Session 24:

SPARK STREAMING

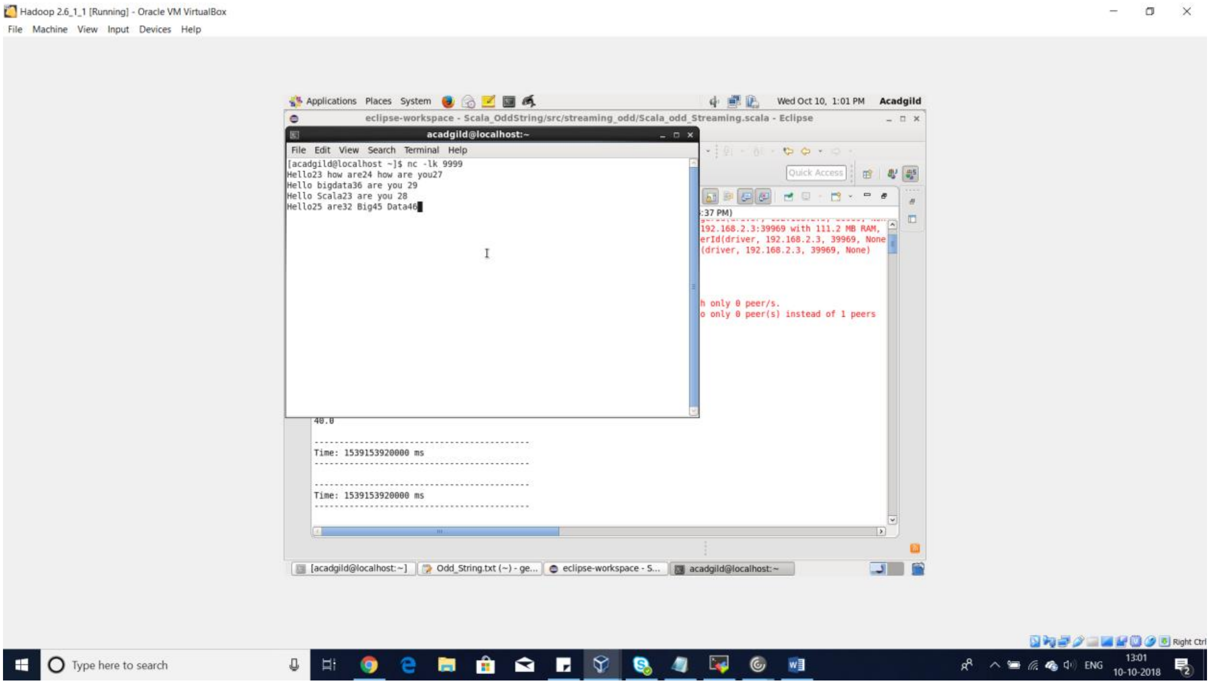
Assignment 1

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.

Output: − Before running the application, start “netcat” in terminal using the command “nc -lk 9999” − Run the spark application − Put some string with numbers on the shell as shown below

