Computing Basics for Bioinformatics on the Cloud

Session 2

Yaoyu Wang

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Overview

- Introduction to cloud computing
- Get on to the cloud
- Get around on the cloud with linux
- Free bioinformatics resources on the cloud

- Download slide at:

http://tinyurl.com/cccbtrain

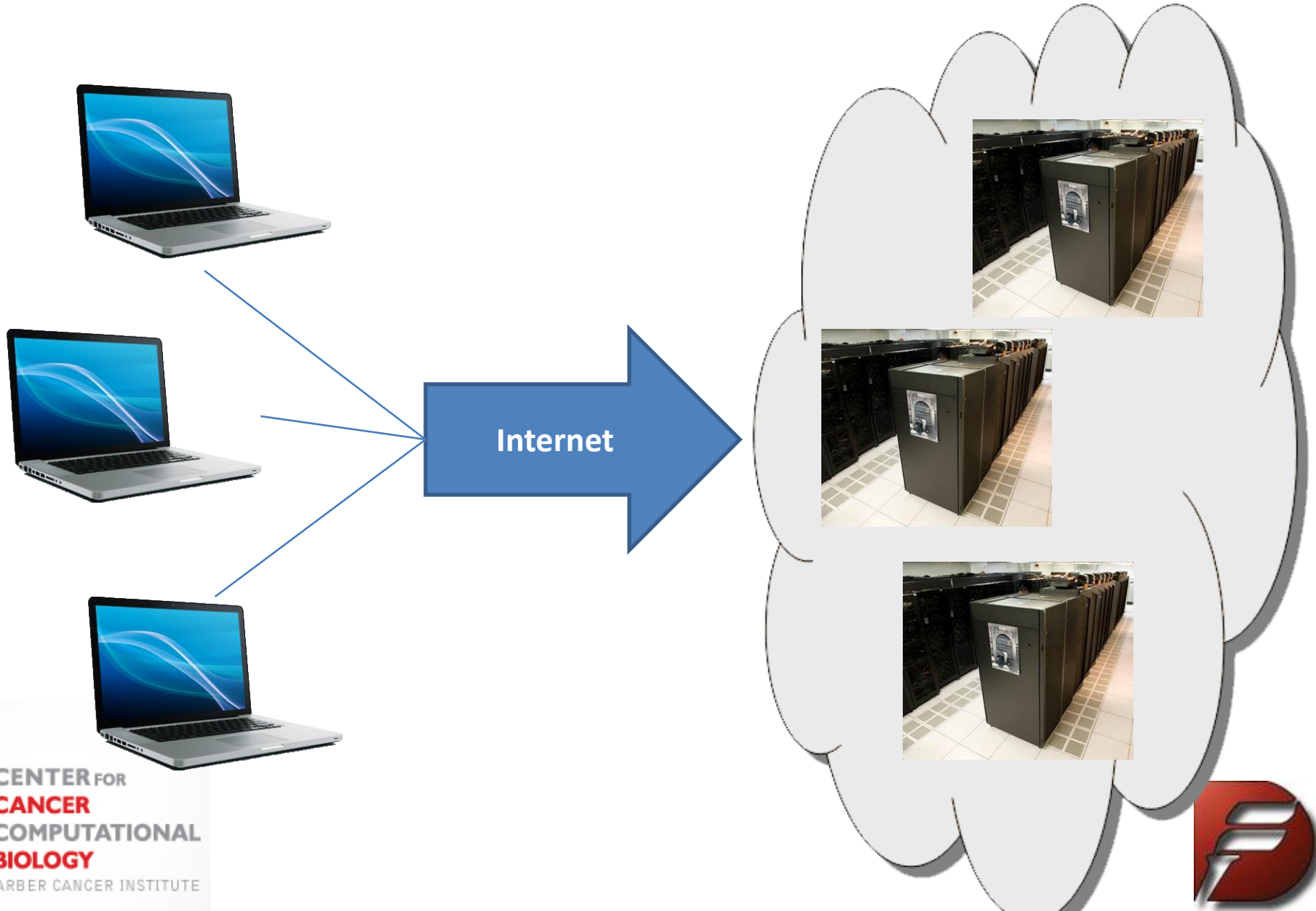
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

What is Cloud Computing?

Cloud Computing is a general term used to describe a network based computing that takes place over the Internet

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Overly Simplistic Cloud Computing Schema



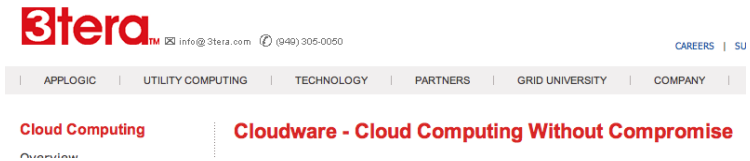
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Cloud Characteristics

- The "flexibility and elasticity" allows these systems to scale up and down at will – utilising the resources of all kinds (CPU, storage, server capacity, load balancing, and databases).
- The "pay as much as used and needed" type of utility computing and the "always on!, anywhere and any place" type of network-based computing.
- The "no-need-to-know" in terms of the underlying details of infrastructure, applications interface with the infrastructure via the APIs.

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Some Commercial Cloud Offerings



Amazon Web Services

- Elastic Compute Cloud (EC2)
  - Rent computing resources by the hour
  - Basic unit of accounting = instance-hour
  - Additional costs for bandwidth
- Elastic Block Storage (EBS)
  - Persistent storage
  - Charge by provision block
  - Additional costs for bandwidth

Cloud Computing

Quick-start an Amazon EC2 instance…

Cloud Computing

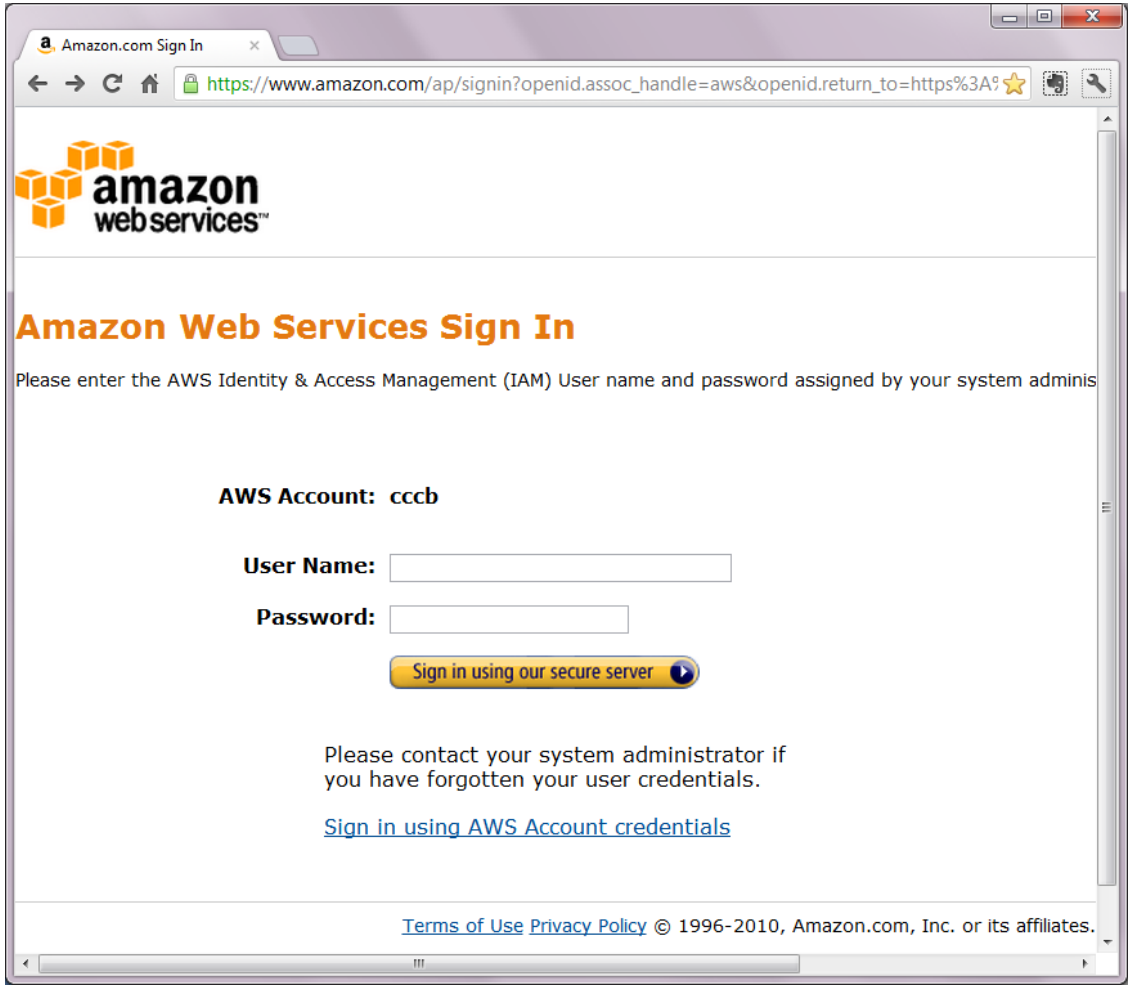
Running a server on Amazon Web Service (AWS) requires you to go through the following steps:

1. Create your account on Amazon (if you do not already have one) and setup your credentials
2. Install Amazon EC2 API Tools
3. Instantiate your **image** (i.e. your account)
4. Configure your **instance** (i.e. your virtual machine)

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Cloud Computing: Start the Machine

http://aws.amazon.com/

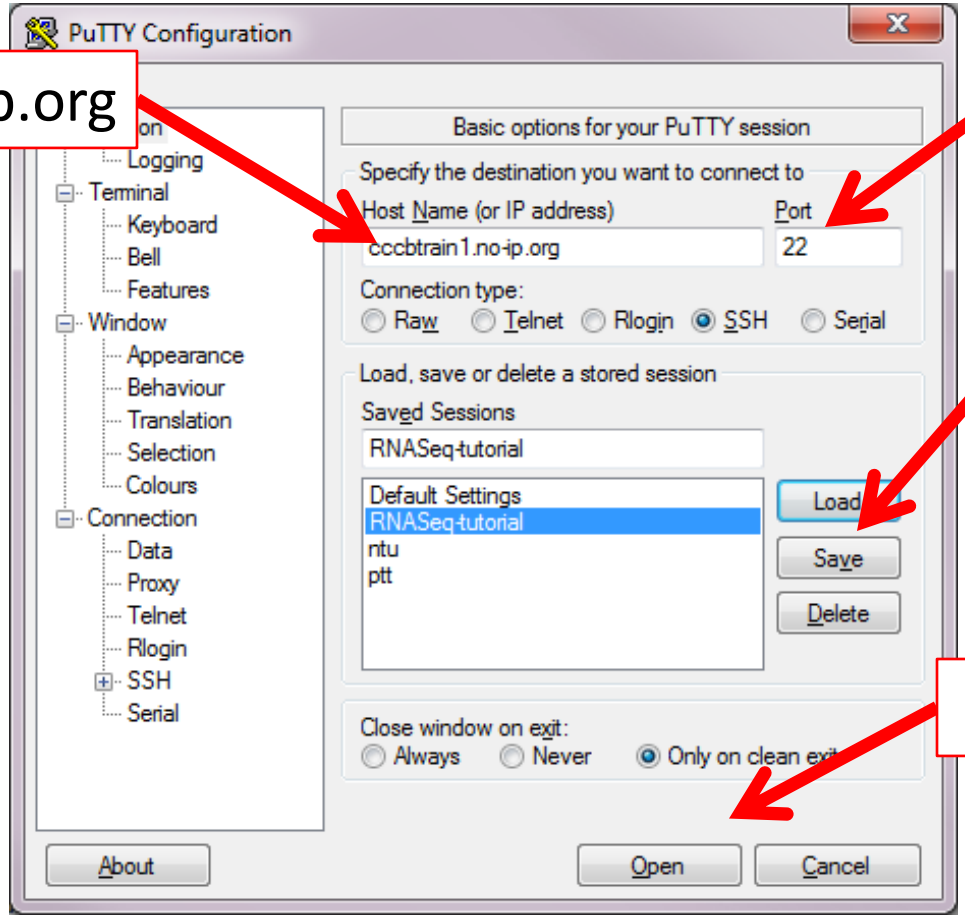


CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Cloud Computing: Logging in

For Windows

Download PuTTY at: http://tinyurl.com/putty-d

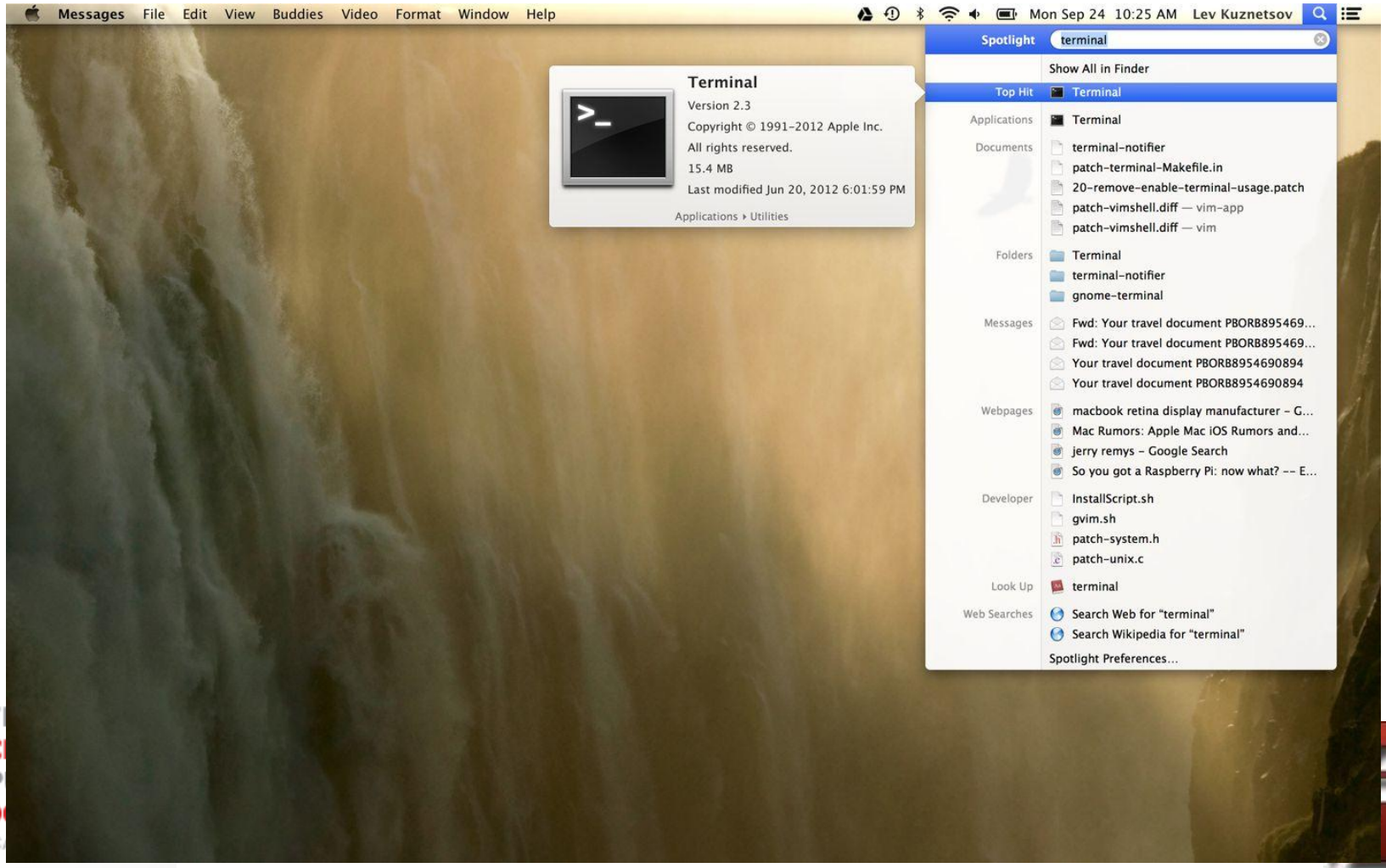


CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Cloud Computing: Logging in

