Enterprise E-mails and File Archival System

Ketan Shevale

E&TC SAOE, Pune, India

Abstract-**Archival data storage plays a critical role in data preservation as almost all current data will eventually be archived. Businesses need a plan to store and archive content, including files, that enables them to search quickly, yet provides data security. Content has to be stored and managed efficiently while providing the ability for users to search the content and retrieve it as needed. Enterprise Archival System addresses archiving issues by providing an integrated solution for archiving emails and files. Enterprise archival system provides an enterprise solution to the archival of email attachments at centralized location. This paper will explain the basics of enterprise archival solutions. It also discusses the feature of Enterprise Archival System. The EAS feature allows archiving and managing the files attached to emails, effective searching through the archived data, reporting of archived data and storage usage**.

Keywords -*Anonymous blacklisting; Privacy; Revocation*

I. INTRODUCTION

In the age of information, e-mail has become the mode of communication. Attachments to e-mail have served as a business solution in the communication process, and they are now a norm in the day-to-day business process. Once information is stored electronically, finding an efficient method of storing, archiving, and managing it efficiently is a challenge. Developing software to archive content by optimizing the use of storage yet providing management simplification was a challenge. It provides a method for quick implementation, a system that is transparent to users with an open application programmable interface (API). We are implementing Enterprise Archival System. Archival of emails and data is an important activity in any Organization. The Enterprise Archival Application can provide a solution to archive this information at a centralized location. This system is the application where you can keep backup of your important emails. To reduce the mail box size, handling complex work with ease, to keep track of several activities, reduce the mail box size. [1]

Enterprise Archival System lets you reduce overall storage requirements, and keep all content fully searchable and instantly accessible to the end user. With an option to retain multiple copies of the same file or document, forward the archived mails, delete mails for end users it really reduces the overload. With Enterprise System Archiving, your business will benefit from the following features:
Reporting-Analysis of active and archived contents.
Archival-Archival of important email attachments (documents) on secondary storage
Retention- Deletion of emails

• Reporting and decision analysis—Analysis of active and archived contents at centralized location.
•Archiving—Flexible, policy-controlled archiving to secondary storage based on criteria
•Retention management—Deletion of files based on user request
 Enterprise email & file Archiving saves critical space on file servers by seamlessly moving files to alternate storage devices without affecting the end user.
Using an enterprise email and file archival the mails and the attachments to mail are searched as well as retrieve transparently without affecting end users. The system also provides an option to the analysis of usage of storage by each system user. The mails are search by the subjects of mails.

A. *Reporting*

Enterprise Archival system provides information about the storage utilization and consumption to the administrators so that they can manage their space efficiently. Archival system also provides reports about the users of system and the storage usage by each user on storage. The Enterprise Archival system reporting engine generates a wealth of information about active data already in the archive.
Examples of reports include:
•Reports about the projects in organization and their storage utilization.
• Storage Summary

B. *Extensive and transparent user search and retrieval*

Enterprise Vault software archives emails, but users retain access to emails transparently. This enables users to instantly search and access archived content

C. *Retention Management*

The system has provision to delete the emails& files archived at centralized location.

D. *Configuring the archive stores*

While databases are not good for storing email (or at least, aren't as good as file-systems at doing so), email is not particularly good for storing in databases, for several reasons. Perhaps the most understandable of these reasons is that email is not relational data. While email does have some structured meta-data (like sender, receiver, subject, date, and so forth), it's general structure is as a blob of unstructured text. Standard files in file-systems have a similar amount of meta-data (name, access times, modify times, owner, etc.), and are similarly unsuited for relational databases. Typical relational databases are designed for storage of large collections of relations between relatively small blocks of data . A typical size for a frequently used database is in the several hundred megabytes.

