Bzip2 mini-HOWTO

David Fetter, *david@fetter.org* `mailto:david@fetter.org` v2.00, 22 agosto 1999

Questo documento descrive come usare il nuovo programma di compressione bzip2. Il documento sgml originale è *qui* `http://fetter.org/Bzip2-HOWTO/` . Traduzione a cura di Gerardo Di Iorio, arete(at)luxna2.na.infn.it, e revisione a cura di Claudio Cattazzo, claudio(at)pluto.it.

Indice

1 Introduzione

Bzip2 è un nuovo algoritmo per la compressione dei dati. Generalmente crea file che sono il 60-70% della dimensione dei corrispondenti ottenuti usando gzip.

Questo documento vi guiderà attraverso alcune comuni applicazioni per bzip2.

Le future versioni di questo documento conterranno applicazioni della libreria C libbzip2 gentilmente scritta dall'autore di bzip2, *Julian Seward* `mailto:Julian_Seward@muraroa.demon.co.uk` . Il manuale di bzip2, che include le informazioni di basso livello della libreria, può essere trovato *qui* `http://www.bzip2.org/bzip2/docs/manual_toc.html` .

Le future versioni di questo documento potranno includere un riassunto del dibattito su come e se usare bzip2 nel kernel di Linux.

1.1 Cronologia delle revisioni

1.1.1 v2.00

Cambiata la sezione 5 (Usare bzip2 con less) in modo che i file .tar.bzip2 possano essere realmente letti. Grazie a *Nicola Fabiano* `mailto:mc8644@mclink.it` per la correzione.

Aggiornata l'utilità buzzit.

Aggiornate le informazioni su tar.

1.1.2 v1.92

Aggiornata la sezione 2.2 (Ottenere i binari di bzip2), inclusi quelli di S.u.S.E.

1.1.3 v1.91

Corretto un carattere e chiariti alcuni idiomi della shell 4 (nella sezione sull'uso di bzip2 con tar). Grazie per questo ad Alessandro Rubini.

Aggiornato lo strumento buzzit in modo da non danneggiare l'archivio originale bzip2.

Aggiunto bgrep, uno strumento simile a zgrep.

1.1.4 v1.9

Chiarito il problema di gcc 2.7.* Grazie a Ulrik Dickow per averlo indicato.

Aggiunto l'elegante sistema di Leonard Jean-Marc per lavorare con tar.

Aggiunta la traduzione in Svedese di Linus Ãkerlund.

Corretta la sezione wu-ftpd con i suggerimenti di Arnaud Launay.

Spostate le traduzioni in una loro sezione.

1.1.5 v1.8

Buzzit e tar.diff sono stati messi nell'sgml a cui appartengono. Corretta la punteggiatura e la formattazione. Grazie a Arnaud Launay per il suo aiuto nella correzione della mia copia. :-)

Il progetto xv è stato eliminato per il momento a causa della mancanza di interesse popolare.

Aggiunti dei rompicapo per le future versioni del documento.

1.1.6 v1.7

Aggiunta l'utilità buzzit. Corretta la patch per GNU tar.

1.1.7 v1.6

Aggiunto il Netscape enabler di TenThumbs.

Modificato anche lesspipe.sh secondo i suoi suggerimenti. Adesso dovrebbe funzionare meglio.

1.1.8 v1.5

Aggiunta la traduzione francese di Arnaud Launay e il suo file per wu-ftpd.

1.1.9 v1.4

Aggiunta la traduzione giapponese di Tetsu Isaji.

1.1.10 v1.3

Aggiunto .emacs di Ulrik Dickow per 19.30 e superiori.

(Su suo consiglio, corretta anche la patch jka-compr.el per emacs. Oops! Bzip2 non ha ancora(?) un'opzione append.)

1.1.11 v1.2

Modificata la patch per emacs così i file .bz2 vengono riconosciuti automaticamente.

1.1.12 v1.1

Aggiunta patch per emacs.

1.1.13 v1.0

Prima versione.

2 Ottenere bzip2

La pagina di Bzip2 si trova presso il *sito inglese* `http://www.bzip2.org/` . Il mirror negli Stati Uniti è *qui* `http://www.digistar.com/bzip2/index.html` .

