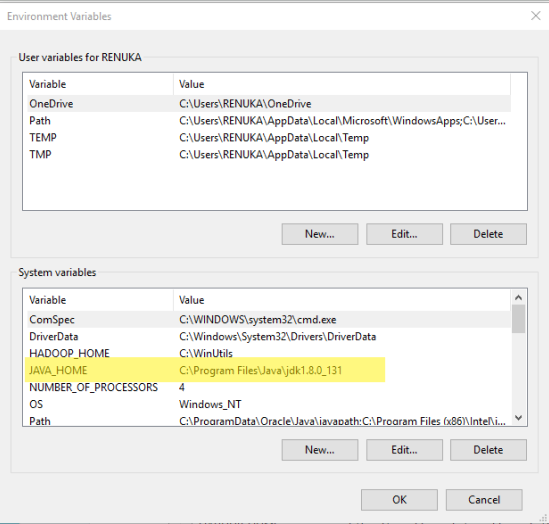
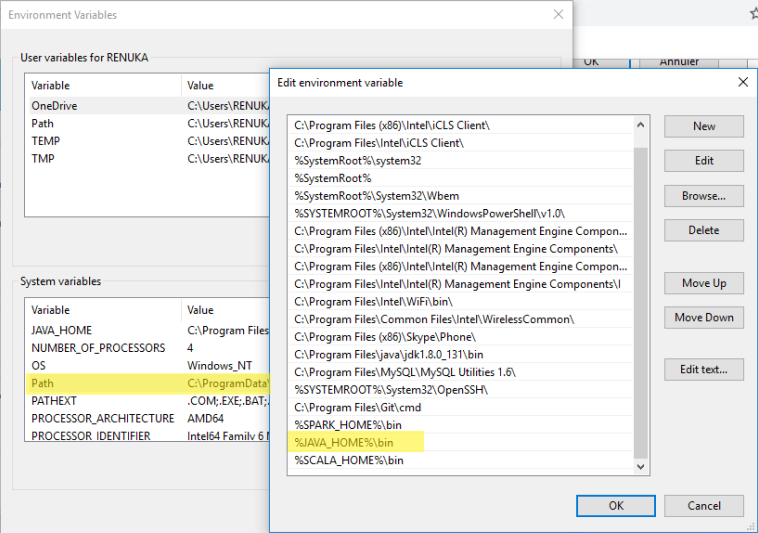
**Installation of Apache Spark on Windows 10**

1. Install Java
   1. First, we need to install Java to execute Spark applications
   2. To do that, go to this [page](https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html) and download the latest version of the JDK.
   3. After installing it, add the JAVA\_HOME variable to your system variables and make sure its path is pointing to the JDK parent folder.



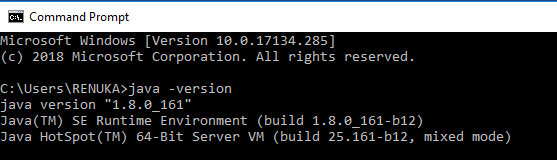
* 1. After adding this variable, modify the **Path** system variable and add a new entry like this: **%JAVA\_HOME%\bin.**This will let Windows command line recognize Java commands



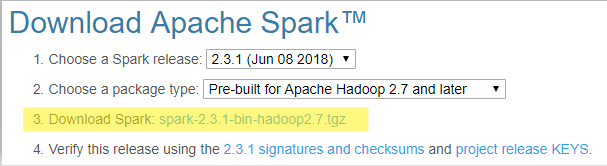
* 1. Now, start the command line and type

Java -version

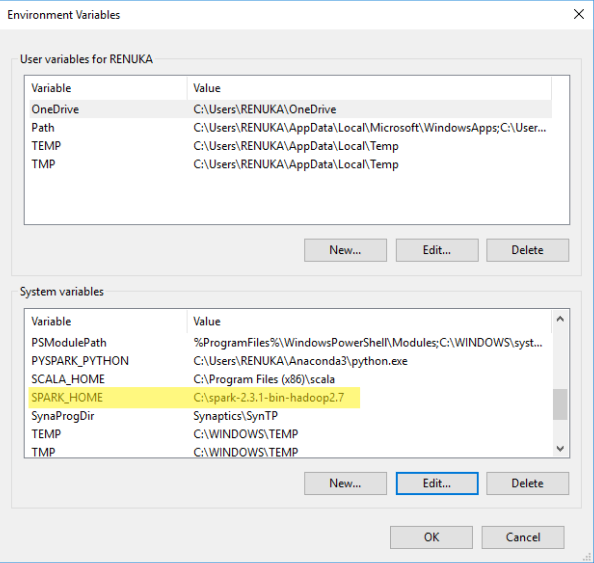
to check if java was installed correctly



