

HPC Data Lake Platform from Lenovo

WekaFS and Cloudian HyperStore for AI, ML, and Advanced Analytics

The Challenge

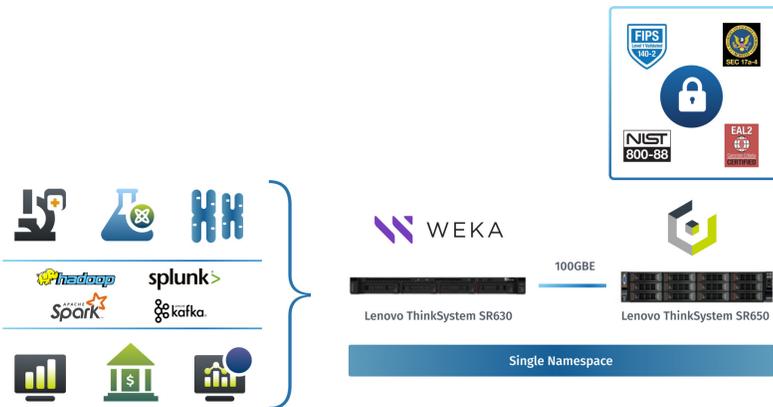
Organizations are consuming and creating more data than ever before and many are applying AI/ML on ever larger data sets, to make better decisions in near/real-time and unlock new revenue streams. Traditional storage systems simply can't handle the processing needs or the scalability required for iterative analytic workloads and introduce bottlenecks to productivity and data-driven decision making. The ability to unify data silos to accommodate diverse workloads along with high performance compute speed is crucial for today's modern organizations.

The Solution

The Lenovo ThinkSystem series running Cloudian HyperStore and WekaFS offers a state-of-the-art optimized storage solution for high-performance and scale-out workloads.

Deployed on Lenovo ThinkSystem, Weka file system (WekaFS) and Cloudian HyperStore™ provide an integrated storage solution allowing you to overcome the challenges associated with accelerating and scaling your data pipeline, while lowering overall storage costs associated with data analytics. WekaFS is a distributed, scale-out and POSIX-compliant file system, built on a modern architecture using NVMe Flash and supports Ethernet or Infiniband transport, with low latency, multi-protocol (POSIX, NFS, SMB, S3, GPUDirect Storage, CSI) high performance access. WekaFS distributes metadata throughout the cluster via patented mechanisms that prevent hot spots to maximize performance levels. Performance is predictable, consistent and scales linearly as more nodes are added to the storage cluster.

Weka is industry-leading storage systems for performance workloads on the Lenovo SR630 running the Weka file system (WekaFS). An entry-level cluster size for Weka requires eight server nodes, which are then presented as a single storage cluster.



Cloudian HyperStore compliments WekaFS and is integrated through Weka's cloud tiering function, adding a cost-effective, exabyte-scale, software-defined object storage to the solution. Cloudian offers modular growth, letting you expand from terabytes to an exabyte without disruption. Embedded data redundancy features provide up to 14 nines of data durability, removing the need for a separate data backup process.

Cloudian HyperStore supports capacity storage workloads using multiple Lenovo server and storage configurations. These include SR650, SR530 with D3280 HDD populated servers for infinite scale-out capacity storage.

Solution Benefits

- POSIX / SMB / NFS / S3 / GPU Direct Storage / CSI protocols access
- Enterprise-grade on-prem object storage platform that scales limitlessly
- Industry leading storage infrastructure with Lenovo ThinkSystem
- Native S3-APIs with industry-leading compatibility
- Scale-out modular design with centralized data management and Single global namespace
- Start small and expand without downtime
- Military-grade security and regulatory compliance certifications
- Hybrid and multi-cloud ready
- Single platform for all applications with 1/3 savings

Use Cases

- High performance file and object storage
- Machine Learning, AI, Advanced Analytics and Big Data
- Life Sciences Research, Genomics
- Financial Services, High Frequency Trading, Compliance



Award-Winning

Proven at over 700 enterprises and managed service providers worldwide with nearly two exabytes of capacity under management, Cloudian Scored Highest for All Use Cases in the Gartner 2020 Critical Capabilities for Object Storage Report and was named a Gartner Peer Insights Customers' Choice in 2020, 2021, and 2022.

