



Archivemail e Hypermail Archiviazione della posta su Web

Introduzione

Un problema che spesso si presenta a chi gestisce servizi di posta elettronica è rappresentato dai volumi della posta stessa e dalla loro difficile gestione.

Spesso la grande quantità di posta che resta nella INBOX di un utente ne rende difficile la gestione e lenta la consultazione. Ma come risolvere il problema archiviando la posta e permettendo all'utente di consultarsi comodamente l'archivio magari direttamente con il browser?

Il mio obiettivo era quello di poter archiviare (estrarre) della posta da numerose e voluminose mailbox gestite tramite un IMAP server, originariamente in formato MAILDIR, caso piuttosto comune, e permettere agli utenti di accedere comunque con il browser al loro archivio personale.

La mia soluzione, forse artigianale ma comunque efficace, ha sfruttato un paio di programmi presente nel repository di Debian Sarge, ma utilizzabili su qualsiasi piattaforma.

La loro combinazione e qualche piccola istruzione in shell scripting hanno risolto il problema. Le fasi affrontate sono state:

1. Archiviazione della posta da maildir in formato mbox
2. Conversione della posta in formato HTML per la consultazione dell'archivio via Web
3. Configurazione di htdig come motore di ricerca interno alla posta

Vediamo nel dettaglio i passaggi.

Installazione software necessari

I software utilizzati sono due **archivemail** per l'archiviazione vera e propria **hypermail** per la conversione in formato HTML.

Cominciamo con il primo passaggio, l'archiviazione della posta:

```
# apt-get install archivemail
```

```
# archivemail
```

```
Usage: archivemail [options] mailbox [mailbox...]
```

```
Moves old mail in IMAP, mbox, MH or maildir-format mailboxes to an mbox-format mailbox compressed with gzip.
```

```
Options are as follows:
```

```
-d, --days=NUM      archive messages older than NUM days (default: 180)
-D, --date=DATE      archive messages older than DATE
-o, --output-dir=DIR directory to store archives (default: same as original)
-s, --suffix=NAME    suffix for archive filename (default: '_archive')
-S, --size=NUM       only archive messages NUM bytes or larger
-n, --dry-run        don't write to anything - just show what would be done
-u, --preserve-unread never archive unread messages
  --delete           delete rather than archive old mail (use with caution!)
  --include-flagged messages flagged important can also be archived
  --no-compress      do not compress archives with gzip
```



```
--warn-duplicate  warn about duplicate Message-IDs in the same mailbox
-v, --verbose      report lots of extra debugging information
-q, --quiet        quiet mode - print no statistics (suitable for crontab)
-V, --version      display version information
-h, --help         display this message
```

Example: `archivemail linux-kernel`

This will move all messages older than 180 days to a 'mbox' mailbox called 'linux-kernel_archive.gz', deleting them from the original 'linux-kernel' mailbox. If the 'linux-kernel_archive.gz' mailbox already exists, the newly archived messages are appended.

To archive IMAP mailboxes, format your mailbox argument like this:

```
imap://username:password@server/mailbox
```

Website: <http://archivemail.sourceforge.net/>

Come si vede dalle opzioni del programma l'archiviazione è possibile sia su messaggi in formato MBOX che MAILDIR oltre che a connessione attraverso il server IMAP, con tanto di autenticazione.

Le politiche di archiviazione sono varie e specificate da alcune opzioni:

```
-d <num days> → numero di giorni al di sotto dei quali i messaggi vengono archiviati. Ad esempio --d 180 archivia tutti i messaggi più vecchi degli ultimi 180 giorni.
-D <data> → permette di specificare una data da usare come limite dell'archiviazione. Il formato per la data può essere di tipo ISO format ad esempio '2005-12-12', Internet format ad esempio '12 Dec 2005' oppure Internet format con il nome esteso per i mesi ad esempio '12 December 2005'.
-S, --size=NUM → archivia solo messaggi con dimensione superiore ad certo numero di bytes.
-u → evita di archiviare i messaggi non letti o comunque marcati come tale dall'utente.
-o --output-dir=DIR → specifica la directory dove verranno salvate gli archivi in formato MBOX
```

Vediamo in dettaglio l'operazione svolta, partendo da un volume in cui erano archiviati diverse GB di messaggi in formato MAILDIR, gestiti da un server postfix+courier IMAP.

