COMPSCI 1XC3 (Winter 2026)

Lab 4-1: Compression, Archiving & Search

By: Nicole Sorokin

Today's Agenda

1. Lab setup
2. Why compression?
3. Compressing files (`gzip` , `bzip2`)
4. Archiving directories (`tar`)
5. Searching for files (`locate` , `find`)
6. TA check-off

> **Labs = Practice & Learning, not testing. Mistakes are expected and welcome!**

Create Your Lab Workspace

Run the following commands:

```
mkdir -p ~/1xc3/lab4-1
cd ~/1xc3/lab4-1
```

1. Why Compression?

Compression is used to:

- Reduce file size
- Save disk space
- Speed up file transfers
- Make large files easier to move or store

Compression does not combine files - it only shrinks them.

2. gzip & gunzip

Create a large file:

```
ls -l /bin /usr/bin > foo.txt
ls -lh foo.txt
```

- `foo.txt` contains directory listings
- This makes it large enough to see compression effects clearly

Compress the File

```
gzip foo.txt
ls -lh
```

What happened to `foo.txt`?

- `foo.txt` is replaced by `foo.txt.gz`
- The original file no longer exists
- This is `gzip`'s default behavior

Uncompress the File

```
gunzip foo.txt.gz
ls -lh
```

- `foo.txt.gz` is removed
- `foo.txt` is restored to its original form

Viewing Compressed Output

```
gzip -c foo.txt > foo_copy.gz
gunzip -c foo_copy.gz | less
```

- `-c` writes output to standard output
- Original files are not replaced
- Useful for pipelines and inspection

3. bzip2 & bunzip2

```
bzip2 foo.txt
ls -lh
```

- Similar behavior to `gzip`
- Uses a different compression algorithm
- Often produces smaller files, but is slower

Uncompress

```
bunzip2 foo.txt.bz2
```

4. Archiving with tar

```
mkdir playground
touch playground/file_{1..5}.txt
tar cvf playground.tar playground
```

- `c` → create archive
- `v` → verbose output
- `f` → archive filename

Extracting Archives

```
rm -r playground
tar xvf playground.tar
```

- `x` → extract archive
- Directory and files are restored exactly

View Archive Contents

```
tar tvf playground.tar
```

- Lists files inside the archive
- Does not extract anything

5. Searching for Files

Searching by Pathname — locate

```
locate bin/zip
```

- Searches file paths, not file contents
- Uses a prebuilt database
- Very fast, but may miss recent files

Filtering Results with grep

```
locate zip | grep bin
```

- `locate zip` finds all paths containing `zip`
- `grep bin` filters results that contain `bin`

Searching Directories — find

- Searches directories recursively
- Does not use a database
- Can be slower but always up-to-date

```
find .
```

- Lists files and directories under current directory

```
find ~ | wc -l
```

- Counts all files and directories under your home directory

Specifying Search Criteria

`find` allows searching based on attributes such as:

- File name
- File type
- File size
- Permissions

Searching by Type

```
find ~/1xc3 -type d | wc -l
```

- Counts directories under `~/1xc3`

```
find ~ -type f | wc -l
```

- Counts regular files under home directory

Searching by Name and Size

```
find ~ -type f -name "*.png" -size +50k | wc -l
```

- Finds PNG files larger than 50 KB
- `*` is quoted to prevent shell expansion

TA Check-off — Task 1 (`gzip`)

Goal:

Understand file compression and restoration using `gzip`.

Task (Figure the commands out and show them to me):

1. Create a large text file
2. Compress it using `gzip`
3. Verify the original file is replaced
4. Restore the file using `gunzip`
5. Verify the original file is back

TA Check-off — Task 2 (`bzip2`)

Goal:

Compress and uncompress a file using `bzip2`.

Task (Figure the commands out and show them to me):

1. Compress a text file using `bzip2`
2. Verify the `.bz2` file exists
3. Uncompress the file using `bunzip2`
4. Verify the original file is restored

TA Check-off — Task 3 (`tar` — Archiving)

Goal:

Archive and extract a directory using `tar`.

Task (Figure the commands out and show them to me):

1. Create a directory with multiple files
2. Archive the directory into a `.tar` file
3. Remove the original directory
4. Extract the archive
5. Verify the directory and files are restored

TA Check-off — Task 4 (`locate`)

Goal:

Search for files using pathname lookup.

Task (Figure the commands out and show them to me):

1. Use `locate` to find paths related to `zip` or `bin`
2. Filter results to only show paths containing `bin`
3. Show one valid result

TA Check-off — Task 5 (`find` with Criteria)

Goal:

Use `find` to search directories with constraints.

Task (Figure the commands out and show them to me):

1. Search for PNG files under your home directory
2. Restrict results to files larger than 50 KB
3. Count how many matching files exist