University of Tsukuba | Center for Computational Sciences

Software Researches for Big Data and Extreme-Scale Computing

Gfarm/BB – Gfarm File System for Node-local burst buffer

http://oss-tsukuba.org/en/software/gfarm

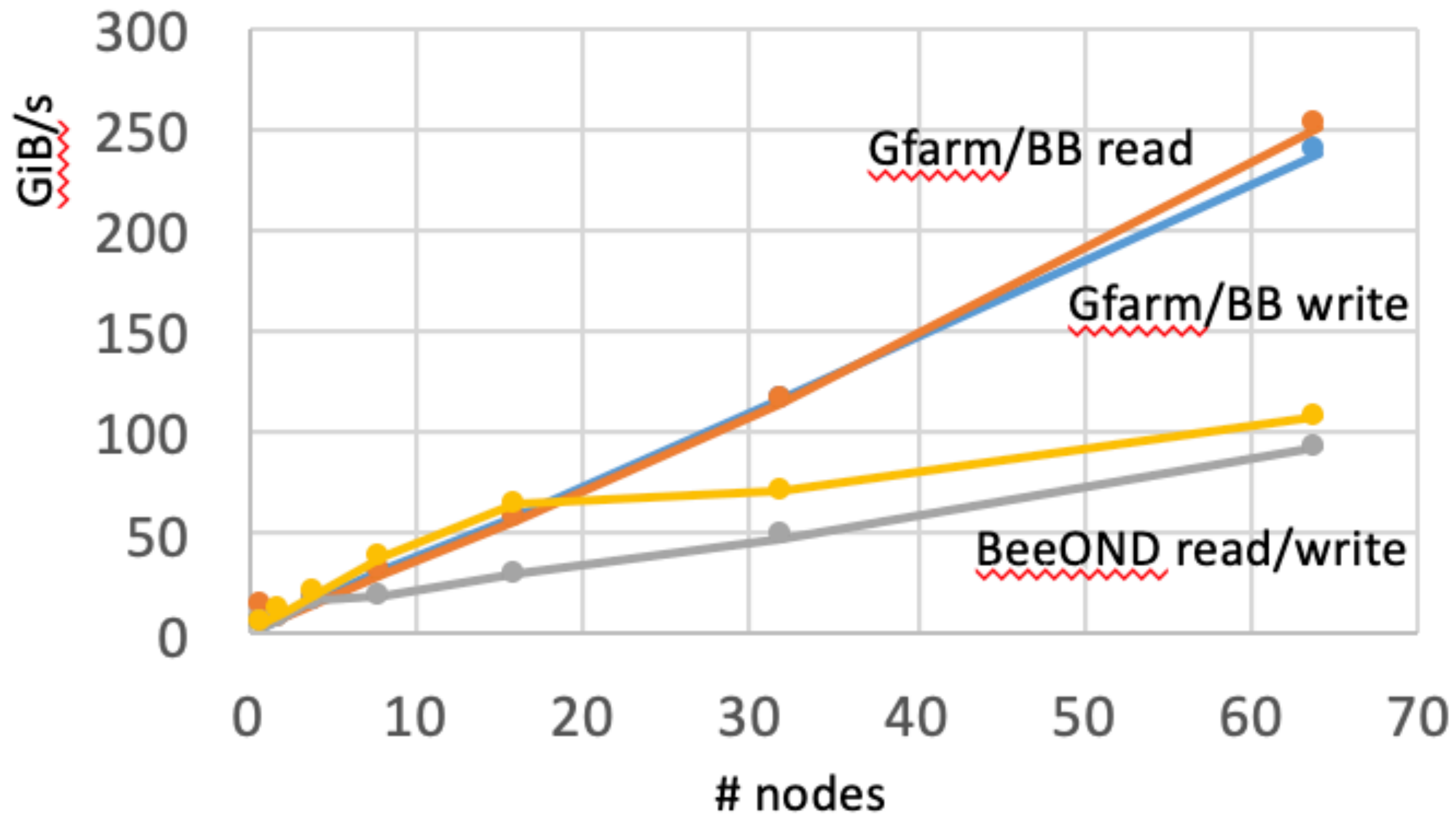

Fig. 1: IOR file-per-process read/write performance on Cygnus supercomputer

```
gfarmbb –h hostfile –m mount_point start
…
gfarmbb –h hostfile stop
```

Features include

- Open source
- Exploit local storage and data locality for scalable I/O performance
- InfiniBand support
- Data integrity is supported for silent data corruption
- Production systems: 8PB JLDG, 100PB HPCI Storage, etc.

