

Abims

Linux for Beginners

ABiMS

Training Module 2024

Mark HOEBEKE

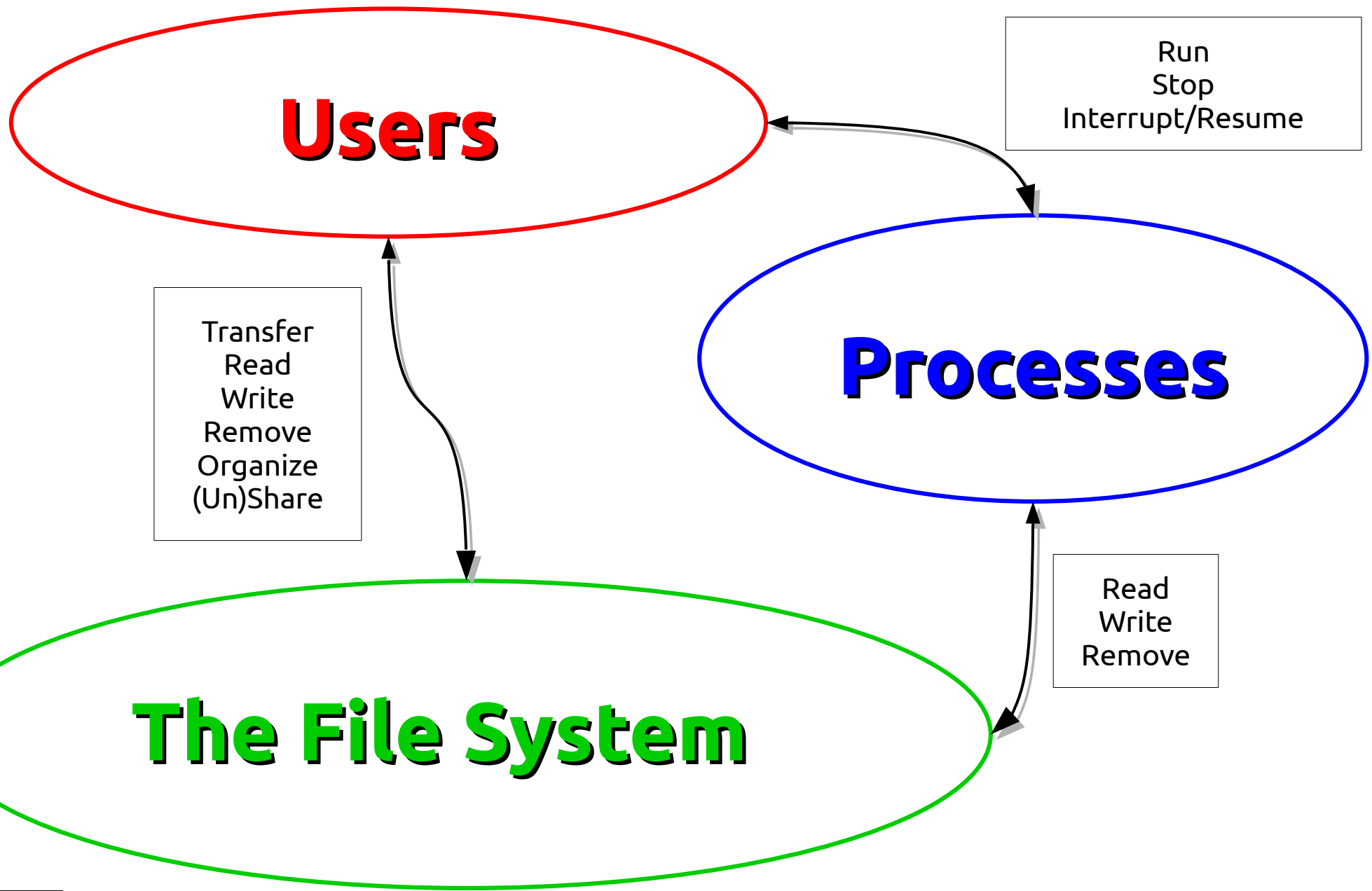
Jean-Michel AROUMOUGOM



My {colleague|competitor} told me about a highly interesting tool for analysing data quite related to mine. It's some kind of thingy running on *Lynuks* and installed somewhere on a server.

How do I:

- Connect (**myself**) to the server?
- Transfer **my files** holding the data on the server?
- Launch the **program**?
- Specify where to find the **files** with the data it will process?
- Specify where to write the **files** with the results it will generate?
- Organize all these **files** in **folders** not to get lost later on ?
- Run several **programs** "simultaneously" (or several "instances" of the same **program**) with different parameters or input data?
- Stop an instance of the **program** if it runs for too long?
- {Share|protect} **my files** containing input data or results {with my colleagues|from my competitors}?

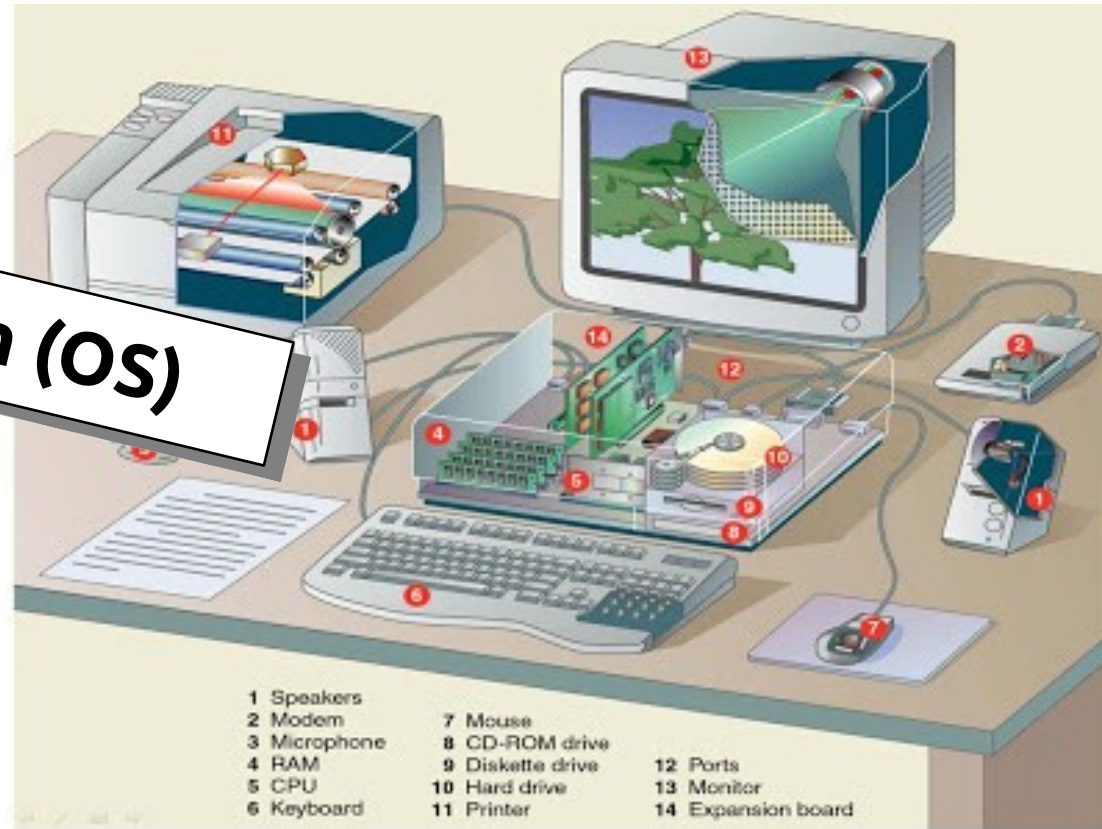


- 1 Purpose of an Operating System – why Linux ?**
- 2 Establishing a connection and transferring files**
- 3 The Command Line Interface**
- 4 The File System**
- 5 Manipulating File Contents**
- 6 Users, Groups and Access Control**
- 7 Processes**

- 1 Purpose of an Operating System – why Linux ?**
- 2 Establishing a connection and transferring files
- 3 The Command Line Interface
- 4 The File System
- 5 Manipulating File Contents
- 6 Users, Groups and Access Control
- 7 Processes

Purpose of an Operating System

The Operating System (OS)



- An OS is a “privileged” program run when the machine is switched on and that:
- Loads other programs in memory (RAM),
 - Allocates the resources (memory, CPU time, disk space) they request,
 - Handles their communications (input/output) with peripheral devices (screen, keyboard, mouse, network, printer...)
 - Halts their execution
 - Reclaims the allocated resources.

- Preemptive multitasking & multi-user system
 - Durability & stability with a track record back to 1994
- **Open-source and free (as in beer)**
 - Its code can be freely copied, modified and redistributed
- Offering a vast software catalogue :
 - Office suites : LibreOffice
 - Internet tools : browsers (Firefox, Chrome), e-mail programs (Thunderbird, Evolution)
 - Multimedia : audio/video playback tools (VLC, Totem)
 - Graphics : image manipulation (Gimp), 3D modeling (Blender)
 - Software development : languages (Python, Java, C/C++...), environments (Eclipse, IDLE, PyDev, DDD)
- Scientific disciplines included :
 - Bioinformatics : blast, emboss, phylip, mafft, clustal, trimal...

Linux Distributions

A Linux distribution includes:

- Some flavour of the Linux kernel.
- A portfolio of prepackaged software
- Administration tools facilitating the installation and update of these *packages*.



The main differences between distributions are:

- On a technical level :
 - The *packaging* format used to package(!) software.
 - The tools to administer these *packages*.
- Business model wise :
 - Technical Support : community-based vs. commercial.
 - The licenses for the software they provide.



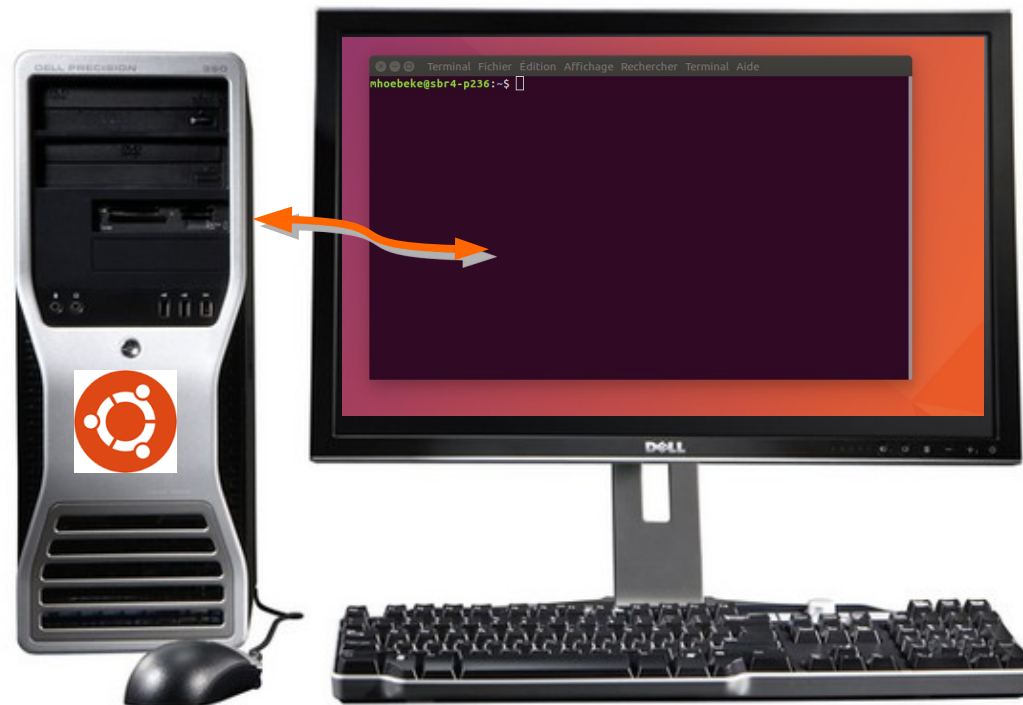
- 1 Purpose of an Operating System – why Linux ?
- 2 Establishing a connection and transferring files**
- 3 The Command Line Interface
- 4 The File System
- 5 Manipulating File Contents
- 6 Users, Groups and Access Control
- 7 Processes

The Terminal :

- A means to "communicate" with **a machine** relying on a **command line** in the context of a **session**

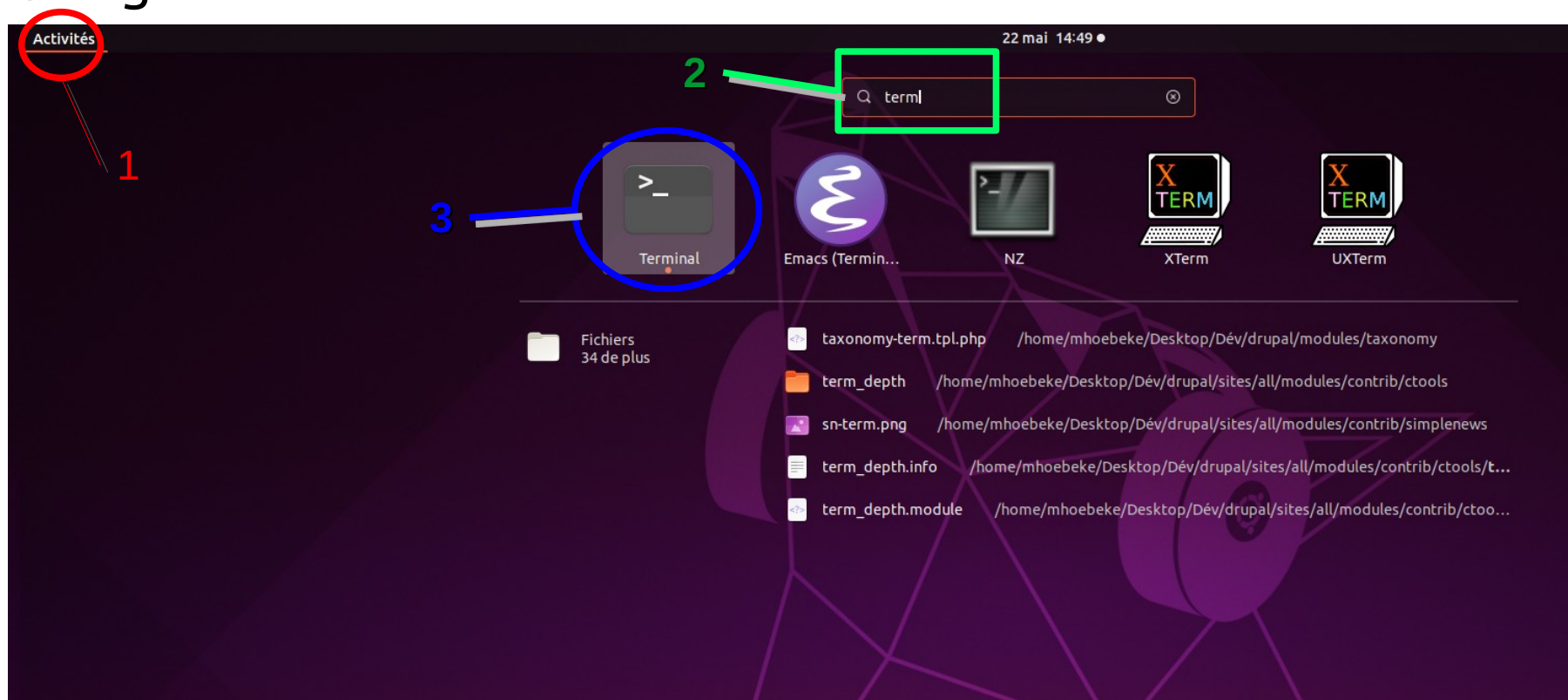
Use Case 1 : Local Sessions

- Type commands (programs) that will run (use memory, CPU and disk space) on **your workstation**



Opening a local session (on a workstation running Ubuntu) :

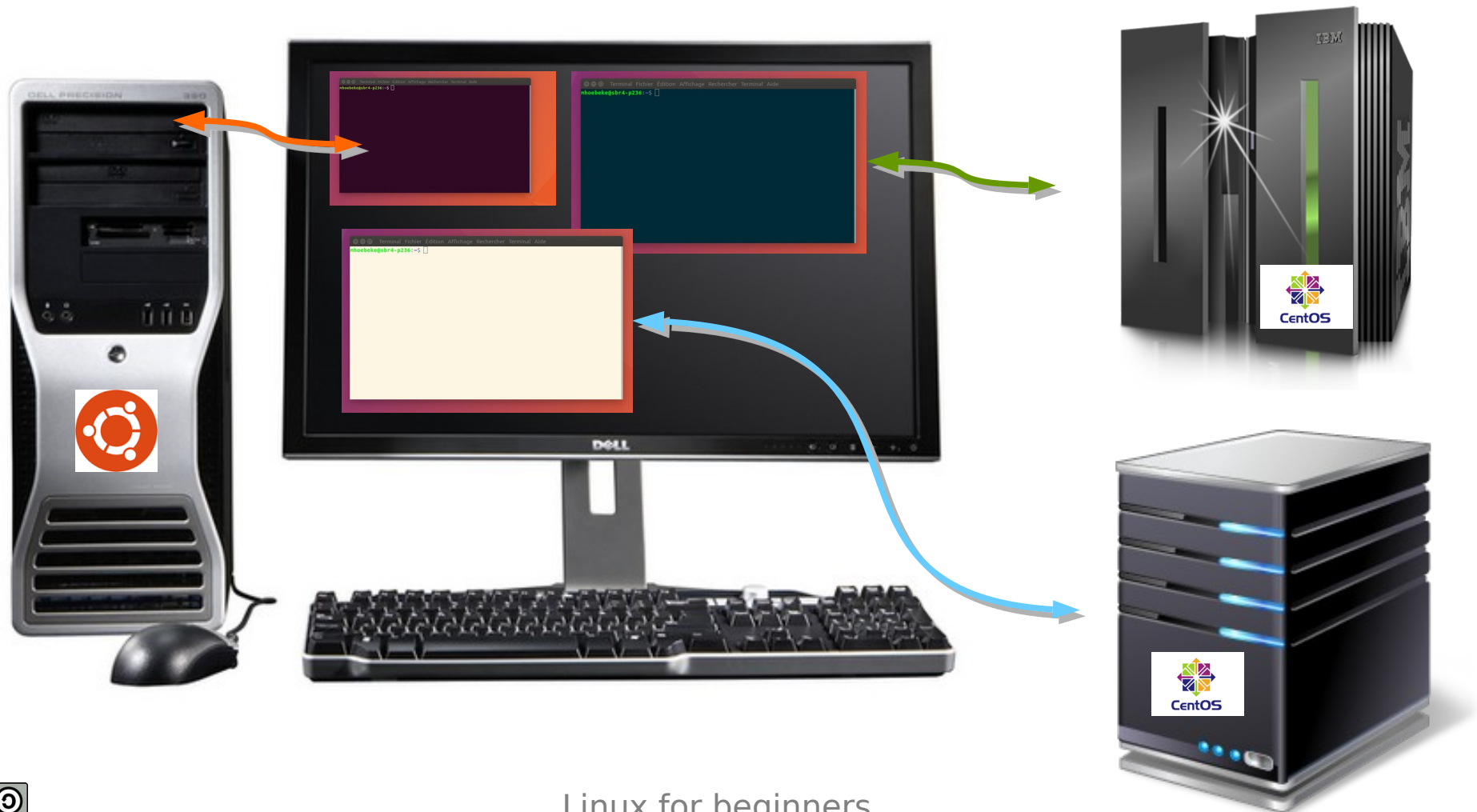
- Using the keyboard : **Ctrl+Alt+T**
- Using the mouse:



Connecting | Remote Sessions

Remote Sessions :

- To work on a machine accessible through the network
- Potentially opening additional windows on the remote machine



Linux for beginners

Using the Secure (SSH) Shell Protocol:

- Encrypted (secure) communications protocol opening a channel allowing information exchange between two machines.
- Requires knowledge of:
 - The name (or address) of the remote machine,
 - The user name (login) on the remote machine.

