Technical Report

VMware vSphere for Element Software Configuration Guide

Guide for NetApp HCI and SolidFire

Michael White and Daniel Elder, NetApp HCI
August 2020 | TR-4806

NetApp®

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

NetApp® Element is the underlying software for NetApp SolidFire® storage and NetApp HCI.

This document explains how to configure and connect a VMware vSphere host to a SolidFire iSCSI target using the native multipath plug-ins that are included with the vSphere software. It also provides design and implementation recommendations for disk alignment procedures, the network load balancing configuration, network redundancy, and performance control.

Note: This document assumes an intermediate to advanced level of VMware knowledge. You can simplify many functions presented here by using the Element vCenter Plug-In for VMware. Not all possible scenarios are explored in this document.

This guide covers multiple versions of VMware ESXi and conforms to the features available in ESXi version 6.7. Significant differences relative to prior ESXi releases are not noted. Earlier versions of ESXi might require a different workflow.

NetApp SolidFire Element software version 9 (Fluorine) or later, introduces support for the VMware Virtual Volumes (vVols) feature. See the VMware vSphere Virtual Volumes for SolidFire Storage Configuration Guide for information on configuring SolidFire vVols.

This document assumes that you have a basic ESXi configuration and a working SF-series storage cluster. For information on configuration and administration of an SF-series cluster, see the SolidFire and Element Documentation Center available on the SolidFire customer support portal. You can also use this document to manually configure NetApp HCI hosts or any third-party hosts connected to NetApp HCI.

The first three sections of this guide contain information on the remaining end-to-end configuration necessary for provisioning and consuming SolidFire storage in a vSphere cluster. The fourth section contains information on advanced tuning that might be required for high-performance workloads. You can use the following links to navigate to any of the sections:

- Prepare the vSphere Environment
- Storage System Environment Preparation
- Connect vSphere to SolidFire Storage
- VAAI Integration with SolidFire Storage
- Performance Control
- vSphere Storage Distributed Resource Scheduler and SolidFire

2 Prepare the vSphere Environment

This document assumes you are using the software iSCSI adapter provided by the VMware ESXi hypervisor. NetApp HCI does not support using independent hardware adapters for vSphere.

Note: Due to how ESXi hosts interpret valid 0x85 SCSI command responses, ‘state in doubt’ messages may be seen in the vmkernel.log files. NetApp recommends disabling VMware’s SMART daemon to prevent this from occurring. Refer to KB article: SolidFire: Turning off smartd on the ESXi hosts makes the cmd 0x85 and subsequent “state in doubt” messages stop.

2.1 Adding the Software iSCSI Adapter

The software iSCSI adapter is disabled by default; you must enable it before continuing with configuration. After it is enabled, the software iSCSI adapter cannot be disabled.

Note: If ESXi is configured to boot from iSCSI, then the iSCSI initiator is automatically enabled.

Prerequisites

- VMware vSphere Web Client version 6.7+. Previous versions that are EOL with VMware are not covered by this guide.

Procedure

1. In the VMware vSphere web client, open the ESXi host.
2. Click the Manage tab.
3. Click Storage.
4. Click Storage Adapters.
5. Click the Add () icon.
6. In the resulting dialog, click OK.

A new iSCSI software adapter group is created with an iSCSI adapter (typically labeled vmhba3x for ESXi 6.0 and earlier hosts and vmhba64 for ESXi 6.5 and later hosts).

2.2 Basic Networking Configuration

NetApp recommends using multiple network paths between a vSphere host and the SF-series cluster. With multiple paths, there are two options for multipathing: network-based or storage-based. Network-based multipathing makes the decision of which path to use at the network layer. Storage-based multipathing builds multiple distinct network connections between the host and the SolidFire cluster and uses the VMware storage implementation to choose the path. VMware recommends storage-based multipathing.

SolidFire Network Configuration

For the best performance of the SF-series nodes, NetApp recommends configuring the 10GbE or 10/25GbE network interfaces for LACP. For more information on configuring the 10GbE network for SF-series storage, see the SolidFire and Element Documentation Center.

Creating a vSwitch and Adding Uplinks

You should use a vSwitch or VMware vSphere Distributed Switch (VDS) with at least two configured physical network uplinks (vmnic) for redundancy. Although you can configure the iSCSI initiator with a single network uplink, this creates a single point of failure. A VDS is used in this example, but VMware standard switches (VSS) are also supported.

Prerequisites

- A vSphere vSwitch or a VDS
- One or more network connections between the ESXi host and SolidFire storage
- If you are using VDS, the switch and corresponding distributed port group must already be created.

Procedure

1. In the VMware vSphere web client, open the ESXi host.
2. Click the Manage tab.
3. Click Networking.
4. Click VMkernel Adapters.
5. Click the Add () icon.
6. Under Select Connection Types, select the VMkernel Network Adapter option and click Next.
7. Under Select Target Device, select an appropriate option and click Next.

8. Under IPv4 Settings, select the Use Static IPv4 Settings option and enter the following based on your environment:
   a. IPv4 address
   b. Subnet mask
9. Click Next.
10. Review the settings and click Finish.

    The configured VMkernel is shown in the VMkernel adapters list.

    Note: Without defining any explicit binding between the software iSCSI initiator and a VMkernel interface, the vSphere host is still able to initiate single-path iSCSI communication with the SolidFire cluster. It can do so if the destination address is accessible from one or more VMkernel interfaces. The actual vmknic used for storage traffic in this case is determined by the routing table of the ESXi host. The configured VMkernel is shown in the VMkernel adapters list.

11. Create another VMkernel adapter in the same subnet.

    NetApp strongly recommends using two VMkernel ports in the same subnet for iSCSI storage access. This document assumes that this is the case. If you miss this step, you will not have two paths to storage. It is important to understand that this must be on the same vSwitch because this cannot span vSwitches.

