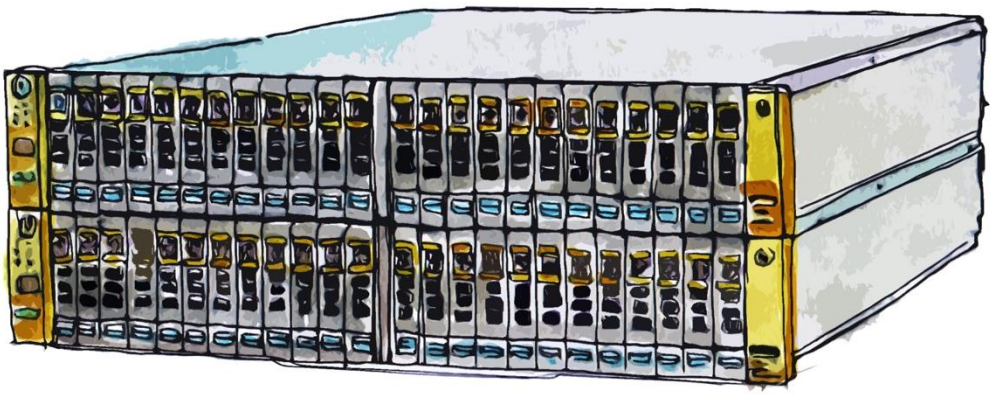
Langton Blue

Author: Chris M. Evans

HPE 3PAR ADAPTIVE DATA REDUCTION

A competitive comparison of array-based data reduction



First Published: January 2017
Latest Update: January 2017
HPE Document ID – 4AA6-6256ENW
LB Document ID – LB1CD0095
Release 1

Table of Contents

1- Executive Summary

Modern storage arrays can't compete on price without a range of data reduction technologies that help reduce the overall total cost of ownership of external storage. Unfortunately, there is no one single data reduction technology that fits all data types and we see savings being made with both data deduplication and compression, depending on the workload. Typically, OLTP-type data (databases) work well with compression and can achieve between 2:1 and 3:1 reduction, depending on the data itself. Deduplication works well with large volumes of repeated data like virtual machines or virtual desktops, where many instances or images are based off a similar "gold" master.

Storage appliance vendors need to be considering the use of a range of data reduction technologies (including compression and deduplication) as part of a data reduction strategy, especially with hybrid or all-flash based systems. Data reduction reduces the overall effective cost of deploying a flash-based array, without compromising on the performance capabilities. Efficient data reduction can reduce the cost of all-flash systems to a level that is comparable to or cheaper than traditional HDD-based arrays, especially when considering TCO that includes space, power and cooling costs.

With the release of 3PAR OS 3.3.1, HPE has again extended the 3PAR platform, introducing compression as another feature of what is now called Adaptive Data Reduction (ADR). The four stages of ADR progressively improve the efficiency of data stored on physical media, through the use of Zero Detect, deduplication, compression and Data Packing. Each of these features uses a combination of a 5th generation hardware ASIC, in-memory metadata and efficient processor utilization to deliver optimal physical space utilization for today's hybrid and all-flash systems.

HPE continues to demonstrate the ability to extend and enhance the 3PAR platform, providing ongoing value for customers through one of the most effective storage architectures in the industry.

2- Introduction

2.1- Objective

This report looks at the implementation of Adaptive Data Reduction and specifically data compression on the HPE 3PAR StoreServ storage platform and compares the features and functionality to that offered by equivalent products in the market place today, with reference to all-flash solutions.

2.2- Audience

Decision makers in organizations looking to deliver highly efficient storage solutions based on centralized all-flash storage will find this report provides an understanding of the benefits of using compression technologies to reduce physical data storage.

2.3- Contents of This Report

- **Executive Summary** – a summary of the background and conclusions derived from Langton Blue research and analysis.
- **The Need for Data Reduction** – a discussion on the need to implement efficient data storage functionality.
- **Deduplication versus Compression** – an outline discussion on space reduction technologies and their practical application.
- **HPE 3PAR compression: Deep Dive** – a detailed discussion on the implementation of compression within the 3PAR platform.
- **Competitive Analysis** – a comparison with competing vendor products.
- **Conclusions** – a summary of the findings comparing HPE 3PAR StoreServ with competing platforms.

3- The Need for Data Reduction

There's no doubt that data stored and retained by enterprises is growing at an incredible rate, with growth being seen in both structured and unstructured content. Storage system vendors are making every effort to make their products more attractive on a $/GB basis, as this is typically the way in which customers look at the cost of acquiring storage infrastructure. As a result, techniques such as data deduplication and compression have been used to make raw storage more attractive when looking at effective capacity figures. The main reasons for implementing data reduction techniques include:

- **Physical Space Reduction** – storage systems can store more logical data as the host sees it, compared to physical data on disk.
- **Cost reduction** – reducing the amount of physical storage in use translates directly into hardware cost savings.
- **Cost Avoid** – implementing data reduction within the storage array avoids paying for data reduction licences higher up the stack (e.g. within the database platform).
- **Wear Reduction** – flash media has a limited write lifespan so reducing the number of physical updates to SSD media means drives last longer.

3.1- Optimising with Flash

Flash storage provides significant benefits when implementing data reduction technologies. The low latency and high throughput characteristics of flash means data reduction can be achieved with minimal impact to the application. The ability to implement data reduction has also been made possible due to the significant increase in processing power of today's Intel processors used in modern storage systems, in conjunction with the performance of internal backplanes and the reducing cost (and greater use) of system memory (DRAM).

3.2- Choosing the Right Location

The implementation of data reduction techniques is not limited to external storage arrays. It's also possible to implement both compression and deduplication at the application or operating system levels (where supported). Software features such as Oracle's Advanced Compression provides the ability to implement data compression within an Oracle database. However, this feature isn't free; in addition, it consumes host processor resources and may mean purchasing a higher specification of server, also increasing Oracle licence costs. Bear in mind also that data reduction technology isn't generally available in operating systems or applications (although there are some exceptions, like NTFS compression on Windows).

Host-based data reduction can also be used to reduce the amount of used space at the file system level. With this implementation, data reduction software has to be deployed and managed on each host, including tracking licence usage. Host-based data reduction also consumes resources (both processor and memory) and doesn't provide the ability to make savings across many servers with similar data. Global deduplication within the storage array, for example, can result in significant storage savings, because multiple O/S instances will deduplicate down to a single instance.

3.3- Optimisation Techniques

Within the HPE 3PAR architecture there are four main techniques that can be used to reduce the amount of physical space used compared to the logical space assigned to applications. Collectively they are part of an overall strategy known as Adaptive Data Reduction.

- **Zero Detect** – when a data stream contains zero data, Zero Detect technology can be used to identify and eliminate the need to write this data to disk. This can happen when (for example) a data volume or database file is initially formatted or when data is logically released with commands like UNMAP (sdelete command).
- **Deduplication** – the process of deduplication looks to identify duplicated pieces of information within a data set and reduce it to a single copy. Each logical reference to the data points to only a single physical copy.
- **Compression** – this process also looks to identify information in a data set that can be optimised and to reduce the amount of data stored, by using multiple algorithms (for example pattern recognition) through reducing the size of the original data content.
- **Data Packing** - a process that optimises the storage of compressed pages by storing multiple compressed page images within a single 16KiB page, thereby increasing capacity efficiency, performance and endurance.

Thin provisioning and Zero-Detect technologies have already been covered in other white papers (see References on page 16 for details). In this paper we will focus on compression and deduplication technologies and their application to workload profiles in the data centre.

4- Deduplication versus Compression

Deduplication and compression both look to reduce the amount of physical space stored on disk compared to the logical amount of data written or seen by the host. However, the way in which these technologies work is subtly different and this affects the savings that can be achieved from differing workloads.

4.1- Deduplication

The deduplication process typically looks to identify incoming blocks of data that are identical to data already stored on the storage system. If a match can be made, then the new data won't be stored to disk. Instead, metadata will be used to create pointers that match the new logical reference of the data within the volume to the existing copy of the data on disk. The more references to a single page that can be made, the greater the space savings that can be achieved.

Deduplication algorithms use system metadata to keep track of data already stored. This is achieved by creating a hash value for each write I/O, either based on fixed boundaries (e.g. 4KB, 16KB) blocks or using variable block sizes. As data is ingested into the system, a hash is created and compared to the existing hash values already stored. A match indicates the data is already stored in the system, no match means the data has to be physically stored on disk and the list of hashes stored is updated. The way in which hash values are generated can be computationally expensive. In the 3PAR architecture, hash value calculations are offloaded to the embedded hardware ASIC, with collisions managed in metadata.

4.2- Compression

Compression looks to reduce the amount of data in a data stream by applying compression algorithms that use a range of techniques, including lookup table, run-length encoding and dictionaries. The compression process can be either lossless or lossy, meaning the original data is either re-creatable or not re-creatable from the compressed copy. Lossless compression is obviously desirable in storage systems, whereas lossy compression is used on media (e.g. audio or images) where the original uncompressed data isn't required or some loss in quality is acceptable.

Compression algorithms are typically processor intensive, whereas deduplication is both processor and memory intensive, requiring more metadata to track the hash references.

4.3- Data Reduction and Workload Types

Deduplication and compression processes work best on different types of data. Typically, compression works best with OLTP transactional based data that has already been normalised at the application layer. This is because opportunities exist to find repeated patterns in the data but those patterns aren't repeated elsewhere (the normalisation process in databases reduces repeated records in tables to a single instance).

By comparison, deduplication works well on data with lots of duplication. This includes virtual server and virtual desktop environments where components such as the operating system are typically cloned or very similar in structure.

In mixed workload environments, which account for the majority of modern environments, benefits are achieved from implementing both compression and duplication on the same data sets, as some data will contain duplicate data and some will naturally compress well.

5- HPE 3PAR Adaptive Data Reduction: Deep Dive

5.1- Background

The HPE 3PAR StoreServ architecture is based on a Mesh-Active Cluster that is comprised of multiple controller nodes and disk shelves. All system resources participate in access to all volumes, whether that's host ports, CPUs, cache or even drives which also have active access to multiple controllers.

HPE 3PAR StoreServ is a completely abstracted architecture that divides physical storage devices (hard disk drives or solid-state drives) into 1GB units called chunklets. Chunklets are combined together across physical disks to create logical disks (LD) that apply RAID protection. Each LD can have a different data protection mechanism that is independent of the physical devices, meaning there is no rigid RAID structure in place. Virtual volumes (VVs) are created from logical disks and use a page size granularity of 16KiB. The high level of abstraction in the 3PAR StoreServ design means that all components are involved in data access, reducing the risk of bottlenecks and making use of every available resource.

HPE 3PAR Adaptive Data Reduction defines a number of features that include Zero Detect, deduplication, compression and Data Packing. Hardware acceleration of HPE 3PAR Adaptive Data Reduction technologies is achieved through the use of a custom ASIC chip, now in its 5th generation. The 3PAR ASIC provides line speed features such as RAID parity calculations, cluster interconnects, Zero Detect, hash calculations for deduplication and other functions used in delivering the 3PAR capabilities.

5.2- 3PAR Deduplication

Before we look at 3PAR compression, it's worth taking a moment to look at the implementation of 3PAR deduplication and how this works in conjunction with Thin Provisioning technology. Deduplication identifies data writes that would duplicate data already written to the 3PAR system. The 3PAR ASIC is used to create a hash key and validate whether newly written data already exists on the array. Duplicate data is discarded (subject to the write being tracked in metadata associated with the user volume), while unique data is written to SSD. A feature known as Express Indexing is used to ensure fast metadata lookup to validate the hash values calculated by the ASIC. This uses multi-level metadata tables to quickly (and with excellent scalability) determine if a data block has been seen previously. If data is identified as a duplicate, the ASIC is used to do a bit-level XOR comparison of the new and existing data to be 100% sure the data is unique.

5.3- 3PAR Compression Walkthrough

3PAR compression complements the existing deduplication technology to provide a further layer of data reduction. Once new data has gone through the deduplication process and been identified as unique, it is processed for compression. There is no benefit in compressing data before duplicates have been identified, as this wastes CPU resources on data that may be subsequently discarded.

The 3PAR compression implementation uses a feature called Express Scan. This looks at the data to be compressed and determines whether the content can offer compression savings. If it can, the compression process continues. If it can't then compression is abandoned in favor of saving CPU cycles, resulting in a significant reduction in the processor load associated with compression.

Compression uses an algorithm known as lz4 to perform the compression process. This algorithm is extremely efficient compared to other techniques, especially at decoding, which is where performance is important. The writing and storing of data can be done as an asynchronous process to the host I/O and has always been done like that as part of the caching of write I/O in shared storage systems. However, read requests, particularly random reads are synchronous and need to be completed immediately, so a fast decode (or inflate) algorithm is essential. Lz4 is a mature technology that is widely used across the industry in operating system distributions, file systems, databases and some storage arrays.

5.4- Data Packing

As discussed previously, 3PAR StoreServ uses 16KiB pages to store virtual volume data. When pages are compressed, the resulting data will obviously be shorter in size than the original page and not media efficient (e.g. 4KiB blocks for flash). 3PAR Data Packing optimizes physical space by storing multiple compressed pages within a single 16KiB page and using the existing exception tables structure to track the data. Compressed pages are stored in an area called SD space; metadata at the beginning of the SD page tracks the offset of each piece of compressed content.

In addition to the space savings made, Data Packing allows overwrites by the host to be stored in place, reducing the amount of system garbage collection that is needed. By aligning on 16KiB, data is read and written in a "flash friendly" way, improving both performance and media endurance and positioning 3PAR efficiency for the use of future storage class memory products.

5.5- Selective Data Reduction

3PAR StoreServ systems provide the capability to enable data reduction technologies at the LUN or volume level with a feature known as Selective Data Reduction. The ability to allow the customer to choose whether to implement space saving features allows flexibility in determining the best balance of the overhead of using space saving (such as a slight increase in latency) versus the savings made. For example, certain data types, like encrypted data or data that has already been compressed (media files) will see little or no benefit from being placed on a compressed volume. It makes sense to turn off this feature for this kind of data and not waste CPU cycles trying to gain additional savings. Having a single solution with this flexibility provides customers with choice and removes the need to deploy multiple hardware platforms.

5.6- Integration with Existing Data Reduction technologies

3PAR compression is supported with existing data reduction technologies including Thin Provisioning and 3PAR deduplication. Replication of compressed volumes is supported, as the data is only compressed on back-end media. This enables volumes to be replicated to targets that are not flash-based or running on older hardware that doesn't support compression.

5.7- Enabling 3PAR Compression

3PAR compression is available for all HPE 3PAR Gen5 platforms (8000 and 20000 series systems). The initial implementation is only supported for volumes stored on flash, whether as part of a hybrid or all-flash system. Compression is available to customers at no additional cost and is included in the base OS license for the 3PAR platform.

5.8- Space Savings

The amount of space saved through data reduction is dependent on the data itself. Typically, for each data type, different space savings processes can be applied. As an example, the following data types will see benefits from specific technologies although you may experience different results:

- Databases (e.g. Oracle) – compression (around 2:1 saving)
- Exchange Server – compression and deduplication (around 1.5:1 saving)
- Compressed Video – Thin Provisioned
- Virtual Desktop – compression and deduplication (2:1 and greater saving).

6- Competitive Analysis

Data compression has been a key feature of all-flash arrays that have been developed and released onto the market over the last few years. The use of data reduction technologies is seen to provide an advantage in pricing in an industry where customers look at $/GB as one of the basic TCO measurements of storage systems. Ultimately this makes data reduction a "table stakes" feature set for any vendor wanting to enter the all-flash market. Of course this also applies to hybrid systems, or at least the part of those systems deploying volumes on flash media.

In order to perform a comparison, a number of the leading vendors in all-flash solutions have been chosen for evaluation. These include mix of start-ups and established companies in the industry.

6.1- EMC XtremIO

EMC acquired Israeli start-up XtremIO in 2012 and the first GA products were brought to market in 2013. The platform was rapidly developed, with compression added as a feature in the XtremIO operating system, XIOS in version 3.0. The implementation of compression after initial product release was clearly a retro-fit and significant rewrite for the company as both the underlying data block size and metadata structures were changed. This required customers to offload data before deploying XIOS 3.0 as the upgrade was destructive.

Data optimisation features in the XtremIO platform are always-on and cannot be disabled based on the data type or volume. In fact, the deduplication process is inherent to the way XtremIO stores and distributes data across its high resiliency multi-node architecture. As a result, data that isn't suited to compression will still go through the proprietary algorithms that are implemented in the XtremIO platform (EMC provides no details on the compression algorithm used within XtremIO, however it is believed to be proprietary in nature).

Compression occurs after data is deduplicated in XIOS, in line with the process used by most vendors. EMC claims compression savings of between 2:1 and 3:1, depending on the data type, however no breakdown of these figures is provided.

6.2- Pure Storage

Pure Storage came to market with FlashArray in 2012. Since then, the company has moved on to become public and revamped the FlashArray hardware architecture with a new line of products called FlashArray//m. The later systems are still scale-up technology based on a dual-controller architecture and multiple disk shelves, but have started the transition internally to a PCIe/NVMe-based

architecture. NVRAM (write cache) has already moved, however drives are still connected through 12Gb SAS.

FlashArray uses a number of data reduction techniques that together are referred to as "FlashReduce". This includes basic pattern removal (looking for simple repeated patterns and zeroed data), compression and deduplication. Compression is implemented as a two-stage process. The first pass performs what is called "medium weight" compression on new data after deduplication. Data that is committed to the back-end storage is then subsequently post-processed to perform a more intensive compression process, called Deep Reduction. The Purity operating system for FlashArray uses a combination of a lightweight LZO algorithm for first pass compression and a patent-pending version of Huffman coding for the post-processing.

Pure has decided that the two stage process offers a better balance between performance and space savings. In addition, during periods of heavy system load, certain optimisation processes can be curtailed to ensure a consistent level of performance. Outside of this, FlashReduce features are always-on and cannot be changed on a per-volume/LUN basis.

6.3- IBM

IBM has developed a range of all-flash products, collectively known as FlashSystem. The basis of the solutions is technology acquired from Texas Memory Systems in 2013. The base FlashSystem 900 hardware is a "no-frills" dual controller with custom flash drives called MicroLatency modules. Features are added through the use of other IBM technology, including Spectrum Virtualise (formerly SVC) and Spectrum Accelerate (software from XIV).

In FlashSystem V9000 (SVC-based) and A9000 (XIV-based) systems, compression is implemented through an IBM feature called Real-Time Compression, which is used across a range of IBM storage products. The technology is based on RACE or Random Access Compression Engine, a combination of software and a hardware Add-In Card (Intel Quick Assist) that is deployed in pairs into both the V9000 and A9000 hardware. An additional 8-core processor and 32GB of dedicated memory is also required for V9000.

RACE itself is based on a form of the Lempel-Ziv (LZ) lossless compression algorithm that uses a sliding window technique to identify compressible content, typically based around a 32KB window, although the actual block size compressed is variable in length. A pre-decide mechanism is used to determine whether content will compress well, before committing to compress data. In addition, for data that appears to offer better levels of compression a Huffman-based algorithm is used

in place of LZ. The output from compression is coalesced into a fixed size write (32KB) in order to optimise the re-reading of compressed data.

IBM includes a licensee for Real-Time Compression for storage internal to the V9000 system, however if compression is used on external storage then this usage attracts a chargeable license per terabyte of virtual capacity. In terms of savings, IBM quotes around 50-60% saving on general purpose data, around 50-80% for databases and 45-75% for virtual server infrastructures. Finally, it's worth noting that FlashSystem V9000 doesn't currently support deduplication (in line with the lack of this feature in SVC). Compression in both V9000 and A9000 platforms is always enabled and cannot be turned off selectively by the customer.

7- Conclusions

The specific implementation of compression in external storage arrays is important because the process incurs processor overhead on both write and read operations. New data has to be compressed and stored when written; existing data has to be decompressed when read. This means the compression process should be targeted at only the data that needs it. In the evaluations and comparisons made in this paper, HPE 3PAR was the only vendor that provided the ability to explicitly turn off compression and deduplication for certain workloads, at the volume or LUN level. Some vendors have chosen to do either lightweight initial compression or (as with 3PAR) validate whether the savings from compressing data will result in sufficient savings to make the process worthwhile doing.

From the public information available, EMC XtremIO seems the least effective implementation. Compression is always-on for all data and doesn't appear to be optimized to validate whether data will compress well or not. The relatively small block size (8KB) of XtremIO writes means compression isn't as effective as the other platforms considered.

Both IBM and Pure Storage attempt to optimize the compression process, using either dedicated hardware (in the case of IBM) or a two-stage process (in the case of Pure). Post processing can be an issue in terms of the overhead of consuming additional CPU cycles and back-end drive IOPS. Both platforms use a variable-block approach with block size up to 32KB. In IBM's case, a significant amount of additional hardware is required to implement compression – two cards per V9000 controller (plus CPU and memory) and two cards per controller in A9000. Unlike 3PAR StoreServ's custom ASIC, this hardware isn't used for anything else.

On balance, HPE 3PAR StoreServ provides a balance of space savings through compression, optimized with custom hardware. At the same time, the customer can choose whether to deploy compression and deduplication on a workload by workload basis. All this is delivered non-disruptively for existing 8000 and 20000 series customers as a code upgrade at no cost.

In terms of an overall solution, 3PAR compression integrates into the Adaptive Data Reduction strategy, eliminating zeroed data (Zero Detect), eliminating duplicates (Deduplication) and then selectively compressing data where savings are made. Those compressed components are then efficiently stored on physical media with Data Packing. The result is an end-to-end highly efficient solution for optimising data on disk.

8- References

The following public documents, videos and websites were used as references for this paper.

Previous HPE 3PAR Papers

- 4AA5-1671ENW - HPE 3PAR Priority Optimisation: A Competitive Comparison
- 4AA5-3223ENW - HPE 3PAR Thin Deduplication: A Competitive Comparison
- 4AA5-5593ENW - HPE 3PAR Adaptive Cache: A Competitive Comparison
- 4AA5-8758ENW - HPE 3PAR StoreServ Replication

Vendor Papers

- Introduction to The EMC XtremIO Storage Array (Ver. 4.0) – A Detailed Review
- Introducing and Implementing IBM FlashSystem V9000
- IBM FlashSystem A9000R and IBM FlashSystem A9000 Architecture, Implementation and Usage

Videos

- Pure Storage: FlashArray and Purity OE Architecture Whiteboard Session

9- More Information

For additional technical background or other advice on the use of flash in the enterprise, contact enquiries@langtonblue.com for more information.

Langton Blue Ltd is hardware and software independent, working for the business value to the end customer. Contact us to discuss how we can help you transform your business through effective use of technology.

Website: www.langtonblue.com
Email: enquiries@langtonblue.com
Twitter: @langtonblue
Phone: +44 (0) 330 220 0128

Langton Blue Ltd
Kemp House
152-156 City Road
London EC1V 2DW
United Kingdom

9.1- The Author

Chris M Evans has worked in the technology industry since 1987, starting as a systems programmer on the IBM mainframe platform, while retaining an interest in storage. After working abroad, he co-founded an Internet-based music distribution company during the .com era, returning to consultancy in the new millennium. In 2009 he co-founded Langton Blue Ltd (www.langtonblue.com), a boutique consultancy firm focused on delivering business benefit through efficient technology deployments. Chris writes a popular blog at http://blog.architecting.it, attends many conferences and invitation-only events and can be found providing regular industry contributions through Twitter (@chrismevans) and other social media outlets.

HPE Publication: 4AA6-6256ENW