

Virtualizing SAN Connectivity with VMware Infrastructure 3 and Brocade Data Center Fabric Services

How the VMware Infrastructure platform can be deployed in a Fibre Channel-based shared storage environment and how support for NPIV enables critical fabric services to be extended to Virtual Machines (VMs) in the data center.



version 1.0

Table of Contents

Introduction..... 3

 NPIV in a Virtualized Environment 3

Shared Storage Services at the Core..... 4

VMware Infrastructure Capabilities Relying on Shared Storage..... 5

Fibre Channel Technology for the Virtualized Data Center 6

 Fibre Channel Enhancements in ESX Server 3.5..... 7

 Infrastructure Required to Deploy NPIV 9

Brocade Adaptive Networking in the Virtualized Data Center 9

Two Case Studies..... 10

 Case Study: Using QoS to Guarantee Application Performance during I/O-Intensive Backups 10

 Case Study: Optimizing VM Placement for Maximum I/O Throughput..... 13

Conclusion..... 14

Introduction

The first phase of virtualization deployment in the data center realized significant cost savings on hardware, power consumption, and rack space through physical server consolidation.

The next phase of virtualization adoption is being driven by the following business requirements:

- Guaranteed service availability during planned server maintenance and system failures
- Quick turnaround time for service requests from automating common management tasks, such as server provisioning and application deployment
- Fast response to rapidly changing business conditions from leveraging a flexible and adaptive computing environment where applications are not longer bound to specific hardware resources
- Rapid recovery with minimal business impact in the event of a major disaster
- Reduced operational cost through IT automation

The services provided by a high-performance Data Center Fabric (DCF) infrastructure are critical to the success of this next phase. The DCF provides fast and reliable access from the virtualization platform to centralized storage resources and enables the decoupling of servers and storage.

This paper looks at how the VMware® Infrastructure platform leverages Fibre Channel (FC)-based shared storage to provide advanced virtualization services in the data center and how using ESX 3.5 deployed with Brocade® DCF architecture enables enterprise customers to reap the full benefits of their investment in virtualization. It also demonstrates how support for N_Port ID Virtualization (NPIV) introduced with VMware ESX 3.5 enables critical fabric services to be extended to individual Virtual Machines (VMs).

NPIV in a Virtualized Environment

NPIV is an extension to the Fibre Channel standard that enables a single physical Host Bus Adapter (HBA) port to register multiple World Wide Port Names (WWPNs) with the fabric.

Each physical FC port in a fabric has a World Wide Port Name (WWPN) assigned to it by the equipment manufacturer, which uniquely identifies each port. WWPNs play a critical role in determining the visibility and accessibility of storage LUNs by servers connected to the fabric. Zoning is the mechanism by which FC ports are grouped together to restrict interference, add security, and simplify management. Zoning utilizes WWPNs to allow access to storage. A server can see and access only storage LUNs that share a common zone with that server. Besides Zoning, WWNs play a critical role in FC traffic monitoring and the ability to charge back storage requests to application owners.

Support for NPIV in ESX Server 3.5 is accomplished via Raw Disk Mapping (RDM) storage LUNS.

Shared Storage Services at the Core

VMware® Infrastructure 3 provides a number of options to handle Virtual Machine (VM) storage. At a very basic level, local storage, or disk space installed locally in the ESX Server host server itself, can be used to store the VM data. Although this option is inexpensive, it creates dependencies between ESX Servers and Virtual Machines (VMs) and prevents live migration of VMs across ESX Servers. Choosing local storage to store VMs precludes the use of critical capabilities, such as VMware® VMotion™, VMware® Distributed Resource Schedule (DRS), and VMware® High Availability (HA).

The use of shared storage with VMware Infrastructure 3 is vastly preferable, as it removes a VM dependency on any one particular ESX Server. With shared storage, VMs and their data stored entirely on the shared storage and ESX Servers become simply a set of resources on which a VM can run. No ESX Server “owns” a particular VM and any ESX Server can boot any VM as long as it can access the storage Logical Unit Number (LUN) on which the virtual machine is stored through the Storage Area Network (SAN).

A SAN presents a set of shared storage devices (depending on access privileges) in the form of a LUN to multiple ESX Servers instances simultaneously. Each server has access to the data store in which the VM is stored (LUN) as if it were directly attached to that server.

