Bzip2 Howto

David Fetter, *david@fetter.org* <mailto:david@fetter.org>
Vertaald door: Ellen Bokhorst *bokkie@nl.linux.org* <mailto:bokkie@nl.linux.org> v2.00 Zo Aug 22 1999

Dit document vertelt je hoe je het nieuwe bzip2 compressie-programma kunt gebruiken. De lokale kopie van de sgml <sgml/Bzip2-HOWTO.sgml> staat momenteel op de huidige site, en de sgml van de "auteur" staat op <http://fetter.org/Bzip2-HOWTO/>.

Inhoudsopgave

1 Introductie

Bzip2 is een blits nieuw algoritme voor het comprimeren van gegevens. Het maakt gewoonlijk bestanden aan, die uit 60-70% van de grootte van hun met gzip gecomprimeerde tegenhangers bestaan.

Dit document leidt je door een paar algemene applicaties voor bzip2.

Toekomstige versies van dit document zullen de applicaties van libbzip2 en de bzip2 C library behandelen, waarvan de auteur *Julian Seward* `<mailto:Julian_Seward@muraroa.demon.co.uk>` zo vriendelijk is geweest, ze te schrijven. De bzip2 handleiding, met low-level informatie over de library kan worden gevonden bij `<http://www.bzip2.org/bzip2/docs/manual_toc.html>`.

Toekomstige versies van het document kunnen ook een samenvatting bevatten van de bespreking over of (en hoe) bzip2 in de Linux kernel zou moeten worden gebruikt.

1.1 Revisie Historie

1.1.1 v2.00

De sectie 5 (Gebruiken van bzip2 met less) is gewijzigd, aangezien .tar.bzip2 bestanden nu werkelijk kunnen worden gelezen. Met dank aan Nicola Fabiano `<mailto:mc8644@mclink.it>` voor de correctie.

De buzzit utility bijgewerkt.

De informatie over tar bijgewerkt.

1.1.2 v1.92

De sectie 2.2 (Verkrijgen van bzip2 binaries) bijgewerkt, waaronder het toevoegen van die van S.u.S.E.

1.1.3 v1.91

Een typo gecorrigeerd en een aantal uitdrukkingen van de shell verduidelijkt in de 4 (sectie over het gebruiken van bzip2 met tar). Met dank aan Alessandro Rubini hiervoor.

De buzzit tool bijgewerkt niet op het originele bzip2-archief te stampen.

bgrep toegevoegd, een op zgrep lijkende tool.

1.1.4 v1.9

Het gcc 2.7.* probleem verduidelijkt. Met dank aan Ulrik Dickow om me hierop te wijzen.

De elegante wijze van Leonard Jean-Marc om met tar te werken, toegevoegd.

De Zweedse vertaling van Linus Ãkerlund toegevoegd.

De wu-ftpd sectie door een suggestie van Arnaud Launay gecorrigeerd.

Vertalingen naar een eigen sectie verplaatst.

1.1.5 v1.8

buzzit en tar.diff in de sgml geplaatst waar ze horen. Leestekens en formattering hersteld. Met dank aan Arnaud Launay voor zijn hulp bij het corrigeren van mijn kopie. :-)

xv project laten vervallen vanwege gebrek aan voldoende belangstelling.

lokartikelen toegevoegd voor toekomstige versies van dit document.

1.1.6 v1.7

buzzit utility toegevoegd.

De patch hersteld met het oog op gnu tar.

1.1.7 v1.6

Toegevoegd TenThumbs' Netscape enabler.

Door zijn suggestie ook lesspipe.sh gewijzigd. Het zou nu beter moeten werken.

1.1.8 v1.5

De Franse vertaling van Arnaud Launay, en zijn wu-ftpd bestand toegevoegd.

1.1.9 v1.4

De Japanse vertaling van Tetsu Isaji toegevoegd.

1.1.10 v1.3

Toegevoegd Ulrik Dickow's .emacs voor 19.30 en hoger.

