connectivity

For connecting DGX SuperPOD to customer edge for uplink and customer corporate network access, we recommend at least 2x 100GbE links with DR1 single-mode connectivity to cope with the growing demand on high-speed data transfer into and from DGX SuperPOD.

A dedicated VTEP for uplink, the default hand over to customer edge is based on BGP peering and providing routes from in-band to customers' edge. The building management system (BMS), which is responsible to provide datacenter-scale integration for power and cooling telemetry, is required to flow through the uplink interfaces as part of the untrusted zone.

For route handover, we prepare eBGP protocol to peer with customer's network. Routes to/from in-band, out-of-band, and building management system are announced.

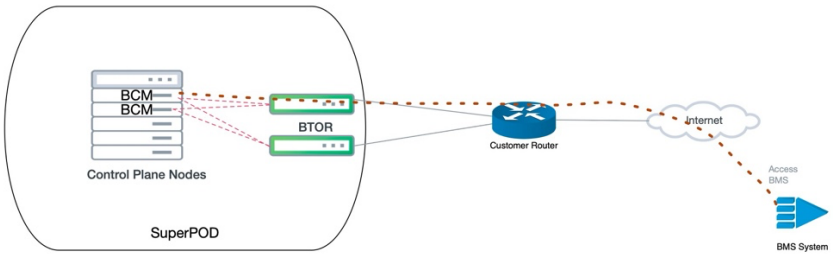

Figure 14: Uplink Connectivity

With DGX SuperPOD GB300 – the default configuration also assumes that customer would provide dedicated access to their OOB network. The last pair of the OOB switch becomes the "OOB edge TOR" switches, which would also be responsible of OOB uplink.

In case customers choose not to have dedicated OOB link, the spine SN5600D switches can be still connected to the In-band spine network, and routes to the OOB interfaces can be leaked towards customer edge.

Storage Architecture

The compute nodes must be paired with HPS utilizing NVMe technology, which provides a balanced I/O to maximize overall system performance with the following requirements:

- The storage system is based on file and optimized for both read and write operations across multiple nodes.
- File storage system must support RDMA to accelerate data and I/O access.
- Seamless linear scalability in both capacity and performance is essential to accommodate growing workloads without disruption.

Reliability, availability, and serviceability (RAS) are critical pillars of HPS. The architecture must be highly resilient, capable of maintaining uninterrupted data access even during component failures. Non-disruptive upgrades to storage software and firmware are essential to minimize downtime and maintain operational continuity.

File-based storage with RDMA is ideal for various workloads from training through inference, and while providing fast checkpointing with high-performance access to the data. This type of storage is particularly effective in environments where low latency and throughput are essential.

Developing accurate deep learning (DL) models relies on processing large and ever-growing volumes of data. The iterative nature of DL training means data must be accessed and reused rapidly, making both read and write storage performance critical. Efficient checkpointing—saving model states during training—ensures recoverability but can introduce bottlenecks unless handled by high-bandwidth, low-latency storage systems. Asynchronous checkpointing has become more common, allowing training to continue while writing checkpoints in the background and reducing the impact on throughput.

Model fine-tuning builds on pre-trained models, adapting them to specific enterprise needs with less resource intensity than full training. However, it still requires fast access to both model checkpoints and training data to support rapid experimentation and frequent updates. High-performance, scalable storage

is essential to maintain throughput and enable efficient iterative cycles during fine-tuning, supporting agile development environments.

Inference pipelines, particularly for large language and multimodal models, demand extremely fast and consistent access to model weights, embeddings, and feature data to ensure responsiveness in real-time applications. As these workloads grow in scale and complexity, storage systems must provide high bandwidth and predictable low-latency performance. The ability to scale storage performance with compute resources is key to maintaining efficient inference as demand increases.

Advanced AI workflows also depend on optimizations like key-value (KV) caching, which reduces latency and computational overhead by storing intermediate model computations. Query and retrieval operations, such as those in Retrieval-Augmented Generation (RAG) pipelines, require low-latency, high-throughput access to vector indexes for timely, context-aware responses. Storage infrastructure supporting these advanced use cases must offer not only robust performance but also features like secure access controls, efficient updates, and compliance with governance policies to protect sensitive enterprise data.

The preceding metrics assume a variety of workloads, datasets, and need for training and inference. It is best to characterize workloads and organizational needs before finalizing performance and capacity requirements. Table 4 and Table 5 are provided to help determine the I/O levels required for different types of models.

