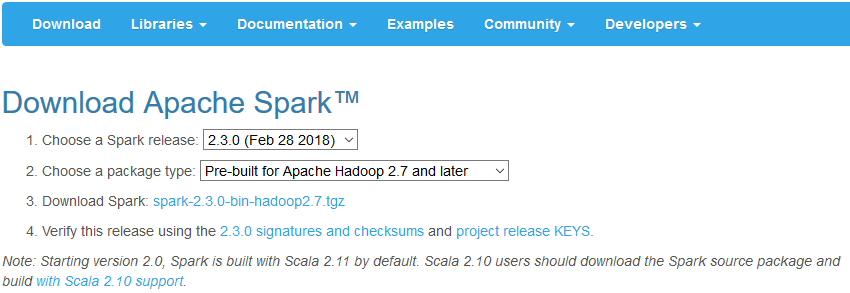
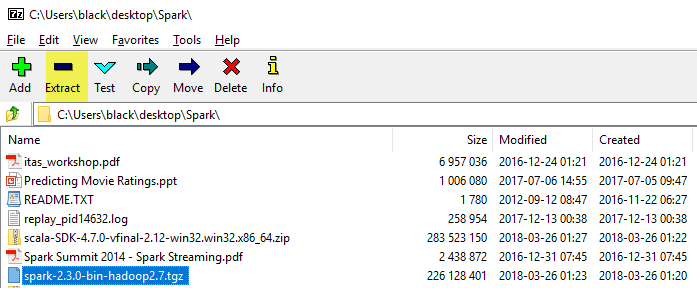
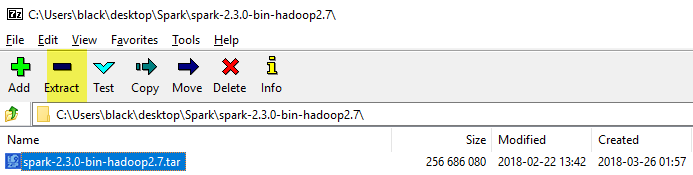
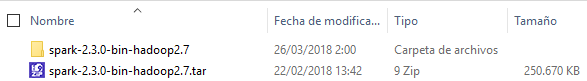
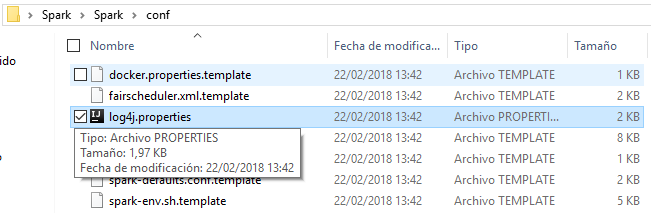
**Steps to install Apache Spark in STANDALONE Mode**

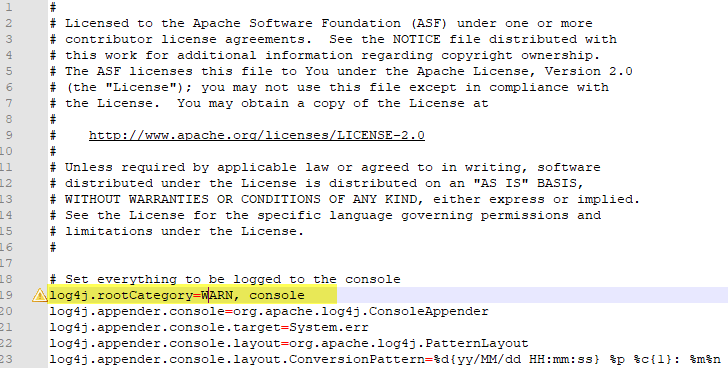
1. Go to Apache Spark [Website](http://spark.apache.org/downloads.html).
2. Select version of Apache Spark. In this case, download version 2.3. Click on [spark-2.3.0-bin-hadoop2.7.tgz](https://www.apache.org/dyn/closer.lua/spark/spark-2.3.0/spark-2.3.0-bin-hadoop2.7.tgz)
3. It will direct you to the Apache Software Foundation. Click on [**http://www-us.apache.org/dist/spark/spark-2.3.0/spark-2.3.0-bin-hadoop2.7.tgz**](http://www-us.apache.org/dist/spark/spark-2.3.0/spark-2.3.0-bin-hadoop2.7.tgz). This will download a tgz file.
4. Move the tgz file to your desktop.
5. Use any tool to unzip the tgz file. In this course, it will be use 7 zip. Open 7 zip, select spark-2.3.0 and click on extract.
6. This will create a new folder with the name of the tgz. Inside the spark-2.3.0-bin-hadoop2.7 folder, it will be a tar file. Select it and click on Extract.
7. This will create a new folder with the name of spark-2.3.0-bin-hadoop2.7.

