



SANbox 8000 Deployment Guide

Deploying VMware[®] ESX Server[™]

In the SANbox[™] 8000/SVM Environment

April 23, 2007

Contents

Overview	3
Supported VMware ESX Versions.....	3
SANbox 8000/SVM Software Versions.....	3
VMware Configuration	4
VMware Advanced Settings.....	4
Disk.UseDeviceReset.....	4
Disk.UseLunReset.....	4
Disk.MaxLUN	4
Most Recently Used Multipathing	4
VMware Startup Profile.....	4
QLogic Driver Parameters	4
SVM Configuration.....	6
User-defined Host SCSI Personality.....	6
Primary Switch.....	6
SANbox 8000 Configuration	7
Front-side paths.....	7
VMware Administration	7
Rescan SAN Operations.....	7
Avoiding ESX Server Reboot When Resizing Volumes	8
VMware ESX Disk Naming.....	8
VMware ESX Reboot and Active Paths.....	9
Installing and Booting VMware ESX Server 2.5 from the SAN.....	9
VMware ESX 3.0 Requirements	10
Known Issues With VMware	11
Trademarks.....	12
Notices	12
Contacting Support	12

Overview

SANbox 8000 interoperates with a wide variety of SAN switches, host platforms, and storage devices. The deployment guide provides specific recommendations to help you deploy servers running VMware ESX Server virtual infrastructure software in a SANbox 8000/SVM environment.

This document assumes that you are familiar with the use of both the SANbox 8000/SVM solution and the VMware product. The *SANbox 8000 Storage Services Platform Installation and Configuration Guide* contains general information on SANbox 8000/SVM concepts and operation. The SANbox 8000 SSP release notes provide the most up-to-date information on platform capabilities, supported devices, and known issues.

Consult the “Warranty and Technical Support” section of the *SANbox 8000 Storage Services Platform Installation and Configuration Guide* for details on obtaining technical support for any issues you encounter while using the SANbox 8000/SVM solution.

Supported VMware ESX Versions

The following VMware ESX and Virtual Center versions are supported by the current SANbox 8000 ReadyPath software release:

- VMware ESX Server 3.0.1, build 32039
- VMware ESX Server 2.5.2, build 16390
- Virtual Center 2.0.1, build 33643

Older VMware ESX releases (2.1.2, 2.1.1, 2.1.0 and 2.0) are not supported but may work. It is recommended that VMware ESX users upgrade to version 2.5.2/3.0.1 or later.

SANbox 8000/SVM Software Versions

The following QLogic ReadyPath software versions (or newer) are required for proper interoperability with VMware ESX Server:

- ReadyPath Base OS 1.1.9-11
- ReadyPath Patch 1.1.9-11 Patch C
- ReadyPath Maintenance 2.0.1-0

The following application software versions (or newer) are required:

- SANbox 8000 SVM Agent 4-4.59-3 (installed on SANbox 8000 SSP Platform)
- SVM Appliance Software V4R4-73 (installed on SVM Appliance)

VMware Configuration

Some of the VMware setting must be changed for proper operation with SANbox 8000.

VMware Advanced Settings

For VMware ESX to function properly with SANbox 8000, several of the VMware “Advanced Settings” must be changed from their default settings.

Disk.UseDeviceReset

The Disk.UseDeviceReset setting must be set to ‘0’. This setting causes VMware to not send a target reset to the SANbox 8000 port when initiating a failover, allowing the failover to be done on a more granular, per-LUN basis (see Disk.UseLunReset below).

Disk.UseLunReset

The Disk.UseLunReset setting must be set to ‘1’. This setting causes VMware to send a LUN reset when initiating a failover.

Disk.MaxLUN

This value defines the largest LUN value VMware will scan for on each remote target port. The Disk.MaxLUN setting should be set to the largest LUN exported from the SANbox 8000 to the VMware server (up to the maximum ‘128’). If this is not set correctly, the VMware server will only discover the first 8 (the default) volumes.

Most Recently Used Multipathing

Currently SANbox 8000 only supports VMware with ‘Most Recently Used’ setting for Multipathing. The ‘Preferred Path’ setting is not supported.

