White Paper
December 2021
Data Center | Storage

Why Intel® Optane™ SSDs Are a Better Option than NAND Flash SSDs in the Storage Cache Tier

Consistently high performance, low latency, and excellent endurance enable Intel Optane SSDs to outshine NAND Flash SSDs

Author
Flavio Fomin
Solutions Architect, Intel Corporation

Table of Contents

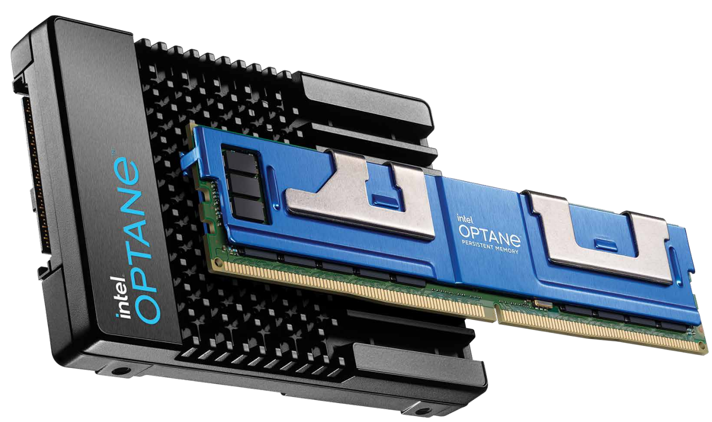
Executive Summary

This paper explores how NAND Flash and Intel® Optane™ SSDs' capabilities affect a VMware vSAN configuration—in particular, the storage cache tier. Considerations include cost, performance, footprint, and risk of data loss.

Generic performance and endurance guidelines are commonly used to select cache drives. These guidelines help prevent the selection of inadequate drives, but when they are used as recommendations for which drive to use, the opportunity to optimize for cost and performance can be missed. And in some cases, there's a risk of using a drive that does not provide acceptable endurance for certain workloads, increasing exposure to data loss events.

In this paper we provide additional information that is not included in the generic sizing guidelines. We describe how Intel Optane technology adds value to a multi-tier distributed storage implementation, such as VMware vSAN, when deployed as a cache tier. This value results from the Intel Optane SSD's higher throughput and endurance, compared to NAND Flash SSDs:

- When processing a 50 percent read/write load, Intel Optane SSDs can provide up to 2.5X more throughput.[1]
- The Intel Optane SSD's lifetime endurance is 16X more than a NAND Flash SSD.[2]



Intel® Optane™ technology adds value to a multi-tier distributed storage implementation when deployed as a cache tier.

Introduction

When configuring a new VMware vSphere and vSAN cluster, one common goal is to match the workload requirements to the characteristics of the hardware components in the host configuration, with the intention of optimizing cost and performance and achieving optimal service levels.

All host components influence this sizing process, but sizing NAND Flash SSDs can be more challenging than other components because the I/O capabilities—MB/s and IOPS—can change with workload profiles. For instance, the read/write throughput and read latency of a NAND Flash SSD varies as we add a concurrent write I/O activity. These differences should be considered when sizing a storage solution, especially when sizing the cache tier.

Drive endurance is another factor that should not be overlooked when sizing a cache drive. If we simply follow guidelines for minimum endurance levels in terabytes written (TBW), without matching the workloads to the drive endurance, we risk using drives that can wear out before the warranty lifetime. Although the supplier may replace the faulty drive, the failure can still expose the business to data loss risk and service-level effects. For instance, in a VMware vSAN cluster, if a cache drive fails, the entire disk group becomes unavailable—increasing the data loss risk until the entire disk group is replaced.

In this study, we compare the differences in VMware vSAN performance using a NAND Flash SSD versus an Intel® Optane™ SSD. We evaluate how the NAND Flash SSD performance capabilities vary with different read/write I/O mixes. That is to say, the gradual increase in cache-to-disk de-staging rates can affect overall performance; this explains why the Intel Optane media can sustain higher throughput than a NAND Flash SSD when processing the same mixed read/write load.[3] We look at the different endurance capabilities and draw attention to the growing risks of data loss if a drive is selected using outdated endurance guidelines. And finally, we summarize our findings, establishing a comparison of cost, performance, footprint, and risk of data loss illustrated by two cluster configurations.

VMware vSAN Architectural Components and I/O Flow

To facilitate the interpretation of the data results presented in the next sections, we provide a high-level review of the vSAN I/O flow (see Figure 1) on the core hardware components.

1. **Write requests.** In a VMware vSAN configuration that exclusively uses SSDs in the capacity tier, vSAN buffers all write I/O requests in the cache SSDs, with reads also coming from data still resident in cache. Up to vSAN 7.0U3, the write buffer is limited to 600 GB.
2. **Data replication.** Writes directed to one host must be replicated to other hosts, according to the protection method used. Note that protection methods amplify the cache and capacity drive activity. For instance, using RAID 1 for every VM write request results in two write I/O requests in the cache tier.
3. **De-staging.** De-staging from the cache tier to the capacity tier happens at a variable rate (related to what percentage of the cache is full):
   - 0-30 percent full: no de-staging (effectively, a 100 percent cache hit rate).
   - 30-70 percent full: throttled de-staging, attempting to reach an equilibrium between front-end and back-end I/O requests.[4]
   - >70 percent full: unthrottled de-staging while the incoming rate to the cache is throttled to balance with de-staging rate.
4. **Read requests.** Reads come from the capacity tier drives, except when the data is still resident in the cache tier.

