INTRODUCTION TO COMMAND LINE INTERFACE

Klebsiella Workshop

Sep 2024

BIG DATA INSTITUTE
UNIVERSITY OF OXFORD
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE
SANBI
South African National Bioinformatics Institute
UNIVERSITY of the WESTERN CAPE
ASLM
AFRICAN SOCIETY FOR LABORATORY MEDICINE
AFRICA CDC
PUBLIC HEALTH ALLIANCE FOR GENOMIC EPIDEMIOLOGY
African Union

Content

Aims

- Introduce the advantages of CLI
- Cover the Basics of CLI Commands
- Implementation of commands
- Installation & execution of tools
- Analyses of outputs
- Complete hands-on exercise

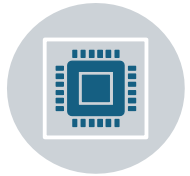
Introduction

YOU ARE HERE

COMMAND LINE TOOLS ARE FASTER AND MORE EFFICIENT FOR HANDLING LARGE DATASETS

COMMAND LINE ALLOWS REMOTE ACCESS TO POWERFUL COMPUTATIONAL RESOURCES (E.G., CLUSTERS, CLOUD SERVERS)

TOOLS AND PIPELINES CAN BE EASILY DEVELOPED, CUSTOMIZED AND COMBINED



COMMANDS CAN BE DOCUMENTED AND SHARED, ENSURING THAT ANALYSES ARE REPRODUCIBLE BY OTHERS.

WHILE THERE IS A STEEP LEARNING CURVE, MASTERING THE COMMAND LINE IS INVALUABLE FOR BIOINFORMATICIANS AND OFFERS LONG- TERM BENEFITS.

COMMAND LINE INTERFACES OFTEN INTEGRATE WELL WITH VERSION CONTROL SYSTEMS (E.G. GIT)

CLI INTRO

UNIX & BIOINFORMATICS

NETWORK

SHELL

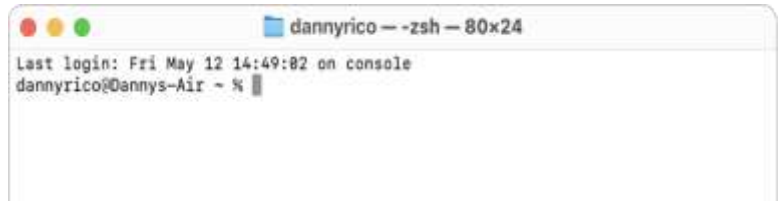
FILTERS

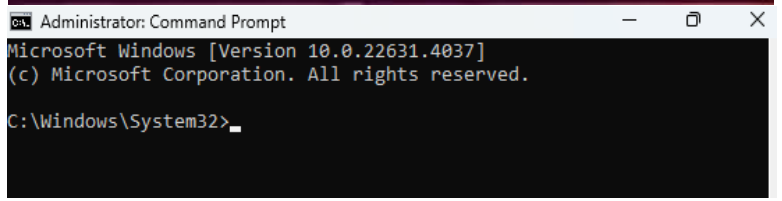
FILES & SYSTEMS

Command Line Interface

YOU ARE HERE

- Text based techniques which allows users to interact with operating system OS
- CLI is provided by a Shell – program which interprets user commands & executes them.
- Excellent control and flexibility over:
  - System management
  - Launch Programs
  - Modify Files & Directories
  - Automated processing
  - Scripting





Linux Environment

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

- The Bash shell (the **B**ourne **A**gain **Sh**ell) is the most popular Unix shell.

- Advantages of using Linux:
  - Software Security
  - Stability
  - Extensive networking capabilities
  - Software updates in the hands of the user

Linux for Bioinformatics

YOU ARE HERE

- Linux is Open Source and has attracted many different contributors.
  - Linux's CLI enables bioinformaticians:
    - Handle large quantities and varieties of data
    - Automate complex analyses
    - Run complex research pipelines
- Linux-optimized bioinformatics tools are plenty:
  - Alignment
  - Sequence analysis
  - Genomics
  - Proteomics
  - Statistical analyses
- Scalability allows quicker data analyses.

CLI INTRO

UNIX & BIOINFORMATICS

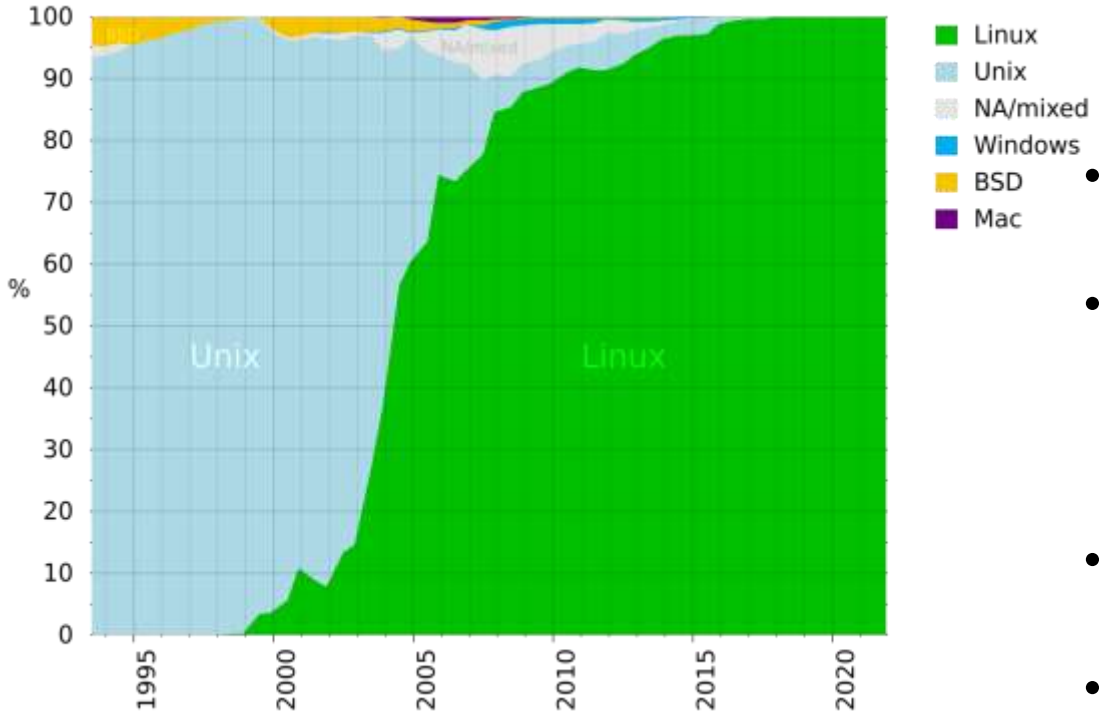
NETWORK

SHELL

FILTERS

FILES & SYSTEMS

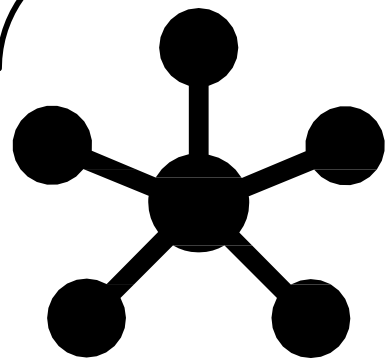
Linux for Super Computing



- Over 70% of web servers run on Linux.
- 90% of all cloud infrastructure runs on Linux.
- More than 95% of the world's top one million servers run on Linux.
- Linux runs on over 90% of the world's supercomputers.
- The Linux operating system has thousands of developers worldwide contributing to its codebase.
- Linux has been around for over 25 years
- Linux continues to evolve with regular updates.

YOU ARE HERE

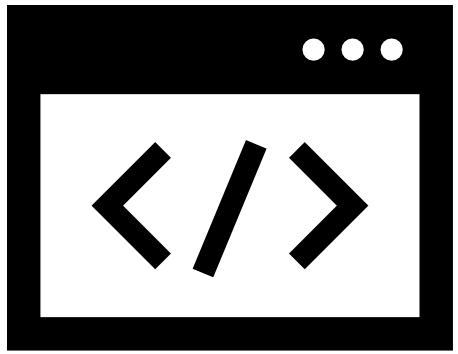
Network Commands



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

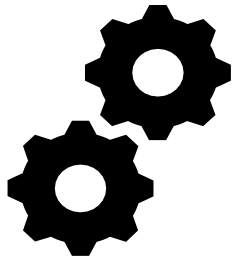
Shell Commands

- **watch:** Monitor changes in command output over time
- **clear:** Empties the current terminal display
- **history:** view previously executed commands
- **echo:** Displays text or variables on the terminal.
- **info:** Provides command documentation.
- **free:** Displays system memory usage.
- **date:** Prints or sets the system date and time.
- **cal:** Shows a month or year's calendar.
- **df:** displays disk space usage.



YOU ARE HERE

Filters Commands



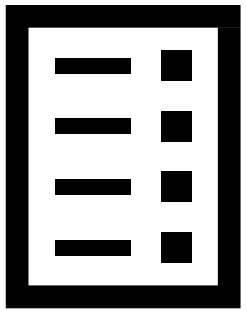
- **grep:** Searches for text patterns in files or command output.
- **egrep:** Extended version of grep with more advanced pattern matching capabilities.
- **more:** Displays output one page at a time.
- **less:** Allows scrolling and searching through output.
- **head**: Displays the beginning portion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

YOU ARE HERE

File Systems

- **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.

File & 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/ touch:** 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