Assignment 24.1 Spark Streaming.

Task 1

Read a stream of Strings, fetch the words which can be converted to numbers. Filter out the rows, where the sum of numbers in that line is odd.

Provide the sum of all the remaining numbers in that batch.

Code:

```
import org.apache.spark._
import org.apache.spark.streaming._
import org.apache.spark.streaming.StreamingContext._


val EvenLines = sc.accumulator(0)


val wordstonumbers = Map("Hi"->1, "This"->2, "is"->3, "Assignment"->4, "number"->5,"Twenty"->6,"it"->7, "about"->8, "spark"->9, "Streaming"->10)
val wordstonumbersbroadcast = sc.broadcast(wordstonumbers)


def lineWordNumberSum(line:String):Int = {
var sum:Int = 0
var words = line.split(" ")
for (word <- words) sum += wordstonumbersbroadcast.value.get(word).getOrElse(0)
sum
}


val ssc = new StreamingContext(sc, Seconds(5))
val stream = ssc.socketTextStream("localhost", 9999)


stream.foreachRDD(line => {
val lineStr = line.collect().toList.mkString("")
if (lineStr != "") {
var numTotal = lineWordNumberSum(lineStr)
```

if (numTotal % 2 == 1) println(lineStr)

else{

EvenLines += numTotal

println("Sum of lines with even word number so far =" + EvenLines.value.toInt)}}

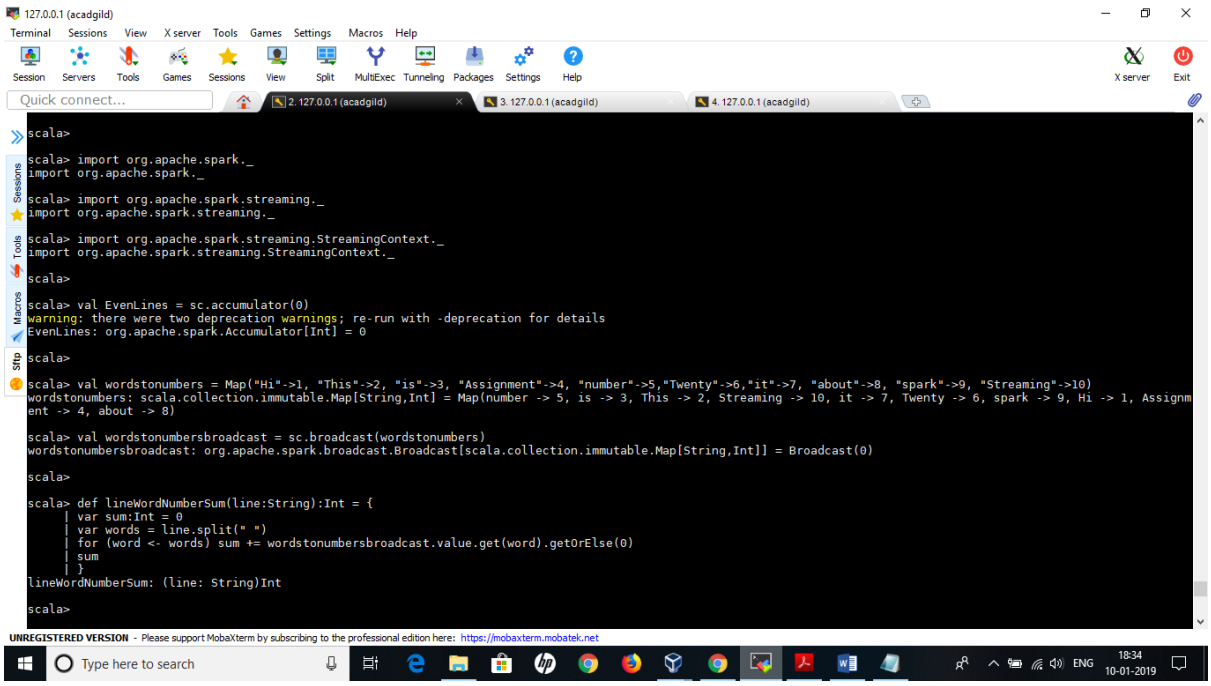
})