For Mac

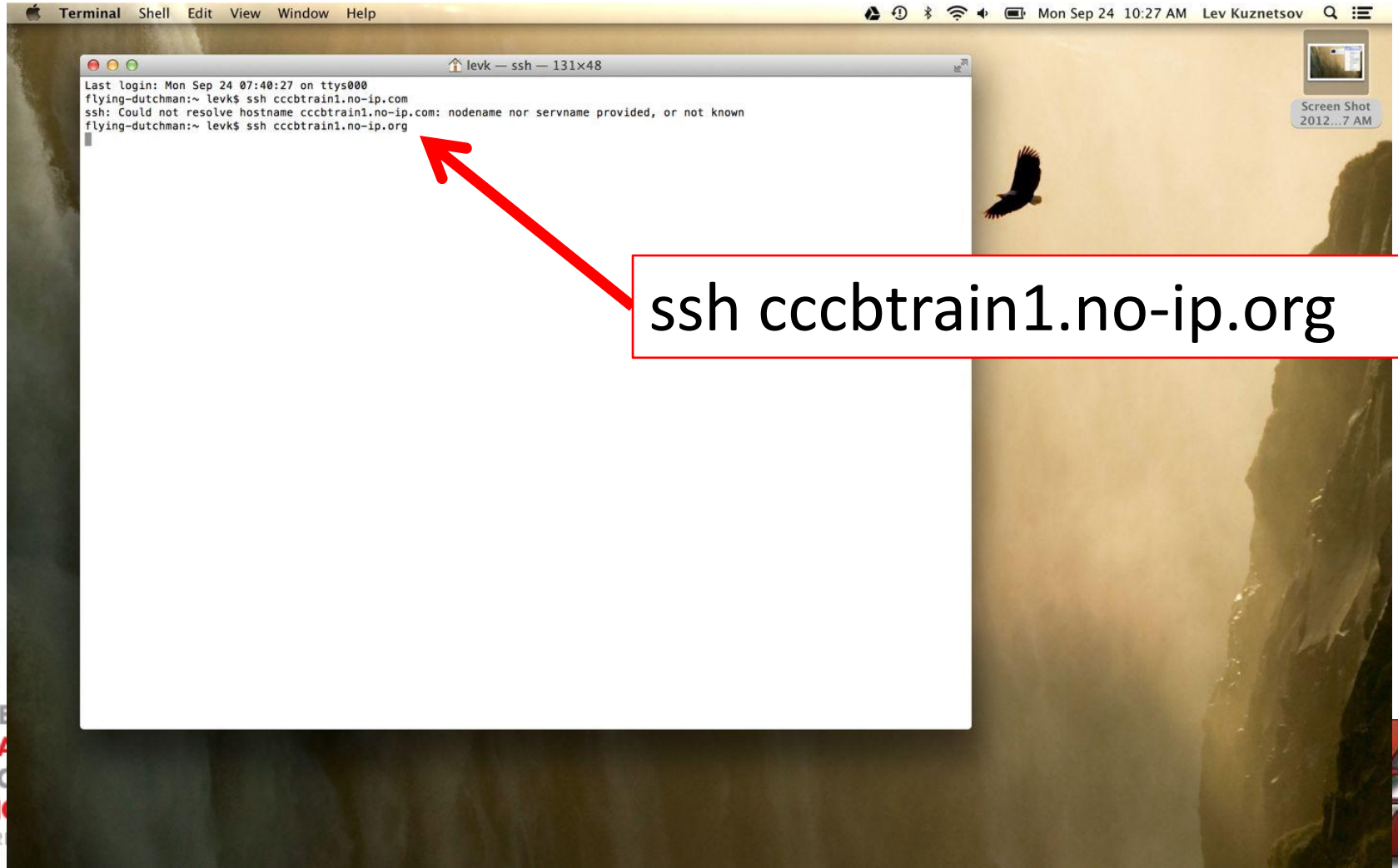
You already have the necessary program pre-installed



Cloud Computing: Logging in

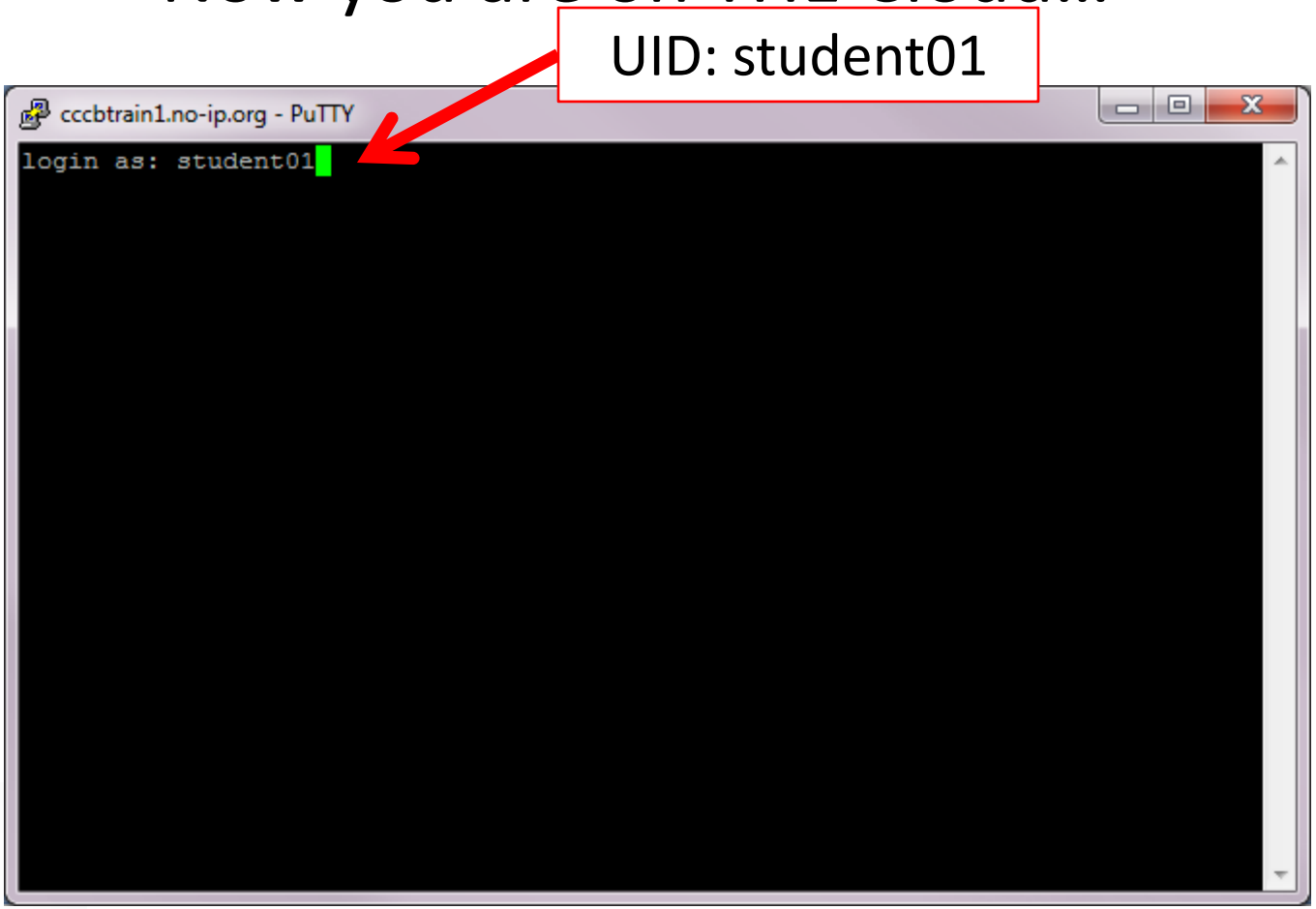
For Mac

You already have the necessary program pre-installed



Cloud Computing: Logging in

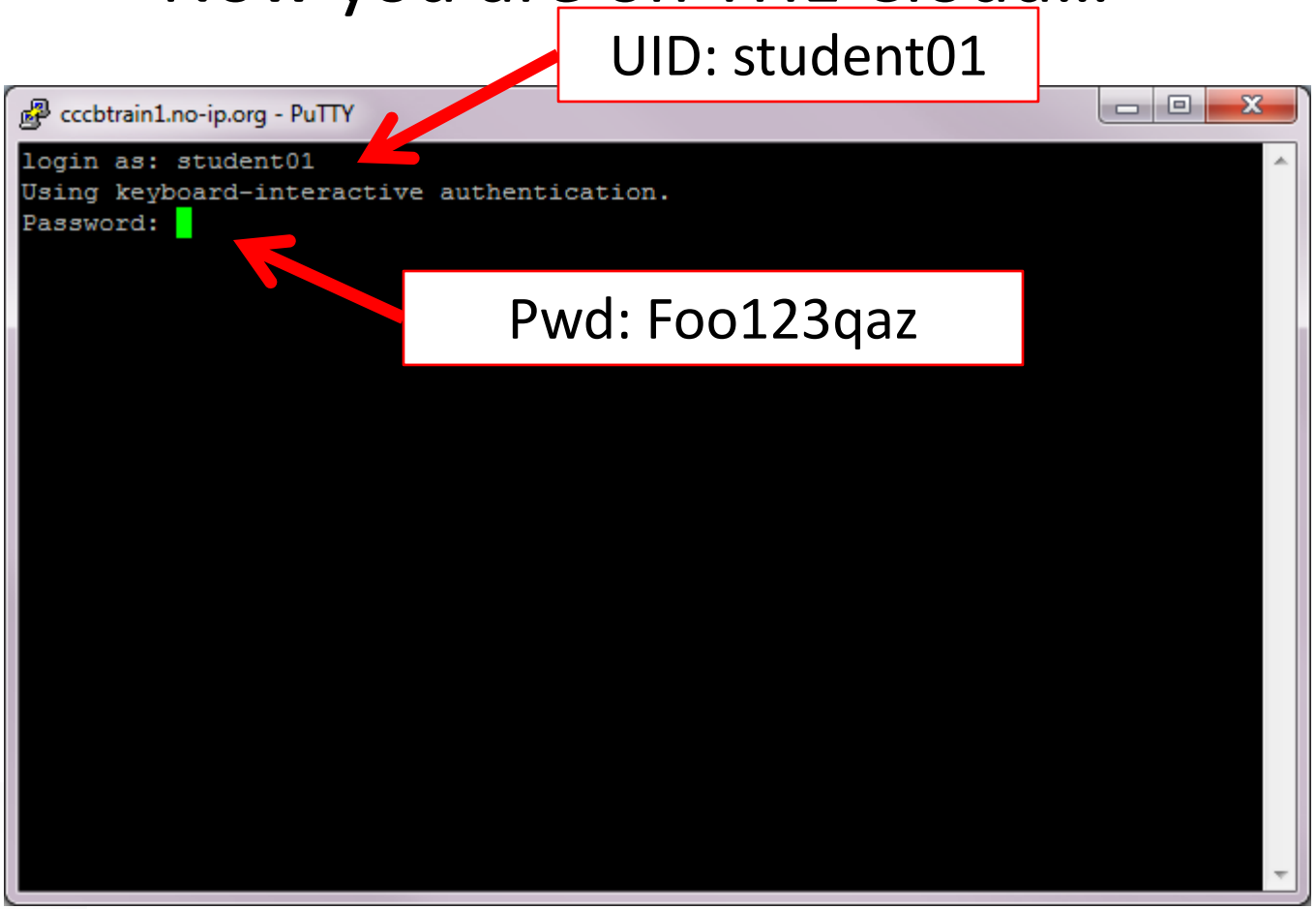
Now you are on THE Cloud…



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Cloud Computing: Logging in

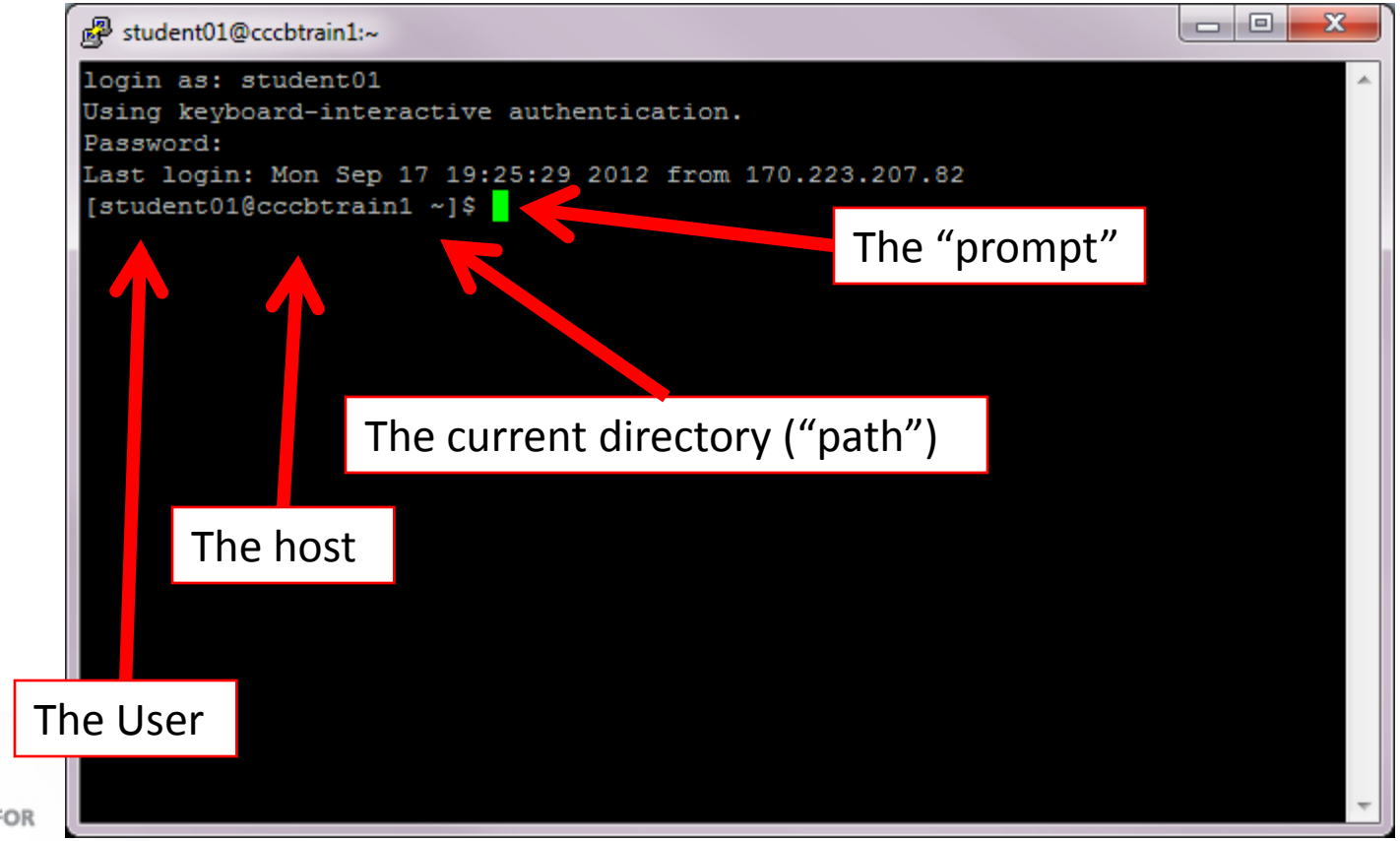
Now you are on THE Cloud…



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Cloud Computing: Logging in

Now you are on THE Cloud…



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Cloud Computing

How to navigate around the Cloud?

