Solution Brief
Data Center | Virtualization

Tiered Memory Can Boost Virtual Machine Memory Capacity and Lower TCO

Increase VM density, enable larger VMs, or lower capital and operational expenditure by using high-capacity Intel® Optane™ persistent memory in Memory Mode for main system memory, and DRAM for memory cache

Solution Benefits

- Intel® Optane™ PMem can offer **greater memory capacity per host** than DRAM for virtualized data center infrastructures like VMware vSphere. Using Intel Optane PMem leaves more headroom for hosting future virtualized workloads that require larger memory capacity, instead of having to run those demanding workloads on bare metal.
- Implementing tiered memory solutions with Intel Optane PMem can **improve overall TCO**. Savings can accrue from server consolidation, or from lower CapEx when configuring systems with lower amounts of expensive DRAM.
- Accessing the storage subsystem can incur unacceptable latencies. Keeping more data in larger system memory can often **improve workload performance**.

Executive Summary

As the data explosion continues, VMs with larger memory capacities are necessary in today's data centers (see Figure 1). But DRAM DIMM capacity is not scaling at the same rate as memory demand, and is not keeping pace with modern high-core-count processors. Not to mention that buying TBs of DRAM can be cost-prohibitive.

For VMware vSphere deployments, the solution is memory tiering. Similar to the well-accepted concept of tiered storage, memory tiering uses affordable large-capacity memory for main volatile system memory, coupled with traditional DRAM used as memory cache. Most virtualized applications and databases are ideal candidates for a tiered memory architecture using DRAM and Intel® Optane™ persistent memory (PMem) because they often consume host memory that is not actively used by the VM.

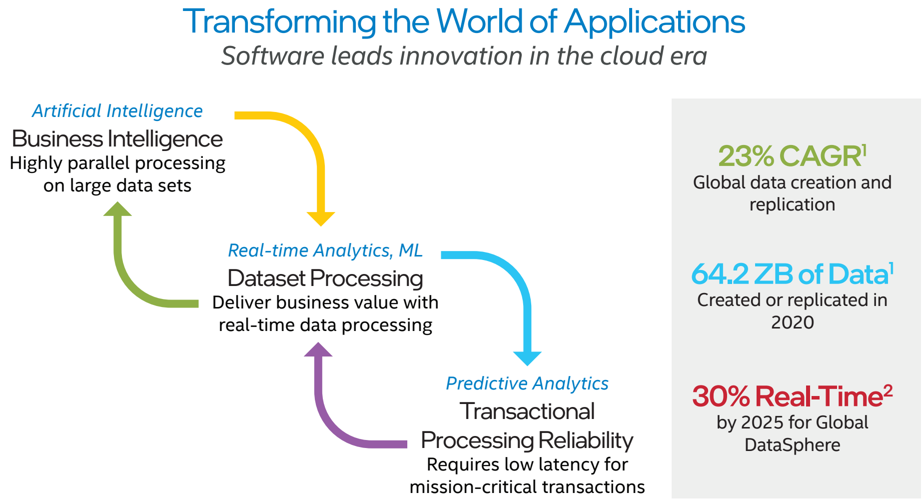

Figure 1. Due to digital transformation across enterprises, data is growing exponentially.[1,2]

