NetApp Verified Architecture

NetApp EF-Series AI with NVIDIA DGX A100 Systems and BeeGFS

NVA Design

Abdel Sadek, Tim Chau, Joe McCormick and David Arnette, NetApp

March 2021 | NVA-1156-DESIGN

Abstract

This document describes a NetApp Verified Architecture for machine learning (ML) and artificial intelligence (AI) workloads using NetApp® EF600 NVMe storage systems, the BeeGFS parallel file system, NVIDIA DGX™ A100 systems, and NVIDIA® Mellanox® Quantum™ QM8700 200Gbps IB switches. This design features 200Gbps InfiniBand (IB) for the storage and compute cluster interconnect fabric to provide customers with a completely IB-based architecture for high-performance workloads. This document also includes benchmark test results for the architecture as implemented.

In partnership with

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Executive summary

This document contains validation information for the NetApp EF-Series AI reference architecture for ML and AI workloads. This design was implemented using a NetApp EF600 all-flash NVMe storage system, the ThinkParQ BeeGFS parallel filesystem, eight DGX A100 systems, and QM8700 switches for both the compute cluster interconnect and storage connectivity. The operation and performance of this system was validated using industry-standard benchmark tools and proven to deliver excellent training performance. Customers can easily and independently scale compute and storage resources from half-rack to multi-rack configurations with predictable performance to meet any machine learning workload requirement.

Program summary

The NetApp Verified Architecture program provides customers with reference configurations and sizing guidance for specific workloads and use cases. These solutions are:

- Thoroughly tested
- Designed to minimize deployment risks
- Designed to accelerate time to market

This document is for NetApp and partner solutions engineers and customer strategic decision makers. It describes the architecture design considerations that were used to determine the specific equipment, cabling, and configurations required to support the validated workload.

NetApp EF-Series AI solution

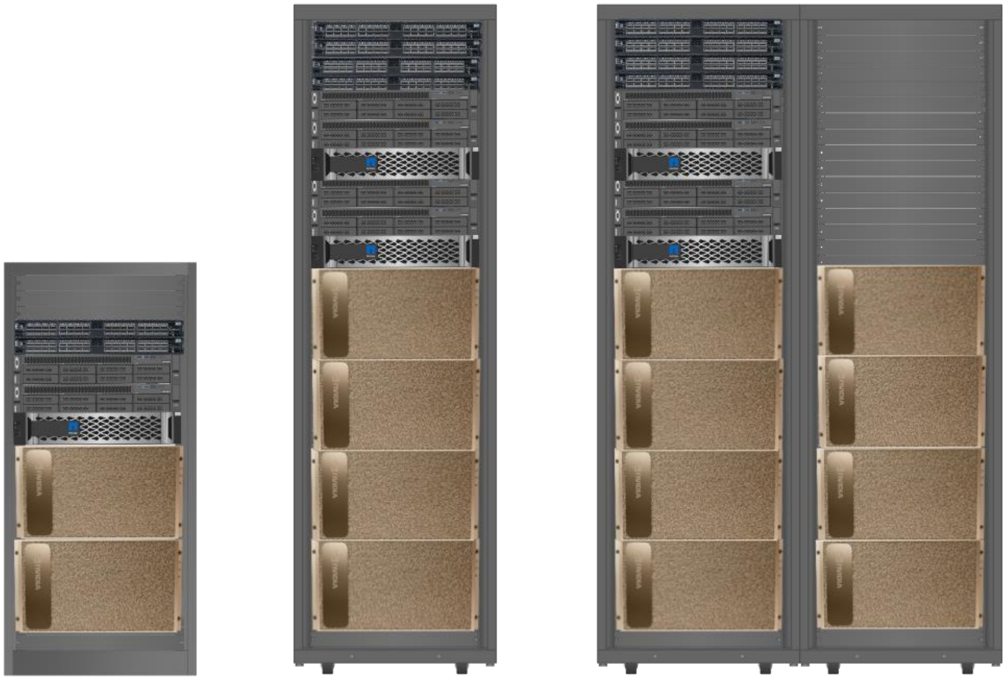
The NetApp EF-Series AI reference architecture, powered by DGX A100 systems and NetApp cloud-connected storage systems, was developed and verified by NetApp and NVIDIA. It gives IT organizations an architecture that:

- Eliminates design complexities
- Allows independent scaling of compute and storage
- Enables customers to start small and scale seamlessly
- Offers a range of storage options for various performance and cost points

NetApp EF-Series AI tightly integrates DGX A100 systems, NetApp EF600 NVMe storage systems, and the BeeGFS parallel file systems with state-of-the-art IB networking. NetApp EF600 AI simplifies artificial intelligence deployments by eliminating design complexity and guesswork. Customers can start small and scale seamlessly from science experiments and proof-of-concepts to production and beyond.

Figure 1 shows several variations in the EF-Series AI family of solutions with DGX A100 systems. The EF600 powered BeeGFS building blocks have been verified with up to eight DGX A100 systems. By adding more of these building blocks, the architecture can scale to multiple racks supporting many DGX A100 systems and petabytes of storage capacity. This approach offers the flexibility to alter compute-to-storage ratios independently based on the size of the data lake, the deep learning (DL) models that are used, and the required performance metrics.

Figure 1) NetApp EF-Series-based BeeGFS building blocks for AI with NVIDIA DGX A100 systems.