(a.k.a Linux Basics)

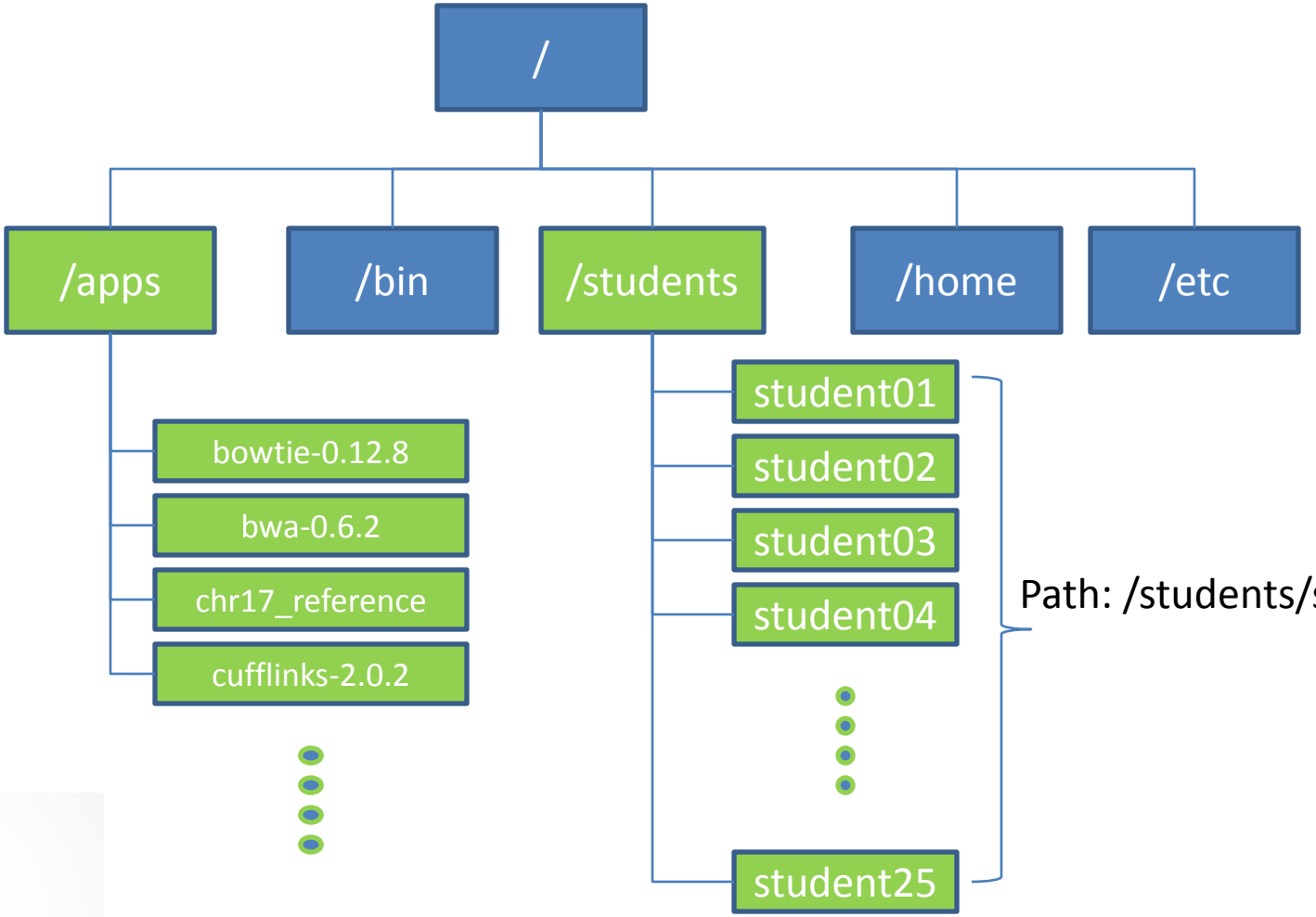
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

What exactly is a “shell”?

- After logging in, Linux/Unix starts another program called the **shell**
- The shell interprets commands the user types and manages their execution
    - The shell communicates with the internal part of the operating system called the **kernel**
    - The most popular shells are: tcsh, csh, korn, and bash
    - The differences are most times subtle
    - For this tutorial, we are using bash
- Shell commands are **CASE SENSITIVE!**

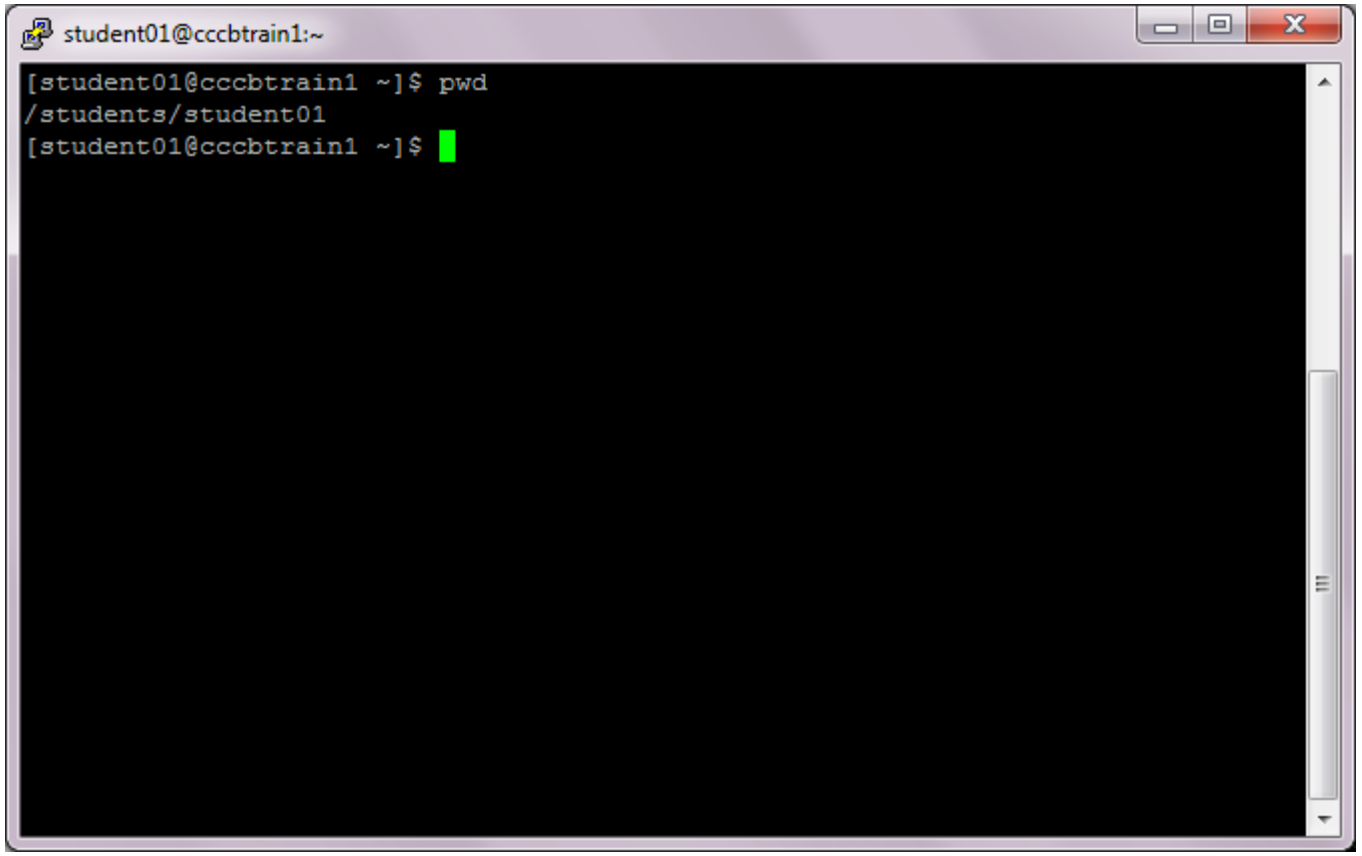
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