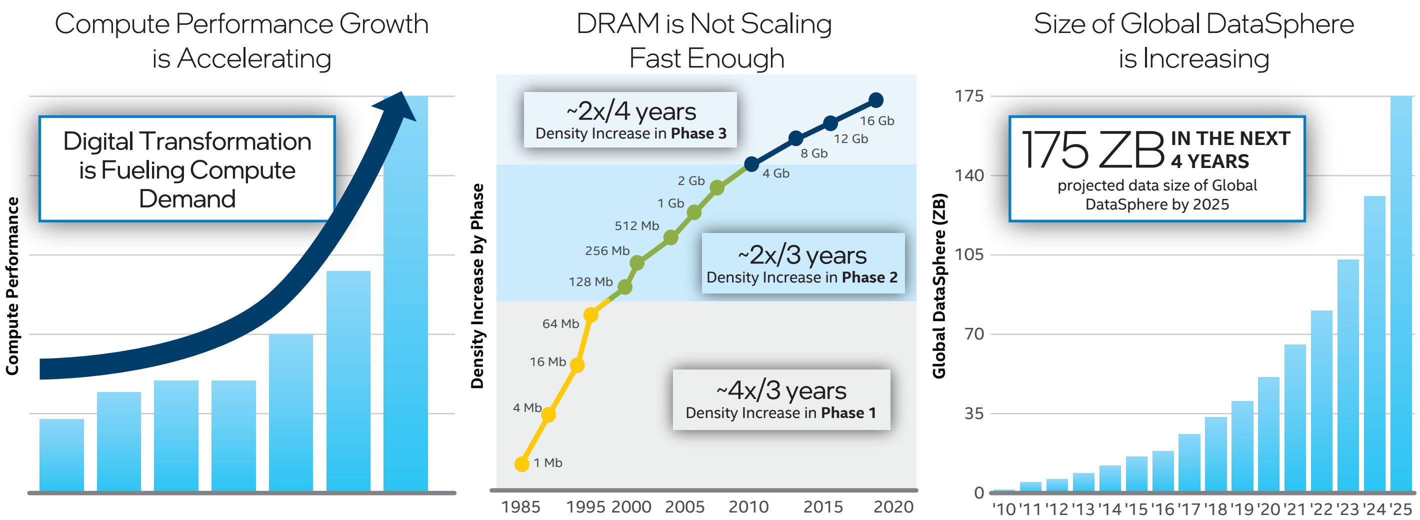

Figure 2. Compute performance and data demand is scaling rapidly; memory capacity is not.[3,4]

Business Challenge: DRAM Capacities Not Keeping Up with Workload Demands

As an increasing number of businesses engage in digital transformation, three important trends emerge (see Figure 2):

- Digital transformation fuels demand for compute, and with it, increasing demand for the memory to support that compute.
- Business data is also increasing exponentially. However, traditional DRAM is not scaling to meet this demand. The growth rates for DRAM density have slowed over time because it has become costly and complex to scale to increasingly higher capacities. In other words, business data growth is increasing demand for compute power but DRAM can be a bottleneck.[3]
- Even as the growth in DRAM capacity slows, memory demand continues to increase, creating a widening—and accelerating—gap.[4]

Simultaneous with these trends, today's mission-critical, data-intensive workloads running on VMware vSphere need to store more hot data in memory—but DRAM at scale is increasingly expensive and limited in capacity. Examples of such workloads include large in-memory databases and real-time workloads. For these types of real-time workloads, bringing data as close to the CPU as possible is critically important so that latencies remain low.

These realities are converging into a need for a new memory tier. A tiered memory system uses the same approach as tiered storage (see Figure 3). In a tiered memory scenario, Intel® Optane™ persistent memory (PMem) in Memory Mode is used as the large capacity tier, while a relatively small amount of DRAM serves as a memory cache tier to maintain high speed and low latency.

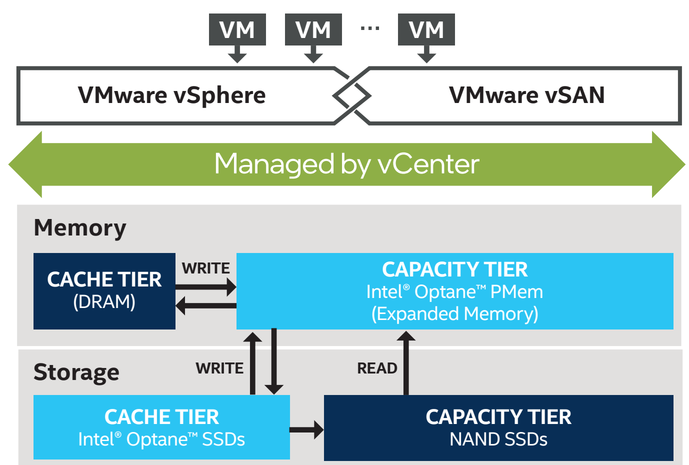

Figure 3. In a tiered memory system, Intel® Optane™ persistent memory (PMem) serves as the main system memory, while the system's DRAM serves as a memory cache.

Intel® Optane™ Persistent Memory (PMem) in Memory Mode Explained

Historically, there have been two separate places to store data: in volatile memory (DRAM) or in persistent storage (such as SSDs). But as discussed above, DRAM isn't big enough or affordable enough for today's larger memory demands, and even NVMe-based SSDs have unacceptable latencies for modern real-time workloads.

Intel Optane PMem fills that large gap between DRAM and storage; it's a totally new type of media with characteristics of both memory and storage. An Intel Optane PMem module looks very much like a DRAM DIMM and plugs into the same physical DIMM connector on the same memory bus.