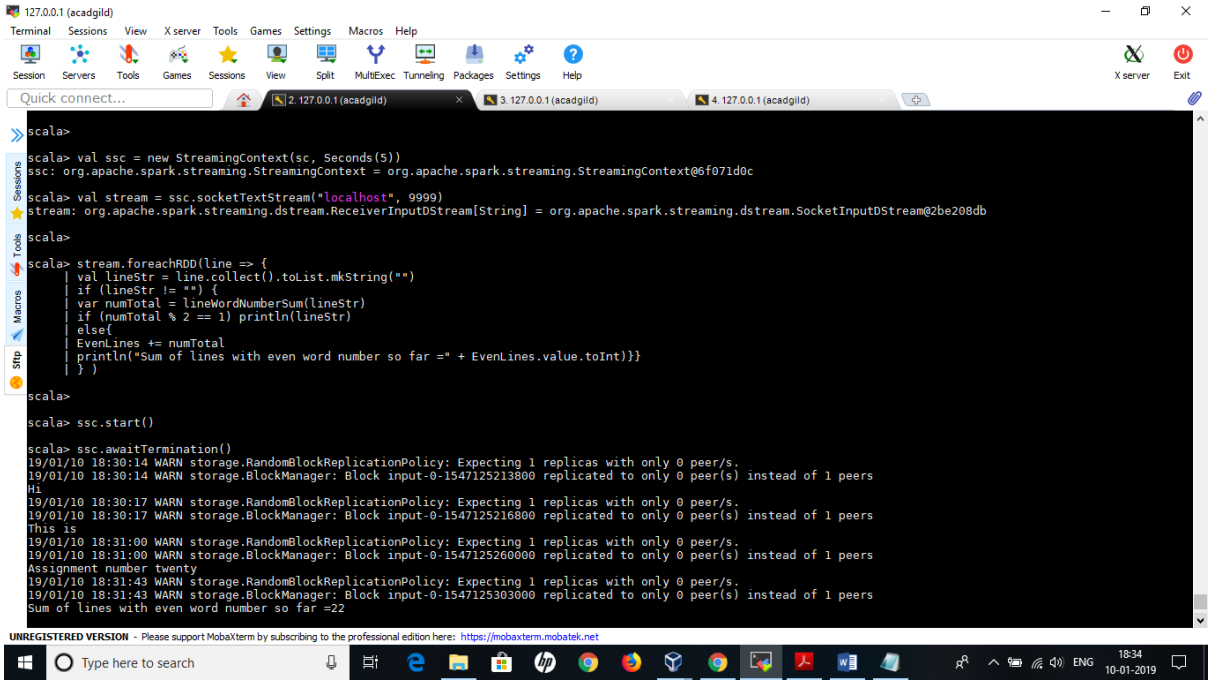
ssc.start()

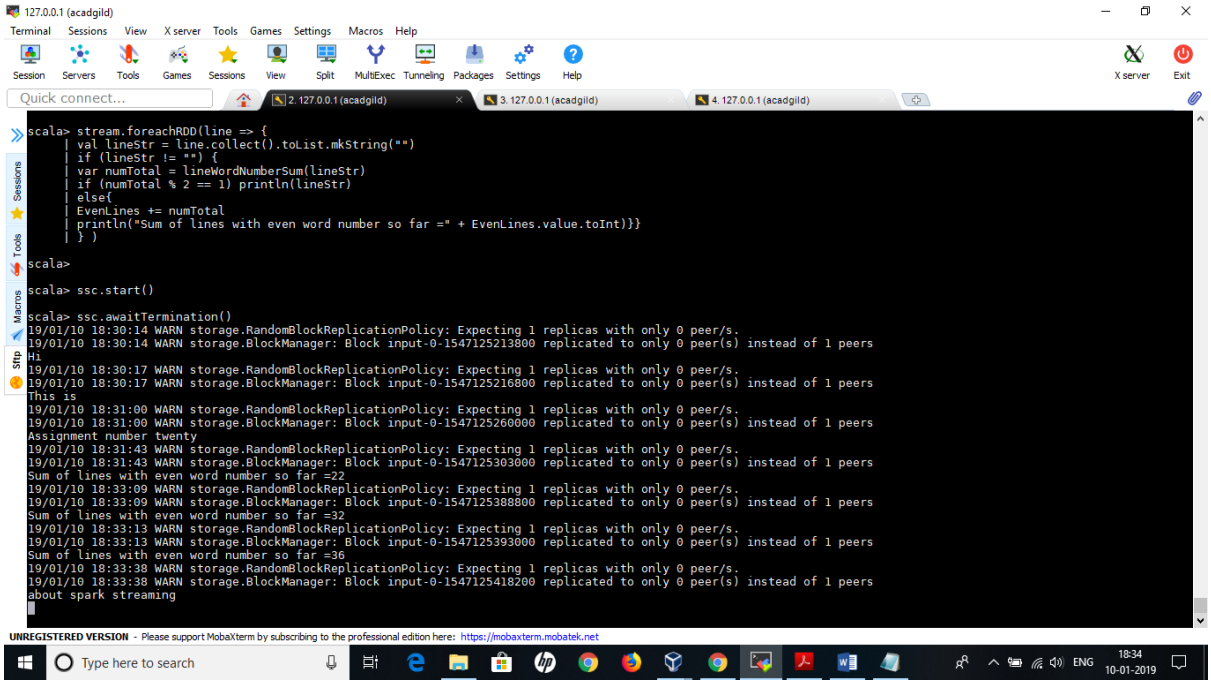
ssc.awaitTermination()

