



MIMOSA
SYSTEMS

White Paper

By Barry Murphy
and Bob Spurzem

May 2008

*From Email Archiving to Integrated
Content Archiving*

Mimosa NearPoint™ File System Archiving

CONTENTS

Introduction 3

Benefits of NearPoint File System Archiving 4

Key Product Capabilities 5

 Unified Back End for All Content Types 5

 Granular File Capture Capabilities 6

 Efficient Crawl of File System Content 7

 Versioning 7

 Archive and Index 7

 NearPoint Grid Architecture 8

 File Extension 8

 Search and eDiscovery 9

 Monitoring 10

 Content Monitoring for Risk Mitigation 10

 Retention/Disposition Policies 11

 Next-Generation Architecture 12

Example Use Case 13

Conclusion 14

Introduction

The widespread use of unstructured electronic files (such as Microsoft® Office, Adobe® PDF, and others) creates a critical management challenge for enterprise organizations. System administrators responsible for protecting and managing unstructured information face rapid storage growth and shrinking backup windows, both of which reduce overall data protection and increase storage costs. Attempts to manage unstructured files and documents manually are proving to be largely unsuccessful. A leading research firm revealed that 47 percent of open systems capacity is available but in the wrong place, and 55 percent of unplanned server outages occur from out-of-control disk space consumption¹. Another study found that management of unstructured files is complicated by the fact that 51 percent of unstructured data is unnecessary, duplicate, or non-business related, and 68 percent of data has not been accessed for 90 days or more². Without any way to improve this predicament, administrators continue to add storage and back up the same files over and over again—making the situation tenuous at best.

As file volume grows, organizations struggle to preserve the right content based on applicable compliance regulations or on pending litigation or investigations. Stringent legal guidelines from the amended Federal Rules of Civil Procedure (FRCP) put the burden squarely on the shoulders of the corporation—companies have a duty to preserve potentially relevant electronically stored information. Creating broad, enterprise-wide retention policies is extremely complex and difficult. As a result, many organizations desire to retain content based on role, such as finance or human resources. Other organizations have compliance concerns and need to not only retain content by role, but also copy the content and keep it on separate servers to ensure immutability. For all of these reasons, organizations increasingly turn to archiving vendors.

Over the past few years, hype has been built around the growing volume of email and the difficulty of managing it and discovering it. As a result, email archiving became all the rage—and rightly so given the benefits it delivers in the form of storage optimization and reduction in email eDiscovery costs (e.g., eliminating the need to restore backup tapes). Now the growth of file system content presents the same challenges and drives file system archiving. However, first-generation file system archiving solutions failed to gain user adoption because their heavy footprint stubbing solutions virtually always led to corrupt stubs. Even worse, trying to back up the file system with a spaghetti-like mess of stubs proved to be an administrative nightmare.

Today's opportunity is to leverage a next generation integrated archive that manages emails, instant messages, and files. With an integrated archive, search and discovery of electronic information becomes simpler because there is a single, unified repository to search for all of this otherwise unmanaged content. There is also a single eDiscovery

¹ Source: Strategic Research Corporation

² Source: SNIA/Source Consulting

application that enables review of the content in response to eDiscovery or investigations. In addition, the integrated archive eases the complexity of retention management by providing a single place to apply retention policies.

Mimosa Systems is an independent software vendor based in Santa Clara, California, that develops and markets solutions to enable administrators to manage unstructured information, including email, files, and documents. Its flagship product is Mimosa NearPoint™ for Microsoft Exchange Server; as its name indicates, Mimosa NearPoint supports Microsoft Exchange Server for email archiving, eDiscovery, recovery, and storage optimization. Now Mimosa NearPoint also offers the File System Archiving Option as part of its core capabilities. With File System Archiving (FSA), file server storage can be reduced and corresponding backup times are also reduced.

This white paper describes Mimosa NearPoint File System Archiving v1.0. We will cover high-level product goals and key product capabilities to give the reader a clear picture of this new product's benefits for storage management and legal discovery. An example case is included to demonstrate the product's benefits in a sample setting.

Benefits of NearPoint File System Archiving

The Mimosa NearPoint File System Archiving Option delivers multiple benefits. First, it provides significant storage cost savings and improves storage management and backup efficiency of file servers that contain thousands of unstructured information files. Under policy control, administrators can crawl file servers and index and archive files selected by various criteria such as type, size, and age. The crawl is performed without the need to place any agents on the file servers.

