Bzip2 Howto

作者:
David Fetter, *dfetter@best.com* <mailto:dfetter@best.com>
譯者:
謝崑中 HSIEH KUN-CHUNG, 台灣 中華民國 Taiwan, Republic of China (NOT People's Republic of China) (電子郵件 *linuxer.bbs@cis.nctu.edu.tw* 或 *linux4tw@geocities.com* 網頁 "Linux for Taiwan 台灣人的 Linux" *http://members.xoom.com/linux4tw/*
版本 v1.92，1998年 8月18日

這份文件在教你如何使用新的 bzip2 壓縮程式。目前的 sgml 版本被放在 EURO 這裏 <Bzip2-HOWTO.sgml>，還有 "作者用的" sgml 檔放在這兒 <http://www.best.com/~dfetter/Bzip2-HOWTO/Bzip2-HOWTO.sgml>。

目錄

1 簡介

Bzip2 是個很棒的壓縮資料用之新型演算法。大致上能夠達到以它的對手 gzip 所產出的檔案之 60-70% 左右的壓縮率。

這份文件將教你幾個 bzip2 的應用軟體。

本文件的未來版將會討論即將出現的 bzip2 程式庫，該程式庫 原bzip2 的作者 *Julian Seward* `<mailto:Julian_Seward@muraroa.demon.co.uk>` 描述如下:

現在正致力於 `bzip2` 的下一個版本，它也將採用相同的 `.bz2` 檔案格式；
主要增加的功能是一個像 `zlib` 的程式庫。讓我們可以從程式裏頭來讀取和書寫資料到這種格式的檔案裏。

本文件的未來版本也會包含是否應該 (以及如何)讓 Linux kernel 使用 bzip2之議題的結論。

1.1 沿革歷史

1.1.1 v1.92 更新 2.2 (取得 bzip2 二進位程式碼) 章節，加入 S.u.S.E.的。

1.1.2 v1.91

在 4 (使用 bzip2 配合 tar 那章裏)修正一個範例以及 弄清楚一些 shell 句子。感謝 Alessandro Rubini。

更新 buzzit 工具不會去 stomp 原來的 bzip2 壓縮檔。

增 bgrep, 像 zgrep-的工具。

1.1.3 v1.9

澄清 gcc 2.7.* 的問題。感謝 Ulrik Dickow 的熱心指出。

加入 Leonard Jean-Marc 很棒的配合 tar

加上 Linus 攀erlund 的瑞典文版本。

依照 Arnaud Launay 的建議修正了 wu-ftpd 章節

把譯文放到它們自己的位章節上。

1.1.4 v1.8

把 buzzit 和 tar.diff 放到 sgml 文件裏它們應該被放置的地方。 修正拼字與格式。感謝 Arnaud Launay 的協助我修正這份文件。:-)

殺掉 xv 計畫，因此大家不感興趣。

加上 teasers 爲本文件的未來版本。

1.1.5 v1.7

加入 buzzit 公用程式。修正對 gnu tar 的修補。

1.1.6 v1.6

加入 TenThumbs 的 Netscape 致能者。

以及依據它的建議更改 lesspipe.sh。它現在可以工作的更好了。

1.1.7 v1.5

加入 Arnaud Launay 的法文譯本，以及他的 wu-ftpd 檔。

1.1.8 v1.4

加入 Tetsu Isaji 的日文譯本。

1.1.9 v1.3

加入 Ulrik Dickow 給 19.30 或更高版本用的 .emacs 檔

(還有修正給 emacs用的 jka-compr.el 修補檔。糟了! Bzip2 還沒有(?) "append" 的旗標選項。)

1.1.10 v1.2

加入對 emacs 的修補，讓它可以自動地辨識出 .bz2 檔。

1.1.11 v1.1

加入對 emacs

1.1.12 v1.0

第一回合。

2 取得 bzip2

Bzip2 的主網頁在 英國的主網站 <http://www.muraroa.demon.co.uk/>。 美國的映射站在 這裏 <http://www.digistar.com/bzip2/index.html>。你也可以在 Red Hat 的ftp 站裏 找到 <ftp://ftp.redhat.com/pub/contrib>。