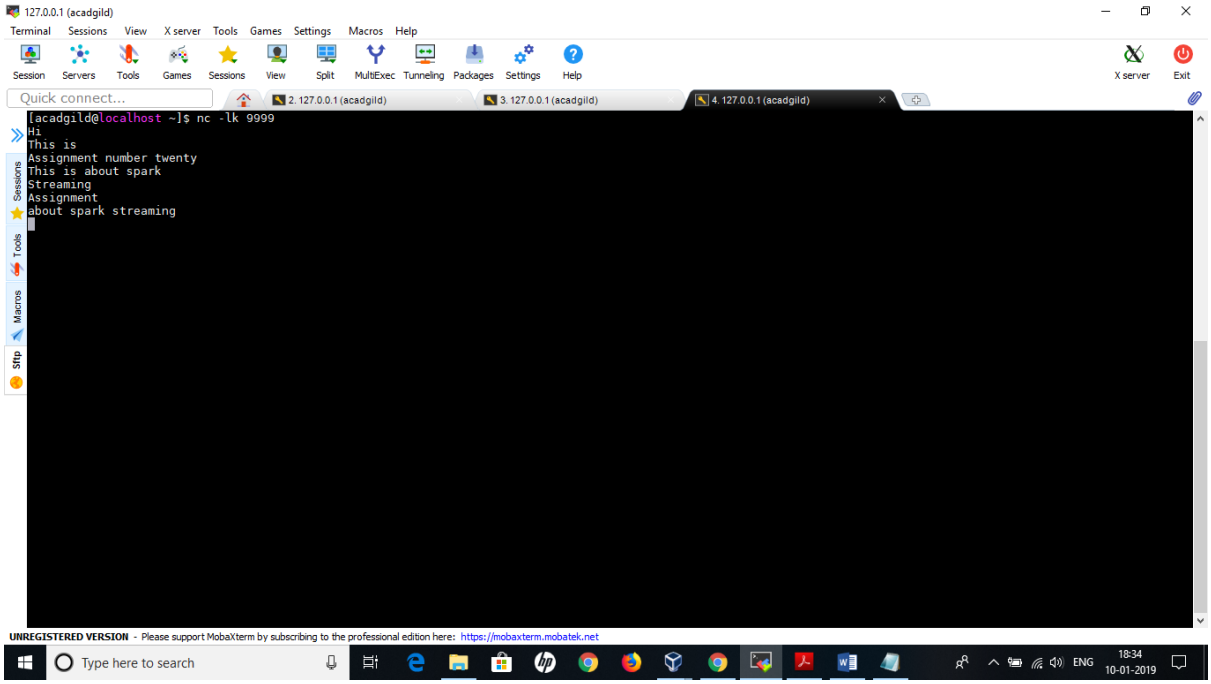
For Netcat: ***nc –lk 9999***

Output:




127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
2. 127.0.0.1 (acadgild)
3. 127.0.0.1 (acadgild)
4. 127.0.0.1 (acadgild)
scala>
scala> val ssc = new StreamingContext(sc, Seconds(5))
ssc: org.apache.spark.streaming.StreamingContext = org.apache.spark.streaming.StreamingContext@6f071d0c
scala> val stream = ssc.socketTextStream("localhost", 9999)
stream: org.apache.spark.streaming.dstream.ReceiverInputDStream[String] = org.apache.spark.streaming.dstream.SocketInputDStream@2be208db
scala>
scala> stream.foreachRDD(line => {
| val lineStr = line.collect().toList.mkString("")
| if (lineStr != "") {
| var numTotal = lineWordNumberSum(lineStr)
| if (numTotal % 2 == 1) println(lineStr)
| else{
| EvenLines += numTotal
| println("Sum of lines with even word number so far =" + EvenLines.value.toInt)}}
| })
scala>
scala> ssc.start()
scala> ssc.awaitTermination()
19/01/10 18:30:14 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:30:14 WARN storage.BlockManager: Block input-0-1547125213800 replicated to only 0 peer(s) instead of 1 peers
Hi
19/01/10 18:30:17 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:30:17 WARN storage.BlockManager: Block input-0-1547125216800 replicated to only 0 peer(s) instead of 1 peers
This is
19/01/10 18:31:00 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:31:00 WARN storage.BlockManager: Block input-0-1547125260000 replicated to only 0 peer(s) instead of 1 peers
Assignment number twenty
19/01/10 18:31:43 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:31:43 WARN storage.BlockManager: Block input-0-1547125303000 replicated to only 0 peer(s) instead of 1 peers
Sum of lines with even word number so far =22
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
Type here to search
ENG 18:34 10-01-2019


127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
2. 127.0.0.1 (acadgild)
3. 127.0.0.1 (acadgild)
4. 127.0.0.1 (acadgild)
scala> stream.foreachRDD(line => {
| val lineStr = line.collect().toList.mkString("")
| if (lineStr != "") {
| var numTotal = lineWordNumberSum(lineStr)
| if (numTotal % 2 == 1) println(lineStr)
| else{
| EvenLines += numTotal
| println("Sum of lines with even word number so far =" + EvenLines.value.toInt)}}
| })
scala>
scala> ssc.start()
scala> ssc.awaitTermination()
19/01/10 18:30:14 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:30:14 WARN storage.BlockManager: Block input-0-1547125213800 replicated to only 0 peer(s) instead of 1 peers
Hi
19/01/10 18:30:17 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:30:17 WARN storage.BlockManager: Block input-0-1547125216800 replicated to only 0 peer(s) instead of 1 peers
This is
19/01/10 18:31:00 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:31:00 WARN storage.BlockManager: Block input-0-1547125260000 replicated to only 0 peer(s) instead of 1 peers
Assignment number twenty
19/01/10 18:31:43 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:31:43 WARN storage.BlockManager: Block input-0-1547125303000 replicated to only 0 peer(s) instead of 1 peers
Sum of lines with even word number so far =22
19/01/10 18:33:09 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:33:09 WARN storage.BlockManager: Block input-0-1547125388800 replicated to only 0 peer(s) instead of 1 peers
Sum of lines with even word number so far =32
19/01/10 18:33:13 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:33:13 WARN storage.BlockManager: Block input-0-1547125393000 replicated to only 0 peer(s) instead of 1 peers
Sum of lines with even word number so far =36
19/01/10 18:33:38 WARN storage.RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
19/01/10 18:33:38 WARN storage.BlockManager: Block input-0-1547125418200 replicated to only 0 peer(s) instead of 1 peers
about spark streaming
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
Type here to search
ENG 18:34 10-01-2019