Table 4. Storage performance requirements

Level	Work Description	Data Set Size
Standard	Multiple concurrent LLM or fine-tuning training jobs and periodic checkpoints with inference, where the compute requirements dominate the data I/O requirements significantly.	Most datasets can fit within the local compute systems' memory cache during training or inference. The datasets are single modality, and models have millions of parameters.
Enhanced	Multiple concurrent multimodal training jobs and periodic checkpoints with inference, where the data I/O performance is an important factor for end-to-end workload timing.	Datasets are too large to fit into local compute systems' memory cache requiring more I/O during training or inference, not enough to obviate the need for frequent I/O. The datasets have multiple

		modalities and models have billions (or higher) of parameters.

Table 5. Guidelines for storage performance

Performance Characteristic	Standard (GBps)	Enhanced (GBps)
Single SU aggregate system read	90	280
Single SU aggregate system write	45	140
4 SU aggregate system read	360	1,120
4 SU aggregate system write	180	560

DGX SuperPOD Software

NVIDIA Mission Control software delivers a full-stack data center solution engineered for enterprise infrastructure deployments, like NVIDIA DGX SuperPOD. It integrates essential management and operational capabilities into a unified platform, providing enterprise customers with seamless control over their infrastructure deployments at scale.

Figure 17 shows the detailed decomposition of DGX GB300 SuperPOD software stack.

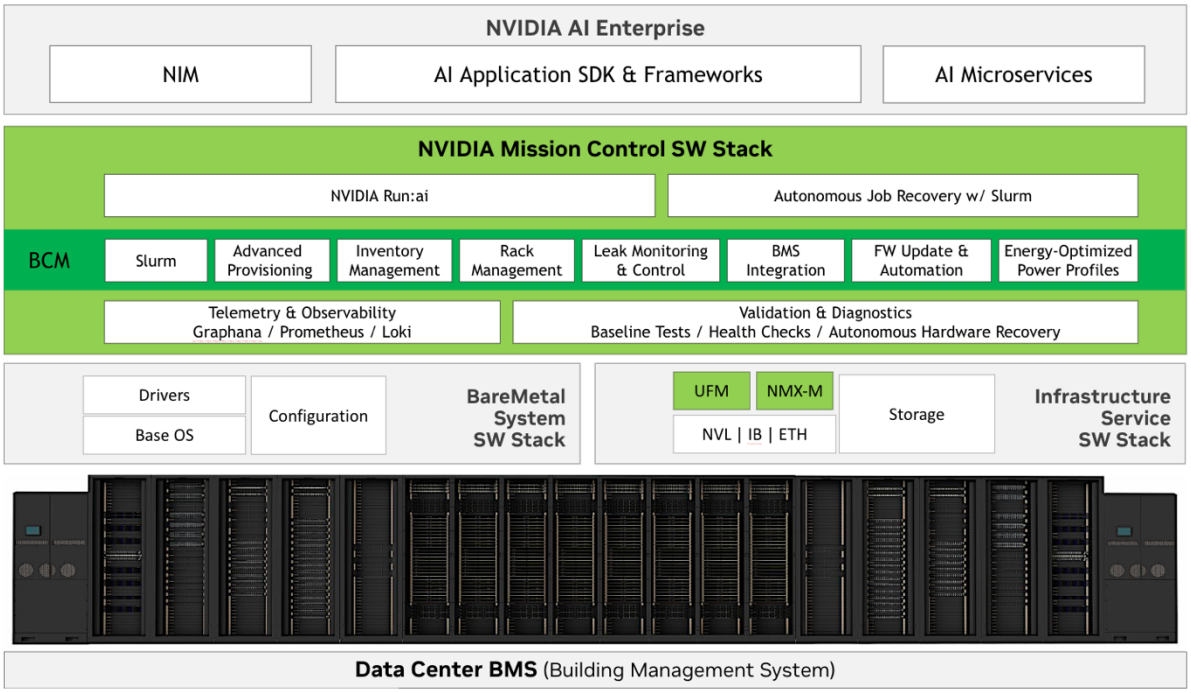

Figure 17. NVIDIA DGX GB300 SuperPOD Software Stack

NVIDIA Mission Control software stack comprises a five-layer architecture that builds from foundational system health and diagnostics to advanced NVIDIA DGX SuperPOD cluster management operations. It leverages NVIDIA Base Command Manager (BCM) technology and NVIDIA Run:ai functionality to provide scheduler access with SLURM and Kubernetes for AI and HPC workload orchestration. The Telemetry and Observability layer leverages proprietary diagnostic tools and health-checks for system insights, while the Validation and Diagnostics layer,

