

High Availability and Disaster Recovery features in Microsoft® Exchange Server 2007 SP1

Product Group - Enterprise

Dell White Paper

By

Farrukh Noman

Ananda Sankaran

April 2008

Contents

Introduction.....	3
High Availability Features	4
Exchange Server 2007 RTM.....	4
Single Copy Cluster	4
Local Continuous Replication.....	4
Cluster Continuous Replication	5
Exchange Server 2007 SP1	5
Clustered Continuous Replication (CCR) in SP1	7
Passive Node Performance Improvement.....	7
Standby Continuous Replication (SCR).....	12
Standby Continuous Replication (SCR).....	12
Deployment Scenarios	14
SCR replication on Stand-alone Mailbox server	14
SCR replication with LCR enabled Mailbox server	15
SCR replication with SCC Mailbox server	16
SCR replication with CCR Mailbox server	17
Performance	18
Conclusion	22

Introduction

Data and service availability mechanisms are crucial for IT organizations in preventing application outages due to a variety of failures. Enterprises are considering efficient ways to reduce the downtime of their messaging infrastructure which is critical for business operations. Microsoft® Exchange Server 2007 RTM (Release to Manufacturing) introduced new mailbox availability options like Local Continuous Replication (LCR) and Cluster Continuous Replication (CCR) as compared to Exchange Server 2003 which was limited to the shared storage clustering based on Microsoft Cluster Services. The new release of Exchange Server 2007 Service Pack 1 (SP1) complements Exchange Server 2007 by enhancing the operational performance of high availability features, adding disaster recovery options and adding support for Windows Server® 2008. This article will discuss the Exchange Server 2007 RTM and SP1 high availability features and associated benefits. Performance analysis conducted at Dell labs to characterize the behavior of new high availability options and improvements will also be discussed.

High Availability Features

Achieving high availability for the mailbox server application and mailbox data is a crucial requirement of every Enterprise messaging system. Exchange Server 2007 provides various availability options which can be deployed based on the specific needs. The following sections describe the various availability options in Exchange Server 2007 RTM version and the new additions and improvements available in the Exchange server 2007 SP1 release.

Exchange Server 2007 RTM

The RTM version of Exchange Server 2007 provides the following built-in high availability options for mailbox server and data:

Single Copy Cluster

Single Copy Cluster (SCC) is based on the shared storage Microsoft Cluster Services (MSCS) clustering model that existed with previous Exchange versions. It follows an active-passive and shared-nothing architecture wherein a single copy of the storage groups and databases reside on shared external storage. Two or more Exchange mailbox servers are connected to this shared storage to form a cluster. The active node hosts the database residing on the shared storage and fails it over to the passive node during a failure. It is recommended that each active mailbox server in the cluster should have a corresponding passive node. This is because at any time a cluster node can host only one active clustered mailbox server and, when multiple active node failures occur, there may not be sufficient passive nodes to handle the failover. Windows Server 2003 supports up to eight server nodes in a cluster. Compared with previous versions, Exchange Server 2007 RTM provides improved deployment setup and management experience with this clustering model. SCC ensures availability of only the mailbox server role. In SCC deployments, no other server role can be consolidated with the Mailbox role on the clustered servers.

Local Continuous Replication

Local Continuous Replication (LCR) is a single-server solution that provides availability by creating and maintaining a copy of the Exchange storage groups' logs and database on a second set of disk volumes connected to the same mailbox server. The copy is maintained asynchronously using transaction log copy and replay on the passive target databases. The passive copy is initially created by copying the storage group from active copy through a "seeding" process and the subsequent updates are synchronized via log copy process. The copied transaction logs are then replayed on the passive storage group's database. In case of an active copy failure, the mailbox server can be pointed manually to start using the passive copy as the production version. The passive database copies can be used to offload the required backup activities from the active databases with minimal impact to the end user response time. The

availability of passive copy also allows modifying the backup schedule and reducing the full-backup frequency from daily to weekly.

Cluster Continuous Replication

Cluster Continuous Replication (CCR) is based on the Microsoft Cluster Services (MSCS) Majority Node Set (MNS) cluster model. The active and passive mailbox server nodes within the cluster maintain their own copy of mailbox databases on non-shared storage. The database and storage group on the passive node is created by copying the active node's database copy through a seeding process. After initial seeding, the passive copy of storage group is kept consistent with the active copy through data replication. The replication takes place in the form of asynchronous transaction log copy and replay. After an active node failure, the passive node automatically takes over hosting mailboxes and becomes the active node via failover process. A third server, called witness file share, provides arbitration mechanism and makes sure that only one server functions as active node at any time. The file share witness server must be hosted by a machine that is not part of the CCR cluster but should be part of the Active Directory® domain containing the cluster nodes, and it is recommended to be the Hub Transport server.

CCR offers enhancements to data backup and recovery strategy similar to LCR. The passive database copy in CCR can be utilized for offloading required database backup activity and also to reduce backup frequency. The Hub Transport server includes a feature called Transport Dumpster which is utilized to avoid data loss during failover of the active mailbox server node. The Transport Dumpster, maintained in the Hub Transport server, is a queue of recently delivered messages to the active mailbox server from clients and other services. At the time of a failure, the active mailbox server may not have completed processing these recent messages. Thus to ensure their completion after a failover, all Hub Transport servers in the Active Directory site are requested to resubmit the mail in their Transport Dumpster queue to the new active mailbox server. This ensures that all recently delivered messages get recorded and prevents data loss during failures. Certain use cases or scenarios exist where this feature does not provide complete recovery from user data loss.

Exchange Server 2007 SP1

Exchange Server 2007 SP1 further enhances Exchange Server 2007 by improving the performance of existing availability features. The Cluster continuous replication (CCR) in SP1 has been enhanced by supporting redundant separate cluster networks for log shipping and database seeding. This reduces the network congestion from log copying and client traffic which utilizes the same public network. One or more mixed networks in the cluster allow proper distribution of traffic. The failover time for clustered mailbox server in CCR environment has also been improved by not purging the database cache when the database goes offline during a cluster failover. Retaining the cache contents allows connectivity to the clients while the clustered mailbox server is going through a down-time period. Further modifications have been done to improve the performance of passive node during cluster continuous replication, which are described in more detail in the next section.

The LCR feature in Exchange Server 2007 SP1 has also been improved by adding the transport dumpster mechanism which was supported only with CCR in the RTM version. As mentioned earlier, the transport dumpster in CCR keeps a queue of recently delivered messages in the Hub Transport Server which are automatically recalled by the passive mailbox server during the failover process. In LCR the process of recalling these messages is manual due to the manual recovery process associated with LCR copies.

There is a new disaster recovery option included with Exchange 2007 SP1 which is called as Standby Continuous Replication (SCR). Unlike LCR and CCR, SCR allows multiple copies of standby database at different target locations to increase redundancy and handle worst possible disaster scenarios. The SCR feature and associated deployment scenarios are further discussed in later sections.

Windows Server 2008 operating system is also supported with the SP1 release besides Windows Server 2003. Windows Server 2008 brings new and improved high-availability options. One of the improved features includes support for different subnets for the cluster nodes to address geographically dispersed cluster scenarios. Windows Server 2008 offers new quorum models for avoiding split-brain scenarios and improving quorum resiliency.

