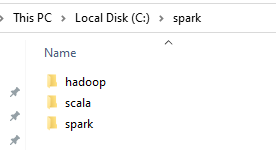
Step 1:

* Within “drive C”

Create folder “spark”

Within folder “spark”, create three folders

* hadoop
* scala
* spark



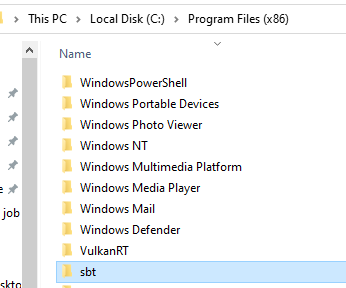
Step 2:

* Download **\*.msi** file in this link

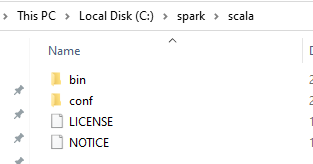
<https://www.scala-sbt.org/download.html?_ga=2.216452502.1266178000.1614007509-1351715761.1614007509>

* Install the msi file

Note: file will be automatically installed in the location “C:\Program Files (x86)\sbt”

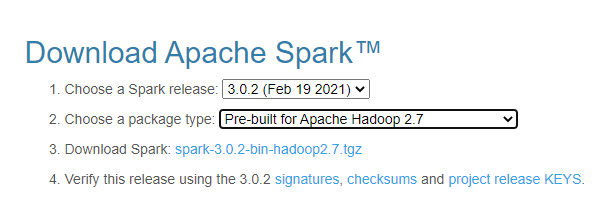


* Download **\*.zip** under the “All platform” section and unzip it. Copy the contents from the unzipped folder and paste in the empty “scala” folder we created under the folder “spark” in “c drive”

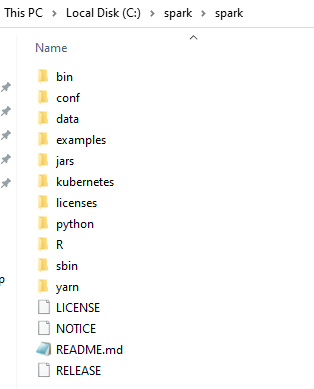


Step 3:

* Go to this page: <https://spark.apache.org/downloads.html>



* Click on the 3rd step to download the file after selecting spark release and package type in step 1 and step 2
* Unzip the file and copy the contents and paste them in the “spark” folder we created under the “spark” folder in the “c drive”.

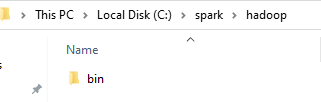


Step 4:

* Go to the <https://github.com/steveloughran/winutils>
* Depending on the package type you select in Step3
* Download the folder from above link

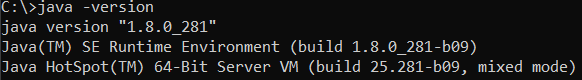
For ex, since I selected “prebuilt for apache Hadoop 2.7” , I downloaded “Hadoop-2.7.1” folder.

* Unzip the folder, copy its all contents and paste in the “hadoop" folder we created within the “spark” folder within the “c drive”



Step 5:

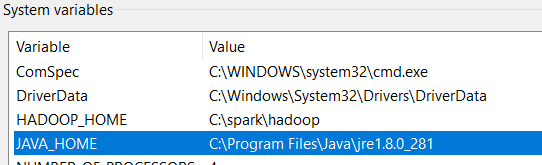
* Check if JAVA is installed
* Open cmd (windows command prompt) , or anaconda prompt, from start menu and run



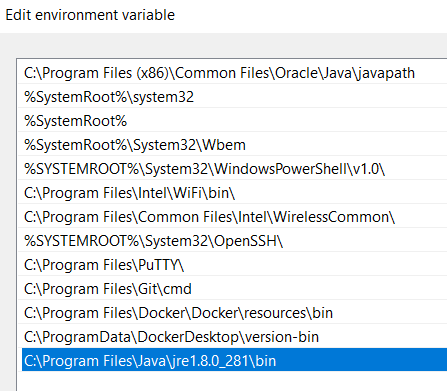
* If you get error in this step, download java from the below link

<https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html>

* Add the following environment variable,



* Add to the path



Step 6:

* Install anaconda from this link <https://docs.anaconda.com/anaconda/install/windows/>

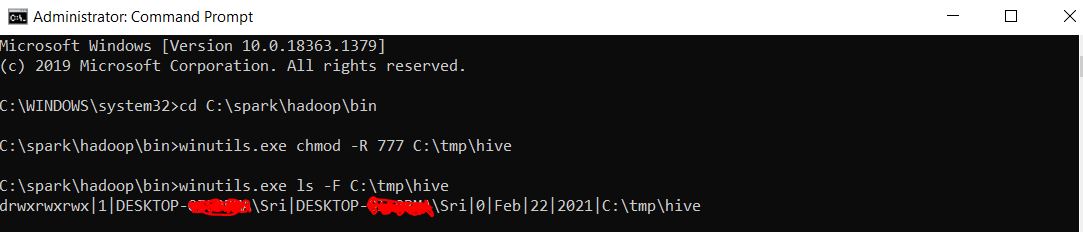
Step 7:

* Create the folder C:\tmp\hive
* Execute the following command in cmd started using the option Run as administrator.

winutils.exe chmod -R 777 C:\tmp\hive

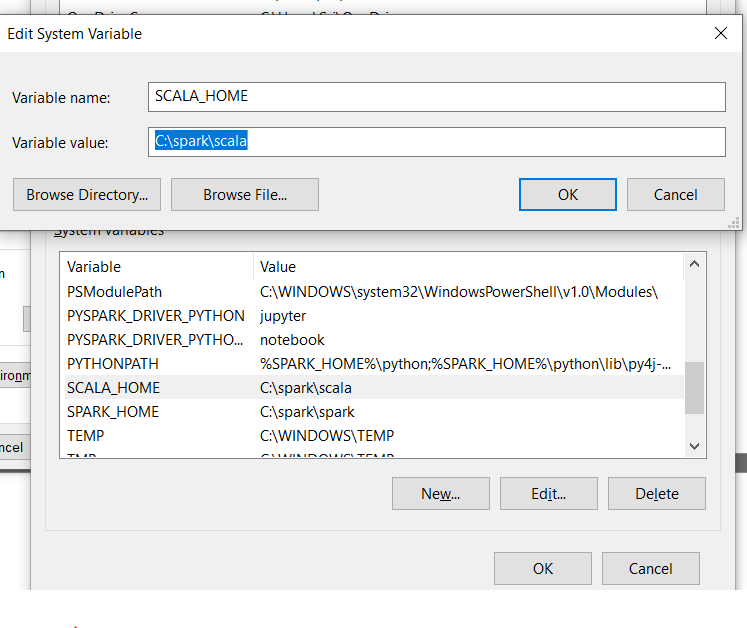
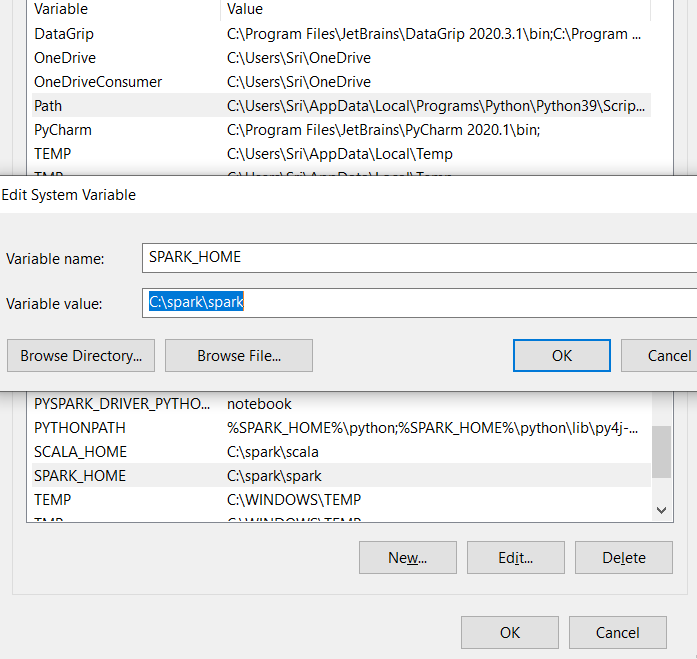
winutils.exe ls -F C:\tmp\hive

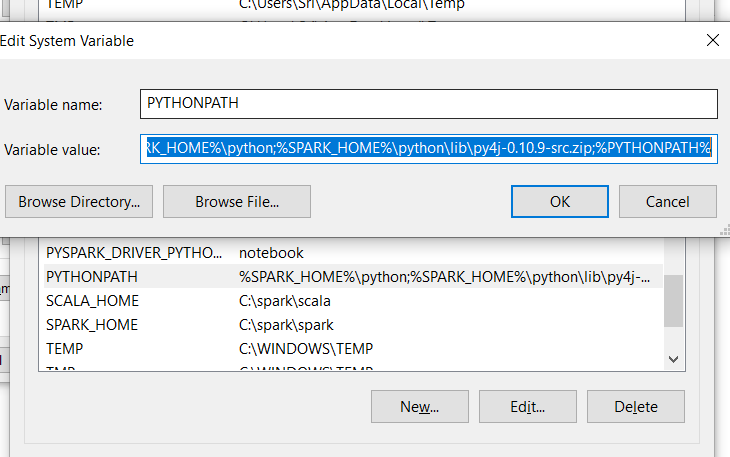
* You must see the output as follows

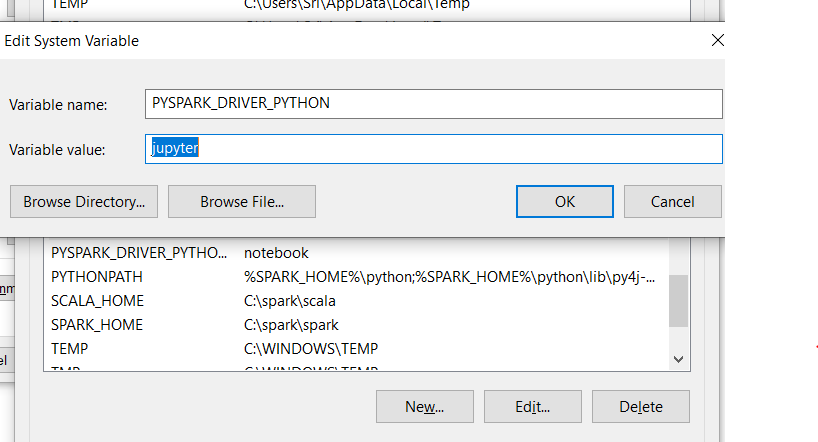
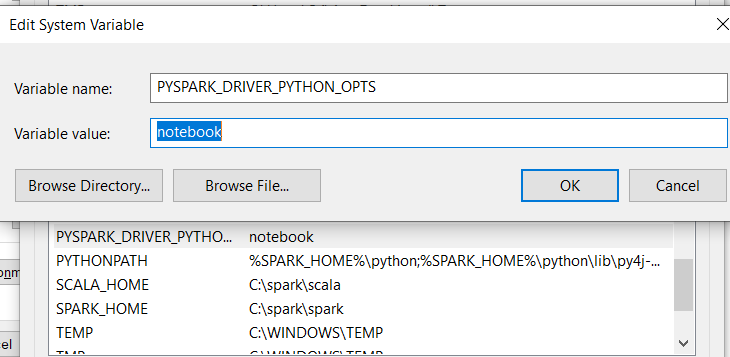