(Ook de jka-compr.el patch voor emacs door zijn suggestie. Oeps! Bzip2 heeft nog(?) geen äppend"flag.)

1.1.11 v1.2

De patch voor emacs gewijzigd zodat het automatisch .bz2 bestanden herkent.

1.1.12 v1.1

Een patch voor emacs toegevoegd.

1.1.13 v1.0

Ronde 1.

2 Verkrijgen van bzip2

De homepage van Bzip2 homepage bevindt zich op *De UK homesite* `<http://www.bzip2.org/>`. De mirrorsite in de United States is *hier* `<http://www.digistar.com/bzip2/index.html>`.

2.1 Bzip2-HOWTO in je eigen taal

Mensen die Frans spreken willen zich misschien wenden tot de Franse documenten van Arnaud Launay. De webversie is te vinden op `<http://www.freenix.fr/linux/HOWTO/mini/Bzip2.html>`, en je kunt ftp `<ftp://ftp.lip6.fr/pub/linux/french/docs/HOWTO/mini/Bzip2.gz>` gebruiken. Er kan met Arnaud contact worden gezocht via elektronische mail op `<mailto:zoro@mygale.org>`

Mensen die Japans spreken willen zich misschien wenden tot de Japanse documenten van Tetsu Isaji `<http://jf.gee.kyoto-u.ac.jp/JF/JF.html>`. Isaji kan worden bereikt via zijn homepage op `<http://www2s.biglobe.ne.jp/~kaien/>`, of via elektronische mail op `<mailto:isaji@mxu.meshnet.or.jp>`

Mensen die Zweeds spreken willen zich misschien wenden tot de Zweedse documenten van Linus Ãkerlund `<http://user.tninet.se/~uxm165t/linux_doc.html>`. Linus kan worden bereikt via elektronische mail op `<mailto:uxm165t@tninet.se>`

2.2 Verkrijgen van de voorgecompileerde binaries van bzip2

Zie de homesites

2.3 Verkrijgen van de sources van bzip2

Ze zijn afkomstig van de Officiële sites (zie 2 (Verkrijgen van Bzip2) voor waar.

2.4 bzip2 voor je computer compileren

Als je gcc 2.7.* hebt, wijzig dan de regel waarin staat

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

in

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

dat wil zeggen, vervang -O3 door -O2 en laat -funroll-loops weg. Misschien dat je ook enige -m* flags toe wilt voegen (zoals bijvoorbeeld -m486), die je gebruikt als je kernels aan het compileren bent.

Het voorkomen van de -funroll-loops is het belangrijkste gedeelte, aangezien dit ervoor zorgt dan vele gcc 2.7's de verkeerde code genereren, en alle gcc 2.7's genereren langzamere en grotere code. Voor andere compilers (lcc, egcs, gcc 2.8.x) zijn de standaard CFLAGS prima.

`make` en installeer het daarna gewoon volgens de README.

3 Gebruiken van bzip2 op zichzelf

Lees de Prima Manual page :)

4 bzip2 met tar gebruiken

Hieronder staan drie manieren om bzip2 met tar te gebruiken, te weten

4.1 Gemakkelijkst in te stellen:

Deze methode vereist in 't geheel geen setup. Om het met bzip2 ingepakte tar-archief in de huidige directory uit te pakken, tik je in

```
/path/naar/bzip2 -cd foo.tar.bz2 | tar xf -
```

of

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

Deze werken, maar het kan een PITA zijn om het vaak te typen.

4.2 Makkelijk in te stellen, tamelijk makkelijk te gebruiken, geen root-privileges nodig:

Met dank aan *Leonard Jean-Marc* `<mailto:leonard@sct1.is.belgacom.be>` voor de tip. Ook met dank aan *Alessandro Rubini* `<mailto:rubini@morgana.systemy.it>` voor onderscheid te maken tussen bash en csh.

