Installing software on Alpine with Conda and Mamba

Research Computing
UNIVERSITY OF COLORADO BOULDER

Installing software on Alpine with Conda and Mamba

Date: February 13, 2024

Instructor: Andrew Monaghan

Contributors: Layla Freeborn, Trevor Hall, Brandon Reyes

- Website: www.rc.colorado.edu
- Documentation: https://curc.readthedocs.io
- Helpdesk: rc-help@colorado.edu
- Survey: http://tinyurl.com/curc-survey18

Slides

https://github.com/ResearchComputing/alpine_conda_mamba_primer

Session Overview

Introduction
- Installing software on CURC systems
- Description of Conda

Setting up Conda on Alpine
- Logging in
- Using conda for the first time: creating the ~/.condarc file
- Starting an interactive session and activating conda

Creating and Modifying Virtual Environments with Conda
- Creating/activating/modifying a python environment
- Useful conda commands and paths

Using Conda Virtual Environments
- In HPC jobs
- In OnDemand Jupyter

Strategies for installing complex Virtual Environments (Discussion only)
- Channels
- Resolving conflicts upon environment creation
- Mamba

Building Software on Alpine

- There are numerous ways to install software on Alpine:
  - grab pre-compiled binaries
  - compile from source
  - **within virtual environments (via Conda, Miniconda, or Mamba)**
  - using containers (Apptainer/Singularity)
  - using a package manager for HPC systems (Spack)

Additional information:
https://github.com/ResearchComputing/research-software-curc

Virtual Environments With CONDA

- Conda is a package (software) management system
  - installs, runs, and updates packages *and their dependencies*
  - creates, saves, loads, and switches between virtual environments
  - created for Python programs, but can package and distribute software for any language

Additional information:
https://curc.readthedocs.io/en/latest/software/python.html

Logging into CU Research Computing

login to CURC via your terminal:

```
ssh monaghaa@login.rc.colorado.edu
```

…or login to CURC via your browser:

https://ondemand-rmacc.rc.colorado.edu

(once logged in, navigate to **Clusters -> Alpine shell**)

Additional information:
https://curc.readthedocs.io/en/latest/access/logging-in.html
https://curc.readthedocs.io/en/latest/gateways/OnDemand.html

Setting up Conda for the first time

Create a new ~/.condarc configuration file in your editor:

```
[monaghaa@login11 ~]$ nano ~/.condarc
```

Paste the following text in the file:

```
pkgs_dirs:
  - /projects/$USER/.conda_pkgs
envs_dirs:
  - /projects/$USER/software/anaconda/envs
```

Save and exit the editor by typing **CTRL-o** then **CTRL-x**

Additional information:
https://curc.readthedocs.io/en/latest/software/python.html#configuring-conda-with-condarc
https://conda.io/projects/conda/en/latest/user-guide/configuration/use-condarc.html

Start a session and activate conda

Start a session on an Alpine compute node with **acompile**:

```
[monaghaa@login11 ~]$ acompile --help
[monaghaa@login11 ~]$ acompile --time=90:00
…
[monaghaa@c3cpu-a5-u28-1 ~]$ module load anaconda
(base) [monaghaa@c3cpu-a5-u28-1 ~]$
```

Note: when you login to CURC you'll be on a ***login*** node. You'll need to be on a ***compute*** node to use anaconda. The **acompile** command allows you to quickly start an interactive job on a compute node.

Additional information:
https://curc.readthedocs.io/en/latest/clusters/alpine/alpine-hardware.html#partitions

Create your first conda environment!

- Environments are created and programs are installed in a few simple steps

```
(base) [monaghaa@c3cpu-a5-u28-1 ~]$ conda create -n my_first_env python==3.10
(base) [monaghaa@c3cpu-a5-u28-1 ~]$ conda activate my_first_env
(my_first_env) [monaghaa@c3cpu-a5-u28-1 ~]$ python
```

Don't install packages in your base environment!

Additional information:
https://curc.readthedocs.io/en/latest/software/python.html#create-your-own-custom-environment

Install packages with “conda install”

- Packages are installed within **activated** environments

using **conda install** to install latest version or specific version:

```
(my_first_env) [monaghaa@c3cpu-a5-u28-1 ~]$ conda install pandas
(my_first_env) [monaghaa@c3cpu-a5-u28-1 ~]$ conda install pandas==0.20.3
```

Additional information:
https://curc.readthedocs.io/en/latest/software/python.html#using-conda

Install packages with “pip”

- Packages are installed within **activated** environments

using **pip** to install latest version or specific version:

```
(my_first_env) [monaghaa@c3cpu-a5-u28-1 ~]$ pip install --no-cache-dir pandas
```

`--no-cache-dir` is crucial on CURC systems!

Useful Conda Commands

```
conda env list                          # list all environments
conda list                              # list packages in active env
conda env remove -n <envname> --all # remove an environment
conda config --show channels            # view configured channels
conda deactivate                        # deactivate environment
conda create --name <clonedenv> /   # clone an environment
      --clone <envtoclone>
```

Additional information:
https://curc.readthedocs.io/en/latest/software/python.html#basic-conda-commands-to-get-you-started

Useful conda file paths on Alpine

```
# location of python libraries
/projects/$USER/software/anaconda/<env>/lib/python3.10/site-packages

# location of package executables
/projects/$USER/software/anaconda/<env>/bin

# location of .condarc file
/home/$USER/.condarc
```

Running Alpine batch jobs with conda

```
[monaghaa@login11 ~]$ nano runconda.sh  #Step 1: open new job script in editor
```

```
#!/bin/bash
# job script name: runconda.sh

#SBATCH --partition=amilan
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --time=10:00

module purge
module load anaconda
conda activate my_first_env
python my_python_code.py
```

```
# Step 2: Write job script
```

https://curc.readthedocs.io/en/latest/running-jobs/batch-jobs.html

```
[monaghaa@login11 ~]$ sbatch runconda.sh  #Step 3: Schedule job
```

Using your conda environment in Jupyter

Step 1: create a kernel *within* your **activated** environment:

```
(my_first_env) [monaghaa@c3cpu-a5-u28-1 ~]$ conda install -c conda-forge jupyterlab
(my_first_env) [monaghaa@c3cpu-a5-u28-1 ~]$ python -m ipykernel install --user --
name myfirst_env --display-name my_first_env
```

Step 2: Use the environment in the OnDemand Jupyter app in your browser:

* https://ondemand-rmacc.rc.colorado.edu
* Navigate to "Interactive Apps" then "Jupyter Session (presets)"
* Launch a session then open a notebook with "my_first_env"

Additional information:
https://curc.readthedocs.io/en/latest/gateways/OnDemand.html#creating-a-jupyter-session-conda-environment

Strategies for complex environments

- Using channels and channel order
- Resolving conflicts upon environment creation
- Using Mamba to accelerate installation

Additional information:
https://curc.readthedocs.io/en/latest/software/python.html
https://curc.readthedocs.io/en/latest/software/python.html#mamba-package-manager

Thank you!

Survey and feedback

http://tinyurl.com/curc-survey18