2027

The volume of email, on the other hand, is typically measured in gigabytes. While there are databases that can handle even multiple terabytes of data (such as Oracle), such databases require serious hardware, complex setup, and significant amounts of administration to keep running at their peak. The meta-data about your email is indeed relational data: it is related to the email itself, and relates email together. This is precisely what relational databases are good at. Some IMAP server use small, simple database files as caches and indexes into email collections. Note that they do not normally store the email itself in a database. Databases can be used to improve the speed of searching mail, but that doesn't mean that a database should be used to store email. A database can be used as a cache of mail meta-data, and thus provide the indexing features generic filesystems lack without all of the penalties of exclusively using a database.

This section will show you how to configure the archive stores where you want Enterprise Mail Archiver to store the archived emails Emails in compressed format. This is stored on disk. Metadata related to the emails, this is always stored in a database. The metadata about the emails stores the subject and the pointer to the file on the file system where the file containing email is stored. When a user want to send data to a list of users, the mail will be stored in files corresponding to each user. Search index entries updated with email data. The search index is always stored on disk.

II.   METHODOLOGY FOR SOLVING THIS PROPOSED THEME

SMTP server is an email server that supports one account per user recognized by the archival system. Any open source SMTP Server (e.g. sendmail) can be used. A module that intercepts the incoming SMTP request and stores the contents of the email in the DB/file system needs to be developed. The database can store the metadata about archived emails (e.g. subject, user name, date of archival).The actual content of the email can be stored on the file system and the file path can be a field in the database table. In addition to the above table for storing archived emails, there will be another table to store information about existing users. The exact schema of various tables needs to be developed. Application server provides a log in for administrator and normal user. Admin can generate several reports about the user's activity. In application server administrator can delete a user and normal user can fire query to search a file or email in system. Normal user can forward archived mail and delete a mail/file from system .user can also download mail/file.
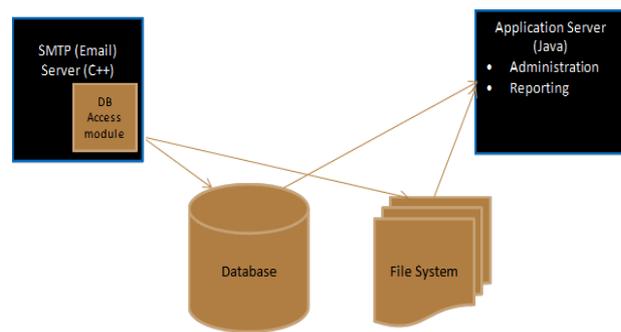
III.   SYSTEM ARCHITECTURE

Figure1.1   System Architecture Diagram of Enterprise Archival System

The Enterprise Archival Application can provide a solution to archive this information at a centralized location.

A.   SMTP Server

This is an email server that supports one account per user recognized by the archival system. Any open source SMTP Server (e.g. sendmail) can be used. A module that intercepts the incoming SMTP request and stores the contents of the email in the DB/file system needs to be developed.

Sendmail is a general purpose internetwork email routing facility that supports many kinds of mail-transfer and -delivery methods, including the Simple Mail Transfer Protocol (SMTP) used for email transport over the Internet.Sendmail to incorporate great flexibility, but it can be daunting to configure for novices.Standard configuration packages delivered with the source code distribution require the use of the M4 macro language which hides much of the configuration complexity. The configuration defines the site-local mail delivery options and their access parameters, the mechanism of forwarding mail to remote sites, as well as many application tuning parameters.Sendmail supports a variety of mail transfer protocols, including SMTP, ESMTP, DECnet's Mail, HylaFax, QuickPage and UUCP.Sendmail itself incorporated a certain amount of privilege separation in order to avoid exposure to security issues. As of 2009[update], current versions of Sendmail, like other modern MTAs, incorporate a number of security improvements and optional features that can be configured to improve security and help prevent abuse.

