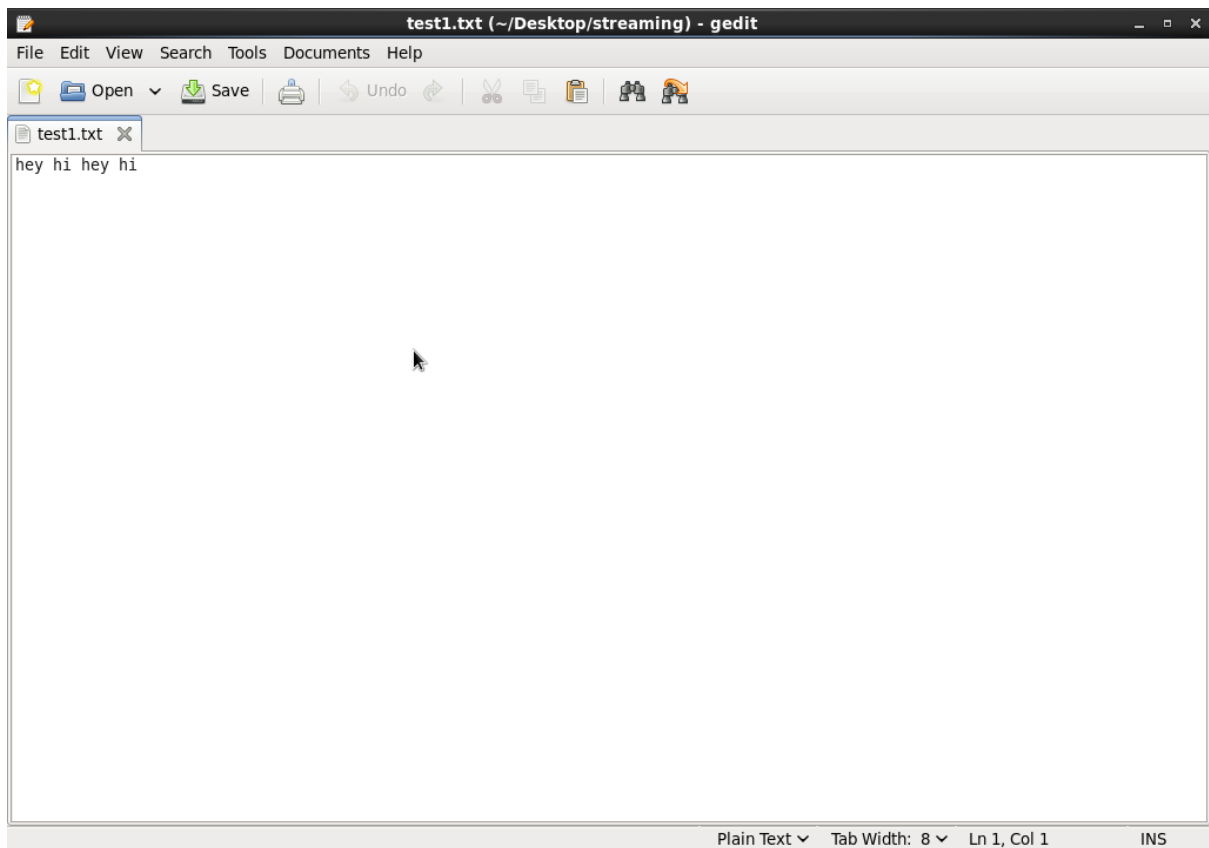


Case Study Spark Streaming.