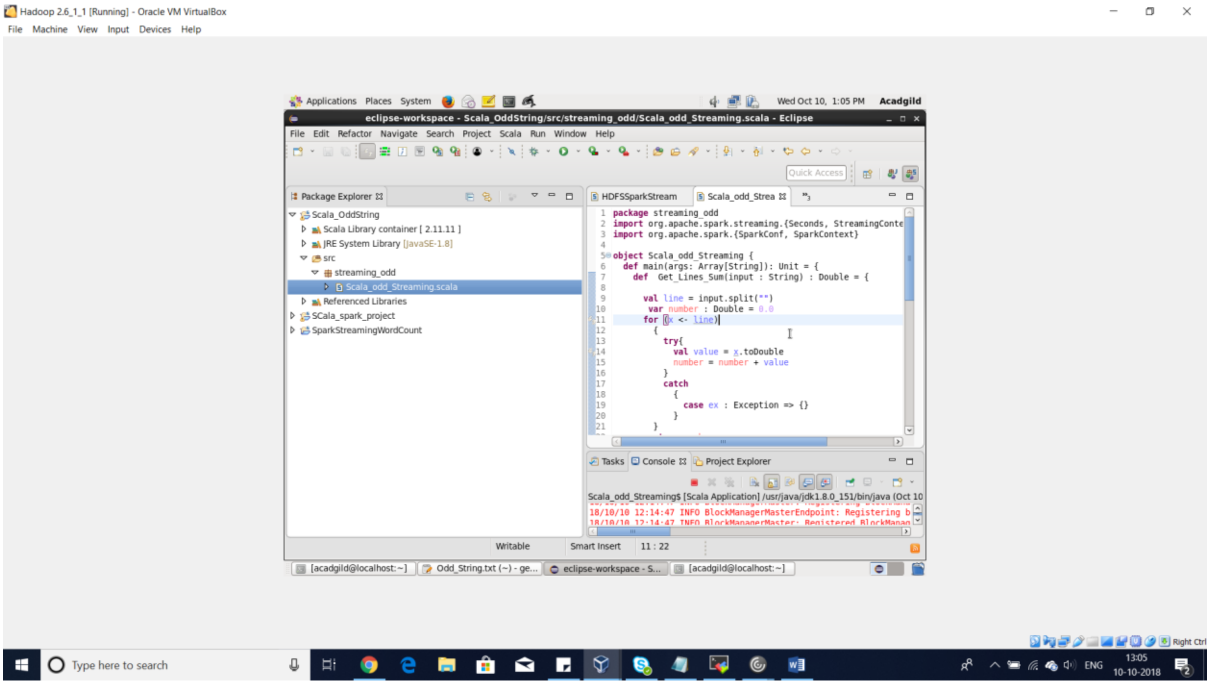


Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Wed Oct 10, 1:01 PM Acadgild
eclipse-workspace - Scala_OddString/src/streaming_odd/Scala_odd_Streaming.scala - Eclipse
acadgild@localhost:~
File Edit View Search Terminal Help
Quick Access
[acadgild@localhost:~]
Odd_String.txt (~) - ge...
eclipse-workspace - S...
acadgild@localhost:~
Type here to search
ENG
13:01
10-10-2018
Right Ctrl

```
[acadgild@localhost ~]$ nc -lk 9999
Hello23 how are24 how are you27
Hello bigdata36 are you 29
Hello Scala23 are you 28
Hello25 are32 Big45 Data46
```

```
:37 PM)
192.168.2.3:39969 with 111.2 MB RAM,
erId(driver, 192.168.2.3, 39969, None
(driver, 192.168.2.3, 39969, None)

h only 0 peer/s.
o only 0 peer(s) instead of 1 peers
```

