PROTECTING MICROSOFT® EXCHANGE

Whitepaper
Double-Take Software, Inc.
Published: October, 2009

Introduction

Though the technological reasons to protect Exchange systems may be self evident, there are quantifiable fiscal reasons to protect them. The dollar value of any given data set may be difficult to calculate, but the cost savings of avoiding even a single Exchange outage can easily be determined. In most organizations, there are at least one or two subsets of end-users who cannot continue to work without these systems. Even if these groups might not regularly produce revenue in the form of direct sales or billable engagements - salaries, benefits and fixed costs still accrue during an outage. Therefore, the loss of the messaging systems for even a few hours could easily result in thousands of dollars in budget outlay without recouping a single dollar in productivity. For revenue-generating groups, the cost of this downtime is more easily quantifiable. Avoiding even one of these outages is not only a good idea for the IT department, but for the CFO's office as well.

Complete protection of Exchange requires a lot more than management of on-site and off-site tape backups. While traditional tape backup is an excellent tool for long-term archiving, true recovery for the sake of business continuity requires real-time data protection which enables disaster recovery and high availability. Having a solution that is cost-effective, hardware independent and scalable is something every IT manager should consider.

What Exchange data needs to be protected?

Exchange can run on various configurations of Windows, and for each of these Exchange/OS configurations, you must also consider the additional complexity of service packs and hardware platforms available. The result of these permutations of configuration and deployment options is a wide variety of systems to be protected. The only common denominator in all of these configurations is that the various Exchange files are stored on Microsoft Windows file systems.

Each Exchange system is actually a complex configuration of multiple databases that are used to store and manipulate data along with a message transport system to move e-mail and other information in and out of these databases. Protecting the database files, along with their log files and checkpoint files, is critical to the recovery of any Exchange server. Generally speaking, each group of databases (Storage Group) will have a set of log files and a checkpoint file. They are identified by the extensions .LOG and .CHK, respectively. Each database unit (Mail or Public Folder Store) will have an e-mail database (.EDB) and a streaming data file (.STM). Only by protecting all of these files can you be sure you can resurrect an Exchange system.

When a recovery from tape is attempted, the tape systems will replace these files onto the original server or a new server, where management tools can be invoked from within the Exchange System and from the command line to re-constitute the databases back to the state they were in when the tape backup was taken. This is an acceptable way to recover from something like a virus attack or human error where a point-in-time copy is needed to restore a good copy of the data. For a majority of outages like hardware failures and site-wide disasters, however, it is not the most effective means of restoring a failed server.

In many environments, solutions such as e-mail archiving are also deployed as an integral part of a company's e-mail architecture. These tools are designed to remove attachments and outdated e-mails from the Exchange server in order to free up disk space and enhance overall performance. This should be considered as part of any company's recover plan for Exchange as it may be possible that not all of your vital data actually resides on the Exchange server itself. These additional systems will also need to be adequately protected in order to fully restore services to end-users. The same theory applies to BlackBerry, GoodLink and other mobile information systems which also integrate with Microsoft Exchange to provide additional functionality such as remote access to e-mail. Without a disaster recovery plan that accommodates these systems, restoring the Exchange server itself is only one step on the road to recovering the entire messaging system.

Finally, as with many other database systems, Exchange requires that each transaction it writes to disk is performed in an explicit order. This is tracked continuously so that Exchange can maintain which changes have been requested and which have actually been committed to the physical database files. Generally referred to as a "transactional database system" and specifically called the JET database engine in Exchange, these databases do not tolerate replication or a backup system that cannot guarantee that write-order integrity of the data is maintained.

Exchange Server 2007

Exchange Server 2007 represents a new platform for messaging, collaboration and communication from Microsoft®. Interweaving with Microsoft Groove 2007, SharePoint®2007 and other products, this version of Exchange Server creates new opportunities for e-mail data protection and availability.

Instead of the traditional stand-alone or front end/back end configurations for Exchange, Microsoft has created a number of server roles to handle distinct portions of messaging and collaboration functionality. The roles are:

- **Mailbox Server** (MS) - contains user data and public folders
- **Hub/Transport** (H/T) - replaces the Message Transfer Agent role and controls flow of messages and other data between components of the Exchange system
- **Client Access Server** (CAS) - allows connections to Outlook, Outlook Web Access and other client systems
- **Unified Messaging Server** (UMS) - creates a portal into voicemail and other forms of enterprise communication systems
- **Edge Servers** (ES) - allows users outside the corporate firewall to connect to a non-secure server that can communicate safely with the Exchange Server 2007 systems that reside within the corporate secure network.

