

LINUX DEBIAN

Les archives et la compression

Cours TSSR — Commandes, options et exemples pratiques

TAR

GZIP / BZIP2

ZIP / RAR / CPIO

Sommaire

1. Les archives
2. La compression
3. Différence entre archivage et compression
4. TAR et les tarballs
5. CPIO
6. GZIP et BZIP2
7. Lire un fichier compressé sans l'extraire définitivement
8. Les fichiers ZIP
9. Les fichiers RAR
10. Tableau récapitulatif
11. À retenir absolument
12. Sources techniques

1. Les archives

Une archive est un fichier qui rassemble un ou plusieurs fichiers et répertoires dans un seul ensemble.

Une archive peut contenir différentes métadonnées :

- Le nom des fichiers
- Leur chemin
- Leur taille
- Leurs permissions
- Leur propriétaire et leur groupe
- Leurs dates de modification
- La structure des répertoires

Une archive n'est pas obligatoirement compressée. Son premier objectif est de regrouper plusieurs éléments dans un seul fichier.

Les archives peuvent être utilisées pour :

- Regrouper plusieurs fichiers
- Sauvegarder des répertoires
- Sauvegarder des disques ou des bandes magnétiques
- Transporter plus facilement un ensemble de fichiers
- Conserver les permissions et l'organisation des fichiers
- Réaliser des sauvegardes

Les archives multivolumes

Une archive multivolume est découpée en plusieurs fichiers ou plusieurs volumes.

Cela permet notamment de répartir une archive trop volumineuse sur plusieurs supports ou plusieurs fichiers.

Attention

- GNU tar permet les archives multivolumes.
- Une archive multivolume créée avec GNU tar ne peut pas être compressée directement.

2. La compression

La compression sert à réduire la taille occupée par les données.

Il existe deux grandes catégories de compression.

Compression non destructive — compression sans perte

Toutes les informations d'origine sont conservées.

Après la décompression, le fichier obtenu est identique au fichier d'origine.

Exemples :

- ZIP
- GZIP
- BZIP2
- XZ
- 7z lorsqu'un algorithme sans perte est utilisé

Cette compression est adaptée aux fichiers système, aux documents, aux programmes, aux bases de données et aux sauvegardes.

Compression destructive — compression avec perte

Une partie des informations est supprimée afin de réduire davantage la taille du fichier.

Après la décompression ou le décodage, il n'est pas possible de reconstituer exactement le fichier d'origine.

Exemples courants :

- JPEG pour les images
- MP3 pour l'audio
- De nombreux codecs audio et vidéo utilisés dans des fichiers MP4 ou AVI

Précision importante

- MP4 et AVI sont avant tout des conteneurs.
- Le type réel de compression dépend du codec contenu dans le fichier.

3. Différence entre archivage et compression

Archivage

L'archivage rassemble plusieurs fichiers dans un seul fichier.

```
tar -cf archive.tar dossier/
```

→ Le fichier archive.tar contient le dossier, mais il n'est pas nécessairement compressé.

Compression

La compression réduit la taille d'un fichier.

```
gzip archive.tar
```

→ Le fichier obtenu est archive.tar.gz.

Le fichier .tar a d'abord regroupé les éléments, puis gzip a compressé l'ensemble.

Formats courants

Format	Fonction
.tar	Archive non compressée
.cpio	Archive
.gz	Compression GZIP d'un fichier ou d'un flux
.bz2	Compression BZIP2 d'un fichier ou d'un flux
.xz	Compression XZ d'un fichier ou d'un flux
.tar.gz ou .tgz	Archive TAR compressée avec GZIP
.tar.bz2 ou .tbz2	Archive TAR compressée avec BZIP2
.tar.xz ou .txz	Archive TAR compressée avec XZ
.zip	Archivage et compression
.rar	Archivage et compression
.7z	Archivage et compression

Avec tar.gz, tar.bz2 ou tar.xz, la compression est généralement appliquée à l'ensemble du flux de l'archive TAR. Avec ZIP, chaque fichier peut être compressé séparément dans l'archive.

4. TAR et les tarballs

Présentation de TAR

tar signifie historiquement « Tape Archiver ».

