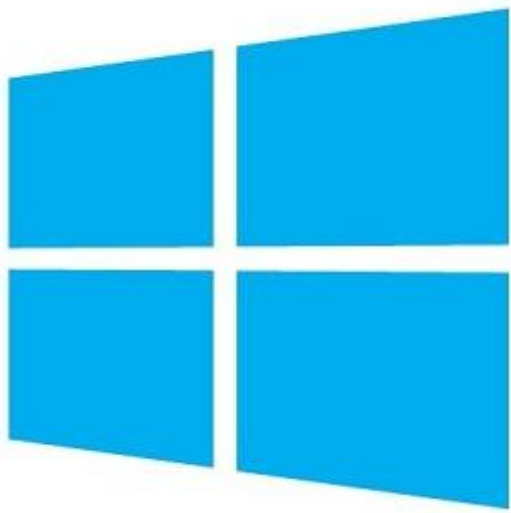


[Get started](#)[Open in app](#)

## Jose Marcial Portilla

[Follow](#)

4.3K Followers

[About](#)

# Installing Scala and Spark on Windows



Jose Marcial Portilla Dec 16, 2016 · 3 min read

Hi! I'm Jose Portilla and I teach over 200,000 students about programming, data science, and machine learning on Udemy! You can check out all my courses [here](#).

If you're interested in learning Python for Data Science and Machine learning, [check out my course here](#). (I also teach [Full Stack Web Development with Django!](#))

Quick guide to installing a basic Scala and Spark set-up on Windows. Spark is written in Scala, which is written in Java, which means we need all three of these things to make sure everything works out! Here are the general steps (if you're enrolled in my

course: [Scala and Spark for Big Data and Machine Learning](#) you can always follow along with the video lecture).

Step 1: Download the latest Java Development Kit that matches your system (32-bit vs 64-bit). You can find the download [website from Oracle here](#) or just Googling “Java Development Kit”.

Step 2: Go to [apache.spark.org](http://apache.spark.org) and download a pre-built version of Spark (pre-built for Hadoop 2.7 and Later) and preferably Spark 2.0 or later.

Step 3: Download winutils.exe in order to make sure that Hadoop works correctly on your computer for Windows. You can find this file as a resource in the video lecture, but you may need to Google for another version if you are running an older version of Windows. (Just google “Spark+winutils” and you should see plenty of links with different sources for the download.

Step 4: Go to your downloaded jdk file and run the installation program, just use all the defaults.

Step 5: Extract the downloaded spark-2.0.2-bin-hadoop2.7-tar.gz file. You may need to extract this twice in order to get the full folder to show.

Step 6: Once you have this folder, go to your C drive and create a new folder called **Spark** and copy and paste the contents of the unzipped spark-2.0.2-bin-hadoop2.7-tar.gz file to this new Spark folder you just created.

Step 7: Create a new folder under your C drive called **winutils**. Then inside of this folder create a new folder called **bin**. Inside of this bin folder place your downloaded winutils.exe file.

Now its time to tell your Windows machine where to find everything, which means we need to edit our environment variables.

Step 8: Go to Control Panel > System and Security > System > Advanced System Settings

Step 9: In the window that pops up, click on the button **Environment Variables**

Step 10: You should see two panels, User Variables and System Variables, click **New...** on the User Variables to create a new variable name and value combination. You'll create the following variables (your full paths may differ slightly):

Variable name: SPARK\_HOME

Variable value: C:\Spark

Variable name: JAVA\_HOME

Variable value: C:\Program Files\Java\jdk1.8.0\_101

Variable name: HADOOP\_HOME

Variable value: C:\winutils

Step 11: Then under your User Variables you should see a variable called **PATH** that is already there. Select it and click Edit

Step 12: You should see a bunch of environment variables already there, but now we're going to add our own paths. Click on new and enter:

**%SPARK\_HOME%\bin**

Then repeat it again and add:

**%JAVA\_HOME%\bin**

Everything should now be installed. Let's test it. Open a command prompt and use **cd** to change directory to C:\Spark and then type:

**spark-shell**

then hit Enter and you should eventually see the Spark Shell display. You can type `:q` to exit out of this.

Hope this was helpful!

A quick note: It is common once you edit and do changes in the PATH that some changes are not recognized. If you follow the steps as shown here and get the 'not recognized' error when trying to run spark-shell, just restart Windows, that should fix the problem.

Programming   Spark

[About](#) [Help](#) [Legal](#)

Get the Medium app