Accelerating Python Applications with Persistent Memory

Python is one of the most popular general-purpose programming languages, and persistent memory (PMEM) is a new device which can accelerate data-intensive computing. There is a strong demand to use persistent memory from Python easily. Therefore, we focus on pmemkv, which is a key-value store optimized for persistent memory, and its python bindings. We are currently evaluating pmemkv’s python bindings in detail for efficient use of PMEM in Python.

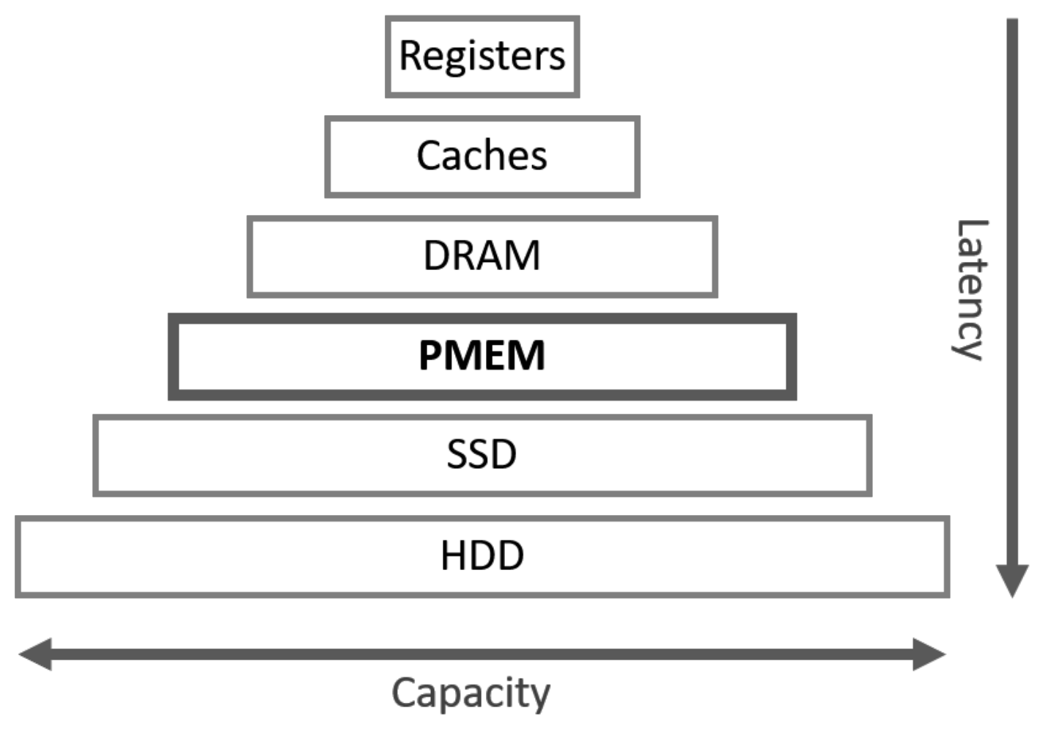

Fig.2a: Memory-storage hierarchy with persistent memory

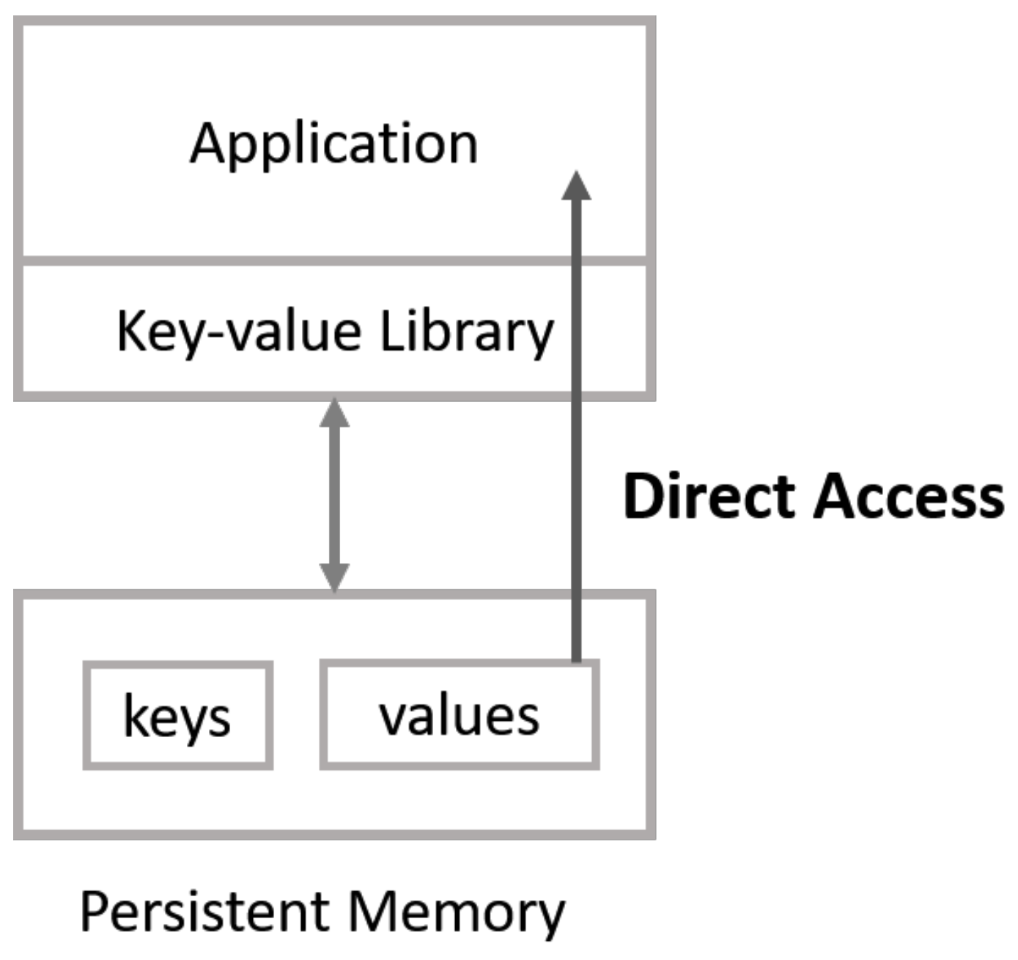

Fig2b: Applications can directly access the persistent memory resident data structures without using buffers.

