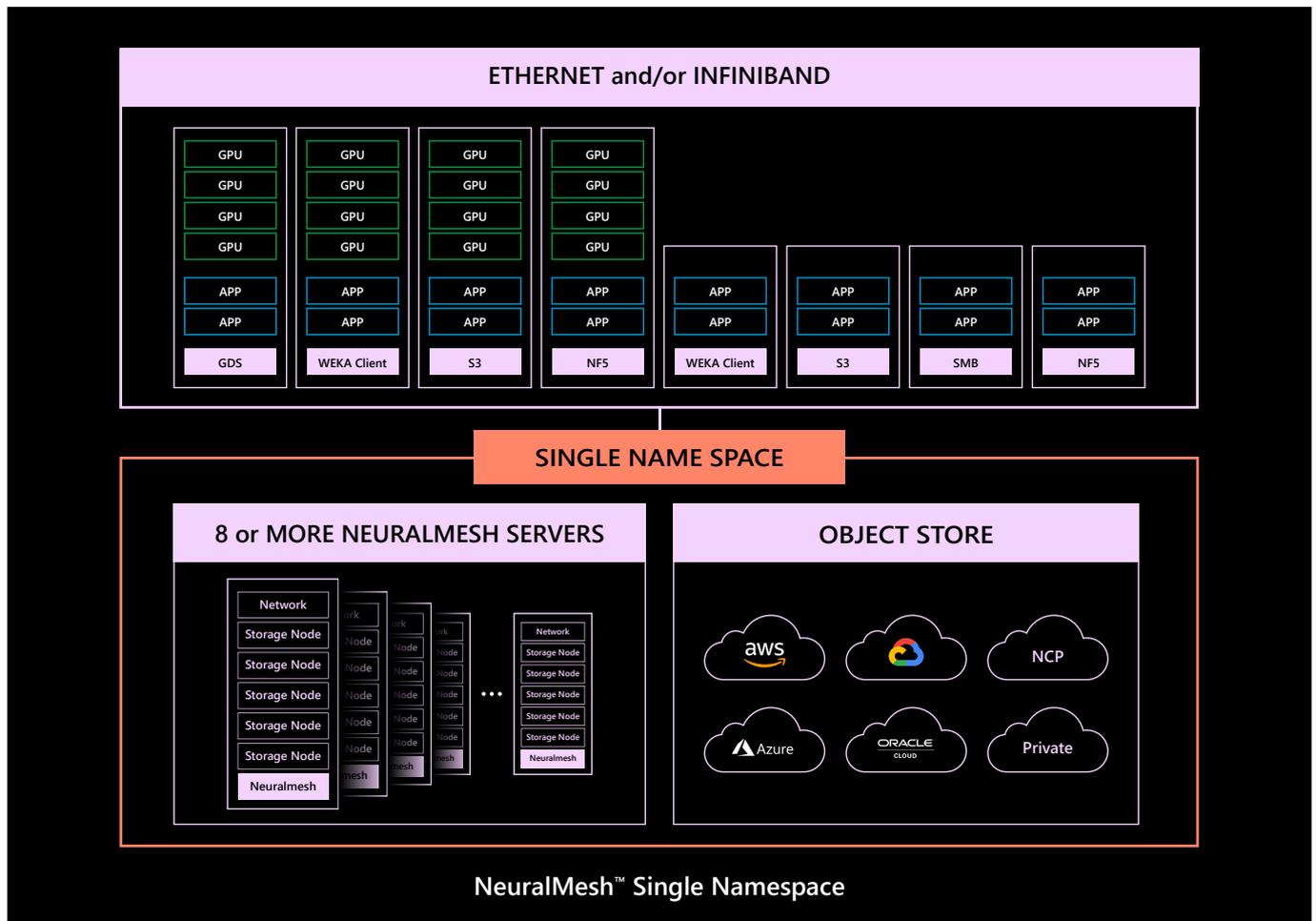


# WEKA Storage Solution with GIGABYTE

The High-Performance Data Platform for the AI Era



## Maximize GPU Cluster Throughput and Efficiency

WEKA's high-performance data platform delivers fast, unified, and scalable data access for AI and other compute-intensive workloads. Combined with GIGABYTE's high-throughput NVMe servers and automated management software, the joint solution ensures that GPUs receive data quickly and consistently — so they operate at full efficiency — accelerating every workflow from model training to inference, MLOps to data pools, all while delivering optimized TCO.

## Why Choose WEKA High-Performance Data Platform?

### Extreme Performance

Distributed POSIX-compliant file system built for AI and HPC. Delivers consistently low latency for both small random I/O and large sequential operations — keeping GPUs always fed.

### Unified, Cloud-Native Data Platform

One platform, one namespace, zero friction. WEKA delivers seamless data flow and consistent access control across on-prem, cloud, and hybrid environments — with native support for high-performance file I/O and S3-compatible object storage tiering built in.

### Elastic Scalability

Expand capacity and throughput linearly by simply adding nodes — zero downtime, zero data migration. WEKA NeuralMesh™ auto-optimizes and self-heals so your AI pipeline never misses a beat.

# Technical Specifications



## Protocol Support

POSIX (native), NFS/SMB (via protocol services), S3 (compatible)



## Data Protection

Snapshots, replication, quotas, and multi-tenancy isolation



## Data Tiering

Hot tier (NVMe) ↔ Warm/Cold tiers (object storage) with automated tiering



## Zero-Copy Path

GPUDirect Storage for accelerated GPU data access



## Data Consistency

Strong consistency with real-time metadata indexing



## Observability

Real-time telemetry and APIs integrated with GPM and third-party monitoring

## Large-Scale AI Model Training

WEKA's ultra-low-latency NVMe tier with GIGABYTE Gen5 servers delivers continuous high-bandwidth data, reducing training time and maximizing GPU utilization.

## Real-Time Inference & MLOps Pipelines

NeuralMesh™ auto-balances load and self-heals, ensuring inference pipelines maintain SLA-level responsiveness 24/7.

## Research Institutes

Provides seamless data mobility across on-premises and cloud via a unified namespace, enabling collaboration and faster research cycles.

## Semiconductor Industry

Optimized for EDA workloads, supports massive concurrent file operations with low latency, improving simulation efficiency and yield.

## WEKA Approved Storage Servers



Rack Server

### R263-Z36-AAV1

AMD EPYC™ 9005/9004 Series Processors  
2U UP 24-Bay Gen5 NVMe Titanium

[Learn More](#)



Rack Server

### R163-Z35-AAH1

AMD EPYC™ 9005/9004 Series Processors  
1U UP 12-Bay Gen5 NVMe/SATA/SAS-4 Titanium

[Learn More](#)



## GET A QUOTE!

For more details, please contact the Giga Computing Sales Department or visit our website:  
[www.gigabyte.com/solutions](http://www.gigabyte.com/solutions)

Designed by