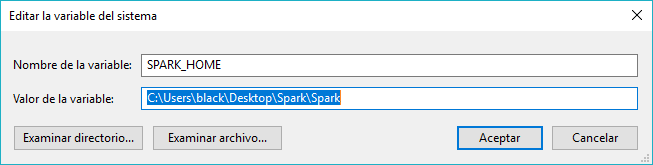
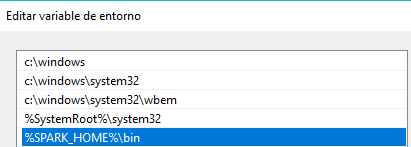
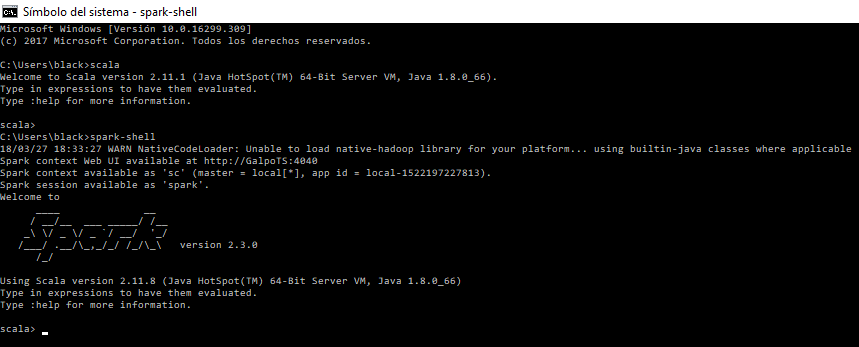


1. Click on the folder and there will be another folder with the same name. Click on it one more time and finally there will another folder with the same name (spark-2.3.0-bin-hadoop2.7).

Example: **C:\Users\black\Desktop\Spark\spark-2.3.0-bin-hadoop2.7\spark-2.3.0-bin-hadoop2.7**

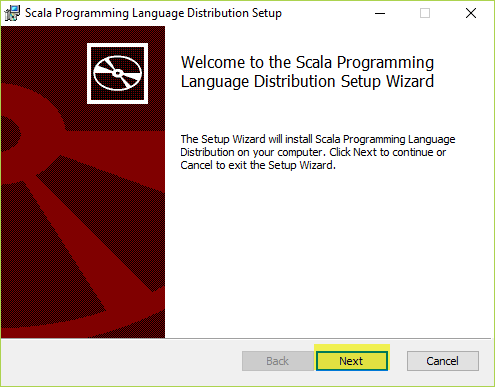
1. Rename the folder to “**Spark”** and move it to your desktop.
2. Inside the Spark folder, Go to the Conf folder and rename the log4j.properties.template to log4j.properties.
3. Open the properties and change log4j.rootCategory=INFO (line 19) to log4j.rootCategory=WARN

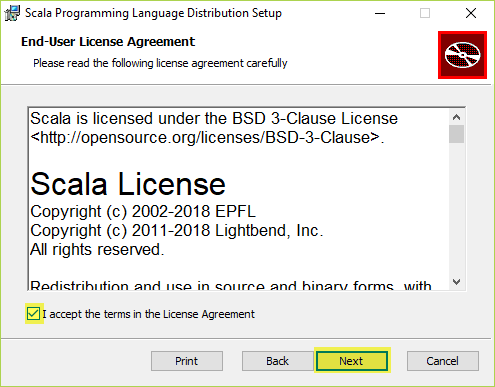


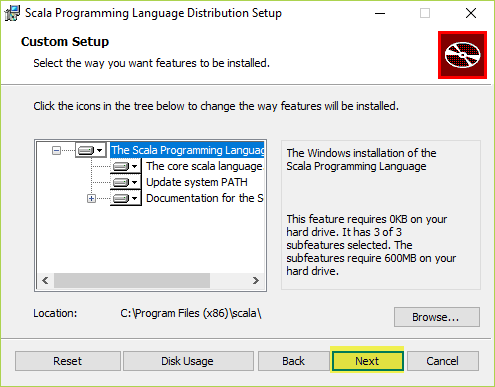
1. Go to Environment variables and create new System Variable. The new variable will have the name SPARK\_HOME and the value of the path of the spark folder.
2. Add the %SPARK\_HOME%\bin variable to PATH variable
3. Open a cmd window and enter **“spark-shell”**

