PAC

STORAGE

PAC Storage Scale-Out NAS Family Technical Overview

This whitepaper describes technical designs and details of the PAC Storage scale-out NAS

Version: 1.1

Updated: December 2019

Contents

PAC Storage Scale-Out NAS System Overview

PAC Storage Scale-Out NAS is a system with high performance and capacity expansion capabilities. By adding more nodes to the system, you can easily get high performance and capacity in a single namespace file system.

PS Scale-Out NAS offers multiple data protection designs to guarantee that enterprise customers' data are highly protected. It provides RAID for disk protection, Erasure Code and Replica for node protection, and remote replication for cluster failure protection. The Self-healing function can further maintain the same protection level when node failure occurs.

PS Scale-Out NAS features complete integration of software and hardware, with an easy-to-use web management interface and swappable modular hardware design, user is able to deploy the system in 30 minutes and easily maintain it.

PAC Storage Scale-Out NAS Nodes

PS Scale-Out NAS cluster consists of multiple nodes. A node is an appliance in a rack form which includes CPU, memory, networking, and drives that provide network file storage capability by clustering multiple nodes in a single cluster; more compute power, networking, and storage capacity can be provided.

PS Scale-Out NAS provides multiple hardware models, depending on performance and capacity needs, a user can choose hardware model suitable for his applications.

The table below shows each PS Scale-Out NAS series' corresponding number of CPU core and available form factors.

Series	CPU Core	Model
PS 6400	12	PS 6416 (G)
		PS 6424 (G)
		PS 6425
		PS 6460 (D) / PS 6460 (G)

PAC Storage Scale-Out NAS Family Model Name

- G indicates there is 1 node in the appliance, and D indicates there are 2 nodes in one appliance.
- PS 6425 supports SSD only.
- Only PS 6425 can be configured as SSD node.

A PS Scale-Out NAS cluster can scale out to 144 nodes for now, and minimum deployment node number is one node for Distributed mode, and three nodes for Erasure Code and Replica configuration. Currently, PS Scale-Out NAS provides the HDD/SSD Node configurations. Drives in PS Scale-Out NAS node should be fully populated to provide maximum performance and capacity, and both HDD and SSD are supported.

Each Node can also scale-up the node capacity through the SAS expansion channel (12 Gb/s), so that the user can expand system capacity with minimal initial configuration. The expansion enclosure will be automatically configured when connected to head node, no further configuration is required. A node can support up to 76 drives for the 3U16 bay model/84 drives for the 3U24 bay model/100 drives for the 2U25 bay model/120 drives for the 4U60 bay model.
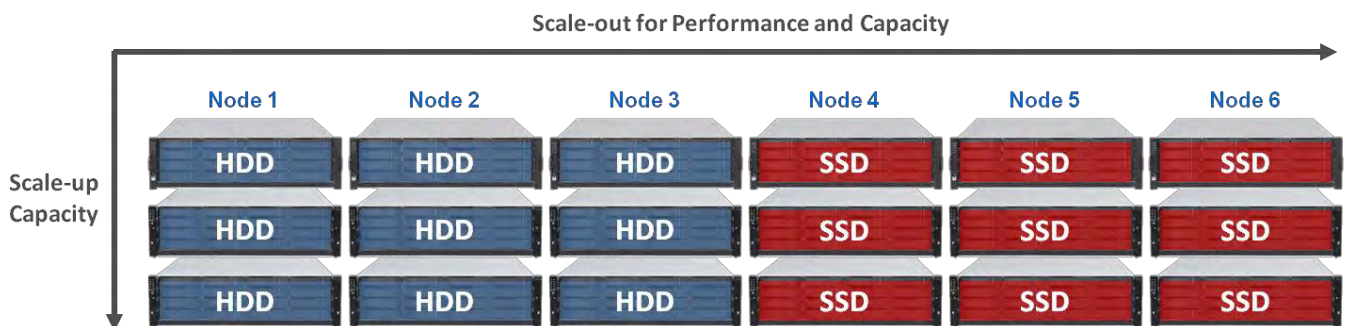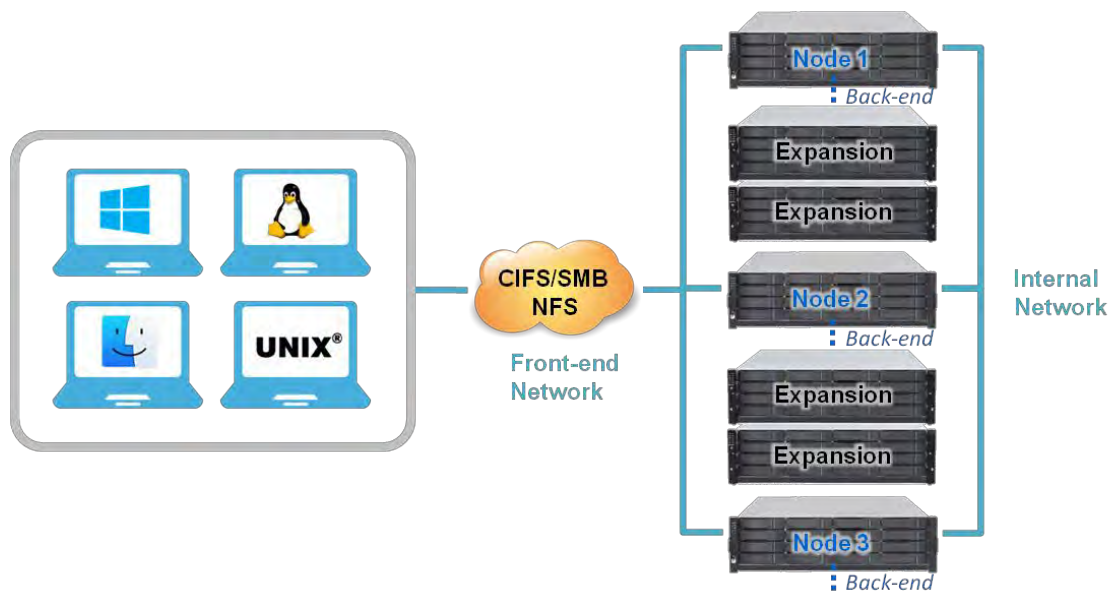

