EMC²
where information lives®

EMC Archiving for Microsoft Exchange 2007

Enabled by EMC CLARiiON CX4, EMC Centera, EMC SourceOne, and VMware vSphere 4

Reference Architecture

EMC Global Solutions

EMC² PROVEN SOLUTIONS

Microsoft GOLD CERTIFIED Partner

vmware® TECHNOLOGY ALLIANCE PARTNER

Published October 2009

Part number: H6441.1

Table of Contents

Reference architecture overview

Document purpose

This document details the reference architecture for the EMC® Archiving for Microsoft Exchange 2007 enabled by EMC CLARiiON® CX4, EMC Centera®, EMC SourceOne®, and VMware vSphere 4 solution.

Solution purpose

The purpose of this solution is to:

- Develop a suggested, EMC SourceOne architecture for 10,000 users in the context of the virtual data center
- Test and document the performance of historical archiving and shortcutting of e-mail from Microsoft Exchange 2007 into the EMC SourceOne native archive (extended via DiskXtender® to EMC Centera)
- Test and document the performance of realtime journaling activities from the Microsoft Exchange 2007 journaling server to the EMC SourceOne native archive (extended via DiskXtender to EMC Centera)

In this solution, EMC's CLARiiON CX4-480 array is used to provide consolidated storage, while EMC SourceOne is used to improve operational efficiency, mitigate risk, and enable compliance with regulatory and governance requirements.

The EMC Centera array provides a second tier, secure platform for data archiving that ensures static and/or infrequently changing information is available online for fast access. An enterprise can regain control of an ever-growing information environment, accelerate backups, and facilitate day-to-day business, compliance, and regulatory needs.

VMware vSphere 4 enables customers to bring the power of cloud computing to their IT infrastructures. Building on the power of VMware Infrastructure, VMware vSphere dramatically reduces capital and operating costs, and increases control over IT infrastructures while preserving the flexibility to many different OS types, applications, or hardware[1].

The solution is engineered to allow customers to:

- Introduce e-mail archiving into their environment with consideration for storage configuration, design, sizing, software, and backup operations.
- Reduce operational costs while enforcing e-mail record-keeping policies in compliance with internal governance as well as industry and government regulations.
- Deliver the highest levels of application service agreements with the lowest total cost per application workload.
- Achieve the flexibility of a solution that scales to meet large enterprise needs, yet offers a simple footprint for midsize organizations.

[1] VMware website

The solution provides information on:

- Creating a well-performing storage design for a virtualized Microsoft Office Exchange 2007 environment on an EMC CLARiiON CX4-480 with a large and very active database (0.48 IOPS, 11.2 TB of Exchange data).
- Designing and deploying EMC SourceOne.
- Documenting the observed environment performance and the suggested methods of growing performance as needs dictate.

The business challenge

Protecting a company from data loss is a strategic imperative. Customers are facing the following business challenges on a daily basis, and many are looking for the easiest way to achieve:

- Optimized IT infrastructure through server and connectivity consolidation
- Decreased downtime and improved reliability through high availability and automated disaster recovery
- Increased energy efficiency through fewer servers and the ability to dynamically power down unused servers with green IT solutions
- Proactive messaging management
- Increased user productivity through seamless access to archived content
- Optimized e-mail archiving, litigation readiness, and compliance to information governance standards

The challenges in meeting these criteria are:

- Regulatory requirements that mandate secure data
- Corporate data security policies
- Compliance reporting
- Unknown resource requirements
- Undetermined infrastructure design to provide high availability

---

The technology solution

The following table describes the environment (including the virtual allocation of servers) and how the roles within it were configured.

