

FauxConda: ways to use Conda without Conda

Wednesday, March 24, 2021

We will be starting at 12:05 PM

Disclaimer

The title was just to get you here, we will NOT use Conda at all !!!

- Except in the singularity image

Here we assume that:

- You can connect to Compute Canada clusters
- You are comfortable with python and python environments
- You understand how to load and search for modules in Compute Canada systems.

Why should you avoid Conda?

Conda is a useful solution but to ensure compatibility and optimal performance, should not be used. Also (list is not comprehensive) :

- Anaconda might suboptimally install software we already have
- Use inappropriate binaries for our architecture.
- Makes incorrect assumptions about the location Libraries.
- Is enormous filling your `$HOME`
- Installation is slower than using Python wheels.
- It modifies the `.bashrc` file, which can cause conflicts.

What options do you have?

In Compute Canada systems you have two main options:

1. Generate a python virtual environment:
 - a. Check python dependencies
 - b. Check program dependencies
 - c. Load and install dependencies

2. Install your package with conda within a singularity image
 - a. Create a singularity image
 - i. If available from dockerhub, you are done!
 - b. Install conda
 - c. Install your program

Python virtual environment

- Does not manage packages
- Provides a user-writable space to install python packages (mainly through pip)
- Sets a temporary user space (when activated) with the paths exported.
- DOES NOT CLASH WITH THE SYSTEM!!!!
- It uses optimal compilation of programs

Step 0: Download the source of the program you want

For this webinar we will be using a bioinformatics pipeline called [Anvi'o](#). You can follow their instructions. In brief [clone git](#):

```
[jshleap@cedar1 scratch]$ git clone --recursive https://github.com/merenlab/anvio.git
Cloning into 'anvio'...
remote: Enumerating objects: 246, done.
remote: Counting objects: 100% (246/246), done.
remote: Compressing objects: 100% (144/144), done.
remote: Total 77549 (delta 144), reused 179 (delta 102), pack-reused 77303
Receiving objects: 100% (77549/77549), 394.58 MiB | 2.44 MiB/s, done.
Resolving deltas: 100% (59056/59056), done.
Updating files: 100% (1175/1175), done.
```

Step 1: Identify python dependencies

Good packages will have a requirements.txt file, others, you will have to scrape conda's meta.yaml.

```
[jshleap@gra-login1 scratch]$ cd anvio/  
[jshleap@gra-login1 anvio]$ ls  
anvio  AUTHORS.txt  bin  CHANGELOG.md  conda-recipe  Dockerfile  LICENSE.txt  
MANIFEST.in  NEWS.md  README.md  requirements.txt  sandbox  setup.py  
[jshleap@gra-login1 anvio]$ ls conda-recipe/anvio*  
conda-recipe/anvio:  
meta.yaml  
  
conda-recipe/anvio-minimal:  
meta.yaml
```


Step 1.1: Create a virtual environment and install python dependencies

```
[jshleap@gra-login1 anvio]$ module load StdEnv/2020 gcc/9.3.0 python/3.8 r/4.0.2

lmod is automatically replacing "intel/2020.1.217" with "gcc/9.3.0".

lmod is automatically replacing "intel/2020.1.217" with "gcc/9.3.0".

[jshleap@gra-login1 anvio]$ virtualenv --no-download anvio_env
created virtual environment CPython3.8.2.final.0-64 in 2184ms
  creator CPython3Posix(dest=/scratch/jshleap/anvio/anvio_env, clear=False, global=False)
  seeder FromAppData(download=False, pip=latest, setuptools=latest, wheel=latest, via=copy, ap
p_data_dir=/home/jshleap/.local/share/virtualenv/seed-app-data/v1.0.1)
  activators BashActivator,CShellActivator,FishActivator,PowerShellActivator,PythonActivator,X
onshActivator
[jshleap@gra-login1 anvio]$ source ./anvio_env/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ pip install --no-index -r requirements.txt

(anvio_env) [jshleap@gra-login1 anvio]$ pip install --no-index numpy scikit-learn pandas --upg
rade
```