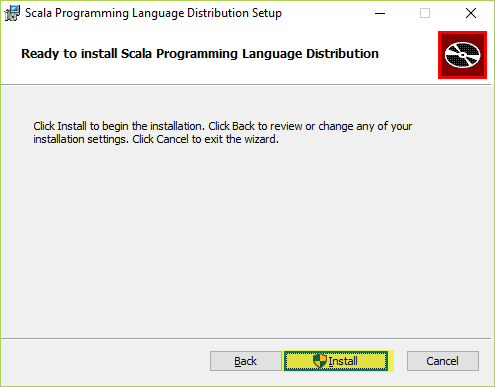
**Steps to install Scala and Scala IDE**

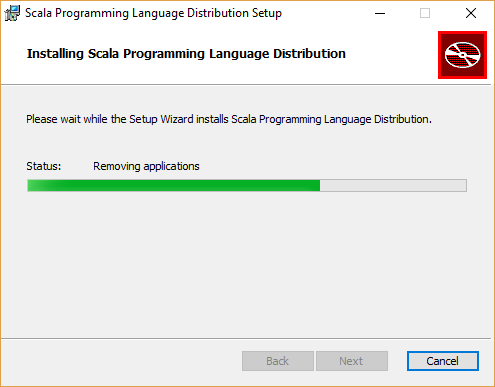
1. Go to scala-lang [Website](https://www.scala-lang.org/download/).
2. Go to Other ways to install Scala and click on [Download the Scala binaries for windows](https://downloads.lightbend.com/scala/2.12.5/scala-2.12.5.msi). This will download the scala-2.12.5.msi file.
3. Install Scala Follow the steps to install it.

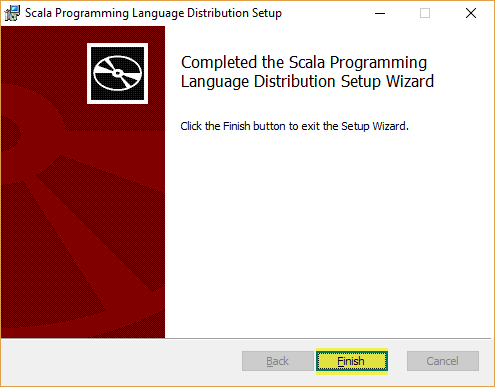




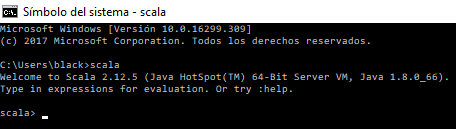








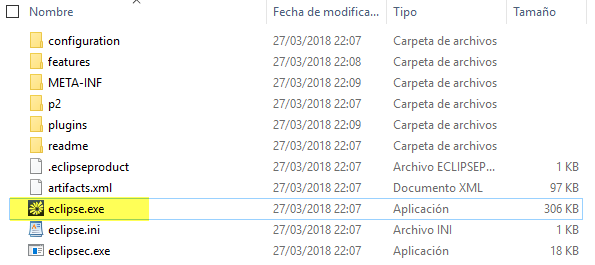
1. Open a cmd window and enter **“spark-shell”**



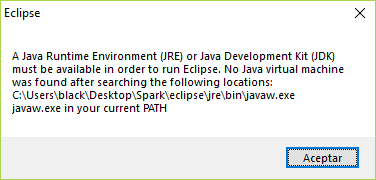
1. Go to Scala-ide [website](http://scala-ide.org/download/sdk.html).
2. Click on [Download IDE Windows - 64 bit](http://downloads.typesafe.com/scalaide-pack/4.7.0-vfinal-oxygen-212-20170929/scala-SDK-4.7.0-vfinal-2.12-win32.win32.x86_64.zip)



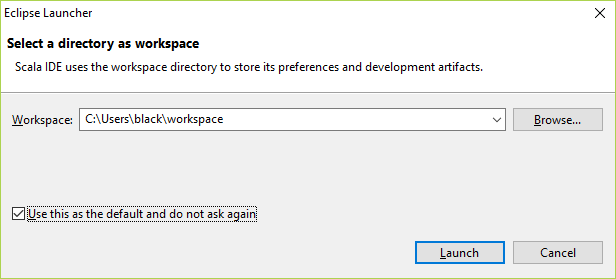
1. That will download a zip file. Unzip it and execute eclipse.exe



