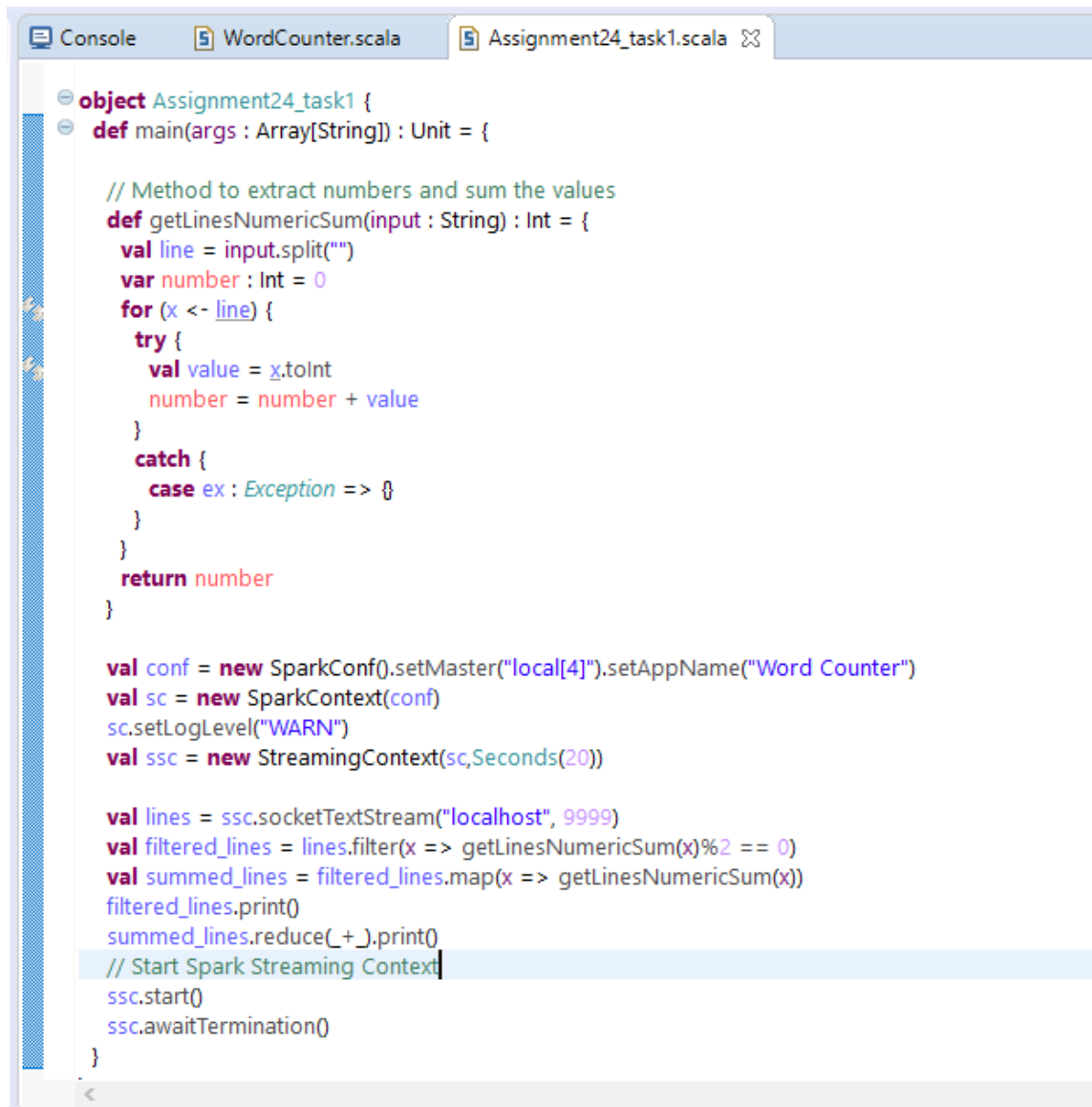


Session 24: SPARK STREAMING Assignment 1

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.

Program: (Assignment24_task1.scala)



```
object Assignment24_task1 {  
  def main(args : Array[String]) : Unit = {  
  
    // Method to extract numbers and sum the values  
    def getLinesNumericSum(input : String) : Int = {  
      val line = input.split(" ")  
      var number : Int = 0  
      for (x <- line) {  
        try {  
          val value = x.toInt  
          number = number + value  
        }  
        catch {  
          case ex : Exception => {}  
        }  
      }  
      return number  
    }  
  
    val conf = new SparkConf().setMaster("local[4]").setAppName("Word Counter")  
    val sc = new SparkContext(conf)  
    sc.setLogLevel("WARN")  
    val ssc = new StreamingContext(sc, Seconds(20))  
  
    val lines = ssc.socketTextStream("localhost", 9999)  
    val filtered_lines = lines.filter(x => getLinesNumericSum(x)%2 == 0)  
    val summed_lines = filtered_lines.map(x => getLinesNumericSum(x))  
    filtered_lines.print()  
    summed_lines.reduce(_+_).print()  
    // Start Spark Streaming Context  
    ssc.start()  
    ssc.awaitTermination()  
  }  
}
```

Execution:

While executing, the program will listen to port number 9999 in localhost machine for data that will form the DStream of data.

