

Unix/Linux Tutorial for Beginners

Session VII

Mihaela Martis

NBIS & Faculty of Medicine and Health Sciences
Division Cell Biology, IKE

history

- enables the repeating of commands entered earlier in the session
- the GNU History library keeps track of all lines typed in the terminal

```
$ history
1997 find . -name '*.tex' | sort -n | wc -l
1998 exit
1999 cd myTeaching/linux_introduction/
2000 kile slides/linux_session6.tex&
2001 cd ..
...
2017 man history
2018 history
```

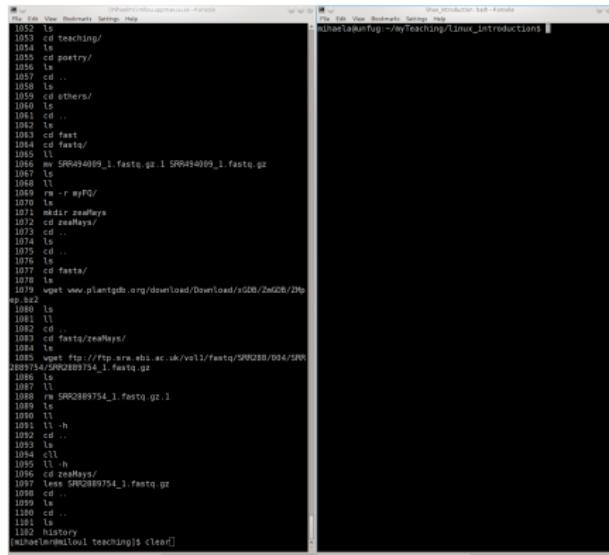
- use  and  to navigate through the history

clear

- clears the screen and shows only the prompt

```
$ clear
```

- is equivalent to **Ctrl + L**



The image shows two side-by-side terminal windows. Both windows have a title bar that reads "MihaelaL@Milo: ~ / My Teaching/Linux_Introductions" and a menu bar with "File", "Edit", "View", "Devices", "Settings", and "Help". The left terminal window contains a series of command-line entries from line numbers 1052 to 1092. These commands include navigating through directories like "Desktop/", "poetry/", "ethers/", "zeaMays/", and "fastq/", listing files with "ls", and extracting files with "tar -xvf". It also includes commands to download files from "www.plantgdb.org" and "ftp.eva.mpg.de". The right terminal window is mostly black, indicating it has been cleared by the "clear" command.

```
1052 ls
1053 Desktop/
1054 ls
1055 cd poetry/
1056 ls
1057 cd ..
1058 ls
1059 cd ethers/
1060 ls
1061 cd ..
1062 ls
1063 cd fastq
1064 cd fastq
1065 ll
1066 SRP0434099_1.fastq.gz.l SRP0434099_1.fastq.gz
1067 ls
1068 ll
1069 tar -xvf SRP0434099_1.fastq.gz.l
1070 ls
1071 mkdir zeaMays
1072 cd zeaMays/
1073 ls
1074 ls
1075 cd ..
1076 ls
1077 cd fastq/
1078 ls
1079 wget www.plantgdb.org/download/Download/xGDB/ZmGDB/Zmp
1080 xz2
1080 ls
1081 ll
1082 cd ..
1083 cd fastq/zeaMays/
1084 ls
1085 wget ftp://ftp.eva.mpg.de/vol1/fastq/SRR280/004/SRR
2808754/SRR2808754_1.fastq.gz
1086 ls
1087 ll
1088 SRP089754_1.fastq.gz.l
1089 ls
1090 ll
1091 rm -h
1092 cd ..
1093 ls
1094 ll
1095 ll -h
1096 cd zeaMays/
1097 rm -h SRP089754_1.fastq.gz.l
1098 cd ..
1099 ls
1100 cd ..
1101 ls
1102 history
[mihaela@Milo teaching]$ clear()
```

Checking disk space

- **du** – shows how much disk space is taken by your files

```
$ cd /proj/g2015039/nobackup/nov2015/tuesday  
$ du -hs .  
4,2M .
```

→ displays combined size of all files in the current directory and recursively in all its subdirectories

Checking disk space

- **du** – shows how much disk space is taken by your files

```
$ cd /proj/g2015039/nobackup/nov2015/tuesday  
$ du -hs .  
4,2M .
```

→ displays combined size of all files in the current directory and recursively in all its subdirectories

→ displays combined size + the size of each subdirectory

```
$ du -h --max-depth=1 .  
32K    ./TEST  
224K   ./molecules  
448K   ./plain_text  
3,5M   ./sequences  
4,2M   .
```

