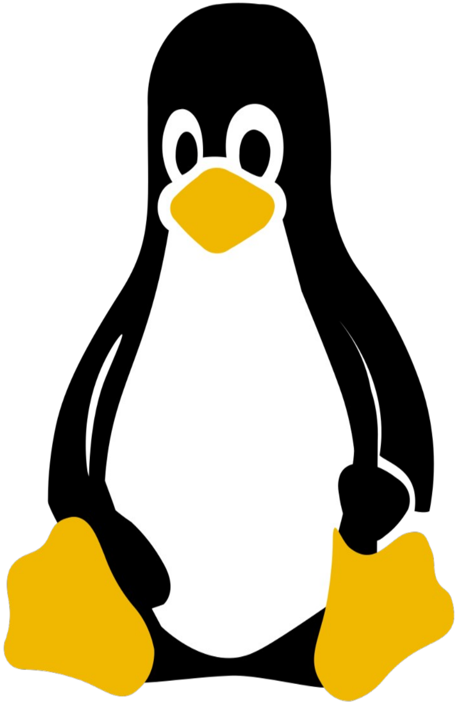
Linux™



Introduction to Linux

- **What is Linux ?**

Linux is operating system that is widely used in scientific computing (including bioinformatics)

- **How is Linux different to other OS ?**

In Linux, the user can access the source code of the kernel and change the code according to their needs.

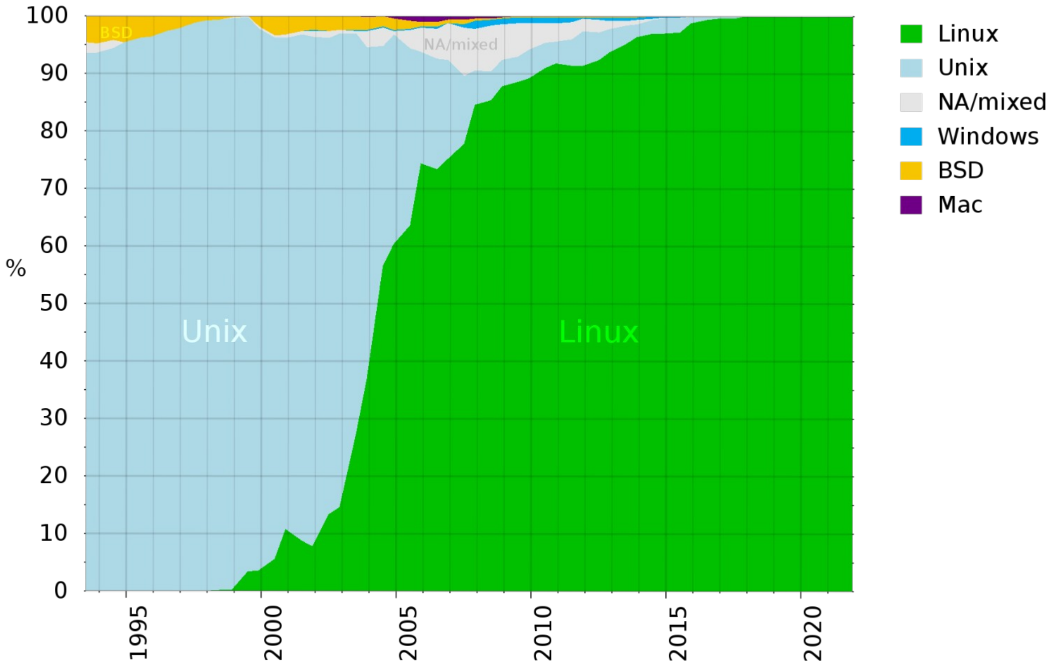
- **What is kernel?**

The kernel is the central program in an operating system and has ultimate authority over the computer.

- **What are some advantages of using Linux?**

The security it provides eliminates the need for antivirus software, it is stable, takes up minimal disk space, has extensive networking capabilities, and puts software updates in the hands of the user.

Usage of the Linux in Supercomputer



OS used on top 500 supercomputers

Why is Linux commonly used by bioinformaticians?

- Linux is commonly used in scientific computing because it is Open Source and has attracted many different contributors.
- Linux's command line interfaces CLI enables bioinformaticians to handle data, automate tasks, and run complex research pipelines. The CLI processes large volumes of data, text files, and bioinformatics applications.
- Linux-optimized bioinformatics tools are plenty. Alignment, sequence analysis, genomics, proteomics, and statistical analysis methods are used by researchers.
- Scalability and speed allows Linux to perform computationally difficult tasks including Next Generation Sequencing (NGS) and genomic analysis. Bioinformaticians can analyze data quicker using multi-core processors and distributed computing systems.

Linux Distributions

What is a Linux Distribution?

A Linux distribution, often known as a Linux distro, is a whole operating system based on the Linux kernel. It is made up of the Linux kernel, core system libraries, user applications, utilities, and a variety of software packages that work together to offer a functional computer environment.

Commonly used Linux Distributions

- Ubuntu: Ubuntu is one of the most widely used Linux distributions. It focuses on ease of use and provides a user-friendly interface. It has both a desktop version and a server version.

- Debian: Debian is a stable and highly customizable distribution. It emphasizes free and open-source software. Many other Linux distributions, including Ubuntu, are derived from Debian.

- Fedora: Fedora is a community-driven distribution sponsored by Red Hat. It aims to showcase the latest technologies and software in the Linux ecosystem. It is known for its strong focus on security and new features.

