Best Practices for Implementing VMware ESX Server

Planning and Operations in the Datacenter

Abstract

Today's data center is an extremely complicated blend of varied hardware, operating systems, data and applications that have accumulated over time. One of the ways to combat the growing requirements is virtualization of the server. Server virtualization abstracts IT services (such as e-mail) from their respective network, storage and hardware dependencies -- all the factors that are traditionally overly complex, and costly to operate.

The maturity of server virtualization technologies, particularly VMware Inc.'s ESX Server software for data center architectures, is prompting an increasing number of companies to place virtualized servers directly into production or sometimes use them with data replication to secondary or tertiary sites for disaster recovery. Because data centers are complex infrastructure comprised of the servers, storage and the underlying network, they need a proper methodology to contend with interoperability of the diverse array of vendor equipment.

This paper provides insights into some of the best practices for implementing iSCSI, NAS and SAN configurations using VMware ESX Server.

Introduction

Enterprise data centers are evolving into architectures where networks, computer systems, and storage devices act in unison, transforming information into a significant asset and at the same time increasing the organization's operational complexity. As IT organizations transform their data centers into more cost-effective and virtualized, operations, they will need to take into account the requirements for data center consolidation, business continuance, virtualization, application optimization and future systems' enhancements.

One thing that has been a growing challenge is keeping up with all the hardware, operating system and software upgrades. Upgrades and new releases are a constant factor presented by every system vendor. As data center managers struggle with server consolidation and server virtualization, systematic planning plays a major role in optimizing operational resources, and, at the same time, increasing the system efficiency.

Server virtualization permits multiple virtual operating systems to run on a single physical machine, yet act logically distinct with consistent hardware profiles. VMware provides a production-ready server virtualization suite, VMware Infrastructure. The VMware ESX Server is the

building block of VMware Infrastructure. It creates the illusion of partitioned hardware by executing multiple "guest" operating systems, thus providing a virtualization platform for the data center.

Planning

The ESX Server offers a very sturdy virtualization platform that enables each server to host multiple secure and portable virtual machines running side by side on the same physical server. Virtual machines are completely encapsulated in virtual disk files that can be either stored locally on the ESX Server or centrally using shared SAN, NAS or iSCSI storage. The ESX Server is managed by a VirtualCenter which can centrally manage hundreds of ESX servers and thousands of Virtual Machines.

The following section provides an introduction to the various possible configurations in storage and provides a solution to a few show stoppers that could evolve during the installation process. Later sections describe a few scenarios during the failover process.

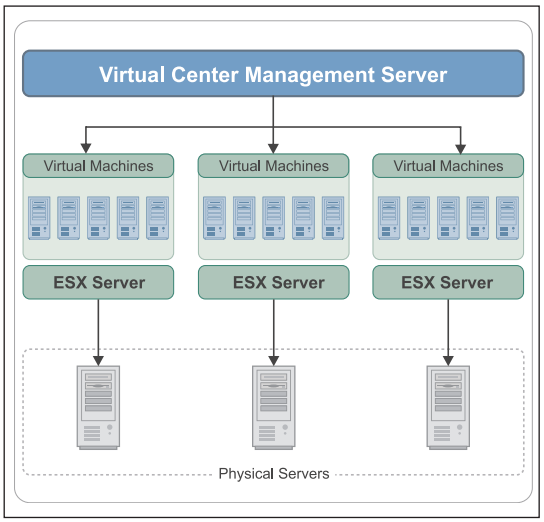

Figure 1: VMware Infrastructure

Best Practices in Installation

Hostname Naming Convention

It is recommended that the hostname should not contain an underscore. Labels cannot start nor end with a hyphen. Special characters other than the hyphen (and the dot between labels) are not advisable, although they are sometimes used. (1)

Boot from SAN

If an ESX server is installed on a SAN LUN then this LUN should be mapped exclusively for this server only.

Discovering the added LUNs without rebooting the Linux VM:

To discover the added LUN in the Linux machine without rebooting the machine the following solution can be adopted

```
#cd /sys/class/scsi_host/hostX
        - X is 0 in our Linux VMs
#echo "- - -" > scan
# lsscsi -- This should display all the attached disks.
# fdisk -l
```

Formatting and adding of RDM Luns in /etc/fstab file in the Linux machines:

Format the RDM LUN's using the 'fdisk' command and make a file system (ext3) on it. Once done edit the /etc/fstab file and add the following entries to automatically mount these LUN's even after a reboot of the VM.

```
# vi /etc/fstab
/dev/sdb1 /root/VD_PT auto defaults 0 0
/dev/sdc1 /root/VD_NPT auto defaults 0 0
```

Where

- `/dev/sdb1 and /dev/sdc1` are the RDM mount drive
- `/root/VD_PT and /root/VD_NPT` is the mount directory mounted on /dev/sdb1 /dev/sdc1 respectively.
- Reboot the VM and give the 'df -h' command in the /root directory to verify the devices are indeed mounted

Virtual Machine Naming Convention

The name should provide all the details regarding the VM. This makes it user friendly and the configuration of the virtual machine can be understood without looking into the configuration page.

