DDN A^{3}I® SOLUTIONS WITH NVIDIA DGX™ A100 SYSTEMS

Fully-integrated and optimized infrastructure solutions for accelerated at-scale AI, Analytics and HPC

EXECUTIVE SUMMARY

DDN A^{3}I Solutions are proven at-scale to deliver optimal data performance for Artificial Intelligence (AI), Data Analytics and High-Performance Computing (HPC) applications running on GPUs in a DGX A100 system. This document describes fully validated reference architectures for scalable NVIDIA DGX POD™ configurations. The solutions integrate DDN AI400X appliances with NVIDIA DGX A100 systems and NVIDIA® Mellanox® InfiniBand network switches.

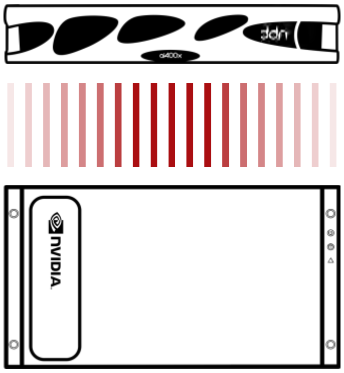
1. DDN A^{3}I End-To-End Enablement for NVIDIA DGX POD

DDN A^{3}I solutions (Accelerated, Any-Scale AI) are architected to achieve the most from at-scale AI, Data Analytics and HPC applications running on DGX systems and DGX POD. They provide predictable performance, capacity, and capability through a tight integration between DDN and NVIDIA systems. Every layer of hardware and software engaged in delivering and storing data is optimized for fast, responsive, and reliable access.

DDN A^{3}I solutions are designed, developed, and optimized in close collaboration with NVIDIA. The deep integration of DDN AI appliances with DGX systems ensures a reliable experience. DDN A^{3}I solutions are highly configuration for flexible deployment in a wide range of environments and scale seamlessly in capacity and capability to match evolving workload needs. DDN A^{3}I solutions are deployed globally and at all scale, from a single DGX system all the way to the largest NVIDIA DGX SuperPOD™ in operation today.

DDN brings the same advanced technologies used to power the world's largest supercomputers in a fully-integrated package for DGX systems that's easy to deploy and manage. DDN A^{3}I solutions are proven to maximum benefits for at-scale AI, Analytics and HPC workloads on DGX systems.

This section describes the advanced features of DDN A^{3}I Solutions for DGX POD.



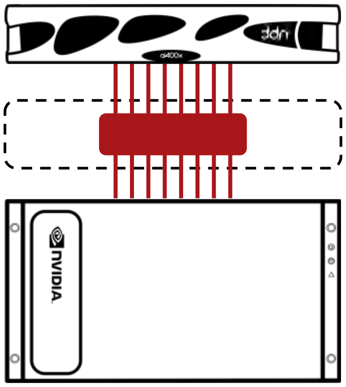
DDN A^{3}I Shared Parallel Architecture

The DDN A^{3}I shared parallel architecture and client protocol ensures high levels of performance, scalability, security, and reliability for DGX systems. Multiple parallel data paths extend from the drives all the way to containerized applications running on the GPUs in the DGX system. With DDN's true end-to-end parallelism, data is delivered with high-throughput, low-latency, and massive concurrency in transactions. This ensures applications achieve the most from DGX systems with all GPU cycles put to productive use. Optimized parallel data-delivery directly translates to increased application performance and faster completion times. The DDN A^{3}I shared parallel architecture also contains redundancy and automatic failover capability to ensure high reliability, resiliency, and data availability in case a network connection or server becomes unavailable.

DDN A^{3}I Streamlined Deep Learning

DDN A^{3}I solutions enable and accelerate end-to-end data pipelines for deep learning (DL) workflows of all scale running on DGX systems. The DDN shared parallel architecture enables concurrent and continuous execution of all phases of DL workflows across multiple DGX systems. This eliminates the management overhead and risks of moving data between storage locations. At the application level, data is accessed through a standard highly interoperable file interface, for a familiar and intuitive user experience.

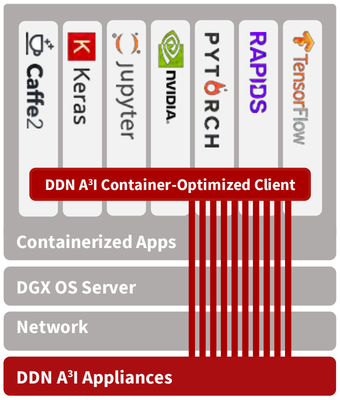
Significant acceleration can be achieved by executing an application across multiple DGX systems in a DGX POD simultaneously and engaging parallel training efforts of candidate neural networks variants. These advanced optimizations maximize the potential of DL frameworks. DDN works closely with NVIDIA and its customers to develop solutions and technologies that allow widely-used DL frameworks to run reliably on DGX systems.



DDN A^{3}I Multirail Networking

DDN A^{3}I solutions integrate a wide range of networking technologies and topologies to ensure streamlined deployment and optimal performance for AI infrastructure. Latest generation InfiniBand (IB) and Ethernet provide both high-bandwidth and low-latency data transfers between applications, compute servers and storage appliances. For DGX POD, DDN recommends an IB Network with NVIDIA network switches. DDN A^{3}I Multirail greatly simplifies and optimizes DGX system networking for fast, secure, and resilient connectivity.

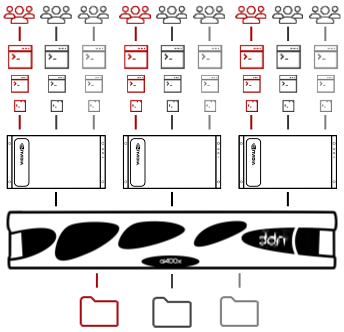
DDN A^{3}I Multirail enables grouping of multiple network interfaces on a DGX system to achieve faster aggregate data transfer capabilities. The feature balances traffic dynamically across all the interfaces, and actively monitors link health for rapid failure detection and automatic recovery. DDN A^{3}I Multirail makes designing, deploying, and managing high-performance networks very simple, and is proven to deliver complete connectivity for at-scale infrastructure for DGX POD deployments.



DDN A^{3}I Container Client

Containers encapsulate applications and their dependencies to provide simple, reliable, and consistent execution. DDN enables a direct high-performance connection between the application containers on the DGX A100 system and the DDN parallel filesystem. This brings significant application performance benefits by enabling low latency, high-throughput parallel data access directly from a container. Additionally, the limitations of sharing a single host-level connection to storage between multiple containers disappear. The DDN in-container filesystem mounting capability is added at runtime through a universal wrapper that does not require any modification to the application or container.

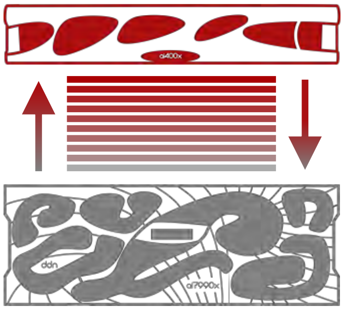
Containerized versions of popular DL frameworks specially optimized for the DGX systems are available from NVIDIA. They provide a solid foundation that enables data scientists to rapidly develop and deploy applications on the DGX system. In some cases, open-source versions of the containers are available, further enabling access and integration for developers. The DDN A^{3}I container client provides high-performance parallelized data access directly from containerized applications on the DGX system. This provides containerized DL frameworks with the most efficient dataset access possible, eliminating all latencies introduced by other layers of the computing stack.



DDN A^{3}I Multitenancy

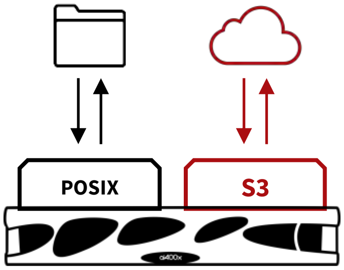
Container clients provide a simple and very solid mechanism to enforce data segregation by restricting data access within a container. DDN A^{3}I makes it very simple to operate a secure multitenant environment at-scale through its native container client and comprehensive digital security framework. DDN A^{3}I multitenancy makes it simple to share DGX systems across a large pool of users and still maintain secure data segregation. Multi-tenancy provides quick, seamless, dynamic DGX system resource provisioning for users. It eliminates resource silos, complex software release management, and unnecessary data movement between data storage locations. DDN A^{3}I brings a very powerful multitenancy capability to DGX systems and makes it very simple for customers to deliver a secure, shared innovation space, for at-scale data-intensive applications.

Containers bring security challenges and are vulnerable to unauthorized privilege escalation and data access. The DDN A^{3}I digital security framework provides extensive controls, including a global *root_squash* to prevent unauthorized data access or modification from a malicious user, and even if a node or container are compromised.



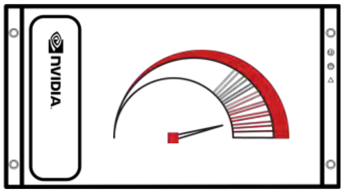
DDN A³I Hot Pools

Hot Pools delivers user transparent automatic migration of files between the Flash tier (Hot Pool) to HDD tier (Cool Pool). Hot Pools is designed for large scale operations, managing data movements natively and in parallel, entirely transparently to users. Based on mature and well tested file level replication technology, Hot Pools allows organizations to optimize their economics – scaling HDD capacity and/or Flash performance tiers independently as they grow.



DDN A³I S3 Data Services

DDN S3 Data Services provide hybrid file and object data access to the shared namespace. The multi-protocol access to the unified namespace provides tremendous workflow flexibility and simple end-to-end integration. Data can be captured directly to storage through the S3 interface and accessed immediately by containerized applications on a DGX system through a file interface. The shared namespace can also be presented through an S3 interface, for easy collaboration with multisite and multicloud deployments. The DDN S3 Data Services architecture delivers robust performance, scalability, security, and reliability features.



DDN A³I Advanced Optimizations for DGX A100 System Architecture

The DDN A³I client's NUMA-aware capabilities enable strong optimization for DGX systems. It automatically pins threads to ensure I/O activity across the DGX system is optimally localized, reducing latencies and increasing the utilization efficiency of the whole environment. Further enhancements reduce overhead when reclaiming memory pages from page cache to accelerate buffered operations to storage. The DDN A³I client software for DGX A100 systems has been validated at-scale with the largest DGX SuperPOD deployment currently in operation.

2. DDN A^{3}I Solutions with NVIDIA DGX A100 Systems

The DDN A^{3}I scalable architecture integrates DGX A100 systems with DDN AI shared parallel file storage appliances and delivers fully-optimized end-to-end AI, Analytics and HPC workflow acceleration on GPUs. DDN A^{3}I solutions greatly simplify the deployment of DGX A100 systems in DGX POD configurations, while also delivering performance and efficiency for maximum GPU saturation, and high levels of scalability.

This section describes the components integrated in DDN A^{3}I Solutions for DGX POD.

2.1 DDN AI400X Appliance

The AI400X appliance is a fully integrated and optimized shared data platform with predictable capacity, capability, and performance. Every AI400X appliance delivers over 50 GB/s and 3M IOPS directly to DGX A100 systems in the DGX POD. Shared performance scales linearly as additional AI400X appliances are integrated to the DGX POD. The all-NVMe configuration provides optimal performance for a wide variety of workload and data types and ensures that DGX POD operators can achieve the most from at-scale GPU applications, while maintaining a single, shared, centralized data platform.

The AI400X appliance integrates the DDN A^{3}I shared parallel architecture and includes a wide range of capabilities described in section 1, including automated data management, digital security, and data protection, as well as extensive monitoring. The AI400X appliances enables DGX POD operators to go beyond basic infrastructure and implement complete data governance pipelines at-scale.

The AI400X appliance integrates with DGX POD over IB, Ethernet and RoCE. It is available in 32, 64, 128 and 256 TB all-NVMe capacity configurations. Optional hybrid configurations with integrated HDDs are also available for deployments requiring high-density deep capacity storage. Contact DDN Sales for more information.

Figure 1. DDN AI400X all-NVME storage appliance.

2.2 NVIDIA DGX A100 System

The NVIDIA DGX A100 system is the universal system for all AI workloads, offering unprecedented compute density, performance, and flexibility in the world's first 5 petaFLOPS AI system. Built on the revolutionary NVIDIA A100 Tensor Core GPU, the DGX A100 system unifies data center AI infrastructure, running training, inference, and analytics workloads simultaneously with ease. More than a server, the DGX A100 system is the foundational building block of AI infrastructure and part of the NVIDIA end-to-end data center solution created from over a decade of AI leadership by NVIDIA. The DGX A100 system integrates exclusive access to a global team of AI-fluent experts that offer prescriptive planning, deployment, and optimization expertise to help fast-track AI transformation.

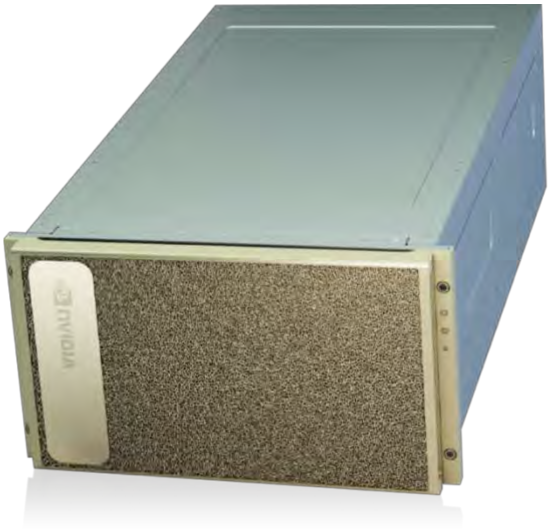


Figure 2. NVIDIA DGX A100 system.

2.3 NVIDIA Mellanox Switches

NVIDIA Mellanox network switches provide optimal interconnect for DGX POD. DDN recommends IB technology for data-intensive compute and storage networks in the DGX POD. The NVIDIA Mellanox QM8700 HDR 200Gb/s IB Switch provides 40 ports of connectivity in a 1U form factor. Every switch is capable of up to 16Tb/s of non-blocking bandwidth with sub 130ns port-to-port latency. The QM8700 is the ideal modular network unit to architect scalable solutions with DGX POD, with no compromise.

Figure 3. NVIDIA Mellanox QM8700 HDR 200Gb/s InfiniBand Switch

For customers that require ethernet or RoCE, the NVIDIA Mellanox Spectrum SN3700 Open Ethernet switch provides 32 ports of 200 GbE in a compact 1U form factor. It enables connectivity to endpoints at different speeds and carries a throughput of 12.8Tb/s, with full port speed flexibility from 10 GbE to 200 GbE per port. This makes it simple to integrate DGX POD within any existing Ethernet environment and achieve optimal performance for data-intensive compute and storage networks.

Figure 4. NVIDIA Mellanox Spectrum SN3700 Open Ethernet Switch.

The NVIDIA Mellanox AS4610 Data Center Open Ethernet Switch is a Gigabit Ethernet Layer 3 switch family featuring 54 ports with 48 10/100/1000BASE-T ports, 4 x 10G SFP+ uplink ports. It provides robust capabilities for critical low-intensity traffic like DGX POD component management.

Figure 5. NVIDIA Mellanox AS4610 Data Centre Open Ethernet Switch.

2.4 NVIDIA AI Software

The value of the DGX POD architecture extends well beyond its hardware. The DGX POD is a complete system providing all the major components for system management, job management, and optimizing workloads to ensure quick deployment, ease of use, and high availability (HA). The software stack begins with the DGX Operating System (DGX OS), which is tuned and qualified for use on DGX A100 systems. The DGX POD contains a set of tools to manage the deployment, operation, and monitoring of the cluster. NVIDIA NGC™ is a key component of the DGX POD, providing the latest DL frameworks. NGC provides packaged, tested, and optimized containers for quick deployment, ease of use, and the best performance on NVIDIA GPUs. Lastly, key tools like CUDA-X, Magnum IO, and RAPIDS provide developers the tools they need to maximize DL, HPC, and data science performance in multi-node environments.

DGX OS and POD Management

NVIDIA AI software (Figure 6) running on the DGX POD provides a high-performance DL training environment for large scale multi-user AI software development teams. In addition to the DGX OS, it contains cluster management, orchestration tools and workload schedulers (DGX POD management software), NVIDIA libraries and frameworks, and optimized containers from the NGC container registry. For additional functionality, the DGX POD management software includes third-party opensource tools recommended by NVIDIA which have been tested to work on DGX POD racks with the NVIDIA AI software stack. Support for these tools is available directly from third-party support structures.

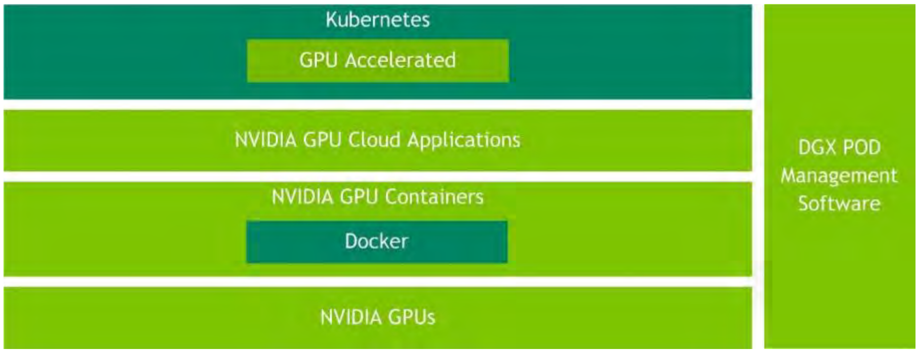


Figure 6. NVIDIA AI Software Stack

The foundation of the NVIDIA AI software stack is the DGX OS, built on an optimized version of the Ubuntu. For customers who prefer RedHat Linux, NVIDIA software packages needed to tune the operating system specifically for the DGX hardware. Whether DGX OS or RedHat, the DGX software includes certified GPU drivers, a network software stack, pre-configured local caching, NVIDIA data center GPU management (DCGM) diagnostic tools, and GPU-enabled container runtime, all certified to work with NVIDIA NGC containers.

The DGX POD management software (Figure 7) is composed of various services running on the Kubernetes container orchestration framework for fault tolerance and HA. Services are provided for network configuration (DHCP) and fully automated DGX OS software provisioning over the network (PXE). The DGX OS software can be automatically re-installed on demand by the DGX POD management software.

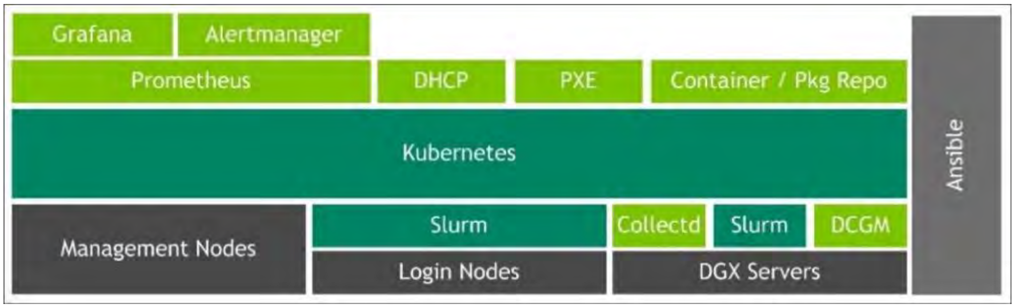


Figure 7. NVIDIA DGX POD Management Software

The DGX POD management software DeepOps provides the tools to provision, deploy, manage, and monitor the DGX POD. Services are hosted in Kubernetes containers for fault tolerance and HA. The DGX POD management software leverages the Ansible configuration tool to install and configure all the tools and packages needed to run the system. System data collected by Prometheus is reported through Grafana. Alertmanager can use the collected data and send automated alerts as needed.

For sites required to operate in an air-gapped environment or needing additional on-premises services, a local container registry mirroring NGC container, as well as OS and Python package mirrors, can be run on the Kubernetes management layer to provide services to the cluster.

The DGX POD management software can deploy Slurm or Kubernetes as the orchestration and workload manager. Slurm is often the best choice for scheduling training jobs in a shared multi-user, multi-node environment where advanced scheduling features such as job priorities, backfill, and accounting are required. Kubernetes is often the best choice in environments where GPU processes run as a service, such as inference, large use of interactive workloads through Jupyter notebooks, and where there is value to having the same environment as is often used at the edge of an organization's data center.

NVIDIA NGC

NVIDIA NGC (Figure 8) provides a range of options that meet the needs of data scientists, developers, and researchers with various levels of AI expertise. These users can quickly deploy AI frameworks with containers, get a head start with pre-trained models or model training scripts, and use domain specific workflows and Helm charts for the fastest AI implementations, giving them faster time-to-solution.

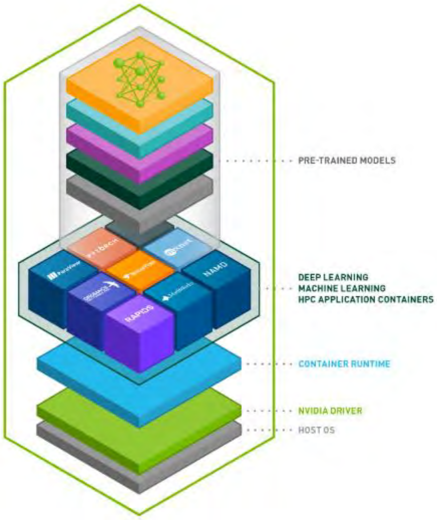


Figure 8. NVIDIA NGC Software Stack

Spanning AI, data science, and HPC, the container registry on NGC features an extensive range of GPU-accelerated software for NVIDIA GPUs. The NGC hosts containers for the top AI and data science software. Containers are tuned, tested, and optimized by NVIDIA. Other containers for additional HPC applications and data analytics are fully tested and made available by NVIDIA as well. NGC containers provide powerful and easy-to-deploy software proven to deliver the fastest results, allowing users to build solutions from a tested framework, with complete control.

NGC offers step-by-step instructions and scripts for creating DL models, with sample performance and accuracy metrics to compare your results. These scripts provide expert guidance on building DL models for image classification, language translation, text-to-speech and more. Data scientists can quickly build performance-optimized models by easily adjusting hyperparameters. In addition, NGC offers pre-trained models for a variety of common AI tasks that are optimized for NVIDIA Tensor Core GPUs and can be easily re-trained by updating just a few layers, saving valuable time.

3. DDN A^{3}I Reference Architectures for DGX POD

DDN proposes the following reference architectures for multi-node DGX POD configurations. DDN A^{3}I solutions are fully-validated with NVIDIA and already deployed with several DGX POD customers worldwide.

The DDN AI400X appliance is a turnkey appliance for at-scale DGX deployments. DDN recommends the AI400X appliance as the optimal data platform for DGX POD designs with the DGX A100 system. The AI400X appliances delivers optimal GPU performance for every workload and data type in a dense, power efficient 2RU chassis. The AI400X appliance simplifies the design, deployment, and management of a DGX POD and provides predictable performance, capacity, and scaling. The AI400X appliance arrives fully configured, ready to deploy and installs rapidly. The appliance is designed for seamless integration with DGX systems and enables customers to move rapidly from test to production. As well, DDN provides complete expert design, deployment, and support services globally. The DDN field engineering organization has already deployed dozens of solutions for customers based on the A^{3}I reference architectures.

As general guidance, DDN recommends an AI400X appliance for every four DGX A100 systems in a DGX POD (Figure 9). These configurations can be adjusted and scaled easily to match specific workload requirements. For the storage network, DDN recommends HDR200 technology in a non-blocking network topology, with redundancy to ensure data availability. DDN recommends use of at least two HDR200 connections per DGX A100 system to the storage network.

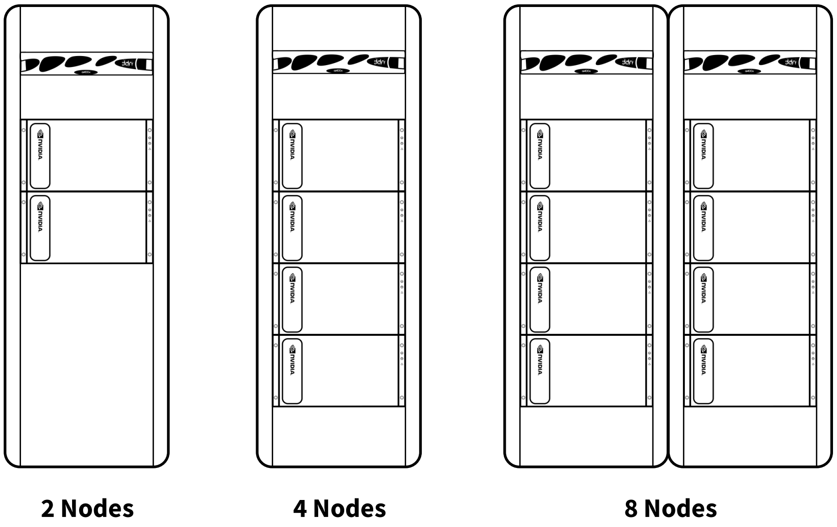


Figure 9. Rack illustrations for DDN A^{3}I POD reference configurations *(network switches not shown).*

3.1 DGX POD Network Architecture

The DGX POD reference design includes four networks:

Storage network. Provides data transfer between the AI400X appliance and the compute nodes. Connects eight ports from each AI400X appliance. Connects two ports from each DGX A100 system, one each from two dual-port NVIDIA Mellanox ConnectX®-6 HCAs. These may be configured as IB or Ethernet mode. DDN recommends for optimal performance and efficiency.

Compute network. Provides inter-node communication. Connects the eight single-port ConnectX-6 HCAs from each DGX A100 system. These may be configured in IB InfiniBand or Ethernet mode.

Management Network. Provides management and monitoring for all DGX POD components. Connects the 1 GbE RJ45 Management port and 1 GbE RJ45 BMC port from each DGX A100 system and each AI400X appliance controller to an Ethernet switch.

(Optional) Cluster Network. Provides provisioning and job scheduling. Uses 100 GbE QSFP56 port on each DGX A100 system and optional external servers.

An overview of the DGX POD network architecture is shown in Figure 10, recommended network connections for each DGX A100 system on Figure 11, and recommended network connections for each DDN AI400X appliance on Figure 12.

