



## V I R T U A L A U

L U N E D I 1 1 G E N N A R C H I V I O

## Why VMware PSA is helping me "to save the day" (part 1)

With the coming of vSphere 4.0 I have the opportunity to rethink my approach to storage multipathing in vmware world. To justify my post title, I need to share some storage background.

Following concepts and descriptions are grabbed from the well written [vSphere Fibre Channel SAN configuration guide](#) that is a "must read"

## Backgrounds

### Storage System Types

Storage disk systems can be active-active and active-passive. ESX/ESXi supports the following types of storage systems:

- An active-active storage system, which allows access to the LUNs simultaneously through all the storage ports that are available without significant performance degradation. All the paths are active at all times, unless a path fails.
- An active-passive storage system, in which one port is actively providing access to a given LUN. The other ports act as backup for the LUN and can be actively providing access to other LUN I/O. I/O can be successfully sent only to an active port for a given LUN. If access through the primary storage port fails, one of the secondary ports or storage processors becomes active, either automatically or through administrator intervention.

### ALUA, Asymmetric logical unit access

ALUA is a relatively new multipathing technology for asymmetrical arrays. If the array is ALUA compliant and the host multipathing layer is ALUA aware then virtually no additional configuration is required for proper path management by the host. An Asymmetrical array is one which provides different levels of access per port. For example on a typical Asymmetrical array with 2 controllers it may be that a particular LUN's paths to controller-0 port-0 are active and optimized while that LUN's paths to controller-1 port-0 are active non-optimized. The multipathing layer should then use paths to controller-0 port-0 as the primary paths and paths to controller-1 port-0 as the secondary (failover) paths. Pillar AXIOM 500 and 600 is an example of an ALUA array. A Netapp FAS3020 with Data ONTAP 7.2.x is another example of an ALUA compliant array.

### Understanding Multipathing and Failover

To maintain a constant connection between an ESX/ESXi host and its storage, ESX/ESXi supports multipathing. Multipathing lets you use more than one physical path that transfers data between the host and external storage device.

In case of a failure of any element in the SAN network, such as an adapter, switch, or cable, ESX/ESXi can switch to another physical path, which does not use the failed component. This process of path switching to avoid failed components is known as path failover.

In addition to path failover, multipathing provides load balancing. Load

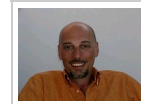
▼ 2010 (6)

► marzo (1)

▼ gennaio (5)

[The case of the ghost LUN 0](#)[Why VMware PSA is helping me "to save the day" \(pa...](#)[Why VMware PSA is helping me "to save the day" \(pa...](#)[Upgrading VAIO VGN-SZ3XP SATA HD with Intel X25-M ...](#)[My first post!](#)

I N F O R M A Z

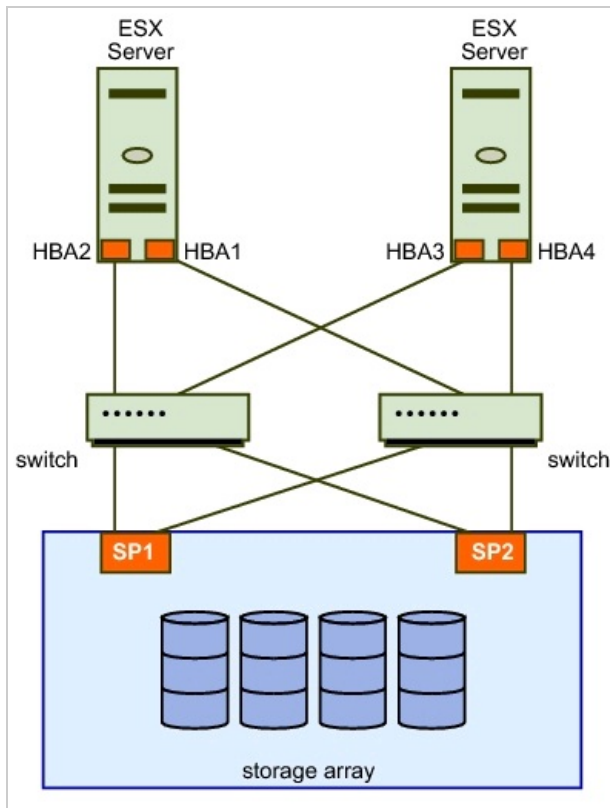
F A B I O  
V I S U A L I  
P R O F I L O

balancing is the process of distributing I/O loads across multiple physical paths. Load balancing reduces or removes potential bottlenecks.

### Host-Based Failover with Fibre Channel

To support multipathing, your host typically has two or more HBAs available. This configuration supplements the SAN multipathing configuration that generally provides one or more switches in the SAN fabric and the one or more storage processors on the storage array device itself. In Figure 1-1, multiple physical paths connect each server with the storage device. For example, if HBA1 or the link between HBA1 and the FC switch fails, HBA2 takes over and provides the connection between the server and the switch. The process of one HBA taking over for another is called HBA failover.

Figure 1-1. Multipathing and Failover



Similarly, if SP1 fails or the links between SP1 and the switches breaks, SP2 takes over and provides the connection between the switch and the storage device. This process is called SP failover. VMware ESX/ESXi supports HBA and SP failover with its multipathing capability.

### Managing Multiple Paths inside VMware PSA (Pluggable Storage Architecture)

To manage storage multipathing, ESX/ESXi uses a special VMkernel layer, Pluggable Storage Architecture (PSA). The PSA is an open modular framework that coordinates the simultaneous operation of multiple multipathing plug-ins (MPPs).

The VMkernel multipathing plug-in that ESX/ESXi provides by default is the VMware Native Multipathing Plug-In (NMP).

