The HP 3PAR Architecture

Returning simplicity to IT infrastructures

Technical white paper

Table of contents

Introduction

IT managers today face ever-evolving IT requirements. They need to leverage business information, consolidate storage assets, and support measurable service levels while dealing with the old problems of mushrooming corporate data and a shortage of skilled storage specialists. Traditional storage solutions have not effectively adapted to the new IT requirements that have evolved over the last decade or more. As a result, companies have had to add layers of hardware and software to meet their needs—a costly and complex proposition. IT managers need a solution that can bring simplicity and efficiency back to the storage infrastructure.

HP 3PAR Utility Storage is the only virtualized storage platform that delivers 100% of the simplicity, efficiency, and agility demanded by today's virtual and cloud data centers. Designed from the ground up to exceed the economic and operational requirements of today's most demanding IT environments, HP 3PAR Utility Storage also delivers the performance, scalability, and availability required of Tier 1 Storage along with unique technology benefits not available with traditional platforms.

The HP 3PAR Storage System family is the hardware foundation of HP 3PAR Utility Storage. Unlike modular and monolithic (or cache-centric) storage arrays, HP 3PAR Storage Systems use a cluster-based approach and feature fourth-generation HP 3PAR Thin Built In ASICs in each clustered Controller Node. The modularity of the system delivers a single HP Converged Storage platform that scales continuously from the small to the very large and offers complete fault tolerance of both hardware and software as part of an HP Converged Infrastructure.

HP 3PAR Software, with the HP 3PAR InForm Operating System (InForm OS) as its foundation, is the intelligence behind HP 3PAR Utility Storage. The HP 3PAR InForm OS has advanced capabilities that provide:

- Fine-grained virtualization and "wide striping" capabilities that deliver massively parallel performance levels as well as the flexibility to configure various levels of service
- Industry-leading, pioneering thin technologies for efficiency and capacity reduction
- Sophisticated resiliency features to protect against hardware, software, and site failures
- Storage federation capability to enable seamless migration of data and workloads between arrays without impact to applications, users, or services
- Uncompromising security, including secure workload segregation to enable multi-tenancy
- Autonomic management to eliminate manual, repetitive, and error-prone administrative tasks and deliver automatically load-balanced storage

This white paper provides an overview of the HP 3PAR Architecture, including system hardware and software.

Figure 1: HP 3PAR Storage Systems

HP 3PAR Architecture overview

The HP 3PAR Architecture, the foundation of HP 3PAR Utility Storage, combines best-in-class, open technologies with extensive innovations in hardware and software design. Each HP 3PAR Storage System features a high-speed, full mesh, passive system backplane that joins multiple Controller Nodes (the high-performance data movement engines of the HP 3PAR Architecture) to form a cache-coherent, Mesh-Active cluster. This low-latency interconnect allows for tight coordination among the Controller Nodes and a simplified software model.

Within this architecture, Controller Nodes are paired via Fibre Channel connections from each Node in the pair to the dual-ported Drive Chassis (or Drive Cages) owned by that pair. In addition, each Controller Node may have one or more paths to Hosts (either directly or over a Storage Area Network, or SAN). The clustering of Controller Nodes enables the system to present to Hosts a single, highly available, high-performance storage system.

Volume management software on the Controller Nodes allows users to create Virtual Volumes (VVs), which are then exported and made visible to hosts as Logical Unit Numbers (LUNs). Within the system, VVs are mapped to one or more Logical Disks (LDs), which implement RAID functionality over the raw storage in the HP 3PAR Storage System's physical drives (PDs). Because the cluster of Controller Nodes presents itself to hosts as a single system, servers can access VVs over any host-connected Fibre Channel port—even if the physical storage for that data (on the PDs) is connected to a different Controller Node. This is achieved through extremely low-latency data transfer across the high-speed, full-mesh backplane.

The HP 3PAR Architecture is currently available in six different HP 3PAR Storage System models to meet customer scaling requirements. The mid-range F-Class storage systems (F400 and F200) are a scaled-down implementation of the same architecture as the high-end T-Class storage systems (T800 and T400), while the new HP P10000 3PAR Storage Systems (V800 and V400) extend the high end with double the capacity and 2.6 times the bandwidth of the T-Class. The V800 and T800 models accommodate up to eight Controller Nodes; the V400, T400, and F400 accommodate up to four Controller Nodes; and the F200 has two Controller Nodes. Examples in this paper are based on the specifications of the V800 unless otherwise noted.

The HP 3PAR Architecture is modular and can be scaled from 4.8 to 1,600 TB, making the system deployable as a small, remote or very large, centralized system. Until now, enterprise customers were often required to purchase and manage at least two distinct architectures to span their range of cost and scalability requirements.

The high performance and scalability of the HP 3PAR Architecture is well suited for large or high-growth projects, consolidation of mission-critical information, demanding performance-based applications, and data lifecycle management and the ideal platform for virtualization and cloud computing environments.

An HP 3PAR V800 Storage System offers peak internal bandwidth of 112 gigabytes per second (GB/s), significantly more than is required by today's Controller Node implementations. The bandwidth and latencies of the HP 3PAR Architecture exceed bus, switch, and even Infiniband-based architectures.

In every HP 3PAR V800 or V400 Storage System, each Controller Node has a dedicated link to each of the other Nodes which operates at 2 GB/s in each direction, roughly eight times the speed of 4 Gb/s Fibre Channel. In an HP 3PAR V800, a total of 28 of these links form the array's full-mesh backplane.

High availability is also built into the HP 3PAR Architecture. Unlike other approaches, the system offers both hardware and software fault tolerance by running a separate instance of the InForm OS on each Controller Node, ensuring the availability of customer data. With this design, software and firmware failures—a significant cause of unplanned downtime in other architectures—are greatly reduced.

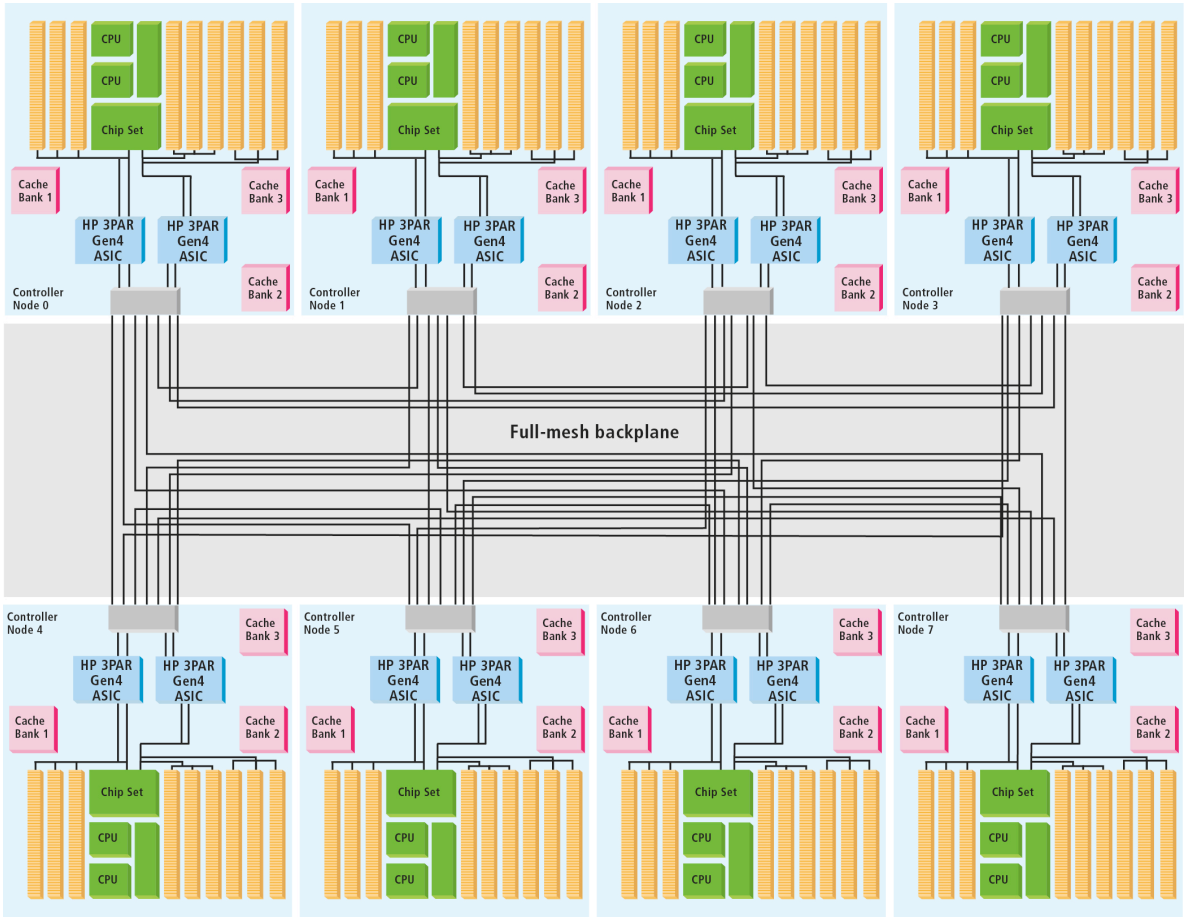
The HP 3PAR Thin Built In ASICs feature a uniquely efficient, silicon-based zero-detection mechanism that gives HP 3PAR Storage Systems the power to remove allocated but unused space without impacting performance. The HP 3PAR ASICs also deliver mixed-workload support to alleviate performance concerns 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 HP 3PAR Storage Systems boost virtual machine density so you can cut physical server purchases in half.