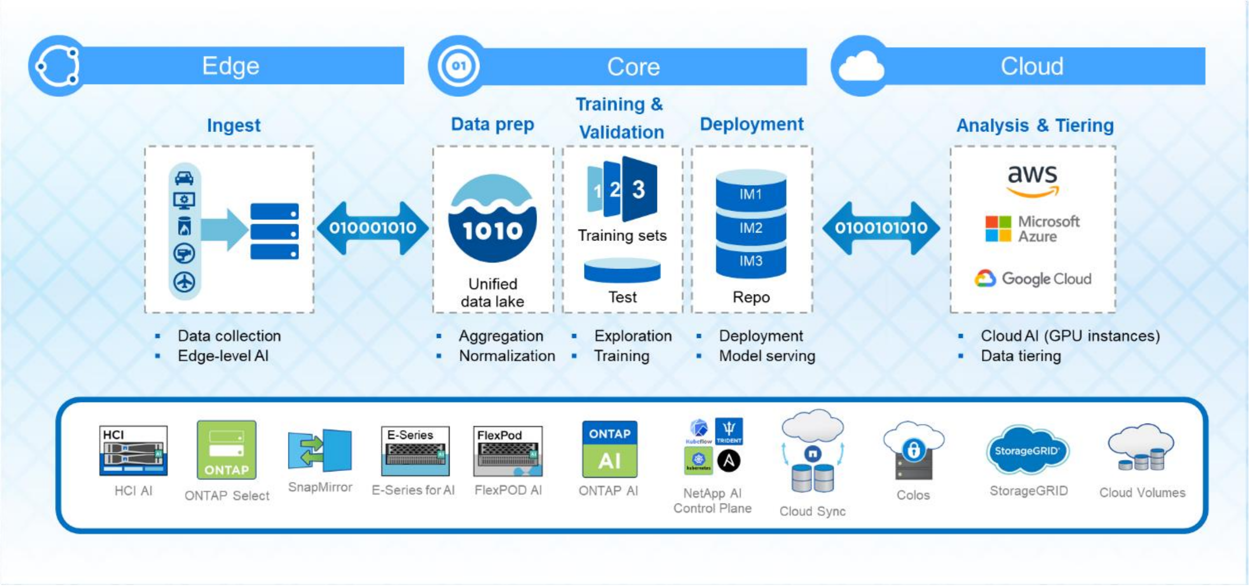
The number of DGX A100 systems and EF600 systems per rack depends on the power and cooling specifications of the rack in use. Final placement of the systems is subject to computational fluid dynamics analysis, airflow management, and data center design.

Deep learning data pipeline

DL is the engine that enables businesses to detect fraud, improve customer relationships, optimize supply chains, and deliver innovative products and services in an increasingly competitive marketplace. The performance and accuracy of DL models are significantly improved by increasing the size and complexity of the neural network as well as the amount and quality of data that is used to train the models.

Given the massive datasets required, it is crucial to architect an infrastructure that offers the flexibility to deploy across environments. At a high level, an end-to-end DL deployment consists of three phases through which the data travels: the edge (data ingest and inferencing), the core (training clusters and a data lake), and the cloud (archive, tiering, and dev/test). This is typical of applications such as the Internet of Things (IoT) for which data spans all three realms of the data pipeline. Figure 2 presents an overview of the components in each of the three realms.

Figure 2) Components of the edge-core-cloud data pipeline.



The following list describes some of the activities that occur in one or more of these areas.

- **Ingest** Data ingestion usually occurs at the edge, for example, by capturing data streaming from autonomous cars or point-of-sale devices. Depending on the use case, an IT infrastructure might be needed at or near the ingestion point. For example, a retailer might need a small footprint in each store that consolidates data from multiple devices.
- **Data prep.** Preprocessing is necessary to normalize and cleanse the data before training. Preprocessing takes place in a data lake, possibly in the cloud, in the form of an Amazon S3 tier or in on-premises storage systems such as a file store or an object store.
- **Training and validation.** For the critical training phase of DL, data is typically copied from the data lake into the training cluster at regular intervals. The servers that are used in this phase use GPUs to parallelize computations, creating a tremendous appetite for data. Meeting the raw I/O bandwidth needs is crucial for maintaining high GPU utilization.
- **Deployment.** The trained models are tested and deployed into production. Alternatively, they could be fed back to the data lake for further adjustments of input weights, or, in IoT applications, the models could be deployed to smart edge devices.
- **Analysis and tiering.** New cloud-based tools become available at a rapid pace, so additional analysis or development work might be conducted in the cloud. Cold data from past iterations might be saved indefinitely. Many AI teams prefer to archive cold data to object storage in either a private or a public cloud. Based on compute requirements, some applications work well with object storage as the primary data tier.

Depending on the application, DL models work with large amounts of structured and unstructured data. This difference imposes a varied set of requirements on the underlying storage system, both in terms of size of the data that is being stored and the number of files in the dataset.

High-level storage requirements include:

- The ability to store and retrieve millions of files concurrently.
- Storage and retrieval of diverse data objects such as images, audio, video, and time-series data
- Delivery of highly parallel performance at low latencies to meet the GPU processing speeds.

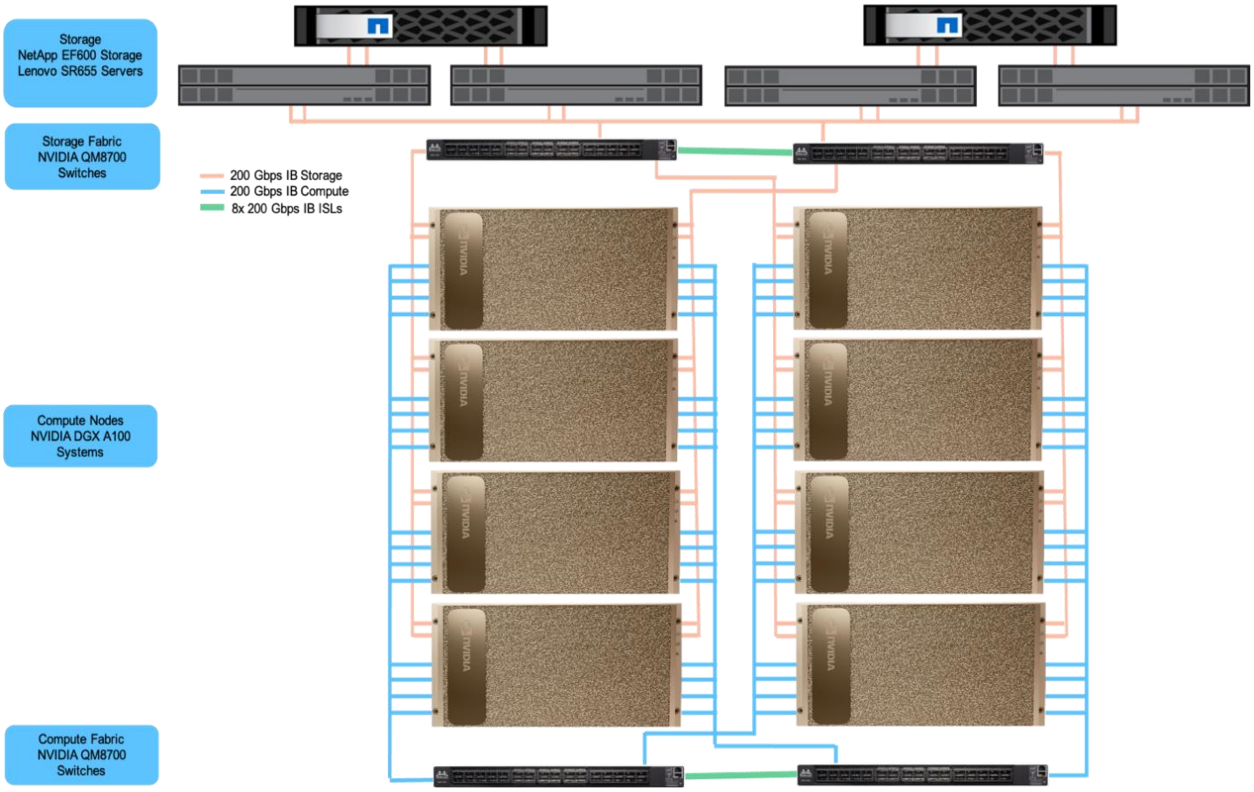
This document focuses on solutions for the training and inference components of the data pipeline.

