Python and Conda Environments in HPC: *From Basics to Best Practices*

Hui (Julia) Zhao
NJIT High Performance Computing

Outline

- Why High Performance Computing
- How to access Python on Wulver at HPC
- Introduction to Conda environments
- Install, uninstall and upgrade packages
- Best Practices for managing conda environments
- Common Python libraries for scientific computing

Why High Performance Computing?

- Handling Complex Problems
- Big Data Analysis
- Speeding up Research
- Parallel Computing
- Resource Sharing and Collaboration

Why Use Python for HPC?

- **Clear Syntax** – Simple, readable, and easy to learn
- **Extensive Libraries** – Optimized packages for scientific computing
- **Multi-language Integration** – Works seamlessly with C, C++, and Fortran
- **Parallel Computing Capabilities** – Supports multi-threading & distributed computing
- **Strong Community Support** – Actively maintained & widely adopted

Python on Wulver

Software	Version	Dependent Toolchain	Module Load Command
Python	3.9.6	foss/2021b	`module load foss/2021b Python/3.9.6`
Python	3.11.5	foss/2023b	`module load foss/2023b Python/3.11.5`
Python	3.10.8	foss/2022b	`module load foss/2022b Python/3.10.8`

Installing Python packages

Method 1: Installing Python Packages from Source

1 Clone the repository

$ git clone https://github.com/pandas-dev/pandas.git

2 Navigate into the package directory

$ cd pandas

3 Install the package to a custom location

$ python setup.py install --prefix=/project/$GROUP/$USER/python_pkg/

Possible Installation Error

Error Message:

Traceback (most recent call last):

File "/usr/lib64/python3.6/site-packages/numpy/core/__init__.py", line 16, in <module>

from . import multiarray

ImportError: libopenblasp.so.0: cannot open shared object file: No such file or directory

Reason: The required shared library (libopenblasp.so.0) is missing or not found.

Installing Python packages - PiP

Method 2: pip

pip stands for **“Preferred Installer Program”**
A package manager for Python packages **only**
Installs packages from the **Python Package Index (PyPI)**

$ python -m pip install --user <python-module-name> --no-cache-dir

-m <module-name>: Always use python -m pip instead of just pip
Ensures pip runs using the **correct Python interpreter**
Avoids conflicts with multiple Python installations

--user Flag: Installs packages **to user account only**
Ensures installation **without admin/root privileges**
Useful on shared HPC systems

--no-cache-dir Option: Prevents pip from storing package caches in the home directory
Saves disk space, especially in HPC environments

Method 3: Conda

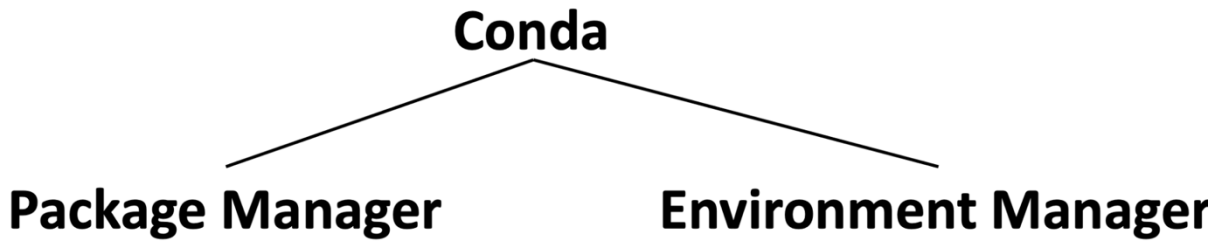
Conda on HPC

- **Introduction to Conda**
- Conda channels
- Conda environment
- Conda packages
- Sharing environments

Introduction to Conda

- **Conda is an open-source package and environment manager**
  Supports **Python and non-Python** packages
  Works across **Windows, macOS, and Linux**

- **Conda is a powerful package & environment manager**



Why use Conda?

