

Hariharan Devarajan, Anthony Kougkas, Xian-He Sun  
Illinois Institute of Technology, Department of Computer Science  
Acknowledgement to Larry Knox and Neil Fortner of The HDF Group

## ABSTRACT

- To increase I/O performance and reduce the pressure for external storage, HPC systems are adding extra layers to the memory and storage hierarchy: Deep Memory And Storage Hierarchy (DMSH).
- Hermes:** a new, heterogeneous aware, multi-tiered, dynamic, and distributed I/O buffering system in HDF5:
  - VOL plugin to intercept HDF5 API calls.
  - Innovative Data Placement Engine to buffer data in the hierarchy.
  - Multiple I/O clients to support heterogeneous hardware.
  - Multi-tiered prefetching to improve read performances.

**Hermes can significantly accelerate I/O and outperforms by more than 12x state-of-the-art buffering platforms.**

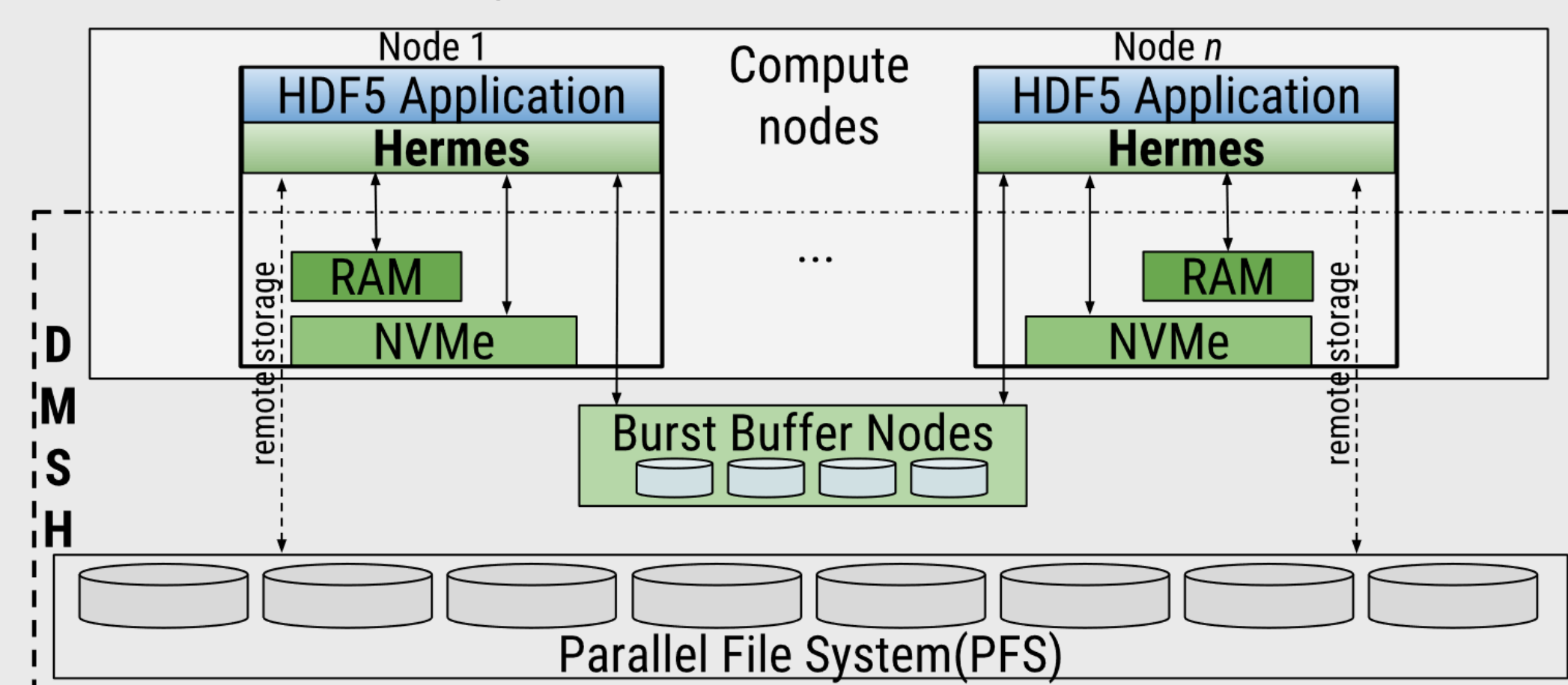
## CONTACT

**Hariharan Devarajan**  
Illinois Institute of Technology  
[hdevarajan@hawk.iit.edu](mailto:hdevarajan@hawk.iit.edu)  
<https://hariharan-devarajan.github.io/>

## HERMES DESIGN

- Hermes captures HDF5 I/O calls and redirects them to different layers of DMSH.
- Legacy applications can easily connect to Hermes by utilizing the VOL infrastructure.
- Hermes' goal is to maximize user productivity by making I/O buffering transparent.
- Vertical hierarchy:** data movement within a compute node and all the way down to the burst buffers and PFS.
- Multi-Layer Prefetching:** for performing smart prefetching on all layers of DMSH.

