

Backup and Recovery for Microsoft Exchange Server

Data Domain Deduplication Storage Best Practices Guide

Abstract

Protecting Microsoft Exchange data is now more critical than ever. IT administrators need to exploit modern disk backup technologies to effectively deal with the data growth, retention requirements and recovery service levels that are essential to businesses. Data Domain's industry-leading deduplication technology provides a powerful backup, archiving and DR solution that can scale with the most demanding data center requirements. This technical brief explores the various practices and considerations for backing up Microsoft Exchange data to Data Domain systems and how to effectively exploit this technology.

Backup and Recovery for Microsoft Exchange Server

Data Domain Deduplication Storage Systems Best Practices Guide

Table of Contents

1 INTRODUCTION	3
1.1 BACKGROUND – PURPOSE AND SCOPE	3
1.2 INTENDED AUDIENCE	3
1.3 CAVEATS	3
2 BEST PRACTICES RECOMMENDATIONS – OVERVIEW.	4
3 DATA DOMAIN TECHNOLOGY OVERVIEW	5
4 EXCHANGE SERVER ARCHITECTURE OVERVIEW.	5
4.1 DATABASE STRUCTURE	5
4.2 EXCHANGE 2007 FEATURES	6
4.2.1 VSS	6
4.2.2 DPM AND FUTURES	7
4.3 PLANNING THE EXCHANGE DATA PROTECTION	7
4.3.1 DATA PROTECTION STRATEGIES	7
4.3.2 EXCHANGE SERVER DATA	8
4.3.3 DELETED ITEM RETENTION	8
4.3.4 CLIENT-SIDE DATA	9
4.4 ADMINISTRATIVE TASKS	10
4.4.1 DEFRAGMENTATION	10
4.5 E-MAIL ARCHIVING	11
5 USING BACKUP APPLICATIONS	12
6 SUMMARY	12
APPENDIX A. DOCUMENTS AND LINKS.	13

List of Tables

TABLE 1 – DAILY ADMINISTRATIVE TASKS	10
--	----

List of Figures

FIGURE 1 – EXCHANGE DATABASE ARCHITECTURE.	5
FIGURE 2 – E-MAIL WITH ARCHIVING STORAGE ARCHITECTURE	11

1 Introduction

In many companies, Exchange is now commonly viewed as a mission-critical application. Its use within an organization has grown from a simple e-mail tool to include contact management, scheduling and file sharing. If the corporate messaging system is not available, productivity goes down, and business opportunities can be lost. Even if e-mail is not critical to your business, chances are that the loss of messaging services, even for short periods would create a substantial disruption to your organization. Loss of the data maintained in Exchange is unacceptable.

The ability to recover all or part of your Exchange data when required is essential. The 24*7 uptime expectation of Exchange messaging systems in enterprise environments is now the norm. With global operations and advanced telecommunications (e.g., smart phones, etc.), individuals and corporations are communicating around the clock. This leaves administrators with the challenge of tighter data protection and system maintenance windows.

Protecting Exchange information is critical and backup is still the primary method used. Combining traditional backup with other methods such as clustering, replication and continuous data protection provide for a robust Exchange protection strategy. This paper focuses on the traditional backup and recovery practices for Exchange along with solutions from Data Domain.

Advances in storage technology have made the relative performance, reliability and ease of use of disk storage systems a popular method of data protection for many data centers. There are many ways to use disk for data protection. This paper focuses on the use of disk-based backup solutions for Exchange. These disk-based solutions can be implemented as either file-based NAS (CIFS or NFS) or virtual tape library (VTL) backup targets or application-specific interfaces such as Symantec's OpenStorage initiative. For a longer term level of compliance, data protection or disaster tolerance such as multi-year retention, physical tape may still play an integral part of any backup solution. When used in conjunction with disk backup targets, tape provides for a complete and robust disaster recovery solution.

1.1 Background - Purpose and Scope

When it comes to backup, Data Domain storage systems provide unique and powerful disk-based solutions with high-speed inline data deduplication functionality that is perfect for simplifying backup and recovery solutions. The WAN-optimized replication capabilities of Data Domain storage systems provide an effective off-site alternative for DR scenarios as well. This paper focuses on backup and recovery of Microsoft Exchange systems and how to best leverage the capabilities of the Data Domain deduplication systems. Microsoft Exchange practices will be explored and mapped onto the specifics of the Data Domain solution as applicable.

1.2 Intended Audience

This paper is intended for solution architects and storage administrators involved in the planning and deployment of data protection solutions for Microsoft Exchange Server (2003 & 2007) environments using disk based or virtual tape technology. A familiarity with the Exchange architecture and basic backup and recovery practices is assumed.

1.3 Caveats

This paper focuses on the backup and recovery practices of Exchange environments utilizing the Data Domain solutions. It will not go into detail on Exchange primary storage layout unless it affects the practices described for the backup/recovery discussions.

There are many new features in Exchange 2007 that deal with availability and recovery mechanisms such as clustering and replication configurations. These will only be addressed as they relate to or affect the typical backup and recovery processes that most administrators face on a daily basis.

The information in this paper has been collected from several sources including Microsoft, several 3rd party backup solutions and other Data Domain technical papers. For the most current information please refer to your vendor's specific solution or the links provided at the end of this document.

2 Best Practices Recommendations

The following is an overview of recommended best practices for a Microsoft Exchange backup and recovery with Data Domain. These specific recommendations will be covered in detail throughout the remainder of this paper.

Perform full backups whenever possible.

Size the Data Domain storage system for enough capacity and backup throughput to meet backup window and retention requirements for performing daily full backups.

Use a dedicated Exchange-aware backup application on a separate server to manage the backup operations.