2.3 Jumbo Frames in vSphere Networking

For better efficiency, NetApp recommends that you enable jumbo frames on all devices between ESXi hosts and the SF-series cluster and nodes. SolidFire storage is configured to use an MTU size of 9000 bytes by default. Most Ethernet devices use an MTU size of 1500 bytes by default. When you enable jumbo frames, make sure that the setting is applied on the designated storage-facing interfaces on each ESXi host.

When connecting network infrastructure used for iSCSI, you should verify that the MTU size is set higher than the host MTU size to allow for the jumbo-frame size plus the Ethernet header. Typically, network infrastructure is configured for the MTU size such as 9214 bytes for Arista Networks devices or 9216 bytes for Cisco devices.

2.4 NIC Teaming and Load Balancing Policies

The NIC teaming policies are included in the configuration of each vSphere vSwitch and port group. NetApp recommends using a vSwitch or VDS with at least two 10GbE physical network uplinks (vmnic) for redundancy. In this example, a VDS is used.

Prerequisites

- vSphere vSwitch or VDS
- One or more network connections between the ESXi host and SolidFire storage

NetApp recommends keeping the default route-based load-balancing method on the originating virtual port ID for iSCSI traffic.

By default, port groups inherit settings from the parent vSwitch object, but you can override this behavior by individually selecting the appropriate options in the configuration wizard. This allows a unique load balancing and redundancy configuration for each port group or virtual interface.

Note: NetApp recommends using the route based on the originating virtual port ID policy for port groups used to facilitate iSCSI connections to SolidFire storage. Using the route based on the IP hash load-balancing policy for iSCSI traffic on switch interfaces that are configured for the Link Aggregation Control Protocol (LACP) is not recommended.

iSCSI Multipathing

iSCSI multipathing provides I/O load balancing and redundancy through the storage initiator stack using multipath I/O (MPIO), rather than the network communication stack using LACP. MPIO has several benefits over LACP for iSCSI traffic:

- Faster I/O path failure detection
- The use of multiple paths for storage I/O
- Less complicated network setup

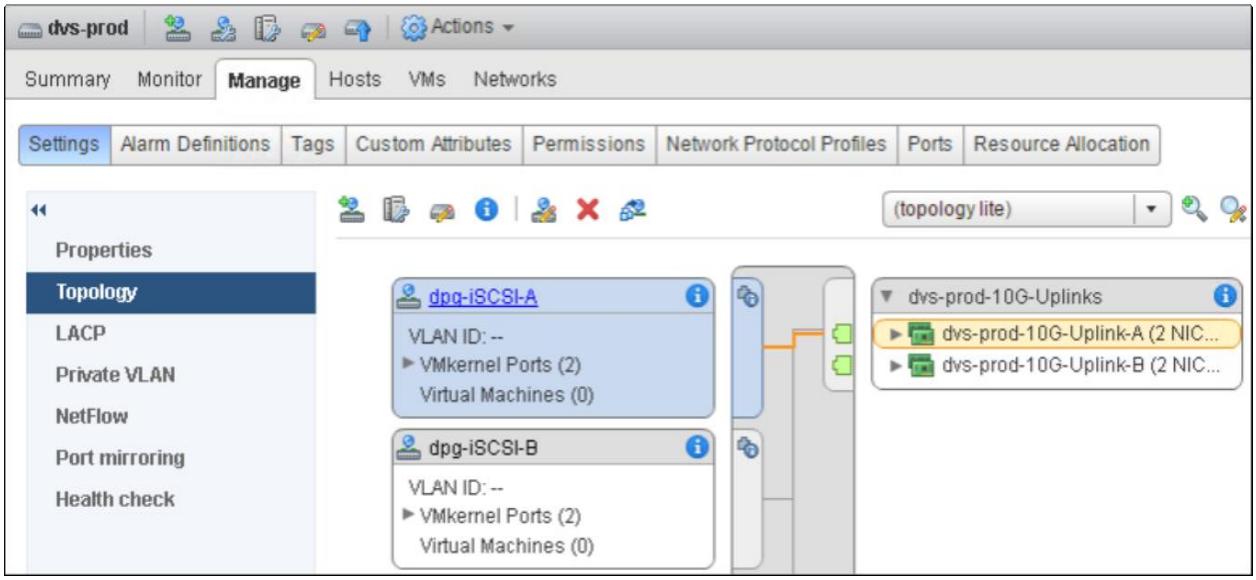
iSCSI multipathing has the following configuration requirements:

- The vSphere host must have two or more physical network interfaces.
- The vSphere host must have two or more VMkernel virtual interfaces in the same addressable network or subnet as the storage target.
- iSCSI initiator port binding must be configured.

You cannot use iSCSI initiator port binding in scenarios involving routed storage traffic between the initiator and target.