Investigate DAOS architecture for metadata operation

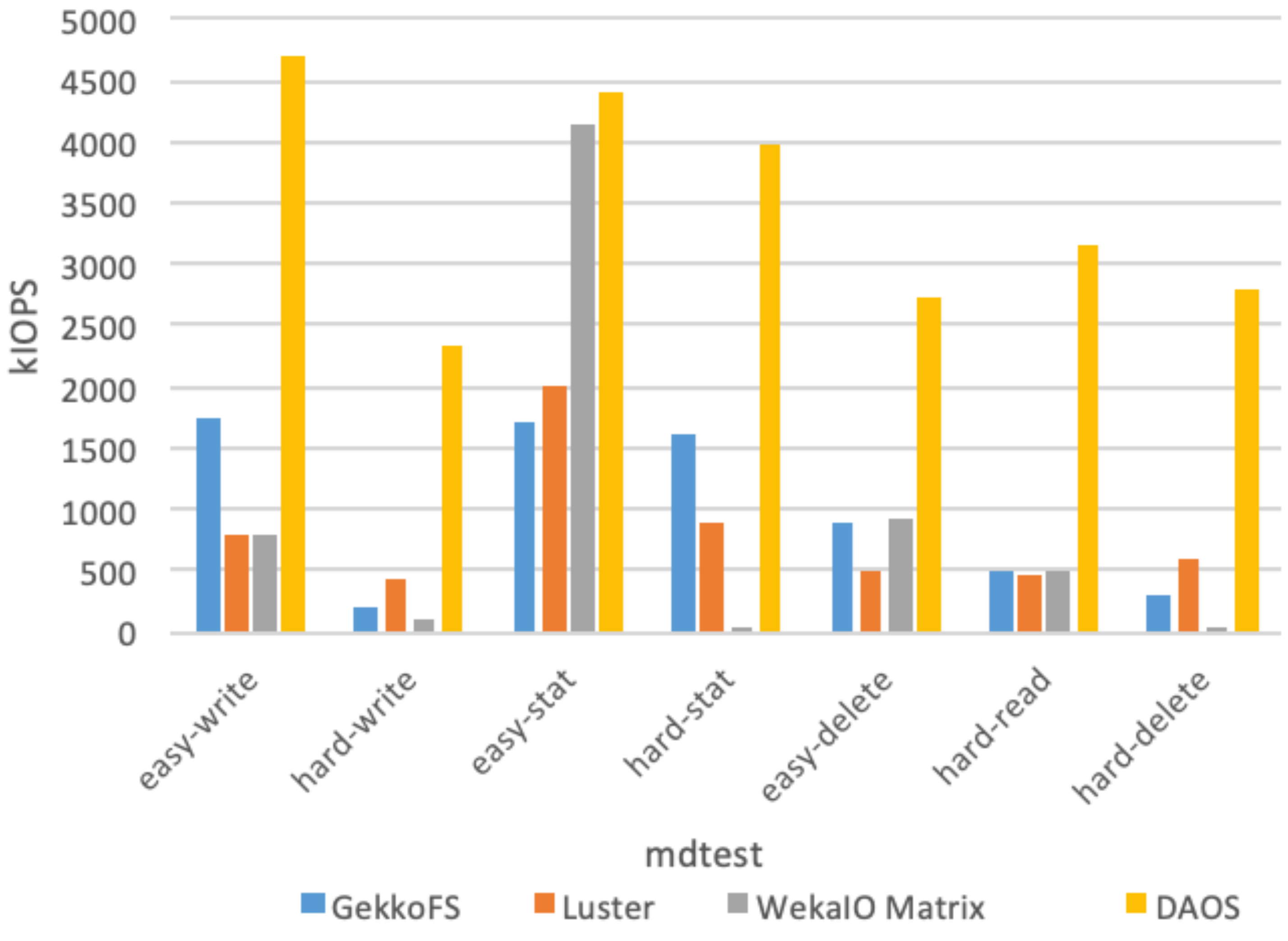

Fig. 3: mdtest performance comparison of IO-500 10 node challenge scores

The open-source DAOS – Distributed Asynchronous Object Storage – is notable for its rank on the IO-500 list and its use of Intel® Optane™ Persistent Memory. In particular, metadata performance is remarkable compared to other systems.

We investigate the reason for DAOS remarkable metadata performance on its architecture and consider to integrate DAOS ways to an existing system or develop a new storage system with persistent memory.