Solution overview

DL systems leverage algorithms that are computationally intensive and that are uniquely suited to the architecture of GPUs. Computations that are performed in DL algorithms involve an immense volume of matrix multiplications running in parallel. Advances in individual and clustered GPU computing architectures leveraging DGX systems have made them the preferred platform for workloads such as high-performance computing (HPC), DL, video processing, and analytics. Maximizing performance in these environments requires a supporting infrastructure, including storage and networking, that can keep GPUs fed with data. Dataset access must therefore be provided at ultra-low latencies with high bandwidth.

This reference architecture was validated with two BeeGFS building blocks. Each block is comprised of one NetApp EF600 system connected using NVMe/IB to two Lenovo x86 servers running BeeGFS server services. Eight DGX A100 systems running BeeGFS client services connected to two NVIDIA Mellanox QM8700 200Gb IB switches for compute and storage operations. Figure 3 shows the basic solution architecture.

Figure 3) NetApp EF-Series with BeeGFS AI verified architecture.



NVIDIA DGX A100 systems

The DGX A100 system is a fully integrated, turnkey hardware and software system that is purpose-built for DL workflows. Each DGX A100 system is powered by eight NVIDIA A100 GPUs that are configured in a hybrid cube-mesh topology that uses NVIDIA NVLink® and NVIDIA NVSwitch® technologies. This configuration provides an ultra-high bandwidth, low-latency fabric for inter-GPU communication within the DGX A100 system. This topology is essential for multi-GPU training, eliminating the bottleneck that is associated with PCIe-based interconnects that cannot deliver linearity of performance as the GPU count increases. The DGX A100 system is also equipped with high-bandwidth, low-latency network interconnects for multinode clustering over IB.

NVIDIA NGC

The DGX A100 system leverages NVIDIA NGC, a cloud-based container registry for GPU-accelerated software. NGC provides containers for today's most popular DL frameworks such as Caffe2, TensorFlow, PyTorch, MXNet, and TensorRT, which are optimized for NVIDIA GPUs. The containers integrate the framework or application, necessary drivers, libraries, and communications primitives, and they are optimized across the stack by NVIDIA for maximum GPU-accelerated performance. NGC containers incorporate the NVIDIA CUDA Toolkit, which provides the CUDA Basic Linear Algebra Subroutines Library (cuBLAS), the CUDA Deep Neural Network Library (cuDNN), and much more. The NGC containers also include the NVIDIA Collective Communications Library (NCCL) for multi-GPU and multinode collective communication primitives, enabling topology-awareness for DL training. NCCL enables communication between GPUs inside a single DGX A100 system and across multiple DGX A100 systems.

NetApp EF600 systems

In the highly competitive world of business, speed is everything. However, even the fastest supercomputer cannot meet expectations if it does not have equally fast storage to support it. The NetApp EF600 all-flash array gives customers consistent, near-real-time access to data while supporting any number of workloads simultaneously. To enable fast, continuous feeding of data to AI applications, EF600 storage systems deliver up to two million cached read IOPS, response times of under 100 microseconds, and 42GBps sequential read bandwidth in one enclosure. With 99.9999% reliability from EF600 storage systems, data for AI operations is available whenever and wherever it is needed.

Key benefits

The key benefits of EF600 storage systems include:

- **Accelerate time to Insight.** Enable blazing fast streaming of data to AI applications with high-throughput, low-latency storage.
- **Future-proof your investment.** Quickly respond to changing workload demands and exponential data growth with a building block architecture that enables seamlessly scaling of performance and capacity as needed.
- **Maximize cost efficiency.** Reduce operating costs with high-density drives and price/performance optimized storage building blocks to ensure you can spend more on GPUs than storage.
- **Reduce risk and enable success.** Rely on a fully integrated, validated AI infrastructure from industry leaders to help gain a competitive edge. Maximize productivity with 99.9999% availability.

ThinkParQ BeeGFS parallel file system

BeeGFS is a parallel file system with an architecture based on the following four main services:

- **Management service.** Registers and monitors all other services.
- **Storage service.** Stores the distributed user file contents known as data chunk files.
- **Metadata service.** Keeps track of the file system layout, directory and file attributes, and so on.
- **Client service.** Mounts the file system to access the stored data.

This design provides flexibility that is key to meeting diverse and evolving AI workloads. NetApp EF-Series storage systems supercharge BeeGFS storage and metadata services by offloading RAID and other storage tasks including drive monitoring and wear detection.

Key benefits

The key benefits of BeeFGS parallel file system include:

- Allows optimization for diverse AI workloads within a single storage namespace.

Do your GPUs each need to access a large number of small files? Do they each need to access a single large file? Do they all need to access the same set of small or large files? Don't know? Many storage solutions are only good at some of these. BeeGFS does it all.

- Designed and developed for ease of use, straightforward installation, and simple management.

