

WEKApod™

Data Platform Appliance

Product Overview

- Certified for NVIDIA DGX SuperPOD™ on NVIDIA DGX H100 Systems
- Powered by the WEKA Data Platform®
- NVMe and NVMe-oF technology (POSIX, RDMA, and TCP)
- Magnum IO™ GPUDirect Storage (GDS)
- NFS, S3, & SMB protocol support

WEKApod™ Performance

THROUGHPUT

- 720GB/s sustained Read Bandwidth
- 186GB/s sustained Write Bandwidth

IOPS

- 18.3 Million 4kB Random Read IOPS
- 4.3 Million 4kB Random Write IOPS

LATENCY

- 162µs 4kB Read Latency
- 116µs 4kB Write Latency

Simplicity

- REST API for orchestration
- Non-disruptive upgrades (NDU) and capacity expansions
- Native cloud burst to all major hyperscalers (AWS, Azure, GCP, OCI)
- NVIDIA Base Command Manager Integration

Scalability

- Scale to 100's of servers
- Linear TB, BW, IOPS at-scale
- Trillions of inodes per deployment
- Billions of inodes per directory tree

Sustainability

- 10-50X better stack efficiency
- 4-7X lower data center footprint
- 260 tons of CO2e saved per PB annually

The WEKA Data Platform™ Foundation

WEKA was founded on the idea that current storage solutions have only provided incremental improvements to legacy designs, allowing for a widening gap between compute performance and data storage performance. Storage remains a leading bottleneck to application performance, and with the continued densification of compute in areas such as GPU-based applications, has become even more problematic. In today's hyper-competitive market, organizations need flexible infrastructure; application workloads are becoming increasingly complex and data sets are continuing to grow unchecked—all forcing enterprises to architect overly complicated and costly systems that reduce IT agility.

WEKA's unique architecture is radically different from legacy storage systems, appliances, and hypervisor-based software-defined storage solutions because it not only overcomes traditional storage scaling and file sharing limitations but also allows parallel file access via POSIX, NFS, SMB, S3, and Magnum IO™ GPUDirect Storage. Our patented WekaFS™ filesystem lays out the data in a manner that allows for application-level latency that rivals SAN, however, distributed in a way that provides object store levels of scalability. The WEKA Data Platform provides a rich enterprise feature set, including local snapshots, automated tiering, dynamic cluster rebalancing, private cloud multi-tenancy, backup, encryption, authentication, key management, user groups, quotas, and much more. WEKA also enables a wide variety of hybrid cloud workflows including bursting to the cloud, running workflows that span locations, as well as using the cloud to protect and archive data.

Fully Integrated, GPU-Optimized Data Platform for AI

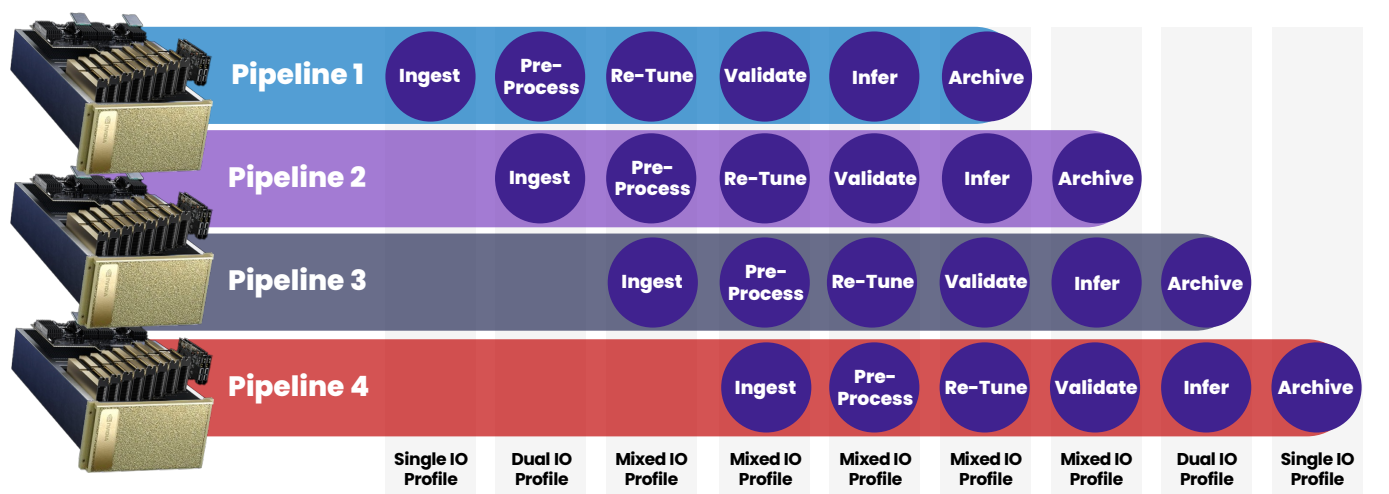
Get all the benefits of the WEKA Data Platform in a validated, turnkey data platform appliance certified for NVIDIA DGX SuperPOD™ on NVIDIA DGX H100 Systems. Starting with a minimum of 8 nodes, WEKApod scales to hundreds of nodes and will seamlessly integrate with NVIDIA Base Command™ Manager for centralized observability, including WEKA Data Platform performance metrics, and seamlessly integrates with Run.ai and other orchestration tools.

Feeding GPUs from a Single Namespace








WEKA collapses the typical GPU-starving “multi-hop” AI data pipeline by using a single namespace where your entire data set is stored. This zero-copy architecture eliminates the multiple steps needed to stage data before training. Your GPUs gain fast access to data needed for training, while WEKA automatically manages tiering of data between high-performance, NVMe-based storage, and low-cost object storage. Incorporating the WEKA Data Platform for AI into deep learning data pipelines increases the utilization rates for NVIDIA GPU systems. It eliminates wasteful data copying and transfer times between storage silos to significantly increase the number of training data sets that can be analyzed per day.

