EMC²
where information lives®

Save Costs on Email Infrastructure and Obtain Value-added Benefits of Email Retention & Archiving

Satheesh MC

EMC Proven Professional Knowledge Sharing 2010

EMC²
PROVEN
PROFESSIONAL

Satheesh MC
Storage Consultant
Satheesh.mc2@gmail.com

Table of Contents

Disclaimer: The views, processes or methodologies published in this compilation are those of the authors. They do not necessarily reflect EMC Corporation's views, processes, or methodologies

Abstract

Over the past years, Email has become a primary channel of business communication. It provides organizations with a fast medium to convey business correspondence such as purchase orders, quotations and sales transactions, in virtually any geographical area with the least physical effort. The drastic escalation in storage requirements is mainly attributed to the increased use of email in general, added to the upsurge of attachments up to 20 MB. This doubles primary storage requirements in IT environments

Also due to email predominance in the business industry, various pieces of legislation have been enacted to protect personal privacy, enforce corporate governance standards and maintain ethical conduct. The Sarbanes-Oxley Act (SOX), Gramm-Leach Bliley act (GLBA) and the Freedom of information act (FOIA) are some examples. They require keeping Email for a longer duration for compliance purposes.

These requirements present two problems, one is the ever Increasing Primary Storage requirement and other one is how to conduct Email retention and archiving for compliance purposes. In addition, corporate policies and regulations require us to save data for longer periods of time. That means storage, translating into more floor space, more power, more staff, and more complexity.

This article explores a solution using EMC SourceOne Email Management and EMC Centera® to help organizations reduce their primary storage requirements and gain the added benefit of Email archiving and retention. It explores issues such as a cost-effective solution, decreasing primary storage requirements, deduplication, intelligent way of storing data, site disaster recovery model, and implementing effective Email retention and archive policies.

Introduction

A majority of business users are relying on Email more than the telephone for business communications. This means that an ever-increasing amount of critical information is passing through a company's Email system. On average, 60% of business critical information is stored within corporate messaging systems.

Email, in today's corporate world, holds a company's critical information. Financial expectations are discussed, stocks might be suggested, sales are made, customers make complaints, companies make promises and all of this is done via Email.

A growing number of businesses are realizing that they need to save these Emails. They cannot just pack them away like receipts you toss in a box at home. They need to be organized and easily accessible. When an archived Email is needed, you don't want to have to call in half the IT team to find it.

There's a lot of buzz in the industry about Email archiving being driven by regulations requiring you to save and retrieve certain kinds of data. It's difficult because if you haven't been saving this data in an organized way, it is difficult to identify and compile certain messages from certain days five years ago. You need a huge amount of backend Disk Storage (either NAS or SAN) to save all these Emails. Many organizations are running out of disk space due to the growing volume of mail files. They are doubling their data storage network to deal with this increasingly high mail volume. Unless you have sophisticated software and storage to manage this process, it is a herculean task.

EMC Solution

EMC's SourceOne provides automated archiving, granular search and retrieval, retention management, and monitoring for Email systems such as Exchange, Exchange/Domino etc.

SourceOne is a comprehensive, policy-based system that automatically collects, organizes, retains, and retrieves messages and attachments. SourceOne reduces the cost of Email storage by creating and managing a central repository of Email and other messages by compressing and eliminating storage of duplicate Emails. It improves Email server response

times, reduces backup and recovery time, boosts end-user and administrator productivity, and controls risk by supporting compliance with regulations and corporate governance policies.

SourceOne also provides a mechanism to save desktop storage by eliminating the need for individual Exchange/Domino .nsf archive files or Exchange .pst files - these can be migrated into the SourceOne archive. The SourceOne Extract facility allows .nsf/pst files to be imported into the Email Archive (utilizing the same classification, indexing, deduplication and compression utilised by the journaling of new Email and the extraction of Email from the Exchange/Domino/Exchange/Lotus servers).

All incoming and outgoing Emails are captured in real-time and managed as a record of business in accordance with rules of evidence. A copy of every incoming or outgoing Email is routinely captured on the Exchange/Domino server, then polled by and moved to the SourceOne server. Life-cycle management is deployed to classify, retain, and dispose of Email by volume according to the organizations' policies and the relevant regulatory requirements

Email Communication Life Cycle management

There are three phases to Email communications' life cycle management.