Use the backup application's Exchange API to get best data streaming results from the Exchange server to the backup server system or use the Exchange VSS capability to get the best volume based backup image.

Where possible, break down backup jobs first by 'Storage Groups' and then by databases within a Storage Group.

Direct backups from a Storage Group or database to the same Data Domain system if multiple systems are deployed. Databases within a Storage Group should be backed up together to ensure the associated log file truncation is properly handled.

Use multiple streams to the Data Domain systems to improve overall aggregate backup throughput.

Limit the maximum number of concurrent backup jobs (streams) to approximately 10 per Data Domain storage system. The actual number depends on the structure of the Exchange configuration and the Data Domain system used – some smaller systems may be slightly lower, some larger systems may be slightly higher.

Increase the 'Deleted Item Retention' levels to allow for more recovery directly from the Exchange online databases.

Analyze and plan the primary and backup storage accordingly for the potential increase in retained data.

Limit the amount of manual and automatic administrative actions (e.g., defragmentation, auto archiving, etc) that can regularly rearrange the internal Exchange databases.

This disrupts potential deduplication benefits as data sets are internally reorganized.

Do not use any form of compression settings within Exchange or the backup software application for data being sent to the Data Domain system.

Do not use any form of encryption settings within Exchange or the backup software application for data being sent to the Data Domain system.

Do not configure any form of multiplexing or stream interleaving.

Use more streams to disk targets. If VTL is used then allocate more virtual tape devices.

Do not perform any read back verification from the backup application during the backup operations.

Backup set verification can be performed after the backups have completed if necessary.

Tune the backup server buffer settings based on specific backup application recommendations and current Exchange environment resources.

Backup individual mailboxes only where you already have a full database backup of the Exchange server; the priority or service level agreement (SLA) requires it and the backup application supports it.

Use a separate policy to manage this. If possible, use backup application capabilities for mailbox level (granular) restores from full database backup images instead.

Selectively backup .pst files to the same Data Domain storage system as the storage groups and databases.

Do not use compression or encryption settings on .pst files that are candidates for inclusion in the client side file system backups.

Configure separate backup directory (or tape pool) structures on the Data Domain system for Exchange backups.

This will help with space / data reduction management and replication control. If desired, further subdivide the backup targets by folder type, storage group, database, or mailboxes.

3 Data Domain Technology Overview

A Data Domain Deduplication Storage System is a storage appliance typically used as a target for backup or archived data. The characteristics that make it good for this include:

- Support for most conventional backup applications through multiple protocols for Network Attached Storage (NAS) interfaces over Ethernet, a VTL interface option over Fiber Channel, and product-specific interfaces such as Symantec NetBackup OpenStorage (OST);
- High-speed, inline deduplication using small, variable-sized sequences to identify and eliminate redundant data;
- Integrated data protection technologies such as RAID6, post-backup data verification, and periodic validation checks of existing data sets;
- Deduplicated replication over Ethernet to secondary systems to automate DR.

Data Domain deduplication storage systems come in a range of sizes and performance levels able to target almost any Exchange backup configuration. Usable backup data set sizes start at just over 300GB and scale up to over 30TB capacity. Throughput rates can range from 50MB/sec to over 300MB/s per appliance depending on model and configuration. With data deduplication rates that can range from 10x to 30x, these systems are well suited to maintain multiple weeks of Exchange backups in an online disk based solution.

Actual usage scenarios may vary so it is important to work with your local Data Domain team to understand the details.

4 Microsoft Exchange Server Architecture Overview

The Microsoft Exchange Server solution is a messaging system that contains an underlying transactional database. Exchange uses the Extensible Storage Engine (ESE) to maintain these transaction-based databases. It uses write-ahead transaction log files to ensure that Exchange data is efficiently processed.

There are now two major versions of Microsoft Exchange Server (2003 and 2007) operational in most data centers today. Though they may still exist in some environments, this paper will not address details about Exchange 5.5 or Exchange 2000 Server versions. Both of these newer versions (2003 & 2007) of Exchange have many similarities with respect to basic data protection strategies – especially as it pertains to the fundamental backup and restore practices. They will be treated as a single solution for most of this discussion and differences will be pointed out specifically for 2003 or 2007 as needed.

4.1. Database Structure

Exchange Server supports the use of two types of database stores: ‘mailbox stores’ and ‘public folder stores.’ You can split the overall Exchange server ‘information store’ into multiple storage groups and database stores. A storage group is a group of databases that share a single transaction log. A single database contains the mailbox or public folder contents. Figure 1 depicts the basic Exchange information architecture.

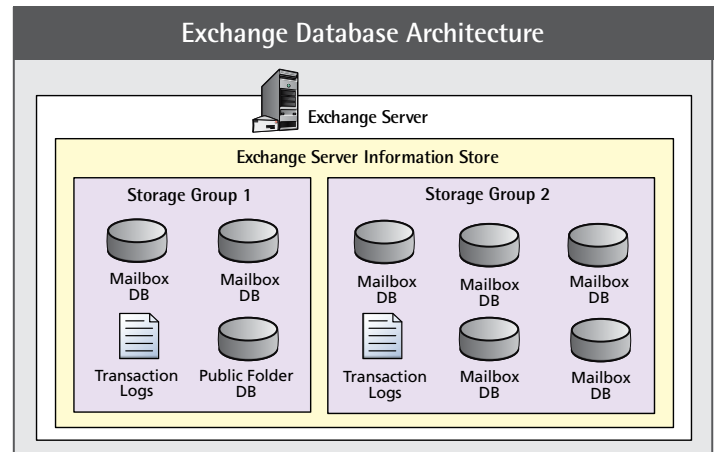


Figure 1. Exchange Database Architecture

Exchange Server 2003 has the following storage group and database limits:

- Standard Edition: 1 storage group, 1 database per storage group
- Enterprise Edition: 4 storage groups, 5 databases per storage group
- Log files are 5MB each

Exchange Server 2007 has the following storage group and database limits:

- Standard Edition: 5 storage groups, 5 databases per storage group
- Enterprise Edition: 50 storage groups, 50 databases per storage group
- Log files are 1MB each – reduced in size to facilitate replication configurations

The primary reason for deploying multiple storage groups and databases is to reduce the size of each individual database while still supporting many users on one server. Having multiple smaller databases enhances Exchange Server backup and recovery. Distributing users across a larger number of mailbox databases can lessen the impact of the loss of a single database, and allow for quicker restores when needed.