For instance, to name "Red Hat Enterprise Linux 4, LSI logic SCSI controller, 64-bit", it would be RHEL4_SRV_ENT_LSI_b64.

Checking the number of luns from the console (1)

The graphical interface to the ESX platform viz.Virtual Infrastructure client may slow down due to an excess number of luns created or some network issues. The number of luns can be checked through the console using the following command instead:

If the failover policy is Fixed, then

```
# esxcfg-mpath -l | grep Fixed | wc -l
```

If the failover policy is MRU, then

```
# esxcfg-mpath -l | grep "Most
Recently Used"| wc -l
```

Time synchronization between the ESX Server and the Virtual machines [1]

VMware emphasizes synchronizing the time between the virtual machines and the ESX Server. Time synchronization helps in debugging issues that were faced during the use of the product. Use the following commands to set the date and time on the ESX Server

`# date -s "02/01/2008 11:34"` - To set the date to Feb 1 2008 11:34am.

`# hwclock --systohc` - To sync the hardware clock with the system time.

After setting the time and date in ESX Server use the command "vmware-toolbox" on Linux VM's to sync its time with ESX server. On Windows, double click on the VMware Tools Icon and select the option to sync the time between VM and ESX.

Faster method for creating Virtual machines by cloning

Cloning a virtual machine is mostly a matter of copying its virtual disks (possibly resetting system attributes before doing so) and then building virtual machine configurations around the copies.

Example: Suppose four virtual machines need to be created, all running Windows 2000 Server. One is to be a file server, one a mail server, one a DHCP server, and one an intranet Web server. The first two have locally stored data, and the last two do not. One way to create these virtual machines would be to build each from scratch as shown in Fig 2. This approach is very labor-intensive. It is also highly error-prone, because many steps are repeated. There is high risk of missing some steps, performing them incorrectly, or in the wrong order. A better approach uses cloning.

In this example, we do each package installation only once, and we clone the virtual machine at appropriate junctures as depicted.

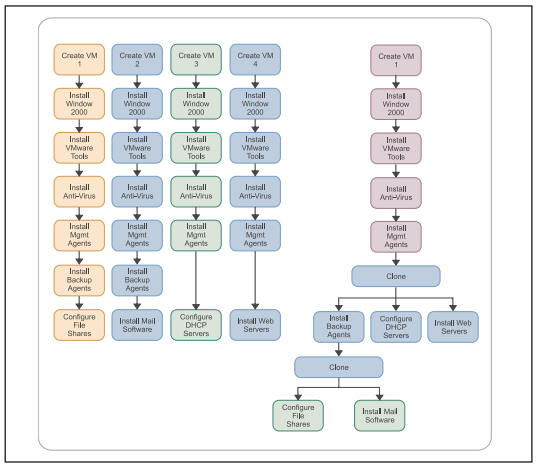

Figure 2. Creating VMs using the OS Installation CD; Cloning from the basic image.

There are three ways in which virtual machines can be created using cloning: [1]

Method 1

a. First install the basic VMs of all types ie. different OS with different SCSI controller types (say LSI Logic or Bus Logic).

b. To create further VM's of the same type, just create a New Virtual Machine (say VM-New) using VI Client. Instead of installing the GOS from CD, it is enough to overwrite the *-flat.vmdk file with an existing *-flat.vmdk. Overwrite the VM-New-flat.vmdk file under /vmfs/volume/storage1/VM-New directory in the ESX Server with the *-flat.vmdk of the basic VM.

```
# pwd /vmfs/volumes/storage1/New-VM
# cp -r
/vmfs/volumes/storage1/Basic-
VM/Basic-VM-flat.vmdk
/vmfs/volumes/storage1/New-VM/New-
VM-flat.vmdk
```

c. Overwrite existing files.

d. Therefore using VI client, we can create all the VMs and run the copy operation in a script. This will avoid manual intervention while creating the VM.

Method 2

a. First install the basic VMs of all types ie different OS with different SCSI controller types (say LSI Logic or Bus Logic).

b. Now create a template for all the basic VMs. Power off the VM. Using VI client, click on "Convert to Template" option. Now make a tar of the VM directory under /vmfs/volumes and upload the tar to some other place i.e., tar cvf VM1.tar /vmfs/volumes/storage1/VM1/.

c. We can use this tar under /vmfs/volumes directory in the ESX Server to create a VM. But after it is untarred, all the files need to be changed to represent the new VM. If the new name should be VM2, then we need to replace all the entries corresponding to VM1 as VM2 from all the files.

Creating the template will remove the entry from the VI client. To register the VM back on the VI client, browse the storage on the ESX Server and go through this directory and double click on the vmx file. Using this method we can clone across ESX but need to be very careful while editing the entries.

Method 3

The third method is to use the "'Clone"' option in Virtual Center.

The following section deals with some of the common scenarios faced during the overall setup and administration of the data center.

VMware tools out of status

Check for the VMware tools corresponding to the ESX version that is installed. The VMware tools have to be updated according to the ESX Server.

