



Contribution ID: 1670 Contribution code: THPL008

Type: **Poster Presentation**

Upgrades to logging and ml analytics architecture at APS

Thursday, 11 May 2023 16:30 (2 hours)

Several machine learning (ML) projects on anomaly detection and optimization were recently started at the Advanced Photon Source (APS). To improve training data quality, and accommodate the upcoming APS Upgrade changes, a large increase in the number and size of log files is expected. Recent studies found performance bottlenecks in the current log analysis architecture, especially for large ML analytics tasks. We explored several approaches to improve both data density and throughput. First, we swapped lzma compression algorithm for modern alternatives like zstd and lz4, scanning presets to find an optimal one that increased decompression throughput by 10x for a 20% file size increase. Several lossy compression schemes were attempted to take advantage of limited device resolution and ML quantization, yielding further size decreases with reasonable fidelity losses. Finally, we tested several analytics and time-series databases, finding them faster for both linear and random-access reads while maintaining good compression ratios. They also enabled offloading analytics computations to server nodes, reducing network load. Our results indicate that with some effort, it is possible to increase the amount of logged data significantly while improving ML analytics performance.

Funding Agency

The work is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

Footnotes

I have read and accept the Privacy Policy Statement

Yes

Primary author: KUKLEV, Nikita (Argonne National Laboratory)

Co-authors: LOBACH, Ihar (Argonne National Laboratory); SHANG, Hairong (Argonne National Laboratory); SOLIDAY, Robert (Argonne National Laboratory)

Presenters: KUKLEV, Nikita (Argonne National Laboratory); LOBACH, Ihar (Argonne National Laboratory)

Session Classification: Thursday Poster Session

Track Classification: MC6: Beam Instrumentation, Controls, Feedback and Operational Aspects: MC6.T22: Reliability, Operability