Additionally, Intel Optane PMem is byte-addressable, which helps lower latency (every byte can be immediately set or reset, unlike storage access, which uses blocks and pages).

Intel Optane PMem can be used in two main modes. Here, we'll focus on Memory Mode, used for hardware memory tiering. Using Intel Optane PMem in Memory Mode does not require any software application changes; simply plug the Intel Optane PMem modules into DIMM slots on the server and VMware ESXi can immediately use them. Note the following:

- In Memory Mode, Intel Optane PMem is actually not persistent—it is volatile just like DRAM, but has a much greater capacity.
- VMware vSphere recognizes the Intel Optane PMem capacity as vRAM; the DRAM cache is not counted toward total system memory capacity.
- All of the memory—DRAM cache and Intel Optane PMem capacity—is managed by the CPU's memory controller to achieve low latencies (see Figure 4). The OS, kernel, and hypervisor have no control over the DRAM cache.

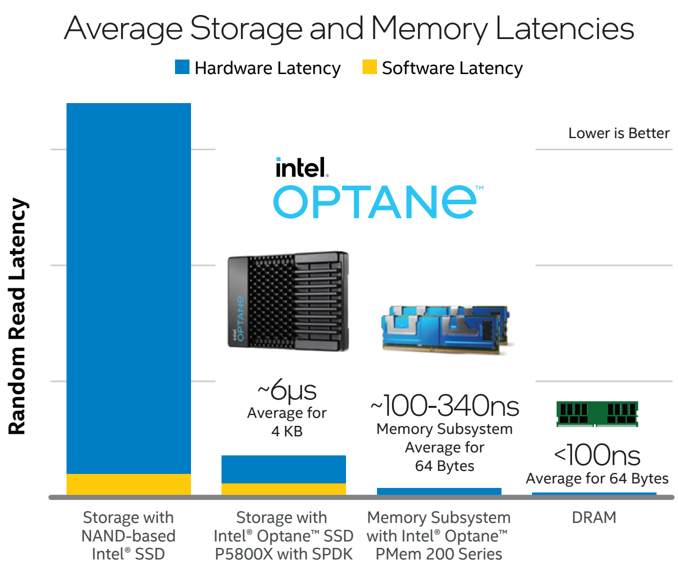

Figure 4. Because it resides on the memory bus, Intel® Optane™ persistent memory (PMem) latency is much closer to that of DRAM than to block storage latency.[5]

To understand how to best use Intel Optane PMem in Memory Mode, consider Figure 5. Each VM has a certain amount of memory allocated to it. Of that **allocated** memory, some of it is being **actively** accessed or used by the applications. The rest of the allocated memory is either **idle** (hasn't been recently accessed or used) or **free** (not being accessed or used by any application on the VM). The combination of active and idle memory comprises the VM's **consumed** memory.

On a tiered memory system equipped with Intel Optane PMem, the data in DRAM is the active subset of data stored in Intel Optane PMem. It is best practice to size the DRAM so that all of the workload's active memory fits within the DRAM footprint. Generally, hosts and applications with active memory less than 25 percent of the consumed memory are good candidates to deliver the same performance of a host with a single tier of DRAM.

To summarize Intel Optane PMem in Memory Mode:

- Intel Optane PMem functions as volatile memory.
- DRAM functions as memory cache.
- Organizations can achieve affordable high-capacity memory with no software application modification.

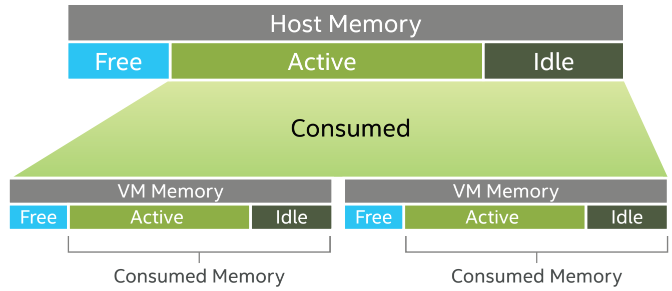

Figure 5. Active and idle memory make up a VM's consumed memory, which together with free memory, comprise the total allocated memory for that VM.

Solution Value: Cost Efficiencies and Larger VM Possibilities with Tiered Memory

Intel Optane PMem offers excellent business value along two primary vectors: increasing VM density (with a lower cost per VM) or saving capital expenditures (CapEx) at the same VM density by simply right-sizing DRAM and taking advantage of memory tiering. The following sections provide more detail on each scenario. Intel Optane PMem proof points are also available.