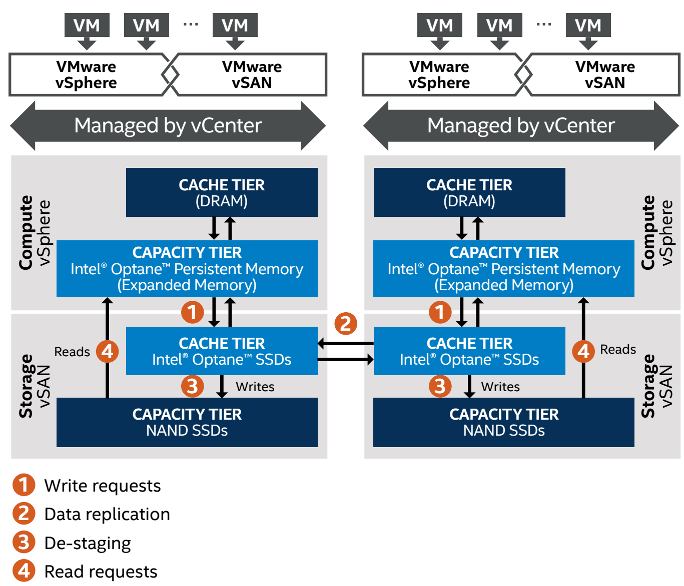


Figure 1. Hardware architecture components and I/O flow.

Test Environment

In this study, we compare the differences in VMware vSAN performance resulting from using two SSD models as cache drives:

- 800 GB NAND Flash SSD with a SAS interface and an endurance of 3 drive writes per day (DWPD).
- 400 GB Intel Optane SSD with a PCIe/NVMe interface and an endurance of 100 DWPD.

We chose the 800 GB drive based on input from OEMs that indicated this type of drive is currently a popular choice. The performance specs of a PCI Gen4 1.6 TB NAND MU flash drive are better than a SAS 800 GB MU drive, but not better than the 400 GB Intel Optane SSD P5800X.[5] Considering that the 800 GB 3 DWPD SSD is an entry-level drive, we expect to illustrate two performance extremes in our study.

Hardware and Software Configuration

The tests were executed on a four-node cluster with the hardware and software listed in Table 1.

Table 1. Test Cluster Hardware and Software Bill of Materials

Host Component	Model
CPU	2x Intel® Xeon® Gold 6348 processor (2x 28 cores @ 2.6 GHz)
Memory	256 GB (16x 16 GB DIMMS, 3200 MT/s)
NIC	1x Intel® Ethernet Adapter E810 100 GbE
Capacity Drives	8x Intel® SSD D7-P5510 3.84 TB
Cache Tier	
Host Config 1 (NAND Flash SSD)	2x Samsung PM1645 800 GB MU (Mixed Use) SAS 3 DWPD
Host Config 2	2x Intel® Optane™ SSD P5800X 400 GB NVMe 100 DWPD
Software	Product Name/Version
Hypervisor	vSphere 7.0U2
SDS	vSAN 7.0U2
Benchmark Tools	HCIBench 2.5.3/VDBench as the I/O driver

VM Configuration

The test environment consisted of the virtual machine (VM) configuration options provided in Table 2.

Table 2. VM Configuration Details

Test Environment Element	Configuration Option
Number of VMs	8 VMs per host/32 VMs total
vCPUs per VM	8
Memory per VM	8 GB
vDisks per VM	2 vDisks, each with 150 GB
Allocated Storage Capacity	9.6 TB usable/19.2 TB raw (based on RAID 1)

Workloads and Methods

The comparison benchmarked the SSDs using the following I/O workloads:

- HCIBench Disk preparation, which writes zeros to the 9.6 TB area allocated to the 64 vDisks used in the test configuration.
- The cache was cleared before each test.
- 70/30 32 KB VDBench ramp-up curve.

The VDBench "curve" option was used in the tests to produce a ramp-up curve. When this option is used, VDBench runs an initial uncontrolled I/O rate test to determine the maximum I/O rate that can be executed with the test I/O profile. It then runs 10 additional tests, starting with 10 percent of the maximum I/O requests observed in the first test, and then growing in increments of 10 percent until it reaches the maximum value (see Figure 2).

Important: The write buffer has a near-zero utilization when the maximum I/O rate is determined. As the test gradually ramps up from 10 percent to 100 percent, de-staging activity gradually increases as the VM I/O requests escalate.

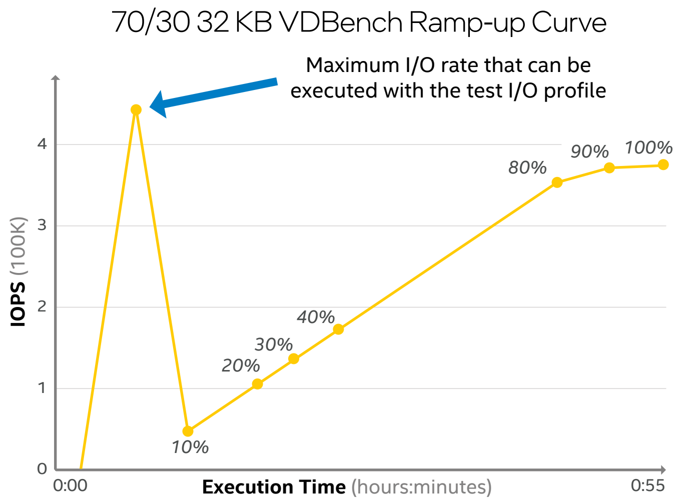


Figure 2. VDBench ramp-up curve option.

