Important Considerations When Migrating Archives to the Cloud

An Osterman Research White Paper

Published July 2016

TransVault™

EXECUTIVE SUMMARY

Email archiving has been a best practice for many years across all industries, but particularly in heavily regulated industries like financial services, healthcare, energy, life sciences, government and several others. Organizations have invested significantly in legacy archiving platforms, many of which no longer adequately serve their needs in terms of performance, scale, compliance or feature sets. Consequently, organizations must find a way to migrate their archived content to new platforms, either on-premises or in the cloud. The latter has become particularly important as a driver for migration of archived content, since a growing proportion of organizations are moving to cloud-based email platforms, most notably Microsoft Office 365.

However, archive migration is not a simple task: in fact, it's rather complex and requires considerable forethought and planning in order to migrate data reliably while also maintaining the integrity and preserving the full chain-of-custody for migrated data. Further complicating any migration effort is the need to maintain access to archived content for end users and other stakeholders, all while migrating data within a reasonable timeframe and with as little impact as possible on normal business operations.

Migrating from Exchange journal archives or other single instance archive formats to Office 365 – one of the most common types of archive migrations occurring today – necessitates a paradigm shift because Office 365 does not use the single instance format used for the journal archive in Exchange. While a migration to Office 365 can be performed reliably and in a way that maintains chain-of-custody, it requires a unique approach that will minimize the impact on network bandwidth and that will permit an organization to retain its existing investments in eDiscovery and other tools. (It is important to note that journal archive migration is different than migration of user archives, a distinction we discuss later in this paper.)

ABOUT THIS WHITE PAPER

This white paper was sponsored by TransVault – information about the company and its relevant solutions is provided at the end of this paper.

WHY YOU MIGHT CONSIDER MIGRATING YOUR EMAIL ARCHIVES

OVERVIEW

Email archiving is an essential best practice for many organizations. A key element of an organization's long-term archiving strategy is the migration from one archiving solution to another as organizational needs change, companies merge or acquire one another, compliance obligations evolve, and as new email platforms are adopted.

Archive migration is a critical issue: decision makers take an archive migration lightly at their peril, because a failure to protect the security, integrity and authenticity of content can result in a variety of negative legal, compliance and user productivity consequences. This section considers the primary drivers that necessitate archive migrations.

THE FIVE KEY REASONS TO CONSIDER MIGRATION OF LEGACY EMAIL ARCHIVES

There are five primary reasons to consider migrating to a new archive or Cloud for your email, although the importance of each issue will vary by company and industry:

- **Migration to cloud-based archiving**
  Archiving content in the cloud has become an increasingly popular alternative to on-premises archiving solutions. A cloud archive can offer an optimized storage solution for long term retention, along with various other cloud managed services, such as more advanced access security, regulatory policy management,

data management, and more efficient litigation support services. The cost savings of cloud archiving versus on-premises archiving can be significant when the upfront costs of hardware, software, additional experienced personnel, and annual support are taken into consideration. The recent shift in more and more businesses choosing to migrate to Microsoft Office 365 is a good example of where the market is headed in this regard. Because of these differentiators, many organizations are moving their archiving needs to the cloud, requiring current on-premises archive data stores to be migrated as well, to realize the benefits from cloud archiving adoption.

- **Data consolidation**
  Data consolidation is another important driver for archive migration. For example, as storage resources age and become less efficient, they are replaced with higher capacity and more technically advanced storage devices. Storage consolidation can reduce the number of devices under management, reduce hardware support costs, and can increase overall system performance. When starting a storage consolidation project, migrating existing data to the new storage resources requires a carefully considered and planned data migration process so that reliable file and metadata conversion, full data access, and performance is assured.

  Another driver for migration in the context of data consolidation is that organizations often find themselves with multiple, incompatible legacy archives that might include an email archive, a file system archive, several SharePoint instances, and a content management solution. This increases costs, increases legal hold and eDiscovery risk, and reduces productivity. Consolidation into a single, higher performance archive can produce a positive return on investment (ROI) based on improved search performance and fewer hardware and software support costs.

- **Upgrading to improved archive technology**
  Many archiving solutions have become outdated because of new file formats, access control requirements, new security/privacy requirements, and other reporting requirements and eDiscovery capabilities, as well as advances in both hardware and software. Legacy archiving solutions are often a barrier to benefiting from new operating systems, hardware performance and functional improvements, especially when moving to a platform as robust as Office 365 which allows for a one-stop-shop for Compliance (see *Making Office 365 One-Stop-Shop for Email Records Compliance*[1]).

