



UiO : **Department of Biosciences**
University of Oslo

BIOS3010 - Bioinformatics - Spring2020

Introduction to Unix

Anders K. Krabberød
a.k.krabberod@ibv.uio.no



Program for week 19

- Introduction to UNIX and Linux
- Briefly on the Unix shell, basic commands
- Basic Unix exercises
- Setting up your computers
- Logging on to Saga!
- Prepare for what's to come – working with “big data”

Why learn about UNIX and Linux?

- Bioinformatics tools and algorithms are very often developed for UNIX/Linux computers.
- Easy to set up pipelines and automate workflows.
- Powerful for handling text files, large files, and large number of files – very common in modern biology.
- Able to work on High Performance Computing platforms (Supercpomputers) e.g. Saga
- Access to much more computing power than from a single computer.

UNIX? Linux? Computer cluster? Saga???



HPC (High Performance Computing) cluster

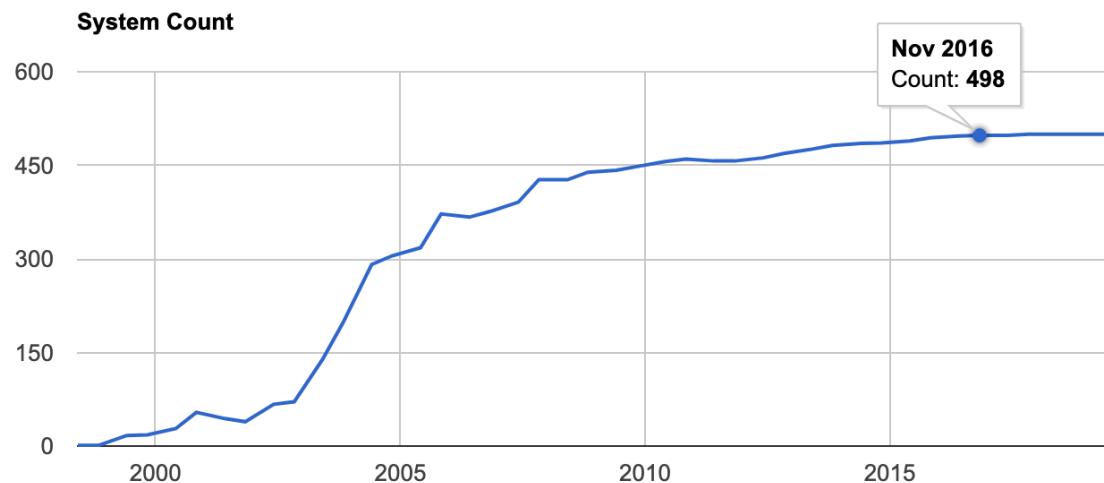
- One of the motivations for learning Linux is that most *supercomputers* or *computing clusters* run a version of Linux or UNIX as the OS.
- Most (if not all) of the supercomputer's on the list of the top 500 most powerful computers use Linux.



MareNostrum in Barcelona. 40 000 kg of computers inside an old church. Annual cost in electricity alone is € 1.4m. www.bsc.es

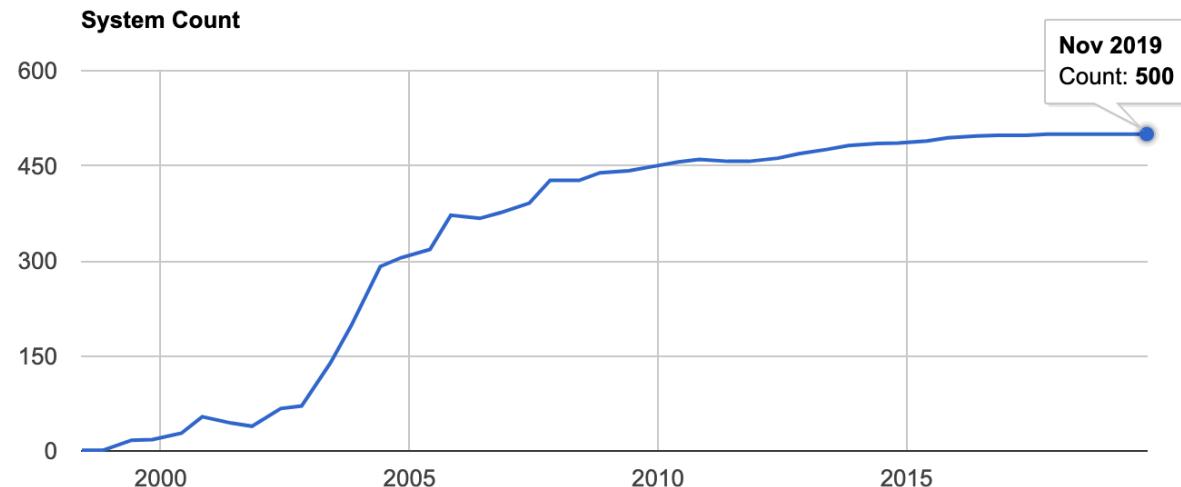
The prevalence of Linux

- Linux OS on the top500 supercomputers in the world



The prevalence of Linux

- Linux OS on the top500 supercomputers in the world



HPC (High Performance Computing)

- In Norway Sigma2 runs several HPC (High Performance Computing) clusters as well as other e-infrastructure facilities. <https://www.sigma2.no/services-overview>
- Current clusters
 - **Saga**
 - Fram
 - Vilje
 - Stallo
 - Betzy (under construction)

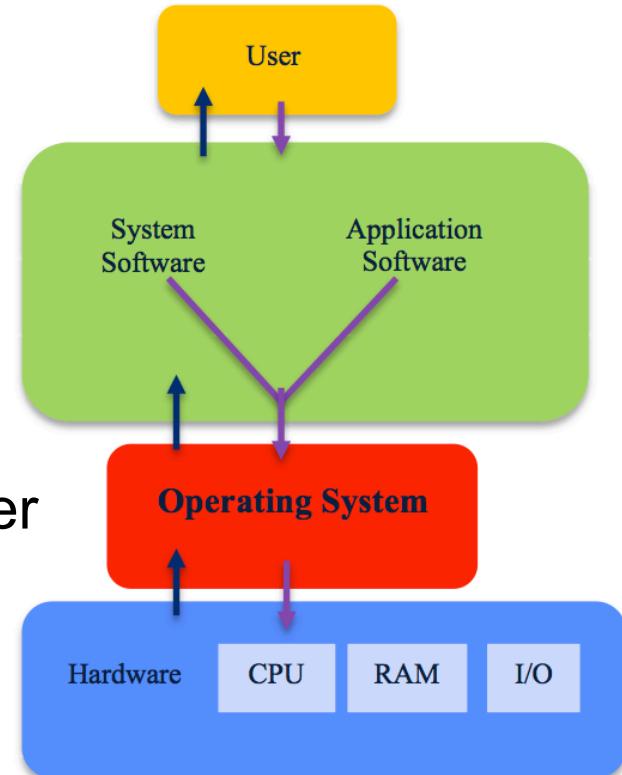
Saga

- We will use Saga for this course
- Online as of fall 2019
- A computing cluster with 200 standard compute nodes, with 40 cores and 192 GiB memory each,
- Total 9824 cores
- Access to resources is managed with a queue system: SLURM
- OS on Saga Linux

Details	Saga
System	Hewlett Packard Enterprise - Apollo 2000/6500 Gen10
Number of Cores	9824
Number of nodes	244
Number of GPUs	32
CPU type	Intel Xeon-Gold 6138 2.0 GHz (normal) Intel Xeon-Gold 6130 2.1 GHz (bigmem) Intel Xeon-Gold 6126 2.6 GHz (accel)
GPU type	NVIDIA P100, 16 GiB RAM (accel)
Total max floating point performance, double	645 Teraflop/s (CPUs) + 150 Teraflop/s (GPUs)
Total memory	75 TiB
Total NVMe+SSD local disc	89 TiB + 60 TiB
Total parallel filesystem capacity	1 PB

Operating systems (OS)

- Software that manages computer hardware
 - input from keyboard and pointing devices.
 - output to screen (or file).
 - keeps track of, reads and writes files.
- Provides common services for computer programs



UNIX: the mother of (almost) all OS

- UNIX was Initially developed in 1969 (Bell Labs / AT&T)
- Coded mostly in C (also developed by AT&T)
- **Multi-tasking** (resources are shared)
- **Multiuser** (can handle many users simultaneously)



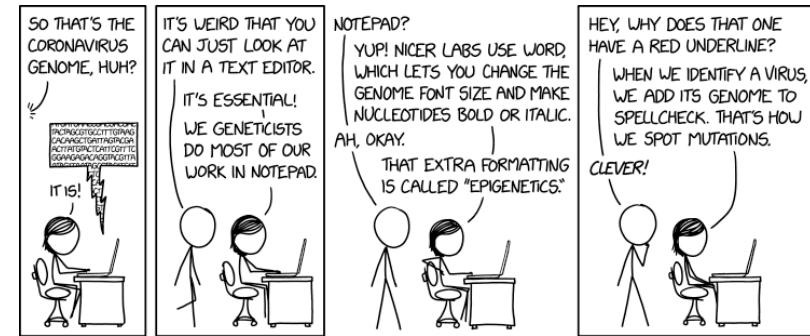
[Ken Thompson](#) and [Dennis Ritchie](#), two of the UNIX developers

The UNIX philosophy

- The success of the UNIX framework is because of the philosophy behind the development
 - Write programs that do **one** thing and **do it well**.
 - Write programs that will **work together**.
 - Write programs to handle **text** streams, because that is a universal interface.

Demystifying files

- Many files used in bioinformatics are just **text** files.
- Extension given to the file reflects the format, or how the “internal layout” of that file is.
- Often the format is formulated as set of rules that tells the user what the file should contain and how it should be organised



```
>Seq1  
ATGGTATGACCATGT  
>Seq2  
ATGGTGGCTTGCTG
```

Demystifying files - examples

- Fasta
 - Lines starting with > is the sequence identifier
 - Lines that follow *without* > is the sequence (DNA, RNA or Amino Acids)
 - Fastq -
 - Sequence with quality score
 - 4 lines per sequence
 - Line 1 begins with a '@' followed by header
 - Line 2 is the sequence
 - Line 3 begins with a '+' and is *optionally* followed by the sequence identifier and description
 - Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.
-
- The diagram illustrates the structure of FASTA and FASTQ files. Arrows point from the list items to specific lines in the example sequences. In the FASTA example, arrows point from the sequence identifiers to the first line of each sequence. In the FASTQ example, arrows point from the sequence identifiers to the first line, from the '+' symbol to the third line, and from the quality scores to the fourth line.
- FASTA Example:**
- ```
>First_sequence_id
ATGGTATGACCATGT
>Next_sequence_id
ATGGTGGCTTGCTGGGCCGGT
```
- FASTQ Example:**
- ```
@First_sequence_id
ATGGTATGACCATGT
+
! ' ' * ( ( ( ( ***+ ) ) %1
@Next_sequence_id
ATGGTGGCTTGCTGGGCCGGT
+
%%) . 1 ***-+* ' ') ) **55) KM
```