Storage federation capability introduced in the latest version of the InForm OS and enabled with HP 3PAR Peer Motion Software meet the needs of the Instant-On Enterprise by enabling customers to move data and workloads between arrays without impact to applications, users, or services. With this innovation, customers can simply and non-disruptively shift data between any model HP 3PAR Storage System without additional management layers or appliances.

Full-mesh controller backplane

Backplane interconnects within servers have evolved dramatically over the last ten years. Recall that most if not all server and storage array architectures employed simple bus-based backplanes for high-speed processor, memory, and I/O communication. With the growth of SMP-based servers came a significant industry investment in switch architectures, which have since been applied to one or two enterprise storage arrays. The move to a switch from buses 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. However, HP 3PAR Utility Storage System arrays represent the first storage platform to apply this interconnect. This design has been incorporated into HP 3PAR Storage Systems to reduce latencies and address scalability requirements.

Figure 02: Full-mesh backplane



The HP 3PAR Storage System backplane is a passive circuit board that contains slots for Controller Nodes. Each Controller Node slot is connected to every other Controller Node slot by a high-speed link (2 GB/s in each direction, or 4 GB/s total), forming a full-mesh interconnect between all Controller Nodes in the cluster. There are two HP P10000 3PAR backplane types: a 4-Node backplane (V400 model) that supports 2 or 4 Controller Nodes and an 8-Node backplane (V800 model) that supports 2 to 8 Controller Nodes. In addition, a completely separate full-mesh network of RS-232 serial links provides a redundant low-speed channel of communication for exchanging control information between the Nodes in the event of a failure of the main links.

Advantages of a tightly-coupled, clustered architecture

All HP 3PAR Storage Systems feature a unique Mesh-Active controller technology as part of a next-gen architecture designed for virtual and cloud data centers. This architecture combines the benefits of monolithic and modular architectures while eliminating price premiums and scaling complexities.

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 mesh controller in the system. This design delivers robust, load-balanced performance and greater headroom for cost-effective scalability, overcoming the tradeoffs typically associated with modular and monolithic storage. The high-speed, full-mesh, passive system backplane joins multiple Controller Nodes to form a cache-coherent, active-active cluster that represents the next generation of Tier 1 Storage.

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) is challenging because active processing elements not only have to be implemented redundantly, but they are 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 advantages of a central global cache (e.g., LUNs can be coherently exported from multiple ports). However, they typically must bear 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, since each element is multi-function in design—handling host, cache, and drive management processes. The tradeoff for this cost-effectiveness is the cost or complexity of scalability. Since 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 tradeoff 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. Modular designs typically use interconnect technologies that are not optimized for clustering (e.g., Fibre Channel or Ethernet) and are therefore not well suited to provide the bandwidth and latencies required for truly active-active processing.

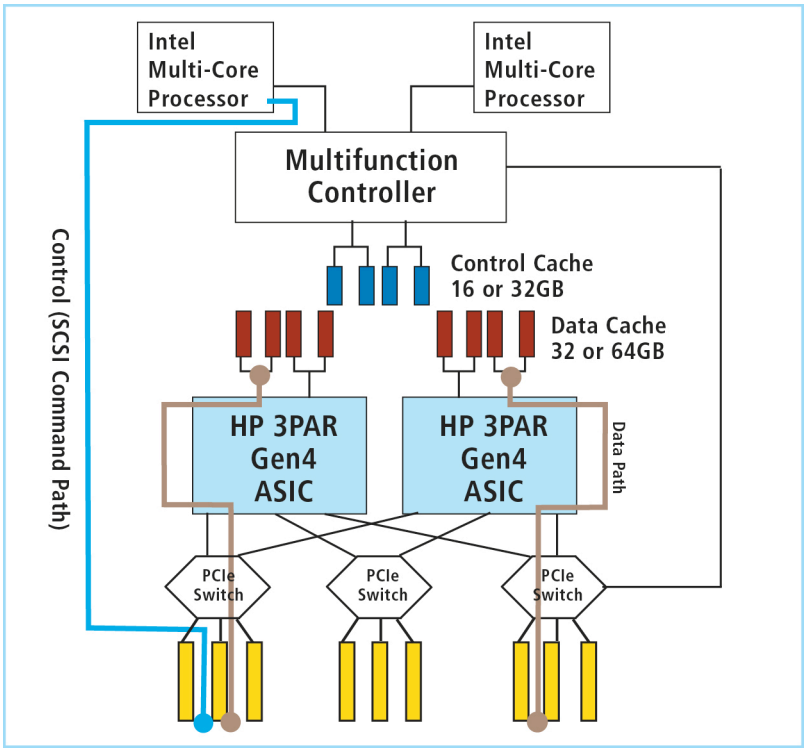
The HP 3PAR Architecture was designed to provide cost-effective, single-system scalability through a cache-coherent, multi-node, clustered implementation. This architecture begins with a multi-function node design and, like a modular array, requires just two initial Controller Nodes for redundancy. However, unlike traditional modular arrays, an optimized interconnect is provided between the Controller Nodes to facilitate Mesh-Active processing. With Mesh-Active controllers, volumes are not only active on all controllers, but they are autonomically provisioned and seamlessly load-balanced across all systems resources to deliver high and predictable levels of performance. The interconnect is optimized to deliver low latency, high-bandwidth communication and data movement between Controller Nodes through dedicated, point-to-point links and a low overhead protocol which features rapid inter-node messaging and acknowledgement.

For scalability beyond two Controller Nodes, the backplane interconnect accommodates more than two Nodes (up to eight in the case of the HP P10000 3PAR V800 Storage System). Of critical importance is that, while the value of this interconnect is high, its cost is relatively low. Because it is passive and consists of static connections embedded within a printed circuit board, it does not represent a large cost within the overall system and only one is needed. Through these innovations, the HP 3PAR Architecture can provide the best of traditional modular and monolithic designs in addition to massive load balancing.

Controller Nodes

An important element of the HP 3PAR Architecture is the Controller Node, a proprietary and powerful data movement engine designed for mixed workloads. Controller Nodes deliver performance and connectivity within the HP 3PAR Storage System. A single system can be modularly configured as a cluster of two to eight of these Nodes. Customers can start with two Controller Nodes in a small, "modular array" configuration and grow incrementally to eight Nodes in a non-disruptive manner—providing powerful flexibility and performance.

Figure 03: HP P10000 3PAR V800 and V400 Controller Node design



This modular approach provides flexibility, a cost-effective entry footprint, and affordable upgrade paths for increasing performance, capacity, connectivity, and availability as needs change. The system can withstand an entire Controller Node failure without data availability being impacted, and each Node is completely hot-pluggable to enable online serviceability.

Mixed workload support

Unlike legacy architectures that process I/O commands and move data using the same processor complex, the HP 3PAR Controller Node design separates the processing of control commands from data movement. This innovation eliminates the performance bottlenecks of existing platforms when serving competing workloads like OLTP and data warehousing simultaneously from a single processing element. Within each HP P10000 3PAR Storage System, control operations are processed by up to 16 high-performance Intel® Quad-Core processors (for an 8-Node V800 system), with dedicated control cache up to 256 GB. All data movement is handled by the specially designed HP 3PAR Gen4 ASICs (two per Controller Node), and dedicated data cache of up to 512 GB.

Abundant, multi-protocol connectivity

For host and back-end storage connectivity, each HP 3PAR Controller Node is equipped with nine high-speed I/O slots (72 slots system-wide on a V800). This design provides powerful flexibility to natively and simultaneously support adapters of multiple communication protocols. Fibre Channel and/or iSCSI TOE adapters can be configured as desired on each Controller Node for multi-protocol host connectivity. In addition, embedded Gigabit Ethernet ports can be configured for remote mirroring over IP, eliminating the incremental cost of purchasing Fibre Channel-to-IP converters. All back-end storage connections use Fibre Channel.

