



GFI White Paper

Email archiving

There's more to archiving than keeping copies of old emails

Contents

Introduction.....	3
The constant challenge of email archiving and retention.....	3
The regulation question.....	4
Beyond archiving – Unlocking value from email data.....	5
Summary.....	6

Introduction

Email is the lifeblood of modern business, without which productivity and day-to-day communications soon break down. Incidents such as the failure of BlackBerry maker RIM's back-end systems¹ which crippled millions of BlackBerry email smartphones serve as an easy reminder of just how important email is today.

On top of our dependency on email comes an additional challenge – managing the exponential growth in the amount of email content that every user receives. The endless growth in email volumes makes email archiving a critical business consideration, helping companies to ensure that sensitive and commercially valuable information is not lost to the Recycle Bin but, at the same time, does not create a performance and storage headache for the end user. Users of popular email clients such as Microsoft Outlook will know all too well of the issues associated with allowing their email database (the .pst file) to grow too large – crashes, sluggish performance, slow searches and corruption of the database itself.

However, while the task of archiving this data from the client to a server or central storage repository is critical, it is only one component of a wider issue – ensuring that the data, once archived, does not become dormant and valueless. Just because information has been archived does not mean it cannot or should not be used as a valuable business asset.

Archiving technology has expanded in recent years to address this important issue, implementing advanced searching to enable businesses to more effectively use, recover and extract value from unstructured email archives and repositories, while enabling IT departments to keep on top of storage, network performance and data compliance requirements without incurring excessive cost.

The next step, however, is to be able to use the information contained in the corporate email archive, to glean data on users' behavior that, when processed, provides a wealth of valuable information for management and HR purposes, for example.

The constant challenge of email archiving and retention

To put into perspective the size of the email archival challenge, first consider the volume of email traffic. In 2010 alone, 107 trillion emails were sent at a rate of 294 billion a day.²

With 1.8 billion email users worldwide and 2.9 billion active email accounts in operation that year, up 480 million from the previous year, the prospects for growth in 2011 are huge, particularly with the rise of mobile-email-enabled devices making email sending and receiving easier than ever. The growth in free email services, such as Google's Gmail and Hotmail is also having a profound impact on email volume, with the major webmail services accounting for over a third of email users.³

The predictions for email growth in the next five years make for equally startling and impressive reading, with data from the Radicati Group revealing that by 2015, upwards of 4.1 billion email accounts will be in active use⁴, a seven per cent year-over-year rise in users based on its data.

This research shows not only the volume, but also the geographical split, with 49 percent of users in Asia, 22 percent in Europe and 14 percent in North America. This immediately presents a network issue for many companies looking to retain email archives centrally, but manage users across multiple sites and countries. Networks can quickly become overwhelmed by large volumes of archival and recovery traffic, not to mention requiring significant storage to retain the emails, many of which are duplicates sent to multiple people in the same organization.

Deduplication is a critical email archiving innovation, allowing companies to archive only one instance of a message, but return it to all the related parties should it ever need to be recalled. Permissions can also be maintained to prevent compromise of privacy and confidentiality by keeping that one email in a central storage repository, rather than maintaining individual email archive silos for each user.

Not only is this a more efficient way of managing email retention, it is more cost effective. By deduplicating emails, especially ones containing large attachments such as presentations, videos, audio and large images, the cost saving on storage space can quickly add up. For example, offloading an Exchange server to a dedicated storage repository for archive material in this way can deliver anywhere from a 20 percent to 80 percent improvement in overall storage costs, depending on the individual circumstances and numbers of users within an organization.

Pooling the archive into a single repository divided by permissions and access rights can also simplify maintenance and monitoring, as well as enable faster and more accurate searches for information and mining of collective data for customer relationship management, business intelligence and compliance purposes.