Key Benefits

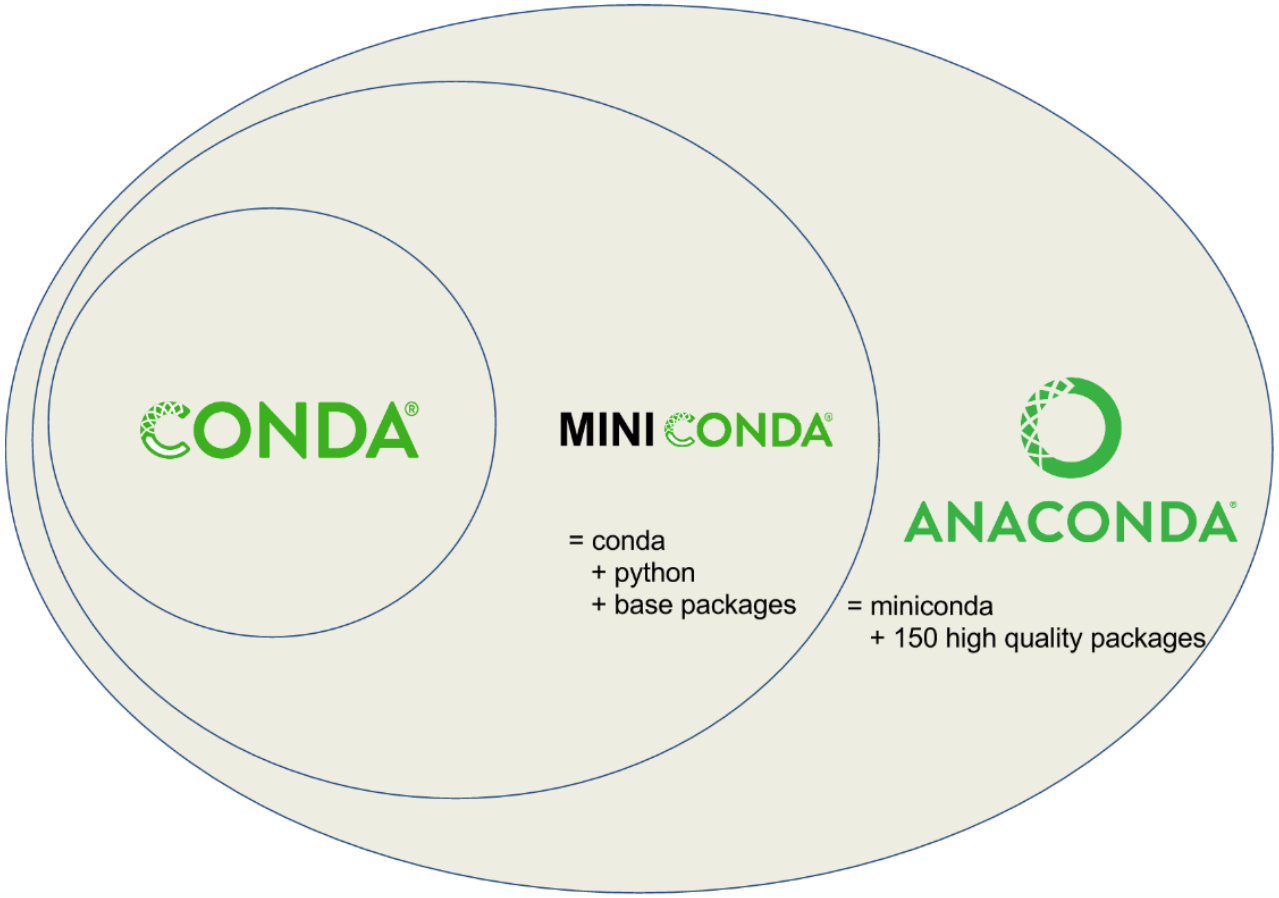
- Simplifies installation, dependency resolution, and reproducibility
- Ideal for scientific computing, data science, and HPC
- Manages dependencies automatically
- Creates isolated environments to prevent package conflicts
- Supports multiple programming languages, including R, C, and Java

Additional Advantages

- Ensures smooth package management across different platforms
- Prevents version conflicts with isolated environments
- Optimized for performance and scalability in scientific applications

NJIT

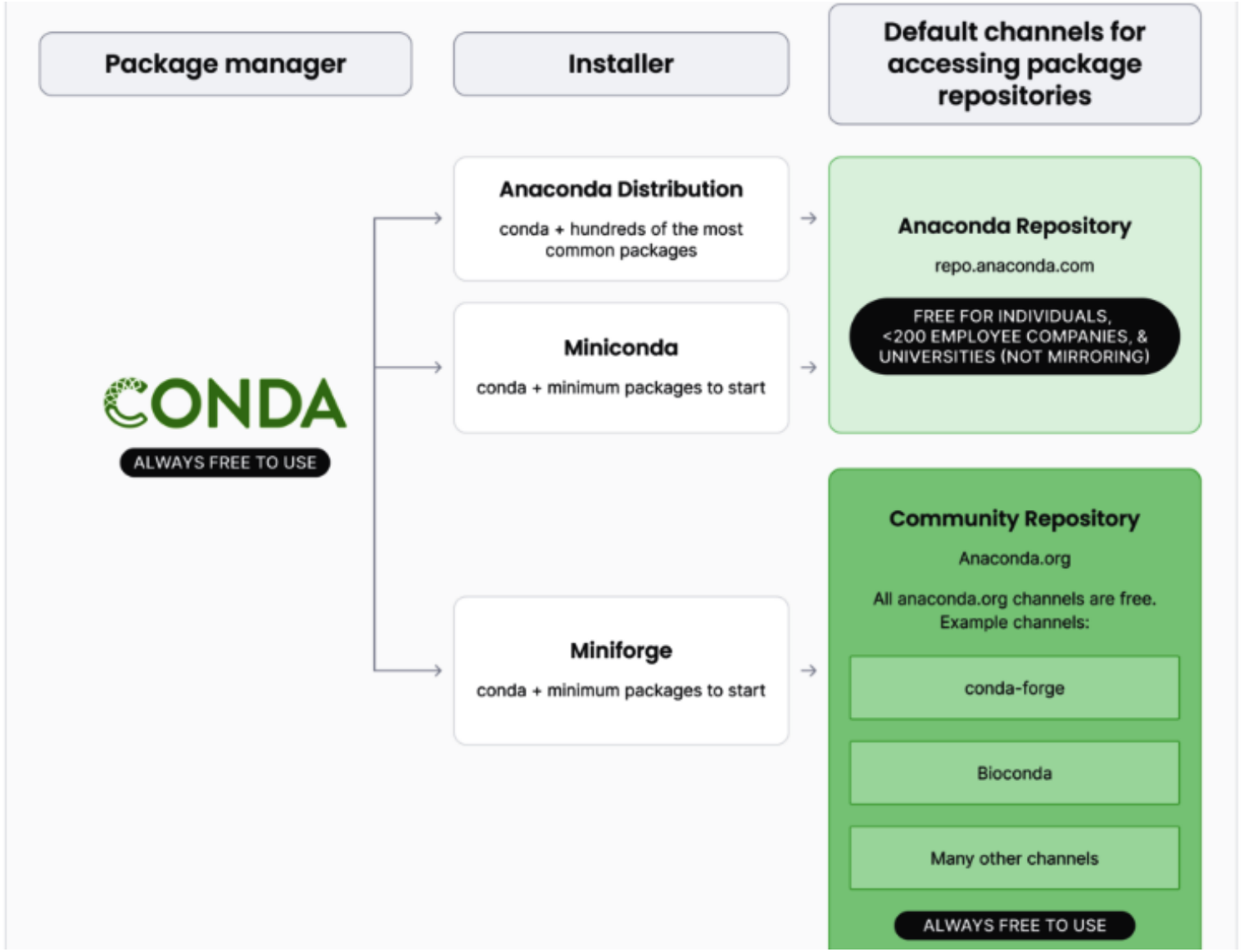
Anaconda vs Miniconda vs Conda



- Conda: Open-source package manager
- Anaconda: A software distribution: open-source (personal) and Commercial
- Miniconda: minimal installer for conda

NJIT

Anaconda Portfolio



Load Miniforge Module on Wulver

Load Miniforge3 Module

$ module load Miniforge3

Check Loaded Modules

$ module list

Currently Loaded Modules:

1) easybuild 3) slurm/wulver 5) Miniforge3/24.1.2-0

2) wulver 4) null

What is Miniforge

$ module whatis Miniforge3

Miniforge3/24.1.2-0 : Description: Miniforge is a free minimal installer for conda and Mamba specific to conda-forge.

Miniforge3/24.1.2-0 : Homepage: https://github.com/conda-forge/miniforge

Miniforge3/24.1.2-0 : URL: https://github.com/conda-forge/miniforge

Conda info

```
g07396@login02:~
$ conda info

     active environment : None
            shell level : 0
       user config file : /home/g07396/.condarc
 populated config files : /mmfs1/apps/easybuild/software/Miniforge3/24.1.2-0/.condarc
          conda version : 24.1.2
    conda-build version : not installed
         python version : 3.10.14.final.0
                 solver : libmamba (default)
       virtual packages : __archspec=1=skylake_avx512
                          __conda=24.1.2=0
                          __glibc=2.28=0
                          __linux=4.18.0=0
                          __unix=0=0
       base environment : /mmfs1/apps/easybuild/software/Miniforge3/24.1.2-0  (read only)
      conda av data dir : /mmfs1/apps/easybuild/software/Miniforge3/24.1.2-0/etc/conda
  conda av metadata url : None
          channel URLs : https://conda.anaconda.org/conda-forge/linux-64
                         https://conda.anaconda.org/conda-forge/noarch
         package cache : /mmfs1/apps/easybuild/software/Miniforge3/24.1.2-0/pkgs
                         /home/g07396/.conda/pkgs
       envs directories : /home/g07396/.conda/envs
                         /mmfs1/apps/easybuild/software/Miniforge3/24.1.2-0/envs
             platform : linux-64
           user-agent : conda/24.1.2 requests/2.31.0 CPython/3.10.14 Linux/4.18.0-372.26.1.el8_6.x86_64 rhel/8.6
glibc/2.28 solver/libmamba conda-libmamba-solver/24.1.0 libmambapy/1.5.7
              UID:GID : 580857:580857
           netrc file : None
         offline mode : False
```

Conda on HPC

- Introduction to Conda
- **Conda channels**
- Conda environment
- Conda packages
- Sharing environments

Configuring Conda channels

A conda channel is a repository of conda packages

$ conda config **--help**

$ conda config **--show**

$ conda config **--show channels**

channels:
- conda-forge
- defaults

$conda config **--describe channels**

$conda config **--add channels conda-forge**

This would add the conda-forge channel to the top of the channel list.

$conda config **--append** channels conda-forge

This would add the conda-forge to the end of the channel list, giving it the lowest priority.

Conda on HPC

- Introduction to Conda
- Conda channels
- **Conda environment**
- Conda packages
- Sharing environments

Why create a Conda environment?

A conda environment is a directory that contains a specific collection of conda packages.

Isolation from other projects

Control Over Packages

- Manage versions and dependencies.

Reproducibility

- Consistent setups across systems.

Dependency Management

- Handles Python and non-Python dependencies.

Python Versatility

- Manage and switch Python versions easily.

Ease of Use

- User-friendly commands for project management.

Cross-Platform

- Works on Linux, Windows, and macOS.

Commonly used Conda commands

Task	Command
Activate environment:	`conda activate [environment_name]`
Deactivate environment:	`conda deactivate [environment_name]`
Show the list of environments:	`conda env list`
Delete environment:	`conda remove [environment_name]`
Export environment:	`conda env export > [environment_name].yml`
Import environment from YAML:	`conda env create -f [environment_name].yml`
Import environment to different location:	`conda env create -f [environment_name].yml -p [PATH]`

Conda cheat sheet - Link to Conda Doc for more helpful commands

Creating Conda Environment

Creating a new conda environment

$ conda create --name my_env

Creating a new conda environment with a specific python version

$ conda create --name my_env python=3.9

Creating a new conda environment with a specific python version and scipy package

$ conda create --name my_env python=3.9 scipy=0.15.0

Creating a new conda environment in difference location with **--prefix** or **-p**

$ conda create --prefix /project/$GROUP/$USER/conda_env AAA

Enter, Exit and Remove conda environment

Entering a Conda environment

$ **conda activate my_env**

$ **conda activate /project/$GROUP/$USER/conda_env/AAA**

Exiting a Conda environment we are currently in

$ **conda deactivate**

Removing a Conda environment

$ **conda env remove -n my_env**

Renaming a Conda environment

$conda rename -n old_env_name new_env_name

List conda environments

A user may list all shared virtual environments and your own private virtual environments

```
[n0088:~ hz3$ conda info --envs
# conda environments:
#
base                     /apps/easybuild/software/Anaconda3/2023.09-0
my_env                   /home/hz3/.conda/envs/my_env
tensorflow               /home/hz3/.conda/envs/tensorflow
tf                       /home/hz3/.conda/envs/tf
tf-gpu                   /home/hz3/.conda/envs/tf-gpu
                         /project/hpcadmins/hz3/conda_env/my_env

[n0088:~ hz3$ conda env list
# conda environments:
#
base                     /apps/easybuild/software/Anaconda3/2023.09-0
my_env                   /home/hz3/.conda/envs/my_env
tensorflow               /home/hz3/.conda/envs/tensorflow
tf                       /home/hz3/.conda/envs/tf
tf-gpu                   /home/hz3/.conda/envs/tf-gpu
                         /project/hpcadmins/hz3/conda_env/my_env
```

Conda on HPC

- Introduction to Conda
- Conda environment
- Conda channels
- **Conda packages**
- Sharing environments

Conda packages

A conda package is a compressed tarball file

List All Installed Packages:

- $ **conda list**
- Displays all packages installed in the active Conda environment.

List Packages in a Specific Environment:

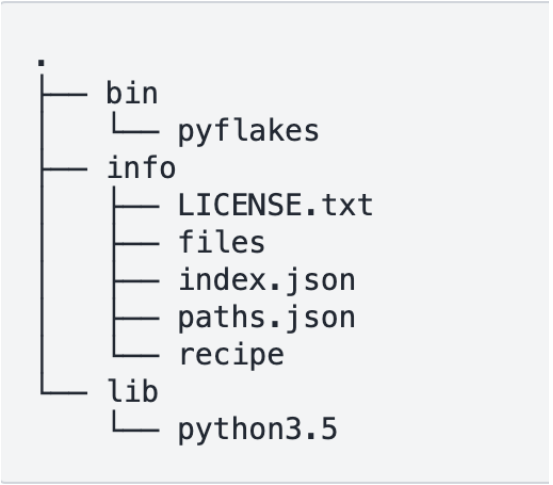
$ **conda list -n env_name or conda list -p /path/to/environment**

Search for a Package:

- $ **conda search package_name**
- Searches for a package across all channels in Conda.

Check for Specific Package Installation:

- $ **conda list | grep package_name**
- Filters the list of installed packages to show only the entries related to package_name.



List packages in all environments

```
[n0088:~ hz3$ conda list
# packages in environment at /apps/easybuild/software/Anaconda3/2023.09-0:
#
# Name                    Version                   Build  Channel
_anaconda_depends         2023.09            py311_mkl_1
_libgcc_mutex             0.1                       main
_openmp_mutex             5.1                       1_gnu
abseil-cpp                20211102.0          hd4dd3e8_0
aiobotocore               2.5.0            py311h06a4308_0
aiofiles                  22.1.0           py311h06a4308_0
aiohttp                   3.8.5            py311h5eee18b_0
aioitertools              0.7.1               pyhd3eb1b0_0
aiosignal                 1.2.0               pyhd3eb1b0_0
aiosqlite                 0.18.0           py311h06a4308_0
alabaster                 0.7.12              pyhd3eb1b0_0
anaconda-anon-usage       0.4.2            py311hfc0e8ea_0
anaconda-catalogs         0.2.0            py311h06a4308_0
anaconda-client           1.12.1           py311h06a4308_0
anaconda-cloud-auth       0.1.3            py311h06a4308_0
anaconda-navigator        2.5.0            py311h06a4308_0
anaconda-project          0.11.1           py311h06a4308_0
```

List packages in an environment

```
[n0088:~ hz3$ conda list -n my_env
# packages in environment at /home/hz3/.conda/envs/my_env:
#
# Name                    Version                   Build  Channel
_libgcc_mutex             0.1                 conda_forge    conda-forge
_openmp_mutex             4.5                       2_gnu    conda-forge
alm                       2.0.0_dev.2     py312h63811a6_8    conda-forge
blas                      1.0                         mkl
bzip2                     1.0.8                h7b6447c_0
ca-certificates           2024.2.2             hbcca054_0    conda-forge
expat                     2.5.0                h6a678d5_0
icu                       73.2                 h59595ed_0    conda-forge
intel-openmp              2023.1.0          hdb19cb5_46306
ld_impl_linux-64          2.38                 h1181459_1
libblas                   3.9.0           1_h86c2bf4_netlib    conda-forge
libboost                  1.82.0               h6fcfa73_6    conda-forge
libboost-python           1.82.0          py312hfb10629_6    conda-forge
libexpat                  2.5.0                hcb278e6_1    conda-forge
libffi                    3.4.4                h6a678d5_0
```

List the installed packages for the present environment

(myenv) $ conda list

Installing Conda packages

- Entering a Conda environment
  - $ **conda activate my_env**
  - (my_env) $: **conda install scipy=1.6 --channel conda-forge**

- Create an environment called my_biowork-env and install blast from the bioconda channel:
  - $ **conda create --name my_biowork-env blast --channel bioconda**

- The name flag can be used to specify the environment in which we install the package
  - $ **conda install -n my_env scipy**

$ **conda install conda-forge::tensorflow --prefix /project/$GROUP/$USER/my_env**

Mamba

Mamba is a reimplementation of the conda package manager in C++ for maximum efficiency

- Parallel downloading of repository data and packages files using multi-threading

- Libsolv for much faster dependency solving
- a *drop-in* replacement for conda
- Same commands as conda
- Robust and fast but not 100% drop-in replacement yet (especially for conda-env commands)

https://mamba.readthedocs.io/en/latest/

Mamba on Wulver

```
module load Miniforge3

# create new environment
mamba create --name env_name python numpy pandas
# install a new package into an existing environment
conda activate env_name
mamba install scipy
```

Mamba on wulver

Conda on HPC

- Introduction to Conda
- Conda environment
- Conda channels
- Conda packages
- **Sharing environments**

Exporting Conda environment

Export a conda environment to a new directory or a different machine

1. activate the environment first that you intend to export.
2. export it to a YAML file:

$ conda env export > my_environment.yml

```
name: my_env
channels:
- defaults
dependencies:
- _libgcc_mutex=0.1=main
- _openmp_mutex=5.1=1_gnu
- blas=1.0=mkl

<ouput snipped>

#the last line is the path of the env
prefix: /home/a/abc3/.conda/envs/my_env.
```

Create an Conda environment from yml file

- First load Miniforge
- Create the environment from the YAML file:

```
conda env create -f my_environment.yml
Collecting package metadata (repodata.json): done
Solving environment: done

<ouput snipped>

Downloading and Extracting Packages
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
# $ conda activate my_env
#
# To deactivate an active environment, use
#
# $ conda deactivate
```

Importing Conda environment to a new location

Use the **--prefix** or **-p** option to **specify the environment location**

- $ **conda env create -f my_environment.yml -p /project/$GROUP/$USER/conda_env/my_env**
- This will create the environment in the specified directory instead of the default conda environment directory.

Provide the full path of the environment to activate it.

- $ **conda activate /project/$GROUP/$USER/conda_env/my_env**
- $ **conda env list**
  - # conda environments:
  - base /mmfs1/apps/easybuild/software/Miniforge3/24.1.2-0
  - * /project/$GROUP/$USER/conda_env/my_env

Updating a Conda environment

When to update your conda environment?

- One of your core dependencies just released a new version
- You need an additional package for data analysis (add a new dependency).
- You have found a better visualization package and no longer need to old visualization package

Update the contents of your environment.yml file and run the following command:

$ conda env update **--file** environment.yml **--prune**

- **--prune** option tells Conda to remove any dependencies that are no longer required from the environment

Best practices

- **Use interactive sessions on a compute node**
  - Use an interactive session on a compute node to install software with Conda
  - **$ srun -p general -n 1 --qos=standard --account=$PI_ucid --mem-per-cpu=2G --time=59:00 --pty bash** *#modify sl options as desired*
  - **$ interactive -a $PI_UCID -q standard -j cpu**

- **Use /project directory with large quotas**
  - Use /project/$PI/$USER directory other than the home directory for conda environments and packages. Using your home directory can fill its limited space.
  - Managing Conda Cache and changing the default caching behavior

- **Avoid installing packages into your base Conda environment**

Configuring Conda Package Cache

Default Location: **$HOME/.conda/pkgs**

Check Current Cache Directory: **conda info**

Change Cache Location**:**

- Edit **.condarc**
  pkgs_dirs:
  - /path/to/desired/cache/directory
- Use Conda Command:
  conda config --add pkgs_dirs /project/$GROUP/$USER/conda_env/pkgs_dirs
- Set Environment Variable:
  export CONDA_PKGS_DIRS=/path/to/desired/cache/directory

Verify Change: **conda info**

More Options: **official .condarc user guide**

PiP vs Conda

✅ **Favor Conda over Pip** whenever possible
✅ **Use Conda first**, then Pip only if necessary

Why Choose Conda?

- **Pre-compiled packages** – No need to build from source
  Automatic dependency resolution – Handles package conflicts
  Better for scientific computing – Optimized for numerical libraries

When to Use Pip?

- If the package **is not available** in Conda
  If you need **the latest version** of a package $ pip install latest-package

Pip drawbacks

- Dependencies may need manual resolution
  Possible compatibility issues with Conda-installed packages

NJIT

Pip installs in a Conda environment

Recommend

✅ **Use Conda environments for isolation**
✅ **Always install Conda packages first**, then use **pip**
⚠ **Avoid installing Conda packages after using pip**

Create and activate a Conda environment

$ **conda create --name my_env pandas**
$ **conda activate my_env**
Install additional packages with **pip**
(my_env)$ **python -m pip install --user multiregex**

Recreate the environment if you need to modify packages after using pip
⚠ **Use --no-cache-dir** to prevent pip from filling your home directory
(my_env)$ **python -m pip install --no-cache-dir package_name**

Refer to Conda guide for using pip in a Conda environment

Common Python libraries for scientific computing

Library	Key features	Common Use Case
Numpy	Multidimensional arrays, Broadcasting, Vectorization	Mathematical operations, Basic statistics
SciPy	Numerical integration, Optimization, Linear algebra	Solving differential equations, Signal processing
Matplotlib	2D and 3D plotting, Customizable plots	Visualizing data, Scientific charts
Pandas	DataFrame and Series, Data manipulation, Cleaning	Data analysis, Time series analysis
Scikit-learn	Machine learning algorithms, Data preprocessing tools	Classification, Regression, Clustering
TensorFlow	Computational graph, Automatic differentiation	Building deep learning models, Neural networks
PyTorch	Dynamic computational graph, TorchScript for deployment	Machine learning, Computer vision

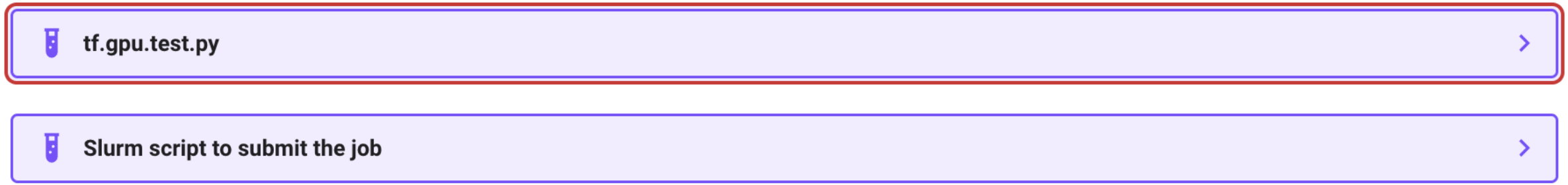
Example - install tensorflow-gpu

$conda create --name tensorflow python=3.9

$conda activate tensorflow

$conda install -c anaconda tensorflow-gpu numpy=1.21.6

Simple TensorFlow test program to make sure the virtual env can access a GPU.



https://hpc.njit.edu/Software/programming/python/conda/#install-tensorflow-with-gpu

Example - Install PyTorch with GPU

$conda create --name torch-cuda python=3.10

$conda activate torch-cuda

$conda install -c "nvidia/label/cuda-12.2.0" cuda-toolkit

$conda install -c pytorch -c nvidia pytorch torchvision torchaudio pytorch-cuda -y

A simple PyTorch test program is given below to check whether PyTorch has been installed properly. Program is called

User can use the following job script to run the script.

https://hpc.njit.edu/Software/programming/python/conda/#install-tensorflow-with-gpu

Connect with Us

- Open a ticket using email: hpc@njit.edu
- Request Software: HPC Software Installation
- Consult with Research Computing Facilitator: HPC User Assistance
- Further information: HPC at NJIT
- System updates
    - Read Message of the Day on login
    - Visit NJIT HPC News

Questions?