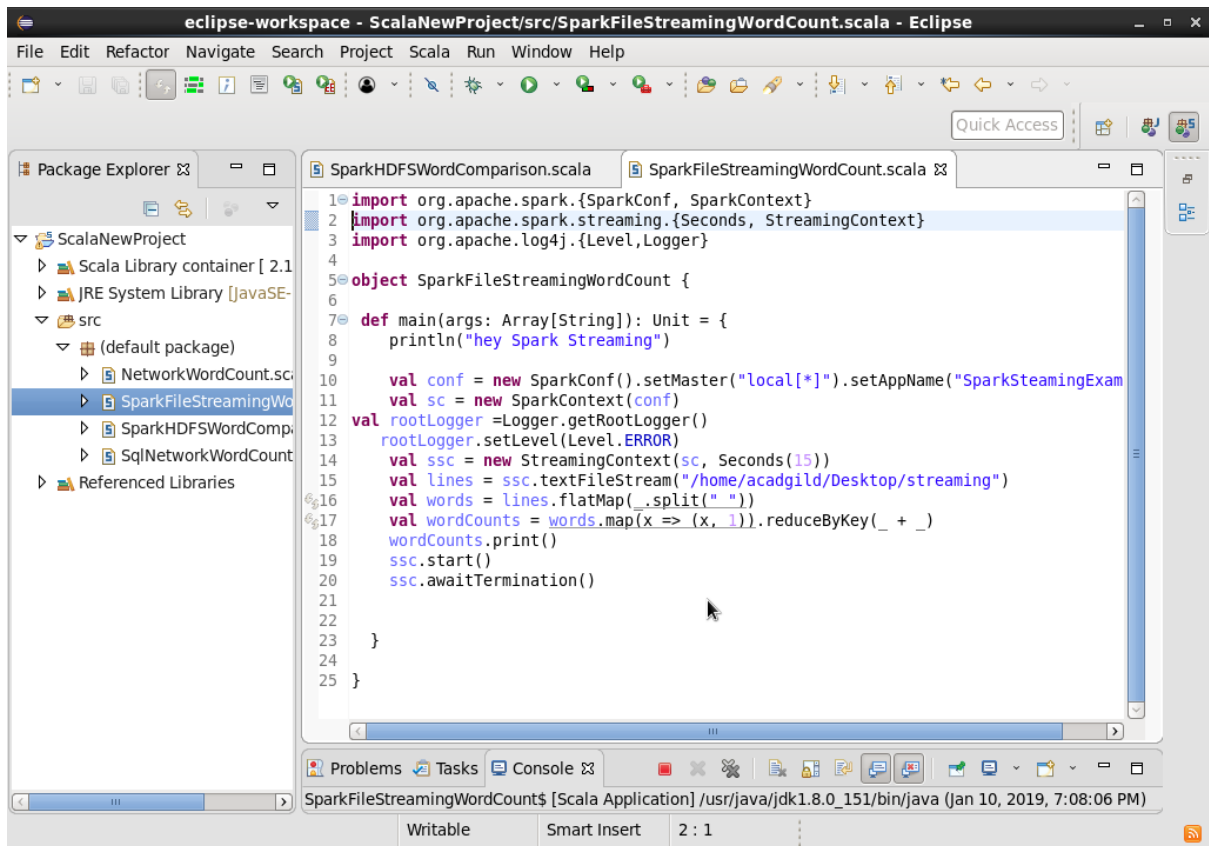
There are two parts this case study

- **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.

Solution: First see the txt file.



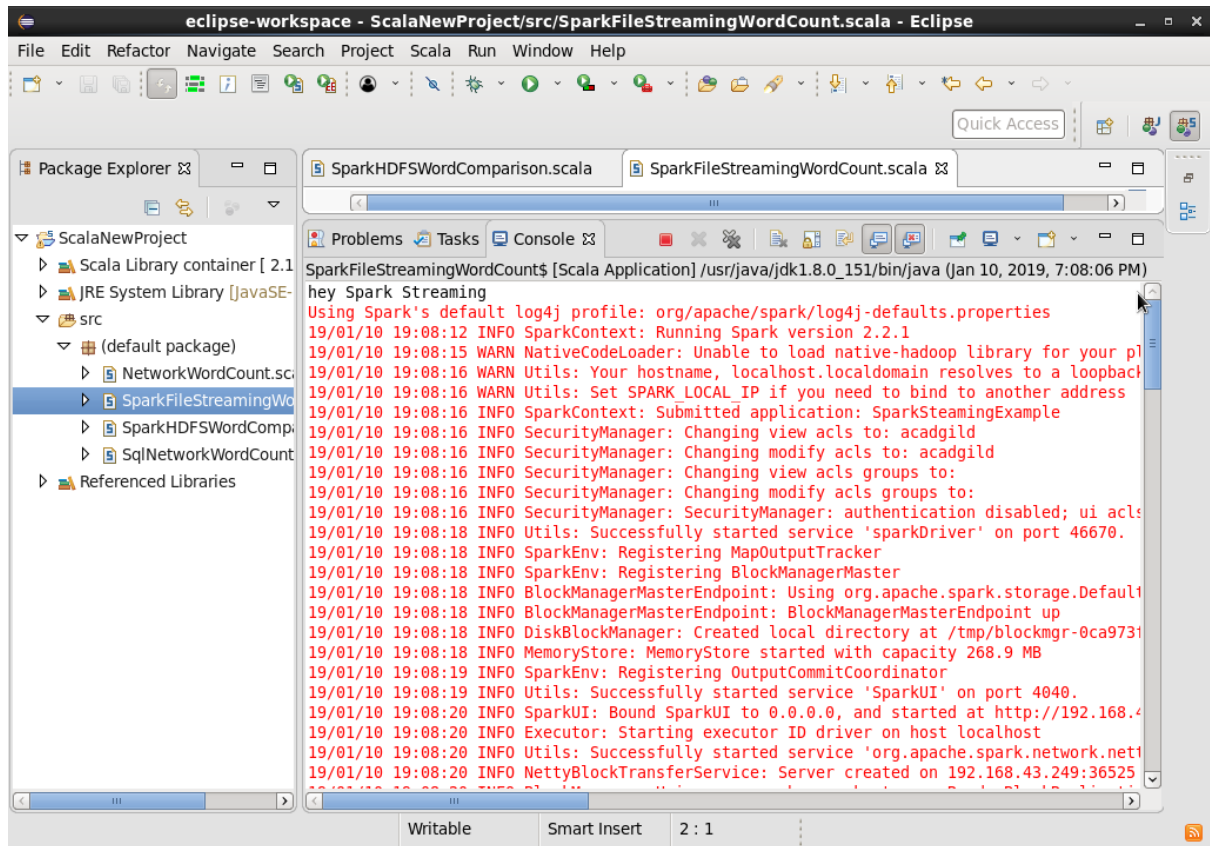
Code:



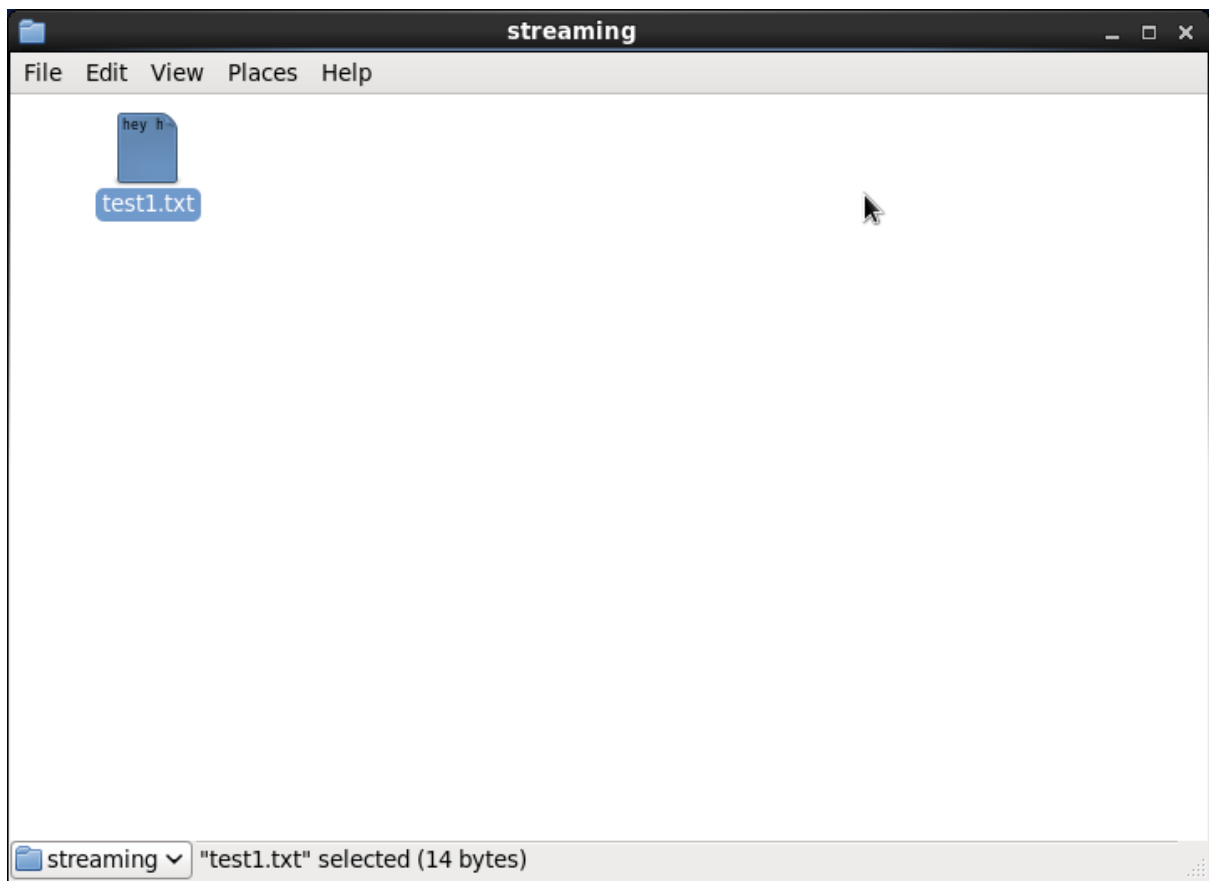
Directory:



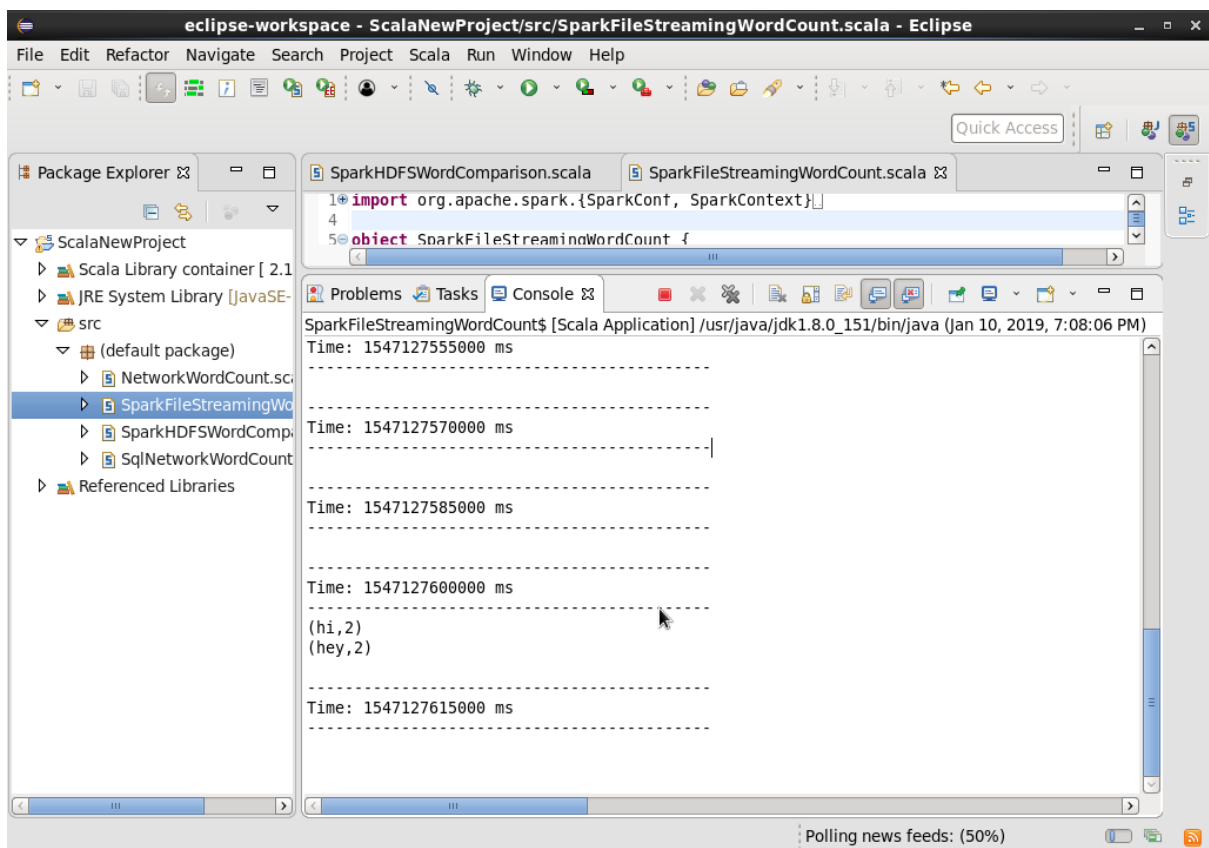
Code execution:



Put the file in directory

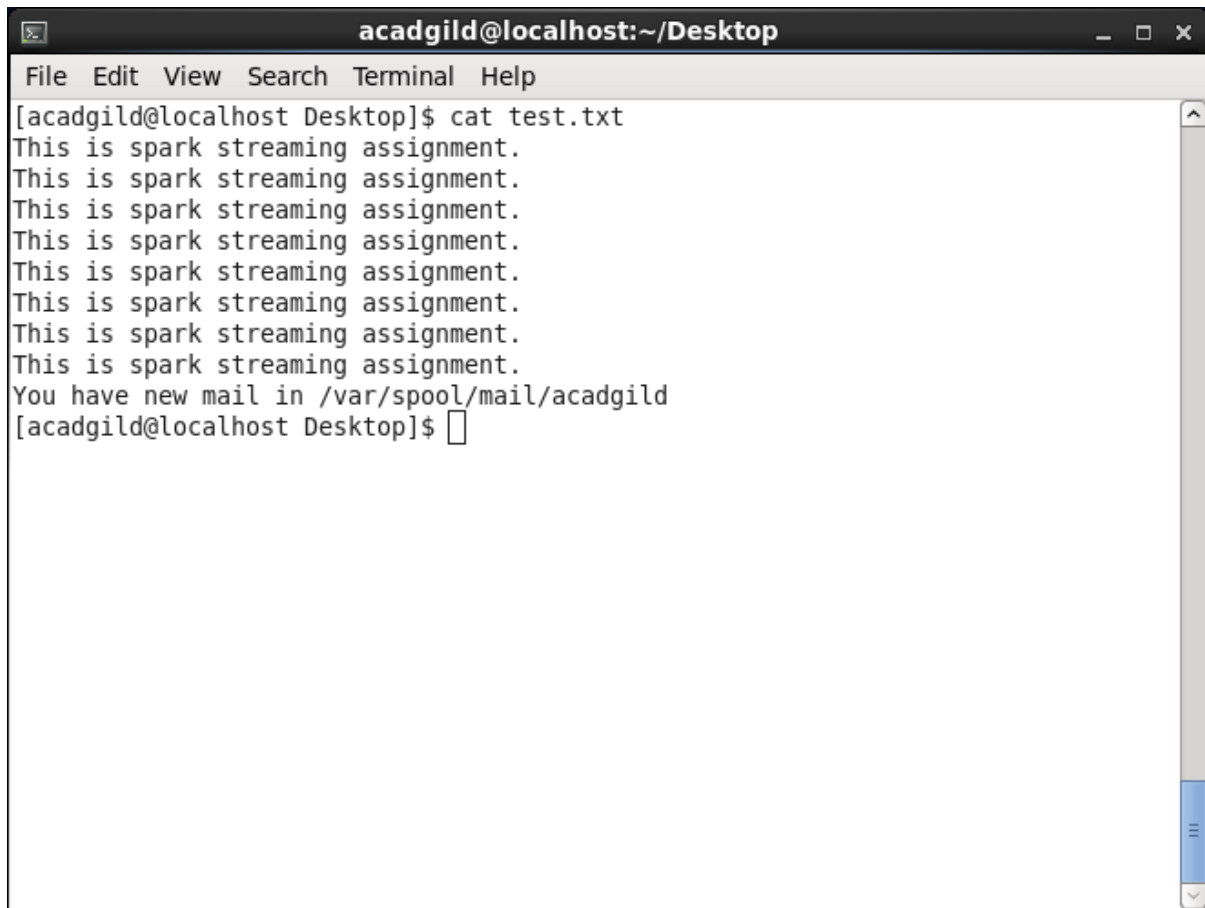


Output:



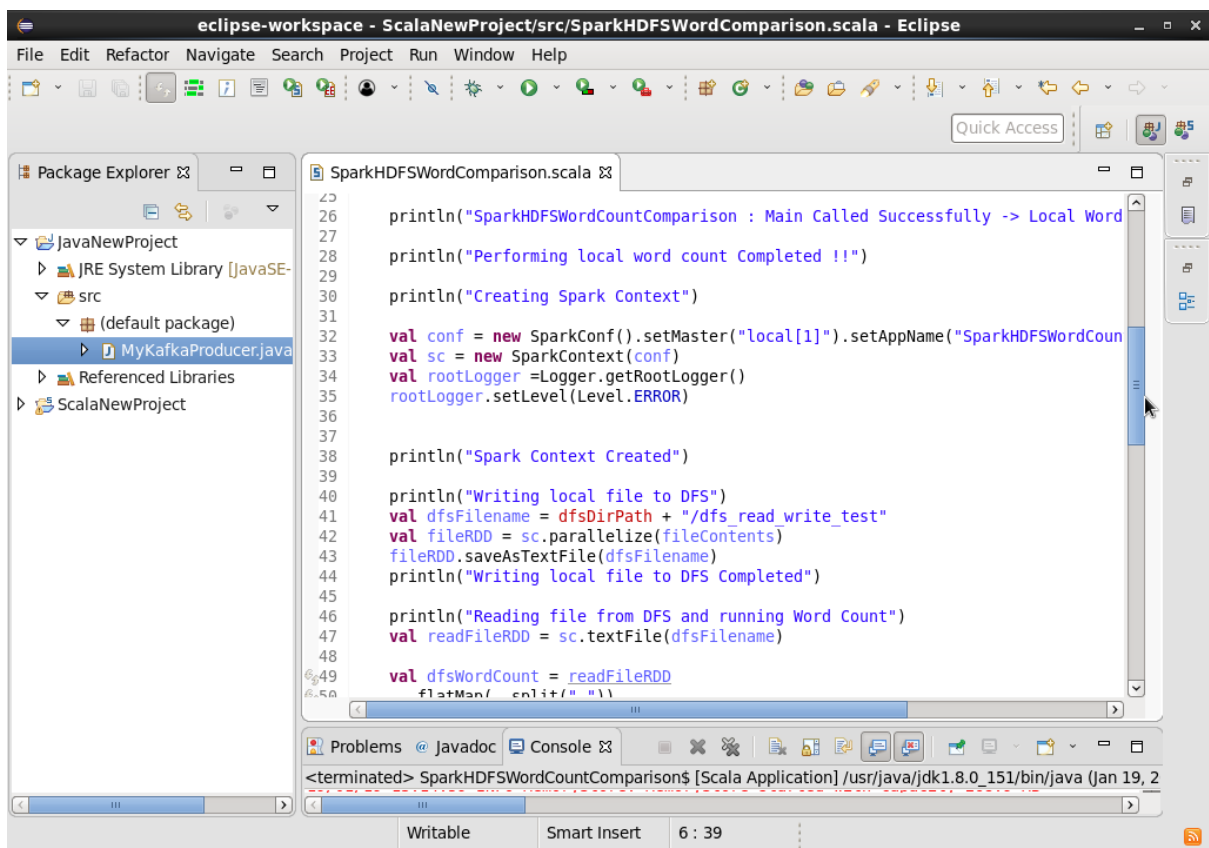
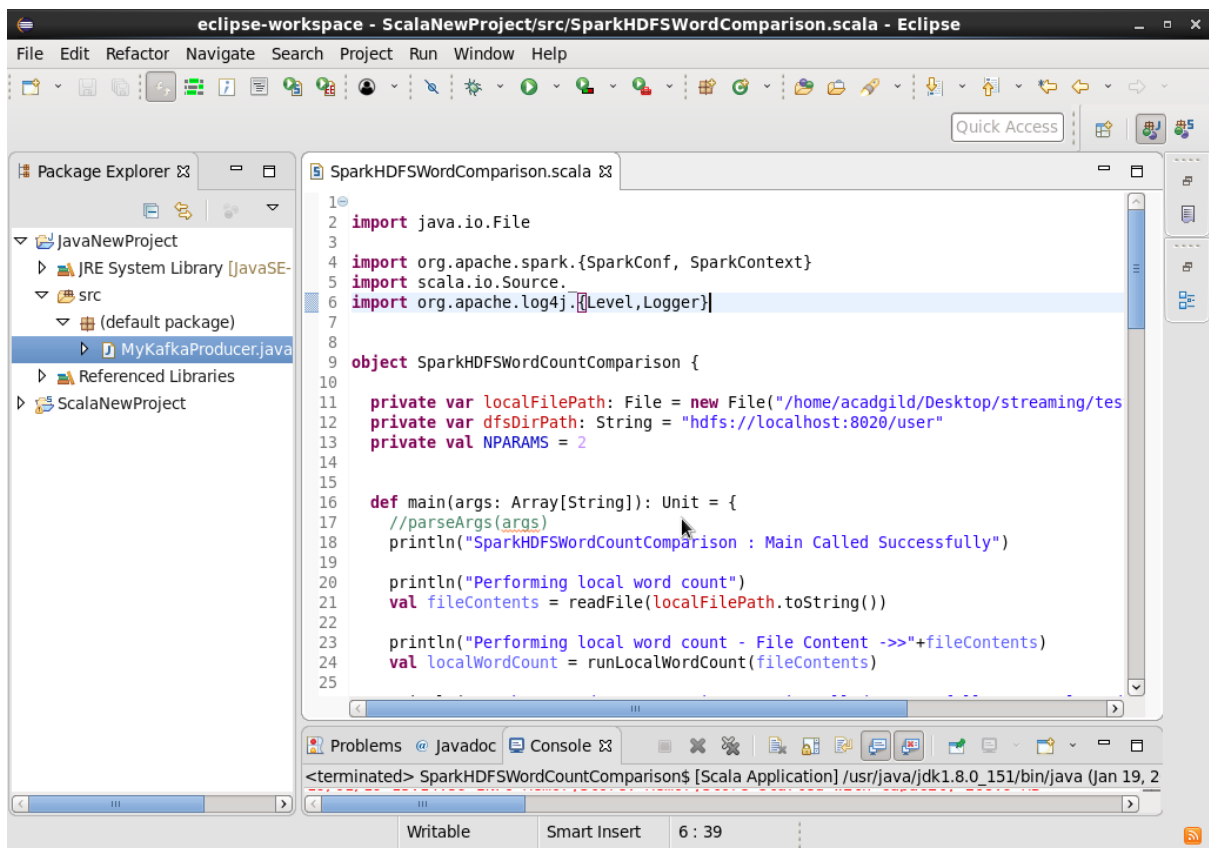
- **Second Part** - In this part, you will have to create a Spark Application which should do the following
 1. Pick up a file from the local directory and do the word count
 2. Then in the same Spark Application, write the code to put the same file on HDFS.
 3. Then in same Spark Application, do the 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

