CONNECTED™

› Connected ArchiveStore/EM

Updated: March 18, 2004

❯ C O N T E N T S

IS AUTOMATED ARCHIVING OF EMAIL REALLY NECESSARY?

Yes. Business use of email is now as pervasive as the telephone and fax. And email is often the more critical communication tool for daily business operations. Reliance on electronic messaging is so firmly entrenched, in fact, as to warrant special attention and regulation from government bodies and industry watchdogs.

This newfound attention, combined with the problem of managing ever-growing stores of electronic messages, is elevating the importance of email archival. Demand for an automated, active email archiving solution has now attained "must have" status in every organization.

BACKING UP EMAIL WON'T KEEP AWAY REGULATORY TROUBLE

Distinguishing between email backup and email archival is critical if regulatory problems are to be avoided. Email backup systems are designed to provide wholesale recovery of the email server, should a disaster befall the production environment. These systems are not designed for compliance or legal discovery-related record retention.

Simple email system backups have no provision for the review of individual email records. Backup processes format the data to reduce storage space and speed future recovery processing. This formatting works against attempts to review and retrieve individual messages.

A true email archiving and retention system ensures, at a minimum, that companies have ready access to any given email record, whenever it is needed. Maximizing the

evidential weight of email records also requires a secure audit trail capable of tracking every action against every archived email message.

The Pressure of Legal Issues

Securities and Exchange Commission (SEC) Rule 17a-4, Sarbanes-Oxley, The Health Insurance Portability and Accountability Act (HIPAA), and The Patriot Act are high-profile examples -- among many dozens of other rules and regulations -- that now govern how a business must manage electronic communications such as email. Retention periods, acceptable types of long-term storage, search and retrieval facilities, and auditing are variously dealt with according to specific industry characteristics.

The Sarbanes-Oxley Act (HE.3763) of 2002, for example, holds CEOs and CFOs personally responsible for corporate wrong-doing. The act requires organizations to demonstrate that auditing is performed, and that records cannot be altered, concealed or destroyed. For electronic messages, with their susceptibility to tweaking and manipulation, this type of requirement poses a significant challenge for IT managers.

Even without the prodding of government and regulatory agencies the need to search through archived email records with very specific selection criteria is fast becoming a business prerequisite. The cost of discovery -- the pre-trial process of investigation during which parties to a lawsuit request access to documentation about a particular event -- can be prohibitive when the only form of email message archival is a raw data backup. Without an alternate means of searching through archived email, discovery costs can be so high that they effectively limit the organizations response during legal disputes, leaving the business exposed to frivolous lawsuits.

Email Lifecycle Management

Large organizations now generate enormous quantities of email. And the rate of growth shows no signs of slowing. Industry analysts predict future email storage needs will increase by 65%, or more, annually. For IT managers struggling to address today's email headaches, this is not welcome news. Email administrators spend a large amount of time manually enforcing email quotas and policies and juggling available storage resources. Incessant demand for more inbox space from end-users only highlights the urgent need for effective email archiving mechanisms.

Striving for Efficient Data Management

Predictably, the tremendous growth in email is having an impact on end-user productivity. Surveys of user habits show that managing email consumes several hours of an employee's day. Finding ways around quota restrictions, such as offloading messages to locally stored .pst files, and searching for old messages are time consuming activities.

Email administrator productivity is also on the wane. Growing stores of email data make it difficult to find a time window to backup mail servers. And, the volume of data stored aggravates even simple daily administration tasks.