  Legacy archiving systems rely on older indexing and search technology, resulting in less reliable and slower search response times as data stores grow. For many companies, search performance has become a major bottleneck and liability. The only sure way to improve this situation is to upgrade to a higher performance archiving solution. A key part of this upgrade process is the migration of the existing archive data set in such a way that takes advantage of the new system capabilities with the existing archived data set.

  Another important consideration is scalability, which is often overlooked or, at least, underestimated. Many legacy archiving customers have found out too late that the archiving vendor's goal of unlimited scalability and performance has not met expectations. In many instances, these shortcomings have arisen as the number of customers has grown at a rapid pace, thereby increasing the amount of content archived. In some cases, organizations were faced with an especially large eDiscovery requirement and found that eDiscovery searches were taking days or weeks to finish. These situations can place the organization at risk and dramatically increase their legal and other costs. The only option is to move to a higher performance archive that offers improved scalability.

---

[1] Source: http://bit.ly/1Zh2Zje

2

In today's business climate, organizations need to take advantage of any and all improvements to stay competitive so that they do not violate their legal and regulatory requirements – an upgrade to a new archiving solution can be an effective way to do that.

- **Improving compliance with litigation requirements**
  An essential litigation preparedness requirement is that all potential parties to an actual or anticipated lawsuit find and secure all relevant content with a secure legal hold, thereby ensuring that evidence destruction (spoliation), including inadvertent deletions, cannot take place.

  An essential phase in the eDiscovery process is Early Case Assessment (ECA), the process of reviewing potentially relevant case data and evidence to estimate risk, cost and time requirements. ECA is used to set the appropriate go-forward strategy to prosecute or defend a legal case, addressing the issue of whether or not an organization should contest or settle a case.

  With older archiving solutions that offer only basic indexing and slow search performance, ECA can be incomplete and time consuming. Because of the timeframes involved, these solutions may not leave decision makers with enough time to properly review evidence to create an appropriate case strategy. In organizations with normal or heavy litigation profiles, migrating archived data stores to higher performance archive repositories can quickly produce cost savings and reduce the risk of inappropriate case strategies. Further, organizations with multiple repositories to search – such as a legacy archive and a new Office 365 environment – face higher costs and a greater risk of not producing required content.

- **Improving compliance with regulatory obligations**
  Highly regulated companies like broker-dealers, investment advisors, healthcare providers, energy companies, and pharmaceutical manufacturers face strict regulatory oversight. New, more onerous regulatory obligations focused on corporate data are becoming more common. Many organizations have employed archiving platforms over the years to specifically address regulatory retention requirements. As those laws were updated and new ones adopted, data retention became just one of the requirements to be met. New laws require companies to actively protect and audit the access of customers' personally identifiable information (PII) and protected healthcare information (PHI). Many older archiving platforms were not architected with these higher levels of security in mind. For those companies that cannot (or will not) protect that information, lawsuits, large fines, and reputational damage can be the result. Migrating archived data to more secure repositories is the logical step to reduce that liability. Moreover, many organizations have a need to review a large number of messages on a regularly scheduled basis and there is a need to enable workflows that can make this process more efficient – an archive migration can be the best way to ensure that this happens.

KEY ISSUES TO CONSIDER

As organizations consider their options related to migrating to a new archive, there are a number of important questions that decision makers need to ask of vendors that may be selected to assist them with the migration process. While the specific answers to this checklist of questions – and their relative importance – will vary based on a number of factors, all of these questions should be considered by organizational decision makers charged with overseeing and managing the migration effort.

INGESTION CONSIDERATIONS

Because archives can grow to enormous size, it is essential to plan for the methods by which data will be ingested into the new archiving solution:

- If the transfer of data will take place over the Internet, how fast can the new solution ingest data?

- What limitations exist on ingestion capabilities?

- What is the bandwidth available for the ingestion process?

- How reliable is this bandwidth?

- Does the vendor offer the ability to receive physical media (e.g., DVD-ROMs or hard drives) that contain archived data as another avenue for ingestion?

FUNCTIONAL CONSIDERATIONS

There are a number of functional requirements that must be considered about how the data will be transferred from the old to the new archive:

- In what format(s) has the existing archive(s) data been stored?

- Can archived data be converted to another format?

- How old is the data in the current archive?

- Where is this data stored?

- To what location will the data be transferred?

