



Evaluating Email Archiving Solutions

Analyst: Dianne McAdam

Management Summary

Years ago, business was conducted primarily over the phone. Today, email has replaced the phone as the primary means for communicating with people inside and outside of the company. Business quickly grinds to a halt when the email system is down. Email has become a mission-critical application.

Many customers negotiate the majority of their contracts and agreements via email; those same companies exchange invoice and payment information by email. Email messages may be the only record of these transactions, and they must be protected and retained.

Email systems are periodically backed up by traditional backup software products that allow the systems to be restored in the event of an outage, such as a disk system failure. However, these backup products do not allow customers to easily search through thousands of emails to find, for example, the messages outlining the payment information that another company had agreed to.

Many software vendors have developed *email archiving* products that allow customers to search their email to find specific messages. In fact, there are so many products that it can be difficult to choose the right one for your environment. This paper outlines the features that you should consider when evaluating email archiving products.

First . . . The Core Requirements

All email archiving solutions must fulfill certain core requirements. Simply put, they must be able to store an email message and any associated attachments, such as a Word document or a PowerPoint presentation, for a specified period. Retention periods must be defined, and not all emails should be treated the same. For example, emails sent to and received from CxO executives may have longer retention periods than those of regular employees. Stock or bond traders, who are involved in regulated industries, will have their retention periods dictated by the regulating committee. Financial traders' policies are dictated by the Securities Exchange Commission. Corporate policy may dictate that personnel records be kept for several years after the employee has left the company. Emails from certain departments, such as payroll, may require a different retention period.

While retention periods should be set and never changed, there may be situations where retention periods must be extended. For example, emails about broker trades have a defined retention period of seven years. But, audits can occur at any time, and it is possible that two months before these emails are scheduled to be deleted, an investigation into a particular broker's trading activities could be initiated. Since the audit will take several months, all emails for this particular broker must be retained until the audit is completed. This extension is sometimes referred to as a litigation hold; it is a core requirement for regulated industries. However, non-regulated companies are still subject to internal audits, external audits by outside accounting firms, and discovery processes by attorneys. For these reasons, companies evaluating email archiving solutions should determine if retention periods can be extended.

IN THIS ISSUE

➤ First ... The Core Requirements	1
➤ Beyond the Core Requirements	2
➤ Future Concerns	2
➤ More Questions to Ask	2
➤ Conclusion	4

While, archiving emails is important to retain the information for future reference, the ability to intelligently search and retrieve those emails is even more important. Email archives can grow very large very quickly. Enterprises may have to search through hundreds of thousands or even millions of messages. Email archiving software must be able to index all emails by standard text fields such as *To*, *From*, *CC*, *BCC*, *Subject*, and *Date*. Enterprises with large email archiving storage also require full text indexing to limit the number of emails returned from a query. For example, if a broker's activities have triggered an audit, a search of all of the broker's email messages over the last several years could result in the return of thousands of emails. However, a full text index can reduce the search to a manageable one hundred emails that discuss the suspicious transactions concerning *XYZ Corporation*.

Beyond the Core Requirements

Some email solutions offer additional features that make them more efficient and cost-effective.

Taming Email Storage Growth

The number and size of email messages grows every year. As they increase, so do the costs associated with purchasing and managing email storage.

How can you tame the monster?

Some vendors compress emails that are archived. Others separate the attachments from emails, storing only one instance of attachments sent multiple times. Each original attachment is replaced by a link to the single stored instance of that attachment. If a 10-MB attachment sent to 20 people is only stored once (10 MB) rather than 20 times (200 MB), you save 190 MB of storage. If several changes are made to a presentation and each version is circulated to a list of people, the storage savings quickly add up. Some vendors both - compress emails and remove redundant attachments.

What About Personal Folders?

Some end users prefer to save emails locally by moving these emails to their personal computer; for Microsoft Exchange, these would be saved as PST files. Corporations where this is a common practice should determine if the archiving product will discover and archive emails stored in PST files.

What About Instant Messages?

