



MIMOSA
SYSTEMS

White Paper

By Bob Spurzem
Mimosa Systems, Inc.

October 2007

Mimosa™ NearPoint™ File and Documentation Archiving

CONTENTS

| | |
|-------------------------------|----|
| Introduction..... | 3 |
| Goals..... | 4 |
| Key Product Capabilities..... | 5 |
| Example Use Case..... | 9 |
| Conclusion..... | 11 |
| For More Information..... | 11 |

Introduction

The widespread use of unstructured electronic files (such as Microsoft® Office, Adobe® PDF, and others) is creating a critical management challenge for enterprise organizations. System administrators responsible for protecting and managing unstructured information are facing rapid storage growth and shrinking backup windows that reduce overall data protection and increase storage costs. Attempts to manually manage unstructured files and documents 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 consumption.¹ 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 more.² 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.

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. Mimosa Systems is currently developing an option for NearPoint that adds support for file and documents. When added to NearPoint, this option will bring together, in one solution, complete information management of email (and attachments), files, and documents. File server storage can be reduced and corresponding backup times are also reduced. Search and discovery of electronic information becomes simpler because there is only a single repository to search for litigation support.

This white paper describes Mimosa NearPoint File and Document 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.

¹ Source: Strategic Research Corporation

² Source: SNIA/Source Consulting

Goal

The primary goal of the NearPoint File and Document Archiving Option is to improve 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 type, size, and age. Selected files are captured, processed, and optionally removed from the source (a process known as stubbing). Captured files are managed in the archive according to retention and disposition policies to meet rules for corporate governance. File 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.

For legal discovery and adherence to the Federal Rules of Civil Procedure (FRCP)³ the NearPoint File and Document Archiving Option 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 identify and access electronically stored information for litigation. Should a lawsuit be filed, the automated retention and disposition policies can be put on hold, preventing any loss of evidence.

The NearPoint platform is a fully integrated archival storage solution that leverages off-the-shelf hardware and the latest low-cost storage technology. Global single instancing across all email, attachments, and files ensures the highest level of archive storage efficiency. NearPoint offers a single point of management of all archive information, a single control of retention and disposition, and a consolidated view for performing eDiscovery searches.

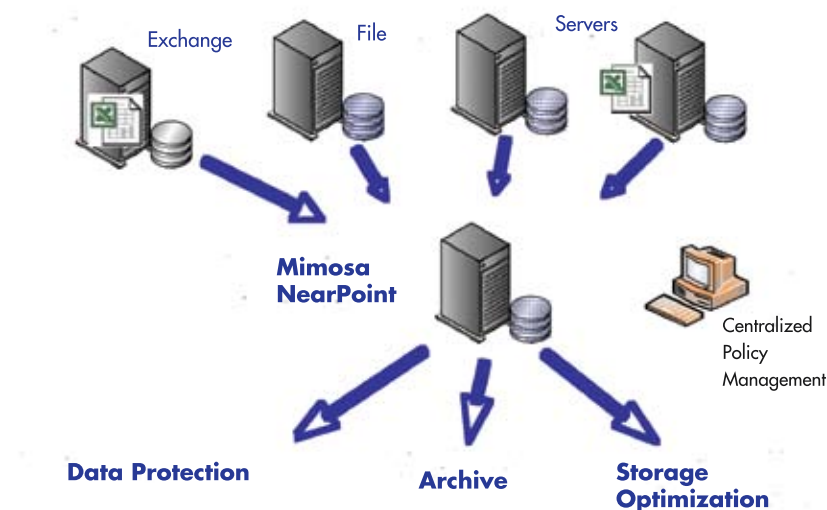


Figure 1. Mimosa NearPoint Archival Platform

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

Key Product Capabilities

Unified Backend for All Content Types

The Mimosa NearPoint archive platform supports all electronic messaging information for Microsoft Exchange Server, including attachments. With the addition of file and document 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—as much as 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 (assets) 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 asset as well as all metadata for the archive. This unified backend 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.
- **Accurate search.** Search is performed in one step, reducing errors.
- **Retention compliance.** Retention is managed with a single policy, reducing errors.
- **Legal compliance.** Litigation holds are managed with a single policy, reducing errors.

Asset Policy Control

For easy application of archival policies and to assign access rights to auditors, the Mimosa NearPoint File and Document Archiving Option manages asset control by Asset Group. Assets are defined as files and documents located on servers, shared files, and folders. Asset Groups manage multiple assets in a single logical group. Asset Groups are defined based on geography, department, or project, and up to two levels of nesting of groups are supported. The Asset Group manages the application of asset capture rules, archival policies, and assignment of access rights to auditors. If two levels of nesting are defined for groups, the rules/policies/rights are applied only 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. The rules are system dependent and are intended to capture key files from the most commonly used folders (e.g., \My Documents\ on NTFS). Capture rules are defined by asset type, asset size, and asset age (creation date, modification date, and access date).

