Rich Operators for Advanced In Situ Processing

Gnosis Research Center @ ILLINOIS TECH

Hua Xu Anthony Kougkas Xian-He Sun
hxu40@hawk.iit.edu,akougkas,sun@iit.edu

Sandia National Laboratories
Oak Ridge National Laboratory
U.S. Department of Energy

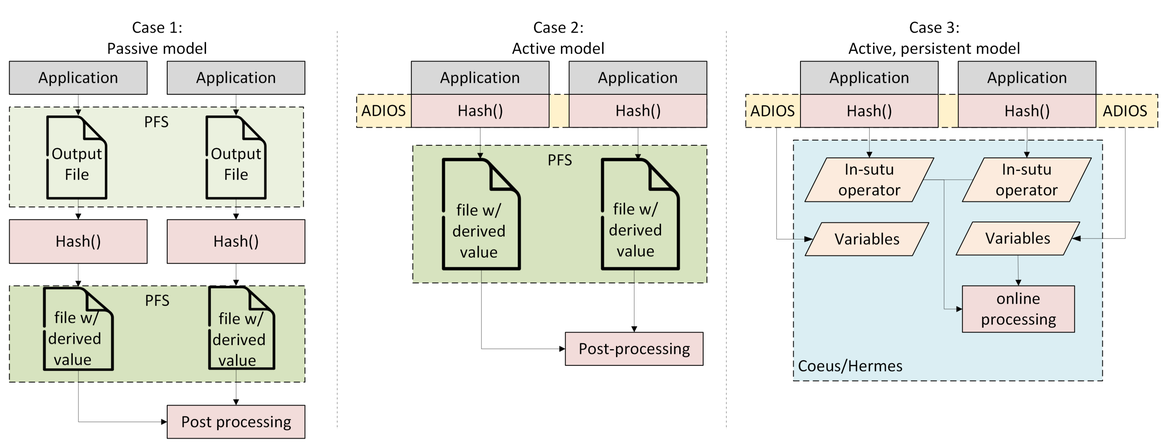
Introduction

In situ operations[1] have become increasingly important as data volumes generated exceed the acceptable time for reading and writing. With base science codes being complex and multi-use, incorporating in situ operations directly into the science code is less than ideal. Instead, this work looks to high-level I/O libraries as a place for pluggable, complex operators isolating the in situ components from the host code, but still enabling full data structure knowledge.

Design

To fully understand the design of our I/O-centric operator, we adopt a run validation scheme from the Recup project, utilizing hash operations stored as ADIOS2-derived variables[2]. We trace the data flow from its initial generation, through the computation of derived quantities, to its efficient handling and online validation. In particular, we outline how our approach addresses the following:

1. The initial handling of simulation data and the activation of the I/O layer to define and compute derived quantities
2. The hashing mechanism used for efficient data summarization and validation.
3. Hierarchical storage management strategies enabling optimized storage and retrieval of derived data.
4. The runtime control mechanisms that orchestrate online process operations, leveraging stored derived data for rapid validation.



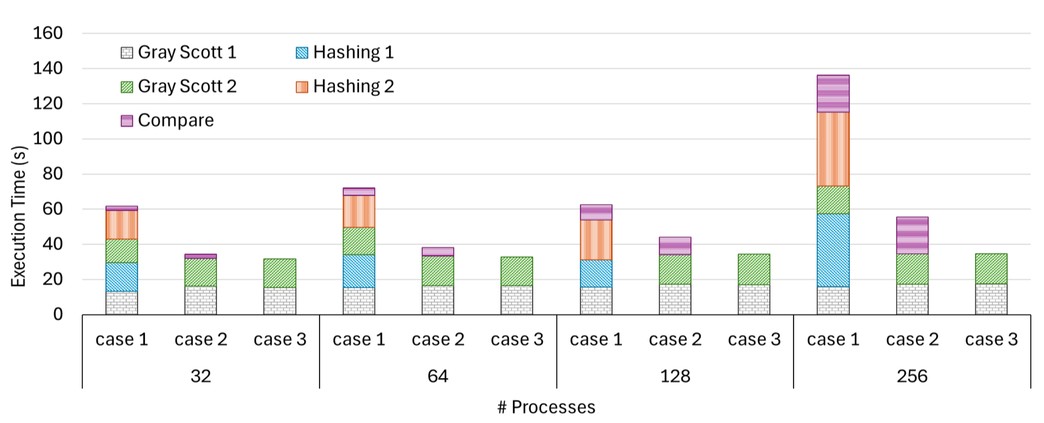
Objective

1. Develop our I/O-focused operator, highlighting complex functionality, such as reading from storage, and optimization and scalability using additional tools without requiring changes to the application.
2. Investigate fully offline, hybrid, and fully online deployments with extensive experiments to understand their trade-offs.

Evaluation

(a) Ares from IIT

(b) Doom from Sandia



(c) Frontier from Oak Ridge

Execution time on three platforms for I/O scaling from 32 to 256 processes in Gray-Scott

Conclusion

We conducted a comprehensive evaluation across three scenarios, each exhibiting varying degrees of online and active capabilities. We demonstrated that incorporating operators into the I/O layer can mitigate these bottlenecks and reduce computational overhead.

Future work

For future work, we plan to expand the active I/O layer of our Coeus system by introducing more advanced and versatile operators. By embedding these operators directly into the I/O layer, we can develop generic and portable solutions that not only reduce I/O overhead but also make intelligent use of the contextual knowledge available within the I/O subsystem.

References

[1] Fang Zheng, Hasan Abbasi, Jianting Cao, Jai Dayal, Scott Schwan, and Norbert Podhorszki. In-situ i/o processing: a case for location flexibility. In *Proceedings of the Sixth Workshop on Parallel Data Storage*, PDSW '11, page 37–42, New York, NY, USA, 2011. Association for Computing Machinery.

[2] William F. Godoy, Norbert Podhorszki, Ruonan Wang, and Scott Klasky. ADIOS 2: The Adaptable Input Output System. A framework for high-performance data management. *SoftwareX*, 12:100561, 2020.