**Assignment 48**

**Task 1**

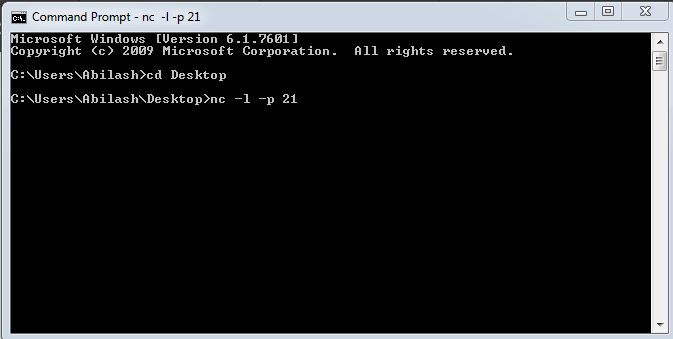
**Read a stream of Strings; fetch the words which can be converted to numbers. Filter out the rows, where the sum of numbers in that line is odd. Provide the sum of all the remaining numbers in that batch.**

Solution:

To read a stream, we need to create localhost on a port and read that data in the stream. To create localhost netcat is used.

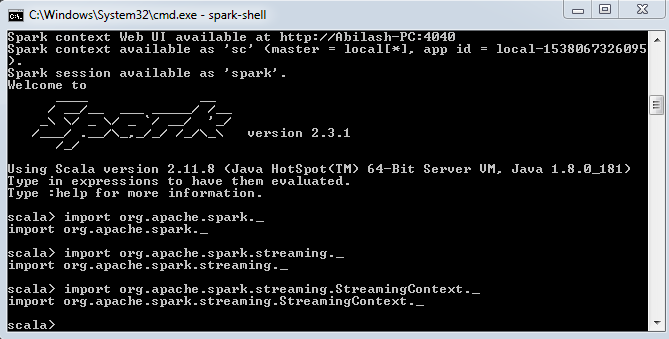
After installing netcat of windows below command is called to create a local host on port 21

***nc –l –p 21***



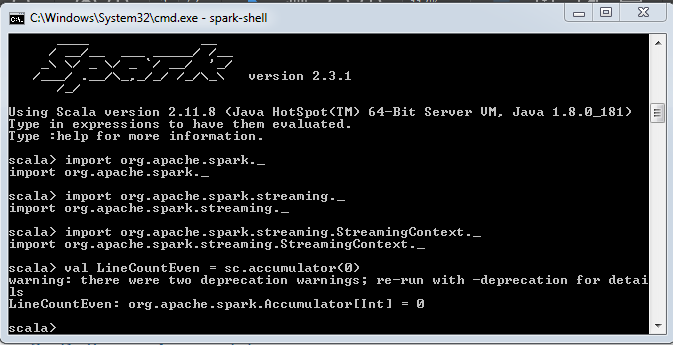
**STEP-1**

Import all Streaming Packages..



**STEP-2**

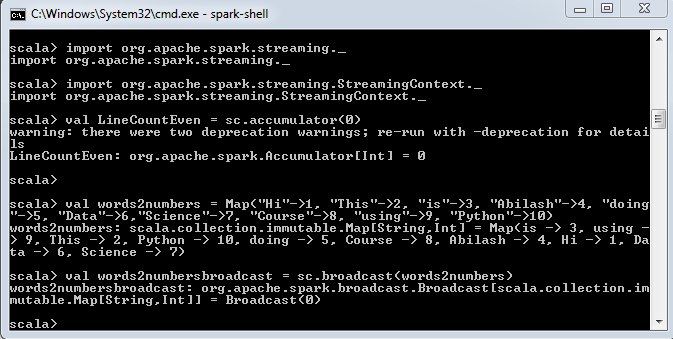
Create an accumulator “LineCountEven” which keeps track of sum of words when equal to Even number.



**STEP-3**

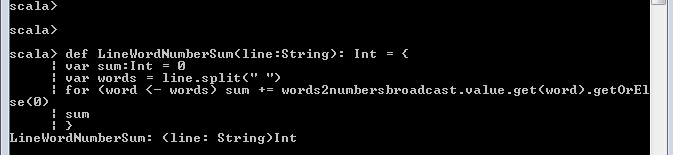
Create and RDD **“words2numbers”** which maps Strings to corresponding Numbers, any other words other than in the mapping will return 0.

Store the Broadcasted **words2numbers** to **words2numbersbroadcast**



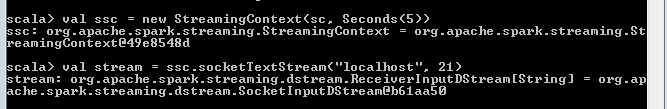
**STEP-4**

Create a function “LineWordNumberSum” where we are splitting a line based on blank space to get all the words in next. In the lookup value, we are determining corresponding numbers for a word in the words2numbersbroadcast and we adding the all the numbers.