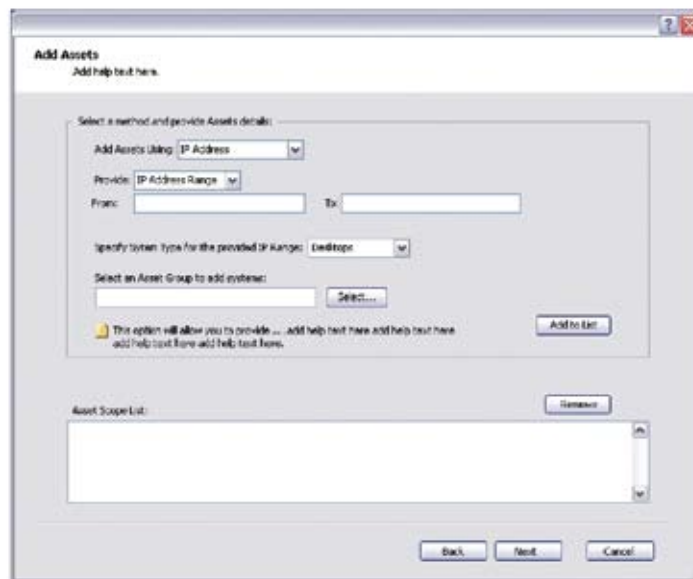


Figure 2. Add Asset User Interface

Crawl Policy

The NearPoint File and Document Archiving Option crawls and captures files within Asset Groups according to administrator-defined schedules. The crawler is agentless and collects files and file metadata from file servers based on asset capture rules. These rules define the scope of files to be captured. Only the files that match the asset 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. Credentials may give read-only or read/write access to partial sets of folders on the shared volumes. 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 or share or system. Files that have been stubbed by NearPoint are automatically ignored during the crawl. By default, NearPoint archives all versions of captured files. Optionally, the administrator can configure how many versions are 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.

Index and Archive Policy

The NearPoint File and Document Archiving Option indexes all captured files within an Asset Group. Captured files that are stubbed (based on group-specific stubbing rules) will also be indexed by default, but indexing can be turned off on a per-group basis. This feature is useful when the intended purpose is only to reduce storage and not to index captured files. Indexing is performed on the file content and metadata, or it can be performed on metadata only. When a search is performed, the item content as well as the item context can be specified in the search criteria. By doing so, the search can be refined to search only within subfolders of a group.

Retention/Disposition Policies

The NearPoint File and Document Archiving Option provides the ability to set retention periods for files 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 an Asset Group and to subsets of Asset Groups. 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, then 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.

File Extension (Stubbing)

The NearPoint File and Document Archiving Option supports stubbing for NTFS and NetApp file systems. For NTFS, the administrator can choose between Internet-style shortcuts (URL) or no stub. The Internet-style shortcut places a URL link in a small stub file. To access the stubbed file, the user double-clicks the stub file to access the URL. The advantages of this approach are simplicity and system independence. Applications (such as backup applications) operate on the stub file and not on the original file. A disadvantage of this method is that a large number of small files can result, causing backups to slow down. Optionally, the administrator can choose to have no stub at all.

The files are deleted from the file system and placed in the archive, without a stub. For NetApp file systems, a seamless-style stub is available in addition to the Internet-style shortcut and the no-stub policy. The NetApp file system has a built-in feature for stubbing called FPolicy. FPolicy is a seamless-style stub accessible via CIFS. It works by sending requests from the Filer via RPC to NearPoint, where the request is satisfied based on access permissions. In all situations, the File Extension Policy can be defined per Asset Group by asset type, asset size, and asset 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. Automatic archive de-duplication improves storage efficiency when duplicate files are archived from different sources.

Search

The NearPoint File and Document Archiving Option is integrated with the NearPoint eDiscovery Option for archive search capability. Administrators grant auditors permission to search Asset Groups, and the scope of the search is limited to files captured within the groups. Search criteria are based on file content and file context. Search criteria include Asset Group, asset, file type, file size, author, date, and keyword. Search results are displayed in a standard format that combines file results and email results in a single view (Figure 3). Search results can be saved and shared among auditors, and individual results can be tagged with predefined tags. For litigation support, results can be put on litigation hold. For export, results are downloaded into standard ZIP file format.

