



What's New with VMware Virtual SAN 6.2

Technical White Paper

February 2016

Contents

Introduction.....	3
Architecture Overview	3
Space Efficiency.....	4
Deduplication and Compression.....	4
RAID-5/6 (Erasure Coding).....	6
Swap Efficiency.....	7
Quality of Service	8
Performance Service.....	8
Cluster Metrics	9
Host Metrics	11
Virtual Machine Metrics.....	13
Health Service	15
Application Support	17
Summary	17
About the Authors.....	17

Introduction

VMware Hyper-Converged Software™ transforms industry-standard x86 servers and directly attached storage into radically simple Hyper-Converged Infrastructure (HCI) to help eliminate high IT costs, management complexity, and performance limitations. VMware Hyper-Converged Software enables the lowest cost and highest performance HCI solutions available. The tightly integrated software stack includes VMware vSphere®, the market-leading hypervisor; VMware Virtual SAN™, radically simple, enterprise-class native storage; and VMware vCenter Server™, a unified and extensible management solution.

Virtual SAN is natively integrated with vSphere and it is configured with just a few mouse clicks. Since disks internal to the vSphere hosts are used to create a Virtual SAN datastore, there is no dependency on external shared storage. Virtual machines can be assigned specific storage policies based on the needs of the applications. These workloads benefit from dependable shared storage with predictable performance and availability characteristics.

The new release of Virtual SAN, version 6.2, further reduces TCO by delivering up to 10x greater storage efficiency. Virtual SAN 6.2 is optimized for modern all-flash storage, delivering efficient near-line deduplication, compression, and erasure coding capabilities that enable high performance all-flash systems for as low as \$1 per GB of usable capacity—up to 50% less than the cost of lower-performing hybrid solutions from the leading competition.

The following list represents the primary new features in Virtual SAN 6.2:

- Deduplication, compression, and erasure coding
- Quality of Service (QoS) using IOPS limits
- New performance service and improved health service

This document provides a brief Virtual SAN architecture overview and covers the new features introduced in Virtual SAN 6.2.

Architecture Overview

Virtual SAN clusters contain two or more physical hosts that contain either a combination of magnetic disks and flash devices (hybrid configuration) or all flash devices (all-flash configuration) that contribute cache and capacity to the Virtual SAN distributed datastore.

In a hybrid configuration, one flash device and one or more magnetic drives are configured as a disk group. A disk group can have up to seven magnetic drives. One or more disk groups are utilized in a vSphere host depending on the number of flash devices and magnetic drives contained in the host. Flash devices serve as read-and-write cache for the Virtual SAN datastore while magnetic drives make up the capacity of the datastore. By default, Virtual SAN will use 70% of the flash capacity as read cache and 30% as write cache.

For all-flash configurations, the flash device(s) in the cache tier are used for write caching only (no read cache) as read performance from the capacity flash devices is more than sufficient. Two different grades of flash devices are commonly used in an all-flash Virtual SAN configuration: Lower capacity, higher endurance devices for the cache layer and more cost effective, higher capacity, lower endurance devices for the capacity layer. Writes are performed at the cache layer and then de-staged to the capacity layer, only as needed. This helps extend the usable life of the lower endurance flash devices in the capacity layer.

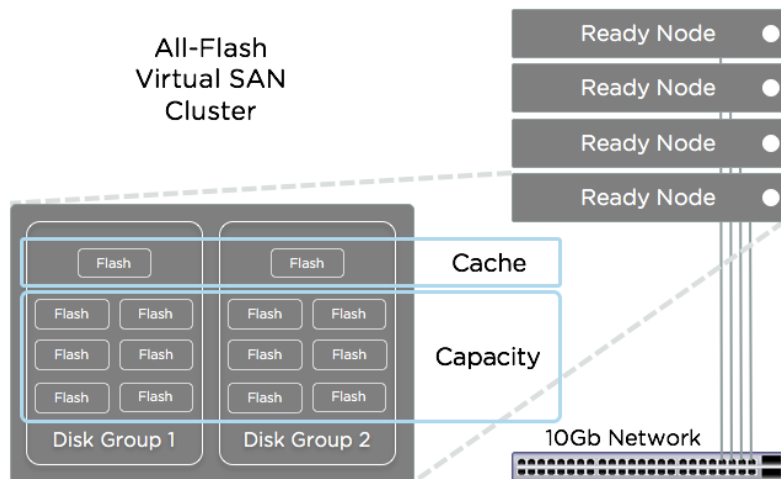


Figure 1. All-Flash Virtual SAN Cluster

When considering hardware configurations for a Virtual SAN cluster, the easiest approach in many cases is using Virtual SAN Ready Nodes. These servers are preconfigured and installed with vSphere and Virtual SAN to help ensure compatibility and faster implementation times. Another option is to build a custom solution using jointly validated components from a number of OEM vendors. The [Virtual SAN Hardware Quick Reference Guide](#) provides some sample server configurations as directional guidance and all components should be validated using the [VMware Compatibility Guide for Virtual SAN](#).

Space Efficiency

Virtual SAN 6.2 introduces space efficiency technologies optimized for all-flash configurations to minimize storage capacity consumption while ensuring performance and availability. These new features include deduplication, compression, and RAID5/6 erasure coding to reduce capacity consumption while ensuring the same levels of availability and performance for a lower total cost of ownership.

Deduplication and Compression

Enabling deduplication and compression can reduce the amount of storage consumed by as much as 7x. Actual reduction numbers will vary as this depends primarily on the types of data present, number of duplicate blocks, how much these data types can be compressed, and distribution of these unique blocks. For example, video files do not compress well while documents and spreadsheets typically yield more favorable results. Figure 3 shows deduplication and compression efficiency in an all-flash Virtual SAN environment where multiple versions and distributions of Windows and Linux servers are deployed along with vCenter Server, VMware vRealize Operations, and vSphere Replication.

