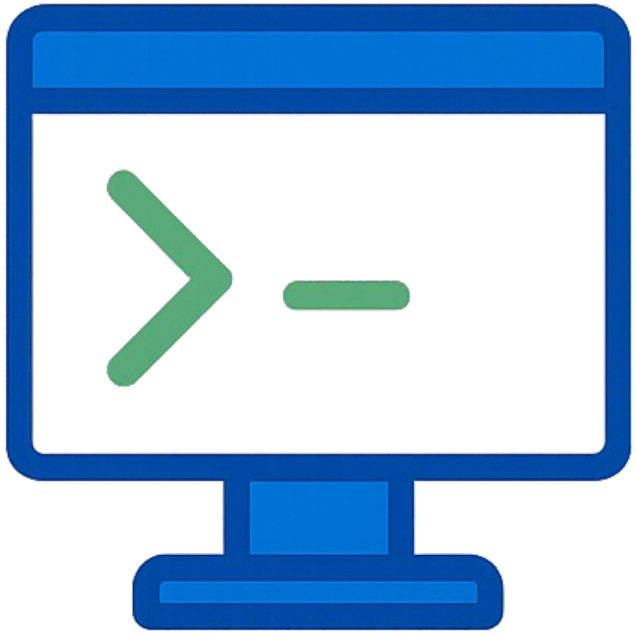
Computer Operating Systems

Computer fundamentals for bioinformatics

Kristine Lacek, MS
Bioinformatics Scientist
US CDC – Influenza Division



Disclaimer

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Use of trade names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

References to non-CDC sites on the Internet do not constitute or imply endorsement of these organizations or their programs by CDC or the U.S. Department of Health and Human Services. CDC is not responsible for the content of pages found at these sites.

Module Objectives

- Basic computer architecture
  - BIOS/UEFI

Module Objectives

- Basic computer architecture
  - BIOS/UEFI
- Introduce Operating Systems (OS)

Module Objectives

- Basic computer architecture
  - BIOS/UEFI
- Introduce Operating Systems (OS)
- Understand how to use Linux on Windows
  - What is a virtual machine is?
  - How do operating systems differ?
  - What does *nix means for bioinformatics?

