

EMC Extends Benefits of File Virtualization into the Archive

June 2007



File virtualization is critical to successful file networking. However, we believe that file virtualization users do not often consider the impact that archiving will have on a virtualized file environment.

The ability to abstract files from their physical locations is the crux of file virtualization, but the procedure has serious implications for a recallable archive. Vendors differ in their approach to file virtualization, and some virtualization products can play fast and loose with archived data. Archives that are subject to compliance demands and e-discovery operations cannot risk slow restoration, unavailable access, performance hits, or application insistence on physical paths – all of the ills that some virtualization products are heir to.

Well aware of the issues around virtualizing archived data, EMC has actively pursued the means to solve these challenges with an approach called *Virtualized File Archiving*. Virtualized File Archiving combines EMC Rainfinity, its file virtualization platform, with EMC's content-addressable storage (CAS) Centera platform. This combined approach integrates the advantages of global file virtualization with the strengths of CAS, while preserving high levels of scalability and performance.

In this paper, we discuss some of the challenges around virtualizing archives and explore how EMC has leveraged Rainfinity file virtualization with Centera CAS. We believe that the result is Virtualized File Archiving: a highly efficient virtualized archivefile management solution that retains all of its compliance-driven integrity and capabilities.

Why Virtualize Files?

There are many benefits of file virtualization including increased utilization/higher ROI, management flexibility, migration flexibility, and data protection gains.

And nowhere are these virtualization benefits more needed than in the world of unstructured file-based data, which is growing at a huge rate. Large volumes of unwieldy files force IT administrators to

expand storage resources and to migrate files among them. If users and applications are solely relying on physical paths to locate files, then this process can create havoc.

File virtualization creates a logical pool that enables logical names instead of hard physical paths. The physical locations remain on the physical storage layer, but the virtual storage pool enables great flexibility in usage and management. Administrators can physically move files

TECHNOLOGY BRIEF

without interrupting user access, and as far as the user is concerned, nothing has changed.

This capability allows administrators to better set and meet Service Level Agreements (SLAs) for their file-based data: active data can be on high-performance Tier 1 disk while less active data can easily be migrated to Tier 2 and 3 online storage. This capacity underlies tiered storage architecture and networked storage efficiency for files and filesystems, all at an excellent ROI.

Yet for all its undoubted benefits, file virtualization faces serious challenges in a surprising place: the archive environment.

Archiving: A Challenge to File Virtualization

Individual file placement in a virtualized file environment is the heart of the issue. Once the logical and physical locations are separated in a virtual file environment, supporting the placement of files across tiers is fundamental to how a virtualization solution scales, ensures data integrity and leverages underlying storage. It is a new challenge for the file virtualization field.

Managing archives for SLAs, data protection, and compliance gives rise to some sticky issues for virtualized files. These issues include the need to ensure access to data when locations have changed, rapid recall, application access, fast performance, data integrity checking, and centralized management.

- **Challenge: Uninterrupted data access.** The whole point of virtualizing data is to abstract the logical view from the physical one, but if a logical file is suddenly unable to locate its physical location, then data is put at serious risk. This is an instance where virtualization approaches with single points of failure risk unacceptable downtime at the archive level.
- **Challenge: Rapid recall.** In case of disaster recovery, legal discovery or compliance audits, archived data must be quickly retrieved. The virtualization technology must not retard the recovery process nor compromise its integrity.
- **Challenge: Application access.** It is not only users who depend on logical views, but applications as well. In a virtualized file environment, if an application requires a physical path but is only presented with a newly abstracted logical location, that mismatch can lead to the application's failure, or unsuccessful data request.
- **Challenge: Fast performance.** Tier 2 and Tier 3 archive storage may not have the same performance requirements as heavy transactional environment storage, but they must still observe performance SLAs. As with fast recall, the virtualization technology must not play havoc with acceptable performance speed on archive systems.
- **Challenge: Data integrity checking.** Compliance and e-discovery procedures often require proof that archived data

T E C H N O L O G Y B R I E F

has not been tampered with. Virtualization technologies must be capable of providing such proof, or risk the company being out of compliance or facing stiff court fines -- both unattractive situations.

- **Challenge: Centralized management.** Virtualization products that only virtualize files on a single storage system might be acceptable for very small businesses, but swiftly become unmanageable in larger environments. Virtualization in mid-range and larger companies must be capable of centrally managing virtualization operations across different storage systems and tiers.

Ultimately a file virtualization solution must enable users to pursue non-disruptive data migrations, capacity balancing, and file archiving. This embraces management across inactive and active files and allows the administrator to pursue these key management tasks in any order.

EMC Virtualized File Archiving

EMC is well aware of these challenges, and has addressed them with Virtualized File Archiving. By leveraging Rainfinity and Centera, Virtualized File Archiving works both in terms of automated file placement or archiving, and also with the storage target.

Let us begin this discussion with a brief look at these two platforms.

Product Call-Out: EMC Rainfinity

Rainfinity is EMC's file virtualization platform. Rainfinity virtualizes unstructured data environments and non-disruptively moves data -- including active, open files -- across NAS, CAS, and file storage environments.

