NETAPP TECHNICAL REPORT

Microsoft Exchange 2007 SP1 Continuous Replication Best Practices Guide

Robert Quimbey, NetApp
March 2008 | TR-3651

TABLE OF CONTENTS

1 INTRODUCTION

This is a follow-up to the NetApp document Microsoft Exchange 2007 Continuous Replication Best Practices Guide. Microsoft® Exchange Server 2007 SP1 includes new features that increase availability, decrease disk I/O workload, and improve transport performance. Messaging is a mission-critical service that consumes an ever-increasing percentage of an information worker's day. Messaging availability can be reduced by logical corruption in the database and by server, network, storage, and site failures. This document explores the new features of Exchange Server 2007 SP1 that impact storage and provides guidance in the storage design when using continuous replication features with NetApp® storage systems. This guidance applies to all three flavors of continuous replication: local continuous replication (LCR), clustered continuous replication (CCR), and standby continuous replication (SCR).

See the NetApp document Microsoft Exchange Server 2007 Best Practices Guide for general Exchange Server 2007 best practices, and see the Microsoft document Technical Architecture of Exchange Server 2007 for in-depth Exchange 2007 information.

Intended audience

This technical report is intended for messaging and storage professionals who design, test, deploy, and manage their corporate messaging infrastructure. For methods and procedures mentioned in this technical report, it is assumed that the reader has working knowledge of the following:

- Microsoft Exchange 2007 architecture
- Exchange storage architecture and administration
- NetApp Data ONTAP®

2 SERVICE PACK 1 (SP1) OVERVIEW

Unlike past Exchange service packs, which focused solely on bug fixes, Exchange 2007 SP1 includes dozens of new features and many significant performance improvements. SP1 is a full product that can be directly installed on the target server, without first needing the RTM installation.

2.1 WINDOWS SERVER 2008

Exchange 2007 SP1 supports Windows® Server 2008; however, upgrading an existing Exchange 2007 server running Windows 2003 to Windows 2008 is not supported. The only supported method to deploy Exchange 2007 SP1 on Windows Server 2008 is to install it on a Windows Server 2008 system that does not have Exchange installed. Microsoft documentation for moving Exchange 2007 on Windows 2003 to a server running Windows 2008 will be posted on TechNet. Formatting the system drive on an Exchange 2007 RTM installation on a Windows 2003 server and then installing Windows Server 2008 running Exchange 2007 SP1 setup using the /recoverserver setup switch is not supported, nor is using a Windows 2008 SCR target to migrate, since both the source and target operating system must be the same when using any flavor of continuous replication. Supported methods for upgrading to Windows Server 2008 are described in Microsoft's Exchange blog.

One of the most talked about features for Windows Server 2008 is multisubnet clustering, which is supported for both single copy clusters (SCCs) and clustered

continuous replication (CCR). Multisubnet clustering eases the configuration of NetApp's MetroCluster, which uses synchronous replication to a disaster recovery site, providing virtually zero data loss and instant availability. Windows Server 2008 also brings DHCP (IPv4) to clusters and IPv6, though the restrictions and configuration are very precise. See Microsoft's document on IPv6 Support in Exchange 2007 SP1.

The recommended quorum configurations include:

- Node and file share majority is recommended for CCR.
- Node majority is recommended for SCC with an odd number of nodes.
- Node and disk majority is recommended for SCC with an even number of nodes.

2.2 TRANSPORT FEATURES (HUB AND EDGE)

BACK PRESSURE

Back pressure is a system resource monitoring feature on transport servers that monitors important system resources. When default thresholds are exceeded, new connections are rejected until system resource utilization falls below those thresholds. Algorithm improvements were made in SP1, and the free disk space requirement was reduced from 4GB to 500MB, which is important for transport sizing. For more information, see Microsoft's document Understanding Back Pressure.

PRIORITY QUEUING

The categorizer will now act on the message priority that users set, whereas in previous Exchange versions it did not. Messages that are flagged with high importance will be queued before those with normal importance, which will be queued before those with low importance. High-importance messages that exceed the MaxHighPriorityMessageSize threshold (250kB) will be set to normal importance. For more information, see Microsoft's document Understanding Priority Queuing in Exchange 2007 SP1.

