# Cisco MDS and Emulex Virtual HBA Solutions for VMware Infrastructure 3

*A Technology Overview for SAN Connectivity using NPIV in a VMware ESX Server 3.5 Environment*

*Prepared by:*

*Cisco Systems, Inc., Emulex Corporation, and VMware, Inc.*

Scope of this Paper:
This co-authored technology overview provides data center users with valuable information for understanding the benefits of NPIV technology in a VMware environment and practical use cases and deployment scenarios for maximizing these benefits.

This guide is intended to introduce the concepts of NPIV in a VMware environment, using Emulex and Cisco products. General use cases are presented and possible benefits are mentioned. Actual customer configurations can vary widely along with results, and these deployment scenarios may not be possible or practical in any given case. Please consult with your Cisco, Emulex and VMware representative on specific questions based on your specific needs.

## Introduction

Virtualization. This one word speaks volumes and is first on the minds and the agendas of most IT professionals today – and it has to be. For the ever increasing service and resource demands placed on IT departments by users to remain manageable – or to *become* manageable – the abstraction that virtualization technologies provide is required. Virtualization technologies make complex systems manageable and virtualization inherently adds strong deployment and re-deployment flexibilities to data center operations.

IT organizations are well aware of the tremendous benefits to be realized in deploying server virtualization technology. According to InfoWorld, over 74% of IT organizations rated server virtualization as either critical or highly important technology. Among server virtualization solutions, VMware has become the de facto standard owing in large part to VMware's commitment to innovation.

The implementation of VMware servers is highly correlated with the adoption of storage area networks (SANs) to provide shared data access necessitated by shared server environments. Emulex and Cisco, two recognized leaders in Fibre Channel connectivity, are working closely with VMware to make sure that server virtualization leverages new capabilities that evolve in the Fibre Channel connected environments.

One of these new technologies that is available as part of VMware ESX Server 3.5, is support for the N-Port ID Virtualization standard (NPIV). Support of NPIV enables each virtual machine (VM) on a VMware ESX Server to have a unique Fibre Channel Worldwide Port Name (WWPN) providing an independent data path to the SAN through the Virtual HBA Port. By providing a unique Virtual HBA port, storage administrators will be able to implement SAN best practices such as LUN-masking and zoning for individual virtual machines. Administrators will also be able to take advantage of powerful SAN management tools and simplified migration of virtual machines and their storage resources.

Emulex and Cisco are collaborating with VMware on interoperability and end user education to help both the industry and end users understand the benefits of NPIV technology in a VMware environment. The purpose of this paper is to provide some fundamental insight into NPIV and to help end users discover the practical use cases and deployment scenarios for maximizing the benefits of this new capability.

## NPIV Technology

NPIV is the acronym for “N_Port ID Virtualization”, a T11 ANSI standard which was developed by Emulex and IBM, to provide the capability for a fabric switch to register several WWPNs on the same physical HBA port.

With the advent of server virtualization, we gain the benefit of multiple virtual machines replacing independent physical hosts, but we lose the direct port-to-host relationship. Using the capabilities of NPIV, that relationship can be restored and represented as a virtual port. In VMware Infrastructure 3 (VI3) ESX Server 3.5, each virtual machine can be assigned a unique WWPN, enabling NPIV and establishing the hosts’ WWPN as a unique Fibre Channel endpoint on the SAN. The per-VM WWPN restores the ability to employ traditional methods of zoning and LUN-masking in a virtual server environment.

## VMware in an NPIV environment

The following key technologies are required to support NPIV in a VMware ESX Server environment deploying Emulex Fibre Channel HBAs and Cisco MDS switch technology:

- VMware VI3 ESX Server 3.5 which has the capability to generate and assign a set of unique virtual WWPN to each VM running on a single ESX Server. By default it will try to assign 4 WWNs attaching to different physical ports for failover multipathing.
- Emulex LightPulse® Fibre Channel HBAs with the Virtual HBA technology feature. This functionality is available for all Emulex midrange and enterprise class LightPulse 4Gb/s Fibre Channel Host Bus Adapters installed with the appropriate Emulex VMware ESX Server drivers that will ship with VMware VI13 ESX Server 3.5. All Emulex 4Gb/s enterprise and midrange host bus adapters listed in the VMware I/O compatibility guide are supported, including LPe11000, LPe11002, LPe1150 LP11000, LP11002, LP1150, and equivalent OEM branded models and mezzanine cards for blade servers.
- Cisco MDS NPIV aware fabric switches with NPIV support. This functionality is available in Cisco’s MDS 9000 Family of fabric switches configured with SAN OS 3.0.(1) and above.