Task 2

Read two streams

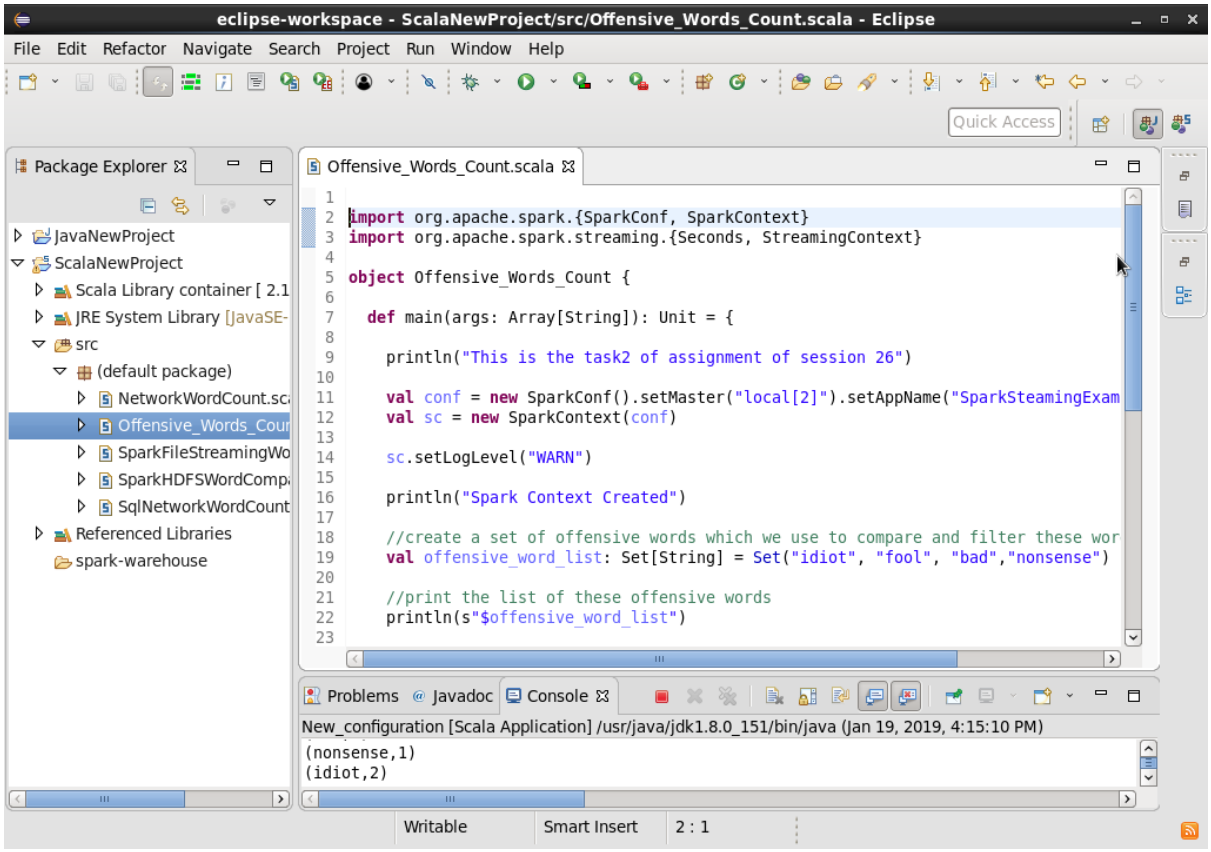
1. List of strings input by user

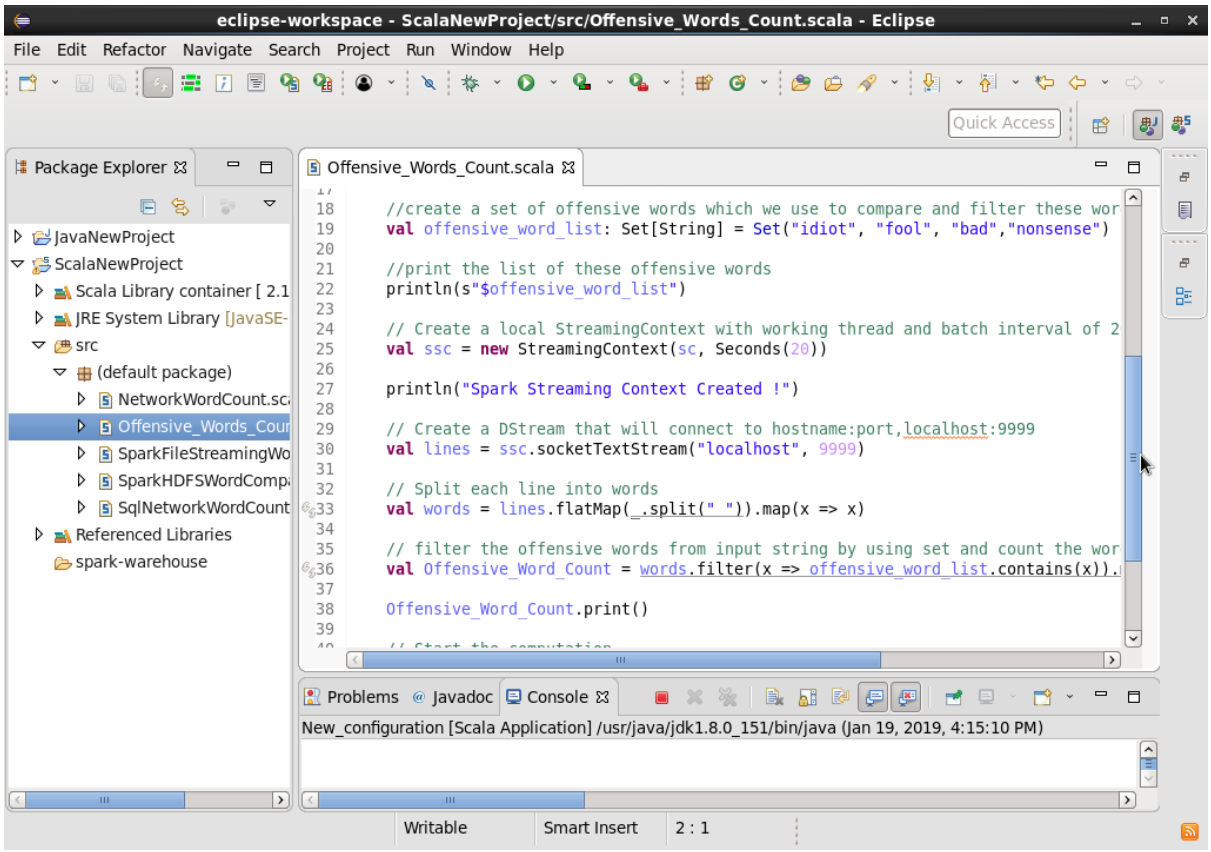
2. Real-time set of offensive words

