WEKA Data Platform for NVIDIA DGX BasePOD with H100 Systems

Scalable Performance for Highly Parallelized Enterprise AI Workloads

REFERENCE ARCHITECTURE

Contents

Executive
Summary

Organizations of various sizes, use cases, and technical skills are looking for easy to deploy infrastructure solutions to accelerate their artificial intelligence (AI), machine learning (ML), and deep learning (DL) initiatives.

WEKA worked with NVIDIA® to validate a high-performance, scalable AI storage solution to support NVIDIA DGX BasePOD™ requirements that can be made accessible to everyone. This document contains information for the WEKA Data Platform reference architecture solution for DGX BasePOD.

This design was implemented using up to sixteen NVIDIA DGX™ H100 systems and NVIDIA Quantum™ switches. The operation and performance of this system were validated by NVIDIA and WEKA using NVIDIA benchmarking tools.

This architecture delivers excellent linear scaling for highly parallized, compute-intensive workloads based on the validation testing results. With this architecture, organizations can start small and quickly and independently scale compute and storage resources up to multi-rack configurations with predictable performance to meet any AI or ML workload requirement.

Tested Solution Details

Product Name	Product Version	WEKA Version	Source	WEKA OS
NVIDIA DGX	H100	4.2	On-Premises	CentOS

1. Introduction

Audience

We wrote this reference architecture for those who design, manage, and support the WEKA® Data Platform. Consumers of this document should already be familiar with NVIDIA DGX and the WEKA platform.

We have organized this document to address critical items for enabling successful design, implementation, and transition to operation.

Purpose

This document covers the following subject areas:

- Overview of the WEKA solution.
- Overview of NVIDIA DGX.
- The benefits of NVIDIA DGX with WEKA.
- Recommendations for architecting a complete NVIDIA DGX solution on the WEKA platform.

Document Version History

Version Number	Published	Notes
1.0	September 2021	Original publication.
1.1	September 2022	Updated format and BasePOD
1.2	January 2024	Updated for DGX BasePOD with H100
1.3	March 2024	New perfomance data

2. WEKA Data Platform

The WEKA Data Platform is deployed on commercially available NVMe servers which were used for the DGX BasePOD validation testing. An entry-level WEKA cluster for DGX BasePOD requires eight servers for full availability with the ability to survive up to a two-node failure. Each server has a CPU, NVMe storage, and high-bandwidth networking. The exact configuration for the RA is detailed in the Technology Requirements section. The cluster can be easily scaled to hundreds of nodes.

Performance at Scale

At the core of the WEKA Data Platform is WekaFS™, the world's fastest and most scalable POSIX-compliant parallel file system. It is designed to transcend the limitations of legacy file systems that leverage local storage, NFS, or block storage, making it ideal for data-intensive AI and HPC workloads. WekaFS is a clean sheet design integrating NVMe-based flash storage for the performance tier to the GPU-accelerated servers, object storage, and ultra-low latency interconnect Infiniband or RoCE fabrics such with 100, 200, and 400GbE NICs , creating a highly performant scale-out storage system. WekaFS performance scales linearly as more servers are added to the storage cluster allowing the infrastructure to scale with the increasing demands of the business.

Multi-Protocol Ready

In addition to POSIX access, WEKA supports standard file access protocols, including NFS, SMB, and S3, for maximum compatibility and interoperability. Hadoop and Spark environments also benefit from the performance of a shared file system through a fully integrated connector that allows WEKA to replace HDFS and function as a single, easy-to-manage data lake for all forms of analytics.

Expandable Global Namespace over S3 Object Store

WEKA delivers best-of-breed performance from the NVMe flash tier, and the namespace can expand to any S3 object store, on-premises or in the cloud. This optional hybrid storage model with the ability to develop the global namespace to lower-cost hard disk drives in an object store delivers a cost-effective data lake without compromising performance. The integrated tiering to multiple S3 targets enables cost-effective data lifecycle management for older or less used training datasets.

Advanced Durability and Security

Large and agile datasets in AI/ML frequently require a data versioning capability. This is achieved using WEKA's instant and space-efficient snapshots capability for experiment reproducibility and explainability. The Snap-to-Object feature captures a point-in-time copy of the entire, unified (flash and object store) file namespace that can be presented as another file namespace instance in a private or public cloud. With crucial management integration, WEKA's integrated snapshots and end-to-end encryption features ensure that data is always backed up and secure throughout its lifecycle. WEKA also provides immutability and data mobility for these datasets with instant recovery. WEKA can seamlessly back up to multiple cloud targets providing backup, DR, and data governance capability.