Cet outil a été conçu à l'origine pour effectuer des sauvegardes sur des bandes magnétiques.

Le format TAR n'applique pas de compression par lui-même. Il rassemble simplement les fichiers et leurs métadonnées dans une archive.

Une archive TAR est souvent appelée une « tarball ».

tar est l'un des outils d'archivage les plus utilisés sous Linux et Unix.

Il peut :

- Créer une archive
- Extraire une archive
- Lister son contenu
- Archiver récursivement des répertoires
- Conserver des métadonnées
- Être associé à gzip, bzip2, xz et d'autres compresseurs

GNU tar reconnaît automatiquement le format de compression lors de la lecture de nombreuses archives. Lors de la création, il faut normalement préciser le compresseur avec -z, -j, -J, ou utiliser -a pour choisir automatiquement le compresseur en fonction de l'extension.

Syntaxe générale

```
tar [options] archive [fichiers ou répertoires]
```

Une opération principale doit être indiquée :

- -c pour créer
- -x pour extraire
- -t pour lister

Options principales

Option	Explication
-c	Active le mode création
-x	Active le mode extraction
-t	Liste les fichiers présents dans l'archive
-f	Spécifie le fichier d'archive à utiliser
-v	Active le mode verbeux et affiche les fichiers traités
-z	Utilise la compression GZIP
-j	Utilise la compression BZIP2
-J	Utilise la compression XZ
-a	Choisit automatiquement le compresseur selon l'extension lors de la création
-C	Change de répertoire avant l'extraction ou l'opération
-M	Active le mode multivolume

L'option -f est suivie du nom de l'archive :

```
tar -cf archive.tar dossier/
```

Dans cette commande :

- c = créer
- f = utiliser le fichier archive.tar

Lister le contenu d'une archive TAR.GZ

Version détaillée :

```
tar -tvzf archive.tar.gz
```

- t : liste le contenu
- v : affiche les informations détaillées
- z : utilise GZIP
- f : indique le nom de l'archive

Version courte avec détection automatique :

```
tar -tf archive.tar.gz
```

GNU tar peut reconnaître automatiquement la compression lors de la lecture de l'archive.

Créer une archive TAR non compressée

```
tar -cvf archive.tar dossier/
```

Version sans affichage détaillé :

```
tar -cf archive.tar dossier/
```

Extraire une archive TAR dans le répertoire courant

```
tar -xvf archive.tar
```

Version courte :

```
tar -xf archive.tar
```

Les fichiers sont extraits dans le répertoire courant.

Extraire dans un répertoire précis

Le répertoire de destination doit déjà exister.

```
mkdir extraction
```

```
tar -xf archive.tar -C extraction/
```

Pour une archive compressée :

```
tar -xf archive.tar.gz -C extraction/
```

Créer une archive TAR.GZ pour un répertoire

```
tar -cvzf archive.tar.gz directory/
```

Version sans mode verbeux :

```
tar -czf archive.tar.gz directory/
```

Créer une archive TAR.GZ contenant plusieurs éléments

```
tar -cvzf archive.tar.gz d1/ d2/ f1 f2
```

Les répertoires d1, d2 et les fichiers f1, f2 sont regroupés et compressés dans archive.tar.gz.

Extraire une archive TAR.GZ

Avec l'option GZIP explicite :

```
tar -xvzf archive.tar.gz
```

Version courte avec détection automatique :

```
tar -xf archive.tar.gz
```

Créer une archive TAR.BZ2

```
tar -cvjf archive.tar.bz2 directory/
```

- c : création
- v : mode verbeux
- j : compression BZIP2
- f : nom de l'archive

Extraire une archive TAR.BZ2

Version détaillée :

```
tar -xvjf archive.tar.bz2
```

Version courte avec détection automatique :

```
tar -xf archive.tar.bz2
```

Créer une archive TAR.XZ

```
tar -cvJf archive.tar.xz directory/
```

Extraire une archive TAR.XZ

```
tar -xvJf archive.tar.xz
```

Version courte :

```
tar -xf archive.tar.xz
```

Création avec détection de l'extension

GNU tar peut sélectionner le compresseur selon l'extension avec -a :

```
tar -caf archive.tar.gz directory/
```

```
tar -caf archive.tar.bz2 directory/
```

```
tar -caf archive.tar.xz directory/
```

L'option -a est principalement utile lors de la création. Pour la lecture et l'extraction, GNU tar détecte généralement le format automatiquement.