Linux

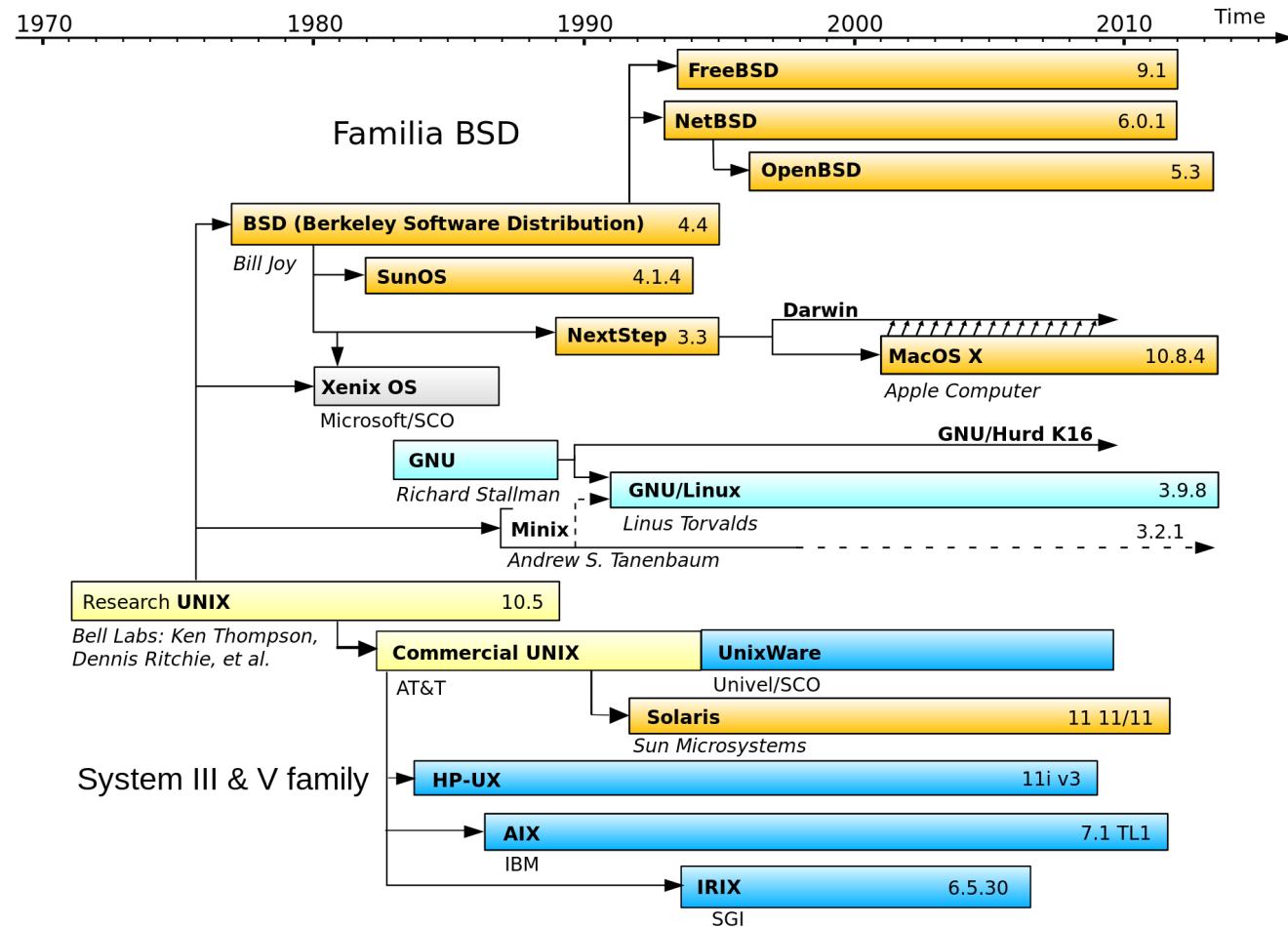
- Development started by Linus Torvalds in 1991
- UNIX clone
- Free and open Source
- Now comes in many variants
 - Ubuntu
 - Debian
 - Centos
 - +++++
- Some variants are portable enough to be installed on a usb pendrive.



Linus Torvalds with the first Linux kernel in 1991



Timeline of UNIX/ Linux development



For more about the relation between UNIX and Linux: https://youtu.be/jowCUo_UGts

- Advantages with Linux
 - Stability
 - Security
 - Good control of users and processes
 - Can handle large computing load
 - Open Source (you can see and modify the code)
 - Several free options (Ubuntu, CentOS, Debian, etc.)
 - Large community of developers
 - Large amount of free software (specially for bioinfo)
- Disadvantages
 - Some hardware may not be supported
 - Can be less user friendly than other options
 - (For CLI) It's not point-and-click

The prevalence of UNIX and Linux

- Android phones runs Linux
- The MacOS is based on a UNIX clone
- The iOS on iPhones is also based un UNIX
- The World Wide Web is basically a network of Linux-server
- Windows computers do **not** use UNIX

Poll

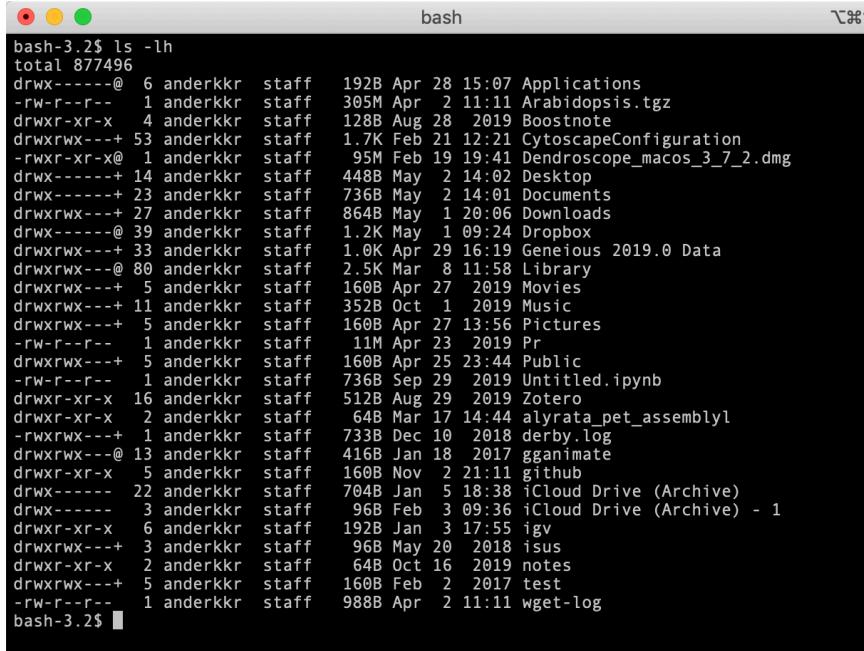
- What kind of computer do you use?
- What kind of OS do you run?
 - Windows
 - Mac
 - Linux?
- Have you used Linux before?

Terminology - recap

- **UNIX** is the original OS
- **Linux** is not UNIX, but it is based on the same underlying philosophy (more on this later).
- **GNU** is not UNIX
 - which incidentally is the intention of the acronym GNU: GNU is a recursive acronym for “GNU's Not Unix”
- A **shell** is an access interface for the operating system. It is a program used for controlling and running programs.
- **Bash** is a shell program (others include sh, ash, zsh, ksh...)
- A **terminal** is a program that lets the user interact with the shell. It interprets the keystrokes from the user and passes it on to the shell.
- A **kernel** is the central part of an **operating system**. It manages the operations of the computer and the hardware, most notably memory and CPU time
- **POSIX** stands for Portable Operating System Interface, and is an attempt by a consortium of vendors to create a single standard version of UNIX

Questions?

UNIX is text based



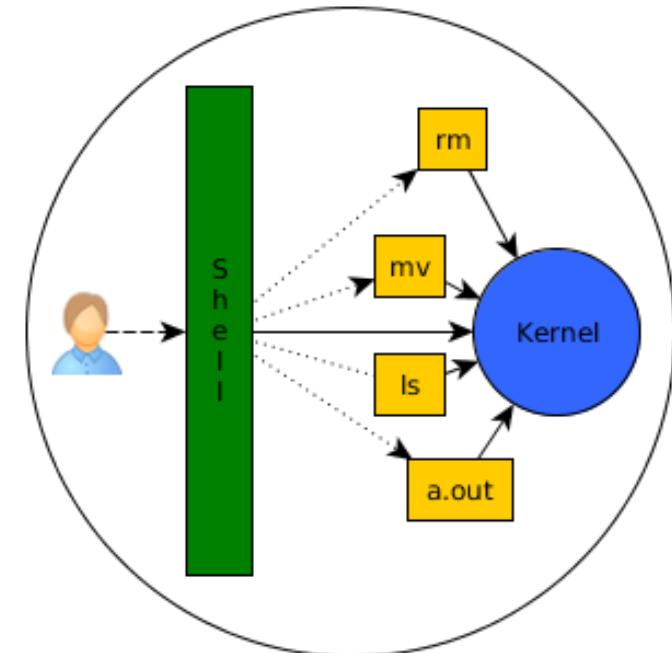
A screenshot of a macOS terminal window titled "bash". The window shows the output of the command "ls -lh". The output lists various files and directories in a long-form listing format, showing permissions, owner, group, size, date modified, and name. The terminal window has a dark background with light-colored text. The title bar and window controls are visible at the top.

```
bash-3.2$ ls -lh
total 877496
drwx-----@ 6 anderkkr  staff  192B Apr 28 15:07 Applications
-rw-r--r--  1 anderkkr  staff  305M Apr  2 11:11 Arabidopsis.tgz
drwxr-xr-x  4 anderkkr  staff  128B Aug 28  2019 Boostnote
drwxrwx---+ 53 anderkkr  staff  1.7K Feb 21 12:21 CytoscapeConfiguration
-rwxr-xr-x@ 1 anderkkr  staff  95M Feb 19 19:41 Dendroscope_macos_3_7_2.dmg
drwx-----+ 14 anderkkr  staff  448B May  2 14:02 Desktop
drwx-----+ 23 anderkkr  staff  736B May  2 14:01 Documents
drwxrwx---+ 27 anderkkr  staff  864B May  1 20:06 Downloads
drwx-----@ 39 anderkkr  staff  1.2K May  1 09:24 Dropbox
drwxrwx---+ 33 anderkkr  staff  1.0K Apr 29 16:19 Geneious 2019.0 Data
drwxrwx---@ 80 anderkkr  staff  2.5K Mar  8 11:58 Library
drwxrwx---+ 5 anderkkr  staff  160B Apr 27 2019 Movies
drwxrwx---+ 11 anderkkr  staff  352B Oct  1 2019 Music
drwxrwx---+ 5 anderkkr  staff  160B Apr 27 13:56 Pictures
-rw-r--r--  1 anderkkr  staff  11M Apr 23 2019 Pr
drwxrwx---+ 5 anderkkr  staff  160B Apr 25 23:44 Public
-rw-r--r--  1 anderkkr  staff  736B Sep 29 2019 Untitled.ipynb
drwxr-xr-x  16 anderkkr  staff  512B Aug 29 2019 Zotero
drwxr-xr-x  2 anderkkr  staff  64B Mar 17 14:44 alyrata_pet_assemblyl
-rwxrwx---+ 1 anderkkr  staff  733B Dec 10 2018 derby.log
drwxrwx---@ 13 anderkkr  staff  416B Jan 18 2017 ganimate
drwxr-xr-x  5 anderkkr  staff  160B Nov  2 21:11 github
drwxrwx---+ 22 anderkkr  staff  704B Jan  5 18:38 iCloud Drive (Archive)
drwx-----+ 3 anderkkr  staff  96B Feb  3 09:36 iCloud Drive (Archive) - 1
drwxr-xr-x  6 anderkkr  staff  192B Jan  3 17:55 igv
drwxrwx---+ 3 anderkkr  staff  96B May 20 2018 isus
drwxr-xr-x  2 anderkkr  staff  64B Oct 16 2019 notes
drwxrwx---+ 5 anderkkr  staff  160B Feb  2 2017 test
-rw-r--r--  1 anderkkr  staff  988B Apr  2 11:11 wget-log
bash-3.2$
```

- You need a way to communicate with the OS
- A **terminal** will interpret your keystrokes
- A **shell** interprets the commands

The Shell

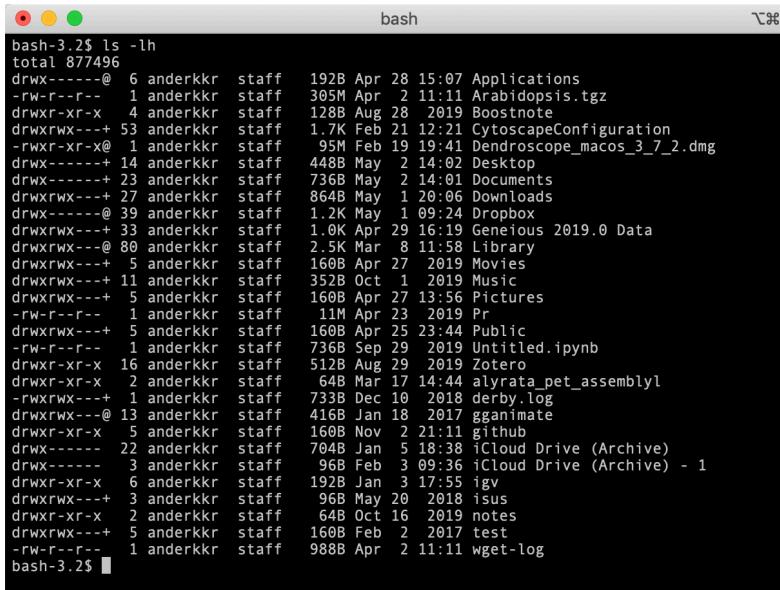
- The **shell** is a user interface that provides access to the operating systems services.
- It is called a shell because it is the outermost layer around the operating system.
- Generally comes in two flavours:
 - **Command-line interface (CLI)**
 - Graphical user interface (GUI)



Command-line interface (CLI) shell

- You type in the commands and press enter. The shell figures out how to interpret the commands and which programs to run.
- The shell is the user interface for communicating with the computer. There are many different kinds of shells and I won't go into detail about this. And for most basic use they will do mostly the same. But the most popular Shell is the Bash (Bourne again shell). It is default on most Linux implementations today.

