Software for Bioinformatics

Oliver Stringham

Contents

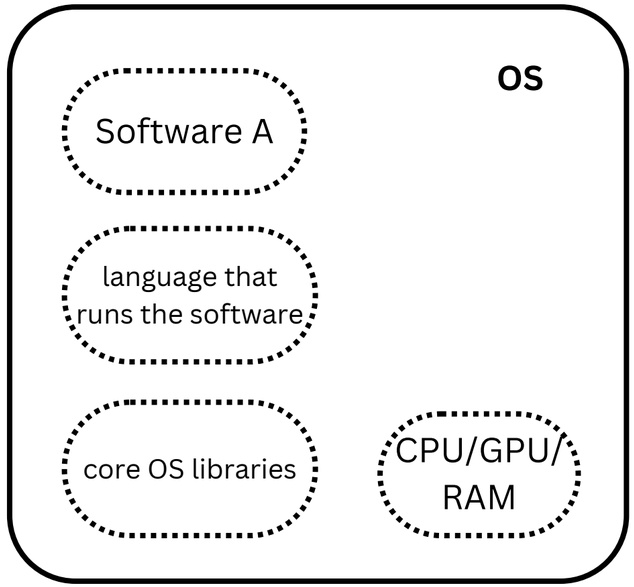
- **Installing software**
  - Logging on to Annotate2
  - Command line basics
  - ‘Installing’ images
    - quick check it worked

- **Running Sofware**
  - Transfering files to server
  - Syntax of running sofware
  - Run length filtering

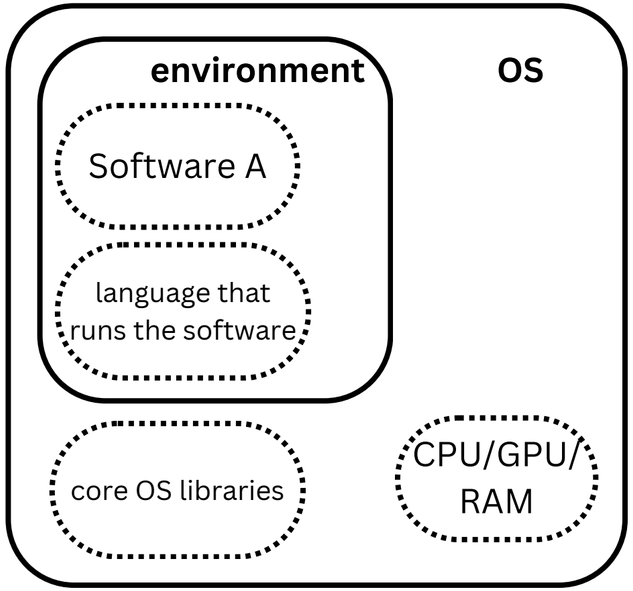
Installing software options*

*Details not important if you don’t care; just use containers

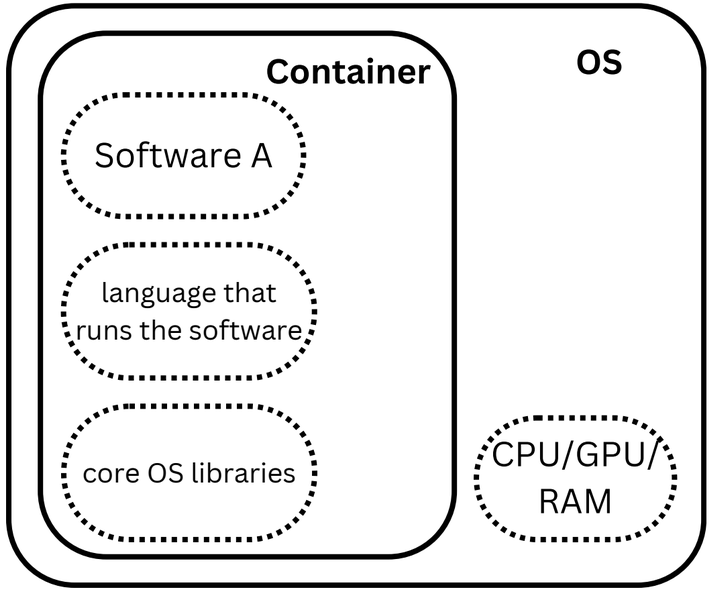
1. installing to your Operating System (OS)