Scale-out and Scale-up of PS Scale-Out NAS Cluster

Network

There are two networks for the cluster, which can be monitored through the management page.

● Front-end network – This network is for user clients to access cluster nodes for data IO, since PS Scale-Out NAS supports single namespace, any client can connect to any node to access the same file. Clients can use 10GbE (SFP+)/25GbE (SFP28)/40GbE (QSFP+), and NFS/CIFS/FTP to connect to the nodes.

● Internal network – This network is for nodes internal communication and should be isolated to the user client network. Through 10GbE (SFP+), 25GbE (SFP28), and 40GbE (QSFP+) interface, the nodes are all connected for metadata sync and data forward. The trunk function is enabled by default, and high availability architecture can be achieved through the redundant switch.

Network of PS Scale-Out NAS Cluster

I/O Load Balancing

PS Scale-Out NAS can scale out to multiple nodes to provide service for multiple clients, to make sure all nodes can provide service to maximize overall system performance, and data will be evenly distributed to all nodes. Moreover, to effectively balance IO loads between the nodes, PS Scale-Out NAS supports DNS Round–robin for clients to access to different nodes whenever they connect to a node, this will make IO balance for all nodes and, therefore, the system can provide its best performance.
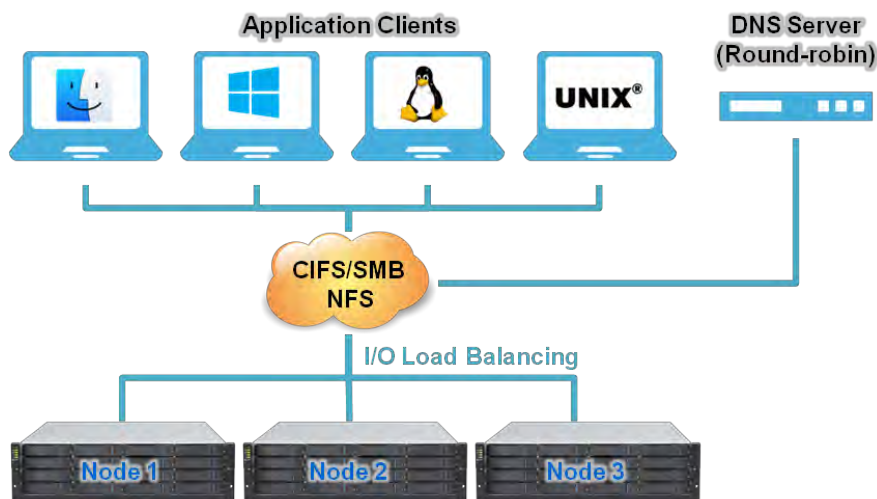
DNS Round-Robin for I/O Load Balancing

Provisioning Structure

Disk Pools

PS Scale-Out NAS supports disk pools for different application needs, and user is allowed to create disk pool based on different protection levels or disk type. For example, user can create Replica disk pool for critical data, and Distributed disk pool for application requiring high performance. PS Scale-Out NAS also supports cloud pool, multiple public cloud providers are supported, including Amazon, Microsoft, and Alibaba; user can store backup and archive data to cloud pool.