Reporting Tools

In addition to the HCIBench reports, which summarize the I/O activity and latency for the workloads tested, this study also used data from the vSAN performance database, such as write buffer utilization, cache-to-disk de-stage rates, and the physical activity of the cache SSDs and capacity SSDs (see Figure 3).

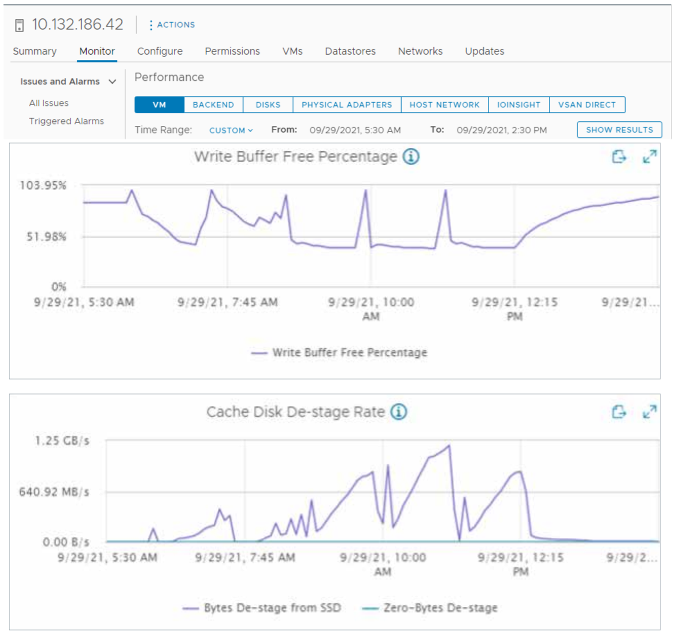


Figure 3. vSAN performance database information.

Test Results

The following sections describe four different methods of examining VMware vSAN performance:

- Elapsed time for vDisk preparation
- A 70/30 read/write workload using a block size of 32 KB
- Effect of cache-to-disk de-staging on VM latency
- I/O activity at the physical layer

vDisk Preparation—Elapsed Time

As part of the test environment preparation, the vDisks must be fully written to ensure that the thin-provisioned space is allocated prior to the tests. In this test configuration, the vDisk preparation wrote a total of 9.6 TB. This operation took 1 hour and 55 minutes to complete when using the 800 GB NAND Flash SSD as a cache drive. By comparison, the same operation took only 30 minutes to complete when using the 400 GB Intel Optane SSD as a cache drive—**nearly 4X faster** (see Figure 4).[6] We can extrapolate these results to other write-intensive workloads, such as video editing, backup-and-restore jobs, and buffer flushes. Note that this write throughput difference is reduced to 1.8X faster if comparing the Intel Optane SSD P5800X 400 GB to a PCIe Gen4 1.6 TB NAND Flash SSD.[7] Although this is a reduction, 1.8X faster is still a significant difference.

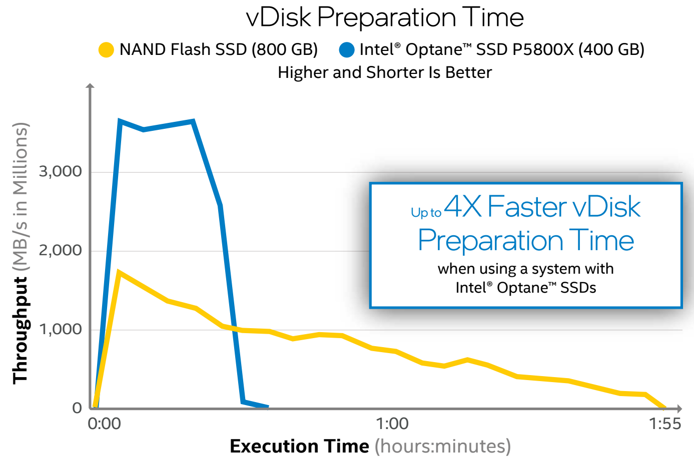


Figure 4. The Intel® Optane™ SSD completed vDisk preparation almost 4X faster than the NAND Flash SSD.

70/30 Read/Write 32 KB

Figure 5 illustrates an I/O versus latency curve for a random I/O workload with 70 percent reads and 30 percent writes, with a 32 KB block size. The SSD used as the cache drive is the only hardware component changed between tests, and we can observe that the test with the 400 GB Intel Optane SSD P5800X **performs 2X more IOPS** than the test with the 800 GB NAND Flash SSD.[8]

We also observe that the 800 GB NAND Flash SSD curve presents erratic behavior when approaching the maximum I/O rate. This happens because when the vSAN system reaches a point at which it cannot de-stage the cache drive fast enough, it starts to throttle the VM I/O requests to balance VM requests and de-staging activity. In contrast, the Intel Optane SSD curve demonstrates a **predictable behavior as the workload escalates**.

In the next two sections, we examine the vSAN layer data and the physical layer data to better understand this difference in I/O performance.