Selected files are captured, processed, and optionally stubbed from the source. Captured files are managed in the archive according to retention and disposition policies to meet rules for corporate governance. File extension (stubbing) is used to reduce file server storage and reduce backup times. Using NearPoint, administrators can move old files or files not recently accessed to the archive, thus reducing the burden on file servers and removing content from high-end production servers. NearPoint FSA also provides global single instancing across all archived content—messages, attachments, files—dramatically reducing the amount of content to be stored. As users modify files on the file server, the changed files are kept as newer versions of these files on NearPoint, with only the deltas from the original version of the file being stored on NearPoint. Administrators can use the version control capabilities of NearPoint FSA to specify the number of file versions to store. Auditors can subsequently search across different versions of the file, administrators can recover specific versions of the file, with end users continuing to have seamless access to their information.

NearPoint FSA also eases the burden of consistent retention management and helps to combat the rising costs of eDiscovery. As a unified archive of emails, instant messages, attachments, and files, NearPoint provides a one-stop shop for applying retention and disposition rules across all content. For legal discovery and adherence to the Federal Rules of Civil Procedure (FRCP)³, NearPoint FSA provides a single repository to search for all electronic information from file servers and Microsoft Exchange Server. Using built-in search tools, end users and auditors can quickly perform keyword search of the entire archive. Search results are displayed that combine email and files in a single view, simplifying the overall discovery process. With all electronic information right at your fingertips, NearPoint satisfies the critical FRCP requirement to be able to quickly identify and access electronically stored information for litigation. Should a lawsuit be filed, NearPoint can easily execute a legal hold on the necessary content, preventing any loss of evidence.

In addition to lowering storage costs, easing eDiscovery and retention management, and allowing seamless end-user access to information, NearPoint FSA also enables organizations to gain visibility into content risks within historical data. For example, an organization might not know that 25 percent of files stored on its file system contain proprietary prices lists. But if that information were to leak out, the organization would lose a great deal. By archiving that file system into NearPoint, the organization can now monitor the content and apply the appropriate policies to it.

Key Product Capabilities

Unified Back End for All Content Types

The Mimosa NearPoint archive platform supports all electronic messaging information for Microsoft Exchange Server, including attachments and items such as calendar entries and tasks (Figure 1). With the addition of file system archiving, the NearPoint platform is capable of supporting all content types. Based on tests performed by Mimosa, we found that a significant number of files and documents pass through Microsoft Exchange as attachments—more than 50 percent. For file and document archiving, a major reduction in archive storage is gained when file objects are de-duplicated across the entire archive repository. NearPoint stores all files and documents in flat files, which are managed in a folder hierarchy that NearPoint creates and manages. A Microsoft SQL database stores the digital signature for each file as well as all metadata for the archive. This unified back end has the following important benefits:

- **Storage efficiency.** A single copy of email, files, and documents is stored in the archive.
- **Fast search.** Search time is dramatically reduced, compared with tape-based search.

³ http://www.uscourts.gov/rules/EDiscovery_w_Notes.pdf

- **Accurate search.** Search is performed in one step, reducing errors.
- **Retention compliance.** Retention is managed with a single policy on a single set of content, reducing errors.
- **Legal compliance.** Litigation holds are managed with a single policy on a single set of content, reducing errors.