A Storage Area Network provides a wide range of benefits including:

- Effective utilization of storage resources across the data center with a shared storage pool available to all servers connected to the SAN
- Simplified, centralized management of storage, reducing administrative workload as storage no longer needs to be managed on a server-by-server basis
- Increased flexibility and scalability through any-to-any storage and server connectivity
- Improved throughput performance to shorten data backup and recovery time
- Reduced Local Area Network (LAN) congestion due to removal of backups from production IP networks
- Higher data availability for business continuance through a resilient network design
- Excellent scalability and investment protection allowing you to easily add more storage according to your business needs
- Superior security (VM data partitioning) for storage environments
- Non-disruptive business operations when you add or redeploy storage resources
- Proven Return On Investment (ROI) in virtualization infrastructure

Direct-attached or local storage is compared to SAN-based shared storage in Figure 1.

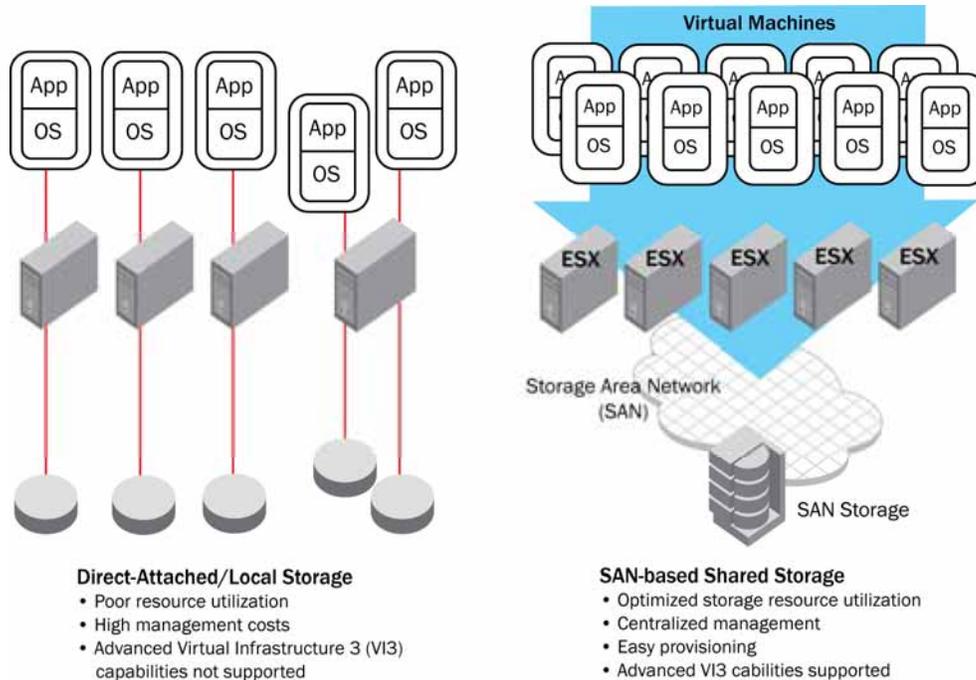


Figure 1. Configuration of direct-attached and local storage vs. SAN-based shared storage

VMware Infrastructure Capabilities Relying on Shared Storage

The following key VMware Infrastructure capabilities require a shared storage infrastructure to operate:

VMware® vMotion™ allows virtual machines to be moved from one ESX Server host to another on the fly, with no interruption in service.

VMware HA provides easy-to-use, cost-effective high availability for all applications running in VMs. If a VM is impacted by server failure, VMware HA can power on the Virtual Machine on another available ESX Server host.

VMware DRS dynamically allocates and balances computing capacity across collections of hardware resources aggregated into logical resource pools. It continuously monitors utilization across resource pools and intelligently allocates available resources among VMs based on performance thresholds set by an administrator.

VMware® Consolidated Backup enables LAN-free backup of VMs from a centralized proxy server location, reducing administrative overhead by centralizing backup management and eliminating the need for backup agents on each VM. Consolidated Backup also eliminates network traffic on the network by backing up VMs over the shared storage.

VMware Site Recovery Manager makes it possible to build, automate, and test data center disaster recovery plans by leveraging Infrastructure 3 core capabilities and integrating with market-leading, shared-storage-based data replication solutions.

The majority of VMware customers who have deployed virtual infrastructures today are using shared storage for VM storage. Many medium to large organizations already have SANs deployed in their environment, and these SANs can be leveraged for new VMware Infrastructure 3 deployments.

Fibre Channel Technology for the Virtualized Data Center

Fibre Channel (FC) is the predominant protocol on which shared storage implementations are designed. Over the past several years, FC shared storage solutions have become widely deployed in corporate infrastructure and are the preferred choice for enterprise data centers. Fibre Channel is a set of advanced data transport standards that allow large amounts of data to be moved reliably at multigigabit speeds between computers, servers, disk arrays, and other devices.

The Fibre Channel standard is accredited by many standards bodies, technical associations, vendors, and industry-wide consortiums. There are many products on the market that take advantage of FC's high-speed, high-availability characteristics. An FC-based fabric provides a wide range of benefits to virtualized data centers.

Proven reliability

Fibre Channel was initially adopted by a majority of the world's financial institutions, Fortune 500 companies, and other organizations running critical applications requiring guaranteed reliable, on-time data delivery. Today this same mission-critical data center reliability is available, at a much lower cost, to any business or organization. Fibre Channel has become the de-facto standard for SAN deployment in the enterprise.

High speed and low latency

Today 4, 8 and 10 Gbit/sec products are available. This rate of data transmission speed is projected to double in 2009, keeping up with the foreseeable needs of the most-demanding applications. The Fibre Channel road map calls for incremental bandwidth increases up to 128 Gbit/sec over the next decade.

Guaranteed delivery

Guaranteed in-order delivery of raw block data is a critical requirement for reliable and predictable transfer of storage Input/Output (I/O) requests across the SAN between servers and storage arrays and guaranteed data integrity.

Multipathing

Allows fault tolerance connections through multiple physical paths to storage devices. This not only protects against the single point of failure but it also enables dynamic load balancing across ports for optimum performance.

Optimized congestion management

Fibre Channel's credit-based flow control mechanism delivers data as fast as the destination buffer is able to receive it in order to meet high throughput data transfers. This facilitates applications such as backup, restore, remote replication, and other business continuance capabilities.

Scalability

It is common to see hundreds of enterprise-wide servers and storage arrays connected to an FC fabric with thousands of ports spanning multiple geographically-distant data centers. An FC SAN performs with equal reliability and predictable data rates, regardless of the number of connected nodes.

High efficiency

Real-world network performance is directly correlated to the efficiency of the technology. Fibre Channel has very little transmission overhead. Most important, the FC protocol is specifically designed for highly efficient operation using hardware-accelerated implementations.

The ability of Fibre Channel to maintain consistent high performance and low latency access to storage as fabric utilization increases makes it uniquely positioned to serve the needs of the virtualized data center.

The consolidation of hardware resources that drives server virtualization adoption typically results in increased fabric utilization. In a virtualized environment, a few physical servers are running a large number of VMs performing a large number of concurrent storage I/Os through a limited number of physical ports. FC technology guarantees deterministic delivery of storage I/O and enables VMware Infrastructure to consistently and reliably sustain optimum performance during peak server and storage utilization.

Fibre Channel Enhancements in ESX Server 3.5

VMware introduced support for NPIV virtualization in version 3.5 of ESX Server. The benefits of NPIV support are detailed in this section. VMware ESX Server 3.5 leverages NPIV and assigns individual WWPNs to each VM, so that each VM can be recognized as a specific end point in the fabric, as shown in Figure 1.

The benefits of this approach are numerous:

Better accounting

Since every I/O can be tied back to an individual VM, it is easier for the storage administrator to charge back the line of business owner of the Virtual Machine.

Quality of Service (QoS) to the VM

Since I/O requests originating from each VM can be identified on the fabric, they can be prioritized differently by the fabric switches allowing different QoS levels to be assigned to individual VMs, as shown in Figure 2.

Easier monitoring and troubleshooting

The same monitoring and troubleshooting tools used with physical servers can be used with VMs, since the WWN and the fabric address that these tools rely on to track frames are now uniquely associated to a VM.

