Hewlett Packard
Enterprise

DAOS TESTING @HPE

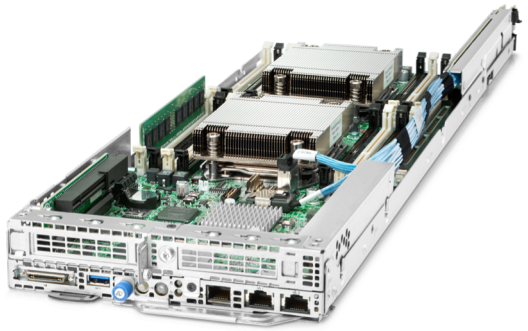
Cedric Milesi HPE

November 2019

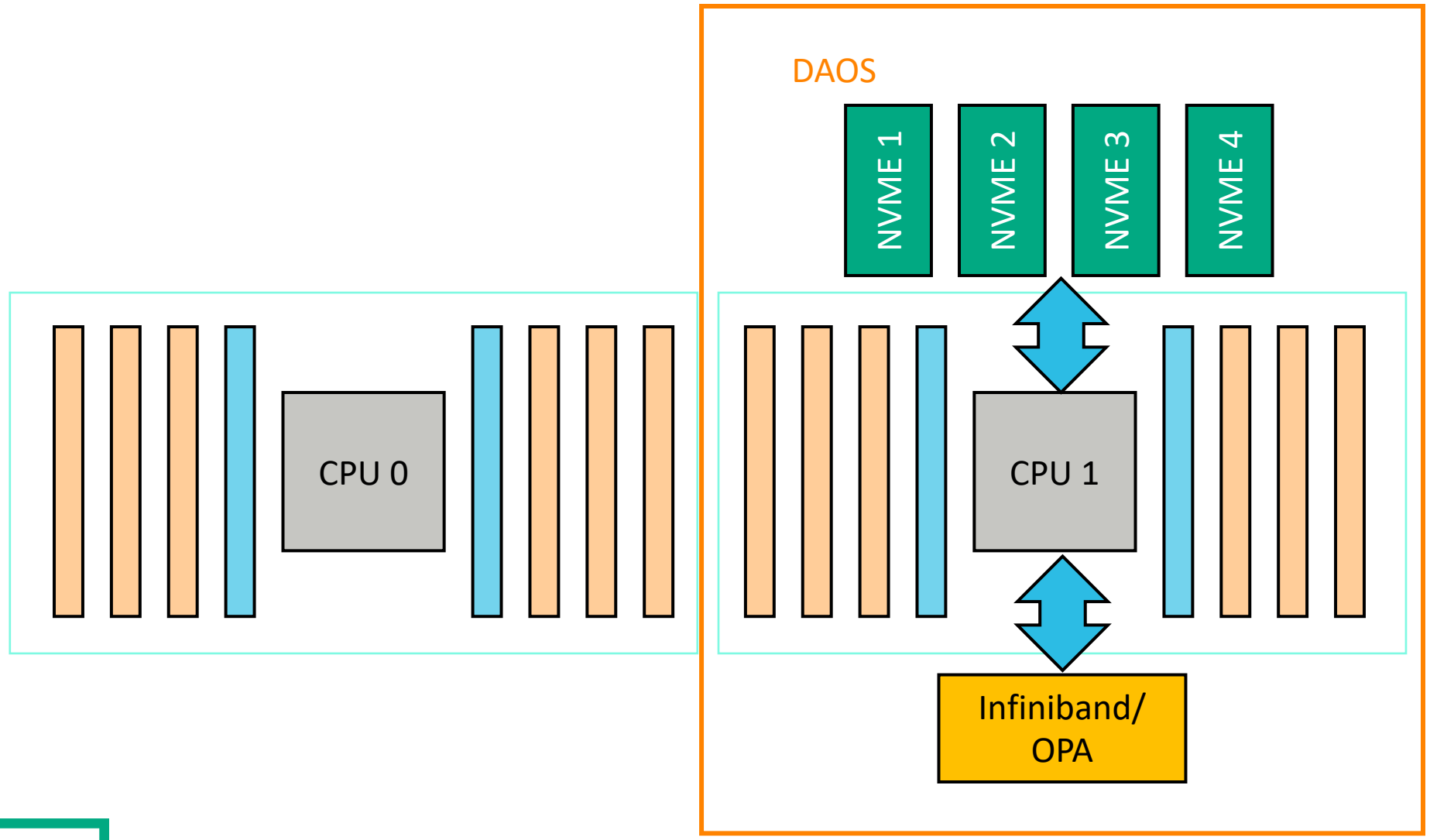
TESTBED V1

- 7x storage nodes
  - Dual socket
    - –Xeon(R) Gold 6134 @ 3.20GHz
    - –8 cores per socket
  - 96GB of DRAM
  - 4x 3.2TB P4600 SSD
    - –https://ark.intel.com/content/www/us/en/ark/products/97002/intel-ssd-dc-p4600-series-3-2tb-2-5in-pcie-3-1-x4-3d1-tlc.html
    - –R@2.8GB/s and W@1.9GB/s
  - 1x OPA card

- 8x compute nodes
  - Dual socket
    - –16x tasks max per client
      - – Limited by number of OPA contexts
  - 1x OPA card