ESX Server status showing as "disconnected" or "not responding"

- If the ESX Servers are added by host name, make sure you have them entered in the Windows machine in `C:\Windows\system32\drivers\etc\hosts` file as well.
- If the ESX Servers were added using the hostname by host name, you need to enter all the ESX Server names in `/etc/hosts` of all ESX systems and the Linux virtual machines.
- In the VI Client make sure that the ports are open by checking in `configuration->Security->open all ports for access`.
- Try making some change to a GOS changes like power off/on.
- Reboot ESX. The VC might then discover the ESX as online.

Make sure that you disable the firewall (Windows and/or any other operating system) on the system running VC.

Cloning Windows

Cloning a Windows bus machine generates an error "Specified key, name, identifier already exists":

- Execute the following commands in the service console in the ESX Server.
  - `service mgmt-VMware restart`
  - `service VMware-vpxa restart`
  - `service VMware-vmkauthd restart`
- Remove all the VM configuration files like vmx, vmdk for the particular VM
- Rename the source VM before cloning and then clone.
- Reboot the Server.

Solaris installation gets stalled

Increase the memory size on the Solaris virtual machine. Ideally the advised memory is 1 GB on the host.

SUSE Virtual Machine not being dynamically assigned an IP

To bring up the network on SLES VM, do the following:

- `ifconfig -a` -- Note the Mac address say <mac1>
- `cd /etc/sysconfig/network/`
- Here you will see a file ifcfg-eth-id-<Mac Address> (say mac2)
- Copy the contents of the file to another file and name this file as ifcfg-eth-id-mac1 ie

```
# cp ifcfg-eth-id-<mac1> ifcfg-
  eth-id-<mac2>

# service network restart
```

- Now you should be able to ping the gateway and identify an IpAddress for the above eth-id. Now if you enter just ifconfig, it will show you both lo and the eth<id>

Invalid VMs shown in the VI Client

Virtual machines could become inaccessible in the VI Client. The virtual machines can be retrieved in the following way:

Remove from inventory the "Invalid" system and re-add this to the inventory. You can browse the datastore (either local or SAN, depending on where your VM resides) and right-click the VMX file, then select "Add to Inventory".

A-Synchronization of Clocks

The clocks in virtual machines run in an unpredictable manner. Sometimes they run too quickly, other times they run too slowly, and sometimes they just stop. (2)

7.1 64-bit AMD Systems

This problem can occur on some 64-bit AMD multiprocessor (including multi core) systems, if VMware products are being run on them.

Timestamp counters (TSCs) on 64-bit AMD systems should ideally remain synchronized because these systems run all CPUs from a common clock oscillator. However, some 64-bit AMD systems have power management features that can cause the TSCs on some processor cores to lose time in relation to other cores leading to unpredictable clock behavior in the virtual machines run on one of these systems. Disabling the following features could help:

- In the system's BIOS setup windows, look for settings labeled PowerNow or Cool'N'Quiet, and disable them.
- If the Linux host operating system is running, look for a system daemon or applet called cpufreqd, cpuspeed, powernowd, cpudyn, speedy, or cpufreq, and disable it. For example, on some systems the command service cpuspeed stop might work. The instructions to disable the daemon or applet found on a system vary. Refer to your system's documentation for more information.
- If these features are required or disabling them is not possible, each virtual machine should be

assigned to a subset of processor cores on which the TSCs remain synchronized. In some cases this may need to be done after turning off power management in the system's BIOS; this occurs if the system only partially disables the power management features involved.

- If Windows XP Service Pack 2 is running as the host operating system on a multiprocessor 64-bit AMD host that supports processor power management features, the hotfix in the following operating systems may need to be applied.

- Microsoft Windows Server 2003, Standard and Enterprise x64 Editions
  - Microsoft Windows XP Service Pack 2, when used with Microsoft Windows XP Home and Professional Editions
  - Microsoft Windows XP Tablet PC Edition 2005
  - No hotfix is needed for Microsoft XP Media Center.

7.2 Intel Systems

This problem can occur on some Intel multiprocessor (including multi core) systems. After a Windows host performs a "stand by" or "hibernation", the TSCs may be unsynchronized between cores. If the hotfix cannot be applied for some reason, as a workaround, do not use Windows "stand by" and "hibernation" modes.

VI Client setup

Disconnected \ Not Responding signal

If the ESX Server status shows as "disconnected" or "not responding" the following steps will help solve the problem:

- If the ESX Servers were added by host name, make sure they are entered in the c:\windows\system32\drivers\etc\hosts file in the windows virtual machine as well.
- If the ESX Servers were added by host name, the ESX Server and the Linux virtual machine names need to be entered in /etc/hosts file.
- In the VI client go to `configuration->Security->open all ports for access.`
- Try making some GOS change like power off/on
- Reboot ESX.

VC might then discover the ESX as online. At this stage make sure that the firewall is disabled (Windows and/or any other SW) on the system running VC.