VMware Startup Profile

When configuring the VMware startup profile, the Fibre Channel HBA used to access the SANbox 8000 virtual volumes should be set to ‘Dedicated To Virtual Machines’ and not ‘Shared with Service Console’.

QLogic Driver Parameters

Several QLogic driver settings may need to be modified for VMware to work correctly with the SANbox 8000 virtual volumes. This is done through the following procedure on ESX 2.x:

1. Login as ‘root’ to the ESX server command-line console
2. Edit ‘/etc/vmware/hwconfig’

WARNING: Incorrectly editing this file may cause your VMware subsystem not to start. Please proceed with caution.

3. Search for the first instance of the QLogic HBA and note the device number (e.g. ‘1.1.0’)

4. Locate the line 'device.lilo_name.x.y.z.owner = "VM"' where lilo_name is the name of the VMware ESX server and x.y.z is the device number from the previous step. For example:

```
device.esx.1.1.0.owner = "VM"
```

Add the following line underneath this line:

```
device.esx.1.1.0.options = "qlport_down_retry=10 qlloop_down_time=90"
```

5. Save the changes to this file and reboot the server.
6. Repeat this procedure on each ESX Server that is accessing SANbox 8000 virtual volumes.

This is done through the following procedure on ESX 3.x:

1. Login as 'root' to the ESX server command-line console
2. Edit '/etc/vmware/esx.conf'

WARNING: Incorrectly editing this file may cause your VMware subsystem not to start. Please proceed with caution.

3. Search for lines like this for each HBA:

```
/device/010:01.0/name = "QLogic Corp QLA231x/2340 (rev 02)"
```

4. Add option settings to the following line:

```
/device/010:01.0/options = "qlport_down_retry=10 qlloop_down_time=90"
```

5. Save the changes to this file and reboot the server.

Tip: Both values are by default set to 30. By decreasing the time the driver tries to reestablish the link if the port goes down, the failover occurs sooner in case of a path failure.

Tip: You can verify if the settings have been applied successfully using the following command:

```
cat /proc/scsi/qla2300/x
```

Where x is the number of your vmhba (typically 0 and 1 or 1 and 2).

SVM Configuration

User-defined Host SCSI Personality

Each VMware ESX server must be defined as a “User Defined Host” (UDH). A UDH is a grouping of discovered HBAs that are installed in a single physical server and can have a specific “SCSI Personality” defined for the virtual volumes that are presented to it. A “SCSI Personality” defines the way in which the virtual volume behaves to certain commands, especially with regard to volume failover. The correct SCSI personality to use with VMware ESX Server is the ‘IBM FAStT Personality’ (not ‘AVT’).

Follow the following procedure to set the SCSI personality for volumes exported to the VMware ESX servers:

1. From the SVM GUI, select Volume Manager->Host->Manage
2. Click on the ‘Manage’ icon (small icon on right-hand side) for the UDH of the VMware server
3. Click on the ‘SCSI Personality’ button
4. Choose ‘IBM FAStT Personality’ from the pull-down box and then click ‘Confirm’

This will cause all virtual volumes exported to this server to be presented with the FAStT personality.

The SANbox 8000 operates like an IBM FAStT. Refer to the IBM Redbook for additional details regarding implementing VMware ESX server with IBM FAStT, especially section 8.10:

<http://www.redbooks.ibm.com/redbooks/pdfs/sq246434.pdf>

Primary Switch

When creating virtual volumes, the SVM alternates the default ‘primary switch’ between the two available SANbox 8000’s. When exporting multiple virtual volumes to a single VMware server this results in the server seeing half of the volumes as active on one SANbox 8000 and half active on the other SANbox 8000. During testing this was found to cause problems for the VMware server, particularly during reboots and failover.

To avoid this situation, the primary switch for all volumes exported to one VMware server should be set to the same SANbox 8000 unit. This is done through the following procedure:

1. From the SVM GUI, select Volume Manager->Volume->Manage
2. For each volume exported to the VMware server, select the ‘Manage’ icon (small icon on right-hand side)
3. Click on the ‘Set Switch Primary/Secondary’ button
4. Change the ‘Primary’ and ‘Secondary’ radio buttons as appropriate and click ‘Continue’ and then ‘Confirm’
5. Once the primary switch has been changed, the volume may be indicated in orange color; select the ‘Manage’ icon again and click on the ‘Fail Back’ button

Note that operations such as removing and adding back permissions for a volume will often result in a different ‘Primary switch’ setting and requires correction.