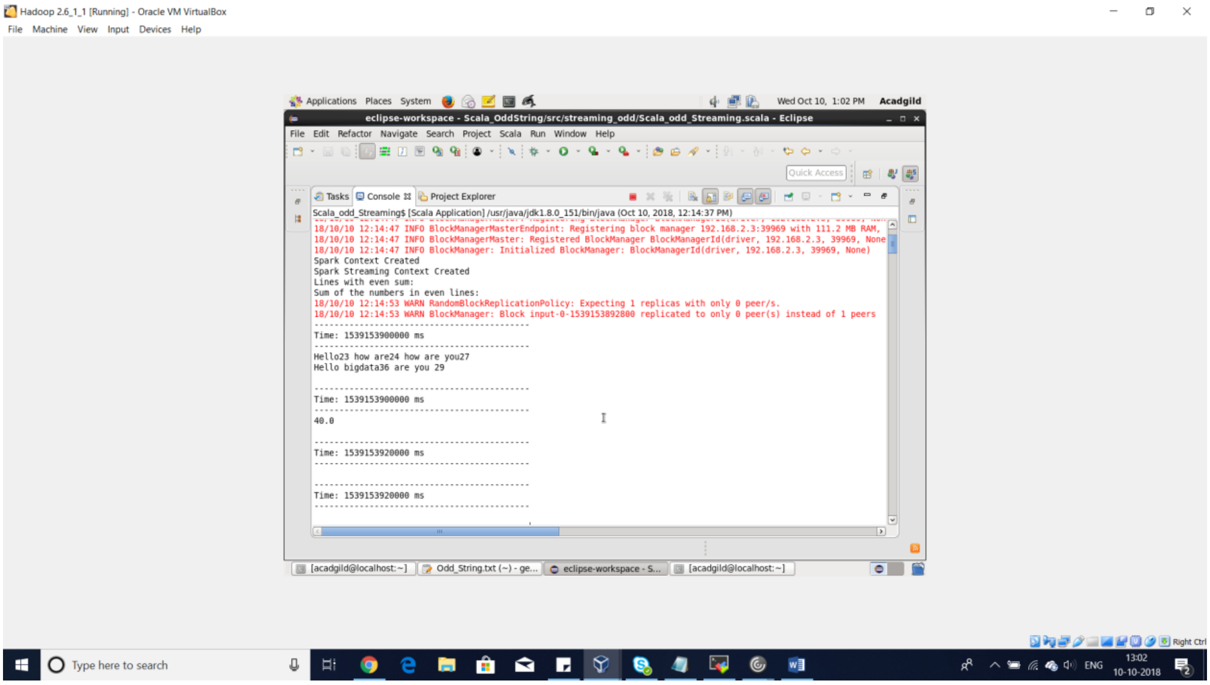
```
40.0

-------------------------------------------
Time: 1539153920000 ms
-------------------------------------------

-------------------------------------------
Time: 1539153920000 ms
-------------------------------------------
```



```
package streaming_odd
import org.apache.spark.streaming.{Seconds, StreamingConte
import org.apache.spark.{SparkConf, SparkContext}

object Scala_odd_Streaming {
  def main(args: Array[String]): Unit = {
    def  Get_Lines_Sum(input : String) : Double = {

      val line = input.split("")
       var number : Double = 0.0
      for (x <- line)
        {
          try{
            val value = x.toDouble
            number = number + value
          }
          catch
            {
              case ex : Exception => {}
            }
        }
```

```
Scala_odd_Streaming$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Oct 10
18/10/10 12:14:47 INFO BlockManagerMasterEndpoint: Registering b
```

Given below screenshot – we are able to see that lines containing sum of odd numbers are filter, even number is displayed, and sum of the number is displayed in the next line.



