

EMC POWERPATH/VE FOR VMWARE vSPHERE

Best Practices Planning

Abstract

EMC® PowerPath®/VE is a path management solution for VMware and Microsoft Hyper-V servers. This paper focuses on PowerPath/VE for VMware vSphere™. Similar to PowerPath Multipathing for other operating systems, PowerPath/VE provides the highest level of dynamic load balancing, path failover, path restoration, path testing, and automated performance optimization. This white paper will cover the best practices for PowerPath/VE connectivity and management.

August 2011

Copyright © 2009, 2011 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is”. EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware, ESX, ESXi, vMotion, VMware vCenter, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part Number h6340.1

Table of Contents

Executive summary.....	4
Introduction	4
Audience.....	4
PowerPath/VE integration with VMware vSphere	4
Managing and monitoring with PowerPath/VE	6
Licensing.....	8
Management access and security.....	9
PowerPath/VE in vCenter Server	11
Update Manager.....	11
LUN ownership.....	12
Virtual Storage Integrator	13
PowerPath/VE and NMP co-existence.....	16
VMware Native Multipathing.....	18
PowerPath/VE features	19
Availability and connectivity	20
Zoning.....	20
Connectivity to VMAX	21
Connectivity to Symmetrix	23
Connectivity to Vblock Infrastructure Packages.....	23
Connectivity to VNX	24
Host registration.....	25
Connectivity to CLARiiON	26
Unisphere Server Utility.....	26
Connectivity to EMC Celerra.....	26
NPIV and vSphere.....	26
VMware advanced features and PowerPath/VE.....	26
Conclusion.....	27

Executive summary

Data center managers are looking to virtualization as a means to scale without increasing overhead, simplify management, and reduce application downtime. Physical server consolidation results in reclaiming valuable data center space, realizing higher utilization rates, increasing operational efficiencies, and improving availability of resources and applications. Used in small and large data centers alike, VMware vSphere™ provides a platform to virtualize data centers. As customers continue to expand the use of virtualization in their data centers, the availability of hardware, network, and SAN storage resources connected to those servers is critical.

EMC® PowerPath®/VE for VMware® vSphere is a multipathing extensions module for vSphere that provides software that works in combination with SAN storage to intelligently manage Fibre Channel, iSCSI, and Fibre Channel over Ethernet I/O paths. PowerPath/VE is installed on the vSphere host and will scale to the maximum number of virtual machines on the host, improving I/O performance. The virtual machines do not have PowerPath/VE installed nor are they aware that PowerPath/VE is managing I/O to storage. PowerPath/VE dynamically load balances I/O requests and automatically detects, and recovers from, path failures. This new addition to the PowerPath family standardizes path management across heterogeneous physical and virtual environments—servers, operating systems, and virtual platforms. PowerPath/VE is the best solution for ensuring application uptime by providing availability and optimization of the server, storage, and SAN.

Introduction

This white paper identifies best practices for understanding and using PowerPath/VE with VMware vSphere (ESX® 4.x and ESX 5.x). It should be noted that no two customers have identical data center configurations. This paper highlights the best practices and specific requirements of PowerPath/VE with vSphere.

Audience

This paper is intended for storage administrators, storage architects, and EMC employees. Individuals who are familiar with PowerPath Multipathing will learn the similarities and differences of configuring, managing, and monitoring with PowerPath/VE. The best practices outlined in this paper provide options for fine-tuning and enhancing a production deployment. EMC pre-sales and deployment personnel will be better able to position, define, and design PowerPath/VE for a customer's SAN.

PowerPath/VE integration with VMware vSphere

VMware release ESX 4.x and ESX 5.x (vSphere) introduced an enhancement to network storage. New in vSphere, the vStorage APIs for Multipathing provide a

framework that enables storage vendors to develop Multipathing Extensions Modules (MEM) that plug in to vSphere. These plug-in modules provide multipathing software that can be used in place of vSphere's native multipathing functionality to deliver advanced multipathing capabilities for storage arrays. It is one of several vStorage APIs that has been published to enable storage partners like EMC to integrate their vendor-specific capabilities to provide more optimal performance and to give joint customers more choice in how they configure storage resources for virtualization.

PowerPath/VE works with VMware ESX and ESXi™ as a MEM that provides enhanced path management capabilities to ESX and ESXi hosts. PowerPath/VE is supported with vSphere (ESX 4.x and ESX 5.x) only. Previous versions of ESX do not have the vStorage APIs for Multipathing that are required by PowerPath/VE.

PowerPath/VE installs as a kernel module on the vSphere host. PowerPath/VE will plug in to the vSphere I/O stack framework to bring the advanced multipathing capabilities of PowerPath – dynamic load balancing and automatic failover – to the VMware vSphere platform. This is shown in Figure 1.

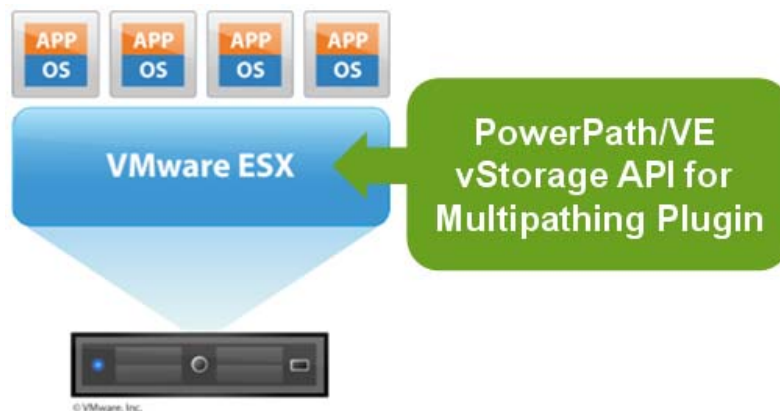


Figure 1. PowerPath/VE vStorage API for multipathing plug-in

At the heart of PowerPath/VE path management is server-resident software inserted between the SCSI device-driver layer and the rest of the operating system. This driver software creates a single pseudo device for a given array volume (LUN) regardless of how many physical paths on which it appears. The pseudo device, or logical volume, represents all physical paths to a given device. It is then used for creating virtual disks, and for raw device mapping (RDM), which is then used for application and database access.

PowerPath/VE's value fundamentally comes from its architecture and position in the I/O stack. PowerPath/VE sits above the HBA, allowing heterogeneous support of operating systems and storage arrays. By integrating with the I/O drivers, all I/Os run through PowerPath and allow for it to be a single I/O control and management point. Since PowerPath/VE resides in the ESX kernel, it sits below the virtual machines and the virtual machines, applications, databases, and file systems running them. PowerPath/VE's unique position in the I/O stack makes it an infrastructure

manageability and control point – bringing more value going up the stack. This is shown in Figure 2.

Note: On other PowerPath platforms, the pseudo devices (for example, *emcpower0*) may be used by applications rather than a device's physical address (for example, */dev/dsk/c4t5006048452A4EB97d9s0*). In the case of PowerPath/VE for vSphere, applications running on the guest operating systems do not refer to the pseudo device. Instead the device UID (in *naa.xxxx* format) is used for guest operating systems' applications.