THE SOLUTION: CONNECTED ARCHIVESTORE/EM
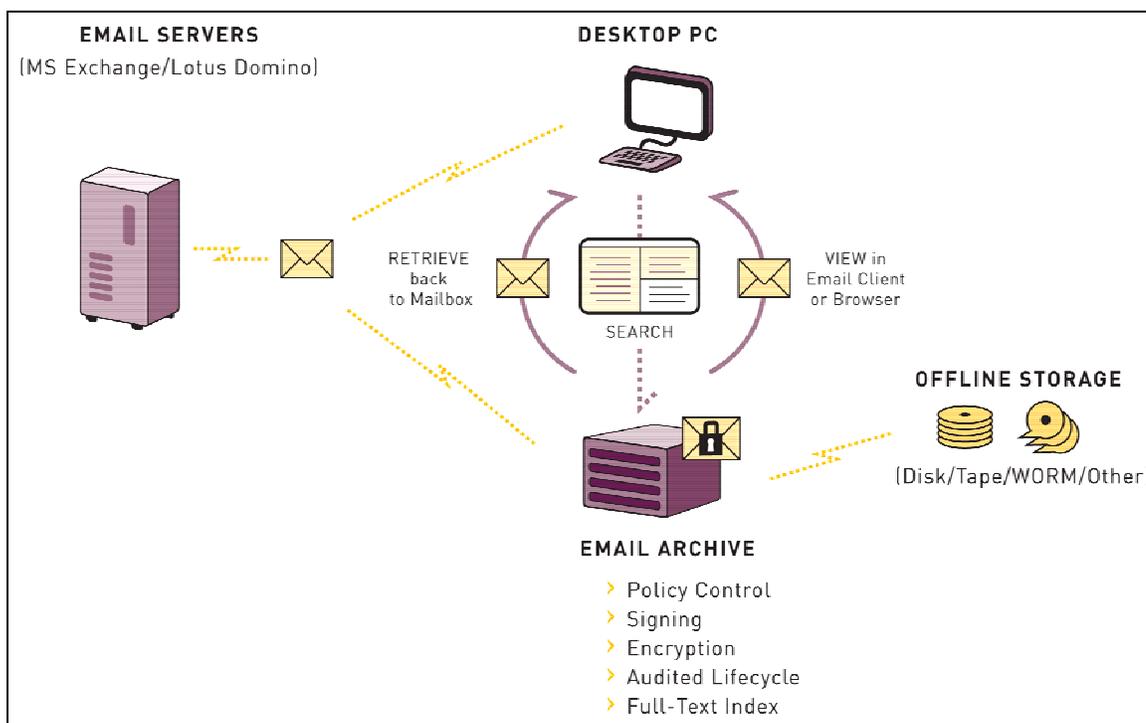
Connected's ArchiveStore/EM provides corporations and their IT managers, email administrators, and end-users with much needed relief in the management of email messages. Offering automatic, secure, searchable, and scalable archiving of email messages for Microsoft Exchange environments, ArchiveStore/EM dramatically simplifies email administration. ArchiveStore/EM delivers the foundation for cost-effective compliance with electronic communication regulations and supports fast, effective legal discovery.

Automatic Archival

ArchiveStore/EM works with the native mail server journaling features of Microsoft Exchange to automatically populate a special IMAP mailbox with copies of every message and attachment processed by the mail server. Each message is retrieved from the mailbox, using IMAP or POP3, by the ArchiveStore/EM Archiver process.

Highly Secure and Scalable by Design

The Archiver captures every email and attachment and immediately compresses the data. Each message is then assigned a unique Advanced Encryption Standard (AES) 256-bit encryption key and encrypted. Finally, the message is digitally signed, using an RSA 1024 bit key. The compressed, encrypted, and signed messages and attachments, normalized for single-instance, are then written to a relational database. Only after the archived message is successfully stored in the database is it deleted from the archive inbox.

Unlike solutions that troll mail servers, ArchiveStore/EM processes messages as they pass through the server. This real-time processing provides airtight auditing and leaves no window for the messages to be tampered with prior to being encrypted and archived. The distributed configuration of the Archiver, which runs as a Windows service, also eliminates the potential for degraded mail server performance due to archiving. ArchiveStore/EM is able to run multiple Archiver processes simultaneously, each accessing a different mail server.

The Email Lifecycle

Each mail message processed by the Archiver is classified according to retention policies that dictate which messages are archived, where they are archived, and for how long. Retention policies are configured by the system administrator and apply to all messages, corporate-wide. Individual users cannot influence which emails are

archived and which are not, ensuring compliance with even the most rigid regulatory controls.

After encrypting a message the Archiver encrypts the encryption key and stores it in a relational database, along with expanded RFC-822 header information and metadata for the message. If ArchiveStore/EM EventSync is used -- to supplement journal processing for Microsoft Exchange 2000 and 2003 mail servers -- the entire SMTP envelope is expanded, allowing a broader range of header information to be stored with each message, for example, BCC addresses.

Searching For Archived Data

The compressed and encrypted email messages and attachments are stored in a relational database prior to being moved to more cost-effective off-line storage. While in the database the Java-based Lucene full-text indexing engine catalogs the subject-line, body-text, and attachments to allow fast text-based search and retrieval of archived data. Native Java open source parsers and Microsoft-specific parsers are used to perform indexing of a wide variety of attachment files including, Microsoft Office documents, PDF files, etc.

Policy Enforcement

The ArchiveStore/EM Policy Enforcement Engine, also configured as a Windows service, is scheduled to run on a regular basis. The Policy Enforcement Engine flags messages destined for near-line or off-line storage to be processed by the Job Queue and Offline Storage Systems. Messages can also be retained in the relational database for fast user access.

If the Policy Enforcement Engine determines that a message has reached end-of-life, the message and metadata for the message are deleted from the relational database. Destroying the encryption key makes it impossible to retrieve the message, effectively erasing the email.

All email destruction is audited. And, after an amount of time specified by the administrator, audit trail records applicable to the destroyed message can also be designated for destruction.