  Eliminate complexity associated with traditional parallel and distributed file systems while taking full advantage of the performance benefits.
- Get data to GPUs faster by using remote direct memory access (RDMA) over IB.

  For servers that don't support RDMA, BeeGFS can serve files over TCP/IP and RDMA concurrently ensuring no GPU is left out.
- Intelligently distributed file contents and metadata optimized for highly concurrent access.

  Avoid fundamental architectural limitations imposed by the design of some storage solutions.

For detailed information about a BeeGFS deployment, see TR-4755: BeeGFS with NetApp E-Series.

NVIDIA IB networking

The demand for more computing power, efficiency, and scalability continues to grow in the HPC, cloud, Web 2.0, ML, data analytics, and storage markets. To address these demands, NVIDIA networking provides complete end-to-end solutions supporting IB networking technologies. NVIDIA Mellanox IB products deliver the highest productivity, enabling compute clusters and converged data centers to operate at any scale.

NVIDIA Mellanox Quantum switches—the right choice for deep learning workloads

Mellanox Quantum HDR 200Gbps IB Smart Switches deliver a complete chassis and fabric management that enables managers to build highly cost-effective and scalable switch fabrics ranging from small clusters up to thousands of nodes, reducing operational costs and infrastructure complexity.

The QM8700 series has the highest fabric performance available in the market with up to 16Tbps of nonblocking bandwidth with sub 130ns port-to-port latency. Advanced features such as adaptive routing, congestion control, and enhanced quality of service (QoS) ensure the maximum fabric performance under all types of traffic conditions, and its best-in-class design supports low power consumption.

World's most advanced networking

In-network computing is the offloading of standard applications to run within network devices. This is synonymous with NVIDIA Networking Scalable Hierarchical Aggregation and Reduction Protocol (SHARP). SHARP technology improves upon the performance of Message Passing Interface (MPI) and ML collective operations, by offloading the operations from the CPU or GPU to the network and eliminating the need to send data multiple times between endpoints.

This innovative approach decreases the amount of data traversing the network and dramatically reduces the collective operations processing time. Implementing collective communication algorithms in the network also has additional benefits, such as freeing up valuable CPU resources for computation rather than using them to process communication.

Mellanox IB switching also provides native self-healing autonomy which enables failed communications links to be corrected in hardware without intervention of the subnet manager. This novel approach improves system resiliency and results in saving communications from expensive which can often result in application failure.

NVIDIA Mellanox Unified Fabric Manager

To easily manage scale-out IB computing environments, Mellanox IB switches can be coupled with NVIDIA Mellanox Unified Fabric Manager (UFM).

UFM platforms empower research and industrial data center operators to efficiently provision, monitor, manage, and preventatively troubleshoot and maintain the modern data center fabric, to realize higher utilization of fabric resources and a competitive advantage, while reducing OPEX.

From workload optimizations and configuration checks, to improving fabric performance through AI-based detection of network anomalies and predictive maintenance, UFM platforms provide a comprehensive feature set to meet the broadest range of modern scale-out data center requirements.

Technology requirements

This section covers the hardware and software that was used for the testing described in the “Solution verification” section.

Hardware requirements

Table 1 lists the hardware components that were used to verify this solution.

Table 1) Hardware requirements.

Hardware	Quantity
DGX A100 systems	Eight
EF600 storage system	Two high-availability (HA) pair, includes 24x 1.92TB NVMe SSDs on each
QM8700 IB switches	Two for compute cluster interconnect and to BeeGFS servers
Lenovo SR655 servers	Four running BeeGFS server services

Software requirements

Table 2 lists the software components that were used to validate the solution.

Table 2) Software requirements.

Software	Version
SANtricity OS	11.62.00.9012
BeeGFS	7.2
BeeGFS servers OS	RedHat Enterprise Linux RHEL 7.8
BeeGFS servers OFED	5.1-0.6.6
BeeGFS servers multipathing	Device Mapper MultiPath (DMMP)
IB switch OS	3.9.0606
DGX OS	4.99.10
Docker container platform	19.03.8
Container version	nvcr.io/nvidia/mxnet:20.06-py3 – MLPerf test tensorflow:20.05-tf2-py3 – other tests
DGX A100 OFED	5.3.0-59
NCCL test version	https://github.com/NVIDIA/nccl-tests/tree/ec1b5e22e618d342698fda659efdd5918da6bd9f
FIO version	3.1

Solution architecture

This reference architecture is verified to meet the requirements for running DL workloads. It enables data scientists to deploy DL frameworks and applications on a prevalidated infrastructure, eliminating risks and allowing businesses to focus on gaining valuable insights from their data. This architecture also delivers exceptional storage performance for other HPC workloads without any modification or tuning of the infrastructure.

Network topology and switch configuration

This reference architecture is based 200Gbps IB HDR fabrics for compute-cluster interconnect and storage access. Both fabrics use a pair of NVIDIA Mellanox QM8700 IB switches operating as independent redundant fabrics. Each DGX A100 system is connected to the switches using 10 single-ported ConnectX-6 cards at 200Gbps, with even-numbered ports connected to one switch and odd-numbered ports connected to the other switch. Eight of the IB ports connected to the IB switches are used for GPU-to-GPU communications between the DGX A100 compute nodes. The other two IB ports connected to the same switches are used for storage traffic between DGX A100 and the BeeGFS servers building blocks.

Storage system configuration

Regardless of whether you’re considering a white box server or enterprise storage system, there are many factors to consider such as RAID level, optimal drive count per service, collocation of BeeGFS storage and metadata on the same underlying storage media, and more. Extensive testing by the NetApp team of certified BeeGFS systems engineers, eliminates guess work when deploying BeeGFS on E-Series.

