Reimaging the Data Centre Memory and Storage Hierarchy

Andrey Kudryavtsev,
SSD Solution Architect, NSG, Intel Corp.
Notice and Disclaimers

Intel technologies’ features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. **No product can be absolutely secure.**

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. For more complete information about performance and benchmark results, visit [http://www.intel.com/benchmarks](http://www.intel.com/benchmarks).

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit [http://www.intel.com/benchmarks](http://www.intel.com/benchmarks).

Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

The benchmark results may need to be revised as additional testing is conducted. The results depend on the specific platform configurations and workloads utilized in the testing, and may not be applicable to any particular user's components, computer system or workloads. The results are not necessarily representative of other benchmarks and other benchmark results may show greater or lesser impact from mitigations.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at [http://www.intel.com/go/turbo](http://www.intel.com/go/turbo).

© 2018 Intel Corporation. Intel, the Intel logo, and Intel Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as property of others.
Changing Data Needs Have Exposed Storage & Memory Gaps

- **Working set scaling is limited**
  - **DRAM**
    - Capacity Gap
  - **NAND SSD**
    - Performance Gap
    - Lower performance/$ limits value of stored data
    - Cost Gap

Scale Your Innovation
Intel® Optane™ Technology+ Intel® QLC Technology

Fill the Gaps

- Enable new insights with bigger, more affordable memory
- Break through bottlenecks to increase value of storage data
- Cost-optimized SSDs enable storage consolidation and acceleration

Intel® Optane™ DC persistent memory

Intel® QLC 3d NAND SSD
HPC. Recognize the BEST fit for SSDs.
HPC. Recognize the BEST fit for SSDs.

1. **Local compute storage** – use Optane SSDs for certain workloads requiring large scratch/temp storage
HPC. Recognize the BEST fit for SSDs.

1. **Local compute storage** – use Optane SSDs for certain workloads requiring large scratch/temp storage

2. **IO nodes** – deploy Optane SSDs to accelerate data transfer to/from compute node and/or burst buffer for usages such as memory snapshot across multiple compute nodes
HPC. Recognize the BEST fit for SSDs.

**Local compute storage** – use Optane SSDs for certain workloads requiring large scratch/temp storage

**IO nodes** – deploy Optane SSDs to accelerate data transfer to/from compute node and/or burst buffer for usages such as memory snapshot across multiple compute nodes

**Metadata** – accelerate meta data and access to frequently accessed small files with Optane SSDs
HPC. Recognize the BEST fit for SSDs.

1. **Local compute storage** – use Optane SSDs for certain workloads requiring large scratch/temp storage.

2. **IO nodes** – deploy Optane SSDs to accelerate data transfer to/from compute node and/or burst buffer for usages such as memory snapshot across multiple compute nodes.

3. **Metadata** – accelerate meta data and access to frequently accessed small files with Optane SSDs.

4. **Storage** – reduce storage TCO with space and power efficient, highly manageable ruler FF. Opportunity for NAND based NVMe SSDs.

---

**HPC Cluster**

- **Optane SSD**
- **P4510/P4610 SSDs**

**Parallel Storage**

- **Meta data**
  - Optane SSDs
- **Tiered Storage**
  - P4510 16TB U.2 or “Ruler” EDSFF SSDs

**Local compute storage** – use Optane SSDs for certain workloads requiring large scratch/temp storage.

**IO nodes** – deploy Optane SSDs to accelerate data transfer to/from compute node and/or burst buffer for usages such as memory snapshot across multiple compute nodes.

**Metadata** – accelerate meta data and access to frequently accessed small files with Optane SSDs.

**Storage** – reduce storage TCO with space and power efficient, highly manageable ruler FF. Opportunity for NAND based NVMe SSDs.
HPC. Recognize the BEST fit for SSDs.

1. **Local compute storage** – use Optane SSDs for certain workloads requiring large scratch/temp storage

2. **IO nodes** – deploy Optane SSDs to accelerate data transfer to/from compute node and/or burst buffer for usages such as memory snapshot across multiple compute nodes

3. **Metadata** – accelerate meta data and access to frequently accessed small files with Optane SSDs

4. **Storage** – reduce storage TCO with space and power efficient, highly manageable ruler FF. Opportunity for NAND based NVMe SSDs.