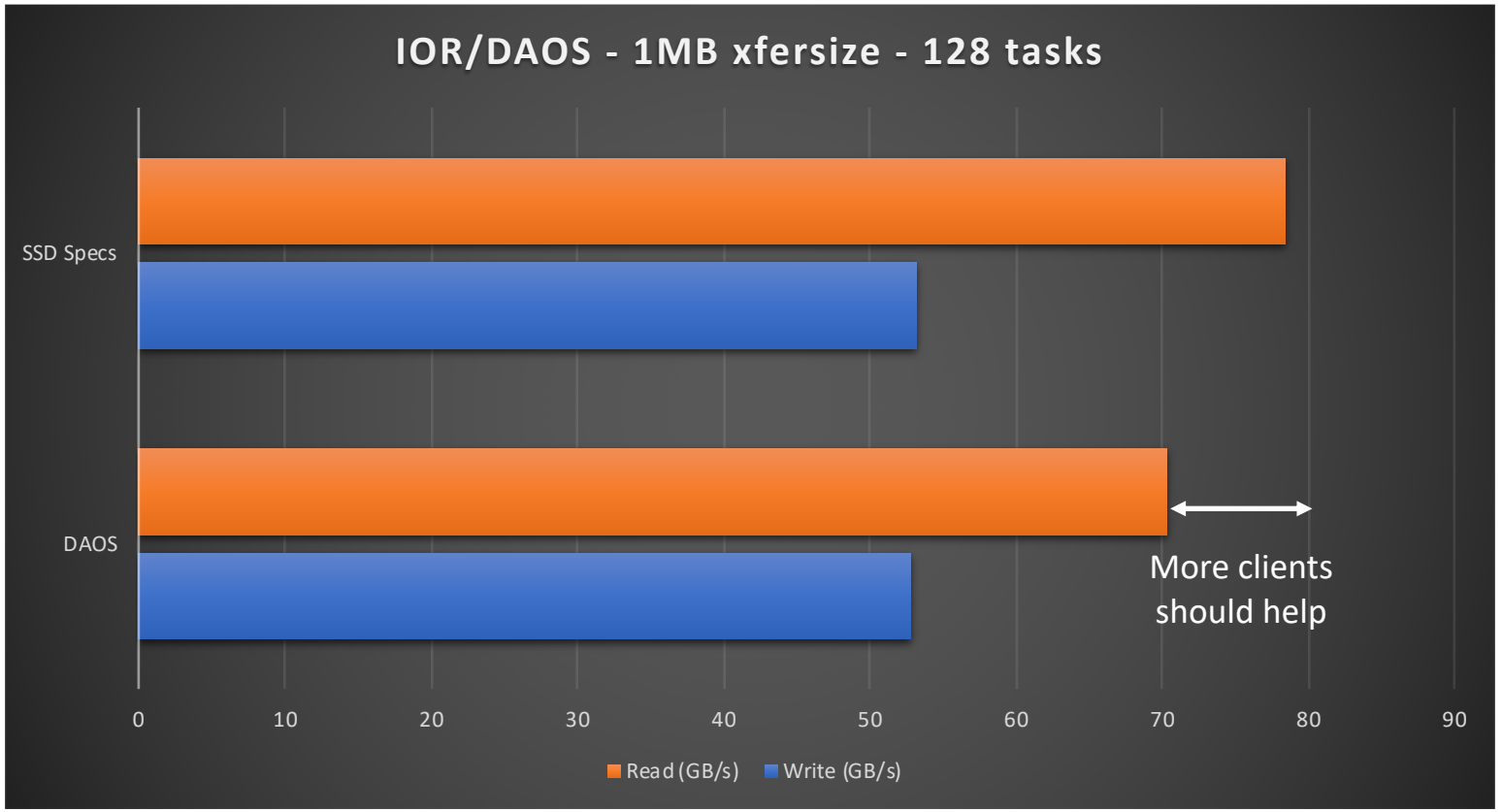


OPTIMIZING PERFORMANCE ON SERVER



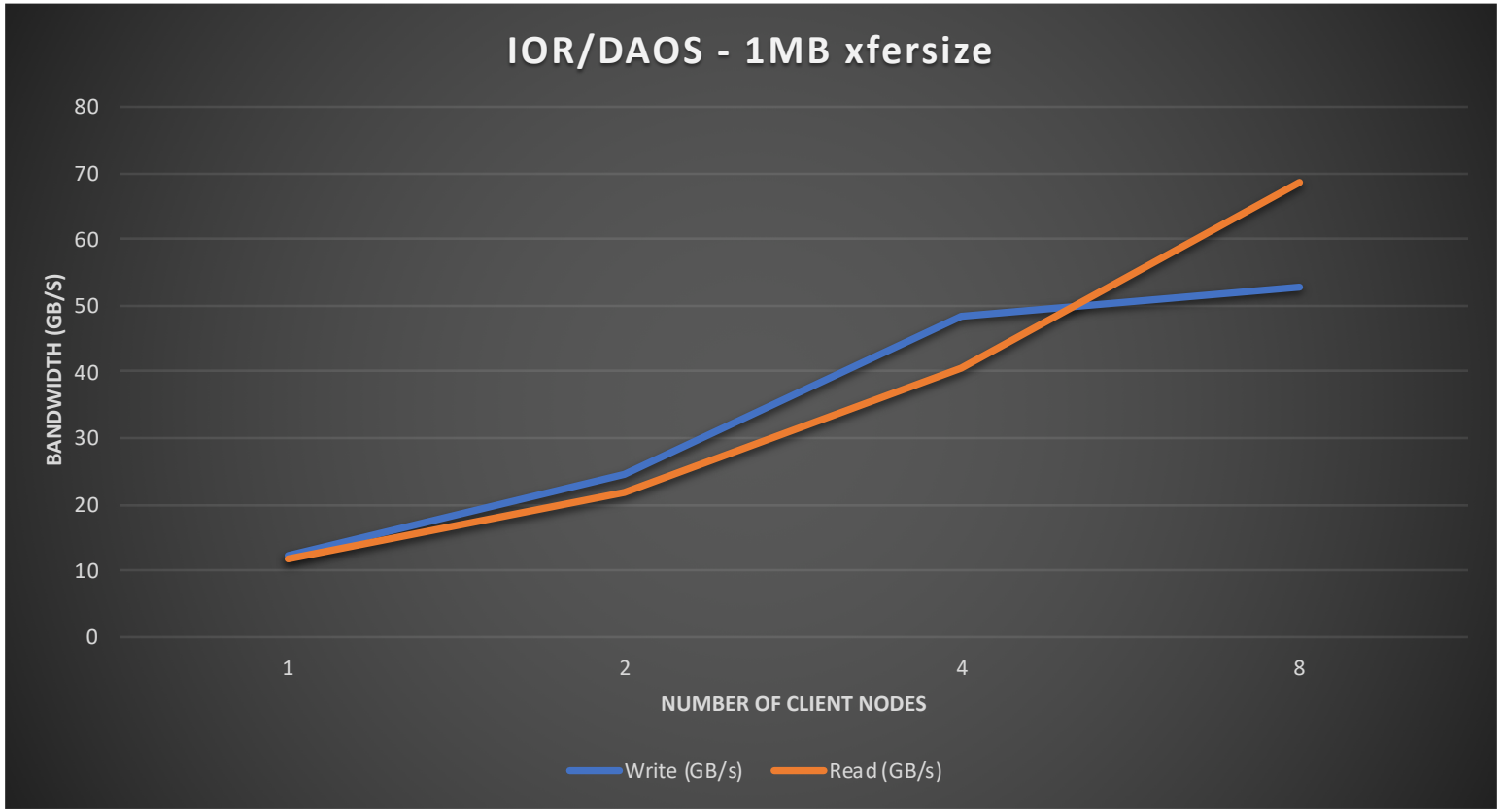
BENCHMARK RESULTS

DAOS & LARGE I/O



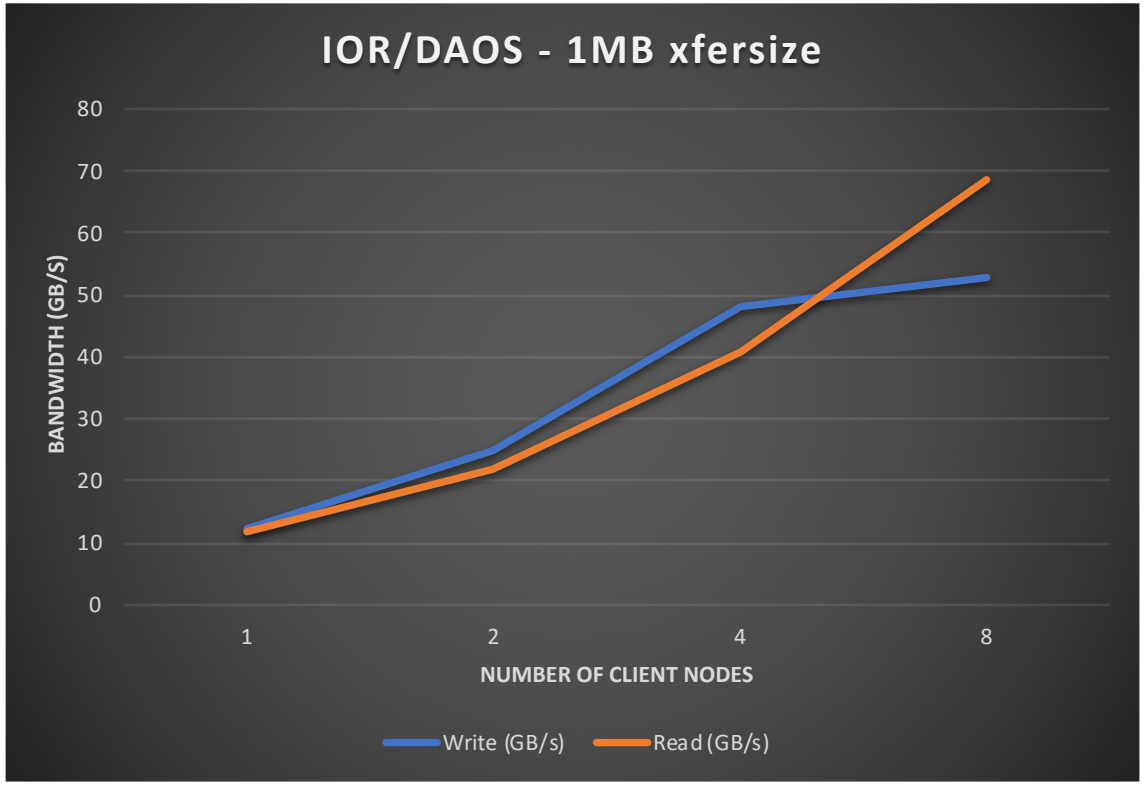
- DAOS delivers the maximum SSD bandwidth on write according to the SSD specficiations
- For reads, a few more clients would be required to saturate the SSDs.

