

Bzip2 HOWTO

David Fetter dfetter@best.com

Русский перевод Ilgiz Kalmetev ilgiz@mail.rb.ru

v1.92 Tue Aug 18 1998

Этот документ рассказывает, как использовать новую программу сжатия bzip2. Локальная копия sgml в настоящее время располагается в Bzip2-HOWTO.sgml, а ведущий sgml - на <http://www.best.com/~dfetter/Bzip2-HOWTO/Bzip2-HOWTO.sgml>.

Содержание

1 Введение	1
2 Получение bzip2	3
3 Использование bzip2	4
4 Использование bzip2 со tar	4
5 Использование bzip2 с less	6
6 Использование bzip2 с emacs	6
7 Использование bzip2 с wu-ftpd	8
8 Использование bzip2 с grep	8
9 Использование bzip2 с Netscape под X.	9
10 Использование bzip2 для повторного сжатия других форматов компрессии.	10

1 Введение

Bzip2 - это совершенно новый алгоритм для сжатия данных. Обычно он позволяет достичь 60-70%-ного сжатия файлов по сравнению с gzip. Этот документ подскажет вам как связать некоторые прикладные программы с Bzip2.

В будущих версиях документа будет обсуждаться библиотека bzip2, которую автор bzip2'a Julian Seward < [Mailto: Julian_Seward@muraroa.demon.co.uk](mailto:Mailto:Julian_Seward@muraroa.demon.co.uk) > описывает следующим образом:

Я работаю над следующей версией bzip2, который использует тот же самый Bz2 формат файла; основное добавление - zlib-подобная библиотека для чтения и записи данных в таком же формате внутри программ.

Будущие версии документа могут также включать резюме обсуждения, должен ли (и как) bzip2 использоваться в ядре Linux.

1.1 Хронология версий

1.1.1 V1.92

Обновлен раздел "Получение скомпилированного bzip2", включен S.u.S.E.

1.1.2 V1.91

Исправлены tyro и разъяснены некоторые идиомы shell в разделе о использовании bzip2 с tar. Благодарю Alessandro Rubini за это.

В первоначальном архиве bzip2 утилита buzzit заменена на stomp.

Добавлен bgrep - zgrep-подобная утилита.

1.1.3 V1.9

Объяснены проблемы с gcc 2.7.* . Благодарю Ulrik Dickow за указания.

Добавлен элегантный способ работы с tar от Leonard Jean-Marc.

Добавлен шведский перевод от Linus Gkerlund.

Исправлен раздел wu-ftpd в соответствии с предложением Arnaud Launay's.

Переводы перемещены в их собственный раздел.

1.1.4 v1.8

buzzit и tar.diff в sgml перемещены на их место. Исправлены пунктуация и форматирование.

Благодарю Arnaud Launay за его помощь в исправлении моей копии :-)

Выкинут xv проект из-за недостатка интереса публики.

Добавлены задания для будущих версий документа.

1.1.5 v1.7

Добавлена утилита buzzit. Исправлена заплатка против gnu tar.

1.1.6 v1.6

Добавлено взаимодействие в Netscape от TenThumbs.

Также изменен lesspipe.sh в соответствии с его предложением. Теперь должно работать лучше.

1.1.7 v1.5

Добавлен французский перевод Arnaud Launay, и его файл для wu-ftpd.

1.1.8 v1.4

Добавлен японский перевод Tetsu Isaji.

1.1.9 v1.3

Добавлен .emacs от Ulrik Dickow для версий 19.30 и выше.

(Также исправлен патч jka-compr.el для emacs в соответствии с его предложением. Oops! Bzip2's пока(?) не имеет флажок "append".)

1.1.10 v1.2

Изменена заплатка для emacs так, что он автоматически распознает .bz2 файлы.

1.1.11 v1.1

Добавлена заплатка для emacs.

1.1.12 v1.0

То же что и 1.

2 Получение bzip2

Домашняя страница Vzip2 - домашний сайт в Великобритании <http://www.muraroa.demon.co.uk/>. В Соединенных Штатах есть зеркало <http://www.digistar.com/bzip2/index.html>. Вы можете также обратиться к Ftp RedHat <ftp://ftp.redhat.com/pub/contrib>.