Clustered Continuous Replication (CCR) in SP1

Exchange Server 2007 SP1 offers improved performance of I/O operations on the CCR passive node due to passive node database cache optimizations, besides other improvements to CCR described in earlier sections. In the previous RTM version, the replication process created new database cache for every log replay activity on the passive node. This process of populating and purging the database cache frequently generates heavy I/O on the passive side. The new architectural modification in Exchange Server 2007 SP1 allows database cache to retain data during replay cycles and reduces I/O.

Passive Node Performance Improvement

A set of performance tests were conducted to compare the CCR improvement between Exchange Server 2007 SP1 and RTM versions. Simulations were conducted on CCR configuration as illustrated in Figure 1. The resource utilization of active/passive mailbox server nodes and their associated storage subsystem were recorded. The configuration details of Figure 1 are given below.

Configuration Details:

- Mailbox servers (Active and passive)
- Dell™ PowerEdge™ 2950 with 2 x Dual Core Intel™ Xeon® 5160, 3.00 GHz processors; 8 GB system RAM
- Windows Server 2003 R2 Enterprise x64 Edition with SP2; Exchange Server 2007 Enterprise Edition RTM; Exchange Server 2007 Enterprise Edition SP1
- Hub Transport/Client Access server
- Dell PowerEdge 2950 with 2 x Dual Core Intel Xeon X5160, 3.00 GHz processors; 8 GB system RAM
- Windows Server 2003 R2 Enterprise x64 Edition with SP2; Exchange Server 2007 Enterprise Edition RTM; Exchange Server 2007 Enterprise Edition SP1
- External mailbox storage
- 2 x Dell PowerVault™ MD1000 (each on active and passive node)
- Database volume: RAID 10 with 10 x 146 GB 15K RPM SAS drives (MD1000)
- Log volume: RAID 1 with 2 x 73 GB 15K RPM SAS drives (MD1000)
- Database copy volume: RAID 10 with 10 x 146 GB 15K RPM SAS drives (MD1000)
- Log copy volume: RAID 1 with 2 x 73 GB 15K RPM SAS drives (MD1000)

- Microsoft LoadGen Simulation Tool
- Build version: 08.01.0177.000
- User profile: 1000 heavy users executing 94 tasks per 8 hour user day in Outlook® 2007 online mode

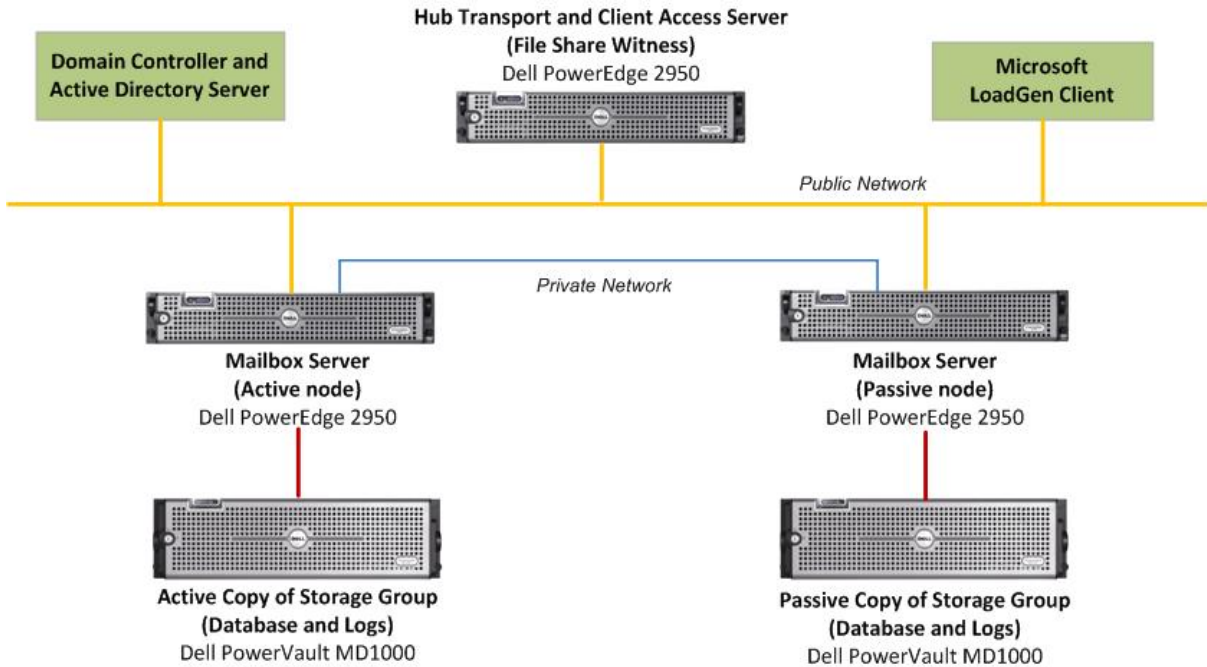


Figure 1: CCR Test Configuration

User load was simulated on the test setup with Exchange Server 2007 RTM version installed on cluster nodes. The same user load was simulated on the test setup with Exchange Server 2007 SP1 installed on cluster nodes. As shown in Figure 2, the database read IOPS on the passive node with RTM version appears significant and amounts to more than double the read IOPS on the active node with RTM version. This is due to the same phenomenon mentioned earlier about the database cache purging and populating for every replay activity. With SP1, the CCR database reads and writes on active node remains almost the same as compared with the RTM version but the database I/Os on passive node show a significantly large improvement. There is a decrease of approximately 78% in the database read IOPS on the passive node with SP1. Similarly a reduction of approximately 12% with the database read IOPS was observed on the passive node.

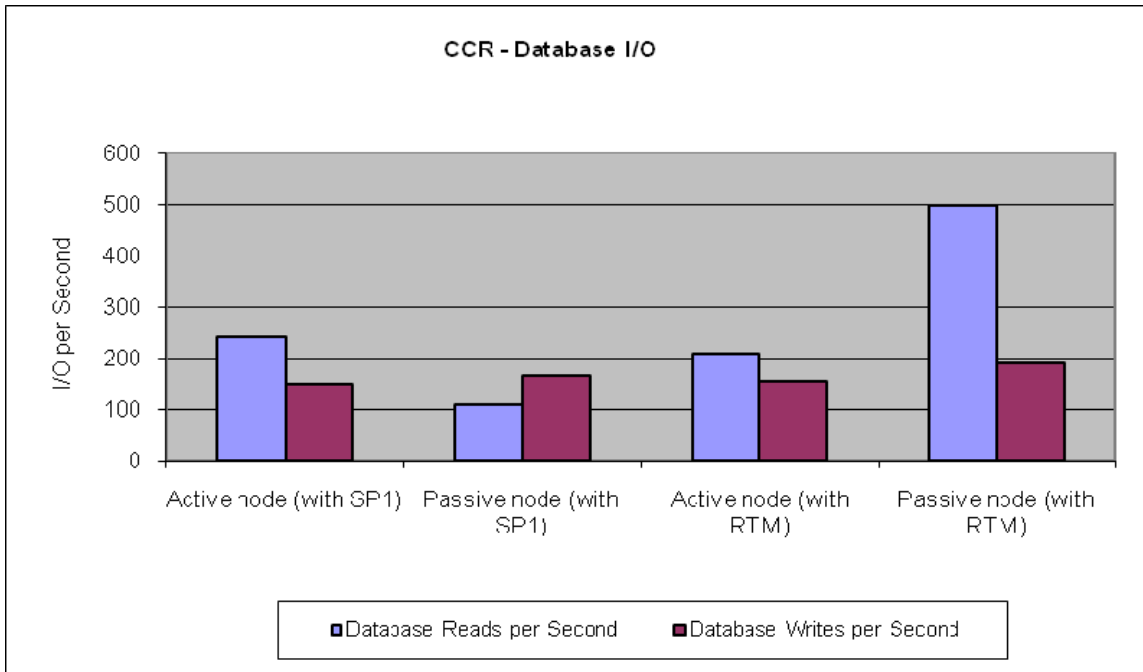


Figure 2: CCR Database I/O comparison for Exchange Server 2007 RTM and SP1

The log I/O in Figure 3 also shows a similar reduction in Log I/Os on Microsoft Exchange Server 2007 SP1 passive node, with an approximate decrease of 36% in log read IOPS, as compared to that of Microsoft Exchange Server 2007 RTM version. There is an overall reduction in database and log I/O operations which will allow relaxing the storage resource requirements on the passive node.

