

# Discovering the linux world through jupyter



Christine Tranchant-Dubreuil and François Sabot

# The objectives

Run your own analysis using jupyter book and Linux !



After this course, you will be able to :

- Know the main Linux commands
- Move into the Linux file tree : `pwd` , `ls` , `cd` , `mkdir`
- Work with text files: `head` , `tail` , `wc` , `grep`
- Chain and combine commands
- Run programs from the command line

# What is Linux ?

An operating system well known for :

- its security and stability
- its frequent updates
- its (no) fees and openSource (mostly) softwares

**Created in 1991 by Linus Torvalds**

Linux source code **opensource** and **free** : copy, modify, redistribute



# What is Linux ?

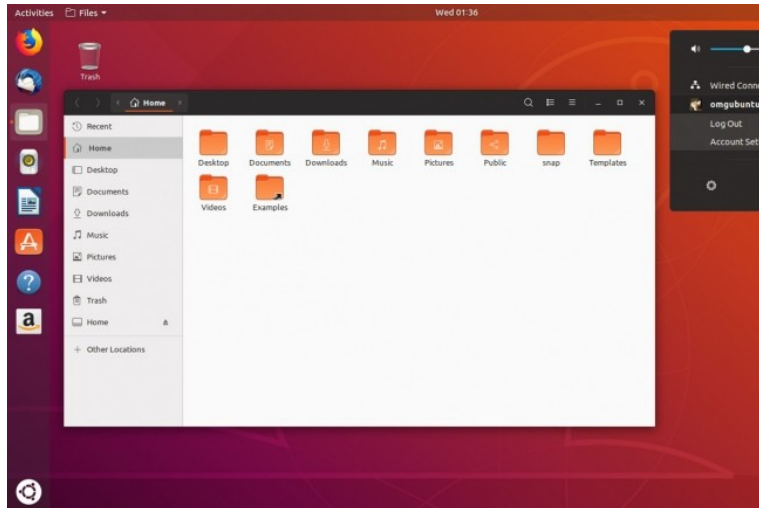


- ***Robust et multi-plateform OS*** : computer, server, android...
- ***Multi-users system*** : Several users can work simultaneously
- ***Multi-tasking system (processes/programs)*** Every user can run several programs at the same time

# How to use Linux ?

- 2 classical ways :

## Graphical User Interface



# How to use Linux ?

- 2 classical ways :

## Graphical User Interface

## Command-Line Interface through a terminal

```

root@localhost:~# ss -s -t -a -i www.wikipedia.org
NG text.pmpa.wikipedia.org (208.80.152.2) 56(84) bytes of data.
...
root@localhost:~# ss -s -t -a -i www.wikipedia.org
NG text.pmpa.wikipedia.org ping statistics ---
 0 packets transmitted, 1 received, 0% packet loss, time 0ms
 rtt min/avg/med = 549.528/549.528/549.528/0.698 ms
root@localhost:~# cd /var
root@localhost:~# cd /var
root@localhost:var# ls -la
total 72
drwxr-xr-x, 18 root root 4096 Jul 30 22:04 .
drwxr-xr-x, 23 root root 4096 Sep 14 20:42 ..
drwxr-xr-x, 2 root root 4096 May 14 00:15 account
drwxr-xr-x, 11 root root 4096 Jul 31 22:26 cache
drwxr-xr-x, 3 root root 4096 May 18 16:03 db
drwxr-xr-x, 2 root root 4096 May 18 16:03 empty
drwxr-xr-x, 2 root root 4096 May 18 16:03 games
drwxr-xr-x, 2 root gdm 4096 Jun 2 18:39 gdm
drwxr-xr-x, 36 root root 4096 May 18 16:03 lib
drwxr-xr-x, 2 root root 4096 May 18 16:03 local
drwxrwxrwx, 1 root root 11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x, 14 root root 4096 Sep 14 20:42 log
drwxrwxr-x, 1 root root 10 Jul 30 22:42 mail -> spool/mail
drwxr-xr-x, 2 root root 4096 May 18 16:03 nis
drwxr-xr-x, 2 root root 4096 May 18 16:03 opt
drwxr-xr-x, 2 root root 4096 May 18 16:03 preserve
drwxr-xr-x, 2 root root 4096 Jul 1 22:11 report
drwxrwxr-x, 1 root root 6 May 14 00:12 run -> ../run
drwxr-xr-x, 14 root root 4096 May 18 16:03 spool
drwxrwxrwt, 4 root root 4096 Sep 12 23:50 tmp
drwxr-xr-x, 2 root root 4096 May 18 16:03 yp
root@localhost:var# yum search wki
advised-plugins lampstacks, presto, refresh-packagekit, remove-with-leaves
updates/free-updates/primary_db 2.7 kB 00:00
updates/free-updates/primary_db 296 kB 00:04
updates/non-free-updates 2.7 kB 00:00
updates/retail/updates 5.9 kB 00:00
updates 4.7 kB 00:00
updates/primary_db 73% 62 kB/s | 2.6 MB 00:15 ETA

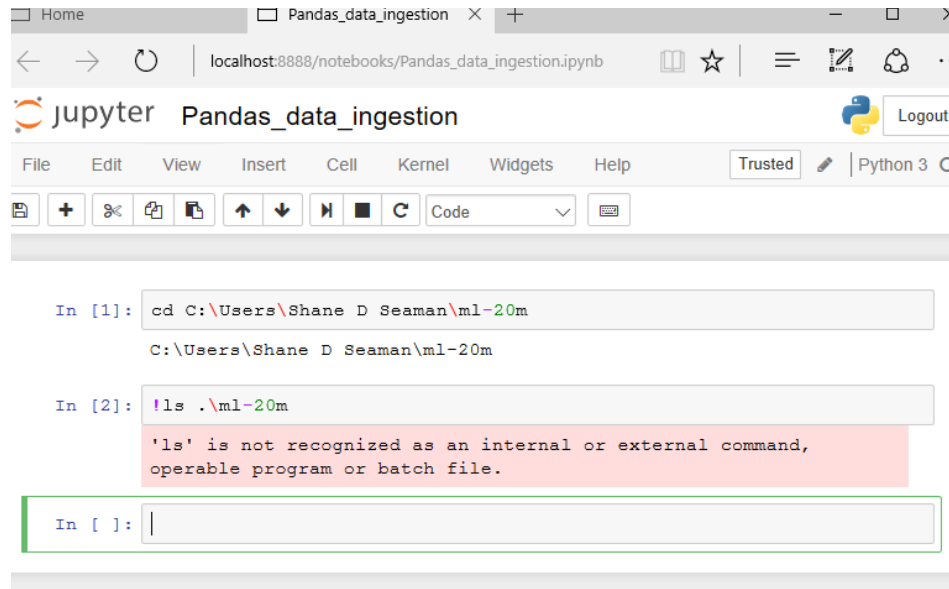
```



