SPARK With Spyder (with Python)

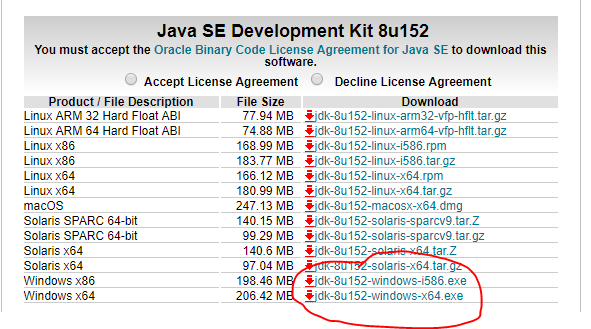
**Prerequisites:**

1. Spyder
2. Java SDK - 1.8.0\_152
3. Spark 2.2.0
4. Pyspark Python package
5. Winutils
6. Setting Environment variables

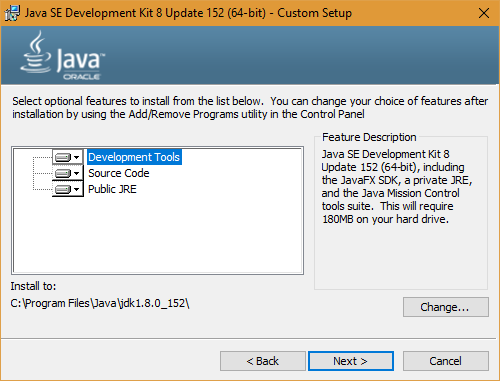
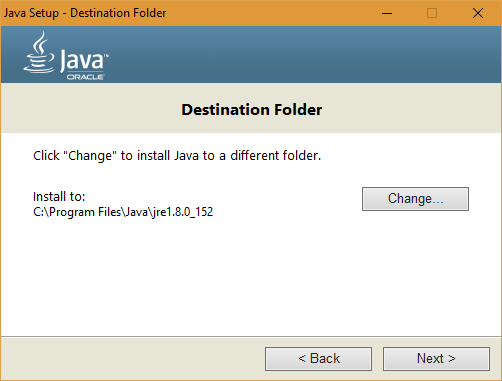
2. Installing Java:

Download the Java SDK (1.8.0\_152) setup file from following site.

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>



Run the installer to install. Click Next to continue.