The Shell



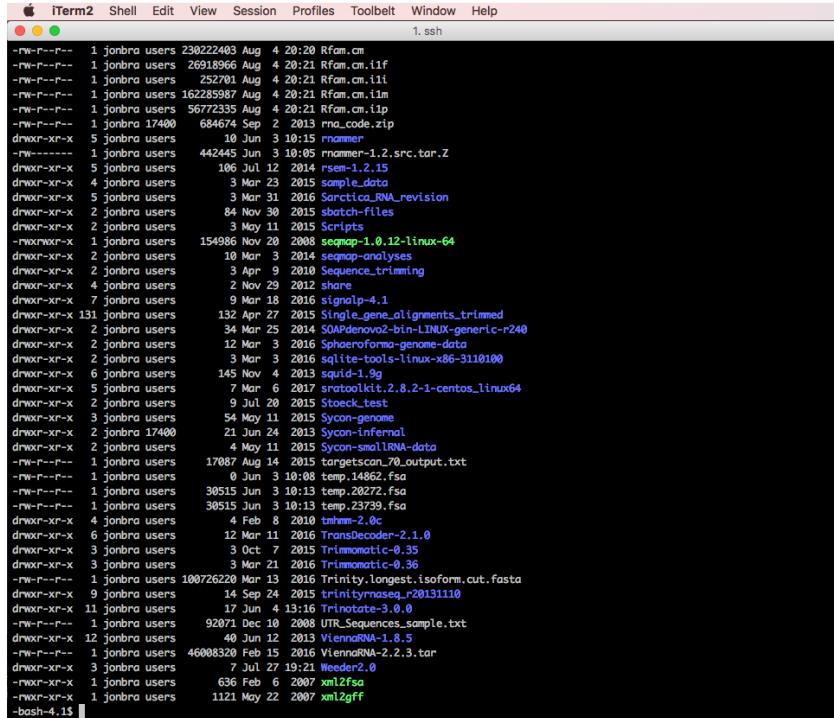
A screenshot of a terminal window titled "bash". The command "ls -lh" is run, displaying a detailed list of files and their metadata. The output includes columns for permissions, owner/group, file size, modification date, and name. Notable entries include "Arabidopsis.tgz", "Boostnote", "CytoscapeConfiguration", "Dendroscope_macos_3.7_2.dmg", "Dropbox", "Geneious 2019.0 Data", "Library", "Movies", "Music", "Pictures", "Pr", "Public", "Untitled.ipynb", "Zotero", and several log files like "derby.log" and "wget-log".

```
bash-3.2$ ls -lh
total 877496
drwx-----@ 6 anderkkr staff 192B Apr 28 15:07 Applications
-rw-r--r-- 1 anderkkr staff 305M Apr 2 11:11 Arabidopsis.tgz
drwxr-xr-x 4 anderkkr staff 128B Aug 28 2019 Boostnote
drwxrwx---+ 53 anderkkr staff 1.7K Feb 21 12:21 CytoscapeConfiguration
-rwxr-xr-x@ 1 anderkkr staff 95M Feb 19 19:41 Dendroscope_macos_3.7_2.dmg
drwx-----+ 14 anderkkr staff 448B May 2 14:02 Desktop
drwx-----+ 23 anderkkr staff 736B May 2 14:01 Documents
drwxrwx---+ 27 anderkkr staff 864B May 1 20:06 Downloads
drwx-----@ 39 anderkkr staff 1.2K May 1 09:24 Dropbox
drwxrwx---+ 33 anderkkr staff 1.0K Apr 29 16:19 Geneious 2019.0 Data
drwxrwx---@ 80 anderkkr staff 2.5K Mar 8 11:58 Library
drwxrwx---+ 5 anderkkr staff 160B Apr 27 2019 Movies
drwxrwx---+ 11 anderkkr staff 352B Oct 1 2019 Music
drwxrwx---+ 5 anderkkr staff 160B Apr 27 13:56 Pictures
-rw-r--r-- 1 anderkkr staff 11M Apr 23 2019 Pr
drwxrwx---+ 5 anderkkr staff 160B Apr 25 23:44 Public
-rw-r--r-- 1 anderkkr staff 736B Sep 29 2019 Untitled.ipynb
drwxr-xr-x 16 anderkkr staff 512B Aug 29 2019 Zotero
drwxr-xr-x 2 anderkkr staff 64B Mar 17 14:44 alyrata_pet_assembly1
drwxrwx---+ 1 anderkkr staff 733B Dec 10 2018 derby.log
drwxrwx---@ 13 anderkkr staff 416B Jan 18 2017 ganimate
drwxr-xr-x 5 anderkkr staff 160B Nov 2 21:11 github
drwx----- 22 anderkkr staff 704B Jan 5 18:38 iCloud Drive (Archive)
drwx----- 3 anderkkr staff 96B Feb 3 09:36 iCloud Drive (Archive) - 1
drwxr-xr-x 6 anderkkr staff 192B Jan 3 17:55 igv
drwxrwx---+ 3 anderkkr staff 96B May 20 2018 isus
drwxr-xr-x 2 anderkkr staff 64B Oct 16 2019 notes
drwxrwx---+ 5 anderkkr staff 160B Feb 2 2017 test
-rw-r--r-- 1 anderkkr staff 988B Apr 2 11:11 wget-log
bash-3.2$
```

- The user writes commands on the command line usually in a terminal window
- Line-by-line
- Step-by-step
- Logical work flow

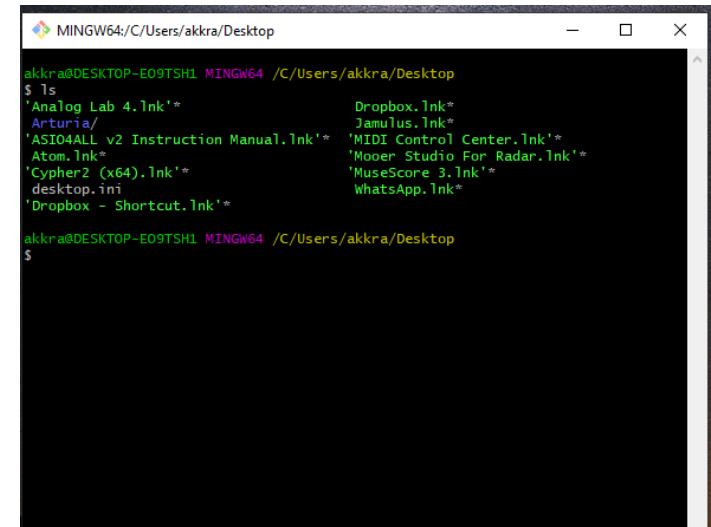
Linux is case-sensitive!

The Shell



```
Apple iTerm2 Shell Edit View Session Profiles Toolbelt Window Help
1. ssh

-rw-r--r-- 1 jonbra users 230222403 Aug 4 20:20 Rfam.cm
-rw-r--r-- 1 jonbra users 26918966 Aug 4 20:21 Rfam.cm.tif
-rw-r--r-- 1 jonbra users 252701 Aug 4 20:21 Rfam.cm.tif
-rw-r--r-- 1 jonbra users 162285987 Aug 4 20:21 Rfam.cm.ilm
-rw-r--r-- 1 jonbra users 56772335 Aug 4 20:21 Rfam.cm.lip
-rw-r--r-- 1 jonbra users 6846574 Sep 2 2013 rna_code.zip
drwxr-xr-x 5 jonbra users 10 Jun 3 10:15 rnammer
-rw----- 1 jonbra users 442445 Jun 3 10:05 rnammer-1.2.src.tar.Z
drwxr-xr-x 5 jonbra users 106 Jul 12 2014 rsem-1.2.15
drwxr-xr-x 4 jonbra users 3 Mar 23 2015 sample_data
drwxr-xr-x 5 jonbra users 3 Mar 31 2016 Sarictica_RNA_revision
drwxr-xr-x 2 jonbra users 84 Nov 30 2015 sketch_files
drwxr-xr-x 2 jonbra users 3 May 11 2015 Scripts
drwxr-xr-x 1 jonbra users 154986 Nov 20 2008 seqmap-1.0.12-linux-64
drwxr-xr-x 2 jonbra users 10 Mar 3 2014 seqmap-analyses
drwxr-xr-x 2 jonbra users 3 Apr 9 2010 Sequence_trimming
drwxr-xr-x 4 jonbra users 2 Nov 29 2012 share
drwxr-xr-x 7 jonbra users 9 Mar 18 2016 signalp-4.1
drwxr-xr-x 131 jonbra users 132 Apr 27 2015 Single_gene_alignments_trimmed
drwxr-xr-x 2 jonbra users 34 Mar 25 2014 SOAPdenovo-2-bin-LINUX-generic-r240
drwxr-xr-x 2 jonbra users 12 Mar 3 2016 Sphaeroverpa-genome-data
drwxr-xr-x 2 jonbra users 3 Mar 3 2016 sqlite-tools-linux-x86-3110100
drwxr-xr-x 6 jonbra users 145 Nov 4 2013 squid-1.9g
drwxr-xr-x 5 jonbra users 7 Mar 6 2017 sratoolkit.2.8.2-1-centos_linux64
drwxr-xr-x 2 jonbra users 9 Jul 20 2015 Stoack_test
drwxr-xr-x 3 jonbra users 54 May 11 2015 Sycon-genome
drwxr-xr-x 2 jonbra users 21 Jun 24 2013 Sycon-infernal
drwxr-xr-x 4 May 11 2015 Sycon-smallRNA-data
drwxr-xr-x 2 jonbra users 17087 Aug 14 2015 targetscan_7_0_output.txt
-rw-r--r-- 1 jonbra users 9 Jun 3 10:08 temp_14862.fsa
-rw-r--r-- 1 jonbra users 30515 Jun 3 10:13 temp_20272.fsa
-rw-r--r-- 1 jonbra users 30515 Jun 3 10:13 temp_23739.fsa
drwxr-xr-x 4 jonbra users 4 Feb 8 2010 tmhmm-2.0c
drwxr-xr-x 6 jonbra users 12 Mar 11 2016 TransDecoder-2.1.0
drwxr-xr-x 3 jonbra users 3 Oct 7 2015 Trimmomatic-0.35
drwxr-xr-x 3 jonbra users 3 Mar 21 2016 Trimmomatic-0.36
-rw-r--r-- 1 jonbra users 100726228 Mar 13 2016 Trinity.longest.isoform.cut.fasta
drwxr-xr-x 9 jonbra users 14 Sep 24 2015 trinityrnaseq_r20131110
drwxr-xr-x 11 jonbra users 17 Jun 4 13:16 Trinityote-3.0.0
-rw-r--r-- 1 jonbra users 92071 Dec 10 2008 UTR_Sequences_sample.txt
drwxr-xr-x 12 jonbra users 40 Jun 12 2013 ViennaRNA-1.8.5
-rw-r--r-- 1 jonbra users 46008320 Feb 15 2016 ViennaRNA-2.3.tar
drwxr-xr-x 3 jonbra users 7 Jul 27 19:21 Weeder2.0
-rw-r--r-- 1 jonbra users 636 Feb 6 2007 xm12fsa
-rw-r--r-x 1 jonbra users 1121 May 22 2007 xm12gff
-bash-4.1$
```



```
akkra@DESKTOP-E09TSH1 MINGW64 /C/Users/akkra/Desktop
$ ls
'Analog Lab 4.lnk'*           Dropbox.lnk*
'Arturia/*                   Jamulus.lnk*
'ASIO4ALL v2 Instruction Manual.lnk'* 'MIDI Control Center.lnk'* 
'Atom.lnk*'                   'Moer Studio For Radar.lnk'* 
'Cypher2 (x64).lnk'*          'MuseScore 3.lnk'* 
'desktop.ini'                 'WhatsApp.lnk'* 
'Dropbox - Shortcut.lnk'* 

akkra@DESKTOP-E09TSH1 MINGW64 /C/Users/akkra/Desktop
$
```

The Bash shell is default on
most Linux implementations.
Bourne-again shell

The Shell

- Getting a shell
- **Windows:** Gitbash for windows (<https://gitforwindows.org/>),
 - click download, scroll down and choose the latest 64-bit .exe file. Install then start *gitbash* from the start menu.
 - *Other options for shell exists and if you have something you are comfortable with using already installed, that is fine.*
- **Mac and Linux:** A shell is already installed (launch the Terminal application).