# How to use Linux ?

- a new and alternative ways :

## jupyter book

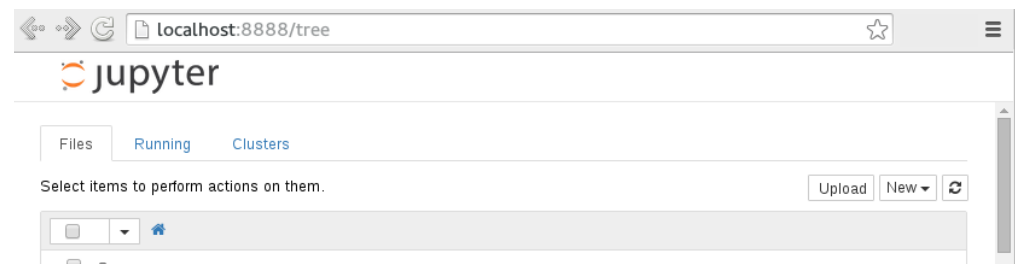


The screenshot shows a Jupyter Notebook window titled 'Pandas\_data\_ingestion'. The browser address bar shows 'localhost:8888/notebooks/Pandas\_data\_ingestion.ipynb'. The notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar. The code cell contains the following text:

```
In [1]: cd C:\Users\Shane D Seaman\ml-20m
C:\Users\Shane D Seaman\ml-20m

In [2]: !ls .\ml-20m
'ls' is not recognized as an internal or external command,
operable program or batch file.
```

The error message indicates that the 'ls' command is not recognized, which is a common mistake when using Windows command prompt instead of Linux terminal.





<input type="checkbox"/>	 figures
<input type="checkbox"/>	 Tutorial_Evoked_Responses_In_The_Visual_Cortex.ipynb
<input type="checkbox"/>	 anatomy_of_a_region_simulation.ipynb
<input type="checkbox"/>	 anatomy_of_a_surface_simulation.ipynb
<input type="checkbox"/>	 connectivity_normalisation_modes.ipynb
<input type="checkbox"/>	 corrcoeff_analyse_view_region.ipynb
<input type="checkbox"/>	 display_sensors_locations.ipynb
<input type="checkbox"/>	 display_subject_dataset.ipynb
<input type="checkbox"/>	 display_surface_local_connectivity.ipynb

# Why using Linux ?

- Numerous fast and powerful programs
- Easy to link commands and programs (workflow)
- Numerous bioinformatics softwares available
- 90% of servers on Linux

# Why using Linux ?

Need to practice!!!

**Important investments to have good results !!!**



# Let's go through jupyter

IFB Cloud x mydatalocal/ x hello-day1 - Jupyter No... x myFirstJupyterBook - Ju... x

