



Benchmarking High Performance File Systems

Eshan Bhargava

Luca Atzori, Krzysztof Michal Mastyna, Joaquim Santos

14/09/2022

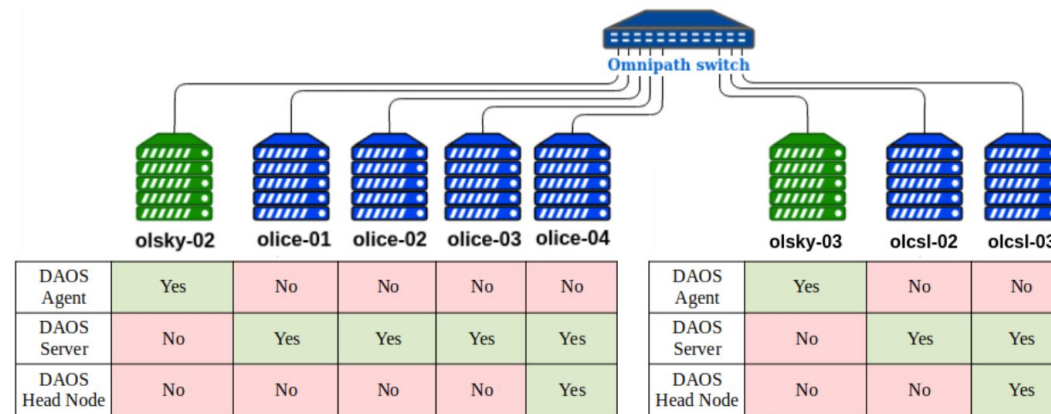
Introduction

- Artificial Intelligence and High Performance Computing have caused exponential data growth.
 - As the scale of processing systems increase, the data access from storage systems needs to increase commensurately.
- Companies are quickly finding solutions to the increase in data by developing new storage systems.
- In this project, we benchmark two installations of the DAOS file system in order to see its viability for future use cases.

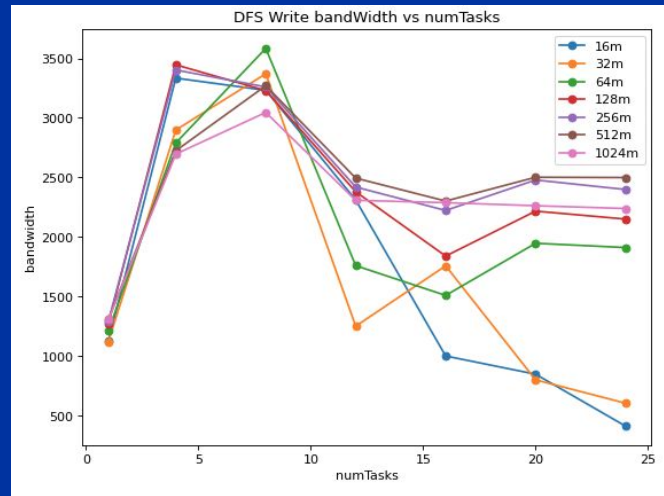
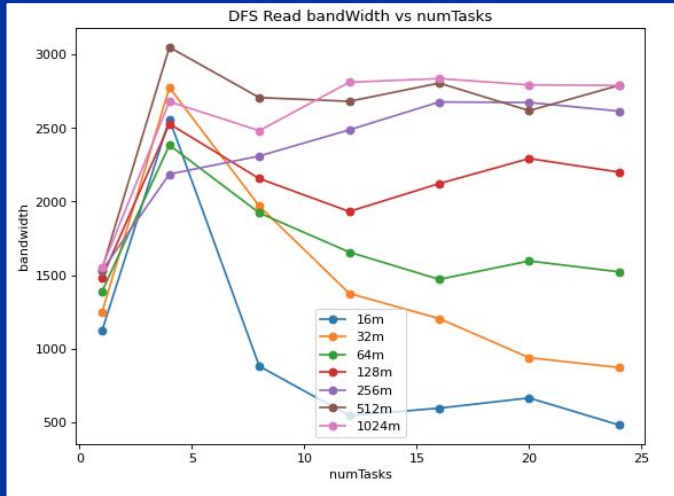
DAOS: Distributed Asynchronous Object Storage

- DAOS is an Intel R&D file system, which has great potential for the use cases at CERN and tech companies.
 - Overcomes limitations of traditional distributed storage
 - Open source
 - Software-defined scale-out object store
 - Low latency
 - High message rate user space communications that bypass the Operating System