Flexible provisioning and upgrade

Zoning and other services are no longer tied to the physical WWN (hard-wired to the HBA), so it is easier to replace an HBA. You do not have to reconfigure the SAN storage, because the new server can be pre-provisioned independently of the physical HBA WWN.

Workload mobility

The virtual WWPNs associated with each VM follow the VM when it is migrated across physical servers, No SAN reconfiguration is necessary when the workload is relocated to a new server.

Granular security

Access to specific storage LUNs can be restricted to specific VMs using the VM virtual port WWN for zoning, in the same way that they can be restricted to specific physical servers. (Note that ESX 3.5 still requires the physical HBA WWPN to be zoned with the storage WWPN.)

Applications identified in the SAN

Virtualized applications tend to be run on a dedicated VM, so the WWN of the VM now identifies the application to the SAN.

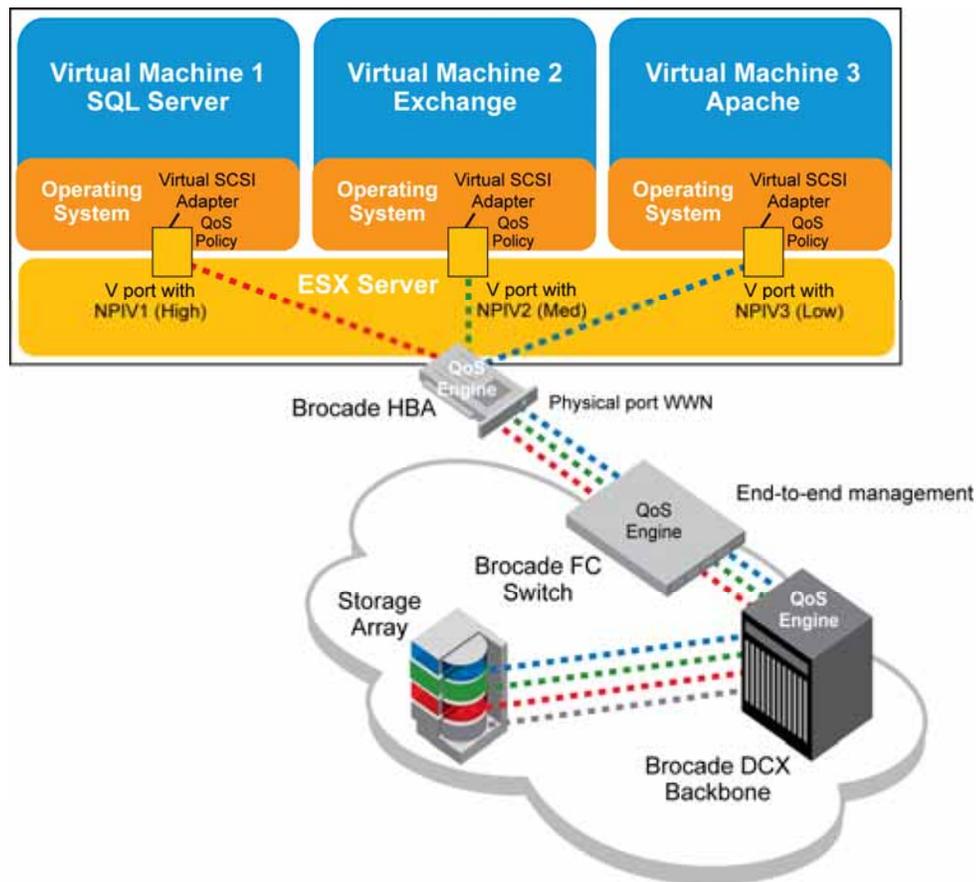


Figure 2. Identifying and prioritizing VM traffic with NPIV

Infrastructure Required to Deploy NPIV

The following hardware and software components are required to deploy NPIV in the fabric:

Fabric

NPIV introduces extensions to the core FC protocol that need to be supported on all switches connected to the fabric. All Brocade FC switches and enterprise-level platforms running Fabric OS® (FOS) 5.1.0 or later support NPIV.

HBAs

HBAs must support NPIV as well. Additionally, the HBA driver needs to expose an API for the VM monitor to create and manage virtual fabric ports. Most FC HBAs today have this capability.