https://134.158.247.8/tree/mydatalocal 133 %

jupyter Quit Logout

Files Running Clusters

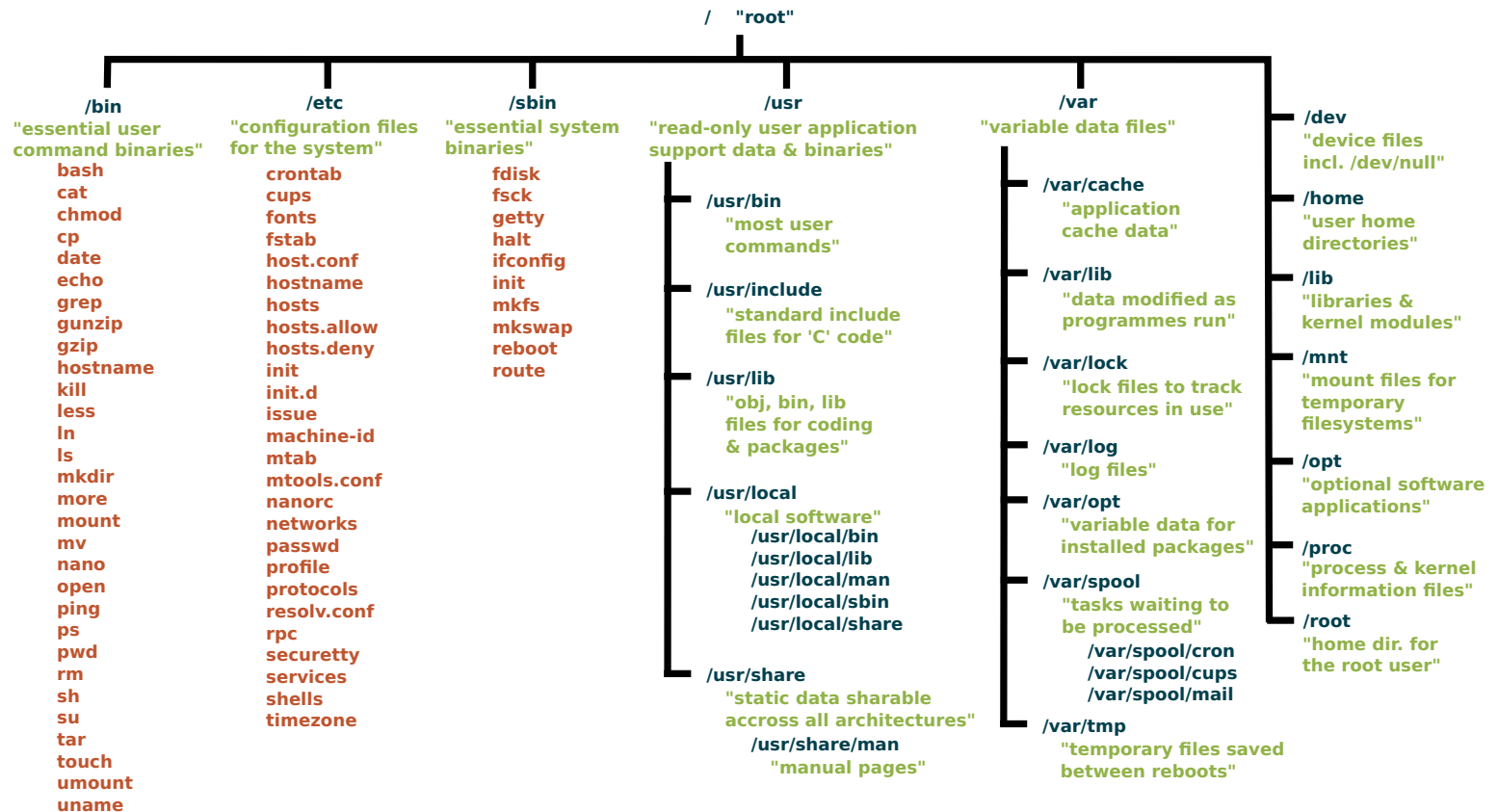
Select items to perform actions on them. Upload New Refresh

<input type="checkbox"/>	0	/ mydatalocal	Name ↓	Last Modified	File size
<input type="checkbox"/>		..		il y a quelques secondes	
<input type="checkbox"/>		fastqDirLR		il y a une minute	
<input type="checkbox"/>		fastqDirSR		il y a une minute	
<input type="checkbox"/>		hello-day1.ipynb	Actif	il y a quelques secondes	3.81 kB
<input type="checkbox"/>		linuxLessonPractical.ipynb		il y a 21 heures	5.61 kB
<input type="checkbox"/>		myFirstJupyterBook.ipynb	Actif	il y a 3 minutes	1.73 kB
<input type="checkbox"/>		reference.fasta		il y a une minute	1.04 MB

1079760dd85a9bbc43e10354 Tout surligner Respecter la casse Respecter les accents et diacritiques Mots entiers Expression non trouvée

# First steps on Linux

Commands for *moving around the file system* and *manipulating files/folders*



# Linux command syntax

`command [ -options ] [ arguments ]`

# Your first command *pwd*

**pwd** print the name of the current directory (the full path)

# Your first command *pwd*

**pwd** print the name of the current directory (the full path)

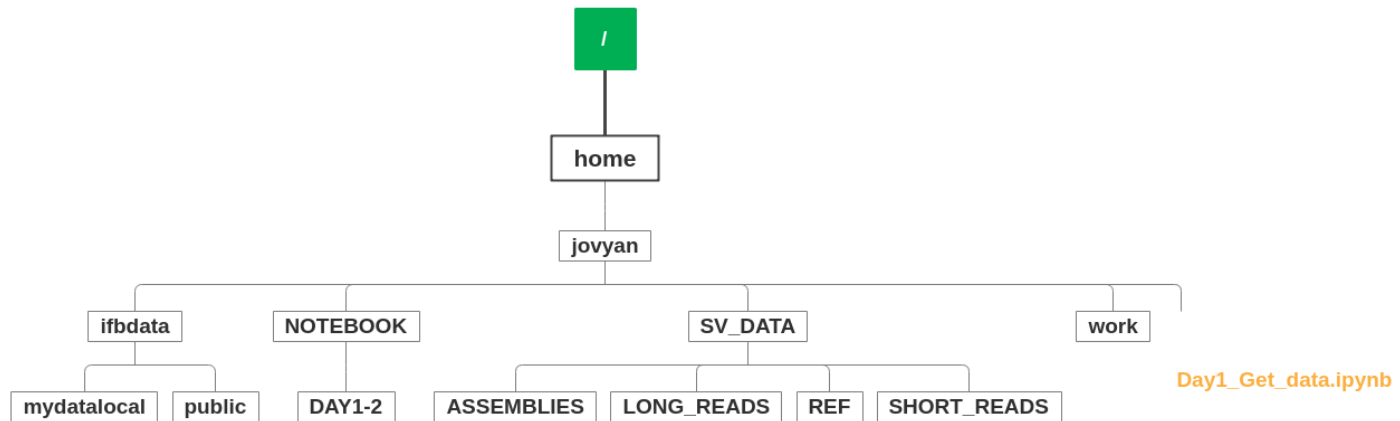
In [ ]:



# Your first command *pwd*

**pwd** print the name of the current directory (the full path)

In [ ]: `pwd`



# 2nd command `/s`

- without argument : list all the files in the current directory (by default)

# 2nd command `ls`

- without argument : list all the files in the current directory (by default)

In [ ]: `ls`

# ls + option -l

ls -l : *list files with more information about each file* (long)

# ls + option -l

ls -l : *list files with more information about each file* (long)

In [ ]: `ls -l`

# 2nd command + argument

## */s*

with directory path as argument

# 2nd command + argument

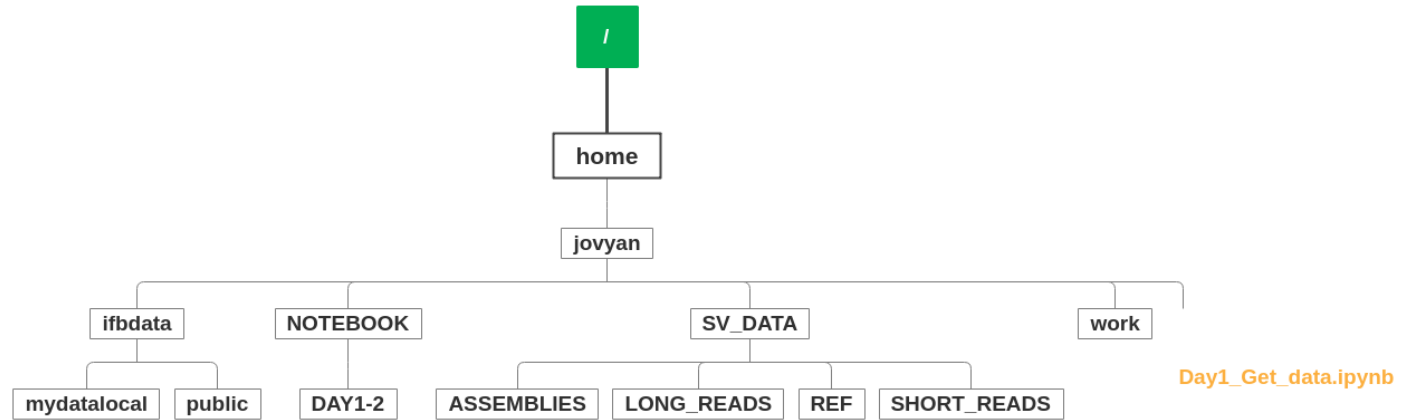
## */s*

with directory path as argument

```
In [ ]: ls /home/jovyan/SV_DATA/REF
```



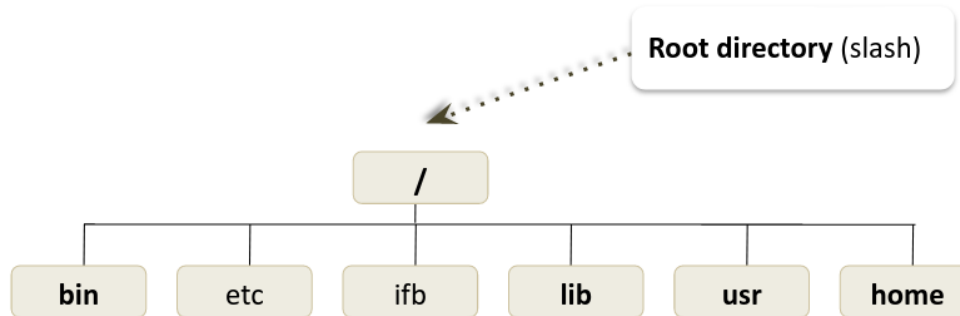
- What is a file tree for Linux?



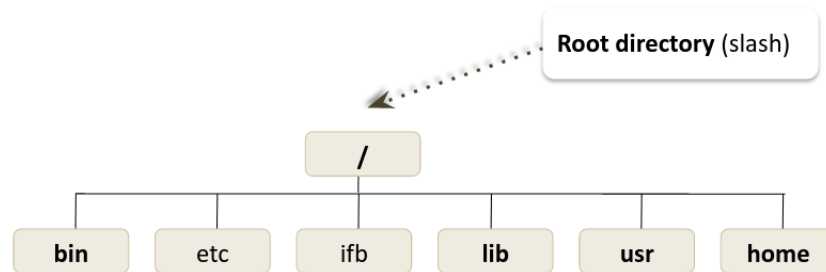


# The file tree - file system

- Directory structure starts at the root directory called “/” (slash)



# The file tree - Main directories



<b>/bin</b>	<b>Main commands, shell, programs</b>
<b>/usr, /opt</b>	<b>Applications and user librairies</b>
<b>/usr/bin</b>	<b>Other commands</b>
<b>/home</b>	<b>User directory (one per user, name= login)</b>

# Path of a file



- Path = location to a file/directory in the file system
- On linux filetree, you can use absolute or relative path
- What is the difference between the absolute and relative path ?