Virtual machine creation

The virtual machine created in a VI client will have a default LSI logic driver. If a virtual machine with a bus logic driver needs to be installed the following steps are required. Virtual Machine Installation for Windows with Bus Logic driver:

In VI-Client

- Right-click on the ESX machine on which to create the new virtual machine and select the "Create New Virtual Machine" option.
- Choose the "Custom"option in the New Virtual Machine wizard. Then, name the virtual machine as per the naming convention requirements.
- Choose the Datastore on which to create the new virtual machine. Then, choose the Operating system as Microsoft Windows.
- Choose the following options of the number of processors, memory and network connections as per the requirement and topology.
- In the I/O adapters section of the wizard, select the Bus Logic SCSI Adapter and proceed by choosing to create a new virtual disk in the Disk Capacity section of the Wizard.
- Specify the required Disk size and select the "Store with Virtual machine" for disk location. Keep default values in the advanced options menu and complete the wizard process.
- Once the virtual machine is created, on the newly created virtual machine, go to right click->Edit Settings->CDROM and select the Device type as Host device and check the "Connect at Power On" option and click OK.
- Power on the new virtual machine and click on the console. A Windows setup blue screen will appear and prompt to press function key F6 to specify any additional drivers. Watch out for this prompt and press the F6 button on time.
- Proceeding further after the F6 button is pressed, setup will continue until it reaches a point where it will prompt to specify additional devices. At this time, go to the virtual machine on VI client and right click->Edit settings->Floppy Drive.
- Check the "Use existing floppy image in datastore" option under the device type option and point to the floppy image for installing the Bus Logic driver. This image will be available under "/vmimages/floppy/vmscsi-1.2.0.2.flp" .Select the "Connect at Power On" and "Connect options" and click on OK
- Return to Windows Setup menu and continue

setup (by pressing Enter). Now follow the setup and installation process as usual to complete the installation.

Configuration

4.1 ISCSI Configuration

The section provides a brief overview of the VMware iSCSI implementation using either a software initiator or a hardware initiator and basic deployment steps covering both software and hardware initiator options.

iSCSI presents SCSI LUNs (targets) to iSCSI initiators (requesters), which unlike NAS makes block devices available via the network. Therefore you can mount block devices (disks) across an IP network to your local system, and utilize those as you would with any other block device.

The figure below depicts an iSCSI SAN configuration. With the release of ESX Server 3.x, VMware added support for iSCSI with both software initiator and hardware initiator implementations. The software initiator iSCSI plugs into the ESX Server storage stack as a device driver similar to other SCSI and Fibre Channel drivers.

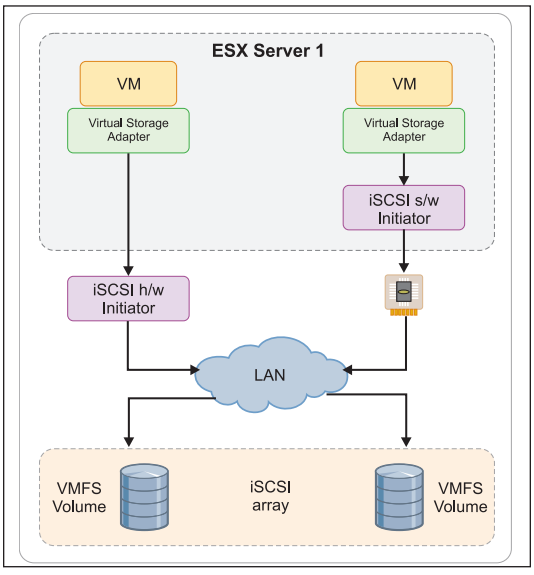

Figure 3: iSCSI SAN Configuration

When the ESX Server host is connected to storage via iSCSI, be sure that the following steps have been carried out to ensure that the communication between the storage array and the ESX Server is in order:

- Installation of VMotion and IP Storage licenses on VC
- Host and storage ports have to be on the same subnet
- A supported Ethernet NIC OR a QLogic40 50or other card on the Hardware Compatibility List (Experimental)
- A supported iSCSI Storage

The following figures illustrate the topology of an iSCSI SAN configuration with a software initiator and hardware initiator respectively.