TRANSPORT DUMPSTER

The hub transport servers in an organization can send any mail lost during a continuous replication failure from the hub transport dumpster. In Exchange 2007 SP1, LCR-enabled storage groups join CCR-enabled storage groups in being able to utilize this dumpster. In most cases, the only data at risk is client activity that does not go through transport, such as changing a message property, tasks, or many calendaring tasks. The hub transport servers cannot be clustered, so in a site disaster, hub transport servers that are lost are not available to recover mail from, and that mail might be lost. For more information, see the Microsoft document Working with the Queue Database on Transport Servers. NetApp has both a host-based replication solution in ReplicatorX™ and a storage-based replication solution in SnapMirror® to replicate the hub transport server data; see the NetApp document Exchange 2007 Disaster Recovery Model Using NetApp Solutions.

BACKUP AND RESTORE

Exchange does not provide a mechanism to back up a hub transport server. One strategy is to utilize VSS to back up the hub transport database and transaction logs. The mailbox server role must be installed on the hub transport server, and the default database must be placed on the same NetApp LUN as the hub transport database. Next, use SnapManager® 4.0 for Exchange to perform a VSS backup of

the mailbox database. The VSS writer will quiesce the mailbox databases and will create a Snapshot™ copy of the entire LUN. Proper freezing and quiescing of the hub transport database cannot occur. The best solution is to stop the transport service and then create the backup.

2.3 MAILBOX SERVER ROLE

ONLINE DEFRAGMENTATION

Online defragmentation was seldom monitored because it was difficult to do so prior to Exchange 2007 SP1. In SP1, detailed information is presented in the event viewer as a single event: 703. This event shows how long the defragmentation pass took and how often the database has been fully defragmented since the installation of SP1.

This is very important information that can be used to adjust the online maintenance window. Microsoft recommends that defragmentation be completed in 14 days, so if the defragmentation is completing much sooner, shortening the online maintenance window can reduce the amount of disk churn and reduce the Snapshot space consumption on NetApp storage.

For more precision, two performance counters should be logged to measure the online defragmentation trend:

- MSExchange Database ==> Instances\Online Defrag Pages Freed/Sec
- MSExchange Database ==> Instances\Online Defrag Pages Read/sec.

Reduce the online maintenance window if the read-to-freed ratio is greater than 100 to 1 and increase the online maintenance window if the read-to-freed ratio is less than 50 to 1. For example, if the read counter is averaging 600, and the freed counter is averaging 5, the ratio would be 120 to 1, and the recommendation would be to reduce the online maintenance window. These counters can be enabled with the registry key "Show Advanced Counters."

ONLINE MAINTENANCE DATABASE SCANNING, AND PAGE ZEROING

Prior to Exchange 2007 SP1, the only way to ensure a live database was not corrupt, and to zero deleted pages in the database without taking the database offline, is to perform an online streaming backup. During the streaming online backup, each page in the database is read, and the checksum integrity is verified. Page zeroing only occurs during streaming online backup, as the setting is ignored for a VSS backup.

Exchange 2007 SP1 forces page zeroing to create log files, propagating the page zeroing to all continuous replication targets.

When enabled in the registry (Online Maintenance Database Scanning in Exchange 2007 SP1), database scanning and page zeroing will run during the online maintenance window at the server level. Online database scanning addresses the CCR scenario where the backup is performed on the passive, and the active database is never completely scanned. Online maintenance will split the maintenance window effectively in half, between online defragmentation and database scanning. Databases should have page zeroing enabled when created; otherwise, the first time it is run against the database, the disk I/O workload (sequential read) will severely impact performance. Throttling can be enabled to reduce the impact the first time page zeroing is run.

2.4 HIGH-AVAILABILITY FEATURES

STANDBY CONTINUOUS REPLICATION

SCR is a new feature introduced in Exchange 2007 SP1, and an SCR target can be configured for any source mailbox server role except LCR. The server that is configured as an SCR target must have the mailbox server role installed and cannot have any LCR-enabled databases. The SCR server can be in a clustered configuration; however, no configured clustered mailbox servers can be on that cluster (only SCR targets). SCR storage groups (like LCR and CCR) cannot have more than one database.

