November 2016

Demartek Evaluation of HPE 3PAR Persistent Checksum

The role of end-to-end data integrity in flash storage environments

Executive Summary

Data integrity is the term applied to the reliability of data written to storage and later retrieved for processing. The IT industry at large has several mechanisms to help make sure that data integrity is maintained in the software and hardware layers of application environments. Transaction or redo logging and checksums are common examples of application-level mechanisms, although their use cases are different. Hardware, such as storage systems, also employ integrity checks, which may include error correction codes (ECC), cyclic redundancy checks, and also checksum logic (similar to that performed by applications), along with other algorithms.

However, modern storage systems, in particular powerful, high-speed, flash-enhanced and all-flash systems, have increased in speed and capacity to such a degree that they can effectively outperform the rate at which media errors are caught and corrected. Likewise, storage network protocols like Gen 5 and Gen 6 Fibre Channel support such high bandwidths that a large number of recurring errors in the SAN can be delivered to storage media or applications very quickly, which may exacerbate problems. The challenge IT vendors face is identifying and remediating any data integrity issue before it manifests in detectable problems.

Some Tier-1 storage products are now equipped with integrity checksum processes. These processes are designed to verify that data transmitted from the server is identical to data actually received by the storage array and vice versa. HPE 3PAR StoreServ Storage arrays are one such solution. All 3PAR StoreServ Storage models, such as the all-flash HPE 3PAR StoreServ 8450 Storage array, feature built-in, end-to-end data integrity checking to eliminate the risk of bad data—a technology called 3PAR Persistent Checksum.

Unlike some other storage vendors, HPE builds Persistent Checksum logic into its 3PAR StoreServ Storage array controllers, HPE 3PAR StoreServ 16Gb Fibre Channel Adapters, and HPE StoreFabric Fibre Chanel Host Bus Adapters (HBAs) for media-to-host protection. Other vendors limit data integrity checking to high-end, Tier-1 hardware only, while expecting operating system and application vendors to write complementary checksum verification code into their own products. By contrast, HPE brings the entire checksum process into the HPE 3PAR StoreServ Storage and HPE StoreFabric storage networking solution, thereby lowering risk and making this technology available to midrange storage customers as well.

This report presents the results of evaluation work that HPE commissioned Demartek to perform with 3PAR Persistent Checksum to demonstrate the effectiveness of this solution at detecting and correcting data integrity issues. Demartek examined the function and performance of an all-flash HPE 3PAR StoreServ 8450 Storage array in an HPE StoreFabric Gen 5 SAN while intentionally introducing data transmission errors into the fabric.

HPE 3PAR Persistent Checksum

Demartek
November 2016

Key Findings

> Data integrity handling must change in response to technology and performance improvements in enterprise storage.

> HPE 3PAR Persistent Checksum effectively identifies and corrects data integrity issues in the SAN, preventing potential application impacts.

> HPE 3PAR Persistent Checksum logic had virtually no impact on the all-flash performance of the HPE 3PAR StoreServ 8450 Storage array.

Data Integrity Protection

Data is protected in many ways. At the software layer, intrusion detection and prevention, security protocols, rigorous quality testing, application monitoring, logging, checksums, and backups all play a part. Hardware vendors have also contributed by building even lower levels of integrity validation into their products. It is probably safe to assume that most storage product consumers take these protections for granted, and that might be a good thing—it means they work pretty well—but they are not foolproof. It is useful to have at least a basic awareness of what is going on as data moves across the SAN, to and from storage systems, media, and application servers to appreciate the value of robust integrity validation.

Uncorrectable Bit Error Rates on drives

There are many potential sources where errors may occur in the data stream. One that has been rigorously evaluated is the incidence of bit errors on drives. Bit errors are random inaccuracies which may occur in the drives when writing data, or bits that might randomly change over time. All drives are susceptible to bit errors and include logic to detect and correct them when they occur. However, this checking doesn't catch every error. The metric of bad bits to data written is called the uncorrectable bit error rate (UBER) and has been measured by drive vendors to be between $1x10^{14}$ bits and $1x10^{17}$ bits, depending on drive types (i.e.: consumer grade HDDs to enterprise SSDs). That is roughly one uncorrected bit error per 12 TB written all the way up to 12000 TB of data, again depending on drive type.

This may seem like a lot of data on the drive before a bit error will occur, but remember the UBER is an average, and devices are susceptible at any time. It's also important to point out that the larger the drive, the greater the likelihood that an error might occur. UBERs are per drive; enterprise storage systems with tens or hundreds of drives will see errors even more frequently, simply due to multiplying the number of drives in a single device.

As flash replaces spinning media in the enterprise data center, it is tempting to look at the lower incidence of UBER on enterprise SSDs and assume that this alone mitigates the risk to data integrity. However, this would be a mistake. While it is true that high-end flash drives have error rates that are orders of magnitude lower than HDDs, SSDs are also orders of magnitude faster than HDDs. At the same time, SSD capacities are increasing dramatically, with multiple terabyte drives becoming more and more common. A single SSD, or an all-flash array, that is being utilized anywhere near its potential capacity can hit its UBER just as quickly as a spinning drive, or even faster if it is busy with heavily transactional I/O. Quite simply, flash will do its job very quickly but is ultimately indifferent as to whether the data it is storing or serving is good data or bad data. Flash may be the future of the enterprise data center but it is not a panacea when it comes to error detection and prevention. Not on its own, anyway.

Bit errors in the Fibre Channel SAN

Drive errors aren't the only place for potential errors to creep into the data stream—they are just the easiest to measure. For reasons mentioned earlier, bit errors can also occur at just about any other point within a SAN. Fibre Channel is commonly considered to be a lossless protocol. Perhaps this is true in an idealized environment where every cable is tightly coupled, power never fluctuates, temperature and humidity remain strictly constant, the entire data center is shielded from external radiation and vibration, and system components never degrade over time. Of course, this data center doesn't exist and experienced storage administrators and architects know that FC does indeed experience occasional loss and error, although in all fairness to the protocol, it is a rare occurrence outside of component failure or misconfiguration.

Ensuring data integrity via T10 DIF

In 2003, the T10 committee (the committee for SCSI Storage Interfaces) of the International Committee for Information Technology Standards[1] proposed additional end-to-end data integrity protection (from host CPU to media). The standard includes the addition of a Data Integrity Field (DIF) in computer storage devices as a way for computer systems to detect changes in data that occur during transmission from a host to media or during writes to the media. DIF added an additional 8 bytes to the then standard 512-byte drive sectors for this error checking (or *checksum*) data.

Since this time, some enterprise storage vendors have incorporated the use of DIF into their high-end products, but the presence of these extra bytes is only half the solution. Complete end-to-end validation requires that the host side of a computer system also support DIF checking, to create the checksum data on writes to storage and to validate the checksum when reading data. As of the date of this report, operating system support is somewhat sporadic. A handful of enterprise applications also have DIF support that can be enabled to verify data integrity between the application engine and the storage target, but its use is not universal.

3PAR Persistent Checksum—HPE's approach to ensuring end-to-end data integrity

3PAR Persistent Checksum is a built-in feature of the HPE 3PAR OS and HPE 3PAR Gen5 Thin Express ASIC, 3PAR StoreServ Fibre Channel Adapter, and StoreFabric HBA firmware and drivers (supported on common enterprise operating systems). All HPE 3PAR StoreServ 8000 and 20000 Storage systems support end-to-end protection (from the host), while previous generations of 3PAR, like the HPE 3PAR StoreServ 7000 and 10000 support protection on the drives.

3PAR Persistent Checksum is unique as HPE works with its storage networking supply chain vendors to integrate HPE's implementation of T10 DIF into the HPE SAN fabric via HPE StoreFabric FC HBAs and driver software. These HBAs work in concert with HPE 3PAR StoreServ Storage arrays (Figure 1) to automate data integrity verification, end-to-end, from the hosts through to the point where data is written to storage media. Persistent Checksum is one of a number of native, end-to-end error checking features delivered by an HPE 3PAR StoreServ Storage environment[2] and another component of HPE's infrastructure aware flash storage strategy[3].

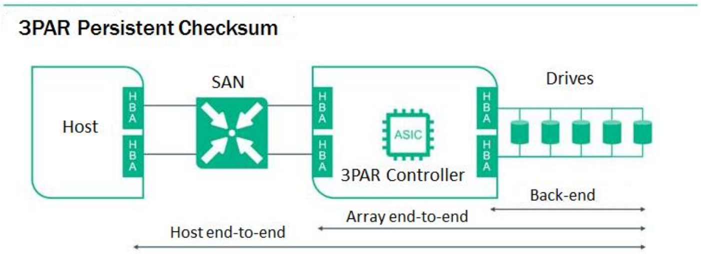


Figure 1 – All HPE 3PAR StoreServ 7000, 8000, 10000, and 20000 systems have back-end T10 DIF protection. 3PAR StoreServ 8000 and 20000 systems also include built-in array end-to-end protection and provide full host-to-media protection when supported HBAs are installed on the server.

[1] http://incits.org/

[2] HPE, "HPE 3PAR StoreServ Storage: designed for mission-critical high availability", https://www.hpe.com/h20195/v2/GetPDF.aspx/4AA3-8316ENW.pdf (December 2015)

[3] https://community.hpe.com/t5/Around-the-Storage-Block/Infrastructure-Aware-Flash-Storage-A-Healthy-Choice-for-Your/ba-p/6898894#.V-By9q1xGtZ

Putting 3PAR Persistent Checksum to the test

We put HPE 3PAR Persistent Checksum to the test by injecting bit errors into an active FC data stream. Inducing errors on drives is difficult, so we opted to introduce errors by running the Fibre Channel data stream through a Viavi Solutions Inc. Xgig® 1000 analyzer/jammer and intentionally changing the bits in bi-directional data frames while ensuring that SCSI CRC checks would not detect the changes. If this is difficult to visualize, then in layman's terms, we intentionally scrambled the bits in certain FC data frames in a way that did not trigger default error correction policies. The frames affected were random, but the interval that these bad frames were inserted was a constant.

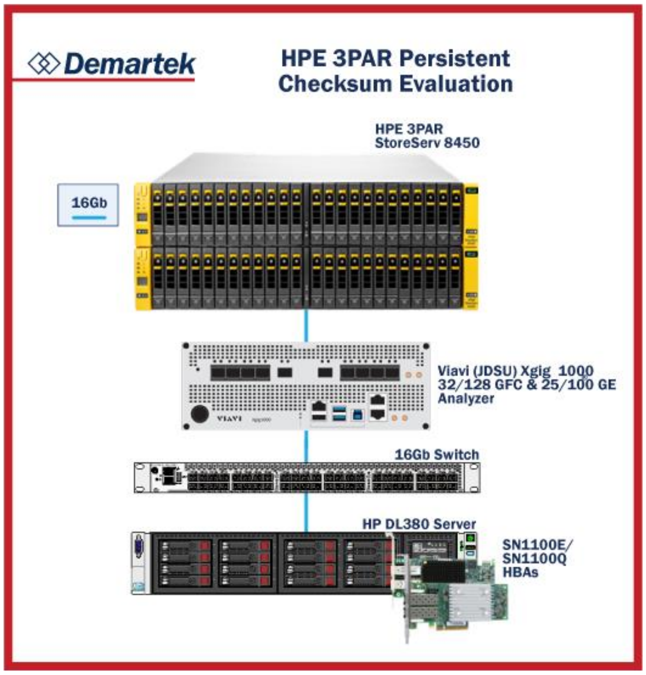
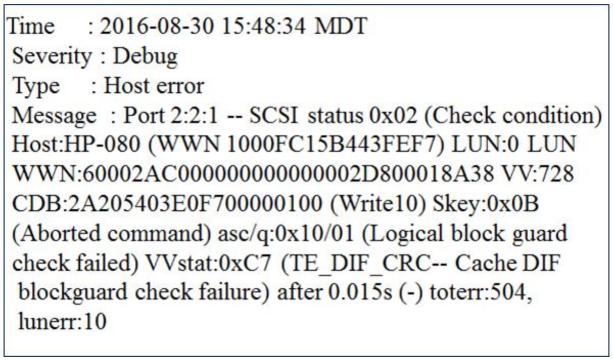

Demartek
HPE 3PAR Persistent Checksum Evaluation
HPE 3PAR StoreServ 8450
16Gb
Viavi (JDSU) Xgig 1000 32/128 GFC & 25/100 GE Analyzer
16Gb Switch
HP DL380 Server
SN1100E/ SN1100Q HBAs

Figure 2 – Evaluation environment topology

To create a stream of data for this analysis, a Microsoft Windows Server virtual machine supporting an OLTP database application was deployed in a VMware ESXi environment. The VMware datastores holding the virtual machine and application files were both served from the HPE 3PAR StoreServ 8450 Storage. This allowed the scrambled frames to affect either the database or the virtual machine operating system.

With the 3PAR Persistent Checksum feature enabled on the storage system and the StoreFabric FC HBAs, the HPE 3PAR OS detected and reported damaged FC frames. A sample error message is displayed in Figure 3. These bad I/Os were then prevented from disrupting the integrity of the data written or processed by the application.



```
Time     : 2016-08-30 15:48:34 MDT
Severity : Debug
Type     : Host error
Message  : Port 2:2:1 -- SCSI status 0x02 (Check condition)
Host:HP-080 (WWN 1000FC15B443FEF7) LUN:0 LUN
WWN:60002AC00000000000002D800018A38 VV:728
CDB:2A205403E0F700000100 (Write10) Skey:0x0B
(Aborted command) asc/q:0x10/01 (Logical block guard
check failed) VVstat:0xC7 (TE_DIF_CRC-- Cache DIF
blockguard check failure) after 0.015s (-) toterr:504,
lunerr:10
```

Figure 3 – HPE 3PAR event log entry for a T10 DIF error

When 3PAR Persistent Checksum was disabled, these errors were not detected and the bad data was not corrected in the application or the virtual machine datastores. In fact, when we intentionally injected errors in the data frames while 3PAR Persistent Checksum protection was manually disabled, the application virtual machine operating system both crashed within about two and half minutes of beginning the error insertion (Figure 4).

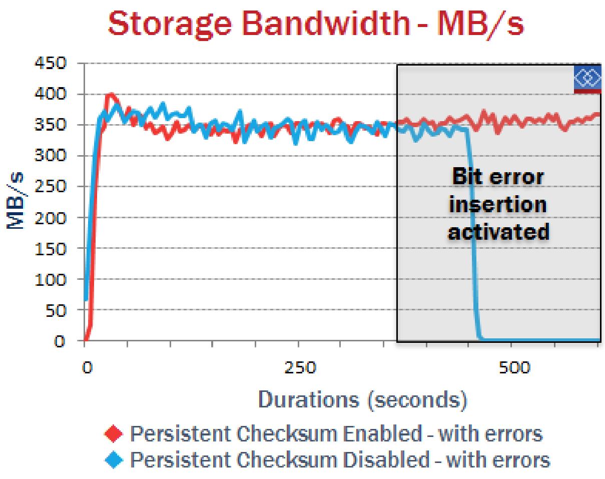


Figure 4 – Application and VM failure caused by bit error

By moving error handling to a low level in the storage stack, HPE 3PAR Persistent Checksums prevented integrity errors from even reaching the software stack. The application and VM data were protected by the 3PAR array and StoreServ HBA from the induced integrity failures.

In the spirit of full disclosure, we configured the analyzer/jammer to send an altered frame every five seconds, a bit error rate on the order of 1x10^{12}, which is extremely high. Placed in a context that is easier to conceptualize, this is along the lines of one error for every 1.75 GB of data (or, in terms of throughput, we were forcing one bad frame roughly every 210,000 IOPS). This example serves as an effective demonstration of a worst case scenario, and the preventative benefit of 3PAR Persistent Checksum technology.

To make an application owner's bad day even worse, with 3PAR Persistent Checksum disabled, there is obviously no record of any error in the storage system's event log. The owner of the data would have had no indication of where problems were coming from, only that the database was beginning to have transaction errors for some reason, followed by an application failure. This eliminates a tool that the storage administrator and application owner would otherwise have at their disposal for analysis. Or perhaps more accurately stated, this eliminates a tool that automatically corrects a type of data integrity issue, allowing bit errors on the drive or within the SAN to be removed from consideration in the event that something is causing the application or operating system to error.

What is the impact of 3PAR Persistent Checksum on storage system performance?

Flash is a powerful storage media and has clearly changed the way businesses think about data storage. It's probably safe to assume that most all-flash systems are not pushed to their operational limits by a single application. Cost-conscious businesses may choose to leverage that excess capacity by deploying additional applications or services. One of the great advantages of all-flash arrays like the HPE 3PAR StoreServ 8450 Storage is the ability to support multiple workloads simultaneously without compromising service levels. Of course, deploying multiple workloads on the same storage system, and storage media, increases not only the amount of data read and written, but the number of places from which these activities occur, which also increases the potential sources of bit error within the SAN.

Usually, the implementation of any storage enhancement feature is going to include some impact to the performance delivered by a storage system. Application owners, particularly those of applications with heavy I/O demands, or who share a storage system with other tenants, should be curious about whether taking advantage of a feature like 3PAR Persistent Checksum is worth the trade off in performance.

We evaluated this by running the same OLTP database workload we used to create the data we intentionally damaged earlier. This workload didn't place much stress on the HPE 3PAR StoreServ 8450 Storage array controllers in the first place, which speaks well to the processing power of the array in general. There also was no measureable difference in average controller CPU use regardless of whether 3PAR Persistent Checksum was enabled or not (Figure 5). We attribute this to HPE's choice to offload the checksum logic to the 3PAR ASIC.

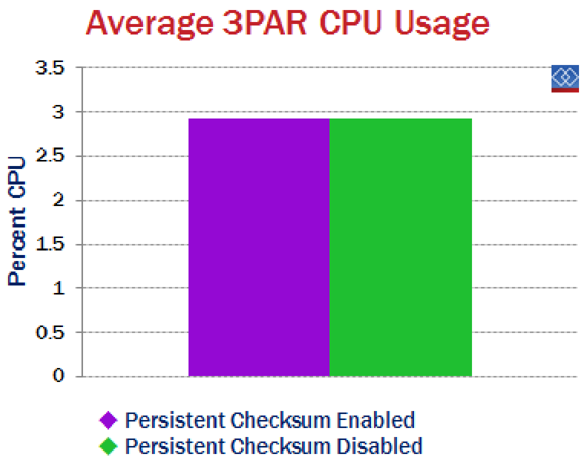


Figure 5 – 3PAR CPU use during workload execution

We considered I/O impact as well by measuring the storage system performance with 3PAR Persistent Checksum enabled and disabled while both inserting bit errors and with unadulterated data. In all cases, storage bandwidth was essentially unchanged (Figure 6), regardless of whether the system was performing checksum calculations or not, or correcting bit errors when they occurred.

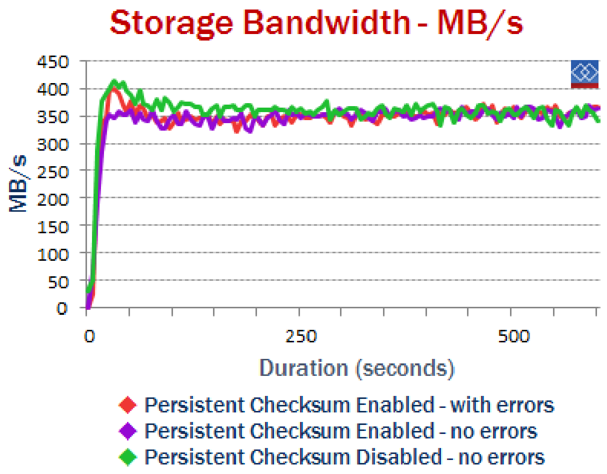


Figure 6 – Bandwidth of OLTP workload with Persistent Checksum enabled and disabled

As expected, IOPS performance demonstrates the same pattern as bandwidth, with no appreciable impact in steady-state IOs per second regardless of whether 3PAR Persistent Checksums are on or off (Figure 7).

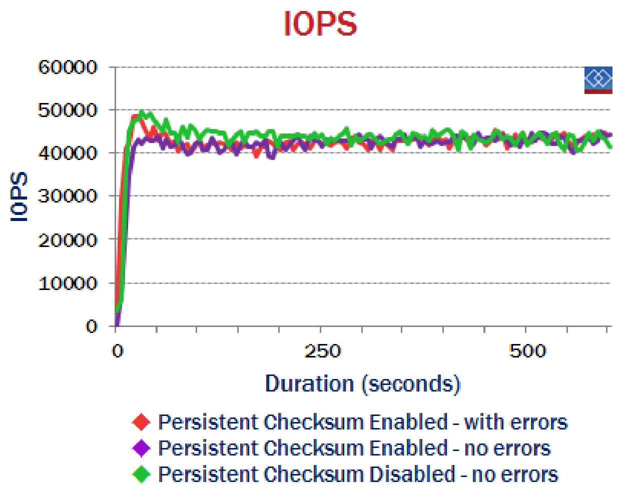


Figure 7– IOPS of OLTP workload with Persistent Checksum enabled and disabled

Bandwidth and IOPS aren't the only measure of storage system performance. Latency (or the lack thereof) has become increasingly important as a service level metric for transactional applications. Therefore, we also evaluated the time it took for the system to process the I/O requests. There was no statistically significant difference in average I/O latency between I/O running with Persistent Checksum or without (Figure 8).

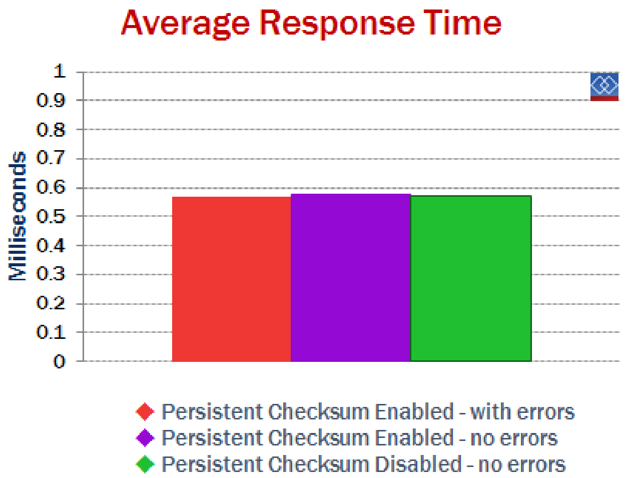


Figure 8 – I/O latency with Persistent Checksum enabled and disabled

From this data, we conclude that for the workload evaluated, 3PAR Persistent Checksum has no net negative impact on either the amount of work done by the 3PAR StoreServ 8450 Storage array or the time in which it took to perform each I/O comprising that work.

Summary and Conclusion

HPE 3PAR Persistent Checksum is a built-in feature that is enabled by default on HPE 3PAR StoreServ Storage arrays and can be extended to HPE 3PAR StoreServ 16Gb Fibre Channel Adapters on the host. This technology leverages the T10 DIF standard, as do a number of high-end storage products from several vendors. However, HPE includes this feature standard on its entire 3PAR StoreServ Storage family—spanning from midrange to high-end systems—effectively elevating the level of service that all customers receive from their storage.

Flash storage is widely considered a panacea for application performance. While the validity of this viewpoint does depend on some application planning and tuning, it is undeniable that flash has changed the way the industry thinks about data storage. When combined with high-speed storage networking, such as Gen 5 Fibre Channel (16GFC), flash storage systems can have a radical impact on business operations. However, some of the best features of all-flash storage systems also argue the strongest for safeguards like 3PAR Persistent Checksum.

The extreme performance of flash devices, coupled with the increasing size of drives and modules, may increase the frequency of bit errors on the media. Add in the trend of co-hosting legacy storage clients onto fewer, more robust all-flash systems, and the potential for errors increases even more simply because of the number of places on the SAN where things can go wrong. Keeping the fabric healthy negates the impact of a portion of infrastructure-related failures. 3PAR Persistent Checksum is one of many attributes of HPE's infrastructure-aware approach to the all-flash data center. This approach takes into consideration the broader implications of flash in the enterprise data center rather than approaching this technology as simply a point solution for application acceleration. In many cases, 3PAR Persistent Checksum comes with the storage an HPE 3PAR StoreServ customer already owns, with no licensing or activation required.

The IT industry at large recognizes that data demands are growing at a geometric rate, and inevitably no system is completely lossless or without potential for data errors (not even considering possible malicious sources of data manipulation). Hardware vendors are taking action and some operating systems and applications are now T10 DIF-aware, but this is by no means universal. HPE's solution stands out by the nature of its end-to-end design. Operating systems and applications do not need to be T10 DIF-aware in a StoreFabric SAN and yet are still assured of protection from data integrity errors.

Should 3PAR Persistent Checksum drive a business's storage system purchasing decision? That is ultimately for each business to decide and is probably the wrong question to be asking. The real question should be "How valuable is data, and what solution is best positioned to protect that data?" Mid-range, all-flash storage customers get an extra layer of protection by default with HPE 3PAR StoreServ Storage without sacrificing performance. And performance does drive purchasing decisions. We would consider this at least worthy of some consideration when evaluating upcoming storage acquisitions.

As a closing thought, if a business already has a current model 3PAR StoreServ Storage array, they've already got 3PAR Persistent Checksum, whether they know it or not. If this is you, it might be worth examining your event logs to see if HPE's foresight has already provided you with extra data protection you didn't even know you had.

The most current version of this report is available at http://www.demartek.com/Demartek_HPE_3PAR_Persistent_Checksums_2016-11.html on the Demartek website.