



AIStor for HDFS modernization

AIStor eliminates HDFS bottlenecks while delivering the scale, speed, and cost optimization required for modern data-driven enterprises.

Uncompromised Performance

- 19.2 TiB/s throughput demonstrated at 1.0 EiB scale via ExaPOD Reference Architecture.
- Sub-millisecond latency for many object operations, enabling real-time analytics.
- Linear scaling up to 800Gbps per storage server with high availability.
- Superior small file performance with direct PyTorch and TensorFlow integration.

Economics that Scale

- Up to 60% cost reduction .
- 67% less storage overhead through erasure coding vs HDFS 3x replication
- Independent scaling of compute and storage eliminates overprovisioning waste.
- Substantial productivity gains. Millions already know S3 APIs vs specialized Hadoop training.

Zero-Friction Operations

- Kubernetes-native design engineered for container orchestration.
- Native S3 API unlocks thousands of modern tools without custom integration.
- Global enterprises manage petabytes with minimal operational overhead.
- Non-disruptive migration. S3A connector enables gradual transition while existing Hadoop jobs continue.

The growing gap between HDFS and modern data lakehouses.

Every day, your data grows. Every day, in HDFS your storage volume grows three times faster. That's the reality of 3x replication: 30 petabytes of storage to protect 10 petabytes of data. While competitors train AI models with object storage at exabyte scale, you're managing NameNode memory limits and scheduling weekend maintenance windows for manual rebalancing. While they leverage elastic, cloud-native architectures, you're constrained by HDFS's tightly coupled compute-storage design that forces overprovisioning.

HDFS infrastructure costs can reach significantly higher levels than modern storage alternatives. The opportunity cost compounds daily. While you provision new HDFS capacity with unused compute nodes, competitors have already deployed their next AI feature using disaggregated, cost-optimized infrastructure.

Cloudera ended support for CDH in March 2022. If your data doubles every 18-24 months (the industry average), you're already storing twice as much as you had when support ended. Every dollar spent maintaining legacy infrastructure is a dollar not invested in AI innovation.

The seven critical advantages of AIStor v. HDFS

HDFS Reality

1. Architectural scaling limitations

NameNode architecture creates hard limits. With 1GB of heap memory supporting approximately 1 million files, Cloudera recommends staying under 300 million files for optimal performance. While HDFS Federation can provide horizontal scaling, it requires complex namespace management and application-level coordination.

2. Enterprise multi-tenancy complexity

Secure multi-tenancy in HDFS requires complex Kerberos configurations, custom quota management, and careful HDFS permission orchestration. Teams often deploy separate clusters to achieve true isolation, multiplying infrastructure costs.

3. Operational scaling bottlenecks

HDFS's tightly coupled architecture forces you to scale compute and storage together, resulting in significant resource underutilization. Combined with 3x replication overhead, storing one exabyte requires three exabytes of capacity plus compute nodes you don't need.

4. Geographic distribution complexity

Cross-datacenter HDFS deployments require complex architectures with DistCp jobs, eventual consistency challenges, and significant bandwidth overhead. Disaster recovery often means accepting hours of data loss.

5. Non-container-native by design

Running HDFS on Kubernetes requires extensive custom engineering: StatefulSets with headless services, complex network discovery, and custom identity management solutions. Integration with modern CI/CD pipelines requires additional abstraction layers.

MINIO **AI*STOR** Solution

1. Architectural scaling without limits

Fully distributed architecture with no single points of failure. Scale horizontally from terabytes to exabytes without architectural bottlenecks. Enterprises routinely migrate multi-exabyte workloads without the namespace complexity of Federation.

2. Simplified Enterprise multi-tenancy

Native bucket-level isolation with IAM policies, quotas, and encryption keys per tenant. Consolidate multiple workloads on a single cluster while maintaining strict security boundaries. Built-in encryption at rest and in transit protects data without performance impact.

3. Operational scaling simplicity

Kubernetes-native automation eliminates manual operations. Your existing cloud-native team can manage AIStor without specialized training. Native integration with Spark, Presto, Jupyter notebooks, Kubeflow, and MLflow significantly accelerates development cycles. Scale storage and compute resources independently and as needed.

4. True geographic distribution

Active-active replication across sites with configurable consistency. Deploy globally with automatic failover and minimal RPO/RTO. Geographic distribution is configuration, not custom engineering.

5. Cloud-native by design

Deploy with standard Kubernetes manifests, scale with native orchestration, and integrate using S3 APIs your teams already know. Direct integration with GitHub Actions, GitLab CI, and Jenkins. No custom engineering required.

6. Complex compliance adherence

Meeting SEC 17a-4(f), GDPR, or HIPAA requirements with HDFS requires additional architectural layers and third-party tools. Organizations must implement custom solutions for WORM functionality, comprehensive audit trails, and immutable storage.

7. File, not Table Architecture

HDFS requires separate metadata stores, complex schema evolution, and manual partition management. Time travel and ACID transactions require additional tooling.

6. Compliance without complexity

Enterprise-grade compliance out of the box. Native object versioning, WORM-compliant immutable storage for SEC 17a-4 requirements, comprehensive lifecycle management, and automated bitrot detection with self-healing.

7. Native Iceberg Datalakehouse Architecture

Built-in Iceberg catalog provides schema evolution, time travel and ACID transactions without external dependencies. Native integration eliminates metadata management overhead while enabling both AI and analytics in one unified data store.

Real customer success stories

Global Energy Leader

SmartGrid with 66% Less Infrastructure

Challenge: 45+ power plant operator needed 240 Hadoop nodes and 14 months to support smart meter analytics for millions of customers.

Solution: Deployed complete data lake platform in 10 weeks including OT integrations, using just 48 MinIO AIStor nodes (24 per site across 2 HA sites) with fully automated deployment and operations.

Impact: 66% infrastructure reduction, <1 FTE operations through automation, 4-day PoC to validate performance

Global Financial Institution

60% Cost Reduction at Double Scale

Challenge: Growing Hadoop environment was causing performance degradation and stability issues for critical compliance, fraud detection, and risk analytics.

Solution: Dual-site MinIO AIStor with active-active replication supporting 100+ projects.

Impact: 60% cost savings while doubling capacity, 30% ML performance improvement, zero-downtime DR.

Your risk-free modernization path.



This approach eliminates "rip and replace" risk while delivering immediate value through improved developer productivity and reduced operational overhead.

Your organization faces a simple choice: continue investing in HDFS infrastructure that grows 3x faster than your data, or adopt the cloud-native architecture already proven at scale. MinIO AIStor delivers the performance, economics, and operational simplicity that modern workloads demand