Dell Storage PS Series Architecture: Hybrid Array Load Balancer

The Dell PS Series architecture is a virtualized, self-tuning system that uses intelligent load balancers to maintain system performance at optimum levels. This technical report describes the hybrid array load balancer and how it works.

Dell Engineering
February 2017

A Dell EMC White Paper

Revisions

Date	Description
March 2013	Initial release
February 2017	Update to industry changes

Table of contents

Executive summary

In many organizations, significant effort is devoted to tuning resources: modifying network settings on switches, changing various parameters on servers, re-configuring QoS on the WAN, and continually altering the layout of data on the SAN. Estimates indicate that 80% of most IT budgets are dedicated to operations and maintenance, leaving only 20% for technical innovation to enhance the business. The Dell™ PS Series array solutions help reduce the operational cost associated with storage by automating management tasks and optimizing resource utilization. By leveraging PS Series solutions, organizations can maximize storage efficiency by ensuring that the right data is in the right place at the right cost.

1 PS Series Storage

Dell PS Series storage is an award winning, virtualized, scale-out SAN architecture that provides simplified management, ease of use, a comprehensive set of data protection and recovery tools and great ROI for both large and small organizations. PS Series arrays are designed and built to be highly available and work together to provide virtualized SAN resources called a group. The inclusive software model means never having to decipher a complex licensing scheme, and never having to purchase software upgrades to receive new features.

1.1 The benefit of automatically balancing SAN resources

The IT resources available to most organizations are limited. The realities of budgets, staffing, knowledge, and the business needs of the organization are often competing and conflicting, forcing IT managers to optimize the environment for the proper balance of these items. The usual result is that tuning for optimal performance is deprioritized in favor of more pressing items. Unfortunately, the demands that are placed on a SAN after installation are rarely the same as those that were anticipated at the time the SAN was installed. For example, applications may have been added or removed from the environment, and user loads may have increased or decreased based on the evolving needs of the business. Typically, due to data growth, the number of disks or other storage devices increases, often dramatically.

For a classic non-virtualized storage architecture there are normally only two options to optimize performance once these inevitable changes have occurred and neither of them are simple. Either re-layout all of the data on the storage array manually or replace the existing array in its entirety and begin the process of performance tuning from scratch.

However, PS Series arrays provide an alternative strategy for tuning that does not strain IT resources in either of those ways. Because of the virtualized scale-out architecture of the PS Series arrays, adding resources to meet the demands of the business is easy – administrators simply add a new member to an existing PS Series group and more space is immediately available. The system will automatically rebalance the load across the new and existing group members, with no downtime and in a way that is completely transparent to the servers, applications and users.

Re-deploying or retiring older equipment is equally simple, just choose the member to remove, and it will automatically re-locate its data to other members in its pool and extract itself from the group and be immediately ready for re-deployment or retirement. This, too, occurs seamlessly and transparently with no downtime.

2 PS Series hybrid array and load balance

One line of PS Series arrays is aimed at increasing performance over HDD-only arrays while balancing cost associated with SSD-only arrays. These hybrid models are available in several configurations, combined SSD devices with HDDs creating a tiered platform within a single array.

The current hybrid arrays available are:

- PS6210XS
- PS6610ES
- PS4120XS
- PS-M4110XS

A hybrid array supports only the RAID 6 - accelerated RAID type. Both the SSDs and HDDs are configured in RAID 6 configurations, using one of the HDDs as a universal hot spare for protection. The PS6610 is configured with both SSD and HDD hot spares. In addition, a small portion of the SSD RAID set is dedicated to accelerating random writes that are targeted to the HDD RAID set by temporarily buffering these writes for the HDDs. This space is used automatically should the HDDs be unable to keep up with the incoming write requests.

The hybrid array load balancer was created to facilitate automatic movement of data between SSD and HDD media within the array to enable SSD performance for the active subset of data and HDD capacity for the remainder.