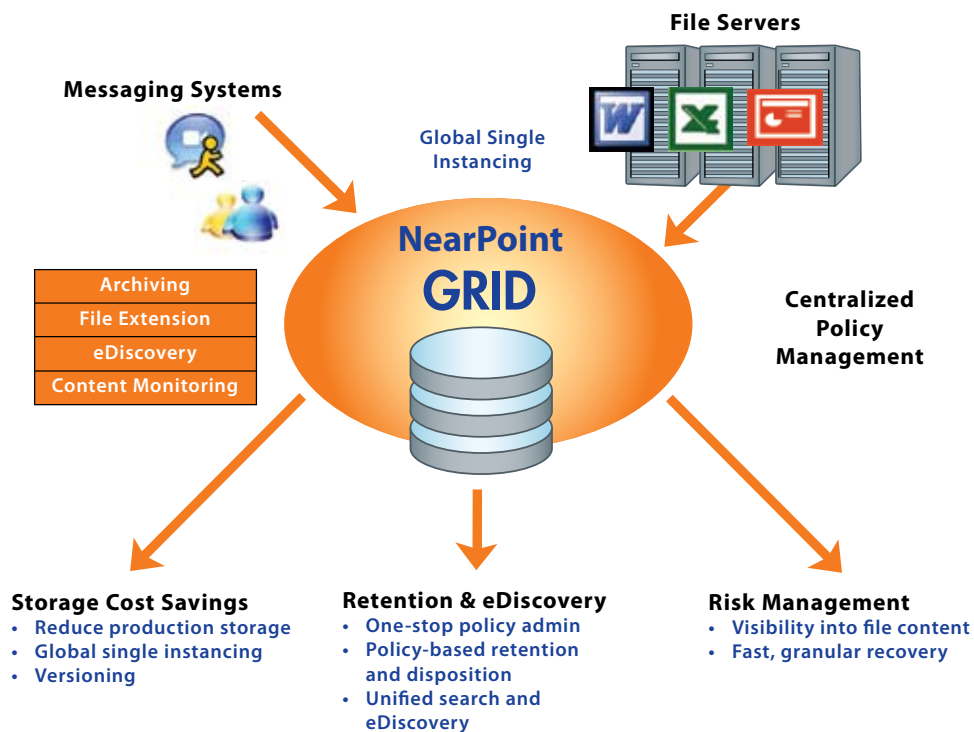


Figure 1. Mimosa NearPoint Archival Platform

Granular File Capture Capabilities

For easy application of archival policies and to assign access rights to auditors, Mimosa NearPoint FSA manages asset control by logical groups known as Cabinets. Assets are defined as files and documents located on servers, shared files, and folders. Cabinets manage multiple assets in a single logical group. Cabinets may be defined by any grouping criteria, such as geography, department, or project, and up to two levels of nesting of groups are supported. Cabinets manage the application of asset capture rules, archival policies, and assignment of access rights to auditors. Two levels of nesting are defined for Cabinets, with the policies and rights applied to the second level. NearPoint supplies standard inclusion/exclusion rules “out of the box” to be used if the administrator does not customize the rules for a group. Capture rules are defined by asset type, asset size, and asset age (creation date, modification date, and access date).

NearPoint provides comprehensive policies for archiving, stubbing, versioning, and retention/disposition based on policies. NearPoint performs an efficient multi-pass scan of metadata, looking at the lowest-cost metadata first, examining the properties of that metadata, and then determining if the file is a candidate for archiving before copying the file. It's important to note that NearPoint FSA maintains as small a footprint on the file server as possible. The only agent installed is an optional one for seamless stubbing.

Efficient Crawl of File System Content

NearPoint FSA crawls and captures files within Cabinets according to administrator-defined schedules. The crawler does not require agents to be deployed on the file server, and it collects files and file metadata from file servers based on file capture rules. These rules define the scope of files to be captured. Only the files that match the file capture rules will be copied from the file server to NearPoint.

Open files, encrypted files, and password-protected files are handled as exceptions and are displayed as part of monitoring. The crawler gains access to shared folders based on credentials supplied by the administrator. Crawls are performed on scheduled intervals or on a perpetual schedule while avoiding blackout periods. If errors are encountered during the crawl, they are noted and the crawl continues to the next file, share, or system.

Versioning

By default, NearPoint archives all versions of captured files. As an option, the administrator can configure the number of versions kept in the archive for storage savings. During a search, auditors can view results at a particular point in time and retrieve the latest version of the file at that point. When a file is renamed or moved within a file share, NearPoint FSA tracks this file and versions any changes to the moved/renamed file rather than treating it as a new file, thus enabling proper lifecycle management.

Archive and Index

NearPoint FSA indexes all captured files within a Cabinet, thereby making a full-text index of all content available for search, user access, and eDiscovery. All captured files, including ones that are optionally stubbed (based on group-specific stubbing rules), will be indexed. This feature is useful when the intended purpose is only to reduce storage and not to index captured files. NearPoint creates a digital signature to files on the way into the archive for indexing and on the way out of the archive for search, and ensures that the two signatures match. Indexing is performed on the file content and metadata, or it can be performed on metadata only. When a search is performed, file content as well as properties can be specified in the search criteria.