Using SSH from a local terminal

```
mhoebeke@n221:/projet/fr2424/sib/mhoebeke
File Edit View Search Terminal Help
[mhoebeke@n221 mohoebeke] ssh -Y mohoebeke@slurm0.sb-roscoff.fr
The authenticity of host 'slurm0.sb-roscoff.fr (192.168.1.47)' can't be established.
RSA key fingerprint is 9e:62:b2:95:ae:fe:15:96:fc:75:c7:87:c4:83:a0:d1.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'slurm0.sb-roscoff.fr' (RSA) to the list of known hosts.

4
ABIMS
Analysis and Bioinformatics for Marine Science
http://abims.sb-roscoff.fr - support: abims@sb-roscoff.fr

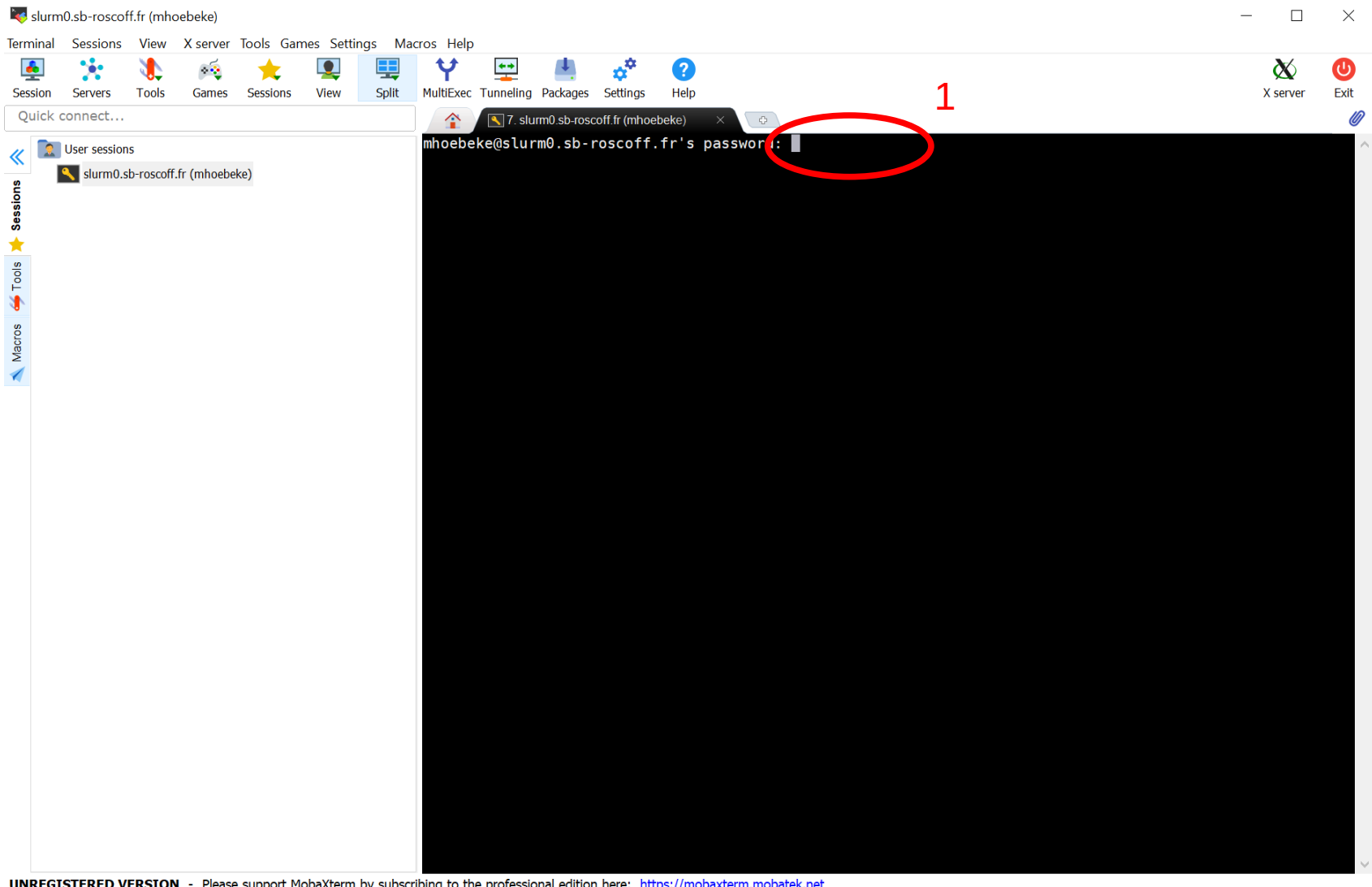
Command 'keychain' not found, but can be installed with:
apt install keychain
Please ask your administrator.

mhoebeke@slurm0:~$
```

Using SSH on a Windows workstation with MobaXTerm

The screenshot shows the MobaXTerm application interface. A red circle labeled '1' highlights the 'Sessions' icon in the top-left toolbar. A dialog box titled 'Session settings' is open, with a red circle labeled '2' around the 'SSH' icon in its top toolbar. The 'Basic SSH settings' tab is active, showing 'Remote host' as 'lurm0.sb-roscoff.fr' (circled with a red circle labeled '3') and 'Specify username' as 'mhoebeke' (circled with a red circle labeled '4'). The 'Port' is set to '22'. At the bottom of the dialog, the 'OK' button is circled with a red circle labeled '5', and the 'Cancel' button is visible to its right. The main window background shows a sidebar with 'Sessions', 'Tools', and 'Macros' sections, and a main area with a 'Quick connect...' field and a 'User sessions' list.

Using SSH on a Windows workstation with MobaXTerm



UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Using SSH on a Windows workstation with MobaXTerm

slurm0.sb-roscoff.fr (mhoebeke)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

/home/fr2424/sib/mhoebeke/

| Name | Size (KB) | Last modify |
|-----------------|-----------|-------------|
| .. | | |
| .agisoft | | 2020-09-25 |
| .cache | | 2020-01-06 |
| .compiz | | 2016-10-20 |
| .compiz-1 | | 2013-04-09 |
| .conda | | 2019-09-19 |
| .config | | 2020-09-17 |
| .dbus | | 2011-04-01 |
| .emacs.d | | 2010-09-17 |
| .fltk | | 2016-01-27 |
| .fontconfig | | 2019-04-10 |
| .gconf | | 2020-11-02 |
| .gconfd | | 2020-11-02 |
| .gnome | | 2014-06-12 |
| .gnome2 | | 2019-05-20 |
| .gnome2_private | | 2010-08-26 |
| .gnupg | | 2020-01-06 |
| .gstreamer-0.10 | | 2016-10-18 |
| .gvfs | | 2011-04-01 |
| .idlerc | | 2020-01-08 |
| .ipython | | 2020-09-17 |
| .java | | 2019-06-26 |
| .libreoffice | | 2015-09-04 |
| .local | | 2020-01-06 |

Follow terminal folder

Remote monitoring

```

- Video: https://asciinema.org/a/zZrSazw5Fh7YmpHvUfvV0nzNi
- SGE2Slurm: https://frama.link/2020-09-SGE2Slurm

IMPORTANT:
- slurm0: Never launch job on this server -> Use srun
- /home: Never launch job from this space
- /projet: Use your /projet folder for its performance, its volumetry
and its independence from the /home space
- /scratch: For your huge temporary files, please use /scratch
but note that files older than 90 days are automatically deleted
- installed softwares can be listed by using:
  - module load avail

CITATION: Please cite the plateform ABiMS in the Acknowledgement of your future publication

Slurm

IMPORTANT -> current Slurm Reservation(s) (scontrol show reservation):
ReservationName=bbmose_7 StartTime=2020-11-02T16:40:00 EndTime=2020-11-17T12:00:00 Duration=14-19:20:00
Nodes=n100 NodeCnt=1 CoreCnt=64 Features=(null) PartitionName=(null) Flags=SPEC_NODES
TRES=cpu=128
Users=(null) Accounts=bbmose Licenses=(null) State=ACTIVE BurstBuffer=(null) Watts=n/a

TIPS: Maybe, you can still submit jobs in the meantime by indicate to Slurm the smaller job duration: exam
ple with 5 days long
sbatch --time=5-00:00:00 -p bigmem myscript.sbatch

Last login: Fri Nov 6 10:24:57 2020 from 82.252.141.107

Command 'keychain' not found, but can be installed with:

apt install keychain
Please ask your administrator.

mhoebeke@slurm0:~$

```

slurm0 0% 3,09 GB / 125,69 GB 0,40 Mb/s 0,32 Mb/s 148 days lecorguille garczare lecorguille mhoebeke

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

File Transfer using FileZilla

Defining the connection parameters Establishing the connection

The screenshot shows the FileZilla interface with the 'Gestionnaire de Sites' dialog box open. The dialog box is divided into several sections:

- 1**: The FileZilla main window title bar and menu bar.
- 2**: The 'Nouveau site' button in the 'Gestionnaire de Sites' dialog.
- 3**: The 'slurm0.sb-roscoff.fr' site entry in the 'Mes Sites' list.
- 4**: The 'SFTP - SSH File Transfer Protocol' dropdown menu in the 'Général' tab.
- 5**: The 'slurm0.sb-roscoff.fr' text input field for the host.
- 6**: The 'Demander le mot de passe' dropdown menu for the authentication type.
- 7**: The 'mhoebeke' text input field for the username.
- 8**: The 'Connexion' button at the bottom of the dialog.

Other visible elements include the 'Général', 'Avancé', 'Paramètres de transfert', and 'Jeu de caractères' tabs, and the 'Couleur de fond' dropdown menu set to 'Aucune'.

File Transfer using FileZilla

Validating the connection

The screenshot shows the FileZilla client interface. At the top, the title bar reads "slurm0.sb-roscoff.fr - sftp://mhoebeke@slurm0.sb-roscoff.fr - FileZilla". The main window displays the connection status as "Connexion à slurm0.sb-roscoff.fr...". A dialog box titled "Clé de l'hôte inconnue" is overlaid on the interface. The dialog contains the following text:

La clé du serveur hôte est inconnue. Vous n'avez aucune garantie que ce serveur est bien le bon.

Détails

Hôte : slurm0.sb-roscoff.fr:22
Algorithme de la clé de l'hôte : ssh-ed25519 255
Empreintes : SHA256: oG6Od0UEMQ+3ZpHAKlNuRsTylTgQf8Dn4w5ilUP3znI=
MD5: 15:66:9f:52:f9:aa:be:0b:ca:92:be:73:c2:22:1a:36

Approuver ce serveur et l'associer à la connexion ?

Toujours faire confiance à cet hôte, ajouter cette clé au cache

Buttons: OK, Annuler

The background interface shows a local site at "C:\Users\..." and a list of files and folders including ".anaconda", ".conda", ".config", ".DbSchema", ".docker", ".eclipse", and ".freemind". The status bar at the bottom indicates "Déconnecté." and "Fichiers en file d'attente".

File Transfer using FileZilla

- Selecting the source folder
- Selecting the destination folder
- Selecting the file(s) to transfer
- Transfer through *drag 'n drop*

The screenshot shows the FileZilla interface with the following details:

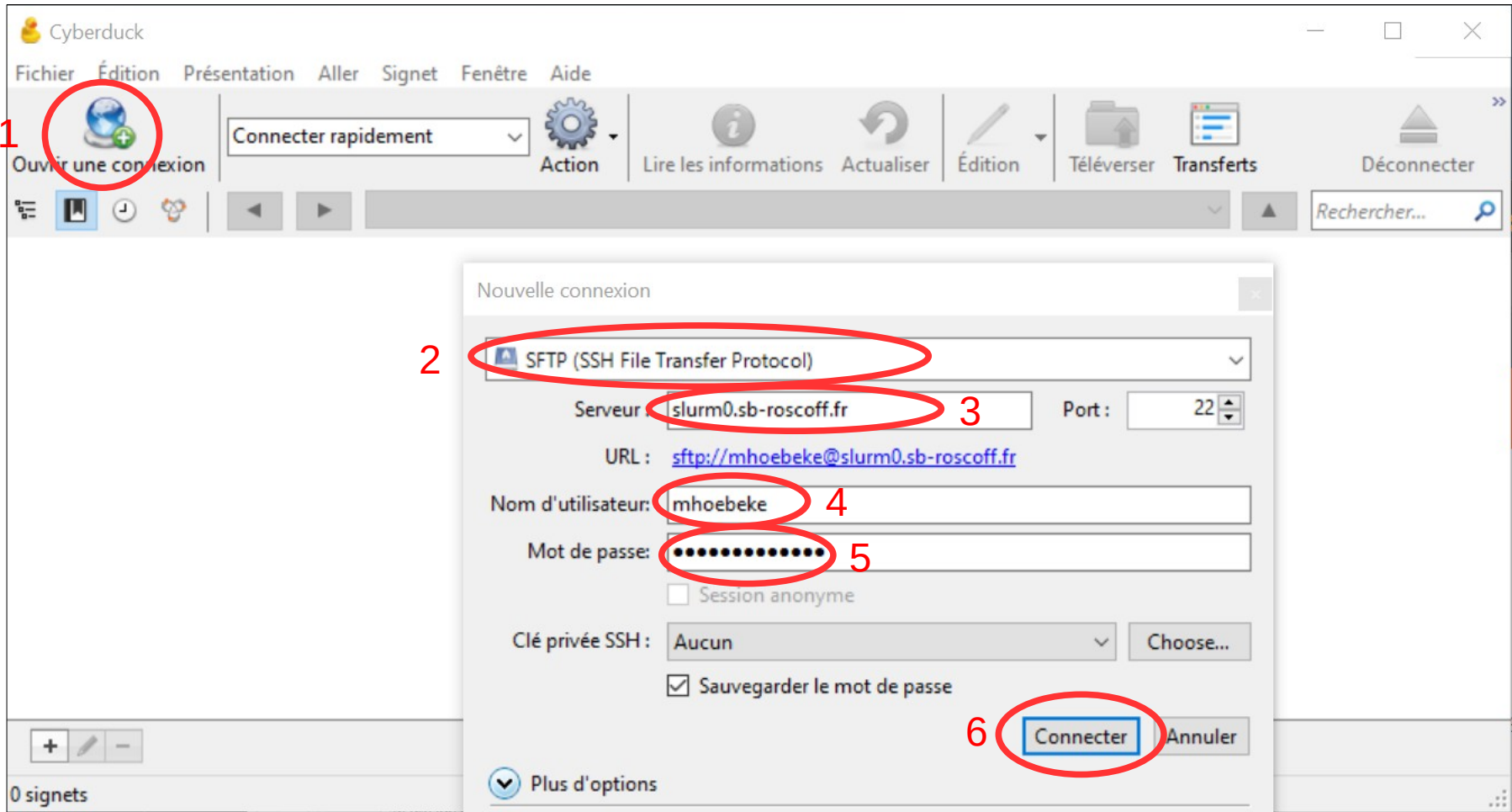
- Host:** slurm0.sb-roscoff.fr - sftp://mhoebeke@slurm0.sb-roscoff.fr
- Site local:** C:\Users\mhoebeke\ (Annotated with red box 1)
- Site distant:** /home/fr2424/sib/mhoebeke (Annotated with red box 2)
- Local File List:**

| Nom de fichier | taille d... | Type de fich... | Dernière m... |
|----------------|-------------|-----------------|---------------|
| .. | | | |
| .anaconda | | Dossier de f... | 29/05/2019... |
| .conda | | Dossier de f... | 08/10/2019... |
| .config | | Dossier de f... | 17/03/2020... |
| .DbSchema | | Dossier de f... | 07/04/2020... |
| .docker | | Dossier de f... | 03/11/2020... |
| .eclipse | | Dossier de f... | 09/01/2019... |
| .freemind | | Dossier de f... | 31/03/2020... |
- Remote File List:**

| Nom de fichier | Taille ... | Type d... | Dernière ... | Droits ... | Proprié... |
|----------------|------------|------------|--------------|------------|------------|
| .. | | | | | |
| .agisoft | | Dossier... | 25/09/20... | drwx-... | mhoeb... |
| .cache | | Dossier... | 06/01/20... | drwx-... | mhoeb... |
| .compiz | | Dossier... | 20/10/20... | drwx-... | mhoeb... |
| .compiz-1 | | Dossier... | 09/04/20... | drwxr... | mhoeb... |
| .conda | | Dossier... | 19/09/20... | drwxr... | mhoeb... |
| .config | | Dossier... | 17/09/20... | drwx-... | mhoeb... |
| .dbus | | Dossier... | 01/04/20... | drwx-... | mhoeb... |
- Annotations:**
 - 1:** Red box around the local file tree.
 - 2:** Red box around the remote file tree.
 - 3:** Red oval around the local file list.
 - 4:** Red arrow pointing from the local file list to the remote file tree.

