MMB8052 - Practical 1

Introduction to the Linux Command Line

Simon J Cockell

2022-06-07

Contents

Introduction to the Module

Module Organisation

Welcome to *MMB8052 - Bioinformatics for Biomedical Scientists*. This module is mostly practical in nature - 10 computer lab sessions will introduce you to the fundamental computing tools used throughout much of modern bioinformatics, and will also provide case studies of the "read world" use of these tools. The lectures in the module will be delivered by a range of scientists from across the Faculty of Medical Sciences, and will highlight the *application* of bioinformatics in modern biomedical research.

Module Assessment

Due to the practical emphasis of the module, the assessment will also focus on these practical aspects. The module is assessed through two exercises - one short answer quiz in which the solutions to the posed problems should be derived computationally (more on this later). The second assessment will be a lab report style write-up of the application of some of the tools you will learn how to use in the practicals. Assessment 1 is worth 25% of the module mark. Assessment 2 makes up the remaining 75%.

In addition to these assessed, summative exercises, there will also be a range of formative tests throughout the module. These will mostly take the form of multiple choice quizzes, integrated into the practical sessions or provided separately on Canvas. While these components are not assessed, or compulsory to complete, they will aid and reinforce your understanding of the material presented.

About This Handbook

The practical sessions will each be accompanied by a handbook like this one. These handbooks will provide you with a lot of background information relevant to the practical at hand, and will provide you with walk-through instructions for what you are supposed to do in each session. Computer code (usually intended to be typed in to the appropriate computational interface - we'll get to what this means later) will be presented in chunks styled like this:

```
# this are some command line instructions:
$ pwd
/home/student
$ ls -l -h ~
```

Exercises, which will direct you to accomplish some computational task, will be presented like this:

Estimated time: 2 mins

Exercise 1

The instructions to follow will be in this block of text.

- First instruction
- Second instruction
- etc.

Linux

Linux is an umbrella term which describes a family of open-source, “Unix-like” computer operating systems which are based on the Linux kernel. The first Linux operating systems were released in the mid-90s and today they are used throughout computing, particularly in the infrastructure which runs the World-Wide Web, and in scientific computing and virtually all supercomputers. The Android smartphone operating system is a Linux system. Popular Linux distributions:

- Ubuntu
- Fedora
- Debian
- MX Linux

The Kernel

A kernel is the computer program at the heart of an operating system (OS), and is responsible for the control of everything in the system. The kernel facilitates interactions between hardware and software (via *drivers*) and optimises the use of system resources such as CPU time and RAM usage. The main kernels in use in modern computing are the Linux kernel, the Windows NT kernel and the MacOS kernel. All of the Linux distributions listed above use the same kernel at their core.

Unix

The Unix operating system is ancient, in computing terms. It was conceived and implemented in 1969 at AT&T's Bell Labs. It is modular by design, with a number of robust tools each designed to perform a limited, well-defined function. A program, known as the Unix "Shell" provides a text-based interface to these tools, and allows the user to combine them in order to perform complex workflows. Thanks to its efficient and robust nature, this computing paradigm persists today in the modern, Unix-like operating systems, Linux and MacOS.

> This is the Unix philosophy: Write programs that do one thing and do it well. Write programs to work together. Write programs to handle text streams, because that is a universal interface.
>
> – Doug McIlory (Bell Labs)

The Shell

A Unix shell is a command-line user interface for Unix-like operating systems. It provides a programmable environment for controlling the OS, and running executable programs. It is typical for the user of a Unix-like OS to interact with the shell via a *terminal emulator* - a program which simulates the features of a hardware video terminal interface. Feature-rich terminal emulators are available for all modern operating systems.

There are many different shell programs, all of which have different features. The shells you will encounter most often are bash (the "Bourne-Again Shell") which is the default in most major Linux distributions, and zsh (Z Shell) which is the default in MacOS. Zsh is backwards compatible with bash (so any code written for bash will work in zsh, but not necessarily vice versa).

Logging in to a Linux Server

Rather than install Linux directly onto your desktops, we will use a terminal emulation program to log into a virtual cloud server which has been configured for this practical. It should be noted that you'll be using your computer as a dumb terminal (client) that will display the information generated by programs running on the cloud VM (this is an example of a *client-server* model).

You should have received an email from Microsoft Azure, inviting you to register for the lab - click the "Register for the lab" button in the email, and log in to Azure Labs with your University credentials.

Once logged in, you should see a heading "My virtual machines", with a single entry underneath. Click the slider in the bottom-left of the VM box to power the machine on. Once it's started up (this will take a couple of minutes), click the icon in the bottom-right of the VM box which looks like a small grey monitor. Pick “Connect via SSH” from the menu which appears, and set your password when prompted. Once the password has been set (again, this takes a couple of minutes) select “Connect via SSH” again, copy the text in the popup box which appears - this is your SSH invocation to log in to your VM. For example:

```
student@ml-lab-77568ef7-c936-416a-a101-
 ↪  5e2874043ea1.uksouth.cloudapp.azure.com
```

Logging in from Windows

Estimated time: 2 mins

Exercise 1a

- In the “Port” box, type the number given in the −p argument of your SSH invocation (this is likely to be a large number, greater than 60000 - from the example above 65432)

PuTTY will then open and prompt you for a password. Enter the password you set when starting your VM for the first time.

Logging in from MacOS

Estimated time: 2 mins

Exercise 1b

- Open “Terminal.app” (located in /Applications/Utilities/Terminal.app).
- Paste in the connection string shown when you click “Connect via SSH” on the Azure Lab (above). As per the example above:

```
ssh -p 65432 student@ml-lab-77568ef7-c936-416a-a101-5e2874043ea1.uksouth.cloudapp.azure.com
```

- Enter your password when prompted.

The Linux Filesystem