All disk pools can be managed in a single namespace, this eliminates management difficulty when using NAS system. When creating a shared folder, user can specify a disk pool to store data and change to another disk pool later on with just a single click, the folder path will not be changed, and data will be migrated automatically from one pool to another.

Create shared folder on selected disk pools

RAID Logical Drive

PS Scale-Out NAS uses PAC Storage RAID Protection technology to provide system high availability and high performance. When first-time deployment, user can specify RAID level to protect the disk, the system will automatically create Logical Drives (LDs) for all the node's disks and provide as the base unit of the disk pool. The RAID level cannot be changed after deployment when adding a new node or JBOD, the same RAID level will be applied.

Data Protection

PS Scale-Out NAS supports multiple data protection mechanisms to protect the system from disk, node or even cluster failure. We will go into details for each protection mechanism.

cluster.

Erasure Coding / Replica

PS Scale-Out NAS provides Erasure code and Replica to distribute data across multiple nodes, so when there is a node or LD failure, data can still be protected and accessed. Erasure code is more space-efficient, it is recommended to use Erasure Code when there are more nodes in the cluster, and Replica provides higher read performance compared to Erasure code, you can also use it when there are only a few nodes in the cluster.

File Write/Read

When data is written to the node in the cluster, it will first be distributed to other nodes depending on disk pool protection level, and only when data is completely written to all nodes, the system will respond to the client that write has been completed. For better performance, a memory cache is used to store the data, and it is protected by the battery.

When the client reads data from the node in the cluster, data may be retrieved from other nodes via the internal network, and then it assembles the file if necessary and returns to the client.
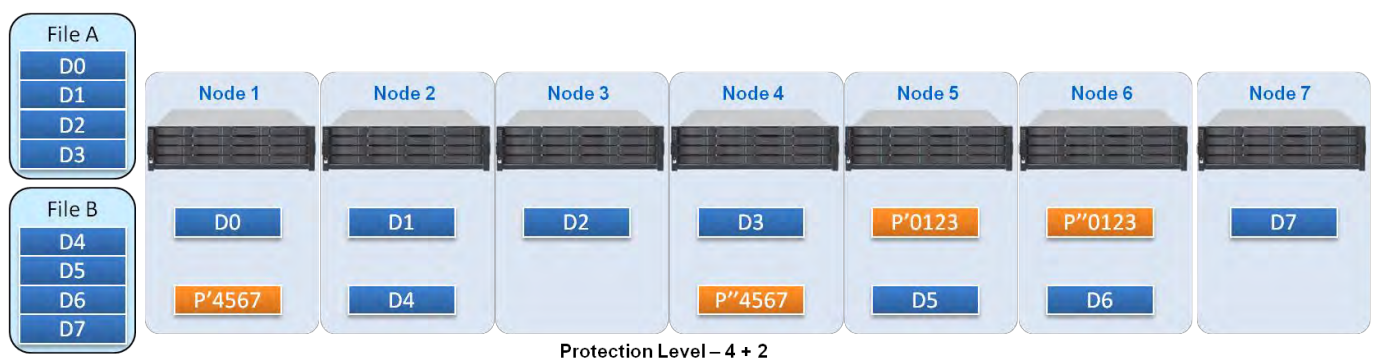
Protection Level

PS Scale-Out NAS supports the following data protection levels: Distributed, Erasure Code, and Replica.

In 'Distributed' configuration, only one copy of a file is stored in the cluster, and, therefore, higher performance is provided. When a client tries to store a file, the system will decide to store the file in which node by a hash algorithm, and then store the whole file within the chosen node. There is no failure tolerance for nodes, but the files are still protected by RAID.

On 'Erasure Code' level, user is allowed to specify N+M level protection, while N is the number of nodes to stay alive, and м represents the number of nodes that can be lost for the cluster to continue to operate.

Protection Type	Protection Level (N+M)	Description
Distributed	1	Configure with 1 node. Each data will be stored in single node.
Erasure coding	2+1	Needs at least 3 nodes to configure. Allows losing 1 node.
	4+1	Needs at least 5 nodes to configure. Allows losing 1 node.
	4+2	Needs at least 6 nodes to configure. Allows losing 2 nodes.
	8+1	Needs at least 9 nodes to configure. Allows losing 1 node.
	8+2	Needs at least 10 nodes to configure. Allows losing 2 nodes.
Replica	X2	Needs at least 3 nodes to configure, replicates 2 copies. Allows losing 1 node.
	X3	Needs at least 3 nodes to configure, replicates 3 copies. Allows losing 2 nodes.

