Data Domain's Global Compression — Greased Lightning For Data Protection

Analyst: Anne MacFarland

Management Summary

People have always wanted duplicate copies of data. Medieval monks copied manuscripts, ostensibly to spread the word of God, but many were kept sequestered *for safekeeping*. With the advent of the photocopier, the human propensity to *keeping a copy for my files* became visible as a rapidly expanding need for filing cabinet space, a problem later technology was supposed to solve. It solved the space problem but not the propensity.

Electronic data is in many ways more fragile than its paper precursor, and any damage may leave no residue or evidence of its occurrence. There remains a need for periodic physical data replication with log files to ensure an accurate data set. Bandwidth has increased, as have the clock cycles of processors - but so has the volume of data. There is software to automate the process, and point-in-time copies let the process trickle in the background less obtrusively than old batch processes. **Yet, still, data protection takes time that is scarce and space that we do not want to waste. It is still a pain in the data center, and any lessening is valuable.**

A new answer is full-integrity data compression. This works by recognizing duplicate data, removing the redundant values and replacing them with less bulky pointers to a single instance of the data. It can be compared to taking all the traditional campaign promises in a political campaign and replacing them with numbers (or letters) for each promise to shorten the length and the bombast of political speeches (of course, no politician would allow this to happen). This is different from some music and video compression algorithms, where 85% data integrity is considered *good enough*, and which might be more like news reports on those political speeches.

Much of corporate data is stored multiple times as attachments, graphics, and document versions. Full-integrity data compression can reduce the replication data stream significantly. And the story gets even better. To keep protected data up-to-date, an enterprise must replicate continuously. Incremental-change backups lessen data vulnerability, but make rebuilds more tedious, since log files must be used. **Most data centers do periodic "full backups" to provide a new baseline from which to restore. These full backups are chock full of duplication found in earlier backups. If this redundancy can be recognized by the compression algorithms, duplication can be avoided and compression rates soar.**

Data Domain Corp. has just such a set of full-integrity data compression algorithms, brought to market in an appliance called the *DD200*. They make it possible to protect more enterprise data more comprehensively without a larger backup workload. For more details, read on.

Data Domain's Mission

Data Domain seeks to make the comprehensive data protection provided by physical redundancy more affordable. Data is intransigently bulky and growing fast. The demand for documentation becomes more strident, not less fierce. Reliance on images for documentation, makes granularity and color quality not just nice but crucial for analysis. Even with plans for 10 Gb/s bandwidth building on the horizon, the need for data compression algorithms will not lessen. While bandwidth will be plentiful, it will never be free.

The Data Domain DD200

Armed with a data-protection mission, Data Domain has optimized its *Data Domain Recovery Appliance (DD200)* for sequential data flow, such as full physical data replication or other large data movement workloads. Compression is at least ten fold. Throughput is rated at 150 GB/hour, aided by SMP and hyper-threading techniques. Compression rates depend on the nature of the data and the congruence with what is already stored.

Data Domain algorithms can compress data of any format. Ongoing consistency checks (checksums) before and after compression assure data integrity throughout the process. Their *Restore Protection Manager* software recognizes all of the connections to the compressed data elements and assures that no elements are deleted until all links to them have expired. It can enforce retention periods.

Data Domain sees no need to create what others do well. It does not do the entire backup, just the compression. The DD200 works with enterprise backup software from Legato and Veritas at first release, and Data Domain will expand the list. Their appliance sits behind the backup server, configured as a back-up target, and appearing to the backup-server as a file system. Data Domain's disk storage permits user-driven restores, simultaneous backups and restores, and multiple slow (low priority) back-up streams, all within a unified, enterprise-wide data redundancy strategy. The DD200 does not directly replace tape, but acts more as a cache (rather like

virtual tape servers) for data that, as it less frequently used, may be further transferred to tape. **With DD 200, an enterprise can start to simplify data services.**

The 4U Data Domain Recovery Appliance 200 has 2 *Xeon* processors, whose memory structure is key to the high throughput. The appliance includes 4 TB raw (23 TB compressed) of ATA storage, in the form of sixteen 250 GB, 5400 RPM drives.

The heart of Data Domain (and the bulk of its intellectual property) is its Restore Protection Manager software. This contains the algorithms that speedily recognize data that has been compressed on a previous backup iteration. Restore Protection Manager, like NetApps' *WAFL* file system, is a non-overwriting process. Unlike WAFL, RPM writes only to the heads of a log, making consistency management simpler.

Data Domain's Roadmap

This is the first release from a small company. There are some functional limitations. The initial appliance handles only NFS files; it will handle CIFS files (used by Microsoft) in the next release. The next release also will include local and remote replication capabilities to complement the compression. There may be a two-box product (compression head plus storage array), if customers demand it. **The crux of Data Domain is its software, which can be responsive to the demands of its customers.**

At first release, the list price of the hardware is $25,300 and the software license is $32,700, totaling $58,000 for the whole appliance. **If you suffer from the pain of bulky backups, this may be what you need.**

Conclusion

This is a nifty device – and a good data management tool that signals a trend in data services. **Data protection is, alas, not optional. As a necessity, the process should be optimized.** Like RAID, full-integrity data-compression algorithms are a basic part of this optimization. Check it out.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

➢ *The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.*

About the Author

Anne MacFarland is Director of Enterprise Architectures and Infrastructure Solutions for The Clipper Group. Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

➢ *Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 28. (Please dial "1-28" when you hear the automated attendant.)*

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.