Je kunt als volgt een regel in .bashrc plaatsen:

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

De analoge regel ziet er in .tcshrc, of .cshrc ongeveer zo uit:

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

4.3 Ook makkelijk te gebruiken, maar hiervoor heb je root-toegang nodig.

Update je tar tot de laatste versie van GNU, die is op het moment 1.13.10. Het kan worden gevonden bij *de ftpsite van GNU* `<ftp://alpha.gnu.org/gnu/tar/>` of iedere mirror.

5 bzip2 met less gebruiken

Om bzip2 bestanden al doende te decomprimeren, d.w.z. om "less"te kunnen gebruiken zonder ze eerst te decomprimeren, kun je als volgt een lesspipe.sh (man less) maken:

```
#!/bin/sh
# Dit is een preprocessor voor 'less'. Het wordt gebruikt wanneer de
# omgevingsvariabele ervan is ingesteld:
# LESSOPEN="|lesspipe.sh %s"

  case "$1" in
  *.tar) tar tvvf $1 2>/dev/null ;; # Bekijk de inhoud van diverse tar-bestanden
  *.tgz) tar tzvvf $1 2>/dev/null ;;
# Deze werkt voor de ongewijzigde versie van tar:
  *.tar.bz2) bzip2 -cd $1 $1 2>/dev/null | tar tvvf - ;;
# Deze werkt voor die versie van tar waarop de patch is toegepast:
# *.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 ;; # Bekijk gecomprimeerde bestanden
  *.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` ; # Controleer of het een binary is, zo ja -- bekijken met '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 Bzip2 met emacs gebruiken

6.1 emacs voor iedereen wijzigen:

Ik heb de volgende patch geschreven voor jka-compr.el welke bzip2 aan de auto-compression-mode toevoegt.

Disclaimer: Ik heb dit alleen met emacs-20.2 getest, maar ik heb geen reden om aan te nemen, dat een vergelijkbare aanpak niet met andere versies zou werken.

Om het te gebruiken,

1. Ga je naar de source-directory van emacs-20.2/lisp (waar je het ooit hebt uitgepakt)
2. Plaats de hieronder staande patch in een bestand met de naam jka-compr.el.diff (dit zou het enige in dat bestand moeten zijn ;).
3. Tik in

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

4. Start emacs, en tik in

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

5. Ga uit emacs.
6. Verplaats je originele jka-compr.elc naar een veilige plek voor het geval van bugs.
7. Vervang het door de nieuwe jka-compr.elc.
8. Veel plezier!

```
--- 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 @@
 ;; De variabele, jka-compr-compression-info-list kan worden gebruikt om
 ;; jka-compr aan te passen dat het met andere compressieprogramma's werkt.
 ;; De standaardwaarde van deze variabele staat om jka-compr te laten werken met
-;; Unix compress en gzip.
+;; Unix compress en gzip. David Fetter voegde bzip2 ondersteuning toe :)
 ;;
 ;; Als je je ongerust maakt over de stderr uitvoer van gzip en andere
 ;; compressie/decompressie programma's die in je buffers terecht komt, je
@@ -121,7 +121,9 @@


 ;;; Ik heb dit zodanig gedefinieerd in de veronderstelling dat .Z bestanden
 ;;; in unix
-;;; compress formaat zijn; en .gz bestanden in gzip formaat.
+;;; compress formaat zijn; en .gz bestanden in gzip formaat, en .bz2 bestanden,
+;;; in het snelle nieuwe bzip2 formaat van http://www.muraroa.demon.co.uk.
+;;; Ga door met het goede werk, mensen!
 (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 emacs voor één persoon wijzigen:

Met dank hiervoor aan Ulrik Dickow, *ukd@kampsax.dk* `<mailto:ukdATkampsax.dk>`, Systeem Programmeur bij Kampsax Technology:

Om het zodanig te maken dat je bzip2 automatisch kunt gebruiken als je geen systeembeheerder bent, voeg je het volgende toe aan je .emacs bestand.

```
;; Automatische (de)compressie bij het laden/bewaren van bestanden
(gzip(1) en gelijken)
;; We starten het in de uit stand, zodat bzip2(1) ondersteuning kan worden
;; toegevoegd.
;; Bij elkaar gezette code door Ulrik Dickow voor ~/.emacs met Emacs 19.34.
;; Zou ook met vele oudere en nieuwere versies van Emacs moeten werken.
;; Echter geen garantie.
;;
(if (fboundp 'auto-compression-mode) ; Emacs 19.30+
    (auto-compression-mode 0)
  (require 'jka-compr)
  (toggle-auto-compression 0))
;; Voeg nu bzip2 ondersteuning toe en zet auto compressie weer aan.
(add-to-list 'jka-compr-compression-info-list
             ["\\.bz2\\(~\\|\\.~[0-9]+~\\)?\\'"
              "zipping"       "bzip2"         ()
              "unzipping"     "bzip2"         ("-d")
              nil t])
(toggle-auto-compression 1 t)
```

7 bzip2 met wu-ftpd gebruiken

Met dank aan Arnaud Launay voor deze bandbreedte besparing. Het volgende zou in /etc/ftpconversions geplaatst moeten worden om tijdens het werken compressies en decompressies met bzip2 uit te voeren. Zorg er wel voor dat de paths (zoals /bin/compress) juist zijn.

```
 :.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 bzip2 met grep gebruiken

Het volgende utility dat ik bgrep noem, is een kleine aanpassing op zgrep dat met Linux wordt meegeleverd. Je kunt het gebruiken om, met grep, bestanden te doorzoeken zonder ze eerst met bzip2 te decomprimeren.

```
#!/bin/sh

# bgrep -- een omhulsel rondom een grep-programma waarmee bestanden zonodig
```

```
worden gedecomprimeerd
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 zit vol fouten met -e op 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 "doorzoek bzip2 bestanden met grep"
  echo "usage: $prog [grep_options] pattern [files]"
  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 bzip2 met Netscape onder X gebruiken

tenthumbs@cybernex.net zegt:

> Ik vond ook een manier om bzip2 met Linux Netscape te gebruiken voor Content-Encoding net als het gzip gebruikt. Voeg dit toe aan $HOME/.Xdefaults of $HOME/.Xresources:

Ik gebruik de optie -s omdat ik eerder wat snelheid bij het decomprimeren zou verruilen tegen het gebruik van RAM. Je kunt de optie achterwege laten als je dat wilt.

```
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 bzip2 gebruiken om andere compressieformaten opnieuw te comprimeren

Het volgende perl-programma accepteert gecomprimeerde bestanden in andere formaten (.tar.gz, .tgz. .tar.Z, en .Z) en pakt ze opnieuw in voor betere compressie. De perl-source heeft alle soorten handige documentatie over wat het doet en hoe het dit doet. Deze laatste versie accepteert bestanden als invoer op de commando-regel. Zonder commando-regel argumenten, probeert het ieder bestand in de huidige werkdirectory opnieuw in te pakken.

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

#########################################################
#                                                       #
# Dit programma zet ingepakte en met gzip gecomprimeerde #
# programma's in de huidige directory om in het bzip2    #
# formaat. Het handelt de .tgz extensie op een redelijke #
# manier af door er een .tar.bz2 bestand van te maken.   #
#                                                       #
```

```
##########################################################
$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 "Repacking $_ ($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 is about $factor% of the size of $_. :",($factor<100)?')':'(',"\n";
        unlink $_;
    } else {
        print "Arrgghh!  Er is iets gebeurd met $_: $!\n";
    }
}
print "Je hebt "
    , ($saved_bytes>=0)?"bespaard ":"verloren "
    , abs($saved_bytes)
    , " bytes aan opslagruimte :"
    , ($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 "Kan system-wide totaalniet openen: $!";
$machine_bzip2_total += $saved_bytes;
print TOTAL $machine_bzip2_total;
close TOTAL;

print "Dat is een totaal van ",`cat $totals_file`," bytes bespaard.\n";
```