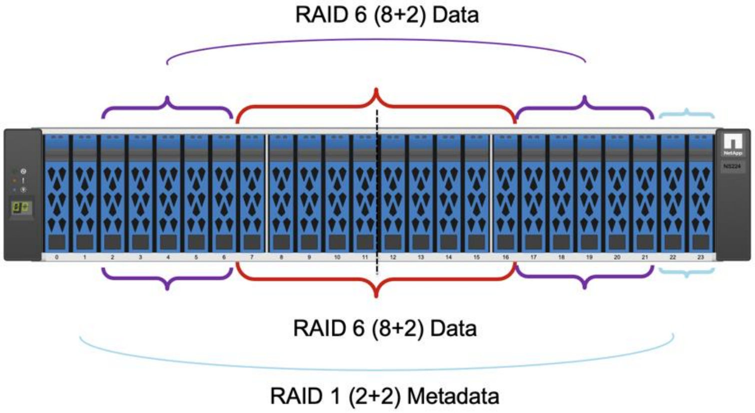
This solution was validated with 24x 3.8TB NVMe SSDs, which deliver 55TB of usable storage capacity. Currently, EF600 systems are available with multiple SSD sizes up to 15TB for a maximum capacity of 335TB. For optimal performance, each EF600 system was deployed with the following storage configuration:

- Storage volumes: Two 8+2 RAID 6 volume groups with four volumes each and eight total volumes.
- Each BeeGFS storage service is assigned two 6143.78GiB volumes with a 256KiB segment size (amount of data written to each drive before moving to the next).
- Metadata volumes: One 2+2 RAID 10 volume group with four volumes.
- Each BeeGFS Metadata service is assigned one 1535.94GiB volume with a 64KiB segment size (optimized for small I/O).
- Each volume group is set at 14% optimization capacity to optimize SSD performance and wear life.
- Each volume has read/write caching enabled with write cache being mirrored between controllers.

  Note: Dynamic cache read prefetch was disabled.

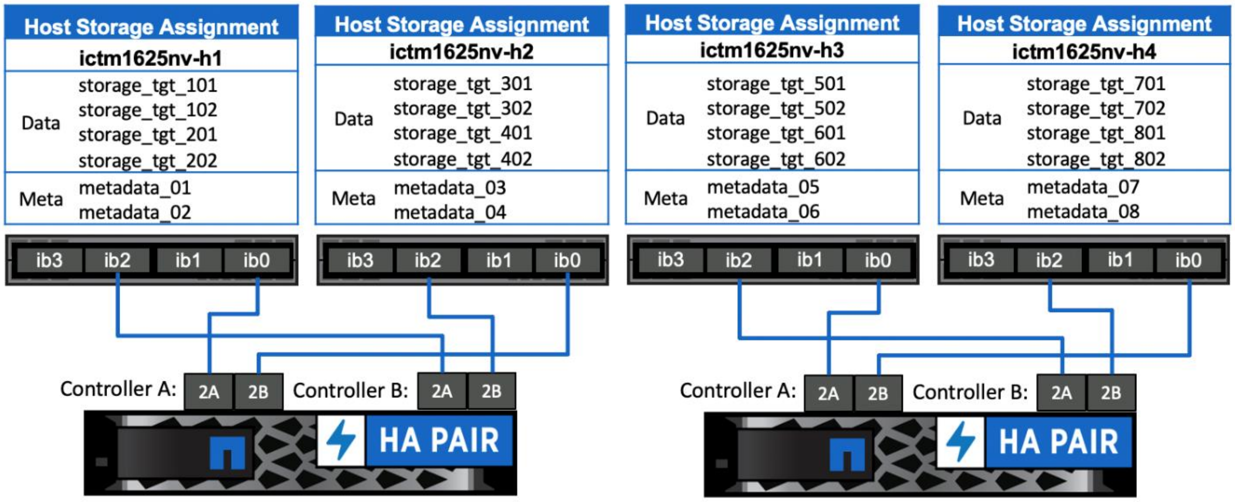
To help the storage controllers use both drive-side PCIe buses more effectively, drives are evenly assigned to each volume group from drives 0–12 and 12–23 (see Figure 4).

Figure 4) Optimized EF600 volume group configuration for BeeGFS.



As tested, the EF600 storage system includes two controllers with four IB ports per controller. For this validation, one port from each controller was connected to each BeeGFS server over NVMe/IB, providing a total of 400Gbps bandwidth (Figure 5).

Figure 5) EF600 to BeeGFS storage volume assignments and connectivity.



Host Storage Assignment	
ictm1625nv-h1	
Data	storage_tgt_101 storage_tgt_102 storage_tgt_201 storage_tgt_202
Meta	metadata_01 metadata_02

Host Storage Assignment	
ictm1625nv-h2	
Data	storage_tgt_301 storage_tgt_302 storage_tgt_401 storage_tgt_402
Meta	metadata_03 metadata_04

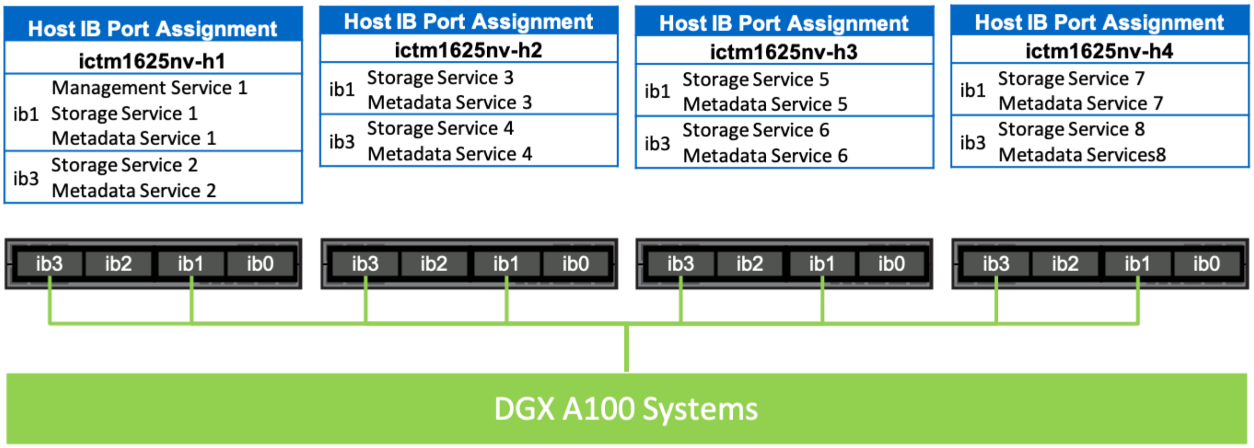
Host Storage Assignment	
ictm1625nv-h3	
Data	storage_tgt_501 storage_tgt_502 storage_tgt_601 storage_tgt_602
Meta	metadata_05 metadata_06

Host Storage Assignment	
ictm1625nv-h4	
Data	storage_tgt_701 storage_tgt_702 storage_tgt_801 storage_tgt_802
Meta	metadata_07 metadata_08

Host configuration

As storage overhead such as RAID is offloaded to the EF600 systems, each BeeGFS server can run multiple services to maximize performance and network utilization while reducing rack space. Each BeeGFS service is configured with preferred IB ports, though services can fall back to the secondary IB port/host channel adapter (HCA) if needed. Figure 6 shows the BeeGFS services running on each server along with the preferred IB ports for each service, and physical connections to the Mellanox QM8700 switches.