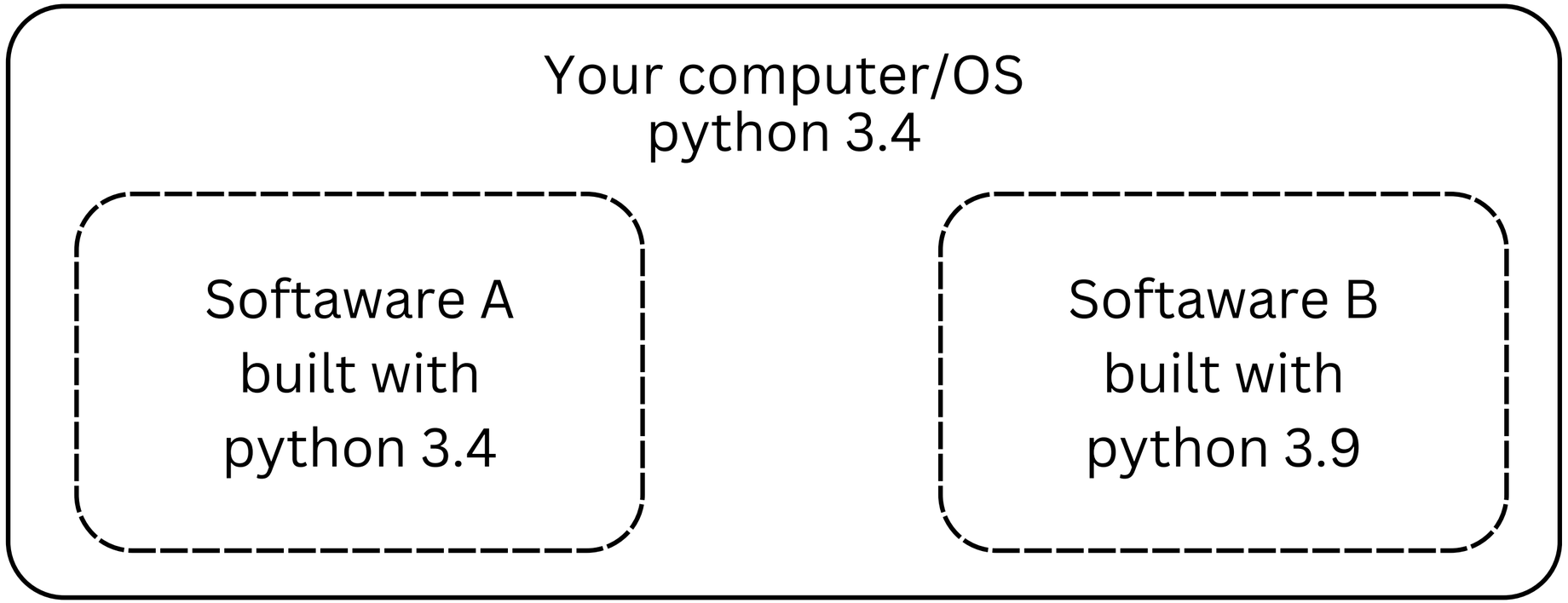
2. installing via an environment



3. ‘installing’ via a container



Dependency Hell



What happens when you run Software B?