All source mailbox server roles (except LCR) can have multiple SCR targets for each storage group, though the practical (and tested by Microsoft) limit is four SCR targets per source. An SCR server can be configured as the SCR target for storage groups from multiple (non-LCR) source servers.

Microsoft does not provide a backup mechanism for SCR. However, one solution to this is to stop the replication service and to create a NetApp VSS Snapshot copy of the volumes housing the Exchange data.

EXCHANGE PATH

With the new continuous replication features (LCR/CCR/SCR), the dependency on the database and log path between servers is becoming more important. A few moments of careful consideration can save an enormous amount of pain in the future. The path of the exchange binaries, where you choose to install Exchange 2007, is also important for current and future SCR scenarios where that path must be the same between servers.

With CCR and SCR, the target Exchange transaction logs and databases that should be housed on separate storage from the active must use the same path as the active. This can cause frustration if care isn't used in the design of the active path, particularly with SCR if the SCR target will host databases from multiple sources or active servers. One strategy is to use mountpoints and ensure each database path and storage group name is original.

For example, select an Exchange data drive (P:) and create a separate subfolder for databases and logs (P:\log and P:\db). In our example, we have two mailbox servers, a standalone mailbox server (MBX1) and a CCR cluster (CCR1). By naming the mountpoint under log and db uniquely, you can create an SCR target from any source; only having to ensure the drive letter (P:) is available on the SCR
server.