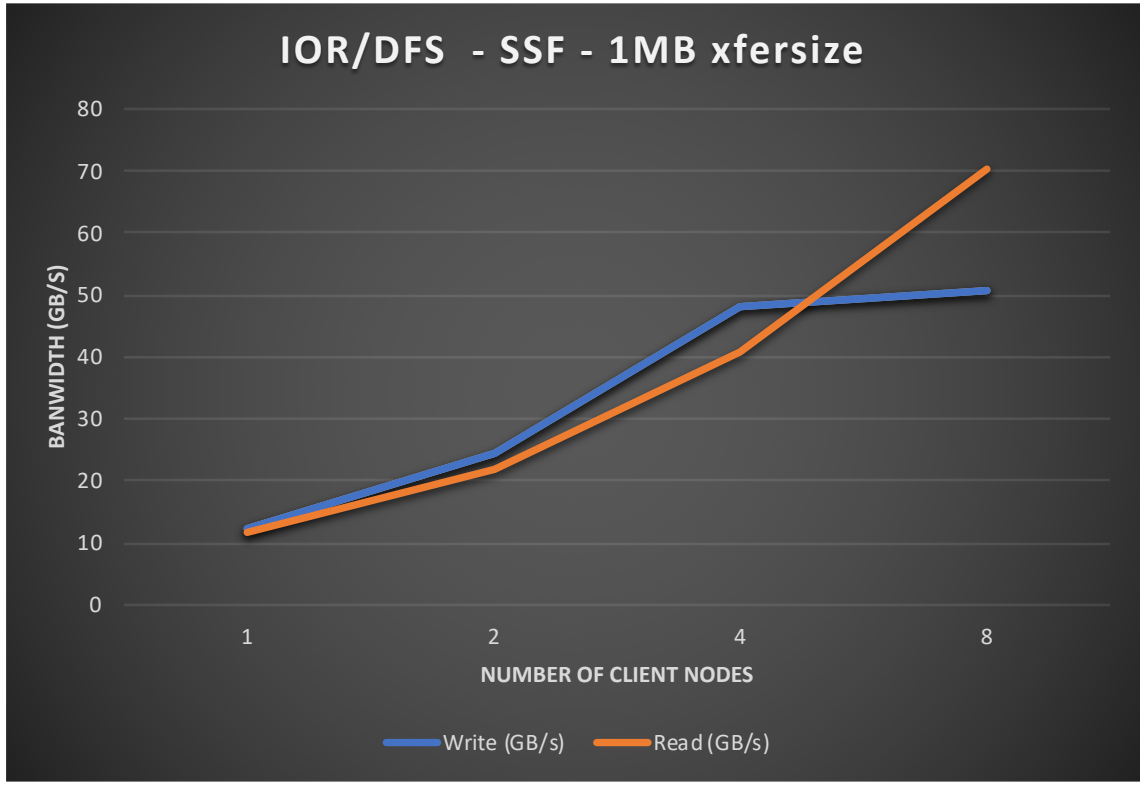
CLIENT SCALABILITY



- 12GB/s max per client due to fabric limit
- Linear scalability for both reads and writes
- Writes are limited by the fabric up to 4 client nodes, then limited by SSD bandwidth
- Reads still scale linearly and might still go higher with more client nodes

