

Deployment Guide: Emulex Virtual HBA Solutions and VMware vSphere 4



How to Deploy VMware vSphere 4 and ESX 4.0 Using Emulex
Virtual HBA and CNA Technology



Table of Contents

Introduction	3
Fibre Channel NPIV	3
NPIV Use Cases and Benefits	4
Requirements for NPIV	5
Raw Device Mapping	5
Creating a Virtual Machine with NPIV	6
Zoning with NPIV-Based Virtual Ports	8
LUN Masking with NPIV-Based Virtual Ports	9
Discovering VPorts Using HBAnyware	10
Conclusion	10
Appendix A - Converting a VMFS Disk Image to an RDM	11
Appendix B – Troubleshooting	12

Introduction

IT organizations can use VMware vSphere 4 to create a virtualized infrastructure of server resources which are dynamically provisioned and optimized. By consolidating servers, dramatic savings can be realized for capital expense (CapEx) for new equipment and operating expense (OpEx) for power, cooling, maintenance and system management.

The portability and recovery capabilities of VMware implementations rely on external shared storage, and are most effective with a Storage Area Network (SAN). A SAN provides scalable storage resources that support compelling VMware vSphere 4 features, including VMware VMotion, Distributed Resource Scheduler (DRS), High Availability (HA), and Consolidated Backup. Extending the SAN to remote sites also enables disaster recovery and business continuity.

The most popular SAN solution for VMware ESX Server and enterprise data center storage consolidation is Fibre Channel. The high performance delivered by the Fibre Channel protocol is best positioned to serve the higher I/O requirements for multiple virtual machines (VMs) running on a single server. SAN connectivity helps enable server virtualization, while server virtualization drives an increased need for SAN connectivity.

Fibre Channel NPIV

A major challenge for VMware storage administrators is maintaining the traditional Fibre Channel best practice to separate storage access using a combination of fabric zoning with the switch and LUN masking with the storage array. Both are configured using the Worldwide Node Name (WWNN) and Worldwide Port Name (WWPN) of host bus adapters (HBAs) or converged network adapters (CNAs) that connect host servers to the SAN. When VMs share the WWPN identity of a physical HBA or CNA port, there is no option to uniquely isolate, monitor and manage storage for an individual VM.

The solution to this challenge is N_Port ID Virtualization (NPIV), which allows a single physical HBA or CNA port to function as multiple virtual ports (VPorts) with each VPort having a unique identity in the SAN fabric. Using the Fibre Channel NPIV option in VMware vCenter Server, server administrators can create VMs that are identified with virtual WWNNs and WWPNs.

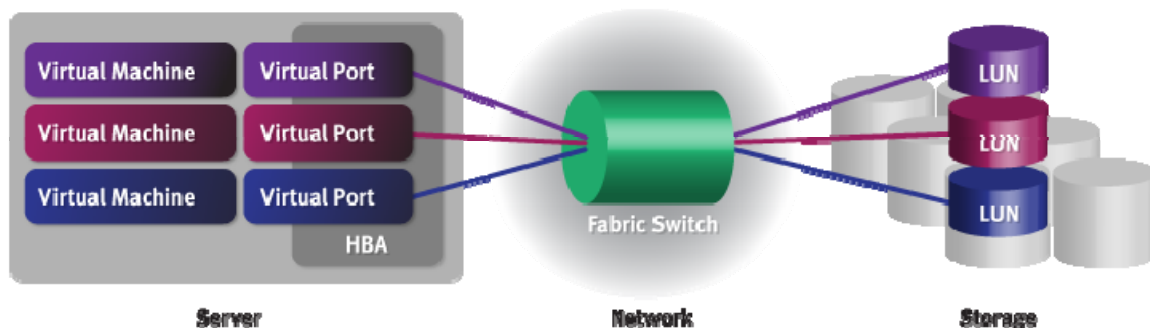


Figure 1: Virtual Port Attachment using NPIV

NPIV Use Cases and Benefits

NPIV is most valuable in managing storage access for mission-critical or SLA-driven applications. Use cases and benefits of using NPIV with VMware vSphere 4 include:

- **Application or User Level Chargeback** - I/O throughput, storage traffic and utilization can be tracked to the VM using the virtual WWPN, enabling application or user-level chargeback. Without NPIV, the SAN and ESX Server can only see the aggregate usage of all VMs that share the same physical Fibre Channel port (F_Port). The only exceptions are vendor-specific LUN-based tools.
- **LUN Customization and Tracking** - VMs can be associated to devices mapped under RDM, enabling LUN customization and tracking suited to the needs of an application. SAN tools can report VM-specific performance or diagnostic data based on the VPort associated with a VM. Switch-side reporting tools and array-side tools can also report diagnostic and performance-related data on a VM basis.
- **Bi-Directional Association of VMs with Storage** – Without NPIV, SAN administrators can only trace from a VM to an R, DM. NPIV provides the enhanced ability to also trace back from an RDM to a VM.
- **Support for SAN Management Tools** - Storage provisioning for ESX Server hosted VMs can use the same methods, tools and expertise in place for physical servers. With the VM uniquely related to a virtual WWPN, traditional methods of zoning and LUN masking can be used, enabling unified administration of virtualized and non-virtualized servers. Fabric zones can restrict target visibility to selected applications hosted by VMs. Configurations which required unique physical adapters based on an application can now be remapped to unique NPIV instances on the ESX Server.
- **Inter Virtual SAN Routing (IVR)** - Storage administrators can configure IVR in ESX Server environments based on an individual VM, enabling large end users to reconfigure their fabrics, aggregating islands of storage, fragmenting massive SANs into smaller, more manageable ones and assigning resources on a logical basis.
- **VM Migration** – VMware VMotion supports preservation of the VPort ID when a VM is moved to a new ESX server, which also improves tracking of RDMs to VMs. Access to storage can be restricted to a group of ESX Servers (VMware cluster) on which the VM can run or be migrated to. If the VM is moved to a new ESX Server, no changes in SAN configuration are required to use a different physical Fibre Channel port. When the zones and LUN masking are set up correctly, the VPort name will stay with the VM after migration to a new ESX server.
- **HBA Upgrades** – HBA upgrades, expansion and replacement are seamless. When SAN zoning and LUN masking are based on a virtual connection, physical adapters can be replaced or upgraded with minimal changes to the SAN configuration

Requirements for NPIV

- NPIV is supported with Emulex LightPulse® 4Gb/s and 8Gb/s Fibre Channel HBAs running Firmware Version 2.72a2 (or higher). NPIV is also supported with Emulex 10Gb/s CNAs.
- Fibre Channel SAN switches must be NPIV-enabled.
- The physical HBAs or CNAs on an ESX Server 4.0 host must have access to all of the LUNs accessed by VMs running on the host.
- VMs that use NPIV must access storage with Raw Device Mapping (RDM). RDM is described in more detail in the following section.

Raw Device Mapping

VMware ESX Server offers two options for managing disk access:

- VMware Virtual Machine File System (VMFS)
- Raw Device Mapping (RDM)

VMFS is a virtual, clustered file system that allows multiple ESX servers to access the same storage. With this architecture, multiple VMs on multiple ESX servers can share the same datastore and associated LUN. On-disk locking is used to ensure that a VM can only run on one server at a time. With multiple VMs using a common LUN, the effort required for LUN management is minimized.

RDM uses a mapping file inside VMFS that acts as a proxy for a raw device, allowing direct block-level access from a VM to a LUN. The mapping file is presented to the management software as an ordinary disk file, available for the usual file system operations. A separate LUN is required for each VM.

A VM can access a LUN presented through either RDM or VMFS. Both file systems can be used in parallel on the same ESX server.

Because the ESX implementation of NPIV requires RDM disk access, it's helpful to review the benefits and limitations of RDM. RDM provides the advantages of direct access to a physical device while keeping some of the benefits of a virtual disk in the VMFS file system. In effect, RDM combines VMFS manageability with raw device access. VMware VMotion, VMware DRS, and VMware HA are all supported with RDM.

Benefits of RDM include:

- **Storage-controlled isolation**— Each VM has its own a LUN, which is managed by traditional storage and fabric security tools.
- **Performance**—RDM allows disk access using array cache algorithms that are interacting with one VM only, as opposed to shared-LUN environments which diminish array cache effectiveness. Additionally, different storage arrays and/or disk types (e.g. Fibre Channel or

SATA, enterprise or midrange arrays) can be assigned to specific VMs, optimizing asset utilization and reducing total cost of ownership.