**STEP-5**

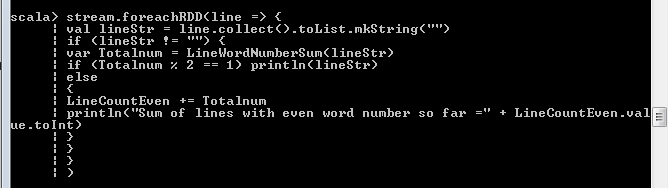
In this step, we are streaming the data as a string in a 5 seconds interval and return the stream. The streams are reading in a port 9999 which is listened



**STEP-6**

Process each RDD in stream, we are converting the RDD to string.

If RDD is not blank calculate word’s number and sum them using the function LineWordNumberSum and put as variable Totalnum. If Totalnum is odd, print the provided line in the output, else add Totalnum to accumulator LineCountEven and print the sum.



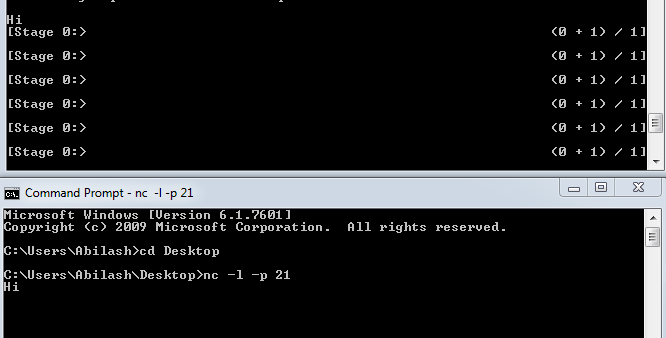
STEP-7

Start the Stream

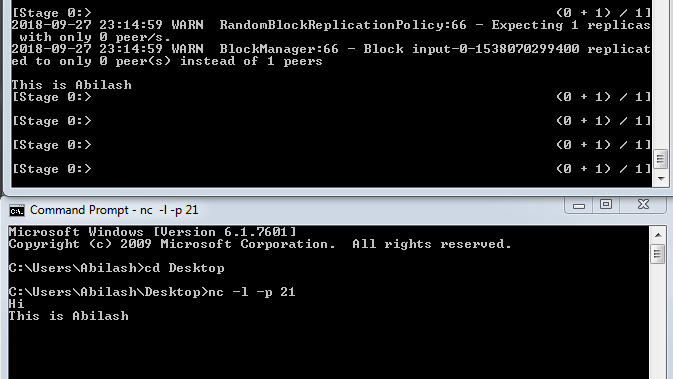


**OUTPUT**

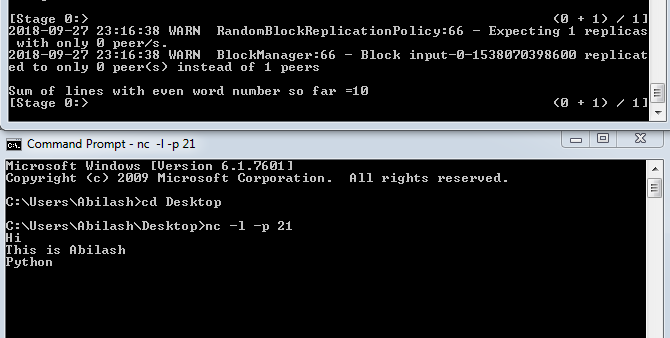
Here “Hi” is entered in the Stream and the Output Displayed is “Hi”, because the map of value “Hi” is 1 which is ODD number



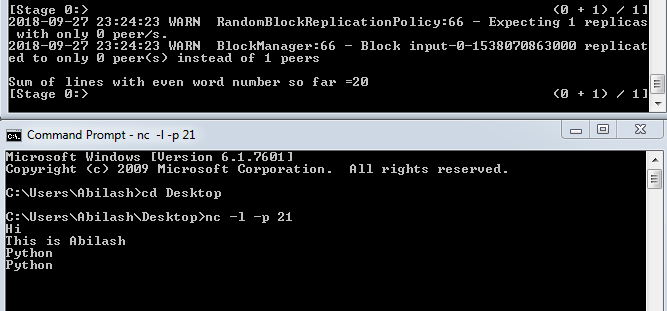
Here “This is Abilash” is entered in the Stream and the Output Displayed is “This is Abilash”, because the map of value (2+3+4 =9) is 9 which is ODD number



Here “Python” is entered in the Stream and the Output Displayed is “Sum of lines with even word count so fat = 10”, because the lines entered with sum of even is only this line Python and its map value is 10



Entering “Python” again will give total number of even counted lines.



**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.

**Step-1**

Import the Streaming Context – netcat has been used as explained in the above task.