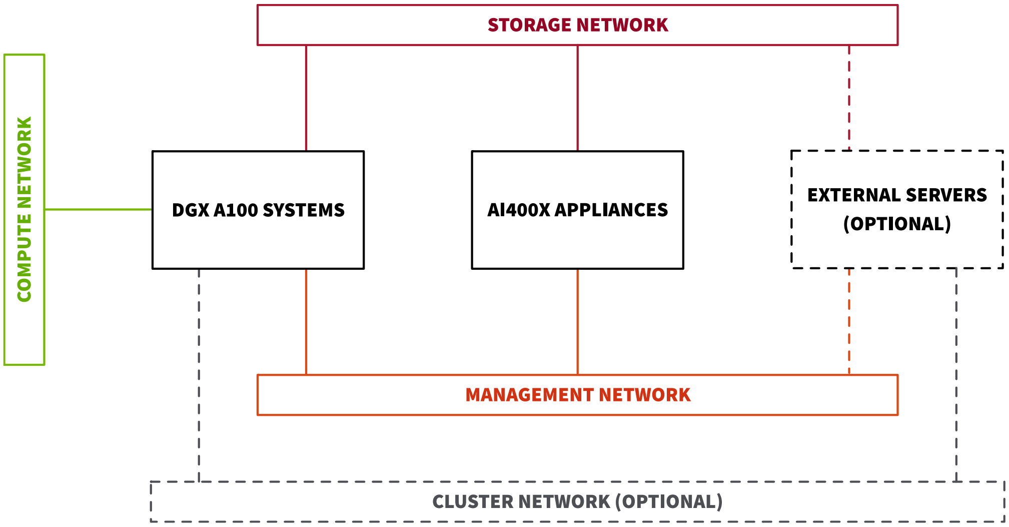


Figure 10. Overview of the DGX POD network architecture.

DGX A100 Systems Network Connectivity

For DGX POD, DDN recommends ports 1 to 8 on the DGX A100 systems be connected to the compute network. DDN recommends that all DGX A100 systems in the DGX POD be equipped with the optional HCA card. ports 9 and 11 be connected to the storage network. For DGX A100 systems without the optional HCA card installed, DDN recommends ports 9 and 10 be connected to the storage network. As well, the management ("M") and BMC ("B") ports should be connected to the management network.

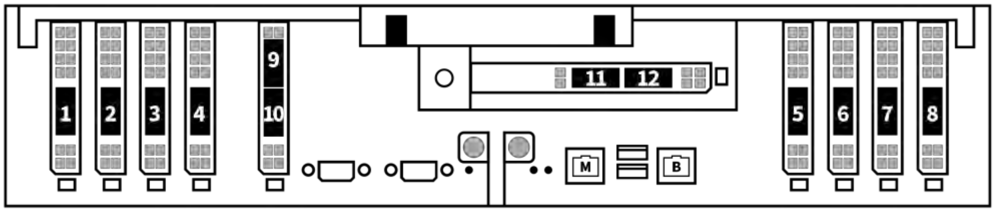


Figure 11. Recommended DGX A100 system network port connections.

AI400X Appliance Network Connectivity

For DGX POD, DDN recommends ports 1 to 8 on the AI400X appliance be connected to the storage network. As well, the management ("M") and BMC ("B") ports for both controllers should be connected to the management network. Note that each AI400X appliance requires two inter-controller network ports connections ("I1" and "I2") using short ethernet cables supplied.

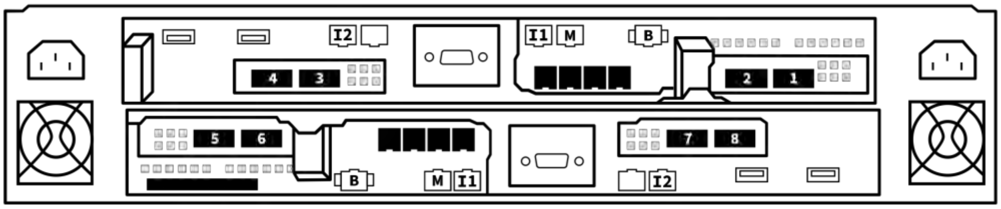
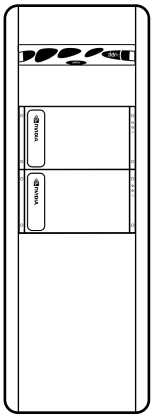


Figure 12. Recommended AI400X appliance network port connections.



3.2 DGX POD with Two DGX A100 systems

Figure 13 illustrates the DDN A^{3}I architecture in a 1:2 configuration in which two DGX A100 system are connected to an AI400X appliance through a pair of network switches that are configured for HA. Every DGX A100 system connects to each of the storage network switches via one HDR or 100 GbE links. The AI400X appliance connects to each of the storage network switches via four HDR 200Gb/s IB or 100 GbE links. The storage network switches are interconnected with four dedicated links. This ensures non-blocking data communication between every device connected to the network. The HA design provides full-redundancy and maximum data availability in case of component failure in one of the devices.

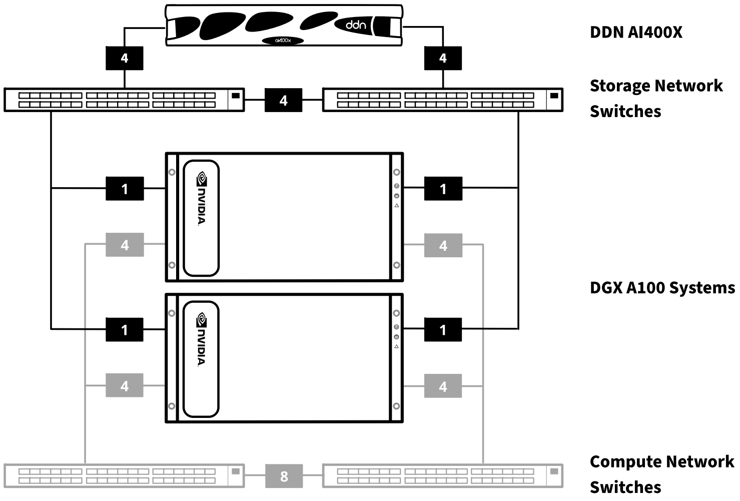


Figure 13. DDN A^{3}I POD reference architecture with two DGX A100 systems (management network not shown).

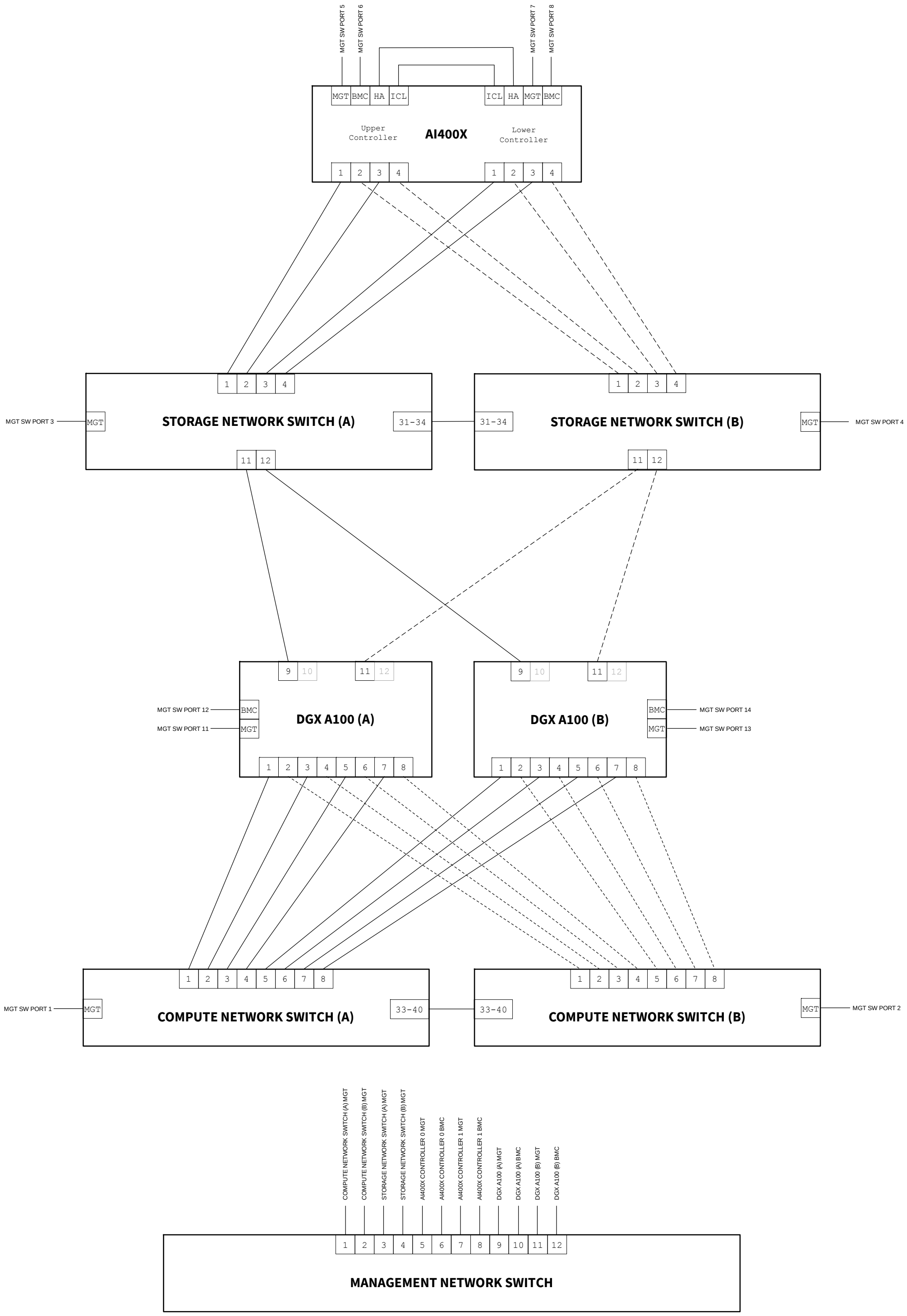
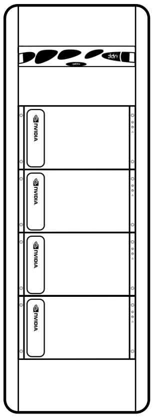


Figure 14. DDN A^{3}I POD network diagram with two DGX A100 systems.