- If you've never used the terminal, you can start with simple commands and a good amount of patience ☺
- There's plenty of resources and documentation online
 - Goolge is your friend!
- A set of instructions and exercises made for the course will help you:
- <https://github.com/krabberod/bios3010-unix-linux>

Poll

- Have you used a terminal/shell/command line before?
- Do you know how to log on to a server?

- Coffee break!

Login into Saga

- Open terminal (mac/linux) or start gitbash (windows)
- Type `ssh <username>@saga.sigma2.no`
- Type password

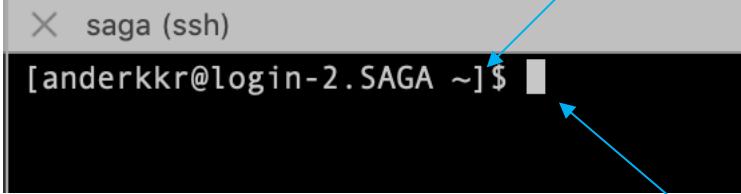
```
(base) [anderkkr:~]$ ssh anderkkr@saga.sigma2.no
anderkkr@saga.sigma2.no's password:
Last login: Fri May  1 19:13:14 2020 from 84.211.16.85

Welcome to saga.sigma2.no!

Documentation:  https://documentation.sigma2.no/
Support email: support@metacenter.no
Request resources:
https://www.sigma2.no/content/apply-e-infrastructure-resources/
~~~~~
Latest news from:      https://opslog.sigma2.no/

o 2020-04-29: Stallo problems / urgent maintenance
o 2020-04-27: Stallo RAM upgrade
o 2020-04-24: Downtime 20th - 24th of April is over. Services are back in prod
o 2020-04-16: Stallo slurm problem
o 2020-04-08: tos-project3 on NIRD is read only
~~~~~
WARNING: Please save all your important files.
          Backup is taken *only* for $HOME which has quota in place.
          For more info see: https://documentation.sigma2.no/storage/backup.html

-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
[anderkkr@login-1.SAGA ~]$ █
```



The screenshot shows a terminal window titled "saga (ssh)". Inside the window, the command line prompt "[anderkkr@login-2.SAGA ~]\$" is displayed, followed by a small black square cursor. The window has a light gray header bar and a dark gray body.

The dollar sign is a **prompt**, shows that the shell is waiting for input.

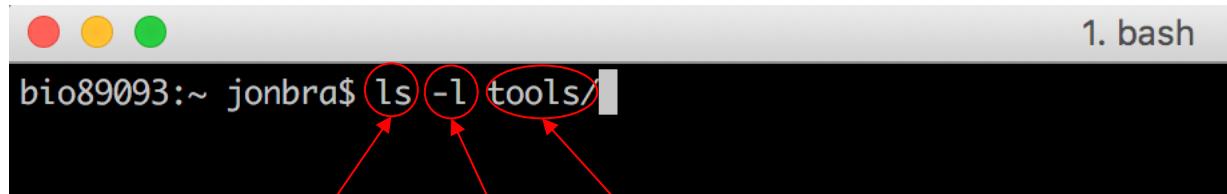
The cursor. This is where you write commands

General comments on commands

- The commands operates on text. Either files containing text or input from the keyboard (called *standard input*).
- Commands in Linux usually takes the form
 $\textit{command} [\textit{-options}] \textit{inputfile}$
- The philosophy of Linux (and UNIX) programs is that they are supposed to be simple and efficient with a standardised way of taking inputs and giving outputs. Because of the standardisation, the output of one command can (often) be used as the input of another command. The advantage of this will be clear later.
- The full manual for a command can be accessed with $\textit{man command}$

Unix command syntax

command [-option(s)] [argument(s)]



A screenshot of a terminal window titled "1. bash". The window shows the command "ls -l tools/" entered by the user "jonbra\$". The command "ls" is highlighted with a red oval, followed by the option "-l" and the argument "tools/".

command

option

argument

There must be spaces
between commands,
options and arguments

Remote login program (ssh)

- `man ssh`

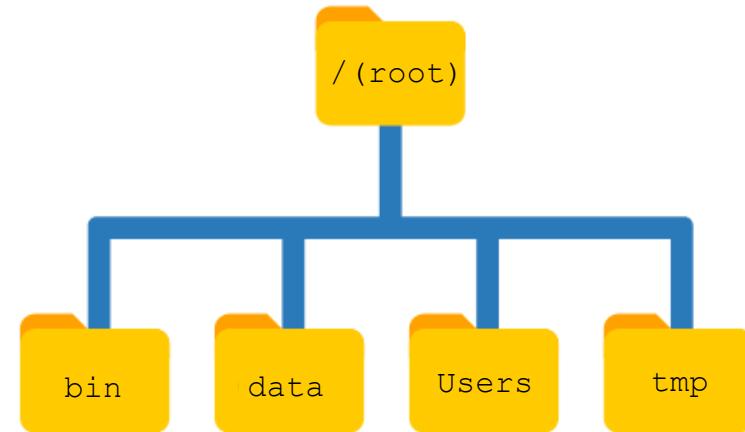
`ssh` (SSH client) is a program for logging into a remote machine and for executing commands on a remote machine. It is intended to provide secure encrypted communications between two untrusted hosts over an insecure network. X11 connections, arbitrary TCP ports and UNIX-domain sockets can also be forwarded over the secure channel.

`ssh` connects and logs into the specified destination, which may be specified as either [user@]hostname or a URI of the form `ssh://[user@]hostname[:port]`.

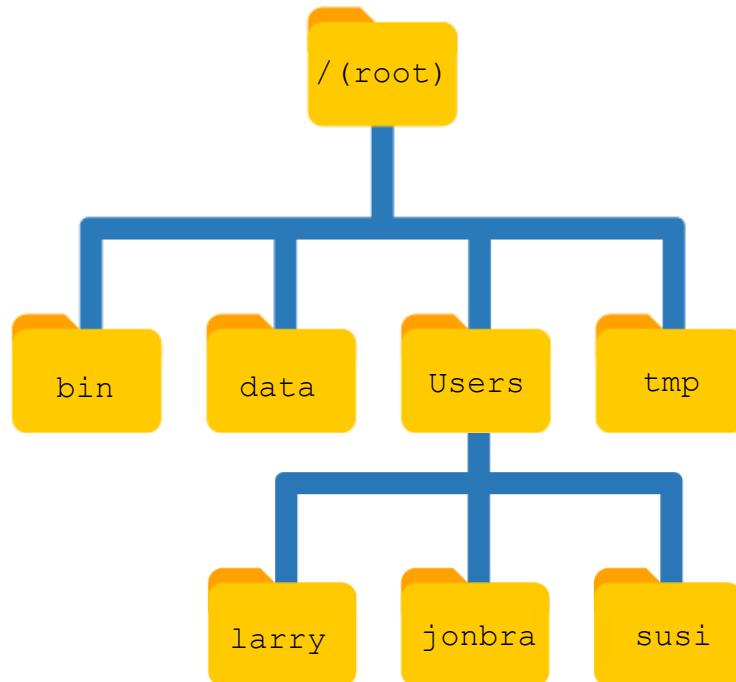
File structure

- Unix has a hierarchical file system.
- At the top is the **root directory** that holds everything else. We refer to it using a slash character / on its own
- Inside that directory are several other directories:
 - *bin* (which is where some built-in programs are stored),
 - *data* (for miscellaneous data files),
 - *Users* (where users' personal directories are located),

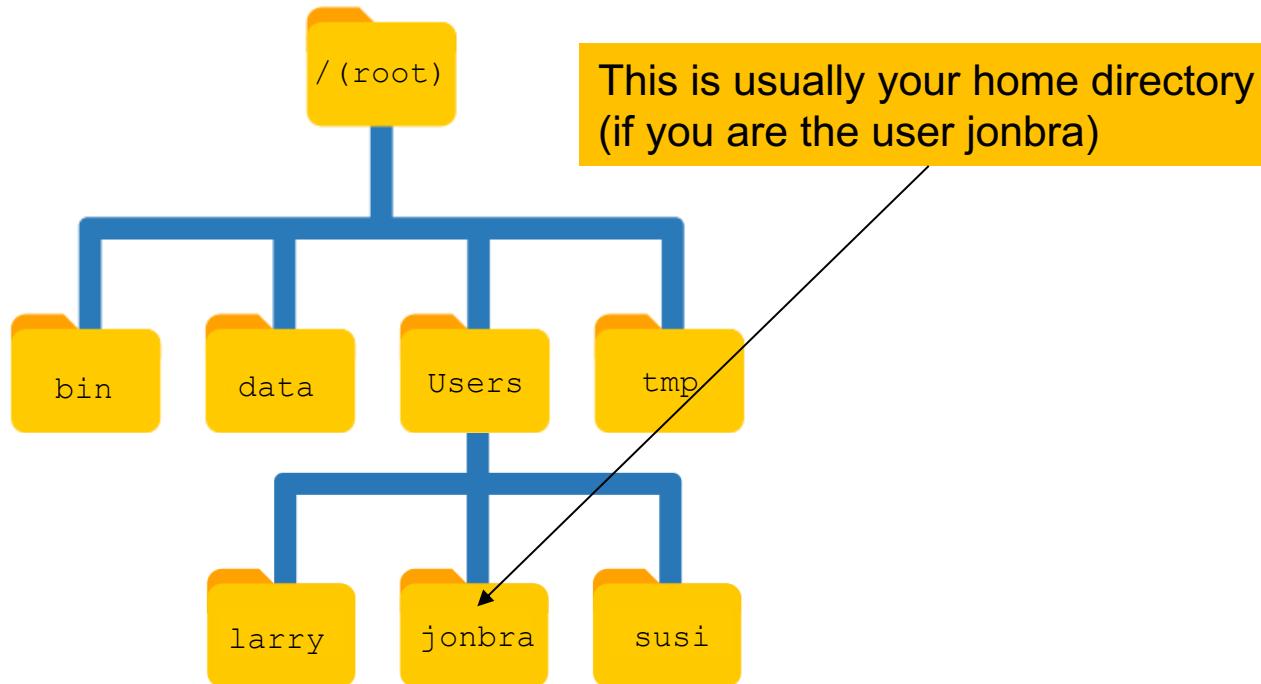
```
1. bash
bio89093:/ jonbra$ ls /
Applications          Volumes
Library               bin
Network               cores
PackageInfo           dev
Payload               etc
Scripts              home
System                installer.failure.requests
Users                 macOS_SDK
bio89093:/ jonbra$
```



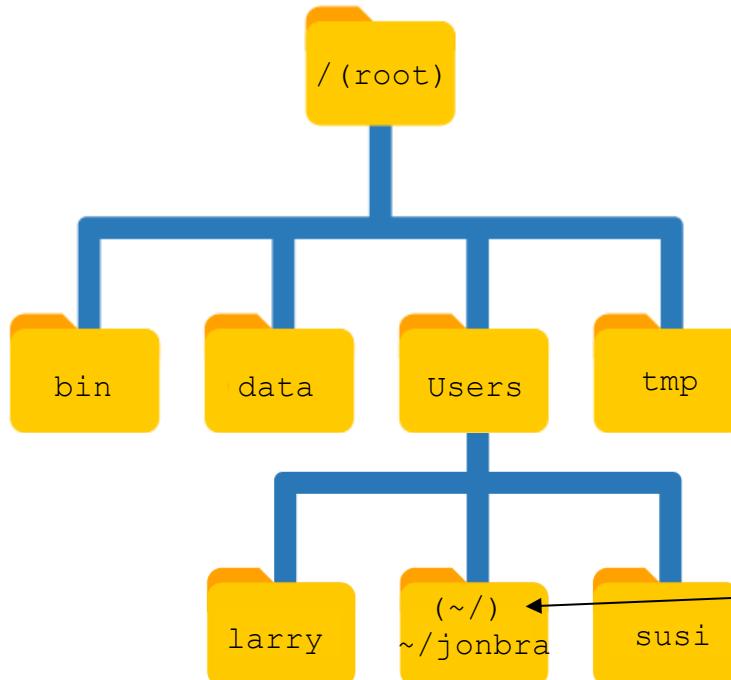
File structure



The home directory path may look different depending on the OS.
E.g. Windows MS-DOS: C:\Documents\jonbra or C:\Users\jonbra
Mac/Linux/gitbash: /Users/jonbra (or something similar)



