

Realtime
publishers

The Essentials Series:
Operations Benefits of Email Archiving

Reducing the Time Required for Data Recovery

sponsored by



by Jim McBee

Reducing the Time Required for Data Recovery.....	1
Email Archiving Reduces Email Storage Requirements	1
Shrinking the Exchange Server Database	2
Reduce Your Recovery Time Objective	3
Improve Your Recovery Point Objective	4
Dial-Tone Recovery	5
What Is the Right Approach?.....	5

Copyright Statement

© 2008 Realtime Publishers, Inc. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtime Publishers, Inc. (the “Materials”) and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtime Publishers, Inc or its web site sponsors. In no event shall Realtime Publishers, Inc. or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtime Publishers and the Realtime Publishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtime Publishers, please contact us via e-mail at info@realtimepublishers.com.

Reducing the Time Required for Data Recovery

Many users feel that they could go without telephones for an extended period of time but couldn't work without their email. Email has become the preferred communication method for most users, and their Inboxes contain information that is necessary for the user to communicate with customers and management, process internal requests, handle accounting matters, and more. Almost every aspect of a typical user's job revolves around incoming and outgoing email and referencing existing email. Consequently, IT is under constant pressure to provide better system availability and, in the event of any type of failure (from database corruption to wholesale server failure), a plan to rapidly recover services.

At first glance, disaster recovery and email archiving do not seem synonymous, but consider that the archive contains some (or all) of the users' email data. Further, the archive takes a potentially tremendous burden off the mail server's storage, thus reducing the amount of data that needs to be restored in the case of failure. An email archiving system can aid disaster recovery in several ways:

- Reducing the amount of data on the production mail servers
- Reducing the time required to recover data
- Improving the efficiency of data recovery

This article will reveal some of the hidden benefits of email archiving from the perspective of data and disaster recovery.

Email Archiving Reduces Email Storage Requirements

The primary goals of any email archival system are usually to reduce the total amount of data on the production email servers and move older data out of users' mailboxes while still allowing the data to be retrieved. Achievement of these goals allows Exchange Server administrators to reduce the amount of data that must be stored on the Exchange Server. Less data on the Exchange Server brings about a number of advantages:

- Reduction in the amount of time necessary to perform nightly or weekly backups
- Reduction in the length of time required for nightly online maintenance
- Improvement of Exchange Server performance due to less email data
- Reduction in the amount of time for restoring data from backup

Shrinking the Exchange Server Database

If an Exchange Server database file (for example, Database1.edb) is 100GB in size, and you archive 80GB of data from that database, the database will not shrink to 20GB automatically. Exchange Server will eventually clean up all the space that archived messages were consuming, though this database reduction might take 1 day or as long as 30 days depending on how long your organization keeps deleted items. The default for Exchange Server 2007 is 14 days, but this value can be verified or changed on the Limits property page of each Exchange Server mailbox database.

You can verify the amount of free disk space by examining the Windows Application event log for event ID 1221 (see Figure 1). In this figure, you can also see that the database MBDB-HNLEX04-03 has 22962MB of free space in the database file.