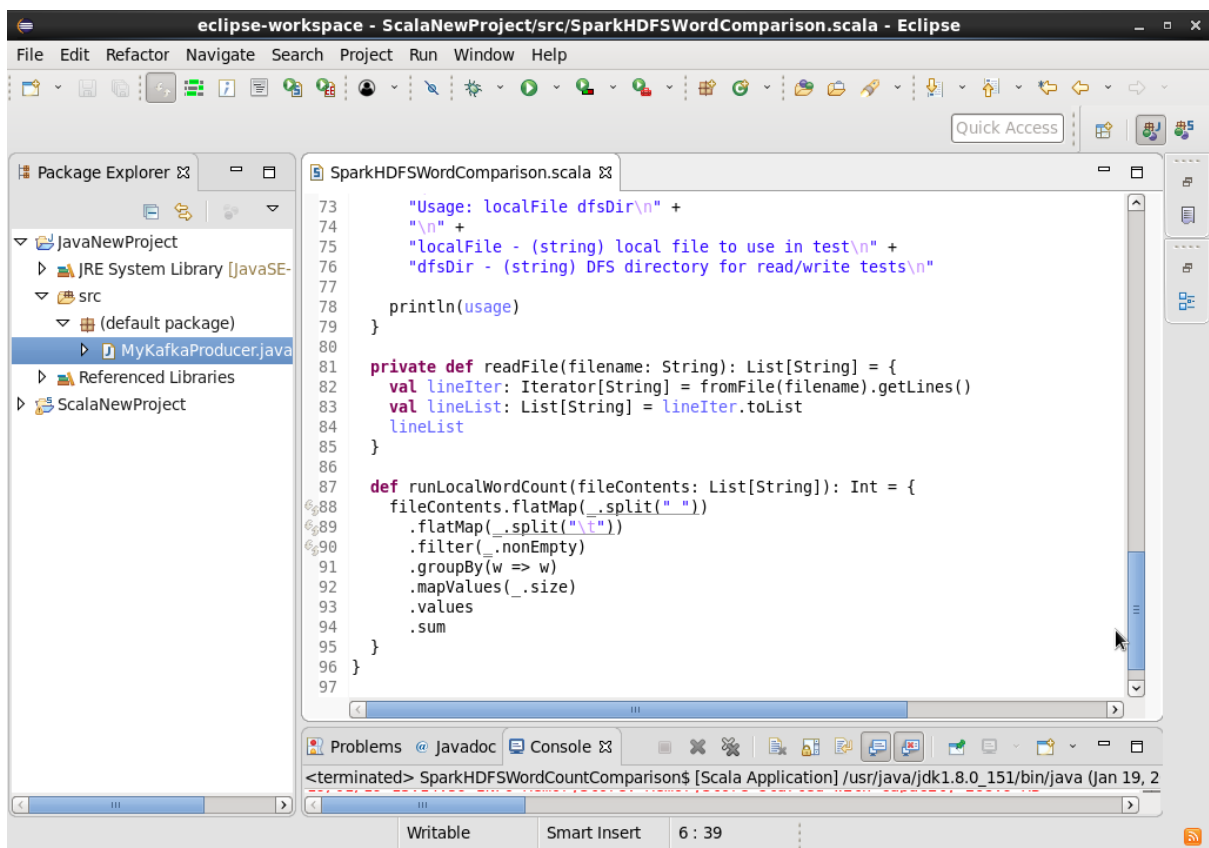
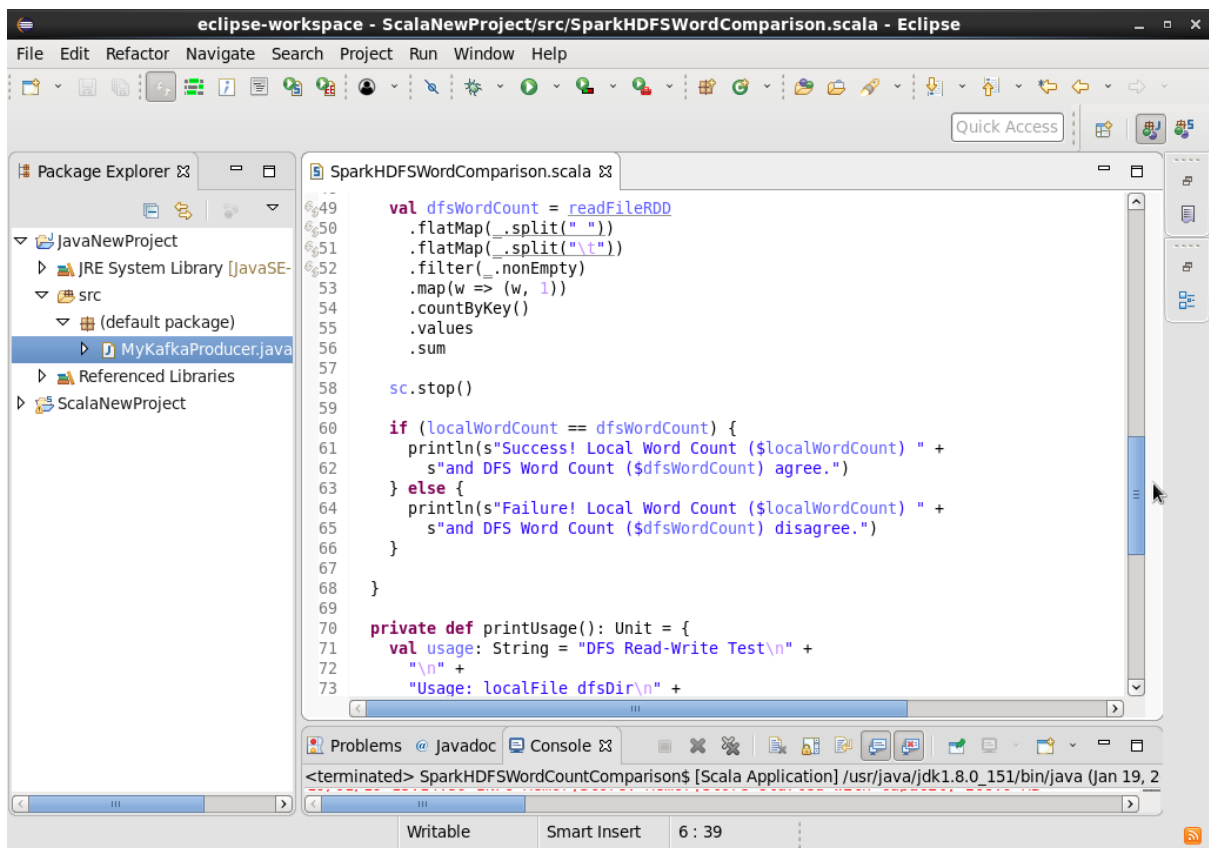
Solution: Input File

