HP 3PAR Storage and VMware vSphere 5 best practices

Table of contents

Executive summary 3

Configuration 4
- Fibre Channel 4
- Multi-pathing considerations 6
- Allocating storage 8
- Virtual volumes 8
- Boot from SAN 9

Overview and configuration of VMware vSphere Storage API Integration 9
- VAAI 9
- VASA 16
- Configuring HP VMware vCenter Server integration 17

Thin provisioning 18
- Configuring 18
- 3PAR Thin Provisioning vs. vSphere Thin Provisioning 21
- Thin Conversion 22
- Thin Persistence 22
- Zero Detect 23

HP 3PAR Adaptive Optimization 23
- Manual and auto optimization on HP 3PAR Storage 23
- Implementing Adaptive Optimization 24
- Best practices 24

Performance tuning 29
- I/O sizing 29
- SPC-1 benchmark results 30
- Alignment considerations 30
- Virtual SCSI adapters and virtual disk types 30
- Wide striping 32
- Storage I/O Control 32

HP 3PAR Adaptive queue depth throttling 34
Recovery Manager 37
Overview 37
Architecture 37
Benefits 38
Usage and best practices 39
HP 3PAR Integration with VMware vCenter Site Recovery Manager (SRM) 40
Overview 40
Architecture 40
Best practices and caveats 41
Summary 43
For more information 44

Executive summary

When supported with the correct underlying storage platform, server virtualization delivers greater consolidation, administrative efficiency, business continuity and cost savings. As a result, server virtualization is not only transforming the data center, but also the businesses that those data centers fuel. However, these transformative results depend on enterprise storage to deliver the performance, availability, and flexibility to keep up with the dynamic and consolidated nature of virtualized server environments.

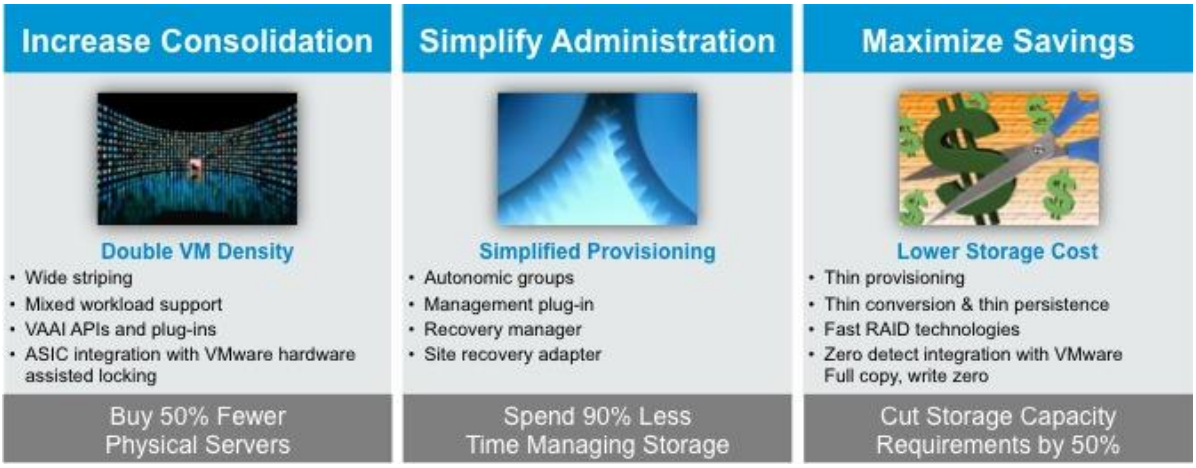
HP 3PAR Storage is the next generation of federated Tier 1 storage and was built from the ground up to exceed the economic and operational requirements of virtual data centers and cloud computing environments by providing the SAN performance, scalability, availability and simplified management that clients need. It does this through an innovative system architecture that offers storage federation, secure multi-tenancy, built-in thin processing capabilities, and autonomic management and storage tiering features that are unique in the industry.

When deployed together, VMware vSphere and HP 3PAR Storage deliver a compelling virtual data center solution that increases overall resource utilization, provisioning agility, application availability, administrative efficiency, and reduces both capital and operating costs.

Implementing HP 3PAR Storage systems with VMware vSphere 5 enables its users the unique ability to:

- Increase consolidation savings by doubling virtual machine density
- Maximize savings through lower storage costs by up to 50%
- Simplify storage provisioning and management time by up to 90%

Figure 1. HP 3PAR Storage for VMware vSphere Environments



These benefits in VMware environments are delivered through a combination of HP 3PAR advanced features and integration with VMware storage technologies (Figure 1).

Increase Consolidation: Integrating HP 3PAR Storage systems with VMware vSphere 5 enables its users to double virtual machine density on physical servers through 3PAR wide striping, mesh-active clustered architecture, mixed workload support, and HP 3PAR hardware assisted support of VMware vSphere Storage APIs for Array Integration (VAAI).

Simplify Administration: Managing storage in VMware environments is simplified through unique 3PAR capabilities such as Autonomic Groups, Recovery Manager for VMware software, and integrated management through VMware vCenter Server.

Maximize Savings: HP 3PAR thin technologies including Thin Provisioning, Thin Conversion, and Thin Persistence and in-line zero detect capability of the 3PAR ASIC deliver the ultimate in storage efficiency in VMware environments.

HP 3PAR arrays continue to capture the SPC-1 benchmark results due to the 3PAR architecture with multiple storage controller nodes and wide striping over available disk drives. Capturing SPC-1 is crucial to VMware's demand for low latency I/O response and HP 3PAR's ability to double VM density with fewer servers compared to other arrays on the market today.

This white paper outlines best practices on how to set up HP 3PAR Storage with VMware vSphere 5 as well as how to take advantage of HP 3PAR's unique features such as vSphere integration, HP 3PAR Thin Provisioning technologies, Dynamic and Adaptive Optimization, and Recovery Manager for VMware to create a world class virtualized IT and application infrastructure. The information contained in this document should be used along with the documentation set provided by HP for the HP 3PAR Storage system, HP 3PAR Storage operating system (InForm), and the documentation provided by VMware for vCenter, Site Recovery Manager (SRM), and other related products.

Target audience: IT Administrators and Solution Architects planning to leverage HP 3PAR Storage within a VMware vSphere 5 environment.

This white paper is based on testing performed in August 2012.

Configuration

Fibre Channel

Target port limits and specifications

To ensure an optimal configuration, observe the following limitations on ESXi host server HBA ports and HP 3PAR Storage target ports:

- Maximum of 16 host server ports per 2 Gb HP 3PAR Storage port
- Maximum of 32 host server ports per 4 or 8 Gb HP 3PAR Storage port
- Maximum total of 1,024 host server ports per HP 3PAR Storage system

The I/O queue depth for each HP 3PAR Storage system HBA model is shown in Table 1. Note that the I/O queues are shared among the connected host server HBA ports on a first-come, first-served basis. For recommendations on managing I/O queues for optimal performance, please see the Storage I/O Control and HP 3PAR Adaptive queue depth throttling sections in this document.

Table 1. I/O queue depth for HP 3PAR Storage HBAs

HP 3PAR Storage HBA model	I/O queue depth
QLogic 2Gb	497
LSI 2Gb	510
Emulex 4Gb	959
HP 3PAR HBA 4Gb	1638
HP 3PAR HBA 8Gb (V400/V800 systems only)	3276

Zoning recommendations

Fabric zoning controls which Fibre Channel end-devices have access to each other on the SAN fabric. You can set up fabric zoning by associating the device World Wide Names (WWNs) or the switch ports with specified zones in the fabric. Although you can use either the WWN method or the port zoning method with HP 3PAR Storage, the WWN zoning method is recommended because the zone survives the changes of switch ports when cables are moved around on a fabric.

Employ fabric zoning, using the methods provided by the switch vendor, to create relationships between host server HBA ports and storage server ports before connecting the host server HBA ports or HP 3PAR Storage system ports to the fabric(s).

HP 3PAR Storage arrays support the following zoning configurations:

- One initiator to one target per zone (Figure 2).

- One initiator to multiple targets per zone (zoning by HBA). This zoning configuration is recommended for HP 3PAR Storage. Zoning by HBA is required for coexistence with other HP Storage arrays.

Figure 3 shows an example where each host HBA port is zoned to two target ports on the 3PAR array. Note that this provides failover and protection against any single failure in the fabric or array. In addition, it provides the following benefits:

- If one controller node goes offline, both host HBA ports still have one path to the array.
- If the host HBA ports have higher speeds than the array HBA and the array is lightly loaded, this configuration may be able to sustain higher data rates to the array.

Figure 2. One initiator (host HBA port) to one HP 3PAR target port (one-to-one zoning)

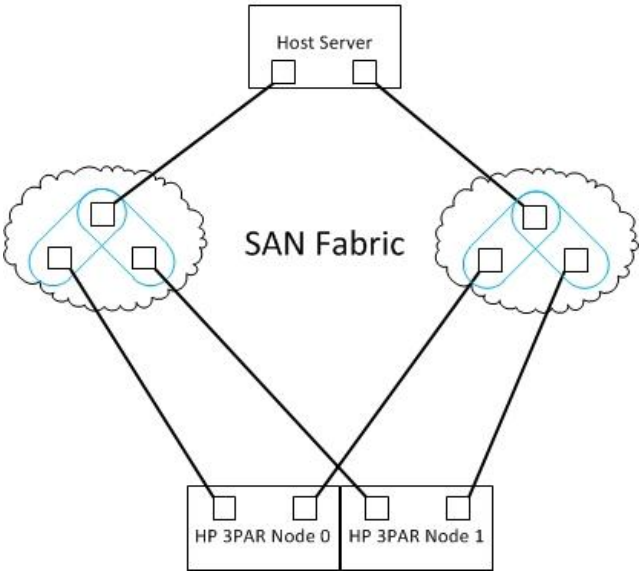
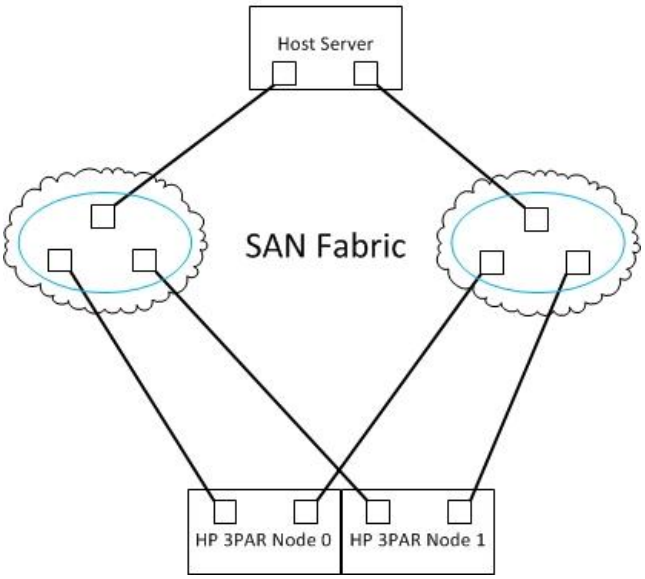


Figure 3. One initiator (host HBA port) to two 3PAR target ports (zoning by HBA).



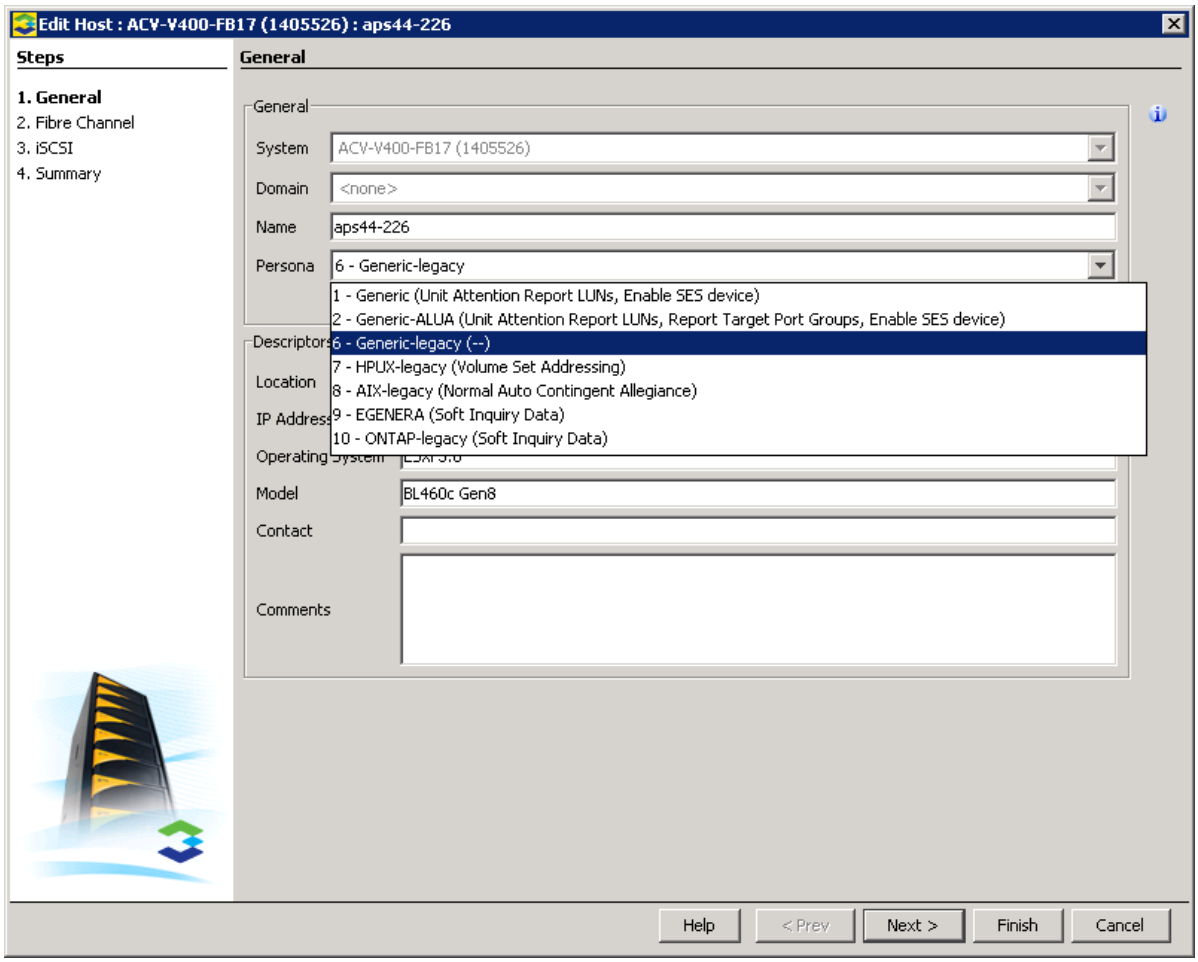
Note that the storage targets in the zone can be from the same HP 3PAR Storage system, multiple HP 3PAR Storage systems, or a mixture of HP 3PAR and other HP storage systems.

After configuring zoning and connecting each host server HBA port and HP 3PAR Storage port to the fabric(s), verify the switch and zone configurations using the InForm OS CLI showhost command, to ensure that each initiator is zoned with the correct target(s).

Creating the host definition (host persona)

Before exporting VLUNs from HP 3PAR Storage to the ESXi host, you need to create a host definition that specifies a valid host persona for each host that is to be connected to the HP 3PAR Storage. ESXi uses the generic legacy host personality of 6 for both the QLogic and Emulex HBAs. The host persona may be specified when creating the host definition, or the host definition may be edited as shown in Figure 4.

Figure 4. Set host persona to 6



Alternatively, the createhost or sethost command may be used to set or change the persona from the 3PAR InForm CLI:

```
# createhost –persona 6 <hostname> [WWN...]

# sethost –persona 6 <hostname>
```

Multi-pathing considerations

By default, the QLogic and Emulex drivers for the VMware ESXi host support failover. To take advantage of failover support, virtual volumes should be exported to multiple paths to the host server. To do this, create a host definition on

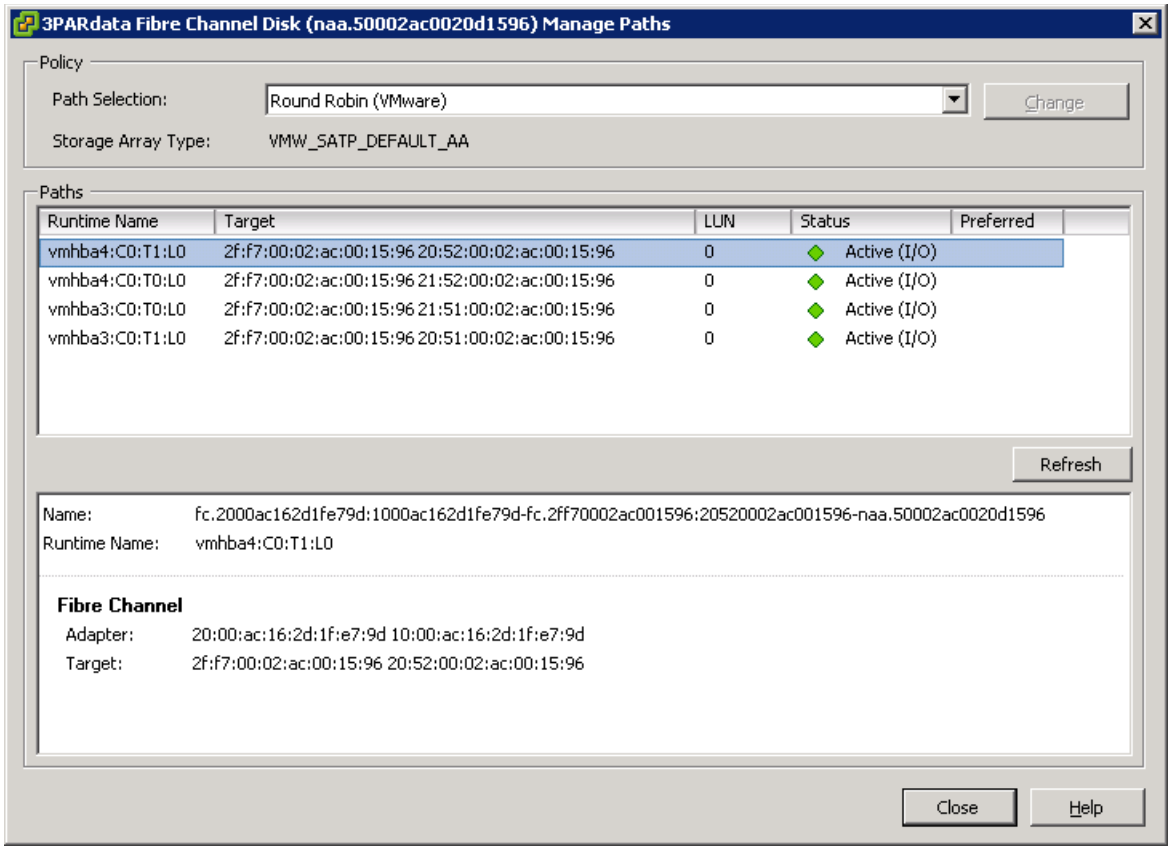
the HP 3PAR Storage system that includes the World Wide Names (WWNs) of multiple HBA ports on the host server and then export the VLUNs to that host definition. For an ESXi cluster, the VLUNs must be exported to all of the host definitions for the cluster nodes, or a host set may be created containing all of the servers and the VLUNs can be exported to the host set.