All of the Exchange server roles (except for Edge Servers) can be installed on a single server, but it is more likely that these roles will be distributed between multiple servers for better performance. It is important, whether by architecting an out-of-the-box solution or augmenting Exchange capabilities with 3rd party products like Double-Take® Availability, to ensure that there is not a single point of failure for any of these roles. This is paramount to ensuring the availability of Exchange servers to end users.

With Exchange Server 2007, Microsoft has also introduced Exchange-specific replication capabilities for the first time with the inclusion of Local Continuous Replication and Continuous Cluster Replication for Exchange data. Local Continuous Replication (LCR) and Continuous Cluster Replication (CCR) are each capable of protecting the Exchange database information using unique methods offering two different levels of protection.
LCR provides replication of database information on a Mailbox Server via a form of log shipping. The LCR

replication creates a second copy or data set of the database information on the same server. For recovery, LCR offers the ability to recover the data from the replicated dataset or from a Volume ShadowCopy (VSS) snapshot in the event of malicious intent (virus attack, deliberate data destruction) or other hardware-independent data loss issues. LCR only creates a copy on the same physical server device; the data is not protected against the loss of the entire Exchange server.

CCR provides replication of database information via the same methods as LCR, but also allows for replication of the logs to one or more nodes of a Majority Node Set Microsoft Cluster. The Majority Node Set configuration provides for replication of the quorum disk resource and CCR provides for replication of the Exchange-specific data. Instead of the traditional shared disk cluster architecture, CCR creates a copy of the database information on the second node allowing for failover between nodes as if the system was running under the Shared Disk (now called Single Copy Cluster or SCC) model. As LCR provides replication within a single server, CCR replicates within a single cluster typically providing replication only within the same site.

In the Service Pack 1 for Exchange 2007, Server Continuous Replication (SCR) will also be introduced. SCR allows replication of Exchange database information from one Exchange 2007 Server to another via a similar technology to LCR. While this will allow for a copy of your data in another physical location, it does not provide any form of availability service, and requires a restoration of the data to the original or a new Exchange 2007 Server; or else the use of a third-party tool to provide failover. While it addresses similar customer problems as LCR and CCR, Double-Take Availability from Double-Take Software provides a complimentary solution which can be used either with LCR and CCR or separately to provide disaster recovery and high availability for Exchange Server 2007 environments.

How Double-Take Availability Protects Exchange Data

Double-Take Software has been providing solutions for Exchange Server since before Exchange 5.5 Server was released. Double-Take Software is a Gold-Certified, Independent Software Vendor (ISV) Partner of Microsoft and has gained the Microsoft Partner Program Advanced Infrastructure Solutions Competency for Exchange Server. This partnership, combined with the unique experience we have protecting Exchange servers, ensures that you can effectively and safely protect all versions of Exchange, including Exchange Server 2007.

For stand-alone Exchange servers, Double-Take Availability can be used either by itself or in combination with LCR to provide multiple levels of Exchange protection and recovery. Alone, Double-Take Availability can provide either local or remote failover for Exchange data and services and disaster recovery to one or more backup sites. When combined with LCR and its ability to provide a second local copy of Exchange data, Double-Take Availability can provide a third copy of Exchange data replicated to a remote server for disaster recovery or remote availability purposes. In a distributed Exchange environment, Double-Take Availability can provide an additional level of protection for Exchange 2007 server Hub and Transport, Mailbox Server and Client Access Server roles by replicating the data using asynchronous, byte-level replication to standby systems. This allows for the movement of an entire server role (or roles) to another server in the same or a different data center as necessary.

For remote availability and disaster recovery of Microsoft Clusters running Exchange Server 2007, Double-Take Availability can provide a reliable, proven method to protect and Exchange cluster regardless if it is a "traditional" Failover Clustering shared storage scenario, or a cluster created using the new CCR cluster capabilities. In enterprises that are using clustering to provide local high-availability, Double-Take Availability can provide replication and failover of the cluster to another cluster at another physical site, even if that site is on an entirely different IP subnet. This provides both local availability through the clustering system, and the ability to resume services at another site in the event of a site-wide outage.