When VMware VI3 is deployed in a blade-based environment, Emulex Virtual HBA technology, providing VM-level connectivity within the ESX Server environment on a given blade, can be combined with Cisco NPV (N-Port Virtualizer) on a 9124e or equivalent switch, NPV providing managed connectivity of multiple blades within one or several chassis. The combination of Virtual HBA and NPV, although available, is not discussed in this document, which focuses on connectivity within a single VMware server or blade. Please refer to your Cisco contacts for more detail about NPV.

## NPIV Value Proposition

Clearly a key driver for any IT initiative must include the technical justification for implementing a new technology. So what is the value proposition for leveraging NPIV in a virtual server environment?

Below is a list of advantages when using NPIV within the VMware ESX Server. These benefits are listed in order of the timeframe in which they are available with ESX Server 3.5 and future versions and the relative benefit to the IT community.

- I/O throughput, storage traffic and utilization can be tracked to the virtual machine level via the WWPN, allowing for application or user-level chargeback. As each NPIV entity is seen uniquely on the SAN, it is possible to track the individual SAN usage of a virtual server. Prior to NPIV, the SAN and ESX Server could only see the aggregate usage of the physical FC Port by all of the virtual machines running on that system.
- Virtual machines can be associated to devices mapped under RDM to allow for LUN tracking and customization to the application needs. SAN tools tracking WWPN could report virtual machine specific performance or diagnostic data. As each NPIV entity is seen uniquely on the SAN, switch-side reporting tools, and array-side tools, can report diagnostic and performance-related data on a virtual machine basis.
- Bi-directional association of storage with virtual machines gives administrators the ability to both trace from a virtual machine to an RDM (available today) but also be able to trace back from an RDM to a VM (significantly enhanced with NPIV support).
- Storage provisioning for ESX Server hosted virtual machines could use the same methods, tools, and expertise in place for physical servers. As the virtual machine is once again uniquely related to a WWPN, traditional methods of zoning and LUN masking could continue to be used.
- Fabric zones can restrict target visibility to selected applications. Configurations which required unique physical adapters based on an application can now be remapped on to unique NPIV instances on the ESX Server.
- Virtual machine migration supports the migration of storage visibility. Access to storage can be limited to the ESX Server actively running the virtual machine. If the virtual machine is migrated to a new ESX Server, no changes in SAN configuration would be required to adjust for the use of different physical Fibre Channel ports. Additionally, the ESX Server requirement to open-zone all ESX Servers that may host the virtual machines, which also means all systems have access to all storage, is no longer required.
- HBA upgrades, expansion and replacement are now seamless. As the physical HBA WWPNs are no longer the entities upon which the SAN zoning and LUN-masking is based, the physical adapters can be replaced or upgraded without any change to SAN configuration.

Simply stated:

> *NPIV will enable storage and SAN fabric administrators to manage connections from virtualized machines in the same way, and with the same tools, as traditional physical hardware based servers.*

But how does this value proposition translate into the day-to-day management and operations of a data center?  This section takes a deeper look into some SAN management processes and tools, exploring the performance and management benefits available and the tools to track and account for NPIV WWPNs are provided. Helpful Hints provide additional insight as to practical considerations for production deployment of NPIV enabled solutions.  Diagrams provide visual profiles of typical implementations.

## *1. Fabric QoS*

Quality of service or QoS is an important requirement for data centers having specific commitments to their user community.

Using the Cisco MDS Fabric Manager console the administrator can assign a traffic priority level of high, medium or low, to any initiator WWPN (physical or virtual). In a local SAN where congestion issues are generally minimal, setting different priority levels to different ports will not bring any visible results.

In cases where traffic is routed to remote storage (for instance using DWDM or FCIP) this capability may be of great interest. For example, the same server may use remote mirroring for one VM, requiring very short response times, while another VM will be set up to send backup files to a remote backup server or virtual tape appliance. In this case, the user will want to assign a high priority to the virtual WWPN engaged in remote mirroring, and a low priority to the WWPN for backup.