VMware ESXi 5 includes active/active multipath support to maintain a constant connection between the ESXi host and the HP 3PAR Storage array. Three path policies are available, “Fixed”, “Most Recently Used” and “Round Robin”. For ESXi 5, Round Robin is the recommended policy; however, it is not enabled by default. The default I/O path policy is Fixed. The path policies can be modified from the vSphere Client on a per datastore basis as follows:

1. In the vSphere Client inventory panel, select the datastore.
2. Click the **Configuration** tab, select an ESXi host, and click **Properties** for the datastore.
3. In the pop-up window, click the **Manage Paths** button.
4. Select **Round Robin** from the Path Selection drop-down menu.
5. Click the **Change** button to save the new setting.
6. Repeat steps 2 through 5 for each ESXi host.

Figure 5 shows an example of a 3PAR Fibre Channel VLUN that has the Round Robin path policy. Note that the status for all four paths to the LUN is “Active (I/O)”.

Figure 5. LUN set to Round Robin path policy



Note that in Figure 5, the Storage Array Type is VMW_SATP_DEFAULT_AA. VMware native multipathing has two key plugins: a storage array type plugin (SATP) that determines how path failover is handled, and a path-selection plugin (PSP) that decides which path is used to issue an I/O request to a storage device. The default SATP, VMW_SATP_DEFAULT_AA, is mapped to the default PSP, VMW_PSP_FIXED, which is the Fixed path policy. SATP rules can be edited or created using esxcli commands on the ESXi host to automatically achieve a round robin path policy for newly discovered LUNs.

Use the following command to create a custom SATP rule that will allow the ESXi host to configure the HP 3PAR LUNs to use round robin multipath policy. The command must be executed on each ESXi host that is connected to the HP 3PAR array.

```
# esxcli storage nmp satp rule add -s "VMW_SATP_DEFAULT_AA" -P "VMW_PSP_RR" –O "iops=1000" -c
"tpgs_off" -V "3PARdata" -M "VV" -e "HP 3PAR Custom Rule”
```

Verify the new rule using the following command:

```
# esxcli storage nmp satp rule list | grep "3PARdata"
```

The new rule will be effective when adding new devices to the ESXi host, but not for existing devices. Existing devices continue to use VMW_SATP_DEFAULT_AA with VMW_PSP_FIXED by default. For existing LUNs, either a host reboot is required, or the path policy must be set for each LUN.

Allocating storage

In general, a single or few large LUNs should be configured as a VMFS datastore for use by multiple VMs. A larger LUN gives VMware vSphere administrators the most flexibility by not requiring storage administrators to provision new storage every time a new VM is created. A larger LUN also minimizes rescans for new LUNs on the ESXi host, which can impact VMFS I/O. LUNs presented to ESXi hosts should be striped across many physical drives leveraging HP 3PAR’s parallel architecture to maximize performance and virtual machine density.

With its massively parallel architecture, HP 3PAR Storage uses all available drives of the same type behind its virtual volumes, which are presented as LUNs to the ESXi hosts. Virtual machine workloads are automatically spread across all internal resources (nodes, drives, ports, etc.), delivering high application service levels and predictable levels of performance (high IOPS and low latencies). With Wide striping and autonomic distribution of workloads, storage administrators need not spend time figuring out what drives are available in an array or how best to lay out a LUN. Instead, creating a virtual volume on HP 3PAR Storage takes less than a minute and is a matter of a few clicks with the HP 3PAR InForm Management Console or a simple command from the command line interface.

With traditional arrays, storage administrators generally create LUNs in the 500 GB to 2 TB range for VMFS volumes, with a 1 TB LUN being a very common size. This seems to be the acceptable tradeoff between creating a large enough LUN to satisfy the performance requirements of VMware vSphere without wasting too much unused space in the LUN. However, HP 3PAR Thin Provisioning software eliminates this tradeoff altogether.

With HP 3PAR Thin Provisioning software, one can allocate as much logical capacity to a VMFS volume as needed over the lifetime of that volume without actually dedicating any physical capacity up front. Physical capacity is allocated seamlessly on an as needed basis and is consumed only when vSphere writes to the thin provisioned virtual volumes. There is no wasted capacity in the LUN that is presented to the ESXi host. Further details may be found in the Thin provisioning section of this document.

Virtual volumes

Before creating virtual volumes, you must first create Common Provisioning Groups (CPGs) to allocate space for the virtual volumes. A CPG creates a virtual pool of logical disks that allows up to 4,095 virtual volumes to share the CPG's resources and allocates space on demand. The RAID type, space allocation, growth increments and other logical disk parameters can be set when you create a CPG or modified later. You can create fully-provisioned virtual volumes and Thinly-Provisioned Virtual Volumes (TPVVs) that draw space from the CPG's logical disk pool. CPGs enable fine-grained, shared access to pooled logical capacity. Instead of pre-dedicating logical disks to volumes, the CPG allows multiple volumes to share the buffer pool of logical disks. By default, a CPG is configured to auto-grow new logical disks when the amount of available logical disk space falls below a configured threshold. The initial buffer pool of logical disks starts off at a fraction of the exported virtual capacity of mapped volumes and automatically grows over time as required by application writes.

Volumes draw their resources from Common Provisioning Groups (CPGs), and volumes are exported as Logical Unit Numbers (LUNs) to hosts. Virtual volumes are the only data layer visible to hosts. You can create physical copies or virtual copy snapshots of virtual volumes for use if the original base volume becomes unavailable.

For information about configuring Thin-Provisioned Virtual Volumes, see the Thin provisioning section of this document.

Boot from SAN

In a boot from SAN environment, each ESXi 5 host's operating system is installed on the HP 3PAR Storage system, rather than on the host's local disk. It is required that a separate virtual volume be created for each ESXi 5 host's boot image to use boot from SAN.

In order to properly configure and allocate the ESXi 5 host to use an HP 3PAR virtual volume for its boot volume, the following steps must be taken:

1. Perform the required and recommended zoning as outlined in the Zoning recommendations section.
2. Create a virtual volume and export it as a VLUN to the ESXi 5 host.
3. Disable the HP Smart Array or other local boot device if installed in the server's BIOS.
4. Boot the HBA BIOS by selecting Alt-Q or Ctrl-Q for QLogic HBAs or Alt-E for Emulex HBAs during POST.
    a. Reset all of the HBA NVRAM settings to their default values.
    b. Exit.
5. Reboot the server once POST is complete and reenter the HBA BIOS as done in step 4.
    a. Enable the HBA port to boot.
    b. Discover the LUN and designate it as bootable through the HBA BIOS.
    c. Save and Exit.
6. Start installation of VMware ESXi 5.
    a. During the installation, you will be provided an option to select the HP 3PAR Storage LUN presented to the host.

For more information, refer to the HP Boot from SAN Configuration Guide:
http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c01861120/c01861120.pdf

For more information on booting ESXi from Fibre Channel SAN, refer to the ESXi and vCenter Server 5 Documentation's Booting ESXi from Fibre Channel SAN: http://pubs.vmware.com/vsphere-50/topic/com.vmware.vsphere.storage.doc_50/GUID-9004389B-E2C0-4BE5-811C-E4886E3B7450.html

Overview and configuration of VMware vSphere Storage API Integration

HP and VMware deliver advanced integration between HP 3PAR Storage and VMware vSphere 5 with vSphere Storage APIs for Storage Awareness (VASA) and vSphere Storage APIs for Array Integration (VAAI). This section will provide an overview of the APIs as well as how to configure and leverage them.

VAAI

The vSphere Storage APIs are a set of technologies and interfaces that enable vSphere to leverage storage resources to deliver the efficiency, control, and ease of customization that clients demand of their virtual environment. The vSphere Storage APIs for Array Integration (VAAI) is one of these technologies. Under the VAAI initiative, APIs have been introduced to improve performance, resource utilization, and scalability by leveraging more efficient array-based operations.

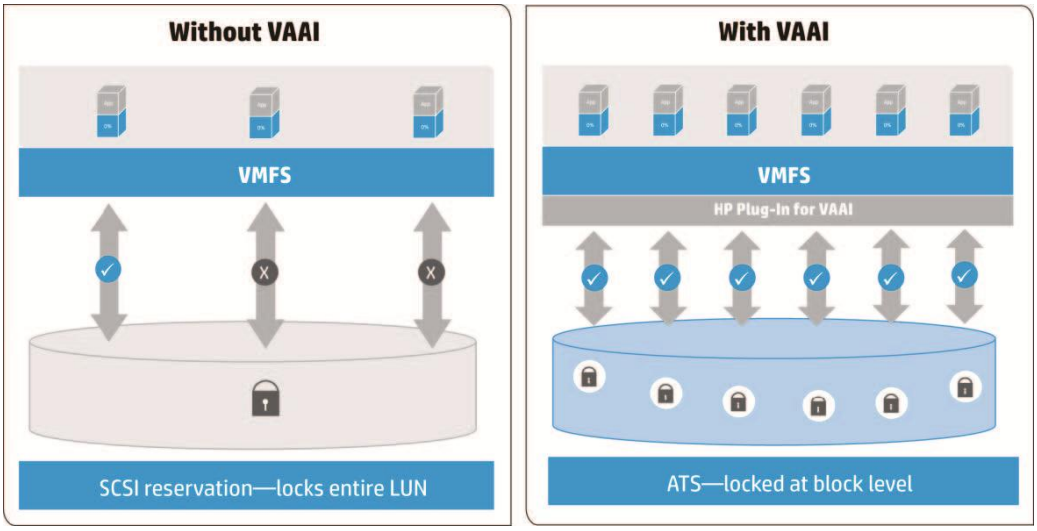
HP developed the HP 3PAR Management Software Plug-In for VMware VAAI to deliver enhanced performance, agility, and scalability using vSphere commands first introduced in vSphere 4.1. Initial support of the SCSI (T10) standard was introduced in vSphere 4.1 through the use of block level commands. These standard commands were enabled by a standard VMware plug-in which enabled Hardware Assisted Locking, Fast Copy, and Block Zeroing.

vSphere 5 provides enhanced support for the T10 standards without the need to install a plug-in, enabling vSphere to directly utilize more advanced features of the storage array. HP 3PAR InForm OS version 2.3.1 does not offer native T10 support and will still require the installation of the 3PAR VAAI plug-in. However, InForm OS version 3.1.1 and later includes native T10 support and no longer requires the plug-in to take advantage of the enhanced VAAI functionality within vSphere 5.

Some of the important hardware commands that VAAI enables are documented below:

Hardware Assisted Locking eliminates SCSI reservation contention by providing a fast, fine-grained locking mechanism. The ATS ("Atomic Test and Set") command verifies that a block of metadata is what is expected (test) and then replaces it with an updated block (set) in a single, atomic operation. Using this command, the ESXi host can lock a portion of a LUN related to a single VM instead of locking the whole LUN as seen in Figure 6, thereby allowing other VMs on the same LUN to continue operating normally. The implementation of ATS on HP 3PAR Storage arrays uses the HP 3PAR ASIC to further improve performance. The combination of ATS and the HP 3PAR ASIC allows an increase in VM density per LUN and greater scalability for vSphere deployments.

Figure 6. Diagram representing old method before VAAI on the left and hardware-assisted locking method available with VAAI on the right.



Fast Copy uses the XCOPY command to improve the performance of common storage operations like VM cloning and Storage vMotion by performing large data movement operations directly within the storage array. By not requiring each block to make a round-trip to the host, the time required for these operations is significantly reduced and storage network traffic minimized. When combined with HP 3PAR Thin Persistence Software, drive I/O and storage capacity can also be reduced since blocks of zeros are not written due to the array's Zero Detect capability, which is integrated into the HP 3PAR ASIC.

Block Zeroing uses the standard SCSI command WRITE_SAME to offload large, block-level write operations of zeros from the host to the storage array. Block zeroing improves host performance and efficiency when allocating or extending Eager Zeroed Thick (EZT) virtual disks, or on initial access to a block on a non-EZT virtual disk. When combined with built-in Zero Detect and EZT virtual disks, storage array bandwidth, disk I/O bandwidth, and disk consumption is minimized. Initializing EZT virtual disks in seconds rather than minutes eliminates the tradeoff between fast VM creation and fast run-time performance.

Note

Due to performance issues, the new automatic space reclamation feature (UNMAP) in vSphere 5 has been disabled by VMware in ESXi 5.0 Patch 02 and ESXi 5.0 U1. The recommendation is to manually initiate a reclaim of space from a thin provisioned volume using the command:

```
vmkfstools -y <n>% <datastore>.
```

For more information, see
Reclaiming VMFS deleted blocks on HP 3PAR thin-provisioned virtual volumes

Or refer to VMware Knowledge Base entries
http://kb.vmware.com/kb/2014849 and http://kb.vmware.com/kb/2007427

Checking to see if VAAI is enabled

As stated previously, HP 3PAR OS version 3.1.1 and later include native SCSI T10 support and no longer require the VAAI plug-in to take advantage of the enhanced VAAI functionality within vSphere 5. VAAI is enabled by default in VMware vSphere 5 and an ESXi 5 host will automatically try and use VAAI primitives to see if a LUN supports them. If the use of the VAAI primitives fail, the LUN will be marked as non-VAAI capable. An ESXi host will then automatically try again every 16,384 I/Os to see if a LUN supports VAAI in order to automatically detect if VAAI has been enabled on the LUN, and thus start to use the VAAI primitives if enabled. To see if VAAI is enabled using the VMware Client:

1. In the vSphere Client inventory panel, select the host.
2. Click the Configuration tab, and click Advanced Settings under Software.
3. Check that these options are set to **1** (enabled):

   DataMover.HardwareAcceleratedMove

   DataMover.HardwareAcceleratedInit

   VMFS3.HardwareAcceleratedLocking

You should see something similar to Figure 7 and Figure 8.

Figure 7. DataMover Settings set to 1 (enabled)

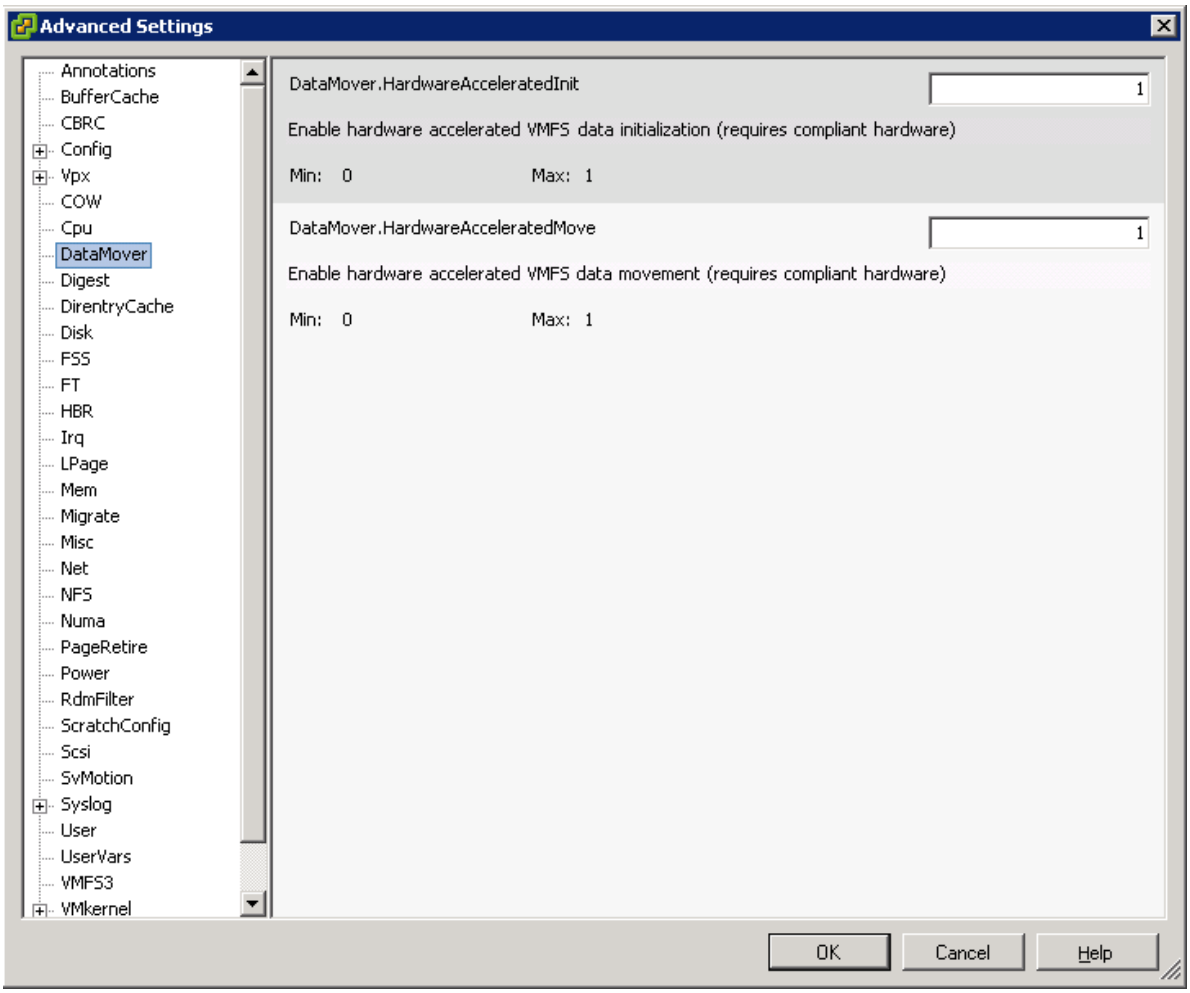
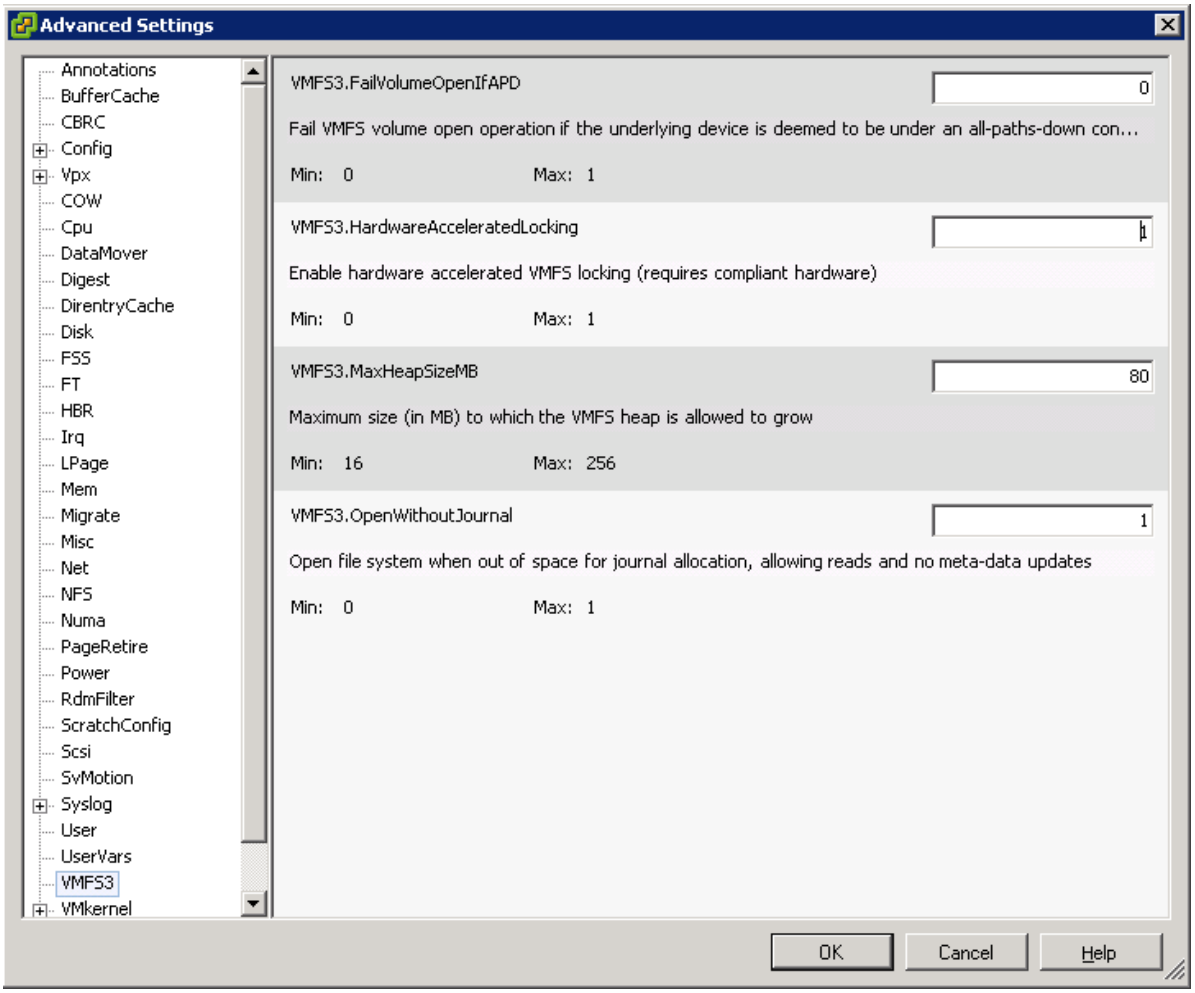


Figure 8. VMFS3.HardwareAcceleratedLocking setting set to 1 (enabled)



You can also check on the ESXi 5 host by running the following CLI commands and ensuring that the **Int Value** is set to **1**.

```
# esxcli system settings advanced list -o /DataMover/HardwareAcceleratedMove
# esxcli system settings advanced list -o /DataMover/HardwareAcceleratedInit
# esxcli system settings advanced list -o /VMFS3/HardwareAcceleratedLocking
```

You will see an output similar to below for each command:

Path: /VMFS3/HardwareAcceleratedLocking

Type: integer

Int Value: **1** ← **set to 1 if enabled**

Default Int Value: 1

Min Value: 0

Max Value: 1

String Value:

Default String Value:

Valid Characters:

Description: Enable hardware accelerated VMFS locking (requires compliant hardware)

Enabling/Disabling VAAI on a ESXi 5 host

If you need to enable, or for some reason disable, VAAI, you can perform the following steps by setting the variables to **1** for enable or **0** for disable. This is useful to see the performance benefits first hand of storage operations when executed with VAAI enabled and disabled.

