

WHITE PAPER

Improving File System Storage Efficiency with Unified Storage

Sponsored by: EMC

Steven Scully
March 2009

EXECUTIVE SUMMARY

IT organizations of all sizes are struggling to deal with the tremendous growth in data of all types, especially file-based, unstructured data. Even in difficult economic times, the number of digital applications and the amount of data they create continue to grow at a rapid pace. IDC forecasts that more than 75% of all new disk storage capacity shipped will be consumed by file-based, unstructured data in 2012. Managing all these files is becoming a bigger task for storage administrators.

Unified storage is one of the technologies that IT managers are using to improve their storage efficiency as well as to improve business flexibility and lower the total cost of storage. The benefits of unified storage include:

- ☒ Unified block- and file-level storage through SAN, NAS, and iSCSI interfaces, allowing storage administrators to consolidate multiple workloads onto a single storage platform
- ☒ Advanced software features for provisioning, data protection, and system management that improve storage efficiencies and utilization rates and simplify management
- ☒ Multiple tiers of disk to provide cost-effective storage of files at various stages of their life cycle
- ☒ Integrated hardware and software solutions (often in the form of appliances) that simplify the implementation

EMC, a leader in storage and software solutions, recently enhanced its Celerra unified storage platform to more efficiently handle the growing need for file system management. For IT managers who want to further automate file management activities, EMC has interfaced Celerra with the policy-based management capabilities of its Rainfinity File Management Appliance (FMA).

SITUATION OVERVIEW

The tremendous growth in data and the complexity of storage management challenges affect IT organizations of all sizes. Today's IT organizations are supporting a growing array of data-intensive applications (e.g., high-performance computing and business analytics) and content-rich applications (e.g., digital images and video). They are also improving application availability and disaster recovery through great use of disk-based data protection/replication. Challenging economic times bring only more focus on these requirements.

IDC predicts that CIOs will be especially focused on the proliferation of files on datacenter storage systems in the coming year or two. Digital content, data retention rules (existing and potential new rules), and users' unwillingness to throw away anything continue to drive spending on file servers and NAS systems. All of these files also consume valuable backup time and resources. Companies will be looking for storage solutions that simplify the management of data and take advantage of the benefits of tiered storage.

CHALLENGING TIMES FOR FILE SYSTEM MANAGEMENT

IT organizations face a number of challenges as they look for ways to better manage their file-based data. The number of files continues to grow and file systems can be difficult to manage, but budgets and resources are not necessarily keeping pace.

File Systems Keep Growing

The march of digital data continues at a rapid pace. Even in difficult economic times, applications are generating tremendous amounts of data. Corporate reorganizations and mergers along with increased governmental regulations contribute to the growth in data. Through the rest of this decade, IDC forecasts that the amount of external disk storage capacity shipped will grow 50% a year. By 2012, IDC forecasts that over 75% of all new disk storage capacity will be consumed by file-based, unstructured data.

A variety of applications across a broad set of industries is driving this file-based data growth. In many industries and specific market segments, the shift to digital content is now beyond the point of no return. Examples of this unstructured data include:

- ☒ Digital images used in industries such as healthcare and entertainment. Not only are there more digital images, but they are all growing larger and more detailed (i.e., high definition).
- ☒ Compliance and archiving applications. Whether they are used for regulatory compliance or not, copies of files (sometimes multiple) are archived for safekeeping.
- ☒ Collaborative applications such as email, SharePoint, and product life-cycle management create more files that need to be shared, protected, and possibly restored.

☒ Video-based applications such as surveillance and video-on-demand services create or deliver files 24 hours a day.

☒ Web 2.0 applications for Internet commerce or for internal applications boost productivity and improve customer support.

IT organizations are also adapting to a major shift in the use of disk-based backup and recovery solutions, which provide improved data protection with greater availability and continuity than previous solutions (i.e., tape). While these solutions improve application and data availability by shortening backup windows and file restore times, they add significantly to the number of files to be stored and managed on a daily basis.

File Systems Are Hard to Manage

Block (structured) data is typically managed by an application (e.g., databases), but files are generally considered unstructured data. Their management is typically left to the users who create the files. Many organizations don't know how much information they have, who (or what) created it, and where (or how many times) it is all stored.

IT managers do know that over 80% of the data stored on general-purpose file servers and dedicated NAS systems never changes and is rarely accessed after the first 90 days. However, they often continue to keep the data on high-performance/high-cost storage systems for a variety of reasons. They also have to back up all this data on a regular basis, consuming valuable backup windows and resources.

For industries that are subject to regulatory compliance issues (e.g., finance, healthcare, life sciences), compliance adds another layer of management complexities. Files have to be retained for specified amounts of time and may have to be stored on devices with WORM capabilities. In addition, multiple copies of each file may be required.

IT Budgets Are Not Growing

While the need for storage capacity hasn't slowed significantly, many IT budgets are under pressure. In more challenging economic times, organizations cannot afford to ignore the up-front or long-term costs of storage. Many IT departments have the same number of, or fewer, people to manage more and more storage while also supporting an ever more diverse array of applications that generate data and use that storage.