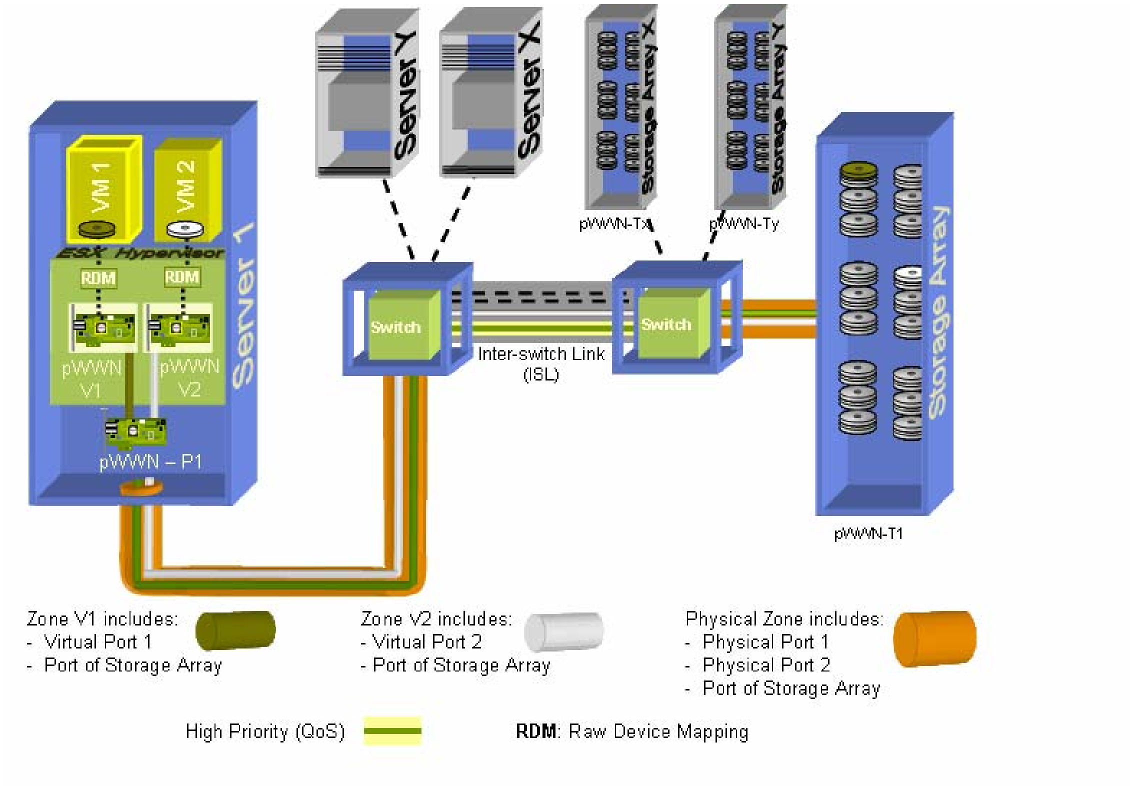

**Figure 1: QoS for VMs Using NPIV.**

QoS has become a valued SAN and network management capability in complex physical server environments. In the virtual data center world it becomes a fundamental capability in the SAN. The ability to assign service levels to VMs running on a single ESX Server via NPIV affords the SAN administrator yet another point of control, aiding management and service level assignments.

Figure 1 shows a test configuration where QoS is used to prioritize bandwidth allocation in cases where this bandwidth may be restricted, such as remote SAN access by IP or DWDM gateways. In this case the traffic generated by VM1 has higher priority with respect to the traffic generated by VM2, or by other devices sharing the common infrastructure. In a test configuration, a SCSI traffic generator like IOmeter can be used to evaluate whether the performance requirements are met.

## *2. Port-and-VM Level Statistics*

Virtual HBA technology offers the underpinnings for tracking traffic, packet sizes, error rates and other statistics at the virtual machine level. These capabilities could be integrated into future versions of VirtualCenter and other management frameworks, enabling users to monitor, troubleshoot and predict I/O patterns and evolution and thereby support service level agreements on I/O intensive applications even more effectively. This capability would also allow storage administrations to provide charge back based on I/O throughput as well as conduct performance monitoring of workloads running in virtual environments.

## 3. Bi-Directional Association of Storage with Virtual Machines

A common request by storage administrators is to be able to trace back which users or workloads are accessing a particular LUN. In a virtual environment, there may be many workloads running on a single physical server and this can make this process much more difficult to navigate. When a virtual machine is using a dedicated LUN, that is a raw device map (RDM), there is an easy way to determine which LUN is associated with a given VM. However, there is no simple way to go from a LUN back to which VM is associated with it prior to NPIV.

Support for NPIV with ESX Server provides bi-directionally traceability. Specifically, administrations will not only be able to trace from a virtual machine to an RDM (available today) but also be able to trace back from and RDM to a VM (significantly enhanced with NPIV support). Bi-directional association facilitates better resource allocation visibility for both SAN and virtualization administrations teams; thereby providing SAN administers with the ability to use best practice approaches to storage allocation and management, decreasing disruption, reducing training costs and enhancing administrative efficiency.