3.3 DGX POD with Four DGX A100 systems

Figure 15 illustrates the DDN A^{3}I architecture in a 1:4 configuration in which four DGX A100 systems are connected an AI400X appliance through a pair of network switches that are configured for HA. Every DGX A100 system connects to each of the storage network switches via one HDR 200Gb/s IB or 100 GbE links. The AI400X appliance connects to each of the storage network switches via four HDR 200Gb/s IB or 100 GbE links. The storage network switches are interconnected with four dedicated links. This ensures non-blocking data communication between every device connected to the network. The HA design provides full-redundancy and maximum data availability in case of component failure in one of the devices.

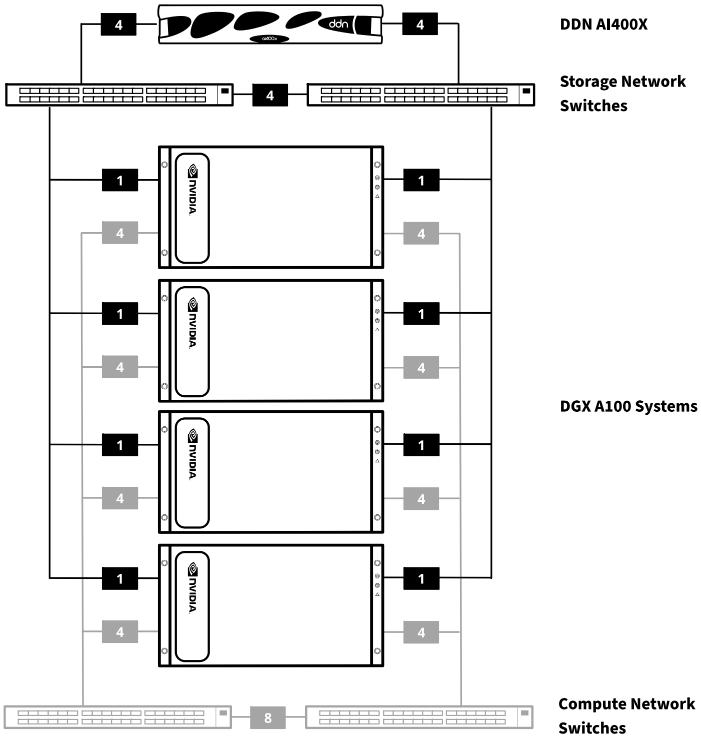


Figure 15. DDN A^{3}I POD reference architecture with four DGX A100 systems (*management network not shown*).

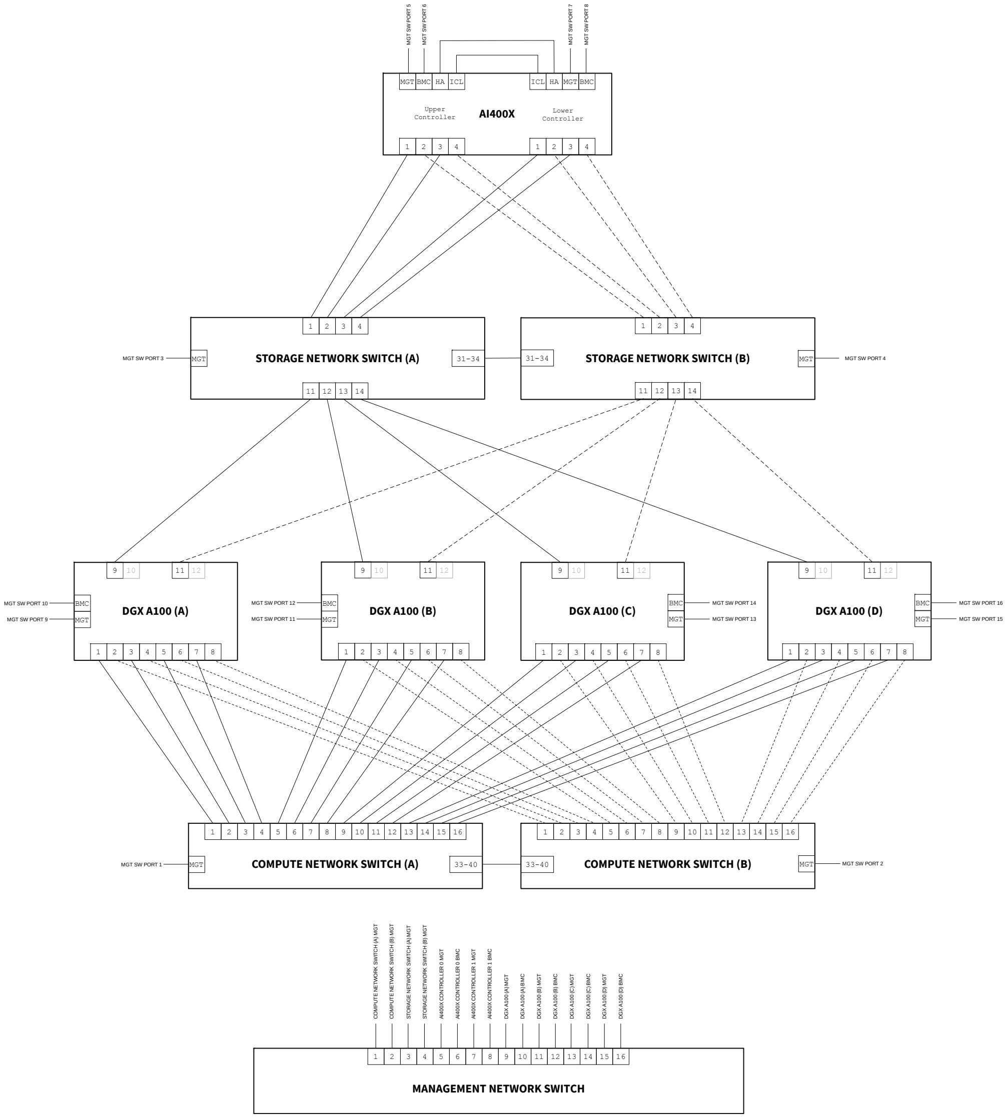
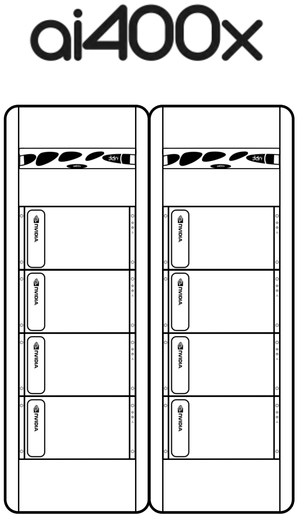


Figure 16. DDN A^{3}I POD network diagram with four DGX A100 systems.



3.4 DGX POD with Eight DGX A100 Systems

Figure 17 illustrates the DDN A^{3}I architecture in a 2:8 configuration in which eight DGX A100 systems are connected with two AI400X appliances through a pair of network switches that are configured for HA. Every DGX A100 system connects to each of the storage network switches via one HDR 200Gb/s IB or 100 GbE links. The AI400X appliance connects to each of the storage network switches via four HDR 200Gb/s IB or 100 GbE links. The storage network switches are interconnected eight dedicated links. This ensures non-blocking data communication between every device connected to the network. The HA design provides full-redundancy and maximum data availability in case of component failure in one of the devices.

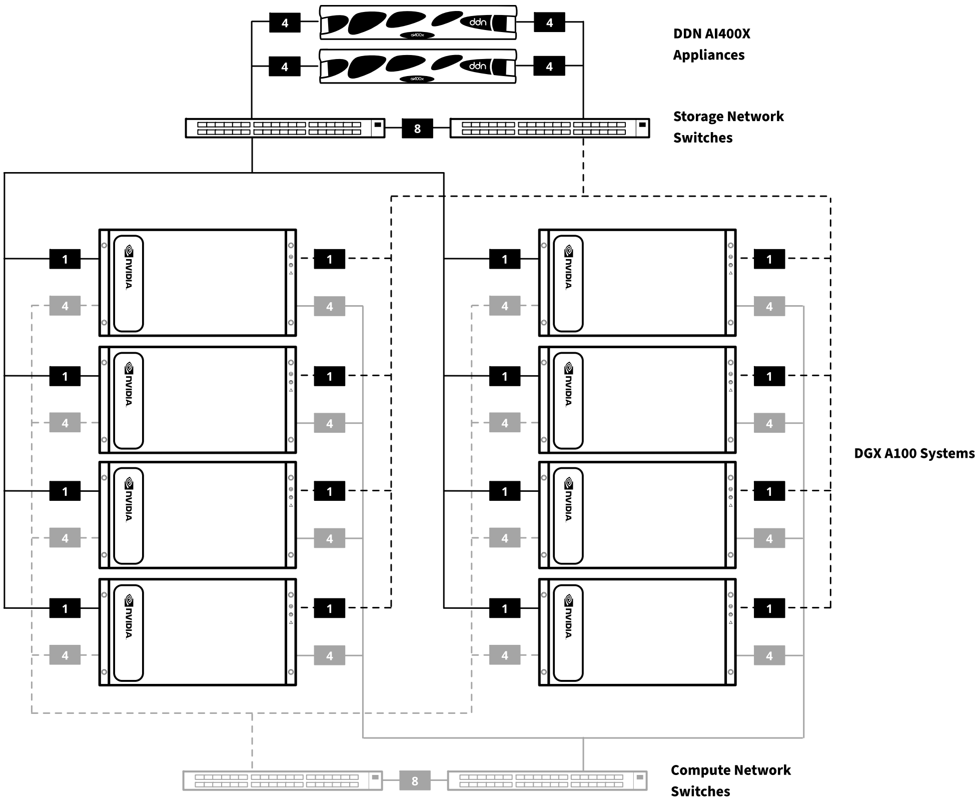


Figure 17. DDN A^{3}I POD reference architecture with eight DGX A100 systems (management network not shown).