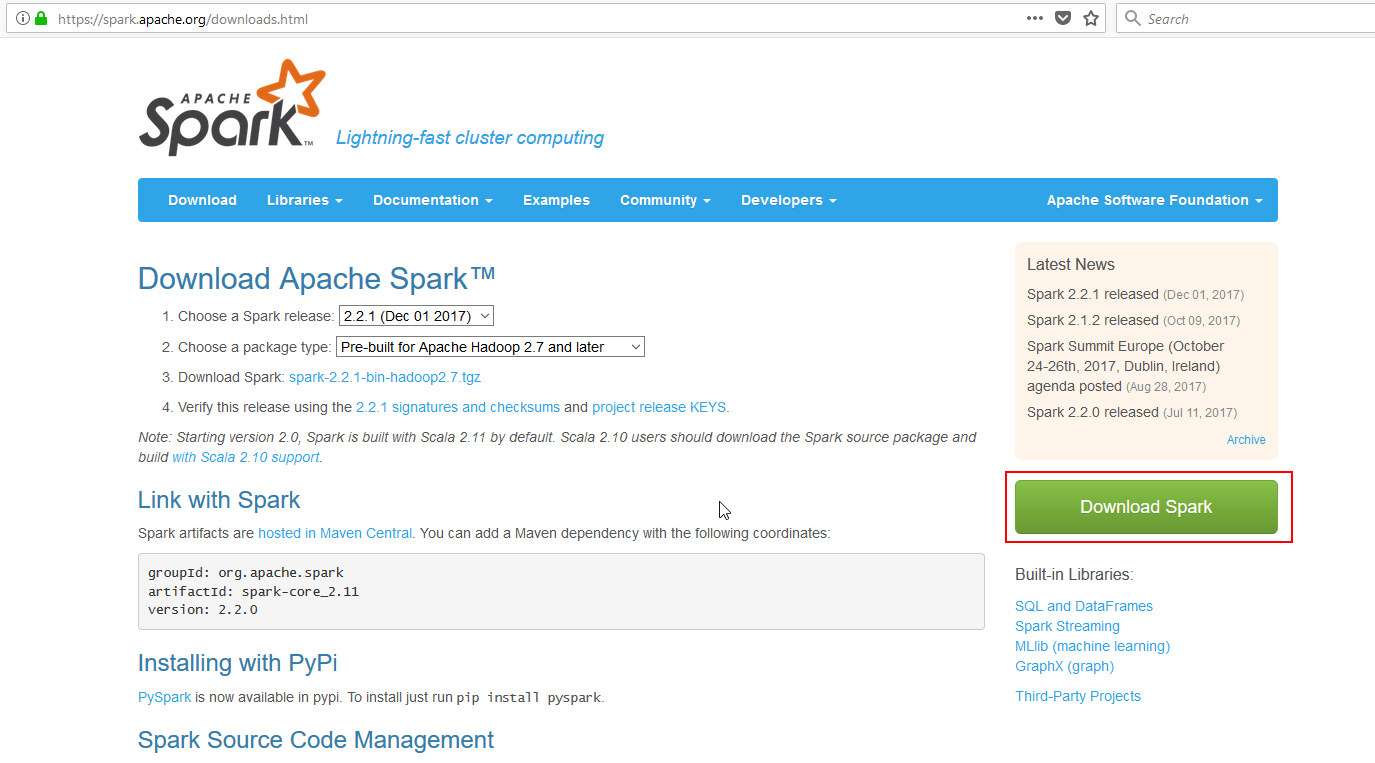
 

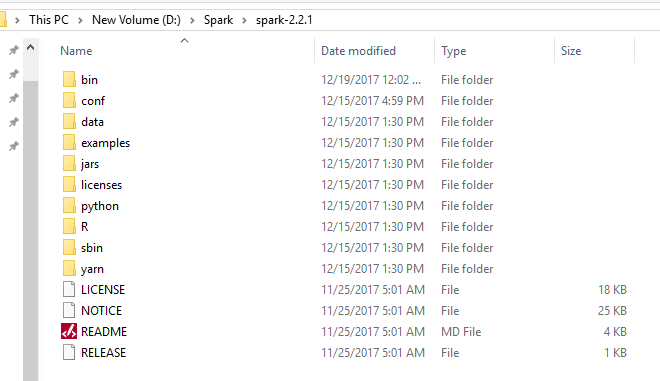
2. Installing Spark:

Download a **pre-built** version of Apache Spark from

<https://spark.apache.org/downloads.html>

Note: If required, download and install WinRAR so you can extract the .tgz file you downloaded. <http://www.rarlab.com/download.htm>

Extract the Spark archive, and copy its **contents** into **D:\Spark\spark-2.2.1** after creating that directory. You should end up with directories like



4. Installing Pyspark package:

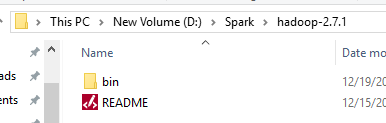
Open Anaconda command prompt (Go to start and type Anaconda) and run following command