Note: In a VMware storage environment using VMFS, the traditional one-to-one association between HBA WWN and LUN WWN does not apply as there are many VMs that might be sharing a single shared LUN.

## 4. Fabric Zoning with NPIV

A key SAN management best practice is fabric zoning. Fabric zoning or the partitioning of a Fibre Channel fabric into smaller subsets increases security and simplifies management. Zoning provides access control of the SAN. When a SAN is configured for zoning, the devices outside a zone are not visible to the devices inside a zone. Also SAN traffic within each is isolated from the other zones. In case a storage device doesn’t support LUN mapping and masking, zoning is the only available tool for the storage administrator to limit access to a given target to a given initiator. In case the storage device supports LUN masking and mapping, the common practice is to use both zoning and LUN mapping and masking.

Zoning requires use of the HBA Worldwide Port Name to identify the data path. In a future version of VI3 with NPIV support, zoning will be used to isolate VMs running on the same ESX Server from one another.

**<u>Helpful Hint:</u>**
Since all physical HBAs and array controllers must be visible to VMkernel, an array port will require two zones:

1. A working zone including:

    - The virtual port linked to a VM
    - The array port for the LUN used by the VM

2. A control zone including:

- All physical HBA ports on the ESX Server
- All array ports attached to that ESX Server

Also note that with zoning and VMotion the control zone must include all physical ports you want to migrate to. More details on VMotion are available at http://www.vmware.com/products/vi/vc/vmotion.html.

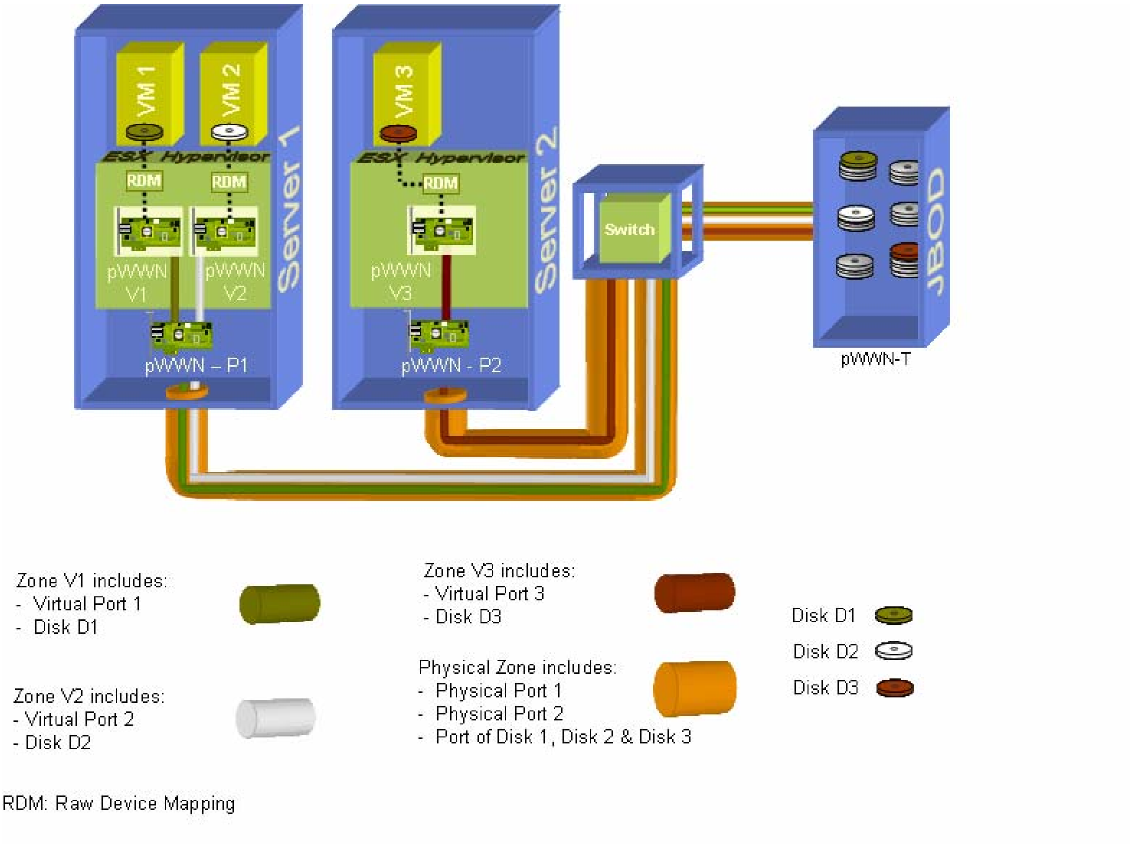

**Figure 2: Virtual Machines using NPIV. Configuration based on zoning for a device (e.g. a disk inside a JBOD) that doesn't support LUN mapping and masking.**

