BEST PRACTICES FOR PROTECTING MICROSOFT EXCHANGE DATA

Bill Webster

September 25, 2003

VERITAS ARCHITECT NETWORK

TABLE OF CONTENTS

Introduction

For many organizations Microsoft Exchange messaging and collaboration applications are now as critical to business success as electronic commerce and online transaction processing (OLTP). Safeguarding the business information in Exchange databases against loss and corruption requires the same level of protection available to more traditional business applications.

For administrators charged with maintaining the near-continuous availability demands of the Exchange messaging environment, data protection means backup and recovery. Although frequent backups generally provide faster recovery, the need for minimal end-user downtime during the backup process, and fast recovery in the event of an outage - be it the loss of a single mailbox, a local database problem, or a system-wide disaster - has administrators looking to total storage management solutions.

This article takes a look at the complexities of availability, backup, and recovery encountered when deploying and maintaining Microsoft Exchange Server. A set of storage management best practices is discussed that provide, along with the VERITAS solutions for Microsoft Exchange software, a consistent backup and recovery approach for the entire Exchange environment.

Exchange Data Protection Best Practices

Protecting the data in Microsoft Exchange Server databases requires thorough planning and preparation of both the backup and restore processes. Performing frequent backups is essential. Exchange supports a variety of backup options, including, online full, differential incremental, and cumulative incremental – each applicable to the entire Exchange database – and an optional, but highly recommended, mechanism for backing up and recovering individual mailboxes. Knowing when to deploy the various backup options is important.

Microsoft Exchange Server offers three categories of protection for user data and system files:

1) Application Protection

Providing backup and recovery of the Exchange application files, clustering support, and disaster recovery procedures.

2) Database Protection

The backup and recovery procedures for database volumes within Exchange storage groups and databases.

3) Mailbox-level Protection

The ability to safeguard individual mailbox data, including both mail messages and attachments, for quick restore with minimal impact on the system or network.

Application Protection

Application level protection of Exchange data focuses on the files and settings necessary for day-to-day operation of Exchange Server. These files and settings support the functionality of the Exchange application at the highest level. Procedures to protect files and settings are generally put in place when Exchange is initially deployed, to maximize application effectiveness and tailor the implementation to the specific needs of a user environment.

Backing Up the Host Server For Exchange

The single most important task in protecting the Exchange Server is to plan for regular, verified, online backups. This task is critical, forming the basis for any disaster recovery plan.

In order to provide a successful restore point, the backup of Exchange Server files and settings must be coordinated with the backup of Windows 2000 operating environment data. Storage volumes holding the Windows OS, Windows System State, and Exchange Server software must be backed up together.

Protecting Active Directory

The Windows 2000 Active Directory contains Exchange Server configuration information and must be backed up regularly. Deploying Exchange Server in a redundant configuration requires installation into an Active Directory domain containing two or more domain controllers. Multiple controllers allow for Active Directory replication. If a controller fails or becomes corrupted Exchange data can be restored from a backup and missing transactions replicated from the surviving controller. The Exchange Server must not be installed on the domain controller, however, as this would complicate any recovery effort by requiring Active Directory be restored first.

If Exchange is deployed with Active Directory, the backup of Windows System State data will also capture Active Directory data. In Exchange environments not running Active Directory, a backup of the system state data, on a server running Active Directory, must be coordinated with the backup of Exchange data to ensure a consistent recovery point.

The Microsoft Internet Information Store (IIS) metabase must also be backed up with the Exchange Server databases. If a disaster event prompts a full restore of Exchange, the IIS metabase must be recovered to the Windows 2000 server before the Exchange Server application data is recovered.

Defining Exchange Data Protection Requirements and Plans

Microsoft Exchange Server backup and recovery must be a key component of any enterprise disaster recovery plan, owing to the critical nature of enterprise messaging data. Clearly defining backup and restore resource requirements, and continually reviewing and testing recovery plans, is the only way to ensure successful recovery when disaster strikes. Administrators will find it very helpful to maintain a duplicate Exchange Server configuration solely for the purpose of disaster recovery testing.

Obtaining a consistent backup of Exchange operational data requires the coordinated copying of the following components of the application and operating environment:

- Windows Operating System (OS) files
- Windows System State files
- Microsoft IIS metabase
- Exchange Server application files and database
- Exchange Key Management Services (KMS) and Site Replication Services (SMS) databases

Backing up the Exchange Server operational information must be coordinated to ensure all relevant data is available for recovery. Both backup and restore will benefit from storing Exchange databases across multiple storage groups – a logical set of physical storage devices – and by limiting the size of mailboxes and public folders. The use of open file agents, however, should be avoided to prevent unnecessary recovery complications.

Databases and Logs

Microsoft Exchange Server provides two main databases of user information: the Directory Store and the Message Store. Prior to Exchange 2000, the Directory Store was a stand-alone database, but this information now resides in Active Directory. Although the user data in Active Directory does not change as frequently as messaging data, it is still necessary to coordinate the backup of the Directory Store data with the Message Store.