5. CPIO

Présentation de CPIO

cpio signifie « Copy In / Copy Out ».

Il s'agit d'un très ancien outil d'archivage Unix. Il a été en grande partie remplacé par tar pour les usages courants.

Il reste néanmoins utilisé pour certaines tâches spécifiques, notamment pour certains fichiers initramfs.

Dans son fonctionnement Unix classique, cpio ne reçoit pas directement une liste de fichiers sous la forme habituelle.

Il utilise les flux standards :

- Il lit la liste des fichiers depuis l'entrée standard stdin
- Il écrit l'archive vers la sortie standard stdout
- L'entrée ou la sortie est généralement reliée à un fichier avec une redirection
- Une commande comme find peut lui fournir la liste des fichiers avec un pipe

GNU cpio possède trois modes principaux :

Option	Mode
-o	Copy-out : crée une archive
-i	Copy-in : lit ou extrait une archive
-p	Pass-through : copie directement des fichiers vers un autre répertoire

GNU cpio peut également utiliser des options permettant de lire ou écrire directement un fichier d'archive, mais son utilisation traditionnelle repose sur les flux standards.

Créer une archive CPIO

Version simple :

```
find . -print | cpio -ov > archive.cpio
```

- `find . -print` fournit la liste des fichiers
- Le pipe `|` transmet cette liste à `cpio`
- `-o` crée l'archive
- `-v` affiche les fichiers traités
- `>` redirige l'archive vers `archive.cpio`

Version plus sûre pour les noms contenant des espaces ou des caractères particuliers :

```
find . -print0 | cpio --null -ov --format=newc > archive.cpio
```

Lister le contenu

```
cpio -itv < archive.cpio
```

- `-i` : lecture de l'archive
- `-t` : liste le contenu
- `-v` : mode détaillé
- `<` : envoie le fichier vers l'entrée standard

Extraire une archive

```
cpio -idmv < archive.cpio
```

- `-i` : extraction
- `-d` : crée les répertoires nécessaires
- `-m` : conserve les dates de modification
- `-v` : affiche les fichiers traités

6. GZIP et BZIP2

`gzip` et `bzip2` sont principalement des outils de compression de fichiers ou de flux.

Ils ne regroupent pas directement plusieurs fichiers dans une archive unique. Pour compresser plusieurs fichiers ensemble, ils sont généralement associés à `tar`.

Extensions :

- GZIP : `.gz`
- BZIP2 : `.bz2`

Par défaut, `gzip` et `bzip2` remplacent généralement le fichier d'origine par sa version compressée. L'option `-k` permet de conserver le fichier source.

BZIP2 obtient généralement une compression plus importante que les compresseurs classiques basés sur LZ77, mais utilise davantage de mémoire et de temps de calcul. GZIP est généralement choisi lorsque la rapidité et une faible consommation de ressources sont prioritaires.

GZIP — Compression

```
gzip -9 file
```

→ Résultat : `file.gz`

`-9` demande le niveau de compression maximal de GZIP.

Le fichier original est remplacé par défaut.

Pour conserver le fichier original :

```
gzip -k -9 file
```

GZIP — Décompression

```
gzip -d file.gz
```

Commande équivalente :

```
gunzip file.gz
```

Pour conserver également le fichier compressé :

```
gzip -dk file.gz
```

BZIP2 — Compression

```
bzip2 -9 file
```

→ Résultat : file.bz2

Pour conserver le fichier original :

```
bzip2 -k -9 file
```

Avec BZIP2, le niveau -9 sélectionne une taille de bloc de 900 Ko. Il s'agit également du comportement par défaut des versions actuelles de BZIP2.

BZIP2 — Décompression

```
bzip2 -d file.bz2
```

Commande équivalente :

```
bunzip2 file.bz2
```

Pour conserver également le fichier compressé :

```
bzip2 -dk file.bz2
```

7. Lire un fichier compressé sans l'extraire définitivement

Les commandes zcat, bzcat, zless et bzless permettent de consulter le contenu d'un fichier compressé sans créer automatiquement un fichier décompressé sur le disque.

Elles évitent donc de devoir :

1. Décompresser le fichier
2. Le consulter
3. Supprimer ensuite le fichier décompressé

ZCAT — GZIP vers la sortie standard

```
zcat /proc/config.gz
```

Cette commande affiche le contenu décompressé sur la sortie standard.

Le fichier /proc/config.gz n'existe que sur les systèmes dont le noyau expose sa configuration à cet emplacement.

Commande équivalente :

```
gzip -dc fichier.gz
```

zcat peut reconnaître un fichier GZIP grâce à sa signature interne, même si son nom ne se termine pas par .gz.

BZCAT — BZIP2 vers la sortie standard

```
bzcat file.bz2
```

Commande équivalente :

```
bzip2 -dc file.bz2
```

bzcat écrit le contenu décompressé vers la sortie standard. L'extension .bz2 n'est pas nécessaire pour nommer le fichier, à condition que son contenu soit réellement au format BZIP2.

Rediriger le contenu décompressé vers un fichier

```
zcat /proc/config.gz > kconfig
```

Cette commande :

- Lit /proc/config.gz
- Décompresse son contenu
- Crée le fichier kconfig
- Ne supprime pas le fichier compressé d'origine

Pour BZIP2 :

```
bzcat file.bz2 > file
```

ZLESS — Lire un fichier GZIP page par page

```
zless /proc/config.gz
```

zless transmet le contenu décompressé à un pageur comme less.

BZLESS — Lire un fichier BZIP2 page par page

```
bzless file.bz2
```

Ces commandes sont particulièrement utiles pour consulter de gros fichiers de logs ou de configuration compressés.

8. Les fichiers ZIP

Le format ZIP regroupe les fonctions d'archivage et de compression.

Sous Linux :

- zip crée ou modifie une archive ZIP
- unzip liste, teste ou extrait une archive ZIP

Il faut distinguer les syntaxes des deux commandes.

Syntaxe de ZIP

```
zip [options] archive.zip [fichiers]
```

Syntaxe de UNZIP

```
unzip [options] archive.zip
```

Options importantes

Commande	Option	Explication
zip	-r	Ajoute récursivement un répertoire et son contenu
unzip	-d répertoire	Choisit le répertoire d'extraction
unzip	-l	Liste le contenu sans extraire
unzip	-t	Teste l'intégrité de l'archive

Attention

- -d /chemin est une option de unzip.
- Avec la commande zip, l'option -d possède une autre fonction : elle sert à supprimer des éléments d'une archive.

Créer une archive ZIP avec plusieurs fichiers

```
zip source.zip *.c *.h
```

Cette commande insère tous les fichiers se terminant par `.c` et `.h` dans `source.zip`.

Créer une archive ZIP contenant un répertoire

```
zip -r photos.zip photos/
```

L'option `-r` est nécessaire pour parcourir récursivement le répertoire.

Lister le contenu sans extraire

```
unzip -l photos.zip
```

Extraire dans le répertoire courant

```
unzip photos.zip
```

Extraire dans un répertoire précis

```
unzip photos.zip -d /tmp/
```

La syntaxe recommandée place l'option `-d` et le répertoire de destination après le nom de l'archive.

9. Les fichiers RAR

RAR est un format d'archive et de compression associé aux outils RAR et WinRAR.

Les outils officiels permettent de créer et d'extraire des archives RAR. Sous Debian, l'outil disponible dépend des paquets installés. `unrar-free` est une implémentation libre consacrée à l'extraction, tandis que l'outil officiel RAR est distribué par RARLAB.

Commandes principales

Commande	Argument	Explication
<code>unrar</code>	<code>x</code>	Extrait les fichiers en conservant leur chemin complet
<code>unrar</code>	<code>e</code>	Extrait les fichiers dans le répertoire courant sans recréer leur chemin complet
<code>unrar</code>	<code>l</code>	Liste les fichiers présents dans l'archive
<code>rar</code>	<code>a</code>	Ajoute des fichiers à une archive ou crée l'archive

Extraire dans le répertoire courant sans conserver les chemins

```
unrar e photos.rar
```

Les fichiers contenus dans les sous-répertoires sont placés directement dans le répertoire courant.

Cela peut provoquer des conflits si plusieurs fichiers portent le même nom.

Extraire en conservant l'arborescence

```
unrar x photos.rar
```

Les répertoires et sous-répertoires enregistrés dans l'archive sont recréés.

Lister le contenu

```
unrar l photos.rar
```

Créer une archive RAR

```
rar a archive.rar fichier1 fichier2 dossier/
```

Dans cette commande :

- a signifie « ajouter »
- archive.rar est l'archive à créer ou à modifier
- Les éléments suivants sont ajoutés dans l'archive

10. Tableau récapitulatif

Objectif	Commande
Créer une archive TAR	<code>tar -cvf archive.tar dossier/</code>
Lister une archive TAR	<code>tar -tf archive.tar</code>
Extraire une archive TAR	<code>tar -xvf archive.tar</code>
Créer une archive TAR.GZ	<code>tar -cvzf archive.tar.gz dossier/</code>
Extraire une archive TAR.GZ	<code>tar -xf archive.tar.gz</code>
Créer une archive TAR.BZ2	<code>tar -cvjf archive.tar.bz2 dossier/</code>
Extraire une archive TAR.BZ2	<code>tar -xf archive.tar.bz2</code>
Créer une archive TAR.XZ	<code>tar -cvJf archive.tar.xz dossier/</code>
Extraire une archive TAR.XZ	<code>tar -xf archive.tar.xz</code>
Extraire vers un répertoire	<code>tar -xf archive.tar.gz -C destination/</code>
Compresser avec GZIP	<code>gzip fichier</code>
Décompresser un GZIP	<code>gunzip fichier.gz</code>
Lire un GZIP sans l'extraire	<code>zcat fichier.gz</code>
Lire un GZIP page par page	<code>zless fichier.gz</code>
Compresser avec BZIP2	<code>bzip2 fichier</code>
Décompresser un BZIP2	<code>bunzip2 fichier.bz2</code>
Lire un BZIP2 sans l'extraire	<code>bzcat fichier.bz2</code>
Lire un BZIP2 page par page	<code>bzless fichier.bz2</code>
Créer un ZIP	<code>zip archive.zip fichiers</code>
Créer un ZIP récursif	<code>zip -r archive.zip dossier/</code>
Lister un ZIP	<code>unzip -l archive.zip</code>
Extraire un ZIP	<code>unzip archive.zip</code>
Extraire un ZIP ailleurs	<code>unzip archive.zip -d destination/</code>
Lister un RAR	<code>unrar l archive.rar</code>
Extraire un RAR avec chemins	<code>unrar x archive.rar</code>
Extraire un RAR sans chemins	<code>unrar e archive.rar</code>
Créer un RAR	<code>rar a archive.rar fichiers</code>

11. À retenir absolument

Option	Signification
-c	Créer
-x	Extraire
-t	Lister
-v	Afficher les détails
-f	Préciser le fichier d'archive
-z	GZIP
-j	BZIP2
-J	XZ
-C	Choisir un autre répertoire

Exemple simple à mémoriser :

```
tar -cvzf sauvegarde.tar.gz dossier/
```

→ Création d'une archive compressée.

```
tar -tf sauvegarde.tar.gz
```

→ Vérification de son contenu.

```
tar -xf sauvegarde.tar.gz
```

→ Extraction dans le répertoire courant.

12. Sources techniques

- GNU Tar Manual — Free Software Foundation : <https://www.gnu.org/software/tar/manual/>
- GNU Gzip Manual — Free Software Foundation : <https://www.gnu.org/software/gzip/manual/>
- Bzip2 Manual — Sourceware : <https://sourceware.org/bzip2/manual/manual.html>
- Debian Manpages — cpio : <https://manpages.debian.org/>
- Debian Manpages — zip / unzip : <https://manpages.debian.org/>
- Debian Manpages — unrar-free : <https://manpages.debian.org/>
- MDN Web Docs — compression avec perte : <https://developer.mozilla.org/>