White Paper

EMC[2] where information lives

EMC Backup and Recovery for Microsoft Exchange 2010 Enabled by EMC Unified Storage, EMC Data Domain, EMC NetWorker Plug-in for Exchange 2010, and Microsoft Hyper-V

A Detailed Review

EMC Information Infrastructure Solutions

Abstract

This white paper describes the design, deployment, and validation of Microsoft Exchange Server 2010 in a Microsoft Hyper-V virtualized application environment incorporating EMC® NetWorker® and EMC NetWorker Plug-in for Exchange 2010 as backup technologies and EMC Data Domain® as the deduplication technology on an EMC CLARiiON® CX4-480.

March 2011

Part number: H8047.2

Table of Contents

Executive summary

Business case

Microsoft Exchange Server 2010 introduces a new feature called database availability groups (DAGs), which provides high availability (HA) and site resiliency for Exchange-specific server and storage components. DAG replaces previously existing Exchange and Windows failover clustering-based technologies used for HA and site resiliency, such as single copy clusters (SCC), continuous cluster replication (CCR), and standby continuous replication (SCR).

Performing backup and recovery in Exchange 2010 with DAG copies is a new concept. Deduplication considerations with Exchange backups are also important in order to minimize disk space. EMC, with its rich portfolio of hardware, software, and partner offerings, is well positioned to offer a solution that combines recommendations and best practices for creating a robust backup and deduplication solution for Exchange 2010.

Product solution

This solution involves building a virtualized Exchange 2010 environment using DAG on an EMC® CLARiiON® platform. Additionally, the EMC NetWorker® family and EMC Data Domain® deduplication storage systems are leveraged to establish a backup and recovery environment for Exchange 2010 databases.

This solution does not provide a comprehensive guide to every aspect of an Exchange Server 2010 business-ready solution. Instead, it constructs a replica of a common customer environment and validates the environment for both performance and functionality.

Key results

This white paper shows the following benefits:

- **Reduced backup timeframes**: Backing up terabytes of data takes a long time with a traditional tape-based backup solution. For example, backing up 3.6 TB of data with LTO-5 (140 MB/s), the most advanced tape technology, requires 18 hours for conventional tape backup. However, it takes only 7 hours to back up the same amount of data through 1 Gb Ethernet in this solution, which means the backup time is reduced by up to 60 percent. Additionally, the Data Domain Boost software option improves the backup throughput at a rate of about 70 percent.
- **Reduced backup storage requirements**: Compared with the traditional 100 percent backup storage capacity requirement on tape, only a relatively small amount of storage capacity is required on Data Domain for full and incremental backups with data deduplication. In this solution, the total deduplication ratio is about 3.8:1 after two full backups and two incremental backups.

  Note Longer backup retention periods result in better deduplication ratios.
  Note Your mileage may vary depending on how many duplicates are in your specific environment.

- **Reduced physical infrastructure footprint**: By deploying the domain controller, Exchange servers, and the NetWorker server on a Microsoft Hyper-V virtualized platform, the number of physical servers required in this solution is significantly decreased, which leads to a higher server utilization rate and reduced energy and operational costs. In this solution, the entire environment only consists of three physical servers with Hyper-V enabled, while the conventional deployment without virtualization requires at least seven physical servers.

- **Validated and tested performance:** The test result shows that the designed architecture satisfies all recommended performance guidelines, as provided by Microsoft for Exchange 2010. For more details about these Microsoft recommendations, visit the Microsoft TechNet website: Performance and Scalability Counters and Thresholds

Introduction

Overview

This solution builds out and tests Exchange 2010 backup, recovery and deduplication functionalities with NetWorker 7.6 Service Pack 1 (SP1), the NetWorker Plug-in for Exchange, and Data Domain DD690.

This solution simulates multiple backup and recovery scenarios for a single mailbox, for a single item, and for the entire mailbox store. VSS snapshots are taken from DAG passive copies to offload the backup workload from the production environment. In addition, Data Domain Boost is utilized to add performance capabilities to the deduplication functionality.

The testing determines projected backup windows, throughput, and measure overhead on the mailbox server roles in both full and incremental backup scenarios.

Purpose

The purpose of this white paper is to:

- Demonstrate rapid and efficient backup and recovery in a multi-terabyte Exchange environment using VSS and backup deduplication technologies
- Validate the backup and recovery performance for Exchange 2010 by integrating Data Domain deduplication storage and NetWorker 7.6 SP1 with Data Domain Boost
- Validate the deduplication function and document the deduplication ratio of Data Domain, which makes an ideal long-term backup solution for Exchange
- Provide the design and architecture on Hyper-V virtualization deployment to reduce the physical footprint
- Determine best practices, including Data Domain and NetWorker design overview and considerations

Scope

The scope of this solution is bound by the following parameters:

- Build an Exchange 2010 environment with DAG configuration on a CLARiiON CX4-480 platform
- Use Windows Server 2008 R2 Hyper-V for virtualization
- Use the following EMC components to provide Exchange 2010 data protection and recovery:
  - EMC Data Domain DD690 with OS version 4.9 and Data Domain Boost
  - NetWorker Server 7.6 SP1 with NetWorker Plug-in for Exchange 2010 version 1.1
- Show the overall backup and recovery performance and deduplication ratio of Data Domain for Exchange 2010, leveraging the software VSS feature enabled by NetWorker Plug-in for Exchange
- Show backup performance improvement when using Data Domain Boost
- Show backup performance impact on the Exchange 2010 production environment under a specified user load

- Identify the steps of designing and implementing an Exchange 2010 server backup and recovery solution integrating EMC software and hardware

Audience

This white paper is intended for Microsoft Exchange database administrators and storage architects who are involved in planning, architecting, or administering an environment with EMC CLARiiON as the storage platform and also for those who are planning to implement backup and replication solutions.

Key components

Introduction

This section briefly describes the key solution components. For details on all the components that make up the environment, see the “Environment profile” section in this white paper.

EMC NetWorker and NetWorker Plug-in for Microsoft Exchange

EMC NetWorker Server provides a single, unified solution that not only protects Hyper-V and Windows Server 2008 R2, but also enables nondisruptive backup of applications such as Microsoft Exchange servers running on virtual machines.

NetWorker Plug-in for Exchange 2010 version 1.1 can support Microsoft VSS-based full backup and incremental backup for Exchange Server 2010 in both stand-alone and DAG configurations.

EMC Data Domain and Data Domain Boost

EMC Data Domain deduplication storage systems reduce the amount of disk storage by identifying redundant files and data as they are being stored and providing a storage footprint that is 10x-30x smaller, on average, than the original dataset. Data deduplication is performed on incoming data streams and allows only the new and changed segments of data to be identified and stored as unique instances within the Data Domain file system.

EMC Data Domain Boost significantly increases performance by distributing parts of the deduplication process to the backup server. Without Data Domain Boost, the backup server sends all data—unique or redundant—to an EMC Data Domain system for deduplication processing. With Data Domain Boost, only unique segments are sent to a Data Domain system.

EMC CLARiiON CX4-480

EMC CLARiiON CX4-480 is a versatile and cost-effective solution for organizations seeking an alternative to server-based storage. It delivers performance, scalability, and advanced data management features in one easy-to-use storage solution.

Microsoft Exchange Server 2010

Microsoft Exchange Server 2010 is designed to meet today’s communication and collaboration challenges. It provides advanced email and scheduling while delivering new methods of access for employees, greater productivity for IT administrators, and increased security and compliance capabilities for organizations.

Microsoft Windows Server 2008 R2 Hyper-V

Hyper-V, the Microsoft hypervisor-based server virtualization technology, provides the cost savings of virtualization that enables customers to make the best use of their server hardware investments by consolidating multiple server roles as separate virtual machines running on a single physical machine.

Windows Server 2008 R2 introduces a new version of Hyper-V. It includes several core improvements for creating dynamic virtual data centers, including increased availability and improved management for virtualized data centers, increased performance and hardware support for Hyper-V virtual machines, and improved virtual networking performance.

Physical architecture

Overview

The following section identifies and briefly describes the technologies and components that are used in the environment.

Physical environment

Figure 1 illustrates the overall physical architecture of the environment.