# What is an absolute path of a file or a directory ?

- Path = location to a file/directory in the file system

## **ABSOLUTE**

---

Complete path of a file  
starting from the root directory /

---

starts always with /

---

always good wherever user is working

# What is a relative path of a file or a directory ?

- Path = location to a file/directory in the file system

## **RELATIVE**

---

Path related to the present working directory  
(where the user is working)

---

Never starts with /

---

Depends on where the user is working

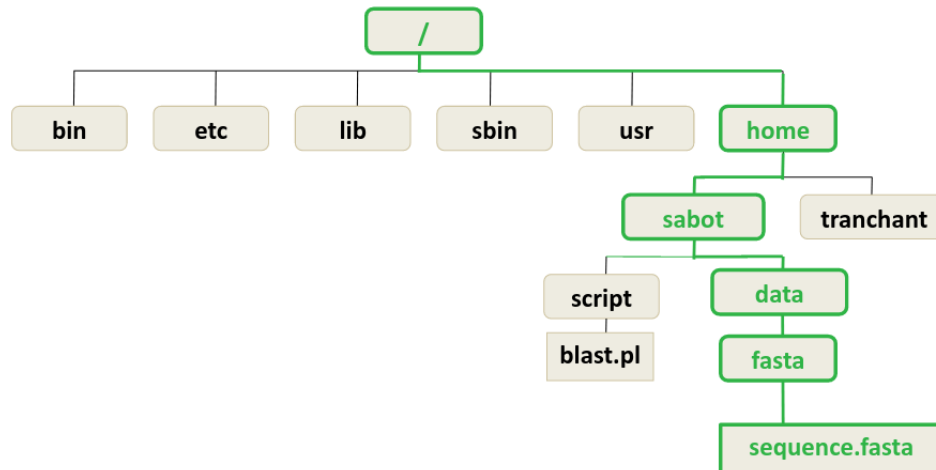
# In brief

- Path = location to a file/directory in the file system

<b>ABSOLUTE</b>	<b>RELATIVE</b>
Complete path of a file starting from the root directory /	Path related to the present working directory (where the user is working)
starts always with /	Never starts with /
always good wherever user is working	Depends on where the user is working

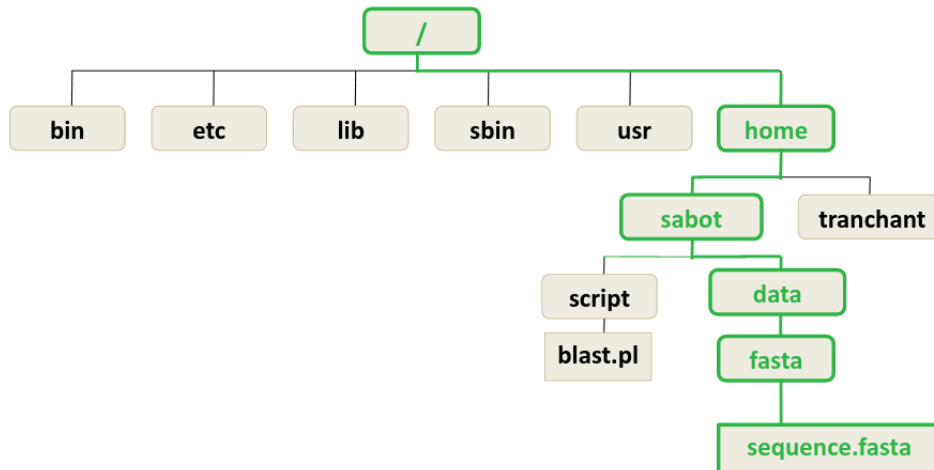
# Absolute Path

- Always start with / (root directory)
- Always works wherever user is working on the server or vm



# Absolute Path

- Always start with / (root directory)
- Always works wherever user is working on the server or vm

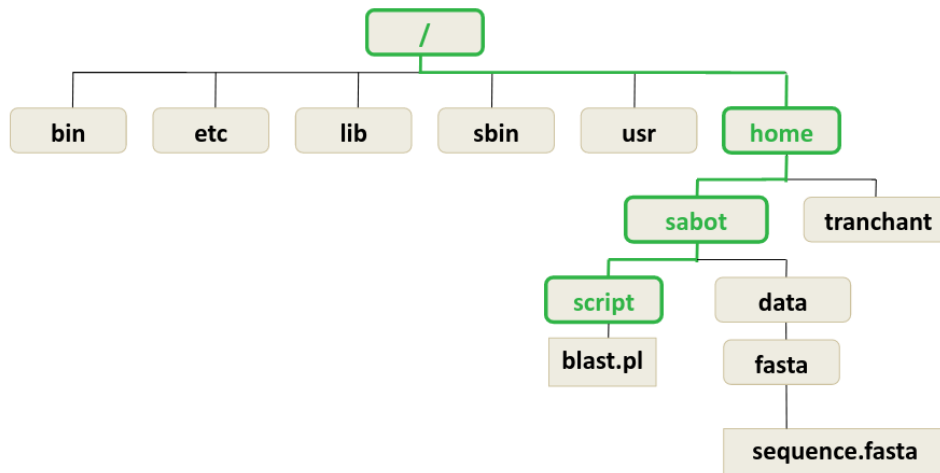


- What is the absolute path of the file sequence.fasta ?