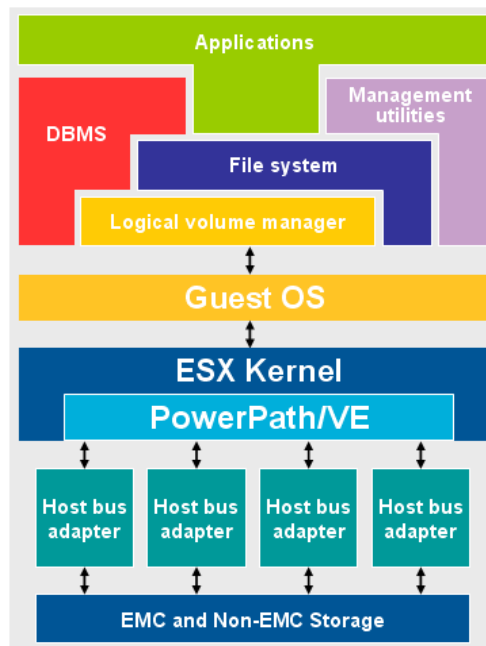


Figure 2. PowerPath/VE I/O stack

Managing and monitoring with PowerPath/VE

PowerPath/VE for vSphere provides expanded management and monitoring capabilities from what an administrator may be accustomed to with other PowerPath platforms. Table 1 highlights the differences.

Table 1. PowerPath/VE installation and management comparison

	Installation, Removal, and Upgrade		Managing, Monitoring, and Configuring		Licensing
	CLI	GUI	CLI	GUI	
PowerPath/VE for ESX	Remote via vCLI	VMware Update Manager	Remote via RTOOLS	VMware vCenter™	Electronic licensing
PowerPath for Windows	Local installer	N/A	Local <i>powermt</i>	Microsoft Management Console Plug-in	License key
PowerPath for Linux and UNIX	Local installer	N/A	Local <i>powermt</i>	N/A	License key

PowerPath /VE for ESX requires the administrator to install a new PowerPath management CLI, called “RTOOLS,” that is designed specifically for the ESX environment. It facilitates remote licensing, monitoring, managing, and configuration of PowerPath/VE hosts. RTOOLS uses the *rpowermt* commands. The syntax, arguments, options, and outputs are almost identical to the traditional *powermt* commands used on all the other PowerPath-supported operating system platforms. There is one significant difference. RTOOLS is a remote CLI that consolidates the management and configuration of all PowerPath/VE for ESX hosts at one location.

Not all vSphere platforms have a service console interface. (vSphere 5 has no version with a service console.) In order to manage an ESXi host, customers have the option to use vCenter Server or vCLI (vCLI is also referred to as VMware Remote Tools) on a remote server. PowerPath/VE for ESX uses the RTOOLS command line utility for both ESX and ESXi. PowerPath/VE for vSphere cannot be managed on the ESX host itself.

Developing a remote CLI to accommodate ESXi hosts has provided an additional benefit to server and storage administrators. RTOOLS monitors and manages multiple PowerPath/VE for ESX hosts across the entire SAN from a single management server. Administrators do not have to log in to individual hosts to access the PowerPath CLI as they would normally do with other PowerPath-supported platforms.

Administrators should designate one virtual machine or a physical machine to manage one or multiple ESX hosts. RTOOLS is supported on Windows and Red Hat. These operating systems may be installed as a virtual machine or physical server. As seen in Figure 3, there are four management options available:

- Windows 2003, 2008, and 2008 R2 on a virtual machine
- RHEL 5.2, 5.3, 6.0 and 6.1 on a virtual machine
- Windows 2003, 2008, and 2008 R2 on a physical host
- RHEL 5.2, 5.3, 6.0 and 6.1 on a physical host

The RTOOLS package contains an rpm and a zip file for RHEL 5/6 and Windows, respectively. Neither the physical server nor the virtual machine has to be dedicated to RTOOLS. They can be shared with other applications.

Although it is technically supported to have multiple RTOOLS management consoles, it is not recommended. Choosing one option reduces the number of potential access points to PowerPath/VE hosts in the SAN.

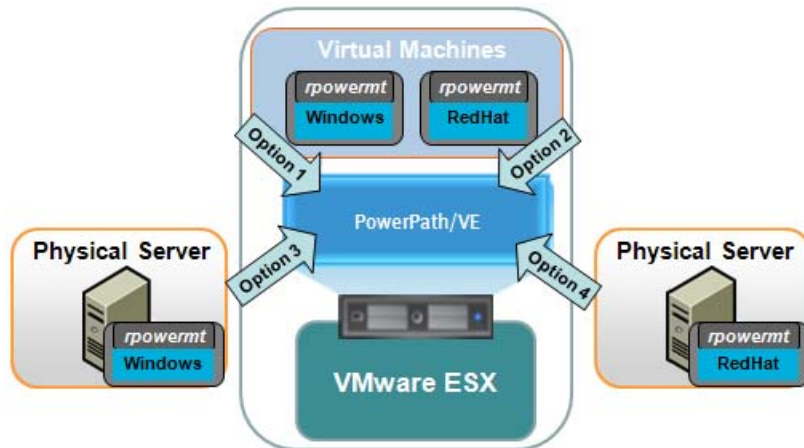


Figure 3. PowerPath/VE RTOOLS management options

Licensing

PowerPath/VE for vSphere uses electronic license management. Consult the *EMC PowerPath/VE for vSphere Installation and Administration Guide* for more detailed information on use cases, installation, and configuration.

PowerPath/VE has two modes of licensing, Served and Unserved. The Installation and Administration Guide has more information on choosing the best options when ordering PowerPath/VE. There is no cost difference between them.

From a PowerPath/VE feature and functionality point of view (on the vSphere host) and RTOOLS, there is no difference. If the license source is the License Server (Served) or an Unserved license file (Unserved), the ESX host will be fully licensed for multipathing.

From a deployment point of view, the Served option requires that the PowerPath/VE License Server (from Flexera) be installed. This means more time for deployment and configuration. The required Flexera code is included in the PowerPath/VE for ESX download bundle on Powerlink.

Note: Use only the version in the PowerPath/VE software bundle.

From a flexibility point of view, the Served option is the best. In a typical SAN, servers are commonly added and removed from the infrastructure. PowerPath/VE can be easily registered (licensed) on new servers and unregistered (unlicensed) on decommissioned servers as the infrastructure changes. This is important when the environments changes frequently. As an example, a technology refresh could result

in adding and removing servers. Another example would be in test labs. Non-productions ESX hosts with PowerPath/VE are regularly imaged.

An Unserved license option is less flexible in this environment. It would require a new license file every time a new server is added or an existing server is reimaged. This requires the customer to go to Powerlink to have a new license file generated.

From a simplicity point of view, the Unserved option is a good one in a small environment (e.g. <10 hosts) that does not change.

Note: PowerPath licenses are not included with EMC Symmetrix VMAX™ /VMAXe or VNX™/VNXe systems. PowerPath is an extra-charge item for these systems and must be ordered separately.

Management access and security

PowerPath/VE for vSphere uses a CIM-based (Common Information Model) solution that allows communication between the RTOOLS server and the ESX host. A CIM client must first authenticate before it can access data or perform operations on a vSphere host.

There are 3 options for authentication when remotely connecting from the RTOOLS server to the vSphere server.

- 1) Root login and password are entered at the command line once and stored in a lockbox (encrypted file) on the RTOOLS server
- 2) Root login and password are entered at the command line every time an *rpowermt* command is executed. No login or password data is stored in the lockbox
- 3) A CIM ticket is issued every time an *rpowermt* command is executed. No login or password data is stored in the lockbox

Option1

When connecting to an ESX host for the first time, *rpowermt* prompts the administrator for the username and password of that host. Each new ESX host accessed by an *rpowermt* command generates a prompt for the username and password. The administrator is prompted for a lockbox password once. *rpowermt* securely stores the following information in a lockbox (encrypted file):

ESX host being accessed

Username and password of ESX host that is accessed

Storing the hosts and passwords in a lockbox enables them to be maintained across system reboots. To guard against unauthorized access, if a lockbox is copied from one RTOOLS server to another, the user is prompted to enter the lockbox password again before *rpowermt* commands can be issued.