2.1 Bzip2-HOWTO nella propria lingua

Coloro che parlano francese potrebbero riferirsi ai documenti in francese di Arnaud Launay. La versione web è *qui* `http://www.freenix.fr/linux/HOWTO/mini/Bzip2.html` e si può usare ftp *qui* `ftp://ftp.lip6.fr/pub/linux/french/docs/HOWTO/mini/Bzip2.gz` . Arnaud può essere contattato attraverso la posta elettronica a *questo indirizzo* `mailto:zoro@mygale.org` .

Chi parla giapponese può far riferimento al documento giapponese di Testsu Isaji *qui* `http://jf.gee.kyoto-u.ac.jp/JF/JF.html` . Isaji può essere contattato attraverso la *sua pagina* `http://www2s.biglobe.ne.jp/~kaien/` o attraverso la posta elettronica a *questo indirizzo.* `mailto:isaji@mxu.meshnet.or.jp`

Chi parla svedese può far riferimento ai documenti in svedese di Linus Ãkerlund *qui* `http://user.tninet.se/~uxm165t/linux_doc.html` . Linus può essere contattato con la posta elettronica a

questo indirizzo. `mailto:uxm165t@tninet.se`

2.2 Ottenere i binari precompilati di bzip2

Vedere i siti.

2.3 Ottenere i sorgenti di bzip2

Sono reperibili dal sito ufficiale (per sapere dove si veda 2 (Ottenere Bzip2)).

2.4 Compilare bzip2 per la propria macchina

Se si ha gcc 2.7.*, cambiare la riga

```
CFLAGS = -O3 -fomit-frame-pointer -funroll-loops
```

in

```
CFLAGS = -O2 -fomit-frame-pointer
```

ovvero, sostituire la parte -O3 con -O2 e togliere -funroll-loops. Si potrebbe anche voler aggiungere i parametri -m* usati durante la compilazione del kernel (come -m486, ad esempio).

La parte più importante consiste nell'evitare -funroll-loops, poiché genera codice errato su molte versioni di gcc 2.7 e codice più lento e grande su tutte le versioni di gcc 2.7. Per gli altri compilatori (lcc, egcs, gcc 2.8.x) vanno bene le opzioni CFLAGS predefinite.

Dopo di questo basta eseguire make e installarlo come spiegato nel README.

3 Usare bzip2 da solo

Leggere la pagina di manuale :)

4 Usare bzip2 con tar

Sono elencati di seguito tre modi per usare bzip2 con tar, ovvero:

4.1 Il più facile da impostare

Questo metodo non richiede alcuna configurazione. Per estrarre un archivio tar bzip2, foo.tar.bz2, nell'attuale directory, fare

```
/percorso/per/bzip2 -cd foo.tar.bz2 | tar xf -
```

o

```
tar --use-compress-prog=bzip2 xf foo.tar.bz2
```

Questo metodo funziona ma può essere un po' scomodo da scrivere spesso.

4.2 Facile da impostare, molto facile da usare, nessuna necessità dei privilegi di root

Grazie a *Leonard Jean-Marc* `mailto:leonard@sct1.is.belgacom.be` per il suggerimento. Grazie anche ad *Alessandro Rubini* `mailto:rubini@morgana.systemy.it` per la distinzione tra bash e csh.

Nel proprio .bashrc si può mettere una riga simile a questa:

```
alias btar='tar --use-compress-program /usr/local/bin/bzip2 '
```

Nel proprio .tcshrc o .cshrc, la riga analoga è simile a questa:

```
alias btar 'tar --use-compress-program /usr/local/bin/bzip2'
```

4.3 Sempre facile da usare, ma necessita di accesso da root

Aggiornare la propria versione di tar a quella più nuova di GNU, che attualmente è la 1.13.10. Può essere trovata nel *sito ftp GNU* `ftp://alpha.gnu.org/gnu/tar/` o in qualunque mirror.