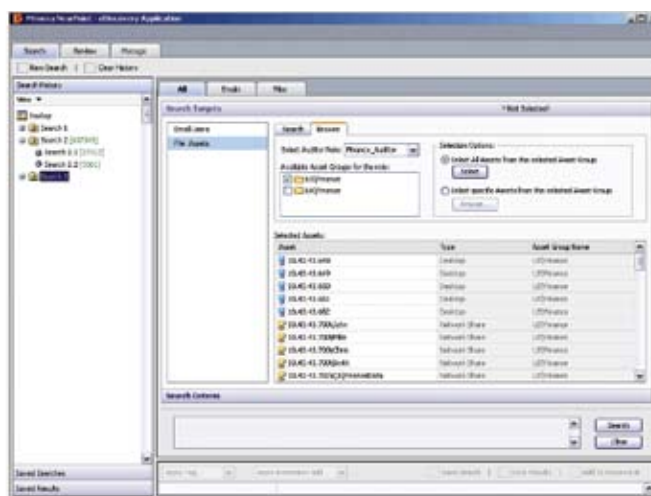


Figure 3. eDiscovery Search User Interface

Monitoring

The NearPoint File and Document Archiving Option 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 Asset Groups by file type. It will also display the number of files archived and indexed, as well as stubbed. Alerts and errors for unreachable systems or files during a crawl are grouped as follows:

- 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)

Example Use Case

Our example company uses the Mimosa NearPoint File and Document Archiving Option to accomplish a variety of tasks using the company's Windows file servers and NetApp Filer that is shared across three branch offices (Boston, New York, and Miami). Each server is shared by several departments (Finance, Accounting, and HR) and has assets (files and documents) spread across shares and folders. For the Boston office, the administrator would like to identify all Microsoft Office files that are greater than 1MB in size, were created less than three years ago, and were accessed less than five months ago from the "finance_2007" folder in the "Finance" share on the "NetApp1" Filer. In addition, the administrator wants to exclude all subfolders starting with the string "acct" from the finance_2007 folder. The crawl policy will archive and index all Microsoft Office documents to enable search. The retention/disposition period is defined as Word and Excel documents with a retention period of five years and all other Microsoft Office files for three years. URL stubs are created for PowerPoint files and Visio files that are greater than 20MB in size (defaults are used for other Microsoft Office files).

For the Accounting department, the administrator would like to identify all files from the "Accounts" share on a Windows NTFS file server, excluding ones in the "Sys" folder, based on Mimosa defaults. The administrator would also like to archive and index all files, setting the retention/disposition period for audio and video files to one year and to the default values for the remaining files. All subfolders starting with the string "acct" from the finance_2007 folder under the Finance share on the NetApp1 Filer are included, and the retention/disposition period for all files under the subfolders is set to three years. Default stubs are created for all files that are more than 1MB in size.

All files in the Accounts share are identified, except for text files that are greater than 2MB in size, and default stubs are used to stub all such files and documents. The data protection period for stubbed files is changed to six months.

For the HR department, the administrator would like to identify all Microsoft Office files from the UNIX file server named "hr_server" from the "hr" share, as well as all files from the "hr_rec" folder under the "common" share, based on Mimosa defaults. All files other than binary, audio, and video files from the "common" share from all other folders are also identified, and all such files are archived, indexed, and stubbed. The retention/disposition period for all identified files under the hr_rec folder is set for five years, and the period for the remaining files is set for two years. An auditor role is created for review of the Finance department in the Boston office, and users are added in that role. An auditor role is created for review of the Accounting and HR departments, and a perpetual crawl is scheduled with specified blackout periods.

Incremental progress is monitored on a per-departmental group basis to ensure that progress is being made and to track the rate at which progress is being made. Questions are answered: How many files were captured? How many files were archived, indexed, and stubbed on NearPoint? Were there any issues that need to be tracked? Was an attempt made to fix the issues if they were not transient? Were there any issues with server/share/folder/file access on the source file system?

Months later, following a re-org, the Finance and Accounting groups merged and so did their assets. Now all subfolders starting with the string "acctn" from the finance_2007 folder under the Finance share on the NetApp1 Filer are included in one of the Finance departmental groups. Microsoft Word and Excel documents that had a retention period of three years now have a retention period of five years. All PowerPoint and Visio files greater than 1MB in size were being stubbed with default stubs; now, URL stubs are created for PowerPoint and Visio files that are greater than 20MB in size (defaults are used for other Microsoft Office files).

Conclusion

The Mimosa NearPoint File and Document 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 File and Document Archiving Option performs automated rule-based archival 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, email, and attachments 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 and Document 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.



© 2007 Mimosa Systems, Inc. All rights reserved worldwide. Mimosa and NearPoint are 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