On initial volume setup, choose the switch connected to the lower numbered VMware 'vmhba' as the 'Primary switch'. This helps produce consistent behavior on the first VMware scan for the volumes.

1. From VMware management console, bring up Options->Storage Management and select the 'Adapter Bindings' tab.
2. Note the MAC address of the lower numbered vmhba. (8 colon separated hex numbers in parentheses.)
3. From the SVM GUI, select Volume Manager->Host->HBA Manage. Note the switch name connected to that HBA and set it as the 'Primary Switch' after all volume operations.

Load balancing can be achieved with multiple VMware servers on a SAN if connections are arranged so that the 'Primary switch' alternates between the two SANbox 8000s.

SANbox 8000 Configuration

Front-side paths

Each VMware ESX server may have one or two paths between it and each SANbox 8000. Two paths imply cross connections between the frontside SAN switches and SANbox 8000s. VMware denotes paths as "preferred" in some places but this is not relevant when using the MRO path selection method. VMware will only force path changes when doing an initial scan or when the active path has gone down.

In an HA configuration (with 2 SANbox 8000s) using non-cross connected frontside, each VMware ESX server will see only one active and one standby path for all volumes.

In an HA configuration with cross connected frontside paths, each ESX server will see one active and three standby paths for all volumes.

Cross connected frontside connections can be used to prevent a ping-pong situation with shared volumes when one server can see a volume through one SSP and a second server sees it through the other SSP.

VMware Administration

This section describes "best practices" for storage administration with SANbox 8000 and VMware ESX Server.

Rescan SAN Operations

It is recommended that when ever a change is made to the front-side SAN that the 'Rescan SAN' operation be performed on all ESX servers. This is particularly important after recovery of a path failure or when an SSP is replaced. If Rescan SAN is not performed, the ESX server may not know about new available paths and will operate in a single path mode.

The ESX GUI or `vmkmultipath` command should be used to verify that all expected paths have been discovered and are marked as available.

Avoiding ESX Server Reboot When Resizing Volumes

SVM volumes can be extended in a way that avoids rebooting the ESX server and integrates well with Windows disk management. These instructions are for ESX 2.5, but ESX 3.0 has GUI operations that are very similar.

The following operations have only been tested on Windows Server 2003 hosts but may be applicable to other host OSes. Only virtual disks in “persistent” data mode have been tested.

Virtual machines running on the server must be powered off if they have hard disks that access SVM volumes being modified (VMware requirement). The two resize methods described here depend on whether SVM volumes are presented directly to virtual machines or are used to create VMFS volumes on the ESX server.

The first method is for SVM volumes added to virtual machines as “System LUN/disk” devices. Simply expand an SVM volume while it is online. Ignore the fact that “Rescan SAN” in ESX does not immediately show increased storage. Power up the virtual machine that uses the volume and open “Disk Management”. New storage will show up as “Unallocated” space on the corresponding disk entry (*) and can be formatted as a new partition or used as dynamic volume space. “Rescan SAN” will report the proper server disk size once the space is allocated.

SVM volumes used to create VMFS volumes need additional steps. After powering off the virtual machine(s), remove volume permissions to the ESX server in the SVM. “Rescan SAN” and verify that the server disk (vmhba:y:z) has gone away. Expand the volume in the SVM and give permissions back to the ESX server. “Rescan SAN” and verify that the original server disk is back with free space equaling the amount added to the volume.

You can seamlessly use the new space to extend an existing VMFS volume and disk image files on that volume. First, format the new space as a VMFS volume. Edit an existing VMFS volume (presumably on the same server disk) and carefully check the “span” box next to the new volume to merge the two volumes. Increase the size of disk image files on that volume with the following console command:

```
vmkfstools --extendfile 12000M /vmfs/V_stripe_2/D_stripe_2.dsk
```

(This command takes the path to an image file and the total desired size, not the incremental size. ESX 2.5 uses a .vmdk extension for image files.)

Once powered up, the virtual machines will show new, unallocated space on the disk (*).