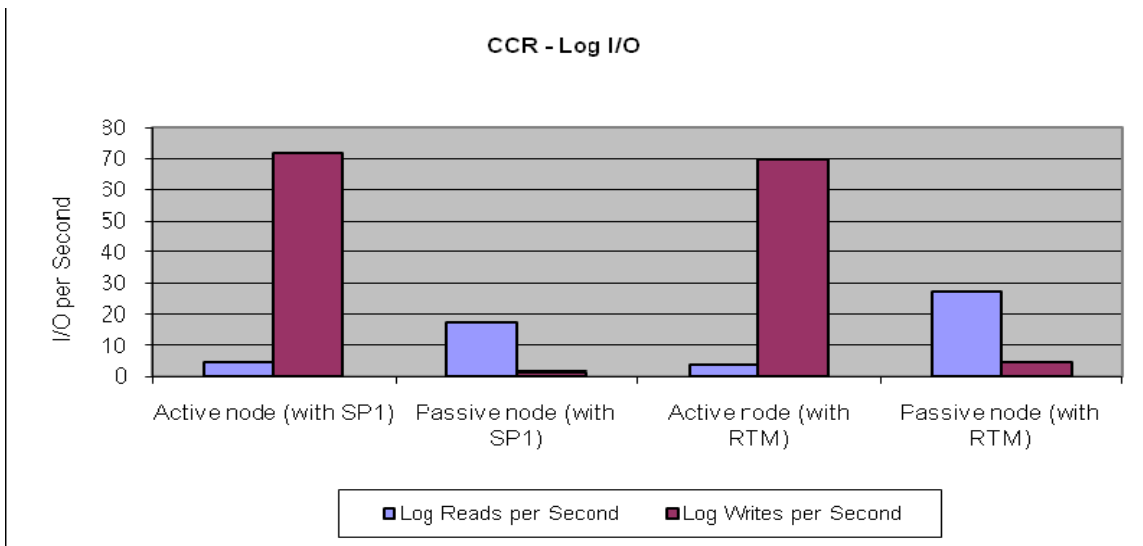


Figure 3: CCR Log I/O comparison for Exchange Server 2007 RTM and SP1

The memory utilization of mailbox servers on active nodes for both SP1 and RTM versions are approximately equal (as shown in Figure 4) but it is almost quadrupled on the passive node with SP1. The higher memory utilization occurs at the expense of reducing the database I/Os during the replay cycles and retaining the data for efficient use. The same reason accounts for slightly higher CPU utilization on the passive node with SP1 compared to that of RTM version (as shown in Figure 5).

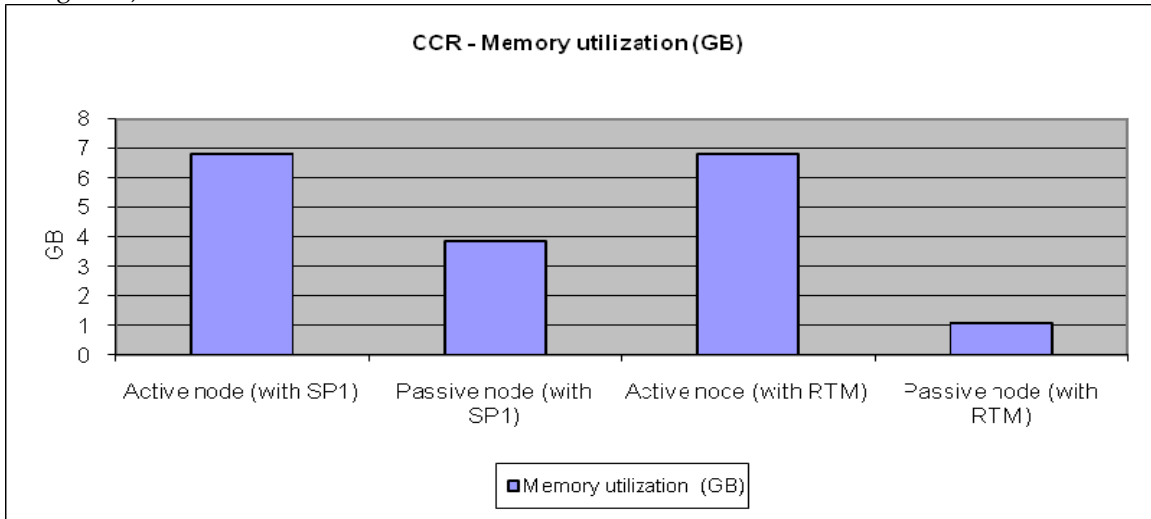


Figure 4: CCR Memory Utilization comparison for Exchange Server 2007 RTM and SP1

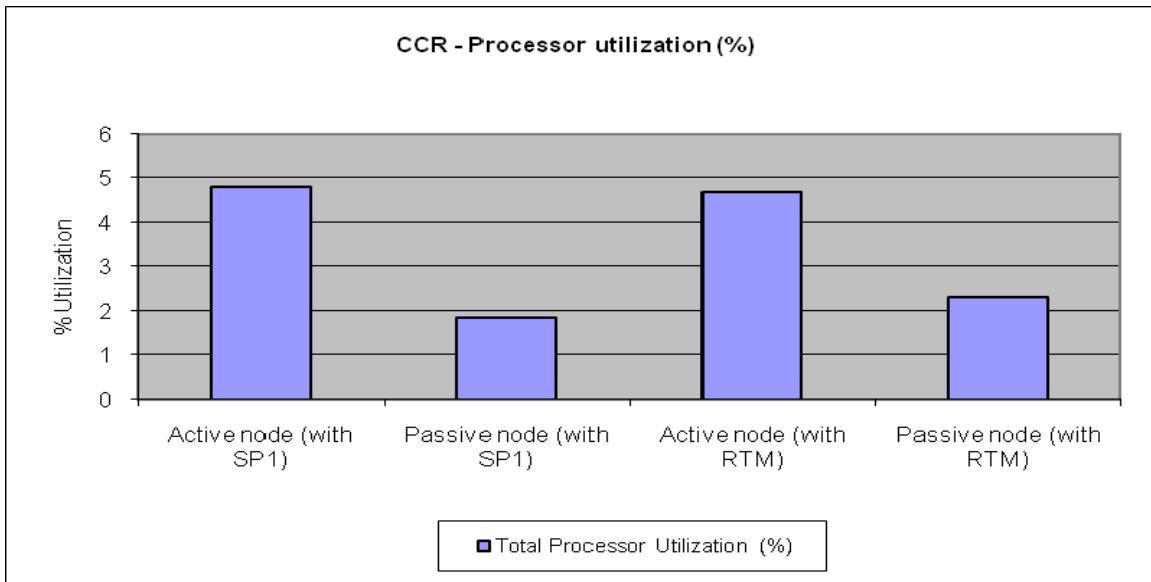


Figure 5: CCR Processor Utilization comparison for Exchange Server 2007 RTM and SP1

The memory and CPU utilization in SP1 passive node increases but is still significantly less than the usage incurred on the active node and therefore does not require any major server hardware upgrades on the passive node. Generally, the same set of processor and memory resources are provided for both active and passive nodes in a CCR deployment.