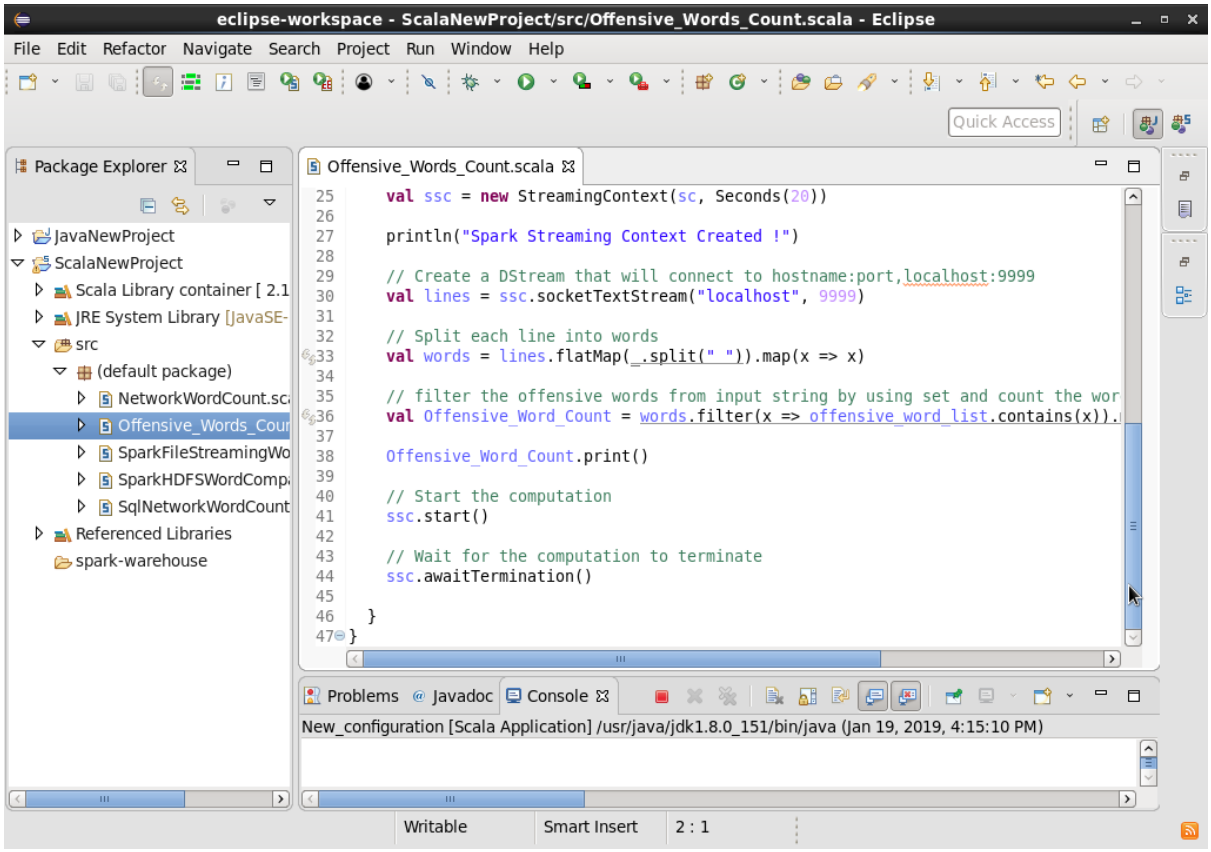
Find the word count of the offensive words inputted by the user as per the real-time set of offensive

Words.