File structure



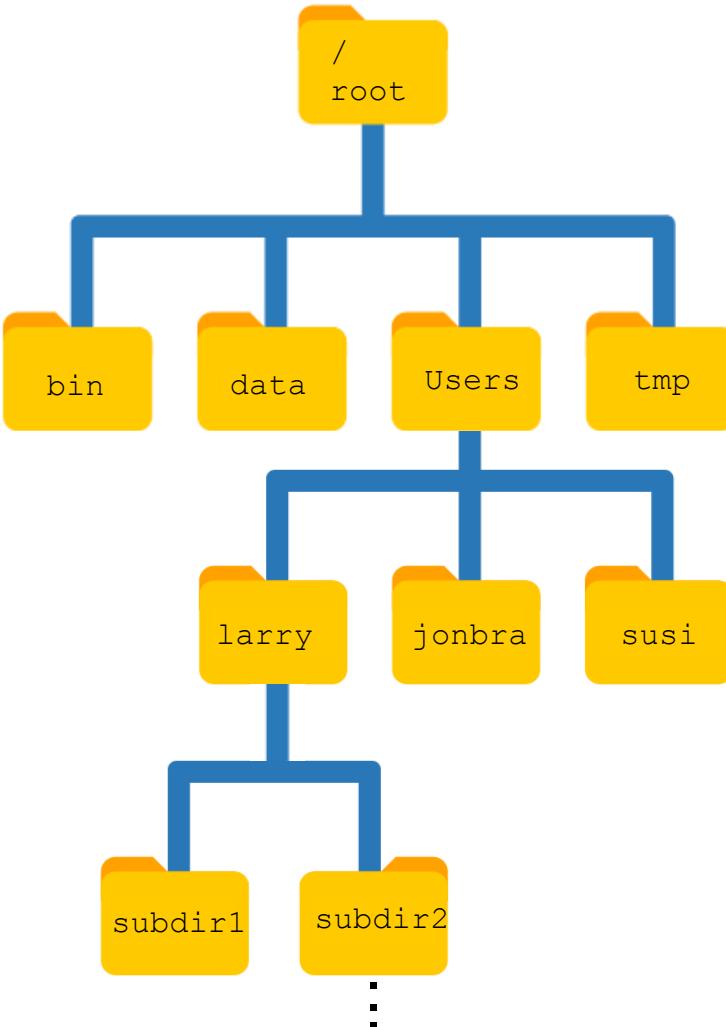
The home directory is also `~/` or `$HOME`

“tilde” character!

If you are logged in as `jonbra`,
`~/jonbra` will also be your
home directory

Navigating the filesystem

- `cd` changes directory
- `ls` lists the files and directories in the current directory
- `pwd` prints the path of the current directory
- `~/` or `$HOME` is your home directory (`cd` will take you there)
- `.` is your current directory
- `..` is the directory directly above the current



- To make a new directory: `mkdir`
 - `mkdir larry`

To make subdirectories

- `mkdir larry/subdir1`
- `mkdir larry/subdir2`
- `mkdir larry/subdir2/subdir1`

To navigate use `cd` (change directory)

- `cd larry`
- `cd ..` (takes you one level up)
- `cd .` (you stay where you are)
- `cd` (shortcut directly to your home directory)

To find out where you are, use `pwd` (print working directory)

To list the files and directories in the current directory use `ls`

```

[jonbra@freebee ~]$ mkdir larry
[jonbra@freebee ~]$ mkdir larry/subdir1
[jonbra@freebee ~]$ mkdir larry/subdir2
[jonbra@freebee ~]$ mkdir larry/subdir2/subdir1
[jonbra@freebee ~]$ cd larry/
[jonbra@freebee larry]$ cd ..
[jonbra@freebee ~]$ cd .
[jonbra@freebee ~]$ cd larry/subdir2/subdir1/
[jonbra@freebee subdir1]$ cd
[jonbra@freebee ~]$ cd larry/subdir1/
[jonbra@freebee subdir1]$ pwd
/usr/abel/u1/jonbra/larry/subdir1
[jonbra@freebee subdir1]$ ls
[jonbra@freebee subdir1]$ cd ..
[jonbra@freebee larry]$ ls
subdir1 subdir2
[jonbra@freebee larry]$
  
```

Transfer files

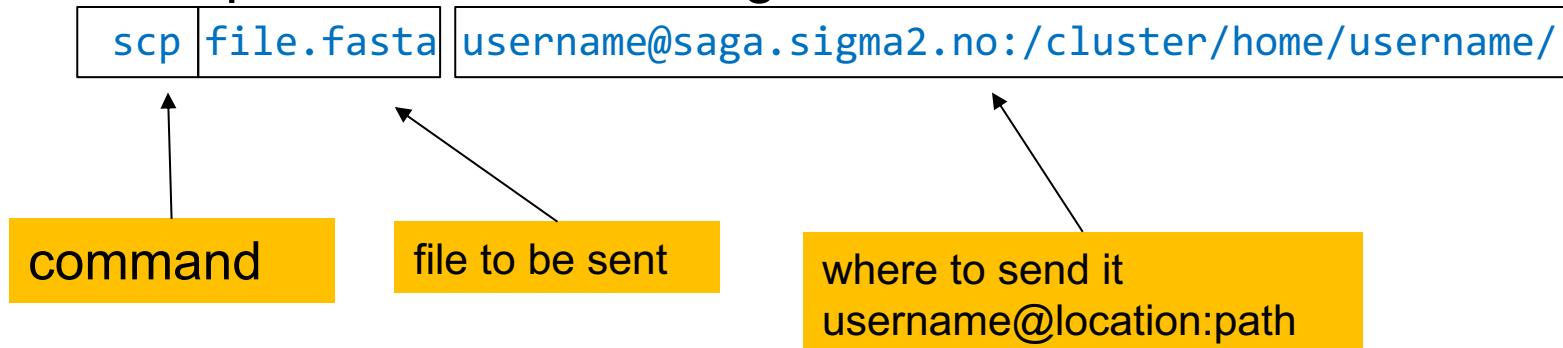
- You will learn to transfer files between Saga and your laptop (both ways) in the exercises
- We will use the command line with *scp*
- You can use software such as
 - WinSCP (for windows)
 - Fetch (Mac)
 - Cyberduck (Mac)

Transfer files with scp

- `scp` is used to copy files from one machine to a remote machine over a secure SSH connection.
- General structure:

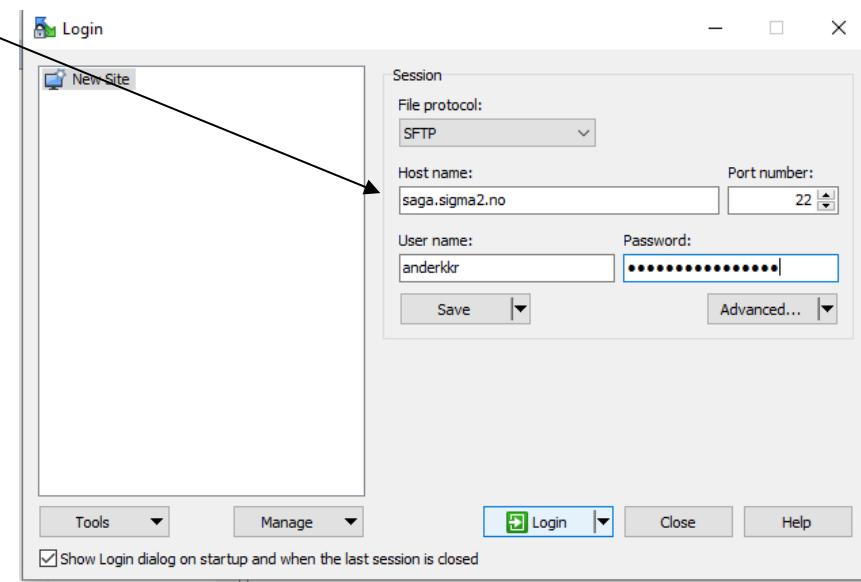
`scp <from> <to>`

- Example a files sent to saga:



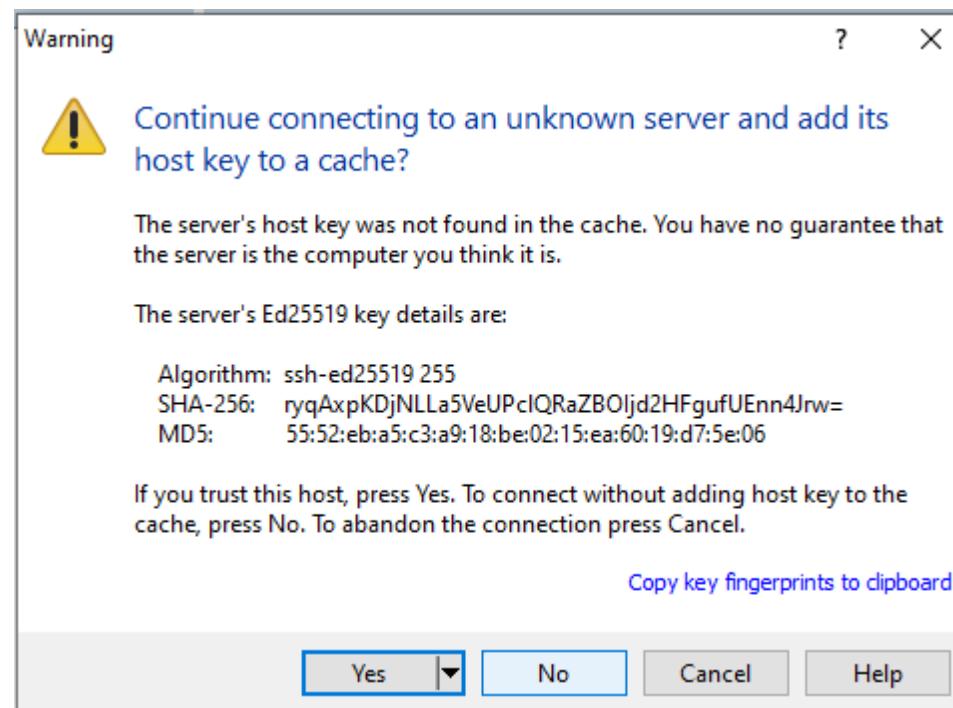
Transfer files WINScp

- Windows users only:
- WinSCP <https://winscp.net/eng/download.php>
- host = saga.sigma2.no



Transfer files WINScp

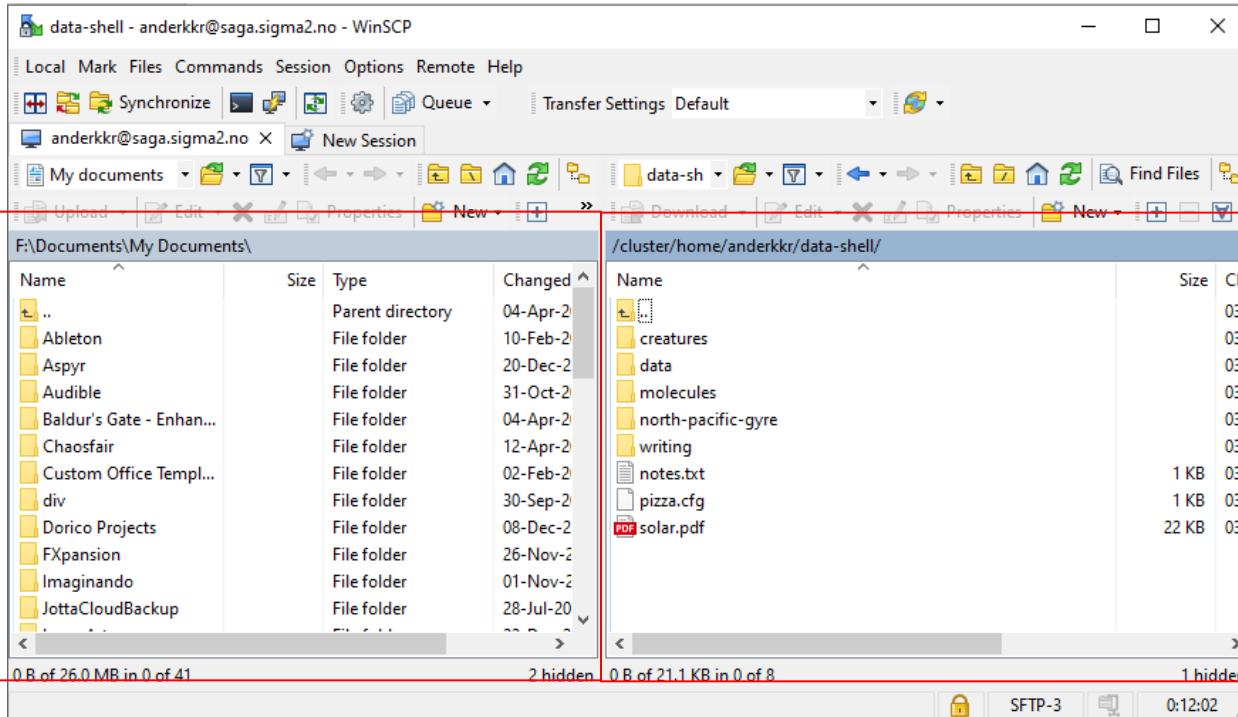
Press “yes”



Transfer files WINScp

Your computer!

