An introduction to the conda ecosystem

Jonathan Helmus, Senior Software Engineer

June 4, 2020

CONDA®

- Conda is a cross platform package and environment management system
- Written and maintained by Anaconda, Inc
- Open Source, BSD licensed
- Installs reusable, relocatable packages for software written in any language
- Many Python and R data science, machine learning and AI frameworks
- Does not require administrator privileges
- Available by installing the **Anaconda Individual Edition** or **Miniconda**

package management with conda

- Packages are binaries, no compiler or libraries are needed
- Does not require administrator privileges
- Uses a SAT solver for dependency resolution

Common commands:

- conda list : list the installed packages
- conda install : install one or more package(s)
- conda remove : remove a package
- conda update : update a package

conda environments

Conda can create isolated environments that have their own set of packages.

- conda create : create a new conda environment
- conda activate : activate a conda environment
- conda deactivate : deactivate the current conda environment

Great when you need to work with different versions of a library or application.

Environment specification can be exported to a file and recreated.

conda channels

- Conda packages are provided from different repositories, called channels.
- Out of the box, conda installs packages from the “defaults” channel.
- Other channels can be enabled to access additional collections of packages

Some key channels are:

- defaults : packages from Anaconda, Inc
- conda-forge : large community led collection of packages
- bioconda : community specializing in bioinformatics packages
- rapidsai : GPU accelerated data science libraries

CONDA-FORGE

RAPIDS

BIOCONDA®

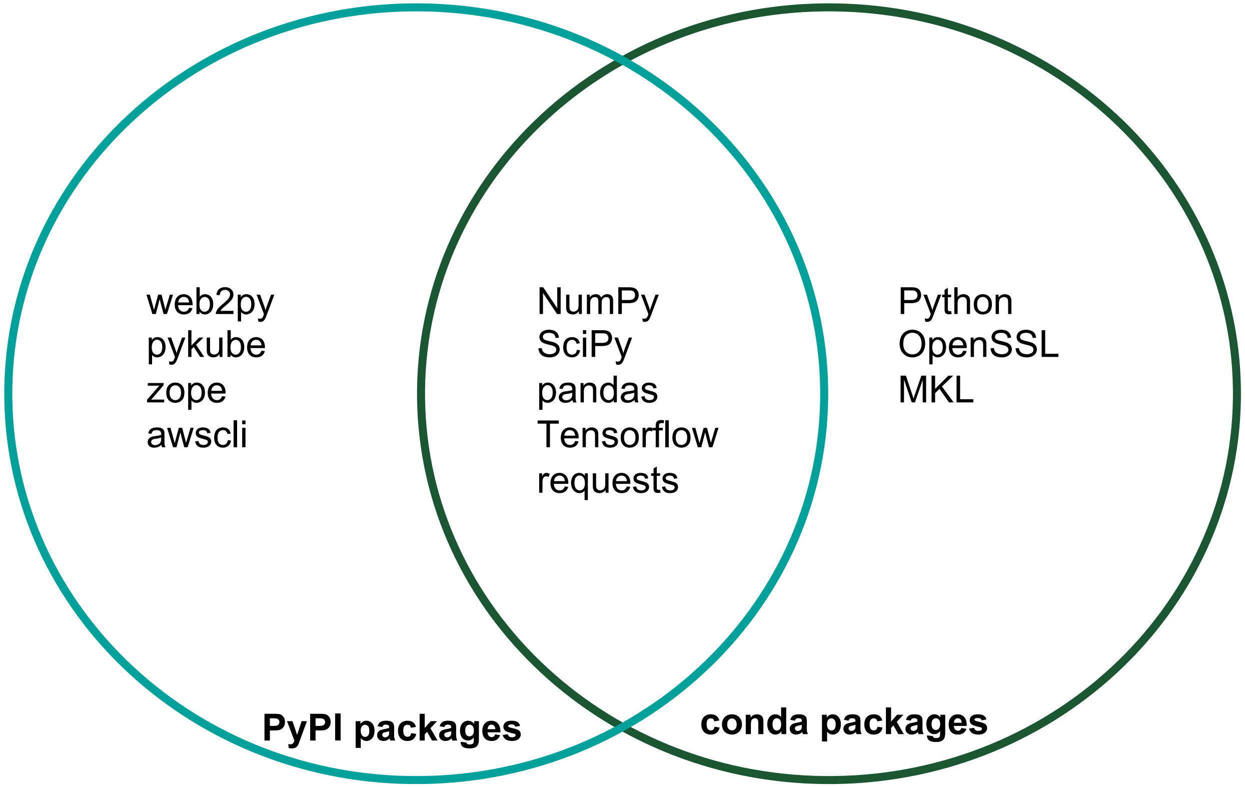
conda vs. pip

pip

- Installs Python packages
- Binaries (wheels) and source distributions
- virtualenv or venv needed to create isolated environment.
- Resolves dependencies recursively

conda

- Installs software written in any language
- Packages are binaries
- Built in support for creating isolated environments
- Uses a SAT solver for dependency resolution



web2py
pykube
zope
awscli

NumPy
SciPy
pandas
Tensorflow
requests

Python
OpenSSL
MKL

PyPI packages

conda packages

conda vs. system package managers

system package manager

- Installs software written in any language
- Packages are binaries
- Various algorithms for resolving dependencies
- Typically no isolated environments
- Often requires admin permissions
- Unique to a single platform

conda

- Installs software written in any language
- Packages are binaries
- Uses a SAT solver for dependency resolution
- Support for isolated environments
- No administrator privileges needed
- Cross platform

conda : packages

A conda package is a compressed archive containing system libraries, Python modules, executable programs or other components.

Packages contain binaries, no compilation is needed at install or run time.

Metadata is included with the package in the “info” directory.

conda build is the tool for creating conda packages:

conda build

conda build is a tool to create reusable, relocatable packages for software written in any language.

Packages are built from **recipes** which specify package metadata and build steps.

Process

- create an isolated build environment
- execute build steps
- bundle files
- create a test environment
- execute tests

Package storage

Anaconda Cloud (anaconda.org) offers hosting of packages.

- accounts are free to create
- packages are publically available to anyone

Anaconda Team Edition

- Enterprise-grade package manager
- Mirror publically available packages
- Host private packages
- Curate packages based upon license, version or security concerns

Anaconda Individual Edition

The Anaconda Individual Edition is a curated collection of data science libraries.

- ~300 packages are installed automatically
- Rigorously tested for compatibility
- Additional packages can be install using conda
- Released quarterly
- Latest release is 2020.02 in March 2020.

conda env

conda env is a sub-command included with conda that can be used to record and create conda environments

- To record the packages which make up an environment:

```
conda env export > environment.yml
```

- To create an environment from a stored environment file:

```
conda env create -f environment.yml
```

conda list

conda list can also be used record and re-create environments in a text form

- To record an environment:

```
conda list --explicit > spec-list.txt
```

- To recreate the environment:

```
conda create --name python-course --file spec-list.txt
```

constructor

constructor is a tool for creating installers from conda packages

Input is a YAML file with:

- name
- version
- channels
- list of packages

Creates a stand alone installer, output depends on target platform

- MacOS: pkg or shell
- Windows : exe
- Linux : shell

conda pack

conda pack is a tool for creating relocatable conda environments.

- The deployment artifact in a compressed archive (tar.gz or zip)
- To create a packed environment:

```
conda pack -n example_env
```

- To unpack the environment:

```
tar -xzf example_env.tar.gz -C example_env
source example_env/bin/activate
conda-unpack
```

Thank You!

Jonathan Helmus

ANACONDA.