Cloud Bursting and Data Mobility

In addition to providing versioning, WEKA's Snap-to-Object feature offers additional benefits beyond backup and DR to the public cloud; it enables secure data portability from on-premises to the public cloud for organizations that require access to on-demand GPU resources in the public cloud.

Container Support

Organizations are increasingly adopting containers deployed on Kubernetes (K8s) for AI workloads. Using the WEKA Kubernetes CSI plug-in, organizations now have flexibility in how and where they deploy containerized applications. It provides easy data mobility from on-premises to the cloud and back while delivering the best storage performance and latency. Figure 1 provides an overview of WEKA architecture in a typical production deployment.

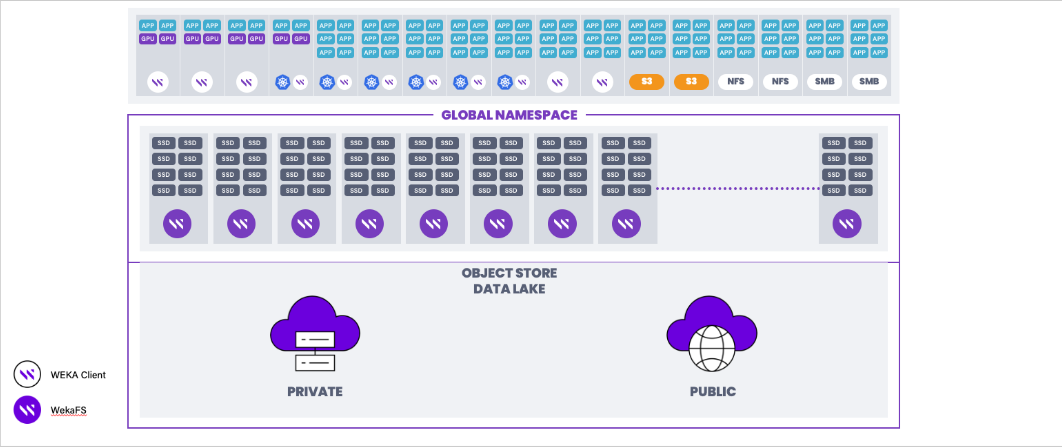

FIG. 1 WEKA Architecture

3. NVIDIA DGX Architecture

Figure 2 shows an exploded view of the significant components in the NVIDIA DGX H100 system.

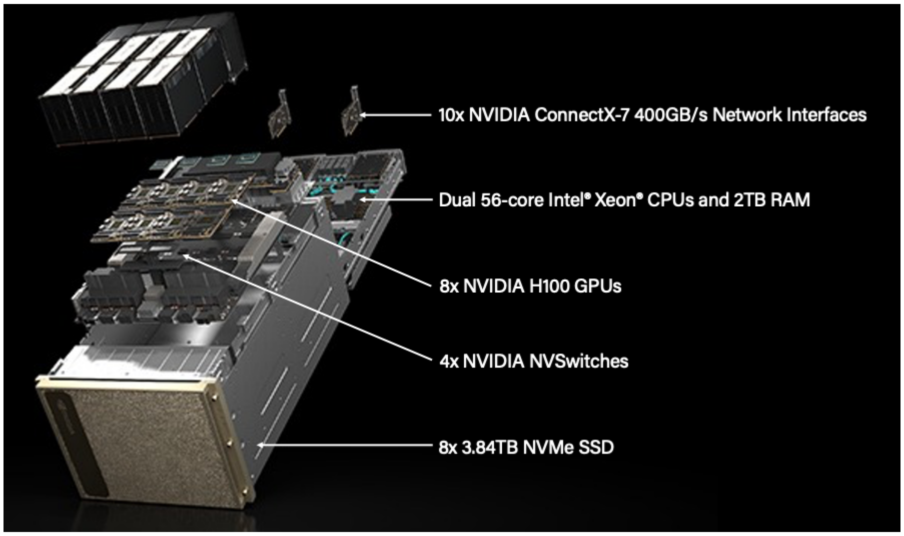

FIG. 2 DGX H100 components

NVIDIA H100 Tensor Core GPU

At the core, the NVIDIA DGX H100 system leverages the NVIDIA H100 Tensor Core GPU (Figure 3), designed to accelerate large and complex AI workloads and multiple small workloads, including enhancements and new features for increased performance over the NVIDIA H100 GPU. The H100 GPU in DGX H100 features 80 (GB) of high-bandwidth HBM3 memory and more extensive and faster caches and is designed to reduce AI and HPC software and programming complexity.