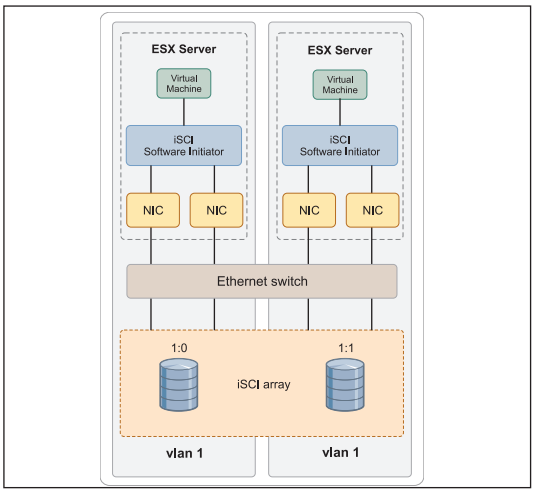

Figure 4: S/W iSCSI initiator configuration

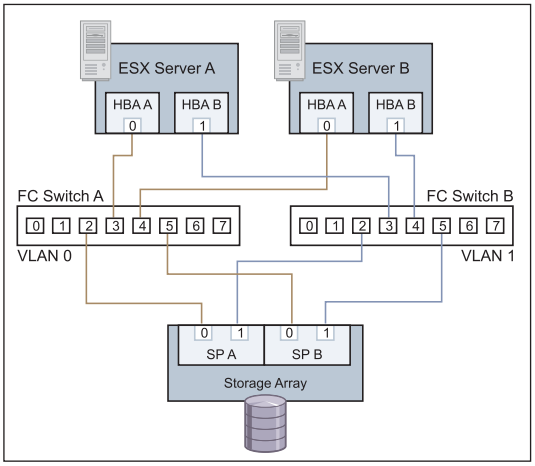

Figure 5: H/W iSCSI initiator configuration

A storage network consists of two types of equipment: initiators and targets. Initiators are data consumers, such as hosts. Targets are data providers, such as disk arrays or tape libraries. Initiators and targets are collectively referred to as end points, which can be software, software with hardware assistance, or hardware. It is important to assign the same Target and LUN (logical unit number) and name from each of the storage processor ports to all LUNs. This ensures that the ESX Server has multiple paths to the LUNs, which are essential for failover.

The key steps involved in the iSCSI configuration and installations are:

- Authentication: Check for CHAP-per HBA or per target.
- Target Discovery
  - Send Targets
  - Send Targets discovery (Dynamic)
  - No SLP discovery
  - No iSNS discovery
- Administration: Configuring the HW initiator and the SW initiator

4.2 NAS Configuration

The following topology is recommended by VMware for NAS.

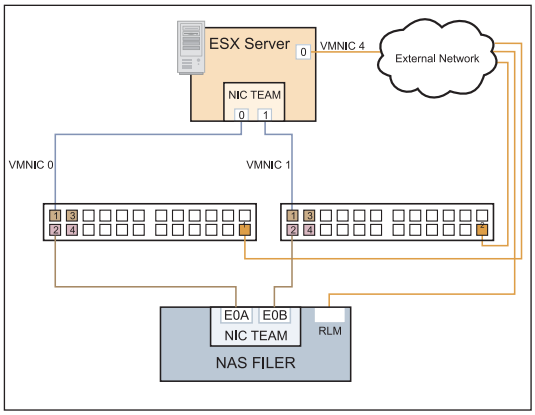

Figure 6: NAS Configuration

Connect the ESX Server to the NAS filer through two IP switches as shown in Figure above. Ensure that all ESX paths can connect to the NAS filer and its respective data stores.

This configuration ensures that the shared NFS volumes can be accessed through two separate IP networks which is required for the failover mechanism.

SAN Configuration

Connect ESX Servers to storage through redundant Fiber Channel (FC) fabrics as shown in figure below. Ensure that each LUN has the same LUN number assigned from all storage processor (SP) ports. In each fabric, assign all hosts, (i.e., FC HBA ports) and targets (i.e., SP ports) to a single zone .Use this zoning configuration for all certification tests, unless otherwise specified. This configuration ensures that four paths are available from each ESX Server to all LUNs.

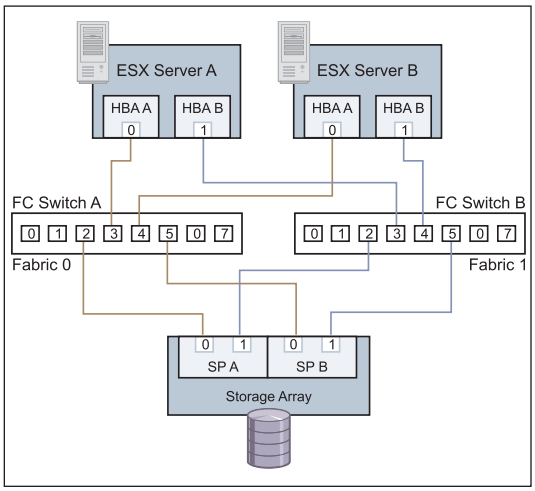

Figure 7: SAN Configuration

Best Practices in Configuration

There are a number of potential errors that can occur during each of the above mentioned steps of the configuration process. If the iSCSI array has different ports and the port has to be on different subnets, it would require two vmkernels, each comprising an IP address from one subnet.

Configuring the software initiator

While enabling the software initiator, the VI client may display an error as shown in the figure below.

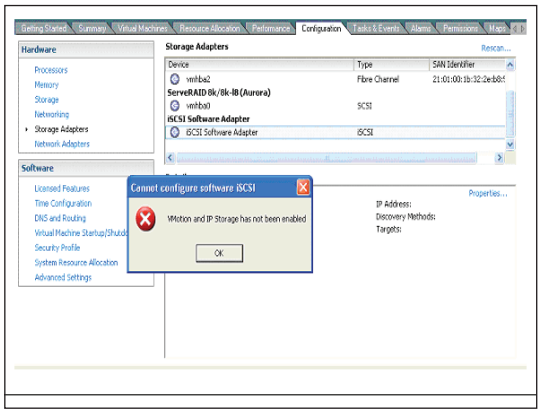

Figure 8: Error while enabling the s/w initiator

The error primarily occurs because the Vmkernel is required for iSCSI/SAN traffic. This error can be avoided by:

- Checking whether you have "'vmkernel"' added into your vswitch.
- Enabling VMotion and IP Storage licenses.