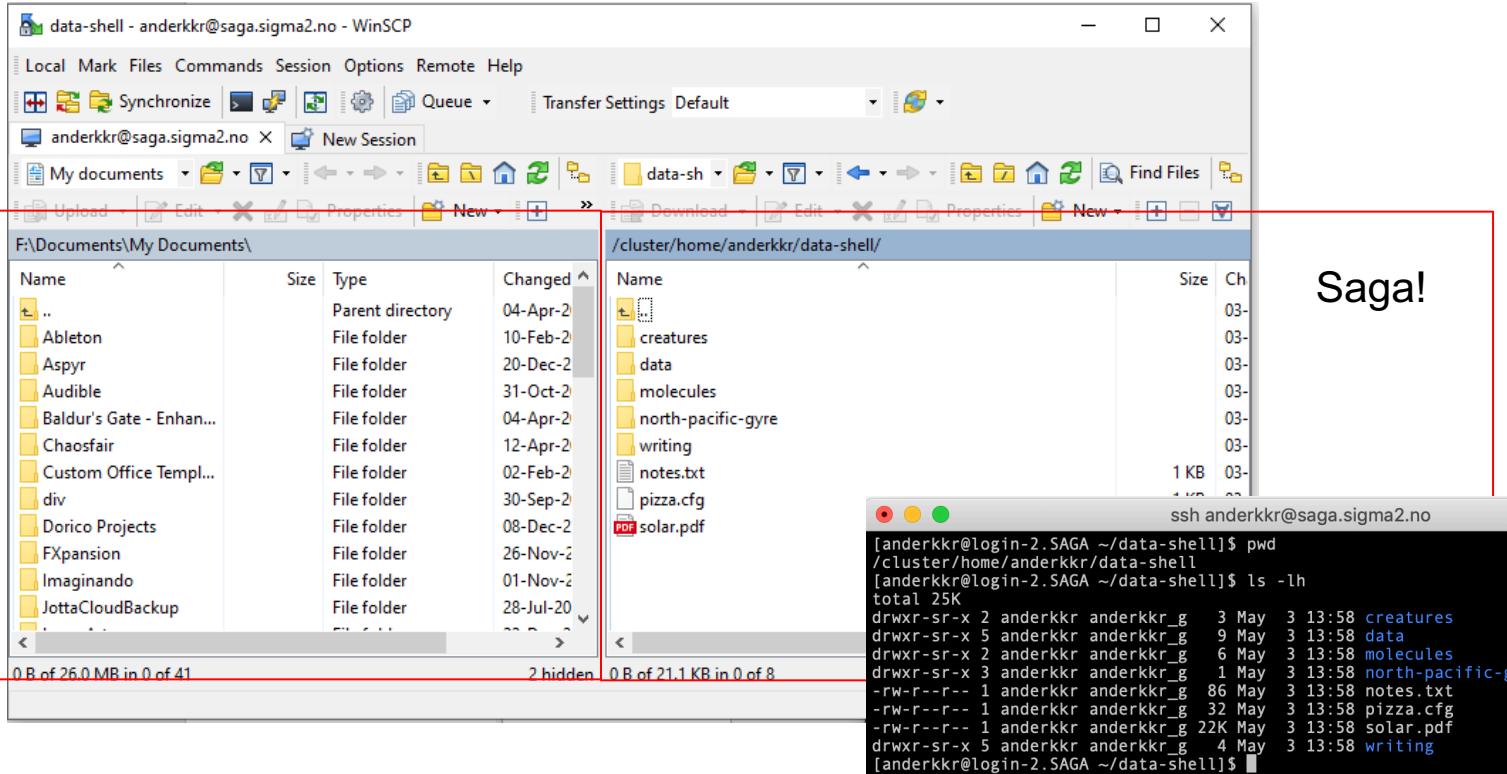
Saga!



Transfer files WINScp

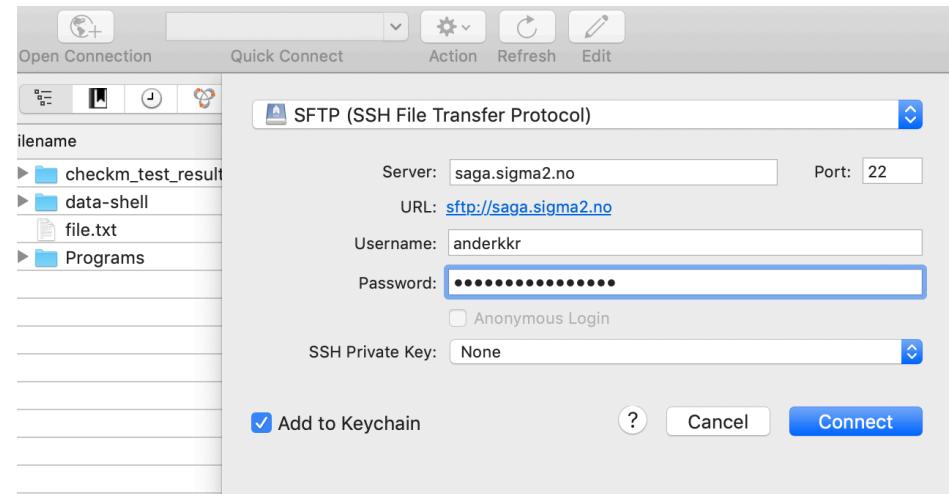
Your computer!

Saga!



Transfer files with Cyberduck (Mac)

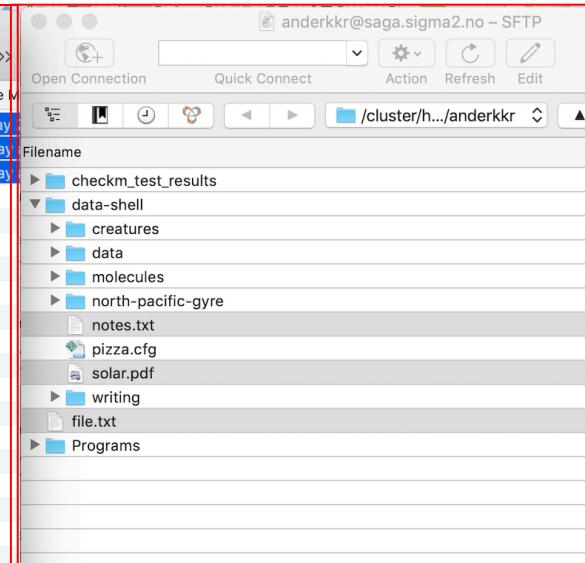
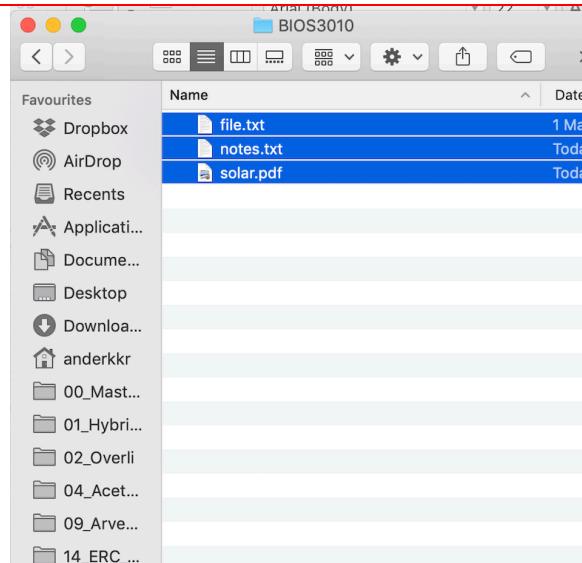
- <https://cyberduck.io/>
- Download, unzip and open Cyberduck
- Click Open Connection
- Choose SFTP (SSH File Transfer Protocol)
- Enter servername
 - saga.sigma2.no
- Enter username and password
- Click connect



Transfer files with Cyberduck (Mac)

- Now you can drag-and-drop files between the Cyberduck window (Saga) and a finder window (your computer)

Local
computer
(Finder
window)



Saga
Cyberduck

Some of the most useful commands which will be covered in the exercises

- `ssh` - Secure log on
 - `ls` - List content
 - `pwd` - Print the working directory
 - `cd` - change directory
 - `mkdir` - make directory
 - `cp` - Copy files and directories
 - `scp` - Secure copying of files and directories
 - `rsync` - Synchronise and transfer files and directories
 - `rm` - Remove files and directories
 - `less` - Print content of file to screen
 - `more` - Print content of file to screen
 - `head` - Print the head of file(s)
 - `tail` - Print the end (tail) of file(s)
 - `cat` - Print and concatenate several files
 - `grep` - Search through text files
 - `sort` - Sort the content of files
 - `uniq` - Print unique (lines, or words)
- Special operators:**
- `>` (IO redirection)
 - `>>` (IO append)
 - `|` (the vertical line is called *pipe*)
 - Useful cheat sheet:
<https://cheatography.com/davechild/cheat-sheets/linux-command-line/>

Remember

- Linux commands are case sensitive!
 - `ssh`
 - is **NOT** the same as
 - `Ssh`
- Filenames and paths are also case sensitive!
 - `huge_file.fasta`
 - is **NOT** the same as:
 - `Huge_file.fasta`
- Protip: Use the tab-key to autocomplete

Moving files with the `mv` command

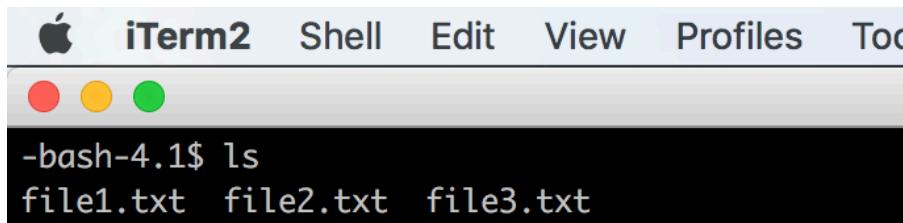
The screenshot shows a terminal window titled "iTerm2" with a session titled "2. jonbra@ulrik:~/MBV-INFX410 (ssh)". The terminal displays a file listing from the current directory (~) and then demonstrates the use of the `mv` command to move a file.

```
-bash-4.1$ ls -lrt
total 204403
drwxr-xr-x 5 jonbra users      7 Jul 14 21:32 sratoolkit.2.7.0-ubuntu64
-rw-r--r-- 1 jonbra users 63707890 Jul 14 21:33 sratoolkit.tar.gz
-rw-r--r-- 1 jonbra users     297 Oct 11 13:41 getRawData.sh
-rw-r--r-- 1 jonbra users     546 Oct 11 13:51 sra.slurm
-rw-r--r-- 1 jonbra users    730 Oct 11 14:03 slurm-15702476.base
-rw-r--r-- 1 jonbra bio   72051869 Oct 14 23:39 gbbct100.seq.gz.1
-rw-r--r-- 1 jonbra bio   72051869 Oct 14 23:39 gbbct100.seq.gz
-rw-r--r-- 1 jonbra users 1496824 Nov  7 10:00 genes.fpkm_tracking
drwxr-xr-x 2 jonbra bio       1 Nov 15 09:36 Ensembl-demo
drwxr-xr-x 2 jonbra bio       0 Nov 15 15:15 mvtest
-rw-r--r-- 1 jonbra bio       0 Nov 15 15:15 file1.txt
-bash-4.1$ mv file1.txt mvtest/
-bash-4.1$ cd mvtest/
-bash-4.1$ ls
file1.txt
-bash-4.1$ mv file1.txt file2.txt
-bash-4.1$ ls
file2.txt
-bash-4.1$
```

A green callout box contains the text "Move file1.txt to directory mvtest". A green arrow points from this text to the `mv` command in the terminal output where it moves the file to the `mvtest` directory.

