



Contribution ID: 165

Type: Poster

The High-throughput Data I/O framework for HEPS

The High Energy Photon Source produces vast amounts of diverse, multi-modal data annually, with IO bottlenecks increasingly limiting scientific computational efficiency. To overcome this challenge, our approach introduces a threefold solution. First, we develop daisy-io, which has a unified IO interface designed for cross-disciplinary applications, which integrates accelerated data retrieval techniques such as parallel processing and prefetching to optimize access speeds across heterogeneous datasets. Second, we construct a data streaming platform that eliminates disk read/write bottlenecks through real-time data handling. This platform incorporates three core components: a stream ingestion module for dynamic data reception, a stream parsing module for on-the-fly structural processing, and a stream buffering module for temporary data staging. Finally, to further enhance data transmission efficiency, we implement a lightweight serialization protocol and domain-specific compression algorithms, minimizing latency and bandwidth demands. Collectively, these innovations not only accelerate data read/write operations but also abstract complexities arising from disparate data sources and formats, enabling seamless integration into scientific workflows while maintaining adaptability across experimental scenarios.

Significance

References

Experiment context, if any

Authors: FUSY, FU Shiyuan; Dr SUN, Hao-Kai (IHEP, CAS); LIU, Jianli (IHEP); HU, Yu; WANG, lei (Institute of High Energy Physics); QIFAZHI, 齐法制

Presenter: FUSY, FU Shiyuan

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research