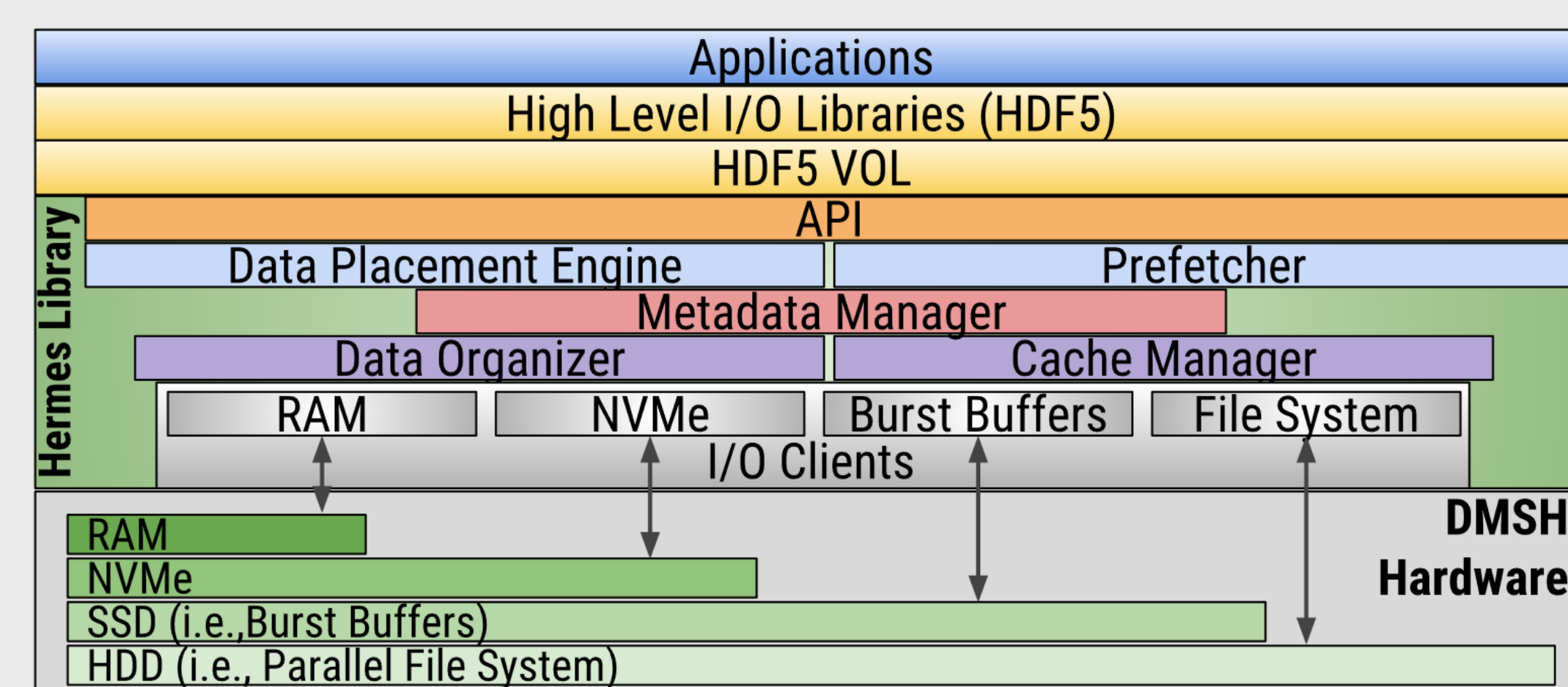
Fig. 1: Hermes Architecture



## HERMES COMPONENTS

- API:** intercepts I/O calls (HDF5 VOL)
- Data Placement Engine:** Maps data to DMSH
- Data Organizer:** Moves data between levels.
- Cache manager:** Manages caching. LRU and LFU
- Prefetcher:** Typical prefetching algorithms.
- I/O Clients:** Modular, extensible