- **Physical-to-Virtual (P2V) Migration**—Once the VMware P2V utility builds the VM and loads the OS and application into the new VM, RDM allows the contents of the LUNs supporting the physical server to be accessed without copying them to a new location or converting them to a new format.

Creating a Virtual Machine with NPIV

The steps for creating a VM with NPIV are:

1. Using VMware vSphere Client, select the **Getting Started** tab and click **Create a new virtual machine**. You can also use the File option in the toolbar.
2. The Name and Location screen is displayed. Enter the name for the VM and click **Next**.
3. The Datastore screen is displayed. The datastore should be accessible from any server that might host the VM in the future. Select the datastore and click **Next**.
4. The Virtual Machine Version screen is displayed. Information on limitations and compatibilities with each version is displayed. Version 7 provides better I/O performance and is recommended if you will be exclusively using ESX 4.0 or later. Select an option and click **Next**.
5. The Guest Operating System, CPUs, Memory, Network and SCSI Controller screens follow in succession. Select an option and click **Next** for each screen.
6. The Select a Disk screen is displayed. Select **Raw Device Mappings** and click **Next**.
7. The Select and Configure a Raw LUN screen is displayed. With RDM, each VM is exclusively mapped to a specific LUN and a list of available LUNs is displayed. Select a LUN and click **Next**.
8. The Select a Datastore screen is displayed. Select the datastore for the LUN mapping and click **Next**.
9. The Compatibility Mode screen is displayed. The options for Physical or Virtual mode are displayed. NPIV can be used with either mode. Select a mode and click **Next**.
10. The Advanced Options screen is displayed. In most cases, there will be no need to choose any advanced options. Click **Next** to continue.
11. The Ready to Complete screen is displayed. Check **Edit the virtual machine settings before completion** to enable NPIV. Click **Continue**.

12. A new screen will display. Select the **Options** tab and then select **Fibre Channel NPIV**. The following screen will display:

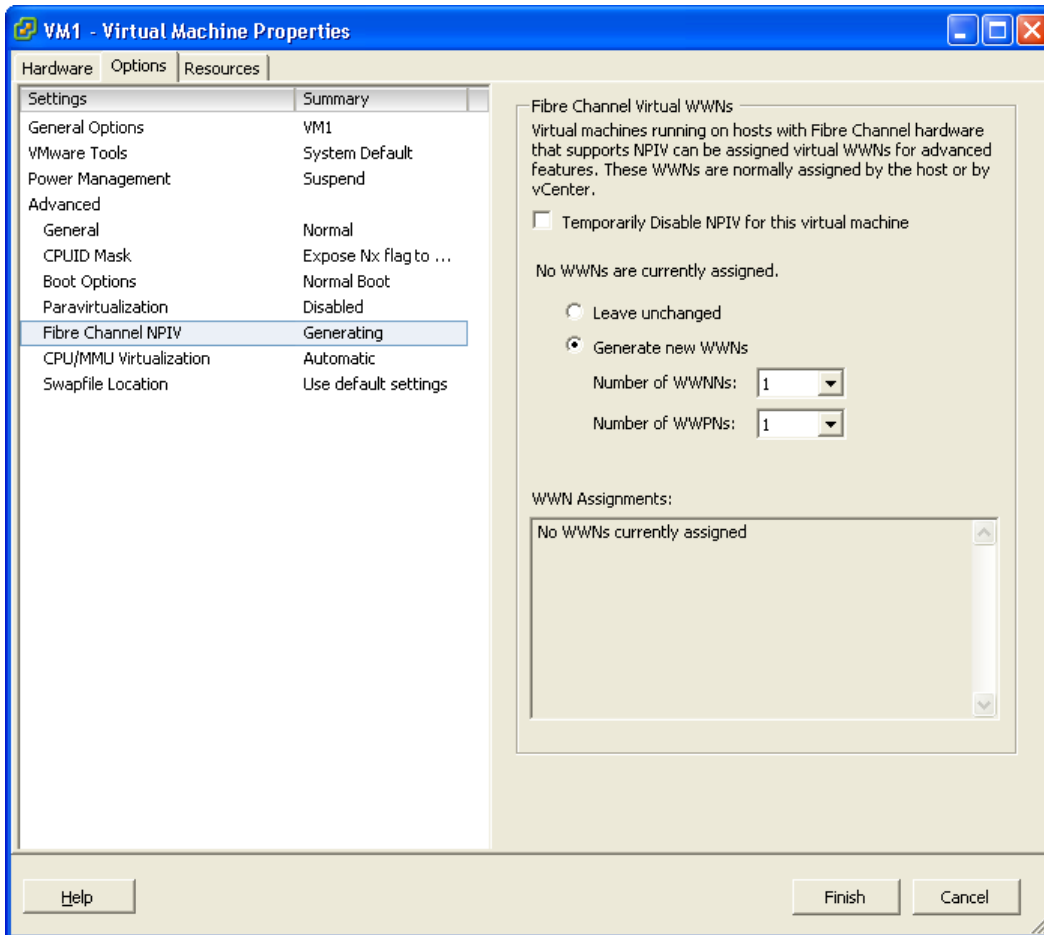


Figure 2: NPIV creation with VMware vSphere Client

Select **Generate new WWNs** and specify the number of virtual WWNNs and WWPNs to be assigned. For most cases, one virtual WWNN and one virtual WWPN will be sufficient. With multipathing, the virtual WWNN and virtual WWPN will be used with either of the physical paths (only one path is used at a time).

Note that the actual values for the WWNN and WWPN are not assigned until the VM is created. You will need to edit the virtual machine settings after the VM has been created to see the WWNN and WWPN values.

13. Click **Finish** to create the VM.

Zoning with NPIV-Based Virtual Ports

Zoning provides access control in a SAN topology by specifying the HBA or CNA ports that can connect to storage processors. There are two types of zoning:

- Hard zoning – Devices outside the zone are prevented access to devices inside the zone and isolation is based on the physical port connection.
- Soft zoning – Filtering is done in the Fibre Channel switch to prevents ports from being seen by devices outside the zone. Filtering is based on the WWPN, and the ports are still accessible if a user knows the WWPN.

With NPIV, VMs are uniquely related to a virtual WWPN, allowing the same zoning methods to be used for physical and virtual ports.

The following basic steps are used for VM-based zoning:

1. Group the physical ports for the host adapters and array targets (typically storage array controllers) into the same zone. This allows the host to have access to storage that will be used by the VMs on the host.
2. Create a zone using the VPort (virtual WWNN or WWPN) for one or more VMs and the storage array targets.

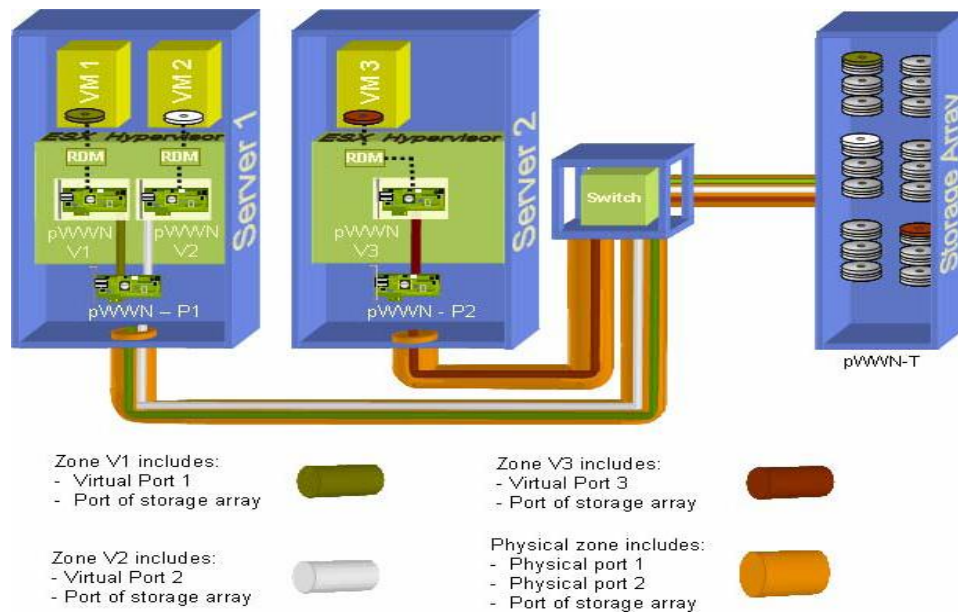


Figure 3: Implementing Fabric Zoning Using Virtual Ports

LUN Masking with NPIV-Based Virtual Ports

After configuring zoning, the next step is to use LUN masking on the array to control and manage storage for a VM. The application that's used for LUN masking will be specific to the array manufacturer, but the basic steps are:

- Mask all LUNs used by all VMs so they're available to the physical ports for HBAs or CNAs.
- Mask specific LUNs used by an individual VM using the VPort (virtual WWNN or WWPN).

It's critical to have the VPort-assigned LUN path visible to the VM at the time it powers up. If the LUN paths are not visible, ESX Server will destroy the VPort causing the driver to drop its login to the Fabric. To prevent this cycle, VPort WWNs should be programmed into the host groups and LUN masking configurations at the storage array prior to powering on an NPIV-enabled VM.

The following is a LUN masking example using an EMC CLARiON CX3 storage array. The LUN must be masked to the ESX server's physical HBA node and port address and to the NPIV-based VPort. Masking to the VPort must be configured manually since the array does not recognize the path automatically.

To manually mask an NPIV VPort:

1. Make sure the physical HBA on the ESX server is masked to desired LUNs on the array.
2. Create the VM and configure the RDM storage.
3. Enable NPIV for the VM in the configuration options.
4. Go into the ESX CLI and record both the WWNN and WWPN from the VM.vmx file (copy/paste works well if using a remote SSH shell).
5. Using the Navisphere GUI:
 - a. Right click on the array icon and select **Connectivity Status**.
 - b. Click **New**.
 - c. For **Initiator Name**, enter the NPIV WWNN and WWPN from the VM.vmx file using the format WWNN:WWPN as in the following example:
`2e:f4:00:0c:29:00:01:73:2e:f4:00:0c:29:00:02:73`
 - d. Select **Existing Host** and use the same host name that is currently used for the physical HBA path.
 - e. From the Storage Groups tab, select the storage group name and the host right click on the host and select "Connectivity Status". Click on the new host initiator path and select "Reconnect".

Steps 1 through 3 apply regardless of the array model. Please refer to array vendor-specific documentation on LUN masking for detailed configuration information.

Discovering VPorts Using HBAnyware

Storage and server administrators can use the Emulex HBAnyware® centralized management tool to view the storage path and server information related to a virtual port. Emulex HBAnyware is a powerful management platform that enables secure, centralized discovery, monitoring, reporting and administration of Emulex HBAs and CNAs on local and remote hosts. This cross-platform tool is available for download from Emulex's website.

HBAnyware 4.1 offers three views: Host View, Fabric View, and Virtual Port View. VPort information can be accessed from all three frameworks by expanding the adapter ports. VM information can be viewed for each VPort. This allows administrators to know VMs that will be affected by any changes to storage.