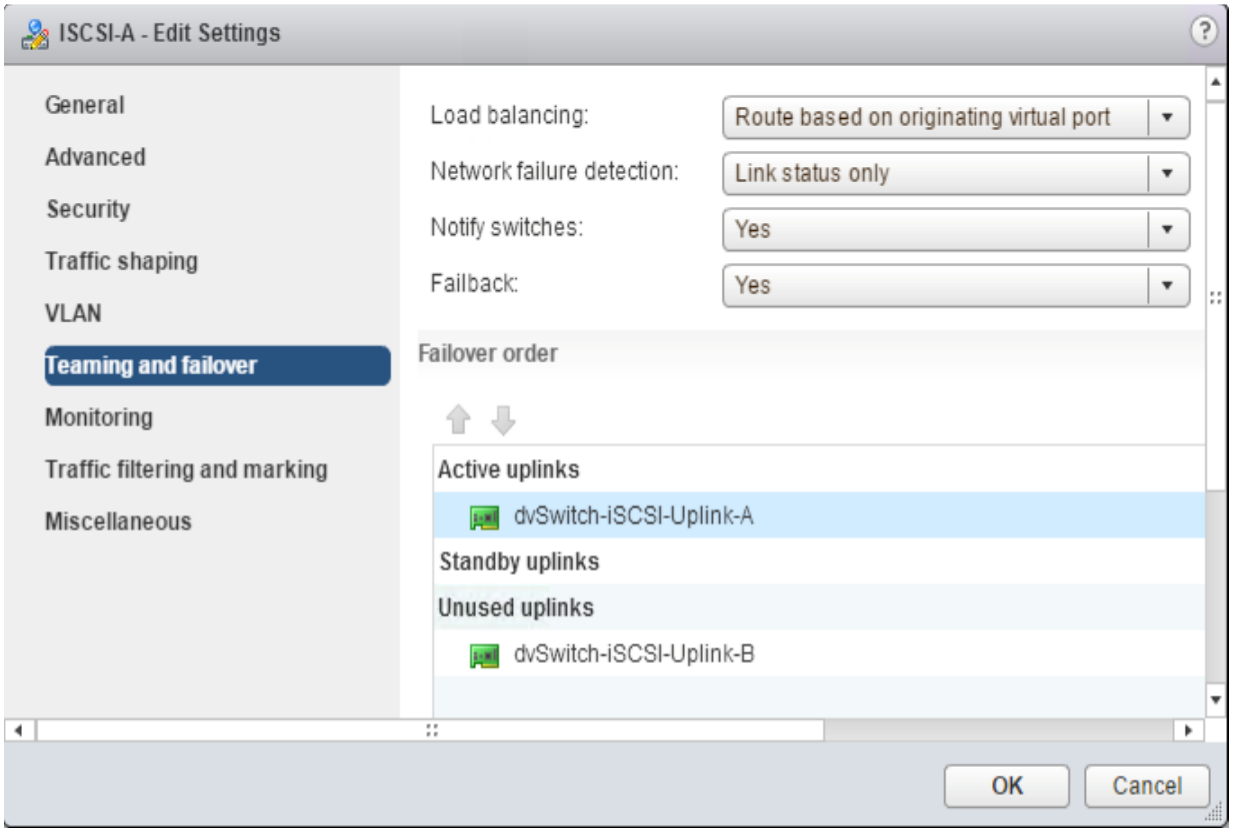
Configure a VMkernel interface for each physical network interface to be included in the multipathing configuration. By default, all uplinks are active for each port group. You should configure each port group used for iSCSI and its VMkernel interface to override the vSwitch physical interface failover order to configure a single active uplink per iSCSI port group. This is illustrated in Figure 1.

Figure 1) Configuring port groups.



To override the default failover order, you can edit the properties of the iSCSI port group and select the Teaming and Failover section as shown in Figure 2. You can configure a single active uplink for each iSCSI port group. All other uplinks must be in the Unused Uplinks list to be eligible for port binding. Standby is not adequate.

Figure 2) Override the default failover order.



You can configure more than two uplinks for this configuration pattern. In this scenario, each VMkernel interface uses a single, unique, active physical network adapter; all other physical adapters are marked as unused.

Table 1) VMkernel port group configuration.

VMkernel Port Group Interface	Active Uplink	Failover Order Unused Uplink
dpg-ISCSI-A	dvs-prod-10G-Uplink-A	dvs-prod-10G-Uplink-B
dpg-ISCSI-B	dvs-prod-10G-Uplink-B	dvs-prod-10G-Uplink-A

After adding two or more appropriately configured VMkernel interfaces, you must bind the interfaces to the iSCSI initiator. For more details, see the section “Binding VMkernel Interfaces to the iSCSI Adapter.”

Note: If you configure jumbo frames, test them to ensure that they are configured end-to-end before continuing by using the `vmkping` command: `vmkping <SolidFire SVIP> -I vmk2 -s 8972 -d`.

The 8972-byte MTU size is required because the ICMP/ping implementation does not encapsulate the 28-byte `ICMP (8) + TCP (20) (ping + standard transmission control protocol packet)` header. Therefore, for ping tests, it must be subtracted as follows:

9000 - 8 - 20 = 8972

Binding VMkernel Interfaces to the iSCSI Adapter

After adding two or more appropriately configured VMkernel interfaces, you must bind the interfaces to the iSCSI adapter.

Prerequisites

- VMkernel interfaces bound to vSwitches or VDS. For more details, see the section “Creating a vSwitch and Adding Uplinks.”
- vSphere iSCSI software initiator enabled and configured. For more details, see the section “Adding the Software iSCSI Adapter.”

Procedure

1. In the VMware vSphere web client, select the ESXi host.
2. Click the Manage tab.
3. Click Storage.
4. Click Storage Adapters.
5. Select the adapter under iSCSI Software Adapter.

   Note: This adapter is often assigned the name vmhba3x or vmhba64.
6. Click the Network Port Binding tab.
7. Click the () icon. A VMkernel port binding is added for each VMkernel interface previously configured for iSCSI multipathing.

Best Practices

Port binding is used when multiple VMkernel ports for iSCSI reside in the same broadcast domain. Port binding has the following requirements:

- Cluster target iSCSI ports must reside in the same broadcast domain and IP subnet as the VMkernel ports.
- All VMkernel ports used for iSCSI connectivity must reside in the same vSwitch.
- Currently, port binding does not support network routing.

If any of these four criteria are not met, one of three things happen when configuring port binding (messages and menus can vary depending on the ESXi version):

- When you attempt to add network adapters under Storage Adapters > iSCSI Software Adapter > Properties > Network Configuration, this error appears:

  ```
  Only VMkernel adapters compatible with the iSCSI port binding
  requirements and available physical adapters are listed.
  ```

  If a targeted VMkernel adapter is not listed, you can go to Host > Configuration > Networking to update the failover policy for the VMkernel interface.
- Selecting the VMNIC for binding causes this message to appear:

  ```
  The selected physical network adapter is not associated with VMkernel
  with compliant teaming and failover policy. VMkernel network adapter
  must have exactly one active uplink and no standby uplinks to be
  eligible for binding to the iSCSI HBA.
  ```
- On the iSCSI Adapter Properties page under Network Configuration, the Port Group Policy on the VMkernel Port Bindings shows as Non-Compliant.