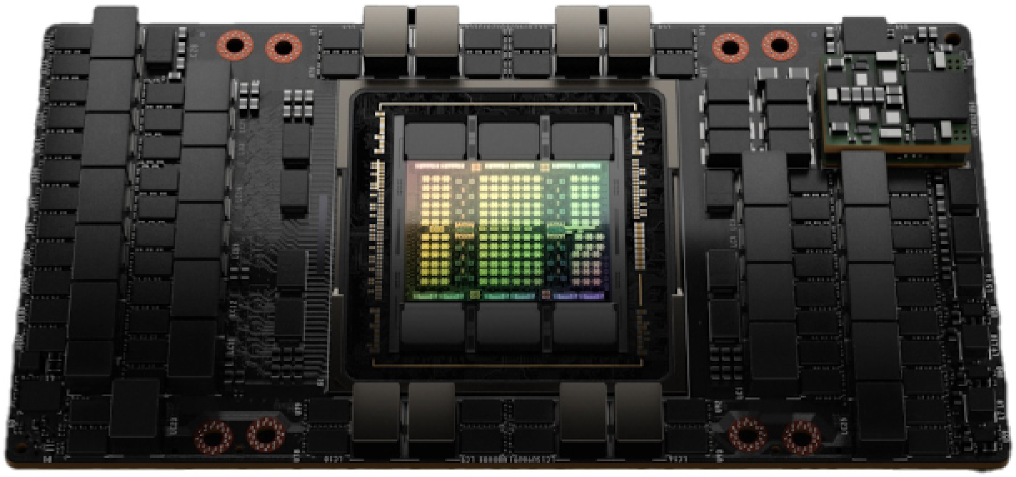

FIG. 3 NVIDIA H100 GPU

The NVIDIA H100 GPU includes new features to accelerate AI workload and HPC application performance further.

- Fourth-generation Tensor Cores
- New Transformer Engine
- Up to 7 Multi-Instance GPUs (MIG) instances

You can find additional product information online.

Multi-Instance GPU (MIG)

The NVIDIA H100 GPU incorporates a second-generation partitioning capability called Multi-Instance GPU (MIG) for increased GPU utilization. MIG uses spatial partitioning to carve the physical resources of a single H100 GPU into as many as seven independent GPU instances. With MIG, the NVIDIA H100 GPU can deliver the guaranteed quality of service with up to 3x higher compute capacity and 2x higher memory bandwidth than A100 with simultaneous instances per GPU.

On an NVIDIA H100 GPU with MIG enabled, parallel compute workloads can access isolated GPU memory and physical GPU resources as each GPU instance has its memory, cache, and streaming multiprocessor. This allows multiple users to share the same GPU and run all instances simultaneously, maximizing GPU efficiency.

MIG can be enabled selectively on any number of GPUs in the DGX H100 system - not all GPUs need to be MIG-enabled. However, if all GPUs in a DGX H100 system are MIG enabled, up to 56 users can simultaneously and independently take advantage of GPU acceleration.

Typical use cases that can benefit from MIG are

- Multiple inference jobs with batch sizes of one that involves small, low-latency models and that doesn't require all the performance of a full GPU
- Jupyter notebooks for model exploration
- Resource sharing of the GPU among multiple users

Fourth Generation NVLink and Third Generation NVSwitch

The DGX H100 system contains four third-generation NVIDIA® NVSwitch™ fabrics interconnecting the H100 GPUs using fouth-generation NVIDIA NVLink® high-speed interconnects. Each H100 GPU uses eighteen NVLink interconnects to communicate with all four NVSwitches, which means there are multiple links from each GPU to each switch. This provides maximum bandwidth to communicate across GPUs over the links.

