SPARK STREAMING ASSIGNMENT 22.1

===

Problem Statement

//Running below code in sparkshell

// Create a local StreamingContext with batch interval of 10 second

/* Create a DStream that will connect to hostname and port, like localhost 9999. As stated earlier, DStream will get created from StreamContext, which in return is created from SparkContext. */

// Using this DStream (lines) we will perform transformation or output operation.

// Start the computation

// Wait for the computation to terminate

//CTRL C at spark-shell will terminate streaming

Input Commands for STATELESS SPARK STREAMING:

//Install below software

```
sudo yum install nc.x86_64
```

```
nc -lk 9999
```

//Invoking spark shell for streaming

```
[acadgild@localhost ~]$ spark-shell --master local[4]
```

//Importing the spark packages for streaming

```
Scala> import org.apache.spark._
```

```
Scala> import org.apache.spark.streaming._
```

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

//Declaring and defining the Stream Context object with the help of SPARK CONTEXT object sc and defining batch interval as 10 sec

```
Scala> val ssc = new StreamingContext(sc, Seconds(10))
```

//Defining the streaming source with the help of Spark Streaming object and pass the localhost details and port number as arguments.

```
Scala> val lines = ssc.socketTextStream("localhost", 9999)
```

//performing flatmap operation to flatten the input data

```
Scala> val words = lines.flatMap(_.split(" "))
```

//performing map operation to generate value 1 for each key/word and

```
Scala> val pairs = words.map(word => (word, 1))
```

//perform reduce operation to sum up the all the values of each key pair post successful map operation

```
Scala> val wordcounts = pairs.reduceByKey(_ + _)
```

//printing the word count for each word streamed from localhost 9999 port nr.

```
Scala> wordcounts.print()
```