Figure 2 illustrates three VMs on two servers. Zoning is used to provide isolation for different users, applications or departments. In this case each individual disk is placed in an individual zone together with the related virtual initiator. The zone P includes physical devices, that are the physical HBAs (pWWN-P1, pWWN-P2) and the disks (pWWN-D1, pWWN-D2, pWWN-D3). Each zone Vx includes the virtual HBA (pWWN-Vx) and the physical disk Dx (pWWN-Dx), only.

For best results in HBA swapping scenarios, it is further advised to configure the zone P using “soft zoning”, meaning that the zone is based on port numbers rather than burnt-in WWN.

## 5. VSAN (IVR) Routing and Trunking

Virtual SANs (VSANs) enable large end users to reconfigure their fabric in many important ways: aggregating islands of storage, fragmenting massive SANs into smaller, more manageable ones (and avoiding limitations such as switch count and number of hops), and assigning resources on a logical basis (e.g., department, or application such as backup) . Furthermore, a variety of fabric events such as RSCNs are contained within one VSAN, isolating incidents and increasing overall fabric uptime. These and other reasons have resulted in a very high deployment rate of VSANs especially across very large users and fabrics.

In a VMware environment, different VMs on a given physical server may need to access storage on different VSANs. Prior to Virtual HBAs, this was only possible by configuring access to all VSANs from all virtual machines, a cumbersome solution that also negates some of the benefits of VSANs.

Virtual HBA technology enables selective routing of Virtual Machines to the desired VSAN.

- Physical HBAs and all virtual ports belong to one VSAN (as VSANs are based on the physical switch ingress port). Whenever possible, users will likely select a default VSAN which also hosts storage for a majority of VMs.

- Selected virtual ports can be configured for routing to different VSAN using IVR.

- VSAN-based fault isolation and enhanced security is enabled at the individual VM level.

End users will be able to leverage NPIV based Inter-VSAN Routing in conjunction with a future release of ESX Server.