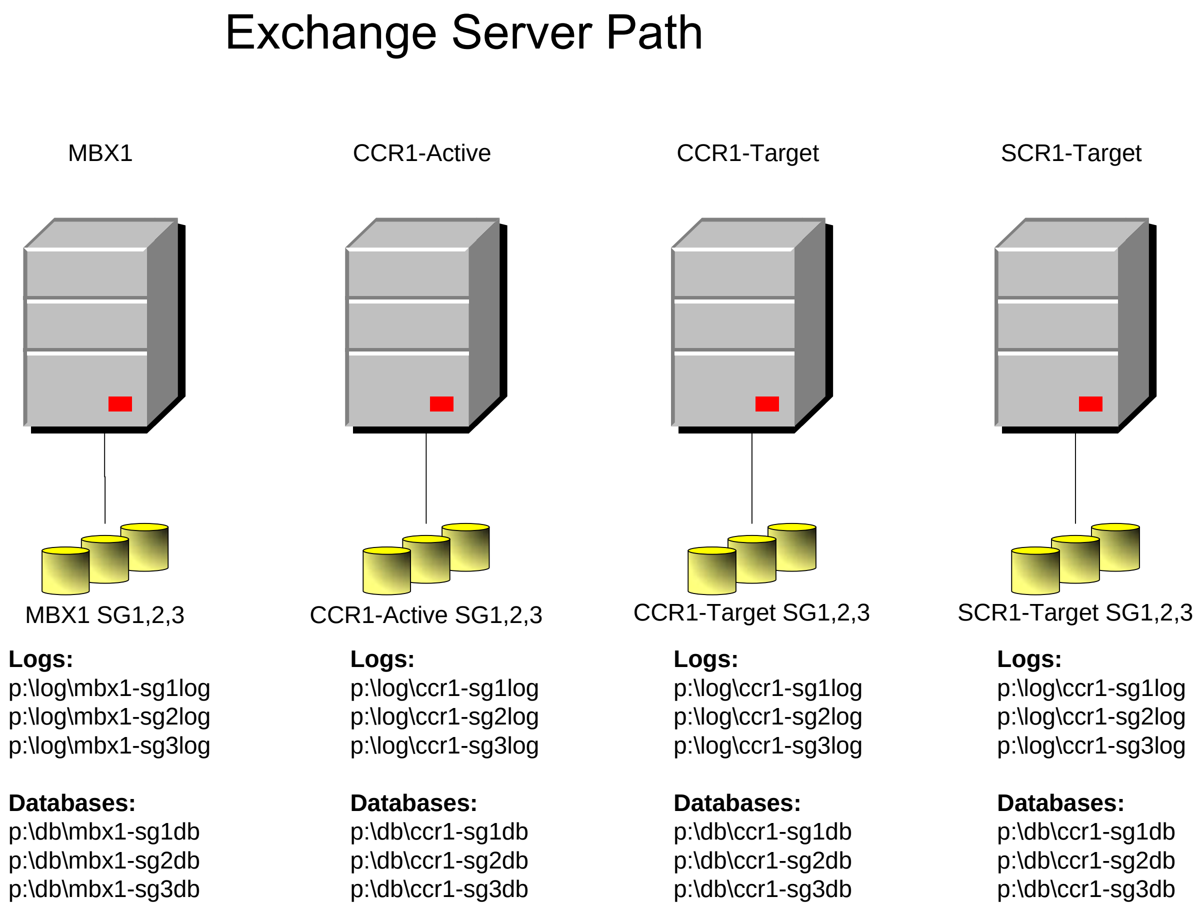

Figure 1) Exchange Server Path

It might seem redundant to include the hostname in the path; however, if you decide to use SCR in the future, it helps the administrator identify the storage group and source server instantly, eliminates path conflicts, and reduces the risk of error when creating the mountpoints. Each subfolder (mbx1-sg1log) is a mountpoint with a separate FlexVol® volume behind it.

SNAPDRIVE

For optimal performance and manageability ease, NetApp SnapDrive® should be used to create LUNs to properly align the primary partition. Using diskpart to manually align and configure a LUN is not recommended.

BACKUP WITH NETAPP SNAPMANAGER 4.0 FOR EXCHANGE

SnapManager for Exchange 4.0 is LCR and CCR aware and is integrated with many new Exchange 2007 features, such as up to 50 storage groups and Powershell integration. Powershell is the scripting interface that Exchange 2007 uses to automate tasks. SnapManager for Exchange 4.0 is tightly integrated with Powershell, and a command-line reference can be found in the NetApp document SnapManager for Exchange 4.0 Installation and Administration Guide. By default, with LCR, running a VSS backup on the target affects the production LCR server CPU when running an integrity check. With SnapManager for Exchange 4.0, the integrity check can be

offloaded to a remote verification server. With CCR, the administrator has the flexibility to back up the active node, the target node, or both. With CCR, rather than enabling online database scanning, which will cause a large disk I/O impact every time online maintenance is run, many NetApp customers choose to periodically perform a copy backup and checksum integrity to validate database health. This can be scheduled more robustly than the online database scanning.

Unlike many software-based VSS backups, SnapManager for Exchange 4.0 can create more than 250 Snapshot copies and keep them on disk with negligible performance impact for very quick recovery (minutes). Other vendors' solutions take a VSS backup, stream it off to tape, and then destroy the volume shadow copy, as keeping more than a couple of shadow copies adversely affects performance on their storage systems. When these tape solutions must be restored, it will be from tape, a lengthy and painful process. See the NetApp document SnapManager 4.0 for Microsoft Exchange for more information.

CONTINUOUS REPLICATION DISK I/O

Introducing continuous replication to a production Exchange Server slightly changes the active log workload, because those log files must be read when copied to the target. More interesting is that the target LUN workload varies significantly from the active LUNs. LoadGen 2007 was used to simulate 3,000 Outlook® 2007 online-mode users with 250MB mailboxes. For scenario details, see "Server and Storage Configuration" later in this technical report.