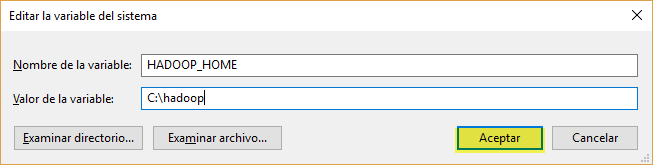
1. Add javaw.exe to your PATH system variable. Example: C:\Program Files\Java\jdk1.8.0\_66\bin. This will help to avoid the next error:



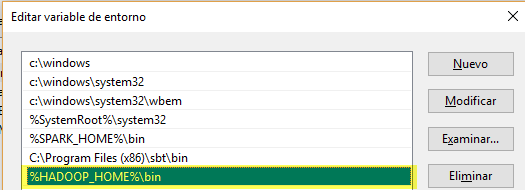
1. Select the workspace. Click Launch.



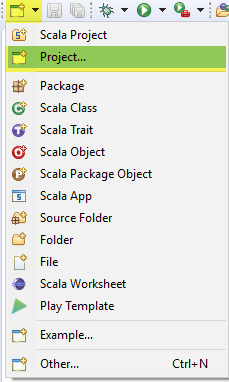
1. Create a new System Variable with name HADOOP\_HOME.



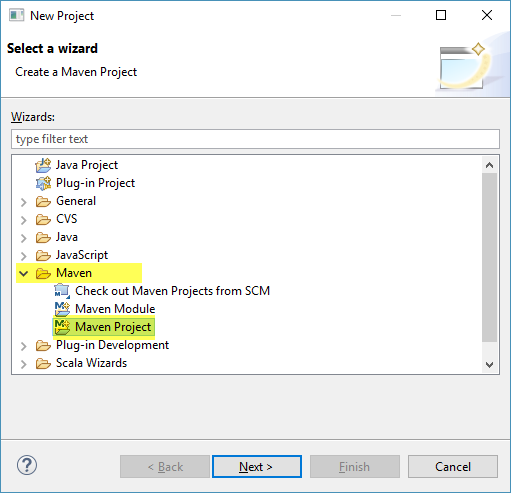
1. Add the HADOOP\_HOME variable to the PATH System Variable. %HADOOP\_HOME\bin%



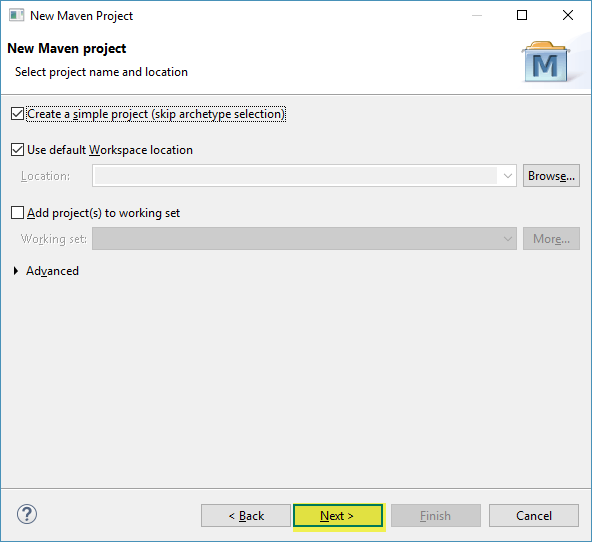
1. Create a new Maven project. Click in new and the select Project.



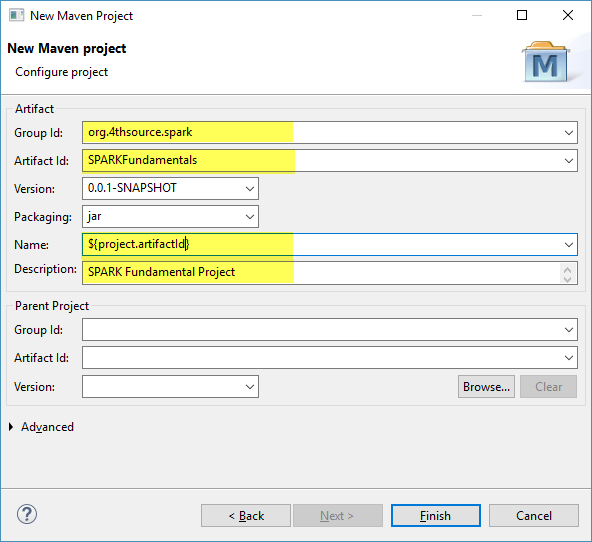
1. Go to Maven folder and select Maven Project.