3 Storage System Environment Preparation

Following the successful installation of a SolidFire storage system, you must configure the following elements to provision storage to a vSphere cluster:

- At least one tenant account (typically one per vSphere cluster)
- Volumes
- Volume access groups

Note: This guide covers requirements for provisioning virtual machine file system (VMFS) volumes for ESXi hosts. For information about configuring vSphere vVols for use with SolidFire storage systems, see the VMware vVols Configuration Guide available mysupport.netapp.com. The SolidFire storage system fully supports simultaneous use of vVols and traditional VMFS storage.

3.1 Storage Provisioning

You must configure and provision the SolidFire storage system for access in the vSphere environment so that you can access iSCSI volumes on the SolidFire storage system. To do this, assign SolidFire volumes to accounts and optionally volume access groups.

Creating a SolidFire Account

Each SolidFire account represents a unique volume owner and receives its own set of Challenge-Handshake Authentication Protocol (CHAP) credentials. You can access volumes assigned to an account either using the account name and the relative CHAP credentials or through a volume access group.

Optionally, you can use the VMware vSphere web client configured with the SolidFire Plug-In for the VMware vCenter Server client to complete this step. For more details, see the Element Plug-In for vCenter Server User Guide.

Procedure

1. Log into the Element web UI.
2. Go to Management > Accounts.
3. Click Create Account.
4. Enter the CHAP user name to be used with the ESXi host in the Username field.
5. In the CHAP Settings section, complete the following fields:
   a. Enter the initiator secret for CHAP node session authentication.
   b. Enter the target secret for CHAP node session authentication.

   Note: Leave these fields blank to auto-generate the secrets. To view them, click Actions > View Details.
6. Click Create Account.

Creating a SolidFire Volume

After provisioning an account, you must create volumes and associate them with the account. This enables iSCSI initiators using the provided CHAP credentials to discover and mount iSCSI devices that are associated with that account in Element OS. In the case of volumes that are connected via Fibre Channel (FC), the account serves as a container for the volumes providing per-account statistics and segmentation.

Note: The ESXi hypervisor requires 512-byte emulation. 4k-device block sizes are not supported by ESXi.

Optionally, you can use the VMware vSphere web client configured with the SolidFire Plug-In for the VMware vCenter Server client to complete this step. For more details, see the SolidFire Plug-In for VMware vCenter Server Client User Guide (https://docs.netapp.com/sfe-115/topic/com.netapp.doc.sfe-mg-vcp/home.html).

1. Go to Management > Volumes.
2. Click Create Volume.
3. In the Create a New Volume dialog, enter the Volume Name (can be 1 to 64 characters in length).
4. Enter the total size of the volume.

   Note: The default volume size selection is in GB. Volumes can be created with GB or GiB:

   - 1 GB = 1 000 000 000 bytes
   - 1 GiB = 1 073 741 824 bytes
5. Select a block size for the volume.

   This option is necessary to support operating systems and applications that do not recognize native 4k drives such as VMware ESXi.
6. Click Account and select the account from the list that should have access to the volume. If an account does not exist, click the Create Account link, enter a new account name, and click Create. The account is created and associated with the new volume.

   Note: If there are more than 50 names, the list does not appear. Begin typing and the auto-complete function displays possible values for you to choose from.
7. Set the quality-of-service values or accept the default values.

   Use the spin box in each column to select the desired IOPS values.

Caution: Volumes with MaxIOPS and BurstIOPS greater than 15,000 are allowed to accommodate higher bandwidths. Achieving greater than 15,000 small-block IOPS on a single volume requires a high queue depth and might require special MPIO configuration. For details, see the NetApp Element Release Notes.

8. Click Create Volume.

Creating an Access Group

Volumes can be discovered through volume access groups that allow ESXi initiators access to volumes without requiring CHAP authentication. Volume access groups provide a simple method for mapping a set of storage system volumes to a set of ESXi host initiators.

You can also use the vSphere web client configured with the SolidFire Plug-In for VMware vCenter Server to complete this task. For details about using the SolidFire Plug-In for VMware, see the NetApp Element Plug-In for vCenter Server User Guide.

When creating a volume access group, note the following guidelines:

- A maximum of 2000 volumes can be in a volume access group.
- An IQN or WWPN can belong to only one volume access group.
- A single volume can belong to a maximum of four volume access groups.
- Do not use volume access groups for boot from iSCSI volumes.

Prerequisites

Access to the vSphere web client and the SolidFire Element UI or vCenter plug-in.

Procedure

1. In the vSphere web client, determine the host IQN:
   a. In the ESXi host web interface, click Manage.
   b. Click Storage.
   c. In the left pane, click Storage Adapters.
   d. Select the adapter under iSCSI Software Adapter.

   Note: This adapter is often assigned the name vmhba3x or vmhba64.

   e. Under the Adapter Details section, click the Properties tab.
   f. Under the General section, take note of the iSCSI name (which is the host IQN).
2. Log in to the SolidFire Element web UI.
3. Go to Management > Access Groups.
4. Click Create Access Group.
5. Enter a name for the volume access group in the Name field.
6. To add an iSCSI initiator to the volume access group, complete the following steps:
   a. Under Add Initiators, select an existing initiator from the Initiators list.
   b. Click Add Initiator.

   Note: You can create an initiator during this step by clicking the Create Initiator link, entering an initiator name, and clicking Create. The system automatically adds the initiator to the Initiators list after you create it.

   The accepted format of an initiator IQN is `iqn.yyyy-mm` where `y` and `m` are digits, followed by text which must only contain digits, lower-case alphabetic characters, a period (.), a colon (:), or a dash (-).

   See the following formatted example:

```
iqn.2010-01.com.solidfire:c2r9.fc0.2100000e1e09bb8b
```

   TIP: You can find the initiator IQN for each volume by selecting View Details in the Actions menu for the volume on the Management > Volumes > Active list.
7. (Optional) Repeat step 6 to add more initiators as needed.
8. Under Attach Volumes, select a volume from the Volumes list and click Attach Volume.
9. (Optional) Repeat step 8 to attach more volumes as needed.
10. Click Create Access Group.