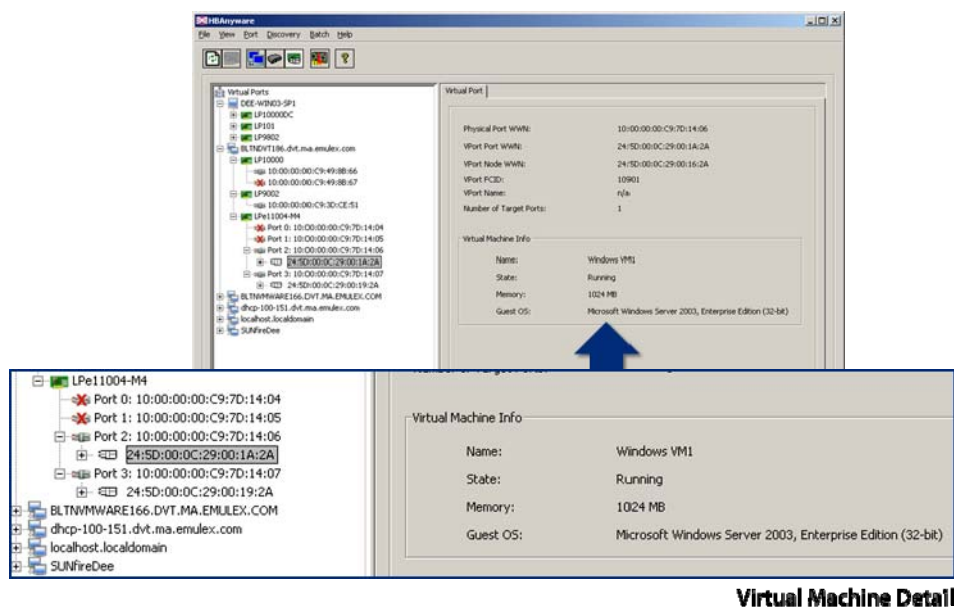


Figure 4: HBAnyware Virtual Port View

Conclusion

Using VMware vSphere 4, data centers can dramatically reduce CapEx and OpEx costs by consolidating applications from underutilized, and often outdated, servers to a much smaller number of virtualized servers. A Fibre Channel SAN provides scalable, high-performance storage that fully enables the agility and flexibility needed to optimize virtual server resources.

Emulex HBAs and CNAs work with VMware vSphere to provide SAN best practices support for individual VMs. Using NPIV, system and storage administrators can isolate and manage storage based on a virtual connection that is unique for each VM.

Appendix A - Converting a VMFS Disk Image to an RDM

NPIV requires an RDM-based virtual machine. A typical scenario will be converting a VMFS-based disk image to an RDM-based virtual machine disk (vmdk) that is assigned to a new NPIV-based VM. The following steps detail how this is done.

When the VM owning the vmfs file is powered-off, use vmkfstools to perform the vmdk disk creation using the following command:

```
# vmkfstools -i <from_disk> <to_disk> -d <rdm: | rdmp:> <device>
```

<from_disk>	Name of the existing vmfs disk file to be cloned
<to_disk>	Name of the new RDM-based disk to be created with the disk image
<rdm: rdmp:>	Disk type to map via vmfs
<device>	Raw device name of the SAN-based disk to which the contents of the disk image are written

The raw name is formatted as:

```
/vmfs/devices/disks/vmhba<Instance>:<Tgt>:<Lun>:<Partition>
```

vmhba<Instance>	VMware host bus adapter instance corresponding to the physical port that can see the SAN disk
<Tgt>	SCSI target ID of the SAN disk
<Lun>	SCSI LUN number of the SAN disk
<Partition>	Disk partition number

The following example is a single command line in which the VM to be updated is named “rhel3”:

```
# vmkfstools -i /vmfs/volumes/storage1/rhel3/rhel3.vmdk
/vmfs/volumes/storage1/rhel3-rdm.vmdk -d
rdm:/vmfs/devices/disks/vmhba4:0:0:0
```

Note: it is assumed that “storage1” is a shared vmfs store in the ESX cluster

Appendix B – Troubleshooting

The focus of this section is troubleshooting common VPort set-up problems.

