Dell Storage PS Series Arrays: Scalability and Growth in Virtual Environments

Dell EMC Engineering
January 2017

A Dell EMC Technical White Paper

Revisions

Date	Description
November 2011	Initial release
January 2017	Updated to reflect new branding and formatting

Acknowledgements

Engineering: Chuck Armstrong

The information in this publication is provided "as is." Dell Inc. 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 software described in this publication requires an applicable software license.

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

Table of contents

Executive summary

This document provides VMware® and Dell™ PS Series storage administrators with recommended best practices when configuring a virtual environment. This paper covers Dell recommendations on how to plan and scale virtual environments with PS Series arrays, specifically with regard to making connections to PS Series volumes and managing connection counts to a PS Series Group. This paper also covers connectivity and high availability as well as some common performance-tuning best practices. In addition, there is a discussion on various methods to implement for data-drive connectivity for the virtual machines.

Audience

The information in this guide is intended for VMware and PS Series storage administrators and installers.

1 Introduction

The PS Series SAN architecture allows for mobility of storage volumes across the peer storage arrays. This architecture also includes a Network Load Balancer (NLB) to redirect iSCSI login sessions to provide better efficiency and bandwidth management to the SAN. As new array members are added to a storage pool, new and existing volumes will load balance across the new members and the NLB will take advantage of newly introduced controller network ports to load balance connections into the SAN. This flexibility gives the PS Series SAN the ability to move data volumes seamlessly to any member or members in the group and maintain iSCSI sessions to the members that share the volumes. When not managed properly, this flexibility can also introduce growth and scaling issues that may not be of concern when initially deploying a PS Series SAN.

In firmware 9.0.x, the connection count limitation that has been tested and validated is 1,024 per pool with 4,096 per group (with 4 pools). For a list of supported connections per firmware version, see Table 1. This limitation is regardless of the actual number of members in the pool. Each member has to share all of the page table information as well as connection information, so this number does not scale by adding or subtracting members. This number is not a hard limit but a soft limit that was tested and validated. Exceeding 1,000 connections will generate errors in the logs and potentially dropped connections to volumes. Due to scale and sizing, keeping these connection counts lower will generally result in easier growth of the SAN.

This document is designed to assist in planning and deploying hypervisor environments. This document also addresses steps toward re-architecting an environment to help lower the overall number of connections to a PS Series Group, and provides suggested improvements to current environments.

Table 1 Supported PS Series Group connection counts by firmware version

PS Series firmware	PS6xxx storage pool connection limit	PS6xxx Group connection limit	PS4xxx storage pool connection limit	PS4xxx Group connection limit
Prior to v5.x.x	512	2,048	256	512
v5.x.x and newer	1,024	4,096	512	1,024

Note: Refer to http://www.eqlsupport.dell.com for more information on the technical specification of each array family and the supported connection limits they offer.

2 Virtual, hypervisor-based environments

Virtualization adds a new set of planning and scaling challenges to the data center. For example, the advanced virtual machine availability features in VMware, such as High Availability, Distributed Resource Scheduling, Fault Tolerance, and VMware vSphere® Storage vMotion®, depend on the ability of all the VMware ESXi™ hosts in a cluster to see all of the volumes. The same is true for Microsoft® Hyper-V® clusters utilizing Clustered Shared Volumes. Many customers start their virtual environments small. However, because these environments can scale easily by adding new servers and more storage, this often results in a redesign to accommodate the growth of the environment. This growth and scalability can also leave customers in situations where the growth of the virtual server environment exceeds the allowable active connections limits to the PS Series pool or Group.

For clusters in general, the formula for calculating the connection count to a PS Series pool is:

Total connections = hosts x connections per volume (sessions) x volumes x PS Series members in pool

Note: The number of VMs does not have anything to do with the connection-count formula but overhead must be taken in account for things like iSCSI-initiated connections from VMs, volume snapshots, or backup volumes.