Module Objectives

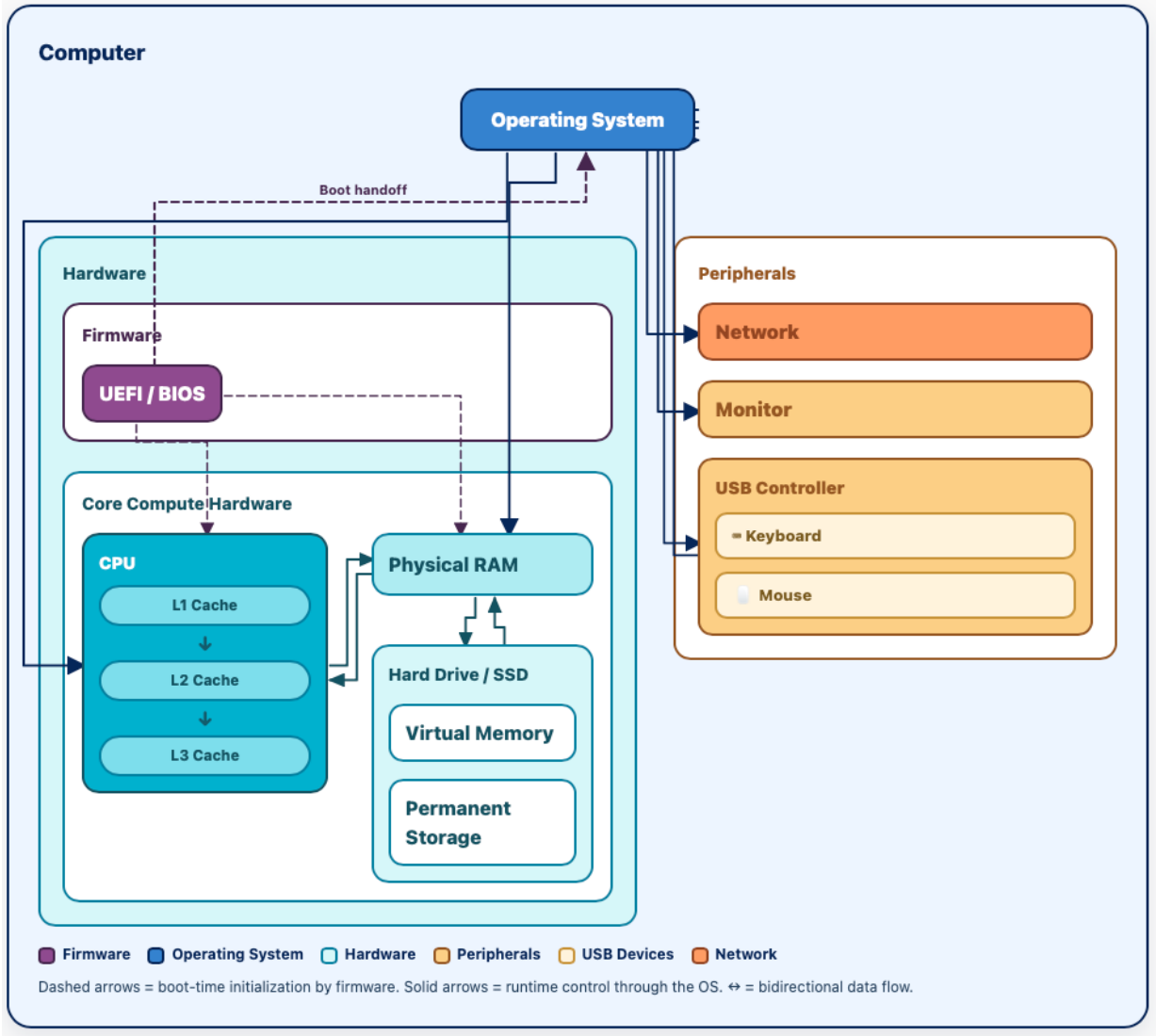
- Basic computer architecture
  - BIOS/UEFI
- Introduce Operating Systems (OS)
- Understand how to use Linux on Windows
  - What is a virtual machine is?
  - How do operating systems differ?
  - What does *nix means for bioinformatics?
- Install WSL on windows machines, access command prompt from MacOS

Computer architecture

- Basic computer architecture

Core Hardware Components

- **CPU (Processor)**
  - Executes instructions
  - Determines compute capacity
  - More cores = more parallel processing
- **RAM (Memory)**
  - Temporary working space for running programs
  - Fast, volatile storage
  - Limits how much data can be processed at once
- **Hard Drive (Storage)**
  - Long-term data storage
  - Slower than RAM
  - Used for files, datasets, logs, and outputs



Computer architecture

- Basic computer architecture

In Bioinformatics, We Primarily Monitor:

1. Memory (RAM usage)

Large datasets can exhaust memory

Insufficient RAM causes crashes or slowdowns

2. I/O (Disk Read/Write Speed)

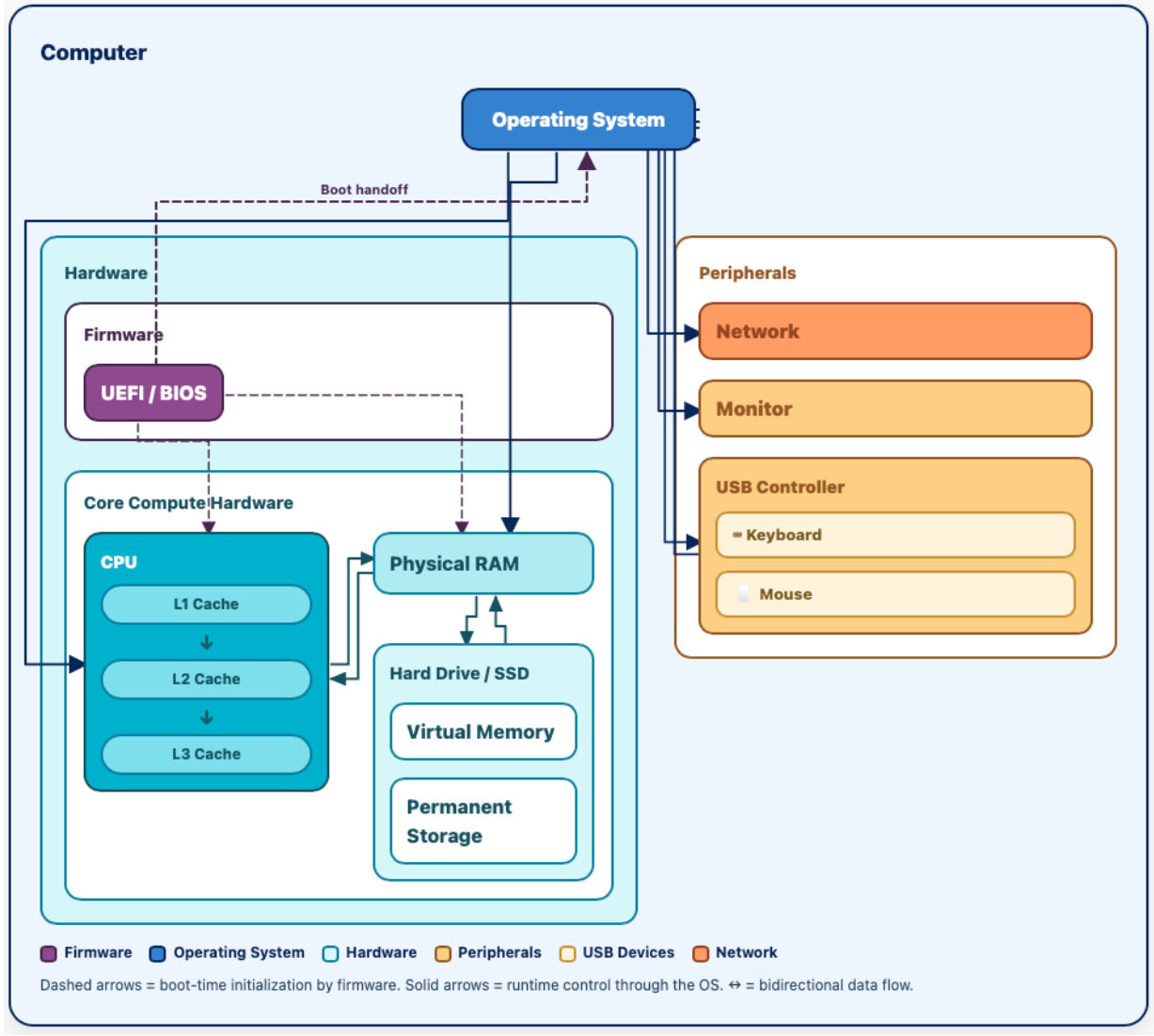
Moving large files can become a bottleneck

High disk usage slows pipelines

3. CPU Count (Core Availability)

Many tools support multithreading

More cores can reduce runtime



Operating Systems (OS)

- An **operating system** (OS) is the core software that manages all hardware and software resources on a computer
- Provides a **user interface**, allowing people to interact with applications, files, and the system itself
- Handles essential tasks
  - memory management
  - process scheduling
  - storage access
  - security controls
- Allows applications to run by managing how programs access hardware resources

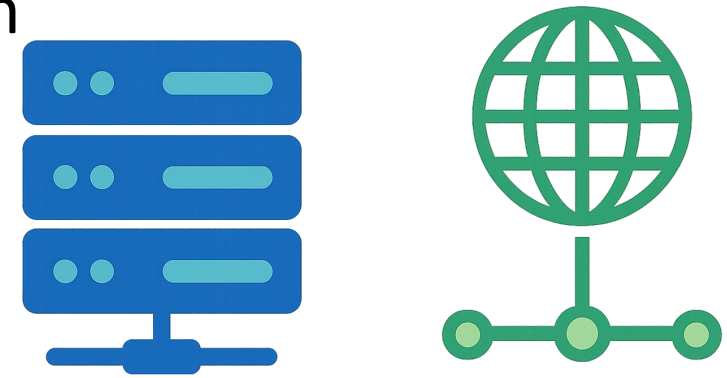
- Common OS examples
  - Windows
  - macOS
  - Linux
  - Android
  - iOS

Virtual Machines (VMs)

- A **virtual machine** (VM) emulates a full computer system, allowing multiple operating systems to run on a single physical machine
- Run two different operating systems on the same hardware for flexibility:
  - Compatibility needs
  - Using tools available only on a specific OS
  - Ease of use
- VMs provide strong isolation between multiple operating systems