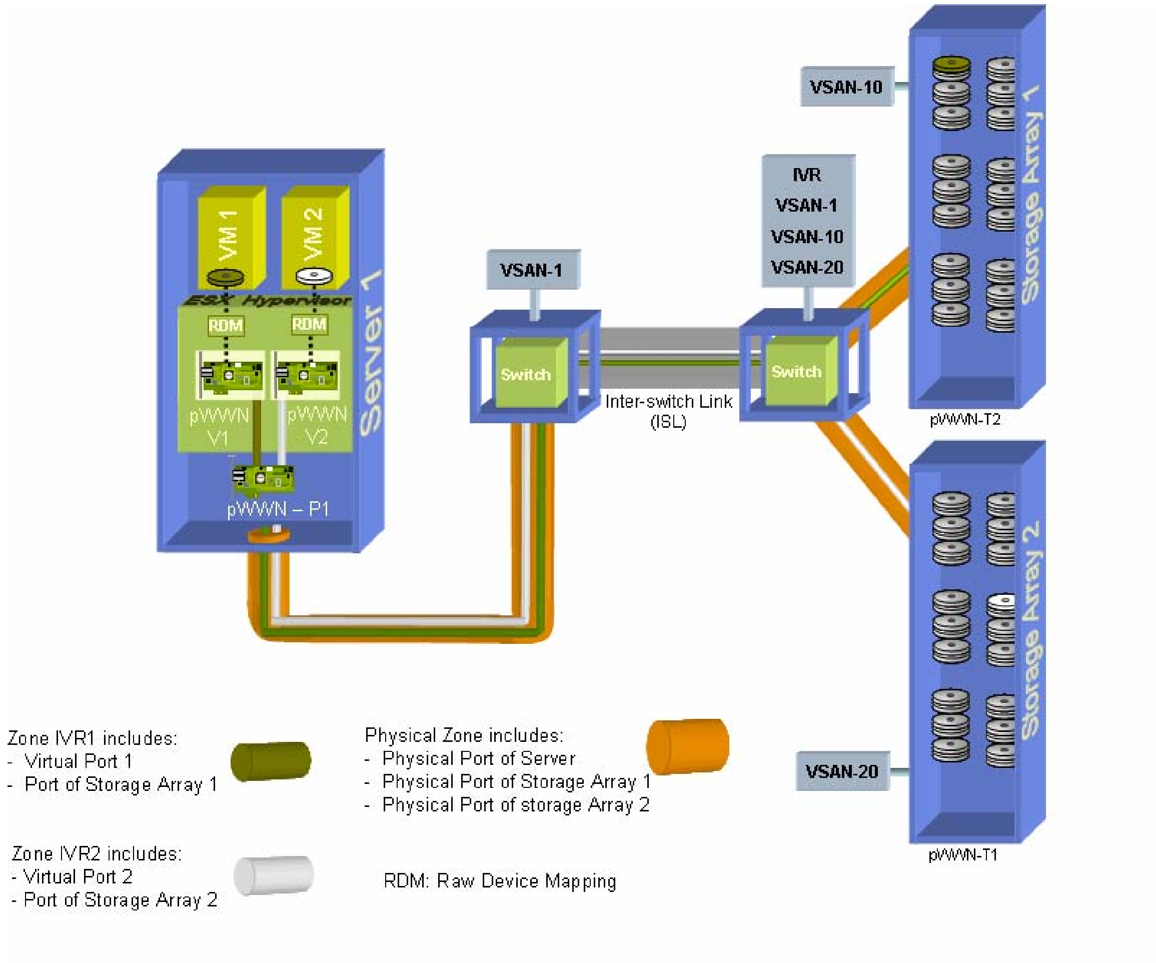

**Figure 3: Routing Virtual Machines across VSANs using NPIV and IVR. In this figure the targets are in different VSANs.**

Figure 3 illustrates how NPIV, coupled with IVR, enables multiple VMs, sharing the same physical server, being routed to different VSANs for security, resiliency and incident isolation.

**Inter-VSAN Routing Zoning:**
The IVR zone IVR-P includes the physical devices, that are the physical HBAs (pWWN-P1) and the physical port of the storage arrays (pWWN-T1, pWWN-T2)

Each IVR zone IVRx includes the virtual machine “x” (pWWN-Vx) and the physical port of the storage arrays (pWWN-Tx) only.

**LUN Mapping and Masking:**
Each LUN “x” is exposed to the physical initiator pWWN-P and to virtual machines “x” pWWN-Vx only.

## 6. Array Level LUN Masking, Mapping and VMotion

SAN best practices dictate the use of a common model to best manage, protect and secure data in a Fibre Channel connected storage area network. Key practices include:

- **Logical Unit Number Masking** or LUN masking is an authorization process that makes a Logical Unit Number available to some hosts and unavailable to other hosts. LUN masking is usually used to protect data corruption caused by misbehaving servers.

- **LUN Mapping** refers to the conversion between the physical SCSI device (storage) logical unit number and the logical unit number of the storage presented to operating environments.

Implementation of these SAN best practices typically requires the use of an HBA Worldwide Port Name as a means of specifying the data path between the server and the storage LUN. With earlier versions of VMware ESX Server, as in other operating systems, this data path was predicated on the physical WWPN of the HBA installed in the server. As this physical WWPN was shared across a number of virtual machines, SAN tools had no ability to associate a data path to a given VM.

One of the very compelling features of VMware Infrastructure 3 is VMotion. VMotion lets you move live, running virtual machines from one host to another while maintaining continuous service availability. Neither users nor the application's VM know it has been moved. Applications don't have to be taken off line. VMotion is also a huge advantage when doing hardware maintenance. With ESX Server leveraging NPIV, a Virtual HBA or VPort can be created for each VM. In VMotion, virtual machines can be migrated between different ESX Servers together with their virtual ports. Figure 4 and Figure 5 illustrate how LUN mapping works in this ESX Server environment with NPIV-aware VMotion enabled. Note that the requirements and behaviors for LUN masking vary by array.

**Helpful Hint:**
VMkernel must have visibility to all the LUNs used by all the virtual machines on the system. However, VMkernel does not have write access to any LUNs because all writing is performed by VMs exclusively. For practical purposes it is recommended to set up and mask each LUN to make them visible to the physical HBA WWPN (and thereby to VMkernel) and the virtual WWPN associated with the VM that will use it, blocking out all other virtual WWPNs and VMs.

It is common practice to combine the use of LUN masking and mapping with zoning to achieve an additional level of isolation between users or applications.

