The Design and Implementation of Email Archiving System Based on J2EE

Hu Hongwei, Yu Ping, Zhou Nan

HAL Id: hal-01420228
https://inria.hal.science/hal-01420228

Submitted on 20 Dec 2016

The Design and Implementation of Email Archiving System Based on J2EE

1 Hu Hongwei, 2 Yu Ping, 3 Zhou Nan

[1] Net Center, China Agricultural University，Beijing 100083,China
[2] Net Center, China Agricultural University，Beijing 100083,China
[3] Net Center, China Agricultural University，Beijing 100083,China
[1] Tel.: 62737471, fax: 62736848
[1] E-mail:huhongwei@cau.edu.cn

Abstract: With the increasingly widespread use of email and the increasing importance of email data, email archiving has become necessary for mail management. In view of this, we propose a solution of the design and implementation of email archiving system based on J2EE. This solution researches and analysis some key technologies of mail archiving such as mail backup, full-text indexing, email recovery an so on first ,and then designs a mail archiving system based on J2EE , resolves the mail real-time backup, full-text search and others.Finally,this system has been applied in actual mail archiving work . Experimental results show that the mail archiving system based on J2EE can achieve real-time email backup and rapid email recovery , this solution can effectively help enterprise users to resolve the problem of email archiving .

Key words:email archiving;J2EE; Journaling; full text search;email recovery

1. Introduction

With the rapid development of information technology and the increasingly widespread use of internet , email has become one of the most widely used communication tools in the world because of it's fast, efficient and low cost.

The increasingly widespread use of e-mail makes the email data more important. In modern societ,email data has become critical information assets for government, enterprises and individuals,and plays an increasingly important role in finanical report, commercial negotiations, statistical analysis , decision support and so on.

However,there are some information security risk in using email because of the failures of the email system and the user's faulty operation.In recent years,many problems frequent frequent occurrence while in using email, such as data corruption,data loss and so on, and it Seriously affecting the use of secure e-mail.

The increasing importance of e-mail data and email security incidents occur frequently has raised new demands for the security and integrity of the mail system.

Therefore, building a mail archiving system which can achieve real-time email backup and rapid email recovery has become an important part of modern email management work.

Email archiving system is a solution which can achieve mail data migration, protection and management target, it contains a plurality of functional modules,such as mail backup, full-text indexing, email recovery, Statistical Analysis and so on.

In order to achieve the target of long-term store users' email data ,the email backup module backup mail data to mail archiving system from mail server using the mail backup technoloy.

In the full-text indexing module, mail content and attachments can be indexed and added to the index database,which can improve the retrieval efficiency of e-mail data.

The mail recovery module achieves the restore function, which can restroe the data message archiving server to the user's mailbox, or export email data file to the users.

The information of mail server and email archiving server can be analyzed by the statistical analysis module , This information can be displayed in various forms, such as charts ,graphics, tables and so on.

Application and research shows that the mail archiving system can resolve some problems ,such as offsite backup,long-term storage, category management, real-time retrieval, Rapid Restore and so on.This makes it possible to long-term store and effective use of email data,it also can help manager to analyze email data and make effective decisions.

Generally speaking, email archiving system currently used broadly divided into three classes: email archiving system based on backup technology;email archiving system based on email gateway system and pure software email archiving system.In this paper,we first study and analyze email archiving technology and J2EE paltform,then we design and implement an email archiving system based on J2EE,than can backup,index,restore the email data,and last this system has been tested ,compared and analyzed.Application and research show this system can effectively solve backup and restore problem of the mass mail,and it provides security guarantee for e-mail application.

2 Analyze of email archiving system

2.1 System Structure Analyze

From a structural point of view,the email archiving system are divided into two main parts,data bankup part and bankuped data manage part.The main function of backup part is to apply the bacukup mechanism to backup the email data,which are received by email server to mail archiving server, and deletes expired backup mail data regularly;the bankuped data manage part achieves the function of bankuped data online query, statistics and restore and so on. The overall structure of the email archiving system is shown in Figure 1.
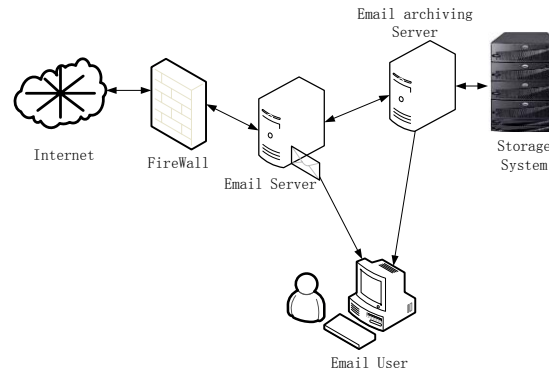