Networks and Servers

- A **network** is a system that connects computers and devices, allowing them to **share** data, resources, and services through wired or wireless communication links
- A **server** is a computer system designed to provide resources or services—such as data, applications, or network functions—to other devices over a network
  - $$$$
  - Optimized for reliability and continuous operation
  - Handle requests from clients, managing tasks
    - Data storage
    - Application hosting
    - Authentication
    - Communication between systems



Command Line Interface

- Much of bioinformatics is done in a Linux (or unix) OS: ***nix**
- Unix and Linux are built around a command-line–focused design: system control, automation, and scripting
  - Windows is oriented toward a graphical user interface (GUI)
    - also provides command tools like PowerShell and Command Prompt
- Unix/Linux environments are widely used in servers, bioinformatics, and high-performance computing
    - command-line tools
    - **shell scripts**

Command Line Interface

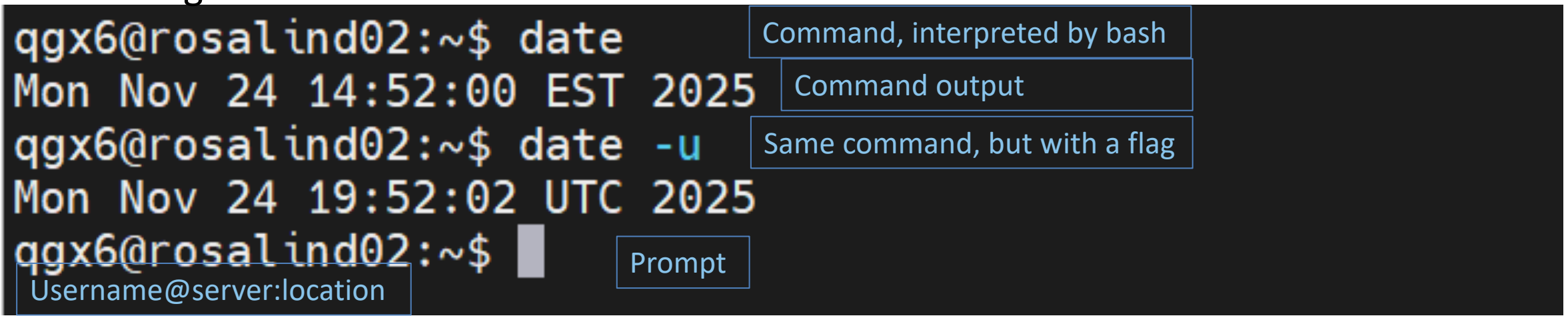
- Much of bioinformatics is done in a Linux (or unix) OS: ***nix**

```
[root@localhost ~]# ping -q fa.wikipedia.org
PING text.pmtpa.wikimedia.org (208.80.152.2) 56(84) bytes of data.
^C
--- text.pmtpa.wikimedia.org ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 540.528/540.528/540.528/0.000 ms
[root@localhost ~]# pwd
/root
[root@localhost ~]# cd /var
[root@localhost var]# ls -la
total 72
drwxr-xr-x. 18 root root 4096 Jul 30 22:43 .
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.  3 root root 4096 May 18 16:03 empty
drwxr-xr-x.  2 root root 4096 May 18 16:03 games
drwxrwx--T.  2 root gdm  4096 Jun  2 18:39 gdm
drwxr-xr-x. 38 root root 4096 May 18 16:03 lib
drwxr-xr-x.  2 root root 4096 May 18 16:03 local
lrwxrwxrwx.  1 root root   11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x. 14 root root 4096 Sep 14 20:42 log
lrwxrwxrwx.  1 root root   10 Jul 30 22:43 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
lrwxrwxrwx.  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 wiki
Loaded plugins: langpacks, presto, refresh-packagekit, remove-with-leaves
rpmfusion-free-updates                                   | 2.7 kB     00:00
rpmfusion-free-updates/primary_db                        | 206 kB     00:04
rpmfusion-nonfree-updates                                | 2.7 kB     00:00
updates/metalink                                         | 5.9 kB     00:00
updates                                                  | 4.7 kB     00:00
updates/primary_db                  73% [=============        ]  62 kB/s | 2.6 MB     00:15 ETA
```

