SARDINA

AI, ML, and the Ceph Advantage: Scalable Storage for Smarter Workflows

by Dr Kenneth Tan,
Sardina Systems

Session Highlights

- Why scalable storage matters for AI/ML
- Ceph's role and scale-up challenges in AI/ML workflows
- Hybrid, edge, and multi-cloud AI with Ceph
- Optimizing and tuning AI pipelines
- Ensuring performance, reliability, and scale
- Hardware and deployment best practices for AI/ML workflows

Dr Kenneth Tan
Sardina Systems, Executive Director

- Over 20 years of experience in the technology sector, supercomputing, defense, investment banking
- PhD in Computer Science
- Cloud economics, technology trends analysis, sustainable data center operations, and international business development
- Leads a top-notch European team of Sardina Systems

Sardina Systems

11 years
of operation

12 countries
where we have
clients and partners

90%
customer retention rate

UK-headquartered | Operating Globally

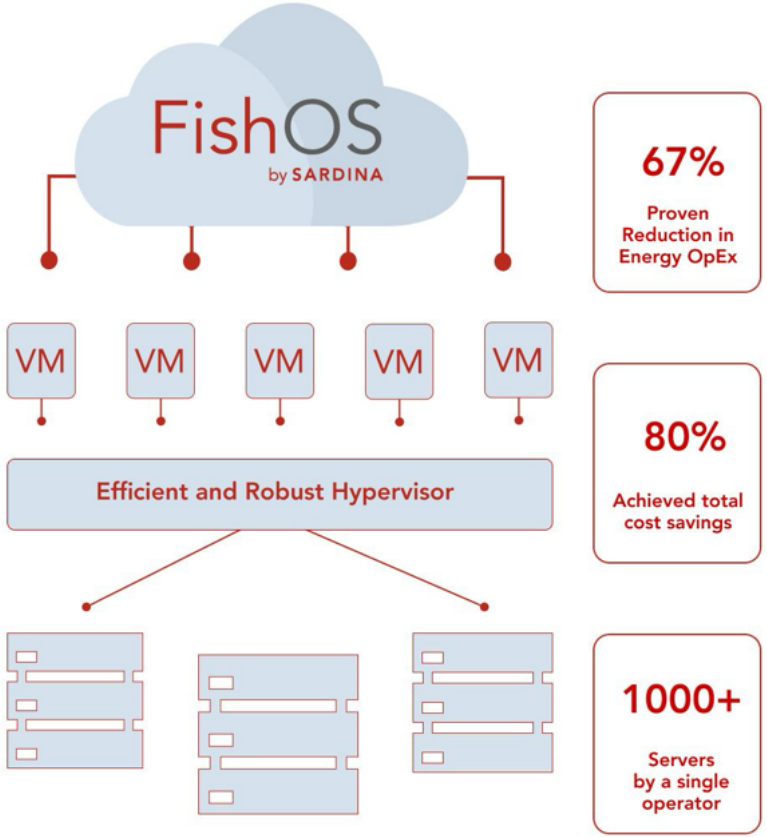
What is FishOS?

FishOS® is a cutting-edge cloud management platform built on OpenStack, Kubernetes, and Ceph.

It offers a smart, efficient, and fully-automated solution for managing cloud infrastructure.

Includes 11 uniquely developed components and utilizes a flexible, license-based pricing model.

FishOS has been a three-time winner of international awards.



Why Storage Matters in AI/ML

AI is not just about GPUs. Data is the foundation.

AI/ML success depends on:

- Scalable, high-throughput data pipelines
- Low-latency data access
- Resilient, flexible infrastructure

Traditional storage systems often limit innovation — they're monolithic, siloed, and difficult to scale.

Without intelligent storage, GPU clusters are left idle, waiting for data, wasting both time and money.

ceph – Built for Scale and Intelligence

Ceph is trusted by enterprises, cloud providers, and research institutions worldwide — for a good reason.

- Software-defined, distributed architecture
- Unified storage (Block, Object, File)
- Cloud-native ready — hybrid/multi-cloud compatible
- Scales horizontally with commodity hardware
- Distributed, software-defined storage
- Self-healing, fault-tolerant and no vendor lock-in

Ceph scale-up challenges are real, but manageable with the right expertise, automation, and hardware tuning.