Designing of the CCR storage subsystem with RTM version required careful consideration due to more database I/Os on the passive node compared to the active node. In most cases, it was provisioned with the same set of disk resources as the active node. This is because the passive node does not handle production load until after failover. Exchange Server 2007 SP1 provides greater performance on the passive node by reducing the database I/O operations, without any need for additional hardware resources.

Standby Continuous Replication (SCR)

Standby Continuous Replication (SCR) is the new high availability and disaster recovery feature provided in Exchange Server 2007 SP1. It overcomes the restrictions posed in existing continuous replication technologies by supporting multiple replication targets per storage group and multiple source server storage groups per target server. Mailbox data from source servers are replicated to target servers via the same built-in log copy and replay technology available with LCR and CCR. SCR source can be standalone mailbox servers with or without LCR enabled or CCR clusters or SCR clusters. SCR target can be a standalone server without LCR or passive node in a SCC cluster with no clustered mailbox role installed. SCR requires a standalone mailbox role installed in the target and supports one database per storage group for replication. The SCR target must be part of the same active directory domain, but can be part of a different active directory site and supports site resilience. SCR is supported with Exchange Server 2007 SP1 Standard and Enterprise Editions. CCR and SCC require Enterprise Edition due to clustering requirements.

The SCR technology copies the log data from a source node to a single or multiple target nodes and provides administrator flexibility to replay these logs at a later time. The replay lag time which can be easily varied from instant replay to seven days, is used to prevent logical corruption being replicated to the target node. In such situations, the SCR target or standby node is manually promoted or failed over to the production environment. The SCR targets are therefore recommended to have similar set of mailbox server and storage sub-system resource as the source node in order to address failure scenarios. There is an additional built-in replay delay of 50 log files with SCR. The Exchange replication service chooses the maximum of the two values; the user provided replay lag time or the 50 logs replay delay. So if the user sets the replay lag time to zero seconds, the SCR process will still wait for 50 logs to be generated before initiating the replay. Waiting for at least 50 logs ensures a safety lock on the target database for recovery purposes.

SCR uses a slightly modified policy for truncating transactional logs. Usually the log truncation in a continuous replication environment does not happen until all logs have been copied and replayed. Since SCR allows multiple targets with varying replay delays, the transactional logs on the source can take considerably higher disk space if the source waits for all target logs to complete replay. To overcome this problem, SCR source allows truncation of logs once they are checked and copied to the all the targets. The log truncation on the target node does not happen until the completion of log replay lag time, actual log replay process and log truncation lag time set by the administrator. The truncation lag time like replay lag time has a maximum grace period of seven days which is carefully chosen to make efficient use of the log disk space.

Table 1 below provides a high level comparison of all the mailbox high availability features described in earlier sections along with SCR. Specific deployment rules apply for deploying these features. Appropriate product and feature documentation should be consulted before planning these deployments.

Feature	SCC	LCR	CCR	SCR
Availability level	<i>Application</i>	<i>Data</i>	<i>Application and Data</i>	<i>Application and Data</i>
Automatic Failover	<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>
Native Data Replication	<i>No</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Disaster Recovery	<i>No</i>	<i>No</i>	<i>Yes</i>	<i>Yes</i>
Replication Targets per Storage Group	<i>None</i>	<i>Single</i>	<i>Single</i>	<i>Multiple</i>
Replay delay option	<i>N/A</i>	<i>No</i>	<i>No</i>	<i>Yes</i>
Microsoft Windows Server Catalog Listing for cluster solution hardware	<i>Yes</i>	<i>No</i>	<i>No</i>	<i>No</i>
Backup improvements	<i>No</i>	<i>Yes – offload to passive copy</i>	<i>Yes – offload to passive copy</i>	<i>No –Target copies cannot be used for backup</i>

Table 1: Mailbox Availability feature Comparison

SCR can also be deployed in combination with existing replication technologies like LCR, SCC and CCR to increase redundancy and site resiliency. The deployment scenarios are further described in the next section. The process of enabling SCR is similar in most deployments but the recovery option may vary depending on the type of failure and presence of cluster configuration. During disaster recovery of a target, the storage groups are restored manually on the target node using “Setup /m:RecoverServer” command for non-clustered mailbox servers or “Setup /RecoverCMS” for a clustered mailbox server deployments. Database portability can be used for recovery from logical corruptions on the source node, where the source data is cleanly shutdown and the target database is brought online manually. User mailbox configuration can also be pointed to the new server during this process. After recovery, the target node becomes the production or source node and replication can be enabled on it with SCR to restore redundancy.

Deployment Scenarios

In SCR, a source node can have multiple standby targets and similarly a single target can be configured to receive data from multiple sources. A single target can hold up to 50 storage groups in Exchange 2007 Enterprise Edition deployment and up to 5 in Standard Edition deployment. The source node is recommended not to deploy more than 4 targets for performance degradation reasons. SCR can be used in combination with existing replication technologies to increase redundancy and availability. A few deployment scenarios are illustrated below.

SCR replication on Stand-alone Mailbox server

In SCR a standalone mailbox server can act as a source (host production databases) and also as a target (host SCR copy) for another source. Thus SCR can be used in bi-directional mode for two Stand-alone mailbox servers, where each server can act as source and target at the same time. The only restrictions in such an environment are that the targets should not have LCR enabled for any storage group and it should not be a clustered mailbox server.

The unidirectional and bi-directional SCR replication processes with stand-alone Mailbox servers are shown in Figure 6 and 7 respectively.