Because all of the databases in a storage group share a transaction log, each storage group should be backed up as a whole using the

same backup policy or operation. An Exchange-aware backup application is the best way to manage this. If your backup infrastructure supports multiple backup streams, you can backup multiple databases or storage groups at the same time. As long as you do not exceed the disk I/O throughput of your servers, controllers, and backup hardware, you can save time by simultaneously running multiple instances of backups or restores. The Data Domain system can easily handle multiple backup input streams and deduplicate the data inline simultaneously across the incoming streams.

Microsoft Exchange Server databases can be backed up while users are online, even as new data is written to them. This capability exists because of Exchange's transaction logging mechanism. As you begin an online backup using the Exchange API, the backup program streams the database file to the backup medium. Changes to the database continue, even to parts of the database that have already been backed up. These missed changes will later be reconstructed from transaction log files.

Another advantage of the online backup API is that it simplifies selecting databases for backup or restore operations. Rather than having to know which Exchange files require backup for a particular database, the administrator selects the appropriate server and databases to be backed up, and Exchange automatically collects and backs up the appropriate files. Transaction logs will typically be sent to the same backup set as the databases.

After the database file has been backed up, Exchange copies at least one transaction log (and usually several of them) to the backup set. These are the transaction logs generated from the time the backup starts until just after it finishes. Backing up the Exchange during periods of low activity will limit the number of associated log files and make recovery simpler. One nuance of taking a backup of a storage group is that an Exchange-aware backup program will modify the header of the database, typically adding information about the time of the last backup of the database.

In very large environments, if multiple Data Domain systems are being deployed, make sure the backups from a particular Storage Group always go to the same target Data Domain system. This will improve the data deduplication results for that configuration.

When you use the backup application to restore Exchange databases, API calls are made to the Exchange Extensible Storage Engine (ESE) to restore Exchange database files and their associated log files. You can use Exchange database backups to restore one or more damaged mailbox stores or public folder stores. You can also use Exchange database backups to restore every mailbox and public folder on the server.

Individual databases in a storage group may be restored while all other databases remain online. This method is the preferred means of replacing a single failed database. When the database is remounted, pertinent transactions are automatically played back from the storage group's log files to bring the restored database back to the time of the disaster.

4.2 Exchange 2007 Features

A new capability - continuous replication - is available in Exchange 2007 through two features called Local Continuous Replication (LCR) and Cluster Continuous Replication (CCR). These features use built-in asynchronous replication technology to create a copy of a storage group and keep it up-to-date using log shipping and replay. The replication facilitates this by applying the log files of a production database to a copy of that database which runs on the same server for LCR or separate passive/standby server for CCR.

In Exchange 2007, backups can be run against the copy of the production database on either the local server (LCR) or the alternate server (CCR) node, decreasing the performance impact on the production Exchange environment. One of the other benefits of using LCR or CCR is the ability to offload VSS-based backups from the active production storage groups to the alternate storage groups. In architectures intended to leverage VSS-based backups (see VSS below), the primary storage volumes (LUNs) are typically "snapped" and presented to another server for data access. With LCR and CCR modes of Exchange, the alternate server can be the passive system of the replication pair or a separate backup server.

Note: You cannot back up a target storage group in a Standby Continuous Replication (SCR) environment. Backups of storage group copies are available for LCR and CCR environments only.

4.2.1 VSS

Windows Server 2003 provides a storage management function called Volume Shadow Copy Service (VSS) which provides a consistent point in time (PIT) reference of the file system. From this consistent PIT reference, the storage subsystem or backup application can access or make a copy of the live Windows data.

Microsoft also provides a VSS API (Volume Shadow Service Application Program Interface) specifically for backup applications to hook into these Exchange databases. There is also an Exchange streaming API that can be used to access the Exchange database files independent of the VSS PIT data.

If the Exchange Server is being protected with a storage based copy solution (e.g., SAN snapshots) utilizing VSS functions, then it is recommended that the storage volume be backed up in its entirety (e.g., full volume level image). Methods using SAN storage backup techniques are often more complicated than the backup application's API-based service. For this reason, using a 3rd party application to backup the Exchange databases through more traditional methods is recommended.

Exchange-aware VSS backups are supported for both the active and passive storage groups and databases. The passive copy backup support is only for VSS, and it is implemented by the Exchange Replica VSS Writer that is part of the Microsoft Exchange Replication service. Streaming backups are only supported from active storage groups. You cannot use streaming backup APIs to back up the database on passive storage groups.

To perform a VSS-based backup of a passive storage group, you must use a third-party backup application that supports Exchange VSS.

4.2.2 DPM and Futures

Microsoft has recently introduced new approaches to protecting critical Windows system data using DPM (Data Protection Manager) and Continuous Replication (CCR, LCR, SCR) services. This paper will only focus on the current mainstream backup practices within industry. As DPM or Continuous Replication becomes more prevalent, best practices for using Data Domain systems to address this or future needs will be explored.