Increasing VM Density

VMware ESXi hosts are often constrained by memory, as opposed to being I/O- or CPU-bound. With more allocated memory available, it is possible to add more VMs to a VMware ESXi host. More VMs per server improves server consolidation, drives up CPU utilization for better data center efficiency, and can enable data centers to deploy fewer servers. The result can reduce energy as well as operational, and software licensing costs for significant total cost of ownership (TCO) benefits.

Right-sizing DRAM to Save CapEx and Enable Larger VMs

By dramatically increasing VMware ESXi host system memory, organizations can affordably accommodate larger VMs. Possible scenarios include:

- Applications and databases that are already virtualized but have increasing memory requirements.
- Large bare-metal applications and databases that have not previously been good candidates for virtualization due to memory requirements.
- Increased ability to provision new large VMs to keep pace with business demand.

It is common practice to over-provision DRAM on a given VM for future memory growth needs (as data is growing exponentially) and to handle periodic workload spikes (examples include holiday online shopping traffic increases and month/quarter/year-end close activity).

However, this practice can result in scenarios where a significant amount of system memory is idle (Figure 6). This snapshot shows a 30-day use chart for VM memory resources. In this real-world example, the VM's applications demanded only about 14 percent of the VM's allocated memory for active memory. Even accommodating more than 50 percent above this level for load spikes and a safety buffer, a recommended amount of DRAM for this VM's active data would be only about 22 percent of the allocated memory—meaning that approximately 77 percent of this VM's allocated DRAM is idle. In this scenario, Intel Optane PMem could be used for the over-provisioned memory, lowering the amount of DRAM necessary and helping lower TCO.

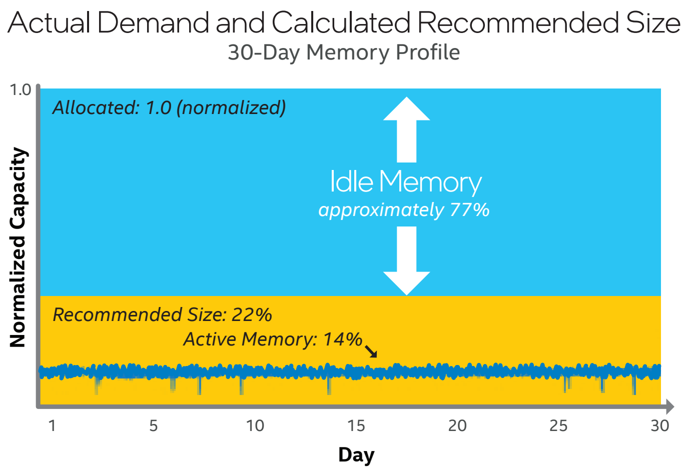

Figure 6. The cluster's cumulative VM memory utilization profile shows active data use is well below the allocated DRAM capacity.

Solution Architecture: Big Memory for the New Data Frontier

VMware ESXi deployments using 3rd Generation Intel® Xeon® Scalable processors can take advantage of the second generation of Intel Optane PMem—Intel Optane persistent memory 200 series. With eight memory channels, these processors support up to 8 TB of Intel Optane PMem in Memory Mode, compared to a two-socket DRAM-only platform, which can accommodate only 4 TB of memory. To achieve optimal bandwidth and preserve memory-per-core or memory-per-VM performance, it makes sense to keep all eight memory channels active.

Figure 7 illustrates several possible tiered configurations of Intel Optane PMem and DRAM. Organizations that are currently configuring servers with 384 GB of DRAM may consider upgrading to 512 GB as their workloads' memory needs increase. Many of our customers are already using 768 GB or more memory and they are moving towards 1 TB or more memory. But with Intel Optane PMem, system memory can be increased to 1 TB, 2 TB, or more for significantly less outlay of budget. It is also possible to use Intel Optane PMem 100 series (with 2nd Gen Intel Xeon processors), as shown in the bottom half of Figure 7.