Archiviazione formato Maildir di cartelle e sottocartelle

Proviamo ad archiviare tutti i messaggi più vecchi di 180 giorni per una specifica mailbox.

```
# archivemail -d180 -u --no-compress paolo@test.it/*
paolo@test.it/.: archived 50 of 191 message(s) in 0.4 seconds
paolo@test.it/..: archived 0 of 0 message(s) in 0.0 seconds
paolo@test.it/.Drafts: archived 0 of 1 message(s) in 0.0 seconds
paolo@test.it/.Mailbox: archived 27 of 27 message(s) in 0.1 seconds
paolo@test.it/.Sent: archived 8 of 70 message(s) in 0.1 seconds
paolo@test.it/.Trash: archived 0 of 10 message(s) in 0.0 seconds
```

Con i messaggi in formato maildir sotto la home delle user si formano dei file in formato MBOX con il nome `_archive` che formano il nostro archivio effettivo, sottratto come volume alla mailbox originale:

```
# cd paolo@test.it
# ls -la .*_archive
-rw----- 1 postfix postfix 5996245 2006-10-19 16:29 ._archive
-rw----- 1 postfix postfix 1555971 2006-10-19 16:29 .Mailbox_archive
-rw----- 1 postfix postfix 198633 2006-10-19 16:29 .Sent_archive
```



```
-rw----- 1 postfix postfix 178633 2006-10-19 16:29 .Trash_archive
-rw----- 1 postfix postfix 298633 2006-10-19 16:29 .Draft_archive
```

Questi archive sono per ogni nostra cartellina le mail in formato file unico (MBOX). Il file `._archive` è la nostra INBOX, spesso la cartellina più importante.

A questo punto se si vuole poi passare alla conversione in formato HTML, come detto nell'introduzione, conviene usare il comando `cat` per accorparli in un unico file mailbox per comodità:

```
# cat .*_archive > archvio_TOT
# ls -lha archvio_TOT
-rw-r--r-- 1 root root 37M 2006-10-19 16:44 Archvio_posta
```

Adesso abbiamo il nostro archivio definitivo possiamo passare alla fase successive

Generazione dell'archivio in formato HTML

Adesso che abbiamo il nostro "sorgente" convertiamolo in formato HTML. Per fare questo entra in gioco il programma `hypermail`.

La conversione è semplice ed efficace, anche su volumi di grandi dimensioni.

```
# mkdir /var/www/archivio/paolo@test.it
# hypermail -L it -m paolo@test.it/archvio_TOT -d /var/www/archivio/paolo@test.it -p
```

Le Opzioni ed il suo funzionamento in generale sono molto semplici:

```
-L it → specifica di usare la traduzione italiana
-m → indica il file in formato MBOX da utilizzare
-d → indica la directory in cui esportare l'archivio in file HTML
-p → è la modalità verbose
```

Sullo schermo si potrà seguire tranquillamente la conversione per vedere se tutto va per il meglio.

```
# hypermail -L it -m paolo@test.it/archvio_TOT -d /var/www/archivio/paolo@test.it
-Caricamento casella di posta "paolo@test.it/archvio_TOT"...
Creazione directory "www//att-0001", modalità 755.
 1 Creato il file per l'allegato www//att-0001/ZTL_def.ppt Creazione directory "www//file-0003", modalità 755.
 3 Creato il file per l'allegato www//att-0003/01-part Creazione directory "www//file-2004", modalità 755.
 4 Creato il file per l'allegato www//att-0004/01-part Creazione directory "www//file-1005", modalità 755.
10 Creato il file per l'allegato www//att-0010/copertina.doc Creazione directory "www//att-0014", modalità 755.
14 Creato il file per l'allegato www//att-0014/riunione.doc Creazione directory "www
Creazione directory "www//att-0288", modalità 755.
292 Creato il file per l'allegato www//att-0292/Scheda_intervento_PISL.doc
295 messaggi.
Scrittura messaggi su "www/"...
Scrittura indice delle date su "www/date.html"...
Scrittura indice delle discussioni su "www/index.html"...
Scrittura indice degli oggetti su "www/subject.html"...
Scrittura indice degli autori su "www/author.html"...
bytes "www/attachment.html"...
```

Il log mostrato è solo un esempio in pratica, viene creata una pagina per ogni messaggio con tanto di link agli allegati, il tutto ordinabile per autore, data o allegato.