4.3 Planning Exchange Data Protection

The first thing to plan is the backup application. Using an Exchange-aware backup application to back up the Exchange database on a regular basis is recommended. Follow the specific guidelines for setting up that application's integration with Exchange and any special use of the Data Domain system.

The next thing to determine is what types of failures you want to protect against. Many failures and disasters may require that you repair or restore one or more parts of your messaging system. It is important that you have a strategy in place to recover from the following situations:

- Lost mail item (permanently deleted mail)
- Lost mailbox
- Lost database or Storage Group
- Lost server that is running Exchange (Exchange databases and transaction log files intact)
- Lost server that is running Exchange (Exchange database and transaction log files also lost)

The last thing to determine is what is going to be protected and where that data is located. The next sections examine this detail.

4.3.1. Data Protection Strategies

The recovery strategy you select also influences your backup strategy. Consider implementing one or more of the following strategies to help you be prepared to recover the desired items:

Recovery Storage Group

With the recovery storage group feature in Exchange 2007, you can mount a second copy of an Exchange mailbox database on the same server as the original database or on any other server that is running Exchange in the same Exchange administrative group. This action can be performed while the original database is still running and serving clients. With this capability, you can recover data from an older backup copy of the database without disturbing user access to current data. The recovery storage group can also be useful in various disaster recovery scenarios, most notably the messaging dial tone scenario. You can only use the recovery storage group

to recover mailbox stores, and not to recover public folder stores. Performing full backups at the Storage Group level make using this recovery method much simpler and effective.

Third-party Brick-level Backups

Some third-party backup tools let you back up and restore Exchange at the level of individual mailboxes (sometimes called brick level). This can take considerable server and storage resources and time, as each mailbox must be processed separately. In general, performing mailbox level backups as the only method of protection is not recommended. There are also indications that there is less than ideal deduplication from the mailbox only backups.

If you are using a Data Domain storage system, the extra logical storage required to capture both full database backups as well as some (critical user) mailbox level backups will be reduced due to the deduplication of the data in the mailboxes with the content of the full database backups. Administrators may choose both processes with the idea that database backups protect against system failures (i.e., restore the Exchange Server) and mailbox backups protect against user failures (i.e., single user deletes items). However, many 3rd party Exchange aware backup applications are beginning to offer the ability to perform mailbox level (granular) restore from a full database backup image. In either case, where possible and based on performance and space requirements, if a combination of database and mailbox level backups are going to be performed, target the subsets (e.g., the database and the mailboxes in that database) to the same Data Domain system. This will allow the sharing of deduplication segment information for the common content within each database context.

As part of your recovery strategy, determine the time required to restore and replay transaction log files. Performance in your environment may vary significantly from the average, and you should consider log replay in addition to restore time. If you perform weekly full backups and daily incremental backups, you may have hundreds of transaction log files that require replay after a restoration. More frequent full backups may require more processing on a daily basis but can make recovery much easier and faster without the need to coordinate incremental data restores. The extra backup storage and time required by a full backup is mitigated by the high speed inline data deduplication capabilities of the Data Domain storage system as the target backup device.

To minimize the time that it takes to recover an individual database, configure storage limits for the mailbox and public folder stores to constrain databases to a maximum size limit. The sizes will directly affect your backup and restore SLAs. There are many factors to consider when designing the layout and sizes of your Exchange configuration. The details of such design considerations are outside the scope of this paper.

4.3.2 Exchange Server Data

The Exchange database files, including both mailbox and public folder databases and the Exchange transaction log files that are specific to each storage group should be backed up and restored with an Exchange-aware backup application. The database is stored as the following files:

Extensible Storage Engine (.edb file)

The Extensible Storage Engine (ESE), formerly referred to as the Joint Engine Technology (JET) database, stores all data submitted by Exchange clients. One .edb file is associated with each database.

For mailbox stores, the .edb files will mostly contain the message content, metadata and any attachments included with each message. It is not uncommon for the attachments to a message to comprise the bulk of the storage requirements for an Exchange environment. Deduplication of data in attachments is a key benefit of the Data Domain technology. The message body is usually smaller than the segment size used (4-12K variable size segments) and affects overall deduplication rates based on the messaging use case for each environment. You can typically expect greater data deduplication effects for Exchange environments that contain a significant amount of attachments in the messaging traffic.

Depending on the size of the Exchange environment, one or more databases or storage groups may exist. Although these backups can be directed to the same storage device (as defined by the backup application), it is recommended for the backup administrator to partition the Data Domain system into subject specific subdirectories if possible. The data deduplication process applies globally to all directories on a single Data Domain system, but it is possible to view and report specific data set compression characteristics if the structure is divided as such. For example, backups of exchange can be directed to `/backup/exchange/storage-group1` or `/backup/exchange/storage-group2`. For information on the specifics of how to accomplish this, contact your Data Domain team for more information.

Transaction Log Files (.log files)

All changes made to the database are first committed to transaction log files. Any time a user modifies data stored in a mailbox or data is added to the mailbox, that change is written to a transaction log file before it is written to the database. The change is immediately committed to the in-RAM database cache and then copied back to disk when the system's load permits. Transaction log files are created sequentially. This sequence is referred to as the log stream. The transaction log files are relatively small (e.g., 1 MB for Exchange 2007 and 5 MB for Exchange 2003). The number of transaction log files created depends on the client load on the server.

After transaction log files are committed, they are then protected by backups. After a successful backup, Exchange deletes (truncates) transaction log files from the file system. If regular backups are not performed or backups fail, transaction

log files will accumulate on the file system. If transaction log files are not committed and deleted, they consume space on your hard disk. If your hard disk fills, Exchange dismounts your database and stops accepting data until you make more space available on the hard disk.