Research of caching file system to exploit node local storages

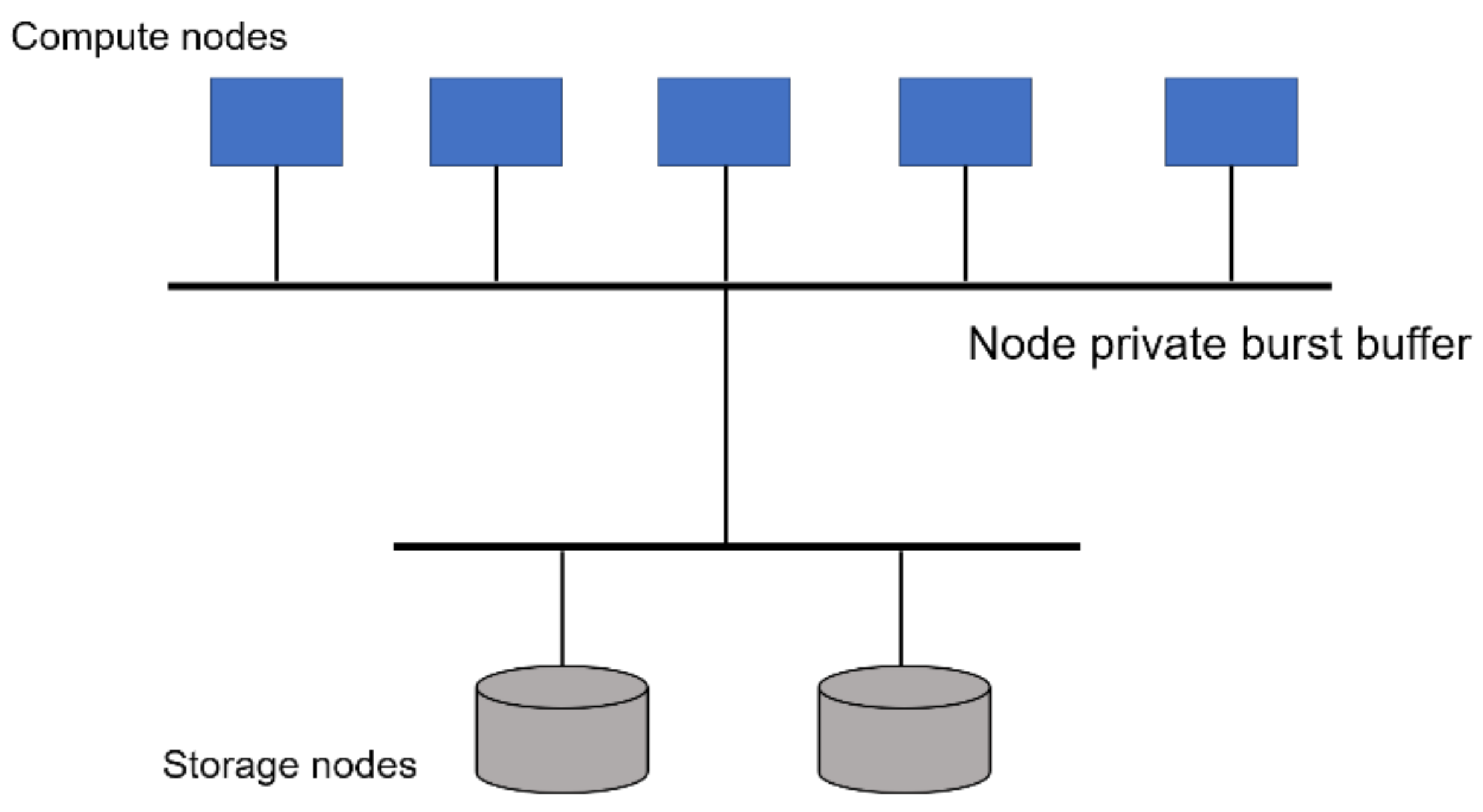

Fig. 4: Automation of construction/destruction a swarm cluster

The performance gap between processors and disk-based storage is growing in modern HPC systems. To reduce the gap, SSDs attached to compute nodes has been used as a “node local burst buffer”. We are implementing distributed file system that uses local SSDs as a caching layer of the storage nodes. The system uses fuse-library for system call replacing and mochi-framework for RPC data transfer.

Acceleration of Deep Learning using pytorch with persistent memory

Persistent memory offers greater capacity than DRAM and significantly better performance than storage. We use it for deep learning with pytorch. Usually, before performing deep learning using the GPU, the training data is copied to the main memory from the storage. We exploit the persistent memory to improve the performance.

Acknowledgment
This work is partially supported by Multidisciplinary Cooperative Research Program in CCS, University of Tsukuba, New Energy and Industrial Technology Development Organization (NEDO), and Fujitsu Laboratories Ltd.

contact address: pr@ccs.tsukuba.ac.jp

https://www.ccs.tsukuba.ac.jp/