Figure 6) DGX A100 to BeeGFS server networking.



The topology highlighted in Figure 5 and Figure 6 result in the EF600 systems being directly connected to one port of each HCA (ib0/ib2) on the Lenovo SR655 servers, while the DGX A100 systems are connected to other ports (ib1/ib3). Because HCAs have separate send/receive buffers, this deliberate design choice allows each HCA (in the case of writes) to effectively use the bidirectional bandwidth on PCIe to handle the incoming traffic (RECV) from the DGX systems while simultaneously handling outgoing traffic (SEND) to the EF600 systems (inversely on reads).

The two IB ports, ib4 and in10, in each of the DGX A100 systems is cabled across two QM 8700 IB switches for the storage fabric. The switches have eight Inter-Switch Links (ISLs).

A separate set of two QM8700 IB switches is used as compute fabric with eight ISLs connecting them. Each DGX A100 system is connected to the two QM8700 IB switches using eight single-ported ConnectX-6 cards at 200Gbps (ib0 through ib3 and ib6 through ib9). Those IB ports connected to the IB switches are configured to enable the lowest possible latency for GPU-to-GPU communications between the DGX A100 compute nodes.

Solution verification

This reference architecture was validated using synthetic benchmark utilities and DL benchmark tests to establish baseline performance and operation of the system. Each of the tests described in this section was performed with the specific equipment and software listed in “Technology requirements.”

Infrastructure validation

The following tests were performed with one, two, and four DGX A100 systems to validate basic operation and performance of the deployed infrastructure:

- NVIDIA nvsm stress test

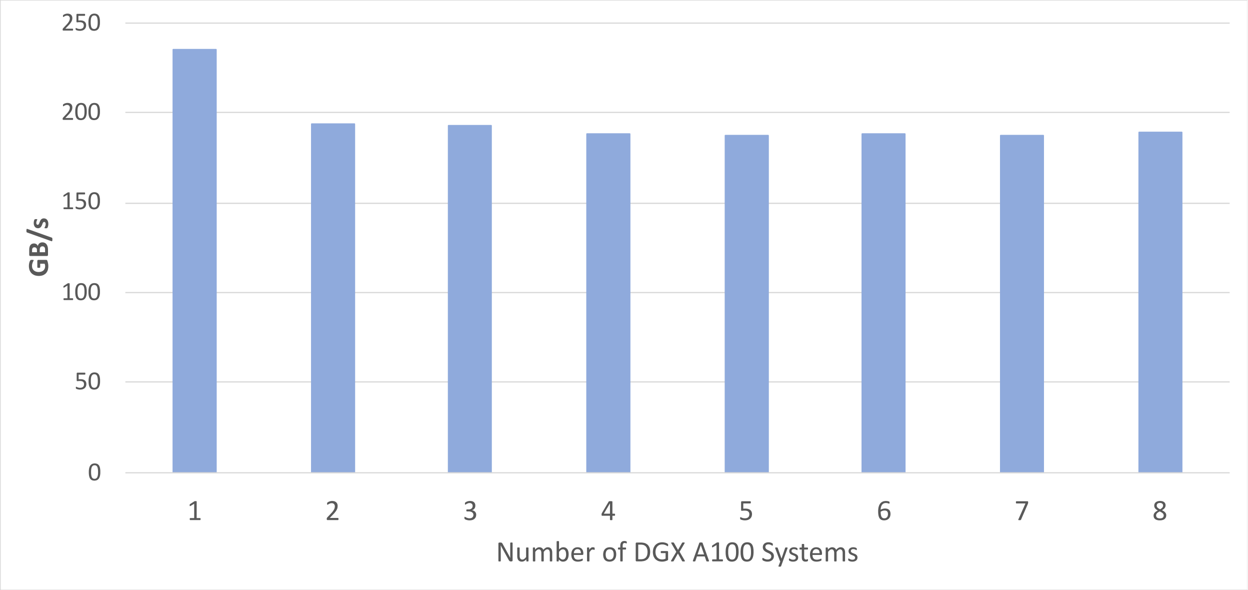
  This test suite performs a pass/fail verification of many crucial DGX A100 systems. All systems should report a passing status for the tests in this group.
- NVIDIA NCCL all_reduce_perf
- FIO bandwidth test
- FIO I/O operations per second (IOPS) test

The following sections provide details and results for these tests.

NVIDIA NCCL all_reduce_perf test

This test validates the performance of the interconnects between GPUs. For single-node systems, the bottleneck should be the NVIDIA NVLink connection between GPUs. For multinode systems, the bottleneck should be the Ethernet or IB connections between DGX A100 systems. This test measures the combined bandwidth between systems using all eight available physical connections. Figure 7 shows the results of the NCCL all_reduce_perf test.

Figure 7) NCCL bandwidth test result with a BeeGFS stripe setting of 16 targets per file and a chunk size of 2MB.



FIO bandwidth and IOPS tests

These tests are intended to measure the storage system performance using the synthetic I/O generator tool FIO. Two separate configurations were used, one optimized to deliver maximum bandwidth and the other optimized for IOPS. Each configuration was run with both 100% reads and 100% writes, and the files used by FIO were created as a separate step to isolate those activities from the actual test results. Here are the specific FIO configuration parameters for these tests:

- ioengine = posixaio
- direct = 1
- blocksize = 1024k for bandwidth test, 4k for IOPS test
- numjobs = 64 for bandwidth test, 64 for IOPS test
- iodepth = 64
- size = 4194304k

Figure 8 shows the results of the FIO bandwidth tests with up to eight DGX A100 systems. Performance scales in a linear manner with up to three nodes achieving 55GBps and peaking at 65.7GBps at eight nodes.