Vediamo un esempio di quello che si ottiene:



archivio_TOT Per discussione

289 messaggi: *Inizio* gio ott 10 2002 - 11:32:54 CEST, *Fine* mer apr 19 2006 - 13:44:01 CEST

Received on: [Messaggi più recenti](#)

Other periods: [[autore](#)] [[data](#)] [[oggetto](#)] [[allegato](#)]

- [Re: richiesta per il materiale 10001.pdf](#) erwan@maggiola.it (gio ott 10 2002 - 12:32:54 CEST)
- [Ciao, ciao. "L'unico controllo & archivio"](#) erwan@maggiola.it (gio ott 10 2002 - 12:36:50 CEST)
- [materiali Tra ciao](#) erwan@maggiola.it (gio ott 10 2002 - 12:37:49 CEST)
- [ciao Erwan](#) erwan@maggiola.it (gio ott 10 2002 - 12:38:44 CEST)
- [Re: richiesta](#) erwan@maggiola.it (sab ott 14 2002 - 08:43:19 CEST)
- [ciao Erwan](#) erwan@maggiola.it (ven ott 18 2002 - 11:51:09 CEST)
- [L'unico controllo & archivio - 1." richiesta](#) erwan@maggiola.it (sab ott 21 2002 - 10:46:21 CEST)
- [Ciao, ciao. "L'unico controllo & archivio"](#) erwan@maggiola.it (gio ott 24 2002 - 16:25:56 CEST)
- [Re: Erwan](#) erwan@maggiola.it (sab ott 24 2002 - 16:34:23 CEST)
- [L'unico controllo & archivio - 4." richiesta](#) erwan@maggiola.it (mar nov 29 2002 - 08:45:52 CEST)
- [FBI ciao](#) erwan@maggiola.it (gio nov 11 2002 - 13:02:43 CEST)
- [FBI ciao. "il mio controllo"](#) erwan@maggiola.it (sab nov 04 2002 - 17:05:53 CEST)
- [FBI ciao. "proposta & chat"](#) erwan@maggiola.it (mar nov 05 2002 - 08:53:09 CEST)
 - [Re: FBI ciao. "proposta & chat"](#) Valentina.Morici (mar nov 05 2002 - 08:58:40 CEST)
- [ciao FBI ciao](#) erwan@maggiola.it (mar nov 05 2002 - 17:10:01 CEST)
- [FBI ciao. "materiali & chat"](#) erwan@maggiola.it (mar nov 06 2002 - 08:02:30 CEST)
- [FBI ciao. "materiali & chat"](#) erwan@maggiola.it (mar nov 06 2002 - 12:01:33 CEST)
- [FBI ciao. "il mio chat"](#) erwan@maggiola.it (gio nov 07 2002 - 12:06:57 CEST)
- [FBI ciao](#) erwan@maggiola.it (ven nov 08 2002 - 11:03:23 CEST)
- [FBI ciao. "chat"](#) erwan@maggiola.it (sab nov 09 2002 - 16:01:10 CEST)

Automatizzare la conversione

Si capisce facilmente come si possa archiviare la posta di n mailbox, utilizzando dei semplici script bash o perl.