The third-generation NVSwitch [Figure 4] is 50% faster than the previous version, which was introduced in the NVIDIA A100 system. The combination of four NVSwitches and fourth-generation NVLinks enables individual GPU to GPU communication to peak at 900 GB/s, which means that if all GPUs communicate with each other, the total amount of data transferred peaks at 7.2 TB/s for both directions.

FIG. 4 NVIDIA NVSwitches

NVIDIA ConnectX-7

Multi-system scaling of AI deep learning and HPC computational workloads requires strong communications between GPUs in multiple systems to match the significant GPU performance of each design. In addition to NVLink for high-speed communication internally between GPUs, the DGX H100 is purpose-built for multi-system AI scaling using NVIDIA ConnectX-7 400Gb/s NDR InfiniBand ports (also configurable as 400Gb/s Ethernet ports), providing 8 Tb/s of peak bi-directional bandwidth from a single system that can be used to immediately build a high-speed cluster of DGX H100 systems such as NVIDIA DGX SuperPOD.

The most common methods of moving data to and from the GPU involve leveraging the onboard storage and using the NVIDIA ConnectX-7 network adapters through Remote Direct Memory Access (RDMA). The DGX H100 incorporates additional switching between the IO cards and the GPUs, allowing each GPU to communicate with external sources without blocking other GPUs' access to the network.

The NVIDIA ConnectX-7 I/O cards offer flexible connectivity as they can be configured as NDR InfiniBand or 400Gb/s Ethernet. This allows the NVIDIA DGX H100 to be clustered with other nodes to run HPC and AI workloads using low latency, high bandwidth InfiniBand, or RDMA over Converged Ethernet (RoCE).

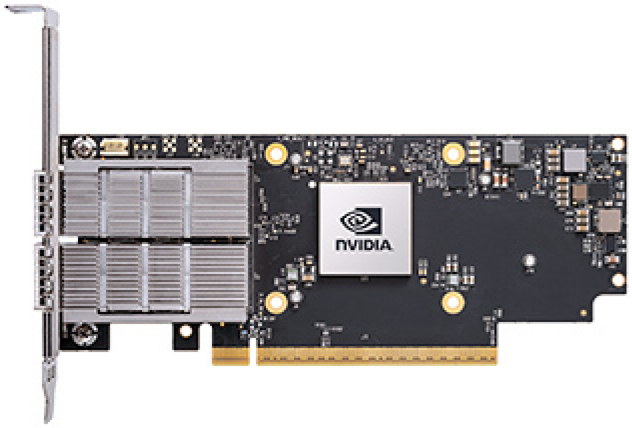

FIG. 5 NVIDIA Single-port ConnectX-7

NVIDIA Base Command

NVIDIA Base Command™ provides enterprise-grade orchestration and cluster management, libraries that accelerate compute, storage, and network infrastructure, and system software optimized for running AI workloads.

Base Command works with the NVIDIA AI Enterprise™ software suite, which is now included with every DGX system. The NVIDIA AI software enables end-to-end AI development and deployment with supported AI and data science tools, optimized frameworks, and pretrained models.

Base Command also offers enterprise-workflow management and MLOps integrations with DGX-Ready Software providers like Domino Data Lab, Run.ai, and Weights & Biases. It also includes libraries that optimize and accelerate compute, storage, and network infrastructure — while ensuring maximized system uptime, security, and reliability.

4. WEKA Solutions Built on DGX BasePOD

The WEKA Data Platform powered by NVIDIA DGX BasePOD will enable us to extend these benefits to our customers at a much larger scale. Key highlights of the solution include:

- The WEKA Data Platform and NVIDIA DGX BasePOD are now directly applicable to mission-critical enterprise AI workflows, including natural language processing and larger-scale workloads for customers in the life sciences, healthcare, and financial services industries, among many others. WEKA can efficiently serve large and small files across various workload types.
- WEKA's continued innovation and support of NVIDIA Magnum IO™ GPUDirect Storage technology provides low-latency, direct access between GPU memory and storage. This frees CPU cycles from servicing the I/O operations and delivers higher performance for other workloads.