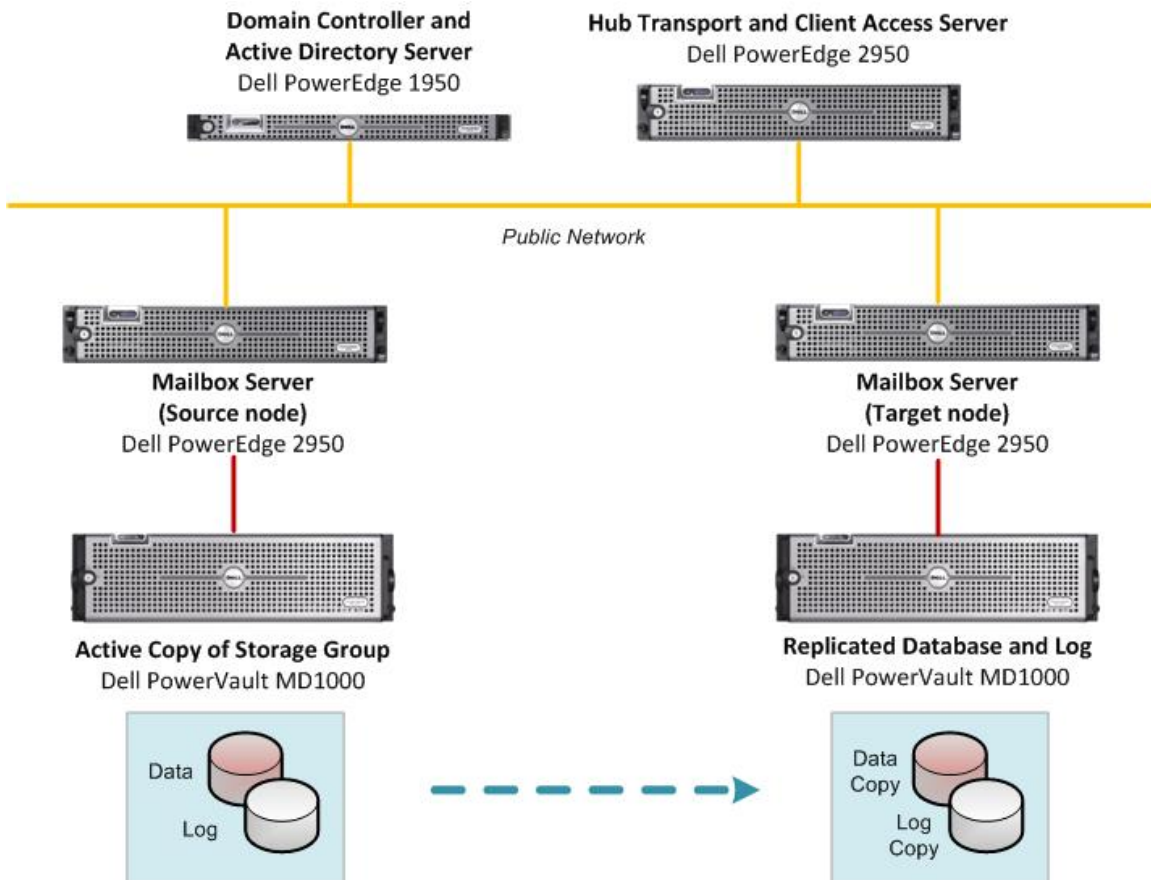


Figure 6: SCR Unidirectional replication on Stand-alone Mailbox server

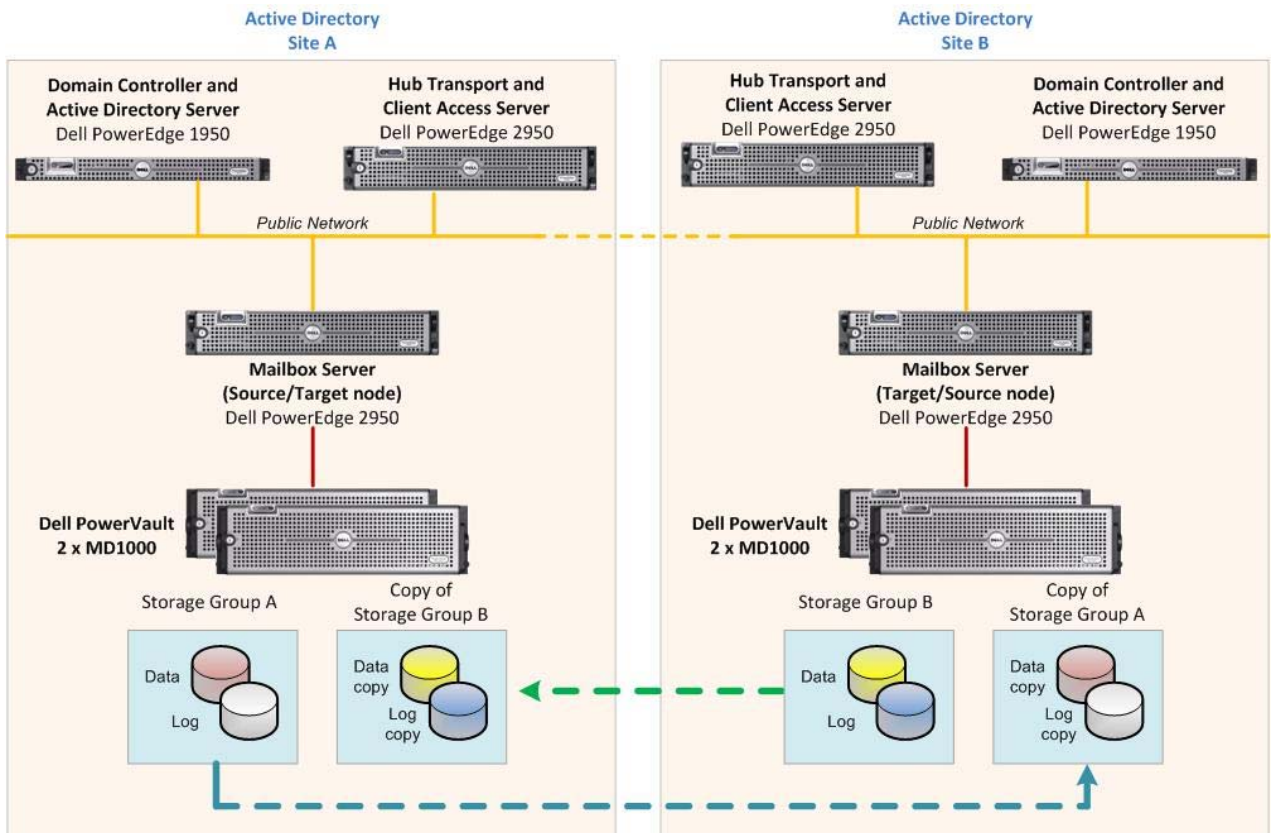


Figure 7: SCR Bidirectional replication on Stand-alone Mailbox servers

SCR replication with LCR enabled Mailbox server

Local continuous replication (LCR) creates an asynchronous backup copy of transactional database and logs on a separate set of volume disks connected to the same mailbox server. The SCR feature on an LCR configuration expands the scope of backup from local to remote by replicating the source logs to remote target(s) and adding site resiliency to data. Figure 8 illustrates a simple one-to-one LCR configuration with SCR.

