



Break the AI storage bottleneck

HPE Solutions with WEKA is the **fastest** file storage for AI¹

70%

of a GPU's AI training time is spent waiting for data^{2,3}

AI demands more from data storage

Artificial intelligence (AI) has reshaped how industries use data, driving the need for higher throughput, lower latency, and massive scalability in data workflows. With the rise of large language models (LLMs) and generative AI (GenAI), the ability to handle vast datasets and complex models at the highest speeds defines competitive advantage.

Data-intensive applications in financial analytics, genomics, life sciences, and high-performance computing (HPC) are pushing infrastructure to its limits. Traditional storage architecture is now the weakest link, unable to keep up with rapidly escalating performance demands while scaling across core, multi-cloud and edge data centers.

The storage bottleneck starving GPUs of data

Optimizing compute and networking alone isn't enough—GPUs are only as fast as the data feeding them. Conventional storage struggles with the sheer volume of unstructured data, causing slow read/write speeds from multiple stages of data copying, file staging, and metadata processing.

The result? What should be fast flowing data piles up. GPUs sit idle, waiting for checkpoints to finish, costing time and resources. To unlock full GPU performance, the storage bottleneck must be eliminated.

¹SPECstorage® Solution 2020_Results. SPEC Standard Performance Evaluation Corporation. (Jan 2025).

²Murray, D. G., Simsa, J., Klimovic, A. & Indyk, I. (2021). *tf data: A Machine Learning Data Processing Framework*. Google™, Microsoft.

³Mohan, M., Phanishayee, A., Raniwala, A. & Chidambaram V. (2021). *Analyzing and Mitigating Data Stalls in DNN Training*. Microsoft, VMware®.

HPE Solutions with WEKA

An end-to-end solution for your AI pipeline

Traditional storage architectures are not designed for the extreme I/O demands, massive concurrency, and dynamic data patterns current AI/ML demands. Modern AI workloads require intelligent storage that manages the entire data lifecycle—from ingestion and preprocessing, to training, inference, and archival. **HPE Solutions with WEKA** is designed to meet these demands with a zero-copy, zero-tuning, parallel file system that ensures seamless, high-speed data flow throughout the AI pipeline.

Whether ingesting data at the edge, preprocessing in the core, or training large-scale models in the cloud, get consistent, high-throughput, IOPS, and ultra low-latency performance that surpasses NVIDIA® Cloud Partner (NCP) certification standards. Reduce prefill cycles and drive faster, accurate AI reasoning with lower cost of innovation.

Features & benefits



Unlock maximum GPU speed with AI-era storage

Process billions of files at microsecond latency across environments, regardless of data type. Zero-copy, parallel architecture enables GPUs to read and write data directly from storage, eliminating duplication and silo bottlenecks.

- Achieve 1TB/s read bandwidth with <1 rack—triple the IOPS compared to other solutions occupying 9x more space.
- Get up to 400 Gbps Ethernet or InfiniBand network cards and full NVIDIA Magnum IO Software Suite support.



Gain transformative scale wherever your data lives

Support up to 1 EB of flash across on-prem, cloud, and hybrid environments. Extend file system namespace into the multi-exabytes with cost-effective object storage tiers, such as HPE Solutions with Scalify or any S3-compatible platform.

- Manage multiple customers, teams, or projects at the desired privacy, security, and performance levels with composable clusters on shared hardware.
- Scale as priorities evolve without impacting the entire cluster, optimizing costs and management overhead.
- Native simultaneous, multi-protocol support: NVIDIA GDS (GPUDirect Storage), POSIX, S3, NFS, SMB, and HDFS.



Simplify with fast file storage built for enterprise

Deploy fast storage without the complexity. Auto-optimize performance for any workload with zero-tuning intelligence. Just set it and forget it—no storage specialist required.

- Effortlessly configure, monitor, and manage clusters in real time with an intuitive, built-in console.
- Get consistent data availability with a single, global namespace that supports billions of files.
- Automated hot/cold data tiering allows GPUs to gain rapid access to training data, maintaining top performance.



Protect your AI investments

Get enterprise-grade resiliency and security. High-load servers with built-in redundancy, failovers, and silicon root of trust, ensure uptime and firmware integrity. Advanced behavioral analytics and system monitoring strengthen holistic protection.

- Protect against rogue actors with end-to-end encryption (at-rest and in-transit), authentication, Snap-to-object (S2O), and automatic cloud backups.
- Extend the life of hardware investments with non-disruptive upgrades and mixed-generation node compatibility.
- Pay only for the storage you use with the HPE GreenLake Flex consume-as-a-service model.

