AfricaCDC
Centres for Disease Control and Prevention

Module

B01

Bioinformatics Foundational Course

Introduction to Unix/Linux, Command Line and Shell Scripting

NGS Academy for the Africa CDC

ASLM
AFRICAN SOCIETY FOR LABORATORY MEDICINE

H3ABioNet
Pan African Bioinformatics Network for H3Africa

ASLM ACADEMY

Module
B01

Introduction to Unix/Linux, Command Line and Shell Scripting

Module last updated:
December 2024

Suggested or approximate number of sessions	2-3
Suggested or approximate total learning time	5-6 hours for each section (a and b)
Target audience	Bioinformaticians and IT personnel
Delivery format	Lectures, videos, practicals
Level of the module	Introductory

👤 Contributors

Hocine Bendou, George Githinji, Shahiid Kiyaga, Tony Yiqun Li, Perceval Maturure, Nicola Mulder and Verena Ras.

📖 Module description

This module explores the concepts related to multiuser operating systems, with a specific focus on Linux. It is divided into 2 sections, a) for basic Linux/unix and b) for shell scripting. Linux is a freely available operating system that falls under this category. Some popular distributions of Linux include Red Hat Enterprise Linux, Ubuntu, and CentOS. Multiuser operating systems like Linux are utilized in various environments, such as high-end workstations, database servers, web servers, and for managing shared resources. They offer a range of standard features, including enhanced security measures, reliability, and scalability. This module will cover use of command-line utilities and use of command-line tools to organize and analyze data. The module content will include the intricacies of Linux as a multiuser operating system, examining its core functionalities and benefits, understanding how it is utilized in different settings, data transfers/management on a Linux machine, some performance measurement and analysis tools. In the second part, shell scripting is introduced for basic data or text manipulation. It is recommended that the required tools are installed prior to the course and that a Linux operating system is used. If using the Carpentries course, it is designed to be taught over two full days of instruction. In this module, participants are introduced to the following topics and/or concepts:

a. Introduction to Linux/unix
   ○ Installing Linux
   ○ Basic commands
   ○ Understanding your file structure and creating directories
   ○ Navigating your file system using the command line
   ○ Working with files, command history, examining files, file formats, creating, moving, copying, and removing files
   ○ Basic commands for viewing and manipulating files (cat, more, head), sorting contents, searching for patterns in files, comparing files
   ○ Using text editors such as nano
   ○ Using help files to identify useful command options.
   ○ Using tab completion, and its advantages.
b. Introduction to bash and shell scripting
   ○ Key reasons for learning shell.
   ○ Accessing and reading help files for bash programs
   ○ Writing files, writing shell scripts, converting a script into a program
   ○ Basic data analytics using Shell
   ○ Using pipes, grep, working with text
   ○ Moving and downloading data, transferring data between a local machine and the Cloud.
   ○ Introduction to HPC, file systems, bash scripting, tools, schedulers e.g. Slurm and PBS.

Module learning outcomes

On completion of this module, the participants will have a basic knowledge of, or will be able to:
a. Introduction to Unix
   ○ Explain the vocabulary of Linux command lines.
   ○ Navigate linux systems.
   ○ Navigate files and directories
   ○ View, search and manipulate file contents
   ○ Use the nano text editor to modify text files.
b. Shell scripting:
   ○ Write a basic shell script, use the bash command to execute a shell script.
   ○ Use chmod to make a script an executable program.
   ○ Search files, redirect output.
   ○ Write loops using Basename.

Module assessments

Module practical: Practical available on the ASLM platform
Module quiz: Assessment questions available on the ASLM platform

Module resources

- Open UCT | ZivaHub - IBT Module 2: Linus Training Material Archive
- Bioinformatics Workbook | Command Line Basics and Useful Programs
- ZhikunWu GitHub | Bioinformatic Resources
- The Carpentries | The Unix Shell
- The Carpentries | Introduction to the Command Line for Genomics
- The Carpentries | Introducing the Shell
- The Carpentries | Writing Scripts and Working with Data
- SIB SWISS | GitHub - Training Collection: Collection of Bioinformatics Training Materials

References

- ClaudeAI. (2024). ClaudeAI response on Unix/linux Multiple Choice Questions Retrieved Aug 06, 2024.

Acknowledgements