This type of internally-tiered array is able to respond to dynamic, tiered workloads faster and with less overhead than moving data over the SAN to other array types.

The workloads targeted with the hybrid arrays are tiered in nature and have medium to high random I/O, and low sequential I/O requirements. Examples of workloads that behave in this fashion are OLTP databases and VDI environments.

2.1 How the load balancer works

PS Series, like many virtualized storage products, use a concept known as pages to logically separate the volume that is presented to the hosts from the physical resources of the hardware devices. In a PS Series array, a page represents 15MB of physical storage. The goal of the hybrid array load balancer (HALB) is to move the most active pages to the SSD tier of storage within the array.

As data is written to the hybrid array, writes to volumes are initially placed within the SSD tier until it is 2/3 full, at which point writes to volumes are spread across both SAS and SSD drives.

Pages are monitored and page heat is tracked within seconds for each page in the array. Every read or write to a page will increases the page heat. The HALB determines how likely a page will be accessed based on the page heat for that page. As page heat increases, the load balancer moves pages to an SSD from the HDD.

The levels of page heat that are tracked, and the action of the load balancer based on each type, is based on a combination of factors described in Table 1.

Table 1 HALB Decision Matrix

Heat Index	If the page is on an SSD	If the page is on an HDD
Hot	Do nothing, page is optimized	Attempt to move to SSD
Warm	Do nothing, page is optimized	Move to an SSD if there are cold pages on the SSD to move and if no hot pages have been identified to be optimized.
Cold	Move to HDD if hot or warm pages exist on HDD and space is needed to allow them to move to SSD	Do nothing, page is optimized

Balancing operations of pages occur as a background task every two minutes within the array. If there are a large number of pages to be moved, only a portion of them are moved in any one cycle in order to ensure that user I/O is given priority.

2.2 How the PS Series hybrid array HDD write accelerator works

Normally, as with any write-back cache, writes destined for the HDD tier are cached by the memory of the hybrid array. These writes are protected by mirroring the cache of the dual controllers, as they are in any other PS Series array. Writes are then optimized and destaged to disk. However, there are times when an application may be sending writes faster than the HDD tier, where the page resides, can absorb the write activity. In order to ensure high application performance, the hybrid array HDD write accelerator can store short-term, high-volume bursts of random writes destined for the HDD tier in the SSD tier to act as an extended write cache.

In order to accomplish this, a modest portion of the SSD tier is reserved by the array to act as the HDD Write Accelerator. The size of this space is determined based on the array type and firmware version and cannot be viewed directly or modified.

If a burst of random writes to pages stored on HDD media is being serviced, data is written directly to this portion of the SSD and mapped to the intended HDD location in order to enhance random write performance of the application. As a background operation, and when the array performance threshold allows, that data is then destaged to the HDDs, to make room in the SSD buffer for additional write burst caching.

In the hybrid array, write caching is performed in the following priority:

1. Service writes in controller cache until it is exhausted.
2. Service writes in SSD cache until it is exhausted.
3. Service writes to HDD.

Using the Automatic Performance Load Balancer (APLB) with hybrid arrays

Since the Capacity Load Balancer treats hybrid arrays like any other member in the PS Series group, capacity is assigned to them in relation to their overall capacity contribution to the pool. The APLB is able to work in conjunction with the hybrid arrays, permitting other, non-hybrid arrays to migrate active data to the hybrid arrays in exchange for inactive data. In the case where more active data is placed on a hybrid array than the

SSD tier is able to hold, the APLB is able to migrate active data from the HDD tier to another member such as a second hybrid or standard array that may have more resources available to better service the I/O requests.

Load Balancing of arrays within pools is discussed more extensively in *Dell Storage PS Series Architecture: Load Balancers*.

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 EMC software, hardware and services.

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

A.1 Related resources

The following list includes documents referred to in this paper:

Dell EqualLogic Group Manager Administrator's Guide available on eqlsupport.dell.com (login required)

TR1070: *Dell Storage PS Series Architecture: Load Balancers*