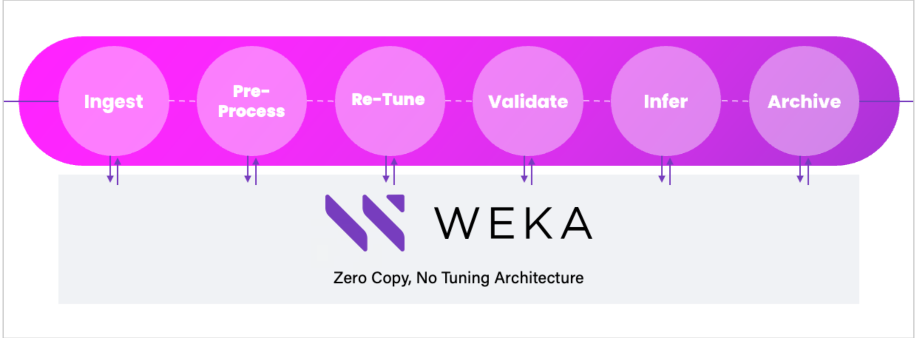

FIG. 6 WEKA Data Platform for AI Pipelines

5. Solution Design

Design Decisions

The following tables cover design decisions and rationale for NVIDIA DGX BasePOD with WEKA.

General Design Decisions: WEKA Servers

Item	Details	Rationale
Minimum Size	8-Node WEKA Cluster	Minimum size requirements
Scale approach	Incremental, modular scale	Allow for growth from Proof-of-Concept to massive scale
Scale Units	Servers	Granular scale to precisely meet the capacity demands

General Design Decisions: NVIDIA DGX Systems

Item	Details	Rationale
Minimum Size	1 NVIDIA DGX System	Minimum size requirements
Scale approach	Incremental, modular scale	Allow for growth from Proof-of-Concept to massive scale
Scale Units	Servers	Granular scale to precisely meet the capacity demands

General Design Decisions: Networking

Item	Details	Rationale
Compute/DGX	InfiniBand is recommended	The GPU-to-GPU network needs to be fast with optimized offloading
Storage	400GbE or InfiniBand	WEKA recommends 400GB to match to the ConnectX-7 speeds in the DGX H100 System

NVIDIA DGX BasePOD Sizing

AI is powering mission-critical use cases in every industry—from healthcare to manufacturing to financial services. The NVIDIA DGX BasePOD™ reference architecture provides the critical foundation on which business transformation is realized and AI applications are born.

NVIDIA DGX BasePOD is an all-inclusive system that includes the following components:

- NVIDIA DGX H100 systems
- Networking for the DGX compute layer and the storage layer
- Accelerated Storage Fabric

NOTE: *It's always a good practice to add a buffer for contingency and growth.*

Network

The solution tested for this RA consisted of NVIDIA DGX H100 systems connected to two NVIDIA QM9700 Ethernet switches with two 400 Gb/s network connections per DGX H100 system. Each NIC was connected to a separate NVIDIA QM9700 switch for eight network links to the storage system. Each DGX H100 system also had eight single-port NVIDIA ConnectX-7 400Gb/s NDR InfiniBand ports (also configurable as 400Gb/s Ethernet ports) for inter-GPU system communication.

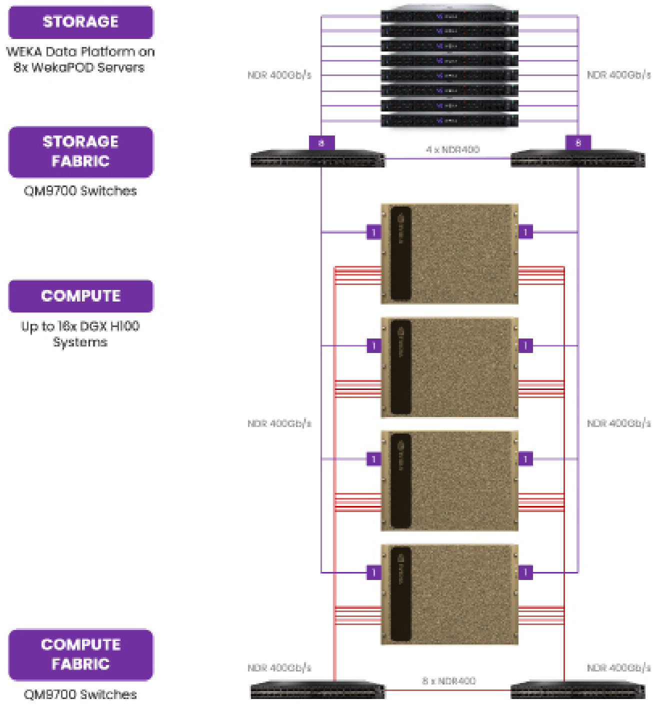

FIG. 7 Network Architecture

The WEKA servers were each configured with two ConnectX-7 400Gb/s NICs. Each NIC was connected to a separate NVIDIA QM9700 switch for 16 400Gb/s network links.