Rather than write all log information to a single large file, Exchange uses a series of log files. When a log file is full, Exchange closes it and renames it with a sequence number. The first log filled ends with the name `Enn00001.log`. The `nn` refers to a two-digit number known as the base name or log prefix.

An Exchange server can have multiple storage groups, and the log files for each storage group are distinguished by file names with numbered prefixes (for example, `E00`, `E01`, `E02`, or `E03`). The log file currently open for a storage group is simply named `Enn.log`—it does not have a sequence number until it has been filled and closed.

The checkpoint file (`Enn.chk`) tracks how far Exchange has progressed in writing logged information to the database files. On a busy database, the checkpoint typically lags three or four log files behind the currently open log. There is a checkpoint file for each log stream and a separate log stream for each storage group. Within a single storage group, all the databases share a single log stream. Thus, a single log file often contains operations for multiple databases.

Some other miscellaneous server side data items to consider are:

- The Exchange Search information that is specific to each mailbox database in a storage group has no defined backup method – this information needs to be rebuilt rather than restored.
- The Offline Address Book (OAB) is typically protected through normal file system backup and restore. It can also be rebuilt if needed. Protection through Public Folders and replication is also a consideration.
- The Windows registry (i.e., `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Exchange HKLM\SYSTEM\currentcontrolset\Services`) is protected through the backup / restore of the System State, or the export of the System State and file level backup.

4.3.3 Deleted Item Retention

Unless primary storage space for the Exchange databases is a critical issue, choosing less aggressive deleted item retention policies will provide greater recoverability of data from directly within the Exchange Server environment. Rigorous pruning of this content becomes less of an issue when combined with the advantages of data deduplication for the backup storage. The extra deleted item content should not adversely affect overall capacity requirements of the backup storage solution if Data Domain deduplication storage is being used.

When a user deletes an item, it appears deleted to the user. However, a copy of the deleted item is kept in the user's mailbox database for a specified time, which allows the item to be recovered if it was deleted unintentionally.

When the Exchange database receives a request to delete a message, it determines if the message should be soft deleted or hard deleted. Soft deletion is also referred to as logical deletion, and hard deletion is also referred to as physical deletion.

When a message is hard deleted, the message reference is immediately removed from the `MsgFolder` table. During the next background cleanup process, the entries in the Deleted Messages table are examined and the corresponding entries in the messages table are deleted. This process occurs every hour by default. However, you can control this schedule by editing the following registry entries:

- `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\MSExchange\Parameters\Public\Background Cleanup` (value in milliseconds)
- `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\MSExchange\Parameters\Private\Background Cleanup` (value in milliseconds)

If you enable and increase the deleted item retention time, you may need to perform additional capacity planning for the primary storage. If you can support the additional workload, running background cleanup on a daily basis will allow more deleted item information to be backed up to the Data Domain system before being permanently removed.

A soft deletion is performed if none of the criteria for a hard deletion are met. A flag is set on the entry in the `MsgFolder` table that indicates that the message has been soft deleted from the folder. At this point, the message is available for deleted item recovery. During the next scheduled Exchange database maintenance process, each folder is examined to determine if any of the soft-deleted messages that the folder contains have passed the deleted item retention time. If such a message is found, the message is hard deleted.

By default, deleted items are stored in an Exchange database for a certain number of days before they are permanently deleted by Exchange. You can set the length of the deleted item retention period by either using the database defaults, or by selecting the number of days that a deleted item is kept before it is permanently deleted.

It is recommended that you configure this setting to 14 days or more. If the deleted item retention period is set to 0, the deleted items are permanently removed from the server immediately. Unless disk space is an issue, we recommend that you do not disable the deleted item retention feature.

Deleted item retention can be configured on a per-database and a per-user basis. Individual user settings override database settings.

By default, Microsoft Outlook enables deleted item recovery from the Deleted Items folder only.

In Exchange 2007, deleting a mailbox does not mean that it is permanently deleted (or purged) from the Exchange mailbox database immediately. Instead, the mailbox is flagged for deletion, and users cannot access it. At the end of the mailbox retention period, the mailbox is permanently deleted from the database. You can also permanently delete the mailbox by choosing to purge it at any time. If you mistakenly delete a mail-enabled user account, you can re-create that user object, and then reconnect that mailbox during the mailbox retention period.

By default, the deleted mailbox retention period is 30 days. You can configure the deleted mailbox retention period at the mailbox database level. If the mailbox retention period is set to 0, deleted mailboxes are permanently removed from the server immediately. Unless disk space is an issue, we recommend that you do not disable the deleted mailbox retention feature.

4.3.4 Client-Side Data

Most of the critical corporate information that gets backed up is on the Exchange server – at least at some point in its lifecycle. In some situations, it is useful to also consider protecting data that is found on the client side (Outlook) of the Exchange environment.

Outlook .pst Files

The .pst file is a storage file for Outlook information that does not require an Exchange server. This is where Microsoft Outlook stores locally saved end user data. It can be backed up and restored using traditional file system backup functionality. Unlike the .ost file, the .pst file is not synchronized with the server, and it is not associated with a mailbox. A .pst file can be accessed with a MAPI application. A .pst file can be used in the following ways:

- By Outlook to store mailbox data.
- As a common e-mail data exchange format. Many commercial e-mail clients are able to import from and export to a .pst file.

Losing a .pst file frequently means losing data because the .pst file might contain the only copy of that data. This is true for message traffic that passed through the Exchange server to the .pst file and was completely deleted during the interval between backup jobs. The .pst file is updated whenever new data is stored in this file. An incremental file system backup of this file will essentially be the same as a full backup even though only a small percentage of data has actually been added to this file since the last backup.

Protecting user's .pst files has always been problematic for several reasons:

- They typically reside on the end user's system which must be online and part of the backup solution (e.g., backup client) in order to be backed up.

- They tend to become relatively large over time as users file away message content that they no longer want or are not allowed to keep in their active Outlook/Exchange folders
- They do not fall into a convenient full/incremental backup strategy since a small change updates the whole .pst file. This makes backup data set sizes balloon over time as well.

With a backup storage platform like Data Domain that provides deduplication of recurring data segments, storing .pst files becomes much more viable, since less total storage will be required for backup data. Data deduplication is compounded by the fact that the message data and more importantly, attachments in the .pst files were likely also in the core Exchange environment (depending on rules and policies) and backed up there through normal procedures.

New rules can be established within an organization for protecting centralized .pst storage or scheduled access and protection by the backup application at the file system level. Another alternative to better manage .pst data discussed later is to use e-mail archiving to help solve some of the .pst file challenges.

Outlook .ost Files

The .ost file is an offline copy of all the end-user data that exists on a user's mailbox on an Exchange server. This is where Outlook stores locally saved end user data. It can be backed up and restored using traditional file system backup functionality.

The .ost file is created when Outlook is running in cached mode and associated with a specific mailbox. The .ost file can be used for the following:

- A mobile user can use an .ost file to work with an offline copy of e-mail while disconnected from the network.
- An .ost file is used by Office Outlook 2007 and Outlook 2003 to reduce the number of MAPI calls from the client to a server that is running Exchange.

The .ost file is a synchronized copy of that data on the server that is running Exchange. Losing the .ost file will only result in a loss of any changes (for example, in Drafts or Outbox) that have not been synchronized with the server. Remote users are at the most risk of losing data in this manner because they can be offline for extended periods of time. The .ost file can be re-created by synchronizing all the data on the server that is running Exchange.

This type of backup requires backup application access to the end user (client) system. If the .ost file is lost or damaged, the information can be recovered through resynchronization of the end-user's system with the Exchange Server. The .ost files can be just as easily resynched with the Exchange server as they can be backed up by the backup application.

4.4 Administrative Tasks

Exchange administrators can choose to implement certain house-keeping options within an Exchange configuration. The following section will address some of the ones which could affect the backup and recovery solutions recommended.

On a daily basis, there are some Store Maintenance tasks (see Table 1) that are performed to keep the underlying database consistent and well tuned. The administrator can set a window each day (default is 12-5 AM) for this processing to occur. If the list is not completed, it is continued during the next day's window. The background tasks can be monitored through 'Event Logging Level' and application logs. It is recommended to let this task list complete every 1-2 weeks. More frequent administration may not yield significant benefits. Note that the processing stops when a backup operation is running.

Task	Mailbox Store	Public Store
Purges indexes	Yes	Yes
Performs tombstone maintenance	Yes	Yes
Purges deleted Items cache (Dumpster cleanup)	Yes	Yes
Purges outdated public folder content	No	Yes
Purges expired tombstones in public folders	No	Yes
Resolves public folder conflict aging	No	Yes
Updates server versions	No	Yes
Cleanup secure folders	No	Yes
Purges deleted mailboxes	Yes	No
Checks for obsolete messages	Yes	Yes
Defragments store	Yes	Yes

Table 1. Daily Administrative Tasks

4.4.1 Defragmentation

The online daily defragmentation of the Exchange database is part of the database maintenance process. The task is controlled by the maintenance interval defined for the individual database. This can also be controlled for a collection of mailbox stores through database policy settings. The online defragmentation process detects and removes database objects that are no longer being used. This process provides more usable space but does not change the size of the database file. The daily defragmentation of the Exchange databases can potentially have some subtle effects on the data deduplication levels that can be obtained. As objects are removed and data reorganized, the existence and location of redundant segments will also decline to some degree.

Although not tied directly to defragmentation, the nature of use and actual content of the active Exchange environment can also play a major role in the overall deduplication rates that can be achieved. For messaging environments where a predominant amount of the Exchange data is lightweight and short-lived (e.g., instant messaging, scheduling, task coordination, etc.), the negative effects of regular defragmentation (and database reorganization) on deduplication will yield lower compression rates, typically in the 6-10x ranges. However, in messaging environments which use Exchange to share documents and users maintain higher retention practices (“e-mail filing”), the effects of defragmentation are far less and the overall deduplication rates can increase to a range of 15-20x or greater.

Outside of automatic online processes, the database can also be manually defragmented with the ESEUTIL utility from Microsoft. Additional manual defragmentation is only recommended when the database configuration undergoes significant external restructuring (e.g., moving a large number of users/mailboxes).

4.5 E-mail Archiving

E-mail archiving is becoming a serious consideration for many Exchange environments. Using an e-mail archiving solution helps IT departments deal with operational efficiencies of ever growing configurations as well as records retention and compliance support for corporate and legal purposes.

Today’s e-mail archiving applications perform a fairly straightforward process of grooming the message and attachment information out of the active Exchange Information Store and managing it separately in an archival repository with links to the archive from the original Exchange mailbox/message items. The specifics of the e-mail archiving implementations may differ, but the contents of the archive are driven from the messaging traffic that originally flows into the Exchange server.