- Configuring a VMotion and IP Storage connection in the networking configuration.
- Adding the service console port to the same virtual switch.

Make sure both the VMotion and IP Storage Network and the service console port connection have appropriate IP addresses and are routed properly to the array.

Unable to discover LUNs

When a hardware initiator is configured, there may be situation where it may not able to access/discover LUNs. The possible solutions to make the LUNs available to the ESX Server are as follows:

- Restart the iSCSI host appliance.
- Disable and Enable the active storage port.
- Vmkiscsi-tool –D –l vmhba1 which will display the static discovery page wherein we can see whether the target is getting discovered.
- Vmkiscsi-tool –T vmhba1 which will display whether the targets are configured or not.

Checking for established SW-iSCSI Session

To check the list of iSCSI session which ESX server has established with the array, enter "cat /proc/scsi/vmkiscsi/<file name>" in the ESX Server service console. This will display the list of LUNs and whether the session is active or not. A "?" indicates the session is not established. An IP address indicates the session has been established.

Authentication

- If the rescan on ESX didn't show up any CHAP Authenticated LUNs, reboot the ESX after enabling the CHAP.

GOS Unstable

There may be instability problems of the guest operating systems installed on the NFS partitions. Possible solution to the problem is as follows:

- Check for the firmware.
- Check the total memory of the VMs. If the total memory exceeds the total actual physical memory available, then it may cause instability in the Guest OS. Try decreasing the memory size.
- Recommended memory size for Windows-256 MB and Linux-512 MB and Solaris is 1 GB.
- Set the MTU in the NAS storage to 1500.If you have Windows 2003 standard 32 bits as one of the OS installed, upgrade it to SP2.
- Increase the Net->Net Heap size to 30 and reboot the ESX Server.

Unstable network

The network may be unstable because of the presence of a NAS or NFS partition. The problem can be mitigated by checking for any issues with the network switch by connecting a desktop to the switch directly and running a ping in long incremental loops. No packet drops implies that the switches are fine. Check the storage by connecting with loop-back cables and ping again. Again, no packet drops implies that the storage is working fine. The problem may be with a wrong network configuration.

NFS data store could not be created

While creating a NFS data store on the NAS array, creating a datastore could be a problem. Possible solutions adopted by Cognizant are as follows:

- Check whether the NFS daemon is running in the ESX Server using the command service –status -all. If the daemon is not running it can be started using "service nfs start" or "service nfs restart" command.
- Check whether the /etc/exports, /proc/fs and the /var/lib/nfs/xtab directories in the ESX Server are empty. They all should actually list the volume mounted on NFS.
- Check for vmkernel logs in the ESX Server. If it shows vmkernel error #13 then the mount point is not determined correctly. Try to mount the nfs partition once again.

Unable to access NFS data stores

A possible condition that could be faced when there are NFS with two different failover paths is the datastore is not accessible.

- NFS share is always associated with an IP address. It is not possible to have a failover on two different networks. We should probably go by host name and then resolve the name internally to two different IP addresses, by using IP aliases.

ESX showing core dump

When trying to mount a maximum number of NFS mount (32), there is a chance that the ESX may crash and show core dump. The possible solution may be:

- Check the ESX Server booting screen, if it shows a NMI: (non-maskable Interrupt) and a HEARTBEAT error (PCUPO didn't have a heartbeat for 18 sec) Physical CPU #0 seems to have died.
- Recommended solution is to contact the hardware vendor and run diagnostics on this machine. It may be a problem related to CPU #0.

ESX Features: Failover

Failover is a backup operational mode in which the functions of a system component (such as a processor, server, network, or database, for example) are assumed by secondary system components when the primary component becomes unavailable through either failure or scheduled down time. Failovers are used to make systems more fault-tolerant, and are typically an integral part of mission-critical systems that must be constantly available. Failover involves automatically offloading tasks to a standby system component so that the procedure is as seamless as possible to the end user. Storage networks use many paths -- each consisting of complete sets of all the components involved -- between the server and the system. A failed path can result from the failure of any individual component of a path. Multiple connection paths, each with redundant components, are used to help ensure that the connection is still viable even if one (or more) paths fail. The capacity for automatic failover means that normal functions can be maintained despite the inevitable interruptions caused by problems with equipment. In a dedicated data center environment, improper failover can cause several issues. Some of the possible issues faced during an S/W iSCSI, H/W iSCSI, NAS failover are addressed below along with the possible solutions implemented successfully by Cognizant to various customers.