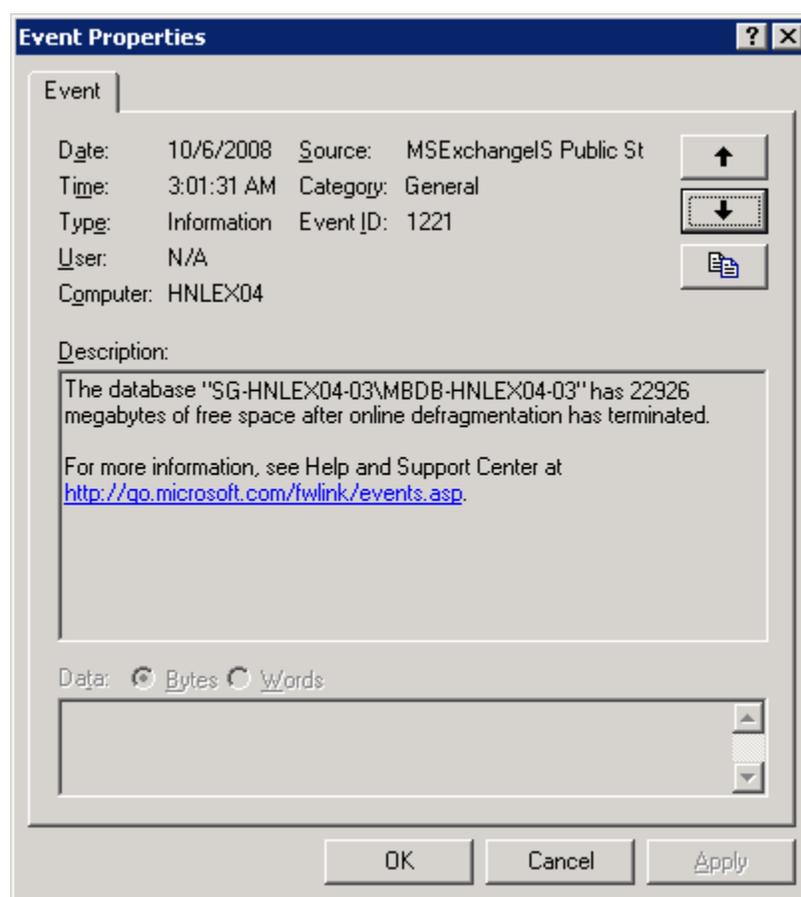


Figure 1: Viewing the amount of empty database space.

Although you cannot tell from event ID 1221, the database whose properties are shown in Figure 1 is 34GB in size; thus, almost 65 percent of the database file is actually empty space. However, the database file size will remain at 34GB indefinitely (unless all that space is reused). To actually shrink the database file, the server administrator must dismount the database and use the ESEUTIL.EXE utility with the /D switch to perform an offline defragmentation and compaction of the database file. After that process runs, the file will be closer to 12GB in size.

Taking a database offline for maintenance may not be desirable if you have to maintain a certain Service Level Agreement (SLA). Another approach is to create a new, empty database, move all the mailboxes off the database during off hours, and once the database has no active users, perform the offline compaction (or simply delete the new empty database) and move the mailboxes back. This approach will reduce the likelihood that users will be affected and ensures that the administrator does not need to work until midnight several nights in a row.

How Much Reduction?

An organization that I work with installed an email archive system with the stated goal of reducing the amount of data on their production servers. Users were demanding more and more mailbox space. For just over 1300 users, their total Exchange Server data storage was nearly 500GB plus an additional estimated 1TB of PST data.

The goal was to immediately copy archived data to the email archive so that data could be recovered from the archive and to ensure that the organization met specific legal requirements for data retention and eDiscovery requirements, but to leave mail in the users' mailboxes for 30 days before permanently removing the data from the users' mailboxes. The company's backup-to-disk software required nearly 7 hours to complete the nightly full backup, but the backup had to begin at 5:00PM—while many users were still working—in order to allow the backup to complete and allow time for online maintenance to complete.

After the archive system had been put in place and had removed all the historical data older than 30 days, the mailbox databases were compacted. The resulting Exchange Server databases' total size were 75GB. The backup-to-disk process then took just under an hour each night and ran at a later time in the evening so as not to interfere with people working late.

The SLA required a Recovery Time Objective (RTO) of 8 hours, which the organization could not meet before the archive system implementation given the volume of data that they had to maintain. After the archive system was installed, the estimated recovery time from a complete outage was reduced to just over 2 hours.

Reduce Your Recovery Time Objective

Organizations that strive to meet stated levels of service in an SLA usually also define a recovery Time Objective (RTO) in their SLA. The RTO specifies the maximum amount of time that restoration of services will take given specific types of outages. RTOs are usually defined in concert with management; factors that are considered include risk analysis of downtime as well as capabilities of a particular system for timely restoration.

It is not uncommon for Exchange Server configurations (number of servers, maximum database sizes, backup methodologies, and server hardware) to be changed because an original design would not permit the desired RTO. By reducing the amount of data that is actually on the Exchange Server systems, an email archival system can help reduce the RTO because an Exchange Server can be restored more quickly if there is less data to restore after a complete outage. Further, the archive may also contribute to the restoration of service because users may be able to employ the archive while full data services are restored or the email archival system may even allow data to be recovered from the archive back to the mail server.