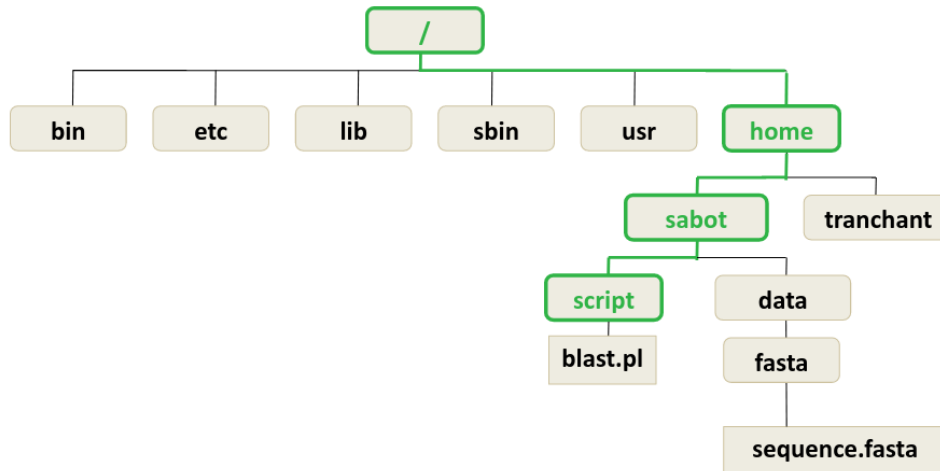
# Absolute Path

- Always start with / (root directory)
- Always works wherever user is working on the server or vm



# Absolute Path

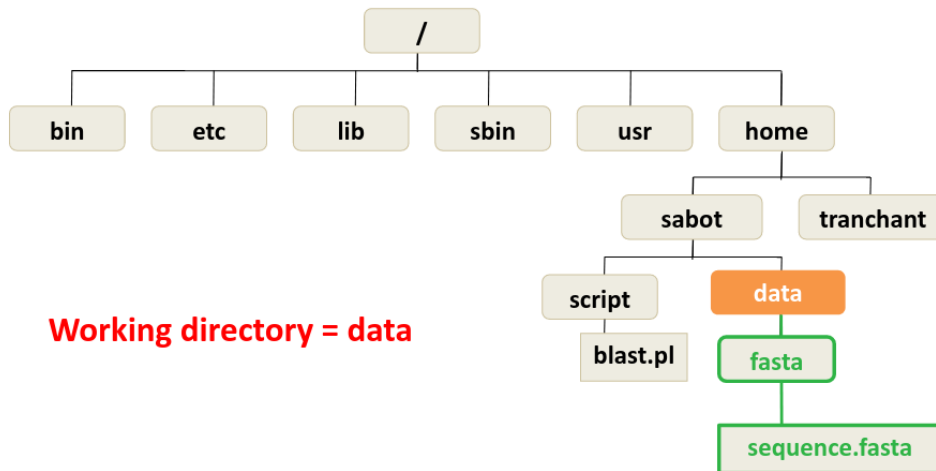
- Always start with / (root directory)
- Always works wherever user is working on the server or vm



- What is the absolute path of the directory script ?

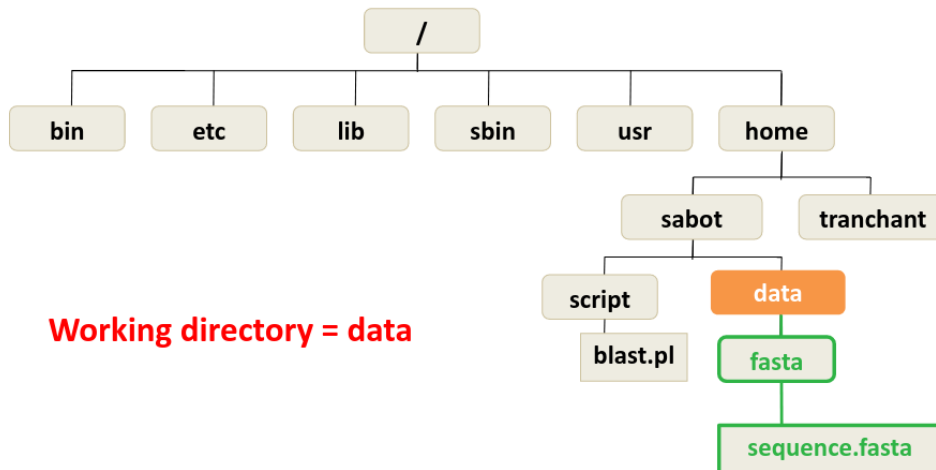
# Relative Path

- Path related to the present working directory
- Never starts with /



# Relative Path

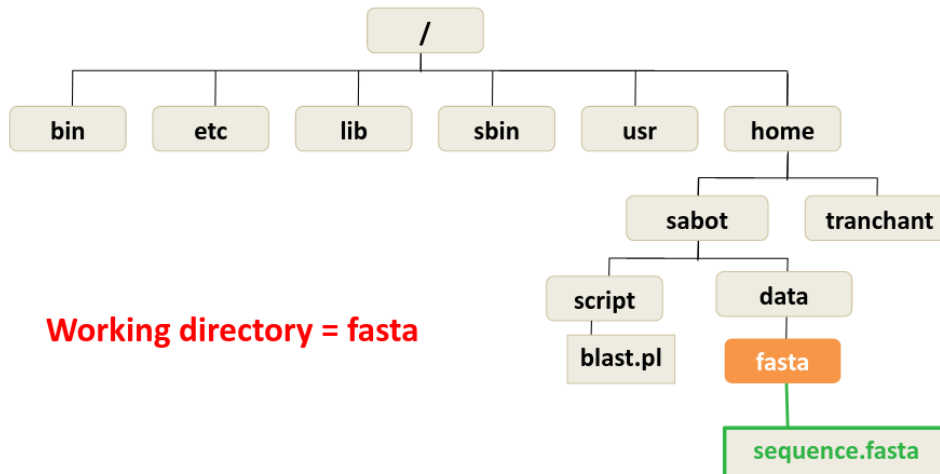
- Path related to the present working directory
- Never starts with /



- What is the relative path of the file sequence.fasta ?