Filename	Function
/etc/mail/access	sendmail access database file
/etc/mail/aliases	Mailbox aliases
/etc/mail/local-host-names	Lists of hosts sendmail accepts mail for
/etc/mail/mailer.conf	Mailer program configuration
/etc/mail/sendmail.cf	sendmail master configuration file
/etc/mail/virtusertable	Virtual users and domain tables

Table1: send mail configuration files

Sendmail is quite simple and efficient, and so can handle many times more simultaneous messages than more complex systems such as Microsoft's Exchange System or IBM's Lotus Notes system. Also, it is quite easy to tell sendmail to stop listening for new messages if the CPU load is getting too high, preventing small problems running away into big ones. The delivery mechanisms that sendmail uses are very flexible. It can be configured to send messages using any kind of arbitrary program. This lets sendmail route messages for any kind of protocol or medium that tools exist for.[9]

Enterprise contains different groups corresponding to the different projects going on in enterprise. Enterprise archival system provides a solution to maintain the groups information and the data related to each group by providing an account for each group on email server. All members of group has an authority to forward the mails to group and download mails related to their group.

Algorithm for DB Access module:
1. while (true)
2. do_operation()
3. sleep(800)
4. end while

do_operation()
1. Query the database and find all groups and their respective email addresses.
2. Find out offset value for all groups using hash table(hash table contains group email addresses and their offset values).
3. For each email in hash map call function read_and_store(offset).

Read_and_store(int)
1. Construct filename
2. Open the file
3. Seek to offset
4. While(fgets(file_ptr))
5. Copy contents of file into user/group account specific file
6. End.

B.   Database and File System
The database can store the metadata about archived emails (e.g. subject, user name, date of archival). The actual content of the email can be stored on the file system and the file path can be a field in the database table. In addition to the above table for storing archived emails, there will be another table to store information about existing users. The exact schemas of various tables need to be developed.

C.   Application Server
It can provide the following functionality –
Functionality for the admin user
Administrator has the authority for management of user accounts that means creation of new users , deletion of existing users, modification of user account.
-Reports about the total used size and size limit per user.
 Functionality for normal users
-Basic queries to browse through the archive (e.g. query to list subjects of all emails archived between a date ranges).

-Option to delete archived files/emails.

IV.   ADVANCED FEATURES

A.   Removes pressure on Department servers
The proposed system helps to reduce load on department servers as the backup is provided at centralized location.

B.   Reduces mailbox size
With the EAS the emails need not to be keep at mailbox for long time.The important   mails are moved to backup and hence the mailbox size is reduced.

C.   Anywhere access to archived items
The archived items are store at centralized location.

D.   Transparent to users
User can see what are mails and files archived and can query through the archival.

E.   Powerful search for finding archived files
The archived files are efficiently search with an effective search technique.

F.   Mails can be forwared,deleted,replied to.
Use of the proposed system has provision for the forwarding, deleting, replying to the emails.With this options the load on administrator is reduced.

V.   CONCLUSION

Clearly, e-mail use within the corporate environment will only continue to rise. It is estimated that worldwide e-mail traffic increased by 35 percent in 2004, totaling 76.8 billion messages per day. Corporate e-mails accounted for 83 percent of this traffic. If left unchecked, corporate e-mail can leave a business vulnerable. Thus by using this system it is possible to overcome the above situation. EAS system can reduce the mail box size as it stores the path of the file instead of actual file. Data is stored at centralized location any person of organization can access the data if administrator has given him the authority. Mails/Files can be viewed, forwarded easily no need of downloading.

REFERENCES

[1]   Sloan, J.L. Nat. Center for Atmos. Res., Boulder, CO O'Lear, B.T. ,Kitts, D.L ; Irwin,   B.L. MaSSIVE: The Mass Storage System IV Enterprise , IEEE 1993

[2]   DongJin Lee: Measurement for Improving the Design of Commodity Archival Storage Tiers , IEEE 2008

[3]   A Model for Digital Archival of Municipal Documents.

[4]   "Archiving: Finding Data In 2050"July 1, 2008 "The Forrester Wave™:

[5]   Message Archiving Software, Q1 2008" February 14, 2008

[6]   "The Forrester Wave™: Message Archiving Hosted Services

[7]   Magic Quadrant for E-Mail Active Archiving @ Gartner

[8]   Best Practices: Email Archiving by Forrester Research