Command Line Interface (CLI)

- Linux includes a robust CLI that allows users to interact with the operating system by giving commands.
- CLI provides excellent control and flexibility over system management, automated processes, and scripting.
- CLI in Linux is a text-based technique of communicating with the operating system. It allows users to enter text strings as orders to accomplish various operations, execute applications, modify settings, and manage the system.
- The CLI in Linux is provided by a shell, which is a program that interprets user commands and executes them. The default shell in most Linux distributions is Bash (Bourne Again SHell), but other **shells** like Zsh, Ksh, and Csh are also available.

What is the shell?

- The shell is a **CLI** or text-based user interface that lets users submit commands to interact with the operating system.
- The shell interprets and executes commands, manages the execution environment, and provides system resources and tools.
- The shell lets you launch programs, modify files and directories, and administer the system.
- Shells let users automate processes using scripts. The command-line environment is crucial for system administrators, developers, and power users to efficiently and flexibly operate the computer system.

Shell For Bioinformatics

- The Shell is the **CLI** environment in genomics that allows researchers and bioinformaticians to interact with and manipulate genetic data through the use of various tools and scripts. The shell allows you to execute genetic analysis, data processing, and bioinformatics operations rapidly and reproducibly.
- Shell environments provide access to a number of bioinformatics software and tool packages to conduct a variety of genomic research utilizing command-line tools, including sequence alignment, variant calling, gene expression analysis, and more

How to access the Bash shell?

- In Linux and MacOS you already have the Bash shell
    - MacOS might need some changes to make Bash your default shell
- In Windows the “Git for Windows” package provides a Bash shell
- For instructions on Shell software setup, consult the Software Carpenties guide
    - You can find it in the Software Carpentries Bash Shell Lesson

Useful Commands

- Network: **ssh, scp, ping, telnet, nslookup, wget**
- Shells: **BASH, watch, clear, history, echo**
- System Information: **w, whoami, man, info, which, free, echo, date, cal, df, free, man, info**
- Command Information**: man, info**
- Symbols: |**, >, >>, <, &, >&, 2>&1, ;, ~, ., .., $!, !:<n>, !<n>**
- Filters: **grep, egrep, more, less, head, tail**
- Hotkeys: **ctrl+c , ctrl+d**
- File System: **ls, mkdir, cd, pwd, mv, ln, touch, cat, file, find, diff,**

  File Editors: **vim, gvim, nano**

Network

- Secure Shell: Remote login and secure communication.
- **scp**: Secure Copy for local and remote file transmission.
- **ping**: Sends ICMP echo queries to an IP or domain to test network connection.
- **telnet**: Interactive server communication.
- **nslookup**: Retrieves domain IP addresses and other DNS information.
- **wget**: Command-line utility for HTTP, HTTPS, and FTP file downloads.

System information

- **w:** Displays currently logged-in users.
- **whoami:** Prints the current user's username.
- **man:** Displays a command's manual page.
- **info:** Provides command documentation.
- **which**: Shows command or executable location.
- **free:** Displays system memory usage.
- **echo:** Displays text or variables on the terminal.
- **date:** Prints or sets the system date and time.
- **cal:** Shows a month or year's calendar.
- **df** displays disk space usage.

Filters

- **grep:** Searches for text patterns in files or command output.
- **egrep:** Extended version of grep with more advanced pattern matching capabilities.
- **more:** Displays output one page at a time.
- **less:** Allows scrolling and searching through output.
- **head:** Displays the beginning portion of a file or command output.
- **tail:** Displays the end portion of a file or command output.

File System

- **ls:** Lists files and directories in a directory.
- **mkdir:** Creates a new directory.
- **cd:** Changes the current directory.
- **pwd:** Prints the current working directory.
- **mv:** Moves or renames files and directories.
- **ln:** Creates a hard or symbolic link to a file or directory.
- **touch:** Creates an empty file or updates the timestamp of an existing file.
- **cat:** Displays the contents of a file.

- **df:** Shows disk space usage of file systems.
- **du:** Displays disk usage of files and directories.
- **mount:** Mounts a filesystem to a specific directory in the file hierarchy.
- **file:** Determines the type of a file.
- **find:** Searches for files and directories based on various criteria.
- **diff:** Compares files line by line.
- **cmp:** Compares two files byte by byte.
- **chmod:** Changes the permissions of files and directories.
- **find**: Searches for files and directories based on various criteria.

File and line editor

Line Editors:

- **awk:** A versatile text processing language for extracting and manipulating data.
- **sed:** Stream Editor, used for text transformation and editing.
- **File Editors:**
- **vim:** A highly configurable and powerful text editor.
- **gvim:** Graphical version of Vim.
- **nano:** popular text editor for Unix-like systems that simplifies terminal text file editing.

Shell script or .Sh Script File

- **Shell script**, is a text file that contains a script in a scripting language like Bash (Bourne Again SHell). Shell scripts simplify the process of doing repetitive or difficult activities by automating them and executing commands in a predetermined order.

- **Shell script creation**: Launch Nano, Vim, or any other txt editor. Name a new file with an essential **".sh"** extension. Like **"script.sh"**. Start the file with the shebang line to designate the interpreter. Bash shebang lines are: **#!/bin/bash**

- **Shellscript execution**: Run the script by typing its name or path in the terminal:

```
$ ./script.sh
```