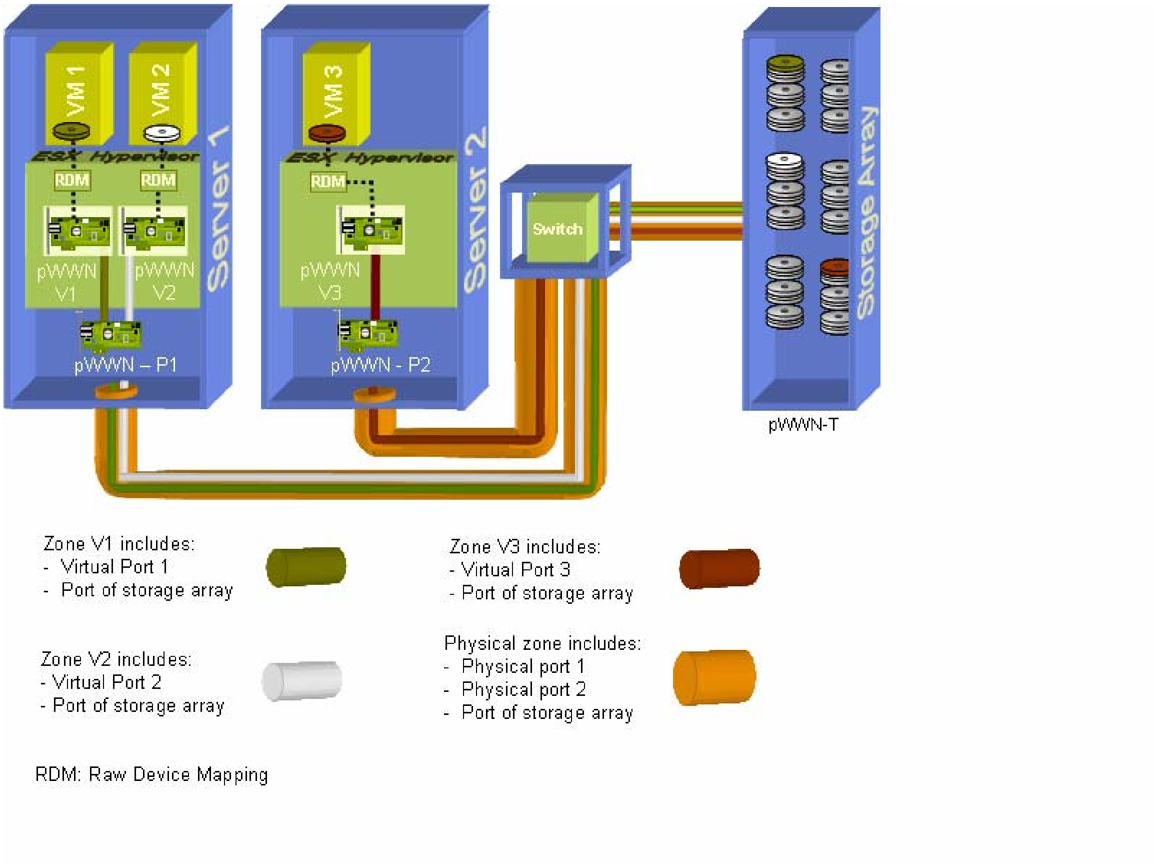

**Figure 4: Virtual machines using NPIV: Configuration based on Zoning, LUN Masking and Mapping.**

**LUN Mapping and Masking**
Each LUN "x" is exposed to the physical initiators (pWWN-P1, pWWN-P2) and to virtual machine VMx (pWWN-Vx) only.

**Zoning**
The zone P includes the physical devices, that are the physical HBAs (pWWN-P1, pWWN-P2) and the storage array port (pWWN-T). Each zone Vx includes only the VMx virtual HBA (pWWN-Vx) and the storage array port (pWWN-T).

