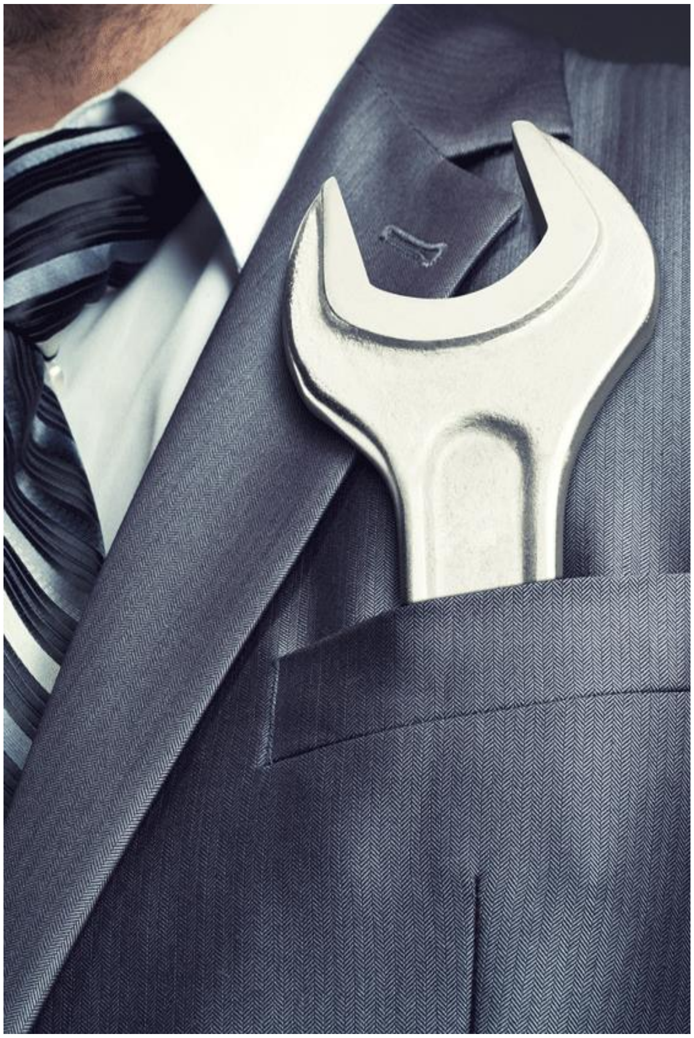
Huge Benefits With All-Flash and vSAN 6.5

VMware's New vSAN 6.5 Shows Substantial Gains With All-Flash Configurations



Overview

According to VMware, their release of vSAN 6.5 brings real, tangible benefits, especially with all- flash configurations[1].

We wanted to measure these benefits on our own Micron Accelerated Solutions all-flash vSAN platforms. These platforms are built on industry-standard systems that are high value, flash optimized, fully tested and documented.

As part of that value proposition, we tested the small, random IOPS and average latency benefits vSAN 6.5 brings.

With the release of vSAN 6.5, VMware enables substantial improvements in both 4K IOPS and average latency across both the cache and capacity tiers. While a 100% read IO profile showed no real difference (in either test), all other tested IO profiles showed double-digit benefits (higher IOPS and lower average latency).

In this tech brief, we show our test results and compare vSAN 6.5 advantages (on the same platform).

Testing IOPS and Latency

Small, random IOs are common in virtualized deployments due to the "IO blender effect" associated with virtualization[2]. Because of this, exceptional small, random IOPS performance and low latency across a variety of IO profiles (workloads) is essential to ensure good virtual machine (VM) responsiveness.

We wanted to determine what (if any) IOPS and average latency improvements could be found in vSAN 6.5 compared to 6.2 when deployed on an all-flash configuration. We tested 4K IOPS and average latency across five IO profiles using a four-node, all-flash system with vSAN 6.5, and again with vSAN 6.2. We used VMware's HCIBench[3] test tool. (HCIBench is a publicly available test tool that leverages the ability to deploy large virtual machine configurations with very specific working sets and workload types. Because of its deterministic approach, HCIBench enables consistent, repeatable performance measurement so direct comparisons can be made among different configurations.)

vSAN uses a two-tiered storage mechanism. Each node has a cache tier and a capacity tier. With some deployments, the vast majority of the IOs may reside in the cache tier; with others, the IOs may span the cache tier and the capacity tier.