Security protocols differ among customers. There may be cases where administrators are not permitted to save passwords. In such cases, the lockbox can be bypassed by setting an environment variable or with an inline *rpowermt* command option. Please contact EMC Customer Support for more details on how to do this.

Host-level security is important. On other PowerPath platforms, *powermt* commands require root access. PowerPath/VE *rpowermt* commands do not require root-level access to run the executable; however, the ESX root password or account with root privileges is required as an entry in the lockbox. Access to the *rpowermt* command executable is based on the native access controls of the server (Windows or RHEL) where *rpowermt* is installed.

As discussed earlier, *rpowermt* provides the administrator a new capability with PowerPath, the ability to centralize vSphere management and monitoring. However, this also means that if an *rpowermt* server is not locked down properly, any user could issue PowerPath/VE commands to all vSphere hosts. The lockbox is designed to secure vSphere usernames and passwords. It does not protect against access to the *rpowermt* server commands. Figure 4 shows the lockbox with the RTOOLS management option.

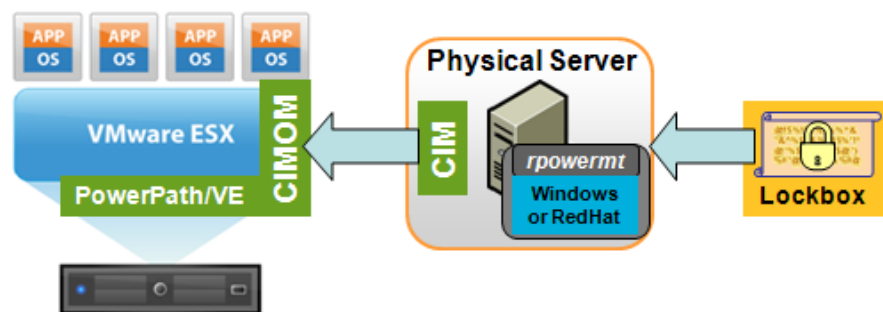


Figure 4. PowerPath/VE secured communication with the lockbox

Option 2

When both login name and password are entered at the command line, the lockbox is automatically bypassed. This requires the administrator to specify the login and password for every *rpowermt* command executed.

The lockbox may also be disabled by setting an environment variable on the RTOOLS server. Refer to PowerPath/VE Installation and Administration Guide for settings.

Option 3

CIM ticket authentication is a function of VMware vCenter. As stated earlier, a CIM client must first authenticate before it can access data or perform operations on a vSphere host. This feature is configured on vCenter server and activated with *rpowermt* using a command line argument. This feature also works when the lockdown mode is enabled. The vSphere host login and password are not required. The RTOOLS lockbox is not utilized. For more information, refer to PowerPath/VE Installation and Administration Guide and VMware user documentation.

PowerPath/VE in vCenter Server

PowerPath/VE for vSphere is managed, monitored, and configured using *rpowermt* commands as discussed in the [previous section](#). This CLI-based management is common across all PowerPath platforms (except for Windows, which has both CLI and GUI) with slightly different command usage. Virtual Storage Integrator (VSI) displays PowerPath path status information. The additional monitoring functionality has allowed administrators to free themselves of issuing CLI commands in order to view PowerPath multipathing status.

Update Manager

VMware Update Manager compares the operating systems and applications running in the Virtual Infrastructure deployment against a set of standard updates and patches. Specified updates can be applied to the hypervisor, to guest operating systems, and to the applications on scanned ESX Server hosts and virtual machines.

PowerPath/VE for vSphere leverages Update Manager to provide a GUI-based tool for managing the installation or upgrade of PowerPath/VE on all vSphere servers managed by a vCenter instance. This deeper management integration allows administrators to bypass installs and upgrades using vCLI. PowerPath/VE is installed on the vSphere hypervisor. It is not installed on the virtual machines. Update Manager also automates the movement of virtual machines through Distributed Resource Scheduler, which allows a nondisruptive installation or upgrade in a cluster environment. New PowerPath/VE code can be measured against a user-defined baseline, staged, and finally remediated.

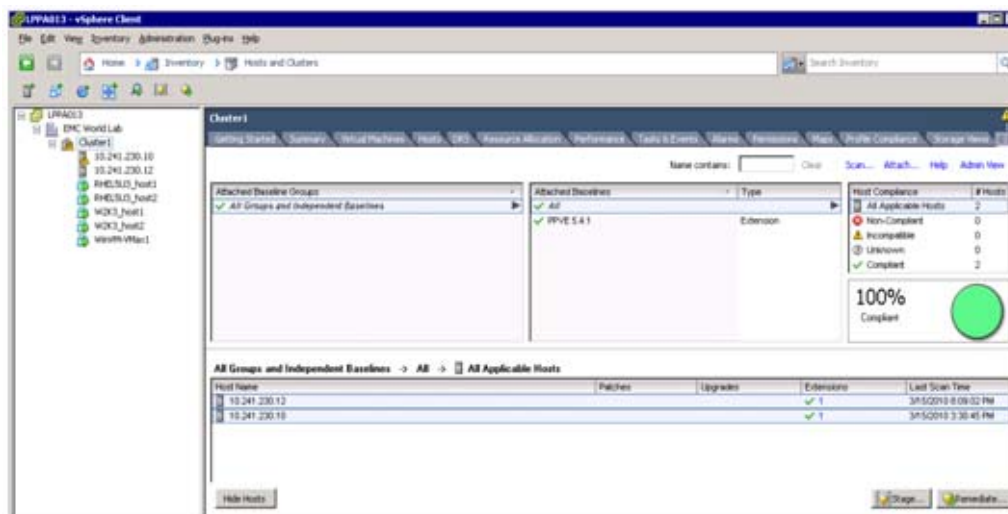


Figure 5. A fully remediated vSphere cluster running PowerPath/VE

vCLI Installation, Upgrade and Removal

PowerPath/VE may be installed or upgraded using VMware's vCLI. With ESX5, the online bundle package installation is no longer available. VIB files must be uploaded to the ESX hypervisors (via SCP for example). Administrators have two options to start the installation. They may use *esxcli* commands from a remote host or they may SSH to a specific ESX host and execute *esxcli* commands.

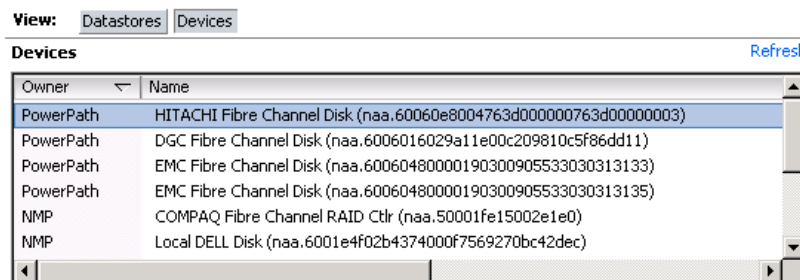
To uninstall PowerPath/VE, administrators must use local or remote *esxcli* commands. The host must be manually rebooted by the administrator.

Refer to the PowerPath/VE Installation and Administration Guide for details on installing, upgrading, and removing.

LUN ownership

vCenter allows administrators to view LUN ownership. In regards to multipathing, ownership refers to the multipathing product that manages I/O on the data paths to the LUNs. To view ownership:

1. Open vCenter Client.
1. Highlight the ESX host.
2. Click the **Configuration** tab.
3. Click **Storage** under the Hardware window.
4. Click **Devices** for the View window (Figure 6).