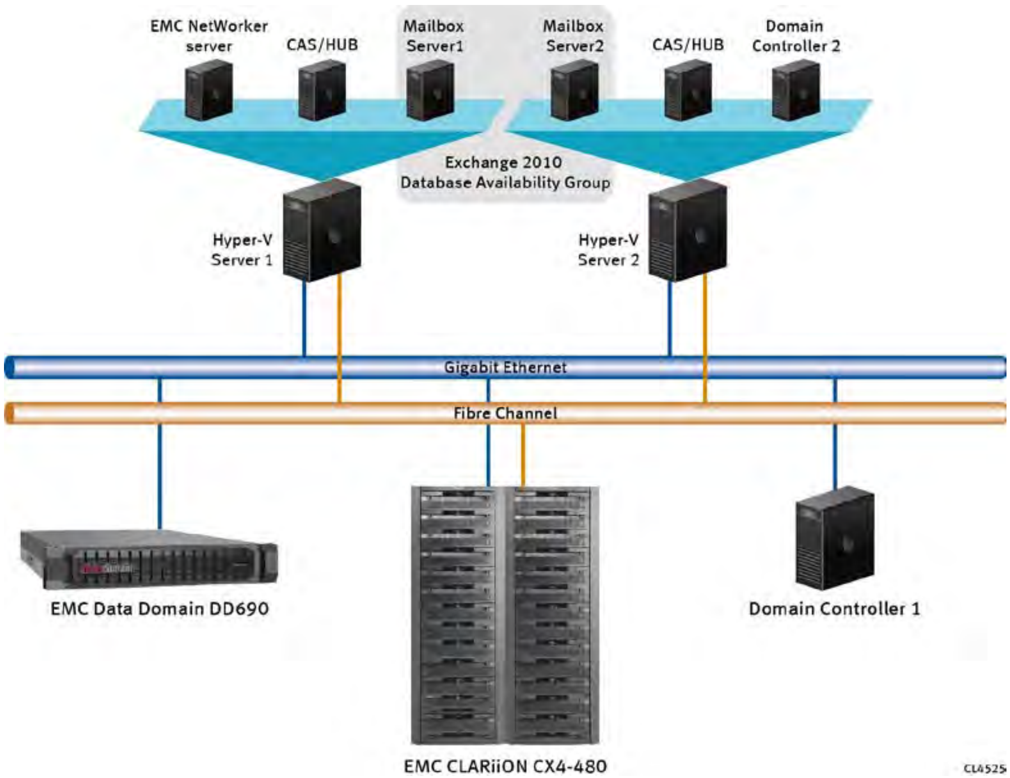

Figure 1. Overall physical architecture of the environment

Table 1 lists the hardware used to validate the solution.

Hardware resources

Table 1. Hardware resources

Equipment	Quantity	Configuration
CLARiiON CX4-480	1	24 SATA II disks (1 TB/7200 rpm)
SAN switch	1	4 Gb
IP switch	1	1 Gb
Data Domain DD690	1	16 TB raw devices

Hyper-V server, 24 cores, 128 GB RAM, 2 dual-port, 8 Gb/s HBA	1	The following virtual machines are deployed: One Exchange mailbox server (4 vCPUs/20 GB) One Exchange HUB/CAS server (4 vCPUs/8 GB) One domain controller server backup (4 vCPUs/4 GB)
Hyper-V server, 24 cores, 128 GB RAM, 2 dual-port, 8 Gb/s HBA	1	The following virtual machines are deployed: One Exchange mailbox server (4 vCPUs/20 GB) One Exchange HUB/CAS server (4 vCPUs/8 GB) One NetWorker server (4 vCPUs/4 GB)
Domain controller server, 8 cores, 32 GB RAM	1	One domain controller server

Table 2 lists the software used to validate this solution.

Software resources

Table 2. Software resources

Equipment	Quantity	Configuration
Windows Server 2008 R2 Enterprise Edition	11	Two Hyper-V servers Two Exchange mailbox servers Two Exchange HUB/CAS servers Two domain controller servers One NetWorker server Two LoadGen client virtual machines
Exchange 2010 Enterprise Edition	4	Two Exchange mailbox servers Two Exchange HUB/CAS servers
JetStress 2010	2	N/A
LoadGen 2010	2	N/A
NetWorker Server 7.6 SP1	1	N/A
NetWorker Plug-in for Exchange 2010 v 1.1	2	N/A
Data Domain OS 4.9	1	N/A
EMC PowerPath	2	N/A

Microsoft Exchange design

Exchange 2010 design in a virtualized environment

Introduction

The design and testing principles applied to this virtualized environment demonstrate how Exchange 2010 with large mailboxes can achieve high availability and meet performance metrics by using minimal resources. Testing is based on virtualized Exchange 2010 servers with two HA DAG copies implemented.

This solution is intended not only to meet the basic functionality requirements when deploying an efficient, repeatable backup and recovery design on a virtualized Exchange 2010 platform, but also to provide a solid foundation for future growth and development of the environment.

Storage design for Exchange database and log LUNs

Because many variables and factors vary from organization to organization, sizing and configuring storage for use with Exchange 2010 might be a complicated process. One of the methods used to simplify the sizing and configuration of storage for use with Exchange 2010 is to define a unit of measure – a building block.

What is a building block?

A building block represents the amount of disk and server resources required to support a specific number of Exchange 2010 users. The amount of required resources is derived from:

- A specific user profile type
- Mailbox size
- Disk requirements

Using the building block approach removes the guesswork and simplifies the implementation of the Exchange 2010 mailbox server. After the initial building block is designed, an organization can take this block of work and multiply it by a factor until the desired number of Microsoft Exchange server users (Microsoft Messaging API (MAPI) Outlook users) has been properly met or configured to satisfy the Microsoft Exchange Server recommended performance metrics.

EMC's best practices involving the building-block approach for Exchange Server design have proven to be very successful throughout many customer implementations.

Exchange 2010 Building block design process

The high-level building block design process for Exchange 2010 is similar to that used for Exchange 2007. Review Table 3 to understand the process flow used for developing and validating the test environment's storage design.

Table 3. Building block design process

Phase	Description
1	**Collect user requirements** In this phase, the Exchange administrator identifies: • Number of users • User I/O profile = Send/Receive and message size per user • Mailbox size • Deleted item retention • Concurrency • Replication required, number of DAG database copies • Backup/Restore requirements (RTO or RPO) • Third-party software that effects space or I/O (BlackBerry, antivirus software)
2	**Design the storage architecture based on user requirements** In this phase, the Exchange design is developed using the following tools: • EMC Exchange building-block methodology • EMC Exchange 2010 solutions • EMC and Microsoft best practices • Published Exchange 2010 Exchange Solution Review Program (ESRP) documentation: Microsoft Exchange Solution Reviewed Program (ESRP) – Storage v3.0
3	**Validate the design** In this phase, the design is validated with the following tools: • Jetstress is used to validate storage • LoadGen is used to validate storage and Exchange server performance

Applying the building block design process to Exchange 2010

The following sections walk you through the storage design process applied.

Phase 1—Collect user requirements

The user requirements used to validate both the building block storage design methodology and CLARiiON CX4-480 performance are detailed in Table 4.

Table 4. User requirements

Profile characteristic	Value
Number of users	3,000
Exchange 2010 profile	100 messages sent/received per day (0.10 IOPS per user)
Read/write ratio	3:2
Mailbox server	2 (virtual machines)
Number of DAG copies	2

User count per server	3,000 (1,500 active and 1,500 passive users)
Mailbox size	1 GB
Number of databases per server	10 (five active and five passive databases)
User count per database	300
RAID type	RAID1_0, 1 TB 7,200 rpm SATA II

As seen from the previous table, the test environment needs to support 3,000 users at 0.1 IOPS per user and a 1 GB mailbox quota.

Phase 2—Design the storage architecture based on user requirements

The following basic formula can be used to calculate storage requirements for the Exchange 2010 database and logs.

((IOPS * %R) + WP (IOPS * %W)) / Physical Disk Speed = Required Physical Disks

Where	**Is**
IOPS	The number of input/output operations per second
%R	The percentage of I/Os that are reads
%W	The percentage of I/Os that are writes
WP	The RAID write penalty multiplier (RAID 1_0=2, RAID 5=4)
Physical disk speed	For the CLARiiON CX4 series: • 130 for 10,000 rpm FC drives • 55 for 7,200 rpm SATA II drives

Important

Calculate the user IOPS for each building block first by using this formula: **(Users * IOPS per user) + 20% (overhead and spikes).** Then, add in the write penalty and divide by I/O per spindle for Exchange 2010.

Phase 3—Validate the design

For a complete summary of Jetstress and LoadGen findings, see the "Testing and validation" section.

Storage requirements calculated for the Exchange 2010 test environment

The calculations for the Exchange 2010 test environment are summarized as follows. These requirements are calculated using the formula detailed in the previous section.