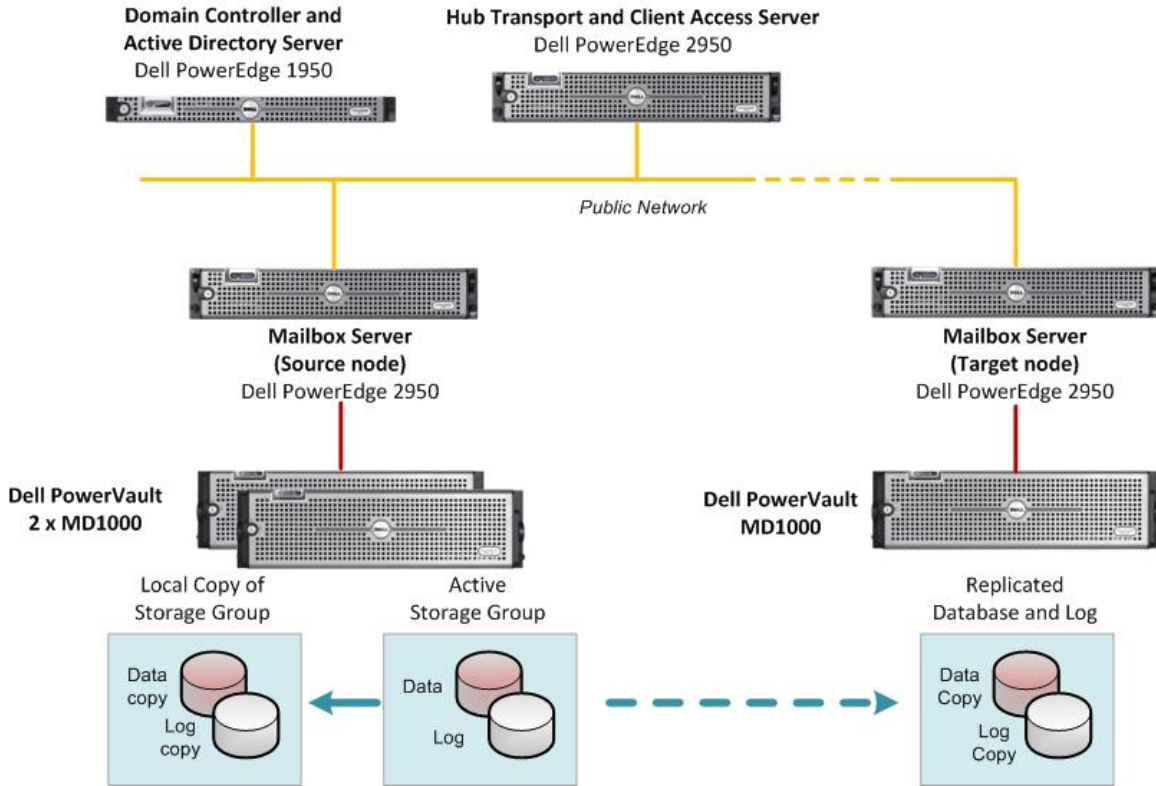


Figure 8: SCR Replication with LCR enabled Mailbox server

SCR replication with SCC Mailbox server

An SCC deployment uses Microsoft cluster service (MSCS) deployed in a shared database model and providing high-availability to the mailbox server. Since the data is shared, there is only one copy of the database and logs. The SCR replication on the top of an SCC environment adds the data redundancy besides mailbox server availability provided by SCC. Figure 9 shows a simple one-to-one SCC configuration with SCR.

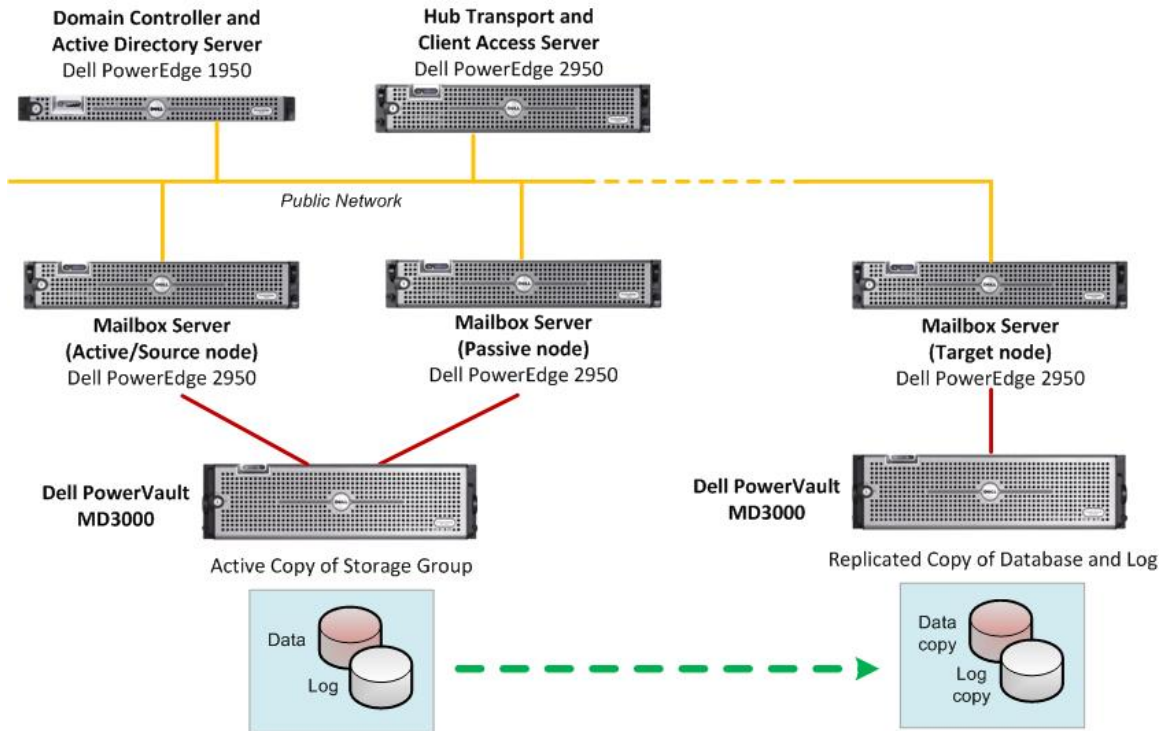


Figure 9: SCR Replication with SCC Mailbox server

SCR replication with CCR Mailbox server

The combination of SCR with CCR configuration provides enhanced availability and site resiliency at two tiers. The CCR provides a non-shared storage model with availability to both mailbox server and database but the replication at the cluster level is limited to the same active directory site. CCR is supported with cluster nodes hosted at different physical locations but not the same active directory logical site. The SCR configuration on top of it adds resiliency by copying data to a remote active directory site. Figure 10 shows a simple one-to-one CCR configuration with SCR. The same configuration can be changed to one-to-many by adding multiple SCR targets.