1. In the vSphere Client inventory panel, select the host.
2. Click the Configuration tab, and click Advanced Settings under Software.
3. Set the options to the appropriate values. See Figure 7 and Figure 8 to see an example of a VAAI enabled configuration

   DataMover.HardwareAcceleratedMove

   DataMover.HardwareAcceleratedInit

   VMFS3.HardwareAcceleratedLocking

Alternatively, to set these values from the command line on the ESXi 5 hosts, run the following commands to enable/disable VAAI by setting the variable "--int-value" to **1** for enable or **0** for disable:

```
# esxcli system settings advanced set --int-value 1 --option /DataMover/HardwareAcceleratedMove

# esxcli system settings advanced set --int-value 1 --option /DataMover/HardwareAcceleratedInit

# esxcli system settings advanced set --int-value 1 --option /VMFS3/HardwareAcceleratedLocking
```

Whether this is done from the vSphere Client or the ESXi 5 host, the changes will be active after running the above commands and a reboot of the host is not required.

Note

VAAI should be either enabled or disabled on all ESXi hosts connected to the storage. Having ESXi hosts not all in the same state is not supported.

Checking to see if the storage array supports VAAI

To check if the disk, in this case the 3PAR disk, fully supports VAAI, run the following command. An option having the value of **supported** means that that piece of the VAAI functionality is enabled. You can optionally provide the **–d <naa>** parameter to display only one device.

```
# esxcli storage core device vaai status get
```

You will see an output similar to below for the command:

```
naa.50002ac0020d1596      ← This is the 3PAR disk
  VAAI Plugin Name:
  ATS Status: supported
  Clone Status: supported
  Zero Status: supported
  Delete Status: supported
naa.600508b1001cb5424a979d461c291bfa      ← This is the local disk
  VAAI Plugin Name:
  ATS Status: unsupported
  Clone Status: unsupported
  Zero Status: supported
  Delete Status: unsupported
```

You can also run the following command to see more information on all your storage devices. You can optionally provide the **–d=*naa*** parameter to display only one device.

```
# esxcli storage core device list -d=naa.50002ac0020d1596
naa.50002ac0020d1596
  Display Name: 3PARdata Fibre Channel Disk (naa.50002ac0020d1596)
  Has Settable Display Name: true
  Size: 307200
  Device Type: Direct-Access
  Multipath Plugin: NMP
  Devfs Path: /vmfs/devices/disks/naa.50002ac0020d1596
  Vendor: 3PARdata
  Model: VV
  Revision: 3111
  SCSI Level: 5
  Is Pseudo: false
  Status: on
  Is RDM Capable: true
  Is Local: false
  Is Removable: false
  Is SSD: false
  Is Offline: false
  Is Perennially Reserved: false
  Thin Provisioning Status: yes
  Attached Filters:
  VAAI Status: supported
  Other UIDs: vml.020000000050002ac0020d1596565620202020
```

Hardware acceleration plug-ins and filter

With other storage arrays, including HP 3PAR Storage before the release of HP 3PAR OS version 3.1.1, that do not natively support the T10 SCSI standard, a VAAI plug-in is needed to use the VAAI capabilities VMware offers. To manage a VAAI capable device, your host attaches the VAAI filter and vendor-specific VAAI plug-in to the device. Because of the native T10 support built into HP 3PAR OS 3.1.1 and greater, this is not needed and as a result your ESXi host can communicate directly to the HP 3PAR Storage and does not require the VAAI plug-ins.

Reclaiming VMFS deleted blocks on HP 3PAR thin-provisioned virtual volumes

Leveraging the VAAI Thin Provisioning Block Space Reclamation (UNMAP) Primitive, it is possible to reclaim deleted space to meet continuing storage needs. HP 3PAR Storage can leverage the UNMAP commands sent by an ESXi 5 host for space reclamation. VAAI needs to be enabled on the host and **Delete Status** for the device has to have the value of **supported.**

```
# esxcli storage core device vaai status get -d naa.50002ac0020d1596
naa.50002ac0020d1596
  VAAI Plugin Name:
  ATS Status: supported
  Clone Status: supported
```

Zero Status: supported

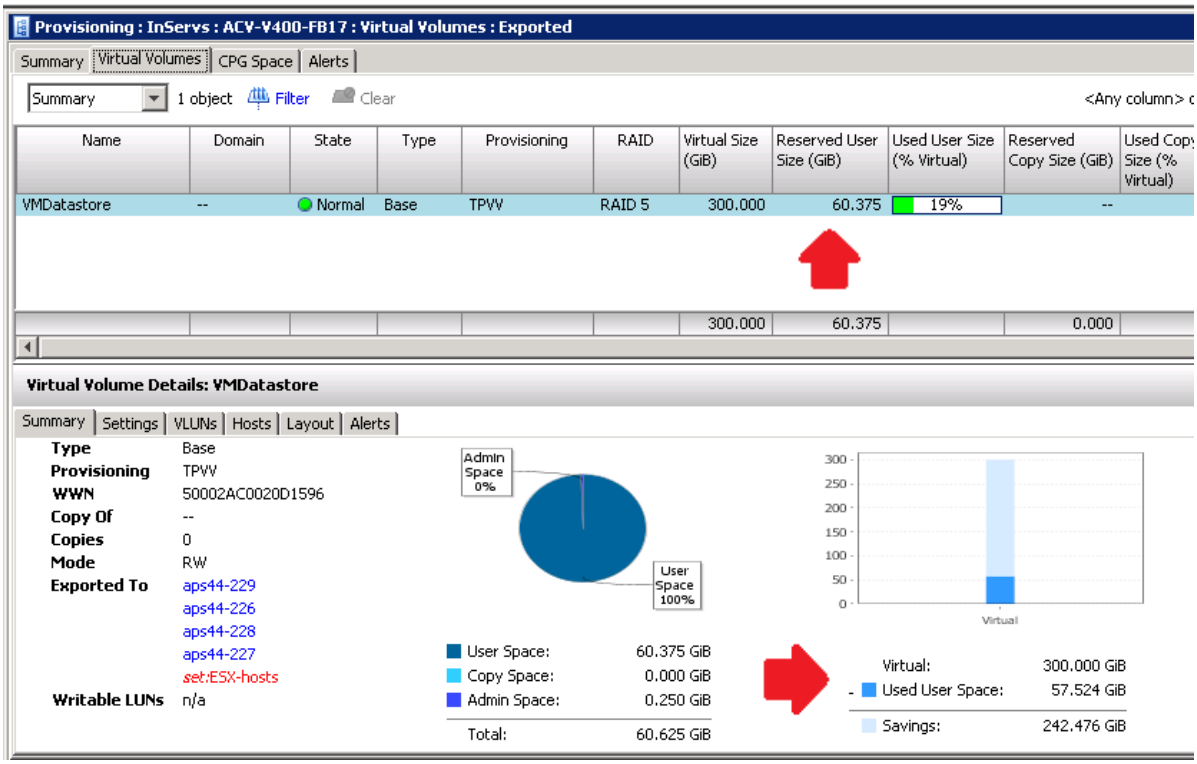
Delete Status: supported

Note
VMware recommends disabling automatic space reclamation using UNMAP on ESXi 5 hosts with thin-provisioned LUNs. ESXi500-201112001 and ESXi 5.0 Update 1 disable issuing UNMAP commands by default. If you continue to use an un-patched ESXi 5 host, you need to manually disable UNMAP on all hosts. For instructions on disabling UNMAP, refer to VMware Knowledge Base entry http://kb.vmware.com/kb/2007427.

ESXi 5.0 Update 1 includes an updated version of vmkfstools that provides an option (-y) to send the UNMAP command to the storage arrays regardless of the ESXi host's global setting.

You can see the space used on the HP 3PAR Storage in the Virtual Volume (VV) VMDatastore by looking at Figure 9.

Figure 9. HP 3PAR Virtual Volume Allocation

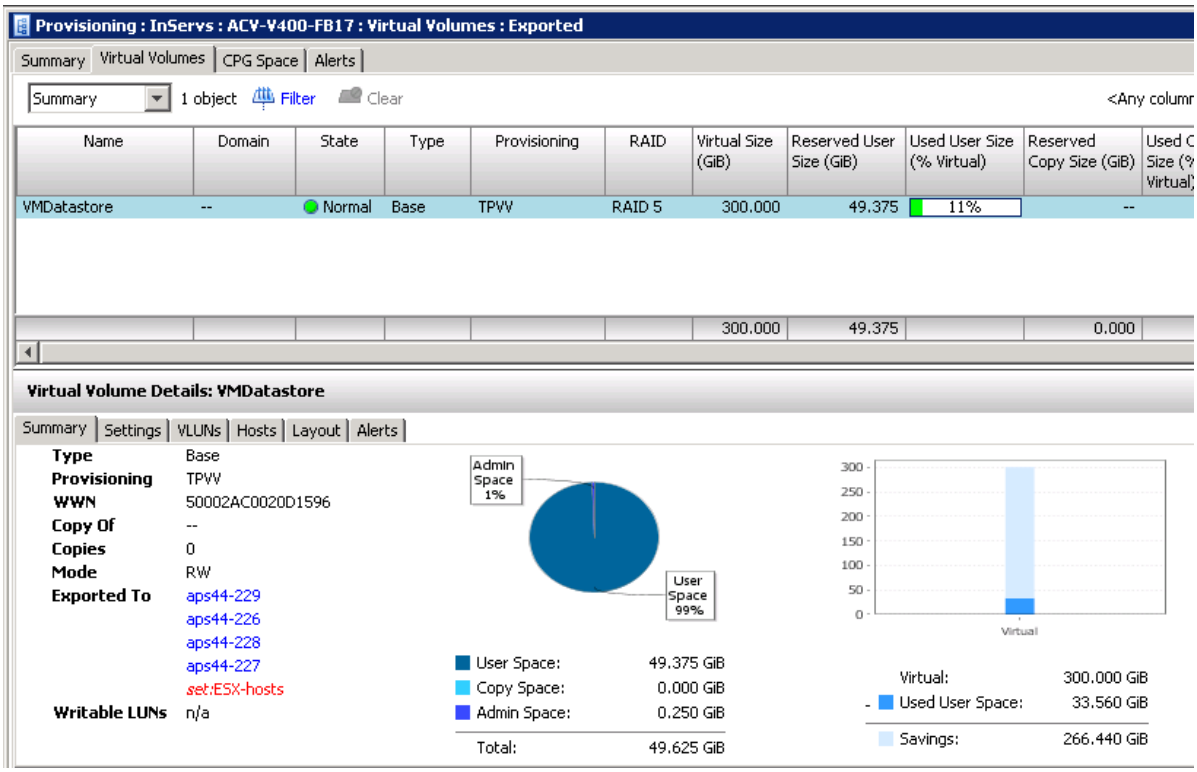


To reclaim space, on an ESXi 5 host, change directory to the datastore you want to reclaim space from, and run the vmkfstools command similar to below to reclaim a defined percentage of the free space. It is recommended by VMware to use a number of 60 to ensure that the operation is done in a time efficient manner. In this case 90 was used to demonstrate the capabilities of the command.

 # cd /vmfs/volumes/<volume-name>
 # vmkfstools -y 90
Attempting to reclaim 90% of free capacity 173.7 GB (156.3 GB) on VMFS-5 file system 'VM Datastore' with max file size 64 TB.
Create file .vmfsBalloon3YB3Mx of size 156.3 GB to reclaim free blocks.
Done.

Upon completion of the command, the space in the example above on the HP 3PAR devices was reduced from 19% to 11%, or 11GB (Used User Size). You can see the reduced size of the VMDatastore VV in Figure 10.

Figure 10. HP 3PAR Virtual Volume after UNMAP operation



As the operation is time consuming, consider running it during a maintenance window as the high I/O generated by the SCSI UNMAP may impact storage performance on the array thus impacting running virtual machines.

VASA

The vSphere Storage APIs for Storage Awareness (VASA) is a set of APIs introduced with vSphere 5 that enables VMware vCenter Server to detect the capabilities of the storage array LUNs and their datastores. This visibility into the array's configuration of its datastores and their capabilities, simplifies vSphere administration with HP 3PAR Storage. Capabilities such as RAID level, thin or thick provisioned, device type (SSD, Fibre Channel, or Nearline) and replication state can now be made visible from within vCenter Server's disk management interface. This allows vSphere administrators to select the appropriate disk for virtual machine placement based on its needs. VASA eliminates the need for maintaining complex spreadsheets detailing the storage capabilities of each LUN previously required to guarantee the correct Service Level Agreement (SLA).

The concept of a storage profile, introduced in vSphere 5, extends the base VASA functionality. These profiles are used in conjunction with the capabilities of the LUN to determine which LUNs meet the needs of a VM. vSphere 5 can use this information to migrate virtual machines between LUNs for load balancing using Storage Distributed Resource Scheduler (DRS) while maintaining the storage performance and availability needs (RAID level, etc.) of the virtual machine. These profiles also allow vSphere to make placement decisions automatically based on the needs of the VM and the available datastores, further reducing the administration impact.

Beginning with vSphere 5, the HP 3PAR Management Software Plug-In for VMware vCenter also includes a vSphere Storage APIs for Storage Awareness (VASA) plug-in which allows vSphere to display detailed (device type, RAID level, etc.) information on the 3PAR LUNs directly from with the VMware vCenter Client. You can read more about the HP 3PAR Management Software Plug-In for VMware vCenterwhich leverages VMware's VASA capabilities in the HP 3PAR Management Software Plug-In for VMware vCentersection below.

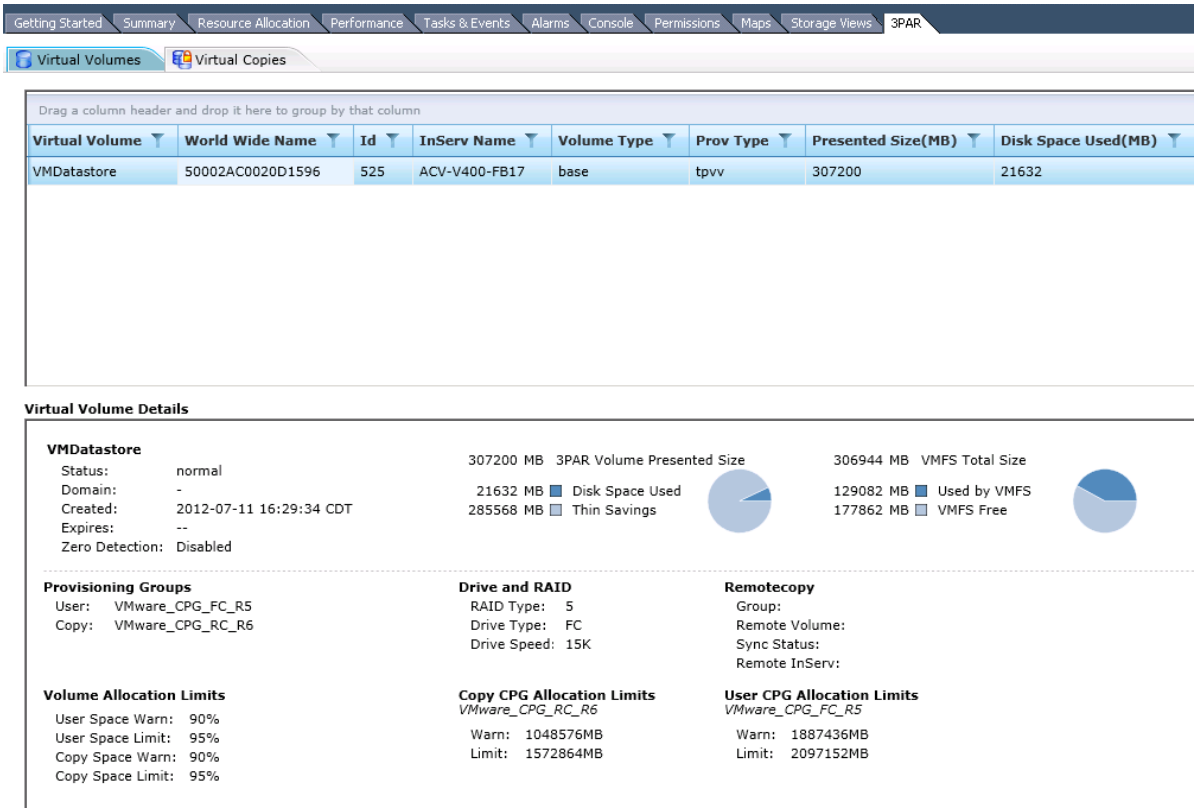
Configuring HP VMware vCenter Server integration

HP 3PAR Management Software Plug-In for VMware vCenter

The HP 3PAR Management Software Plug-In for VMware vCenter is a plug-in that allows easy identification of HP 3PAR virtual volumes used by VMs and datastores. It provides an integrated view of the VMs and associated storage resources. Properties such as volume type (Thick or Thin Provisioned Virtual Volume (TPVV)), device type (FC disk, NL disk, or SSD), RAID level, etc. are displayed via the 3PAR tab in the vSphere Client.

