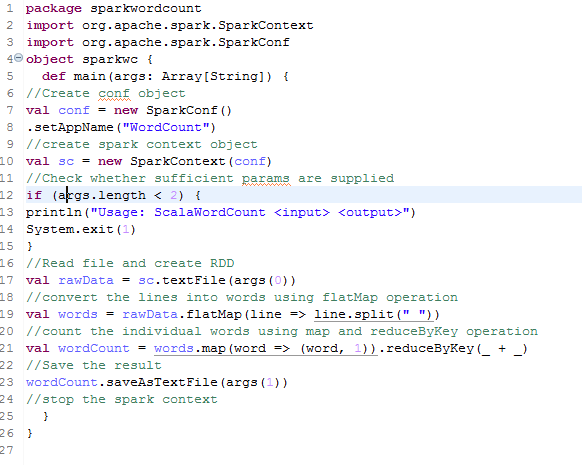
Spark Submit:

We are creating the Maven project in Eclipse.

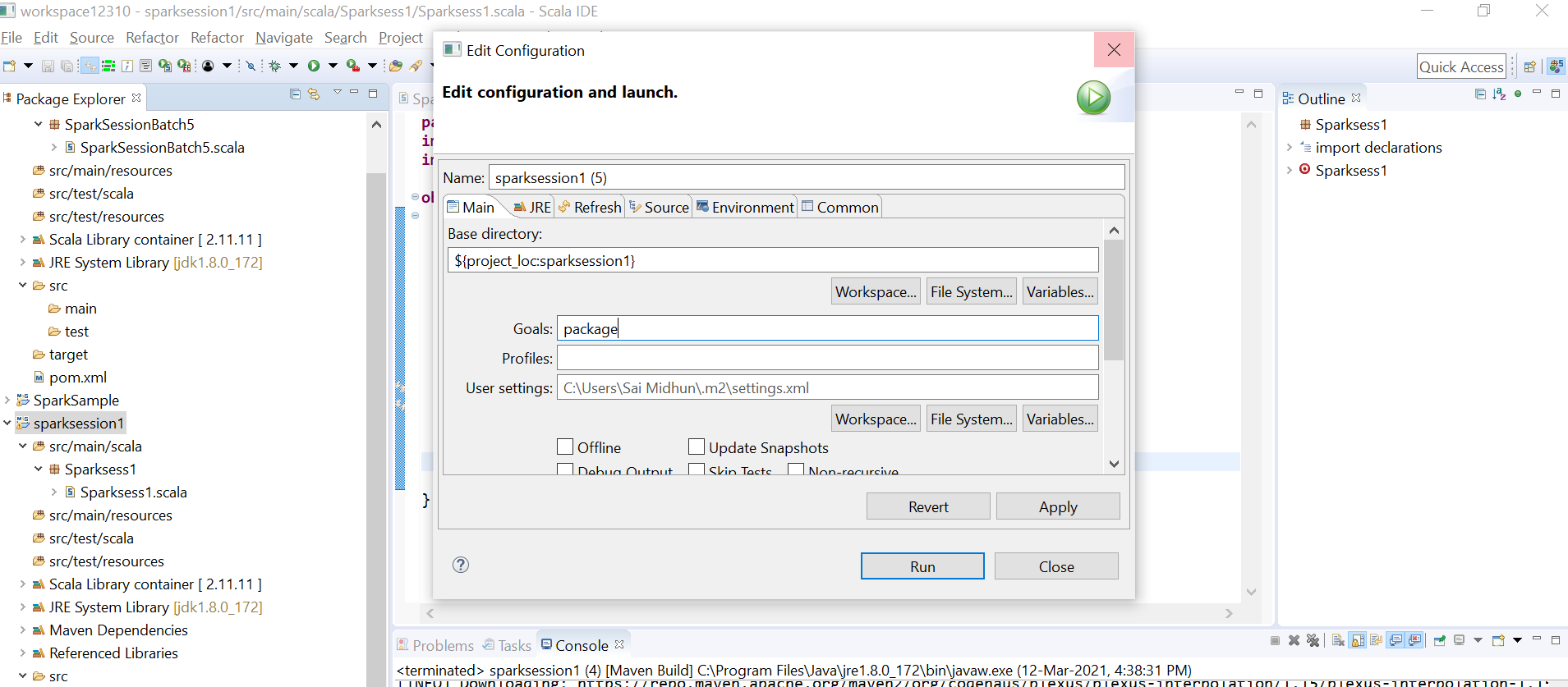
Normal Procedure .



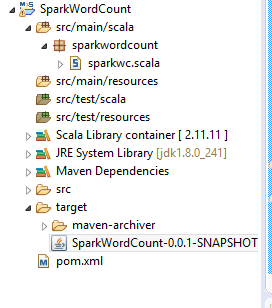
Now we need to save it and run as Maven Build.