The Exchange Message Store

The Message Store contains all Exchange messaging data. The data is divided between two databases: the public database, containing public folder data; and the private database, housing individual user mailboxes.

Exchange supports multiple storage groups for the Message Store, allowing better scalability, greater recovery flexibility, and improved support for clustering configurations. Every change to Exchange user data is recorded in the Exchange transaction logs, which operate at the storage group level. If a problem occurs on a device in a storage group, the Message Store can be restored, and the transaction log data re-applied, to bring user information back to a point immediately before the failure.

Online Backup of the Exchange Message Databases

Exchange supports online full, differential incremental, and cumulative incremental backups. Full backups form the foundation of a disaster recovery plan and involve making a complete copy of all Exchange database and message store information. This type of backup is performed often - at least weekly - with either differential or cumulative incremental backups running periodically during the day.

What Exchange Backup Method To Use – And When		
Full	Performed daily, with larger sites performing full backup 2-3 times per week, along with some form of incremental backup	Per each Exchange group. Captures Exchange database and stream files and truncates logs.
Differential	Differential backups should be performed at least once per day depending on service level agreements. Often performed several times a day to minimize backup impact.	Workload dependent. This backup method captures Exchange databases and truncates transaction logs. Not to be used with incremental backup methods configured for the same policy. Use either Differential or Cumulative, not both.
Cumulative	Frequency based on service level agreements or restore requirements.	Fastest restore performance.

Figure 1. Microsoft Exchange Server backup methods.

The key difference between the differential and cumulative incremental backup is in the restore process. Both incremental backups capture changes to the Exchange, KMS, and SRS databases, user mailboxes, and the Public store, but the differential method deletes the transaction logs after each execution whereas the cumulative does not. This means that after restoring the full copy backup, each differential backup must be reapplied in the correct sequence. Cumulative backups, on the other hand, leave the transaction logs intact. Each cumulative backup is able to scan the entire transaction log for changes since the last full backup was executed. During recovery the most recent full backup is restored followed by only the most recent cumulative as it will contain all the changes since the full backup. Execution times for cumulative backups can increase during the day, as the amount of transaction log data to be processed increases.

The Small Office

For office environments with a relatively small number of messages being processed, a nightly full backup may provide adequate data protection and the fastest recovery. If log file growth becomes an issue, incremental copies can be added during the day to provide additional recovery points.

Medium-Sized Business

Many medium-size organizations running Microsoft Exchange Server schedule full backups to run on weekends, supplemented by multiple incremental backups run during the day. If the incremental backup method used is cumulative sufficient disk space must be available to host an entire week of transaction logs. Using differential backups requires less storage space for logs. Scheduling the backup processes, whether full or incremental, to run at the same time of day, week in week out, will help simplify disaster recovery.

The Enterprise

Large-scale Exchange Server deployments require special consideration due to the volume of messages being processed daily. Using storage groups to separate mailboxes, by department or last name, allows backups to be run in parallel, improving overall backup performance and maximizing server I/O utilization. Exchange Server supports up to 4 storage groups, using up to 5 databases per group. These configuration options give administrators tremendous flexibility in the scheduling of backups and recovery of individual storage groups.

Mailbox Level Backup

With many industries subject to increasing attention from regulators, tracking the flow of email that moves in and out of the organization has become a major concern. Documenting the path of a transaction through the business now involves sifting through mountains of corporate email as well as traditional hard copy memos. Without efficient mechanisms to archive and retrieve individual email messages, the cost of locating electronic business communications becomes prohibitive.

The daily backups designed to protect Exchange Server databases in the event of a disaster do not make recovery of individual messages or mailboxes easy. Providing the granular level of access necessary to retrieve individual message requires mailbox-level backups.

Mailbox-level backups use the messaging application programming interface (MAPI) to read Exchange message data and then archive the information to tape. The MAPI interface is ideal for the systematic retrieval of individual Exchange messages, but it does not lend itself to the high-performance, bulk data copy needs of disaster recovery. Using mailbox-level backups invariably means that Exchange data is backed up twice: once for disaster recovery and once for archival and retrieval.

Optimizing Mailbox-level Backups

Although backing up Exchange data twice is unavoidable, steps can be taken to reduce the burden on the messaging system. Not all mailboxes will require message-level restore capabilities, for example, and restricting retrieval functionality to senior management and executives will likely reduce the amount of data being backed up significantly. As with database level backups, the incremental backup of mailboxes will also improve performance by backing up only new email messages. Mailbox size restrictions also offer a common sense means of reducing the volume of data being backed up.

Every Exchange user’s mailbox contains non-critical data that does not require protection through the backup and recovery process. Excluding messages in the Deleted and Sent Items folders for each user will eliminate a substantial number of messages from the backup and speed processing.

Staggering the execution of mailbox backups offers another technique to alleviate the processing burden. Instead of backing up every mailbox every day, mailboxes can be grouped, one group being backed up each day. This technique may not be feasible, depending on end-user needs, but if the aim of backing up mailbox data is solely for long-term data retention, copying every mailbox every day may be overkill.