2.1 Bzip2-HOWTO 各種語言版本

說法語的人可以看看 Arnaud Launay 的法文文件。 網頁版本在 這裏 <http://www.freenix.fr/linux/HOWTO/mini/Bzip2.html>，你可以用 ftp 來取得 <ftp://ftp.lip6.fr/pub/linux/french/docs/HOWTO/mini/Bzip2.gz> 你可以用電子郵件 <mailto:zoro@mygale.org>來連絡Arnaud

說日語的人可以看看 Tetsu Isaji的 日文譯本 <http://jf.gee.kyoto-u.ac.jp/JF/JF.html>。 你可以用上 Isaji的 網站 <http://www2s.biglobe.ne.jp/~kaien/>，或是用 電子郵件 <mailto:isaji@mxu.meshnet.or.jp>的方式。

說瑞典語的朋友可以參閱 Linus 攀erlund 的瑞典文譯本 按這 <http://user.tninet.se/~uxm165t/linux_doc.html>。你可以用 電子郵件 <mailto:uxm165t@tninet.se> 與 Linus 連絡。

2.2 取得 bzip2 未編譯的二進位檔

查一下主網站。

Debian 的Intel 二進位程式碼在 這裏 <ftp://ftp.debian.org/debian/dists/stable/main/binary-i386/utils/bzip2_0.1pl2-5.deb>。

Red Hat的 alpha 二進位程式碼在 這裏 <ftp://ftp.redhat.com/pub/redhat/redhat-5.1/alpha/RedHat/RPMS/bzip2-0.1pl2-1.alpha.rpm>。

Red Hat的 Intel 二進位程式碼在 這裏 <ftp://ftp.redhat.com/pub/redhat/redhat-5.1/i386/RedHat/RPMS/bzip2-0.1pl2-1.i386.rpm>。

Red Hat的 SPARC 二進位程式碼在 這裏 <ftp://ftp.redhat.com/pub/redhat/redhat-5.1/sparc/RedHat/RPMS/bzip2-0.1pl2-1.sparc.rpm>。

Slackware的 Intel 二進位程式碼在 這裏 <ftp://www.cdrom.com/pub/linux/slackware-3.5/slakware/a1/bzip2.tgz>。

S.u.S.E.的 Intel 二進位程式碼在 這裏 `<ftp://ftp.suse.com/pub/SuSE-Linux/5.2/suse/ap1/bzip.rpm>` 。

你也可以從許多相類似的映射站得取得檔案。

2.3 取得 bzip2 原始程式

你可以從官方的網站 (看 2 (Getting Bzip2)，或是 Red Hat 也有 它 `<ftp://ftp.redhat.com/pub/contrib/SRPMS/bzip2-0.1pl2-1.src.rpm>`)。

2.4 編譯 bzip2 給你的機器使用

如果你有 **gcc 2.7.***，將下面這行

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

改成

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

也就是，把 -O2 改成 -O3 並把 -funroll-loops刪掉。你也可以加入任何你在編譯 kernel 時的 -m* 旗標 (例如像 -m486)。

避免使用 -funroll-loops 是最重要的了，因爲它會造成許多的 gcc 2.7產生錯誤的程式碼，以及全部的 gcc 2.7 產生速度慢又較胖的程式碼。 對其他的編譯器 (lcc, egcs, gcc 2.8.x) 使用內定的 CFLAGS 就沒問題。

之後，只需 `make` 它，然後依照 README 來安裝它。

3 使用 bzip2

請參閱很棒的手冊 :)

4 配合 tar來使用 bzip2

下面有三種配合 tar 使用 bzip2 的方法，也就是

4.1 容易設立:

這個方法跟本就不必做設定。執行解開bzip2'd 的 tar包裏， 在現行目錄底下有個 foo.tar.bz2，執行

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

成功了，但可能有常需輸入的小缺點。

4.2 容易設立，容易使用，不需要 root 的特權:

感謝 *Leonard Jean-Marc* <mailto:leonard@sct1.is.belgacom.be> 貢獻這個小技巧。也謝謝 *Alessandro Rubini* <mailto:rubini@morgana.systemy.it> 從 csh裏分類 出 bash 的。