File System Organization



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

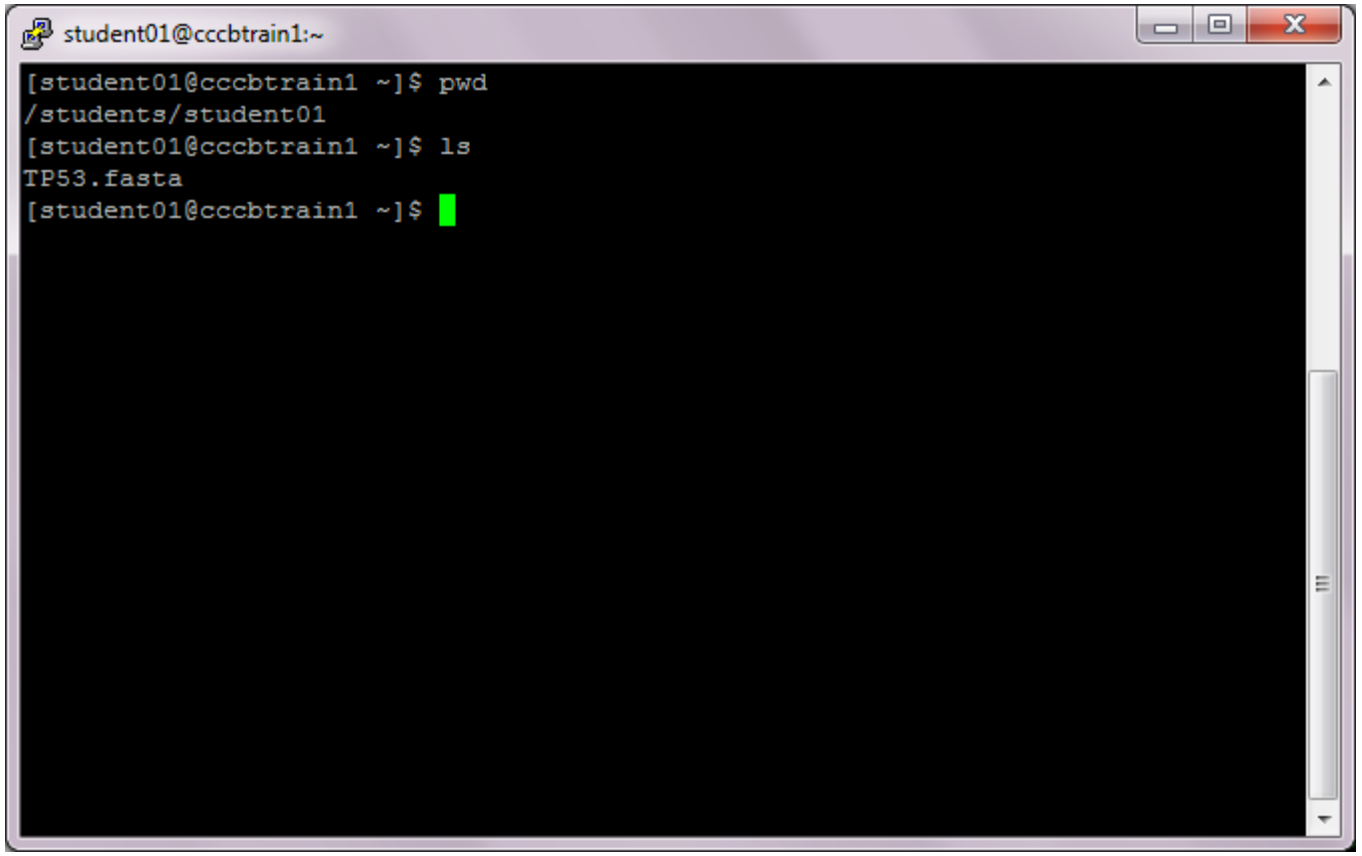
Basic Linux Command



pwd - print name of current/working directory

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

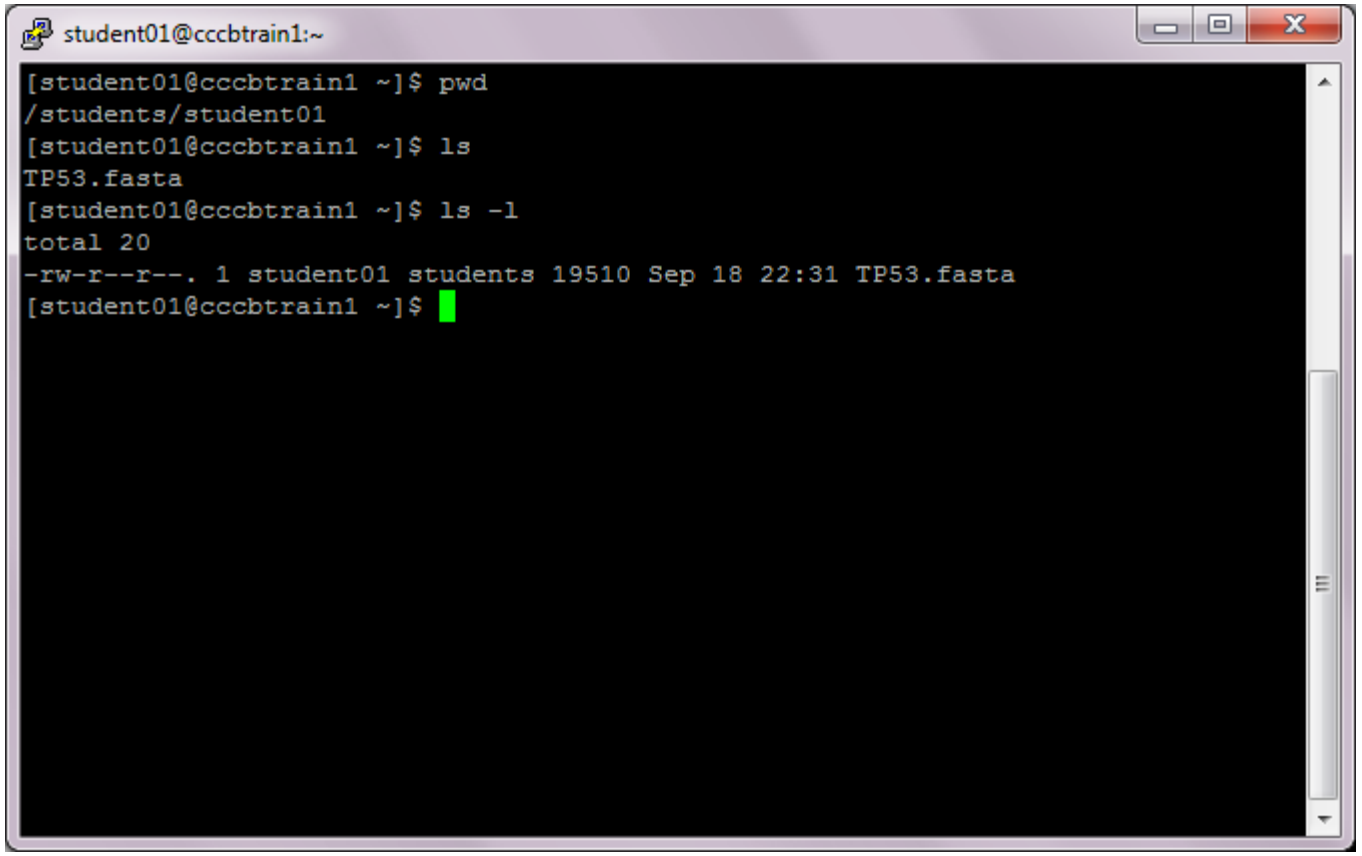
Basic Linux Command



ls - list directory contents

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

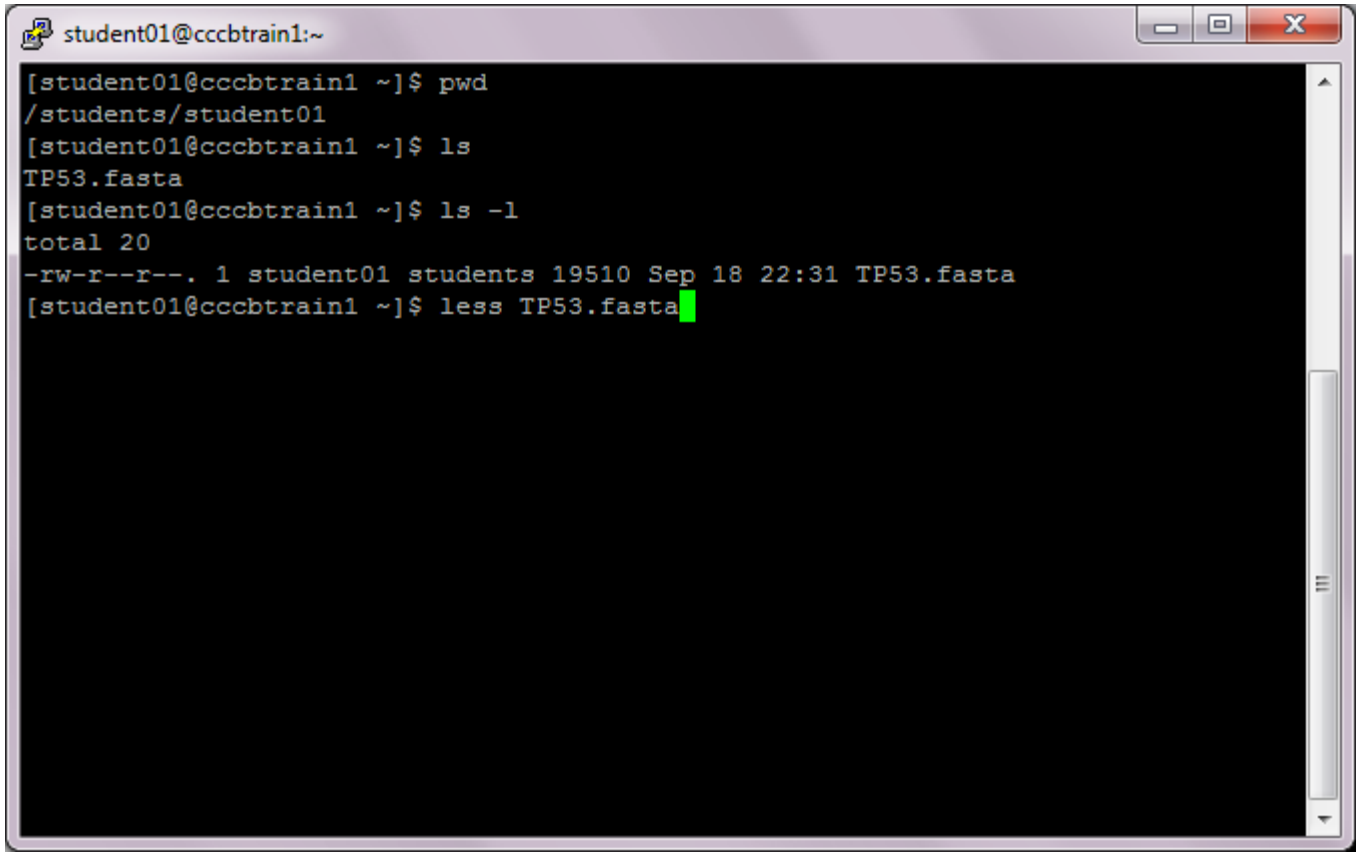
Basic Linux Command



ls –l use a long listing format

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

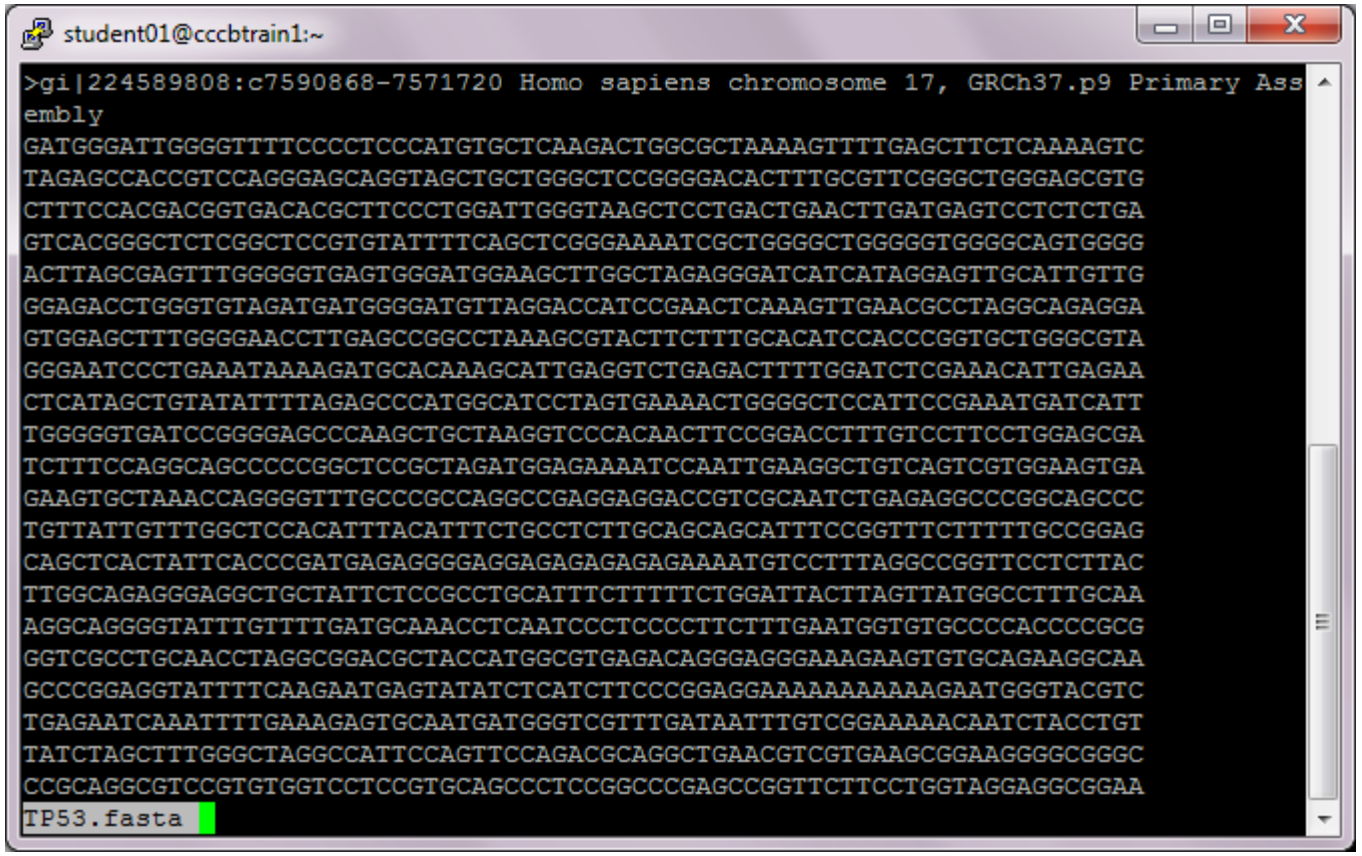
Basic Linux Command



less – quick file viewer

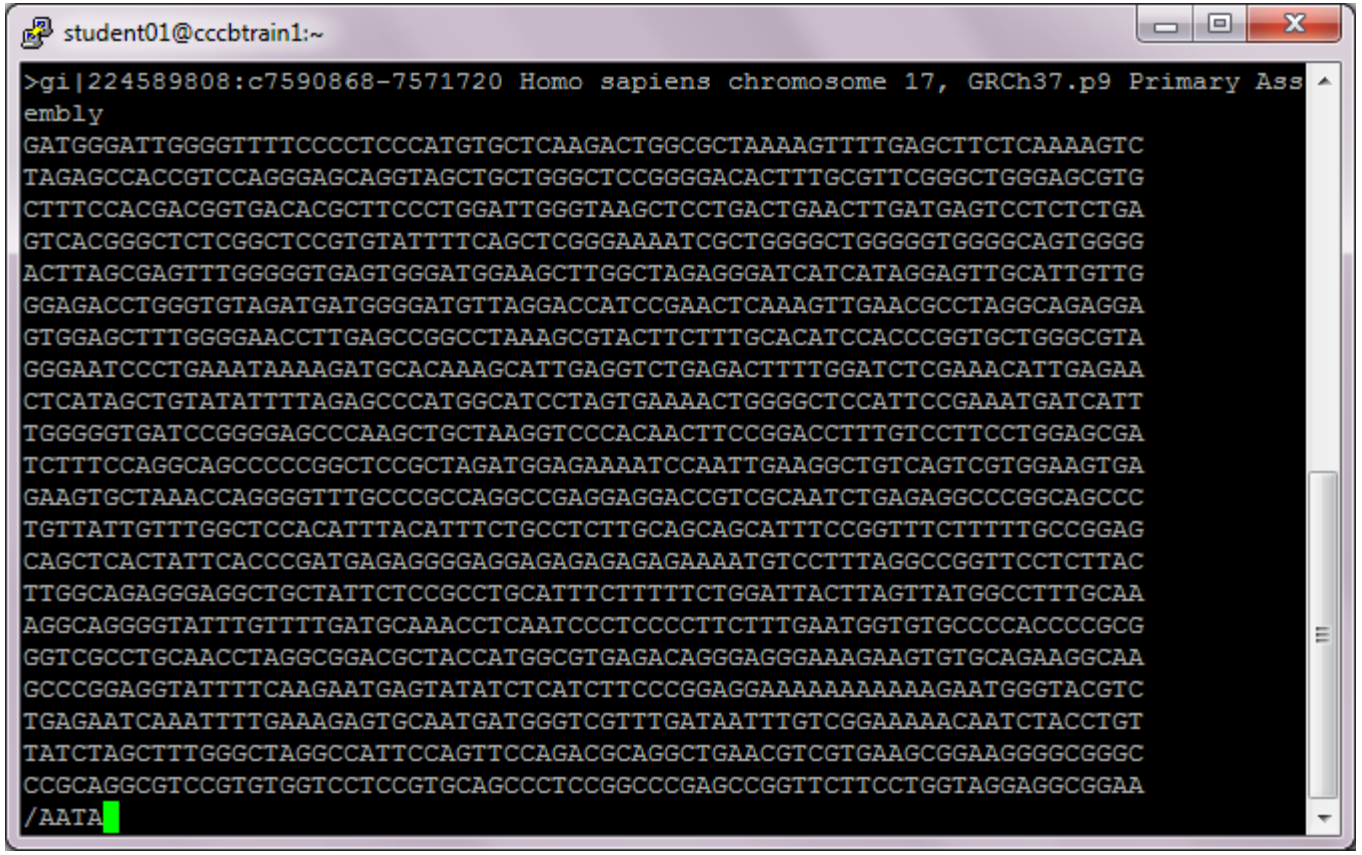
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