IOPS Requirements

Calculations are based on the targeted user profile listed previously and on the availability of 1 TB 7,200 rpm SATA II drives on the CLARiiON CX4-480. It is essential to calculate IOPS first and then the capacity.

In this solution, each mailbox server hosts 1,500 active mailboxes and 1,500 passive mailboxes. In the event that one mailbox server is down, the other mailbox server will serve all 3,000 users. Therefore, each server must be able to handle the IOPS of 3,000 users.

3,000 users generate 360 IOPS (that is, 3,000 * 0.1 IOPS + 20% headroom). In RAID 1_0, it requires at least 10 (9.16 round up to 10) spindles to complete the tasks:

((360 * 0.6) + 2 * (360 * 0.4)) / 55 = 9.16

Capacity Requirements

Per Microsoft best practice, the database files and transaction logs of the active database copies and passive database copies should be located on different disk drives. Therefore, the capacity for active databases and passive databases needs to be calculated separately.

For the database size of the active database copies, at least 2,025 GB formatted capacity is required for 1,500 mailboxes (that is, 1,500 * 1 GB + 35% = 2,025 GB, where 35 percent reservation is for deleted items retention).

For the transaction log size of the active database copies, at least 210 GB formatted capacity is required (that is, 1,500 * 20 logs/per user/per day * 7 days retention = 210 GB). Therefore, the total space of the active database copies required based on capacity is 2,235 GB (2,025 GB + 210 GB).

Six 1 TB SATA II disks are grouped as one RAID 1_0 group (3+3), which provides 2.75 TB formatted capacity.

Similarly, six more 1 TB SATA II disks which group as one RAID 1_0 group (3+3) is required for the passive database copies.

In conclusion, twelve 1 TB SATA II disks, grouping as two RAID 1_0 groups (3+3), are required for each mailbox server.

Number of disks required

Based on the previous calculations, capacity requirements supersede IOPS requirements. In total, twelve 1 TB SATA II drives are grouped as two RAID 1_0 groups (3+3) to fulfill both IOPS and capacity requirements.

Building block design

The next step is to identify how many databases to configure per Exchange server. This involves determining how large the databases need to be. Based on the capacity of each RAID group in this solution, each database and log LUN is configured to support 300 users. Five databases and five log LUNs are accommodated per RAID group. Therefore, each RAID group will host 1,500 users in total.

In summary, a building block is created and provides all the necessary requirements for performance, capacity, and data protection to support 3,000 users. Table 5 summarizes the final building block created for this configuration.

Table 5. Building block information

Item	Description
Number of users supported	3,000
User profile supported	100 messages sent/received per day (0.10 IOPS per user)
Mailbox size	1 GB
Disk size and type	1 TB 7,200 rpm SATA II drives
RAID type	RAID 1_0
Database LUN size	500 GB
Log LUN size	50 GB
Total disks required	Two RAID 1_0 (3+3) groups – six disks for active database copies and six disks for passive database copies

Scaling the configuration up to 3,000 users with two DAG copies will require two of these building blocks (four RAID 1_0 groups with a total of 24 disks).

In this solution, RAID Group 1 and RAID Group 2, which host the active copy and passive copy of mailbox server 1, use the storage on Bus 1. RAID Group 3 and RAID Group 4, which host the active copy and passive copy of mailbox server 2, use the storage on Bus 2. By using this design, no matter which database copy is activated on each mailbox server, you can ensure that the two mailbox servers use different buses, and therefore the performance is optimized.

Figure 2. Storage configuration for Exchange

Exchange DAG configuration

High availability is provided in this solution with the use of DAG. Within a DAG, a set of mailbox servers performs continuous replication to provide automatic recovery in the event of failures.

In this environment, each database has two DAG copies deployed on two Exchange 2010 mailbox servers. Each Exchange 2010 mailbox server hosts five active database copies and five passive database copies.

In this way, the passive DAG copies will be used for backup, thus minimizing the performance impact to the production environment. For detailed information about Exchange server backup, refer to the “Backup and recovery design for Exchange” section of this white paper.

Figure 3 illustrates the database copy layout in this solution.

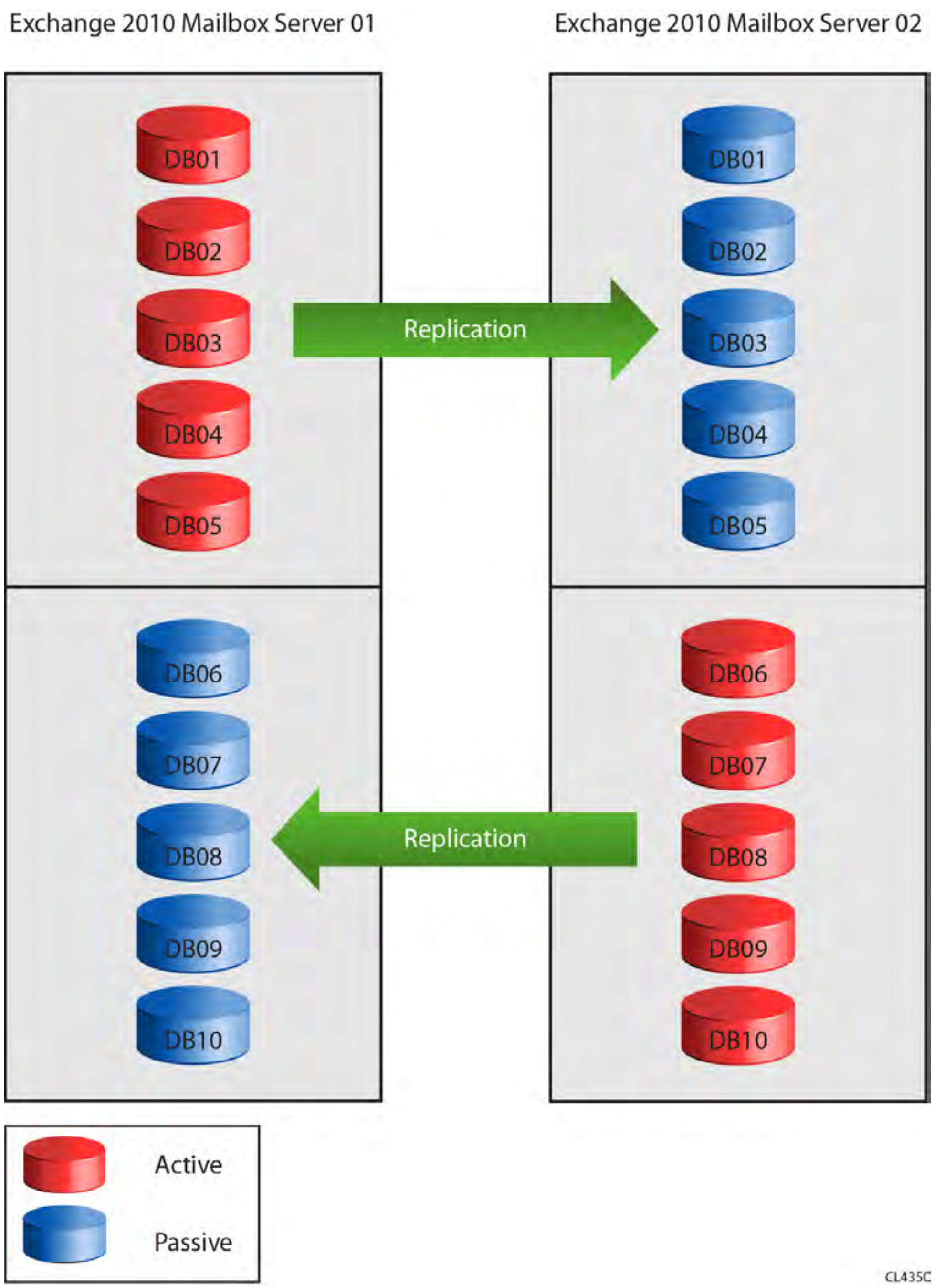

Figure 3. Database copy layout

Exchange virtualization resource allocation

The Exchange servers are deployed on virtual machines. The virtualization allocation of this solution is detailed in Table 6.