When the protection level is Erasure Code N+M, data will be sliced into N pieces and M data parity will be generated. The N pieces of data and M parity will then be randomly written to N+M nodes based on node capacity. When there is a node loss, data can be calculated through the parity. The following is an example of the stripe unit storage of Erasure coding 4+2.

Erasure Coding N+M Protection

On the other hand, when the protection level is 'Replica', data will be replicated to other nodes in the cluster to make multiple copies. The following is a sample stripe unit storage example of Replica X2.
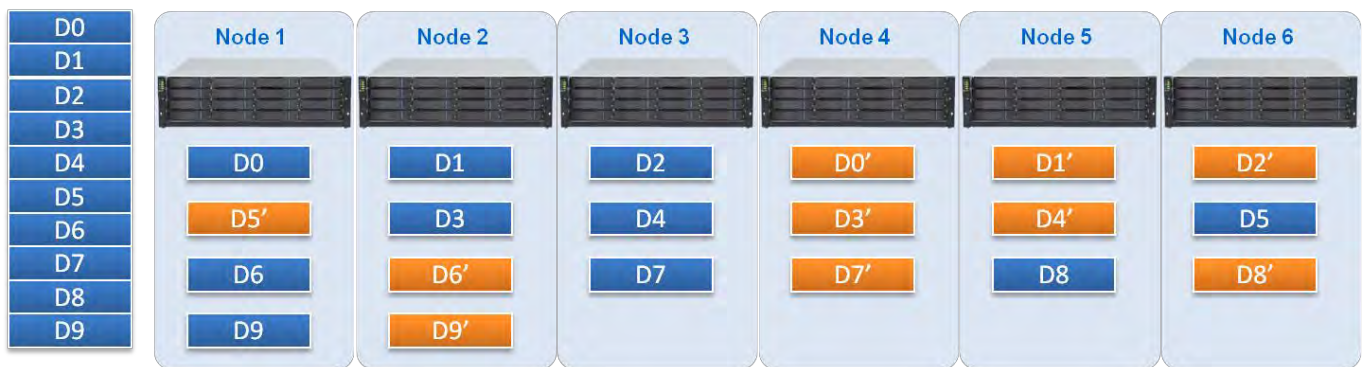
Protection Level – X2

Replica Protection

Healing

PS Scale-Out NAS supports self-healing function. When there is node or LD failure, the system can start the healing process to maintain the system protection level. The followings are details of the healing process.

- Node failure healing – When there is node or expansion unit failure, the system will fall into degraded status, if node or expansion unit cannot come back in some time, self-healing process will be triggered to write data on the failed node or expansion unit to other active nodes. Once the self-healing process is completed, the system will return to a normal status and data protection level will be the same as before.
- LD failure healing – Where there is LD failure, the system will fall into degraded status, the self-healing process will be started to write data on failed LD to other normal LDs after some time. Normally the LD in the same node will first be used to write data, and only when there is no capacity left, the data will be written to LDs in other nodes.

PAC Storage RAID Technology

PS Scale-Out NAS uses PAC Storage RAID technology to protect disk in the node. RAID (Redundant Array of Independent Disks) is a storage virtualization technology that integrates multiple hard disks to provide better performance and data protection. Different RAID levels can be selected based on the user's capacity and protection level requirements. There are three configurations for the PS Scale-Out NAS RAID level setting.

(1) Optimal Performance – This configuration provides the best system performance; it uses RAID 5 for disk protection. Several LDs will be created to maintain certain level disk protection, and the exact LD number is dependent on the enclosure drive number.

(2) Balance – The same as Optimal Performance mode, Balance mode also uses RAID 5 for disk protection, the only difference is that in Balance mode, there will be a spare drive in each enclosure, this can provide higher level protection, but on the other hand, usable capacity will be smaller.

(3) Optimal Protection – This configuration uses RAID 6 for disk protection and, therefore, allows 2 drives failure for each LD. Compared to Optimal Performance mode, there will be slight performance downgrade, but it can offer better drive failure protection.

Unlike the architecture using Erasure coding/Replica across nodes, RAID technology can achieve a single node's scale-up stable extension and reduce the failure domain problem.

