

0) pip install conda (contient jupyter)

sudo apt-get install python-pip

1) Exécuter la commande suivante pour télécharger anaconda :

wget https://repo.continuum.io/archive/Anaconda2-4.4.0-Linux-x86_64.sh

2) Sous l'utilisateur HDFS :

Editer le fichier .bashrc situé dans le home :

nano ~/.bashrc

3) Ajouter les lignes suivantes à la fin du fichier puis sauvegarder :

Cloudera :

```
export SPARK_HOME=/opt/cloudera/parcels/SPARK2-2.2.0.cloudera1-1.cdh5.12.0.p0.142354/lib/spark2
```

Horton :

```
export SPARK_HOME=/usr/hdp/2.6.1.0-129/spark2
export PYTHONPATH=$SPARK_HOME/python/:$PYTHONPATH
export PYTHONPATH=$SPARK_HOME/python/lib/py4j-0.10.4-src.zip:$PYTHONPATH
export ANACONDA_HOME=~/.anaconda2
export PYSPARK_DRIVER_PYTHON=$ANACONDA_HOME/bin/jupyter
export PYSPARK_DRIVER_PYTHON_OPTS="notebook --no-browser --port=7777"
export PYSPARK_PYTHON=$ANACONDA_HOME/bin/python
export PYSPARK_SUBMIT_ARGS="--master local[2] spark-shell"
export PATH=$ANACONDA_HOME/bin:$SPARK_HOME/bin:$PATH
```

4) Déployer la configuration du bashrc :

source .bashrc

5) Rendre le fichier d'installation d'anaconda exécutable avec la commande suivante

(répondre no quand l'installation demande de modifier le bashrc) :

chmod +x Anaconda2-4.4.0-Linux-x86_64.sh

6) Installer Anaconda :

./Anaconda2-4.4.0-Linux-x86_64.sh

7) Faire un redirect de port localhost :

[http://www.learn4master.com/tools/jupyter-notebook/how-to-access-ipython-notebook-runningon-](http://www.learn4master.com/tools/jupyter-notebook/how-to-access-ipython-notebook-runningon-remote-server)

remote-server

amiral\$ ssh -N -f -L localhost:7775:localhost:7777 ubuntu@145.239.15.48

7) lancer la commande pyspark --master local

8) Copier l'URL dans votre navigateur web

Lancer le navigateur anaconda :

anaconda-navigator (edited)

"/opt/cloudera/parcels/SPARK2-2.2.0.cloudera1-

1.cdh5.12.0.p0.142354/lib/spark2/data/mllib/sample_libsvm_data.txt"