Fig. 2: Hermes Components

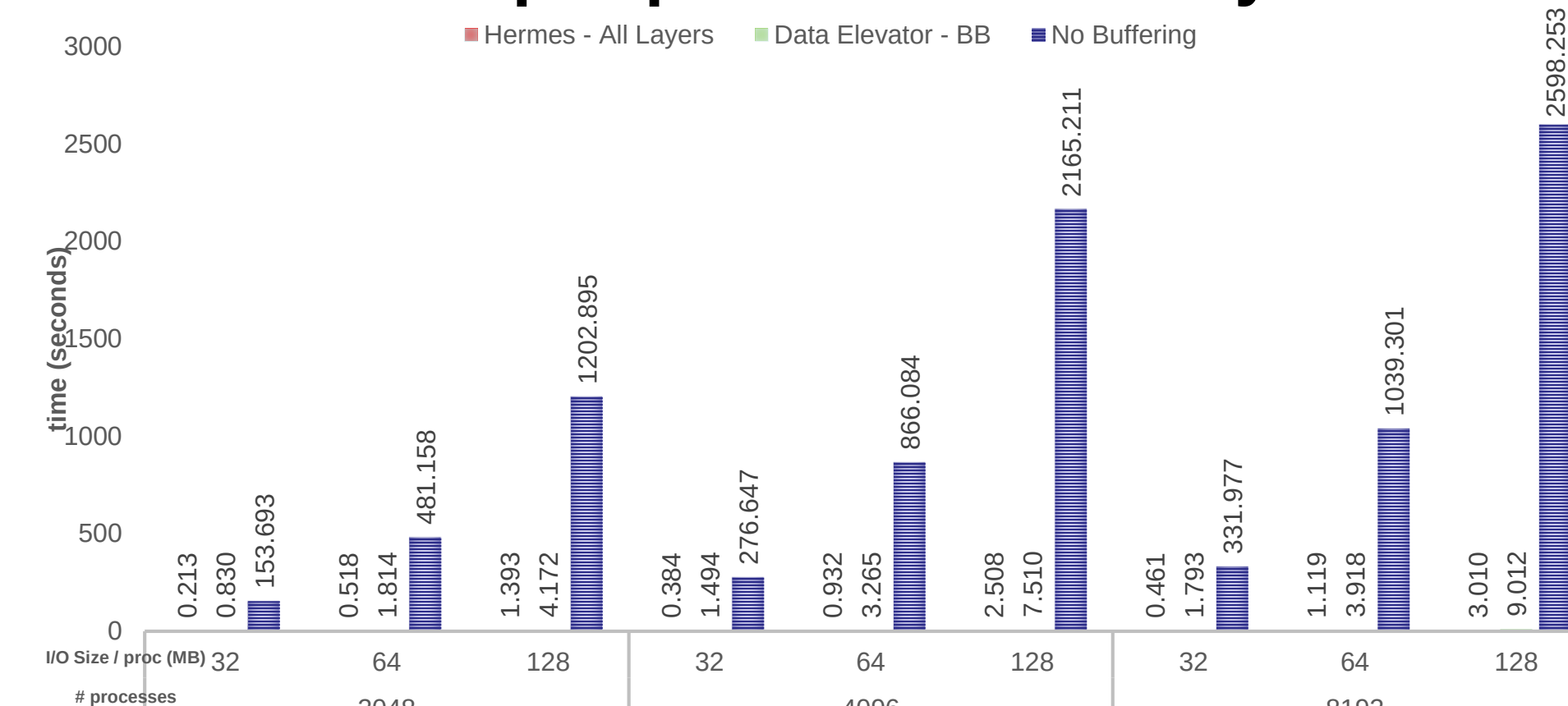


## TESTBED

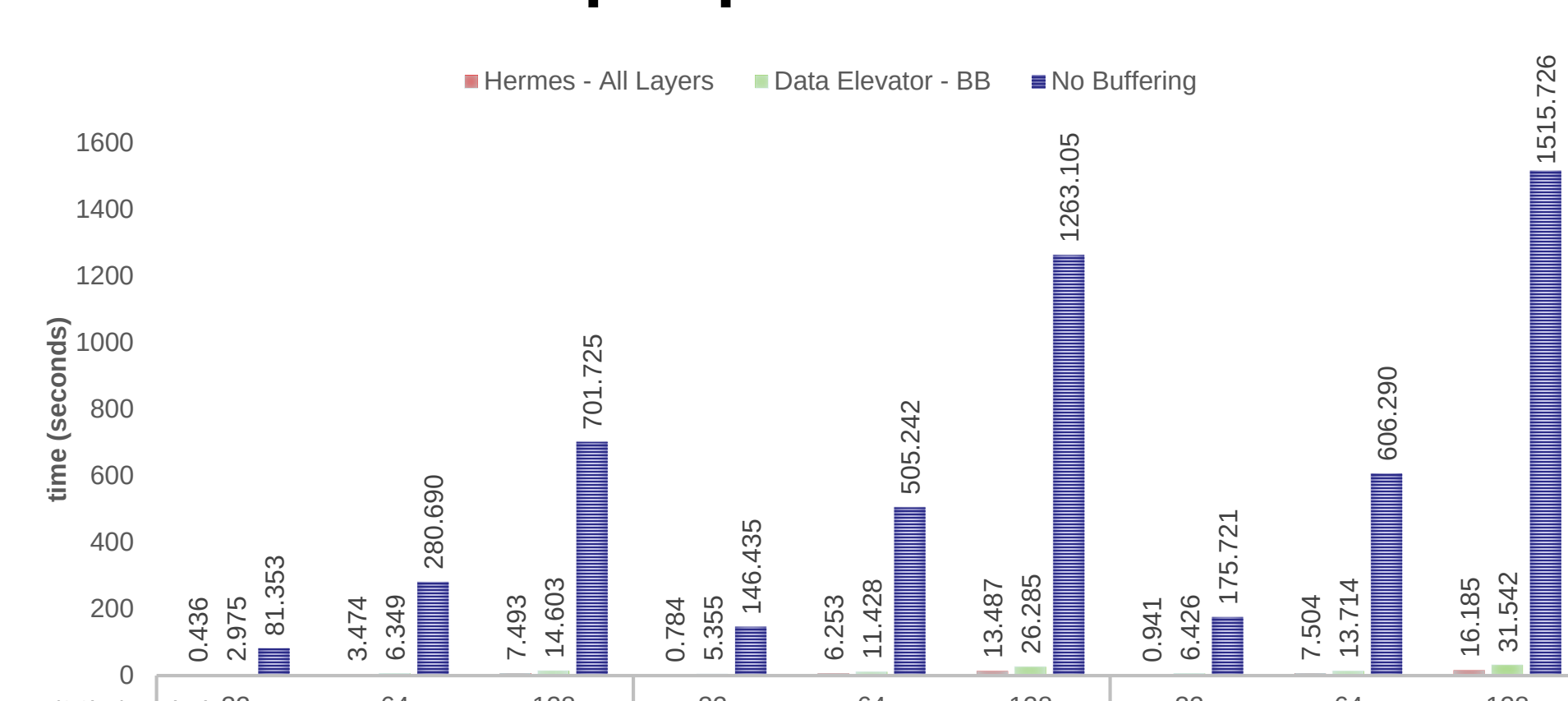
- Hardware:** NERSC Cori
- Configuration:** 128 Haswell nodes
- RAM:** 128 GB per node
- File System:** Lustre
- Burst Buffer:** Cray Datawarp
- Comparison:** LBNL Data Elevator

