BEST PRACTICES

Configuring iSCSI in a VMware ESX Server 3 Environment

vmware®

foedus

Contents

Introduction

VMware Infrastructure 3 is the industry's first full Infrastructure virtualization suite that empowers enterprises and small businesses alike to transform, manage, and optimize their IT systems infrastructure through virtualization. VMware Infrastructure 3 delivers comprehensive virtualization, management, resource optimization, application availability, and operational automation capabilities in an integrated offering.

With the release of VMware ESX Server 3, VMware has expanded support for storage options available for the VMware VMFS cluster file system. In addition to Fibre Channel SAN storage, VMware Infrastructure 3 adds support for iSCSI SAN storage. The adoption of iSCSI storage is quickly expanding throughout IT datacenters worldwide. The benefits of an iSCSI storage solution are evident:

- ISCSI SAN provides enterprise-class storage at a fraction of the price of a Fibre Channel SAN
- iSCSI SAN can be integrated into an existing Ethernet infrastructure.
- iSCSI storage permits IT departments to deploy their storage solutions using existing resources without the need for additional hardware or IT staff.

The support of iSCSI SAN in a virtual environment provides IT administrators a low-cost, highly available solution for deploying virtual machines. For the first time, IT departments that could not afford a Fibre Channel SAN will now be able to take advantage of advanced VMware Infrastructure functionality such as:

- Redundant storage paths
- Centralized storage environment
- Live migration of running virtual machines with VMware VMotion™
- Dynamic allocation and balancing of computing capacity with VMware DRS

This paper provides information on best practices for configuring and using iSCSI storage with VMware ESX Server 3. To get the most out of this paper, you should understand the basic concepts of virtualization and be familiar with the process of configuring VMware ESX Server.

iSCSI Explained

Internet Small Computer Systems Interface (iSCSI — RFC 3720) is a relatively new SAN technology. Simply put, iSCSI is a means of transport for the SCSI protocol, encapsulating SCSI commands within TCP packets. The commands are encapsulated at the block I/O level, rather than the file I/O level, making the storage appear local to the host operating system. Since all the commands are at the block level, all SCSI rules still apply. For instance, a SCSI storage device cannot be shared without a file system that allows SCSI command sharing.

Now that flexible and inexpensive storage can be connected over a standard Ethernet infrastructure, IT Administrators have a world of new possibilities for deploying storage solutions.

Initiators

An iSCSI initiator provides a means for an iSCSI client (the ESX Server host) to access storage on the iSCSI target (the storage device). There are two implementations of iSCSI initiators: hardware and software.

The software initiator is a driver that interacts with the ESX Server host's TCP/IP stack to contact the iSCSI target via an existing Ethernet adapter. This adds a significant amount of workload to the host's CPUs because the iSCSI protocol needs to be unpackaged and read by the host CPUs, resulting in a performance decrease under any type of significant I/O load. Implementations of a software initiator should be restricted to areas where performance is not a requirement, such as for development. No additional software is needed to configure a software initiator on VMware ESX Server.

A hardware initiator is an adapter card — commonly referred to as a Host Bus Adapter, or HBA — that implements connectivity from the iSCSI client to the iSCSI target but does so more efficiently. Rather than utilizing the host's CPU cycles to process the iSCSI protocol, this approach offloads the traffic to the HBA, which does the necessary processing.

Hardware-initiated iSCSI is supported experimentally with ESX Server 3.0. In particular, the QLogic 4010 adapter can be used for test and development environments, but it is not currently recommended for a production environment.

Software-initiated iSCSI,, on the other hand, is supported fully and is the recommended configuration for production iSCSI deployments with ESX Server 3.0.

Discovery and Logging On

Discovery allows an initiator to find the iSCSI target or targets to which it has access. This requires minimal user configuration. To discover an iSCSI target using a software initiator, the initiator sends a query (SEND TARGET) to the iSCSI target on a specified TCP port (TCP port 3260 by default).

Once an iSCSI target is discovered, the initiator logs on and a TCP/IP connection is created. This connection provides the means for iSCSI data transfer, the negotiation of parameters, and (optionally) authentication. After it logs on, the initiator can begin sending SCSI commands to the target.

Authentication

iSCSI implementations commonly use the Challenge Handshake Authentication Protocol (CHAP) for authentication. This method is supported for the VMware ESX Server implementation of iSCSI storage. During the CHAP authentication process, the target sends its ID along with a random key to the initiator. The initiator replies with the proper credentials that allow for authentication. The initiator also sends a hash value containing the same random key, the initiator's ID, and a CHAPsecret, or password. If the CHAPsecret is correct, the target grants access to the initiator. VMware ESX Server supports one set of CHAP credentials per software initiator.

The goal of CHAP is to authenticate the iSCSI client at the target and to prevent any illegitimate access to the target's storage. This is known as unidirectional CHAP. Bidirectional CHAP is the process of authenticating on both the client and target side. ESX Server supports only unidirectional CHAP authentication. In addition, only one set of authentication credentials can be sent from the ESX Server host. ESX Server does not support per-target authentication credentials.

