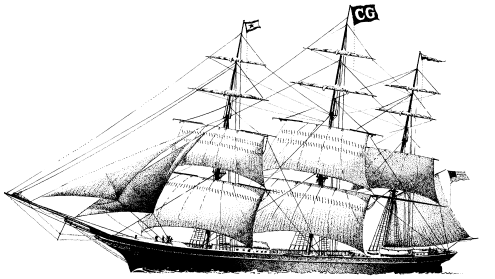
THE CLIPPER GROUP

Navigator™



Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2003015 April 18, 2003

The Value of Guaranteed-Authentic Information — The Expanding Role of EMC's *Centera*

Analyst: Anne MacFarland

Management Summary

Once, you could spot changes to a document by visible erasures or the telltales of *white-out*. You could identify what typewriter was used to create a document by character irregularities. Computer-generated fonts and printers have vanquished subtle clues to authenticity. **Digital records can be falsified far more convincingly - with less effort - than paper records. Never has it seemed easier to change history.** It can be tempting to "neaten up" the records – to erase initiatives that failed and which now might be considered "a waste of time." An individual may purge files of business value that seem useless to his future. **Never has the integrity of corporate documentation been more at risk.**

Government regulations address this danger by specifying, for public companies and many industries, the retention and disposal requirements for data. Over the years, these regulations have expanded to cover computerized records and, now, e-mail. In the past year, fines levied on some large enterprises for not preserving e-mail records signal that compliance will be enforced. **Most executives are considering how best to comply with these regulations during a period of budgetary constraint.**

Once the shock of the requirement wears off, some see that it can be beneficial to the enterprise in other ways. One of the biggest assets of an enterprise is the knowledge of how that enterprise functioned –when it entered markets and what initiatives failed – and why. This knowledge lives largely in *soft records* - like e-mail, calendars, and personal notes - and provide much insight into the *harder data* of reports, proposals, spreadsheets, and transaction and financial records. **The soft records can provide the *why* to the *what* that is shown in the hard data that is traditionally preserved.** Frequent changes and reorganizations threaten the continuity of this corporate knowledge. Geographic dispersion also may challenge the building of a common corporate knowledge base. More than ever, enterprises need assured access to genuine data assets to extract both broad and sharply focused knowledge of the markets they serve. **Guaranteed-authentic data has a premium value.**

Last year, EMC launched *Centera*, establishing a new category of data storage called *CAS* (content addressed storage). With CAS, an application retrieves fixed content by means of a content address. "Fixed content" comprises a range of digital assets retained for active reference and long-term business value, such as e-mail, documents, checks, medical images, broadcast content, completed engineering documents, etc. The content address for a fixed content object is a digital fingerprint, derived from the content itself and assuring its authenticity. Customers quickly recognized the value of Centera, and demand for the product has accelerated quarter over quarter. Now, EMC has over 125 Centera partnerships with independent software vendors (ISVs) for content management, medical and document imaging and archiving (e-mail, video, etc.), workflow, back-up, media, surveillance, e-learning, and oil and gas applications.

In addition to updating Centera's hardware, EMC has added software features to address the need to retain guaranteed-authentic information according to enterprise policies. It developed two compliance "personalities" that give Centera an additional value to all enterprises subject to government regulations. For more details, read on.

IN THIS ISSUE

Centera - Take Two

Denser, faster components and more software options

The architecture of EMC's new *Centera* platform is fundamentally the same as was introduced last year.[1] In version two, the ATA disk drives are bigger (250 GB instead of 160 GB) and the processors are faster (1 GHz, up from 866 MHz). What is new is that in version two there is an AC power transfer switch for each node, supporting content parity protection (or CPP, explained below), and failover between the two power rails.

Version two nodes[2], or entire cabinets of them, can be added non-disruptively to version one Centera environments: there is full backward compatibility between generations. Most capabilities of version two can be added to version one via an upgrade. There is one limitation – the Centera version one cannot support CPP.

Access Security

EMC has added software to allow systems operators to set access and activity parameters, both by application and by server[3]. They can specify read, write, delete, purge, and query functionality. This security access is token-based, and does not require recoding of the applications. It is fully backwards compatible with applications integrated with earlier versions of the Centera API.

Content Parity Protection

In version one, Centera stored information and mirrored the content to another node for redundancy (and possibly to a remote Centera for remote redundancy). Content mirroring requires the same amount of storage space for the mirror as the content itself. **For objects larger than 20 KB, Centera version two offers the option of a parity RAID-5-like approach it calls *Content Parity Protection (CPP).*** This process splits a content object in six parts, calculates a parity bit, and stores the seven elements in stripes across different nodes. Content retrieval needs any six of the seven elements. This takes less space than content mirroring (1.2 times vs. 2 times the original data stored). This option requires a version two enclosure (with the AC transfer switch) and a minimum of 16 nodes.

Deletion Functions

Shredding (Permanent Data Deletion)

Enterprise prudence often demands that data be quickly deleted after regulatory and corporate retention periods have been met. Deletion means that the data is truly unrecoverable.[4] Traditional WORM (write once, read many) formats of tape and optical media permanently preclude erasure, which in today's climate is seen as a source of risk for data that no longer requires retention. **Centera's deletion algorithms include a government-approved seven-time delete-and-overwrite cycle, known as *shredding*.**

Garbage Collection

Garbage collection deals with the recovery of storage space when a content object is no longer referenced by various users' metadata (known as the content descriptor file). It completes the deletion process and releases storage space for reuse. In version one of Centera, this garbage collection was an invoked process. In version two, it is a background process, like Centera's integrity checking.

Monitoring

Centera now features SNMP support. Metrics and alerts about system health can be fed to storage management software (such as EMC's *Control Center*). This allows Centera to be managed remotely, and as part of a larger storage environment. This capability is disabled in the Centera *Compliance Edition Plus* (see discussion on the next page), as part of its high security features.