Vediamo ad esempio una serie di script da lanciare nella directory delle mailbox da archiviare:

```
# Archivia le email dalle mailbox
#!/bin/bash
cd /root/archivia_posta

for i in `ls`
do
    if [[ -d $i ]]
    then
        archivemail -d 180 -u --no-compress $i/*
    fi
done
```

```
#raggruppa in unica mailbox per ogni utente
for i in `ls`
do
    if [[ -d $i ]]
    then
        cat $i/*_archive >>$i/archivio_TOT
    fi
done
```

```
#process le mailbox in formato HTML
```



```
#!/bin/bash
cd /root/archivia_posta
for i in `ls`
do
if [[ -d $i ]]
then
mkdir $i/archive
hypermail -L it -m $i/archivio_TOT -d $i/archive -p
fi
done
```

Sono forse banali, ma comunque perfettamente aderenti al mio problema. Basta lanciarli in sequenza (magari con qualche adattamento) per avere la conversione automatica delle mailbox in formato HTML.

Archiviazione tramite IMAP

Nel caso in cui si voglia archiviare la posta di un altro sistema, che fornisca comunque il servizio IMAP, si può usare una sintassi apposta, molto utile nel caso in cui le mailbox non siano sul sistema gestito direttamente da noi.

```
#archivemail -d10 imap://paolo@test.it:testpwd@mail.test.it/*. * -o archivio/
```

Ad esempio con questo sistema potete archiviare la posta di un server Exchange. Ovviamente diventa più lento per via del download della posta ma le funzioni restano le stesse.

Script di Archiviazione

Vediamo infine un semplice script che effettua l'archiviazione notturna della nostra posta, in modo da avere sempre una mailbox snella.

```
#!/bin/sh
#Backup di posta con archivemail: archivia la posta più vecchia di una settimana

#file temp
File_RAPPORTO=rapporto.txt

#touch per file temp
touch $File_RAPPORTO

#Archivio della Posta
archivemail -d7 -u -o /archivio_posta /usr/local/virtual/paolo@test.it/*. * >$ File_RAPPORTO

#Invio rapporto per email
cat $REPORT_FILE | mail -s "Archivio Posta $(date'+%d/%m/%y')" paolo@netlink.it
```

Letture da ThunderBird

Un'altra interessante funzione di questo metodo di archiviazione è rappresentata dal fatto che copiando l'archivio informato mbox sotto la cartella locale di Thunderbird i nostri messaggi compaiono magicamente. Come se fossero stati scaricati via pop3 in modo selettivo.



Si può infatti scompattare l'archivio nel local Folder di ThunderBird il file verrà letto come cartella con all'interno i suoi messaggi.

Note Finali

Si può facilmente intuire l'utilità di questo metodo di archiviazione. La posta archiviata secondo i nostri criteri di filtro può essere data in consultazione via web, oppure masterizzata su un CD, zippata oppure aperta con un client di posta come TBird.

Nel caso in cui la voleste tenere su web, divisa magari per directory potete usare **Htdig** per indicizzare la directory ed avere anche un motore di ricerca per cercare tra i messaggi.

Ho trovato molto comodo l'abbinamento di archivemail con hypermail, archiviando volumi davvero grandi di posta, utenti con GB di posta tanto per intenderci, ed il sistema si è rivelato efficace e abbastanza veloce da costituire uno strumento completo di archiviazione. Per chi avesse esigenze più complesse di esportazione in formato HTML può usare **Mhonarc** al posto di hypermail, altro ottimo programma utile a questa funzione.

Chi avesse risolto il problema dell'archiviazione e della consultazione della posta, utilizzando altri strumenti opensource c'è lo faccia sapere, gli daremo spazio su SI.

*Doc: **archive_mail.pdf***

Dott. Paolo PAVAN [Netlink Sas]- pavan@netlink.it

Con il supporto di **Alberto Vincenzi** -alberto.vincenzi@gmail.com

Data: Settembre 2006

Note finali

- Il presente documento è a semplice scopo divulgativo
- L'autore non si assume la responsabilità di eventuali danni diretti o indiretti derivanti dall'uso dei programmi, o dall'applicazione delle configurazioni menzionate nel seguente articolo
- I marchi citati sono di proprietà dei rispettivi proprietari e sono stati utilizzati solo a scopo didattico o divulgativo.
- Il documento viene rilasciato sotto Licenza Creative Commons.
- Sono possibili errori o imprecisioni, segnalatemele a pavan@netlink.it
- Chi volesse integrare il presente documento, può scrivere a pavan@netlink.it.