

ASSIGNMENT 24 Spark Streaming

To solve the both the tasks, we create the spark application, which has two Scala files with object Even_Number_Line.scala and Offensive_Words_Count.scala.

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.

Below screen shot shows the spark application to filter the lines containing even numbers

```
import org.apache.spark.{SparkConf, SparkContext}
import org.apache.spark.streaming.{Seconds, StreamingContext}
object Even_Number_Line{
  def main(args: Array[String]):Unit = {

    // create a function which
    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 => {}
        }
      }
      return number
    }

    println("This is the task1 of assignment session 26")
  }
}
```

```

println("This is the task1 of assignment session 26")

val conf = new SparkConf().setMaster("local[2]").setAppName("EvenLines")
val sc = new SparkContext(conf)

sc.setLogLevel("WARN")
println("Spark Context Created")

// Create a local StreamingContext with working thread and batch interval of 20 seconds
val ssc = new StreamingContext(sc, Seconds(30))

println("Spark Streaming Context Created")

// Create a DStream that will connect to hostname:port,localhost:9999
val lines = ssc.socketTextStream( hostname="localhost", port= 9999)

//filter the even string from input line by using Get_Lines function
val lines_filter = lines.filter(x => Get_Lines_Sum(x)%2 == 0)

//add all the numbers the even string from input line by using Get_Lines function
val lines_sum = lines_filter.map(x => Get_Lines_Sum(x))

```

```

44
45 //filter the even string from input line by using Get_Lines function
46 val lines_filter = lines.filter(x => Get_Lines_Sum(x)%2 == 0)
47
48 //add all the numbers the even string from input line by using Get_Lines function
49 val lines_sum = lines_filter.map(x => Get_Lines_Sum(x))
50
51 println("Lines with even sum:")
52 lines_filter.print()
53
54 println("Sum of the numbers in even lines:")
55 lines_sum.reduce(_+_).print()
56
57 // Start the computation
58 ssc.start()
59
60 // Wait for the computation to terminate
61 ssc.awaitTermination()
62 }
63
64

```

Writable Smart Insert 51 : 36

eclipse-workspace - sp...

Net cat command to send the stream data

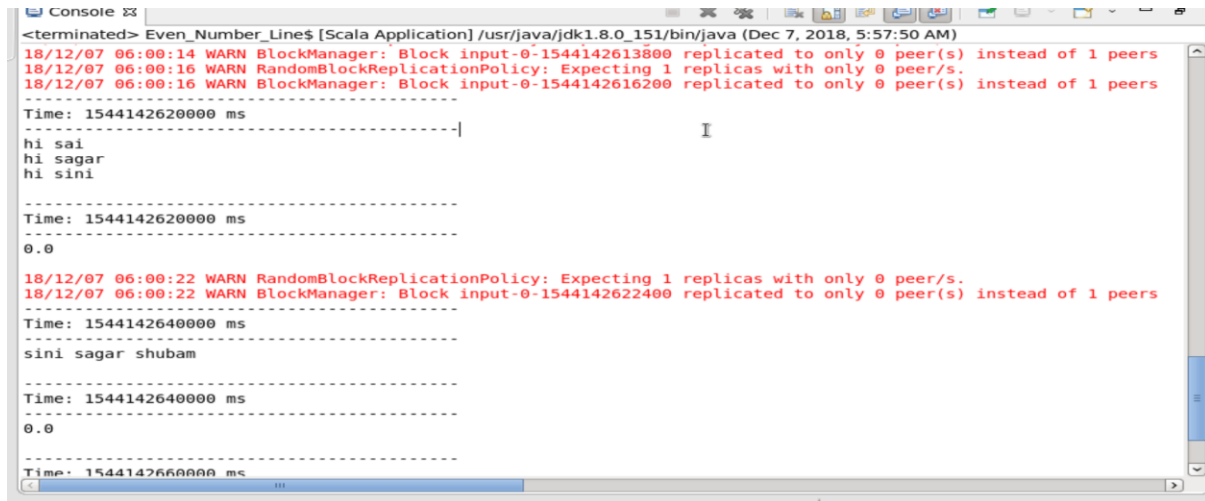
```

[acadgild@localhost Desktop]$ nc -lk 9999
hi sai
hi sagar
hi sini
sini sagar shubam
^C
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$