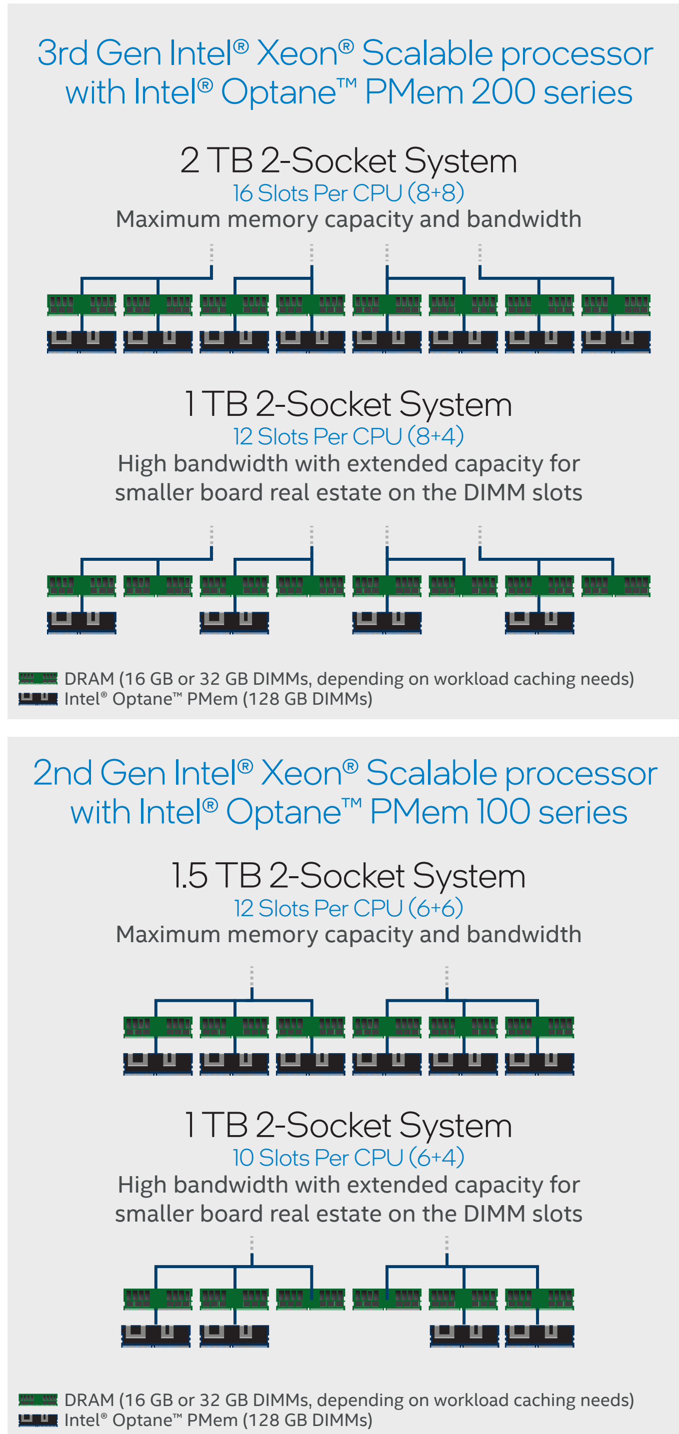

Figure 7. Tiered memory configurations using Intel® Optane™ persistent memory (PMem) can help organizations meet their performance and TCO goals.

Best Practices for Using Intel Optane PMem with VMware vSphere

VMware vSphere (vSphere 6.7 EP10 and later) supports Intel Optane PMem 100 series with 2nd Gen Intel Xeon Scalable processors. vSphere 7.0U2 and later supports Intel Optane PMem 200 series with 3rd Gen Intel Xeon Scalable processors. Many configurations are supported, and VMware provides several best practices for using Intel Optane PMem with VMware software. For example:

- The recommended DRAM sizes are as follows:
  - A 1:8 ratio of DRAM to Intel Optane PMem, which is 12.5 percent of Intel Optane PMem capacity
  - A 1:4 ratio of DRAM to Intel Optane PMem, which is 25 percent of Intel Optane PMem capacity
- The system must be populated with at least four Intel Optane PMem DIMMs per socket.
- Maximum performance is achieved when active memory of the host in steady state fits within the amount of DRAM configured in the host. Even when DRAM cache misses occur, memory latency is still in the hundreds of nanoseconds range.[6]
- Verify that the server platform is running with the Balanced Profile BIOS setting recommended by the server OEMs.

Read the VMware Knowledge Base article, "vSphere Support for Intel's Optane Persistent Memory (PMem)," for a full discussion of best practices and supported configurations. For more information on server platforms that meet these requirements, please contact your OEM server vendor for supported configurations.

Conclusion

VMware ESXi or similar environments can experience a wide array of benefits from using tiered memory to expand memory capacities beyond DRAM limitations. Most virtualized, general-purpose workloads—either on-premises or in the hybrid cloud—with low active memory utilization are excellent candidates for a tiered memory structure. Examples of such workloads include virtual desktop infrastructure (VDI) and databases.

