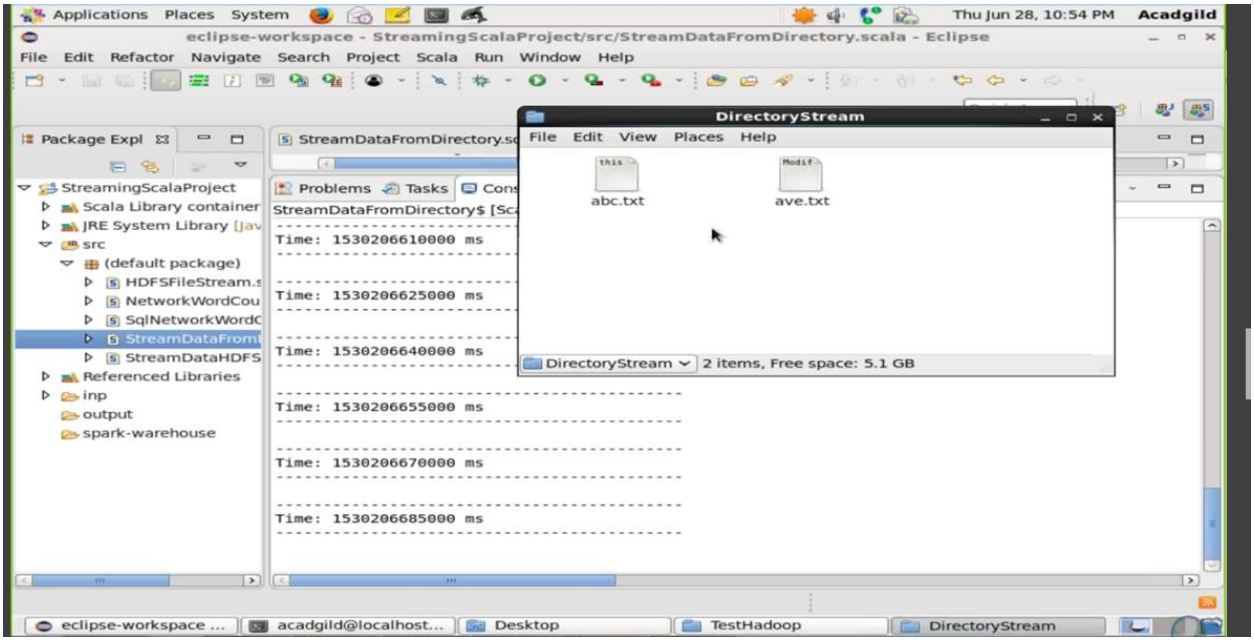
There are two parts this case study

- **First Part -** You have to create a Spark Application which streams data from a file on local directory on your machine and does the word count on the fly. The word should be done by the spark application in such a way that as soon as you drop the file in your local directory, your spark application should immediately do the word count for you.

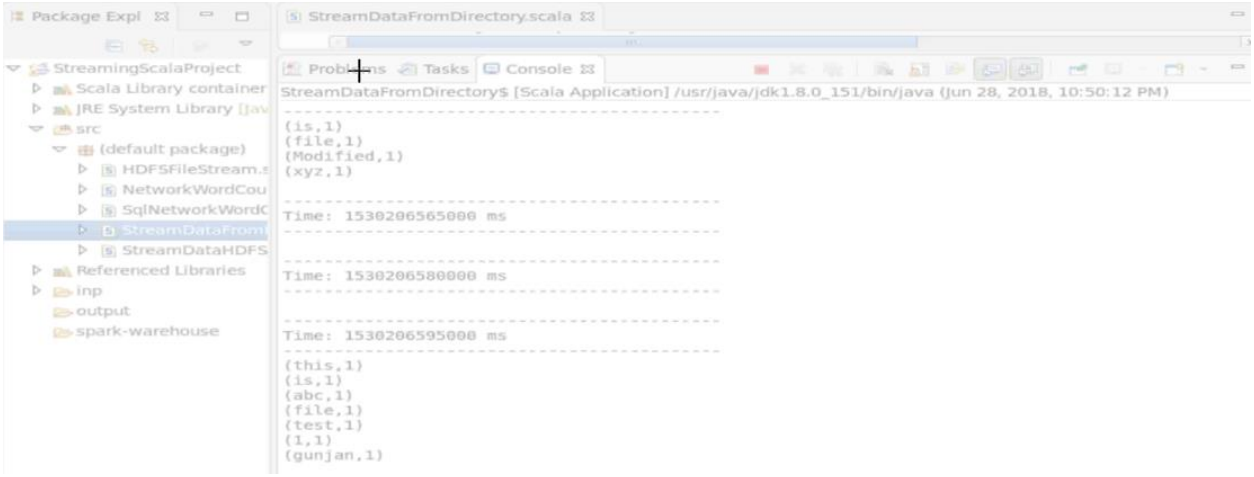
Second Part - In this part, you will have to create a Spark Application which should do the following

1. Pick up a file from the local directory and do the word count
2. Then in the same Spark Application, write the code to put the same file on HDFS.
3. Then in same Spark Application, do the word count of the file copied on HDFS in step 2
4. Lastly, compare the word count of step 1 and 2. Both should match, other throw an error

Task 1 :



These files putting to get on the fly count:


Package Expl
StreamingScalaProject
Scala Library container
JRE System Library
src
(default package)
HDFSFileStream.s
NetworkWordCou
SqlNetworkWordC
StreamDataFrom
StreamDataHDFS
Referenced Libraries
inp
output
spark-warehouse
StreamDataFromDirectory.scala
Problems
Tasks
Console
StreamDataFromDirectory$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jun 28, 2018, 10:50:12 PM)
(is,1)
(file,1)
(Modified,1)
(xyz,1)
Time: 1530206565000 ms
Time: 1530206580000 ms
Time: 1530206595000 ms
(this,1)
(is,1)
(abc,1)
(file,1)
(test,1)
(1,1)
(gunjan,1)