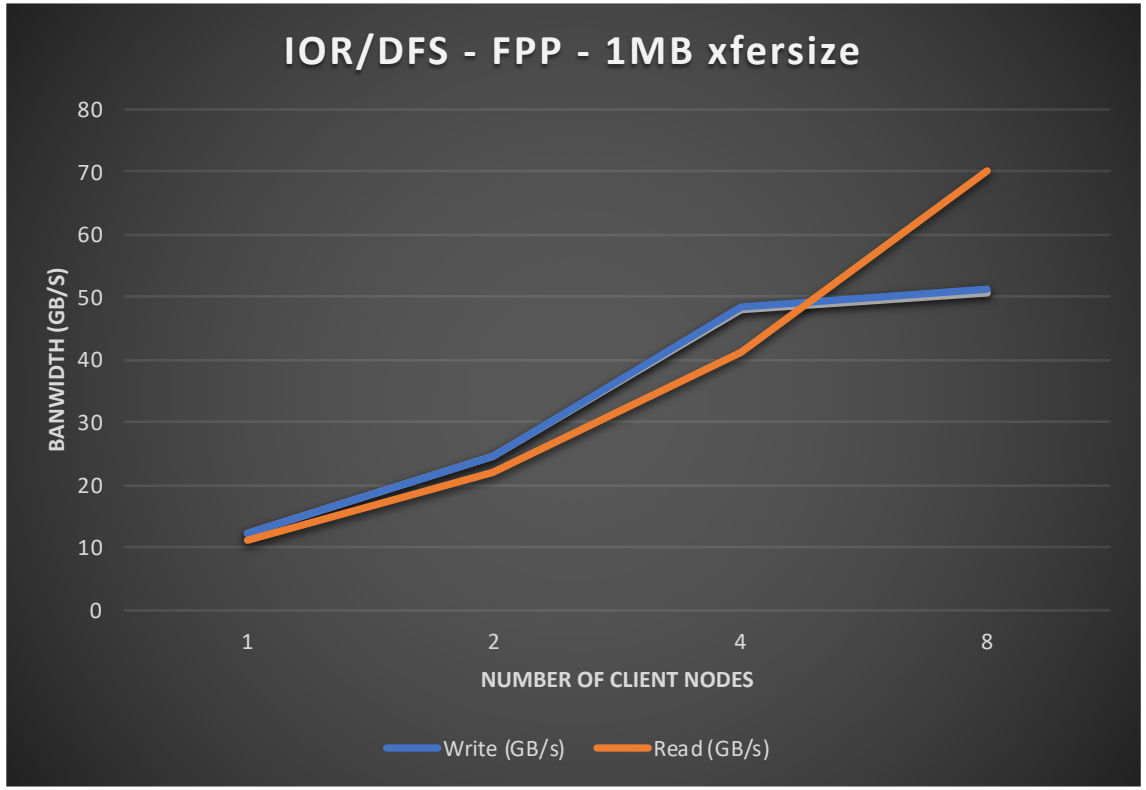
DAOS VS POSIX/DFS API

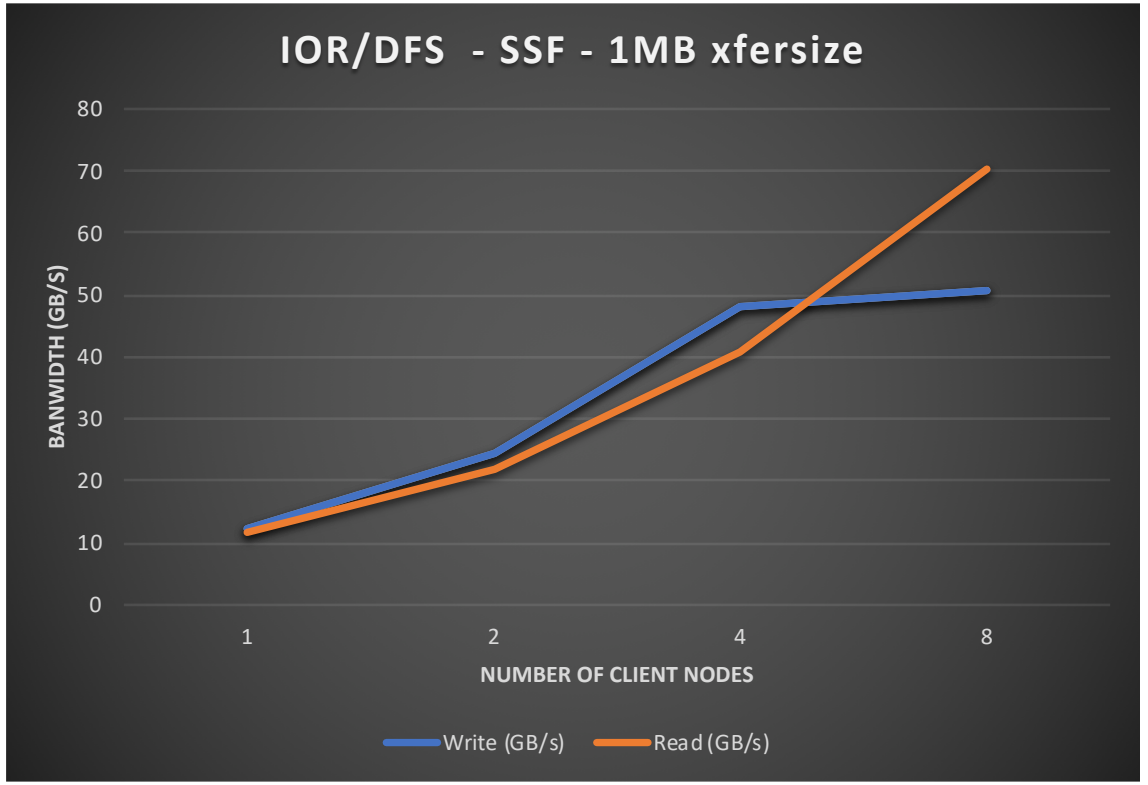




- Same performance with native DAOS API and DFS API
- Both APIs rely on native array objects with a 1-byte cell size