Real-world results

- Faster GPU-based AI learning model solutions with greater accuracy
- Geometric increase in number of training data sets that can be analyzed per day
- Increased time-to-value for GPU-driven life sciences, AI modeling, and financial applications
- More efficient GPU usage, lowered cost of innovation, and faster project initiation





Discover the right infrastructure for your AI workloads

HPE Solutions with WEKA is a key product in HPE's comprehensive portfolio of unstructured data and AI storage solutions. Choose from optimized HPE platforms purpose-built for WEKA with highest performance and data availability:

HPE ProLiant DL325 Gen11 for WEKA

An NVIDIA Cloud Partner-ready, 1U 1P AMD-based, storage server. Designed for limited spaces that need capacity-dense storage and high performance for data-intensive workloads.



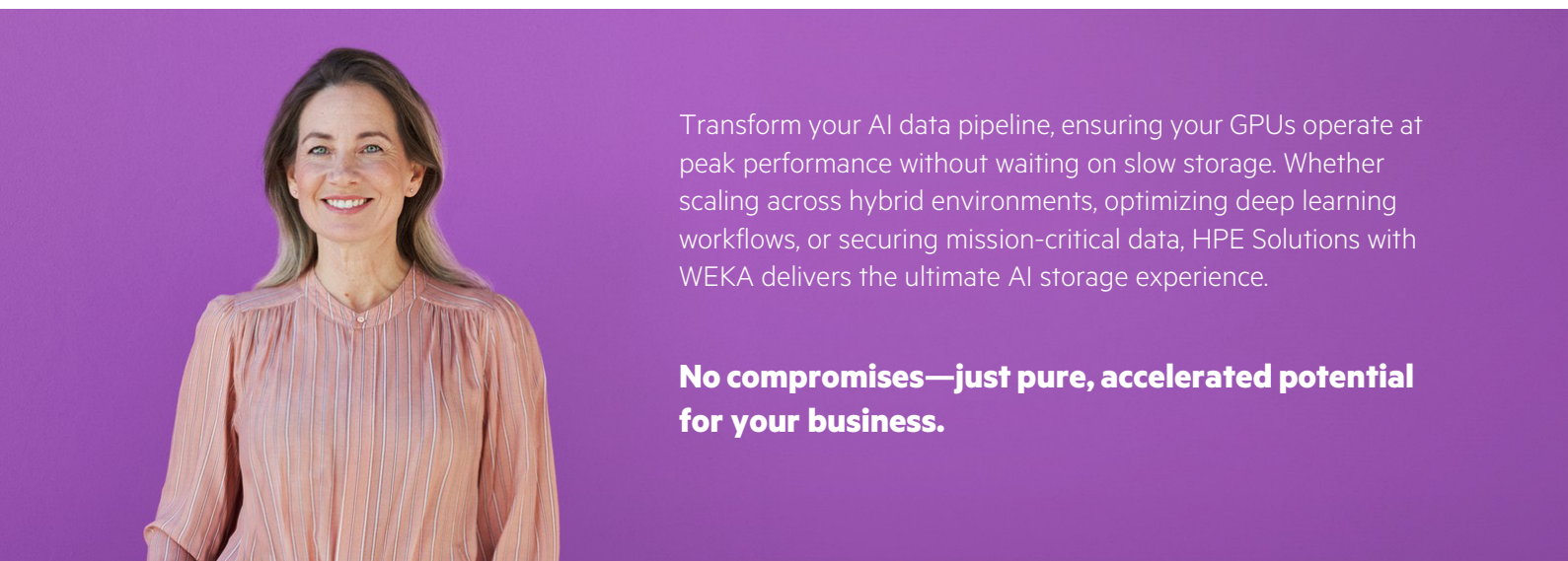
New! HPE Alletra Storage Server 4210 for WEKA

A powerful Intel®-based server ideal for data storage-intensive applications. The successor to the record-breaking Alletra Storage Server 4110, upgraded with faster CPUs, memory, and support for advanced air and direct liquid cooling.



4 new world records

HPE Alletra Storage Server 4110 set four new world records in the SPECStorage 2020 benchmarks, delivering industry-leading speed with NeuralMesh™ by WEKA¹

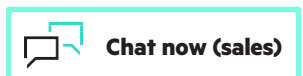


Transform your AI data pipeline, ensuring your GPUs operate at peak performance without waiting on slow storage. Whether scaling across hybrid environments, optimizing deep learning workflows, or securing mission-critical data, HPE Solutions with WEKA delivers the ultimate AI storage experience.

No compromises—just pure, accelerated potential for your business.

Learn more at

[HPE Solutions with WEKA](#)



© Copyright 2025 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

AMD is a trademark of Advanced Micro Devices, Inc. Google is a registered trademark of Google LLC. Intel is a trademark of Intel Corporation or its subsidiaries in the U.S. and/or other countries. Microsoft is either a registered trademark or trademark of Microsoft Corporation in the United States and/or other countries. GPUDirect and NVIDIA are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All third-party marks are property of their respective owners.

a00142745ENW