Note: The following authentication protocols are not supported: Kerberos, Secure Remote Protocol (SRP), or public key authentication methods for iSCSI. Additionally, IPsec authentication and encryption are not supported with VMware ESX Server. The following discovery and management protocols are not supported: iSNS and SLP.

Designing the Environment

The following suggestions for implementing an iSCSI solution in a VMware ESX Server environment will help you design a system that provides the highest level of services, improving performance and providing redundancy in the event of a hardware failure.

Network Considerations

The proper network architecture is essential to a healthy, high-performance environment. There are several different ways of approaching network architecture and fine-tuning it to meet the demands of a virtualized environment.

Use the following recommendations to guide you in configuring the iSCSI network:

- For iSCSI configurations, ESX Server supports only isolated networks. Inherently, Fibre Channel SANs have an added security advantage over iSCSI SANs because they are based on a physically isolated fabric. In a typical network every server, workstation, printer, etc. communicates using an Ethernet-based network, the same network used by the iSCSI nodes and hosts. This opens iSCSI traffic to a wide array of security vulnerabilities as well as degraded performance.
- VMware recommends dedicated Ethernet switches for iSCSI connections.
- Use Gigabit Ethernet network adapters for iSCSI initiators. 100Mb adapters are not supported for iSCSI connectivity.
- Ensure that the iSCSI interface is set to full duplex or configured to negotiate at full duplex.
- To prevent oversubscribing (input from multiple network interfaces being sent to fewer network interfaces) in heavily utilized networks, utilize multiple network connections on the iSCSI target. Ensure that you do not have multiple virtual machines performing large write operations to a single storage point.
- Use static IP addresses for initiators. ESX Server does not support DHCP for iSCSI connections.

  Note: If DHCP must be used and the storage is on a public LAN, be sure CHAP authentication is implemented.

Disk Considerations

Use the following recommendations for configuring the iSCSI storage:

- iSCSI vendors offer a variety of RAID implementations to meet the performance and redundancy needs of different applications. Each LUN should have the right RAID level and storage characteristic for the specific applications running in the virtual machines allocated to the LUN.
- If a virtual machine needs more space than can be allocated from the existing VMware VMFS file system, avoid extending VMware VMFS volumes. Instead, create a new LUN with a new VMware VMFS volume. An extended volume will not balance data across the two physical participants, resulting in hot and cold spots at the target.
- Enable read/write cache on the iSCSI target.
- Where possible, dedicate disk or RAID groups to LUNs that will host VMware VMFS volumes. Remember that multiple hosts will be requesting I/O from the disk or RAID group simultaneously.

Initiator Considerations

By enabling redundant iSCSI initiators and switches, you greatly reduce the possibility of a service outage caused by a hardware failure. It is a good idea to diagram the proposed environment to

identify any single points of failure. Figure 1 is a sample diagram of a fully redundant iSCSI implementation.

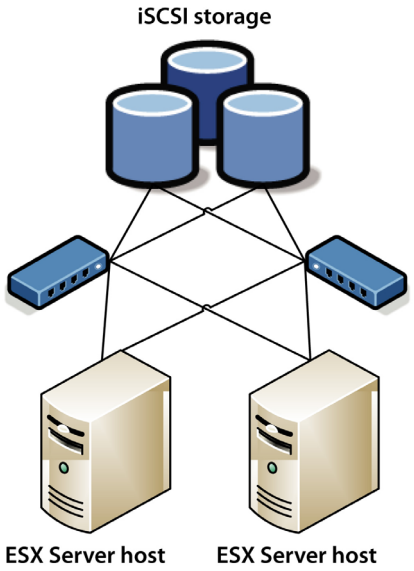

Figure 1: A topology providing redundant paths to iSCSI storage

Software Initiator Specific

The following recommendations are important when you are using a software initiator:

- When using a software initiator, use a dedicated virtual switch to lower chances of having network traffic intercepted by potential attackers during transmission. This configuration will physically segment virtual machine network traffic and iSCSI traffic.
- You can configure only one software initiator on an ESX Server host. When configuring a virtual switch that will provide iSCSI connectivity, bind multiple network connections to the switch to provide redundancy.
- Ensure the network adapters that are bound to the virtual switch originate from separate network switches (see Figure 1) to eliminate any single points of failure.

Hardware Initiator Specific

The following recommendations are important when you are using a hardware initiator:

- To enable path failover for any given LUN, each initiator must have access to that LUN. If you define a fixed failover path policy, the storage LUN uses the preferred path you specify whenever available. If the preferred path becomes unavailable, the ESX Server host chooses the next available path to use, then fails back to the preferred path once it becomes available again. Alternatively, the most recently used (MRU) failover policy continues to use the same path until a service interruption. With an MRU policy, there is no fail back action. Failover policies are configured from either the Virtual Infrastructure Client or the ESX Server command line.