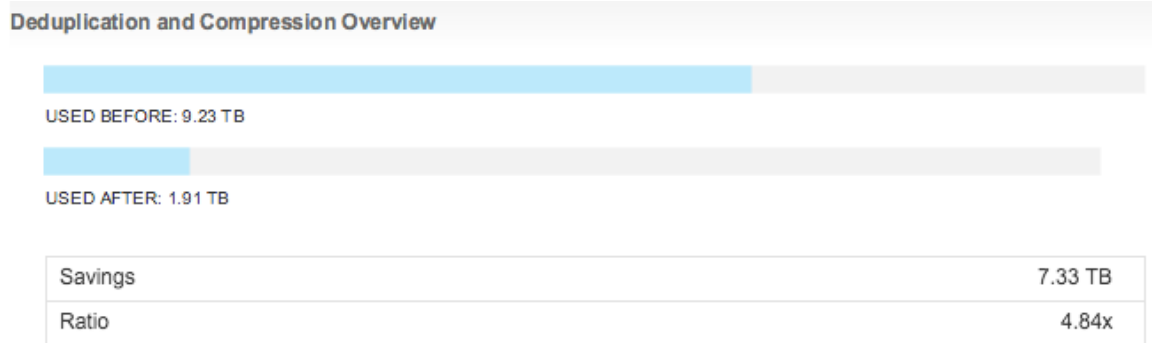


Figure 2. Deduplication and Compression Overview

Deduplication and compression is a cluster-wide setting that is disabled by default and can be enabled using a simple drop-down menu. Note that a rolling reformat of every disk group on every host in the Virtual SAN cluster is required, which can take a considerable amount of time. However, this process does not incur virtual machine downtime. Deduplication and compression are enabled as a unit. It is not possible to enable deduplication or compression individually.

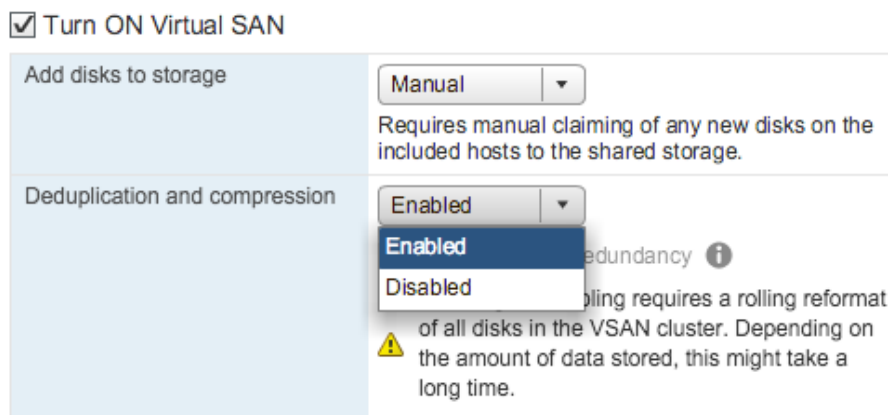


Figure 3. Enabling Deduplication and Compression

Deduplication occurs when data is de-staged from the cache tier to the capacity tier of an all-flash Virtual SAN datastore. The deduplication algorithm utilizes a 4K-fixed block size and is performed within each disk group. In other words, redundant copies of a block within the same disk group are reduced to one copy, but redundant blocks across multiple disk groups are not deduplicated. Deduplication at the disk group level using a 4K-block size provides a good balance of efficiency and performance.

The compression algorithm is applied after deduplication has occurred just before the data is written to the capacity tier. Considering the additional compute resource and allocation map overhead of compression, Virtual SAN will only store compressed data if a unique 4K block can be reduced to 2K or less. Otherwise, the block is written uncompressed.

There is a storage policy implication to be aware of when deduplication and compression are enabled. This is especially true when upgrading from a previous version of Virtual SAN where one or more storage policies contain an “Object space reservation” rule with a value other than 0% or 100%. When deduplication and compression are enabled, “Object space reservation” rules must be set to 0% or 100%. Values from 1% to 99% are not supported when deduplication and compression are enabled. An object that is assigned a storage policy containing an “Object space reservation” rule of 100% will be analyzed for deduplication, but no space savings will be realized since capacity for the entire object is reserved. Before upgrading Virtual SAN, all policies containing an explicit “Object space reservation” rule should be configured to 0% or 100%. Note: The implicit default value for this rule is 0% so there is no need to adjust a policy that does not have the “Object space reservation” explicitly defined.

Naturally, the processes of deduplication and compression on any storage platform require additional CPU cycles and potentially impact performance in terms of latency and maximum IOPS. Virtual SAN is no exception. However, considering deduplication and compression are only supported in all-flash Virtual SAN configurations, these effects will be negligible in the majority of use cases due to the extreme performance of flash devices. For guidance on what workloads will benefit the most from deduplication and compression, see [VMware Virtual SAN 6.2 Space Efficiency Technologies](#).

RAID-5/6 (Erasure Coding)

Virtual SAN 6.2 introduces a new Storage Policy Based Management (SPBM) rule, called Fault Tolerance Method, which allows virtualization administrators to choose which method of fault tolerance to use.

Prior to Virtual SAN 6.2, RAID-1 (Mirroring) was used as the failure tolerance method. Virtual SAN 6.2 adds RAID-5/6 (Erasure Coding) to all-flash configurations. While mirroring techniques excel in workloads where performance is the most important factor, they are expensive in terms of capacity required. RAID-5/6 (Erasure Coding) data layout can be configured to help ensure the same levels of availability, while consuming less capacity than RAID-1 (Mirroring). Use of erasure coding reduces capacity consumption by as much as 50% versus mirroring at the same fault tolerance level. This method of fault tolerance does require additional write overhead in comparison to mirroring as a result of data placement and parity.

RAID-5/6 (Erasure Coding) is configured as a storage policy rule and can be applied to individual virtual disks or an entire virtual machine. Note that the failure tolerance method in the rule set must be set to RAID5/6 (Erasure Coding).

Rule-Set 1

Select rules specific for a datastore type. Rules can be based on data services provided by datastore or based on tags. The VM storage policy will match datastores that satisfy all the rules in at least one of the rule-sets.

The screenshot displays the configuration for a storage policy rule set. Under "Rules based on data services", the "Failure tolerance method" is set to "RAID-5/6 (Erasure Coding) - Capacity". The "Number of failures to tolerate" is set to 2. The "Storage Consumption Model" section on the right indicates that a 100 GB virtual disk would consume 150 GB of storage space, with 0 B of initially reserved storage space and 0 B of reserved flash space.

Figure 4. Creating a VM Storage Policy with RAID6-like Erasure Coding

RAID-1 (Mirroring) in Virtual SAN employs a 2n+1 host or fault domain algorithm, where n is the number of failures to tolerate. RAID-5/6 (Erasure Coding) in Virtual SAN employs a 3+1 or 4+2 host or fault domain requirement, depending on 1 or 2 failures to tolerate respectively. RAID-5/6 (Erasure Coding) does not support 3 failures to tolerate. Table 1 details the host and capacity requirements of RAID-1 (Mirroring) vs. RAID-5/6 (Erasure Coding).