For environments considering an e-mail archiving addition to their Exchange solution, the Data Domain storage solution provides

additional benefits when used in conjunction with regular Exchange server backup operations. As long as the archiving application is not permuting the data sets from their original format and content (e.g., indexing, compressing or encrypting), then the information extracted from the Exchange databases into the e-mail archiving repository should have significant amounts of data segment redundancy with that encountered during backup operations. Figure 2 below depicts a configuration with the backup files and the archive repository residing on the same Data Domain system. It is recommended to use separate directories on the Data Domain system for backup and archiving. The data deduplication will be managed globally across the system but the two data sets can then be managed and measured independently.

E-mail archiving solutions may also be used to eliminate or reduce the amount of data in user’s .pst files by incorporating them into an overall e-mail archiving solution.

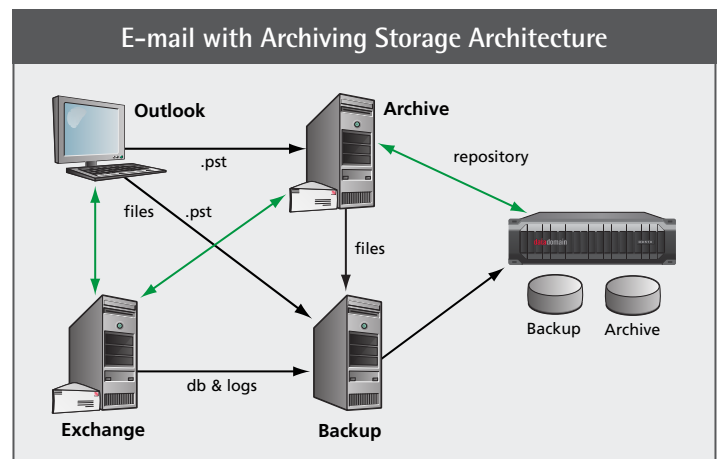


Figure 2. E-mail Backup with E-mail Archiving Architecture

5 Using Backup Applications

You can back up and restore the following items on an Exchange server:

- **Entire storage groups.** Storage groups, including all the log files and database files.
- **Exchange mailbox databases.** One database or groups of databases.
- **Exchange mailboxes.** One or more critical mailbox level sets.
- **Exchange public folder databases.** The public folder database on any server.

There are four kinds of online backups of an Exchange database: normal, copy, incremental, and differential. You may be familiar with these terms; however, the meaning that Exchange assigns to each of them differs from conventional usage:

Normal

The backup program backs up the database files (.edb and .stm), patch data, and at least one log file. After the backup is complete, the backup application deletes all log files prior to the checkpoint at the time that the backup began; this prevents log files from accumulating until they use up all available drive space. This is the preferred method. Whenever possible perform full backups. This yields a recovery option of a single image operation. It also allows for the purge of transaction logs for the Exchange environment. As long as the backup window permits a full backup, using the Data Domain system will alleviate the capacity challenges of having extra space for several full copies of the Exchange databases through deduplication of redundant data.

Copy

The same as a normal backup, except that the backup application does not delete old log files, and does not update the database header to indicate that a backup has taken place.

Incremental

The backup program only backs up log files, not database files. Log files since the last normal backup are copied to the backup medium and are then purged from disk. To restore an incremental backup, you must also restore an earlier normal or copy backup in order to get the database files. Transaction logs from the incremental backup can be replayed after the logs from the full backup are replayed, assuming that there is an unbroken sequence of logs between the two backups and that all logs were successfully recovered from backup. There are usually two types of incremental backups – one clears the logs and one does not. If at all possible use full (normal) backups over incremental backups. Transaction logs files which contain data which is predominantly new or changed, will not deduplicate nearly as well as full database copies. The multistep restore process can also be problematic.

Differential

The same as an incremental backup, except that the backup program does not delete old log files from disk.

In terms of the files actually placed on the backup medium, there is no difference between a normal and copy backup, and no difference between an incremental and differential backup.

Do not perform backup image validation during the backup process. The main reason for validating backups is to validate the media. This is essentially eliminated with disk-based systems that continually check themselves, such as Data Domain deduplication storage.

6 Summary

Proper protection of Microsoft Exchange data through regular backup practices is critical for all organizations today. In choosing to use disk-based technologies in the backup environment, Exchange administrators can realize the benefits of performance and reliability for both backup and restore functions.

For environments where the Exchange users rely on the messaging system to regularly share content (attachments) and organize their information (public folders) it is not unusual to accumulate growing amounts of data – often with a high degree of redundancy. Some of the redundant data is handled within Exchange through built in single instance storage techniques. However, this does not help if mail attachments are saved first or copy and pasted into new message threads. The ability to exploit data deduplication technology on the backup data allows organizations to keep more data for longer periods of time without incurring the linear cost of storage associated with standard disk or tape solutions that do not deduplicate the data.

In an organization, Exchange backups do not happen in a vacuum. When the Exchange backup sets are sent to the same deduplication storage systems that are holding the backups of other Windows based user files, the overall deduplication effects can often be compounded. This is partly due to the fact that much of the data moving through Exchange is usually similar to the data that exists or is being saved on the users' systems.

When combined with Exchange-aware backup applications, Data Domain deduplication storage systems provide an excellent platform for the backup and restore of Exchange environments. By following some simple guidelines presented in this paper in Section 2, Exchange administrators can make the backup and restore process simple and effective at the same time leveraging the capabilities of the Data Domain deduplication storage systems. By avoiding practices that can alter the original data set (e.g., compression, encryption, defragmentation) and by relaxing the practices that try to minimize the amount of data handled (e.g., deleted items, incremental) it is possible to leverage the disk based solutions from Data Domain without sacrificing performance or capacity considerations.