Fig 1 The overall structure of the email archiving system

2.2. System Modules Analyze

From the perspective of functional modules,mail archiving system consists of several major subsystems as follows: backup subsystem, index and backed email query subsystem, restore subsystem, statistical analysis subsystem, user management subsystem and configuration subsystem, as shown in Figure 2. In this paper,we mainly discuss four system: email data backup subsystem,mail index and backed email query system,email restore subsystem, statistical analysis subsystem and user management subsystem.
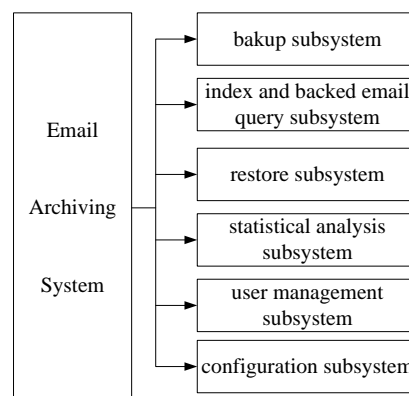

Fig 2 email archiving system functions Framework

3. Implementation of email archiving system

3.1. Email backup subsystem

The mail backup subsystem mainly achieve the function of email backup,and it was dived into three modules ,email data acquisition, data store and backed data management.

In this paper,we apply the journaling method to acquire email data from email server.We first change some email manage system parameter to make the mails which are send and received by mail server are all copied to the journaling mailbox.Then the mails which were copied to the journal mailbox will be dumped into the mail archiving server.And last the email which was backuped will be deleted from the journaling mailbox to reduce the pressure of the mail server.

In the mail archiving system, archived email data was stored in a mixed- methods of database and file system.The meta data of the mail was stored in the database ,the content and the attachment of the mail was stored in the data file.This method has two advantages.on the one hand,it can improve stored efficiency of archived email and meet the demand of managing massive data ,on the other hand,it can achieve fast retrieval of archived messages and impore retrieval efficiency of archived email data.

Taking into account the demand of massive mail data long-term backup and the capacity limitations of the storage system,we conduct a mail backup duplicate checking processing.The email send to multiple users was retain only a single copy.In order to further save storage space, this email archiving system also performed compression process the mail contents and attachments.

In order to achive email backup function,we develop a backup management software which was called "emailbackup",this software apply multiple threads method to read and store email data,and Significantly improves the backup efficiency archived mail.

After setting the parameter of interval time, threads number and backup time of the "mailbackup"software,it will regularly read mail from the journal mailbox and stored data to the email archiving server,and also delete expired backups mail in order to save storage space.

In this paper,we design two methods to set system parameters: the firt method is manual operation, administator can manul set the interval time,threads number and backup time of the "mailbackup"software based on frequencies of email transmission and performance of email archiving server .

3.2. Index and query subsystem

Research shows that a prerequisite for the effective management and use of archiving email is to achieve fast retrieval of archived email. However, because the enterprise-class email archiving system stores tens of millions or even billions of email,so it will reduce the efficiency of retrieval archived email if we adopt a common approach to retrieve archived da,and it will not meet the need fast retrieval of archived email.

In this paper, we design a archiving email full-text indexing system based on Luncene in order to achieve fast retrieval of archived email.

Lucene is an open-source full-text search engine toolkit[1], and has been widely used in many files. After analyzing the performance requirements of emal archving system,we design a full-text retrieval subsystem of archived mail basen on Lucene.

This subsustem consists of two parts.The first part is the module of creating and maintaining archived email index;the seconde part is the module of index database retrieving.The structure of the archived mail full-text retrieval subsystem is shown in figure 3.
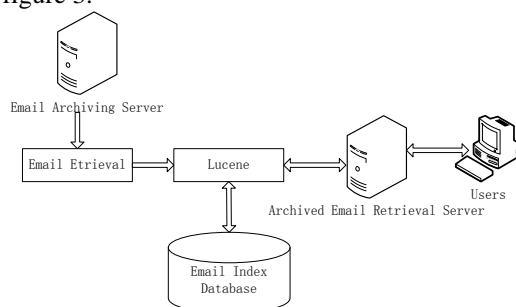
Fig. 3. The structure of the full-text retrieval subsystem

The establishment of archived email index is procude of adding a index record to the index database.

In this paper,we achieve the function of archived email index combined the process of email backup.The email information will be supplied to the email index system when the email was backed up,and then the email index system will achieve full-text and different attributes index by establish index with email header, sender, recipient, delivery time, email content and attachments.

The creation and maintenance of mail index is achieved by three classes of Lucene:indexWriter,document and filed[2].