Ceph & AI/ML — A Natural Fit

Ceph fits naturally into the modern AI/ML development cycle.

Separation of compute and storage
➔ Flexible pipeline design

Erasure coding and self-healing
➔ Cost-effective reliability

Integrates with K8s, CI/CD
➔ DevOps-friendly

CephFS
➔ Shared training datasets

RADOS and S3
➔ Model checkpoints, logs, data lakes

While Ceph has a learning curve, its strong ecosystem, devops tooling, and unified architecture make it highly suitable for AI/ML teams.

Hybrid & Edge AI with ceph

Unified storage layer across edge, on-prem, and cloud

Smart factories, autonomous systems, healthcare

Support for federated learning, edge inference

Ceph spans environments — AI lives everywhere

Optimized AI Pipelines with Ceph

AI Pipeline Needs	How Ceph Delivers
Data ingestion & staging	High-throughput object, block, file storage
Model training	Fast read/write (eg: RBD, CephFS)
Inference at edge/cloud	S3-compatible API support for seamless access
Data versioning & backups	Scalable & efficient snapshots, replication
Automation	API-driven management & integrations

Busting the Myths: Is Ceph Slow?

- Slowness often due to unfair benchmarks or outdated hardware
- NVMe, BlueStore, and smart cluster design deliver high throughput

Other misconceptions:

❌	✅
"Ceph is only for storage admins"	DevOps-friendly
"Ceph doesn't support cloud-native AI"	Supports Kubernetes, S3, and OpenStack
"Ceph is difficult"	Its versatility means more configuration options, not complexity

Ceph is AI-ready and scalable for real workloads. You don't need a massive team — you need the right partner.

Ensuring Reliability & Fault Tolerance at Scale

- Erasure coding: data recovery with minimal overhead
- Self-healing: detects and repairs failed objects automatically
- Multi-site replication + object locking
- No single point of failure — critical for high-availability AI ops

When you're training models on weeks of data, reliability isn't optional — it's essential.

Scaling Ceph for Large AI Workloads

Erasure coding
For large objects

Adequate RAM
For better cache performance

NVMe/SSDs
For training data pools

GPU adjacency
To help with throughput

Ceph thrives at scale with proper planning

Sardina Systems helps clients deploy massive AI/ML storage without trade-offs.

Performance Tuning for AI Workloads

- Use BlueStore with tuned journal settings
- Read-ahead for large training datasets
- Customize pools: replication vs erasure coding
- Tune based on task (inference = low latency, training = throughput)
- Storage tuning is just as important as model tuning

How to Seamlessly Deploy Ceph in AI/ML Workflows

1. Deploy and manage Ceph natively with minimal overhead
2. Integrate with Helm, Terraform, and Ansible for streamlined workload setup and scaling
3. Leverage built-in dashboards and observability tools for real-time monitoring
4. Work with OpenStack, FishOS, and hybrid/multi-cloud infrastructure

Even with initial expertise required, Ceph's integration ecosystem is a major boost for devops pipelines.

Hardware Best Practices for AI+Ceph

Ceph is hardware-agnostic — but making the right choices unlocks massive gains for AI/ML performance, scalability, and reliability.

NVMe and SSDs for Performance-Critical Workloads

- NVMe for hot data (e.g., training sets, active models)
- Ceph BlueStore to reduce overhead and unlock full disk performance

Plan for Storage Tiering

- Combine SSD/NVMe (performance tier) with HDDs (capacity tier)
- Use Ceph pools to manage datasets across tiers intelligently

Optimize Networking

- Dual 25/40/100GbE or InfiniBand for low-latency data flow
- Isolate public (client-facing) and cluster (replication/backfill) networks

Leverage GPU-Accelerated Nodes Where Needed

- Optimize for high bandwidth, low latency: dual 25/40/100GbE or InfiniBand
- Keep storage and compute loosely coupled to scale independently

Key Takeaways

- AI/ML needs smart storage, not just fast GPUs
- Ceph is scalable, resilient, cloud-native
- Myths around complexity & speed are outdated
- FishOS Ceph simplifies production use
- DevOps-friendly, performance-tuned, future-proof

ceph days

SARDINA

Thank You!

Let's Stay Connected!