vSAN Space Efficiency Technologies

Space efficiency capabilities with vSAN using VMware Cloud Foundation 9.1

May 11, 2026

vmware by Broadcom

Table of Contents

Introduction

Space efficiency technologies in enterprise storage play an important role in improving value and decreasing costs. VMware vSAN has several technologies in place to help improve storage efficiency.

Space efficiency techniques can be categorized into the following:

- **Opportunistic.** These space efficiency techniques are dependent on conditions of the data, and not guaranteed to return a predetermined level of savings. vSAN offers several types of opportunistic space efficiency features such as deduplication, compression, TRIM/UNMAP space reclamation, thin provisioning and snapshot storage efficiencies.
- **Deterministic.** These space efficiency techniques can be relied upon to deliver a guaranteed level of capacity savings. vSAN offers deterministic space efficiency capabilities through data placement schemes that are optimized for storing data in a resilient but efficient manner. This includes RAID-5/6 erasure codes.

In vSAN, opportunistic and deterministic space efficiency features can be used independently or together. The specific considerations when doing so will be discussed in this document.

Scope of Topics

The information provided in this document will assume the use of vSAN 9.1, and/or VMware Cloud Foundation (VCF) 9.1. VCF deployments may have additional requirements and support limitations that fall outside of the scope of this document. **For VCF environments, please refer to the Administration Guide for VMware Cloud Foundation for guidance as it relates to VCF.** The release of vSAN in VCF 9.1 marks the 6th release of vSAN ESA, and as a result, this document will primarily focus on space efficiency features in vSAN ESA. For space efficiency features related to the vSAN Original Storage Architecture (OSA), see earlier versions of this document.

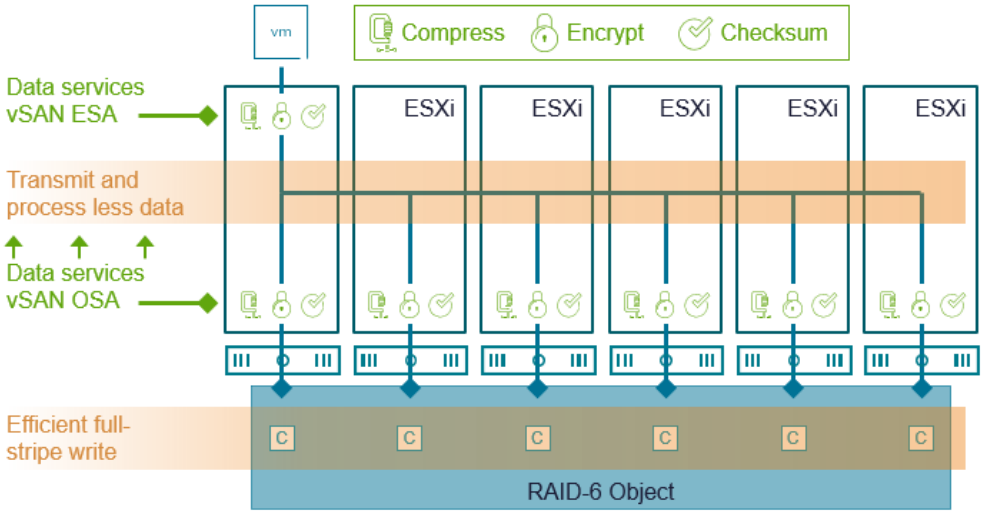
Opportunistic Space Efficiency Features

Opportunistic space efficiency techniques do just as the name implies. If a given set of conditions is ideal for saving capacity, it will do so based on that given set of conditions. The degree to which the savings will occur will be highly dependent on the technology in place, the workloads, and even the host hardware configuration.

How much space savings can one expect from using these opportunistic space efficiency features? The answer really depends on the workload and the characteristics of the data stored. Fortunately, vSAN in VCF 9.1 has a new way of measuring efficiency savings, and is described later in this document.

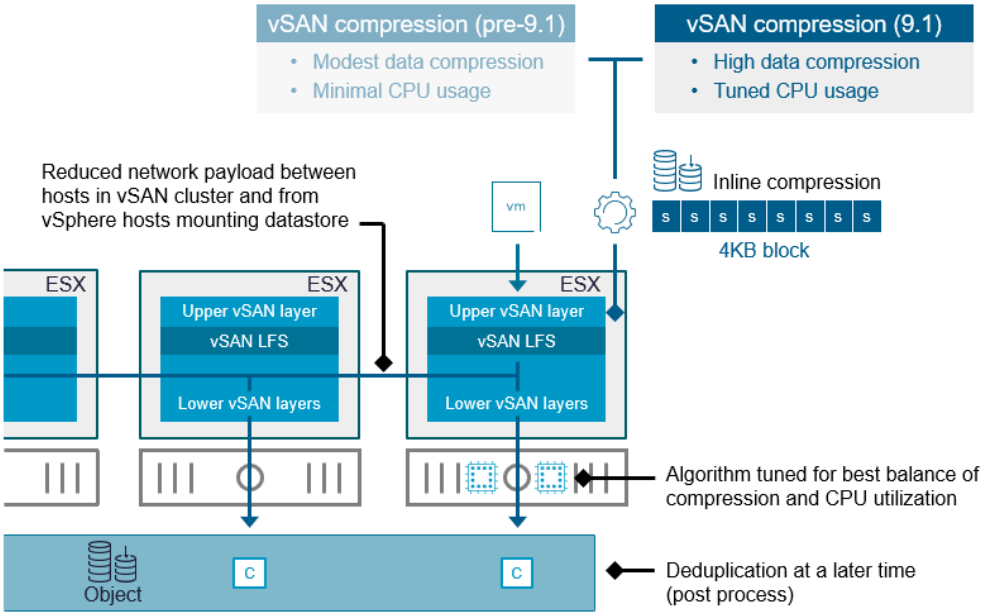
Compression

The Express Storage Architecture (ESA) introduced in vSAN 8 included a new data compression capability. With vSAN ESA, data compression (and other services such as encryption, and checksum processing) occurs in line, near the top of the storage stack. When a guest VM issues a write operation, it will compress the data as it enters the top of the vSAN storage stack. This single compression step means that all subsequent processing or network transmission works with the compressed data, which saves CPU cycles, and network resources. This approach can even improve performance and reduce resources when using the vSAN ESA in a stretched cluster topology.



Using data compression in vSAN ESA, each incoming 4KB block is evaluated on a 512 Byte sector size. With 8 sectors to a 4KB block, this means that the 4KB block can be reduced in increments of 512 bytes, depending on how compressible the 4KB block is. For example, a 4KB block could be compressed down to 7/8ths its original size if it is not very compressible, or all the way down to 1/8th its original size, if it is highly compressible.

vSAN in VCF 9.1 enhances data compression by using a new highly space efficient algorithm. Moving from the common LZ4 algorithm to a highly tuned version of ZSTD will yield better compression ratios across all workload types while keeping CPU resources to a minimum.



While this can offer much better data compression than past versions, it is entirely dependent on the customer's data, and how compressible it may or may not be. For example, an already compressed image or video file format will not be compressed beyond the native file format. As a result, we recommend taking a conservative approach in estimating what your compression rates may be for real world data.

In versions of vSAN prior to 9.1, data compression was controlled by storage policy. It was on by default, with the option to disable it. In vSAN for VCF 9.1, it is a cluster-based feature that is enabled by default and cannot be turned off in the UI. In

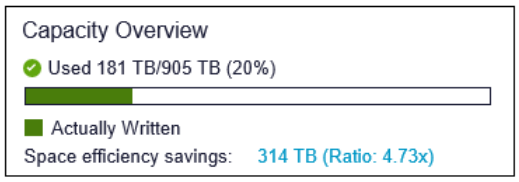
past cases, it was discovered that turning off data compression hindered performance because of the other processing efficiencies gained (Checksum, encryption, deduplication, and network transmission) thus the reason why it is an always-on capability now in vSAN. For more information on the new data compression capabilities in vSAN, see the post: "More Capacity with VMware vSAN Compression and Global Deduplication in VCF 9.1."