2.1 Vzip2-HOWTO на вашем языке

Франкоговорящие могут обратиться к документам на французском языке от Arnaud Launay. Веб версия на <http://www.freenix.fr/linux/HOWTO/mini/Bzip2.html>, и Вы можете использовать <ftp://ftp.lip6.fr/pub/linux/french/docs/HOWTO/mini/Bzip2.gz> Arnaud можно найти по электронной почте. Адрес <mailto:zoro@mygale.org>

Японоговорящие могут желать обратиться к японским переводам от Tetsu Isaji на <http://jf.gee.kyoto-u.ac.jp/JF/JF.html>. Isaji можно найти на его домашней страничке <http://www2s.biglobe.ne.jp/kaien/>, или по электронной почте по адресу <mailto:isaji@mxu.meshnet.or.jp>

Шведы могут обратиться к шведским переводам от Linus Gkerlund на http://user.tninet.se/~uxm165t/linux_doc.html. Linus доступен по электронной почте по адресу <mailto:uxm165t@tninet.se>

2.2 Получение скомпилированного bzip2

См. домашние странички.

Бинарники Intel Debian

ftp://ftp.debian.org/debian/dists/stable/main/binary-i386/utils/bzip2_0.1pl2-5.deb>.

Альфа-бинарник RedHat лежит на

<ftp://ftp.redhat.com/pub/redhat/redhat-5.1/alpha/RedHat/RPMS/bzip2-0.1pl2-1.alpha.rpm>>.

Бинарник Intel RedHat на

<ftp://ftp.redhat.com/pub/redhat/redhat-5.1/i386/RedHat/RPMS/bzip2-0.1pl2-1.i386.rpm>>.

Бинарник SPARC RedHat на

<ftp://ftp.redhat.com/pub/redhat/redhat-5.1/sparc/RedHat/RPMS/bzip2-0.1pl2-1.sparc.rpm>>.

Бинарник Intel Slackware на

<ftp://www.cdrom.com/pub/linux/slackware-3.5/slakware/a1/bzip2.tgz>>.

Бинарник S.u.S.E. Intel на

<ftp://ftp.suse.com/pub/SuSE-Linux/5.2/suse/ap1/bzip.rpm>>.

Вы можете также получить их в аналогичных местах на различных зеркалах.

2.3 Получение исходных текстов bzip2

Они приходят с официальных сайтов (см. "Получение Vzip2", или RedHat раздает их по URL <ftp://ftp.redhat.com/pub/contrib/SRPMS/bzip2-0.1pl2-1.src.rpm>>).

2.4 Компилирование bzip2 на вашей машине

Если у вас gcc 2.7. *, смените сторку

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

на

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

То есть замените -O3 на -O2, и уберите) -funroll-loops. Вы можете по желанию добавить любые из -m * флагов (-m486, например), которые вы используете, когда компилируете ядро.

Удаление -funroll-loops наиболее важно, так как он заставляет некоторые версии gcc 2.7 генерировать неправильный код, и все версии gcc 2.7 вырабатывают медленный и большой код. Для других трансляторов (lcc, egcs, gcc 2.8.x) значение по умолчанию CFLAGS вполне нормально. После этого, только соберите его и установите его по README.

3 Использование bzip2

Читайте Замечательную Map Страницу :)

4 Использование bzip2 со tar

Ниже перечислены три способа использовать bzip2 с tar, а именно

4.1 Самый простой в настройке:

Этот метод не требует никакой установки вообще. Чтобы растарить в текущий каталог сжатый bzip2'ом архив tar foo.tar.bz2, введите

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

Этот способ нормально работает, но может надоесть при частой работе с архивами.

4.2 Простая установка, довольно легкая в использовании, не нужно иметь привелегии root:

Благодарю Leonard Jean-marc (mailto:leonard@sct1.is.belgacom.be) за совет. Благодарю также Alessandro Rubini (mailto:rubini@morgana.systemy.it) за различия bash от csh.