## BENCHMARKS

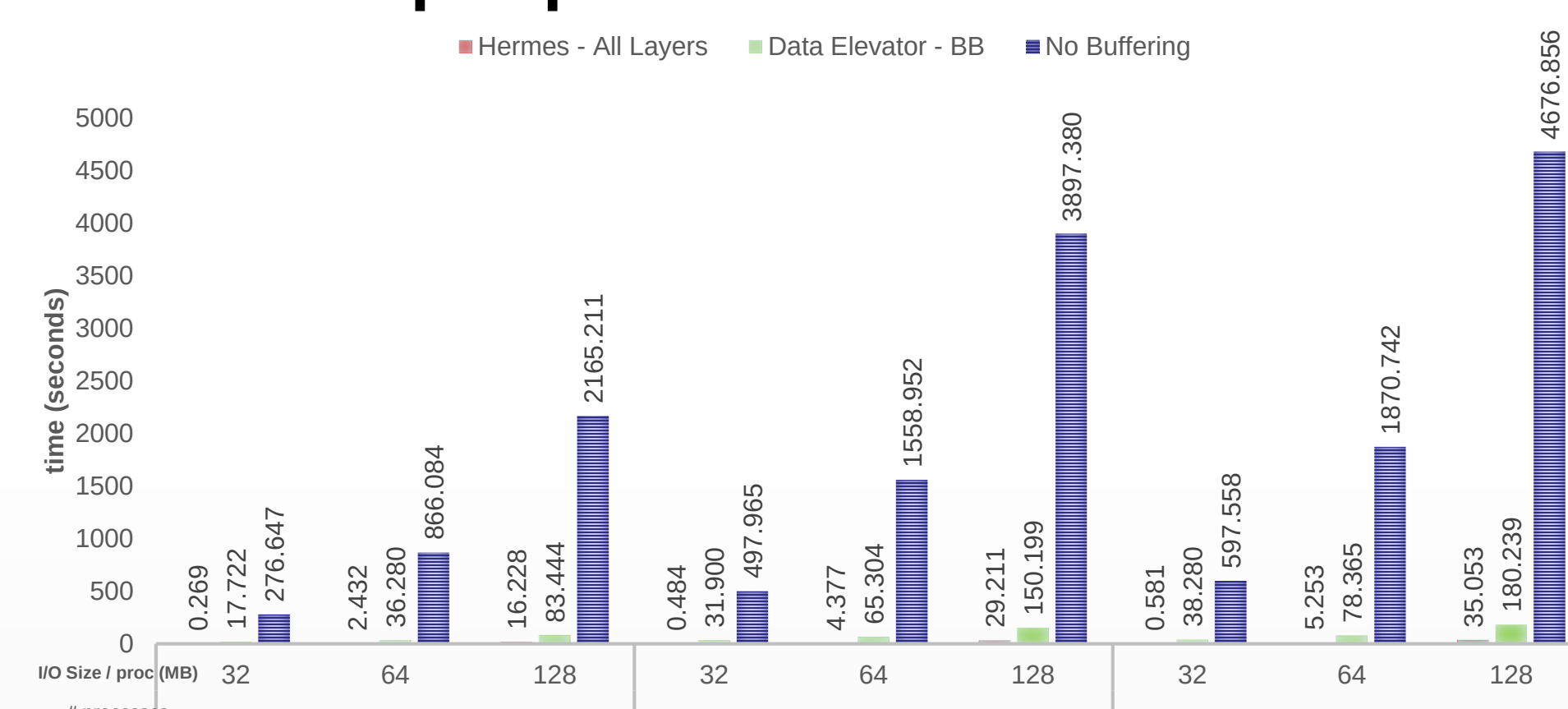
## File-per-process Write-only



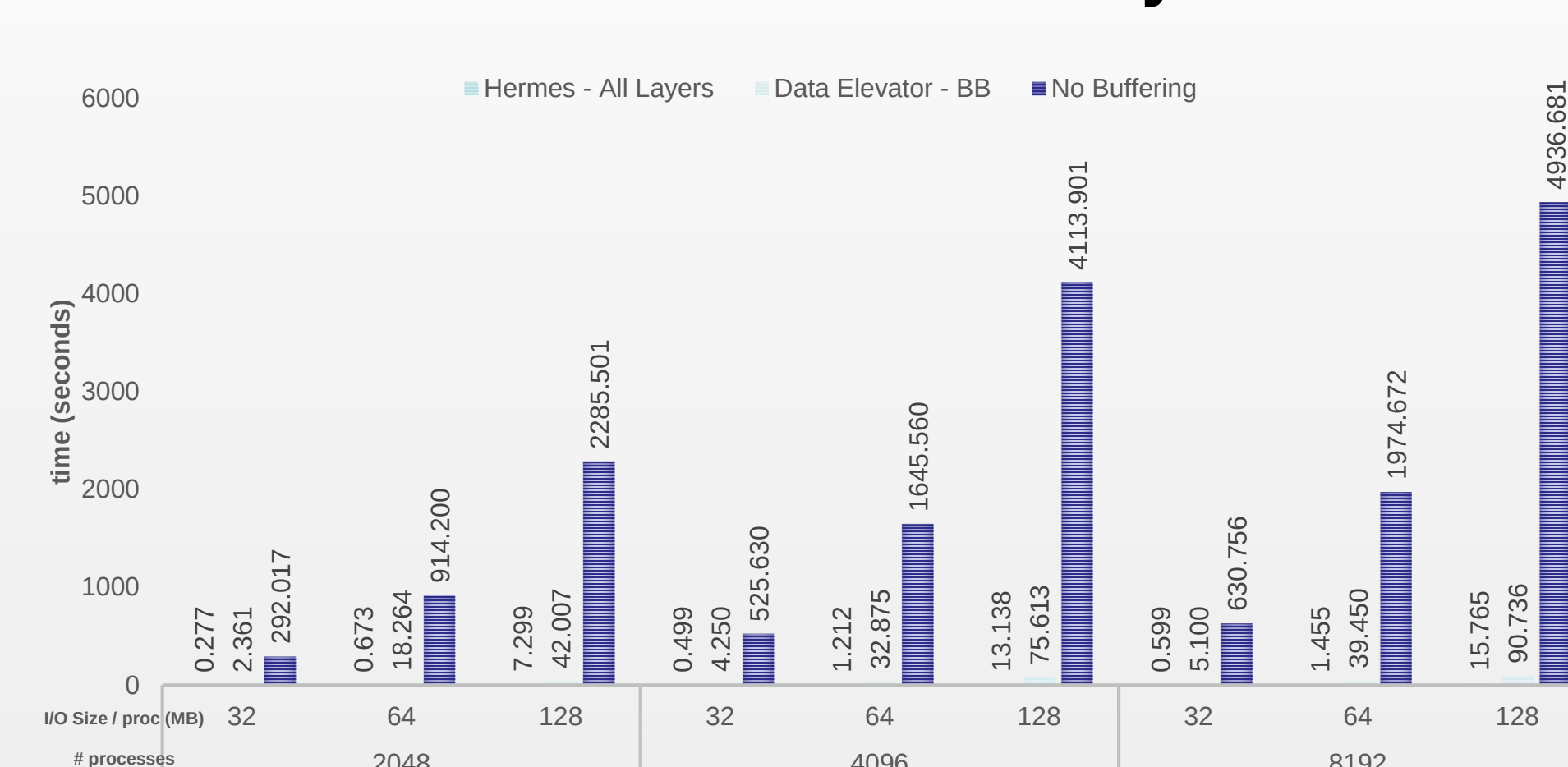