DATABASE I/O

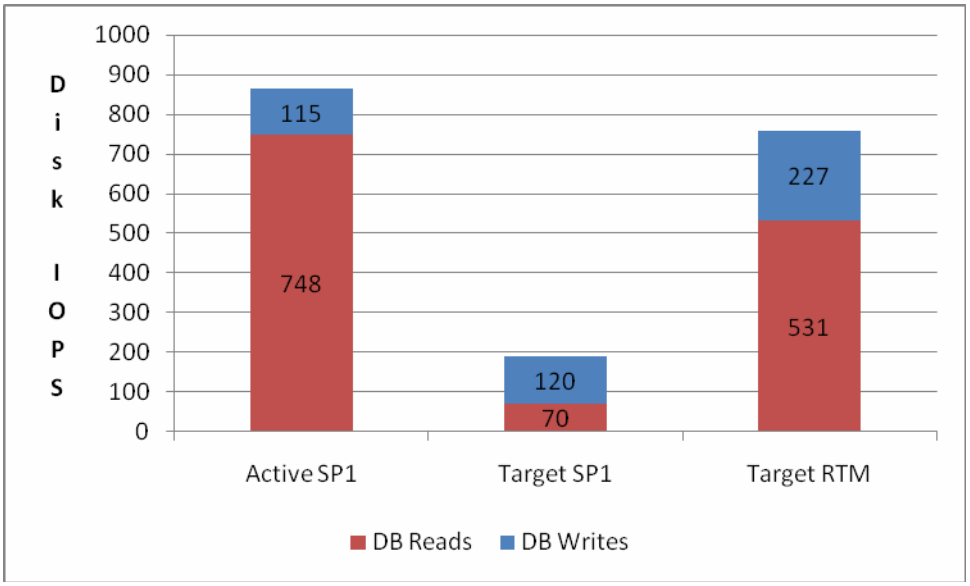

Figure 2) Active versus target DB I/O.

In Exchange 2007, the target I/O was higher than the active. In Exchange 2007 SP1 the target database I/O is much lower (–78%) than the active node. The target I/O from Exchange 2007 is included in the graph to show how significantly the I/O workload has changed in SP1. This huge improvement in the target database I/O enables the storage administrator to combine the target databases from multiple CCR clusters in the same disk aggregate. Without SP1, the storage administrator had to isolate the target storage for each CCR target node in a separate aggregate.

In production, with some user configurations, the target DB I/O can be two to three times higher than measured with LoadGen, yet with SP1 it still is less than the active DB I/O, allowing for the storage design change, removing the isolation requirement.

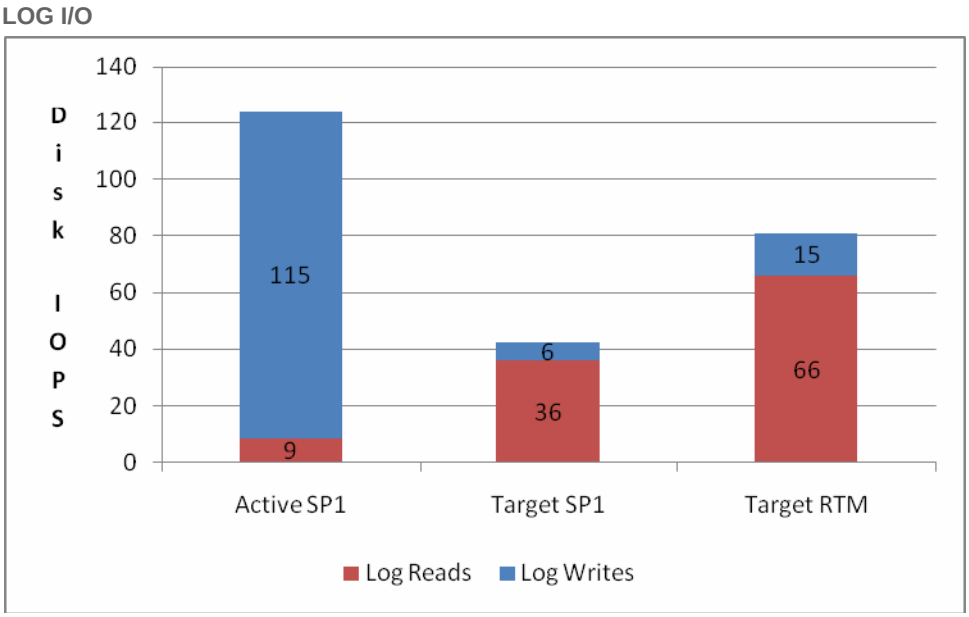

Figure 3) Active versus target log I/O.

In Exchange 2007 the target log I/O is less than the active log I/O. In Exchange 2007 SP1 the target log I/O is 64% less than the active log I/O, a big improvement. In SP1, the amount of data being read on the log LUN is almost half the behavior measured in Exchange 2007.