Improve Your Recovery Point Objective

The RTO defines the maximum permissible time for recovery from some type of outage; the Recovery Point Objective (RPO) specifies the maximum amount of acceptable data loss that is permissible. In the past, many mechanisms have been put in place to reduce the maximum amount of data that might be lost in the event of an outage. Some of these approaches include:

- Running multiple backups each day (possibly even multiple backups per hour)
- Placing databases and transaction logs on separate storage area networks (SAN) or server hardware or using SAN-level replication
- Performing backups to disk frequently
- Installing replication solutions to keep copies of data in a recoverable location

Running backups at multiple points during a business day is the most common solution to helping to reduce the RPO. For example, if the RPO specifies that the maximum amount of data that could be lost after a significant outage of the mail servers is equivalent to what is lost over a period of 2 hours, a mail server backup would have to be scheduled for every 2 hours. However, all these solutions increase overhead. Exchange Server database backups that are performed during the day are usually noticed by the end user community due to the additional load that streaming backups, snapshots, or backups to disk place on the Exchange Server. Replication technologies involve additional hardware, possibly software, and considerable additional storage. In some instances, performing even hourly backups of Exchange Servers may require additional physical Exchange Servers to handle the backup load and the required user load. The actual recovery times specified for an Exchange Server are usually agreed upon after management and IT review the types of data being backed up, the risk associated with losing that data, the costs associated with reproducing the data, and the various costs associated with lowering or raising a proposed RPO.

Depending on the type of email archiving system and the frequency at which data is copied to the archive, the email archiving system can help reduce the RPO because data more recent than the last backup will be contained in the archive. However, email archive systems that copy data in near real-time using transaction log copies will be better adapted to disaster recovery.

However, MAPI-based systems that crawl through mailboxes one at a time will place a significant additional load on an Exchange Server. A MAPI crawl that runs frequently or during the business day will incur a noticeable performance delay on the Exchange Server.

MAPI crawls are very similar to brick-level backups of Exchange Server; a brick-level backup requires that the software open each mailbox using the same functions that a Microsoft Outlook client would use. Each folder and message is opened and backed up one at a time. Brick-level backups are not a desirable approach to Exchange Server backups because of the time needed to perform the backup and the load that this places on the Exchange Server.

Archive systems that perform MAPI crawls through mailboxes suffer these problems. In addition, MAPI-level crawls may not be able to back up all the email metadata—data that might be required to be archived by internal policy or legal regulations.

Dial-Tone Recovery

When faced with rebuilding an Exchange Server from a wholesale server failure, the email administrator often has to make a decision to restore email services immediately and then restore the data as soon as is practical. This *dial-tone restore* recovers a minimum level of functionality, such as the ability to send and receive email without having the older email messages, calendar, and contact information.

Having an email archive system in place can make a dial-tone restore even more palatable for the end user community because they can immediately begin to send and receive messages and can access their older emails in the archive. Depending on how recent the data is in the archive (such as an archive that is built using transaction logs in near real-time), users' entire mailboxes, calendars, and contacts may be immediately available in the archive.

What Is the Right Approach?

As you evaluate email archiving systems, consider features that may be useful during a serious outage or with disaster recovery. Factor into your evaluation the type of load the archive system will place on your existing Exchange Server infrastructure when they system copies data into the archive and how frequently the archive system needs to be run to keep the archive updated.

In order for an archive system to be considered valuable in disaster recovery scenarios, the archive must not only be up to date but also contain all parts of a user's mailbox. Some archive systems might not archive important information within the mailbox, such as calendar items, contacts, tasks, personal journals, and mail metadata. Mail metadata includes information such as whether the message was read, forwarded, replied to, flagged, or categorized. Archive systems that move data to the archive using log shipping technologies rather than MAPI crawls provide better performance and a complete image of a user's data because the data is moved over at the transaction layer rather than "after the fact" using a MAPI crawl. In the event of a server failure, an effective archive system provides built-in disaster recovery features that enable an administrator to restore data from the archive quickly and efficiently.

Reducing the amount of data on the production Exchange Servers will help reduce your RTO. However, carefully weigh the benefits of immediately (or very quickly) copying data to the archive, as the performance of different systems varies greatly as does the effect the copy process will have on the Exchange Servers.

As email has taken over as one of the most prolific applications in business use, the requirement to provide higher availability and improved disaster recovery solutions has grown as well. A mail server outage cannot be tolerated. Email archival systems aid data recoverability and disaster recovery to address the mission-critical need for business access to email.