

FFUSE: I/O Fault Injection and Resilience Scoring for HPC and AI Workloads

Authors:

Anila Ghazanfar, Julian Kunkel

As HPC and AI/ML workloads increasingly co-locate on shared storage infrastructure, I/O reliability becomes critical to both domains. Filesystem faults including errors, partial reads, latency spikes, and silent corruption can cause crashes or undetected data loss. No prior work systematically compares how HPC and AI workloads respond to identical filesystem faults or provides unified cross-domain resilience metrics. We present FFUSE, a FUSE-based fault injection framework that injects seven categories of filesystem faults at the application I/O layer without modifying code. We introduce FIORE (Fault Injection Observed Resilience Evaluation), a metric quantifying failures without error reporting to assess fault awareness. Testing 52 fault injection scenarios reveals fundamental differences: LAMMPS (HPC) demonstrated 53.8% crash rate versus ResNet-50 (AI) that showed 69.2% crash rate. Both show universal vulnerability to error codes (100% crash) and silent corruption (100% undetected without checksums). These findings reveal that checkpoint format and execution model, not application code, determine resilience. This first cross-domain characterization of HPC and AI resilience under FUSE-layer faults provides practitioners with concrete guidance for improving checkpoint reliability in converged HPC-AI systems.