



Best Practices in Email Archiving

reducing the total cost of ownership of email archiving solutions



Introduction

Business cannot function without email. With over 40% of users checking their email every few minutes, IT managers have had to improve the availability and speed of their messaging platforms, as well as manage increasing email volumes. Email has also emerged as the de facto information management platform, which presents a serious business challenge: although email systems excel at quickly distributing information, most simply fail at reliably storing and providing immediate access to the 83% of business-critical data that they now contain. Surprisingly, the massive investment in enterprise knowledge and document management applications over the past decade has done little to address this challenge.

In addition to being a core business function, expedient access to an organization's knowledge capital found in email messages is also a matter of legal and regulatory compliance. Some regulations require indefinite retention of certain types of data, and organizations without proper tools to store and retrieve critical information contained in emails now routinely face either escalating retrieval costs or severe compliance fines.

Email archiving solutions address these stringent requirements by providing reliable long-term storage of massive amounts of email messages, as well as immediate and secure access to them. In this white paper, we review the best practices in email archiving solutions, with a subsequent focus on reducing their total cost of ownership. We also discuss managed email archiving services, which allow enterprises to quickly implement email archiving, while eliminating upfront capital expenses, reducing maintenance costs, and improving scalability over in-house solutions.

Components of Email Archiving Solutions

Email archives must implement the following four key business functions:

1. Ensure archive completeness
2. Provide efficient and reliable long-term storage of messages
3. Ensure security and integrity of messages in the archive
4. Provide immediate access to archived data for authorized users

Ensuring Archive Completeness

The first task of an email archiving solution is to capture the complete set of business records—messages that contain information “required in the company's day-to-day decision making, financial and business analysis, forecasting and reporting, customer service, resource management, compliance with state and federal laws and regulation, and legal interests” . Therefore, before any archiving solution can be designed or evaluated, the organization's unique mix of business and compliance requirements must be defined and documented in a data retention policy.

The four key questions that must be addressed by the policy are:

1. What information must be archived and where is this information stored?
2. What are the categories to which this information belongs?
3. How long should the various types of data remain accessible?
4. Who is responsible for the archiving and retrieval of the various categories of data, and who has access to it?

Since the completeness of the archive may need to be proven in court, it is extremely important to keep the data retention policy current, and also document the message flow that demonstrates that the archive has access to all the pertinent communications.

In the simplest scenario when all emails are routed through a single server, archiving all the messages on the centralized system is an easy task. However, this is obviously not possible for organizations that have geographically dispersed (federated) messaging systems; in this situation, the archive must be able to capture information from multiple sources simultaneously. For companies that operate in multiple jurisdictions, the archive (and the underlying data policy) may also need to accommodate divergent data retention requirements for each of the locations. Furthermore, since large amounts of information now reside in users' personal stores (e.g., PST files), the solution must support loading offline data into the archive.

Although a minor delay between the time that an email arrives and the time that it is archived is usually acceptable, lengthy archiving interruptions threaten the integrity of the archive. Therefore, to guarantee completeness, archiving systems must be resilient to network, software and hardware failure, and must include high availability features.

In addition, since spam now accounts for up to 70% of all email volume, and computer viruses are on the average found in 1 of every 25 messages, it is important that the message archiving system be tightly integrated with content-filtering solutions, or risk being overwhelmed by irrelevant data.

Finally, what the archive does not contain is often as important as its contents. Email messages that linger in the archive after they are no longer required to be stored—or are specifically required by law to be destroyed—present an additional risk to the organization. Therefore, the message archiving solution of choice must implement automatic policy-driven record deletion.

Providing Efficient and Reliable Long-Term Storage of Messages

The second function of an email archiving system is to provide long-term storage of an ever-growing volume of business communications. This is far from a trivial task, since in certain cases data must be stored for up to 30 years (as with certain industrial and medical records required by the US Federal agency OSHA) or even indefinitely! There are two components required to ensure continued availability of archived data: the archive system must be fault-tolerant and it must support open storage formats.