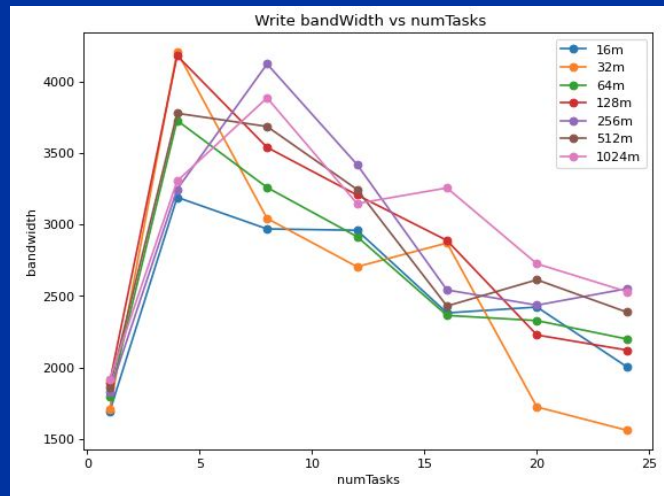
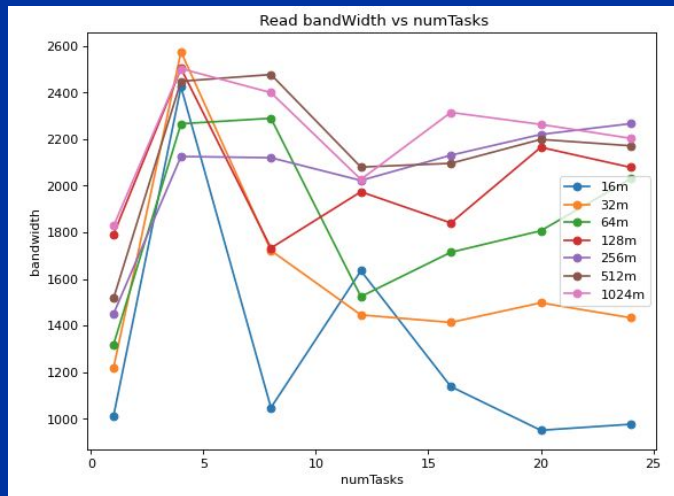
- DAOS installation CERN
 - Ice Lake Cluster
 - Cascade Lake Cluster



Ice Lake Cluster



Cascade Lake Cluster



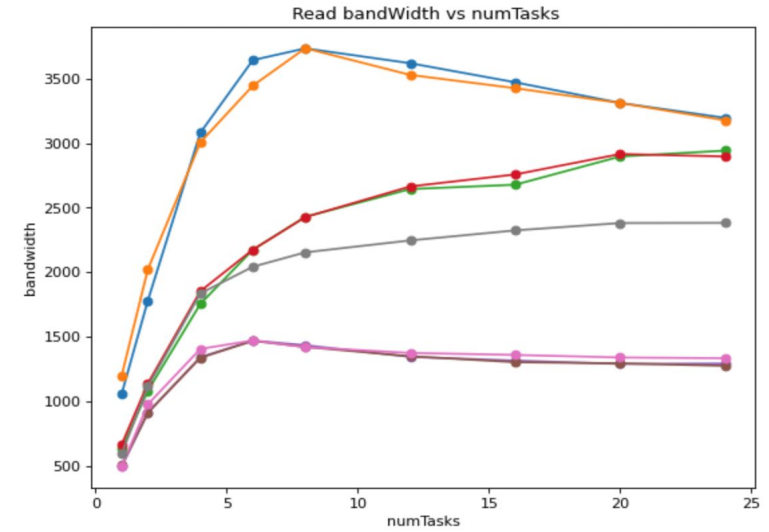
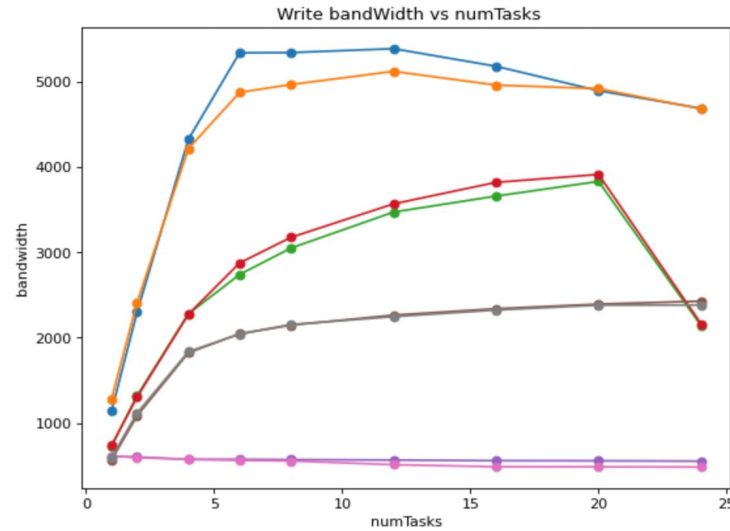
IOR Benchmark