5 Usare bzip2 con less

Per decomprimere al volo file bzip2, ovvero usare less senza prima decomprimerli, si può creare un file lesspipe.sh (man less) come questo:

```
#!/bin/sh
# Questo è un preprocessore per 'less'. Viene usato quando questa
# variabile d'ambiente esiste:   LESSOPEN="|lesspipe.sh %s"
```

```
  case "$1" in
  *.tar) tar tvvf $1 2>/dev/null ;; # Mostra il contenuto di vari file tar
  *.tgz) tar tzvvf $1 2>/dev/null ;;
# Questo funziona per la versione non modificata di tar:
  *.tar.bz2) bzip2 -cd $1 $1 2>/dev/null | tar tvvf - ;;
# Questo funziona con la versione con patch di tar:
# *.tar.bz2) tyvvf $1 2>/dev/null ;;
  *.tar.gz) tar tzvvf $1 2>/dev/null ;;
  *.tar.Z) tar tzvvf $1 2>/dev/null ;;
  *.tar.z) tar tzvvf $1 2>/dev/null ;;
  *.bz2) bzip2 -dc $1  2>/dev/null ;; # Mostra correttamente i file compressi
  *.Z) gzip -dc $1  2>/dev/null ;;
  *.z) gzip -dc $1  2>/dev/null ;;
  *.gz) gzip -dc $1  2>/dev/null ;;
  *.zip) unzip -l $1 2>/dev/null ;;
  *.1|*.2|*.3|*.4|*.5|*.6|*.7|*.8|*.9|*.n|*.man) FILE=`file -L $1` ; # groff src
    FILE=`echo $FILE | cut -d ' ' -f 2`
    if [ "$FILE" = "troff" ]; then
      groff -s -p -t -e -Tascii -mandoc $1
    fi ;;
  *) cat $1 2>/dev/null ;;
#  *) FILE=`file -L $1` ; # Controlla se è un binario e se lo è lo mostra con 'strings'
#    FILE1=`echo $FILE | cut -d ' ' -f 2`
#    FILE2=`echo $FILE | cut -d ' ' -f 3`
#    if [ "$FILE1" = "Linux/i386" -o "$FILE2" = "Linux/i386" \
#         -o "$FILE1" = "ELF" -o "$FILE2" = "ELF" ]; then
#      strings $1
#    fi ;;
  esac
```

6 Usare bzip2 con emacs

6.1 Cambiare emacs per tutti

Ho scritto la seguente patch per jka-compr.el che aggiunge bzip2 all'auto-compression-mode.

Avvertenza: io l'ho solo provata con emacs-20.2 ma non c'è nessuna ragione per credere che un simile approccio non funzioni con altre versioni.

Per usarla,

1. Andare nella directory dei sorgenti emacs-20.2/lisp (oppure dove li si ha estratti con tar)

2. Mettere la seguente patch in un file chiamato jka-compr.el.diff (deve essere l'unico file ;).

3. Fare

   ```
   patch < jka-compr.el.diff
   ```

4. Avviare emacs e fare

   ```
   M-x byte-compile-file jka-compr.el
   ```

5. Uscire da emacs.
6. Mettere il proprio jka-compr.elc originale in un posto sicuro in caso di bachi.
7. Sostituirlo con il nuovo jka-compr.elc.
8. Buon divertimento!

```
--- jka-compr.el        Sat Jul 26 17:02:39 1997
+++ jka-compr.el.new    Thu Feb  5 17:44:35 1998
@@ -44,7 +44,7 @@
 ;; The variable, jka-compr-compression-info-list can be used to
 ;; customize jka-compr to work with other compression programs.
 ;; The default value of this variable allows jka-compr to work with
-;; Unix compress and gzip.
+;; Unix compress and gzip.  David Fetter added bzip2 support :)
 ;;
 ;; If you are concerned about the stderr output of gzip and other
 ;; compression/decompression programs showing up in your buffers, you
@@ -121,7 +121,9 @@


 ;;; I have this defined so that .Z files are assumed to be in unix
-;;; compress format; and .gz files, in gzip format.
+;;; compress format; and .gz files, in gzip format, and .bz2 files,
+;;; in the snappy new bzip2 format from http://www.muraroa.demon.co.uk.
+;;; Keep up the good work, people!
 (defcustom jka-compr-compression-info-list
   ;;[regexp
   ;; compr-message  compr-prog  compr-args
@@ -131,6 +133,10 @@
      "compressing"    "compress"     ("-c")
      "uncompressing"  "uncompress"   ("-c")
      nil t]
+    ["\\.bz2\\'"
+     "bzip2ing"        "bzip2"         ("")
+     "bunzip2ing"      "bzip2"         ("-d")
+     nil t]
    ["\\.tgz\\'"
      "zipping"        "gzip"         ("-c" "-q")
      "unzipping"      "gzip"         ("-c" "-q" "-d")
```

6.2 Modificare emacs per una sola persona

Un ringraziamento per questo va a Ulrik Dickow, *ukd@kampsax.dk* `mailto:ukdATkampsax.dk` , programmatore di sistema presso Kampsax Technology.

