



WekaFS™ on AWS

Faster Deep Learning for AI and Analytics



EPIC PERFORMANCE AND LIMITLESS SCALE
 10 GB/sec to a single P3dn.24 GPU Client ensures that applications never have to wait for data



LINEAR PERFORMANCE SCALING ACROSS THE CLUSTER
 Performance scales linearly as more storage nodes are added to the cluster



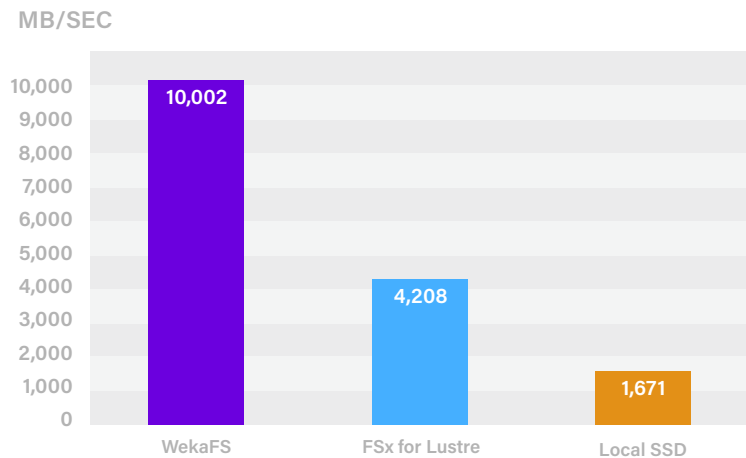
GLOBAL NAMESPACE
 Single namespace between flash and AWS S3 unifies all your data



BEST PERFORMANCE AND ECONOMICS
 Combine flash performance and S3 storage for best performance and cost

PREVENT GPU STARVATION AND GET TO THE ANSWERS FASTER

Modern analytics platforms are GPU-intensive and require large data sets to deliver the highest levels of accuracy to the training or analytics models. The workloads demand a high-bandwidth, low-latency storage infrastructure to ensure that the GPU cluster is fully saturated with as much data as the application needs. Typical data sets can span from terabytes to tens of petabytes, and the data access pattern for each Epoch is unique and unpredictable. This calls for a data infrastructure that can instantaneously and consistently feed large amounts of random data to multiple GPU nodes in real time, all emanating from a single shared data pool. WekaFS is a modern file system that is uniquely built to meet the performance and scalability needs of data-intensive applications leveraging P3 GPU instances in the AWS cloud. WekaFS on AWS was ranked #1 for supercomputer storage as measured by the IO-500 benchmark. WekaFS has proven scalable performance of over 10 GB/sec bandwidth to a single P3dn.24 GPU instance, delivering 2.4x the performance of FSx for Lustre and 6x the performance of local NVMe SSDs across a shared training set.



Read Performance to a Single P3dn.24 GPU Client

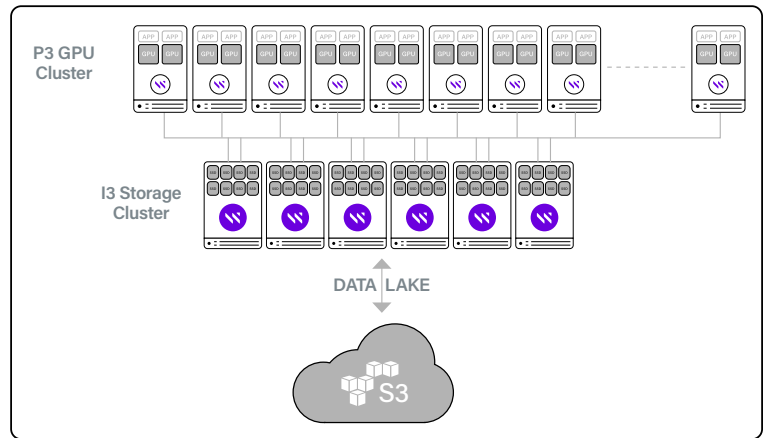
GET PREDICTABLE LOWEST LATENCY PERFORMANCE

Managing large amounts of data is challenging when the AI training system spans multiple GPU nodes. WekaFS runs on I3en EC2 instances, presenting a shared POSIX file system to the GPU servers and delivering sufficient performance to keep data-intensive applications compute-bound. WekaFS is a fully-parallel file system that distributes both data and metadata across the entire storage infrastructure to ensure massively parallel access. The software has an optimized network stack that leverages AWS networking and delivers very low-latency and high-bandwidth performance, resulting in a solution that can handle the most demanding data and metadata-intensive operations.

DEPLOY THE WEKA FILE SYSTEM ON DEMAND

WekaFS runs natively in AWS and can be deployed as a high-performance file system on any I3en instance. WekaFS leverages the performance of NVMe flash technology to deliver the highest performance and lowest latency for highly random data access patterns to both small and large files. An integrated tiering layer means that most data can be stored on lower cost AWS S3 infrastructure while data is still delivered with the highest performance from NVMe-based I3 instances to the GPU Cluster.

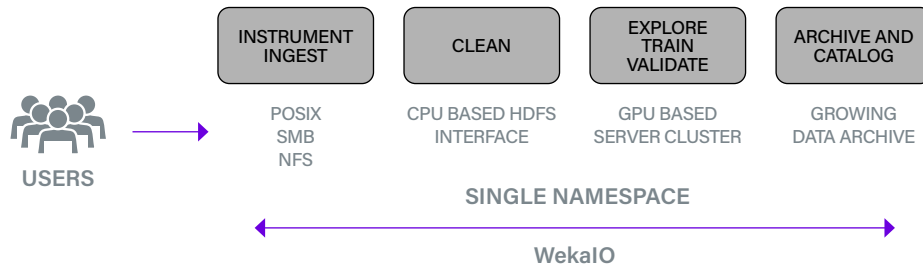
WekaFS offers full snapshot-to-S3 capability so that the file system does not have to remain on when the GPUs are idle. Simply snapshot the file system to an S3 bucket and shut down the I3en instances to save on computing cost. The file system can be re-hydrated at a later time when new GPU workloads are scheduled.



Weka Delivers Full Bandwidth from S3 Data Lake to the P3 GPU Instances

SINGLE SOLUTION FOR END-TO-END AI DATA WORKFLOW

WekaFS is the only solution that provides end-to-end data management for data-intensive AI and analytics workloads. A single global namespace supports and spans high-performance flash for ingest and inference, scaling to petabyte capacities, and AWS S3 for long term data growth and preservation. There is no need to copy or move data from one service to another as WekaFS maintains a single copy of data across the entire storage infrastructure and throughout the data lifecycle. Administrators and users have instant access to, and complete visibility of, the corporate-wide data set under management. A patented data protection scheme distributes data across the entire file system, and overall system reliability increases as the system scales.



WekaIO Manages All Stages of the AI Training Cycle in a Single Namespace

To find out more or to arrange for a free trial, go to <https://www.weka.io/get-started> or contact us at info@weka.io. For details on the IO-500 benchmark results, go to <https://www.vi4io.org/io500/start>.