Together WekaFS and Cloudian Hyperstore on Lenovo ThinkSystem unify and simplify the data pipeline for performance-intensive workloads and accelerated DataOps, all at 1/3rd lower TCO than traditional storage systems.

Solution Advantages

High Performance

WekaFS is an ultra-low-latency, high-throughput solution that is purpose built for environments running concurrent workloads. It is architected to eliminate compute cluster bottlenecks and reduce valuable processing times, making it ideal for iterative workloads like AI/machine learning. As the world’s fastest shared parallel file system, WekaFS is 3x faster than local file systems and 10x faster than traditional NAS. Running on the industry’s best, Lenovo SR 630, WekaFS provides storage for the highest performance workloads.

Scalable

Cloudian HyperStore brings the flexibility and elasticity of the cloud within your data center. With multiple Lenovo ThinkSystem deployment possibilities, customers can start small – as small as a 3-node configuration -- and scale out to thousands of nodes as needed. These nodes can be physical or virtual, running on industry-best Lenovo ThinkSystem SR650 amongst other options. In addition, unlike some systems that require all nodes to be identical, HyperStore lets you add heterogeneous nodes of any size, providing scalability across multiple data centers or facilities anywhere in the world.

Secure

The combined solution on Lenovo ThinkSystem provides extensive security features to deploy and operate a protected storage solution that is FIPS, CFTC 4511, SEC 17 a-4, Common Criteria compliant and certified at the capacity tier. Security features include:

- Data encryption and transparent key management
- AES-256 server-side encryption for data stored at rest
- SSL encryption for data in transit (HTTPS)
- Role-based access controls with specified levels of access
- Fine-grained storage policies and Audit trail logging
- WORM (Write Once Read Multiple) for storage of immutable data
- Extremely short RTO providing near-instant recovery of files due to Ransomware
- Flexible RPO scheduling to meet any file protection requirement

Resilient

The solution provides high data durability with the option to protect and distribute data using replication or erasure coding. Administrators can configure the number of replicas or type of erasure code scheme required to meet SLA and cost objectives. Storage policies also provide fine grain control of data placement across data centers, taking into consideration factors such as cost efficiency, security levels, and proximity.

Multi-tenant

The solution allows multiple users on shared infrastructure without compromising security. Granular access control and audit logging capabilities control and logically separate data access. Users can securely access data from the same nodes without impacting operations. Administrators can also control quality of service (QOS) by limiting usage rates and setting quotas on a per-group, per-user basis.

Economical

Running on cutting edge Lenovo ThinkSystem with local NVMe SSD’s, the solution drives down the cost of storage for analytics workloads by one-third as compared to traditional storage.



Cloudian is the leader in data management software for hybrid/multi-cloud. With military-grade security, limitless scalability, and seamless cloud integration, Cloudian’s S3-compatible object storage lets enterprises and managed service providers optimize data access, meet data sovereignty requirements, and cut costs by consolidating information to a single, cloud-like platform. Cloudian’s geo-distributed architecture manages and protects object and file data at the edge, core, and in the cloud, for both conventional and modern applications

More at <https://www.cloudian.com>.



Lenovo (HKSE: 992) (ADR: LNVGY) is a US\$70 billion revenue global technology powerhouse, ranked #171 in the Fortune Global 500, employing 75,000 people around the world, and serving millions of customers every day in 180 markets. Focused on a bold vision to deliver smarter technology for all, Lenovo has built on its success as the world’s largest PC company by further expanding into key growth areas including server, storage, mobile, solutions and services. This transformation together with Lenovo’s world-changing innovation is building a more inclusive, trustworthy, and sustainable digital society for everyone, everywhere.

To find out more, visit <https://www.lenovo.com>.



WEKA delivers the industry’s first multi cloud data platform for artificial intelligence (AI), machine learning (ML), and other next-generation workloads. The WEKA® Data Platform’s advanced architecture is optimized to solve complex data management challenges and provide consistently epic performance at scale across on-premises, edge, hybrid, and multi-cloud environments. WEKA is accelerating innovation, research, and discovery for the world’s leading enterprises and research organizations – including eight of the Fortune 50 – and helping them to achieve first-to-market results with AI. WEKA is backed by nine world-class strategic investors, including Hitachi, HPE, and NVIDIA, and operates in more than 20 countries worldwide.

For more information, visit <https://www.weka.io>.