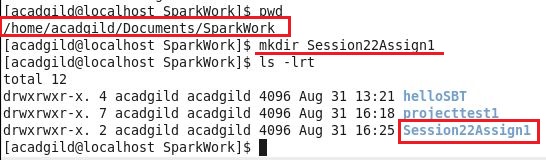
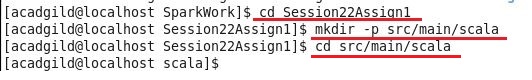
**Spark Streaming using TCP Socket**

**Step 1: Create Project Directory “Session22Assign1” inside /home/acadgild/Documents/SparkWork/**

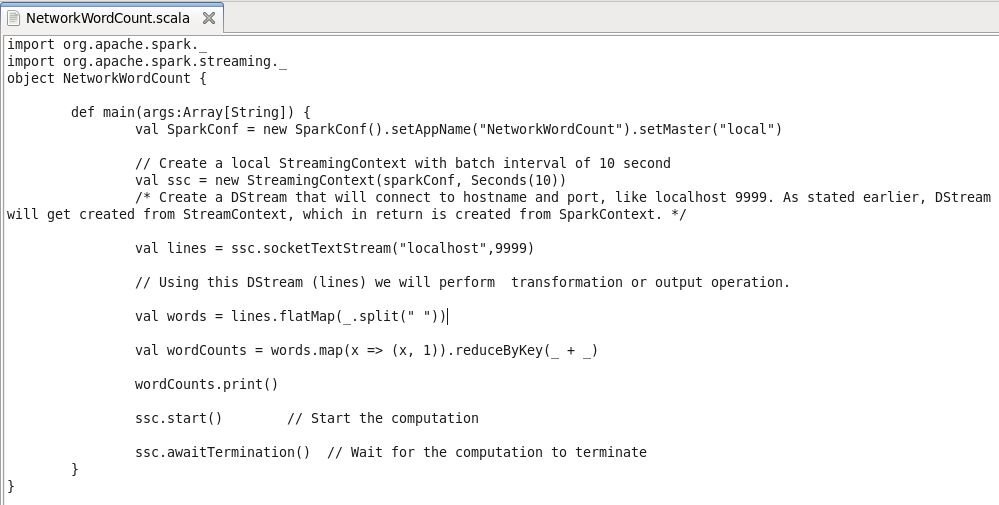
****

**Step 2: Change to project directory “Session22Assign1” and create “src/main/scala” directory structure inside it, after that, change to “scala” directory and create “NetworkWordCount.scala” file inside it which contains code to be run:**

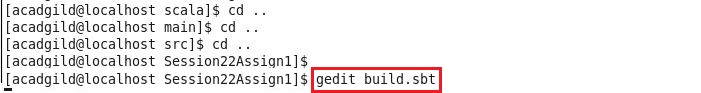
****

**G:\ACADGILD\course material\Hadoop\Sessions\Session 22\Assignments\Assignment1\Screenshots\2.1.JPG**

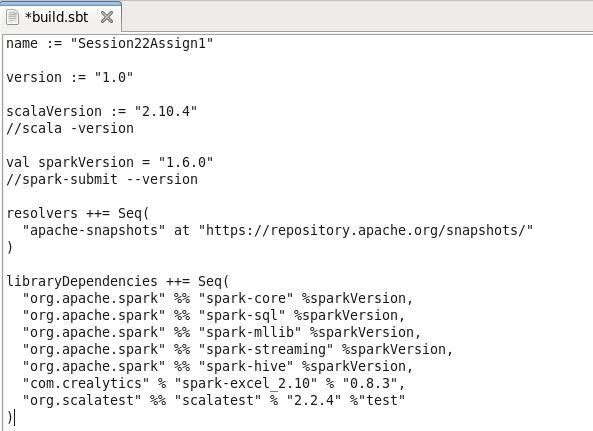
**Step 3: Write below contents inside “NetworkWordCount.scala” file**

****

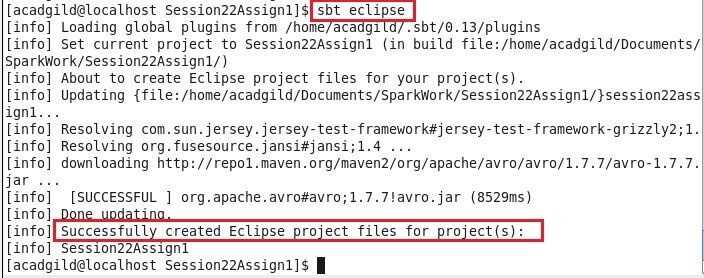
**Step 5: Now change to main project directory “Session22Assign1” to create build.sbt file inside it**

****

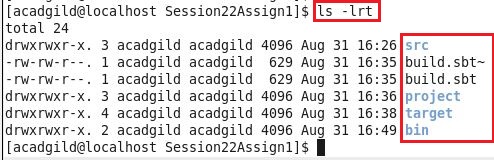
**Write following contents inside build.sbt**

****

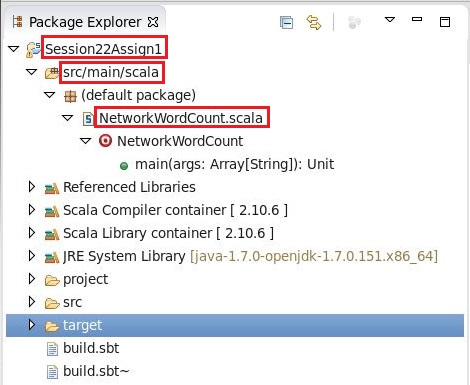
**Step 6: Run “sbt eclipse” command inside “Session22Assign1” project directory**

****

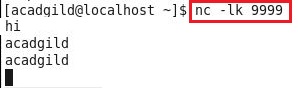
**Step 7: We can see the following directory structure inside “Session22Assign1” project directory after firing “sbt eclipse” command**

****

**Step 8: Now import the project inside eclipse, after import, following directory structure will be shown in eclipse**

****

**Step 9: Open new terminal and type “nc –lk 9999” command to run “netcat” as a data server, after that, type few words**

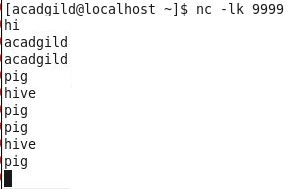
****

**This terminal acts as a server where words are fed continuously, and our Spark Streaming code counts the number of occurrences (in a batch interval of 10 sec).**

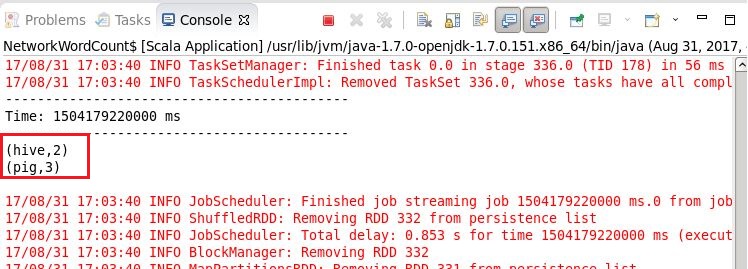
**Step 10: In eclipse, where project is imported, we can check the output in console,**

****

**Step 11: Again, type some words in data server terminal**

****

**Step 12: In eclipse, check the output in console,**

****

**As the interval has been set at 10 sec, that’s why output is captured like above.**