Optimal Performance

Optimal Protection

Balance
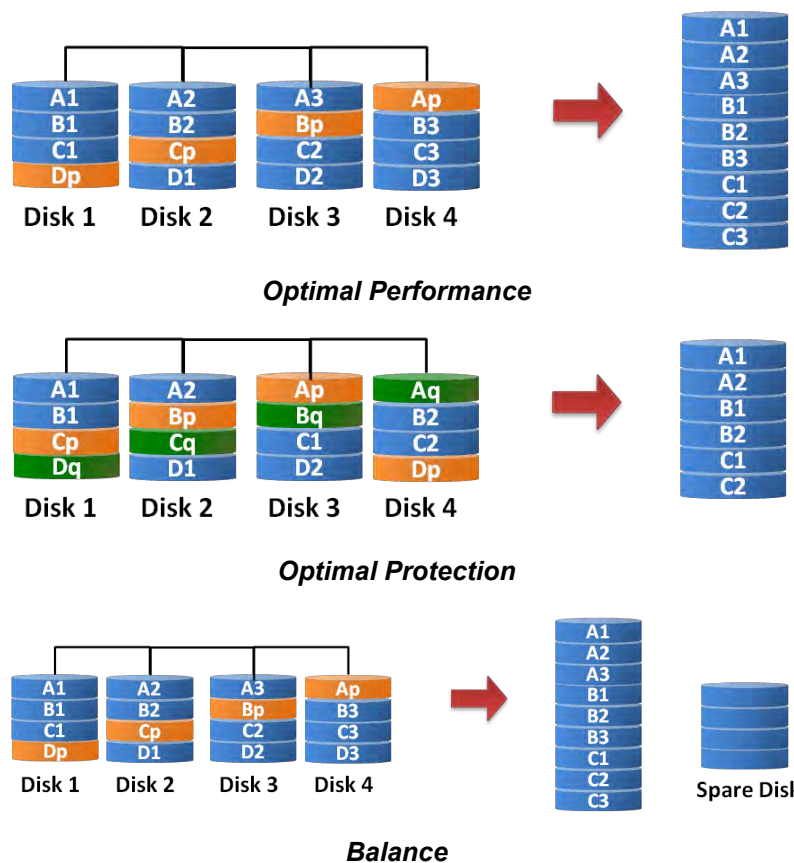
Online Expanding

PS Scale-Out NAS supports both scale-out and scale-up online expansion. During the expansion process, the system does not need to be shut down and, therefore, maintains high availability.
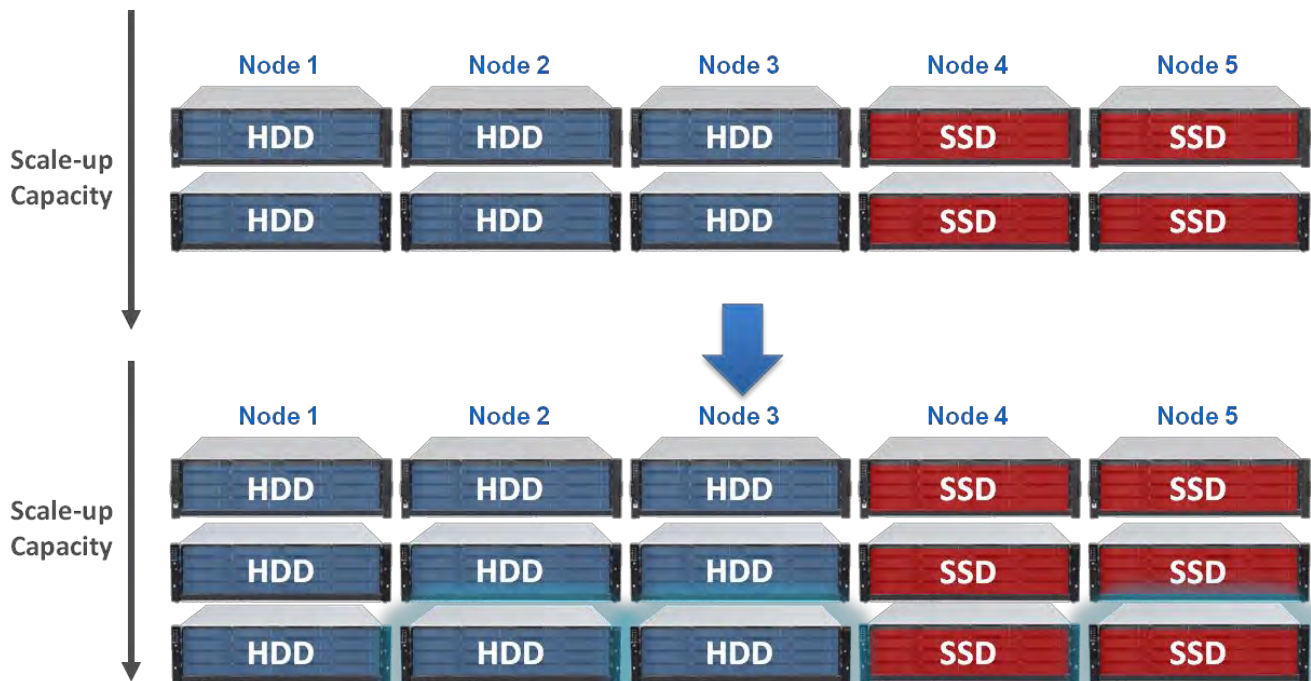
Scale Out

By adding nodes to the cluster, you can linearly increase system performance and capacity. Each node can provide extra CPU, network bandwidth, and storage capacity. Only with a few clicks, a node can be added to the cluster, and user is allowed to re-balance data in the cluster to make sure all nodes anticipate providing IO.

