



# IX International Course of Massive Data Analysis FOR GENOMICS

Introduction to Linux Shell



Ignacio Medina  
imedina@ebi.ac.uk

**Presentation**

# Index

---

- Introduction
- Most useful shell commands

# Introduction

## What is GNU/Linux and the Shell?

---

- GNU/Linux is a **free** and **open-source** operating system developed by thousands of contributors and led by *Linus Torvalds* since the beginning in 1991
- Linux **shells** (commonly *Bash*) allow users to execute more than 200 commands and to write pipelines in the *Shell Script* programming language to automatize tasks
- Linux is widely used in research and super computers, more than 96% of super computers use Linux:
  - <http://www.top500.org/statistics/list>
- It's an essential tool for bioinformatics and big data analysis and research

# Introduction

## Why we use Linux?

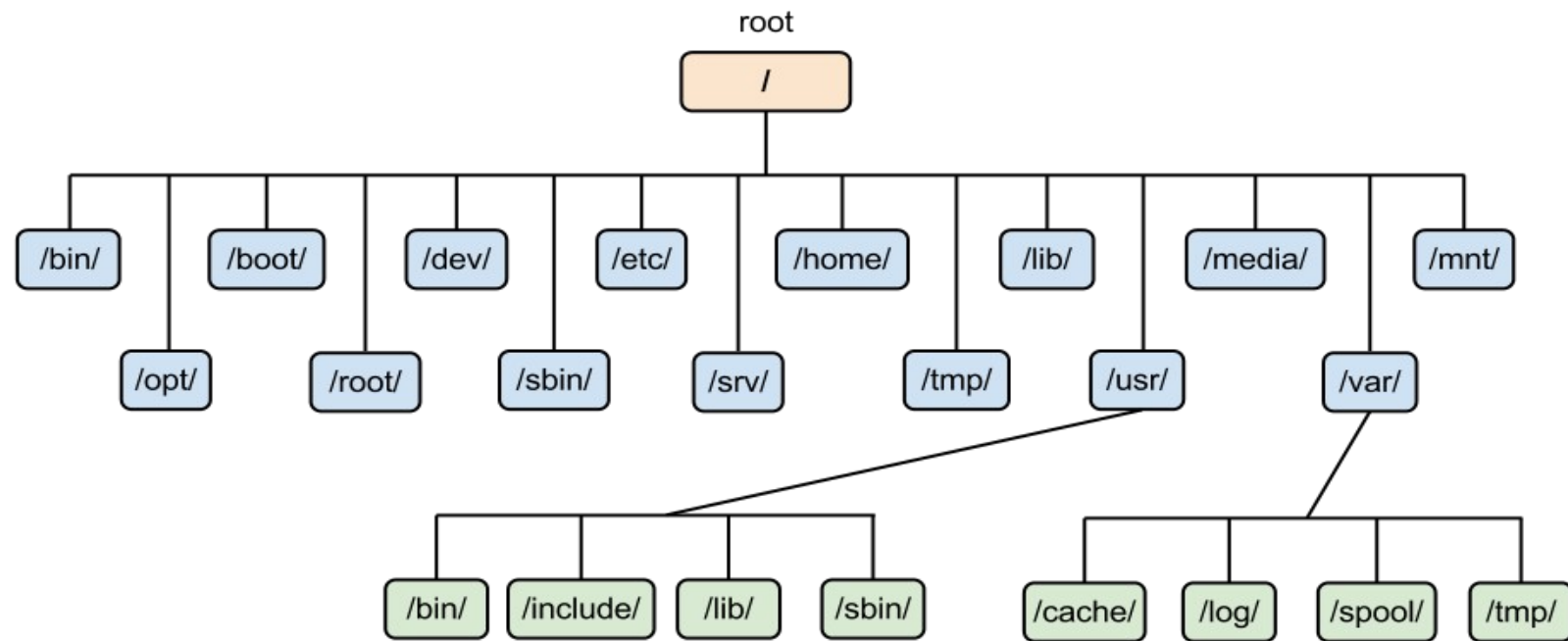
---

- Linux offers:
  - Ease to program and work with big data
  - Stability
  - Security
  - Low price
  - High-performance computing
  - ...
- Usually visual graphical interfaces to software does not exist

# Introduction

## Linux filesystem

- All paths start at '/' called **root**, no C: or D: like in Windows. Similar to other Unix-like systems such as Mac OS
- Users home folders under '/home', ie. */home/participant*



# Introduction

## Understanding *paths*

---

- A ***path*** identifies uniquely a file or directory in the file system.
- The character **'/'** is used to concatenate directories
- **IMPORTANT** - Two types of paths:
  - **Absolute**: Always start with **'/'** which is the root folder. Example:
    - *ls /home/participant/Desktop*
  - **Relative**: All paths not beginning with **'/'** but with a file or folder name. Example
    - *ls Desktop*
- **'Tab'** key is your friend, auto-completes the paths for you if you press twice

# Introduction

## Understanding *\$PATH* variable

---

- A *\$PATH* is a environment variable that list all directories in the system with *binaries* (executable programs)
- You can see the list by executing:
  - *echo \$PATH*
- All binaries in one of these directories can be executed automatically in the shell, no absolute path is needed
- You can use '*Tab*' key twice to autocomplete the binary name

# Most useful commands

## Working with files and directories

---

- Many tutorials and documentation:
  - [http://linuxcommand.org/learning\\_the\\_shell.php](http://linuxcommand.org/learning_the_shell.php)
- Inline help in the shell using command '**man**', ie. *man ls*
- Some useful **shell commands**:
  - cd
  - ls
  - mkdir and rm
  - pwd
  - mv
  - cp
  - less, head and tail
  - tree
  - Output redirection: '>'
  - ...