Access Control

Archived email is securely tagged by the Archiver process, restricting access to only authorized users. The ArchiveStore/EM UserSync process automatically

synchronizes user access privileges with the authoritative source of user data - Microsoft Active Directory or an LDAP server. This ensures that end-users are authorized to access only email messages that they were originally party to. For example, if a user is removed from a distribution list ArchiveStore/EM recognizes the change and allows the user to view archived distribution list email before the removal date, but not after.

Archived email messages are accessed via a web browser or a mail client application. The ArchiveStore/EM browser interface provides Google-like text querying of messages and allows retrieved email to be viewed online or recovered to the original inbox. The ArchiveStore/EM system administrator is able to recover mail to any mailbox. Users of Microsoft Exchange 2000 and 2003 mail servers access archived email directly from a Windows mail client using a unique ArchiveStore/EM feature called the stub manager.

Reducing Email Server Storage Requirements

For users of Microsoft Exchange 2000 and 2003 servers, ArchiveStore/EM can reduce mail server storage requirements by up to 70%. The ArchiveStore/EM stub manager searches Exchange Server message stores and replaces archived email with a short stub. The stub is an iframe-based HTTP reference that redirects the mail client to the ArchiveStore/EM database for message retrieval.

Stub processing is completely transparent to the end-user. Access to an archived email message, via a stub, prompts the automatic return of an HTML version of the email from the ArchiveStore/EM web interface. The message appears in the preview pane of the end-user's mail client.

By replacing mail server messages and attachments with HTTP references ArchiveStore/EM dramatically reduces the storage requirements of the Exchange Server. With less storage to manage mail server backup processing is faster and administrators gain greater flexibility in consolidating multiple servers. The use of iframe-based stubs also eliminates the need to impose quotas on end-user inboxes.

Secure Audit Trail

ArchiveStore/EM performs comprehensive auditing of every event in the lifecycle of an email message. Each time a message is stored, viewed, retrieved, or deleted the audit system tracks the change, logging the activity in the database. Any changes made to policy configurations affecting an archived message are also audited.

The encryption and digital signing of all emails and attachments, as soon as they enter the Archiver process, eliminates any possibility of the audit trail being

circumvented. Without comprehensive encryption, this guarantee cannot be made. The combination of strong encryption and a bullet-proof audit trail allows administrators to vouch for organizational compliance with auditing requirements and regulations with complete confidence.

Mail Store Database and Mail Migration Tools

The database used by ArchiveStore/EM is designed for scalability and performance. Currently running on a Microsoft SQLServer2000 database, the ArchiveStore/EM processes access data using JDBC calls. The database is designed for connection pooling, and uses short transactions with minimal locking. Maintenance routines are supplied to perform database backup, index rebuilds, statistics recalculation, and integrity checking.

ArchiveStore/EM provides tools to import offline archives of mail messages into the main database. Messages in the Exchange message store are imported using the Microsoft migration facility and .pst files on user desktops can be processed using a mail import tool.

Technology Standards

Supported Email Servers

The initial release of ArchiveStore/EM supports Microsoft Exchange mail servers. Future releases will support other market leading email servers/clients, such as Lotus Domino/Notes.

Exchange 5.5 provides native journaling facilities allowing ArchiveStore/EM to capture all messages passing through the server. Exchange 2000 and 2003 require the ArchiveStore/EM EventSync process to supplement native journaling.

Supported Protocols

Journaled messages are processed using IMAP and POP3.

User access privileges are synchronized with authoritative LDAP directories or Active Directory sources using LDAP .

Messages recovered from the ArchiveStore/EM database are returned to the originating mailbox through SMTP.

Advanced Encryption Standard (AES) 256 bit encryption. AES is the standard US government cryptographic algorithm for encoding unclassified information.

Support for interfacing with HSM technologies.

Open Source Components

Index Server full-text catalog and search facilities are provided by the Open Source Java-based Lucene full-text indexing and search engine.

Microsoft Office documents (.doc, .xls, .ppt), Adobe PDF files, html documents, and zip files can be indexed using the parsers included with ArchiveStore/EM.

Plugable third-party filters are supported to extend indexed file types.

ABOUT CONNECTED CORPORATION

Connected Corporation is the leading provider of storage software for automated protection, archiving and recovery of distributed data. The company's powerful, yet easy-to-use solutions help customers protect and centrally manage data from PCs, Windows servers and email systems, while eliminating costs and burdens associated with capturing and recovering highly distributed data. Customers range from mid-size businesses to more than 600 of the world's largest corporations, including Boeing, Deloitte & Touche, Deutsche Telekom, EDS, Hewlett-Packard, SAP, Schlumberger and Toyota.

For information phone (US toll-free) 1.800.934.0956 or +1.508.808.7300. www.connected.com