Table 6. Virtualization allocation

Server role	vCPU	Memory (GB)	Boot disk (GB)	Raw device mapping disk
Domain controller Server x 1	4	4	80	N/A
Exchange 2010 HUB/CAS Server x 2	4	8	80	N/A
Exchange 2010 mailbox server x 2	4	20	80	500 GB x 10 (Database LUNs) 50 GB x 10 (Log LUNs)
NetWorker Server	4	4	80	N/A

Note Microsoft provides detailed information on how to calculate memory and CPU requirements for Exchange Server 2010:

Understanding the Mailbox Database Cache

Mailbox Server Processor Capacity Planning

Backup and recovery design for Exchange

Overview

EMC NetWorker 7.6 SP1 provides integration with Data Domain deduplication storage systems, and also provides the support for Data Domain Boost. Therefore, backup data can be deduplicated on NetWorker storage nodes before it is sent for storage on a Data Domain system. This capability dramatically decreases the amount of data that is sent and stored on the NetWorker Data Domain device, and thus reduces the requirement for broad bandwidth to the storage device.

Backup design for Exchange 2010

Figure 4 illustrates the Exchange 2010 backup design used in this solution. The two Exchange 2010 mailbox servers are also the two NetWorker storage nodes, and they have access to the Data Domain system. The backup data (Microsoft VSS snapshots of DAG passive copies) is deduplicated on NetWorker storage nodes first, and then transferred to the Data Domain system, thereby reducing the total backup window distinctly.

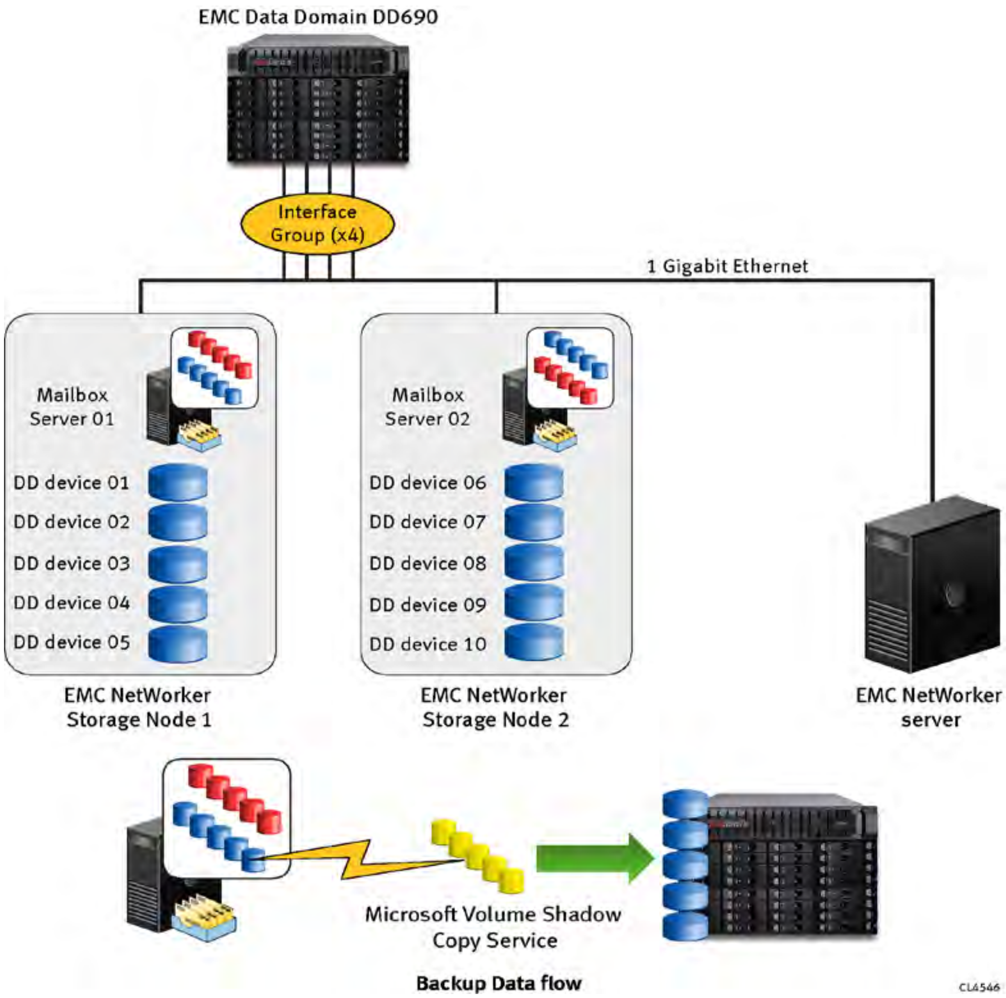

Figure 4. Exchange 2010 backup design

Table 7 lists the workflow of the topology for Exchange 2010 backup:

Table 7. Workflow of Exchange 2010 backup

Stage	Description
1	The NetWorker server initializes the backup request to the Exchange 2010 DAG passive copies.
2	Enabled by NetWorker Plug-in for Exchange 2010, the VSS writer uses the Microsoft Windows VSS framework with Exchange 2010 to create a software snapshot of the DAG passive copies.
3	By default, the backup solution invokes the system to perform internal Microsoft Exchange database consistency checks prior to the backup. This is an optional stage and can be disabled.
4	The NetWorker client (mailbox server) performs the backup. You do not have to transfer data to the NetWorker storage node because they are on the same server.
5	Data Domain Boost allows backup data to be deduplicated on NetWorker storage nodes before it is sent for storage on the Data Domain system.
6	The backup status is reported to the native Microsoft Exchange 2010 VSS writer.
7	After the backup is completed successfully, Exchange performs a transaction log truncation. If the backup fails, the logs are not truncated.
8	All backup steps are logged and reported in the savegroup completion report of the NetWorker server.

The NetWorker server, the NetWorker storage node (the Exchange 2010 mailbox server in this case), and the Data Domain system communicate through a 1 Gb LAN. By using Data Domain Boost, an efficient Exchange 2010 backup and recovery solution can be supported with limited Ethernet bandwidth.

Backup throughput optimization design and considerations

With the 1 Gb LAN, the following three optimization methods are used to improve the total backup throughputs in this solution.

More NetWorker storage nodes

For Data Domain Boost, using more NetWorker storage nodes means more offload capability for the Data Domain system when deduplicating backup data. In this solution, there are two storage nodes designed and each serves five Exchange 2010 database backups.

Place NetWorker storage nodes on backup clients

In this solution, a NetWorker storage node is installed to each backup client, therefore it is not necessary to transfer backup data from the backup client to the NetWorker storage node over the LAN. This design is strongly recommended for backup performance enhancement purposes.

An interface group that creates a private network within the Data Domain system

You can configure an interface group that creates a private network within the Data Domain system, comprised of the IP addresses designated as a group interface. The group interface uses the Advanced Load Balancing and Failover feature to improve data transfer performance and increase reliability. In this solution, four 1 Gb network interfaces form an interface group.

NetWorker client configuration

To obtain the optimized backup performance, consider the following methods:

- Balance the backup load by appropriately setting up parallelism in NetWorker
- Balance the data flow during parallel backup

To achieve these targets, configure the NetWorker client resource attributes. There are two NetWorker clients, and each of them backs up five Exchange 2010 database passive copies.

In this particular solution, a total of 10 Exchange 2010 mailbox databases need to be backed up, therefore the client parallelism value is set to 5 for each NetWorker client. This means that 10 databases can be backed up simultaneously to 10 Data Domain devices that are accessed evenly by two storage nodes. This design ensures that there is sufficient backup load. In the meantime, the backup data flow for each backup session can be appropriately balanced.

The following values are specified in the **Save Set** attribute of the client resource based on the description above.

For client mailbox 01:

```
APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 06

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 07

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 08

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 09

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 10
```

For client mailbox 02:

```
APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 01

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 02

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 03

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 04

APPLICATIONS:\Microsoft Exchange 2010\Mailbox Database 05
```

Data Domain design and configuration

Data Domain system overview

Data Domain systems are disk-based deduplication appliances and gateways that provide data protection and disaster recovery (DR) for organizations. The Data Domain operating system (DD OS) provides both a CLI for performing all system operations, and Enterprise Manager (a GUI) for configuration, management, and monitoring.

Data Domain device considerations

The NetWorker 7.6 SP1 software enables the integration of Data Domain storage systems with NetWorker software features and storage node processes. The integration uses special NetWorker Data Domain storage devices created by the NetWorker software. Each device represents a storage volume and appears as a folder in the Data Domain system.

Two best practices for NetWorker Data Domain devices usage are as follows:

- Use more NetWorker Data Domain devices to achieve better performance
- Use fewer streams on each NetWorker Data Domain device to achieve better performance by setting the appropriate **Target sessions** value

A NetWorker storage node supports up to 16 read/write NetWorker Data Domain devices and the corresponding 16 read-only devices.