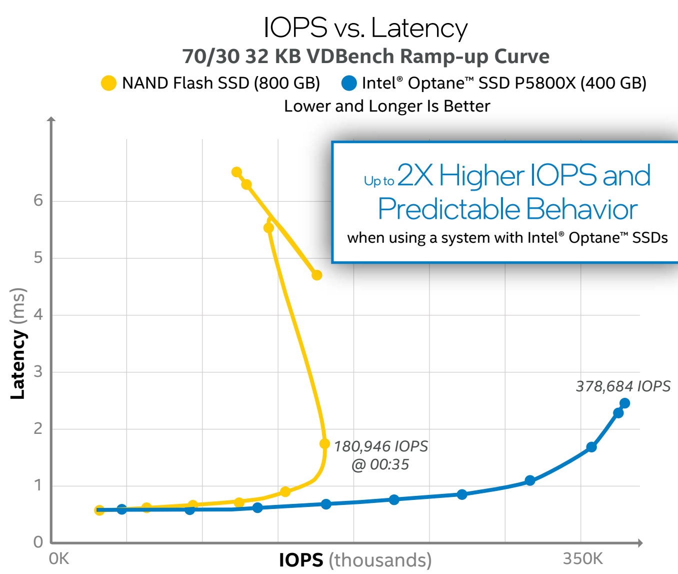


Figure 5. The Intel® Optane™ SSD provides predictable I/O behavior even as the I/O workload escalates.

Effect of Cache-to-Disk De-staging on VM Latency

In the charts shown in Figure 6, we observe the write buffer utilization and the cache-to-disk de-staging rates, when processing the same 70/30 32 KB workload discussed in the previous section. We can correlate the cache activity with the VM I/O activity.

As described in the "Workloads and Methods" section, the first six minutes of this test is an uncontrolled maximum I/O activity that runs with the purpose of identifying the maximum I/O rate in the configuration. This information is later used to generate a ramp-up curve. In this section, we refer to this task as a "ratemax." When ratemax is executed, the write buffer is nearly 100 percent free, indicating that there is no competing de-staging activity.

In the test with the 800 GB NAND Flash SSD, the system reached 305,000 IOPS at ratemax. Because we had a 600 GB write buffer size on the NAND Flash SSD, the write buffer reached nearly 40 percent utilization, which triggered a light de-staging activity of 8 MB/s. As the I/O activity progressed, the de-staging activity increased together with the I/O workload, and the de-stage rate plateaued near 280 MB/s. At this point, near 60 percent of ratemax, the VM write I/O latency jumped to 15.5ms and the VM I/O activity started to decline, as vSAN throttled I/O requests to avoid overrunning the cache. This explains the erratic behavior described in the previous section.

When looking at the charts for the Intel Optane SSD, we observe a 50 percent higher I/O rate during ratemax, reaching 450,000 IOPS. Because of the higher write activity, vSAN triggered a higher initial de-stage rate of 434 MB/s, which kept the buffer utilization at 60 percent. The cache consumption from ratemax was quickly drained; the de-stage rate was reduced to 146 MB/s as ramp-up test started, and then increased linearly following the increase in the VM I/O workload.

In this exercise, we compare the effects of the cache-to-disk de-staging on the system I/O capability, observing the importance of the cache drive performance to the overall vSAN I/O performance. We also take the opportunity to briefly discuss the differences in the write buffer size.

As described in the "VMware vSAN Architectural Components and I/O Flow" section, when the cache utilization is between 30 and 70 percent, vSAN applies a throttled de-staging; this implies that the de-stage rate is expected to increase to keep up with the increase in VM I/O activity.