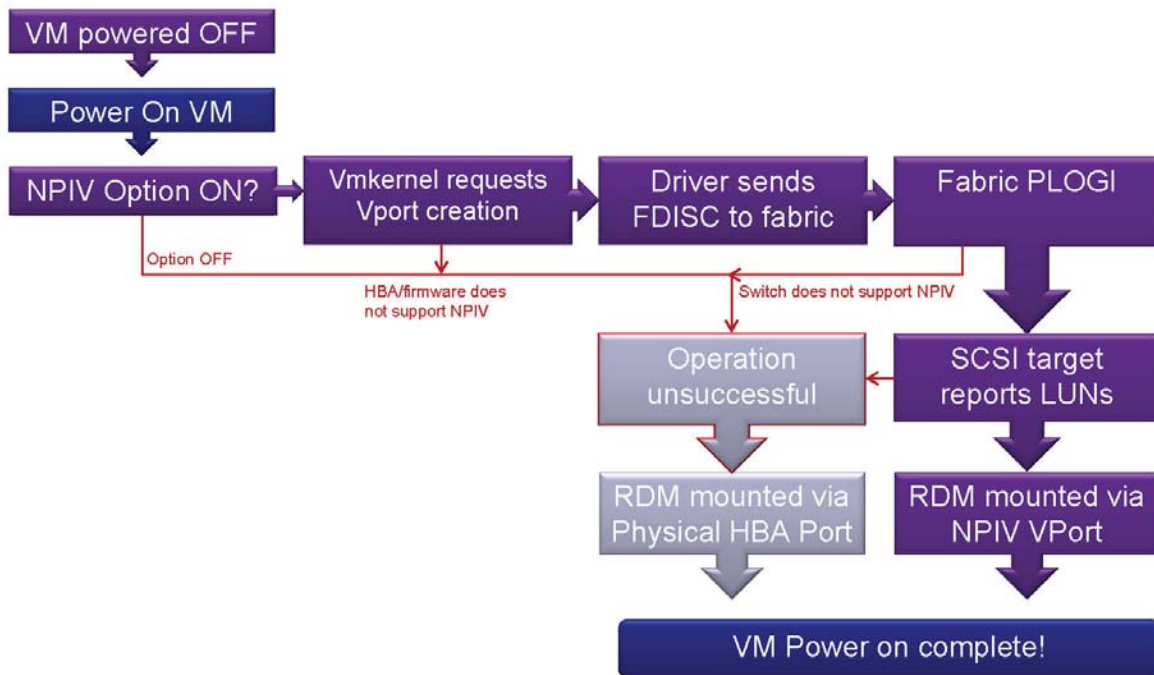


Figure 5: VPort Creation Flow

From the VMware ESX command line interface (CLI), cat the file `/proc/scsi/lpfc/x` (where x is the vmhba number). The output should look like the following:

```

Emulex LightPulse FC SCSI elx_7.4.0.13
Emulex LPe1150 4Gb lport FC: PCIe SFF HBA on PCI bus 05 device 00 irq 121
BoardNum: 5
SerialNum: VM61080357
Firmware Version: 2.70A2 (W3F2.70A2)
Hdw: 2057706d VendorId: 0xf0e510df
Portname: 10:00:00:00:c9:53:a3:59  Nodename: 20:00:00:00:c9:53:a3:59
  
```

SLI Rev: 3

```

NPIV Supported: VPIs max 7 VPIs used 1
RPIs max 512 RPIs used 15
  
```

Vports list on this physical port:

```

Vport DID 0x610b01, vpi 1, state 0x20
  
```

```

Portname: 24:94:00:0c:29:00:07:0f  Nodename:24:94:00:0c:29:00:06:0f
  
```

Link Up - Ready

```

PortID 0x610b00
  
```

```

Fabric
  
```

```

Current speed 4G
  
```

```
Current Mapped Nodes on Physical Port:
lpfc5t00 DID 610e00 WWPN 50:00:1f:e1:50:09:e2:9c WWNN
50:00:1f:e1:50:09:e2:90
lpfc5t01 DID 620400 WWPN 50:00:1f:e1:50:09:e2:9d WWNN
50:00:1f:e1:50:09:e2:90
```

Notice that the SLI section indicates whether or not the fabric supports NPIV. If the fabric did not support NPIV, the output would be:

```
SLI Rev: 3
NPIV Unsupported by Fabric
RPIs max 512 RPIs used 8
```

Frequently Asked Questions

I have enabled my switch for NPIV but the /proc/scsi/lpfc/x file still says NPIV is unsupported by the fabric?

This issue is typically resolved by rebooting the ESX server. It has also been observed that on some Brocade switches, a soft firmware reset after the firmware load may not enable the NPIV feature and that a switch power-cycle reboot may be required for new NPIV capable firmware to become effective.

How can I troubleshoot the NPIV fabric login process?

A Fibre Channel analyzer can be used to examine on-the-wire traffic. To determine which operations are being used on the ESX host, HBA logging can be enabled. Use HBAnyware to turn on the driver parameter `log_verbose` and set this parameter to `0x4047`. Then, view the log by using the ESX CLI using the following command:

```
#tail -f /var/log/vmkernel
```



VMware, Inc. 3401 Hillview Drive Palo Alto CA 94304 USA Tel 650-427-5000
www.vmware.com

Emulex Corp. 3333 Susan Street Costa Mesa CA 92626 USA Tel 714-662-5600
www.emulex.com

© Emulex Corporation. All Rights Reserved. No part of this document may be reproduced by any means or translated to any electronic medium without prior written consent of Emulex.

Information furnished by Emulex is believed to be accurate and reliable. However no responsibility is assumed by Emulex for its use; or for any infringements of patents or other rights of third parties which may result from its use. No license is granted by implication or otherwise under any patent, copyright or related rights of Emulex. Emulex, the Emulex logo, LightPulse and SLI are trademarks of Emulex.

VMware, the VMware "boxes" logo and design, Virtual SMP and VMotion are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