3 CONTINUOUS REPLICATION LUN DESIGN

Creating LUNs that provide adequate performance is only half of the solution. The LUNs must be provisioned with enough capacity, as well. For detailed capacity planning information, see the Microsoft Exchange document Planning Disk Storage.

The first best practice is to ensure that the transaction logs and databases are in separate aggregates. From both a performance (mixing workloads) and reliability perspective, do not place the transaction logs and databases on the same physical disks. In some cases, mixing workloads will not impact the performance of the storage. From a reliability perspective, do not place the transaction logs and database files from the same storage group on the same physical disks. Proper testing should be run to validate any configuration. The general Microsoft Exchange best practice is that similar workloads can share spindles, even between Exchange Servers.

Another best practice when creating volumes inside the aggregate is to use NetApp FlexVol volumes and to create a separate flexible volume for each storage group. FlexVol volumes are equally spread across every disk in the aggregate, and they are the layer where Snapshot copies are created. When a separate FlexVol volume is created for each storage group inside the database aggregate and the log aggregate, the VSS restores do not affect any other storage group.

Due to the target I/O reduction in SP1, it is no longer necessary to isolate the physical disks for multiple CCR clusters. For example, if there are two CCR clusters, create one log aggregate for all active nodes and one database aggregate for all active nodes per sizing recommendations. Then, create one log aggregate for all target nodes and one database aggregate for all target nodes. Figure 3 illustrates the disk configuration for the active nodes on two separate CCR clusters. Notice that each storage group has its own FlexVol volume. This same configuration should be repeated for the target nodes at the DR site on the DR storage.