Because we wanted to see how vSAN 6.5 might affect either type of deployment, we tested the capabilities of each vSAN version using two working sets:

> **Cache Tier Only Testing:** We used a working set equivalent to 50% of the cache tier capacity. This ensured that the working set fit entirely within the cache tier and that the capacity tier was not involved in testing.
>
> **Capacity Tier (with Cache Tier) Testing:** To test across both the cache and capacity tiers, we used a working set equivalent to 75% of the total vSAN capacity. This extended the test data well beyond any caching, reflecting the performance of the complete architecture.

For each IO profile, we compared both the 4K IOPS and average latency between vSAN 6.2 and 6.5. In each of the sections that follow, we show measured IOPS and average latency, as well as the percentage difference from 6.2 to 6.5.

Cache Tier: 4K Random IOPS and Average Latency

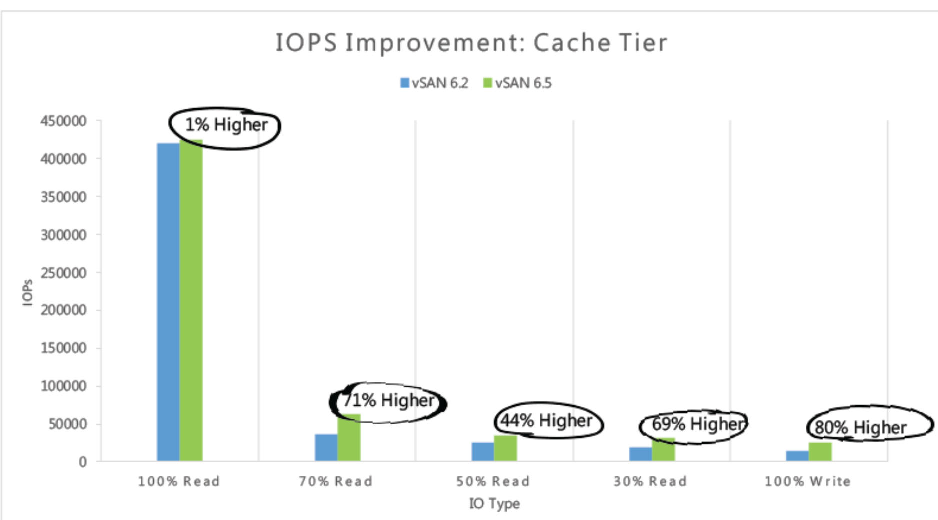


Figure 1a: Cache Tier IOPS

Figure 1a shows the IOPS difference across the five measured IO profiles.

The vSAN 6.2 data is shown in blue, the vSAN 6.5 data in green, with the IOPS along the left axis.

For each IO profile, the percentage difference is circled. While the difference in the read only IO profile is negligible, the differences in all other measured IO profiles are not.