Virtualization Software

VMware ESX Server 3.5 manages the relationships between virtual NPIV ports, also called "VPorts," and Virtual Machines. VMware ESX Server 3.5 assigns a unique set of four WWPNs to each VM (one VPort per virtual SCSI adapter). Support for NPIV in ESX Server 3.5 is limited to Raw Disk Mapping (RDM) storage LUNS.

Note: NPIV is completely transparent to storage arrays, so no specific support is required on the storage side.

Brocade Adaptive Networking in the Virtualized Data Center

Brocade Adaptive Networking services extend fabric intelligence to the application, enabling fabric-wide application QoS-level monitoring and management that automatically reacts to changes in virtual server workloads. This approach enables the fabric to dynamically allocate shared resources as changes occur in the data flows between virtual servers and virtual storage. If congestion occurs, the fabric can adjust bandwidth and other resources according to defined QoS levels, helping to ensure that higher-priority workloads receive the resources they need. Adaptive Networking introduces new services in the fabric, as described in the following section.

Quality of Service

Quality of Service is the granular allocation of fabric resources to applications based on the relative importance of the application as defined by the assigned level—High, Medium, or Low. When applications become dynamic, the QoS priority must follow the applications as they move between physical server and fabric connections. The key technology connecting virtual servers to virtual storage is Brocade virtual channels technology, which enables Adaptive Networking services to monitor resource usage, detect (or predict) congestion in the data path, and dynamically adjust resources to avoid congestion based on the QoS priority.

Traffic Management services

Provide congestion management to support application service levels. They can also provide automatic Ingress Rate Limiting and advanced queuing algorithms to remove congestion and dedicate bandwidth to specific applications.

Fabric Dynamic Profiling services

Provide end-to-end analysis of individual application data flows and resource usage, supplying in-depth information about the impact on overall fabric performance. These services identify points of congestion, and monitor and report on numerous statistics counters for physical resource utilization—useful information for provisioning, capacity planning, and end-to-end fault isolation tools that simplify fabric management.

Policy Management services

Prevent buffer credit exhaustion (buffer credits provide fabric flow control) and detect underutilized shared physical resources, reclaiming them or reallocating them to optimize application flow according to defined policies.

All Adaptive Networking services are available on the Brocade DCX® Backbone running FOS 6.0.0 and later and on Brocade 5300, 5100, and 300 Switches running FOS 6.1.0 or later. NPIV support in VMware ESX Server 3.5 enables the benefits of Brocade Adaptive Networking services to be extended to each individual VM rather than the physical server running the VM.

Two Case Studies

The following real-world use cases showcase NPIV and Brocade Adaptive Networking services.

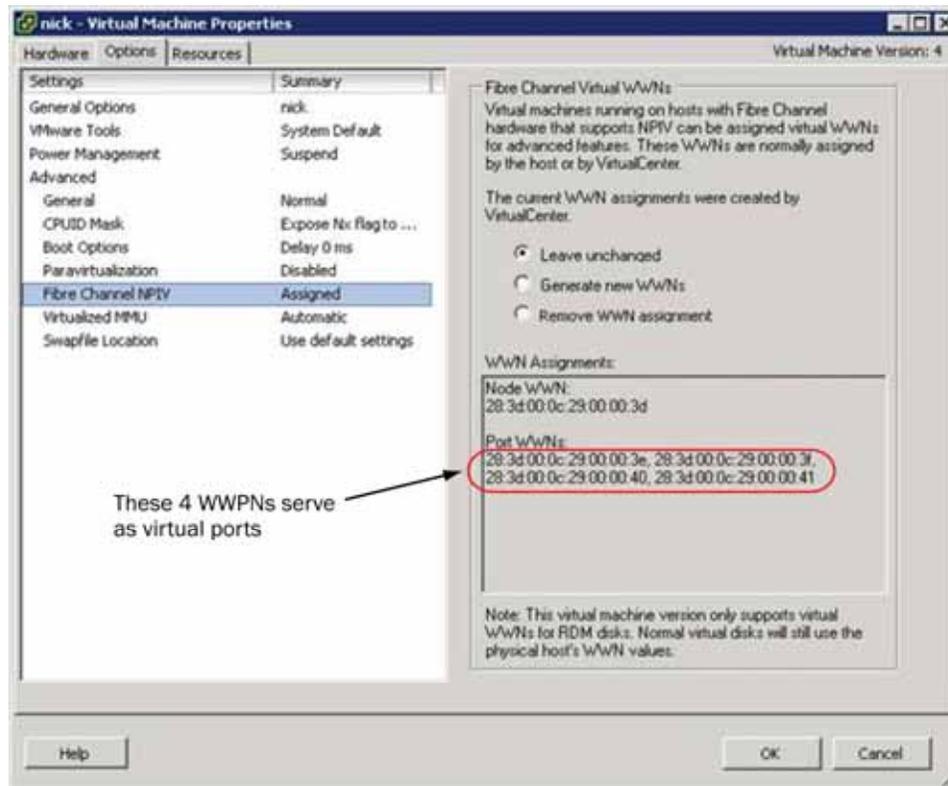
Case Study: Using QoS to Guarantee Application Performance during I/O-Intensive Backups