For some people, even email messages are not fast enough, as instant messaging (IM) has become the preferred method of communications in many organizations. Corporations in regulated industries must save not only email messages but also instant messages. *Can the email archiving product you are evaluating also archive IM messages, or is an additional product required?*

More Sophisticated Searches

All email archive products can search through email archives to retrieve emails sent or received by specific individuals, or they can search by date range. But it is possible that the email message itself contains very little information - the good stuff may be in an attachment. Some solutions can scan these attachments for specific text strings while others cannot.

Remote Replication

While archiving emails to disk or tape within the main data center can satisfy a corporation's email retention requirements, it provides no protection against an outage in the main data center. Sending a second copy of the archive to a remote location is essential to allow archives to be accessed after a data center failure. Some archive solutions have policies that can direct data to both local and remote locations.

Future Concerns

By its very definition, email archives store messages for a long time. Say your emails are being created under *Exchange 2003*. When these emails are retrieved ten years from now, it is probable that the corporation will no longer be running *Exchange 2003*. It may have converted to a different email system altogether. Enterprises need to determine the best approach to ensure that older, archived emails can be read. Some may decide to retain older versions of the email software. Some archiving software stores emails in a non-Exchange format, such as *XML* or *HTML*. Enterprises carefully need to check existing regulations to determine if emails must be retained in native format or if the regulations allow non-native formats.

More Questions to Ask

It can be difficult to choose the right email archiving solution for your environment. The following questions may help to narrow down the choices.

Supported Environments

1. Which email servers (e.g., Exchange, Domino) are supported? Which email clients (e.g., Outlook 2000, Outlook 2003)? Are these solutions certified by the email vendors?
2. Does the solution support end users who access email remotely through various means, such as Web-based access?
3. Is support available for end users who regularly use home computers to read their email? Is support available for remote users who normally access email "on the road"?

Architecture

4. Is journaling required?¹
5. Is an agent required on the email server? Is an agent required on the client? How are agents installed? Can they be installed remotely?
6. How is the solution implemented?

Performance, Scalability, and Availability

7. What is the performance impact of the archive solution on the email application?
8. How many end users can the archiving solution support?
9. How many messages can be stored per hour? If the amount of messages increases, can the email solution adequately support the additional growth? Do you need to install more servers? What is the upgrade plan?
10. How many email messages can it store? A thousand? A hundred thousand? More?
11. Indexes are created when each message is stored to provide quick access. Where are the indexes stored? How much storage is required to store them? If the index is corrupted, can it be rebuilt?
12. How quickly can the software search through, say, 100,000 messages to find all of the messages sent by "Bob Smith"? When the num-

ber of stored messages increases from 100,000 to ten million or 100 million, is search performance degraded?

13. How many email servers can be supported by the archiving system? How much storage can be supported?
14. Is the archive engine on a clustered server? If the server fails and it is not clustered, what is the impact on the email application? Can the email application continue processing? How do you recover operations when the failed archive engine is replaced or repaired?
15. Does the email archiving solution compress data to reduce storage requirements? If so, what compression rates can be expected? What compression rates have other customers experienced? Are messages compressed? Are attachments compressed?
16. Does the solution detect and remove multiple copies of the same message? Does the solution detect and remove multiple copies of the same attachments? Storing only one instance of the same message or attachment can significantly reduce storage requirements.

Hardware

17. Does the email archive solution use industry-standard hardware? Has it been built to vendor specifications? What is the cost of this hardware? If the hardware fails, can it be easily and quickly replaced? What is the vendor's escalation procedure when hardware fails?
18. What storage devices are supported? Are WORM devices supported?

Administration

19. What happens when a former employee's mailbox is deleted?
20. When a new employee is hired, will the archive solution detect the new mailbox?
21. Can an entire mailbox be restored, if needed?
22. Are PST folders detected and archived? Are public folders detected and archived?
23. Are Instant Messages supported? If not, how will be Instant Messages be archived?
24. Are logs or reports generated that track which employees have retrieved messages from archives?

Policies

25. How granular are the policies? Can different policies be set for different departments or for different groups of workers?
26. Can policies be set to determine when messages are sent to the archive? By age of message? By size? By a combination of both?