The HP 3PAR Management Software Plug-In for VMware vCenter provides the VMware administrator a view into the HP 3PAR Storage via a single pane (Figure 11). There is no need to login to the storage array to identify space consumption or determine how a volume maps to a datastore. This information is easily visible via the plug-in, as is capacity usage, allocation limits, and other data. This information can then be used to define storage profiles ensuring the storage meets the needs of the virtual machine.

Figure 11. Viewing virtual volume mapping information via the HP 3PAR plug-in



HP 3PAR Recovery Manager Software for VMware vSphere

HP 3PAR Recovery Manager Software for VMware vSphere enables the protection and rapid recovery of VMs and datastores. It provides virtual copy management and allows the administrator to take LUN-level snapshots of VMs and datastores via the vSphere management console. HP 3PAR Recovery Manager Software provides array-based snapshots that are fast, space-efficient, and VM-aware. This plug-in solves the issues associated with traditional, agent-based backup schemes that are typically slow, complex, and fail to offer the flexibility and granularity that array-based snapshots can provide.

HP 3PAR Recovery Manager Software makes it possible to create hundreds of virtual copies. The number of virtual copies to retain and the retention period for each virtual copy can easily be specified. Once a virtual copy has been created, this plug-in allows the flexibility of granular restores at the VMFS layer, the VM layer, or the individual file level.

For more detailed information on the HP 3PAR Management Software Plug-In for VMware vCenter see http://h18006.www1.hp.com/storage/software/3par/mpvs/index.html and for HP 3PAR Recovery Manager Software for VMware vSphere, see http://h18006.www1.hp.com/storage/software/3par/rms-vsphere/index.html.

HP 3PAR Replication Adapter Software for VMware vCenter SRM

VMware vCenter Site Recovery Manager (SRM) provides end-to-end management of array-based replication, virtual machine failover, and automated disaster recovery management for environments that use VMware vCenter Server. HP 3PAR Replication Adapter Software for VMware vCenter SRM was developed to provide integration between VMware vCenter Site Recovery Manager (SRM) and HP 3PAR Remote Copy Software.

HP 3PAR Storage Replication Adapter (SRA) Software for VMware vCenter Site Recovery Manager (SRM) integrates VMware SRM with HP 3PAR Storage and replication software to provide a complete and integrated Business Continuity solution. The solution offers centralized management of recovery plans, non-disruptive testing, and automated site recovery, failback and migration processes. The HP 3PAR SRA software combines HP 3PAR Remote Copy Software and HP 3PAR Virtual Copy Software with VMware SRM to ensure the highest performing and most reliable disaster protection for all virtualized applications.

HP 3PAR Dynamic Optimization Software and VMware vSphere

HP 3PAR Dynamic Optimization Software is a software product that offers an online and non-disruptive way to make changes to volumes on the HP 3PAR Storage System. Storage administrators can move volumes between different drive types (Fibre Channel, Nearline, SSD), convert between RAID levels (RAID 1, RAID 5, or RAID 6/RAID MP), and/or rebalance volumes as new drives are added, all without impacting any hosts that the system is busy serving.

In VMware vSphere environments, HP 3PAR Dynamic Optimization Software can be used to move running VMs between different drive tiers without impacting what the VMs are doing. Similarly, as new drives are added to the array, the LUN that the ESXi host is using can be striped across the new drives on the fly without creating an outage at the ESXi host level.

VMware vMotion offers somewhat similar functionality, but at the hosting layer. HP 3PAR Dynamic Optimization Software works at the storage layer, bypassing any application-level overheads. Therefore, it can be used to optimize data service levels while VMware vMotion can be used to optimize CPU utilization across multiple hosts.

Thin provisioning

HP 3PAR Thin Provisioning allows for creating Thinly-Provisioned Virtual Volumes (TPVVs) as an alternative to fully-provisioned volumes. A TPVV uses logical disks (LD) that belong to a logical disk pool, or the Common Provisioning Group (CPG).

All TPVVs associated with the same CPG draw user space from a shared LD pool as needed and are allocated space on demand in one chunklet increments, either 256 MB for HP 3PAR F-Class and T-Class, or 1 GB for HP 3PAR P10000, per controller node. As the volumes that draw space from the CPG require additional storage, the system automatically creates additional logical disks and adds them to the pool until the CPG reaches the user-defined growth limit that restricts the CPG's maximum size. The maximum TPVV volume size limit is 16 TB. These allocations are adaptive since subsequent allocations are based on the rate of consumption for previously allocated space. For example, if a TPVV is initially allocated 256 MB per node but then consumes that space in less than sixty seconds, the next allocation becomes 512 MB per node. However, if the initial 256 MB per node is consumed more slowly, the next allocation increment remains at 256 MB per node. Under this provisioning scheme, the maximum allocation increment is 1 GB per controller node supporting the TPVV.

As the TPVV reaches either its exported size or its user-defined allocation limit, the system allows for allocation of an additional 128 MB per node beyond these limits in order to ensure that the exported TPVV address space is usable. With VMware ESXi 5 and HP 3PAR OS 3.1.1 MU2, HP 3PAR Thin Provisioning simplifies management by allowing creation of large VMFS datastores without impacting VM performance, while also increasing ROI by not having to pay for more storage than actually used. For more information on performance enhancements with HP 3PAR Inform OS 3.1.1, refer to the VAAI section in this document.

Configuring

As previously discussed, when implementing HP 3PAR Thin Provisioning, LDs are added on demand to a CPG allowing storage allocation as required to all the TPVVs associated with the CPG. CPGs provide the dynamic framework to balance storage utilization and performance with Thin Provisioning. However, it is imperative that growth increments be adaptive to the CPG characteristics and monitored to meet workload demand. The following points should be considered when defining CPG growth limits.

- Define and use different CPGs for differing service qualities (e.g., RAID 1 or RAID 5 VVs; Nearline, Fibre Channel and SSD).
- Define and use different CPGs for VVs restricted to specific controller nodes within the HP 3PAR storage, (e.g., VVs associated with nodes 0 and 1 and VVs associated with nodes 2 and 3).
- Define and use different CPGs for VVs for which it is desirable to use different disks to minimize interaction with each other for performance or availability reasons (e.g., to separate databases and their transaction logs).
- Define, set, and monitor over time CPG growth warnings and growth limits to restrict the CPG's growth and maximum size.

Note
Proper planning and monitoring is required when using TPVV as it is possible to set the growth warnings or limits to exceed the amount of available storage on a system. If a TPVV associated with a CPG consumes all available system space, new writes will fail and/or snapshot volumes associated with the CPG may become invalid or stale. Under these conditions, some host applications will not handle write failures gracefully and may produce unexpected failures.

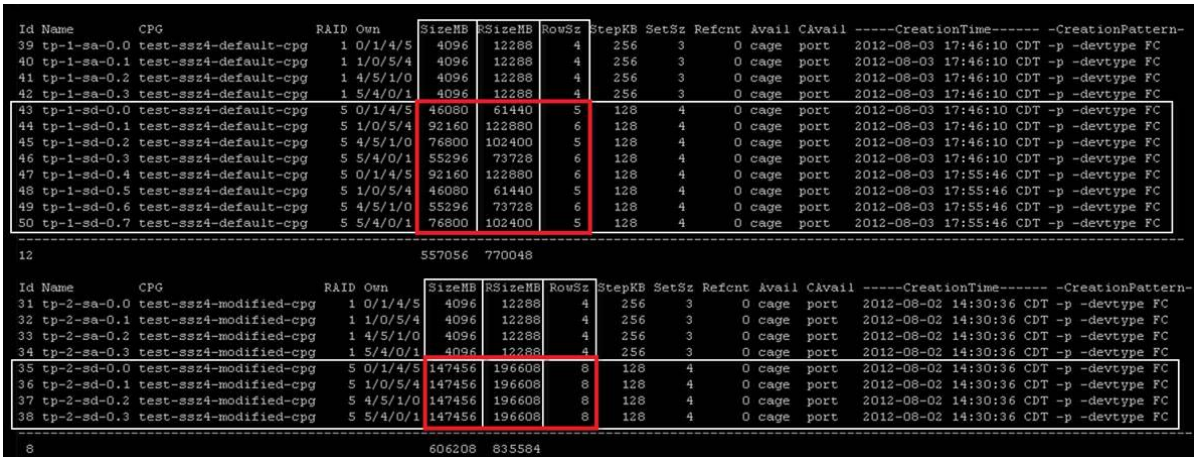
When creating a CPG, whether using thin or thick provisioning, the following two criteria should always be observed:

- The CPG growth increment should be as small as possible to minimize the amount of growth space.
- For optimal performance, the CPG growth increment needs to be tuned to evenly distribute LDs across all 3PAR Storage System Physical Disks (PDs) per CPG policy.

For a better understanding of CPG growth increments and the benefits of properly sizing them, take note of Figure 12 and note the following:

1. The default CPG has 8 LDs compared to 4 for the tuned CPG.
2. The default CPG's LDs are unevenly allocated to PDs (sizeMB) when compared with the tuned CPG's evenly sized LDs.
3. The default CPG row size (RowSz) is smaller/not balanced when compared to the tuned CPG; keeping in mind larger row sizes are a best practice in themselves.

Figure 12. Default vs. Tuned LD distribution on disk



Id	Name	CPG	RAID	Own	SizeMB	RSizeMB	RowSz	StepKB	SetSz	Refcnt	Avail	CAvail	-----CreationTime------	-CreationPattern-
39	tp-1-sa-0.0	test-ssz4-default-cpg	1	0/1/4/5	4096	12288	4	256	3	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
40	tp-1-sa-0.1	test-ssz4-default-cpg	1	1/0/5/4	4096	12288	4	256	3	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
41	tp-1-sa-0.2	test-ssz4-default-cpg	1	4/5/1/0	4096	12288	4	256	3	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
42	tp-1-sa-0.3	test-ssz4-default-cpg	1	5/4/0/1	4096	12288	4	256	3	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
43	tp-1-sd-0.0	test-ssz4-default-cpg	5	0/1/4/5	46080	61440	5	128	4	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
44	tp-1-sd-0.1	test-ssz4-default-cpg	5	1/0/5/4	92160	122880	6	128	4	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
45	tp-1-sd-0.2	test-ssz4-default-cpg	5	4/5/1/0	76800	102400	5	128	4	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
46	tp-1-sd-0.3	test-ssz4-default-cpg	5	5/4/0/1	55296	73728	6	128	4	0	cage	port	2012-08-03 17:46:10 CDT	-p -devtype FC
47	tp-1-sd-0.4	test-ssz4-default-cpg	5	0/1/4/5	92160	122880	6	128	4	0	cage	port	2012-08-03 17:55:46 CDT	-p -devtype FC
48	tp-1-sd-0.5	test-ssz4-default-cpg	5	1/0/5/4	46080	61440	5	128	4	0	cage	port	2012-08-03 17:55:46 CDT	-p -devtype FC
49	tp-1-sd-0.6	test-ssz4-default-cpg	5	4/5/1/0	55296	73728	6	128	4	0	cage	port	2012-08-03 17:55:46 CDT	-p -devtype FC
50	tp-1-sd-0.7	test-ssz4-default-cpg	5	5/4/0/1	76800	102400	5	128	4	0	cage	port	2012-08-03 17:55:46 CDT	-p -devtype FC
12					557056	770048								

Id	Name	CPG	RAID	Own	SizeMB	RSizeMB	RowSz	StepKB	SetSz	Refcnt	Avail	CAvail	-----CreationTime------	-CreationPattern-
31	tp-2-sa-0.0	test-ssz4-modified-cpg	1	0/1/4/5	4096	12288	4	256	3	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
32	tp-2-sa-0.1	test-ssz4-modified-cpg	1	1/0/5/4	4096	12288	4	256	3	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
33	tp-2-sa-0.2	test-ssz4-modified-cpg	1	4/5/1/0	4096	12288	4	256	3	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
34	tp-2-sa-0.3	test-ssz4-modified-cpg	1	5/4/0/1	4096	12288	4	256	3	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
35	tp-2-sd-0.0	test-ssz4-modified-cpg	5	0/1/4/5	147456	196608	8	128	4	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
36	tp-2-sd-0.1	test-ssz4-modified-cpg	5	1/0/5/4	147456	196608	8	128	4	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
37	tp-2-sd-0.2	test-ssz4-modified-cpg	5	4/5/1/0	147456	196608	8	128	4	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
38	tp-2-sd-0.3	test-ssz4-modified-cpg	5	5/4/0/1	147456	196608	8	128	4	0	cage	port	2012-08-02 14:30:36 CDT	-p -devtype FC
8					606208	835584								

Please see Table 2 for an illustration of the impact on physical disks in both test scenarios.

Table 2. StatPd Comparison - Default vs. Tuned Auto Growth

StatPd (write service time) Comparison – RAID5 TPVV Default vs. Tuned Growth Increment – 256PDs					
Milliseconds (95th pctl)	**SSZ4 Default**	**SSZ4 Tuned**	**Milliseconds (PD Max)**	**SSZ4 Default**	**SSZ4 Tuned**
5ms	0	1	>20 <50ms	31	14
6ms	68	107	>50 <100ms	14	8
7ms	113	115	>100 <200ms	4	7
8ms	49	21	>200ms	9	0
9ms	15	10	-	-	-
10ms	6	2	-	-	-
11ms	3	-	-	-	-
12ms	2	-	-	-	-

The tuned RAID 5 TPVV achieves usage of approximately 87% of all PDs with less than 7ms write service time. However, the PDs in the untuned TPVV are only capable of using 70% of the available PDs with less than 7ms write service time. More dramatically is the number of 'hot disks' in the default CPG, with over 10% recording over 50ms write service times (3.5% >200ms) compared with 6% (0% >200ms) in the tuned CPG. See Table 3 for both default and minimum CPG growth increments.

Table 3. CPG Growth Increments

Number of Nodes	Default	Minimum
1-2	32G	8G
3-4	64G	16G
5-6	96G	24G
7-8	128G	32G

The optimal CPG growth increment depends on several factors, including but not limited to the following:

- Total physical drives per node pair
- Number of system nodes
- RAID type
- Drive type (FC, NL, SSD)

Growth Increment explained: A 64 GB growth increment using RAID1 (set size = 2) on a 4 node system will distribute 16 chunklets per node, or 32 mirrored chunklets per node. This assumes all PDs are evenly distributed across all nodes. For a 4 node system with 128 PDs, this CPG growth increment would be optimal, resulting in allocation of LDs across all available PDs (32 mirrored chunklets per node x 4 nodes = 128 chunklets on disk). The same 64GB growth increment using RAID5 (set size = 3) would use the same 16 chunklets per node, but instead would require an additional 8 parity chunklets for 24 in total per node. In this scenario, the newly allocated LD would be striped across 96 PDs (24 chunklets x 4 nodes = 96 chunklets on PD), leaving 32 PDs unused. Ideally, this RAID5 CPG should be created with a growth increment of 87380MB which with parity overhead (2D+1P or 87380MB + 43690MB = 131070MB/1024MB = 128), which would allocate LDs across all 128 PDs.

For best storage performance, HP recommends that CPGs be tuned based upon number of physical drives, RAID type, and drive type for well-balanced and efficient disk utilization and system performance.

Beyond aligning growth increments with available PDs and RAID type, for best performance, scheduling regular CPG compactions is recommended. CPG compaction allows for capacity allocated to a CPG not currently in use to be returned to the pool of free chunklets. This activity should be scheduled during low activity periods to minimize the performance impact of chunklet initialization (zeroing) which happens automatically when chunklets are freed.

Tech tip

You can schedule regular compaction of CPGs with the **createsched** command

```
% createsched compactcpg <CPG_name> <taskschedule> <schedname>
```

The following example will perform a compactcpg every Saturday at 2:00 A.M.

```
% createsched compactcpg CPG_FC_15K_RAID5 “0 2 * * 6”
```

For more information refer to the HP 3PAR OS 3.1.1 Command Line Interface Reference guide or run “help createsched” from the CLI.

The following are high-level guidelines to ensure maximum performance and optimal reliability in the volumes supported by the logical disks created from a CPG:

- To provide the highest availability, CPGs should use cage availability, which is the default if available. If cage availability is not available, then magazine availability should be used.
- Chunklets in the same row (RAID set or set size) should be from different physical disks. In other words, a physical disk should not appear twice in the same row (RAID set).
- The system should use as many physical disks as possible.
- The load on all physical disks should be balanced.
- The system should use the largest possible row size.
- Provision virtual volume’s user space and snapshot space from different CPGs. If the virtual volume’s user and snapshot space are on different CPGs, the user space remains available to the host if the CPG containing the snapshot space becomes full.

When creating VMs, there are a number of options that are available for the VMDK files. VMware vSphere creates VMs using the “Lazy Zeroed Thick” option by default. With this option, when a new VM is created, the full size of the VMDK is not immediately zeroed. Instead, zeros are returned upon reads from unwritten areas, but not actually backed by physical storage until actual write operations. For performance-intensive environments and security concerns, VMware recommends using “Eager Zeroed Thick” (EZT) virtual disks. EZT disks have the smallest overhead but require zeros to be written across all of the capacity of the VMDK at the time of creation. Unlike many other storage vendors, HP 3PAR Thin Persistence Software and HP 3PAR Zero Detect enabled virtual volumes allow clients to retain the thin provisioning benefits when using Eager Zeroed Thick VMDKs without sacrificing any of the performance benefits offered by this VMDK option. Please see the Virtual SCSI adapters and virtual disk types section of this document for a comparison of the available disk types with VMware vSphere 5.

3PAR Thin Provisioning vs. vSphere Thin Provisioning

When implementing HP 3PAR TPVVs, administrators often ask whether implementing vSphere Thin Provisioning for VMDK files makes any sense. In general, Thin Provisioning with 3PAR and vSphere accomplish the same end-result, albeit at different logical layers. With VMware vSphere Thin Provisioning, administrators realize greater VM density at the VMFS layer, at the cost of some CPU and disk I/O overhead as the volume is incrementally grown on the ESXi hosts. By implementing HP 3PAR TPVVs, the same VM density levels are achieved, however the thin provisioning CPU work is offloaded to the HP 3PAR ASIC. If the goal is to reduce storage costs, maximize storage utilization, and maintain performance, then use HP 3PAR Thin Provisioning Software to provision VMFS volumes. If performance is not a concern but over-provisioning VMs at the VMFS layer is important, then administrators can consider implementing both Thin Provisioning solutions. However, administrators should realize that there are no additional storage savings realized by using VMware Thin Provisioning on top of 3PAR TPVVs and in fact, implementing both solutions adds more management complexity to the environment. For a better understanding of the performance tradeoffs implementing both Thin Provisioning solutions, see the Virtual SCSI adapters and virtual disk types topic in the Performance tuning section of this document.

Thin Conversion

HP 3PAR Thin Conversion Software is an optional feature that converts a fully-provisioned volume to a Thinly-Provisioned Virtual Volume (TPVV). Virtual volumes with large amounts of allocated but unused space are converted to TPVVs that are much smaller than the original volume. During the conversion process, allocated but unused space is discarded and the result is a TPVV that uses less space than the original volume. The conversion process has four steps.