pip install pyspark

5. Downloading Winutils:

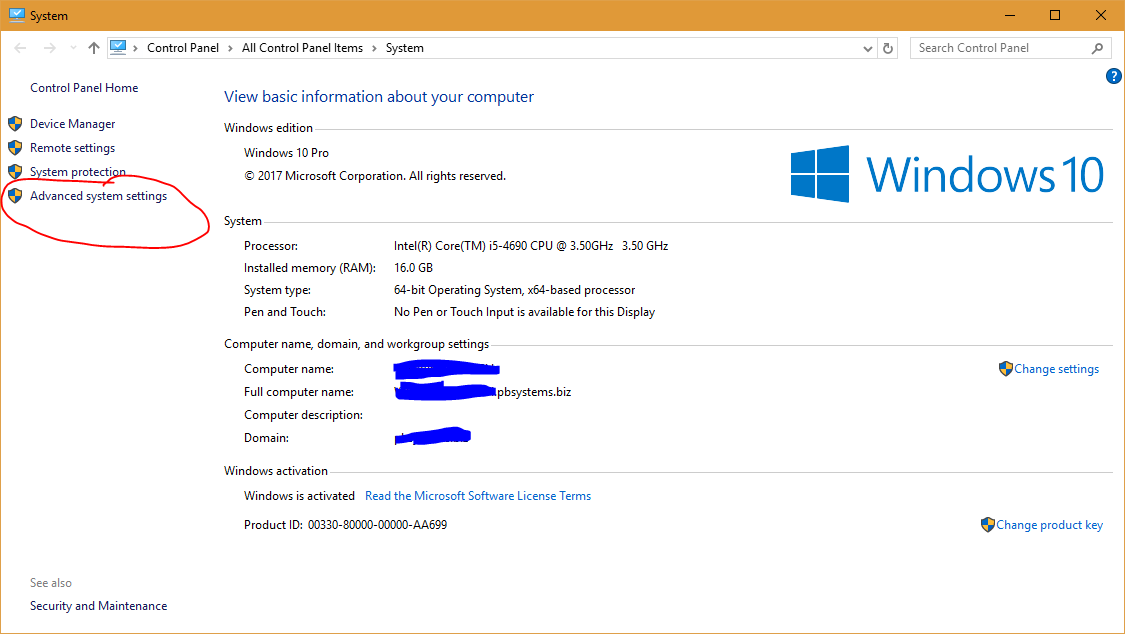
Download “hadoop-2.7.1” from below link

<https://github.com/steveloughran/winutils>

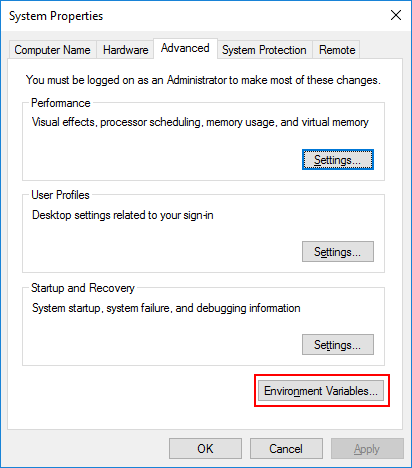
And move folder contents into a **D:\Spark\hadoop-2.7.1** folder that you’ve created. Folder should look like 

6. Setting environment variables:

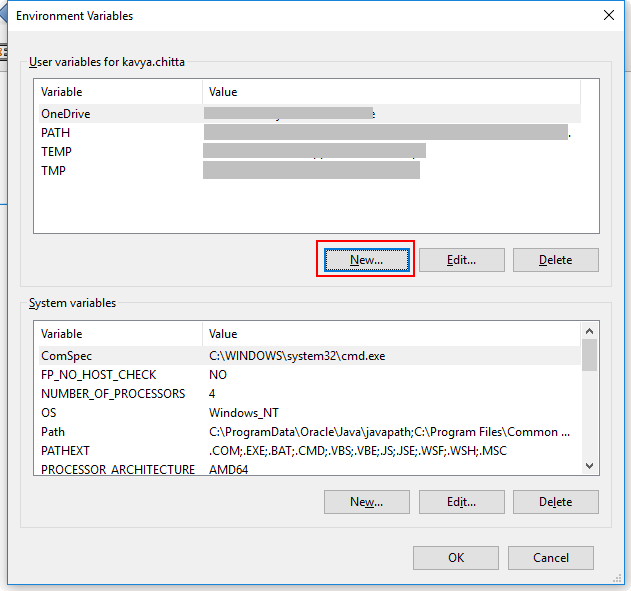
1. Right-click your Windows menu, select System.
2. Click on “Advanced System Settings” and then the “Environment Variables” button.



1. Click on “New” and then the “Environment Variables” button.



1. Click on “New” and then the “Environment Variables” button.

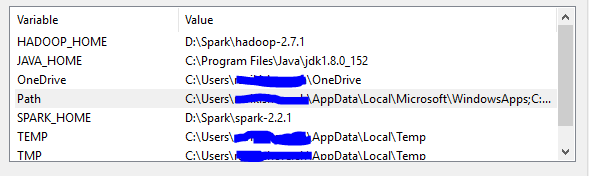


Enter name: SPARK\_HOME

Value: <<Your spark extracted path>> Ex: D:\Spark\spark-2.2.1

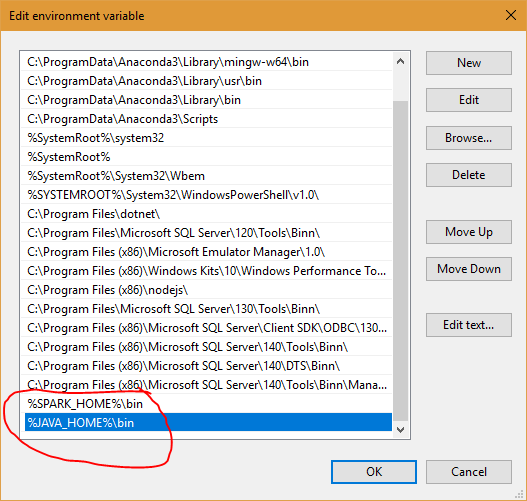
1. Similarly repeat step 4 for following variables,

|  |  |
| --- | --- |
| **Name** | **Value** |
| JAVA\_HOME | C:\Program Files\Java\jdk1.8.0\_152 |
| HADOOP\_HOME | D:\Spark\hadoop-2.7.1 |



1. Now double click on “Path” under system Variables and click “Add” button and enter following

%SPARK\_HOME%\bin



1. Again click “Add” button and enter following

%JAVA\_HOME%\bin

1. Finally create a temp directory “D:\tmp\hive” and run following command. Navigate to bin folder in winutils downloaded above.

winutils.exe chmod 777 D:\tmp\hive