Owner	Name
PowerPath	HITACHI Fibre Channel Disk (naa.60060e8004763d000000763d00000003)
PowerPath	DGC Fibre Channel Disk (naa.6006016029a11e00c209810c5f86dd11)
PowerPath	EMC Fibre Channel Disk (naa.60060480000190300905533030313133)
PowerPath	EMC Fibre Channel Disk (naa.60060480000190300905533030313135)
NMP	COMPAQ Fibre Channel RAID Ctr (naa.50001fe15002e1e0)
NMP	Local DELL Disk (naa.6001e4f02b437400f7569270bc42dec)

Figure 6. Device ownership in vCenter Server

Figure 6 shows that there are VMAX, VNX, and Hitachi devices owned by PowerPath/VE. The VMware Native Multipathing (NMP) module owns any device that PowerPath cannot manage or any device that the user chooses not to be managed by PowerPath; for example the COMPAQ array is an older active/passive model not supported by PowerPath. (See [PowerPath/VE and NMP co-existence](#) regarding the management of LUN types not supported by PowerPath/VE.) A set of *claim rules* is added via the vStorage APIs for Multipathing to determine LUN ownership. For a complete list of supported EMC and third-party arrays, consult the *EMC Support Matrix* or E-Lab Navigator on [EMC Powerlink](#).

Note: LUN ownership can also be viewed in the ESX CLI with the *esxcfg-mpath* command.

Virtual Storage Integrator

EMC Virtual Storage Integrator (VSI) for VMware vSphere is a plug-in to VMware vCenter that provides a single management interface used for managing EMC storage within the vSphere environment. Features can be added and removed from VSI independently, providing flexibility for customizing VSI user environments. VSI Feature Manager manages the features. VSI provides a unified user experience, allowing each of the features to be updated independently, and new features to be introduced rapidly in response to changing customer requirements.

When PowerPath/VE is installed on an ESX/ESXi host, VSI presents important multipathing details for a given device, such as the load-balancing policy, the number of active paths, the number of dead paths, and others. For VSI to obtain and present PowerPath/VE details for the underlying storage, the PowerPath/VE RTOOLS for Windows package must be installed. It is part of the PowerPath/VE for vSphere installation kit. Both VSI and RTOOLS are free downloads. In Figure 7, VSI recognizes PowerPath/VE is installed and displays the version.

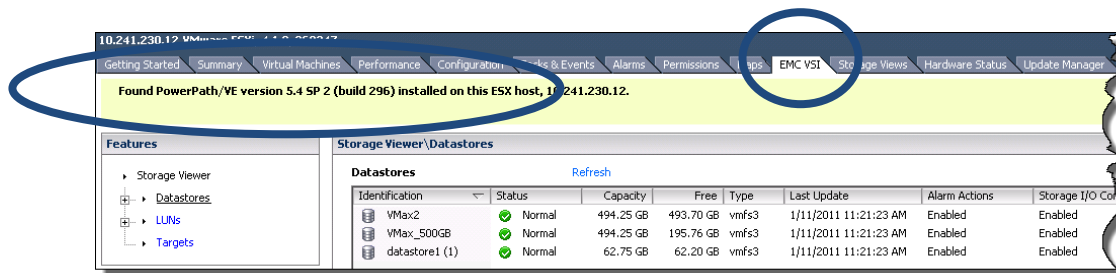


Figure 7. PowerPath version displayed through the VSI tab

For storage and server administrators, PowerPath/VE has more visibility in VMware vCenter than any other multipathing product including VMware NMP. With each release of PowerPath/VE, management functionality has deepened its integration. This has allowed administrators to decrease their reliance on CLI tools and use a single management application for monitoring. Figure 8 is a screenshot of the PowerPath/VE path details in vSphere. This information can also be found in the output of the `rpowermt display dev=x` command. The `rpowermt` command displays additional LUN and path information. However, VSI displays most of the information that administrators typically need.

Note: If VMware's NMP managed those LUNs, the details window in Figure 8 would not exist. Only PowerPath-managed devices have this level of path information.

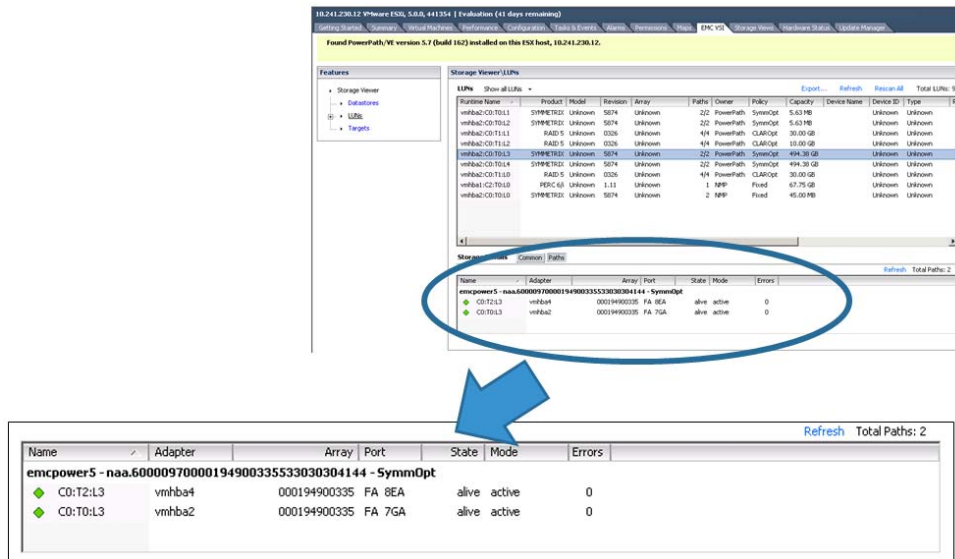


Figure 8. Path details to a VMX device managed by PowerPath/VE

PowerPath VSI details

This section highlights the meaning of some of the fields in VSI and their importance to the administrator. Administrators using VSI will gain valuable insight into path status. With heightened path awareness, potential I/O issues can be avoided. The example in Figure 8 will be used as a reference.

emcpower5 – This is the pseudo device that PowerPath/VE creates for every managed device. One or more native paths to the same storage LUN are presented as a single pseudo device. A pseudo device provides load balancing and failover. Below the pseudo device name is the physical path information to the device represented by the CTD (Controller, Target, Disk) numbers. The *esxcfg-mpath* CLI commands in ESX display the CTD information for each device. In the case of VMware, the pseudo devices are not directly accessible by the application as is common on other PowerPath platforms. The hypervisor abstracts this information for usage by the virtual machines.

The WWID is displayed. Each physical device on the storage array has a unique identifier. For the VMX device in this case, it is **naa.60000970000194900335533030304144**. PowerPath/VE discovered this device and assigned a pseudo device name.

The load-balancing policy is also displayed. By default, PowerPath/VE will always automatically assign the optimized policy for a particular array's LUN. In this case, **SymmOpt** is selected. Multiple load-balancing policies are available; however, customers are advised to use the default.

Total Paths – Through zoning and masking, physical devices are presented to the hosts. Usually, the number of ports will be dependent on the throughput requirements of the applications to that device. In this example, the number of paths is 2. Per VMX connectivity best practices, both HBAs (vmhba2 and

vmhba4) have access to different FA ports. A greater number of paths translates to higher availability and increased throughput. It is important for the administrator to know the number of paths to the device in case of an improper configuration or fault.

Each path has an icon, which indicates the state and mode of the path. The possible icons and their respective state and mode are:



The solid green icon indicates that the path is active.



The black outline icon indicates that the path is active but unlicensed.



The red outline icon indicates that the path is dead.

The path information can be refreshed with the **Refresh** link at the top of the sub-panel. When refreshing path information, VSI connects to the PowerPath/VE service on the ESX/ESXi host to gather new path information.

State – A path state is *alive* or *dead*. The states are reflections of path availability and not user configuration. If the path is alive, PowerPath/VE considers this path for I/O. If the path is dead, no I/O will be sent down this path until it is restored after the successful completion of a path test.

