StorageIO Group Information Technology Industry Analysts and Consultancy

StorageIO Industry Trends and Perspective Solutions Brief
Real-time Data Compression for On-line Active Data

Author: Greg Schulz – Sr. Analyst

Compliments of Storwize

July 23 2008

For organizations wanting to address power, cooling, footprint and environmental (PCFE) "green" issues, a holistic data footprint reduction strategy should be pursued. A data footprint strategy should include real-time data compression of changing and static on-line data while using de-duplication of recurring backup data and archive for in-active data. This solution brief looks at how and where real-time data compression of active on-line data, including primary NAS and secondary storage, can be used to achieve data footprint reduction benefits across a wide range of applications in organizations of all sizes.

Background and Issues

In the big picture, if approximately 5-10% of data is changing being backed-up and de-duped, that leaves 90-95% of the remaining active on-line data as an opportunity for data footprint reduction as well.

To meet power, cooling, footprint and environmental (PCFE) and green challenges, real-time compression of on-line active and static reference data is an opportunity to achieve energy savings, boost storage capacity and enhance performance of existing primary storage systems. Real-time compression of on-line storage enables IT organizations of all size to address data and storage management issues including:

- Support more files and databases
- Manage more data efficiently
- Sustain business and information growth
- Avoid or address performance bottlenecks
- Enable existing storage store more data
- Enhance performance of existing storage

High-availability (HA), business continuance (BC) and disaster recovery (DR) of critical business data are also enhanced by real-time data compression. A smaller data footprint or amount of data to be transmitted over busy networks requires less storage capacity on storage systems and less network bandwidth.

Value Proposition

Benefits of a reduced data footprint extend beyond backup data. For example benefits can be achieved by reducing the amount of storage required for on-line data using real-time data compression. Other business and technology value proposition benefits of real-time data compression of actively changing data include:

- Data center-wide data footprint reduction
- Boost effective storage performance
- Increase usable storage system capacity
- Enable PCFE issues to be addressed
- Off-load application servers including:
  - Database and data warehouse systems
  - File and document management servers
  - Personal email files including PSTs
  - Web and entertainment systems
  - Reduce footprint of virtual servers
- Move more data for BC/DR in less time
- Reduce data traffic on congested networks
- Improve data management efficiency
- Enhance value of existing IT resources
- Faster backup and restores

The Technology

Real-time compression of on-line storage enables data to be compressed on the fly with no noticeable performance degradation or time delays to users or applications.

www.storageio.com P.O. Box 2026 Stillwater, MN 55082 651-275-1563 info@storageio.com

StorageIO Industry Trends and Perspective Solutions Brief
Real-time Data Compression for On-line Active Data

Real-time compression is in use all around us every day at home and at the office. For example, many tape drives support some form of streaming real-time compression to boost storage capacity utilization without negatively impacting performance. Databases, file systems and email solutions can also compress data on the fly though at the expense of server CPU cycles.

Application performance demands, including updates and reading of data, have an impact on storage systems and real-time compression solutions. Real-time compression for on-line storage should be able to retrieved data from storage systems uncompressing data without causing application and response time delay. Similarly data being updated or written by applications should be able to be stored in real-time in compressed format on storage systems without causing applications to slow down or pause.

Real-time and lossless compression algorithms can affect how processing and I/O resources are utilized to avoid performance delays or impacts to application and storage systems. Efficient algorithms enable wire or line speed real-time data compression without impact to users of applications of data being read or written. The result is that more data can be read or written, boosting effective storage performance in addition to enabling more data to be stored in a given footprint.

The quantity of available processing resources, including the size of processors, amount of memory, off-load processors combined with networking interconnects; enable real-time compression algorithms to compress changing data on the fly without negative performance impacts. The benefit is to off-load application servers and storage systems from using valuable compute and performance resources for real-time compression of active data.

How transparent the solution is to the existing environment, including avoiding the need for special host server based software or drivers, co-existing with servers and storage, eliminating need for additional post-processing to avoid additional overhead to storage systems, is also an important consideration of a compression solution

Another requirement of algorithms for real-time compression of on-line and actively changing data is to ensure 100% data integrity with lossless compression.

Strategies and Recommendations

Candidate applications for real-time on-line compression of active data include energy, general file sharing, web, seismic, pharmaceutical, entertainment, surveillance, financial and other performance sensitive environments. Environments that also need bulk storage at a lower cost can benefit from on-line compression where de-dupe or post processing may be a challenge.
Since modified data is being written on the fly, no extra processing is required to go back and re-compress changed data as would result in extra workload or performance impacts to a storage system.

Real-time data compression solutions require less data to be written to and processed by a storage system. The result is maxim usage of existing storage systems capabilities in an energy efficient manner and doing more with less, including boosting the available effective cache for performance on storage systems.

For read operations, less data needs to be retrieved, making read-ahead caches and algorithms more effective by transferring more data in less time, enabling more usable performance on a storage system. The end result is an improved effective storage system

www.storageio.com P.O. Box 2026 Stillwater, MN 55082 651-275-1563 info@storageio.com

update performance on reads and writes, while a smaller data footprint requires less storage.

The benefits are that more performance can be obtained from existing storage systems while boosting storage system cache effectiveness and storing more data in a smaller footprint.

For example, instead of using 750GB 7.2K RPM SATA disk drives, a NAS filer, such as an EMC Celerra or NetApp FAS, could be configured with 300GB 15.5K FC or SAS disk drives that have a similar power draw on a spindle to spindle basis.

Assuming a 2-3 to one or better compression ratio, the NAS filer would have the similar or better capacity with the same number of HDDs. However the performance benefit of the disk drives that are 2x the performance in RPMs yielding better IOPs as opposed to the slower SATA drives that could result in a performance bottleneck.

Closing Comment

Using real-time compression that can provide performance optimization benefits of existing storage and networking resources is a good technique to enable doing more with less. For actively changing data real-time compression compliments de-dupe for archive and static data scenarios.

For environments that are already interested in using de-duplication to improve data footprint associated with backup data or duplicate data associated with virtual servers, applying real-time data compression solutions for changing data across all applications can achieve even greater benefits, including addressing PCFE or green related issues and boosting overall data and storage management effectiveness.

Where to learn more:

An example of a solution that enables real-time compression of active data on NAS based primary on-line storage without performance compromise is the Storwize STN-6000 appliance. Learn more about the STN-6000 and its capabilities at www.storwize.com.

Additional material pertaining to data footprint reduction including the StorageIO Industry Trends and Perspective report "***Business Benefits of Data Footprint Reduction***", companion solutions brief for real-time data compression and other topics can be found at www.storageio.com.