A terminal window titled 'acadgild@localhost:~/Desktop' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the command 'cat test.txt' and its output: 'This is spark streaming assignment.' repeated eight times, followed by a system message 'You have new mail in /var/spool/mail/acadgild'. The prompt '[acadgild@localhost Desktop]\$' is followed by a cursor.

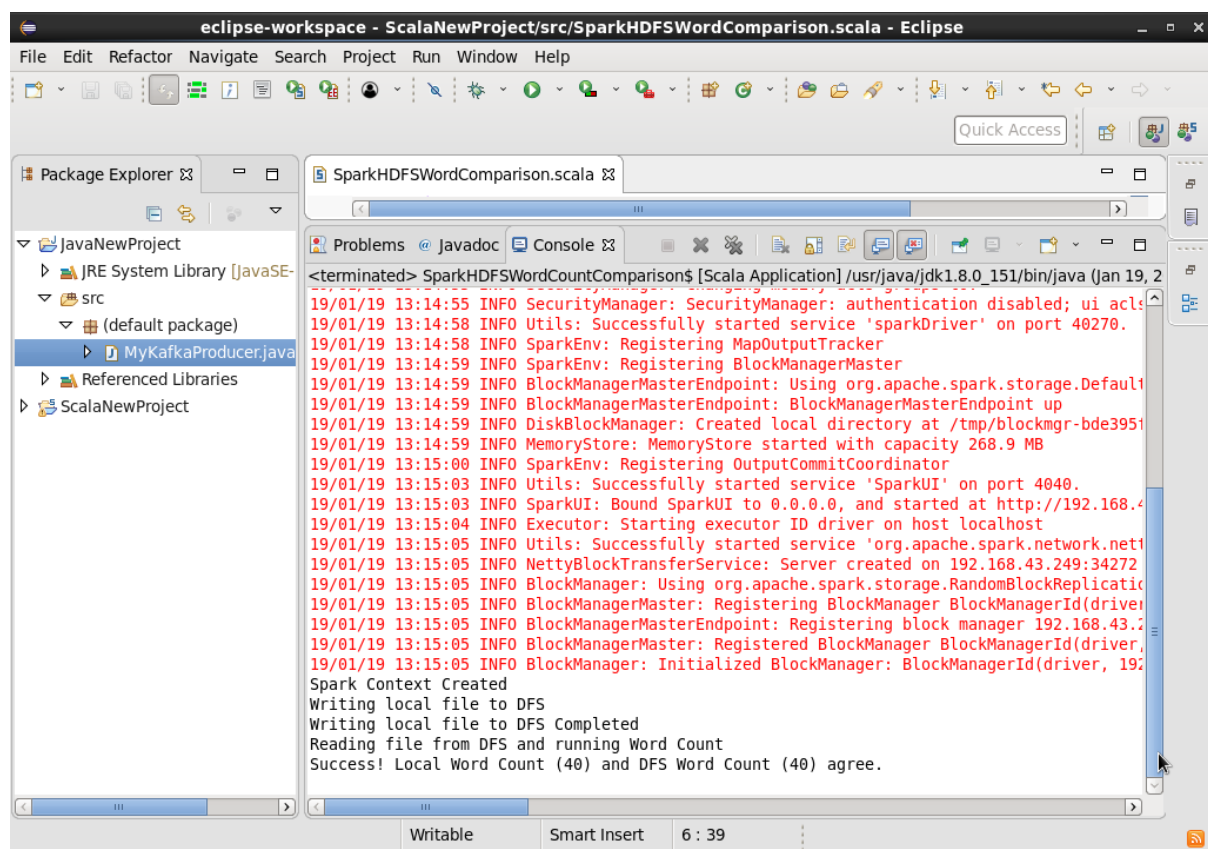
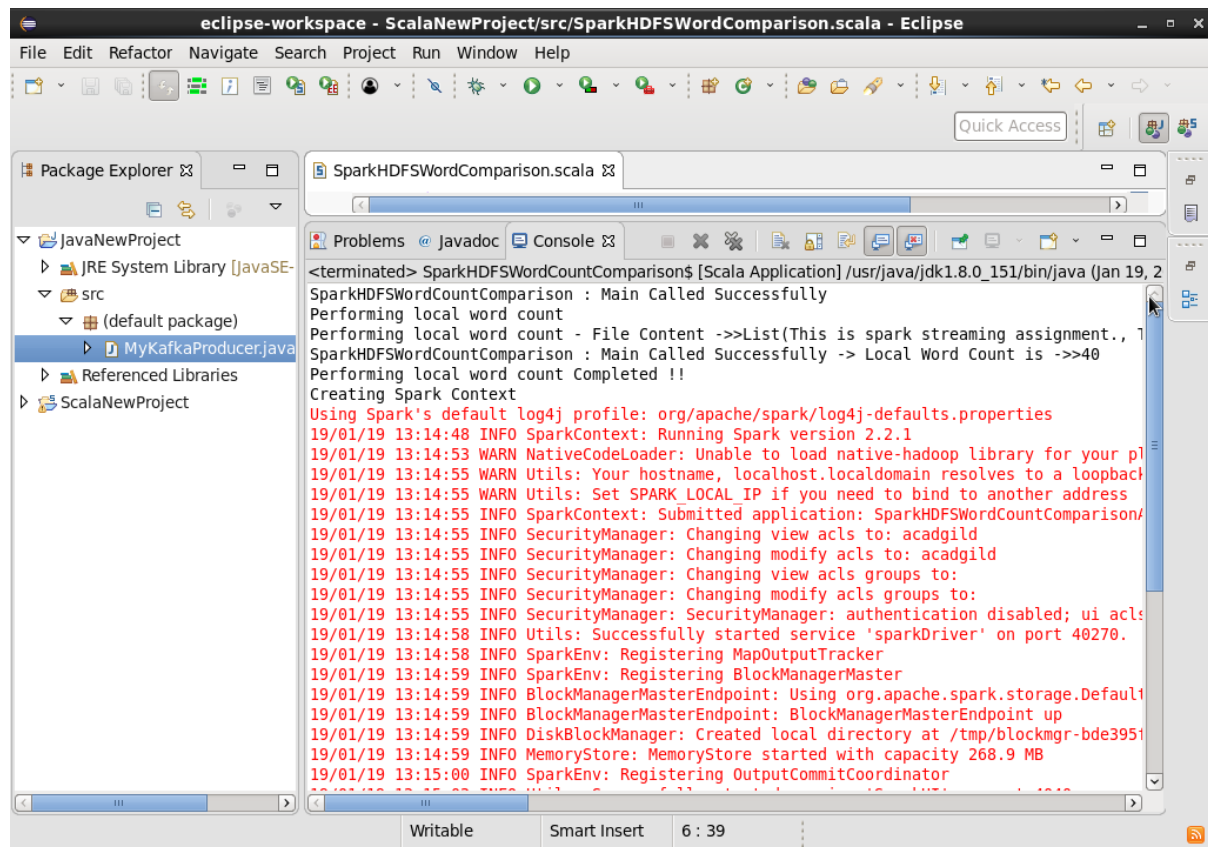
```
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
[acadgild@localhost Desktop]$ cat test.txt
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

Code:





Output:



Output file on HDFS:

127.0.0.1 (acagild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

2 127.0.0.1 (acagild)