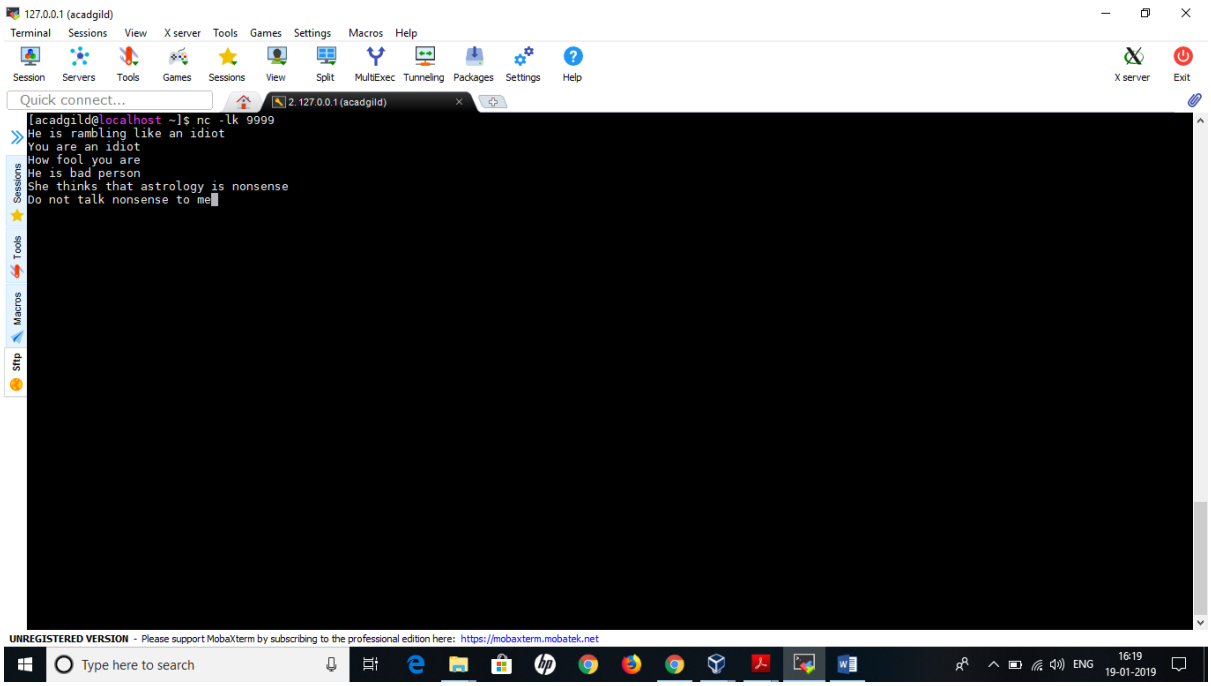
Code:


eclipse-workspace - ScalaNewProject/src/Offensive_Words_Count.scala - Eclipse
File Edit Refactor Navigate Search Project Run Window Help
Quick Access
Package Explorer
JavaNewProject
ScalaNewProject
Scala Library container [2.1
JRE System Library [JavaSE-
src
(default package)
NetworkWordCount.sc
Offensive_Words_Cou
SparkFileStreamingWo
SparkHDFSWordCompa
SqlNetworkWordCount
Referenced Libraries
spark-warehouse
Offensive_Words_Count.scala
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
import org.apache.spark.{SparkConf, SparkContext}
import org.apache.spark.streaming.{Seconds, StreamingContext}
object Offensive_Words_Count {
def main(args: Array[String]): Unit = {
println("This is the task2 of assignment of session 26")
val conf = new SparkConf().setMaster("local[2]").setAppName("SparkSteamingExam
val sc = new SparkContext(conf)
sc.setLogLevel("WARN")
println("Spark Context Created")
//create a set of offensive words which we use to compare and filter these wor
val offensive_word_list: Set[String] = Set("idiot", "fool", "bad","nonsense")
//print the list of these offensive words
println(s"$offensive_word_list")
Problems Javadoc Console
New_configuration [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jan 19, 2019, 4:15:10 PM)
(nonsense,1)
(idiot,2)
Writable Smart Insert 2 : 1