(*) This can be a “basic” or “dynamic” disk. A basic disk can be converted to dynamic in order to add the new space (now a dynamic volume) to existing dynamic volumes.

VMware ESX Disk Naming

VMware ESX 2.5 assigns a disk name (canonical path) to each SVM volume that can change when the server reboots. This is especially the case if reboot occurs with degraded SAN paths. Some precautions should be taken to minimize disruption this can cause.

The disk name takes the form: vmhba<#>:<#>:<LUN #>. For every SAN volume in a multipath setup, there are multiple equivalent names where LUN # is constant and the other part can change.

A good way to document current names is with the console command: `vmkmultipath -q`. Save results of this command in a file once storage is configured and stable.

Review disk names after a reboot and investigate the following two problem areas for changed disk names.

Virtual machine configurations (GUI or /root/vmware/*/*.vmx) can have stale name references for hard disk definitions (scsi0:*). “System LUN” virtual disks are most at risk but virtual disk files (*.dsk and *.vmdk) can also be wrong if symbolic VMFS names were not available during initial configuration. (Take the opportunity to implement symbolic references during VM downtime.)

The virtual machine swap file definition (Options->Swap Configuration) may break if the file is on a SAN volume. An incorrect name will prevent activation of swap and cause an error message after boot. Redefine swap using the new disk name and the original swap filename (*.vswp). SAN boot configurations are most prone to this.

Automatically booting virtual machines is not recommended unless naming has been confirmed to be stable through reboot.

It is useful for new names to match existing names when adding volumes to the server. This can be influenced before “Rescan SAN” by executing the console command: `vmkfstools -s vmhba<#>`, with appropriate number. Then execute “Rescan SAN” to complete the operation and see if the desired names got used.

VMware ESX Reboot and Active Paths

When a server reboots, disk names (see VMware ESX Disk Naming) reflect preferred paths and are forced to be the active paths. (This is the only time ESX seems to force a path.) These initial VMware preferred paths may not match the SVM primary switch setting.

Non-matches cases show up “degraded” in the SVM GUI. The best remedy for this is to keep Primary Switch selections in the SVM set for optimum load balancing and perform the Change Active Switch operation on any volumes that are degraded by a server reboot.

Installing and Booting VMware ESX Server 2.5 from the SAN

The following experiences were gained from installation on a system using Qlogic adapters. Future server ESX versions and other hardware configurations may behave differently. This must be tested in a non-production setting for each environment.

- At least one adapter in a pair must be enabled to run a boot BIOS on power up. Both can be enable if redundancy is desired and the considerations detailed below are taken.
- The ESX CDROM installer must be intercepted at the initial text prompt and continued with the “bootfromsan” option.
- Adapters must be visible to the SAN early in the install process so the SVM can be used to define the host and create/export the boot volume. This can be done when the installer is waiting at the first graphical prompt (after adapter drivers get loaded). Set the primary switch according to the adapter used by the boot BIOS.
- The installer creates a complex partition structure on the boot disk that does not allow the included vmfs partition to be resized. Unless accurate provisioning can be done at install time, size the boot disk vmfs partition for handling VM swap files and add manageable SVM volumes later for virtual machine disks.
- Removing SVM volumes from the server will cause the “Rescan SAN” operation in the ESX GUI to generate spurious error messages until the server is rebooted. These can be ignored.
- Creating a Virtual Disk in “System LUN/Disk” mode does not work. Trying to create a virtual machine hard drive with metadata fails in the ESX GUI and creating it without metadata fails on virtual machine power up (Windows).

Dual boot paths must be used with care. Refer to the section on ESX 2.5 Volume Management for important information about ESX disk naming and multipathing. A change in boot path can cause the “SAN Disk” name of the boot disk to change however the virtual machine swap file definition does not get updated automatically. It needs to be manually corrected if the disk name for the volume containing swap comes up changed after a reboot. The need to do this manually means virtual machines should NOT be set to power up automatically on reboot.

VMware ESX 3.0 Requirements

VMware 3.0 has a new management console. Most of the differences are cosmetic but one is critical to the operation of shared volumes or volumes that might need to move from one server to another. The following configuration should be policy for all servers since this potentially applies to all SAN volumes,

Under ConfigurationTab/Advanced settings/LVM there are two settings related to snapshotting and resignaturing. LVM.DisallowSnapshotLun must be changed to “0” (default is “1”) and LVM.EnableResignature must remain set to “0”.

