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

Vijayalakshmi Saravanan, Sai Karthik Navuluru, Khaled Z. Ibrahim

University of Texas at Tyler | University of Texas at Dallas

Lawrence Berkeley National Laboratory

PDSW Workshop - SC'25

Motivation & Problem Statement

Exascale Data Crisis

- Petabyte-scale data from simulations and experiments
- Unprecedented storage and I/O bandwidth demands

Traditional Limitations

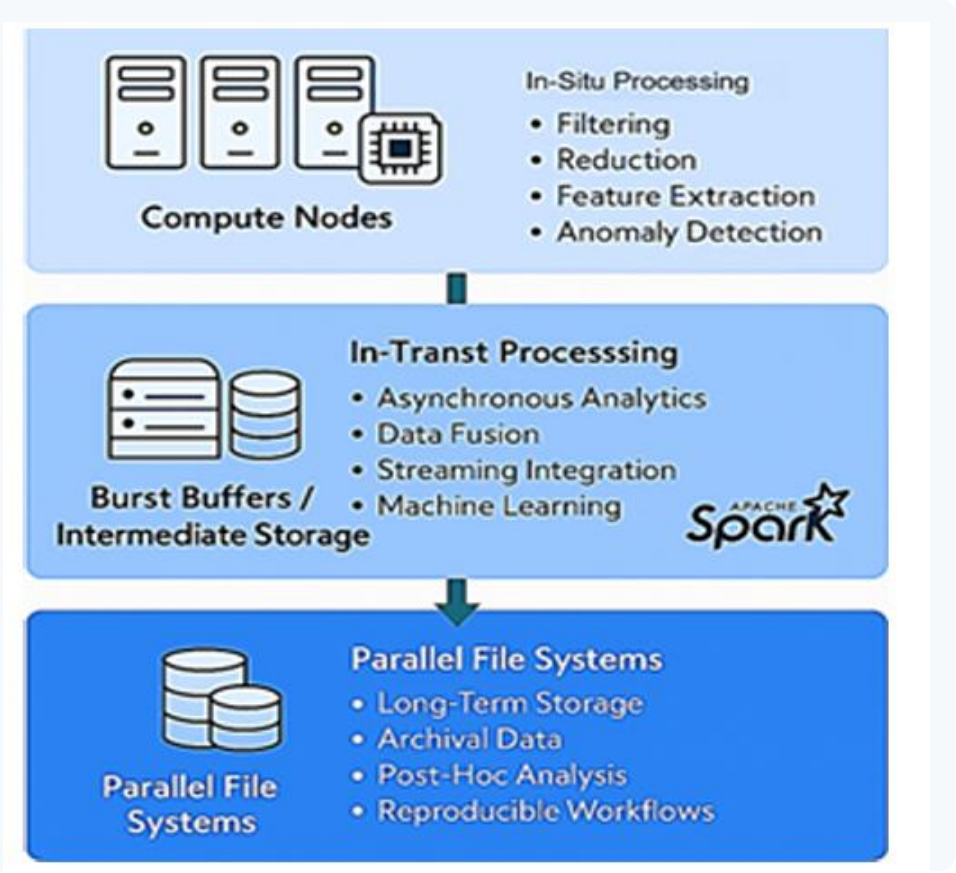
- High latency between generation and analysis
- I/O bottlenecks and inefficient resources
- Inability to deliver timely insights

Our Solution

In-Situ & In-Transit Processing

- Embed computation into memory/storage hierarchy
- Process at source and intermediate tiers
- Real-time analytics and ML-driven insights

Multi-Tier In-Situ & In-Transit Framework



Key Components

- Apache Ignite: In-memory data grid
- Apache Spark: Distributed computation
- Containerized microservices
- ADIOS2: High-performance I/O

Enabled Analytics

Anomaly Detection • Change Points • Uncertainty Quantification • ML Patterns

Technical Implementation

In-Situ Processing

Computation executes at data source using node-local memory and accelerators. Data filtered, reduced, and analyzed before leaving compute node. Minimal I/O overhead.

In-Transit Processing

Asynchronous analytics at intermediate storage layers. Dedicated analysis resources decouple simulation from analytics. Scalable ML pipeline execution.

Storage Integration

Multi-tier hierarchy: node-local memory → burst buffers → parallel file systems. Seamless data flow with intelligent placement strategies.

Software Architecture

- **Apache Ignite:** O(1) key-value lookups, ACID transactions, SQL interface
- **Spark RDDs:** Fault-tolerant distributed datasets, DAG execution
- **MLlib:** Scalable machine learning algorithms
- **Containers:** Reproducible environments, dynamic allocation

Case Study Domains

Molecular Dynamics: NWChem simulation trajectories

Climate Modeling: E3SM multivariate datasets

Performance Results & Impact

↓ Latency
Decreased I/O latency

✓ Efficient
Optimized resources

Real-Time
In-workflow analytics

Key Achievements

- Eliminated disk I/O bottlenecks
- Asynchronous resource optimization
- Strong workflow scalability

Scientific Impact

Enables real-time analytics on exascale systems

Paradigm shift from post-hoc to in-workflow analysis

Future Directions

Extended case studies • Workflow tuning • Emerging storage integration

Our framework represents a fundamental shift from storage-centric to computation-centric HPC workflows, enabling real-time scientific insight extraction at exascale.