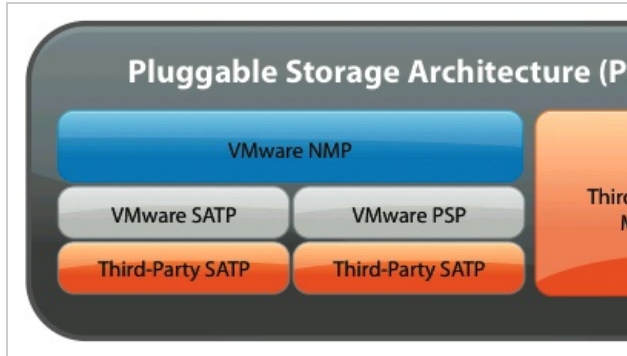
The NMP is an extensible module that manages sub-plug-ins. There are two types of NMP sub-plug-ins, Storage Array Type Plug-Ins (SATPs), and Path Selection Plug-Ins (PSPs). SATPs and PSPs can be built-in and provided by VMware, or can be provided by a third party.

If more multipathing functionality is required, a third party can also provide an MPP to run in addition to, or as a replacement for, the default NMP.

When coordinating the VMware NMP and any installed third-party MPPs, the PSA performs the following tasks:

- Loads and unloads multipathing plug-ins.
- Hides virtual machine specifics from a particular plug-in.
- Routes I/O requests for a specific logical device to the MPP managing that device.
- Handles I/O queuing to the logical devices.
- Implements logical device bandwidth sharing between virtual machines.
- Handles I/O queuing to the physical storage HBAs.
- Handles physical path discovery and removal.
- Provides logical device and physical path I/O statistics.

Figure 1-2. Pluggable Storage Architecture



The multipathing modules perform the following operations:

- Manage physical path claiming and unclaiming.
- Manage creation, registration, and deregistration of logical devices.
- Associate physical paths with logical devices.
- Process I/O requests to logical devices:
- Select an optimal physical path for the request.
- Depending on a storage device, perform specific actions necessary to handle path failures and I/O command retries.
- Support management tasks, such as abort or reset of logical devices.

#### VMware Multipathing Module

By default, ESX/ESXi provides an extensible multipathing module called the Native Multipathing Plug-In (NMP).

Generally, the VMware NMP supports all storage arrays listed on the VMware storage HCL and provides a default path selection algorithm based on the array type. The NMP associates a set of physical paths with a specific storage device, or LUN. The specific details of handling path failover for a given storage array are delegated to a Storage Array Type Plug-In (SATP). The specific details for determining which physical path is used to issue an I/O request to a storage device are handled by a Path Selection Plug-In (PSP). SATPs and PSPs are sub-plug-ins within the NMP module.

#### VMware SATPs

Storage Array Type Plug-Ins (SATPs) run in conjunction with the VMware NMP and are responsible for array specific operations.

ESX/ESXi offers an SATP for every type of array that VMware supports. These SATPs include an active/active SATP and active/passive SATP for non-specified storage arrays, and the local SATP for direct-attached storage. Each SATP accommodates special characteristics of a certain class of storage arrays and can perform the array specific operations required to detect path state and to activate an inactive path. As a result, the NMP module can work with multiple storage arrays without having to be aware of the storage device specifics. After the NMP determines which SATP to call for a specific storage device and associates the SATP with the physical paths for that storage device, the SATP implements the tasks that include the following:

- Monitors health of each physical path.
- Reports changes in the state of each physical path.
- Performs array-specific actions necessary for storage fail-over. For example, for active/passive devices, it can activate passive paths.

## VMware PSPs

Path Selection Plug-Ins (PSPs) run in conjunction with the VMware NMP and are responsible for choosing a physical path for I/O requests.

The VMware NMP assigns a default PSP for every logical device based on the SATP associated with the physical paths for that device. You can override the default PSP.

By default, the VMware NMP supports the following PSPs:

### Most Recently Used (MRU)

Selects the path the ESX/ESXi host used most recently to access the given device.

If this path becomes unavailable, the host switches to an alternative path and continues to use the new path while it is available.

### Fixed

Uses the designated preferred path, if it has been configured. Otherwise, it uses the first working path discovered at system boot time. If the host cannot use the preferred path, it selects a random alternative available path. The host automatically reverts back to the preferred path as soon as that path becomes available.

### Round Robin (RR)

Uses a path selection algorithm that rotates through all available paths enabling load balancing across the paths.

### VMware NMP Flow of I/O

When a virtual machine issues an I/O request to a storage device managed by the NMP, the following process takes place.

1. The NMP calls the PSP assigned to this storage device.
2. The PSP selects an appropriate physical path on which to issue the I/O.
3. If the I/O operation is successful, the NMP reports its completion.
4. If the I/O operation reports an error, the NMP calls an appropriate SATP.
5. The SATP interprets the I/O command errors and, when appropriate, activates inactive paths.
6. The PSP is called to select a new path on which to issue the I/O.

I hope this could help to have the initial background needed to understand why PSA architecture provides an incredible steps forward for customers, like me, that are using ALUA storage arrays.

In my next post I'll share my experiences moving from a manually FIXED PSP path to a new automatically balanced ROUND ROBIN path.

P U B L I C A T I O N . D A  
E T I C I P S S A T G R V A I R E U A, V S P A I R C E I

0 C O M M E N T I :

P O S T A U N C O M M

Commenta come: Seleziona profilo...

**Posta commento** Anteprima



[Post più recente](#)

[Home page](#)

[Post più vecchio](#)

Iscriviti a: [Commenti sul post \(Atom\)](#)