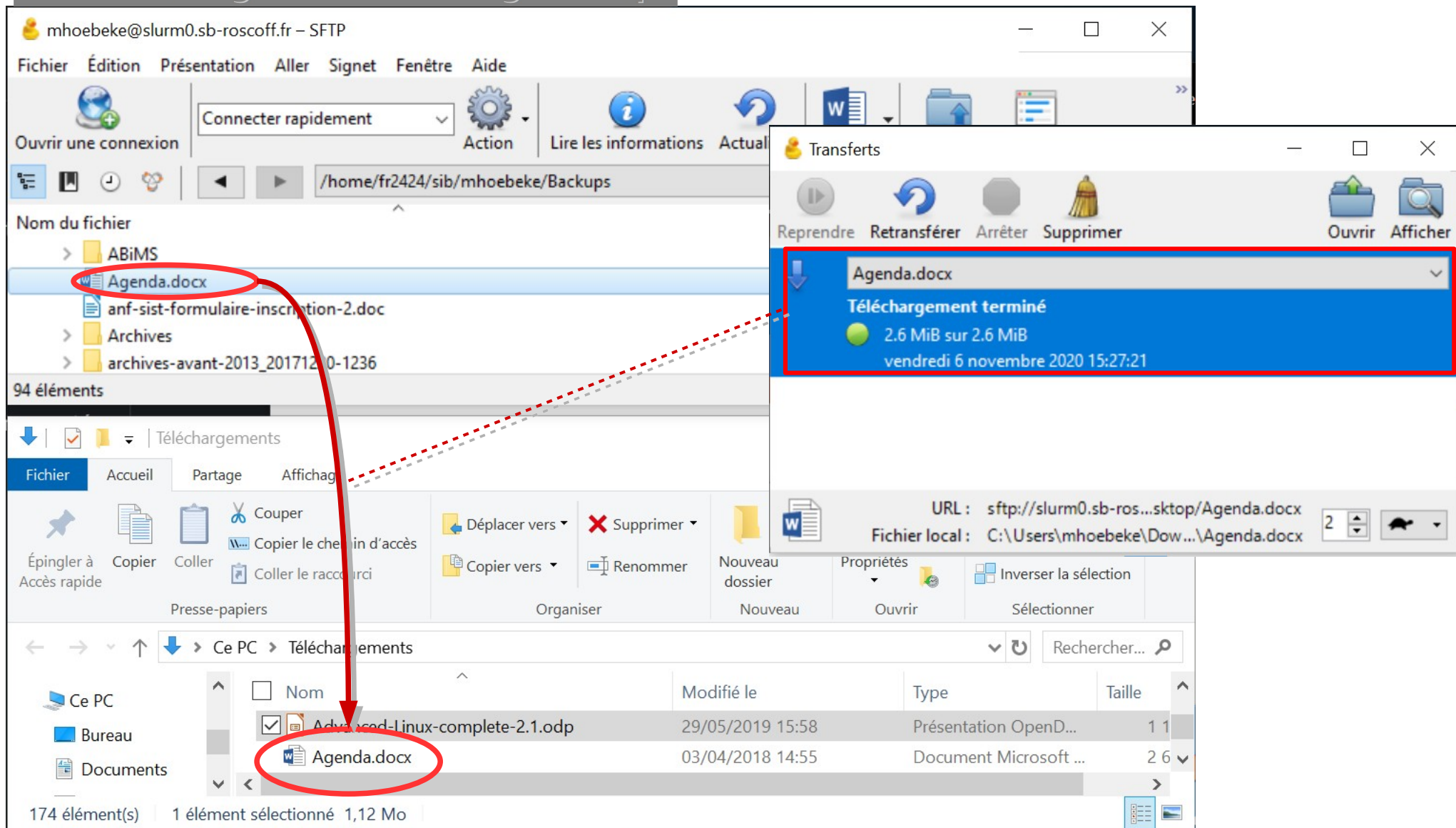
File Transfer using Cyberduck

Defining the connection parameters Establishing the connection



File Transfer using Cyberduck

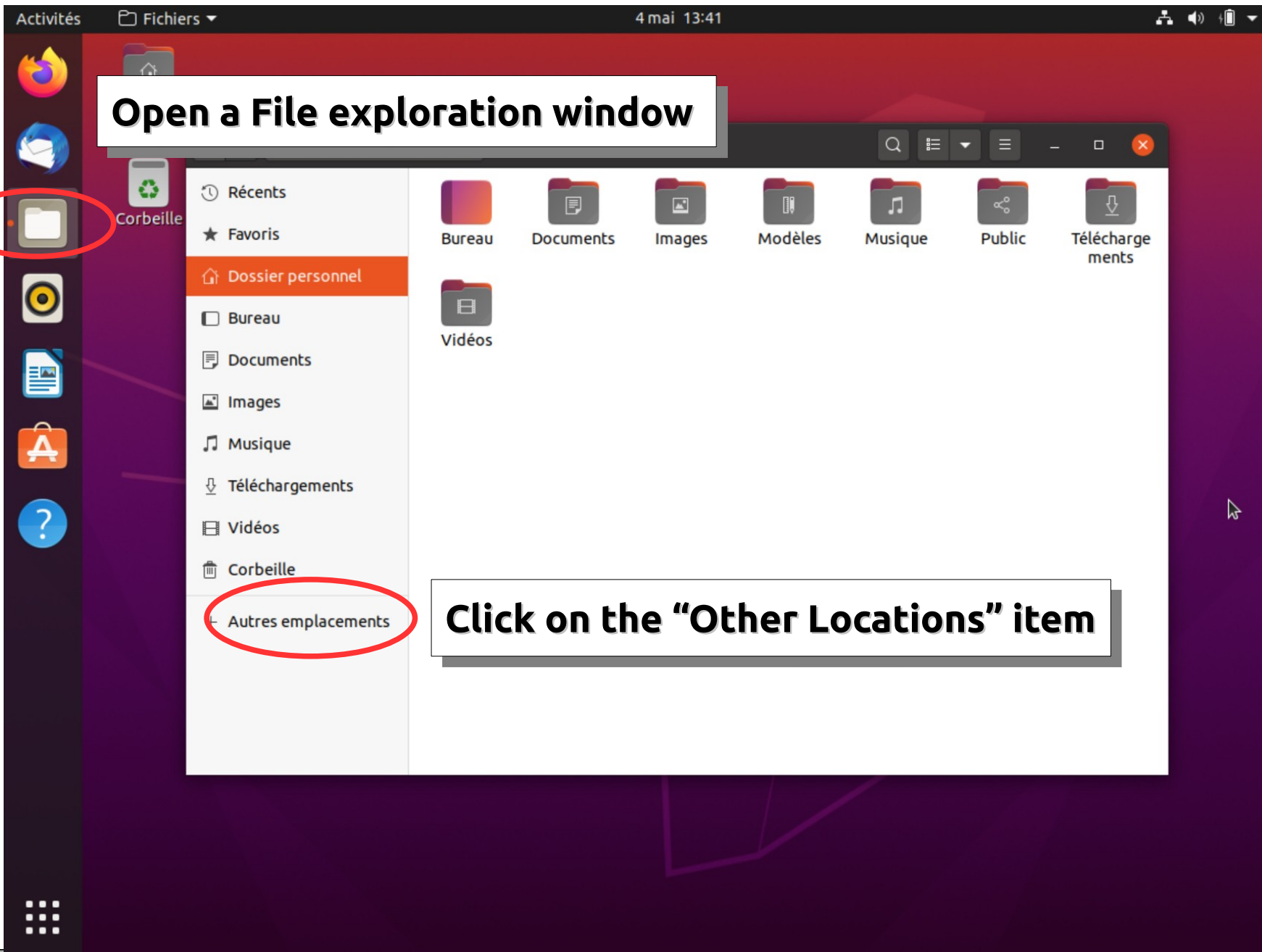
Transferring files with drag 'n drop



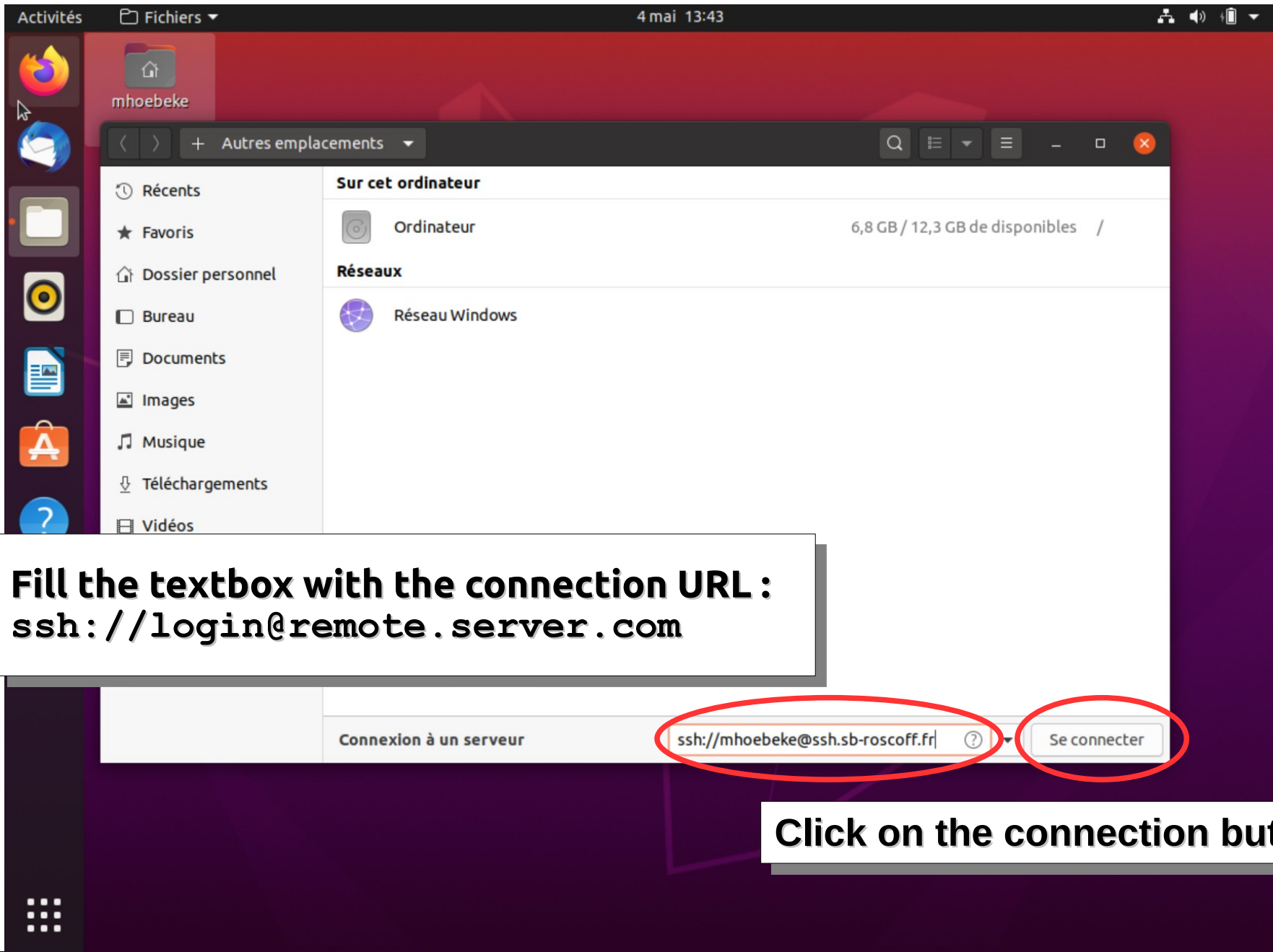
The screenshot illustrates the file transfer process in Cyberduck. The main window shows a remote SFTP connection to `mhoebeke@slurm0.sb-roscoff.fr`. The local pane shows the file `Agenda.docx` being dragged from the remote location to the local `Téléchargements` folder. A secondary window titled `Transferts` displays the progress of the download, indicating that the file is `2.6 MiB` in size and has been successfully transferred.

| Nom | Modifié le | Type | Taille |
|---------------------------------|------------------|------------------------|--------|
| Advanced-Linux-complete-2.1.odp | 29/05/2019 15:58 | Présentation OpenD... | 1 1 |
| Agenda.docx | 03/04/2018 14:55 | Document Microsoft ... | 2 6 |

File Transfer using GNOME/Ubuntu



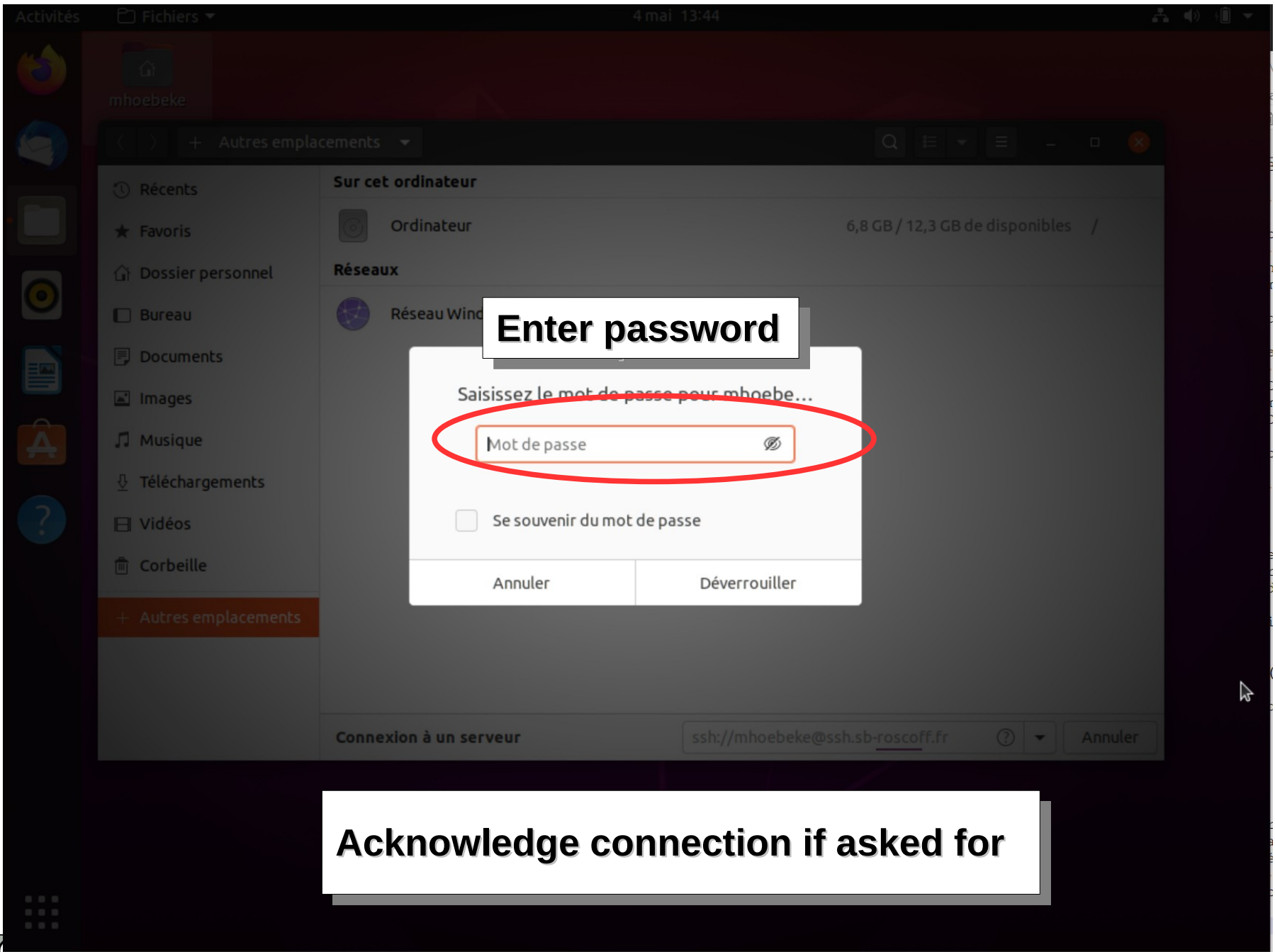
File Transfer using GNOME/Ubuntu



Fill the textbox with the connection URL :
`ssh://login@remote.server.com`

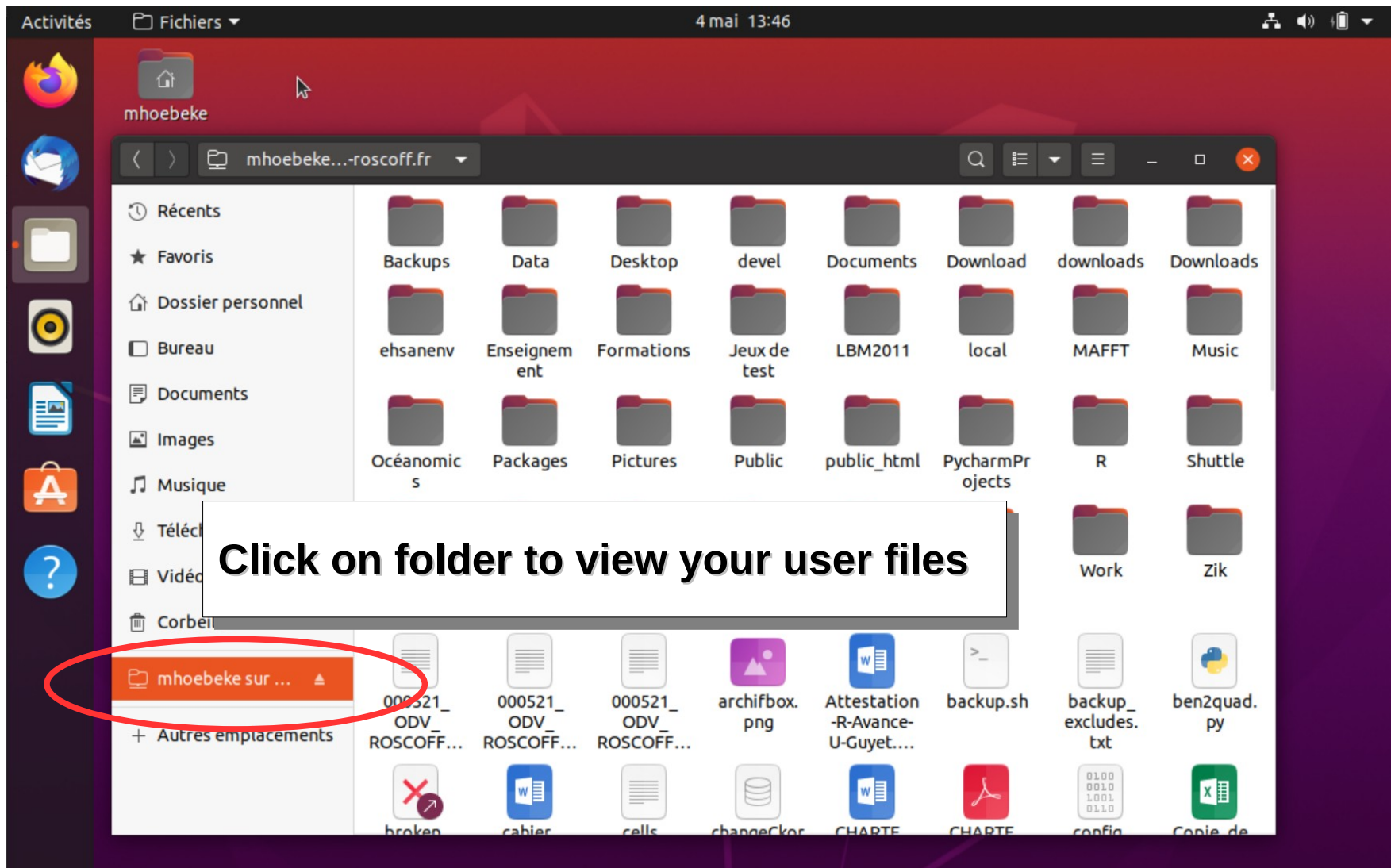
Click on the connection button

File Transfer using GNOME/Ubuntu



Acknowledge connection if asked for

File Transfer using GNOME/Ubuntu



Use the ordinary copy/paste/drag/drop between local files and remote files.