That number is compounded when MPIO technologies like the PS Series Device Specific Module (DSM) for Windows[1] or the Multipathing Extension Module (MEM) for VMware vSphere[2] are enabled. These advanced MPIO modules install a connection management service with intelligent path-selection capabilities that maintains a table of contents on each host of all the relevant volumes on the PS Series Group members. This table of contents allows the MPIO service on the host to know exactly where the data blocks for an attached volume reside on the SAN and the path-selection policy assures the appropriate path is used to access the data. This allows for extremely efficient least queue depth MPIO with performance benefits.

The DSM for Windows and MEM for vSphere MPIO modules are based on the same algorithm that will, by default, create two connections per volume slice (volume slice = distribution of volume data across arrays in the same pool) supporting up to six total connections per volume. When using the PS Series DSM and MEM, it is important to understand how volume connections are created and managed. All PS Series MPIO modules are built to provide performance benefits by creating connections to all members that a volume spans, thereby providing a direct path from the host to the data being requested.

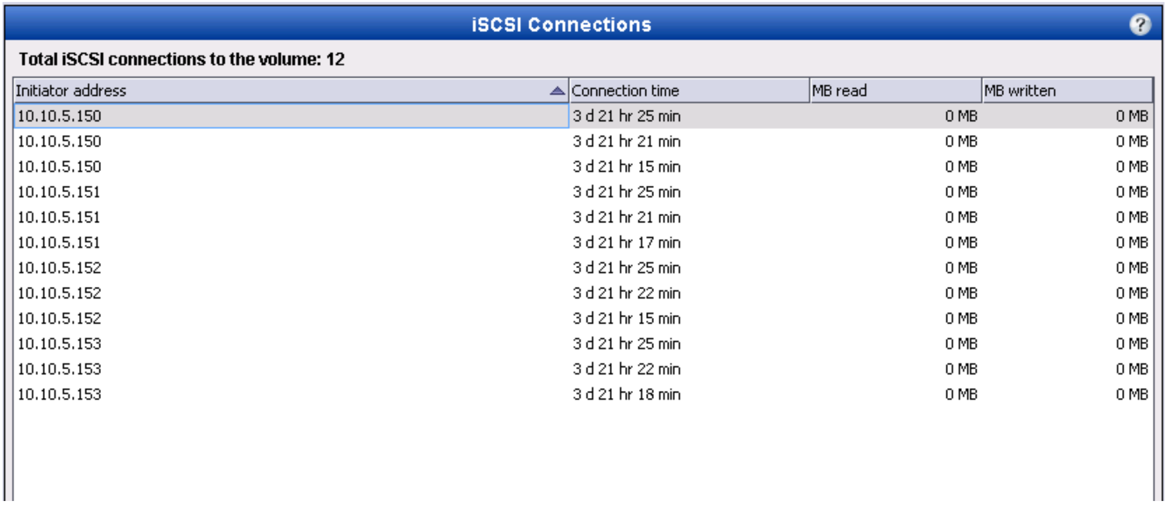
[1] See *Configuring and Deploying the Dell PS Series Multipath I/O Device Specific Module with Microsoft Windows*.

[2] See *Configuring and Installing the PS Series Multipathing Extension Module for VMware vSphere and PS Series SANs*.

In larger environments, these additional MPIO connections need to be taken into account at design time and as the environment scales. Here is an example of how quickly connections can be consumed: Assume there is a 16 node cluster connecting to 20 volumes in a single storage pool with 3 PS arrays. That equation looks like the following:

16 hosts x 6 session connections (2 sessions per volume slice x 3 members) x 20 volumes = **1920 connections to the pool**

This far exceeds the validated number of connections per pool. To control and avoid excessive consumption of connections, the MEM and DSM can be customized to scale down the number of connections made to volumes. As mentioned, by default these MPIO modules will make two connections per member and up to six connections per volume. Sometimes it may be necessary to scale down these default limits to reduce the number of connections made to a PS Series pool or Group. In this case, it is recommended to start by changing the **Max sessions per volume slice** setting from 2 to 1. This will reduce the amount of connections made to a volume in half, with only minimal impact on performance. See the following example for a 2-node cluster volume with the default setting of 2, compared to changing the setting to 1.