Appendix

Microsoft Links

Exchange Server 2003 Disaster Recovery Guide

<http://www.microsoft.com/technet/prodtechnol/exchange/2003/library/disrecopgde.msp>

Exchange Server 2003 Technical Reference Guide

[http://technet.microsoft.com/en-us/library/aa996429\(EXCHG.65\).aspx](http://technet.microsoft.com/en-us/library/aa996429(EXCHG.65).aspx)

Offline Backup and Restore Procedures

<http://go.microsoft.com/fwlink/?LinkId=3052&kbid=296788>

What Needs to Be Protected in an Exchange Environment?

[http://technet.microsoft.com/en-us/library/bb124780\(EXCHG.80\).aspx](http://technet.microsoft.com/en-us/library/bb124780(EXCHG.80).aspx)

Exchange Store Maintenance

<http://technet.microsoft.com/en-us/library/aa996226.aspx>

Using Exchange Server 2003 Recovery Storage Groups Guide

<http://technet.microsoft.com/en-us/library/aa998782.aspx>

Exchange Server Database Utility Guide

<http://technet.microsoft.com/en-us/library/aa996953.aspx>

Eseutil /D – Defragmentation Mode

[http://technet.microsoft.com/en-us/library/bb123761\(EXCHG.65\).aspx](http://technet.microsoft.com/en-us/library/bb123761(EXCHG.65).aspx)

How to Set the Maintenance Schedule for a Database (2007)

[http://technet.microsoft.com/en-us/library/bb629590\(EXCHG.80\).aspx](http://technet.microsoft.com/en-us/library/bb629590(EXCHG.80).aspx)

How to Monitor Online Defragmentation

[http://technet.microsoft.com/en-us/library/bb691410\(EXCHG.80\).aspx](http://technet.microsoft.com/en-us/library/bb691410(EXCHG.80).aspx)

Data Domain Links

Appliance Series

<http://www.datadomain.com/products/appliances.html>

Data Invulnerability Architecture

<http://www.datadomain.com/products/DIA.html>

Data Domain Technology

<http://www.datadomain.com/products/technology.html>

Glossary

Many of the terms and references pertain to specific storage technologies and require knowledge of storage design and architecture. Some of the key terms are outlined here for completeness.

API Application Programming Interface

CCR Cluster Continuous Replication

CCR combines automatic management of redundancy and application-level data replication. CCR is a solution that can be deployed without a single point of failure in a single datacenter or between two datacenters. Transaction log replication is used to copy the databases and maintain the concurrency of the data among cluster nodes. The scheduled outage functionality in CCR is designed to make sure that all log data on the active node is successfully copied to the passive node.

Database

In the context of disaster recovery, database is a generic term that refers to either a mailbox store or a public folder store. An Exchange database is composed of both information in memory and the database files on the disk. If the information in memory is lost before it is written to the database files on the disk, it can be replayed from the transaction log files.

ESE Extensible Storage Engine

The database engine that Exchange uses. ESE is a multi-user Indexed Sequential Access Method (ISAM) table manager with full data manipulation language (DML) and data definition language (DDL) capabilities. Applications such as Exchange use ESE to store records and create indexes. An Extensible Storage Engine (ESE) database is used by Exchange. The logs used by ESE are to provide atomicity, consistency, isolation, and durability (ACID) characteristics.

Transaction Log Files

Files that contain a record of the changes that were made to an Exchange database. All changes to the database are recorded in the transaction log files before they are written into the database files. If a database shuts down unexpectedly, unfinished transactions can be restored by replaying the transaction log files into the database.

LCR Local Continuous Replication

LCR lowers the total cost of ownership for Exchange 2007 by reducing the number of regular backups that are required for data protection. Although LCR does not eliminate the need to take backups (data backup are important to have if a disaster strikes), it does significantly reduce the need to take regular, daily backups. LCR provides fast recovery with current data, as well as a single-server solution for transaction log copying and replaying.

PIT Point in Time

A particular version of a copy created by a VSS snapshot or clone operation for the purposes of verification or data access.

RLO Retention Level Objective

This SLA defines the desired retention levels for the backup data sets. Typically this will address how long backup data is kept and whether

the data is kept on-site, off-site or a combination of both. Retention levels for full and incremental backup data are included in this SLA.

RPO Recovery Point Objective

The amount of data that could be lost measured in time. It determines the last known point in time that a backup of the data was captured.

RTO Recovery Time Objective

The amount of time that passes before the business is restored to an operational level with the backup data defined by the RPO.

SLA Service Level Agreement

Combinations of requirements that help define a business' data protection or any other service offering.

Storage Group

An Exchange concept that identifies a set of databases.

VSS Volume Shadow Copy Service

The Volume Shadow Copy Service (VSS) is a service in Windows Server 2003 and Windows Server 2008. It is a framework that facilitates communication between applications such as Exchange 2007, storage subsystems, and storage management applications (including backup applications) in order to define, persist, and exploit point-in-time copies of storage data.

Data Domain
2421 Mission College Blvd.
Santa Clara, CA 95054
866-WE-DDUPE; 408-980-4800
sales@datadomain.com
22 international offices: datadomain.com/company/contacts

Copyright © 2008 Data Domain, Inc. All rights reserved.

Data Domain, Inc. believes information in this publication is accurate as of its publication date. This publication could include technical inaccuracies or typographical errors. The information is subject to change without notice. Changes are periodically added to the information herein; these changes will be incorporated in new additions of the publication. Data Domain, Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time. Reproduction of this publication without prior written permission is forbidden.

The information in this publication is provided "as is". Data Domain, Inc. makes no representations or warranties of any kind, with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Data Domain and Global Compression are trademarks of Data Domain, Inc. All other brands, products, service names, trademarks, or registered service marks are used to identify the products or services of their respective owners.
WP-MES-0608