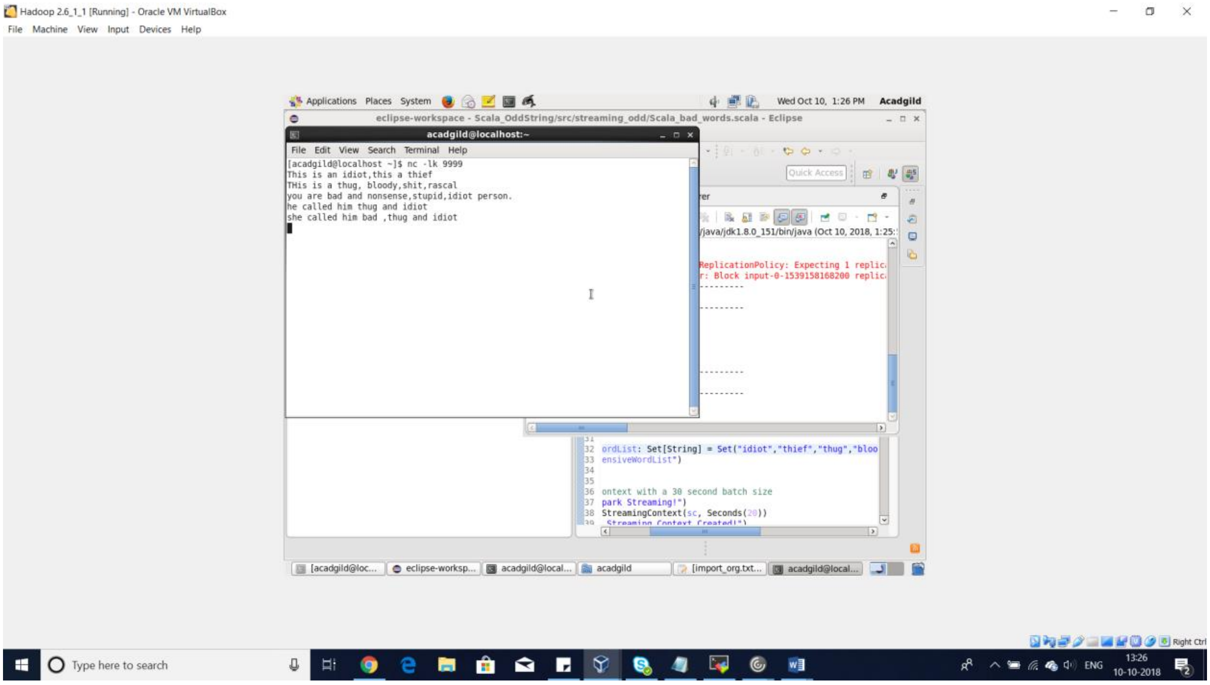
```
Scala_odd_Streaming$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Oct 10, 2018, 12:14:37 PM)
18/10/10 12:14:47 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.2.3:39969 with 111.2 MB RAM,
18/10/10 12:14:47 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.2.3, 39969, None
18/10/10 12:14:47 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.2.3, 39969, None)
Spark Context Created
Spark Streaming Context Created
Lines with even sum:
Sum of the numbers in even lines:
18/10/10 12:14:53 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/10/10 12:14:53 WARN BlockManager: Block input-0-1539153892800 replicated to only 0 peer(s) instead of 1 peers
-------------------------------------------
Time: 1539153900000 ms
-------------------------------------------
Hello23 how are24 how are you27
Hello bigdata36 are you 29

-------------------------------------------
Time: 1539153900000 ms
-------------------------------------------
40.0

-------------------------------------------
Time: 1539153920000 ms
-------------------------------------------

-------------------------------------------
Time: 1539153920000 ms
-------------------------------------------
```

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 realtime set of offensive words Solution: Note: Source code file is provided along with this assignment report.

Solution:

In the spark application, we have a set of words that we considered as offensive words "idiot", "fool", "bad", "nonsense", "shit", "damn", "stupid", "dash", "bloody", "rascal",thief,thug

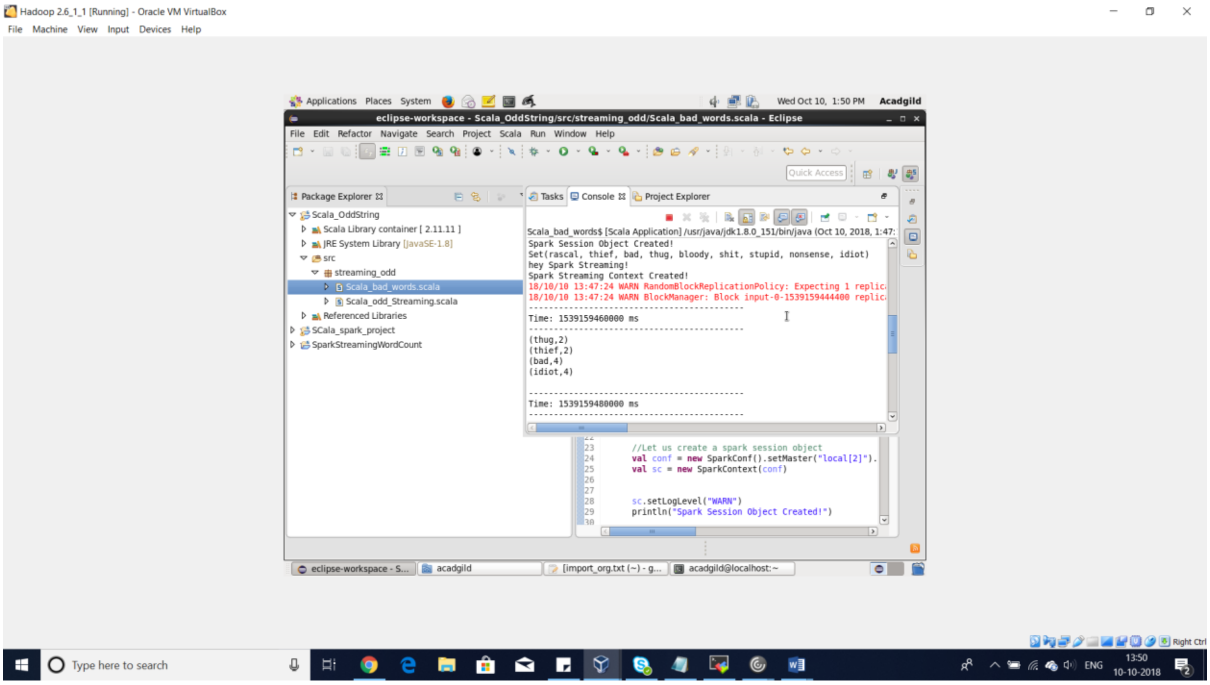
In screenshot below, we are able to see that spark application has counted the number of offensive words occur as per the input string provided by the user.


Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Wed Oct 10, 1:26 PM Acadgild
eclipse-workspace - Scala_OddString/src/streaming_odd/Scala_bad_words.scala - Eclipse
acadgild@localhost:~
File Edit View Search Terminal Help
Quick Access
[acadgild@loc... eclipse-worksp... acadgild@local... acadgild [import_org.txt... acadgild@local...
Type here to search
ENG
13:26
10-10-2018
Right Ctrl

```
[acadgild@localhost ~]$ nc -lk 9999
This is an idiot,this a thief
THis is a thug, bloody,shit,rascal
you are bad and nonsense,stupid,idiot person.
he called him thug and idiot
she called him bad ,thug and idiot
```

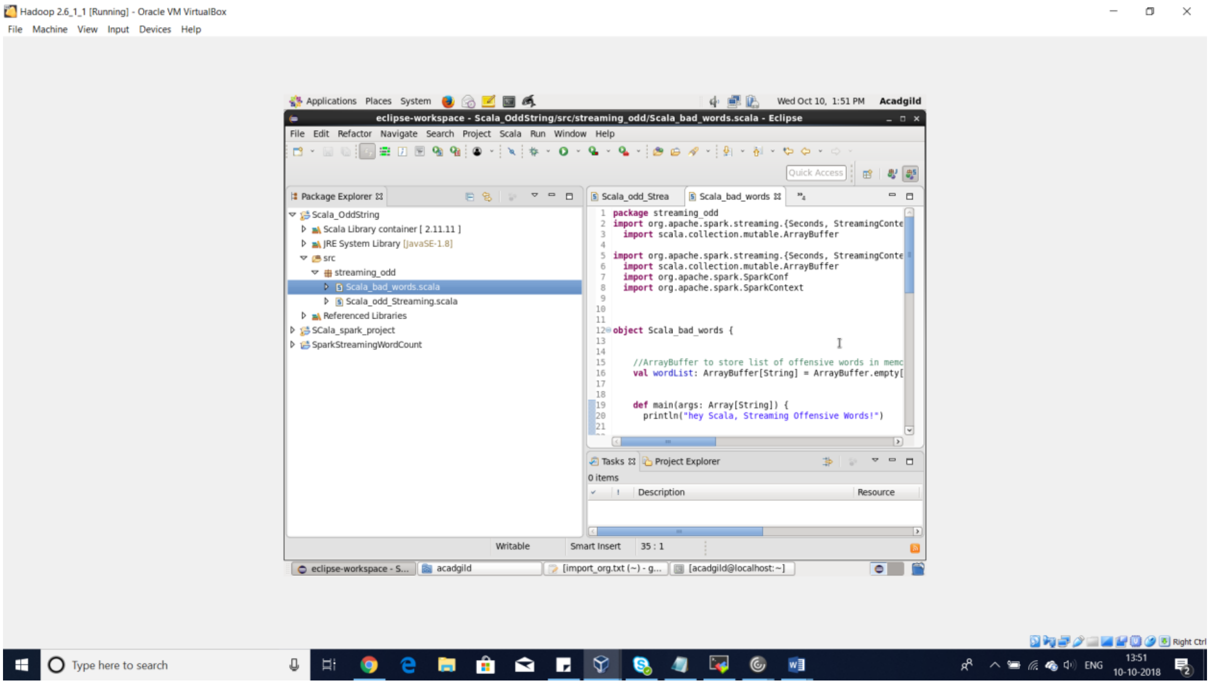
```
/java/jdk1.8.0_151/bin/java (Oct 10, 2018, 1:25:

ReplicationPolicy: Expecting 1 replic
r: Block input-0-1539158168200 replic
----------

----------



----------

----------
```

```
ordList: Set[String] = Set("idiot","thief","thug","bloo
ensiveWordList")


ontext with a 30 second batch size
park Streaming!")
StreamingContext(sc, Seconds(20))
 Streaming Context Created!")
```


Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Wed Oct 10, 1:50 PM Acadgild
eclipse-workspace - Scala_OddString/src/streaming_odd/Scala_bad_words.scala - Eclipse
File Edit Refactor Navigate Search Project Scala Run Window Help
Quick Access
Package Explorer
Scala_OddString
Scala Library container [2.11.11]
JRE System Library [JavaSE-1.8]
src
streaming_odd
Scala_bad_words.scala
Scala_odd_Streaming.scala
Referenced Libraries
SCala_spark_project
SparkStreamingWordCount
Tasks
Console
Project Explorer
eclipse-workspace - S...
acadgild
[import_org.txt (~) - g...
acadgild@localhost:~
Type here to search
ENG
13:50
10-10-2018
Right Ctrl

```
Scala_bad_words$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Oct 10, 2018, 1:47:
Spark Session Object Created!
Set(rascal, thief, bad, thug, bloody, shit, stupid, nonsense, idiot)
hey Spark Streaming!
Spark Streaming Context Created!
18/10/10 13:47:24 WARN RandomBlockReplicationPolicy: Expecting 1 replic
18/10/10 13:47:24 WARN BlockManager: Block input-0-1539159444400 replic
-------------------------------------------
Time: 1539159460000 ms
-------------------------------------------
(thug,2)
(thief,2)
(bad,4)
(idiot,4)

-------------------------------------------
Time: 1539159480000 ms
-------------------------------------------
```

```
        //Let us create a spark session object
        val conf = new SparkConf().setMaster("local[2]").
        val sc = new SparkContext(conf)


        sc.setLogLevel("WARN")
        println("Spark Session Object Created!")
```


Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Wed Oct 10, 1:51 PM Acadgild
eclipse-workspace - Scala_OddString/src/streaming_odd/Scala_bad_words.scala - Eclipse
File Edit Refactor Navigate Search Project Scala Run Window Help
Quick Access
Package Explorer
Scala_OddString
Scala Library container [2.11.11]
JRE System Library [JavaSE-1.8]
src
streaming_odd
Scala_bad_words.scala
Scala_odd_Streaming.scala
Referenced Libraries
SCala_spark_project
SparkStreamingWordCount
Scala_odd_Strea
Scala_bad_words
Tasks
Project Explorer
0 items
Description
Resource
Writable
Smart Insert
35 : 1
eclipse-workspace - S...
acadgild
[import_org.txt (~) - g...
[acadgild@localhost:~]
Type here to search
ENG
13:51
10-10-2018
Right Ctrl

```
package streaming_odd
import org.apache.spark.streaming.{Seconds, StreamingConte
  import scala.collection.mutable.ArrayBuffer

import org.apache.spark.streaming.{Seconds, StreamingConte
  import scala.collection.mutable.ArrayBuffer
  import org.apache.spark.SparkConf
  import org.apache.spark.SparkContext



object Scala_bad_words {


    //ArrayBuffer to store list of offensive words in memo
    val wordList: ArrayBuffer[String] = ArrayBuffer.empty[


    def main(args: Array[String]) {
      println("hey Scala, Streaming Offensive Words!")

```