Index retrieving is the process to retrieve email information from index database based on the web interface supplied by the retrieving system.The retrieving interface of Luene consists of QueryParser,IndexSearcher and Hists[3].

In the full-text retrieval subsystem,we can get the email serial number through email header, sender, recipient, delivery time and other information,than we will get email content and attachment stored in email archiving server.

3.3. Archiving email restore and exprot

Because of the failures of the email system and the user's faulty operation,the user's email often lose, accidentally deleted or damaged issues Therefore,the restore of the archived email to the email user has to be a major function of email archchiving system.

Tmail archchiving system can restore users' email by sending archived email to the users or export email file to users.

In the process of email restore,we can achieve the email restore by JavaMail interface where retrieving the email informtion.

JavaMail is a java interface usede to send,receive,reade mail which was supplied by Sun Company[4], and it has been widely used in email system[5].In this paper,we desgin a email restore subsystem used JavaMail,which was called "MailSender".This subsystem achieve the funtion of restoring email to email management from the email archiving system.

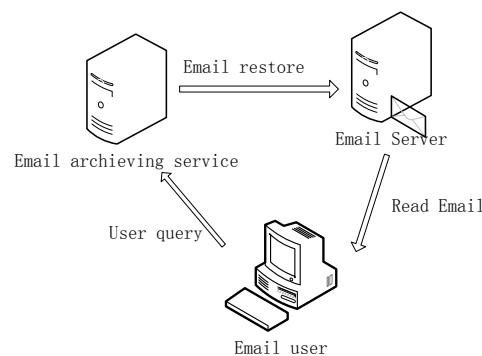The structure of the email restore sybsystem is shown in figure 3.

Fig. 4. The structure of the email restore sybsystem

3.4. User Management and limitation

In the email archiving system,the safe of archieved email is very important.The safe levels of the email archiving system consists of serveral levels, sucn as software level,operation system level,hardware level and network safe level.

In the process of software design,properly setting user limitations is a important work for software level safe.

We dived the users of email archiving system into two categories, comman user and administrator user.

The comm user is matched to the email user,he can query and restore the archived email.

The administrator user is the administrator of the email archiving system,administrator is dived into three categories:the system operator,email auditor and the super administrator.

The responsibilities of system operator is to manage common users and query email basic information,he can manage uses that are from one or some email groups.The system operator can only read email's basic information,such as header, sender, recipient, delivery time and so on, but he can't read email's content.

The email auditor can read email content and audit email informaiton within the scope of relevant laws and regulations,the audit users was assigned the super administrator.

The responsibilities of super administrator is to manage the system operators and email auditors ,and he wil set proper parameter information of the email archiving system.

4. Test and Analyze

In order to test performance of the email archiving system,we carryid out test on a university anda company.The parameters of email archiving system was set below,the threads number of backup system is three and the interval time is thirty seconds.

The thirty day's test shows that the email archiving system can achieve the demand of email real-time bacup,and the full-text index subsystem can accurately retrieve mail from more than ten milion email data in second-class time,and it also achieve the function of restore email data in sub-second time.

Analysis showed that the email archiving system can meet the need of enterprise-class email archiving work

5. Conclusions

Email archiving system has become an important tool in email management work,and plays an important role in the email security and audit work.

In the paper,we design a pure sofware email archiving system based on J2EE after studying the technology of email archiving.The software can work on windows,linux,unix and other operating systems,and can applied to a variety of databases,such as Mysql,Oracle,Sql Server and so on.

The email archiving system designde by this paper can meet the demand of email archiving in the the premise of reducing the dependence on hardware and software environment,it also improves the applicability of email archiving products and takes a new way for the development and promotion of email archiving system.

References

[1].LI Xiao-li,DU Zhen-long Investigation on Personalized Search Engine Based on Lucence, Computer Engineering, Vol.16, Issue 19, 2010, pp. 258-260.

[2].GUAN Jian-he, GAN Jian-feng Design and implementation of web search engine based on Lucene, Computer Engineering and Design, Vol.28, Issue 2, 2007, pp. 489-491.

[3].Bai Kun Geng Guohua Study And Application Of Vertical Search Engine Based On Lucene And Heritrix, Computer Applications and Software, Vol.26, Issue 1, 2009, pp. 258-260.

[4]. ZHU An-qing, WANG Hui-jin, GAO He-fu Design and implementation of message system inMIS based on J2EE, Computer Engineering and Design, Vol.26, Issue 10, 2005, pp. 2775-2777.

[5]. ZHONG Luo , LIU Ling, XIA Hong-xia Development Research of Web Mail System Based on JavaMail API, JOURNAL OF WUHAN UNIVERSITY OF TECHNOLOGY, Vol.28, Issue 6, 2006, pp.84-86.