¹ There are several different implementations of email archiving on the market today. For example, "Journaling" is a Microsoft-provided feature that archives email messages. When IT administrators turn on the journaling feature, a copy of each message is sent to a different location or recipient. Some email archiving solutions require the journal feature to be available since these solutions are recipients of the journals. Buyer beware! Turning on journaling will affect performance. In fact, Microsoft estimates that journaling will cause 15 to 35 percent performance degradation. (Microsoft's estimates can be viewed in the chapter, "Journaling Impact on User Mailbox Servers," located at <http://www.microsoft.com/technet/prodtechnol/exchange/guides/E2k3Journal/89728bb7-c6ec-47ec-97ed-e21f2e44bc74.mspx>) Other implementations do not require journaling. They use the transaction log to update the archive. This method has much less impact on performance.

27. For regulated industries, can policies be established that require managers to review employees' emails?
28. Can retention periods be extended when necessary?

Security

29. What levels of security are provided? Can employees view and retrieve all messages in the archives or are they restricted to viewing only certain messages?
30. Who has the authority to delete messages in the archives?

End User Interface

31. When a message is archived, how does it appear to the end user? Is the message flagged with an icon to signify that it is archived?
32. When an end user needs to search the archive for messages, is the search menu easy to use?
33. Can end users be restricted to viewing and retrieving only certain messages?

Replication to a Remote Location

34. Does the email solution have policies to route emails both locally and to a remote location? If not, can existing replication products be used to replicate the contents of the archive to a secure, remote location?
35. What are the distance limitations?
36. Will replication impact performance?

Pricing

37. How is the product priced? By number of seats? By the capacity of the archive? By the number of email servers?

Operations and Support

38. Does the product integrate with existing backup/restore applications? Email archiving solutions do not replace existing backup processes, but should integrate with those functions.
39. How long will it take to implement? Is it customer-installable? If not, are there additional charges for installation?
40. Is training required or recommended? What is the cost of that training?
41. What support is available from the vendor? Are support costs included in the purchase price?

Market Acceptance

42. How many solutions have been sold?
43. How many are installed in production environments? Which email applications are in production?
44. Is the product a recent addition to the market or has it been available for months or years?

Future Concerns

45. If the email system is upgraded to a newer release level (Exchange 2000 to Exchange 2003, for example), are any changes required to the archive system? When the email system is upgraded to the latest version two years from now, will the archive vendor support these new releases? What are the archive vendor's plans to maintain compatibility with future releases?
46. Are the messages stored in native format or XML format? If an audit occurs seven years from now, will you be able to read the email messages archived today?

Conclusion

Email archiving solutions allow companies to store emails quickly and easily and to retrieve them when required. But there are other benefits as well. Archiving emails reduces the size of the main email storage. IT administrators (and end users) are no longer required to spend lots of time managing individual mailboxes to ensure that individual quotas are not exceeded. Smaller primary email storage is easier to manage, provides better performance to end users, and is faster to back up. More expensive high-performance disk storage can be dedicated to the primary storage, while less expensive higher-capacity storage can store archived items, saving money.

Email archiving solutions are not just for regulated companies. They provide benefits to all enterprises. Are they worth the cost? Definitely!



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Dianne McAdam is Director of Enterprise Information Assurance for the Clipper Group. She brings over three decades of experience as a data center director, educator, technical programmer, systems engineer, and manager for industry-leading vendors. Dianne has held the position of senior analyst at Data Mobility Group and at Illuminata. Before that, she was a technical presentation specialist at EMC's Executive Briefing Center. At Hitachi Data Systems, she served as performance and capacity planning systems engineer and as a systems engineering manager. She also worked at StorageTek as a virtual tape and disk specialist; at Sun Microsystems, as an enterprise storage specialist; and at several large corporations as technical services directors. Dianne earned a Bachelor's and Master's degree in mathematics from Hofstra University in New York.

- ***Reach Dianne McAdam via e-mail at dianne.mcadam@clipper.com or at 781-235-0085 Ext. 212. (Please dial "212" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, and "clipper.com" are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "Navigating Information Technology Horizons", and "teraproductivity" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.