Hewlett Packard Enterprise

Technical white paper

HPE 3PAR ARCHITECTURE

CONTENTS

INTELLIGENT, TIER-1 STORAGE FOR A HYBRID CLOUD WORLD

HPE 3PAR is AI-driven storage for proven Tier-1 performance and resiliency

Powered by AI, HPE 3PAR Storage offers a Tier-1 all flash foundation with the flexibility to move data seamlessly between on-premises and public cloud. Built to meet the extreme requirements of massively consolidated cloud service providers, HPE 3PAR Storage is ideal for mission-critical workloads that require the highest levels of resiliency. Start fast and stay fast at the speed of memory with intelligent storage that anticipates and prevents issues across the infrastructure stack.

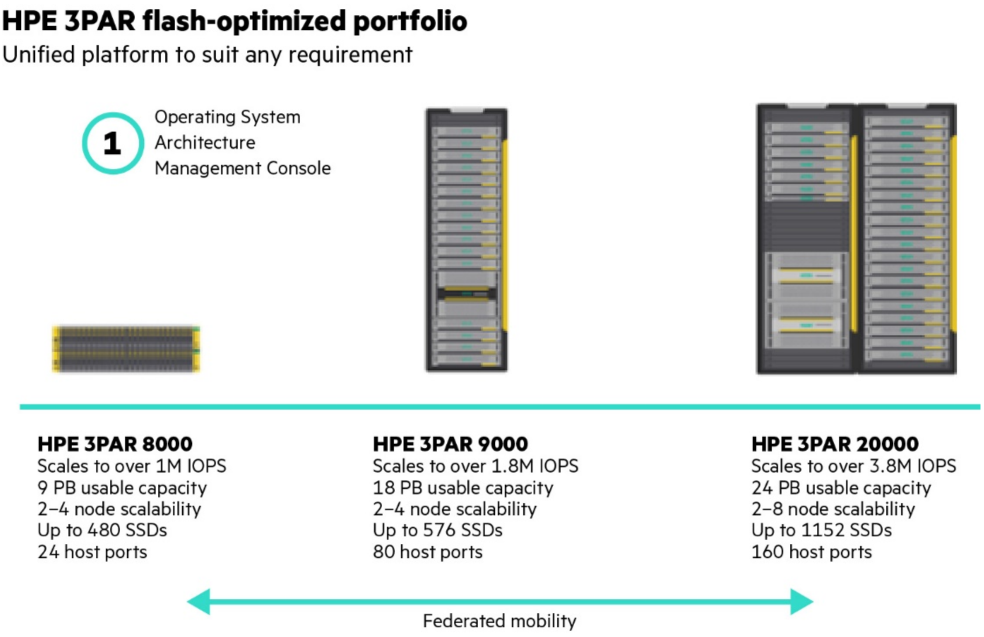


FIGURE 1. HPE 3PAR portfolio

All HPE 3PAR models are built on a single flash-optimized architecture, run the exact same HPE 3PAR Operating System, and offer a common set of enterprise data services. HPE 3PAR arrays can natively replicate and federate amongst each other without the need for any external replication or virtualization appliance. All models also offer features such as support for unification of block and file protocols and the use of spinning media to further optimize costs.

In addition, the HPE 3PAR 20000 Storage system features an eight-node-capable backplane that supports two to eight-controller nodes. This enterprise class of flash storage offers the flexibility to start with a few TBs and grow to over 20 PBs usable for massive consolidation. The HPE 3PAR 8000 and 9000 Storage systems feature a quad-node-capable architecture, which eliminates the trade-offs associated with most midrange storage solutions. Dual controller architectures suffer from “cache write-through mode” with the loss of a controller. In this scenario, writes are not signaled complete to the host until the data is on the back-end disk, which significantly impacts performance. HPE 3PAR systems with 4 nodes or more maintain high and predictable service levels even in the event of a cache or controller node failure by avoiding cache write-through mode.

This white paper describes the architectural elements of the HPE 3PAR Storage family.

HPE 3PAR HARDWARE ARCHITECTURE OVERVIEW

Each HPE 3PAR Storage system features a high-speed, full-mesh passive interconnect that joins multiple controller nodes (the high-performance data movement engines of the HPE 3PAR Architecture) to form a Mesh-Active cluster. This low-latency interconnect allows for tight coordination among the controller nodes to create a system-wide cache which is global, coherent and fault tolerant.

In every HPE 3PAR Storage system, each controller node has a dedicated link to each of the other nodes that operates at 4 GiB/s in each direction. In an HPE 3PAR 20800 Storage system, a total of 56 of these links form the array’s full-mesh backplane. In addition, each controller node may have one or more paths to hosts—either directly or over a storage area network (SAN). The clustering of controller nodes enables the system to present hosts with a single, highly available, high-performance storage system. This means that servers can access volumes over any host-connected port—even if the physical storage for the data is connected to a different controller node. This is achieved through an extremely low-latency data transfer across the high-speed, full-mesh backplane.

The HPE 3PAR Architecture can be scaled from 1.2 TiB to 9.6 PB of raw capacity, making the system deployable as a small, remote office system, or a very large centralized system. Until now, enterprise customers were often required to purchase and manage at least two distinct primary architectures to span their range of cost and scalability requirements, well that changes now. HPE 3PAR Storage is the ideal platform for business-critical applications, for virtualization and for cloud computing environments. The high performance and scalability of the HPE 3PAR Architecture is well suited for large or high-growth projects, consolidation of mission-critical information, demanding performance-based applications, and data lifecycle management. High availability is also built into the HPE 3PAR Architecture through full hardware redundancy. Controller node pairs are connected to dual-ported drive enclosures. Unlike other approaches, the system offers both hardware and software fault tolerance by running a separate instance of the HPE 3PAR Operating System on each controller node, thus facilitating the availability of customer data. With this design, software and firmware failures—a significant cause of unplanned downtime in other architectures—are greatly reduced.

HPE 3PAR Gen5 ASIC

The HPE 3PAR 20000, 9000, and 8000 systems use the fifth and latest generation of the HPE 3PAR ASIC. The HPE 3PAR Gen5 ASIC is engineered and designed for solid-state performance. The ASIC enables the new 20000, 9000, and 8000 series to deliver up to 5X improvement in system bandwidth and faster XOR operations compared to previous generations. It works in parallel with the CPU, evenly processing the I/O workload across the node Active—Mesh scale—out architecture, ensuring lower latency resulting in better system bandwidth.

The HPE 3PAR Gen5 ASICs also feature a uniquely efficient, silicon-based zero-detection and deduplication mechanism that gives HPE 3PAR Storage systems the power to perform inline deduplication and remove allocated but unused space with minimal impact to performance. The HPE 3PAR Gen5 ASICs also deliver mixed-workload support to alleviate performance concerns and cut traditional array costs.

Transaction- and throughput-intensive workloads run on the same storage resources without contention, thereby cutting array purchases in half. This is particularly valuable in virtual server environments, where HPE 3PAR Storage boosts virtual machine density so you can cut physical server purchases.

The Gen5 ASIC also enables Persistence Checksum that delivers T10-PI (Protection Information) for end-to-end data protection (against media and transmission errors) with no impact to applications or host operating systems.

Full-mesh controller backplane

Backplane interconnects within servers have evolved dramatically over the years. Most, if not all, server and storage array architectures have traditionally employed simple bus-based backplanes for high-speed processor, memory, and I/O communication. Parallel to the growth of SMP-based servers, significant investments were also made to switch architectures, which have been applied to one or two enterprise storage arrays.

The move from buses to switches was intended to address latency issues across the growing number of devices on the backplane (more processors, larger memory, and I/O systems). Third-generation full-mesh interconnects first appeared in the late 1990s in enterprise servers. Figure 2 shows the full-mesh HPE 3PAR Architecture.

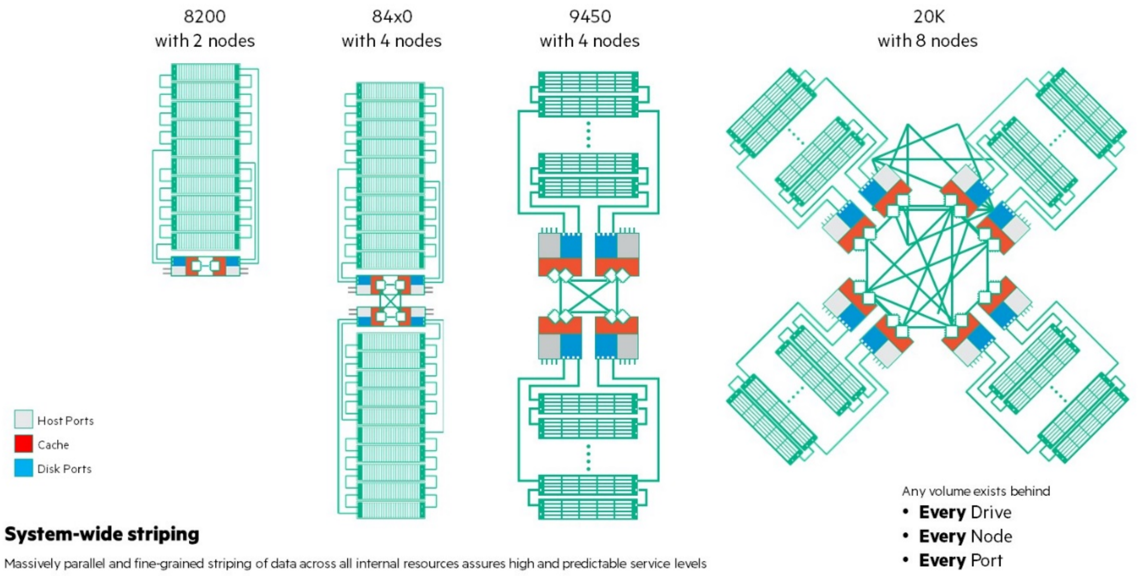


FIGURE 2. Full-mesh Active-Active Cluster Architecture

The HPE 3PAR full-mesh backplane is a passive circuit board that contains slots for up to four or eight controller nodes, depending on the model. As noted earlier, each controller node slot is connected to every other controller node slot by a high-speed link (4 GiB/s in each direction, or 8 GiB/s total), forming a full-mesh interconnect between all controller nodes in the cluster—something that Hewlett Packard Enterprise refers to as a Mesh-Active design. These interconnects deliver low-latency, high-bandwidth communication and data movement between controller nodes through dedicated point-to-point links and a low overhead protocol that features rapid inter-node messaging and acknowledgment. It's important to note that, while the value of these interconnects is high, the cost of providing them is relatively low. In addition, a completely separate full-mesh network of serial links provides a redundant low-speed channel of communication for exchanging control information between the nodes.

The HPE 3PAR 20000 features an eight node capable backplane that supports two to eight controller nodes. HPE 3PAR 9000 and 8000 Storage systems feature either a dual-node or quad-node-capable systems that is essentially an equivalent of what was used in erstwhile enterprise class arrays which offer the same high-speed links between nodes.

Active/Active vs. Mesh-Active

Most traditional array architectures fall into one of two categories: monolithic or modular. In a monolithic architecture, being able to start with smaller, more affordable configurations (i.e., scaling down) presents challenges. Active processing elements not only have to be implemented redundantly, but they are also segmented and dedicated to distinct functions such as host management, caching, and RAID/drive management. For example, the smallest monolithic system may have a minimum of six processing elements (one for each of three functions, which are then doubled for redundancy of each function). In this design—with its emphasis on optimized internal interconnectivity—users gain the Active/Active processing advantages of a central global cache (e.g., LUNs can be coherently exported from multiple ports). However, these architectures typically involve higher costs relative to modular architectures.

In traditional modular architectures, users are able to start with smaller and more cost-efficient configurations. The number of processing elements is reduced to just two, because each element is multifunction in design—handling host, cache, and drive management processes. The trade-off for this cost-effectiveness is the cost or complexity of scalability. Because only two nodes are supported in most designs, scale can only be realized by replacing nodes with more powerful node versions or by purchasing and managing more arrays. Another trade-off is that dual-node modular architectures, while providing failover capabilities, typically do not offer truly Active/Active implementations where individual LUNs can be simultaneously and coherently processed by both controllers.

The HPE 3PAR Architecture was designed to provide cost-effective single-system scalability through a cache-coherent, multinode clustered implementation. This architecture begins with a multifunction node design and, like a modular array, requires just two initial controller nodes for redundancy. However, unlike traditional modular arrays, enhanced direct interconnects are provided between the controllers to facilitate Mesh-Active processing. Unlike legacy Active/Active controller architectures—where each LUN (or volume) is active on only a single controller—this Mesh-Active design allows each LUN to be active on every controller in the system, thus forming a mesh. This design delivers robust, load-balanced performance and greater headroom for cost-effective scalability, overcoming the trade-offs typically associated with modular and monolithic storage arrays.

System-wide striping

Through a Mesh-Active design and system-wide striping, the HPE 3PAR Architecture can provide the best of traditional modular and monolithic architectures in addition to massive load balancing.

The HPE 3PAR Mesh-Active design not only allows all volumes to be active on all controllers, but also promotes system-wide striping that autonomically provisions and seamlessly stripes volumes across all system resources to deliver high, predictable levels of performance. System-wide striping of data provides high and predictable levels of service for all workload types through the massively parallel and fine-grained striping of data across all internal resources (disks, ports, loops, cache, processors, etc.). As a result, as the use of the system grows—or in the event of a component failure—service conditions remain high and predictable. Unlike application-centric approaches to storage, HPE 3PAR Storage provides autonomic rebalancing that enables the system to evenly balance and use all available physical resources. This is particularly important with hardware upgrades since existing data should be rebalanced and stripped across new available resources. On HPE 3PAR Storage, this is done without service disruption or preplanning.

For flash-based media, fine-grained virtualization combined with system-wide striping drives uniform I/O patterns by spreading wear evenly across the entire system. Should there be a media failure, system-wide sparing also helps guard against performance degradation by enabling a many-to-many rebuild, resulting in faster rebuilds. Because HPE 3PAR Storage autonomically manages this system-wide load balancing, no extra time or complexity is required to create or maintain a more efficiently configured system.

A detailed discussion of resource allocation, including the system's virtualized tri-layer mapping methodology, is provided in the section "Highly virtualized storage operating system."

Controller node architecture

An important element of the HPE 3PAR Architecture is the controller node and it is a powerful data movement engine that is designed for mixed workloads. As noted earlier, a single system, depending on the model, is modularly configured as a cluster of two to eight controller nodes. This modular approach provides flexibility, a cost-effective entry footprint, and affordable upgrade paths for increasing performance, capacity, connectivity, and availability as needs change. In addition, the minimum dual-controller configuration means that the system can withstand an entire controller node failure without impacting data availability. Controller nodes can be added in pairs to the cluster non-disruptively, and each node is completely hot-pluggable to enable online serviceability.

Unlike legacy architectures that process I/O commands and move data using the same processor complex, the HPE 3PAR Storage controller node architecture separates the processing of control commands from data movement, which helps ensure that CPU bandwidth is available for control processing and is not used for bulk data transfer. This innovation eliminates the performance bottlenecks of existing platforms that use a single processing element to serve competing workloads, for example, online transaction processing (OLTP) and data warehousing workloads.

The HPE 3PAR ASIC within each controller node performs parity calculations on the data cache. The Zero-Detect mechanism built into the ASIC allows a hardware-assisted fat-to-thin volume conversion in conjunction with HPE 3PAR Thin Conversion software that enables users to take “fat” provisioned volumes on legacy storage and convert them to “thin” provisioned volumes on the HPE 3PAR Storage system, this takes place inline and non-disruptively during the migration. This Zero-Detect capability also removes streams of zeroes present in I/O prior to writing data to the back-end storage system in order to reduce capacity requirements and prolong SSD life span. The HPE 3PAR ASIC is also a crucial element of the system’s ability to perform inline, block-level Thin Deduplication with Express Indexing (see the “Deduplication with Express Indexing” section for more details).

HPE 3PAR 20800 offers scalability in its truest sense not only with support for 2304 drives and 9.6 PB raw capacity but also with a port count of up to 160 FC ports for host connectivity and for consuming other advanced data services like replication, host persona, federation or to be used for Storage Class Memory. Each of these ports is connected directly on the I/O bus, so all ports can achieve full bandwidth up to the limit of the I/O bus bandwidths that they share.

Each HPE 3PAR 9000 and 20000 series controller node can have a maximum of the following ports per node:

- (10) Dual-port 32 Gbps Fibre Channel adapter
- (20) Quad-port 16 Gbps Fibre Channel adapter
- (10) Dual-port 10 Gbps iSCSI or Fibre Channel over Ethernet (FCoE) converged network adapter
- (6) 10 Gbps Ethernet adapter for file and object access services using File Persona

Additionally, the all-flash HPE 3PAR 9000 and 20000 series controllers support NVMe Storage Class Memory Modules in PCI slot 3 as a way to enhance the performance in specific small block workloads. One card per node can be installed.

With all the host ports available on an 8-controller configuration, HPE 3PAR 20000 series systems offer abundant multiprotocol connectivity. For back-end connectivity, each node can have up to (3) Quad-port 12 Gbps SAS Adapters.

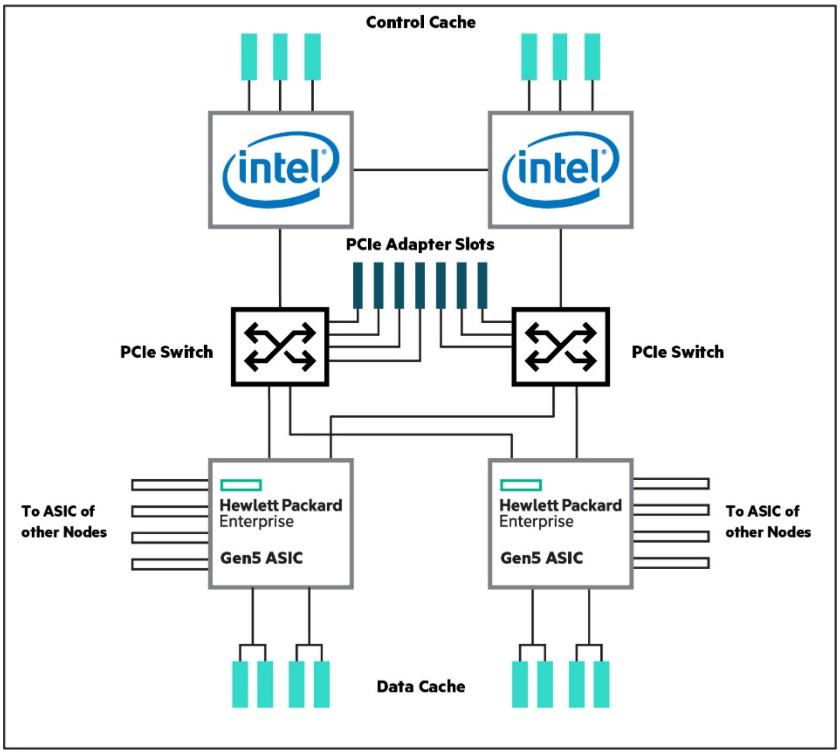


FIGURE 3. Controller node design of the HPE 3PAR 9450 and 20000 Storage system

Each HPE 3PAR 8000-series controller node has two built-in 4 port 16 Gbps Fibre Channel ports and one PCIe expansion slot. This slot can hold one of the following adapters:

- Dual-port 32 Gbps Fibre Channel adapter
- Quad-port 16 Gbps Fibre Channel adapter
- Dual-port 10 Gbps iSCSI or Fibre Channel over Ethernet (FCoE) converged network adapter
- Quad-port 16 Gb Fibre Channel/10 Gb NIC Combo Adapter
- Quad-port 10 Gb iSCSI/10 Gb NIC Combo Adapter
- Dual-port 10 Gbps Ethernet adapter for file and object access services using HPE 3PAR File Persona
- Quad-port 1 Gbps Ethernet adapter for file and object access services using HPE 3PAR File Persona

With up to 24 ports available on a quad-controller configuration, HPE 3PAR 8000-series systems offer abundant multiprotocol connectivity. For back-end connectivity, each node has two built-in 2 x 4-lane 12 Gbps SAS ports.

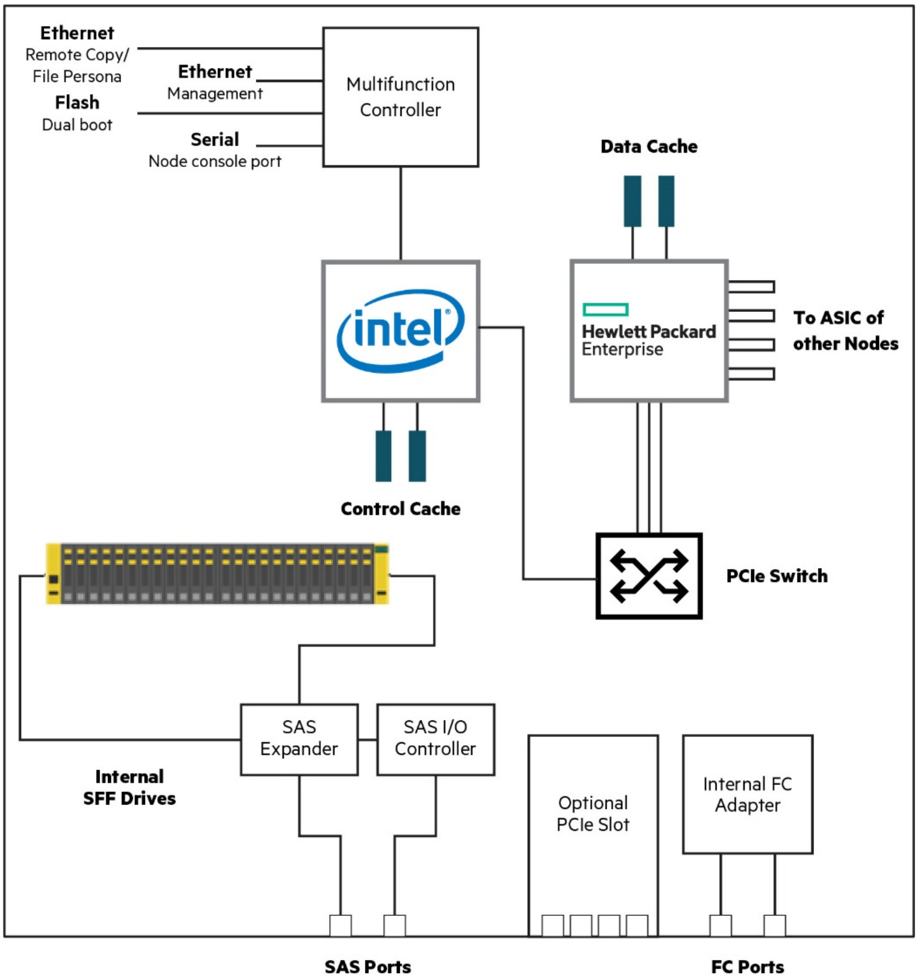


FIGURE 4. HPE 3PAR 8000 Controller Node

Across all 8000-series models, this controller node design extensively leverages commodity parts with industry-standard interfaces to achieve low costs and keep pace with industry advances and innovations. At the same time, the HPE 3PAR ASICs add crucial bandwidth and direct communication pathways without limiting the ability to use industry-standard parts for other components. Processor specifications by HPE 3PAR model are shown in Table 1.

TABLE 1. Processor specifications by HPE 3PAR Storage system model

Model	CPUs	Controller nodes	Total on-node cache	Total cache including optional flash cache
8200	Intel® 6 core 2.2 GHz/per controller node	2	Up to 64 GiB	832 GiB
8400	Intel 6 core 2.2 GHz/per controller node	2 or 4	Up to 128 GiB	1664 GiB
8440	Intel 10 core 2.4 GHz/per controller node	2 or 4	Up to 384 GiB	8384 GiB
8450	Intel 10 core 2.4 GHz/per controller node	2 or 4	Up to 384 GiB	Not applicable
9450	2 x Intel 10 core 2.4 GHz/ per controller node	2 or 4	Up to 896 GiB	Not applicable
20450	2 x Intel 8 core 2.5 GHz/per controller node	2 or 4	Up to 1.8 TiB	Not applicable
20800	2 x Intel 8 core 2.5 GHz/per controller node	2, 4, 6, or 8	Up to 2.6 TiB	34.5 TiB
20840	2 x Intel 10 core 2.4 GHz/per controller node	2, 4, 6, or 8	Up to 3.6 TiB	51.6 TiB
20850	2 x Intel 10 core 2.4 GHz/per controller node	2, 4, 6, or 8	Up to 3.6 TiB	3.6 TiB

Drive enclosures

Another key element of the HPE 3PAR Storage system is the drive enclosure or drive chassis, which serves as the capacity building block within the system. This section looks in more detail at the different drive enclosures supported by HPE 3PAR Storage systems.

TABLE 2. Capacity building block for HPE 3PAR 20000 systems

	HPE 3PAR 20800	HPE 3PAR 20840	HPE 3PAR 20850	HPE 3PAR 20450
Controller Nodes	2, 4, 6, 8	2, 4, 6, 8	2, 4, 6, 8	2, 4
Max capacity (raw)	9600 TiB	9600 TiB	8043 TiB	4021 TiB
Max usable file capacity	1024 TiB	1024 TiB	512 TiB	256 TiB
Hard drives	8–2304	2304	Not applicable	Not applicable
SSDs	6–1152	6–1152	6–1152	6–576

TABLE 3. Capacity building block for HPE 3PAR 8000 systems

	HPE 3PAR 8450	HPE 3PAR 8440	HPE 3PAR 8400	HPE 3PAR 8200
Controller Nodes	2, 4	2, 4	2, 4	2
Max capacity (raw)	3351 TiB	4000 TiB	2400 TiB	1000 TiB
Max usable file capacity	2–512 TiB	2–512 TiB	2–512 TiB	2–256 TiB
Hard drives	Not applicable	6–960	6–576	6–240
SSDs	6–480	6–480	6–240	6–120

TABLE 4. Capacity building block for HPE 3PAR 9450 systems

	HPE 3PAR 9450
Controller Nodes	2, 4
Max capacity (raw)	6000 TiB
Max usable file capacity	512 TiB
Hard drives	Not applicable
SSDs	6–576

NOTE
For further details, please refer to SPOCK.

HIGHLY VIRTUALIZED STORAGE OPERATING SYSTEM

HPE 3PAR Storage uses the same highly virtualized storage operating system across all models—including high-end, midrange, hybrid, and all-flash arrays. To help ensure performance and improve the utilization of physical resources, the HPE 3PAR Operating System employs a tri-level mapping with three layers of abstraction (including HDD, chunklets, LDs).

Fine-grained approach to virtualization

The tri-level abstraction methodology imposed by the HPE 3PAR Operating System relies on a fine-grained virtualization approach that divides each physical disk into granular allocation units referred to as chunklets, each of which can be independently assigned and dynamically reassigned to different logical disks that are used to create virtual volumes.

Multiple layers of abstraction

As shown in Figure 5, the physical disk abstraction layer breaks physical drives of any size into a pool of uniform-sized, 1 GiB chunklets. The fine-grained nature of these chunklets eliminates underutilization of precious storage assets.

Complete access to every chunklet eliminates large pockets of inaccessible storage. This fine-grained structure enhances performance for all applications as well, regardless of their capacity requirements. For example, while a small application might only allocate a small amount of physical capacity, this capacity will be virtualized and striped across dozens or even hundreds of drives. With this approach, even a small application can leverage the performance resources of the entire system without provisioning excess capacity.

The first layer of abstraction employed by the OS breaks media devices into 1 GiB chunklets to enable higher utilization and avoid stranded capacity. This fine-grained virtualization unit also enables mixed RAID levels on the same physical drive, thereby eliminating dedicated RAID groups and seamlessly supporting new media technologies such as SSDs.

The second layer of abstraction takes the 1 GiB chunklets created from abstracting physical disk capacity and creates logical disks (LDs) striped across the system's physical drives and implementing specified RAID levels. Multiple chunklet RAID sets from different PDs are striped together to form an LD. All chunklets belonging to a given LD will be from the same drive type. LDs can consist of all NL, FC, or SSD chunklets. There are no mixed-type LDs, although Fast Class (Fibre Channel or SAS) LDs, may consist of both 10K and 15K drive chunklets. The association between chunklets and LDs allows LDs to be created with template properties based on RAID characteristics and the location of chunklets across the system. LDs can be tailored to meet a variety of cost, capacity, performance, and availability characteristics. In addition, the first- and second-level mappings taken together serve to parallelize work massively across physical drives and their Fibre Channel or SAS connections. LDs are divided into "regions," 128 MB of contiguous logical space from a single LD.

The third layer of abstraction maps LDs to Virtual Volumes (VVs), with all or portions of multiple underlying LDs mapped to the VV. VVs are the virtual capacity representations that are ultimately exported to hosts and applications as virtual LUNs (VLUNs) over Fibre Channel, iSCSI, or FCoE target ports. A single VV can be coherently exported through as few as two ports or as many as ports as desired (no fewer than two, one from each of two different nodes as a minimum). This layer of abstraction uses a table-based association—a mapping table with a granularity of 128 MB per region and an exception table with a granularity of 16 KB per page—as opposed to an algorithmic association. With this approach, a very small portion of a VV associated with a particular LD can be quickly and non-disruptively migrated to a different LD for performance or other policy-based reasons, whereas other architectures require migration of the entire VV. This layer of abstraction also implements many high-level features such as snapshots, caching, pre-fetching, and remote replication.

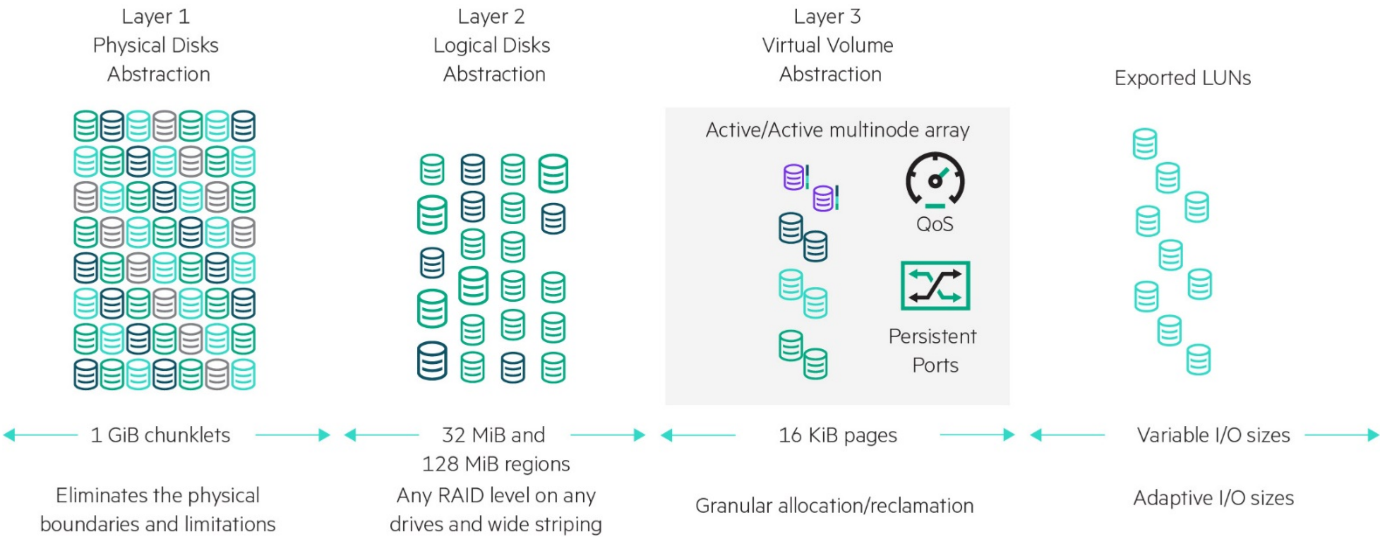


FIGURE 5. Virtualization with a tri-level mapping methodology that provides three layers of abstraction

One-stop allocation, the general method employed by IT users for volume administration, requires minimal planning on the part of storage administrators. By an administrator simply specifying virtual volume name, RAID level, and size, the HPE 3PAR Operating System autonomically provisions VVs at the moment that an application requires capacity. This process is also known as "just-in-time" provisioning. Contrast this to traditional architectures where the storage administrator must assign physical disks to RAID sets when the array is installed, which can be difficult or impossible to change later on and makes it difficult to respond to changing requirements.

The three-layer abstraction implemented by the HPE 3PAR Operating System can effectively utilize any underlying media type. This means that HPE 3PAR Storage is able to make the most efficient use of SSDs through massive load balancing across all drives to enable ultra-high performance and prolong flash-based media life span.

Logical disks

There are three types of logical disks (LDs):

- User (USR) LDs provide user storage space to fully provisioned VVs.
- Snapshot data (SD) LDs provide the storage space for snapshots (or virtual copies), thinly provisioned (TPVV) and thinly deduplicated (TDVV) virtual volumes.
- Snapshot administration (SA) LDs provide the storage space for metadata used for snapshot and TPVV and TDVV administration.

As mentioned earlier, RAID functionality is implemented at the LD level, with each LD mapped to chunklets in order to implement RAID 1+0 (mirroring + striping), RAID 5+0 (RAID 5 distributed parity + striping), or RAID MP (multiple distributed parity, with striping).

The HPE 3PAR Operating System will automatically create LDs with the desired availability and size characteristics. In addition, several parameters can be used to control the layout of an LD to achieve these different characteristics:

- **Set size:** The set size of the LD is the number of drives that contain redundant data. For example, a RAID 5 LD may have a set size of 4 (3 data + 1 parity), or a RAID MP LD may have a set size of 16 (14 data + 2 parity). For a RAID 1 LD, the set size is the number of mirrors (usually 2). The chunklets used within a set are typically chosen from drives on different enclosures. This helps ensure that a failure of an entire loop (or enclosure) will not result in data becoming unavailable until the drive enclosure is repaired. It also helps ensure better peak aggregate performance because data can be accessed in parallel on different loops.
- **Step size:** The step size is the number of bytes that are stored contiguously on a single physical drive.
- **Row size:** The row size determines the level of additional striping across more drives. For example, a RAID 5 LD with a row size of 2 and set size of 4 is effectively striped across 8 drives.
- **Number of rows:** The number of rows determines the overall size of the LD given a level of striping. For example, an LD with 3 rows, with each row having 6 chunklets' worth of usable data (+2 parity), will have a usable size of 18 GiB (1 GiB/chunklet x 6 chunklets/row x 3 rows).

NOTE
Hewlett Packard Enterprise recommends that the storage administrator allow the array to decide the best combination of step size, row size, and row number.

LD Ownership

Every LD has an "owner" and a "backup" owner. Using the traditional layout, chunklets from any given PD are owned by a single node with the partner node as the backup owner; thus every node creates LDs from the PDs it "owns." Express Layout alters ownership for flash drives. If the set size configured requires more than 50 percent of the drives behind the node pair then the LD will be created using chunklets from PDs behind the node pair, allowing each node to create LDs larger than traditionally possible. This allows smaller flash systems to create larger set sizes, reducing RAID overheads and improving capacity efficiency.

Common provisioning groups

A common provisioning group (CPG) creates a virtual pool of LDs that allows VVs to share the CPG's resources and allocate space on demand. You can create fully provisioned VVs, thinly provisioned VVs and thinly deduplicated VVs that draw space from the CPG's logical disk pool.

CPGs enable fine-grained, shared access to pooled logical capacity. Instead of pre-dedicating logical disks to volumes, a CPG allows multiple volumes to share the buffer pool of LDs. For example, when a TPVV is running low on user space, the system automatically assigns more capacity to the TPVV by mapping new regions from LDs in the CPG to the TPVV. As a result, any large pockets of unused but allocated space are eliminated. Fully provisioned VVs cannot create user space automatically, and the system allocates a fixed amount of user space for the volume at the time it is created.

Virtual volumes

There are two kinds of VVs: "base volumes" and "snapshot volumes." A base volume can be considered to be the "original" VV and is either a fully provisioned virtual volume, a thinly provisioned virtual volume or a thinly provisioned deduplicated virtual volume. In other words, it directly maps all the user-visible data. A snapshot volume is created using HPE 3PAR Virtual Copy software. When a snapshot is first created, all of its data is mapped indirectly to the parent volume's data. When a block is written to the parent, the original block is copied from the parent to the snapshot data space and the snapshot points to this data space instead. Similarly, when a block is written in the snapshot, the data is written in the snapshot data space and the snapshot points to this data space.

VVs have three types of space:

- The **user space** represents the user-visible size of the VV (i.e., the size of the SCSI LUN seen by a host) and contains the data of the base VV.
- The **snapshot data space** is used to store modified data associated with snapshots. The granularity of snapshot data mapping is 16 KB pages.
- The **snapshot admin space** is used to save the metadata (including the exception table) for snapshots.

Each of the three space types is mapped to LDs, with all of these LDs striped across all controller nodes; thus, VVs can be striped across multiple nodes for additional load balancing and performance.

The size limit for an individual virtual volume is 16 TiB. A VV is classified by its provisioning type, which can be one of the following types:

- **Fully provisioned VV (FPVV):** Has either no snapshot space or deprecated, statically allocated snapshot space.
- **Thinly provisioned VV (TPVV):** TPVV has space for the base volume allocated from the associated CPG and snapshot space allocated from the associated snapshot CPG (if any).

On creation, 256 MB per node is allocated to a TPVV. Storage is allocated on demand in the snapshot data area as required by the host operation being performed. The snapshot admin area contains the metadata indexes that point to the user data in the SD area. Because the SA metadata needs to be accessed to locate the user data, the indexes are cached in policy memory to reduce the performance impact of the lookups.

TPVVs associated with a common CPG share the same LDs and draw space from that pool as needed, allocating space on demand in small increments for each controller node. As the volumes that draw space from the CPG require additional storage, the HPE 3PAR Operating System automatically extends existing LDs or creates new LDs until the CPG reaches the user-defined growth limit, which restricts the CPG's maximum size.

- **Thinly deduped VV (TDVV):** TDVVs behave similarly to TPVV volumes with the fundamental difference that TDVVs within the same CPG will share common pages of data. The data shared is determined via the inline deduplication mechanism described later in this paper. TDVVs are supported only on CPGs that use SSDs as a tier of storage.
- **Commonly provisioned VV (CPVV):** The space for this VV is fully provisioned from the associated CPG, and the snapshot space is allocated from the associated snapshot CPG.

VLUNs and LUN masking

VVs are only visible to a host once the VVs are exported as VLUNs. VVs can be exported in three ways:

- To specific hosts (set of Worldwide Names or WWNs)—the VV is visible to the specified WWNs, regardless of which port(s) those WWNs appear on. This is a convenient way to export VVs to known hosts.
- To any host on a specific port—this is useful when the hosts (or the WWNs) are not known prior to exporting, or in situations where the WWN of a host cannot be trusted (host WWNs can be spoofed).
- To specific hosts on a specific port.

Metadata handling

For metadata, HPE 3PAR implements a mechanism of fast lookup tables that store location pointers to accelerate data access. This process relies on three layer addresses translation mechanism akin to Virtual Memory lookup tables. The VV metadata also known as Snapshot Admin space, contains bitmaps indicating which pages of the shared snapshot data areas are used along with the exception tables associated with each volume.

The exception tables provide the offset into the shared data area where the shared data is located. There are three levels (L1, L2, and L3) of exception tables for each Thin Volume or snapshot. The bits of the I/O request's logical block address (LBA) are used as indexes into the L1, L2, and L3 exception tables to determine the 16 KiB page being accessed.

Metadata management is shared across all cluster resources from CPU memory (control cache).

When HPE 3PAR Adaptive Flash Cache is configured, all Snapshot Admin space associated with any TPVV (or snapshot) on the array will always be also cached to Flash Cache.

For inline deduplication, where fast lookups are extremely important, HPE 3PAR implements an innovative mechanism to detect duplicate pages called Express Indexing. The technology uses the computed hash signature as an index to determine whether a match already exists using the three level translation mechanisms described earlier.

When new a write I/O request comes in, the Logical Block Address (LBA) is used as an index into three different page tables as per a regular TPVV. However, instead of allocating a new page, the hash signature of the incoming data page is computed by the HPE 3PAR ASIC and compared to the signatures of the TDVV data already stored in the CPG.

If a match is found then the L3 page table entry will be set to point to the existing copy of the data page. Only if no match is found, a new page is allocated.

System-wide sparing

The HPE 3PAR Operating System has a logical volume manager that handles volume abstraction at the VV layer and also handles sparing. This logical volume manager reserves a certain number of chunklets as spare chunklets depending on the sparing algorithm and system configuration. Unlike many competitive arrays that reserve dedicated spare drives that then sit idle, system-wide sparing with HPE 3PAR Storage means that spare chunklets are distributed across all drives. This provides additional protection and enables a balanced load that extends the SSD life span by providing even wearing. It also protects against performance degradation by enabling a “many-to-many” rebuild in the event of a failure.

Flash-optimized innovations

HPE 3PAR Mesh-Active Architecture, system-wide striping, fine grain virtualization, advanced metadata handling, and system-wide sparing are just some of the pillars of HPE 3PAR architecture that enhance flash-based media. Flash-based media can deliver many times the performance of conventional spinning HDDs and it can do so at very low, sub-millisecond latency. However, it is important to understand that these advantages can only be realized by an architecture that has optimized its entire I/O path to be performance-centric. If the storage controllers that sit between servers and back-end flash devices can’t keep up with the performance of the flash drives, they become performance bottlenecks.

To work with flash-based media in the most performance-optimized manner, the HPE 3PAR Architecture includes features designed to handle it in a substantially different way than spinning media. It also exploits every possible opportunity to extend flash-based media life span by reducing factors that contribute to media wear. On top of that, all SSDs currently available for HPE 3PAR Storage are enterprise-grade SAS SSDs, giving you a choice between the latest planar NAND and emerging 3D NAND technologies. HPE 3PAR Storage currently offers several NAND flash options that leverage the latest enterprise Multi-level cell (MLC) technology.

This flash-optimized architecture relies on several new and unique HPE 3PAR Storage innovations that accelerate performance and extend flash-based media life span:

Deduplication with Express Indexing: Deduplication is a technology designed to eliminate duplicate information from being committed to disk. The 3PAR ASIC features a dedicated, high performance, low latency hashing engine used for deduplication that can lead to not only massive savings over standard deployment methodologies but a much smaller performance overhead when deduplication is enabled. Deduplication employs Express Indexing, a mechanism that provides extremely high performance lookup tables for fast detection of duplicate write requests.

When a new write request enters cache, a hash (or fingerprint) of the data is generated in order to draw a match against other data stored on the array. Generating a hash of every data write is an extremely CPU-intensive task and many software-implemented hashing algorithms commonly found in all-flash platforms add a significant performance overhead to write performance. With HPE 3PAR deduplication software, the CPU-intensive jobs of calculating hash signatures for incoming data and verifying reads are offloaded to the ASICs, freeing up processor cycles to perform other critical data services.

Once the hash has been calculated, Express Indexing performs instant metadata lookups in order to compare the signatures of the incoming request to signatures of data already stored in the array. If a match is found, the system flags the duplicate request and prevents it from being written to the back end. Instead, a pointer is added to the metadata table to reference the existing data blocks. To prevent hash collision (when two write pages have the same signature but different underlying data), HPE 3PAR deduplication software leverages the controller node ASICs once again to perform a high-performance bit-to-bit comparison before any new write update is marked as a duplicate, preventing incorrect data matches. This is an important step to prevent data corruption and should be at the core of any deduplication implementation.

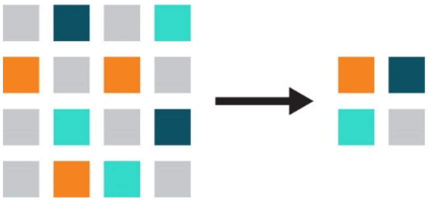


FIGURE 6. ASIC-based hash signature generation for inline deduplication

This hardware-assisted approach enables an inline deduplication implementation that carries multiple benefits, including increased capacity efficiency, flash performance protection, and flash media lifespan extension. The combination of hardware-assisted hashing and Express Indexing is so powerful, that HPE 3PAR offers performance comparable to—or exceeding—other storage systems with over eight times more CPU cores and memory resources. This allows HPE 3PAR to offer significantly reduced power and cooling costs.

Compression: While deduplication looks for opportunities to remove entire blocks of data by comparing them against each other, compression works by looking for opportunities to reduce the size of pages before they're written to flash. When compression and deduplication are enabled together, data is first deduplicated then compressed.

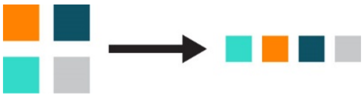


FIGURE 7. Compression reduces data footprint

Compression is unlike many other 3PAR-implemented technologies as for optimal performance, compression needs to be run across as many parallel threads as possible instead of fewer, extremely fast, dedicated engines. This results in CPU cores being optimal for compression rather than dedicated ASIC engines. However, this doesn't mean that the 3PAR ASIC doesn't have a part to play in compression: since virtually every other CPU-intensive operation is offloaded to the ASIC, CPU cores in 3PAR arrays are very lightly used meaning that CPU resources are available for use with compression without worrying about resource contention. This leads to high performance without requiring excessive system resources. Additionally, 3PAR implements Express Scan, a technology specific to 3PAR that further reduces the CPU overhead associated with compression. This is achieved by inspecting blocks to ensure data is compressible, rather than wasting CPU cycles attempting to compress data we can identify is incompressible.

HPE 3PAR compresses data using an extremely fast, modern compression algorithm that's very fast for compression tasks and extremely fast for decompression tasks while offering excellent compression savings. Read and write performance profiles are very important with compression; write performance needs to be high to support incoming write streams of data but since writes are cached in the system before they're committed to flash, compression is essentially an asynchronous task so doesn't impact write latency as heavily. However, reads are far more sensitive because not all reads are delivered from cache; whenever a read "miss" occurs (when a read is requested and it's not in cache) the array must read the back-end data, decompress it and return it to the host. Performance here is key as latency will increase as decompression performance reduces. The compression algorithm 3PAR implements offers extremely high decompression tasks, virtually eliminating any latency difference between thin provisioned volumes and compressed volumes.

Data Packing: Once data has been compressed, the result is a number of data blocks that are smaller in size but are also odd sizes (1.3 KiB, 4.2 KiB, 5.6 KiB, etc.). These pages are not only odd sizes, but they're very hard to write to flash since flash pages are fixed in size—writing these pages directly to flash will result in lost performance and reduced efficiency, neither of which are desirable. Data Packing addresses this issue by packing these odd-sized pages together into a single page before they're written to flash.

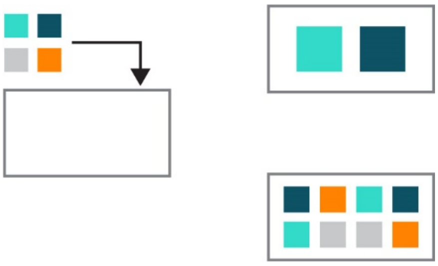


FIGURE 8. Multiple pages stored together through Data Packing

The result of Data Packing is that the system maintains fixed-size flash-native block sizes that increase both performance and system efficiency. While there are many different approaches to addressing this issue, Data Packing carries some innovative benefits. One of the largest of these is that 3PAR systems require no post-processing tasks to deal with excessive amounts of garbage collected—simply because Data Packing creates very little garbage in the first place. Overwrites are generally in-place, meaning we don't generate "stale" data and have to collect it later—the system simply updates the overwritten data where it's held and moves on to the next write. Many approaches append many writes together into a huge stripe before committing them to flash, but as data is overwritten this results in "holes" in the data which impacts the efficiency of the system—until housekeeping defragments these stripes, the system's usable space is reduced so these systems typically hide some free space from the administrator to cover up the lost capacity. HPE 3PAR systems will never hide capacity from the administrator as Data Packing prevents 3PAR systems from getting into this scenario in the first place.

The result of Data Packing is that 3PAR Adaptive Data Reduction technologies are 100% inline, requiring no post-process tasks, resulting in consistent, high levels of performance with consistent space usage.

Compaction: In addition to Adaptive Data Reduction, 3PAR systems offer additional capacity-efficiency technologies that include market-leading hardware-accelerated thin provisioning, Thin Clones, Thin Reclaim, Virtual Copy, and other technologies. Savings from these technologies are not included in Adaptive Data Reduction (or data reduction ratios) and are instead reported in the Compaction ratio—a ratio that exposes the full suite of efficiency technologies. Therefore, Compaction is the combination of Thin technologies and Adaptive Data Reduction.

- **Express Layout:** This unique technology born from the HPE 3PAR three-layer virtualization technology allows 3PAR controller nodes to share access to SSDs in order to further drive efficiency. Replacing traditional layouts for flash, Express Layout allows each SSD to be actively accessed by both controllers in a node pair at the same time. This allows a node pair to use capacity from every drive to build logical capacity. For smaller configurations, like an eight drive system, Express Layout allows the nodes to significantly reduce the overhead historically associated with parity RAID layouts and can result in more than 10% reduction in overhead which increasing performance by allowing more than one controller to drive I/O through the drive.
- **Adaptive Read:** When reading data into cache from spinning disks, HPE 3PAR systems in fixed-block sizes of 16 KB. This is useful in a spinning disk world because seek times are slow and once a read opportunity is available reading additional data into cache increases the chance of future cache hits, lowering the latency of future I/O. With flash media, random reads are extremely fast as there are no spinning mechanical parts so the penalty of a read miss is minimal and reading additional data actually imposes a penalty. Given this attribute of reading from flash, HPE 3PAR systems read only the data required into cache when reading from SSDs. As the size of the cached data is adapted to the originating host I/O request (as opposed to the system cache page size)—the resulting benefits are:

  Greatly optimized read latency—with I/O-intensive workloads, typically true with flash environments, for a 4 KB I/O, only a 4 KB block is read in cache, not 16 KB. This reduces the time required to complete a read operation.

  More efficient back-end throughput—a smaller cache page read also optimizes back-end throughput, enabling higher IOPS without hindering the back-end throughput handling. For example, a thousand 4 KB reads from flash result in 4 Mbps of back-end throughput instead of 16 Mbps of throughput, which would be the case if minimum cache page read granularity was 16 KB.
- **Express Layout:** This unique technology born from HPE 3PAR's three-layer virtualization technology allows HPE 3PAR controller nodes to share access to SSDs in order to drive efficiency. Replacing traditional layouts for flash, Express Layout allows each physical drive (PD) to be "owned" by both controllers at the same time allowing both nodes to use chunklets from every drive to build logical disks (LDs). This technology allows smaller systems to realize reduced capacity overhead.
- **Adaptive Writes:** As cache pages are 16 KB in HPE 3PAR systems, Adaptive Writes works by detecting when a sub-16 KB write hits cache, identifying the part of the cache page that holds data and committing only that part of the cache page to disk, discarding the whitespace stored in the rest of the cache page. The HPE 3PAR architecture achieves this by keeping a bitmap for each cache page and writes only the changed part of the cache page to flash media. This optimization results in fewer writes to flash media which not only increases the effective endurance of flash media also reduces back-end throughput as it is only writing changed data to the back end.
- **Autonomic Cache Offload:** Autonomic Cache Offload is a flash optimization that eliminates cache bottlenecks by changing the frequency at which data is offloaded from cache to flash media based on system utilization. This ensures consistently high performance levels as the system scales performance to hundreds of thousands and even millions of IOPS. New writes coming into the array are acknowledged to the host as soon as the I/O gets written to cache in two nodes for protection. The in-cache write then gets flushed to the storage media at a rate based on cache utilization. At higher levels of utilization, HPE 3PAR increases the frequency at which flushes take place which allows the system to deliver consistent performance without running into cache bottlenecks, even at extreme performance levels.

  Another important aspect of the cache offload algorithm is the decision around which cache data should be flushed to the back end. HPE 3PAR systems maintains a log of read cache hits and keeps more frequently accessed pages in cache, lowering latencies.

Another important aspect of the HPE 3PAR caching algorithm is how it handles large write requests. To enable cache-consistency, writes are acknowledged to the host as soon as they are written to the cache on two nodes. If the writes are large, the HPE 3PAR caching algorithm allocates cache pages (16 KB) as the pages become available instead of waiting for all the necessary pages to become available before starting to commit data. For example, a 128 KB write request requires eight cache pages but writes can be started even when just one cache page is available and is completed when the eighth cache page has been written to. This allows HPE 3PAR to improve write latency which is crucial when delivering millions of IOPS.

- **Multi-tenant I/O processing:** Multi-tenant I/O processing enables performance improvement for mixed workloads or virtual desktop infrastructure (VDI) deployments by breaking large I/O into smaller chunks so that small read requests don't get held up or stuck behind larger I/O requests, which also helps ensure reduced latency expected of flash-based media.
- **Adaptive Sparing:** Using Adaptive Sparing technology, Hewlett Packard Enterprise has collaborated with SSD suppliers to extend usable capacity per drive by up to 20 percent. This is achieved by reducing capacity typically reserved by media suppliers for wear management and then using that space more efficiently. At a system level, increasing usable drive capacity also helps spread writes more broadly to extend SSD endurance.

WORKLOAD-CENTRIC STORAGE PERSONAS

HPE 3PAR Storage is expressed by workload-centric Storage Personas. A persona is **the aspect of one's character that is presented to or perceived by others**. Storage Personas are thus comprised of data access protocols and data services for the presentation of storage to hosts and clients. Specifically, HPE 3PAR features a Block Persona and a File Persona that are engineered into the core of the HPE 3PAR OS and system architecture, and are managed seamlessly together via the HPE 3PAR Management Console and scriptable HPE 3PAR CLI. Through these Storage Personas, HPE 3PAR provides truly converged block, file, and object access to simultaneously support an expanse of workloads while allowing the best storage approach to be employed for a given workload.

Block Persona

HPE 3PAR Storage is expressed by the Block Persona in the form of block volumes to server hosts via Fibre Channel, iSCSI, and FCoE. The block data services are as mentioned throughout this white paper.

File Persona

File Persona can be enabled on a HPE 3PAR storage system node pair with an optional license. It requires either a 2 port 10GbE or a 4 port 1GbE NIC to be installed in the system or the on-board 1GbE RCIP port to be enabled for File Persona.

File Persona is designed for client workloads such as home directories and user shares; content management and collaboration; data preservation and governance; and custom cloud applications by presenting file shares via SMB (CIFS) and NFS as well as object shares via the Object Access API to client devices. File data services include: User Authentication Services; capacity and user/group Quota Management; File Store Snapshots with user-driven file restore; and Antivirus Scan Services for integration with third-party antivirus software.

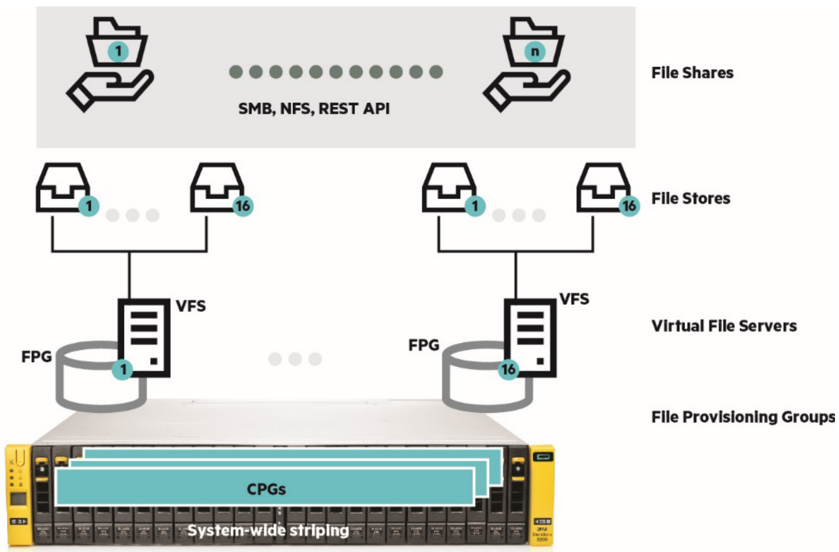


FIGURE 9. Logical view of HPE 3PAR File Persona managed objects

The HPE 3PAR File Persona has four types of managed objects. File Provisioning Groups, which are instances of the HPE intellectual property Adaptive File System, control how files are stored and retrieved. Each File Provisioning Group is transparently constructed from one or multiple Virtual Volumes. Virtual File Servers (which present virtual IP addresses), participate in User Authentication Services, and can have properties for such things as user/group Quota Management. File Stores are the slice of a Virtual File Server and File Provisioning Group at which File Store Snapshots are taken, capacity Quota Management can be performed, and Antivirus Scan Services policies customized. File Shares are presented to clients via SMB, NFS, and the Object Access API, and are where share permissions are set. File Stores and File Provisioning Groups are typically only explicitly managed for advanced operations.

MULTI-TENANT ARCHITECTURE BENEFITS

With HPE 3PAR Storage, you can securely partition resources within a shared infrastructure in order to pool physical storage resources for lower storage costs without compromising security or performance.

The HPE 3PAR Storage platform was built from the ground up to deliver multi-tenant capacity that supports massive consolidation with ultra-high performance. The multicontroller scalability and extreme flexibility built into HPE 3PAR Storage makes deploying and maintaining separate storage silos to deliver different QoS levels a thing of the past. Unlike application-centric approaches to storage, one-click autonomic rebalancing on HPE 3PAR Storage enables you to enhance QoS levels without service disruption, pre-planning, or the need to purchase separate arrays to support different service levels. To support multiple tenants and workloads, HPE 3PAR Storage provides secure administrative segregation of users, hosts, and application data. The following sections provide insight into the architectural elements that support each of these core capabilities.

Tier-1 resiliency

HPE 3PAR Storage is designed to support massive consolidation by supporting mixed workloads and secure administrative segregation of users, hosts, and application data. Multi-tenancy allows IT organizations to deliver higher performance levels, greater availability, and next-generation functionality to multiple user groups and applications from a single storage system.

Today’s IT realities—including virtualization, cloud computing, and ITaaS—demand the ability to deliver predictable service levels in an inherently unpredictable world, and make system resiliency the single most important requirement for multi-tenancy. Traditionally, Tier-1 storage has been characterized by hardware redundancy, advanced replication capabilities, and massive scalability in capacity and host connectivity.

However, in order to enable the consolidation of multiple tenants onto a single system, hardware and software fault tolerance, as well the ability to predictably prevent downtime and handle failures in a way that is non-disruptive to users and applications, become critical. The HPE 3PAR Architecture supports multi-tenancy by allowing you to consolidate with confidence and achieve higher service levels for more users and applications with less infrastructure.

Advanced iSCSI connectivity enhancements

HPE 3PAR Storage offers a unified solution for IP-only environments for both block (over IP with iSCSI) and file (over IP with SMB/NFS). Although historically IP environments have not been considered suitable for flash, recent iSCSI connectivity enhancements to the platform including VLAN tagging give you the flexibility to deploy flash using Ethernet while preserving sub-millisecond latencies and Tier-1 resiliency and data services.

Latest iSCSI enhancements include:

- IPv6 support
- Support for more than 8,000 iSCSI initiators per system
- Support for VLAN tagging
- Enterprise iSCSI (iSCSI over DCB/lossless Ethernet)
- Send Targets discovery support
- Persistent Ports support

iSCSI VLAN tagging is an initiative that allows iSCSI ports to be configured with multiple IP addresses and VLAN tags, within HPE 3PAR Storage. With support for up to 500 VLAN tags and 256 initiators per port, the traffic to these ports is well streamlined on VLAN membership. This is important in an iSCSI architecture, which is governed by SCSI standards. Handling the network traffic and ensuring all the security protocols are intact becomes bit of a challenge. Especially when SCSI domain consists of more than one iSCSI initiators and iSCSI target; traditional iSCSI traffic is pretty static when frequent changes are to be made to underlying network, on the go. iSCSI VLAN tagging offers network simplification and makes HPE 3PAR iSCSI traffic efficient by smoothening SCSI flows and giving that traffic higher priority.

Support for IPv6 ensures that network activity of each device can potentially be tracked, thus offering the added security boost to the existing infrastructure.

Hardware and software fault tolerance

HPE 3PAR Storage delivers Tier-1 resiliency with an architecture designed to eliminate any single point of failure (hardware or software) in the system. To mitigate single points of failure at the hardware layer, the system is designed with redundant components, including redundant power domains. In fact, to raise the bar with the fault tolerance mechanism, HPE 3PAR 20000 Storage system is configured with two self-encrypting boot drives that work in redundant mode.

An independent copy of HPE 3PAR OS runs on every controller node, so even in the smallest configuration, with two controller nodes, the system offers resiliency also for the software stack.

HPE 3PAR Storage components such as storage nodes, disk- and host-facing host bus adapters (HBAs), power supplies, batteries, and disks all feature N+1 and in some cases N+2 redundancy so that any of these components can fail without system interruption. The only non-redundant component in the system is a 100 percent completely passive controller node backplane/backplanes[1] that, given its passive nature, is virtually impervious to failure.

HPE 3PAR Storage offers up to four current load-balanced power distribution units (PDUs) per rack, which provide a minimum of two separate power feeds. The system can support up to four separate data center power feeds, providing even more power resiliency and further protection against power loss as well as brownouts. Redundant power domains help ensure that as many as two disk chassis power supplies can fail without power being lost to back-end disk devices.

Controller nodes in an HPE 3PAR Storage system includes a local physical drive that contains a separate instance of the HPE 3PAR Operating System as well as space to save cached write data in the event of a power failure.

The controller nodes are each powered by two (1+1 redundant) power supplies and backed up by two batteries. Each battery has sufficient capacity to power the controller nodes long enough to flush all dirty data in cache memory into the local physical drive in the event of a complete power failure to the node. Although many architectures use battery backed RAM as cache (to maintain the data in cache while waiting for power to be restored) these are not suitable for extended downtimes that are usually associated with natural disasters and unforeseen catastrophes.

Another common problem with many battery-powered backup systems is that it is often impossible to ensure that a battery is charged and working. To address this problem, the HPE 3PAR Storage controller nodes are each backed by at least two batteries. Batteries are periodically tested by slightly discharging one battery while the other remains charged and ready in case a power failure occurs while the battery test is in progress. Following a power failure HPE 3PAR Operating System keeps track of battery charge levels and limits the amount of write data that can be cached based on the ability of the batteries to power the controller nodes while they are recharging following the power failure.

The HPE 3PAR Storage power failure protection mechanisms eliminates the need for expensive batteries to power all of the system’s drive chassis, while dirty cache data is destaged to disks on the back end of the array. Note that, because all cached write data is mirrored to another controller node, a system-wide power failure would result in saving cached write data onto the internal drives of two nodes. This offers further protection following a power failure in the event a node in the cluster ID damaged by the power failure. The second node containing the data can be used for recovery of the saved data. Because each node’s dual power supplies can be connected to separate AC power cords, providing redundant AC power to the system can further reduce the possibility of an outage due to an AC power failure.

Advanced fault isolation and RAID protection

Advanced fault isolation and high reliability are built into the HPE 3PAR Storage system. The drive chassis, drive magazines, and physical drives themselves all report and isolate faults. A drive failure will not result in data being unavailable.

HPE 3PAR Storage constantly monitors drives via the controller nodes and enclosures, and isolates faults to individual drives, then “offlines” only the failed component.

HPE 3PAR Storage is capable of RAID 1+0 (dual or triple mirror or striped), RAID 5+0 (RAID 5 distributed parity, striped in an X+1 configuration where X 2 <= X <= 8), or RAID MP (multiple distributed parity, and currently striped with either a 4:2; 6:2; 8:2; 10:2; 14:2 configuration). In an appropriately configured HPE 3PAR array, all available RAID options allow HPE 3PAR Storage to create parity sets on different drives in different drive cages with separate power domains for greater integrity protection.

Each drive enclosure is divided into two redundant I/O modules cages that plug into the drive chassis midplane. The drive chassis components—power supplies, Fibre Channel or SAS Adapters, and drives—are serviceable online and are completely hot-pluggable.

[1] HPE 3PAR 20000 and 9450 systems have a single backplane. 4-Node HPE 3PAR 8000 systems have backplanes and cables.

Redundant power supply/fan assemblies hot-plug into the rear of the midplane. Should the drive chassis midplane fail for any reason, partner cage or cages will continue to serve data for those volumes that were configured and managed as “High Availability (HA) Cage” volumes. If the “HA Cage” configuration setting is selected at volume creation, the controller node automatically manages the RAID 1+0, RAID 5+0, or RAID MP data placement to accommodate the failure of an entire cage without affecting data access.

Controller node redundancy

The HPE 3PAR Operating System instance running on each of the controller nodes is both statefully managed and self-healing, providing protection across all cache-coherent, Mesh-Active storage controller nodes should one or more processes fail and restart. Write cache is mirrored across controllers, and the system offers RAID 1+0 (mirroring + striping), RAID 5+0 (RAID 5 distributed parity + striping), and RAID MP (multiple distributed parity with striping) across the drives behind the nodes.

In addition, controller nodes are configured in logical pairs whereby each node has a partner. The partner nodes have redundant physical connections to the subset of physical drives owned by the node pair. Within the pair, each serves as the backup node for the LDs owned by the partner node.

If a controller node were to fail, data availability would be unaffected. In the event of a node failure, that node’s partner takes over the LDs for the failed node. It then immediately flushes data in the write cache on other nodes in the array that belongs on the LDs it has taken over.

Furthermore, under certain circumstances, the system is capable of withstanding a second Node failure (however rare) without affecting data availability. After the Node failover recovery process for the initial Node failure is complete, a second Controller Node from the remaining Node pairs can fail without causing system downtime.

Data integrity checking

In addition to hardware fault tolerance, all HPE 3PAR Storage systems offer automated end-to-end error checking during the data frames’ journey through the HPE 3PAR Storage array to the disk devices to help ensure data integrity in support of Tier-1 resilience. In addition to this HPE 3PAR Thin Express Gen5 ASIC comes with the Persistent Checksum feature known as T10-DIF that ensures end-to-end data protection, from host HBA to physical drives.

Embedded Cyclical Redundancy Checking (CRC) checking includes, but is not exclusive to, the following layers within all HPE 3PAR Storage systems:

- CRC/parity checks on all internal CPU and serial buses
- Control cache ECC checks
- Data cache ECC checks
- PCIe I2C bus CRC/parity checks
- HPE 3PAR ASIC connection CRC/parity checks
- Protocol (Fibre Channel/iSCSI/FCoE) CRC checks at the frame level (hardware accelerated via the HPE 3PAR ASICs)
- Disk devices CRC checks at the block level, occurring once the data has landed and throughout the lifecycle of the data once it’s stored to disk

CRC error checking is also extended to replicate data with HPE 3PAR Remote Copy software, which help ensure that potential cascaded data issues do not occur. HPE 3PAR Storage replication includes a link pre-integration test to verify the stability of Remote Copy replication links in advance of adding these links within the HPE 3PAR Storage system for use with HPE 3PAR Remote Copy over an IP network (RCIP).

All drives in the HPE 3PAR 20000, 9000, and 8000 storage systems are formatted with 520-byte blocks in order to provide space to store a CRC Logical Block Guard, as defined by the T10 Data Integrity Feature (T10-DIF) for each block. This value is computed by the HPE 3PAR HBA before writing each block and is checked when a block is read. NL SAS does not support 520-byte blocks, so on Enterprise NL SAS drives, data blocks are logically grouped with an extra block to store the CRC values. The CRC Logical Block Guard used by the T10-DIF is automatically calculated by the host HBAs to validate data stored on drives without additional CPU overhead.

HPE 3PAR Storage continuously runs a background “PD scrubber” process to scan all blocks of the physical drives in the system. This is done to detect any potential issues at the device block layer and trigger RAID rebuilds down to 512-byte granularity if necessary. This is particularly important when it comes to flash media because it allows the system to proactively detect and correct any low-level CRC and bit errors.

Furthermore, Self-Monitoring, Analysis and Reporting Technology (SMART) predictive failures mean that any disk device crossing certain SMART thresholds would cause the storage controller nodes to mark a drive as “predictive failure,” identifying it for replacement before it actually fails.

When inline deduplication is used, the controller node ASICs will perform a bit-by-bit comparison before any new write is marked as a duplicate. This helps ensure data integrity by introducing this critical check into the deduplication process to support mission-critical environments.

HPE 3PAR Storage systems also issue logical error status block (LESB) alerts if a frame arriving in the storage interface has CRC errors beyond a certain threshold. This indicates that a cable or component between the host and storage device needs replacing or cleaning.

Memory fencing

HPE 3PAR Storage is able to correct single-bit (correctable) errors and detect double-bit (uncorrectable) errors.

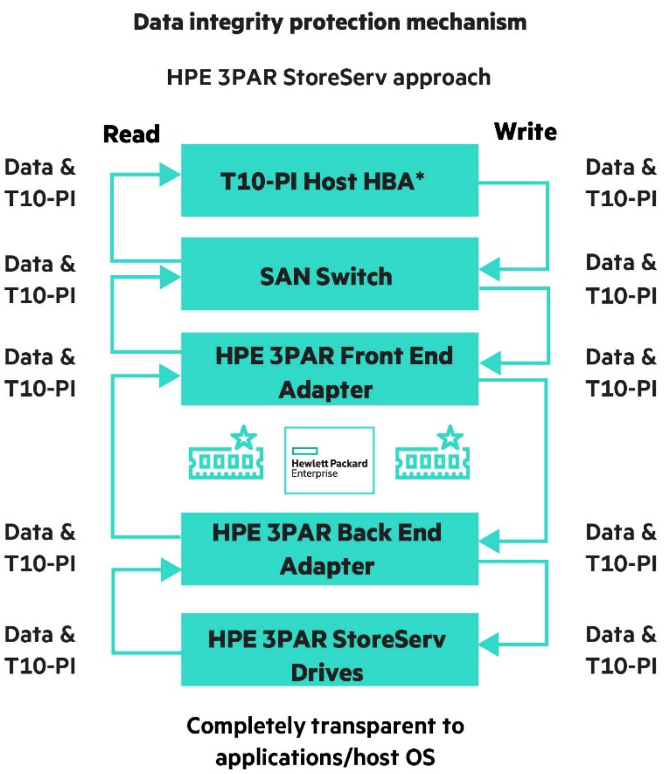
It achieves this by using a thread (memory patrol) that continuously scans the memory and keeps track of correctable errors at a 16 KB page granularity. If, during the scan, the thread detects uncorrectable errors, areas of the memory are fenced and put onto a “do not use” list. The system will raise a service alert when the threshold of correctable errors is reached and/or memory is fenced such that replacing the DIMM is recommended.

Persistent technologies

No one has time for downtime, which is why modern Tier-1 resiliency requires that data access and service levels be maintained during failure recovery, maintenance, and software upgrades. Tier-1 resiliency demands that failures not only be prevented, but that the system can recover quickly in the event that something goes wrong. Not only is HPE 3PAR Storage designed to be non-disruptively scalable and upgradable, but the system also has several advanced features to prevent unnecessary downtime and to maintain availability and performance levels during planned as well as unplanned outage events. These features are collectively known as persistent technologies.

Persistent Checksum

This feature, Persistent Checksum addresses media and transmission errors that can be caused by any component in the I/O stack all the way from the server HBA through the SAN switches and into the HPE 3PAR HBAs making the data secure all the way from the hosts right to the drives and providing additional protection above CRC transmissions for Fibre Channel alone. Persistent Checksum is server and application independent (it does require server HBAs that support the feature) and offers elaborate host OS support. When using unsupported HBAs T10-DIF tags are added and verified on the array target ports, inter-node copies and back-end HBAs. When using supported HBAs, T10-DIF tags are added by the Host HBAs and verified throughout the HPE 3PAR Storage system, making the data secure all the way from the hosts to the drives. Where Persistent Checksum detects media or transmission errors, graceful error recovery will take place, avoiding impact on the host application.



*For a list of supported T10-PI HBA see HP (now Hewlett Packard Enterprise) SPOCK

FIGURE 10. Persistent Checksum

Persistent Cache

HPE 3PAR Persistent Cache is a resiliency feature built into the HPE 3PAR Operating System that allows graceful handling of an unplanned controller failure or planned maintenance of a controller node. This feature eliminates the substantial performance penalties associated with traditional modular arrays and the cache "write-through" mode they have to enter under certain conditions. HPE 3PAR Storage can maintain high and predictable service levels even in the event of a cache or controller node failure by avoiding cache write-through mode via HPE 3PAR Persistent Cache.

Under normal operation on an HPE 3PAR Storage system, each controller has a partner controller in which the controller pair has ownership of certain logical disks. As mentioned earlier, LDs are the second layer of abstraction in the system's approach to virtualization of physical resources. In the event of controller node failure, HPE 3PAR Persistent Cache preserves write caching by having the surviving node from a node pair, where one node has failed, mirror writes that would have gone to the failed node to other nodes in the cluster instead. By doing this, the surviving node does not have to go into write-through mode for the LDs it owns to ensure data integrity in the unlikely event it should fail too.

For example, in a quad controller configuration (where Node 0 and Node 1 form a node pair and Node 2 and Node 3 form a second node pair), each node pair might own 100 LDs with each node within the pair fulfilling the role of the primary node for 50 of those LDs. If Node 2 fails, the system will transfer ownership of its 50 LDs to Node 3, and Node 0 and Node 1 will now be the backup (and thereby the cache mirroring partner) for the 100 LDs that Node 3 is now responsible for. The mirroring of write data coming into Node 3 for those 100 LDs will distributed across Node 0 and Node 1.

Persistent Ports

Another persistent technology, HPE 3PAR Persistent Ports allows for a non-disruptive environment (from the host multipathing point of view) where host-based multipathing software is not required to maintain server connectivity in the event of a node or link outage on any SAN fabric. This applies to firmware upgrades, node failures, and node ports that are taken offline either administratively, or as the result of a hardware failure in the SAN fabric that results in the storage array losing physical connectivity to the fabric.

From a host standpoint, connections to HPE 3PAR Storage systems continue uninterrupted with all I/O being routed through a different port on the HPE 3PAR Storage array. This helps you achieve an uninterrupted service level for applications running on HPE 3PAR Storage systems.

Persistent Port functionality works for Fibre Channel, FCoE, and iSCSI transport layers and provides transparent and uninterrupted failover in response to the following events:

- HPE 3PAR OS firmware upgrade
- HPE 3PAR node maintenance or failure
- HPE 3PAR array "loss sync" to the FC fabric or iSCSI network
- Array host ports being taken offline administratively
- Port laser loss for any reason (applies to FC only)

Multi-System Software

HPE 3PAR Multi-System Software brings a rich set of features that can be used to design disaster-tolerant solutions that cost-effectively address disaster recovery challenges. It is a uniquely easy, efficient, and flexible replication technology that allows you to protect and share data from any application.

Implemented over a native IP network[2] (through the built-in Gigabit Ethernet interface available on all nodes) and native Fibre Channel, users may flexibly choose one of four different data replication modes—**Asynchronous Streaming** or Asynchronous Periodic (for asynchronous replication), **Synchronous**, or **Synchronous Long Distance**—to design a solution that meets their solution requirements for recovery-point objectives (RPOs) and recovery-time objectives (RTOs).

With all four of these modes, HPE 3PAR Multi-System software allows you to mirror data between HPE 3PAR Storage systems eliminating the incompatibilities and complexities associated with trying to mirror data between traditional vendors' midrange and enterprise array technologies. Source and target volumes may also be flexibly and uniquely configured to meet users' needs (e.g., different RAID levels, the use of FPVVs vs. TPVVs, or different drive types). HPE 3PAR Remote Copy is "thin aware" in that it is able to replicate both thin and thick volumes by using TPVV target volumes to provide the same cost savings associated with thin-provisioned source volumes created with HPE 3PAR Thin Provisioning software.

[2] Asynchronous Streaming is not support on the built-in GigE interface at the release of HPE 3PAR OS version 3.2.2.

Synchronous Replication
Provides zero data loss in the event of failure for the ultimate RPO but can have an impact on host performance. As spinning media solutions measure performance in tens of milliseconds, creating an exact copy of data over an extended distance adds some latency but it's generally acceptable in terms of meeting service-level agreements. All-flash systems are far more sensitive to latency overheads as performance is now measured in microseconds so any overhead measured in milliseconds can significantly increase latency. The overhead associated with replicating every write request over an IP links twice (round trip) has this impact.

Periodic Asynchronous Replication
Based on snapshots and delta syncs, has minimal impact on host performance but it does require compromise as RPOs are measured in minutes, not seconds or milliseconds. This may be suitable for many environments where RPOs in minutes are acceptable but data compliance and business requirements can often drive the need for lower RPOs. Changed data within an HPE 3PAR Remote Copy Volume Group is transferred only once—no matter how many times it may have changed—between synchronization intervals. Additionally, efficiencies in the initial copy creation of the target volumes that do not require replication of "zero" data across the replication network (regardless of target volume type, thick or thin) result in a faster initial synchronization and better network utilization.

Asynchronous Streaming Replication
HPE 3PAR offers the most complete set of disaster recovery solutions in the industry including support for various forms of replication through the Multi-System Software suite of products. The latest addition to this extensive arsenal is Asynchronous Streaming replication. This extends resiliency options with a replication technology that is ideal for use in all-flash systems as it not only offers near-zero RPOs but also doesn't impact host performance.

Synchronous Long Distance
SLD delivers a disaster recovery solution across long distances with a potential for zero data loss RPO and an RTO of minutes. This is achieved with a replication configuration that uses three sites (3DC) to simultaneously replicate a virtual volume from the primary array in synchronous mode to a target HPE 3PAR Storage array located at a synchronous site (within a metropolitan area) and in asynchronous periodic mode to an HPE 3PAR Storage array located at an asynchronous site (across a long distance). In addition to the HPE 3PAR Remote Copy connections from the primary array to the two backup arrays, a passive asynchronous periodic link is configured from the synchronous target array to the disaster recovery target array. Under the Synchronous Long Distance mode algorithm, the synchronous target array intelligently tracks the delta set of I/Os that have been acknowledged to the host but which have not yet been replicated to the asynchronous target array. In the event that a disaster takes the primary storage array down, the user has the flexibility to recover either from the synchronous target array or the asynchronous target array.

Privacy, security, and multi-tenancy

Virtual Domains
HPE 3PAR Virtual Domains software is an extension of HPE 3PAR virtualization technologies that delivers secure segregation of virtual private arrays (VPAs) for different user groups, departments, and applications while preserving the benefits delivered by the massive parallelism architected into the HPE 3PAR platform. It supports HPE 3PAR's Multi-Tenancy Paradigm.

By providing secure administrative segregation of users and hosts within a consolidated, massively parallel HPE 3PAR Storage system, HPE 3PAR Virtual Domains allows individual user groups and applications to affordably achieve greater storage service levels (performance, availability, and functionality) than previously possible.

HPE 3PAR Virtual Domains is completely virtual and represents no physical reservation of resources. To use HPE 3PAR Virtual Domains, a master administrator first creates a virtual domain, and then assigns logically defined entities to it. These include one or more host definitions based on Worldwide Name (WWN) groupings, one or more provisioning policies (RAID and disk type), and one or more system administrators (who are also granted role-based privileges by the master administrator).

Depending on the level of access, users can create, export, and copy standard volumes or thin-provisioned volumes. HPE 3PAR Virtual Domains is ideal for enterprises or service providers looking to leverage the benefits of consolidation and deploy a purpose-built infrastructure for their private or public cloud.

Data encryption
Data is perhaps the most important asset for organizations in today's digital age. Companies are looking to protect data against theft and misuse while meeting compliance requirements. The HPE 3PAR Storage complies with the standards set forth by the National Institute of Standards and Technology (NIST) and FIPS 140-2 (Federal Information Processing Standard) and features Data-at-Rest (DAR) encryption that helps protect valuable data through self-encrypting drive (SED) technology. SED drives are HDDs and SSDs with a circuit (ASIC) built into the drive's controller chipset that automatically encrypts and decrypts all data being written to and read from the media.

HPE 3PAR Storage supports Full Disk Encryption (FDE) based on the Advanced Encryption Standard (AES) 256 industry standard. The encryption is part of a hash code that is stored internally on physical media. All encryption and decryption is handled at the drive level and needs no other external mechanism.

General Data Privacy Regulation: The General Data Protection Regulation (GDPR) is a new European privacy law which came into force on 25 May 2018 and significantly increased the risks for companies that fail to use and protect personal data in compliance with the law. The GDPR introduces significant monetary penalties of up to a maximum of 20 million Euros or 4% of the annual worldwide turnover of a corporate group. The GDPR requires organizations to implement appropriate technical and organizational measures to secure data and introduces new breach notification requirements.

HPE 3PAR Storage by its inherent design and architecture with security that is built into the product will assist customers in meeting their GDPR security requirements HPE 3PAR security categories can be identified as the following:

- Authorization
- Authentication
- Availability
- Encryption
- Integrity
- Auditing

These categories are all fundamental security aspects by which HPE 3PAR continues to enhance and harden overall product architecture. HPE 3PAR has already and will continue to adopt security by design into its operating system, appliances and tools which support the array.

Transport Layer Security (TLS) 1.2-only support: HPE 3PAR OS 3.3.1, 3.2.2 MU6 Patch 107 and 3.2.2 MU4 Patch 106 and above enables TLS 1.2-only configurations which eliminate any potential impact of security vulnerabilities by preventing TLS 1.0/1.1 connections, which allows 3PAR customers to enhance their Payment Card Industry Data Security Standard (PCI DSS) 3.2 compliance strategy.

Authentication keys are set by the user and can be changed at any time. The Local Key Manager (LKM) included with the HPE 3PAR Storage encryption license is used to manage all drive encryption keys within the array and provides a simple management interface. In the event of a drive failure or the theft of a drive, a proper key sequence needs to be entered to gain access to the data stored within the drive. When an SED drive is no longer powered on, the drive goes into a locked state and requires an authentication key to unlock the drive when power is restored. Without the key, access to the data on the SED is not possible.

Hewlett Packard Enterprise also offers the enhanced encryption support on the HPE 3PAR Storage systems by offering FIPS 140-2 compliant SED drives that provides the ability to use an external Enterprise Secure Key Manager (ESKM). ESKM is deployed whenever customers use encrypted storage or communication methods to protect their sensitive information. Herein, they store and serve keys to unlock the data stored on FIPS 140-2 compliant drives within the HPE 3PAR Storage systems with strong access controls and security.

FIPS 140-2 compliance provides the customer the satisfaction of knowing their data is securely stored on the HPE 3PAR array. Key Management on the array with either LKM or ESKM coupled with FIPS drives, offers customers a safe environment in which to securely store their data.

2-Factor Authentication

3PAR (3PAR OS 3.3.1 and SSMC 3.1 and later) enabled 2-Factor Authentication support in SSMC through Common Access Cards (CAC) and with SSMC 3.2 support for Virtual Smart Card as the second factor of authentication has been made available in subsequent releases.

For more information, see the HPE 3PAR Data-At-Rest Encryption

MAINTAINING HIGH AND PREDICTABLE PERFORMANCE LEVELS

The ability of HPE 3PAR Storage to maintain high and predictable performance in multi-tenant environments is made possible through architectural innovations that utilize all available array hardware resources at all times thereby eliminating resource contention, support mixed workloads, and enhance caching algorithms to accelerate performance and reduce latency.

Load balancing

Purpose-built for the Enterprise as well as virtual and cloud data centers, the HPE 3PAR Architecture is unlike legacy controller architectures in that the Mesh-Active system design allows each volume to be active on any controller in the system via the high-speed, full-mesh interconnection that joins multiple controller nodes to form a cache-coherent Active/Active cluster. As a result, the system delivers symmetrical load balancing and utilization of all controllers with seamless performance scalability by adding more controllers and disk drives to the system.

Mixed-workload support

Unlike legacy architectures that process I/O commands and move data using the same processor complex, the HPE 3PAR Storage controller node design separates the processing of SCSI control commands from data movement. This allows transaction-intensive and throughput-intensive workloads to run on the same storage resources without contention, thereby supporting massive consolidation and multi-tenancy. This means that, for example, the system can easily handle an OLTP application and an extremely bandwidth-consuming data warehousing application concurrently with ease.

This capability is made possible by the HPE 3PAR ASIC, which offloads data processing from the control processor, where metadata is processed. By pathing and processing data and metadata separately, transaction-intensive workloads are not held up behind throughput-intensive workloads. As a result, the HPE 3PAR Storage platform, as compared to the architectures of traditional storage vendors—including many of today's all-flash arrays—delivers excellent performance consistently, even in mixed-workload scenarios. Figure 11 illustrates the benefits of mixed-workload support.

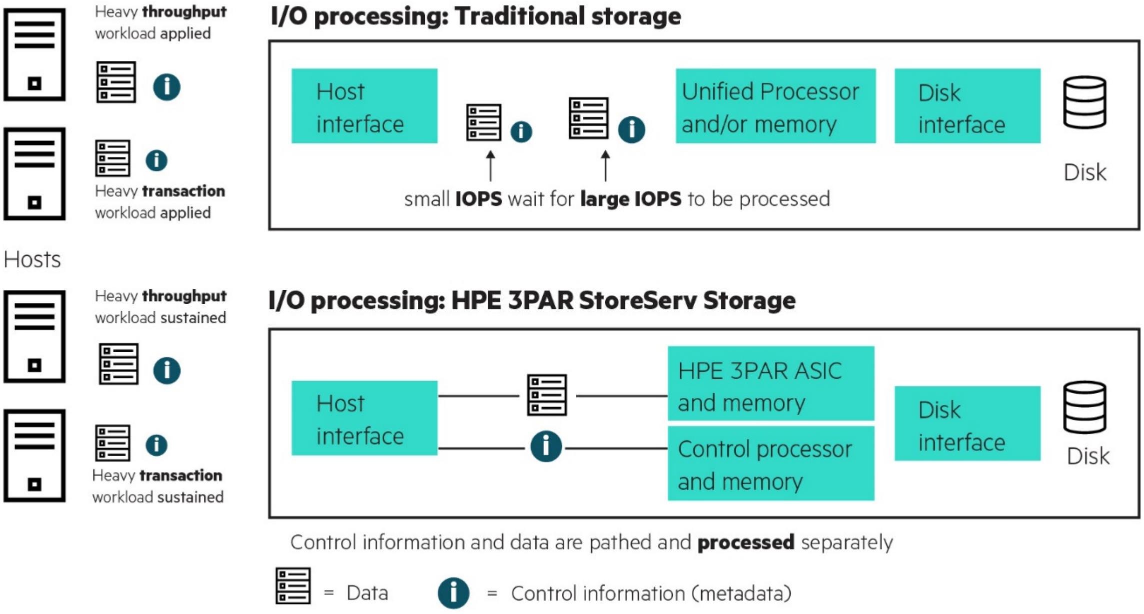


FIGURE 11. HPE 3PAR Storage with mixed-workload support

Control operations are handled as follows:

- With the HPE 3PAR 20800 Storage system, control operations are processed by up to 16 high-performance Intel Six-Core processors.
- With the HPE 3PAR 20840 Storage system, control operations are processed by up to 16 high-performance Intel Eight-Core processors.
- With the HPE 3PAR 20850 Storage system, control operations are processed by up to 16 high-performance Intel Eight-Core processors.
- With the HPE 3PAR 20450 Storage system, control operations are processed by up to 8 high-performance Intel Eight-Core processors.
- With the HPE 3PAR 8450 and 8440 Storage systems, control operations are handled by up to four Intel 10-core processors.
- In the case of the HPE 3PAR 8400 Storage system, control operations are handled by up to four Intel Hexa-core processors.
- For the HPE 3PAR 8200 Storage system, control operations are handled by up to two Intel Hexa-core processors.

Data movement is handled as follows:

- For the HPE 3PAR 9450 and 20000 Storage system, all data movement is handled by the specially designed HPE 3PAR ASICs (two per controller node).
- For the HPE 3PAR 8000 Storage system, all data movement is handled by the HPE 3PAR ASICs (one per controller node).

Priority Optimization

Quality of service (QoS) is an essential component for delivering modern, highly scalable multi-tenant storage architectures. The use of QoS moves advanced storage systems away from the legacy approach of delivering I/O requests with "best effort" in mind and tackles the problem of "noisy neighbors" by delivering predictable tiered service levels and managing "burst I/O" regardless of other users in a shared system. Mature QoS solutions meet the requirements of controlling service metrics such as throughput, bandwidth, and latency without requiring the system administrator to manually balance physical resources. These capabilities eliminate the last barrier to consolidation by delivering assured QoS levels without having to physically partition resources or maintain discreet storage silos.

HPE 3PAR Priority Optimization software enables service levels for applications and workloads as business requirements dictate, enabling administrators to provision storage performance in a manner similar to provisioning storage capacity. This allows the creation of differing service levels to protect mission-critical applications in enterprise environments by assigning a minimum goal for I/O per second and bandwidth, and by assigning a latency goal so that performance for a specific tenant or application is assured. It is also possible to assign maximum performance limits on workloads with lower service-level requirements to make sure that high-priority applications receive the resources they need to meet service levels.

Priority Optimization feature and industries leading "Latency goal" feature that enables the storage administrator to set SLAs as low as 500 μs for volumes residing on SSD storage. It also makes it possible to configure service-level objectives in terms of KB/s and I/O bandwidth on a virtual volume set (VVset) or between different virtual domains.[3] All host I/Os on the VVset are monitored and measured against the defined service-level objective. HPE 3PAR Priority Optimization control is implemented within the HPE 3PAR Storage system and can be modified in real time. No host agents are required, and physical partitioning of resources within the storage array is not necessary.

Adaptive Flash Cache

HPE 3PAR Adaptive Flash Cache is built-in functionality of the HPE 3PAR that allows SSDs to act as Level-2 read cache holding random read data for spinning media that has aged out of DRAM read cache. Adaptive flash cache reduces application response time for read intensive I/O workloads and can improve write throughput in mixed-workload environment. Adaptive Flash Cache effectively increases the amount of random read data cached on high-speed media on a node. By doing this, Adaptive Flash Cache can increase the overall random read IOPS an array can deliver by "unloading" the back end of the array potentially resulting in increased write throughput on the array. Using SSDs as Level-2 read cache to hold random read data that has aged out of DRAM cache is a cost-effective way of keeping more random read data on very fast media to improve overall system performance.

For more information on HPE 3PAR Adaptive Flash Cache, please see the HPE 3PAR Adaptive Flash Cache white paper.

NVMe Storage Class Memory Module

Select HPE 3PAR All Flash storage now can take advantage of storage technology innovations to accelerate application performance with an NVMe Storage Class Memory module, which enhances the performance of All-Flash storage systems in specific small block workloads by decreasing the latency for specified volumes. Intelligent caching algorithms are used to extend dram cache to SCM devices through NVMe transport. NVMe SCM module is available as an easy to install add-on card and simply takes advantage of an existing node PCI slot-3.

Express Writes

HPE 3PAR Express Writes represent a series of optimizations aimed to improve host write latency. This is achieved via the HPE 3PAR OS by optimizing SCSI transactions and thus improving the number of interrupts per transaction, which results in improved array CPU utilization and lower host write latency. Express Writes technology enables HPE 3PAR array to achieve synchronous write latency below 200 microseconds. Depending on the workload, hosts may see an overall decrease in write latency of up to 20 percent, which may result in an increase in IOPS and throughput.

With Express Writes disabled 90 percent of write I/O was committed in 0.250 microseconds or more and 10 percent was committed in 0.125 microseconds or less. With Express Writes Enabled 50 percent of write I/O is committed in 0.125 microseconds or less.

[3] A VVset may contain a single volume or multiple volumes. A virtual volume may also belong to multiple virtual volume sets allowing users to create hierarchical rules.

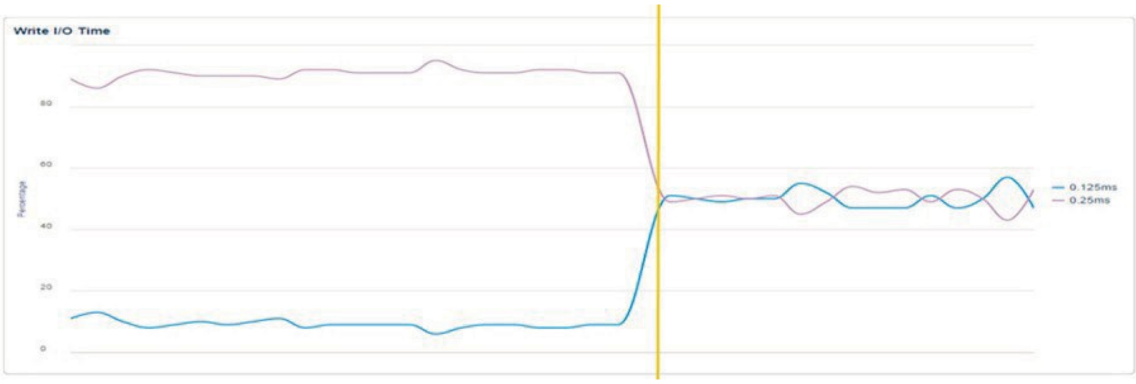


FIGURE 12. The following picture is from a system with Express Writes disabled and then enabled (after the line)

Performance benefits of system-wide striping

In a traditional storage array, small volumes either suffer from poor performance by using few drives or waste expensive resources by using more drives than required for capacity in order to obtain sufficient performance. On HPE 3PAR Storage systems, even modest-sized volumes will be widely striped using chunklets spread over multiple drives of the same type. Wide striping provides the full performance capabilities of the array (nodes, CPUs, buses, cache, disk drives) to small volumes without provisioning excess capacity and without creating hotspots on a subset of physical drives.

Physical drives can hold a mix of RAID levels because RAID groups with HPE 3PAR Storage are constructed from chunklets rather than from whole drives. Different chunklets on a physical drive can be used for volumes with different RAID levels. On a traditional array, a storage administrator might be forced to use RAID 1 for an archival volume in order to use space that is available on a RAID 1 disk even though RAID 5 would deliver adequate performance and better space utilization. The chunklet-based approach employed by the HPE 3PAR Operating System allows all RAID levels to coexist on the same physical drives, using the better RAID level for each volume. Additional details about striping are provided in the "Highly virtualized storage operating system" section.

Bandwidth and communication

The ASICs within each HPE 3PAR 20800, 20840, 20850, and 20450 Storage controller node to serve as the high-performance engines that move data between three I/O buses, a four memory-bank data cache, and seven high-speed links to the other controller nodes over the full-mesh backplane. These ASICs perform RAID parity calculations and inline zero-detection to support the system's data compaction technologies. CRC Logical Block Guard used by T10-DIF is automatically calculated by the ASIC's to validate data stored on drives with no additional CPU overhead. An HPE 3PAR 20800 Storage system with eight controller nodes has 16 ASICs totaling 224 GiB/s of peak interconnect bandwidth.

The single ASIC within each HPE 3PAR 8450, 8400, and 8200 Storage controller node serves as the high-performance engine that moves data between three I/O buses, a dual memory-bank data cache, and three high-speed links to the other controller nodes over the full-mesh interconnect. As with the HPE 3PAR 20000 Storage series, the ASICs for the HPE 3PAR 8000 Storage series models perform parity RAID calculations and inline zero-detection; CRC Logical Block Guard used by the T10-DIF is automatically calculated by the ASIC's to validate data stored on drives with no additional CPU overhead. An HPE 3PAR 8450 Storage system with four nodes has 4 ASICs totaling 48 GiB/s of peak interconnect bandwidth.

Data transfer paths

Figure 13 shows an overview of data transfers in an HPE 3PAR Storage system with two simple examples: a write operation from a host system to a RAID 1 volume (lines labeled W1 through W4), and a read operation (Gray lines labeled R1 and R2). Only the data transfer operations are shown, not the control transfers.

The write operation consists of:

- **W1:** Host writes data to cache memory on a controller node
- **W2:** The write data is automatically mirrored to another node across the high-speed backplane link so that the write data is not lost even if the first node experiences a failure; only after this cache mirror operation is completed is the host's write operation acknowledged
- **W3** and **W4:** The write data is written to two separate drives (D1 and D1'), forming the RAID 1 set

In step W2, the write data is mirrored to one of the nodes that owns the drives to which data will be written (in this example, D1 and D1'). If the host's write (W1) is to one of these nodes, then the data will be mirrored to that node's partner.

HPE 3PAR Persistent Cache allows a node to mirror the write data to a node that does not have direct access to drives D1 and D1' in the event of a partner node failure.

The read operation consists of:

- **R1:** Data is read from drive D3 into cache memory
- **R2:** Data is transferred from cache memory to the host

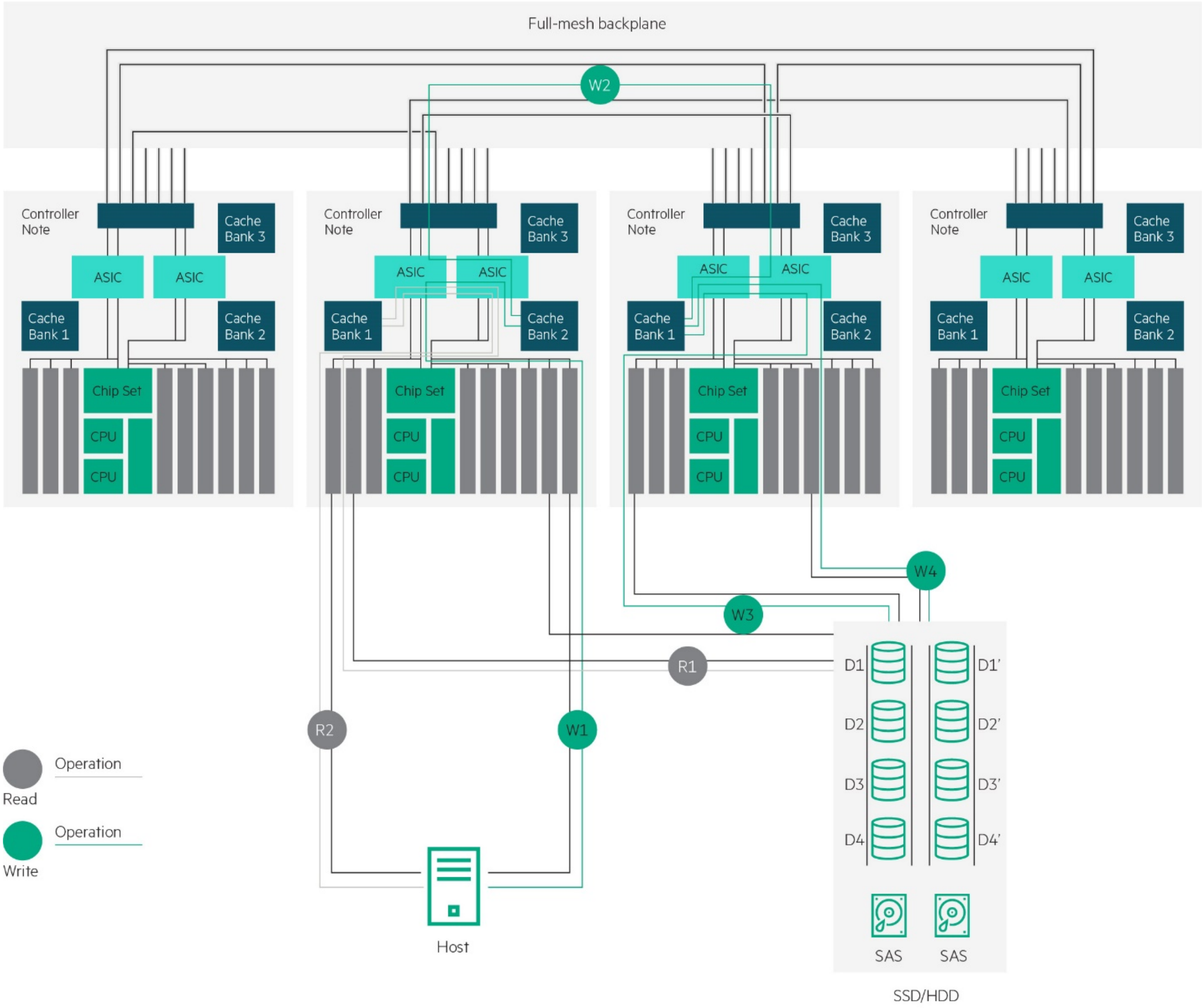


FIGURE 13. Data transfer paths

I/O bus bandwidth is a valuable resource in the controller nodes, and is often a significant bottleneck in traditional arrays. As the data transfer example in Figure 13 illustrates, I/O bus bandwidth is used only for data transfers between the host-to-controller node and controller node-to-drive transfers. Transfers between the controller nodes do not consume I/O bus bandwidth.

Processor memory bandwidth is another significant bottleneck in traditional architectures, and is also a valuable resource in the controller nodes. Unique to the HPE 3PAR Storage system, controller node data transfers do not consume any of this bandwidth. This frees the processors to perform their control functions far more effectively. All RAID parity calculations are performed by the ASICs and do not consume processor or processor memory bandwidth.

Sharing and offloading of cached data

Because much of the underlying data associated with snapshot volumes is physically from the base VVs, data that is cached for the base VV can often be used to satisfy read accesses for a snapshot of that base VV.

In the event that two or more drives that underlay a RAID set become temporarily unavailable (or three or more drives for RAID MP volumes)—for example, if all cables to those drives are accidentally disconnected—the HPE 3PAR Operating System automatically moves any “pinned” writes in cache to dedicated Preserved Data LDs. This helps ensure that all host-acknowledged data in cache is preserved so that it can be properly restored once the destination drives come back online without compromising cache performance or capacity with respect to any other data by keeping cache tied up.

On flash-based systems, autonomic cache offload mitigates cache bottlenecks by automatically changing the frequency at which data is offloaded from cache to flash media. This helps ensure high performance levels consistently as workloads are scaled to hundreds of thousands of IOPS.

Pre-fetching

The HPE 3PAR Operating System keeps track of read streams for VVs so that it can improve performance by “pre-fetching” data from drives ahead of sequential read patterns. In fact, each VV can detect up to five interleaved sequential read streams and generate pre-fetches for each of them. Simpler pre-fetch algorithms that keep track of only a single read stream cannot recognize the access pattern consisting of multiple interleaved sequential streams.

Pre-fetching improves sequential read performance in two ways:

- The response time seen by the host is reduced.
- The drives can be accessed using larger block sizes than the host uses, resulting in more efficient operations.

Write caching

Writes to VVs are cached in a controller node, mirrored in the cache of another controller node, and then acknowledged to the host. The host, therefore, sees an effective response time that is much shorter than would be the case if a write were actually performed to the drives before being acknowledged. This is possible because the mirroring and power failure handling help ensure the integrity of cached write data.

In addition to dramatically reducing the host write response time, write caching can often benefit back-end drive performance by:

- Merging multiple writes to the same blocks so that many drive writes are eliminated
- Merging multiple small writes into single larger drive writes so that the operation is more efficient
- Merging multiple small writes to a RAID 5 or RAID MP LD into full-stripe writes so that it is not necessary to read the old data for the stripe from the drives
- Delaying the write operation so that it can be scheduled at a more suitable time

Additional benefit to write caching is provided via Adaptive Flash Cache; while the technology is aimed to accelerate reads, the fewer random read request to the HDD tier translate to more pages available for writes.

Fast RAID 5

The architectural design of the HPE 3PAR Storage systems enables RAID 5 redundancy with performance levels that are on par with RAID 1 mirroring. This implementation combines the HPE 3PAR ASIC, a large, memory cache, and wide striping for reducing spindle contention to offer performance that approaches that of RAID 1, thus reducing the performance impact typical of RAID 5 on legacy storage architectures.

For certain workloads, Fast RAID 5 can provide higher performance than RAID 1. The write-back cache in HPE 3PAR Storage systems allows sequential writes (as generated by transaction journals, logs, and similar performance-sensitive workloads) to be collected until a full parity group can be written, reducing disk I/O traffic and possible back-end bottlenecks. Given its layout algorithm, Fast RAID 5 is appropriate for volumes that are dominated by read activity. HPE 3PAR Storage systems allow selection of the number of data blocks per parity block (N+1) to suit different needs. For RAID 5, 3+1 is the default, but any value from 2+1 to 8+1 can be selected. Higher values of N result in higher storage efficiency, but can reduce the chances for full-stripe writes. Hewlett Packard Enterprise customers using HPE 3PAR Storage arrays typically choose HPE 3PAR Fast RAID 5 for most or all volumes, as Fast RAID 5 provides greater storage efficiency with minimal performance degradation compared to RAID 1.

HPE 3PAR RAID MP (Fast RAID 6)

Exponential growth in HDD capacity without commensurate improvements in reliability or performance results in greater risk of data loss. For example, consider the 600 GiB SAS disks and 6 TiB nearline SAS disks available on HPE 3PAR Storage systems. The capacity difference alone implies that reconstruction of a failed disk upon replacement can be expected to take more than six times as long with the 6 TiB disk. The nearline disks are slower, too, which further increases the mean time to repair (MTTR) relative to smaller, faster SSD or SAS disks. A longer MTTR creates a larger window during which a second disk failure could cause data loss when using RAID 1 or RAID 5. Fast RAID 6 was created to address this problem. Like Fast RAID 5, Fast RAID 6 uses distributed parity, but it stores two different parity values in a manner that allows the data to be reconstructed, even in the event of two drive failures.

HPE 3PAR RAID MP (multiple, distributed parity) initially supports dual parity and is capable of supporting higher parity levels in the future.

Environments such as highly consolidated virtual host environments tend to have unusually high data protection requirements due to the large number of users that could be affected by data loss, and so demand the highest level of data protection.

High I/O loads make RAID 6 problematic on traditional arrays; the implementation of RAID MP on HPE 3PAR Storage arrays is the only choice that provides the extra level of data protection without compromising I/O performance.

HPE 3PAR Gen5 ASIC for inline data reduction

At the heart of every HPE 3PAR node there is an HPE 3PAR ASIC that features an efficient, silicon-based zero-detection and hashing mechanism. This unique hardware capability gives HPE 3PAR Storage the power to perform inline deduplication and remove allocated but unused space inline and non-disruptively.

For thinly allocated volumes (and thin dedupe volumes) Zero-Detect capability can recognize an incoming write request of 16 KiB of zeros and either not allocate space for the zero block or free the space if it was already allocated for that block. All this happens in cache, and therefore no zeroes are written to the back end. When a read request comes in for a block that is unallocated, the HPE 3PAR will immediately return zeros back to the host. Built-in zero-detection can be controlled per Thin Volume, and it is enabled by default.

Many other storage arrays do not detect blocks of zeroes on write. Instead, the zeros are written to disk and a scrubbing process later detects these zeroed blocks and discards them. With this approach, the zeroed blocks consume space until they're scrubbed, and therefore they may not be available for use by other volumes when needed. Also, there is increased load placed on the storage as the scrubbing process examines the block contents on the physical storage.

The HPE 3PAR Gen5 ASIC is also the only solution in the industry with a built-in, silicon-based hash calculation engine. With HPE 3PAR Deduplication, the process of calculating the hash signatures for incoming data and verifying reads are offloaded to the HPE 3PAR ASICs, freeing up processor cycles to deliver advanced data services and service I/O requests. This hardware-assisted approach, together with a unique Express Indexing feature, enables extremely efficient, extremely granular block-level inline deduplication that carries multiple benefits, including increased capacity efficiency, flash performance protection, and flash media life span extension. Unlike other approaches, HPE 3PAR Deduplication performs collision detection on all data before marking it as duplicated, which is essential to ensuring data integrity for mission-critical environments.

Thin technologies highlights

- Thin technologies are completely automated.
- HPE 3PAR Operating System (OS) software uses a reservationless, dedicate-on-write approach to thin volumes and virtual copies that draws and configures capacity in fine-grained increments from a single free space reservoir without pre-dedication of any kind.
- Thin technologies use an allocation unit size of just 16 KiB, so you don't have to worry about small writes consuming megabytes or even gigabytes of capacity.
- HPE 3PAR is a storage platform built from the ground up to support Thin technologies by reducing the diminished performance and functional limitations that plague bolt-on thin solutions.

Thin Provisioning

Since its introduction in 2002, 3PAR has been widely considered the gold standard in thin provisioning. 3PAR's thin provisioning solution leverages the system's dedicate-on-write capabilities to make storage more efficient and more compact, allowing customers to purchase only the disk capacity they actually need and only as they actually need it.

Thin Persistence

Thin Persistence is an optional feature that keeps TPVVs and read/write snapshots of TPVVs small by detecting pages of zeros during data transfers and not allocating space for the zeros. This feature works in real time and analyzes the data before it is written to the source TPVV or read/write snapshot of the TPVV. Freed blocks of 16 KB of contiguous space are returned to the source volume, and freed blocks of 128 MB of contiguous space are returned to the CPG for use by other volumes.

Thin Conversion

With Thin Conversion, a technology refresh does not requires terabyte-for-terabyte replacement, but instead offers the opportunity to eliminate a significant amount of legacy capacity through fat-to-thin conversion made possible by leveraging the zero-detection and inline deduplication capabilities within the HPE 3PAR ASIC and a unique virtualization mapping engine for space reclamation that powers the simple and rapid migration of inefficient “fat” volumes on legacy arrays to more efficient, higher-utilization “thin” volumes on the HPE 3PAR Storage array. Virtual volumes with large amounts of allocated but unused space are migrated to TPVVs or TDVVs that are much smaller than the original volumes. During the migration process, allocated but unused space is discarded and the result is a TPVV that uses less space than the original volume. If TDVVs volumes are chosen all duplicate data will be detected during the migration and only one copy of the data will be written to the HPE 3PAR.

Thin Copy Reclamation

An industry first, Thin Copy Reclamation keeps storage as lean and efficient as possible by reclaiming the unused space resulting from deleted virtual copy snapshots. This solution builds on a virtualization mapping engine for space reclamation called HPE 3PAR Thin Engine, which is included as part of the HPE 3PAR Operating System.

Thin Copy Reclamation is an optional feature that reclaims space when snapshots are deleted from a system. As snapshots are deleted, the snapshot space is reclaimed from a TPVV or fully provisioned VV and returned to the CPG for reuse by other volumes. Deleted snapshot space can be reclaimed from virtual copies, physical copies, or Remote Copy volumes.

Virtual Copy

Virtual Copies are HPE 3PAR snapshot implementation used to provide a point-in-time virtual copy of a VV and ensure that the original data can always be obtained should a problem occur when updating the data on a VV. Virtual Copies implements both an efficient variant of copy-on-write (COW) and redirect-on-write (ROW) mechanisms. For COW, the 3PAR OS uses a delayed copy-on-write (DCOW) that eliminates any performance impact to host I/O, DCOW is used for snapshots of Fully Provisioned and Thinly Provisioned volumes. HPE 3PAR delayed COW, relegates the reading of the original data, updating of the base volume with the new data, and the copy of the original data to a background process, after the write update has been acknowledge to the host.

For Thinly Deduped Virtual Volumes (TDVVs), the HPE 3PAR OS relies on the redirect-on-write to create snapshots and store the original data. This process allow the HPE 3PAR to use a single write I/O at most to store the host I/O to an existing block of data to the base TDVV. Furthermore, the in-line dedupe engine will check if the new write if dedupable through the fast and efficient metadata lookup. In case of a match against any data block in already in the dedupe store, then only metadata update will take please and I/O are send to the back end.

Virtual Copies are always thin, non-duplicative (only one copy of the change blocks if kept and then referenced) and reservationless, they can also have variable quality of service as the user can specify a different CPG for snapshot data; for example user data (SD) on SSD and snapshots data (SD2) on a more cost efficient SAS tier. Snapshot of Thinly Deduped Virtual Volumes (TDVVs) that reside on an SSD tier will deduplicate snap data.

Thanks to efficient metadata handling, users can configure thousands of read-only and read-write snapshots and enables flexible management as promotion from any snapshot without destroying other snapshots.

Autonomic storage management

The HPE 3PAR Operating System helps simplify, automate, and expedite storage management by handling provisioning, tiering, and change management autonomically and intelligently, at a subsystem level, and without administrator intervention.

The system’s user interfaces have been developed to offer autonomic administration, which means that the interfaces allow an administrator to create and manage physical and logical resources without requiring any overt action. Provisioning does not require any pre-planning, yet the system constructs volumes intelligently based on available resources, unlike manual provisioning approaches that require planning and the manual addition of capacity to intermediary pools.

Ease of use

The HPE 3PAR Operating System reduces training and administration efforts through the simple, point-and-click HPE 3PAR StoreServ Management Console (SSMC) and the scriptable HPE 3PAR Command Line Interface (CLI). Both management options provide uncommonly rich instrumentation of all physical and logical objects for one or more storage systems, thus eliminating the need for the extra tools and consulting often required for diagnosis and troubleshooting.

Open administration support is provided via SNMP, Storage Management Initiative Specification (SMI-S), and Web Service API.

Provisioning is managed intelligently and autonomically. Massively parallel and fine-grained striping of data across internal resources assures high and predictable service levels for all workload types. Service conditions remain high and predictable as system use grows or in the event of a component failure, while traditional storage planning, change management, and array-specific professional services are eliminated.

Managing performance is paramount to ensuring critical applications have the resources needed at the times of peak demands. These resources rely on a tool or tools to report metrics in a timely and accurate fashion. The monitoring of performance allows the storage administrator the flexibility to view real time data as well as performance trends by using historical data stored within a database. The tool which is used with HPE 3PAR arrays is called System Reporter.

HPE 3PAR Performance Insights eliminate the guesswork and time out of diagnosing bottlenecks and optimizing application performance. With Performance Insights, customers are provided automated details that show the root cause to complex anomalies with application-aware insights eliminating tuning of infrastructure with painful trial-and-error methods.

Available in HPE SSMC 3.4, Performance Insights offers machine-learned algorithm that are trained in the cloud and deployed on-premises. This capability enables a fast time-to-response and extends HPE InfoSight to dark sites. Performance Insights offer the following benefits:

- Identify when performance issue is due to saturation
- Plan workloads better with knowledge of available headroom
- Identify root cause to complex anomalies with application-aware insights

HPE InfoSight Integration with SSMC enables you to get access to predictive analytics done in HPE cloud. This enables you to know about problems, such as data availability or data loss issues, and performance degradation that are likely to occur shortly by having the HPE InfoSight cloud based service analyze the HPE Storage system logs. The HPE InfoSight service does this by running machine learning algorithms and community benchmarks to arrive at various insights that are made available to the SSMC storage administrator to act upon. The repertoire of intelligence grows in HPE InfoSight and new anomalies (signatures) are detected gradually, making it a powerful tool in Storage Administration.

HPE 3PAR Autonomic Groups takes autonomic storage management a step further by allowing both hosts and VVs to be combined into groups or sets that can then be managed as a single object. Adding an object to an autonomic group applies all previously performed provisioning actions to the new member. For example, when a new host is added to a group, all volumes that were previously exported to the group are autonomically exported to the new host with absolutely no administrative intervention required. Similarly, when a new volume is added to a group, this volume is also autonomically exported to all hosts the group has previously been exported to—intelligently and with no administrator action required.

In fact, management of the HPE 3PAR Storage system requires only knowledge of a few simple, basic functions:

- Create (for VVs and LDs)
- Remove (for VVs and LDs)
- Show (for resources)
- Stat (to display statistics)
- Hist (to display histograms)

Although there are a few other functions, these commands represent 90 percent of the console actions necessary, returning simplicity to the storage environment. Both the CLI and the HPE 3PAR Management Console communicate with the corresponding server process on the HPE 3PAR Storage system over TCP/IP over the on-board Gigabit Ethernet port on one of the controller nodes.

Management of the HPE 3PAR Storage system benefits from very granular instrumentation within the HPE 3PAR Operating System. This instrumentation effectively tracks every I/O through the system and provides statistical information, including service time, I/O size, KB/s, and IOPS for VVs, LDs, and physical drives (PDs). Performance statistics such as CPU utilization, total accesses, and cache hit rate for reads and writes are also available on the controller nodes that make up the system cluster.

These statistics can be reported through the HPE 3PAR Management Console or through the CLI. Moreover, administrators at operation centers powered by the leading enterprise management platforms can monitor MIB-II information from the HPE 3PAR Storage system. All alerts are converted into SNMP Version 2c and Version 3 traps and sent to any configured SNMP management station.

An even more powerful and flexible way to manage HPE 3PAR Storage systems than through the CLI and the HPE 3PAR Management Console is through the use of the HPE 3PAR Web Services API. This API enables programmatic management of HPE 3PAR Storage systems. Using the API, the management of volumes, CPGs, and VLUNs can be automated through a series of HTTPS requests. The API consists of a server that is part of the HPE 3PAR Operating System and runs on the HPE 3PAR Storage system itself and a definition of the operations, inputs, and outputs of the API. The software development kit (SDK) of the API includes a sample client that can be referenced for the development of customer-defined clients.

HPE's commitment to the OpenStack community brings the power of OpenStack to the enterprise with new and enhanced offerings that enable enterprises to increase agility, speed innovation, and lower costs. HPE has been a top contributor to the advancement of the OpenStack project. HPE's contributions have focused on continuous integration and quality assurance, which support the development of a reliable and scalable cloud platform that is equipped to handle production workloads. To support the need that many larger organizations and service providers have for enterprise-class storage, Hewlett Packard Enterprise has developed the HPE 3PAR Block Storage Drivers, which support the OpenStack technology across both iSCSI and Fibre Channel (FC) protocols. This provides the flexibility and cost-effectiveness of a cloud-based open source platform to customers with mission-critical environments and high resiliency requirements.

For more information about HPE 3PAR support for OpenStack, please see the OpenStack HPE 3PAR Block Storage Driver Configuration Best Practices.

SmartSAN

SAN plays a critical role in any data center by providing access and connectivity between storage arrays and servers via a dedicated network. Fibre Channel (FC) has been the dominant storage protocol that enjoys significant SAN market share. FC is popular for storage because of its enterprise-class performance, availability, and security. Fibre Channel zoning is a key feature that adds to security and better management of the SAN by providing necessary segregation and allowing controlled communication among selected devices within a large fabric. However, configuring zones still is a complex, tedious, and error-prone operation in a majority of SAN installations. Thus, signifying a need for automating these operations as much as possible to avoid human errors and reduce potential SAN downtime.

HPE SmartSAN for HPE 3PAR comes with a set of innovative features, one of which is automated zoning to address the above issues. In addition, it also supports standards-based device registrations and diagnostic data collection for better configuration, visibility, and diagnostic purposes. Automated zoning as implemented on HPE 3PAR as part of HPE SmartSAN 1.0, utilizes peer zoning as defined in FC standards, thus empowering HPE 3PAR Array to configure zones automatically whenever hosts are provisioned on the target side.

Data optimization

HPE 3PAR storage offers several products that can be used for service-level optimization. These solutions match data to the most cost-efficient resource capable of delivering the needed service level at any given time.

Dynamic Optimization is a powerful software feature that enables storage administrators to perform several online volume optimizations:

- Online data movement of existing volumes to different RAID levels and different tiers of storage. For example, application that requires high performance only during certain windows can be tuned to RAID 1 on SSDs and during lower activity periods moved to more cost-effective RAID 6 storage on nearline disks.
- Convert existing volumes to a different volume type. For example, Thinly Provisioned volumes on a HDD CPG to a Thinly Deduped volume on an SSD tier. Or the conversion of a full volume to a thinly provisioned volume. This conversion happens transparently and non-disruptively.

Adaptive Optimization software is another autonomic storage tiering tool that takes a fine-grained, policy-driven approach to service-level optimization. HPE 3PAR Adaptive Optimization works by analyzing performance (access rates) for sub-volume regions, then selects the most active regions (those with the highest I/O rates) and uses the proven sub-volume data movement engine built into the HPE 3PAR Operating System to relocate those regions to the fastest storage tier available. It also moves less active regions to slower tiers to help ensure space availability for newly active regions.

Traditional storage arrays require storage administrators to choose between slow, inexpensive storage and fast, expensive storage for each volume—a process that depends on the knowledge of the application's storage access patterns. Moreover, volumes tend to have hotspots rather than evenly distributed accesses, and these hotspots can move over time.

Autonomic rebalancing

The HPE 3PAR Operating System automatically creates a balanced system layout by mapping VVs to many LDs, which are composed of chunklets drawn from many physical disks. As new hardware, drives, drive chassis, and nodes are added to a system, the existing data can be laid out to use the new components and benefits from system-wide striping.

The HPE 3PAR Autonomic Rebalance feature provides the ability to analyze how volumes on the HPE 3PAR Storage system are using physical disk space and makes intelligent, autonomic adjustments to help ensure better volume distribution when new hardware is added to the system. This rebalancing is achieved via the "tunesys" command.

STORAGE FEDERATION

HPE 3PAR Storage Federation is a suite of technologies that solves one of the major paint-point when it comes to tech-refresh, the act of migrating the data itself, that enables customers to move off a third-party array to 3PAR and then by streamlining the entire 3PAR asset lifecycle management and technology refresh. HPE 3PAR Federation technologies elevate data to newer, current hardware technologies—while avoiding the pain associated with "fork-lift" upgrades that require extensive planning and preparation (and may include such difficulties as application downtimes or other impactful events). And unlike other technologies, 3PAR federation gives the option to either repurpose legacy arrays, extend their value through further federation or replication, or simply retire them. Most importantly, HPE 3PAR Storage Federation is an Enterprise solution, at no moment in time it will present a SPOF (like running on a single controller) and supports online migration of volumes that are exported to clusters and/or replicated.

Storage federation is a native functionality built into the HPE 3PAR Operating System that enables users to move data and workloads between arrays without impacting applications, users, or services. Simply and non-disruptively shift data between any model of HPE 3PAR Storage system without additional overhead impact to hosts management layers or appliances. With a very broad support matrix of Host Operating Systems, seamless bidirectional data mobility on HPE 3PAR Storage systems can also improve availability in not just physical hosted environments but also Hypervisor based environments like VMware®, Microsoft Hyper-V etc.

Storage federation is the delivery of distributed volume management across self-governing, homogeneous peer storage arrays, all managed by a single management interface call HPE 3PAR Management Console. Federated data mobility allows live data to be easily and non-disruptively moved between HPE 3PAR Storage systems without any impact to host performance of any complexity of an added appliance in the datapath. This is very similar to the virtual machine mobility enabled by products like VMware vSphere® vMotion®, Microsoft Hyper-V Migration etc., but in the case of HPE 3PAR Storage Federation, data volume mobility is enabled between storage systems. Storage federation on HPE 3PAR Storage systems is peer-to-peer relationship, something that is native to the system itself.

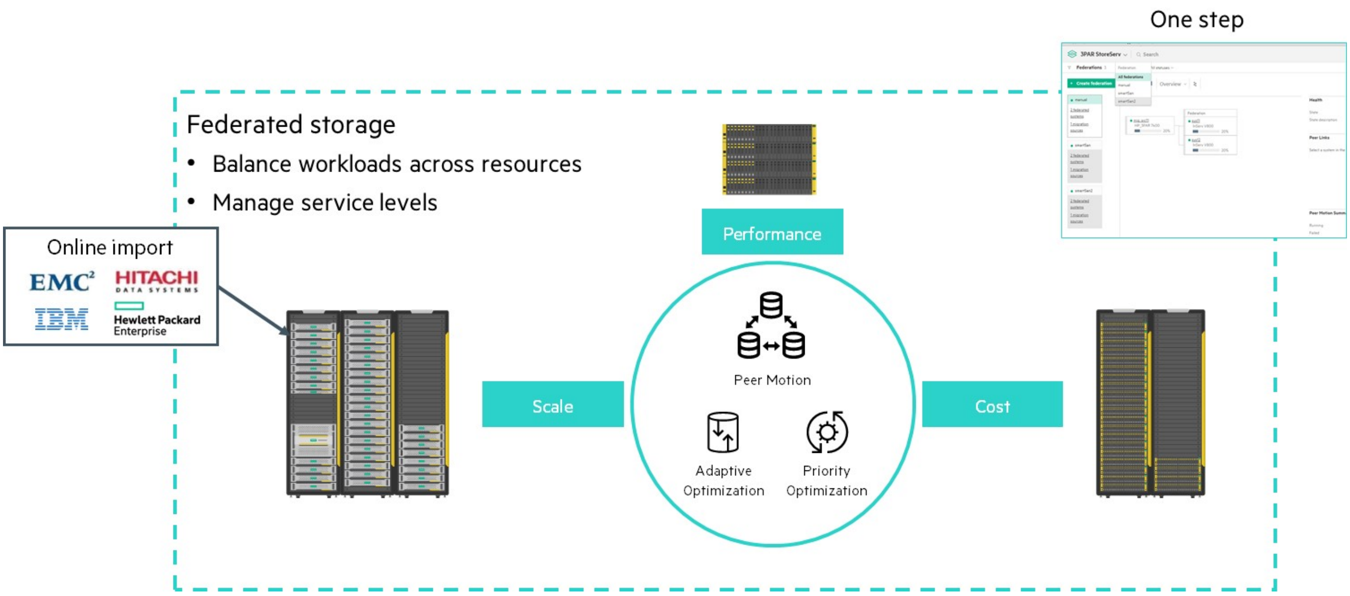


FIGURE 14. Storage Federation

NOTE
That Storage Federation is different from hierarchical virtualization, which is the delivery of consolidated or distributed volume management through appliances that hierarchically control a set of heterogeneous storage arrays. Hierarchical virtualization, also sometimes referred to as external storage virtualization, adds a new layer that has to be purchased and managed. A new layer not only introduces additional fault domains that need to be handled but also compromises the overall functionality of the entire system to be at the lowest common denominator.

In contrast, storage federation on HPE 3PAR Storage systems delivers the following benefits:

- Keeps costs low (no redundant layer of intelligent controllers)
- Reduces failure domains (no additional layers)
- Maintains functionality of each of the peers
- Simplifies administration
- Reduced complexity of Interoperability between storage systems

Storage federation has emerged as a way to address and improve storage agility and efficiency at the data center and even metropolitan area level. The primary building blocks of HPE 3PAR Storage Federation are:

- HPE 3PAR Peer Motion
- HPE 3PAR Online Import
- HPE 3PAR Peer Persistence

Peer Motion

Peer Motion software is the first non-disruptive, do-it-yourself data mobility tool for enterprise block storage that does not require any external appliance to be included in the datapath, nor does it introduce any additional overhead on the host resources. Unlike traditional block migration approaches, HPE 3PAR Peer Motion enables non-disruptive data mobility of storage volume(s) between any HPE 3PAR Storage system that are a part of the Storage Federation, without complex planning or dependency on extra tools. HPE 3PAR Peer Motion offers bidirectional data mobility which helps setup your federation upfront and create a setup with back and forth data movement between two HPE 3PAR Storage systems within the same data center.

Federated multiarray bidirectional data mobility between HPE 3PAR Storage systems is easy to implement and manage via the HPE 3PAR Management Console (SSMC). Simple predefined workflows have been implemented within the SSMC that give you the ability to move data—for example, moving all volumes associated with a host from one HPE 3PAR Storage system to another or moving individual volumes for workload balancing purposes-all with just one click.

HPE 3PAR Peer Motion software leverages the same built-in technology that powers the simple and rapid inline thin conversion of inefficient fat volumes on source arrays to more efficient, higher-utilization thin volumes on the destination HPE 3PAR Storage system.

HPE 3PAR Peer Motion software allows all HPE 3PAR Storage systems to participate in peering relationships with each other in order to provide the following flexibility benefits:

- **Federated workload balancing**—Moves workloads from over-utilized assets to underutilized ones
- **Federated asset management**—Non-disruptively adds new storage to the infrastructure or migrates data from older systems to newer ones
- **Federated thin provisioning**—Manages storage utilization and efficiency at the data center level, not the individual system level

For more information about HPE 3PAR Peer Motion, please see the Data mobility between HPE 3PAR StoreServ and third-party storage.

Online Import

Based on HPE 3PAR Peer Motion technology, HPE 3PAR Online Import software leverages federated data mobility on the HPE 3PAR Storage array to simplify and expedite data migration from non-3PAR arrays. With HPE 3PAR Online Import software, migration from these platforms can be performed in only five stages:

1. Set up the online import environment
2. Zone the host to the new system
3. Configure host multipathing

4. Unzone from the source, and start the migration
5. Validate the application

HPE 3PAR Online Import for EVA uses HPE EVA Command View as the orchestration platform to enable direct migration of data from a source HPE EVA Storage system to a destination HPE 3PAR Storage array without requiring host resources for data migration. The entire process can be completed with only minimal to no disruption (depending on host OS), sending EVA virtual disk and host configuration information to the HPE 3PAR Storage array without the need to change host configurations or interrupt data access in most cases.

For more information on HPE 3PAR Online Import for HPE EVA, see the HPE EVA P6000 to HPE 3PAR Online Import Best Practices. For more information on supported non-3PAR arrays for Online Import, see the HPE 3PAR Online Import.

MULTI-SITE RESILIENCY

HPE 3PAR Peer Persistence

HPE 3PAR Peer Persistence software enables HPE 3PAR Storage systems located within a metropolitan distance to act as peers to each other for delivering a high-availability, transparent failover solution for the connected VMware vSphere®, Microsoft Hyper-V, and Microsoft Windows clusters. HPE 3PAR Peer Persistence allows an array-level, high-availability solution between two sites or data centers where failover and failback remains completely transparent to the hosts and applications running on those hosts. Unlike traditional disaster recovery models where the hosts (and applications) must be restarted upon failover, HPE 3PAR Peer Persistence allows hosts to remain online serving their business applications, even when the serving of the I/O workload migrates transparently from the primary array to the secondary array, resulting in zero downtime.

In an HPE 3PAR Peer Persistence configuration, a host cluster would be deployed across two sites and an HPE 3PAR Storage system would be deployed at each site. All hosts in the cluster would be connected to both of the HPE 3PAR Storage systems. These HPE 3PAR systems present the same set of VVs and VLUNs with same volume WWN to the hosts in that cluster. The VVs are synchronously replicated at the block level so that each HPE 3PAR Storage system has a synchronous copy of the volume. A given volume would be primary on a given HPE 3PAR Storage system at any one time. Using Asymmetric Logical Unit Access (ALUA), HPE 3PAR Peer Persistence presents the paths from the primary array (HPE 3PAR Storage system on which the VV is primary) as "active/optimized" and the paths from the secondary array as "standby" paths. Issuing a switchover command on the array results in the relationship of the arrays to swap, and this is reflected back to the host by swapping the state of the paths from active to standby and vice versa. Under this configuration, both HPE 3PAR Storage systems can be actively serving I/O under normal operation (albeit on separate volumes).

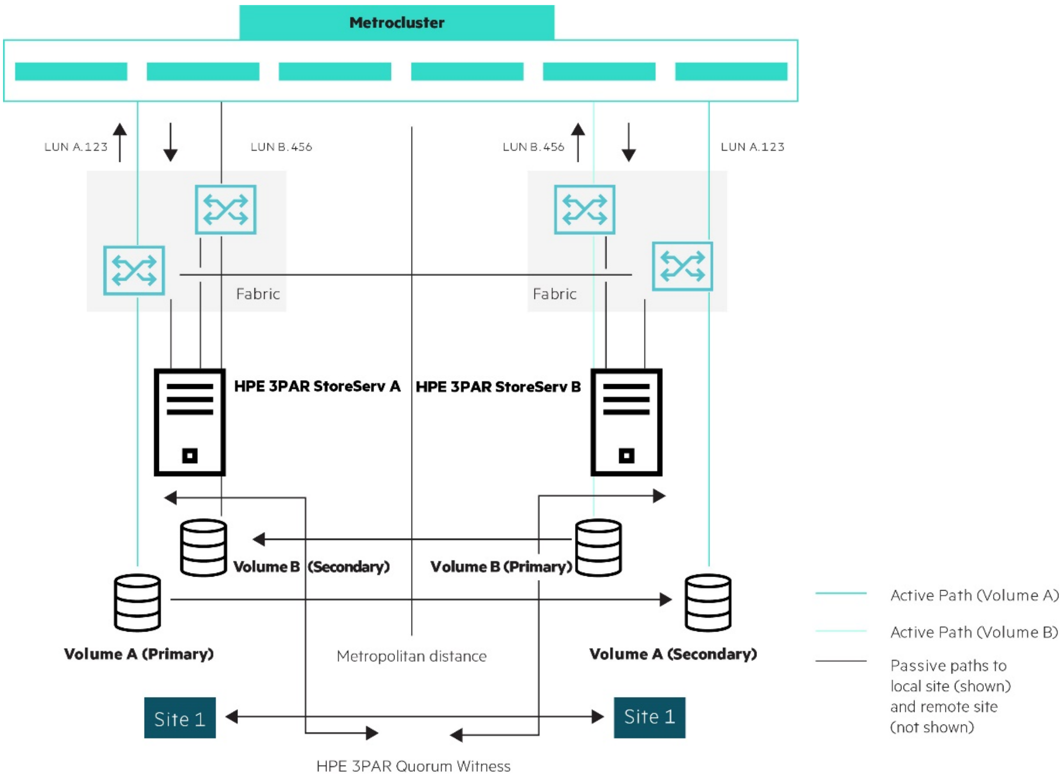


FIGURE 15. Transparent failover with HPE 3PAR Peer Persistence software

With HPE 3PAR Peer Persistence, Oracle RAC stretch cluster support is now available on Linux® (RHEL 6) over FC, FCoE, and iSCSI protocols.

HPE 3PAR 3DC Peer Persistence

HPE 3PAR 3DC Peer Persistence feature augments disaster recovery to high availability delivered by Peer Persistence, by adding a third 3PAR array as periodic asynchronous replication target to an Peer Persistence configuration.

In an HPE 3PAR 3DC Peer Persistence configuration, a volume that is replicated synchronously between the two arrays in Peer Persistence configuration is also replicated using periodic asynchronous Remote Copy replication, to a third 3PAR array on a faraway location. This way there are two exact copies of the data between the arrays in Peer Persistence relationship and a third copy of the data on the third array which might lag in time by at least 5 minutes.

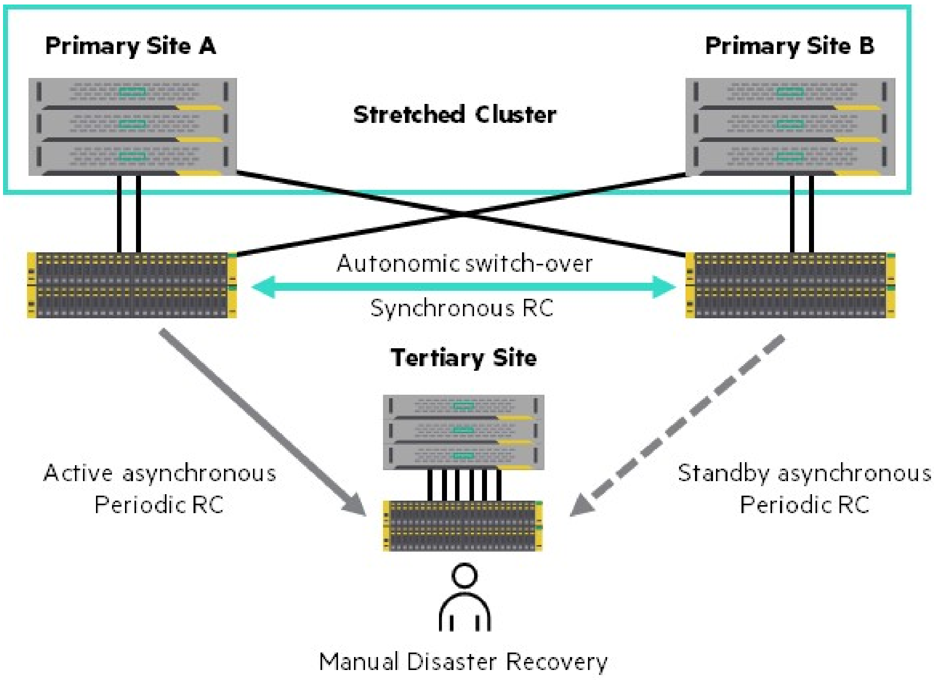


FIGURE 16. HPE 3PAR 3DC Peer Persistence

PROACTIVE SUPPORT

HPE Support for HPE 3PAR Storage provides a global support infrastructure that leverages advanced system and support architectures for fast, predictive response and remediation. The HPE 3PAR Secure Service Architecture provides secure service communication between the **HPE 3PAR Storage systems** at the customer site and HPE Support, enabling secure diagnostic data transmission and remote service connections. Key diagnostic information such as system health statistics, configuration data, performance data, and system events can be transferred frequently and maintained centrally on a historical basis. As a result, proactive fault detection and analysis are improved and manual intervention is kept to a bare minimum.

This implementation provides automated analysis and reporting that delivers accuracy and consistency, full system information in hand that reduces on-site dependencies, and fully scripted and tested automated point-and-click service actions that reduce human error.

HPE 3PAR Storage systems include a dedicated service processor, a server that monitors one or more.

HPE 3PAR systems enable remote monitoring and remote servicing of the array. The service processor is a physical server that is external to the HPE 3PAR Storage system and communicates to it via TCP/IP. A virtual service processor is available for the HPE 3PAR 8000 Storage series. Service processor with a redundant power supply is also available.

The requirement for a dedicated server outside of the storage sub-systems separates the fault domains and ensures greater reliability for the storage array without compromising serviceability. The service processor is a physical server with the necessary OS already pre-installed from the factory. It is ready-to-go racked in a factory integrated storage setup that minimizes complexity in setup, installation and usage for the customer.

The service processor functions as the communication interface between a customer's IP network and HPE by managing all service-related communications. It leverages the industry-standard Hypertext Transfer Protocol Secure (HTTPS) to secure and encrypt data for all inbound and outbound communications. The information collected via the service processor is sent to HPE. This information includes system status, configuration, performance metrics, environmental information, alerts, and notification debug logs. No customer data is sent.

The data sent is used by Hewlett Packard Enterprise support teams to proactively monitor the array and contact the customer if potential issues are discovered. Customers are warned proactively about potential problems before they occur. In the case of switch issues, the customer is advised of an issue and replacement parts are dispatched. Trained Hewlett Packard Enterprise service personnel can service the system at the customer's convenience. If the service processor cannot dial Hewlett Packard Enterprise for any reason, both the HPE 3PAR Storage system and Hewlett Packard Enterprise support centers will receive alerts.

The service processor is also used to download new patches, maintenance updates, new firmware revisions and diagnostics; it will store them and push them to the HPE 3PAR Storage system for software upgrades. If remote access is needed for any reason, the customer can configure inbound secure access for OS upgrades, patches, and engineering access. If the customer's data center does not permit "phone home" devices, then all alerts and notifications will be sent to the customer's internal support team. The customer can then notify Hewlett Packard Enterprise support of an issue or suspected issue, either over the phone or via the web.

In addition, all HPE 3PAR Storage systems support a complimentary support service known as Over-Subscribed System Alerts (OSSA) in addition to and concurrent with automated remote monitoring, alerting, and notification. This automated monitoring tool performs proactive utilization checks on key system elements utilizing data that resides at HPE. This data is collected periodically from the system and sent to Hewlett Packard Enterprise. The intent is to provide valuable information such as storage node CPU utilization, disk IOPS, the number of host initiators per port, and other metrics to keep the HPE 3PAR Storage system running optimally.

SUMMARY

In the digital era, new applications and workloads are creating a massive growth in data being created and actioned across hybrid cloud. Data is transformative only when it can be refined and accessed at the right place and at the right time, driving actionable insights into new revenue streams. However, extracting maximum value out of it is much easier said than done.

HPE 3PAR delivers intelligent storage with a tier-1 all-flash foundation to unlock the potential of your data. With HPE 3PAR, your storage is:

- AI-driven: HPE 3PAR uses advanced analytics and machine learning through HPE InfoSight to not only remove the burden of managing infrastructure, but also serves as the foundation to provide context-aware intelligence about how your data should be managed. Overall, HPE InfoSight has predicted and auto-resolved 85% of more than thousands of complex, priority cases across the HPE 3PAR installed base.
- Built for cloud: HPE 3PAR applies intelligence to see, manage, and automate your storage no matter where your data lives. For example, powerful toolsets are available to automate and manage your HPE 3PAR for the cloud, DevOps, virtualization, and container environments. You can also effortlessly orchestrate intelligent, multi-tiered data protection from on-premises arrays to the public cloud—driven by policy and business need.
- As-a-Service experience. HPE 3PAR delivers the flexibility to align to your specific consumption and investment needs. With as-a-Service, on-premises storage solutions from HPE GreenLake, you get scalability and simplified IT operations—even operated for you by HPE—all in a pay per use model. Eliminate overprovisioning to save up to 30% on storage cost, deploy workloads when they are needed, and free up your staff to focus on core business initiatives.

HPE 3PAR Storage does all this with a Tier-1 all flash foundation to support your mission-critical applications and beyond. Built to meet the extreme requirements of massively consolidated cloud service providers, HPE 3PAR enables you to confidently consolidate mixed and unpredictable workloads with ease. All HPE 3PAR models are built on a single flash-optimized architecture, run the exact same HPE 3PAR Operating System, and offer a common set of enterprise data services. Get ready for anything by starting small and scaling big with HPE 3PAR Storage.

For more information

For detailed and up-to-date specifications on each of these products, please refer to the product QuickSpecs:

- HPE 3PAR 20000 Storage QuickSpecs
- HPE 3PAR 8000 Storage QuickSpecs
- HPE 3PAR 9000 Storage QuickSpecs
- HPE 3PAR Software Products QuickSpecs

LEARN MORE AT

hpe.com/storage/3parstoreserv

4AA3-3516ENW, December 2019, Rev. 11