A second green callout box contains the text "Pro tip:
`mv` command can also rename files!
("moves" file to a new filename)". An arrow points from this text to the second `mv` command in the terminal output, which renames the file `file1.txt` to `file2.txt`.

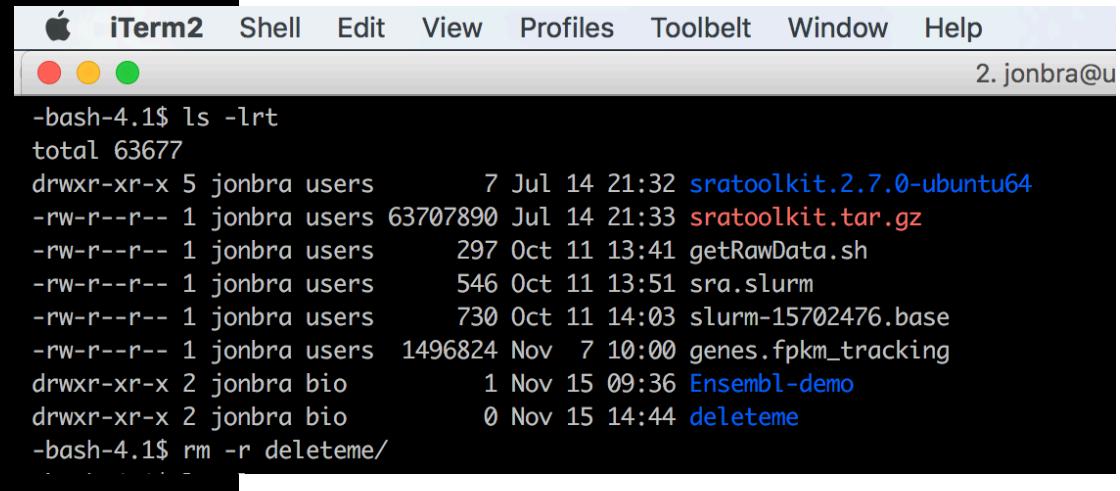
Delete files with the `rm` command



```
-bash-4.1$ ls
file1.txt file2.txt file3.txt
-bash-4.1$ rm file1.txt
-bash-4.1$ ls
file2.txt file3.txt
-bash-4.1$ rm file*
-bash-4.1$ ls
-bash-4.1$
```

Delete empty directories with `rm -r`

Delete directories AND
files within with `rm -rf` Use with
caution!! (`rm -i`)



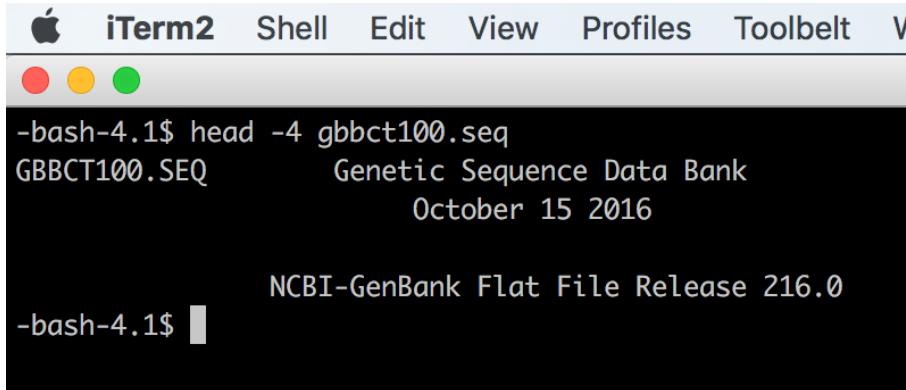
```
-bash-4.1$ ls -lrt
total 63677
drwxr-xr-x 5 jonbra users      7 Jul 14 21:32 sratoolkit.2.7.0-ubuntu64
-rw-r--r-- 1 jonbra users 63707890 Jul 14 21:33 sratoolkit.tar.gz
-rw-r--r-- 1 jonbra users     297 Oct 11 13:41 getRawData.sh
-rw-r--r-- 1 jonbra users      546 Oct 11 13:51 sra.slurm
-rw-r--r-- 1 jonbra users    730 Oct 11 14:03 slurm-15702476.base
-rw-r--r-- 1 jonbra users 1496824 Nov  7 10:00 genes.fpkm_tracking
drwxr-xr-x 2 jonbra bio        1 Nov 15 09:36 Ensembl-demo
drwxr-xr-x 2 jonbra bio        0 Nov 15 14:44 deleteme/
-bash-4.1$ rm -r deleteme/
```

NB! There's no trash can on Unix!
Your files are immediately deleted!

Viewing files

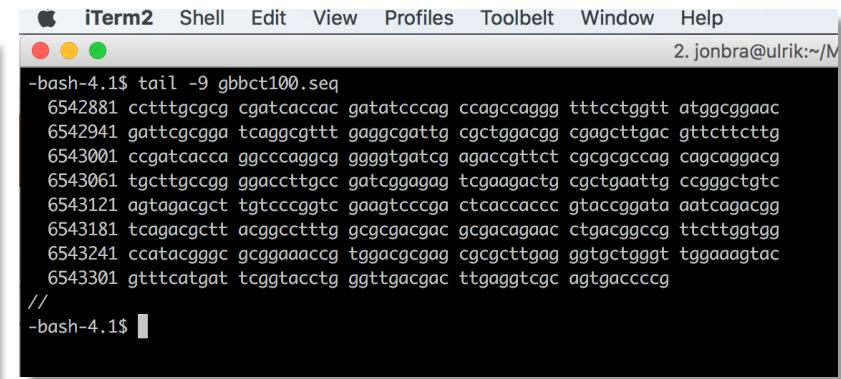
Pro tip:
Use the different
options to tweak the output

- `cat` displays the entire content on the screen
- `less` displays the content on the screen one file/line at a time (press "q" to exit).
- `head/tail` displays the first or last lines of a file



```
-bash-4.1$ head -4 gbbct100.seq
GBBCT100.SEQ      Genetic Sequence Data Bank
                  October 15 2016

NCBI-GenBank Flat File Release 216.0
-bash-4.1$
```



```
iTerm2  Shell  Edit  View  Profiles  Toolkit  Window  Help
2. jonbra@ulrik:~/M
-bash-4.1$ tail -9 gbbct100.seq
6542881 cttttgcgca cgatcaccac gatatccag ccagccaggg tttctgggtt atggcgaaac
6542941 gattcggga tcaggggtt gagggattt cgctggacgg cgagcttgcac gttttcttg
6543001 ccgatcacca ggccccggcg ggggtgtatcg agaccgttct cgccgcggcag cagcaggac
6543061 tgcttgcggg ggaccttgcc gatcggagag tcgaagactg cgctgaattt ccgggtgtc
6543121 agtagacgct tgtcccggtc gaagtcccga ctcaccaccc gtaccggata aatcagacgg
6543181 tcagacgctt acggcctttgc gcgcgacgac ggcgcacaaat ctgcggccg ttcttggtag
6543241 ccatacgggc gcggaaacgg tggacgcgag cgccgttgag ggtgtgggt tgaaaatgtac
6543301 gtttcatgtat tcggatctgt gtttgcgac ttgagggtcgc agtgcaccccg
//-
-bash-4.1$
```

Downloading files from the internet

A screenshot of a Firefox browser window. The address bar shows 'https://ftp.ncbi.nih.gov/genbank'. The main content area displays a file list titled 'Index of /genbank'. The table has columns for 'Name', 'Last modified', and 'Size'. The 'Name' column contains links to various sub-directories like 'catalog/', 'daily-nc/', 'docs/', etc. The 'Last modified' column shows dates ranging from 14-Oct-2016 to 05-Dec-2013. The 'Size' column shows sizes like '28K', '0M', '6M', '9M', '67M', '68M', and '69M'. A green box highlights the 'catalog/' link.

Name	Last modified	Size
catalog/	14-Oct-2016 14:43	-
daily-nc/	14-Nov-2016 01:33	-
docs/	05-Dec-2013 12:39	-
genomes/	02-Dec-2015 16:58	-
livelists/	14-Nov-2016 00:31	-
quality_scores/	14-Oct-2016 17:41	-
release.notes/	14-Oct-2016 17:41	-
tls/	11-Nov-2016 09:25	-
tools/	01-Jul-2013 11:20	-
tsa/	13-Nov-2016 05:51	-
wgs/	13-Nov-2016 05:51	-
GB_Release_Number	14-Oct-2016 17:39	4
README_genbank	14-Oct-2016 17:40	28K
		0M
		6M
		9M
gbbct101.seq.gz	14-Oct-2016 17:39	67M
gbbct102.seq.gz	14-Oct-2016 17:39	68M
gbbct103.seq.gz	14-Oct-2016 17:39	69M

Right-click and copy the link!

A screenshot of an iTerm2 terminal window. The title bar says 'iTerm2' and the session name is '2. jonbra@ulrik:~/MBV-INFX410 (ssh)'. The terminal output shows the execution of the 'wget' command to download 'gbbct100.seq.gz' from 'https://ftp.ncbi.nih.gov/genbank/'. The progress bar indicates a download speed of 19.8M/s over 5.8s. The final message shows the file was saved successfully at [72051869/72051869].

```
-bash-4.1$ wget https://ftp.ncbi.nih.gov/genbank/gbbct100.seq.gz
--2016-11-15 14:59:18-- https://ftp.ncbi.nih.gov/genbank/gbbct100.seq.gz
Resolving ftp.ncbi.nih.gov... 130.14.250.13, 2607:f220:41e:250::7
Connecting to ftp.ncbi.nih.gov|130.14.250.13|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 72051869 (69M) [application/x-gzip]
Saving to: `gbbct100.seq.gz'

100%[=====] 72,051,869 19.8M/s in 5.8s

2016-11-15 14:59:25 (11.8 MB/s) - `gbbct100.seq.gz' saved [72051869/72051869]

-bash-4.1$
```

Redirect and pipe

- Since Linux (and UNIX) use streams of text it is easy to
- write the output of a command to a file.
 - Two versions:
 - The **greater than symbol** (>) is used to write a new files (will overwrite)
`ls -lh > list_of_files.txt`
 - Two symbols (>>) will append the content to a file
`ls -lh >> list_of_files.txt`

Redirect and pipe

- Since Linux (and UNIX) use streams of text it is easy to
- Use the output of one command as input in another with the horizontal line called pipe |

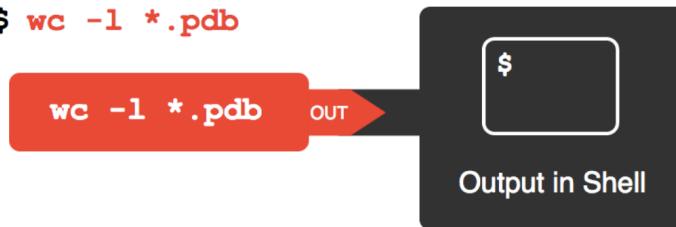
```
wc -l *fasta | sort -n
```

First command

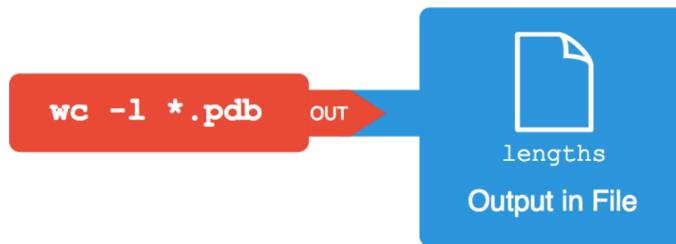
Pipe

Second command

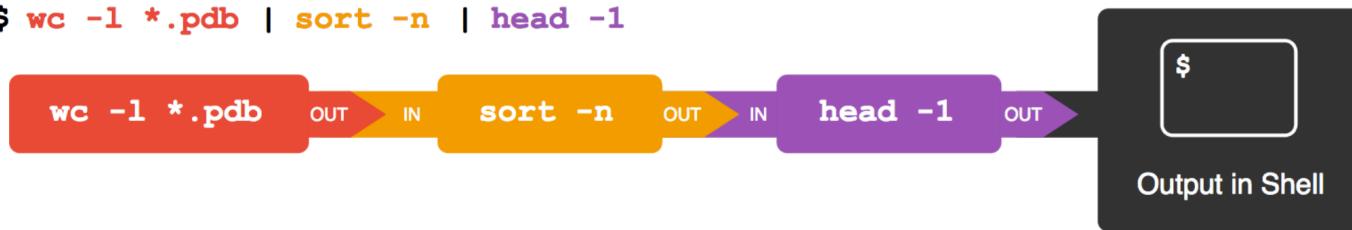
```
$ wc -l *.pdb
```



```
$ wc -l *.pdb > lengths
```



```
$ wc -l *.pdb | sort -n | head -1
```



Piping commands together

Several commands in a row:

```
wc -l * | sort -n | head -n 3 > results.txt
```

More on this in Exercise 3.