	Tolerated Failures	RAID-1 (Mirroring)		RAID-5/6 (Erasure Coding)		Erasure Coding Space Savings vs. Mirroring
		Minimum Hosts Required	Total Capacity Requirement*	Minimum Hosts Required	Total Capacity Requirement*	
FTT=0	0	3	1x	n/a	n/a	n/a
FTT=1	1	3	2x	4	1.33x	33% less
FTT=2	2	5	3x	6	1.5x	50% less
FTT=3	3	7	4x	n/a	n/a	n/a

*Without Deduplication/Compression taken into account.

Table 1. Space Requirements for Number of Failures to Tolerate and Tolerance Method

Erasure coding can provide significant capacity savings over mirroring, but it is important to consider that erasure coding incurs additional overhead. This is common among any storage platform today. Because erasure coding is only supported in all-flash Virtual SAN configurations, effects to latency and IOPS are negligible in most use cases due to the inherent performance of flash devices.

For more guidance on which workloads will benefit from erasure coding, see [VMware Virtual SAN 6.2 Space Efficiency Technologies](#).

Swap Efficiency

Virtual swap files are created when virtual machines are powered on. In cases where physical host memory is exhausted, the virtual swap file is used in place of physical memory for a virtual machine. Virtual swap files are sized according to the allocated memory minus reserved memory.

A virtual machine with 4GB of RAM allocated, with a 2GB memory reservation will create a 2GB virtual swap file. On Virtual SAN, that swap file will be created with a mirrored policy, resulting in 4GB of space consumed. 100 virtual machines with the same configuration will consume 400GB of capacity. In larger deployments of hundreds or thousands of virtual machines, this additional capacity could be substantial and require significant capacity.

In addition to the use of deduplication and compression, as well as erasure coding, it can be advantageous to use the Virtual SAN “SwapThickProvisionDisabled” advanced host setting for additional space savings.

Enabling this setting creates virtual swap files as a sparse object on the Virtual SAN datastore. Sparse virtual swap files will only consume capacity on Virtual SAN as they are accessed. The result can be significantly less space consumed on the Virtual SAN datastore, provided virtual machines do not experience memory over commitment, requiring use of the virtual swap file.

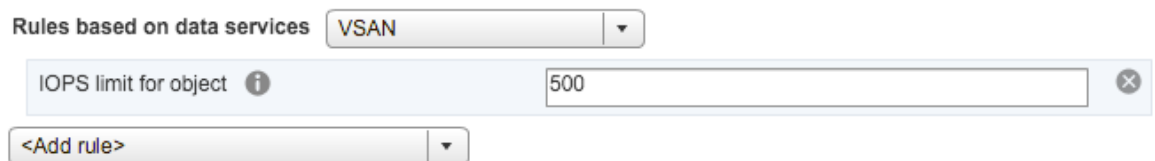
Quality of Service

Virtual SAN 6.2 adds a quality of service feature that limits the number IOPS an object may consume. In underutilized configurations, limits may or may not be necessary, as objects likely have sufficient resources to effectively meet the needs of their workload. While it is entirely desirable to have more than enough resources, it does not come without cost. Efficiently sized configurations are typically a good mix of cost and available resources. The metrics of appropriate resources for workloads can change over time, especially as utilization grows, or as workloads are added over the lifecycle of a platform.

There are several situations where it might be advantageous to limit the IOPS of one or more virtual machines. The term “noisy neighbor” is often used to describe when a workload monopolizes available I/O or other resources, which negatively affect other workloads on the same platform. In environments where there is a mix of both low and high utilization, it could be detrimental for a virtual machine with low utilization during normal operations to change its pattern and start to consume massive resources, in turn starving others for enough to operate properly. Particularly at a larger scale the impact of this situation may affect multiple business units, tenants, or customers.

A good example of a “noisy neighbor” situation includes end of month reporting. Consider those virtual machines generating reports residing on the same 4 node Hybrid Virtual SAN cluster as other Tier 1 applications that have stringent Service Level Agreement requirements. What would happen if one or more large reports were generated, consuming a larger percentage of IOPS than were available for the minimum service level requirements of the Tier 1 applications? It is quite possible the Tier 1 applications would not be able to satisfy the requirements of the Service Level Agreement. Implementing IOPS limits on the reporting solution VM’s, could prevent the situation where the reporting solution starves the Tier 1 applications for IO.

With the Quality of Service addition to Virtual SAN 6.2, IOPS limits are now available. Quality of service for Virtual SAN 6.2 is a Storage Policy Based Management (SPBM) rule. Because quality of service is applied to Virtual SAN objects through a Storage Policy, it can be applied to individual components or the entire virtual machine without interrupting the operation of the virtual machine.



The screenshot shows a configuration interface for a Storage Policy Based Management (SPBM) rule. At the top, it says "Rules based on data services" with a dropdown menu set to "VSAN". Below this is a text input field labeled "IOPS limit for object" with an information icon (i) and a close icon (x). The value "500" is entered in the field. At the bottom, there is a button labeled "<Add rule>" with a dropdown arrow.

Figure 5. Creating a Storage Policy limiting IOPS to 500

Quality of service for Virtual SAN 6.2 is normalized to a 32KB block size, and treats reads the same as writes. An example with an IOPS limit of 500 (regardless of block size up to 32KB) will result in 500 IOPS, while a block size of 64KB would result in 250 IOPS. It is important to consider the workload profile when configuring IOPS limits.

Performance Service

A healthy Virtual SAN environment is one that is performing well. Virtual SAN 6.2 includes a number of new graphs and data points that provide performance information at the cluster, host, virtual machine, and virtual disk levels. Time Range can be modified to show information from the last few hours or a custom date and time range.

The performance service is enabled at the cluster level, and it is important to remember that the performance history database is not a component of vCenter Server. The performance history database is stored as a Virtual SAN object, independent of vCenter. A storage policy is assigned to the object to control space consumption, availability, and performance of that object. If the object becomes unavailable, performance history for the cluster will be unavailable until access to the object is restored. The performance history database can consume up to 255GB of capacity on the Virtual SAN datastore.

Performance Service is Tuned ON		Turn off	Edit storage policy ...
Stats object health	✓ Healthy		
Stats object UUID	5b14a856-aab0-6a63-40ea-002590e14756		
Stats object storage policy	📁 Virtual SAN Default Storage Policy		
Compliance status	✓ Compliant		

Figure 6. Virtual SAN - Performance Service Status and Object Policy

Cluster Metrics

At the cluster level, the performance monitoring service shows performance metrics for virtual machines running on Virtual SAN as well as the Virtual SAN backend. These metrics provide quick visibility to the entire Virtual SAN cluster, showing how Virtual SAN and the objects that reside on it are performing.

Figure 7 shows metrics such as IOPS, throughput, latency, and outstanding IO for all Virtual SAN objects in the cluster, not only virtual disks.

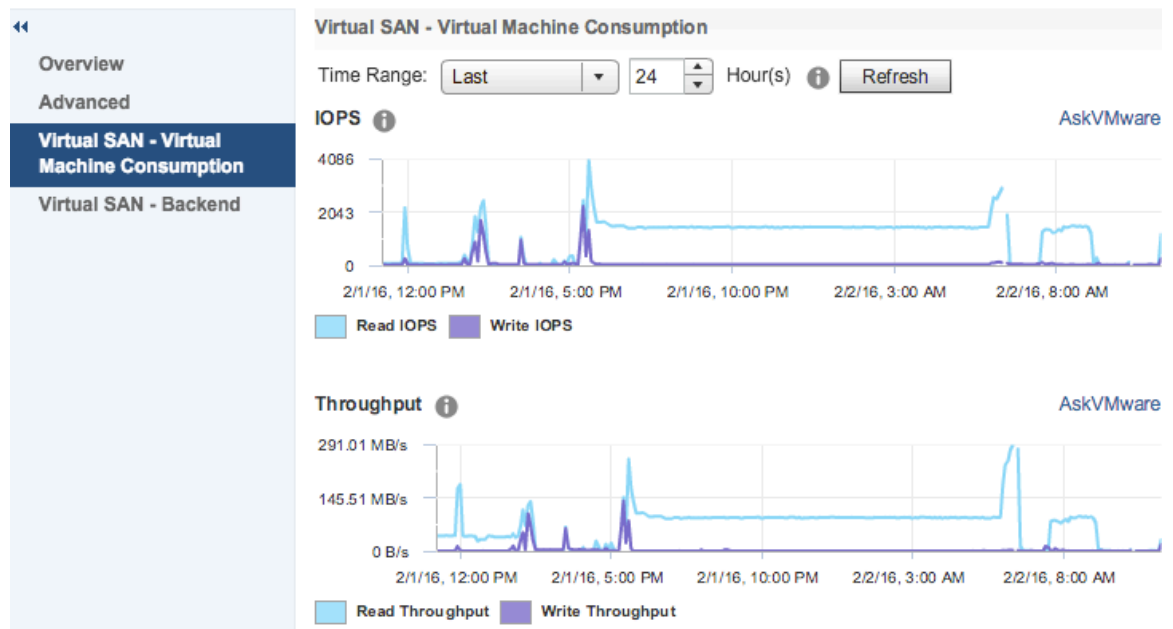


Figure 7. Virtual SAN - Virtual Machine Consumption

Backend Virtual SAN metrics show what is required to deliver what is visible at the virtual machine and object level. Overhead from different storage policies can show significantly different backend metrics than what is observed at the virtual machine or object level.

Take the example of a virtual machine with a number of failures to tolerate of 1, with a failure tolerance method of RAID-1 (Mirroring). For every write IO to a virtual disk, 2 will be seen on the backend. This is because two mirrors both receive a write, despite the fact that the virtual disk only had a single write. Changing the fault tolerance method to RAID-5/6 (Erasure Coding) will now require 4 writes, because data and parity in a RAID-5 configuration is comprised of 3 data writes and 1 parity write.

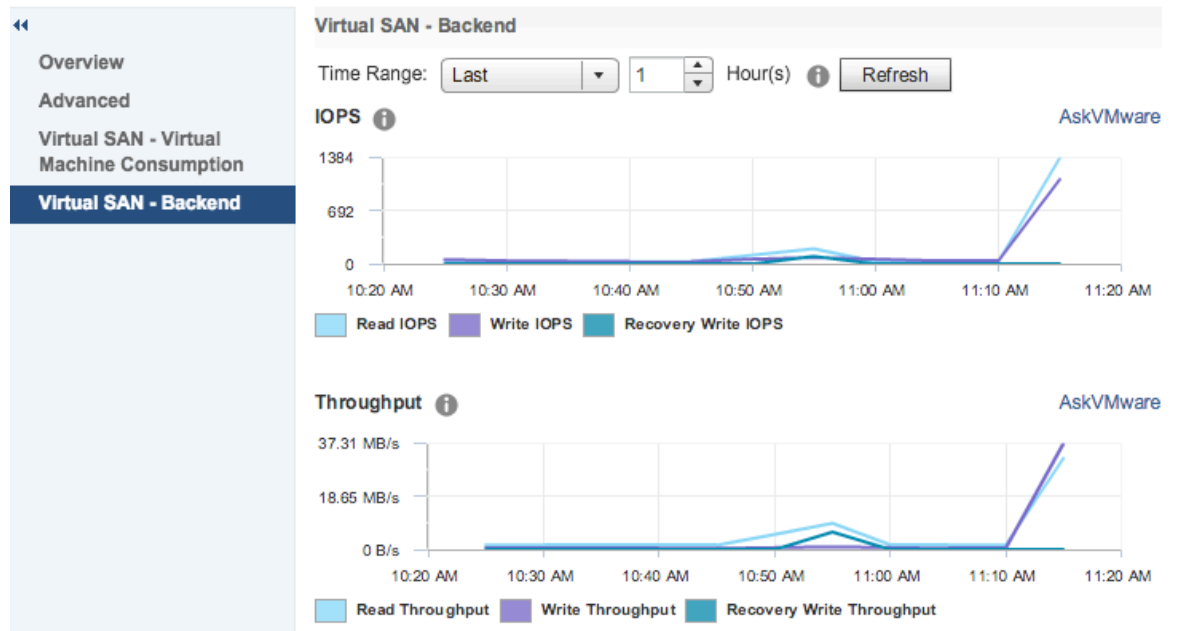


Figure 8. Virtual SAN - Backend

Host Metrics

As visibility of cluster performance is important, so is performance at the host. Virtual machine consumption as well as backend metrics are visible specific to the host selected, as well as additional metrics.

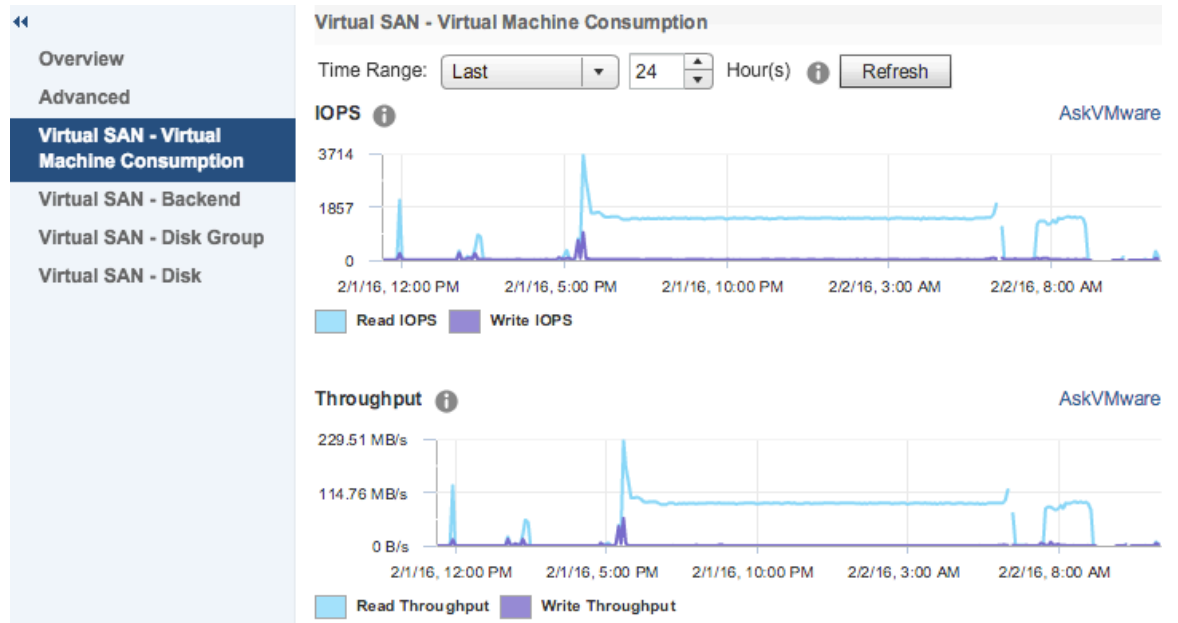


Figure 10. Host-level Virtual Machine Consumption Performance Data

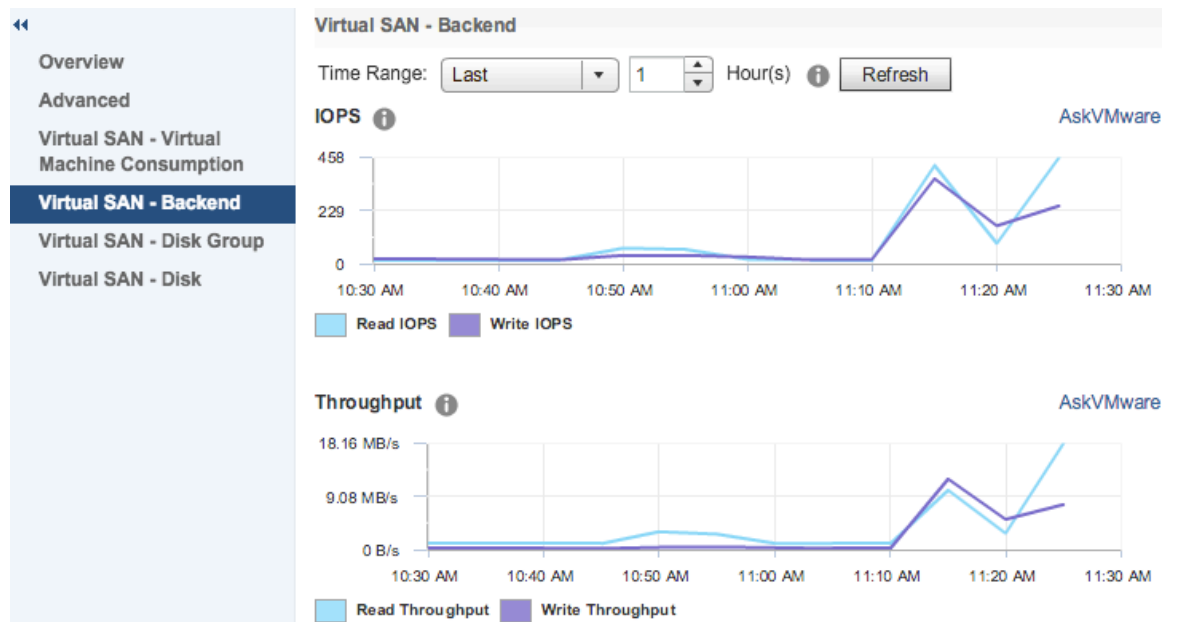


Figure 11. Host-level Backend Performance

Virtual SAN requires at least 1 disk group, and may be configured with up to 5 disk groups. It is important to be able to distinguish how each disk group is performing, independent of any other disk groups on the same host.