The network switch was configured with a message transmission unit (MTU) size of 9000.

WEKA does not require RoCE to be configured on any portion of the design and does not require priority flow control (PFC) to be configured on the network switch, greatly simplifying the network deployment.

6. Validation

This reference architecture has undergone NVIDIA DGX BasePOD certification, which requires the completion of a suite of tests that measure performance using an automated IO test suite.

Hardware Configuration

Hardware	Quantity
DGX H100 systems	1-16
WEKA Servers	8, includes per server AMD EPYC™ 9454P 48 core CPU, 384GB of DDR5 RAM, 14x gen5 NVMe SSDs per server, 2 x 400Gb/s NVIDIA ConnectX-7 NICss
QM9700C switch (storage fabric)	2
QM9700 switch (compute fabric)	2

Software Configuration

Software	Version
WekaFS	4.2.6
Server OS	CentOS 8.6
DGX OS	4.99.9
QM9700 switch firmware	3.11.2016
DGX OS	6.1.0

Software	Version
OFED version	5.8-3.0.7.0
Base Command	Version 10.0

7. Performance

Performance validation of the DGX BasePOD environment consisted of an 8-node WEKA cluster configured as shown in figure 7, above. One change compared to the figure 7 diagram was connection to 16 DGX H100 systems instead of the four shown. FIO was used to generate IO across multiple parameters for read and write for this system with no tuning across any of the runs.

IO Profile	Cluster Results	Perfromance per Node
1MB sequential reads, 200 job count	765 GB/s	95.5 GB/s
1MB sequential writes, 200 job count	186 GB/s	23.25 GB/s
4KB random read latency, 1 job count	162us	n/a
4KB random write latency, 1 job count	116us	n.a
4KB random read IOPs, 200 job count	18.25M IOPs	2.28M IOPs
4KB random write IOPs, 200 job count	4.3M IOPs	535K IOPs

8. Conclusion (Getting Started)

nal compute infrastructures are unsuitable for demanding AI workloads due to slow legacy CPU architectures and varying system requirements. This drives up complexity, increases cost, and limits scale. Engineers at WEKA and NVIDIA collaborated to architect a scalable and robust infrastructure that pushes the boundaries of AI innovation and performance.

The validation showed robust linear performance scalability from one to sixteen DGX H100 systems, allowing organizations to start small and grow seamlessly as AI projects ramp. The results demonstrate that scaling GPU infrastructure to accelerate time to insights will be well supported by WEKA. The validated WEKA configuration makes it easy for teams to focus on developing new products and gain new faster insights with AI/ML.

9. Appendix

References

Product & Technology	
Weka File System	
WEKA Data Platform	ttps://www.weka.io/data-platform/how-our-data-management-platform-works/
WEKA Architecture White Paper	https://www.weka.io/resources/white-paper/wekaio-architectural-whitepaper/
NVIDIA DGX H100 System	
NVIDIA DGX H100 System	https://www.nvidia.com/en-us/data-center/dgx-h100/
NVIDIA H100 Tensor Core GPU	https://www.nvidia.com/en-us/data-center/h100/
NVIDIA Networking	
NVIDIA Quantum QM9700	https://nvdam.widen.net/s/k8sqcr6gzb/infiniband-quantum-2-qm9700-series-datasheet-us-nvidia-1751454-r8-web
Machine Learning Frameworks	
TensorFlow	https://www.tensorflow.org/
Horovod	https://eng.uber.com/horovod/

10. About WEKA

WEKA offers a modern subscription software-based data platform delivering the massive performance and scale demanded by today's cloud and AI workloads. With the simplicity of NAS, the performance of a parallel filesystem, and the scale of object storage, WEKA eliminates any compromises between Simplicity, Speed, and Scale. Learn more at www.weka.io or follow us on X/Twitter @WekaIO.

WKA???-01 03/2022