**Word Count Execution in Spark using sbt**

Code for word count problem:

import org.apache.spark.SparkConf

import org.apache.spark.SparkContext

import org.apache.spark.rdd.RDD.rddToPairRDDFunctions

object WordCount {

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

val conf = new SparkConf()

.setAppName("WordCount")

val sc = new SparkContext(conf)

val test = sc.textFile("hdfs:///wc.txt")

test.flatMap( line => line.split(" "))

.map( word => (word, 1) )

.reduceByKey(\_ + \_ )

.saveAsTextFile("hdfs:///sampleOp")

sc.stop

}

}

Please follow the below steps:

**Step 1 :** Go to Terminal, and in the desktop, create a folder with the project name **wordpro** using the below commands:

**cd Desktop**

**mkdir wordcount**

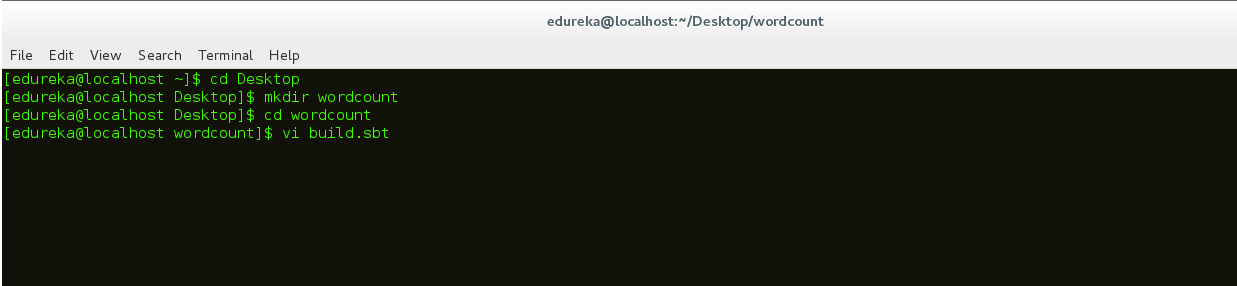
**Step 2:** Now, inside the wordpro, we will create a file build.sbt and add the dependencies in it.

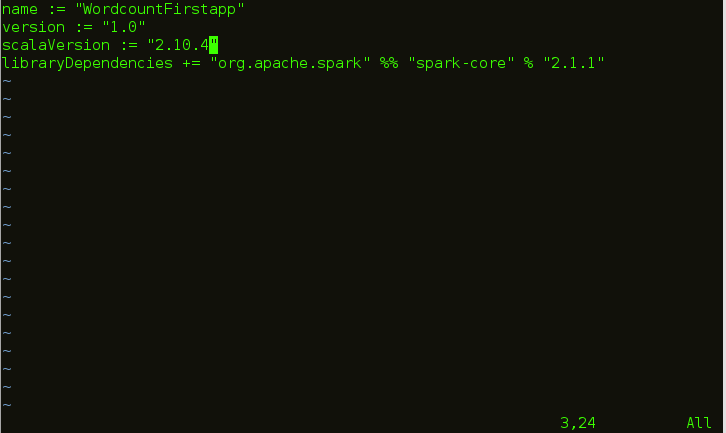
**cd wordcount**

**vi build.sbt**

Then we will add the dependencies in the build.sbt file.

Refer to the below screenshot:





Now, we will save the build.sbt file using the below command:

Press **esc** button and then type **:wq**

**Step 3 :** Then we will create another folder inside wordcount named **src** by using the below command **:**

**mkdir src**

**Step 4:** Inside **src,** we will create another folder **main** using the below command:

**cd src**

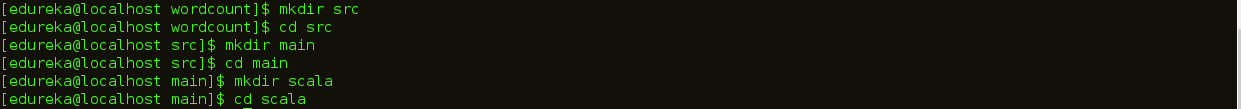
**mkdir main**

**Step 5:** Inside **main,** we will create another folder **scala** using the below command:

**cd main**

**mkdir scala**

Please refer to the below screenshot:

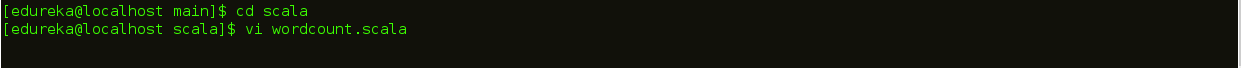


**Step 6 :** Inside scala, we need to add the wordpro.scala file which consists of the code for word count

**cd scala**

**vi wordcount.scala**

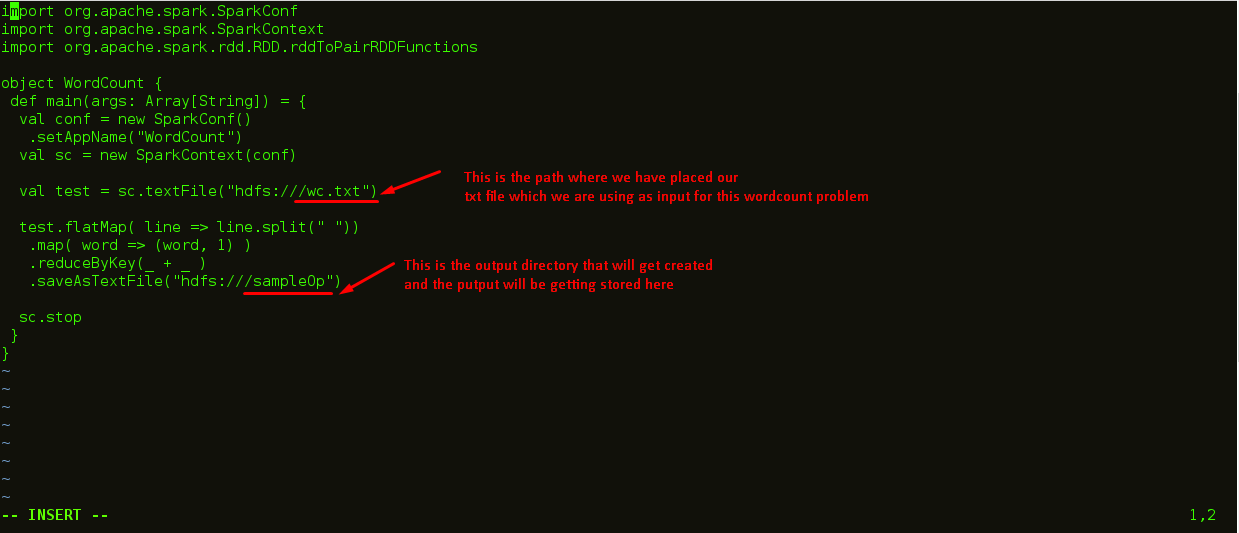
Please refer to the below screenshot



Now we need to give the code for wordcount.

The code has already been provided at the top of this documentation.

Please refer to the below screenshot:



Now, save the code using the below,

Press **esc** button and type **:wq** and press **enter**

**Step 7 :** After completion of creation of project structure, we execute the below command from the project folder using the below command:

**cd .**.

**cd .**.

**cd .**.

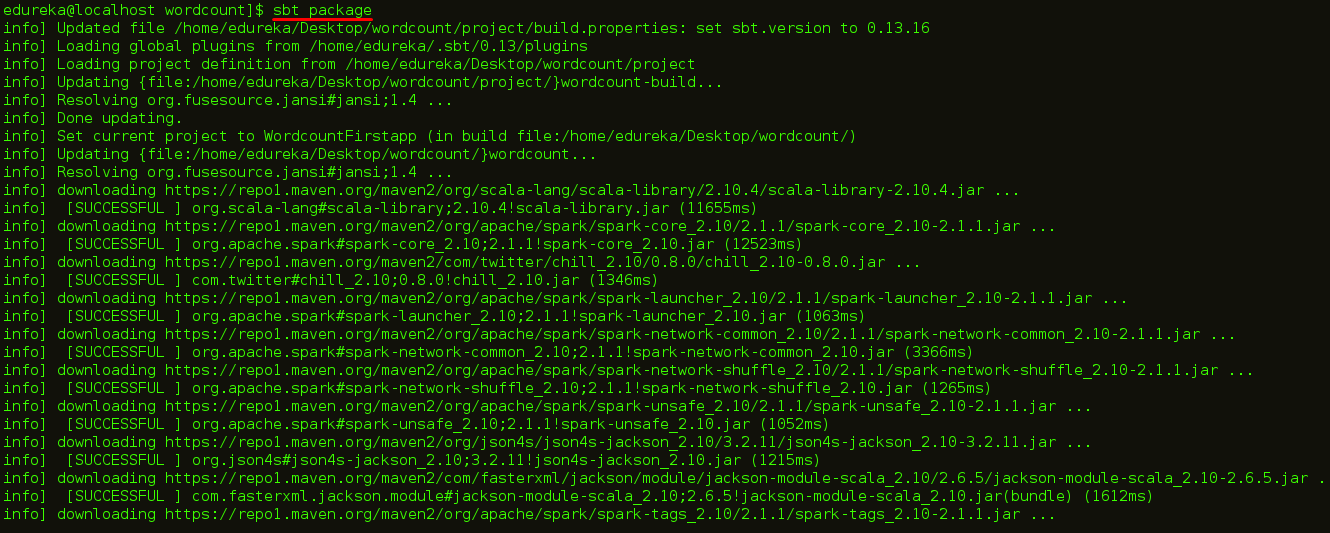
The above commands will take us to the **wordcount** folder

Then, we give the below command:

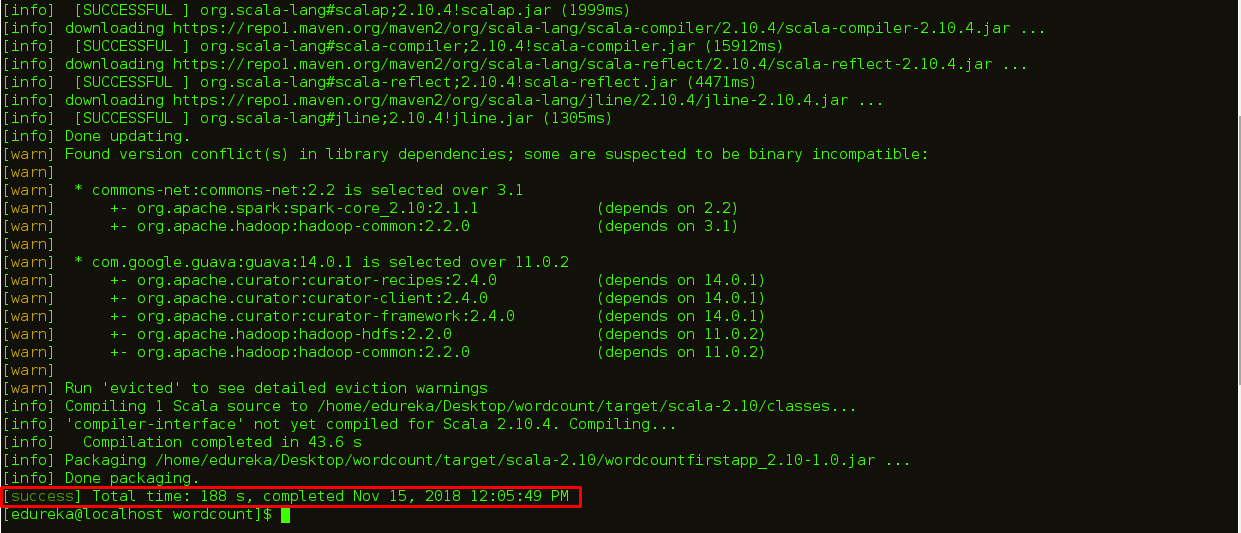
**sbt package**

Please refer to the screenshot below for clarity:



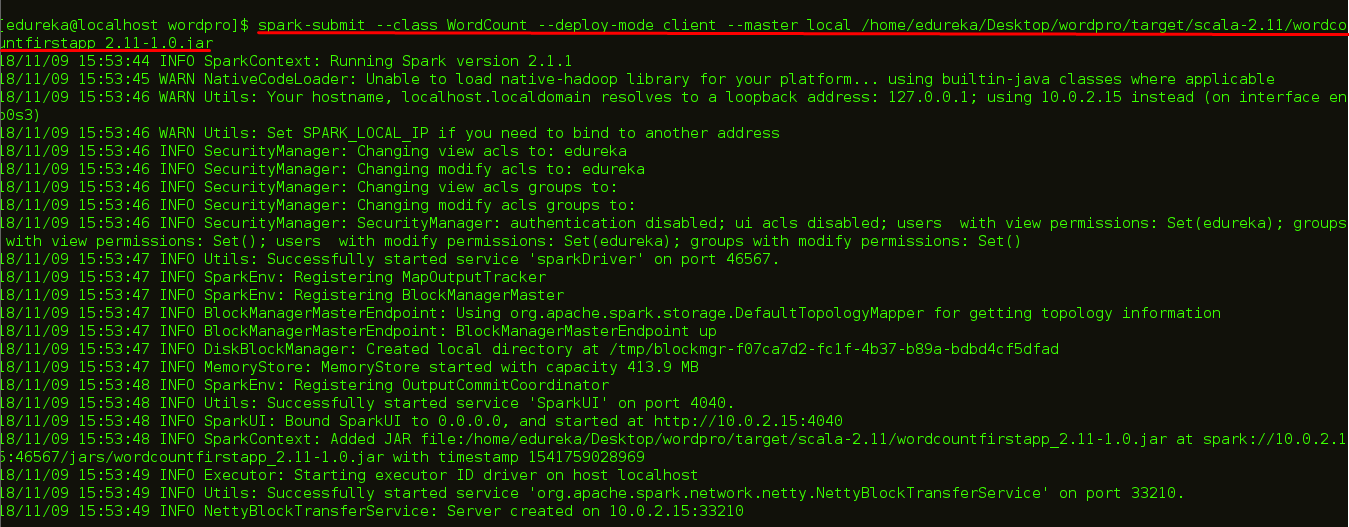


**Step 8 :** After execution of the sbt package is successful, we will be getting the below success message:



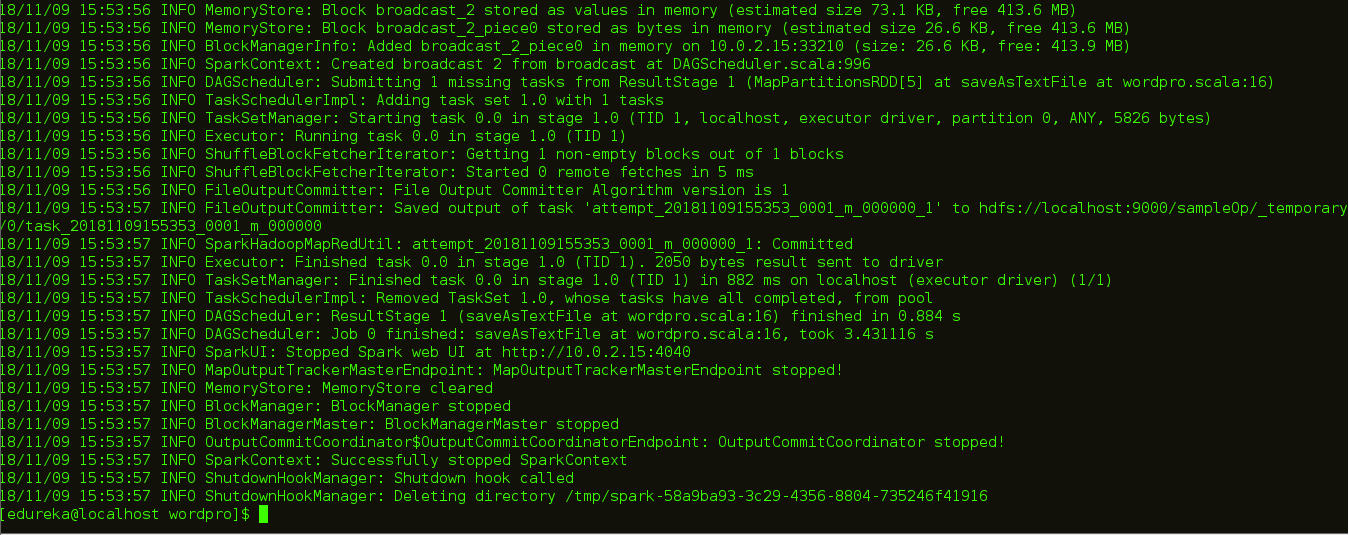
**Step 9 :** Now we will execute the below spark-submit command.

Please refer to the below screenshot:



**Step 5 :** Now , we see below that the execution was successful.

Please refer below:



**Step 5 :** Now we can view the output in the following manner.

Please refer to the below screenshot:

