Introduction to Conda

Ammar Aziz – VIDRL

Doherty Institute

THE UNIVERSITY OF MELBOURNE

The Royal Melbourne Hospital

A joint venture between The University of Melbourne and The Royal Melbourne Hospital

Objectives

- What is Conda
- What problems Conda solves
- How to use Conda to install bioinformatics tools/software
- Learn briefly about virtual environments

Software (tool) Management

- Installing bioinformatics software was often very challenging.
- Each tool or program can have many dependencies
  - Snippy has 17 software dependencies
  - Each of those can have their own dependencies!
- Updating software adds even *more* complexities
- Different software has different dependencies – leading to conflicts – how do you resolve these?
- The default install mode was installing system-wide*
  - Creates complex dependencies between your research projects that shouldn't really exist!

* virtual environments (Environment Modules) on linux have been around since 1990s but required advanced Linux knowledge to setup and maintain.

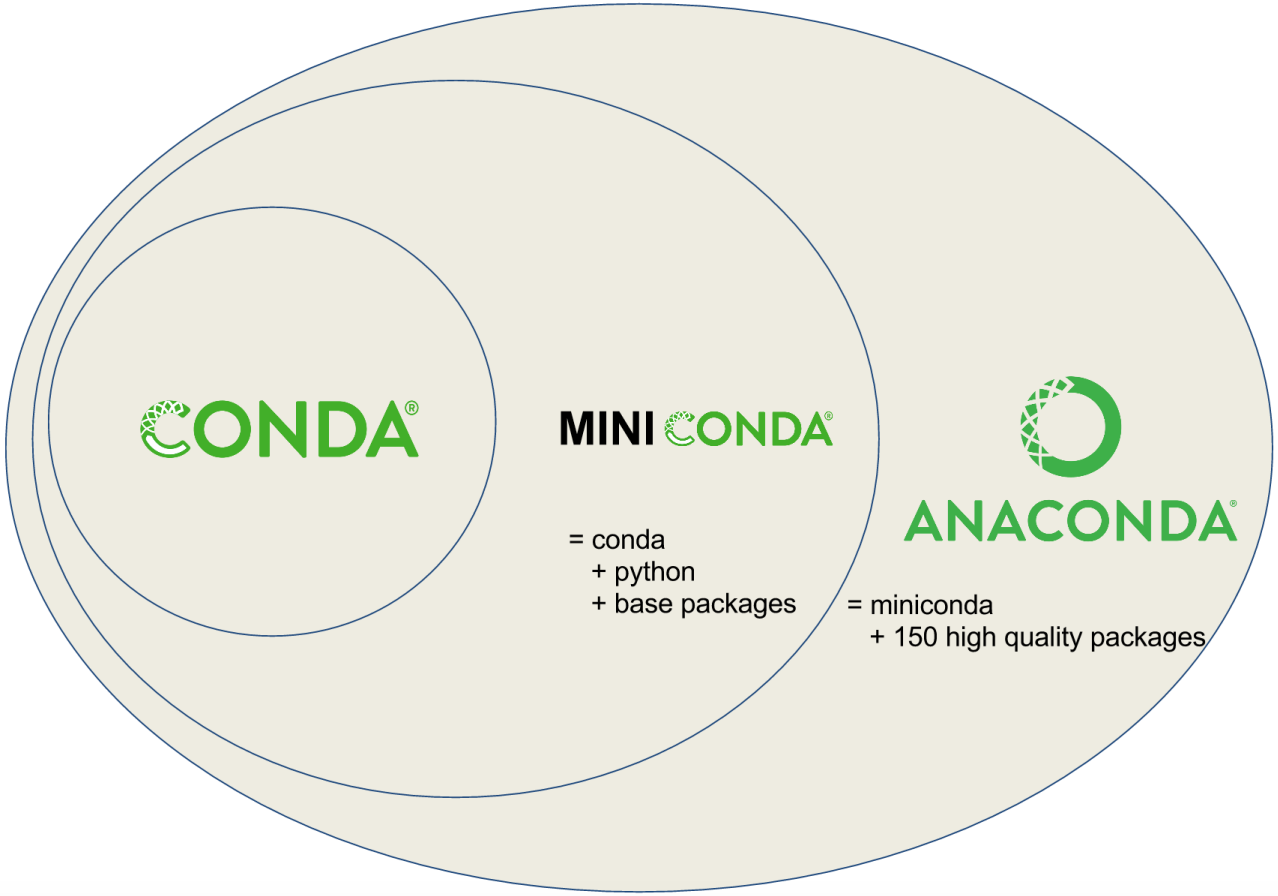
Conda – the scientific software manager

- Conda can quickly install, run, and update packages and their dependencies.
- Conda can create, save, load, and switch between project specific software environments on your local computer.
- Although Conda was created for Python programs, Conda can package and distribute software for any language such as R, Ruby, Lua, Scala, Java, JavaScript, C, C++, FORTRAN.

Why use Conda

- Conda provides prebuilt packages, avoiding the need to deal with compilers
- Conda is cross platform, with support for Windows, MacOS, GNU/Linux, and support for multiple hardware platforms, such as x86 and ARM.
- Conda allows for using other package management tools (such as `pip`) inside Conda environments
- Anaconda (software distribution) provides commonly used data science libraries and tools, such as R, NumPy, SciPy and TensorFlow built using optimised, hardware specific libraries
  - This requires commercial license – use with care

Conda vs. Miniconda vs. Anaconda



- Conda is a tool for managing environments and installing packages. Preloaded with many packages
- Miniconda combines Conda with Python and a small number of core packages
- Anaconda is the company behind Conda software
- Anaconda is also the name of a mega distribution with lots of software including R, Jypter, and many more
- Mamba – faster implementation of Conda

Conda distributions

Miniconda

Miniconda is a minimal installer provided by Anaconda. Use this installer if you want to install most packages yourself.

Anaconda Distribution

Anaconda Distribution is a full featured installer that comes with a suite of packages for data science, as well as Anaconda Navigator, a GUI application for working with conda environments.

Miniforge #

Miniforge is an installer maintained by the conda-forge community that comes preconfigured for use with the conda-forge channel. To learn more about conda-forge, visit their website.

'Distributions' refers packaging/delivering the same or similar software, each distribution has the same 'base' software with different optional software added. In the case of Conda distributions, they all perform the same task : install software

Which Conda distribution to use?

- Use **Miniforge** – open source, does not include 'defaults' channel, comes preinstalled mamba
- Read more on why:
  - https://ubinfie.github.io/2024/10/15/anaconda-defaults.html
- https://conda-forge.org/download

The way of the Mambalorian is miniforge distribution. This is the way.

Installing Conda (miniforge) - Linux/MacOS

- Go to https://conda-forge.org/download/
- Select your operating system to download
- Navigate to where you downloaded the file
- Run this command:

```
bash Miniforge3-$(uname)-$(uname -m).sh
```

What about Conda on Windows?

- Several options:
  - Install the mega distribution "Anaconda" which installs Navigator softare for managing environment using GUI
    - No personal experience with this
  - Install WSL2+Ubuntu

    - Best experience for CLI bioinformatics
    - Runs inside Windows, allows for accessing your data directly – no funny business
    - Negative: might require admin rights to install WSL2
- Some help docs here:
  - https://protocols.hostmicrobe.org/conda#block-d64fdca6e37e4997ae8671282562d9ac

Conda basics – Installing software

- Find if the package is in conda – simply search the internet for

```
mamba install {software name}
```

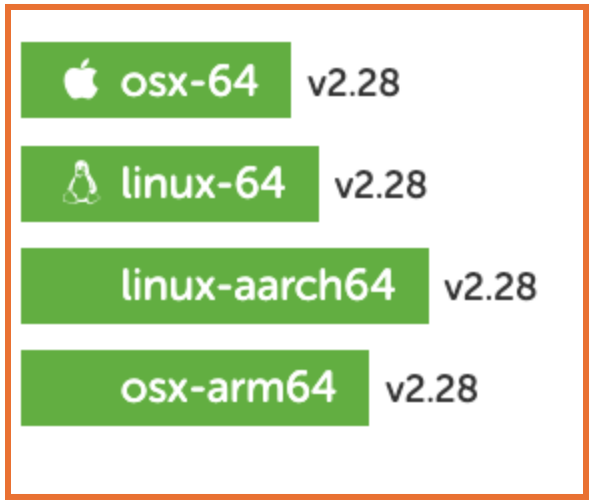
- For example:

```
mamba install minimap2
```

ANACONDA.ORG

Search Anaconda.org

bioconda / packages / minimap2 2.28



Tells you what version is available on which systems.

If you search on the internet "Conda + {software}" you will find instructions.