- 1 Purpose of an Operating System – why Linux ?
- 2 Establishing a connection and transferring files
- 3 The Command Line Interface**
- 4 The File System
- 5 Manipulating File Contents
- 6 Users, Groups and Access Control
- 7 Processes

Structure of the *Command Line*

```
[stage01@slurm0 ~]$ head -n 20 insulin.fas #print the first 20 lines
```

[stage01@slurm0 ~] \$ The *prompt* : displays the current user's login (**stage01**), the host (or machine) name (**slurm0**), the current directory (working directory) (**~**)

head The name of the program to run (first word following the prompt)

-n 20 A command *option* (**-n**) possibly with an associated value (**20**).

insulin.fas A command *argument*

print (...) Comments (ignored by the command)

- The space character is used to separate the various fields of the command line.
- **The character case (upper/lower) is important** (**head** is not the same as **HEAD**)
- Each command has its own set of arguments and options
- **The Enter key (↵) is used to run the program**

Each (self respecting) command is documented

To display a short documentation on how to use a command, it is possible to use the `-help` or `-h` options

```
[stage01@slurm0~]$ ls --help
```

```
Utilisation : ls [OPTION]... [FICHIER]...
```

```
Afficher des renseignements sur les FICHIERS (du répertoire actuel par défaut).
```

To get a more detailed documentation, it is possible to use the `man` (i.e. *manual*) command Giving it as an argument the name of the command for which to display the documentation.

```
[stage01@slurm0~]$ man ls
```

```
LS(1)
```

```
User Commands
```

```
LS(1)
```

```
NAME
```

```
ls - list directory contents
```

```
SYNOPSIS
```

```
ls [OPTION]... [FILE]...
```

```
DESCRIPTION
```

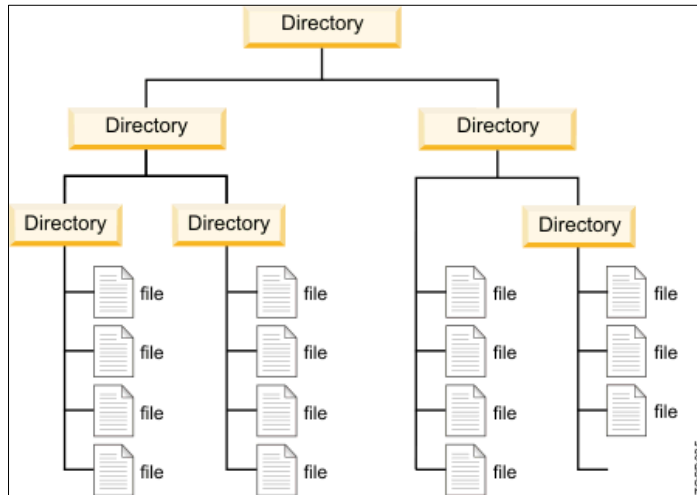
```
List information about the FILES (the current directory by default). Sort entries alphabetically if none of -cftuvSUX nor --sort is specified.
```

- 1 Purpose of an Operating System – why Linux ?
- 2 Establishing a connection and transferring files
- 3 The Command Line Interface
- 4 The File System**
- 5 Manipulating File Contents
- 6 Users, Groups and Access Control
- 7 Processes

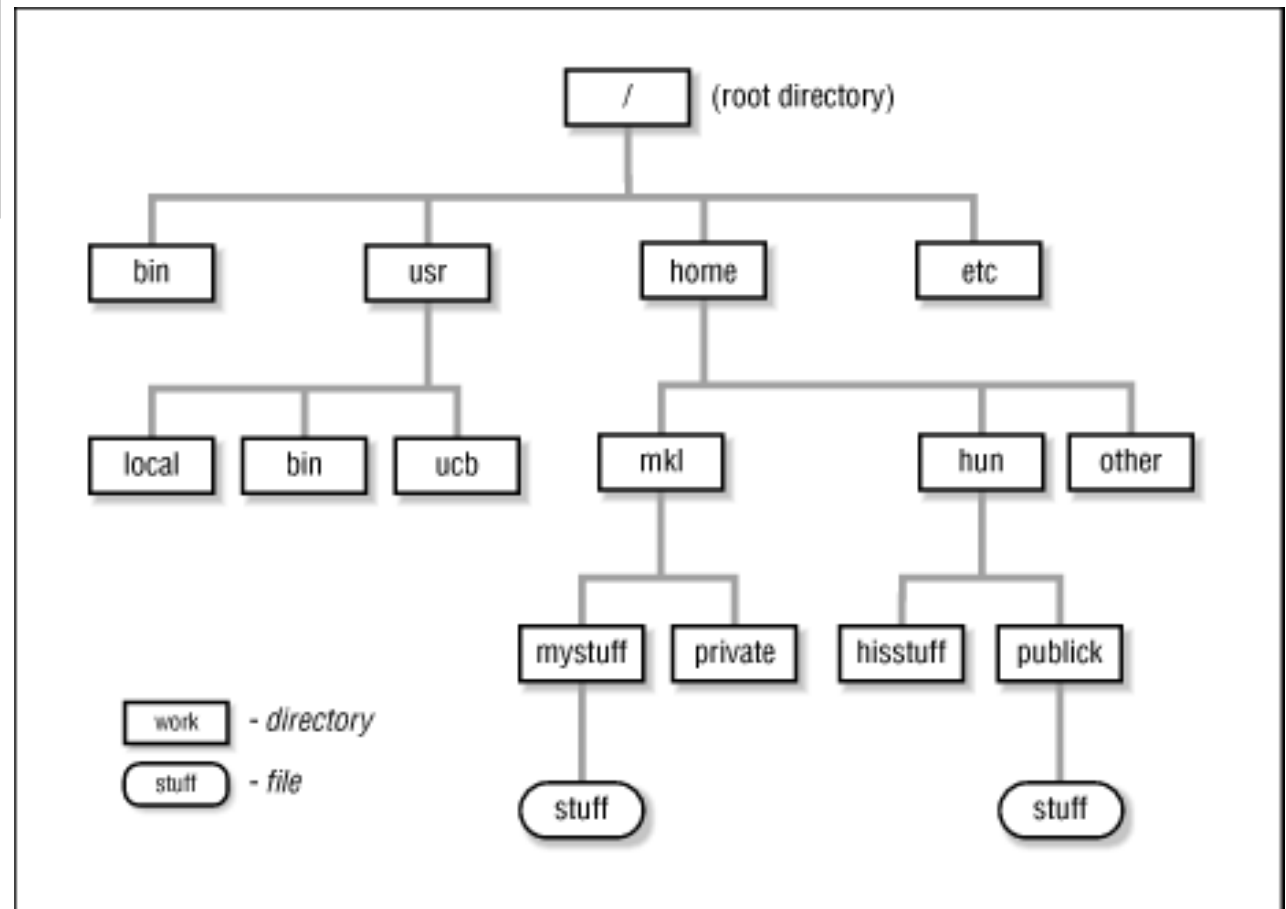
In most Operating Systems, data are stored in **files** (text, images, tables, sequences, measurement series...). Quickly, the number of files increases and it becomes necessary to organize them to avoid getting lost. This is done by grouping them in **folders or directories**. Folders can be stored in other folders, which in turn can be stored in other folders, which in turn can be stored in other folders...



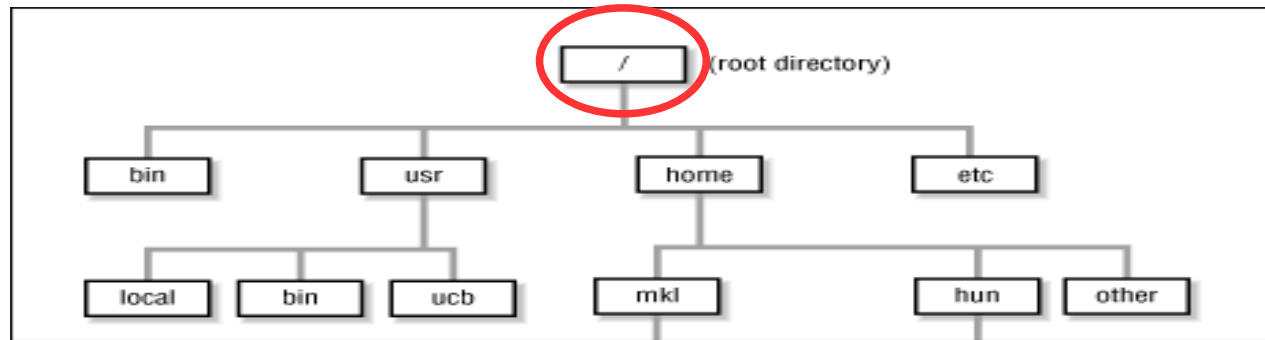
The Concept



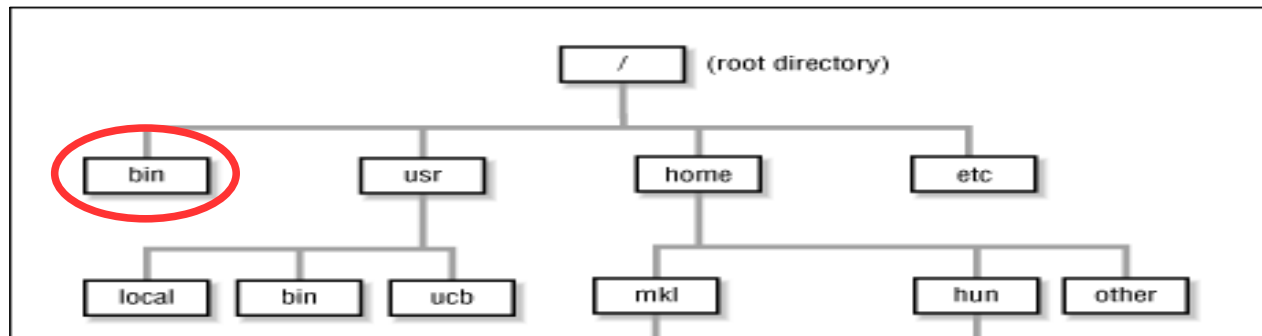
The way it has been implemented in Linux (Unix)



The File System | Some important directories

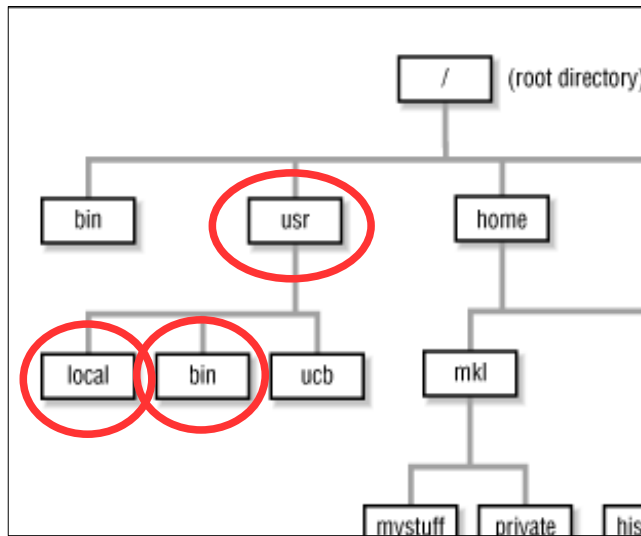


Slash or root directory : the **unique** (top level) entry point for the whole file system. A *path* leading to a file or a directory can always be specified starting from the root directory



/bin : directory containing the major part of system commands.

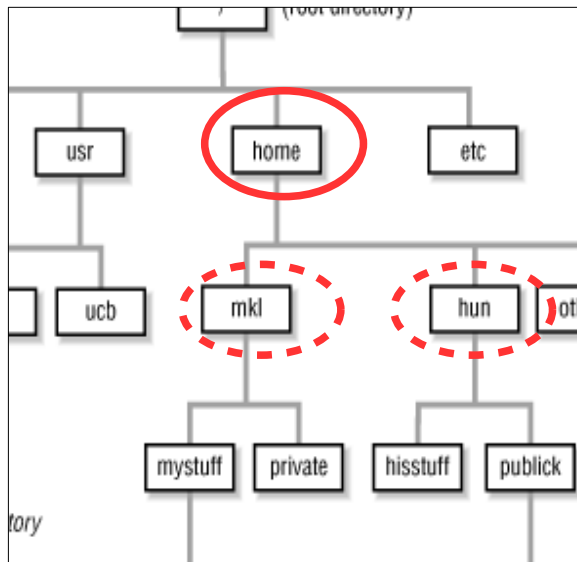
The File System | Some important directories



/usr : directory containing sub-directories with “user” commands.

/usr/bin : directory with frequently used commands (not as important to run the system as those in **/bin**).

/usr/local : directory with subdirectories containing tools installed on this specific machine (in particular **/usr/local/bin**)



/home : top level directory of the user data directory tree.

Its organization depends on the number of users having an account on the machine :

- few users (on workstations) : each users subdirectory is located directly in **/home**.
- many users : user subdirectories are located in sub(sub(sub))folders.

At the SBR : the organisation matches the research unit and team layout, as in : **/home/fr2424/sib/mhoebeke**.
The user directories are accessible on all (Linux) machines of the campus.

/tmp : a directory where anyone can read and write files (but, only a file's creator has the rights to remove her own stuff). Handy to share data with colleagues **on the same machine** (but beware the volume of data).

SBR Specifics

/shared/projects : toplevel directory for project directories, destined to hold project data than needs to be backed-up (initial data sets, final results).

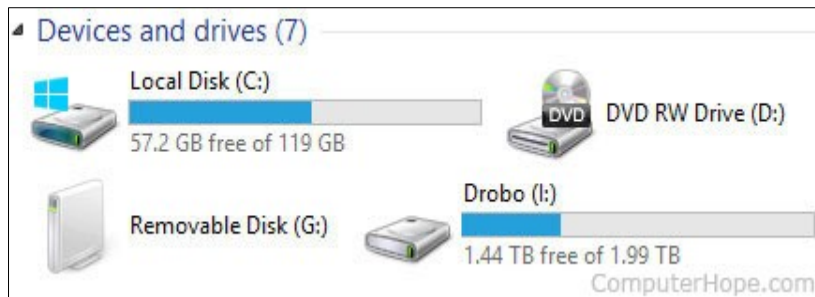
/shared/projects/tp_aa_XXXXXX : the main place you'll be working in today

/shared/home/tpYYYYYY : your home directory

/shared/scratch : toplevel directory of a directory tree mimicking the project directory tree destined to hold "work" data such as those generated by ongoing computations, and not needing backups. **"Old" files are automatically removed after a certain inactivity delay.**

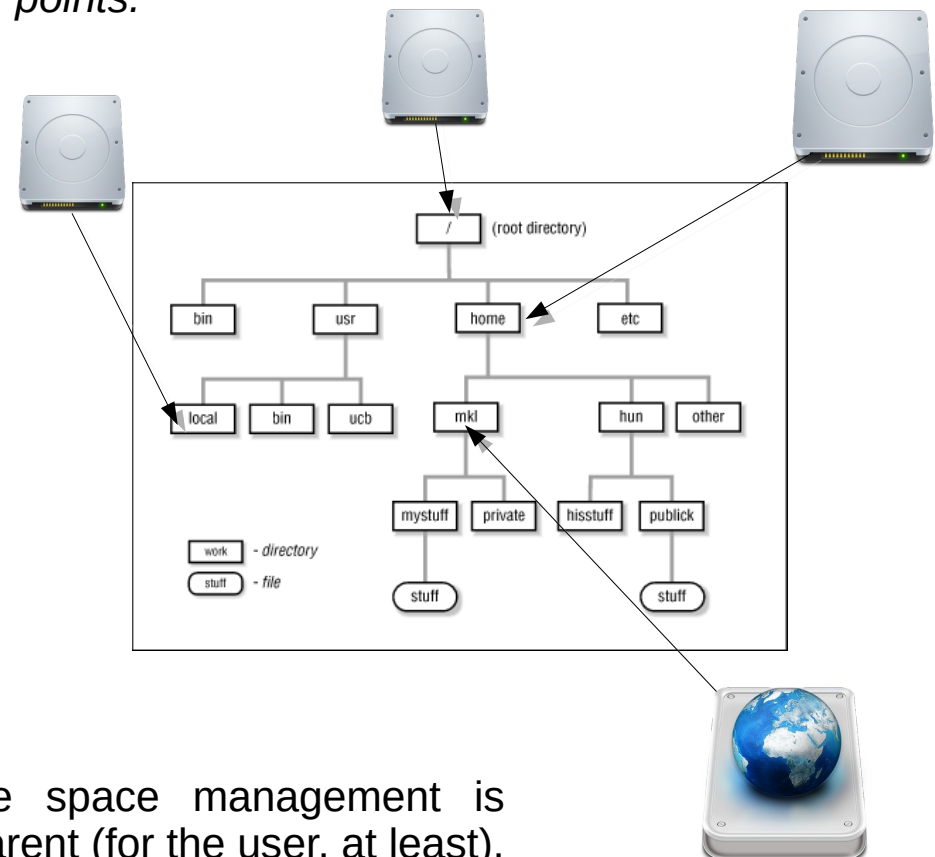
The File Systems | So where are the actual disks ?

Windows-like systems define a single letter for each storage device (hard disk, DVD reader/writer, USB key, network drive), generating a **non-unique top level directory tree**.



In order to use a file, one has to know on which drive (letter) it is physically located. Disk space extension by adding new disks leads to the definition of new letters. Moving data between disks may entail application reconfiguration.

UNIX/Linux like systems associate all devices (not only storage devices), both local and remote, to directories in the **single rooted directory tree**, through so-called *mount points*.



Storage space management is transparent (for the user, at least).

Command execution takes place in the context of a **session**, defining at all times a **current user** (the user running the command) and a **current directory** or **working directory** (the directory where the command has been typed in).

When starting a new session, the current directory is always the **home directory**.

Who am I (who is the current user) ?

```
[stage01@slurm0 ~]$ whoami  
stage01
```

Where am I (what is the current directory) ?

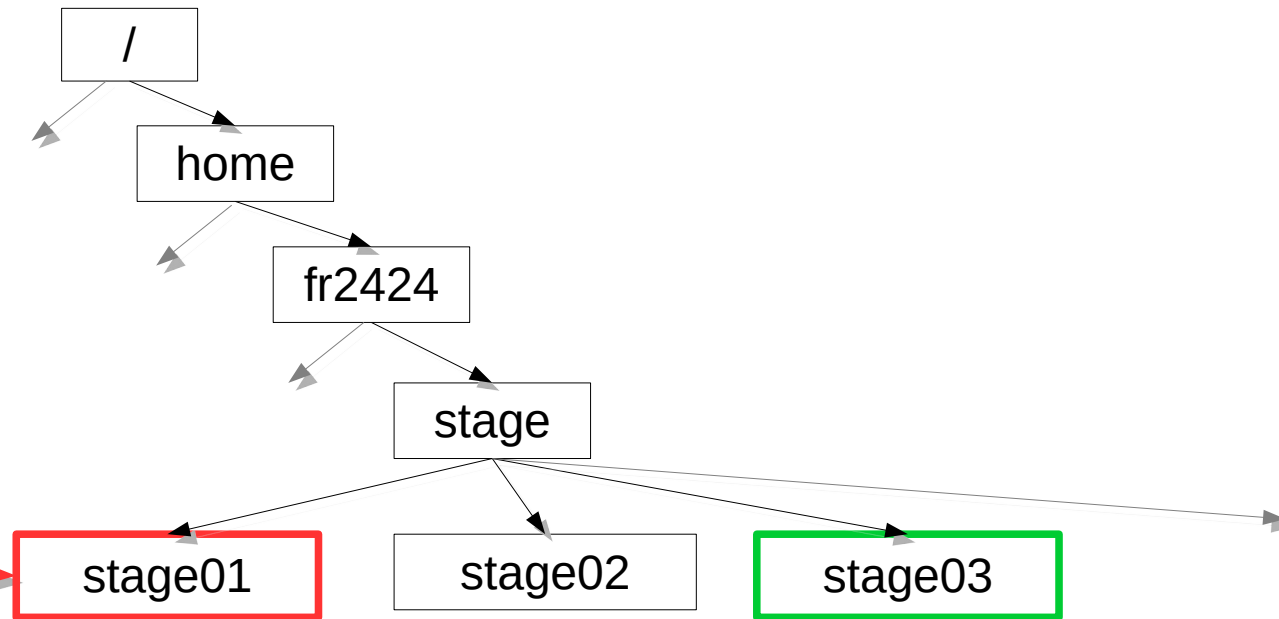
```
[stage01@slurm0 ~]$ pwd # print working directory  
/home/fr2424/stage/stage01
```

What can be found “here” (what are the contents of the current directory) ?

```
[stage01@slurm0 ~]$ ls #list (contents of working directory)  
Bureau Documents Images Modèles Musique Public  
Téléchargements Vidéos
```

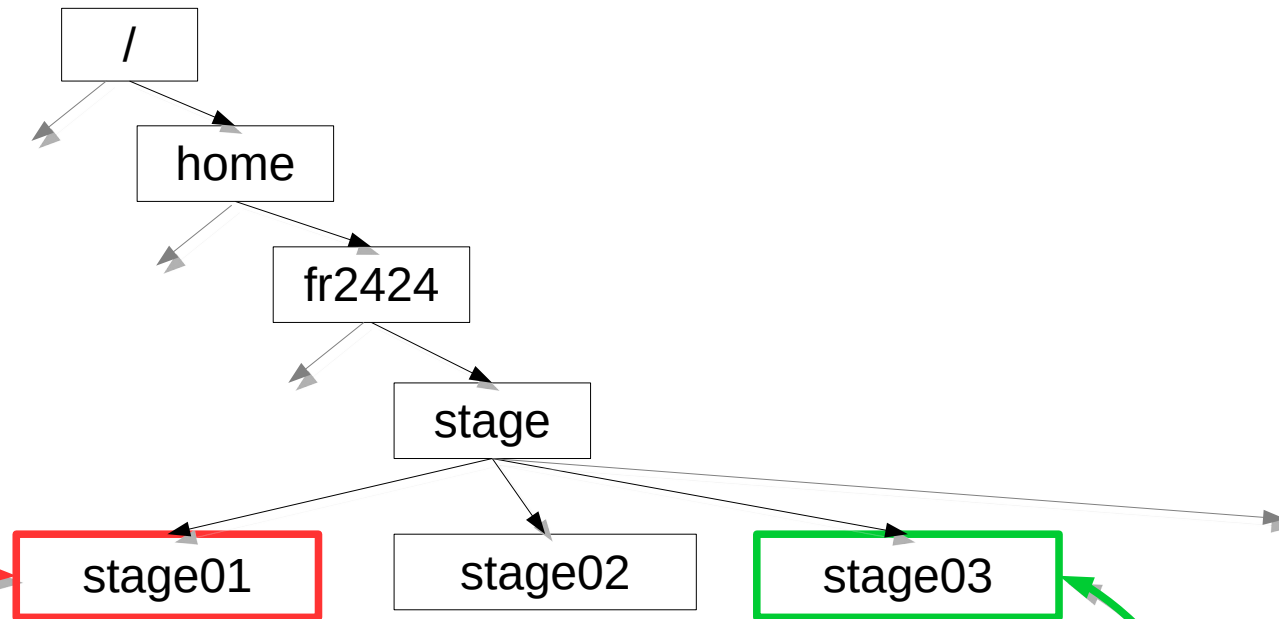
To change the current directory (a.k.a “move around” in the directory tree)

```
[stage01@slurm0 ~]$ pwd # print working directory  
/home/fr2424/stage/stage01
```



