

Using Anaconda on the Tufts HPC

Rebecca Batorsky
Senior Bioinformatics
Specialist
July 2019

- Python versions and packages
- Anaconda
- Customizing conda defaults so you don't fill up your home directory
- Creating and activating environments

Python versions

To view Tufts HPC available python versions:

```
tutln01@m3n09:~$ module av python
```

```
----- /opt/shared/Modules/modulefiles-rhel6 -----  
python/2.6.5      python/2.7.3(default)  
python/2.7.4      python/2.7.6      python/3.5.0      python/3.6.0
```

Python versions

To view Tufts HPC available python versions:

```
tutln01@m3n09:~$ module av python
```

```
----- /opt/shared/Modules/modulefiles-rhel6 -----  
python/2.6.5      python/2.7.3(default)  
python/2.7.4      python/2.7.6      python/3.5.0      python/3.6.0
```

Command entered

Bash prompt for user
tutln01 on compute
node m309

Result of command

Python versions

To view Tufts HPC available python versions:

```
tutln01@m3n09:~$ module av python
```

```
----- /opt/shared/Modules/modulefiles-rhel6 -----  
python/2.6.5      python/2.7.3(default)  
python/2.7.4      python/2.7.6      python/3.5.0      python/3.6.0
```

To load a python version:

```
tutln01@m3n09:~$ module load python/3.6.0
```

```
tutln01@m3n09:~$ python --version
```

```
Python 3.6.0
```

Python packages

Next, install a package

```
tutln01@m3n09 :~$ pip install matplotlib
```

```
Requirement already satisfied: matplotlib in /opt/shared/python/3.6.0/lib/python3.6/site-packages (2.2.3)
```

```
...
```

This means that matplotlib version 2.2.3 is already installed for you on the system

Great! But what if you need an updated matplotlib version 3.0.3?

Python packages

Next, install a package

```
tutln01@m3n09 :~$ pip install matplotlib
Requirement already satisfied: matplotlib in /opt/shared/python/3.6.0/lib/python3.6/site-packages (2.2.3)
...
```

This means that matplotlib version 2.2.3 is already installed for you on the system

Great! But what if you need an updated matplotlib version 3.0.3?

```
tutln01@m3n09 :~$ pip install matplotlib==3.0.3
...
Found existing installation: matplotlib 2.2.3
  Uninstalling matplotlib-2.2.3:
    Could not install packages due to an EnvironmentError: [Errno 13] Permission denied:
    '/opt/shared/python/3.6.0/lib/python3.6/site-packages/__pycache__/pylab.cpython-36.pyc'
    Consider using the `--user` option or check the permissions.
```

Python packages

Next, install a package

```
tutln01@m3n09 :~$ pip install matplotlib
```

```
Requirement already satisfied: matplotlib in /opt/shared/python/3.6.0/lib/python3.6/site-packages (2.2.3)
```

```
...
```

This means that matplotlib version 2.2.3 is already installed for you on the system

Great! But what if you need an updated matplotlib version 3.0.3?

```
tutln01@m3n09 :~$ pip install matplotlib==3.0.3
```

```
...
```

```
Found existing installation: matplotlib 2.2.3
```

```
Uninstalling matplotlib-2.2.3:
```