NearPoint Grid Architecture

Mimosa NearPoint proved its scalability for Exchange email archiving with its multi-node grid architecture. That same technology supports FSA. Utilizing a superscalar or super-pipelined grid, NearPoint can support tens of thousands of file systems in a single system. A hot-plug modular architecture allows servers and storage to be added or taken away as required to match performance challenges, without breaking the logical consistency of the archive information—an important capability for eDiscovery. The modular architecture also provides load balancing and avoids single points of failure. Archive storage capacity grows on demand automatically, and default configuration and wizard-driven menus simplify deployment and management (Figure 2).

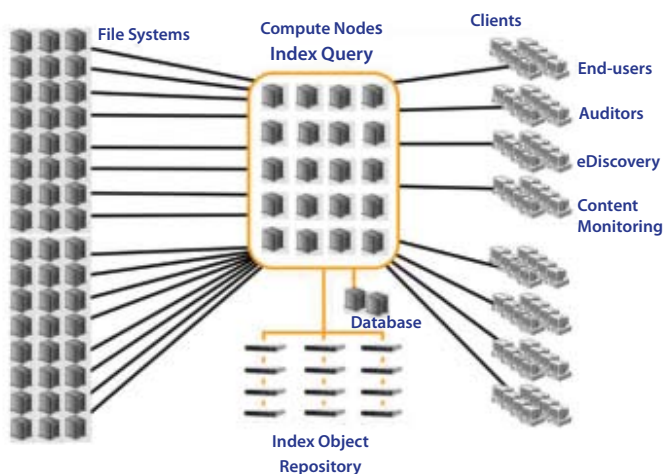


Figure 2. Grid Architecture

File Extension

NearPoint FSA supports file extension for NTFS file systems. For NTFS, the administrator can choose between seamless stubs and Internet-style shortcuts (URL). The seamless stubs look exactly like the original file to a user or an application, whereas Internet-style shortcuts place a URL link in a small stub file. To access the stubbed file, the user double-clicks the stub file to access the URL. With NearPoint, putting an agent on the file server is optional (required only in the case of seamless stubs), allowing administrators to deploy FSA with zero footprint on the file system.

In all situations, the File Extension Policy can be defined per Cabinet by file type, file size, and file age and freshness. Stubbed files can be restored from the archive and returned to the file system in situations where file-level recovery is desired. The most significant benefit of stubbing is storage reduction on file servers and reduction of backup times. NearPoint’s file extension capabilities avoid the “follow-the-stub” backup nightmares of

first-generation products. If using NearPoint for backup purposes, administrators need only to configure NearPoint to not back-up stubs. If using smart back-up software, it is possible to configure NearPoint to allow that backup solution to easily follow the stub and backup the correct file system content.

Search and eDiscovery

The Mimosa NearPoint eDiscovery Option is a robust yet easy-to-use search and discovery application used from an auditor's or legal counsel's desktop. In the past, discovery could take days or weeks because of the need to restore from backup tapes. With the eDiscovery Option, which leverages the disk-based NearPoint archive, search times can be reduced to minutes or seconds. Using eDiscovery, auditors define new searches by selecting target Exchange mailboxes and/or FSA Cabinets and by defining search criteria. Full-text search of messages, attachments, and files is supported in a single, unified view (Figure 3).

With the eDiscovery Option, teams can work together to respond to a discovery request or complete an internal audit. Search criteria can be saved and shared among users, ensuring consistent and accurate results for frequently used search patterns. Content assets can be placed on legal hold to prevent deletion and then shared with other auditors or legal staff to facilitate a review workflow. Assigned reviewers can easily browse through and tag items to aid subsequent searches and to reduce unnecessary duplication of effort when matters overlap. Upon completion, review results are exportable in Outlook® PST files suitable for external counsel review or court-ready production.