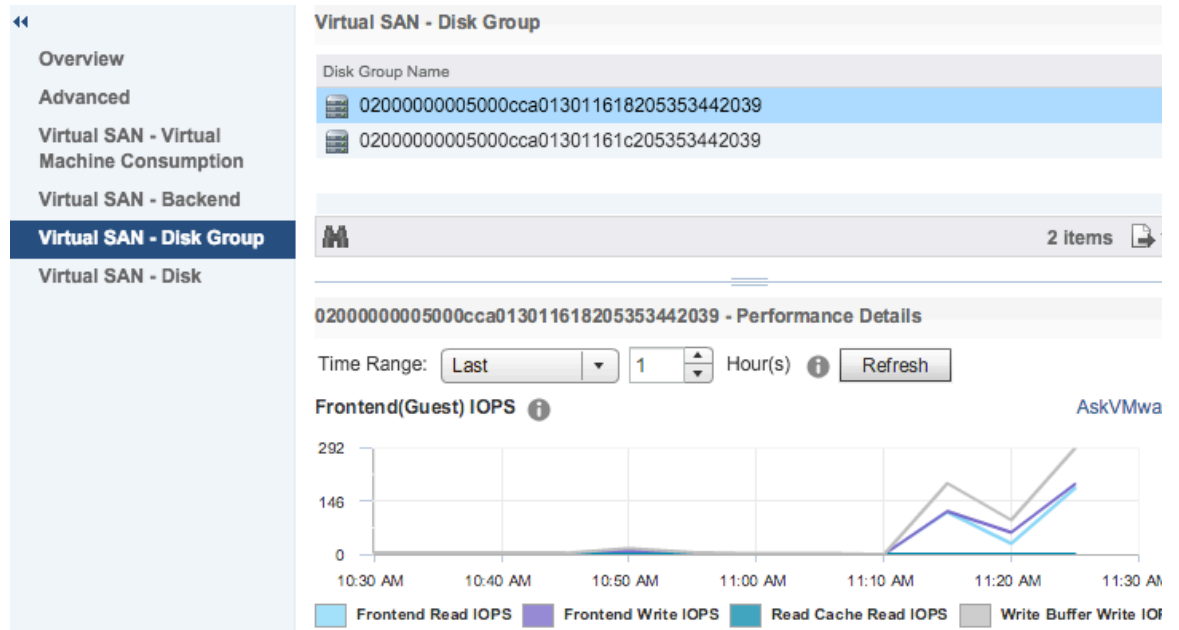


Figure 12. Host-level Disk Group Performance

Disk groups in a Virtual SAN host require at least 1 capacity device, and may contain up to 7 capacity devices. Just as it is important to be able to distinguish different disk groups' performance, it is also important to have visibility to the performance metrics of a single device.

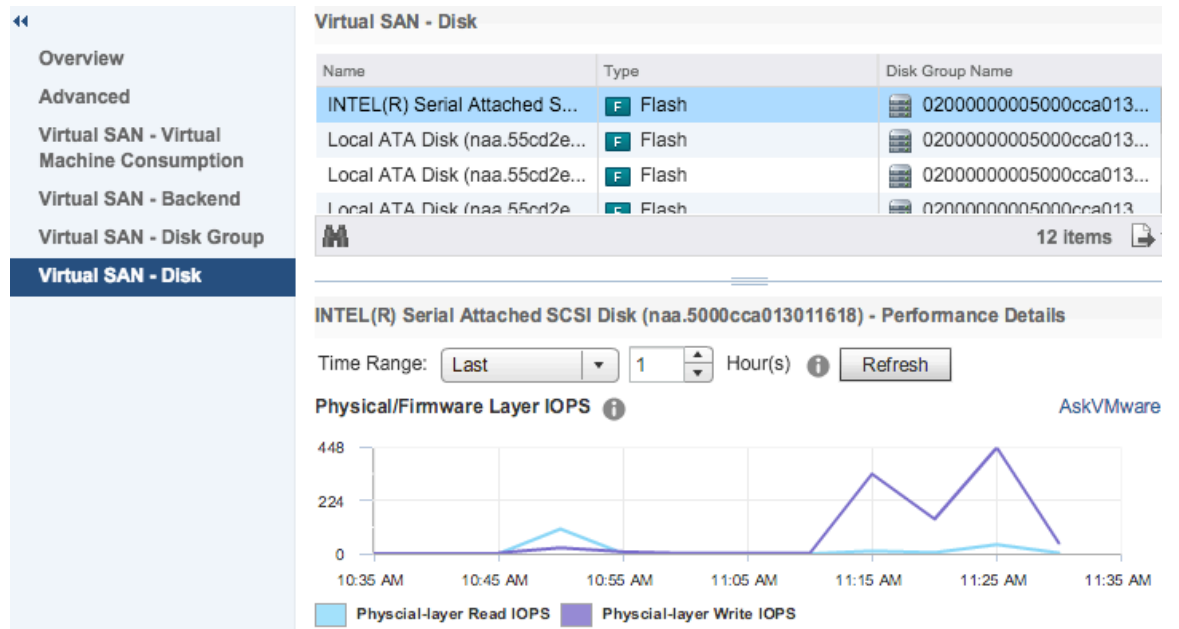


Figure 13. Individual Disk Performance

Virtual Machine Metrics

Performance metrics are now available for all of the upper level components of Virtual SAN. We can now view performance data at the cluster, host, disk group, and devices. It is also important to have the same visibility for specific virtual machines and their disks.

As before virtual machine consumption shows that the virtual machine is experiencing from an application perspective. Figure 14 shows virtual machine consumption specific to all the objects the virtual machine contains.