Command Line Interface Terminology

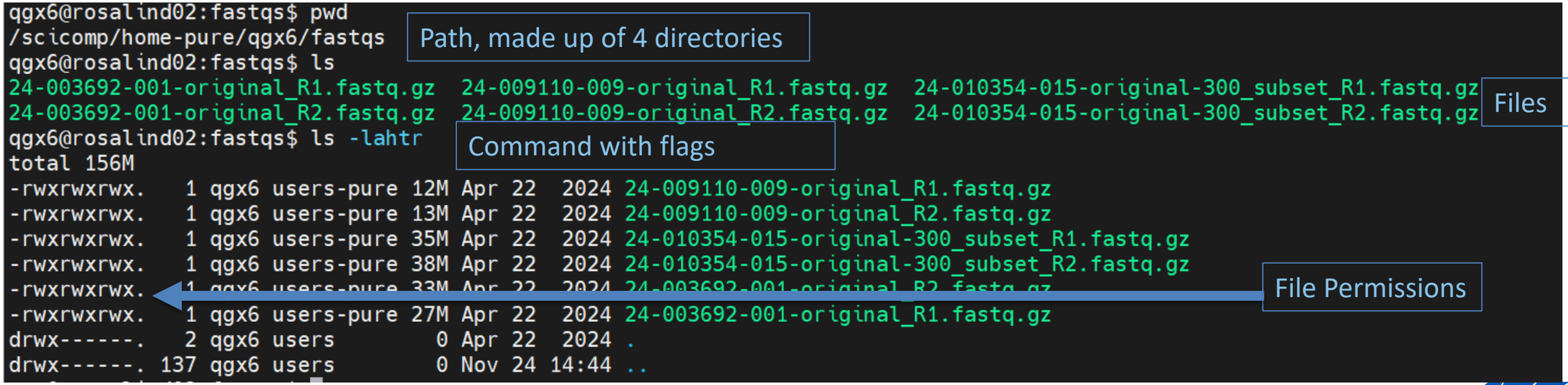
- **Shell** — The program that interprets command-line input and runs commands
  - bash (Linux)
  - zsh (Unix)
  - PowerShell (Windows)

  Very similar → *nix
- **Prompt** — The text displayed by the shell indicating it is ready to accept a command.
- **Command** — **A program** or instruction typed into the shell to perform an action, such as listing files or moving data
- **Flag (or option)** — A modifier added to a command (often starting with - or --) that changes how the command behaves.



Command Line Interface Terminology

- **Directory** — A location in the filesystem used to organize files; similar to a “folder” in graphical interfaces.
- **Path** — The exact location of a file or directory in the filesystem, written as a sequence of directory names.
- **Permissions** — Rules that control who can read, write, or execute a file or directory.



Linux and Filesystem

Linux treats most things like files: one consistent way to access resources

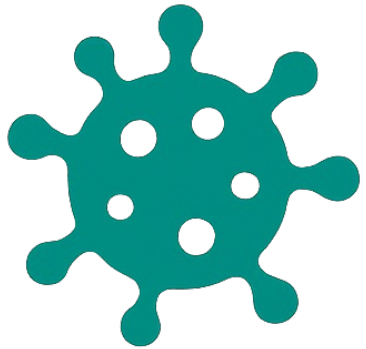
- Regular files (documents, programs)
- Directories
- Devices (/dev)
- System information (/proc)

- All files exist under one root directory

```
/
├── home/
├── etc/
├── var/
└── dev/
```

Bioinformatics and CLI

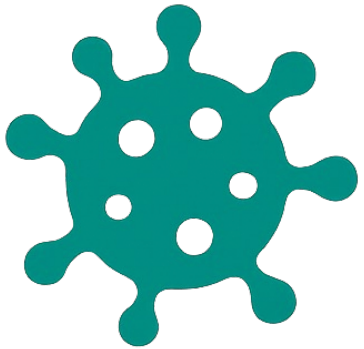
Much of Bioinformatics is file manipulation and stringing together inputs and outputs from command-line tools



Bioinformatics and CLI

Much of Bioinformatics is file manipulation and stringing together inputs and outputs from command-line tools

The CLI allows efficient handling of large datasets, enabling researchers to run pipelines, filter data, and automate repetitive tasks using powerful text-based tools