S/W iSCSI Failover

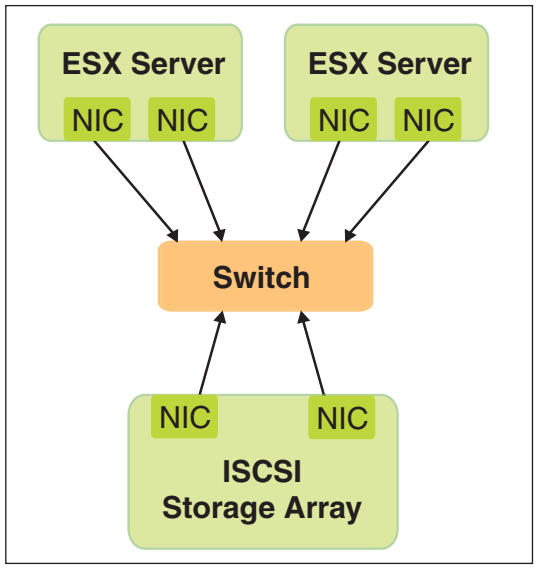

Figure 9: Configuration for failover during S/W iSCSI configuration

The above diagram depicts the configuration setup for a S/W iSCSI NIC/SP failover.

H/W iSCSI Failover

ESX Server supports multipathing to maintain a constant connection between the server machine and the storage device in case of the failure of an HBA, SP port, or switch.Multipathing support does not require specific failover drivers. To support path switching, the server typically has two or more HBAs available from which the storage array can be reached using one or more switches. Alternatively, the setup could include one HBA and two storage processors so that the HBA can use a different path to reach the disk array.

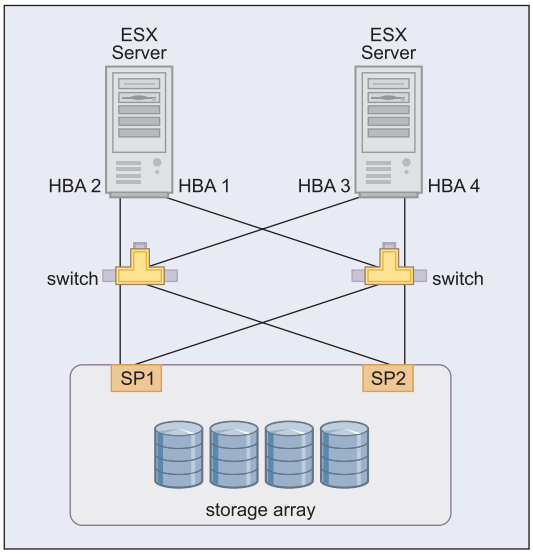

Figure 10: Configuration for failover during H/W iSCSI configuration

In the above figure multiple paths connect each server with the storage device. For example, if HBA1 or the link between HBA1 and the switch fails, HBA2 takes over and provides the connection between the server and the switch. The process of one HBA taking over for another is called HBA failover. Similarly, if SP1 fails or the links between SP1 and the switches breaks, SP2 takes over and provides the connection between the switch and the storage device. This process is called SP failover. VMware ESX Server supports both HBA and SP failover with its multipathing capability. You can choose a multipathing policy for your system, either Fixed or Most Recently Used depending upon the failover policy of the storage device. If the policy is Fixed, a preferred path can be specified. Each LUN (disk) that is visible to the ESX Server host can have its own path policy. Virtual machine I/O might be delayed for at most sixty seconds while failover takes place, particularly on an active/passive storage array. This delay is necessary to allow the SAN fabric to stabilize its configuration after topology changes or

other fabric events. In the case of an active/passive array with path policy Fixed, path thrashing might be a problem.

6.3 NAS Failover

As per the topology diagram in Figure 6 when failover happens on an active NIC, the other NIC should take over and the data transfer should happen without any disruption.

6.4 SAN Failover

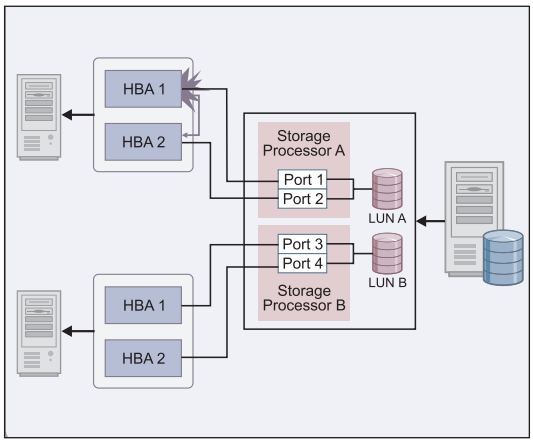

Figure 11: Configuration for failover during FC SAN configuration

The above configuration is that of a failover scenario in a FC SAN configuration. The implementation is very similar to that of H/W iSCSI failover. If one of the HBA cards gives way, the alternate HBA card takes over thus providing an alternate path for data transfer. Similarly if any of the storage ports go down, the alternate port becomes active and continues the flow of data.

Best Practices in Failover

In a typical failover environment when one of the active NICs fail, I/O will continue to flow in the inactive NIC. Some of the issues faced in a failover and the solutions adopted are as follows:

Windows VMs reboots itself during failover

The possible solution is as follows:

- For the Windows 2000 and Windows Server 2003 guest operating systems, standard disk Timeout Value may need to be increased so that Windows will not be extensively disrupted during failover.
- For a VMware environment, the Disk Timeout Value must be set to 60 seconds.
- Select Start > Run, type regedit.exe, and click OK. In the left panel hierarchy view, double-click HKEY_LOCAL_MACHINE->System->CurrentControlSet->Services->disk.
- Select the Timeout Value if it exists and set the Data value to x03c (hexadecimal) or 60 (decimal).
- By making this change, Windows waits at least 60 seconds, for delayed disk operations to complete, before generating errors.
- If the Timeout Value does not exist, select New from the Edit Menu and then DWORD value.
- In the Name field type Timeout Value and then set the Data value to x03c (hexadecimal) or 60 (decimal).
- Click OK and exit the Registry Editor program.

