

HP Smart Array controller technology

Technology brief, 4th edition

Introduction	2
Smart Array performance factors	2
Smart Array processing engine	2
Hardware tuning	3
Cache module performance benefits	3
RAID performance enhancements	5
Background RAID creation	5
RAID 5 and RAID 6 read-modify-write	5
Striping across arrays	5
RAID 1 load balancing	5
Data availability	5
Adjusting strip size with the Array Configuration Utility (ACU)	6
RAID migration with ACU	6
Drive roaming	6
Mirror splitting and recombining with the ACU	6
Online drive flash	7
Recovery ROM	7
Pre-Failure Warranty using S.M.A.R.T technology	7
Automatic data recovery with rapid rebuild technology	7
Online spare	8
Dynamic sector repair	8
ECC protection	8
Write caches	8
Software RAID	12
Zero Memory RAID	12
Smart Array Advanced Pack	12
RAID 6	13
RAID 60	13
Advanced Capacity Expansion (ACE)	13
Mirror splitting and recombining	13
Drive erase	13
Storage pathway redundancy	14
Video on Demand (VOD)	14
Storage management	14
Array Configuration Utility	15
Option ROM Configuration for Arrays	16
CPQONLIN	16
HP Systems Insight Manager	16
Array Diagnostics Utility	16
Summary	17
For more information	18
Call to action	19



Introduction

In 1989, HP (then Compaq) was the first company to introduce RAID subsystems in the network server marketplace. Today, RAID is an industry-standard technology used for most online network data storage. Our Smart Array controllers support RAID levels 0, 1, 1+0, 5, 6, 50, and 60.

HP engineers continue to enhance Smart Array performance, expansion, migration, and data availability capabilities. We provide administrators with tools for Smart Array configuration, storage management, and diagnostics. These tools make it easy to use Smart Array products as well as making interfaces and feature sets consistent between product generations. Consistent features sets ensure that you can move data between servers and external storage enclosures, and between models of Smart Array controllers.

This paper discusses the Smart Array processor engine, RAID level selection, data availability, fault tolerance, and recovery mechanisms. These capabilities allow Smart Array controllers to sustain high I/O throughput in a reliable, predictable way.

References made in this technology brief to the “present generation” of Smart Array controllers refer to the P410i, P410, P411, P212, P712, and P812. We began releasing these SAS-based, PCIe 2.0-compliant Smart Array controllers in 2009.

Smart Array performance factors

HP Smart Array hardware performance depends on multiple, interdependent factors.

We’ve enhanced Smart Array hardware performance with:

- Multi-core multi-processing engine
- Hardware tuning that includes customized connection settings
- Advanced caching capabilities

Standard technologies that contribute to Smart Array performance include:

- PCIe bus
- Serial Attached SCSI (SAS) and its interoperability with Serial Advanced Technology Attachment (SATA) devices
- SAS-2 enhancements to cabling and connectors
- SAS topologies and zoning

Smart Array processing engine

The Smart Array multi-core multi-processing engine manages the RAID system. It transforms high-level read or write requests from an application into the individual instructions required for the RAID array. The current generation Smart Array controllers use an embedded RAID-on-Chip (RoC) processor running at 600 MHz. While it is not a direct measure of overall RAID performance, the new processor is capable of supporting up to 60,000 4 KiB random IOPS compared to 35,000 for the previous generation engines.

NOTE:

KiB, MiB, and GiB are measurement units established by the International Electrotechnical Commission (IEC) and accepted by all major standards organizations. They replace the use of KB, MB, and GB. You may see both acronym types, along with other acronym replacements, as we update our tools, user interfaces, and documentation.

The Smart Array processing engine is responsible for processing all operations, but its capabilities are particularly critical to complex RAID functions such as write operations. Both RAID 5 and RAID 6 use mathematical XOR (Exclusive or) operations to calculate the parity data written to the drive array for data recovery. The processing engine is essential for high performance operations at these RAID levels, particularly write performance. Newer Smart Array controllers managing arrays with larger drive counts display the most performance improvements. With smaller drive counts, logical drive array performance tends to be constrained by the aggregate I/O of the drives and not the bandwidth of the Smart Array processing engine.

Hardware tuning

RAID controllers with poor signal integrity between devices and without an optimized PCIe bus can have poor performance. Our engineering teams tune connection settings on every Smart Array controller with every connected device, and even the PCIe bus, to reduce error rates.

HP signal integrity requirements for the Smart Array controllers are far more stringent than the industry standard. The generally accepted industry standard for the Unrecoverable Bit Error rate is typically specified at 1 bit error in 10^{12} for enterprise-class disk drives in Fibre Channel and SAS environments. We feel that an error rate of even 10^{15} is unacceptable.

Our engineers customize controller link settings to maximize signaling margins and reduce error rates. We adjust PHY¹ transmitter and receiver connections to provide optimum settings for each Smart Array controller with every connected device. Controllers and connected devices are tested at 1.5 Gb/s SATA, 3 Gb/s SATA, 3 Gb/s SAS, and 6 Gb/s SAS. HP assigns “loss factors” to each device slot to indicate optimum PHY tuning using the results of these tests. Unique IDs identify all devices and loss factor data. The IDs and associated loss factor data provide the information needed to precisely adjust the controller PHYs for highest performance and signal margins. We stress test our controllers by intentionally degrading the signal to the point of failure. This determines the exact signal margins. We also test Smart Array controllers with all compatible operating systems, HP servers, server backplanes, and compatible HP devices.

The Smart Array controller powers up using a default configuration that lets it function efficiently until the retrieval of device-unique IDs and associated loss factor data. If the Smart Array controller cannot retrieve any loss factor data, or if it detects compromised signal integrity by an event like cable failure, Smart Array controllers can “step down” operating speed to maintain a connection. iLO management intervention accomplishes the performance step down. The Smart Array controller resumes normal operations once it no longer detects an issue.

The Smart Array controller PCIe connector has a unique ID and customized settings. HP engineering adjusts the settings for the PCIe connector to account for the PCIe topology where the controller resides. These customized settings apply to both embedded and add-in cards attached to the PCIe bus.

Cache module performance benefits

With advanced read-ahead and write-back caching capabilities, Smart Array controller cache modules significantly improve I/O performance.

Read-ahead caching

The HP Smart Array controller family uses an adaptive read-ahead algorithm that anticipates data needs and reduces wait time. It can detect sequential read activity on single or multiple I/O threads and predict when sequential read requests will follow. The algorithm then reads ahead from the disk drives. When the read request occurs, the controller retrieves the data from high-speed cache memory in microseconds rather than from the disk drive in milliseconds. This adaptive read-ahead scheme provides excellent performance for sequential small block read requests.

¹ The physical layer device, a circuit block that includes a PMD (physical media dependent), a PMA (physical media attachment), and a PCS (physical coding sublayer)

The controller disables read-ahead when it detects non-sequential read activity. HP Smart Array controller adaptive read-ahead caching eliminates issues with fixed read-ahead schemes that increase sequential read performance but degrade random read performance.

Write-back caching

HP Smart Array controllers use a write-back caching scheme that lets host applications continue without waiting for write operations to complete to the disk. A controller without a write-back cache returns completion status to the OS after it writes the data to the drives. A controller with write-back caching can “post” write data to high-speed cache memory and immediately return “back” completion status to the OS. The write operation completes in microseconds rather than milliseconds. The controller writes data from the controller’s write cache to disk later, at an optimal time for the controller.

Once the controller locates write data in the cache, subsequent reads to the same disk location come from the cache. Subsequent writes to the same disk location will replace the data held in cache. This is a “read cache hit.” It improves bandwidth and latency for applications that frequently write and read the same area of the disk.

The write cache will typically fill up and remain full most of the time in high-workload environments. The controller uses this opportunity to analyze the pending write commands to improve their efficiency. The controller can use write coalescing that combines small writes to adjacent logical blocks into a single larger write for quicker execution. The controller can also perform command reordering, rearranging the execution order of the writes in the cache to reduce the overall disk latency. With larger amounts of write cache memory, the Smart Array controller can store and analyze a larger number of pending write commands, increasing the opportunities for write coalescing and command reordering while delivering better overall performance.

Logical drives in RAID 5 and RAID 6 configurations gain higher write performance by combining adjacent write requests to form a full stripe of data (“full-stripe write”). Write operation for RAID 5 and RAID 6 normally requires extra disk reads to compute the parity data. But if all the data required for a full stripe is available in the cache, the controller does not require the extra disk reads. This improves write bandwidth for sequential writes to a logical drive in a RAID 5 or RAID 6 configuration.

Error checking and correcting (ECC) DRAM technology protects the data while it is in cache. Smart Array battery-backed or flash-backed cache backup mechanisms protect the cache data against a server crash and power loss. The controller disables caching when battery-backed or flash-backed cache is an option but the battery-backed or flash-backed cache is not installed. You can override this behavior but doing so opens a window for possible data loss. Disk drives provide an option to enable write caching that is not battery backed. We advise against enabling disk drive write cache because a power failure could result in data loss.

Cache width

Present generation Smart Array controllers support 256 MiB, 512 MiB, and 1 GiB cache modules. The 512 MiB and 1 GiB modules use a 72-bit wide (64 bits data + 8 bits parity) cache instead of the 40-bit wide (32 bits data + 8 bits parity) cache used in the 256 MiB modules. This doubles the bandwidth for moving cache data to and from the storage system, further increasing overall array performance.

For more information on Smart Array cache modules, see the [“Data Availability”](#) section later in this paper.

RAID performance enhancements

Smart Array controllers use several enhancements to increase RAID performance.

Background RAID creation

When you create a RAID 1, RAID 5, or RAID 6 logical drive, the Smart Array controller must build the logical drive within the array and initialize the parity before enabling certain advanced performance techniques. Parity initialization takes several hours to complete. The time it takes depends on the size of the logical drive and the load on the controller. The Smart Array controller creates the logical drive, initializing the parity whenever the controller is not busy. While the controller creates the logical drive, you can access the storage volume which has full fault tolerance.

RAID 5 and RAID 6 read-modify-write

After parity initialization is complete, writes to a RAID 5 or RAID 6 logical drive are typically faster because the controller does not read the entire stripe to update the parity data. Since the controller knows that the parity data is consistent with all the member drives in the stripe, the controller needs to read from only two disk drives during a RAID 5 write or three disk drives for a RAID 6 write to compute the parity data, regardless of array size.

Striping across arrays

RAID 50 and 60 methods stripe the data across multiple RAID/JBOD sets with different levels of parity. These nested RAID types let you configure arrays across HP Modular Smart Arrays (MSAs).

RAID 50 (RAID 5+0) uses RAID 0 block-level striping across RAID 5 arrays with distributed parity. RAID 50 will tolerate one drive failure in each spanned array without loss of data. RAID 50 configurations require a minimum of six drives and require less rebuild time than single RAID 5 arrays.

RAID 60 (RAID 6+0) uses RAID 0 block-level striping across multiple RAID 6 arrays with dual distributed parity. With dual parity, RAID 60 will tolerate the failure of two disks in each spanned array without data loss. RAID 60 configurations require a minimum of eight drives.

RAID 60 is available as an option with the [Smart Array Advanced Pack](#) (see section later in this paper) and is not supported on all HP Smart Array controllers.

RAID 1 load balancing

RAID 1 logical drives contain two copies of the data. During reads to RAID 1 logical drives, the Smart Array controller issues read requests to either drive in the mirrored set. The Smart Array controller uses RAID 1 load balancing to balance the number of requests between the two disk drives during a heavy read load to achieve higher read bandwidth.

Data availability

Smart Array controllers support online array expansion, logical drive extension, strip² size migration, and RAID migration. These technologies protect data and let you modify the array without interrupting data access. Smart Array controllers can monitor I/O activity, track key parameters, predict potential problems,

² When a Smart Array controller makes an array, the unit of data that it manipulates is defined as a "strip" (ranging in size from 64 KiB to 256 KiB). These strips are distributed across the physical drives in the array. A "stripe" is one set of strips. Our Smart Array controllers actually configure strips not stripes. You can calculate the stripe size from the strip size, the number of physical drives in the logical drive, and the RAID level.

take corrective actions, provide automatic recovery, and deliver full fault management to protect against downtime.

Adjusting strip size with the Array Configuration Utility (ACU)

You can change the strip size of a logical drive, a process called strip-size migration, without any downtime by using the Array Configuration Utility, a browser-based graphical application that helps configure Smart Array controllers.

You can use strip size migration to improve logical drive performance if your read/write environment changes. All RAID levels supported have a default strip-size of 256 KiB. The default provides good performance across many types of applications. You can set the strip-size to 8, 16, 32, 64, 128, or 256 KiB for all supported RAID levels. You can also set the strip-size to 512 KiB for RAID 50. These values can change for certain controllers and storage device implementations. You should consult configuration documentation for your specific controller and storage device.

Certain applications, especially those performing mostly one type of transaction, write transactions for example, may require you to adjust the strip-size to achieve optimal performance. Table 1 lists recommended strip sizes for general types of server applications. You can start with these general recommendations and then fine-tune to determine the best overall performance for a particular application.

The ACU lets you make these changes online without disruption or data loss. For more about the ACU, see the "[Storage management](#)" section later in this paper.

Table 1. Recommended strip sizes

Type of server application	Suggested strip size
Mixed read/write	Default value
Mainly sequential read (such as audio/video applications)	Larger strip sizes
Mainly write (such as image manipulation applications)	Smaller strips for RAID 5, RAID 6 Larger strips for RAID 0, RAID 1, RAID 1+0

RAID migration with ACU

Using the ACU, you can also change the RAID level of the logical drive without downtime, maintaining data availability. You can increase raw data storage capacity, increase the number of spindles in a logical drive, or change fault tolerance (RAID) configurations with a RAID migration.

Drive roaming

Drive roaming lets you move disk drives and arrays while maintaining data availability. You can move one or more disk drives in a configured logical drive to a different bay position as long as the new bay position is accessible by the same controller. In addition, you can move a complete array from one controller to another, even if controllers are in different servers.

Drive roaming is an offline feature. There is no method for removing an array while the server is online and then moving it to a new physical location.

Mirror splitting and recombining with the ACU

You can use the ACU to split an array with one or more RAID 1 or RAID 1+0 logical drives into two identical new RAID 0 arrays. This is useful if you want to replicate a configuration or need to build a backup before performing a risky operation.

You can also recombine a split-mirrored array with the ACU. When you recombine a split-mirrored array, you lose access to all data on the second array.

The Smart Array Advanced Pack (SAAP) may be required to use the mirror splitting and recombining functions. For support information regarding mirror splitting on specific controllers, see the controller QuickSpecs.

NOTE:

You must split or re-mirror arrays when the server is offline and operating in the standard configuration mode of ACU.

Online drive flash

The present generation Smart Array controllers support online drive flashing, which saves time when updating disk drive firmware. Instead of taking the hard disk drive (HDD) offline before loading a new firmware image, you can download an updated HDD firmware image to the Smart Array controller and update all of the HDDs the next time you reboot the server.

Recovery ROM

HP Smart Array controllers store a redundant copy of the controller firmware image to protect against data corruption. If the active firmware image becomes corrupt, Smart Array controllers use the redundant firmware image and continue operating. The recovery ROM provides protection against power outages during firmware flashing.

Pre-Failure Warranty using S.M.A.R.T technology

We pioneered failure-prediction technology for disk drives by developing monitoring tests run by Smart Array controllers. Called Monitoring and Performance (M&P) or Drive Parameter Tracking, Smart Array controllers externally monitor disk drive attributes such as seek times, spin-up times, and media defects to detect changes that could indicate potential failure.

We worked with the disk drive industry to help develop a diagnostic and failure prediction capability known as Self-Monitoring Analysis and Reporting Technology (S.M.A.R.T.). As S.M.A.R.T. matured, we used both M&P and S.M.A.R.T. to predict disk drive failures.

S.M.A.R.T. has matured to the point that we rely exclusively on this technology to predict disk drive failure to support Pre-Failure Warranty. Since 1997, all HP SCSI, SAS, and SATA server-class disk drives have incorporated S.M.A.R.T. technology. S.M.A.R.T. disk drives inform the host when a disk drive is experiencing abnormal operation likely to lead to drive failure.

S.M.A.R.T. places the monitoring capabilities within the disk drive itself. These monitoring routines are more accurate than the original M&P tests because they have direct access to internal performance, calibration, and error measurements for a specific drive type.

HP Smart Array controllers scan the disk drive media during idle time and report, or repair, any media defects detected. The controllers recognize S.M.A.R.T. error codes and notify HP Systems Insight Manager (SIM) of a potential problem. HP SIM notifies administrators of drive failures.

Automatic data recovery with rapid rebuild technology

When you replace a disk drive in an array, Smart Array controllers use the fault-tolerance information on the remaining drives in the array to reconstruct the missing data and write it to the replacement drive. If a drive fails during the rebuild, the reconstruction fails and the data is likely to be permanently lost.

The present generation of Smart Array controllers includes rapid rebuild technology for accelerating the rebuild process. Faster rebuild time helps restore logical drives to full fault tolerance before a subsequent drive failure can occur, reducing the risk of data loss.

Generally, a rebuild operation requires approximately 15 to 30 seconds per gigabyte for RAID 5 or RAID 6. Actual rebuild time depends on several factors, including the amount of I/O activity occurring during the rebuild operation, the number of disk drives in the logical drive, the rebuild priority setting, and the disk drive performance. ACU lets you view the rebuild progress and set the priority for the rebuild operation.

Online spare

Smart Array controllers let you designate an unlimited number of drives as online spares. Multiple arrays can have the same spare drive as a global spare. Smart Array configuration utilities ensure that SAS disk drives are only spares for SAS arrays and likewise SATA disk drives for SATA arrays. During system operation, these spare drives remain up and running but perform no I/O operations. They are available in case one of the active drives in the array fails. Then an online spare drive may replace the failed drive.

If an active drive fails during system operation, the controller automatically begins rebuilding each fault-tolerant logical drive onto the online spare; no administrator action is required. Once the rebuild operation is complete, the system is fully fault-tolerant once again. You can replace the failed drive at a convenient time. Once you install a replacement drive, the controller will restore data automatically from the failed drive to the new drive. At that point, the original online global spare will return to standby mode.

Dynamic sector repair

Disk drive media can develop defects caused by variances in the drive mechanisms under normal operating conditions. To protect data from media defects, HP built a dynamic sector repair feature into Smart Array controllers.

Smart Array controllers perform a background surface analysis during inactive periods, continually scanning all drives for media defects. Smart Array controllers can also detect media defects when accessing a bad sector during busy periods. If a Smart Array controller finds a recoverable media defect, the controller automatically remaps the bad sector to a reserve area on the disk drive. If the controller finds an unrecoverable media defect and you have configured a fault-tolerant logical drive, the controller automatically regenerates the data and writes it to the remapped reserved area on the disk drive.

ECC protection

HP Smart Array cache modules use ECC technology to protect cache data. The ECC scheme generates 8 bits of check data for every 32 or 64 bits of regular data transferred. The cache module uses this information to detect and correct data errors originating inside the DRAM chip or across the memory bus.

Write caches

Using either the Flash-backed write cache (FBWC) or the battery-backed write cache (BBWC), Smart Array controllers can acknowledge a data transfer as “complete” before the data is physically stored in the disk drive. To improve disk write performance, data is temporarily stored in the write cache, which uses DRAM and is substantially quicker when compared to the disk drive.

Back-up power (FBWC or BBWC) is required for RAID controllers to perform operations such as write-back cache, array expansion, logical drive extension, strip size migration, and RAID migration.

Recovering data from battery-backed cache

If an unexpected server shutdown occurs while data is in the BBWC, Smart Array controllers automatically signal the memory chips to enter a self-refresh state and the controller moves to battery power or system

auxiliary power if available. An amber LED on the cache module or battery pack begins flashing to indicate that data is in the cache. Smart Array controllers automatically write this data to disk when the system powers up. The posted-write data in the cache will be lost if the batteries drain completely before the system powers up. The batteries automatically recharge when the system has power. Battery recharge takes between 30 minutes and 2 hours, depending on the remaining capacity level.

If the server fails, you can move the Smart Array controller and all of the drives to another server to allow writing the data in the cache to the drives.

If the controller fails, you can move the cache module containing posted-write data to a new Smart Array controller. You must attach the new Smart Array controller to the original drives to preserve cache data.

You should be aware of a special concern when using an embedded RAID controller with battery-backed cache. If the server board fails, the replacement board must be the same model server board. The controller type and drive bays must be the same. The cache module, battery pack, and drives move to the replacement system to extract the data from the battery-backed cache.

HP selection criteria for battery-backed cache

We designed the Smart Array battery cells, battery enclosures, and contacts to preserve the integrity of your data beyond the minimum specified backup duration. Battery life is about three years in typical server environments.

A dedicated battery microcontroller continuously monitors the HP Smart Array battery pack for signs of damage, including an open battery terminal, partial battery short, charge timeouts, and over discharge conditions. You can see the battery status with:

- LED
- Power-on self-test (POST) messages
- Event messages to the host
- ACU information pages
- Array Diagnostic Utility (ADU)
- HP SIM

The battery microcontroller automatically disables the battery-backed cache features any time it detects battery damage or the charge level falls below the required limits to achieve the specified backup duration. The battery microcontroller automatically restores battery-backed cache features when the microcontroller detects a replacement battery or when battery recharging is complete. High-end HP Smart Array controllers contain two batteries to protect against a single battery cell failure.

Types of batteries

HP Smart Array controllers use rechargeable Nickel Metal Hydride (NiMH) button cell batteries specifically designed for long life at server temperatures. The HP 650 mAh P-Series battery has the same form factor as previous versions and extends battery life up to 48 hours before you need to recharge it. Typical capacity for the HP Smart Array battery pack reduces by 5 to 10 percent over a 3-year period, depending on server temperature and number of discharge cycles.

NiMH cells do not contain harmful lead, mercury, or cadmium material. Additionally, NiMH chemistry does not suffer capacity memory effects that can lower battery capacity. For example, short discharge cycles reduce the memory capacity of Nickel Cadmium (NiCd) batteries. Lithium Ion (Li-Ion) batteries are typically smaller than NiMH batteries, but their capacity is permanently reduced at high temperatures and they are usually limited to 100 full discharge cycles.

Battery replacement

HP Smart Array controllers include serviceable battery packs that allow tool-free battery pack replacement with no need to replace either the Smart Array controller or the detachable cache module.

If it is not possible or desirable to replace the batteries, you have three options to disable write-back cache and avoid losing critical data:

- Use the ACU to adjust the read-and-write cache ratio to 100 percent read cache.
- Use the ACU to disable the array accelerator for each logical drive, which disables both read-ahead and write-back cache.
- Replace an existing RAID controller with a newer Smart Array controller model.

NOTE:

If the write cache is off, some write performance degradation may occur.

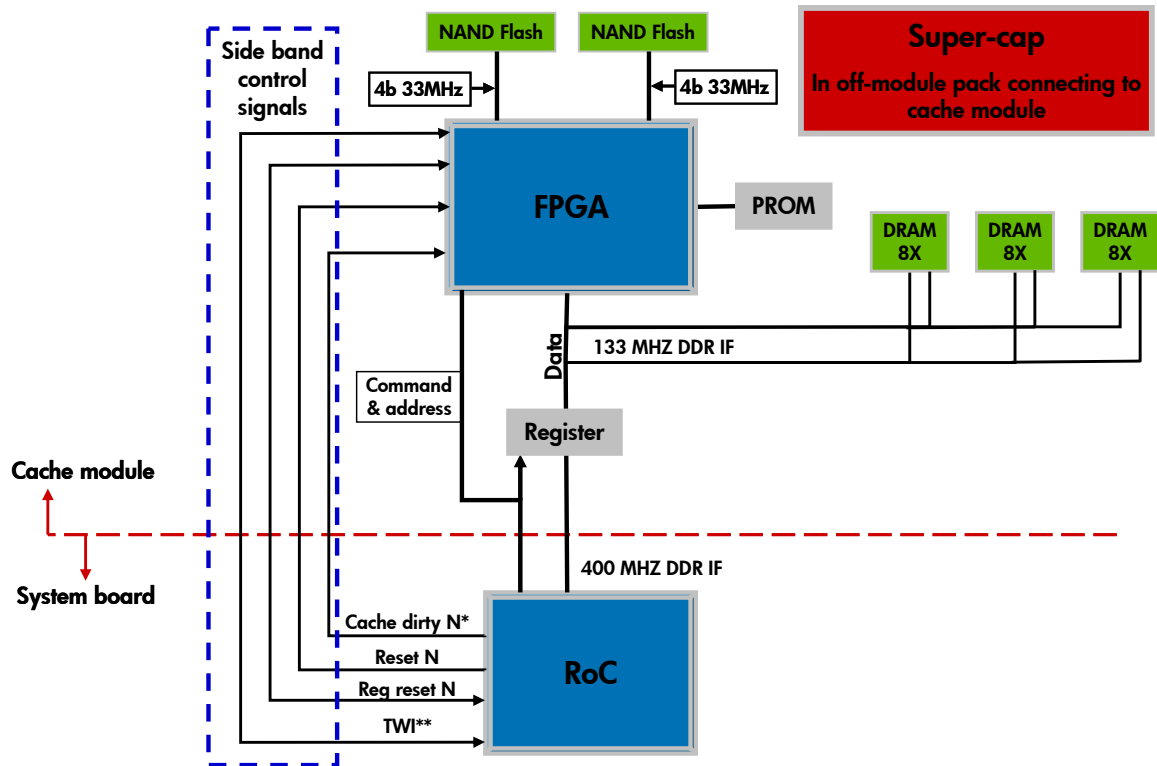
Flash-backed write cache

HP introduced the flash-backed write-cache (FBWC) system in the fourth quarter of 2009. The FBWC uses flash devices to retain cache data and super-capacitors (Super-caps) instead of batteries to provide power during a power loss. A BBWC must provide power during the entire power loss, while a FBWC only needs to provide power during the time to backup from DRAM to flash. The FBWC offers significant advantages over the HP Battery-backed write-cache (BBWC) system. Since the FBWC writes the contents of memory to flash devices, there is no longer a 48-hour battery life limitation, and the data posts to the disk drive on the next power up.

FBWC architecture

The FBWC DDR2 mini-DIMM cache module is designed for the current generation of PCIe2.0, SAS-based Smart Array controllers based on the PMC PM8011 max SAS SRC 8x6G RAID on a chip (RoC). The primary FBWC components consist of the cache module, Super-caps with integrated charger, and a RoC located on the system board, shown in Figure 1.

Figure 1. FBWC block diagram



* Cache tracks written over are "dirty."

** Two-wire interface (TWI)

The FBWC cache module contains a field programmable gate array (FPGA), DDR2 DRAMs, and NAND flash devices that can support up to 1 GB of DDR2 memory and up to 72 data bits (64 data bits plus 8 ECC bits). The FBWC can support up to 800 Mb/s data rate when the Smart Array controller is driving the DDR2 bus. When the FPGA is driving the bus during data recovery, the data rate is 266 Mb/s. The FBWC module connects to the Smart Array controller through a 244-pin mini-DIMM connector.

Super-capacitor

The Super-cap sub-assembly consists of two 35-Farad 2.7V capacitors, configured in series, providing 17 Farads at up to 5.4V. The charger maintains the Super-cap at 4.8V, providing the required amount of power to complete backup operations while extending the life of the Super-cap. The charger monitors Super-cap health and activates LED indicators on the FBWC module to warn of impending failure. The Super-Cap module uses the same form factor and housing as the HP 650 mAh P-Series battery used in the HP BBWC.

Recovering data from the FBWC

When system power is present, the FPGA on FBWC is in its idle state. In the idle state, the FPGA simply monitors the voltage statuses, the resets, and the control signals managed by the Smart Array controller. The FPGA's DDR2 I/O pins are in "tri-state," equivalent to a disabled mode, to avoid bus contention. When system power is lost, the FPGA waits for the Smart Array controller's clock enable signal to transition to low, signaling that the controller has stopped driving the DDR2 bus. The FPGA then assumes control of the bus and begins moving data from the DRAMs to the non-volatile flash memory. Upon the next power up, the FPGA then restores the cache by moving data from the flash memory to the DRAMs.

The Smart Array controller moves the data from the cache to the DRAMs, verifies that the data is correct, and then transfers it to the disk drives.

Software RAID

Today's operating systems offer basic support for RAID disks (called software RAID) to create and manage logical drives that do not contain the OS. Software RAID can require a significant amount of the server's resources and harm performance.

HP has developed the B110i software RAID solution based on the Smart Array firmware and using standard SATA controllers. The B110i SATA Software RAID supports ACU, ACU-CLI (command line tool), SNMP agents, and Web-Based Enterprise Management (WBEM) providers.

The Smart Array B110i also includes the following:

- Support for solid state disks
- LED support
- SATA drive firmware flashing (offline)

The B110i features an OS-specific driver from HP that uses the embedded ICH10R controller. The B110i supports RAID 0, 1, and 1+0 and a maximum of two logical drives. The B110i supports up to four 1.5 Gb/s or 3 Gb/s SATA drives. Administrators can move drives to a Smart Array controller in a seamless procedure that maintains the user data and RAID configuration.

The HP Smart Array RAID Hot Plug Advance Pack for the B110i provides hot-plug RAID support for the embedded SATA controller. The Hot Plug Advance Pack is available as a license key on cold plug and hot plug models.

Zero Memory RAID

Using Zero Memory RAID (ZMR), you can create a RAID 0+1 configuration without additional memory. ZMR uses memory embedded in the controller. ZMR has limited configurations: up to eight drives in Zero Memory Mode, or seven drives and one tape drive. ZMR does not include caching; but you can upgrade supported systems with a BBWC memory module that can increase performance.

Current generation (P410, P410i, P212, and P712) Smart Array controllers support ZMR for internal, direct connections only. The P212 controller does not include ZMR on the external connector. Modular Smart Array (MSA) products do not support ZMR mode. You can update the P212 controller to the 256 MB BBWC but not the 512 MB BBWC. Smart Array Advanced Pack is not available on Zero Memory configurations.

Smart Array Advanced Pack

HP Smart Array Advanced Pack (SAAP) firmware provides advanced functionality for the present generation of SAS/SATA-based Smart Array controllers. It enhances performance, reliability, and data availability. SAAP requires a license key for activation.

NOTE:

At a minimum, a 256 MB cache and battery kit, or FBWC, is required to enable the SAAP license key. SAAP is not available on Zero Memory Configurations.

RAID 6

RAID 6 protects against failure of any two drives. It requires a minimum of four drives, but only two are available for data. RAID 6 can tolerate multiple simultaneous drive failures without downtime or data loss. It is ideal for applications requiring large logical volumes, because it can safely protect a single volume of up to 56 disk drives. RAID 6 also offers lower implementation costs and greater usable capacity per U than RAID 1.

RAID 60

RAID 60 (RAID 6+0) is a nested RAID method that uses RAID 0 block-level striping across multiple RAID 6 arrays with dual distributed parity. RAID 60 lets you split RAID storage across multiple external boxes. It requires a minimum of eight drives, but only four are available for data.

Advanced Capacity Expansion (ACE)

Advanced Capacity Expansion complements the conventional capacity expansion feature of Smart Array controllers by letting you perform two new operations to either shrink or move existing arrays.

The Shrink Array operation allows customers to remove drives from an existing array. The Move Array operation allows you to transfer the contents of a disk array from one set of physical drives to a second set of physical drives.

Mirror splitting and recombining

There are two methods for safely breaking a RAID 1+0 mirror and rebuilding drives. The first method uses the ACU. The ACU splits an array that consists of one or more RAID 1+0 logical drives into two identical new arrays consisting of RAID 0 logical drives. The other option is a manual method, in which the volumes remain in their original RAID 1 or RAID 1+0 configurations after the split. Using the ACU does not require splitting other arrays attached to the controller at the same time. This functionality is useful when replicating a configuration or building a backup before performing a risky operation.

The Array Configuration Utility will show the Split Mirror task as an available option only under all of the following conditions:

- The version of ACU and the selected controller support the action
- The server is offline
- Any applicable SAAP keys have been installed

Drive erase

Securely sanitizing disk drives involves completely overwriting the drive data at the lowest level, below the OS file system and partition tables. With SAS and SATA drives, this means overwriting all of the logical blocks on the drive. HP Smart Array controllers with SAAP include an integrated Erase Drive feature that can erase data without installing additional software. The Smart Array Erase Drive function is available through the ACU for any logical or physical drive in an array.

The Erase Drive function operates by writing zeroes to every logical block on the logical or physical drive. This overwrites all file contents as well as the metadata, including all RAID controller, partition, and file system metadata. At a simplified level, erasing a drive is a serial write process, because the average sequential write throughput of the drive governs the speed. As a result, a drive erase can take several hours to complete on a moderate size drive such as a 500 GB midline SATA drive.

Storage pathway redundancy

Dual Domain support for SAS creates redundant pathways from servers to storage devices. These redundant paths reduce or eliminate single points of failure within the storage network and increase data availability. Dual Domain SAS implementations can tolerate HBA failure, external cable failure, expander failure, or failure in a spanned disk (JBOD).

SAAP-enabled Dual Domain is available on the P411 and P812 Smart Array controllers (SAAP license key required) when used with the HP StorageWorks D2600 or D2700 Disk Enclosures.

Other controllers do not support SAAP, but do support Dual Domain (no license key required). For example, the P700m Smart Array controller attached to either the HP StorageWorks 600 MDS or the HP StorageWorks 2000sa MSA storage enclosure supports Dual Domain. For these configurations you must install two HP 3G SAS BL Switch modules in a c7000 or c3000 blade enclosure.

Dual Domain support is available for the 3 Gb/s HP Smart Array P800 when attached to an HP StorageWorks MSA60/70 with the HP Dual Domain I/O Module Option. Dual Domain support requires HP Smart Array firmware v5.10 or higher and dual-port SAS drives.

Video on Demand (VOD)

Servers delivering video streaming services typically require significant amounts of disk storage. The HP Smart Array engineering teams have made significant strides in using HP technology to optimize Smart Array-based storage for VOD applications or any latency-dependent data flow. These improvements include optimizing for large block random I/O and decreased maximum latency.

Optimizing for large block random I/O improves the application server's ability to grab enough video data on each pass to maintain all the video streams in a rotating data retrieval scheme where a large number of simultaneous video streams exist. VOD performance improvement during this operation is a result of tuning the drive array, the file system, and the VOD application server so that each block request is stripe-aligned.

Decreasing the maximum latency of block requests to the Smart Array is one of the key goals for VOD. Physical Drive Elevator Sorting lets the controller re-order block requests to the physical drives to minimize the drive seek time. While this improves total performance, it can increase the maximum latency. The controller may schedule the execution of an earlier block request behind subsequent requests and execute the requests outside the VOD server's required time window. Turning off Physical Drive Elevator Sorting decreases the latency variability, which is what the VOD software wants. In a related manner, disabling Monitor and Performance and extending the Background Surface Scan Interval help eliminate spikes in latency that might occur during these background tasks.

Other improvements include changing the cache ratio to be 0 % read and 100 % write. Since VOD operations are 99% random, any read-ahead operation would penalize performance. You want to post the writes so that they have the least impact on the reads. We also found that in a multi-stream environment like VOD, the queuing mechanism in SAS drives delivers significantly better performance than SATA drives.

Another important way to optimize VOD performance in Smart Array-based storage is to enable Degraded Mode Performance (DPO). This decreases changes in latency between normal and degraded modes of operation. We added DPO specifically for VOD.

Storage management

HP Smart Array controllers include built-in intelligence to improve data storage management. This makes it easier for administrators to configure, modify, and manage storage. HP provides six utilities for managing an array on a Smart Array controller:

- The Array Configuration Utility is the main configuration tool for HP Smart Array controllers. It exists in three interface formats—the ACU GUI, the ACU CLI, and ACU Scripting. All three formats have separate

executables. Starting with version 8.28.13.0, ACU Scripting is a stand-alone application that is distributed with the ACU CLI application. The new ACU scripting executable (hpacuscripting) replaces the former executable (cpqacuxe) in all scripts.

- ORCA (Option ROM Configuration for Arrays) is a simple ROM-based configuration utility
- CPQONLIN (Compaq Online Configuration Utility) is a menu-based configuration utility specifically for servers using Novell NetWare
- HP System Insight Manager is a management suite for server and storage environments
- ADU (Array Diagnostic Utility) is a diagnostic and reporting utility for Smart Array controllers
- HP Smart Array SAS/SATA Event Notification Service (CISSESRV) is a service that provides event notification to the Windows Server 2003 and Windows Server 2008 system event log and HP ProLiant integrated management log

Table 2. Summary of HP configuration utility features

Utility features	ACU	ORCA	CPQONLIN
Creating and deleting arrays and logical drives	Yes	Yes	Yes
Assigning RAID level	Yes	Yes	Yes
Executable at any time	Yes	No	Yes
Uses a wizard to suggest the optimum configuration for an unconfigured controller	Yes	No	Yes
Sharing spare drives among several arrays	Yes	No	Yes
Assigning multiple spare drives per array	Yes	No	Yes
Setting strip size	Yes	No	Yes
Migrating RAID level or strip size	Yes	No	Yes
Configuring controller settings	Yes	No	Yes
Expanding an array	Yes	No	Yes
Creating multiple logical drives per array	Yes	No	Yes
Uses a graphical interface	Yes	No	No
Available in languages other than English	Yes	No	No
Describes configuration errors	Yes	No	No
Setting of boot controller	No	Yes	No

Array Configuration Utility

The ACU utility supports the HP MSA family of entry-level SAN products. This single tool seamlessly manages direct-attached storage (DAS) and SAN-attached storage.

The ACU runs online on Microsoft® Windows® Server and Linux® operating systems. If you use other operating systems, you can run the utility offline by booting the system from the SmartStart or Server Storage Support Software disc.

The ACU also contains a command line interface (ACU-CLI), offering a quicker way to deploy multiple servers by automating creation of arrays and logical drives.

Option ROM Configuration for Arrays

ORCA is an alternative method of viewing, creating, and deleting multiple arrays and logical drives during system power up. Although ORCA performs all array configuration tasks, it works best with simpler configurations. We recommend the ACU for more advanced array configurations.

CPQONLIN

CPQONLIN is a configuration utility that runs online on Novell NetWare. It functions like the ACU in this environment. Refer to the HP Smart Array controller documentation for more information about using CPQONLIN.

HP Systems Insight Manager

HP Systems Insight Manager is a client/server tool for integrated server environment management. Based on SNMP, it is capable of monitoring more than 1,200 system-wide parameters for performance and other operational characteristics of Smart Array controller storage. The program displays:

- Configuration information
- Device driver versions
- Controller firmware versions
- Pre-Failure Warranty information
- Operating statistics

HP SIM lets you look at low-level performance characteristics of Smart Array controllers. It monitors three basic Smart Array controller performance parameters:

- I/O commands per second
- Average command latency
- Local processor utilization

Analyzing these key parameters can help you fine-tune your configurations. HP SIM can chart performance over time for each of these parameters. A background task monitoring these parameters notifies HP SIM if a drive fails to meet certain factory-preset criteria. HP SIM alerts the administrator to the potential problem.

Array Diagnostics Utility

ADU is an in-depth diagnostic and reporting utility for all Smart Array controllers. The ADU quickly identifies problems such as incorrect versions of firmware, improperly installed drives, inappropriate error rates, and failed batteries on the array accelerator board.

The ADU displays a detailed analysis of the system configuration. If the cause is not apparent, the ADU can generate a full report for administrators to fax or e-mail to HP customer service for phone support.

Summary

HP Smart Array controllers are powerful I/O solutions for today's most demanding storage requirements. Our controllers provide solutions for all four primary data storage requirements:

- Capacity growth
- High performance
- Data availability
- Manageability

HP is the only server provider with a seamless storage solution set that spans the range from embedded Smart Array controllers in servers to plug-in PCI Smart Array controllers to SAN-attached Modular Smart Arrays storage. The tools used for managing and configuring storage are the same in all of those environments. Data sets are compatible across all of those environments. Smart Array controllers and HP universal drives allow easy migration of data from DAS to SAN.

For more information

Resource description	Web address
Server drive technology	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01071496/c01071496.pdf
HP Smart Array controllers	www.hp.com/products/smartarray
HP disk drive products	www.hp.com/products/harddiskdrives
HP Modular Smart Array external storage systems	www.hp.com/go/msa
HP ProLiant servers	www.hp.com/go/proliant
HP Smart Array Controllers and basic RAID performance factors	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c02249094/c02249094.pdf
HP SAS Technology	www.hp.com/go/serial
Configuring Arrays on HP Smart Array Controllers Reference Guide	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00729544/c00729544.pdf
Smart Array Advanced Pack	www.hp.com/go/SAAP
Serial Attached SCSI technology	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01613420/c01613420.pdf
Redundancy in enterprise storage networks using dual-domain SAS configurations	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01451157/c01451157.pdf
HP Smart Array controllers for HP ProLiant Servers User Guide	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01608507/c01608507.pdf?jumpid=reg_R1002_USEN
HP One-Button Disaster Recovery (OBDR) Solution for ProLiant	http://h18006.www1.hp.com/products/storageworks/drs/index.html
RAID 1+0: breaking mirrors and rebuilding drives	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00378986/c00378986.pdf

Call to action

Send comments about this paper to TechCom@HP.com



Follow us on Twitter: <http://twitter.com/ISSGeekAtHP>



© Copyright 2010 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

TC110202TB, February 2011