//Start the computation

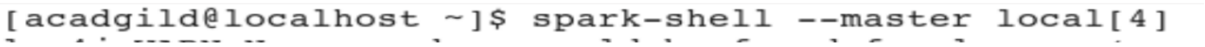
```
Scala> ssc.start()
```

//Wait for the computation to terminate

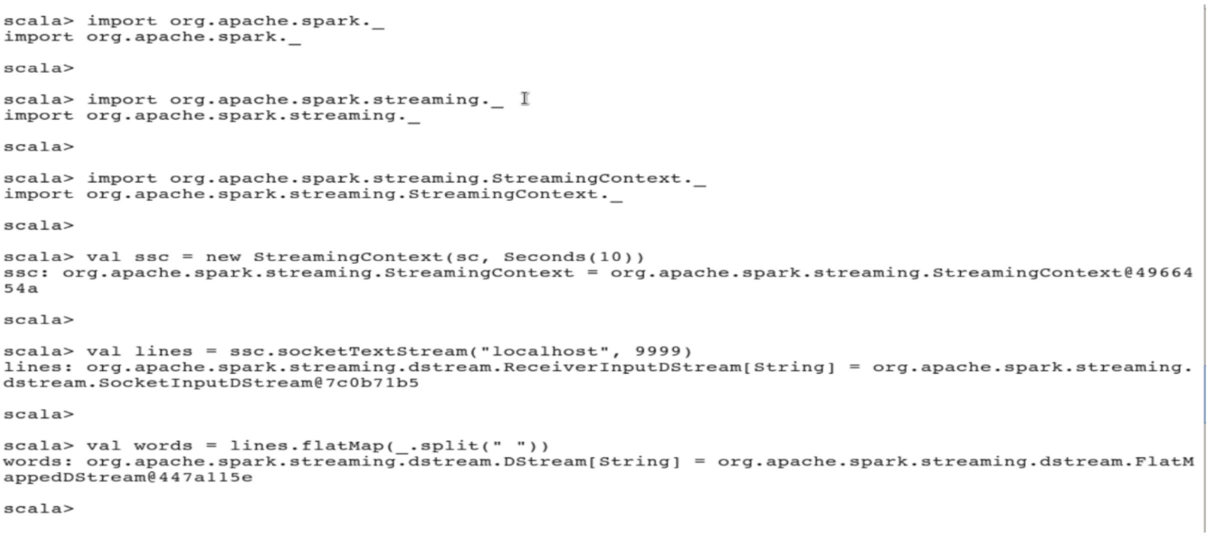
```
Scala> ssc.awaitTermination()
```

===

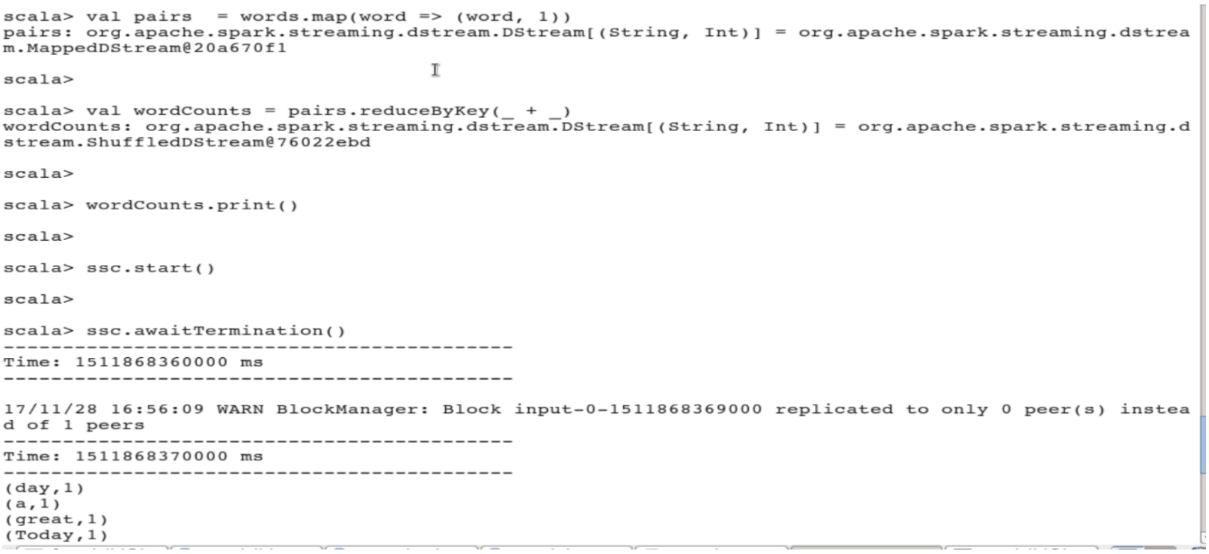
Screenshots:



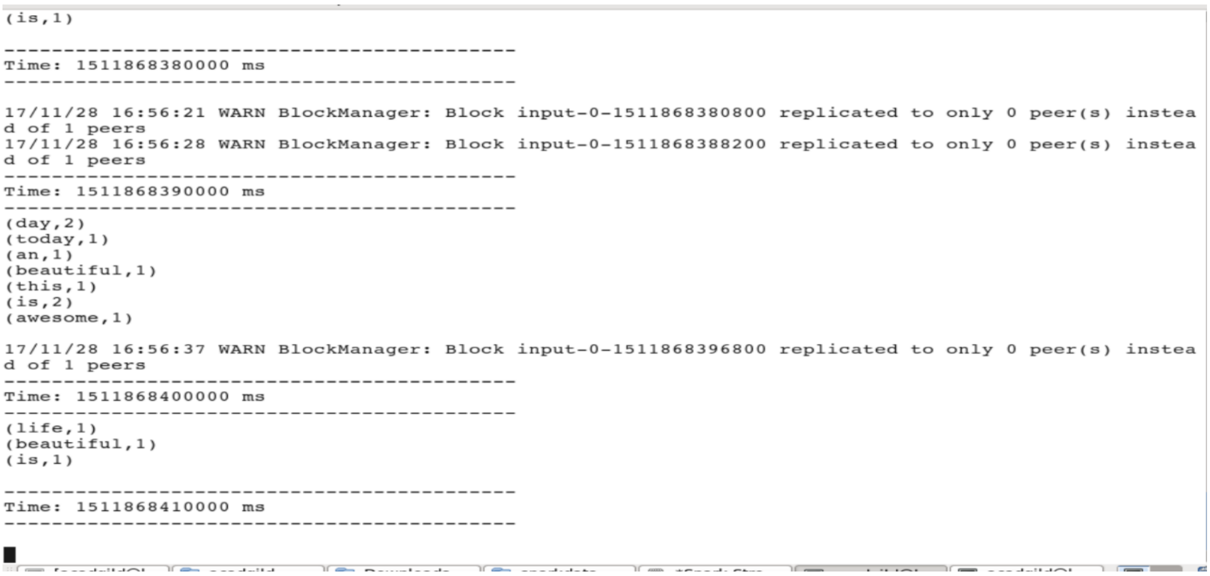
```
[acadgild@localhost ~]$ spark-shell --master local[4]
```



```
scala> import org.apache.spark._
import org.apache.spark._

scala>

scala> import org.apache.spark.streaming._
import org.apache.spark.streaming._

scala>

scala> import org.apache.spark.streaming.StreamingContext._
import org.apache.spark.streaming.StreamingContext._

scala>

scala> val ssc = new StreamingContext(sc, Seconds(10))
ssc: org.apache.spark.streaming.StreamingContext = org.apache.spark.streaming.StreamingContext@49664
54a

scala>

scala> val lines = ssc.socketTextStream("localhost", 9999)
lines: org.apache.spark.streaming.dstream.ReceiverInputDStream[String] = org.apache.spark.streaming.
dstream.SocketInputDStream@7c0b71b5

scala>

scala> val words = lines.flatMap(_.split(" "))
words: org.apache.spark.streaming.dstream.DStream[String] = org.apache.spark.streaming.dstream.FlatM
appedDStream@447a115e

scala>
```