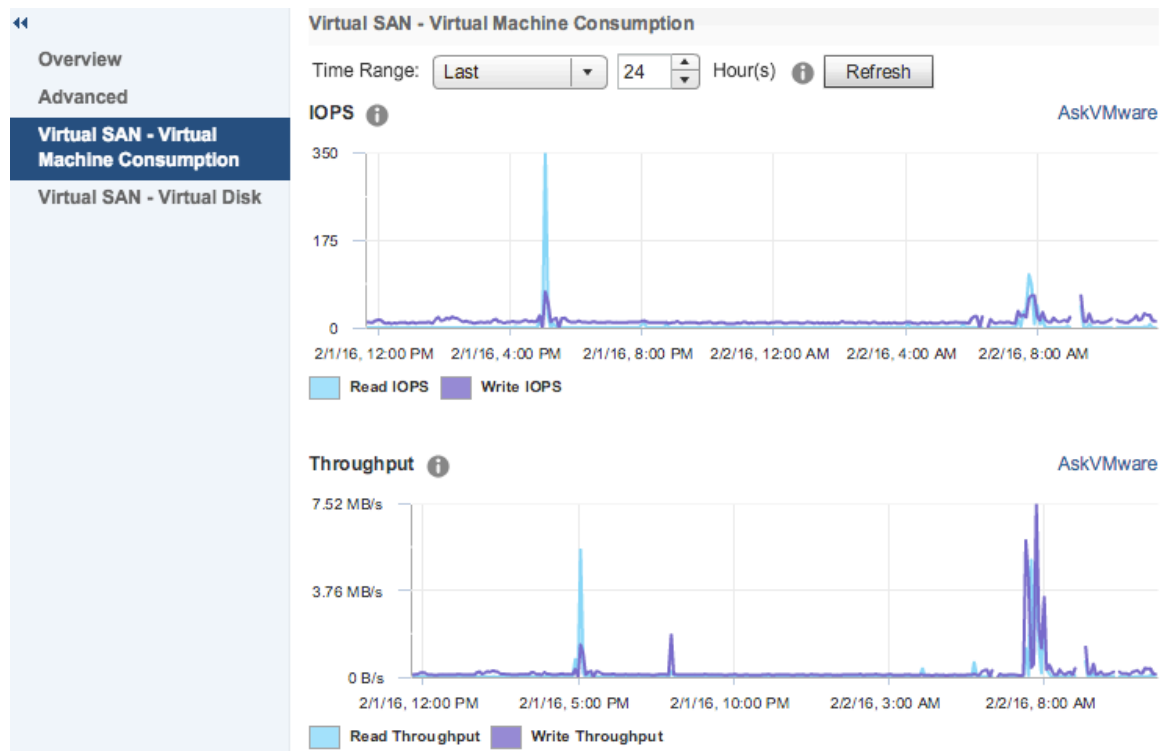


Figure 14. Virtual Machine-level Virtual Machine Consumption

Going even deeper, individual metrics are available on a per disk basis. It is very common to have virtual machines configured with more than one virtual disk. With Virtual SAN leveraging Storage Policy Based Management (SPBM), it is entirely possible to have completely different storage policies applied to different virtual disks.

Different properties like SPBM policy, workload, capacity, and more can contribute to the performance characteristics of a virtual disk.

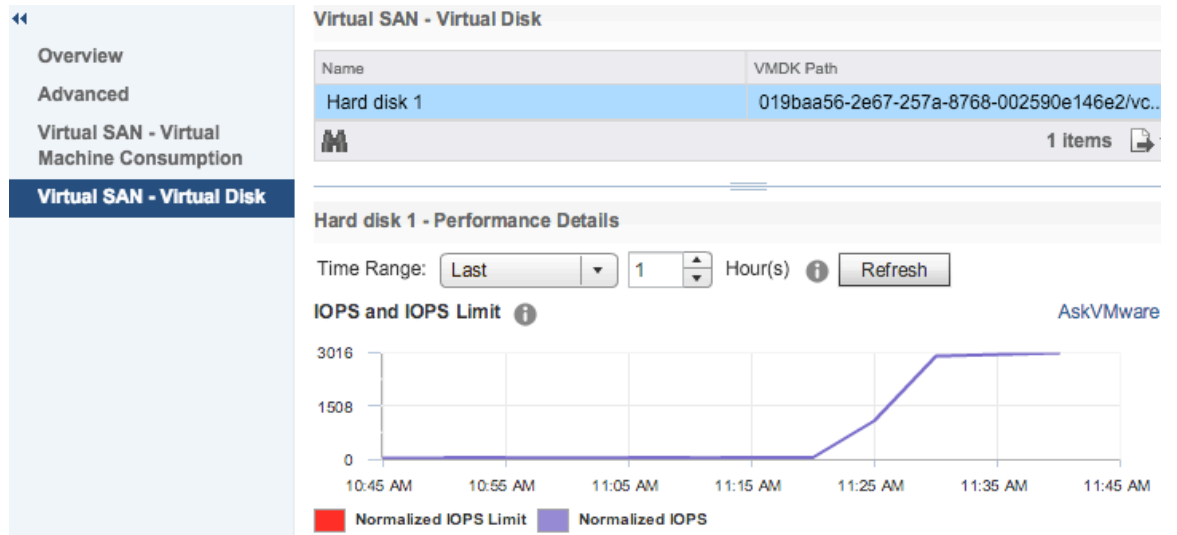


Figure 15 Virtual Machine - Virtual Disk

Capacity details are also available in the Virtual SAN user interface making it easy for administrators to understand how much capacity various object types are consuming.

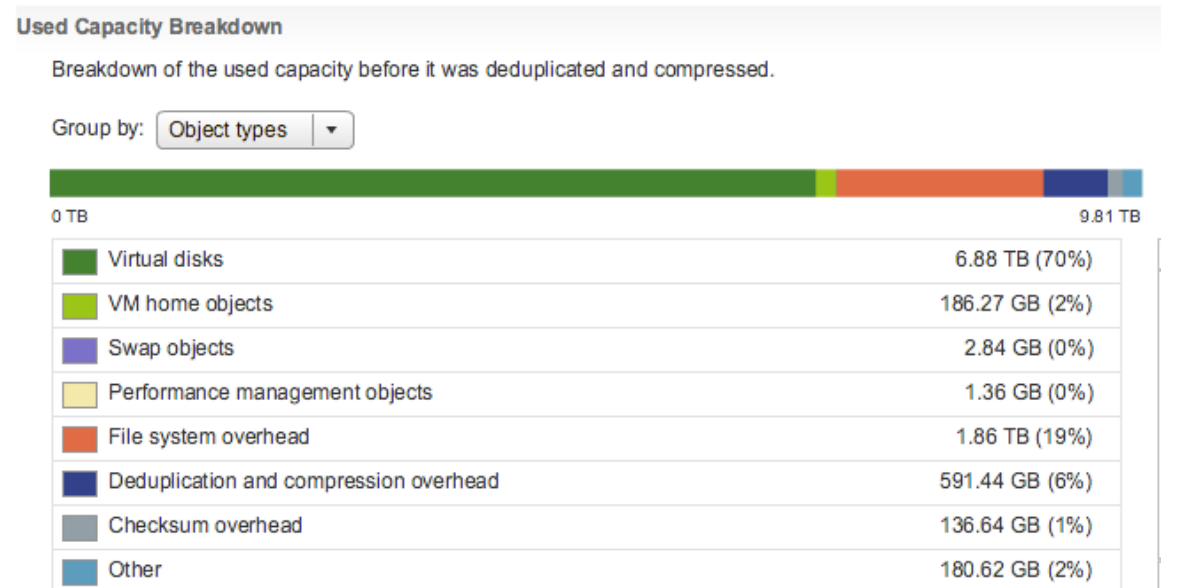


Figure 16. Used Capacity Breakdown

Virtual SAN 6.2 includes the Virtual SAN API, which is an extension of the vSphere API. The Virtual SAN API centers on a small set of managed objects, which enable administrators to query runtime state, as well as, configure Virtual SAN. The API is exposed as a Web service, running on both VMware vCenter Server systems and VMware ESXi systems. Managed objects are available for cluster-level and host-level operations.