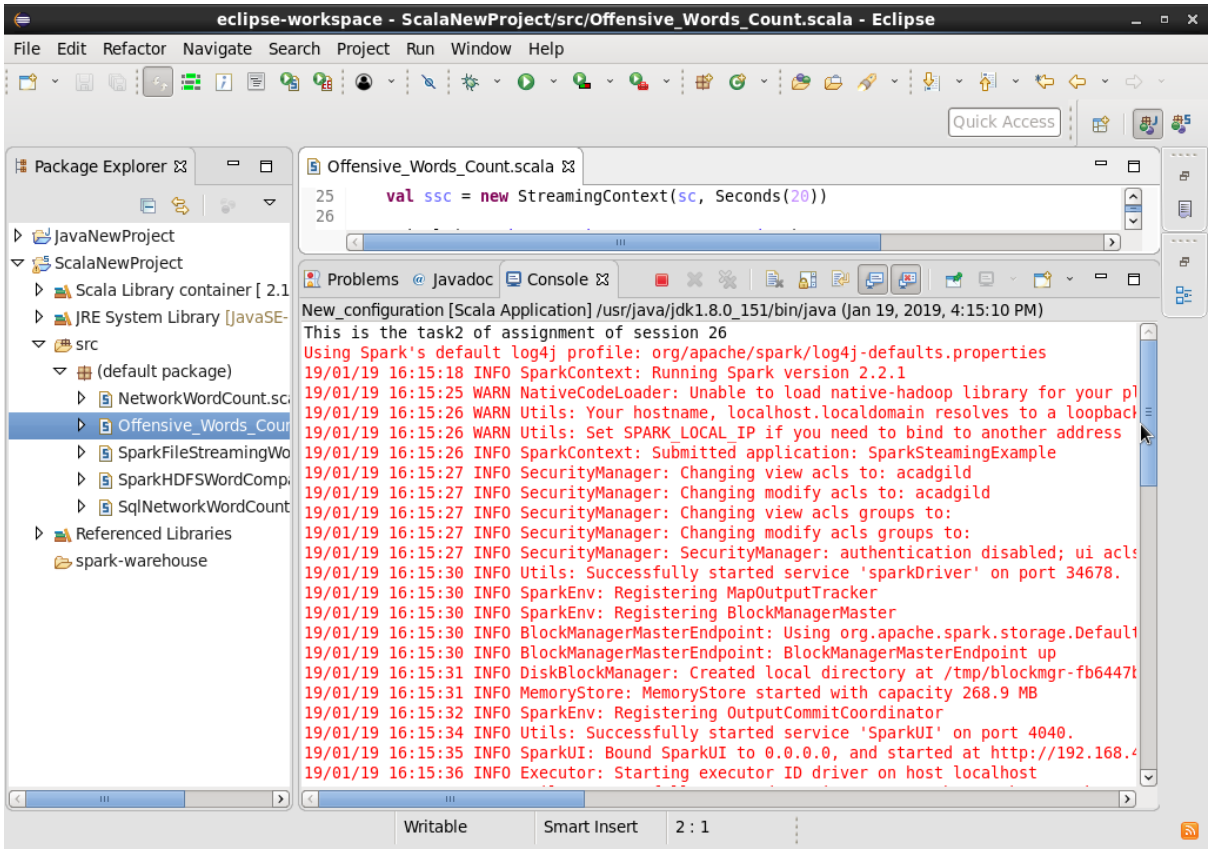

eclipse-workspace - ScalaNewProject/src/Offensive_Words_Count.scala - Eclipse
File Edit Refactor Navigate Search Project Run Window Help
Quick Access
Package Explorer
JavaNewProject
ScalaNewProject
Scala Library container [2.1
JRE System Library [JavaSE-
src
(default package)
NetworkWordCount.sc
Offensive_Words_Cou
SparkFileStreamingWo
SparkHDFSWordCompa
SqlNetworkWordCount
Referenced Libraries
spark-warehouse
Offensive_Words_Count.scala
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
//create a set of offensive words which we use to compare and filter these wor
val offensive_word_list: Set[String] = Set("idiot", "fool", "bad","nonsense")
//print the list of these offensive words
println(s"$offensive_word_list")
// Create a local StreamingContext with working thread and batch interval of 2
val ssc = new StreamingContext(sc, Seconds(20))
println("Spark Streaming Context Created !")
// Create a DStream that will connect to hostname:port,localhost:9999
val lines = ssc.socketTextStream("localhost", 9999)
// Split each line into words
val words = lines.flatMap(_.split(" ")).map(x => x)
// filter the offensive words from input string by using set and count the wor
val Offensive_Word_Count = words.filter(x => offensive_word_list.contains(x)).
Offensive_Word_Count.print()
Problems Javadoc Console
New_configuration [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jan 19, 2019, 4:15:10 PM)
Writable Smart Insert 2 : 1

