

Archivierung und Komprimierung

LPI Essentials

Andreas B. Mundt
a.mundt@lehrerfortbildung-bw.de

31. Mai 2022



3.1 Dateien mithilfe der Befehlszeile archivieren

Gewichtung: 2

Beschreibung: Dateien im Heimatverzeichnis archivieren

Hauptwissensgebiete:

- Dateien, Verzeichnisse
- Archive, Komprimierung

Auszugsweise Liste wichtiger Dateien, Begriffe und Hilfsprogramme:

- tar und seine verbreiteten Optionen
- gzip, bzip2, xz
- zip, unzip

Aus- und Überblick

① Archivierung

Archivierung mit tar
Aufbau eines tar-Archivs

② Komprimierung

Komprimierungsarten
Anwendungsbeispiele
gzip, bzip2, xz

③ Archive komprimieren

tar-Archive komprimieren
zip, unzip

Archivierung

Archivierung: Viele Dateien werden in einer Datei zusammengefasst.

Vorteile bei der Verwaltung (Download, Transfer, ...) eines System/Dokument-Zustands, der sich auf viele Dateien erstreckt.

Beispiele:

- Backup
- Source Code
- Programm-Pakete
- Komplexe Dokumente

Manche Streaming-Geräte (z.B. Bandlaufwerke) kommen mit einer großen gepackten Datei besser zurecht, als mit vielen kleinen Dateien.

Archivierung mit dem **tape archiver tar**

Beispiele eines tar-Aufrufs¹:

```
tar -c [-f ARCHIVE] [OPTIONS] [FILE...]
```

```
tar -x [-f ARCHIVE] [OPTIONS] [MEMBER...]
```

```
tar -t [-f ARCHIVE] [OPTIONS] [MEMBER...]
```

-c **create**: erzeugt ein Archiv

-f **file**: Archiv-Datei

-v **verbose**: mehr Ausgabe

-x **extract**: packt ein Archiv aus

-t listet den Inhalt auf

¹Aufgrund der langen Geschichte von tar gibt es unterschiedliche Schreibweisen der Optionen.

Wie ist ein tar-Archiv aufgebaut?

Dateien und anschließend Archiv erzeugen:

```
cd /tmp && mkdir Verzeichnis
```

```
echo "Wir senden ein paar Daten in eine Datei!" \  
                                         > Verzeichnis/DateiNr1.txt
```

```
echo "Noch mehr Bytes angehängt an die zweite Datei." >> DateiNr2.txt
```

```
tar vcf archiv.tar Verzeichnis/DateiNr1.txt DateiNr2.txt
```

Wie ist ein tar-Archiv aufgebaut?

Mal reingucken mit `hexdump -C archiv.tar2:`

```
00000000 56 65 72 7a 65 69 63 68 6e 69 73 2f 44 61 74 65 |Verzeichnis/Date|
00000010 69 4e 72 31 2e 74 78 74 00 00 00 00 00 00 00 00 |iNr1.txt.....|
00000020 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
*
00000060 00 00 00 00 30 30 30 30 36 34 34 00 30 30 30 31 |....0000644.0001|
00000070 37 35 30 00 30 30 30 31 37 35 30 00 30 30 30 30 |750.0001750.0000|
00000080 30 30 30 30 30 35 31 00 31 34 32 34 33 31 33 36 |0000051.14243136|
00000090 34 33 34 00 30 31 34 31 32 37 00 20 30 00 00 00 |434.014127. 0...|
000000a0 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
*
00000100 00 75 73 74 61 72 20 20 00 61 6e 64 69 00 00 00 |.ustar .andi...|
00000110 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
00000120 00 00 00 00 00 00 00 00 00 61 6e 64 69 00 00 00 |.....andi...|
00000130 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
*
00000200 57 69 72 20 73 65 6e 64 65 6e 20 65 69 6e 20 70 |Wir senden ein p|
00000210 61 61 72 20 44 61 74 65 6e 20 69 6e 20 65 69 6e |aar Daten in ein|
00000220 65 20 44 61 74 65 69 21 0a 00 00 00 00 00 00 00 |e Datei!.....|
00000230 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
*
00000400 44 61 74 65 69 4e 72 32 2e 74 78 74 00 00 00 00 |DateiNr2.txt....|
00000410 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
*
00000460 00 00 00 00 30 30 30 30 36 34 34 00 30 30 30 31 |....0000644.0001|
....
```