Per fare in modo di usare bzip2 automaticamente, senza essere un amministratore di sistema, aggiungere ciò che segue nel proprio file .emacs.

```
;; Automatic (un)compression on loading/saving files (gzip(1) and similar)
;; We start it in the off state, so that bzip2(1) support can be added.
```

```
;; Code thrown together by Ulrik Dickow for ~/.emacs with Emacs 19.34.
;; Should work with many older and newer Emacsen too.  No warranty though.
;;
(if (fboundp 'auto-compression-mode) ; Emacs 19.30+
    (auto-compression-mode 0)
  (require 'jka-compr)
  (toggle-auto-compression 0))
;; Now add bzip2 support and turn auto compression back on.
(add-to-list 'jka-compr-compression-info-list
             ["\\.bz2\\(~\\|\\.~[0-9]+~\\)?\\'"
              "zipping"        "bzip2"         ()
              "unzipping"      "bzip2"         ("-d")
              nil t])
(toggle-auto-compression 1 t)
```

7 Usare bzip2 con wu-ftpd

Ringrazio Arnaud Launay per questo salva banda. Ciò che segue dovrebbe andare in /etc/ftpconversions per comprimere e decomprimere al volo con bzip2. Assicurarsi che i percorsi (come /bin/compress) siano giusti.

```
 :.Z:  :  :/bin/compress -d -c %s:T_REG|T_ASCII:O_UNCOMPRESS:UNCOMPRESS
 :   : :.Z:/bin/compress -c %s:T_REG:O_COMPRESS:COMPRESS
 :.gz: :  :/bin/gzip -cd %s:T_REG|T_ASCII:O_UNCOMPRESS:GUNZIP
 :   : :.gz:/bin/gzip -9 -c %s:T_REG:O_COMPRESS:GZIP
 :.bz2: :  :/bin/bzip2 -cd %s:T_REG|T_ASCII:O_UNCOMPRESS:BUNZIP2
 :   : :.bz2:/bin/bzip2 -9 -c %s:T_REG:O_COMPRESS:BZIP2
 :   : :.tar:/bin/tar -c -f - %s:T_REG|T_DIR:O_TAR:TAR
 :   : :.tar.Z:/bin/tar -c -Z -f - %s:T_REG|T_DIR:O_COMPRESS|O_TAR:TAR+COMPRESS
 :   : :.tar.gz:/bin/tar -c -z -f - %s:T_REG|T_DIR:O_COMPRESS|O_TAR:TAR+GZIP
 :   : :.tar.bz2:/bin/tar -c -y -f - %s:T_REG|T_DIR:O_COMPRESS|O_TAR:TAR+BZIP2
```

8 Usare bzip2 con grep

La seguente utilità, che chiamo bgrep, è una leggera modifica di zgrep che viene distribuito con Linux. La si può utilizzare per poter usare grep con i file senza decomprimerli con bunzip2.

```
#!/bin/sh

# bgrep -- un wrapper del programma grep che decomprime file
PATH="/usr/bin:$PATH"; export PATH

prog=`echo $0 | sed 's|.*/||'`
case "$prog" in
        *egrep) grep=${EGREP-egrep}     ;;
        *fgrep) grep=${FGREP-fgrep}     ;;
        *)      grep=${GREP-grep}       ;;
esac
```

```
pat=""
while test $# -ne 0; do
  case "$1" in
  -e | -f) opt="$opt $1"; shift; pat="$1"
           if test "$grep" = grep; then  # grep è bacato con -e su SVR4
             grep=egrep
           fi;;
  -*)      opt="$opt $1";;
   *)      if test -z "$pat"; then
             pat="$1"
           else
             break;
           fi;;
  esac
  shift
done

if test -z "$pat"; then
  echo "Fa il grep di file bzip2"
  echo "Uso: $prog [opzioni_grep] modello [file]"
  exit 1
fi

list=0
silent=0
op=`echo "$opt" | sed -e 's/ //g' -e 's/-//g'`
case "$op" in
  *l*) list=1
esac
case "$op" in
  *h*) silent=1
esac

if test $# -eq 0; then
  bzip2 -cd | $grep $opt "$pat"
  exit $?
fi

res=0
for i do
  if test $list -eq 1; then
    bzip2 -cdfq "$i" | $grep $opt "$pat" > /dev/null && echo $i
    r=$?
  elif test $# -eq 1 -o $silent -eq 1; then
    bzip2 -cd "$i" | $grep $opt "$pat"
    r=$?
  else
    bzip2 -cd "$i" | $grep $opt "$pat" | sed "s|^|${i}:|"
    r=$?
  fi
```

```
  test "$r" -ne 0 && res="$r"
done
exit $res
```