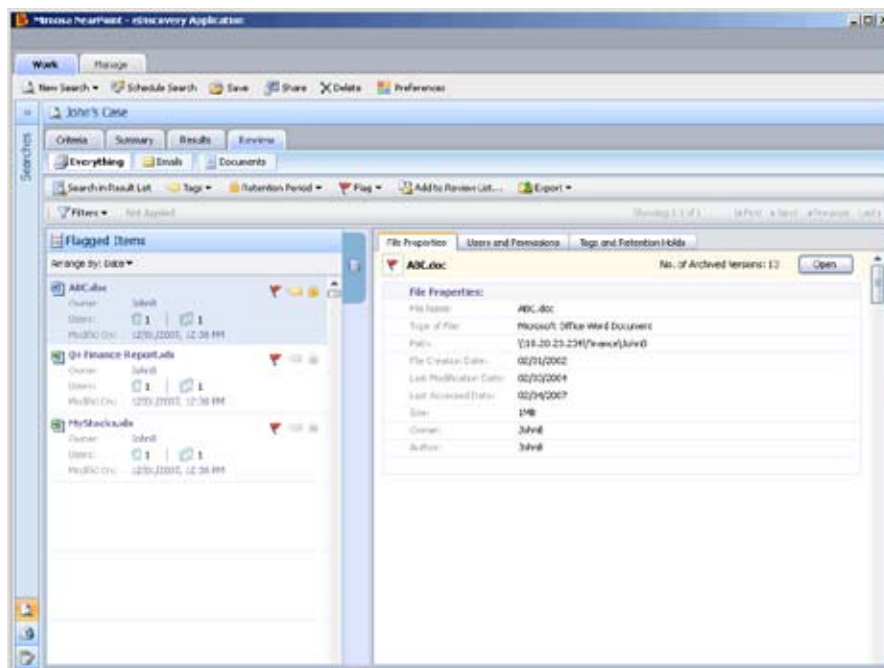


Figure 3. Unified eDiscovery Search User Interface

Monitoring

NearPoint FSA has a built-in monitoring capability to display activity statistics for crawled and archived/indexed files. Monitoring shows the status of a crawl in progress, and it will incrementally display the number of files captured in the various Cabinets by file type. It will also display the number of files archived and indexed, as well as stubbed. The administrator is notified of exceptions for unreachable systems or files during a crawl. Examples of such exceptions are:

- Server not accessible because of network issues
- Share not accessible because of authentication failure (based on supplied credentials)
- Folder not accessible because of insufficient privileges
- Files not accessible (e.g., open files, lack of read permission)

Content Monitoring for Risk Mitigation

Many companies are regularly subjected to litigation and discovery based on preventable actions by employees. The new amendments to the Federal Rules of Civil Procedure specifically designate corporate email as a discoverable record that is subject to litigation hold. An effective strategy to lessen this type of legal risk is to manage these activities before they become an outside legal matter requiring expensive eDiscovery processes. A content monitoring solution facilitates this strategy, allowing organizations to become “content-aware” via visibility into otherwise unmanaged content.

A content monitoring solution, along with a live content archiving solution, will watch for violations of corporate policy in real time within the archive. This solution will also allow the organization to head off questionable or inappropriate activity before it becomes a major problem.

The Mimosa NearPoint Archiving platform with the Content Monitoring Option provides a powerful, cost-effective means of enabling policy violation alerts—such as inappropriate content coming into the organization or sensitive company information being sent to external recipients—ultimately monitoring adherence to corporate communications and code of conduct policies. For example, a company might have a policy to review all content that could potentially violate insider trading regulations. With the NearPoint Content Monitoring Option, organizations can search the archive for existing phrases that might indicate a violation, and then work with the offending party to take corrective action (Figure 4).