²[https://en.wikipedia.org/wiki/Tar_\(computing\)](https://en.wikipedia.org/wiki/Tar_(computing))

Aus- und Überblick

1 Archivierung

Archivierung mit tar
Aufbau eines tar-Archivs

2 Komprimierung

Komprimierungsarten
Anwendungsbeispiele
gzip, bzip2, xz

3 Archive komprimieren

tar-Archive komprimieren
zip, unzip

Komprimierung

Betrachtet man tar-Archive, so enthalten sie Blöcke einer festen Größe, die gegebenenfalls mit Nullen aufgefüllt werden. → Komprimierung³ bietet sich an.

Man unterscheidet:

- ① verlustbehaftete Komprimierung (lossy) → irrelevante Infos verwerfen
 - Audio: nicht hörbare Frequenzen werden entfernt
 - Audio: sich wiederholende, fast identische Samples, nur einmal speichern
 - Bilder/Video: Farbcodierung wird auf weniger Bits verteilt
 - Bilder/Video: ähnliche Bereiche im Bild werden zusammengefasst
- ② verlustfreie Komprimierung (lossless) → Information effizienter codieren
 - Blöcke gleicher Daten nur einmal speichern und dann referenzieren

Beispiel:

Ausgangstext: AUCH EIN KLEINER BEITRAG IST EIN BEITRAG

Komprimiert: AUCH EIN KLEINER BEITRAG IST -4 -2

³<https://de.wikipedia.org/wiki/Datenkompression>

Komprimierung

Betrachtet man tar-Archive, so enthalten sie Blöcke einer festen Größe, die gegebenenfalls mit Nullen aufgefüllt werden. → Komprimierung³ bietet sich an.

Man unterscheidet:

- ① verlustbehaftete Komprimierung (lossy) → irrelevante Infos verwerfen
 - Audio: nicht hörbare Frequenzen werden entfernt
 - Audio: sich wiederholende, fast identische Samples, nur einmal speichern
 - Bilder/Video: Farbcodierung wird auf weniger Bits verteilt
 - Bilder/Video: ähnliche Bereiche im Bild werden zusammengefasst
- ② verlustfreie Komprimierung (lossless) → Information effizienter codieren
 - Blöcke gleicher Daten nur einmal speichern und dann referenzieren

Beispiel:

Ausgangstext: AUCH EIN KLEINER BEITRAG IST EIN BEITRAG

Komprimiert: AUCH EIN KLEINER BEITRAG IST -4 -2

³<https://de.wikipedia.org/wiki/Datenkompression>

Bedeutung der Komprimierung und Beispiele

Es ist oft schneller, eine komprimierte Datei herunterzuladen oder auch von der Platte zu lesen und zu dekomprimieren, als die ursprüngliche Datei zu verwenden:

- `/boot/vmlinuz`, `/boot/initrd.img`
- HTTP⁴:
 - Client request: `Accept-Encoding: gzip, deflate`
 - Server: `Content-Encoding: gzip`
- Log-Dateien: Viele identische Meldungen, Platte zugemüllt
- tar-Archive!

⁴https://en.wikipedia.org/wiki/HTTP_compression

Programme⁵: gzip, bzip2, xz, zip

gzip → GNU zip, meist unter GNU/Linux verwendet, übliche Dateiendung: .gz

bzip2 → Alternative unter Linux mit anderem Kompressionsalgorithmus – langsamer aber bessere Kompression; übliche Dateiendung: .bz oder .bz2

xz → weitere Alternative unter Linux mit anderem Kompressionsalgorithmus der bei der Kompression zeit- und speicheraufwendiger, jedoch bei der Dekompression wesentlich sparsamer als bzip2 ist; übliche Dateiendung: .xz

zip → Komprimierung und Archivierung unter Windows. Wird unter GNU/Linux dann verwendet, wenn die Dateien auch auf Windows genutzt werden sollen (optimierter Dateiaustausch); übliche Dateiendung: .zip

⁵ https://de.wikipedia.org/wiki/Liste_von_Datenkompressionsprogrammen und jeweilige Einzelseiten

Beispiel und Optionen für gzip

Datenbank Backup:

```
mysqldump --all-databases | gzip > db_backup.gz
```

Auswahl von Optionen (vergl. `gzip --help`):