在你的 .bashrc裏，你可以加入像下面這行:

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

在你的 .tcshrc 或 .cshrc裏，可用下面這行相同的:

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

4.3 同樣是容易使用，但需要 root 來存取。

用下面來方法來修補 gnu tar 1.12：

```
cd tar-1.12/src; patch < /path/to/tar.diff
```

編譯、安裝它，你已準備就緒了。用"which tar" 和 "which bzip2"來確認這兩個 tar 和 bzip2 有在你的 $PATH 裏。要使用新的 tar，只需

```
tar xyf foo.tar.bz2
```

來壓縮檔案。

要製作新的壓縮檔，同樣地:

```
tar cyf foo.tar.bz2 檔名1 檔名2 檔名3...目錄1 目錄2...
```

下面是修補的地方:

```
*** tar.c       Thu Jun 11 00:09:23 1998
--- tar.c.new   Thu Jun 11 00:14:24 1998
***************
*** 196,201 ****
--- 196,203 ----
    {"block-number", no_argument, NULL, 'R'},
    {"block-size", required_argument, NULL, OBSOLETE_BLOCKING_FACTOR},
    {"blocking-factor", required_argument, NULL, 'b'},
+   {"bzip2", required_argument, NULL, 'y'},
+   {"bunzip2", required_argument, NULL, 'y'},
    {"catenate", no_argument, NULL, 'A'},
    {"checkpoint", no_argument, &checkpoint_option, 1},
    {"compare", no_argument, NULL, 'd'},
***************
*** 372,377 ****
--- 374,380 ----
                PATTERN                at list/extract time, a globbing PATTERN\n\
```

```
    -o, --old-archive, --portability   write a V7 format archive\n\
        --posix                        write a POSIX conformant archive\n\
+   -y, --bzip2, --bunzip2             filter the archive through bzip2\n\
    -z, --gzip, --ungzip               filter the archive through gzip\n\
    -Z, --compress, --uncompress       filter the archive through compress\n\
        --use-compress-program=PROG    filter through PROG (must accept -d)\n"),
***************
*** 448,454 ****
     Y  per-block gzip compression */

  #define OPTION_STRING \
!   "-01234567ABC:F:GK:L:MN:OPRST:UV:WX:Zb:cdf:g:hiklmoprstuvwxz"

  static void
  set_subcommand_option (enum subcommand subcommand)
--- 451,457 ----
     Y  per-block gzip compression */

  #define OPTION_STRING \
!   "-01234567ABC:F:GK:L:MN:OPRST:UV:WX:Zb:cdf:g:hiklmoprstuvwxyz"

  static void
  set_subcommand_option (enum subcommand subcommand)
***************
*** 805,810 ****
--- 808,817 ----
        case 'X':
        exclude_option = 1;
        add_exclude_file (optarg);
+       break;
+
+       case 'y':
+       set_use_compress_program_option ("bzip2");
        break;

        case 'z':
```

5 使用 bzip2 來配合 less

馬上解壓縮 bzip2 檔，例如，在不必先 bunzip2 檔案的前提下即可用 "less" 到它們身上，你可以做個 lesspipe.sh (man less)，像下面：

```
#!/bin/sh
# This is a preprocessor for 'less'.  It is used when this environment
# variable is set:   LESSOPEN="|lesspipe.sh %s"
```

```
  case "$1" in
  *.tar) tar tvvf $1 2>/dev/null ;; # View contents of various tar'd files
  *.tgz) tar tzvvf $1 2>/dev/null ;;
# This one work for the unmodified version of tar:
  *.tar.bz2) bzip2 -cd $1 $1 2>/dev/null | tar tzvvf - ;;
#This one works with the patched version of 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 ;; # View compressed files correctly
  *.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` ; # Check to see if binary, if so -- view with '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 使用 bzip2 來結合 emacs

6.1 爲每個人改變 emacs:

我寫了下面的給 jka-compr.el 用的修補，它會加入 bzip2 成爲自動壓縮模式。

除外條款: 我只在 emacs-20.2上試過這個，但是我也找不到有 啥麼理由來相信同樣的方法卻不能在其他版本上使用的。

使用它，

1. 到 emacs-20.2/lisp 原始程式碼目錄裏 (看你在哪兒 untarred 它的)
2. 把下面的修補做成叫 jka-compr.el.diff 的檔案(全放在這檔案裏)。
3. 然後執行

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

4. 啓動 emacs，再執行

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

5. 離開 emacs。
6. 把它原來的 jka-compr.elc 放到安全的地方以防有臭虫時備用。
7. 用新的 jka-compr.elc來代替它。
8. 好好玩吧!