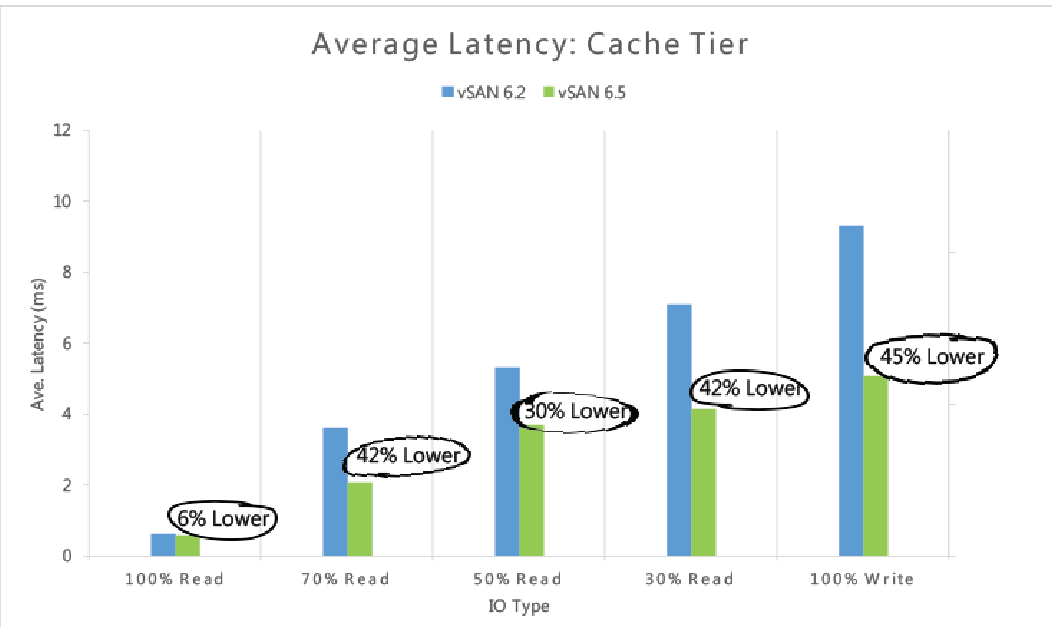
From Figure 1a we can see that the IOPS performance for 100% read traffic is essentially the same with both vSAN versions (the difference is about 1%).

Note that for all other IO loads, the improvement with 6.5 is much greater, as summarized in Table 1.

IO Load	vSAN 6.5 IOPS Improvement
100% Read / 0% Write	1% Higher
70% Read / 30% Write	71% Higher
50% Read / 50% Write	44% Higher
30% Read / 70% Write	69% Higher
0% Read / 100% Write	80% Higher

Table 1: Cache Tier IOPS Summary

Average latency showed similar results. Figure 1b shows the average latency differences for the same test and IO profiles. While the 100% read IO profile showed little difference (just 6%), the mixed IO loads shows substantial improvements, as summarized in Table 2.



IO Load	vSAN 6.5 Average Latency Improvement
100% Read / 0% Write	6% Lower
70% Read / 30% Write	42% Lower
50% Read / 50% Write	30% Lower
30% Read / 70% Write	42% Lower
0% Read / 100% Write	45% Lower

Table 2: Cache Tier Average Latency Summary

Capacity Tier: 4K Random IOPS and Average Latency

As noted earlier, this test uses a data set equal to about 75% of the vSAN's total capacity. This ensures that the cache tier and capacity tier are both tested and reflects the overall architecture performance with larger deployments.

Figure 2a shows the IOPS difference across the five measured IO profiles with this larger data set. In these figures the vSAN 6.2 data is again shown in blue and the vSAN 6.5 data in green. IOPS are along the left axis. For each IO profile, the percentage difference is also shown (circled). Figure 2b shows the average latency differences for the same test and IO profiles.