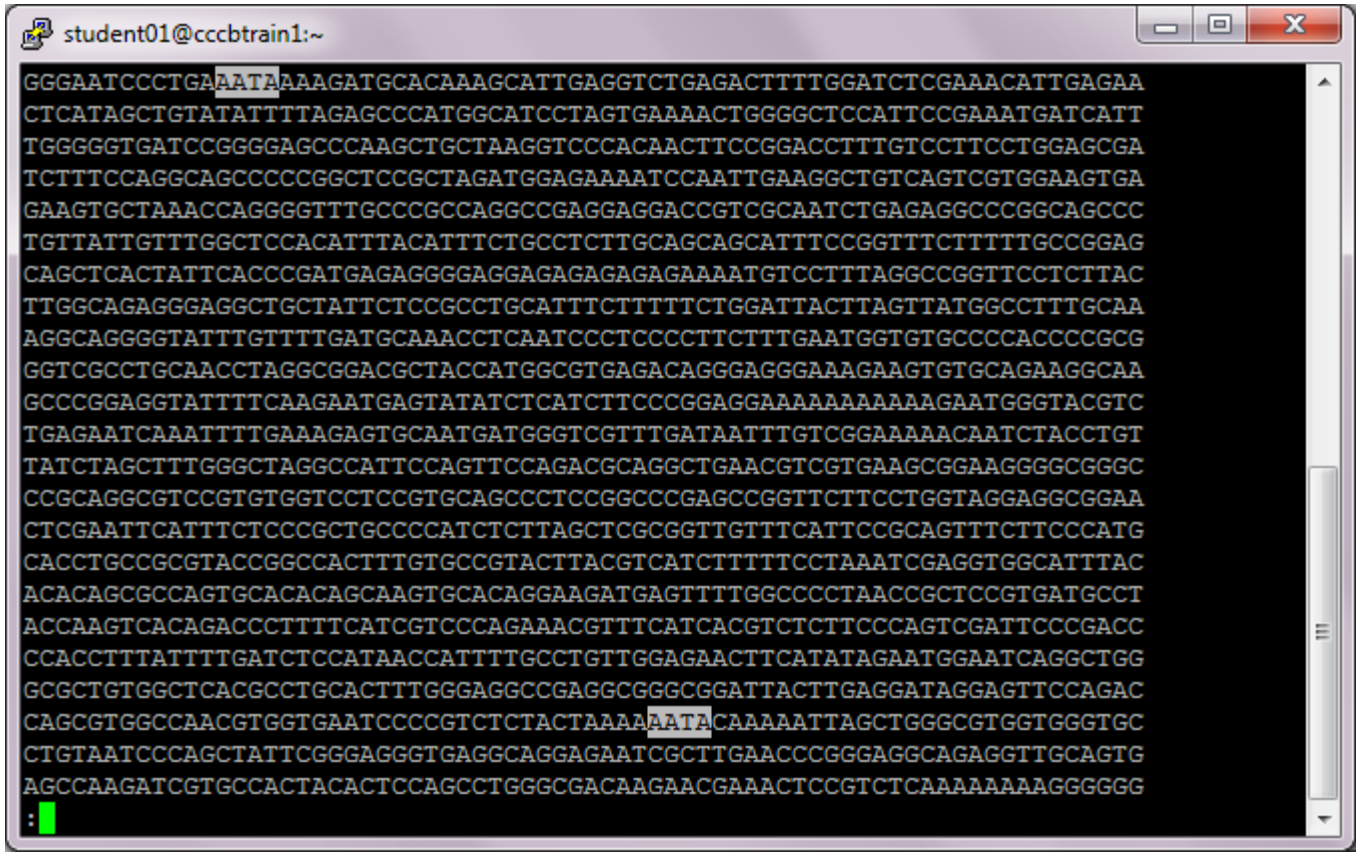
Basic Linux Commands



less – allows for simple text searching

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

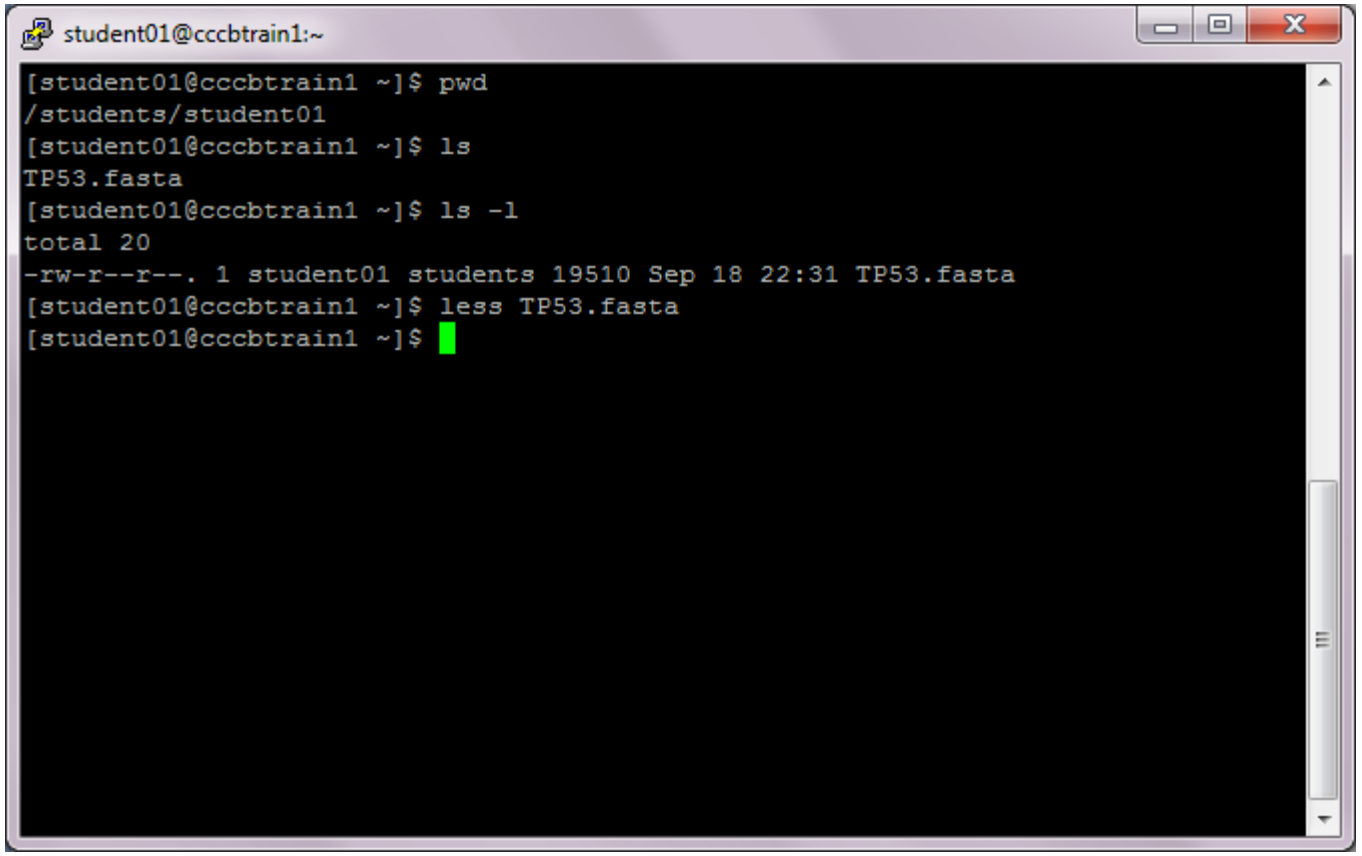
Basic Linux Command



less – allows for simple text searching

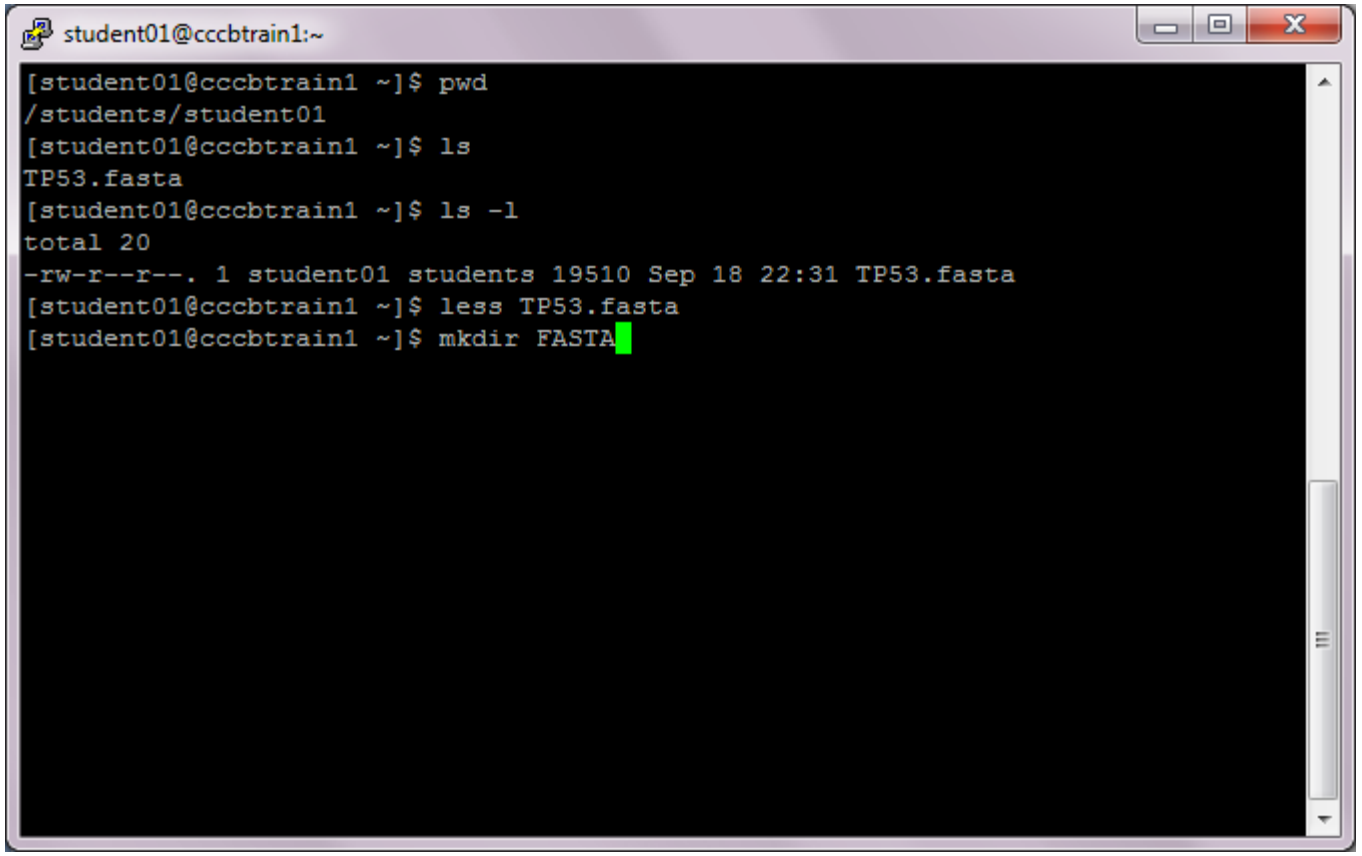
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

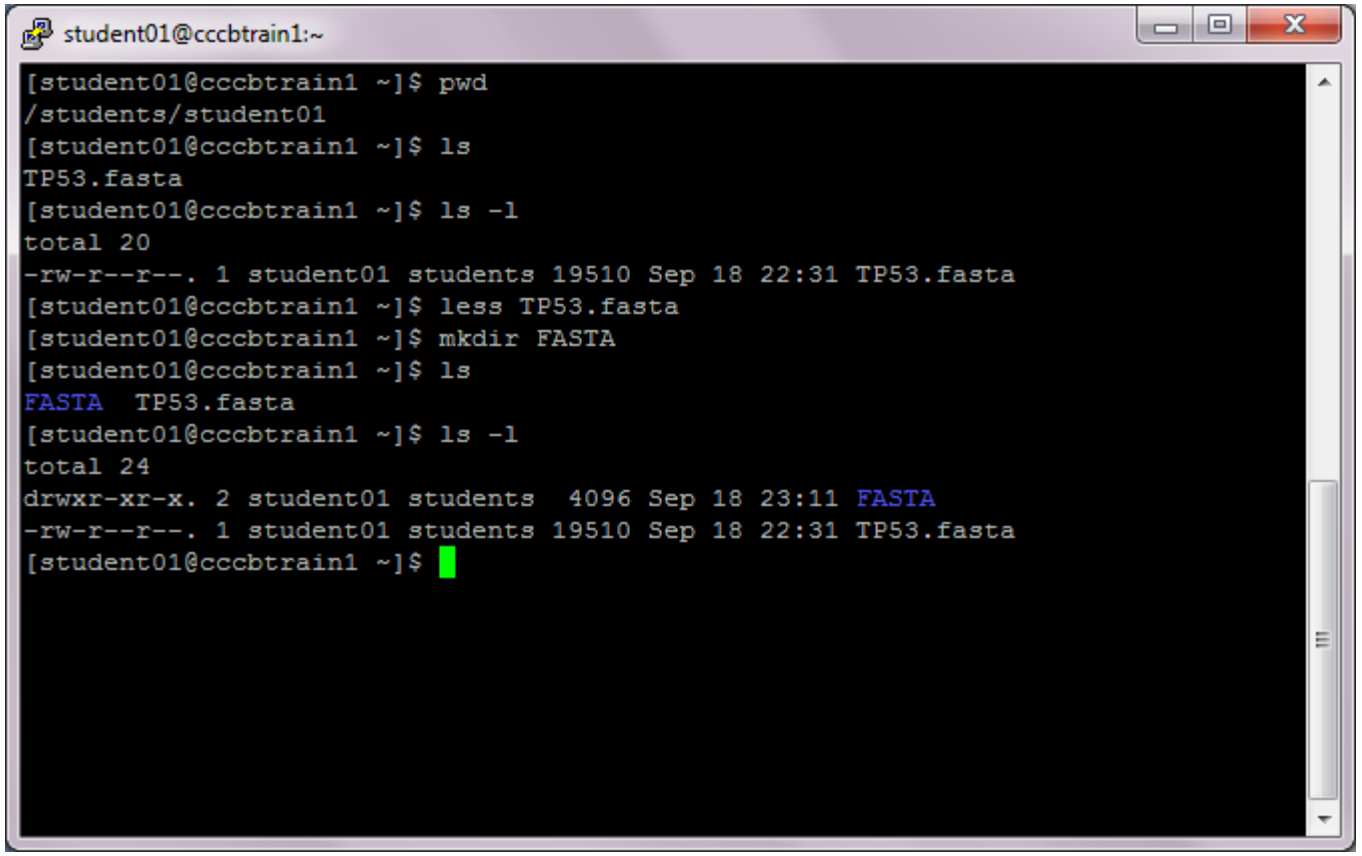
Basic Linux Command



mkdir – make directories

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

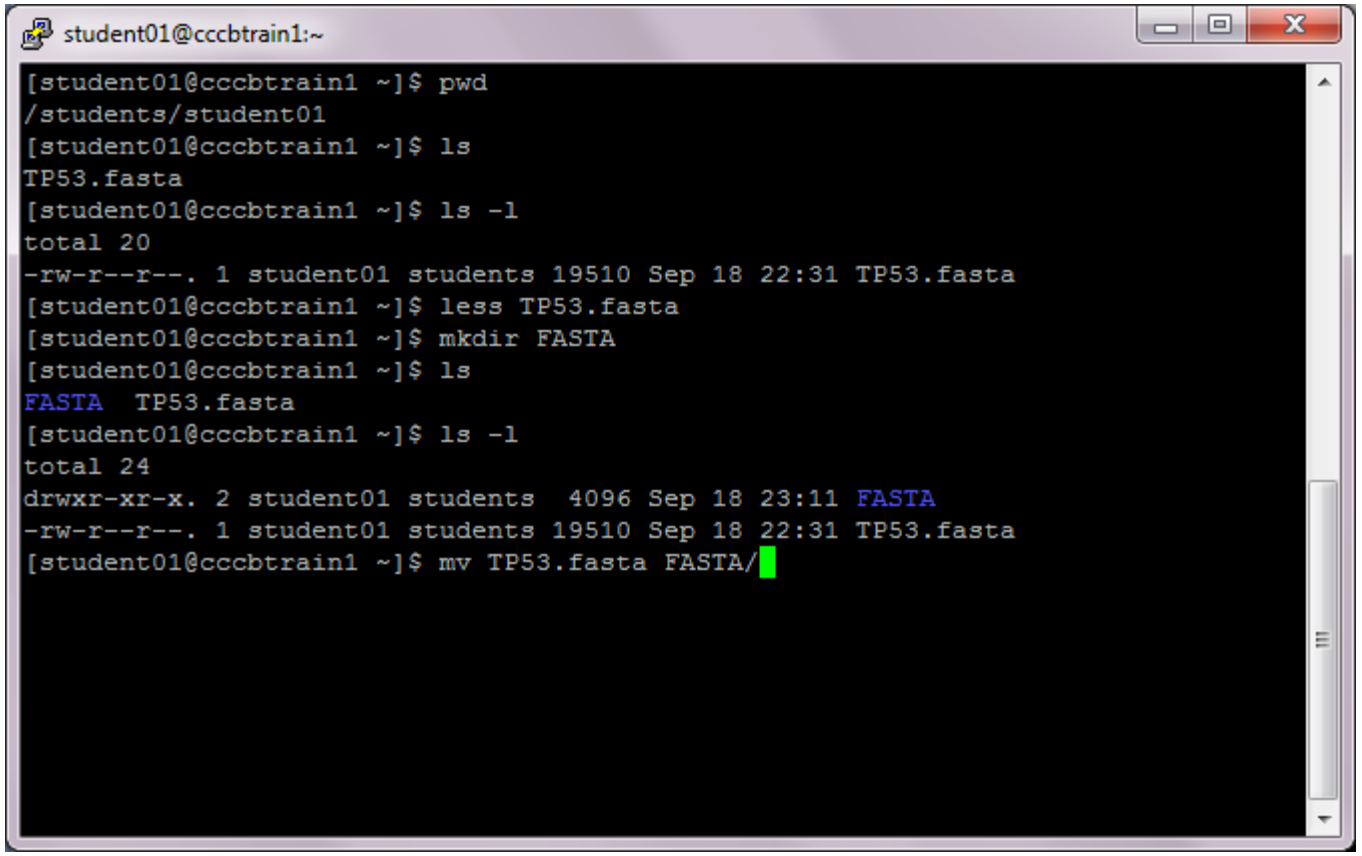
Basic Linux Command



mkdir – make directories

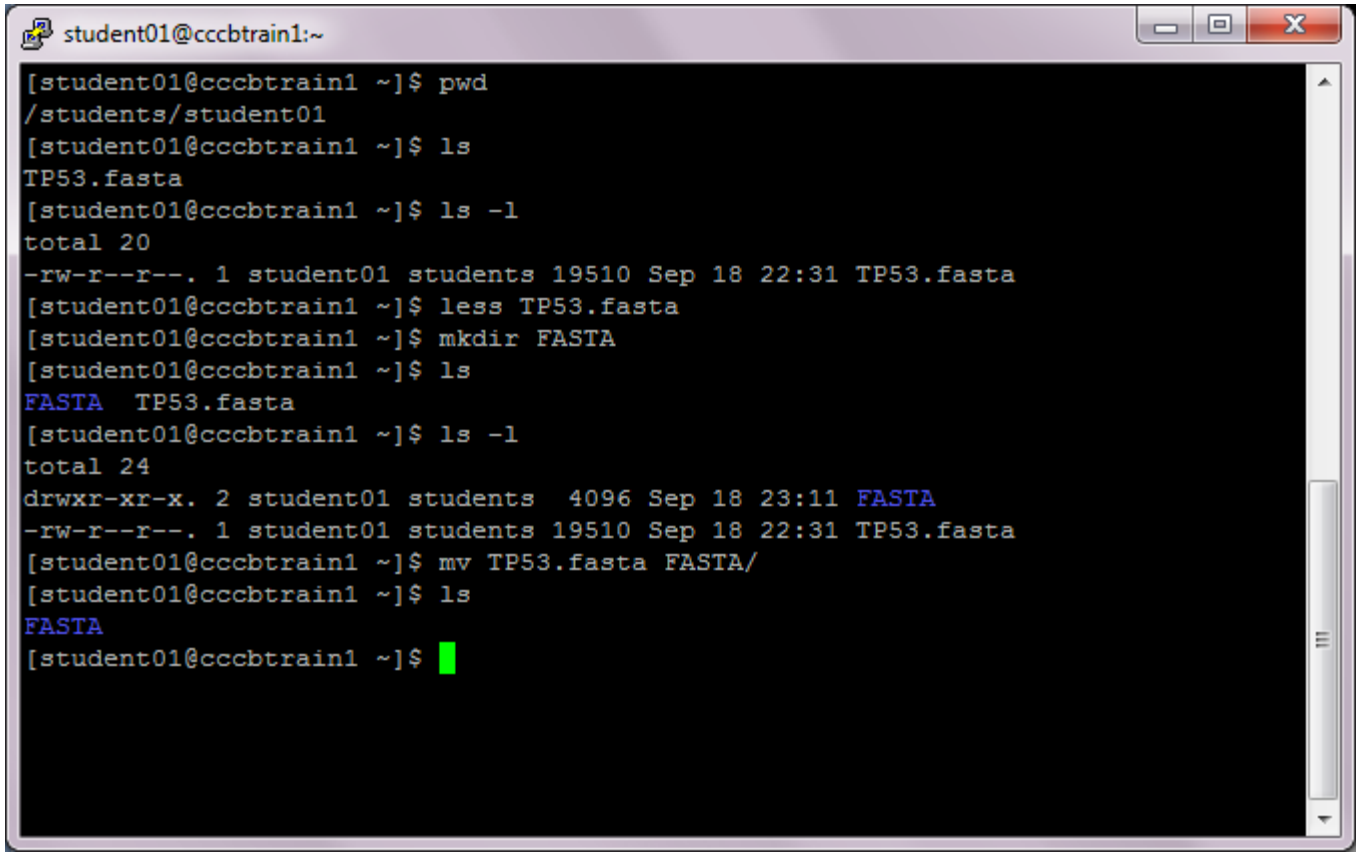
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