Mode – A path mode can be *active*, *standby*, or *unlic*. The modes are reflections of the administrator's configuration choices. If the path is active, it can accept I/O. It is available for load balancing and failover. If the path is standby, it is being held in reserve. An administrator has manually configured the path to standby mode or a failover policy has been assigned that places the path in this mode. If the path is unlic, the administrator has not installed the PowerPath Multipathing license. Only one HBA port will be used for I/O. No HBA failover is available. This is not a recommended configuration because it can lead to data unavailability in a multipath fault scenario.

Note: Mode cannot be configured in VSI. The *rpowersmt* commands must be used to apply licenses, set load-balancing policies, and configure standby paths.

Errors – This is the total number of times this path transitioned from *alive* to *dead*. This is always equal to or less than the total number of HBA I/O path errors. Note that the error count may be different for different LUNs using the same physical paths.

Setting multipathing policies

The Path Management feature for EMC Virtual Storage Integrator for VMware vSphere includes support for both PowerPath/VE and the VMware Native Multipathing Plugin. The Path Management feature allows a VSI user to change the multipath policy on a multitude of devices, based on both storage class and virtualization object. For

example, all Symmetrix® devices for a given ESX host can be managed, or all CLARiiON® devices (“CLARiiON” in the UI includes VNX devices) for a given vSphere cluster.

PowerPath/VE automatically assigns an optimized load-balancing policy for all PowerPath-managed devices. EMC recommends that customers use the default optimized load-balancing and failover policies. Therefore, it is unlikely that administrators will use this management feature. These policies can also be changed using *rpowermt* commands. Figure 9 shows VSI’s Path Management change window.

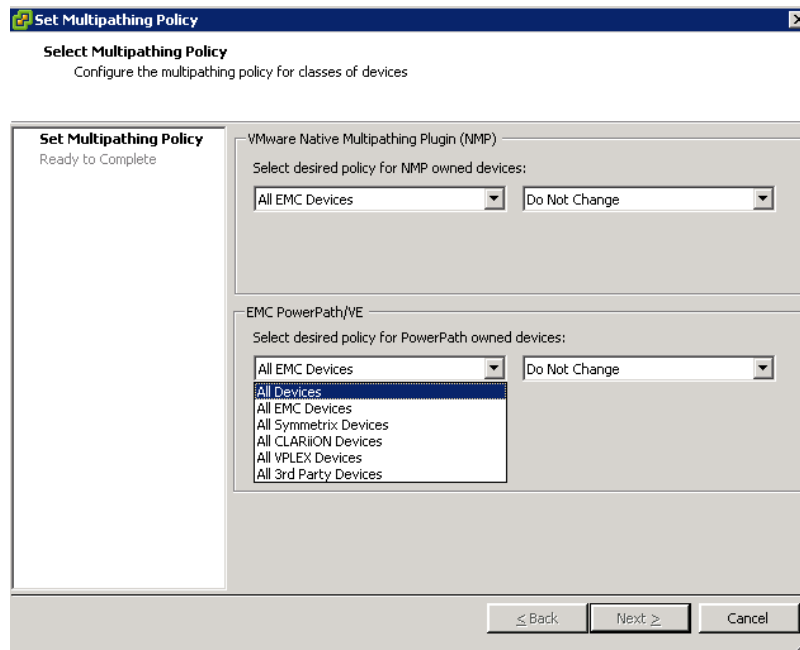


Figure 9. Path management with VSI

PowerPath/VE and NMP co-existence

Both PowerPath/VE and VMware NMP are supported on the same ESX host. PowerPath/VE supports path management for EMC and many non-EMC arrays. NMP supports path management for the remainder. PowerPath/VE and NMP can share the same HBAs, but they cannot manage the same device simultaneously.

Claim rules are used to assign storage devices to either PowerPath/VE or NMP devices. When an ESX host boots or performs a rescan, the ESX host discovers all physical paths to the storage devices visible to the host. Using the claim rules defined in the `/etc/vmware/esx.conf` file (for ESX), the ESX host determines which multipathing module will be responsible for managing a specific storage device. Claim rules are numbered. For each physical path, the ESX host processes the claim rules starting with the lowest number first. The attributes of the physical path are compared with the path specification in the claim rule. If there is a match, the ESX host assigns the MEM plug-in specified in the claim rule to manage the physical path.

This assignment process continues until all physical paths are claimed by a MEM. Figure 10 has a sample claim rules list post-PowerPath/VE installation.

```
# esxcli storage core claimrule list
Rule   Rule   Class  Type   Plugin      Matches
Class
-----
MP     0     runtime transport NMP        transport=usb
MP     1     runtime transport NMP        transport=sata
MP     2     runtime transport NMP        transport=ide
MP     3     runtime transport NMP        transport=block
MP     4     runtime transport NMP        transport=unknown
MP     101   runtime vendor   MASK_PATH vendor=DELL model=Universal Xport
MP     101   file    vendor   MASK_PATH vendor=DELL model=Universal Xport
MP     250   runtime vendor   PowerPath endor=DGC model=*
MP     250   file    vendor   PowerPath vendor=DGC model=*
MP     260   runtime vendor   PowerPath vendor=EMC model=SYMMETRIX
MP     260   file    vendor   PowerPath vendor=EMC model=SYMMETRIX
MP     270   runtime vendor   PowerPath vendor=EMC model=Invista
MP     270   file    vendor   PowerPath vendor=EMC model=Invista
MP     280   runtime vendor   PowerPath vendor=HITACHI model=*
MP     280   file    vendor   PowerPath vendor=HITACHI model=*
MP     290   runtime vendor   PowerPath vendor=HP model=*
MP     290   file    vendor   PowerPath vendor=HP model=*
MP     300   runtime vendor   PowerPath vendor=COMPAQ model=HSV111 (C)COMPAQ
MP     300   file    vendor   PowerPath vendor=COMPAQ model=HSV111 (C)COMPAQ
MP     310   runtime vendor   PowerPath vendor=EMC model=Celerra
MP     310   file    vendor   PowerPath vendor=EMC model=Celerra
MP     320   runtime vendor   PowerPath vendor=IBM model=2107900
MP     320   file    vendor   PowerPath vendor=IBM model=2107900
MP     65535 runtime vendor   NMP        vendor=* model=*
```

Figure 10. Claim rules post-PowerPath/VE installation (vSphere5)

Claim rules do not go into effect until after the vSphere host is rebooted following a PowerPath/VE installation. If changes to the claim rules are needed after installation, devices can be manually unclaimed and the rules can be reloaded without a reboot as long as the devices are not in use. It is a best practice to make claim rule changes after installation but before the immediate post-installation reboot. An administrator can choose to modify the claim rules in order to have NMP manage EMC or non-EMC devices. For instructions on changing claim rules, consult the *EMC PowerPath/VE for VMware vSphere Installation and Administration Guide*.

Note: In order to achieve the highest availability and best performance, PowerPath/VE should be used in place of NMP whenever possible

Note: Changes were made to the claiming of devices in vSphere5. As a result, some third-party that are not supported by PowerPath may be claimed by the PowerPath claim rule. Refer to the *PowerPath/VE for vSphere Installation and Administration*

VMware Native Multipathing

Prior to the release of vSphere 4, ESX administrators had only the NMP for I/O multipathing. This module provided basic failover and basic management path configuration. Even with multiple HBAs, ESX would assign the first path as active and any remaining paths would be passive failover paths. In such configurations, all I/Os from the ESX host queue on the first available HBA in the system while other HBAs are not actively used until a failure on the primary HBA is detected. This behavior leads to an unbalanced configuration on the ESX host and on the storage array. A Round Robin load-balancing setting was introduced in ESX 3.5; however, it did not provide the numerous and advanced load-balancing features of PowerPath/VE.