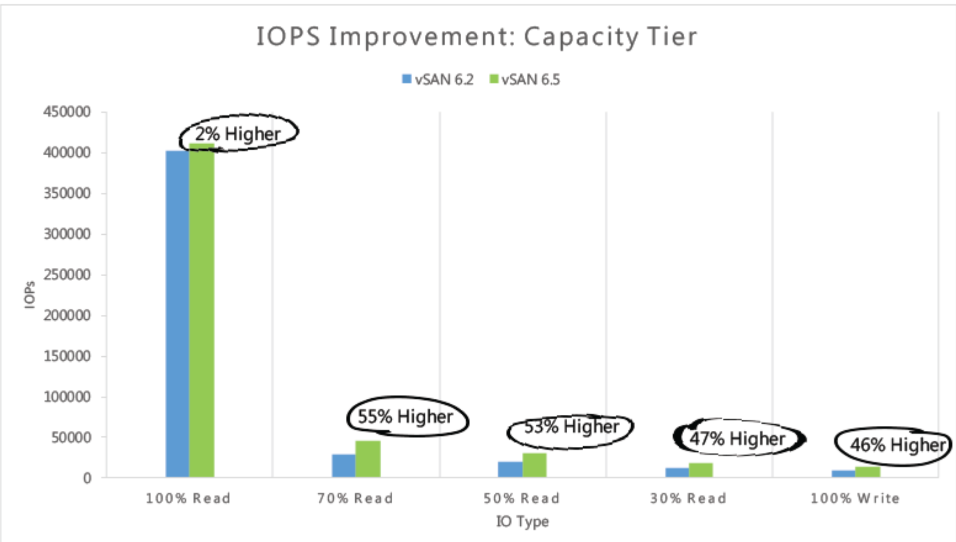


Figure 2a: Capacity Tier IOPS

From Figure 2a we can see that the IOPS performance for 100% read traffic is again essentially the same with both vSAN versions (the difference is about 2%). Note that for all other IO loads the improvement with 6.5 is much greater, as summarized in Table 3.

IO Load	vSAN 6.5 IOPS Improvement
100% Read / 0% Write	2% Higher
70% Read / 30% Write	55% Higher
50% Read / 50% Write	53% Higher
30% Read / 70% Write	47% Higher
0% Read / 100% Write	46% Higher

Table 3: Capacity Tier IOPS Summary

Average latency showed similar results as seen in Figure 2b. While the 100% read IO profile showed little difference (just 2%), the mixed IO loads shows substantial improvements, as summarized in Table 4.

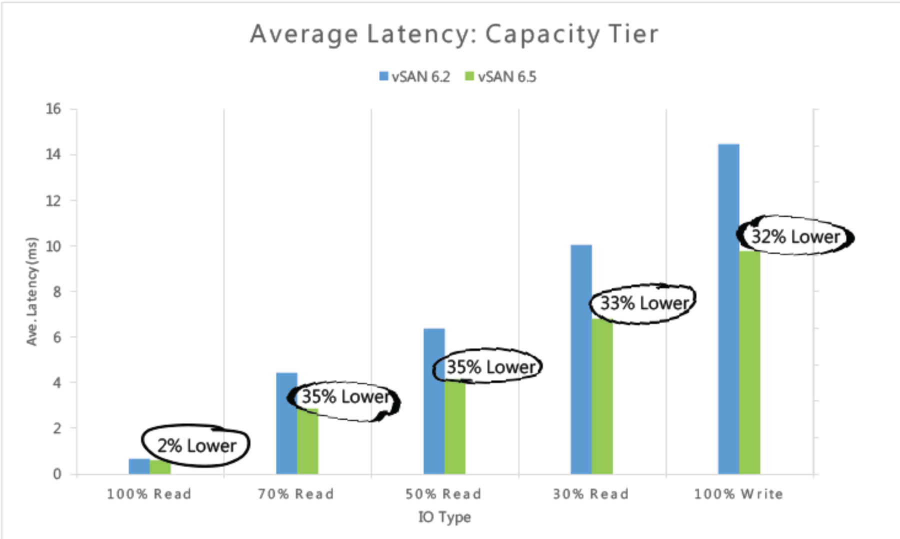


Figure 2b: Capacity Tier Average Latency

IO Load	vSAN 6.5 Average Latency Improvement
100% Read / 0% Write	2% Lower
70% Read / 30% Write	35% Lower
50% Read / 50% Write	35% Lower
30% Read / 70% Write	34% Lower
0% Read / 100% Write	32% Lower

Table 4: Capacity Tier Average Latency Summary

Why Is This Important?

With the release of vSAN 6.5, VMware enables substantial improvements in both 4K IOPS and average latency across both the cache and capacity tiers. While a 100% read IO profile showed no real difference (in either test), all other tested IO profiles showed double-digit benefits (higher IOPS and lower average latency).

These small, random IOs are typical of virtualized environments due to the IO blender effect. Enhancing IOPS performance and average latency can bring substantial benefits to your VMs.

vSAN 6.5 is a large step forward in that direction, providing substantially greater benefit with all-flash configurations like ours (compared to vSAN 6.2).

Check Micron's Accelerated Solutions page for additional details on our Virtual SAN Ready Nodes.

1. See VMware's web site for additional details
2. For additional information on the IO blender effect, see this VMware article
3. For additional information on HCIBench, see the VMware Labs site