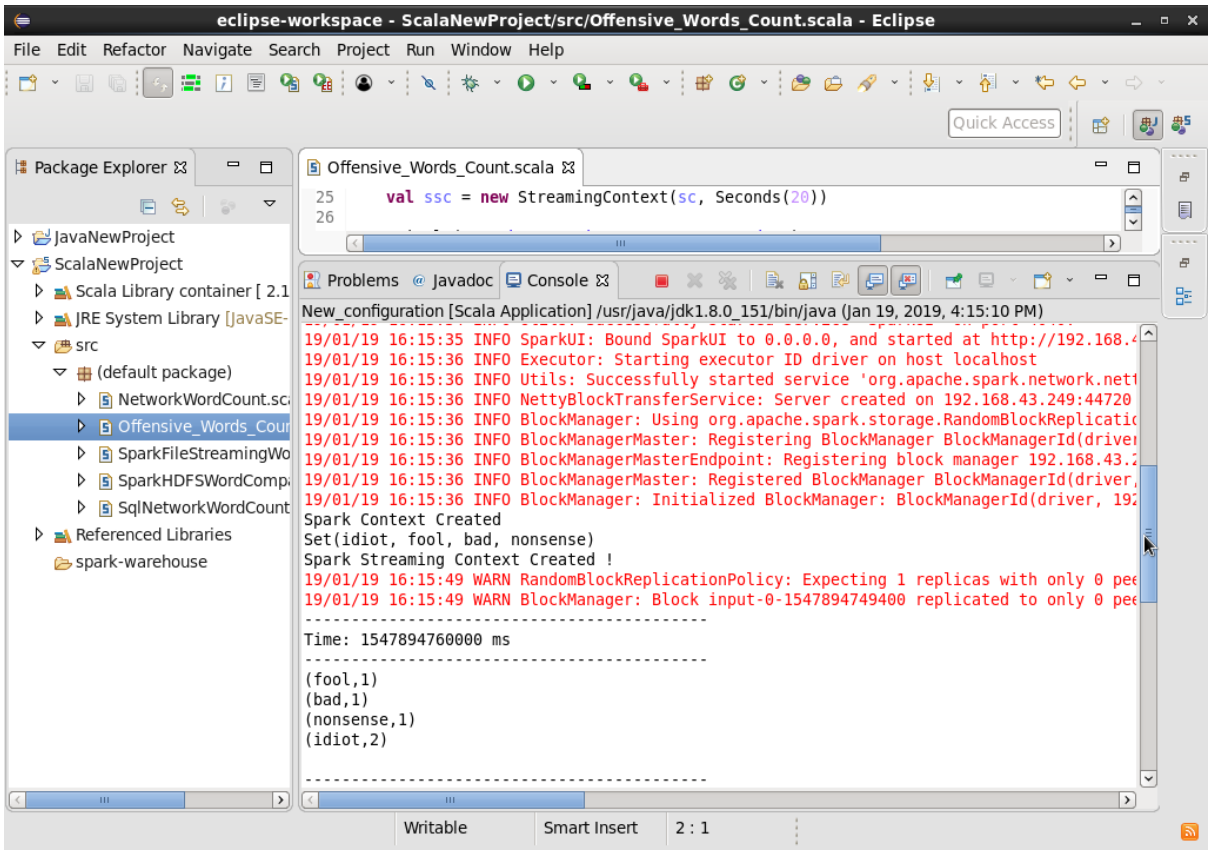


Netcat Input:



Output:


eclipse-workspace - ScalaNewProject/src/Offensive_Words_Count.scala - Eclipse
File Edit Refactor Navigate Search Project Run Window Help
Quick Access
Package Explorer
JavaNewProject
ScalaNewProject
Scala Library container [2.1
JRE System Library [JavaSE-
src
(default package)
NetworkWordCount.sc
Offensive_Words_Cou
SparkFileStreamingWo
SparkHDFSWordCompa
SqlNetworkWordCount
Referenced Libraries
spark-warehouse
Offensive_Words_Count.scala
25 val ssc = new StreamingContext(sc, Seconds(20))
26
Problems Javadoc Console
New_configuration [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jan 19, 2019, 4:15:10 PM)
This is the task2 of assignment of session 26
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
19/01/19 16:15:18 INFO SparkContext: Running Spark version 2.2.1
19/01/19 16:15:25 WARN NativeCodeLoader: Unable to load native-hadoop library for your pl
19/01/19 16:15:26 WARN Utils: Your hostname, localhost.localdomain resolves to a loopbac
19/01/19 16:15:26 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
19/01/19 16:15:26 INFO SparkContext: Submitted application: SparkSteamingExample
19/01/19 16:15:27 INFO SecurityManager: Changing view acls to: acadgild
19/01/19 16:15:27 INFO SecurityManager: Changing modify acls to: acadgild
19/01/19 16:15:27 INFO SecurityManager: Changing view acls groups to:
19/01/19 16:15:27 INFO SecurityManager: Changing modify acls groups to:
19/01/19 16:15:27 INFO SecurityManager: SecurityManager: authentication disabled; ui acls
19/01/19 16:15:30 INFO Utils: Successfully started service 'sparkDriver' on port 34678.
19/01/19 16:15:30 INFO SparkEnv: Registering MapOutputTracker
19/01/19 16:15:30 INFO SparkEnv: Registering BlockManagerMaster
19/01/19 16:15:30 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.Default
19/01/19 16:15:30 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
19/01/19 16:15:31 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-fb6447
19/01/19 16:15:31 INFO MemoryStore: MemoryStore started with capacity 268.9 MB
19/01/19 16:15:32 INFO SparkEnv: Registering OutputCommitCoordinator
19/01/19 16:15:34 INFO Utils: Successfully started service 'SparkUI' on port 4040.
19/01/19 16:15:35 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.
19/01/19 16:15:36 INFO Executor: Starting executor ID driver on host localhost
Writable Smart Insert 2 : 1


eclipse-workspace - ScalaNewProject/src/Offensive_Words_Count.scala - Eclipse
File Edit Refactor Navigate Search Project Run Window Help
Quick Access
Package Explorer
JavaNewProject
ScalaNewProject
Scala Library container [2.1
JRE System Library [JavaSE-
src
(default package)
NetworkWordCount.sc
Offensive_Words_Cou
SparkFileStreamingWo
SparkHDFSWordCompa
SqlNetworkWordCount
Referenced Libraries
spark-warehouse
Offensive_Words_Count.scala
25 val ssc = new StreamingContext(sc, Seconds(20))
26
Problems Javadoc Console
New_configuration [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jan 19, 2019, 4:15:10 PM)
19/01/19 16:15:35 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.
19/01/19 16:15:36 INFO Executor: Starting executor ID driver on host localhost
19/01/19 16:15:36 INFO Utils: Successfully started service 'org.apache.spark.network.net
19/01/19 16:15:36 INFO NettyBlockTransferService: Server created on 192.168.43.249:44720
19/01/19 16:15:36 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicatio
19/01/19 16:15:36 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(drive
19/01/19 16:15:36 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.43.
19/01/19 16:15:36 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver,
19/01/19 16:15:36 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 19
Spark Context Created
Set(idiot, fool, bad, nonsense)
Spark Streaming Context Created !
19/01/19 16:15:49 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 pee
19/01/19 16:15:49 WARN BlockManager: Block input-0-1547894749400 replicated to only 0 pee
Time: 1547894760000 ms
(fool,1)
(bad,1)
(nonsense,1)
(idiot,2)
Writable Smart Insert 2 : 1