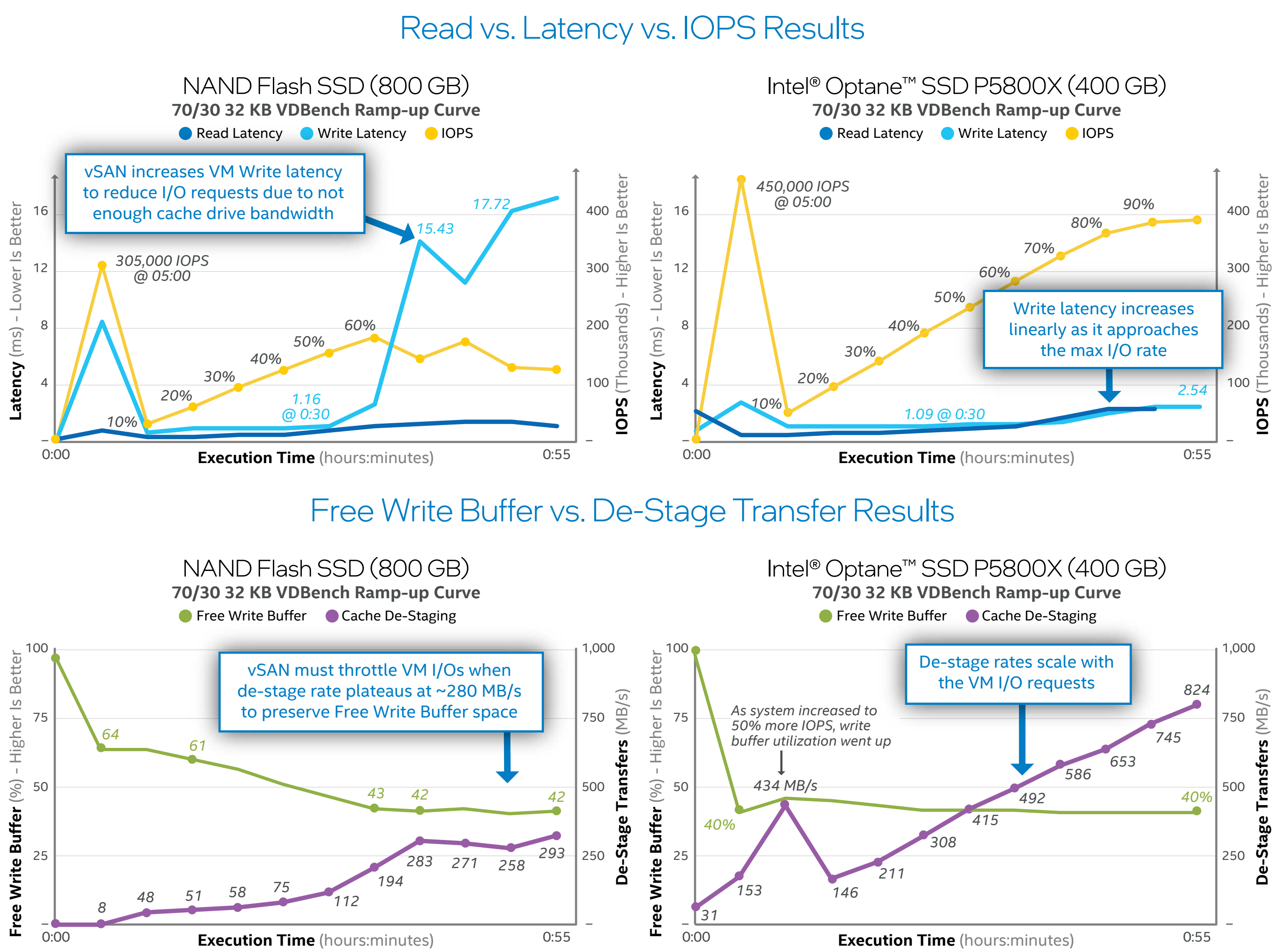


Figure 6. The Intel® Optane™ SSD provided superior performance, with higher IOPS, lower latency, and higher predictability.

When using the 800 GB SSD, the write buffer was configured at its maximum size of 600 GB. With the Intel Optane SSD, the write buffer was smaller (400 GB), and the ramp-up test started with a 50 percent higher I/O rate, which led to a higher de-stage rate from the very beginning of the test. But we can observe that even with a significantly higher de-stage rate on the Intel Optane SSD, the system had higher VM I/O activity with lower latency.

This comparison addresses a common question: Does a buffer size smaller than 600 GB have a negative effect on vSAN I/O performance? We observed that **the ability to sustain a higher read/write bandwidth is the core differentiator, not the buffer size. The de-stage is a significant gating factor for vSAN I/O scalability, and this rate is directly affected by the drive capabilities and by vSAN's internal limits**. Figure 6 shows the de-stage rate with the Intel Optane SSD reached ~820 MB/s while the NAND Flash SSD was limited to ~280 MB/s.

I/O Activity at the Physical Layer

One of the core reasons why the Intel Optane SSD reaches a much higher write throughput is its write-in-place technology.

When data is written to a NAND Flash SSD, a program-erase (P/E) cycle is required. A P/E cycle is the event of writing a memory cell from the programmed state to the erased state, and back to the programmed state. P/E cycles are time-consuming events that directly affect drive performance. Intel Optane media's write-in-place technology does not require a P/E cycle when data is written to the media, enabling significantly higher write I/O density.

In the left chart in Figure 7, we can observe that the NAND Flash SSD reached a write throughput around 750 MB/s when the workload was almost entirely writes. But as the de-staging activity increased and the activity on the drive approached a 50 percent read/write mix, the write throughput peaked at 439 MB/s and the read throughput at 368 MB/s.

In the right chart in Figure 7, the Intel Optane SSD reached a write throughput near 1,100 MB/s when processing mostly writes, and could sustain almost the same write throughput rate at 50 percent read/write, reaching 1,019 MB/s writes and 965 MB/s reads. Note: The Intel Optane SSD can scale significantly further than 1 GB/s writes, but in this test, we reached internal vSAN limits at this point.

Note that increases in latency correlate to the drive I/O limits. On the NAND Flash SSD, test latency went up to 8 ms in the ratemax test, when the write buffer was still clear—indicating the saturation of the drive's capabilities. Latency increased again near 35 minutes into the test when the VM I/O requests had to be throttled to find an equilibrium between reads and writes. But on the Intel Optane SSD, latency remains the same until the test approached vSAN internal limits near 100 percent of the max IOPS.

This comparison adds perspective to the cache-to-disk de-staging discussion in the previous section, further illustrating the relevance of Intel Optane technology's inherent media differences. The 400 GB Intel Optane SSD delivered a **2.5X higher throughput rate** than the 800 GB NAND Flash SSD (1,984 total MB/s versus 772 total MB/s). As described in the Best Practices Guide, "Deploying a Cost-Efficient, High-Performance vSAN Cluster," this performance benefit of Intel Optane SSDs can be extended to other comparisons, where a 400 GB Intel Optane SSD provides nearly 8X higher write throughput per GB than a 1.6 TB NAND Flash SSD with a NVMe/PCIe Gen4 interface.[9]

Physical Cache Disk Activity Results

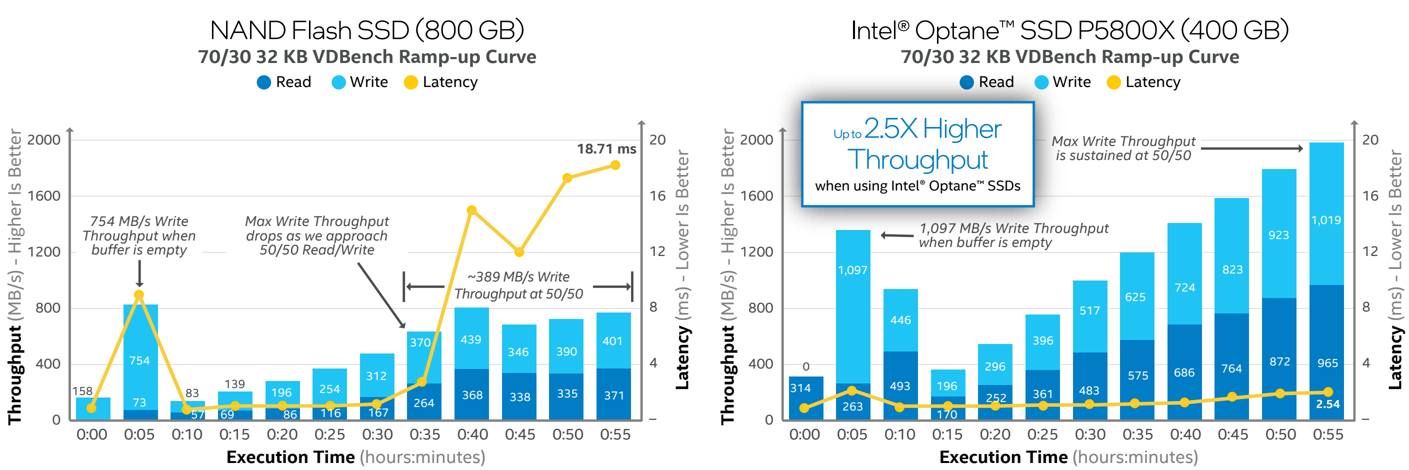


Figure 7. The 400 GB Intel® Optane™ SSD delivers a 2.5X higher throughput rate in this comparison than the 800 GB NAND Flash SSD.

Risk of Using Lower-Endurance Cache Drives

Drive lifetime endurance is reported commercially as DWPD, or the number of times a drive can be written to per day for five years. It can also be presented as TBW over the same time interval.

It is quite common to use endurance guidelines, such as VMware's general SSD endurance guidelines for vSAN, to determine which SSD to use in the cache tier. In this vSAN guideline, we observe two endurance classes for All-Flash configurations: class C drives requiring ≥ 3650 TBW, and class D drives requiring ≥7300 TBW.

TBW over 5 years = drive size x DWPD x 365 days x 5 years

These guidelines are useful, in that they help prevent the selection of cache drives that are totally inadequate; but when they are used as prescriptions, we risk choosing drives that cannot sustain the target workload's demand.

Since these TBW guidelines were proposed, circa 2016, the NAND Flash SSD industry has introduced several larger and less expensive drives, to satisfy cost and capacity demands and, as this document is written, we observe that the 800 GB SAS 3 DWPD (4300 TBW) drive has become a popular option as a cache drive. This is something that requires further discussion because it suggests that an entry-level endurance recommendation made years ago continues to be valid when configuring systems that can have a significantly higher VM and I/O density due to the introduction of new CPU generations and today's more intense hybrid cloud I/O workloads.

We sometimes hear that this is a minor issue because if the drive wears out before the 5-year warranty, the supplier is responsible for replacing it. But we should not ignore the risk and service-level effects that these failures (and subsequent replacements) can bring to large-scale deployments. For instance, the entire vSAN disk group becomes unavailable if the cache drive wears out. This scenario increases exposure to data loss. It also creates an operational impact since all drives in this disk group must be removed and subsequently recreated. The operational costs can be quite meaningful when considering large-scale implementations.

Therefore, we propose that when evaluating a cache drive, it's important to consider what is the maximum write I/O workload that the drive endurance supports. For instance, the 800 GB 3 DWPD drive reaches its 5-year lifetime limit with a write load of 28.4 MB/s.

(3 DWPD × 800 GB × 1024) ÷ (3600 secs × 24 hours) = 28.44 MB/s

It is important to also consider the write I/O amplification in drive evaluation. In our use case with the 800 GB 3 DWDP drive, the 28.4 MB/s value becomes 14.2 MB/s when we account for the I/O amplification from the RAID 1 protection method.

Using this reference value, we evaluated different I/O profiles. For instance, 1,500 IOPS of a workload with 70 percent reads and 32 KB I/O sizes generate a write load of 14 MB/s.

In Table 3, we compare the endurance of 800 and 1600 GB NAND Flash SSDs and the 400 GB Intel Optane SSD to the I/O activity that can be supported for five years.

Table 3. Endurance As It Relates to I/O Activity

Drive Size	Endurance DWPD/TBW	VM Writes MB/s	70/30 32 KB IOPS
800 GB NAND SSD	3/4,300	28	1,500
1600 GB NAND SSD	3/8,600	56	3,000
400 GB Intel® Optane™ SSD	100/73,000	470	25,000

Implications of SSD Performance and Endurance to vSAN Cluster Sizing

In this last section, we work on a cluster sizing exercise—taking into consideration the characteristics of the cache drives—and evaluate the differences in cost, performance, footprint, and risk of data loss. We performed two comparisons: one with a light I/O workload and another with a moderate I/O workload.

Sizing for a Light I/O Workload

In this first exercise, we configure a cluster to satisfy 150 VMs. The characteristics of the VMs are described in Table 4. This workload reaches 30 percent of the effective maximum I/O requests that we observed in our ramp-up test with the 800 GB NAND Flash SSD. A 4-node cluster with the characteristics described in Table 5 can satisfy the workload requirements.