Optimum performance is achieved when the number of concurrent save streams running on each device is minimized. It ensures that each different save set from the storage node attaches to a new device, if available, or to the device that is currently running the fewest save streams. To support this function, the **Target sessions** value for a device is set to a default value of one. The **Max sessions** value is set to a default value of four.

Data Domain deduplication ratio and efficiency

Deduplication is typically measured by the deduplication ratio. This ratio represents the reduction in storage space that results from the data deduplication and compression technology. It is possible to achieve a ratio of 20:1. However, even a ratio of 5:1 qualifies as a successful result.

A number of factors can contribute to the deduplication ratio, including retention periods, the type of data being backed up, change rate, the frequency of full backups, and the use of encryption and compression.

Other than these factors, for optimal use of storage space, customers also need to consider performing the periodic clearing of expired storage space and removal of unused target pools. Specifically, customers need to consider the following factors:

- Retaining data for longer periods of time increases the chance that common data already exists in storage. This results in greater storage savings and a better deduplication ratio.
- Backups of Exchange 2010 are known to contain redundant data and are good deduplication candidates.

- After the first full backup, the data change rate affects the deduplication ratio for these consecutive backups.
- Frequent full backups result in not only high deduplication ratios but also increased data processing on the NetWorker storage node. A daily full backup schedule sends a greater amount of data from the client to the storage node for processing than a weekly full backup schedule with daily incremental backup.
- Data compression and encryption of source data diminish with the ability of Data Domain to find duplicate data and are not recommended for deduplication.
- Run space reclamation weekly as per the default setting. This feature can be scheduled or run manually.
- Periodically review and remove unused pools if they are no longer relevant to the storage environment.

Network and memory considerations

Because NetWorker Data Domain devices are network-based and are supported on Ethernet IP networks, consider the network connectivity and capacity.

Note Currently, NetWorker Data Domain devices support only an IP network and do not support SAN (Fibre Channel) data transport.

It is recommended to use at least two 1 Gb network links for connectivity, one of which is dedicated to administration and the other to data backup. Customers can maximize throughput of the Data Domain system by using multiple or broader bandwidth connections. In environments where 10 Gb connectivity is not available or cost-prohibitive, two alternatives are available:

- Use a dedicated 1 Gb or 10 Gb connection from a storage node directly to the Data Domain system. This provides a private, robust data connection and avoids the latency and complexity of a shared Ethernet connection. However, a separate traditional Ethernet connection is also required for administration and NetWorker Management Console (NMC) access.
- Use two network connections aggregated together. This method will provide increased capacity and can offer some resiliency. The Data Domain system provides automatic Advanced Load Balancing and Link Failover for NIC connections.

It is recommended that you use at least 8 GB of RAM for a storage node.

Data Domain configuration

In this particular solution, Data Domain is integrated with NetWorker 7.6 SP1 and configured as NetWorker Data Domain devices connected to the NetWorker Storage Node. Other configuration details are as follows:

- 10 NetWorker Data Domain devices are created and evenly accessed by two NetWorker storage nodes
- Interface group enabled (four 1 Gb Ethernet links)

Testing and validation

Overview

This section describes the design validation and performance results of this solution. The backup and recovery features are validated in different scenarios, and the performance is measured for Exchange 2010. An EMC Data Domain DD690 appliance is used for data deduplication. Microsoft LoadGen 2010 is used to generate mailbox data and simulate a MAPI workload. Exchange data grows on a daily basis (20 transaction logs per user). Afterwards, weekly full backup and daily incremental backup tests are performed to validate the solution design.

Environment validation with Microsoft Jetstress

Introduction

Exchange 2010 storage designs should be validated for expected transactional IOPS before it is placed into a production environment. To ensure that it functions appropriately, it is recommended that you use the Microsoft Jetstress tool to validate the Exchange storage design.

The Jetstress tool simulates Exchange I/O at the database level by interacting directly with the Extensible Storage Engine (ESE) database technology (also known as Jet) on which Exchange is built.

You can configure Jetstress to test the maximum I/O throughput available to your disk subsystem within the required performance constraints of Exchange. Jetstress can accept a simulated profile of specific user count and IOPS per user to validate that the disk subsystem is capable of maintaining an acceptable performance level by the metrics defined in that profile.

EMC strongly recommends that you use Jetstress testing to validate storage reliability and performance prior to the deployment of your Exchange production environment.

You can download the 64-bit version of Jetstress 2010 on the following Microsoft website:

Microsoft Exchange Server Jetstress 2010

Note The Jetstress documentation describes how to configure and execute an I/O validation or evaluation on server hardware.

Validation results

In this solution, Jetstress version 14.01.0180.003 is used to simulate an I/O profile of 0.1 IOPS per user. The building blocks are validated using a 2-hour performance test. Table 8 shows the sum of I/Os and the average latency across all the servers, which reflects the aggregate performance of this solution.

Table 8. Aggregate performance

Database I/O	Target values	Two mailbox servers (normal operating condition 1,500 users per mailbox server)
Achieved transactional IOPS (I/O database reads/sec + I/O database writes/sec)	150 (single server) / 300 (both servers)	181 (single server) / 362 (both servers)
I/O database reads/sec	N/A	101(single server) /202 (both servers)
I/O database writes/sec	N/A	80 (single server) / 160 (both servers)
I/O database reads average latency (ms)	< 20 ms	14
I/O database writes average latency (ms)	This counter is not a good indicator for client latency because database writes are asynchronous.	4
Transaction log I/O	**Target values**	**Two mailbox servers (normal operating condition 1,500 users per mailbox server)**
I/O log writes/sec	N/A	71 (single server) / 142 (both servers)
I/O log reads average latency (ms)	< 10 ms	0
Total I/O	**Target values**	**Two mailbox servers (normal operating condition 1,500 users per mailbox server)**
(DB+Logs+BDM+ Replication)/sec	N/A	403 (single server) / 806 (both servers)

Environment validation with Microsoft Exchange Server Load Generator

Introduction

After you have completed the storage validation with Jetstress and have determined that storage is sized and performs as expected, the next step in the validation process is to use the Microsoft Exchange Server Load Generator (LoadGen) tool to simulate a MAPI workload against the entire Exchange infrastructure. LoadGen testing is necessary to determine how each Exchange component performs under close-to-production user load.

LoadGen requires full deployment of the Exchange environment for validation testing. You should perform all LoadGen validation testing in an isolated lab environment where there is no connectivity to production data. LoadGen generates users and workloads against the entire Exchange environment including network and storage components.

LoadGen simulates the entire email flow and locates any bottlenecks in the solution. LoadGen also assists in determining CPU and memory resources that are required to sustain the load for which the Exchange environment is designed.

Validation profile

In this solution, LoadGen 2010 is used to simulate Outlook 2007 online mode mailboxes with the following characteristics:

- The action profile is 100 messages per mailbox per day.
- Each mailbox is 1 GB in size.
- Each database contains 300 mailboxes.
- Each mailbox server contains five active database copies with 1,500 active users in total.

The validity of each test run is determined by comparing the results of selected performance counters to a Microsoft specified criteria. Performance counter data is collected at one-minute intervals for the duration of each test run. The results of the first and last hours are discarded. Results are averaged over the remainder of the test.

For additional information about monitoring Exchange 2010 performance and other key performance counters, visit the following Microsoft TechNet website:

Performance and Scalability Counters and Thresholds

Table 9 shows the LoadGen tests that are performed to measure the performance of the Exchange infrastructure of this solution.

Table 9. LoadGen tests

Test	Description
1	Normal operation - 8 hrs, 100% concurrency test under normal operating conditions with 100 messages MAPI profile
2	Mailbox server failure - 8 hrs, 100% concurrency test during the failure of one mailbox server, all databases are brought active on the other mailbox server.
3	Hyper-V server failure - 8 hrs, 100% concurrency test during the failure of one Hyper-V server, containing one mailbox server, one CAS/HUB server, and one domain controller. The load on the remaining mailbox and CAS/HUB servers doubles.

Test 1 – Normal operating condition

Test 1 objective

In this test, the objective is to validate the entire Exchange environment under normal operating conditions. The performance of the Hyper-V host and virtual machines is measured against Microsoft recommended performance targets and thresholds.

Test 1 performance results and analysis

As shown in Table 10, the performance achieved by Hyper-V servers and Exchange virtual machines meets target metrics.