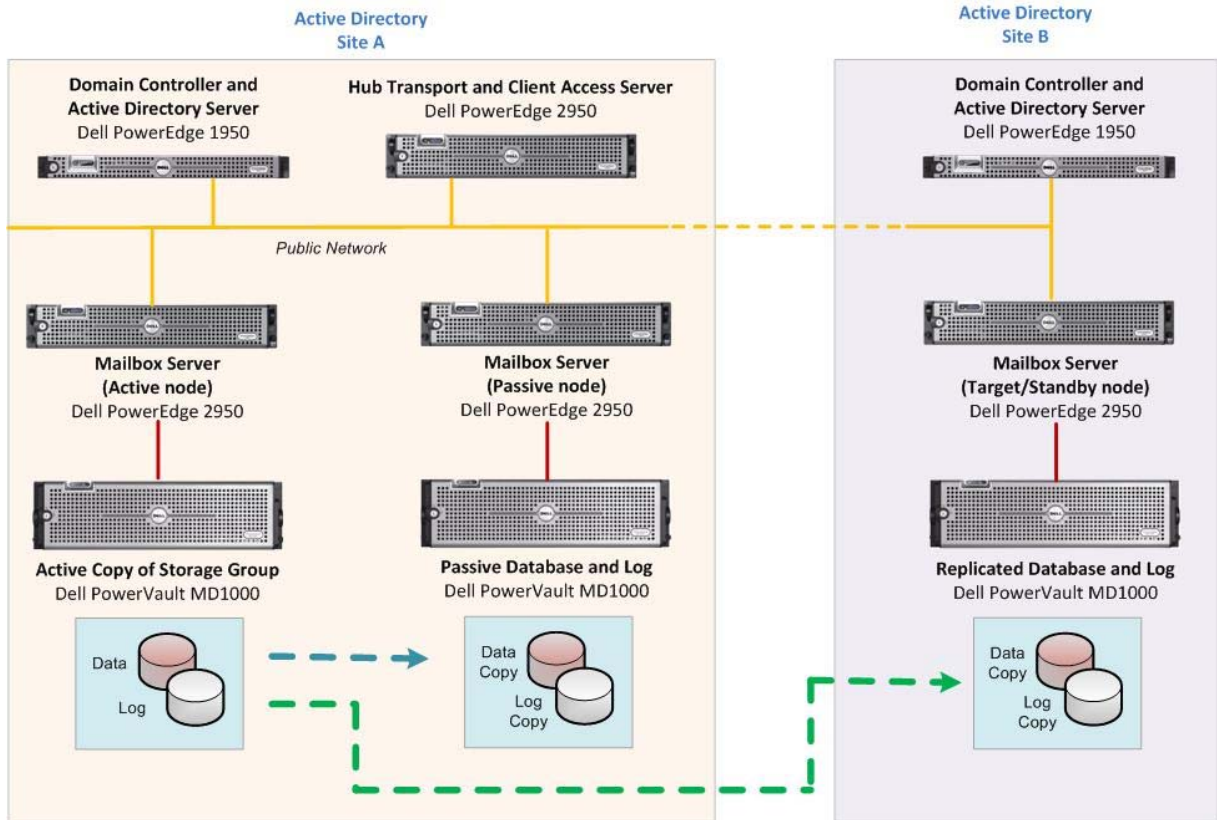


Figure 10: SCR Replication with CCR Mailbox server

Performance

A set of performance tests was conducted on a SCR configuration with Stand-alone mailbox servers (shown in Figure 6). The configuration details are given below. The objective was to understand the performance impacts of SCR on the source and target mailbox server nodes and on their associated storage subsystem. Simulations were conducted on the SCR configuration with two different settings for the passive replay lag time: The default value (24 hours lag) and the instantaneous replay (which includes 50 logs lag time). The utilization levels of various system resources were recorded, and their average values are displayed in following graphs.

Configuration Details:

- Mailbox servers (Source and Target)
- Dell PowerEdge 2950 with 2 x Dual Core Intel Xeon 5160, 3.00 GHz processors; 8 GB system RAM
- Windows Server 2003 R2 Enterprise x64 Edition with SP2; Exchange Server 2007 Enterprise Edition RTM; Exchange Server 2007 Enterprise Edition SP1
- Hub Transport/Client Access server
- Dell PowerEdge 2950 with 2 x Dual Core Intel Xeon X5160, 3.00 GHz processors; 8 GB system RAM

- Windows Server 2003 R2 Enterprise x64 Edition with SP2; Exchange Server 2007 Enterprise Edition RTM; Exchange Server 2007 Enterprise Edition SP1
- External mailbox storage
- 2 x Dell PowerVault MD1000 (each on source and target node)
- Database volume: RAID 10 with 10 x 146 GB 15K RPM SAS drives (MD1000)
- Log volume: RAID 1 with 2 x 73 GB 15K RPM SAS drives (MD1000)
- Database copy volume: RAID 10 with 10 x 146 GB 15K RPM SAS drives (MD1000)
- Log copy volume: RAID 1 with 2 x 73 GB 15K RPM SAS drives (MD1000)
- Microsoft LoadGen Simulation Tool
- Build version: 08.01.0177.000
- User profile: 1000 heavy users executing 94 tasks per 8 hour user day in Outlook® 2007 online mode

As shown in Figure 11, the database I/O levels are almost about the same for both tests with 24 hours replay lag and 50 logs replay lag time (i.e. instant replay). The log I/Os graph (in Figure 12) on the other hand shows a significant difference on the target nodes. The 50 logs replay lag compared to the 24 hours lag, recorded more than double the log disk reads on the target node. This is due to the logs being copied from the active node and replayed on the target node at the same time whereas in the 24 hour lag test, the copy phase and replay phase occur separately in a non-overlapping timeframe (24 hours apart). Thus the replay lag time should be carefully chosen to reduce the log I/O impact on the target side.

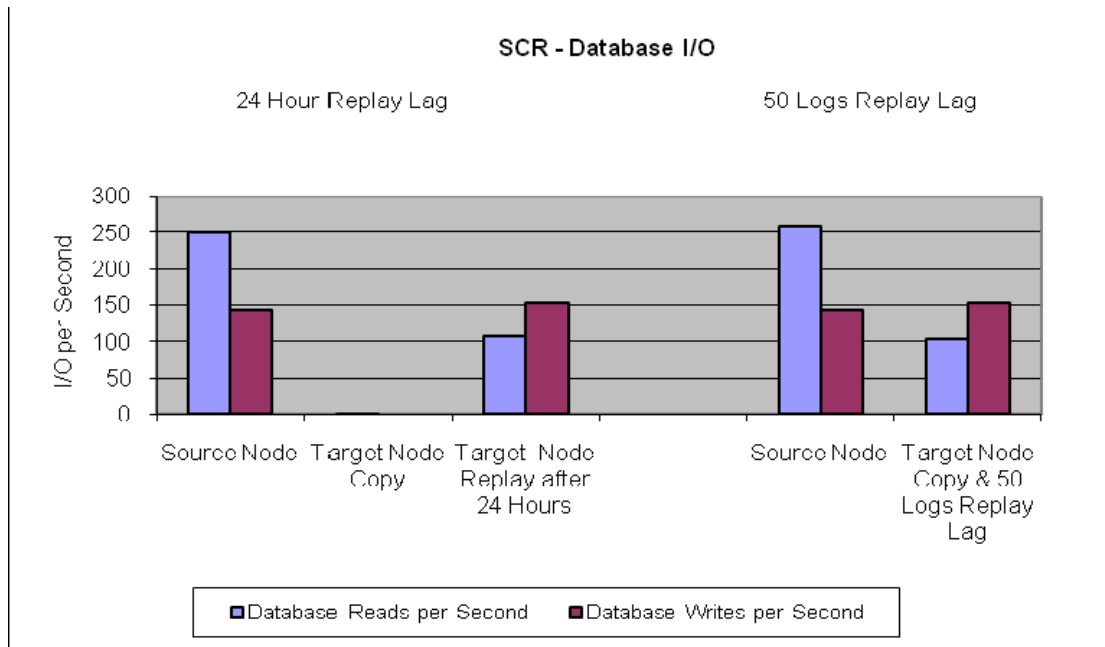


Figure 11: SCR Database I/Os comparison for 24 hour Replay lag and 50 Logs Replay lag