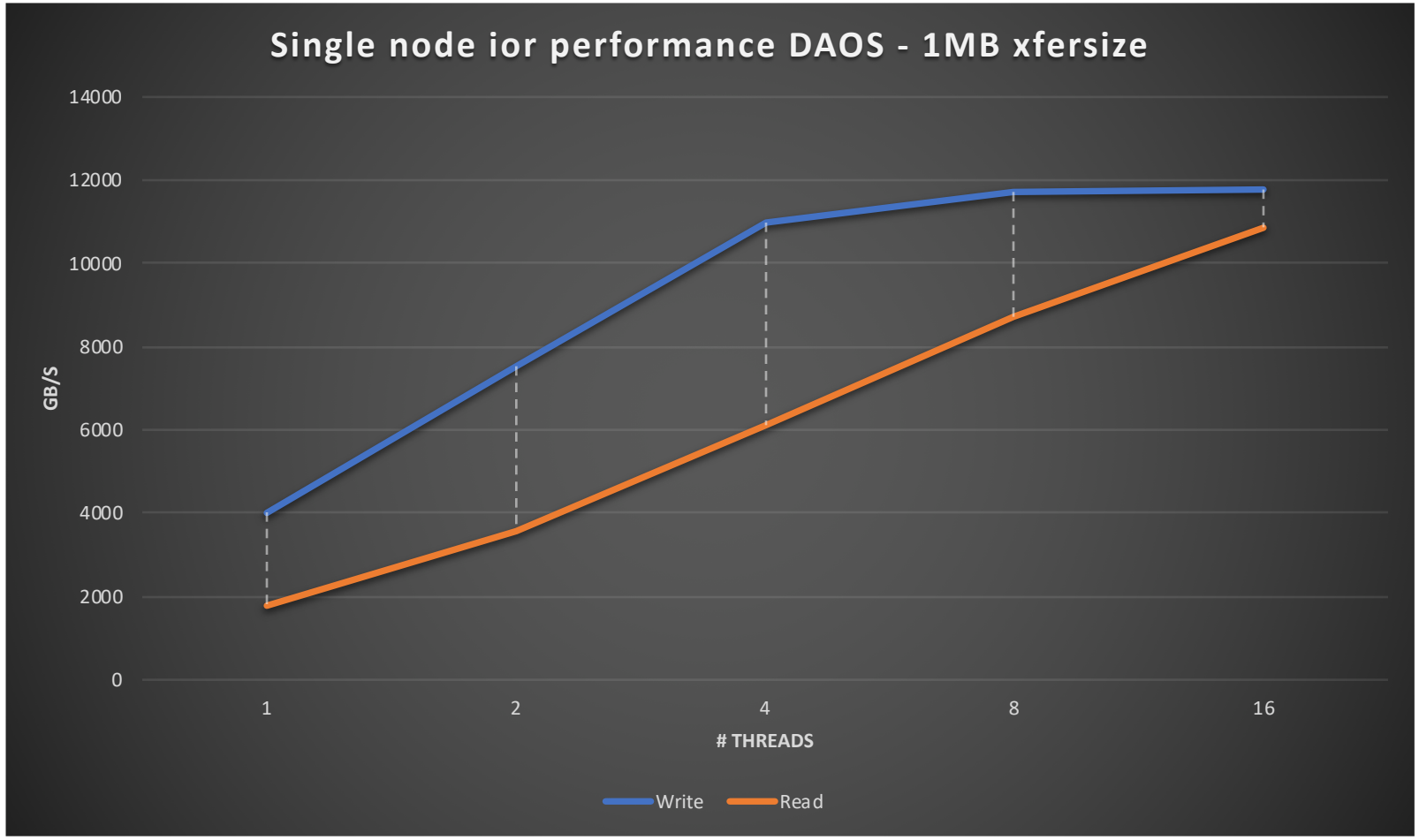
FPP VS SSF





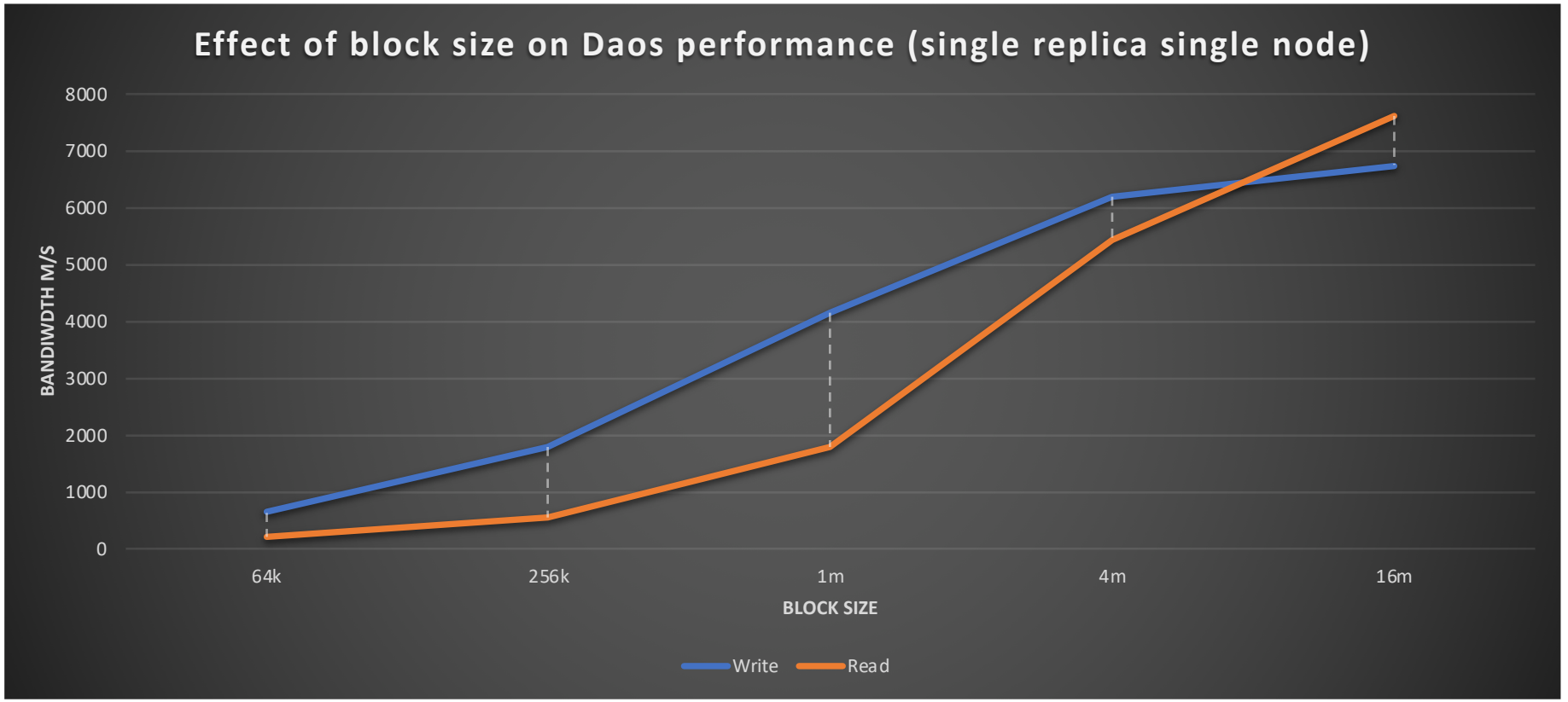
- Same performance with single-shared file (SSF) and file-per-process (FPP)

SINGLE NODE SCALABLITY



- Scale linearly with the number of tasks until client OPA link is saturated.

SINGLE TASK PERFORMANCE WITH DIFFERENT BLOCK SIZE

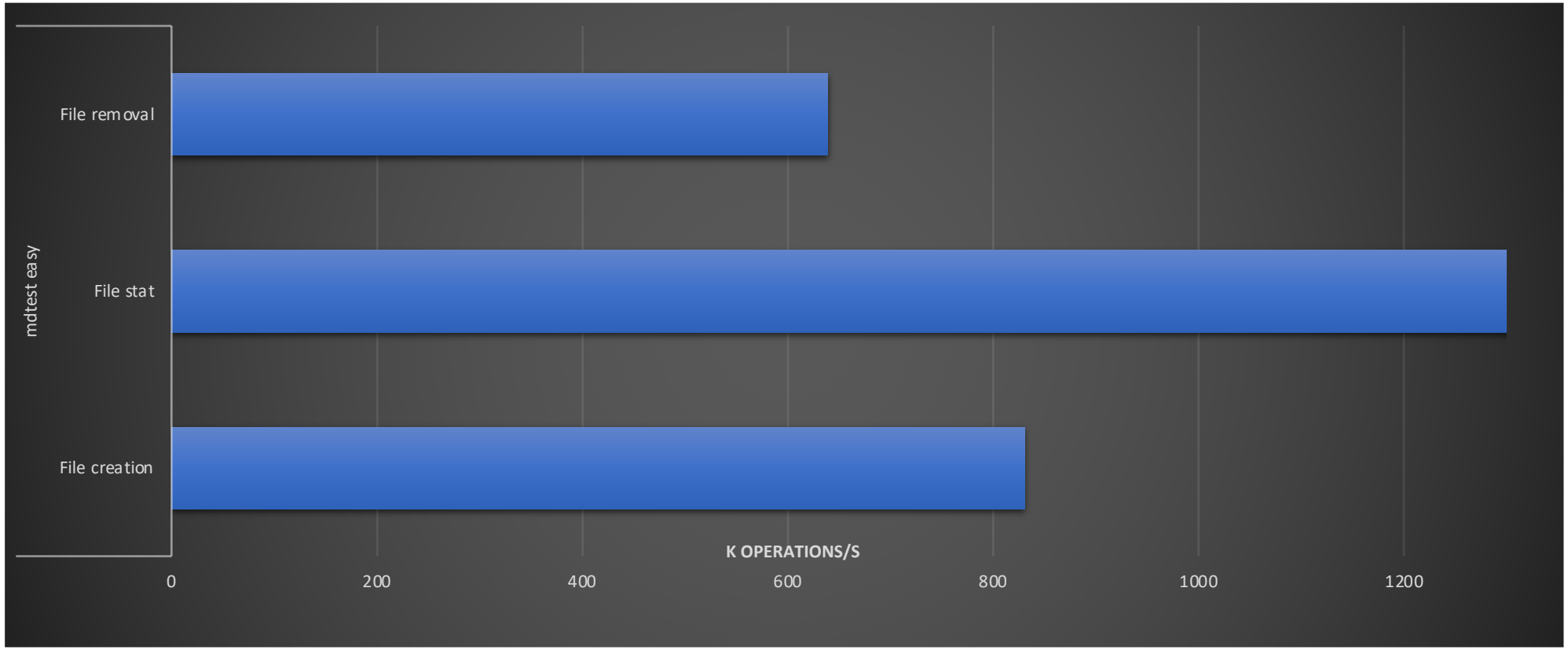


- Up to 8GB/s with a single process! Could be very useful for some applications

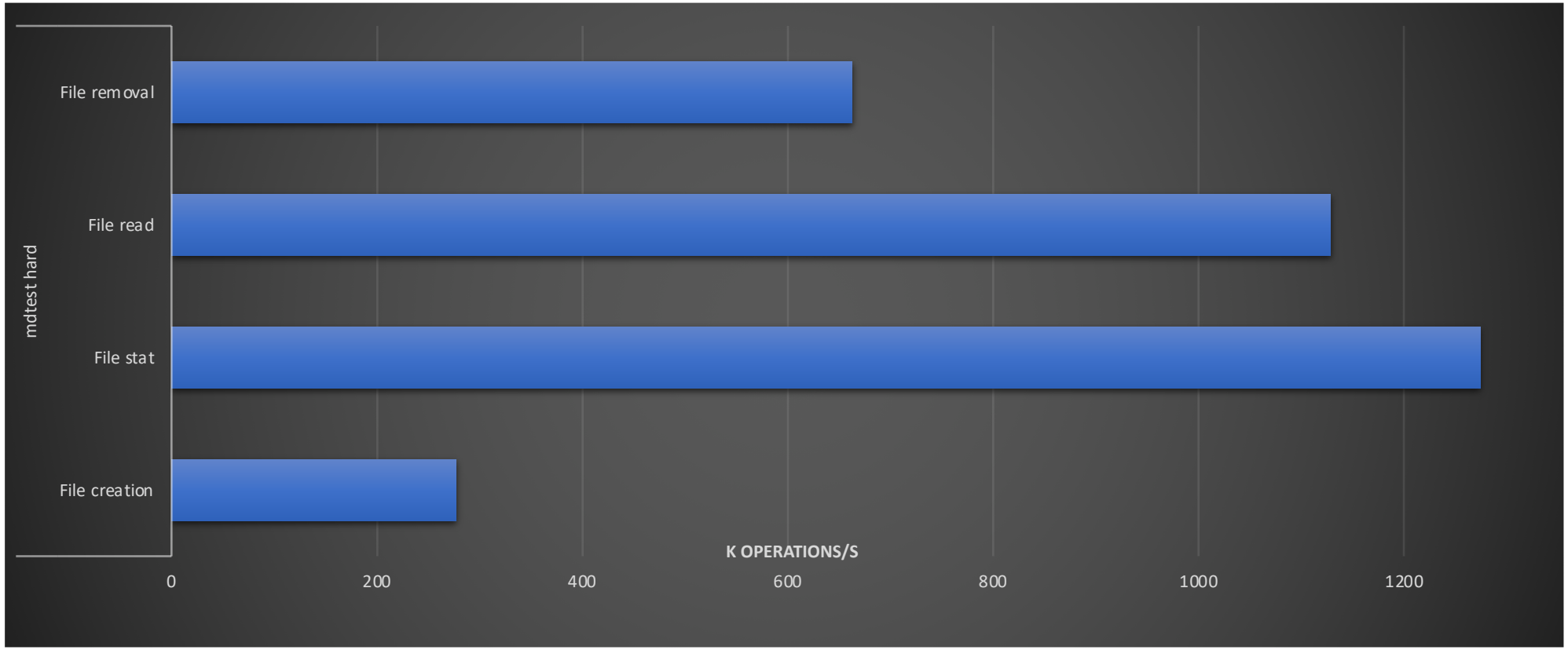
CPU FREQUENCY IMPACT

- No significant impact of the compute node frequency

MDTEST EASY (WITHOUT PMEM)