Adding Volumes to an Access Group

You can add volumes to a volume access group. Each volume can belong to more than one volume access group, and you can see the groups that each volume belongs to in the Active volumes area.

Procedure

1. Log in to the SolidFire Element OS Web UI.
2. Go to Management > Access Groups.
3. Choose an access group and click the Actions button.
4. In the resulting menu, click the Edit button.
5. Under Add Volumes, select a volume from the Volumes list.
6. Click Attach Volume.
7. Repeat steps 5 and 6 to add more volumes as needed.

8. Click Save Changes.

Adding Multiple Volumes to an Access Group

You can add multiple volumes to a volume access group. Use this process when you want to bulk add one or more volumes to a volume access group.

Note: You can also use this procedure to add volumes to an FC volume access group.

Procedure

1. Log in to the SolidFire Element OS Web UI.
2. Go to Management > Volumes.
3. In the list of volumes, check the box next to any volumes you would like to add.
4. Click Bulk Actions.
5. In the resulting menu, click Add to Volumes Access Group.
6. Select the access group to which you want to add the volumes.
7. Click Add.

4 Connect vSphere to SolidFire Storage

After you have configured both the vSphere and SolidFire environments, you can create vSphere datastores.

4.1 Storage Adapter Configuration

After one or more volumes and associated accounts has been provisioned from Element, you need to add the storage targets to the iSCSI adapter on the ESXi host. You can either add iSCSI targets dynamically (`SENDTARGETS`), or manually using an iSCSI static discovery target. Authentication is handled either through SolidFire volume access groups (recommended) or through CHAP credentials.

Configuring Dynamic Discovery (SENDTARGETS)

When you configure dynamic discovery, ESXi hosts become dynamically aware of new SolidFire volumes added to an access group or volumes created under a tenant account with the correct CHAP credentials.

Note: vSphere does not permit configuration with multiple CHAP credentials on a single dynamic-discovery target address. If you configure multiple tenant account volumes from the initiator on the same SolidFire Storage Virtual IP (SVIP), you must also configure static discovery targets. For more details, see the section “Configuring Static Discovery.”

Prerequisites

- SVIP
- CHAP credentials

Procedure

1. In the VMware vSphere Web Client, open the ESXi host.
2. Click the Manage tab.
3. Click Storage.
4. Click Storage Adapters.
5. Select the adapter under iSCSI Software Adapter.

   Note: This adapter is often assigned the name vmhba3x or vmhba64.

6. Under Adapter Details, click the Targets tab.
7. Click Dynamic Discovery.
8. Click Add.
9. In the iSCSI Server field, enter the SVIP.
10. Under Authentication Settings, clear the Inherit Settings from Parent check box.
11. If you are using volume access groups, select None for Authentication Method.
12. (Optional) If you are using one-way CHAP authentication, complete the following steps:
    a. Select Use Unidirectional CHAP from the Authentication Method drop-down list.
    b. Under the Outgoing CHAP Credentials section, enter into the Name field the CHAP user name for the tenant account previously created in the section “Creating a SolidFire Account.”
    c. In the Secret field, enter the initiator secret created previously in the section “Creating a SolidFire Account.”

    Note: You can retrieve these values from the Element web UI under the Management > Accounts page by selecting View Details for the individual account.
13. (Optional) If you are using mutual CHAP authentication, complete the following steps:
    a. Select Use Bidirectional CHAP from the Authentication Method drop-down list.
    b. Under the Outgoing CHAP Credentials section, enter into the Name field the CHAP user name for the tenant account previously created in the section “Creating a SolidFire Account.”
    c. In the Secret field, enter the initiator secret previously created in the section “Creating a SolidFire Account.”
    d. Under the Incoming CHAP Credentials section, enter the same CHAP username used for outgoing CHAP authentication.
    e. In the Secret field, enter the target secret previously created in the section “Creating a SolidFire Account.”
14. Click OK.

Configuring Static Discovery

You can use static discovery to configure multiple volumes associated with multiple tenant accounts. Unlike volume access groups, static discovery requires that you configure each target individually.

Prerequisites

- SVIP
- CHAP credentials

Procedure

1. In the VMware vSphere web client, open the ESXi host.
2. Click the Manage tab.
3. Click Storage.
4. Click Storage Adapters.
5. Select the adapter under iSCSI Software Adapter.

    Note: This adapter is often assigned the name vmhba3x or vmhba64.
6. Under Adapter Details, click the Targets tab.
7. Click Static Discovery.
8. Click Add.

9. In the iSCSI Server field, enter the SVIP.
10. In the iSCSI Target Name field, enter the volume IQN.

    TIP: You can retrieve the IQN value from the Element web UI by selecting the View Details option.
11. Under Authentication Settings, clear the Inherit Settings from Parent check box.
12. (Optional) Complete the following steps if you are using one-way CHAP authentication:
    a. Select Use Unidirectional CHAP from the Authentication Method drop-down list.
    b. Under the Outgoing CHAP Credentials section, enter into the Name field the CHAP user name for the tenant account previously created in the section “Creating a SolidFire Account.”
    c. In the Secret field, enter the SolidFire initiator secret created previously in the section “Creating a SolidFire Account.”

    Note: You can retrieve these values from the SolidFire Element web UI under the Management > Accounts page by selecting View Details for the individual account.
13. (Optional) If you are using mutual CHAP authentication, complete the following steps:
    a. Select Use Bidirectional CHAP from the Authentication Method drop-down list.
    b. Under the Outgoing CHAP Credentials section, enter into the Name field the CHAP user name for the tenant account previously created in the section “Creating a SolidFire Account.”
    – In the Secret field, enter the initiator secret previously created in the section “Creating a SolidFire Account.”
    c. Under the Incoming CHAP Credentials section, enter the same CHAP user name used for outgoing CHAP authentication.
    d. In the Secret field, enter the target secret previously created in the section “Creating a SolidFire Account.”
14. Click OK.

