Why We Replaced S3FS

Lessons from building reliable POSIX storage for AI workloads on Kubernetes

KubeCon CloudNativeCon North America 2025

Motivation

Pain Points of S3FS in Real AI/ML Workloads

- **Semantic Mismatches, Not a Real Filesystem**
  Apps like PyTorch misbehave.
- **Performance and Scalability Bottlenecks**
  Throughput drops and distributed training jobs stall.
- **Multi-Tenant Pitfalls**
  Concurrent workloads become unreliable and hard to manage.
- **Hidden Cost Traps**
  Storage bills skyrocket as datasets grow.

S3FS is easy to mount, but fails under real AI workloads.

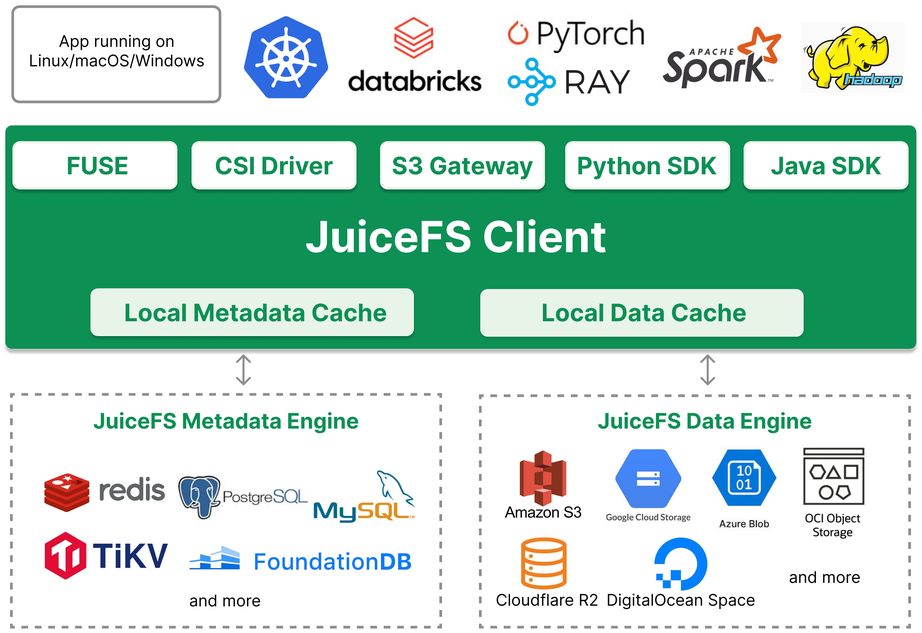
Approach

A Scalable POSIX Filesystem on Object Storage

- **Designed for Cloud, Not the Data Center**
  Optimized for elasticity and cloud-native operations.
- **Decoupled Metadata and Data**
  Scale metadata independently from data throughput to eliminate central bottlenecks.
- **Data Stored on Object Storage**
  Leverage cost-efficient, durable object storage without sacrificing POSIX compatibility.
- **Stay in User Space (FUSE)**
  Faster iteration, simpler upgrades, and Kubernetes-friendly deployment model.

A cloud-native design that unifies object storage with POSIX semantics.

Architecture & Core Features



Core Features

- **Multi-Protocol Access** - One namespace for POSIX, S3 and HDFS.
- **Intelligent Caching** - Accelerates hot data access up to 10x.
- **100 Billion Files in a Single Volume** - Proven in production.
- **Mature Kubernetes CSI Driver** - Native integration for automation.

We bring POSIX guarantees and high performance to object storage, without changing your AI stack.

JuiceFS **GitHub** 12.5K stars

Results



Large-Scale Validation

Test with 100 GCP clients:
- Random read: 7.3 M IOPS
- Sequential read: 1.2 TB/s throughput

MLPerf 2.0 benchmark:
Outstanding results on Ethernet-based storage.
Read more https://bit.ly/3IPWfwJ

Across workloads, JuiceFS consistently cuts training time and cost.

Real-World Case Studies

RNA sequencing on AWS
Switching from S3FS to JuiceFS reduced runtime from 26 h → 7 h.
Read more https://bit.ly/4omec52

LeptonAI
Cut cloud storage cost by 98% for AI workloads.
Read more https://bit.ly/47sz1Gc

BentoML
Reduced LLM model loading time from 20+ min → a few minutes.
Read more https://bit.ly/474NC9r

Key takeaway: JuiceFS eliminates metadata bottlenecks and unpredictable I/O, enabling scalable, cost-efficient, and production-ready storage for AI/ML pipelines on Kubernetes.

Across industries, JuiceFS consistently cuts training time and cost.

Lessons & Ecosystem Impact

Lessons Learned:
"Object storage ≠ POSIX."
"Metadata scalability is the true bottleneck."
"Kubernetes-native automation unlocks massive scale."

Benefits to Ecosystem:
This work goes beyond a single implementation, contributing reusable patterns and proven practices to the community:
- **Reusable Architecture for POSIX on object storage.**
- **CSI & Operator Extensions for user-space filesystems.**
- **Scalable AI Storage design for hybrid clouds.**

Try it yourself:
- GitHub: github.com/juicedata/juicefs
- Community: https://go.juicefs.com/slack

Follow us on LinkedIn

Rui Su, co-founder of JuiceFS
https://juicefs.com | rsu@juicedata.io