Performance Testing

First Published On: 11-05-2016
Last Updated On: 11-07-2016

Table of Contents

1. vSAN Performance Testing

Performance Testing

1.1 Performance Testing Overview

Performance Testing

Performance testing is an important part of evaluating any storage solution. Setting up a desirable test environment could be challenging, and customers may do it differently. Customers may also select from a variety of tools to run workloads, or choose to collect data and logs in different ways. These all add complexity to troubleshoot performance issues claimed by customers, and lengthen the evaluation process.

vSAN Performance will depend on what devices are in the hosts (SSD, magnetic disks), on the policy of the virtual machine (how widely the data is spread across the devices), the size of the working set, the type of workload, and so on.

A major factor for virtual machine performance is the virtual hardware: how many virtual SCSI controllers, VMDKs, outstanding I/O and how many vCPUs can be pushing I/O. Use a number of VMs, virtual SCSI controllers and VMDKs for maximum performance.

vSAN's distributed architecture dictates that reasonable performance is achieved when the pooled compute and storage resources in the cluster are well utilized. This usually means a number of VMs each running the specified workload should be distributed in the cluster and run in a consistent manner to deliver aggregated performance. vSAN also depends on VSAN Observer for detailed performance monitoring and analysis, which as a separate tool is easy to become an afterthought of the testing.

Use vSAN Observer

vSAN ships with a performance-monitoring tool called vSAN Observer. It is accessed via RVC – the Ruby vSphere Console. If you're planning on doing any sort of performance testing, plan on using vSAN Observer to observe what's happening.

Reference VMware Knowledgebase Article 2064240 for getting started with VSAN Observer – *http://kb.vmware.com/kb/2064240*. See detailed information in

Monitoring VMware vSAN with VSAN Observer.

Performance Considerations

There are a number of considerations you should take into account when running performance tests on vSAN.

Single vs. Multiple Workers

vSAN is designed to support good performance when many VMs are distributed and running simultaneously across the hosts in the cluster. Running a single storage test in a single VM won't reflect on the aggregate performance of a vSAN-enabled cluster. Regardless of what tool you are using – IOmeter, VDbench or something else – plan on using multiple "workers" or I/O processors to multiple virtual disks to get representative results.

Working Set

For the best performance, a virtual machine's working set should be mostly in cache. Care will have to be taken when sizing your vSAN flash to account for all of your virtual machines' working sets residing in cache. A general rule of thumb is to size cache as 10% of your consumed virtual machine storage (not including replica objects). While this is adequate for most workloads, understanding your workload's working set before sizing is a useful exercise. Consider using VMware Infrastructure Planner (VIP) tool to help with this task – http://vip.vmware.com.

Sequential Workloads versus Random Workloads

- Sustained sequential write workloads (such as VM cloning operations) run on vSAN will simply fill the cache and future writes will need to wait for the cache to be destaged to the spinning magnetic disk layer before more I/Os can be written to cache, so performance will be a reflection of the spinning disk(s) and not of flash. The same is true for sustained sequential read workflows. If the block is not in cache, it will have to be fetched from spinning disk. Mixed workloads will benefit more from vSAN's caching design.

Outstanding IOs

Most testing tools have a setting for Outstanding IOs, or OIO for short. It shouldn't be set to 1, nor should it be set to match a device queue depth. Consider a setting of between 2 and 8, depending on the number of virtual machines and VMDKs that you plan to run. For a small number of VMs and VMDKs, use 8. For a large number of VMs and VMDKs, consider setting it lower.

Block Size

The block size that you choose is really dependent on the application/workload that you plan to run in your VM. While the block size for a Windows Guest OS varies between 512 bytes and 1MB, the most common block size is 4KB. But if you plan to run SQL Server, or MS Exchange workloads, you may want to pick block sizes appropriate to those applications (they may vary from application version to application version). Since it is unlikely that all of your workloads will use the same block size, consider a number of performance tests with differing, but commonly used, block sizes.

