

Managing Software Packages with CONDA Virtual Environment

Jason Li

HPC User Services

LSU HPC / LONI

sys-help@loni.org

Louisiana State University

Baton Rouge

Mar 29, 2023

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

1. Why Conda?

- 1) Scenarios
- 2) Concepts

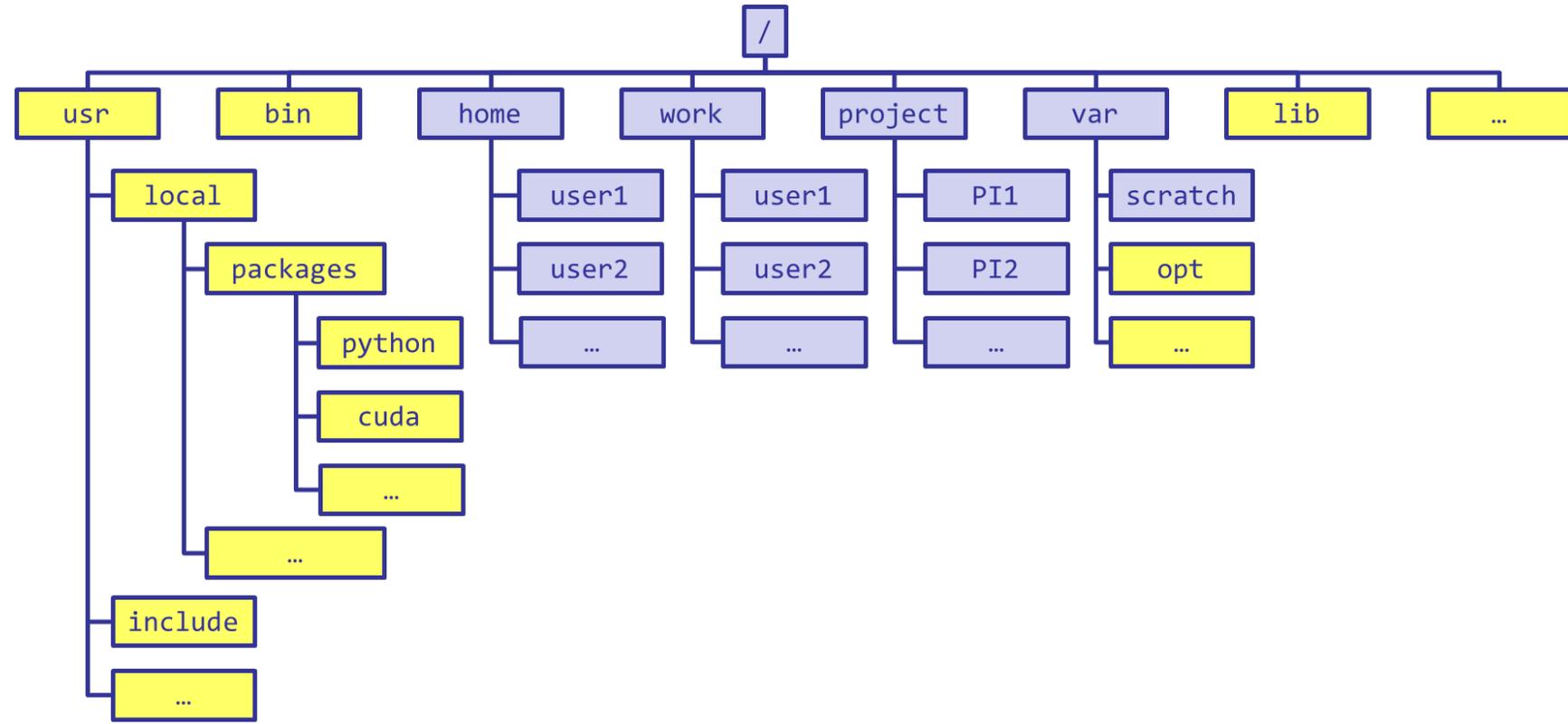
2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

- Facing difficulties installing / managing your software packages on HPC?
- Some typical scenarios...

a) Permission denied



a) Permission denied

```
[jasonli3@smic2 ~]$ module load python/3.6.2-anaconda-tensorflow
[jasonli3@smic2 ~]$ module li
Currently Loaded Modulefiles:
 1) python/3.6.2-anaconda-tensorflow
```