1. Select “Create a simple project (skip archetype selection)” and click Next.



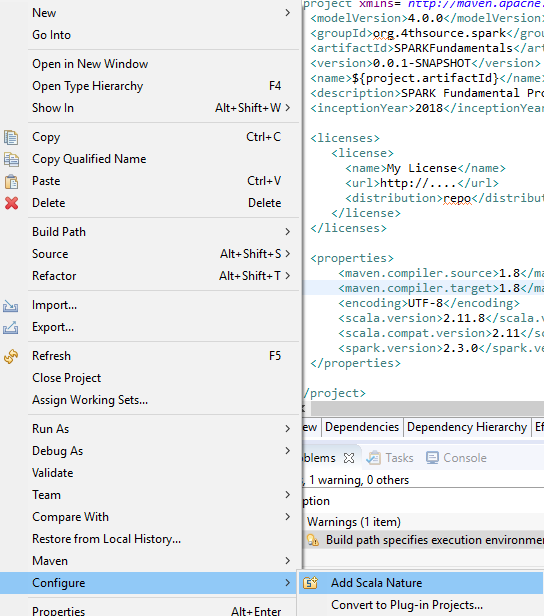
1. Enter the next field. Anything else leaves it blank. Click Finish.

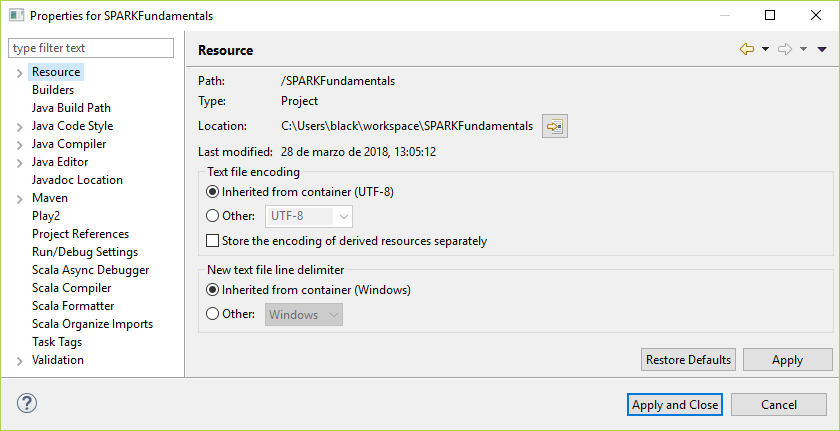


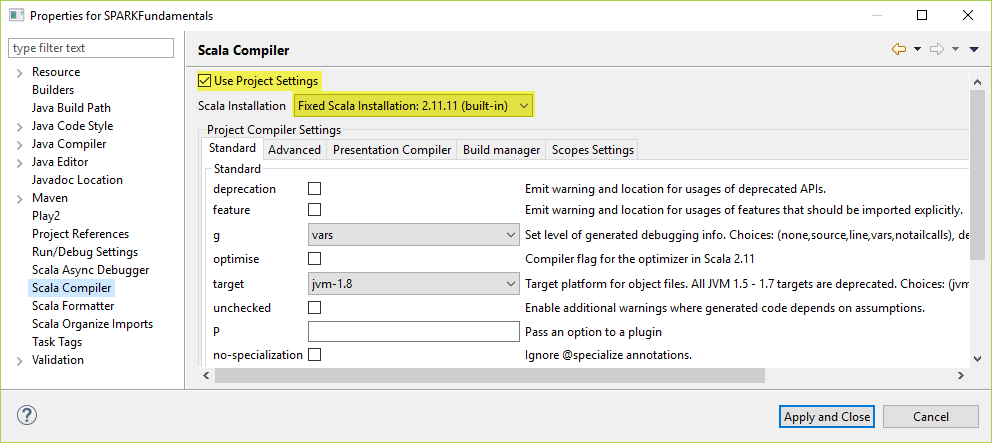
1. Overwrite the pom.xml with this pom.xml. Before doing that, make sure to put the GroupId and ArtifactID that you wrote in step 15.



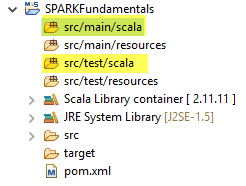
1. Next, we need to add a Scala Nature to our project. Right click under your project, go to Configure and click on Add Scala Nature.



