Lab sheet 4 – Spark Streaming

Name: J Viswaksena **RollNo:** AM.EN.U4AIE21035

Q1) Count the number of words in text data received from a data server listening on a TCP socket.

(i.e converting a stream of lines to words)

Result should be total number of words.

Code:

```
from pyspark import SparkContext
from pyspark.streaming import StreamingContext

sc = SparkContext("local[2]", "NetworkWordCount")
ssc = StreamingContext(sc, 1)

lines = ssc.socketTextStream("localhost", 9999)

words = lines.flatMap(lambda line: line.split(" "))

word_pairs = words.map(lambda word: (word, 1))
word_counts = word_pairs.reduceByKey(lambda x, y: x + y)
pairs = words.map(lambda word: (word, 1))

wordCounts = pairs.reduceByKey(lambda x, y: x + y)
total_word_count = word_counts.reduce(lambda x, y: ("Total", x[1] + y[1]))
wordCounts.pprint()
total_word_count.pprint()

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

Input:

```
(base) viswaksenajayam@Viswaksenas-MacBook-Pro ~ % ncat -l -p 9999
hi How are you
```

Output:

```
-------------------------------
('How', 1)
('are', 1)
('you', 1)
('hi', 1)

-------------------------------
Time: 2024-01-07 20:20:13
-------------------------------
('Total', 4)
```

Q2) Reverse the words in a sentence which is received from a data server listening on a TCP socket.

Code:

```
from pyspark import SparkContext
from pyspark.streaming import StreamingContext

def reverse_words(sentence):
    words = sentence.split(" ")
    reversed_sentence = ' '.join(reversed(words))
    return reversed_sentence

sc = SparkContext("local[2]", "ReverseWordCount")
ssc = StreamingContext(sc, 1)

lines = ssc.socketTextStream("localhost", 9999)

reversed_lines = lines.map(reverse_words)
reversed_lines.pprint()

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

Input:

```
(base) viswaksenajayam@V
My name is Viswak sena
```

Output:

```
-------------------------------------------
sena Viswak is name My
```

Result :

Input: *s = "i love programming very much"*
Output: *s = "much very programming love i"*