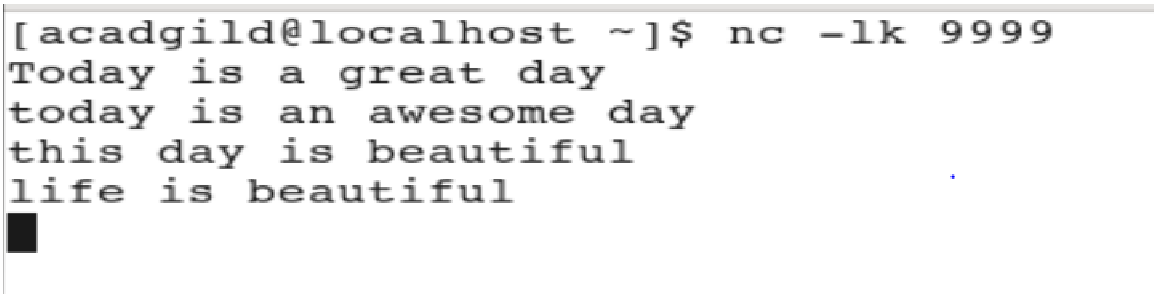


```
scala> val pairs  = words.map(word => (word, 1))
pairs: org.apache.spark.streaming.dstream.DStream[(String, Int)] = org.apache.spark.streaming.dstrea
m.MappedDStream@20a670f1

scala>

scala> val wordCounts = pairs.reduceByKey(_ + _)
wordCounts: org.apache.spark.streaming.dstream.DStream[(String, Int)] = org.apache.spark.streaming.d
stream.ShuffledDStream@76022ebd

scala>

scala> wordCounts.print()

scala>

scala> ssc.start()

scala>

scala> ssc.awaitTermination()
-------------------------------------------
Time: 1511868360000 ms
-------------------------------------------

17/11/28 16:56:09 WARN BlockManager: Block input-0-1511868369000 replicated to only 0 peer(s) instea
d of 1 peers
-------------------------------------------
Time: 1511868370000 ms
-------------------------------------------
(day,1)
(a,1)
(great,1)
(Today,1)
```



```
(is,1)

-------------------------------------------
Time: 1511868380000 ms
-------------------------------------------

17/11/28 16:56:21 WARN BlockManager: Block input-0-1511868380800 replicated to only 0 peer(s) instea
d of 1 peers
17/11/28 16:56:28 WARN BlockManager: Block input-0-1511868388200 replicated to only 0 peer(s) instea
d of 1 peers
-------------------------------------------
Time: 1511868390000 ms
-------------------------------------------
(day,2)
(today,1)
(an,1)
(beautiful,1)
(this,1)
(is,2)
(awesome,1)

17/11/28 16:56:37 WARN BlockManager: Block input-0-1511868396800 replicated to only 0 peer(s) instea
d of 1 peers
-------------------------------------------
Time: 1511868400000 ms
-------------------------------------------
(life,1)
(beautiful,1)
(is,1)

-------------------------------------------
Time: 1511868410000 ms
-------------------------------------------

█
```

Streaming Input Source Screenshot:



```
[acadgild@localhost ~]$ nc -lk 9999
Today is a great day
today is an awesome day
this day is beautiful
life is beautiful
```