Step 2: Identify software dependencies and load the modules

We carry a long list of software, scrape the names on the conda's meta.yaml, and use module spider to check if we have it:

```
(anvio_env) [jshleap@gra-login1 anvio]$ module spider prodigal
```

```
prodigal: prodigal/2.6.3
```

```
Description:
  Prodigal (Prokaryotic Dynamic Programming Genefinding Algorithm) is
  a microbial (bacterial and archaeal) gene finding program developed
  at Oak Ridge National Laboratory and the University of Tennessee.

Properties:
  Bioinformatic libraries/apps / Logiciels de bioinformatique

You will need to load all module(s) on any one of the lines below before t
he "prodigal/2.6.3" module is available to load.

StdEnv/2020
nixpkgs/16.09  gcc/5.4.0
nixpkgs/16.09  gcc/7.3.0
```

Step 2.1: Add module load to activate (Optional)

If you'd like that all available requirements are loaded on activation:

```
(anvio_env) [jshleap@gra-login1 anvio]$ echo 'module load StdEnv/2020 gcc/9.3.0 scipy-stack/2020b \' >> ${VIRTUAL_ENV}/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ echo '  prodigal/2.6.3 mcl/14.137 hmmer/3.2.1 diamond/2.0.4 \' >> ${VIRTUAL_ENV}/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ echo '  blast+/2.11.0 spades/3.15.1 megahit/1.2.9 \' >> ${VIRTUAL_ENV}/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ echo '  bowtie2/2.4.1 bwa/0.7.17 samtools/1.11 centrifuge/1.0.4-beta \' >> ${VIRTUAL_ENV}/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ echo '  trimal/1.4 fastani/1.32  trnascan-se/2.0.7 fasttree/2.1.11 r/4.0.2' >> ${VIRTUAL_ENV}/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ tail -n 10 ${VIRTUAL_ENV}/bin/activate
# be called to get it to forget past commands.  Without forgetting
# past commands the $PATH changes we made may not be respected
if [ -n "${BASH-}" ] || [ -n "${ZSH_VERSION-}" ] ; then
    hash -r 2>/dev/null
fi
module load StdEnv/2020 gcc/9.3.0 scipy-stack/2020b \
  prodigal/2.6.3 mcl/14.137 hmmer/3.2.1 diamond/2.0.4 \
  blast+/2.11.0 spades/3.15.1 megahit/1.2.9 \
  bowtie2/2.4.1 bwa/0.7.17 samtools/1.11 centrifuge/1.0.4-beta \
  trimal/1.4 fastani/1.32  trnascan-se/2.0.7 fasttree/2.1.11 r/4.0.2
```

Step 2.2: Install the missing R components

Use local install:

```
(anvio_env) [jshleap@gra-login1 anvio]$ mkdir -p ${VIRTUAL_ENV}/R/lib
(anvio_env) [jshleap@gra-login1 anvio]$ export R_LIBS=${VIRTUAL_ENV}/R/lib
(anvio_env) [jshleap@gra-login1 anvio]$ echo "export R_LIBS=${VIRTUAL_ENV}/R/lib" >> ${VIRTUAL_ENV}/bin/activate
(anvio_env) [jshleap@gra-login1 anvio]$ R -e "install.packages(c('stringi', 'magrittr', 'optparse', 'rlang', 'Rcpp', 'tidyverse', 'BiocManager'), dependencies=TRUE, lib='${VIRTUAL_ENV}/R/lib', repos='http://cran.r-project.org', quiet=TRUE)"
```

```
(anvio_env) [jshleap@gra-login1 anvio]$ R -e "BiocManager::install('qvalue', lib='${VIRTUAL_ENV}/R/lib')"
```


Step 3: Install the missing software

Sometimes we do not have the required software, so you can install it in your virtual environment. In our case IQTree:

```
(anvio_env) [jshleap@gra-login1 anvio]$ cd ${VIRTUAL_ENV}
(anvio_env) [jshleap@gra-login1 anvio_env]$ git clone https://github.com/Cibiv/IQ-TREE.git
Cloning into 'IQ-TREE'...
remote: Enumerating objects: 861, done.
remote: Counting objects: 100% (861/861), done.
remote: Compressing objects: 100% (474/474), done.
remote: Total 30979 (delta 626), reused 565 (delta 387), pack-reused 30118
Receiving objects: 100% (30979/30979), 65.23 MiB | 23.96 MiB/s, done.
Resolving deltas: 100% (23141/23141), done.
Updating files: 100% (638/638), done.
(anvio_env) [jshleap@gra-login1 anvio_env]$ cd IQ-TREE && mkdir build && cd build
```

```
(anvio_env) [jshleap@gra-login1 build]$ module load eigen/3.3.7
(anvio_env) [jshleap@gra-login1 build]$ cmake -DIQTREE_FLAGS=omp -DEIGEN3_INCLUDE_DIR=${EBROOT
EIGEN} -DCMAKE_INSTALL_PREFIX=${VIRTUAL_ENV} ..
```

```
(anvio_env) [jshleap@gra-login1 build]$ make -j 4
```

```
(anvio_env) [jshleap@gra-login1 build]$ chmod +x iqtrees
(anvio_env) [jshleap@gra-login1 build]$ cp iqtrees ${VIRTUAL_ENV}/bin
(anvio_env) [jshleap@gra-login1 build]$ cd ../../
```

Step 4: Install Anvi'o

Now that we have all the requisites, installing avio is simple:

```
(anvio_env) [jshleap@gra-login1 anvio]$ pip install .  
Ignoring pip: markers 'python_version < "3"' don't match your environment  
Looking in links: /cvmfs/soft.computecanada.ca/custom/python/wheelhouse/gentoo/avx2, /cvmfs/so  
ft.computecanada.ca/custom/python/wheelhouse/gentoo/generic, /cvmfs/soft.computecanada.ca/cust  
om/python/wheelhouse/generic  
Processing /scratch/jshleap/anvio
```

We can test it, but first deactivate and reactivate the environment:

```
(anvio_env) [jshleap@gra-login1 anvio]$ deactivate  
[jshleap@gra-login1 anvio]$ source anvio_env/bin/activate  
  
Lmod is automatically replacing "intel/2020.1.217" with "gcc/9.3.0".  
  
Lmod is automatically replacing "intel/2020.1.217" with "gcc/9.3.0".  
  
(anvio env) [jshleap@gra-login1 anvio]$ anvi-self-test
```

Step 1: Build a singularity image

Some programs can be found in [dockerhub](#) like [Anvi'o](#). In those cases your life just about to become a lot easier:

```
[jshleap@gra-login1 jshleap]$ module load singularity
[jshleap@gra-login1 jshleap]$ singularity build anvio.sif docker://meren/anvio
INFO: Starting build...
```

If not, don't worry, we can still build a singularity image:

```
[jshleap@gra-login1 jshleap]$ singularity build --sandbox anvio-dir docker://centos
INFO: Starting build...
```

Singularity Image with conda

- Self contained space
 - All variables and binary paths are set for you
 - Easier to set up
-
- It is suboptimal to run in our systems (compilation)

Step 2: If not in docker hub, install conda

```
[jshleap@gra-login1 scratch]$ singularity build --sandbox anvio_from_source-dir docker://centos
INFO: Starting build...
```

```
[jshleap@gra-login1 jshleap]$ singularity shell -e -w -B /usr anvio-dir
Singularity> wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
--2021-03-23 15:11:49-- https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 2606:4700::6810:8303, 2606:4700::6810:8203,
104.16.131.3, ...
Connecting to repo.anaconda.com (repo.anaconda.com)|2606:4700::6810:8303|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 94235922 (90M) [application/x-sh]
Saving to: 'Miniconda3-latest-Linux-x86_64.sh'

100%[=====>] 94,235,922 84.2MB/s in 1.1s

2021-03-23 15:11:50 (84.2 MB/s) - 'Miniconda3-latest-Linux-x86_64.sh' saved [94235922/94235922]

Singularity> bash Miniconda3-latest-Linux-x86_64.sh -b -p /opt/conda
```

Step 2.1: Activate conda

```
Singularity> mkdir -p /localhome
Singularity> echo 'export HOME=/localhome' >> /environment
Singularity> source /environment
Singularity> source /opt/conda/bin/activate
(base) Singularity> conda init
no change      /opt/conda/condabin/conda
no change      /opt/conda/bin/conda
no change      /opt/conda/bin/conda-env
no change      /opt/conda/bin/activate
no change      /opt/conda/bin/deactivate
no change      /opt/conda/etc/profile.d/conda.sh
no change      /opt/conda/etc/fish/conf.d/conda.fish
no change      /opt/conda/shell/condabin/Conda.psm1
no change      /opt/conda/shell/condabin/conda-hook.ps1
no change      /opt/conda/lib/python3.8/site-packages/xontrib/conda.xsh
no change      /opt/conda/etc/profile.d/conda.csh
modified       /localhome/.bashrc

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

(base) Singularity> echo ". /localhome/.bashrc" >> /environment
```


Step 3: Download, install, and activate your recipe

```
(base) Singularity> curl https://merenlab.org/files/anvio-conda-environments/anvio-environment-7-LINUX.yaml --output anvio-environment-7.yaml
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
100  9650  100  9650    0     0  13049      0 --:--:-- --:--:-- --:--:-- 13040
(base) Singularity> conda env remove --name anvio-7
(base) Singularity> conda env create -f anvio-environment-7.yaml
Collecting package metadata (repodata.json): \
```

```
[jshleap@gra-login1 jshleap]$ singularity shell -e -w anvio-dir
Singularity> echo "conda init" >> /environment
Singularity> echo "conda activate anvio-7" >> /environment
Singularity> . /environment
no change      /opt/conda/condabin/conda
no change      /opt/conda/bin/conda
no change      /opt/conda/bin/conda-env
no change      /opt/conda/bin/activate
no change      /opt/conda/bin/deactivate
no change      /opt/conda/etc/profile.d/conda.sh
no change      /opt/conda/etc/fish/conf.d/conda.fish
no change      /opt/conda/shell/condabin/Conda.psm1
no change      /opt/conda/shell/condabin/conda-hook.ps1
no change      /opt/conda/lib/python3.8/site-packages/xontrib/conda.xsh
no change      /opt/conda/etc/profile.d/conda.csh
no change      /localhome/.bashrc
No action taken.
(anvio-7) Singularity> conda install -y -c bioconda hmmer=3.2.1
```

Step 3.1: Keep following instructions

Most programs will have a set of steps to follow... just follow those. For Anvi'o:

```
Singularity> cd /opt/  
Singularity> wget https://github.com/merenlab/anvio/releases/download/v7/anvio-7.tar.gz
```

```
Singularity> pip install anvio-7.tar.gz  
Processing ./anvio-7.tar.gz
```

```
Singularity> pip install mistune==0.8.4  
Requirement already satisfied: mistune==0.8.4 in ./conda/envs/anvio-7/lib/python3.6/site-packages (0.8.4)
```

Step 4: Make it an image (optional) and test

```
[jshleap@gra-login1 jshleap]$ chmod -R 755 anvio-dir/  
[jshleap@gra-login1 jshleap]$ singularity build anvio.sif anvio-dir  
INFO: Starting build...  
INFO: Creating SIF file...  
INFO: Build complete: anvio.sif  
[jshleap@gra-login1 jshleap]$ singularity run -e anvio.sif anvio-self-test
```

```
[jshleap@gra-login1 jshleap]$ singularity run -e -w anvio-dir anvio-self-test
```

Q&A

Help can also be reached at support@compute canada.ca or help@sharcnet.ca.

Commands can be found at:

1. Python environment

https://staff.sharcnet.ca/jshleap/GI_webinars/FauxConda/Commands_environment.sh

2. Singularity

https://staff.sharcnet.ca/jshleap/GI_webinars/FauxConda/Commands_singularity.sh

This presentation (minus the show and tell) can be found:

https://staff.sharcnet.ca/jshleap/GI_webinars/FauxConda/FauxConda.pdf

Recording of this presentation can be found at (but give it some time):

<https://youtube.sharcnet.ca>