**Instlling Anaconda on Linux**

Cd /root/Desktop #### to go to the path/location

Ls #### to list all the files in that location

Continue with installation

Conda update spyder #### updating

You can call spyder

**Installing Spark**

**Step 1: Installing Java**

Check to see if Java is already installed by typing:

java -version

If you see something like “The program ‘java’ can be found in the following packages…”

It probably means you actually need to install Java.

First, update the package index by in your terminal typing:

sudo apt-get update

After entering your password it will update some stuff.

Now you can install the JDK with the following command:

sudo apt-get install default-jdk

**Step 3: Install Spark**

Next its time to install Spark. We need git for this, so in your terminal type:

sudo apt-get install git

Next, go to <https://spark.apache.org/downloads.html> and download a pre-built for Hadoop 2.7 version of Spark (preferably Spark 2.0 or later). Then download the .tgz file and remember where you save it on your computer.

Then in your terminal change directory to where you saved that .tgz file (or just move the file to your home folder), then use

tar xvf spark-2.0.2-bin-hadoop2.7.tgz

(Your version numbers may differ).

Then once its done extracting the Spark folder, use:

cd spark-2.0.2-bin-hadoop2.7.tgz

then use

cd bin

and then type

./spark-shell

and you should see the spark shell pop up

this is where you can load .scala scripts. You can confirm that this is working by typing something like:

println(“Spark shell is running”)

That’s it! You should now have Spark running on your computer.

**Start A Jupyter Session**

Windows

Go to CMD and type “Jupyter notebook” then enter

Linux

Go to terminal and type “Jupyter notebook” the enter if it does not open type “Jupyter notebook –allow-root”

Instlling Anaconda on Linux

Cd /root/Desktop #### to go to the path/location

Ls #### to list all the files in that location

Continue with installation

Conda update spyder #### updating

You can call spyder

**Evironemnt Variables**

User variables path has No spark or hadoop

**Path**

C:\Users\Wesh\Anaconda3;C:\Users\Wesh\Anaconda3\Library\mingw-w64\bin;C:\Users\Wesh\Anaconda3\Library\usr\bin;C:\Users\Wesh\Anaconda3\Library\bin;C:\Users\Wesh\Anaconda3\Scripts;C:\Program Files\Intel\WiFi\bin\;C:\Program Files\Common Files\Intel\WirelessCommon\

HADOOP\_HOME===C:\HADOOP

JAVA\_HOME===C:\Java\jdk1.8.0\_151

SPARK\_HOME====C:\sparkhadoop

**System Variables**

**Path** C:\ProgramData\Oracle\Java\javapath;C:\app\Datascientist\product\12.1.0\dbhome\_1\bin;%SystemRoot%\system32;%SystemRoot%;%SystemRoot%\System32\Wbem;%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\;C:\Program Files\Intel\WiFi\bin\;C:\Program Files\Common Files\Intel\WirelessCommon\;C:\Program Files (x86)\Microsoft SQL Server\100\Tools\Binn\;C:\Program Files\Microsoft SQL Server\100\Tools\Binn\;C:\Program Files\Microsoft SQL Server\100\DTS\Binn\;C:\Program Files (x86)\Microsoft SQL Server\100\Tools\Binn\VSShell\Common7\IDE\;C:\Program Files (x86)\Microsoft SQL Server\100\DTS\Binn\;C:\Program Files (x86)\Microsoft Visual Studio 9.0\Common7\IDE\PrivateAssemblies\;C:\Java\jdk1.8.0\_151\bin;C:\Java\jre1.8.0\_151\bin;C:\HADOOP\bin;C:\Program Files (x86)\sbt\bin;C:\Program Files (x86)\scala\bin