LEGAL AND REGULATORY REQUIREMENTS

Another essential element of the migration process involves the legal and regulatory requirements, or corporate governance to which the transferred data is currently subject and to which it will be subject once transferred:

- How important will chain-of-custody be for the data during the migration process?

- How will an organization's regulatory obligations be met during the migration?

- Is there a potential for spoliation of evidence, risk of losing chain-of-custody, or somehow risking the authenticity/integrity of archived data during the migration?

- If so, what processes does the migration vendor have in place to mitigate these risks?

CONSIDERATIONS FOR END USER ACCESS TO THE ARCHIVE

Another important consideration for an archive migration is ensuring that users have access to archived data during the migration process:

- What will be the impact on end users during the archive migration process?

- Will users be able to access archives during the migration?

- Will users be able to access and navigate migrated data easily after the migration?

- Will any end user training focused on accessing their archived content in the new archive be required?

- If so, what impact will this have on user productivity, training requirements, etc.?

OTHER ISSUES TO CONSIDER

An archive migration project includes a number of other issues that are important for decision makers to consider as they evaluate vendors of migration solutions:

- Will the archive migration be able to occur within the time frame allotted to it?

- What impacts, if any, will a delay or unanticipated problems in the migration process have on the various stakeholders that require access to archived data, including legal, compliance, senior management, end users and other staff members?

- Have decision makers determined what data can properly be disposed of before the migration occurs? In other words, has the organization established a defensible deletion policy that will guide their decision-making process in this regard?

- Who makes the final decision about what data will be retained and what will be discarded before the migration process begins?

- Does a mechanism exist to accurately determine if every instance of a message or other content marked for deletion has actually been deleted?

- Among decision makers charged with managing the migration process, has a discussion occurred about whether or not the organization should simply keep everything instead of discarding some data?

THE IMPORTANCE OF MAINTAINING COMPATIBILITY FOR COMPLIANCE

A key element in the migration from one email platform to another is maintaining compatibility during and after the migration. The most common archiving migration discussion today is arguably the one focused on migrating from on-premises Microsoft Exchange to Office 365, given the large installed base of the former and the increasing popularity of the latter.

Compatibility in the migration from Exchange to Office 365 is a critical consideration because of the differences in the way that content is archived in both platforms:

- Exchange employs a journal mailbox that maintains an immutable copy of every email sent to, sent from, or sent within an organization. Because a journal mailbox (sometimes called a "compliance archive") maintains a complete set of all emails that flow into, out of and within an organization, it is an essential and effective tool for eDiscovery and other regulatory compliance requirements. Mailbox/user archiving, on the other hand, involves the process of moving content from individual users' mailboxes to archival storage for the purpose of maintaining a searchable archive on less expensive storage. In short, a journal mailbox is used to store a true and accurate copy of email for strategic, compliance-related purposes; whereas mailbox/user archiving is designed to satisfy tactical, user-driven archiving requirements.

- The journal in Exchange, as well as that of some other email platforms, stores metadata that is essential for the legal, regulatory and other compliance obligations to which stored messages might be subject. Data is stored as a single instance in order to maximize storage efficiency. If single instance storage was not employed for the journal mailbox, storage requirements would be significantly greater.

- Archiving in Office 365 is substantially different than archiving in Exchange. Because Microsoft needed a way for Office 365 to scale more efficiently than

would have been possible using the archiving paradigm in Exchange, single instance storage is no longer used. Instead of a single repository used to store messages, journaled content has vital compliance data stored as metadata in special fields that are unique to Office 365 and that must be retained for each individual sender/recipient as per the compliance requirements of the organization or industry. As a result, Office 365 discovery does not recognize where to locate compliance data when held in the Exchange single journal message format. Moreover, Microsoft's terms of use specifically prohibit using a single mailbox to store emails that belong to multiple users[2].

THE SOLUTION WHEN MIGRATING FROM EXCHANGE TO OFFICE 365

Because of the fundamental differences in the way messages are archived in Exchange and Office 365, the solution when migrating from the former to the latter is to expand existing archives to the new format in a multi-step process:

- Analyze the individual messages in the central journal.

- Identify all of the valid recipients in each message.

- Create individual journal messages by taking each single instance message and creating as many instances of each message as necessary for each recipient. For example, while a message in Exchange that was sent to four recipients would be stored in a single instance, the migration of this archived message to Office 365 would be converted to four individual messages plus the sender's copy.

- Each of the individual messages will then be stored in each recipient's In-Place Archive or Recoverable Items Folder (RIF).