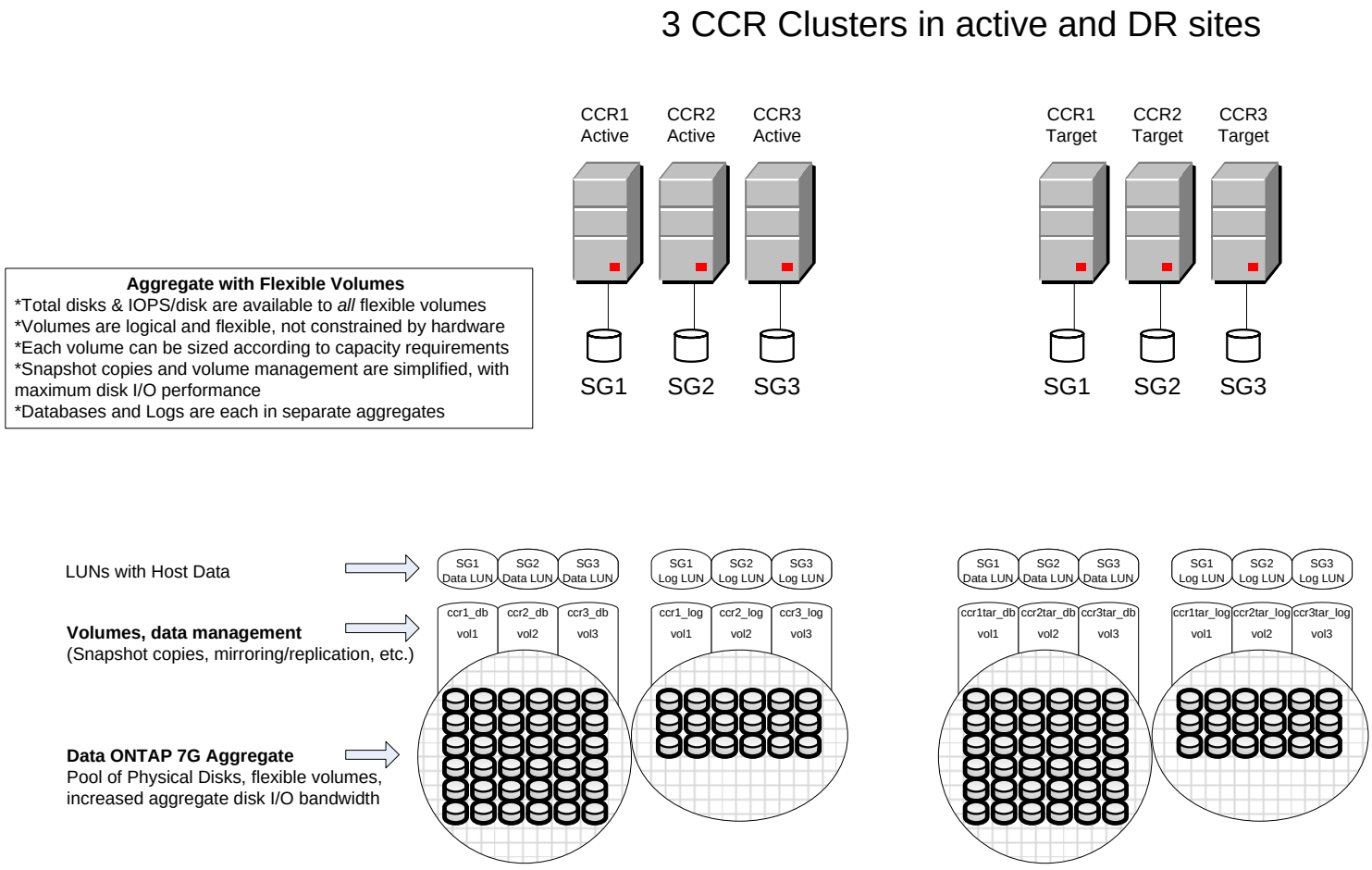

Figure 4) Disk configuration for active nodes on two separate CCR clusters.

The workloads are different from the active LUNs on the target LUNs. The good news is that if the active and target nodes are provisioned with the same capacity and performance, the target will keep up.

4 CONTINUOUS REPLICATION BEST PRACTICES

- Isolate active and target servers on separate storage arrays.
- Separate logs and databases in their own aggregates.
- Create a separate NetApp FlexVol volume for each storage group.

- Use NetApp RAID-DP® for superior performance and protection. For more information on RAID-DP, see Using NETAPP RAID-DP in Exchange Server 2007 Storage Designs.
- Run SnapManager for Exchange backups on the target node and adjust the online maintenance window on the active node.
- Consider using a copy backup (which does not truncate logs) and checksum integrity to validate database health versus online database scanning.
- Reduce the online maintenance window if it is completing a full pass within two weeks and the read-to-freed ratio is greater than 100 to 1.
- Be sure to plan for enough capacity and performance; see the Microsoft document Planning Disk Storage.
- Provision the active and target LUNs identically with regard to capacity and performance.
- Consider NetApp ReplicatorX and/or SnapMirror to achieve an RPO of less than five minutes to replicate Exchange databases, logs, and hub transport data.

5 SERVER AND STORAGE CONFIGURATION

The CCR cluster was configured so that the nodes were isolated from each other in the following configuration:

- Active and passive CCR nodes were HP DL140G3s with 16GB RAM each.
- Each node was connected using iSCSI with Gigabit Ethernet to a separate FAS3050 controller.
- 3,000 Outlook online-mode heavy users were split between two LoadGen 2007 clients.
- There were six storage groups, with 500 users each. DB and log LUNs were presented as mountpoints.
- Each FAS3050 (active and target) had two aggregates, one log and one DB, using NetApp RAID-DP:
    - Log aggregate: five 144GB FC disks (three data, two parity)
    - DB aggregate: 20 144GB FC disks (18 data, two parity)
- Each aggregate had six NetApp FlexVol volumes, one for each storage group.
- Each database was ~134GB in size, with a 30GB content index.

6 SUMMARY

Microsoft Exchange is a business-critical application, and there are many solutions available to increase availability. It is important to spend the time to identify the service-level agreement and to then utilize technologies in the storage design to meet those goals. Once the solution is designed, it is critical to test it (see Jetstress and LoadGen in Tools for Exchange Server 2007) and, once deployed, to monitor both the Exchange Server and the NetApp storage.

NetApp has proven data protection and disaster recovery tools for Microsoft Exchange Server. SnapManager for Exchange backup and restore capabilities, combined with SnapMirror technologies, provide a solid and robust solution for protecting and recovering Exchange data while meeting stringent RPO and RTO objectives.