9 Usare bzip2 con Netscape sotto X

tenthumbs@cybernex.net dice:

> Ho trovato un modo per usare da Linux Netscape bzip2 per Content-Encoding nello stesso modo con cui viene usato gzip. Aggiungere quanto segue a $HOME/.Xdefaults oppure $HOME/.Xresources. Io uso l'opzione -s perché preferisco sacrificare un po' di velocità di decompressione piuttosto che la memoria RAM. Si può togliere questa opzione se non se ne ha bisogno.

```
Netscape*encodingFilters:      \
        x-compress :  : .Z     : uncompress -c  \n\
        compress   :  : .Z     : uncompress -c  \n\
        x-gzip     :  : .z,.gz : gzip -cdq      \n\
        gzip       :  : .z,.gz : gzip -cdq      \n\
        x-bzip2    :  : .bz2   : bzip2 -ds \n
```

10 Usare bzip2 per ricomprimere altri formati di compressione

Il seguente programma perl prende file compressi in altri formati (.tar.gz, .tgz, .tar.Z e .Z) e li rimpacchetta per una migliore compressione. Il sorgente perl ha tutti i tipi di documentazione su cosa fa e come lo fa. Quest'ultima versione riceve i file come input dalla riga di comando. Senza argomenti da riga di comando tenta di ricomprimere tutti i file contenuti nell'attuale directory.

```
#!/usr/bin/perl -w

#######################################################
#                                                     #
# Questo programma prende i programmi compressi       #
# nella directory attuale e li converte nel formato   #
# bzip2. Gestisce l'etensione .tgz in un modo         #
# ragionevole, producendo un file .tar.bz2.           #
#                                                     #
#######################################################
$counter = 0;
$saved_bytes = 0;
$totals_file = '/tmp/machine_bzip2_total';
$machine_bzip2_total = 0;

@raw = (defined @ARGV)?@ARGV:<*>;

foreach(@raw) {
    next if /^bzip/;
    next unless /\.(tgz|gz|Z)$/;
    push @files, $_;
```

```
}
$total = scalar(@files);

foreach (@files) {
    if (/tgz$/) {
        ($new=$_) =~ s/tgz$/tar.bz2/;
    } else {
        ($new=$_) =~ s/\.g?z$/.bz2/i;
    }
    $orig_size = (stat $_)[7];
    ++$counter;
    print "Ricompressione $_ ($counter/$total)...\n";
    if ((system "gzip -cd $_ |bzip2 >$new") == 0) {
        $new_size = (stat $new)[7];
        $factor = int(100*$new_size/$orig_size+.5);
        $saved_bytes += $orig_size-$new_size;
        print "$new è circa il $factor% della dimensione di $_. :",($factor<100)?')':'(',"\n";
        unlink $_;
    } else {
        print "Arrgghh! È successo qualcosa a $_: $!\n";
    }
}
print "Si sono "
    , ($saved_bytes>=0)?"guadagnati ":"persi "
    , abs($saved_bytes)
    , " byte di spazio :"
    , ($saved_bytes>=0)?")":"("
    , "\n"
    ;

unless (-e '/tmp/machine_bzip2_total') {
    system ('echo "0" >/tmp/machine_bzip2_total');
    system ('chmod', '0666', '/tmp/machine_bzip2_total');
}


chomp($machine_bzip2_total = `cat $totals_file`);
open TOTAL, ">$totals_file"
     or die "Impossibile aprire il totale su tutto il sistema: $!";
$machine_bzip2_total += $saved_bytes;
print TOTAL $machine_bzip2_total;
close TOTAL;

print "Sono stati guadagnati un totale di ",`cat $totals_file`," byte su tutta la macchina.\n";
```