At a time when technology trends demand VMs to support ever-larger sizes without sacrificing performance, Intel's memory breakthrough provides an opportunity to surpass traditional architecture limitations, consolidate server resources, and significantly reduce TCO. Intel Optane PMem is an affordable solution that drops right into existing DDR4 memory slots. IT can right-size DRAM investments and expand total memory per server, accommodate much larger workloads, increase VM density, and dramatically improve resource utilization.

These benefits fit perfectly with enterprises in the midst of their digital transformation. Companies working to modernize their infrastructure and increase their efficiency to meet the mounting wave of data demands can benefit from tiered memory systems that take advantage of Intel Optane PMem.

Learn More

You may also find the following resources useful:

- 3rd Gen Intel® Xeon® Scalable processors
- Intel® Optane™ Persistent Memory
- Intel and VMware collaboration page
- Intel and VMware Virtualization Solutions
- VMware solutions with Intel® Optane™ technology
- VMware Knowledge Base article - vSphere Support for Intel Optane PMem

Find the solution that is right for your organization. Contact your Intel representative or visit **VMware Solutions with Intel® Optane™ technology.**

[1] CAGR of 23% and data replication claims: https://www.idc.com/getdoc.jsp?containerId=prUS47560321

[2] Real-time data claim: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

[3] "DRAM scaling: 3D NAND Technology – Implications for Enterprise Storage Applications" by J. Yoon, IBM; 2015 Flash Memory Summit.

[4] Data scaling: Data Age 2025 – sponsored by Seagate with data from IDC Global DataSphere, November 2018.

[5] **NAND:** Based on Intel testing of a 6.4 TB Intel® SSD P5600 as of March 16, 2020: Intel® Xeon® Gold 6139 processor (2.30 GHz), BIOS: SE5C620.86B.00.01.0014.070920180847 (Intel® Server Board S2600WFT), 2 sockets, 32 GB RAM (DDR4-2137), RAM stuffing: 1 of 4 channels, four DIMM slots populated, PCIe attach: CPU (not PCH lane attach), chipset: Intel® C612 Chipset, switch/ReTimer model/vendor: Intel® G4SAC switch (PCIe Gen4), NVMe driver: kernel 4.17.74 (native), C-states: disabled, Intel® Hyper-Threading Technology (Intel® HT Technology) disabled, CPU governor (through OS): performance mode, OS: CentOS 7.5, kernel: 4.14.74, FIO version: 3.5.
Intel® Optane™ SSD: Based on Intel testing of a 1.6 TB Intel® Optane™ SSD P5800X as of September 25, 2020: Intel® Xeon® Gold 6254 processor (3.10 GHz, 30 MB, 160 W, 18 cores per socket, 2 sockets), BIOS: SE5C620.86B.02.01.0009.092820190230, 32 GB RAM (DDR4), RAM stuffing: NA, DIMM slots populated: 4 slots, PCIe attach: CPU (not PCH lane attach), chipset: Intel® C610 Chipset, switch/ReTimer model/vendor: Intel® G4SAC switch (PCIe Gen4), OS: CentOS 7.5.1804, kernel: 4.14.74, FIO version: 3.5; NVMe driver: Inbox, C-states: Disabled, Intel® HT Technology disabled, CPU governor (through OS): performance mode. Enhanced Intel SpeedStep® Technology, Intel® Turbo Boost Technology, and P-states disabled; IRQ balancing services (OS) off; SMP affinity set in the OS; queue depth 1 (QD1) utilizes I/O polling mode with ioengine=pvsync2/hipri.
Intel® Optane™ persistent memory: Intel Optane PMem 200 series tested by Intel on a single-DIMM configuration as of September 25, 2020. 3rd Generation Intel® Xeon® Scalable processor; chipset LBG B1; 26 cores, 1 socket; DDR speed: 2,666 MT/s; 256 GB Intel Optane PMem module, 15 W memory configuration, 1 channel, 32 GB DDR4 (six per socket); Intel Optane PMem firmware: 2.2.0.1516; BIOS 0017.P23; best-known configuration (BKC) version WW38 BKC, Linux OS Fedora release 29 4.20.6-200.fc29.x86_64; Spectre/Meltdown patched (1,2,3a,4); performance tuning quality of service disabled, IODC=5(AD). Intel Optane PMem uses DDRT memory interface and MLC for performance measurements.
DRAM: https://gist.github.com/jboner/2841832 (see line 7)

[6] See endnote 5.