The combination of the award-winning, proven protection and recovery capabilities of Double-Take Availability with the recovery and availability features of Microsoft Exchange Server can provide an end-to-end solution for Exchange protection, eliminate single points of failure and provide the utmost flexibility and redundancy for your business-critical Exchange messaging data. One of the strengths of Double-Take Availability replication technology is that it protects files at the byte-level regardless of the application. In this case, when Exchange writes data to any of its files, the actual byte-level changes it makes to the Windows file system are sent to another Windows server. Once the data is protected to another server, multiple options are available for achieving availability and disaster recovery goals. The first capability this method of data protection offers is a truly hardware-independent, version-independent, and OS-independent data protection solution. Simply put, Exchange resides on a Windows file system and Double-Take Availability can protect those file systems.

Additionally, Double-Take Availability can ensure that each Windows I/O transaction for a protected data-set is not only sent to one or more DR systems with full data integrity, but also that it will be committed in the exact same order that the original Exchange server committed the changes. Double-Take Availability affords both data and write-order integrity, allowing the Exchange and the JET database system to immediately recognize a consistent copy of the data on the recovery system(s).

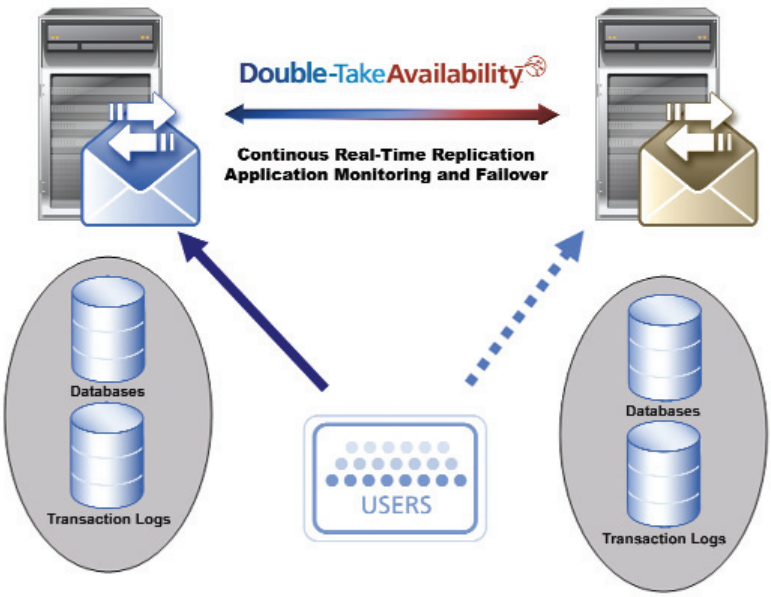
Double-Take Availability Application Manager

Double-Take Availability Application Manager automates the setup and configuration of replication and failover monitoring of Exchange servers for rapid and successful deployment. Features such as auto-discovery of servers and key Exchange data files simplify the process down to just four steps and reduce the risk of human error. Exclusive features such as the pre-flight check help to ensure that all necessary settings within the environment are configured correctly for replication and failover of Exchange. Any errors identified are listed along with suggestions for resolution. The Application Manager checks over 50 different configuration criteria and can automatically correct a majority of them on the administrator's behalf. This helps achieve a seamless, error-proof deployment. The Application Manager performs recovery and failback of Exchange servers with a minimal window of downtime. By performing recovery tasks while users remain online, Exchange downtime is reduced to just the few moments that it takes for Double-Take Availability to stop and start the necessary Exchange services and relegate processing back to the production server. Double-Take Availability is at the core of our Exchange solutions, making available the extensive set of features and functions it offers. All advanced Double-Take Availability features including intelligent data compression and flexible bandwidth scheduling are available for fast and efficient replication of Exchange data.

How to Ensure Exchange Availability

With Stand-Alone Exchange servers:

Exchange systems can come in a variety of flavors and configurations to meet nearly every business need and budget. By far the most common system employed today is the Stand-Alone Exchange server (SAES). SAES systems are a single server running all components of the Exchange system on the same physical or virtual system. This means the server acts as a mail-transport system, routing system, SMTP server and gateway and mail and public folder database server. It may also provide spam filtering, anti-virus scanning or any other function related to Exchange.