```
[acagild@localhost ~]$ hadoop fs -ls /user/
19/01/19 13:20:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 10 items
-rw-r--r-- 1 acagild supergroup          932 2019-01-07 22:41 /user/S20_Dataset_Holidays.txt
-rw-r--r-- 1 acagild supergroup           45 2019-01-07 22:42 /user/S20_Dataset_Transport.txt
-rw-r--r-- 1 acagild supergroup          119 2019-01-07 22:42 /user/S20_Dataset_User_details.txt
-rw-r--r-- 1 acagild supergroup           423 2018-12-23 00:55 /user/Spark_word_file.txt
-rw-r--r-- 1 acagild supergroup           424 2018-12-23 01:00 /user/Spark_word_file1.txt
-rw-r--r-- 1 acagild supergroup           437 2018-12-23 00:49 /user/dataset_Session_14.txt
-rw-r--r-- 1 acagild supergroup           626 2018-12-23 01:30 /user/dataset_Session_19.txt
drwxr-xr-x - acagild supergroup           0 2019-01-19 13:15 /user/dfs_read_write_test
drwxr-xr-x - acagild supergroup           0 2018-02-09 14:50 /user/hive
drwxr-xr-x - acagild supergroup           0 2019-01-10 19:28 /user/streaming
You have new mail in /var/spool/mail/acagild
[acagild@localhost ~]$ hadoop fs -ls /user/dfs_read
19/01/19 13:21:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: /user/dfs_read: No such file or directory
[acagild@localhost ~]$ hadoop fs -ls /user/dfs_read_write_test
19/01/19 13:21:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 3 acagild supergroup           0 2019-01-19 13:15 /user/dfs_read_write_test/_SUCCESS
-rw-r--r-- 3 acagild supergroup          288 2019-01-19 13:15 /user/dfs_read_write_test/part-00000
You have new mail in /var/spool/mail/acagild
[acagild@localhost ~]$ hadoop fs -cat /user/dfs_read_write_test/part-00000
19/01/19 13:27:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
This is spark streaming assignment.
You have new mail in /var/spool/mail/acagild
[acagild@localhost ~]$
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search

ENG 13:27 19-01-2019