1. Assessment
2. Data preparation
3. Zeroing unused space
4. Creating a physical copy

Assessment

Before converting volumes the benefits of the conversion process must be determined. The potential benefits of zeroing free space prior to copying or migrating the data to a TPVV depends on the amount of allocated but unused space. If there is relatively little unused space in the allocated physical space then there is little benefit to zeroing the free space to recapture this relatively small amount of space. If there is a large amount of unused space then zeroing the data prior to copying will result in a substantial reduction in the amount of used space on the HP 3PAR Storage.

Data preparation

Prepare your data for copying by removing unnecessary data. Perform clean-up tasks on the source volume by:

- Emptying trash cans or permanently deleting files.
- Archiving unused files.
- Shrinking databases.
- Deleting temporary files.

Zeroing unused space

Use a guest OS application (sdelete or dd) to write zeros to the allocated but unused volume space. The HP 3PAR Storage Systems detect and discard the zeros during the volume copy operation.

Creating a physical copy

After writing zeros to the allocated but unused space, the source volume is ready for the final phase of conversion. You create a TPVV physical copy of the source volume to convert the source volume to a TPVV. When you create a physical copy, 3PAR Storage Systems automatically detect the zeros and do not allocate space for them in the physical copy. The result is a TPVV that is much smaller than the original volume.

Thin Persistence

HP 3PAR Thin Persistence Software is an optional feature that keeps TPVVs and read/write snapshots of TPVVs small by detecting pages of zeros during data transfers and not allocating space for those pages. This feature works in real-time and analyzes the data before it is written to the source TPVV or read/write snapshot of the TPVV. Freed blocks of 16 KB of contiguous space are returned to the source volume and freed blocks of 128 MB of contiguous space are returned to the CPG for use by other volumes.

Thin Copy Reclamation Software, which is included with every HP 3PAR storage system, reclaims space when snapshots are deleted from a system. As snapshots are deleted, the snapshot space is reclaimed from a Thinly-Provisioned Virtual Volume (TPVV) or fully-provisioned virtual volume and returned to the CPG for reuse by other volumes. Deleted snapshot space can be reclaimed from virtual copies, physical copies, or remote copies. The 3PAR OS automatically reclaims snapshot space if the Virtual Copy, Remote Copy, or Thin Provisioning license is enabled.

Note

Reclaiming space when snapshots are deleted with the Thin Copy Reclamation feature requires the Virtual Copy, Remote Copy, or Thin Provisioning license.

Zero Detect

Zero Detect is currently disabled by default in HP 3PAR Inform OS 3.1.1. However, it is a best practice to enable zero detect on TPVVs. Zero Detect can be enabled by using the “Advanced options” checkbox of the 3PAR Management Console. Zero Detect enables Thin Persistence and achieves space reclamation. For example, when an administrator deletes a vmdk file on an HP 3PAR TPVV the HP 3PAR ASIC detects those zeros as the same pattern (SCSI write_same). Those blocks having been earlier utilized triggers HP 3PAR Thin Persistence reclamation of those blocks and releases them back to the CPG. Thin persistence can reclaim space in chunks of 128MB of contiguous unused or zero-filled space on the LD. Please see Figure 13 below and take note of approximately 80GB total space reclaimed.

Figure 13. Thin Persistence Reclaiming Space

```
reclaim starts for test-r5-custom 0 [81c0000, 9500000) <--- Reclaim operation begins

Modifying maps for VV test-r5-custom (30)(VT_RECLAIM_SD), offset 0x81c0000 length 0x40000 [118 tp-9-sd-0.0, 0x23c0000]->[10000037 zerold.4.5.0, 0x3800000]
Modifying maps for VV test-r5-custom (30)(VT_RECLAIM_SD), offset 0x8200000 length 0x40000 [118 tp-9-sd-0.0, 0x2400000]->[10000037 zerold.4.5.0, 0x3840000]

... entries removed for brevity ...

Modifying maps for VV test-r5-custom (30)(VT_RECLAIM_SD), offset 0xa900000 length 0x80000 [120 tp-9-sd-0.2, 0x2a00000]->[10000055 zerold.6.7.0, 0x3e80000]
Modifying maps for VV test-r5-custom (30)(VT_RECLAIM_SD), offset 0xa980000 length 0x80000 [120 tp-9-sd-0.2, 0x2a80000]->[10000055 zerold.6.7.0, 0x3f00000]

Volume 'test-r5-custom' sd id 1 takes 0 seconds to reclaim 20224 MB space, total reclaimed history is 84608 MB. <--- reclaim operation completes
```

HP 3PAR Adaptive Optimization

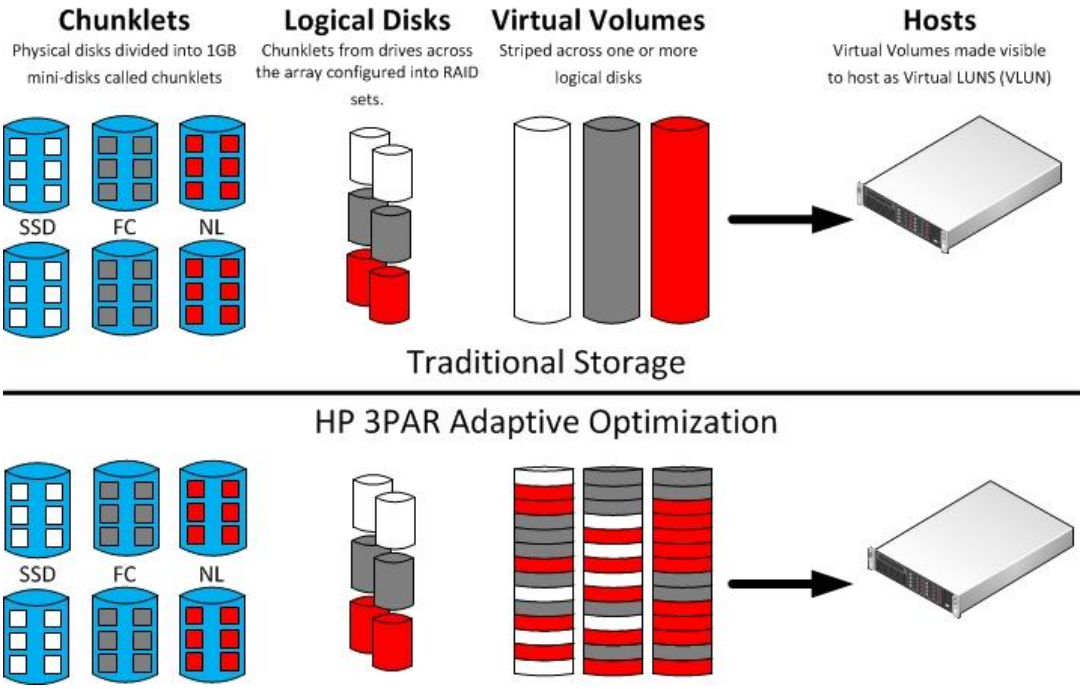
Manual and auto optimization on HP 3PAR Storage

Tiered storage is a data storage environment consisting of two or more kinds of storage. With HP 3PAR, tiers are delineated by drive type, Solid State Drives (SSD), Fibre Channel (FC) drives and Nearline (NL) SATA, as well as RAID level and drive speed. HP 3PAR has two implementations to deal with the tiering of storage, HP 3PAR Dynamic Optimization and HP 3PAR Adaptive Optimization. Both optimization technologies enable storage administrators to optimize their storage resources, without downtime, extensive configuration, or risk of not meeting SLAs.

HP 3PAR Dynamic Optimization enables users to manually and non-disruptively alter service levels associated with a storage volume by RAID level, subsystem failure protection level, drive type, stripe width, and/or radial placement, and rebalance storage volumes, each with a single click.

HP 3PAR Adaptive Optimization Software takes capabilities of HP 3PAR Dynamic Optimization to the next level by taking a fine-grained, highly automated approach to service level optimization. Policy-driven, granular data movement takes place autonomically, providing highly reliable, non-disruptive, cost-optimized storage tiering at the sub-volume level to deliver the right Quality of Service to the right data at the right time on a large scale. Figure 14 shows how HP 3PAR Adaptive Optimization uses pieces of every tier in an HP 3PAR Storage array to ensure the best performance and cost effectiveness of the array.

Figure 14. HP 3PAR Tiering – Non-Tiered vs. Tiered



Implementing Adaptive Optimization

In order to implement HP 3PAR Adaptive Optimization, HP 3PAR System Reporter needs to be able to collect historical data for all the regions in an array, analyze the data to determine the volume regions that should be moved between tiers, and move the regions from one CPG (tier) to another. To see this activity, System Reporter can provide the user with reports showing the regions moved and the impact of Adaptive Optimization.

The remainder of this section does not include installation and initial configuration of System Reporter but will focus on use and methods in an established environment.

For installation and configuration of System Reporter, see:
http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c03050683/c03050683.pdf

For details on Adaptive Optimization, see: http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-0867ENW

Best practices

HP 3PAR Storage Systems are designed to securely and efficiently support multiple "tenants" – different groups of users from different departments or even different enterprises – as they run diverse workloads on a common storage system. Because of this, multiple Adaptive Optimization (AO) configurations can be created to meet the needs of all or some of the configurations (Virtual Domains) on HP 3PAR Storage.

HP 3PAR Virtual Domains provide secure access and improved storage services for different applications and user groups. Virtual Domains provide secure, administrative segregation of users and hosts within the HP 3PAR Storage system. Individual user groups and applications can achieve greater storage service levels (performance, availability and functionality) than previously possible with our storage arrays.

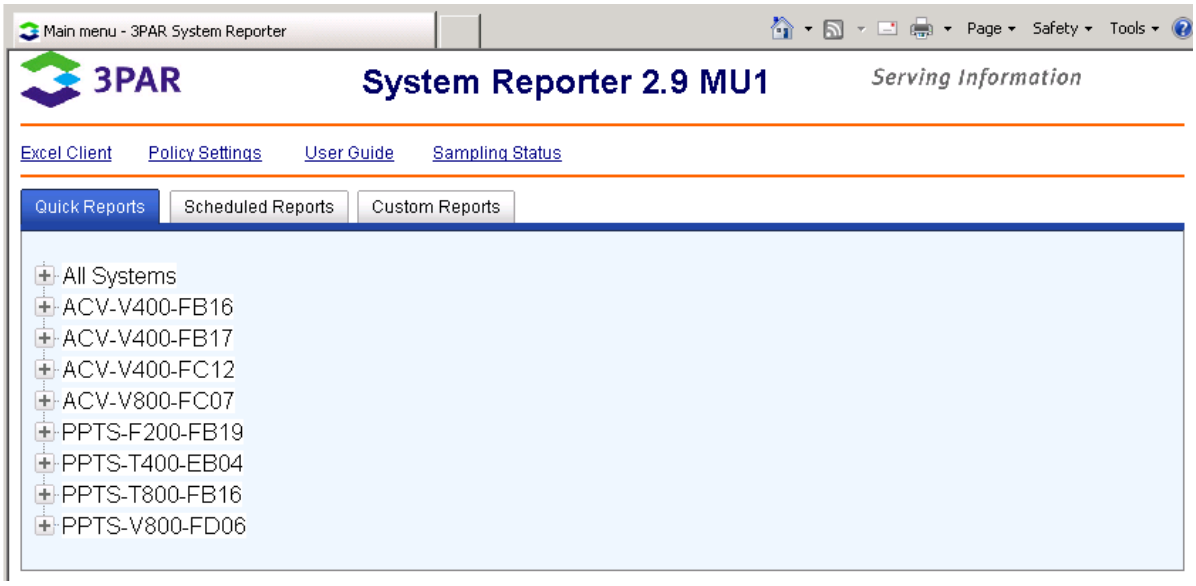
AO Profiles can be created at the storage level of an array or inside a virtual domain which is treated as an array. This is done by creating a Profile in HP 3PAR System Reporter.

HP 3PAR System Reporter

HP 3PAR System Reporter periodically collects detailed performance and space data from HP 3PAR arrays, stores the data in a database, analyzes the data, and generates reports. HP 3PAR Adaptive Optimization leverages HP 3PAR System

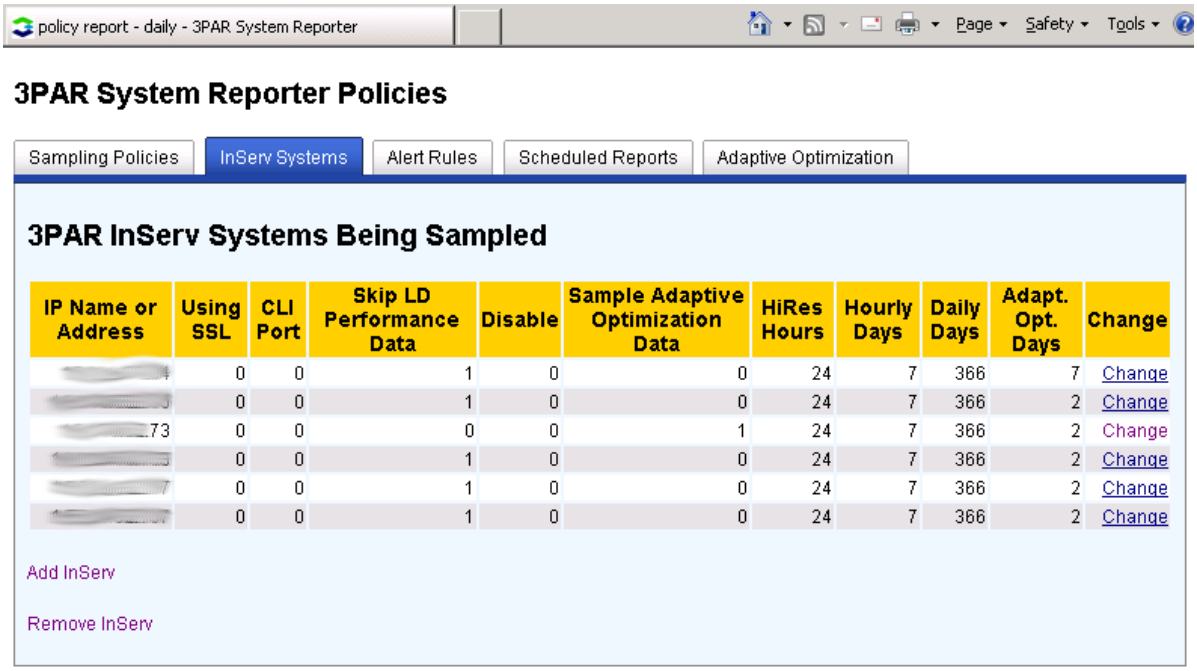
Reporter by adding enhancements to collect region-level performance data, perform tiering analysis, and issue region movement commands automatically as defined in the policies set by the storage administrator. You can also manage several arrays with System Reporter, and thus have multiple policies and configurations for HP 3PAR Adaptive Optimization for multiple arrays from a single pane of glass (Figure 15).

Figure 15. Home screen for HP 3PAR System Reporter



Before defining an Adaptive Optimization policy, you first need to ensure that the HP 3PAR array is configured to collect Adaptive Optimization data. To check, select **Policy Settings** and then select **InServ System** as shown in Figure 16.

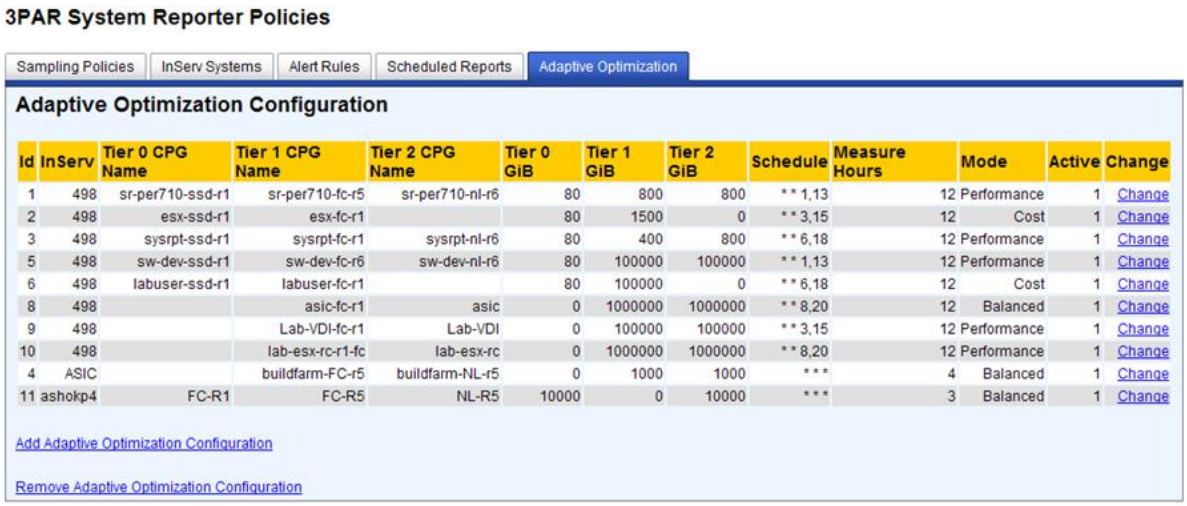
Figure 16. HP 3PAR System Reporter Policies – InServ Systems



In Figure 16 above, the third array in the list (73) is the HP 3PAR Storage system we will be configuring Adaptive Optimization on. For HP 3PAR Adaptive Optimization to collect the data it needs, **Skip LD Performance Data** needs to be set to **0** and **Sample Adaptive Optimization Data** needs to be set to **1**. Both of those setting are not the default, so you might have to select **Change** on the array entry and change those values.

To add, remove, modify or view a summary of all your HP 3PAR Adaptive Optimization policies, select **Policy Settings** and then select **Adaptive Optimization** as in Figure 17.

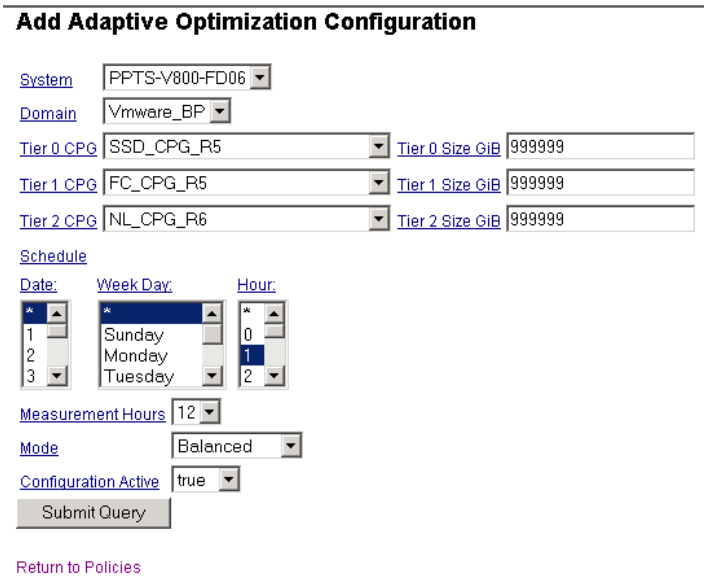
Figure 17. Adaptive Optimization configuration screen



Id	InServ	Tier 0 CPG Name	Tier 1 CPG Name	Tier 2 CPG Name	Tier 0 GiB	Tier 1 GiB	Tier 2 GiB	Schedule	Measure Hours	Mode	Active	Change
1	498	sr-per710-ssd-r1	sr-per710-fc-r5	sr-per710-nl-r6	80	800	800	* * 1,13	12	Performance	1	Change
2	498	esx-ssd-r1	esx-fc-r1		80	1500	0	* * 3,15	12	Cost	1	Change
3	498	sysrpt-ssd-r1	sysrpt-fc-r1	sysrpt-nl-r6	80	400	800	* * 6,18	12	Performance	1	Change
5	498	sw-dev-ssd-r1	sw-dev-fc-r6	sw-dev-nl-r6	80	100000	100000	* * 1,13	12	Performance	1	Change
6	498	labuser-ssd-r1	labuser-fc-r1		80	100000	0	* * 6,18	12	Cost	1	Change
8	498		asic-fc-r1	asic	0	1000000	1000000	* * 8,20	12	Balanced	1	Change
9	498		Lab-VDI-fc-r1	Lab-VDI	0	100000	100000	* * 3,15	12	Performance	1	Change
10	498		lab-esx-rc-r1-fc	lab-esx-rc	0	1000000	1000000	* * 8,20	12	Performance	1	Change
4	ASIC		buildfarm-FC-r5	buildfarm-NL-r5	0	1000	1000	* * *	4	Balanced	1	Change
11	ashokp4	FC-R1	FC-R5	NL-R5	10000	0	10000	* * *	3	Balanced	1	Change

To create an HP 3PAR Adaptive Optimization profile, select **Add Adaptive Optimization Configuration**.