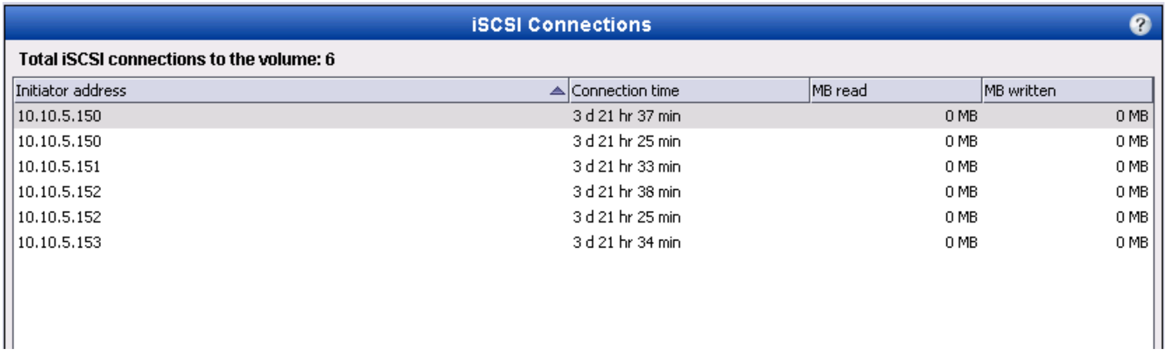


iSCSI Connections

Total iSCSI connections to the volume: 12

Initiator address	Connection time	MB read	MB written
10.10.5.150	3 d 21 hr 25 min	0 MB	0 MB
10.10.5.150	3 d 21 hr 21 min	0 MB	0 MB
10.10.5.150	3 d 21 hr 15 min	0 MB	0 MB
10.10.5.151	3 d 21 hr 25 min	0 MB	0 MB
10.10.5.151	3 d 21 hr 21 min	0 MB	0 MB
10.10.5.151	3 d 21 hr 17 min	0 MB	0 MB
10.10.5.152	3 d 21 hr 25 min	0 MB	0 MB
10.10.5.152	3 d 21 hr 22 min	0 MB	0 MB
10.10.5.152	3 d 21 hr 15 min	0 MB	0 MB
10.10.5.153	3 d 21 hr 25 min	0 MB	0 MB
10.10.5.153	3 d 21 hr 22 min	0 MB	0 MB
10.10.5.153	3 d 21 hr 18 min	0 MB	0 MB

Figure 1 Connections using default MPIO DSM settings



iSCSI Connections

Total iSCSI connections to the volume: 6

Initiator address	Connection time	MB read	MB written
10.10.5.150	3 d 21 hr 37 min	0 MB	0 MB
10.10.5.150	3 d 21 hr 25 min	0 MB	0 MB
10.10.5.151	3 d 21 hr 33 min	0 MB	0 MB
10.10.5.152	3 d 21 hr 38 min	0 MB	0 MB
10.10.5.152	3 d 21 hr 25 min	0 MB	0 MB
10.10.5.153	3 d 21 hr 34 min	0 MB	0 MB

Figure 2 Connections setting **Max sessions per volume slice** changed from 2 to 1

3 Larger volumes

Another tactic for reducing volume connections in virtual environments is to create larger volumes and fewer of them. In many cases, administrators start with a particular volume size to host virtual machines, and as their environment grows, the tendency is to create more volumes of the same size instead of creating or growing larger volumes to handle additional growth. This process only works for so long, because eventually the volume connection count starts to add up. PS Series storage supports the VMware VAAI Hardware Scalable Locking primitive which enables denser VMs per datastore configuration without causing reservation contention. Creating larger volumes or increasing existing volume sizes enables more virtual machines to be placed on an individual volume, resulting in fewer volumes needed, and therefore fewer connections consumed. This will allow environments to scale further without consuming additional connections to the SAN.