Deliver Multi-Dimensional AI Performance

When AI data pipelines are overlapped, a storage system no longer has to deal with just changing IO for every step in a pipeline, but now has to handle the mixed IO from different stages of every pipeline and regular checkpointing to save the state of the pipeline. Kicking off parallel training and/or re-tuning jobs at different times creates a sort of IO blender. The parallelism of jobs blurs the IO patterns to the point where the storage system is dealing with a mixed IO profile that tends to be random in nature.



The WEKA Data Platform™
Zero Copy, Zero Tuning Architecture

- 
 Distributed Data Protection
- 
 Intelligent Fast Rebuild
- 
 Instant Snapshots & Snap Clones
- 
 Snap & Tier to S3 Object Store
- 
 Multiprotocol NFS, SMB, S3, POSIX
- 
 Backup / DR and Cloud Burst
- 
 At-Rest & In-Flight Encryption

WEKApod supports small and large files simultaneously, with both mixed random and sequential I/O patterns while delivering application level 4kB I/O at sub-200 μsecond latency, and 10’s of millions of IOPS. The starting WEKApod configuration delivers up to 720GB/s of sustained read bandwidth, 186GB/s of sustained write bandwidth, and up to 18.3 million random 4kB IOPS. And because WEKApod is powered by the WEKA Data Platform, zero tuning is required for delivering optimal I/O to your applications regardless of I/O size, inode count, file size, etc.

Deliver Sustainable AI

Data centers consume more than 3% of global energy consumption which is projected to rise to 8% by 2030 if left unchecked. Legacy data architectures have a greater environmental impact than contemporary, modern approaches, which can negate sustainability efforts. WEKApod on the WEKA® Data Platform drives 10x-50x better AI/ML stack efficiency reducing annual GPU operating energy. WEKA also lowers the data infrastructure footprint by 4x-7x through data copy reduction and cloud elasticity.

Annual carbon emissions (metric tons CO ₂ e)	Legacy Infrastructure	WEKA	Savings
Hardware CO ₂ footprint (per PB)	152 tons CO ₂ e	30 tons CO ₂ e	122 tons CO ₂ e
Operating CO ₂ footprint (per PB)	164 tons CO ₂ e	26 tons CO ₂ e	138 tons CO ₂ e
Total CO ₂ e (per PB)	316 tons CO ₂ e	56 tons CO ₂ e	260 tons CO ₂ e
Average Customer CO ₂ e (7PB)	2,212 tons CO ₂ e	392 tons CO ₂ e	1,820 tons CO ₂ e

KEY INSIGHT

Siloed applications, extensive data movement, and the need to oversize an environment to meet performance goals all lead to greater energy consumption and as a result, more carbon emissions.

BENEFITS

- Reduced energy consumption while also delivering faster results
- Over 260 tons of CO₂e per petabyte saved compared to a traditional data architecture



WEKApod™ Data Platform Specifications:



Node Type	WEKApod™	
Minimum Node Qty	8	
Maximum Node Qty	100's	
CPUs per Node	1x AMD EPYC™ 9454 Processor 48-core 2.75GHz 256MB Cache (290W)	
Drives per Node	14 x NVMe Read Intensive AG Drive E3s Gen5 with carrier	
SSD Drive Density	7.68 TB	15.36 TB
Total Usable Capacity	484 TB	968 TB
Networking per Node	<ul style="list-style-type: none"> • 2x NVIDIA® ConnectX®-7 400Gb/s NDR IB Single-Port OSFP, PCIe 5.0 x16 • 1x NVIDIA® ConnectX®-6 Lx Dual Port 10/25GbE SFP28, No Crypto, OCP NIC 3.0 	
Software	WekaFS	
Data Protection	<ul style="list-style-type: none"> • Distributed Data Protection (N+2 or N+4) • Drive Hot Sparing • Error Detection: End-to-end Data Protection • In-Flight and At-Rest Data Encryption 	
Protocols	POSIX, NFS, SMB, S3, GPUDirect Storage (GDS)	
Snapshots and Clones	File System Level, Up to 24,576 snapshots	
System Monitoring	Cloud-based Monitoring and Analytics for Application Tuning and Remote Support	
System Management	IP-based Out-of-Band Management Controller	
Minimum 8-Node WEKApod Performance	<ul style="list-style-type: none"> • Sequential read performance up to 720GB/sec • Sequential write performance of up to 186GB/s • Random 4kB read performance up to 18,250,000 IOPS • Random 4kB write performance up to 4,280,000 IOPS • 162µs 4kB read latency • 116µs 4kB write latency 	
8-Node Max Power	7,712 Watts, 30,027.2 BTU/Hour	
Node Dimensions	1 Rack Unit, Max Depth: 787 mm (30.99 in.)	



Expansion Option	WEKApod™ 4-Node Bundle		
Drives per Node	14 x NVMe Read Intensive AG Drive E3s Gen5 with carrier		
SSD Drive Density	7.68 TB	15.36 TB	30.72 TB
Total Usable Capacity	276 TB	552 TB	1,106 TB
4-Node WEKApod™ Performance Expansion	<ul style="list-style-type: none"> • Sequential read performance up to 360GB/sec • Sequential write performance of up to 93.2GB/s • Random 4kB read performance up to 9,120,000 IOPS • Random 4kB write performance up to 2,140,000 IOPS 		
4-Node Max Power	3,856 Watts, 15,013.6 BTU/Hour		
Expansion Dimensions	4 Rack Units, Max Depth: 787 mm (30.99 in.)		



weka.io

844.392.0665