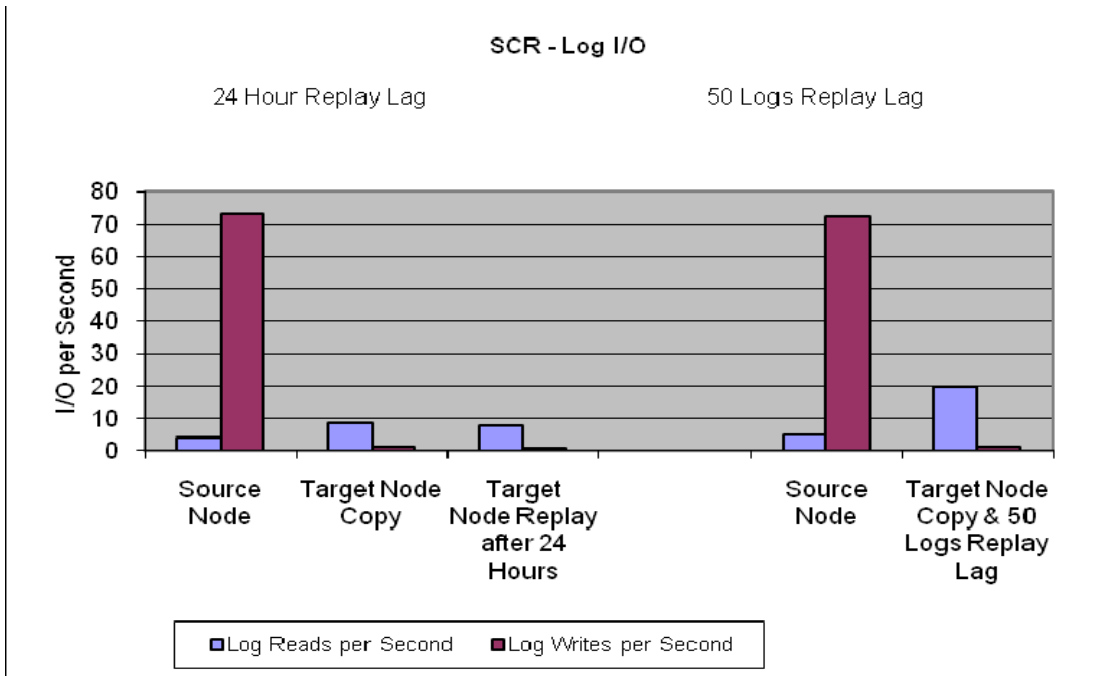


Figure 12: SCR Log I/Os comparison for 24 hour Replay lag and 50 Logs Replay lag

The memory utilizations of mailbox servers between 24 hour lag test and instant replay lag test do not show any significant difference (illustrated in Figure 13). The target node utilizes less memory during copy phase with the 24 hour lag test. During the replay phase it uses almost same amount of memory as 50 logs replay lag test. Similar observation holds with the processor utilization as illustrated in figure 14.

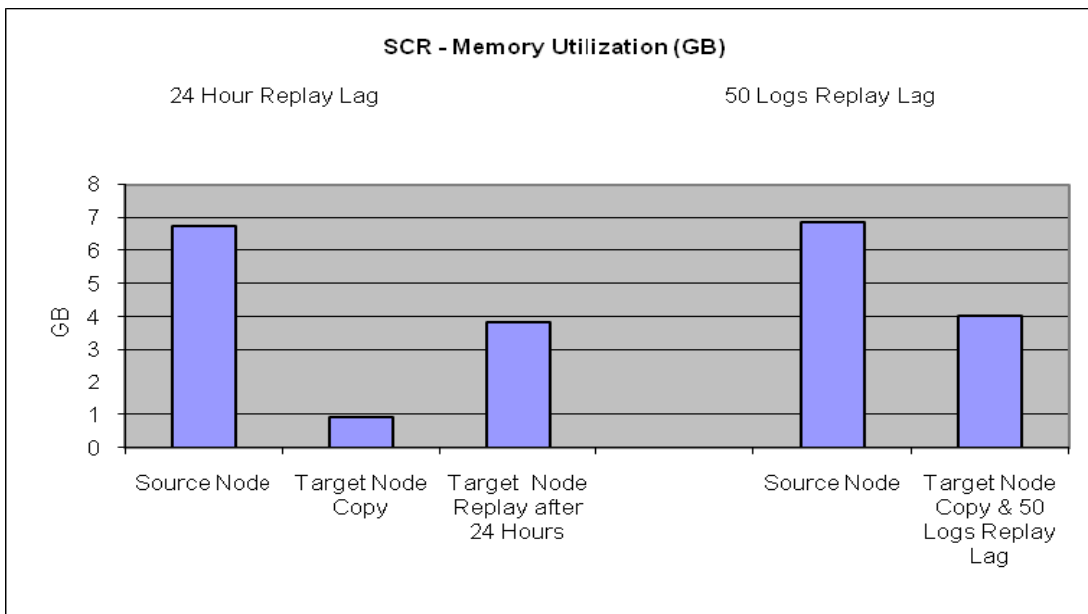


Figure 13: SCR Memory Utilization comparison for 24 hour Replay lag and 50 Logs Replay lag

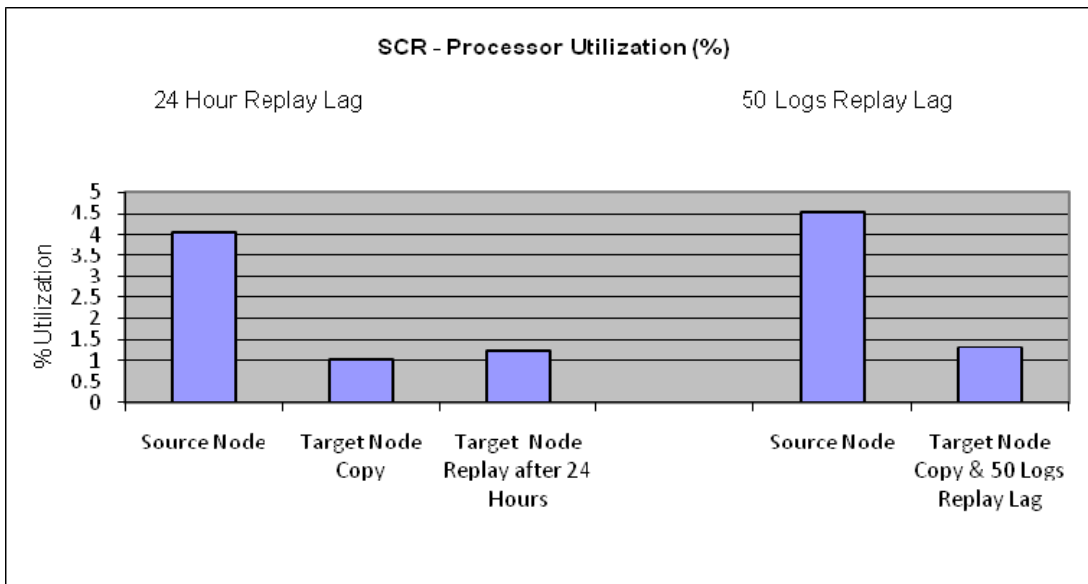


Figure 14: SCR Processor Utilization comparison for 24 hour Replay lag and 50 Logs Replay lag

Even though during SCR activity the target node underutilizes the resources, it is recommended that the target be provided with same set of memory, processor and disk resources as the source node. This is because, after recovery, the target node is required to be operated as the source node serving the actual production user load.

Conclusion

Microsoft Exchange Server 2007 SP1 provides new disaster recovery features and availability options that help businesses to effectively meet their site-resiliency requirements and protect their messaging systems against failures. Administrators should carefully evaluate the level and type of availability required before deciding which option is most appropriate for their environment. The choice of a particular option will depend on various factors such as cost, downtime, geographical redundancy, recoverability, scalability, and manageability. Implementing the required availability and disaster recovery features, and configuring them for optimal performance can help create flexible, highly available systems in enterprise data centers.

Dell PowerEdge servers, Dell PowerVault storage, and Dell/EMC storage provide standard hardware platforms for seamlessly deploying Exchange Server 2007 with the required availability features. More information can be obtained at www.dell.com/exchange. Dell Services include assessment, design, and implementation tailored for those messaging deployments. Dell also offers end-to-end Exchange messaging solutions that include partner offerings for security, archiving, backup, and recovery. More information can be obtained at www.dell.com/secureexchange.

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

© Dell Inc. 2008. All rights reserved. Reproduction in any manner whatsoever without the express written permission of Dell Inc. is strictly forbidden. For more information, contact Dell. Information in this document is subject to change without notice.