1. Install a JDK (Java Development Kit) from http://www.oracle.com/technetwork/java/javase/downloads/index.html . Keep track of where you installed the JDK; you’ll need that later.

2. Download a pre-built version of Apache Spark from https://spark.apache.org/downloads.html

3. If necessary, download and install WinRAR so you can extract the .tgz file you downloaded. http://www.rarlab.com/download.htm

4. Extract the Spark archive, and copy its contents into C:\spark after creating that directory. You should end up with directories like c:\spark\bin, c:\spark\conf, etc.

5. Move winutils.exe (Download from - https://s3.amazonaws.com/sankethadoop/winutils.exe) into a C:\winutils\bin folder that you’ve created. (note, this is a 64-bit application. If you are on a 32-bit version of Windows, you’ll need to search for a 32-bit build of winutils.exe for Hadoop.)

6. Open the the c:\spark\conf folder, and make sure “File Name Extensions” is checked in the “view” tab of Windows Explorer. Rename the log4j.properties.template file to log4j.properties. Edit this file (using Wordpad or something similar) and change the error level from INFO to ERROR for log4j.rootCategory

7. Right-click your Windows menu, select Control Panel, System and Security, and then System. Click on “Advanced System Settings” and then the “Environment Variables” button.

8. Add the following new USER variables:

a. SPARK\_HOME c:\spark

b. JAVA\_HOME (the path you installed the JDK to in step 1, for example C:\Program Files\Java\jdk1.8.0\_101)

c. HADOOP\_HOME c:\winutils

9. Add the following paths to your PATH user variable: %SPARK\_HOME%\bin; %JAVA\_HOME%\bin

10. Close the environment variable screen and the control panels.

11. Install the latest Scala IDE from http://scala-ide.org/download/sdk.html

12. Test it out!

a. Open up a Windows command prompt in administrator mode.

b. Enter cd c:\spark and then dir to get a directory listing.

c. Look for a text file we can play with, like README.md or CHANGES.txt

d. Enter spark-shell

e. At this point you should have a scala> prompt. If not, double check the steps above.

f. Enter val rdd = sc.textFile(“README.md”) (or whatever text file you’ve found)

g. Enter rdd.count()

h. You should get a count of the number of lines in that file! Congratulations, you just ran your first Spark program!

i. Hit control-D to exit the spark shell, and close the console window

j. You’ve got everything set up! Hooray!