PS Series arrays support volume sizes up to 15TB. Microsoft Windows Server® 2012 and newer supports disk partitions up to 16EB (maximum tested size is 16TB) and since the release of VMware vSphere 5, VMware ESXi supports datastores up to 64TB (VMFS 5 removed the 2TB limitation). This adds ability in both hypervisor environments to utilize larger volume sizes to host VMs and ultimately reduce connection counts to the PS Series Group.

4 CHAP authentication for access control

Manipulating MPIO modules and increasing volume sizes are both great ways to manage volume connections but there are other best practices and recommendations that can also help control how connections are made to PS Series volumes. CHAP authentication for access control lists can be very beneficial. In fact, for larger cluster environments, CHAP is the preferred method of volume access authentication, from an ease-of-administration point of view.

Note: Additional information regarding CHAP can be found in the document, *Access Control Policies*.

Using CHAP can help improve scalability and ease management. When a new host server is introduced to a cluster, the administrator needs only to set the CHAP configuration on the iSCSI initiator to enable the new host access to all of the storage resources of the clusters. When a new volume is created, only one ACL entry and an initiator rescan is needed for all the cluster host servers to be able access the additional storage capacity.

Note: In a VMware environment, this also allows rapid movement of datastores between clusters. By simply unregistering the VMs, changing the ACL, and rescanning, entire volumes or datastores can be moved between clusters.

5 Group and pool transparency

When it comes to scaling PS arrays, there are two trains of thought: scale up or scale out. The PS Series architecture uses a scale-out architecture, but there are times when scaling out is not enough.

For example, customer A currently has 3 PS arrays and is purchasing 3 more. Looking at customer A's storage group, they have 3 arrays and 50 volumes deployed in a single storage pool. The new arrays arrive and customer A decides to add the 3 arrays to the same storage pool because they will gain load balancing benefits. Customer A would gain additional load balancing by doing this, but they are already consuming 400 iSCSI connections in the current pool. Instead of placing the 3 arrays in the same pool, it may make more sense to deploy a new storage pool, place the arrays in the new pool, and start creating volumes out of the new pool. This will allow customer A to start at 0 with the connection count in the new pool. Fortunately, for customer A, they can easily create the new pool and perform an online migration of the 3 new arrays from the current pool to the new pool without disrupting the other volumes or arrays in the current pool.

Operating systems do not distinguish between pools as long as the volume is given access to the server. The only difference between using a single pool or many pools is the load balancing or performance tier characteristics of the volume. A volume cannot load balance across storage pools, so to take advantage of PS Series multimember load balancing for a specific volume, there must be more than one array in the pool where the volume resides.

Another option is creating a new PS Series Group. By splitting up arrays into 2 groups, connection availability doubles, but there are some caveats to this approach in some environments. With a new group, a new dynamic discovery address has to be added to host iSCSI initiators and restrictions such as SAN copy offload or online volume migration are not available between groups. These limitations can be mitigated with some thought and proper design, such as using cluster systems that only connect to a single PS Series Group.

6 VMware pods

Creating VMware **pods** can help minimize connections in large VMware environments. The notion of VMware pods simply refers to reducing the number of nodes from large (many node) clusters to smaller clusters that are managed in the overall VMware vCenter® datacenter hierarchy. This allows for the continuation of HA/DRS and also allows better manageability of the volume connection count. Smaller cluster environments allow for continued scale and growth as new virtual resources are needed, without a constant redesign of the whole environment.