Jenny, a VMware administrator working for a global financial institution, has been tasked with provisioning a new Oracle database server to be used as the data tier for a mission-critical trading application. The company's IT policy mandates that all new applications be deployed in Virtual Machines running on VMware ESX Servers. The application is used by traders worldwide. The application business owners want guaranteed I/O performance 24 x7, regardless of the level of fabric utilization. The company performs daily I/O-intensive backups across the FC fabric, consuming a significant amount of bandwidth.

She started by creating a Linux VM and provisioned it with an Oracle database instance. She used a dedicated Raw Device Mapping (RDM) data store on a Symmetrix DMX storage array for the Oracle database data file. Both the DMX array and the ESX Server running the VM are attached to a Brocade DCX Backbone.

Step 1: Assign specific World Wide Names to the VM

Jenny uses the VMware Infrastructure client to generate and assign WWNs to the Virtual Machine after having assigned an RDM LUN to the VM for the database storage. She provides the VM WWNs to the SAN administrator along with his QoS request.



Jenny is now ready to ask the SAN administrator to set up QoS to guarantee I/O performance for the VM.

Step 2: Set QoS level using Brocade tools

Quality of Service, as described earlier, is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow. Priorities are generally identified as High, Medium, or Low. (By default, all flows in a Brocade fabric are Medium QoS.)

To service Jenny's request, the SAN administrator creates a zone that contains the VM VPorts WWNs and the WWN of the storage port used to access the RDM LUN assigned the VM. The zone is then assigned a higher priority. This means that if there is traffic congestion between the HBA used by the server on which the VM is running and the storage array that hosts the LUN, the traffic between the VM and the LUN is given higher priority.

A QoS zone is a special zone that indicates the priority of the traffic flow between a give host/target pair. It has a special name to differentiate it from a regular zone, for example, **QOSH_HighPriorityTraffic**. The switch automatically sets the priority for the host-target pairs specified in the zones based on the priority level in the zone name.

Create a zone that includes the VM VPorts WWPNs and the storage array port WWN:

```
DCX_A_3:root> zoneCreate QOSH_vml_lun1,"10:00:00:00:c9:6b:df:e1;
28:3d:00:0c:29:00:00:3e; 28:3d:00:0c:29:00:00:3f; 28:3d:00:0c:29:00:00:40;
28:3d:00:0c:29:00:00:41"
```

Add the newly created QoS zone to the existing zoning configuration and enable it:

```
DCX_A_3:root> cfgAdd cfg1,QOSH_vml_lun1
DCX_A_3:root> cfgEnable cfg1
You are about to enable a new zoning configuration.
This action will replace the old zoning configuration with the
current configuration selected.
Do you want to enable 'cfg1' configuration (yes, y, no, n): [no] y
zone config "cfg1" is in effect
Updating flash ...
```

Then verify that QoS is enabled on all E_Ports in the fabric:

```
DCX_A_3:root> portCfgShow
Ports of Slot 1  0  1  2  3  4  5  6  7  8  9 10 11  12 13 14 15
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Speed          AN AN AN AN  AN AN AN AN  AN AN AN AN  AN AN AN AN
AL_PA Offset 13 .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
Trunk Port     ON ON ON ON  ON ON ON ON  ON ON ON ON  ON ON ON ON
(output truncated)
NPIV capability ON ON ON ON  ON ON ON ON  ON ON ON ON  ON ON ON ON
QOS E_Port     ON ON ON ON  ON ON ON ON  ON ON ON ON  ON ON ON ON
(output truncated)
```

