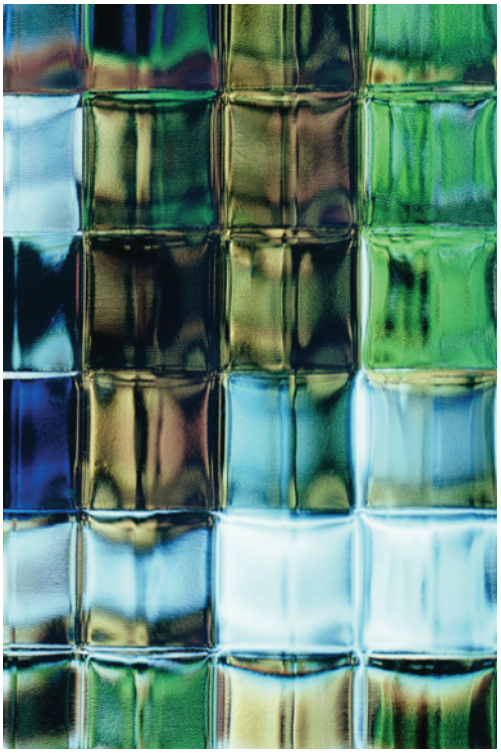
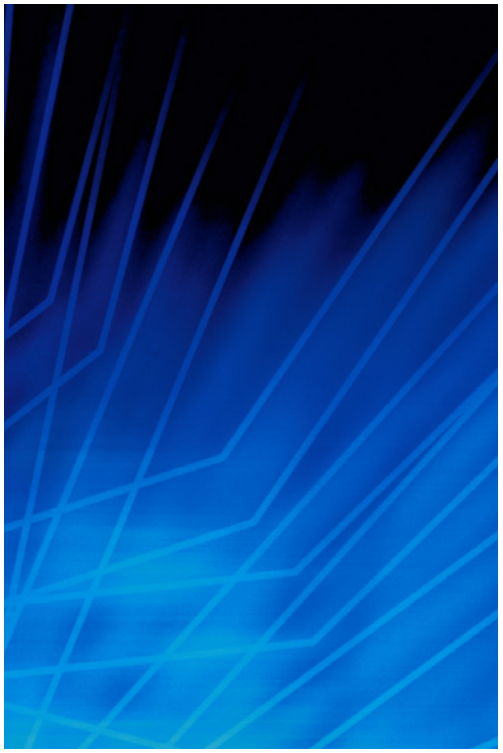




**WHITE PAPER**
**2008**

# The 3PAR InSpire® Architecture
## Returning Simplicity to IT Infrastructures

3PAR
*Serving Information*

# Table of Contents


*Fig. 01*

## 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 these new IT requirements that have evolved over the last decade. As a result, companies have had to add layers of hardware and software to meet their needs—a costly proposition.

IT managers need a solution that can bring the simplicity and elegance back to the storage infrastructure. Enter 3PAR.

3PAR is the global leader in utility storage, a new category of storage systems that enable organizations with multiple lines of business, departments, or customers to securely consolidate storage assets and centralize information for enterprise-scale applications.

3PAR Utility Storage incorporates a tightly tuned system of software, hardware, and mission-critical service. The advanced **3PAR InSpire® Architecture** delivers a modular, highly scalable solution that helps companies reduce storage infrastructure complexity. In fact, it is capable of delivering many times the performance of market-leading monolithic and modular storage architectures.

**The 3PAR InServ® Storage Server family** is the hardware foundation of 3PAR Utility Storage. Unlike modular and monolithic (or cache-centric) storage arrays, 3PAR InServ Storage Servers utilize a cluster-based approach. The modularity of the InServ delivers a single storage platform that scales continuously from the very small to the very large and offers complete fault tolerance of both hardware and software.

**The 3PAR InForm® software family,** with the **3PAR InForm® Operating System** as its foundation, is the intelligence behind 3PAR Utility Storage. The InForm software family uses sophisticated virtualization management capabilities to provide unique advantages in simplification, usability, performance, availability, security, flexibility, and storage efficiency. The InForm software family enables customers to control their storage environment—no matter the size—with dramatically fewer human and capital resources.

This white paper provides an overview of 3PAR InSpire hardware architecture and the 3PAR InForm software architecture.

## 3PAR INSPIRE SYSTEM HARDWARE ARCHITECTURE

### Architecture Overview

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

The Controller Nodes connect using Fibre Channel over two or more paths to each Drive Chassis (i.e. Drive Cage) and to Hosts (either directly or over a Storage Area Network). The cluster of Controller Nodes presents to the Hosts a single, highly available, high performance Storage System. The volume management software on the Controller Nodes creates Virtual Volumes (VVs), which are visible to hosts as storage volumes. VVs are mapped on to one or more Logical Disks (LDs), which implement RAID functionality over the raw storage in the Physical Disks (PDs). VVs can be exported to Hosts as Logical Unit Numbers (LUNs). The cluster of Controller Nodes acts as a single system, so that servers can access VVs over any host-connected Fibre Channel port, even if the physical storage for that data (on physical disks) is connected to a different Controller Node. This is achieved through extremely low latency data transfer across the high-speed, full-mesh backplane.

The 3PAR InSpire Architecture is implemented in three different 3PAR InServ models to meet customer scaling requirements: the InServ T400, T800, and E200. These models accommodate up to two, four and eight Controller Nodes, respectively. This paper will focus on the T-class.

With the 3PAR InSpire Architecture, availability is "built-in" from the start. Unlike other approaches, the InServ offers both hardware and software fault tolerance by running a separate instance of the 3PAR InForm Operating System on each Controller Node, ensuring the availability of customer data. Software or firmware failures—a significant cause of unplanned downtime in other architectures—are greatly reduced.

The 3PAR InSpire Architecture is modular and can be scaled from low to high, deployable as a small remote system or a 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 3PAR InSpire Architecture is well suited for large or high-growth projects, consolidation of mission-critical information, demanding performance-based applications, and data lifecycle management.

A 3PAR InServ T800 offers peak internal bandwidth of 44.8 gigabytes per second (GB/s), significantly more than is required by today's Controller Node implementations. The bandwidth and latencies of the InSpire Architecture supersede bus, switch, and even Infiniband-based architectures.

In every 3PAR InServ Storage Server, each of the Controller Nodes has a dedicated 1.6 GB/s link to each of the other nodes. Each single link is roughly four times the speed of 4 Gb Fibre Channel. In a 3PAR InServ T800, a total of 28 of these links form the array's full-mesh backplane.

### Third-Generation Interconnect: Full-Mesh Controller Backplane

Backplane interconnects within the datacenter 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 to address latency issues across the growing number of devices on the backplane (more processors, larger memory and I/O systems). The third-generation, full-mesh interconnects first appeared in the late 1990s in enterprise servers. However, the 3PAR InServ array represents the first storage platform to apply this interconnect to reduce latencies and address scalability requirements.

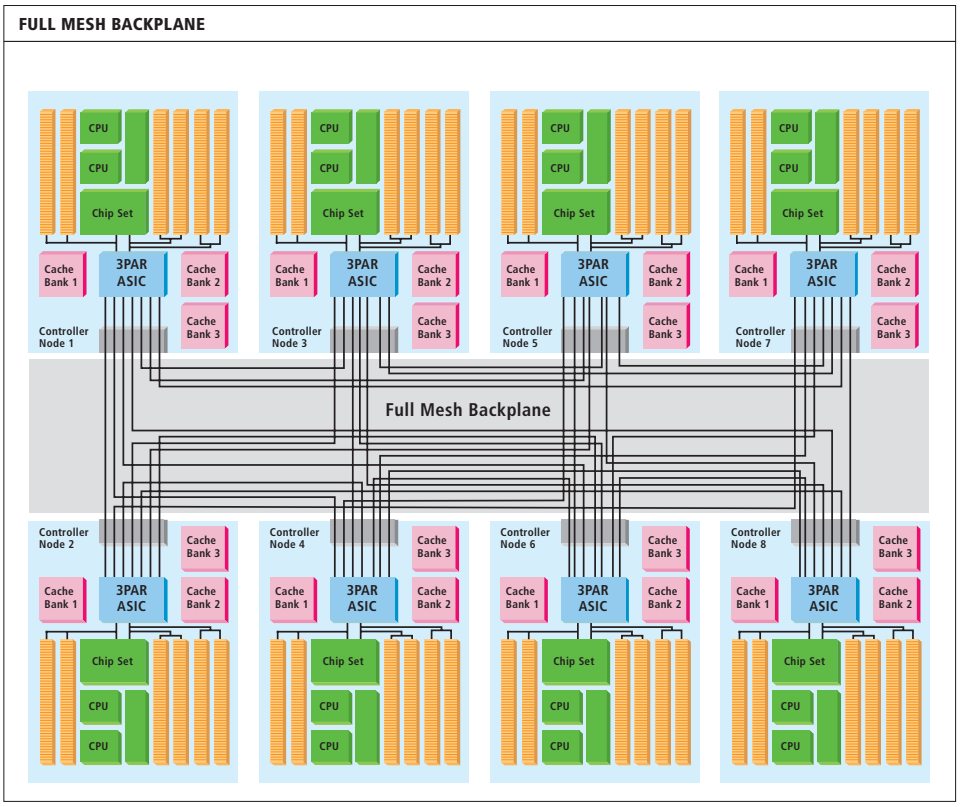



*Fig. 02*

The 3PAR InServ 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 (800 Megabytes per second in each direction, or 1.6 Gigabyte per second total), forming a full-mesh interconnect network between the Controller Nodes. There are two T-class backplane types: a

4-Node backplane (T400 model), which supports 2 to 4 Controller Nodes, and an 8-Node backplane (T800 model), which 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 control information between the nodes which can be used in the event of a failure of the main links.

### Advantages of a tightly coupled cluster architecture

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/disk management. For example, the smallest monolithic system can have a minimum of 6 processing elements (1 for each of 3 functions X 2 for redundancy of each function). In this design, with its emphasis on optimized internal interconnectivity, users gain the active-active advantages of 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 disk 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 be realized only 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 bandwidth and latencies required for truly active-active processing.

3PAR's InSpire Architecture was designed to provide cost-effective single-system scalability through a coherent, multi-node clustered implementation. This architecture begins with a multi-function node design, and like a modular array, requires just two initial nodes for redundancy. However, unlike traditional modular arrays, an optimized interconnect is provided between the nodes to facilitate active-active processing and seamless load balancing of each and all application workloads among the nodes. The interconnect is optimized to deliver low latency, high-bandwidth communication and data movement between nodes through dedicated point-to-point links and a low overhead protocol which features rapid inter-node messaging and acknowledgement. For scalability beyond two nodes, the backplane interconnect accommodates more than 2 controller nodes (up to 8 in the case of the InServ T800 Storage Server). 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 InSpire Architecture is able to provide the best of traditional modular and monolithic designs, in addition to massive load balancing.

## Controller Node

An important element of the 3PAR InSpire Architecture is the Controller Node, 3PAR's proprietary and powerful data movement engine designed for mixed workloads. Controller Nodes deliver performance and connectivity within an InServ Storage Server. 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—giving them powerful flexibility and performance. Each pair of Nodes consumes just four EIA rack units (approximately seven inches) in 3PAR's standard 19-inch cabinet. 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 3PAR InServ Storage Server can withstand an entire Node failure without affecting data availability, and each Node is completely hot-pluggable to enable online serviceability.

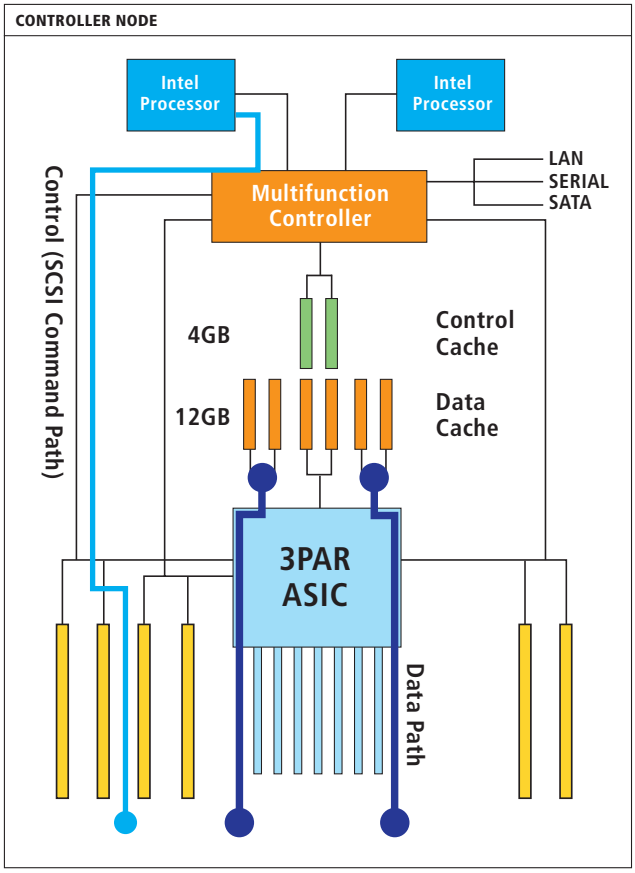



*Fig. 03*

### Mixed Workload Support

Unlike legacy architectures that process I/O commands and move data using the same processor complex, the 3PAR Node design separates the processing of control commands from the 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 3PAR InServ Storage Server, control operations are processed by up to 16 high-performance Intel® Dual-Core processors (for an eight-Node system), with

dedicated control cache up to 32 GBs. All data movement is handled by 3PAR's specially designed ASIC (one per Controller Node), and dedicated data cache of up to 96 GBs.

### Abundant, Multi-Protocol Connectivity

Each Node is equipped with six high-speed I/O slots (48 slots system-wide on an T800) for host and back-end storage connectivity. This design element provides powerful flexibility to support Adapters of multiple communication protocols natively and simultaneously on a 3PAR InServ Storage Server. Fibre Channel and/or iSCSI TOE Adapters can be configured as desired on each Node for multi-protocol host connectivity. In addition, Gigabit Ethernet Adapters 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 Node can deliver a total of 24 ports for a total of 192 ports system-wide. Subject to the InServ Storage Server's configuration, on an T800 up to 128 of these ports may be available for host connectivity, providing abundant connectivity for host connections. Each of these ports is 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.

### 3PAR Controller Node Leverages Commodity Parts

The 3PAR Controller Node design extensively leverages commodity parts with industry-standard interfaces to achieve low costs and the ability to keep pace with industry advances and newer parts. At the same time, the 3PAR Gen3 ASIC adds crucial bandwidth and communication optimizations without limiting the ability to use industry standard parts for other components.

### 3PAR Gen3 ASIC Adds Crucial Bandwidth and Communication Optimizations

As previously mentioned, each Controller Node contains a high-performance Application Specific Integrated Circuit (ASIC) designed by 3PAR. The 3PAR Gen3 ASIC is optimized for data movement between three I/O buses, three memory-bank Data Cache, and the seven high-speed links to the other Controller Nodes over the full-mesh backplane. It performs parity calculations (for RAID 5) on the Data Cache.

3PAR T-Class controller features the world's first storage architecture with Thin Built In™ technology. This new ASIC features a Fat-to-Thin volume conversion algorithm, build right into the silicon. This algorithm, with the help of software planned in the next InForm release, will allow traditionally or "fat" provisioned volumes to be converted to "thin" provisioned volumes nondisruptively. Allocated-but-unused capacity within data volumes is initialized with zeros. The ASIC has built-in zero detection to recognize and virtualize strings of zeros on the fly and the benefit is performance.

An 3PAR InServ T800 Storage Server with eight Controller Nodes has

- 8 ASICs, totaling 44.8 Gbytes per second of peak interconnect bandwidth
- 24 I/O buses, totaling 19.2 Gbytes per second of peak I/O bandwidth
- 24 DDR SDRAM buses for Data Cache and 16 FBDimm buses for Control Cache totaling 123 Gbytes per second of peak memory bandwidth

### Handling Power Failures

Each Controller Node includes a local disk drive that contains the InForm 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 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 disk drive. Although many architectures use "cache batteries," these are not suitable for long downtimes usually associated with natural disasters and unforeseen catastrophes. The InServ's 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 disks of two Nodes. Of course, 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 AC power failure.

A common problem with many battery backup systems is that it is often impossible to be sure that the battery is charged and working. To address this problem, the Controller Nodes in 3PAR InServ Storage Servers are each backed by a string of at least 2 batteries, as previously mentioned. 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 3PAR InForm Operating System keeps track of the battery charge and limits the amount of write data that can be cached based on the ability of the batteries to power the Controller Node long enough to save the data to local disk.

### Data Transfer Paths

Figure 4 shows an overview of data transfers in a 3PAR InServ Storage Server 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 hosts write operation acknowledged.
- W3, W4: The write data is written to two separate disks (D1 and D1') forming the RAID 1 set.

The read operation consists of:

- R1: Data is read from disk D3 into cache memory, and
- R2: Data is transferred from cache memory to the host.

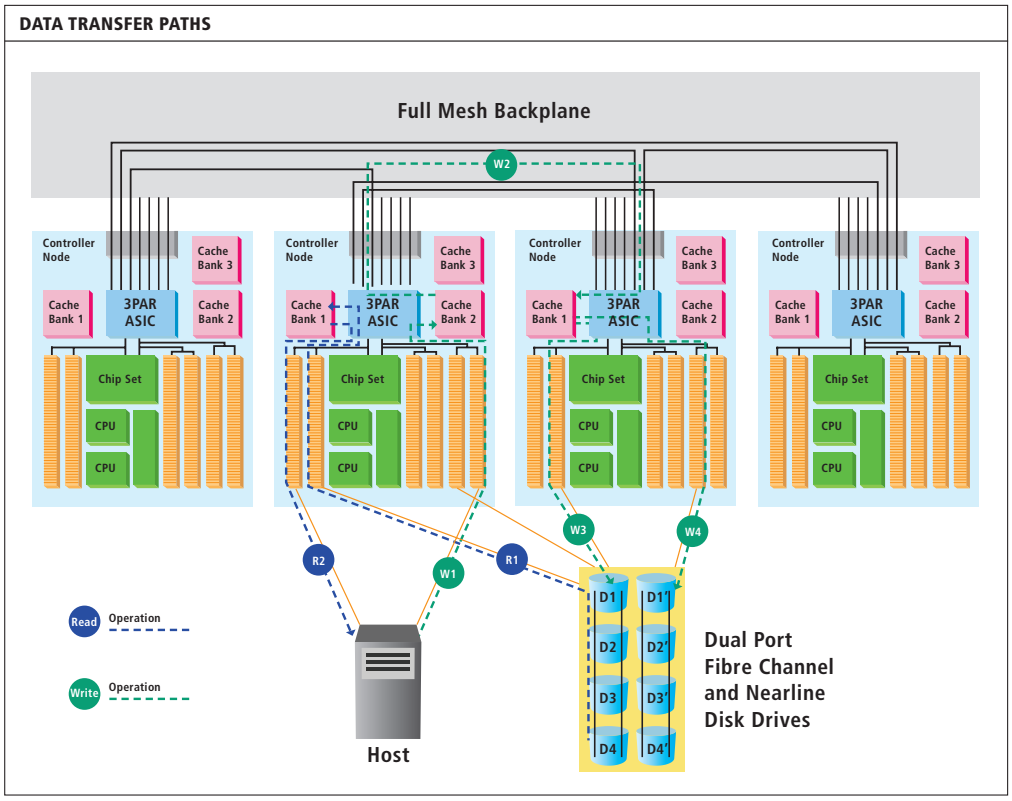



*Fig. 04*

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-disk transfers. Transfers between the Controller Nodes do not consume I/O bus bandwidth.

Processor memory bandwidth is again a substantial bottleneck in traditional architectures, and is also a valuable resource in the Controller Nodes. Uniquely, with the InServ, Controller Node data transfers do not consume any of that bandwidth. This leaves the processors free to perform their control functions far more effectively. All RAID parity calculations are performed by the 3PAR ASIC directly on cache memory and do not consume processor or processor-memory bandwidth.

## Drive Chassis

Another element of the 3PAR InSpire Architecture is the Drive Chassis. Drive Chassis, also referred to as Drive Cages, are intelligent, switched, hyper-dense disk enclosures that serve as the capacity building block within an InServ Storage Server. A single InServ T800 Storage Server can accommodate up to 32 Drive Chassis and scale from 16 to 1,280 drives online and non-disruptively.

### Industry-Leading Density

Drive Chassis have a compact, dense design. Each Drive Chassis (like a pair of Nodes) consumes four EIA rack units in a 19-inch rack. Drive Chassis can be loaded with ten drive magazines, each of which holds four one-inch disk drives. And because each Drive Chassis can hold up to 40

drives, a single Drive Chassis can pack up to 30 terabytes (TBs) of data in just seven inches of rack space when using 750 GB Nearline disk drives. With its compact Drive Chassis design, the InServ platform delivers roughly 2x greater density  over leading alternatives.

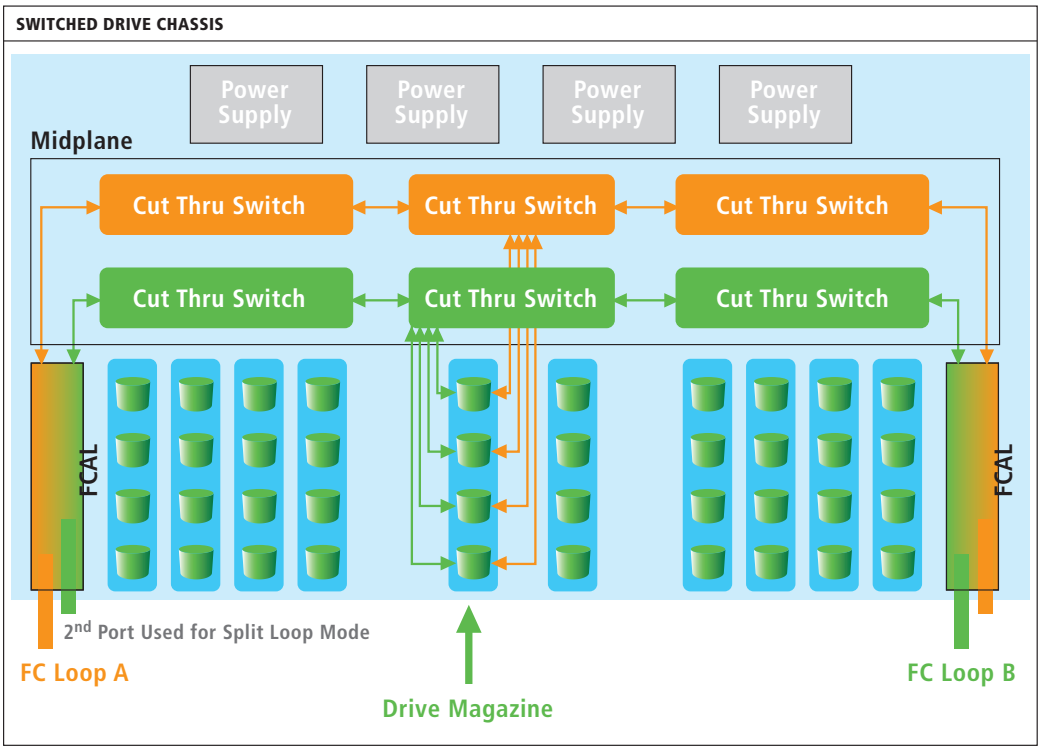



*Fig. 05*

### 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 Gigabit per second, 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 InServ in to the midplane. Redundant power supply/fan assemblies (hot) plug in to the rear of the midplane. Each Fibre Channel disk 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. However rare, should the drive chassis midplane fail, while Drive Chassis midplane is serviced a partner Cage or Cages will continue to serve data for those data 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 10 or RAID 50 data placement to accommodate the failure of an entire Cage without affecting data access.

### Advanced Fault Isolation

Advanced fault isolation and high reliability are built in to the 3PAR InServ. The Drive Chassis, Drive Magazines, and disk drives themselves all report and isolate faults. A drive failure will

not take all disks offline. Through constant monitoring via the Controller Node, Chassis, and drive, the 3PAR InServ Storage Server isolates faults to individual drives and offlines only the failed component.

#### Investment Protection

Drive Chassis provide a common disk enclosure that can house all supported drive types. 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 InServ Storage Server is simplified.

## 3PAR INFORM SOFTWARE ARCHITECTURE

The 3PAR InForm software architecture is packaged as a richly featured suite called the 3PAR InForm Suite. 3PAR InForm Suite works with the 3PAR InServ Storage Server to deliver a new generation of capabilities in storage virtualization, ease of use, security, and service-level reporting while driving down the cost of obtaining and managing enterprise storage resources.

The 3PAR InForm Suite includes the 3PAR InForm Operating System, the core software that delivers unique storage virtualization, virtual volume management, and RAID capabilities. It also includes: 3PAR Access Guard, which enables users to secure hosts and ports to specific volumes within the 3PAR InServ Storage Server; 3PAR Rapid Provisioning, which enables instant, application-tailored volume provisioning; 3PAR Full Copy, which provides a facility for point-in-time physical copies that may be quickly resynchronized with specific base volumes; and 3PAR LDAP which provides centralized authentication and authorization using industry standard lightweight directory access protocol.

### 3PAR InForm Operating System

#### Storage Virtualization: 3PAR Virtual Volume Management

To ensure performance and to maximize the utilization of physical resources, the 3PAR InForm Operating System 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 disk drives of any size into a pool of uniform-sized “chunklets” that are each 256 MB and manages the dual paths to each chunklet and disk 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 chunklets’ fine-grained nature also enhances performance for all applications, regardless of their capacity requirements. While a small application may need only a few chunklets to support its capacity needs, those chunklets may be distributed across dozens or even hundreds of disks. Even a small application can leverage the performance resources of the entire system without provisioning excess capacity. While some vendors stop with this level of virtualization, 3PAR is just getting started.

The second level 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

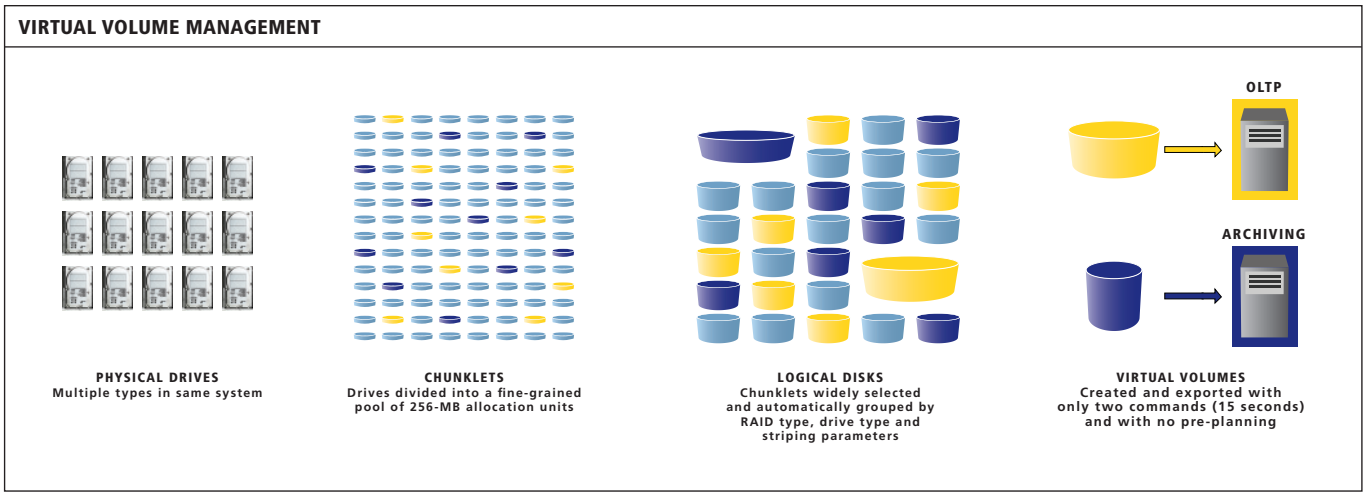



*Fig. 06*

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 required. In addition, the first and second level mappings taken together serve to parallelize work massively across disks and their Fibre Channel connections.

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 Volume LUNs (VLUNs) over Fibre Channel or iSCSI target ports. With the InServ, 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 regions and an exception table with a granularity of 16 KB pages—as opposed to an algorithmic association. Due to 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 simply specifying virtual volume name, RAID type, and size, administrators can have the 3PAR InForm Operating System automatically and intelligently create LDs at the moment capacity needs to be provisioned for a given application.

Separation of the LD and VV layers provides benefits never thought possible based on the limits of traditional array architectures. Consider 3PAR Thin Provisioning, an add-on software product for the 3PAR InForm Operating System. 3PAR Thin Provisioning allows the system administrator to provision VVs several times larger than the amount of physical resources within the storage server. 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.

Further, consider the advanced capabilities that 3PAR Dynamic Optimization, another optional software product, offers. 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 seamlessly behind the scenes. 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.

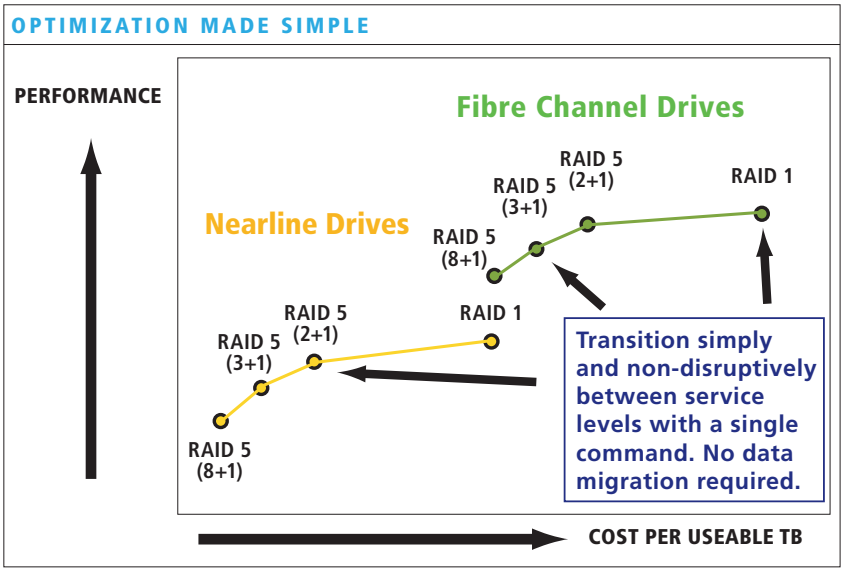



*Fig. 07*

### Physical Disks (PDs), Chunklets, and Drive Cage Firmware

As mentioned, each Physical Disk (PD) is divided into chunklets 256 MB in size. LD allocation refers to chunklets rather than entire disks. This allows great flexibility in LD allocation and permits

- Disks 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 Disks

- 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 InSpire Architecture, LDs implement RAID functionality. Each LD is mapped onto chunklets to implement RAID 10 (mirroring + striping) or RAID 50 (RAID 5 + striping). The InForm System Manager 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 disks that contain redundant data. For example, a RAID 5 LD may have a set size of 4 (3 data + 1 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 disks 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 disk.

- Row size. The row size determines the level of additional striping across more disks. For example, a RAID 5 LD with a row size of 2 and set size of 4 is effectively striped across 8 disks.

- 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 * 6 chunklets/row * 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 3PAR's Virtual Copy facility. 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 meta data (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 host if the VV is exported as a Virtual Volume LUN (VLUN). 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 InServ, VV themselves do not consume LUN numbers as they do on some systems; only VLUNs consume LUN numbers.

### Thin Provisioning

3PAR Thin Provisioning allows companies to maximize capacity utilization by safely de-coupling "allocated" storage from "used" storage, enabling just-in-time delivery of storage to applications. With Thin Provisioning, 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 servers using 3PAR Thin Provisioning VVs. Unlike traditional 3PAR VVs, which are pre-mapped to underlying LDs and ultimately to chunklets, thin VVs are mapped to a logical common provisioning group, which serves as the common storage pool. When writes are made to the thin VV, 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.

3PAR Thin Provisioning allows customers to determine and set capacity thresholds flexibly so that when a threshold is reached, the InServ will generate the appropriate alerts. Over time, as thin VVs utilize capacity within the common provisioning group and as utilization approaches the limit, the system automatically generates several types of warning 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 InServ naturally prevents new writes from occurring until more capacity becomes added.

The InServ T-Class arrays with Thin Built In extend 3PAR's leadership in thin provisioning and thin technologies by introducing the 3PAR Gen3 ASIC. 3PAR is the first in the industry to build thin capabilities into array hardware to power efficient, silicon-based capacity optimization. The revolutionary, zero-detection capable 3PAR Gen3 ASIC within each T-Class Controller Node is designed to deliver simple, on-the-fly storage optimization to boost capacity utilization while maintaining high service levels.

### InForm Command Line Interface

While the flexibility provided by the tri-level virtualization methodology is enormous, management complexity is not. In fact, management of the 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). While there are a few other functions, these commands represents ninety percent of the console actions necessary, returning simplicity to the storage environment.

In addition to simple functions, the InServ'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 InServ, 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 tools. The 3PAR InForm Operating System will intelligently 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, 3PAR saves the system administrator valuable time that could be better spent managing additional terabytes and projects. VV creation requires only two steps, versus dozens with leading monolithic platforms.

The InForm Command Line Interface (CLI) runs on several client platforms including Windows® (2000, 2003, XP), and Sun™ Solaris™. The CLI program on the client communicates with a CLI

server process on the InServ via socket or a Secure Socket Layer (SSL) socket over TCP/IP over the on-board 100BaseT Ethernet port on one of the nodes. Since the InForm CLI commands can run on a remote client, those commands can be used in scripts on the host.

### InForm Management Console

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

Similar to the CLI, the Management Console communicates with a Management Console server process on the 3PAR InServ over TCP/IP over the on-board 100BaseT Ethernet port on one of the nodes.

### Instrumentation and Management Integration

Management of the 3PAR InServ Storage Server benefits from very granular instrumentation within the 3PAR InForm Operating System. 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 Disks. 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 InServ cluster.

These statistics can be reported through the Management Console or through the InForm CLI. Moreover, administrators at operation centers powered by the leading enterprise management platforms can monitor MIB-II information from the 3PAR InServ. 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 Operating System and is sent to the CLI, Management Console and the 3PAR service processor (which either notifies 3PAR Central, 3PAR's centralized support center, or records the alert in a log file). These alerts are used by the InServ 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 disk 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 largest size physical drives)

are identified as "spare" chunklets when the storage server is first set up. Additionally, logging logical disk space is allocated upon storage server set up to log writes for a chunklet that is only temporarily unavailable for some reason. When a connection to a physical disk is lost or when a physical disk fails, all future writes to the disk are automatically written to a logging logical disk until the physical disk 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 disk to other chunklets (free chunklets or allocated spares) starts automatically.

The sparing algorithm for disk replacement offers two alternatives known as "Servicemag-with-Logging" and "Servicemag". With Servicemag-with-Logging, the used chunklets from the failed disk 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 disk 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 on to 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 disk replacement procedure.

With Servicemag, all used chunklets from the valid physical disks on the drive magazine are first relocated to other free or spare chunklets in the system. Similarly, the used chunklets from the failed disk 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 disk replaced. Once the disk has been replaced and drive magazine is installed and back online, the relocated chunklets from the valid physical disks 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 disks 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 disk replacement procedure.

## PERFORMANCE

### Caching and Buffering

#### 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 disks underlying a RAID set become temporarily unavailable (for example, if all cables to those drives were accidentally disconnected), the InForm Operating System 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 3PAR InForm Operating System keeps track of read streams for VVs so that it can improve performance by "pre-fetching" data from disks 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 disks may 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 the write were actually performed to the disks 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 disk performance by:

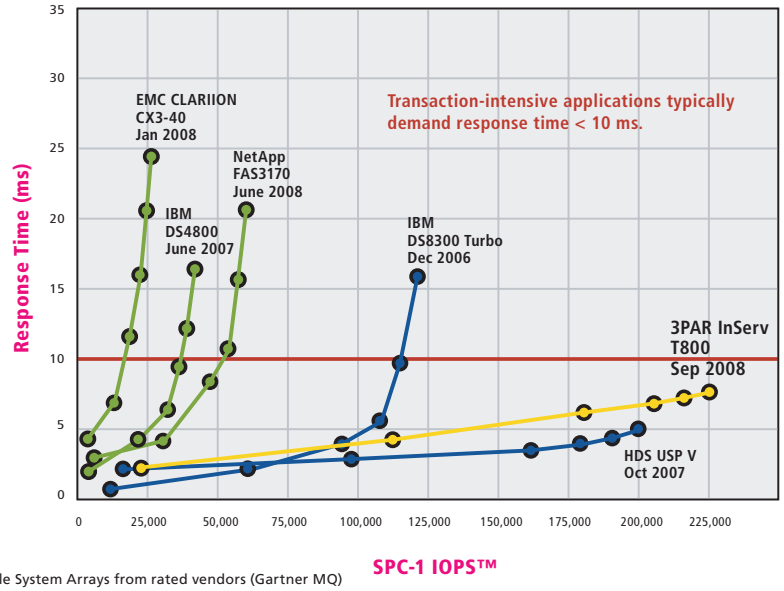
- Merging multiple writes to the same blocks so that many disk writes are eliminated
- Merging multiple small writes into single larger disk writes so that the operation is more efficient
- Merging multiple small writes to a RAID 5 LD into full-stripe writes so that it is not necessary to read the old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

### Performance Benchmarks

#### SPC Benchmark 1™: Example of 3PAR InServ Performance Characteristics

In real-world production environments (as opposed to benchmark environments), delivering abundantly scalable levels of performance for multiple disparate applications in a single system and delivering those levels efficiently and simply is desirable so that organizations can consolidate with confidence and avoid the high human and capital costs traditionally associated with managing performance.

**SPC-1™ SINGLE SYSTEM BENCHMARK COMPARISON**




Single System Arrays from rated vendors (Gartner MQ)
Version 1.8 of SPC-1 results or later. Cache-protected results only. Minimum ASU size 1TB.
Scalability beyond 120 drives. Duplicate results from different vendors eliminated.

*Fig. 08*

**SPC-1™ SMART, THIN, AND READY COMPARISON**

| Tested Storage Configuration | | 3PAR InServ® T800 | EMC CLARiiON CX3-40 | IBM TotalStorage DS4800 | NetApp FAS3170 | IBM TotalStorage DS8300 Turbo | Hitachi Universal Storage Platform V |
|---|---|---|---|---|---|---|---|
| SPC-1 IOPS™ | READY | 224,989.65 | 24,997.49 | 42,254.07 | 60,515.34 | 123,033.40 | 200,245.73 |
| Total ASU* Capacity (GBs) | READY | 77,824.00 | 8,465.02 | 6,871.11 | 19,628.50 | 9,103.36 | 26,000.00 |
| SPC-1 Price/Performance $/SPC-1 IOPS™ | THIN | $9.30 | $20.72 | $17.55 | $10.01 | $18.99 | $17.61 |
| $/ASU TB | THIN | $26,882 | $61,187 | $107,924 | $30,861 | $256,653 | $135,628 |
| TSC** Configuration Script Command Lines | SMART | 142 | 119 | 104 | 225 | 474 | 12800 |
| TSC Configuration Script Command Lines per ASU TB | SMART | 1.8 | 14 | 15 | 11 | 52 | 492 |
| Data Protection Level | | Mirroring | Mirroring | Mirroring | RAID 6 | Mirroring | Mirroring |
| Identifier | | A00069 | A00059 | A00042 | A00066 | A00049 | A000054 |
| Version | | 1.10.1 | 1.10.1 | 1.9 | 1.10.1 | 1.10 | 1.10.1 |

*— Application Storage Unit
**— Tested Storage Configuration

Single System Arrays from rated vendors (Gartner MQ) Version 1.8 of SPC-1 results or later.
Cache-protected results only. Minimum ASU size 1TB.
Scalability beyond 120 drives. Duplicate results from different vendors eliminated.

*Fig. 09*

With 3PAR's massively parallel and automatically load balanced InSpire Architecture, achieving high and predictable levels of performance is dramatically simplified on 3PAR InServ Storage Servers. As discussed earlier, while a small application may need only a few chunklets to support its capacity needs, those chunklets may be distributed across dozens or even hundreds of disks. Each VV supporting an application can leverage all the performance resources—Controller Nodes, cache, ports, I/O buses, and loops—of the entire system automatically without requiring any extensive planning.

3PAR also makes high levels of performance and consolidation affordable, so organizations don't have to overprovision administration or capacity for the sake of performance. Unlike legacy storage architectures, where delivering performance can mean overprovisioning and thereby dramatically underutilizing system resources, 3PAR's massive parallelism and automatic load balancing allows high levels of performance to be achieved with high levels of utilization.

3PAR has posted record-setting SPC-1 benchmark results in which the 3PAR InServ T800 achieved an SPC-1 IOPSTM rate of 224,989.65 and an SPC-1 Price-Performance value of $9.30/SPC-1 IOPS at a total ASU capacity of 77,824 gigabytes. These results used a data protection level of mirrored and received SPC-1 Audit Identifier A00069. The T-Class features the only single-system storage architecture to report 224,989.65 IOPS in a published SPC-1 result, which was achieved with 83% capacity utilization and without complex configuration or performance tuning.

As shown, the 3PAR InServ T800 Storage Server delivers leading SPC-1 IOPS at low latencies (<10ms) that are typically required for transaction-intensive applications. And, with 3PAR's modular architecture, users can start with smaller configurations and scale performance linearly and nondisruptively as dictated by the growth of deployed applications. 3PAR also offers unique mixed workload technology so that transaction- and throughput-intensive workloads can run without contention on the same storage resources, alleviating performance concerns and cutting excessive storage array purchases.

## AVAILABILITY SUMMARY

### Multiple Independent Fibre Channel Links

Each InServ (RAID Unit) can support up to 128 independent Fibre Channel host ports using 3PAR's 4-port cards. These are not switch ports, but rather provide full speed access to the host when any part of the redundant path is failed.

### Controller Node Redundancy

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

If a controller node were to fail, data availability is unaffected. Upon the failure of a controller node, the node failover recovery process automatically flushes the dirty write cache to disk, transfers ownership for the Logical Disks owned by the failed node to its partner node, and puts all Logical

Disks owned by the remaining partner node in write-thru (non-cached) mode. Since a given Virtual Volume on the InServ consists of several Logical Disks that are spread across all configured nodes, the failure of any single node in a system with four or more nodes results in only a portion of Logical Disks going into write-thru mode.

Furthermore, under certain circumstances, the InServ 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 an InServ online and non-disruptively. Similarly, the InForm Operating System and other associated node software can be upgraded online and non-disruptively.

### RAID Data Protection

The 3PAR InServ Storage Server is capable of RAID 10 (mirrored, striped) and RAID 50 (RAID 5, striped in an X+1 configuration, where X can be between 2 and 8). The RAID 50 algorithm allows 3PAR to create parity sets on different disks in different drive cages with separate power domains for maximum integrity protection.

### No Single Point of Failure

There is no hardware or software single point of failure in the InServ. At a minimum, there are two controllers and two copies of the InForm Operating System even in the smallest InServ T400 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.

### Two Separate 4Gb/s Fibre Channel controllers

Each 3PAR InServ offers a minimum of two independent (with respect to bandwidth and latency) Fibre Channel host ports per Controller Node, which translates to 16 ports per InServ. Each of these can independently address all of the data within the unit.

## SUMMARY

3PAR has built the first and only enterprise storage array that delivers true utility storage capabilities, enabling utility computing today. Designed from the ground up to address the needs of utility computing, 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 3PAR InServ addresses all of these needs and provides carrier-class availability that includes full software and hardware fault tolerance.

3PAR InServ has rapidly gained acceptance in mission-critical deployments at Fortune 1000 enterprises in government environments as well as several industries, including financial services, insurance, retail, Internet, Hi-Technology, and the pharmaceutical industry.

The 3PAR InServ Storage Server offers the following key benefits:

**Performance for Large-Scale Consolidation.** The 3PAR InServ delivers high performance and provides a cost effective growth path when needed. Moreover, mixed workloads are supported without impact. Unlike legacy architectures that process I/O commands and move data using the same processor complex, the InServ'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 of existing platforms—for example, when serving competing workloads like OLTP and data warehousing simultaneously—are eliminated.

**Granular Scalability.** The 3PAR InServ 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 their business grows. More importantly, the 3PAR InServ scales easily and risk-free using the granular and non-disruptive upgrades unique to the 3PAR InSpire Architecture.

**Always-On Architecture.** The InServ'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 3PAR full-mesh, passive system backplane joins multiple Controller Nodes to form a cache-coherent, active-active cluster. Each controller node runs a separate instance of the 3PAR InForm Operating System, 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.

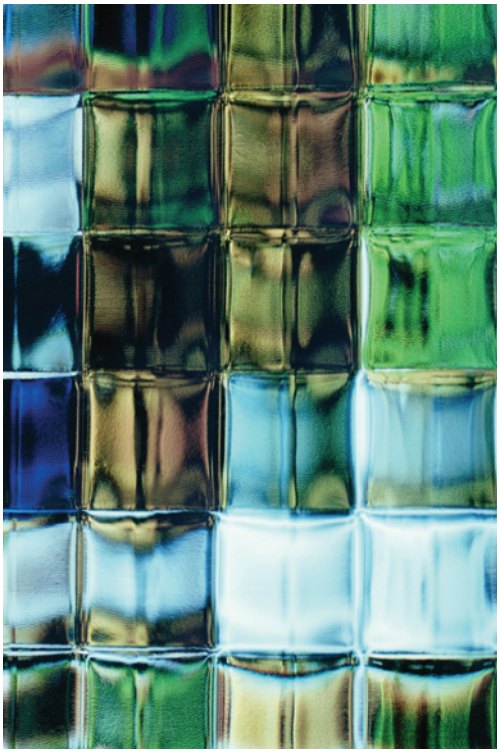
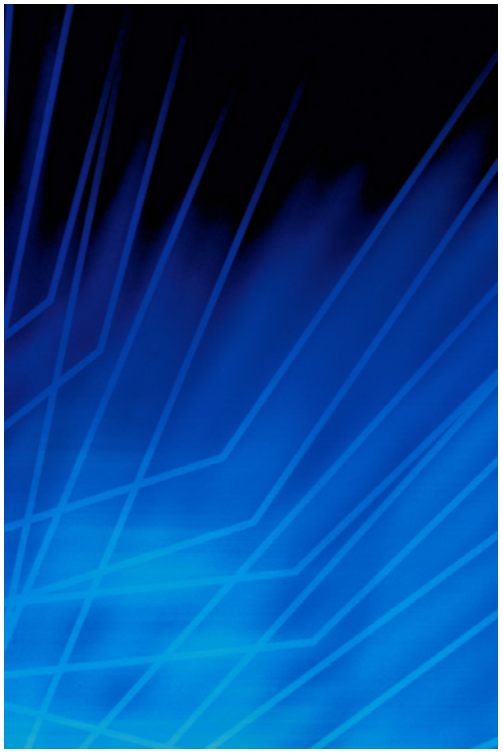
**Simplified Storage Virtualization.** The 3PAR InForm Operating System provides powerful virtualized volume management capabilities that simplify volume creation and LUN exportation. A tri-level mapping methodology similar to the virtual memory architectures of the most robust enterprise operating systems is employed to ensure performance and maximize utilization of physical resources. Thin Provisioning, which safely de-couples "allocated" storage from "used" storage, empowers administrators to maximize capacity utilization by allowing virtualized volumes to appear to have far greater virtual capacity than physical capacity.

**Scalable, Autonomic Administration.** The InForm 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 InForm GUI provides even simpler interactive access to the system.

**Excellent Packaging.** The 3PAR InServ offers the densest packaging of any competitive system by a factor of double or more, enabling customers to consolidate storage and reclaim scarce datacenter space. Moreover, simple, modular packaging is designed with serviceability in mind. The 3PAR InServ uses a standard 19" rack and offers one of the best environmental characteristics.

## ABOUT 3PAR

3PAR® (NYSE Arca: PAR) is the leading global provider of utility storage, a category of highly virtualized, tightly-clustered, and dynamically-tiered storage arrays built for utility computing. Organizations use utility computing to build cost-effective virtualized IT infrastructures for flexible workload consolidation. 3PAR Utility Storage gives customers an alternative to traditional arrays by delivering resilient infrastructure with increased agility at a lower total cost to meet their rapidly changing business needs. As a pioneer of thin provisioning—a green technology developed to address storage underutilization and inefficiencies—3PAR offers products designed to minimize power consumption and promote environmental responsibility. With 3PAR, customers have reduced the costs of allocated storage capacity, administration, and SAN infrastructure while increasing adaptability and resiliency. 3PAR Utility Storage is built to meet the demands of open systems consolidation, integrated data lifecycle management, and performance-intensive applications. For more information, visit the 3PAR Website at: www.3PAR.com.


## U.S. CORPORATE HEADQUARTERS

**3PAR Inc.**
4209 Technology Drive
Fremont, CA 94538
Phone: 510-413-5999
Fax: 510-413-5699

Email: salesinfo@3PAR.com


3PAR
*Serving Information*

ar-wp-08.3