Figure 18. AO configuration in System Reporter



From this example you can see what information is required to define a policy. Table 4 provides descriptions for some of the fields shown in Figure 18.

Table 4. HP System Reporter AO configuration field description and use

Field	Description
System	The HP 3PAR Storage to apply the policy to
Domain	The domain the CPGs are in for use with AO
Tier 0 CPG	Highest performing tier for use in this AO policy
Tier 1 CPG	Next highest performing tier for use in this AO policy
Tier 2 CPG	Lowest performing tier for use in this AO policy

It is recommended and best practice to define tier 0 to be higher performance than tier 1, which in turn should be higher performance than tier 2. For example, RAID 5 with SSDs for tier 0, RAID 5 with FC drives for tier 1 and RAID 6 with NL or SATA drives for tier 2. Another example, could be, RAID 1 with FC 15K for tier 0, RAID 6 with FC 15K drives for tier 1 and RAID 6 with NL or SATA drives for tier 2.

It is also recommended to set AO CPG's sizes to 999999 in each profile in the AO configuration. This allows the HP 3PAR OS provisioning mechanism to control the size of the CPG growth increments which lends to a much more manageable and flexible solution that makes the best use of the capacity available in the system.

Data movement can be scheduled by storage administrators after the specified measurement period has completed. It is important to note that the execution schedule that is defined in an AO policy controls the time at which data is moved, not the beginning of the measurement period as measurements are always collected by HP 3PAR System Reporter. It is also recommended and best practice that all AO policies for a given HP 3PAR System be executed once at the same time every day. This allows the administrator to schedule data movement at times when the additional overhead of that data movement is acceptable, non-peak hours for example.

The mode configuration parameter can be set to one of three values:

- *Performance mode* biases the tiering algorithm to move more data into faster tiers.
- *Cost mode* biases the tiering algorithm to move more data into the slower tiers.
- *Balanced mode* is a balance between performance and cost.

The mode configuration parameter does not change the basic flow of the tiering analysis algorithm, but rather it changes certain tuning parameters that the algorithm uses. It is recommended to start with Balanced by default, then move to a cost or performance model based on needs.

Also your sample time needs to be long enough to characterize the behavior of your application. Samples under four to six hours will not produce enough profile data for accurate optimization. Start with periods of 24 hours even if your application is quiescent off hours.

Note
HP 3PAR Adaptive Optimization supports a maximum of 125 TB of data per AO policy.

You can view a summary of all of your HP 3PAR Adaptive Optimization policies from the **Adaptive Optimization** tab in System Reporter.

Physical disk concerns

When setting up the different tiering levels in an HP 3PAR Adaptive Optimization profile definition in HP 3PAR System Reporter, it is not recommended to use only SSD and NL tiers. Using different RAID levels within the same policy however is acceptable. See Table 5 for examples.

Table 5. Examples of acceptable Adaptive Optimization configurations

Tier 1	Tier 2	Tier 3
SSD	FC	NL
SSD	FC	
FC	NL	
SSD	FC RAID 1	FC RAID 6
FC RAID 1	FC RAID 6	NL

CPG tiering considerations

When creating and using a Solid State Disk CPG, use the following as a best practices guide for your environment:

- Use **RAID5** with a set size of **3 + 1** as this provides the best performance/capacity solution.
- If maximum performance is required, use RAID1.
- Create SSD CPG configurations that will use up to 95% of SSDs.
- Set the growth increment to the minimum values as defined in Table 3. CPG Growth Increments.
- Availability should be set to **cage**. If cage availability is not available, use magazine.
- Leave all other values at their defaults when creating the CPG.

When creating and using a Fibre Channel CPG, use the following as a best practices guide for your environment.

- Use **RAID5** with a set size of **3 + 1** as this provides the best performance/capacity/redundancy solution.
- CPGs that have a high write ratio greater than 50% should leverage RAID1 if maximum performance is needed.
- Leave the growth increment to the default values.
- Availability should be set to **cage**. If cage availability is not available, use magazine.
- Leave all other values at their defaults when creating the CPG.

When creating and using a Nearline CPG, use the following as a best practices guide for your environment:

- Always use **RAID6** by default with a set size of **8**. A larger set size is acceptable as long as you can keep cage availability.
- RAID5 is not recommended due to the long rebuild times of these large drives. The chance of a failure in a second disk before a rebuild is complete is too great.
- Leave the growth increment to the default values.
- Availability should be set to **cage**. If cage availability is not available, use magazine.

Thin Provisioning

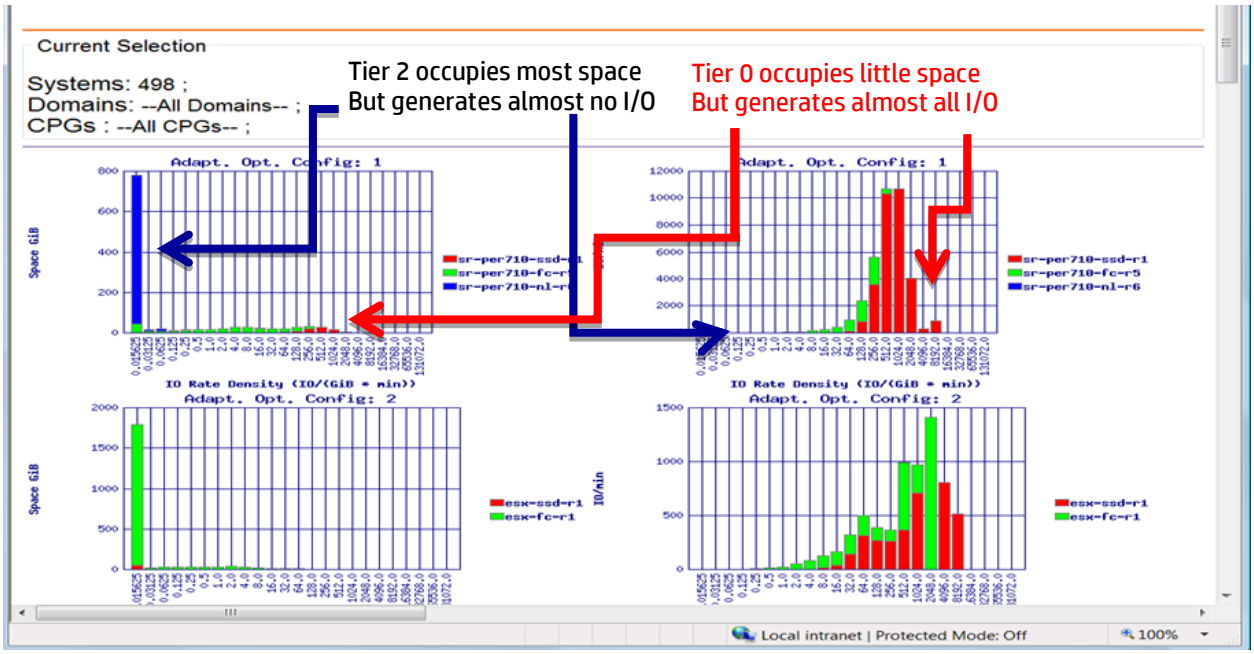
When leveraging HP 3PAR Thin Provisioning, create new volumes on a CPG using FC disks. This ensures that new data being written has good performance and you allow HP 3PAR Adaptive Optimization to manage the SSD tier properly and to migrate regions as needed to the SSD tier.

Adaptive Optimization in action

Once configured, you can see the impact of Adaptive Optimization from the data collected by HP 3PAR System Reporter. In Figure 19 below, the upper two graphs show the region I/O density after HP 3PAR Adaptive Optimization has gathered enough data to implement its changes based on the policies set by the storage administrator. Both charts are histograms with the x-axis showing I/O Rate Density, with the busiest regions to the right and the idlest to the left. The chart on the left shows on the y-axis the capacity for all the regions in each bucket, while the chart on the right shows on

the y-axis the total IOPS/minute for the regions in each bucket. As shown in the charts, the SSD tier (tier 0) occupies very little space but absorbs most of the I/O accesses, whereas the Nearline tier (tier 2) occupies most of the space but absorbs almost no accesses at all.

Figure 19. Summary report for an AO policy



This is precisely what HP 3PAR Adaptive Optimization was created to do, utilize the expensive SSD drives to their fullest potential, and put the least used data in the lower, cheaper tiers.

Performance tuning

While virtual environments certainly increase server consolidation, simplify administration, and maximize ROI, these environments also present unique performance challenges for storage administrators. After consolidation, typically most environments will find memory over utilization a significant factor. In fact, VMware vSphere utilizes several sophisticated techniques for handling memory over-commitment including page sharing, ballooning, compression and swapping. These techniques enable effective memory management which in turn enables more virtual machines to be hosted on a single host. However, several of these memory handling processes place a greater load on the backend storage system. Namely, ballooning and swapping often will page to disk and it is here where high speed and low latency I/O is critical for optimum performance in a VMware vSphere environment. Compounding the reliance on storage resources is the random nature of virtualized workloads, which typically do not find data in storage read cache but must fetch on disk. There is some benefit to enabling read cache in virtualized environments, but primarily, the ability for storage systems to deliver IOPS is going to achieve best performance.

I/O sizing

Traditional storage systems have administrators working through various storage sizing exercises seeking to define volume I/O per any given workload. With a traditional disk based array, administrators must identify storage performance requirements for given workloads using some of the following guidelines:

- Different RAID groups (think physical drives) tailored to specific I/O performance, capacity, and redundancy with volumes assigned to these different groups
- When RAID groups do not meet I/O requirements, additional, dedicated groups must be created for newly provisioned volumes

- Manually balancing/distributing the RAID groups and volumes across the available storage processors for best performance

These management considerations are time consuming, and can be expensive and complex to implement, especially when adding additional storage to meet I/O requirements. With HP 3PAR Storage systems, there is no separation of underlying physical drives and logical RAID groups and all of the planning and spreadsheets to keep track of. Instead an HP 3PAR system will leverage its build in Wide striping capabilities to use every defined physical disk (FC, NL, SSD) on the array. Because of this, I/O sizing is significantly simplified with 3PAR Storage, and storage administrators should consider the I/O requirements of their environment and then align the 3PAR Storage configuration (SPC-1 benchmark results) to meet those specific requirements.

SPC-1 benchmark results

The Storage Performance Council (SPC) defines and administers industry standard benchmarks to characterize storage products. The SPC-1 benchmark was designed to demonstrate the performance of a storage subsystem while performing the typical functions of business critical applications. Those applications are characterized by predominately random I/O operations and require both queries as well as update operations. In October 2011, HP announced that the HP P10000 3PAR V800 Storage System delivered record SPC-1 results, yielding over 450,000 IOPS: storageperformance.org/results/benchmark_results_spc1

At the time of the publication of this paper, this was the number one SPC-1 result for a single storage array.

Alignment considerations

The improper alignment of VMFS file system partitions may impact performance. The recommended practice is to add VMFS storage to ESXi hosts using the vSphere Client, as it automatically aligns VMFS partitions when it creates them. For ESXi 5, VMFS3 and VMFS5 file systems that are created using the vSphere Client are automatically aligned on a 1 MB boundary. VMFS3 file systems created with a previous version of ESX/ESXi used 64 KB alignment.

Partitions that are created using vmkfstools may be aligned manually using the partedUtil tool from the command line. For detailed instructions on using partedUtil, refer to the VMware Knowledge Base entry: http://kb.vmware.com/kb/1036609. Note that when using partedUtil, alignment is determined by the start sector parameter. Specifying a start sector at offset 128 will result in 64 KB alignment, and a start sector at offset 2048 will provide 1 MB alignment.

Virtual SCSI adapters and virtual disk types

In general, with ESXi 5, the default virtual SCSI adapter for any given OS is the best choice for most virtual machine workloads. However, the same statement cannot be made for the virtual disk options available to administrators. These include thick virtual disks which have all space allocated at time of creation and thin virtual disks which have their space allocated but written to upon first write operation. Additionally, two thick virtual disks options are available to choose from, eager-zeroed and lazy-zeroed. Eager-zeroed allocates all the space requested by writing zeros for the entire virtual disk at creation, while lazy-zeroed only zeros at first write. Administrators also have the option of provisioning raw device mapped (RDM) volumes which allow for management access of raw SCSI LUNs as VMDK files. In terms of performance, each virtual disk type has unique characteristics as noted in Table 6 below.

Table 6. Performance comparison - virtual disk types

StatvLun (Hosts) Counters		Lazy Zero Thick (first-write) on TPVV	Lazy Zero Thick (after block zero) on TPVV	vSphere TP and TPVV	RDM on TPVV	Eager Zeroed Thick on TPVV	Fully Provisioned (3PAR)
Write I/O per sec	95th pctl	663	715	758	723	717	718
	Avg	631	674	729	689	690	686
	Max	672	742	779	752	740	735
Write KB per sec	95th pctl	130948	379988	132791	384151	382198	382312
	Avg	124912	358678	127505	366625	366866	365317
	Max	133740	394937	135316	401146	394267	390520
Write Svt ms	95th pctl	2	37	2	31	37	27
	Avg	2	32	2	28	33	24
	Max	3	39	2	34	39	33

Table 6 is limited to large block file transfers only and thus records write operations on disk. Most notably, lazy-zeroed and host thin provisioned volumes (1st and 3rd column results) perform very similarly with first write penalties recorded in both test cases. Subsequent writes to both types of volumes does show a notable increase in performance as seen and recorded in the second column (Lazy Zeroed Thick (after block zero) on TPVV). In all cases, utilizing host-side thin provisioning reduces the underlying volume performance while at the same time incurring additional ESXi host overhead.

See Table 7 below for a comparison of host thin provisioned versus eager zeroed thick I/O characteristics. This comparison illustrates that host-side thin provisioned volumes are not pre-allocated upon creation and must be read from prior to write operations; thus, the read I/O penalty seen below.

Table 7. Read I/O penalty for host-side thin provisioned volumes

StatPort (Host) Counters		vSphere TP and TPVV	Eager Zeroed Thick on TPVV
Read I/O per sec	95thpctl	2.0	0.0
	Avg	1.3	0.0
	Max	4.0	1.0
Read IOSz KB	95th pctl	35	4.1
	Avg	37	2.6
	Max	280	4.1
Read KBytes per sec	95th pctl	60	2.0
	Avg	61	0.5
	Max	1119	2.0
Read Svt ms	95th pctl	1.0	3.9
	Avg	0.7	0.9
	Max	7.8	12

In terms of performance and overhead, RDM, eager-zeroed, and fully provisioned 3PAR VMDKs compare favorably with fully provisioned volumes experiencing the shortest disk service times and least amount of storage CPU cycles. Overall, all three virtual disk types perform remarkably similar and the choice of virtual disk type should be based on specific application requirements.

Wide striping

The unique HP 3PAR Architecture stripes volumes widely across all drives to deliver maximum I/O throughput and minimum latencies, which mitigates server memory bottlenecks and traditional storage constraints. Increased array performance cannot only boost VM-based application performance, but when paired with the superior reliability of the HP 3PAR Storage System and advanced support of VMware's vSphere Storage APIs for Array Integration (VAAI) capabilities, result in higher VM density. This benefit enables organizations to double virtual machine density on physical servers by placing twice as many VMs on physical servers as compared with traditional storage platforms.

Storage I/O Control

Note
vSphere Storage I/O Control (SIOC) is a vSphere feature which manages ESXi device-queue depths, while AO is a 3PAR feature which moves data between storage tiers at scheduled intervals depending upon usage patterns. The two features operate at different layers, so there is no conflict between them.

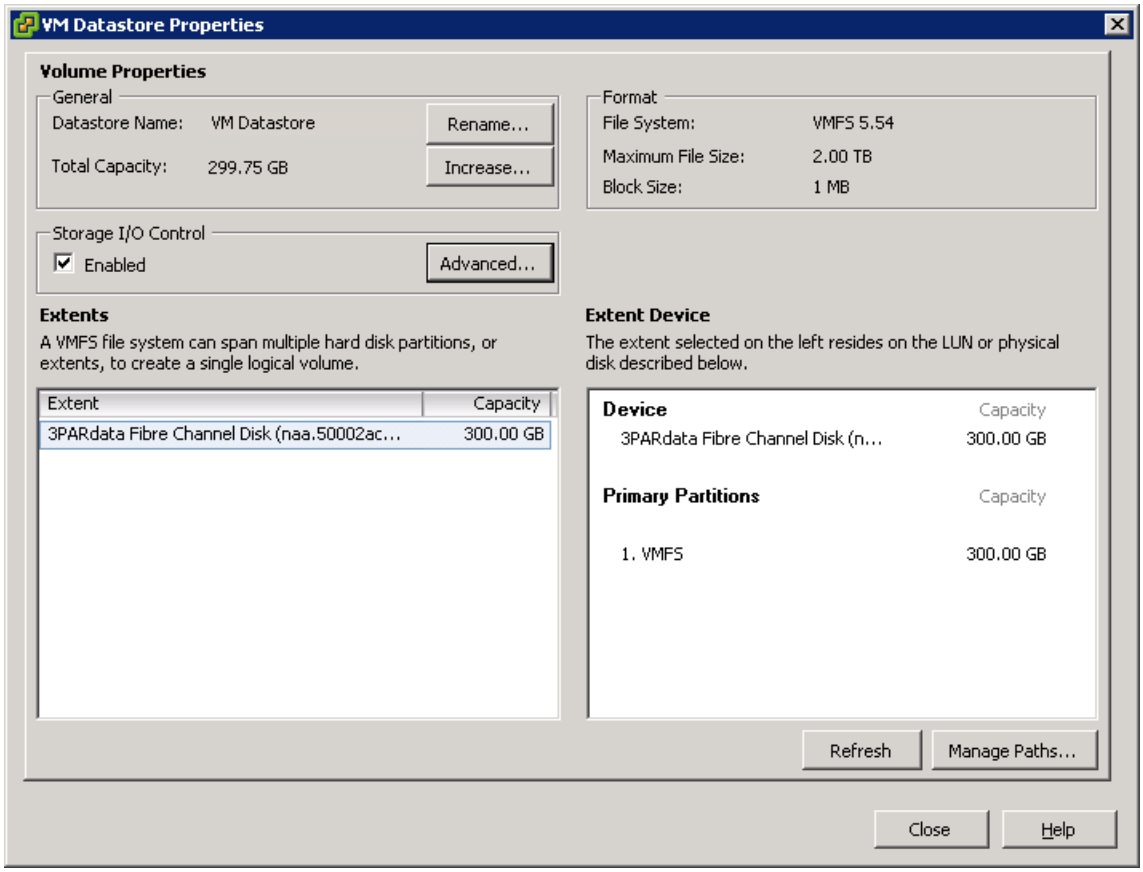
SIOC provides finer-grained control than HP 3PAR Adaptive queue depth throttling, and the latter is not needed if SIOC is enabled.