Scale Up

If more capacity is needed, PS Scale-Out NAS provides a scale-up extension by adding an expansion enclosure. It should be noted that nodes belonging to the same disk pool must extend the same expansion enclosure number to achieve the best data balance. The PS Scale-Out NAS family has some scale-up rules that need users to follow.

Please check the below table for the scale-up configuration rule.

PS model	Maximum Expansion Enclosure (per node)	Total Number of Drives (per node)
PS 3U16 model	PS16J3 * 3	64
	PS60J4 * 1	76
PS 4U24 model	PS16J3 * 3	72
	PS60J4 * 1	84
PS 2U25 model	PS25J2 * 3	100
PS 4U60 model	PS16J3 * 3	78 (D model) 108 (G model)
	PS60J4 * 1	90 (D model) 120 (G model)

Scale-up Rule

4U 60-bay High-density Expansion Enclosure

Expanding Capacity and Performance

Unlike other scale-out NAS products that can only add a node to increase both performance and capacity, PAC Storage Scale-Out NAS has the flexibility to add expansion enclosures to expand capacity only. This is useful when user needs more capacity for the cluster, and it is more cost-effective compared to other scale-out NAS products.

Firmware Upgrade

PS Scale-Out NAS provides firmware upgrade for new features and bug fixes, and there are two types of firmware upgrade.

Rolling Firmware Upgrade

Rolling firmware upgrade will upgrade nodes in the cluster one by one, nodes that are not being upgraded will continue to work and provide service to the clients. There is no downtime for the cluster during the upgrade.

Immediate Firmware Upgrade

Immediate firmware upgrade will stop the service and upgrade all nodes at the same time, there will be a system downtime and, therefore, it is recommended to do this during off-hours.

PAC Scale-Out NAS Management

Management Interface

Web-GUI

PS Scale-Out NAS provides a central web management interface for system management, administrator can connect to the PS Scale-Out NAS management IP to use PAC Management Software to do the management, and no additional management server is needed. The PAC Management interface can manage the whole cluster system and all nodes, and when there is no failure, the PAC Management software can still operate without impact.

SSL Connection

When log in to the management interface, data is encrypted by SSL to ensure connection security.

Deployment Wizard

A Deployment Wizard is provided to guide user step-by-step to build the cluster in just a few minutes.

Authentication and Access Control

PS Scale-Out NAS provides the following user management for client to connect to the cluster for accessing the file system.

- Active Directory (AD)
- LDAP (Lightweight Directory Access Protocol)
- NIS (Network Information Service)
- Local Users & Groups

Data Service Software

PS Scale-Out NAS provides comprehensive data service software features.

Performance Monitor

The PAC Management software interface can display real-time system performance, and administrators can monitor performance based on interfaces, such as cluster, node, and network interface.

Quota Management

The ability to set a capacity quota for the shared folder allows the administrator to plan the usable capacity of the shared folder for different purposes.

Remote Replication (Rsync)

Offsite replication supports Rsync protocol, which can copy data between devices that support Rsync, back up data, and backup can be periodically replicated through scheduling.

Write Once Read Many (WORM)

PS Scale-Out NAS supports WORM: once data is written, it becomes 'read only', and cannot be modified, deleted, or overwritten. WORM can be enabled by disk pool, and once enabled, it cannot be disabled again. WORM can prevent accidental modification or deletion of critical data to meet regulatory or management requirements.

Self-encrypting Drives (SED)

PS Scale-Out NAS supports SED that automatically and transparently encrypts data written to the disk without degrading performance. The SED uses AES 256 to encrypt data. The encryption key is generated randomly and, therefore, is secure enough to protect user data.

PAC Storage Cloud Gateway

With PAC Storage Cloud Gateway, user can easily upload data to public cloud storage. When upload data to the cloud, data will first be stored in PS Scale-Out NAS as a cache buffer, and then automatically transferred to public cloud by secure channel without user interruption. PS Cloud Gateway supports most of the mainstream cloud providers: Amazon S3, Microsoft Azure, Alibaba AliCloud, Baidu Cloud, Tencent Cloud, and OpenStack.

Deployment & Expanding Example

This section will explain how to deploy and expand the PS Scale-Out NAS cluster through several scenarios.