The cost of disk storage has drastically decreased. As a result, many organizations are foregoing the expense and headaches of managing off-site tape storage, and relying on disk-based archives as a method of backup. This approach is not recommended since it does not protect against catastrophic failures (such as the destruction of the site containing both the primary message store and the archive). Instead, a comprehensive archiving solution should include an off-site storage component, either in form of tape backups (which may be less frequent) or a stand-by system with a full copy of the archive housed in a geographically remote location.

Furthermore, the archive system must have the ability to export its data into open, non-proprietary formats, such as XML or flat text files (or industry standards such as PTS files). This functionality is required to migrate the data into next generation archive systems, as will invariably be necessary in the future—three, seven, or thirty years from now.

The amount of data that must be archived is increasing at roughly 30% per year. While today's typical email is only 20KB, messages of 10MB in size are not uncommon. In fact, with the proliferation of video attachments, analysts predict that message sizes will grow from 50 to 200MB, and a typical corporate user will generate almost 20GB of emails over the next 3-year period. Even with the decreasing cost of storage, a reasonable retention and deletion policy is paramount in keeping the storage costs from escalating out of control. In addition, data compression and single instance storage mechanisms should be used to further reduce storage requirements, as well as improve archive performance.

Ensuring Security and Integrity of Messages in the Archive

The third job of a message archiving system is to secure and preserve the integrity of the data under its care. By design, the archiving system has visibility to most of the emails that contain vital business records, and must ensure that messages cannot be intercepted and viewed by a non-authorized party. All data should be encrypted as it travels from the primary message store(s) to the archive. Furthermore, all data, message headers, and indexes should also be stored in the archive in an encrypted format.

Any messaging archive must have the ability to prove the integrity of its records in court. For instance, a company may have to demonstrate that the messages entered into evidence have not been modified, for instance with a "man-in-the-middle" attack. Proving such integrity is typically accomplished with digital signatures (or hashes) such as those produced by the MD5 or SHA1 algorithms. Processing messages (or message components like headers or attachments) with such algorithms generates a unique digital signature for each message. It is extremely difficult to modify data and yet preserve the original digital signature, and courts now routinely accept digital signatures on par with physical ones as proof of data integrity.

Archiving systems must also provide a complete audit log, which can reliably capture all archive activity, and which can be used to further demonstrate the integrity of the stored messages. A database containing message hashes and the audit logs should also be encrypted or digitally signed, to further reduce the risk of tampering.

In general, archiving systems should rely only on open, industry-standard cryptographic algorithms. Unlike many proprietary algorithms, they are admissible in court because they have been—and continue to be—extensively publicly reviewed for vulnerabilities. Furthermore, if cryptographic keys are used to secure access to the archive, security best practices recommend storing copies of the keys with a third-party escrow service.

Providing Seamless Access to Archived Data for Authorized Users

The three core business functions of email archiving discussed above—archive completeness, reliable long-term storage, and message security and integrity—can to a large extent be accomplished with tape backups. However, the ability to provide immediate access to archived data, and specifically to respond to litigation requests in a timely manner, sets true email archiving apart from legacy solutions. This accounts for the tremendous growth in the message archiving arena.

The archive's most important feature is the ability to search through millions of emails in seconds. The best archiving solutions support both office and mobile workers, allowing access to archived messages via a standard web browser as well as from within proprietary messaging clients such as Microsoft Outlook. Not all archiving solutions are up to the task, given the increasing amounts of data and the variety of formats that must be indexed. For instance, systems that rely on "stubbing"—replacing an archived message with a small pointer, and retrieving the message from the archive on demand within Outlook—place tremendous load on the server, and severely curtail the scalability of such solutions.