To change the current directory (a.k.a “move around” in the directory tree)

```
[stage01@slurm0 ~]$ pwd # print working directory  
/home/fr2424/stage/stage01
```

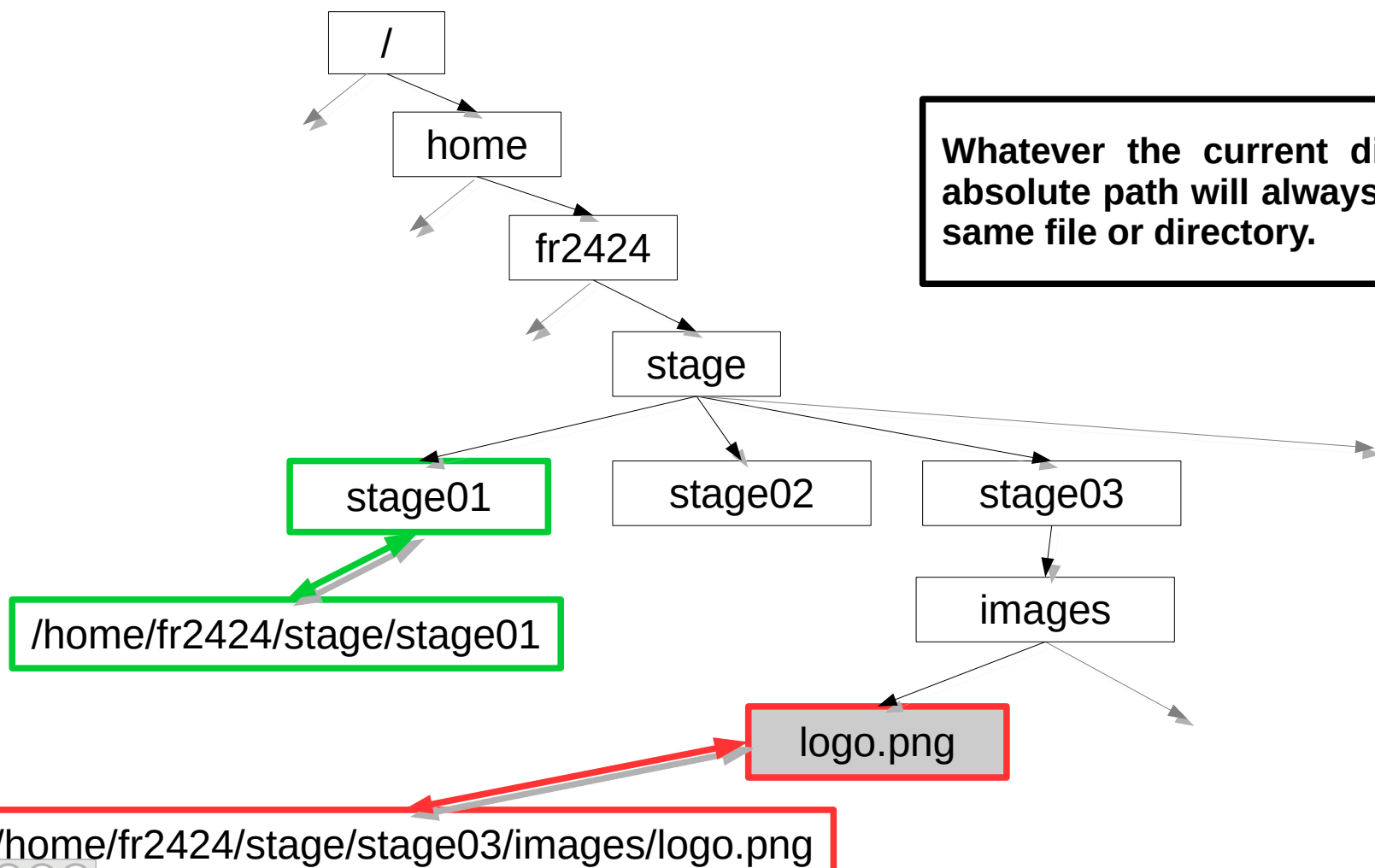


```
[stage01@slurm0 ~]$ cd /home/fr2424/stage/stage03 #change dir  
[stage01@slurm0 ~]$ pwd  
/home/fr2424/stage/stage03
```

Absolute paths

Referring to files and directories located in the file system, as command arguments or options, is done through **paths**.

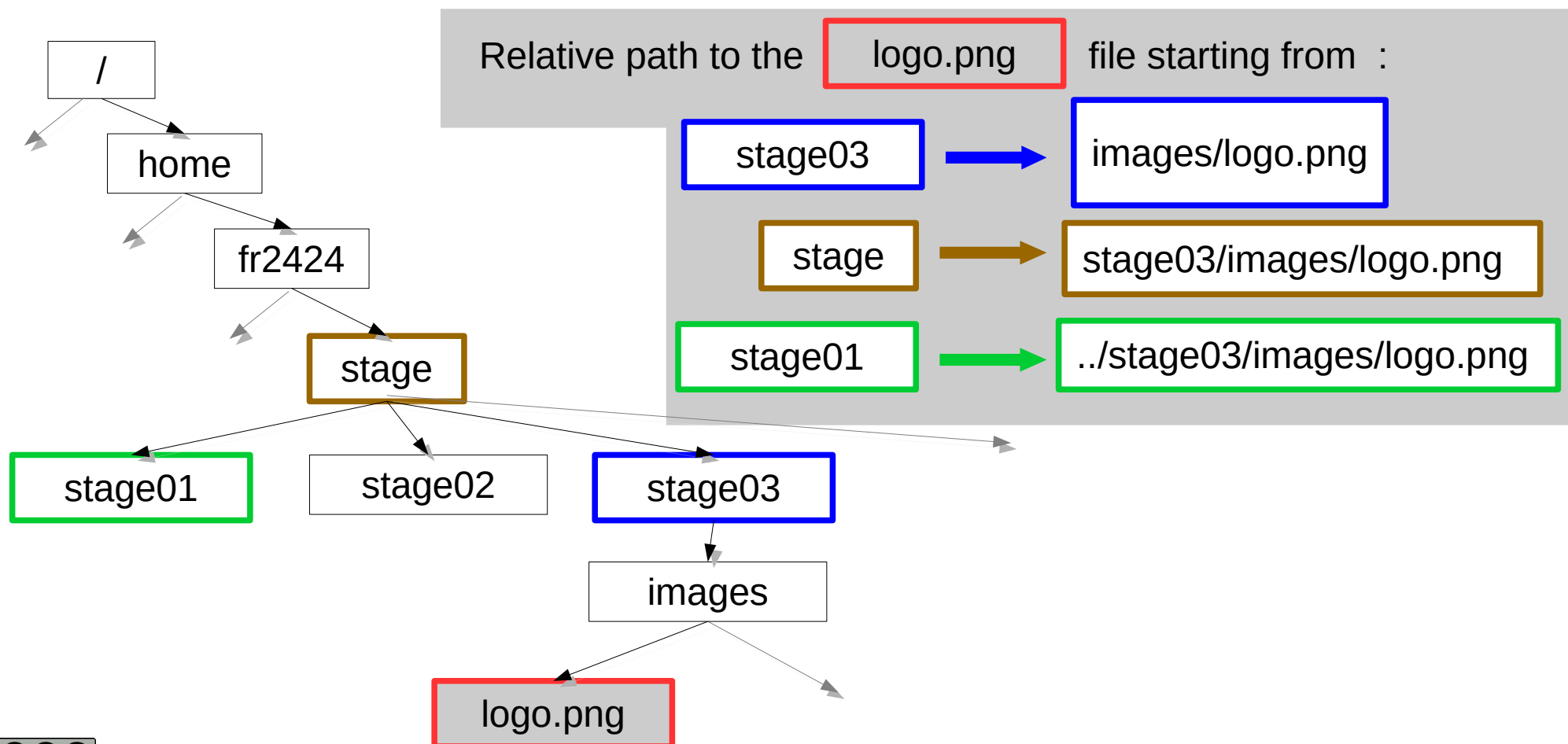
Absolute paths are built starting from the root directory and adding the subdirectories one by one separated by a slash character (/), until the desired file or directory is reached.



Relative paths

Relative paths are built with the **current directory as starting point** and traversing the directory tree upwards or downwards, until the desired directory or file is reached. The successive path components are separated with *slash (/)* characters, and :

- On each **upward step** in the tree, two dots (**..**) are added to the path.
- On each **downward step** in the tree, the name of the directory is added to the path.



A (provisional) conclusion

A shortcut to refer to the current directory :

- The dot ("." character) always refers to the current directory

```
[stage01@slurm0 ~]$ pwd
/home/fr2424/stage/stage01
[stage01@slurm0 ~]$ cd . # ???
[stage01@slurm0 ~]$ pwd
/home/fr2424/stage/stage01
```

Use case : run a command file located in the current directory : `./mycommand`

A shortcut for the home directory :

- The tilde ("~" character) always refers to the home directory of the current user

```
[stage01@slurm0 ~]$ pwd
/home/fr2424/stage/stage03
[stage01@slurm0 ~]$ cd ~ # change to home dir
[stage01@slurm0 ~]$ pwd
/home/fr2424/stage/stage01
```

A shortcut to “your” directory (containing the files you’ll be using today).

```
[stage01@slurm0 ~]$ pwd
/home/fr2424/stage/stage03
[stage01@slurm0 ~]$ cdmystuff #change to dir for today's work
[stage01@slurm0 ~]$ pwd
/shared/projects/stage/stage03
```

Teaser : **cdmystuff** is a so-called *alias* which we’ll learn to define at the end of the day.

Listing the contents of a directory: `ls`

```
[stage01@slurm0 ~]$ ls # no arguments : current dir
Bureau Documents Images Modèles Musique Public Téléchargements
Vidéos
[stage01@slurm0 ~]$ ls /tmp/Linux-Initiation # absolute path
acteur.csv cours insulin.fas insulin_vs_nt.blast tmp
[stage01@slurm0 ~]$ ls .. # relative dir path (parent dir)
common          stage02  stage1   stage17  stage24  stage31  stage6
common.linux-avance stage03  stage10  stage18  stage25  stage32  stage7
(...)
```

“Hidden” files (and directories)

By default, `ls` does not display files having names starting with a dot. The `-a` (all) option needs to be added for them to be included (the `lsa` shortcut can also be used instead of `ls`).

```
[stage01@slurm0 ~]$ ls -a # also show hidden files
.          Bureau Documents .kde      Public    .zshrc
..         .cache  .emacs  .local    Téléchargements
.bash_logout .compiz .gconf   Modèles  Vidéos
(...)
```

Listing files matching a particular pattern

Using the `*` character in an argument of `ls` restricts the list to the files and directories whose names match the pattern formed by the argument:

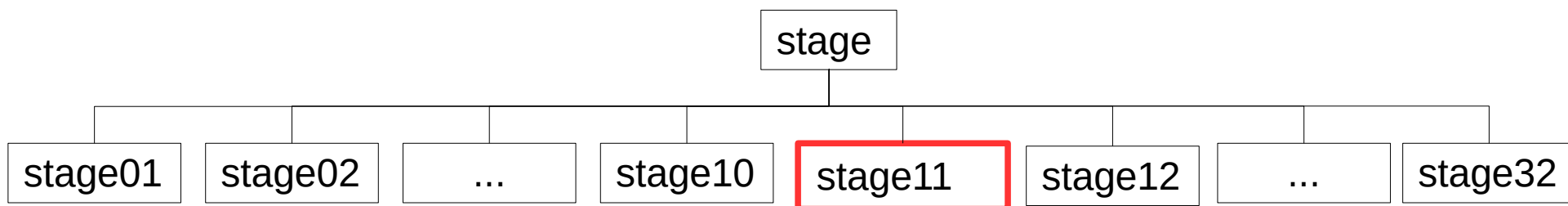
- `image*` : all files starting with the letters *image* (*image-001*, *images-des-vacances*, *imagettes*)
- `*seq*` : all files having the letters *seq* in their names (*sequences*, *mes-sequences*, *maiseqoidon*)
- `*` : each and every file (no restriction)

```
[stage01@slurm0 ~]$ ls Linux-Initiation/*ins*  
Linux-Initiation/insulin.fas  Linux-Initiation/insulin_vs_nt.blast
```

Using *autocompletion*

To avoid to have to type in long filenames it is possible to use the `[TAB]` key. Pressing the `[TAB]` key once launches a file name (or directory) lookup to determine which ones start with what has already been typed.

- If there is a single match, it will be added to the command line,
- If there are several matches another press of the `[TAB]` key will list them all.



```

[stage01@slurm0 ~]$ ls ../sta[TAB]
[stage01@slurm0 ~]$ ls ../stage
[stage01@slurm0 ~]$ ls ../stage[TAB]
stage01/ stage07/ stage12/ stage18/ stage23/ stage29/ stage34/ stage7/
stage02/ stage08/ stage13/ stage19/ stage24/ stage30/ stage35/ stage8/
stage03/ stage09/ stage14/ stage20/ stage25/ stage31/ stage36/ stage9/
(...)
  
```

Displaying the contents of a whole directory (sub)tree with a single command

Through the `ls` command, to which the `-R` (recursive) option is added:

```
[stage01@slurm0 ~]$ ls -R Linux-Initiation
Linux-Initiation/:
acteur.csv  cours  insulin.fas  insulin_vs_nt.blast  tmp
Linux-Initiation/cours:
Linux-Initiation/tmp:
```

With the `tree` command:

```
[stage01@slurm0 ~]$ tree Linux-Initiation
Linux-Initiation
├── acteur.csv
├── cours
├── insulin.fas
├── insulin_vs_nt.blast
└── tmp

2 directories, 3 files
```

Organising data by creating subdirectories

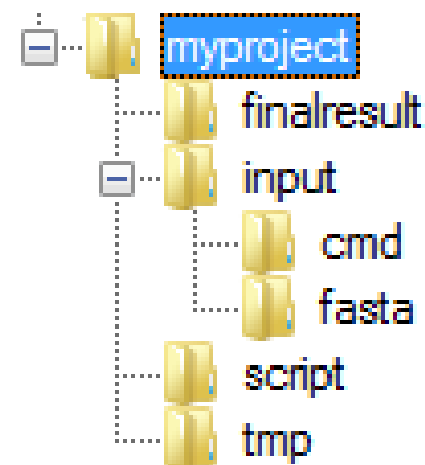
With the `mkdir` (*make directory*) command. By default, only the last directory of the path given as argument is created:

```
[stage01@slurm0 ~]$ mkdir Linux-Initiation/tmp/essais
[stage01@slurm0 ~]$ ls -R Linux-Initiation
Linux-Initiation/:
acteur.csv  cours  insulin.fas  insulin_vs_nt.blast  tmp
Linux-Initiation/cours:
Linux-Initiation/tmp:
Linux-Initiation/tmp/essais:
```

The `-p` option enables the creation of a whole subtree in a single step:

```
[stage01@slurm0 ~]$ mkdir -p
Linux-Initiation/exercices/ex1/data
[stage01@slurm0 ~]$ ls -R Linux-Initiation/
(...)
Linux-Initiation/exercices:
ex1
Linux-Initiation/exercices/ex1:
data
Linux-Initiation/exercices/ex1/data:
```

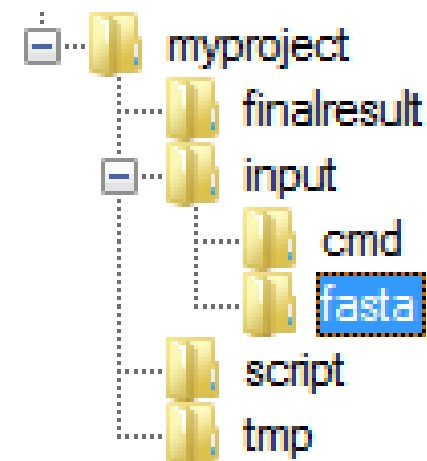
- Create the following directory structure in your home directory:



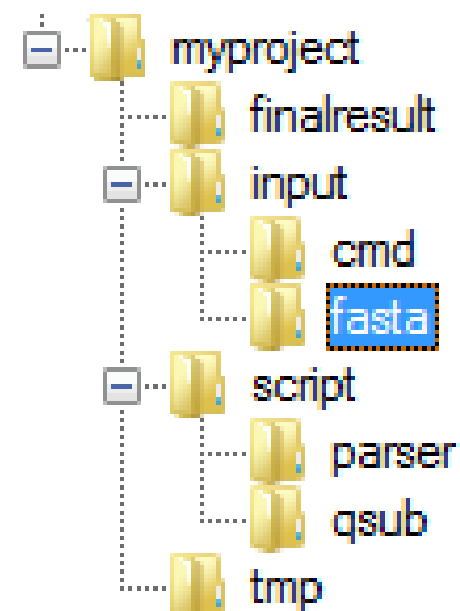
- Check it has been correctly created by displaying it :

Using a single command line for each of the following items :

- Return to your home directory
- Change to the `fasta` directory



- Create a `parser` directory in the `script` directory



Copying data

The `cp` (*copy*) command copies one or more files, and even whole directory trees. It is used as follows : `cp SRC DEST` where `SRC` is the path to the already existing data (the source) and `DEST` is the path the the destination location.

Ex. 1 : copying a single file

```
[stage01@slurm0 ~]$ cp acteur.csv acteur_bak.csv
```

Ex. 2 : copying a single file in another directory

```
[stage01@slurm0 ~]$ cp acteur.csv tmp # keep same filename in DEST  
[stage01@slurm0 ~]$ cp acteur.csv tmp/stars.csv # change filename
```

Ex. 3 : copying a set of files matchin a pattern to another directory

```
[stage01@slurm0 ~]$ cp insulin* tmp  
[stage01@slurm0 ~]$ ls tmp  
insulin.fas  insulin_vs_nt.blast
```

Ex. 4 : copying a complete directory structure using the `-r` (*recursive*) option

```
[stage01@slurm0 ~]$ cp -r ../stage10/exercices/solutions .
```

Move or rename data

The **mv** (*move*) command, depending on its arguments, either renames or moves one or more files, and possibly whole directories.

It is used as follows: **mv SRC DEST** where **SRC** is the path to the already existing data (the source data) and **DEST** either the new name for the file or the directory in which it has to be moved.