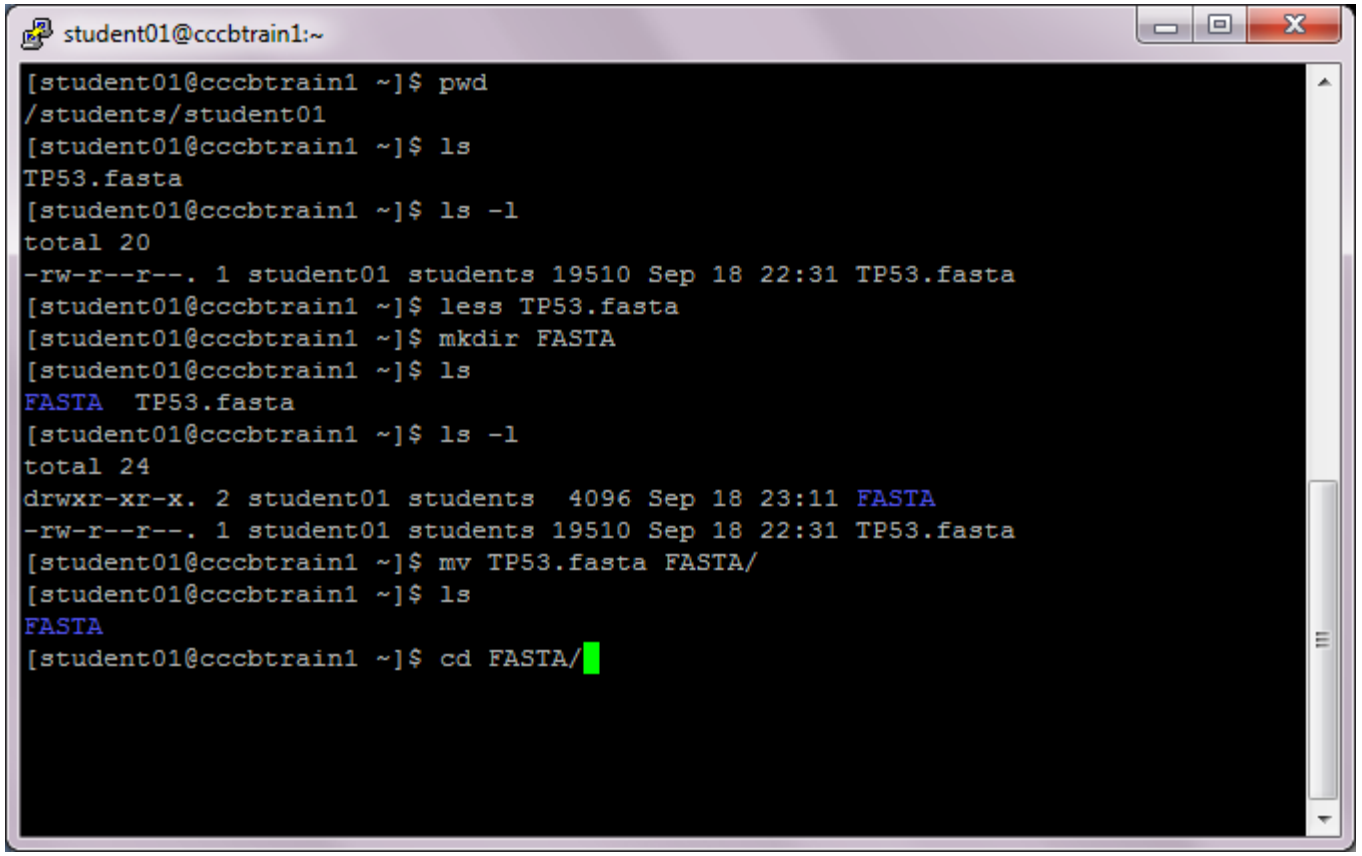
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



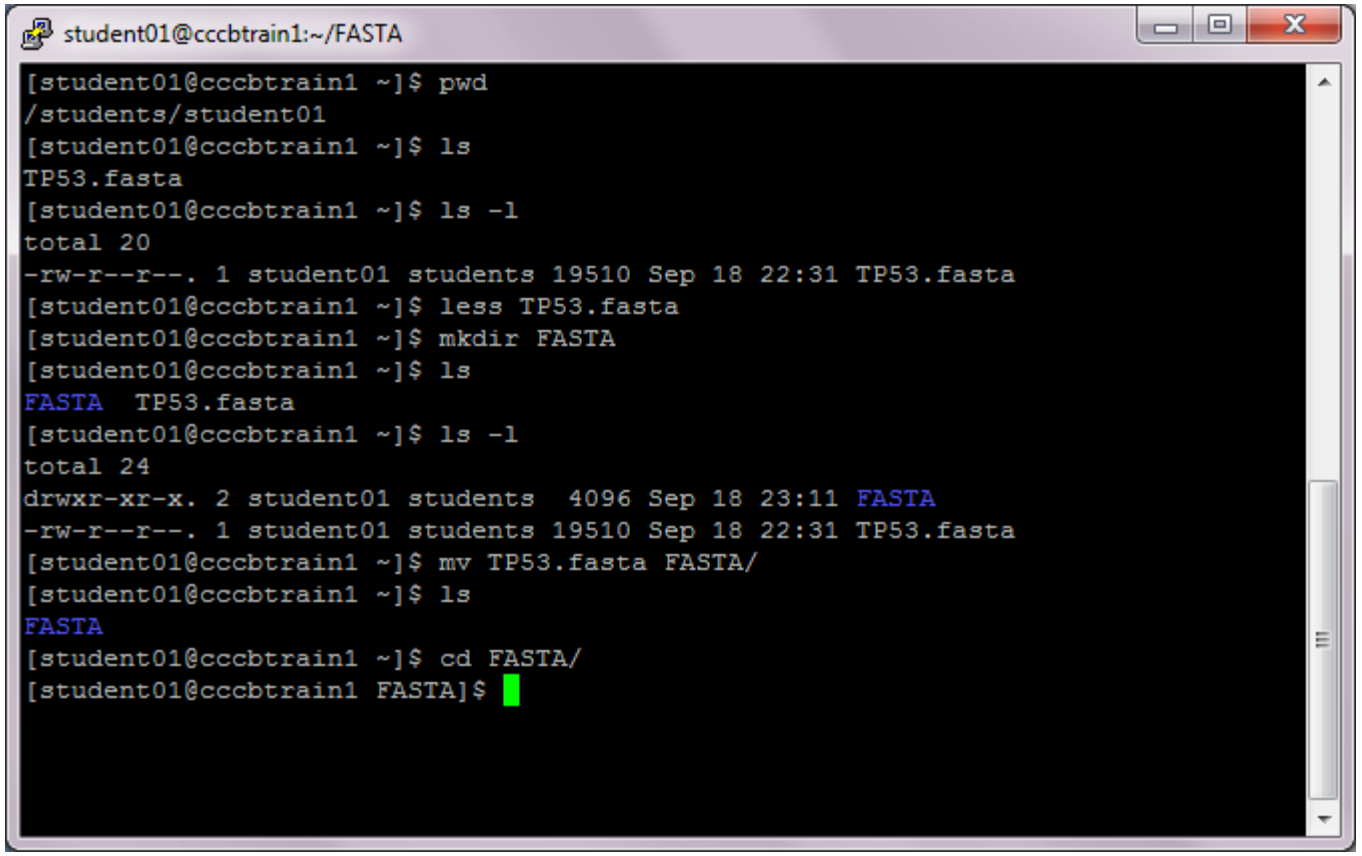
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



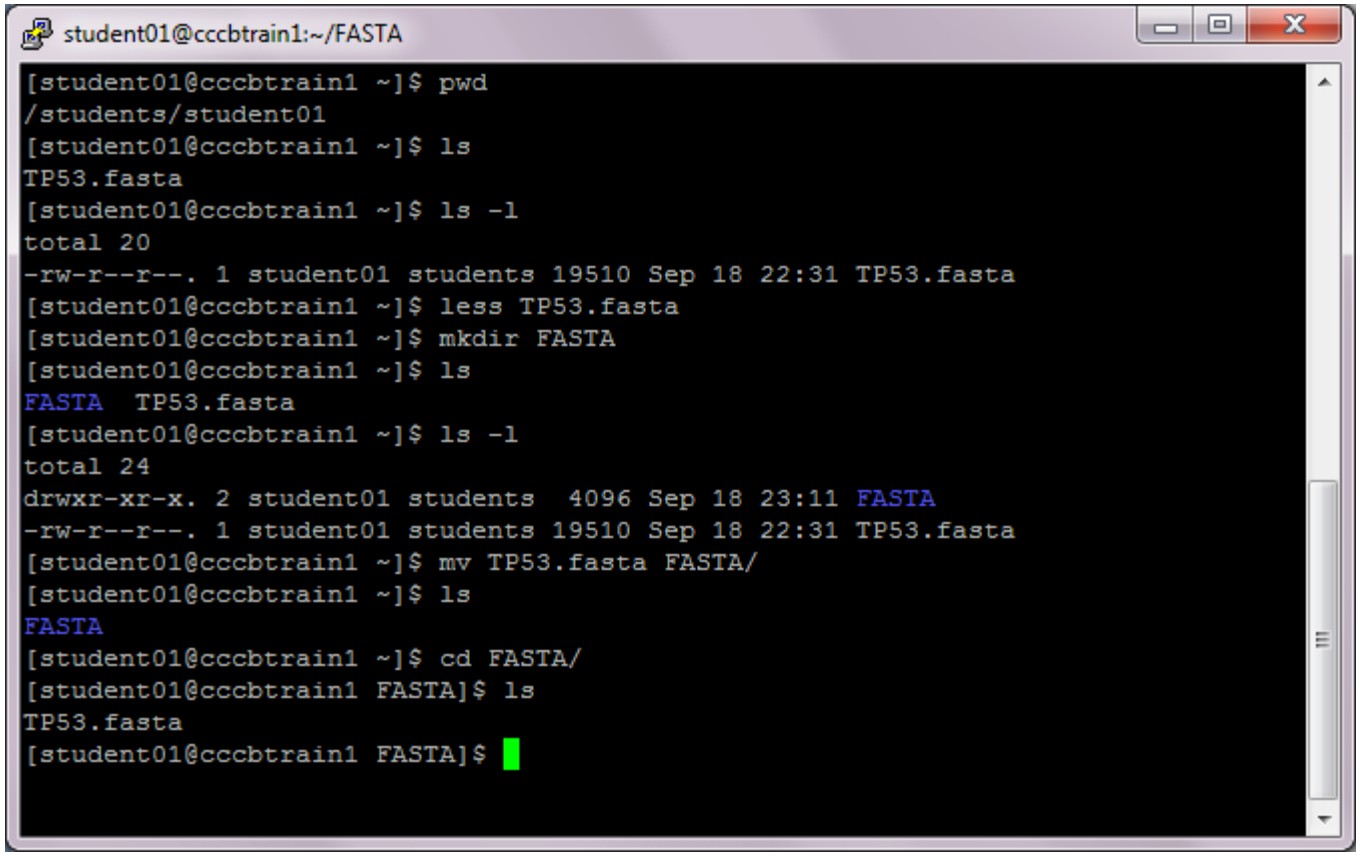
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



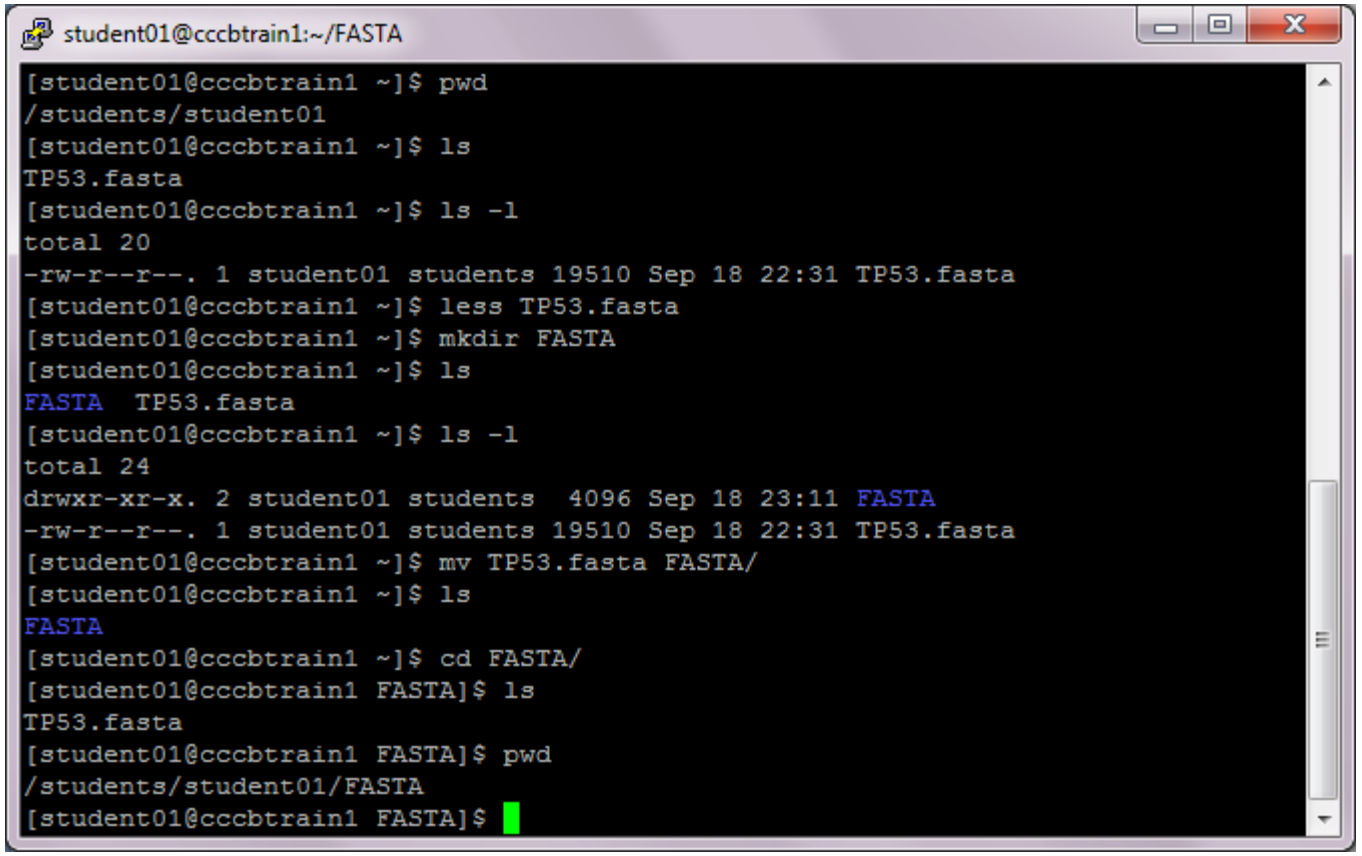
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command