Table 10. Performance result for LoadGen test 1

	Performance counter	Target	Result
Hyper-V root server	Hyper-V Hypervisor Logical Processor(_total)\% Guest Run Time	<75%	9%
	Hyper-V Hypervisor Logical Processor(_total)\% Total Run Time	<80	0%
Mailbox server	Hyper-V Hypervisor Logical Processor (VP0-4)\% Guest Run Time	<80%	19%
	MSExchange database\I/O database reads (attached) average latency	<20 ms	10
	MSExchange database\I/O database writes (attached) average latency	<20 ms <reads avg.	4
	MSExchange database\I/O database reads (recovery) average latency	<200 ms	5
	MSExchange database\I/O database writes (recovery) average latency	<200 ms	9

	MSExchange database\IO log read average latency	<20 ms	6
	MSExchange database\IO log writes average latency	<20 ms	7
	MSExchange Replication(*)\ReplayQueueLength	<2	0
	MSExchangeIS\RPC requests	<70	1
	MSExchangeIS\RPC averaged latency	<10 ms	3
CAS/HUB servers combo	Hyper-V Hypervisor Virtual Processor (VP0-4)\% Guest Run Time	<80%	9%
	MSExchange RpcClientAccess\RPC Averaged Latency	<250 ms	7
	MSExchange RpcClientAccess\RPC Requests	<40	1
	MSExchangeTransport Queues(_total)\Aggregate Delivery Queue Length (All Queues)	<3000	1
	MSExchangeTransport Queues(_total)\Active Remote Delivery Queue Length	<250	0
	MSExchangeTransport Queues(_total)\Active Mailbox Delivery Queue Length	<250	1

Test 2 – Server failure simulation

Test 2 objective

The objective of this test is to validate the environment's performance when one mailbox server is down and all databases are then activated on the other mailbox server.

Test 2 configuration

In this test, one mailbox server is turned off in Hyper-V Manager to simulate a server failure event. The event is detected by the DAG, and the affected databases are activated on the other DAG member. LoadGen is configured to generate the client load of 3,000 users.

Test 2 performance results and analysis

After one mailbox server is turned off to simulate the server failure event, all the five affected mailbox databases are activated on the other DAG member within 20 seconds.

As shown in Table 11, the performance achieved by Hyper-V servers and Exchange virtual machines meets the target metrics. Compared with the results of Test 1, the performance of the CAS server is almost the same. The load on the mailbox server increases because all the users are now hosted on a single mailbox server.

Additionally, the DAG replication stops because only one DAG member is online.

Table 11. Performance result for LoadGen test 2

	Performance counter	Target	Result
Hyper-V Root Server	Hyper-V Hypervisor Logical Processor(_total)\% Guest Run Time	<75%	11%
	Hyper-V Hypervisor Logical Processor(_total)\% Total Run Time	<80%	13%
Mailbox Server	Hyper-V Hypervisor Logical Processor (VP0-4)\% guest run time	<80%	28%
	MSExchange database\I/O database reads (attached) average latency	<20 ms	11
	MSExchange database\I/O database writes (attached) average latency	<20 ms <reads avg.	6
	MSExchange database\I/O database reads (recovery) average latency	<200 ms	0
	MSExchange database\I/O database writes (recovery) average latency	<200 ms	0
	MSExchange database\IO log read average latency	<20 ms	0
	MSExchange database\IO log writes average latency	<20 ms	2
	MSExchange Replication(*)\ReplayQueueLength	<2	0
	MSExchangeIS\RPC requests	<70	2
	MSExchangeIS\RPC averaged latency	<10 ms	3
CAS/HUB servers combo	Hyper-V Hypervisor Virtual Processor (VP0-4)\% Guest Run Time	<80%	10%
	MSExchange RpcClientAccess\RPC Averaged Latency	<250 ms	8
	MSExchange RpcClientAccess\RPC Requests	<40	1
	MSExchangeTransport Queues(_total)\Aggregate Delivery Queue Length (All Queues)	<3000	1
	MSExchangeTransport Queues(_total)\Active Remote Delivery Queue Length	<250	0
	MSExchangeTransport Queues(_total)\Active Mailbox Delivery Queue Length	<250	1

Test 3 – Hyper-V server failure simulation

Test 3 objective

The objective of this test is to validate the environment's performance when a Hyper-V node is down, which causes the loss of one mailbox server, one CAS/HUB server, and one domain controller. This test measures the performance when all 3,000 mailboxes are served by a single mailbox server, a single CAS/HUB server, and a single domain controller.

Test 3 configuration

In this test, CAS Array is configured to provide high availability on the CAS server role. After one Hyper-V node is powered off, the event is detected by the DAG, and the affected databases are activated on the other DAG member. LoadGen is configured to generate the client load of 3,000 users.

Note DAG requires majority voters to be available in order to remain online. If the failed Hyper-V server hosts both the witness server (usually a HUB server) and the mailbox server, automatic recovery is impossible. Administrator effort is required to fix this kind of failure. For additional information, visit the Microsoft TechNet website:

Understanding Database Availability Groups

Test 3 performance results and analysis

After one Hyper-V node is turned off, all five affected mailbox databases are activated on the other DAG member within 20 seconds. As shown in Table 12, the performance achieved by Hyper-V servers and Exchange virtual machines meets the target metrics. Compared with the results of Test 2, the mailbox server's performance is almost the same. The load on the CAS server increases due to losing the other CAS server.

Table 12. Performance result for LoadGen test 3

	Performance counter	Target	Result
Hyper-V Root Server	Hyper-V Hypervisor Logical Processor(_total)\% Guest Run Time	<75%	13%
	Hyper-V Hypervisor Logical Processor(_total)\% Total Run Time	<80%	14%
Mailbox Server	Hyper-V Hypervisor Logical Processor (VP0-4)\% guest Run Time	<80%	27%
	MSExchange database\I/O database reads (attached) average latency	<20 ms	11
	MSExchange database\I/O database writes (attached) average latency	<20 ms <reads avg.	5
	MSExchange database\I/O database reads (recovery) average latency	<200 ms	0
	MSExchange database\I/O database writes (recovery) average latency	<200 ms	0

	MSExchange database\IO log read average latency	<20 ms	0
	MSExchange database\IO log writes average latency	<20 ms	2
	MSExchange Replication(*)\ReplayQueueLength	<2	0
	MSExchangeIS\RPC requests	<70	2
	MSExchangeIS\RPC averaged latency	<10 ms	4
CAS/HUB servers combo	Hyper-V Hypervisor Virtual Processor (VP0-4)\% Guest Run Time	<80%	16%
	MSExchange RpcClientAccess\RPC Averaged Latency	<250 ms	8
	MSExchange RpcClientAccess\RPC Requests	<40	2
	MSExchangeTransport Queues(_total)\Aggregate Delivery Queue Length (All Queues)	<3000	1
	MSExchangeTransport Queues(_total)\Active Remote Delivery Queue Length	<250	0
	MSExchangeTransport Queues(_total)\Active Mailbox Delivery Queue Length	<250	1

Exchange backup scenarios

Introduction

Table 13 lists the Exchange backup scenarios performed in this solution.

Table 13. Exchange backup scenarios

Test Scenario	Description
1	Backup with Data Domain Boost disabled and without user load
2	Backup with Data Domain Boost enabled and without user load
3	Backup with Data Domain Boost enabled and with user load

Note The test results of deduplication ratio are based on the data generated by LoadGen. Because LoadGen is not specialized to test deduplication ratio, the results provided in this section may differ from the actual data in a customer production environment.

Scenario 1: Backup with Data Domain Boost disabled and without user load

This test scenario is used to validate the Exchange 2010 backup performance and deduplication ratio in Data Domain with DD Boost disabled and without user load. Both weekly full backup and daily incremental backup scenarios are tested.

The test results are listed in Table 14.

Table 14. Test results of Scenario 1

Scenario	Total amount of data	Total backup window	Deduplication ratio (reduction %)	Total deduplication ratio (reduction %)
Weekly full backup (no new data)	3771 GB	7 h 42 min	17.9x (94.4)	N/A
Daily incremental	75 GB	21 min	5.5x (81.9)	N/A

The backup throughput to Data Domain for the weekly full backup is 490 GB/h and the backup throughput for the incremental backup is about 214 GB/h.

Scenario 2: Backup with Data Domain Boost enabled and without user load

This test scenario is used to validate the Exchange 2010 backup performance and deduplication ratio in Data Domain with DD Boost enabled and without user load. This test is based on a weekly full backup and daily incremental backup schedule. Compared with backup Scenario 1, Data Domain Boost improves the backup throughput at a rate of about 70 percent.

The test results are listed in Table 15.