Ex. 1 : renaming a single file

```
[stage01@slurm0 ~]$ mv acteur.csv liste_acteurs.csv
```

Ex. 2 : moving a single file to an (already existing) directory

```
[stage01@slurm0 ~]$ mv acteur.csv tmp # keep same filename in DEST  
[stage01@slurm0 ~]$ mv acteur.csv tmp/stars.csv # change filename
```

Ex. 3 : moving a set of files matching a pattern to another (existing) directory

```
[stage01@slurm0 ~]$ mv insulin* tmp
```

Ex. 4 : moving a complete directory structure

```
[stage01@slurm0 ~]$ mv tmp/work/last_stage/output ./finalresults
```

- If **./finalresults** already exists, the **output** directory will be moved into it.
- if **./finalresults** doesn't exist, it will be created and will contain all the data previously located in **output**

Deleting data

The `rm` (*remove*) command deletes the file(s) whose path(s) are given as argument.



What is deleted with `rm` cannot be restored.

(jamais, never, jamas, nie, nooit, gwech ebet, någonsin, никогда, 曾經)

Ex. 1 : removing a single file

```
[stage01@slurm0 ~]$ rm acteur_bak.csv
```

Ex. 2 : removing a set of files matching a pattern

```
[stage01@slurm0 ~]$ rm insulin*
```

Ex. 3 : removing a complete directory structure using the `-r` (*recursive*) option

```
[stage01@slurm0 ~]$ rm -r Linux-Initiation/tmp
```

Ex. 4 : **Armageddon** : forced (`-f`) removal of a whole directory structure

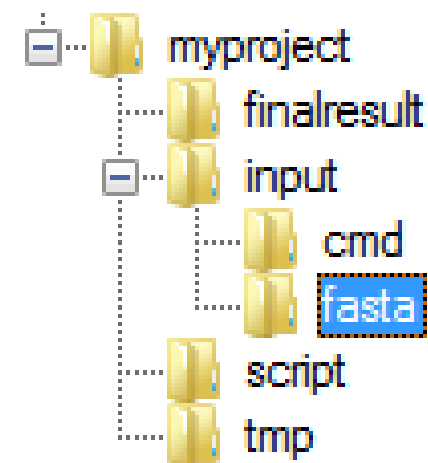
```
[stage01@slurm0 ~]$ rm -rf ~/tmp/worthless_files
```

Special case : removal of an empty directory with the `rmdir` command

```
[stage01@slurm0 ~]$ rmdir ~/tmp/empty_directory
```

Copy the `insulin.fas` file in the `fasta` directory

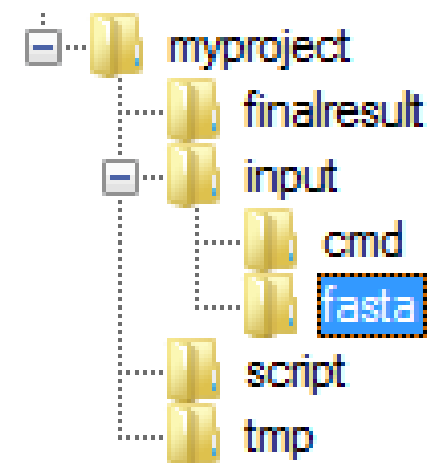
- Go to your home directory



- Copy the file to its destination directory

After making `myproject/finalresult` your current directory

- Move the `insulin.fas` file from the `input/fasta` directory to the `tmp` directory



- Remove the `tmp` directory and all its contents

- 1 Purpose of an Operating System – why Linux ?
- 2 Establishing a connection and transferring files
- 3 The Command Line Interface
- 4 The File System
- 5 Manipulating File Contents**
- 6 Users, Groups and Access Control
- 7 Processes

A few words about file names (1)

Linux is **very** permissive about valid characters in file names (space characters, accents...). It's safer to avoid using them widely. Some recommendations:

- Uppercase & lowercase characters ; digits ; dash ; underscore ; dot.
- Characters with accents or other diacritical signs
- Space characters or other punctuation marks

The "case of the space" character : it can be *despecialized* with the *backslash* (\) or the *double quotes* (")

```
[n00b@slurm0 ~]$ mkdir nouveau dossier # creates 2 dirs, :(
[n00b@slurm0 ~]$ ls .
./:
nouveau
dossier
```

```
[tux@slurm0 ~]$ mkdir nouveau\dossier # creates 1 dir, gg
[tux@slurm0 ~]$ mkdir "nouveau dossier 2" # id.
[tux@slurm0 ~]$ ls .
./:
nouveau dossier
Nouveau dossier 2
```


A few words about file names (2)

Linux doesn't put any requirements on file name extensions (`.txt`, `.csv`, `.pdf`, `.html`, etc.). Any extension can be given to any type of file.

IT IS STRONGLY RECOMMENDED TO REMAIN CONSISTENT

Linux uses other recipes to determine the nature of a file's contents (cf. the `file` command).

Determining the nature of a file (1)

The `file` command displays a hypothesis about a file's nature. It examines the beginning of the file and compares this *fingerprint* to an internal "database" of fingerprints.

Ex. 1 & 2 : Application specific files.

```
[stage01@slurm0 ~]$ file Linux-Initiation.pdf
Linux-Initiation.pdf: PDF document, version 1.4
```

```
[stage01@slurm0 ~]$ file Linux-Initiation.pptx
Linux-Initiation.pptx: Microsoft PowerPoint 2007+
```

Ex. 3 & 4 : Compressed archive files.

```
[stage01@slurm0 ~]$ file Linux-Initiation-supports.zip
Linux-Initiation-supports.zip: Zip archive data, at least
v2.0 to extract
```

```
[stage01@slurm0 ~]$ file Linux-Initiation-supports.tar.gz
Linux-Initiation-supports.tar.gz: gzip compressed data, last
modified: Sat May 7 23:56:36 2017, from Unix
```

Determining the nature of a file (2)

Ex. 5 & 6 : Executable files (binary commands or series of commands in a text file)

```
[stage01@slurm0 ~]$ file /usr/bin/file
/usr/bin/file: ELF 64-bit LSB shared object, x86-64, version 1 (SYSV),
dynamically linked, interpreter /lib64/ld-linux-x86-64.so.2, for GNU/Linux
2.6.32, BuildID[sha1]=a4f09f32eb214a3f9435484fa01d54c939bcf30c, stripped
```

```
[stage01@slurm0 ~]$ file monscript.sh # textfile with commands
monscript.sh: POSIX shell script, ASCII text executable
```

Ex. 7 & 8 : Text files

```
[stage01@slurm0 ~]$ file insulin.fas
insulin.fas: ASCII text, with CRLF line terminators
```

```
[stage01@slurm0 ~]$ file acteur.csv
acteur.csv: ASCII text, with CRLF line terminators
```

Ex. 9 : Files with data in a format unknown to the `file` command.

```
[stage01@slurm0 ~]$ file random.dat
random.dat: data
```

Examining the contents of a (text) file

The `cat` command displays the entire contents of a file.

```
[stage01@slurm0 ~]$ cat acteur.csv
First Name;Last Name;Age
Chuck;Norris;70
Sylvester;Stallone;64
Steven;Seagal;59
```

The `head` command displays the first (10) lines of a file. `head -n` displays the *n* first lines.

```
[stage01@slurm0 ~]$ head -2 acteur.csv
First Name;Last Name;Age
Chuck;Norris;70
```

The `tail` command displays the (10) last lines of a file. `tail -n` displays the *n* last ones.

```
[stage01@slurm0 ~]$ tail -2 acteur.csv
Sylvester;Stallone;64
Steven;Seagal;59
```

Interactively examining the contents of a (text) file

The `more` command displays the contents of a file “one page at a time”. The space bar moves from the current page to the next; and “q” is used to quit.

```
[stage01@slurm0 ~]$ more insulin.fas
>gi|163659904|ref|NM_000618.3| Homo sapiens insulin-like growth factor
1 (somatomedin C) (IGF1), transcript variant 4, mRNA
TTTTGTAGATAAATGTGAGGATTTTCTCTAAATCCCTCTTCTGTTTGCTAAATCTCACTGTCAC
TACTGCTAA
(...)
--More-- 4%
```

The `less` command also displays the contents of a file “one page at a time”. The spacebar moves from the current page to the next; and “q” is used to quit. The ↑ and ↓ arrows allow to move back and forth in the file.

```
[stage01@slurm0 ~]$ less insulin.fas
>gi|163659904|ref|NM_000618.3| Homo sapiens insulin-like growth factor
1 (somatomedin C) (IGF1), transcript variant 4, mRNA
TTTTGTAGATAAATGTGAGGATTTTCTCTAAATCCCTCTTCTGTTTGCTAAATCTCACTGTCAC
TACTGCTAA
(...)
insulin.fas
```

Both `more` and `less` allow to search for a text string in the file. This is done by typing slash (/) followed by the text to search for, and then typing **Enter** ↵ **Ex** : `/variant`

Searching for information in a file (1)

The `grep` command takes two arguments : a *pattern* and a *file name* ; it displays every line of the file containing the pattern.

```

[stage01@slurm0 ~]$ grep transcript insulin.fas
>q|163659904|ref|NM_000618.3| Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1),
transcript variant 4, mRNA
>q|163659900|ref|NM_001111284.1| Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1),
transcript variant 2, mRNA
>g|163659895|ref|NM_001111276.1| Mus musculus insulin-like growth factor 1 (Igf1), transcript
variant 5, mRNA
>g|163659893|ref|NM_001111275.1| Mus musculus insulin-like growth factor 1 (Igf1), transcript
variant 4, mRNA
>g|163659891|ref|NM_010512.4| Mus musculus insulin-like growth factor 1 (Igf1), transcript variant
1, mRNA
    
```

By default, `grep` is case sensitive. To override this behaviour, the `-i` (ignorecase) option can be used.

```

[stage01@slurm0 ~]$ grep TRANSCRIPT insulin.fas # returns nothing
[stage01@slurm0 ~]$ grep -i TRANSCRIPT insulin.fas
>q|163659904|ref|NM_000618.3| Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1),
transcript variant 4, mRNA
>q|163659900|ref|NM_001111284.1| Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1),
transcript variant 2, mRNA
(...)
    
```

Searching for information in a file (2)

`grep` also allows to count the lines matching the pattern with the `-c` (count) option.

```
[stage01@slurm0 ~]$ grep -c transcript insulin.fas  
5
```

As is the case for almost every command, options for `grep` can be combined.

```
[stage01@slurm0 ~]$ grep -c -i TRANSCRIPT insulin.fas  
5
```

By default, `grep` can also display the lines *not containing* the pattern, thanks to the `-v` (invert) option.

```
[stage01@slurm0 ~]$ grep -v -c -i TRANSCRIPT insulin.fas  
511
```

`grep` can be used to search for a pattern in all the files of a directory tree, with the `-r` (recursive) option. In this configuration, the second argument has to be the name of a directory. The lines with the information about the patterns are then prefixed with the filename to which they belong.

```
[stage01@slurm0 ~]$ grep -r -c -i TRANSCRIPT .  
./insulin_vs_nt.blast:144  
./acteur.csv:0  
./insulin.fas:5
```

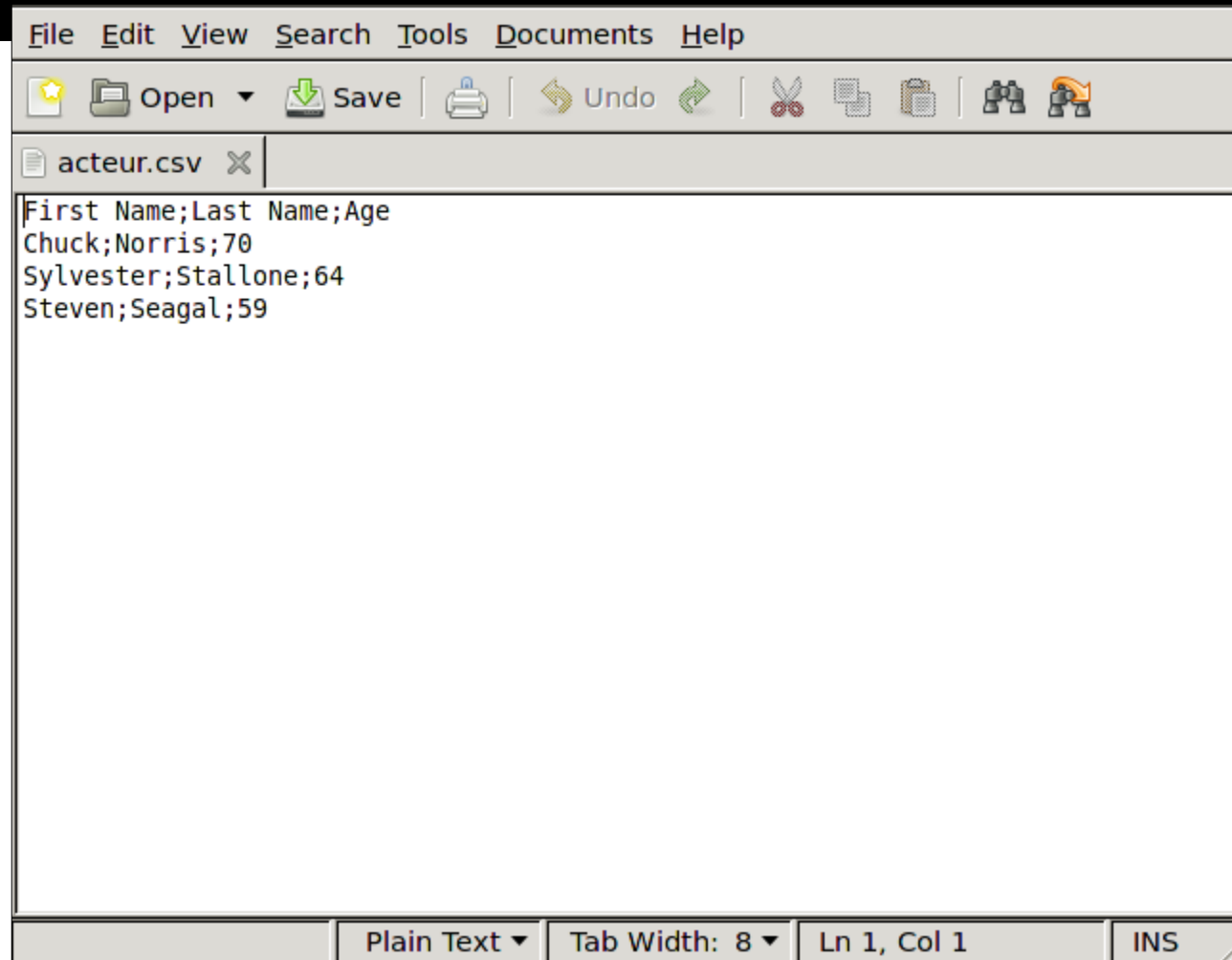
Find two ways of displaying the first line of the `acteur.csv` file (using two different commands)

Find two ways to display the last three lines of the `acteur.csv` file (using two different commands) [Granted, one is quite tricky.]

Changing the contents of a (text) file

The `gedit` command opens a window with a text editor.

```
[stage01@slurm0 ~]$ gedit acteur.csv
```



Changing the contents of a (text) file

The `nano` command opens a text editor in the window of the active session.

```
[stage01@slurm0 ~]$ nano acteur.csv
```

```
GNU nano 2.7.4 Fichier : acteur.csv
First Name;Last Name;Age
Chuck;Norris;70
Sylvester;Stallone;64
Steven;Seagal;59

[ Lecture de 5 lignes (converties du format DOS) ]
^G Aide      ^O Écrire    ^W Chercher  ^K Couper    ^J Justifier ^C Pos. cur.
^X Quitter   ^R Lire fich.^_ Remplacer  ^U Coller    ^T Orthograp.^_ Aller lig.
```

File management & Archiving

It is possible to determine the file size using the `-l` (`-h`) options of the `ls` command. The file size is displayed in the fifth column.

```
[stage01@slurm0 ~]$ ls -l insulin_vs_nt.blast
-rw-r--r-- 1 stage08 stage 30025889 May 03 22:42 insulin_vs_nt.blast
[stage01@slurm0 ~]$ ls -l -h insulin_vs_nt.blast
-rw-r--r-- 1 stage08 stage 29M May 03 22:42 insulin_vs_nt.blast
```

The `wc` (word count) command also displays information about the size of a file : line, word and character count. Using the `-l` option restricts the output to the number of lines.

```
[stage01@slurm0 ~]$ wc insulin_vs_nt.blast
622756 2377511 30025889 insulin_vs_nt.blast
[stage01@slurm0 ~]$ wc -l insulin_vs_nt.blast
622756 insulin_vs_nt.blast
```

File management & Archiving

To determine the size of a directory (and all its contents), the `du` (*disk usage*) command is used, preferably with the `-h` (*human readable*) option.

```
[stage01@slurm0 ~]$ du -h .  
64.0K  ./tmp  
4.0K   ./cours  
29M    .
```

Adding the `-s` (*summary*) option displays the total volume occupied by the directory (and its contents).

```
[stage01@slurm0 ~]$ du -s -h .  
29M    .
```

File management & Archiving

To determine how much disk space is available on a mount point (a disk partition), the `df` command is used (which also comes with an `-h` option).

```
[stage01@slurm0 ~]$ df -h
```

| Filesystem | Size | Used | Avail | Use% | Mounted on |
|----------------------------------|-------|------|-------|------|-------------------------|
| /dev/sda7 | 1008M | 750M | 208M | 79% | / |
| (...) | | | | | |
| /dev/sda1 | 504M | 152M | 327M | 32% | /boot |
| (...) | | | | | |
| /dev/sda2 | 32G | 541M | 30G | 2% | /tmp |
| /dev/sda5 | 16G | 9.5G | 5.5G | 64% | /usr |
| /dev/mapper/VolGroup00-LogVol100 | 4.0G | 3.0G | 795M | 80% | /usr/local |
| (...) | | | | | |
| brazil:/home/umr7139/mma | 247G | 210G | 25G | 90% | /home/umr7139/mmaucture |
| brazil:/home/umr7139/tccd | 591G | 538G | 24G | 96% | /home/umr7139/tccd |
| brazil:/home/umr7144/abice | 1.2T | 440G | 636G | 41% | /home/umr7144/abice |
| (...) | | | | | |

When specifying a directory as argument, `df` displays information about the mounted file system containing the directory.

```
[stage01@slurm0 ~]$ df -h /home/fr2424/sib/mhoebeke
```

| Filesystem | Size | Used | Avail | Use% | Mounted on |
|-------------------------|------|------|-------|------|------------------|
| brazil:/home/fr2424/sib | 2.0T | 1.6T | 242G | 88% | /home/fr2424/sib |
| (...) | | | | | |

File management & Archiving

Linux makes available different commands for compressing files, like `gzip` (older) or `bzip2` (better compression, a little less portable between systems). By default `gzip` replaces the file whose name is given as argument with its compressed version and adds `.gz` to the file name (`.bz2` for `bzip2`).

```
[stage01@slurm0 ~]$ gzip insulin_vs_nt.blast
[stage01@slurm0 ~]$ ls -l -h insulin_vs_nt.blast.gz
-rw-r--r-- 1 stage01 stage 5.0M May 13 20:42 insulin_vs_nt.blast.gz
```

```
[stage01@slurm0 ~]$ bzip2 insulin_vs_nt.blast
[stage01@slurm0 ~]$ ls -l -h insulin_vs_nt.blast.bz2
-rw-r--r-- 1 stage01 stage 3.2M May 13 20:42 insulin_vs_nt.blast.bz2
```

Decompression is achieved with `gunzip` (for `.gz` files) or `bunzip2` (for `.bz2` files).

```
[stage01@slurm0 ~]$ gunzip insulin_vs_nt.blast.gz
```

```
[stage01@slurm0 ~]$ bunzip2 insulin_vs_nt.blast.bz2
```

The compression ratio depends on the file contents : better for text files, quite low for already compressed data (images, sounds, videos).

File management & Archiving

Creating an archive containing several files is possible with the `tar` command.

