Setting up Anaconda 3+

cp ~/Downloads/Anaconda3-4.0.0-Linux-x86\_64.sh ~/

chmod +x Anaconda3-4.0.0-Linux-x86\_64.sh

Run

./Anaconda3-4.0.0-Linux-x86\_64.sh

Press Enter when asked for.

Type Yes when asked for.

Once the installation is done , type the below

source .bashrc

jupyter notebook

Creating a virtual private environment.

Conda

http://bit.ly/condaquickinstall

Download the required version for your OS .

wget https://repo.continuum.io/miniconda/Miniconda-latest-MacOSX-x86\_64.sh

chmod +x Miniconda-latest-MacOSX-x86\_64.sh

# Agree to licenses

./Miniconda-latest-MacOSX-x86\_64.sh

source ~/.bashrc

Create the environment of your choice. Python2 / Python3 / R

conda update conda -y

# Python 2

conda create -n py2 python=2 anaconda jupyter notebook -y

# Python 3

conda create -n py3 python=3 anaconda jupyter notebook -y

# Setup a seperate kernel after you activate a particular env , if this is not done your jupyter could be using default python

conda install notebook ipykernel

# Install R

conda create -n jupyter\_r -c r r-irkernel r-essentials r-recommended rpy2 -y

source activate jupyter\_r

source deactivate

conda env list

Calling Jupyter notebook with Spark configured.

Download Spark .

From Spark directory call the below

Calling Ipython notebook

PYSPARK\_DRIVER\_PYTHON=jupyter PYSPARK\_DRIVER\_PYTHON\_OPTS="notebook " bin/spark/bin/pyspark

For using Python 3 try the below.

PYSPARK\_PYTHON=python3 PYSPARK\_DRIVER\_PYTHON=jupyter PYSPARK\_DRIVER\_PYTHON\_OPTS="notebook " bin/spark/bin/pyspark

create a alias to start spark on Jupyter notebook , add the below in bash\_profile or bashrc file.

alias jspark='PYSPARK\_PYTHON=python3 PYSPARK\_DRIVER\_PYTHON=jupyter PYSPARK\_DRIVER\_PYTHON\_OPTS="notebook " ~/sparkhomedir'

Adding Library to PySpark

PYSPARK\_DRIVER\_PYTHON=jupyter PYSPARK\_DRIVER\_PYTHON\_OPTS="notebook " bin/spark/bin/pyspark --packages com.databricks:spark-csv\_2.11:1.4.0

CSV READER IS ADDED. If you want other packages from maven repository , you can include them after packages.