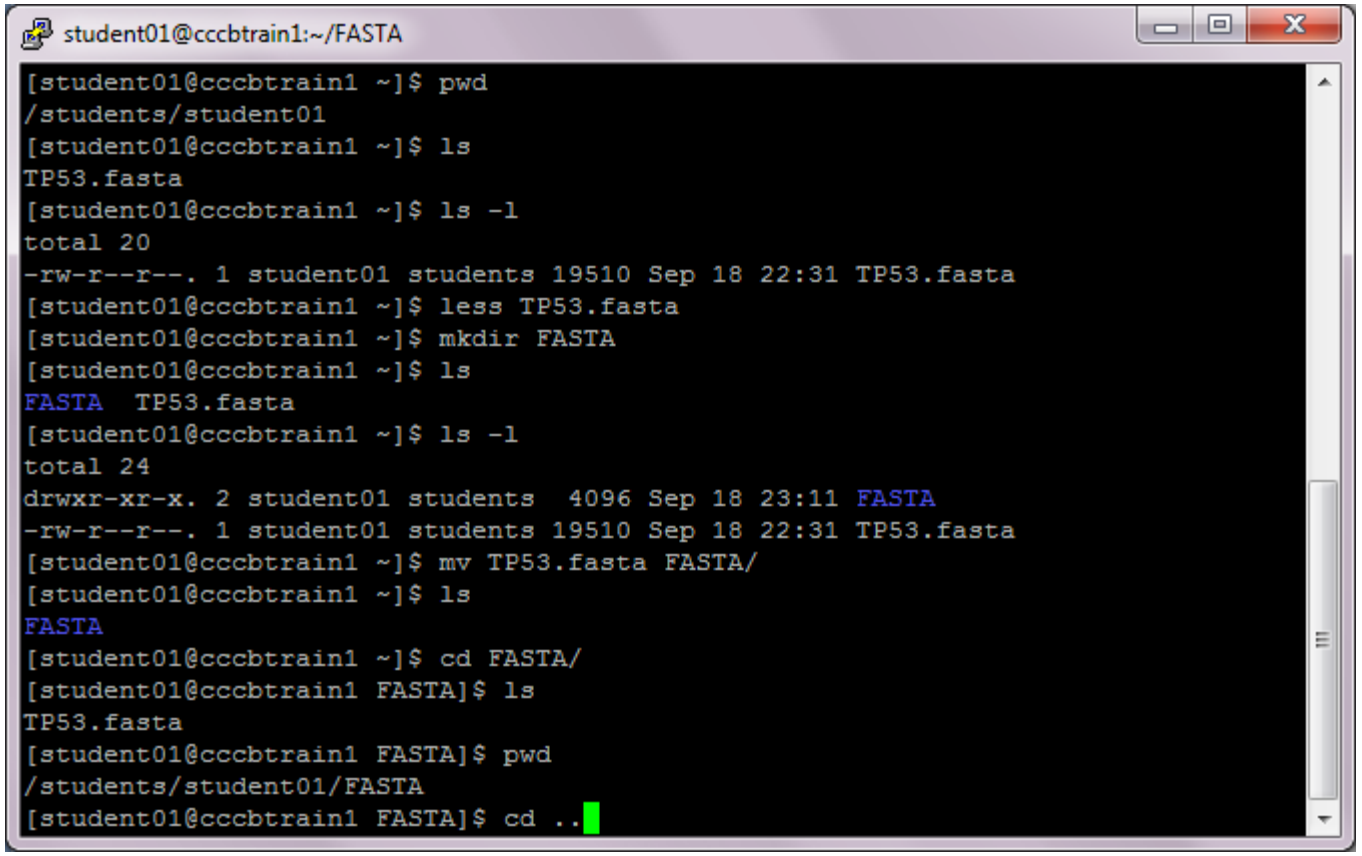


CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command

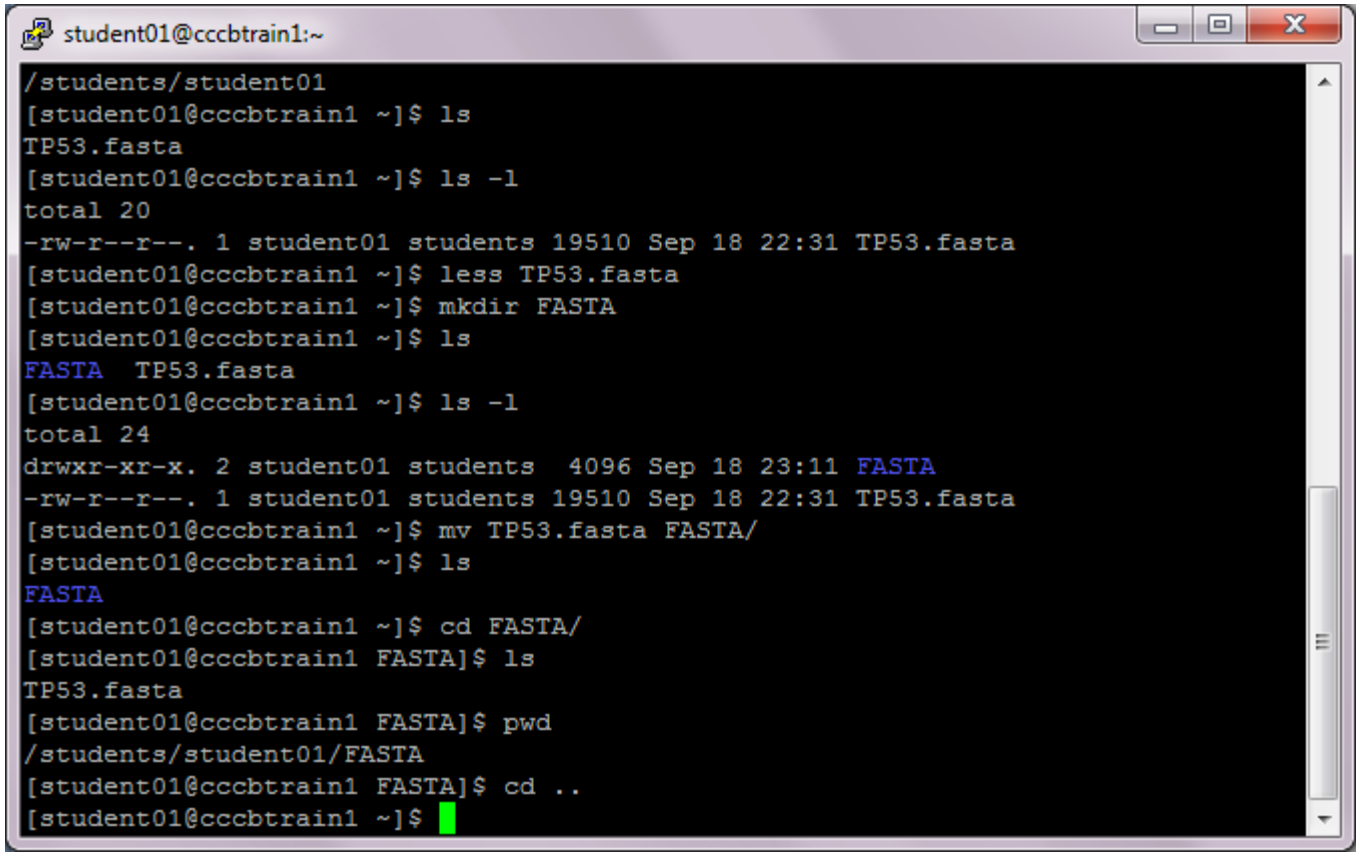


Basic Linux Command



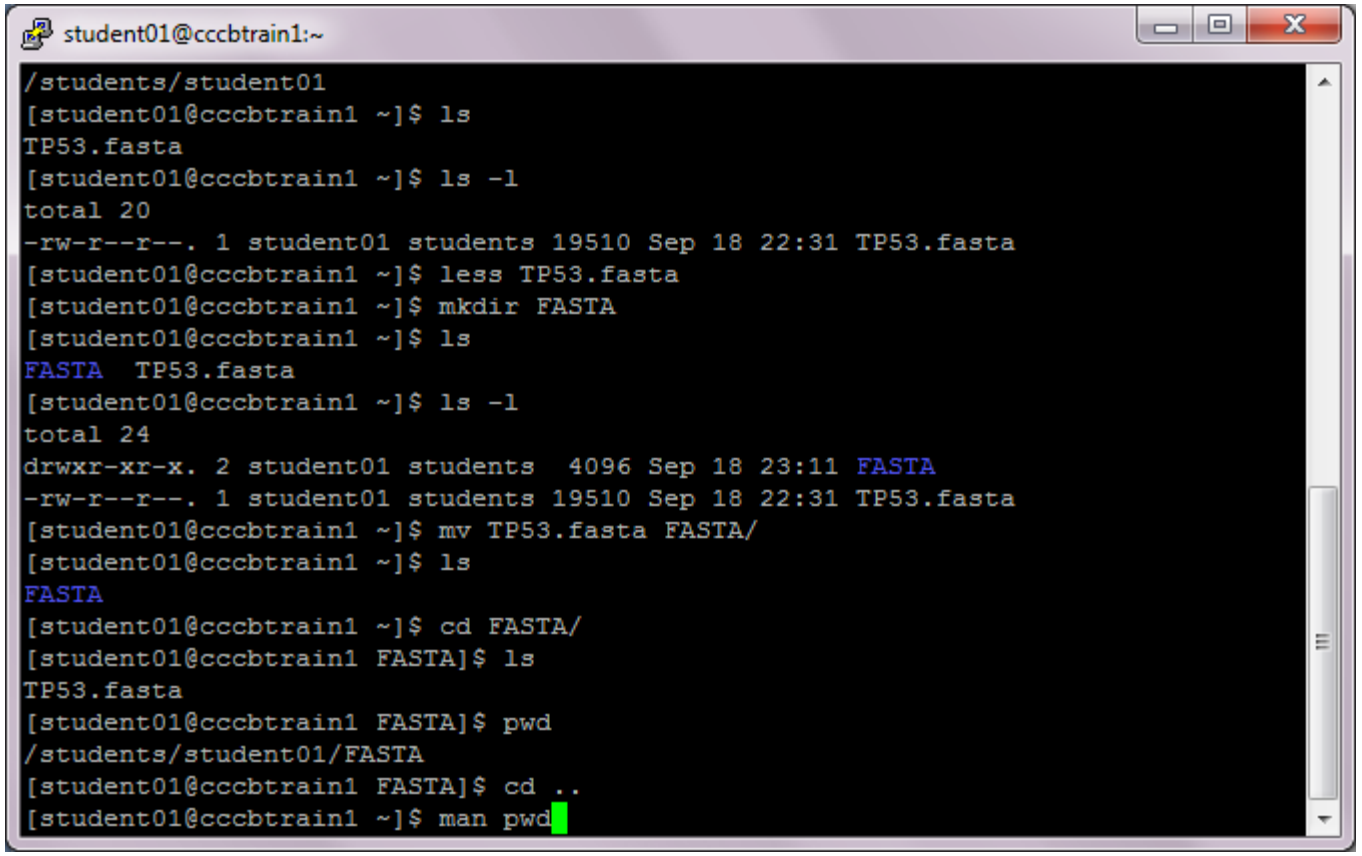
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



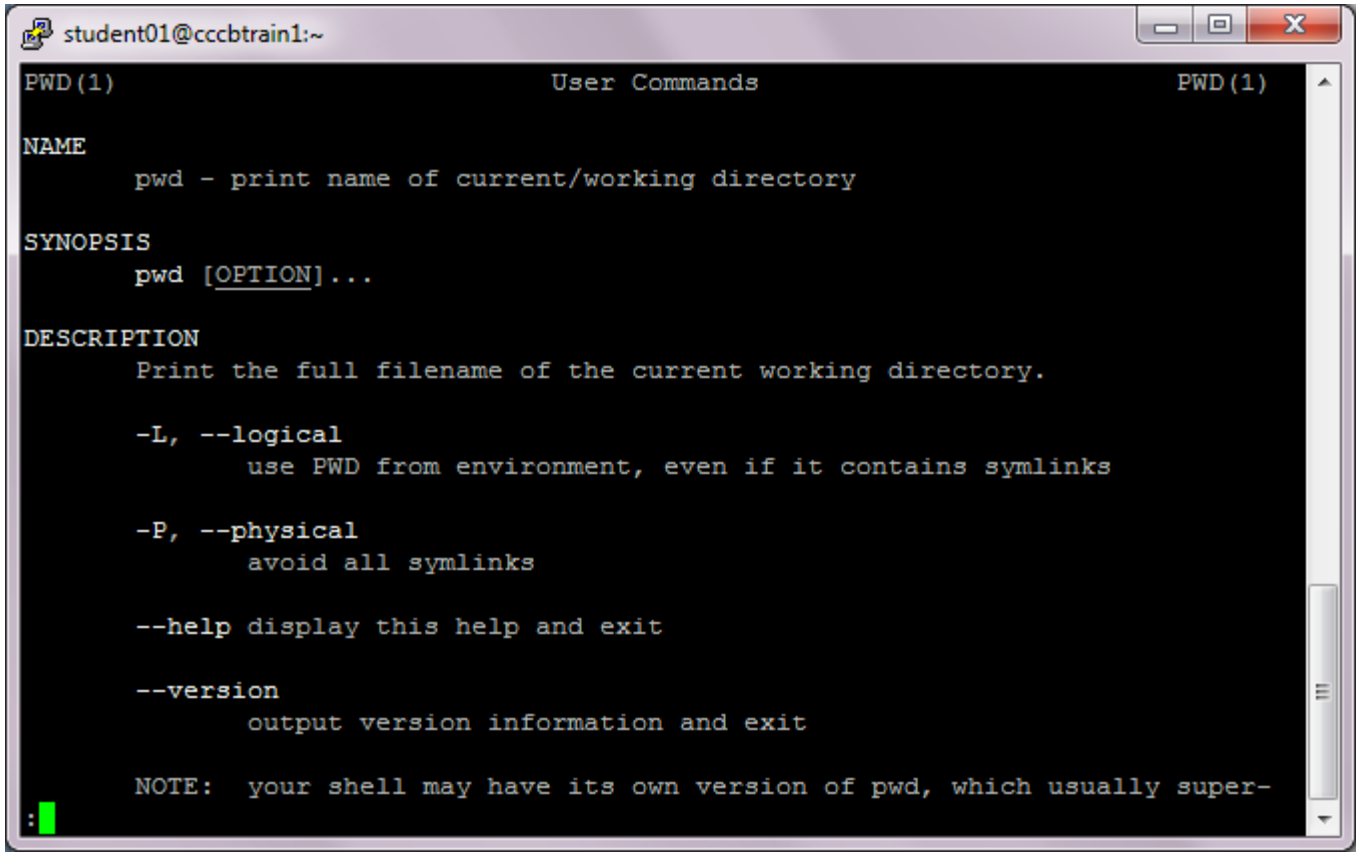
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



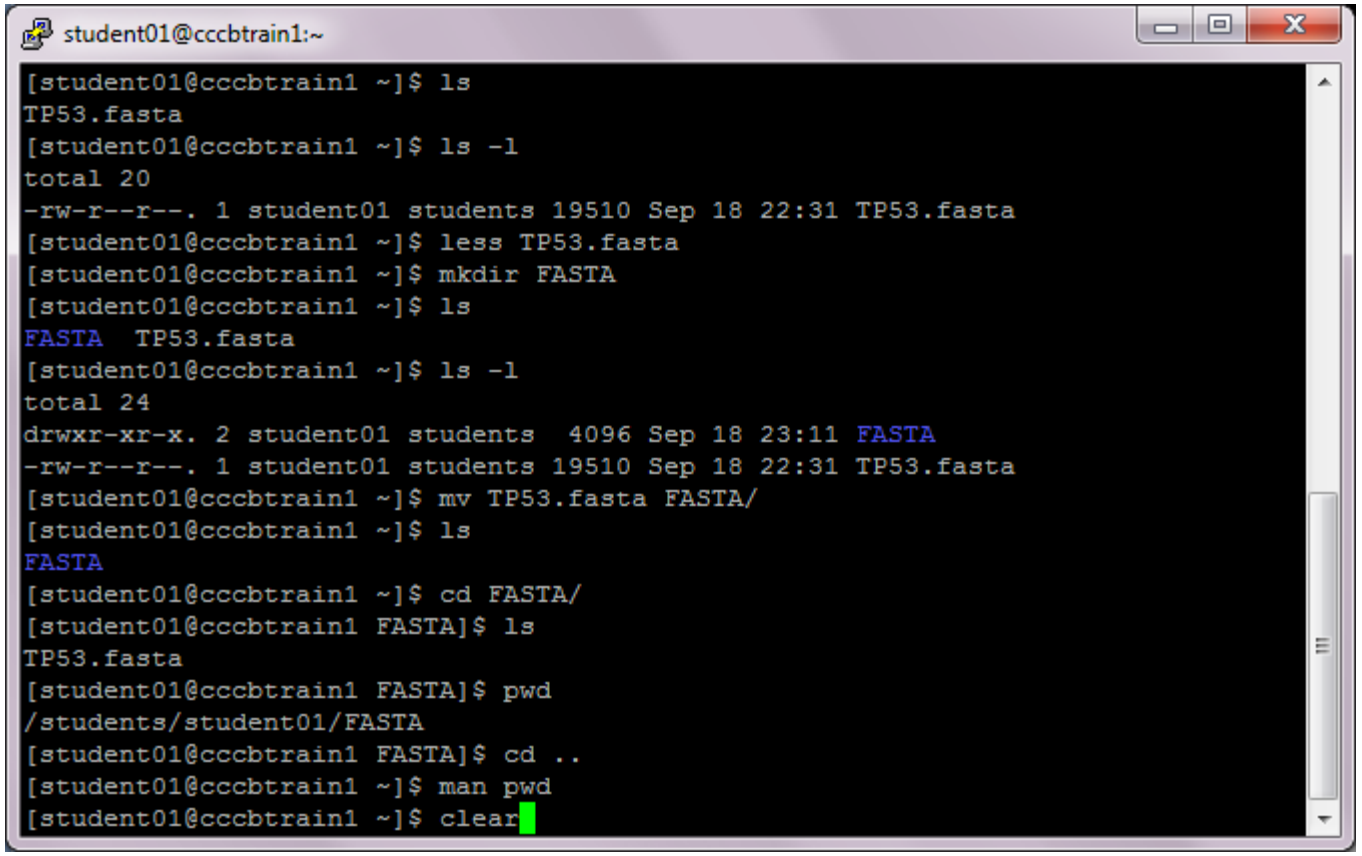
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