a) Permission denied

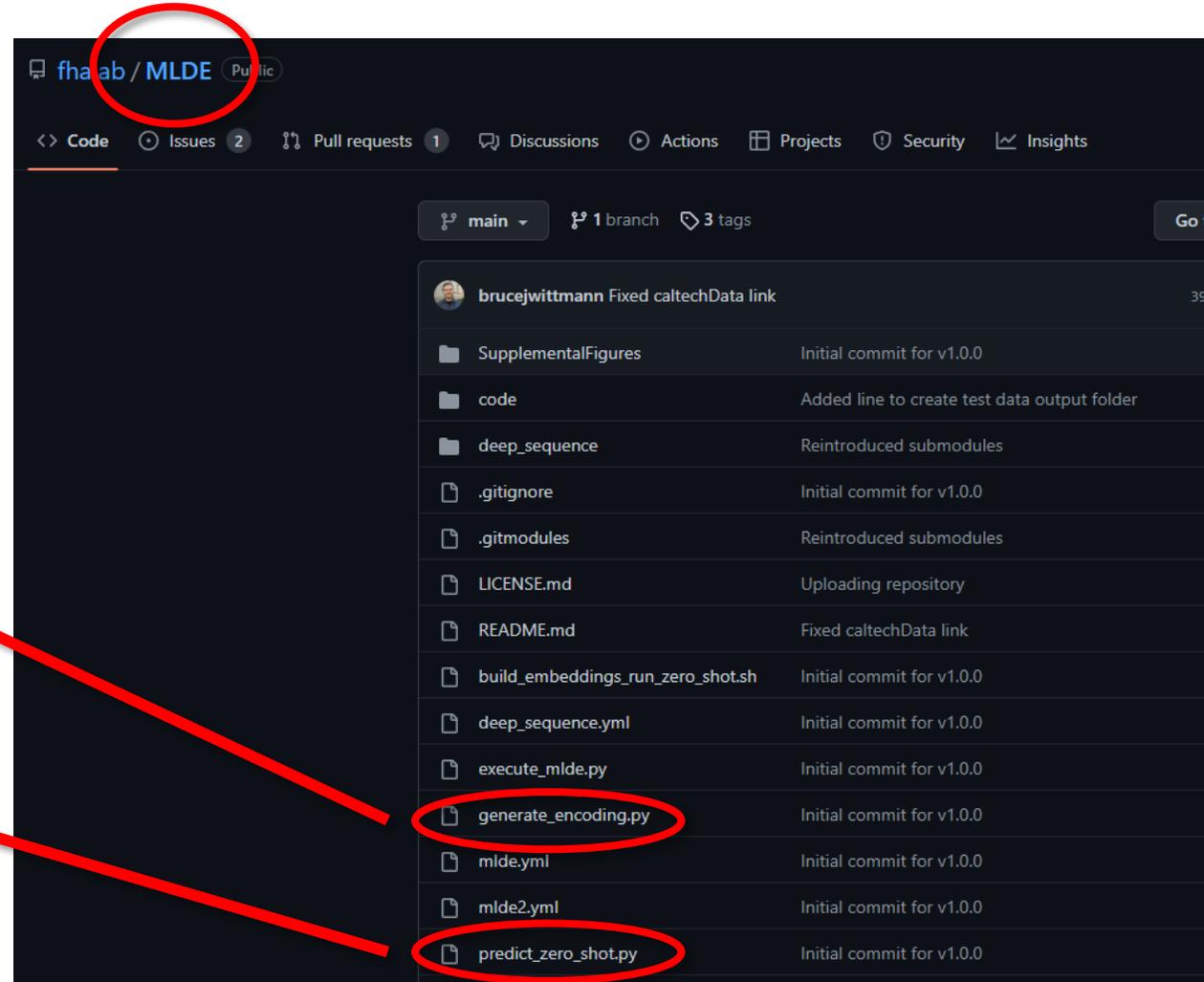
```
$ sudo yum install ...  
$ sudo apt-get install ...  
$ sudo make install
```

b) Conflicted packages

- What if I need two packages w/ conflicted dependencies?

Tensorflow 1.13

PyTorch > 1.5



c) Need a different version?

```
[jasonli3@smic2 ~]$ module av python/ r/ perl/
----- /usr/local/packages/Modules/default/modulefiles/linux-rhel7-ivybridge -----
python/2.7.7-anaconda    python/2.7.13-anaconda-tensorflow  python/3.8.5-anaconda-ood
python/2.7.7/GCC-4.9.0  python/3.6.2-anaconda-tensorflow
python/2.7.10-mkl-mic   python/3.8.5-anaconda

----- /usr/local/packages/Modules/default/modulefiles/linux-rhel7-ivybridge -----
r/4.0.3/intel-19.0.5   r/4.1.2/gcc-9.3.0

----- /usr/local/packages/Modules/default/modulefiles/linux-rhel7-ivybridge -----
perl/5.32.0/intel-19.0.5
```

d) Over complicated dependencies

Table Of Contents

- Installation
 - Packages
 - Docker
 - Source code
 - Download
 - Prerequisites
 - Compilation configuration
 - Compilation and

Prerequisites

Yade compilation and execution rely on a number of mandatory and optional external softwares; the

- cmake build system
- gcc compiler (g++); other compilers will not work; you need g++>=4.2 for openMP support
- boost 1.47 or later
- Qt library
- freeglut3
- libQGLViewer
- python, numpy, ipython, sphinx
- matplotlib
- eigen algebra library (minimal required version 3.2.1)
- gdb debugger
- sqlite3 database engine

e) Sharing / Migrating your environment

- Huge effort & large disk quota to install
 - What if my colleagues want to use?
 - What if I want to migrate a different cluster?

Any of those apply to you?

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips



Virtual Environment

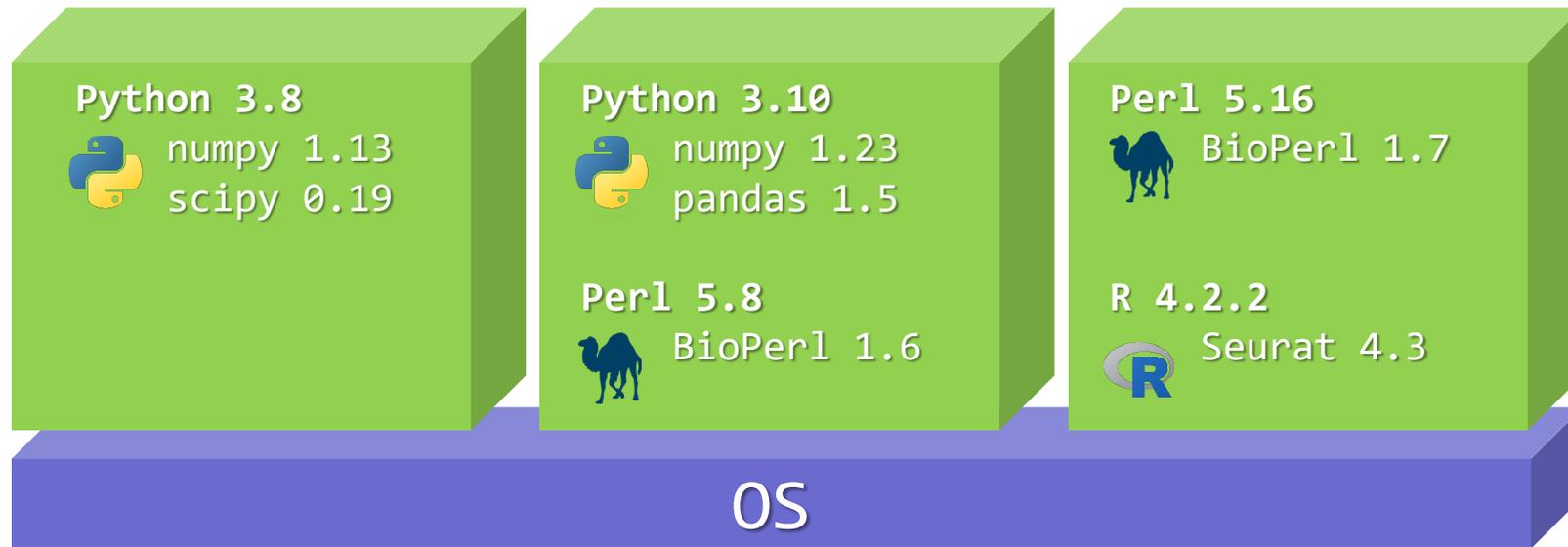
a) Conda



- A **software**
- Installs / updates / manages packages & dependencies
- Creates / loads / switches between virtual environments
- Initially for Python → General purposes
- Advantage: Does **NOT** need sudo permission!

b) Virtual Environment (VE)

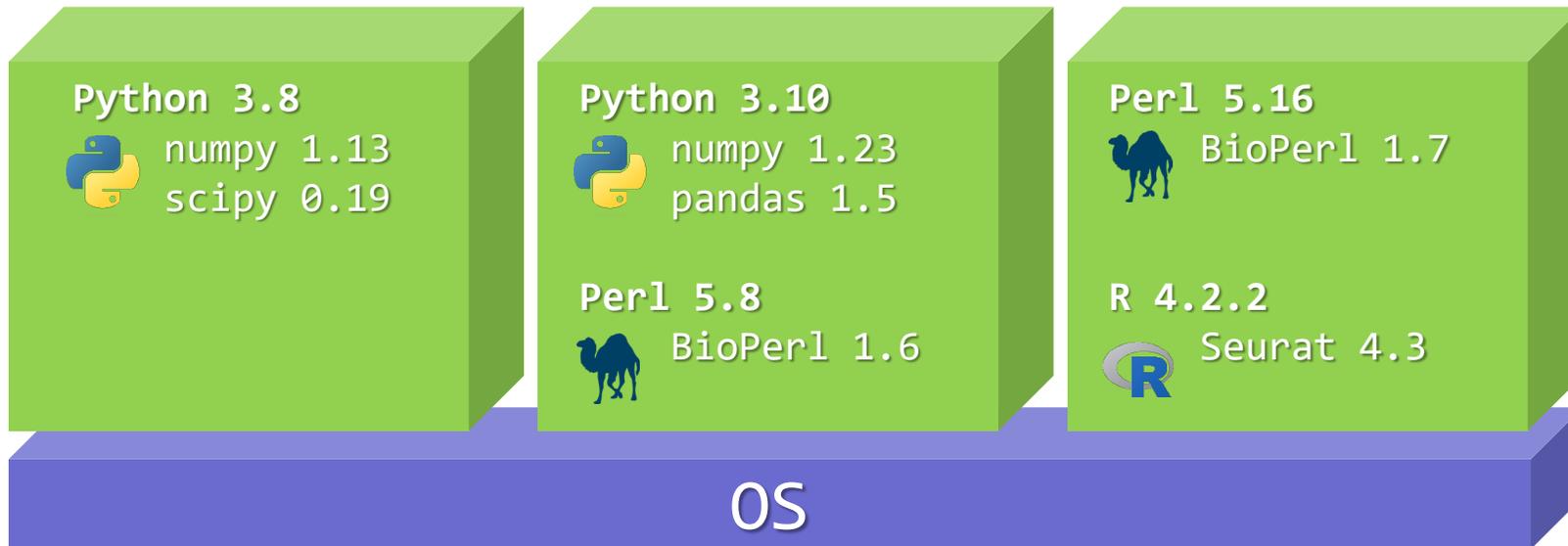
- An **environment**
- **Isolated** and **self-contained** to install and manage packages & dependencies



b) Virtual Environment (VE)

- An **environment**
- **Isolated** and **self-contained** to install and manage packages & dependencies

Whatever happens in a VE stays in that VE...

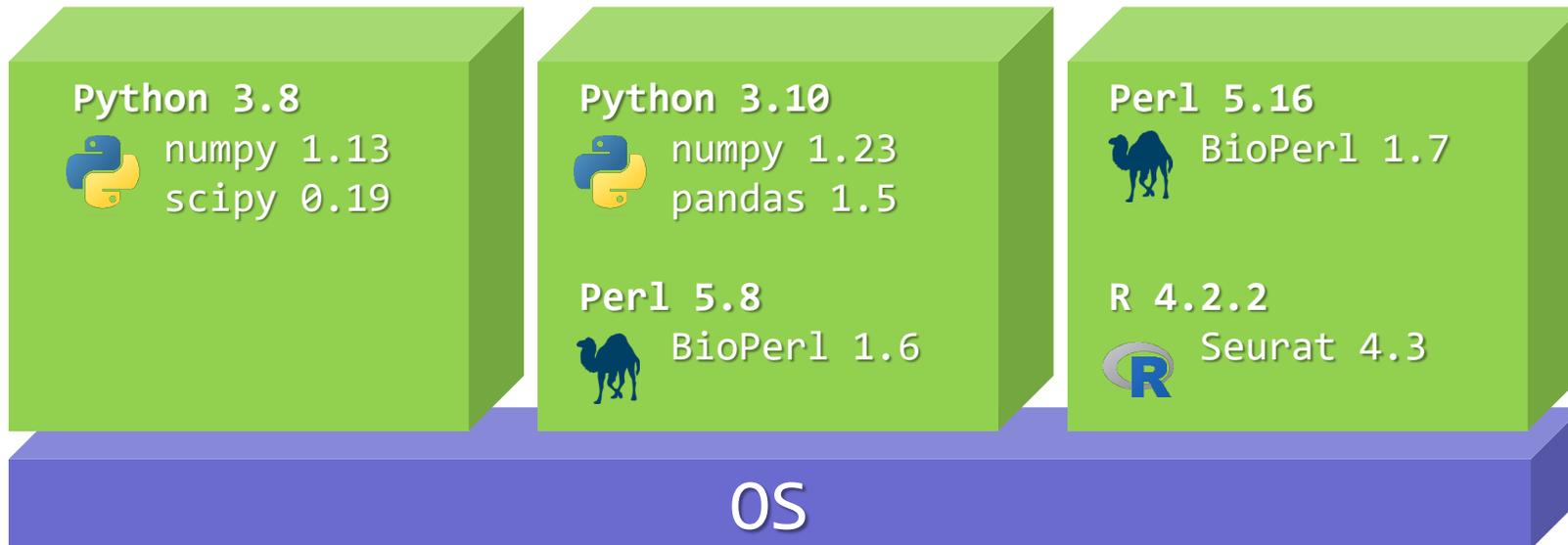


b) Virtual Environment (VE)

- An **environment**
- **Isolated** and **self-contained** to install and manage packages & dependencies



All dependencies are installed within the VE



- Relation

 CONDA

- is **a tool** to create / manage
- is **not the only** tool to create / manage
- **usually** works with

.....
Virtual Environment

- Rule of thumb:

**If a software package you need is managed by Conda,
you (most likely) can install / manage it by yourself without a problem**

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

- ~~• First install a conda in your user environment ... ?~~

a) Use conda that comes with system-wide python module

- No installation / disk quota required.
- Sufficient for most user cases.

```
$ module load python
```

Step 1: Can use Conda

```
$ conda init
```

**Step 2: Can use Conda without loading python module
(recommended)**

b) Install miniconda

- Latest version: https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh

```
$ wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
```

Step 1: Download miniconda

```
$ chmod u+x Miniconda3-latest-Linux-x86_64.sh
```

Step 2: Allow execution

```
$ ./Miniconda3-latest-Linux-x86_64.sh
```

Step 3: Run and follow prompts

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

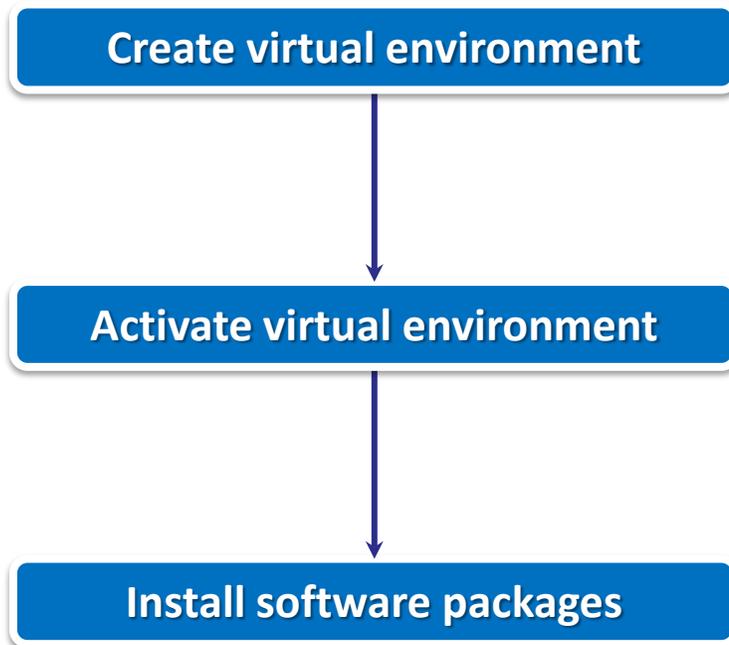
- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

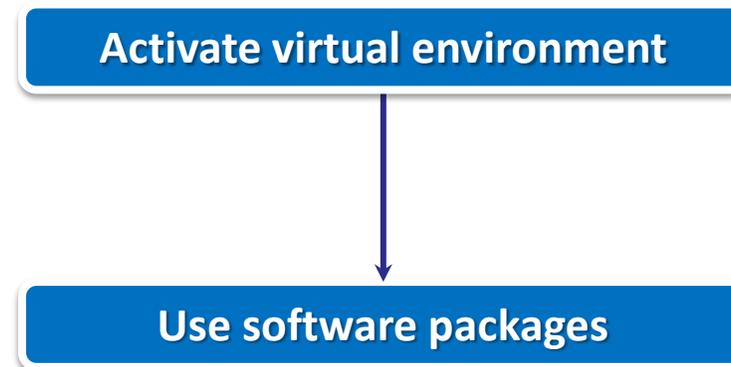
- **Key:**

Always use a virtual environment!

To install ...



To use ...



1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

3) Creating a virtual environment

- Most frequently used commands

Command	
conda create -n ENVIRONMENT	Cre

```
(base) [jasonli3@smic2 ~]$ conda create -n myenv
Collecting package metadata (current_repodata.json): done
Solving environment: done

=> WARNING: A newer version of conda exists. <=
current version: 4.12.0
latest version: 23.1.0

Please update conda by running
$ conda update -n base -c defaults conda

## Package Plan ##

environment location: /home/jasonli3/.conda/envs/myenv

Proceed ([y]/n)?

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#   $ conda activate myenv
#
# To deactivate an active environment, use
#
#   $ conda deactivate
```

[1] <https://docs.conda.io/projects/conda/en/latest/commands.html>



3) Creating a virtual environment

- Most frequently used commands

Command	Description
<code>conda create -n ENVIRONMENT</code>	Create a virtual environment
<code>source activate ENVIRONMENT</code>	Activate a virtual environment

```
(base) [jasonli3@smic2 ~]$ source activate myenv  
(myenv) [jasonli3@smic2 ~]$ █
```

3) Creating a virtual environment

- Most frequently used commands

Command	Description
<code>conda create -n ENVIRONMENT</code>	Create a virtual environment
<code>source activate ENVIRONMENT</code>	Activate a virtual environment
<code>conda deactivate</code>	Deactivate a virtual environment

```
(myenv) [jasonli3@smic2 ~]$ conda deactivate  
(base) [jasonli3@smic2 ~]$
```

3) Creating a virtual environment

- Most frequently used commands

Command	Description
<code>conda create -n ENVIRONMENT</code>	Create a virtual environment
<code>source activate ENVIRONMENT</code>	Activate a virtual environment
<code>conda deactivate</code>	Deactivate a virtual environment
<code>conda env list</code>	List all virtual environments

```
(base) [jasonli3@smic2 ~]$ conda env list
# conda environments:
#
myenv          /home/jasonli3/.conda/envs/myenv
base           * /usr/local/packages/python/3.8.5-anaconda
```

3) Creating a virtual environment

- Most frequently used commands

Command	Description
<code>conda create -n ENVIRONMENT</code>	Create a virtual environment
<code>source activate ENVIRONMENT</code>	Activate a virtual environment
<code>conda deactivate</code>	Deactivate a virtual environment
<code>conda env list</code>	List all virtual environments
<code>conda env remove -n ENVIRONMENT</code>	Remove a virtual environment and all installed packages

CAUTION! NO CONFIRMATION! IRREVOCABLE!

3) Creating a virtual environment

- Most frequently used commands

Command	Description
<code>conda create -n ENVIRONMENT</code>	Create a virtual environment
<code>source activate ENVIRONMENT</code>	Activate a virtual environment
<code>conda deactivate</code>	Deactivate a virtual environment
<code>conda env list</code>	List all virtual environments
<code>conda env remove -n ENVIRONMENT</code>	Remove a virtual environment and all installed packages

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Creating a virtual environment
- 3) Typical workflow
- 4) Installing software packages

3. Advanced Tips

4) Installing software packages

- Before installation...

Make sure a virtual environment is activated!

4) Installing software packages

- Most frequently used commands

Command	Description
<code>conda install NAME</code>	Install a software package

4) Installing software packages

- Most frequently used commands

Command	Description
<code>conda install NAME</code>	Install a software package
<code>conda install NAME=VERSION</code>	Install a specific version

4) Installing software packages

- Most frequently used commands

Command	Description
<code>conda install NAME</code>	Install a software package
<code>conda install NAME=VERSION</code>	Install a specific version
<code>conda install NAME -c CHANNEL</code>	Install from a specific channel (e.g., conda-forge, bioconda, nvidia, ...)

4) Installing software packages

- Most frequently used commands

Command	Description
<code>conda install NAME</code>	Install a software package
<code>conda install NAME=VERSION</code>	Install a specific version
<code>conda install NAME -c CHANNEL</code>	Install from a specific channel (e.g., conda-forge, bioconda, nvidia, ...)
<code>conda install NAME1 NAME2 ...</code>	Install multiple packages at once (let conda work out dependencies)

4) Installing software packages

- Most frequently used commands

Command	Description
<code>conda install NAME</code>	Install a software package
<code>conda install NAME=VERSION</code>	Install a specific version
<code>conda install NAME -c CHANNEL</code>	Install from a specific channel (e.g., conda-forge, bioconda, nvidia, ...)
<code>conda install NAME1 NAME2 ...</code>	Install multiple packages at once (let conda work out dependencies)
<code>conda list</code>	List all installed software package

4) Installing software packages

- Other useful commands

Command	Description
conda search NAME	Search available package versions

4) Installing software packages

- Other useful commands

Command	Description
<code>conda search NAME</code>	Search available package versions
<code>conda search NAME -c CHANNEL</code>	Search available package versions in a specific channel

4) Installing software packages

- Other useful commands

Command	Description
<code>conda search NAME</code>	Search available package versions
<code>conda search NAME -c CHANNEL</code>	Search available package versions in a specific channel
<code>conda search NAME --info</code>	Search available package versions with details

4) Installing software packages

- Other useful commands

Command	Description
<code>conda search NAME</code>	Search available package versions
<code>conda search NAME -c CHANNEL</code>	Search available package versions in a specific channel
<code>conda search NAME --info</code>	Search available package versions with details
<code>conda update/upgrade NAME</code>	Update a package to the latest available version

4) Installing software packages

- Other useful commands

Command	Description
<code>conda search NAME</code>	Search available package versions
<code>conda search NAME -c CHANNEL</code>	Search available package versions in a specific channel
<code>conda search NAME --info</code>	Search available package versions with details
<code>conda update/upgrade NAME</code>	Update a package to the latest available version
<code>conda uninstall/remove NAME</code>	Remove a package

4) Installing software packages

- **Bonus: Hot packages!**

- a) PyTorch (w/ GPU support)

```
$ # Start an interactive job (using qsub or srun)
```

Step 1: Start an interactive job (otherwise will timeout)

```
$ conda create -n torch
```

```
$ source activate torch
```

```
$ conda install -c pytorch -c nvidia pytorch==1.13.1 torchvision==0.14.1 torchaudio==0.13.1 pytorch-cuda=11.6
```

Step 2: Create a VE and install PyTorch & dependencies

4) Installing software packages

- **Bonus: Hot packages!**

- b) Tensorflow (w/ GPU support)

```
$ # Start an interactive job on *GPU* nodes (using qsub or srun)
```

Step 1: Start an interactive job on *GPU* nodes (otherwise will fail)

```
$ conda create -n tf
```

```
$ source activate tf
```

```
$ conda install -c conda-forge -c nvidia tensorflow-gpu=2.11 cudatoolkit=11.6 cuda-nvcc=11.6
```

Step 2: Create a VE and install tensorflow & dependencies

```
$ mkdir -p $CONDA_PREFIX/etc/conda/activate.d
```

```
$ echo 'export XLA_FLAGS=--xla_gpu_cuda_data_dir=$CONDA_PREFIX' >> $CONDA_PREFIX/etc/conda/activate.d/env_vars.sh
```

Step 3: Run these commands to set up environment

- Your workflow should mostly look like...

To install ...

```
$ conda create ...  
  
$ source activate ...  
  
$ conda install ...
```

To use ...

```
$ source activate ...  
  
$ # Do whatever you need  
to do with the packages
```

- Create a virtual environment
- Search for **SciPy** version and install the second-latest version (as well as dependencies)
- After you are done, type in chat the installed **SciPy** and **Python** version

1. Why Conda?

- 1) Scenarios
- 2) Concepts

2. Basic Usage

- 1) Get Conda
- 2) Typical workflow
- 3) Creating a virtual environment
- 4) Installing software packages

3. Advanced Tips

A little more than the basics...

1) Change Conda path

- **Default Conda path:**

- Environments: `~/.conda/envs/`
- Downloaded packages: `~/.conda/pkgs/`

```
[jasonli3@smic1 ~]$ balance
User filesystem quotas for jasonli3 (uid 15827):
  Filesystem      MB used   MB quota
  /home           950      10000
  /work /project  329639   0        6
Storage allocation  MB used   MB quota
```

1) Change Conda path

- **Solution:**

Step 1: Ask your PI to apply for a storage allocation (/project) and add you to it

Step 2: Set up Conda to create / find virtual environments in /project

Step 3: Create your virtual environment and install software package

1) Change Conda path

- **Solution:**

~~Step 1: Ask your PI to apply for a storage allocation (/project) and add you to it~~

Step 2: Set up Conda to create / find virtual environments in /project

~~Step 3: Create your virtual environment and install software package~~

1) Change Conda path

a) Method 1: Command lines

```
$ conda config --add envs_dirs /path/to/envs
```

Add path to environments

```
$ conda config --add pkgs_dirs /path/to/pkgs
```

Add path to downloaded packages

1) Change Conda path

b) Method 2: Configuration file

- Use any text editor to open: `~/.condarc`

```
$ vi ~/.condarc
```

```
envs_dirs:  
  - /project/jasonli3/.conda/envs  
pkgs_dirs:  
  - /project/jasonli3/.conda/pkgs
```

```
~
```

2) Share virtual environment

- **Scenario:**

- I made a huge effort to install an extensive collection of software packages for our group's research needs. I don't want to do it all over again for everyone in our group. Is it possible to just share the virtual environment with them?

2) Share virtual environment

- **Solution:**

Step 1: Ask your PI to apply for a storage allocation (/project) and add **you** and **your colleagues** to it

Step 2: Set up Conda to create / find virtual environments in /project

Step 3: Create your virtual environment and install software packages

Step 4: Ask your colleague to repeat [Step 2]

3) Migrate / clone virtual environment

- **Scenario:**
 - I have been using LSU HPC cluster. But now I want to switch to LONI and run the exactly same software. How can I do that?
 - I am leaving my current position. But I may continue doing similar research. How can I replicate my environment to a different HPC system in a different institute?

3) Migrate / clone virtual environment

- **Solution**

To ...

Run command

- **Solution**

To ...

Export virtual environment recipe to file

```
name: spyder
channels:
  - defaults
dependencies:
  - _libgcc_mutex=0.1=main
  - _openmp_mutex=5.1=1_gnu
  - arrow=1.2.3=py310h06a4308_1
  - astroid=2.14.2=py310h06a4308_0
  - attrs=22.1.0=py310h06a4308_0
  - babel=2.11.0=py310h06a4308_0
  - beautifulsoup4=4.11.1=py310h06a4308_0
  - black=22.6.0=py310h06a4308_0
  - blas=1.0=mkl
  - bottleneck=1.3.5=py310ha9d4c09_0
  - brotli=1.0.9=h5eee18b_7
```

3) Migrate / clone virtual environment

- **Solution**

To ...	Run command
Export virtual environment recipe to file	<code>conda env export > myenv.yml</code>
Create a virtual environment from file	<code>conda env create -f myenv.yml</code>

4) Use virtual environment on Open OnDemand

LSU HPC (SMIC)



OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

Interactive Apps

 Cellranger(beta) System Installed App	 Jupyter Notebook/Lab System Installed App	 RStudio Server System Installed App
--	--	--

Message of the Day

Welcome to the LSU HPC OnDemand portal!

With the OnDemand web portal, you can:

LONI (QB2)



OnDemand provides an integrated, single access point for all of your HPC resources.

Message of the Day

Welcome to the LONI HPC OnDemand portal!

With the OnDemand web portal, you can:

- Manage, download and upload files to the HPC systems (click links in the "Files" on the top of this page)
- Check allocation balances
- Check disk usage and quotas
- Check job status
- Submit jobs using templates
- Access HPC systems via a terminal
- Run interactive apps such as Jupyter Notebook/Lab and Rstudio (click links in the "Interactive Apps" on the top of this page)

Getting started

[1] https://youtu.be/xk5q8p6QQ_k



4) Use virtual environment on Open OnDemand

The screenshot shows the Open OnDemand interface with a Jupyter Notebook open. The notebook contains two code cells:

```
[1]: import sys
print("Hello world!")

[2]: !conda list
```

The output of the second cell shows a list of installed packages in the environment:

```
# packages in environment at /usr/local/packages/python/3.8.5-anaconda:
#
# Name                               Version      Build      Channel
_ipyw_jlab_nb_ext_conf               0.1.0        py38_0
_libgcc_mutex                         0.1          main
alabaster                             0.7.12       py_0
anaconda                              2020.11      py38_0
anaconda-client                       1.7.2        py38_0
anaconda-navigator                    1.10.0       py38_0
anaconda-project                      0.8.4        py_0
argh                                   0.26.2       py38_0
argon2-cffi                           20.1.0       py38h7b6447c_1
asn1crypto                            1.4.0        py_0
astroid                                2.4.2        py38_0
astropy                                4.0.2        py38h7b6447c_0
async_generator                        1.10         py_0
atomicwrites                           1.4.0        nv_0
```

The interface also shows a file browser on the left with files like 'ondemand', 'toWorkFolder', 'setenv.sh', 'testTF.ipynb', and 'Untitled.ipynb'. The 'Untitled.ipynb' file is selected. The status bar at the bottom indicates 'Python 3 | Idle', 'Saving completed', 'Mode: Command', and 'Ln 2, Col 22'.

[1] https://youtu.be/xk5q8p6QQ_k



4) Use virtual environment on Open OnDemand

- **How to:**

Step 1: ssh to the cluster you want to use

LSU HPC	LONI
SMIC	QB2

Step 2: Activate the virtual environment you want to use in Jupyter

```
$ source activate ENVIRONMENT
```

Step 3: Install ipykernel

```
$ conda install ipykernel
```

Step 4: Start a Jupyter session in Open OnDemand, and choose the environment in **kernel**

- **Scenario**

- I need software packages other than Python (R / Perl / Lua / ...)
- I need a different version than the system modules
- I am using the system's R module, but having trouble installing some packages (e.g., rgdal)

```
> install.packages("rgdal")
Warning in install.packages("rgdal") :
  'lib = "/home/packages/r/4.1.2/5k5jengl/rlib/R/library"' is not writable
  would you like to use a personal library instead? (yes/no) [yes]
no
configure: error: gdal-config not found or not executable.
ERROR: configuration failed for package 'rgdal'
* removing '/home/jasonli3/R/x86_64-pc-linux-gnu-library/4.1/rgdal'

The downloaded source packages are in
  '/tmp/Rtmpd2csho/downloaded_packages'
Warning message:
In install.packages("rgdal") :
  installation of package 'rgdal' had non-zero exit status
```

- Solutions

Many non-python packages are managed by Conda too!

To install ...		Run command ...
Languages	R	<code>conda install R</code>
	Perl	<code>conda install perl</code>
	Julia	<code>conda install julia -c conda-forge</code>

- Solutions

Many non-python packages are managed by Conda too!

To install ...		Run command ...
Languages	R	<code>conda install R</code>
	Perl	<code>conda install perl</code>
	Julia	<code>conda install julia -c conda-forge</code>
Dependencies	hdf5	<code>conda install hdf5</code>
	netcdf	<code>conda install libnetcdf -c conda-forge</code>
	FFTW	<code>conda install fftw</code>

- Solutions

Many non-python packages are managed by Conda too!

To install ...		Run command ...
Languages	R	<code>conda install R</code>
	Perl	<code>conda install perl</code>
	Julia	<code>conda install julia -c conda-forge</code>
Dependencies	hdf5	<code>conda install hdf5</code>
	netcdf	<code>conda install libnetcdf -c conda-forge</code>
	FFTW	<code>conda install fftw</code>
		...

- One more cool thing...
 - You can use language specific package management tools

Language	Tool
Python	pip
R	install.packages
Perl	cpan
Julia	Pkg

- Packages will be **isolated** in the virtual environment

5) More than Python

- E.g., Use Conda to solve your R issue

– Use system's R module:

```
$ module load r  
$ R  
> install.packages("rgdal")
```

→ Will fail!

– Use Conda:

```
$ conda create -n rgdal  
$ source activate rgdal  
$ conda install r-rgdal
```

→ Will succeed!

- Scenario

- I need to compile a code from source, but the dependencies are too convoluted



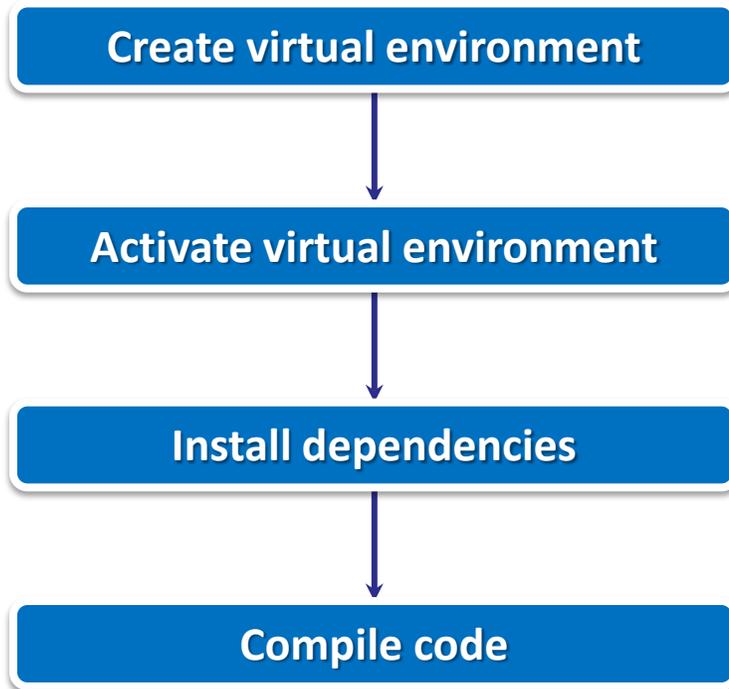
Prerequisites

Yade compilation and execution rely on a number of mandatory and optional external softwares; the

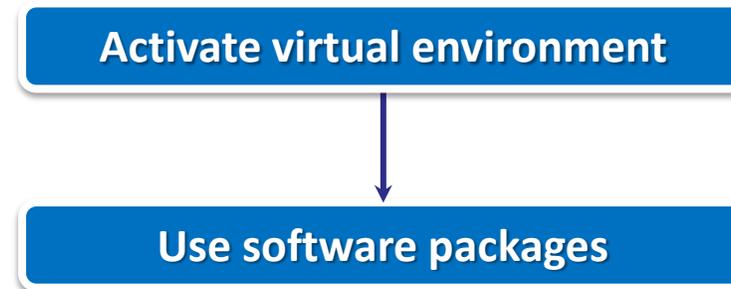
- cmake build system
- gcc compiler (`g++`); other compilers will not work; you need `g++>=4.2` for openMP support
- boost 1.47 or later
- Qt library
- freeglut3
- libQGLViewer
- python, numpy, ipython, sphinx
- matplotlib
- eigen algebra library (minimal required version 3.2.1)
- gdb debugger
- sqlite3 database engine

- **Solution**

To install ...



To use ...



a) Conflict with system module

```
(spyder) [jasonli3@smic2 ~]$ module li
Currently Loaded Modulefiles:
1) python/3.8.5-anaconda
(spyder) [jasonli3@smic2 ~]$ python
Python 3.10.9 (main, Mar  8 2023, 10:47:38) [GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
/usr/local/packages/python/3.8.5-anaconda/lib/python3.8/site-packages/numpy/__init__.p
fore Intel(R) MKL initialization ensuring correct out-of-the box operation under c
ess is not assured. Please install mkl-sec see http://github.com/IntelPyt
from . import _distributor_init
Traceback (most recent call last):
  File "/usr/local/packages/python/3.8.5-anaconda/lib/python3.8/site-packages/numpy/co
from . import multiarray
File "/usr/local/packages/python/3.8.5-anaconda/lib/python3.8/site-packages/numpy/co
```

Fail!

a) Conflict with system module

- Rule of thumb:

Do **NOT** load system module if you are using your own installation!

b) What if I made a mess?

- I mixed conda / pip back and forth, and broke the environment...
- It may be easier to create a new virtual environment and start fresh...

To install ...

```
$ conda create ...  
  
$ source activate ...  
  
$ conda install ...
```

To use ...

```
$ source activate ...  
  
$ # Do whatever you need  
to do with the packages
```

- **Contact user services**

- Email Help Ticket: sys-help@loni.org
- Telephone Help Desk: +1 (225) 578-0900