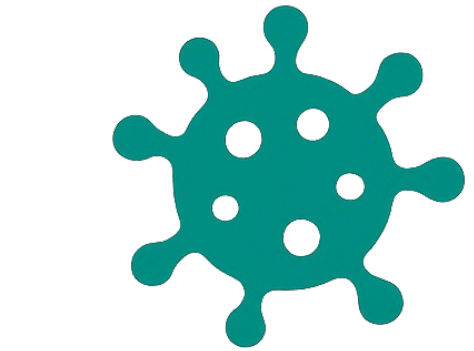


Bioinformatics and CLI

Much of Bioinformatics is file manipulation and stringing together inputs and outputs from command-line tools

The CLI allows efficient handling of large datasets, enabling researchers to run pipelines, filter data, and automate repetitive tasks using powerful text-based tools

Many programs are designed for command-line execution, providing flexibility, scalability, and compatibility with high-performance and cloud-computing environments



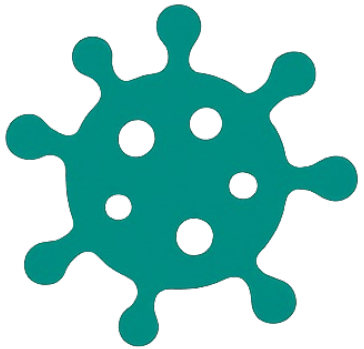
Bioinformatics and CLI

Much of Bioinformatics is file manipulation and stringing together inputs and outputs from command-line tools

The CLI allows efficient handling of large datasets, enabling researchers to run pipelines, filter data, and automate repetitive tasks using powerful text-based tools

Many programs are designed for command-line execution, providing flexibility, scalability, and compatibility with high-performance and cloud-computing environments

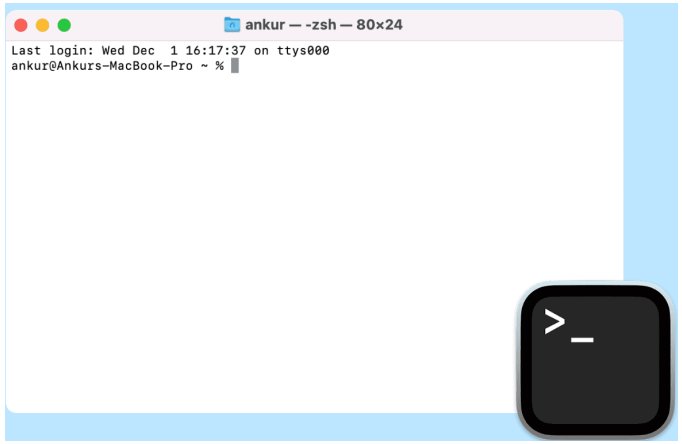
An adept informatician is incredibly comfortable working in the CLI



Practical: Command-line on your Computer

MacOS:

1. Press Command (⌘) + Spacebar to open Spotlight.
2. Type **Terminal**.
3. Press Return to open it.
4. A command prompt will appear on your screen



Windows computer:

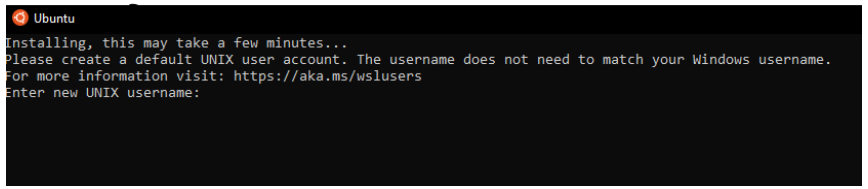
1. Run **Powershell** as Administrator
2. Run the following command in Powershell:

```
wsl --install
```

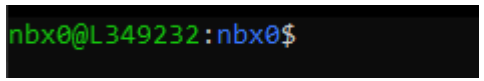
3. Restart your computer
4. Reopen Powershell and enter the following commands:

```
wsl --set-default-version 2
wsl --install -d Ubuntu-24.04
```

Following successful installation, an Ubuntu terminal should pop up



5. Enter a username that will be exclusive for WSL. Press Enter and then enter a password
6. A command prompt will appear on your screen



LinuxOS:

1. Open the Applications or Activities menu (top-left or bottom-left, depending on your system).
2. Search for Terminal, Console, or Xterm (names vary by distribution).
3. Click the icon to launch it.
4. A command prompt will appear on your screen