Table 15. Test results of Scenario 2

Scenario	Total amount of data	Total backup window	Deduplication ratio (reduction %)	Total deduplication ratio (reduction %)
Initial full backup	3,580 GB	4 h 59 min	2.2x (54.4)	2.2x (54.4)
Weekly full backup (no new data)	3,565 GB	4 h 12 min	16.8x (94)	3.9x (74.3)
First incremental backup	118 GB	24 min	2.1x (51.3)	3.8x (73.3)
Second incremental backup	58 GB	11 min	1.9x (48.0)	3.8x (73.8)

- The backup throughput to Data Domain for the weekly full backup is 849 GB/h and the average backup throughput for incremental backups is about 300 GB/h.
- The total deduplication ratio is 3.8x (73.8 percent) after two full backups and two incremental backups. It is predictable that the total deduplication ratio will be increased by adding more weekly full backup data in Data Domain, which means longer retention periods result in increasingly higher deduplication ratios.
- The initial full backup and the incremental backups can be considered as new data backups. For these new data backups, the average deduplication ratio is about 2.1x and the corresponding data reduction is more than 50 percent.

Scenario 3: Backup with Data Domain Boost enabled and with user load

This test scenario is to validate the Exchange 2010 backup performance and deduplication ratio in Data Domain with DD Boost enabled and with user load. Both weekly full backup and daily incremental backup scenarios are tested.

According to the design, when the backup operation of the passive database copies is being performed, there must be user load from the active database copies on the same server. Therefore, it is necessary to test the backup scenario by taking the user load into consideration. The results of this test scenario provide a good reference for customers with similar designs. The test results are listed in Table 16.

Table 16. Test results of Scenario 3

Scenario	Total amount of data	Total backup window	Deduplication ratio (reduction %)	Total deduplication ratio (reduction %)
Initial full backup	3,628 GB	7 h 48 min	2.7x (62.8)	2.7x (62.8)
Weekly full backup (no	3,641 GB	7 h 27 min	59.0x (98.3)	3.8x (73.7)

new data)				
First incremental backup	46 GB	14 min	1.6x (37.6)	3.7x (72.8)
Second incremental backup	39 GB	13 min	2.3x (55.9)	3.8x (73.4)

The backup throughput to Data Domain for the weekly full backup is 489 GB/h and the average backup throughput for incremental backups is about 189 GB/h. Compared with backup Scenario 2, the backup performance impact caused by user load is about 40 percent. Given the limitation of using 1 Gb Ethernet network bandwidth, a backup throughput of about 500 GB/h can be considered as an ideal result.

Performance analysis during the backup

Table 17 consolidates the achieved performance of Hyper-V servers and Exchange virtual machines in all of the three scenarios. The results show that the performance during the weekly full and daily incremental backups meets the target metrics.

Table 17. Consolidated test results

Performance counter	Target	Scenario 1		Scenario 2 (Boost)		Scenario 3 (Boost + Load)	
		Weekly full	Daily Inc.	Weekly full	Daily Inc.	Weekly full	Daily Inc.
Hyper-V Hypervisor Logical Processor(_total)\% Guest Run Time	<75%	7.55	13.14	5.97	18.62	10.23	21.13
Hyper-V Hypervisor Logical Processor(_total)\% Total Run Time	<80%	8.29	13.18	6.38	19.97	11.02	21.59
Memory\available Mbytes	>100 MB	3070	3090	3132	2581	2156	2480
Memory\% committed bytes in use	<80%	44.29	43.96	45.62	46.09	46.74	45.23
Memory\transition pages repurposed/sec	<1000	0	0	0	0	0	0
MSExchange database\I/O database reads (attached) average latency	<20 ms	0	2.85	2	11.74	5.83	14.49
MSExchange database\I/O database writes (attached) average latency	<20 ms <reads avg.	0	7.72	0.22	7.14	2.33	2.35
MSExchange database\IO log read average latency	<20 ms	1.3	1.26	0.44	2.1	1.69	0.91
MSExchange database\IO log writes average latency	<20 ms	0	4.11	6.76	10.37	1.25	1.24
MSExchange Replication(*)\ ReplayQueueLength	<2	0.013	0.02	0.018	0.16	0.53	0.375
MSExchange replication(*)\ CopyQueueLength	<1	0	0	0	0.017	0.06	0

MSExchangeIS\RPC requests	<70	0	0	0	0.19	0.72	0.85
MSExchangeIS\RPC averaged latency	<10 ms	0	0.26	1	7.06	2.35	2.13
MSExchangeIS mailbox(*)\RPC averaged latency	<10 ms	1.93	0.4	0.45	4.36	1.41	1.13
MSExchangeIS mailbox(_total)\messages queued for submission	0	0	0	0	0	0.77	1.13

The following two figures show the Data Domain performance during a full point-in-time backup. Figure 5 is for Scenario 1 where Data Domain Boost is disabled, and Figure 6 is for Scenario 2 where Data Domain Boost is enabled.

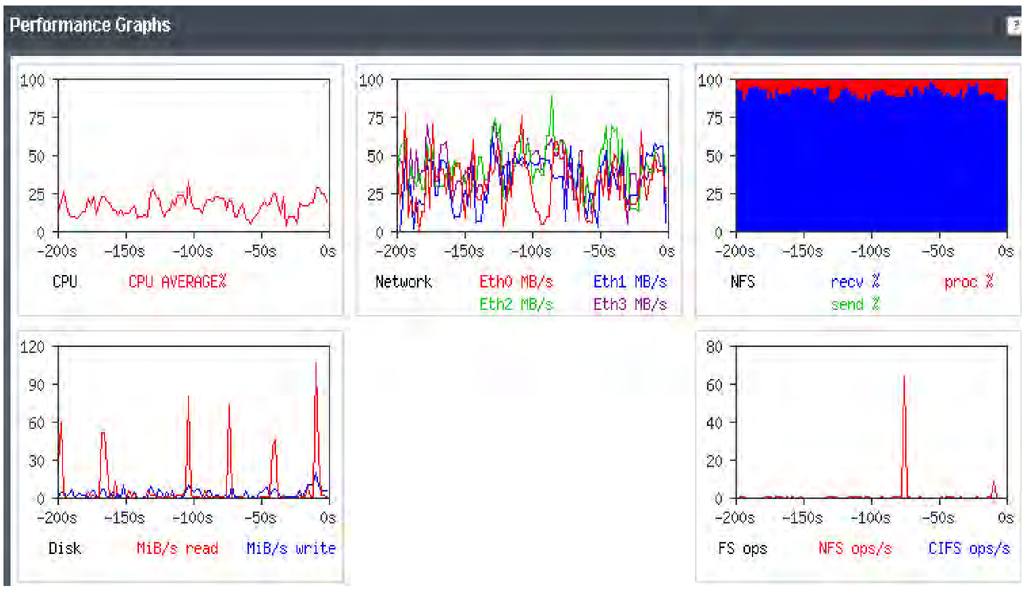

Figure 5. Data Domain performance for Scenario 1 (DD Boost disabled)

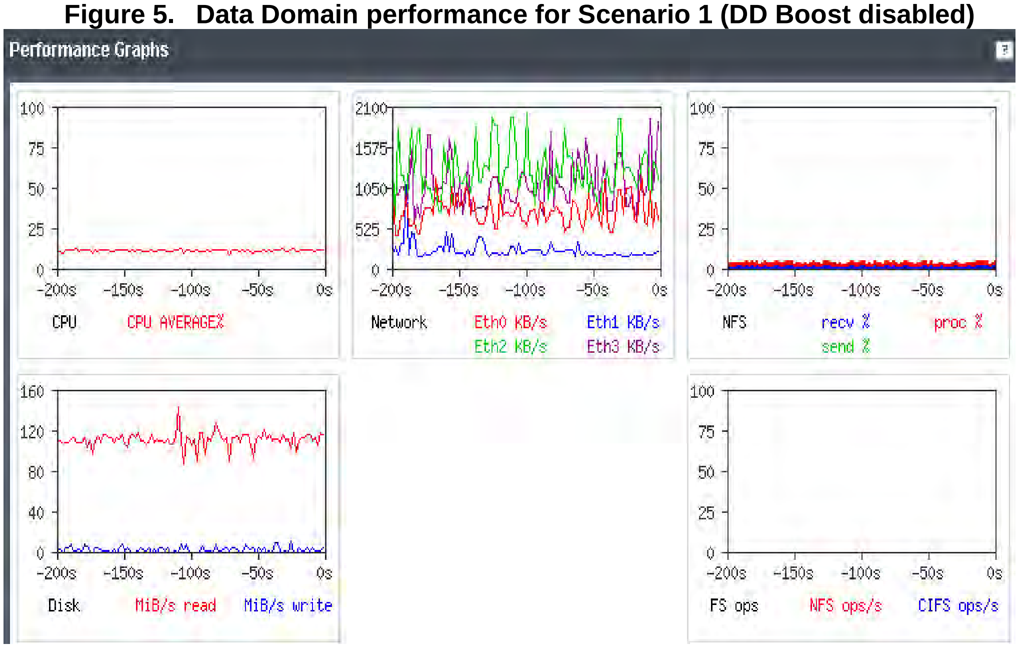

Figure 6. Data Domain performance for Scenario 2 (DD Boost enabled)

Comparison of these two figures in the upper-right section indicates a reduction in received data. As by design, Data Domain Boost does not send the entire data stream to the Data Domain system. Instead, it sends metadata that exactly informs the storage node which unique data to send to the Data Domain system.