Scenario 1 – Deployment

The company purchases scale-out NAS and create a shared folder for marketing department with folder name "Marketing". There will be 3 PS Scale-Out NAS nodes, and all equipped with NL-SAS hard drives.

(1) Follow Deployment Guide to connect network cables to the PS Scale-Out NAS external, internal, and management interfaces, - and power on the nodes.

(2) Use browser to connect to Deployment Wizard to add all three nodes to the cluster and complete the deployment process.

(3) Connect to the PAC Management interface to create a disk pool "Marketing disk pool" and set the protection level to Erasure Code 2+1.

(4) Create a shared folder "Marketing" in the "Marketing disk pool".

(5) Enable the network sharing protocol CIFS/SMB and NFS.

(6) Connect to Active Directory (AD) service and import all users.

(7) Select user from marketing department and assign permissions for "Marketing" shared folder.
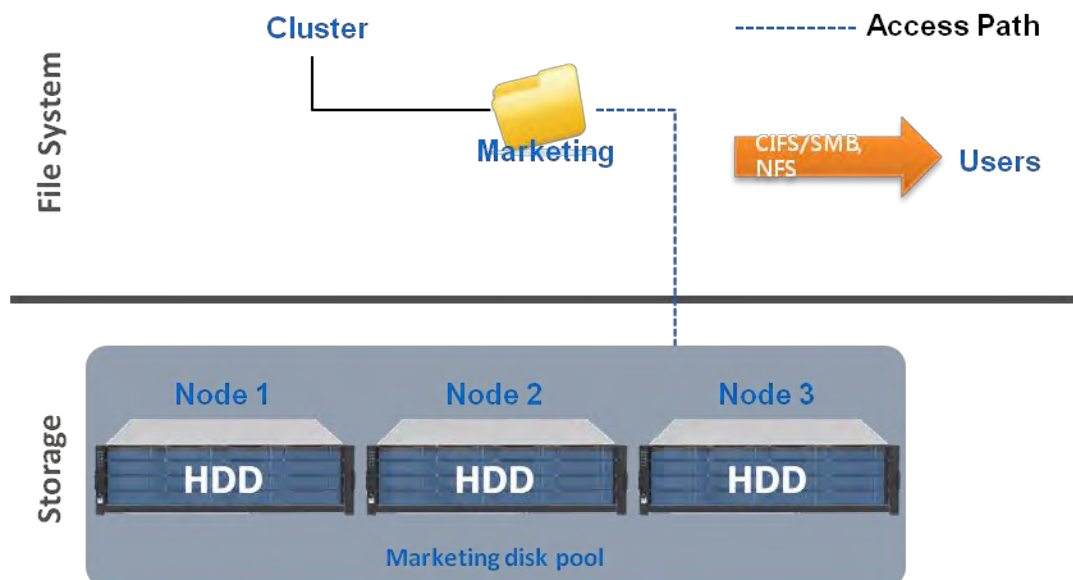
Scenario 2 – Expanding Node

The company expects to extend a scale-out NAS service to the R&D department while the 3 nods cluster has run out of capacity and performance, and, therefore, 3 additional PAC Scale-Out NAS nodes will be purchased. A shared folder "R&D" for the R&D department will be created.

(1) Connect 3 new nodes to the existing cluster networks and power on.

(2) Login to the PAC Management interface and click 'Add Node' to add 3 nodes to the cluster.

(3) Create "R&D disk pool" capacity to use 3 newly added nodes and specify protection level to Replica x 3.

(4) Create a shared folder named "R&D" in "R&D disk pool".

(5) Select users from the R&D group and assign permission for "R&D" shared folder.
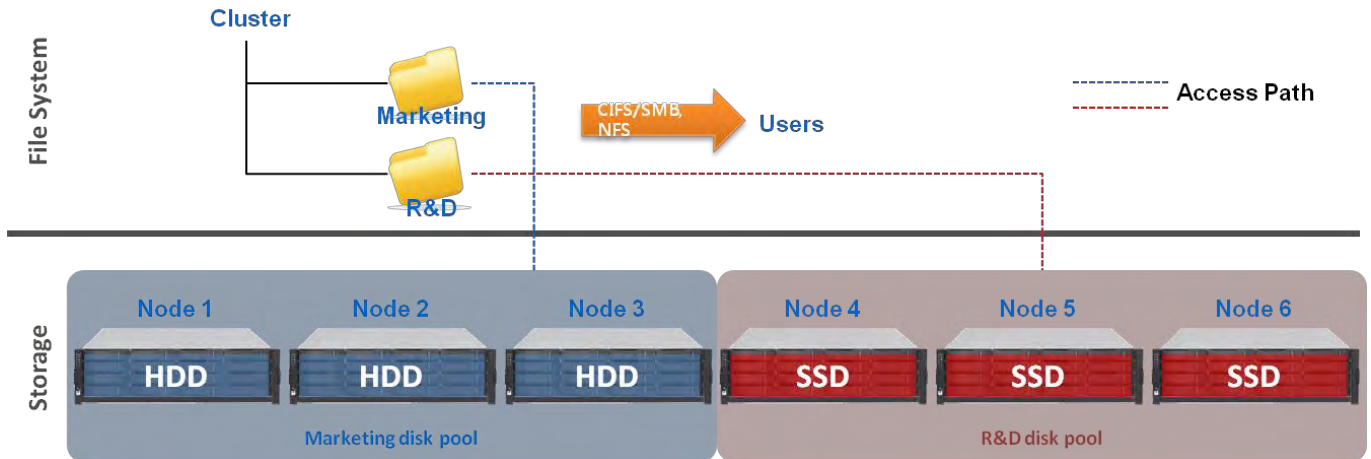
Scenario 3 – Expanding Capacity Only

Demands for the Marketing department has increased, the company expects to add more storage capacity to the marketing department, therefore, 3 expansion enclosure devices, all with NL-SAS hard drives, are purchased according to the capacity requirements.

(1) The administrator connects the expansion enclosures to the back-end network with physical cables to Node 1, Node 2, and Node 3.

(2) Login to the PAC Management interface to expand "Marketing disk pool".
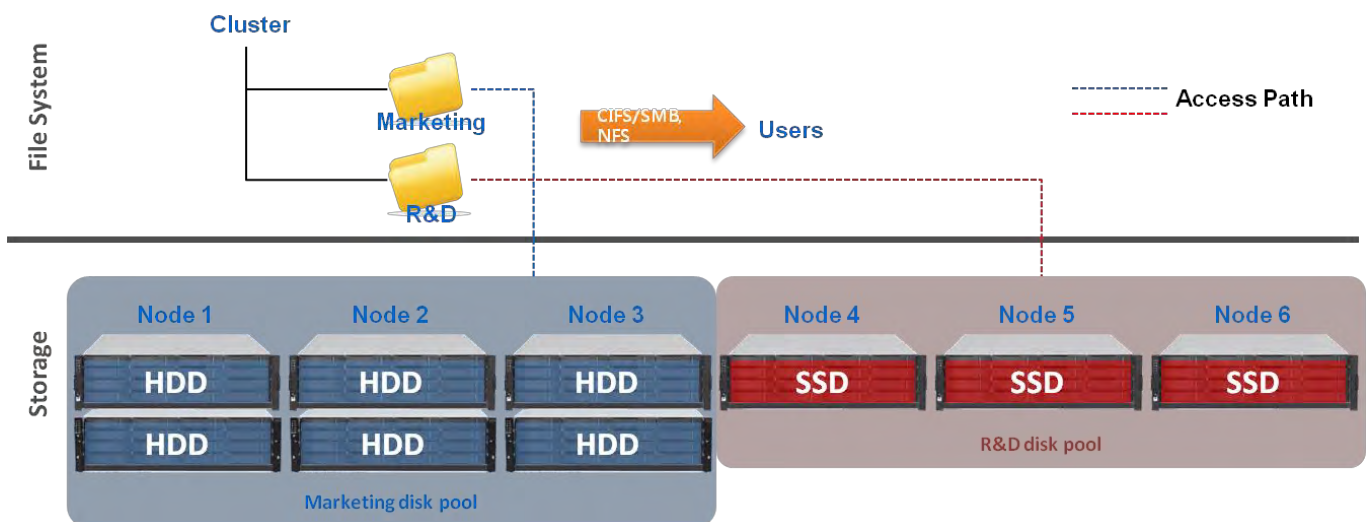
Conclusion

PS Scale-Out NAS delivers high performance and capacity cluster file system, through PS Scale-Out NAS architecture, client can access all nodes data in a single namespace up to 144 nodes and more than 100PB capacity are supported in a single cluster.

With the support of RAID, Erasure Code/Replica, and Remote Replication, PS Scale-Out NAS can provide comprehensive data protection. Through the PAC Management interface, user can manage the whole cluster in a centralized web interface, with a single point of administration.

PS Scale-Out NAS is ideally suited for unstructured data applications in the enterprise – including High Performance Computing (HPC), Media & Entertainment (M&E), large-scale file storage, backup and archive.

Contact Information

Website

For more information on PAC Storage products and services, visit:

https://www.pacstorage.com

Customer Support

Contact your system vendor or visit the following support site.

http://www.pacstorage.com/pac-storage-product-support.html