

# Archivierung / Komprimierung

## Einführung

- Komprimierung wird eingesetzt, um Platzbedarf von Daten zu reduzieren
- **Ziel:** Einsparung von Speicherplatz oder reduzierte Datenmenge bei Übertragung über Netzwerk
- Kompression ersetzt sich wiederholende Muster durch Daten
- **Verlustfrei (lossless)** und **verlustbehaftet (lossy)** sind zwei Varianten der Kompression
- Archivierungswerkzeuge dienen dazu, Dateien und Verzeichnisse in einer einzigen Datei zusammenzufassen
- Archive und Komprimierung gehen meist einher
- Das gebräuchlichste Werkzeug zur Archivierung von Dateien auf Linux-Systemen ist tar
- Komprimierungswerkzeuge unter Linux: bzip2, gzip und xz
- Kompromisse bei Komprimierung: hohe Kompressionsrate führt zu längerer Komprimierung und Dekomprimierung
- Selbständige Komprimierungswerkzeuge sind in der Regel auf Windows-Systemen nicht verfügbar
- Linux-Systeme haben Werkzeuge für Verwaltung von .zip-Dateien: zip und unzip, aber nicht auf allen Systemen standardmäßig installiert

## Komprimierung

- Die eingesparte Speichermenge durch Dateikomprimierung hängt von Art der Daten, Kompressionsalgorithmus und -stufe ab
- Komprimierungstools: `bzip2`, `gzip`, `xz`
- Dateien dekomprimieren mit: `bunzip2`, `gunzip`, `unxz`
- Einige Komprimierungswerkzeuge unterstützen unterschiedliche Komprimierungsstufen. Eine höhere Komprimierungsstufe erfordert in der Regel mehr Speicher und CPU-Zyklen, führt aber zu einer kleineren komprimierten Datei.
- Beispiel: `gzip -1` vs. `gzip -9`, `xz -1` vs. `xz -9`

## Archivierung

- Das Programm `tar` ist das am weitesten verbreitete Archivierungswerkzeug auf Linux-Systemen.
- Der Name "tar" steht für "Tape Archive".
- Dateien, die mit tar erstellt wurden, werden oft als "tar balls" bezeichnet.
- Die GNU-Version von tar hat viele Optionen.
- Die Option "c" erstellt eine neue Archivdatei und "f" gibt den Namen der zu erstellenden Datei an.
- Um den Inhalt eines tar balls zu sehen, verwendet man die Option "t".
- Die Option "-v" gibt die Namen der Dateien aus, mit denen tar arbeitet.
- Um eine Datei aus dem Archiv zu extrahieren, gibt man diese hinter dem Dateinamen des Archivs an.
- Mit Ausnahme von absoluten Pfaden behalten tar-Dateien den gesamten Pfad zu Dateien bei.
- Die Option "P" wird genutzt, um den absoluten Pfad in einer tar-Datei zu verwenden.
- tar kann auch Komprimierung und Dekomprimierung von Archiven on-the-fly durchführen.
- Die am häufigsten verwendeten Optionen für die Komprimierung sind "j", "J" und "z" für bzip2, xz und gzip.
- Bei der Erstellung von komprimierten .tar-Archiven sollte immer eine zweite Dateierweiterung hinzugefügt werden, die den verwendeten Algorithmus angibt.
- Es ist möglich, Dateien zu bereits vorhandenen unkomprimierten tar-Archiven hinzuzufügen über die Option "u".

- Wenn man versucht, Dateien zu einem komprimierten Archiv hinzuzufügen, erhält man eine Fehlermeldung.

## Verwalten von ZIP-Dateien

---

- Windows-Maschinen haben oft keine Anwendungen für tar balls und andere Linux-Kompressionstools
- ZIP-Dateien können verwendet werden, um mit Windows-Systemen zu interagieren
- ZIP-Dateien ähneln komprimierten tar-Dateien
- `zip` und `unzip` können für die Arbeit mit ZIP-Dateien auf Linux-Systemen genutzt werden
- Beispiel für die Verwendung von `zip` und `unzip` gegeben
- Die Option `-r` bewirkt, dass `zip` auch den Inhalt von Verzeichnissen in der ZIP-Datei enthält