- d **--decompress** decompress → alternativ: **gunzip bunzip2 unxz**
- f **--force** force overwrite of output file and compress links
- l **--list** list compressed file contents (info)
- r **--recursive** operate recursively on directories
- t **--test** test compressed file integrity
- v **--verbose**
- 1 **--fast**
- 9 **--best**

Beispiel Komprimierung

```
ls -lh syslog
for BIN in gzip bzip2 xz ; do
    time $BIN syslog
    ls -lh syslog* ; $BIN -d syslog*
done

-rw-r----- 1 andi andi 528K May 24 13:21 syslog

real 0m0.013s
-rw-r----- 1 andi andi 53K May 24 13:21 syslog.gz

real 0m0.057s
-rw-r----- 1 andi andi 33K May 24 13:21 syslog.bz2

real 0m0.078s
-rw-r----- 1 andi andi 31K May 24 13:21 syslog.xz
```

Beispiel Dekomprimierung

```
ls -lh syslog
for BIN in gzip bzip2 xz ; do
    $BIN syslog; ls -lh syslog*
    time $BIN -d syslog*
done

-rw-r----- 1 andi andi 528K May 24 13:21 syslog
-rw-r----- 1 andi andi 53K May 24 13:21 syslog.gz
real 0m0.004s

-rw-r----- 1 andi andi 33K May 24 13:21 syslog.bz2
real 0m0.009s

-rw-r----- 1 andi andi 31K May 24 13:21 syslog.xz
real 0m0.004s
```

Anregungen zur Komprimierung

Untersuchen Sie die Komprimierung für folgende Dateien:

Eine Datei mit Zufallszahlen:

```
dd if=/dev/random of=10M.bin bs=1M count=10
```

Eine Datei mit Nullen:

```
dd if=/dev/zero of=10M.bin bs=1M count=10
```

→ Log-Dateien in `/var/log/` eignen sich besser für realitätsnahe Tests.

Aus- und Überblick

1 Archivierung

Archivierung mit tar
Aufbau eines tar-Archivs

2 Komprimierung

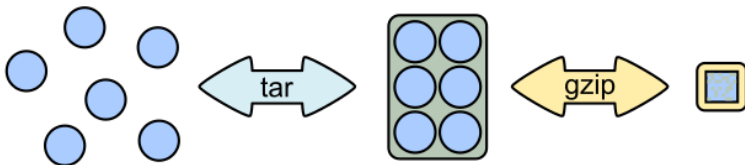
Komprimierungsarten
Anwendungsbeispiele
gzip, bzip2, xz

3 Archive komprimieren

tar-Archive komprimieren
zip, unzip

tar-Archive komprimieren

Es liegt nahe, tar-Archive gleich bei der Erstellung auch zu komprimieren:



6

-c create: erzeugt ein Archiv

-f file: Archiv-Datei

-v verbose: mehr Ausgabe

-x extract: packt ein Archiv aus

-t listet den Inhalt auf

-z wendet gzip auf das Archiv an

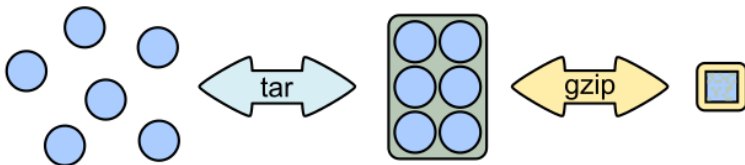
-j wendet bzip2 auf das Archiv an

-J wendet xz auf das Archiv an

⁶Dateien (Kreise) werden mit tar gepackt, anschließend das Archiv mit gzip komprimiert.

tar-Archive komprimieren

Es liegt nahe, tar-Archive gleich bei der Erstellung auch zu komprimieren:



6

-c create: erzeugt ein Archiv

-f file: Archiv-Datei

-v verbose: mehr Ausgabe

-x extract: packt ein Archiv aus

-t listet den Inhalt auf

-z wendet gzip auf das Archiv an

-j wendet bzip2 auf das Archiv an

-J wendet xz auf das Archiv an

⁶Dateien (Kreise) werden mit tar gepackt, anschließend das Archiv mit gzip komprimiert.
<https://commons.wikimedia.org/wiki/File:Targzip.svg>