5. **Memory nodes** – use Optane DC Persistent Memory or Optane SSD with Intel Memory Drive Technology to deploy fat memory nodes.
Low Latency + High Endurance = Greater System Efficiency

1. Source – Intel tested. Average read latency measured at queue depth 1 during 4k random write workload. Measured using FIO 2.15. Common Configuration - Intel 2U Server System, OS CentOS 7.5, kernel 4.17.6-1.el7.x86_64, CPU 2 x Intel® Xeon® 6154 Gold @ 3.0GHz (18 cores), RAM 256GB DDR @ 2666MHz. Configuration – Intel® Optane™ SSD DC P4800X 375GB and Intel® SSD DC P4600 (3D NAND). Latency – Average read latency measured at QD1 during 4K Random Write operations using fio -2.15. System BIOS: 00.01.0013; ME Firmware: 04.00.04.294; BMC Firmware: 1.43.91f7695; FRUSDR: 1.43. The benchmark results may need to be revised as additional testing is conducted. Performance results are based on testing as of July 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.


Intel® Storage Performance Snapshot Tool

Overview
- A lightweight tool for collecting and analyzing system-level performance information with special focus on storage

Data Collector
- Very easy to use, zero configuration, non-intrusive, low overhead
- Based on the Linux® dstat utility and requires dstat to operate
- Outputs a standard CSV file

User Interface (UI)
- HTML-based UI for viewing and analyzing the collected data
- Runs from any modern browser (Chrome, Firefox, IE, Safari)
- Does not require network connection. Data is never uploaded and always stay on the local computer
Introducing Intel® memory drive technology

- Use Intel® Optane™ SSD DC P4800X transparently as memory

- Grow beyond system DRAM capacity, or replace high-capacity DIMMs for lower-cost alternative, with similar performance*

- Leverage storage-class memory today!
  - **No change to software** stack: unmodified Linux* OS, applications, and programming
  - **No change to hardware**: runs bare-metal, loaded before OS from BIOS or UEFI

- Aggregated single volatile memory pool
Intel® memory drive technology delivers big, affordable memory.

**Use Case 1:** Expand beyond limited DRAM capacity

**Intel® Memory Drive Technology**
Expand Insights with Massive Data Pools

**Use Case 2:** Displace dram with Affordable SSDs

**Intel® Memory Drive Technology**
Reduce High-capacity DRAM CAPEX Expenditures

**Note:** Intel® Memory Drive Technology supports Linux® x86_64 (64-bit), kernels 2.6.32 or newer.

*Other names and brands may be claimed as the property of others.*
New Data Solutions. Supporting Data Center Design Flexibility.

Intel® Optane™ DC persistent memory Module

- Capacity: Bigger is Better
- Latency/Bandwidth
- Power
- Persistency: Y

Intel® Optane™ SSD with Intel Memory Drive Technology

- Persistency: N

Graphical representation of product comparison is based on internal Intel analysis, and is provided here for informational purposes only. Any differences in system hardware, software or configuration may affect actual performance.
**HPC. Traditional Cluster Configuration to Flex Memory Nodes**

**HPC Cluster**

- **Compute Nodes**
  - Optane SSD

- **IO Nodes**
  - P4510/P4610 SSD

- **Fat Memory Nodes**
  - DRAM 384GB-3TB

**Additional Info**

- Typically many HPC clusters come with Fat Memory nodes to address the needs of higher memory capacities.
- Examples of workloads: Genomics, Finite element analysis, In-core simulations, Computational Chemistry, etc.
- Historically HPC cluster includes %5 of Fat Memory nodes and between %5 and 10% I/O Nodes, which is based on workload requirements.
- In many cases Fat memory nodes are similar architecture as main compute node just with expanded memory capacity. In other cases, 4-socket systems used to fit that much memory.

**Solution**

- Flexible configuration on demand, where Optane can easily switch between storage and memory mode with **Intel Memory Drive Technology**.
- Flex Memory Node solution delivers more Fat Memory nodes that customer can afford within a DRAM budget, while expanding use cases to the storage needs (temp/scratch) for certain I/O intensive HPC codes.
- Typical usage of that scenario is integrating provisioning into workload scheduler (example SLURM), which scripts Optane either for IMDT installation or file system setup.

---

**Scale Your Innovation**
Demo
SCALE YOUR INNOVATION