MDTEST HARD (WITHOUT PMEM)



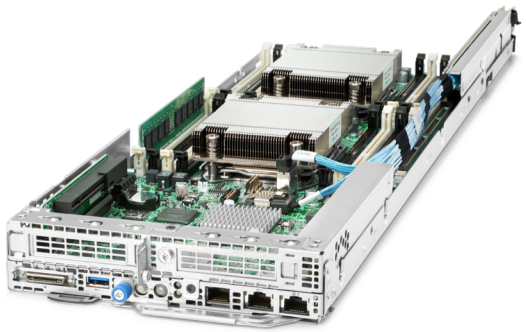
FUTURE TEST

TESTBED V2

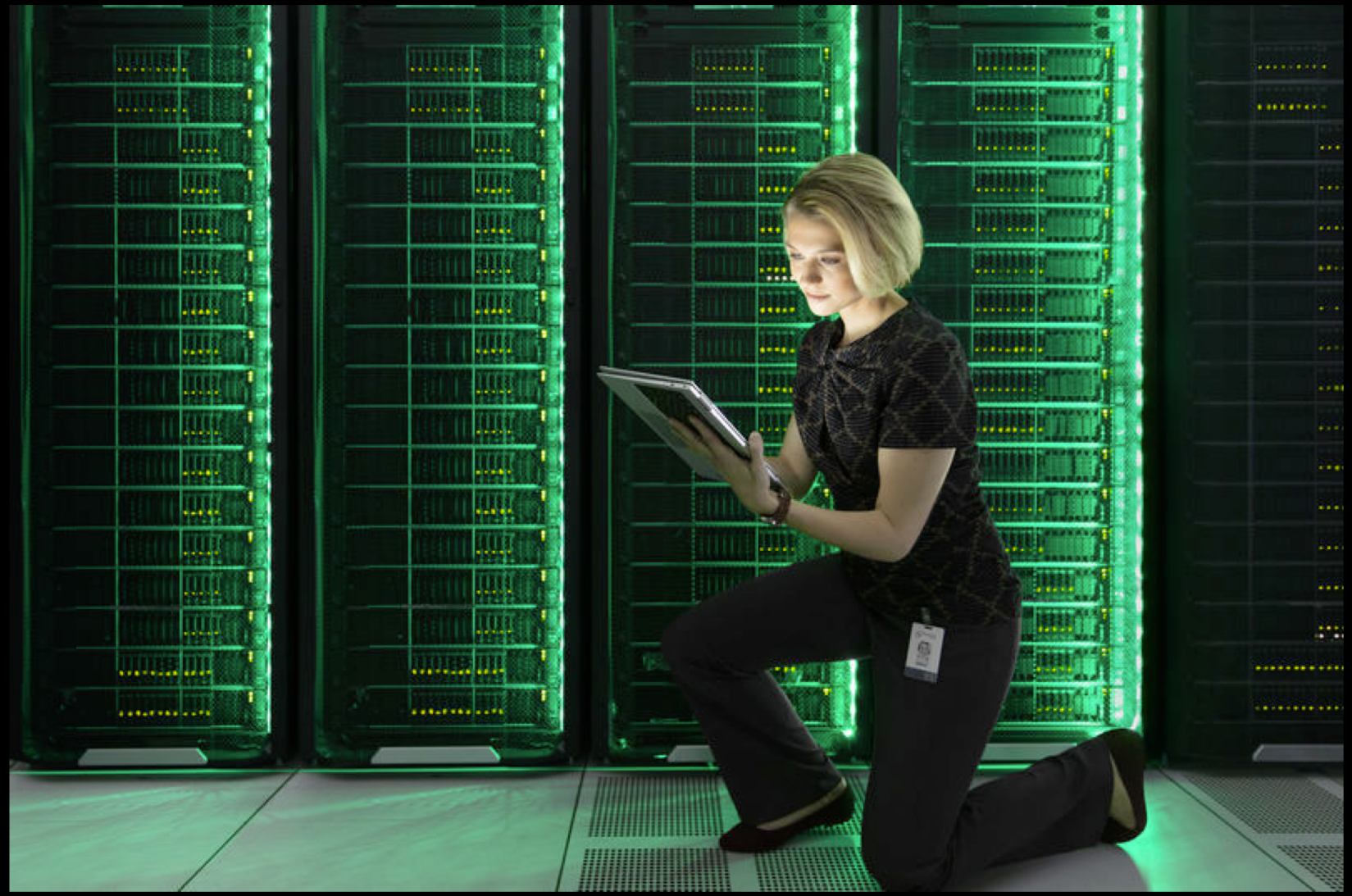
- 7x storage nodes
  - Dual socket
    - Intel(R) Xeon(R) Gold 6242 CPU @ 2.80GHz
    - 16 cores per socket
  - 96GB of DRAM
  - 4x 256GB Optane DC persistent memory
  - 4x 3.2TB P4600 SSD
    - https://ark.intel.com/content/www/us/en/ark/products/97002/intel-ssd-dc-p4600-series-3-2tb-2-5in-pcie-3-1-x4-3d1-tlc.html
    - R@2.8GB/s and W@1.9GB/s
  - 1x IB EDR card

- 8x compute nodes
  - Dual socket Xeon(R) Gold 6148 @ 3.20GHz

  1x IB EDR card



WHAT'S NEXT



- Additional Backend
  - HDF5
  - MPIO
  - other…..
- Core functionality
  - Replica and Erasure Coding
  - Recovery Capability
- Test as scale
- ”Real World” Application

Q&A