The regulation question

A growing overhead imposed by regulation is not only requiring organizations to retain vast amounts of email data, but to recall it quickly when requested for legal or public accountability reasons. In the UK, the Freedom of Information Act⁵ has had a big impact on government departments and private sector organizations undertaking public sector contracts. The Freedom of Information Act requires public bodies to hand over historical data in response to reasonable requests from the general public, and to do so within a legally mandated time. Failure to deliver the information within that timeframe can leave the body open to fines and further prosecution. Elsewhere around the world there is regulation that requires similar, time-sensitive recall of information, particularly email conversations, in order to satisfy legal discovery requirements. In the US, Sarbanes-Oxley and HIPAA have placed such burdens on general businesses and the healthcare sector, making effective archival and reliable, fast recovery of the right information essential to minimize legal costs and avoid further action.

Further legislation requires the minimum retention of email. In the UK, the Data Protection Act⁶ mandates the retention of email for five years, more for specific vertical sectors such as healthcare. In the US, the requirement can vary from state to state, making the process all the more difficult.

Of course, all this is irrelevant if the organization is unable to capture the data in the first place. Effective email archiving is only as good as its ability to capture the data, requiring regular and reliable archiving processes to take place. It is paramount that any email archiving implementation is integrated into the mail server to ensure that the amount of live data on the server is kept to the minimum necessary to serve the business need, and ensure that data is not put at undue risk by being left on the main server for prolonged periods of time, or left to be deleted – intentionally or accidentally – by end users.

Keeping a tidy and compact Exchange server has implications for the client application as well. A compact Exchange mailbox for each user in turn minimizes the size of the local replicated archive, which sits in a Microsoft .pst file. As these files grow in size, they hamper both the performance of Outlook and of the overall machine. With email search now an integral part of the overall search and indexing functions within Windows 7, ensuring Exchange efficiency is paramount to maintaining productivity on the desktop.

Finally, what is archived is as important as how soon and how long it is archived. Data suggest that almost 90 percent of all email sent is spam. While spam filters do an excellent job of blocking the vast majority of this, evolving spam content and techniques mean that some will still get through. Archiving this material is not essential to the business, but may not be practical to screen out. Archiving multiple copies is even more non-essential. Attachments have a major bearing on archive size and efficiency. Certain attachments, such as mp3 music files, photos, videos and executable applications may be useful at the moment of first sending, but are not core to overall email retention needs, or indeed may not be work-related at all. Filtering attachments at the point of archiving can drastically reduce the amount of space needed for an organization's email archive, as well as help reduce the risk of malware-infected files being stored for posterity.

Beyond archiving – Unlocking value from email data

Email archiving has long been viewed as a static technology – taking email from a live email archive and placing it in a storage silo, often placing that silo offline on cheap mass storage such as tape, CD or offline disk.

While that can be a secure and reliable way of capturing a snapshot of email data and restoring it later, it is just that – fixed, static and isolated. The process of accessing and recovering data once it has been archived can be slow, expensive and unproductive. It also prevents users from regularly mining the archive for useful historical information that will aid present and future business activities, such as maintaining customer relationships using historical email data for basic customer relationship management, as well as upselling and cross-selling based on previous transaction information held in email.

Email archiving technology has not only expanded to work within online storage scenarios – where the archival platform is always available for recovering and viewing historical data – it has also added powerful data mining techniques to allow users and administrators to search, examine and collate historical unstructured email in order to extract long-term value from it.

This form of unstructured data management is new in the email archival space, having previously been left to separate applications and platforms to index and enable searching of archives. However, the addition of advanced data mining within the email archival platform delivers new opportunities for all users to tap into their historical email data at will, taking the art of data mining out of the hands of IT experts and placing it in reach of everyone within the organization.

By mining email archives, enables organizations to learn a great deal about the interactions with outside contacts, as well as internal communications.

For example, email analysis in a sales-driven environment can show:

- » How quickly do people reply to email?
- » Are sales leads or support queries being responded to quickly enough?
- » Are service levels being adhered to?

The same techniques can be used to enable the smooth transition of work from one individual or team to another, removing the need for time-consuming manual handover reports to be assembled from scratch.

Using data mining and reporting tools, users and organizations with the appropriate access can immediately and securely see who employees have been regularly communicating with, building an instant list of key contacts, top customers, most relied-on suppliers etc. The same can be used for data loss prevention, allowing the IT team to see if information has leaked outside the organization to competitors, or has been transported outside of the corporate network in breach of IT policy on copying email to personal accounts.