**import org.apache.spark.\_**

**import org.apache.spark.streaming.\_**

**import org.apache.spark.streaming.StreamingContext.\_**

**Step-2**

Declare an **OffensiveWord** string Variable, **wordcount** to count offensive word entered by users, and create two Stream with Port number 9999 and 21.

9999: for Real Time Offensive String

21: for the User input

**var OffensiveWord:String = ""**

**val wordcount = sc.accumulator(0)**

**val ssc = new StreamingContext(sc, Seconds(5))**

**val stream1 = ssc.socketTextStream("localhost", 9999)**

**val stream2 = ssc.socketTextStream("localhost", 21)**

**Step-3**

For stream1 read the offensive words and save it in the variable

**stream1.foreachRDD(line => {val lineStr = line.collect().toList.mkString("")**

**if(lineStr!="") {OffensiveWord += lineStr**

**println("STREAM-9999: These are list of Real Time OffensiveWords " + OffensiveWord)}**

**})**

For stream2 read the offsensive words and display user of the offensive words entered, and also count and display total values entered.

**stream2.foreachRDD(line => {val lineStr = line.collect().toList.mkString("")**

**if(lineStr!="") {**

**for(word <- OffensiveWord.split(",")){**

**if(lineStr==word) {println("STREAM-21:" + lineStr + " YOU HAVE ENTERED AN OFFENSIVE WORD!!!")**

**wordcount +=1**

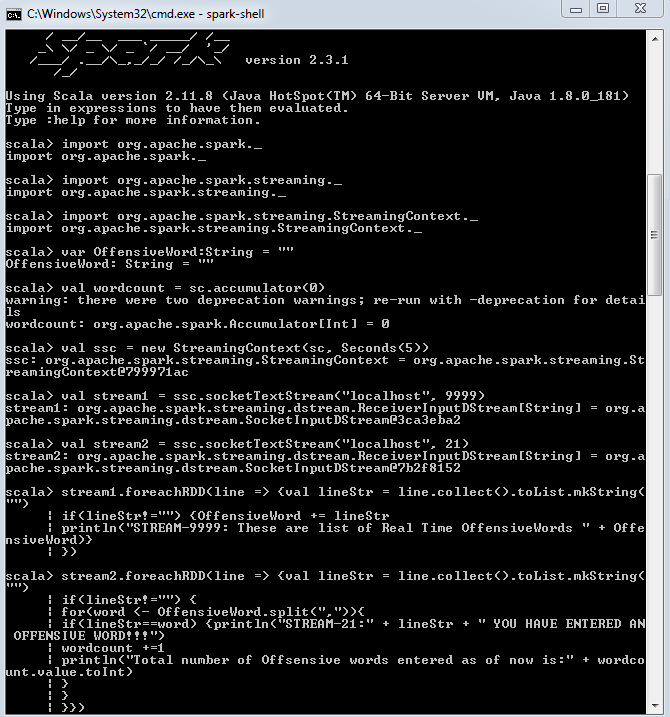
**println("Total number of Offsensive words entered as of now is:" + wordcount.value.toInt)}}}})**