Ex 1 : Creation of an archive with the contents of the `Linux-Initiation` directory: the `-c` option stands for “creation”, the `-f` option allows to specify the file name of the archive file name just afterwards. The final argument is the name of the directory from which to build the archive.

```
[stage01@slurm0 ~]$ tar -cf Linux-Initiation.tar Linux-Initiation
```

`tar` doesn't modify the directory to archive in any way.

Ex 2 : Extraction of all the files of the previously created archive: the `-x` option stands for “eXtraction”, the `-f` option has the same meaning as above. The `-v` option activates the “verbose mode” displaying each file name as it is extracted.

```
[stage01@slurm0 ~]$ tar -xvf Linux-Initiation.tar
Linux-Initiation/
Linux-Initiation/tmp/
(...)
```

Ex 3 : Listing the contents of an archive : the `-t` option (“toc”) displays the list of files in an archive.

```
[stage01@slurm0 ~]$ tar -tf Linux-Initiation.tar
Linux-Initiation/
(...)
```

Ex 4 : Extraction of a single file : the name of the file to be extracted is added as argument.

```
[stage01@slurm0 ~]$ tar -xvf Linux-Initiation.tar Linux-Initiation/insulin.fas
Linux-Initiation/insulin.fas
```

File management & Archiving

The `tar` command can be asked to carry out “on the fly” (de-)compression. To use `gzip` for this, the `-z` option is added. To use `bzip2`, the `-j` option can be used.

Ex1. : On the fly compression/decompression with `gzip`.

```
[stage01@slurm0 ~]$ tar -czf Linux-Initiation.tar.gz Linux-Initiation
[stage01@slurm0 ~]$ file Linux-Initiation.tar.gz
Linux-Initiation.tar.gz: gzip compressed data (...)
[stage08@slurm0 ~]$ tar -xzf Linux-Initiation.tar.gz
```

Ex1. : On the fly compression/decompression with `bzip2`.

```
[stage01@slurm0 ~]$ tar -cjf Linux-Initiation.tar.bz2 Linux-Initiation
[stage01@slurm0 ~]$ file Linux-Initiation.tar.bz2
Linux-Initiation.tar.bz2: bzip2 compressed data (...)
[stage01@slurm0 ~]$ tar -xjf Linux-Initiation.tar.bz2
```


Using shortcuts : symbolic links

It is often handy to be able to access a set of files in a single directory, even if they are originally scattered in several directories. And to do so **without copying any data**.

Use case :

- One of my directories, **allsequences** contains a myriad of sequence files related to various organisms.
- The file names hint to the organism to which the sequences they contain belong (**human_seq*.fasta**, **mouse_seq*.fasta**, **ecto_seq*.fasta**, etc.)
- I wish to apply routines to these sequences whose parameters may depend on the organism. And I wish to group the result files in organism specific (a.k.a separate) directories (**process_human/**, **process_mouse/**, **process_ecto/**, etc.)
- But I want to avoid copying the sequence data in these **process_*/** directories.

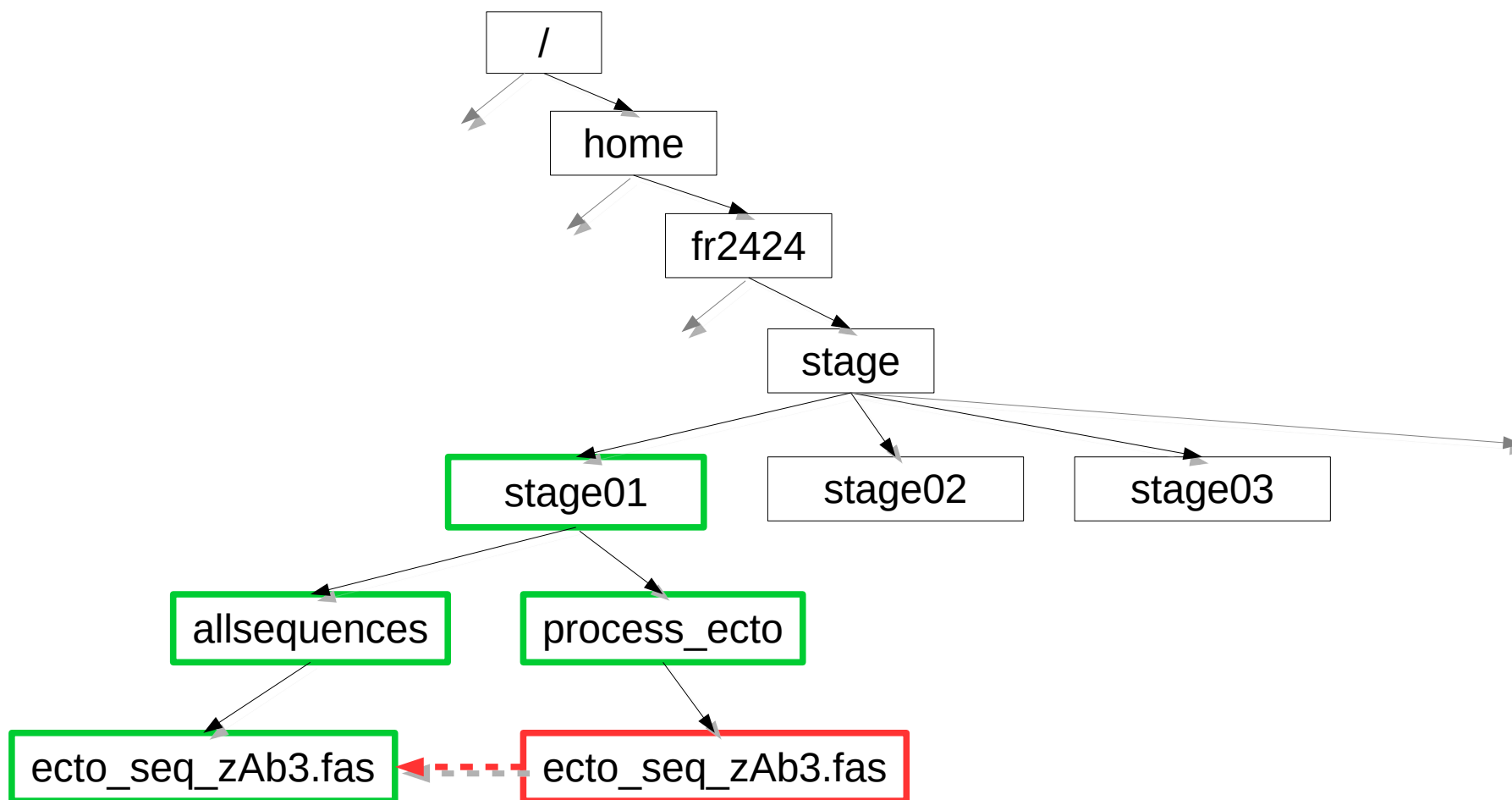
A solution relies on the creation, in the **process_*/** directories, of shortcuts to each sequence file belonging to a given organism, using the **ln -s** (*link, symbolic*) command.

Ex. 1: Creating a symbolic link for a single file : the first argument is the **name of the existing file** (or directory), and the second argument the name of the shortcut to create, or the **directory in which to create a shortcut** with the same name.

```
[stage01@slurm0 ~]$ ln -s allsequences/ecto_seq_zAb3.fas process_ecto/
[stage01@slurm0 ~]$ ls -l process_ecto/
lrwxrwxrwx 1 stage01 stage 30 May 05 08:44 ecto_seq_zAb3.fas ->
allsequences/ecto_seq_zAb3.fas
```

Using shortcuts : symbolic links

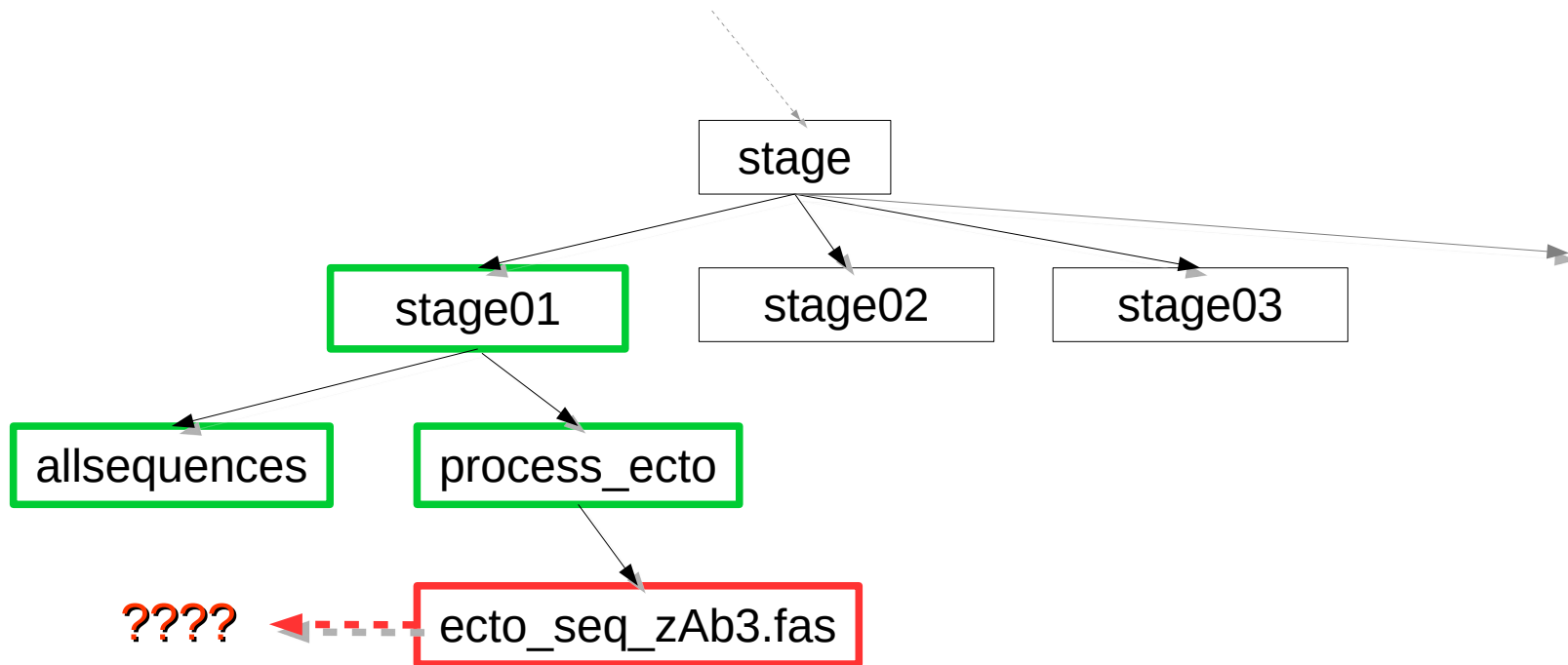
The symbolic link creates a new entry in the file system “pointing” to an already existing entry.



Using shortcuts : symbolic links

Deleting the original file makes the shortcut unusable!

```
[stage01@slurm0 ~]$ rm -f allsequences/ecto_seq_zAb3.fas  
[stage01@slurm0 ~]$ cat process_ecto/ecto_seq_zAb3.fas  
cat: process_ecto/ecto_seq_zAb3.fas: No such file or directory
```



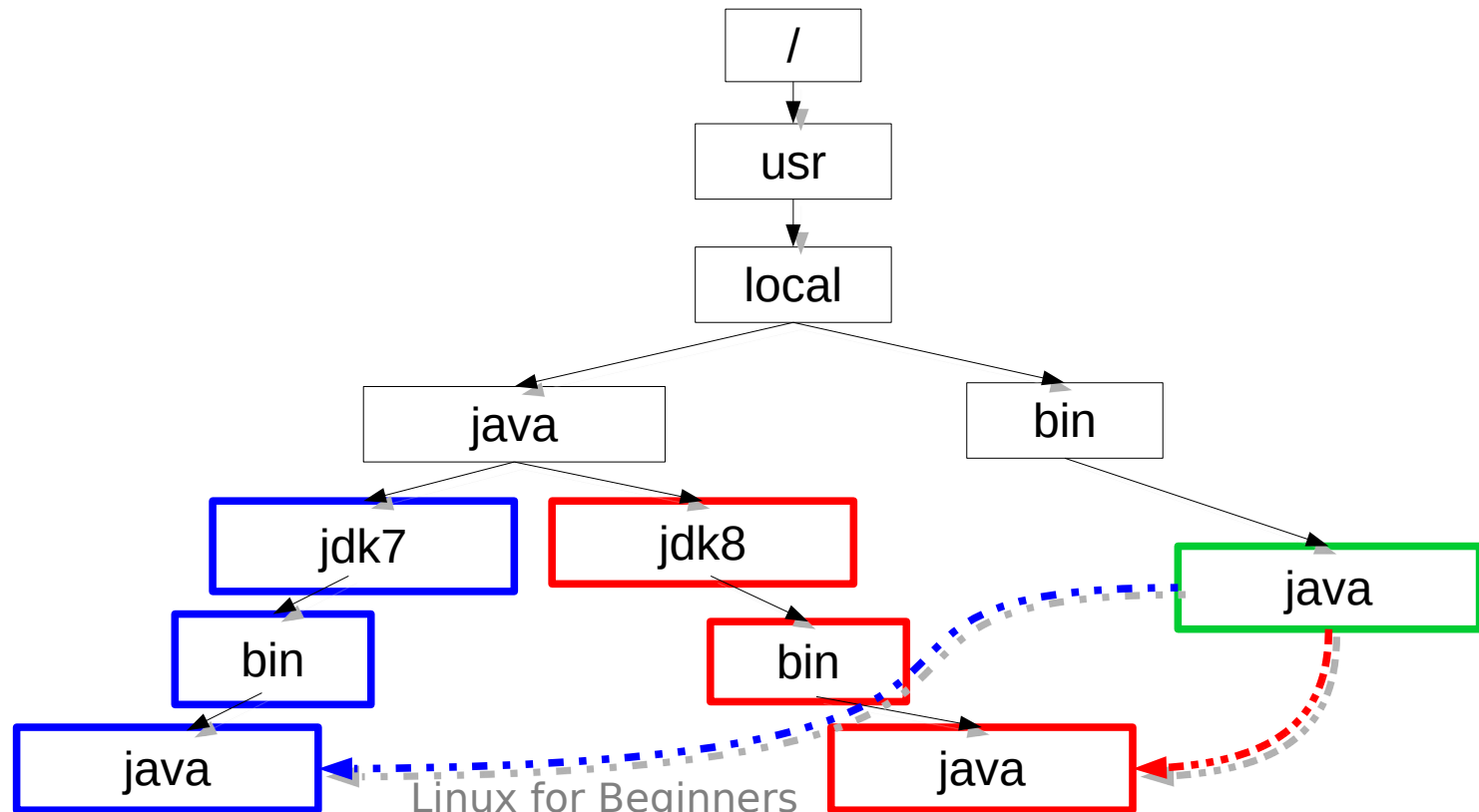
Using shortcuts : symbolic links

It is easy to create series of symbolic links using patterns.

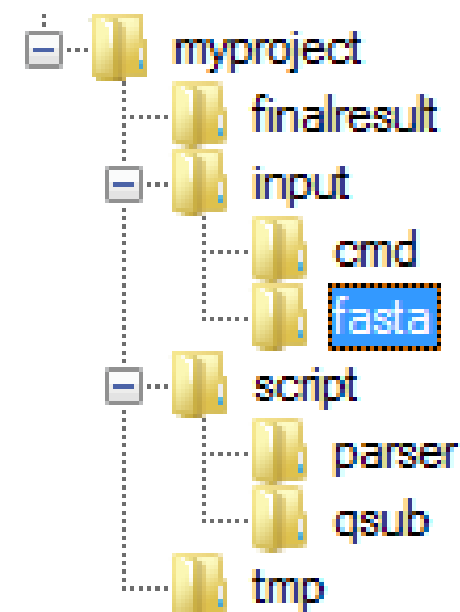
```

[stage01@slurm0 ~]$ ln -s allsequences/ecto_seq*.fas process_ecto/
[stage01@slurm0 ~]$ ln -s allsequences/mouse_seq*.fas process_mouse/
    
```

Symbolic links can also be used to transparently manage software package updates : a command can “point” to a specific version of a tool. When the tool is updated, a new version of the command can be installed alongside the previous one, and the link to the command is adjusted to point to the latest version.



- Create a symbolic link in your *home directory* pointing to the **script** directory.
- Copy the **acteur.csv** file to **finalresult/test-TP.txt**
- Create a symbolic link in your *home directory* pointing to the **test-TP.txt** file located in the in the **finalresult** directory.
- Display the contents of the **test-TP.txt** file located in your *home directory*
- Delete the **finalresult/test-TP.txt** file
- Conclusion ?



Copying Files to/from a Remote Machine

The `scp` command is used to copy files to/from another Linux (UNIX) machine. To use it, it's necessary to have an account (user name and password) on the remote machine. `scp` is used like `cp` sbut one of its arguments (source or destination) includes information about the remote machine as follows:

`username@hostname:`

- If needed, the password on the remote machine will be requested
- Transfers are encrypted: `scp` uses the SSH protocol.

Ex1. : Copying a local file to a remote machine : information about the remote machine is found in the **destination argument** of the command.

```
[stage01@slurm0 ~]$ scp Linux-Initiation.tar.gz stage01@sbr2:cours/
```

Ex2. : Copying a directory structure from a remote machine : the information about the remote machine is in the **source argument** of the command.

```
[stage01@slurm0 ~]$ scp -r stage01@sbr2:cours/ .
```

Copying files from a Web server through a URL

The `wget` command is used to fetch a local copy of a (set of) file(s) located on a Web server, and whose URL is known.

Ex1. : Fetching an ENA entry in FASTA format from its accession number.

```
[stage01@slurm0 ~]$ wget "  
http://www.ebi.ac.uk/ena/data/view/BN000065&display=fasta"  
--2017-04-14 12:42:09-- http://www.ebi.ac.uk/ena/data/view/BN000065&display=fasta  
Resolving www.ebi.ac.uk... 193.62.193.80  
Connecting to www.ebi.ac.uk|193.62.193.80|:80... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: unspecified [text/plain]  
Saving to: "BN000065&display=fasta"  
  
[ <=> ] 320,575 --.-K/s in 0.1s  
  
2017-04-14 12:42:09 (2.07 MB/s) - "BN000065&display=fasta" saved [320575]
```

Ex2. : **Recursively** (`-r` option) fetching contents from an HTML page (handle with care !)

```
[stage01@slurm0 ~]$ wget -r "http://abims.sb-roscoff.fr/training"
```

The remote directory structure is recreated in the directory where the `wget` command is run.

Supposing you are in your project directory, copy, **using a single command**, the `test-TP.txt` file located in the `/tmp` directory of the `ssh` machine in the `myproject/tmp` directory.

- 1 Purpose of an Operating System – why Linux ?
- 2 Establishing a connection and transferring files
- 3 The Command Line Interface
- 4 The File System
- 5 Manipulating File Contents
- 6 Users, Groups and Access Rights**
- 7 Processes

Users & Groups : the Concept

In a Linux (UNIX) system, each resource (file, directory, running program...) is owned by a user having a valid account on the machine. This user is the **owner** of the resource.

Every user belongs to at least one **group**. There is no limit to the number of groups a user can belong to.

At any moment, each user has only one **active** group. This is the group that will be taken into account by the system when the user accesses a resource.

Users and groups have numerical identifiers, called **uid** (user id) and **gid** (group id).

The `id` command displays the information about the identity of the user running the command.

```
[stage01@slurm0 ~]$ id  
uid=7069(stage01) gid=1013(stage) groups=1013(stage)
```

The `ls -l` command displays information about the ownership (which user/group) of files and directories.

```
[stage01@slurm0 ~]$ ls -l acteur.csv  
-rw-r--r-- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Owner

Group

Users & Groups : the Concept

The operations a user can perform on a resource are defined by the rights she has both as user and as member of groups to which she belongs.



- ➔ Every user is limited to what she has access to on the system: file read/write access, program execution, allocated disk or memory space.
- ➔ A single user has no access limitations: the system administrator or `root`

File system Related Access Rights

The `ls -l` command displays information on file/directory ownership and access rights. Access rights are grouped in triplets made of the `r`, `w`, `x` or `-` characters.