Cache Warming Considerations

Flash as cache helps performance in two important ways. First, frequently read blocks end up in cache, dramatically improving performance. Second, all writes are committed to cache first, before being efficiently destaged to disks – again, dramatically improving performance.
However, data still has to move back and forth between disks and cache. Most real-world application workloads take a while for cache to "warm up" before achieving steady-state performance.

Number of Magnetic Disk Drives in Hybrid Configurations

In the getting started section, we discuss how disk groups with multiple disks perform better than disk groups with fewer, as there are more disk spindles to destage to as well as more spindles to handle read cache misses. Let's look at a more detailed example around this.

Consider a vSAN environment where you wish to clone a number of VMs to the vSAN datastore. This is a very sequential I/O intensive operation. We may be able to write into the SSD write buffer at approximately 200-300 MB per second. A single magnetic disk can maybe do 100MB per second. So assuming no read operations are taking place at the same time, we would need 2-3 magnetic disks to match the SSD speed for destaging purposes.

Now consider that there might also be some operations going on in parallel. Let's say that we have another vSAN requirement to achieve 2000 read IOPS. vSAN is designed to achieve a 90% read cache hit rate (approximately). That means 10% of all reads are going to be read cache misses; for example, that is 200 IOPS based on our requirement. A single magnetic disk can perhaps achieve somewhere in the region of 100 IOPS. Therefore, an additional 2 magnetic disks will be required to meet this requirement.

If we combine the destaging requirements and the read cache misses described above, your vSAN design may need 4 or 5 magnetic disks per disk group to satisfy your workload.

Striping Considerations

One of the VM Storage Policy settings is *NumberOfDiskStripesPerObject*. That allows you to set a stripe width on a VM's VMDK object. While setting disk striping values can sometimes increase performance, that isn't always the case.

As an example, if a given test is cache-friendly (e.g. most of the data is in cache), striping won't impact performance significantly. As another example, if a given VMDK is striped across disks that are busy doing other things, not much performance is gained, and may actually be worse.

Guest File Systems Considerations

Many customers have reported significant differences in performance between different guest file systems and their settings; for example, Windows NTFS and Linux. If you are not getting the performance you expect, consider investigating whether it could be a guest OS file system issue.

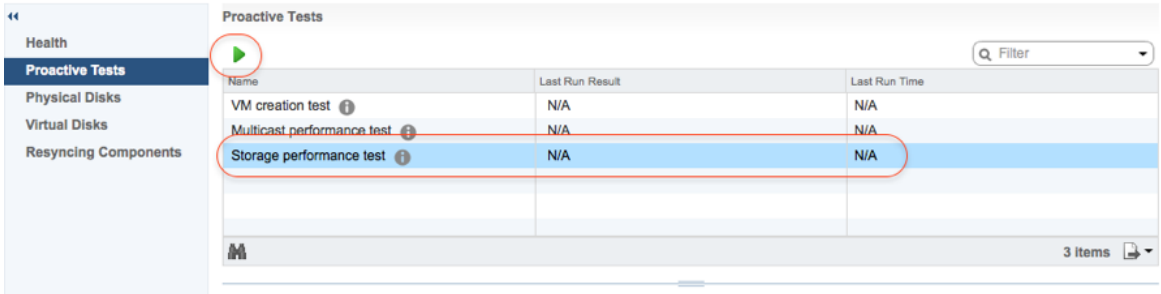
Performance during Failure and Rebuild

When vSAN is rebuilding one or more components, application performance can be impacted. For this reason, always check to make sure that vSAN is fully rebuilt and that there are no underlying issues prior to testing performance. Verify there are no rebuilds occurring before testing with the following RVC command, which we discussed earlier:

- **vsan.check_state**
- **vsan.disks_stats**
- **vsan.resync_dashboard**

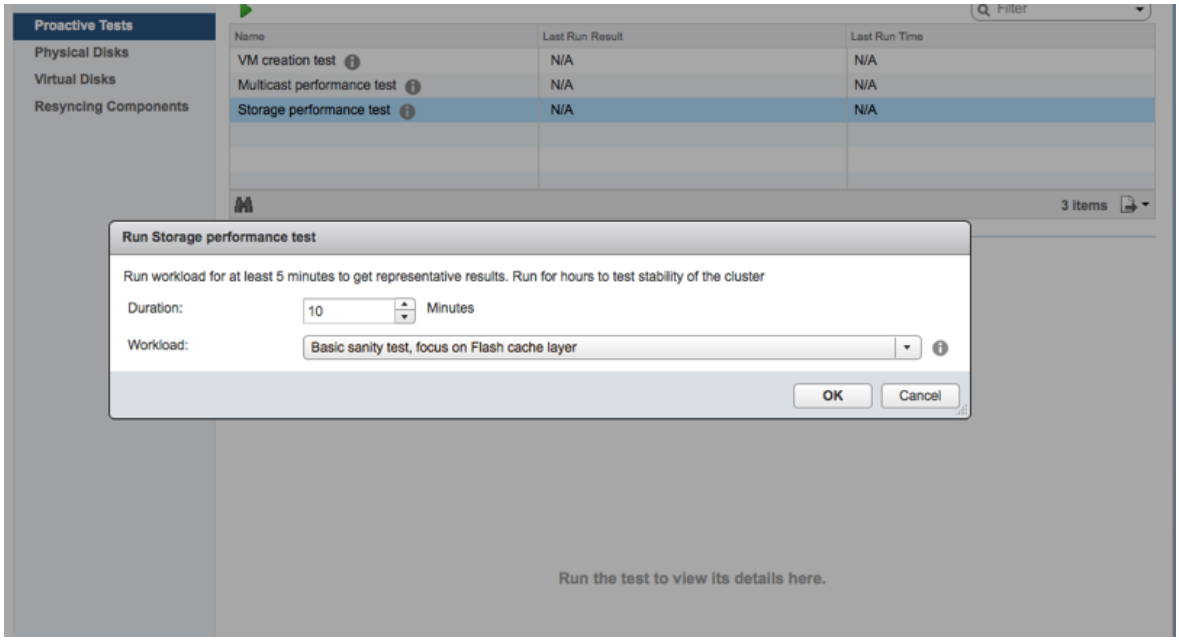
Performance Testing Option 1: vSAN Health Check

vSAN Health Check comes with its own Storage Performance Test. This negates the need to deploy additional tools to test the performance of your vSAN environment. To run the storage performance test is quite simple; navigate to the cluster's Monitor tab > vSAN > Proactive Tests, select Storage Performance Test, then click on the Go arrow highlighted below.



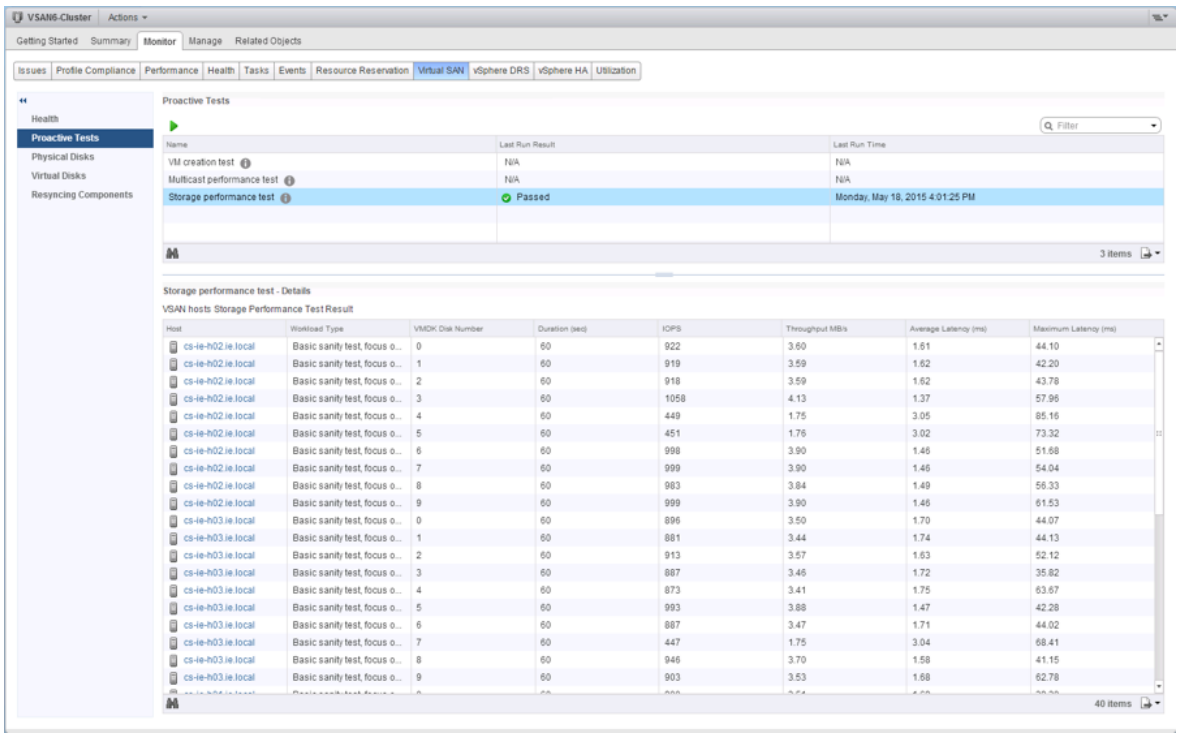
A popup is then displayed, showing the duration of the test (default 10 minutes) along with the type of workload that will be run. The user can change this duration, for example, if a burn-in test for a longer period of time is desired.