## File-per-process Mixed



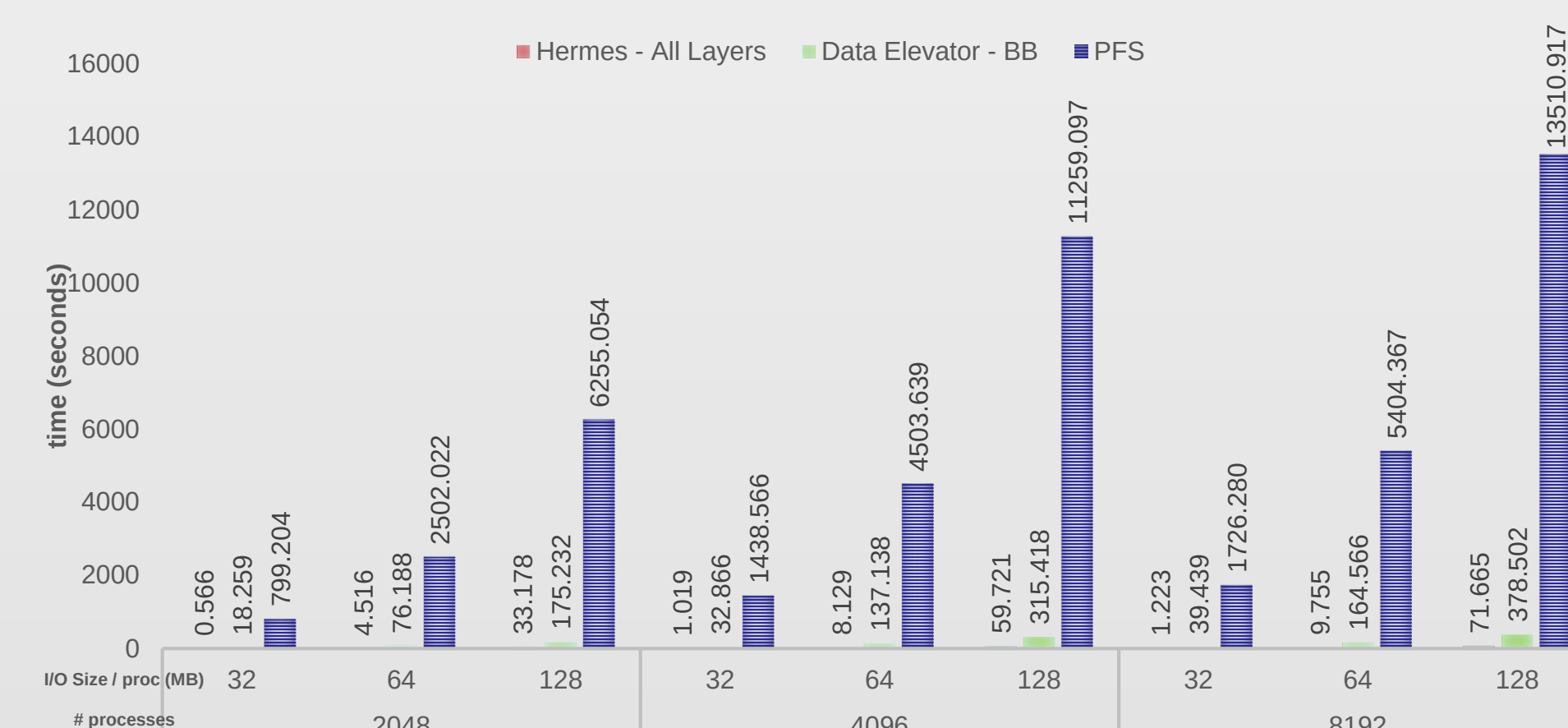
## File-per-process Read After