Rainfinity uses global namespace management to centrally manage and synchronize namespace services across distributed Windows and Unix environments, and to enable multiple filesystems to appear as a single virtual filesystem. In addition to global namespace management, Rainfinity uses six additional purpose-built applications including capacity management, file management, migration and consolidation services, performance management, synchronous IP replication and tiered storage management.

Rainfinity does not provide a single point of failure and is out of the data path when performing file archiving and recall. Rainfinity supports heterogeneous platforms as sources and targets including Celerra, NetApp and Centera. This neutrality ensures a common enterprise-wide archival platform for a virtualized file environment. This is a crucial requirement in today's heterogeneous, multi-vendor settings.

Rainfinity benefits include very high scalability, no single point of failure or risk to data, non-disruptive installation and transparent operations.

T E C H N O L O G Y B R I E F

Product Call-Out: EMC Centera

Centera is EMC's Content Addressable Storage (CAS) archiving platform. CAS enables IT to efficiently archive and manage unchanging digital content (fixed content), which includes data such as medical or financial images, voicemail messages, e-mail archives, and CAD designs. CAS is frequently targeted to companies who must observe compliance regulations and governance procedures, so its ability to safely retain long-term files and to prove integrity is critical. These and other requirements, such as the ability to restore quickly and reliably, make CAS a crucial archival system in regulated industries, litigious arenas, and in companies with large amounts of valuable archived data.

EMC Centera offers key benefits like single-instancing for disk savings, simplified and consolidated storage, provable data authenticity, high scalability, and universal application access.

**Rainfinity and Centera:
Virtualized File Archiving**

Rainfinity provides significant value in heterogeneous, multi-vendor file environments. But it really comes into its own in partnership with EMC Centera. Leveraging Centera's API integration, Rainfinity can provide the first seamless Virtualized File Archiving platform for the archived environment.

The combination of the two is a potent one. Rainfinity provides data mobility at the archiving source, which translates into

seamless access and management into the back-end Centera archive. No matter how aggressively the virtual file environment is managed "above" the archival layer, users do not need to worry where they logically move their files.

The Virtualized File Archiving platform also guarantees file restoration, which is a prime duty in any archive and especially in compliance-serving archives like CAS. Guaranteed restoration ensures that no matter how the file environment is virtualized, Rainfinity will serve up the archived data when users seek it. The data begins at the back-end archive, presents through Rainfinity, and is served to the appropriate hosts. The data will not be lost or orphaned.

An enterprise-wide archival platform must not only be host-neutral, but must also maintain high scalability as data stores grow, retention times lengthen, and more hosts come online. Rainfinity has no database dependencies and does not reside in the data path, nor does it require the creation and continual synchronization of a database to perform recovery. These elements provide tremendous scalability advantages.

Meanwhile, Centera's persistent data index allows Rainfinity to maintain a common interface into the archival platform over time. This ability enables the Virtualized File Archiving platform to maintain full access and data recoverability throughout long retention periods. High scalability also dispenses with the need for multiple archives. Instead, just one virtualized access

TECHNOLOGY BRIEF

into a single platform will efficiently grow and adapt over time to changing corporate and compliance requirements.

High scalability is a necessity, but it must not lose performance in the bargain. Rainfinity's virtualization technology has zero impact on the performance of the archive, meaning that the Virtualized File Archiving platform will perform just as well at scale as it would with a single Centera. Because Centera was designed to provide high performance at petabyte scale ranges, IT teams can know that their Virtualized File Archiving platform will also enjoy these benefits.

Summary of Benefits

- **Benefit #1: Data Mobility.** Rainfinity maintains seamless mobility between back-end archive and hosts.
- **Benefit #2: Guaranteed Restoration.** Data mobility and virtualization do not threaten the archive platform's restoration capabilities.
- **Benefit #3: Platform Independence.** The Virtualized File Archiving platform is host-neutral to serve enterprise-wide archiving environments.
- **Benefit #4: Efficient Scaling.** Centera's persistent data index, coupled with Rainfinity's lack of in-band file placement processing or database dependency, enables Rainfinity to easily

scale and consistently present a common interface over time.

- **Benefit #5: Performance at Scale.** Rainfinity maintains Centera's petabyte-level scaling environment.

Taneja Group Opinion

File virtualization offers compelling advantages to companies struggling to manage their growing stores of unstructured data. However, archive environments present unique challenges to file virtualization. The benefits can be immense, but virtualizing archives – especially compliance-ready CAS -- requires thoughtful integration to meet the challenges.

EMC has done an excellent job in getting ahead of this curve. Their well thought out integration of Rainfinity and Centera is an innovative move in what we believe will become the Virtualized File Archiving space. We like the fact that EMC has taken the effort to think about the entire lifecycle that customers will face in a file virtualization deployment.

We encourage EMC customers to thoroughly explore the implications to their file virtualization deployment for archiving, and to take a look at what EMC is offering with Rainfinity and Centera: the beginning of Virtualized File Archiving.



T E C H N O L O G Y B R I E F

NOTICE: The information and product recommendations made by the TANEJA GROUP are based upon public information and sources and may also include personal opinions both of the TANEJA GROUP and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. The TANEJA GROUP, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document.