As discussed before, archive systems must also offer robust export functionality and a full audit trail, both of which are required for routine business purposes (such as when a person leaves a company or is promoted) as well as legal discovery. Naturally, access to messages must also strictly follow the data retention policy, and the best way to manage individual and role-based permissions is through integration with the enterprise directory services.

Reducing the Total Cost of Ownership

The major benefits of implementing a managed email archiving solution are reduction in the costs associated with:

1. Storage
2. Mail servers
3. Administration
4. Lost productivity
5. Data spoliation

With a properly applied data retention policy, archiving also allows companies to eliminate the wildcard extreme data spoliation costs altogether. Moreover, by utilizing an off-site message archiving service, companies can eliminate upfront capital expenses, reduce maintenance costs, and improve scalability over in-house solutions, while maintaining the required service levels.

Storage Costs

Email volume is expected to double every 3 years. By implementing an intelligent data retention policy and message archiving, organizations can reclaim substantial amounts of disk storage and gain control of storage costs. Email archives store emails more efficiently by implementing compression and eliminating duplicates. Although tape backups are still required for disaster recovery purposes, organizations may also be reduce storage costs by decreasing backup frequency and volume. An added benefit of an off-site message archiving service is the per-user pricing structure, which allows companies to further control their storage costs.

Mail Server Costs

In addition to increased storage requirements, the growth of email volume may require upgraded computing capacity, which translates to more servers, additional software licenses and support contracts. The Radicati Group estimates that the average cost of owning and managing a mail server is \$20,000 per year. By archiving massive amounts of data, organizations may actually decrease the number of servers in operation. In addition, an off-site message archiving service allows companies to forgo the expense of dedicated archiving servers.

Administrative Costs

50% to 75% of the total cost of managing email systems is administrative —network and mail server administrators, desktop support and operations staff. As the number of mail servers decreases, and users start utilizing the archive, IT staff can be reallocated to other areas of the business. Moreover, by outsourcing the installation and management of the email archive, companies can further lower their administrative costs.

Lost Productivity Costs

Ferris Research estimates lost productivity costs associated with managing email at \$120 per user per month. With the growing volume of email, user productivity will further decline as users spend more time on manually archiving messages in PST files, conforming to mailbox quotas, or unsuccessfully searching for old emails. An effective email archive frees users from tedious email management and allows them immediate access to information they need to do their jobs, which instantly improves worker productivity.

Data Spoliation Costs

When vital business data is lost or cannot be quickly and inexpensively retrieved—usually the result of a non-existent data retention policy, or a reliance on off-site backup tapes—organizations face substantial fines and data retrieval costs that spiral out of control. Several recent high-profile examples of multi-million dollar fines that resulted from lost or improperly destroyed data have brought this risk into sharp focus. More importantly, the loss or unavailability of critical business records can severely impact business operations. Data spoliation costs are a risky wildcard for any business without an email archiving solution: a single incident of lost or inaccessible business records can easily result in fines that equal the cost of implementing the archive.

The MessageLabs Solution

The MessageLabs Archiving Service centrally stores all intellectual property found in email communications in a secure, accessible repository, allowing companies to reduce storage requirements on primary email servers by offloading vast amounts of old email to the secure archive, and providing users easy access and management of their archived email. The service uses an on-site appliance to collect messages from Microsoft Exchange via a journaling mailbox, and securely deliver them to the MessageLabs archive. The solution also tightly integrates with MessageLabs Anti-Spam and Anti-Virus services to ensure that only business-related data is archived.

MessageLabs has built its Archiving Service according to the best practices in email archiving. All data resides in the state-of-the-art archive—a highly reliable, secure and scaleable infrastructure based on a distributed network and grid storage architecture. The archiving infrastructure has no single points of failure due to redundant and/or clustered hardware configurations within each data center. MessageLabs ensures data security by encrypting both the data in the archive and all access to it; the encryption keys required to access the archive are stored in the on-site appliance, providing an additional level of security. In addition, multiple copies of encrypted data are maintained on spinning disks at multiple data centers; data is also backed up to tape which is shipped to a secure off-site location.