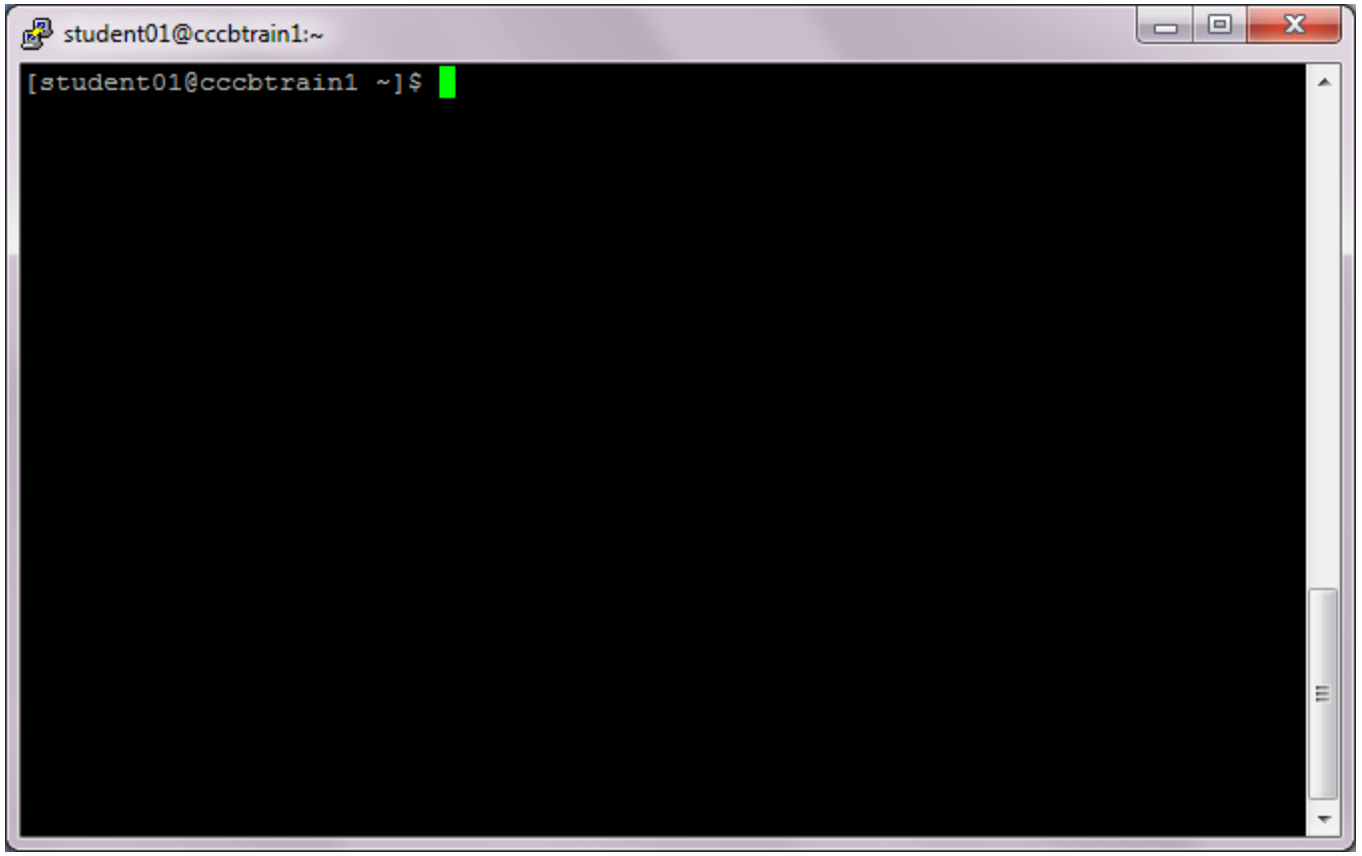
CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Basic Linux Command



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

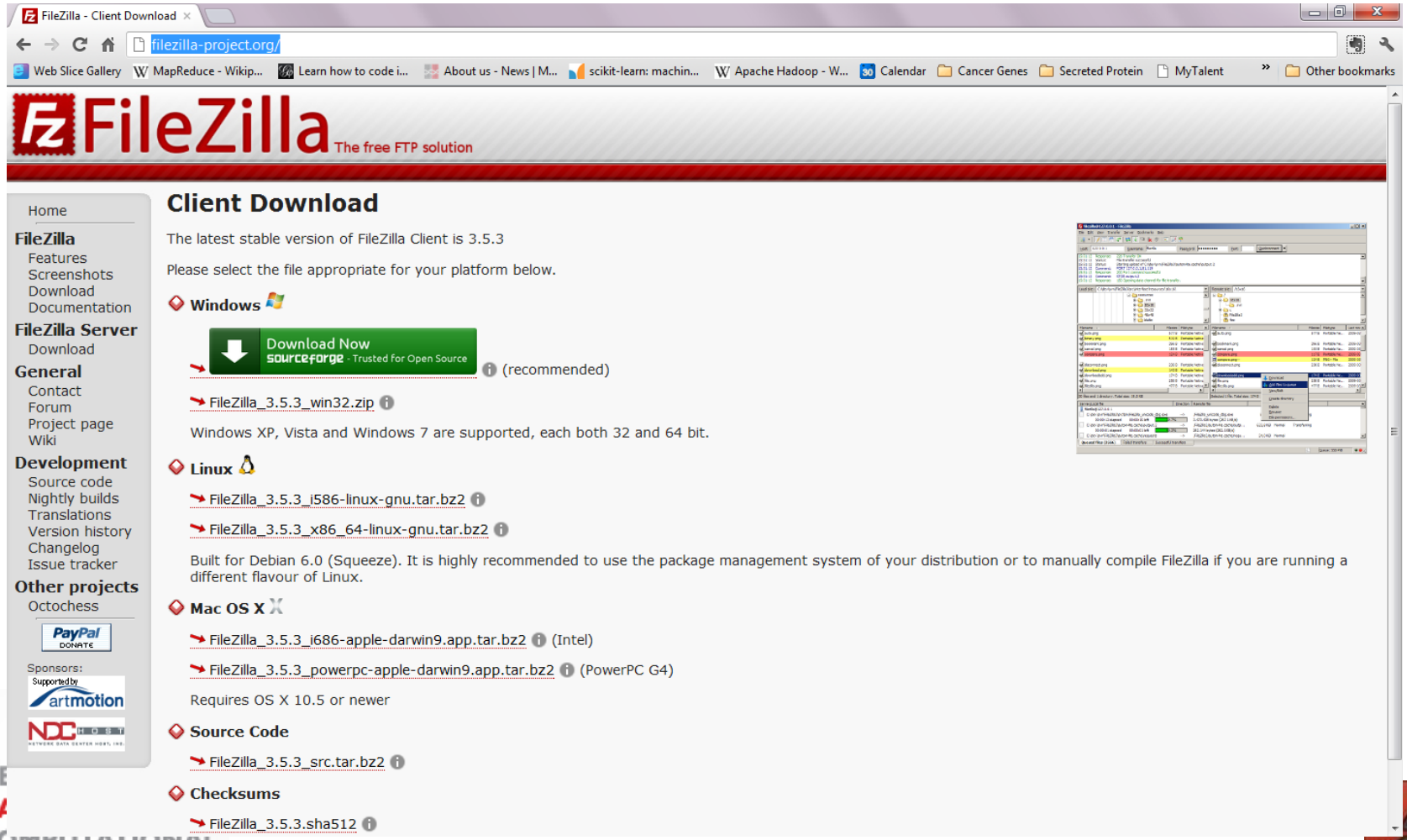
Cloud Computing

How to get files up and down the Cloud?

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

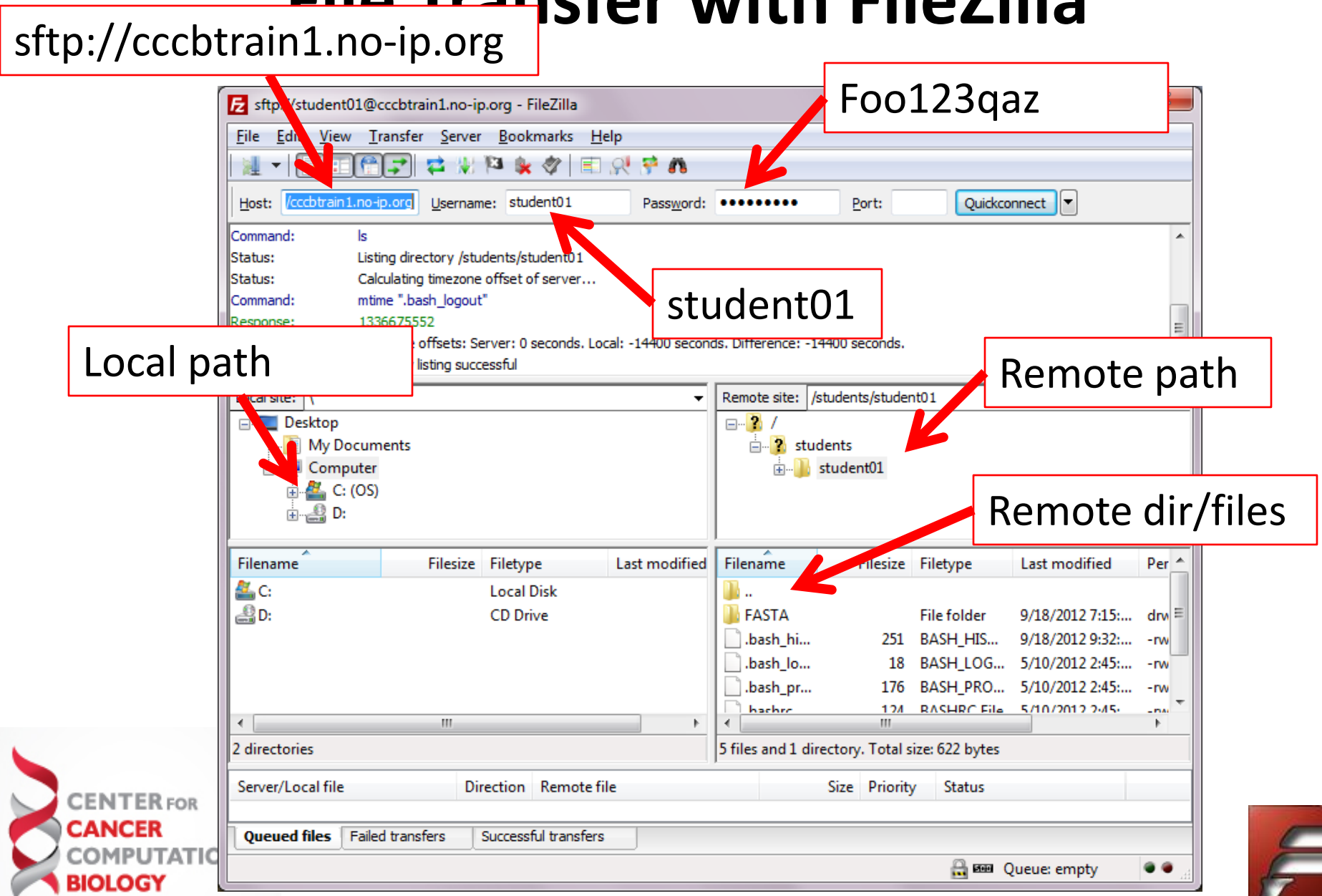
How to get files up and down the Cloud?

Download FileZilla at: http://filezilla-project.org/



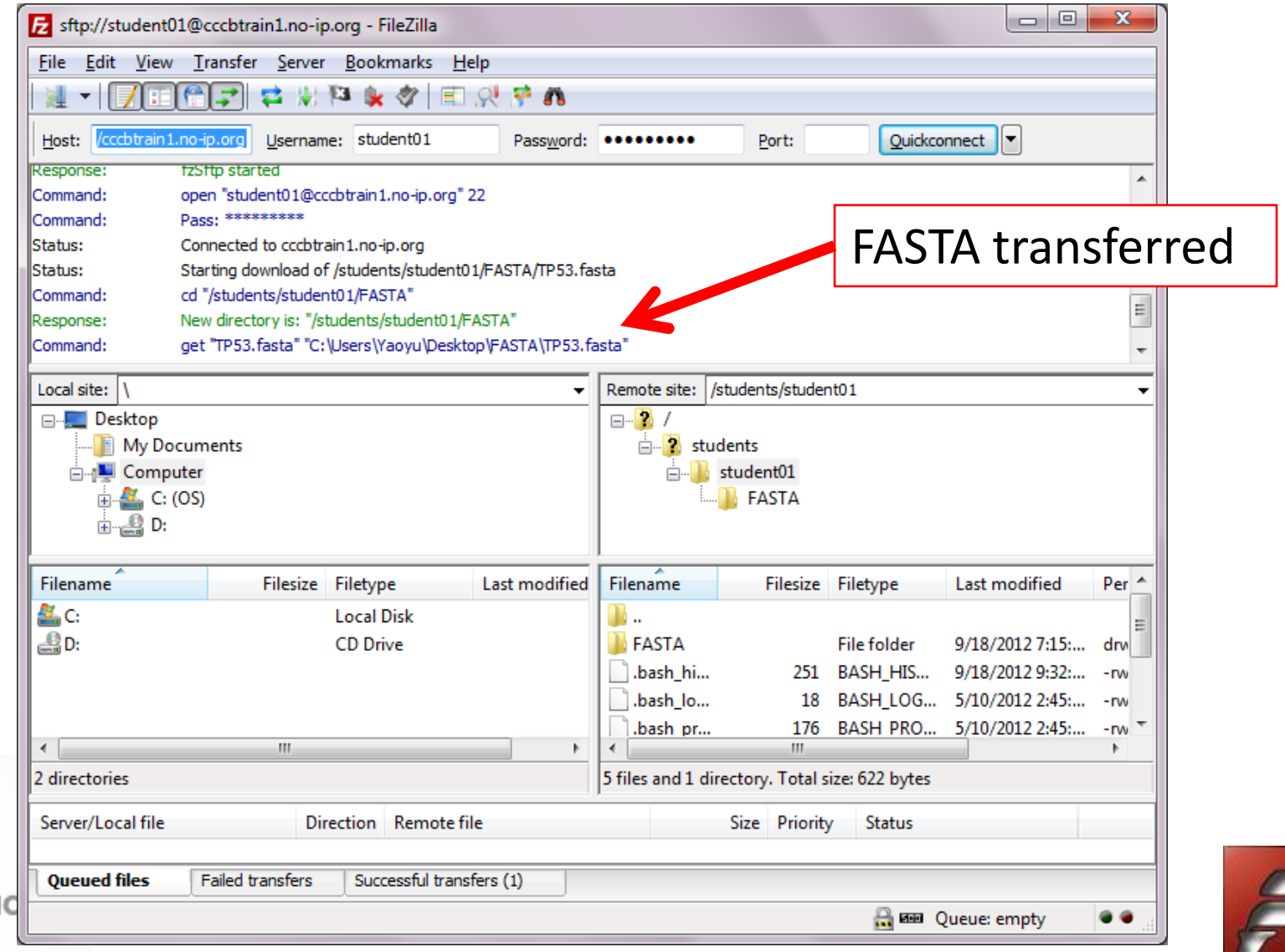
COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

File Transfer with FileZilla



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

File Transfer with FileZilla



CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Session Summary

- Could Computing are remotely hosted computational resources and ubiquitous accessible by internet
- Amazon Web Service (AWS) and Elastic Compute Cloud (EC2) provides cloud computing service and storage by demand
- Linux-based Cloud servers can be accessed through SSH (Secure SHell) by PuTTy
- Files can be transferred up and down the Cloud Server by SFTP using programs such as FileZilla

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE

Download R at:
/apps/R-installation

CENTER FOR CANCER COMPUTATIONAL BIOLOGY
DANA-FARBER CANCER INSTITUTE