Exchange messaging data tends to contain a large amount of redundant information. Eliminating this redundant data from the backup process will improve performance. Messages and attachments sent to multiple recipients make up the largest potion of duplicated data in the Exchange database. Technologies that eliminate the backup of redundant data, such as the single instance storage feature in VERITAS’ data protection solutions software, significantly reduce the amount of mailbox-level backup data processed and speed the execution of the backup.

Multi-streaming offers another technique for mailbox-level backup performance improvement. Depending on the availability of backup devices, splitting up a single backup process into multiple data streams, to be run in parallel against multiple tape devices, will reduce the elapsed time of the backup task.

VERITAS Solutions For Exchange

VERITAS Software offers the highest performing and most flexible backup protection available for Exchange Server with VERITAS Backup Exec™ 9.0 *for Windows Servers* and VERITAS NetBackup™ 4.5. Providing full database- and mailbox-level backup and restore capabilities, including support for embedded objects, attributes, and Outlook components, these market leading Exchange protection solutions offer administrators an optimized approach to the recovery of Exchange environments.

Some of the highlights of VERITAS Solutions for Exchange 2000 are as below:

- **Single Instance Storage of attachments:** 70 percent to 90 percent of Exchange data is attributed to attachments. Techniques for eliminating redundancies in attached files result in considerable improvements in backup performance. The Microsoft Exchange Server Single-Instance Storage (SIS) feature enables message attachments to be stored once, regardless of how many recipients received an attached file. VERITAS uses SIS to eliminate writing multiple redundant attachments to backup media, giving faster backups and reduced storage requirements.

- **Improved Flexibility:** VERITAS Solutions support the non-disruptive backup of all Exchange data. The Exchange Server Information Store, Directory data objects, and transaction log files, are transparently backed up, without impacting end-user access to Exchange data. VERITAS uses the native Exchange APIs - Extensible Storage Engine (ESE) API for database backup and recovery, and the MAPI interface for mailbox-level backup and restore – provided by Microsoft. VERITAS software supports full, differential incremental, and cumulative incremental backup tasks. Scheduling options enable unattended execution on a frequency or calendar basis. All database operations are recorded in the Exchange transaction logs, and VERITAS automatically truncates the log files after committed transactions have been backed up.

- **Individual Public Folder Restore:** VERITAS provides mailbox-level backup capabilities that let administrators use a graphical interface to restore individual Exchange mailboxes, folders, and messages . Multiple storage groups are supported for optimized backup and restore and the VERITAS Shared Storage Option provides fibre channel support for LAN-free backup.

- **Specific Folder Exclusion:** VERITAS supports the exclusion of Deleted and Sent Items folders from backup processing, further reducing the amount of data being backed up. Administrators globally omit these folders, eliminating the tedious task of manually processing each user's exclusions separately.

- **Parallel Backup Streams (multiplexing):** VERITAS NetBackup software uses the NEW_STREAM directive to allow administrators to configure any number of backup data streams. The resulting multiplexed backups can be targeted at multiple tape devices for high-performance parallel processing. This functionality is not available in Backup Exec.

Conclusion

VERITAS Backup Exec™ and NetBackup™ software for Microsoft Exchange offer comprehensive storage management solutions for Exchange. With data protection features specifically tuned to support complex Exchange configurations, VERITAS Software provides administrators with the tools necessary to deploy highly resilient and easily recoverable Exchange implementations.

VERITAS has a reputation for solid, dependable, and innovative storage management solutions that support even the most demanding and complex user configurations. Offering a network of customer support, services, and training to meet the needs of every Exchange installation, VERITAS Backup Exec and NetBackup for Microsoft Exchange is the ideal complement to Microsoft Exchange 2000 Server.

Additional Information

To review independent tests confirming that VERITAS provides the faster backup and restore for Microsoft Exchange, please reference the following documents:

- VeriTest: VERITAS Backup Exec™ 9.0 for Windows Servers: Performance testing – http://www.veritest.com/clients/reports/veritas/backup_performance_2_03.pdf
- VeriTest: VERITAS NetBackup™ DataCenter 4.5: Performance testing – http://www.veritest.com/clients/reports/veritas/netbackup.pdf

VERITAS Backup Exec *for Windows Servers* is the leader in Windows data protection, providing comprehensive, cost effective and certified protection for Microsoft Windows server environments for workgroups and remote branch offices. For more information, please visit: http://www.veritas.com/products/category/ProductDetail.jhtml?productId=bews

VERITAS NetBackup delivers mainframe-class data protection for the largest UNIX, Windows, Linux and NetWare enterprise environments, especially for corporate data centers. VERITAS NetBackup DataCenter provides the most advanced media management available, including dynamic tape sharing, and offers optional database agents to enable online, nondisruptive backup of mission critical applications. For more information, please visit:
http://www.veritas.com/products/category/ProductCategory.jhtml?baseId=2021&categoryId=2003

VERITAS Software Corporation

Corporate Headquarters
350 Ellis Street
Mountain View, CA 94043
650-527-8000 or 866-837-4827

For additional information about VERITAS Software, its products, VERITAS Architect Network, or the location of an office near you, please call our corporate headquarters or visit our Web site at www.veritas.com.