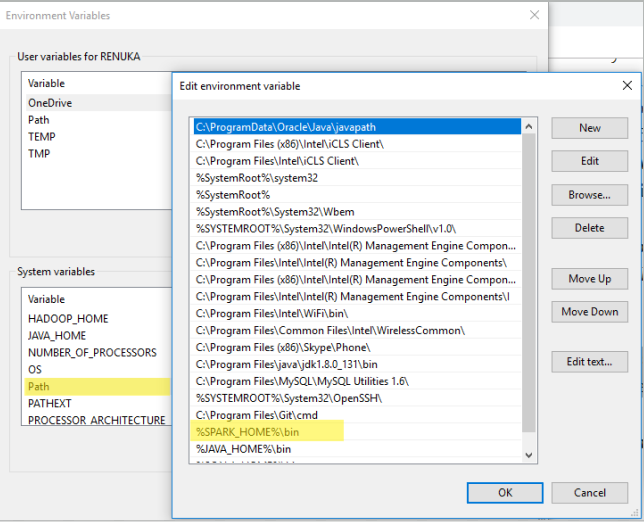
1. Spark Binaries
   1. Go to this [page](https://spark.apache.org/downloads.html" \t "_blank) and choose the latest stable version pre-built for Hadoop 2.7 and later (see figure below)



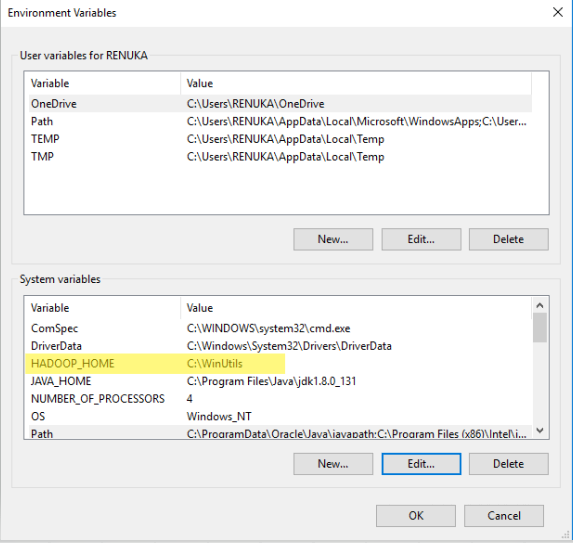
* 1. Extract the contents of*spark-2.3.1-bin-hadoop2.7.tgz* at C:\Spark\
  2. Add a new variable to your system variables and name it: **SPARK\_HOME.**This variable holds the Spark parent directory path (C:\spark-2.3.1-bin-hadoop2.7 for example).



* 1. Add **%SPARK\_HOME%\bin** to **Path** system variable.



1. **Hadoop WinUtils**
   1. Create a new folder and name it **“WinUtils”** and place it in a parent directory of any partition (**C:\WinUtils** for example)
   2. Go to this [page](https://github.com/steveloughran/winutils" \t "_blank) and download this repository by clicking in the right green button and choosing Download ZIP option.
   3. After downloading the zip file, extract it and copy files from this folder **“hadoop-2.7.1”**to **WinUtils** folder (don’t copy the whole directory, just its content, the **bin** folder).
   4. Add **HADOOP\_HOME** variable to your system variables and point it to the WinUtils folder (C:\WinUtils in this case).



1. Install Anaconda and Jupyter
   1. Latest version available at <https://www.anaconda.com/download/>
   2. Add a new variable to your system variables and name it: **PYSPARK\_PYTHON.**This variable holds the path for python.exe (C:\Users\<user\_name>\Anaconda3\python.exe) for example).
2. Run pyspark
   1. Open command line and type: pyspark