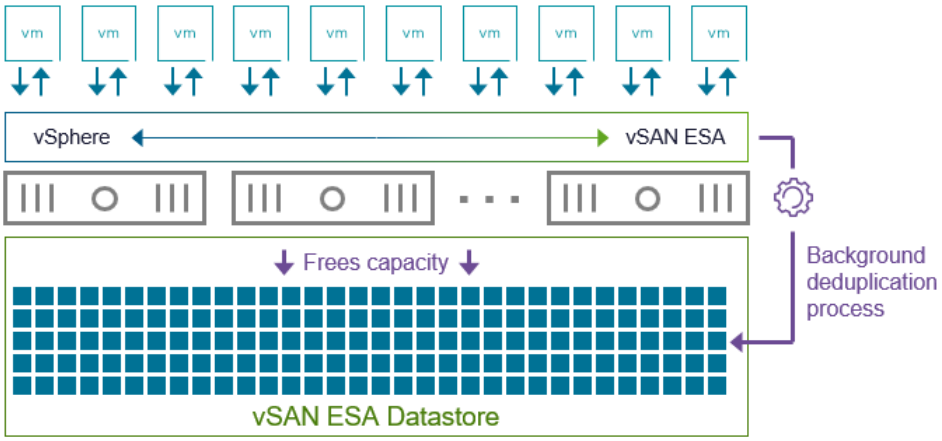
Global Deduplication

vSAN offers a powerful cluster-wide deduplication capability. First introduced under limited availability in vSAN for VCF 9.0, this gave many customers a preview of its ability to deliver extraordinary levels of space efficiency with little to no material impact on performance. It has already proven to drive down the cost of storage for VCF environments to lower than traditional storage offerings. For more information, see the post: "Save Costs and Scale Efficiently with vSAN Deduplication in VMware Cloud Foundation 9.0."

Some of the key characteristics of vSAN Global Deduplication includes:

- **Larger deduplication domain.** The deduplication domain in vSAN ESA is the entire ESA cluster. This dramatically improves the potential of duplicate blocks of data to be found and deduplicated, improving the effective data reduction rates. This deduplication domain will automatically grow as the host count of the cluster grows.
- **Adaptively throttled post-processing.** Performing deduplication outside of the write path helps ensure that there is no interference with ingesting write operations from guest VMs. This helps ensure that guest VM write latency remains low. Since vSAN is fully aware of system resources used at any point in time, advanced algorithms will adaptively throttle the amount of resources used for the purposes of deduplication. During quiet times, the cluster will consume more resources for deduplication than it will during busier times in the day.
- **Minimal performance impact.** The architecture used in ESA's deduplication virtually eliminates all of the performance challenges associated with deduplication in vSAN OSA. Deduplication in vSAN ESA is appropriate for all workload types.





vSAN Global Deduplication is generally available to all customers in vSAN in VCF 9.1. It relaxes most of the initial restrictions with the limited availability release and offers new functionality. Deduplication now supports vSAN HCI and vSAN storage clusters between 3 and 64 hosts. Stretched clusters and 2-Node clusters are not supported at this time.

Deduplication in vSAN for VCF 9.1 also supports vSAN Data-at-Rest Encryption. This gives vSAN the unique ability **deduplicate data that has been fully encrypted without impacting data reduction ratios**. Typically, storage solutions lose

the benefit from deduplication when data is encrypted, as data blocks each use a unique hash when stored. vSAN's post-processing method allows us to quickly unencrypt the block using helper threads low in the stack so that the deduplication process can be completed.

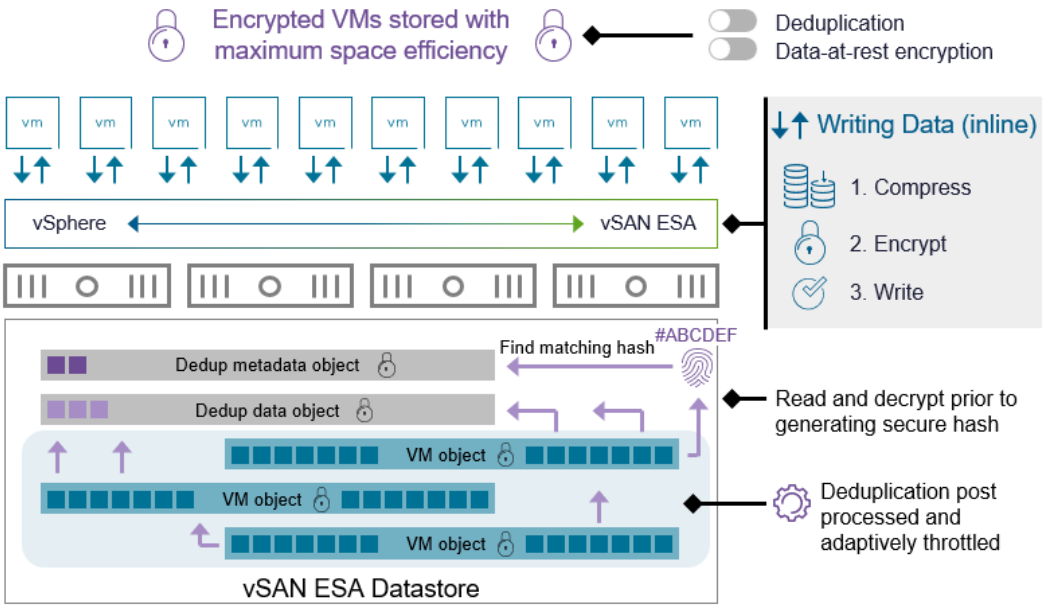
Recommendation: If data encryption is required in your environment, ALWAYS use vSAN encryption services (Data-at-Rest, Data-in-Transit, or both) instead of VMware vSphere Encryption. The former will maintain the full space efficiency capabilities of global deduplication, where the latter will negate all space efficiency benefits from deduplication. For more information, see the document: "vSAN Encryption Services."

Deduplication Post-Processing

All deduplication activities occur after data has been persisted to storage. This type of post-processing helps minimize the impact on guest VM performance. Since ESA's metadata mapping understands recently written data versus cold data, it will always prioritize the deduplication of cold data first to minimize the deduplication of frequently written data that may change rapidly in a short period of time. vSAN's sophisticated set of algorithms will dynamically throttle the amount of resources used for the purpose of deduplication. vSAN will categorize any data movement related to deduplication as resynchronization traffic, where its adaptive resync and adaptive network traffic shaping will manage resources properly to ensure VM workloads maintain priority. There are no administrative tuning or operations to worry about.

The deduplication process will occur in the following manner.

1. vSAN will read a discrete 4KB block and generate a secure cryptographic hash to be stored in the dedup metadata object. This is offloaded to lower layer processes for fast performance.
2. Decrypt and general strong cryptographic hash. This is also offloaded to lower layer processes for fast performance.
3. vSAN will look for a matching hash entry in the dedup metadata object.
    a. If a match has been found with data in the dedup data object, it will update the block with a metadata pointer and reclaim the space
    b. If a match has been found with no data in the dedup data object, it will move both the current data and the original data (using the back-pointer discussed below) to the dedup object, update the blocks with a metadata pointer and reclaim the space.
    c. If no match has been found, it will leave the data as-is. Hash entries will be created in the metadata object with a back-pointer where the data resides so that if a duplicate entry is identified, it can be deduplicated as described above.



Both the data and the metadata used in vSAN ESA's deduplication process reside in unique objects that are instantiated at the time that deduplication is enabled. Depending on the size and other conditions of the cluster, a cluster will have one or more of the following:

- **Deduplication metadata object.** Maintains a hash entry for every chunk in the system. It is the hash entry that helps identify other instances that contain the same data.
- **Deduplication data object.** This will be comprised of all the chunks of data that have been deduplicated. Deduplicated data are moved to these objects to prevent referential hot spots on VMs.

Since vSAN Global Deduplication is a post-processing activity, it may take a few days for the system to learn and execute on deduplication activities, and for those results to show up in the UI. Deduplication processes may also change the advertised compression rates within your cluster. The overall data reduction ratio improves due to it using a combination of both techniques. Even with data that is highly compressible, if vSAN determines that deduplication will result in additional savings, it will do so and report it as such in the UI.

Enabling and disabling vSAN Global Deduplication in vSAN ESA is very different than past implementations in vSAN OSA. When it is enabled in a cluster, it does not perform any rolling reformats or activity that forces large amounts of data movement. Disabling vSAN Global Deduplication in vSAN ESA will only pause the deduplication post-processing. It will not reverse the deduplication of data.