The vSphere Storage I/O Control (SIOC) feature manages shared storage resources across ESXi hosts to provide more predictable performance during periods of congestion. It monitors the latency of I/Os to a datastore for each ESXi host sharing the device. When the average latency for a datastore exceeds a threshold (default is 30 ms), SIOC distributes the storage resources to virtual machines according to their assigned shares. It accomplishes this by reducing the number of I/O queue slots available to lower priority virtual machines and increasing the slots for virtual machines with higher shares. By controlling the ESXi device-queue depths in proportion to the virtual machine shares, SIOC is able to control storage congestion for the datastore and distribute HP 3PAR Storage array resources appropriately. Note that the congestion threshold represents a tradeoff between lower I/O latencies and throughput. When the threshold is set low, I/O throttling will be engaged more aggressively, which will help to maintain a lower I/O latency for the datastore, but will also reduce the overall throughput for the datastore.

To enable SIOC for a datastore:

1. In the vSphere Client inventory panel, select the datastore.
2. Click the **Configuration** tab, and click **Properties** for the datastore.
3. Click the **Enabled** check box under Storage I/O Control, as shown in Figure 20.
4. If you wish to change the congestion threshold, click on the **Advanced** button, enter a value and click **OK**.

Figure 20. VM Datastore Properties



The VMware white paper Storage I/O Control Technical Overview and Considerations for Deployment makes the following recommendations for determining the congestion threshold values for various types of storage, including auto-tiered storage.

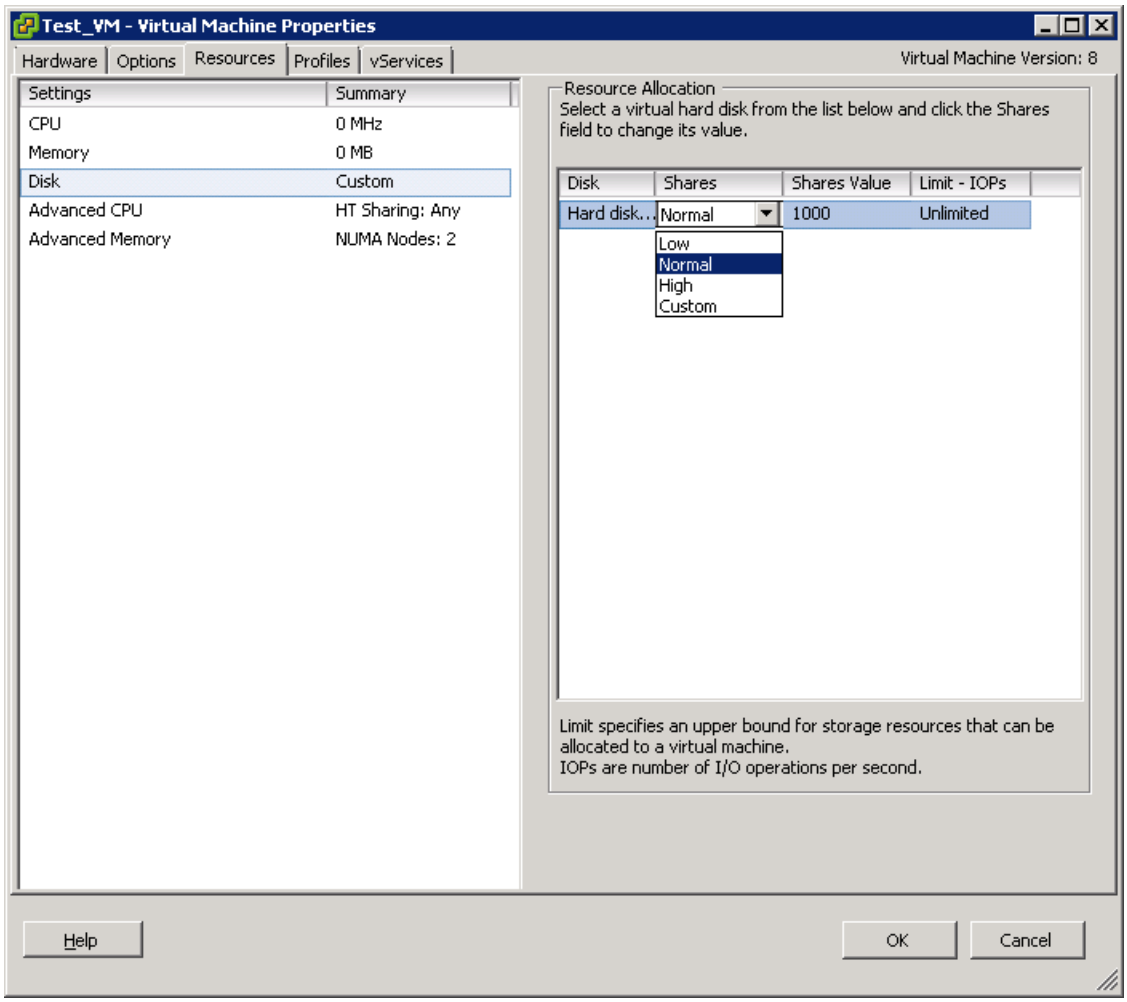
Table 8. Recommended congestion threshold values

Type of storage backing the datastore	Recommended threshold
SSD	10-15 ms
Fibre Channel	20-30 ms
SATA	30-50 ms
Auto-tiered storage	Combine ranges of fastest and slowest storage types

The default threshold of 30 ms should be acceptable for most workloads using Fibre Channel or SATA drives. If Adaptive Optimization is employed, then the threshold should be set to a value that is within the recommended ranges for the fastest and slowest storage types in use. For example, if Fibre Channel and SSD drives are specified by the AO policy, then the threshold for the datastore should be set in the range of 10 to 30 ms. The exact setting should be determined by the requirement to favor low latency (using a low congestion threshold) versus throughput (using a high threshold).

The virtual machine's relative priority on the datastore is determined by its number of disk shares. This is configured by editing the settings for the VM as shown in Figure 21:

Figure 21. Changing disk shares for a VM



Select the **Resources** tab, click on **Disk**, then select the VMDK (Hard disk) of interest, The default number of disk shares is 1000. To increase the priority of a VM relative to other VMs using the same datastore, increase the number of disk shares by clicking on the **Shares** field and change the value to **High**, or select **Custom** and enter a new value in the **Shares Value** field.

HP 3PAR Adaptive queue depth throttling

Note
Adaptive queue depth throttling is not needed if Storage I/O Control is enabled.

Adaptive queue depth throttling is not compatible with Storage DRS.

If adaptive queue depth throttling is enabled, it is important to enable it for all hosts which are accessing the 3PAR Storage.

Each port on the HP 3PAR Storage System has a finite queue depth that depends on the host bus adapter (HBA) model; each server attached to a port shares that port's queue. If a host sends an I/O request to a port with a full queue, the host receives a "queue full" SCSI response from the HP 3PAR array. I/O commands sent to a port in an HP 3PAR array that has reached its maximum queue depth are not processed beyond the "queue full" SCSI response.

Historically, an ESX host's default reaction to this response would be to recognize it as a valid command and to continue sending requests to that port. Lack of I/O responses can result in VMs becoming unresponsive and can lead to a crash of the ESX host. ESX 3.5 Update 4 and later include an adaptive queue depth throttling algorithm which adjusts the LUN queue depth in the VMkernel I/O stack. This algorithm is activated when the storage array indicates I/O congestion by returning a "queue full" SCSI status. When congestion is detected, the VMkernel throttles the LUN queue depth and attempts to gradually restore the queue depth when congestion conditions subside.

Without adaptive queue depth throttling, administrators are forced to limit the number of VMs per physical server so as to reduce the risk associated with any particular VM overrunning I/O queues. Administrators are also forced to manually tune the number of VMs when they detect congestion – a reactive, slow, and error-prone process. By automating congestion control, administrators can confidently create a higher number of VMs per physical server without the need for manual congestion control.

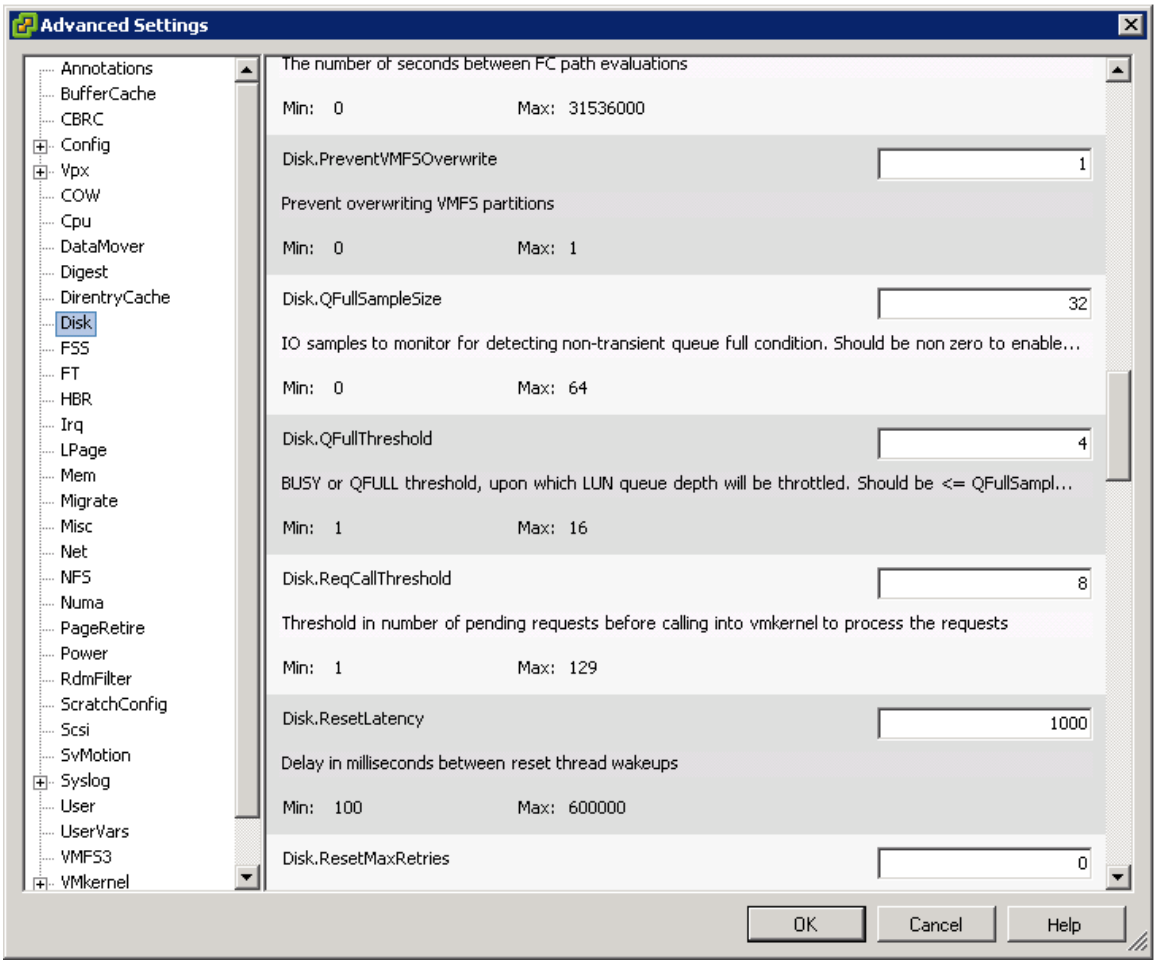
The adaptive queue depth algorithm is disabled by default. It can be enabled by setting two VMware system-wide configuration parameters, QFullSampleSize and QFullThreshold. Setting the QFullSampleSize parameter to a value greater than zero activates the algorithm. The QFullThreshold parameter must be set to a value less than or equal to QFullSampleSize. To set these parameters to optimal values for 3PAR Storage:

1. In the vSphere Client inventory panel, select the host.
2. Click the **Configuration** tab, and click **Advanced Settings** under **Software**.
3. Set the **Disk** to the appropriate values:

   Disk.QFullSampleSize = 32

   Disk.QFullThreshold = 4

Figure 22. Setting of Disk.QFullSampleSize and Disk.QFullThreshold to their recommended values



Alternatively, to set these values from the command line on the ESXi 5 hosts, run the following commands:

```
# esxcli system settings advanced set --int-value 32 --option /Disk/QFullSampleSize
# esxcli system settings advanced set --int-value 4 --option /Disk/QFullThreshold
```

These settings take effect immediately and do not require a reboot of the ESXi hosts. For more information, refer to the VMware Knowledge Base entry: http://kb.vmware.com/kb/1008113.

If it is not possible to enable VMware's adaptive queue depth throttling algorithm, then the HP 3PAR VMware ESX Implementation Guide (http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c03290624/c03290624.pdf) contains details to tune/throttle the queue depth on the ESXi host.

Tech tip

You can monitor the "Qlen" values on the system (using System Reporter or the command `statvlun -ni -rw -host <ESX host>`) to make sure you are not exceeding these values.

Recovery Manager

Overview

HP 3PAR Recovery Manager for VMware vSphere is an array-based, online VM snapshot and recovery solution that gives superior control over data protection and recovery in VMware vSphere environments – including granular, rapid online recovery of files. As with the HP 3PAR Management Plug-In for VMware vCenter, HP 3PAR Recovery Manager gives administrators access to all of this functionality from within the familiar and easy-to-use VMware vCenter Server virtualization management console.

Providing virtual copy management and LUN-level snapshot capabilities, the HP 3PAR Recovery Manager for VMware vSphere delivers array-based snapshots that are fast, space-efficient and VM aware, while traditional backup schemes are slow, complex and lacking in both flexibility and granularity. HP 3PAR Recovery Manager for VMware vSphere snapshots offload the workload associated with performing backups from VMware host to the HP 3PAR Storage.

HP 3PAR Recovery Manager for VMware vSphere makes possible the creation, and subsequent restoration, of hundreds of virtual copies, with the retention periods of those copies easily specified by an administrator. Once a virtual copy has been created, HP 3PAR Recovery Manager for VMware vSphere allows flexible and highly-granular restorations, at the VMFS layer.

As HP 3PAR Recovery Manager for VMware vSphere is not installed on the VMware ESXi hosts, its operation does not impact server performance. Snapshots with HP 3PAR Recovery Manager for VMware vSphere are quick and non-disruptive. Space used for snapshots is minimal and has a smaller footprint than VMware's own snapshots because HP 3PAR Recovery Manager for VMware vSphere leverages HP 3PAR Virtual Copy to perform those tasks.

HP 3PAR Recovery Manager for VMware vSphere includes several software components that lower the cost and time required to manage and protect VMware vSphere environments:

- HP 3PAR Host Explorer for VMware vSphere – discovers VMware host configurations
- HP 3PAR VMware vSphere Storage APIs for Storage Awareness (VASA), allowing VMware vCenter insight into:
  - Provisioning Type
  - Volume Type
  - Drive Type
  - RAID Type
  - Remote Copy
- HP 3PAR Management Plug-in for VMware vCenter – displays virtual volume mapping for easy identification of HP 3PAR volumes used by virtual machines and datastores

The following features of Virtual Copy are key to making HP 3PAR Recovery Manager for VMware vSphere a superior snapshot management product:

- Non-duplicative snapshots reduce the capacity required for disk-to-disk (D2D) backups. When a production volume is changed, a single copy-on-write operation is performed, and little capacity is consumed regardless of the number of snapshots associated with the production volume.
- Reservationless snapshots reduce management overhead and the wasted capacity introduced by snapshot reservations in other technologies.
- Read-write snapshots can be mounted directly by the hosts and used for processing. This extends the benefit of snapshots to their use in test or development environments. Traditional read-only snapshots can be read but not mounted for processing.

Architecture

HP 3PAR Recovery Manager for VMware vSphere utilizes HP 3PAR Virtual Copy technology to create a thin snapshot of the selected data, with copy-on-write de-duplication performed to reduce the size of the thin snapshot. HP 3PAR Recovery Manager for VMware vSphere does not require VM operation to be suspended prior to the creation of the snapshot, allowing for the creation and restoration of hundreds of VM snapshots, with no effect on application performance or availability.

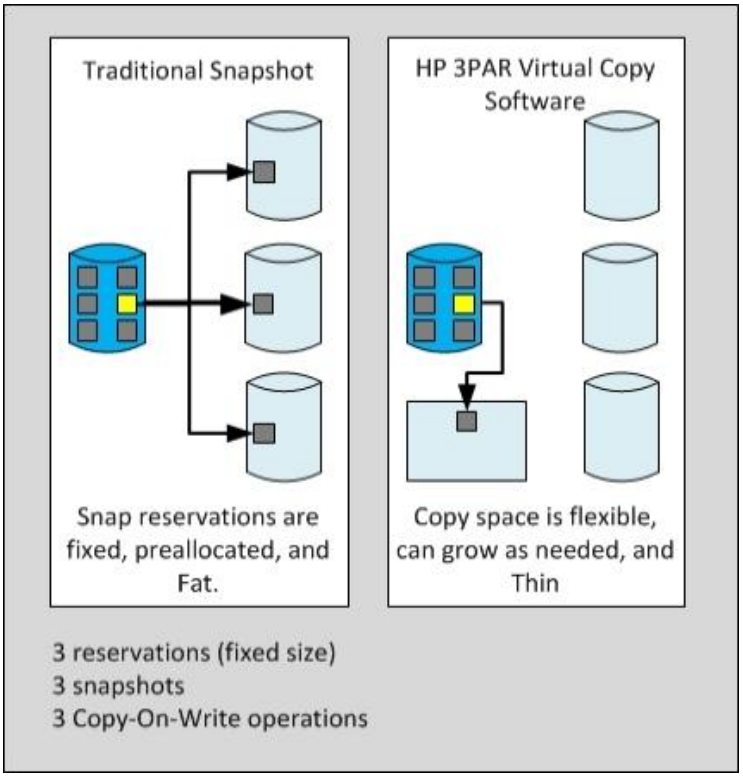
Once implemented, the snapshot process becomes an automated activity, allowing for quick and granular recovery. The data protection and recovery includes:

- Individual VM disks (VMDKs) including any included directories
- Individual files
- Entire virtual data store, comprising VMware vStorage VMFS

HP 3PAR Virtual Copy Software uses a unique system of pointers to increase performance, reduce disk capacity requirements, and enable read-write snapshots that can be mounted for rapid recovery or used in test and development environments.

As shown in Figure 23 , the copy-on-write I/O overhead grows with a traditional array volume as each snapshot requires its own copy of the changed data. I/O is multiplied by every new snapshot created and free capacity is reduced by the storage required for each snapshot reservation. Conversely, with Virtual Copy, the copy-on-write penalty is minimized by a single copy required for all Virtual Copy snapshots of a given volume. I/O overhead is minimized, and there is no wasted capacity consumed by snap reservation space.

Figure 23. Illustration of a traditional snapshot vs. an HP 3PAR snapshot.



Benefits

No more backup window

HP 3PAR Virtual Copy Software snapshots eliminate the need for a backup window by integrating with VMware vSphere to create an instant, non-disruptive point-in-time snapshot.