```
[stage01@slurm0 ~]$ ls -l acteur.csv  
-rw-r--r-- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

```
[stage01@slurm0 ~]$ ls -l acteur.csv  
-rw-r--r-- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Owner access rights (u : user)

```
[stage01@slurm0 ~]$ ls -l acteur.csv  
-rw-r--r-- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Group access rights (g : group)

```
[stage01@slurm0 ~]$ ls -l acteur.csv  
-rw-r--r-- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Access rights for users not members of the file's group (o : others)

File Access Right Management

| Position in the triplet | Character | Matching right | Character | Matching right |
|-------------------------|-----------|----------------------|-----------|---------------------|
| 1 | r | Read access allowed | - | No read access |
| 2 | w | Write access allowed | - | No write access |
| 3 | x | Execution allowed | - | Execution forbidden |

Example

-rw-r--r--

Owner

rw-

Read and write access allowed, execution forbidden

Group

r--

Read access, no writing and execution

Others

r--

Same as for the group

File system Related Access Rights

The `chmod` command is used to modify access rights to files and directories. Its first argument defines the access right modifications to apply. Its second argument defines the file(s) or directory(ies) on which to apply the modifications.

Ex. 1 : Make a file “private” a.k.a remove (using the minus sign -) all rights (letters r, w and x) to the group (letter g) and to others (letter o).

```
[stage01@slurm0 ~]$ chmod go-rwx acteur.csv  
-rw----- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Ex. 2 : Add (+ sign) write access (letter w) for the group (letter g) to a file.

```
[stage01@slurm0 ~]$ chmod g+w acteur.csv  
-rw-rw-r-- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Ex. 4 : Prevent file modification by removing (- sign) write (letter w) access to everyone (including the owner)

```
[stage01@slurm0 ~]$ chmod -w acteur.csv  
-r--r----- 1 stage01 stage 84 May 13 21:19 acteur.csv
```

Ex. 4 : Add (+ sign) execution rights (letter x) to a file.

```
[stage01@slurm0 ~]$ chmod +x monprogramme  
-rwxr-xr-x 1 stage01 stage 84 May 13 21:19 monprogramme
```

File system Related Access Rights - Directories

In the output of `ls -l`, directories are flagged with a `d` letter before the string defining the access rights.

```
[stage08@slurm0 ~]$ ls -l
drwxr-xr-x 4 stage08 stage 4096 May 01 11:41 Linux-Initiation
```

| Position in the triplet | Char. | Matching right | Char. | Matching right |
|-------------------------|----------|------------------------------------------------------------|-------|--------------------------------------------------|
| 1 | r | Reading <i>the list of files</i> is allowed. | - | Reading <i>the list of files</i> is forbidden |
| 2 | w | Creating, renaming and removing files is allowed. | - | File creation, renaming or removal are forbidden |
| 3 | x | Going (with <code>cd</code>) in the directory is allowed. | - | Cd'ing in the directory is forbidden |

The `chmod` command has an `-R` (recursive) option recursively applying the access rights to all files and subdirectories of its destination argument.

File system Related Access Rights – Group Changing

The `chgrp`, command allows to define a new group for a file or directory. The user running the command must be member of the group.

```
[stage01@slurm0 ~]$ chgrp  autregroupe  acteur.csv  
[stage01@slurm0 ~]$ ls -l  acteur.csv  
-r--r----- 1 stage01 autregroupe 84 May 13 21:19 acteur.csv
```

There is a command to change file ownership (`chown`) but its use is restricted to the system administrator...

Authorize all members of the `tps_XXXXXX` group to write in the **Linux-Initiation** directory and its subdirectories. Check with your neighbor that (s)he can deposit and remove files there. But forbid any modification to the `acteur.csv` file.

- 1 Purpose of an Operating System – why Linux ?
- 2 Establishing a connection and transferring files
- 3 The Command Line Interface
- 4 The File System
- 5 Manipulating File Contents
- 6 Users, Groups and Access Rights
- 7 Processes**

Some Definitions

A process is a currently running program. Each time a user issues a command (runs a program), the operating system loads it into memory and starts its execution.

In order to run smoothly, a process needs **memory** and **processor time** (CPU). It is the duty of the operating system to proceed to the optimal allocation of these resources among all the processes running “simultaneously”. The **load** of a machine reflects the activity of all the active processes at any given moment.

As for files, a process has an **owner** (user) and a **group** ; and associated **rights or permissions**.

The user has the ability -to some extent- to control process' execution: she can **stop** them “by force”, **interrupt them** to **resume** them later on, or modify their **priority**.

Running / Stopping / Interrupting processes

Each time a command is issued in the current session, a process is created and executed. Only when the execution ends is it possible to issue new commands.

```
[stage01@slurm0 ~]$ gedit acteur.csv  
[stage01@slurm0 ~]$
```

New commands can be run only after exiting gedit

A process running in the current session can be (brutally) stopped by typing the **Ctrl-C** key combination.

```
[stage01@slurm0 ~]$ gedit acteur.csv  
^C  
[stage01@slurm0 ~]$
```

Ctrl-C “kills” the process. The user regains control.

A process killed by **ctrl-c** frees all the resources (memory, open files) in its possession.

Running / Stopping / Interrupting processes

A process running in the current session can be **interrupted** with the key **Ctrl-Z** combination.

```
[stage01@slurm0 ~]$ gedit acteur.csv
^Z
[1]+  Stopped                  gedit acteur.csv
[stage01@slurm0 ~]$
```

The user regains control after typing **Ctrl-Z**. The process has been **interrupted**.

An interrupted process is “frozen”, and keeps all resources it was allocated (except for processor time). It is assigned a **job identifier** (*not to be confused with the process identifier, cf. following slides*).

The **jobs** command lists all interrupted processes of the current session.

```
[stage01@slurm0 ~]$ gedit insulin.fas
^Z
[2]+  Stopped                  gedit insulin.fas
[stage01@slurm0 ~]$ jobs
[1]-  Stopped                  gedit acteur.csv
[2]+  Stopped                  gedit insulin.fas
```

Resuming an Interrupted Process

The `fg` (*foreground*) command resumes the execution of an interrupted process. Without argument, the most recently interrupted process will be resumed. To resume a specific process, it is possible to use the `%` sign followed by the job identifier.

```
[stage01@slurm0 ~]$ jobs
[1]-  Stopped          gedit acteur.csv
[2]+  Stopped          gedit insulin.fas
[stage01@slurm0 ~]$ fg %1
gedit acteur.csv
```

The `bg` (*background*) command also resumes an interrupted process but **immediately gives back control to the user in in the current session**. Execution of the process continues in the **background**

```
[stage01@slurm0 ~]$ jobs
[1]-  Stopped          gedit acteur.csv
[2]+  Stopped          gedit insulin.fas
[stage01@slurm0 ~]$ bg %2
[2]+ gedit insulin.fas &
[stage01@slurm0 ~]$
```

Running a process in background mode

A process can be run directly in background mode by adding an ampersand (&) to the command line.

```
[stage01@slurm0 ~]$ gedit acteur.csv &  
[3] 26357  
[stage01@slurm0 ~]$
```

Both a **job identifier** and a **process identifier (PID)** are displayed .

Displaying process information : ps

The **ps** (process status) command is used to display more or less detailed information about processes.

Ex. 1 : Listing the processes in the current session.

```

[stage01@slurm0 ~]$ ps
  PID TTY          TIME CMD
 16175 pts/14    00:00:00 bash
 20693 pts/14    00:00:00 dbus-launch
 26357 pts/14    00:00:00 gedit
 27505 pts/14    00:00:00 ps
    
```

The main information are the PID (process identifier) and the name of the command (CMD).

Ex. 2 : Getting the **detailed** list of the processes in the current session with the **-f** (*full*) option.

```

[stage01@slurm0 ~]$ ps -f
UID          PID  PPID  C  STIME TTY          TIME CMD
stage01    16175 16174  0  17:21 pts/14    00:00:00 -bash
stage01    20693     1  0  17:37 pts/14    00:00:00 dbus-launch --autolaunch 9b7328b
stage01    26357 16175  0  17:58 pts/14    00:00:00 gedit insulin.fas
stage01    28638 16175  4  18:06 pts/14    00:00:00 ps -f
    
```

Additional information is shown about the user (UID), the PID of the parent process (PPID), the start time (STIME), the execution time (TIME) and the complete command line (CMD).

Displaying process information : ps

Ex. 3 : Listing the processes of a specific user with the `-u` (*user*) option.

```

[stage01@slurm0 ~]$ ps -fu stage01
UID          PID    PPID  C  STIME TTY          TIME CMD
stage01    16174  16172  0  17:21 ?            00:00:00 sshd: stage08@pts/14
stage01    16175  16174  0  17:21 pts/14      00:00:00 -bash
(...)
stage01    20696     1    0  17:37 ?            00:00:00 /usr/libexec/gconfd-2
stage01    26357  16175  0  17:58 pts/14      00:00:00 gedit insulin.fas
stage01    32327  16175  0  18:20 pts/14      00:00:00 ps -fu stage08
    
```

Ex. 4 : Listing all the processes currently present on the machine with the `-e1f` (*extended, long, full*) options.

```

[stage01@slurm0 ~]$ ps -e1f
0 S  uguyet    29185 29184  0  80    0 - 28353 n_tty_ Apr19 pts/35    00:00:00 /bin/bash
0 S  lgueguen  30113 10845  0  80    0 - 26364 n_tty_ May12 pts/32    00:00:00 less macros.xml
4 S  root      30980  5864  0  80    0 - 28926 unix_s  May12 ?        00:00:00 sshd: nhenry
[priv]
5 S  nhenry    31153 30980  0  80    0 - 28961 poll_s  May12 ?        00:00:01 sshd: nhenry@notty
0 S  nhenry    31154 31153  0  80    0 - 14977 poll_s  May12 ?        00:00:01
/usr/libexec/openssh/sft
0 S  pmandon   31370  2698  0  80    0 - 27641 n_tty_ May12 pts/64    00:00:00 /bin/bash
0 S  lberdjeb  31598     1    0  80    0 - 26526 wait   12:22 ?        00:00:00 /bin/sh
/opt/sge/qlogin.
    
```

Interactive Process Visualisation with top

The `top` command displays and continually refreshes the list of processes running on a machine. By default, the list is sorted according to the process load (%CPU column). The command also summarizes the overall state of the system (uptime, global load, memory availability) above the process list.

```

top - 18:31:24 up 171 days, 5:52, 4 users, load average: 1.19, 1.16, 1.1
Tasks: 650 total, 2 running, 648 sleeping, 0 stopped, 0 zombie
Cpu(s):  0.0%us,  0.0%sv,  3.1%ni, 96.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%
Mem: 131915052k total, 130711176k used, 1203876k free, 203212k buffers
Swap: 1048572k total, 220144k used, 828428k free, 128719136k cached

  PID USER      PR  NI  VIRT  RES  SHR  S  %CPU  %MEM    TIME+  COMMAND
 20471 roze     39   19 233m 223m 1300  R 100.0   0.2   11778:50 multi
     1 root      20    0 23500 1252 1044  S   0.0   0.0     0:02.36 init
     2 root      20    0     0     0     0  S   0.0   0.0     0:20.70 kthreadd
     3 root      RT    0     0     0     0  S   0.0   0.0     2:38.13 migration/0
     4 root      20    0     0     0     0  S   0.0   0.0     0:06.98 ksoftirqd/0
     5 root      RT    0     0     0     0  S   0.0   0.0     0:00.00 stopper/0
     6 root      RT    0     0     0     0  S   0.0   0.0     0:12.25 watchdog/0
     7 root      RT    0     0     0     0  S   0.0   0.0     1:53.64 migration/1
     8 root      RT    0     0     0     0  S   0.0   0.0     0:00.00 stopper/1
     9 root      20    0     0     0     0  S   0.0   0.0     0:09.10 ksoftirqd/1
    10 root      RT    0     0     0     0  S   0.0   0.0     0:12.57 watchdog/1
    
```

Global system load over the last 1, 5 and 15 minutes

Snapshot of CPU usage

Snapshot of available and allocated memory

Process termination : `kill`

The `kill` command sends a signal to the process whose process identifier (PID) is given as argument. By using specific options (`-HUP`, `-TERM`, `-KILL`) the process is more or less “gently” notified.

Ex. 1 : Brutally stopping a process using `kill` with the `-KILL` option

```
[stage01@slurm0 ~]$ gedit acteur.csv &  
[1] 27908  
[stage01@slurm0 ~]$ kill -KILL 27908  
[stage01@slurm0 ~]$
```

The `kill` command can only be used on one's own processes.

Processes and Inheritance

Every process is spawned from a parent process. For instance, open a session yields a new process which will be the parent process of all of the commands typed in on the command line.

The process with PID 1 is the process that, when the machine was started, gave rise to the system's process tree.

Halting a process results in halting all of its child processes (to keep in mind before using `kill`).

When a process ends, its parent process is notified. Until the parent process handles its child's termination, the latter stays in a *zombie* state. Zombie processes do not take up any system resources, except if they proliferate in an uncontrollable way. Eliminating a zombie is done by "killing" its parent process. It is then attached to the process with PID 1 who takes care of eliminating zombies.

Open a connection. Run the **gedit** command in background mode. Then close the connection. What happens ?

Detaching Processes from a Session

It is frequently necessary to execute programs whose running time will exceed the duration of a session. Thus, automatic killing of these processes must be avoided when the session is closed. The **nohup** (*no hang-up*) command allows to detach a process from a session.

```
[stage01@slurm0 ~]$ nohup mon_long_programme_de_bioinformatique.pl  
nohup: ignoring input and appending output to `nohup.out'
```

Any information displayed by the program will be added to a file named **nohup.out** created in the directory where the command was executed.

For programs already running (in background mode) and whose PID is known, the **disown** command allows to detach it from the session.

```
[stage01@slurm0 ~]$ mon_long_programme_de_bioinformatique.pl &  
[1] 29087  
[stage08@slurm0 ~]$ disown 29087
```

Bonus : Customizing Your Environment

The Session's Environment - SHELL

Remember : each command is executed un a **current directory** with the credentials of the **current user**.

When starting a session (local or remote), the process handling what the user types is called the **command interpreter** or **shell**. Its purpose is to wait for the user to hit the Enter key at the end of a command line a to try to make sense of it a.k.a run the commands whose name(s) were typed in.

The **shell** is highly configurable.

Environment Variables - PATH

When the user types a command name, the shell looks for a matching (binary) file in a well defined list of directories. This list is stored in an

environment variable

This specific variable is called **PATH**, and its value can be displayed with **echo** :

```
[stage01@slurm0 ~]$ echo $PATH
```

```
/opt/sge/bin/lx24-amd64:/opt/python/bin:/usr/local/java/bin:/usr/lib64/qt-3.3/bin:/usr/kerberos/sbin:/usr/kerberos/bin:/usr/local/bin:/bin:/usr/bin:/usr/local/sbin:/usr/sbin:/sbin:/opt/dell/srvadmin/bin:/usr/local/public/bin:/usr/local/genome2/bin:/usr/local/genome/bin:/usr/local/adm/bin:/usr/local/admin/script:/usr/local/adm/script:/usr/local/genome/script:/usr/local/genome2/h:/usr/local/genome2/seqclean:/usr/local/genome2/seqclean/bin:/usr/local/genome2/tgicl_linux:/usr/local/genome2/tgicl_linux/bin:/usr/local/genome/emboss/bin:/usr/local/genome/phylip/bin:/usr/local/genome/mgadist:/usr/local/genome/MUMmer:/usr/local/genome/TMHMM/bin:/usr/local/genome/hmmer/bin:/usr/local/genome/fastq/bin:/usr/local/genome/mcl64/mcl-05-321/bin:/usr/local/genome/WoLFPSORT_package_v0.2/bin:/usr/local/genome2/abyss/bin:/usr/local/cristallo/bin:/home/fr2424/stage/stage08/bin:/opt/openmpi/bin:/opt/6.x/matlab/r2013b/bin:/opt/6.x/matlab/r2013b/toolbox/abims/ffca:/opt/6.x/matlab/r2013b/toolbox/abims/mexcdf:/usr/local/cristallo/shelx97:/home/fr2424/stage/stage01/bin
```

Command Lookup Customization

To find out in which of the directories included in the `PATH` a command is actually found, the `which` command can be used :

```
[stage01@slurm0 ~]$ which grep
/bin/grep
[stage01@slurm0 ~]$ which java
/usr/local/java/bin/java
```

To run a command not located in one of the `PATH` directories, its (absolute or relative) path must be given :

```
[stage01@slurm0 ~]$ ls -l myproject/script
-rwxr-xr-x 1 stage01 stage 804 May 4 20:58 myproject/script/supercalcul.sh
[stage01@slurm0 ~]$ supercalcul.sh
-bash: supercalcul.sh: command not found
[stage01@slurm0 ~]$ myproject/script/supercalcul.sh
```

New directories can be added to the `PATH` to allow executing new commands without having to specify their location.

```
[stage01@slurm0 ~]$ export PATH=~ /myproject/script :${PATH}
[stage01@slurm0 ~]$ supercalcul.sh
```

export Propagates the new value of `PATH` to all subsequent processes in the same session

PATH= Means that a new value will be assigned to `PATH`

~/myproject/script The new directory that will be added (at the start) of `PATH`

:\${PATH} The previous value of `PATH` is appended to the new directory.

Using Aliases

It is often handy to avoid repeatedly typing commands with the same options & arguments to define an *alias* for them. This can be done with the `alias` command.

Ex. 1 : Redefining `grep` to display matches in color.

```
[stage01@slurm0 ~]$ alias grep='grep --color'
[stage01@slurm0 ~]$ grep Homo insulin.fas
>gi|163659904|ref|NM_000618.3| Homo sapiens insulin-like growth factor 1 (somatomedin
C) (IGF1), transcript variant 4, mRNA
>gi|163659900|ref|NM_001111284.1| Homo sapiens insulin-like growth factor 1
(somatomedin C) (IGF1), transcript variant 2, mRNA
```

Ex. 2 : Creating an alias to count “human” sequences in a multi-FASTA file.

```
[stage01@slurm0 ~]$ alias hsc='grep -c -i human'
[stage01@slurm0 ~]$ hsc insulin_vs_nt.blast
373
```

Making the Customizations Persistent

Environment variable changes as well as alias definitions are only valid for the duration of the session they were made.

There is however a `~/ .bashrc` (text) file whose contents is read whenever a new session is opened. The contents of this file can be any shell command, including `alias` definitions and `PATH` modifications.

After modifying this file in the current session, the `source` command has to be issued for the new version of the file contents to be taken into account in this session :

```
[stage01@slurm0 ~]$ source ~/.bashrc
```

Displaying Environment Parameters

The `env` command displays all the known environment variables with their values.

```
[stage01@slurm0 ~]$ env
(...)
LANG=fr_FR.UTF-8
GDM_LANG=fr_FR
MANAGERPID=25591
DISPLAY=:0
INVOCATION_ID=964f88f33972481e93c31234b2e4483f
COMPIZ_CONFIG_PROFILE=ubuntu
(...)
```

The `alias` command displays the list of all active aliases.

```
[stage01@slurm0 ~]$ alias
(...)
alias l='ls -CF'
alias la='ls -A'
alias ll='ls -alF'
alias ls='ls --color=auto'
(...)
```

Add a customized version of `grep` to your environment, and reload your configuration in the current session.



After handing filling the evaluation form at http://abims.sb-roscoff.fr/evaluation_formation, you will receive a *cheat sheet* with the essential Linux commands.