MBDP-102

Introduction to Unix

What is Unix?



A family of computer operating systems (OSs)

Linux, MacOS, Solaris, OpenBSD

Key characteristics

Multitasking: multiple software processes can run at the same time

Multiuser: several users can use the same computer at the same time

Multiprocessing: can use more than one computer processor

Portable: can be used in various hardware architectures

The Unix philosophy

"The idea that the power of a system comes more from the relationships among programs than from the programs themselves"

- Use of a large number of simple programs performing a limited, welldefined function
- Use of a command-line interpreter ("shell") to combine these programs to perform complex tasks
- Use of plain text for storing data
- Use of a hierarchical filesystem

The Unix shell (command-line interpreter)

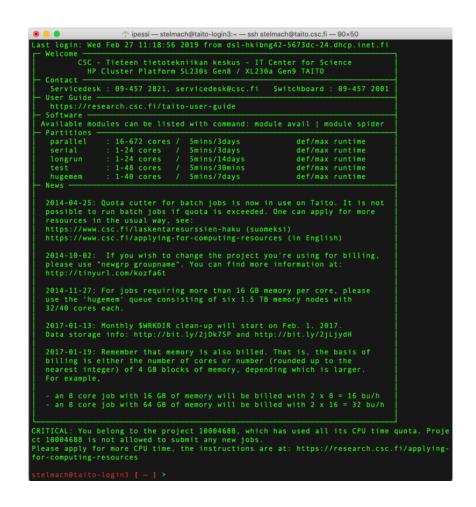
Interprets sequences of text

- Entered by a user
- From a file
- From a data stream

Primary interface before graphical user interfaces (GUIs) appeared

Still widely used today

- Efficient
- Low memory footprint
- Advanced scripting



Some basic Unix commands

```
pwd: print working directory ("where am I?")
ls: list ("show folder contents")
mkdir: make directory (a.k.a. folder)
cd: change directory ("go to folder")
cp: copy
mv: move
rm: <u>rem</u>ove
```

Some additional notes

Case-sensitive

photo.jpg ≠ PHOTO.jpg

Does not like spaces and special characters in file/folder names

- genome report.txt X
- genome_report.txt
- Tromsø.txt X
- Tromso.txt 🗸

Some additional notes

Running commands

- Space after each "word" in the command
- Typed in a single line, one at a time
- After each command, hit "Enter" to execute it
- Lines starting with "#" are comments

Directory navigation:

- One dot (.) means "here"
- To go up one folder: ../
- To go up two folders: ../../

CSC - IT CENTER FOR SCIENCE

- Working with 'omics data is computing-intensive
- Use of high-performance computers (HPC)
 - a.k.a. supercomputers



Disk areas in Puhti:

https://docs.csc.fi/computing/disk

	Owner	Environment variable	Path	Cleaning
home	Personal	\$(HOME)	/users/ <user-name></user-name>	No
projappl	Project	Not available	/projappl/ <project></project>	No
scratch	Project	Not available	/scratch/ <project></project>	Yes - 90 days

These disk areas have quotas for both the amount of data and total number of files:

	Capacity	Number of files
home	10 GiB	100 000 files
projappl	50 GiB	100 000 files
scratch	1 TiB	1 000 000 files

Running jobs:

https://docs.csc.fi/computing/running/getting-started

sinteractive

- Interactive
- Limited to 4 CPUs and 64 GB of RAM

sbatch

- Uses a queing sytem (SLURM) to run jobs in the background
- Higher CPU and RAM availability

```
#!/bin/bash
#SBATCH -- job-name CUTADAPT
#SBATCH --error CUTADAPT %A %a err.txt
#SBATCH --output CUTADAPT %A %a out.txt
#SBATCH --partition small
#SBATCH --nodes 1
#SBATCH --ntasks 1
#SBATCH --cpus-per-task 4
#SBATCH --mem 4G
#SBATCH --time 4:00:00
#SBATCH --account project 2001499
#SBATCH --array 1-4
SAMPLE=Sample0${SLURM_ARRAY_TASK_ID}
module load biokit
cutadapt ../COURSE FILES/RAWDATA/$SAMPLE.NOVASEQ.R1.fastq.gz \
         ../COURSE FILES/RAWDATA/$SAMPLE.NOVASEQ.R2.fastq.gz \
         -o TRIMMED/$SAMPLE.NOVASEQ.R1.fastq.gz \
         -p TRIMMED/$SAMPLE.NOVASEQ.R2.fastq.gz \
         -a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC \
         -A CTGTCTCTTATACACATCTGACGCTGCCGACGA \
         -m 50 \
         -j 4 \
         --nextseq-trim 20 > TRIMMED/$SAMPLE.cutadapt.log.txt
```

How to learn UNIX?

By using it!

- Trial and error
- Don't copy and paste it, type it

Ask the internet

- http://stackoverflow.com/
- http://stackexchange.com/
- http://askubuntu.com/
- Google!

Cheat sheets

• https://www.guru99.com/linux-commands-cheat-sheet.html

Manual ("man") pages

• man ls

Online courses/tutorials

http://codecademy.com

Let's practice the command line a bit

https://github.com/karkman/Metagenomics2021