The MessageLabs Archiving Service allows companies to quickly realize the benefits of email archiving without significant upfront capital or integration expenses, and offers lower maintenance costs and improved scalability over in-house solutions. The service is backed by a robust SLA that guarantees secure and instant access to all archived email messages.

Conclusion

For many organizations, deploying an email archiving solution has become a business requirement as important as providing email services. However, an in-house implementation is a major undertaking, which requires knowledgeable staff, and large up-front investments in hardware, software, storage and support. A managed email archiving service can reduce the upfront costs, and provide businesses better control over the total cost of ownership of message archiving.

About the Author

Dave Zwieback is the Technical Director of inkcom, a boutique consultancy that specializes in infrastructure and security architecture. He can be reached at zwieback@inkcom.com.

About MessageLabs

MessageLabs is the world's leading provider of messaging security and management services to business. Delivered across a globally distributed platform at the Internet level, our fully managed services ensure the integrity of electronic communications, allowing clients to manage and reduce risk while securing their critical infrastructure and information. With 12 data centers on four continents, MessageLabs processes millions of messages every second and provides clients with reassurance and security at a low, predictable cost.

Bibliography

- ⁱ McLaughlin, Laurianne. "E-Mail Volume Expected to Explode." NetworkWorld. 03 Mar. 2006. 30 Apr. 2006 <<http://www.networkworld.com/research/2006/030606-email-explosion.html>>.
- ⁱⁱ Clark, Sue, Mike Davis, Richard Edwards, Andrew Kellett, and Michael Thompson. "Taking the Complexity Out of E-Mail Management." Butler Group E-Mail Management Vendor Solutions Report. June 2005. Butler Group. 30 Apr. 2006 <<http://www.butlergroup.com/reports/emm/>>.
- ⁱⁱⁱ MessageLabs Announces Archiving Service to Meet Growing Compliance and Corporate Governance Requirements." MessageLabs. 30 Jan. 2006. MessageLabs Ltd. 30 Apr. 2006 <http://www.messagelabs.com/publishedcontent/publish/about_us_dotcom_en/news__events/press_releases/DA_150288.html>.
- ^{iv} "Most Companies Neglect Even Basic Email Management." BusinessWorld. 30 Apr. 2006 <<http://itmatters.com.ph/news.php?id=062405a>>.
- ^v Ross, Alan J. "Electronic Records Management." Security Management Online. June 2005. 30 Apr. 2006 <http://www.securitymanagement.com/library/001743_2.html>.
- ^{vi} "Spam Intercepts." MessageLabs. MessageLabs Ltd. 30 Apr. 2006 <http://www.messagelabs.com/publishedcontent/publish/threat_watch_dotcom_en/threat_statistics/spam_intercepts/DA_114633.chp.html>.
- ^{vii} "Virus Intercepts." MessageLabs. MessageLabs Ltd. 30 Apr. 2006 <http://www.messagelabs.com/publishedcontent/publish/threat_watch_dotcom_en/threat_statistics/virus_intercepts/DA_115628.chp.html>.
- ^{viii} Robinson, Dudley. "10/01/1987 - Retention Requirements for Superseded MSDSs." U.S. Department of Labor, Occupational Safety & Health Administration. 1 Oct. 1987. 30 Apr. 2006 <http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_id=19600&p_table=INTERPRETATIONS>.
- ^{ix} Laurie, Ben. "MD5 Collision, Visualised." The Shmoo Group. 31 Aug. 2005. 30 Mar. 2006 <<http://www.shmoo.com/md5-collision.html>>.
- ^x Schwartz, Karen D. "Compliance Worries Drive E-Mail Archiving Market." EWeek.Com. 14 Jan. 2005. 30 Apr. 2006 <<http://www.eweek.com/article2/0,1759,1751654,00.asp>>.