Write



## Shared file Write-only

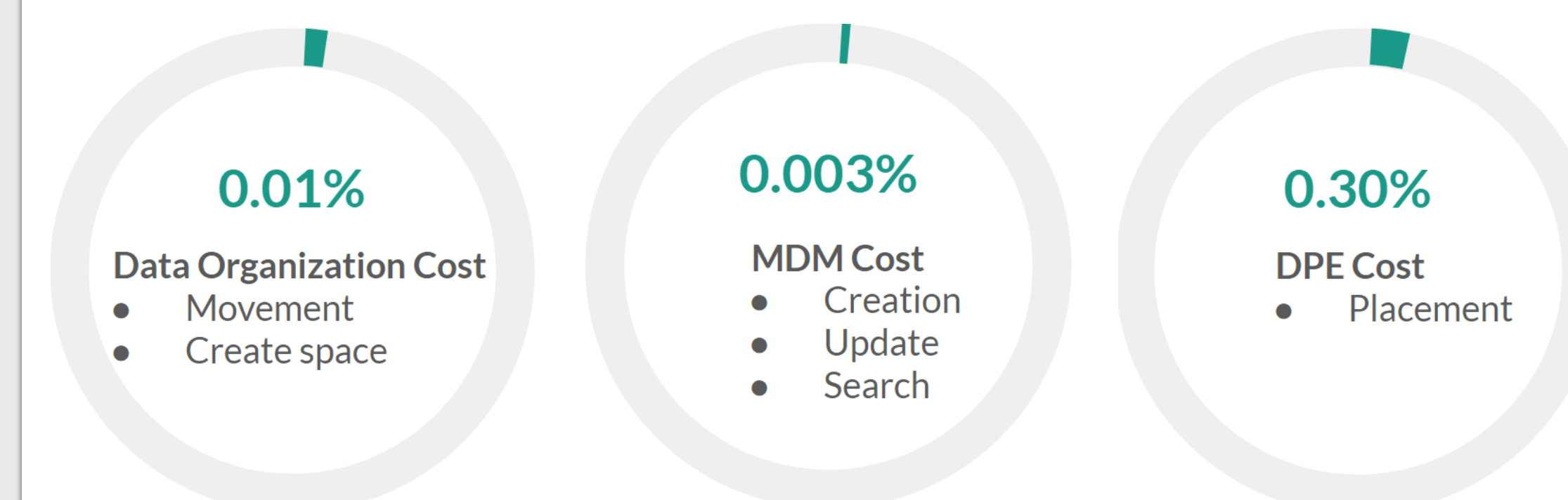


## Shared file Mixed



## LIBRARY