Each pod represents a vSphere cluster of nodes with all the available resource groups (such as HA/DRS) established, although a single pod does not utilize every ESXi host in the entire data center. These pods are still managed by one single vCenter datacenter. Inside the vCenter datacenter, administrators can leverage a **data mover** ESXi host that has access to multiple clusters. The data mover enables seamless virtual machine migration between pods, retaining full flexibility of the environment with additional growth possibilities.

Example of a data mover role:

1. Non-disruptively vMotion a VM to the data mover host.
2. Storage vMotion the VM from the data mover host to the shared datastore of the new pod.
3. vMotion the VM to the ESXi cluster of the new pod.

For example, using the equation described previously with 16 servers and 20 volumes, if 4 pod clusters of 4 ESXi hosts were created with connections to 5 volumes each (still maintaining the relationship of 16 ESXi hosts and 20 datastores), this results in the following reduction in connection counts:

4 ESX hosts x 6 sessions (2 x 3 member pool) x 5 volumes = **120 connections to the pool**

With four pods, this would only consume 480 connections to the pool. When compared to 1,920 connections as shown previously, this is a drastic reduction in total connections needed for this deployment example.

Additional advantages of using small over large clusters include:

- Easier structuring of virtual machines by organizational unit or purpose
- Better utilization of resource pools
- Increased management of data protection and disaster recovery by properly placing VMs on the respective protected datastores

The concept of pods may sound like an extra layer of management, but 16 ESXi hosts load balancing compared to 4 pods of 4 hosts load balancing is essentially the same in practice. Most VMs are **place and forget** in a pod, and the pod will automatically load balance. As the environment continues to grow and pods fill up, they can be grown or VMs can be moved to other pods.

7 Data protection and business continuity

Another important discussion point is data protection and business continuity. PS Series snapshots are done at the volume level, and in most cases, snapshots will not affect storage group connections if they are used to roll back the parent volume to a particular point in time. However, it is not uncommon to use snapshots as side-by-side copies in which the host connects directly to the snapshot instead of recovering the parent volume with a rollback operation. Tools such as Auto-Snapshot Manager VMware Edition (ASM/VE) and Auto-Snapshot Manager Microsoft Edition (ASM/ME) create and manage application and hypervisor consistent point-in-time smart copies (volume snapshots, clones, and replicas) as protection points for recovery operations. These tools also allow easy distribution of application data by allowing transportability of smart copies on the same hosts or other hosts in the data center. Since the entire VM is encapsulated into a set of files stored on a snapshot, the snapshot volumes can be brought online to any cluster for data recovery or even test and development. These operations are common with creating copies of application data for offloading operations, or using copies for selective recovery of object data. When these operations are used, the host(s) connect directly to the snapshot and MPIO tools make appropriate session connections to the snapshot, thus consuming connections to the pool as if it were a standard volume.

As customers grow and scale their environments, a common occurrence is virtual machine sprawl. This is when new VMs are created and placed randomly on datastores or volumes because they happen to have the most available space at the time. This does not impact the performance of the virtual machines because technologies like VMware DRS and Windows Dynamic Memory do a good job of balancing the compute/memory resources. However, this can have implications with data protection and disaster recovery with replication because the service level of protection must be able to protect the most critical machine no matter what else is on the volume. Snapshots and replicas are done at the volume level, and by creating volumes with specific tiers of protection and then assigning the correct VMs accordingly, you can gain a much better protection scheme for the environment. For example, a volume can live on a tier with a protection schedule that takes a snapshot every 4 hours and replicates once daily**,** and VMs that need that level of SLA would be placed there. Another higher priority volume could have a schedule that takes a snapshot every 1 hour, keeps 12 copies, and replicates every 3 hours. By prioritizing the VMs that reside on the volumes, better granular control of snapshot reserve and replication space is maintained and the VMs obtain a level of protection they may not have had before. Even more consistency can be obtained by using the Auto-Snapshot Manager family of tools for VMware and Hyper-V. These protection schemes do not replace traditional backup methods but augment the current schemes by adding additional layers of protection to the virtual machines.