There are a number of different workloads that can be chosen from the drop-down menu.



To learn more about the test that is being run, click on the (i) symbol next to the workload. This will describe the type of workload that the test will initiate.

When the test completed, the Storage Load Test results are displayed, including test name, workload type, IOPS, throughput, average latency and maximum latency. Keep in mind that a sequential write pattern will not benefit from caching, so the results that are shown from this test are basically a reflection of what the capacity layer (in this case, the magnetic disks) can do.



The proactive test could then be repeated with different workloads

As before, when the test completes, the results are once again displayed. You will notice a major difference in results when the workload can leverage the caching layer versus when it cannot.

Performance Testing Option 2: HCIbench

In a hyperconverged architecture, each server is intended to support both many application VMs, as well as contribute to the pool of storage available to applications. This is best modeled by invoking many dozens of test VMs, each accessing multiple stored VMDKs. The goal is to simulate a very busy cluster.

Unfortunately, popular storage performance testing tools do not directly support this model. As a result performance testing a hyperconverged architecture such as vSAN presents a different set of challenges. To accurately simulate workloads of a production cluster it is best to deploy multiple VMs dispersed across hosts with each VM having multiple disks. In addition, the workload test needs to be run against each VM and disk simultaneously.

To address the challenges of correctly running performance testing in hyperconverged environments, VMware has created a storage performance testing automation tool called HCIbench that automates the use of the popular Vdbench testing tool. Users simply specify the parameters of the test they would like to run, and HCIbench instructs Vdbench what to do on each and every node in the cluster.

HCIbench aims to simplify and accelerate customer Proof of Concept (POC) performance testing in a consistent and controlled way. The tool fully automates the end-to-end process of deploying test VMs, coordinating workload runs, aggregating test results, and collecting necessary data for troubleshooting purposes. Evaluators choose the profiles they are interested in; HCIbench does the rest quickly and easily.

This section provides an overview and recommendations for successfully using HCIbench. For complete documentation and use procedures, refer to the HCIbench Installation and User guide which is accessible from the download directory.

Where to Get HCIbench

HCIbench and complete documentation can be downloaded from the following location: HCIbench Automated Testing Tool.

This tool is provided free of charge and with no restrictions. Support will be provided solely on a best-effort basis as time and resources allow, by the VMware vSAN Community Forum.

Deploying HCIbench

Step 1 – Deploy the OVA

To get started, you deploy a single HCIbench appliance called *HCIbench.ova*. The process for deploying the HCIbench OVA is no different from deploying any other OVA.

Step 2 – HCIbench Configuration

After deployment, navigate to http://Controller_VM_IP:8080/ to start configuration and kick off the test.