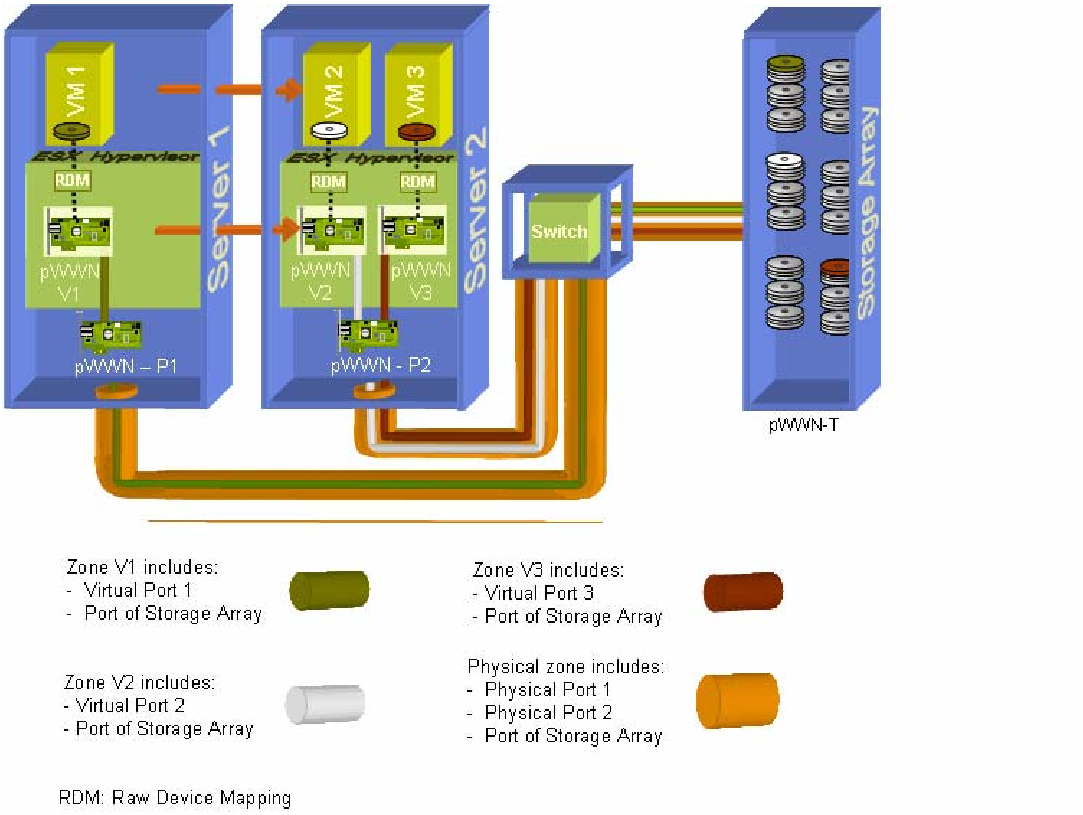

**Figure 5: Virtual machines using NPIV: Zoning, LUN masking and mapping after VM2 VMotion from server 1 to server 2.**

**VMotion: VM-2 from SRV-1 to SRV-2**
SAN Configuration is unchanged.
Zone V2 is unchanged, but now VM2 is located in SRV-2.
LUN mapping and masking is unchanged.

Using NPIV a virtual machine can be relocated from one physical ESX Server to another, without the storage and fabric administrator being requested to change any zoning or LUN mapping and masking setting. At the same time the storage access is protected from any other virtual server by both zoning and LUN masking and mapping, as it is common practice for any physical server.

## *7. Troubleshooting (with fcping, traceroute and End to End Connectivity Analysis tool)*

“Fcping” and “fctraceroute” are the Fibre Channel counterparts to the “ping” and “traceroute” utilities well known in TCP/IP environments.

Fcping is used to determine the basic connectivity between two Fibre Channel network points, as well as monitor and measure network latency. Traceroute reports on a SAN path, including node hops and latency data. Both fcping and fctraceroute can be used for connectivity from either the physical, or with NPIV, a virtual HBA associated with a VM.

Fabric Manager’s End to End Connectivity Analysis tool tests for connectivity of all elements within a zone across the fabric. This tool can report connectivity faults based upon virtual port addresses.

The use of these tools enable SAN administrators to verify that actual connectivity is implemented as planned, and more importantly to trace points of congestion or added latency, in order to optimize fabric efficiency and response time. The use of these tools with Virtual HBAs enables congestion control and optimized response time by individual VM and application, and to better support service-level agreements.

## Conclusion

VMware Infrastructure 3 provides enterprises with unprecedented capabilities for consolidation and infrastructure management. With the support of NPIV as a feature of VMware VI3, data centers can further leverage the benefits of the combined server virtualization with traditional SAN best practices. VMware VI3 and future releases of Cisco SAN OS on MDS switches will deliver these and additional capabilities when combined with NPIV-enabled LightPulse Fibre Channel HBAs from Emulex. The key SAN benefits of increased availability, data protection and manageability enable a new level of efficiency in resource management within the virtualized data center. This paper has addressed a few of the value propositions and solutions we are jointly working to deliver. The partnership will continue to define best practices and enhancements to ensure the maximum benefits for customer solutions based on NPIV and ESX Server.

## For more information

To help you achieve your virtual data center goals, please contact Emulex, Cisco, or VMware for information on Emulex Virtual HBA technology, Cisco MDS fabric switch products employing NPIV technology, or VMware ESX Server.

**Cisco Systems, Inc. 170 West Tasman Drive, San Jose, CA  95134-1706 USA  Tel 408-526-4000**
**www.cisco.com**

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

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