```
--- 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 爲某人改變 emacs:

這次要謝謝 Ulrik Dickow, *ukd@kampsax.dk* <mailto:ukdATkampsax.dk>，他是Kampsax Technology的系

統程式設計師：

了讓非 sysadmin 的朋友可以自動地使用 bzip2，只需在你的 .emacs 檔案裏加上下列東東。

```
;; 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 用 bzip2 來配合 wu-ftpd

感謝 Arnaud Launay 這則節省頻寬的秘訣。下面的內容應該是放到 /etc/ftpconversions 裏面來達成以 bzip2 來做即時壓縮及解壓縮。請確認路徑 (像 /bin/compress) 要正確。

```
 :.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 使用 bzip2 來配合 grep

下面的公用程式，我管它叫 bgrep，事實上是把伴隨在 Linux 的 zgrep 做個小更改 而已。你可以使用它來 grep，而不需先 bunzip2 檔案。

```
#!/bin/sh

# bgrep -- a wrapper around a grep program that decompresses files as needed
```

```
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 is buggy with -e on 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 "grep through bzip2 files"
  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 配合 X 下的 Netscape 來使用。

tenthumbs@cybernex.net 說:

> 我找到一個可以讓 Linux Netscape 來使用 bzip2 做爲 Content-Encoding 就如同它使用 gzip 一樣。把下面的東東加到 $HOME/.Xdefaults 或是 $HOME/.Xresources 我使用 -s 選項，是因爲選擇放棄解壓縮的速度來換得 RAM 的佔用。 你可以把它設成你自己要的。

```
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 使用 bzip2 來再壓縮其他的壓縮格式

下面的 perl 程式會把以下列的壓縮格式檔 (.tar.gz, .tgz. .tar.Z, 和 .Z for this iteration) 重新打包成最佳的壓縮值。這 perl 原始程式都有全方位的好文件 來說明它做什麼以及它如何辦到的。

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

#######################################################
#                                                     #
# This program takes compressed and gzipped programs  #
# in the current directory and turns them into bzip2  #
# format.  It handles the .tgz extension in a         #
# reasonable way, producing a .tar.bz2 file.          #
#                                                     #
#######################################################
```

```
$counter = 0;
$saved_bytes = 0;
$totals_file = '/tmp/machine_bzip2_total';
$machine_bzip2_total = 0;

while(<*[Zz]>) {
    next if /^bzip2-0.1pl2.tar.gz$/;
    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!  Something happened to $_: $!\n";
    }
}
print "You've ",
      ($saved_bytes>=0)?"saved":"lost",
      " $saved_bytes bytes of storage space :",
       ($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 "Can't open system-wide total: $!";
$machine_bzip2_total += $saved_bytes;
print TOTAL $machine_bzip2_total;
close TOTAL;
```

```
print "That’s a machine-wide total of ",‘cat $totals_file‘," bytes saved.\n";
```

11 中譯本

譯者註: 若您發現本文件內容有翻譯不順暢或錯誤的地方，請您來信告訴我，以做更正，謝謝。 譯者電子郵件 *linuxer.bbs@cis.nctu.edu.tw* 或 *linux4tw@geocities.com*

譯註：這份 HOWTO 文件的中文版可在台灣 Linux 使用者組織下的中文 Linux 文件計畫 <http://www.linux.org.tw/CLDP/>
或是"Linux for Taiwan 台灣人的 Linux "*http://members.xoom.com/linux4tw/* 下取得。