

Python environments

Henrik Andersson

2024-08-29



Hey there... How's it going? Yeah, I heard you recently took that Python course. That's, uh, great. Real initiative there. So, since you're all up to speed on that, I'm gonna need you to go ahead and handle a little project for me.

We've got some rainfall data that needs a bit of work, and I'm thinking you're just the person for the job. You see, we need to convert the data from its current format into something, you know, a little more useful. Since you've got those fresh Python skills, this should be right up your alley.

And while you're at it, I'm gonna need you to, uh, remove any outliers. They're really throwing off our averages, and, well, we just can't have that. Then, if you could go ahead and aggregate the data—maybe sum it up or average it out—whatever makes the most sense, that'd be terrific.

Oh, and don't forget about resampling. We're gonna need that data broken down into, like, daily, weekly, and monthly intervals. And, if you could just make sure the column names are all nice and clear afterward, that'd be great.

So, uh, yeah, I'm gonna need this done ASAP. Thanks a bunch. And keep up the good work with that Python stuff, mmmkay?







Hey, how's it going? Yeah, remember that rainfall data project you handled about six months ago? You did a great job on that, by the way. Really top-notch stuff.

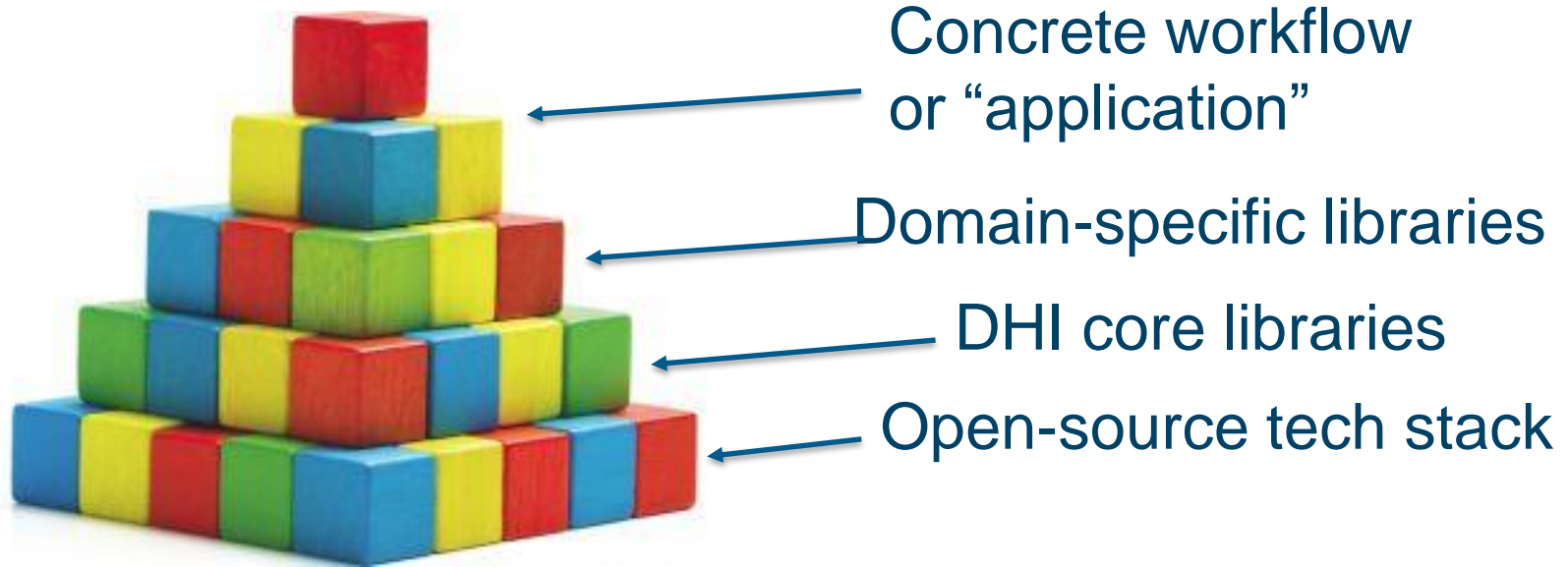
Well, it looks like the client's back with another one of those. Yep, same kind of deal—data conversion, outlier removal, aggregation, resampling, and renaming. You've done this before, so it should be pretty straightforward for you.