Replication Quiescent Mode

Replication Quiescent Mode allows an administrator to defer remote data replication as needed. This refinement is useful where bandwidth to support applications is at a premium during certain times of the day.

---

[1] See *Retrieving the Needle in the Haystack – EMC's Centera Manages by Content* in **The Clipper Group Navigator** dated May 20, 2002 at http://www.clipper.com/research/TCG2002017.pdf

[2] A node is a drawer of disks that is treated as a collective storage group. A Centera array has a minimum of 8 nodes.

[3] This fulfills the requirements of HIPAA, the U. S. Health Insurance Portability and Accountability Act.

[4] As specified in Department of Defense Directive 5015.2.

The Additional Demands of Compliance

For many industries, government regulations mandate how long certain types of corporate records must be kept in an accessible (and searchable) state. The requirements for electronic records span the enterprise from end-user responsibilities to storage media attributes, though the specifics vary by industry and regulation. SEC Rule 17a-4 contains specifications and requirements about the storage media and recording process:

- Data must be time-stamped,
- Data must have verifiable recording accuracy,
- Data integrity during the retention period must be insured and documentable,
- Data retrieval must be prompt, and
- Appropriate data search capabilities must be available.

These requirements involve both the media used and the recording method. **The media must have enduring integrity and the recording process must prevent both accidental overwrites and premature deletions.** Some regulations, such as DoD 5015.2, also contain specifications about the deletion process.

While a particular storage medium will not solve the compliance problem, one's choice of media can affect the price of compliance and can make compliance easier. Traditional WORM tape and optical formats often involve high-touch administration and expensive robotics to be efficient in large implementations. Eventual deletion of records and reuse of space is not part of the traditional WORM concept. Text searches are often impossible. When they are possible, they typically take an excessive amount of time to complete.

Disk is an attractive media alternative, and is certainly more easily searched than removable media. Centera offers disk-based storage that stores and retrieves content via content addressing, providing a fingerprint and timestamp to certify data authenticity. Centera fills the business needs for a scalable, manageable storage solution that is familiar to application administrators, who may have limited experience with the intricacies of optical and tape.

With its *Compliance Editions*, Centera adds an additional element to ensure retention of authentic data for (and only for) as long as it is required. **A data retention attribute has been added to the metadata that accompanies each object.** This is set through the application. Code on the Centera array recognizes this metadata element and enforces the retention period for the data object. No application or user, no matter how privileged, can delete the data until the retention period has expired. Centera will not automatically delete data at the end of the retention period. The user or application must do that. The storage capacity can then be recycled for other needs.

T**he expansion of government regulations to cover more kinds of business records, and the scandals resulting from investigations using those records, have made enterprises more sensitive to the liabilities these records pose.** Enterprises want to delete records when it is legal and desirable to do so, both to manage risk and for reasons of security. Centera's object-oriented access method is considerably friendlier to the process of deleting files whose retention period has expired than is the file system access which supports optical or tape storage. It is not just a matter of the date of creation, but also of the type of file, that determine when the retention period has expired. The attributes of object metadata handle this complexity deftly.

The compliance editions are not a matter of hardware version but, as EMC puts it, of *personality*.

- In the ***Centera Compliance Edition*****,** the retention period must be defined by the application to invoke the inviolate data retention capability, just as workers have done with paper files. For the Compliance Edition, the retention default is "0," which means that the data can be deleted by the application, if no retention period is indicated. This edition will fulfill the compliance regulations of many industries.
- In the ***Centera Compliance Edition Plus*****,** there are still more safeguards,

which fulfill the needs of SEC Rule 17a-4. In *CE Plus*, the application will typically define a retention policy of each individual record. The default retention period enforced by Centera is indefinite, so that if the application fails to define a retention period, all data is automatically kept forever. **Remote administration via SNMP is disabled, there is no SysOps console, and the cabinet is locked, delivering truly high-security permanent storage.**

Disabling the console and remote SNMP monitoring requires physical monitoring of this super-secure array. Appropriate indicator lights are present on the enclosure:

- *Green* for no faults
- *Yellow* indicates that there was a fault, but that Centera is dealing with it
- *Red* for *open me now*.

In this way, the Centera Compliance Edition Plus assures its own health while not compromising the high-security environment that it enables. It is a higher touch environment. Determining the appropriate compliance personality an enterprise needs will depend on the stringency of the applicable regulations, the sensitivity of the data stored, and the capability of the IT environment.

Centera Momentum

Centera deployments have increased quarter over quarter since its launch last year. Over 125 independent software vendors have embraced Centera's CAS approach, with more expected as a result of Centera's increased functionality. EMC's competitors in the object storage space face a considerable catch-up challenge.

With the Centera ability to authenticate data, co-location of data assets should not be a source of concern. Therefore, **Centera provides an opportunity for enterprise or their application service providers to host multiple applications, possibly from multiple companies on the same Centera.** This solves a problem that has beset storage service providers, while government regulations offer new opportunities for them.

Conclusion

Digital content and documents, unless protected, are inherently vulnerable to casual or malicious alteration or deletion. With the proliferation of government regulations to protect such data, enterprises must instill new data retention practices. **Centera's Compliance Editions make this process easier by providing a data access tool to create a guaranteed, persistent environment of guaranteed-authentic information, and one which can be searched in a reasonable amount of time.**

The use of this guaranteed-authentic information – this single version of "truth"– goes beyond compliance. Using such a resource for multiple purposes, fosters enterprise continuity. This continuity underlies the enterprise's power to strategize, to act, and to refrain from actions. It can be the difference between being a market leader and a market also-ran. Think about what Centera version two might do for your enterprise.

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

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.