Health Service

Virtual SAN 6.2 features an improved health service. The health service actively tests and monitors a number of items such as hardware compatibility, network connectivity, cluster health, and capacity consumption. The performance service provides various metrics such as throughput, IOPs, and latency.

The health service is enabled by default and configured to check the health of the Virtual SAN environment every 60 minutes.

Virtual SAN Health (Last checked: Today at 1:30 PM)	
Test Result	Test Name
✓ Passed	▶ Hardware compatibility
✓ Passed	▶ Network
✓ Passed	▶ Physical disk
✓ Passed	▶ Data
✓ Passed	▶ Cluster
✓ Passed	▶ Limits
✓ Passed	▶ Performance service

Figure 17. Virtual SAN Health

The Health service is quite thorough in the number of tests it performs. As an example, proper network configuration is key to a healthy Virtual SAN cluster and there are 11 tests in the “Network” section of the Virtual SAN Health user interface.

✓ Passed	▼ Network
✓ Passed	All hosts have a Virtual SAN vmknic configured
✓ Passed	All hosts have matching multicast settings
✓ Passed	All hosts have matching subnets
✓ Passed	Basic (unicast) connectivity check (normal ping)
✓ Passed	Hosts disconnected from VC
✓ Passed	Hosts with connectivity issues
✓ Passed	Hosts with Virtual SAN disabled
✓ Passed	MTU check (ping with large packet size)
✓ Passed	Multicast assessment based on other checks
✓ Passed	Unexpected Virtual SAN cluster members
✓ Passed	Virtual SAN cluster partition

Figure 18. Network Tests in the Virtual SAN Health Service

If an issue is detected in the environment, a warning is visible in the Virtual SAN user interface. Clicking on the warning provides more details about the issue. For example, a controller driver that is not on the hardware compatibility list for Virtual SAN will trigger a warning. In addition to providing details about the warning, the Virtual SAN Health Service user interface also has an “Ask VMware” button, which brings up the relevant VMware Knowledge Base article.

Controller Driver Ask VMware

Checks if the controller driver is VMware certified. i

Controller List

Host	1 ▲ Device	Driver in use	Driver health
prmh-a09-s...	vmhba4: LSI LSI Logic...	lsi_msgpt3 (06.255.12.00-8v...	Warning
prmh-a09-s...	vmhba2: LSI LSI Logic...	lsi_msgpt3 (06.255.12.00-8v...	Warning

Figure 19. Test Details and the Ask VMware Button

vSphere and Virtual SAN support a wide variety of hardware configurations. The list of hardware components and corresponding drivers that are supported with Virtual SAN can be found in the [VMware Compatibility Guide](#). It is very important to use only hardware, firmware, and drivers found in this guide to help ensure stability and performance of a Virtual SAN environment. The list of certified hardware, firmware, and drivers is contained in a hardware compatibility list (HCL) database. The Virtual SAN user interface makes it very easy to update this information, for use by the Health service tests. If the environment has Internet connectivity, updates can be obtained directly from VMware. Otherwise, HCL updates can be downloaded as a file to enable offline updates.

HCL Database Update from file... Get latest version online

Last updated Today

Figure 20. Updating the Virtual SAN HCL Database

Finally, if an issue does arise that requires the assistance of VMware Support, it is very easy to upload support bundles to help expedite the troubleshooting process. Clicking the “Upload Support Bundles to Service Request...” button enables an administrator to enter an existing support request (SR) number and upload the necessary logs with just a few mouse clicks.

Support Assistant Upload Support Bundles to Service Request...

Last upload time --

Figure 21. Virtual SAN Support Assistant

Application Support

Virtual SAN continues to be a platform of choice for a wide variety of workloads. Virtual SAN is a great fit for all types of workloads, large or small. VMware customers are adopting Virtual SAN at an increasing rate since its release. Virtual SAN has proven to be a robust and effective platform for business critical applications, virtual desktop infrastructure (VDI), test and development, and a multitude of other use cases.

As Virtual SAN is progressively used for business critical applications, it has been important to provide additional certification, support, and validation specific to several key workloads. Examples include support for Microsoft Windows Server Failover Clustering, when using a File Share Witness, as well as, support for Oracle Real Application Clusters. Virtual SAN 6.2 introduces support for core SAP applications.

Business critical applications, virtual desktops, and other application types have increased scalability with Virtual SAN, without the sizing challenges often associated with legacy storage platforms. Virtual SAN, built on vSphere can easily be sized for small workloads, large workloads, and any in between.

Summary

VMware Hyper-Converged Software transforms industry-standard x86 servers and directly attached storage into radically simple Hyper-Converged Infrastructure (HCI) to help eliminate high IT costs, management complexity, and performance limitations. vSphere, vCenter Server, and Virtual SAN are tightly integrated to deliver the leading HCI solution that offers simplicity, reliability, and performance for nearly any use case. Virtual SAN is easy to implement with automated configuration and includes proactive tests to help verify functionality and performance. Virtual SAN is optimized for modern all-flash storage with efficient near-line deduplication, compression, and erasure coding capabilities that lower TCO while delivering incredible performance. New and improved features such as the performance and health services make it easier than ever to verify Virtual SAN configurations and closely monitor key metrics such as IOPs, throughput, and latency at the cluster, host, virtual machine, and virtual disk levels. Quality of service can be managed by using IOPs limits on a per-virtual machines and per-virtual disk basis. Virtual SAN 6.2 is ready for any application with tested and validated deployments of several business critical applications including SAP and Oracle RAC,

About the Authors

Jase McCarty is a Senior Technical Marketing Architect at VMware with a focus on storage solutions. He has been with VMware for a year, as well as in the Information Technology field for over 25 years, with roles on both the customer and vendor side. Jase has Co-Authored two books on VMware virtualization, routinely speaks technology focused user group meetings, and has presented at VMworld and EMC World.

Follow Jase on Twitter: [@jasemccarty](https://twitter.com/jasemccarty)

Jeff Hunter is a Senior Technical Marketing Architect at VMware with a focus on storage and availability solutions. He has been with VMware for more than 8 years, prior to which he spent several years implementing and administering VMware software-defined data centers at two Fortune 500 companies. Jeff has presented at multiple technology conferences including VMworld and EMC World.

Follow Jeff on Twitter: [@jhuntervmware](https://twitter.com/jhuntervmware)