4.2 Mounting SolidFire Block Storage Volumes in ESXi

After you have provisioned volumes and accounts on the storage cluster, configured the network, and created iSCSI targets on the ESXi host, volumes are available to be mounted as devices on the ESXi host.

Note: Optionally, you can use the VMware vSphere web client configured with the SolidFire Plug-In for the VMware vCenter Server client to complete this step. For more details, refer to the Element Plug-In for vCenter Server User Guide.

If the vSphere setting `config.vpxd.filter.hostRescanFilter` is set to False, manual host scanning is required. For more details, see VMware KB 1016873.

Note: ESXi hosts booting from iSCSI should not use volume access groups for boot volumes.

Prerequisites

- SolidFire volumes visible to the ESXi hosts

Procedure

1. In the VMware vSphere web client, open the ESXi host.
2. Select the Host and Clusters tab.
3. Click Manage.
4. Select Storage.
5. Select Datastores.
6. Click the Add Datastore () icon.

7. Select VMFS under Type.
8. Click Next.
9. In the Datastore name field, enter a datastore name.
10. Select a storage device from the list and click Next.
11. Select a partition layout from the Partition Configuration list and click Next.

    Note: This is typically Use All Available Partitions. Set the datastore size to use the entire volume.

12. Review the datastore information and click Finish.

4.3 Virtual Disk Provisioning on a SolidFire Storage System

When using a VMFS datastore partition to provision virtual machine disks (VMDKs) on a SolidFire storage device, there are several options that you can consider to provide the best performance and efficiency. You can choose one of the following options for provisioning virtual disks on a VMFS datastore:

- **Thick-provision lazy zeroed (LZT - vSphere default).** Zeroes ahead in 1M I/O prior to writing data to the volume.
- **Thick-provision eager zeroed (EZT).** Zeros the entire virtual disk prior to creating the virtual machine.
- **Thin-provision.** Provides no advanced zeroing of the volume. Required for automatic guest OS space reclamation. UNMAP passthrough requires the advanced ESXi setting `/VMFS3/EnableBlockDelete` to be set to 1. See the section "VAAI Integration with SolidFire Storage" for more information.

  Note: Zero space in thin provisioned and LZT disks does not consume appreciable block or metadata capacity on SolidFire clusters. Zero space in EZT disks does not consume block capacity on SolidFire clusters. However, it does consume the amount of metadata required to completely allocate the disk, including zero space. By contrast, thin provisioning requires more metadata operations per write, which can affect performance and count against your IOPS credit.

vSCSI Controller Types

Selecting the guest OS type for the virtual machine automatically selects the vSCSI controller based on best fit and driver selection for the guest OS distribution. There are four possible vSCSI controller options:

- **BusLogic.** Legacy emulated vSCSI controllers used for older versions of Windows. Queue depth with this controller is limited to 1 and is not recommended for use in modern operating systems.
- **LSI Logic Parallel (previously LSI Logic).** Emulated vSCSI controller supporting a queue depth of 32 I/O operations.
- **LSI Logic SAS.** Default for Windows 2008 and newer. Required for Microsoft Cluster Service (MCSC) within Windows 2008 and later releases.
- **VMware Paravirtual (PVSCSI).** The highest performing vSCSI controller with a default adapter queue depth of 254 I/O operations and a per-device queue depth of 64. The PVSCSI controller is designed to support very high throughput with minimal processing cost. VMware KB article 2053145 includes instructions on increasing the per-device queue depth to 254 I/O operations for high performance requirements.

For more information about PVSCSI controller compatibility, see VMware KB article 1010398.

Potentially Misaligned Operating Systems

If the operating system you are installing is supported and not on the following list, then it aligns correctly by default during installation. If you would like to install an operating system in the following list on an

ESXi host using a SolidFire volume, you might need to speak with support about alignment prior to installation.

The following is a list of operating systems that, by default, might have an unaligned file:

- Red Hat Enterprise Linux 5.X
- CentOS 5.X
- Ubuntu 10.X
- SUSE Linux Enterprise 11.X
- Debian 5.X
- Microsoft Windows 2003

4.4 Multipathing Options

The storage path selection policy (PSP) determines which path data takes from the host initiator to storage. By default, the SolidFire system is discovered as an active:active array with the generic `VMW_SATP_DEFAULT_AA` storage array type plug-in (SATP). However, for ESXi hosts with multiple initiators (VMkernel ports) configured for VMkernel port binding, you must change this for optimal performance.

Configuring Multipathing Using the GUI

The PSP determines which path data takes from the host initiator to storage.

Prerequisites

- SolidFire volumes that have been configured and added to vSphere
- Configured iSCSI port binding
- Multiple paths available to targets

Procedure

1. In the VMware vSphere web client, open the ESXi host.
2. Open the Storage tab within the Navigator.
3. Select a datastore.
4. Click the Manage tab.
5. Click Settings.
6. In the navigation pane, click Connectivity and Multipathing.
7. Select a target host.
8. Click Edit Multipathing.
9. From the Path selection policy list, select Round Robin (VMware). Round robin enables ESXi to send frames as quickly and evenly as possible to SolidFire, thereby optimizing performance for the majority of instances when using a SolidFire cluster.

   Note: The GUI method of selecting the round-robin PSP does not allow configuration of the optional IOOperations Limit, which defaults to 1000. This means that by default, ESXi sends 1000 I/O operations to one path before switching to the next available path. This does not optimally distribute I/O to all paths. You must use the CLI to configure the round-robin IOOperations Limit value.

10. Click OK.

Multipathing Configuration CLI Example

You can create a custom storage array type plug-in (SATP) rule on the ESXi hosts to automatically claim SolidFire storage devices. This step applies the desired configuration to new SolidFire volumes as they are discovered. The rule should be applied to all hosts in the cluster before any SolidFire volumes are presented to the ESXi hosts. Storage added prior to creating a custom SATP is not claimed until the ESXi host is rebooted.