OVERHEADS

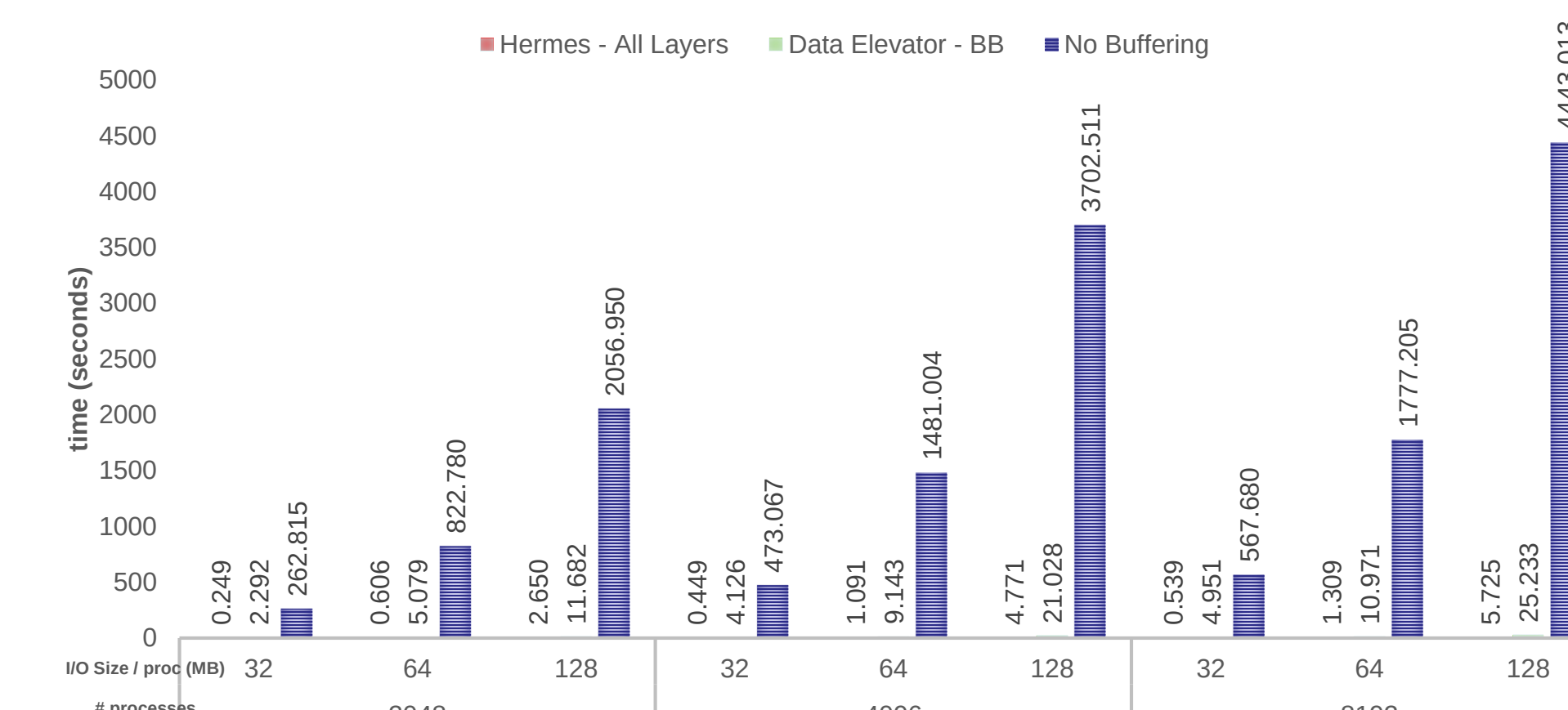
- 1 Million requests
- 1 MB I/O Size per request
- 8096 processes