Maintaining multiple recovery points

Not only can HP 3PAR Recovery Manager for VMware vSphere eliminate the problem of a shrinking backup window, it also enables administrators to maintain multiple recovery points throughout the day.

Improved hot backup

Hot backup capabilities can be used without a snapshot-based backup. The downside of backing up in this fashion is that the applications in "hot backup" mode can have either a processing impact due to the overhead of transaction logging, a disk capacity impact due to long term storage of the state of the disk, or both. However, by integrating Virtual Copy snapshots with VMware vSphere, the duration of time during which the application is quiesced is greatly reduced, and so is the potential impact to performance and capacity consumption. HP 3PAR Recovery Manager for VMware vSphere makes hot backup better, providing the application with multiple points of recovery via high-performance, non-duplicative snapshots.

Low-impact topology

Where traditional backups require added SAN or LAN capabilities in order to handle increased bandwidth at multiple touch points in the data center, with HP 3PAR Recovery Manager for VMware vSphere, snapshots are presented to the backup host directly by the array. Instead of increasing the SAN and LAN bandwidth for every host, HP 3PAR Recovery Manager for VMware vSphere requires only a single backup server. This reduces the CPU and SAN traffic utilization on the production servers.

Flexible RPO and RTO

HP 3PAR Recovery Manager for VMware vSphere's scalable, non-duplicative snapshots allow the administrator to specify a greater number of snapshots for a frequent or extended history of recovery points, giving flexibility when committing to a given Recovery Point Objective (RPO). Flexibility to restore from snapshot or to mount the snapshot directly on the host instantly allows administrators to offer increasingly aggressive Recovery Time Objective (RTO) service level agreements to their internal customers.

Thin snapshot technology reduces sprawl

Non-duplicative snapshots also reduce the redundant copies of data that result from traditional D2D backups without the complexity of additional data deduplication appliances.

Usage and best practices

When faced with needing to achieve the fastest possible Recovery Time Objective (RTO), there are several items that should be taken into account and addressed when recovering from a failure. During a failure, rollback, or corruption of a VM or datastore, the fastest RTO can be met by mounting a Virtual Copy snapshot directly to the original host, replacing the original data volumes. The administrator can operate the application on this snapshot indefinitely, allowing the flexibility to choose a planned outage at a later date in order to promote the snapshot data back into the primary volume.

The ability to accomplish this task is integrated into HP 3PAR Recovery Manager for VMware vSphere. Right-clicking on a dataset presents the "mount" option, simplifying the administrator's duties during the critical time period while the application is down. The later promotion of the snapshot back into the primary volume is a simple step and is also integrated into HP 3PAR Recovery Manager for VMware vSphere.

Because of the ability of Virtual Copy to maintain high-performance read-write snapshots with no special configuration, the ability to promote any given snapshot back into the primary volume can be completed with only a few clicks. Promotion is an internal, array-driven operation that pushes the differences of the altered read-write snapshot back into the original volume. The primary volume is then remounted and the application restarted. The array-based copy of data from the snapshot back to the original volume is referred to as a "Promote." Promotion is performed inside the array for the highest performance and lowest network impact.

Another option that is available with HP 3PAR Recovery Manager for VMware vSphere is rapid recovery from snapshot. This recovery method involves mounting the snapshot and then copying the data to the production volume in a D2D operation. This process is automated via HP 3PAR Recovery Manager for VMware vSphere, integrating with VMware vSphere for a supportable restoration of data. This process takes longer than the first recovery option specified, but does not require planned downtime at a later date.

In addition to proper usage of HP 3PAR Recovery Manager for VMware vSphere, the following best practices should also be observed:

- HP 3PAR Recovery Manager for VMware vSphere should be installed on a dedicated Microsoft® Windows® server residing in the network, so that it may interface with the vSphere clusters and the HP 3PAR Storage.

- HP 3PAR Recovery Manager for VMware vSphere provisioning functions provide an easier way to create volumes from vCenter, but are not a replacement for provisioning using the HP 3PAR InForm Management Console (IMC). If provisioning requires specific attributes, the HP 3PAR IMC should be used.
- In the case of datastore restoration through the promotion of a virtual copy, HP 3PAR Recovery Manager for VMware vSphere does not check if a volume being unmounted is in use. Ensure that the volume is inactive prior to unmounting it in preparation for the restoration of the copy.
- If a VM or datastore is removed, the associated scheduled tasks continue to run but are no longer manageable from the HP 3PAR Management Plug-in and Recovery Manager for VMware vSphere scheduling interface. In this case, the task needs to be manually removed from the Windows scheduler.
- When copying or cloning a virtual machine (VM) to a different datastore, the source VM's Universally Unique Identifier (UUID) is retained in the target VM. When using HP 3PAR Recovery Manager for VMware vSphere to show the virtual copies on the target VM, the virtual copies from the source VM will be displayed. To resolve this problem, the target VM's UUID should be changed by editing the VMX configuration file (.vmx) while the VM is powered-off. The line that needs to be edited or deleted is the uuid.bios="..." line.
- When using the "Copy to Datastore" option to copy a virtual disk to a datastore, it is recommended that the copy and paste functions of the datastore browser be used in order to preserve the VMware Thin Provisioning disk feature.

Note
A virtual copy set may be retained for up to 1,825 days.

HP 3PAR Integration with VMware vCenter Site Recovery Manager (SRM)

Overview

VMware vCenter Site Recovery Manager is a management and automation product that helps build, manage, test and execute disaster recovery plans for a VMware virtual infrastructure. The HP 3PAR Storage System, as the storage component in a VMware virtual infrastructure, holds virtual machine information for a protected site/location and recovery site/location. HP 3PAR Storage Replication Adapter for VMware vCenter SRM 5 is an important integration component that communicates with HP 3PAR Storage Systems to execute specific storage and HP 3PAR Remote Copy functions needed for VMware vCenter Site Recovery Manager operation.

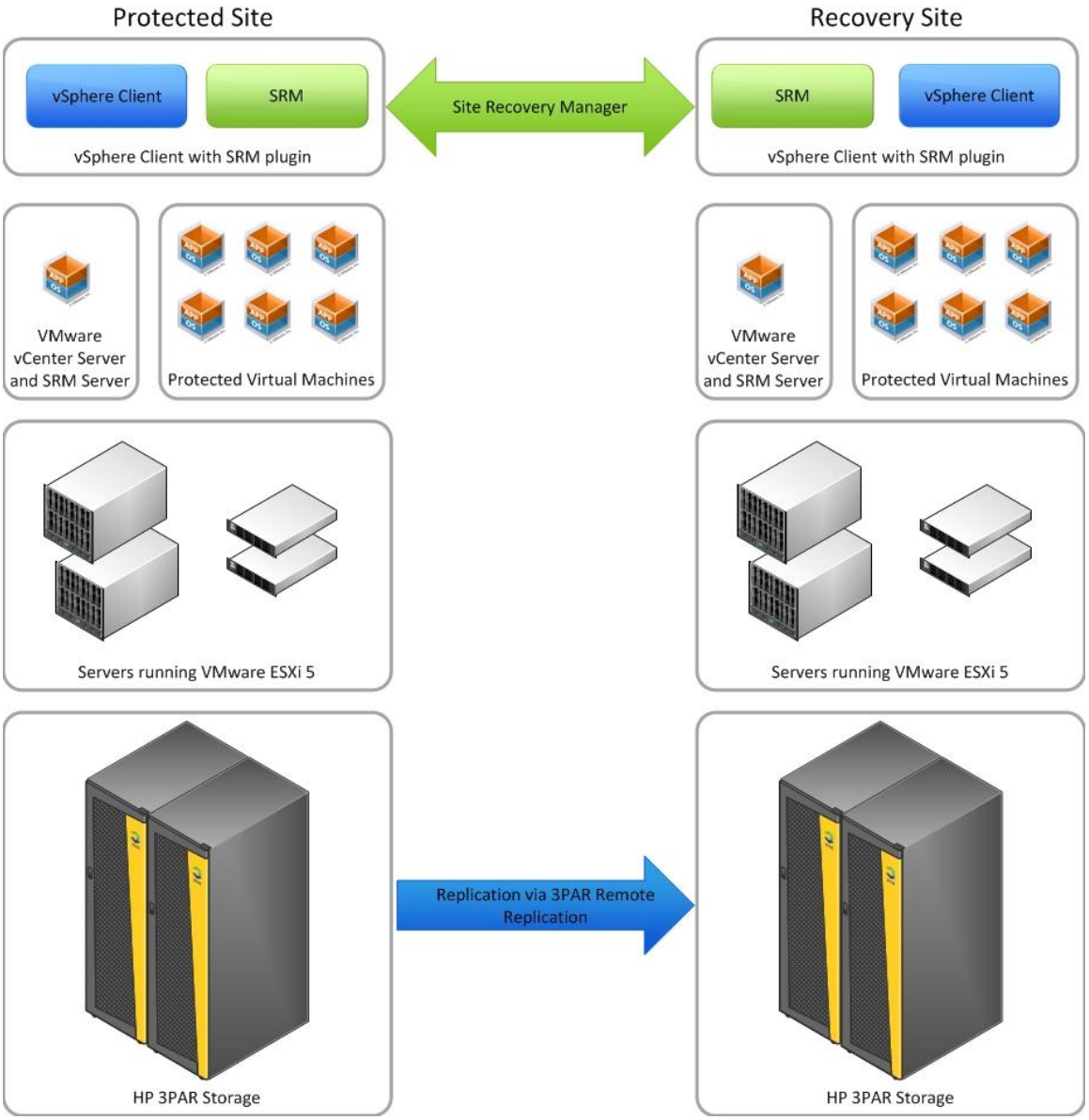
Architecture

VMware vCenter SRM provides business continuity and disaster recovery protection for VMware virtual environments. Protection can range from individual virtual machines (VMs) residing on a single, replicated datastore to all the VMs in a data center. VMware SRM helps IT administrators plan, test, and execute the recovery of virtual machines between the protected site and the recovery site.

As demonstrated in Figure 24, VMware SRM coordinates the recovery process with HP 3PAR Remote Copy Software to ensure that the virtual machines at the protected site are shut down cleanly (in the event that the protected site virtual machines are still available when recovery is invoked) so that the replicated virtual machines can be recovered and powered up at the recovery site. Recovery of protected virtual machines to the recovery site is guided by a recovery plan that specifies the order in which virtual machines are started up. The recovery plan also specifies network parameters, such as IP addresses, ensures the replicated storage holding the protected VMs is brought online and presented to the recovery hosts properly, and can contain user-specified scripts that can be executed to perform custom recovery actions.

After a recovery has been performed, the running virtual machines are no longer protected. To address this reduced protection, SRM supports a reprotect operation for virtual machines protected on array-based storage. The reprotect operation reverses the roles of the two sites after the original protected site is back up. The site that was formerly the recovery site becomes the protected site and the site that was formerly the protected site becomes the recovery site. At this point a planned migration back to the original site can be scheduled, if desired.

Figure 24: VMware SRM Deployment



For more information on Implementing HP 3PAR Remote Copy with VMware vCenter Site Recovery Manager reference the white paper located at http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-2382ENW

Best practices and caveats

In conjunction with the white paper referenced above, the following best practices should also be observed.

- If a single virtual machine sits on two datastores and two virtual volumes, it is recommended to include both virtual volumes in the same Remote Copy group.
- In a disaster recovery scenario, when Remote Copy links are down, the Remote Copy group status could still be *Started*. A failover attempt will not be successful until the Remote Group status becomes *Stopped*.
- If a virtual machine sits on a spanned datastore, all virtual volumes used for the spanned datastore need to be included in a single Remote Copy group.

- SRM might potentially time out if multiple test failover or recovery operations are run simultaneously. Re-run the operation if the time out issue occurs. Alternatively, the time out issue might be avoided if the operations are run sequentially.
- If an ESXi host has both FC and iSCSI definitions created on the HP 3PAR Storage System and vCenter Server also has both FC and iSCSI software adapters configured, per the vCenter Server's request, LUNs will be exposed to both host definitions in the event of failover. However, if only one host definition is presented on the HP 3PAR Storage System (either FC or iSCSI), HP 3PAR SRA will only expose LUNs to whichever is defined on the HP 3PAR Storage System.
- It is strongly recommended to configure one protected group per Remote Copy group.
- If multiple Remote Copy groups are included in one protected group, it is recommended to set the same sync time on all of the periodic Remote Copy groups.
- Remote Copy group reserves *.r* for naming. Do not include the reserved naming in your Remote Copy group name.
- *SRM_RO_<VVID>, SRM_RW_<VVID>, SRM_RECOVER_RO_<VVID>*, and *SRM_TARGETBK_RO_<VVID>* are reserved virtual volume naming conventions for HP 3PAR SRA.
- SRM might potentially run into a virtual volume promote operation during re-protect. If this occurs, retry the re-protect operation.
- HP 3PAR SRA can handle virtual volumes or a virtual volume set exposed to a host set. However, it does not support it. In the event of failover, LUNs will be exposed to the ESXi host(s) as individual LUNs and not a set. If the virtual volumes were exposed to the host set, LUNs will be exposed to an individual host as opposed to a host set.
- If virtual volumes are from a virtual volume set with multiple VMs created, be sure to include all virtual volumes in a single HP 3PAR Remote Copy group and in the same protection group. Otherwise there is a potential of losing connectivity to the VMs if virtual volumes are included in more than one HP 3PAR Remote Copy group and not all Remote Copy groups are included in the protection group.
- If virtual volumes are part of a virtual volume set, LUNs exposed to the ESXi host that are a part of that virtual volume set that do not have HP 3PAR Remote Copy configured and are not managed by SRM will lose the connectivity of the LUNs.
- HP 3PAR SRA can co-exist with a Synchronous Long Distance Remote Copy group with HP 3PAR OS version 3.1.1 or above.
- The same LUN ID should be used for every host in the same Access group. According to VMware's ESXi configuration guide, LUNs must be presented to each HBA of each host with the same LUN ID.
- Multiple Remote Copy groups in one protected group are not recommended. HP 3PAR SRA will log a warning to the user if multiple instances of such configurations are detected during the Test or Recovery operation since this might be an indication that VMs are using virtual volumes from different Remote Copy groups.
- The re-protect operation requires at least one Remote Copy link to be running for the intended storage pair to be up. If a recovery operation is performed as part of the failback process when all Remote Copy links are down, additional manual steps are required on the storage system before the user can re-protect the data:

    1. On the original protected storage system, issue the following command:
       showrcopy groups <RC group name>

       If the group role is not *Primary-Rev*, continue with SRM Reprotect ; otherwise, go to step 2.
    2. Issue the following command to change Remote Copy role:
       setrcopygroup reverse –local –current <RC group name>
    3. Continue with the SRM re-protect operation

- If a failover operation is unsuccessful, make sure to clean up the local disaster state cache. Otherwise, the subsequent SRM operations will fail.

  On both the protected and recovery sites where HP 3PAR SRA is installed, issue the following commands:

  tpdsrm viewstate

  tpdsrm cleanstate -sysid <StorageSystemID> -rcgroup <RCGroupName>

- If HP 3PAR SRA fails the QuerySyncStatus call due to the GetTaskStatus call returning the error, *Malformed InServ Data List: xxx*, perform the following steps to work around this problem:

    1. Get the task ID from the *Malformed InServ Data List: {0 {<TaskID> xxx* error.
    2. Log into the HP 3PAR Storage System at the protected site and issue the *showtask —d <TaskID>* command. Ensure that the task ID is valid.

3. Issue the *removetask —d <TaskID>* command to remove the task detail.
4. Re-run the SRM operation again. If you have multiple failed virtual volumes in the group, you may need to repeat steps 1 through 4 for all volumes.

Summary

Deploying HP 3PAR Storage in VMware vSphere environments helps remove the performance, efficiency, availability, and management headaches associated with traditional storage platforms. Not only does leveraging HP 3PAR Storage enable seamless integration with VMware vCenter to deliver enhanced performance, agility, and scalability in vSphere environments, but is also enables organizations using VMware vSphere with HP 3PAR several other key benefits:

- Tight integration of VMware vCenter Server and the HP 3PAR Recovery Manager Software for VMware vCenter allows administrators to monitor and manage HP 3PAR Storage volumes to create point-in-time, VM-and application-aware, disk-based snapshots from within the vSphere Client.
- Using HP 3PAR Storage with VMware vSphere enables its users to double virtual machine density on physical host servers as compared with traditional storage platforms.
- HP 3PAR Thin Provisioning Software allows physical storage to be consumed only when required for actual written data, rather than when allocated.
- With HP 3PAR Thin Persistence, as the ESXi host writes zeros to the VMDK file, the zeros are detected in-line by the 3PAR ASIC, and no space is allocated for the VMDK in the thin provisioned volume. Also, when a VM is deleted or moved to another data store, that now unallocated storage can be released back to the array rather than keeping it assigned to the LUN.
- HP 3PAR Adaptive Optimization Software can be used to tailor storage performance without disruption to VMware vSphere and contribute new autonomic space reclamation functionality.
- The combination of HP 3PAR Remote Copy and VMware vCenter Site Recovery Manager lets customers build resilient utility computing infrastructures, protect applications at a lower cost, and recover data more quickly and efficiently compared to traditional disaster recovery offerings.

These are among the unique advantages that make HP 3PAR Storage the ideal foundation for building or expanding a virtualized server environment with VMware vSphere as part of a converged infrastructure to meet the needs of the Instant-on Enterprise.

For more information

HP 3PAR Storage Family, hp.com/go/3PAR

HP/VMware technical papers

Adaptive optimization for HP 3PAR Storage,
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-0867ENW

Implementing HP 3PAR Remote Copy with VMware vCenter Site Recovery Manager,
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-2382ENW

Storage I/O Control Technical Overview and Considerations for Deployment,
vmware.com/files/pdf/techpaper/VMW-vSphere41-SIOC.pdf

VMware vSphere VAAI for HP 3PAR Storage performance benefits,
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-2864ENW

HP 3PAR VMware ESX Implementation Guide,
http://bizsupport1.austin.hp.com/bc/docs/support/SupportManual/c03290624/c03290624.pdf

VMware vSphere Storage DRS Interoperability,
vmware.com/files/pdf/techpaper/vsphere-storage-drs-interoperability.pdf

HP 3PAR documentation

HP 3PAR Storage Replication Adapter for VMware vCenter Site Recovery Manager 5.0 Implementation Guide,
http://bizsupport1.austin.hp.com/bc/docs/support/SupportManual/c03074202/c03074202.pdf

HP 3PAR Remote Copy Software User's Guide,
http://h20000.www2.hp.com/bizsupport/TechSupport/DocumentIndex.jsp?lang=en&cc=us&prodClassId=-1&contentType=SupportManual&prodTypeId=18964&prodSeriesId=5044771

HP 3PAR System Reporter Software overview,
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA3-0371ENW

HP Boot from SAN Configuration Guide,
http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c01861120/c01861120.pdf

HP 3PAR InForm OS 3.1.1 Concepts Guide,
http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c02986000/c02986000.pdf

HP 3PAR InForm OS 3.1.1 CLI Administrator's Manual,
http://bizsupport1.austin.hp.com/bc/docs/support/SupportManual/c02985998/c02985998.pdf

To help us improve our documents, please provide feedback at hp.com/solutions/feedback.

Get connected

hp.com/go/getconnected

Current HP driver, support, and security alerts
delivered directly to your desktop

4AA4-3286ENW, Created August 2012