SP failover

Some iSCSI arrays handle controller failover internally. When these arrays are configured with an ESX iSCSI initiator, there will be only one path to ESX. Since this type of array handles controller failover internally, ESX will not know if a controller failover has occurred. We refer to such arrays as supporting internal storage controller failover, or ISCF. When an ISCF failover occurs ESX Server will not see path changes or link loss. The connection may be lost for a short period of time and then restored, with I/O resuming within 60 seconds.

Connection to the storage lost during failover

If an active SP port gets disrupted, this may be due to failure to establish the necessary connections to all LUNs and an inability to access the storage array. Possible solutions include:

- There may be a problem with the Ethernet switch. Check the networking configuration of the switch.
- Power cycle /reboot the storage controller.

Unable to access VMs

During failover access to VMs may be lost. Possible solutions include:

- Check for # cat /proc/scsi/qla4022/2 command to see if the DS is set as 1 and DDBS is set as 4.If DS is set to 1 then failover is not happening and hence the VMS have lost their access.
- Check the driver of the HBA cards.

IO Error during Failover

It may happen in a failover scenario when a active HBA fails due to unavoidable circumstances, the failover does not happen properly due to I/O errors. We faced this issue with some of our iSCSI customers with a Hardware initiator. One of the possible reasons could be due to a hard disk failure (sector error) on iSCSI array. The possible solution adopted is as follows:

- Wipe out the existing RAID configuration and create a new one.

Datastore not restored

Datastore goes down after disabling the active NIC in the NIC Team, the proposed solution is:

- Failover has to happen through the vmnics. If an active NIC is disabled due to unavoidable circumstances data transfer should not be disrupted and it should happen on the alternate NIC.
- Active NIC can be found by doing "'ifconfig"' command.
- Traffic on the NICs can be verified by checking the TX/RX bytes.

Unsuccessful Failover

Failover is not successful due to certain issues. The possible solutions may be:

- Zoning required in case of multi port HBA cards.
- Set the failover policy correctly. It has to be set depending on the storage box being Active/Active or Active/Passive.
- Failover is not successful due to invisibility of the alternate paths from the controllers. The possible solution may be checking the firmware. A firmware upgrade is essential.
- If for any reason Access to the Virtual machines is lost during failover, check the network connectivity and disable all firewalls.

Vmkernel Logs

During failover keep a check on whether the Vmkernel logs a Reservation/IO error in the vmkernel logs. This error indicates that failover is not happening correctly.

Conclusion

Data center planning is an important first step for effective data center management. Particularly in a virtualized data center, conjoining NAS, SAN, and other diverse storage resources onto a networked storage pool can go a long way towards improving efficiencies and overall computing performance.

This paper provided a consolidated dissemination of various aspects of efficiently making use of the virtualization environment in the data center using the VMware ESX Server. These best practices, when implemented correctly, would save time, effort and lower costs in virtual infrastructure implementations. Please note that these solutions are being shared as insights and are not intended to be prescriptive.

References

[1] VMware SAN Configuration Guide's Server 3.0.1 and Virtual Center 2.0.1

[2]http://kb.VMware.com/selfservice/microsites/search.do?cmd=displayKC&docType=kc&externalId=2041&sliceId=2&docTypeID=DT_KB_1_1&dialogID=60294322&stateId=0%200%2060292312

[3] http://i.i.com.com/cnwk.1d/html/itp/VMware_Building_Virtual_Enterprise_VMware_Infrastructure.pdf

About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting and business process outsourcing services. Cognizant's single-minded passion is to dedicate our global technology and innovation know-how, our industry expertise and worldwide resources to working together with clients to make their businesses stronger. With more than 40 global delivery centers and 58,000 employees as of March 31, 2008, we combine a unique onsite/offshore delivery model infused by a distinct culture of customer satisfaction. A member of the NASDAQ-100 Index and S&P 500 Index, Cognizant is a Forbes Global 2000 company and a member of the Fortune 1000 and is ranked among the top information technology companies in BusinessWeek's Info Tech 100, Hot Growth and Top 50 Performers listings.

Start Today

For more information on how to drive your business results with Cognizant, contact us at inquiry@cognizant.com or visit our website at: www.cognizant.com.

Cognizant
Passion for building stronger businesses

World Headquarters
500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277
Email: inquiry@cognizant.com

European Headquarters
Haymarket House
28-29 Haymarket
London SW1Y 4SP UK
Phone: +44 (0) 20 7321 4888
Fax: +44 (0) 20 7321 4890
Email: infouk@cognizant.com

India Operations Headquarters
#5/535, Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060
Email: sdc@chn.cognizant.com