

INF333 2023-2024 Spring Semester

Burak Arslan

Lab II

Feb 23th, 2024

Duration: 120 minutes.

Suggested Reading

Basic Commands:

cd	cd -, cd ..
ln	Manages links
rm	Unlinker
cp	Copier
mv	Mover
grep	The grand filtering machine

Advanced commands:

for	for i in *.bin; do ...;; done;
sha256sum	Returns sha256sum of a file in hex format
sort	sorts <code>stdin</code> into <code>stdout</code>
uniq	writes deduplicated lines from <code>stdin</code> to <code>stdout</code> , expects sorted input <code>uniq -wN</code> only considers the uniqueness of the first N bytes
cut	Basic input manipulation <code>cut -d" " -f1,2,3</code>
stat	Show detailed inode and directory entry information
man	system and shell documentation

Additional Concepts:

glob	*.bin etc.
------	------------

1 Hardlinking Duplicate Files (100pts)

Lab data is here: <https://burakarslan.com/inf333/tp/tp1.tar.gz>

This data contains 100 files that take 128 bytes each.

```
$ ls -1 | wc -l
100
```

```
$ ls -1 | sha256sum
012b7cc03135dde3c5755be19cd9c48ed502933463d3306b6c8423693f6ad781 -
```

```
$ du --bytes
12800 .
```

Question:

- 1.1. Write a bash script that hardlinks duplicate files together so that file data is deduplicated.
- 1.2. Compare the `stat` output for files before and after deduplication. What changed? Explain.

After deduplication, note that the `ls` output remains the same

```
$ ls -1 | wc -l
100
```

```
$ ls -1 | sha256sum
012b7cc03135dde3c5755be19cd9c48ed502933463d3306b6c8423693f6ad781 -
```

...whereas the disk usage is down 90%:

```
$ du --bytes
1280 .
```