There are three main sections in this configuration file:

- vSAN cluster and host information
- vdBench Guest VM Specification
- Workload Definitions

For detailed steps on configuring and using HCIbench refer to the HCIbench User Guide.

Considerations for Defining Test Workloads

Working set

Working set is one of the most important factors for correctly running performance test and obtaining accurate results. For the best performance, a virtual machine's working set should be mostly in cache. Care will have to be taken when sizing your vSAN flash to account for all of your virtual machines' working sets residing in cache. A general rule of thumb is to size cache as 10% of your consumed virtual machine storage (not including replica objects). While this is adequate for most workloads, understanding your workload's working set before sizing is a useful exercise. Consider using VMware Infrastructure Planner (VIP) tool to help with this task – http://vip.vmware.com.

The following process is an example of sizing an appropriate working set for performance testing with HCIbench. Consider a four node cluster with one 400GB SSD per node. This gives the cluster a total cache size of 1.6TB. The total cache available in vSAN is split 70% for read cache and 30% for write cache. This gives the cluster in our example 1120GB of available read cache and 480GB of available write cache. In order to correctly fit the HCIbench within the available cache, the total capacity of all VMDKs used for I/O testing should not exceed 1,120GB.

Designing a test scenario with 4 VMs per host, each VM having 5 X 10GB VMDKs, resulting in a total size of 800GB. This will allow the test working set to fit within cache. The default setting for the number of data disks per VM is 2 and the default size of data disks is 5GB. These values should be adjusted so that the total number of VMs multiplied by the number of data disks per VM multiplied by the size of data disk is less than the size of SSDs multiplied by 70% (read cache in hybrid mode) multiplied by the number of disk groups per host multiplied by the number of hosts. That is:

Number of VMs * Number of Data Disk * Size of Data Disk < Cache tier SSD capacity * 70% read cache (hybrid) * Disk Groups per Host * Number of Hosts

To see the example mathematically:

4 VMs * 5 Data Disks * 10GB = 800GB,

400GB SSDs * 70% * 1 Disk Group per Host * 4 Hosts = 1,120GB

800GB working set size < 1,120GB read cache in cluster

That last statement is true, so this is an acceptable working set for the configuration (and vice versa).

Sequential workloads versus random workloads

Before doing performance tests it is important to understand the performance characteristics of the production workload to be tested. Different applications have different performance characteristics. Understanding these characteristics is crucial to successful performance testing. When it is not possible to test with the actual application or application specific testing tool it is important to design a test which matches the production workload as closely as possible. Different workload types will perform differently on vSAN.

- Sustained sequential write workloads (such as VM cloning operations) run on vSAN will simply fill the cache and future writes will need to wait for the cache to be destaged to the spinning magnetic disk layer before more I/Os can be written to cache, so performance will be a reflection of the spinning disk(s) and not of flash. The same is true for sustained sequential read workflows. If the block is not in cache, it will have to be fetched from spinning disk. Mixed workloads will benefit more from vSAN's caching design.
- HCIbench allows you to change the percentage read and the percentage random parameters. As a starting point it is recommended to set the percentage read parameter to 70 and the percentage random parameter to 30%.

Initializing Storage

During configuration of the workload the recommendation is to select the option to initialize storage. This option will zero the disks for each VM being used in the test, helping to alleviate a first write penalty during the performance testing phase.

Test Run Considerations

As frequently read blocks end up in cache, read performance will improve. In a production environment active blocks will already be in cache. When running any kind of performance testing it is important to keep this in mind. As a best practice performance tests should include at least a 15 minute warm up period. Also keep in mind that the longer testing runs the more accurate the results will be. In addition to the cache warming period HCIbench tests should be configured to for at least an hour.

Results

After the Vdbench testing is completed, the test results are collected from all Vdbench instances in the test VMs. And you can view the results at **http://Controller_VM_IP/results** in a web browser. You can find all of the original result files produced by Vdbench instances inside the subdirectory corresponding to a test run. In addition to the text files, there is another subdirectory named iotest-vdbench-<*VM#*>vm inside, which is the statistics directory generated by vSAN Observer. vSAN performance data can be viewed by opening the stats.html file within the test directory.