Quantity	Purpose	Configuration	Software
2	Active Directory (AD) servers	• 2 vCPUs • 4 GB RAM • 1 vNIC	Windows 2008 (64-bit) SP1
2	Hub/CAS servers	• 4 vCPUs • 16 GB RAM • 1 vNIC	• Windows 2008 (64-bit) SP1 • Exchange 2007 SP1
2	Mailbox servers (users)	• 8 vCPUs • 32 GB RAM • 1 vNIC	• Windows 2008 (64-bit) SP1 • Exchange 2007 SP1
2	Mailbox servers (Journal)	• 4 vCPUs • 16 GB RAM • 1 vNIC	• Windows 2008 (64-bit) SP1 • Exchange 2007 SP1
3	EMC SourceOne native archive (archive and index roles)	• 4 vCPUs • 4 GB RAM • 1 vNIC	• Windows 2003 (32-bit) SP2 • EMC SourceOne native archive
2	EMC SourceOne native archive (search and retrieve roles)	• 4 vCPUs • 4 GB RAM • 1 vNIC	• Windows 2003 (32-bit) SP2 • EMC SourceOne native archive
3	EMC SourceOne worker servers	• 4 vCPUs • 4 GB RAM • 1 vNIC	• Windows 2003 (32-bit) SP2 • EMC SourceOne worker • EMC SourceOne Web services
2	Microsoft Failover Clustering nodes for SQL Server	• 8 vCPUs • 32 GB RAM • 1 vNIC	• Windows 2003 (64-bit) SP2 • SQL Server 2005 SP2
2	Microsoft Failover Clustering nodes for SourceOne master role (also hosts Index and Message Center shares)	• 4 vCPUs • 8 GB RAM • 1 vNIC	• Windows 2003 (32-bit) SP2 • EMC SourceOne master

Key components

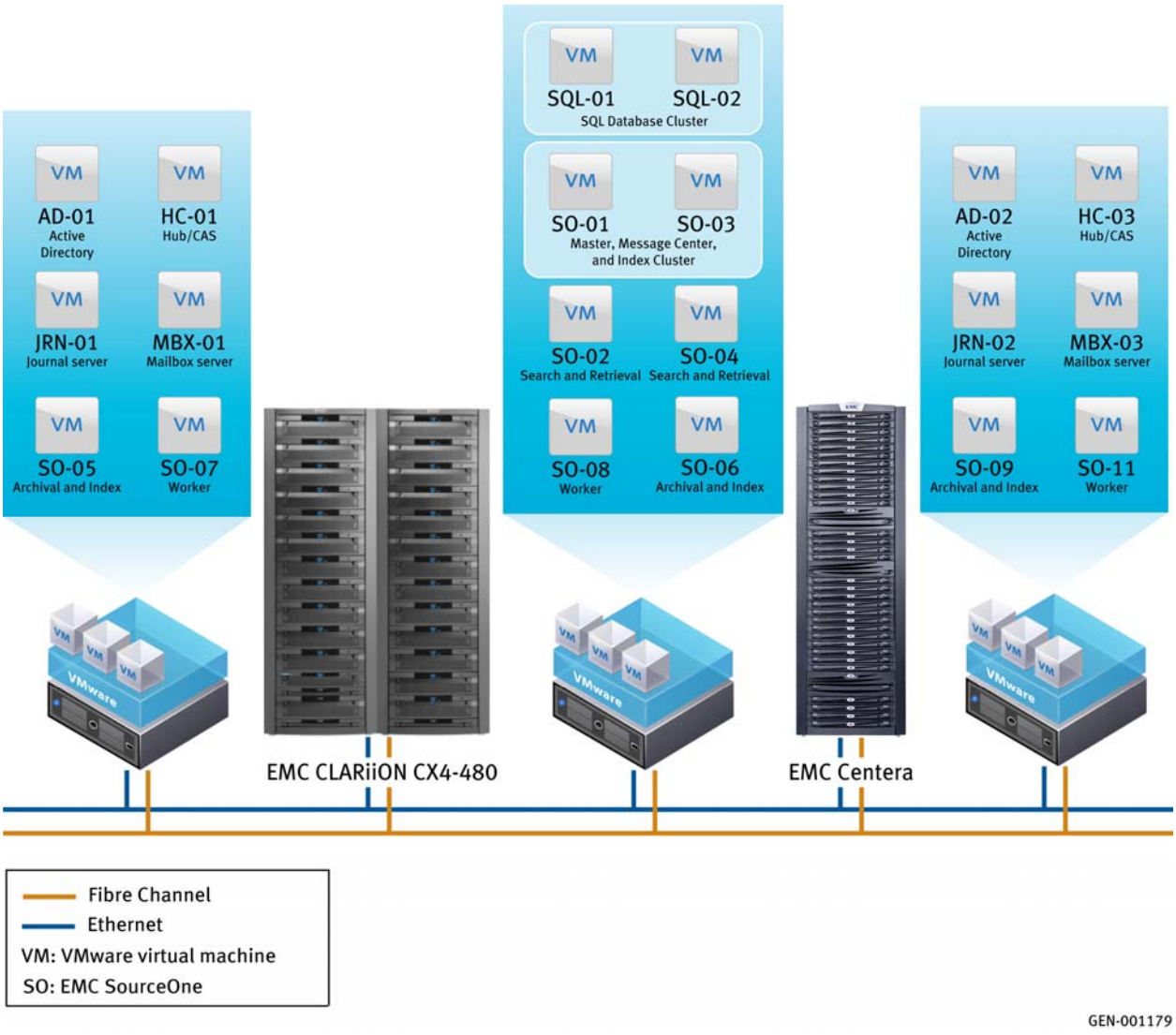
Introduction	This section briefly describes the key solution components. For details on all of the components that make up the reference architecture, refer to the hardware and software resources section.
EMC CLARiiON CX4-480	EMC CLARiiON CX4-480 provides high-capacity networked storage that meets the needs of demanding online transaction processing (OLTP) workloads and large-scale e-mail environments. With the CX4-480 customers can scale seamlessly up to 471 TB of storage capacity and consolidate twice the workloads in one array as is possible with other storage providers.
EMC Centera	EMC Centera delivers active archiving with the capacity, power efficiency, and price point required. Available in a rackable configuration, EMC Centera fits into any EMC rack or industry-standard rack. It provides seamless, affordable access to archived information — including e-mail, electronic documents and images, and call center records. Businesses can regain control of an ever-growing information environment, accelerate backups, and facilitate day-to-day business, compliance, and regulatory needs.
EMC SourceOne	EMC SourceOne e-mail management supports all major messaging environments, including Microsoft Exchange, IBM Lotus Domino, and instant messaging (IM). SourceOne improves the IT efficiency of messaging systems, reduces costs of storage, and improves backup and recovery processes. EMC SourceOne also enables enterprise-wide deployment with a consolidated and centrally managed archive. It provides maximum scalability and litigation readiness for search against the available archive, all at the lowest cost of ownership.
VMware vSphere 4	VMware vSphere is proven to comprehensively virtualize server, storage, and networking resources, reducing capital and operational expenses per application by more than 50 percent. Organizations can achieve consolidation ratios of more than 15:1, while increasing efficiency through management automation and the dynamic allocation of resources to business-critical applications across internal and external cloud infrastructures[2].