8 Replication and scale

It is important to understand how replication plays into the connection-count limitations and design of a DR site. PS Series firmware allows replication reserve space to be selected from a single pool at the destination site. Recovery operations must be planned accordingly because any or all volume(s) from any pool in Site A can be replicated to Site B, but there must be enough replication reserve in the pool designated at Site B to accommodate replication activities from Site A.

Note: If replication space capacity is of concern, PS6610 series arrays are recommended due to their high density storage capacity.

When planning full site replication and recovery, thought must be used in planning the recovery strategy at Site B. With the possibility of replicating volumes from up to 4 concurrent pools from Site A, careful attention must be taken to understand the number of connections that may be needed during a recovery operation at Site B and create the storage environment accordingly.

One approach is to add up the total number of connections to the PS Series SAN at Site A. If the total connection count is within the limits of a single storage pool, nothing else is needed. Be mindful of any additional volumes that may be created in the future and added to the replication plan. If the total connections are above the maximum number supported for a single pool, then the best approach for Site B is to create multiple PS Series Groups to ensure enough connections for each volume to recover.

Another approach is to mimic each pool at Site A with a separate PS Series Group at Site B. What this means is that the PS Series SAN at Site A has pools 1–4. At Site B, there would be 4 separate PS Series Groups created. The design would look similar to: Site A/Pool 1 would replicate to Site B/Group 1 DR, and Site A/Pool 2 to Site B/Group 2 DR and so on. This will allow a similar mirrored environment to be established for DR.

If an active/active (bi-directional replication) environment is configured or if one site has fewer resources than the other, use careful consideration when designing recovery plans to make sure that on a failover, the environment can handle the connection load required.

PS Series arrays, ASM/ME, ASM/VE, and VMware Site Recovery Manager support bi-directional replication between PS Series SANs which means Site A can have virtual machines that are replicated to Site B and Site B can have virtual machines that are replicated to Site A. A disaster at either site will allow for the VMs at the other site to be brought online quickly, but this requires both sites to have sufficient resources to not only host their current environment but the recovery environment as well. This can be factored in by determining which VMs require higher priority as part of the recovery planning step when sizing the two sides.

9 Summary

With virtualized data center environments being the standard, it is important that administrators understand exactly how to manage and control growth and sprawl of resources. Managing multipath configurations, larger volume sizes, CHAP authentication, utilizing group and storage pool transparency, and distributing cluster environments are all ways to scale and manage connections to PS Series SANs for virtualized environments. With a little planning and a good understanding of how volume connections are used and controlled, virtual environments can continue to scale and grow many times over to meet the most demanding data center requirements.

A Technical support and resources

Dell.com/Support is focused on meeting customer needs with proven services and support.

Dell TechCenter is an online technical community where IT professionals have access to numerous resources for Dell software, hardware and services.

Storage Solutions Technical Documents on Dell TechCenter provide expertise that helps to ensure customer success on Dell Storage platforms.

A.1 Related documentation

Articles, demos, technical documentation and more details about the PS Series Storage product family are available at PS Series Technical Documents.

For an updated *Dell Storage Compatibility Matrix*, visit http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/20438558.

For detailed information about PS Series Storage arrays, groups, volumes, array software, and host software, log in to eqlsupport.dell.com.

Table 2 Referenced documentation

Vendor	Document title
Dell	*Dell PS Series Arrays: Advanced Storage Features in VMware vSphere*
Dell	*Configuring iSCSI Connectivity with VMware vSphere 6 and Dell PS Series Storage*
Dell	*Configuring and Deploying the Dell EqualLogic Multipath Device Specific Module with Microsoft Windows*
Dell	*Configuring and Installing the PS Series Multipathing Extension Module for VMware vSphere PS Series SANs*