One area that has seen tremendous growth due to budget constraints in organizations of all sizes is server virtualization. Virtualized servers improve server utilization and performance and make greater use of consolidated, networked storage to deliver strong economic benefits. However, they also risk adding considerable management complexity and can disrupt finely tuned storage networks. IT organizations are looking for every option to make storage more efficient, less costly to acquire and upgrade, simpler to manage, and less costly to power and cool.

IMPROVING FILE SYSTEM EFFICIENCY

The evolution of storage solutions to address the inefficiencies in storing and managing files has been under way for some time but is now receiving increased attention due to the economy. Networked storage, storage tiers, thin provisioning, and storage consolidation are all examples of storage technologies that address the cost of storage and efficiency challenges. The conditions of dramatic file growth, difficult management, and budgetary pressures translate into several goals that vendors have for their unified storage solutions, including:

- ☒ Reducing the complexity of storage systems while expanding their capabilities with storage tiers and advanced capabilities
- ☒ Lowering the overall cost of storage, including initial acquisition costs, ongoing upgrades and maintenance, and associated power and cooling
- ☒ Improving storage flexibility and application availability and therefore increasing business value

In recent years, the concept of a unified storage system such as EMC's Celerra platform has evolved to help address these situations and challenges. Unified storage provides IT organizations with flexible platforms that can address various types of data using multiple storage tiers with a variety of data management features. Implementing unified storage solutions provides IT organizations with numerous benefits, including:

- ☒ The ability to leverage standard hardware to lower product costs and advanced technologies to boost performance and availability
- ☒ The flexibility to store file data with NAS protocols as well as block data with SAN protocols (e.g., iSCSI and Fibre Channel interfaces), allowing storage administrators to consolidate multiple workloads onto a single storage platform
- ☒ Support for multiple storage tiers, ranging from enterprise, flash-based solid state disks to low-power/high-capacity SATA drives to improve the efficiency of storage and lower the overall costs of storage
- ☒ A number of advanced software features for provisioning, data protection, and system management that improve storage efficiencies and utilization rates and simplify management
- ☒ Integrated hardware and software solutions (often in the form of appliances) that simplify implementation and configuration

These types of unified storage systems, often an extension of successful NAS platforms, help customers improve their storage efficiencies and reduce costs.

Finding the Right Storage Partner

Reducing the complexity of implementing and managing consolidated storage for virtualized servers and tiered storage is an important goal for the storage industry. In today's challenging environment, IT managers need solutions that deliver advanced storage capabilities with improved efficiencies, lower costs, simple deployment, and no limitations on expansion while still accommodating their specific business requirements.

The remainder of this white paper looks at the innovations that one leading IT company, EMC, is taking to deliver a family of unified storage solutions along with file management applications that address the ongoing needs of IT managers and storage administrators. The paper also discusses the challenges that EMC must address to further enhance the effectiveness of its offerings.

EMC UNIFIED STORAGE SOLUTIONS

EMC is a leading provider of IT infrastructure and solutions that enable organizations of all sizes to more efficiently store and manage their information. Celerra is EMC's unified storage solution, and it offers multiprotocol storage for file (i.e., CIFS and NFS) and block (i.e., iSCSI and Fibre Channel interfaces) data, a comprehensive set of built-in features at no additional cost (including snapshots, thin provisioning, deduplication, and volume management), and a design for ease of use by IT organizations of all sizes and capabilities.

In February 2009 EMC announced the latest version of Celerra, including several new hardware models and new software capabilities. Available in multiple configurations, Celerra scales from a few drives to almost 1,000. Each Celerra can be configured with multiple tiers of storage based on enterprise flash drives, Fibre Channel drives, or low-power/high-capacity SATA drives. The new Celerras can scale the number of blades as well as the storage capacity, providing independent scalability on both the front end and the back end. They can also be implemented in clustered configurations for high availability. Celerra is also available in several gateway configurations, which provide all the unified storage benefits to existing EMC storage platforms.

EMC further enhanced Celerra's existing features to help address the many challenges faced by storage administrators in managing the explosion of files. Some of the key new capabilities include:

- ☒ **Data deduplication with compression.** Celerra is EMC's first product to provide deduplication at the primary file system level. EMC has integrated technologies from its Avamar and RecoverPoint solutions into the Celerra operating system, which allows the system to improve storage efficiency and reduce storage capacity requirements. The system identifies files that are inactive, compresses them, and then removes all duplicate copies. Deduplication has the potential to save customers significant amounts of storage capacity, including removing the redundant data from the daily backup process. By placing data deduplication with compression in primary storage and moving it closer to the point of data creation, EMC is enabling users to be more storage efficient from the beginning.

- ☒ **Celerra compliance option.** Celerra File-Level Retention (FLR) has been enhanced with a new compliance option that allows users to meet SEC 17a-4(f) regulatory requirements by providing additional protection, such as a trusted clock and write verification, for locked files. Celerra FLR, with its new compliance option, provides for general-purpose primary system retention of file systems for simple but effective archiving and compliance.