1. In the first phase, SourceOne captures a record and keeps a copy of all inbound and outbound mail and attachments in an Exchange/Domino environment through the Exchange/Domino journaling capability. At configurable intervals (such as every 60 seconds), SourceOne pulls the message, including the attachment, from the journal mailbox. It then determines whether the message is unique, content indexes the message text (To, From, Subject, Message Body, including Attachment), and then stores the message to a container file.

2. Second, EMC SourceOne automates the process of managing the "convenience copy" (the original copy in the users mailbox) of the message from within the Exchange/Domino server. The management policies enable the shortcutting /stubbing and archive of original Emails. This reduces and controls the size of end user mailboxes.

3. SourceOne generates a full-text index of all messages and attachments (including zipped attachments) and can quickly and easily retrieve these messages. Web access enables users to search/retrieve their own archived email at any time from any location on the corporate network without help from administrators while also enabling user or group access rights to be assigned to specific archive folders.

   During phase three of the lifecycle, SourceOne policies define message retention within the SourceOne archive. You can set the policy to unlimited retention.

The benefits of EMC SourceOne include:

- Centralized and consolidated Email management; saves IT Support time and effort.

- The ability for end users to simply access their archived Email messages and attachments; saves user and IT Support time and effort.

- Full-text searching for lost or deleted Email or attachments; saves end user time and IT Support effort.

- Elimination of message size and mailbox size restrictions along with a reduction in server downtime and improved access times to users' mail.

- Reduction in Exchange/Domino server back-up and restore times. As all mailboxes will have email attachments older than 90 days removed from the Exchange/Domino servers, there will be a decrease in mailbox sizes on all Exchange/Domino servers and thus a reduction in Exchange/Domino server backup and restore times.

- Software runs non-intrusively on a separate server with minimal performance impact on the Exchange/Domino Email server.

- Reduction in overall archive storage costs by compressing and eliminating duplicate Emails. (Although they do not currently use a recognised archive solution, the users are utilizing the Exchange/Domino servers as de-facto archive storage devices)

- Message-server diagnostic tools aid in the analysis of email traffic patterns and the calculation of on-going storage requirements.

- Record management of archiving (enabling email regulatory compliance and evidential weight in the event of litigation).

EMC Centera

EMC Centera is CAS Device used for archiving data. Content-addressable storage (CAS), also referred to as associative storage, is a mechanism to store information that can be retrieved based on its content rather than its storage location. Centera's unique self-healing and authentication features have prompted its industry-wide recognition as the first disk-based WORM (write once, read many) storage medium and is a compelling solution for regulated storage environments.

EMC SourceOne with EMC Centera produces amazing results. SourceOne takes care of Email management and EMC Centera takes care of long term retention. The single storage foot print removes redundant and duplicate data.

Typical Implementation:

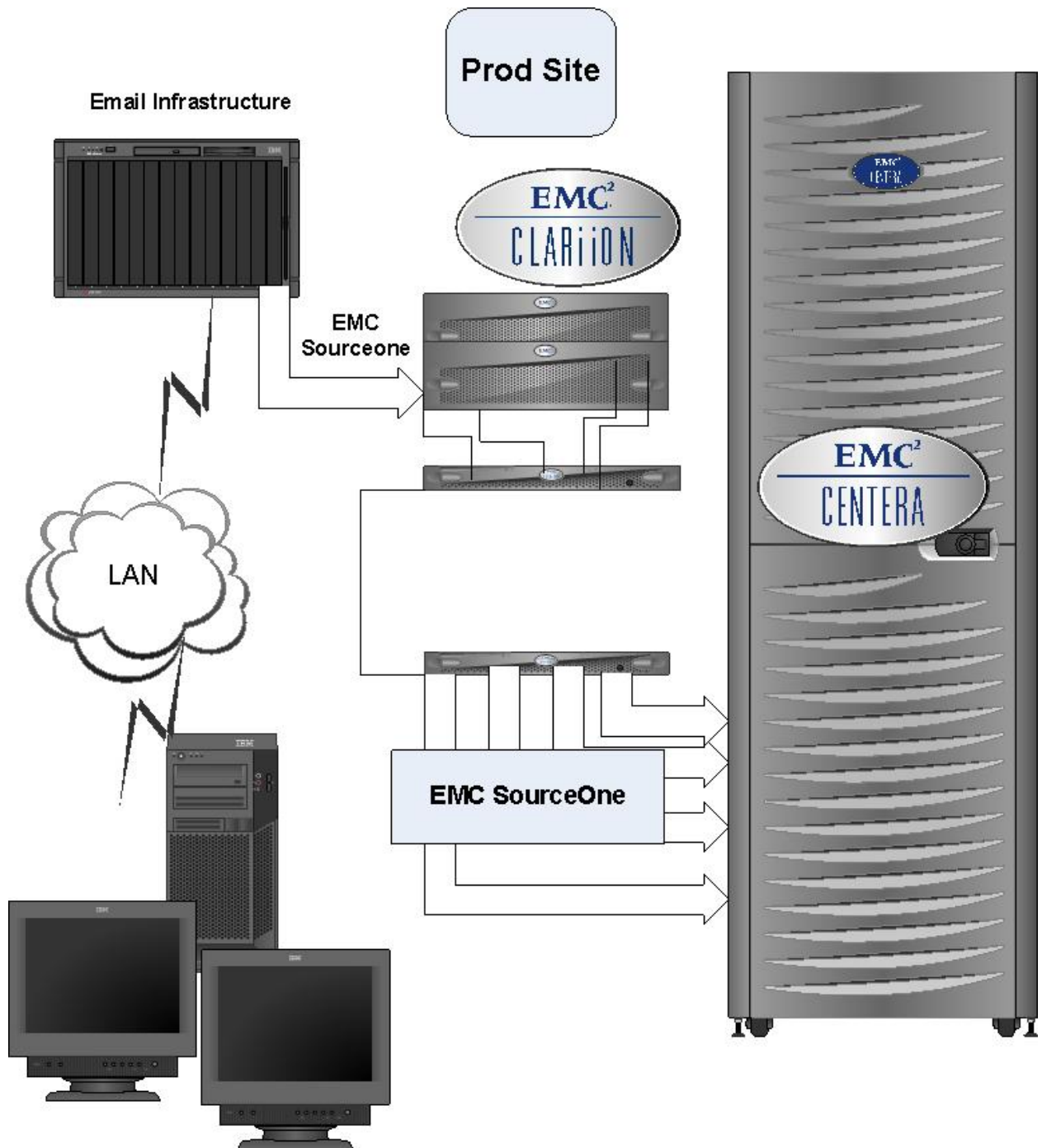Please find below a Typical Implementation of EMC SourceOne with EMC Centera:

Figure 1: Implementation

Here is a sample implementation design coupled with EMC Centera. EMC Centera delivers active archiving with much lower capacity and better power efficiency. It uses Content-Addressable Storage (CAS) as a mechanism to store information that can be retrieved based on its content rather than its storage location.

➢ EMC SourceOne policies check which Email can be archived.

➢ EMC SourceOne transfers the data seamlessly and transparently from your Email Infrastructure to the secondary storage device.

Creating Nearline/Archive Policies

EMC SourceOne makes it very easy to create and test policies via a Graphical User Interface (GUI) environment. The options for inclusion and exclusion are very powerful. Access times, size, dates, age of an email etc., are all available attributes.

Building archive policies is driven by business requirements. The first step to create an effective policy requires that you understand the data you are storing. You can choose to have a different set of policies for different kinds of data.

You can do a "preview" before you apply a policy. This ensures that no data intended to stay on primary storage gets archived. It's often very useful to run several "what if?" scenarios. Administrators can then look at the resulting data and make any necessary changes without affecting production.

The Archive and Recall Process

Let us suppose that a particular Email has been selected for archive due to a policy implemented by EMC SourceOne to copy it to the destination archive device. Metadata is kept for the user to reclaim the archived data. Users still see the same email and size as before but the stub file does not usually take up nearly as much space as the original file.

The administrator can set the following options when a user clicks on an Email:

- Full – All data from the requested file is moved back to primary storage;
- Partial – Only the data blocks requested by the client are moved back to primary storage; or
- Passthrough – Data is sent from the secondary storage to the client and not rewritten to the primary storage.

It's important to note that if a client changes a file, the entire file is moved back to primary storage no matter how you have configured the setting.

Conclusion

EMC SourceOne can significantly reduce primary Data Storage requirements. EMC SourceOne helps organizations address the complex challenges of storing and managing a growing number of Email communications, while reducing risks and providing regulatory and policy compliance, legal discovery, and end-user productivity benefits. EMC SourceOne coupled with EMC Centera provides organizations with single-instance storage, archiving, and long term data preservation. Organizations can significantly save costs by deploying EMC SourceOne to manage and archive Emails.