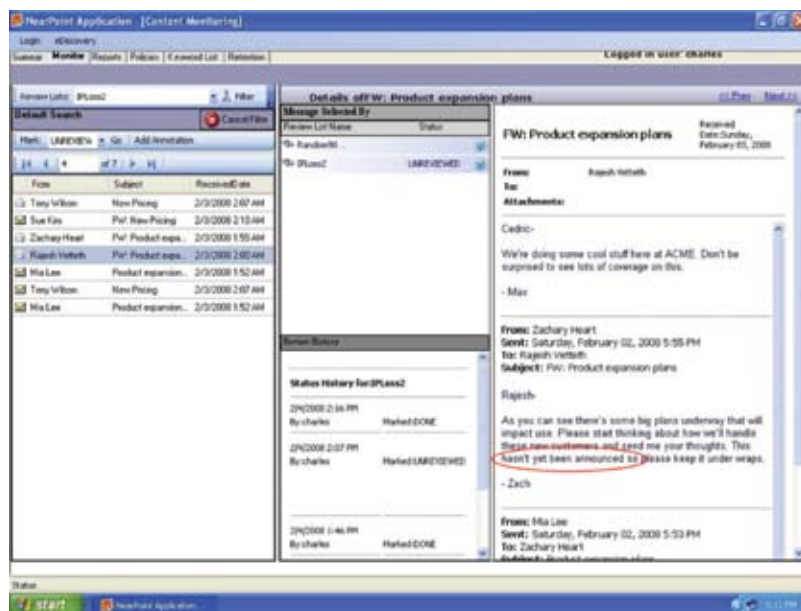


Figure 4. Content Monitoring User Interface

Retention/Disposition Policies

NearPoint FSA provides the ability to set retention periods for files and message items in the archive. NearPoint default retention/disposition policies are supported for various file types. The default policies can be viewed and overridden by the administrator. Retention/disposition policies are applied to all captured files in a Cabinet or to subsets of files within Cabinets. After a retention period expires, the affected files are deleted and indexes are purged from the archive. If a retention period is extended or if the asset was originally assigned a different retention period as part of a different group, the following rules apply:

- If the new retention period is longer, the retention period for the asset will be extended by the delta.
- If the new retention period is shorter but has not expired, the retention period for the asset will be lowered by the delta.
- If the new retention period is shorter and has already expired, the asset will be disposed of immediately.

NearPoint FSA will also support a data protection policy. The data protection policy will retain files affected by that policy on NearPoint for the specified retention period after the file is deleted from the file server.

It's important to note that retention policies can be set and applied globally across all content in NearPoint—email items, instant messages, and files.

Next-Generation Architecture

With the addition of the FSA Option, NearPoint is truly a platform that provides a next-generation architecture for managing user-generated content. What makes NearPoint so special?

- **Service-oriented architecture (SOA).** NearPoint employs an API-based modular design that provides the necessary foundation for external APIs and third-party solutions to extend the platform. It also makes NearPoint appropriate for enterprise deployments with security tiers and suitable for Software as a Service deployment.
- **Unified and extensible archive for email, files, and other content types with XML-based document type definitions.** APIs are available for capturing content from additional sources (e.g., SharePoint). NearPoint has a generalized archiving, indexing, and search infrastructure for adding new document type definitions in the system. And there are minimal/pluggable changes to applications that leverage the generalized document type definitions—localized to preview/view, recovery, and export functions. And, unlike competitive solutions that store content in multiple archive instantiations, NearPoint is a true, single archive—all content is stored in one highly-scalable archive.
- **Extremely scalable with NearPoint Grid Architecture.** NearPoint is scalable to tens of thousands of file systems. It leverages existing NearPoint archive volumes and databases for global single instancing, existing scratch volumes for data capture and export, existing search engine index volumes for new document type-specific indexes, existing resource discovery service and resource management databases, and existing activity tracking databases.
- **Optimized archive volume utilization.** NearPoint provides global single instancing across all document types, delta storage for document versions (content and metadata changes), and versioning policies for ongoing version pruning.
- **Highly scalable search.** NearPoint has metadata-only indexes for high-performance metadata searches, and metadata and content indexes for application queries. Metadata indexes are used for efficient incremental crawls and policy evaluations. The scalable, resilient, and efficient indexing engine supports composite document indexing, provides index splitting and merging for managing index growth, and performs background re-indexing and compaction.
- **Advanced file extension technology.** NearPoint maintains optimal use of the network and storage during stub file operations. Unlike in HSM applications, when a file is opened for reading in NearPoint, the file is not brought over to the file server.
- **Flexible recovery mechanisms for administrators.** Administrators can recover files from the archive if files on file shares are accidentally deleted or corrupted.

Example Use Case

Our example company has several file servers for use by its employees. As a result of growing file system content and litigation risk concerns, the company has deployed the Mimosa NearPoint File System Archiving Option to gain control over the file system content for its finance department. Currently, corporate file systems house five terabytes of content. The company's policy is to keep finance department content for five years. With NearPoint FSA, our example company is able to identify file shares that belong to the finance department and run a crawl of these file shares to find files and put them in Cabinets. The administrator can configure a crawl schedule and indicate peak and off-peak hours as well as planned downtime. The FSA crawl adjusts in a way that puts minimal load on the servers during their peak hours of usage and avoids crawling during planned downtime.

Once this efficient capture runs and the files are added to the archive, NearPoint conducts de-duplication of the files—not just among other files from the file system, but against email attachments in the archive, as well. At this point, the files are indexed and available to be searched either by end users or by auditors or legal staff for investigations and/or eDiscovery.