All previously mentioned HBA driver and configuration settings apply in version 3.0.

Additionally the following hints and behaviors have been observed:

- “Rescan...” in Configuration/StorageAdapters gives the option to scan devices and filesystems in one operation. Do not do this. Scan for devices first and repeat until LUNs show up properly. Then scan for filesystems (if there are any on the new LUNs). In Configuration/Storage, check the Properties of VMFS volumes to make sure the expected paths show up.
- VMFS 3.x filesystems require a lot of overhead space. Experiment with creating SAN volumes that are larger (~1GB) than the total of virtual disk files you wish to create until you understand the overhead requirements.
- Creating a virtual disk file (.vmdk) for a VM from the GUI forces it to be named according to the VM name. This is bad for disk files on shared volumes or ones that might move around. Use the “vmkfstools -c ...” CLI command in order to choose any name. To maximize disk file size, create a file smaller than you want and use “vdf” and “vmkfstools -X” to extend the disk file until it consumes all available space (doing this through the GUI is very difficult).
- As with any ESX version, make the VMFS filesystem name match the SVM volume name if at all possible.

Known Issues With VMware

The following issues have been found during the testing and qualification of VMware ESX Server on the SANbox 8000 platform:

- BUG09983 VMware 2.5 SAN Boot does not support raw LUN
An ESX server booting from a SAN volume (shared HBAs) fails to allow creation of a virtual machine hard disk of type 'System Disk/LUN'. This appears to be a VMware issue.

- BUG09980 VMware 2.5 Volume Removal with SAN boot disk is unclear
Removing SVM volumes from a SAN containing the VMware server's OS boot volume does not happen cleanly. Stale references, notably entries in '/proc/vmware/scsi/vmhba2/0:*', to the deleted volumes remain on the server. This cause popup error messages when "Rescan SAN" is executed from the GUI which can probably be ignored. Rebooting the server clears up the problem. This doesn't happen if the ESX server is not booted from the SAN.

- BUG09795 VMware panic on boot after exporting CG view
When creating a View of a PIT made from a Consistency Group, and exporting that view to a second VMware host, the second host has been seen to occasionally panic on boot. This issue is under investigation with StoreAge (CQ# SVM4490).

- ENH09791 VMware failover takes longer with QLogic HBAs
QLogic HBAs appear to take longer to failover than do Emulex HBAs, especially under heavy I/O load. If you are having problems with long failover times, please insure the QLogic driver parameters are set as described in the VMware Deployment Guide.

- BUG09299 Port doesn't always come up when direct-connecting VMware server
Directly connecting VMware servers to SANbox 8000 (i.e. loop mode) is not currently supported.

- BUG09057 VMware ESX 2.1.x confused by two volumes with identical VMFS data
If VMware is presented with two VMFS volumes with identical data (e.g. a snapshot), one of the volumes will not be recognized properly. This is a VMware issue that will hopefully be resolved in future versions.

Trademarks

QLogic, the QLogic logotype, ReadyPath, and SANbox are trademarks of QLogic, Inc. Other trademarks appearing in this publication are trademark/service marks of QLogic, Inc.

All other brand and product names are trademarks or registered trademarks of their respective owners.

Notices

QLogic makes no warranty of any kind with regard to this material, including but not limited to, implied warranties of merchantability and fitness for a particular purpose and availability of products and/or their features.

Contact a QLogic sales office for information on feature and product availability.

QLogic reserves the right to make improvements and/or changes to its products and programs at any time without notice and shall not be liable for errors contained herein or for incidental, consequential damages in connection with the furnishing, performance, or use of this material.

No part of this document may be photocopied, reproduced, or translated into another language without the prior written consent of QLogic.

Export of technical data contained in this document may require an export license from the United States Government.

Contacting Support

Please feel free to contact your QLogic approved reseller or QLogic Technical Support at any phase of integration for assistance. QLogic Technical Support can be reached by the following methods:

Web: <http://support.qlogic.com>

North America Contact Information

Email: support@qlogic.com

Phone: (952) 932-4040

Support contact information for other regions of the world is available at the QLogic website.