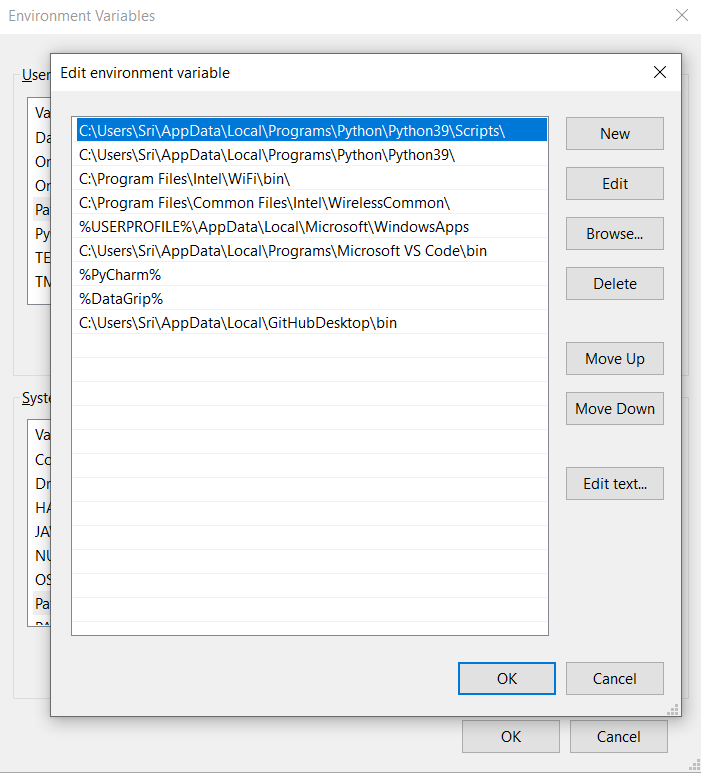
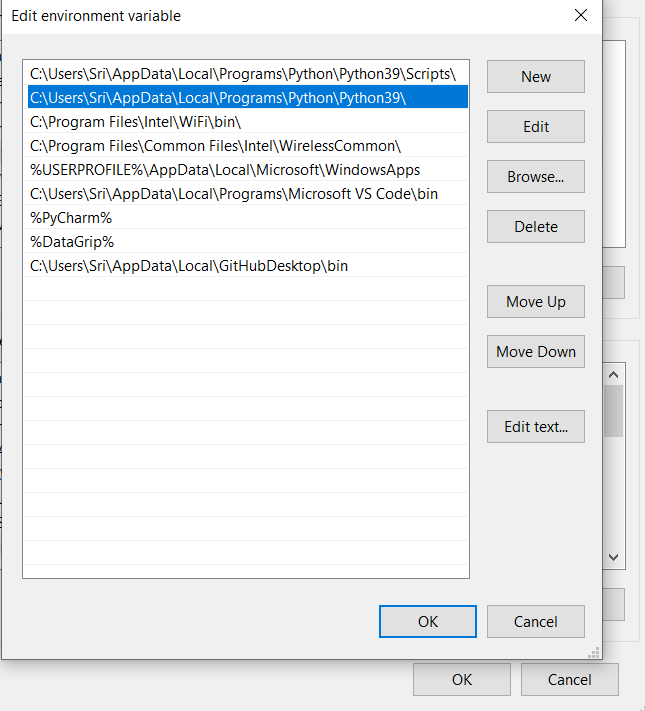
Step8:

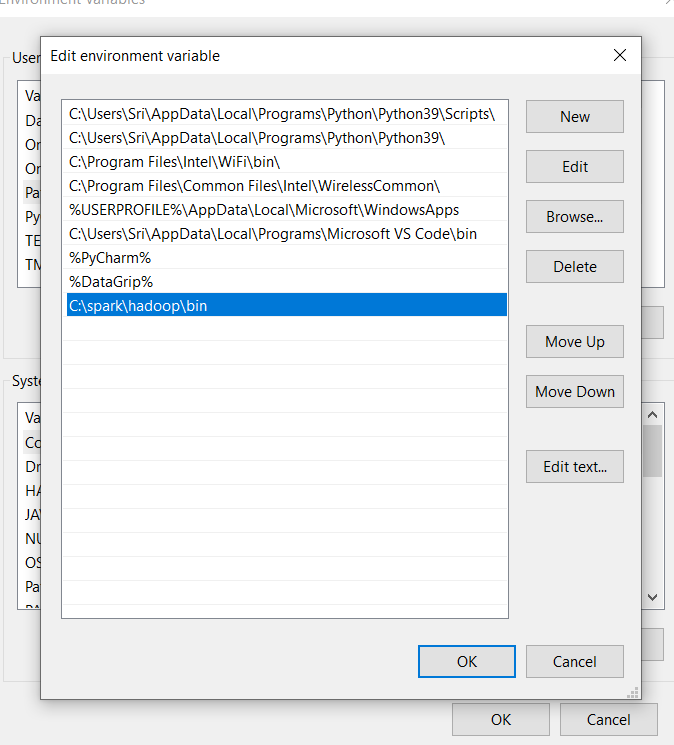
Add these following to the paths



%SPARK\_HOME%\python;%SPARK\_HOME%\python\lib\py4j-0.10.9-src.zip;%PYTHONPATH%







Step 9:

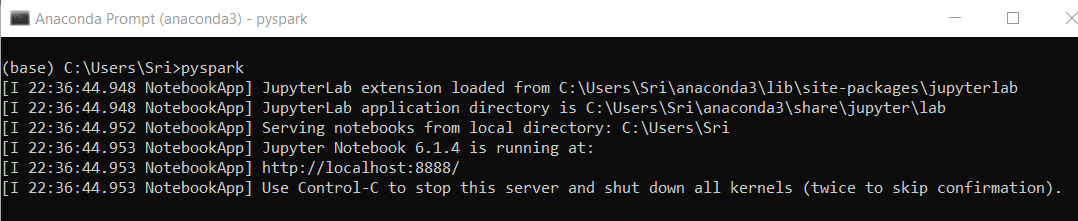
* Click ok and restart the system

Step 10:

* Open anaconda prompt, type

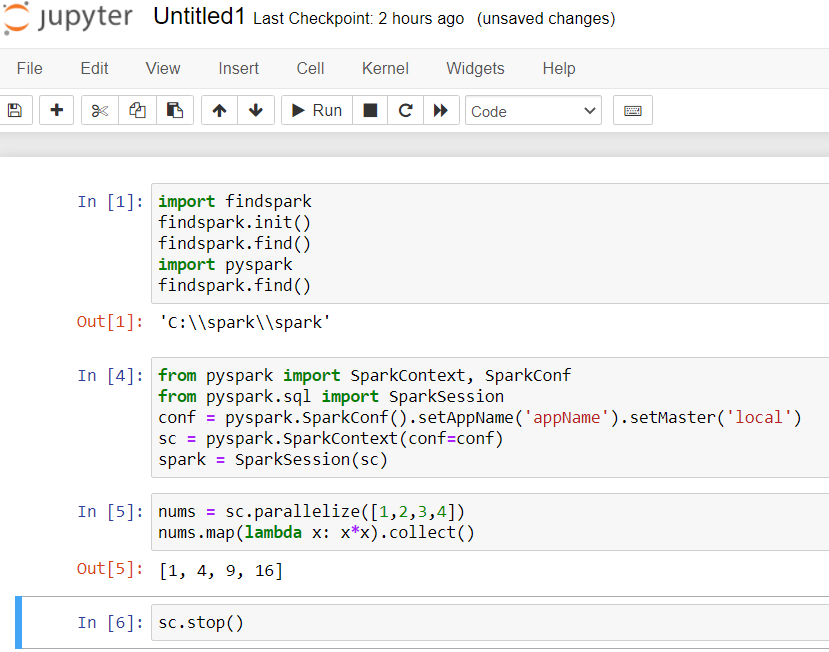
python -m pip install findspark

* If it show some error



You see the Jupter notbook is running

* To verify pyspark operation in jupter notebook



* To run spark in Pycharm, check this

<https://medium.com/big-data-engineering/how-to-install-apache-spark-2-x-in-your-pc-e2047246ffc3>