Don’t do maven clean , strait away convert the code to jar and do the deployment .

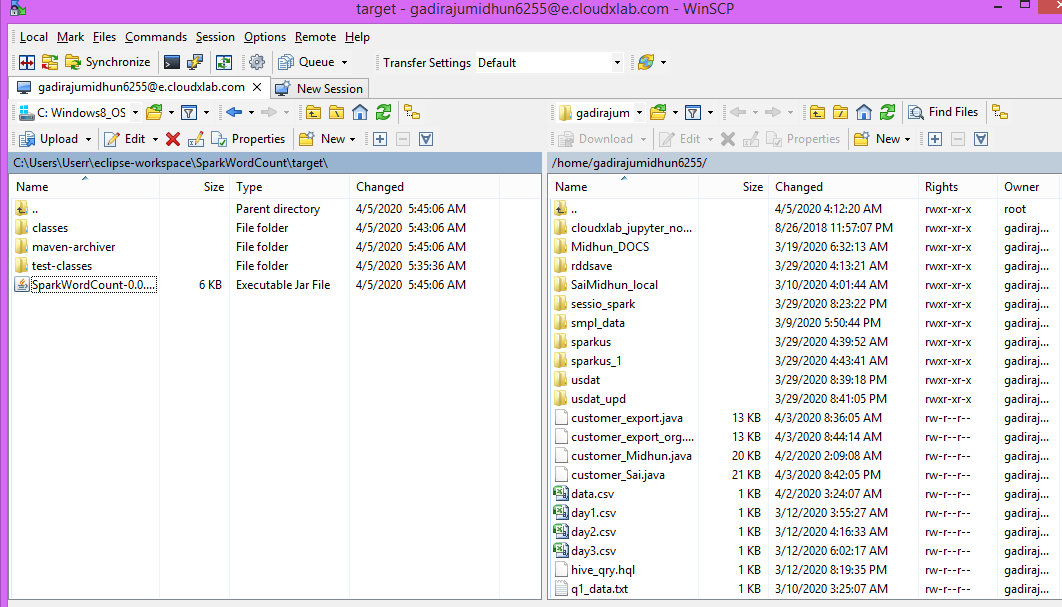
In the maven build , give package to covert your code to jar



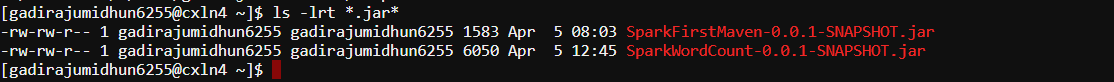
Then the jar will be created in target as shown below .



Now we need to move the jar to cluster using winscp .



Now jar is moved here .



Now we need to do the spark submit .

Since it’s a cluster , we need to place the input file in cluster .



Placing wordcount file in hdfs.

Now execute the spark submit as shown below .



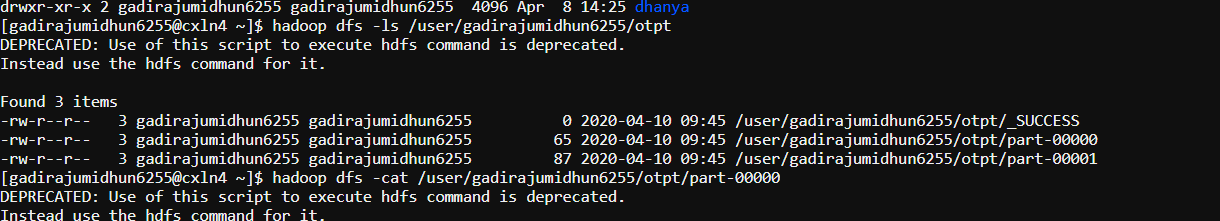
Spark Submit requires 3 inputs

1. Jar file
2. Input path where the input file exists
3. Output where the data has to be stored.

Now see the output .

**Deploying in YARN mode (Client) :**

[gadirajumidhun6255@cxln4 ~]$ spark-submit --class sparkwordcount.sparkwc --master yarn --deploy-mode client SparkWordCount-0.0.1-SNAPSHOT.jar /user/gadirajumidhun6255/wc/wordcount/ /user/gadirajumidhun6255/otpt



**Cluster Mode :**

[gadirajumidhun6255@cxln4 ~]$ spark-submit --class sparkwordcount.sparkwc --master yarn --deploy-mode cluster SparkWordCount-0.0.1-SNAPSHOT.jar /user/gadirajumidhun6255/wc/wordcount/ /user/gadirajumidhun6255/otpt\_cluster

Output:

