11falsenonelisttrue

Connecting to get command line access: `ssh washukey@compute1-client-1.ris.wustl.edu`

Queue to use: `workshop, workshop-interactive`

Group to use: `compute-workshop` (if part of multiple groups)

Video

https://www.youtube.com/watch?v=w1B9mPS8ZsA

pip (and a Linux package manger) vs anaconda

`pip` and `conda` are the two most popular ways to install python packages. There may be instances where you can only find directions on how to install an application with one tool or the either. In that case, the decision has been made for you. In situations where you can find both, using pip and and a your Linux package manager (e.g., `apt-get` in Debian and Ubuntu) can yield quicker build times and smaller final container sizes. However, because pip can only install Python packages, you may find yourself also having to use your package manager (i.e., `apt-get install -y dependency`) to install non Python dependencies. Conda has the advantage of including non-python dependencies. It's disadvantages are the slower build times and significantly larger final containers.

pip install documentation: https://pip.pypa.io/en/stable/reference/pip_install/

conda install documentation: https://docs.conda.io/projects/conda/en/latest/commands/install.html

Case Study: Spacy

spaCy is a free, open-source library for advanced Natural Language Processing (NLP) in Python. Here's some sample code we will call `script.py` from Spacy that iterates over every token in the string `Hello World!`. `script.py` will be created in the same directory as the Dockerfile.

Pip Example: Installing Spacy with pip

Build the image with the image name of spacy and a tag of pip so you'll be able to compare the image sizes on Docker hub

Conda Example: Installing Spacy with conda

Build, tag, and push the image with a name of spacy and a tag of conda;

Compare Image Sizes

Check your Docker hub profile and click on your spacy image. You will see a conda and pip tag under tags. pip is much smaller and you likely noticed it was faster to build.

Try Running Both Versions on Compute1

Which one runs faster and uses less bandwidth?