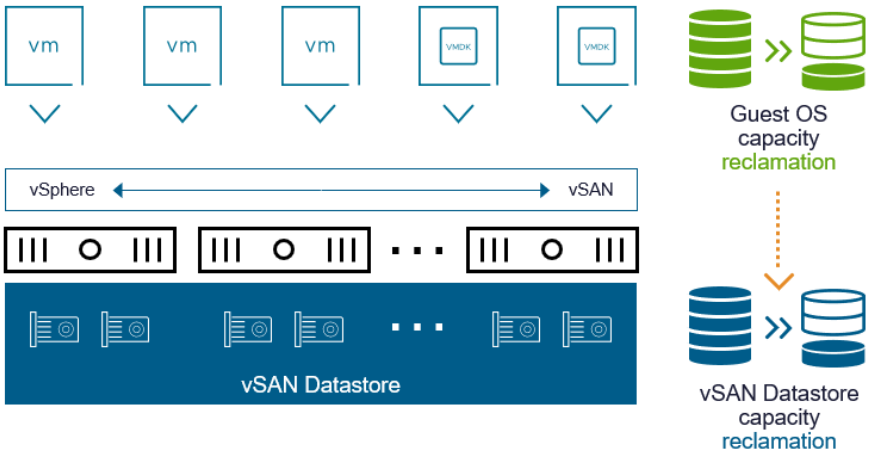
Thin Provisioning

A thin-provisioned storage system provisions storage on an as-needed basis. It does not provision the entire amount of capacity needed for say, a virtual disk, but rather, only the space that has been initially needed by the VMDK. vSAN uses thin provisioning, which provides the minimum amount of storage capacity needed by the vSAN objects created on the datastore. It will then transparently increase the amount of space used as it is needed. As a result, it is entirely possible to initially provide more capacity than the physical datastore is able to provide. This is known as "overprovisioning" and is typical in any storage system that uses thin provisioning. With the new "Effective Capacity" view in vSAN for VCF 9.1, thin provisioning savings will be graphically represented in the UI and expressed as a ratio like other space efficiency techniques.

One of the challenges to thin provisioned systems is that once a given entity has grown (such as a VM's VMDK), it will not shrink when data within the guest OS are deleted - common with databases that use transaction log files. This problem is amplified by the fact that many file systems will always direct new writes into free space. A steady set of writes to the same block of a single small file will eventually use significantly more space at the VMDK level. To solve this problem, automated TRIM/UNMAP space reclamation is available for vSAN and discussed later in this document.

TRIM/UNMAP Space Reclamation

Thin provisioning efficiently allocates storage as needed but cannot automatically reclaim it when data is deleted. To solve this, vSAN uses TRIM/UNMAP commands to recognize when a guest OS no longer needs specific blocks. By processing these commands, vSAN reclaims previously allocated storage as free space, significantly improving overall capacity utilization.



Guest TRIM/UNMAP is enabled by default in more recent editions of vSAN, and its status is visible in the vSAN "Advanced Options" for the configuration of the cluster. This process carries benefits of freeing up storage space but also has other secondary benefits:

- **More accurate data usage reporting.** As noted on the post "The Importance of Space Reclamation for Data Usage Reporting in vSAN" running space reclamation on VMs will help reconcile capacity reporting in your environment.
- **Faster repair** - Blocks that have been reclaimed do not need to be rebalanced or redistributed in the event of a device failure.

While the process can introduce a performance impact, the vSAN performance metrics allow you to track UNMAP activity, including UNMAP IOPS, throughput, and latency. Do not be alarmed by occasional UNMAP latency being relatively high compared to your guest VM latency. This can be due to prioritization of guest VM I/O activity over UNMAP I/O activity. It can also occur when there are little to no UNMAP operations, which does not give a point of reference to determining the latency accurately.

What Opportunistic Space Efficiency Features Should Be Used?

In vSAN for VCF 9.1, data compression, thin provisioning, and TRIM/UNMAP space reclamation features are capabilities that are "always on" and will reduce your capacity usage as much as the data allows. vSAN global deduplication is an optional cluster-based feature that was designed to have little to no performance impact on guest workloads. There may be conditions where deduplication cannot be enabled (as it currently does not support stretched clusters or 2-Node clusters).

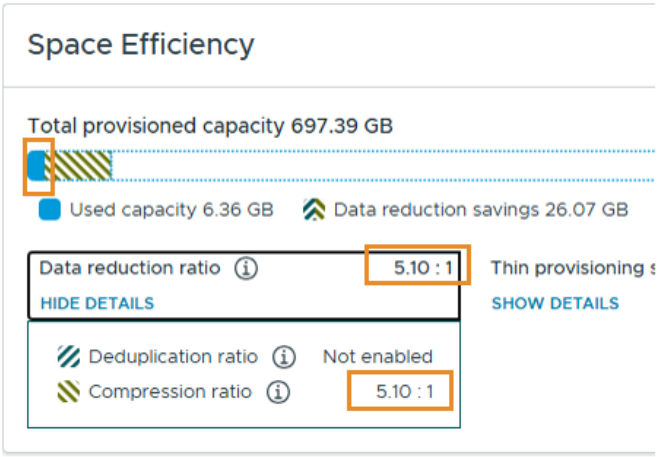
Recommendation: While vSAN global deduplication supports clusters sizes as large as 64 hosts, enabling deduplication on clusters that reside in a single rack may be preferable over clusters scattered across multiple racks. Clusters residing in a single rack typically only depend on a single pair of Top of Rack (ToR) switches to maintain connectivity between hosts – simplifying connectivity. If deduplication is used on larger clusters spanning multiple racks, ensure that your spine and leaf arrangement is highly durable and can maintain connectivity in the event of spine or leaf switch failures.

Comparing Typical Capacity Savings between Compression, and Deduplication

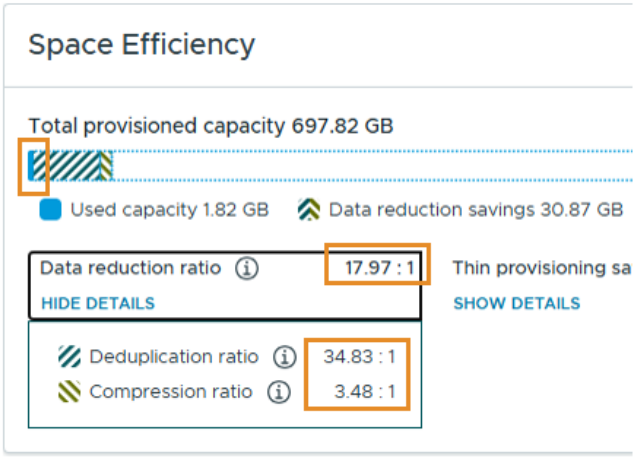
Data compression and deduplication are two independent approaches that strive for the same result: Lower effective capacity usage to drive down your cost of storage. They achieve data reduction in very different ways which yield different "potential" in terms of data reductions. These are both opportunistic space efficiency features that among other things, depend highly on the characteristics of the data, and are not guaranteed.

Deduplication typically has a much higher potential of data reduction than using just data compression. When viewing a cluster that has recently had global deduplication enabled, it is not unusual to see the compression ratio be reduced while this deduplication ratio increases. This is due to how data savings are calculated between the two technologies and is expected.

For example, in the illustration below, we see a cluster that with compression enabled, was yielding a certain compression ratio. But after enabling global deduplication, it shows a lower compression ratio, but a much higher overall data reduction ratio. Therefore, **we highly recommend that when evaluating savings, focus on the overall "Data reduction ratio" instead of discrete savings ratios from each technology.**



Data Compression



Data Compression with Global Deduplication

Deterministic Space Efficiency Features

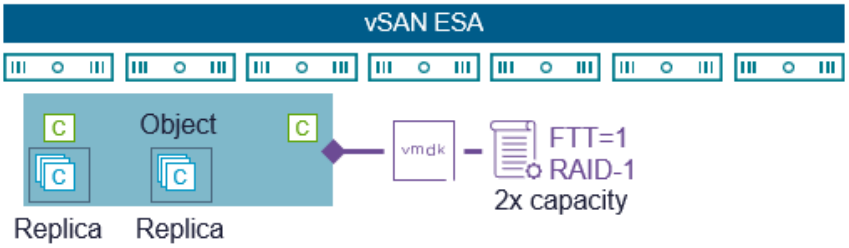
Deterministic space efficiency techniques will result in space efficiency that can be specifically determined. The degree to which the savings will occur will depend on the form of space efficiency used.

Data placement schemes and erasure codes have evolved with vSAN. Basic mirroring was the primary method of providing data resilience in earlier versions of vSAN, but vSAN ESA ushered in erasure coding without any compromise in performance, and is now the dominate method for ensuring resilience of data. The new Auto-RAID feature in vSAN for VCF 9.1 eliminates the need for customers to determine and set the appropriate protection level. It will automatically assign the most resilient and space efficient scheme, based on the characteristics of the cluster. The documentation below will assume the use of Auto-RAID.

The information provided below is based on data placement schemes used in vSAN for VCF 9.1, with an emphasis on methods used most often in vSAN ESA, in VCF 9.1. For more information on data placement schemes, see the document: "vSAN Availability Technologies." Auto-RAID in vSAN for VCF 9.1 has simplified which data placement schemes are used, and their respective overheads. For more information, see the post: "Auto-RAID in VMware vSAN for VCF 9.1 - Comprehensive System-Managed Data Resilience."

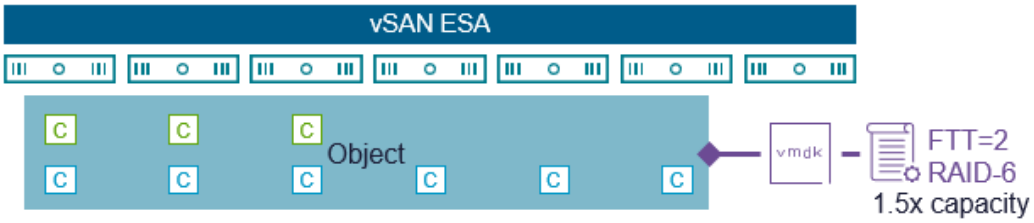
Data Placement Schemes and Erasure Code Concepts

vSAN achieves resilience of data in different ways. One way is by having a copy, or a mirror of a chunk of data (an object in vSAN) to one or more locations, or hosts. This replica of an object resides somewhere else in the cluster to maintain availability in the event of a failure. The level of resilience is defined in the assigned storage policy, and vSAN takes care of the rest, placing it in the cluster to achieve the desired result. A level of failure to tolerate of 1 (FTT=1) when using RAID-1 mirroring creates two copies of that object.



Data mirroring is a simple data placement scheme but comes with the tradeoff of using an equal amount of capacity somewhere else in the cluster to protect the object at the level of resilience you desire. When using Auto-RAID in vSAN for VCF 9.1, **data mirroring is not used except for 2-node and stretched cluster environments**, that must mirror the data to the other host or fault domain respectively.

The more common way **vSAN achieves data resilience is using erasure codes**. Erasure coding is a method of fragmenting data across some physical boundary in a manner that maintains access to the data in the event of a fragment or fragments missing. With vSAN, erasure codes strip this data with parity across hosts, to maintain availability in the event of a host failure. An object assigned an FTT=2 using erasure coding (RAID-6) will maintain availability in the event of a double failure, spreading that data across 6 hosts. **As a reminder, an object using erasure coding is not spread across all hosts.** vSAN's approach offers superior resilience under failure conditions and simplified scalability.



The benefit of erasure codes is predictable space efficiency when compared to mirroring data. In the example above, we see that compared to the RAID-1 mirror, the RAID-6 erasure code is more resilient (FTT=2 versus FTT=1), and consumes less capacity (1.5x the object capacity versus 2x the object capacity).

vSAN has unique capabilities with its erasure coding that make it not only highly space efficient, but often more durable and flexible than erasure codes implemented on traditional storage arrays. For more information, see the post: "Erasure Codes in VMware vSAN versus Storage Arrays."

Implications on Storage Capacity

In versions prior to vSAN in VCF 9.1, vSAN presented available and consumed capacity in the form of raw capacity. This was done in part to help account for vSAN's distributed architecture and giving customers the ability to assign unique resilience settings to discrete VMs through storage policies. As a result, overheads would vary. Rendering available and consumed capacity was challenging for many of our customers, since it was inconsistent with out storage was managed on a traditional storage array: See the posts: "Demystifying Capacity Reporting in vSAN" and: "Understanding "Reserved Capacity" Concepts" in vSAN.

RAID-5 Erasure Coding

RAID-5 erasure coding in vSAN ESA varies somewhat based on the version of vSAN ESA used, and it the topology. Prior to vSAN in VCF 9.1, vSAN ESA has historically used an adaptive RAID-5 erasure code.

- For clusters with 6 or more hosts, it uses a scheme of 4+1, spreading the data and parity fragments across at least 5 hosts. This consumes just 1.25x the capacity of the primary data.
- For clusters with fewer than 6 hosts, it will use a scheme of 2+1, spreading the data and parity fragments across at least 3 hosts. While this consumes a bit more capacity to make the data resilient (1.5x the capacity of the primary

data), it is much more space efficient than mirroring (2x capacity of the primary data), and can run on clusters with as few as three hosts, while also not impacting performance in any way. This makes RAID-5 an ideal choice for smaller clusters powered by the ESA.

With Auto-RAID in vSAN for VCF 9.1, RAID-5 will always (and only) be used in clusters between 3 and 5 hosts, and will ALWAYS use a 2+1 data placement scheme (as opposed to previous versions that may use a 4+1 scheme). One of the benefits of this approach is that it standardizes data resilience overheads (1.5x) to be the same as larger clusters that are capable of using RAID-6.

RAID-6 Erasure Coding

A RAID-6 erasure code in the ESA uses a 4+2 scheme, where data and parity fragments are spread across at least 6 hosts. With Auto-RAID in vSAN for VCF 9.1, RAID-6 will always (and only) be used in clusters with 6 or more hosts.

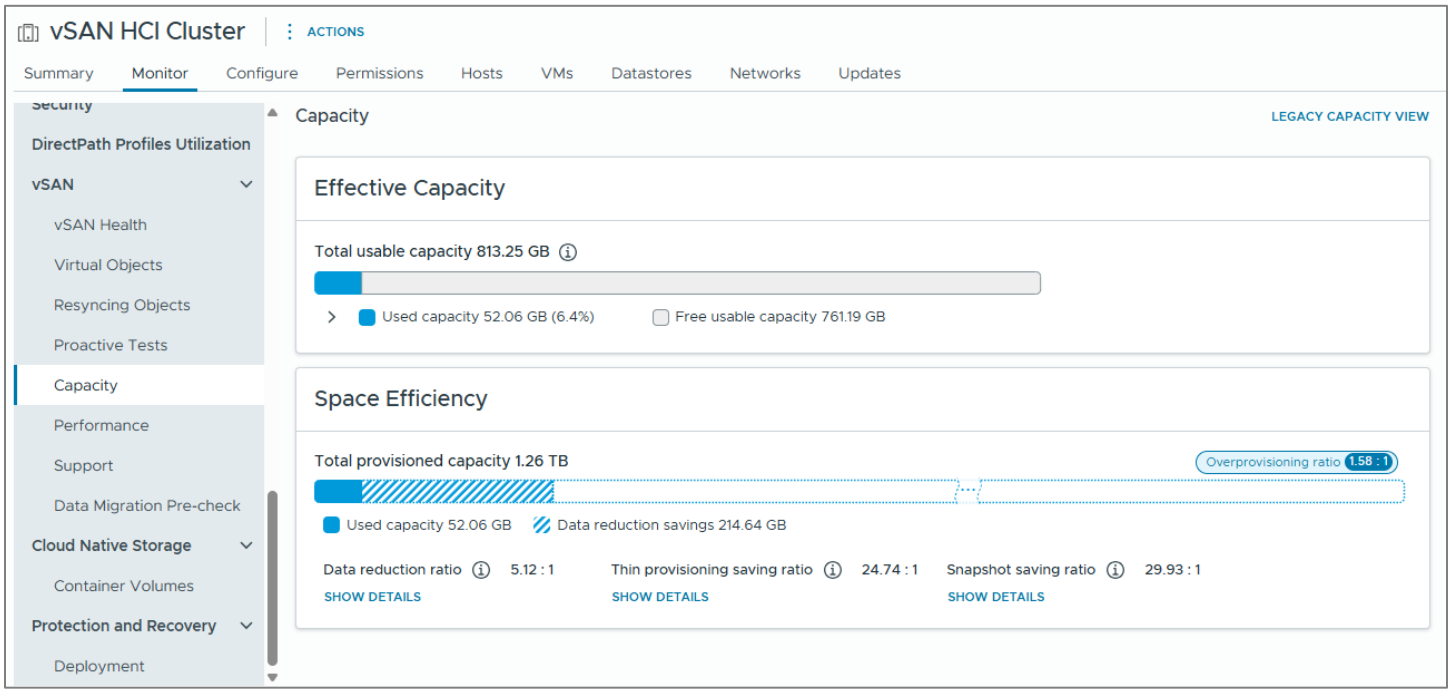
Interpreting Capacity Usage in vSAN

A Change in the Way vSAN Presents Capacity

Prior to vSAN in VCF 9.1, vSAN rendered storage in the form of raw capacity. It would simply aggregate all devices claimed by vSAN within the cluster and show this as the total amount of capacity. Data stored on the devices would not only account for the data itself, but the additional data needed to make it resilient across the cluster. Paired with the ability for administrators to use different storage policies to store data at different levels of resilience, presenting storage in the form of raw capacity provided the best option given the characteristics of vSAN. This approach did present challenges because it was different than how capacity was managed on a traditional storage array.

Thanks to the capabilities of Auto-RAID, vSAN in VCF 9.1 takes a new approach to rendering capacity information in vSAN. The new "Effective Capacity Reporting" view is capacity view to all vSAN 9.1 clusters. It eliminates concepts of raw capacity and other details related to overheads and now presents the actual effective usable capacity of a vSAN cluster. In other words, it conveys capacity information much like traditional storage. For more information, see the post: "Simplifying Storage with the New Effective Capacity View in VMware vSAN for VCF 9.1."

The view is divided into two sections: Effective Capacity and Space Efficiency.



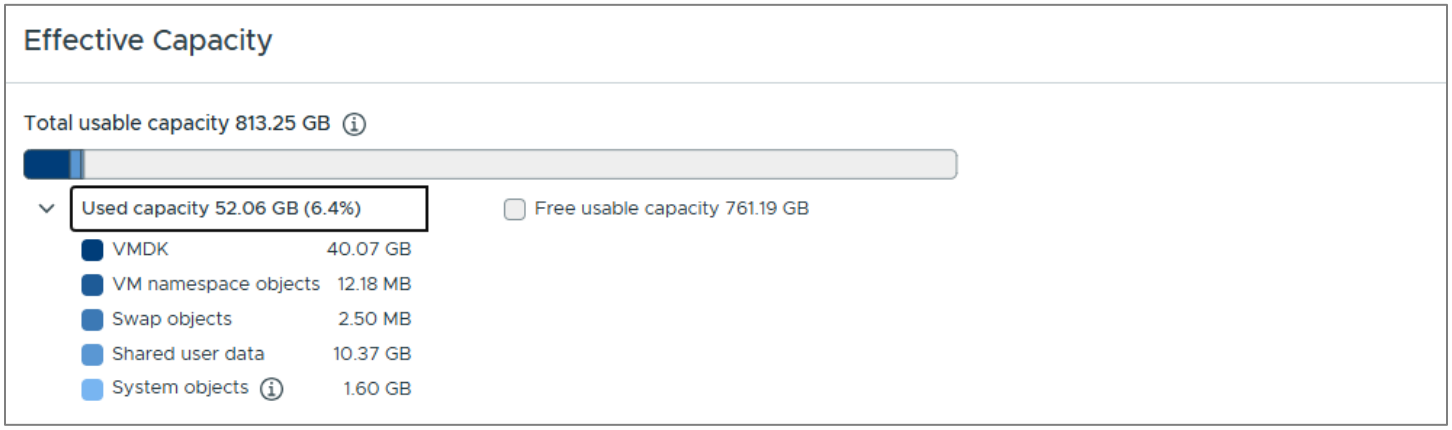
Each one of these sections will be discussed in more detail, below.

Capacity View - Effective Capacity Section

This section shows the total usable capacity of the cluster, in the form of a horizontal bar. It presents the **total usable capacity** after overheads are accounted for.

- Within that bar, the "used capacity" is the **total used capacity, after space efficiency capabilities have taken place on the data stored.**
- Within that bar, the "free usable capacity" presents the **total remaining usable capacity, prior to any potential space efficiency.** It does NOT attempt to forecast free usable capacity from opportunistic space efficiency features like compression and deduplication.

The view can also be expanded to see additional detail on the distribution of data stored.



Capacity View - Space Efficiency Section

The bottom portion of the new "Effective Capacity" view in vSAN will quantify how much space is regained using the specific type of opportunistic space efficiency feature. Capacity savings are provided for the following categories:

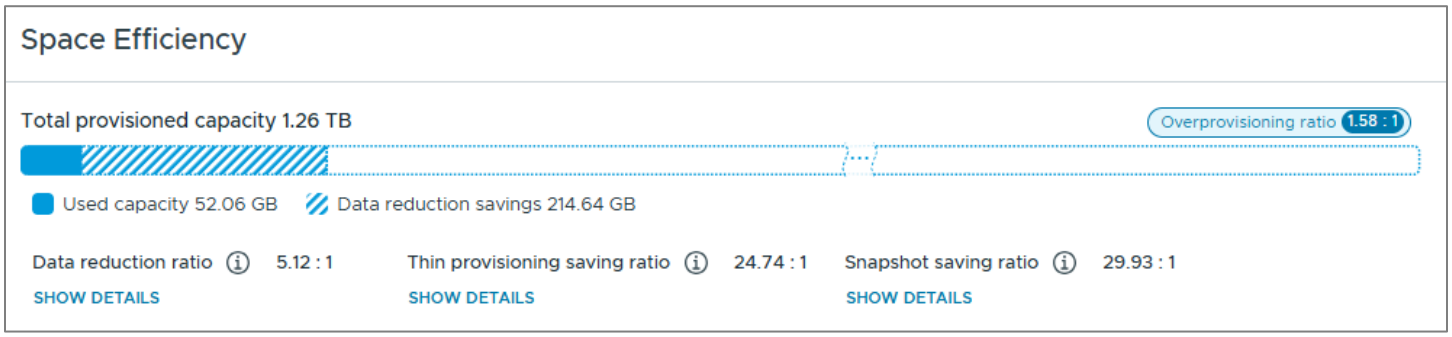
- Data reduction ratio
- Thin provisioning savings ratio
- Snapshot savings ratio

Each category can be expanded in the UI for additional levels of detail. For example, expanding on "Data reduction ratio" will show the discrete savings ratios for both deduplication, and compression. Note that these ratios may change as data is processed, and should be expected. For example, after vSAN writes a large amount of highly compressible data, the compression ratio may increase significantly. But after deduplication processes occur, some of the additional data reductions through deduplication will shift how it is reported. In this scenario, you will typically see the compression ratio decrease, the deduplication ratio increase, and the overall data reduction ratio improve due to it using the most optimal space efficiency technique.

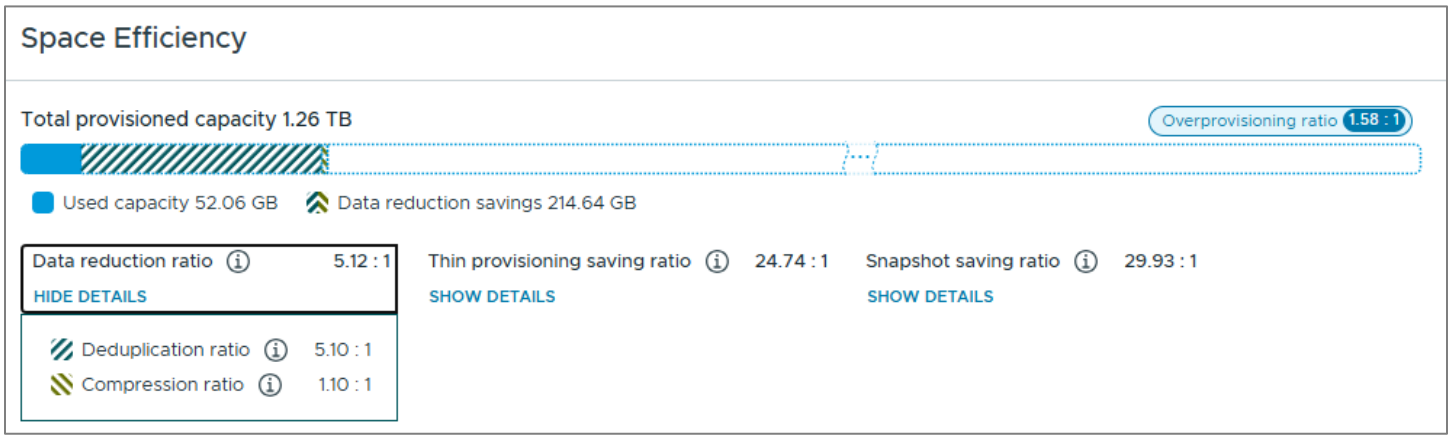
Recommendation: When using both data compression and global deduplication, monitor the overall "Data Reduction Ratio" rather than focusing on the discrete savings ratios of each. This overall data reduction ratio is what matters most in terms of driving down costs of storage.

Improving past editions, all space efficiency savings are expressed as a ratio, by stating a quantity of units of storage before the space efficiency feature followed by the quantity of units after. Examples would include "4:1" "6:1" etc. A higher ratio equals more capacity savings than a lower ratio. More information on data reduction ratios can be found in the post: "Save Costs and Scale Efficiently with vSAN Deduplication in VCF 9.0."

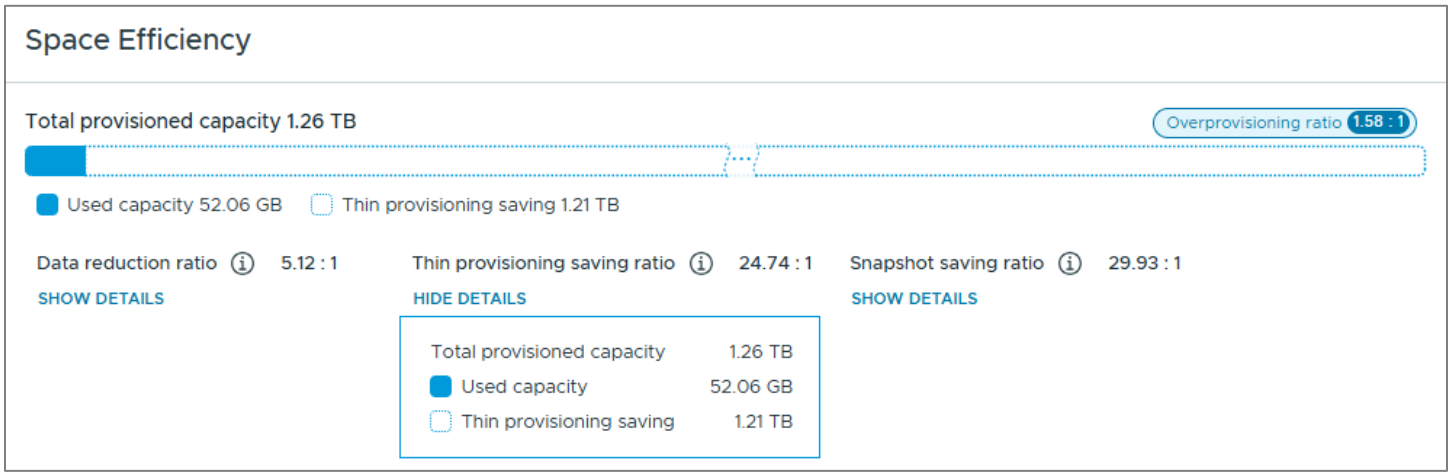
In addition to the capacity savings categories of “data reduction ratio, “thin provisioning savings ratio” and “snapshot savings ratio” the blue, horizontal bar will note the “overprovisioning ratio” on the right side of the image. This indicates that for physical storage provided by the cluster, one would need 1.58 times that amount if all the thin provisioned objects were fully provisioned.



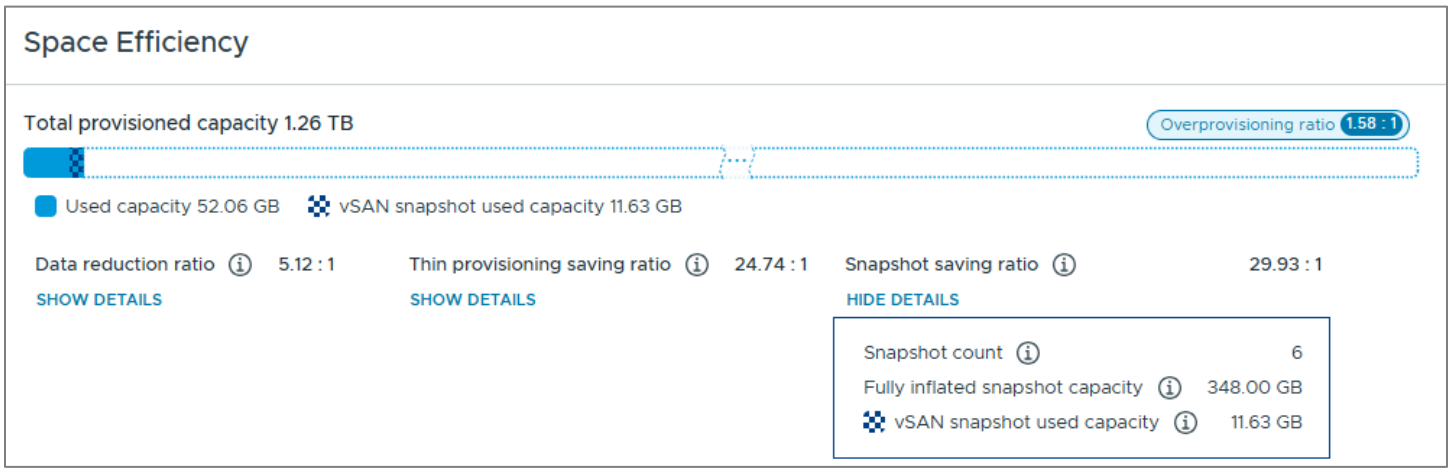
Expanding the “Data reduction ratio” as shown in the image below will present discrete savings ratios for both deduplication and compression. Since these are cluster-based data services, savings from compression and deduplication are only visible at this cluster-level view.



Expanding the “Thin provisioning savings ratio” as shown in the image below will provide details on the amount of capacity saved as a result of thin provisioning



Expanding the snapshot savings ratio as shown in the image below will provide the details on the number of snapshots residing on the datastore, and the used capacity versus the potential capacity used if it were a full copy of original data.



Recommendation: Do not use the "Datastores" view in the vSphere client to interpret capacity consumption or available capacity, or VM capacity statistics. It will show values that are not representative of the capacity statistics for the cluster, or the respective VMs stored in the datastore. Use the vSAN Capacity view for cluster-based statistics, and the VM inventory view for VM-related capacity statistics.

Standardized Capacity Overheads Courtesy of Auto-RAID

Auto-RAID in vSAN for VCF 9.1 will determine the optimal level of resilience for all data residing on each cluster. And since the erasure code overheads have been made consistent, this sets the stage for vSAN to be able to provide the effective capacity of a cluster, rather than the raw storage capacity provided in past editions. With vSAN in VCF 9.1 using Auto-RAID, the following list describes the overheads for various cluster types and sizes.

Standard Single Site Clusters

- 6 or more hosts: FTT=2 using RAID-6. **1.5x** capacity overhead.
- 3 to 5 hosts: FTT=1 using RAID-5. **1.5x** capacity overhead.
- Fewer than 3 hosts (excluding 2-Node clusters): FTT=0. 1.0x capacity overhead.

Stretched Clusters

- 6 or more hosts per site/fault domain: Site Disaster Tolerance using RAID-1 mirror. Secondary level of resilience using FTT=2 (RAID-6). **3.0x** capacity overhead.
- 3-5 hosts per site/fault domain: Site Disaster Tolerance using RAID-1 mirror. Secondary level of resilience using FTT=1 (RAID-5). **3.0x** capacity overhead.
- Fewer than 3 hosts in site/fault domain: Site Disaster Tolerance using RAID-1 mirror. No secondary level of resilience. 2.0x capacity overhead.

2-Node Clusters

- 6 or more storage devices per host: Site Disaster Toler using RAID-1 mirror. (secondary levels of resilience are not currently available for 2-Node clusters using Auto-RAID in 9.1). **2.0x** capacity overhead.
- 3-5 storage devices per host: Site Disaster Toler using RAID-1 mirror. (secondary levels of resilience are not currently available for 2-Node clusters using Auto-RAID in 9.1). **2.0x** capacity overhead.
- Fewer than 3 devices per host: Site Disaster Toler using RAID-1 mirror. 2.0x capacity overhead.

How Much "Effective Capacity" to Expect from a vSAN Cluster

The answer depends somewhat on the characteristics of the cluster, but the following information and example will show you how the new Effective Capacity view will render storage capacity, and demonstrate how you can create quick, simplified estimates of how much realistic capacity a new cluster bill of materials will yield – something that some found difficult to do in past versions.

Estimating Capacity

Here is a simple way to understand capacity capabilities. Data reduction ratios can be achieved using compression, deduplication, or both. Note that vSAN global deduplication is not currently supported in stretched cluster and 2-Node configurations.

Single site vSAN cluster (3 to 64 hosts)	
Cluster Data Reduction Ratio	Approximate Effective Capacity
1.5:1	Roughly 0.75x of the aggregate raw capacity of the cluster
2:1	Roughly equal to the aggregate raw capacity of cluster
3:1	Roughly 1.5x the aggregate raw capacity of the cluster
4:1	Roughly 2x the aggregate raw capacity of the cluster
5:1	Roughly 2.5x the aggregate raw capacity of the cluster
6:1	Roughly 3x the aggregate raw capacity of the cluster
7:1	Roughly 3.5x the aggregate raw capacity of the cluster
8:1	Roughly 4x the aggregate raw capacity of the cluster

Stretched clusters will mirror the data across the two data sites in addition to the host resilience within each site, and will cut the effective capacity in half, compared to a traditional, single site vSAN cluster.

vSAN stretched cluster (with a minimum of 3 hosts in each data site, or larger)	
Cluster Data Reduction Ratio	Approximate Effective Capacity
1.5:1	Roughly 0.38x of the aggregate raw capacity of the cluster
2:1	Roughly .5x the aggregate raw capacity of cluster
3:1	Roughly .75x the aggregate raw capacity of the cluster
4:1	Roughly equal to the aggregate raw capacity of the cluster
5:1	Roughly 1.25x the aggregate raw capacity of the cluster
6:1	Roughly 1.5x the aggregate raw capacity of the cluster
7:1	Roughly 1.75x the aggregate raw capacity of the cluster
8:1	Roughly 2x the aggregate raw capacity of the cluster

2-Node clusters, which currently do not support a secondary level of resilience when using Auto-RAID in vSAN for VCF 9.1, will simply mirror the data between the two hosts.

vSAN 2-Node cluster (no secondary level of resilience)	
Cluster Data Reduction Ratio	Approximate Effective Capacity
1.5:1	Roughly 0.56x of the aggregate raw capacity of the cluster
2:1	Roughly .75x the aggregate raw capacity of cluster
3:1	Roughly 1.13x the aggregate raw capacity of the cluster
4:1	Roughly 1.50x the aggregate raw capacity of the cluster
5:1	Roughly 1.88x the aggregate raw capacity of the cluster
6:1	Roughly 2.25x the aggregate raw capacity of the cluster
7:1	Roughly 2.63x the aggregate raw capacity of the cluster
8:1	Roughly 3x the aggregate raw capacity of the cluster

Example Configuration

10 hosts in a standard single site vSAN cluster, with approximately 40TB per host would provide just under 400TB raw capacity (more specifically, 394TB) as seen in the legacy capacity view. This same cluster, as viewed in the new Effective Capacity view would show just over 200TB in effective capacity (more specifically, 203TB), or approximately half of the raw capacity.

But the actual "used capacity" on the cluster would be after all the space efficiency techniques are applied. Therefore, an estimate of effective usable capacity could be easily derived with the following equation:

- In this example, if the customer was getting a **2:1 data reduction ratio**, they would be able to store up to about 400TB, **roughly equal to the aggregate raw capacity of the hosts**.
- In this example, if the customer was getting a **4:1 data reduction ratio**, they would be able to store up to about 800TB, **about 2x the aggregate raw capacity of the hosts.**
- In this example, if the customer was getting a **6:1 data reduction ratio**, they would be able to store up to about 1.2PB, **about 3x the aggregate raw capacity of the hosts.**
- In this example, if the customer was getting an **8:1 data reduction ratio**, they would be able to store up to about 1.6PB, **about 4x the aggregate raw capacity of the hosts.**

As noted above, when comparing a stretched cluster topology to a single site vSAN cluster, the mirroring of data across two sites will cut the results in half. For example, a 10 host cluster with the same host specifications noted above, but configured as a 5+5+1 stretched cluster with a 2:1 data reduction ratio, would yield about 200TB, roughly 50% of the aggregate raw capacity of the hosts. This overhead remains the same regardless of the underlying erasure code providing that secondary level of resilience.

Note that Reserved capacity concepts such as "Host Rebuild Reserve" and "Operations Reserve" found in previous versions of vSAN are constructs of the old method of capacity management. They are no longer relevant, nor available in the new Effective Capacity view, as all necessary overheads are already accounted for.

Summary

Space efficiency techniques are a common way to improve efficiency with data storage, and as a result, drives down the effective cost of the storage solution. VMware vSAN provides several forms of opportunistic and deterministic space efficiency features that are easy to implement. vSAN will be able to deliver numerous space saving technologies with supreme levels of performance, and simplified management experience.

Additional Resources

vSAN technical blogs. Stay up to date on the most recently published technical information about vSAN. These posts are created by the vSAN Technical Marketing team.

VMware Resource Center. The location for design guides, operations guides and other technical white papers on vSAN. These assets are created by the vSAN Technical Marketing and Product Enablement teams.

Official vSAN documentation. The location for all "how to" documentation on vSAN.

About the Author

Pete Koehler is a Product Marketing Engineer in the VCF division at Broadcom. With a primary focus on vSAN, Pete covers topics such as design and sizing, operations, performance, troubleshooting, and integration with other products and platforms.

Appendix: TRIM/UNMAP Operations

TRIM/UNMAP can be enabled using PowerCLI. Enabling via PowerCLI is shown below.

Status query:

```
Get-Cluster -name R63*|get-VsanClusterConfiguration |ft GuestTrimUnmap
GuestTrimUnmap
--------------
         False
```

Enable:

```
Get-Cluster -name R63*|set-VsanClusterConfiguration -GuestTrimUnmap:$true
Cluster              VsanEnabled  IsStretchedCluster   Last HCL Updated
-------              -----------  ------------------   ----------------
R630-Cluster-70GA    True         True                 25/04/2020 16:03:00
```

Deactivate

```
Get-Cluster -name R63*|set-VsanClusterConfiguration -GuestTrimUnmap:$false
Cluster              VsanEnabled  IsStretchedCluster   Last HCL Updated
-------              -----------  ------------------   ----------------
R630-Cluster-70GA    True         True                 25/04/2020 16:03:00
```

Prerequisites - VM Level

Once enabled, there are several prerequisites that must be met for TRIM/UNMAP to successfully reclaim the capacity no longer used.

- A minimum of virtual machine hardware version 11 for Windows
- A minimum of virtual machine hardware version 13 for Linux.
- disk.scsiUnmapAllowed flag is not set to false. The default is implied true. This setting can be used as a "stop switch" at the virtual machine level should you wish to disable this behavior on a per VM basis and do not want to use in guest configuration to disable this behavior. VMX changes require a reboot to take effect.
- The guest operating system must be able to identify the virtual disk as thin.
- After enabling at a cluster level, the **VM must be powered off and back on**. (A reboot is insufficient). This requirement is a one-time operation on each VM after it is enabled.

Linux Specific Guidance

There are two primary means of reclaiming thin provisioning.

- fstrim is used on a mounted filesystem to discard (or "trim") blocks which are not in use by the filesystem. This is useful for thinly-provisioned storage. Depending on the distribution, this may or may not be included in a cron job, such as /etc/cron.weekly. To manually perform the in-guest reclamation, perform the following:

  ```
  /sbin/fstrim --all || true
  ```

- blkdiscard is used to discard device sectors. Unlike strip(8), this command is used directly on the block device. blkdisacrd is known to have more performance overhead than fstrim. As a result, fstrim is recommended over blkdiscard.

Other Concerns:

- If using encrypted file systems, you may need to add discard to /etc/crypttab.

- If shrinking or deleting LVM volumes, the issue_discards configuration may be needed in /etc/lvm/lvm.conf
- Different options for automating the running of fstrim exist. These range from weekly cron tasks, to fstrim.timer
- The following file systems are reported to work with TRIM: btrfs, ecryptfs, ext3, ext4, f2fs, gfs2, jfs, nilfs2, ocfs2, xfs

Microsoft Specific Guidance

Windows Server 2012 and newer support automated space reclamation. This behavior is enabled by default. To check this behavior, the following PowerShell can be used.

```
Get-ItemProperty
-Path
"HKLM:\System\CurrentControlSet\Control\FileSystem"
-Name
DisableDeleteNotification
```

To enable automatic space reclamation this value the following:

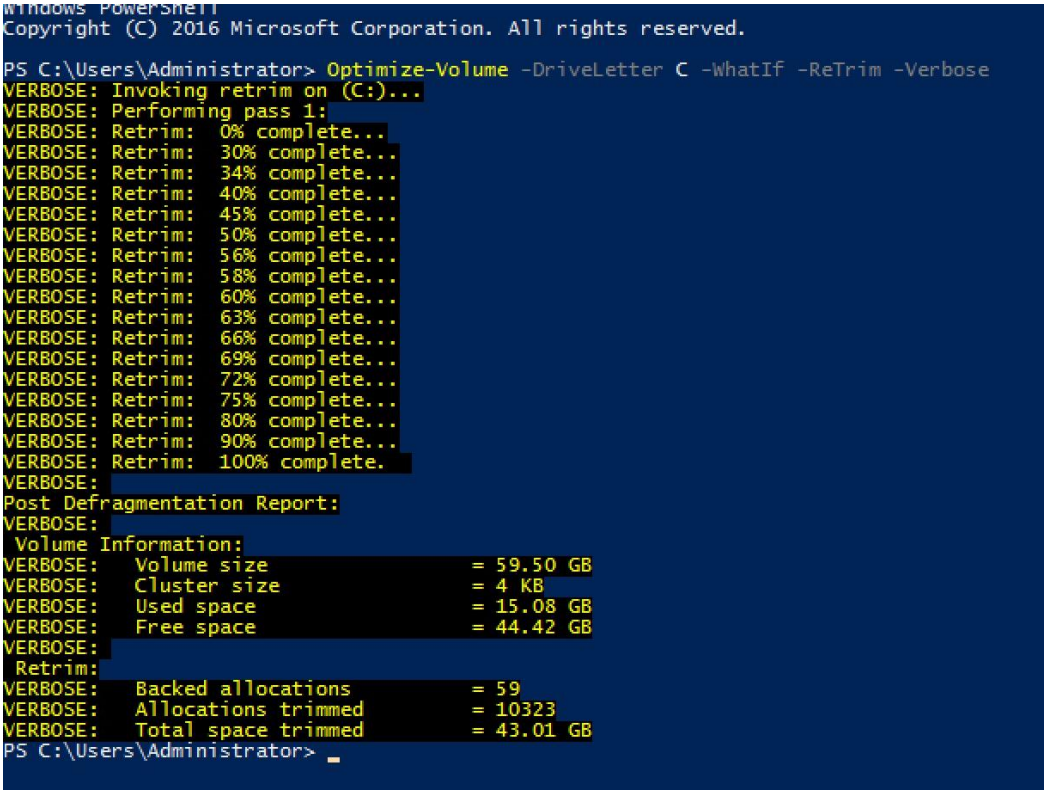
```
Set-ItemProperty
-Path
"HKLM:\System\CurrentControlSet\Control\FileSystem"
-Name
DisableDeleteNotification
-Value
0
```

Windows also offers asynchronous space reclamation, and can be achieved as shown in the following two examples:

Example 1: Perform TRIM optimization using PowerShell

```
PS C:\>Optimize-Volume -DriveLetter H -ReTrim -Verbose
```

This example optimizes drive H by re-sending Trim requests. The -WhatIf flag can be added to test if TRIM commands are being passed cleanly to the backend.



Example 2: Perform TRIM optimization using the command line

defrag /L can also be used to perform the same operation as the Optimize-Storage -ReTrim command.

```
Defrag C: D: /L
```

This Example would reclaim space on both volume C: and D:.

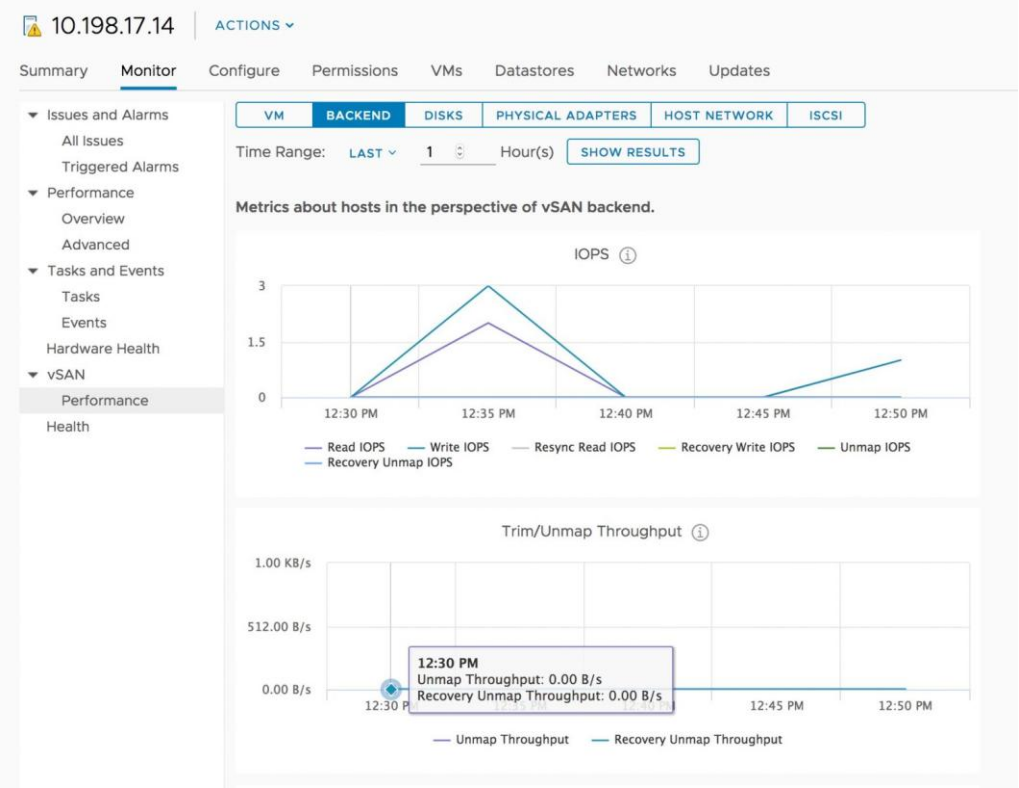
Other Concerns

- Windows when using Optimize-Storage or Defrag /L will report sending TRIM commands to all unused blocks. This reporting should not be relied upon up to determine how much space will be reclaimed.
- It is recommended to primarily use the automatic Delete Notification.

Monitoring TRIM/UNMAP

TRIM/UNMAP has the following counters in the vSAN performance service:

- **UNMAP Throughput.** The measure of UNMAP commands being processed by the disk groups of a host
- **Recovery UNMAP Throughput.** The measure of throughput of UNMAP commands is synchronized as part of an object repair following a failure or absent object.



VADP Backup Considerations

Snapshots are commonly used by backup vendors to provide a known state for backing up VM data at a given point in time. When a snapshot is no longer needed by the backup process, those changes are merged with the base disk. Snapshots and other tasks such as Storage vMotion use VMware's mirror driver to ensure that ongoing changed blocks are mirrored to more than one location. In these circumstances where the mirror driver is used, TRIM/UNMAP commands will not be passed down to the base disk, preventing full reclamation from occurring. One method of accommodating for these conditions is to use a pre-freeze-script.

For Windows:

```
C:\Windows\pre-freeze-script.bat
```

For all other operating systems:

```
/usr/sbin/pre-freeze-script
```

Running the fstrim or DiskOptimize before a snapshot will clean out any deleted files that happened during a previous backup window.

UNMAP with vSAN File Services

vSAN File Services supports UNMAP and will automatically reclaim storage when files are deleted from the NFS or SMB share. TRIM/UNMAP must be enabled on the cluster for this process to take place. Depending on the circumstances, it may take a few minutes for the reclaimed storage to be reported correctly within vCenter Server.