ADVANTAGES AND DISADVANTAGES

This approach to migrating from Exchange to Office 365, while necessary, brings with it both advantages and disadvantages. Among the advantages are:

- It permits an organization moving to Office 365 to fall within Microsoft's license agreement, since dumping journal messages into a single mailbox in Office 365 directly contravenes the Microsoft license agreement.

- However, creating "non-standard" mailboxes will result in confusion and complexity when it comes to future eDiscovery. This is particularly important given that Microsoft is working hard to put the task of eDiscovery into the legal team (i.e., making it a non-technical function). As such, the legal team may be unaware of a historic "workaround" and inadvertently exclude potentially responsive content in their ECA and eDiscovery searches.

- It permits the most cost efficient use of Microsoft licensing because organizations can migrate data that belongs to "leavers" first. This has the benefit of being able to apply "live user" licenses to leaver mailboxes for the duration of the ingestion process, and then reassign those licenses back to the live users for production use. It is essential to note that a read-only mailbox in Office 365 does not require a license to be in place after the migration has finished and no new data will be written to that mailbox.

- It supports a variety of archive formats, including Veritas (formerly Symantec) Enterprise Vault, EMC SourceOne, HP/Autonomy EAS and other proprietary formats. Additionally, dedicated on-premises journal appliances and third-party, cloud-based journal services are supported.

- It supports the migration of .pst, .msg and .eml formats.

---

[2] https://technet.microsoft.com/en-GB/library/exchange-online-limits.aspx

- It maintains compliance with existing eDiscovery tools.

- Finally, organizations can now put all of their archived content into Office 365, thereby delivering benefits in terms of managing and discovering all of their content in one archive. This can result in reduced search time during activities like eDiscovery and early case assessments, as well as reduced costs and less exposure to risk from the maintenance of multiple archives.

However, there are some disadvantages to this approach, as well. For example, it can dramatically increase storage requirements because single instance messages in Exchange can expand into many times the number of messages in Office 365, and it requires the use of higher level/more expensive Office 365 plans (E3 or higher) that provide unlimited storage in the In-Place Archive and its Recoverable Items Folder.

Moreover, while the dramatic increase in the storage of journaled messages can potentially slow ingestion of archived content into Office 365, there are approaches that can mitigate this problem. For example, TransVault's CloudStream offering is a proprietary protocol to ingest data into Office 365 that was built in collaboration with the Office 365 products team. CloudStream uses techniques like traffic flow analysis to ensure that data is moved into Office 365 efficiently, thereby minimizing network traffic and maximizing ingestion rates.

It should be noted that the dramatic increase in storage when migrating content from the single-instance storage format to Office 365 is an issue for Microsoft, not their customers of higher level plans, since Microsoft offers unlimited storage in the In-Place Archive. While some may fear that Microsoft could change its mind and one day limit the amount of storage in the In-Place Archive (akin to what the company did by promising unlimited OneDrive storage for lower level Office 365 plans in October 2014 and then placing a cap on storage in November 2015), Osterman Research does not believe this is a likely scenario for its enterprise-level customers.

SUMMARY AND CONCLUSIONS

There are numerous reasons for organizations to migrate their existing email archives to newer, better, faster or otherwise improved archiving solutions. A key driver for many organizations migrating their archives today is the rapid adoption of cloud-based email platforms, and more specifically Office 365 as a replacement for on-premises Exchange.

The migration of Exchange or other archives to the Office 365 archive format necessitates a fundamental change in the archiving paradigm to which Exchange administrators have become accustomed, but a reliable and proven mechanism exists to do so. This mechanism will not only reliably migrate content from a variety of archiving platforms to Office 365, but will do so while maintaining the integrity and chain-of-custody for migrated content, and it will preserve organizations' investments in eDiscovery and other tools.

ABOUT TRANSVAULT

TransVault's products build on over 23 years' experience with high-end corporate messaging systems and 12 years of implementing and supporting enterprise archiving solutions. This has resulted in a product suite able to address the unique and often very site-specific challenges that can arise during migration.

To date, TransVault is behind more than 1,200 migration projects, including some of the largest in the world.

Its proven track record and status as a Microsoft Gold, Symantec (now Veritas), IBM and HP partner means you can rely on TransVault to ensure that your email records remain a discoverable asset as your IT strategies and infrastructures change.

TransVault products are available exclusively through a specialized international network of certified partners that offer a range of bespoke migration services for their customers.