Using quad-ported Fibre Channel adapters, each Controller Node can deliver a total of 36 ports for a total of up to 288 ports system-wide, subject to the system's configuration. On a V800, up to 192 of these ports may be available for Host connections, providing abundant connectivity. 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.

Leveraging of commodity parts

The HP 3PAR 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 HP 3PAR Gen4 ASICs add crucial bandwidth and communication optimizations without limiting the ability to use industry-standard parts for other components.

Bandwidth and communication optimization

As previously mentioned, each V800 or V400 Controller Node contains two high-performance, proprietary HP 3PAR Gen4 ASICs optimized for data movement between three I/O buses, a three memory-bank Data Cache, and seven high-speed links to the other Controller Nodes over the full-mesh backplane. These ASICs perform parity calculations (for RAID 5 and RAID MP/Fast RAID 6) on the Data Cache, and calculates the CRC Logical Block Guard used by the T10 Data Integrity Feature (DIF) to validate data stored on drives.[1] An HP 3PAR V800 Storage System with 8 Controller Nodes has: 16 ASICs, totaling 112 GB/s of peak interconnect bandwidth and 24 I/O buses totaling 96 GB/s of peak I/O bandwidth.

ASIC-based thin storage

The HP P10000 3PAR Controller Nodes in the V800 and V400 feature Thin Built In technology unique to HP 3PAR Utility Storage and also available in the HP 3PAR T-Class and midrange HP 3PAR F-Class arrays. The HP 3PAR Gen4 ASICs in the V800 and V400 and the HP 3PAR Gen3 ASICs in the T- and F-Class arrays feature a fat-to-thin volume conversion algorithm that is built into silicon. This built-in, fat-to-thin processing capability works with HP 3PAR Software to enable users to take "fat" provisioned volumes on legacy storage and convert them to "thin" provisioned volumes on the system, inline and non-disruptively. During this process, allocated-but-unused capacity within each data volume is initialized with zeros. The Gen4 and Gen3 ASICs use built-in zero detection capability to recognize and virtualize blocks of zeros "on the fly" to drive these conversions while maintaining high performance levels.

Power failure handling

Each Controller Node includes a local physical drive that contains a separate instance of the InForm OS 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 a string of two batteries. Each battery has sufficient capacity to power the Controller Nodes long enough to save all necessary data in memory into the local physical drive.

Although many architectures use "cache batteries," these are not suitable for long downtimes usually associated with natural disasters and unforeseen catastrophes. The HP 3PAR Storage System's Controller Node battery configuration also eliminates the need for expensive batteries to power all of the system's Drive Chassis. Note that, since all write-cached data is mirrored to another Controller Node, a system-wide power failure would result in saving cached write data in the local drives of two Nodes. Since each Node's dual power supplies can be connected to separate AC power cords, providing redundant AC power to the system can reduce the possibility of an outage due to AC power failure.

A common problem with many battery-backup systems is that it is often impossible to be sure that a battery is charged and working. To address this problem, the Controller Nodes in HP 3PAR Storage Systems are each backed by a string of at least two batteries. Batteries are periodically tested by discharging one battery while the other remains charged and ready in case a power failure occurs while the battery test is in progress. The InForm OS 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 long enough to save the data to the local drive.

Data transfer paths

Figure 04 shows an overview of data transfers in an HP 3PAR Storage System with two simple examples: a write operation from a host system to a RAID 1 volume (arrows labeled W1 through W4), and a read operation (blue arrows 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 Controller Node across the high-speed backplane link so that the write data is not lost, even if the first Controller Node experiences a failure. Only after this cache mirror operation is completed is the host's write operation acknowledged.
- W3, 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 normally mirrored to one of the Controller Nodes that owns the drives to be written (D1 and D1' in this example). If the host's write (W1) was to one of these Controller Nodes, then the data would be mirrored to that Controller Node's partner.

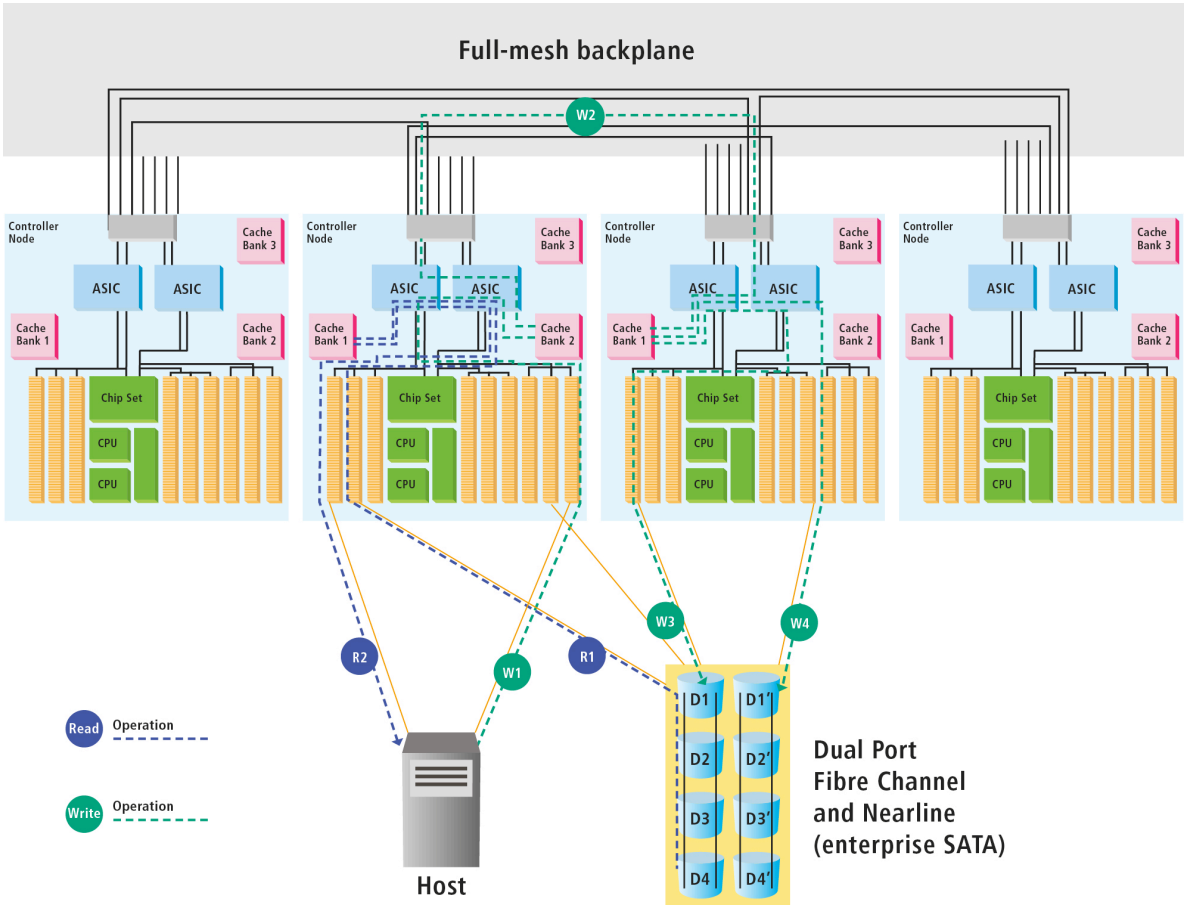
Persistent Cache allows a Node to mirror the write data to a Node that does not have direct access to drives D1 and D1' in the event of a failure of the partner Node.

[1] T-Class and F-Class servers use a single HP 3PAR Gen3 ASIC per Controller Node, which is similar but lacks the T10 DIF support.)

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 04: 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 example data transfers illustrate, 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 again a significant bottleneck in traditional architectures, and is also a valuable resource in the Controller Nodes. Unique to the system, Controller Node data transfers do not consume any of that bandwidth. This frees the processors to perform their control functions far more effectively. All RAID parity calculations are performed by the HP 3PAR Thin Built In ASICs directly on cache memory and do not consume processor or processor-memory bandwidth.

