StorageIO Group Information Technology Industry Analysts and Consultancy

StorageIO Industry Trends and Perspective Solutions Brief
Application Transparency and Co-existence with Real-Time Data Compression

Author: Greg Schulz – Sr. Analyst

Compliments of Storwize

September 4th, 2008

Organizations of all size and applications type, including high-performance file serving and databases with changing data, can realize the benefits of a reduced data footprint using real-time compression.

Background and Issues

Reducing data footprints is a popular concern relating to data and storage management along with power, cooling, floor space, and environmental green issues. Given that the amount of data being generated, used and stored for longer periods of time continues to grow, data footprint reduction techniques including archiving, real-time compression of active data and off-line compression of static data, data de-duplication for backup or archived data are all growing in popularity.

No one approach will address all data footprint issues in most IT organizations For example, one data center can be, at the same time, using archive to preserve data for possible future use or compliance, real-time compression for on-line changing data, off-line compression and de-duplication for static and backup data that seldom changes. Different data access patterns associated with new and emerging applications are changing the traditional rules of data lifecycle management.

An issue with data footprint reduction is that different applications use data differently at various points in time. For example, data de-duplication leveraging heavy thinking and processing algorithms can eliminate recurring data, however, the time delay associated with the heavy thinking can be an issue for on-line high-performance applications. A variation is real-time compression of data that results in a reduced data footprint without causing performance delays on read or write operations for on-line applications and data.

Value Proposition

The value proposition of real-time data compression for on-line and changing data is to store and access more information with existing assets without negatively impacting performance. Data footprint reduction of repeating data, as is the case with backup using data de-duplication, is growing in popularity and in customer adoption. Data de-duplication eliminates recurring data from backed-up data storage or archives reducing associated data footprint requirements for data protection including the amount of storage capacity needed.

Real-time data compression allows the benefits associated with reducing footprints for backup data to be realized across a broader range of applications and storage scenarios. As an example, real-time compression of active and changing data for file serving as well as other high-performance applications allows more data to be read from, or, written to a storage system in a given amount of time. The net result is that storage systems combined with real-time compression can maximize the amount of data stored and processed (read or write) without performance penalties.

The benefits of leveraging real-time compression for on-line active and high-performance applications and data include:

- Single solution for different applications
- Improved effective storage performance
- Increased capacity for fast disk drives
- Enhanced data protection capabilities
- Extended useful life of existing resources

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

StorageIO Industry Trends and Perspective Solutions Brief
Application Transparency and Co-existence with Real-Time Data Compression

The Technology

The ideal approach to real-time compression involves dynamically de-compressing read data and then re-compressing changed data as it is written to a storage system. The benefit is that read data is processed faster by storage systems as a result of a smaller data footprint while modified data is written to disk storage more efficiently and in a compressed footprint. Also, because data is compressed at primary storage, organizations will reduce capacity requirements for all downstream storage (e.g. replicas, archive, and backup). The primary example is Storwize.

One approach often misunderstood as real-time data compression involves dynamically de-compressing data when read with modified data being re-compressed at a later time. The benefit is static data that seldom changes can be reduced, enabling less storage space while allowing applications to read more data in a given timeframe. The downside to this approach is that changing data, including fileservers, email, office documents, design and development along with database files, is written without the benefit of compression. The impact is that more space on disk is required to write the data with no performance improvement benefit during write operations along with the overhead of a subsequent read and re-write operation when data is eventually re-compressed with an example being Ocarina.

The performance benefit should be that storage systems are able to use read and write caches more efficiently as compressed data uses less space on networks, in memory and when written to disks. Also, there is no additional overhead to a storage system for re-reading un-compressed files and subsequent write operations to compress the changed files, overhead that takes I/O and processing cycles that could be used for other productive work.

Strategies and Recommendations

Real-time data compression provides line-speed performance and is ideally suited for primary storage and high-performance applications such as databases. Real-time data compression can be deployed using host application software, network based appliances or in storage devices such as magnetic tape drives. Different approaches to real-time data compression address different needs with various compression rates and compression ratios depending on data types.

When determining which data compression approach to use for which data type several issues should be considered. Learn how and when data is compressed, how performance of storage systems can be enhanced including more efficient use of read and write cache. Consider how different approaches address and support different applications looking at compression and de-compression rates or the amount of data that can be processed in a given time frame as well as the effective compression ratio or data footprint reduction. Also look at applications, data types and deployment scenarios using different data footprint techniques for various tasks coexisting with your current technologies.

Closing Comment

Organizations investigating capacity optimization should consider both primary and secondary storage. Environments that are currently using or considering de-duplication for backup or archiving data should explore additional benefits from applying primary storage optimization as the economic impact can be felt not only in primary storage reduction but also in all downstream storage requirements as well. Learn more about data footprint reduction and related relevant topics at www.thegreenandvirtualdatacenter.com and www.storageio.com.

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