Figure 8) FIO bandwidth test results (GBps) with a BeeGFS stripe setting of one target per file and a chunk size of 512KB.

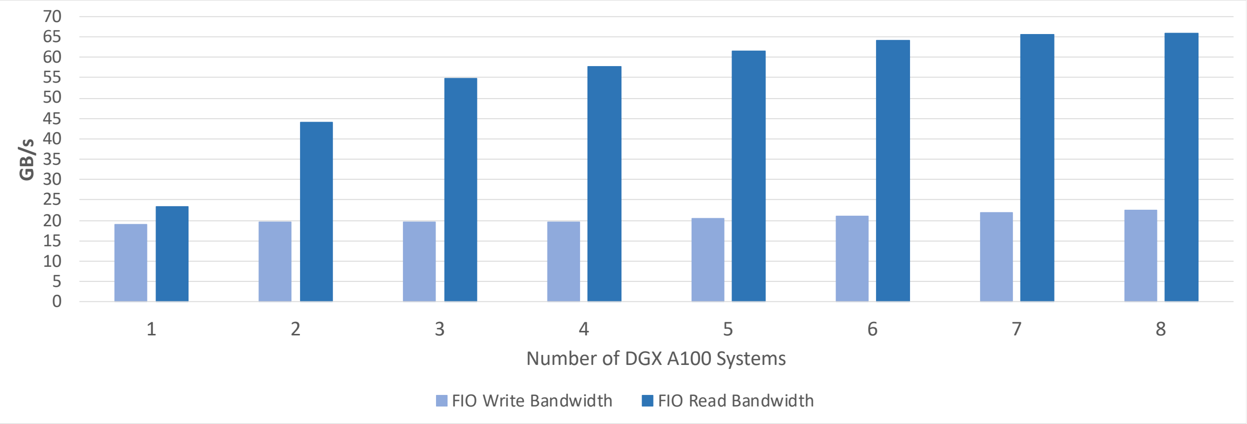
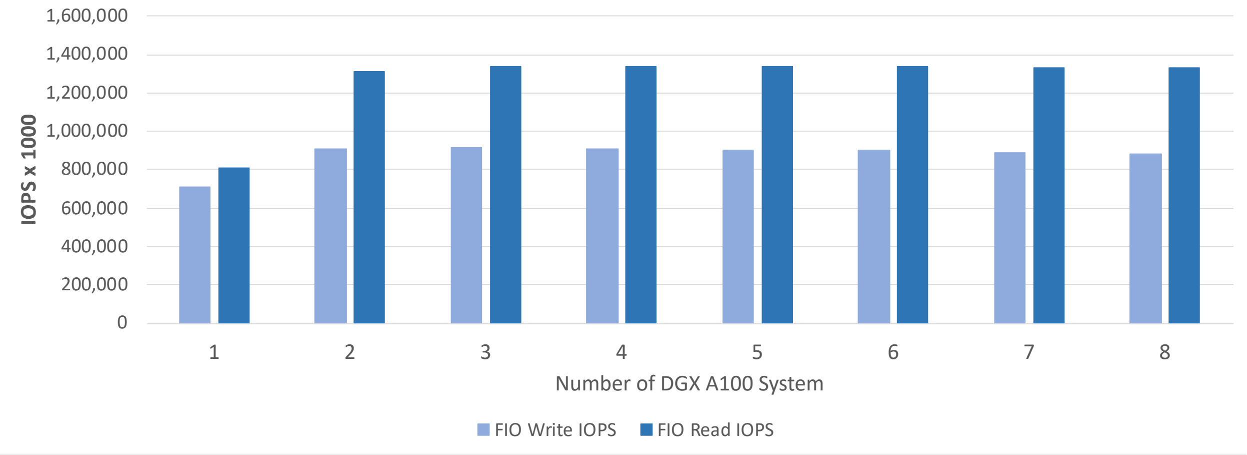


Figure 9 shows the results of the FIO IOPS test.

Figure 9) FIO IOPS test results (operations/second) with a BeeGFS stripe setting of one target per file and a chunk size of 512KB.



Deep learning workload validation

The operation of DL workloads on the deployed infrastructure was validated using the MLPerf Training v0.7 ResNet-50 benchmark test. This test uses the MLPerf v0.7 testing criteria for validating performance of systems using the ResNet-50 model with the parameters and dataset specified in the MLPerf v0.7 testing specification.

The following section provides specific details and results for this test.