Drive Chassis

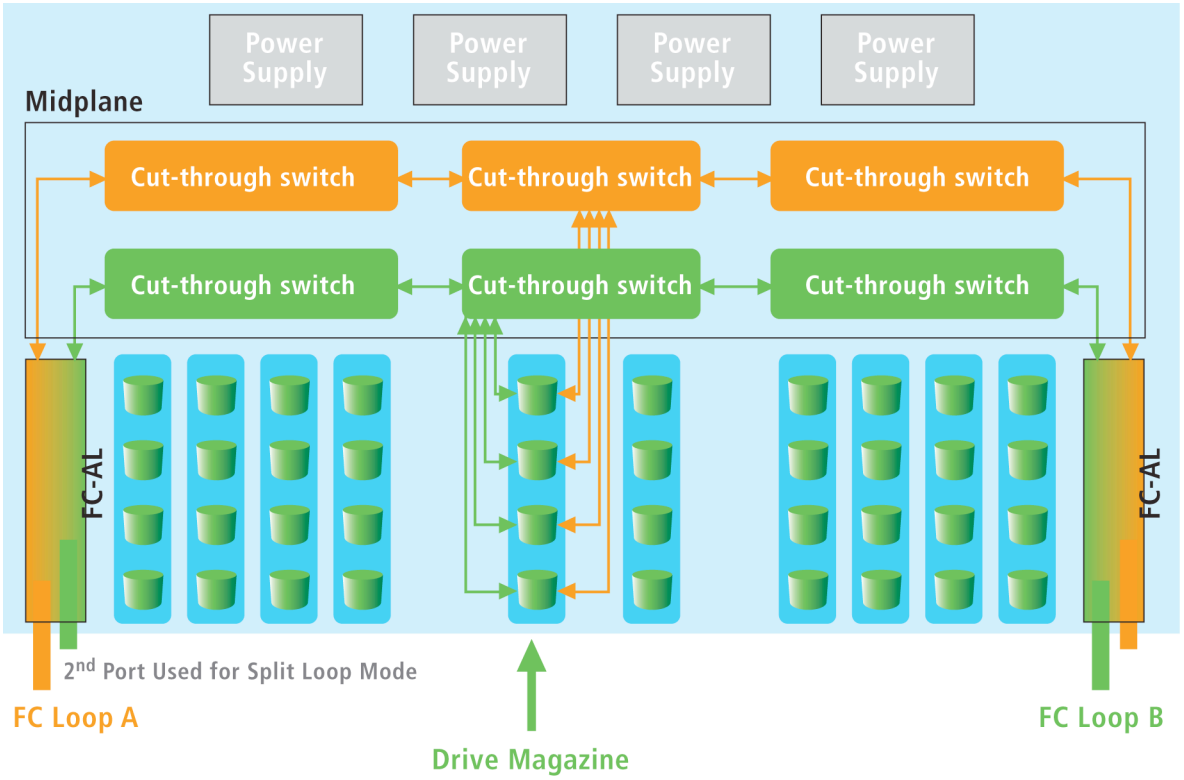
Another element of the HP 3PAR Architecture is the Drive Chassis. Drive Chassis, also referred to as Drive Cages, are intelligent, switched, hyper-dense drive enclosures that serve as the capacity building block within an HP 3PAR Storage System. A single HP P10000 3PAR V800 Storage System can accommodate up to 48 Drive Chassis and scale from 16 to 1,920 drives online and non-disruptively.

Industry-leading density

Drive Chassis have a compact, dense design. Each Drive Chassis consumes four EIA rack units in a 19-inch rack. Each Drive Chassis can be loaded with ten drive magazines holding four one-inch high drives. Because each Drive Chassis can hold up to 40 drives, a single Drive Chassis can pack up to 80 TB of data in just seven inches of rack space when using 2-TB Nearline (enterprise SATA)

disk drives. With its compact Drive Chassis design, the system platform delivers roughly 2x greater density than leading alternatives.

Figure 05: Switched drive chassis



Redundant, hot-pluggable components

Drive Chassis include redundant and hot-pluggable components. Each Drive Chassis includes N+1 redundant power supplies, redundant FC-AL adapters that provide up to four independent, 4 Gb/s, full-bandwidth Fibre Channel ports, and redundant cut-through switches on the midplane for switched point-to-point connections. Drive Magazines hot plug from the front of the system into the midplane. Redundant power supply/fan assemblies hot plug into the rear of the midplane. Each Fibre Channel drive is dual ported and accessible from redundant incoming Fibre Channel connections in an active-passive mode.

The Drive Chassis components—power supplies, Fibre Channel Adapters, and drive magazines—are serviceable online and are completely hot pluggable. Should the drive chassis midplane fail, while it is being serviced, partner Cage or Cages will continue to serve data for those volumes which were configured and managed as "High Availability (HA) Cage" volumes. If the "HA Cage" configuration setting is selected at volume creation, the 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.

Mixed physical drive types

Each Drive Chassis may contain one or more physical drive types:

- Solid State Drives (SSDs) to meet even the most stringent performance demands

- Fibre Channel disk drives to meet high performance or capacity demands
- Nearline (enterprise SATA) disk drives to meet capacity demands at the lowest cost

The HP 3PAR Architecture easily accommodates a mix of physical drive types and sizes within a single Drive Chassis. This unique flexibility eliminates any incremental expense associated with purchasing and managing separate drive chassis for different drive types. Implementing and scaling a tiered storage infrastructure within a single, massively parallel system is thereby simplified.

Data integrity checking

Fibre Channel drives in the HP P10000 3PAR Storage System 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 (DIF) for each block. This value is computed by the HP 3PAR Gen4 ASIC before writing each block, and is checked when a block is read. SATA does not support 520-byte blocks, so on Enterprise SATA drives data blocks are logically grouped with an extra block to store the CRC values.

Advanced fault isolation

Advanced fault isolation and high reliability are built into the HP 3PAR Storage System. The Drive Chassis, Drive Magazines, and physical drives themselves all report and isolate faults. A drive failure will not take all drives offline. The HP 3PAR Storage System constantly monitors drives via the Controller Nodes and Chassis and isolates faults to individual drives, then "off lines" only the failed component.

Physical Drives, chunklets, and Drive Cage firmware

As mentioned, each Physical Drive (PD) is divided into chunklets of 256 MB in size. LD allocation refers to chunklets rather than entire physical drives. This allows great flexibility in LD allocation and permits the following:

- Drives of different sizes to be used within the same RAID sets
- Striping an LD across a large number of PDs
- Fine-grain sparing, migration, and performance data collection

The Drive Cage firmware, which runs on the midplane in each Drive Chassis, serves several functions including:

- Informing the System Manager about environmental conditions (temperature, power supply status, etc.) for the Drive Cages and Physical Drives.
- Informing the System Manager about the physical position of the PDs. This is important because the LD layout takes into consideration the location of the PDs.
- Troubleshooting and taking an offending (failing) PD offline so that other PDs are not impacted.

Logical Disks and RAID types

In the HP 3PAR Architecture, Logical Disks (LDs) implement RAID functionality. Each LD is mapped onto chunklets to implement RAID 1+0 (mirroring + striping), RAID 5+0 (RAID 5 distributed parity + striping), or RAID MP (multiple distributed parity, with striping).* The InForm OS can automatically create LDs with the desired performance, availability, and size characteristics.

Several parameters can be used to control the layout of an LD to achieve 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 Drive Cages. This ensures that a failure of an entire loop or Drive Cage will not result in any loss of data. It also ensures better peak aggregate performance since 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, each row with 6 chunklets' worth of usable data (+ 2 parity) will have a usable size of 4608 MB (256 MB/chunklet x 6 chunklets/row x 3 rows).

An LD has an "owner" and a "backup owner". The owner is the Controller Node that under normal circumstances performs all operations on the LD. If the owner fails, the backup owner takes over ownership of the LD. The owner sends sufficient log information to the backup owner so that the backup owner can take over without loss of data.

The chunklets used in an LD are preferably chosen from PDs for which the owner and backup owner are connected to the primary and secondary path (respectively) so that the current owner can directly access the chunklets.

Virtual Volumes

There are two kinds of VVs: *base volumes* and *snapshot volumes*. A base volume can be considered to be the "original" VV. In other words, it directly maps all the user-visible data. A snapshot volume is created using HP 3PAR Virtual Copy Software. When a snapshot is first created, all its data is mapped indirectly to the parent's data. When a block is written to the parent, the original block is copied from the parent to a separate 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.

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 spaces is mapped to LDs with 32 MB granularity. One or more Controller Nodes may own these LDs; thus VVs can be striped across multiple Controller Nodes for additional load balancing and performance.

Virtual Volume LUN exports and LUN masking

Virtual Volumes are only visible to a host if the VVs are exported as a Virtual LUNs (VLUNs). VVs can be exported in three ways:

- To specific hosts (set of World Wide Names or WWNs). The VV would be visible to the specified WWNs, irrespective of which port 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 their WWNs) are not known prior, or in situations where the WWN of a host cannot be trusted (host WWNs can be spoofed).
- To specific hosts on a specific port.

On the system, VVs themselves do not consume LUN numbers as they do on some systems; only VLUNs consume LUN numbers.

HP 3PAR Software

HP 3PAR Software is comprised of both HP 3PAR Storage System software as well as host-based software that runs on an end-user server.

HP 3PAR Software works with the HP 3PAR Storage System to deliver a new generation of Tier 1 Storage capabilities including next-gen storage virtualization features, ease of use benefits, advanced security features, and service-level reporting while driving down the cost of obtaining and managing enterprise storage resources.

There are three different categories of HP 3PAR Software:

- The HP 3PAR InForm Operating System Software—core software that runs on the system and delivers unique storage virtualization, virtual volume management, and RAID capabilities.
- Additional HP 3PAR Software—optional software offerings that run on the system and offers enhanced capabilities including thin storage technologies, secure partitioning for virtual private arrays, storage federation, and virtual and remote copy capabilities.
- HP 3PAR Host Software—host-based software products that enable the system platform to address the needs of specific application environments, multipathing, and historical performance and capacity management.

HP 3PAR InForm Operating System Software

The software foundation of HP 3PAR Utility Storage is the InForm OS, which utilizes advanced internal virtualization capabilities to increase administrative efficiency, system utilization, and storage performance.

The InForm OS includes the following functionality and features:

- Administration Tools. The InForm OS reduces training and administration efforts through the simple, point-and-click HP 3PAR InForm Management Console and the scriptable HP 3PAR Command Line Interface (CLI). Each interface requires only a handful of intuitive, well-supported actions or commands for complete functionality and system administration. Both management options provide uncommonly rich instrumentation of all physical and logical objects for one or more storage systems, eliminating the need for extra tools and consulting often required for diagnosis and troubleshooting. Open administration support is provided via SNMP and the Storage Management Initiative Specification (SMI-S).
- Rapid Provisioning. The InForm OS eliminates array planning by delivering instant, application-tailored provisioning through the fine-grained virtualization of lower-level components. 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 the use of the system grows or in the event of a component failure while traditional storage planning, change management, and array-specific professional services are eliminated.
- Autonomic Groups. 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 are autonomically exported to that group with absolutely no administrative intervention required. Similarly, when a new volume is added to a group, that volume is also autonomically exported to all hosts in the group—intelligently and without requiring administrator action.
- Scheduler. An HP 3PAR Scheduler also helps automate storage management, reduce administration time, and decrease the chance of administrative error. Scheduler does this by giving users full control over creation and deletion of virtual copy snapshots—a process that is now completely automated with HP 3PAR Utility Storage. When used in conjunction with Autonomic Groups, HP 3PAR Scheduler automates virtual copy snapshots across multiple boot and data volumes with full write consistency across all these different volumes
- Persistent Cache. HP 3PAR Persistent Cache is a resiliency feature built into the InForm OS that allows “always on” application and virtual server environments to gracefully handle an unplanned controller failure. Persistent Cache eliminates the substantial performance penalties associated with traditional arrays and “write-through” mode so that HP 3PAR Storage Systems can maintain required service levels even in the event of a cache or controller node failure. HP 3PAR Persistent Cache leverages the clustered architecture with its unique Mesh-Active design to preserve write-caching by rapidly re-mirroring cache to the other nodes in the cluster in the event of a failure. Persistent Cache is supported on all quad-node and larger HP 3PAR arrays, including the HP 3PAR F400 Storage

System—making HP 3PAR the only vendor to incorporate this industry-leading service level protection capability into midrange as well as high-end arrays.

- RAID MP (Multi-Parity).HP 3PAR RAID MP (Multi-Parity) introduces Fast RAID 6 technology backed by the accelerated performance and Rapid RAID Rebuild capabilities of the HP 3PAR ASIC. HP 3PAR RAID MP is supported on all HP 3PAR Storage System models and delivers extra RAID protection that prevents data loss as a result of double disk failures. RAID MP delivers this enhanced protection while maintaining performance levels within 15% of RAID 10 and with capacity overheads comparable to popular RAID 5 modes. For this reason, HP 3PAR RAID MP is ideal for large disk drive configurations—for example, Serial ATA (SATA) drives above 1 TB in capacity.
- Full Copy. Full Copy is an InForm OS feature that allows you to create point-in-time clones with independent service level parameters. Full Copy offers rapid resynchronizations and is thin provisioning-aware.
- Access Guard. Access Guard is an InForm OS feature that delivers user-configurable volume security at logical and physical levels by enabling you to secure hosts and ports to specific virtual volumes.
- Thin Copy Reclamation. An industry first, Thin Copy Reclamation keeps your storage as lean and efficient as possible by reclaiming unused space resulting from deleted virtual copy snapshots and remote copy volumes.
- LDAP support. Native support for lightweight directory access protocol (LDAP) within the InForm OS delivers centralized user authentication and authorization using a standard protocol for managing access to IT resources. With support for LDAP, HP 3PAR Utility Storage can be integrated with standard, open enterprise directory services. The result is simplified security administration with centralized access control and identity management.

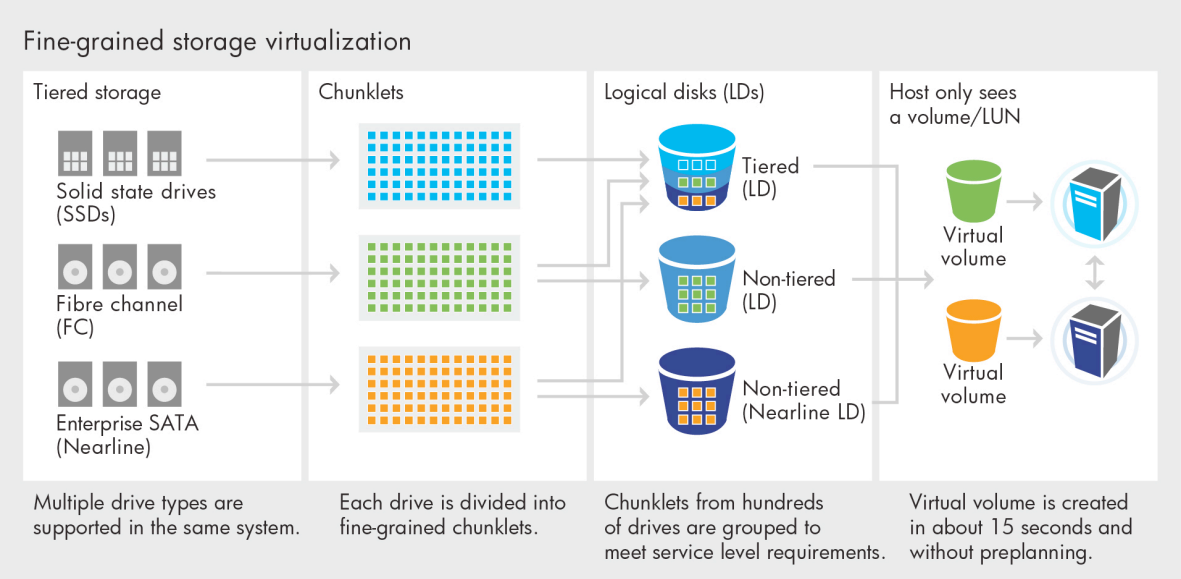
Fine-grained virtualization

To ensure performance and to maximize the utilization of physical resources, the HP 3PAR InForm OS employs a tri-level mapping methodology similar to the virtual memory architectures of the most robust enterprise operating systems on the market today. The first level of mapping virtualizes physical drives of any size into a pool of uniform-sized, fine-grained "chunklets" that are 1 GB each (256 MB on T-Class and F-Class). This level also manages the dual paths to each chunklet and physical drive. The fine-grained nature of these chunklets eliminates underutilization of precious storage assets. Complete access to every chunklet eliminates large pockets of inaccessible storage.

The fine-grained structure also enhances performance for all applications, regardless of their capacity requirements. While a small application might need only a few chunklets to support its capacity needs, those chunklets might be distributed across dozens or even hundreds of drives. Even a small application can leverage the performance resources of the entire system without provisioning excess capacity. While some platforms stop with this level of virtualization, HP 3PAR Utility Storage is just getting started.

The second level of mapping associates chunklets with Logical Disks (LDs). This association allows logical devices 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 depending on the Quality of Service (QoS) level required. In addition, the first- and second-level mappings taken together serve to parallelize work massively across physical drives and their Fibre Channel connections.

Figure 06: Virtual Volume management



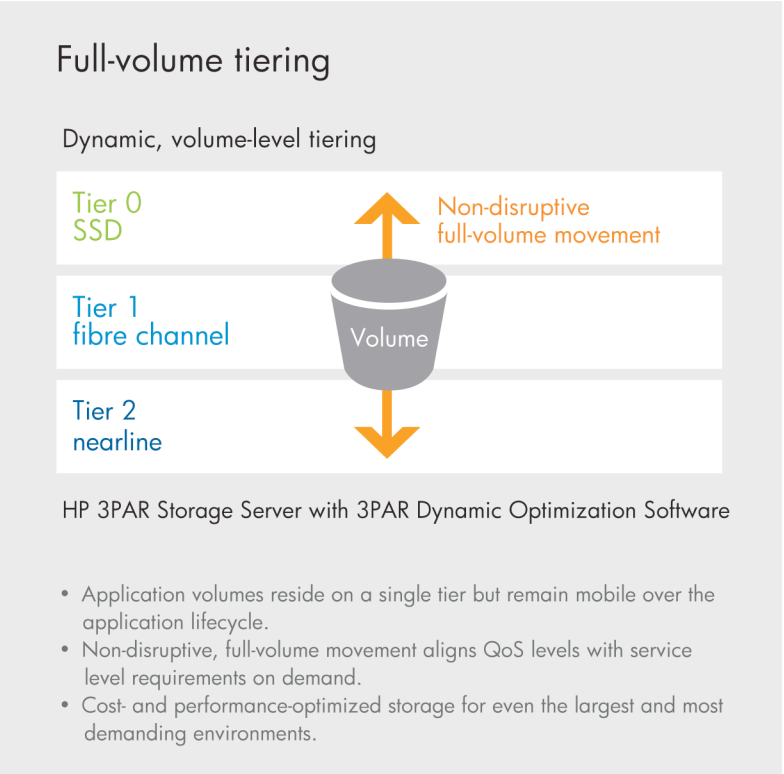
The third level of mapping associates Virtual Volumes (VVs) with all or portions of an underlying LD or of multiple LDs. Virtual Volumes are the virtual capacity representations that are ultimately exported to hosts and applications as Virtual LUNs (VLUNs) over Fibre Channel or iSCSI target ports. With the system, a VV can be coherently exported through as many or as few ports as desired. This level of mapping uses a table-based association—a mapping table with a granularity of 32 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 Virtual Volume associated with a particular LD can be quickly and non-disruptively migrated to a different LD for performance or other policy-based reasons. Other architectures require migration of the entire VV.

One-stop allocation, the general method employed by IT users for volume administration, provides for minimal planning on the part of storage administrators. By an administrator simply specifying virtual volume name, RAID level, and size, the InForm OS autonomically provisions LDs at the moment that an application requires capacity, also known as "just-in-time" provisioning.

Separation of the LD and VV layers provides benefits never thought possible based on the limits of traditional array architectures. Consider HP 3PAR Thin Provisioning Software, an additional HP 3PAR software product for the system that allows the system administrator to provision VVs several times larger than the amount of physical resources within the storage system. This methodology takes advantage of the fact that users or applications generally only fill a VV gradually over a relatively long period of time. For example, by creating and exporting 3 TBs worth of VVs but only utilizing 1 TB of LDs, an organization can dramatically increase asset utilization and defer capital expense–in some cases indefinitely.

As another example, we can consider the advanced capabilities offered by HP 3PAR Dynamic Optimization Software, another additional HP 3PAR software product for the system. Enabled by the separation of the LD and VV layers, Dynamic Optimization allows organizations to align application and business requirements with data service levels easily, precisely, and on demand. With a single command, Dynamic Optimization substitutes source LDs with new target LDs while the VV remains online and available. Data is moved from source LDs to target LDs intelligently and autonomically. In comparison, optimizing data service levels on traditional storage architectures by migrating data, usually between arrays, is prohibitively time-consuming and complex, and in many cases, is simply not done.

Figure 07: Optimization made simple



Command line interface

While the flexibility provided by the tri-level virtualization methodology of the system is enormous, management complexity is not. In fact, management of the HP 3PAR 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); and *hist* (to display histograms). Although there are a few other functions, these commands represent 90% of the console actions necessary, returning simplicity to the storage environment.

In addition to simple functions, the system's user interfaces have been developed to offer autonomic administration. That is, the interfaces allow an administrator to create and manage physical and logical resources without requiring any overt action. With the system, provisioning does not require any pre-planning yet the system constructs volumes intelligently, based on available resources. This stands in contrast to manual provisioning approaches that require planning and manual addition of capacity to intermediary pools. The InForm OS will intelligently and autonomically create the best possible VV given the available resources. This includes built-in performance and availability considerations of the physical resources to which a VV is mapped. By providing this autonomic response, HP 3PAR saves the system administrator valuable time that could be better spent managing additional terabytes and projects. VV creation requires only two steps, as opposed to dozens with leading monolithic platforms.

The HP 3PAR Command Line Interface (CLI) runs on several client platforms including Windows® (2000, 2003, XP, Vista), and Oracle™ Solaris. The CLI program on the client communicates with a CLI server process on the system via a socket or a Secure Socket Layer (SSL) socket over TCP/IP over the on-board Gigabit Ethernet port on one of the Nodes. Since the HP 3PAR CLI commands can run on a remote client, those commands can be used in scripts on the host.

Management console

The HP 3PAR InForm Management Console, a Java-based application, runs on the same client platforms as the HP 3PAR CLI. Administrators can use the Management Console to monitor all physical and logical components of the system, manage volumes, view performance information (IOPS, throughput, and service times for a variety of components), and monitor multiple HP 3PAR Storage Systems all from the same Management Console instance. Additionally, all unacknowledged alerts from each system are reported in a single event window.

Similar to the CLI, the HP 3PAR InForm Management Console communicates with a Management Console server process on the HP 3PAR Storage System over TCP/IP over the on-board Gigabit Ethernet port on one of the Nodes.

Instrumentation and management integration

Management of the HP 3PAR Storage System benefits from very granular instrumentation within the InForm OS. This instrumentation effectively tracks every I/O through the system and provides statistical information, including Service Time, I/O Size, KB/sec, and IOPS for Volumes, Logical Disks, and Physical Drives. 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 Management Console or through the HP 3PAR CLI. Moreover, administrators at operation centers powered by the leading enterprise management platforms can monitor MIB-II information from the HP 3PAR Storage System. All alerts are converted into SNMP Version 2 traps and sent to any configured SNMP management station.

Alerts

When a critical threshold is encountered or a component fails, an alert is triggered by the InForm OS and is sent to the CLI, Management Console, and the HP 3PAR service processor (which either notifies HP 3PAR Central, HP 3PAR's centralized support center, or records the alert in a log file). These alerts are used by the system to trigger automated action and to notify service personnel that action has been taken (or may need to be scheduled).

Sparing

There are three kinds of chunklets within the system: *used, free,* and *spare*. Used chunklets contain user data. Free chunklets are chunklets that are not used by the system. Spare chunklets are designated as the target onto which to "spare" (or move) data from used chunklets when a chunklet or drive failure occurs, or when a drive magazine needs to be serviced.

To ensure that there is always enough free capacity in the system for drive sparing, a small portion of chunklets within the system (usually the equivalent capacity of four of the largest size physical drives) are identified as "spare" chunklets when the storage system is configured. Additionally, logging logical disk space is allocated upon storage system setup to log writes for a chunklet that is only temporarily unavailable for some reason. When a connection to a physical drive is lost or when a physical drive fails, all future writes to the drive are automatically written to a logging logical disk until the physical drive comes back online or until the time limit for logging is reached. This is referred to as *Auto Logging* or *Chunklet Logging*. If the time limit for logging is reached, or if the logging logical disk becomes full, reconstruction and relocation of used chunklets on the physical drive to other chunklets (free chunklets or allocated spares) begins automatically.

The sparing algorithm for drive replacement offers two alternative methods known as *Servicemag-with-Logging* and *Servicemag*. With Servicemag-with-Logging, the used chunklets from the failed drive of a given magazine are reconstructed and relocated to free or spare chunklets, unless Auto Logging has already completed this operation. Meanwhile, the remaining used chunklets on the remaining valid drives of the drive magazine are moved into logging mode (i.e., data writes to these chunklets continue to a logging logical disk). The magazine is then removed and the failed drive replaced. Once the drive magazine is re-installed and back online, the chunklets from the drives that were not replaced are synchronized by using the logging information. Chunklets from the replaced drive are relocated onto the new drive by moving data back from spare chunklets. Allowing writes to continue to the logging logical disk reduces the number of chunklets to be moved, thereby decreasing the time required to perform a drive replacement procedure.

With Servicemag, all used chunklets from the valid physical drives on the drive magazine are first relocated to other free or spare chunklets in the system. Similarly, the used chunklets from the failed drive are reconstructed and relocated to free or spare chunklets, unless auto logging has already completed this operation. Subsequently, the drive magazine can be removed and the failed drive replaced. Once the drive has been replaced and the drive magazine is installed and back online, the relocated chunklets from the valid physical drives on the drive magazine are moved back to their original positions on the drive magazine and chunklets from the replaced drive are relocated onto the new drive. Temporary relocation of all used chunklets from the physical drives on the drive magazine to spare or free space preserves full RAID protection for all used chunklets on these drives. However, depending on the number of chunklets relocated, this process can increase the time required to perform a drive replacement procedure.

Additional HP 3PAR Software

Additional software products available for HP 3PAR Storage Systems offer enhanced capabilities including thin storage technologies, secure partitioning for virtual private arrays, and virtual and remote copy capabilities. These products are additional array-based software that are purchased separately from the InForm OS.

Thin software suite

Mentioned previously, HP 3PAR Thin Provisioning Software allows organizations to maximize capacity utilization by safely de-coupling "allocated" storage from "used" storage, enabling just-in-time delivery of storage to applications. With HP 3PAR Thin Provisioning Software, an administrator can allocate and export any amount of logical capacity to an application without having to reserve the same amount of actual physical capacity. What the application "sees" as physical capacity is different from what is actually purchased and used. More likely than not, what the application "sees" is much greater than the actual physical storage capacity of the system.

Allocated storage is presented to host systems using Thin Provisioning VVs (TPVVs). Unlike VVs, which are pre-mapped to underlying LDs and ultimately to chunklets, TPVVs are mapped to a logical common provisioning group, which serves as the common storage reservoir. When writes are made to the TPVV, the common provisioning group creates the mapping to underlying logical disks and space gets allocated in fine-grained 16 KB increments to accommodate the write.

HP 3PAR Thin Provisioning Software allows customers to determine and set capacity thresholds flexibly so that when a threshold is reached, the system will generate the appropriate alerts. Over time, as TPVVs utilize capacity within the common provisioning group and as utilization approaches the limit, the system generates several types of warnings to provide ample time for the IT administrator to plan for and add the necessary capacity. In the unlikely scenario that the hard limit is reached, the system naturally prevents new writes from occurring until more capacity becomes added.

The HP 3PAR T-Class and F-Class Storage Systems with Thin Built In extended the platform's leadership in thin provisioning and related thin technologies by introducing the HP 3PAR Gen3 ASIC. The HP P10000 3PAR Storage Systems continue to extend this leadership with two, higher-bandwidth HP 3PAR Gen4 ASICs per controller node. HP 3PAR Utility Storage is the first platform in the industry with thin capabilities built into array hardware to power efficient, silicon-based capacity optimization. The revolutionary, zero-detection-capable HP 3PAR Thin Built In ASICs within each Controller Node are designed to deliver simple, on-the-fly storage optimization to boost capacity utilization while maintaining high service levels.

In addition to Thin Provisioning, HP 3PAR Thin Conversion Software allows customers to "get thin" by using the HP 3PAR Thin Built In ASIC to convert legacy storage volumes while preserving service levels and without impacting production workloads. With Thin Conversion, customers can effectively and rapidly "thin" a heterogeneous datacenter to as little as one-quarter of its original size. An industry first, HP 3PAR Thin Persistence Software keeps thin storage as lean and efficient as possible by reclaiming deleted space, so thin volumes stay thin.

Storage federation with Peer Motion

HP 3PAR Peer Motion Software is the first solution to deliver storage federation capability to both midrange and high-end HP 3PAR Storage Systems to dynamically and non-disruptively distribute data and workloads across peer arrays. Peer Motion enables clients to quickly migrate data and workloads between any model of HP 3PAR array without taking applications offline or impacting workflow.

Virtual and Remote Copy Software

Built on thin copy technology, HP 3PAR Virtual Copy Software is a reservationless, non-duplicative snapshot product that allows customers to affordably protect and share data from any application. HP 3PAR Remote Copy Software, also built on thin copy technology, delivers a solution that dramatically simplifies remote data replication and disaster recovery in the most flexible and cost-effective way possible.

Adaptive and Dynamic Optimization Software

As mentioned previously, HP 3PAR Dynamic Optimization Software is an autonomic storage tiering tool that allows organizations to non-disruptively distribute and redistribute application volumes across tiers to align application requirements with data QoS levels on demand. HP 3PAR Adaptive Optimization Software is another autonomic storage tiering tool that takes a fine-grained, policy-driven approach to service level optimization for enterprise and cloud datacenters. More information is provided about these features in the section on autonomic storage tiering.

Virtual Domains and Virtual Lock

HP 3PAR Virtual Domains Software is virtual machine technology that delivers secure segregation of Virtual Private Arrays for different user groups, departments, and applications while preserving the benefits delivered by the massive parallelism architected into the HP 3PAR platform.

Another software solution that delivers additional security, HP 3PAR Virtual Lock Software provides an efficient and cost-effective way to comply with internal governance and lays the foundation for performing electronic discovery (eDiscovery).

System Tuner

HP 3PAR System Tuner Software provides automatic and non-disruptive hotspot detection and remediation to ensure optimal volume performance over time—even as your system grows and you bring new applications online.

Host Software

Host-based software products offered with HP 3PAR Storage Systems address the needs of specific application environments through solutions that include plug-ins for VMware vSphere™ as well as multipathing and historical performance and capacity management software.

GeoCluster Software for Microsoft Windows

By automating storage failover between primary and backup sites and minimizing disruption to running applications, HP 3PAR GeoCluster Software not only simplifies disaster recovery in Windows-based environments, but drastically reduces administration and application recovery times to ensure business continuity and minimize user impact.

Recovery Manager

HP 3PAR Recovery Manager Software for VMware vSphere allows VMware administrators to create hundreds of VM-aware snapshots and initiate rapid online recovery directly from within the VMware vCenter™ Server virtualization management console. HP 3PAR Recovery Manager Software for Exchange is an extension to HP 3PAR Virtual Copy Software that intelligently creates and manages snapshots that can be used to quickly restore Exchange instances or databases (or non-disruptively back them up to tape) for near-continuous data protection. HP 3PAR Recovery Manager Software for Oracle is an extension to HP 3PAR Virtual Copy Software that intelligently creates, manages, and presents time-consistent snapshot images of Oracle® databases for rapid application recovery, near-continuous data protection, data sharing, and non-disruptive backup. HP 3PAR Recovery Manager Software for Microsoft SQL Server is another extension to HP 3PAR Virtual Copy Software that eases costs and administration by providing rapid, affordable online recovery of Microsoft® SQL Server databases from multiple, highly granular point-in-time snapshots. Quickly recover a database to a known point in time, speeding up a variety of operations including rapid recovery of the production SQL server.

VMware plug-ins

The HP 3PAR Management Software Plug-In Software for VMware vCenter gives VMware administrators enhanced visibility into storage resources and precise insight into how individual virtual machines are mapped to datastores and individual storage volumes. When used in conjunction with HP 3PAR Recovery Manager Software for VMware vSphere, this plug-in gives administrators the power of seamless, rapid online recovery from within the vCenter Server virtualization management console.

System Reporter and Host Explorer

HP 3PAR System Reporter is a historical performance and capacity management tool for storage reporting, monitoring, and troubleshooting. Running as an agent on the server, HP 3PAR Host Explorer Software automates host discovery and collection of detailed host configuration information critical to speeding provisioning and simplifying maintenance. Host Explorer automatically and securely communicates host information such as Fibre Channel World Wide Name (WWN) and host multipath data to the system to reduce manual administration.

Multipathing software

HP 3PAR Multipath I/O Software for IBM AIX provides multipathing for IBM®AIX® hosts featuring multiple active/active paths, load balancing, and automatic failover and recovery. HP 3PAR MPIO Software for Microsoft Windows 2003 provides multipathing for Microsoft Windows® hosts featuring multiple active/active paths, load balancing, and automatic failover and recovery.

System performance

Sharing cached data

Because much of the underlying data of snapshot VVs is physically shared with other VVs (snapshots and/or the base VV), data that is cached for one VV can often be used to satisfy read accesses from another VV. Not only does this save cache memory space, but it also improves performance by increasing the cache-hit rate.
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 InForm OS automatically moves any "pinned" writes in cache to dedicated Preserved Data LDs. This ensures that all host-acknowledged data in cache is preserved and properly restored once the destination drives come back online without compromising cache performance or capacity with respect to any other data.

Pre-fetching

The InForm OS 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 would not 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 guarantee integrity of the 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

Autonomic storage tiering

HP 3PAR offers several products that can be used for service level optimization, which matches data to the most cost-efficient resource capable of delivering the needed service level at any given time. HP 3PAR Dynamic Optimization Software allows storage administrators to move volumes to different RAID levels and/or drive types, and to redistribute volumes after adding additional drives to an array. HP 3PAR Adaptive Optimization Software leverages the same proven sub-volume data movement engine used by Dynamic Optimization Software to autonomically move data at the sub-volume level, also non-disruptively and without administrator intervention.

Volume level tiering with HP 3PAR Dynamic Optimization Software

HP 3PAR Dynamic Optimization Software allows storage administrators to convert a volume to a different service level with a single command. This conversion happens within the HP 3PAR Storage System transparently and without interruption. The agility of Dynamic Optimization Software makes it easy to alter storage decisions. For example, a once-hot project that used RAID 1 on high-

performance Fibre Channel disks may be moved to more cost-effective RAID 5 storage on Nearline (enterprise SATA) disks.

Another use of HP 3PAR Dynamic Optimization Software is to redistribute volumes after adding drives to an HP 3PAR Utility Storage array. Using Dynamic Optimization Software, existing volumes are autonomically striped across existing and new drives for optimal volume performance following capacity expansions. The increase in the total disks for the provisioned volume contributes to higher performance.

HP 3PAR Policy Advisor for Dynamic Optimization Software adds intelligent analysis and additional automation features to Dynamic Optimization. Policy Advisor does this by analyzing how volumes on the HP 3PAR Storage System are using physical disk space and automatically making intelligent, non-disruptive adjustments to ensure optimal volume distribution and tiering of storage volumes.

With Dynamic Optimization and Policy Advisor, organizations can achieve virtually effortless cost- and performance-optimized storage across all stages of the disk-based data lifecycle, even in the largest and most demanding environments.

Sub-volume tiering with Adaptive Optimization

HP 3PAR Adaptive Optimization Software is a fine-grained, policy-driven, autonomic storage tiering software solution that delivers service level optimization for enterprises and cloud datacenters at the lowest possible cost while increasing agility and minimizing risk.

Adaptive Optimization analyzes 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 InForm OS to autonomically move those regions to the fastest storage tier. It also moves less active regions to slower tiers to 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 hot spots rather than evenly-distributed accesses, and these hot spots can move over time.

Using Adaptive Optimization Software, an HP 3PAR Storage System configured with Nearline (enterprise SATA) disks plus a small number of Solid State Drives (SSDs) can approach the performance of SSDs at little more than the cost per megabyte of SATA-based storage, adapting autonomically as access patterns change.

Availability summary

Multiple independent Fibre Channel links

Each HP 3PAR Storage System can support up to 192 independent Fibre Channel host ports using HP 3 PAR's 4-port cards. These are not switched ports, but rather provide full-speed access to the host when any part of the redundant path has failed.

Controller Node redundancy

Controller Nodes are configured in logical pairs whereby each Controller Node has a partner. The two partner Nodes have redundant physical connections to the subset of physical drives owned by the Node pair. Within the pair, Nodes mirror their write cache to each other and each serves as the backup Node for the Logical Disks owned by the partner Node.

If a Controller Node were to fail, data availability would be unaffected. Upon the failure of a Controller Node, the Node failover recovery process automatically flushes the dirty write cache to the physical drive, and transfers ownership for the Logical Disks owned by the failed Node to its partner Node.

Persistent Cache is a feature that allows the surviving Node to mirror write data to another Node, enabling systems with four or more nodes to survive a Node failure with minimal performance impact.

In a system with two nodes, the Node failover recovery process puts all Logical Disks owned by the remaining partner Node in write-thru (non-cached) mode.

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.

Controller Nodes are hot-pluggable and can be serviced or added to a system online and non-disruptively. Similarly, the InForm OS and other associated Node software can be upgraded online and non-disruptively.

RAID data protection

The HP 3PAR Storage System is capable of RAID 1+0 (mirrored then striped), RAID 5+0 (RAID 5 distributed parity, striped in an X+1 configuration where X can be between 2 and 8), or RAID MP (multiple distributed parity, currently striped with either a 6+2 or 14+2 configuration). The RAID 5+0 and RAID MP algorithms allow HP 3PAR to create parity sets on different drives in different drive cages with separate power domains for maximum integrity protection.

No single point of failure

There is no single point of failure for hardware or software in the system. At a minimum, there are two Controller Nodes and two copies of the InForm OS even in the smallest system configuration. The only non-redundant component in the system is a 100% completely passive controller backplane which, given its passive nature, is virtually impervious to failure. RMA MTBF hardware calculations include this component and substantiate this claim.

Separate, independent Fibre Channel controllers

Each HP 3PAR Storage System offers a minimum of two independent (with respect to bandwidth and latency) Fibre Channel host ports per Controller Node. Each of these can independently address all of the data within the unit.

Conclusions

As an HP Converged Storage platform, HP 3PAR Utility Storage was designed from the ground up to deliver massive scalability, secure multi-tenancy, high performance, and high availability to fuel enterprise-class virtual data centers and cloud computing environments. Industry-leading software solutions provide unique benefits that make any cloud more agile and efficient while ensuring secure segregation of user groups and applications.

HP 3PAR Utility Storage addresses the key weaknesses with many of today's existing storage architectures. Customers faced with growing capacity requirements, underutilization of existing storage assets, and administrative inefficiency are searching for ways to decrease both cost and complexity. Simplifying the IT infrastructure requires that next-generation storage architectures provide consolidation, bi-directional scalability, and mixed workload support. The HP 3PAR Storage System addresses all of these requirements and provides multi-tenancy and autonomic management capabilities along with carrier-class availability that includes full software and hardware fault tolerance.

The HP 3PAR Storage Systems offer the following key benefits:

- Performance for large-scale consolidation. HP 3PAR Storage Systems deliver high performance and provide a cost-effective growth path. Moreover, mixed workloads are supported without impact. Unlike legacy architectures that process I/O commands and move data using the same processor complex, the platform's unique Controller Node design separates the processing of control commands from the data movement, enabling simultaneous delivery of random I/O and throughput. Performance bottlenecks observed with existing platforms—for example, when serving competing workloads like OLTP and data warehousing simultaneously—are eliminated.
- Granular scalability. HP 3PAR Storage Systems can scale in granular, modular increments from small departmental systems to mission-critical systems requiring high performance, capacity, and connectivity. Customers can start with a small, modular array footprint and grow storage as business grows. More importantly, HP 3PAR arrays scale easily and with minimal risk due to the granular and non-disruptive upgrades unique to the HP 3PAR Architecture.
- Always-on architecture. The platform's entire hardware and software architecture is designed with high availability in mind. Redundancy and online serviceability are designed into every component, including the software. The HP 3PAR full-mesh, passive system backplane joins multiple Controller Nodes to form a cache-coherent, Mesh-Active cluster. Each Controller Node runs a separate instance of the InForm OS, providing software fault tolerance and ensuring availability of user data. Extensive error checks and proactive events and alerts work with the most advanced service tools in the industry to ensure prompt corrective action. Storage federation capability ensures that data and workloads can be seamlessly migrated without disruption.
- Simplified storage virtualization. The InForm OS provides powerful virtualized volume management capabilities that simplify volume creation and LUN exportation. A tri-level mapping methodology

- similar to the virtual memory architectures of the most robust enterprise operating systems is employed to ensure performance and maximize utilization of physical resources.

- Thin technologies. HP 3PAR Thin Provisioning Software safely de-couples "allocated" storage from "used" storage, empowering administrators to maximize capacity utilization by allowing virtualized volumes to appear to have far greater virtual capacity than physical capacity. Additional thin software and hardware innovations unique to HP 3PAR Utility Storage, including the HP 3PAR Gen4 ASIC, drive additional efficiencies by enabling Thin Conversion, Thin Persistence, and Thin Reclamation. HP 3PAR Utility Storage is the only platform with a comprehensive strategy that helps clients start thin, get thin, and stay thin.
- Scalable, autonomic administration. The HP 3PAR CLI provides administrators simple yet powerful commands to control and monitor the system, interactively or through scripts. User interfaces are designed to offer autonomic administration, allowing an administrator to create and manage physical and logical resources without specifying numerous properties. For example, availability and performance rules are implemented intelligently by the system based on available resources. The HP 3PAR Management Console provides even simpler interactive access to the system.
- Minimal footprint. HP 3PAR Storage Systems offer the system density that is double that of competitive systems, enabling customers to consolidate storage and reclaim scarce datacenter space. Moreover, simple, modular packaging is designed with serviceability in mind.
- Green storage savings. With HP 3PAR Utility Storage, customers can purchase up to 75% less capacity—meaning less equipment to house, fewer disks to power and cool, less hardware to downcycle after it has reached end of life, and a reduced carbon footprint. This also means greater CAPEX and OPEX savings.

Due to the platform's superior efficiency and agility, HP 3PAR Utility Storage has rapidly gained acceptance in mission-critical deployments at Fortune 1000 enterprises, in government environments, and in many specialized industries including managed services and hosting, financial services, insurance, retail, Internet, high technology, and the pharmaceutical industry. More importantly, HP 3PAR Utility Storage provides the converged storage foundation necessary to build a converged infrastructure that supports the data center of the future and meets the needs of the Instant-On Enterprise. Such transformation is key to overcoming the inflexibility and high costs created by IT sprawl and enabling organizations to shift their focus to innovation and strategic initiatives that add value to the business.

For more information

Visit www.hp.com and www.hp.com/go/3PAR

4AA3-3516ENW, Created April 2011; Updated June 2011, Rev. 1