Checking disk space

- `df -h` – shows how much disk space is available

```
$ df -h
Filesystem      Size  Used  Avail Use% Mounted on
/dev/mapper/vg-root   32G   14G   17G  44% /
tmpfs            63G   4,0K   63G   1% /dev/shm
/dev/sda1        2,0G  277M  1,6G  15% /boot
/dev/mapper/vg-scratch 234G   6,0G  216G   3% /scratch
gulo@tcp0:/glob    1,1P  598T  452T  57% /gulo
pica1-v3:/pica/v3   228T  169T   60T  74% /pica/v3
pica1-v2:/pica/v2   228T  165T   64T  73% /pica/v2
```

column

- formats the input into multiple columns that are much easier to read
- usage: `column <options> <file>`

```
$ grep -v "^#" mySample_somaticMutations.vcf | cut -f 1-8|head
chr1 10043 . T G . REJECT .
chr1 10055 . T G . REJECT .
chr1 10067 . T G . REJECT .
chr1 10079 . T G . REJECT .
chr1 10157 . T C . REJECT .
chr1 10180 rs201694901 T C . REJECT DB
chr1 10250 rs199706086 A C . REJECT DB
chr1 726859 rs139100483 A G . REJECT DB
chr1 726887 . A G . REJECT .
chr1 726895 rs141325488 A G . REJECT DB
$ grep -v "^#" mySample_somaticMutations.vcf | cut -f 1-8|column -t|head
chr1 10043 . T G . REJECT .
chr1 10055 . T G . REJECT .
chr1 10067 . T G . REJECT .
chr1 10079 . T G . REJECT .
chr1 10157 . T C . REJECT .
chr1 10180 rs201694901 T C . REJECT DB
chr1 10250 rs199706086 A C . REJECT DB
chr1 726859 rs139100483 A G . REJECT DB
chr1 726887 . A G . REJECT .
chr1 726895 rs141325488 A G . REJECT DB
```

Join

- **join** – used to join different files together by a common column
- works only if both files are sorted by the column to be joined on

```
$ cat example.bed
chr1 26 39
chr1 53 84
chr3 32 99
chr1 9 28
chr2 10 19
$ cat example_length.txt
chr1 58352
chr2 39521
chr3 24859

$ sort -k1,1 example.bed > example_sorted.bed
$ sort -k1,1 example_length.txt > example_length_sorted.txt

$ join -1 1 -2 1 example_sorted.bed example_length_sorted.txt
chr1 26 39 58352
chr1 53 84 58352
chr1 9 28 58352
chr2 10 19 39521
chr3 32 99 24859
```

Comparing files

- `diff` – reports differences between files.
- usage: `diff [OPTIONS] FILE1 FILE2`
- useful options:
 - `-b` – ignore blanks
 - `-w` – ignore white spaces and tabs
 - `-i` – ignore case
 - `-r` – recursively compare all files (when comparing folders)
 - `-y` – side by side comparison of files
- normal output shows only the lines that are different between 2 files: <*FROM-FILE-LINE*> *TO-FILE-LINE*

diff example

```
$ diff -y genList1.txt genList2.txt
1APM:I           1APM:I
1APM:E           1APM:E
1AY6:I           1AY6:I
1AY6:K           | 1AY6:H
1AY6:L           1AY6:L
1BH3:A           1BH3:A
1BRR:D           | 1BRR:A
1BRR:B           1BRR:B
1BRR:C           1BRR:C
1BXW:A           1BXW:A
```

Interpreting the output

```
$ diff genList1.txt genList2.txt
gen4c4
< 1AY6:K
_____
> 1AY6:H
7c7
< 1BRR:D
_____
> 1BRR:A
```

- 1st line – **gen4c4**:
 - **c** – changed (**a** – added, **d** – deleted)
 - left number – line numbers of the original file
 - right number – line numbers of the modified file
- 2nd line – < **1AY6:K** – shows lines from the first file that are different from the second file
- 3rd line – a divider
- 4th line – > **1AY6:H** – shows lines from the second file that are different from the first one

Downloading data

- `wget` – is a tool for non-interactive download of files from the Web (http, https, ftp)
- `wget 'http://website.url'`

```
$ wget "http://www.rcsb.org/pdb/files/1ema.pdb"
--2015-08-14 17:50:59--  http://www.rcsb.org/pdb/files/1ema.pdb
Resolving www.rcsb.org (www.rcsb.org)... 128.6.70.10
Connecting to www.rcsb.org (www.rcsb.org)|128.6.70.10|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/plain]
Saving to: 1ema.pdb.1
[=> ] 191 403      389KB/s   in 0,5s
2015-08-14 17:50:59 (389 KB/s) - 1ema.pdb.1 saved [191403]
```

Example curl

- curl – is a tool to transfer data from or to a server (http, https, ftp, imap, pop3)

→ the data is downloaded directly to the screen

```
curl 'http://website.url'
```

```
$ curl "http://www.rcsb.org/pdb/files/1ema.pdb"
% Total % Received % Xferd  Average Speed Time Time Time Current Dload Upload
Total Spent Left Speed
0 0 0 0 0 0 0 0 --:--:-- --:--:-- --:--:-- 0HEADER FLUORESCENT PROTEIN 01-AUG
-96 1EMA
TITLE GREEN FLUORESCENT PROTEIN FROM AEQUOREA VICTORIA
COMPND MOL_ID: 1;
...
```

```
$ curl "http://www.rcsb.org/pdb/files/1ema.pdb" > 1ema.pdb
$ curl "http://www.rcsb.org/pdb/files/1ema.pdb" -o 1ema.pdb
$ curl "http://www.rcsb.org/pdb/files/{1ema,1gfl,1g7k,1xmz}.pdb" -o '#1'.pdb
```

Summary

- `history` – returns a history of used command lines
- `clear` – clears the screen
- `join` – join different files together by a common column
- `column` – formats the input into multiple columns
- `df` – shows how much disk space is available
- `du` – shows how much disk space is taken by your files
- `diff` – reports differences between files
- `wget,curl` – download data from an url from the shell