

Application Agnostic Real-time Data Compression

How Real-time compression across different applications reduces data footprint and energy efficiency without performance compromises

By Greg Schulz

Founder and Senior Analyst, the StorageIO Group

Author “The Green and Virtual Data Center” (CRC)

“Resilient Storage Networks: Designing Flexible Scalable Data Infrastructures” (Elsevier)

February 11th, 2008

Reducing data footprint to boost application performance or streamline data and storage management is an effective approach to addressing address data center power, cooling, and floor space and related environmental or green IT and green storage issues. The key is to achieve a reduction in data footprint without loss of data or incurring performance bottlenecks in the quest to achieve energy efficiency and environmental friendliness. This Industry trends and perspectives brief looks at how real-time data compression can be used for on-line active data to achieve energy efficiency without performance compromise along with enhancing near-line or secondary storage effectiveness.

This Industry Trends and Perspectives Paper is Compliments of:



www.storwize.com

Application Agnostic Real-time Data Compression

Introduction

IT organizations of all size across different industry sectors with diverse application requirements are faced with a common challenge of managing an ever increasing amount of data and storage. More and larger files being stored, more copies of data are being made, stored and protected for longer periods of time to meet compliance or other requirements resulting in a continuing growing data footprint.

The result is that more data storage is needed for access of on-line active data along with the need for more data storage capacity and I/O bandwidth to protect active and in-active data. Faced with rising energy costs, lack of reliable energy, limited floor space and cooling capacity along with new and emerging environmental regulations, IT data centers are faced with balancing the demand to support more data while balancing budgets, application response time and service levels, energy consumption and data management costs.

Background and issues

The combination of growing demand for electricity by data centers, density of power usage per square foot, rising energy costs, strained electricity generating and transmission (G&T) infrastructure and environmental awareness prompted the passage of United States public law 109-431 in 2006. Public law 109-431 instructed the United States Environmental Protection Agency (EPA), part of the Department of Energy (DoE), to report to Congress on the state of IT data centers energy usage in the United States.

In the August 2007 EPA report¹ to Congress, findings included that IT data centers or what is being termed information factories consumed about 61 billion kilowatt hours (kWh²) of electricity in 2006 at an approximate cost of about \$4.5 billion dollars. Also reported is that IT data centers on average consume 15-20 times or more energy per square foot than compared to a typical office building. Without changes in electricity consumption and improved efficiency, the EPA is estimating that IT data centers power consumption will exceed 100 billion kWh by 2011 further stressing an already strained electrical power G&T infrastructure and increasing previously high energy prices.

While there is a growing environmental impact awareness (“The Greening if IT”), the StorageIO Group, through research and regular discussions with IT personal, has found the more pressing problem facing many IT data centers (approximately 85-90%) are growing bottlenecks and approaching ceilings on available power, cooling and floor space.

The result is that while hardware costs continue to decrease with respect to storage capacity, the corresponding costs to manage and protect data continue to rise. Reducing your business and applications data footprint³ without loss of, or timely access to sensitive and information is an effective approach to addressing data and storage management challenges along with achieving energy efficiency.

Data Footprint

Your data footprint is the total data and storage needed to support your various business IT applications needs along with overhead of additional storage requirements to support development, test or quality assurances, decision support or data mining along with infrastructures resource management tasks including data protection, data preservation and business continuance (BC) or disaster recover (DR).

¹ An analysis by the StorageIO group of the 2007 EPA report to Congress along with links to the full EPA report is located at <http://www.storageio.com/reports>

² One kWh is 1,000 watts of energy or the energy usage of a device consuming 1,000 watts per hour

³ Read “Business Benefits of Data Footprint Reduction” at www.storageio.com/xreports.htm

Application Agnostic Real-time Data Compression

The solution to close the growing gap between increased storage management costs and the need to support more data to enable business growth is data footprint reduction across all tiers of storage and applications.

Business and technical benefits of a reduced application agnostic data footprint include:

- Reduce backup and restore time with faster mirroring and replication for disaster recovery
- Faster access to on-line active files and data along with improved energy efficiency
- Maximize IT budgets enabling growth while reducing management support costs
- Improve on the storage capacity and performance gap to remove IT data center bottlenecks
- Do more with your existing resources or, require less IT resources to sustain business growth
- Achieve energy savings and energy efficiency with more effective use of storage resources
- Sustain business growth by enabling more data to be stored in a cost effective manner

On-line and real-time (primary) data compression without performance compromises

In addition to simply moving data to another tier of storage, data footprint reduction can be accomplished using archiving of unused or retired data, compression (on-line and off-line) along with single instance storage (SIS) or de-duplication. Time tested and proven, data compression addresses the growth and demand for more storage capacity across all tiers of storage from on-line primary to secondary near-line or off-line across different types of application categories and data types.

Data compression is a commonly used technique for reducing the size of data being stored or transmitted to improve network performance or reduce the amount of storage capacity needed for storing data. If you have used either a traditional or IP based telephone or cell phone, watched a DVD or HDTV, listened to an MP3, transferred data over the internet or used email you have most likely relied upon some form of compression technology that is transparent to you. Some forms of compression are time delayed such as when you use PKZIP™ to zip files while others are real-time or on the fly based such as when you use a network, cell phone or listen to an MP3.

Two different approaches to data compression that vary in time-delay or impact on application performance along with the amount of compression and loss of data are lossless⁴ (no data loss) and lossy (some data loss for higher compression ratio). In addition to different approaches (lossless and lossy), there are also different implementations of including real-time for no performance impact to applications and time-delayed where there is a performance impact to applications.

Application Agnostic

- ✓ No CPU usage on application servers
- ✓ Operating system transparent
- ✓ Co-exist with existing storage
- ✓ Benefit on-line and off-line data
- ✓ Support database and multi-media data

A new data storage solution introduced into the market a couple of years ago, lossless real-time compression with no performance penalty is done on the fly and not as a background task and is done with no performance impact across different types of applications.

⁴ Lossy compression sacrifices some data quality or integrity for space savings where lossless compression guarantees data integrity with no loss of data during compression.

Application Agnostic Real-time Data Compression

Compression reduces the size, or footprint, of the data, which in turn reduces the I/O traffic to and from the storage system. By being application agnostic, a real-time compression solution particularly one that requires not host or application server drivers or that consumes CPU overhead can be used across a broader set of applications across an organization to maximize data footprint reduction benefits without adding complexity.

In fact, a performance boost can occur as a result of the compression in that less data is being transferred or processed by the storage system off-setting any latency in the compression solution. The storage system is able to react faster during both operations and take up less CPU utilization without causing the host application server to incur any performance penalties associated with host software based compression.

In contrast to traditional “zip” or off-line, time delayed compression approaches that require complete decompression of data prior to modification, on-line compression allows for reading from, or writing to, any location within a compressed file without full file de-compression and resulting application or time delay.

Real-time compression capabilities are well suited for supporting on-line applications including databases, OLTP, email, home directories, web sites and video streaming among others without consuming host server CPU or memory resources or degrading storage system performance.

In many cases, the introduction of appliance based real-time compression provides a performance improvement (acceleration) to I/O and data access operations for database, shared files, web servers along with Microsoft Exchange personal storage (PST) files located in home directories.

Fast relief to exhausted storage capacity management headaches

Scenarios where transparent application agnostic real-time data compression is a benefit include providing fast relief from storage capacity shortages involving unstructured files and data in user home directories. For example, your NAS storage appliances or filer servers are reaching or have exceeded the critical threshold for capacity utilization and your applications availability are about to be impacted due to lack of or no available storage to save data to. It could take hours if not days to identify candidate files for deletion or archiving to other mediums unless you already have some file management or storage resource management tools installed.

The next step would then be to move any data that needs to be saved and contact people that their files are about to be removed or moved. This all takes time and labor to resolve not to mention any resource cost for moving or archiving data before deletion.

An approach to buy yourself some time or to give you some breathing room is to deploy real-time compression in front of your existing NAS storage systems to reduce the data footprint and free up storage capacity. Environments where large volumes of office files including documents, slides, spreadsheets, personal databases, PDFs, PST email files or archives, iTunes and photo galleries exists are prime candidates for real-time compression of on-line data.

Real-time NFS database direct I/O capabilities

Another scenario for using real-time data compression is for time sensitive applications that require large amounts of data including on-line databases, video and audio media servers, web and analytic tools. For example databases such as Oracle support NFS3 direct I/O (DIO) and concurrent I/O (CIO) capabilities to

Application Agnostic Real-time Data Compression

enable random and direct addressing of data within an NFS based file. This differs from traditional NFS operations where a file would be sequential read or written.

To boost storage systems performance while increasing capacity utilizations, real-time data compression that supports NFS DIO and CIO operations expedite retrieval of data by accessing and un-compressing only the requested data. Additionally applications do not see any degradation in performance, CPU overhead off-loaded from host or client servers as storage systems do not have to move as much data.

Beefing up bulk storage without performance compromise

One of the many approaches to addressing storage power, cooling and floor space challenges⁵ is to consolidate the contents of multiple disk drives onto a single larger capacity and slower disk drive. For example, moving the contents of three 300GB 15,000 RPM Fibre Channel disks drives to a single 7,200 RPM 1TB SATA disk drive to avoid power consumption at the expense of performance and cost for data movement. An alternative approach is to use real-time compression to boost the effective capacity of each of the fast 300GB disk drives to approximately the same of the single 7,200 1TB disk drive.

The benefit is that real-time compression boosts the effective storage capacity by several times that of a single 750GB or 1TB HDD without the corresponding 3-4x drop in performance to achieve energy efficiency. This approach is well suited to environments and applications that require processing of large amounts of unstructured data that also need to improve their energy efficiency without sacrificing performance access to data.

Some examples include seismic and energy exploration, simulation, entertainment and video processing of MP3 or MP4 along with JPEG and WAV files, collection and processing of telemetry or surveillance data along with data mining and targeted marketing among others.

Hybrid data footprint reduction - De-dupe and real-time compression

Compression and de-duplication are two popular capacity optimization and data footprint reduction techniques today that can be readily deployed with quick results. While De-duplication or compression implemented as a single solution yield significant savings on storage capacity, using both technologies on the same data files should present even greater storage capacity space savings. The result should be no negative performance delays for on-line time or performance sensitive applications, no waiting for data to be re-inflated during data recovery or restore operations. 100% transparent to storage features (snapshots, replication, mirroring, or continuous data protection-CDP) and other IT infrastructure components.

StorageIO Perspective and Recommendations

The purpose of deploying a data footprint reduction solution is to achieve some level of consistent benefit across as much of your data, storage and applications as possible. Keeping in mind your target use or requirements, look at how various approaches to data footprint reduction achieve and balance data reduction with performance to meet your different tiers of storage and application service requirements.

A common industry misconception is that data compression and de-duplication are competing technologies. Although similar in some respects, the two are actually quite different in both their approaches to storage optimization to achieve data footprint reduction and their practical implementation in real world storage environments to meet different application needs and tiers or categories of storage.

⁵ Learn more about addressing IT data center and storage power, cooling, floor space and related green or environmental topics at www.storageio.com and www.storageioblog.com

Application Agnostic Real-time Data Compression

The StorageIO group sees benefits to using both real-time compression for on-line storage and de-duplication for inactive or backup data to apply data footprint reduction across all types of applications, tiers of storage and usage cases. For example, on-line compression of active data can reduce the amount of data being stored and accessed resulting in improved capacity utilization and I/O performance.

For backup data or data to be replicated, using on-line real-time compression combined with de-duplication solutions can reduce the time required backup or move data maximizing I/O and network bandwidth. The net results is that compression combined with de-duplication enables de-duplication solutions to do more work in a shorter amount of time to increase their overall efficiency and effectiveness.

General considerations for reducing data footprint with real-time compression include among others:

- ✓ How will a solution plug and play into your environment and co-exist with existing technologies?
- ✓ What data integrity guards exist along with availability or resiliency features?
- ✓ Will the solution scale in terms of performance, capacity, connectivity and ease of management?
- ✓ Does the solution work on-line in real-time or off-line in a post processing mode?
- ✓ What client or application server performance impacts are required to achieve data compression?
- ✓ Is the solution application agnostic working across different types of on-line time and performance sensitive application or data types along with across various tiers of storage in addition to supporting secondary, near-line and off-line backup and archive functions?

Storwize as an example of real-time appliance based compression

Storwize is a vendor as an example providing real-time application agnostic compression for on-line as well as near-line or off-line data that is complementary to co-exist with your existing data infrastructure environment along with enhance and compliment data de-duplication and single instance storage solutions.

The Storwize STN-6000 real time compression appliance introduced into the market in 2007 is a cutting edge storage optimization solution built upon widely accepted lossless compression algorithms that have been in use for over three decades on literally hundreds of millions of computers worldwide.

The applicability and benefits of Storwize data compression is not limited to primary on-line storage environments. Although designed for primary and real time data storage use, the Storwize technology provides significant benefits across other applications and storage tiers. For example, secondary storage of infrequently accessed data that needs to remain accessible on-line for long duration of time replicated and snapshot data, backup and archive data among others.

By implementing Storwize compression in the primary data tier, the benefits of storage optimization (and related cost savings) cascade through the storage architecture in the form of reduced data replication, smaller data backups, and reduced backup and restore (no time delays to re-inflate your data) times.

Application Agnostic Real-time Data Compression

Conclusion

Develop a data footprint reduction strategy that addresses on-line active, on-line inactive along with timely recovery of backup and BC/DR data along with removing unused data via archiving. By implementing on-line compression in the primary data storage tier without adding CPU overhead or requiring software on host or client server systems, the benefits of storage optimization (and related cost savings) cascade through the entire storage architecture in the form of reduced data replication, smaller data backups, and reduced backup or restore times.

For on-line active applications and data, the benefits of real-time data compression increases storage capacity without compromising performance across different applications while boosting the overall efficiency and effectiveness of de-duplication solutions

Leveraging the on-line and real-time appliance based compression solution across a diverse set of overall application scenarios and across all storage tiers, providing a significantly lower overall TCO when compared to a de-duplication solution targeted at only data backup and recovery. Using real-time compression for on-line active data is complimentary with use of de-duplication techniques for in-active or backup data as part of a data footprint reduction strategy, thus enabling the entire IT and storage infrastructure stands to benefit.

Real-time compression and de-dupe are complementary for backup in that data is compressed prior to deduping enabling even higher data footprint reduction ratios and storage usage efficiency. The result is a less complex, more efficient environment able to meet performance, storage capacity, application availability and energy efficiency in a cost effective manner.

About the author

Greg Schulz is founder of Server and StorageIO, an IT industry analyst consultancy firm, and author of the books *The Green and Virtual Data Center* (CRC) and *Resilient Storage Network* (Elsevier). Learn more at www.storageio.com, www.storageioblog.com and on twitter @storageio.

All trademarks are the property of their respective companies and owners. The StorageIO Group makes no expressed or implied warranties in this document relating to the use or operation of the products and techniques described herein. The StorageIO Group in no event shall be liable for any indirect, consequential, special, incidental or other damages arising out of or associated with any aspect of this document, its use, reliance upon the information, recommendations, or inadvertent errors contained herein. Information, opinions and recommendations made by the StorageIO Group are based upon public information believed to be accurate, reliable, and subject to change. This industry trends and perspective white paper is compliments of Storwize www.storwize.com