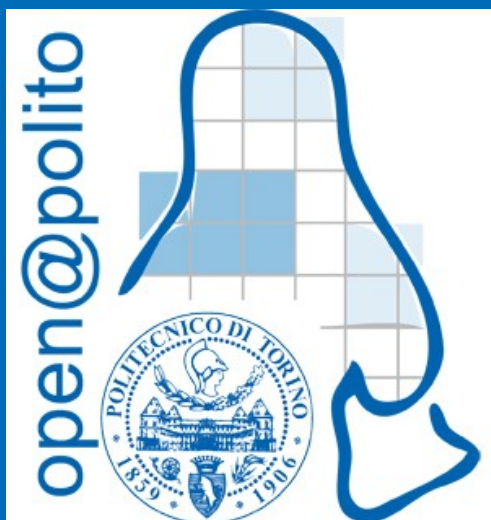




COMPRESSIONE ED ESTRAZIONE DI FILE IN ARCHIVI ZIP E TGZ

Con il supporto di:



Di Chiara Cavallo

Perché archiviare e comprimere dati?

- **Risparmio di Spazio:** La compressione riduce le dimensioni dei file, liberando spazio prezioso sui dischi e sui server.
- **Organizzazione:** Invece di 100 file sparsi, si ha un unico, ordinato pacchetto. È come imballare un trasloco in scatole invece di lanciare oggetti singoli nel furgone.
- **Distribuzione:** È molto più efficiente scaricare un unico file da 100MB che 1000 file da 100KB ciascuno. Riduce i rischi di corruzione e semplifica i trasferimenti.
- **Preservazione dei Permessi:** Su Linux, i file hanno permessi (lettura, scrittura, esecuzione) e proprietà (utente, gruppo). Un buon sistema di archiviazione deve saper preservare queste informazioni, vitali per il corretto funzionamento di script e applicazioni.

Archiviazione

L'archiviazione è l'atto di prendere molti file e/o cartelle e unirli in un unico file.

- **Obiettivo principale:** Organizzazione e convenienza. Trasformare una struttura complessa (centinaia di file in varie cartelle) in un unico oggetto facile da maneggiare, copiare, spostare o inviare.
- **Analogia:** Mettere tutti i documenti di un progetto (fogli, foto, disegni) in un'unica cartella d'archivio con un'etichetta.
- **Cosa succede allo spazio?** Non c'è risparmio. Spesso l'archivio finale è addirittura leggermente più grande della somma dei file originali, perché deve includere metadati e informazioni sulla struttura delle cartelle.
- **Esempi:** *tar* (.tar) e *zip*.

Compressione

La compressione è l'atto di ridurre le dimensioni di uno o più file utilizzando algoritmi che eliminano la ridondanza dei dati.

- **Obiettivo principale:** Risparmio di spazio su disco e banda durante i trasferimenti.
- **Analogia:** Prendere quella cartella d'archivio e usare una pressa per rimuovere l'aria e comprimere la carta, rendendo il pacco più piccolo e denso.
- **Come funziona?** Gli algoritmi di compressione cercano pattern ripetitivi nei dati e li sostituiscono con riferimenti più corti.
- **2 varianti:** “Senza perdita di dati” (lossless) e “con perdita di dati” (lossy).
- **Esempi:** *gzip* (.gz), *bzip2* (.bz2), *xz* (.xz) e *zip*.

Differenze

Caratteristica	gzip	bzip2	xz	zip
Estensione	.gz , .tgz , .tar.gz	.bz2 , .tbz2 , .tar.bz2	.xz , .txz , .tar.xz	.zip
Tipo	Compressore Puro	Compressore Puro	Compressore Puro	Archiviatore Ibrido
Compressione	Bilanciata	Migliore di gzip	La Migliore	Simile a gzip
Velocità Comp.	+++ Veloce	-- Lenta	--- Molto Lenta	++ Veloce
Velocità Decomp.	+++ Veloce	+ Media	++ Veloce	++ Veloce
Risorse CPU	Basse	Alte (comp.)	Molto Alte (comp.)	Basse
Metadati Unix	★★★ (con Tar)	★★★ (con Tar)	★★★ (con Tar)	☆☆☆ (Limitatissimo)
Caso d'Uso Principale	Uso generale, log, velocità	Download più piccoli (obsoleto)	Massima compressione, pacchetti	Condivisione cross-platform

ZIP (.zip)

Nato nel mondo DOS/Windows, è diventato uno standard universale. Tutti i sistemi operativi lo comprendono.

E' un formato che fa entrambe le cose contemporaneamente. È un "tutto-in-uno".

Internamente, ZIP comprime ogni file individualmente e poi li archivia insieme in un unico contenitore.