Using the older Fixed Path and Most Recently Used (MRU) policies, administrators must manually configure preferred paths for datastores in order to reduce the performance impact on the ESX host and storage array. To maximize throughput and avoid wasteful idle equipment (for example, HBAs, SAN ports, storage ports), each datastore has its preferred path set to a different physical path. This is a completely manual process that is time-consuming and never-ending. Each time a virtual machine is added or removed, or an application is changed, or SAN storage is added or removed, the administrator must make changes to the preferred path configurations. HBA throughput must be closely monitored in order to ensure that some ports are not oversaturated while others are underutilized.

When a path failover occurs due to a fault (for example, cable pull or failed fiber optic), I/O increases on the remaining paths as expected. When the fault condition is resolved, the paths will not automatically restore and rebalance if not configured properly (for example, the path is set to MRU).

In vSphere, Round Robin has become the new NMP load-balancing standard. Round Robin is a very simplistic load-balancing algorithm. This policy has advantages over the single path with failover (Fixed and MRU) because all paths are now capable of supporting I/O to the device simultaneously. In older MPIO versions, secondary paths were only capable of transmitting and receiving I/O when the primary path failed. For administrators, the ability to use all paths for load balancing and failover has been a huge benefit. However, not all paths are equal due to issues in the host, network, and storage array.

Round Robin uses all paths in a static configuration. The inherent problem with this approach is that SANs are dynamic by design. Servers, applications, networking components, and storage arrays are added, removed, and modified intentionally and sometimes accidentally. A load-balancing policy must be able to act and react in this type of environment. Round Robin:

- Does *not* dynamically reroute I/O traffic
- Does *not* identify any changes in network activity
- Does *not* consider any changes in the storage array
- Does *not* consider unique array properties

Does *not* track HBA queue depth

Does *not* consider I/O size

PowerPath/VE features

An alternative to NMP Round Robin, PowerPath/VE removes all of the administrative overhead associated with load balancing and failover on native MPIO hosts.

PowerPath/VE provides the following features:

Dynamic load balancing — PowerPath/VE is designed to use all paths at all times. It distributes I/O requests to a logical device across all available paths, rather than requiring a single path to bear the entire I/O burden.

Auto-restore of paths — Periodic auto-restore reassigns logical devices when restoring paths from a failed state. Once restored, the paths automatically rebalance the I/O across all active channels.

Automated performance optimization — PowerPath/VE not only uses all available paths, but it optimizes them as well. PowerPath/VE automatically identifies the type of storage array and sets the highest performing optimization mode by default. For Symmetrix, the mode is SymmOpt (Symmetrix Optimized). For VNX and CLARiiON, the mode is CLAROpt (CLARiiON Optimized). For supported active/active third-party arrays, the mode is Adaptive.

Dynamic path failover and path recovery — If a path fails, PowerPath/VE redistributes I/O traffic from that path to functioning paths. PowerPath/VE stops sending I/O to the failed path and checks for an active alternate path. If an active path is available, it redirects I/O along that path. PowerPath/VE can compensate for multiple faults in the I/O channel (for example, HBAs, fiber-optic cables, Fibre Channel switch, storage array port).

EMC Symmetrix VMAX, VNX, VNXe™, CLARiiON, VPLEX™, and non-EMC array support — PowerPath/VE supports a wide variety of EMC and non-EMC arrays. Consult the *EMC Support Matrix* or E-Lab Navigator for a complete list.

Monitoring and alerting — The integrated management of vCenter or the rpowermt CLI provides detailed path monitoring and alerting. While PowerPath/VE load balances I/O, it maintains statistics for all I/O for all paths.

Automatic path testing — PowerPath/VE periodically tests both live *and* dead paths. By testing live paths that may be idle, a failed path may be identified before an application attempts to pass I/O down it. By marking the path as failed before the application becomes aware of it, timeout and retry delays are reduced. By testing paths identified as failed, PowerPath/VE will automatically restore them to service when they pass the test. The I/O load will be automatically balanced across all active available paths.

Availability and connectivity

The availability of data is arguably the most important aspect of good SAN design. Assuring availability can be achieved via a variety of designs and strategies. Through proper redundancy and failover practices, secure availability can almost be guaranteed.

Availability is a measurement of the amount of time that data can be accessed, compared to the amount of time the data is not accessible because of issues in the environment. Lack of availability might be a result of failures in the environment that may cause a total loss of paths to the device, or it might be an event that caused so much bandwidth congestion that the access performance renders the device virtually unavailable.

Availability is impacted not only by the choice of components used to build the fabric, but also by the ability to build redundancy into the environment. The correct amount of redundancy will allow processes to gracefully fail over to secondary paths and continue to operate effectively. Too little redundancy built into the fabric can cause bandwidth congestion, performance degradation, or (in some cases) a loss of availability.

Corporations spend considerable time and resources to build a redundant infrastructure. In the case of SAN data center connectivity, two, four, eight, or more HBAs are installed on a single server. Dual Fibre Channel fabrics are established to isolate network traffic from hardware faults or accidental user-intervention. Highly available storage arrays (like VMAX and VNX) are connected to the SAN with multiple connections to increase performance as well as ensure availability.

PowerPath/VE is integral to the overall availability and connectivity of vSphere in the SAN. Providing proper connectivity is vital, but having the correct multipathing software on the host to take full advantage of the physical infrastructure is no less important.

Zoning

For vSphere environments or any open systems hosts, each initiator will ideally be in a zone with a single target. However, due to the significant management overhead that this can impose, single initiator zones can contain multiple target ports. (A zone should never contain more than 16 target ports.)

Figure 11 shows a sample zoning configuration from a Fibre Channel switch.

```
zoneset name VMware1 vsan 1
zone name vSphere0110_CX4SPB0_Symm7E0 vsan 1
* fcid 0x6b000e [pwwn 21:00:00:1b:32:10:fe:3d] > vSphere HBA
* fcid 0x6b00ef [pwwn 50:06:01:68:41:e0:a2:c5] > VNX SPB0
* fcid 0x6b03ef [pwwn 50:06:01:61:41:e0:a2:c5] > VNX SPA1
* fcid 0x6b0011 [pwwn 50:06:04:8a:d5:f0:ba:7e] > Symm FA
```

Figure 11. Sample zoning configuration with single initiator and multiple targets

Connectivity to VMAX

Changes in the Symmetrix VMAX architecture and the Enginuity™ operating environment have greatly increased its performance, reliability, and availability. However, as a result of architectural changes there are new recommendations for connecting VMware ESX hosts to the array for both VMware Infrastructure 3 and vSphere 4 environments. The new recommendations ensure the highest level of connectivity and availability to the VMware virtualization platform even during abnormal operations.

Each ESX server in the virtualized environment should have at least two physical HBAs. Each HBA should be connected to at least two different front-end ports on different directors. This configuration ensures continued use of all HBAs on vSphere servers even if a front-end port is brought down for maintenance activities on the director. Furthermore, the additional connectivity avoids unnecessarily idle equipment and reduces the performance impact caused by an offline director.

If the Symmetrix VMAX in use has only one VMAX Engine then each HBA should be connected to the odd and even directors within it. Connectivity to the Symmetrix front-end ports consists of connecting unique hosts to port 0 of the front-end directors before connecting additional hosts to port 1 of the same director and processor. Figure 12 and Figure 13 represent the logical connection between host and storage. A physical representation would display Fibre Channel switches between them.