Initial DFS Results

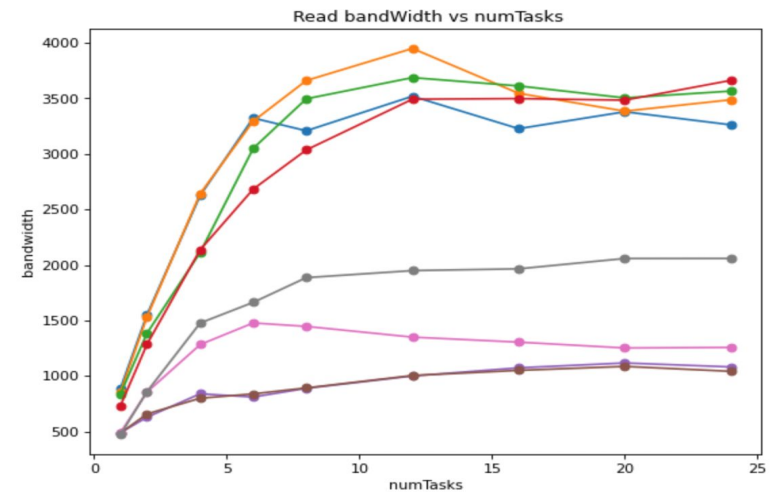
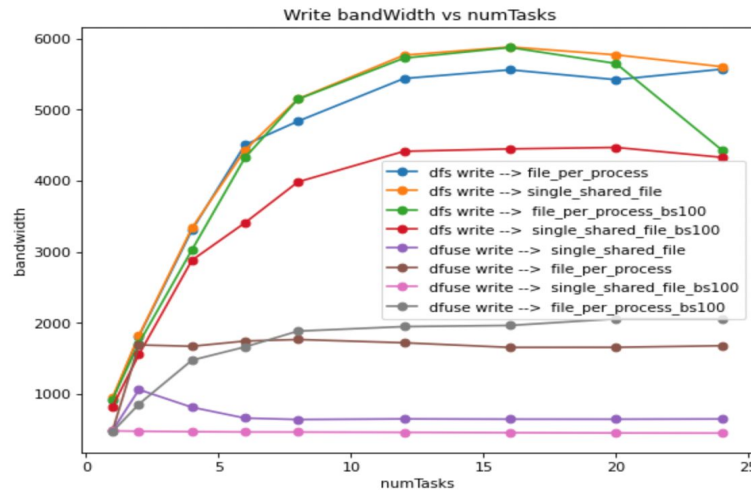
- IOR Measures Latency and Throughput.
- Can specify block size, transfer size, and num tasks (also segments).
- Trial and error with IOR command flags caused delays.

Revised IOR Results

Ice Lake Cluster



Cascade Lake Cluster



Ice Lake Cluster

Write

Read

chep21_summary_gen_lhcb-vPSfCS-zstd		
Page Size	OC_SX	RP_XSF
65536	2.12981	2.12560
524288	2.63672	2.60660
1048576	2.76273	2.76357
4194304.0	2.87566	2.89805

chep21_summary_lhcb-vPSfCS-zstd		
Page Size	OC_SX	RP_XSF
65536	1.18197	1.23968
524288	1.67102	1.77391
1048576	1.97760	1.94331
4194304	2.13033	2.09163

Cascade Lake Cluster

Write

Read

chep21_summary_gen_lhcb-vPSfCS-zstd		
Page Size	OC_SX	RP_XSF
65536	2.15162	2.17508
524288	2.63608	2.56158
1048576	2.82797	2.75375
4194304.0	2.68813	2.73516

chep21_summary_lhcb-vPSfCS-zstd		
Page Size	OC_SX	RP_XSF
65536	1.09623	1.28202
524288	1.64154	1.6142
1048576	1.96568	2.01815
4194304.0	2.16757	1.8203

Root LHCb Case

Zstandard Compression

- Based on writing to a DAOS pool and then reading back a partial dataset.
- This data set is taken from Open Data (LHCb Run 1, B mesons 3 hadrons) and loaded into RNTuple.

Conclusion

- Other Investigations:
 - DAOS
 - DBench
 - MDTest
 - Root LHCb benchmark, no compression
 - Ceph - no actionable results
 - Fio
 - IOR
- DAOS results:
 - The Cascade Lake Cluster outperforms the Ice Lake Cluster.
 - There could be something wrong with the hardware configuration of the Ice Lake Cluster.
 - The optimal ratio of servers to clients should mimic the Cascade Lake Cluster configuration.



Thank You!

Questions?

eshan.bhargava@cern.ch

4eshanb@gmail.com