supports the built-in autonomous recovery engine that ensures rapid failure recovery and system restoration with minimal time to recovery.
NVIDIA Mission Control also handles deployment optimization and system health monitoring, seamlessly cooperating NVIDIA Base Command Manager (BCM) technology for coordinating cluster provisioning & operations. Included in the software stack there is also Network Manager eXperience Manager for monitoring & control of NVLINK Switch trays.

NVIDIA Mission Control delivers critical innovations that directly impact DL and HPC environments, improving efficiency, reducing downtime, and optimizing resource utilization. Here's how these advancements translate into tangible business value:

- Automated Failure Detection & Self-Recovery:
  NVIDIA Mission Control's automated failure detection and recovery mechanisms significantly minimize downtime compared to manual interventions. With faster recovery times, AI training continues without significant disruption, resulting in improved GPU utilization. This reduces the risk of resource wastage.

- Optimized Workload Migration & Resource Allocation:
  NVIDIA Mission Control ensures that workloads are continuously reassigned to healthy nodes, preventing idle GPU time. This leads to increased overall GPU efficiency, enabling the system to process more workloads within the same timeframe. By optimizing resource allocation, NVIDIA Mission Control helps organizations achieve higher throughput for AI training tasks, thereby accelerating time-to-results.

- Unified Diagnostics for Infrastructure & Applications:
  By combining infrastructure and application-level diagnostics, NVIDIA Mission Control helps significantly reduce troubleshooting time. Model builders can identify and resolve issues much faster compared to traditional methods. This considerably decreases the operational burden on SREs and DevOps, leading to savings in personnel costs and reducing the time to production.

- In-Memory Checkpointing for Seamless Job Recovery:
  NVIDIA Mission Control enables recovery of training workloads by automatically restarting failed processes/ranks from the most recent valid checkpoint, ensuring no data loss or rework. In environments with frequent failures, it significantly reduces retraining time, thereby eliminating downtime and preventing interruptions in the training loop.

This results in an increase in model iteration speed and faster go-to-market timelines for AI products.

NVIDIA Run:ai

Included with NVIDIA Mission Control, the Run:ai platform employs a distributed architecture consisting of two primary components: the control plane (backend) and the cluster. The control plane serves as the central management system, capable of orchestrating multiple Run:ai clusters across an organization. For BCM-managed installations, the control plane of Run:AI resides on the “user-space Kubernetes cluster”, while cluster components are deployed directly onto the administrative Kubernetes infrastructure. This separation of responsibilities allows for centralized management while maintaining workload execution close to computational resources.