- ☒ **VMware management tools.** EMC has enhanced the integration of Celerra into VMware environments with several management tools. With the rapid adoption of server virtualization by companies of all sizes, better integration with VMware makes storage administration more efficient in virtualized environments.
 - ☐ VMware Site Recovery Manager (SRM) Automated Failback is a VMware vCenter plug-in that helps users coordinate all the steps needed to fail back to the original virtual environment. This wizard-based tool allows users to define which VM to fail back and when and automates the reverse replication and synchronization of all delta sets.

 - ☐ VMware View Storage Integration is a VMware vCenter plug-in that helps administrators provision thousands of virtual desktops by storing virtual desktops, user data, and application data on a single platform. This consolidated storage is easier to manage and reduces backup requirements.

Managing Files with Policies

Celerra's integrated data tiers provide storage administrators with the ability to move files to the most cost-effective tier for any particular stage of their life cycle. Within the Celerra platform, the movement of files and file systems is not automatic. For IT managers who want to automate the movement of inactive files, EMC's Rainfinity FMA provides a policy-based approach.

Rainfinity FMA, which leverages the Celerra FileMover API, is a policy-based solution that identifies infrequently accessed files and moves them on a per-file basis to a different tier of storage. This tier could be a lower-cost tier of SATA drives integrated into Celerra or it could be another EMC solution such as Centera content-addressed storage. As the name implies, Rainfinity FMA is delivered as an appliance, simplifying the installation.

Rainfinity FMA manages the vitality of files, pruning the less productive files and archiving them. This allows IT managers to better align the file access and retention requirements with the capabilities of the storage. File management complexity is reduced through automation, and a lower total cost of ownership for the storage solution is achieved.

IMPROVING YOUR STORAGE EFFICIENCY

There are a number of ways IT managers can improve the efficiency of their storage implementations. There is no one right way; it will depend on each individual situation. In discussions with IT managers who have successfully improved the efficiency of their IT environments, some common themes emerge:

- ☒ **Virtualize your environment.** Adopt IT virtualization technologies (server and storage) to improve application availability and server and storage utilization. Server virtualization is being successfully deployed by IT organizations of all sizes. It works best with consolidated, networked storage (see below).
- ☒ **Consolidate your storage.** Tremendous efficiencies can be obtained by consolidating storage into networked storage solutions such as Celerra. While the benefits of consolidated storage were traditionally achieved in larger environments, server virtualization (see above) has allowed smaller organizations to achieve the benefits of consolidated storage as well.
- ☒ **Implement advanced storage capabilities.** Thin provisioning and automated volume management capabilities are becoming mainstream and are often included in unified storage solutions such as Celerra. Adding advanced capabilities such as deduplication and compression to primary storage can further improve your storage efficiencies.
- ☒ **Automate file management.** Automating file management begins with understanding how and by whom files are created in your environment. This allows you to create policies for storing, protecting, retaining, and (eventually) deleting files (instead of keeping everything forever). Once you understand the data life cycle more completely, it simplifies the implementation of policy-based file management applications such as Rainfinity FMA.

CHALLENGES

EMC is a leading unified storage and solution vendor. The ability to effectively use the Celerra platforms (and all EMC storage platforms) with Rainfinity FMA allows customers to create and deploy powerful file management solutions. Still, EMC, like all storage vendors, has to address some challenges to make its file management solutions successful, including:

- ☒ **Ensure that solutions are properly positioned.** EMC has a large and robust product line with multiple features and capabilities. This can lead to some product overlap (real or perceived) in the eyes of IT managers, CIOs, and CFOs. EMC needs to continue to focus on properly positioning its unified storage solutions and file management capabilities with its customers, channel partners, and sales force.
- ☒ **Continue to improve storage efficiency.** EMC needs to further integrate the management of all types of data (file and block) in the Celerra unified storage platforms. It also must continue to tightly integrate the Rainfinity FMA capabilities with Celerra to further simplify automated file management for customers.

- ☒ **Further educate the market on the value of automating file management and migration with policy-based solutions.** These types of solutions have been available for some time, and while they are less widely adopted by the market today, some customers realize tremendous benefits from them.

FINAL THOUGHTS AND GUIDANCE

This is a dynamic market with established vendors continuing to improve their offerings and several smaller vendors attempting to build beachfronts by focusing their unified products on specific market verticals or specific applications. When evaluating unified storage solutions, IT managers need to put equal focus on the storage technology and the ability of the storage provider and its business partners to properly deploy and support the solution based on their current and future needs.

Overall, the Celerra platform is an integrated and scalable unified storage solution that is in sync with both current industry trends and end-user demands and well-positioned to address current market needs. By integrating Celerra with policy-based file management from its Rainfinity FMA solution, EMC has created a product and feature set that should help IT organizations improve the efficiencies of storing and managing file systems and the storage solutions containing them.

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2009 IDC. Reproduction without written permission is completely forbidden.