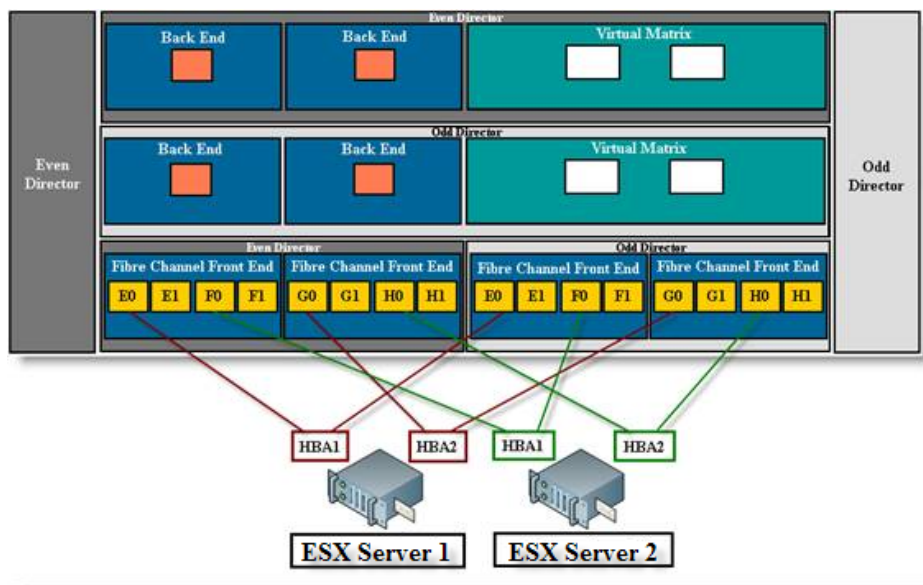


Figure 12. Connecting ESX servers to a single-engine Symmetrix VMAX

If multiple VMAX Engines are available, the HBAs from the vSphere server in the virtualized environment should be connected to different directors on different VMAX Engines. In this situation, shown in Figure 13, the first vSphere server would be connected to four different processors on four different directors over two VMAX Engines, instead of four different processors on two directors in one VMAX Engine.

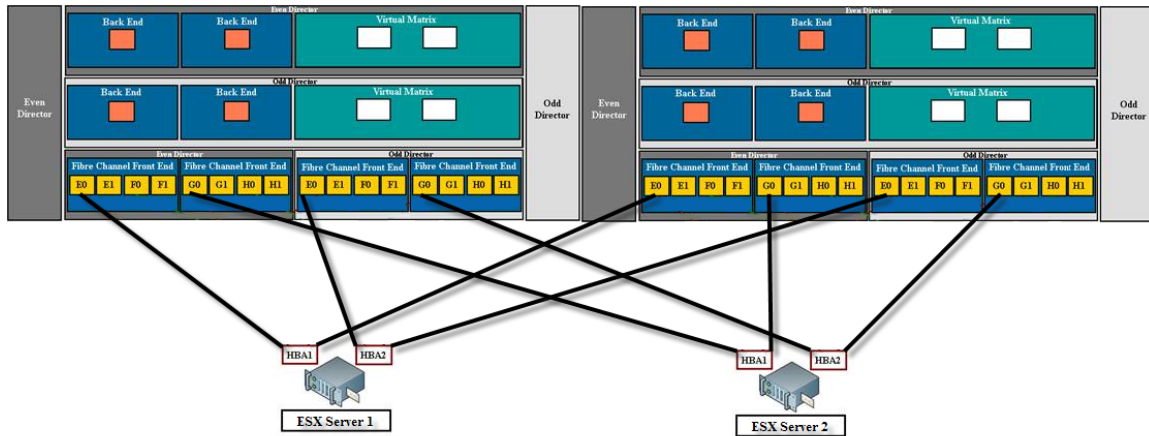


Figure 13. Connecting ESX servers to a Symmetrix VMAX with more than one engine

When the vSphere server is connected to the Symmetrix VMAX array using the best practice recommendation discussed here, the PowerPath/VE kernel module running on the vSphere host will associate four paths to each device presented from the array and associate a pseudo device name (as discussed earlier). An example of this is shown in Figure 14, which shows the output of `rpowermt display host=x.x.x.x dev=emcpower0`. Note in the output that the device has four paths and displays the optimization mode (**SymmOpt**= Symmetrix optimization).

Note: The SPC-2 bit must be enabled on the Symmetrix FA port.

```
C:\>rpowermt display host=x.x.x.x dev=emcpower0
```

```
Pseudo name=emcpower0
Symmetrix ID=000192601265
Logical device ID=03A7
Standard UID=naa.60000970000192601265533030334137
state=alive; policy=SymmOpt; priority=0; queued-IOS=0
=====
----- Host ----- - Stor - -- I/O Path - -- Stats --
### HW Path          I/O Paths  Interf.  Mode  State  Q-IOS  Errors
=====
    2 vmhba1          C0:T4:L10 FA 10eB  active  alive   0     0
    1 vmhba2          C0:T5:L10 FA 10eA  active  alive   0     0
    2 vmhba1          C0:T2:L10 FA 7eB   active  alive   0     0
    1 vmhba2          C0:T3:L10 FA 7eA   active  alive   0     0
=====
```

Figure 14. Output of the `rpowermt display` command on VMAX device

For more information on the `rpowermt` commands and output, consult the *EMC PowerPath/VE for vSphere Installation and Administration Guide*.

As more VMAX Engines become available, the connectivity can be scaled. PowerPath/VE supports up to 32 paths to a device, which allows multiple physical and logical connections across multiple engines. These methodologies for connectivity ensure all front-end directors and processors are utilized, providing maximum potential performance and load balancing for vSphere hosts connected to the Symmetrix VMAX storage arrays in combination with PowerPath/VE.

Connectivity to Symmetrix

Although there are significant improvements to the VMAX architecture and Engenuity code supporting it, the basic connectivity recommendations remain the same for other Symmetrix arrays. Also, the connectivity recommendations for iSCSI, Fibre Channel, and Fibre Channel over Ethernet are the same.

SAN design considerations are also identical. Multiple HBAs connected with dual fabrics attached to two or more Symmetrix FA ports will yield higher performance and availability. As with all storage arrays, PowerPath/VE supports up to 32 paths to a device.

Note: The SPC-2 bit must be enabled on the Symmetrix FA port.

Connectivity to Vblock Infrastructure Packages

Cisco, EMC, and VMware jointly formed the VCE coalition, which represents an unprecedented level of collaboration in development, services, and partner enablement that reduces risk for companies looking to transform their infrastructures for the private cloud. Enabled by [Vblock™ Infrastructure Packages](#), the coalition delivers the industry's first completely integrated IT offering that combines best-of-breed virtualization, networking, computing, storage, security, and management technologies with end-to-end vendor accountability. Rather than assembling individual components, customers can now purchase them as Vblock Infrastructure Packages.

PowerPath/VE is an integral part of all three Vblock Infrastructure Packages. PowerPath/VE supports the I/O path management as it does on other physical and virtualized hosts. Zoning, masking, and connectivity rules are based on Vblock best practices. As VMware manages pools of servers and Fully Automated Storage Tiering (FAST) manages pools of storage, PowerPath/VE manages pools of data paths to assure an optimized virtual infrastructure.

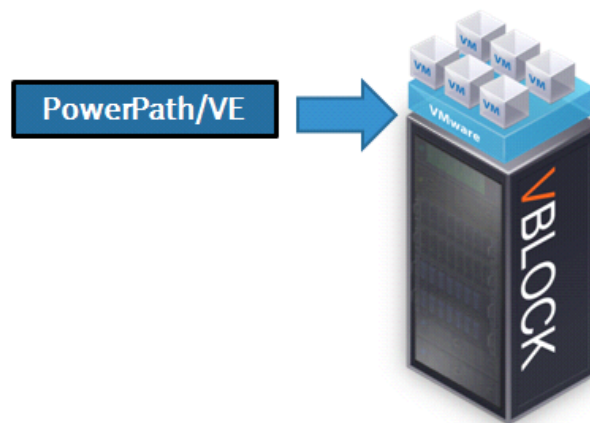


Figure 15. PowerPath's integration point in the Vblock

Connectivity to VNX

VNX and VNXe are EMC's next generation of unified storage. These hardware and software solutions are simple to provision, efficient with lower capacity requirements, affordable for any budget, and powerful enough to handle the demands of virtual applications. The VNX family delivers the world's simplest array to manage and the world's best midtier performance. It is designed to provide five 9s of availability. PowerPath/VE is an integral part of the overall system availability.