The following example creates a custom SATP rule named SolidFire Customer SATP Rule that configures every new SF-series device with the round-robin (`VMW_PSP_RR`) path selection policy (PSP) and changes the IOPS value to 1. NetApp recommends an IOPS value of 1 to provide better I/O balancing versus the default 1000 IOPS per path.

```
[root@vdi1:~] esxcli storage nmp satp rule add -s VMW_SATP_DEFAULT_AA -P VMW_PSP_RR -O iops="1" -
V "SolidFir" -M "SSD SAN" -e "SolidFire custom SATP rule"
```

You can verify creation of the new rule using the following command:

```
esxcli storage nmp satp rule list | grep -i -E 'name|solid'
```

This command gives the following output:

```
~ # esxcli storage nmp satp rule list | grep -i -E 'name|solid'
Name                 Device  Vendor    Model            Driver  Transport  Options                    Rule
Group  Claim Options                        Default PSP  PSP Options  Description

VMW_SATP_DEFAULT_AA          SolidFir  SSD SAN                                                        user
                                       VMW_PSP_RR   iops=10      SolidFire Custom SATP
```

Note: Configuring a custom SATP rule works for both iSCSI and FC-based storage.

5 VAAI Integration with SolidFire Storage

VMware vSphere Storage APIs Array Integration (VAAI) for vSphere was introduced to offload storage-related tasks from ESXi to storage vendor products. This allows for greater performance and efficiency for many common tasks.

SolidFire Element OS version 6 and newer fully support all current VAAI block primitives. In vSphere 5.5+, VAAI primitives are divided into the following groups:

- **Atomic Test & Set (ATS).** Allows for VM block-level locking instead of SolidFire volume-level locking. This configuration allows for greater performance in high utilization environments. Often referred to as Hardware-Assisted Locking.
- **Hardware Accelerated Move/Copy (XCOPY).** Allows for the offload of copy operations from the ESXi host to the SolidFire cluster internally, significantly reducing storage bandwidth and the time needed to complete operations.
- **Block Zero Offload (WRITESAME).** Enables the SolidFire cluster to zero out a large number of blocks through a VAAI command.
- **Thin provisioning.** Enables VMware to interact directly with SolidFire thin-provisioning functionality.
- **Block reclaim (UNMAP).** Enables the SolidFire cluster to reclaim thin-provisioned blocks more easily after deletion through VMware.

6 Performance Control

Performance Control is imperative for vSphere environments. Providing predictable performance from hundreds to thousands of virtual machines across a vSphere environment is critical.

There are three solutions for performance control within a VMware and SolidFire environment:

- SolidFire QoS
- VMware vSphere Storage I/O Control (SIOC; requires vSphere Enterprise Plus)
- SolidFire VMware vCenter Plug-In for QoS and SIOC integration
- VMware vVols (more information in this Technical Report).

The SolidFire QoS solution delivers guaranteed performance on a per-volume basis. This control is delivered on a min, max, or burst basis to each volume on the SolidFire cluster. In addition, SolidFire QoS considers the varying loads that are associated with varying I/O sizes to make sure that, regardless of the I/O size that a VM sends, it delivers the same load on the system. Without the SolidFire Plug-In for VMware vCenter, you must use a 1:1 VM to volume ratio to deliver the full functionality of SolidFire QoS. In this design scenario, you should use SolidFire QoS for high performance, mission critical, and performance-sensitive applications. See the NetApp SolidFire Quality of Service guide to determine what SolidFire QoS setting should be used for each type.

The VMware SIOC solution uses rate limiting and prioritization of VMs to limit their effect on other volumes based on either a congestion threshold setting or a percentage of peak throughput. When a congestion state is identified on a datastore, SIOC is activated and prioritizes VMs based on assigned SIOC shares and the rate-limiting configuration of each VM. VM SIOC priority can become problematic if the VM causing the congestion has a high priority, thereby drowning out VMs with a lower priority. In addition, SIOC can be configured to rate-limit IOPS. However, it does not consider I/O size. If a VM sends larger I/O sizes, it creates more load than it would by sending smaller I/O sizes.

The SolidFire VMware Plug-In integrates SolidFire QoS with VMware SIOC to mitigate many of the limitations of the previously mentioned independent solutions. In this configuration, a 1:1 ratio exists between SolidFire volumes and vSphere datastores. You can optionally enable QoS and SIOC on a per-datastore basis from the vSphere web client. Performance characteristics are configured in vSphere and pushed dynamically to the matching SolidFire volume. When enabled on a datastore, two possible performance control strategies are available:

- Per-datastore control (tiering)
- Per-VM control

See the Element Plug-In for vCenter Server User Guide for more details.

7 vSphere Storage Distributed Resource Scheduler and SolidFire

Storage Distributed Resource Scheduler (DRS) provides placement and load management capabilities for a cluster of VMFS datastores. When using SolidFire storage and implementing Storage DRS, consider the following:

- You do not need Storage DRS for load management when using QoS and SIOC integration made available through the SolidFire vSphere client plug-in. This integration automatically adjusts the datastore IOPS profile based on the Storage I/O control settings. Additionally, Storage DRS sets a threshold of 15ms before it initiates a VMware vSphere Storage vMotion operation to balance load, whereas Storage I/O control uses a threshold of 30ms by default. SolidFire QoS and SIOC integration automatically sets the SIOC congestion threshold to 5ms. Therefore, this setting initiates SIOC and QoS enforcement before Storage DRS has an opportunity to rebalance. Reducing the Storage DRS threshold might cause excessive migration of virtual disks.
- NetApp recommends enabling Storage DRS for initial VM disk placement, disk affinity, and datastore maintenance mode. NetApp recommends configuring Storage DRS to use an automation level of No Automation (manual mode). You may want to adjust Volume IOPS numbers rather than moving VMs around and under no circumstances should it be made automatic.
- If your environment requires full automation with Storage DRS, consider reducing your I/O latency threshold to something less than the storage I/O control threshold.

8 VMware Advanced Settings

You can improve performance for some demanding workloads by adjusting the following parameters. NetApp recommends leaving these settings at default values unless you have specific workloads that are known to benefit from modifying these values. These changes would be done in the SolidFire GUI under advanced settings.