In order to establish connection with port 9999 and key in input to that port, Netcat utility has been used.

First, execute the below command in Command prompt to listen port 9999,

Prompt> `nc64.exe -l -p 9999`

```
C:\Windows\System32\cmd.exe - nc64.exe -l -p 9999
Microsoft Windows [Version 10.0.16299.904]
(c) 2017 Microsoft Corporation. All rights reserved.
D:\TOOLS\netcat-win32-1.12>nc64.exe -l -p 9999
indian45 air7 force2
abhinandhan2 chennai2
vijay3 lakshman6 saravanan1
```

Execute Assignment24_task1.scala program that will receive DStream data from port 9999, extract & split characters in string and convert them to int values. Then sum those values.

```
83
84 Time: 1551527560000 ms
85
86
87 2019-03-02 17:22:52 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
88 2019-03-02 17:22:52 WARN BlockManager:66 - Block input-0-1551527572000 replicated to only 0 peer(s) instead of 1 peers
89
90 Time: 1551527580000 ms
91
92 indian45 air7 force2
93
94
95 Time: 1551527580000 ms
96
97 18
98
99 2019-03-02 17:23:03 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
100 2019-03-02 17:23:03 WARN BlockManager:66 - Block input-0-1551527583400 replicated to only 0 peer(s) instead of 1 peers
101
102 Time: 1551527600000 ms
103
104 abhinandhan2 chennai2
105
106
107 Time: 1551527600000 ms
108
109 4
110
111 2019-03-02 17:23:21 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
112 2019-03-02 17:23:21 WARN BlockManager:66 - Block input-0-1551527601000 replicated to only 0 peer(s) instead of 1 peers
113
114 Time: 1551527620000 ms
115
116 vijay3 lakshman6 saravanan1
117
118
119 Time: 1551527620000 ms
120
121 10
122
```

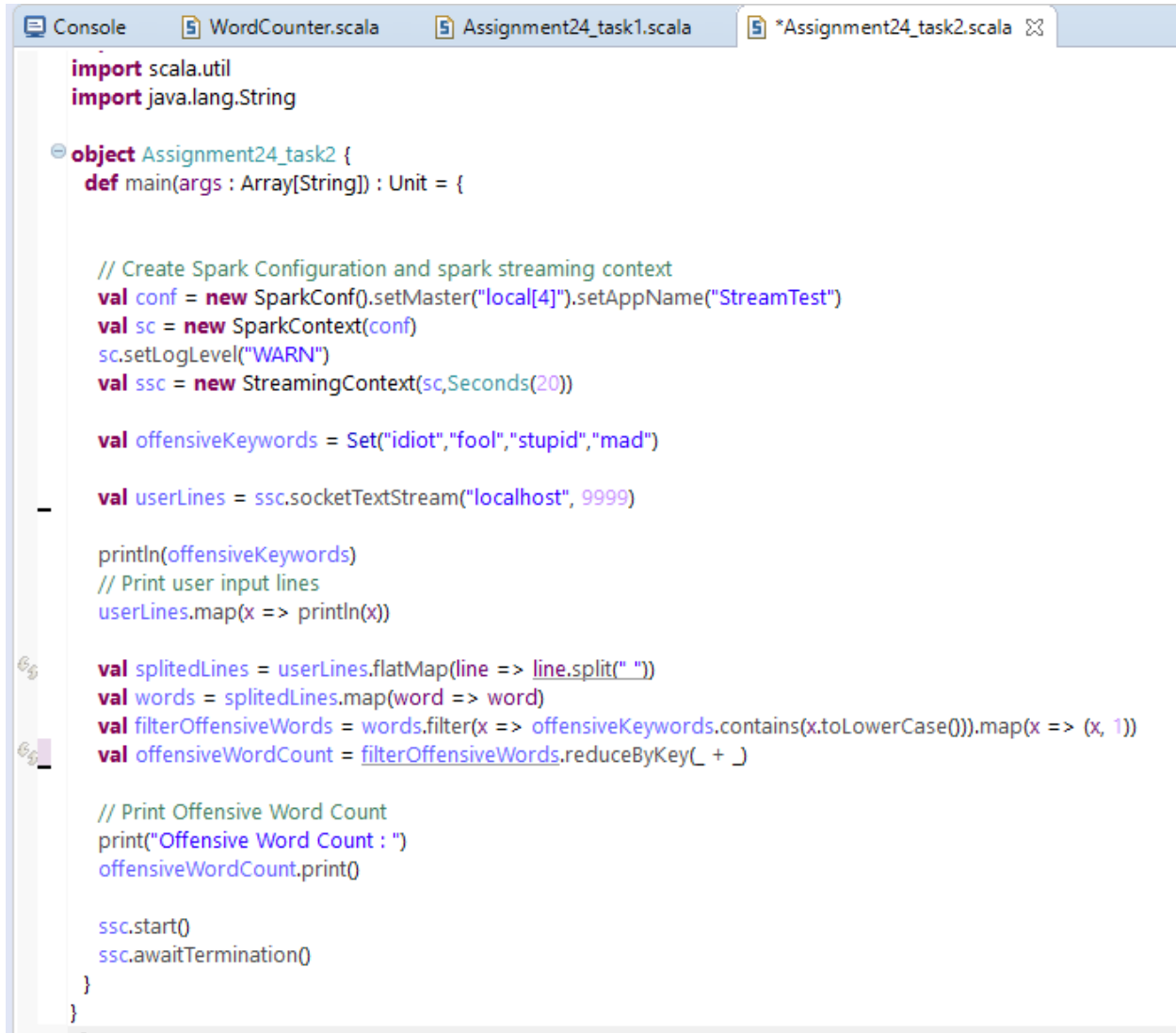
Task – 2:

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

Program: (Assignment24_task1.scala)

The image shows a screenshot of an IDE with several tabs at the top: 'Console', 'WordCounter.scala', 'Assignment24_task1.scala', and '*Assignment24_task2.scala'. The 'Assignment24_task2.scala' tab is active, displaying the following Scala code:

```
import scala.util
import java.lang.String

object Assignment24_task2 {
  def main(args : Array[String]) : Unit = {

    // Create Spark Configuration and spark streaming context
    val conf = new SparkConf().setMaster("local[4]").setAppName("StreamTest")
    val sc = new SparkContext(conf)
    sc.setLogLevel("WARN")
    val ssc = new StreamingContext(sc, Seconds(20))

    val offensiveKeywords = Set("idiot", "fool", "stupid", "mad")

    val userLines = ssc.socketTextStream("localhost", 9999)

    println(offensiveKeywords)
    // Print user input lines
    userLines.map(x => println(x))

    val splitedLines = userLines.flatMap(line => line.split(" "))
    val words = splitedLines.map(word => word)
    val filterOffensiveWords = words.filter(x => offensiveKeywords.contains(x.toLowerCase)).map(x => (x, 1))
    val offensiveWordCount = filterOffensiveWords.reduceByKey(_ + _)

    // Print Offensive Word Count
    print("Offensive Word Count : ")
    offensiveWordCount.print()

    ssc.start()
    ssc.awaitTermination()
  }
}
```

Execution:

Input user data in Netcat

```
C:\Windows\System32\cmd.exe - nc64.exe -l -p 9999

D:\TOOLS\netcat-win32-1.12>nc64.exe -l -p 9999
Are you idiot
How stupid are you
You are mad
Are you stupid
You are a stupid
You are a mad
How stupid are you
You are a mad
I am not a stupid
You are a mad
```

Result:

```
new 1 x new 2 x new 3 x new 5 x new 4 x new 6 x new 7 x
78 2019-03-02 22:40:40 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@63192798{/metrics/json,null,AVAILABLE,@Spark}
79 Set(idiot, fool, stupid, mad)
80 Offensive Word Count : 2019-03-02 22:40:48 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
81 2019-03-02 22:40:48 WARN BlockManager:66 - Block input-0-1551546648000 replicated to only 0 peer(s) instead of 1 peers
82 2019-03-02 22:40:53 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
83
84
85 Time: 1551546660000 ms
86
87 (mad,1)
88 (idiot,1)
89 (stupid,1)
90
91 2019-03-02 22:41:08 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
92 2019-03-02 22:41:08 WARN BlockManager:66 - Block input-0-1551546668200 replicated to only 0 peer(s) instead of 1 peers
93 2019-03-02 22:41:12 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
94
95
96 Time: 1551546680000 ms
97
98 (mad,1)
99 (stupid,2)
100
101
102 Time: 1551546700000 ms
103
104
105 2019-03-02 22:41:48 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
106 2019-03-02 22:41:48 WARN BlockManager:66 - Block input-0-1551546708000 replicated to only 0 peer(s) instead of 1 peers
107 2019-03-02 22:41:52 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
108
109 Time: 1551546720000 ms
110
111 (mad,1)
112 (stupid,2)
113
114 2019-03-02 22:42:02 WARN RandomBlockReplicationPolicy:66 - Expecting 1 replicas with only 0 peer/s.
115 2019-03-02 22:42:02 WARN BlockManager:66 - Block input-0-1551546722400 replicated to only 0 peer(s) instead of 1 peers
116
117 Time: 1551546740000 ms
118
119 (mad,1)
120
```