Beispiel komprimiertes tar-Archiv

Komprimierte Archive erzeugen:

```
tar vcf archiv.tar      Verzeichnis/DateiNr1.txt DateiNr2.txt
tar zvcf archiv.tar.gz  Verzeichnis/DateiNr1.txt DateiNr2.txt
tar jvcf archiv.tar.bz2 Verzeichnis/DateiNr1.txt DateiNr2.txt
tar Jvcf archiv.tar.xz  Verzeichnis/DateiNr1.txt DateiNr2.txt
```

Übliche Dateinamen (o.ä.):

- .tar.gz .tgz
- .tar.bz2 .tbz2 .tbz
- .tar.xz .txz

Auf allen Plattformen verfügbar: zip, unzip

zip is a compression and file packaging utility for Unix, VMS, MSDOS, OS/2, Windows 9x/NT/XP, Minix, Atari, Macintosh, Amiga, and Acorn RISC OS. It is analogous to a combination of the Unix commands tar(1) and compress(1) and is compatible with PKZIP (Phil Katz's ZIP for MSDOS systems).

```
zip archiv DateiNr2.txt Verzeichnis/DateiNr1.txt
```

```
unzip -l archiv.zip
```

```
Archive: archiv.zip
```

Length	Date	Time	Name
-----	-----	-----	----
116	2022-05-24	13:10	DateiNr2.txt
41	2022-05-24	14:20	Verzeichnis/DateiNr1.txt
-----			-----
157			2 files

Beispiel zip-Archiv

OpenDocument-Dateien sind zip-Archive:

```
unzip -l CheatSheet_LinEss_Teil1.odt
```

```
Archive: CheatSheet_LinEss_Teil1.odt
  Length      Date      Time    Name
-----
      39      2020-09-20  06:08    mimetype
       0      2020-09-20  06:08    Configurations2/toolbar/
       0      2020-09-20  06:08    Configurations2/floater/
       0      2020-09-20  06:08    Configurations2/menuubar/
       0      2020-09-20  06:08    Configurations2/popupmenu/
       0      2020-09-20  06:08    Configurations2/accelerator/
       0      2020-09-20  06:08    Configurations2/toolpanel/
       0      2020-09-20  06:08    Configurations2/progressbar/
       0      2020-09-20  06:08    Configurations2/statusbar/
       0      2020-09-20  06:08    Configurations2/images/Bitmaps/
      899      2020-09-20  06:08    manifest.rdf
      984      2020-09-20  06:08    meta.xml
     11876      2020-09-20  06:08    settings.xml
      8131      2020-09-20  06:08    Thumbnails/thumbnail.png
     19714      2020-09-20  06:08    styles.xml
       745      2020-09-20  06:08    Pictures/1000020100000005000000000F5B106A9D7A6352A4.png
     36957      2020-09-20  06:08    content.xml
       1192      2020-09-20  06:08    META-INF/manifest.xml
-----
     80537
                   18 files
```

OpenDocument-Dateien sind zip-Archive

unzip CheatSheet_LinEss_Teil1.odt

```
Archive: CheatSheet_LinEss_Teil1.odt
  extracting: mimetype
    creating: Configurations2/toolbar/
    creating: Configurations2/floater/
    creating: Configurations2/menubar/
    creating: Configurations2/popupmenu/
    creating: Configurations2/accelerator/
    creating: Configurations2/toolpanel/
    creating: Configurations2/progressbar/
    creating: Configurations2/statusbar/
    creating: Configurations2/images/Bitmaps/
  inflating: manifest.rdf
  inflating: meta.xml
  inflating: settings.xml
  extracting: Thumbnails/thumbnail.png
  inflating: styles.xml
  extracting: Pictures/100002010000005000000000F5B106A9D7A6352A4.png
  inflating: content.xml
  inflating: META-INF/manifest.xml
```

Schauen Sie sich mal den Inhalt der Verzeichnisse und Dateien an, z.B.:

```
display ./Pictures/100002010000005000000000F5B106A9D7A6352A4.png
```

① Archivierung

Archivierung mit tar
Aufbau eines tar-Archivs

② Komprimierung

Komprimierungsarten
Anwendungsbeispiele
gzip, bzip2, xz

③ Archive komprimieren

tar-Archive komprimieren
zip, unzip

Quellen der Inspiration

- Michael Krüger <mich@elkrueger.de>
LINUX ESSENTIALS: Archivierung und Komprimierung, Kapitel 9
(19. September 2019)