[2] VMware website

Physical architecture

Architecture diagram

The following diagram depicts the overall physical architecture of the solution.



Validated environment profile

Profile characteristics The solution was validated with the following environment profile.

Profile characteristic	Value
Exchange 2007 users (0.48 IOPS – very heavy user)	10,000 divided between: • 3,000 users x 2 GB mailboxes • 3,000 users x 1 GB mailboxes • 4,000 users x 500 MB mailboxes
Read / Write ratio	1:1
Number of Exchange 2007 users per server	5,000[3]
Number of Exchange 2007 storage groups per server	28 divided between: • 15 storage groups for 2 GB users • 8 storage groups for 1 GB users • 5 storage groups for 500 MB users
Number of Exchange 2007 mail databases per storage group	1 (maximum size 200 GB)
Total size of all Exchange databases in the environment	56 x 200 GB (11.2 TB)
Number of Exchange 2007 users per mail database	Varies based on mailbox size: • 100 users per database where the mailbox size is 2 GB • 188 users per database where the mailbox size is 1 GB • 400 users per database where the mailbox size is 500 MB
Size of Exchange 2007 user mailbox	Varies based on user: • 2 GB • 1 GB • 500 MB
Size of storage group mailbox database LUN (to accommodate the specified number of databases and the deleted item retention on the defrag space)	256 GB

[3] The use of 8 vCPUs to support 5,000 users does not constitute a new building block for Microsoft Exchange 2007 using VMware vSphere. The goal of this change was to ensure that CPU at the Exchange Server was not a bottleneck, allowing any effects of EMC SourceOne activities to be observed during testing. Baseline performance showed CPU utilization remained under 30% on the Exchange mailbox servers. Therefore, it is entirely possible that an additional 8 vCPU mailbox server would support more than 5,000 users, or users with increased IOPS.