Table 4. Light I/O Workload VM Characteristics

Workload Requirements	Per VM	Total
#VMs	N/A	150
Total IOPS–MB/s	360 I/O requests (70/30 32 KB)	54,000/1,912
Usable Storage	250 GB	38.8 TB
CPU Cores/GHz	4 vCPUs	154 cores/325 GHz
Memory	10 GB	1,500 GB

Considerations About the NAND Flash Option

The bandwidth of the MU SAS drives is sufficient to satisfy this light I/O workload. However, the endurance of 3 DWDP for this 800 GB NAND Flash SSD may be inadequate for this workload because it would write each of the 8x cache drives 13.5X per day.

When sizing cache drives for performance, we use the Peak I/O requests metric. But when sizing for endurance, we should use the average write activity instead. The 800 GB NAND Flash SSD can be adequate for this workload, if the average write activity is 4.5X lower than the value used as input (in this case, the total IOPS per MB/s—360 per VM). Otherwise, it increases the risk of data loss in this configuration.

Considerations About the Intel Optane SSD Option

Intel Optane SSDs are more expensive than NAND Flash SSDs, and in this case using the Intel Optane SSD increases the host cost by 4.5 percent.[10] However, the Intel Optane SSD provides up to 2X higher I/O throughput.[11] This difference in performance future-proofs the cluster so that it can absorb larger I/O workloads than initial planned.

The Intel Optane SSD endurance is 16X higher than the NAND Flash SSD's endurance;[12] it far exceeds the 5-year warranty I/O attributes of this light I/O workload, minimizing the risk of data loss due to drive wear-out failures.

Table 5. Configuration Options for a Light I/O Workload

Host Components	NAND Flash SSD Option	Intel® Optane™ SSD Option
CPU	2x Intel® Xeon® Gold 6330N processor (28 cores, 2.2 GHz)	
Memory		
- DRAM	256 GB (16x 16 GB, 3200 MT/s)	
- Intel® Optane™ PMem 200 series	1024 GB (8 x 128 GB)	
Storage		
- Capacity Drives	4x read-intensive SAS drives, 7.68 TB	
- Cache Drives	2x MU SAS drives, 800 GB, 3 DWPD	2x Intel® Optane™ SSD P5800X, 400 GB, 100 DWPD[13]
Other		
Network	1x 25 GbE Network Interface Card	
Number of Hosts	4	4

Sizing for a Moderate I/O Workload

In this second exercise, we continue to configure a cluster to satisfy 150 VMs. But in this use case, the workload reaches 40 percent of the effective maximum I/O requests that we observed in our ramp-up test with the Intel Optane SSD. The characteristics of the VMs are described in Table 6.

In this case, the cost-optimized configuration when using NAND Flash SSDs as cache consists of a 7-node cluster, and the configuration with Intel Optane SSDs as cache remains a 4-node cluster (see Table 7).

Table 6. Moderate I/O Workload VM Characteristics

Workload Requirements	Per VM	Total
#VMs	N/A	150
Total IOPS–MB/s	960 I/O requests (70/30 32 KB)	144,000/4,500
Usable Storage	100 GB	16.6 TB
CPU Cores/GHz	4 vCPUs	154 cores/325 GHz
Memory	8 GB	1,200 GB

Considerations About the NAND FLASH Option

This I/O workload requires fourteen MU SAS 800 GB drives to sustain adequate performance behavior. Seven hosts are configured because the best I/O performance is with two disk groups per host. Because of the larger number of hosts, a single-socket configuration is used to optimize costs. The larger number of hosts has a negative effect on footprint and power/heating.

In this use case, each of the fourteen 800 GB NAND Flash SSDs (3 DWPD) is expected to be written 20x per day, making it substantially inadequate for this configuration. The risks associated with premature drive replacement (prior to the 5-year warranty expiration) increase substantially.

Considerations About the Intel Optane SSD Option

Because the Intel Optane SSD can perform 2X more I/O requests per host than the compared NAND Flash SSD, it is possible to consolidate hosts (a 40 percent lower footprint).

If we had used only DRAM (1 TB DRAM per host) in the Intel Optane SSD option, this option would be 6 percent less expensive[14] and would still be able to support twice the I/O workload requirements. However, because host consolidation increases memory per host, this option makes the use of Intel Optane PMem more attractive, which in this case would enable a 13 percent total cost reduction.[15]

The Intel Optane SSD with 100 DWPD can sustain this workload for more than five years, minimizing the risk of data loss and operational impact due to drive wear-out failures.

Table 7. Configuration Options for a Moderate I/O Workload

Host Components	NAND Flash SSD Option	Intel® Optane™ SSD Option
CPU	1x Intel® Xeon® Gold 6330N processor (28 cores @ 2.2 GHz)	2x Intel Xeon Gold 6330N processor (28 cores @ 2.2 GHz)
Memory		
- DRAM	384 GB (8x 16 GB + 8x 32 GB, 3200 MT/s)	256 GB (16x 16 GB, 3200 MT/s)
- Intel® Optane™ PMem 200 series	N/A	1024 GB (8x 128 GB)
Storage		
- Capacity Drives	6x read-intensive SAS 3.84 TB	4x read-intensive SAS 7.68 TB
- Cache Drives	2x 800 GB MU SAS drives, 3 DWPD	2x Intel® Optane™ SSD P5800X, 400 GB, 100 DWPD
Other		
Network	1x 25 GbE Network Interface Card	1x 25 GbE Network Interface Card
Number of Hosts	7	4

Conclusion

This study has demonstrated that when using Intel Optane SSDs as cache, vSAN reached a 2X higher performance than the tested NAND Flash SSD, due to the drive bandwidth capabilities when processing mixed read/write I/O loads. This higher bandwidth creates opportunities for host consolidation, resulting in cost and footprint savings. It also establishes a much more predictable performance environment, which is important when running mission-critical applications with strict service-level agreements.

