Software Environments

Dr. Christopher S. Simmons

csim@mghpcc.org

Towards Reproducible Computing

- Last time you learned about Git(Hub) and version control
- More than just software development can be tracked
- Majority of scientific workflows consume 3rd party software
- How best to manage software stacks?
    - Find someone else to do it
    - Manually compile from source
    - Use "user-space package managers" and manually configure environment
    - Use a tool to declaratively specify requirements via a config file and version control it
- Many such tools including language specific and agnostic options

Conda, Anaconda, and Miniconda

- conda is an open-source package and environment management system for any programming language; though it is quite popular in the python community.

- Package, dependency and environment management for any language—Python, R, Julia, Ruby, Lua, Scala, Java, JavaScript, C/ C++, FORTRAN, and more.

- Anaconda is a distribution of conda. It is a data science platform that comes with a lot of packages. Anaconda is known for destroying HPC / Parallel File systems

- Miniconda is a minimal conda environment. Unlike Anaconda, Miniconda doesn't come with any installed packages by default.
  - It is a small, bootstrap version of Anaconda that includes only conda, Python, the packages they depend on, and a small number of other useful packages, including pip, zlib and a few others

Caveat: /tmp and secured systems

- /tmp can be set to noexec for security reasons
- Many programs use $TMPDIR which defaults to /tmp if not set
- In order to use conda or spack, first you must create a new "TMPDIR" and point the environment variable to it

$ mkdir $SCRATCH/tmp

$ echo "export TMPDIR=$SCRATCH/tmp" >> ~/.bashrc

- Logout and log back in or source ~/.bashrc before continuing

Getting Miniconda

Download and install from their website

https://docs.conda.io/en/latest/miniconda.html

```
[css180002@europa src]$ wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
--2021-04-08 11:36:04--  https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 104.16.130.3, 104.16.131.3, 2606:4700::6810:8303, ...
Connecting to repo.anaconda.com (repo.anaconda.com)|104.16.130.3|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 94235922 (90M) [application/x-sh]
Saving to: 'Miniconda3-latest-Linux-x86_64.sh'

Miniconda3-latest-Linux-x86_64.sh  100%[===================================================================>]  89.87M   260MB/s    in 0.3s

2021-04-08 11:36:04 (260 MB/s) - 'Miniconda3-latest-Linux-x86_64.sh' saved [94235922/94235922]

[css180002@europa src]$ chmod +x Miniconda3-latest-Linux-x86_64.sh
[css180002@europa src]$ ./Miniconda3-latest-Linux-x86_64.sh
```

Getting Miniconda cont.

```
Preparing transaction: done
Executing transaction: done
installation finished.
Do you wish the installer to initialize Miniconda3
by running conda init? [yes|no]
[no] >>> yes
no change        /home/css180002/miniconda3/condabin/conda
no change        /home/css180002/miniconda3/bin/conda
no change        /home/css180002/miniconda3/bin/conda-env
no change        /home/css180002/miniconda3/bin/activate
no change        /home/css180002/miniconda3/bin/deactivate
no change        /home/css180002/miniconda3/etc/profile.d/conda.sh
no change        /home/css180002/miniconda3/etc/fish/conf.d/conda.fish
no change        /home/css180002/miniconda3/shell/condabin/Conda.psm1
no change        /home/css180002/miniconda3/shell/condabin/conda-hook.ps1
no change        /home/css180002/miniconda3/lib/python3.8/site-packages/xontrib/conda.xsh
no change        /home/css180002/miniconda3/etc/profile.d/conda.csh
modified         /home/css180002/.bashrc

==> For changes to take effect, close and re-open your current shell. <==

If you'd prefer that conda's base environment not be activated on startup,
   set the auto_activate_base parameter to false:

conda config --set auto_activate_base false

Thank you for installing Miniconda3!
[css180002@europa src]$ []
```

Managing Conda

$ conda –version # see which version of conda

$ conda –help # get help and list of commands

$ conda update –help # get help with a specific command

$ conda update -n base -c defaults conda # update conda

$ conda update –all # updates all packages; use with care

- https://conda.io/projects/conda/en/latest/user-guide/getting-started.html

Managing Environments

- Using conda, you can create an isolated python environment for your project.
  - An environment is a set of packages that can be used in one or multiple projects
  - The default environment with Anaconda is the base environment, which contains Anaconda default packages
- There are two ways of creating a conda environment.
  - Manual specifications of packages
  - An environment file in YAML format (environment.yml)

- YAML stands for YAML Ain't Markup Language -- It is a human friendly data serialization standard for all programming languages.

Managing Environments Cont.

Create a file named environment.yml

$ conda env create --file environment.yml

$ conda activate myenv #activate env

$ conda deactivate # returns to base env

$ conda env list # lists environments

```
name: myenv
channels:
  - defaults
dependencies:
  - python=3.9
  - pip
  - pip:
      - pyjokes
```

Creating, sharing and deleting environments

$ conda create -c conda-forge -n test_env python=3.9 numpy matplotlib pandas

$ conda env export -f test_env.yml -n test_env

$ conda create --name live_env --clone test_env

$ conda env remove -n live_env

Bese practices for maintaining and environment
- When starting a new environment, always generate it from an environment file rather than the command line
- Maintain your environment.yml files in git along side the code repo for the science
- As you add packages to the environment, be sure to update the environment file
- Unless you must, try to avoid specifying the version of each package. This will ensure you have the most up to date version that will work across platform.

Conda channels and more info

- https://docs.conda.io/projects/conda/en/latest/user-guide/concepts/channels.html

- https://anaconda.org/intel/repo

- https://conda.io/projects/conda/en/latest/user-guide/cheatsheet.html

- Full documentation-- https://conda.io/docs/