

Assignment - 10

Established EMR connection

[illegible]

4)

```
C:\> Users > suman > OneDrive > Desktop > ass10 > consume.py
1  from pyspark import SparkContext
2  from pyspark.streaming import StreamingContext
3
4  # Create a local StreamingContext with a batch interval of 10 seconds
5  sc = SparkContext("yarn", "NetworkWordCount")
6  ssc = StreamingContext(sc, 10)
7
8  # Create a DStream
9  lines = ssc.socketTextStream("ec2-3-145-37-239.us-east-2.compute.amazonaws.com", 3333)
10
11 # Split each line into words
12 words = lines.flatMap(lambda line: line.split(" "))
13
14 # Count each word in each batch
15 pairs = words.map(lambda word: (word, 1))
16 wordCounts = pairs.reduceByKey(lambda x, y: x + y)
17
18 # Print each batch
19 wordCounts.pprint()
20
21 ssc.start() # Start the computation
22 ssc.awaitTermination() # Wait for the computation to terminate
23
24
```

5,6)

```
suman@Sumanth MINGW64 ~/OneDrive/Desktop/ass10
$ scp -i keypair.pem consume.py hadoop@ec2-3-145-37-239.us-east-2.compute.amazonaws.com:/home/hadoop
consume.py                                100% 697      11.8KB/s   00:00

suman@Sumanth MINGW64 ~/OneDrive/Desktop/ass10
$ scp -i keypair.pem log4j.properties hadoop@ec2-3-145-37-239.us-east-2.compute.amazonaws.com:/home/hadoop
log4j.properties                          100% 3199     88.0KB/s   00:00
```

8,9)

```
[hadoop@ip-172-31-10-195 ~]$ sudo cp ./log4j.properties /etc/spark/conf/log4j.properties
[hadoop@ip-172-31-10-195 ~]$ nc -lk 3333
```

10)

```
[hadoop@ip-172-31-10-195 ~]$ nc -lk 3333
MY NAME IS LAVANYA AND MY SCHOOL NAME IS IIT.
MY NAME IS LAVANYA AND MY SCHOOL NAME IS IIT.
```

```
at java.lang.Thread.run(Thread.java:750) ~[?:1.8.0_392]
[hadoop@ip-172-31-10-195 ~]$ spark-submit consume.py
23/11/13 19:54:08 INFO SparkContext: Running Spark version 3.4.0-amzn-0
23/11/13 19:54:08 INFO ResourceUtils: =====
23/11/13 19:54:08 INFO ResourceUtils: No custom resources configured for spark.driver.
23/11/13 19:54:08 INFO ResourceUtils: =====
23/11/13 19:54:08 INFO SparkContext: Submitted application: NetworkWordCount
23/11/13 19:54:08 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 4, script: , vendor: , memory -> name: memory
, amount: 9486, script: , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
23/11/13 19:54:08 INFO ResourceProfile: Limiting resource is cpus at 4 tasks per executor
23/11/13 19:54:08 INFO ResourceProfileManager: Added ResourceProfile id: 0
23/11/13 19:54:08 INFO SecurityManager: Changing view acls to: hadoop
23/11/13 19:54:08 INFO SecurityManager: Changing modify acls to: hadoop
23/11/13 19:54:08 INFO SecurityManager: Changing view acls groups to:
23/11/13 19:54:08 INFO SecurityManager: Changing modify acls groups to:
23/11/13 19:54:08 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: hadoop; groups with view permissions: EMPTY
; users with modify permissions: hadoop; groups with modify permissions: EMPTY
23/11/13 19:54:09 INFO Utils: Successfully started service 'sparkDriver' on port 36165.
```

```
-----
Time: 2023-11-13 19:57:20
-----
('NAME', 2)
('IS', 2)
('SCHOOL', 1)
('IIT.', 1)
('MY', 2)
('LAVANYA', 1)
('AND', 1)

23/11/13 19:57:20 INFO JobScheduler: Finished job streaming job 1699905440000 ms.0 from job set of time 1699905440000 ms
23/11/13 19:57:20 INFO JobScheduler: Total delay: 0.357 s for time 1699905440000 ms (execution: 0.324 s)
```

```
MY NAME IS LAVANYA AND MY SCHOOL
this is a test of the the system
```

```
23/11/13 20:04:30 INFO DAGScheduler: Job 121 finished: runJob at PythonRDD.scala:179, took 0.073789 s
-----
Time: 2023-11-13 20:04:30
-----
('this', 1)
('is', 1)
('test', 1)
('of', 1)
('a', 1)
('the', 2)
('system', 1)
23/11/13 20:04:30 INFO JobScheduler: Finished job streaming job 1699905870000 ms.0 from job set of time 1699905870000 ms
23/11/13 20:04:30 INFO JobScheduler: Total delay: 0.296 s for time 1699905870000 ms (execution: 0.259 s)
```

- Sailavanya Narthu
- A20516764
- snarthu@hawk.iit.edu