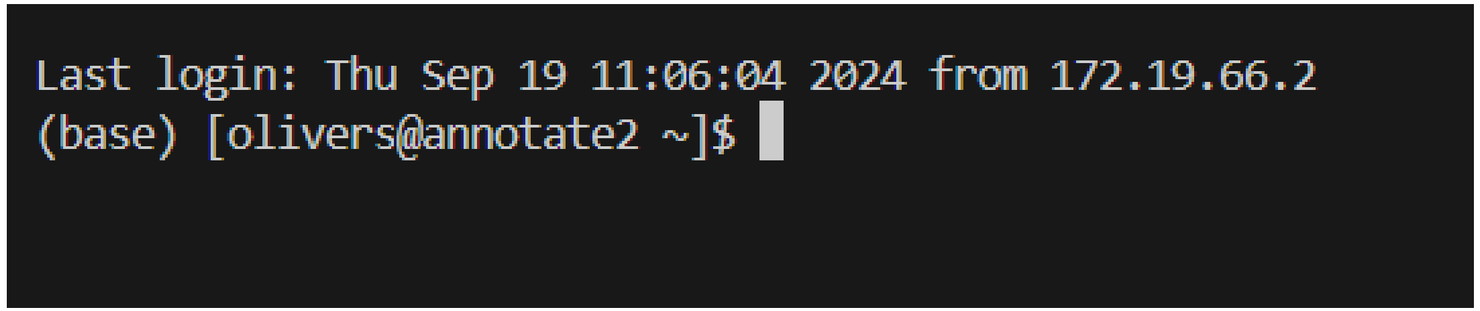
Note: dependency clashes can happen with python/R/etc, their packages, or core operating system libraries

OS Opinion:

Use conatainers where you can

Log onto Annotate2 server

1. Open Vscode (or a terminal)
2. On top bar go to Terminal --> New Termimal
3. type in the Terminal:
    a. `ssh your_username@annotate2.sebs.rutgers.edu`
    b. `ssh olivers@annotate2.sebs.rutgers.edu`
4. if it asks are you sure you want to continue, type yes or y (one time only)
5. type in your password (to paste, use right click) and press Enter
6. If successfull it will look something like this:



~ means your home directory

Some command line basics

- `pwd`
  - will print the current directory you are in
- `ls`
  - Will show all current files and directories in current directory
- `mkdir test`
  - will make a new directory called ‘test’
- `cd test/`
  - will change directories to the test/
- `cd ..`
  - will go up one directory

Rule one of programming always write down code you use even if you think you'll never use it again.

Let's start now.

'Installing' Images

Let's try to install seqkit via a container then use it to length filter a fastq file

- Make a new directory in your home directory called 'images'
  - only needs to be done once
- Look for (official) image (usually Docker Hub)
  - go to docker hub and search seqkit
- Run code in cluster to download & convert to singularity image (.sif)
  - `singularity build location-to-store-image where-image-lives-on-web`
  - `singularity build images/seqkit.sif docker://nanozoo/seqkit:2.6.1--022e008`
- wait until finishes
- check if it worked

2 ways to use containers

- interactive mode - inside the container
  - `singularity shell location-of-image`
  - run a command
  - type exit to exit container
- call the container for the specific line of code
  - `singularity exec location-of-image command`

Let’s try both ways to check the version of seqkit we ‘installed’

`seqkit version`

Let’s run length filtering on a real fastq 😱

But first you need to know how to import files to the server 😒

Import / Export Files from remote server

Can do via terminal but annoying because not GUI

Instead, use WinSCP (windows) or Cyberduck (Mac).
This will let you drag/drop files to/from the server

Go to week 2 folder and download fastq file to your computer
We will then transfer it to the server
check via command line it’s there

Using seqkit

https://bioinf.shenwei.me/seqkit/usage/#seq

- first let’s try to see some basic stats about the fastq file
  - `seqkit stats filename`
- Now let’s run length filtering and save output to a new file
  - `seqkit seq --min-len 50 --max-len 400 filename > new-filename`
- Check stats again

We did so much today 😄

That is all.