Along with other redundant components, the VNX and VNXe series support redundant storage processors in an active/active configuration, in which both storage processors support active workloads. In the event of a storage processor failure, the surviving storage processor will pick up the workload of the failed storage processor until the failed storage processor is replaced. Once replaced, the new storage processor will dynamically assume the original workload of the failed storage processor. PowerPath/VE can quickly respond to changes in the status of the storage processors. It can proactively fail over multiple I/O paths to LUNs to the alternate storage processor when common failures are detected.

PowerPath/VE supports multipathing load balancing and path failover to iSCSI, Fibre Channel, and Fibre Channel over Ethernet protocols to block devices. PowerPath/VE does not support iSCSI access to NFS or CIFS file systems.

Dual, active/active storage processors also support "rolling upgrades," or the ability to upgrade the system software or hardware while the VNX and VNXe systems are running. In a rolling upgrade, each storage processor is upgraded and rebooted separately, while the active storage processor keeps the system online. PowerPath/VE recognizes when a VNX is in the process of an upgrade. It proactively trespasses devices to the online storage processor while the other is completing its upgrade.

[Asymmetric Logical Unit Access \(ALUA\)](#) will be the default failover mode versus optional mode today with CLARiiON. ALUA is Asymmetric Active/Active with both paths presented as active and both paths can receive I/O. The path to the storage processor that owns the LUN is called optimized, while the other is called non-optimized as I/Os are redirected to the owning storage processor through CMI (inter-storage processor link).

Figure 16 shows the output of `rpowermt display host=x.x.x.x dev=emcpower17`. PowerPath/VE automatically recognizes ALUA mode and sets the optimized load-balancing policy, CLAROpt. (This is the optimized mode for VNX and CLARiiON devices.) The PowerPath/VE kernel module running on the vSphere host will associate four paths to each device presented from the array and associate a pseudo device name.

```
C:\>rpowermt display host=x.x.x.x dev=emcpower17
```

```
Pseudo name=emcpower17
CLARiiON ID= CF2G8101900187 [linx110_Spitfire1]
Standard UID= 60060160163029008B8F7037B0FBDF11 [LUN 65]
state=alive; policy=CLAROpt; priority=0; queued-IOS=0
Owner: default=SP A, current=SP A      Array failover mode: 4
```

```

=====
----- Host ----- - Stor - -- I/O Path - -- Stats --
### HW Path          I/O Paths   Interf.  Mode    State  Q-IOs Errors
=====
  2 vmhba3           C0:T2:L9  SP B1   active  alive   0      0
  1 vmhba1           C0:T1:L9  SP B0   active  alive   0      0
  1 vmhba1           C0:T2:L9  SP A1   active  alive   0      0
  2 vmhba3           C0:T1:L9  SP A0   active  alive   0      0
=====

```

Figure 16. Output of the *rpowermt display* command on a VNX device

Note: VNX and CLARiiON have the same output in the *rpowermt* display. A VNX can be distinguished by its “ID.”

LUN 65 in Figure 16 is owned by SPA. All paths to this storage processor are considered optimized paths in terms of ALUA. PowerPath/VE will load balance I/O to a LUN over all available optimized paths. PowerPath/VE will normally not use the non-optimized path due to its higher latency under normal operating conditions.

In Figure 17, HBAs 1 and 2 have access to both VNX storage processors. Since the LUN is owned by SPA, HBAs 1 and 2 will load balance I/O across SPA0 and SPA1. SPB0 and SPB1 will not be used for I/O for LUN 65 unless there is a failure of SPA. In a failure condition, PowerPath/VE will seamlessly fail over to the alternate storage processor and fail back when the path is recovered.

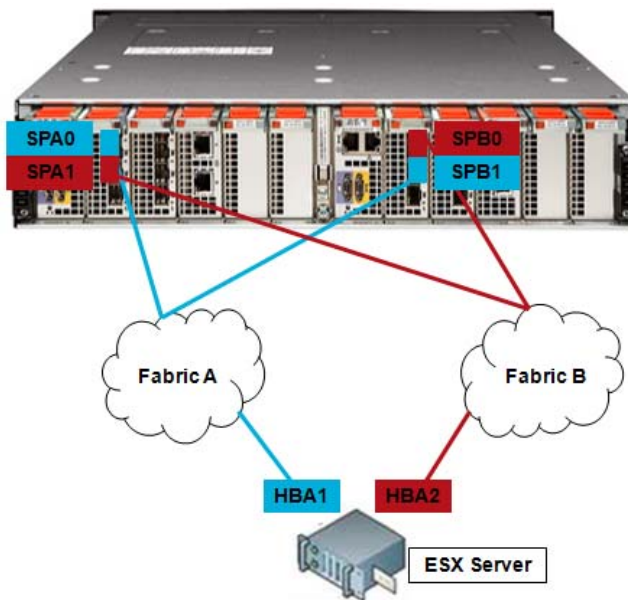


Figure 17. HBA to storage processor port connectivity on a VNX

Host registration

Host registration is a “push” in a VMware environment. PowerPath/VE performs auto-registration. With this functionality, no agent and no manual registration are required.

Connectivity to CLARiiON

As with the VNX, the basic connectivity recommendations remain the same for CLARiiON arrays. Each vSphere host should have at least two physical HBAs. Each HBA should be zoned to at least two ports of each storage processor. This configuration ensures continued use of all HBAs on the host even in the event of HBA, Fibre Channel switch, or SP failure or reboot.

SAN design considerations are also identical. Multiple HBAs connected with dual fabrics attached to two or more CLARiiON ports will yield higher performance and availability. As with all storage arrays, PowerPath/VE supports up to 32 paths to a device.

Unisphere Server Utility

Although not related directly to PowerPath, the Unisphere™ Server Utility is a tool for determining availability. To verify high availability, this utility is included on the Unisphere Server Support CD that ships with the storage system.

The server utility performs the following functions:

- Verify server high availability — Determines if the server is configured for high availability by verifying that the server has at least one connection path to each storage-system SP, and that PowerPath/VE is running.

- Update server information to the storage system

- Use the snapshot feature

Connectivity to EMC Celerra

PowerPath/VE for vSphere using iSCSI (native block) is supported.

NPIV and vSphere

NPIV on VMware works on the guest OS on the virtual machines only. PowerPath/VE installs at the ESX kernel layer and therefore is not impacted by NPIV.

VMware advanced features and PowerPath/VE

PowerPath has a unique location in the I/O stack that allows it to integrate seamlessly with the advanced features and capabilities of VMware. PowerPath/VE is fully compatible with VMFS, RDM, Virtual SMP, High Availability, Distributed Resource Scheduler, VMotion®, Storage VMotion, Consolidated Backup, vCenter Site Recovery Manager, and more.

Conclusion

PowerPath/VE is the best multipathing solution for ensuring application uptime by providing availability and optimization of the server, storage, and SAN. Using intelligent load balancing, proactive path failover, seamless path recovery, and integrated management, PowerPath/VE for vSphere provides server and storage administrators the ability to dynamically scale the infrastructure by optimizing existing data paths, reduce management overhead by simplifying the configuration and using an integrated monitoring tool in vCenter, and increase uptime with proactive path testing and constant I/O monitoring.