We also discussed endurance and illustrated the importance of going beyond general endurance guidelines when choosing a cache drive. The maximum write I/O workload that a drive supports is an important consideration to minimize risks of data loss and operational impact due to wear-out failures.

Learn More

You may find the following resources helpful:

- Best Practices for Deploying a Cost-efficient, High-performance vSAN Cluster
- Intel® Optane™ SSDs
- Supercharge Your VMware vSAN Clusters with the Intel® Optane™ SSD P5800X blog

Find the solution that is right for your organization. Contact your Intel representative or visit Intel Optane SSDs for Data Center.

[1] Testing by Intel as of September 28, 2021.
Common Configuration: 4 nodes, 2x Intel® Xeon® Gold 6348 processor (28 cores, 2.6 GHz), total DRAM 256 GB (16 slots/16 GB/3200 MT/s), storage (capacity) = 8x Intel® SSD D7-P5510 3.84 TB, 1x Intel® Ethernet Network Adapter E810-CQDA2 (100 GbE), Intel® Hyper-Threading Technology = ON, Intel® Turbo Bost Technology = ON, VMware vSAN 7.0U2 with RAID 1.
NAND Flash SSD Cache Configuration: 2x Samsung PM1645 800 GB MU (Mixed Use) SAS.
Intel® Optane™ SSD Cache Configuration: 2x Intel Optane SSD P5800X 400 GB NVMe.

[2] A NAND-based 800 GB drive with an endurance rating of 3 DWPD will last five years only if the sustained write workload is lower than 2,400 GB per day or 28.4 write MB/s. Put another way, even a light sustained I/O load such as 450 32-KB write IOPS activity would fill the drive three times per day, assuming the use of RAID 1 as data protection. In contrast, an Intel® Optane™ SSD P5800X 400 GB drive has an endurance rating of 100 DWPD. It can sustain 40,000 GB written per day or 470 MB/s written per day for five years. That equates to a 16X increase in endurance.

[3] See endnote 1.

[4] Front-end requests come from the VMs. Back-end requests represent all the physical I/O activity performed by VMware vSAN.

[5] **PCI Gen4 1.6 TB NAND MU flash drive specifications:** https://www.samsung.com/semiconductor/ssd/enterprise-ssd/MZPLJ1T6HBJR/

SAS 800 GB MU drive specifications: https://www.dell.com/en-us/shop/dell-800-gb-solid-state-drive-serial-attached-scsi-sas-mixed-use-12gbps-512e-25-inch-hot-plug-drive-pm1645-3-dwpd-4380-tbw-ck/apd/400-azii/storage-drives-media

400 GB Intel® Optane™ SSD P5800X specifications: https://www.intel.com/content/www/us/en/products/sku/201861/intel-optane-ssd-dc-p5800x-series-400gb-2-5in-pcie-x4-3d-xpoint/specifications.html

[6] See endnote 1.

[7] Intel, "Deploying a Cost-Efficient, High-Performance vSAN Cluster," https://www.intel.com/content/www/us/en/architecture-and-technology/deploying-vsan-cluster-best-practices-guide

[8] See endnote 1.

[9] Testing by Intel as of May 10, 2021.

Intel® Optane™ SSD Configuration: 4 nodes, 2x Intel® Xeon® Gold 6348 processor (28 cores, 2.6 GHz), total memory = 256 GB (16 slots/16 GB/3200 MT/s), Intel® Hyper-Threading Technology = ON, Intel® Turbo Boost Technology = ON, 2x Intel® Optane™ SSD P5800X (cache) 400 GB and 8x Intel® SSD D7-P5510 3.84 TB (capacity), 1x Intel® Ethernet Adapter E810C 100 GbE, BIOS = 2.1 (ucode = 05003003), VMware vSphere 7.0U2, HCIBench 2.5.3.

Intel® SSD D7-5600 Configuration: 4 nodes, 2x Intel® Xeon® Gold 6348 processor (28 cores, 2.6 GHz), total memory = 256 GB (16 slots/16 GB/3200 MT/s), Intel Hyper-Threading Technology = ON, Intel Turbo Boost Technology = ON, 2x Intel® SSD D7-P5600 (cache) 1.6 TB and 8x Intel SSD D7-P5510 3.84 TB (capacity), 1x Intel Ethernet Adapter E810C 100 GbE, BIOS = 2.1 (ucode = 0x8d055260), VMware vSphere 7.0U2, HCIBench 2.5.3.

[10] Price at current list prices as of November 1, 2021. Prices change frequently. Your costs and results may vary. Prices obtained from https://www.dell.com/en-us/work/shop/cty/pdp/spd/poweredge-r750/pe_r750_14794_vi_vp?configurationid=19732e1e-1fe3-419e-8b40-d091f06c3f51.

[11] See endnote 1.

[12] See endnote 2.

[13] See endnote 2.

[14] See endnote 11.

[15] See endnote 11.