Copy paste into terminal.

conda install

To install this package run one of the following:

```
conda install bioconda::minimap2
```

```
conda install bioconda/label/cf201901::minimap2
```

Installing packages, software, tools, libs

- To install packages:

```
mamba install -c bioconda minimap2
```

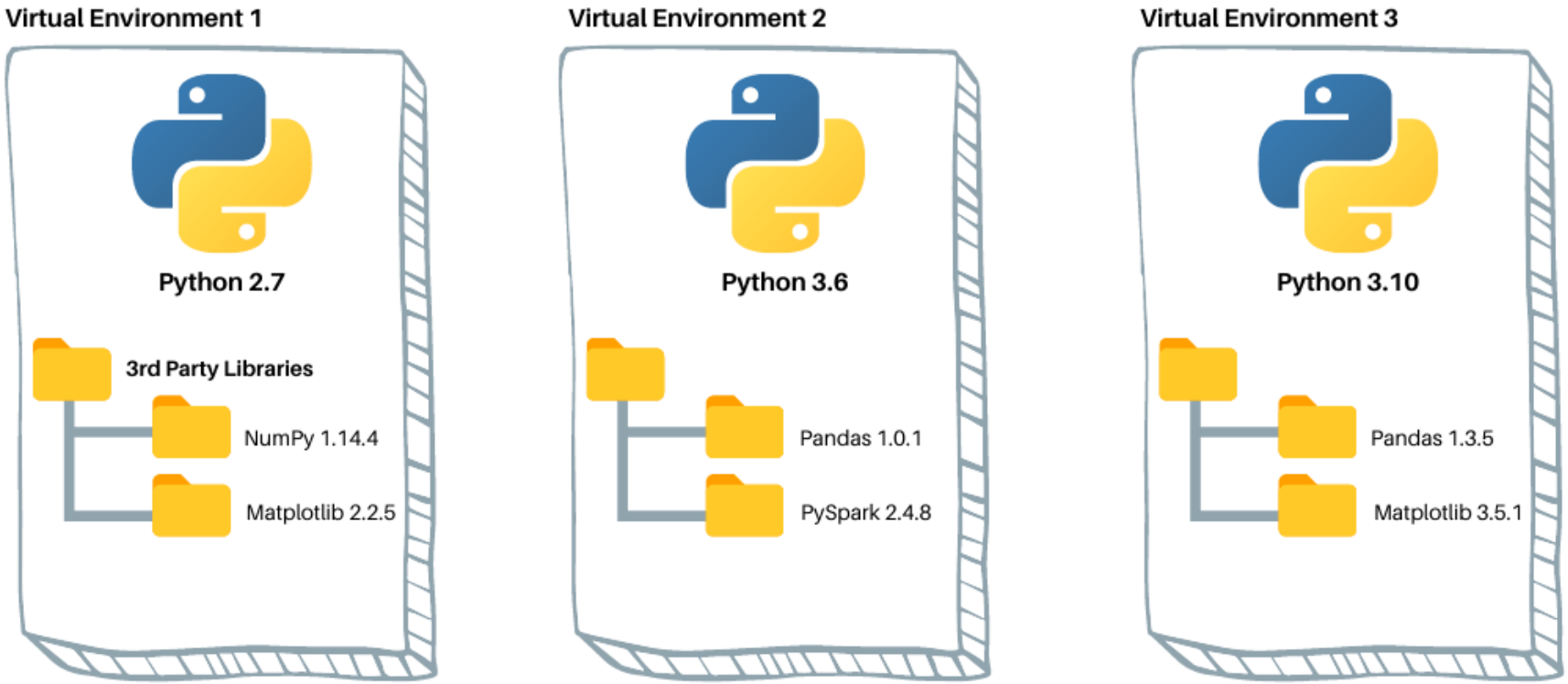
- The tells conda to use the channel 'bioconda'
- Multiple channels can be specified
- If you installed the miniforge distribution, you do not need to specify any channels – the defaults are set for you

- Tip: You can install multiple tools at once:

```
mamba install -c bioconda minimap2 samtools
```

Conda Environments

- A Conda environment is a directory that contains a specific collection of Conda packages that you have installed.



Creating New Environments

- New environments can be activated:

```
mamba activate myEnv
```

- To deactivate:

```
mamba deactivate
```

- Create as many environments as you wish!
- Call them whatever you wish – but be descriptive:
  - 'dengue-seq-pipeline' – good
  - 'pipeline' – bad
  - 'myEnv' - bad

Best practices

- Do not install software or tools in your **(base)** environment – keep this clean
- Create an environment for each project/tool
- Use descriptive but short names
- To show your environments:

```
mamba env list
```

- To export your environment:

```
mamba list –e > req.txt
```

Questions? + Resources

- Official Documentation:
  - https://docs.conda.io/projects/conda/en/latest/user-guide/getting-started.html
- Software Carpentries:
  - https://carpentries-incubator.github.io/introduction-to-conda-for-data-scientists
- Conda for Bioinformaticians:
  - https://protocols.hostmicrobe.org/conda