Accelerating Exascale Scientific Discovery via In-Situ and In-Transit Data Analytics in HPC

Vijayalakshmi Saravanan
vsaravanan@uttyler.edu
vijayalakshmi.saravanan@utdallas.edu
University of Texas at Tyler
Tyler, TX, USA
University of Texas at Dallas
Dallas, TX, USA

Sai Karthik Navuluru
SaiKarthik.Navuluru@utdallas.edu
University of Texas at Dallas
Dallas, TX, USA

Khaled Z. Ibrahim
KZIbrahim@lbl.gov
Lawrence Berkeley National Laboratory
Berkeley, CA, USA

Abstract

The rapid growth of multimodal scientific data from large-scale simulations and experimental instruments is placing unprecedented demands on storage, I/O, and analysis workflows. Conventional post hoc approaches, which depend on disk-based processing, suffer from high latency, bandwidth bottlenecks, and inefficient resource utilization, limiting their ability to deliver timely scientific insights. To overcome these limitations, this work investigates an in-situ and in-transit processing framework that embeds computation directly into the memory and storage hierarchy of high-performance computing (HPC) systems. In-situ processing executes computation at the data source, using node-local memory and accelerators to filter, reduce, or analyze data before it leaves the compute node. In-transit processing complements this by utilizing intermediate storage layers, including burst buffers, or dedicated analysis resources to perform computation asynchronously, balancing workloads between simulations and analytics. Building upon a hybrid architecture that integrates Apache Ignite's in-memory data grid with Apache Spark's distributed computation and containerized microservices, the framework enables real-time ingestion, fusion, and machine learning over heterogeneous scientific datasets, as shown in Figure 1. By embedding analytics into multi-tier storage hierarchies spanning node-local memory, burst buffers, and parallel file systems, the proposed approach minimizes I/O overhead, preserves data fidelity, and facilitates scalable ML-driven techniques such as anomaly detection, change point detection, and uncertainty quantification. Our initial findings show decreased read/write latency, effective utilization of CPU and memory usage, and strong scalability for complex workflows.

This work-in-progress paper presents case studies including molecular dynamics trajectories from NWChem simulations and E3SM climate modeling data, illustrating the framework's adaptability across various scientific fields. This work advances a data-aware HPC paradigm in which computation and storage interact seamlessly through in-situ and in-transit processing, accelerating time-to-insight and enabling exascale-class scientific discovery.

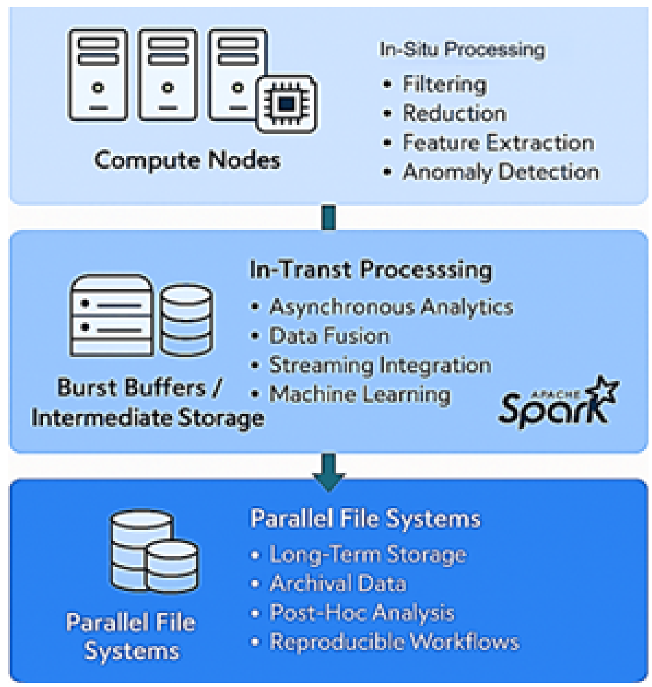

Figure 1. In-situ and in-transit processing integrated into the HPC memory and storage hierarchy, enabling real-time analysis at compute nodes, asynchronous analytics at intermediate storage, and post-hoc analysis in parallel file systems.

CCS Concepts: • **Computing methodologies** → **Modeling methodologies**; **Machine learning algorithms**; • **Mathematics of computing** → **Probabilistic algorithms**; **Mathematical analysis**.

SC Workshops '25, St Louis, MO, USA

ACM ISBN 979-8-4007-1871-7/2025/11
https://doi.org/10.1145/3731599.3767463

Keywords: In-situ analysis, In-transit, HPC and in-memory computing

ACM Reference Format:
Vijayalakshmi Saravanan, Sai Karthik Navuluru, and Khaled Z. Ibrahim. 2025. Accelerating Exascale Scientific Discovery via In-Situ and In-Transit Data Analytics in HPC . In *Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC Workshops '25), November 16–21, 2025, St Louis, MO, USA*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3731599.3767463

Acknowledgments

This material is based upon work supported by the Advanced Scientific Computing Research Program in the U.S Department of Energy, Office of Science, under FAIR Award number DE-SC0026194. This manuscript has been authored by an author at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility using NERSC award ASCR-ERCAP0031060.