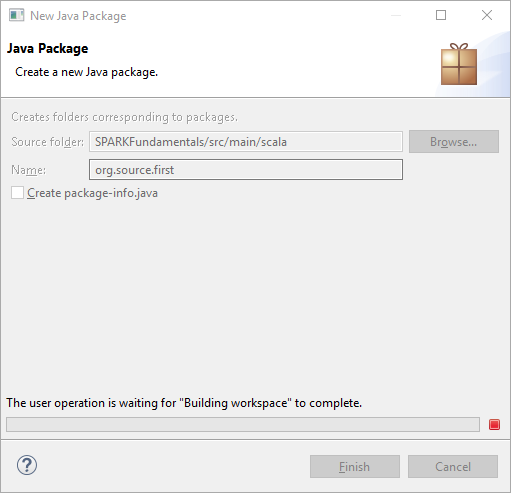
1. To select the Scala Compiler. Right click on the project and go to properties.
2. Select Scala Compiler; select “Use Project Settings”. Finally select the correct version of Scala. In this case is 2.11.11. Apply the changes and then click OK.



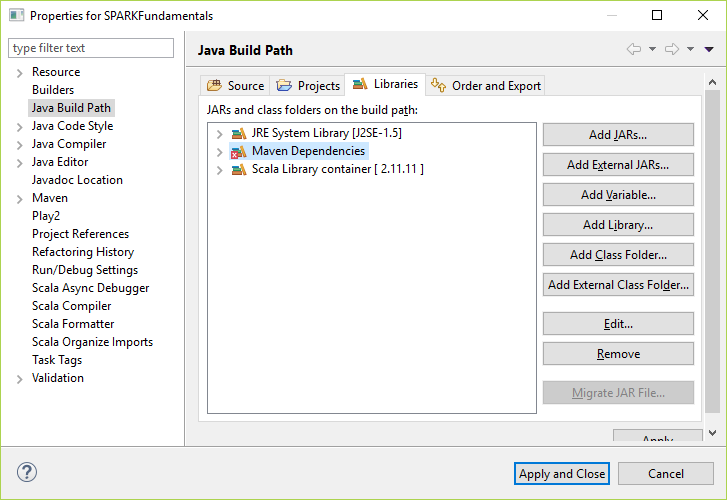
1. Change the src/main/java and src/test/java to src/main/scala and src/test/scala. Right click on each, select Refactor option, change the name and click OK.



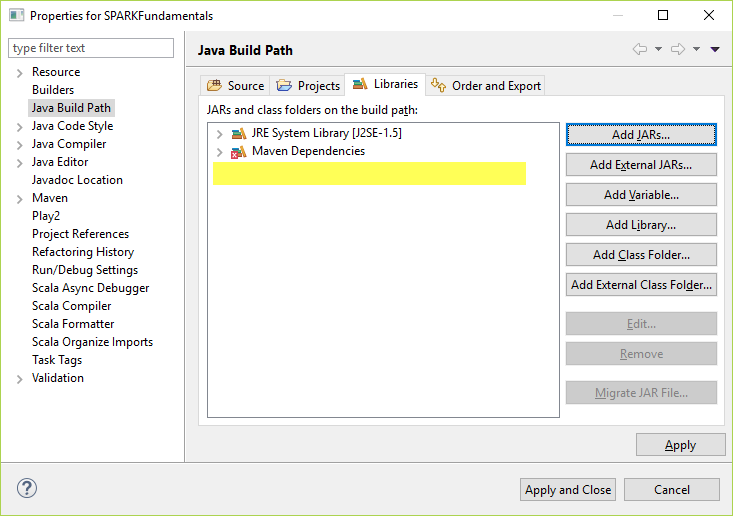
1. Now create a new package under scr/main/scala. Right click under the folder, select New. Finally click on Package. Enter the name of the package and click Finish.



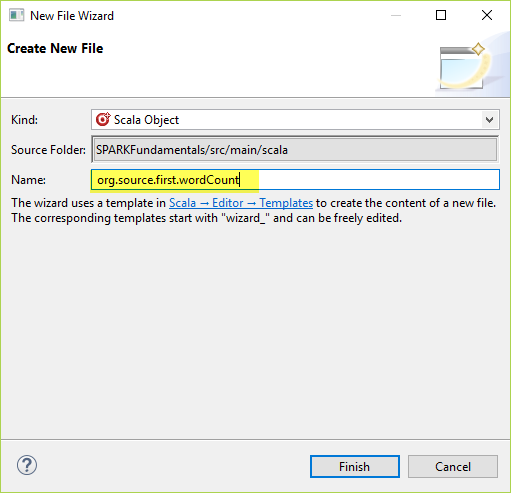
1. Next, go to Configure Build Path. Right click on your project, then Build Path. Finally, Configure Build Path.