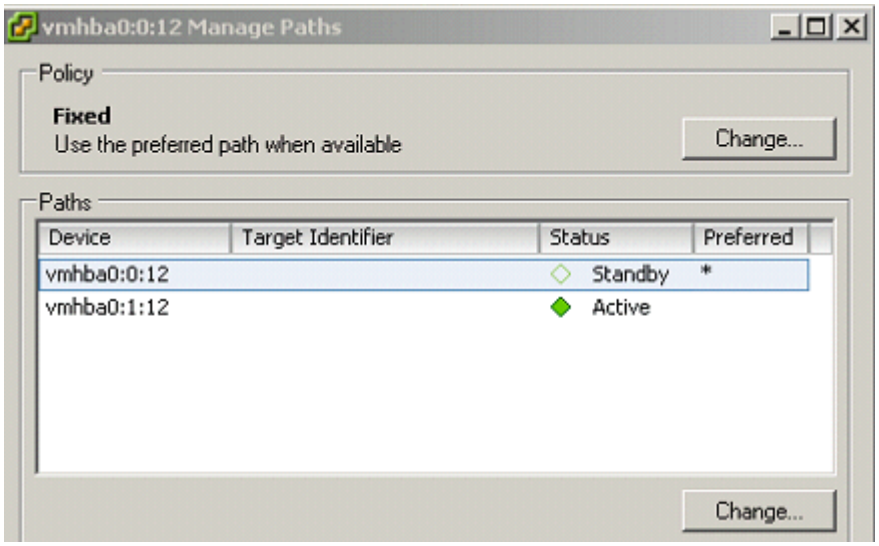

Figure 2: Managing the failover path policy from the Virtual Infrastructure Client

- Many iSCSI targets utilize more than one storage processor. This increases performance and provides redundancy in the case of a storage processor or path failure. Depending on the vendor, the target uses either an active/active or active/passive method of accessing a LUN. When using an active/active iSCSI target, use fixed mode failover policy. With an active/passive target, use MRU.

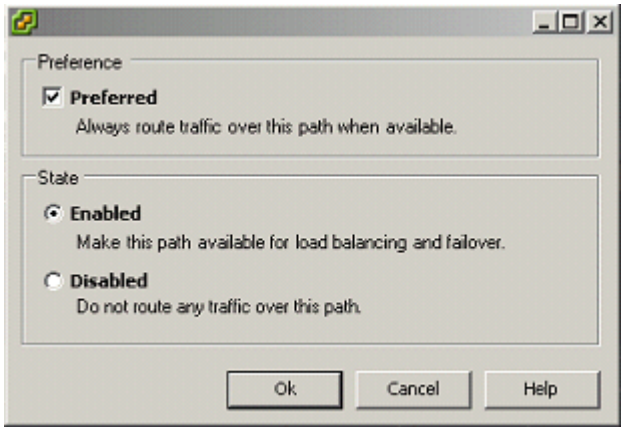

Figure 3: Setting a fixed path policy from the Virtual Infrastructure Client

- Manually spread I/O across all available initiators and storage processors. This helps to balance I/O traffic on both the ESX Server host and the iSCSI target. In the event of a failure, the alternate initiator or the alternate storage processor assumes ownership of the LUN.
- Disconnect the HBAs during ESX Server installation when you install an ESX Server host connected to an existing SAN.
- Set maximum queue depths to prevent bottlenecks at the iSCSI initiator during high I/O periods. This example uses the syntax for the QLogic 4010 HBA. Verify the syntax for other adapters with the manufacturer's documentation. Find the appropriate device name in `/etc/vmware/esx.conf`.

```
/device/001:02:0/vmkname = "vmhba1"
```

Add the following line after the device name:

```
/device/001:02:0/options = "ql4xmaxqdepth=nn"
```

Maximum Configurations

The following list shows supported features and maximum configurations allowed by ESX Server 3.0:

- A maximum of 254 LUNs
- A maximum of 128 VMware VMFS 3 volumes
- A maximum size of 64TB per VMware VMFS 3 volume
- A maximum of 8 targets for iSCSI
- Clustering not supported for iSCSI
- Boot from SAN not possible using software initiator, only using a hardware initiator supported experimentally by VMware

Conclusion

iSCSI is an excellent fit for many VMware ESX Server environments. When iSCSI is implemented correctly, IT staffs can cut costs and save IT resources while surrendering little to no functionality. Implementing the iSCSI environment correctly through careful design and planning can provide a high level of uptime while minimizing costs associated with a virtual infrastructure deployment.

About the Author

Rob Daly is a Senior Systems Engineer for Foedus. He is a VMware Certified Professional and has worked in the IT industry for nine years. Over the last four years, Rob has concentrated his focus on VMware Infrastructure, designing and implementing a wide variety of VMware ESX Server solutions for companies throughout the United States, including several for Fortune 100 companies.

Revision: 20060912 Item: OP-005-INF-002-110

VMware, Inc. 3145 Porter Drive Palo Alto CA 94304 USA Tel 650-475-5000 Fax 650-475-5001 www.vmware.com

vmware®

foedus