Profile characteristic	Value
Storage group log LUN size	40 GB
Exchange 2007 production data RAID type, physical drive size, and speed	RAID 1/0, 450 GB, 15k
Exchange 2007 production clone RAID type, physical drive size, and speed	RAID 5, 450 GB, 15k
Number of storage groups per journaling server	2
Journaling servers	2
Number of journaling mailboxes per production Exchange storage groups	1 (= >4 in total)
Number of journaling mailboxes per journaling server storage group	1

Hardware and software resources

Hardware The hardware used to validate the solution is listed below.

Equipment	Quantity	Configuration
Storage array	1	CLARiiON CX4-480 • 16 GB memory • 8 Fibre Channel, 4 iSCSI ports per storage processor • 16 DAEs • 180 Fibre Channel 450 GB 15k[4] • FLARE® 28
CAS	1	EMC Centera (8 node) • Generation 4 EMC Centera • 500 GB drives • 16 TB raw capacity
SAN	2	32-port fabric switch
Network	2	32-port Ethernet switch
Production servers	3	16-core, 2.9 GHz CPU, 128 GB RAM
Load servers	4	16-core, 2.9 GHz CPU, 32 GB RAM

[4] This use case tested multiple different mailbox sizes on the same Exchange server. As a result, the storage layout for this user count differs from other published material. A standard database configuration was required. Therefore, each database was designed to support the maximum number of users per database (400), rather than the actual user counts that were present on each individual database (for instance 100 users on the databases that contained 2 GB mailboxes).

Software The software used to validate the solution is listed below.

Software	Version
Microsoft Windows	2003 (SP2) Enterprise Edition
Microsoft Windows	2008 (SP1)
Microsoft Exchange	2007 (SP1)
EMC Navisphere® Agent and CLI	6.28 (Agent and CLI)
VMware vSphere	4.0 RTM
EMC SourceOne	1.0 (build 4115)
EMC DiskXtender	6.3
Microsoft Exchange Load Generator	14.00.0582.000

Conclusion

Findings

The following list details the findings noted during the validation of this reference architecture. The described architecture was found to be capable of:

- Real time archiving of e-mail as generated by 10,000 very heavy Exchange users.
- Archiving and shortcutting of between 100,000 and 120,000 messages per hour in conjunction with other environmental load factors such as active users and online maintenance (OLM).

Useful EMC SourceOne role characteristics on the virtual infrastructure were observed during testing, which helped in the design of the EMC SourceOne deployment.

Summary

Companies today are concerned about the overall cost of the storage infrastructure required to support a large and rapidly growing deployment of Microsoft Exchange 2007. While Microsoft provides e-mail management and movement capabilities within Exchange and Outlook, other solutions are available in the market place.

Once such solution for reducing overall storage costs of Exchange 2007 deployment is to deploy a third-party archiving solution such as EMC SourceOne, which will scan the e-mail items and take them out of the Exchange database according to the policies defined. This proactive “scan and hold” process is important for businesses that are under legal pressure to retain and hold e-mail for a certain amount of time due to regulatory requirements.

Independent archive solutions reduce the costs associated with production storage requirements, backup, and bandwidth. This solution also includes several other components that lower IT costs even further, for example, EMC Centera and VMware vSphere 4.

Included in the tested configuration is:

- EMC CLARiiON CX4-480 — an efficient primary storage array that comfortably handles the performance and storage capacity requirement of the 10,000 Exchange 2007 users, including the requirements for their associated journaling and historical archiving needs.
- EMC SourceOne — a reliable e-mail management and archiving software solution that scales to support the required performance and realtime archiving needs of the 10,000 very heavy Exchange 2007 users.
- EMC Centera — a reliable second tier storage array that provides the requisite CAS capacity and performance required to satisfy the archive journaling and compliancy needs of the environment.
- VMware vSphere — the industry’s first cloud operating system that leverages the power of virtualization to transform data centers into dramatically simplified cloud computing infrastructures and enable IT organizations to deliver the next generation of flexible and reliable IT services, using internal and external resources, securely and with low risk[5].

[5] VMware website

Next steps

EMC can help to accelerate assessment, design, implementation, and management while lowering the implementation risks and cost of creating a virtualized Exchange 2007 environment.

To learn more about this Exchange solution, contact an EMC representative, or your local authorized EMC reseller. Visit the Solutions section of EMC at www.emc.com/solutions for more details.