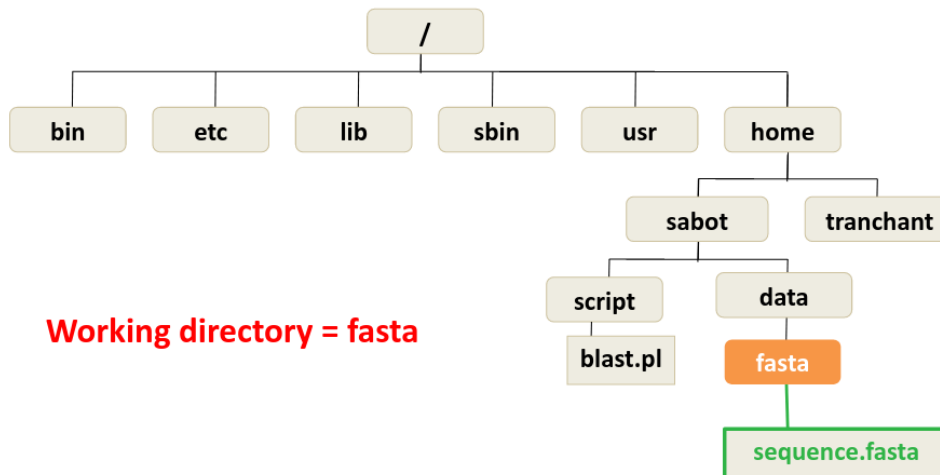
# Relative Path

- Path related to the present working directory
- Never starts with /



# Relative Path

- Path related to the present working directory
- Never starts with /



- What is the relative path of the directory fasta ?



# Rules for naming file and directory names

- linux is case sensitive

```
Sequence.fasta  SEQUENCE.fasta  
sequence.fasta
```

# Rules for naming file and directory names

- linux is case sensitive
- Only ROMAN letters, numbers and \_ -



# Rules for naming file and directory names

- linux is case sensitive
- Only ROMAN letters, numbers and \_ -
- No space, accent or special symbol

```
& ~ # " ' { ( [ | ` \ ^ @ ) ] } $ * % ! / ;  
, ?
```

# Rules for naming file and directory names

- linux is case sensitive
- Only ROMAN letters, numbers and \_ -
- No space, accent or special symbol
- No need to use filename extension (.txt), just to improve readability of filenames.

# A few Basic Commands

---

pwd	Display the full path of the current directory
-----	------------------------------------------------

---

ls	List all files/directories
----	----------------------------

---

ls -l	Display all files (Long listing)
-------	----------------------------------

# A few Basic Commands

How to get help about one command

- with the 'option --help ou -h
- with the command man

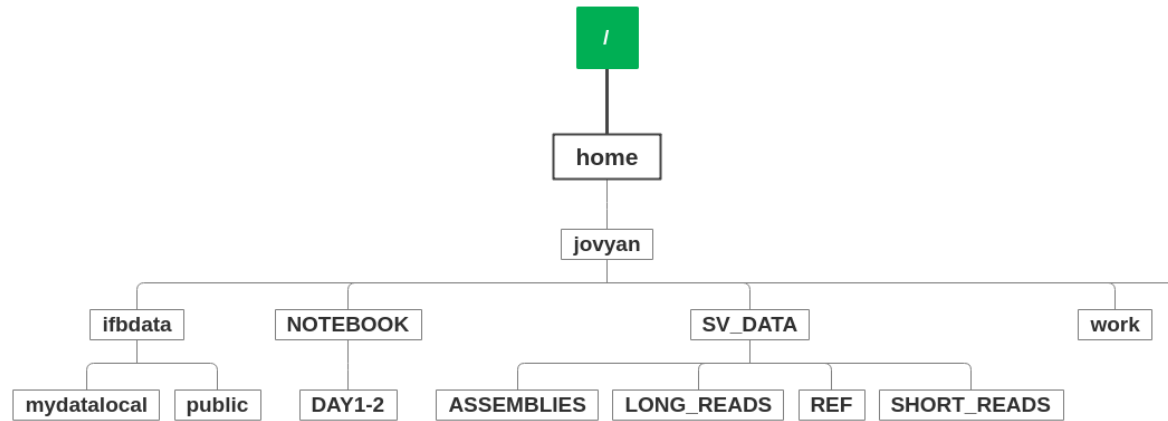
# Command *cd*

- Move from the current directory into a new directory

**cd DIRECTORY NAME (absolute or relative path)**



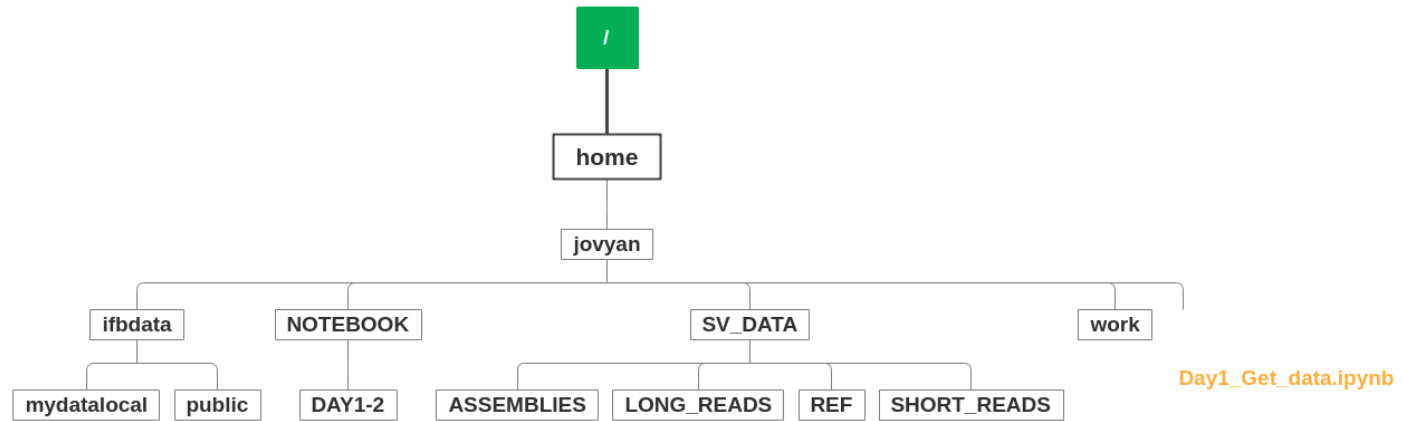
- Move into the REF directory (absolute path)



Day1\_Get\_data.ipynb

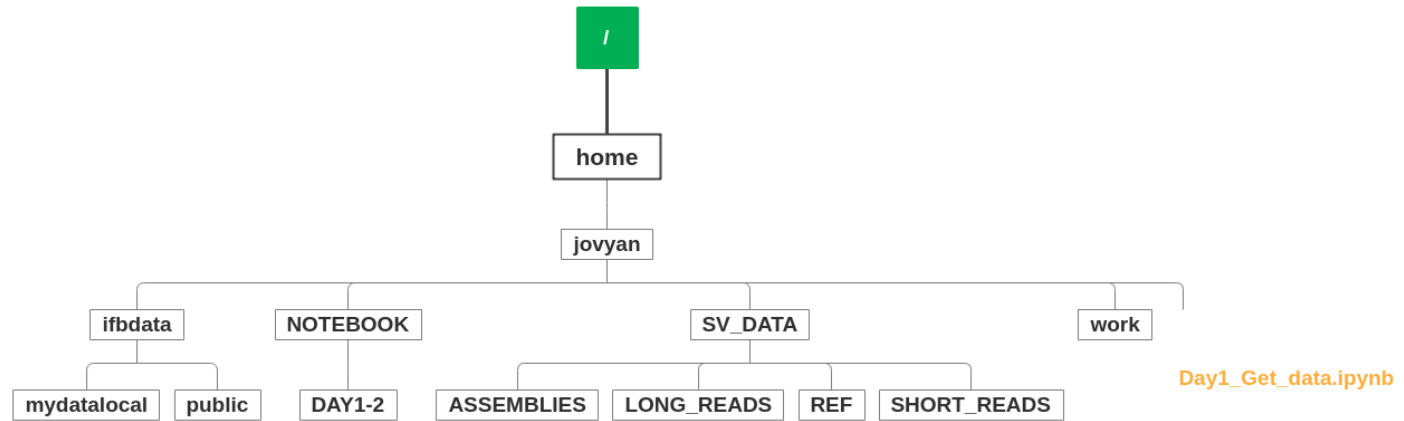


- Move into the SV\_DATA directory (absolute path)





- Move into the SHORT\_READS directory (relative path)





# Some Commands

pwd	Display the full path of the current directory
ls	List all files/directories
cd DIR_NAME	Change the working directory
rm nom_fichier	Remove a directory
cp file1 file2	Make a copy of file1 and calls it file2
cp FILE_NAME DIR_NAME	Copy the file FILE_NAME in the directory DIR_NAME
cp FILE DIR/NEW_FILE	Mix of 2

Commands for displaying  
text files and searching  
word

# command cat

- `cat nom_fichier` : Displays the content of a file on the screen

(Don't use it with big files!!)

# command cat

- `cat nom_fichier` : Displays the content of a file on the screen

(Don't use it with big files!!)

```
In [ ]: cat _PUT_THE_CORRECT_PATH_TO_ASSEMBLIES/assembly-stats.txt_
```

# command wc

- `wc -l` : Displays the number of lines in a file

# command wc

- `wc -l` : Displays the number of lines in a file

```
In [ ]: wc -l PUT_THE_CORRECT_PATH_TO_REF/GCA_002220235.1_ASM222023v1_genomic.gf
```

# command head

- `head nom_fichier` : writes the first ten lines of a file to the screen

# command head

- `head nom_fichier` : writes the first ten lines of a file to the screen

```
In [ ]: head PUT_THE_CORRECT_PATH_TO_REF/GCA_002220235.1_ASM222023v1_genomic.gff
```



# command head

- `head nom_fichier` : writes the first ten lines of a file to the screen

```
In [ ]: head PUT_THE_CORRECT_PATH_TO_REF/GCA_002220235.1_ASM222023v1_genomic.gff
```

```
In [ ]: head -n 20 PUT_THE_CORRECT_PATH_TO_REF/GCA_002220235.1_ASM222023v1_genom
```

# command tail

- `tail nom_fichier` : writes the last ten lines of a file to the screen

# command tail

- `tail nom_fichier` : writes the last ten lines of a file to the screen

```
In [ ]: tail PUT_THE_CORRECT_PATH_TO_REF/GCA_002220235.1_ASM222023v1_genomic.gff
```