1. Go to Libraries and Remove Scala Library container [2.11.11]. Apply the changes and click OK.



1. Create a new Scala Object under your new package. Right click under it and select New. Finally click on Scala Object. Enter the name of the Scala Object and click Finish.



1. Copy the file below to your project.



1. Write your first Spark-Scala code

**package** org.source.first

**import** org.apache.spark.\_

**import** org.apache.spark.SparkContext.\_

**import** org.apache.spark.sql.SparkSession

**object** wordCount {

**val** spark = SparkSession.builder()

.appName("Wordcount")

.config("spark.master","local")

.getOrCreate()

spark.sparkContext.setLogLevel("ERROR")

**def** main(args: Array[String]) = {

// val conf = new SparkConf()

// .setAppName("wordCount")

// .setMaster("local")

//

// val sc = new SparkContext(conf)

**val** textrdd = spark.sparkContext.textFile("first\_scala.txt")

**val** newRDD = textrdd.flatMap(line => line.split(" ")).map(word => (word,1)).reduceByKey(\_ + \_).saveAsTextFile("food.count.txt")

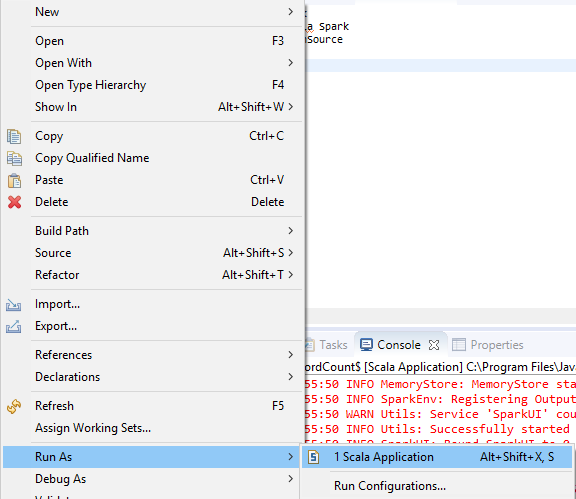
//.take(3)

//map(word => (word,1)).reduceByKey(\_ + \_).saveAsTextFile("food.count.txt")

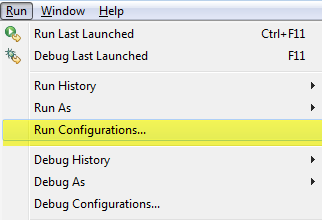
}

}

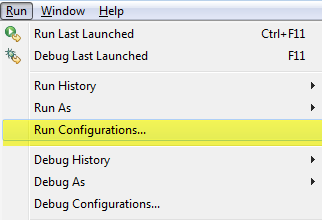
1. Right click on wordCount.scala, go to Run As and click on 1 Scala Application.



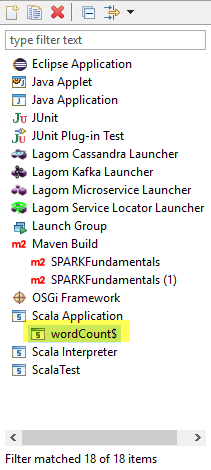
1. asdasas



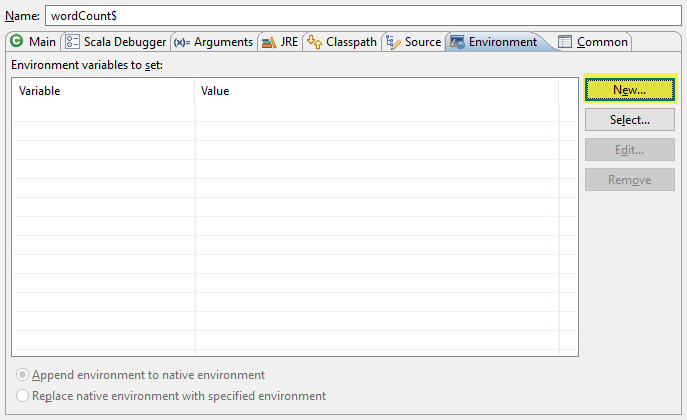
1. If there is a related error to winutil. Follow the next steps.
   1. Go to Run and click on Run Configurations.