If not, enable it using **portCfgQos**:

```
DCX_A_3:root> portCfgQos --enable 1/0
```

Case Study: Optimizing VM Placement for Maximum I/O Throughput

A SAN administrator, Mak, working for a multi-national pharmaceutical enterprise, wants to optimize VM placement across a pool of servers and storage resources. He first needs to understand which VMs are the highest consumers of I/O bandwidth across the fabric so he can advise the VMware administrator on where to locate the VM to insure the best I/O performance and bandwidth utilization.

He is going to leverage the Brocade Top Talkers service to accomplish this task. In an FC fabric, "Top Talkers" is the name given to the top "n" bandwidth-consuming flows on a given switch. Top Talkers flows are identified by the WWPNs pair corresponding to the two end points of the connection. NPIV support in ESX 3.5 enables Virtual Machines with RDM storage LUNS to access the fabric through their own Vport WWPN, so VMs can now be seen as end points by the Top Talker service.

Step 1: Identify the most active VM in the fabric

Use the Brocade Command Line Interface (CLI) to identify the fabric Top Talkers. On the VM F_Port device, monitor for ingress and/or egress traffic:

```
DCX_A_3:root> perfTtMon --add ingress 1/0
```

Then show traffic performance (in the MB/sec column):

```
DCX_A_3:root> perfTtMon --show 1/0
=====
Src_WWN                               Dst_WWN                               MB/sec
=====
28:3d:00:0c:29:00:00:01                20:14:00:a0:b8:29:e4:ba                6.926
28:3d:00:0c:29:00:00:05                20:14:00:a0:b8:29:e4:ba                6.972
28:3d:00:0c:29:00:00:0a                20:14:00:a0:b8:29:e4:ba                6.830
28:3d:00:0c:29:00:00:0f                20:14:00:a0:b8:29:e4:ba                64.573
28:3d:00:0c:29:00:00:14                20:14:00:a0:b8:29:e4:ba                6.772
```

Notice that one port device displays 10x the I/O of the other 4 and is therefore the Top Talker. Thanks to NPIV each VM has its Vport, so it is easy to identify I/Os coming from individual VMs. Use VMware® VirtualCenter to map Top Talker WWNs to virtual WWNs associated with individual VMs.

Step 2: Migrate these VMs to ESX Servers with the highest I/O capacity

Each of the source WWNs in the CLI output above corresponds to a Virtual Machine and each destination WWN correspond to a storage port. The VM with the WWN "28:3d:00:0c:29:00:00:0f" is generating a lot more I/O than all the others. If this I/O is using congested paths in the SAN, then it is a candidate to be moved to another platform that uses less-congested paths.

Conclusion

The combination of VMware infrastructure 3.5 and Brocade DCF architecture, leveraging NPIV to access advanced fabric services, is a major step forward on the path to end-to-end virtualization in the data center. These new capabilities greatly benefit organizations who want to virtualize their business-critical applications without risking degradation of I/O performance. VMware ESX 3.5 NPIV support puts Virtual Machines on an equal footing with physical servers for accessing storage resources and enables critical fabric services to be delivered to individual Virtual Machines, namely:

Quality of Service can be assigned to individual Virtual Machines to provide VMs running I/O-intensive, mission-critical application prioritized access to storage and guaranteed I/O service levels.

Monitoring and troubleshooting I/O traffic on the fabric can be done at the VM level using existing Fibre Channel monitoring tools to identify and troubleshoot VM I/O requests.

Accounting and chargeback for I/O and storage access can now be performed by the individual VM, enabling IT to perform granular accounting and chargeback in a virtualized environment.



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

© Brocade Communications Systems, Inc. All Rights Reserved. 08/08 GA-WP-094-00

Brocade, Fabric OS, File Lifecycle Manager, MyView, and StorageX are registered trademarks and the Brocade B-wing symbol, DCX, and SAN Health are trademarks of Brocade Communications Systems, Inc., in the United States and/or in other countries. All other brands, products, or service names are or may be trademarks or service marks of, and are used to identify, products or services of their respective owners.

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.

