**Installation steps**

**Spark Environment Setup on Window Machine**

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. DO NOT INSTALL JAVA 9 – INSTALL JAVA 8. Spark is not compatible with Java 9. And BE SURE TO INSTALL JAVA TO A PATH WITH NO SPACES IN IT. Don’t accept the default path that goes into “Program Files” on Windows, as that has a space.
2. Install Scala : <https://www.scala-lang.org/>
3. Install python: <https://www.python.org/>
4. Install R: <https://www.r-project.org/>
5. Download a pre-built version of Apache Spark 2.2 (not 2.3 or newer!) from <https://spark.apache.org/downloads.html>
6. If necessary, download and install WinRAR so you can extract the .tgz file you downloaded. <http://www.rarlab.com/download.htm>
7. 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.
8. Please refer the software folder for winutils.exe and move it 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.)
9. 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
10. 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.
11. Add the following new USER variables:

SPARK\_HOME c:\spark

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

HADOOP HOME c:\winutils

Add the following paths to your PATH user variable:

%SPARK\_HOME%\bin

%JAVA\_HOME%\bin

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

Test it out!

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

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

1. Look for a text file we can play with, like README.md or CHANGES.txt
2. Enter spark-shell 1.6 or spark-shell2.cmd for spark 2.0

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

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

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

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

You’ve got everything set up! Good Job!

**Install Scala IDE**

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

**Install Intelij**

<https://www.jetbrains.com/idea/download/#section=windows>