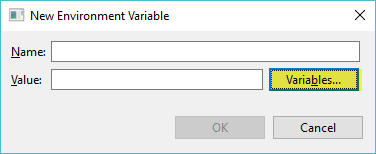
* 1. Go to Scala Application-> wordCount$



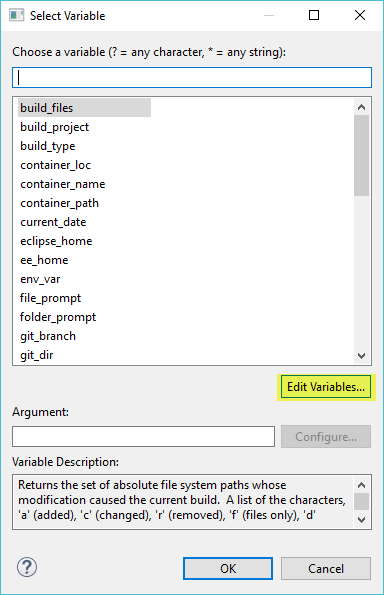
* 1. Go to Environment and create a new variable. Click on New.



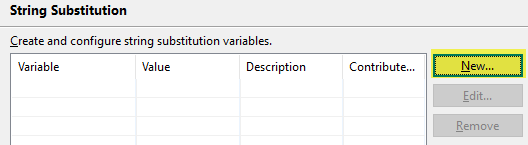
* 1. Click on the Variables button.



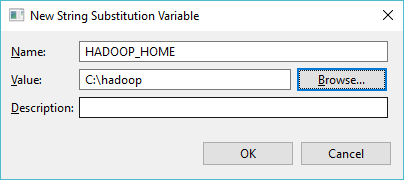
* 1. If this is the first time here, you need to click on Edit Variables. If not, the select the HADOOP\_HOME variable and go to the final step.



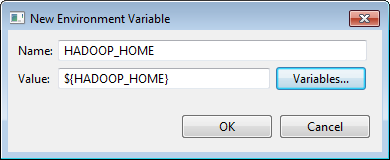
* 1. We are going to create a new String Substitution. Click on New button.



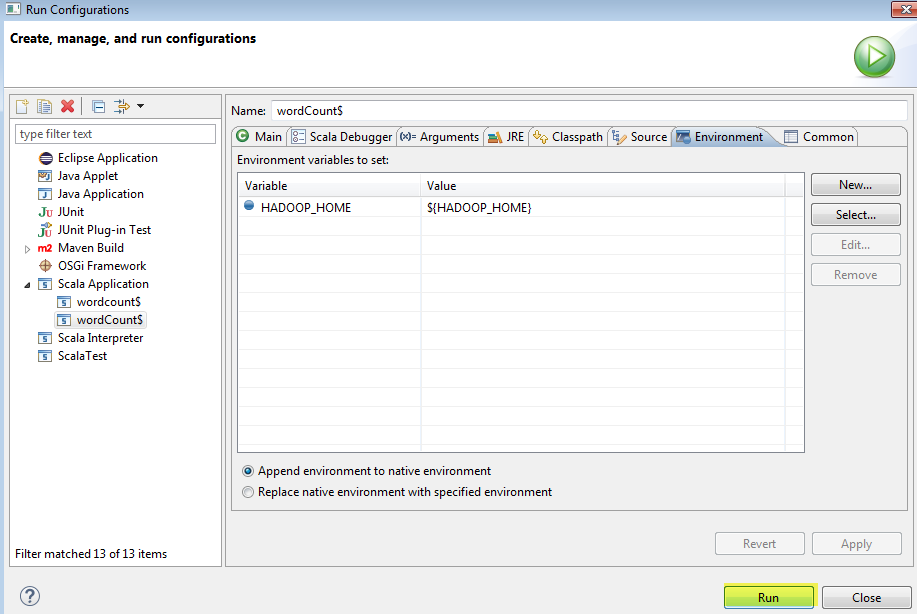
* 1. Enter the name of the new String Substitution Variable, browse for the bin folder. The bin folder has the winutil.exe. After that, click OK.



* 1. Click the button OK that is in the String Substitution. Then Select the new HADOOP\_HOME variable and Click OK.



* 1. Now, the assigned HADOOP\_HOME variable is in our project. Click Apply and then click Run. This will generate a folder with the name you provided in your Scala code.



* 1. Go to the folder, if everything was correct, you will see two files. One with the name \_SUCCESS and the second with the name part-00000. Open the second file.