I'm thinking you can just, you know, follow the same process as last time. Maybe even streamline it a bit, since you're already familiar with the workflow. And, uh, if you could get this wrapped up sooner rather than later, that'd be greaaaaat.

So, yeah, I'm gonna need you to make this your priority. Thanks again, and keep up the good work. We really appreciate it.

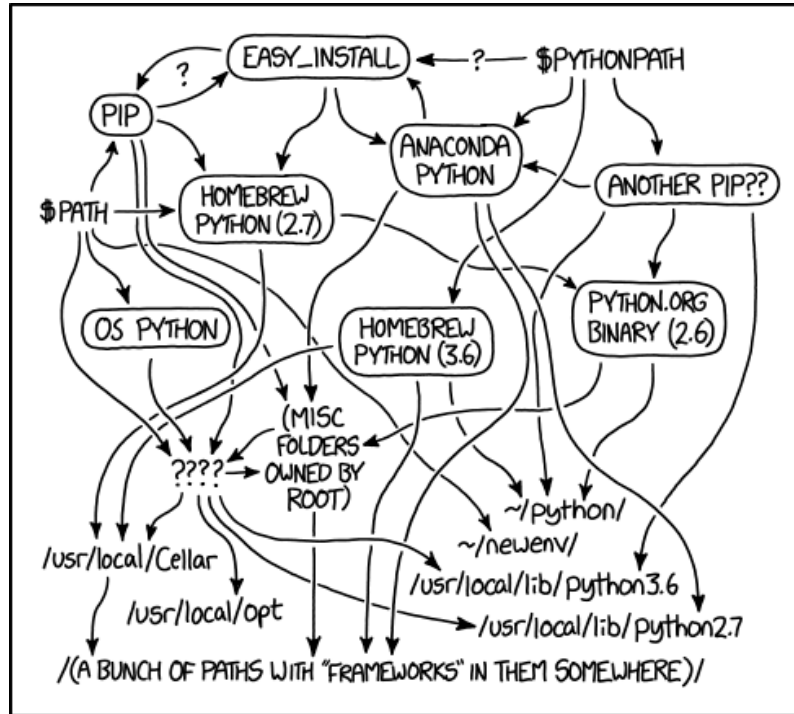


Dependencies



Reproducibility

- Collaboration with colleagues
- Moving computation from your laptop
- Returning to an old project



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

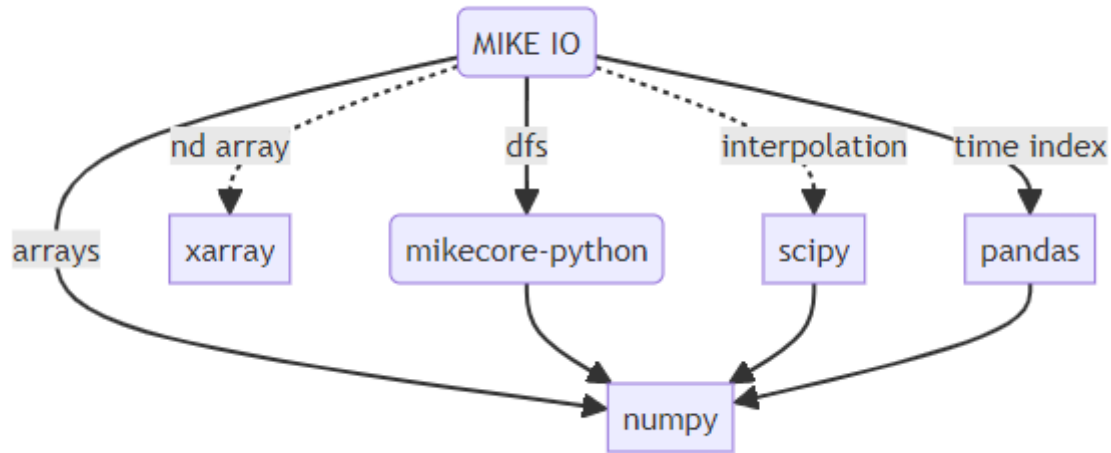
<https://xkcd.com/1987/>

Python package repositories

- Python Packaging Index (PyPI)
- Conda
 - Anaconda (default)
 - Conda-forge

① Use of Anaconda's Offerings at an organization of more than 200 employees requires a Business or Enterprise license. For more information, see our [Terms of Service](#).

Your dependencies also have dependencies



What happens when I type: `pip install pandas`?

- In a fresh environment:
 - installs latest version of pandas
 - install latest version of pandas dependencies (numpy, ...)
- In an environment where pandas is installed:
 - Nothing

What happens when I type: `pip install --update pandas`?

- In a fresh environment:
 - installs latest version of pandas
 - installs latest versions of pandas dependencies (numpy, ...)
- In an environment where pandas is installed:
 - installs latest version of pandas
 - **Nothing** happens to the indirect dependencies 🤖

```
pip install -r requirements.txt
```

```
≡ requirements.txt X  
1  jupyter-book  
2  matplotlib  
3  numpy  
4  pandas  
5  pyarrow  
6  mikeio>=2.0  
7  netcdf4  
8  xarray  
9
```


pip freeze

```
> pip freeze
accessible-pygments=0.0.4
alabaster=0.7.13
asttokens=2.4.1
attrs=23.1.0
Babel=2.13.1
beautifulsoup4=4.12.2
certifi=2023.7.22
cftime=1.6.3
charset-normalizer=3.3.2
click=8.1.7
comm=0.1.4
contourpy=1.2.0
cyclor=0.12.1
debugpy=1.8.0
decorator=5.1.1
docutils=0.18.1
exceptiongroup=1.1.3
executing=2.0.1
fastjsonschema=2.18.1
fonttools=4.44.0
greenlet=3.0.1
idna=3.4
imagesize=1.4.1
importlib-metadata=6.8.0
ipykernel=6.26.0
```

Lists installed packages along with version

Virtual environments

- Isolated installation of python packages
- Located in the project folder
- Part of the python standard library
- ``python -m venv .venv``
- Remember to exclude from source control!

```
C:\Users\jan\projects>mkdir waterworld

C:\Users\jan\projects>cd waterworld

C:\Users\jan\projects\waterworld>python -m venv .venv


C:\Users\jan\projects\waterworld>dir
Volume in drive C is Windows
Volume Serial Number is 4E59-97F8

Directory of C:\Users\jan\projects\waterworld

2024-08-27 12:41    <DIR>          .
2024-08-27 12:41    <DIR>          ..
2024-08-27 12:41    <DIR>          .venv
               0 File(s)                0 bytes
               3 Dir(s)  643,875,381,248 bytes free

C:\Users\jan\projects\waterworld>.venv\Scripts\activate

(.venv) C:\Users\jan\projects\waterworld>python -m pip install pandas
Collecting pandas
  Using cached pandas-2.2.2-cp310-cp310-win_amd64.whl (11.6 MB)
Collecting python-dateutil>=2.8.2
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
  229.9/229.9 kB 1.8 MB/s eta 0:00:00
```





uv

A faster alternative to pip

```
C:\Users\jan\projects\riverdance>uv venv
Using Python 3.8.19
Creating virtualenv at: .venv
Activate with: .venv\Scripts\activate

C:\Users\jan\projects\riverdance>.venv\Scripts\activate

(riverdance) C:\Users\jan\projects\riverdance>uv pip install pandas
Resolved 6 packages in 553ms
Prepared 6 packages in 6.07s
Installed 6 packages in 610ms
+ numpy==1.24.4
+ pandas==2.0.3
+ python-dateutil==2.9.0.post0
+ pytz==2024.1
+ six==1.16.0
```

<https://docs.astral.sh/uv/pip/>

Uv can install python ✨

```
C:\Users\jan\projects>mkdir danube

C:\Users\jan\projects>cd danube

C:\Users\jan\projects\danube>uv venv --python 3.12
Using Python 3.12.5
Creating virtualenv at: .venv
Activate with: .venv\Scripts\activate

C:\Users\jan\projects\danube>.venv\Scripts\activate

(danube) C:\Users\jan\projects\danube>python
Python 3.12.5 (main, Aug 14 2024, 04:23:18) [MSC v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> |
```


uv – one tool to rule them all 🤔

Highlights

- 🚀 A single tool to replace `pip`, `pip-tools`, `pipx`, `poetry`, `pyenv`, `virtualenv`, and more.
- ⚡ 10-100x faster than `pip`.
- 🐍 Installs and manages Python versions.
- 🛠️ Runs and installs Python applications.
- ➕ Runs scripts, with support for inline dependency metadata.
- 📁 Provides comprehensive project management, with a universal lockfile.
- 🔄 Includes a pip-compatible interface for a performance boost with a familiar CLI.
- 📖 Supports Cargo-style workspaces for scalable projects.
- 💾 Disk-space efficient, with a global cache for dependency deduplication.
- ⬇️ Installable without Rust or Python via `curl` or `pip`.
- 💻 Supports macOS, Linux, and Windows.

uv is backed by Astral, the creators of Ruff.

Conda/mamba

- Different source of packages
- Mamba = faster version of conda
- Named global environments
- Good for packages with complex dependencies
 - GIS
 - Deep learning GPU
- MIKE IO is no longer published to conda-forge

```
C:\Users\jan\projects\nemo>mamba create -n fish

mamba (1.4.2) supported by @QuantStack

GitHub: https://github.com/mamba-org/mamba
Twitter: https://twitter.com/QuantStack

Looking for: []

Preparing transaction: done
Verifying transaction: done
Executing transaction: done

To activate this environment, use

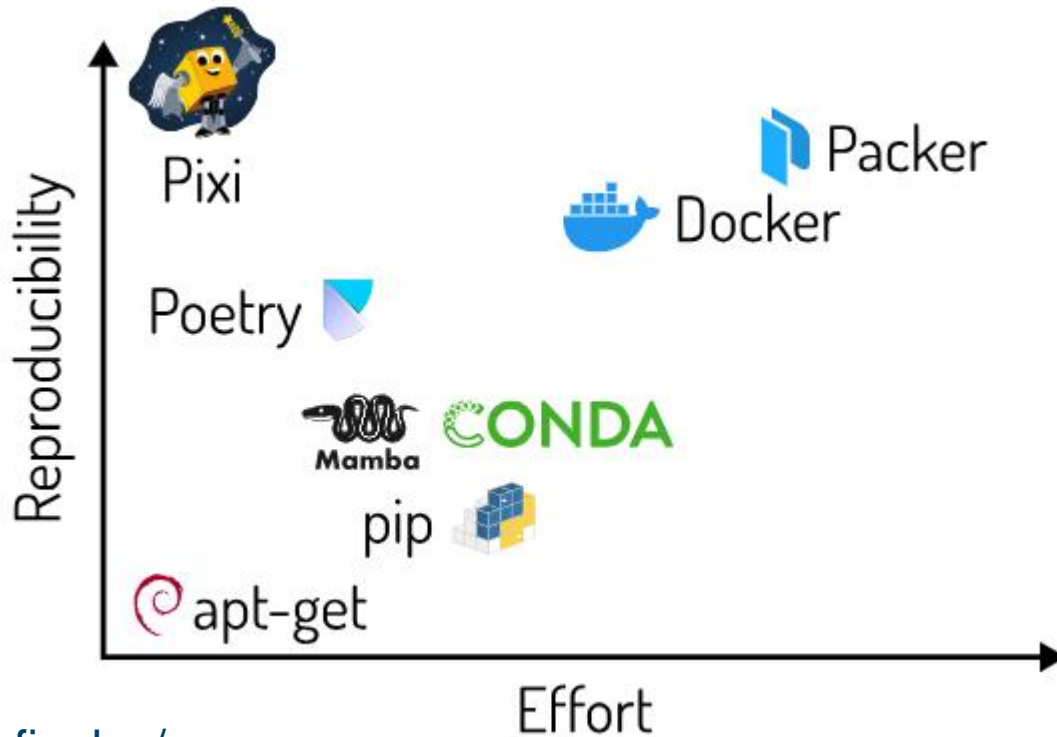
    $ mamba activate fish

To deactivate an active environment, use

    $ mamba deactivate

C:\Users\jan\projects\nemo>mamba activate fish
(fish) C:\Users\jan\projects\nemo>mamba install pandas
```

Pixi – next generation of conda/mamba



<https://prefix.dev/>

Why pixi over conda?

- Complete reproducibility of complex scientific software
 1. Faster than conda
 2. Better integration with PyPI (pip)
 3. No base environment! (1 project == 1 environment)
 4. Defined tasks
 5. Lockfile

Conclusions

- Proper dependency management is not trivial
- Don't rely on installing packages in a single environment
- Pip + venv is the part of the standard library
- **Uv is a modern implementation of pip (and more)**
 - Fast, can install python, extra tools like ``uv pip compile``
- Conda/mamba is a separate universe
- Pixi is a new and promising tool if you need conda packages