INSTALLING JUPYTER FOR BIG DATA AND ANALYTICS

Jupyter and Zeppelin both provide an interactive Python, Scala, Spark, etc. interpreter. Plus they do what the command line cannot, which is support graphical output with graphing packages like matplotlib.

While I personally prefer Zeppelin, it seems more data scientists and big data engineers are using Jupyter (aka iPython). For example most of the interactive blog posts on Kaggle.com are published using Jupyter. Another advantage is that when you push Zeppelin notebooks to github.com they are stored as JSON files. Github is not able to display that in an easy-to-read format. But Github can display Jupyter notebooks, because it can understand and render the Jupyter .ipynb format.

Installing Jupyter

Zeppelin is easy to install as well. You just download it and unzip it. But the Jupyter installation might even be simpler.

As with all things Python, you should set up Jupyter using a virtual environment. That means virtualenv or Anaconda. Jupyter recommends using Anaconda, but I prefer virtualenv because I have found that it is easier to install TensorFlow and Keras with that.

So the first step is to install virtualenv.:

```
sudo pip3 install -U virtualenv
```

Install Python3.4, as that version works with TensorFlow and Keras (So do other versions of Python, but the instructions here have been tested with that.). On Mac this is made easier with:

```
brew install pyenv
export PYENV_ROOT=/usr/local/var/pyenv
pyenv install 3.4.4
```

Make a link to that so that virtualenv can find it:

```
ln -s /usr/local/var/pyenv/versions/3.4.4/bin/python3 /usr/bin/python3
```

Then create the Python3.4 environment:

```
virtualenv -p python3.4 py34
```

Then activate it with:

```
source py34/bin/activate
```
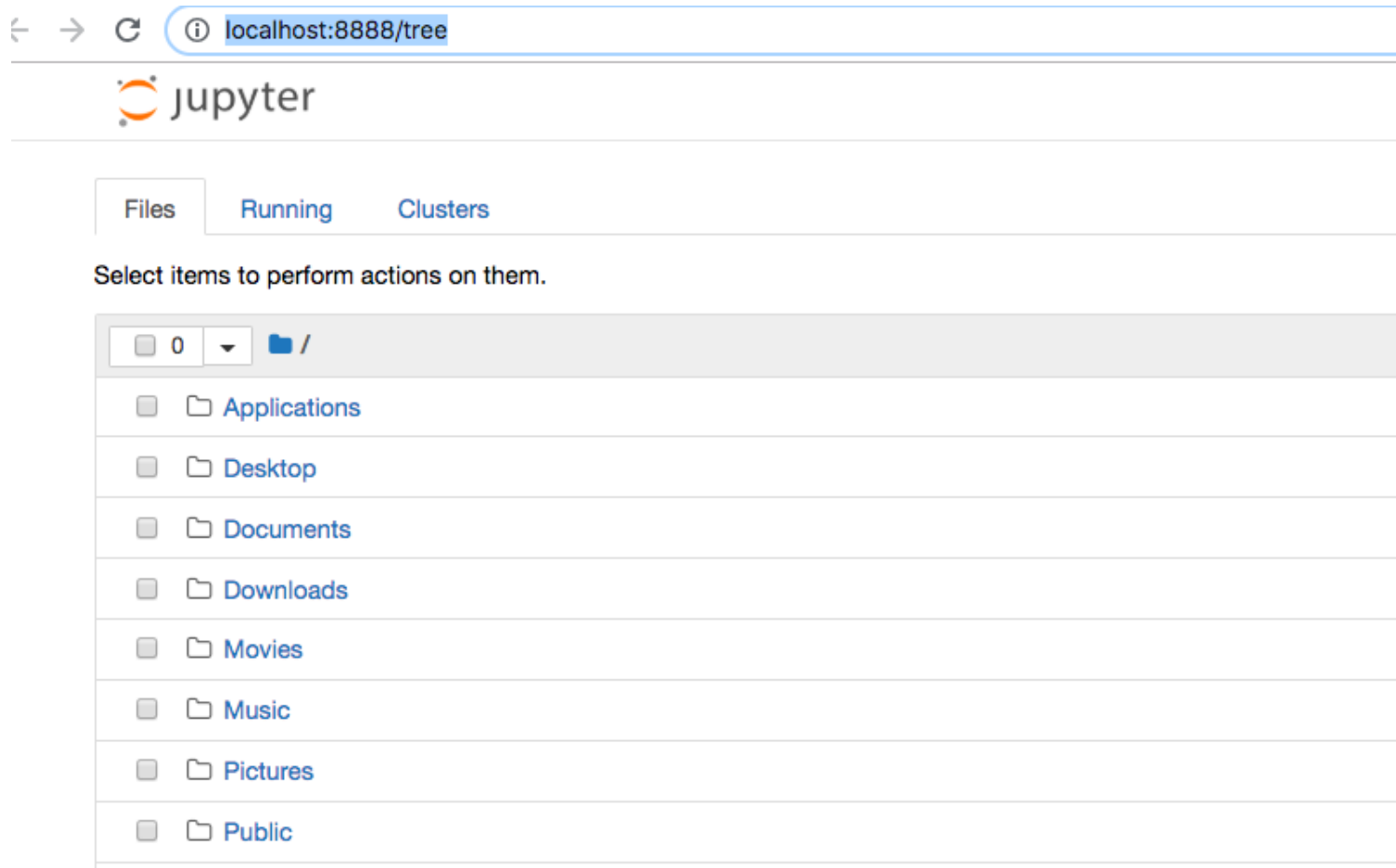
Then launch it with:

```
jupyter notebook
```

If it reports that Jupyter is not found then use:

```
python py34/lib/python3.4/site-packages/jupyter.py notebook
```

It should open Jupyter in the browser at:

```
http://localhost:8888/tree
```

Installing TensorFlow and Keras

Now, you would want to use pip to install TensorFlow and Keras so that you can do something useful with Jupyter. You will also want Pandas, Numpy, etc. Then those will be in scope and available when you run the jupyter notebook command.

```
pip install --upgrade tensorflow
pip install keras
```