<https://github.com/krabberod/bios3010-unix-linux>

Some online resources

- Linux Cheat Sheet
 - <https://www.cheatography.com/davechild/cheat-sheets/linux-command-line/>
 - <https://files.fosswire.com/2007/08/fwunixref.pdf>
- Vim Cheat Sheet
 - <https://www.cheatography.com/nuggles/cheat-sheets/vim/>
- Nano Cheat Sheet
 - <https://www.cheatography.com/bipinthite/cheat-sheets/nano-editor/>

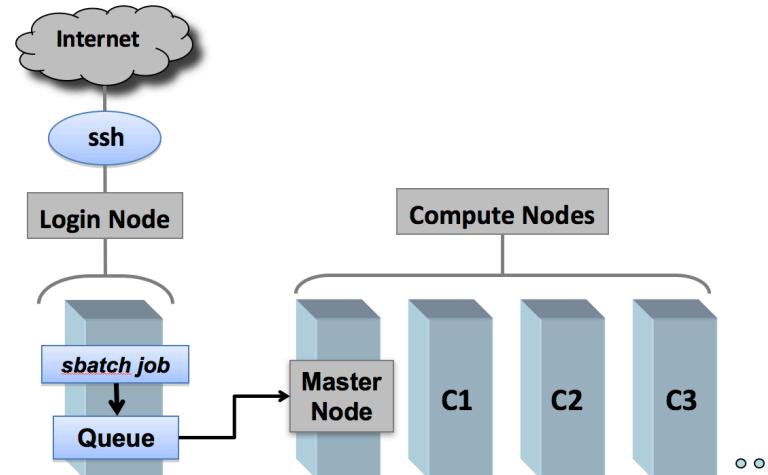
Preinstalled programs on Saga

- Saga has already many preinstalled programs.
- Check available modules with
 - module avail
- These can be loaded by the command
 - module load <program>

```
[anderkkr@login-1] ~ $ module avail
----- /cluster/modulefiles/all
ABYSS/2.0.2-gompi-2019a
AdapterRemoval/2.3.1-foss-2018b
AdapterRemoval/2.3.1-GCC-8.2.0-2.31.1
ADF/2019.103-StaticMKL
AdmixTools/5.1-GCC-7.3.0-2.30
ADMIXTURE/1.3.0
Amber/18-foss-2018b-AmberTools-18-patchlevel-10-8
Amber/18-fosscuda-2018b-AmberTools-18-patchlevel-10-8
AMOS/3.1.0-foss-2018b
Anaconda2/2019.03
Anaconda3/2019.03
Ancestry_HMM/200105-foss-2019a
angsd-wrapper/190321-GCC-7.3.0-2.30
angsd/0.925-foss-2018b
angsd/0.928-GCC-7.3.0-2.30
angsd/0.931-GCC-8.2.0-2.31.1
ARKS/1.0.4-gompi-2019a
Arm-PerfReports/19.1
Armadillo/9.700_2-foss-2019a
arpack-ng/3.7.0-foss-2019a
asd/200218-GCC-8.3.0
ASE/3.17.0-foss-2018b-Python-3.6.6
ASE/3.17.0-intel-2018b-Python-3.6.6
ASE/3.18.0-foss-2019a-Python-3.7.2
ASE/3.18.0-intel-2019a-Python-3.7.2
ASE/3.19.0-foss-2019b-Python-3.7.4
Aspera-CLI/3.9.0.1326.6985b21
AUGUSTUS/3.3.2-intel-2018b-Python-2.7.15
BAMM/2.5.0-foss-2018b
BamTools/2.5.1-foss-2018b
BamTools/2.5.1-GCC-8.2.0-2.31.1
BamTools/2.5.1-iccfourt-2019.1.144-GCC-8.2.0-2.31.1
BamTools/2.5.1-intel-2018b
BayeScan/2.1-foss-2018b
BayeScan/2.1-intel-2018b
BayeScEnv/1.1-foss-2018b
Bazam/1.0.1-Java-1.8
BBMap/38.50b-GCC-8.2.0-2.31.1
BCFtools/1.9-foss-2018b
BCFtools/1.9-intel-2018b
beagle-lib/3.0.2-foss-2018b
beagle-lib/3.1.2-GCC-8.2.0-2.31.1
Beagle/4.1-27Jan18-Java-1.8
Beagle/5.1-24Mar20-Java-11
Beagle/5.1-25Nov19-Java-11
Beast/2.5.2-GCC-8.2.0-2.31.1
BEDOPS/2.4.35-foss-2018b
BEDTools/2.27.1-foss-2018b
BEDTools/2.27.1-intel-2018b
BEDTools/2.28.0-GCC-8.2.0-2.31.1
BEDTools/2.28.0-iccfourt-2019.1.144-GCC-8.2.0-2.31.1
```

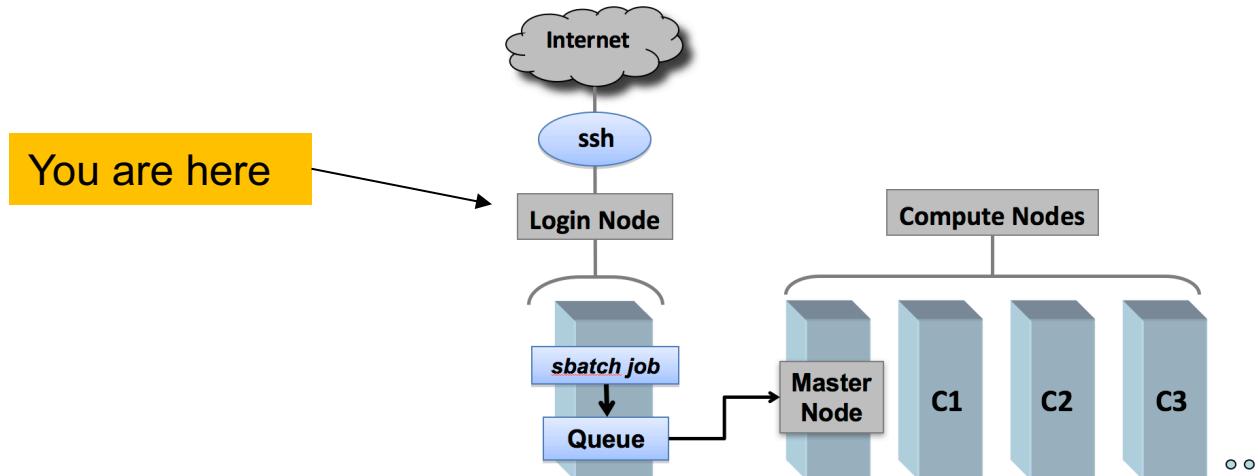
Queue system - Slurm

- Since Saga is used by hundreds of people it needs a queueing system to divide the resources
- The queue system is called Slurm.



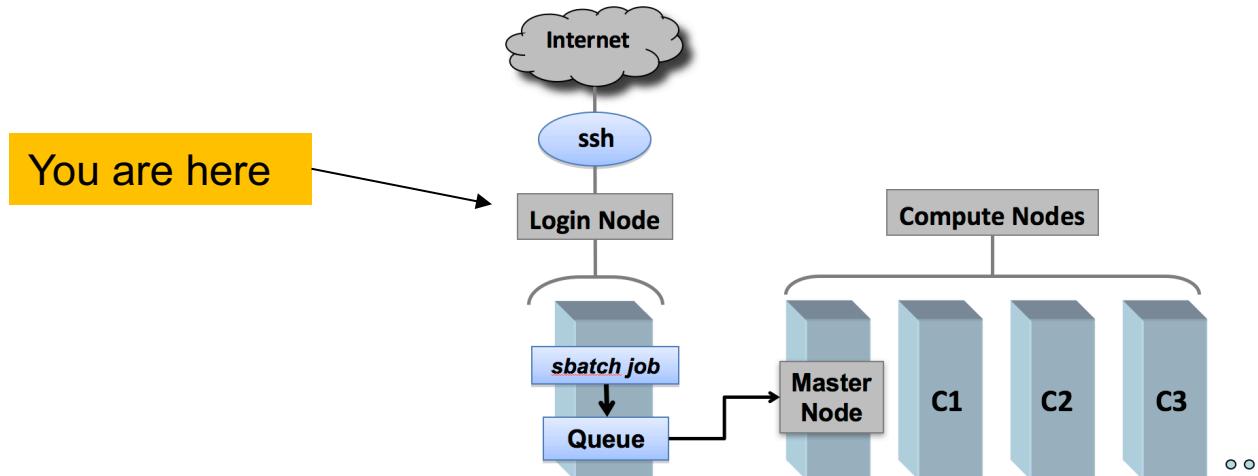
Queue system - Slurm

- Since Saga is used by hundreds of people it needs a queueing system to divide the resources
- The queue system is called Slurm.



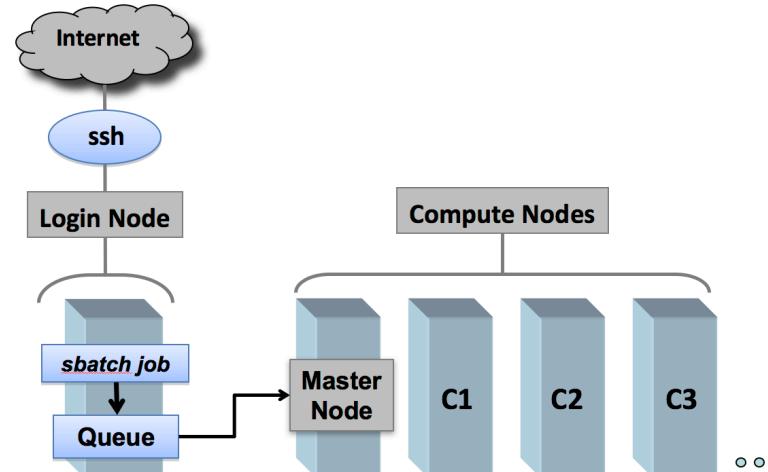
Queue system - Slurm

- Since Saga is used by hundreds of people it needs a queueing system to divide the resources
- The queue system is called Slurm.



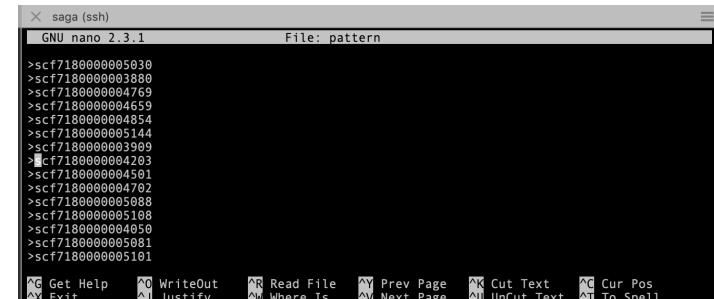
Queue system - Slurm

- You will use the queue system later in the course, when you will do genome assembly.



Text editors in Linux

- Unix has many built-in ways editing text files.
- We will be using *nano* in this course and it will be introduced in the exercises.
- It is a very simple texteditor with simple, basic functionality.
- Start:
`nano filename`
- CTRL-x – Exits the editor. If you are in the middle of editing a file the exit process will ask you if you want to save your work.
- CTRL-o – Save file name and continue working.
- CTRL-w – Search your text.
- CTRL-g – Get help with Nano.



```
>scf7180000005030
>scf7180000003880
>scf7180000004769
>scf7180000004659
>scf7180000004854
>scf7180000005144
>scf7180000003909
>scf7180000004203
>scf7180000004501
>scf7180000004702
>scf7180000005088
>scf7180000005108
>scf7180000004050
>scf7180000005081
>scf7180000005101
```

File: pattern

Get Help Exit WriteOut Read File Prev Page Cut Text Cur Pos
Where Is Next Page UnCut Text To Spell

- Overwhelmed?
- Keep calm and practice the simple stuff.
 - Test on real cases
- Do the exercises:
- <https://github.com/krabberod/bios3010-unix-linux>