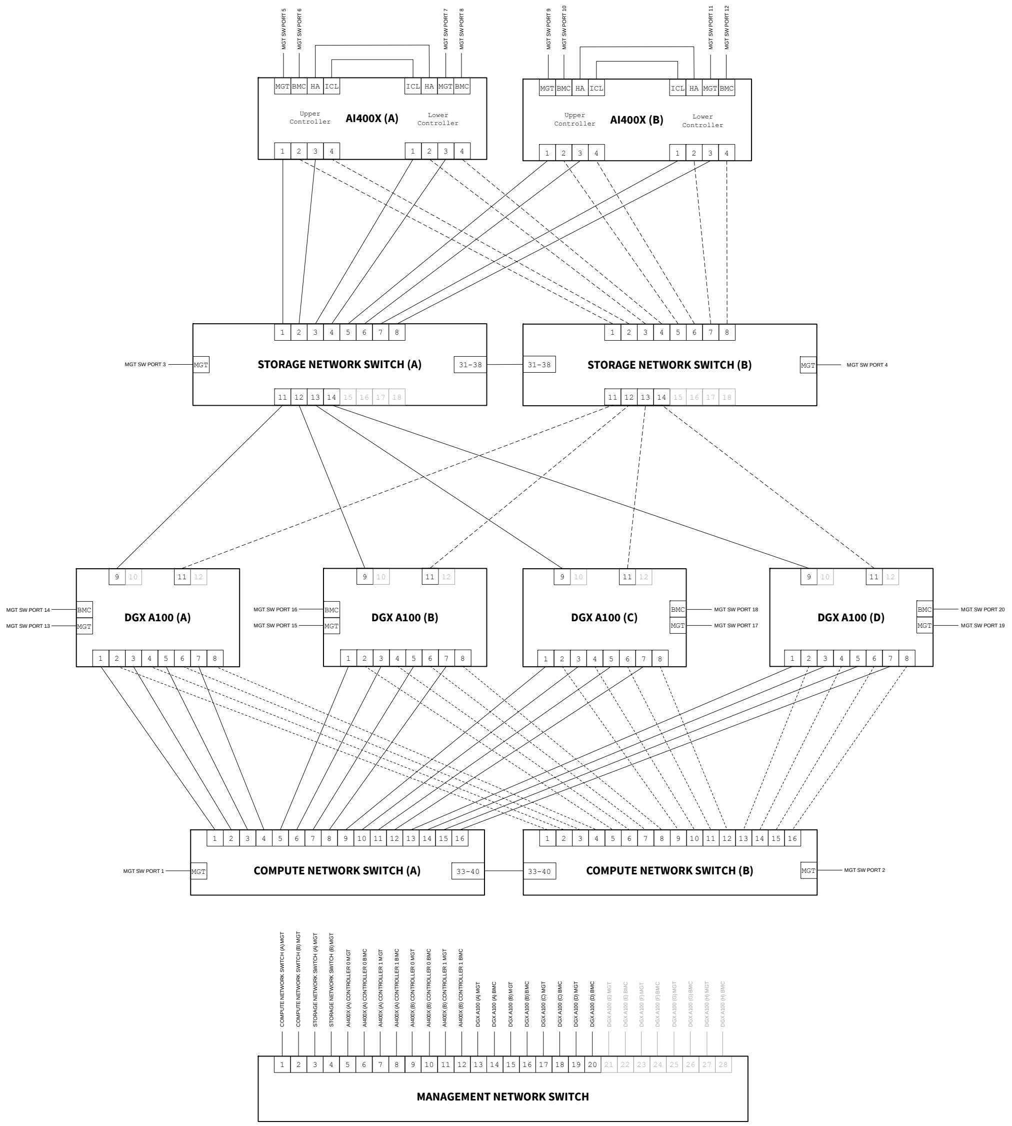


Figure 18. DDN A^{3}I POD network diagram with eight DGX A100 systems (A-D).

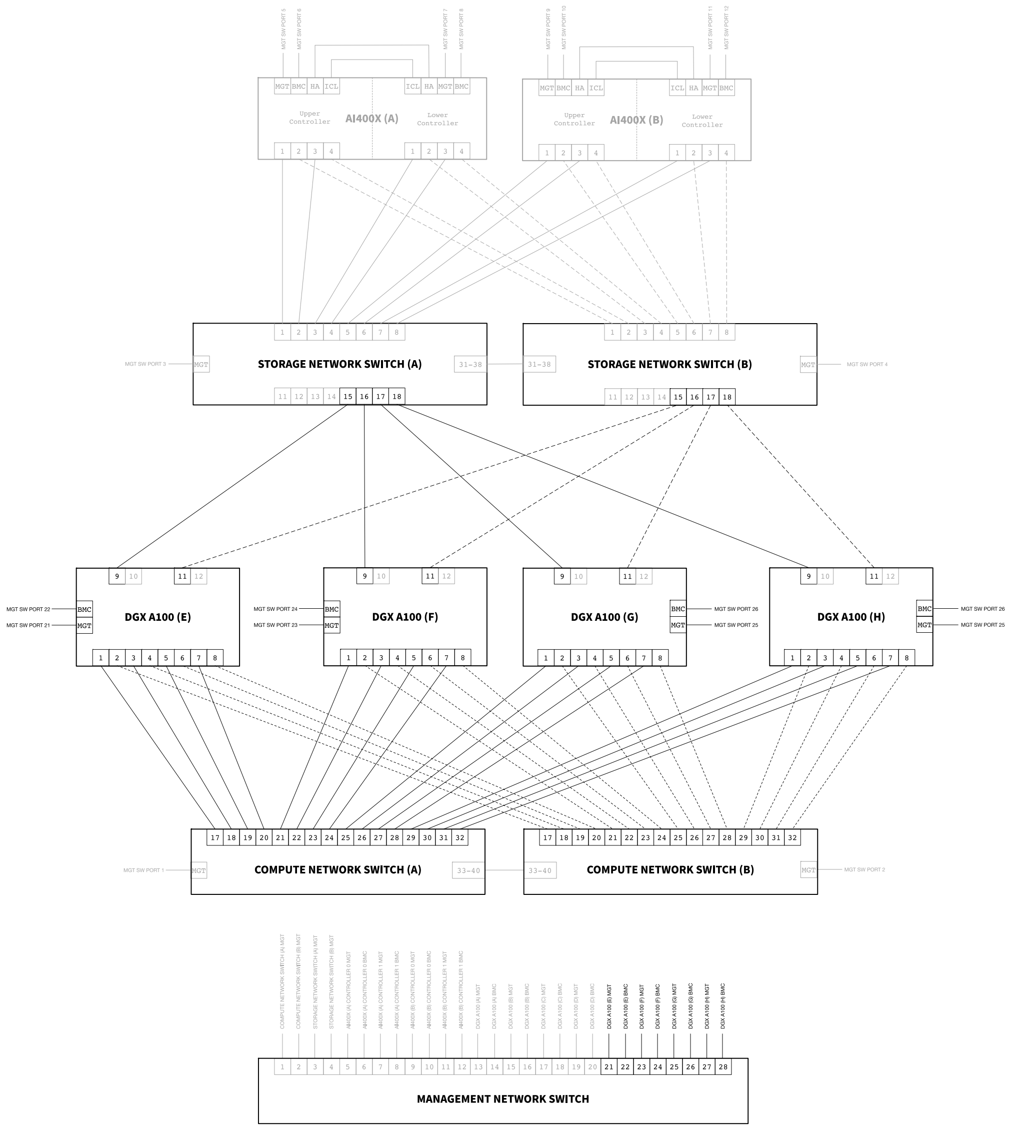


Figure 19. DDN A^{3}I POD network diagram with eight DGX A100 systems (E-G).

3.5 Optional Management Servers

The DGX POD reference design includes three optional servers for advanced cluster management like resource provisioning and job scheduling. The servers integrate seamlessly with the DDN A^{3}I Reference Architectures for DGX POD with two, four and eight DGX A100 systems configurations. The servers connect to the same storage and management network as the DGX A100 systems and AI400X appliances. This enables direct access to the shared filesystem from the servers, and also provides connectivity with DGX A100 systems for cluster operations. For DGX POD configurations that require very high-throughput cluster management, a dedicated 100 GbE network can be deployed to connect the DGX A100 systems with the servers. NVIDIA currently validates designs using the 100 GbE dedicated network. Contact DDN Sales to discuss optional server configurations and ensure optimal end-to-end integration is achieved for both infrastructure and applications.

4. DDN A^{3}I Solutions with NVIDIA DGX POD Validation

DDN conducts extensive engineering integration, optimization, and validation efforts in close collaboration with NVIDIA to ensure best possible end-user experience using the reference designs in this document. The joint validation confirms functional integration, and optimal performance out-of-the-box for DGX POD configurations.

Performance testing on the DDN A^{3}I architecture is been conducted with industry standard synthetic throughput and IOPS applications, as well as widely used DL frameworks and data types. The results demonstrate that with the DDN A^{3}I shared parallel architecture, containerized applications can engage the full capabilities of the data infrastructure and the DGX A100 systems. Performance is distributed evenly across all the DGX A100 systems in the DGX POD, and scales linearly as more DGX A100 systems are engaged.

This section details some of the results from recent at-scale testing integrating AI400X appliances with up to eight DGX A100 systems.

The tests described in this section were executed in an NVIDIA data center on eight DGX A100 system equipped with eight A100 GPUs running DGX OS Server Software 4.99.11 and two AI400X appliance are running DDN EXAScaler v5.2.2.

For the storage network, all eight DGX A100 systems are connected to a Mellanox QM8700 switch with two HDR 200Gb/s IB links, one per dual-ported adapter (see recommendation in section 3.1). The AI400X are connected to the same network with eight HDR100 IB links each. The switch is running Mellanox OS 3.9.0606. For the compute network, all eight DGX A100 systems are connected to a non-blocking HDR 200Gb/s IB network. All eight-single ported network adapters on the DGX A100 systems are connected to the compute network.

This test environment allows us to demonstrate performance with the largest possible DGX POD configuration.

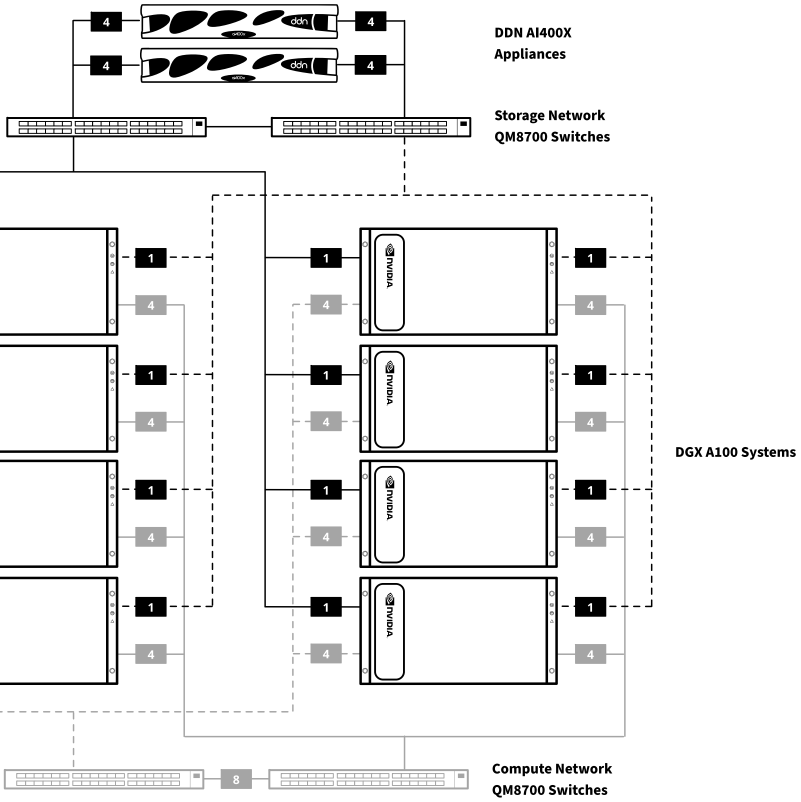

DDN AI400X Appliances
Storage Network QM8700 Switches
DGX A100 Systems
Compute Network QM8700 Switches
4
1
8

Figure 20. Test environment with eight DGX A100 system and two AI400X (management network not shown).

4.1 DGX POD FIO Performance Validation

This series of tests demonstrate the peak performance of the DDN POD reference architecture using the fio open-source synthetic benchmark tool. The tool is set to simulate a general-purpose workload without any performance-enhancing optimizations. Separate tests were run to measure both 100% read and 100% write workload scenarios.

Below are the specific FIO configuration parameters used for these tests:

- blocksize = 1024k
- ioengine = posixaio
- direct = 1
- bw-threads = 255
- iodepth= 128

The AI400X appliance provides predictable, scalable performance. This test demonstrates the architecture's ability to deliver full throughput performance to a small number of clients and distribute the full performance of the DDN solution evenly as a large number of DGX A100 systems are engaged.

In Figure 21, test results demonstrate that DDN solution can deliver over 47 GB/s of read throughput to a single DGX A100 system, and evenly distribute the full read and write performance of the two AI400X appliances with up to eight DGX A100 systems engaged simultaneously. The DDN solution can fully saturate both network links on every DGX A100 system, ensuring optimal performance for a very wide range of data access patterns and data types for applications running on DGX A100 systems in a DGX POD.

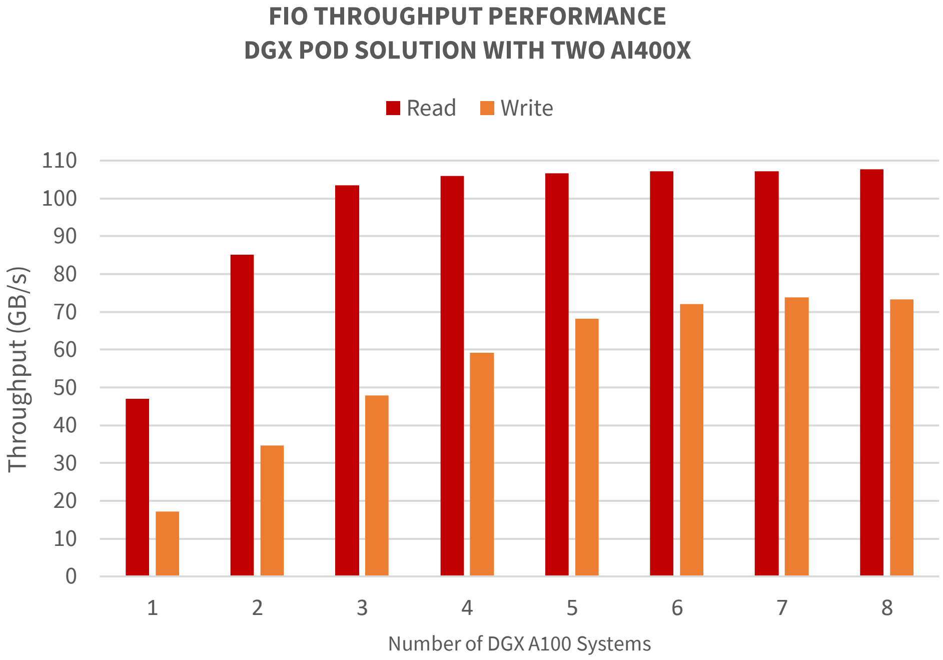


Figure 21. FIO throughput with scaled DGX POD configurations.

4.2 DGX POD MLPERF Performance Validation

Application benchmarks provide important validation and measurements that characterize DGX POD performance beyond infrastructure performance. The MLPerf benchmarks include a variety of different DL models with different I/O requirements, representative of common DL workloads. As demonstrated in the following tests, the AI400X appliance delivers scalable infrastructure and application performance for a wide variety of DL workloads running on DGX POD.

ResNet-50 is the current standard for DL benchmarks. It is an I/O intense workload, processing over 20,000 images per second on a DGX A100 system. The average size of an image in the ImageNet database is 125 KB. This translates into needing to read data at approximately 3 GB/s. For the largest DGX POD with eight DGX A100 systems configurations, the DL workflows requires up to 24 GB/s of read throughput from a shared data repository. A single AI400X appliance delivers nearly twice the required performance and provides ample headroom to meet rapidly evolving application needs.

Figure 22 illustrates application performance as multiple nodes are engaged in the DGX POD. With AI400X appliance, the application on DGX POD achieves linear performance scaling as more nodes are engaged.

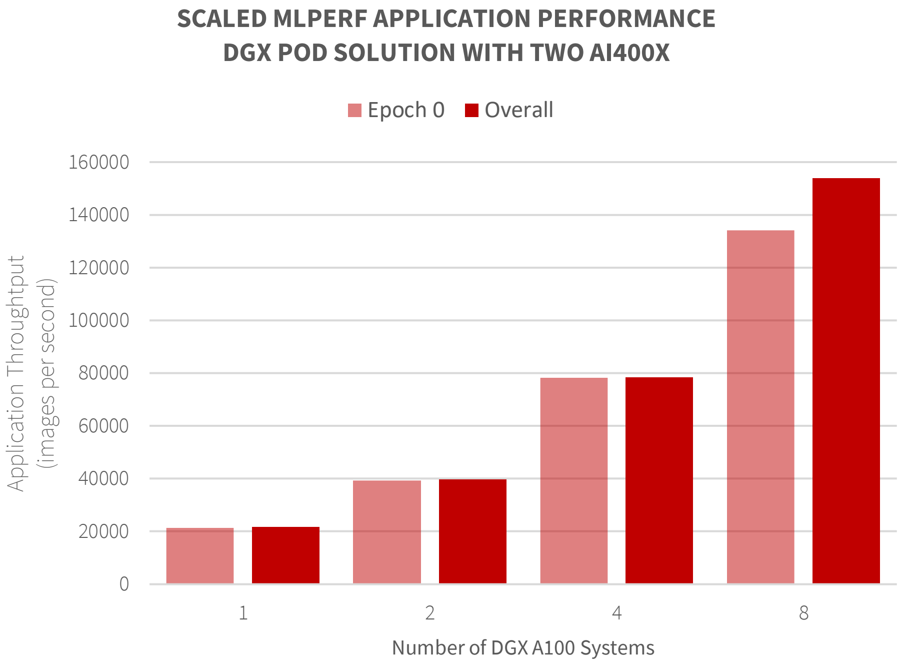


Figure 22. ML Perf application throughput with scaled DGX POD configurations.

4.3 DGX POD NCCL Performance Validation

The NVIDIA Collective Communications Library (NCCL) implements multi-GPU and multi-node collective communication primitives that are performance optimized for NVIDIA GPUs and Networking. This NCCL test is intended to verify scalability across multiple DGX A100 systems to ensure there are no inherent bottlenecks for GPU-to-GPU communication within subsequent benchmarks.

The results of the NCCL test demonstrate that a single node achieves the full internal NVIDIA NVLink™ capabilities of a DGX A100 system, slightly over 235 GB/s. With multiple nodes engaged, the test demonstrate that the compute network is capable of ensuring full network capabilities with up to eight DGX A100 systems engaged simultaneously. This test demonstrates that the compute network in the DDN POD Reference Architecture is fully non-blocking and capable of delivering data to all 64 A100 GPUs with no-compromise to performance and allows limitless flexibility for applications that can take advantage of multiple GPUs within a single DGX A100 system, or across all nodes in the DGX POD.

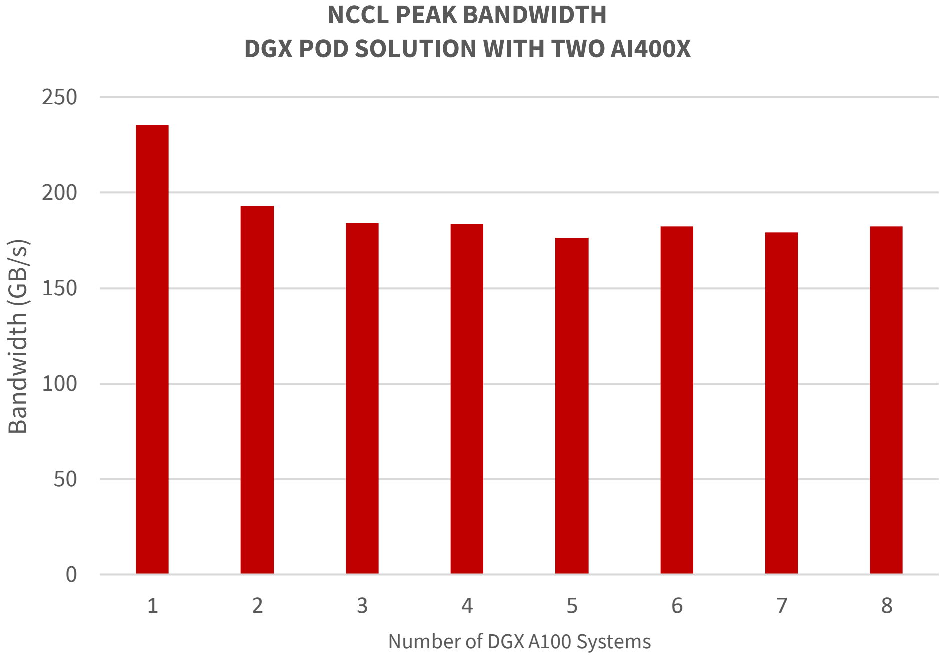


Figure 23. NCCL Bandwidth with scaled DGX POD configurations.

5. Scaling DDN A^{3}I Reference Architectures for DGX A100 Systems

The DDN A^{3}I Reference Architectures for DGX POD are designed to deliver an optimal balance of technical and economic benefits for a wide range of common use cases for AI, Data Analytics and HPC. Using the AI400X appliance as a building block, solutions can scale linearly, predictably and reliably in performance, capacity and capability. For DGX POD applications with requirements beyond the base reference architecture, it's simple to scale the data platform with additional AI400X appliances.

The same AI400X appliance and shared parallel architecture used in the DDN A^{3}I Reference Architectures for DGX POD are also deployed with very large systems. The AI400X appliance has been validated to operate properly with up to 560 DGX A100 systems simultaneously.

In figure 24, we show an fio throughput test performed by NVIDIA engineers similar to the one presented in section 4.1. In this example, up to 128 DGX A100 systems are engaged simultaneously with 10 AI400X appliances. The results of the test demonstrate that the DDN shared parallel architecture scales linearly and fully achieves the capabilities of the ten AI400X appliances, 500 GB/s throughput for read and 350 GB/s throughput for write, with 16 DGX A100 systems engaged. This performance is maintained and balanced evenly with up to 128 DGX A100 systems simultaneously.

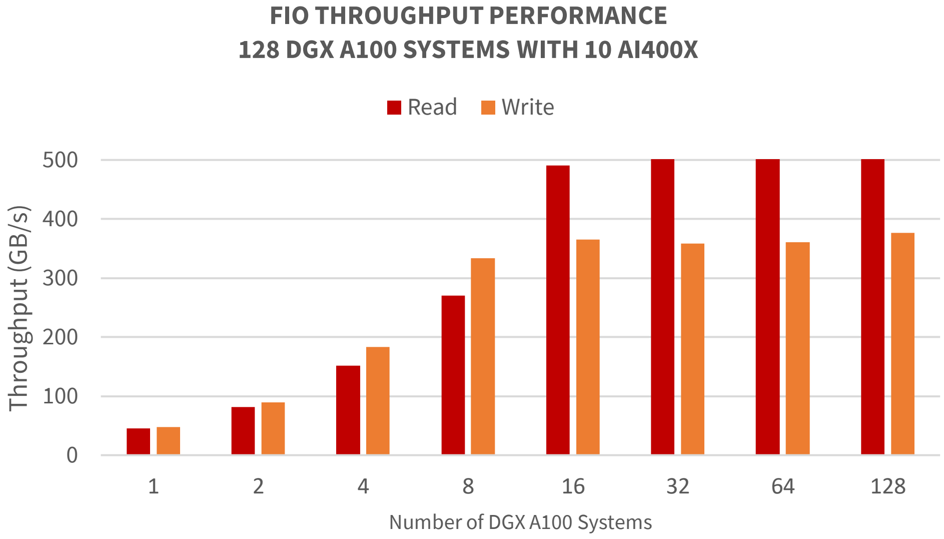


Figure 24. FIO throughput scaling with a very large number of DGX A100 systems.

For more information on at-scale performance and validation with DGX A100 systems and AI400X appliances, consult the NVIDIA DGX SuperPOD: DDN A^{3}I AI400X Appliance (RA-10133-001).

6. Contact DDN to Unleash the Power of Your DGX POD

DDN has long been a partner of choice for organizations pursuing at-scale data-driven projects. Beyond technology platforms with proven capability, DDN provides significant technical expertise through its global research and development and field technical organizations.

A worldwide team with hundreds of engineers and technical experts can be called upon to optimize every phase of a customer project: initial inception, solution architecture, systems deployment, customer support and future scaling needs.

Strong customer focus coupled with technical excellence and deep field experience ensures that DDN delivers the best possible solution to any challenge. Taking a consultative approach, DDN experts will perform an in-depth evaluation of requirements and provide application-level optimization of data workflows for a project. They will then design and propose an optimized, highly reliable and easy to use solution that best enables and accelerates the customer effort.

Drawing from the company's rich history in successfully deploying large scale projects, DDN experts will create a structured program to define and execute a testing protocol that reflects the customer environment and meet and exceed project objectives. DDN has equipped its laboratories with leading GPU compute platforms to provide unique benchmarking and testing capabilities for AI and DL applications.

Contact DDN today and engage our team of experts to unleash the power of your AI projects.

About DDN

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. DDN has designed, developed, deployed, and optimized systems, software, and solutions that enable enterprises, service providers, research facilities, and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud.