Understanding VMware vSphere® 5.1 Storage DRS™

TECHNICAL MARKETING DOCUMENTATION
V1.0/UPDATED MARCH 2013

vmware®

Table of Contents

Introduction to VMware vSphere Storage DRS

Virtual machine storage provisioning historically has imposed operational challenges. Monitoring datastore capacity and I/O load has proven to be very difficult and as a result is often neglected. During the provisioning process for virtual machines, virtual disk datastores are often randomly selected, leading to hot spots and over- or underutilized datastores. VMware vSphere® Storage DRS™ was a new feature introduced in VMware vSphere 5.0 that provides smart virtual machine placement and load-balancing mechanisms based on I/O and space capacity. It helps decrease operational effort associated with the provisioning of virtual machines and monitoring of the storage environment.

There are five key features that vSphere Storage DRS offers:

• Resource aggregation
• Initial placement
• Load balancing
• Affinity rules
• Datastore maintenance mode

Resource Aggregation

Resource aggregation is the key component that all other vSphere Storage DRS features depend on. vSphere 5.0 introduced a new VMware® vCenter Server™ object called a datastore cluster—a collection of datastores aggregated to a single unit of consumption from an administrator's perspective. These datastore clusters form the basis of vSphere Storage DRS. When one is created, vSphere Storage DRS can manage the storage resources in a comparable manner to how VMware vSphere Distributed Resource Scheduler™ (vSphere DRS) manages compute resources in a cluster. As with a cluster of hosts, a datastore cluster is used to aggregate storage resources, enabling smart and rapid placement of the virtual disk files of a virtual machine and the load balancing of existing workloads. Figure 1 depicts a datastore cluster of 12TB formed by four 3TB datastores.
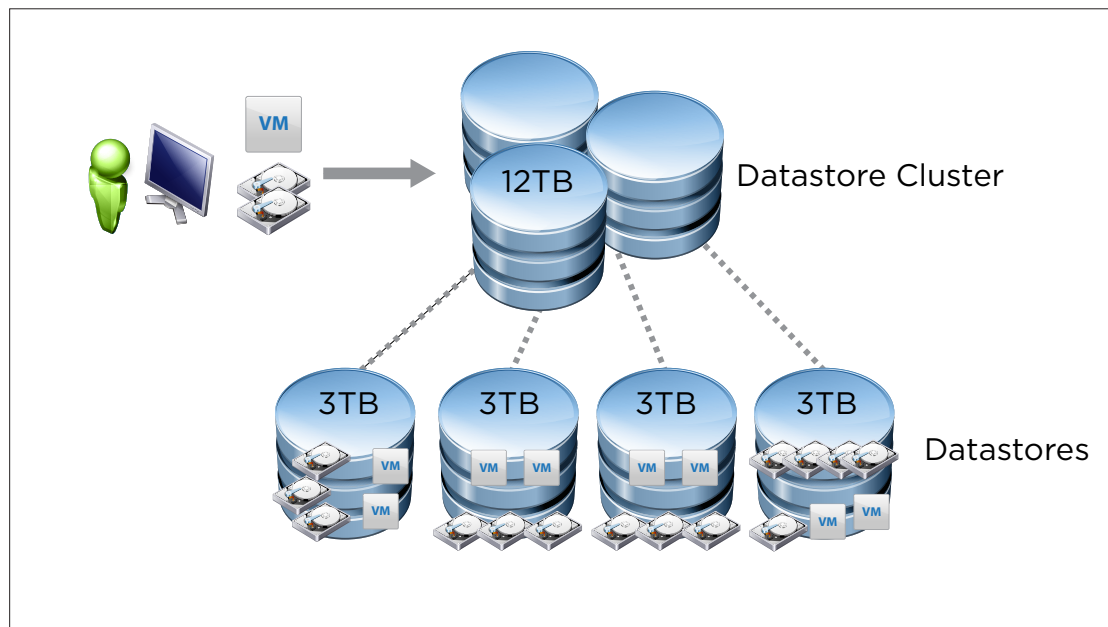

Figure 1. Datastore Cluster

The flexibility of this new abstraction, further separating the physical from the logical, greatly simplifies storage management by enabling datastores to be quickly and dynamically added to or removed from a datastore cluster in response to maintenance, performance or out-of-space issues.

The datastore cluster functions as the vSphere Storage DRS load-balancing domain and should consist of datastores that are similar in size and performance characteristics. The vSphere Storage DRS load balancer performs initial virtual machine disk format (VMDK) file placement and provides efficient and balanced use of datastores.

vSphere Storage DRS 5.1 supports as many as 32 datastores in a single datastore cluster; a vCenter Server instance can support as many as 256 datastore clusters.

A datastore cluster provides a single point of management. vSphere Storage DRS settings are applied per datastore cluster level and relate to each datastore and virtual machine inside the datastore cluster.

A datastore cluster presents an abstraction layer for the provisioning process. During the virtual machine provisioning process, a datastore cluster—rather than an individual datastore—is selected as a target destination. This abstraction layer gives the administrator the ability to add or remove datastores without having to change the virtual machine provisioning process.
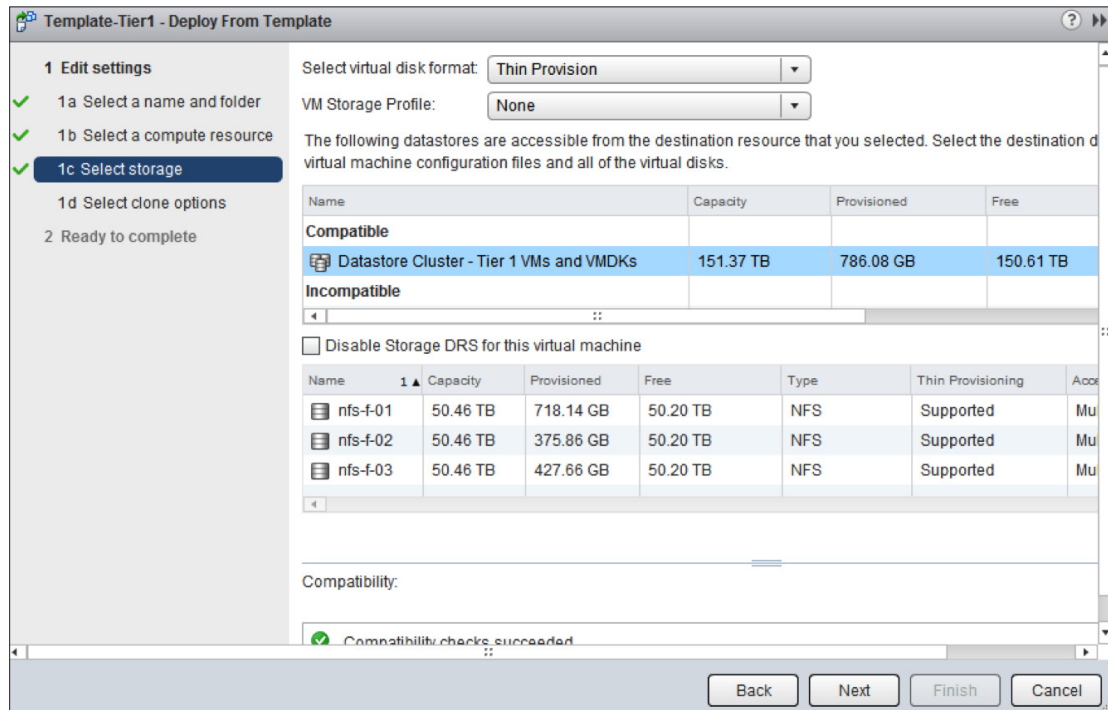
Figure 2. Storage Destination Selection

Resource aggregation of datastores also helps utilize and incorporate new datastores into the virtual machine provisioning process. Before datastore clusters, administrators in charge of virtual machine provisioning had to be made aware of new datastores. Typically a (static) list is used, containing datastores that are suitable for a specific service-level agreement (SLA). Maintaining, controlling and using such a list is a tedious job. With datastore clusters, newly created datastores can be added to the cluster, and vSphere Storage DRS can automatically incorporate them for selection during new virtual machine placement.

Initial Placement

The datastore selection process has proven to be complex during the manual provisioning of a virtual machine. Crucial provisioning factors such as current space utilization and I/O load are often ignored. vSphere Storage DRS provides initial placement and ongoing balancing recommendations, helping vSphere administrators make placement decisions based on space and I/O capacity.

Initial placement is the process whereby vSphere Storage DRS selects a datastore within a datastore cluster on which to place a virtual machine disk. This process occurs when the virtual machine is being created or cloned, when a virtual machine disk is being migrated to another datastore cluster or when a disk is added to an existing virtual machine.

vSphere Storage DRS ensures that initial placement recommendations are made in accordance with space constraints and with respect to the goals of space utilization and I/O load balancing. These goals aim to minimize both the risk of storage I/O bottlenecks and the performance impact on virtual machines.

Initial placement simplifies and speeds up the provisioning process by automating the selection of a datastore. The aggregation of multiple datastores into a single datastore cluster offers administrators a much smaller-scale decision because they now could select only a single datastore cluster instead of picking one datastore out of many.

Load Balancing

Storage systems typically are provisioned based on capacity and reliability needs. Diverse I/O performance from various workloads and hot-spotting can cause significant imbalance across datastores over time. Load balancing can be thought of as a tool that proactively prevents high latencies and out-of-space scenarios resulting from overload on individual datastores. As load imbalances begin to occur, vSphere Storage DRS makes recommendations to correct them. These imbalances might not have been correlated with a particular datastore's having become overloaded or its experiencing hot-spotting, in which case the moves can be considered proactive. Space-utilization load balancing is designed to be reactive, to alleviate bottlenecks or extreme imbalances.

Affinity Rules

vSphere Storage DRS applies smart placement rules, in the form of affinity rules, on the virtual machine files. Affinity rules help prevent placing virtual machines with similar tasks on the same datastore; they also help keep virtual machines together when required.

Datastore Maintenance Mode

Datastore maintenance mode is a feature that typically will not be used often but will be appreciated when needed. It can be compared to host maintenance mode. When a datastore is placed in maintenance mode, all registered virtual machines on that datastore are migrated to other datastores in the datastore cluster. Typical use cases are data migration to a new storage array and VMFS-3–to–VMFS-5 migration.

Requirements

For vSphere Storage DRS to function correctly, the environment must meet the following basic requirements:

• VMware vCenter Server 5.0 or later
• VMware ESXi™ 5.0 or later
• VMware vCenter cluster (recommended)
• VMware vSphere Enterprise Plus Edition™ license
• Shared VMware vSphere VMFS or NFS datastores
• Datastores visible in only one vCenter datacenter object

vSphere Storage DRS Operational Details

Initial Placement Automation Level

vSphere Storage DRS generates placement recommendations that must be manually approved by the administrator, regardless of the load-balancing automation level of the datastore cluster. The user interface (UI) provides the user with the most optimal placement recommendation.
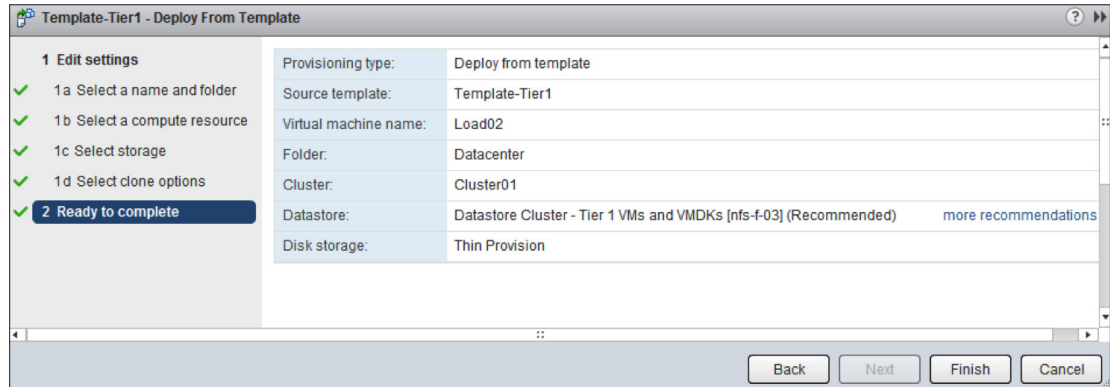

Figure 3. Best Placement Recommendation Included in Provisioning Process

Depicted in Figure 3, **more recommendations** shows the remaining placement recommendations in descending order, listing the optimal placement at the top.
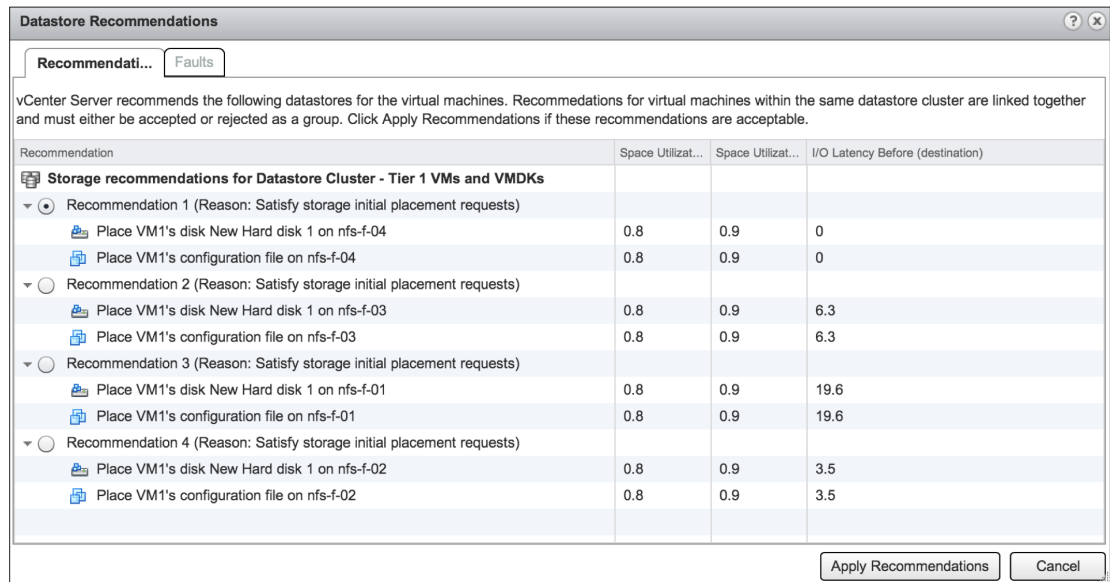
Figure 4. Additional Initial Placement Recommendations

Based on available space and I/O latency metrics, vSphere Storage DRS selects the most appropriate datastore for the new virtual machine. In Figure 4, the UI shows the space utilization of the datastore before and after the new virtual machine is placed onto it. The initial placement UI shows the measured I/O latency before the virtual machine is placed. vSphere Storage DRS does not detect the workload demand of the virtual machine, so the UI does not show an "I/O Latency After" (destination) column.

Space Utilization % Before (destination)	Space Utilization % After (destination)	I/O Latency Before (destination)
0.8	0.9	0
0.8	0.9	0
0.8	0.9	6.3
0.8	0.9	6.3

Figure 5. Initial Placement Metrics

Datastore Cluster Defragmentation

Datastore cluster defragmentation provides additional reduction of complexity in the provisioning process. If the datastore cluster contains adequate free space but each individual datastore does not, vSphere Storage DRS migrates virtual machines to other datastores to make room for the initial placement of a new virtual machine or disk.

Prerequisite Migrations

vSphere Storage DRS attempts to make room for the new virtual machine by searching alternative locations for existing virtual machines. During this analysis, it simulates placing virtual machines on other datastores one by one. This generates a combined set of migration recommendations. The migration of existing virtual machines is called prerequisite migration. A recommendation set includes prerequisite migrations combined with the initial placement recommendation of the new virtual machine.

Depth of Recursion

The initial placement process uses a recursive algorithm for searching alternative placement options. To avoid generating an infinite list of combinations and to reduce the overhead of a recommendation set on the existing environment, vSphere Storage DRS depth of recursion is limited to two steps.

A step is a set of migrations from a single datastore to make room for migration into that datastore. A step can contain one or multiple virtual machine (disk) migrations. A combination of multiple steps is supported. Figure 6 illustrates this process.
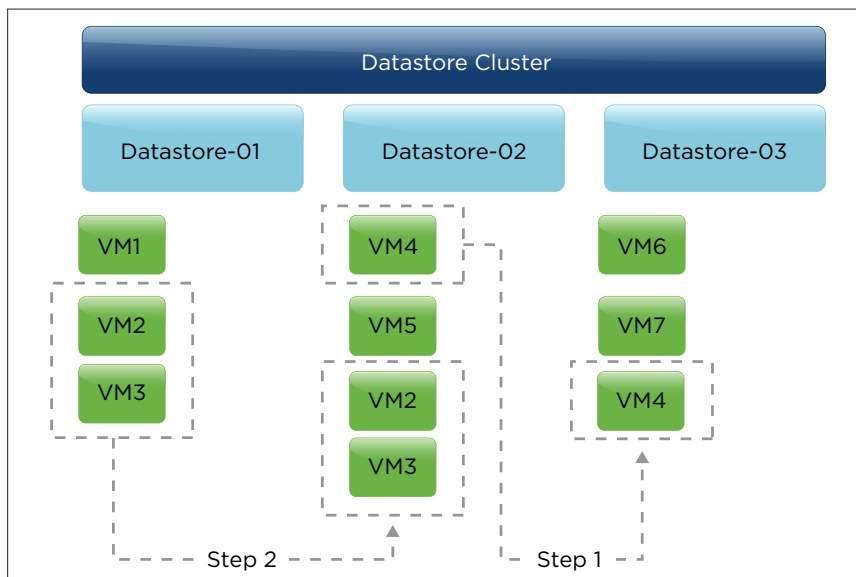

Figure 6. Prerequisite Migration Steps

To place the new virtual machine, free space must be created. **VM2** and **VM3** can be migrated to **Datastore-02**, but this move would violate the datastore utilization threshold. Therefore, vSphere Storage DRS generates an additional prerequisite migration that recommends moving **VM4** to **Datastore-03**.

vSphere Storage DRS generates a recommendation set after analyzing the cost and benefit of each combined set of recommendations. The recommendation set placed at the top will be the one containing the fewest moves and migrating the least data. vSphere Storage DRS affinity rules are respected during the analysis and the migration recommendation process. If the virtual machine is configured with an affinity rule, vSphere Storage DRS keeps the files together in the working directory of the virtual machine. If a virtual machine is configured with a VMDK anti-affinity rule, vSphere Storage DRS can migrate individual VMDK files, enabling it to distribute more granularly.

Thin-Provisioned VMDK

vSphere Storage DRS supports the use of thin-provisioned VMDK files. When placing a virtual machine in a datastore cluster, the initial placement process uses a very conservative algorithm when determining an appropriate datastore. The process applies two different calculations during the analysis.

Provisioned Virtual Machine

vSphere Storage DRS cannot determine the actual space usage for the virtual machine, so the initial placement process uses the provisioned size of the VMDK file in its calculation. In the screenshot in Figure 7, **Hard disk 1** has a provisioned size of **60GB**.
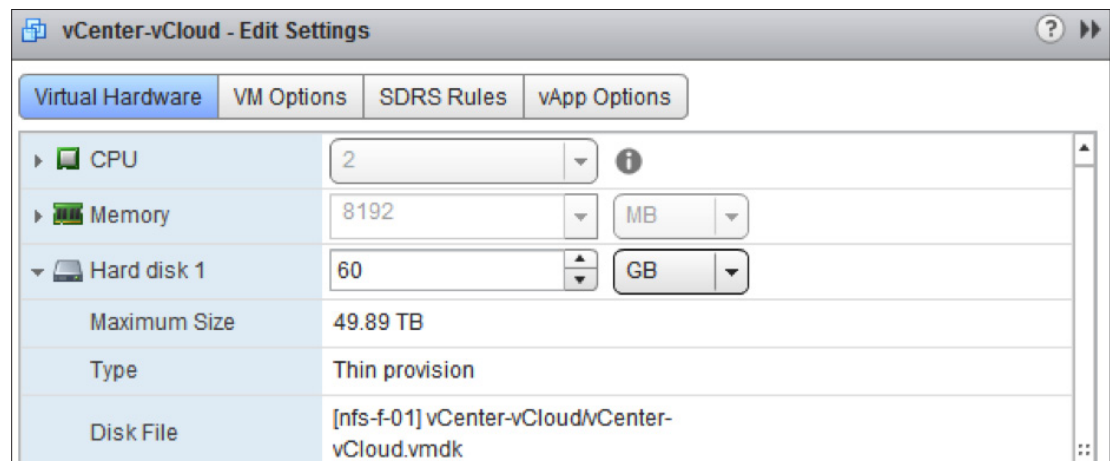

Figure 7. Virtual Machine Provisioned Space – Hard Disk 1

This conservative approach guarantees that the destination datastore is capable of storing the virtual machine in the worst-case scenario of 100 percent space utilization. This prevents vSphere Storage DRS from migrating a virtual machine to another datastore directly after deployment.

Existing Virtual Machine

Datastores report the allocated space of the existing virtual machines stored on them to vSphere Storage DRS during initial placement and also for migration recommendations. To cope with space-utilization growth of the existing thin-provisioned virtual machines during an initial placement or load-balancing operation, vSphere Storage DRS incorporates a safety buffer space in its calculation.

Entitled Space Use

vSphere Storage DRS calculates the entitled space use—the sum of current allocated space plus the safety buffer space—of a virtual machine. The buffer space is determined by the advanced setting **PercentIdleMBinSpaceDemand**. The PercentIdleMBinSpaceDemand default value is set to 25 percent, indicating that vSphere Storage DRS must add 25 percent of the IdleMB—the space difference between the allocated space and the provisioned space—to the allocated space of the VMDK file. In this example, the thin-provisioned VMDK is 60GB in size (provisioned space) and the virtual machine is using 20GB. The IdleMB within the VMDK is 40GB.
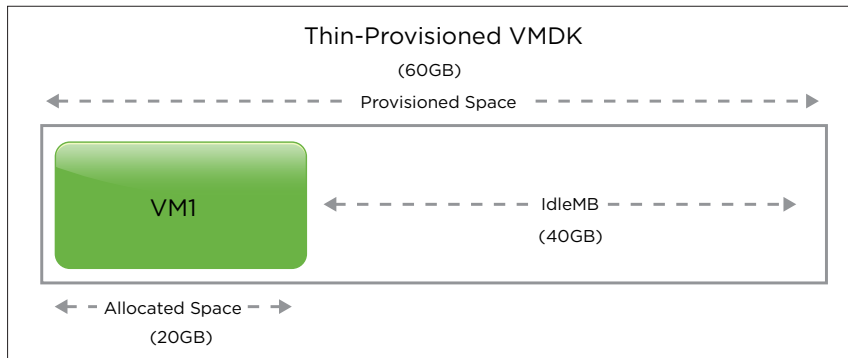
Figure 8. IdleMB Calculation

The PercentIdleMBinSpaceDemand is applied (25 percent) to the unallocated space: 25 percent of 40GB = 10GB. This results in an entitled space of 30GB (20GB + 10GB = 30GB). The entitled use of this virtual machine is 30GB.
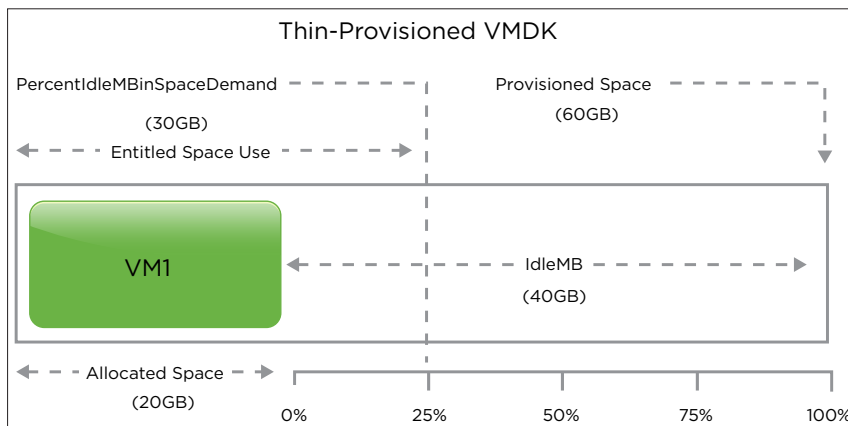
Figure 9. Entitled Space Calculation

In Figure 10, the virtual machine is stored on a datastore of 100GB. vSphere Storage DRS determines that this datastore has 70GB of available free space.
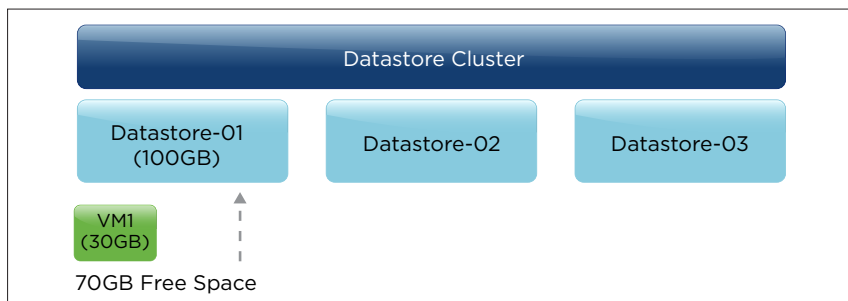
Figure 10. Datastore Free Space

ADVANCED SETTING	DEFAULT SETTING	MINIMUM VALUE	MAXIMUM VALUE
PercentIdleMBinSpaceDemand	25	0	100

Table 1. PercentIdleMBinSpaceDemand

Profile-Driven Storage

Managing virtual machines and ensuring that they are placed on datastores that match their SLA requirements can be a cumbersome and challenging task. vSphere 5.0 introduced Profile-Driven Storage, which enables the capture of underlying storage capabilities—such as space-saving technologies, replication and protocol definitions—either automatically via a vSphere Storage APIs – Storage Awareness (VASA) provider or manually by the user.

Virtual Machine Storage Profile Architecture

Virtual machine storage profiles list the storage capabilities that virtual machine disk files require to run the applications within the virtual machine. A virtual machine storage profile is attached to a storage capability. In turn, a storage capability profile is attached to a datastore. Coupled with virtual machine storage profiles and based on specific SLA requirements, Profile-Driven Storage enables the user to rapidly and intelligently place the virtual machines.
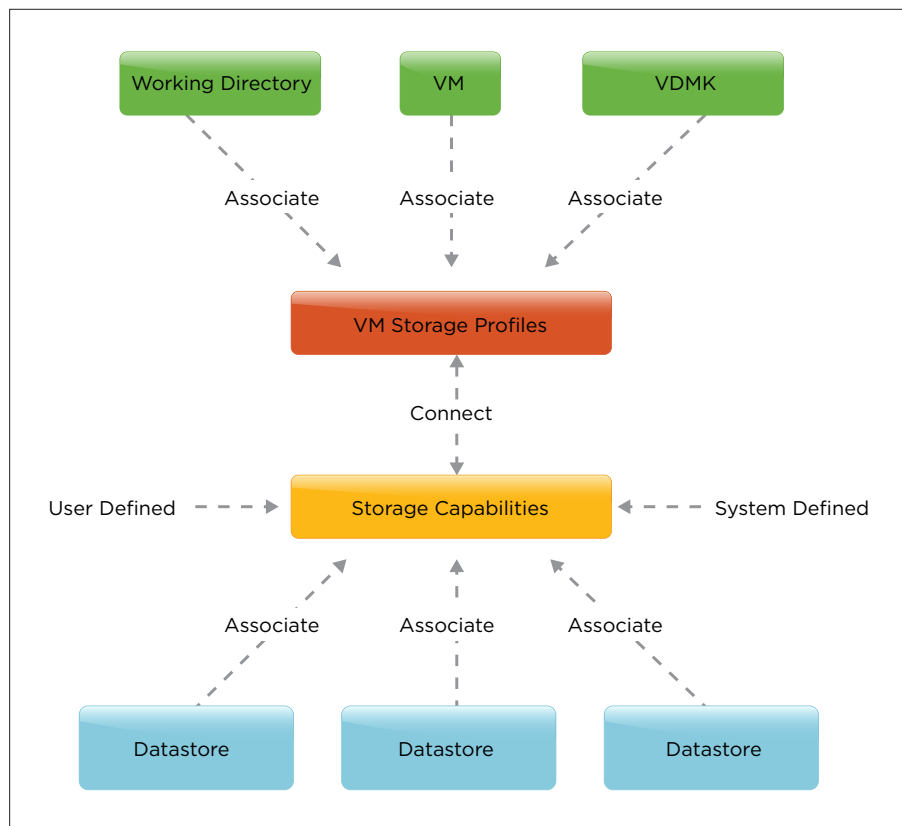

Figure 11. Storage Profile Architecture

vSphere Storage DRS supports the use of virtual machine storage profiles, the configuring of which is outside the scope of this white paper. It requires datastores with identical storage characteristics and capabilities within a single datastore cluster so the cluster is compatible with the virtual machine storage profile. This enables the initial placement process to display datastore clusters as compatible storage destinations.

Distribute a Virtual Machine Across Datastore Clusters

A storage profile can be assigned to a virtual machine but can also be assigned on a more granular, per-VMDK, level. If multiple datastore clusters exist, with each having a different storage capability, a virtual machine can be distributed across them. In each datastore cluster, vSphere Storage DRS can apply both initial placement and I/O load-balancing operations on the virtual machine files. It places and moves the virtual machine files aligned with the configured affinity rule.
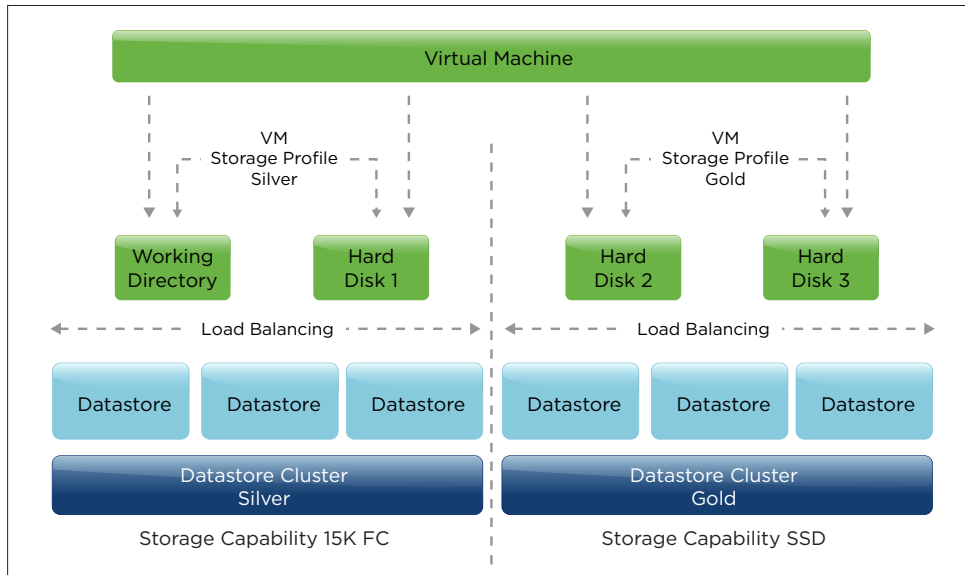
Figure 12. Distributed Virtual Machine Across Multiple Datastore Clusters

Load Balancing

Initial placement reduces complexity in the provisioning process; load balancing addresses imbalances within a datastore cluster. vSphere Storage DRS generates load-balancing recommendations when one or more datastores in the datastore cluster exceed the user-configurable I/O latency or space-utilization thresholds. Depending on the selected automation level, these recommendations are automatically applied by vSphere Storage DRS or are presented to the administrator, who then can decide to apply them manually.

vSphere Storage DRS can operate in two distinct load-balancing modes: **No Automation** (manual mode) or **Fully Automated**.

Figure 13. vSphere Storage DRS Automation Level

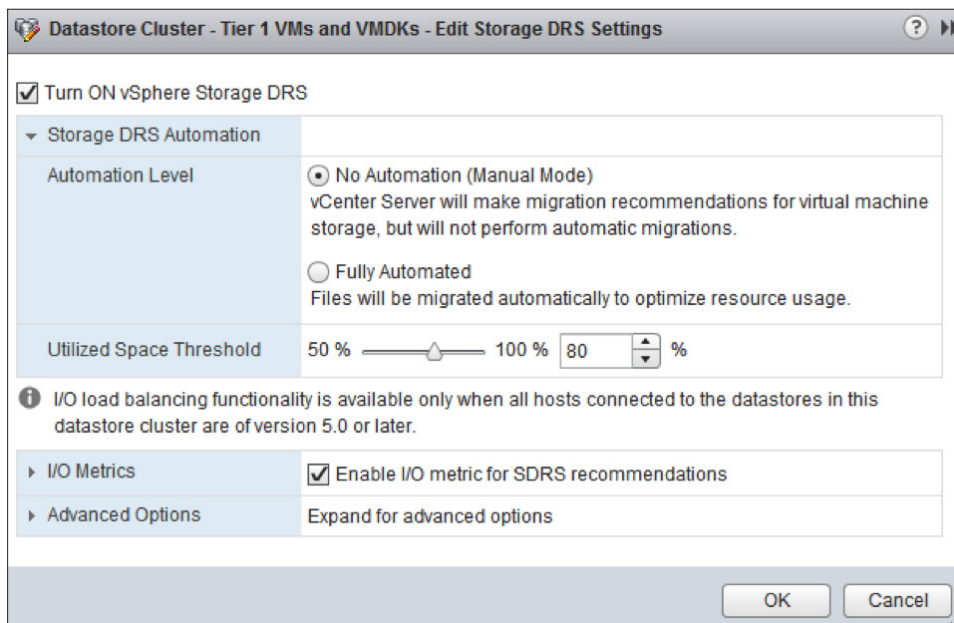When **No Automation** is selected, vCenter Server presents the load-balancing recommendations, which require manual approval for execution. **Fully Automated** mode enables vCenter Server to automatically approve the migration recommendations that vSphere Storage DRS generates. Automation mode applies to both space-utilization load balancing and I/O load balancing.

Load-Balancing Functionality

vSphere Storage DRS load balancing consists of two separately configurable options, space-utilization load balancing and I/O metric load balancing (I/O latency). When activating vSphere Storage DRS on a datastore cluster, both options are enabled. I/O-latency load balancing can be disabled independently of all vSphere Storage DRS features. The datastore cluster properties provide the option to disable I/O load balancing by clearing the **Enable I/O metric for SDRS recommendations** tick box.
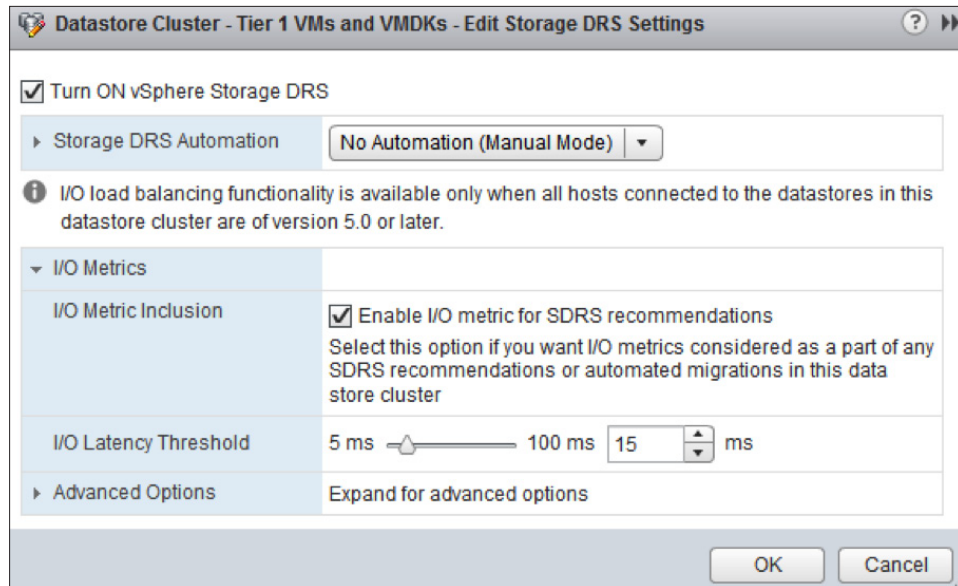
Figure 14. I/O Latency Load Balancing Enabled

Space-utilization load balancing cannot be manually disabled other than by turning off vSphere Storage DRS on the datastore cluster. Each load-balancing algorithm independently generates a candidate migration recommendation that is reviewed, weighted and aggregated before vSphere Storage DRS generates a unified migration recommendation.

Space-Utilization Load Balancing

vSphere Storage DRS load-balances datastores that are violating the space-utilization threshold within a cluster. The default utilized space threshold is 80 percent.

Figure 15. Utilized Space Threshold

The space-utilization load-balancing algorithm comprises three steps before generating a candidate migration recommendation:

1. Determine whether datastores are violating the space-utilization threshold.

2. Analyze datastore usage and the virtual machine disk footprint.

3. Evaluate the minimum space-utilization difference and cost–benefit analysis.
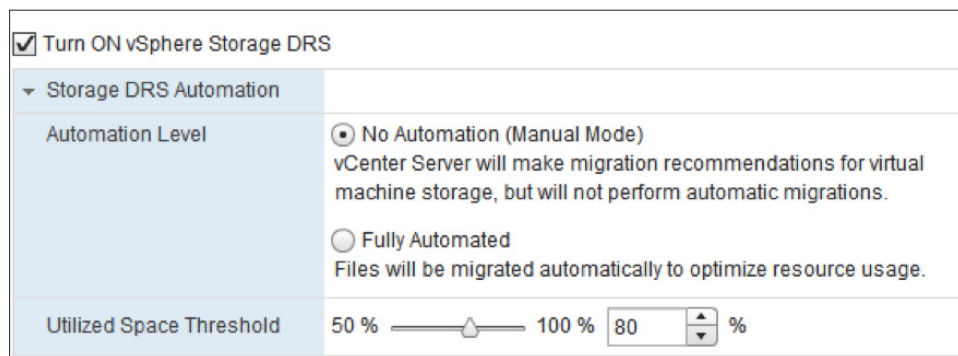
Step 1. Determine Whether Datastores Are Violating the Space-Utilization Threshold
If the space utilization of a datastore exceeds 80 percent, the datastore violates the threshold and the vSphere Storage DRS load-balancing algorithm is invoked. vSphere Storage DRS attempts to avoid an out-of-space situation and therefore runs a load-balancing operation as soon as the datastore exceeds its space-utilization threshold. This operation can be outside of the normal load-balancing interval of every 8 hours.

The space-utilization threshold is a soft limit, enabling vSphere Storage DRS to place virtual machines in the datastore cluster even if all datastores exceed the space-utilization threshold. vSphere Storage DRS attempts to generate prerequisite migrations before virtual machine placement. If this fails, the virtual machine is placed on the datastore that provides the best overall cluster balance.
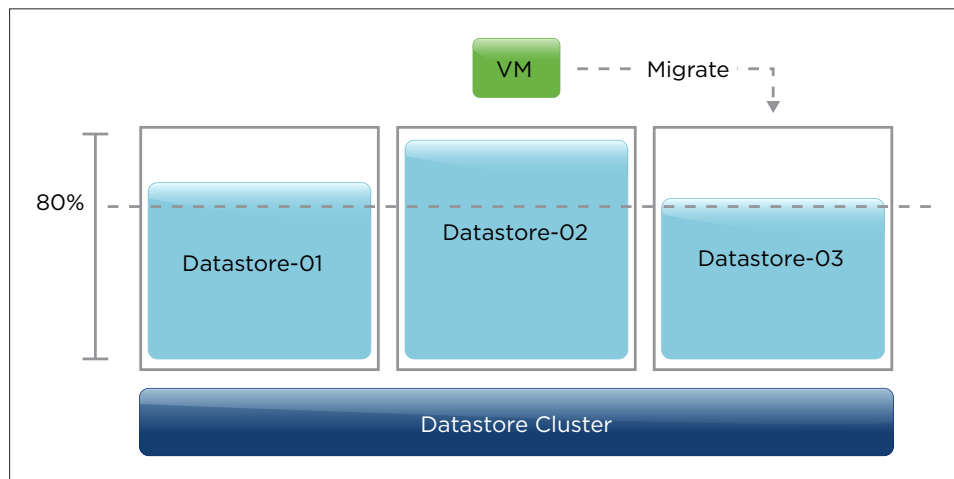
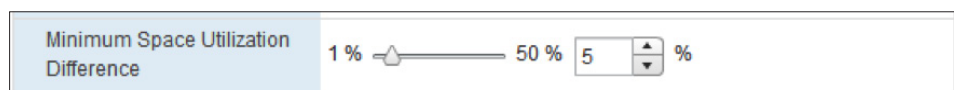**Figure 16.** Space Utilization Balancing in an Overcommitted Datastore Cluster

This performance applies to space-utilization load-balancing operations as well, even if all datastores violate the space-utilization threshold. vSphere Storage DRS tries to keep space utilization near the threshold across all datastores.

Step 2. Analyze Datastore Space Usage and Virtual Machine Disk Footprint
vSphere Storage DRS detects the space usage only of registered virtual machines. However, other files—such as ISOs and user data as well as unregistered virtual machines and vSphere replication replicas—can be stored on the datastores as well. When evaluating datastores that might act as valid destinations, vSphere Storage DRS requests space utilization reports from each one to ensure that there is enough space available.

Step 3. Generate Migration Recommendation Candidates
The **Minimum Space Utilization Difference** setting is accessible in the advanced option section of the datastore cluster. This setting defines the difference in utilization percentage that a destination datastore must be from the source datastore. The default setting is 5 percent. vSphere Storage DRS regards a datastore as a valid destination for load-balancing migrations when the space utilization differs by more than 5 percent from the source datastore. This setting can have a value of between 1 and 50 percent.

Figure 17. Minimum Space Utilization Difference

During the load-balancing operation, vSphere Storage DRS applies a cost–benefit analysis and determines which virtual machine or virtual machine disk files to migrate to correct the space utilization violation. The size of the virtual machine disk file and the activity determine the cost for this migration. Disk activity results in mirrored writes to source and destination during a migration, so vSphere Storage DRS tends to migrate powered-off virtual machines to powered-on virtual machines. When calculating the cost involved in moving a thin-provisioned virtual disk, vSphere Storage DRS uses the allocated space of the thin disk instead of the provisioned space. The benefit is the increase in free space on the source datastore after the virtual machine migration.

Risk is calculated for migration of thin-provisioned virtual disk files. vSphere Storage DRS attempts to avoid migration of thin-provisioned virtual disks that would cause the destination datastore to violate the space utilization threshold in the near future (30 hours). Thin-provisioned disks can experience space growth during the migration. The growth rate is estimated by using and weighing historical data samples; recent samples carry more weight than older samples. If the benefit outweighs the cost and risk, vSphere Storage DRS generates a migration recommendation.

Linked Clones
The space-usage estimation and load-balancing operations of linked clones are somewhat different than for traditional virtual machines. Relocations of linked clones consume varied amounts of space, depending on the availability of the base disk on the destination datastore.

If the base disk used by a virtual machine is available on the destination datastore, vSphere Storage DRS must move only the delta disk of the virtual machine. However, if the virtual machine is moved to a destination datastore that does not store the base disk, both the base disk and the delta disk must be moved.

vSphere Storage DRS tends to move a linked-clone virtual machine to a datastore that contains the base disk, because moving the delta disk not only results in the use of less total space but also minimizes the overhead of the storage migration.

The base disk of a linked clone is retained on a datastore as long as there is at least one linked-clone virtual machine using it.

VASA Provider Thin-Provisioned Pool
VASA is a set of APIs that enables storage arrays to integrate with vCenter for management functionality. vSphere Storage DRS includes information from VASA providers for load-balancing operations. Space-utilization load-balancing operations use the thin-provisioning alarm supplied by the VASA provider, if available. When a thin-provisioned pool on the storage array that furnishes storage capacity to datastores reaches 75 percent utilization, VASA surfaces a **Thin-provisioned volume capacity threshold exceeded** warning to vCenter. vSphere Storage DRS marks the datastores backed by the thin-provisioned pool and does not list them as possible destinations for load-balancing operations.

vSphere Storage DRS does not generate migration recommendations to resolve the fill rate of the thin-provisioned pool.

I/O Load Balancing

The goal of I/O load balancing is to resolve the imbalance of I/O load across datastores in the datastore cluster. By moving virtual machines around inside the datastore cluster, vSphere Storage DRS attempts to enable optimal performance of all virtual machines within the cluster. I/O load balancing is manipulated by two settings: **I/O Latency Threshold** and **I/O Imbalance Threshold**.

I/O Latency Threshold
By default, **I/O Latency Threshold** is set to **15ms**. If a datastore exceeds the threshold for a prolonged period of time, vSphere Storage DRS detects the datastore as being overloaded and migrates virtual machines to other datastores if possible. Unlike space-utilization load balancing, I/O load balancing occurs only during the scheduled invocation period (default every 8 hours).
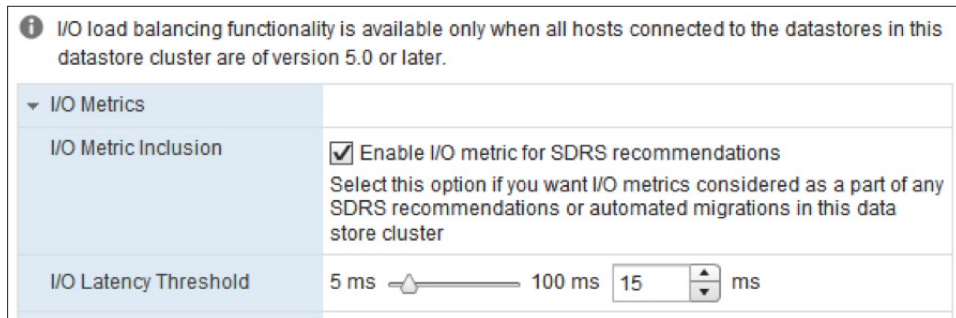
> ⓘ I/O load balancing functionality is available only when all hosts connected to the datastores in this datastore cluster are of version 5.0 or later.

▾ I/O Metrics	
I/O Metric Inclusion	☑ Enable I/O metric for SDRS recommendations Select this option if you want I/O metrics considered as a part of any SDRS recommendations or automated migrations in this data store cluster
I/O Latency Threshold	5 ms ◁────── 100 ms 15 ⬆⬇ ms

Figure 18. I/O Latency Threshold

As with space-utilization load balancing, three steps are taken before generating the migration recommendation:

1. Determine whether datastores are violating the I/O latency threshold.

2. Analyze datastore capabilities and virtual machine workload.

3. Generate migration recommendation candidates.

Step 1. Determine Whether Datastores Are Violating the I/O Latency Threshold
The goal of I/O load balancing is to fix long-term prolonged I/O imbalances, VMware vSphere Storage I/O Control addresses short-term burst and loads. To avoid responding to peak load moments, vSphere Storage DRS collects statistics over a prolonged period of time—at least 16 hours of the current day—to make a proper assessment of the load. It uses the metric VMObservedLatency as input to determine whether the datastore is violating the latency threshold.

VMObservedLatency
VMObservedLatency measures the I/O round trip—from the moment the VMkernel receives the I/O by the virtual machine monitor (VMM), to the datastore, and all the way back to the VMM.
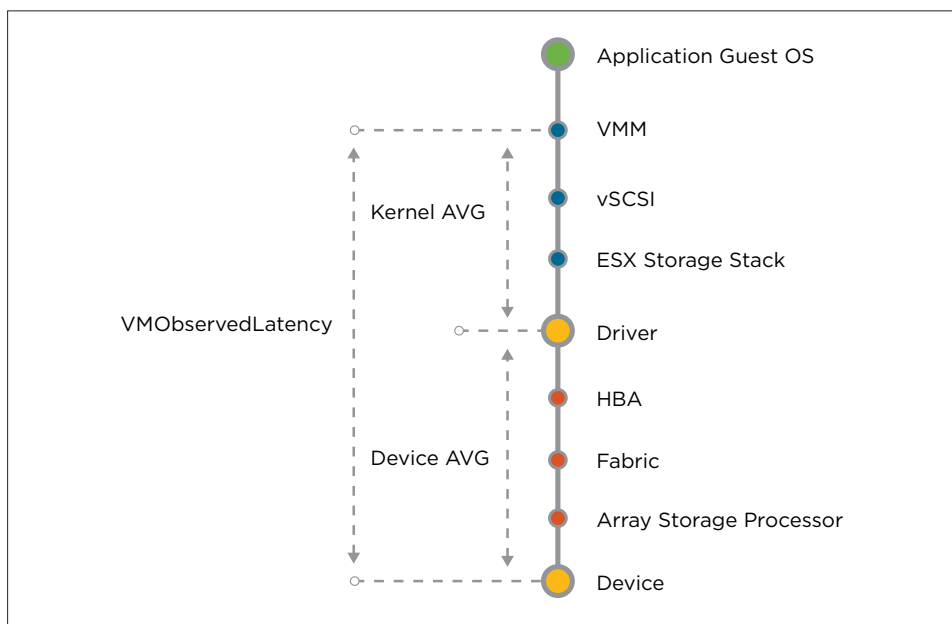

Figure 19. VMObservedLatency

The hosts collect the VMObservedLatency values of the virtual machines and send them periodically to vSphere Storage DRS, which by default stores the statistics as percentile values aggregated over a period of a day. These data points are sorted in ascending order; if the 90th percentile value exceeds the latency threshold, vSphere Storage DRS detects the datastore as overloaded.

The 90th percentile value resembles the lowest edge of the busiest period. Overall, the datastore must experience a sustained load above the threshold for more than 1.6 hours a day. This can be either a workload that is sustained for more than 1.6 hours or one with sustained loads with periodic intervals. vSphere Storage DRS marks a datastore that incurs a threshold violation as a source for migrations.

Step 2. Analyze Datastore Capabilities and Virtual Machine Workload
To resolve an I/O imbalance within a datastore cluster, vSphere Storage DRS requires input for its migration recommendations. This input is generated by a process called online device modeling, which enables vSphere Storage DRS to detect datastore performance and workload modeling. This provides insights as to the virtual machine workload present in the datastore cluster.

Virtual Machine Workload Characterization
Storage workload characterization is an essential first step to providing an intelligent, I/O-aware placement process. For each virtual disk in the datastore cluster, vSphere Storage DRS measures the following metrics:

1. Number of outstanding I/Os

2. I/O size

3. Read/write ratio

4. Randomness percentage

vSphere Storage DRS analyzes the impact of each data point on latency, and it returns an overall workload model metric.

Online Device Modeling
Device-modeling techniques measure the performance of the datastores. Storage arrays expose only a LUN as a logical device. Most of the device-level characteristics such as number of disk spindles servicing the datastore, RPM and average seek delay are hidden from the host and vSphere layer. Without this information, vSphere Storage DRS potentially might migrate a virtual machine from a datastore that is serviced by faster Fibre Channel spindles to a datastore that is serviced by slower SATA spindles.

To assess the performance of each datastore, vSphere Storage DRS leverages vSphere Storage I/O Control, using the workload injector and a reference workload to determine the relative performance capability.

To get a proper performance capability model, vSphere Storage I/O Control injects random read I/O to the disk, using various amounts of outstanding I/O to measure the latency. The duration of the complete cycle is 30 seconds; the procedure is triggered once a day per datastore. Although it's a short-lived process, this workload generates some overhead on the array. vSphere Storage DRS is designed to enable storage performance for your virtual machines, not to interfere with them. Therefore this workload will not run when activity is detected on the devices backing the datastore.

Datastore Correlation Detector
Immediately after the online device–modeling run, vSphere Storage DRS in vSphere 5.1 leverages another vSphere Storage I/O Control module, the datastore correlation detector. When datastores share the same back-end resources, their performance characteristics are tied together. If one datastore is experiencing high levels of latency, the other datastores will experience similar levels of latencies. If datastores are serviced by the same collection of spindles, these datastores are considered "performance related."

vSphere Storage DRS I/O load balancing avoids migrating virtual machines between performance-related datastores because the migration has no effect on the datastore latency.

To determine whether datastores are performance related, vSphere Storage I/O Control runs the datastore correlation detector immediately after the device-modeling operation. After modeling a datastore (for example, datastore A), vSphere Storage I/O Control selects a random datastore (for example, datastore B) to check for correlations. If both datastores are idle, the correlation detector uses the same workload to measure the average I/O latency in the following three simulations:

1. Inject workload on datastore A.

2. Inject workload on datastore B.

3. Inject workload on datastores A and B together.

The first two simulations are to establish a baseline for both datastores. If there is a significant increase in latency detected in the third simulation, both datastores are marked as performance related.

Step 3. Generating Migration Recommendation Candidates

After determining which datastores are overloaded, vSphere Storage DRS reviews the load differences between the datastores in the datastore cluster and compares that result to the I/O imbalance threshold.

The I/O imbalance threshold indicates the aggressiveness of I/O load balancing. The I/O imbalance threshold slider gives the user the ability to change the imbalance tolerance of vSphere Storage DRS. Adjusting the slider to a more conservative setting permits vSphere Storage DRS to tolerate more imbalance across datastores in the datastore cluster. Using an aggressive setting enables vSphere Storage DRS to produce recommendations that correct even small imbalances.
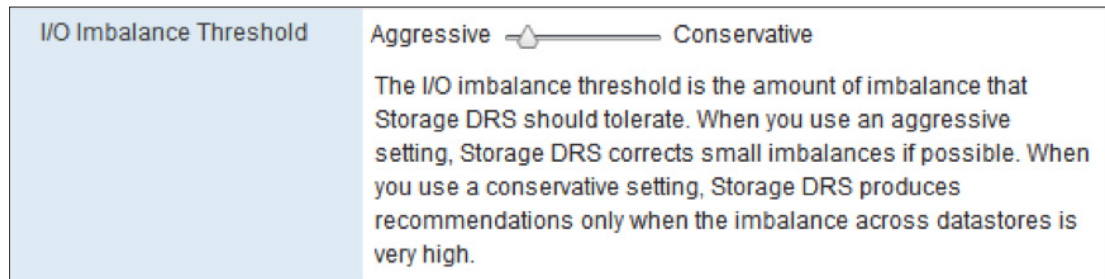
Figure 20. I/O Imbalance Threshold

The cost–benefit analysis is roughly the same as that for the space-utilization load balancer. Cost is related to the duration of the VMware vSphere Storage vMotion® process, which by itself is dependent on the size of the virtual machine disk file and the I/O activity within the virtual machine. Benefit is the estimated improvement of I/O latency that is achieved on the source datastore after the migration. To compute changes in latency as a result of moving I/O load from the source datastore to the destination datastore, vSphere Storage DRS uses the device model.

If after subtracting the cost from the benefit, the benefit is greater than zero, vSphere Storage DRS generates an I/O balance recommendation.

Unified Load-Balancing Recommendation

A migration recommendation should not violate any configured threshold. For example, a recommendation generated to solve a space utilization imbalance should not introduce an I/O load imbalance. To avoid these situations, vSphere Storage DRS calculates space and I/O imbalance for each candidate migration and applies a goodness metric.

Goodness Metric

The goodness metric is a weighted sum of the improvements in the I/O-load and space-utilization load-imbalance metrics. This means that one of the two metrics can become a dominant factor in the generated migration recommendations if a datastore is close to its threshold. For example, if a datastore is close to its space-utilization threshold, space balancing is the dominant factor because it is a more important concern for that datastore. Similarly, the I/O load metric receives more weight than the space metric if the datastore is violating the I/O threshold. Otherwise, both space utilization and I/O load receive an equal weight.

Timescales for vSphere Storage DRS

Unlike vSphere DRS, which operates and generates migration recommendations every 5 minutes, vSphere Storage DRS must operate at much larger timescales. Typically, the duration of a VMDK file migration is longer than the migration of virtual memory, so both the scheduler frequency and load-balance interval take this into account.

Frequency

By default, vSphere Storage DRS load balancing is invoked every 8 hours. The default frequency is 480 minutes (8 hours), but it can be set within a range from 60 minutes to 42,300 minutes (30 days).

Figure 21. Load-Balance Frequency

Interval

A related subject is time required for vSphere Storage DRS to "warm up." This warm-up time is needed for the I/O load balancer to develop an online model, not only for the workloads running on the datastores but also for the capabilities of the disks servicing the datastores. As previously mentioned, if an I/O imbalance is to be detected and acted upon, it must persist for some period of time. Although the default frequency is set to 8 hours, vSphere Storage DRS requires at least 16 hours' worth of data for the current day. This requirement impacts the load-balance interval of vSphere Storage DRS.

The space and I/O load imbalance is checked every 8 hours. For space utilization, vSphere Storage DRS uses the current statistics. For I/O load balancing, it uses previous-day statistics if the day is fewer than 16 hours old. Figure 22 depicts the I/O load-balancing interval.
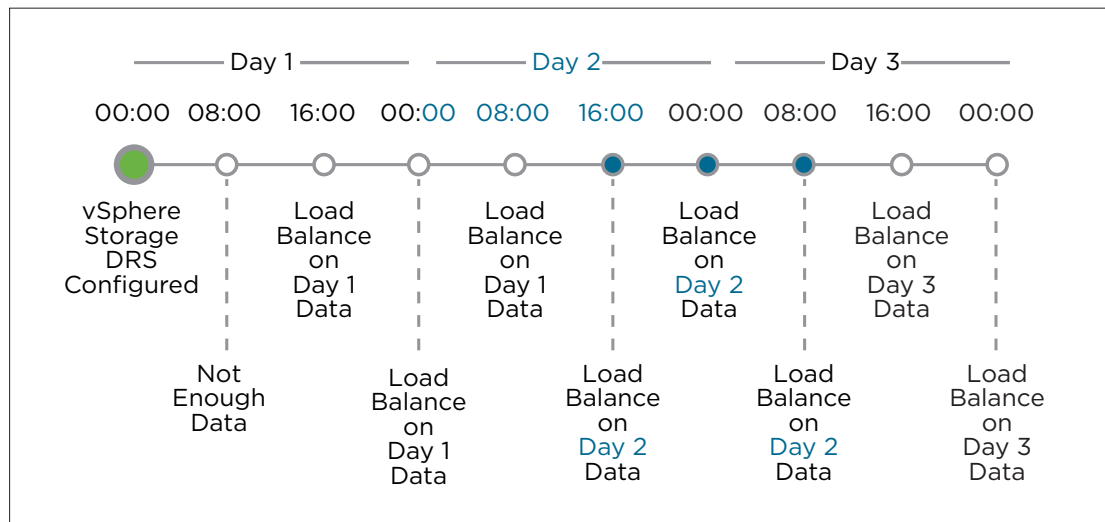

Figure 22. Data Aggregation Mapped to Invocation Period Schedule

In the example depicted by Figure 22, vSphere Storage DRS is enabled at 00:00. At 08:00, vSphere Storage DRS does not initiate I/O load balancing due to absence of a 16-hour dataset. At 16:00, the second invocation, the I/O load-balancing algorithm runs and uses the statistics of the current day. 00:00 marks the beginning of the new day. Collected data points of day 1 are rolled up into a single snapshot and vSphere Storage DRS analyzes the loads of the datastores. At 08:00 on day 2, the previous day's statistics are used because sufficient statistics for the current day are not available.

As a result, vSphere Storage DRS typically recommends I/O load balancing once per day; space balancing is performed every 8 hours and when a datastore is violating the space-utilization threshold. If the user determines that there is an imbalance in the datastore cluster, selecting **Run Storage DRS** generates a load-balancing invocation.

In addition to periodic scheduling, a load-balancing invocation is performed at the following times:

• When the datastore cluster configuration is updated

• When a datastore enters maintenance mode

• When a datastore is introduced in the datastore cluster

• When a datastore exceeds the space-utilization threshold

• When a VMDK file is initially placed

Schedule vSphere Storage DRS
The **Schedule Storage DRS** option enables the user to change the datastore cluster configuration, such as automation mode or I/O latency threshold, for a specific time window. The scheduling option enables the user to define a specific schedule for application of the modified datastore cluster configuration.
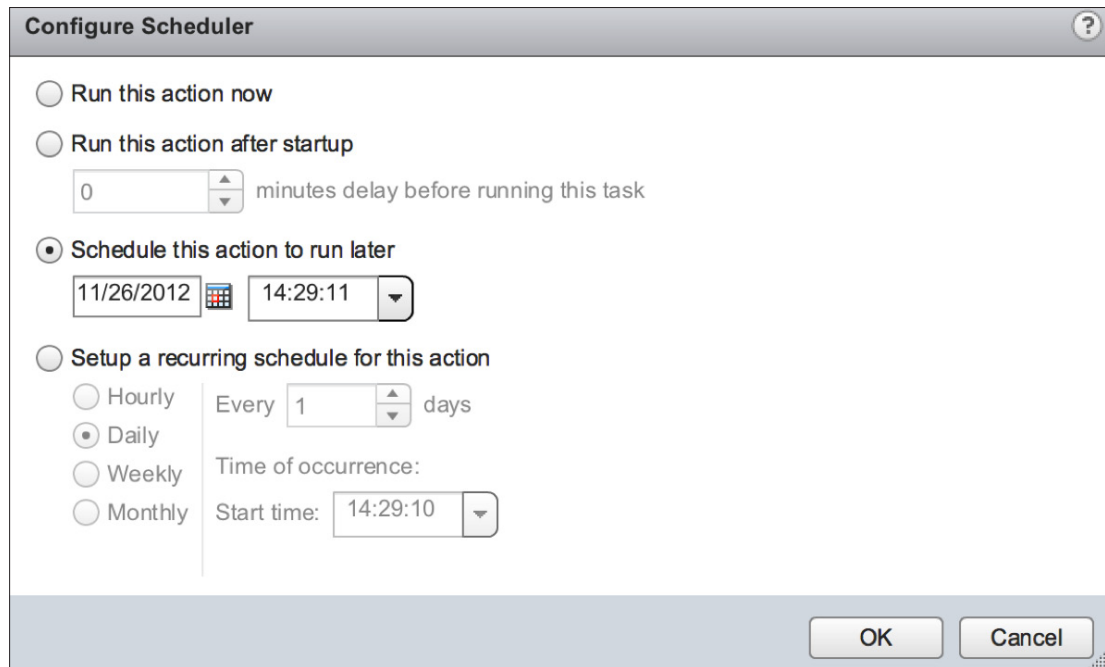

Figure 23. Configure Scheduler

Affinity Rules

Affinity rules in vSphere Storage DRS are conceptually very similar to the affinity and anti-affinity rules that are available in vSphere DRS. They enable control over which virtual disk files should not be placed on the same datastore within a datastore cluster. They also assist the user in keeping the VMDK file configurations in the datastore cluster aligned with the availability requirements. vSphere Storage DRS provides three different affinity rules, as presented in Table 2.

INTRA–VIRTUAL MACHINE LEVEL AFFINITY RULES	
VMDK affinity rule (default)	Forces vSphere Storage DRS to keep the virtual disk files of a virtual machine on a single datastore
VMDK anti-affinity rule	Forces vSphere Storage DRS to keep the virtual disk files of a virtual machine on separate datastores
Virtual machine–virtual machine anti-affinity rule	Forces vSphere Storage DRS to keep the files of the specified virtual machines on separate datastores

Table 2. Affinity Rule Overview

Default Datastore Cluster Affinity Rule

In vSphere 5.1, users can configure the default affinity rule of the datastore cluster via the UI. This feature is available only via VMware vSphere Web Client. VMware vSphere Client™ does not provide this option.

Figure 24. Default Virtual Machine Affinity Rule

By default, vSphere Storage DRS applies an intra–virtual machine VMDK affinity rule, forcing it to place all the files and VMDK files of a virtual machine on a single datastore. By deselecting the **Keep VMDKs together by default** option, the opposite becomes true and an intra–virtual machine anti-affinity rule is applied. This forces vSphere Storage DRS to place the virtual machine files and each VMDK file on a separate datastore.

VMDK Affinity Rule

By default, vSphere Storage DRS applies an intra–virtual machine affinity rule to all new virtual machines in the datastore cluster. The intra–virtual machine affinity rule keeps the virtual machine files—such as VMX file, log files, vSwap and VMDK files—together on one datastore. This facilitates troubleshooting, but vSphere Storage DRS load-balancing algorithms can benefit from distributing the virtual machine across datastores.

DrmDisk

vSphere Storage DRS uses the DrmDisk construct as the smallest entity it can migrate. A DrmDisk represents a consumer of datastore resources. This means that vSphere Storage DRS creates a DrmDisk for each VMDK file belonging to the virtual machine. A soft DrmDisk is created for the working directory containing the configuration files such as the .VMX file and the swap file.

• A separate DrmDisk for each VMDK file

• A soft DrmDisk for system files (VMX, swap, logs, and so on)

If a snapshot is created, both the VMDK file and the snapshot are contained in a single DrmDisk.

VMDK Anti-affinity Rule

When the datastore cluster or the virtual machine is configured with a VMDK-level anti-affinity rule, vSphere Storage DRS must keep the DrmDisk containing the virtual machine disk files on separate datastores.

Impact of VMDK Anti-Affinity Rule on Initial Placement

Initial placement immensely benefits from this increased granularity. Instead of searching a suitable datastore that can fit the virtual machine as a whole, vSphere Storage DRS can seek appropriate datastores for each DrmDisk file separately. Due to the increased granularity, datastore cluster fragmentation—described in the "Initial Placement" section—is less likely to occur; if prerequisite migrations are required, far fewer are expected.

Impact of VMDK Anti-Affinity Rule on Load Balancing

Similar to initial placement, I/O load balancing also benefits from the deeper level of detail. vSphere Storage DRS can find a better fit for each workload generated by each VMDK file.

vSphere Storage DRS analyzes the workload and generates a workload model for each DrmDisk. It then determines in which datastore it must place the DrmDisk to keep the load balanced within the datastore cluster while offering sufficient performance for each DrmDisk. This becomes considerably more difficult when vSphere Storage DRS must keep all the VMDK files together. Usually in that scenario, the datastore chosen is the one that provides the best performance for the most demanding workload and is able to store all the VMDK files and system files.

By enabling vSphere Storage DRS to load-balance on a more granular level, each DrmDisk of a virtual machine is placed to suit its I/O latency needs as well as its space requirements.

Virtual Machine–to–Virtual Machine Anti-Affinity Rule

An inter–virtual machine (virtual machine–to–virtual machine) anti-affinity rule forces vSphere Storage DRS to keep the virtual machines on separate datastores. This rule effectively extends the availability requirements from hosts to datastores. In vSphere DRS, an anti-affinity rule is created to force two virtual machines—for example, Microsoft Active Directory servers—to run on separate hosts; in vSphere Storage DRS, a virtual machine–to–virtual machine anti-affinity rule guarantees that the Active Directory virtual machines are not stored on the same datastore. There is a requirement that both virtual machines participating in the virtual machine–to–virtual machine anti-affinity rule be configured with an intra–virtual machine VMDK affinity rule.

Datastore Maintenance Mode

vSphere Storage DRS introduced the datastore maintenance mode feature, which can be compared to host maintenance mode. When a datastore is placed in maintenance mode, all registered virtual machines on that datastore are migrated to the other datastores in the datastore cluster. Unregistered virtual machines, templates and files such as ISOs are not migrated automatically by vSphere Storage DRS.

Automation Mode

Depending on the automation mode, migration recommendations generated by datastore maintenance mode are automatically accepted by vCenter (fully automated mode) or are presented to the user to validate (no automation mode).

Use Case 1: VMFS-3–to–VMFS-5 Upgrade

Datastore maintenance mode is an excellent feature that can help you migrate virtual machines from VMFS-3 datastores to VMFS-5 datastores aggregated in a vSphere Storage DRS datastore cluster. Place the VMFS-3 datastore in a datastore cluster and enter datastore maintenance mode. vSphere Storage DRS makes migration recommendations based on space and I/O utilization in the datastore (if I/O load balancing is enabled). By accepting the migration recommendations, vSphere Storage DRS smoothly migrates all the virtual machines off the datastore. After completion, reformat the datastore with VMFS-5 and exit datastore maintenance mode. Repeat the procedure on all remaining VMFS-3 datastores.

Use Case 2: Storage Array Migration
Datastore maintenance mode can be used to safely migrate virtual machines out of the datastore for storage array–related maintenance operations such as migrating a LUN to another RAID group or migrating virtual machines from one storage array to another array. Instead of migrating all the virtual machines manually, move the datastore serviced by spindles of the old array into the datastore cluster containing datastores located on the new array. Place the "old" datastore containing the virtual machines into maintenance mode. Enable vSphere Storage DRS to find the virtual machines suitable datastores serviced by spindles on the new array. Repeat until all datastores on the old array are empty.

Conclusion

The goal of vSphere Storage DRS is to help decrease operational effort associated with the provisioning of virtual machines and monitoring of the storage environment. vSphere Storage DRS assists in—or automates—several management tasks such as initial placement of virtual disks, out-of-space avoidance and runtime load balancing of space consumption and I/O load on datastores. It also provides a simple interface to specify business constraints using affinity and anti-affinity rules, and it enforces them while making provisioning determinations.

VMware considers vSphere Storage DRS to be the beginning of the journey to software-managed storage in virtualized datacenters and is planning to accommodate the newer storage architectures and storage disk types in the future to make vSphere Storage DRS even more widely applicable.

Acknowledgments

I would like to thank Mustafa Uysal and Sachin Manpathak of the VMware Cloud Resource Management group for contributing to this white paper, as well as Duncan Epping, Cormac Hogan and James Senicka of VMware for their technical review.

Reference Material

The following VMware knowledge base articles and documents have been used in the development of this case study:

vSphere 5.1 Resource Management

VMware vSphere Storage DRS Interoperability

About the Author

Frank Denneman is senior architect in the Technical Marketing group at VMware. His focus is on resource management, with a specialization in vSphere DRS, vSphere Storage DRS, vSphere Storage I/O Control, VMware vSphere vMotion® and VMware vSphere Storage vMotion. He is the co-author of several books, including best sellers *vSphere 5.x Clustering Deepdive* and *VMware vSphere 4.1 HA and DRS Technical Deepdive*. Frank is the owner and main author of leading virtualization and VMware blog frankdenneman.nl.

• Follow Frank's blogs at http://www.frankdenneman.nl and http://blogs.vmware.com/vSphere.

• Follow Frank on Twitter: @frankdenneman.

vmware®