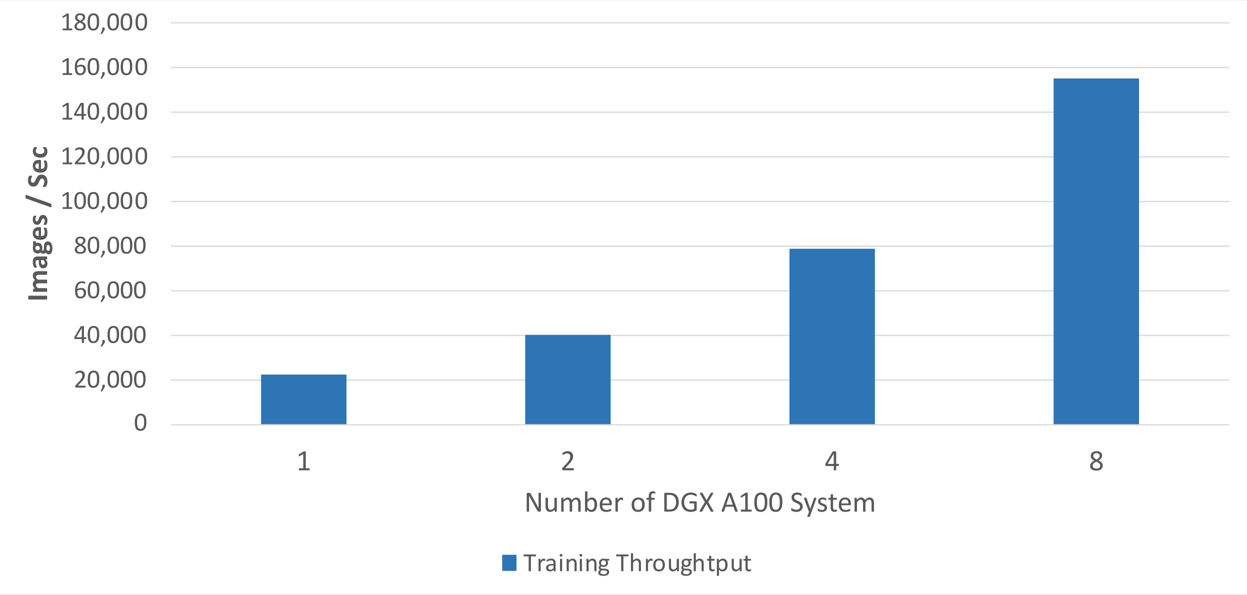
MLPerf Training v0.7 ResNet-50

This reference architecture was tested using a MLPerf Training v0.7 benchmark to validate the operation of DL workloads on the deployed infrastructure. MLPerf is an industry-standard benchmark implementation of various neural networks for validating the performance of DL infrastructure. This test used the MXNet implementation of ResNet-50 in addition to the ImageNet dataset in IORecord format to validate model training performance. DALI was used to accelerate ingest and preprocessing of data, and Horovod was used to distribute the training across multiple DGX A100 systems. The results presented maintained a consistent batch size per system of 408 images as the workload was scaled (weak scaling).

The base container image used for these tests was the 20.06 MXNet image from NGC. MLPerf benchmark tests were deliberately not optimized for any specific hardware implementation so that the overall system performance in these tests could be increased by tuning parameters such as concurrency.

Figure 10 shows the average images per second for the training run duration of 45 epochs.

Figure 10) MLPerf Training v0.7 average images per second with a BeeGFS stripe setting of 16 targets per file and a chunk size of 2MB.



Solution sizing guidance

This architecture is intended as a reference for customers and partners who would like to implement a DL infrastructure with NVIDIA DGX A100 systems and a NetApp EF-Series system.

As is demonstrated in this validation, the EF600 system and BeeGFS easily support the DL training workload generated by eight DGX A100 systems. For larger deployments with higher storage performance requirements, additional BeeGFS building blocks can be added to the storage cluster. While BeeGFS does not publish hard scaling limits, there are customer deployments to 30PB and no known theoretical limits. Although the dataset used in this validation was relatively small, BeeGFS can scale to massive capacities with linear performance scalability, because each building block delivers performance comparable to the level verified in this document.

Other NetApp storage systems such as the NetApp EF300 all-flash array provide performance for smaller GPU footprints at a lower cost. For a low-cost cold tier or when workloads are highly sequential, the E5760 system provides high density building blocks with a maximum capacity of 7.68PB. Using the BeeGFS storage pools feature, a single BeeGFS file system can use multiple classes of storage building blocks. Writing to various classes of storage is as simple as designating some directories as “fast” and others as “slow.”

With industry-proven building blocks from NetApp and the flexibility of BeeGFS storage pools, customers can start with a low-cost initial footprint, and scale out performance and/or capacity seamlessly to allow AI projects to go from proof-of-concept to production without worrying about storage.

Conclusion

The DGX A100 system is a next-generation DL platform that requires equally advanced storage and data management capabilities. By combining a DGX A100 system with BeeGFS building blocks based on NetApp EF600 systems, this verified architecture can be implemented at almost any scale, from a single DGX A100 system paired with a single BeeGFS building block up to potentially 120 DGX A100 systems with a scalable number of BeeGFS building blocks presenting a single storage namespace. Combined with the superior cloud integration and software-defined capabilities of the NetApp product portfolio, NetApp storage solutions enable a full range of data pipelines that span the edge, the core, and the cloud for successful DL projects.

Acknowledgments

The authors gratefully acknowledge the contributions that were made to this technical report by our esteemed colleagues from NVIDIA and NetApp. Our sincere appreciation and thanks go to all the individuals who provided insight and expertise that greatly assisted in the research for this paper.

Where to find additional information

To learn more about the information that is described in this document, review the following documents and/or websites:

NetApp EF-Series systems

- NetApp EF-Series product page
  https://www.netapp.com/us/products/storage-systems/all-flash-array/ef-series.aspx
- EF600 datasheet
  https://www.netapp.com/pdf.html?item=/media/19339-DS-4082.pdf
- NetApp AI and HPC solutions
  https://www.netapp.com/artificial-intelligence/high-performance-computing/
- BeeGFS with NetApp E-Series Solution Deployment
  https://www.netapp.com/pdf.html?item=/media/17132-tr4755pdf.pdf

NetApp Interoperability Matrix

- NetApp Interoperability Matrix Tool
  http://support.netapp.com/matrix

NVIDIA DGX A100 systems

- NVIDIA DGX A100 systems
  https://www.nvidia.com/en-us/data-center/dgx-a100/
- NVIDIA A100 Tensor core GPU
  https://www.nvidia.com/en-us/data-center/a100/
- NVIDIA GPU Cloud
  https://www.nvidia.com/en-us/gpu-cloud/

NVIDIA Mellanox networking

- NVIDIA Mellanox Quantum QM8700 series IB switches
  https://www.nvidia.com/en-us/networking/infiniband/qm8700/

Machine learning frameworks

- TensorFlow: An Open-Source Machine Learning Framework for Everyone
  https://www.tensorflow.org/
- Horovod: Uber's Open-Source Distributed Deep Learning Framework for TensorFlow
  https://eng.uber.com/horovod/
- Enabling GPUs in the Container Runtime Ecosystem
  https://devblogs.nvidia.com/gpu-containers-runtime/

Dataset and benchmarks

- ImageNet
  http://www.image-net.org/
- MLPerf training and inference benchmarks
  https://mlperf.org/

Version history

Version	Date	Document version history
Version 1.0	March 2021	Initial release

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

Data contained herein pertains to a commercial item (as defined in FAR 2.101) and is proprietary to NetApp, Inc. The U.S. Government has a non-exclusive, non-transferrable, non-sublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b).

Trademark Information

NetApp