In this use case, a legal matter does arise and the company has to find files created by senior financial executives that are two years old or less. Using the NearPoint eDiscovery Option, corporate legal staff members identify several senior financial staff and create a search based on metadata that indicates the selected custodians authored content and that the creation date is within the last two years. When the search results are returned, email and files are returned as part of the same search so legal staff can review the files for privilege or responsiveness and tag files appropriately. Or the collected set of files can be exported as a PST file and sent to the law firm for further review. The whole process takes only days; in years past, it would have taken months or even years.

Given the large amount of data in company file systems, our example company also has a policy to archive any files that are greater than two megabytes in size. For files of that size that have been accessed within the last 90 days, the policy indicates that a stub will be left so that users retain seamless access to the file. Again, the administrator uses the granular capture control in NearPoint to run the crawl to look for files greater than two megabytes in size and that have not been accessed within the last 90 days. Those files are all marked for archiving. The crawl also looks for files that are greater than two megabytes in size and that have been recently accessed. These files are marked for archiving, too, but NearPoint leaves a stub in place so that users do not sense a change to their files.

As a result, the company achieves numerous benefits. By archiving large files, especially those with long retention periods, the company reduced the amount of content on production servers and is able to optimize storage costs. In this case, five terabytes of content on production servers are reduced to one terabyte, producing a savings in storage space of 80 percent. As a result of stubbing, backup windows are reduced by close to 50 percent.

Conclusion

The Mimosa NearPoint File System Archiving Option integrates with the Mimosa NearPoint for Microsoft Exchange Server email archiving solution, delivering a fully integrated solution to manage unstructured electronic information. For storage reduction, eDiscovery, and compliance with new FRCP rules for litigation support, the NearPoint FSA Option performs automated rule-based archiving of file servers and optionally replaces files with small stub files. As a result, file server storage is reduced and backup times are shortened. Users can retrieve stubbed files with a simple double-click, and auditors can perform quick searches across admin groups to identify files, instant messages, and email for litigation support. Legal discovery is performed quickly on the disk-based NearPoint archive instead of on tape media, dramatically reducing the cost of discovery. Litigation holds are managed at the file and document level by NearPoint instead of on tape assets, again resulting in dramatic cost savings.

Find Out More

For more information about the Mimosa NearPoint File System Archiving Option and the entire Mimosa NearPoint solution for unstructured information, contact your Mimosa Sales Representative at (408) 970-9070, or visit our web site at www.mimosasystems.com.

About Mimosa Systems

Mimosa Systems, Inc. delivers next-generation content archiving solutions for information immediacy, discovery, and continuity. Mimosa NearPoint is the industry's most comprehensive unstructured information management software solution for email, files, and instant messages, enabling archiving, eDiscovery, storage management, and recovery in a unified solution. Mimosa is a Microsoft Gold Certified Partner, recognized for its competencies in networking infrastructure solutions, ISV/software solutions, and advanced infrastructure solutions. Mimosa is a privately held company whose investors include August Capital, Clearstone Venture Partners, Dot Edu Ventures, JAFCO Ventures, and Mayfield Fund. Mimosa was founded in 2003 and is based in Santa Clara, California, with offices in Germany, the United Kingdom, and India. For more information, see www.mimosasystems.com.



MIMOSA SYSTEMS HEADQUARTERS

United States
 3200 Coronado Drive
 Santa Clara, CA 95054
 T 408-970-9070
 F 408-970-9041

Email: info@mimosasystems.com
 Sales: sales@mimosasystems.com
 Technical Support: support@mimosasystems.com
 Public Relations: pr@mimosasystems.com

WORLDWIDE OFFICES

United Kingdom
 400 Thames Valley Park Drive
 Thames Valley Park
 Reading RG6 1PT
 T +44 (0) 118 963 7860

Germany
 Max-Plank-Strasse 8
 85609 Aschheim-Dornach
 T +49 89 904 7551-0

India
 8th Floor, Amar Genesis
 ITI Road, Aundh
 Pune 411 007 India
 T +91 20 4048596



© 2008 Mimosa Systems, Inc. All rights reserved worldwide. Mimosa, Mimosa Systems, the Mimosa logo, and NearPoint trademarks of Mimosa Systems, Inc. in the United States and other countries. Other product names mentioned herein may be trademarks or registered trademarks of their respective owners. FSAWP_0408_NA