The possibilities for HR, IT and team leads goes further, as email archive data mining can help in identifying ghost or dormant accounts, shutting them down as necessary and diverting active email addresses to users who can respond to them.

Content can be controlled and analyzed, allowing IT to identify instances of email being used for inappropriate material, particularly attachments. As well as allowing organization to identify instances of inappropriate material passing through the email system and taking appropriate action as necessary, the same data analysis can provide an insight into what legitimate content, including attachments, is passed through email on a daily basis. This information can then be used to modify IT policy accordingly to account for new and evolving content requirements, as well as helping to make the case for future investment in storage or for excluding unnecessary attachments from the email archive in future.

Summary

Data is an asset, no different from the computers, filing cabinets, tables and chairs within your organization. Unlike these traditional physical assets, the value of data as an asset can be harder to unlock due to the way companies often store and retain both structured and unstructured data across a myriad of disconnected silos.

Unlocking that data can be difficult at the best of times when it is distributed across multiple silos, but when it is archived, it runs the risk of becoming stale, and being taken offline and out of sight. When that happens, the value of that data soon diminishes, while its ability to keep on contributing to the success of the business fades.

The need to retain information within an organization, regardless of size and regardless of whether it is private or public sector, is one of the most critical IT functions of the moment. It is paramount that information is stored safely, securely and reliably. Also critical is that the information being archived can be retrieved quickly and efficiently, and that a minimum amount of duplication takes place during the archiving process in order to manage storage cost and maximize efficiency.

Email archival solutions are largely differentiated on price, with most offering the key three elements of offloading Exchange, data retention and compliance. The new generation of archival solutions is adding a fourth element – data mining.

Once a reliable and comprehensive data archive is achieved, a variety of opportunities are open to an organization to use that data more effectively, unlocking additional value from the information and giving users access to a richer, more detailed account of previous interactions, transaction and allowing organizations to keep account of liabilities, risk and exposure via email.

1. <http://www.bbc.co.uk/news/technology-15243892>

2. <http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers/>

3. <http://www.email-marketing-reports.com/metrics/email-statistics.htm>

4. <http://www.radicati.com/wp/wp-content/uploads/2011/05/Email-Statistics-Report-2011-2015-Executive-Summary.pdf>

5. <http://www.legislation.gov.uk/ukpga/2000/36/contents>

6. http://www.ico.gov.uk/for_organisations/data_protection.aspx

USA, CANADA AND CENTRAL AND SOUTH AMERICA

15300 Weston Parkway, Suite 104, Cary, NC 27513, USA

Telephone: +1 (888) 243-4329

Fax: +1 (919) 379-3402

ussales@gfi.com

UK AND REPUBLIC OF IRELAND

Magna House, 18-32 London Road, Staines, Middlesex, TW18 4BP, UK

Telephone: +44 (0) 870 770 5370

Fax: +44 (0) 870 770 5377

sales@gfi.co.uk

EUROPE, MIDDLE EAST AND AFRICA

GFI House, San Andrea Street, San Gwann, SGN 1612, Malta

Telephone: +356 2205 2000

Fax: +356 2138 2419

sales@gfi.com

AUSTRALIA AND NEW ZEALAND

83 King William Road, Unley 5061, South Australia

Telephone: +61 8 8273 3000

Fax: +61 8 8273 3099

sales@gfiap.com

For a full list of GFI offices/contact details worldwide, please visit <http://www.gfi.com/contactus>



Disclaimer

© 2011. GFI Software. All rights reserved. All product and company names herein may be trademarks of their respective owners.

The information and content in this document is provided for informational purposes only and is provided "as is" with no warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, and non-infringement. GFI Software is not liable for any damages, including any consequential damages, of any kind that may result from the use of this document. The information is obtained from publicly available sources. Though reasonable effort has been made to ensure the accuracy of the data provided, GFI makes no claim, promise or guarantee about the completeness, accuracy, recency or adequacy of information and is not responsible for misprints, out-of-date information, or errors. GFI makes no warranty, express or implied, and assumes no legal liability or responsibility for the accuracy or completeness of any information contained in this document.

If you believe there are any factual errors in this document, please contact us and we will review your concerns as soon as practical.