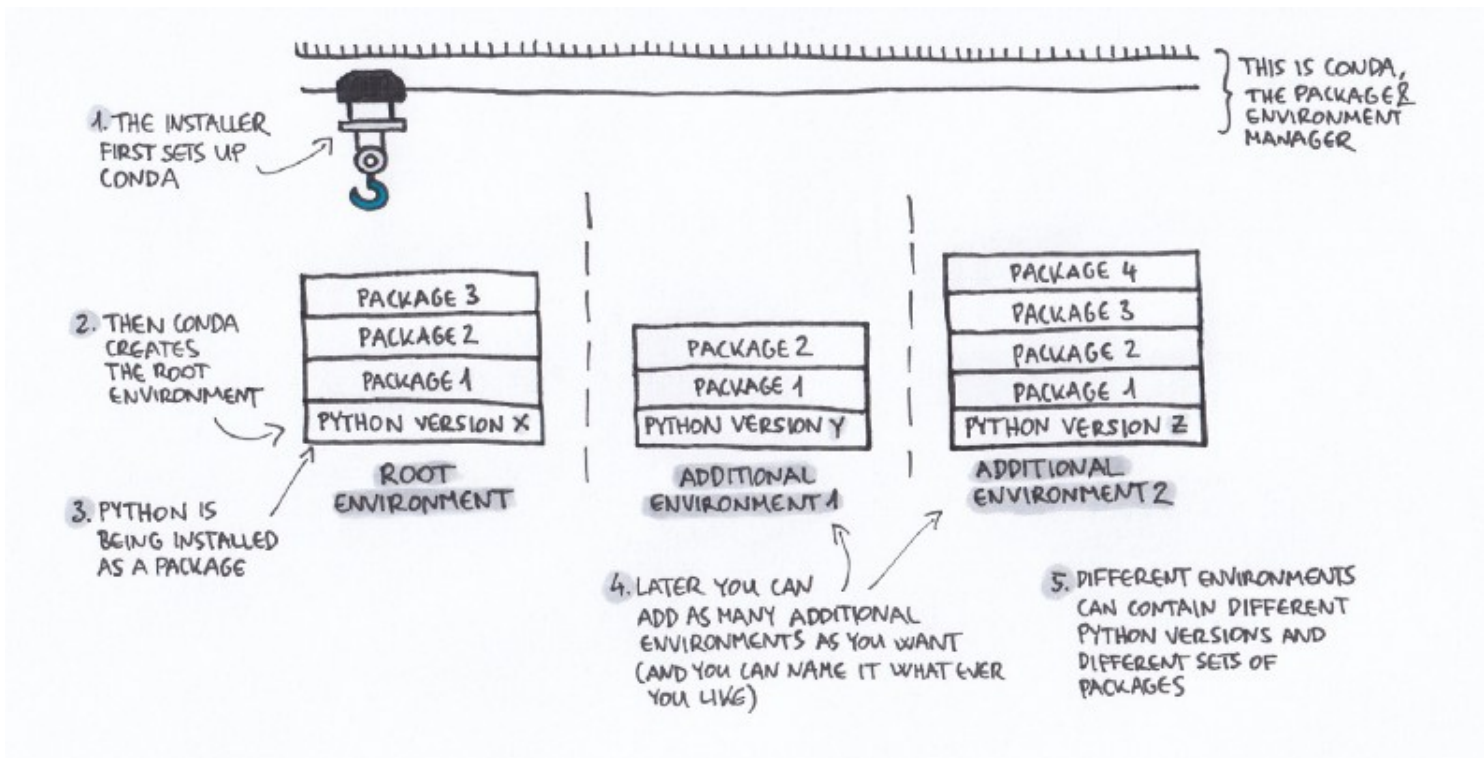
```
Could not install packages due to an EnvironmentError: [Errno 13] Permission denied:  
'/opt/shared/python/3.6.0/lib/python3.6/site-packages/__pycache__/pylab.cpython-36.pyc'
```

```
Consider using the `--user` option or check the permissions.
```

You don't have write privileges in the system python path – you'll need to create an environment in a location where you can write.

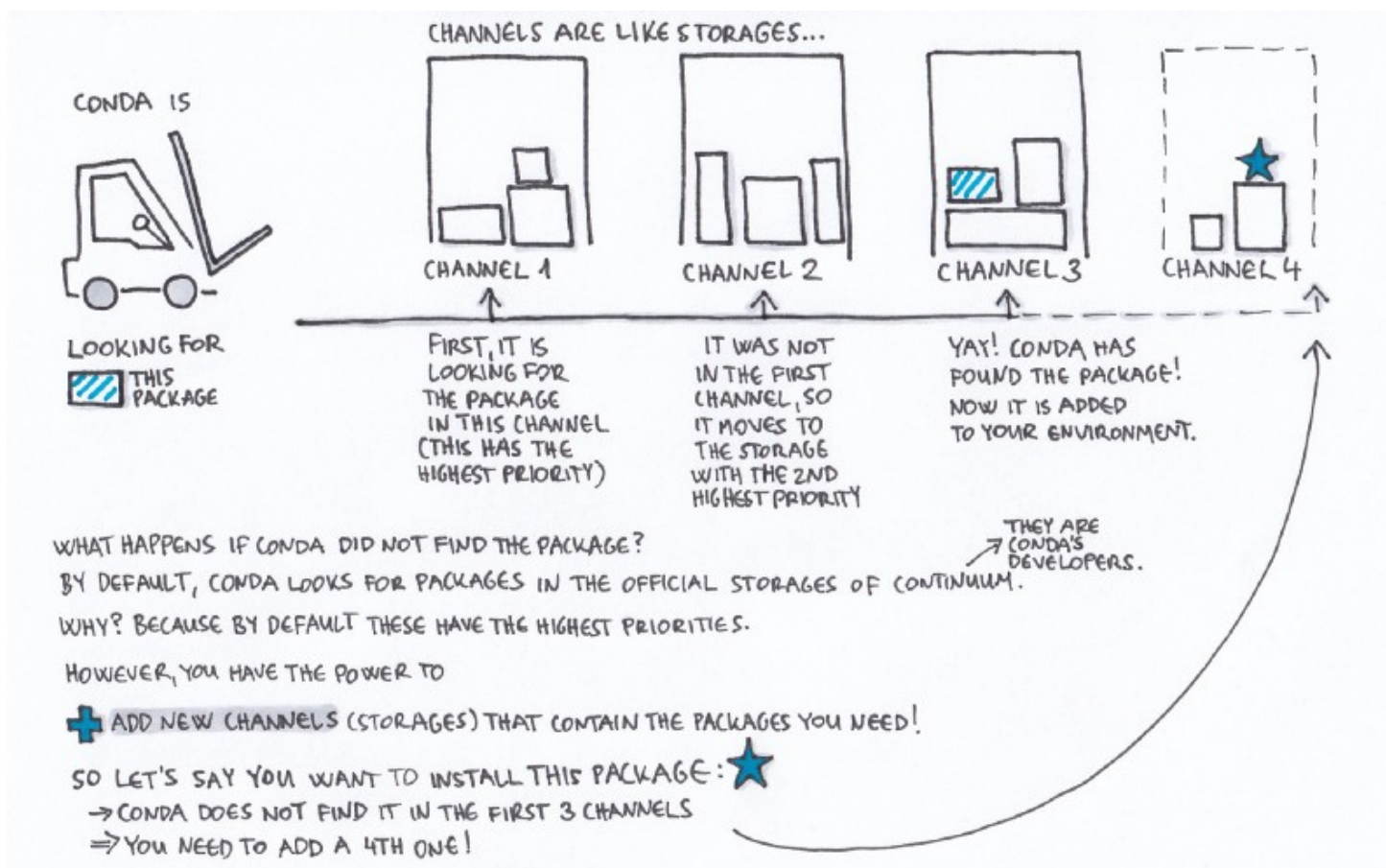
Anaconda

You can maintain python versions and packages in separate **environments** so that they don't conflict with system versions or with each other



<https://www.freecodecamp.org/news/why-you-need-python-environments-and-how-to-manage-them-with-conda-85f155f4353c/>

Channels



<https://www.freecodecamp.org/news/why-you-need-python-environments-and-how-to-manage-them-with-conda-85f155f4353c/>

Anaconda on Tufts HPC

To load Anaconda:

```
tutln01@m3n09:~$ module load anaconda/3
```

Basic information about anaconda:

```
[tutln01@m3n09 ~]$ conda info
active environment : None
user config file : /cluster/home/tutln01/.condarc
populated config files :
  conda version : 4.5.11
  conda-build version : 3.4.1
  python version : 3.6.4.final.0
base environment : /cluster/tufts/software/anaconda3 (read only)
channel URLs : https://repo.anaconda.com/pkgs/main/linux-64
...
package cache : /cluster/tufts/software/anaconda3/pkgs
                 /cluster/home/tutln01/.conda/pkgs
envs directories : /cluster/home/tutln01/.conda/envs
                  /cluster/tufts/software/anaconda3/envs
platform : linux-64
user-agent : conda/4.5.11 requests/2.18.4 CPython/3.6.4 Linux/2.6.32-696.20.1.el6.x86_64 rhel/6.9 glibc/2.12
UID:GID : 31394:7000
netrc file : None
offline mode : False
```

Anaconda on Tufts HPC

To load Anaconda:

```
tutln01@m3n09:~$ module load anaconda/3
```

Basic information about anaconda:

```
tutln01@m3n09: ~$ conda info
active environment : None
user config file : /cluster/home/tutln01/.condarc
populated config files :
  conda version : 4.5.11
  conda-build version : 3.4.1
  python version : 3.6.4.final.0
base environment : /cluster/tufts/software/anaconda3 (read only)
channel URLs : https://repo.anaconda.com/pkgs/main/linux-64
...
package cache : /cluster/tufts/software/anaconda3/pkgs
                 /cluster/home/tutln01/.conda/pkgs
envs directories : /cluster/home/tutln01/.conda/envs
                  /cluster/tufts/software/anaconda3/envs
platform : linux-64
user-agent : conda/4.5.11 requests/2.18.4 CPython/3.6.4 Linux/2.6.32-696.20.1.el6.x86_64 rhel/6.9 glibc/2.12
UID:GID : 31394:7000
netrc file : None
offline mode : False
```

By default packages are installed and package source files cached in your home directory – which won't work on the HPC since our home directories are only ~5Gb

Anaconda on Tufts HPC

To load Anaconda:

```
tutln01@m3n09:~$ module load anaconda/3
```

Basic information about anaconda:

```
tutln01@m3n09: ~$ conda info
active environment : None
user config file : /cluster/home/tutln01/.condarc
populated config files :
  conda version : 4.5.11
  conda-build version : 3.4.1
  python version : 3.6.4.final.0
  base environment : /cluster/tufts/software/anaconda3 (read only)
  channel URLs : https://repo.anaconda.com/pkgs/main/linux-64
...
  package cache : /cluster/tufts/software/anaconda3/pkgs
                  /cluster/home/tutln01/.conda/pkgs
  envs directories : /cluster/home/tutln01/.conda/envs
                   /cluster/tufts/software/anaconda3/envs
  platform : linux-64
  user-agent : conda/4.5.11 requests/2.18.4 CPython/3.6.4 Linux/2.6.32-696.20.1.el6.x86_64 rhel/6.9 glibc/2.12
  UID:GID : 31394:7000
  netrc file : None
  offline mode : False
```

← Config file allows one to change these defaults!

← By default packages are installed and package source files cached in your home directory – which won't work on the HPC since our home directories are only ~5Gb

Anaconda base environment

To see the environments available:

```
tutln01@m3n09:~$ conda info -envs
# conda environments:
#
base          * /cluster/tufts/software/anaconda3
```

To activate the base environment and view installed packages:

```
tutln01@m3n09:~$ source activate base
(base) tutln01@m3n09:~$ conda list
# packages in environment at /cluster/tufts/software/anaconda3:
# Name                Version           Build Channel
_ipyw_jlab_nb_ext_conf 0.1.0             py36he11e457_0
_r-mutex               1.0.0             mro_2
...
```

Anaconda base environment

To deactivate :

```
(base) tutln01@m3n09:~$ source deactivate  
tutln01@m3n09:~$
```

~/condarc

To change where our environments are installed and package source is kept, create and/or edit the file ~/condarc

```
tutln01@m3n09:~$ nano ~/condarc
```

```
channels:
```

- conda-forge
- bioconda
- defaults

```
envs_dirs:
```

- /path/to/project/space/conda_envs/

```
pkgs_dirs:
```

- /path/to/project/space/conda_envs/pkgs/

Note you should type this in a text editor rather than copying and pasting from the slides in order to avoid formatting errors

If you don't have project space, email tts-research@tufts.edu and HPC staff can help find a place for your work

For more condarc modifications: <https://docs.conda.io/projects/conda/en/latest/user-guide/configuration/use-condarc.html>

~/ .condarc

Now you will see a new environment directory:

```
tutln01@m3n09:~$ conda info
```

```
...
```

```
envs directories : /path/to/project/space/conda_envs/  
                  /cluster/home/tutln01/.conda/envs  
                  /cluster/tufts/software/anaconda3/envs
```

```
...
```

Create an Environment

To create an environment “test” by name:

```
tutln01@m3n09:~$ conda create -n test
...
## Package Plan ##
  environment location: /path/to/project/space/conda_envs/test
Proceed ([y]/n)? Y
...
```

To activate:

```
tutln01@m3n09:~$ source activate test
(/path/to/project/space/conda_envs/test) tutln01@m3n09:~$
```

To install matplotlib 3.0.3

```
(/path/to/project/space/conda_envs/test) tutln01@m3n09:~$ conda install matplotlib=3.0.3
```

Create an Enviornment

To install matplotlib 3.0.3

```
(/path/to/project/space/conda_envs/test) tutln01@m3n09:~$ conda install matplotlib=3.0.3
```

To deactivate

```
(/path/to/project/space/conda_envs/test) tutln01@m3n09:~$ source deactivate
```

Loading environments from a path

The Tufts HPC has some bioinformatics-oriented environments installed that are not yet available as modules:

```
tutln01@m3n09:~$ cd /cluster/tufts/bio/tools/conda_envs
tutln01@m3n09: conda_envs $ ls
biopython    deeplabcuts minimap2  opera-ms  prokka      rseqc    scTDA_2.7  transit_2_3_4  vcfkit
conda_scripts  ensembl-vep multiqc  pbttools  python3     salmon   seurat     ucsc_tools    wxpython
cpanm        htseq      mummer   pkgs      rnaseq_course  samtools test      umap-learn
```

To activate an environment:

```
tutln01@m3n09: conda_envs $ source activate /cluster/tufts/bio/tools/conda_envs/salmon/0.13.1
(/cluster/tufts/bio/tools/conda_envs/salmon/0.13.1 ) tutln01@m3n09: conda_envs $
```

To deactivate:

```
(/cluster/tufts/bio/tools/conda_envs/salmon/0.13.1 ) tutln01@m3n09: conda_envs $ source deactivate
```

Removing environments

```
tutln01@m3n09:~$ conda remove -n test --all
```

Summary

- Load python and anaconda as modules on HPC
- Load the base anaconda environment
- Modify the ~/.condarc configuration file to customize environment and package paths and avoid filling up your home directory
- Create an environment and install a tools
- Load an environment from a path

For more information about bioinformatics resources at tufts, see <https://sites.tufts.edu/biotools/>

More information

- For more information about bioinformatics resources at tufts, see <https://sites.tufts.edu/biotools/>
- Questions about using the Tufts HPC: tts-research@tufts.edu
- Documentation on Anaconda: <https://conda.readthedocs.io/en/latest/>
- Helpful anaconda blog: <https://medium.freecodecamp.org/why-you-need-python-environments-and-how-to-manage-them-with-conda-85f155f4353c>