**Step-4**

Start the Stream

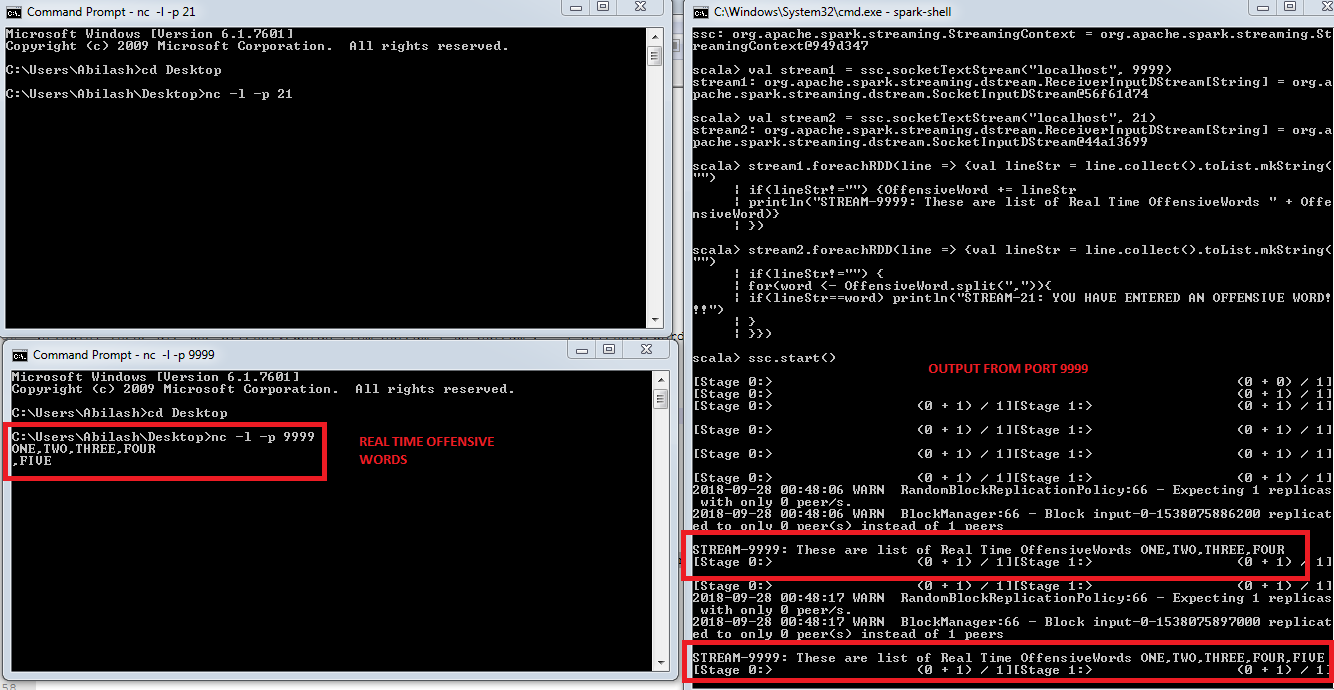
**ssc.start()**

**ENTIRE-CODE**

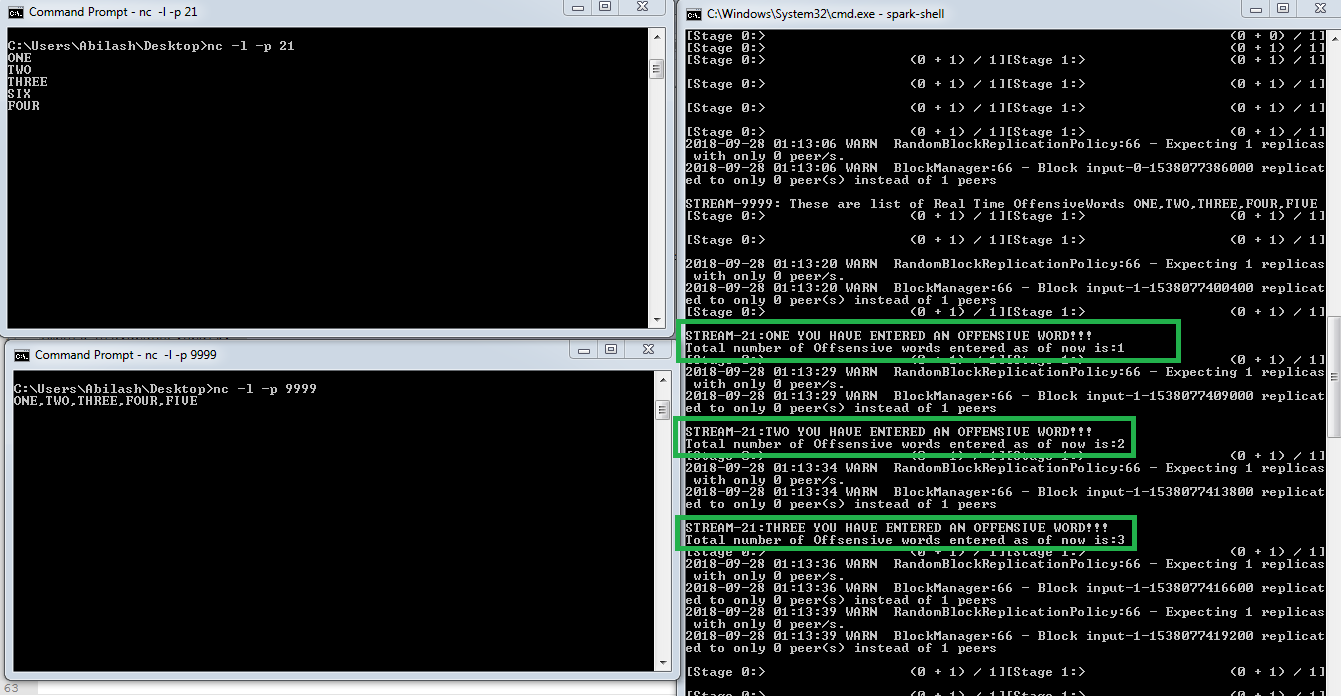


**OUTPUT**

Real time Offensive words are entered in Port 9999 as shown below, and the values entered in this port is stored to the variable and that can be seen in the Output.

****

Users would enter the words in the Port 21, and the words entered in this is checked against the Offensive words list and displays a message as shown below, also number of offensive words user entered would be displayed as well, see the results below

****