The end result of this configuration is that you have a single server that must be protected - all your eggs are quite literally in one basket. Double-Take Availability can provide you with data replication of all key information for the Exchange system and any other systems running on the production server. In addition, the Application Manager can help you prepare a secondary server to take over in the event of a loss of the primary. The Application Manager provides application-specific configuration, availability and management features and is a free toolkit available as part of Double-Take Availability.

The Application Manager will select the appropriate directories and volumes on your production machine which need to be protected, configure Double-Take Availability replication, and configure the secondary server's Exchange configuration to match that of the production machine. It will also prepare the secondary server to execute the necessary commands to start Exchange services during an outage and allow failover to occur. During this configuration process, you can specify what network path the replication systems should use, if data should be compressed for transmission and if you wish to manually initiate the failover process ("one-click failover") or have it happen automatically after a timeout you define.

During an outage, the Application Manager will either automatically initiate a failover if the production server is unreachable for the amount of time you set, or alert you and wait for you to initiate the failover manually. Alerting for either type of failover scenario is available via SNMP, SMTP and the Windows Event Log, in addition to the native Double-Take Availability management tools. Regardless of which failover methodology you choose, the procedure for restoring services for end-users is the same.

First, Double-Take Availability's DNS Fail Over (DFO) component will update Active Directory DNS servers to re-route end-users to the recovery server. You may specify any and/or all DNS record types for update, depending on what systems you need re-directed. Double-Take Software Professional Services can also assist you with providing automated failover for customized DNS or Exchange deployments as well.

After DFO re-routes the end-users, the Double-Take Availability Exchange Failover component will dynamically re-assign all mailboxes and Public Folders from the failed server to the recovery server. This will allow end-users to regain access to their information the next time they attempt to connect to Exchange. Internal testing and real-world feedback has revealed that this failover process is very fast and creates a relatively small load on your existing DNS and Active Directory infrastructure - approximately 14,000 users can be moved to secondary server in around 7 minutes. In fact, while conservative failover estimates are suggested to be placed at 45 minutes for total failover, most customers report being able to failover in less than 20 minutes - even when performing the failover across a WAN connection.

Finally, the Application Manager will prepare and start the appropriate Exchange and 3rd party services (anti-virus, mobile mail, archiving, etc) on the recovery server. This last step re-establishes a live Exchange server for your end-users to connect to. Outlook clients may need to be re-started, but no end-user configuration will be required. In addition, other components such as Outlook Web Access (OWA) and other e-mail integrated system will simply pick up where they left off.

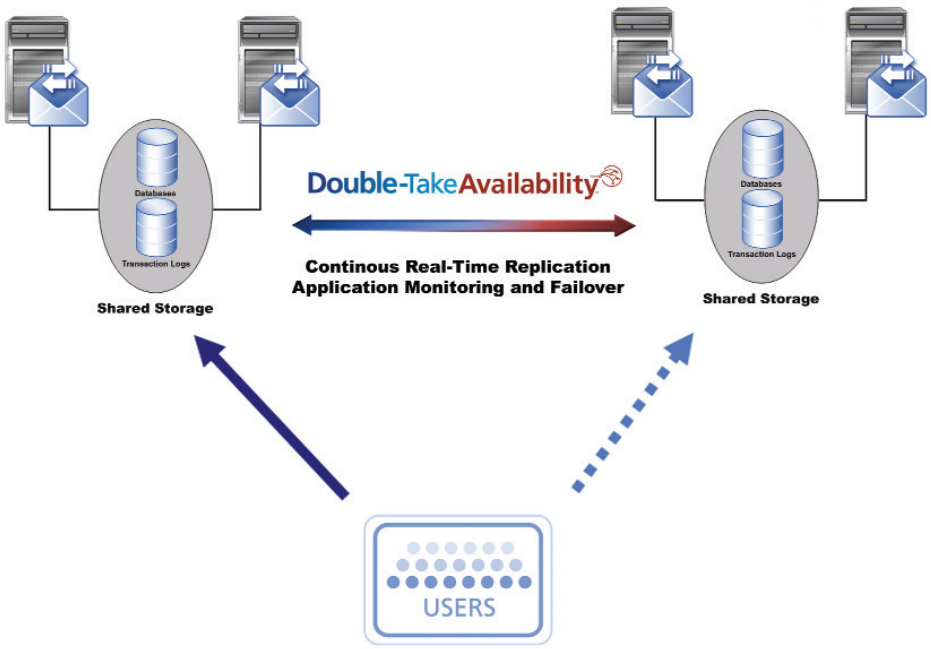
With Front-End/Back-End Exchange servers:

Front-End/Back-End (FE/BE) configurations for Exchange servers are becoming more and more popular as spam filters, virus firewalls and other solution sets make putting those systems on a non-database server much more effective. Front-End servers house mail routing and transfer systems, third-party tools for mail management and often the Outlook Web Access component of Microsoft Exchange. In these configurations, Back-End servers function as repositories for mailbox and public folder data used by end users.

In many ways, the method by which Double-Take Availability protects FE/BE systems is very similar to the method employed with SAE systems described earlier. By leveraging the inherent load balancing capabilities of Exchange services such as Outlook Web Access or the load balancing feature of other Exchange-related solutions, a second Front-End server can be configured to accept mail and perform other functions either when the primary Front-End server fails or dynamically as additional resources are needed. Double-Take Availability is then deployed to the Back-End servers to provide disaster recovery and high availability for the Exchange databases used by the Front-End services. All options available for protecting SAE systems with the Application Manager are also available for protecting Back-End systems.

With Microsoft Cluster Services and Exchange Server:

In some cases, even a 20 minute failover time is simply too long to withstand. Microsoft Failover Clustering can provide your organization with faster failover times in some cases - often failover occurs in a matter of just a few minutes. Though Failover Clustering provides some benefits in terms of failover for Exchange, there are also potential drawbacks to be considered that may affect your decision to implement the technology. Firstly, Failover Clustering requires hardware that is certified to be compatible with Failover Clustering and also requires the purchase of shared disk array. Secondly, the Failover Clustering architecture (two servers, a shared disk array, etc) may not be ideal for situations where you wish to separate the nodes of the customer across great distances. Lastly, due to the nature of Failover Clustering and the shared-disk configuration it uses, two servers will alternately use the same copy of the data. This makes Failover Clustering an exceptionally good system for local availability, but potentially introduces a single point of failure if an entire site experienced a disaster or other outage.



Double-Take Availability used in combination with Failover Clustering, however, allows for the protection of an Exchange Cluster to either another single- or multi-node Exchange cluster at a remote recovery location, or a stand-alone Exchange server. By integrating seamlessly into the Failover Clustering component of Windows Server, it does this without adding additional complexity to cluster administration. This is a best-of-both-worlds solution, offering Failover Clustering failover in the event of a single-server failure, but still allowing for a redundant copy of the data on another system and the ability to fail over via the Application Manager to that secondary system in the event that the entire production cluster is lost.

Double-Take Availability, when used in conjunction with an Exchange cluster failover, performs nearly identically to the Stand-Alone or Front-End/Back-End configurations outlined earlier. In the case of Exchange clusters, however, replication is performed from the "owning node" of the cluster. The "owning node" is the cluster node that is currently running the Exchange services. In non-cluster deployments, Double-Take Availability would replicate from a single, physical machine. In a clustered deployment of Double-Take Availability, however, replication occurs from the "owning node" and switches to another node should Failover Clustering failover Exchange to another node in the cluster. Because Double-Take Availability is tightly integrated with Microsoft Failover Clustering, this happens automatically without administrator intervention.

Other Considerations for Exchange and the Rest of Your Environment:

It is likely that new versions, service packs, hot fixes, and 3rd party add-ons will make protecting Exchange even more difficult in the future. By focusing on the Windows file system and OS, Double-Take Software will continue to provide value to customers with its effective, yet simple, approach to Exchange protection. This approach, along with the stability and scalability of the solution, is a reason for the current leadership position Double-Take Software maintains among Exchange protection technologies.

When considering enterprise technologies, it can be difficult to select vendors that support large areas of complex and heterogeneous organizations. When considering the variety of applications (such as Exchange, SQL, Oracle, and file services) and the different versions of each of these applications in use, the task is even more daunting. However, because Double-Take Software replication technologies focus on data replication and then assist in the pre-configuration of applications independently, the same level of data protection for Exchange is equally viable for any other Windows based application. More simply put, you can standardize on one Windows availability solution, regardless of the myriad of applications in your environment.