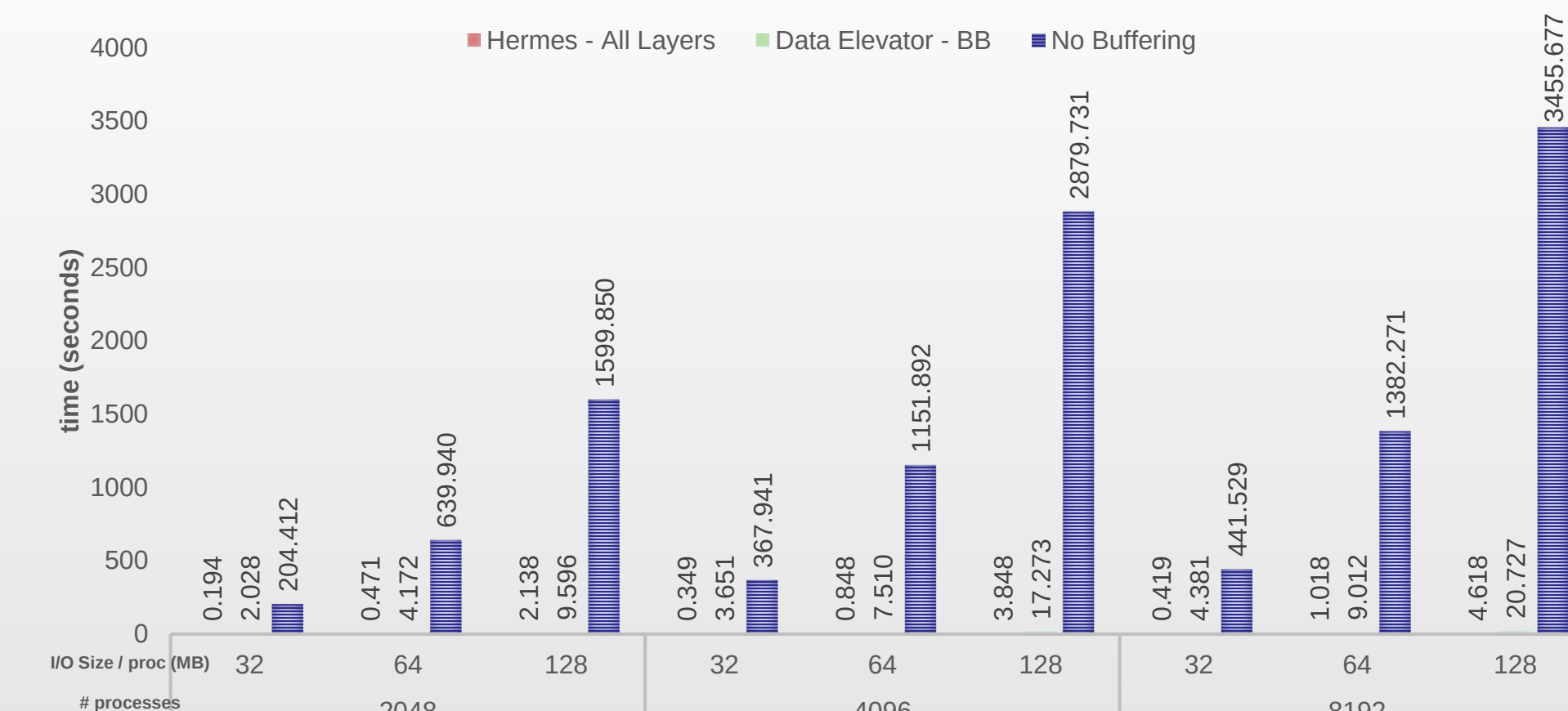
Reporting percentage of library overhead with respect to that of I/O on RAM

## APPLICATIONS

## VPIC



## FLASH



## CONCLUSIONS

- Hermes enables, manages, and supervises I/O buffering into the DMSH and offers a buffering platform that can be application- and system-aware.
- Hermes aims to maximizing productivity, increasing resource utilization, abstracting data movement, maximizing performance, and supporting a wide range of scientific applications and domains.
- Hermes outperforms existing buffering platforms by more than 62x.