Figure 17 showcases the architecture of NVIDIA Run:ai

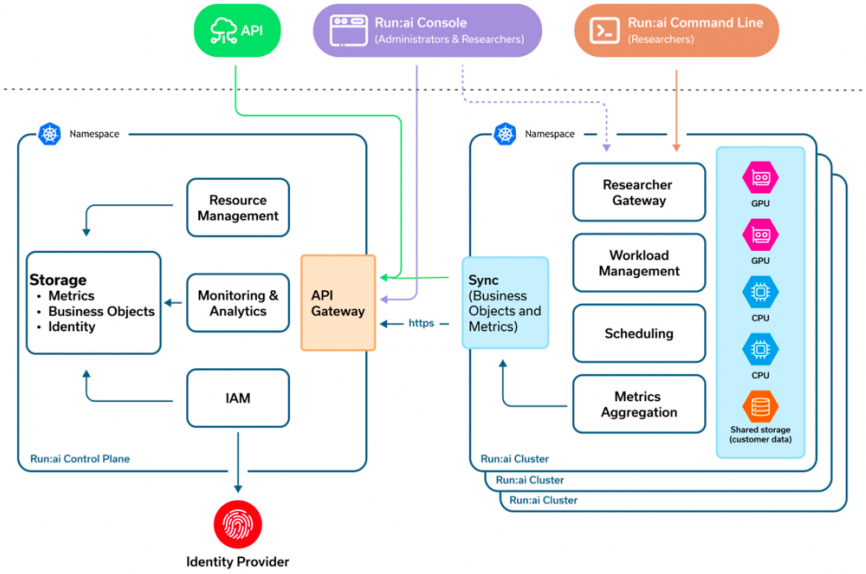

Figure 17. NVIDIA Run:ai

Summary

NVIDIA DGX SuperPOD with NVIDIA DGX GB300 systems is the next generation of data center scale architecture to meet the demanding and growing needs of AI training and inferencing. This reference architecture document for DGX SuperPOD represents the architecture used by NVIDIA for our own AI model and HPC research and development. DGX SuperPOD continues to build upon its high-performance roots to enable training of the largest NLP models, support the expansive needs of training models for automotive applications, and scaling-up recommender models for greater accuracy and faster turn-around-time.
DGX SuperPOD represents a complete system of not just hardware but all the necessary software to accelerate time-to-deployment, streamline system management, proactively identify system issues. The combination of all these components keeps systems running reliably, with maximum performance, and enables users to push the bounds of state-of-the-art. The platform is designed to both support the workloads of today and grow to support tomorrow's applications. These are representatives of the configuration and must be finalized based on actual design.
Your NVIDIA representative can work with you to finalize the exact components list.

Further information on the detailed requirement for DGX SuperPOD, such as the datacenter design guide, detailed software architecture for NVIDIA mission control, and relevant security information, are also available to customers through NVIDIA representative.

NVIDIA Reference Architecture Notice and Disclaimers

NVIDIA Reference Architecture documentation is NVIDIA confidential information and embodies substantial intellectual property and engineering know-how. NVIDIA grants you limited permissions to use NVIDIA Reference Architecture to create NVIDIA platform-enabled data center clusters with NVIDIA technologies. You may not use the NVIDIA Reference Architecture to develop competing products or technologies or assisting a third party in such activities. Except as expressly stated in this Notice & Disclaimer statement, no other license or right is granted to you by implication, estoppel or otherwise. NVIDIA objects to and rejects any additional or different terms you may propose.

NVIDIA Reference Architecture is subject to change without notice. NVIDIA technologies may require enabled hardware, software, or service activation. Your costs and results may vary. This document is not a commitment to develop, release, or deliver any technology, code, or functionality. NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice. It is your sole responsibility to evaluate and determine the applicability of any information contained in this document and ensure that the product is suitable and fit for the application planned by you. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party or a license from NVIDIA.

NVIDIA products or services are subject to the applicable NVIDIA product or sales terms and conditions that are provided in conjunction with the relevant NVIDIA product or service. NVIDIA's provision of these resources does not expand or otherwise alter NVIDIA's applicable warranties or warranty disclaimers for the NVIDIA products or services.

You will not use the NVIDIA Reference Architecture or NVIDIA's confidential information to: (i) identify or support an assertion or potential assertion of any intellectual property rights (including patent, copyright, or trade secret); or (ii) prepare, file, amend, or prosecute any patent applications. You agree to grant NVIDIA a limited, worldwide, nonexclusive, perpetual, irrevocable, non-sublicensable, nontransferable, royalty-free, fully-paid license to any patent claim that you may file after accessing NVIDIA Reference Architecture that covers the subject matter disclosed herein.

Warranty Disclaimer. NVIDIA Reference Architecture materials disclosed are provided "AS IS". To the maximum extent permitted by applicable law, NVIDIA disclaims all warranties and representations of any kind, whether express, implied, or statutory, relating to or arising under this agreement, including, without limitation, the warranties of title, noninfringement, merchantability, fitness for a particular purpose, usage of trade, and course of dealing.

Limitation of Liability. To the maximum extent permitted by applicable law, in no event will NVIDIA be liable for any (i) indirect, punitive, special, incidental, or consequential damages, or (ii) damages for the (a) cost of procuring substitute goods or (b) loss of profits, revenues, use, data, or goodwill arising out of or related to this document, whether based on breach of contract, tort (including negligence), strict liability, or otherwise, and even if NVIDIA has been advised of the possibility of such damages and even if a your remedies fail their essential purpose. Additionally, to the maximum extent permitted by applicable law, NVIDIA's total cumulative aggregate liability for any and all liabilities, obligations, or claims arising out of or related to this document will not exceed one-hundred U.S. dollars (US$100).