Table 2 describes the software iSCSI queue depth options.

Table 2) SW iSCSI queue depth.

SW iSCSI Queue Depth	
Description	This option is also known as disk queue length (DQLEN). The iscsivmk_LunQDepth parameter controls the maximum number of outstanding commands, or queue depth, for each LUN accessed through the software iSCSI adapter.
Default value	128. The valid range is 128 to 4096.
When to set it	Increase the queue depth when the esxtop utility shows a very high %USD value for SF-series disk-device access over iSCSI with low reported DAVG/cmd response times. Increasing the queue depth when the DAVG/cmd counter is elevated might only increase response times under those circumstances.
How to set it	You can use the following command to configure queue depth: `esxcli system module parameters set -m iscsi_vmk -p iscsivmk_LunQDepth=256`
More details	For more details, see the VMware KB article 1008113.

Table 3 describes the summarization size option for VAAI offload commands.

Table 3) MaxHWTransferSize.

MaxHWTransferSize	
Description	Specifies the summarization size for VAAI offload commands.
Default value	4096. The valid range is 1024 to 16384.
When to set it	Increase MaxHWTransferSize whenever the cluster might offload commands from ESXi. MaxHWTransfersize reduces the number of SCSI summarization commands required to complete offloads. Note that setting the hardware transfer size to a value greater than 4094 on SolidFire Element Carbon 6.1435 (Patch 3) or earlier causes the SolidFire cluster to ignore XCOPY offloads, leading to poor clone and vSphere Storage vMotion operation performance.

MaxHWTransferSize	
How to set it	You can use the following command to configure the MaxHWTransferSize value: `esxcli system settings advanced set -i 16384 -o /DataMover/MaxHWTransferSize (reboot host)`
More details	See the VMware vSphere Storage APIs – Array Integration (VAAI) documentation on the VMware website.

Table 4 describes the iSCSI Delayed ACK option.

Table 4) iSCSI Delayed ACK.

iSCSI Delayed ACK	
Description	Allows the initiator to delay acknowledgment of received data packets.
Default value	The default value is on. Valid values are on and off.
When to set it	NetApp recommends enabling this option for highly sequential workloads. If you have latency sensitive random workloads, you can improve performance by disabling this option.
How to set it	You can modify the DelayedAck setting on a discovery address using the following steps: 1. On a discovery address, click the Dynamic Discovery tab. 2. Click the Server Address tab. 3. Click Settings > Advanced. 4. Deselect the DelayedAck setting.
More details	See the VMware KB article 1002598.

Table 5 describes the header digest option.

Table 5) Header digest.

Header Digest	
Description	Increases data integrity on unreliable Ethernet networks. Might degrade performance.
Default value	Prohibited. Valid values are Prohibited, Discouraged, Preferred, and Required.
When to set it	Enable header digest when there are a high number of cycle redundancy check (CRC) errors or frame errors on the storage network.
How to set it	You can modify the header digest setting using the following commands. Note that the syntax for ESXi versions prior to 6.5 requires “digest" in the "value":

Header Digest	
	`esxcli iscsi adapter param set -A vmhba64 -k HeaderDigest -v required` `esxcli iscsi adapter param set -A vmhba64 -k HeaderDigest -v digestRequired`
More details	See Listing and Setting iSCSI Parameters in the ESXi product documentation at the VMware website.

Table 6 describes the data digest option.

Table 6) Data digest.

Data Digest	
Description	Increases data integrity on unreliable Ethernet networks. Might degrade performance.
Default value	Prohibited. Valid values are Prohibited, Discouraged, Preferred, and Required.
When to set it	Enable data digest when there are a high number of CRC errors or frame errors on the storage network.

How to set it	You can modify the data digest setting using the following commands. Note that the syntax for ESXi versions prior to 6.5 requires "digest" in the "value":
	`esxcli iscsi adapter param set -A vmhba64 -k DataDigest -v required`
	`esxcli iscsi adapter param set -A vmhba64 -k DataDigest -v digestRequired`

More details	See Listing and Setting iSCSI Parameters in the ESXi product documentation at the VMware website.

Appendix A: Virtual Volumes Support

Element software version 9 (Fluorine) and later supports VMware vSphere vVols. Please see the VMware VVol Configuration Guide for information on setup and configuration of vSphere vVols.

Where to Find Additional Information

To learn more about the information that is described in this document, review the following documents and/or websites:

- Provisioning SolidFire IOPS for Quality of Service. This document discusses the concepts of provisioning IOPS in a series of use cases https://www.netapp.com/us/media/tr-4644.pdf
- SolidFire Element Release Notes. This document provides information about enhancements, system capabilities, known issues, and workarounds for the latest release of the Element software. https://library.netapp.com/ecm/ecm_download_file/ECMLP2859110

- SolidFire Element 9 User Guide. This document describes how to setup and administer a SolidFire cluster.
  https://docs.netapp.com/sfe-115/index.jsp
- Best Practices for Running VMware vSphere on iSCSI. This paper explains the design considerations and deployment options for deploying vSphere infrastructures using iSCSI storage.
  https://storagehub.vmware.com/t/vsphere-storage/best-practices-for-running-vmware-vsphere-on-iscsi/
- VMware vVols Configuration Guide. This document discusses how to use vVols and storage-based policy management.
  https://www.netapp.com/us/media/tr-4642.pdf
- VMware vVols FAQ. This document addresses frequently asked questions related to the vVols feature.
  https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/virtualvolumes/vmware-virtual-volumes-faq.pdf

Contacting NetApp Support

If you have any questions or comments about SolidFire documents or products in general, contact NetApp support or email support@solidfire.com.

Version History

Version	Date	Document Version History
Version 9.0	Oct 2016	Initial release.
Version 1.1	November 2019	Updates and corrections.
Version 1.2	April 2020	Updates, corrections, and additions.
Version 1.2.1	August 2020	Added recommendation on smartd.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

Trademark Information

TR-4806-0820