В вашем .bashrc, Вы можете поместить строку типа:

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

В вашем .tcshrc, или .cshrc, аналогичная строка:

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

4.3 Также легко в использовании, но требуется уровень доступа root.

Примените патч, приведенный ниже к gnu tar 1.12 следующим образом:

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

Скомпилируйте его, установите, и вы готовы, чтобы следовать дальше. Удостоверьтесь что и tar, и bzip2 находятся в вашем \$PATH, выполнив "which tar"и "which bzip2". Для использования нового tar, выполните

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

чтобы декомпрессировать файл.

Чтобы сделать новый архив:

```
tar cyf foo.tar.bz2 file1 file2 file3 ... directory1 directory2 ...
```

Вот патч:

```
*** 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, _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:hiklmoprstuvwxyz"
    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 Использование bzip2 с less

Чтобы разжать bzip2'ованные файлы на лету, то есть занять возможность использовать для них "less" без предварительной распаковки, вы можете переделать lesspipe.sh (man less) примерно так:

```
#!/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 Использование bzip2 с emacs

6.1 Изменения в emacs для всех:

Я написал следующую заплатку jka-compr.el, которая добавляет bzip2 в auto-compression-mode. Оговорка: я проверил это только с emacs-20.2, но не вижу причин полагать, что этот метод не будет работать с другой версией.

Чтобы использовать ее

1. Идем в каталог исходников emacs-20.2/lisp (туда, где Вы распаковали его)
2. Поместите приведенный ниже патч в файл, называемый jka-compr.el.diff (кроме патча в этом файле ничего не должно быть);).
3. Введите

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

4. Запустите emacs, и введите

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

5. Выйдите из emacs.
6. Переместит ваш оригинальный jka-compr.elc в безопасное место на случай ошибок.
7. Замените его на новый jka-compr.elc.
8. Наслаждайтесь!

```
--- 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 <htmlurl url="http://www.muraroa.demon.co.uk." nam
+;;; 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 Изменения в emacs для одного человека:

Благодарю за это Ulrik Dickow (ukd@kampsax.dk), системного программиста в Kampsax Technology: Чтобы сделать так, чтобы Вы могли использовать bzip2 автоматически, когда Вы не Sysadmin, всего лишь добавьте следующее к вашему .emacs файлу.

```
;; 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 Использование bzip2 с wu-ftpd

Благодарю Arnaud Launay за этот bandwidth saver. В /etc/ftpconversions надо поместить следующее, чтобы происходило непрерывное сжатие и декомпрессия с bzip2. Удостоверитесь, что пути (типа /bin/compress) корректны.

```
:.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 Использование bzip2 с грер

Следующая утилита, которую я называю bgrep, является небольшой модификацией Zgrep, которая поставляется с Linux. Вы можете использовать ее как грер для файлов без предварительного bunzip2'ния.

```
#!/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
```

```

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 Использование bzip2 с Netscape под X.

tenthumbs@cybernex.net говорит:

Я тоже нашел способ заставить Linux Netscape использовать bzip2 для Content-Encoding точно так же, как он использует gzip. Добавьте эти строки к \$HOME/.Xdefaults или \$HOME/.Xresources Я использую опцию -s, потому что я уступаю некоторое быстроедействие декомпрессии для уменьшения использования оперативной памяти. Вы можете убрать эту опцию, если не хотите ее использовать.

```

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

```

```
gzip      : : .z, .gz : gzip -cdq \n\
gzip:: .z, .gz: gzip -cdq \n\
x-bzip2   : : .bz2   : bzip2 -ds \n
x-bzip2:: .bz2: bzip2 -ds \n
```

10 Использование bzip2 для повторного сжатия других форматов компрессии.

Следующая программа на perl берет файлы, сжатые в других форматах (.tar.gz, .tgz, .tar.Z, и .Z для этой итерации), и перепаковывает их для получения лучшего коэффициента сжатия. Исходный текст на perl сам рассказывает о том, что он делает и как заставить его работать.

```
#!/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;
$total_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";
```