The CPU utilization is also decreased when Data Domain Boost is enabled. Because only unique data needs to be sent over the LAN to the Data Domain device instead of the entire payload, the CPU works more efficiently with Data Domain Boost enabled.

These relative efficiencies are achieved by Data Domain Boost enabling the storage node to make the "unique data" decisions further away from the target Data Domain device. Conversely, this also means deduplication processing is closer to the source, with the typically expected efficiencies observed.

Exchange 2010 recovery scenarios

Introduction

Table 18 lists the Exchange recovery scenarios performed in this solution.

Table 18. Exchange recovery scenarios

Test Scenario	Description
1	Single database recovery
2	Single mailbox recovery

Scenario 1: Single database recovery

This test shows the RTO to perform a single Exchange 2010 database recovery. This operation overwrites the original database.

The test results are as follows:

- It takes less than 3 hours (176 minutes) to recovery from one full Exchange 2010 database backup that contains 4,591 files (367 GB).
- It only takes 2 minutes to restore from one incremental Exchange 2010 database backup that contains 4,161 files (4,258 MB).
- The recovery speed is about 125 GB/h.
- Table 19 lists the performance counters captured on mailbox server 1. They are used to measure the impact in the production environment during the Data Domain recovery.

Table 19. Performance of mailbox servers

Mailbox server	CPU usage	Network bandwidth
Mailbox server 1	7%	250 MB/s

Figure 7 shows the Data Domain performance during a full point-in-time recovery.

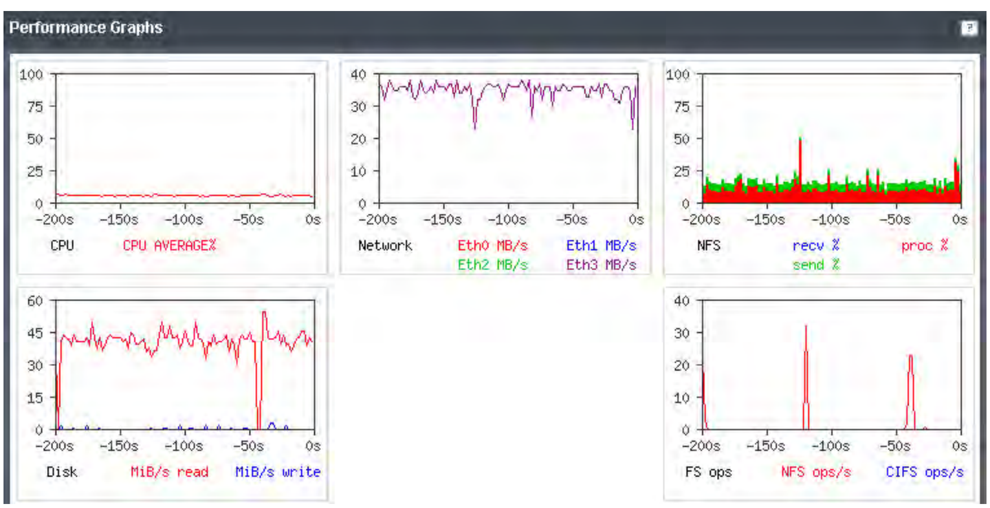

Figure 7. Data Domain performance

Scenario 2: Exchange 2010 mailbox recovery

Microsoft provides a mechanism called recovery database (RDB) to recover data at the mailbox level or item level. RDB is a special kind of mailbox database used to mount a restored mailbox database and to extract data from the restored database as part of a recovery operation.

You can use the following steps to recover data using RDB.

Step	Action
1	Use the NetWorker User program to restore the database to a point-in-time (full and incremental recovery) and to a new location.
2	Play back all of the available transaction logs found in the log directory and apply the restored transaction logs with the `Eseutil` command. For example: *`R:\db1>eseutil /R E02 /I /L “R:\log1” /D`*
3	Create the recovery database with the `New-MailboxDatabase` command and the `–Recovery` switch. Specify the database path and log file path as the predefined folders. For example: *`New-MailboxDatabase –Recovery –Name RDB1 –Server 498MBX01 –EdbFilePath “R:\DB1\Exchange_Snap_29092010-2256-6852\Mailbox database 01.edb” –LogFolderPath “R:\Log1”`*
4	Mount the recovery database and allow restores to the original mailbox database: *`Mount-Database –Identity RDB1`* *`Set-MailboxDatabase –Identity RDB1 -AllowFileRestore:$true`*
5	Use the `Restore-Mailbox` command to recover mailbox-level data or item-level data.

For more details about recovery databases, visit the Microsoft TechNet websites:

- Create a Recovery Database
- Restore Data Using a Recovery Database

The advantage of this recovery method is that RDB is a built-in feature of Exchange 2010.

Conclusion

Summary

This solution demonstrates how to build an Exchange 2010 DAG environment on a CLARiiON CX4-480 platform.

This solution provides the Exchange 2010 backup with a schedule of a weekly full and daily incremental backups. It uses VSS to back up terabytes of data rapidly and stores it efficiently on the data deduplication storage.

Key points

The table below summarizes the key points that this solution addresses.

Key Point	**Solution objective**
Reduced backup timeframes	Backing up terabytes of data takes a long time with a traditional tape-based backup solution. For example, backing up 3.6 TB of data with LTO-5 (140 MB/s), the most advanced tape technology, requires 18 hours for conventional tape backup. However, it takes about only 7 hours to back up the same amount of data through 1 Gb Ethernet in this solution, which means the backup time is significantly reduced by up to 60%. Additionally, Data Domain Boost improves the backup throughput at a rate of about 70%.
Reduced backup storage requirements	Compared with the traditional 100% backup storage capacity requirement on tape, only a relatively small amount of storage capacity is required on Data Domain for full and incremental backups with data deduplication. In this solution, the total deduplication ratio is about 3.8:1 after two full backups and two incremental backups. **Note** Your mileage may vary depending on how many duplicates are in your specific environment.
Reduced physical infrastructure footprint	By deploying the domain controller, Exchange servers, and the NetWorker server on a Microsoft Hyper-V virtualized platform, the number of physical servers required in this solution is significantly decreased, which leads to a higher server utilization rate and reduced energy and operational costs. In this solution, the entire environment only consists of three physical servers with Hyper-V enabled, while the conventional deployment without virtualization requires at least seven physical servers.
Validated and tested performance	The test results show that the designed architecture satisfies all recommended performance guidelines provided by Microsoft for Exchange 2010.

Next steps

To learn more about this and other solutions contact an EMC representative or visit: www.emc.com.

References

White papers

For additional information, see the white papers listed below.

- *EMC Backup and Recovery for Microsoft Applications—Deduplication Enabled by EMC CLARiiON and Data Domain—A Detailed Review*
- *Business Continuity for Microsoft Exchange 2010 - Enabled by EMC Unified Storage, Cisco Unified Computing System, and Microsoft Hyper-V*

Product documentation

For additional information, see the product documents listed below.

- *Microsoft Exchange 2010 Backup and Recovery Support with EMC NetWorker Technical Note*
- *EMC NetWorker Data Domain Deduplication Devices Version 7.6 Service Pack 1 Integration Guide*
- *Data Domain OS 4.9 Administration Guide*
- *Data Domain OS 4.9 Command Reference Guide*
- *EMC NetWorker Release 7.6 Service Pack 1 Administration Guide*
- *EMC NetWorker Release 7.6 Service Pack 1 Error Message Guide*
- *EMC NetWorker Release 7.6 Service Pack 1 Installation Guide*
- *EMC NetWorker Release 7.6 Service Pack 1 Release Notes*

Other documentation

For additional information, see documents at the website listed below:

- Microsoft TechNet