Vantaggi: Compatibilità universale, Flessibilità alle modifiche, Accesso casuale, Comando unico.

Svantaggi: Meno efficiente per certe situazioni (es. tanti piccoli file simili), Perde metadati critici di Unix.

Usare ZIP quando:

- Deve condividere file con utenti Windows.
- L'archivio deve essere letto o modificato facilmente su qualsiasi sistema.

ZIP (.zip)

ZIP è gestito dai programmi *zip* e *unzip*.

Per installarli su Debian: *sudo apt install zip unzip*

Per **creare** un archivio ZIP: *zip -r nome_archivio.zip cartella_da_zippare*

Per **estrarre** un archivio ZIP: *unzip nome_archivio.zip*

Per **estrarre il contenuto di una directory** specifica: *unzip nome_archivio.zip -d path_cartella*

Per **escludere dei file**: *zip -r nome_archivio.zip path_cartella/ -x "pattern"*

Esempio: *zip -r politecnico.zip politecnico -x "*.txt" "docenti/*"*

→ zippa tutto tranne i file .txt e la cartella docenti con il relativo contenuto

Per **vedere il contenuto** di un archivio **senza estrarlo**: *unzip -l nome_archivio.zip*

TAR.GZ (.tgz)

Formato nativo in ambienti Unix/Linux per archiviare e comprimere dati.

Questo è un formato a due fasi:

- **TAR (Tape ARchiver)**: Il suo compito è impacchettare. Prende molti file e directory e crea un unico file (.tar) senza compressione.
- **GZIP**: Il suo compito è comprimere. Prende un file (spesso .tar) e lo comprime, creando un file .gz.

Vantaggi: Preservazione dei Metadati, Elevata efficienza di compressione, Standard in Unix/Linux

Svantaggi: Modifiche inefficienti, Compatibilità limitata, Accesso sequenziale, Doppio strato di elaborazione

Usare TAR.GZ quando:

- Lavora esclusivamente in ambienti Unix/Linux.
- Deve preservare tutti i permessi, i proprietari e i link simbolici.
- Vuole la massima efficienza di compressione per certi tipi di file (es. codice sorgente).

TAR (.tar)

Per **creare** un archivio TAR:

tar -cvf nome_archivio.tar cartella_da_archiviare

-c → Crea. Creazione archivio

-v → Verboso. Mostra i file che vengono processati (opzionale, ma utile per seguire il processo)

-f → File. Specifica il nome dell'archivio (**fondamentale**, deve essere sempre presente come **ultima opzione**)

Per **estrarre** un archivio TAR:

tar -xvf nome_archivio.tar (-C cartella_di_destinazione)

-x → Estrai. Estrazione archivio

-C → Cambia directory prima di effettuare l'estrazione (opzionale). Se omessa l'archivio viene estratto nella directory corrente

Per **leggere** un archivio TAR **senza estrarlo**: *tar -tf nome_archivio.tar*

-t → Lista. Mostra il contenuto dell'archivio senza estrarlo

GZIP (.gz)

Per **comprimere** un archivio TAR con GZIP:

```
gzip nome_archivio.tar
```

Per **decomprimere** un archivio TAR con GZIP:

```
gunzip nome_archivio.tar.gz
```

TAR.GZ (.tgz o .tar.gz)

Ma perché fare due passaggi? TAR è così intelligente da fare tutto in uno!

Per **creare** un archivio TAR e **comprimerlo** con GZIP:

```
tar -czvf nome_archivio.tar.gz cartella_da_archiviare
```

-z → Zip. Comprime l'archivio appena creato

Per **estrarre** un archivio TAR e **decomprimerlo** con GZIP:

```
tar -xzvf nome_archivio.tar.gz (-C cartella_di_destinazione)
```

-z → Zip. Decomprimi l'archivio prima di estrarlo

TAR.GZ (.tgz o .tar.gz)

Per **estrarre il contenuto di una directory o file** specifico:

```
tar -xzvf nome_archivio.tgz (-C cartella_destinazione) path_cartella/
```

Per **escludere dei file**:

```
tar -czvf nome_archivio.tgz --exclude="pattern" (\ --exclude="pattern") cartella_da_archiviare
```

Esempio: *tar -czvf politecnico.tgz --exclude="*.txt" politecnico*

→ archivia e comprime tutto tranne i file .txt

Per **comprimere con diversi algoritmi di compressione**:

```
tar -cavf nome_archivio.tar.estensione_algoritmo cartella_da_archiviare
```

Esempio: *tar -cavf politecnico.tar.xz politecnico*

-a → Auto Compress. Indica a tar di dedurre automaticamente il formato della compressione dall'estensione inserita a seguito di nome_archivio