```

Support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Output of console



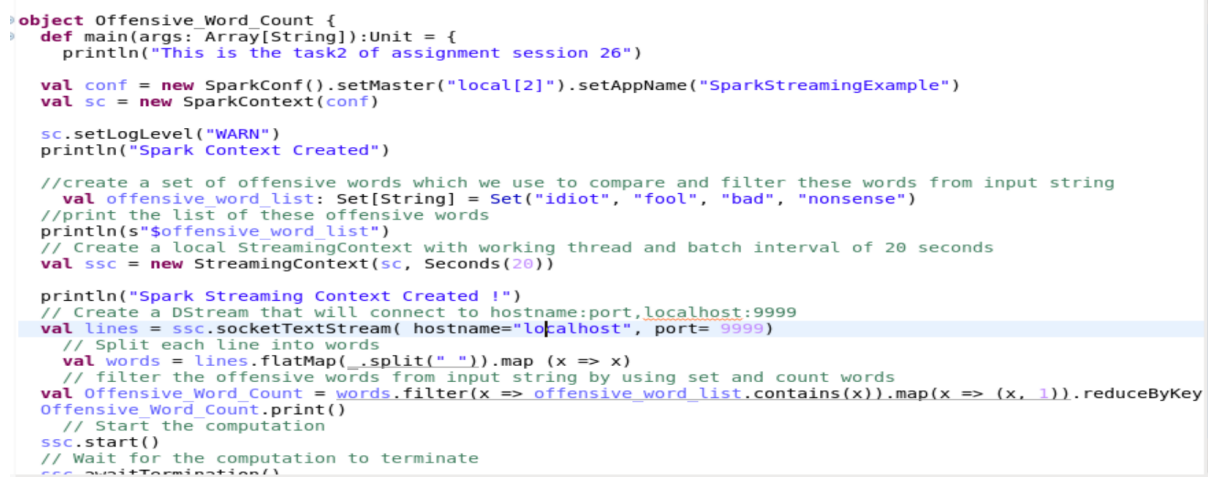
```
<terminated> Even_Number_Line$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Dec 7, 2018, 5:57:50 AM)
18/12/07 06:00:14 WARN BlockManager: Block input-0-1544142613800 replicated to only 0 peer(s) instead of 1 peers
18/12/07 06:00:16 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/12/07 06:00:16 WARN BlockManager: Block input-0-1544142616200 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1544142620000 ms
-----
hi sai
hi sagar
hi sini
-----
Time: 1544142620000 ms
-----
0.0
-----
18/12/07 06:00:22 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/12/07 06:00:22 WARN BlockManager: Block input-0-1544142622400 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1544142640000 ms
-----
sini sagar shubam
-----
Time: 1544142640000 ms
-----
0.0
-----
Time: 1544142660000 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 real-time set of offensive words

Below screen shot shows the spark application to filter the offensive words from input string entered by use



```
object Offensive_Word_Count {
  def main(args: Array[String]): Unit = {
    println("This is the task2 of assignment session 26")

    val conf = new SparkConf().setMaster("local[2]").setAppName("SparkStreamingExample")
    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 words from input string
    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 20 seconds
    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(hostname="localhost", port= 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 words
    val Offensive_Word_Count = words.filter(x => offensive_word_list.contains(x)).map(x => (x, 1)).reduceByKey(
    Offensive_Word_Count.print()
    // Start the computation
    ssc.start()
    // Wait for the computation to terminate
    ssc.awaitTermination()
  }
}
```

Netcat command to send live stream data

In the above spark application, we have a set of words that we considered as offensive words.

“idiot”, “fool”, “bad”, “nonsense”

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ nc -lk 9999
idiot
fol fool sagar
bad bad
```

Output Console

```
<terminated> Offensive_Word_Count$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Dec 7, 2018, 6:07:25 AM)
Time: 1544143080000 ms
-----
(idiot,1)
-----
Time: 1544143100000 ms
-----
18/12/07 06:08:26 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/12/07 06:08:26 WARN BlockManager: Block input-0-1544143105800 replicated to only 0 peer(s) instead of 1 peers
18/12/07 06:08:30 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/12/07 06:08:30 WARN BlockManager: Block input-0-1544143110600 replicated to only 0 peer(s) instead of 1 peers
18/12/07 06:08:37 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/12/07 06:08:37 WARN BlockManager: Block input-0-1544143117200 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1544143120000 ms
-----
(fool,2)|
(idiot,1)
-----
18/12/07 06:08:43 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/12/07 06:08:43 WARN BlockManager: Block input-0-1544143123600 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1544143140000 ms
-----
(bad,2)
-----
```